



DATA SCIENCE AND ANALYTICS DISSERTATION

Classification of bow shock and magnetopause event positions by using the Magnetic Field properties of Saturn

A thesis submitted in fulfillment of the requirement for the degree of M.Sc. in Data Science and Analytics.

Author: Mathews Philip Venattu (20250487)

Supervisor: Dr. Katarina Domijan

Department of Mathematics and Statistics
University: National University of Ireland, Maynooth.

Date Submitted: 10th July, 2021

Abstract

Cassini-Huygens Space Research mission was a joint collaboration of NASA, ESA and Italian Space agency to study about the planet Saturn and its system that includes its Rings and Natural Satellites. During its life span of about 20 years the spacecraft orbited the planet Saturn for 13 years and also frequently changed its shape and tilt. Because of this frequent changes in the orbital trajectory of spacecraft, it passed multiple times through the boundaries of Bowshock and Magnetopause at different latitude, longitude and phases of Solar Cycle. In this Project, based on the Magnetic Field data collected during the year 2005 from MAG (Magnetometer) instrument attached to the spacecraft was used to Classify the different boundary crossings of CASSINI Spacecraft. MAG (Magnetometer) in Cassini recorded the strength and direction of the Magnetic field around the spacecraft. Different Statistical models were used for the classification and those models were carefully examined to get more insights about magnetic field properties during different events. In this project, different trial and error approaches were taken to create a training dataset that gives better test accuracy on test dataset.

Contents

List of tables	1
List of figures	2
1 Introduction	3
2 Background	5
3 Data	6
3.1 Dataset 1: Magnetopause And Bowshock Crossing List	6
3.2 Dataset 2: Magnetometer Dataset	7
3.3 Dataset 3: Combined Data Of Dataset 1 And Dataset 2	8
4 Data Manipulation Section	10
4.1 Removing Time Dependency	10
4.2 Data Imputation	10
5 Exploratory Data Analysis	11
6 References	13
7 Appendix	14
7.1 Supporting code	14

List of tables

List of figures

1	Line Graph of Total Magnetic Field Strength between 72 and 74th day of the year 2005	11
2	Line Graph of Total Magnetic Field Strength between 136 and 138th day of the year 2005	11

1 Introduction

All planets in our solar system have magnetic field like earth. Some of the planets like Uranus, Saturn, Jupiter, and Neptune has large magnetic field than earth. Magnetosphere of an astronomical object is the area surrounding that object where its Magnetic field is present. Like all other planets Saturn also has the similar magnetosphere structure - it has a Bow shock, Magneto sheath, Magneto pause and Magneto tail. Among this structure bow shock is the point at which the Magnetosphere of the Saturn interacts with the solar wind that in turn suddenly reduces its speed and pressure. Magneto pause is the boundary between Saturn's Magnetic field and Solar wind. The magneto sheath exists between the bow shock and the magneto pause, an area of shocked solar wind that is significantly influenced by the changes that occur within the bow shock and whose features can impact the interaction across the magneto pause. The dynamic pressure of Solar wind usually determines boundary and position of Magneto pause and Bow shock [2]. The Cassini Huygens Mission is a joint NASA/ESA/ASI project to make a detailed survey of the ringed planet Saturn and its natural satellites. Cassini spacecraft recorded magnetic field and plasma condition of the environment during its insertion to Saturn's orbit by using the Cassini Magnetometer (MAG), Plasma Spectrometer (CAPS). The magnetic field strength pattern clearly shows some "overshoot" and "foot" when the spacecraft passed through the boundaries of Magnetopause, Bow shock and Magneto sheath [3]. The Magnetometer (MAG) which continuously acquired magnetic field data that is associated with the Plasma Environment and internal source of Saturn are essential to study about the interaction between solar wind and Magnetosphere of Saturn.

Magnetosphere of Saturn act as an obstacle to the Solar wind far away from the planet and the magnetopause of the planet Saturn lies in 20 Rs (Radius of Saturn=60330 km). So, the Solar wind which interact with the magnetosphere is decelerated by the bow shock and the physical properties of the plasma (Sergis (2013)). The magneto disk pressure, which inflates the equatorial magnetosphere considerably more than the high-latitude magnetosphere, affects the geometry of the magnetopause boundary itself, resulting in clear polar flattening [5]. In the case of Earth, basic pressure balance is due to the interaction between upstream solar wind flow against magnetic pressure from the magnetosphere and this helps to draw the approximate location of magnetopause boundary but when it comes to Saturn we must consider the influence of the natural satellite Enceladus, which serves as a huge internal plasma source. The pressure related with the super thermal component of this internally produced plasma serves to inflate the magnetosphere considerably beyond what a basic magnetic pressure calculation would predict [6]. So even in a steady solar wind conditions, Due to the internal plasma mentioned above the magneto pause boundary of Saturn may move significantly.

Lots of studies were done based on the data acquired from the Cassini Spacecraft instruments. (Sergis (2013)) they chose intervals between 2004 and 2011 when the spacecraft was travelling through the magnetic sheath and used the data recorded to explore different properties like plasma, energetic particle, magnetic field density, temperature etc. They concentrated more on the presence of W⁺s ions (water group) and explained the ejection of energetic ions as a function of pitch angle and energy which shows the often flow of heavy energetic ions from bow shock. Analysis of CAPS (Cassini Plasma Spectrometer) by Burkholder et. al [8] shown the significant

ion flow differences prenoon and post noon, and apart from the magnetic field data was used to illustrate the rotation of IMF (Inter Planetary Magnetic Field) vector.

Based on literature survey, I got information regarding the different variables in the dataset like Magnetic field, direction of cross, position of the spacecraft etc. In this project we are only considering the bow shock and magnetopause crossing of the spacecraft during the year 2005. This report explains about the entire project with different sections like Background, Dataset explanation, Data Manipulation, Data Visualization, Models trained to classify the type of crossings, R code used, summary of the results and conclusion. Background section explains about the different terms used and current approaches done by others for classification. In Dataset explanation section it explains about each variable and the different datasets that were used for this project. Data Manipulation section gives a detailed explanation about the transformations and imputations done on the dataset. New dataset made after data manipulation is explained visually on the data visualization section. There are different models tried to achieve better results each model that are used for this project is explained on this part of the report. Remaining portions explains about the code and its results.

2 Background

In the early days, Scientists have very less information regarding the Planet Saturn and its magnetosphere because as we all know that the magnetic fields are invisible, and it needs to be studied from inside. Cassini Huygens mission was a great opportunity for the science world to explore the planet and its behaviour. This mission helped to study the magnetic field and the flow of different gases under the influence of the magnetic field which affects the auroras of Saturn. This mission given some powerful insights about the atmosphere and the surrounding of Planet Saturn. By comparing Saturn with the similar exoplanets will give information regarding the evolution of the Solar System. Different studies were done based on the data gathered from the Cassini spacecraft. Based on this data [9] demonstrated that polar flattening of the magnetopause causes shorter streamline pathways over the poles, resulting in a higher-pressure gradient, which twists the field. This in turn leads to different conditions at the magnetopause when compared to those predicated based on axisymmetric assumptions. From 2004 day 299 through 2012 day 151, a substantial data was compiled by [10] of magneto sheath measurements was collected using data from CAPS, MAG, and MIMI. This data collection enables researchers to investigate things like local temporal dependence of magneto sheath parameters. They also demonstrated a new method for estimating upstream solar wind speed using the same magneto sheath parameters. [11] used the MAG data for research which provides a broad picture of low-frequency waves in Saturn's magnetosphere, which has crucial consequences for how magnetospheric energy leaks.

Both the Bow shock and magneto pause models can be used as a significant tool which gives insights about the solar upstream conditions and its dynamic pressure at which they are associated. All the data associated with Cassini uses KSM coordinate system and this system is Saturn centred where the x-axis is towards the sun [13]. Orbital tour of Cassini around Saturn which started in the month of July 2004 during that time the spacecraft crossed 100 Bow Shock boundaries. A study done by [12] On 11th and 12th of April 2005, Cassini magnetometer readings were made during a typical sequence of Cassini bow shock crossings. The spacecraft began and finished the period downstream of the shock in the magneto sheath solar wind, with two trips into the upstream solar wind, each separated by two shock crossings. The presence of obvious shock ramps and a constant upstream field indicates that these are quasi-perpendicular crossings. During this time magnetic field strength values recorded by the magnetometer were so high. In this report I analysed the magnetometer and position data of Cassini spacecraft during the year 2005 to classify the Magnetopause and Bow Shock events.

3 Data

Cassini Spacecraft orbited around the Saturn for about 9 years. During this period the spacecraft transmitted valuable information regarding Saturn like the magnetic field strength, position at which it was measured to earth. Magnetometer and CAPS are the main instruments that were used for measuring the magnetic field strength and Kinetic Energy of particles at each point. For this project, I am only considering the data that was recorded during the year 2005 by the spacecraft. Mainly two datasets were used in this project to make a final combined useful dataset, first dataset contains a list of Bow Shock and Magneto pause event crossings that occurred during the year 2005 (Jackman et. al, 2019). The second dataset contains the information regarding the position of spacecraft and the vector data of Magnetic field strength.

3.1 Dataset 1: Magnetopause And Bowshock Crossing List

This dataset only contains the data of the year 2005 and that was originally developed by compiling two datasets that are posted in the MAPSView webpage (<http://mapskp.cesr.fr/BSMP/index.php>) which contains the Bow Shock and Magnetopause event crossings between 2004 day 179 and 2007 day 349 (H.J. McAndrews, S.J. Kanani, A. Masters, and J.C. Cutler) through visual identification of CAPS and MAG data. The second list of data has the magnetopause crossings during the year 2004 to October 2010 and May 2012 to February 2013 [6].

This dataset contains seven variables: `year_cross`, `doy_cross`, `doyfrac_cross`, `hour_cross`, `minute_cross`, `type_cross`, `dirn_cross`, `xcrosslist`, `ycrosslist`, `zcrosslist`

```
## Rows: 480
## Columns: 10
## $ year_cross      <int> 2005, 2005, 2005, 2005, 2005, 2005, 2005, 2005, 2005, ...
## $ doy_cross       <int> 3, 4, 4, 4, 6, 15, 22, 24, 24, 24, 27, 27, 42, 42, 42...
## $ doyfrac_cross   <dbl> 3.794445, 4.106945, 4.241667, 4.900000, 6.484028, 15....
## $ hour_cross      <int> 19, 2, 5, 21, 11, 17, 19, 3, 6, 14, 16, 16, 7, 8, 19, ...
## $ minute_cross    <int> 4, 34, 48, 36, 37, 20, 50, 1, 55, 22, 1, 13, 52, 22, ...
## $ type_cross      <chr> "BS", "BS", "BS", "BS", "DG", "DG", "MP", "BS", "BS", ...
## $ dirn_cross      <chr> "I", "O", "I", "O", "S_SW", "E_SP", "O", "O", "I", "O...
## $ xcrosslist      <dbl> 22.840, 23.150, 23.280, 23.880, 25.044, 7.750, -1.200...
## $ ycrosslist      <dbl> -52.74, -52.18, -51.93, -50.62, -46.84, 2.44, -44.41, ...
## $ zcrosslist      <dbl> 5.54, 5.74, 5.82, 6.22, 7.12, 3.79, -4.69, -3.85, -3....
```

Variable Description

- `year_cross`: It contains a numeric value of the year in which spacecraft crossed the event.
- `doy_cross`: It contains a numeric value of the day on which spacecraft crossed the event.
- `hour_cross`: It contains a numeric value of the hour at which spacecraft crossed the event.
- `minute_cross`: It contains a numeric value of the minute at which spacecraft crossed the event.
- `doyfrac_cross`: $\text{doy_cross} + (\text{hour_cross} \times 60 + \text{minute_cross}) / (24 \times 60)$
- `type_cross`: This is a categorical variable contains information about what type of event did the spacecraft crossed.
 - **MP**: Magnetopause
 - **BS**: Bow Shock
 - **DG**: Data gap

- **SC**: SCAS interval which are unreliable data
- **dirn_cross**: This is also a Categorical variable that contains information regarding in which direction did the spacecraft moved. The direction categories in this variable are:
 - **in**: Inbound means the spacecraft is moving towards the planet.
 - **out**: Outbound means the spacecraft is moving away from the planet.
 - **S_SW**: Starts with the solar wind is the region at which spacecraft recorded values at the start of solar wind.
 - **S_SH**: Starts with Magneto sheath is the region at which spacecraft recorded values at the start of magneto sheath.
 - **S_SP**: Starts with Magnetosphere is the region at which spacecraft recorded values at the start of magnetosphere.
 - **E_SW**: Ends with the solar wind is the region at which spacecraft recorded values at the end of solar wind.
 - **E_SH**: Ends with Magneto sheath is the region at which spacecraft recorded values at end start of magneto sheath.
 - **E_SP**: Ends with Magnetosphere is the region at which spacecraft recorded values at the end of magnetosphere.

For an inbound the first event that will occur is a Bow Shock and later followed by Magnetopause. But in the case of an outbound direction the first event that occur will be a Magnetopause and later followed by a Bow Shock. Region of sampling at the start of any data gap will have a S_SW, S_SH and S_SP direction type and Region of sampling at the end of data gap E_SW, E_SH and E_SP. Dimension of this dataset is (480,10).

3.2 Dataset 2: Magnetometer Dataset

In this dataset it contains the magnetometer data of Cassini Spacecraft during the year 2005. Time difference between each Data points is one minute which means each data point represents the data of a particular minute. The data in this dataset are provided in the KSM (Kronocentric Solar Magnetospheric) Coordinate system which is a kind of Saturn centred Coordinate system where direction of X is from Saturn to the Sun and X-Z plane of the Coordinate system contains the Saturn centred axis of Magnetic Dipole 'M'.

```
## Rows: 494,683
## Columns: 12
## $ X                <int> 411138, 411139, 411140, 411141, 411142, 411143, 4111...
## $ Timestamp.UTC.   <chr> "01/01/2005 00:00:30.000", "01/01/2005 00:00:30.000"...
## $ DOY.UTC.         <chr> "2005-001", "2005-001", "2005-001", "2005-001", "200...
## $ SCET.s.          <dbl> 1104537630, 1104537630, 1104537690, 1104537750, 1104...
## $ deltaTime.s.     <dbl> 60, 0, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 6...
## $ X_KG.km.         <dbl> -3583775, -3584179, -3584723, -3584968, -3584868, -3...
## $ Y_KG.km.         <dbl> -94231.023, -76627.055, -43428.027, -8284.628, 26895...
## $ Z_KG.km.         <dbl> -255420.7, -255418.2, -255413.5, -255408.5, -255403....
## $ BTotal.nT.       <dbl> 0.4469750, 0.4498453, 0.4486538, 0.4455171, 0.441104...
## $ BX_KRTP.nT.      <dbl> 0.2975732, 0.3047569, 0.3070252, 0.2960059, 0.300877...
## $ BY_KRTP.nT.      <dbl> 0.11034938, 0.12523502, 0.12184002, 0.12933519, 0.12...
## $ BZ_KRTP.nT.      <dbl> -0.3143709, -0.3055858, -0.3030233, -0.3062884, -0.2...
```

Some of the relevant variables in the dataset are:

- `X_KSM.km.`: This is the X coordinate point value of the spacecraft in KSM Coordinate System.
- `Y_KSM.km.`: This is the Y coordinate point value of the spacecraft in KSM Coordinate System.
- `Z_KSM.km.`: This is the Z coordinate point value of the spacecraft in KSM Coordinate System.
- `Timestamp.UTC.` : It is the timestamp at which data point was recorded by the Magnetometer.
- `DOY.UTC.` : It tells about the day at which the datapoint was recorded in the year 2005.
- `BX_KSM.nT.`: It is the x component of magnetic field strength in Amperes/meter.
- `BY_KSM.nT.`: It is the y component of magnetic field strength in Amperes/meter.
- `BZ_KSM.nT.`: It is the z component of magnetic field strength in Amperes/meter.
- `BTotal.nT.` : It is the resultant vector of Bx, By and Bz

$$B_{Tot} = \sqrt{Bx^2 + By^2 + Bz^2} \quad (1)$$

This dataset contains 494683 rows and 12 columns.

3.3 Dataset 3: Combined Data Of Dataset 1 And Dataset 2

This is a newly created dataset by merging dataset 1 and dataset 2 So, that we can understand the Magnetic field properties during the events like Magnetopause and Bow Shock. For merging the two datasets I used date and time as the key. To format the date in Dataset 1 I used the `doy_cross` variable in each row adding to "2004-12-31" date. By doing so it will generate a date with respect to the reference date. In dataset 2 the variable `Timestamp.UTC.` is in string format inorder to convert it into a data format I used `as.POSIXct()` function with `format="%d/%m/%Y %H:%M"`. Later I have converted the `Timestamp.UTC.` variable into a new format and which is then stored in the `date` variable. Two new variables are also created in the Dataset 2 known as `hour_cross` and `minute_cross`. `left_join()` function was used for merging the two datasets by using the variables `date`, `hour_cross` and `minute_cross` which is common on dataset 1 and dataset 2. In the newly created dataset it contains all the variables of dataset 1 and dataset 2. The dimesnsion of the newly created dataset is 494683 rows and 19 columns. Some of the variables are removed from the dataset because we know that dataset 1 has very less number of datapoints when compared to dataset 2 So, it is better remove the variables like `xcrosslist`, `ycrosslist`, `zcrosslist`, `year_cross`, `doy_cross`, `SCET.s.`, `doyfrac_cross`, `hour_cross` and `minute_cross` from the merged dataset. Data manipulations and Visualizations were done on this newly created dataset.

```
## Rows: 494,683
## Columns: 16
## $ X               <int> 494525, 494526, 494527, 494528, 494529, 494530, 4945...
## $ Timestamp.UTC.  <chr> "01/01/2005 00:00:30.000", "01/01/2005 00:00:30.000"...
## $ DOY.UTC.        <chr> "2005-001", "2005-001", "2005-001", "2005-001", "200...
## $ deltaTime.s.    <dbl> 60, 0, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 6...
## $ X_KSM.km.       <dbl> 1171585, 1171614, 1171670, 1171729, 1171787, 1171846...
## $ Y_KSM.km.       <dbl> -3390681, -3390665, -3390635, -3390603, -3390571, -3...
## $ Z_KSM.km.       <dbl> 219621.2, 219636.3, 219664.8, 219695.0, 219725.3, 21...
## $ BTotal.nT.      <dbl> 0.4469750, 0.4498453, 0.4486538, 0.4455171, 0.441104...
```

```
## $ BX_KSM.nT.      <dbl> -0.13606475, -0.12058760, -0.11886104, -0.12252847, ...
## $ BY_KSM.nT.      <dbl> -0.3754122, -0.3783377, -0.3798774, -0.3700427, -0.3...
## $ BZ_KSM.nT.      <dbl> -0.2002696, -0.2103865, -0.2061500, -0.2149763, -0.2...
## $ TimeStamp       <dtm> 2005-01-01 00:00:00, 2005-01-01 00:00:00, 2005-01-0...
## $ date            <date> 2005-01-01, 2005-01-01, 2005-01-01, 2005-01-01, 200...
## $ type_cross      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ dirn_cross      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ time            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
```

After removing some of the variables, now the modified dataset has 494683 rows and 16 columns.

4 Data Manipulation Section

In this section will explain about the data manipulation that was done on the merged dataset (dataset 3). There were lots of NA values in different predictors, so it is important to impute these values before using it for training the models.

4.1 Removing Time Dependency

Since all the datapoints were recorded by the spacecraft using the instruments over time so, there can be a time dependency. To remove the time dependency, I made the dataset wider which means a thirty-minute window was used for each data point and stored the magnetic field strength values and position of the spacecraft at each minute as a column for each row. Now for each data point there are 219 columns. I have labelled each column in the format (predictor_name{minute_index}) For example, the BX_KSM16 represents the BX_KSM value after one minute of the selected datapoint. Since all time-dependency variables need to be removed So, in this dataset the variable Timestamp.UTC. ' was removed for this purpose.

4.2 Data Imputation

In this dataset there are many NA values in different predictors like **type_cross** and **dirn_cross**. Since **type_cross** and **dirn_cross** are both categorical variables So, it is critical to impute the NA values with relevant short terms. **type_cross** variables represents the type of event at which the spacecraft crossed. Currently **type_cross** variable has values MP, BS, DG and SCAS which represents Magnetopause, Bow shock, Data gap and Unreliable data. All the other data than the above-mentioned categories in the newly created dataset can be categorized as NE which means No Events Occurred.

For **dirn_cross** variable, which represents the direction at which spacecraft is moving. This variable has the categories E_SH, E_SP, E_SW, I, O, S_SH S_SP and S_SW. So, I have imputed all the datapoints which has NA values in **dirn_cross** as UD (Unknown Direction) which means the direction of the spacecraft when that datapoint was recorded is Unknown.

5 Exploratory Data Analysis

The dataset contains the records of more than 490000 magnetometer readings with labels of type of crossing, Magnetic field strength values of fifteen minutes before and after of a datapoint and position of the spacecraft at which the data was recorded. The orbits of the spacecraft covered almost all local hours and gave sufficient dayside coverage. Before getting into further analysis its important to understand whether the data is imbalanced or not. The dataset contains enough datapoints for training different models. Understanding the each variables on data is required before training the Models for classifying the events. Eventhough dataset contains

Inorder to understand the inbound and outbound boundary crossings. From [1] It is stated that from 72th day to 74th day of the year 2005 Cassini Spacecraft was in Outbound on the dawn flank. Figure shown below is the Line graph of Total Magnetic Field recorded during the period of 72 to 74th day of the year 2005. In the Figure below the blue dotted line represents the point at which spacecraft crossed Bow Shock boundary and the red dotted line represents the Magnetopause.

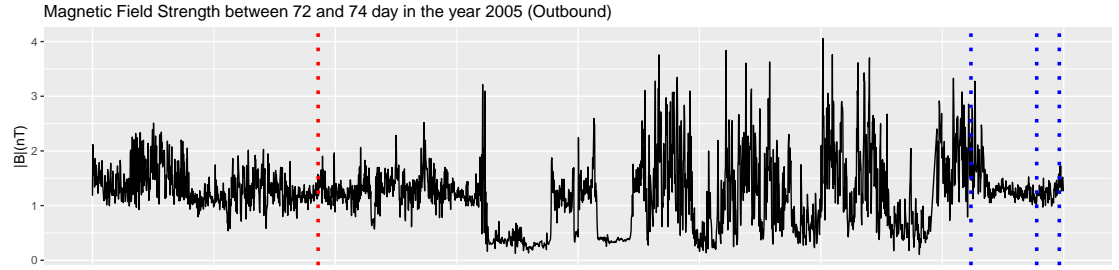


Figure 1: Line Graph of Total Magnetic Field Strength between 72 and 74th day of the year 2005

From the above figure we can clearly see that when the spacecraft approached the Bow Shock boundary there is a large value fluctuation of total magnetic field but when it comes to Magnetopause event there are only small value fluctuations. Since this data was recorded during the Outbound, the first boundary that was crossed by the spacecraft was Magnetopause and later followed by the Bowshock Boundaries.

During the Inbound prenoon of Cassini Spacecraft which is between 136th and 138th day of the year 2005, the spacecraft observed some clean bow shock crossings on 136th day and followed by the Magnetopause Crossings on 137th daya of the year 2005. Figure Shown below is the Magneticfield data between those days and the events like Bow Shock and Magneto pause are marked with blue and red color respectively.

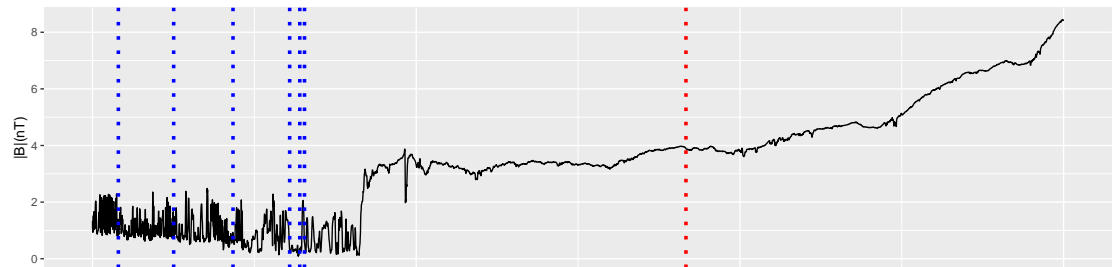


Figure 2: Line Graph of Total Magnetic Field Strength between 136 and 138th day of the year 2005

From the above figure we can clearly see that first boundary crossed by the spacecraft was Bow shock and later followed by the Magnetopause. There are very large fluctuations during when it

crossed the Bow Shock boundary but in the case of Magnetopause region the total magnetic field was kind of slow increase.

6 References

Sergis, Jackman, N. 2013. “Particle and Magnetic Field Properties of the Saturnian Magnetosheath: Presence and Upstream Escape of Hot Magnetospheric Plasma.” *Journal of Geophysical Research: Space Physics*, nos. 118, 1620–1634. <https://doi.org/10.1002/jgra.50164>.

7 Appendix

7.1 Supporting code