CHARACTERIZING SPACE WEATHER DRIVERS VIA IN-SITU DETECTION AND CLASSIFICATION

M.TECH. THESIS

Submitted by

EKNOOR KAUR

2402121003

In partial fulfilment for the award of the degree of

MASTER OF TECHNOLOGY

In

SPACE ENGINEERING



DEPARTMENT OF ASTRONOMY, ASTROPHYSICS AND SPACE ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY

INDORE

May 2026

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **CHARACTERIZING SPACE WEATHER DRIVERS VIA IN-SITU DETECTION AND CLASSIFICATION** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** and submitted in the **DEPARTMENT OF ASTRONOMY, ASTROPHYSICS, AND SPACE ENGINEERING (DAASE), Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2024 to May 2026 under the supervision of Prof. Bhargav Vaidya, Associate Professor, Department of Astronomy, Astrophysics and Space Engineering, IIT Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

07/05/26

Signature of the student with date
(EKNOOR KAUR)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

7/5/2026

Signature of the Supervisor of
M.Tech. thesis (with date)
(BHARGAV VAIDYA)

EKNOOR KAUR has successfully given his/her M.Tech. Oral Examination held on **04th May, 2026**.

Signature(s) of Supervisor(s) of MTech thesis
Date: 7/5/2026

Programme Coordinator, M.Tech.
Date:

Suman Majumder
Convenor, DPGC
Date:

S. Dey
HoD, DAASE
Date:

DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **WIND DATA CLUSTERING USING SELF ORGANIZING MAPS** in the partial fulfilment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** and submitted in the **DEPARTMENT OF ASTRONOMY, ASTROPHYSICS AND SPACE ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2025 to May 2026 under the supervision of Dr. Bhargav Vaidya, Associate Professor, Department of Astronomy, Astrophysics and Space Engineering, IIT Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

18/05/26

Signature of the student with date

(EKNOOR KAUR)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

18/05/26

Signature of the supervisor with date

(Dr. BHARGAV VAIDYA, ASSOCIATE PROFESSOR)

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my thesis supervisor, **Dr. Bhargav Vaidya**, for his invaluable support and guidance throughout the course of this research work. I am also thankful to the faculty and staff of the **Department of Astronomy, Astrophysics and Space Engineering, IIT Indore**, for their support and for providing the infrastructure and facilities essential for carrying out this research work.

I am profoundly grateful to my family members for their unconditional love, patience, and understanding, especially during the times I could not be in touch as often as I should have been. Their enduring support has been the foundation of all my efforts. I also thank my sister for her constant motivation and belief in me throughout this journey.

Lastly, I extend heartfelt thanks to my dear friend, whose thoughtful encouragement, insightful conversations, and emotional support were a steady source of strength during the most challenging moments of this work.

ABSTRACT

The solar wind exhibits significant variability due to changes in coronal source regions and transient structures such as Interplanetary Coronal Mass Ejections (ICMEs). Identifying these regimes is essential for understanding heliospheric processes and improving space weather forecasting. Plasma can be studied using complementary approaches: remote sensing for large-scale observations and in-situ techniques for local measurements.

To investigate in-situ plasma characterization in the solar wind, an unsupervised framework based on Self-Organizing Maps (SOM) is applied using measurements from the WIND spacecraft. A two-dimensional SOM is trained on four key parameters: proton density, bulk velocity, logarithm of plasma beta, and logarithm of proton specific entropy, and clustered using K-Means. The ideal number of clusters is determined using the Kneedle, Calinski-Harabasz, and Silhouette score methods. This results in four physically meaningful regimes: Slow Solar Wind (SSW), Compressed Slow Wind (CSW), Highly Alfvenic Wind (HAW), and Ejecta. The model shows good agreement with an existing source-based classification scheme and successfully identifies ICME intervals, with activation patterns reflecting solar cycle variations.

While the SOM enables in-situ analysis of space plasma, controlled laboratory experiments are performed to investigate plasma diagnostics under well-defined conditions. A DC glow discharge plasma analogous to ionospheric conditions is generated. Cylindrical and planar Langmuir probes are developed as in-situ diagnostic instruments, and a data processing pipeline is constructed to acquire and analyse I-V characteristics. Plasma parameters such as electron temperature, number density, plasma frequency, and Debye length are estimated. Electron temperature obtained using Optical Emission Spectroscopy (OES) shows good agreement with probe measurements highlighting the consistency and reliability of the diagnostic methods across both approaches.

TABLE OF CONTENTS

ACKNOWLEDGEMENT 2

ABSTRACT 3

Chapter - 1 : Introduction 9

1.1. Fundamentals of Plasma 9

1.2. Background Solar Wind 12

1.2.1. Parker Solar Wind Model 12

1.3. Space Weather Drivers 15

1.4. Observing Solar and Space Plasma 16

1.4.1. Remote Sensing 16

1.4.2. In-situ Observation 17

1.5. Motivation and objectives 22

Chapter - 2 : Methodology and Training 23

2.1. Classifying Structures in Solar Wind – Dataset 23

2.1.1. Collection 23

2.1.2. Cleaning 24

2.1.3. Feature Selection and Scaling 24

2.2. Adopted Methodology: Self-Organizing Maps (SOM) 26

2.2.1. Self-Organizing Maps 26

2.2.2. Kneedle and K-Means Algorithm 28

2.3. Training SOM 29

2.3.1. Parameters 29

2.3.2. Results 30

Chapter - 3 : Classifying Solar Wind Structures 33

3.1. Clustering Results 33

3.2. Cluster Analysis 35

3.3. ICME Data within Cluster 39

3.4. Comparison with Empirical Clustering Method 40

Chapter - 4 : Case Studies – Solar Wind Classification 43

4.1. WIND: 2016 43

4.2. Aditya-L1: 2024 & 2025 48

4.2.1. 2024 Data 48

4.2.2. 2025 Data 50
4.3. SWASTi Simulation: May 2024 event 52
Chapter - 5 : Langmuir Probe Design and Data Acquisition 55
5.1. Generating Plasma 55
5.2. Design and Development of Langmuir Probe 56
5.2.1. Cylindrical Langmuir Probe 56
5.2.2. Planar Langmuir Probe 57
5.3. Data acquisition and analysis of I-V characteristic 58
Chapter - 6 : Quantifying Plasma Parameters 62
6.1. Using Remote Sensing 62
6.2. From Langmuir Probe 64
Chapter - 7 : Summary and Outlook 70
7.1. Classification of Solar Wind from In-Situ Data 70
7.2. Developing Instrumentation for In-Situ Plasma Measurements 71
7.3. Outlook 71
REFERENCES 72
APPENDIX – A 73
APPENDIX – B 75
APPENDIX – C 79

LIST OF FIGURES

Figure 1: Debye shielding around a positive charge 10
Figure 2: Occurrence of plasma in nature 12
Figure 3: Five solutions of Parker's model 13
Figure 4: Solar wind at different coronal temperatures 14
Figure 5: Parker Spiral 15
Figure 6: Faraday cup schematic 18
Figure 7: Top-hat analyzer 19
Figure 8: Langmuir probe geometries; a) Cylindrical, b) Spherical, and c) Planar 20
Figure 9: Langmuir Probe I-V Characteristic 21
Figure 10: Solar wind parameters for the month of July 2004 24
Figure 11: Methodology 26
Figure 12: Initial and Output SOM 30
Figure 13: LR, NR, and QE decay during training 31
Figure 14: SOM U-Matrix 31
Figure 15: SOM Feature Maps 32
Figure 16: Ideal number of clusters using Kneedle and CH 33
Figure 17: Silhouette analysis (k = 3 & 4) 34
Figure 18: Cluster map 35
Figure 19: Feature maps with cluster boundaries 36
Figure 20: Clustered timeseries, July 2004 37
Figure 21: Alfvenicity of each cluster 37
Figure 22: Feature distribution for each cluster 38
Figure 23: ICME fraction distribution across SOM nodes 39
Figure 24: Comparison heatmaps 41
Figure 25: Xu & Borovsky timeseries 42
Figure 26: July 2016 WIND data 43
Figure 27: WIND 2016 hitmap 44
Figure 28: Activation per cluster 44
Figure 29: Clustered timeseries (WIND, 2016) 45
Figure 30: Comparison heatmaps (WIND, 2016) 46
Figure 31: ICME duration distribution 47
Figure 32: Predicted ICMEs and ground truth (raw) 47
Figure 33: Predicted ICMEs and ground truth (after processing) 48
Figure 34: SOM implementation on Aditya-L1 2024 data 48
Figure 35: Clustered timeseries (Aditya-L1, May 2024) 49
Figure 36: Comparison heatmaps (Aditya-L1, 2024) 50
Figure 37: Aditya-L1 ICME predictions 50
Figure 38: SOM implementation on Aditya-L1 2025 data 50
Figure 39: Clustered timeseries (Aditya-L1, June 2025) 51
Figure 40: Comparison heatmaps (Aditya-L1, 2025) 52

Figure 41: BMU Hitmaps (SWASTi, May 2024) 53
Figure 42: Activation Barplots (SWASTi, May 2024) 53
Figure 43: Clustered Timeseries (SWASTi, May 2024) 54
Figure 44: Plasma Generation in Laboratory 56
Figure 45: Cylindrical probe design 56
Figure 46: Planar probe design 57
Figure 47: Langmuir probe circuit diagram 58
Figure 48: The input, obtained, raw noise, and cleaned data (Cylindrical probe) 59
Figure 49: Smoothing data to obtain the I-V characteristic 60
Figure 50: Slope and curvature of I-V curve vs. applied voltage 61
Figure 51: Baseline corrected emission spectrum (Argon plasma at 0.300 mbar, 650 V) 62
Figure 52: Boltzmann plot analysis (Argon plasma at 0.300 mbar, 650 V) 64
Figure 53: Smoothed I-V Curve (cylindrical probe) 65
Figure 54: Semilog plot of ln I_e vs. V 66
Figure 55: Linear fit in electron retardation region 66
Figure 56: I-V curve (Planar probe) 68
Figure 57: Electron retardation region analysis 68
Figure 58: September 10, 2004, 12:15 – 13:00 (missing data and nonphysical spikes) 73
Figure 59: October 23, 2004, 04:19 – 04:20 (nonphysical spike in Proton Density) 73
Figure 60: December 17, 2004, 17:43 – 17:44 (nonphysical spike in Proton Density) 74
Figure 61: Clustered scatterplots (Input parameters: V_p, N_p, $\log_{10}T_p$, and B) 76
Figure 62: Clustered scatterplots (Input parameters: V_p, N_p, $\log_{10}T_p$, and $\log_{10}\beta_p$) 77
Figure 63: Clustered scatterplots (Input parameters: V_p, N_p, $\log_{10}S_p$, and $\log_{10}\beta_p$) 78
Figure 64: Decay Curves 80

LIST OF TABLES

Table 1: CUDA-SOM Settings 29
Table 2: ICME data within cluster 39
Table 3: Xu & Borovsky categories 40
Table 4: ICME data within Xu & Borovsky clusters 41
Table 5: Event-wise and datapoint-wise model performance parameters 48
Table 6: Design specifications of cylindrical probe 57
Table 7: Design specifications of planar probe 57
Table 8: Spectroscopic data used for Boltzmann plot analysis 63
Table 9: Plasma parameters obtained from cylindrical and planar probe 69
Table 10: SOM clustering results for different input parameters 75

Chapter - 1 : Introduction

This chapter provides an overview of the key concepts needed to understand the solar wind plasma and its impact on space weather. It starts with a review of the physics of plasma, including key concepts such as Debye shielding, plasma oscillations and sheath formation. It then considers the source and large-scale dynamics of the solar wind, as outlined in the Parker solar wind model and the Parker spiral. Key space weather sources such as coronal mass ejections (CMEs) and interaction regions, are also discussed. It also outlines the observational techniques used to study space plasma, including both remote sensing and in-situ methods. This sets the stage for the physical and observational understanding of the data analysis and modelling methods used in the following chapters.

Space weather is largely determined by the outflow of plasma from the Sun, called the solar wind, and its interaction with the planetary environments. The properties of this plasma and the structures it creates are critical in estimating and avoiding its effects on space and terrestrial technologies.

1.1. Fundamentals of Plasma

Plasma is a quasineutral gas consisting of charged particles, such as ions and free electrons, that exhibit collective behaviour. It is formed when a neutral gas is energized enough to strip electrons from atoms, creating ions. ***Quasineutrality*** implies that, on a macroscopic scale, plasma is electrically neutral i.e. positive and negative charges are equal ($n_e \sim n_i \sim n$). ***Collective behaviour*** refers to the fact that the motion of charged particles is governed by long-range electromagnetic fields, resulting in phenomena such as waves and instabilities.

Debye shielding:

When an external charge is introduced into a plasma, the surrounding charged particles rearrange themselves quickly to reduce its electric field. For example, when a positive charge is placed inside the plasma, electrons quickly accumulate around it, effectively shielding it. The characteristic length scale over which this shielding occurs is the ***Debye length (λ_D)***, which defines how far electric fields can penetrate into plasma before they are neutralized by surrounding charges. The region within a sphere of radius λ_D known as the ***Debye sphere*** contains the particles responsible for this shielding.

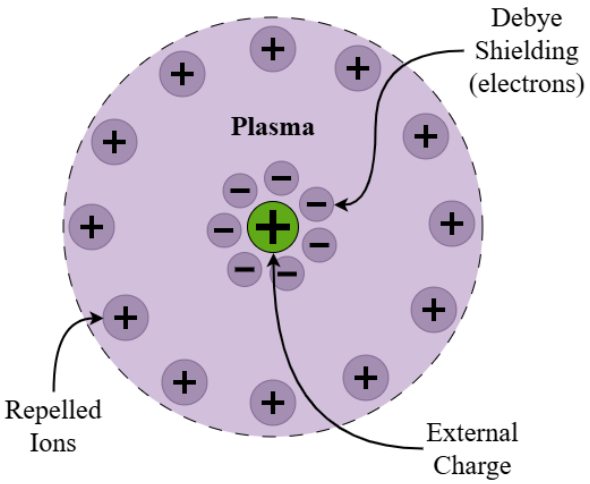


Figure 1: Debye shielding around a positive charge

Plasma oscillation:

In the presence of an electric field, electrons are displaced relative to ions. When the field is removed, the electrons tend to return to their original positions. Due to their inertia, this results in simple harmonic oscillations about the equilibrium position. The angular frequency of these oscillations, known as the plasma frequency, is given by:

$$\omega_p = \sqrt{\frac{n_e e^2}{m_e \varepsilon_0}}$$

Where,

n_e is the electron number density (m^{-3}),

e is the elementary charge (1.6×10^{-19} C),

m_e is the electron mass (9.11×10^{-31} kg), and

ϵ_0 is the free space permittivity (8.854×10^{-12} F/m).

Debye sheath:

A Debye sheath is a thin, non-neutral boundary layer that forms at the interface between a plasma and a material surface. Electrons reach the material surface more rapidly due to their higher mobility as compared to ions, causing it to acquire a negative potential relative to the plasma. This potential difference repels further electrons and accelerates positive ions toward the surface, resulting in a region with a net positive charge. The thickness of this layer is typically a few Debye lengths, depending on plasma parameters. The presence of the sheath is

important in plasma diagnostics, as it strongly influences the current collection characteristics of probes inserted into the plasma.

For an ionized gas to be classified as plasma, the following three conditions must be satisfied:

1. **Debye Length Condition:** To ensure that the local electric fields are neutralized and collective plasma behaviour dominates, the Debye length (λ_D) should be much smaller than the overall size of plasma (L) i.e.,

$$\lambda_D \ll L$$

2. **Sufficient Particle Density:** For effective shielding, the number of particles in the Debye sphere (N_D) must be significantly greater than 1 i.e.,

$$N_D \gg 1$$

3. **Dominance of Electromagnetic Forces:** To ensure that long-range electromagnetic forces dominate over local collisional forces, the product of plasma oscillation frequency (ω_p) and the mean time between particle collisions (τ) must be greater than one.

$$\omega_p \tau > 1$$

In nature, plasmas are found in the Sun, stars, solar wind, auroras, lightning, and interstellar space. These natural plasmas span a wide range of temperatures and charged particle densities.

Figure 2 illustrates the occurrence of different plasmas on temperature-density diagram. *Low-temperature, low-density plasmas* like auroras, solar wind, and interstellar space appear in the lower-left. *Moderate plasmas* such as neon signs, fluorescent lights, flames, and lightning lie in the mid-region. *High-temperature, high-density plasmas* such as the solar core, white dwarfs, and fusion devices (magnetic or inertial confinement) are in the top-right.

The red line in the figure represents the boundary between *weakly coupled (ideal) plasmas* and *strongly coupled (non-ideal) plasmas*. In weakly coupled plasmas, long-range electromagnetic forces dominate and particles behave collectively. In strongly coupled plasmas, inter-particle interactions become significant and classical models break down.

These fundamental plasma properties form the basis for understanding the behaviour of the solar wind, which is discussed in the following section.

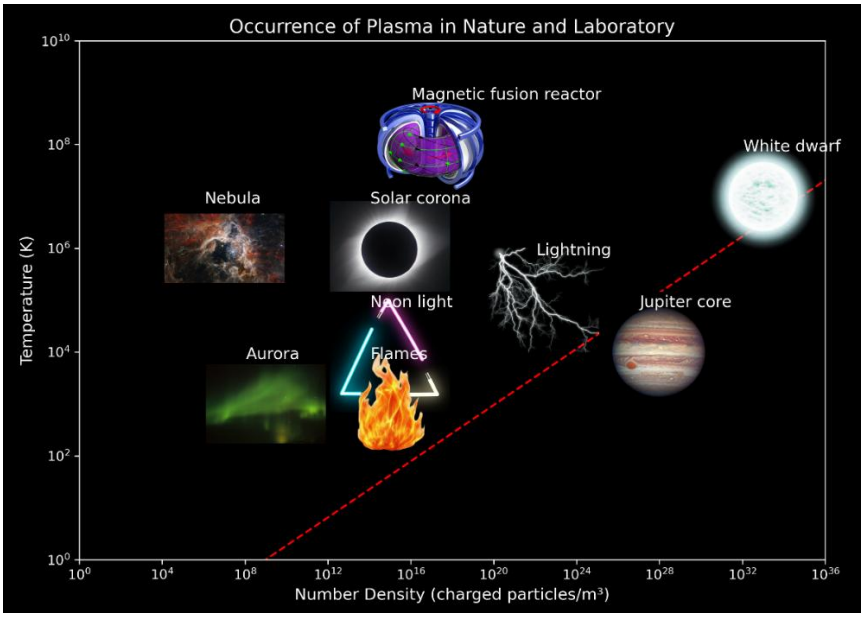


Figure 2: Occurrence of plasma in nature

Red line represents the boundary between ideal and non-ideal plasma.

1.2. Background Solar Wind

A comprehensive understanding of the physical properties of the solar wind and the large-scale structures it forms is crucial for space weather prediction and analysis. These phenomena significantly influence Earth's magnetosphere, contributing to geomagnetic disturbances that can impact satellite operations and other space-based technologies. Historical events such as the 1859 Carrington geomagnetic storm and the 2022 Starlink disruption clearly highlight the importance of monitoring solar energetic events. Continuous research into the behaviour of the solar wind is therefore vital for safeguarding technological infrastructure and advancing our knowledge of space environment dynamics.

Building on this background, the solar wind can be described using theoretical models that explain its origin and large-scale behaviour.

1.2.1. Parker Solar Wind Model

The solar wind is a steady stream of hot, ionized plasma that streams outward from the Sun's corona into the heliosphere at supersonic speeds. Coronal plasma is extremely hot and has low density, making particle collisions rare. As a result, the solar wind behaves as a collisionless, magnetized plasma and is well described by magnetohydrodynamics (MHD). This constant stream results from several physical processes acting together:

a. **Thermal Expansion:** The corona’s extreme temperature does not allow it to be a static atmosphere; it generates a large gas pressure that cannot be confined by gravity alone and keeps it expanding. This outflow is further accelerated by pressure gradients and eventually becomes supersonic at a certain distance from the Sun.

b. **Magnetic Structuring:** Magnetic field lines rooted in the photosphere extend into interplanetary space. Where these field lines are open, plasma is free to escape along them. Coronal holes, which are large regions of unipolar open magnetic field, naturally channel plasma into fast solar wind streams.
c. **Wave Heating:** Alfven waves carry energy that can heat and accelerate the plasma as they travel upwards from the photosphere into the corona.
d. **Magnetic Reconnection:** Plasma trapped in closed magnetic loops is released when reconnection occurs. This mechanism is believed to be a key contributor to the variability of the slow solar wind.

Parker showed that, assuming a steady, spherically symmetrical and isothermal outflow, the governing fluid equations yield five possible solutions (as shown in Figure 3), but only the transonic solution (V) is physically meaningful.

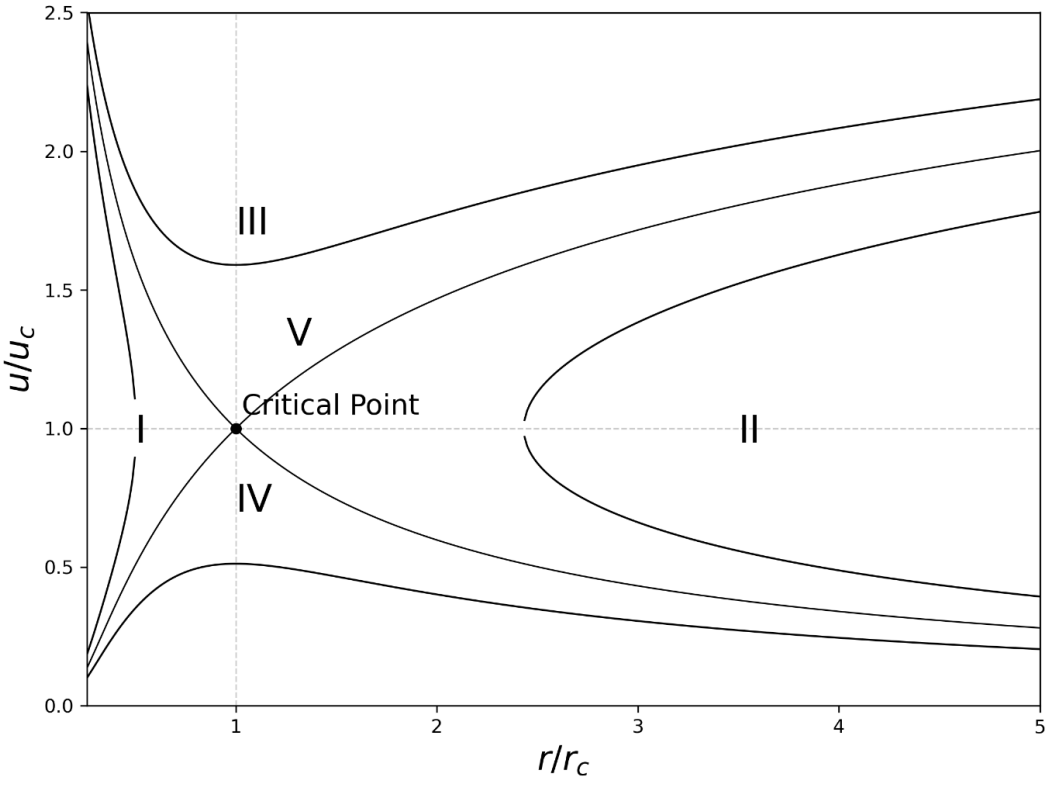


Figure 3: Five solutions of Parker's model

Solution I and II give two possible values of wind velocity at the same radius, which is unphysical. Solution III describes a purely supersonic wind and solution IV (also called the 'Solar Breeze' solution) describes a purely subsonic wind; both failing to match the actual observed conditions. The solar wind solution or the transonic solution defines solar wind to be subsonic near the solar surface that smoothly transitions to supersonic speeds as it moves away from the Sun. Figure 4 shows this solution for different coronal temperatures, highlighting how temperature dictates the speed of the outflow and the location of the sonic point i.e. the point where solar wind speed is equal to sound speed in corona. The following equation shows the Parker's transonic solar wind solution:

$$\left(\frac{u^2}{u_c^2} - 1\right) - ln\left(\frac{u}{u_c}\right)^2 = 4\, ln\left(\frac{r}{r_c}\right) + 4\left(\frac{r_c}{r} - 1\right)$$

Where,

u is the radial expansion speed,

r is the radius,

u_c is the coronal sound speed,

r_c is the critical radius i.e. the radius when solar wind changes from subsonic to supersonic.

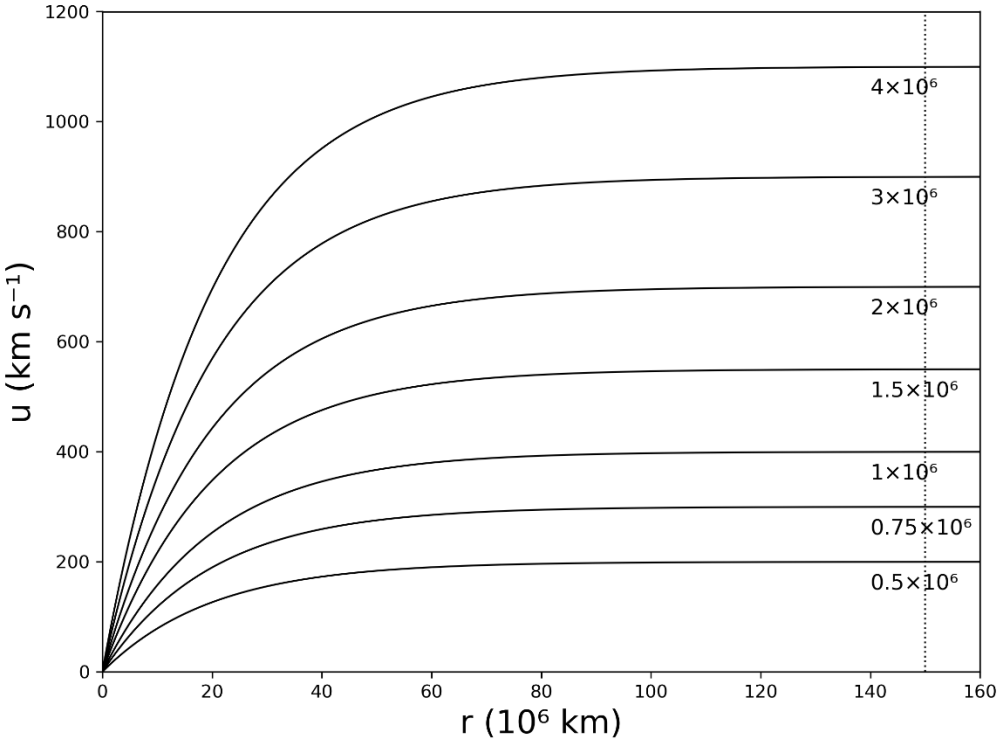


Figure 4: Solar wind at different coronal temperatures

Parker Spiral

While Parker model describes radial outflow, the effect of solar rotation further modifies the structure of interplanetary magnetic field. The magnetic field lines carried by the wind are dragged and wound into a spiral shape as a combined result of the radial outflow of the wind and Sun's rotation [6]. In a simple approximation, the magnetic field lines can be modeled by assuming an ideal Archimedean spiral configuration which is defined in polar coordinates (r, ϕ) by:

$$\phi(r) - \phi_0 = -\frac{\Omega_\odot}{u}(r - r_0)$$

Where,

$\Omega_{\odot}$ is solar rotation rate,

u is solar wind speed,

r_0, ϕ_0 are reference positions near the Sun.

A simple analogy is a rotating bucket filled with water that has small holes in its side. If the bucket was not rotating, the streams of water would exit in straight lines. However, due to rotation, each stream is deflected, tracing out a spiral path just like the magnetic field lines carried outward by the rotating Sun.

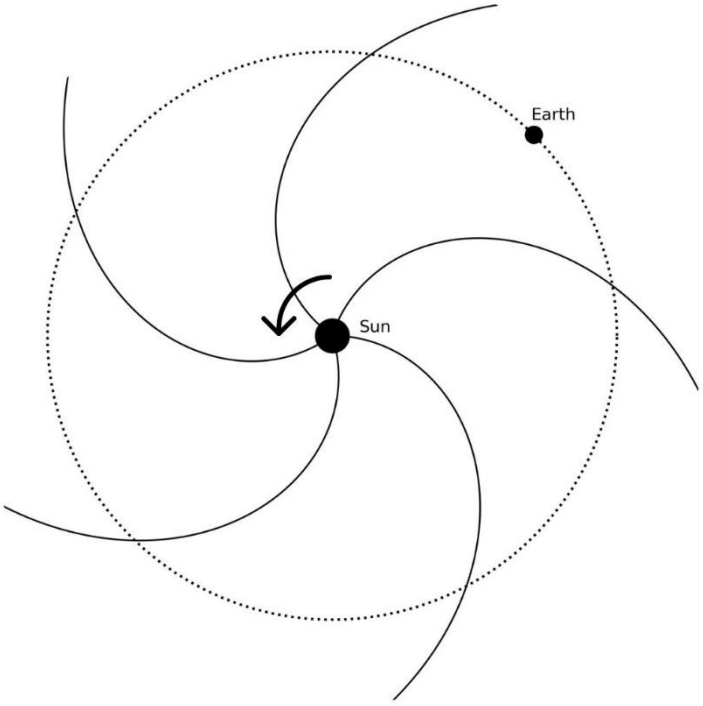


Figure 5: Parker Spiral

1.3. Space Weather Drivers

The large-scale solar wind described above gives rise to several dynamic structures that drive space weather. The solar wind is broadly categorised into two primary states based on its speed: fast and slow solar wind. The slow wind (300 – 500 km/s) originates from streamer belt regions around the Sun's equator while the fast wind (700 – 800 km/s) originates from open magnetic field regions called coronal holes, with slow wind being roughly twice as dense and more turbulent than the fast wind. By the time solar wind reaches Earth, it is largely collisionless and fully ionized. Its dynamics are dominated by electromagnetic interactions and influenced by large-scale solar-cycle behaviour.

A solar cycle typically lasts for 11 years, during which the solar wind changes drastically. During solar minima, steady fast winds from large, stable, polar coronal holes dominate the

heliosphere. Frequent and smaller geomagnetic storms are observed during this period. When the solar activity reaches its maxima, variable slow winds dominate the heliosphere as large coronal holes become irregular and coronal streamers extend to higher latitudes. This is the time when intense geomagnetic storms are caused due to large-scale structures generated in the corona.

Rapid magnetic reconnections in active regions release magnetic energy impulsively as solar flares. While flares do not significantly modify bulk solar wind parameters, they can precede CMEs and indicate periods of strong coronal activity, increasing the likelihood of transient solar wind disturbances.

Coronal mass ejections (CMEs) and Interplanetary CMEs (ICMEs) are the most energetic solar events. A CME is a large-scale expulsion of magnetized plasma from the low corona which usually involves the eruption of a twisted magnetic flux rope. As these CMEs propagate through the heliosphere, they become ICMEs. Fast ICMEs often drive interplanetary shocks identified by abrupt rise in velocity, temperature, density and magnetic field strength.

When fast wind encounters slower, denser wind ahead, a Stream Interaction Region (SIR) is formed where fast plasma compresses slow moving wind. These regions typically have enhanced proton density, temperature, magnetic field strength and velocity gradients. When a coronal hole producing fast wind lasts for multiple solar rotations, a SIR becomes a Corotating Interaction Region (CIR), which is a recurrent, large-scale structure that rotates with the Sun and contributes to alternating slow and fast wind at 1 AU.

To study these solar wind properties and structures, a combination of observational techniques is required.

1.4. Observing Solar and Space Plasma

Space plasma can be studied via remote sensing, including imaging techniques (such as coronagraphs) and spectroscopy. These methods help determine the particle composition and calculate properties like number density and electron temperature of the plasma. The problem with this method is that while it gives information about the plasma globally, it becomes very difficult to understand the processes at work inside plasma on a small, localized scale. Therefore, in-situ measurements are essential. Spacecrafts carry measurement instruments on-board to directly detect the particles present in the plasma, allowing for a more detailed and localized understanding of its behaviour. In-situ detectors can be divided into three categories: electric field probes, magnetic field sensors, and particle analyzers. This section explains the types and working of these instruments.

1.4.1. Remote Sensing

Electromagnetic radiation emitted or scattered by solar and space plasma is often observed using remote sensing techniques. These methods do not require physical contact with the

plasma and thus can be used to study large-scale systems such as the solar corona and heliosphere.

Imaging is a widely used technique, employing instruments like **coronagraphs** and **extreme ultraviolet (EUV) imagers**. Coronagraphs occult the bright solar disk to allow observation of faint coronal structures, facilitating observation of large-scale plasma structures such as coronal mass ejections (CMEs) as they erupt from the Sun. EUV imagers yield information on the temperature structure of the solar plasma.

Another key approach is **spectroscopy**, where the emitted radiation is used to obtain spectral lines of various elements and ions. The intensity, broadening, and shifts of spectral lines provide information about the electron temperature, density, velocity, and composition of the plasma. For instance, Doppler shifts of spectral lines can be used to infer plasma motion, while intensity of spectral lines can be used to calculate temperature.

Magnetograms are also an important remote sensing tool used to measure the magnetic field structure of the solar surface (photosphere). They are used to locate areas of significant magnetic activity (sunspots and active regions) and are important for understanding the origin of solar wind features.

Although remote sensing techniques offer useful global and large-scale insights into plasma behaviour, they do not provide local information about the plasma and its kinetic processes. As such, they are typically accompanied with in-situ measurements, which allow direct sampling of plasma particles and fields.

While remote sensing provides global context, direct measurements are necessary to understand local plasma behaviour and dynamics.

1.4.2. In-situ Observation

1.4.2.1. Particle Detectors

Particle detectors measure the flux, energy, direction, and composition of solar wind particles, allowing derivation of plasma moments such as density, temperature, and bulk velocity. These instruments sense the interaction of particles with a detector material, convert that interaction into a measurable electrical signal and tell us about the density, composition and velocity of the solar wind. Several types of such instruments exist, but this section focuses on those most relevant to solar wind missions.

Faraday Cup

A Faraday Cup (FC) is one of the most simple, robust, and reliable instruments used for in-situ solar wind measurements. It works on the principle that a charged particle produces a measurable current when it strikes a conductive surface. The magnitude of this current is

proportional to the particle flux entering through its aperture. Because of their ability to withstand extreme thermal environments, FCs are a core instrument on long-term missions like WIND, Parker Solar Probe, ACE, and Aditya-L1.

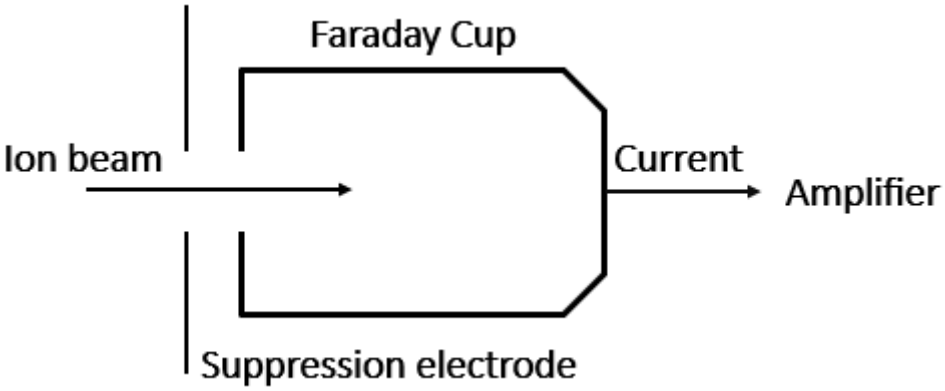


Figure 6: Faraday cup schematic

A typical FC consists of an ***aperture*** which allows the particles to enter the cup and blocks sunlight which would otherwise generate spurious photoelectron currents on the collector plate. These particles fall on the ***collector plate*** and produce current which is then converted into a digitized signal by ***current measurement electronics***. The ***retarding grids*** filter particles by energy-per-charge (E/q) ratio, only particles with kinetic energy higher than the applied bias voltage can pass through to the collector and the ***suppressor grid*** prevents secondary electrons emitted from escaping the detector to ensure accurate current measurement.

Unlike more complex analyzers, the FC separates particles based on E/q ratio instead of mass or direction using the retarding potential technique; a positive bias voltage is applied to the grid and ions with insufficient kinetic energy are repelled from the collector. By sweeping the applied voltage from low to high, the cup successively blocks lower-energy ions, generating a current-voltage (I-V) curve which is then used to calculate key solar wind parameters: proton number density, temperature, and bulk speed.

Electrostatic Analyzer (ESA)

Unlike a FC, ESA provides both energy-per-charge and angular resolution which allows reconstruction of full 3D velocity distribution functions for the solar wind ions and electrons. It is a core instrument on many heliophysics missions, including the WIND/3DP instrument, from which the bulk parameters used in this analysis are obtained. The principle behind the working of an ESA is that charged particles can be selectively filtered by their energy-per-charge ratio using curved electrostatic plates.

A typical ESA consists of an ***entrance aperture*** which defines the field of view, and an ***exit slit*** that removes particles that do not satisfy the analyzer's focusing condition. Once inside the detector, particles travel between two ***curved electrostatic plates*** held at a controlled potential difference. These plates form a radial electrostatic field that bends particle trajectories. Particle

detectors such as ***microchannel plate (MCP)*** or ***solid-state detector (SSD)*** count the number of particles that successfully pass through the analyzer. Figure 7 shows a simple cross-section of a top-hat ESA.

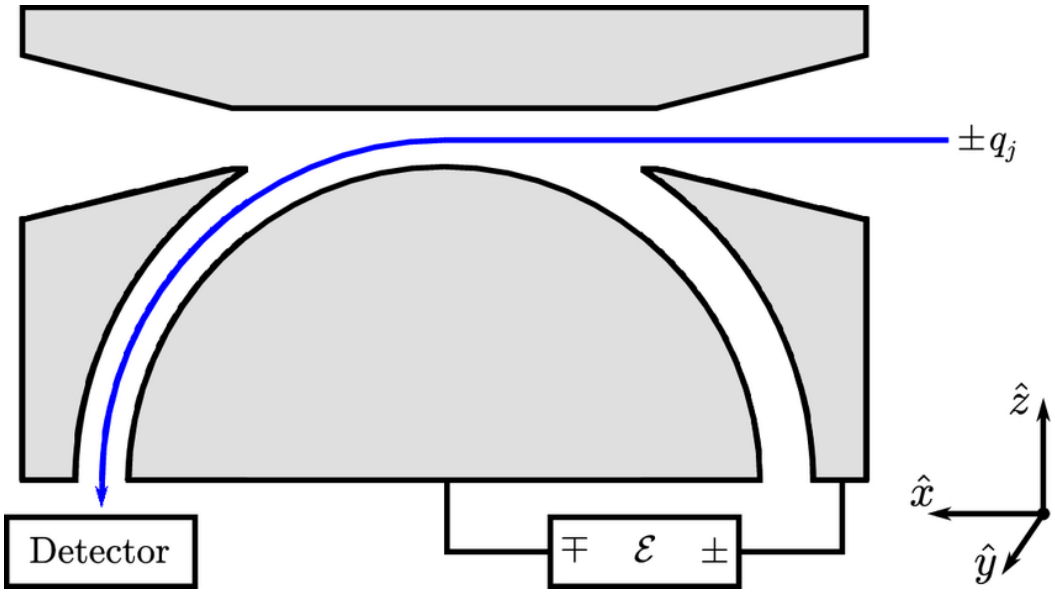


Figure 7: Top-hat analyzer

An ESA filters particles by balancing the centrifugal force from particle's curvature and electric forces from the plates. By adjusting the plate voltage, the instrument sequentially scans through narrow E/q channels. Only particles with very specific E/q ratio follow a stable trajectory through the analyzer to the detector, others strike the walls and are removed. A top-hat ESA provides full 3D angular coverage by segmenting the detector into azimuthal and elevation sectors. From combined multi-angle, multi-energy measurements, an ESA provides plasma bulk parameters and 3D velocity vector for multiple species making it crucial for detailed solar wind studies.

Mass Spectrometer

A mass spectrometer is designed to determine the composition of the solar wind by separating ions according to their mass-to-charge (m/q) ratio. In most heliophysics missions, it is implemented as an extension of an ESA, using the ESA the front-end selector while adding additional subsystems to measure particle speed and residual energy. After the ESA selects ions and reduces background noise, the filtered ions enter a time-of-flight (TOF) chamber where they strike a thin foil or conversion layer that emits secondary electrons, generating a 'start' signal. The ion then travels a fixed flight path before striking a MCP which produces a 'stop' signal. The difference between these times gives the ion's velocity. To improve mass resolution, especially for heavier ions, some instruments include a residual energy or impact detector which measures the ion's residual kinetic energy after passing through the TOF section. By combining these three measurements, the instrument uniquely determines m/q and identifies individual ions and charge states.

Langmuir Probe

A Langmuir Probe is a diagnostic instrument used to determine key plasma parameters such as electron temperature, electron and ion densities, and plasma potential in both laboratory and space plasmas. It operates by inserting a small electrode into the plasma and applying a swept or fixed bias voltage. The resulting current collected by the probe depends on how electrons and ions respond to the electric field surrounding the probe. This relationship is captured in the probe's I-V characteristic, which reflects the physics of the Debye sheath that forms around the electrode. In this sheath, electrons, being faster and lighter, reach the probe more easily than ions, causing the probe to charge negatively until a balance is established. This sheath controls how electrons and ions reach the probe and is therefore central to interpreting Langmuir probe measurements.

Langmuir probes are made of materials such as tungsten and molybdenum which can withstand high temperatures and are classified into three different categories based on their geometry as shown in Figure 8:

1. **Cylindrical probes:** A straight thin wire is used as the probe tip, easiest to construct.
2. **Spherical probes:** A metallic sphere installed on a boom, insensitive to plasma flow direction (omni-directional).
3. **Planar (flat) probes:** A flat conducting surface with insulated back, high signal to noise ratios.

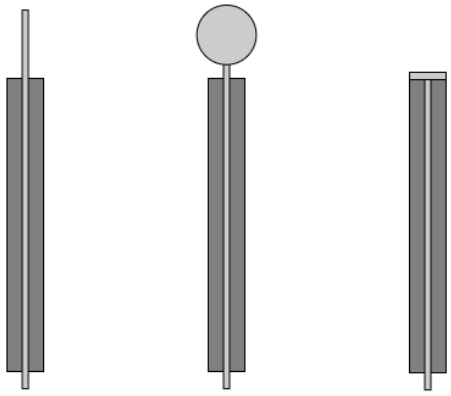


Figure 8: Langmuir probe geometries; a) Cylindrical, b) Spherical, and c) Planar

Plasma parameters are extracted by analysing five specific regions of the I-V curve (Figure 9):

- **Ion saturation region (probe strongly negative):** Ions dominate the collected current.
- **Transition region:** Electron current rises exponentially as the probe voltage reaches plasma potential. The slope of this region gives the electron temperature.
- **Electron saturation region (probe positive):** The collected current is dominated by electrons and relates to electron density.
- **Floating potential (V_f):** The point where net current equals zero i.e. the ion and electron currents are balanced.
- **Plasma potential (V_p):** The true electric potential of the ambient plasma and lies near the 'knee' of the I-V curve where transition region merges into the electron saturation region.

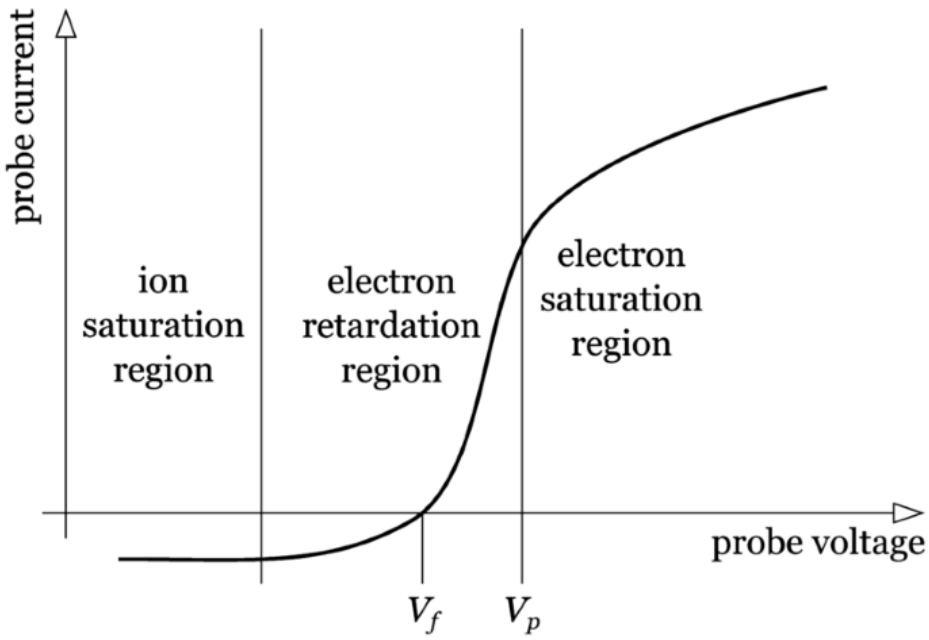


Figure 9: Langmuir Probe I-V Characteristic

1.4.2.2. Magnetic Field Measurements

Magnetic field instruments provide measurements of the DC and fluctuating components of the interplanetary magnetic field (IMF), enabling identification of large-scale structures and wave activity. Magnetic sensors or magnetometers provide direct, high precision measurements of the magnetic field vector.

Fluxgate Magnetometer

A fluxgate magnetometer is the primary instrument used to measure the vector magnetic field in the solar wind with high precision and low noise. It operates by monitoring the non-linear magnetic response of a ***ferromagnetic core*** that is periodically driven into magnetic saturation. An alternating current applied to the ***drive (excitation) coil*** forces the core to oscillate between positive and negative saturation. In the absence of an external field, the magnetization is perfectly symmetric, and the induced voltage in the ***sense (pickup) coil*** contains only even harmonics of the drive signal. When an external field is present, it biases the core and produces an asymmetric saturation response, generating odd-harmonic components proportional to the ambient magnetic field. By demodulating this signal, the instrument determines the magnetic field along the sensor axis. Using three orthogonal sensors, spacecraft obtain the full 3D field vector. Additionally, boom mounted placement minimizes the spacecraft generated magnetic disturbances, enabling long term stable and accurate IMF measurements.

Search Coil Magnetometer

A search coil or induction magnetometer measures time varying magnetic field by detecting voltage induced in a ***coil of wire*** when the magnetic flux through it changes [7]. Its operation is based on Faraday's law of induction, where a changing magnetic field B(t) induces a voltage V in a coil.

$$V(t) = -NA\frac{dB}{dt}$$

Where N is the number of coil turns and A is the area of the coil.

To enhance sensitivity, the coil is wound around a high-permeability ***ferromagnetic core***, and the induced signal is amplified using low noise preamplifiers.

Having established the physical background and measurement techniques relevant to the solar wind, the next section outlines the objectives that guide this study.

1.5. Motivation and objectives

Considering the importance of reliable classification of the solar wind conditions for space weather applications, it is necessary to define the scope, course and the desired results of this thesis with a clear set of objectives. These aims guided the development of Self Organizing Map (SOM) based analysis pipeline, clustering strategy, and comparative and validation analysis done using different spacecraft datasets.

The overall aim of this thesis is to create and test an unsupervised machine learning framework that can be used to classify solar wind plasma conditions and identify transient structures like ICMEs based on them. The specific objectives supporting this goal are:

1. To preprocess, normalise, and train a SOM using high-resolution plasma and magnetic field measurements made by the WIND spacecraft, and to analyse how SOM organises multidimensional features of the solar wind into an interpretable topological map.
2. To cluster the SOM weight vectors using K-Means, determining the optimal number of solar wind regimes using quantitative measures like Kneedle algorithm, Calinski-Harabasz index, Silhouette scores; and to interpret the resulting clusters in terms of known solar wind regimes while comparing with Xu & Borovsky solar wind origin categories.
3. To test the robustness of the trained SOM by applying it to independent datasets from WIND and Aditya-L1 across different solar cycle phases, analyse the activation patterns, flag possible ICME datapoints, and calculate event-wise detection metrics to quantify the model's performance.
4. To develop Langmuir probes of different geometries as a supporting instrument for space plasma measurements relevant to solar wind characterisation and compare their results.

To expand on these objectives, the next chapter describes the complete methodological approach adopted to achieve them.

Chapter - 2 : Methodology and Training

This chapter presents the methodological framework adopted in this study, including data collection, preprocessing, feature selection, and scaling. Solar wind plasma and magnetic field data obtained from the WIND spacecraft are processed to ensure quality and consistency across different solar cycle phases. The chapter describes the selection of relevant physical parameters and the application of robust scaling to account for the large variability in solar wind properties. The core of the methodology is based on the Self-Organizing Map (SOM), an unsupervised machine learning technique that produces a low-dimensional (2D) representation of high-dimensional input data while preserving its topology. The trained SOM is subsequently clustered using the K-Means clustering algorithm, which groups similar plasma states into clusters. Quantitative approaches such as the Kneedle algorithm, Calinski-Harabasz index, and Silhouette analysis are used to determine the optimal number of clusters. Overall, this chapter describes the entire process for data preparation, training and analysis of the solar wind data.

2.1. Classifying Structures in Solar Wind – Dataset

This section describes the dataset used in this study along with the preprocessing steps applied prior to training.

2.1.1. Collection

Solar wind plasma and magnetic field parameters were obtained from NASA's CDAWeb service for the WIND spacecraft. Specifically, the 3-second resolution datasets from the WIND/MFI (magnetic field) and WIND/3DP (velocity, proton density, and proton temperature) instruments were used.

A total of twenty-four months of data corresponding to different phases of the solar cycle were selected for analysis. Instead of using a continuous 11-year dataset, six-month intervals from each phase were considered. This approach ensures representation of all solar cycle phases while reducing computational cost. In addition, the WIND ICME catalogue was used to identify and overlay time intervals corresponding to ICME events on the solar wind time series plots for comparison and event analysis.

Figure 10 shows the solar wind parameters for the month of July, along with four identified ICME intervals: "July 22, 2004, 09:45:00 – July 24, 2004, 05:45:00", "July 24, 2004, 05:32:00 – July 25, 2004, 14:37:00", "July 25, 2004, 17:45:00 – July 26, 2004, 03:34:00", and "July 26, 2004, 22:25:00 – July 27, 2004, 23:58:00".

2.1.2. Cleaning

Before applying the SOM, data quality control was performed to remove missing or corrupted samples. The WI_PM_3DP dataset was cleaned using VALID (0: bad quality and 1: good quality) and GAP (0: no gap and 1: gap) flags, while the WI_H0S_MFI was cleaned by removing the fill values. Manual inspection was also performed to remove spurious outliers and abrupt transitions due to instrumental issues or data transmission errors. Figures in APPENDIX – A show examples of bad quality data observed during manual inspection for the months of September, October and December of year 2004.

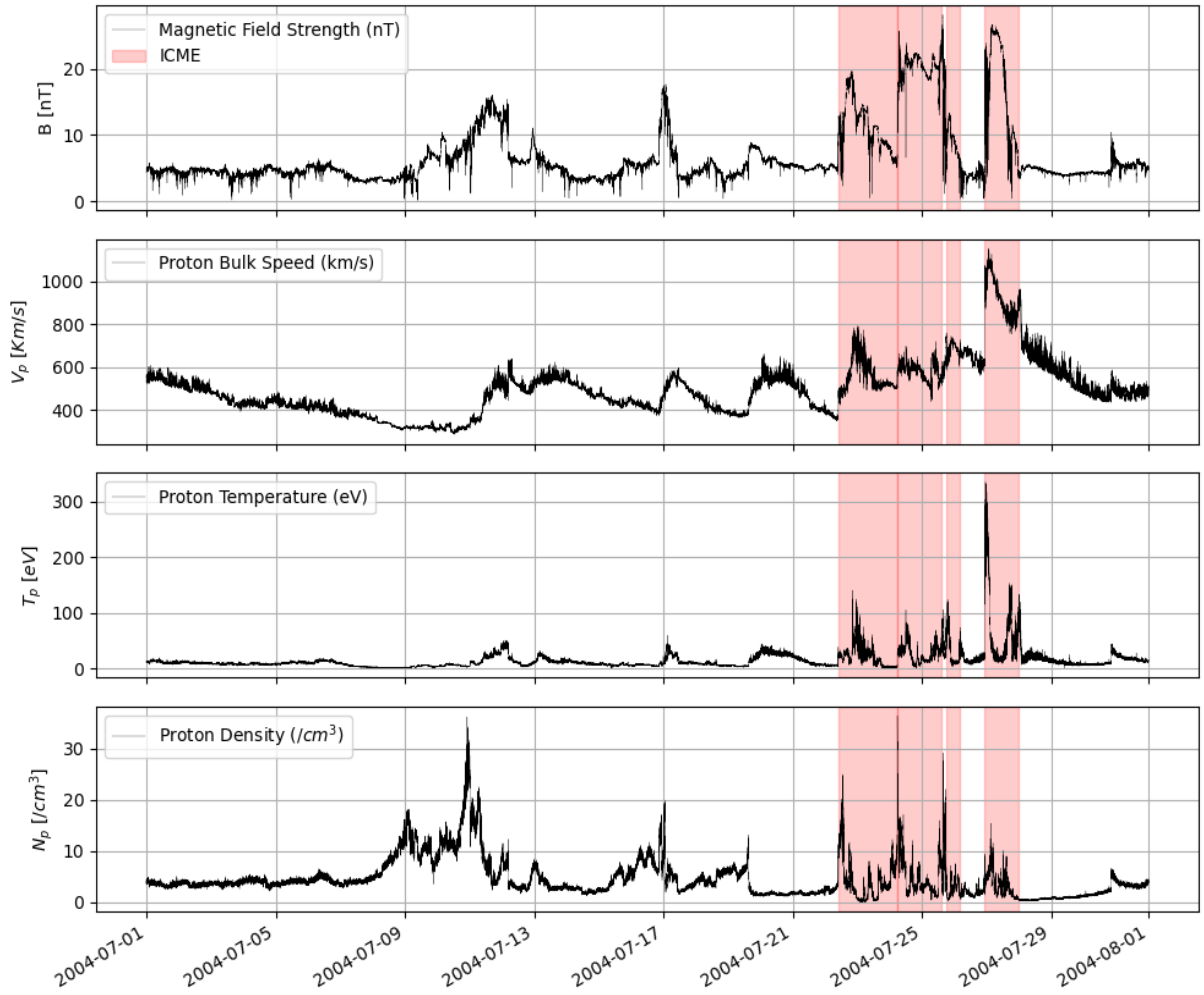


Figure 10: Solar wind parameters for the month of July 2004

2.1.3. Feature Selection and Scaling

Initially, four in-situ parameters - proton bulk velocity (V_p), proton number density (N_p), magnetic field magnitude (B), and logarithm of proton temperature ($\log_{10}T_p$), were used for training. For this purpose, a six-month dataset (July 1, 2004 – January 1, 2005) was considered.

However, after cluster analysis and model implementation, three limitations were observed:

1. The SOM was unable to effectively cluster most ICME events using this set of parameters.
2. The map did not accurately capture solar wind structures across different solar cycle phases (maxima and minima).
3. The final weights of edge nodes after training were unreliable due to ***edge effects***.

To address these issues, a revised set of parameters was identified that could better capture ICME characteristics without degrading the resolution of other solar wind structures. Parameters such as alpha-proton ratio, proton specific entropy, and plasma beta were evaluated (detailed analysis is presented in APPENDIX – B). Based on this analysis, V_p, N_p, $\log_{10}S_p$ and $\log_{10}\beta_p$ were selected as the final feature set, where S_p and β_p are calculated as:

$$S_p = \frac{T_p}{N_p^{2/3}}$$

$$\beta_p = 0.4026\frac{N_p T_p}{B^2}$$

Where:

T_p is proton temperature [eV],

N_p is proton number density [cm^{-3}],

S_p is in eV cm^2, and

B is magnetic field strength [nT].

The logarithm of proton specific entropy and plasma beta was used due to their wide dynamic range, which could otherwise reduce clustering effectiveness.

To ensure that the SOM captures solar wind variability across all phases of the solar cycle, a generalized training dataset was constructed. Instead of using a continuous multi-year dataset, six-month intervals were selected from each phase of the solar cycle:

- January 1, 1998 – July 1, 1998 (Rising phase)
- July 1, 2000 – January 1, 2001 (Solar maxima)
- July 1, 2004 – January 1, 2005 (Declining phase)
- September 1, 2008 – March 1, 2009 (Solar minima)

In the next step, robust scaling was applied to normalize the data using median and interquartile range (IQR). This method is less sensitive to outliers compared to standard scaling methods such as Z-score scaling, which rely on the mean and standard deviation.

The transformation is given by:

$$x_{scaled} = \frac{x - M_x}{IQR_x}$$
$$IQR_x = Q3_x - Q1_x$$

Where:

M_x is the median,

IQR_x is the interquartile range,

Q1 is the first quartile (25th percentile), and

Q3 is the third quartile (75th percentile) of the dataset.

The issue of unreliable edge nodes could not be resolved through feature selection or scaling and was instead addressed by modifying the map's topology, as discussed in Section 2.3.

2.2. Adopted Methodology: Self-Organizing Maps (SOM)

To classify solar wind structures from high-dimensional plasma data, an unsupervised learning approach is required that can preserve the correlation between different parameters. Self-Organizing Maps (SOM) were chosen for this task as they can preserve the topology of multi-dimensional data while mapping the data on a low-dimensional grid. This enables visualization of complex solar wind patterns and classification of solar wind structures. The approach followed in this study, which involves training a SOM and then applying K-Means to cluster the data, is shown in Figure 11.

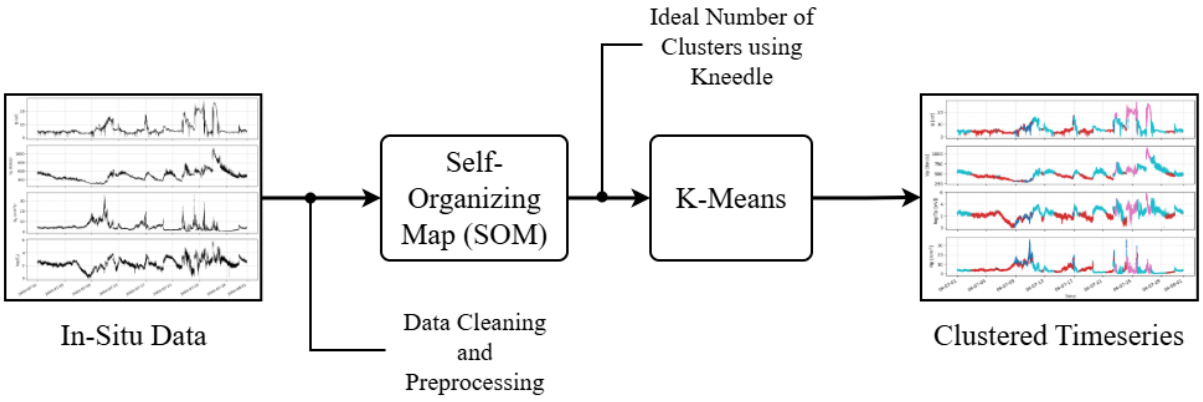


Figure 11: Methodology

2.2.1. Self-Organizing Maps

SOM is an unsupervised, competitive learning algorithm designed to map high-dimensional data onto a low-dimensional (mostly 2D) grid while ensuring topological preservation of the data [1]. In a self-organizing map, each neuron represents a prototype or weight vector that has the same dimensionality as the input data.

During training, the SOM calculates the Euclidean distance between input and weight vector to find the best-matching unit (BMU) and iteratively adjusts the weights of the BMU and neighboring units based on the following weight updating formula:

$$w_i(t+1) = w_i(t) + \eta(t).h_{ci}(t).[x(t) - w_i(t)]$$

Where:

- $w_i(t)$ is the weight vector of neuron i at time t,
- x(t) is the input vector,
- η(t) is the learning rate,
- $h_{ci}(t)$ is the neighbourhood function around the best-matching unit (BMU) c.

The BMU or the "winning neuron" is the neuron which is closest to the input vector.

$$c = \underset{i}{\operatorname{argmin}} (\| x - w_i \|)$$

Where ||.|| denotes the Euclidean distance between the input vector and the weight vector and is given by:

$$\| x - w_i \| = \sqrt{\sum_{j=1}^{d} (x_i - w_{ij})^2}$$

Where d is the number of dimensions of the input, and x_i and w_{ij} are the components of the input vector x and the weight vector w_i, respectively.

The neighbourhood function $h_{ci}(t)$ defines how much influence each neuron in the SOM grid has based on its distance from the BMU at a given time t and typically follows a Gaussian form:

$$h_{ci}(t) = \exp\left(-\frac{\| w_i - \mathrm{c} \|^2}{2\sigma^2(t)}\right)$$

Where σ(t) is the neighbourhood radius that decays with time causing the neighbourhood to shrink as training progresses. Initially, the neighbourhood radius is large, meaning a wide range of neurons are updated, but as training continues, the radius shrinks, and only the BMU and its close neighbours are updated. The Gaussian function is preferred because it provides a smooth, continuous decay in influence with distance. This smoothness is important in SOM training to avoid abrupt changes in the weight vectors, which could lead to erratic learning behaviour.

The Quantization Error (QE) is the average squared Euclidean distance between each input vector and its corresponding BMU. A lower QE means that the SOM has been trained well, with its neurons (weights) being close to the input data points, indicating that the map structure has adapted well to the data. For a dataset with n number of datapoints:

$$QE = \frac{1}{n}\sum_{i=1}^{n} \|x_i - c_{x_i}\|^2$$

2.2.2. Kneedle and K-Means Algorithm

After training, each SOM neuron represents a cluster prototype. To obtain macro-clusters, the K-Means algorithm is applied to the trained SOM weight vectors [4].

During K-Means clustering, the inertia decreases as the number of clusters 'k' increases. Initially, this decrease is sharp, but after a certain point, adding more clusters yields only marginal improvement. The Kneedle algorithm is applied to the inertia vs k curve to automatically detect this "elbow point", where the rate of improvement significantly slows down. This point represents the balance between model complexity and clustering accuracy [2].

The K-Means objective is to minimize the intra-cluster variance (or inertia) given by,

$$ICV(K) = \sum_{k=1}^{K} \sum_{x \in C_k} \|x - \mu_k\|^2$$

Where C_k is the k^{th} cluster and μ_k is the centroid of the k^{th} cluster.

Another method used for estimation of ideal number of clusters 'k' is the Calinski-Harabasz index. The CH index attains its maxima when clusters are dense and clearly separated and it is given by,

$$CH(k) = \frac{Tr(B_k)/(k-1)}{Tr(W_k)/(N-k)}$$

Where $Tr(B_k)$ and $Tr(W_k)$ are the trace of the intra-cluster and inter-cluster dispersion matrices, and N is the number of data points.

Along with this, Silhouette scores are used to analyze clusters obtained from K-Means. It is a measure of cohesion and separation of clusters i.e. how well a data point fits within its assigned cluster compared to other clusters.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Where b(i) is the minimum average distance of i to all other clusters and a(i) is the average distance of i to other datapoints within the same cluster. This analysis consists of measuring scores for all datapoints and the average score for each cluster and the ideal number of clusters is where the average silhouette score is maximized.

Therefore, the final number of clusters is chosen by jointly considering the CH index, Kneedle elbow point, and Silhouette score to ensure a balanced and consistent clustering configuration.

2.3. Training SOM

2.3.1. Parameters

The table below shows the values of parameters used for training SOM along with those used in initial SOM. The calculations of all parameters are attached in APPENDIX – C.

Parameter	Value/Setup used for the 24-month dataset (current)	Value/Setup used for the 6-month dataset (initial)
Input parameters	V_p, N_p, $\log_{10}S_p$, $\log_{10}\beta_p$	V_p, N_p, $\log_{10}T_p$, B
Learning Mode	Online	Online
Number of Rows (x)	189	126
Number of Columns (y)	116	96
Initial Learning Rate (η_0)	0.38	0.25
Initial Neighbourhood Radius (σ_0)	38	25
Maximum Number of Epochs (T_{max})	13	9
Distance Metric	Euclidean	Euclidean
Neighbourhood Function	Gaussian	Gaussian
Initialization Method	Random Input Vectors	Random Input Vectors
Lattice Type	Hexagonal	Hexagonal
Grid Initialization Randomization	On	On
Decay Type for η and σ	Exponential	Exponential
Normalize Distance	Off	Off
Map Geometry	Toroidal	Planar

Table 1: CUDA-SOM Settings

The number of SOM neurons (K) was determined using Kohonen's heuristic [1]:

$$K = 5\sqrt{N}$$

The grid dimensions x and y were selected such that:

$$x \times y \cong K, \qquad and \ \frac{x}{y} = \frac{PC_1}{PC_2}$$

Where PC_1 and PC_2 are the first two principal component eigenvalues, ensuring the map respects data variance ratios and N is the total number of data points in the dataset.

The decay function for both neighbourhood radius (σ) and the learning rate (η) is selected as an exponential function given by:

$$f(t) = f_0 exp\left(-\frac{t}{\sqrt{T_{max}}}\right)$$

where f_0 is the initial value of the parameter, t is the iteration step and T_{max} is the maximum number of epochs. The maximum number of epochs was chosen such that the neighbourhood radius approaches 1 near the end of training:

$$T_{max} \leq \left[\log\left(\frac{\sigma_{crit}}{\sigma_0}\right)\right]^2 = \left[\log\left(\frac{\eta_{crit}}{\eta_0}\right)\right]^2$$

Where σ_0 and η_0 is the initial neighbourhood radius and learning rate, $\eta_0 = 0.01\sigma_0$,

σ_{crit} and η_{crit} is the critical/final neighbourhood radius and learning rate.

The map geometry was taken to be toroidal to deal with edge effects caused due to the neighbourhood function getting cut off and causing unreliable weights near edges in planar geometry. This type of geometry joins the top, bottom and left, right edge neurons of the map, forming a continuous topology without boundary discontinuities.

2.3.2. Results

The training of the SOM took approximately 23 hours and the Quantization Error (QE) associated with the trained map is 0.13 [5].

Visualization of the initial SOM map shows random distribution of input vector across nodes and the final trained structure (after 13 epochs) shows topological organization of the input features as shown in Figure 12 and the trend of neighborhood radius, learning rate, and quantization error throughout training is shown in Figure 13.

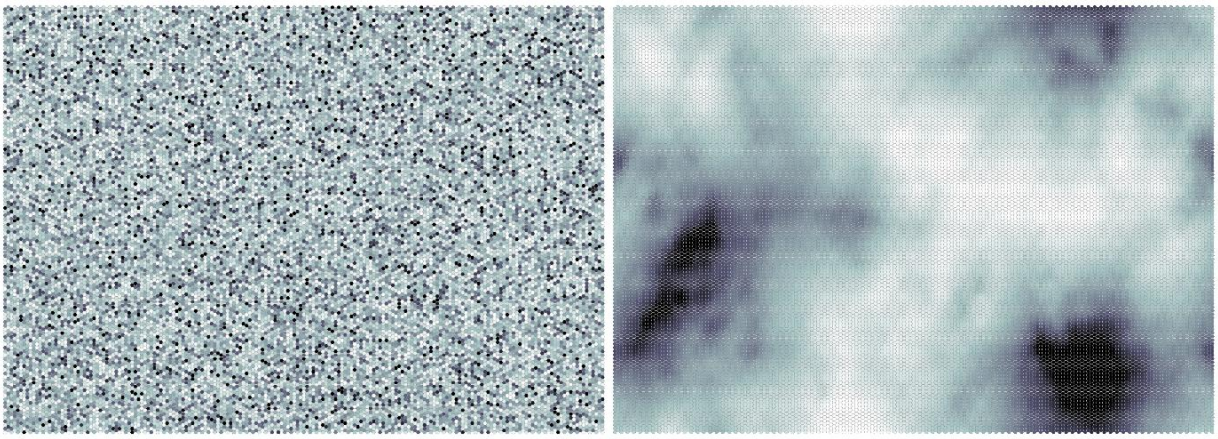


Figure 12: Initial and Output SOM

Figure 14 shows the U-Matrix (Unified Distance Matrix) of the SOM and represents the average distance between neighboring nodes, revealing cluster boundaries and transition regions. The red dots indicate neurons with more than 95% data points corresponding to ICME.

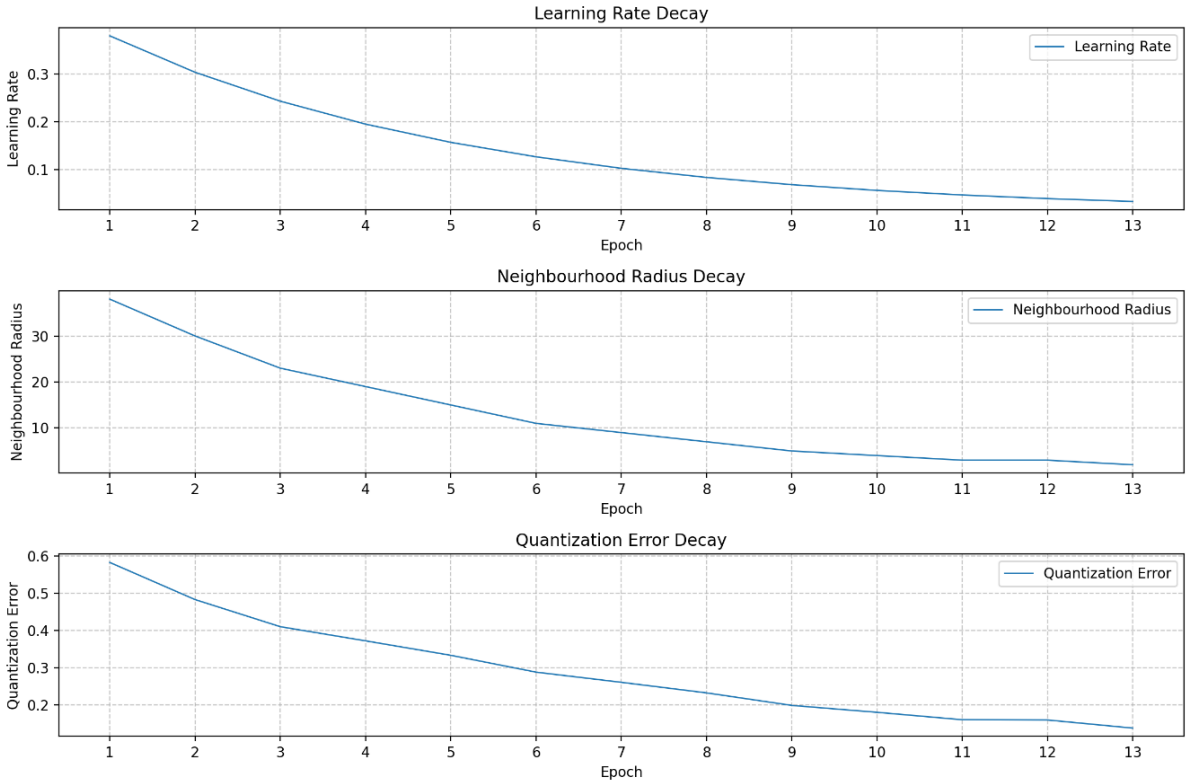


Figure 13: LR, NR, and QE decay during training

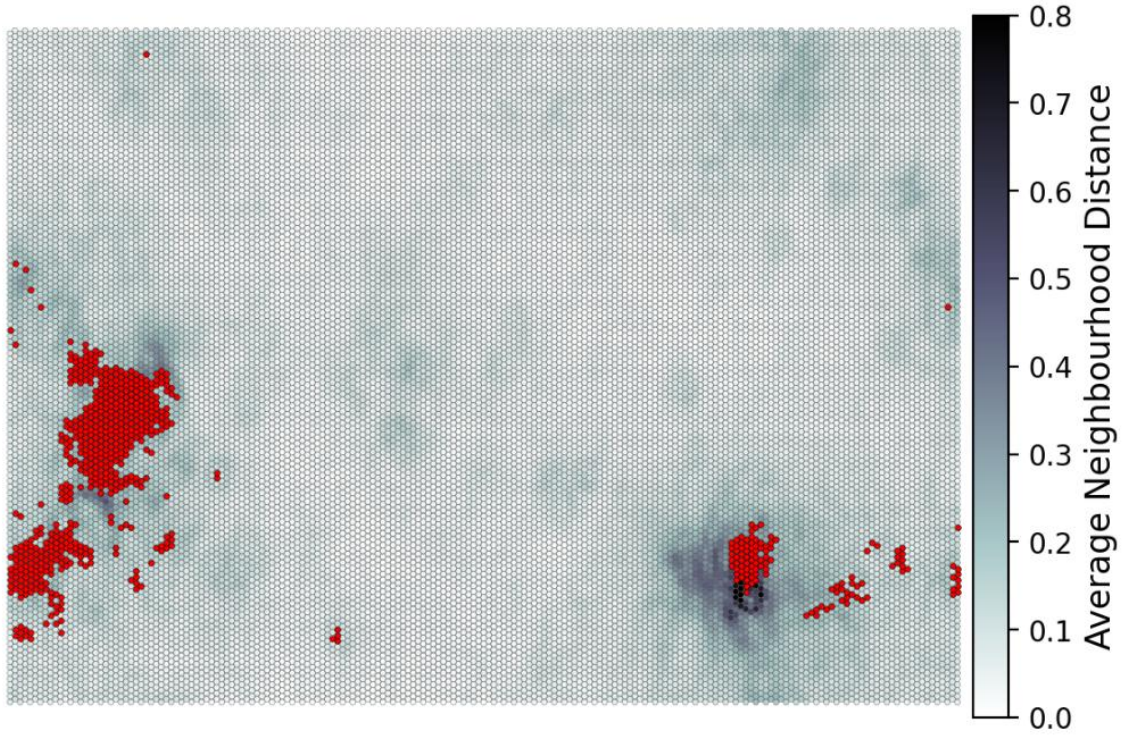


Figure 14: SOM U-Matrix

Figure 15 shows the individual 2D maps for each input feature (velocity, density, log proton specific entropy, and log plasma beta) and represents their distribution over the trained map. The color of the node indicates the magnitude of the corresponding feature for that node. To visualize the actual feature values, the weights were transformed back to original scale. It is observed that ICME nodes appear as clusters in the region with low values plasma beta.

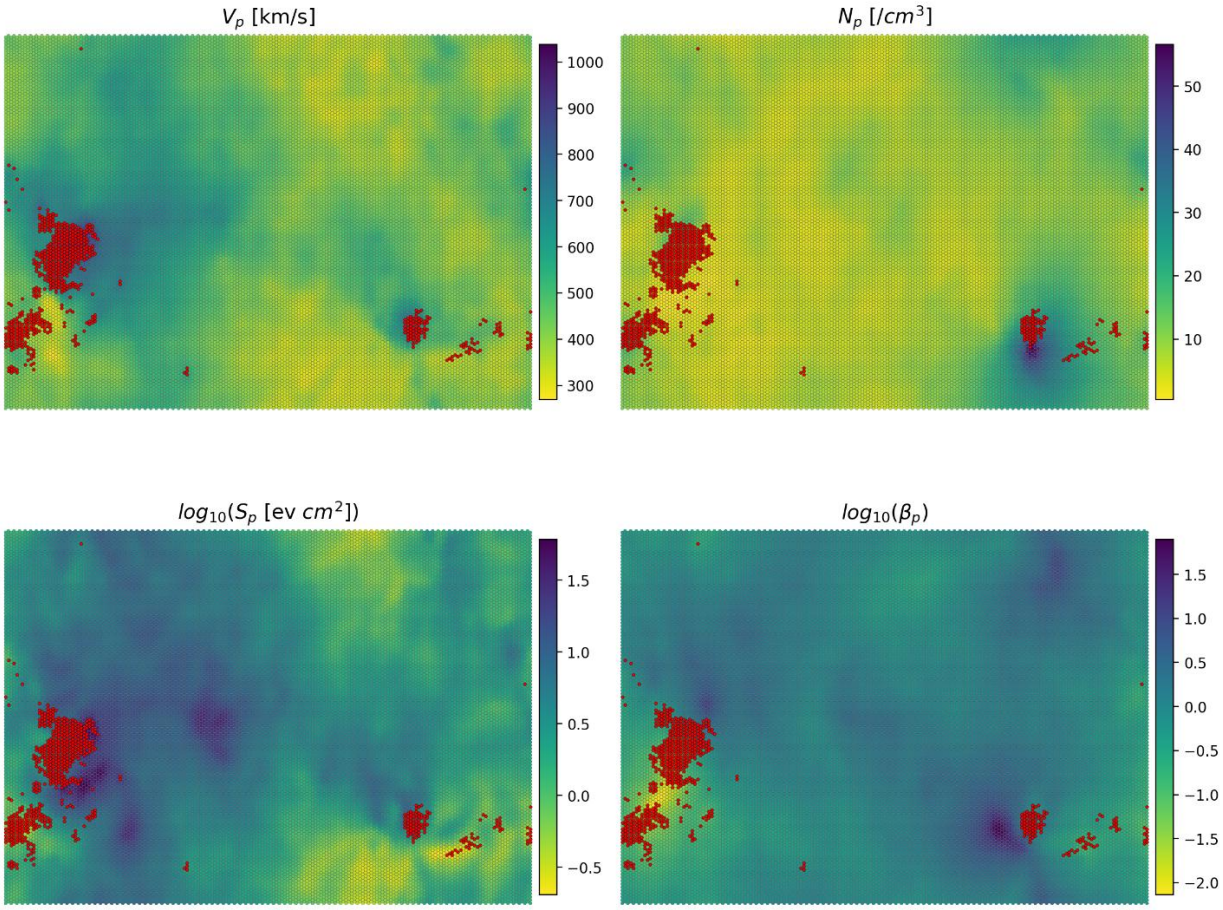


Figure 15: SOM Feature Maps

The trained SOM map provides a structured representation of the solar wind variability. In the next step, clustering techniques are applied to extract physically meaningful solar wind categories.

Chapter - 3 : Classifying Solar Wind Structures

This chapter presents the clustering results obtained from the trained SOM using the K-Means algorithm, with the number of clusters optimised using the Kneedle method, Calinski-Harabasz index, and Silhouette analysis. The resulting clusters are examined based on their physical characteristics, including variations in solar wind velocity, density, and thermodynamic properties, and are compared with known solar wind structures. Particular emphasis is placed on the identification of transient events such as Interplanetary Coronal Mass Ejections (ICMEs) and their distribution across clusters. The chapter further investigates the relationship between the SOM-based clusters and known solar wind regimes, providing a physical interpretation of the clusters based on their characteristics. To validate the results, the clustering is compared with the empirical classification scheme of Xu and Borovsky. This approach provides a measure of consistency and reliability of the SOM clustering in identifying both steady-state and transient solar wind features.

3.1. Clustering Results

To find the ideal number of clusters, multiple evaluation methods were employed, including the Kneedle algorithm, CH, and Silhouette analysis.

In Figure 16, blue is ICV vs. k and purple is CH vs. k plot. It shows the optimal number of clusters as detected by Kneedle algorithm and CH score. The Kneedle algorithm identifies an elbow point at k = 4, whereas CH score reaches its maximum at k = 3.

Figure 17 shows the Silhouette analysis for K-Means clustering with k = 3 and 4. The average silhouette scores for both cases are comparable, but k = 4 provides a more uniform distribution of data points across clusters, indicating improved separation.

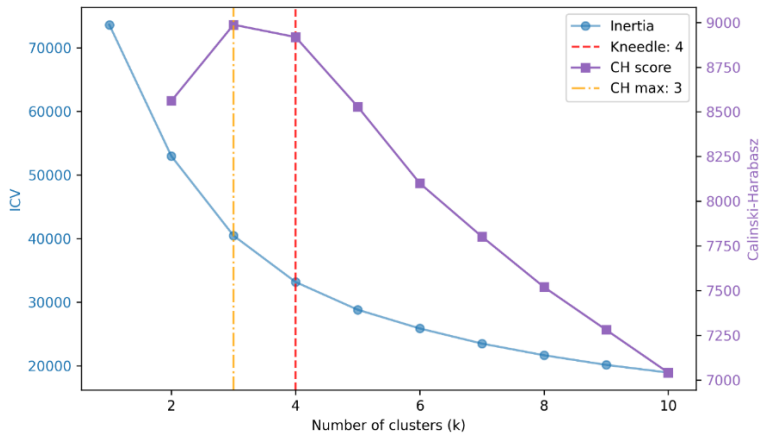


Figure 16: Ideal number of clusters using Kneedle and CH

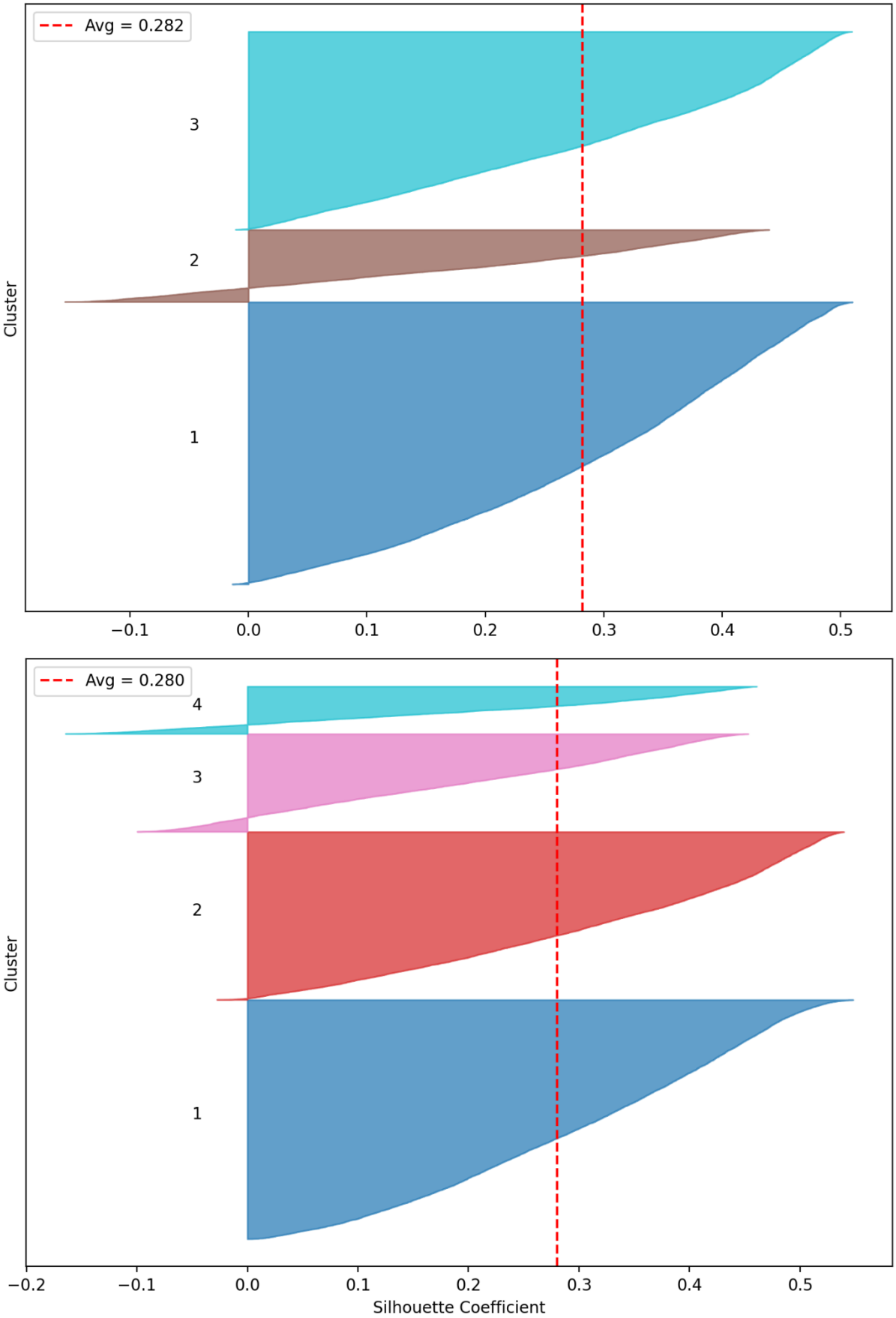


Figure 17: Silhouette analysis (k = 3 & 4)

The discrepancy between these methods arises because they evaluate cluster quality using different criteria. The CH score favors more compact and well-separated clusters, which can lead to a smaller number of clusters. However, a lower number of clusters may oversimplify the underlying structure of the solar wind data.

In the present study, **k = 4** was selected as it provides a better representation of the physical variability in solar wind structures. Accordingly, K-Means clustering was applied to the SOM nodes for 4 clusters, and the resulting clustered map is shown in the figure below.

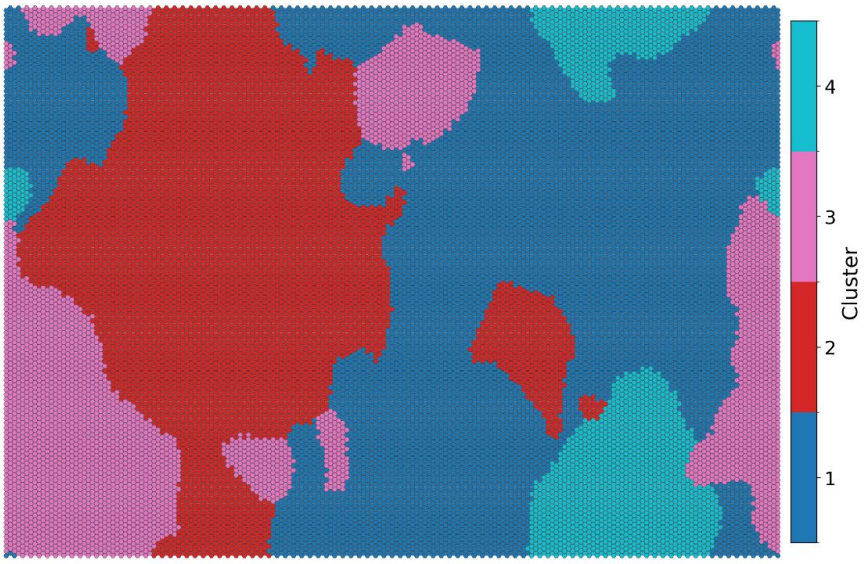


Figure 18: Cluster map

The resulting clusters provide a carse-grained representation of solar wind regimes. To interpret their physical meaning, a detailed analysis of cluster characteristics is presented in the following section.

3.2. Cluster Analysis

Since each SOM node represents the local average of datapoints with similar properties, each solar wind datapoint can be associated with one cluster and therefore, it is possible to map different parts of time series to different clusters. Figure 19 shows the feature maps overlayed with cluster boundaries and ICME nodes, and Figure 20 shows the clustered time series plot for July 2004.

Cluster 1 (Blue) is characterized by low wind velocity, relatively higher proton density, and low proton specific entropy, and is identified as slow solar wind. Cluster 2 (Red) exhibits lower density along with higher wind speed and proton specific entropy compared to Cluster 1, indicating a more dynamic flow regime, identified as high Alfvenic wind. Cluster 3 (Pink) represents extreme solar wind transients such as Coronal Mass Ejection (CME) ejecta and is characterized by notably low plasma beta, suggesting strong magnetic field dominance. Cluster 4 (Cyan) shows very high proton density and low bulk speed, which is typical of compressed plasma regions and is associated with compressed solar wind in Stream Interaction Regions

(SIRs), where slower wind is compressed by faster incoming flows. Figure 22 shows the bar plots and histograms depicting feature distribution in each cluster.

Based on visual inspection of time series and average feature values for each cluster, the clusters were named as follows:

- Cluster 1 (Blue): Slow solar wind (SSW)
- Cluster 2 (Red): Highly Alfvenic wind (HAW)
- Cluster 3 (Pink): Ejecta
- Cluster 4 (Cyan): Compressed slow wind (CSW)

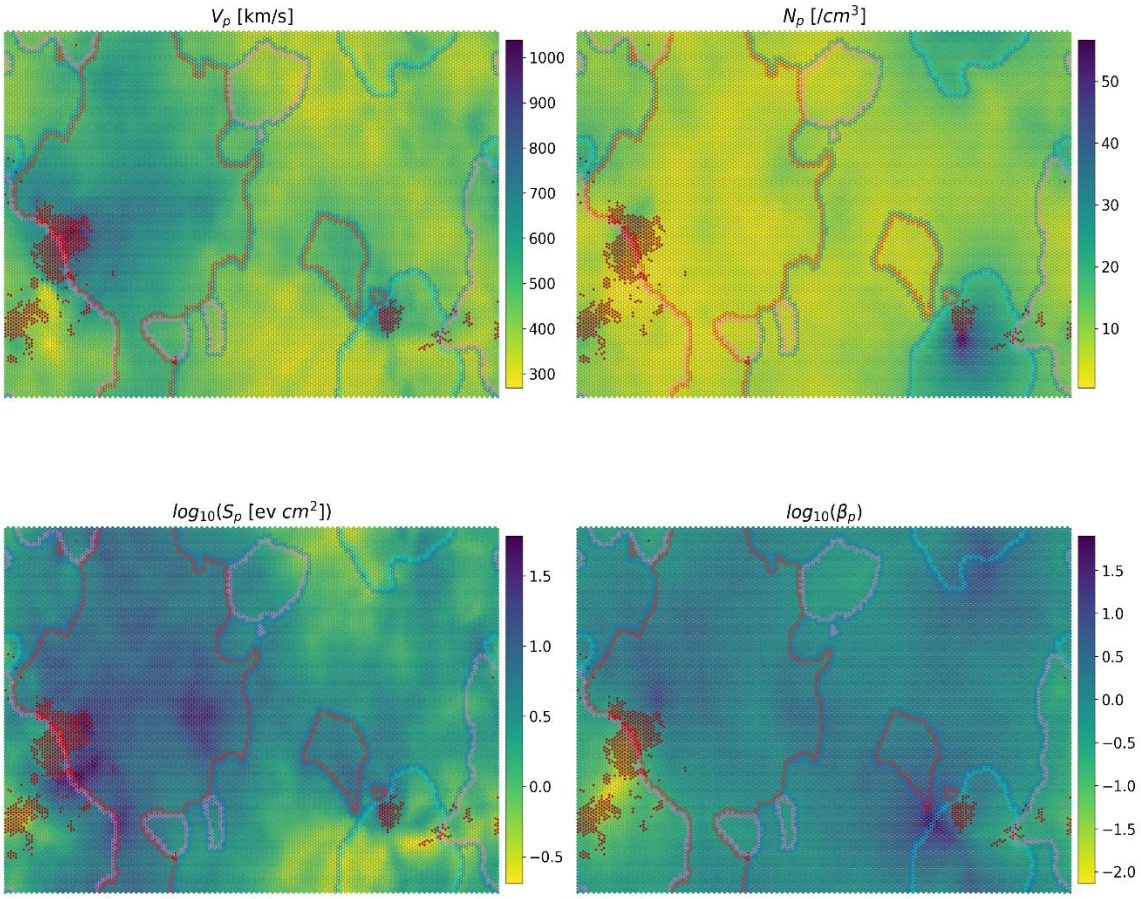


Figure 19: Feature maps with cluster boundaries

To further characterize the dynamical nature of these clusters, Alfvenicity is evaluated as a measure of correlation between the variations in proton velocity and magnetic field components.

$$C_{vb} = \frac{1}{3}\sum_{i} C_{vb,i} \;\; with \; C_{vb,i} = \frac{\sum_j \left(V_{i,j} - \bar{V}_i\right)\left(B_{i,j} - \bar{B}_i\right)}{\sqrt{\sum_j \left(V_{i,j} - \bar{V}_i\right)^2 \left(B_{i,j} - \bar{B}_i\right)^2}}$$

Where $V_{i,j}$ and $B_{i,j}$ are the j^{th} sample of velocity and magnetic field's i^{th} (corresponds to the x, y, and z) component. $\bar{V}_i$ and $\bar{B}_i$ are their values averaged over a time window of 30-minutes. The correlation coefficient i.e. C_{vb} indicates the degree of correlation between velocity and

magnetic field fluctuations, with higher values corresponding to stronger Alfvenic behaviour. As shown in Figure 21, the HAW cluster exhibits the highest value of C_{vb} (approximately 0.7), followed by SSW, while the Ejecta cluster shows the lowest values, indicating weak Alfvenicity.

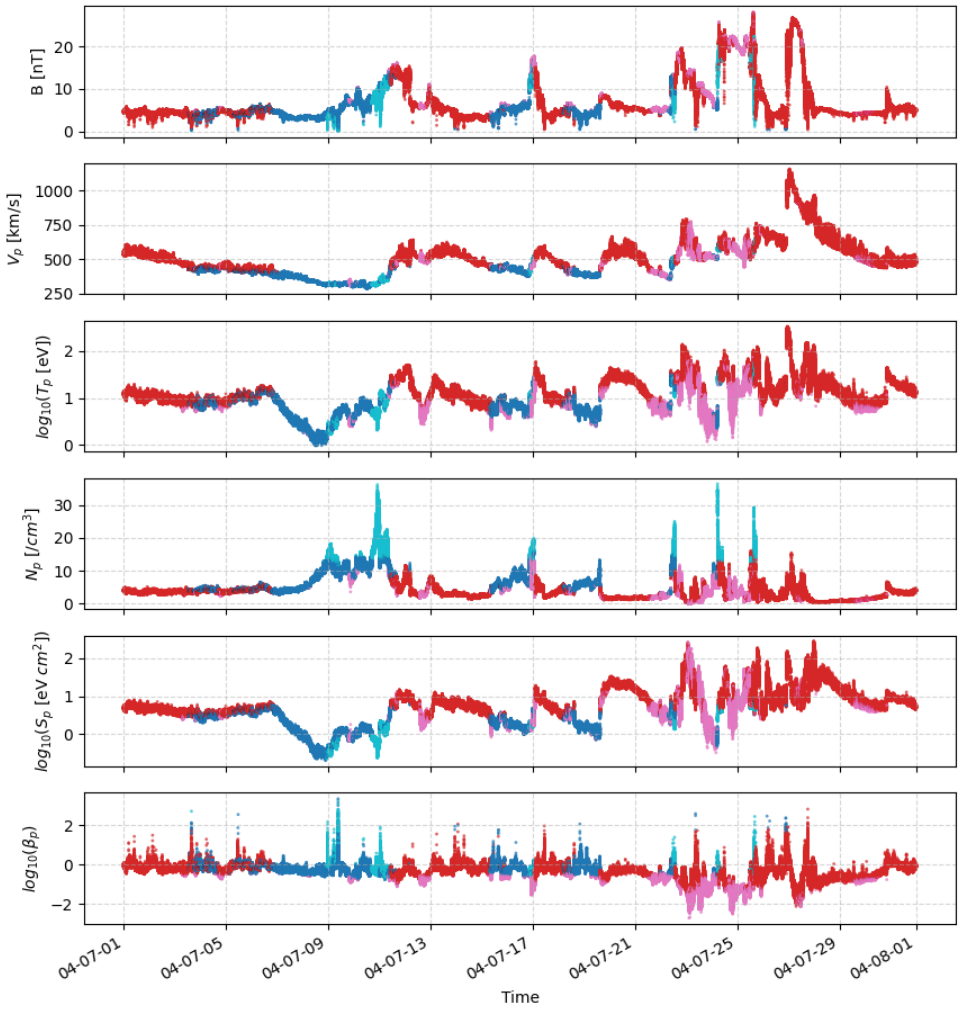


Figure 20: Clustered timeseries, July 2004

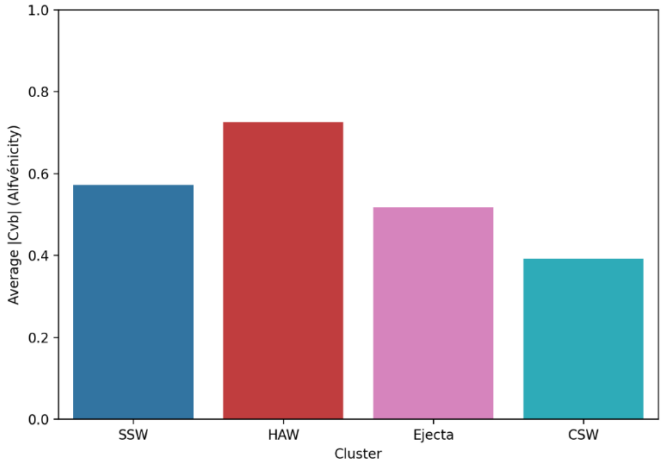


Figure 21: Alfvenicity of each cluster

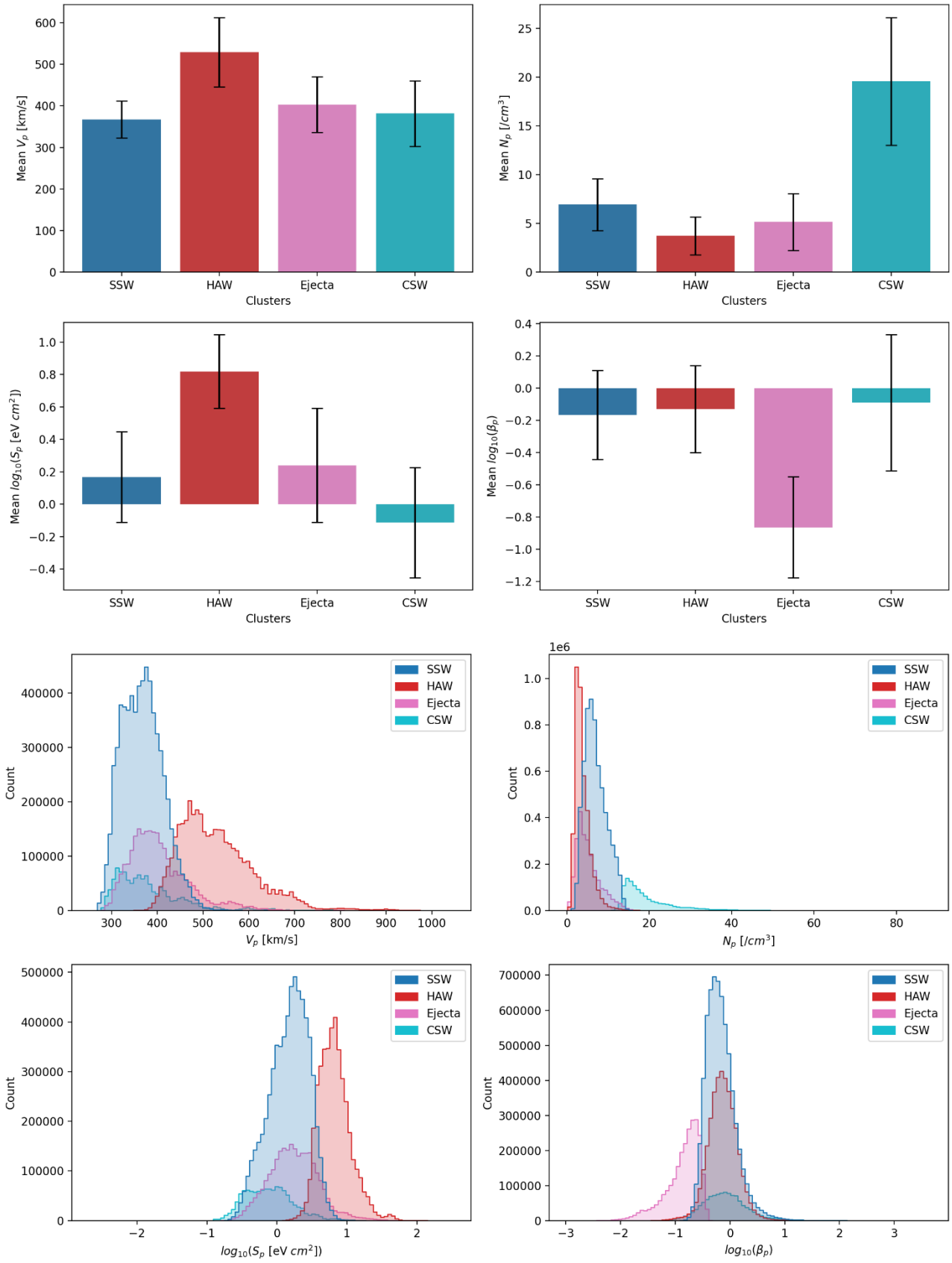


Figure 22: Feature distribution for each cluster

3.3. ICME Data within Cluster

Before quantifying the fraction of ICME datapoints across clusters, it is necessary to define criteria for identifying ICME-dominated SOM nodes. For this, ICME fraction across SOM nodes was visualized using histogram, as shown in Figure 23. The distribution of ICME datapoints across SOM nodes is bimodal, hence it was decided to define two percentage thresholds: one for non-ICME nodes (0.05%) and one for ICME nodes (95%) for better performance of the model.

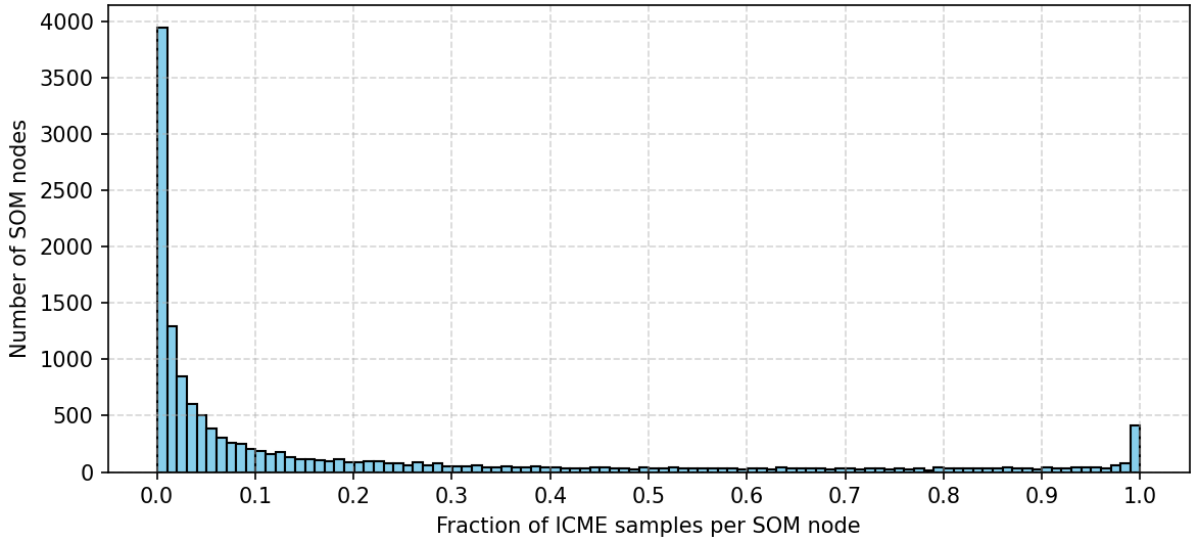


Figure 23: ICME fraction distribution across SOM nodes

Nodes with ICME fraction >= 95% were considered as pure ICME nodes and are highlighted in Figure 14, Figure 15, and Figure 19. On visual inspection of the distribution of ICME nodes across the SOM, it can be seen that the Ejecta (Pink) cluster is populated the most by ICME nodes, followed by HAW (Red) cluster, with CSW (Cyan) and SSW (Blue) clusters being the least populated.

To quantify this data, relative ICME content i.e. how many datapoints of the cluster are ICME w.r.t. total ICME points in all clusters (%) and ICME within cluster i.e. how many points of the cluster are ICME w.r.t. total points in the same cluster (%) were calculated and are given in the table below.

Cluster	ICME within cluster (%)	Relative ICME content (%)
SSW	03.42	13.71
HAW	06.91	17.03
Ejecta	34.61	53.64
CSW	21.82	15.60

Table 2: ICME data within cluster

The results show that ICME datapoints are distributed across all clusters, with the highest concentration in the Ejecta cluster, followed by HAW and CSW, while SSW shows the lowest ICME presence. The Ejecta cluster has 34% of its datapoints associated with ICMEs.

3.4. Comparison with Empirical Clustering Method

The Xu & Borovsky classification method is an established empirical scheme used to classify solar wind plasma based on its origin. It is based on three parameters:

1. Alfven Speed:

$$V_A = \frac{B}{\sqrt{4\pi m_p N_p}}$$

2. Proton Specific Entropy:

$$S_p = \frac{T_p}{N_p^{2/3}}$$

3. Temperature ratio:

$$T_{exp} = \left(\frac{V_p}{283}\right)^{3.113}$$

Where V_p, m_p, N_p and T_p are proton velocity, mass, density and temperature, B is the magnetic field strength [3]. There are four categories in this classification method: Streamer Belt (SB), Sector Reversal (SR), Coronal Hole (CH), and Ejecta, based on the origin of solar wind and defined by conditions given in Table 3. SB denotes wind originating from streamer belts and is associated with slow solar wind, SR is wind originating from sector reversal region, CH is wind originating from coronal holes or fast solar wind, and Ejecta includes magnetic clouds and ICMEs.

Category	Codition
Ejecta	$\log_{10}V_A > 0.277\log_{10}S_p + 0.055\log_{10}(T_{exp}/T_p) + 1.83$
CH	$\log_{10}S_p > -0.525\log_{10}(T_{exp}/T_p) - 0.676\log_{10}V_A + 1.74$, not Ejecta
SR	$\log_{10}S_p < -0.658\log_{10}V_A - 0.125\log_{10}(T_{exp}/T_p) + 1.04$, not Ejecta
SB	Data points which do not satisfy any of the above conditions.

Table 3: Xu & Borovsky categories

The same classification was applied to the WIND dataset, and the results were compared with the SOM-derived clusters. The row-normalized matrix in Figure 24 shows the percentage of each Xu & Borovsky category in our cluster where each row sums up to 1. It can be observed that the HAW cluster mainly maps to Coronal Hole (CH) and partially to Streamer Belt (SB). The Ejecta cluster corresponds very well to Xu & Borovsky's Ejecta category. SSW is mainly composed of SB and CSW is a mixture of SB and Sector Reversal (SR).

The column-normalized matrix, on the other hand, shows the contribution of each SOM cluster within a given Xu & Borovsky category. This representation highlights that CH wind is almost entirely captured by the HAW cluster, while the Ejecta category is predominantly mapped to the SOM Ejecta cluster. SB wind is distributed mainly between SSW and CSW clusters, and SR is primarily associated with CSW.

Together, these results demonstrate a strong overall agreement between the SOM-based clustering and the empirical classification, while also revealing subtle differences in how transitional and mixed solar wind regimes are identified.

To further analyse the differences in SOM and Xu & Borovsky clusters, distribution of ICME datapoints throughout the clusters was calculated. The results are shown in the table below.

Cluster	ICME within cluster (%)	Relative ICME content (%)
SB	05.63	23.26
CH	06.72	11.86
Ejecta	55.96	53.16
SR	06.19	11.72

Table 4: ICME data within Xu & Borovsky clusters

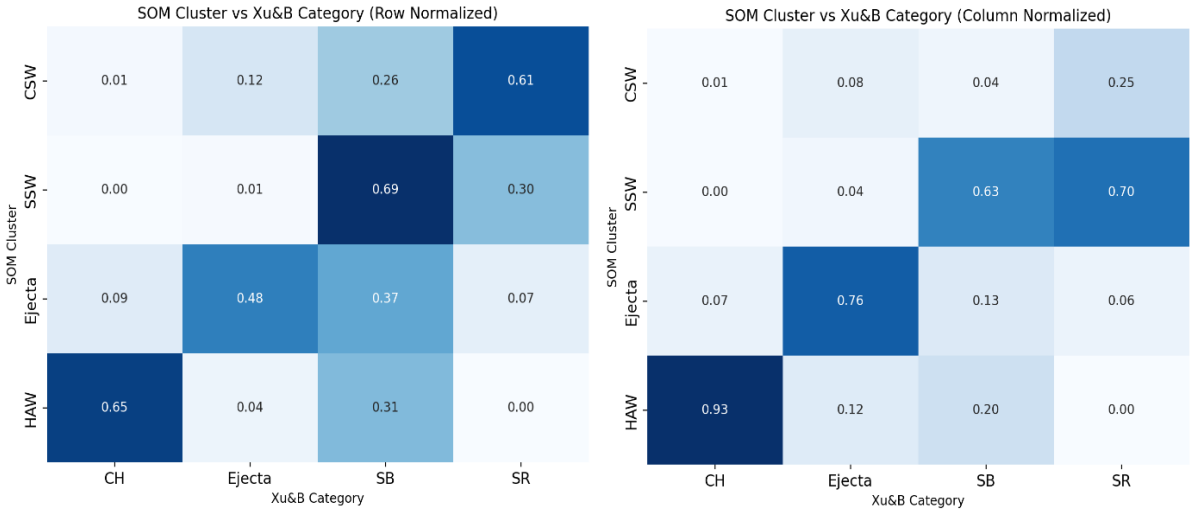


Figure 24: Comparison heatmaps

The SOM clusters show comparable relative ICME content to the Xu & Borovsky classification for the Ejecta category (~53%), where both methods show strong agreement. For the remaining clusters, Som distributes ICME-associated datapoints across HAW (17.03%), SSW (13.71%), and CSW (15.60%), whereas Xu & Borovsky method shows a higher concentration in the SB category (23.26%) and lower concentrations in CH (11.86%) and SR (11.72%). This indicates that the SOM-based approach is able to capture ICME signatures embedded within multiple solar wind regimes.

The ICME fraction within clusters is highest for the Ejecta cluster (34.61%), while remaining relatively low and comparable across CSW (21.82%), SSW (3.42%) and HAW (6.91%). This confirms that the Ejecta cluster is strongly associated with ICME events.

Overall, the clusters obtained using SOM show good agreement with Xu & Borovsky classification. The observed differences arise because SOM is designed to identify patterns based on data structure and transient signatures, whereas the Xu & Borovsky scheme classifies solar wind based on its origin. Thus, SOM provides a complementary perspective that is particularly sensitive to dynamic solar wind features.

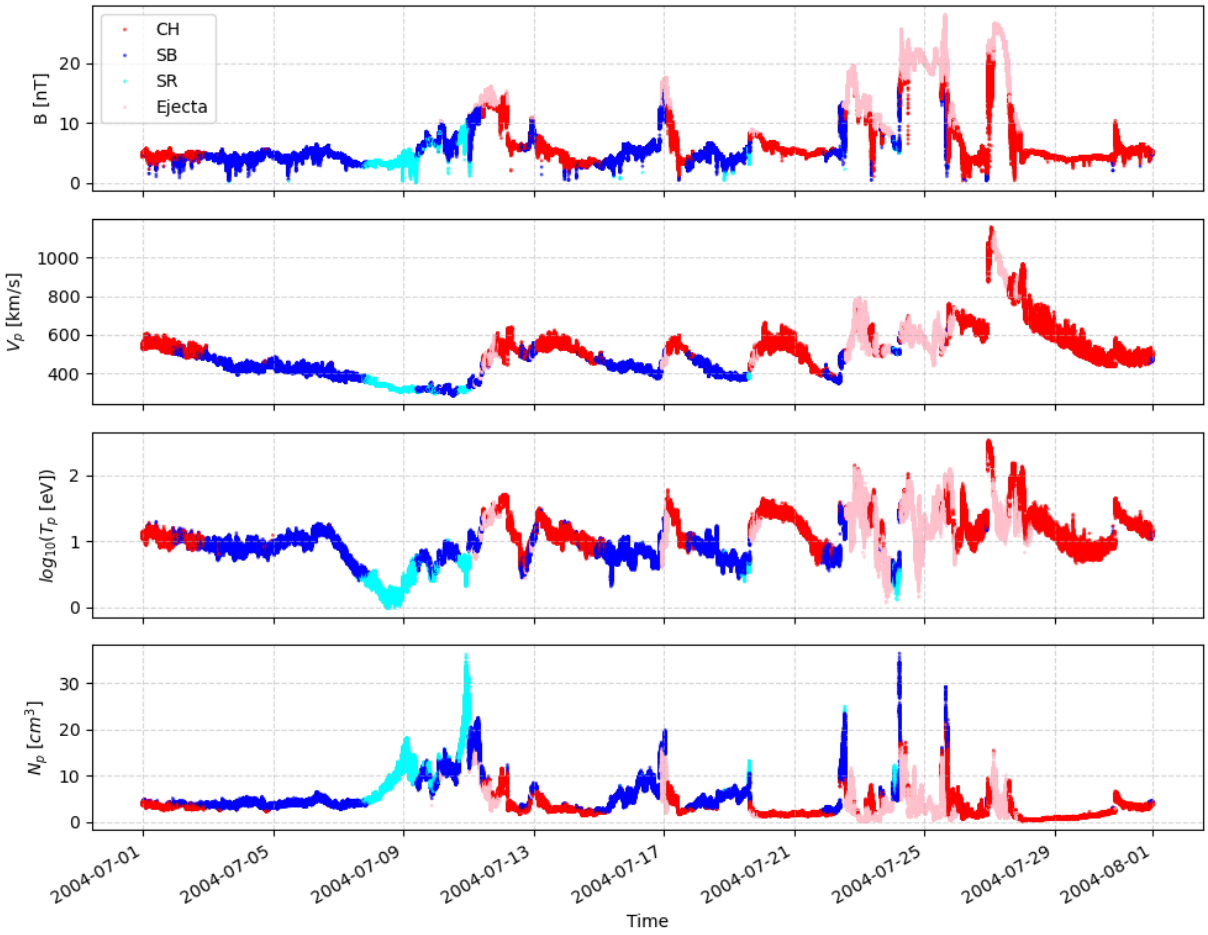


Figure 25: Xu & Borovsky timeseries

Having established the physical interpretation and validity of the SOM clusters, the next chapter focuses on testing the model on independent datasets to evaluate its robustness and generalization capability.

Chapter - 4 : Case Studies – Solar Wind Classification

This chapter evaluates the performance of the trained SOM model on independent datasets to assess its robustness in classifying solar wind structures and detecting ICMEs. The model is first applied to WIND data from 2016 which represents the declining phase of the solar cycle, and its clustering behaviour is analysed through activation patterns, time series plots, and comparison with Xu & Borovsky classification. Model performance is further quantified using event-wise and datapoint-wise metrics such as precision, recall, and F1-score. The trained SOM is then tested on Aditya-L1 data from 2024 and 2025, corresponding to the rising and peak phases of solar cycle 2025, demonstrating its ability to adapt to varying heliospheric conditions. Finally, the model is applied to SWASTi simulation data to evaluate its consistency across multiple vantage points. These analyses establish the generalization capability of the SOM-based classification framework.

4.1. WIND: 2016

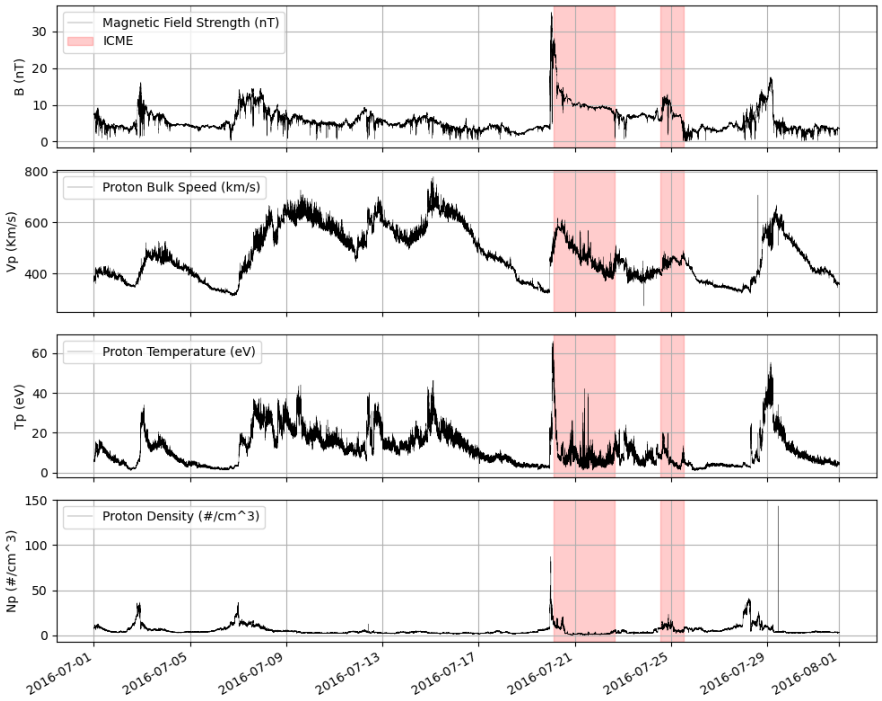


Figure 26: July 2016 WIND data

ICME intervals (red rectangles): "July 20, 2016, 02:53 – July 22, 2016, 15:31", and "July 24, 2016 13:51 – July 25, 2016 13:08"

To test the model's performance, it was applied to WIND data from July to December 2016, which corresponds to the declining phase of solar cycle 24 (post the 2014-15 peak). The BMU for each input vector was identified as the neuron with minimum Euclidean distance to the trained weight vectors. The new dataset was processed the same way as the 2004 dataset. Figure 26 shows the hit density map, i.e. how many times a SOM node was activated, with cluster boundaries and ICME nodes (from trained map) overlaid.

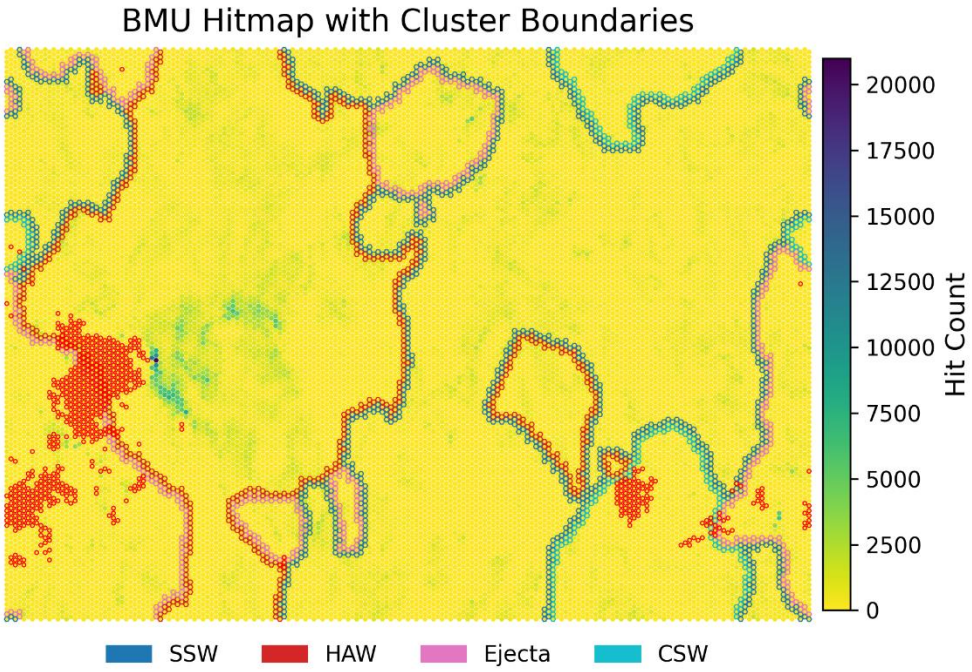


Figure 27: WIND 2016 hitmap

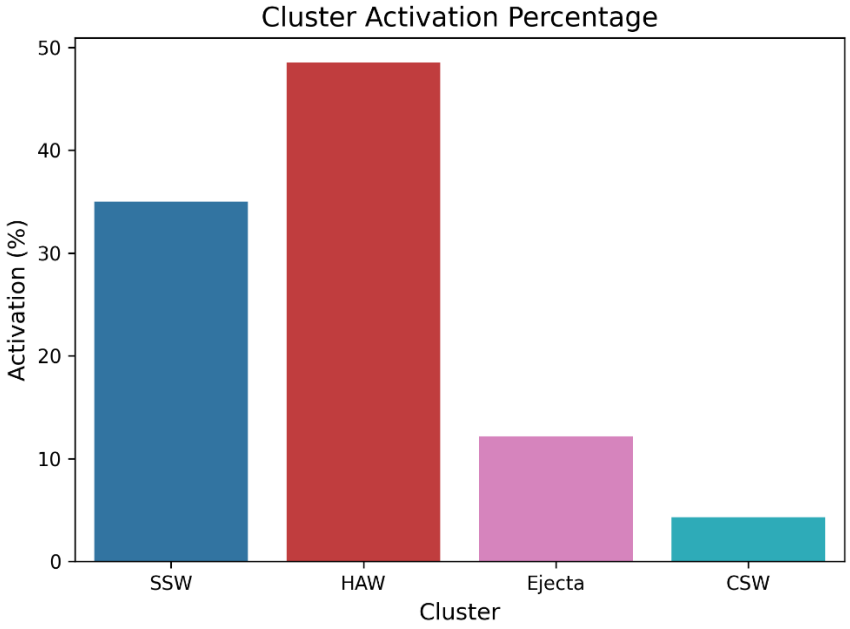


Figure 28: Activation per cluster

The bar plot shows activation percentage of each SOM cluster. Most of the new datapoints fall into the HAW cluster with an activation of ~49%, which is associated with high-speed streams originating from coronal holes. The second most activated cluster was SSW (~35%) representing background slow solar wind, followed by Ejecta (~11%). CSW is least activated (~4%).

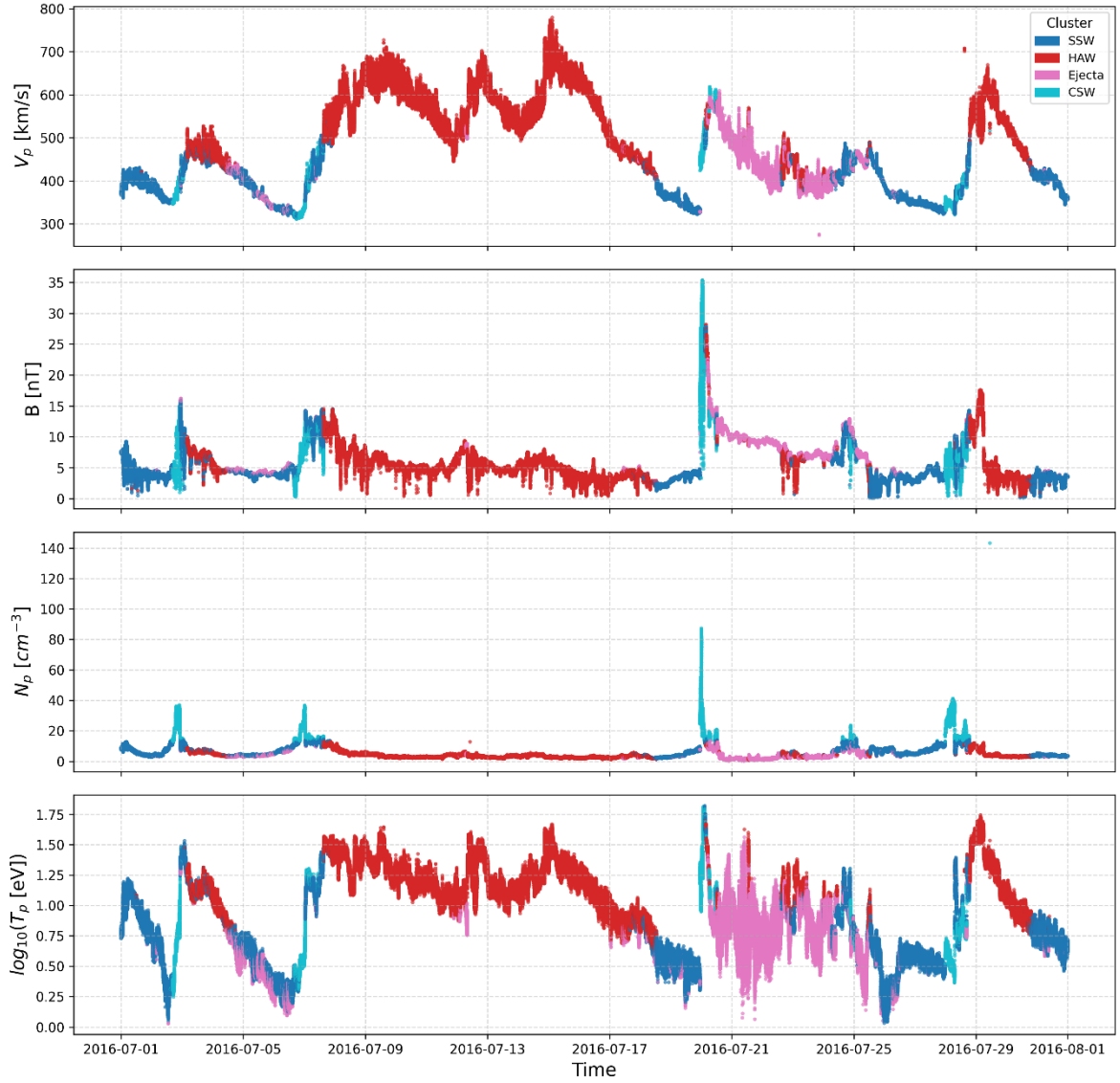


Figure 29: Clustered timeseries (WIND, 2016)

The above figure shows the clustered timeseries obtained from the SOM, the map correctly identifies the two ICME intervals in July 2016 [Figure 26] as ejecta. To further verify this clustering, it was compared to the Xu & Borovsky categories again. From Figure 30, it is clear that the HAW cluster maps to CH very well.

This distribution indicates that the solar wind during this period is dominated by highly Alfvenic wind, with a significant contribution from slow solar wind. This is consistent with the known heliospheric conditions during the declining phase of the solar cycle. During this phase,

the solar wind is dominated by slow solar wind along with coronal hole driven high-speed wind. This agreement indicates that the model captures meaningful physical patterns in the solar wind.

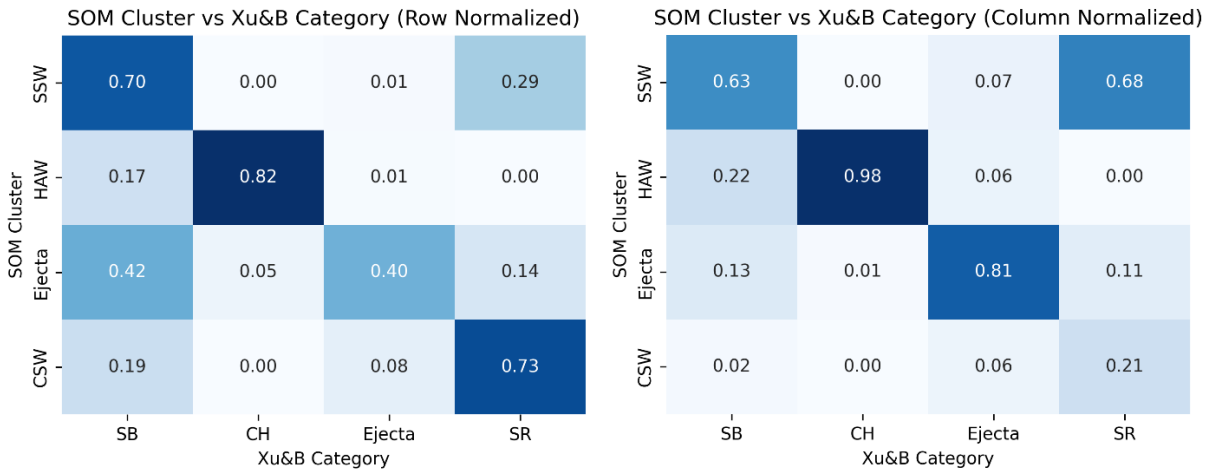


Figure 30: Comparison heatmaps (WIND, 2016)

The datapoints that fell into ICME nodes as defined after training were assigned ICME flags (0: no ICME and 1: ICME) and event-wise recall, precision, and F1-score of the model was calculated which are discussed below.

Model Performance Analysis

Figure 32 shows the raw predicted ICME flags obtained from SOM compared with true ICME flags without any processing. It can be seen that the model predicts some extremely short ICME intervals which are physically not possible. For this, datapoint-wise performance parameters were calculated. To make the output ICME flags given by SOM a continuous series of 1s or 0s to show time series of ICME event and quiet time, the first step was to fill gaps between series of 1s. For this a time window of 5 hours was given such that if another flag 1 was repeated within 5 hours then all 0s within this gap will be changed to 1. The second step was to reduce false positives. For this, a series of flag 1 was removed if it was less than a duration of 6 hours. Figure 33 shows a visualisation of the predicted and true ICME flags after this processing. The model shows 4 out of 6 ICME events detected correctly without any false positives.

These durations were selected based upon the histogram of all ICME durations available in the WIND ICME catalog [Figure 31]. Since no ICMEs in the catalog lasted less than 5.67 hours and only 5% of the total ICMEs lasted for less than 10 hours, ICMEs below 6 hours can be safely removed.

Table 5 shows the following performance parameters:

1. *Accuracy* defines the overall correctness of the model and is given by the fraction of true predictions and total predictions, i.e.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2. *Precision* is defined as the percentage of true positives (or negatives) among all predicted positives (or negatives).

$$Precision = \frac{TP}{TP + FP}$$

3. *Recall* is the percentage of true positive (or negative) in actual class 1 (or 0).

$$Recall = \frac{TP}{TP + FN}$$

4. *F1-score* is computed as the harmonic mean of precision and recall. When F1 is high, both the parameters are high.

$$F1\ Score = \frac{2 \times precision \times recall}{precision + recall}$$

Where true positives (TP) and true negative (TN) are correctly predicted flags (1 or 0) while false positives (FP) and false negatives (FN) are incorrect predictions.

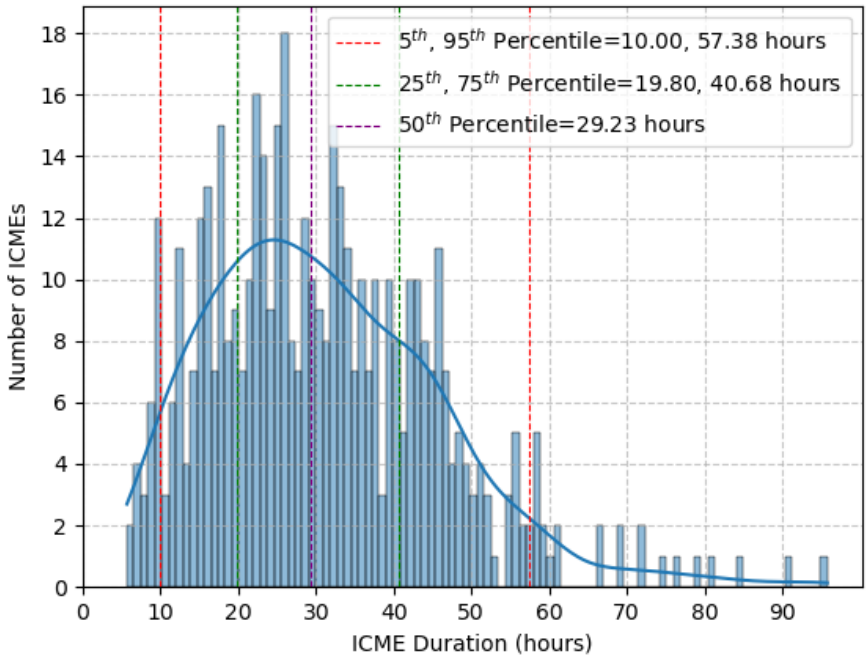


Figure 31: ICME duration distribution

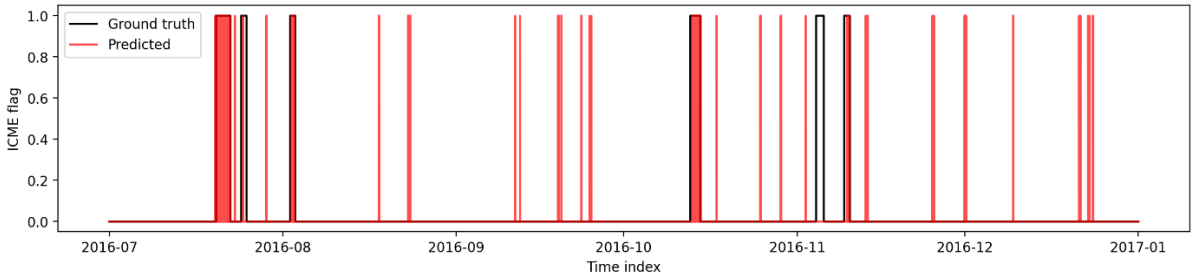


Figure 32: Predicted ICMEs and ground truth (raw)

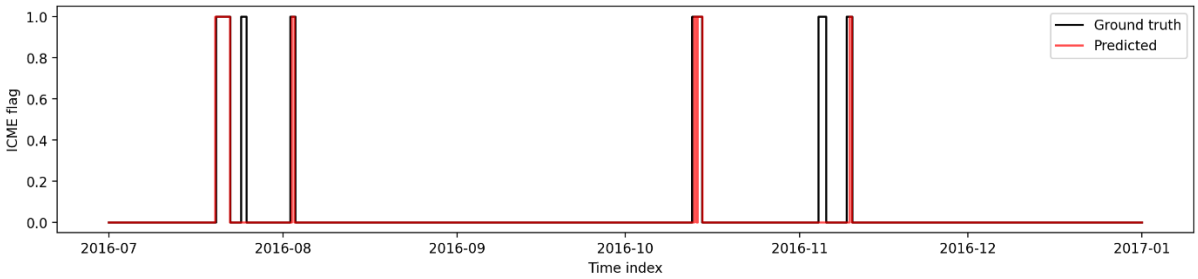


Figure 33: Predicted ICMEs and ground truth (after processing)

S.No.	Parameter	Event-wise (%)	Datapoint-wise (%)
1	Accuracy	NA	95.8
2	Precision	100	94.1
3	Recall	66.67	12.6
4	F1-score	80.02	22.2

Table 5: Event-wise and datapoint-wise model performance parameters

Event-wise accuracy is not defined as true negatives are not meaningful at event scale.

The lower values of recall and F1-score in the raw predictions are expected, as SOM assigns ICME flags at the datapoint level, resulting in fragmented detections and increased false positives and negatives. This is why filling short gaps and removing short ICME intervals become necessary as this makes continuous series of 1s and 0s which help in calculating event-wise parameters.

4.2. Aditya-L1: 2024 & 2025

After analysing the model's performance, it was implemented on SWIS and MAG data from Aditya-L1 for the period May 2024 – November 2025 with a cadence of 10s.

4.2.1. 2024 Data

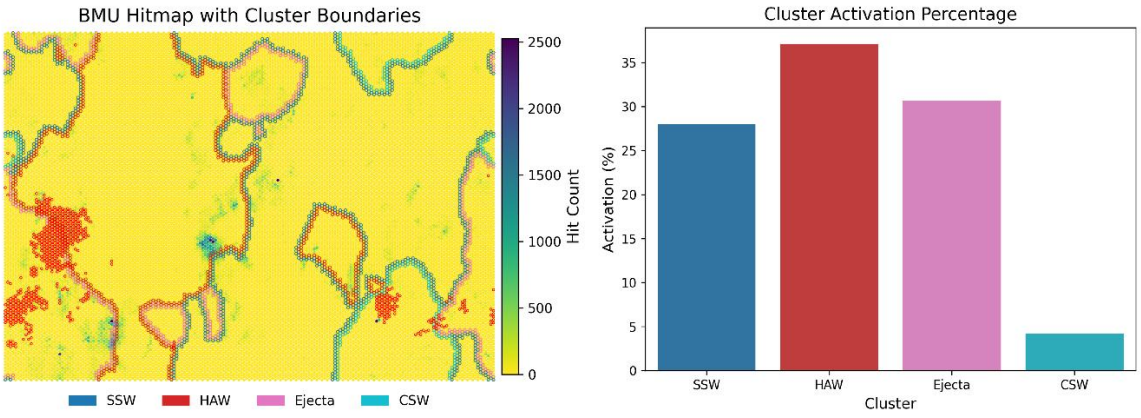


Figure 34: SOM implementation on Aditya-L1 2024 data

From the above figure, the HAW cluster is still the most activated with ~36% activation which is lower than its activation in the 2016 WIND sample. The second most activated cluster is Ejecta (~30%) which is significantly higher compared to its activation in 2016 WIND dataset. The least activated cluster is CSW (~4%) which is comparable to that in 2016 while SSW (~28%) shows a slight increase in its activation. Since this time period corresponds to the rising/peak phase of solar cycle 25, these results confirm that slow solar wind and events like ICMEs increase in the active phase of the cycle along with decrease in HAW and CSW due to unstable coronal holes.



Figure 35: Clustered timeseries (Aditya-L1, May 2024)

Figure 35 shows clustered timeseries as obtained from SOM for the month of May 2024. It shows a mixture of high alfvenic and compressed slow wind during the ICME event of May 10 – May 12, 2024 with some portions identified as Ejecta. This can be attributed to, as mentioned by Xu & Borovsky, ejecta has two populations: true ejecta from the Sun, and one population resembling CH and SR type plasma based on its properties [3]. The comparison

heatmaps shown in the figure below show that the SOM and Xu & Borovsky derived categories are in good agreement with each other.

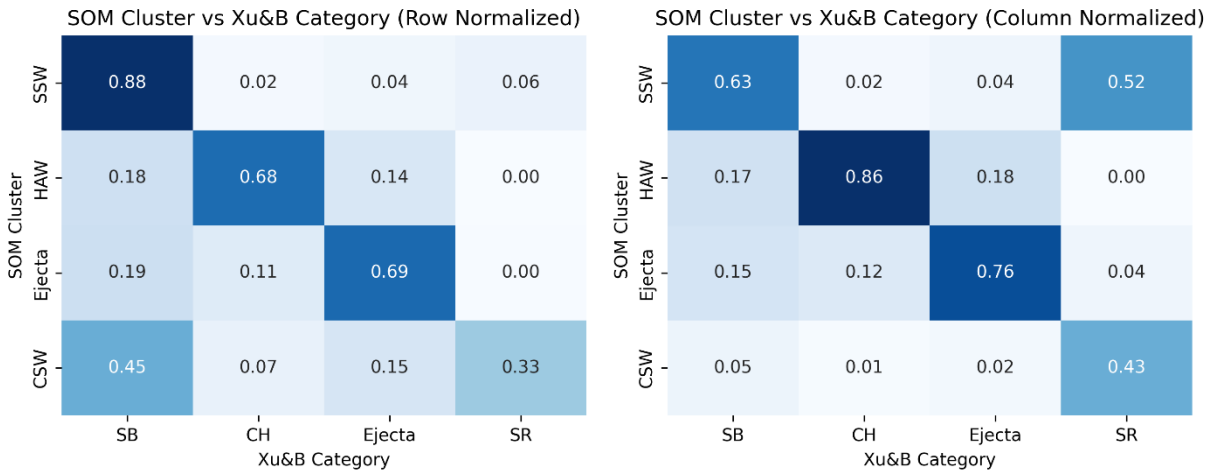


Figure 36: Comparison heatmaps (Aditya-L1, 2024)

Figure 37 shows ICME predictions for Aditya-L1 data. The ground truth line shows one ICME at 10th May 2024 which is one of the last few entries in the catalog. The last entry in the catalog was of 31st July 2024, all ICME related datapoints in between this time interval were removed because of gap or low-quality data during preprocessing. The other ICME intervals predicted by SOM match the ICME intervals mentioned in Richardson & Cane ICME list.

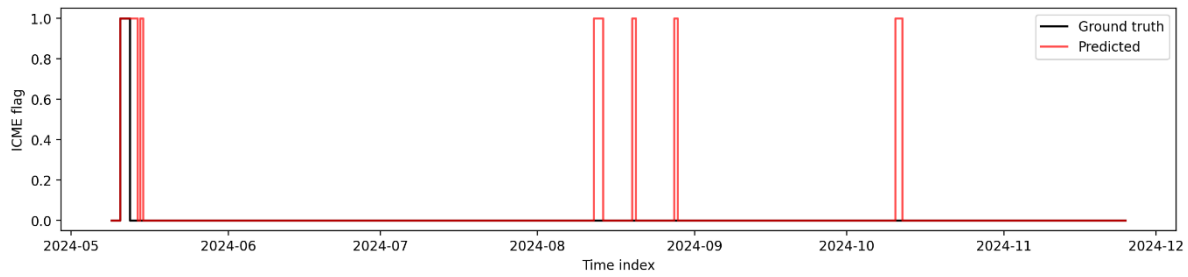


Figure 37: Aditya-L1 ICME predictions

4.2.2. 2025 Data

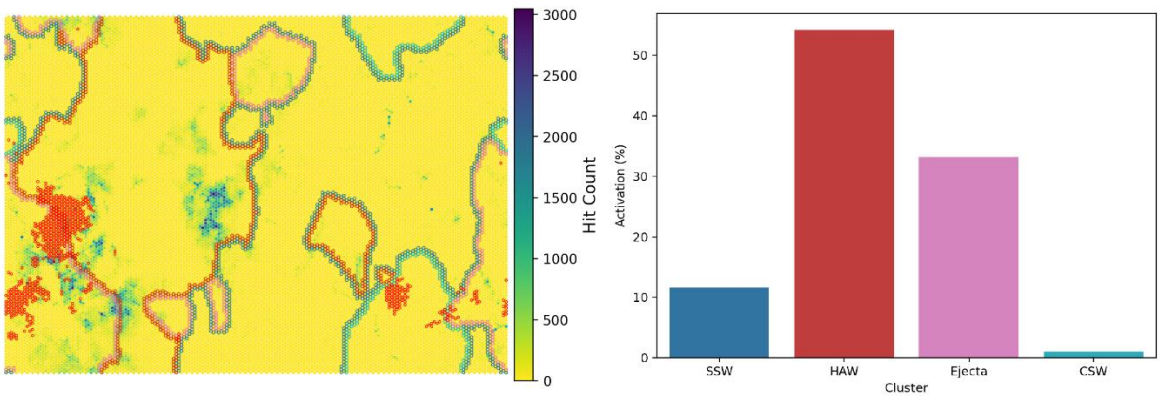


Figure 38: SOM implementation on Aditya-L1 2025 data

Figure 38, Figure 39, and Figure 40 show the BMU hitmap and activation bar chart, clustered timeseries for June 2025, and Xu & Borovsky comparison heatmaps for the 2025 dataset from Aditya-L1. The most activated cluster in this case too, is the HAW (~53%) cluster. Both SSW (~11%) and CSW (~1%) show decreased activation while Ejecta (~32%) remains the same. This is consistent with the heliospheric conditions during solar maxima.

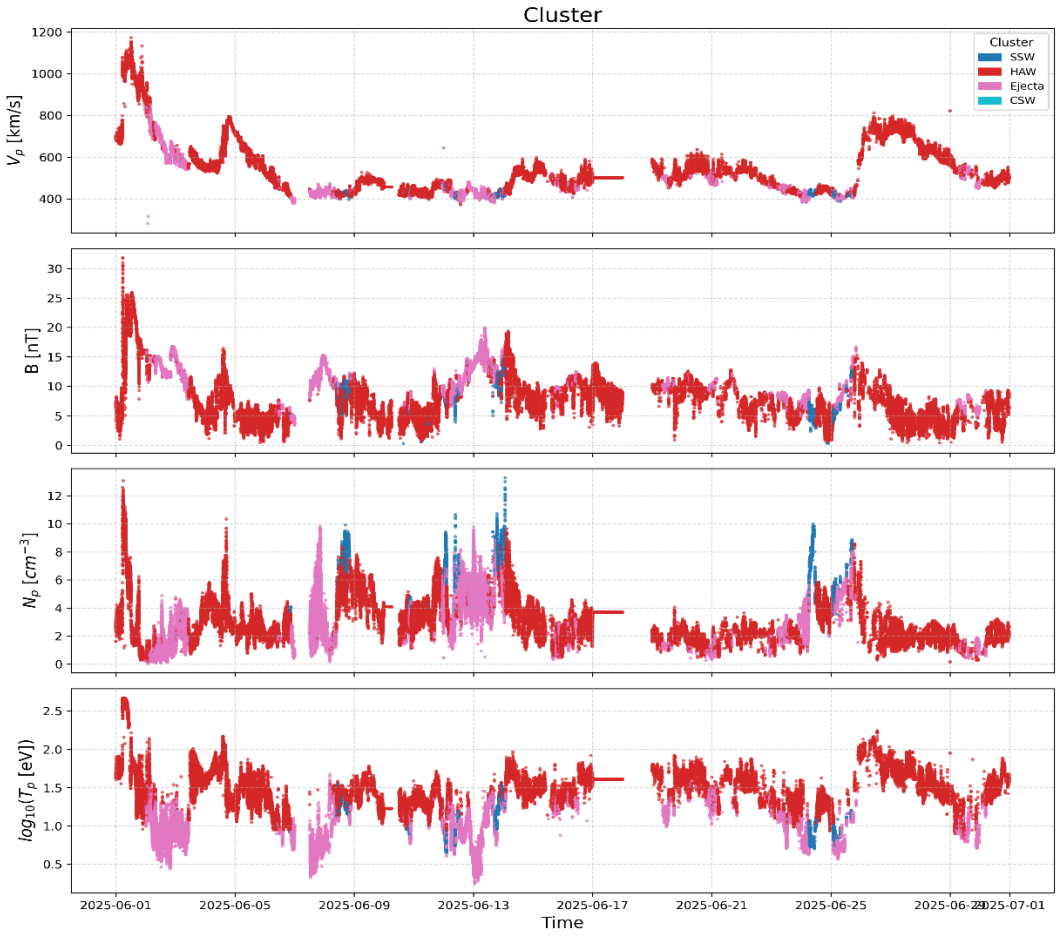


Figure 39: Clustered timeseries (Aditya-L1, June 2025)

The clustered timeseries show three major ICME events which are mentioned in Richardson & Cane ICME list: “June 03, 2025, 02:00 – June 03, 2025, 14:00”, “June 07, 2025, 10:00 – June 08, 2025, 10:00”, and “June 11, 2025, 21:00 – June 13, 2025, 23:00”. The comparison heatmaps given below show good overall agreements between SOM-based and Xu & Borovsky based clusters.

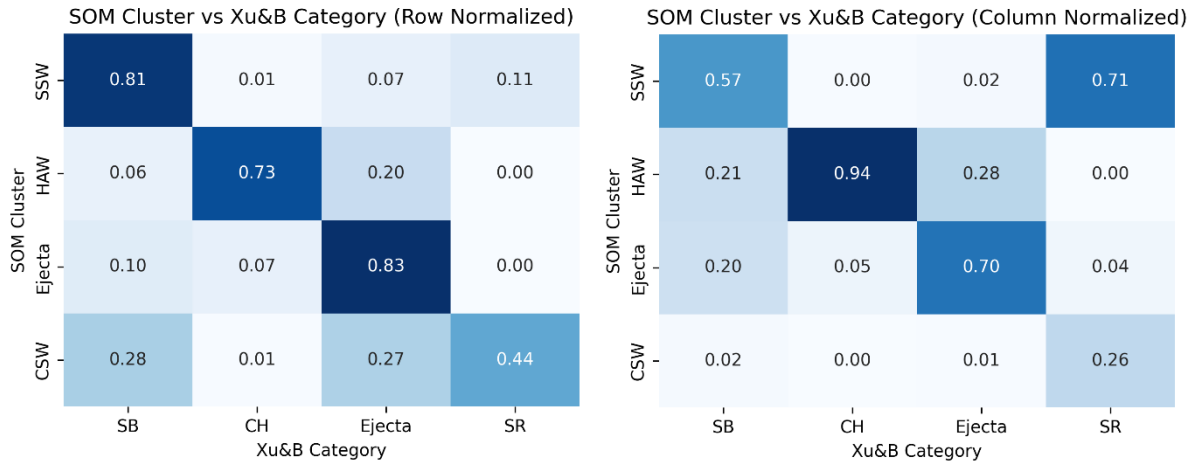


Figure 40: Comparison heatmaps (Aditya-L1, 2025)

Overall, the difference in cluster activation percentages between WIND 2016, Aditya-L1 2024, and Aditya-L1 2025 can be attributed to several factors. First, the two time periods belong to two different phases of the solar cycle with 2016 showing more stable background wind and persistent CH streams, while 2024 – 2025 has more variable slow wind and increased transient activity. Second, both missions have different data cadences which changes the density of input vectors and smooths short duration structures in the Aditya dataset. Finally, the datasets differ in size and continuity. Together, these instrumental and heliospheric differences naturally lead to shifts in percentage distribution across the clusters and in Xu & Borovsky category comparisons.

4.3. SWASTi Simulation: May 2024 event

Space Weather Adaptive SimulaTion (SWASTi) framework is a three-dimensional physics-based heliospheric model intended for space weather analysis and modelling. The SOM was tested on the data obtained from SWASTi simulation of May 2024 event. The dataset presents a multi-vantage point analysis of the CME and solar wind conditions. Four observation points were added: Earth, Parker Solar Probe (PSP), Solar Orbiter, and STEREO-A. The obtained hitmaps, activation barplots, and clustered timeseries are presented in the figures below.

The activation percentages are similar in Earth and STEREO-A simulation dataset. The PSP data shows higher activation percentages for SSW and CSW clusters while that for Ejecta and HAW are lower. The solar orbiter data has no activation for HAW cluster. The activation percentage of SSW and CSW is higher compared to Earth but lower than PSP. The percentage for Ejecta is higher than PSP but lower than Earth.

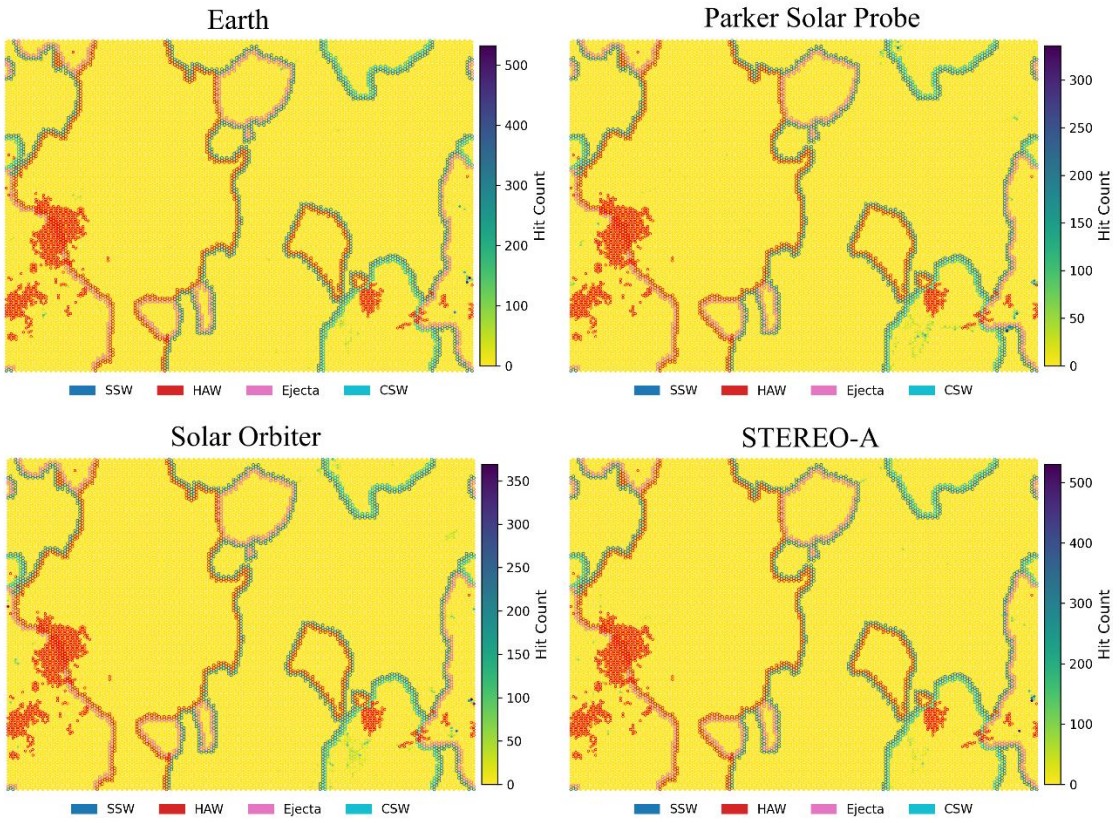


Figure 41: BMU Hitmaps (SWASTi, May 2024)

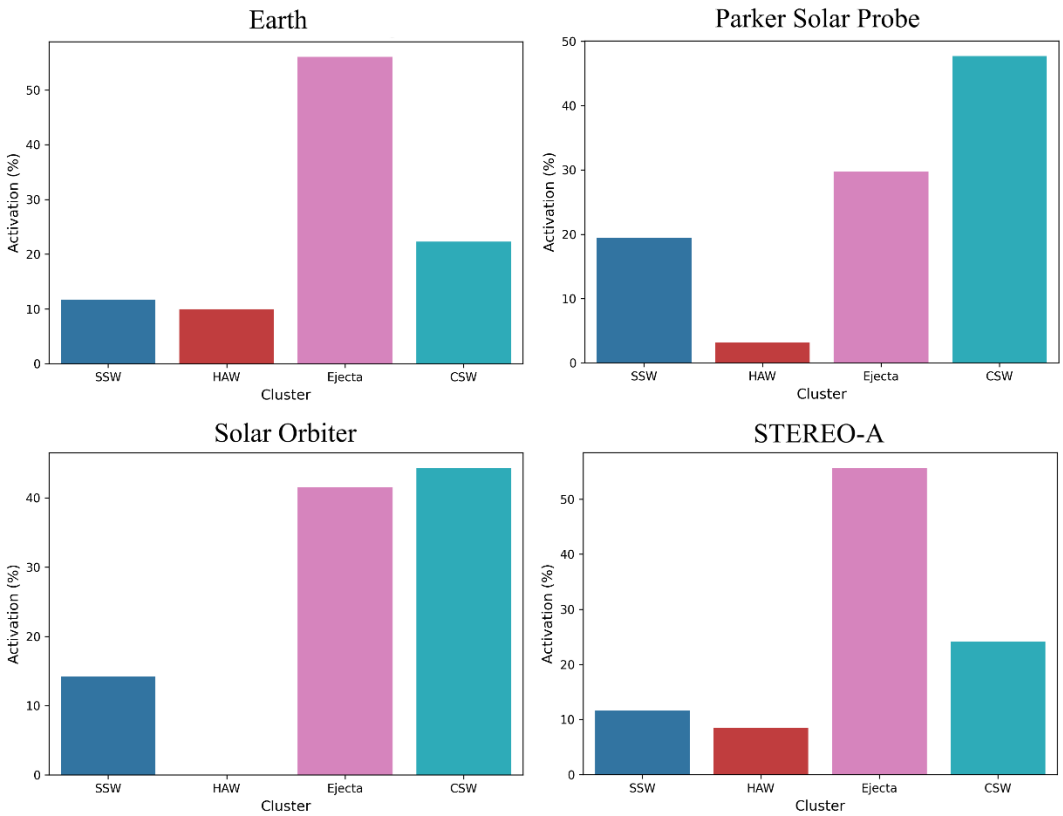


Figure 42: Activation Barplots (SWASTi, May 2024)

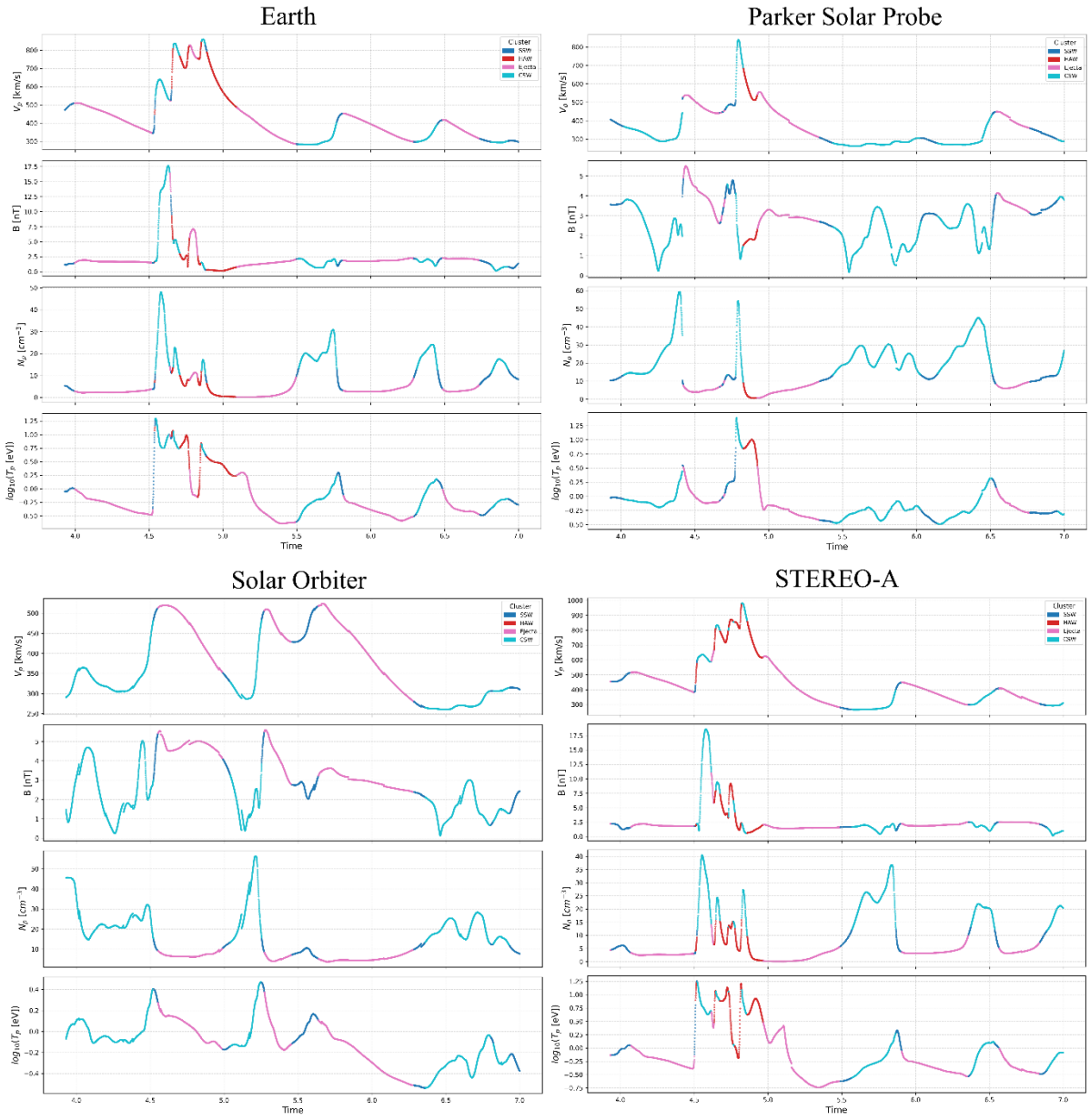


Figure 43: Clustered Timeseries (SWASTi, May 2024)

This highlights the importance of reliable in-situ measurements, which motivates the design and development of diagnostic instruments. Langmuir probes, as discussed in Chapter 1, are among the most widely used particle detectors. They are easy to develop and can be built in a laboratory. The next chapter discusses about the design and fabrication of Langmuir probe and analysis of its I-V characteristic in laboratory.

Chapter - 5 : Langmuir Probe Design and Data Acquisition

This chapter describes the design, development, and experimental implementation of Langmuir probes for plasma diagnostics. It starts with the generation of plasma using a DC glow discharge setup under low-pressure conditions, establishing a stable environment for measurements. The chapter then describes the process of designing and constructing two probe geometries (cylindrical and planar) with emphasis on the design, material selection, and assembly procedures. Next, the approach to measure the I-V characteristics of the probe is described, focusing on the experimental setup and measurement procedure. To remove the noise present in the measured data, data processing methods including baseline correction, Savitzky-Golay (Savgol) filter, and bin averaging are applied. Finally, the I-V characteristics are analysed to determine the key plasma parameters, including plasma potential and floating potential, demonstrating the effectiveness of the developed probes for laboratory plasma diagnostics.

5.1. Generating Plasma

Plasma generation in laboratory settings is typically achieved by ionizing a neutral gas under low-pressure conditions. One of the most fundamental and widely studied plasma sources is the **DC glow discharge**, valued for its simplicity and well-defined structure.

In this method, a direct current (DC) voltage is applied across two electrodes placed inside a vacuum chamber filled with a low-pressure inert gas, commonly Argon or Helium [Figure 44]. The operating pressure usually ranges from 0.1 to 1 mbar. When the applied voltage exceeds the breakdown voltage of the gas, free electrons are accelerated by the electric field and ionize neutral atoms through collisions. This leads to the formation of a self-sustained plasma discharge, characterized by distinct regions such as the cathode glow, dark space, and positive column. The glow discharge provides a stable and reproducible plasma source, making it ideal for diagnostic techniques such as optical emission spectroscopy and Langmuir probe measurements.

In the present work, the plasma chamber was first evacuated to a base pressure of 0.03 mbar using a rotary vacuum pump. The chamber was then filled with **Argon** gas at a working pressure of **0.300 mbar**. A DC voltage supply was applied across the electrodes and increased until plasma breakdown occurred (typically in the range of 300 - 400V). A stable and bright glow discharge was obtained at a discharge voltage of **650 V**, which was used for both remote

sensing and in-situ measurements. The design and development of Langmuir probes used for this work in mentioned in the next section.

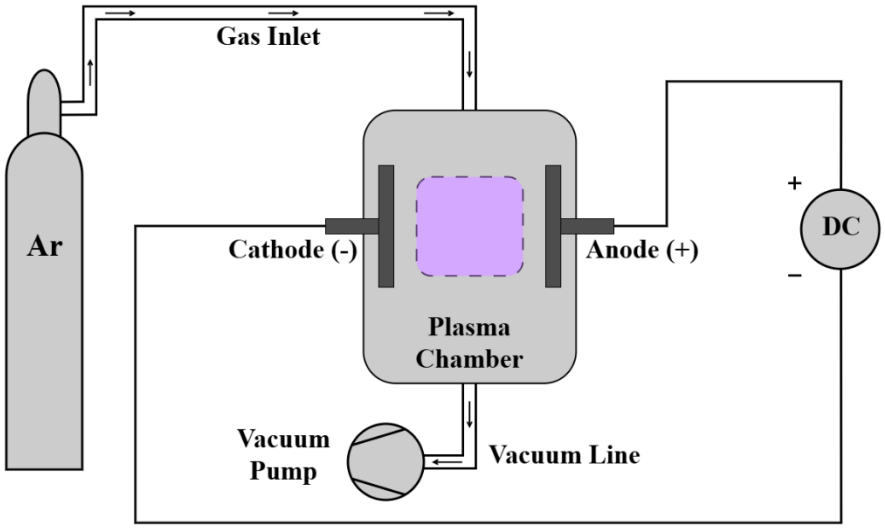


Figure 44: Plasma Generation in Laboratory

5.2. Design and Development of Langmuir Probe

The design and fabrication of the Langmuir probe were carried out with the objective of ensuring mechanical stability, electrical reliability, and minimal perturbation to the plasma. Two probe geometries (cylindrical and planar) were developed to compare their performance in plasma diagnostics.

5.2.1. Cylindrical Langmuir Probe

The cylindrical probe was constructed using a thin tungsten wire as the sensing electrode. The probe tip was connected to an aluminium coaxial cable using a copper wire to ensure good electrical contact. This assembly was then soldered to a BNC mounting flange to facilitate compatibility with measurement instruments. To electrically isolate the probe from the plasma, except at the probe tip, ceramic tubes were used as insulating material. Teflon tape was applied at the junctions to ensure mechanical stability and electrical insulating during probe insertion.

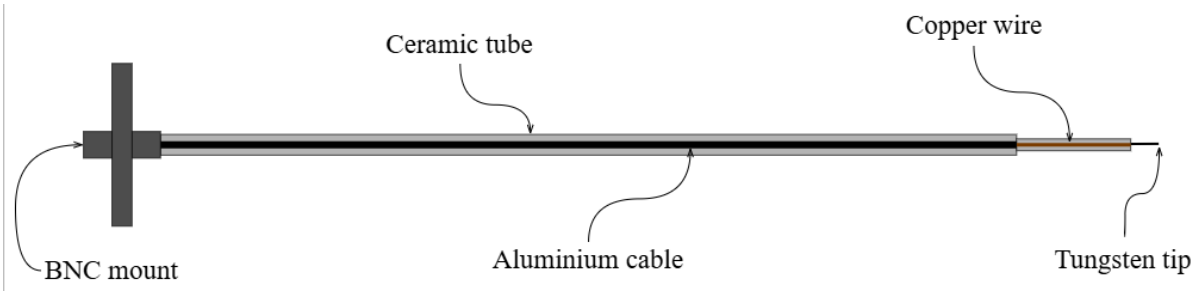


Figure 45: Cylindrical probe design

The overall probe length was measured to be 350 mm. The detailed design specifications are listed in the table below.

Probe Component	Dimension	Material
Coaxial cable	Length = 300 mm	Aluminium
Threading wire	Length = 50 mm	Copper
Probe tip	Length = 8 mm Diameter = 0.1 mm	Tungsten
Main insulating tube	Length = 300 mm	Ceramic
Small insulating tube	Length = 40 mm	Ceramic

Table 6: Design specifications of cylindrical probe

5.2.2. Planar Langmuir Probe

For the planar probe, a mild steel tip of 3 mm diameter was used as the sensing electrode. It was connected to the aluminium coaxial cable using a brass connector which was then soldered to the BNC mounting flange. Similar to the cylindrical probe, ceramic tubing was used for insulation purposes and Teflon tape was used to secure the assembly.
The larger effective area of the planar probe allows it to collect more current and hence improve the accuracy in determining plasma density. However, due to its geometry, it is more sensitive to plasma flow direction and sheath expansion effects compared to the cylindrical probe.

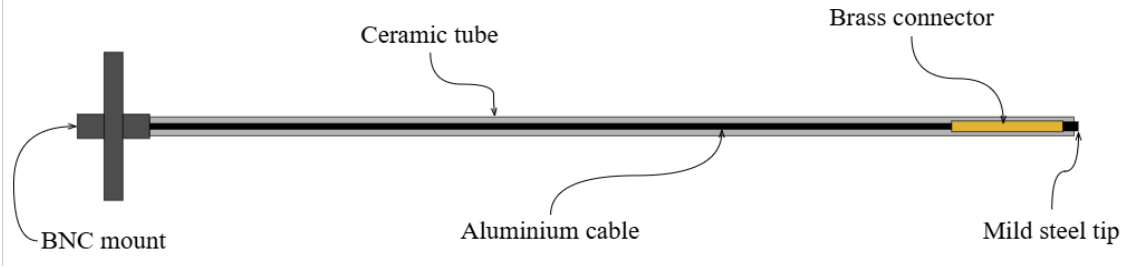


Figure 46: Planar probe design

The overall probe length was measured to be 350 mm. The detailed design specifications are listed in the table below.

Probe Component	Dimension	Material
Coaxial cable	Length = 300 mm	Aluminium
Connector	Length = 40 mm	Brass
Probe tip	Diameter = 3 mm	Mild steel
Insulating tube	Length = 350 mm	Ceramic

Table 7: Design specifications of planar probe

After complete assembly, both probes were tested using a multimeter to verify electrical continuity and ensure proper connections before use in plasma experiments.

5.3. Data acquisition and analysis of I-V characteristic

To obtain plasma parameters the I-V characteristic of the probe was measured using the following experimental setup.

Figure 47 shows the circuit used to obtain the I-V characteristic of a cylindrical Langmuir probe. A ramp voltage generated by a function generator and amplified using an amplifier is applied across a series combination of the probe and a known resistor. This voltage sweep biases the probe from negative to positive potentials with respect to the plasma enabling the collection of both ion and electron currents. When the probe is negatively biased, it repels electrons and attracts ions while when it is positively biased, it attracts electrons. The plasma acts as a conductive medium, completing the circuit and allowing current to flow through the probe.

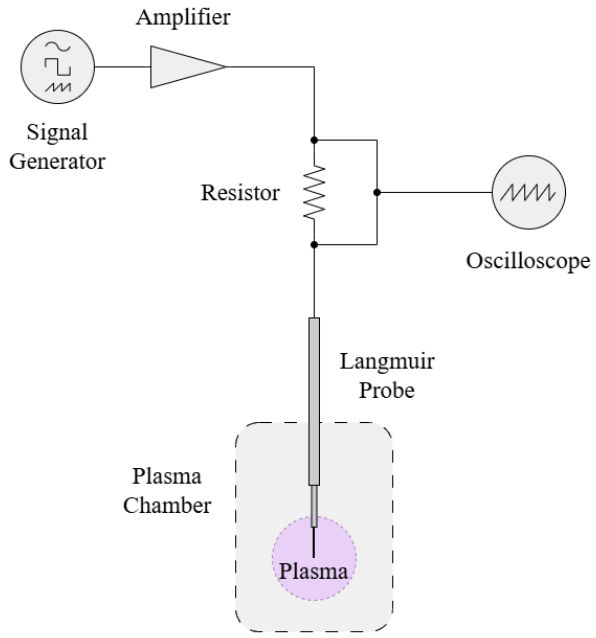


Figure 47: Langmuir probe circuit diagram

The oscilloscope is used to measure the voltage drop across the resistor. Since the resistor value is known, the probe current can be calculated using Ohm's law:

$$I_{probe} = \frac{V_R}{R}$$

Where, V_R is the voltage drop across the resistor of resistance R.

Once the applied bias voltage and corresponding probe current data is obtained from the oscilloscope, they are processed to obtain the probe I-V curve. In this setup, the input ramp voltage applied to the probe ranges from -100 V to 100 V, and the resistor has a value of 100Ω.

Due to the presence of significant noise in the measured data, which resulted in unphysical plasma parameter estimates, a noise correction technique was applied. The baseline noise

signal obtained at 0 V bias was subtracted from the measured probe voltage to improve signal quality. Figure 48 shows the input ramp voltage, the obtained probe voltage, the raw noise signal, and the cleaned probe voltage.

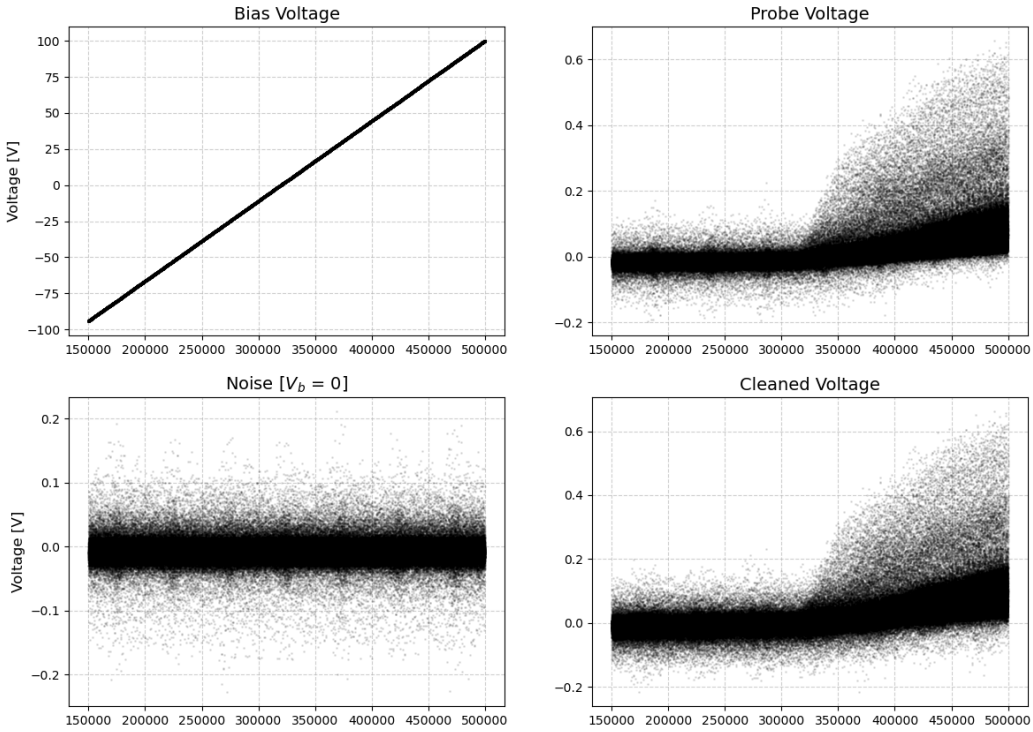


Figure 48: The input, obtained, raw noise, and cleaned data (Cylindrical probe)

x-axis represents data point index.

As the corrected signal was still too noisy, a multi-stage smoothing approach was applied to obtain a reliable I-V characteristic. First, the data were smoothed using a Savitzky-Golay filter to preserve the overall trend while reducing high frequency noise. This was followed by bin averaging to further suppress random variations. Finally, a second Savgol filter was applied to obtain the final smooth I-V curve, as shown in Figure 49. This multi-stage approach ensured preservation of the physical trend while minimizing noise.

A **Savgol filter** works by fitting a low-degree polynomial to a local window of data points, allowing preservation of the shape and height of signal peaks unlike simple moving averages which tend to blur these features. The filter moves a window of fixed size across all data points, performs a least-squares polynomial fit within each window, and replaces the central point with the fitted value. To implement this filter, two primary hyperparameters are chosen:

1. *Window length:* the number of adjacent data points used for the fit (must be an odd integer).
2. *Polynomial order:* the degree of the polynomial (must be less than the window length).

Bin averaging or **data binning** simply groups a range of continuous values into discrete bins and replaces them with a single representative value, such as median or mean. This technique typically results in effective noise reduction, although it may reduce data resolution.

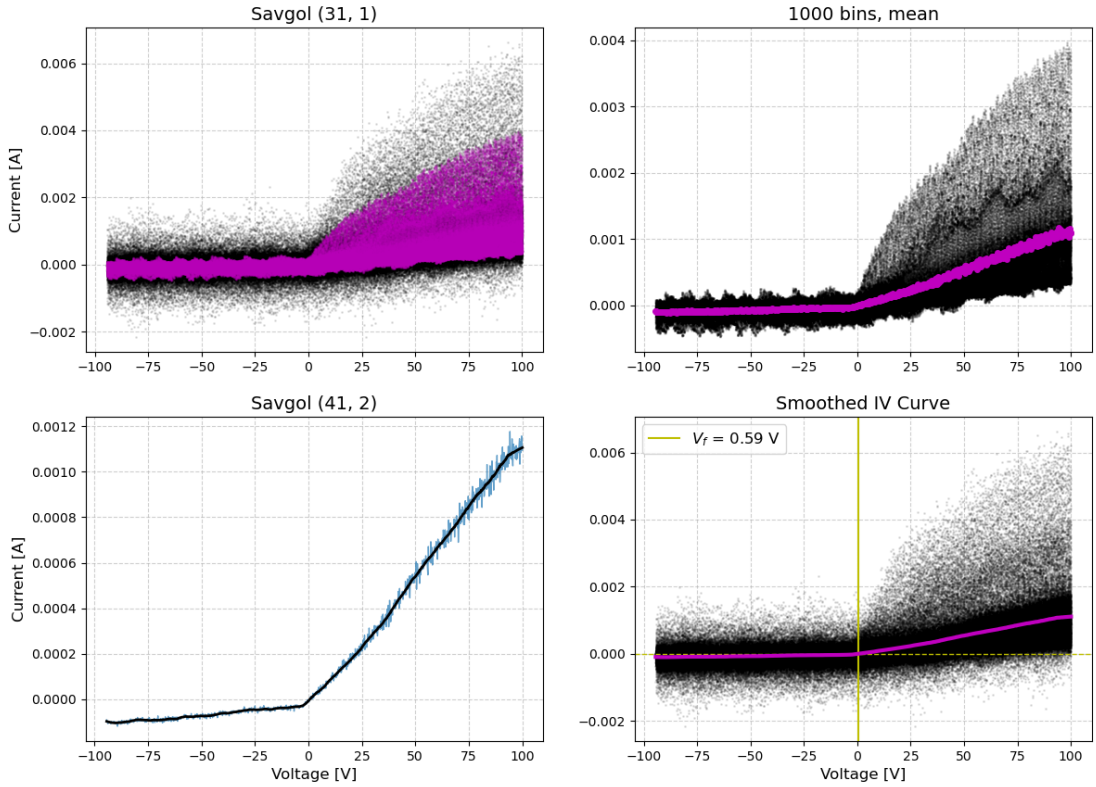


Figure 49: Smoothing data to obtain the I-V characteristic

The window length and polynomial order were chosen as (31, 1) and (41, 2) for the first and second Savgol filters, respectively. For bin averaging, 1000 bins of 0.2 V each were taken to ensure sufficient number of datapoints for further smoothing and in final I-V curve.

From the processed I-V curve, the floating potential (V_f) was determined to be **0.59 V.**

The plasma potential (V_p) was determined by identifying the knee of the I-V curve using derivative-based analysis. The first derivative (slope) and the curvature of the curve were computed, and the plasma potential was identified at the voltage corresponding to the maximum rate of change. It should be noted that curvature used here is not the second derivative of I-V curve, but is defined as:

$$curvature\ (k) = \frac{|I"(V)|}{[1 + (I'(V))^2]^{3/2}}$$

Where, I'(V) and I"(V) are the first and second derivatives of the I-V curve, respectively.

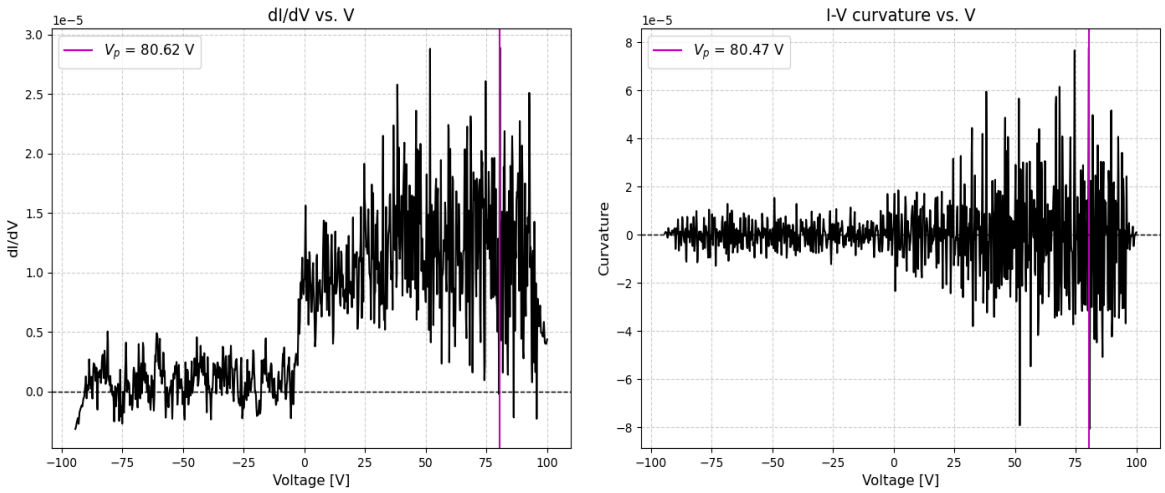


Figure 50: Slope and curvature of I-V curve vs. applied voltage

The figure above shows the variation of slope and curvature with respect to the applied voltage. From this analysis, the plasma potential (V_p) was found to be approximately **80.62 V**. These results demonstrate the successful implementation of the Langmuir probe for plasma diagnostics, providing reliable estimates of key plasma parameters under controlled laboratory conditions.

Chapter - 6 : Quantifying Plasma Parameters

In this chapter, we discuss the estimation of key plasma parameters using remote sensing and in-situ diagnostics. First, it examines the non-intrusive technique of Optical Emission Spectroscopy (OES) to estimate electron temperature from the emission spectra of argon plasma. The obtained spectra are analysed using the Boltzmann plot method, which provides insight into the excitation dynamics of the plasma. The chapter then focuses on Langmuir probe diagnostics, where plasma parameters are determined from experimentally measured I-V curves. A series of steps are followed to distinguish between ion and electron currents, identify characteristic regions, and determine electron temperature and density. Additional plasma parameters such as plasma frequency and Debye length are then calculated. Finally, the results from cylindrical and planar probe measurements are compared, and the agreement between OES and probe-based measurements is presented, demonstrating the effectiveness of the diagnostic methods.

6.1. Using Remote Sensing

Optical Emission Spectroscopy (OES) is a powerful non-intrusive technique used for characterizing plasma parameters. It is based on the principle that when atoms in a sample are excited, their electrons are elevated to higher energy levels. Upon relaxation back to lower energy states, these atoms emit photons at element-specific wavelengths. These emissions form discrete spectral lines, which are specific to the transitions occurring in different elements present in the plasma. The technique provides valuable insights into plasma composition, electron temperature, and electron density without disturbing the plasma environment.

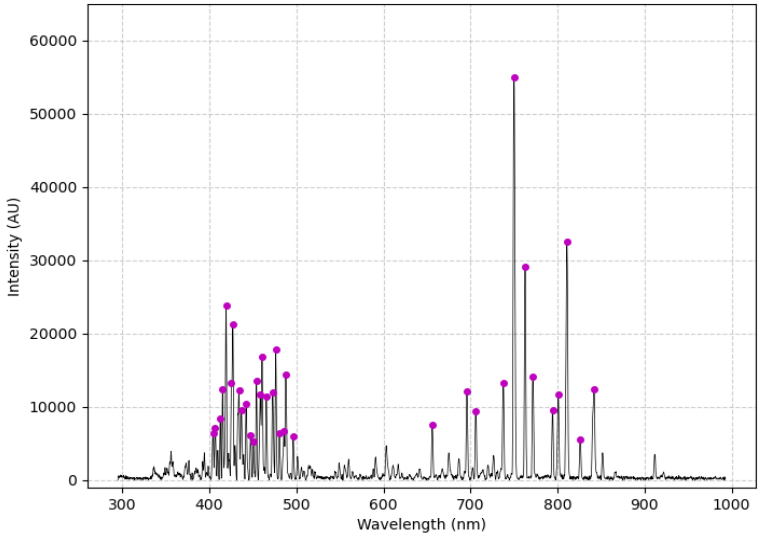


Figure 51: Baseline corrected emission spectrum (Argon plasma at 0.300 mbar, 650 V)

The figure above shows the emission spectra of Argon plasma at a pressure of 0.300 mbar and a discharge voltage of 650 V. The spectrum exhibits prominent emission lines in two wavelength ranges: 400 - 500 nm and 696 - 850 nm with the highest intensity at 750 nm. In addition to Argon lines, weak emission lines corresponding to the Hydrogen Balmer series (H_α, H_β, and H_γ) are also observed, indicating the presence of impurities in the discharge.

The electron temperature is estimated using the **Boltzmann plot method,** which relates the intensities of emission lines to the population of excited states:

$$\ln\left(\frac{I_{ik}\lambda_{ik}}{A_{ki}g_k}\right) = -\frac{E_k}{k_B T_e} + \ln\left(\frac{N}{Z(T)}\right)$$

Where:
- I_{ik} = line intensity
- λ_{ik} = wavelength of the transition
- A_{ki} = Einstein transition probability
- g_k = statistical weight of the upper level
- E_k = excitation energy
- Z(T) = partition function

Plotting the left-hand side against E_k yields a straight line with slope:

$$m = -\frac{1}{T_e}$$

Where T_e is expressed in electron volts (eV).

The experimentally obtained wavelengths were compared with known Argon spectral lines from **NIST Atomic Spectra database**. It was found that the emission lines in the 400 - 500 nm range correspond to Ar II (singly ionized argon), and those in the 696 - 850 nm range correspond to Ar I (neutral argon). This indicates the coexistence of excitation and ionization processes within the plasma.
However, Ar I lines exhibit signs of self-absorption, making them unsuitable for Boltzmann plot analysis. Therefore, only Ar II lines were selected for estimating the electron temperature. The spectroscopic parameters used for the analysis are listed in Table 8, and the corresponding Boltzmann plot is shown in Figure 52.

Peak Wavelength [nm]	**Peak Intensity**	**A_{ki} [s^{-1}]**	**g_k**	**E_k [eV]**	**Ion**
406.63	7171.00	58000000	6	21.49804987	Ar II
425.27	13241.79	16400000	6	19.549011775	Ar II
458.66	11774.04	66400000	6	21.127033505	Ar II
460.68	16916.83	78900000	8	21.143075448	Ar II
476.32	17845.30	64000000	4	19.867157038	Ar II
480.37	64499.46	78000000	6	19.222902151	Ar II

Table 8: Spectroscopic data used for Boltzmann plot analysis

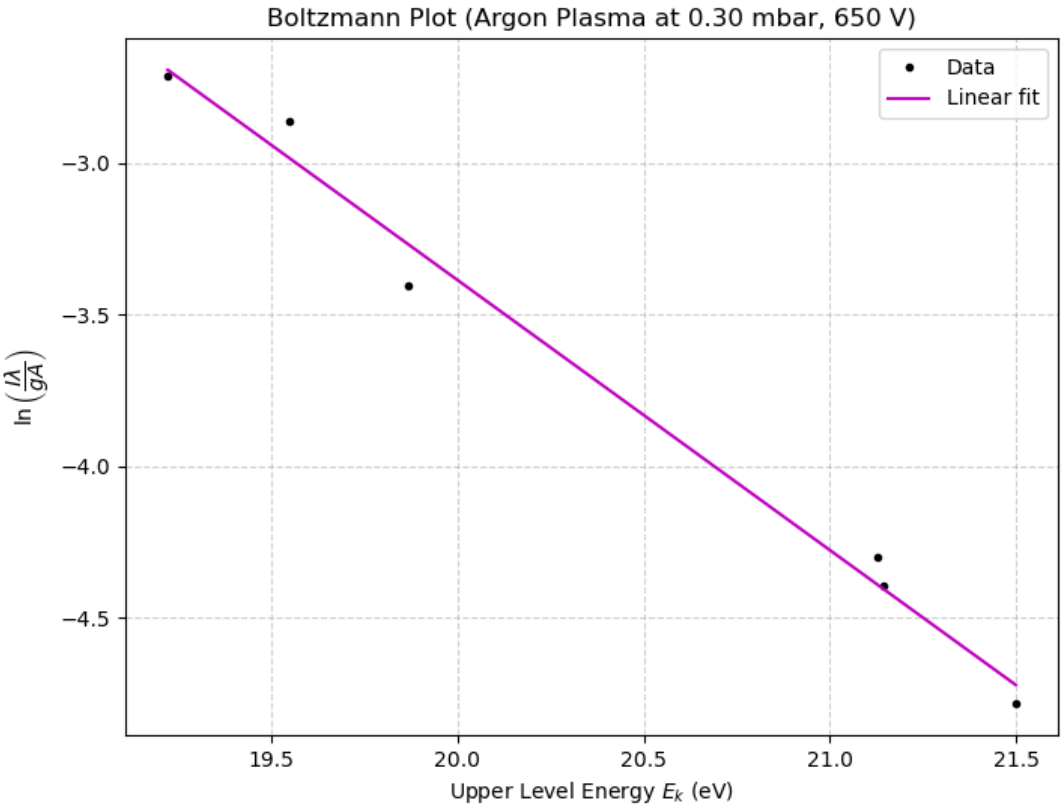


Figure 52: Boltzmann plot analysis (Argon plasma at 0.300 mbar, 650 V)

Fit Equation: y = -0.8913x + 14.4411

From the slope of the fitted line, the electron temperature is estimated to be: **T_e = 1.12 eV**.

6.2. From Langmuir Probe

The current obtained using Ohm's law represents the total current, i.e., sum of ion and electron current components. In order to determine the plasma parameters such as electron temperature and density, the electron current must be separated from ion current.

The first step in this process is to identify the ion saturation region and perform a linear fit. As seen in Figure 9, the ion saturation region appears flat in an ideal I-V. In practical measurements, it exhibits a finite slope due to sheath expansion and collisional effects. Therefore, ion current cannot be approximated as a constant average value in this region, as this would lead to inaccurate values of electron current and false estimate of plasma parameters.

In this experiment, ion saturation region was identified in the voltage range of -25 V to -100 V. A linear fit was performed in this region and extrapolated over the full voltage range. The ion saturation current was estimated as the mean value of the fitted ion current in this region, resulting in I_{is} = **-0.0786 mA**.

The electron current was then calculated as:

$$I_e = I_{probe} - I_i(V)$$

Where I_e, I_{probe}, and $I_i(V)$ are electron current, total probe current, and fitted ion current at a given voltage, respectively.

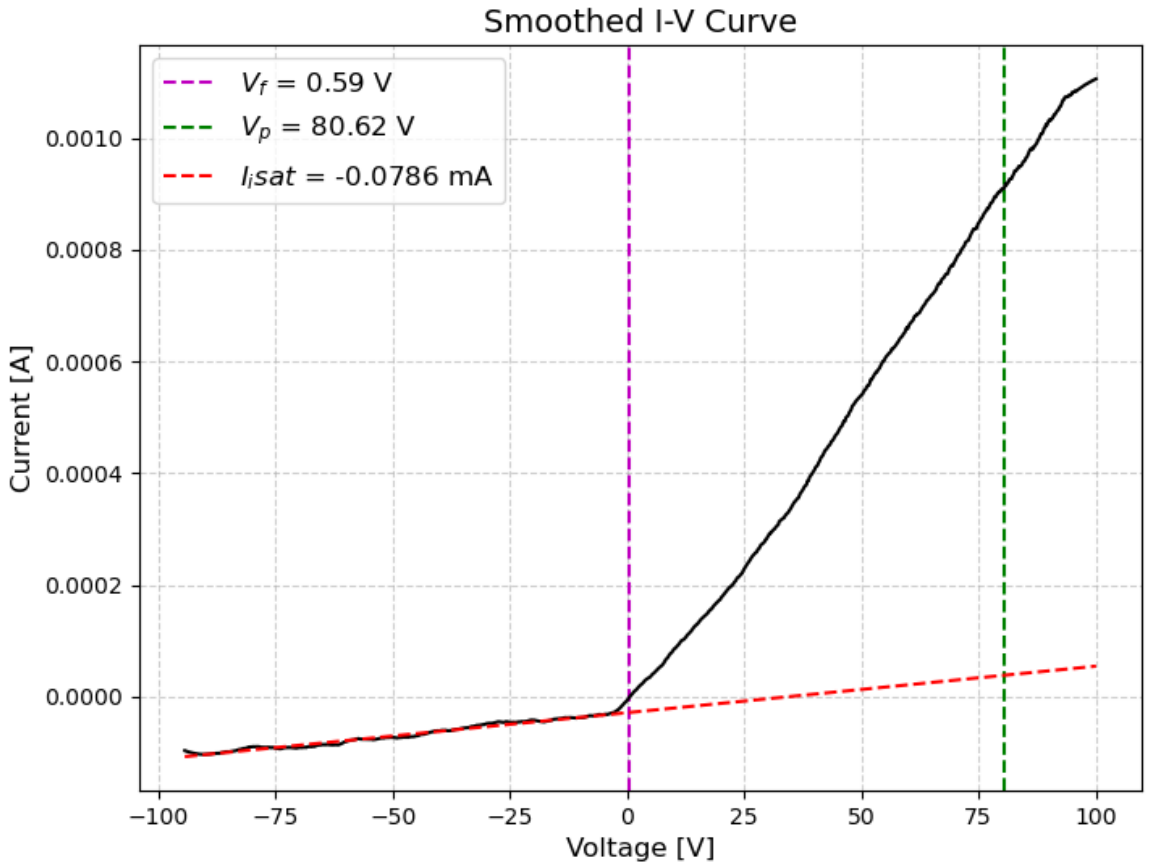


Figure 53: Smoothed I-V Curve (cylindrical probe)

Shows floating potential (V_f, magenta), plasma potential (V_p, green) and linear fit in the ion saturation region (red).

To determine the electron temperature, the natural logarithm of the electron current was plotted against the probe bias voltage as shown in Figure 54. In electron retardation region, this plot exhibits a linear behaviour. A linear fit was performed over this region [Figure 55], and the electron temperature was obtained from the inverse of the slope:

$$T_e = \frac{1}{m} \ [eV]$$

Where m is the slope of the linear fit.

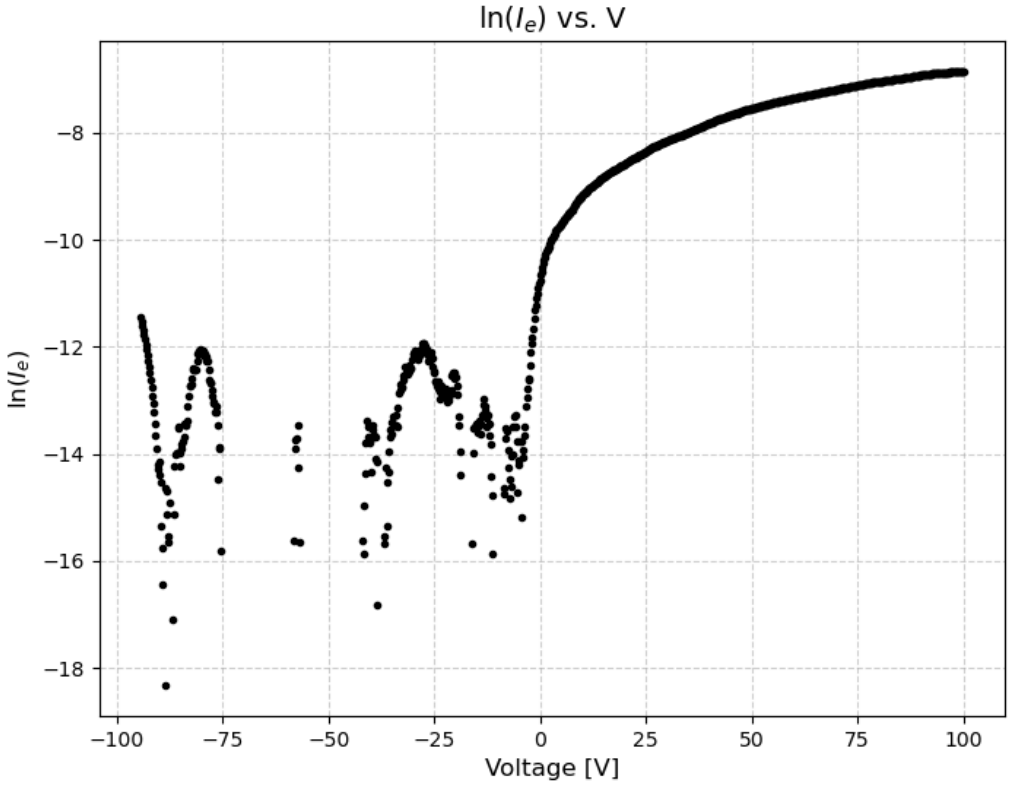


Figure 54: Semilog plot of ln I_e vs. V

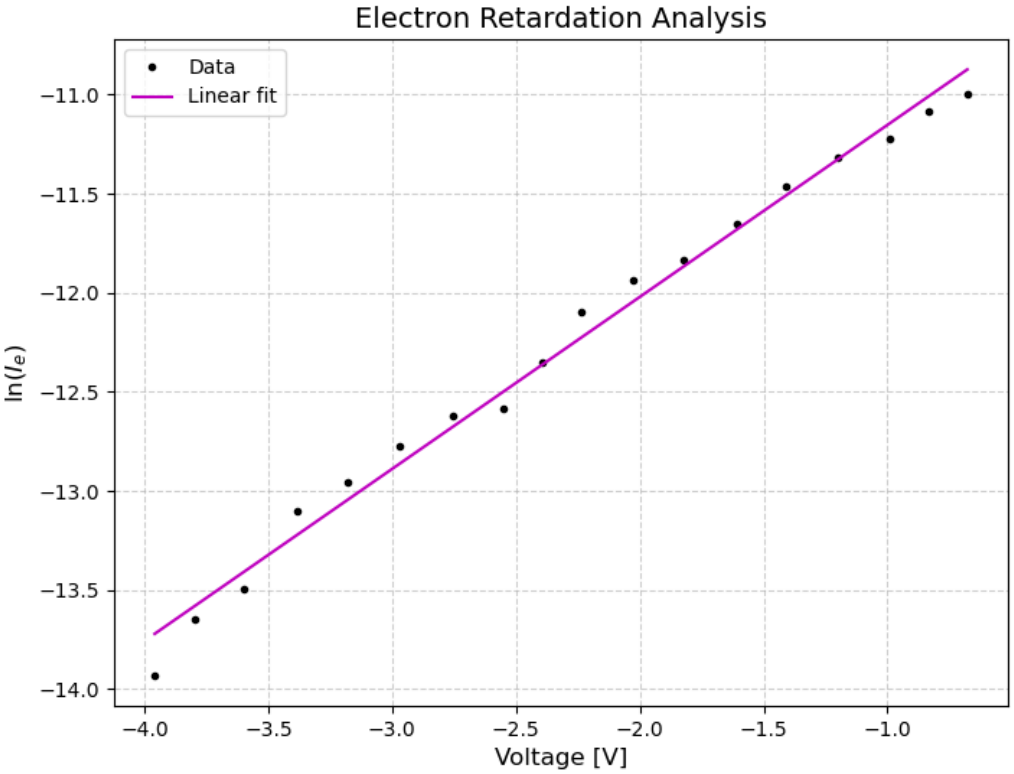


Figure 55: Linear fit in electron retardation region

Fit equation: y = 0.8678x - 10.2855

From this, the electron temperature was determined to be **1.15 eV**.

The electron number density (n_e) was calculated using the formula:

$$n_e = \frac{I_{es}}{eA}\sqrt{\frac{2\pi m_e}{eT_e}}$$

Where:

I_{es} is electron saturation current i.e. the value of electron current at plasma potential in amperes (A),

T_e is the calculated electron temperature in electron volts (eV),

m_e is the electron mass (9.11×10^{-31} kg),

e is the elementary charge (1.6×10^{-19} C), and

A is the probe tip surface area given by:

$$A = 2\pi rl$$

Where r and l are probe tip radius and length, respectively.

Using these values, the electron number density was found to be **1.214×10^{16} m^{-3}**.

The plasma frequency was then calculated as:

$$f_p = \frac{1}{2\pi}\sqrt{\frac{n_e e^2}{\epsilon_0 m_e}}$$

Where, ϵ_0 is free space permittivity (8.854×10^{-12} F/m).

In this case, f_p = **0.9878 GHz**

The Debye length was calculated using:

$$\lambda_D = \sqrt{\frac{\epsilon_0 T_e}{en_e}}$$

Where T_e is in eV.

In this case, λ_D = **72.34 µm**

Similar diagnostics were performed using the planar probe. Figure 56 and Figure 57 show the corresponding I-V curve and electron retardation region analysis, respectively.

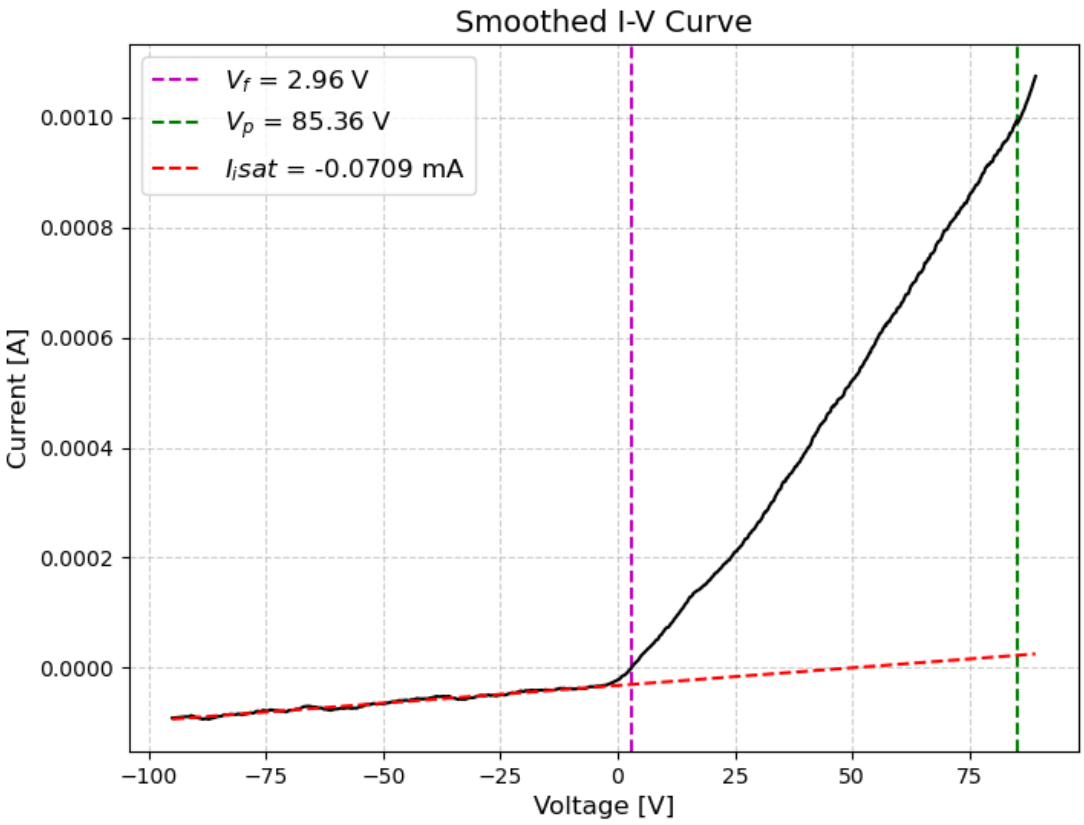


Figure 56: I-V curve (Planar probe)

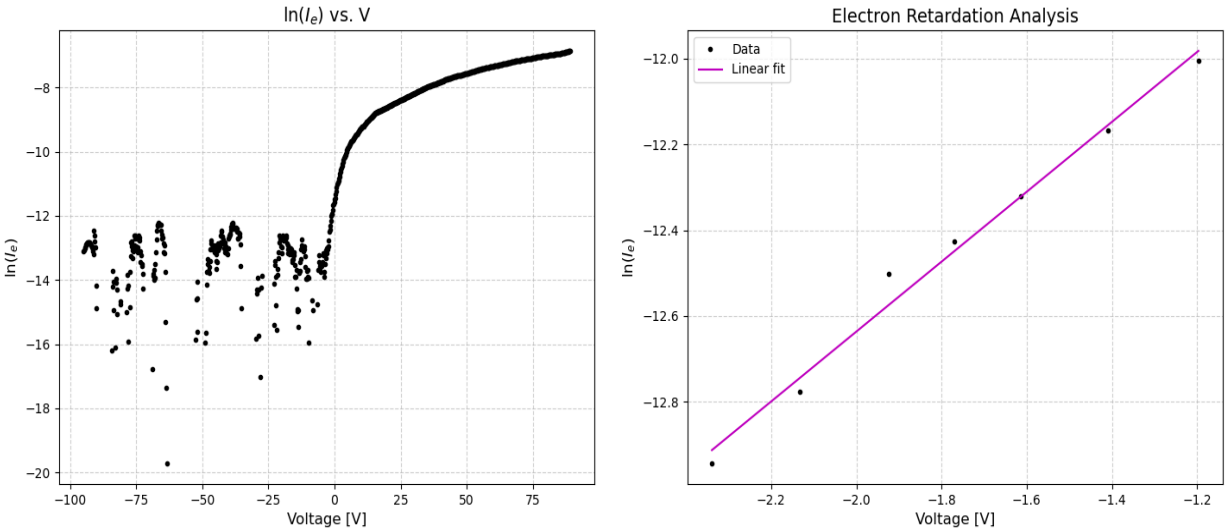


Figure 57: Electron retardation region analysis

Fit equation: y = 0.8136x - 11.0088

The comparison of results from both probes is mentioned in the table below.

Parameter	Langmuir probe geometry Langmuir probe	
	Cylindrical	**Planar**
Floating potential (V_f)	0.59 V	2.96 V
Plasma potential (V_p)	80.62 V	85.36 V
Ion saturation current (I_{is})	-78.6 μA	-70.9 μA
Electron saturation current (I_{es})	875.9 μA	968.7 μA
Electron temperature (T_e)	1.15 eV	1.23 eV
Electron number density (n_e)	1.21×10^{16} m^{-3}	4.62×10^{15} m^{-3}
Plasma frequency (f_p)	0.99 GHz	0.61 GHz
Debye length (λ_D)	72.34 μm	121.31 μm

Table 9: Plasma parameters obtained from cylindrical and planar probe

The floating potential is observed to be positive in both cases, which deviates from the ideal expectation of a slightly negative value. This can be attributed primarily to experimental factors such as baseline offsets in oscilloscope and grounding conditions, which can introduce a shift of the order of 1-3 V. Additionally, secondary electron emission from the probe surface can contribute to an increase in collected electron current, effectively shifting the floating potential toward positive values. This effect is expected to be stronger for planar probe due to its larger area and higher ion bombardment. Furthermore, plasma oscillations and sheath fluctuations may distort the I-V characteristic, leading to a shift in the zero-current crossing point.

The plasma potential values obtained are relatively high compared to typical glow discharge conditions. This can be attributed to the method of estimation, as derivative-based techniques are highly sensitive to noise. Although multi-stage smoothing was applied to stabilize the signal, it can introduce slight shifts or broadening in the peak of derivative or curvature, affecting the accuracy of the identified plasma potential.

The electron temperature values obtained from both the probes are almost equal.

The electron number density, Debye length, and plasma frequency exhibit noticeable differences between the two probe geometries. These discrepancies arise due to geometric effects such as sheath structure, effective collection area, and edge effects along with the limitations of ideal Langmuir probe theory when the probe dimensions become comparable to the Debye length. Additionally, the assumption of a collisionless sheath is not strictly valid at the operating pressure of 0.300 mbar, further contributing to deviations in the estimated plasma parameters.

The electron temperature obtained from OES is in good agreement with the values obtained from the Langmuir probe. Since OES is a non-intrusive technique, this agreement supports and validates the probe measurements. This validates the reliability of plasma parameters obtained from the Langmuir probe.

Chapter - 7 : Summary and Outlook

This chapter summarizes the key findings of the present work and discusses potential directions for future work. The study combines a data-driven approach for solar wind classification with experimental plasma diagnostics, providing a comprehensive understanding of plasma behaviour across space and laboratory environments. The SOM-based framework demonstrated the ability to identify distinct solar wind regimes and detect transient structures such as ICMEs, while maintaining consistency with established classification schemes. In parallel, the development of Langmuir probes and the use of OES enabled estimation of plasma parameters in controlled laboratory conditions. Together, these approaches highlight the effectiveness of integrating machine learning techniques with experimental diagnostics. The insights gained from this work not only validate the methodology used but also provide a foundation for extending such techniques to real-time space weather monitoring and advanced plasma characterisation studies.

7.1. Classification of Solar Wind from In-Situ Data

This study has presented an unsupervised machine learning approach using Self-Organising Maps (SOM) to classify the solar wind plasma conditions from in-situ data of the WIND spacecraft. Using SOM combined with K-Means clustering, four clusters of solar wind were identified:

- Slow Solar Wind (SSW),
- Compressed Slow Wind (CSW),
- High Alfvenic Wind (HAW), and
- Ejecta

These clusters were found to be well aligned with physically interpretable solar wind structures and highly consistent with the Xu & Borovsky model.

The approach was able to capture both the steady-state and transient structures in the solar wind, including ICMEs. The ejecta cluster had the strongest association with ICMEs, and other clusters were found to include embedded transient features, indicating the propagation of ICMEs through different background solar wind conditions. Testing of the trained SOM using independent data from WIND (2016) and Aditya-L1 (2024-2025) validated the model's generalization ability and its applicability for data across different phases of the solar cycle.

Overall, the SOM-based method offers a robust, effective, and interpretable technique for solar wind classification and ICME detection.

7.2. Developing Instrumentation for In-Situ Plasma Measurements

Besides data-driven analysis, this work also included the design and fabrication of Langmuir probes for plasma measurements. Cylindrical and planar Langmuir probes were developed while ensuring robust mechanical structure, reliable electrical contacts, and minimal disturbance of the plasma. A DC glow discharge plasma was generated using Argon gas at controlled pressure and voltage of 0.300 mbar and 650 V.

The I-V curves of the probes were processed using noise correction, smoothing and binning to determine the plasma parameters. Langmuir probe theory was used to calculate the electron temperature, number density, plasma frequency, and Debye length. A non-invasive technique, Optical Emission Spectroscopy (OES), was used for independent evaluation of electron temperature.

The measurements from both the probes were in good agreement, while variations were attributed to geometric sheath effects. The close agreement between OES and probe measurements demonstrate the validity of the experimental methodology and confirmed the obtained plasma parameters.

7.3. Outlook

The findings of this work provide opportunities for future research in data analysis and experimental plasma diagnostics. In data analysis, the SOM approach can be further refined by adding other parameters such as magnetic field fluctuations and composition rations to enhance classification accuracy. A key advancement would be to integrate SOM algorithm on space missions to enable real-time analysis of large datasets, allowing autonomous classification of solar wind regimes and early detection of ICMEs without the need for ground-based data processing.

On the experimental side, additional efforts can be devoted to reduce the size of the Langmuir probe system for space applications. This involves miniaturizing the design, lowering power consumption, and enhancing durability to withstand extreme space conditions. These compact probes could allow for in-situ continuous plasma diagnostics on satellites.

Overall, combining onboard machine learning with miniaturized diagnostic tools offers a promising approach to real-time space weather monitoring and plasma diagnostics.

REFERENCES

1. T. Kohonen, Self-Organizing Maps, 3rd ed., Springer, 2001.
2. V. Satopää et al., Finding a 'knee' in a curve, KDD Workshop, 2011.
3. F. Xu and J. E. Borovsky, JGR Space Physics, 120, pp.70–100, 2015.
4. F. Carella et al., Astronomy & Astrophysics, 2025.
5. A. Mistriello, CUDA-SOM GitHub Repository, 2020.
6. J. Gieseler el al., Solar-MACH, Frontiers in Astronomy and Space Sciences, 2023.
7. Verscharen, D., Klein, K.G. & Maruca, B.A, The multi-scale nature of the solar wind, Living Reviews in Solar Physics, 5, 2019.

APPENDIX – A

Examples of Bad Quality Data

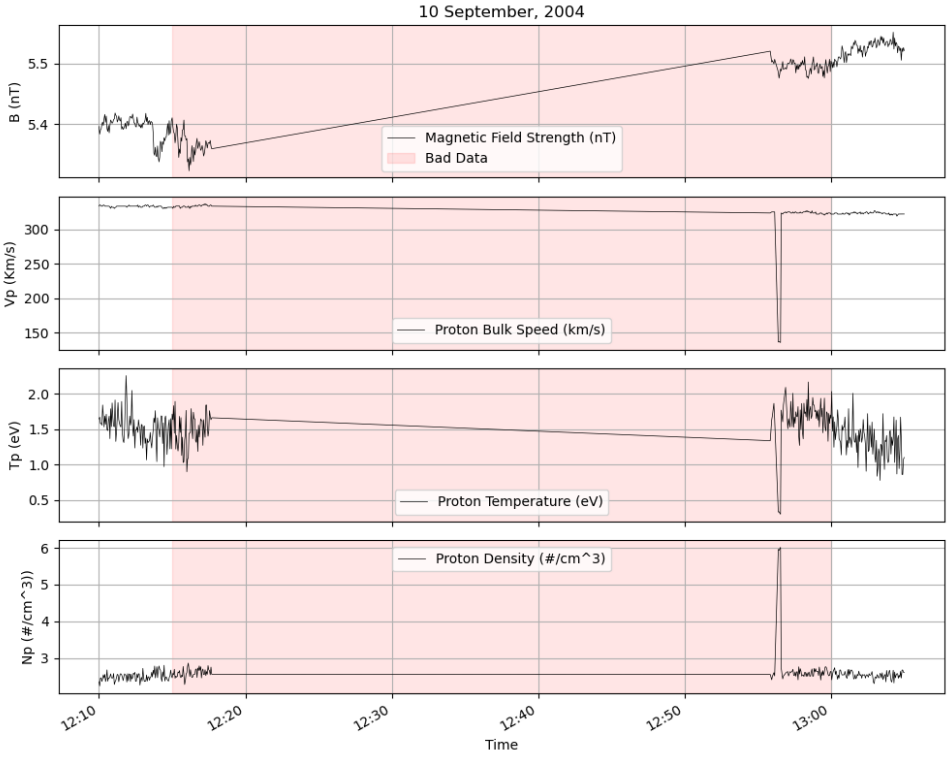


Figure 58: September 10, 2004, 12:15 – 13:00 (missing data and nonphysical spikes)

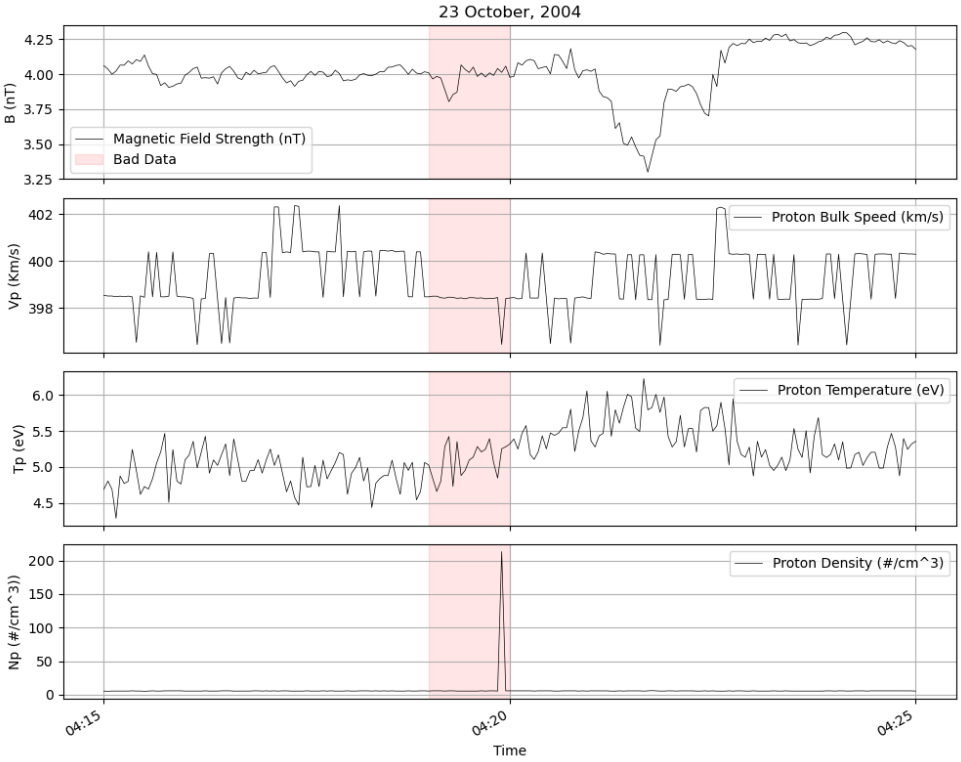


Figure 59: October 23, 2004, 04:19 – 04:20 (nonphysical spike in Proton Density)

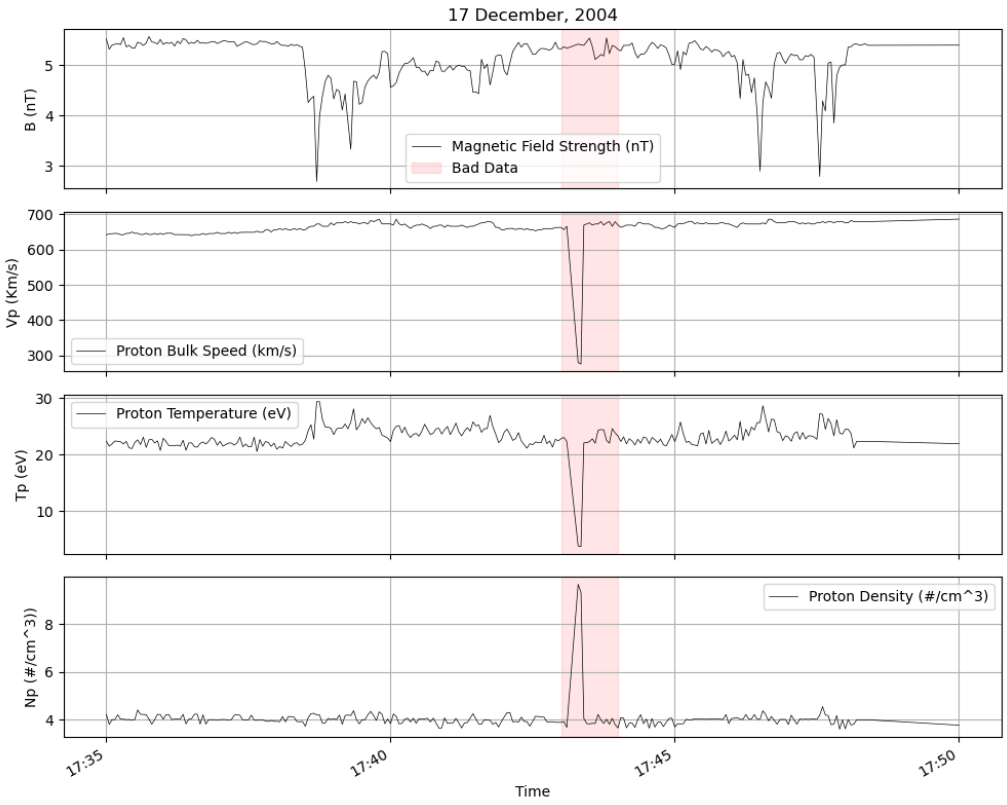


Figure 60: December 17, 2004, 17:43 – 17:44 (nonphysical spike in Proton Density)

APPENDIX – B

Input Data Feature Selection

S. No.	Input	Dataset size and SOM parameters	QE (%)	k	SOM Cluster	Xu and Borovsky Comparison							
						Row Normalized				Column Normalized			
						SB	SR	Ejecta	CH	SB	SR	Ejecta	CH
1	V_p, N_p, $\log_{10}T_p$, B **(Initial)**	5 mil datapoints, 11346 nodes	11.09	4	SSW	**0.72**	0.16	0.07	0.05	**0.66**	0.63	0.36	0.09
					CSW	0.51	**0.40**	0.08	0.01	0.11	**0.37**	0.09	0
					Ejecta	0	0	**0.93**	0.07	0	0	**0.23**	0.01
					HAW	0.31	0	0.08	**0.61**	0.23	0	0.32	**0.90**
2	V_p, N_p, $\log_{10}T_p$, N_α/N_p	5 mil datapoints, 11288 nodes	13.16	3	SSW	**0.71**	0.15	0.08	0.06	**0.74**	0.69	0.46	0.13
					CSW	0.39	**0.52**	0.09	0.01	0.05	**0.31**	0.06	0
					Ejecta	-	-	-	-	-	-	-	-
					HAW	0.28	0	0.12	**0.60**	0.20	0	0.48	**0.87**
3	V_p, N_p, $\log_{10}T_p$, B, N_α/N_p	5 mil datapoints, 11346 nodes	19.05	4	SSW	**0.73**	0.13	0.06	0.08	**0.72**	0.54	0.34	0.16
					CSW	0.44	**0.49**	0.07	0	0.09	**0.45**	0.08	0
					Ejecta	0.02	0	**0.91**	0.06	0	0	**0.22**	0.01
					HAW	0.28	0	0.10	**0.62**	0.19	0	0.36	**0.83**
4	V_p, N_p, $\log_{10}T_p$, $\log_{10}\beta_p$	5 mil datapoints, 11316 nodes	12.40	4	SSW	**0.78**	0.17	0.02	0.03	**0.65**	0.63	0.09	0.05
					CSW	0.39	**0.53**	0.07	0.01	0.06	**0.36**	0.06	0
					Ejecta	0.15	0.01	**0.58**	0.26	0.04	0.01	**0.74**	0.12
					HAW	0.37	0	0.03	**0.60**	0.25	0	0.11	**0.82**
5	V_p, N_p, $\log_{10}T_p$, $\log_{10}\beta_p$, $\log_{10}S_p$	5 mil datapoints, 11310 nodes	13.53	4	SSW	**0.73**	0.16	0.09	0.02	**0.56**	0.53	0.36	0.03
					CSW	0.33	**0.61**	0.06	0.01	0.06	**0.47**	0.05	0
					Ejecta	0.02	0	**0.60**	0.39	0	0	**0.48**	0.12
					HAW	0.46	0	0.02	**0.52**	0.38	0.01	0.10	**0.85**
6	V_p, N_p, $\log_{10}\beta_p$, $\log_{10}S_p$	5 mil datapoints, 11398 nodes	11.97	4	SSW	**0.78**	0.18	0.04	0	**0.61**	0.62	0.16	0
					CSW	0.27	**0.65**	0.07	0.01	0.03	**0.37**	0.05	0
					Ejecta	0.13	0	**0.56**	0.31	0.03	0	**0.73**	0.15
					HAW	0.42	0	0.02	**0.56**	0.32	0.01	0.06	**0.84**
7	V_p, N_p, $\log_{10}\beta_p$, $\log_{10}S_p$ **(Current)**	19 mil datapoints, 21924 nodes	13.85	4	SSW	**0.69**	0.30	0.01	0	**0.63**	0.70	0.04	0.01
					CSW	0.26	**0.61**	0.12	0.01	0.04	**0.25**	0.08	0
					Ejecta	0.37	0.07	**0.48**	0.09	0.13	0.06	**0.76**	0.07
					HAW	0.31	0	0.04	**0.65**	0.20	0	0.12	**0.93**

Table 10: SOM clustering results for different input parameters

QE is quantization error and k is the number of clusters obtained from Kneedle algorithm.

In order to select the best set of parameters for training, multiple models with different set of input parameters were trained and the clustering was compared with Xu & Borovsky classification. The problem with the initial model was that while SOM-based Ejecta had 93% Xu & Borovsky Ejecta, it was capturing only 23% of the actual Ejecta category. This was because Ejecta has two populations based on plasma properties:

1. Pure Ejecta from the Sun – this is the real ejecta.
2. Population with similar properties as CH and SB plasma.

SOM is a statistical model that depends completely on the input data. If one population had similar properties to another population, SOM merges them into single cluster. This is why it was unable to detect ICMEs without any magnetic obstacles. This set of parameters did not show any clear cluster boundaries in the scatterplot, as shown in Figure 61, for SOM to able to

identify these clusters. This resulted in a change of input parameter set to something that could capture these transients better.

First choice was to use the alpha-proton number density ratio but SOM could not see Ejecta in this case as alpha-proton ratio is very noisy. When both alpha-proton ratio and magnetic field strength were included in input dataset, the CSW cluster improved. This was then related to the fact that the SIR and CIRs have lower values of alpha-proton ratio compared to the other categories.

Since the Xu & Borovsky categorizing scheme is based on observed data, their analysis on the values of parameters for each category was used to identify new parameters. The only parameter where ejecta stood out from other categories was the plasma beta (β_p) which has distinguishably low values in ICMEs. When plasma beta was included in training dataset, Ejecta cluster improved significantly. When the scatter chart for this model was plotted [Figure 62], it was observed that SSW, CSW, and HAW clusters were still mixed up and not classified properly.

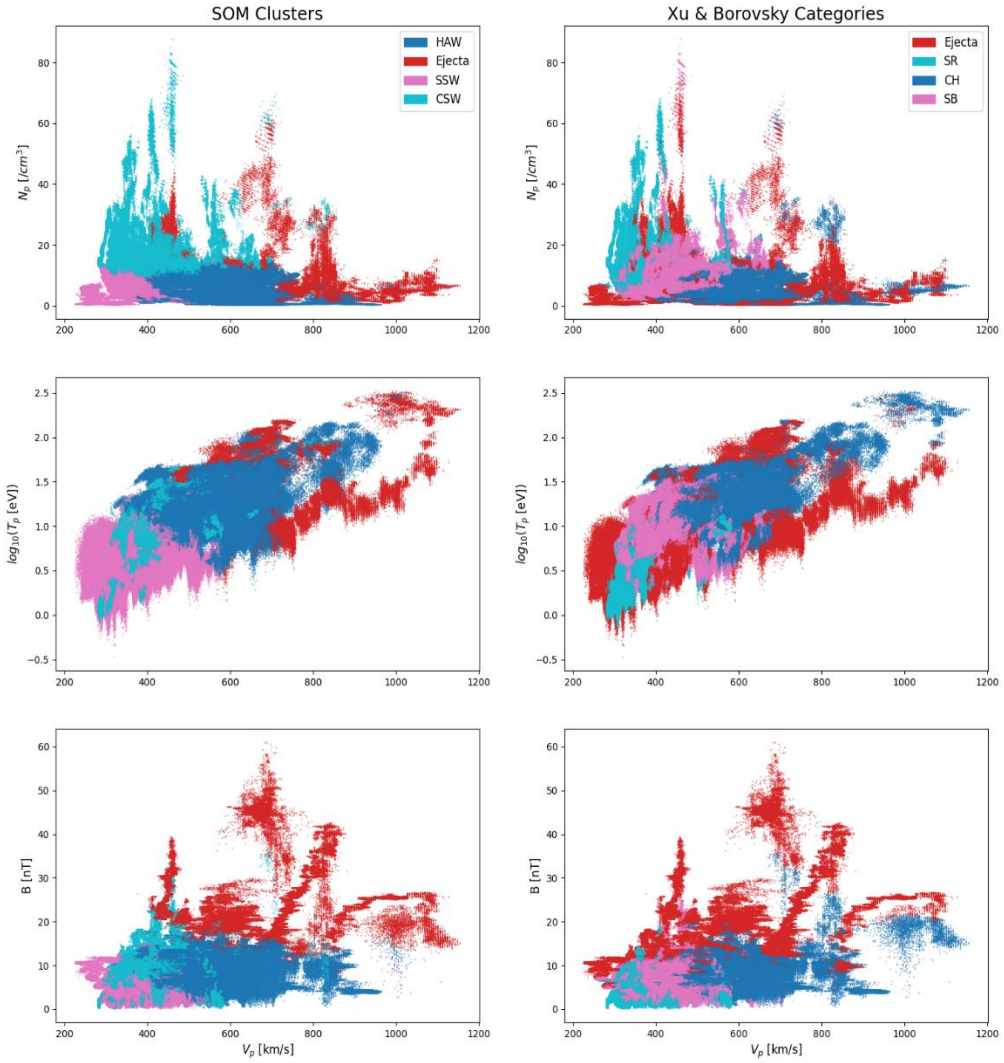


Figure 61: Clustered scatterplots (Input parameters: V_p, N_p, $log_{10}T_p$, and B)

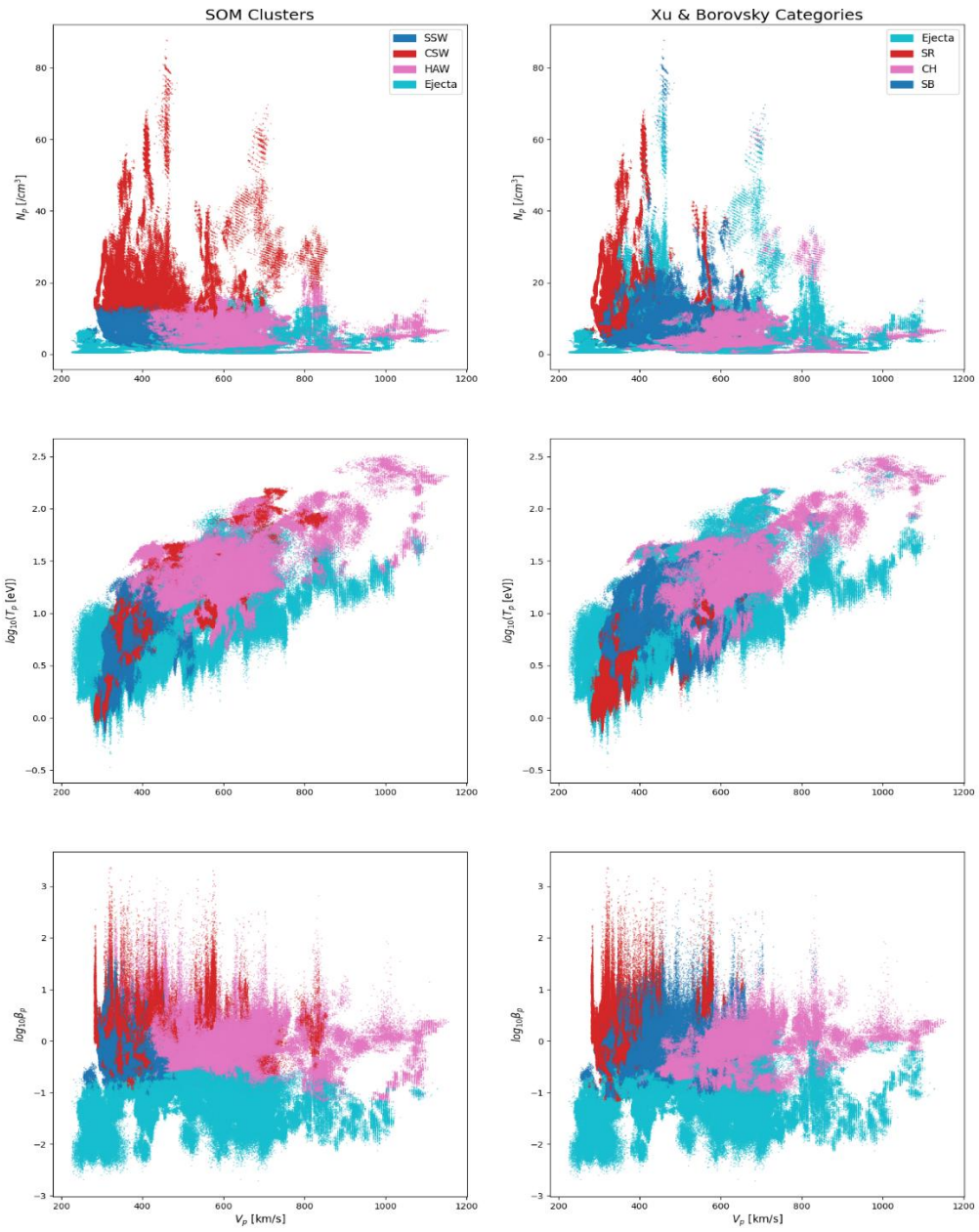


Figure 62: Clustered scatterplots (Input parameters: V_p, N_p, $log_{10}T_p$, and $log_{10}\beta_p$)

To resolve this issue, proton specific entropy (Sp) was used instead of logarithm of proton temperature as it showed slightly better cluster clarity. Figure 63 shows the scatterplot for the finalized model, it shows better cluster resolution in SSW, CSW, and HAW clusters as compared to the previous model. All these scatter charts were plotted only for July 2004 data to limit the number of datapoints and better visibility of clusters.

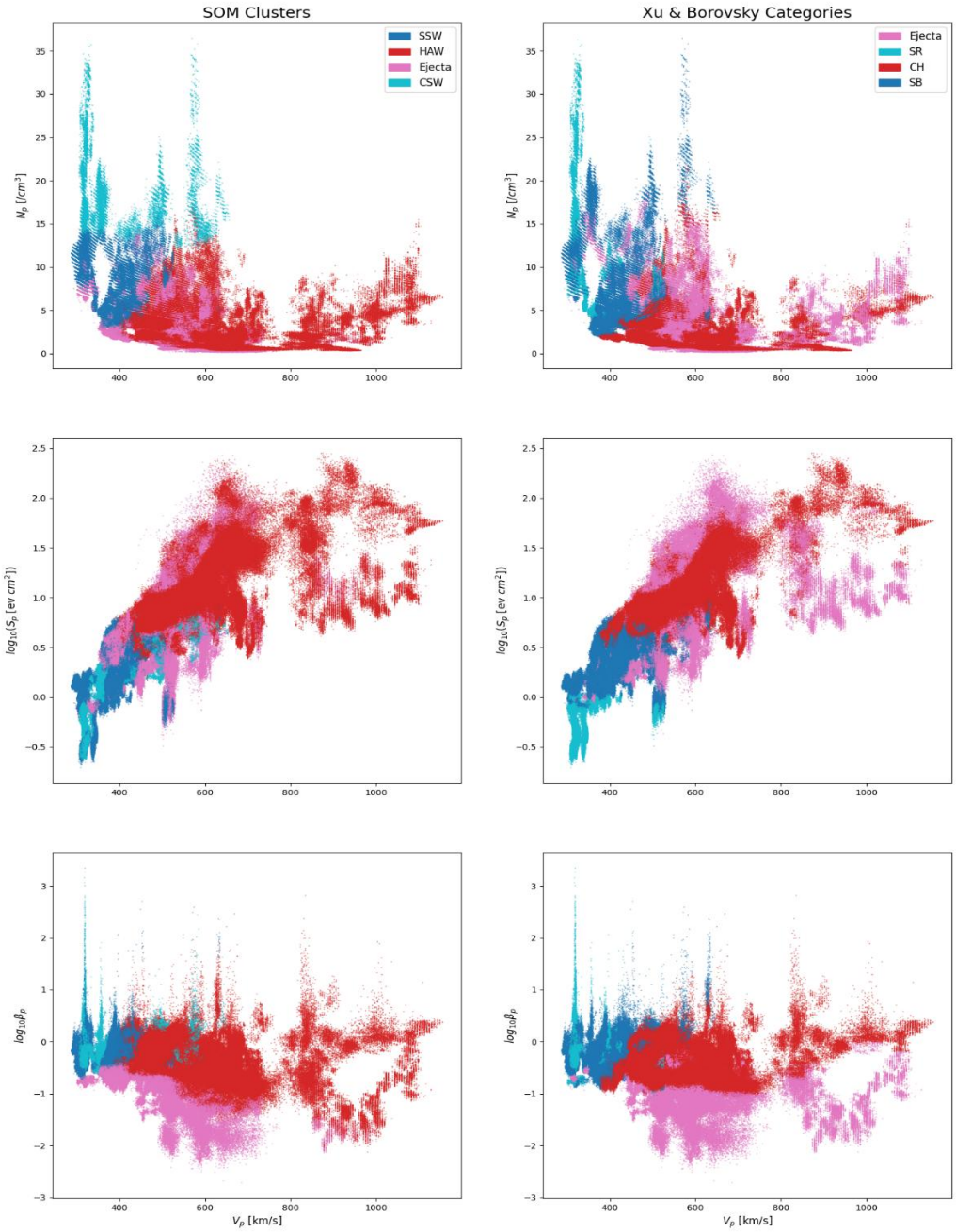


Figure 63: Clustered scatterplots (Input parameters: V_p, N_p, $log_{10}S_p$, and $log_{10}\beta_p$)

APPENDIX – C

SOM Training Parameter Calculations

a. Grid size:

Number of data points in dataset, $N = 19{,}254{,}708$

Number of neurons, $K = 5\sqrt{N} = 21{,}940$

Principal component analysis was performed on the dataset using PCA from sklearn library in python to calculate the required number of rows and columns in SOM grid.

$$x \times y = 189 \times 116 = 21{,}924 \text{ neurons}$$

b. Initial and critical parameters:

Initial neighbourhood radius, $\sigma_0 = 38$

Initial learning rate, $\eta_0 = 0.01\sigma_0 = 0.38$

Critical neighbourhood radius, $\sigma_{crit} = 1$, so that the neighbourhood kernel effectively includes only the BMU and its immediate neighbors at the end of the training.

Critical learning rate (η_{crit}) is calculated such that:

$$\left[\log\left(\frac{\sigma_{crit}}{\sigma_0}\right)\right]^2 = \left[\log\left(\frac{\eta_{crit}}{\eta_0}\right)\right]^2 \Rightarrow \eta_{crit} = 0.01$$

c. Maximum Number of Epochs:

$$T_{max} \leq \left[\log\left(\frac{\sigma_{crit}}{\sigma_0}\right)\right]^2 = 13.23$$

Figure 64 illustrates the exponential decay of hyperparameters and quantization error (QE) of SOM calculated theoretically over epochs. The red line highlights the critical values when radius goes to 1. T_{max} for training SOM was taken to be 13.

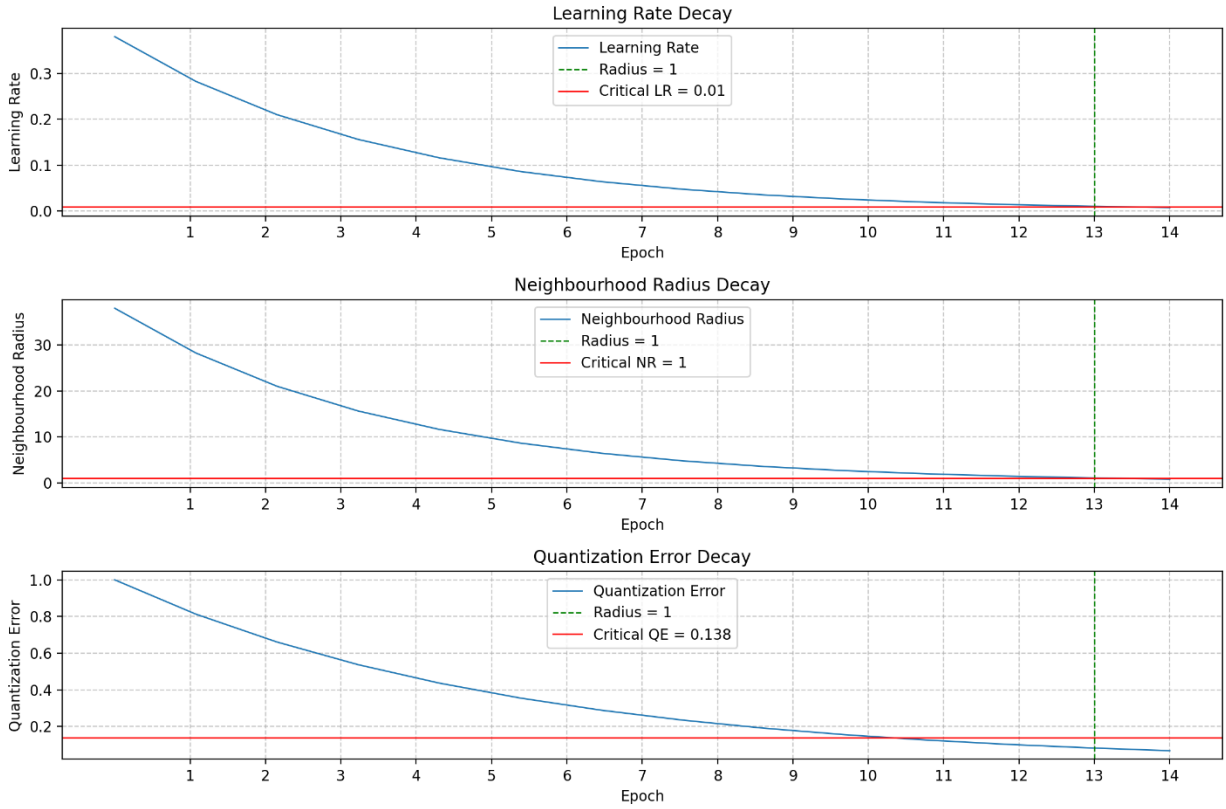


Figure 64: Decay Curves