# Individual Project Final Report

# Characterisation of Compressive Fluctuations in the Solar Wind: Solar Orbiter Data with Machine Learning

Adam Stocker

Space Science and Engineering: Space Technology

08/07/2022

**ᴹUCL**

## 1 ABSTRACT

The solar wind is a plasma stream propagating through space. It originates in the solar corona, and is responsible for forming the heliosphere. Understanding the Solar wind is vitally important to fully understand the Solar Cycle, which fuels space weather and greatly affects the Earth. A major unknown surrounding the Solar Wind is its temperature profile. The observed temperature of the solar wind remains higher than predicted values (Verscharen u. a., 2019). There are many theories about what provides this thermal energy to the solar wind stream; a prominent view is that the dampening of fluctuations within the flow dissipates energy into the plasma. A type of compressive fluctuation within the solar wind is the Slow Mode wave. This wave mode is very efficient at Landau damping and through this mechanism could be a contender in explaining some of this extra thermal energy within the solar wind (Verscharen u. a., 2016). In this project, we have successfully used machine learning techniques to find slow mode waves using data from the solar orbiter spacecraft. The distribution of these waves was then compared to other wind parameters such as wind temperature and velocity. We found that the number of this wave mode decreases when the wind speed increases or the temperature rises. Finally, this wave mode was found throughout the solar wind at different distances from the Sun, and therefore, its dampening does provide heating to the solar wind. More analysis is needed to find how the found slow mode waves are produced and to find the significance of this dampening.

## CONTENTS

## 2 INTRODUCTION

There are three layers to the Sun's atmosphere: the Photosphere, the Chromosphere and the Corona. The temperature is approximately 6000K at the Photosphere. This temperature rises rapidly to over 1,000,000K in the transitional region of the Corona (Sakurai und Sato, 2017). The Corona consists of a hot plasma comprising a discrete collection of ionised particles. The plasma behaviour arises from particle interactions through the long-range coulomb potential. This Coulomb potential is defined as:

$$\psi = \frac{q}{4\pi\epsilon_0 r} \tag{1}$$

Where $q$ is the particle charge, $r$ is the distance from a charged particle, $\epsilon_0$ is the dielectric constant of free space. The electric potential of a system of point charges is the sum of the Coulomb potentials of the individual species within the plasma. However, if there are equal numbers of positively and negatively charged species within the plasma individual particles will be shielded from their neighbouring particles. Therefore in a neutral plasma, individual charged particle potentials are shielded by surrounding particles. This shielding is called Debye shielding (Parks, 2003). This Deybe shielding is described by equation (2).

$$\psi = \frac{q}{4\pi\epsilon_0 r} e^{\frac{-r}{\lambda_D}} \tag{2}$$

Where $\lambda_D$ is the Debye length. If $r > \lambda_D$ then $\psi$ will decrease exponentially. Therefore, one can visualise the Debye length as a sphere of influence, where if $r$ is greater than the $\lambda_D$, the effects of individual particle interactions are cancelled by the collective effect of particles of opposite charge. The Debye length is defined as:

$$\lambda_D^2 = \frac{kT_e\epsilon_0}{n_0 q_e^2} \tag{3}$$

Where $k$ is the Boltzmann constant, $T_e$ is the electron temperature, $n_0$ is the equilibrium density of the plasma, and $q_e$ is the charge on the

electron.

Suppose a group of electrons are displaced, resulting in a local charge density. A force is then applied to restore the local charge density back to a neutral state. This movement causes an oscillatory motion with the electrons within the plasma flow. If the plasma flow behaviour being mapped is greater than the $\lambda_D$, these local oscillations can not maintain charge separation. The plasma, therefore, must be quasi-neutral on scales greater than $\lambda_D$.

As the Corona exists at such a high temperature, some electrons will have a thermal velocity greater than the Sun's escape velocity. This velocity thrusts electrons into a hyperbolic orbit of the Sun, and because of the plasma's quasi-neutrality, protons and positively charged heavy ions are thrust out with the electrons. This mechanism is called ambipolar diffusion and significantly contributes to solar wind formation. As this wind permeates free space, expands and accelerates until approximately 10 solar radii into the heliosphere, fueled by the decreasing temperature (Parks, 2003).

This Solar wind is a turbulent plasma flow at supersonic speeds. The speed of the wind is commonly used to categorise the wind. The categories are: slow wind at a velocity of between 300 $km\,s^{-1}$ to 500 $km\,s^{-1}$; the fast wind which has a velocity of between 500 $km\,s^{-1}$ and 800 $km\,s^{-1}$; and the wind of eruptive events such as coronal mass ejections, where the velocity can reach as high as 2000 $km\,s^{-1}$ (Verscharen u. a., 2019).

## 2.1 Temperature of the Solar Wind

Using the double-adiabatic invariant equations, or Chew-Goldberger-Low (CGL) invariants (Baranov und Kartalev, 1972), it is possible to predict the expected temperature trend of the solar wind. In figure 1, Verscharen u. a. (2019) plot the parallel and perpendicular temperatures predicted by these CGL equations against the actual temperature of the wind. The plot includes the perpendicular proton temperature, $T_{\perp p}$, the parallel proton temperature, $T_{\parallel p}$, the perpendicular electron temperature, $T_{\perp e}$, and the parallel electron temperature, $T_{\parallel e}$. In addition, the plot includes these temperature distributions for the fast wind and the slow wind. The trend shows that for this temperature data, the cooling rate is less than predicted by the CGL equations. Therefore, a heating mechanism within the solar wind must provide this higher-than-expected temperature.

## 2.2 Plasma Dynamics in the Solar Wind

A fundamental equation to describe a collisionless plasma such as the solar wind is the Vlasov equation: (Vlasov, 1968)

$$\frac{df(\mathbf{r}, \mathbf{p}, t)}{dt} = 0 \tag{4}$$

Equation (4) can be expanded as:

$$\frac{\partial f_j}{\partial t} + \mathbf{v} \cdot \frac{\partial f_j}{\partial \mathbf{x}} + \frac{q_j}{m_j}\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) \cdot \frac{\partial f_j}{\partial \mathbf{v}} = 0, \tag{5}$$

Where $f_j$ is defined as the particle velocity distribution function for a particle species.

The $q_j$ is the charge of the species of particle, $m_j$ is the mass of the species of particle, $\mathbf{E}$ is the electric field strength, and $\mathbf{B}$ is the magnetic field.

The Vlasov equation describes micro-physical properties and, therefore, can fully describe the evolution of a collisionless plasma (Parks, 2003). The solar wind fluctuations that this project is concerned

with involve changes in the bulk parameters of the flow. Therefore, it is possible to extract sufficient information using moment equations for the flow. For example, the $n^{th}$ moment of $f$ can be found by taking the integral of $v^n f$ with respect to the velocity. The zeroth moment is the number density and is:

$$n_j = \int f_j d^3\mathbf{v}, \tag{6}$$

The first moment is the bulk velocity of the plasma flow:

$$\mathbf{U}_j = \frac{1}{n_j} \int \mathbf{v} f_j \, d^3\mathbf{v} \tag{7}$$

These moment equations can then be subbed back into the moments of the Vlasov equation replacing the $f_j$ term.

$$\int \frac{d}{dt} f_j d^3v \tag{8}$$

This gives the continuity equation of a plasma.

$$\frac{\partial n_j}{\partial t} + \nabla \cdot (n_j \mathbf{U}_j) = 0 \tag{9}$$

Next, replacing the $f_j$ value in the Vlasov equation with the first moment gives the momentum equation.

$$\frac{d}{dt} \int \mathbf{v} f_j \, d^3\mathbf{v} \tag{10}$$

Equation (10) can be expanded to give:

$$n_j m_j \left(\frac{\partial}{\partial t} + \mathbf{U}_j \cdot \nabla\right) \mathbf{U}_j = -\nabla \cdot \mathsf{P}_j + n_j q_j \left(\mathbf{E} + \mathbf{U}_j \times \mathbf{B}\right) \tag{11}$$

Where the $\nabla \cdot \mathsf{P}_j$ term is the divergence of the pressure sensor, this can be derived using the second velocity moment being subbed into the Vlasov equation.

## 3 MAGNETOHYDRODYNAMIC WAVE ASSUMPTION

Magnetohydrodynamics (MHD) studies moving conductive fluids and their interactions with a magnetic field. It is concerned with the bulk properties of this conducting fluid and is the summation of the individual species' properties. The solar wind plasma, being a conducting fluid, and large-scale fluctuations within it can be approximated using these MHD equations. MHD dynamics can be applied when 1) the timescale for interactions in the plasma being investigated is greater than the plasma's gyro-period and collisional relaxation time, 2) the characteristic scale of the interactions in the plasma being investigated is greater than the plasma's gyro-radius, and 3) when the plasma velocities are non-relativistic (Priest, 1982).

The bulk properties of a magnetofluid under the MHD conditions are therefore:

$$\rho \equiv \sum_j m_j n_j \tag{12}$$

with (12) denoting the mass density of the fluid.

$$\mathbf{U} \equiv \frac{1}{\rho} \sum_j m_j n_j \mathbf{U}_j \tag{13}$$

(13) is the bulk velocity of the fluid.

$$P \equiv \frac{1}{3} \sum_j \mathsf{P}_j \tag{14}$$

(14) is the total scalar pressure of the fluid. Using the bulk property equations (12), (13) and (14), the continuity equation for a plasma (9) becomes
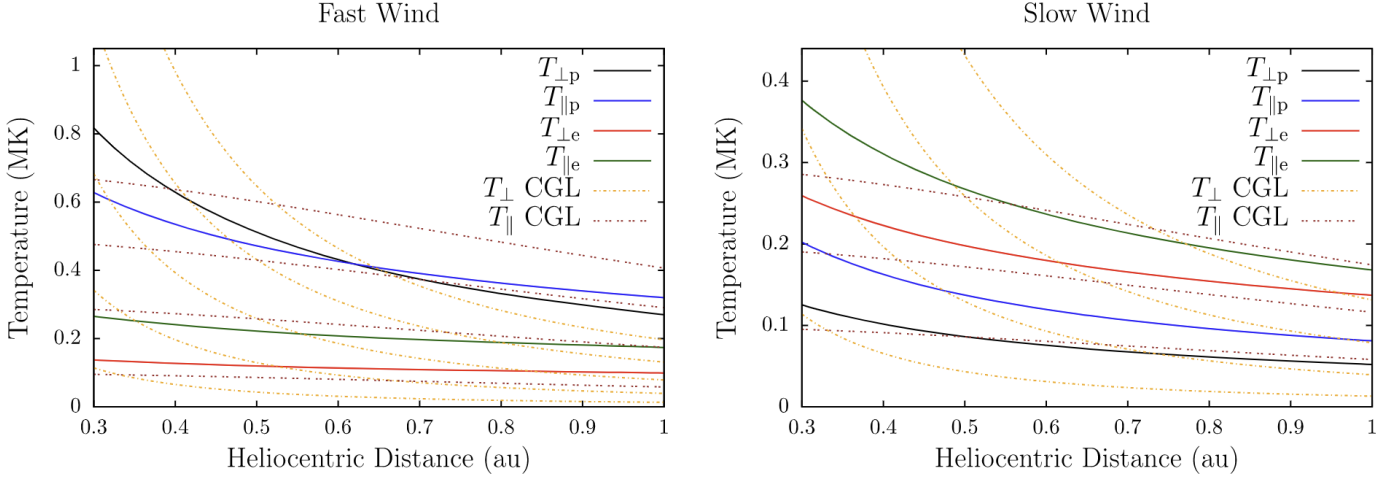
**Figure 1.** A graph showing the temperature profiles within the solar wind, generated by Verscharen u. a. (2019). $T_{\|}$CGL and $T_{\perp}$CGL are the predicted temperature profiles predicted by the CGL invariants. $T_{\perp p}$ is the perpendicular proton temperature, $T_{\|p}$ is the parallel proton temperature, $T_{\perp e}$ is the perpendicular electron temperature, and $T_{\|e}$ is the parallel electron temperature.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{U}) = 0 \qquad (15)$$

under MHD assumptions. The MHD equations assume $r > \lambda_D$, so quasi-neutrality holds. This ensures that there is no electric field component in the MHD momentum equation, thus using (11) it is:

$$\rho \left( \frac{\partial}{\partial t} + \mathbf{U} \cdot \nabla \right) \mathbf{U} = -\nabla \cdot \mathsf{P} + (\mathbf{J} \times \mathbf{B}) \qquad (16)$$

Here, $\mathbf{J}$ is the current density. The equation for Ohm's law within MHD is as follows: (Verscharen u. a., 2019)

$$\mathbf{E} = -\mathbf{U} \times \mathsf{B} \qquad (17)$$

Finally, equation (18) is Faraday's Law.

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \mathbf{E} \qquad (18)$$

Equation (17) and (18) describe Alfvén's frozen in theorem (Alfven, 1963). This theory states that the magnetic flux tubes move with a perfectly conducting fluid as though frozen to it (Roberts, 2007). The implications to the plasma when frozen in a flow are that the same plasma parcels will stay connected to the same field line and therefore do not mix with plasma parcels other field lines.

### 3.1 MHD wave types

Plasma waves within the solar wind can exist as compressive or non-compressive fluctuations.

### Alfvén Waves

Suppose the Solar Wind moves with a bulk velocity of $\mathbf{U}$, as defined in (13), and the magnetic field remains frozen in the flow, as stated in Alfvén's theorem. If any section of this plasma has a local velocity relative to $\mathbf{U}$ - a unique electric field will be generated as stated in Ohm's law for MHD in (17). The curl of this electric field will induce a magnetic field perturbation. If this perturbation creates a curvature in the magnetic field there will be a magnetic tension in the plasma, which will provide a restoring force to the moving plasma, with force being in the direction of $\mathbf{J} \times \mathbf{B}$. The force accelerates the plasma and thus the frozen-in field lines in this direction, this causes a build up of tension in

the opposite direction. This process repeats, and the resulting oscillatory motion is called an Alfvén Wave. These waves are non-compressive fluctuations within plasma and can exist on large enough scales to be described by the MHD equations (Cramer, 2001). First, we define the perturbation magnetic field vector and perturbation velocity vector, $\delta \mathbf{B}$ and $\delta \mathbf{U}$, respectively.

$$\mathbf{B} = \mathbf{B_0} + \delta \mathbf{B} \qquad (19)$$

$$\mathbf{U} = \mathbf{U_0} + \delta \mathbf{U} \qquad (20)$$

Where the $\mathbf{B_0}$ and $\mathbf{B}$ are the background, or average, magnetic field vector and the local magnetic field vector, respectively, $\mathbf{U_0}$ and $\mathbf{U}$ are the average velocity and the local velocity of the wind, respectively. For this wave type $\delta \mathbf{B}$ is perpendicular to to $\mathbf{k}$ and $\mathbf{B_0}$ and there exists a correlation between $\delta \mathbf{U}$ and $\delta \mathbf{B}$. $\mathbf{k}$ is defined as the wave-number in the direction of $\mathbf{B_0}$ (Verscharen u. a., 2019). This can be expressed as:

$$\frac{\delta \mathbf{U}}{v_a} = \mp \frac{\delta \mathbf{B}}{\mathbf{B_0}} \qquad (21)$$

These Alfvén waves also exist on smaller scales at lower wavelengths; at these sizes, however, the MHD assumption breaks down, and kinetic theory is required.

### Slow and Fast Mode Waves

These modes are compressive fluctuations within the solar wind, and both are primarily acoustic. An anti-correlation characterises the slow mode between the plasma density, $\rho$, and the fluctuations in the magnetic field strength, $\delta|\mathbf{B}|$. These slow mode waves are the subject of this project and will be investigated in greater detail in section 3. The fast mode is categorised from a positive correlation between the density, $\rho$, and $\delta|\mathbf{B}|$. Both of these acoustic modes have the total-pressure-gradient force as their restoring forces. However, the fast mode wave also has a magnetic-tension force acting to restore it as well - like with the Alfvén Wave (D'amicis u. a., 2021).

### 3.2 Dampening of plasma waves

Investigation into potential heating mechanisms in the solar wind is critical to understanding the temperature distribution of the wind. These

heating mechanisms result from the dissipation of energy from the dampening of waves within the solar wind. The main dampening mechanisms within the solar wind are quasilinear diffusion, non-linear phase mixing and stochastic heating. An integral aspect of quasilinear diffusion is Landau dampening (Verscharen u. a., 2019).

## Landau Dampening

Landau dampening is when the collective mode of oscillations between the charged particles of a plasma are damped. Collisions between particles are assumed to be rare enough to be neglected. It plays a part in the dampening of all hydrodynamic systems and MHD systems. Under MHD conditions the dampening effects of Landau dampening are significant when the phase velocity of the wave is close to the particle velocity. A particle with a velocity of close to but just under the phase velocity will be accelerated by the wave; thus taking energy out of the MHD wave. This energy is then dissipated through the flow as thermal energy. (Ryutov, 1999)

## 4  SLOW MODE WAVES

### 4.1  How to find Slow Mode Waves

The following section includes quoted passages from Stocker (2022).

Slow mode waves exist primarily within the inertial Range of the solar wind turbulence. (Verscharen u. a., 2019) Further to the anti-correlation between $\delta|\mathbf{B}|$ and $\rho$, the modes $\delta \mathbf{U}$ in the parallel direction to $\mathbf{B_0}$ is anti-correlated with the $\mathbf{B}$. Furthermore, the phase speed of a slow mode wave is equal to the ion acoustic speed of the plasma. The implications of this are that slow mode waves are very efficient at Landau damping. Because of this efficiency at Landau damping, detecting and characterising these waves modes is the project's focus. This local damping could provide thermal energy to the solar wind and explain some of the local heating. Slow mode waves could therefore provide some of this thermal energy - hence their importance.

In figure 2, the magnitude of the magnetic field is negatively correlated with the proton density. In figure 1(a), at 06:23 there is a gradual increase in the magnetic field and a gradual decrease in the proton density. The magnetic field decreases while the proton density increases until at 06:32. There is a fast increase in the Magnetic Field at the same time there is a fall in the proton density. In Figure 2(b) there is a clear spike and dip in the magnetic field and proton density respectively at 08:57. At 09:18 the magnetic field increases simultaneously with a dip in the proton density. Therefore, as both these plots demonstrate anti-correlation between the proton density and the magnitude of the magnetic field, a slow mode wave was passing the Solar Orbiter during these time periods.

### 4.2  Power Spectrum analysis

In figure 2, Verscharen u. a. (2019), show a power spectrum of solar-wind magnetic fluctuations. They take data from the ACE and Cluster 4 spacecraft, ranging from a 58-day time interval to a 1hr time interval. It is clearly shown that the power law for the inertial range is $f^{-\frac{5}{3}}$. As stated, slow mode waves exist in this inertial range, therefore using this frequency spectrum one can determine that slow mode waves can exist on periods of 10 seconds to 7000 seconds.

## 5  SOLAR ORBITER

The Solar Orbiter spacecraft launched from Cape Canaveral on 10 February 2020. It was launched by an Atlas V rocket into orbit around the Sun. The minimum distance the craft will fly to the Sun within the mission is 0.28 AU, limited only by the maximum temperature of the solar arrays recycled from the BepiColumbo. The primary science objective for this mission is "How does the Sun create and control the Heliosphere - and why does solar activity change with time?" (Marirrodriga u. a., 2021) This primary goal is split into four more specific science goals. These goals are concerned with the formation of the Sun's magnetic field and the driving force of the solar wind; by what mechanism drives solar transients; by what mechanism do solar eruptions produce energetic particles; and the working principals behind the solar dynamo.

The Solar Orbiter will build upon previous solar research by exploring regions of the solar environment not explored by spacecraft. The craft will spend greater than five consecutive days at solar latitudes above $25^{o}$ to perform in-situ observations (Müller u. a., 2020). These observations will be the first-ever in-situ measurements outside the solar system's ecliptic. For these observations, the Orbiter has four in-situ instruments: the Energetic Particle Detector (EPD), the Magnetometer (MAG), the Solar Wind Analyzer Suite (SWA) and the Radio and Plasma Waves suite (RPW). Its remote-sensing instruments are the Extreme Ultraviolet Imager (EUI); Metis, which is a coronagraph; the Polarimetric and Helioseismic Imager (PHI); the Solar Orbiter Heliospheric Imager (SoloHi); the Spectral Imaging of the Coronal Environment (SPICE); and the X-ray Spectrometer/Telescope (STIX).

Expanding on the driving force of the solar wind science goal, like the Solar Wind's temperature distribution, the solar Corona's heating mechanism remains unknown. It is theorised that possible convective motions in the Sun's photo-sphere could transfer energy into the Corona through magnetic reconnection (Marirrodriga u. a., 2021). The Corona's temperature is mainly responsible for the Solar Wind and, therefore, the MDH waves within it. The answer to the temperature of the Solar Corona remains relevant to answering the question of heating mechanisms behind the solar wind.

Slow mode waves can be found based on the relationship between the perturbation of the magnitude of the magnetic field, $\delta|\mathbf{B}|$, and the perturbation of the plasma density, $\delta \rho$. Furthermore, due to the quasi-neutrality, the wind density perturbation can only be taken from a specific particle species within the wind. Therefore, for this project, the magnetic field data are taken from the MAG instrument, and the density data are taken from the PAS instrument.

### 5.1  Magnetometer

Solar Orbiter is installed with two fluxgate magnetometers to measure the local magnetic field. Two sensors, MAG-IBS and MAG-OBS, are provided for redundancy - installed on a boom at varying distances of 1m for the IBS instrument and 3m for the OBS instrument (Horbury u. a., 2020). This boom is positioned in the radial plane to the solar Orbiter, pointing away from the Sun. This position ensures that the MAG instruments are always located within the spacecraft's heat shield shade. The instrument remains similar to the magnetometer installed on Cassini and Double star missions, and thus at Solar Orbiters' proximity to the Sun must remain in this shadow. The fluxgate principle was selected for the MAG instrument on Solar Orbiter because of the overall reliability and stability - ideal for a space-based environment. It performs in a way that can measure the local magnetic field directly in vector form. Other magnetometer designs can fulfil this criterion of measuring the field strength directly - namely, the SQUID (superconducting quantum interference devise) magnetometer and the Hall effect magnetometer. The SQUID instrument must be cooled by liquid helium - increasing complexity and reducing reliability, and the Hall effect instrument can
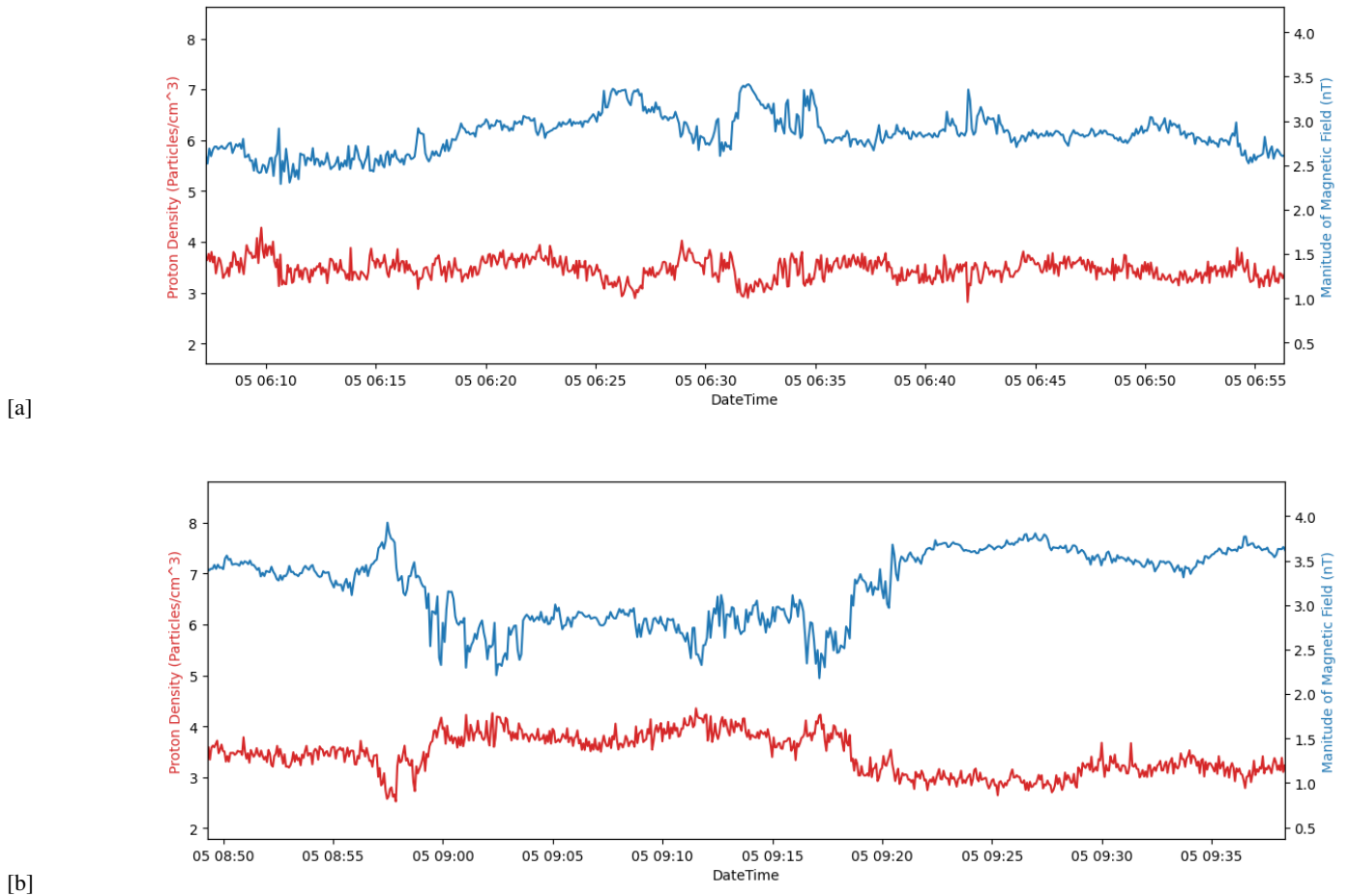
**Figure 2.** [a] is dated between 06:05 on the $5^{th}$ of July 2021, until 06:55 on the same day. [b] is dated between 08:50 until 09:40 on the same day as [a]. Graph is illustrating the anti-correlation between magnetic field strength and proton density in the solar wind. The proton density is in red and the magnetic field data is in blue. Data, as with all from this project, are taken from the Solar Orbiter Archive [a].

[a] https://soar.esac.esa.int/soar/

not reach the sensitivity of the fluxgate design.

The fluxgate principle involves two wires and a magnetically permeable alloy. Solar Orbiter's MAG instrument uses duel magnetically permeable alloys. These alloys are orthogonal to each other and cemented into position by a Macor ceramic block - Ultra Electronics provided this unit (Horbury u. a., 2020). One of the wires is termed the drive winding, and the other the sense winding. They are both wrapped around the core at different orientations. The drive wire has an alternating current flowing through it. If there is no external magnetic field both cores are saturated simultaneously, not inducing a current in the sense wire. However, in the presence of an external magnetic field, the orthogonal magnetically permeable alloys reach saturation at different times and a detectable current is induced in this sense wiring (Cao u. a., 1979). This induced current is then converted into magnetic field readings by the MAG's front-end electronics, located in the MAG electronics box within the spacecraft.

## 5.2 Solar Wind Analyser

The Solar Wind Analyser (SWA) is comprised of four sensors. These are the electron Analyser System (SWA-EAS), which measures the velocity distribution function of the electrons within the flow. The Proton-Alpha Sensor (SWA-PAS) measures ion species' velocity distribution within the flow and their moments. The Heavy Ion Sensor (SWA-HIS) will measure the velocity distribution of certain ion species, including their respective numbers and charges within the solar wind flow. It also contains a data processing unit termed the SWA-DPU. All sensors can take data with a cadence of below 10 seconds (Owen u. a., 2020).

This project is concerned with proton and ion density, so all data is taken from the SWA-PAS sensor. The instrument outputs are the total proton and ion count, the density $[particles \, cm^{-3}]$, the velocity $[km \, s^{-1}]$ in the stellar reference frame (SRF), the velocity $[km \, s^{-1}]$ in the radial, tangential and normal (RTN) frame, the pressure tensor $[j \, cm^{-3}]$ in the SRF frame, the pressure tensor in the RTN frame and the temperature of the local plasma. The SWA-PAS contains an electrostatic analyser and an array of conventional channel electron multipliers (CEM), which detect the species of plasma being measured. The SWA-PAS sensor must face the incoming plasma and be positioned in the spacecraft heat-shield cutout. This sensor has a curved-out backing to allow solar radiation to pass through and independent heat shielding. Two charged deflector plates follow the entrance slit of the sensor; these deflect the relevant particles through the electrostatic analyser towards the
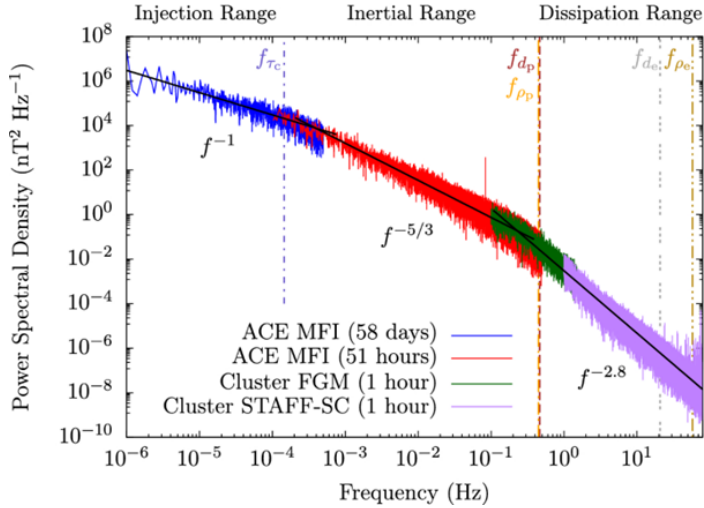
**Figure 3.** Verscharen u. a. (2019) generated a power spectral density analysis of the solar wind magnetic field fluctuations. Magnetic field data is taken from the ACE and Cluster 4 spacecraft, with data intervals ranging from 58-days to 1hr.

CEM to be measured. The field of view of the sensor is $-22.5^o$ to $22.5^0$, deflection from the deflector plates is dependent on this input angle and velocity, and the count on the CEM array is a function of the particles' velocity distribution (Owen u. a., 2020).

## 6 MACHINE LEARNING

There are three categories of machine learning: supervised, unsupervised, and reinforcement learning. Supervised learning is when an algorithm 'learns' from labelled data sets - to either cluster data or build a regression model to predict future values. Unsupervised learning is when an algorithm 'learns' from patterns within an unlabeled data set. This unlabelled data are commonly placed into groupings called clusters. Finally, reinforcement learning is when an algorithm attempts to interact with its environment with initially random behaviour, and it 'learns' through being rewarded or punished based on its success at the required task (Géron, 2019). This project is attempting to characterise slow mode waves within the solar wind , so supervised and unsupervised learning would both be effective at achieving this. For a supervised model to work, one would first need to provide sufficient training data; this is an example of when there is a negative correlation between the magnitude of the magnetic field and the proton density. Once the supervised algorithm had finished learning with the data, its accuracy would be validated. This accuracy validation would take more labelled data, but this time the algorithm would be attempting to predict if there was a slow mode wave or not in the data set. The predictions would then be compared with the labels, and if the algorithm's accuracy proved above a predetermined threshold - the algorithm could then be fed new unlabeled data and would return where slow mode waves are within this data set.

The unsupervised alternative has a prepossessing stage with the time series data converted into points within a defined parameter space. The algorithm then finds clusters within this space. The clusters will be mapped according to patterns, thus grouping points with similar characteristics. Therefore, if clusters are found with parameters expected for slow mode waves, it is likely that the algorithm has found slow mode waves. The parameters to be displayed in this space would have to include the correlation between the magnitude of the magnetic field and the plasma density, as this is critical to defining a slow mode wave.

Table (1) displays the key positives and negatives of using supervised or unsupervised learning for this task. The critical positives for the supervised machine learning algorithm are that there is no need for a prepossessing stage with most algorithms; time-series data can be used. Furthermore, once the algorithm has completed training and validation, no more model updates need to be completed. This gained simplicity means that using the algorithm is very easy when testing novel data. However, supervised algorithms require an extensive training data set to learn from. This training set means that thousands of instances of slow mode wave data must be manually labelled, a very time-consuming task and one that is very prone to human error. If some of these slow mode waves are improperly labelled, the accuracy of the model's findings would be severely reduced. Finally, great care would need to be made to ensure that the data set used for this training data included all wave speeds. For example, if the algorithm is only trained on labelled instances of the wave mode during the slow wind, it might not spot a slow mode wave during the fast wind.

On the other hand, unsupervised learning does not require a labelled data set as it finds its patterns within the data set. However, there needs to be a prepossessing stage for this machine learning method. This process is slow and memory intense each time new solar wind data are analysed. After this prepossessing stage is complete, the model's user-defined parameters, or hyperparameters, must be updated for that specific data set. Trends within the parameter space will vary depending on wind types and the length of the investigated period. Even with the negatives with unsupervised learning, because it does not require training data, a potential source of error, this is the preferable machine learning type for finding slow mode waves within the solar wind .

### Why use Machine Learning?

As a slow mode wave is defined by the anti-correlation between the magnitude of the magnetic field and the plasma density, a feasible method to find this wave type could be as follows. Find the correlation of these two parameters over a selected period, and if the correlation is below a predetermined value, there is a slow mode wave in the data set. However, this method is inferior to finding slow mode waves as outlined above in the unsupervised machine learning method because the algorithm determines the parameters that define this wave mode itself. It determines this based on wave characteristics within the selected parameter space. Implications of this are that the exact values which define the slow mode wave depend on the wind's characteristics rather than being generic for all wind types.

### 6.1 Machine Learning in Solar Physics

The following section is directly quoted from Stocker (2022).

Machine Learning is rising in prevalence in categorising and predicting solar events. For example, Colak und Qahwaji (2009) used a supervised Machine Support-Vector algorithm was used to predict when a solar flare is likely to happen and how big it will likely be. This algorithm was trained with more than a year of sunspot and flare data.

Reiss u. a. (2021) used training data from STEREO-A and STEREO-B was used to predict all three magnetic field components in the solar wind . These spacecraft measure solar events by remote sensing as well as taking in-situ measurements like the spacecraft in Lagrange-point 1. Due to the large delay on in-situ data, extracting more data from remote sensing sensors can dramatically reduce response time. Therefore using this method, there is more warning on solar events to earth than when using in-situ spacecraft.

Camporeale u. a. (2017) used the classification of different wave types of the solar wind is attempted. They use a Gaussian Process classification

**Table 1.** Table to highlight the positives and negatives between different machine learning algorithms

|  | Supervised | Unsupervised |
|---|---|---|
| **Positives** | Time series data could be used to train the algorithms; no need for preprocessing. The algorithm is looking for trends so it can be used on fresh data without needing to retrain. | Does not need labelled data. Finds patterns in data so define a slow mode wave based on the data. |
| **Negatives** | Required a labelled data set. Likely to miss slow mode waves if a wind parameter changes, eg for a CME. | Data needs to be preprocessed before it can be analysed. The algorithm's user-defined parameters need to be altered for new data. |

algorithm and classify the data into four subgroups: ejecta, coronal holes, sector reversal and streamer belts. For this classification, seven input dimensions have a median accuracy of greater than 90% for all categories.

Li u. a. (2020) is another example of using machine learning to categorise solar wind . They also use four categories based on how the specific plasma was produced. These categories are coronal-hole-origin plasma, streamer-belt-origin plasma, sector-reversal-region plasma, and ejecta. They used ten different supervised classification algorithms. K-nearest neighbour performed best with a 92.8% accuracy. They also concluded they could get the most reliable results when eight input dimensions were used.

Finally, different solar wind states were categorised by D. Aaron Roberts und Lepri (2020) to trace the states' origins back to the sun. They used an unsupervised K-means algorithm to determine these states.

## 6.2 Selected Algorithms

The following section is directly quoted by Stocker (2022).

*6.2.1 KMeans and mini-batch KMeans* K-means clustering is an unsupervised algorithm which separates sample data into N groups of equal variance. It must be told how many categories the data will be sorted into, labelled C. It takes this value and assigns C data points randomly as the centroids. It then assigns each data point into one of the C groups based on physical proximity to the randomly selected centre points. The average position is then mapped, and the data are re-categorised now based on its proximity in euclidean space to these average points. This process is iterated until the centres no longer move. Then the within-cluster sum-of-squares criterion is applied to the newly found clusters: (Hubert und Arabie, 1985)

$$\arg \min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \mu_i\|^2 \tag{22}$$

This process is then repeated a predetermined number of times, and the final clusters are the ones which have the lowest value in the within-cluster sum-of-squares criterion. As each cluster will have a single scalar variance value, their shapes must be circular. This regularity presents a weakness if the desired cluster has an elongated shape.

Elkan (2003) improved this method. The new algorithm, termed accelerated K-means, exploits triangle inequality which removes unnecessary distance calculations. It is also possible with K-means to input a NumPy array as the rough centres, leading K-means to select the random points only in these selected areas. This user selection is called the centroid initialisation method. This method could be beneficial in finding slow mode waves as one of these rough centre points could be positioned at a region on the graph of a negative correlation - forcing K-means to look for clusters where the slow mode waves should be located.

A technique first suggested in Arthur und Vassilvitskii (2006).

Minibatch K-means, first proposed by Sculley (2010), moves the centre slightly with each iteration. This movement increases the algorithm's speed by up to a factor of 4.

*6.2.2 Gaussian mixture* Gaussian mixture generates a probabilistic model with the input data, assuming that the clusters are Gaussian distributions whose parameters are unknown. This distribution includes the variance of the data set and the cluster mean enabling ellipsoid-shaped clusters, each having unique shapes, sizes, densities and orientations. Similarly to K-means, the number of clusters is a user-defined variable, but unlike K-means - because of this ellipsoidal shape - elongated clusters can be correctly categorised. Each cluster is modelled with a mean value and a variance as described by equation (23), where the observation, $\boldsymbol{x^{(i)}}$, within a cluster, $c^{(i)}$, with $\mathbf{k}^{(i)}$ clusters. Each of these clusters has a mean value in euclidean space, $\mu^{(j)}$, and covariance matrix, $\sum^{(j)}$ (Bouman, 1995). It states that this observation, given its cluster, is proportional to the normal distribution of the mean value and the covariance matrix.

$$\boldsymbol{x^{(i)}} \| c^{(i)} = \mathbf{k}^{(i)} \sim \mathcal{N} \left( \mu^{(j)}, \sum{}^{(j)} \right) \tag{23}$$

*6.2.3 Agglomerative cluster* Agglomerative clustering is an example of hierarchical clustering. It begins by assigning each data value to its cluster and then grouping the data based on proximity in euclidean space. This process repeats until there is only one cluster remaining. This process results in a hierarchy tree containing each cluster's formation called a dendrogram (Hubert und Arabie, 1985). The number of clusters is a user-defined variable, N, so the algorithm will output the stage of the dendrogram from when N clusters remain. For example, if there are 60 data points to be clustered by an Agglomerative clustering algorithm, it begins by assigning all data points to a cluster. There are now 60 clusters forming the first level of the dendrogram. The closest clusters are combined with each iteration, and the new clusters are recorded on the dendrogram. Further iterations follow until there is only 1 cluster remaining. It then takes the user-defined variable, N, and exports the dendrogram at the stage where N clusters remain.

*6.2.4 DBSCAN* DBSCAN is an algorithm which clusters based on the density of the input data. There are two hyperparameters, $\epsilon - neighborhood$ and minimum samples. It begins with a random point in the data set, and if it has greater than the minimum samples hyperparameter within $\epsilon - neighborhood$ distance, including itself, this data point is considered a core instance. It then tests all data points within that neighbourhood. All instances within $\epsilon - neighborhood$ distance of a core instance are clustered together. If a data point is not

$\epsilon - neighborhood$ distance away from a core instance, it is termed an anomaly, and the point is discarded (Khan u. a., 2014).

## 6.3 Prepossessing

The preferred method for finding slow mode waves is an unsupervised machine learning algorithm that clusters data points within a parameter space. This parameter space, therefore, has to be defined. We used the correlation between the magnitude of the magnetic field and the plasma density over a designated period as one of these parameters. We used as a second parameter the average value between the magnitude of the magnetic field and the plasma density. As under MHD conditions, quasi-neutrality holds for the plasma, the perturbation density of the solar wind, $\delta \rho$, is approximately the same as the perturbation density of the protons in the wind, $\delta \rho_p$. This similarity allows for the density data can be taken only from Solar Orbiter's SWA-PAS instrument[1]. The magnetic field density is taken from Solar Orbiter's magnetometer instrument.

We first download the density and magnetic field data using the Solar Orbiter's Sunpy API. The MAG and PAS instruments have different cadences and do not take data with the same epoch. Therefore, the first stage of prepossessing is to match the epochs of these two data sets and convert the EPOCH time into a DateTime format. We do this with Pandas' API. We then set the DateTime as the index of a data frame containing both magnetic field data and proton density data. We then define which time windows to average and correlate the data into. In section 3.2, it was found that the inertial range of the power spectral density exists between periods of 10 seconds to 7000 seconds. Therefore the time windows within these limits must be chosen. We chose time windows of 900, 1800, 2700 and 3600 seconds to average and correlate the data over. We test more than one time window because slow mode waves exist on this inertial range. Next, we had to find the perturbation of the density and the magnetic field, and then both values were normalised. For simplicity, these perturbation numbers will, from now on, be called deltas. This stage is expressed below.

$$\delta|\mathbf{B}| = \frac{\mathbf{B} - \mathbf{B_0}}{B_0}, \tag{24}$$

$$\delta \rho = \frac{\rho - \rho_0}{\rho_0} \tag{25}$$

As the average field and density values, $B_0$ and $\rho_0$, are being subtracted from the field and density figures, negative values for the magnetic field delta and the proton density delta are possible. We found the values of $B_0$ and $\rho_0$ by averaging over the same time window used to calculate the average and correlation parameters.

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2(y_i - \overline{y})^2}} \tag{26}$$

$$a = \frac{\overline{x}\,\overline{y}}{2} \tag{27}$$

Finally, we find the correlation between the magnitude of the magnetic field and the proton density over the selected time windows. This was done using the Pearson correlation equation given in equation (26). Where $r$ is the correlation coefficient, $x_i$ are the values within a time window of the density delta, $\overline{x}$ is the mean of the density delta data within the time window, $y_i$ is the magnetic field delta data within the time window, and $\overline{y}$ is the average magnetic field delta within the time window. These two columns were then averaged over the same time window as with the correlation. This is given in equation (27), where $a$ is the combined average. This process yields a new data frame, with columns

---

[1] https://soar.esac.esa.int/soar/

of average, $a$, and correlation, $r$, of the magnitude of the magnetic field delta and the proton density delta, indexed with the date-time of the first data point within the time window. Therefore the higher the time window, the fewer rows there are in this correlation-average data frame.

As the velocity delta, $\delta \mathbf{U}$, in the parallel direction to the magnetic field average, $\mathbf{B_0}$, is anti-correlated with the magnitude of the magnetic field, $\delta|\mathbf{B}|$, we investigated this relationship. The $\delta|\mathbf{B}|$ value had already been found. So, to get the parallel velocity perturbation, we took the velocity readings also from Solar Orbiter's SWA instrument. First, the velocity deltas are defined in the same way as the magnetic field delta and the proton density delta.

$$\delta \mathbf{U} = \frac{\mathbf{U} - \mathbf{U_0}}{U_0} \tag{28}$$

Next, we found the velocity component in the parallel plane to $\mathbf{B_0}$; this was done with the following equation.

$$\delta \mathbf{U_{par}} = \frac{\delta \mathbf{U} \cdot \mathbf{B_0}}{|\mathbf{B_0}|} \tag{29}$$

The Pearson Correlation between this $\mathbf{U_{par}}$ was then found with the magnetic field delta - using equation (26). This correlation would have worked with either the proton density or the magnetic field delta. When a slow mode wave is present, there will be an expected anti-correlation between the parallel velocity perturbation and the magnetic field, but the field is anti-correlated with the proton density.

## 6.4 Model Limitations

The outlined method successfully characterises the solar wind into clusters within parameter space. The locations of the found clusters indicate characteristics of the solar wind at that instant - these parameters are then to be compared with the expected wind characteristics to determine if these clusters contain slow mode waves. However, there are some limitations to the proposed model. Firstly, because only a single wave type is being characterised, the interference of this and other wind fluctuations can change the characteristics of the wave - leading to a potential false classification. These missed waves are critical because mapping slow mode waves are being done to determine if they could dissipate enough heat to heat the solar wind. If this wave mode is being missed because of other fluctuations, such as Alfvén waves, the findings could be weakened.

Secondly, we use two months of time series data to collect the results for this project. The data was downloaded, and the DateTime of the magnetic field and proton density data were matched. However, the Solar Orbiter Archive occasionally can, on rare occasions, miss a day of data. This gap in the data set happened on average six times per time series analysed and was more common for the MAG data. The structure of the algorithm stitched these discontinuous time series together. The averages, $a$, and correlations, $r$, were then created over their respective time windows, and if these were taken over a point where there was missing data, the point in question would be corrupted. This error could be fixed by creating a new data frame if there is a time gap in the DateTime data of more than 1 minute, far greater than the 4-second cadence of the PAS instrument. The averages and correlations would then be taken over each data frame. This way, data would not be used, but no correlation or average data would be corrupted by this means.

## 7   RESULTS

Figures 3 to 15 were created with time series data between the dates of 01 January 2022 and 01 March 2022, which was the last available date
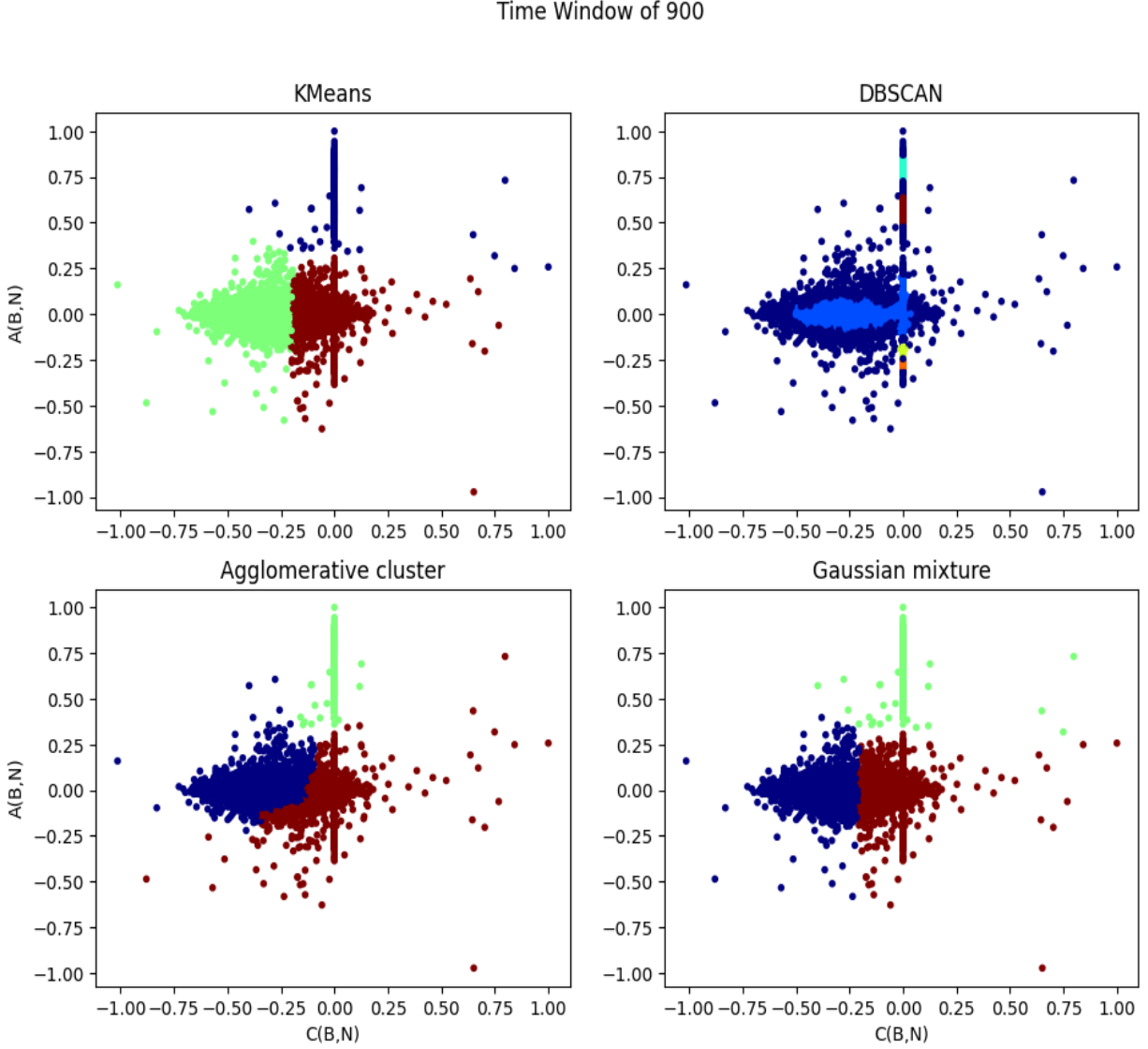
**Figure 4.** A 2-dimensional parameter space of comparing A(B,N) and C(B,N). The time window used for this data was 900 seconds. Clusters to the left side of this parameter space have an anti-correlation between the proton density delta and magnetic field delta, and so have characteristics of a slow mode wave.

for data as of writing on 8 July 2022. During this time, the Solar orbiter was approaching its closest perihelion to date. The distance from the Sun ranges from 0.99 AU to 0.57 AU. Figures 4 to 7 have axes labeled A(B,N) and C(B,N). A(B,N) refers to the average, $a$, between the magnetic field delta and the density delta over the time window. C(B,N) refers to the correlation, r, between the magnetic field delta and the density delta over the time window. Figures 4, 5, 6 and 7 all have A(B,N) on the y-axis and C(B,N) on the x-axis. Therefore, clusters on the left side of this presented parameter space show a negative correlation between the magnetic field deltas and the proton density deltas over the chosen time window. These are the clusters which, therefore, likely contain slow mode waves.

Firstly, the data distribution under a 900-second time window, as seen in figure 3, includes a line of data points extending up the 0 correlation axis. This line extends to the average being equal to 1 but only runs as low as A(B,N) < -0.5 axis. This location suggests a large quantity of high-density wind with a strong magnetic field, but these two parameters

are uncorrelated. Potentially, they have an oscillation $\pi$ radians out of phase. There is also a distinct diamond shape of high-density data points which extends from approximately -0.75 to 0.25 on the C(B,N) axis and is essentially symmetric about A(B,N) = 0 axis, ranging from +0.25 to -0.25. This distribution suggests that most time-series data between the two months investigated have an anti-correlation between the magnetic field delta and the plasma density delta. Some outliers exist, notably 14 data points with a correlation of greater than 0.5. Suggesting these are fast-mode waves within the wind.

### 900 Second Time Window

The hyperparameters chosen for the KMeans algorithm were the number of clusters as three and a maximum iteration number of 600 iterations. The line extending up the 0 C(B,N) axis provides the foundation for the first cluster, which is in blue. The diamond shape has been split into two clusters, critically with one in the negative C(B,N) region, coloured in

green. This green cluster's boundary runs approximately vertically down the -0.25 on the C(B,N) axis.

The hyperparameters chosen for the DBSCAN algorithm were a neighbourhood of 0.17, and the minimum number of samples to a cluster was 40. This algorithm is the only one to work based solely on the points' density in the parameter space. The implications of this are the centre of the diamond shape, being a high-density point, has been placed in its cluster. Other clusters are located along the 0 C(B,N) line. The dark blue points are what DBSCAN has deemed as outliers.

The hyperparameters chosen for the Agglomerative cluster are setting the number of clusters to 3 and the models linkage to *ward*. *Ward* combines different clusters based on the variance of individual clusters, with the clusters with the lowest variance combined. This model has produced results similar to KMeans, with the line in C(B,N) = 0 being clustered and the diamond shape being split into two separate clusters. Unlike KMeans, however, the Agglomerative cluster has not split the diamond shape by a horizontal line. For this algorithm, the dark blue colour is the cluster with the points which demonstrate the most anti-correlation between the magnetic field delta and the proton density delta.

The hyperparameters chosen for the Gaussian mixture model were the number of components as three and a covariance type as *tied*. The *tied* ensures that all clusters share the same general covariance matrix. The Gaussian mixture model has clustered data in a very similar way to the KMeans algorithm. The line at C(B,N) = 0 is a cluster, and the diamond shape is split in half with a vertical line at approximately C(B,N) = -0.25. As with the Agglomerative cluster, the dark blue cluster is the one which demonstrates the most anti-correlation between the magnetic field delta and the proton density delta.

## 1800 Second Time Window

In figure 4, with a time window of 1800 seconds, similar data trends can be seen as in figure 3.

The only hyperparameter changes for all algorithms were lowering the number of clusters for KMeans and Agglomerative cluster to 2 and decreasing the number of components of the Gaussian mixture model to 1. KMeans, Agglomerative Cluster and Gaussian mixture all found similar clusters on the left-hand side of the parameter space. These are coloured blue, red and blue, respectively. The agglomerative cluster was the most conservative, followed by KMeans and Gaussian mixture having the most points in this left cluster. Finally, DBSCAN found a single cluster, and it is located on the left-hand side of the parameter space. This cluster, in red, is located in the centre of the high-density region of the data spread.

## 2700 Second Time Window

The time window of 2700 seconds, as shown in figure 5, shows another clear line running up C(B,N) = 0. The primary data distribution also exists on the left-hand side of the figures. KMeans, Agglomerative Cluster and Gaussian Mixture have all found clusters which exist only on the left-hand side of the graphs, coloured blue, green and green, respectively. These algorithms have also found two other clusters within the parameter space. DBSCAN has found three clusters within this space, with the dark blue points classed as outliers.

## 3600 Second Time Window

For the time window of 3600 seconds, in figure 6, the data distribution interestingly contains 7 data points with a correlation between the

magnetic field delta and proton density delta above 0.5. There are also 7 data points with this level of correlation for the time windows 2700 and 1800. These points lead to speculation that out of the 13 high correlation points for the 900-second time window - potential fast mode waves - 7 of them exist on scales of greater than 1 hour. With this time window, KMeans, Agglomerative Cluster and Gaussian mixture model all found three categories, with one existing only on the left-hand side of the parameter space. For this time window, the Agglomerative cluster was the most conservative, followed by KMeans - with Gaussian mixture clustering some points with a correlation of 0, suggesting that for this time window, the Gaussian Mixture may demonstrate some inaccuracies with the categorisation of slow mode waves. DBSCAN again found a cluster in the middle of the high-density period of data points, with all the dark blue points labelled as outliers. This sole cluster was only on the left-hand side of the parameter space.

## Analysis of all Time Windows

When investigating the effectiveness of all the algorithms over the different time windows, the results must be compared. Firstly, since DBSCAN clusters are based on the data points' proximity in euclidean space and the data distribution for all time windows having a distinct area of high-density points, DBSCAN commonly found a cluster in the centre of this high-density region. Although, in most cases, this cluster falls exclusively on the left side of the parameter space - in every time window, many data points fall further left of this cluster. From observing the distribution of the data points, most of the points to the left of the cluster do not appear to be outliers and follow the general pattern observed for all distributions over all time windows. Therefore, although a cluster is being found on the left-hand side of the parameter space, suggesting a cluster of potential genuine slow mode waves, too many other points are being - wrongly - concluded as outliers. The conclusion is that DBSCAN is not an effective algorithm for finding slow mode waves within this data and thus will not be included in further analysis.

In analysing the remaining algorithms, each found three clusters for time windows of 900, 2700 and 3600 seconds. These clusters can be categorised based on their general shapes. The far-left cluster, coloured green for the KMeans with a 3600-second time window, can be termed cluster 1. The cluster with the highest average coloured red for KMeans with a 3600 second time window is termed cluster 2. Moreover, the final cluster with a position lower on the A(B,N) axis, coloured in blue for the KMeans algorithm for a 3600-second time window, is termed cluster 3. Cluster 1 generally falls exclusively on the left side of the parameter space. Its general distribution lies largely symmetrical around the A(B,N) = 0 axis. Cluster 2 exists largely symmetrical about the C(B,N) = 0 axis while being high on the A(B,N) axis. This high average indicates that for these time windows, it is common for the magnetic field and the proton density to be higher than $B_0$ and $\rho_0$ while not showing a correlation between their oscillations over the time window. Cluster 3 appears centered along the C(B,N) = 0 and A(B,N) = 0 axis', with a slight bias towards the left side of the parameter space.

Finally, for the time window of 1800, the data distribution was not such that the remaining algorithms could split the data into only two clusters. The left-hand cluster is negative on the C(B,N) axis, suggesting it contains slow mode waves, while the cluster to the right contains all other wave types. Finding this cluster on the left-hand side of the parameter space suggests that there is a clear distinction between slow mode waves and other data for this time window. The effectiveness of the algorithms at this time window lead to further analysis. For this 1800-second time window, the prepossessing stage has cut two months of
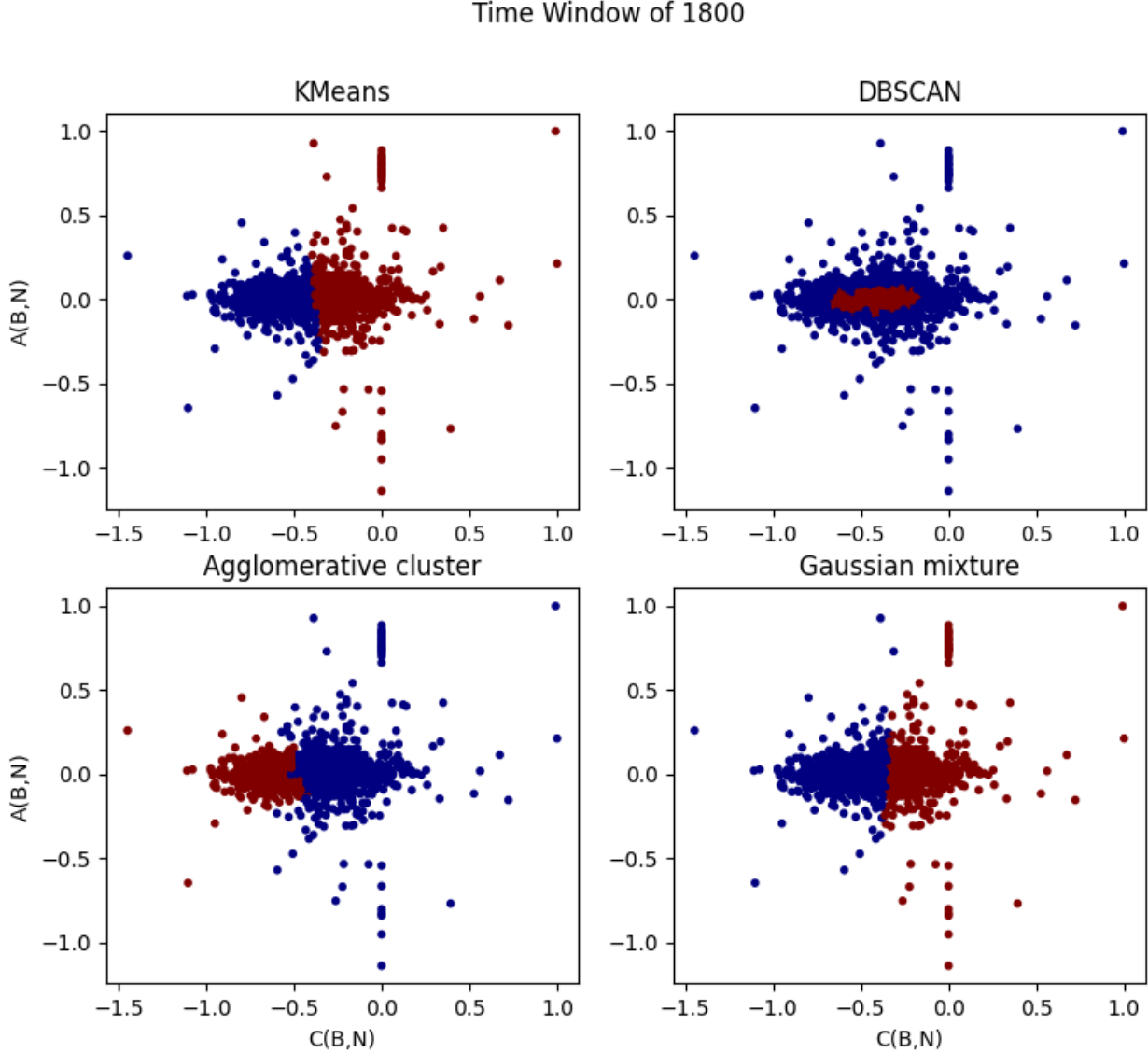
## Time Window of 1800



**Figure 5.** A 2-dimensional of comparing A(B,N) and C(B,N). The time window used for this data was 1800 seconds. Clusters to the left side of this parameter space have an anti-correlation between the proton density delta and magnetic field delta, and so have characteristics of a slow mode wave.

data into 2603 data points. In the left-hand cluster, the algorithms found 1) 730 points for KMeans, 2) 847 points for DBSCAN, 3) 529 points for Agglomerative cluster, and 4) 833 points for Gaussian Mixture. So neglecting DBSCAN, Gaussian Mixture has found the most potential slow mode waves, followed by KMeans and Agglomerative Cluster is the most conservative, only clustering points with a strong correlation.

Figure 8 illustrates the patterns these algorithms find during the two-month time series. Each slow mode wave that the algorithms are finding are represented by a vertical line. The x-axis of this plot is the date signalling the times slow mode waves are located within the solar wind. From this graph, the slow mode waves are grouped together times of a high density of this wave mode and other times where there are none of this wave mode present. Each algorithms attempts at categorising slow mode waves are represented by a different colour. These areas of high and low density appear at irregular intervals throughout the time series. DBSCAN (orange) is visibly finding slow mode waves at times the other

algorithms saying there is not a slow mode wave present. As the left cluster on the Gaussian Mixture from figure 4 is the largest cluster, all of the points within the KMeans and Agglomerative cluster fit within it. The same applies to the KMeans cluster and Agglomerative cluster. The left-hand cluster in the Agglomerative cluster algorithm includes all points within the KMeans cluster, meaning it would fit perfectly within the left-hand cluster from KMeans. The implications of this are that in figure 7 all of the lines for the Agglomerative cluster are present within the KMeans lines. Then all of the KMeans lines are included in the Gaussian Mixture's lines.

### 7.1 Parallel Velocity Delta and Slow Mode Waves

As stated in section 3.1, the velocity delta component in the plane parallel to the average field, $\mathbf{B_0}$, is anti-correlated with the magnetic field delta when a slow mode wave is present. We tested this correlation using the clusters already created by the unsupervised machine learning algorithms. This check was done by plotting a third dimension to the
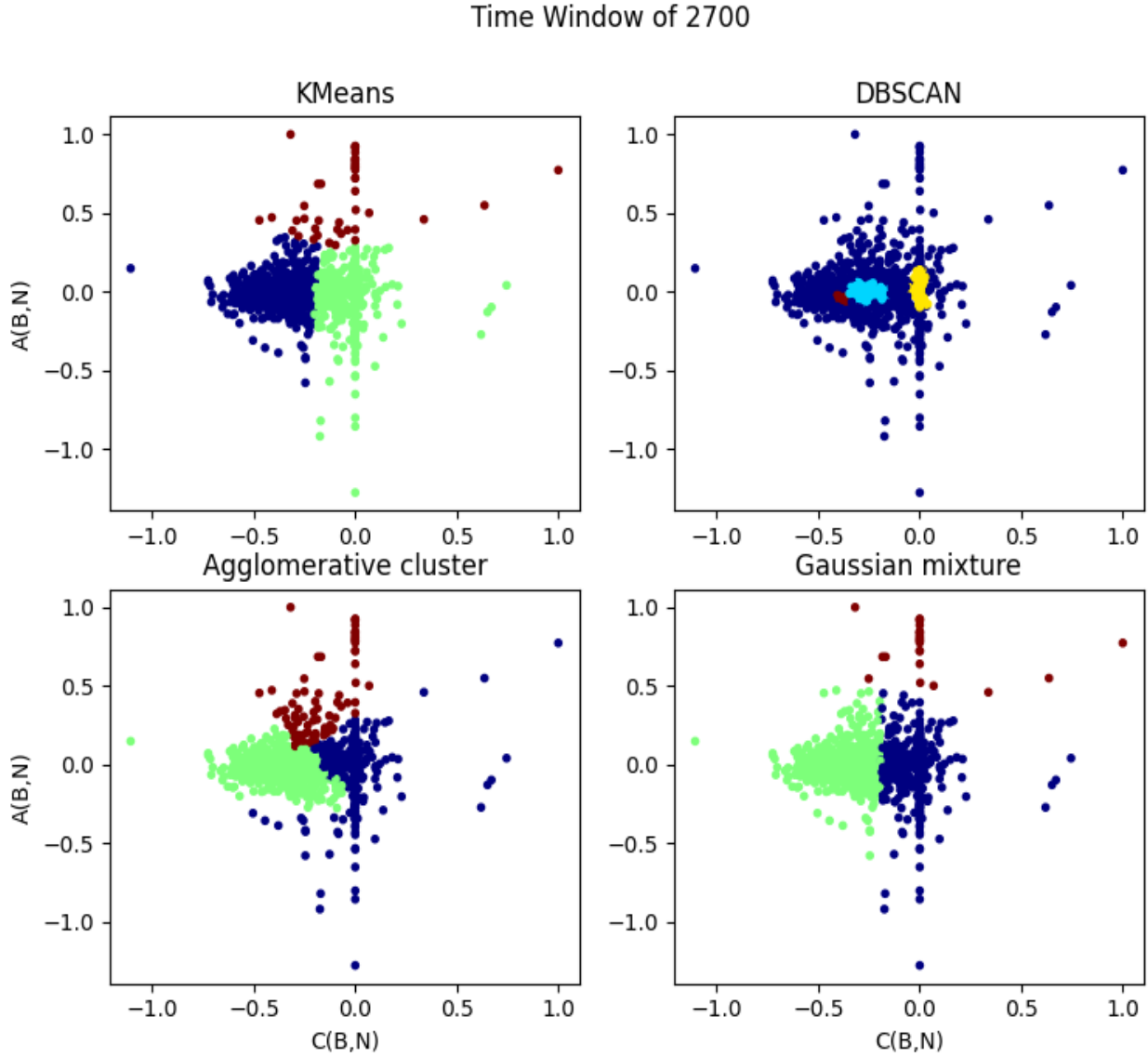
**Figure 6.** A 2-dimensional of comparing A(B,N) and C(B,N). The time window used for this data was 900 seconds. Clusters to the left side of this parameter space have an anti-correlation between the proton density delta and magnetic field delta, and so have characteristics of a slow mode wave.

plots in Figures 5 and 7. We do not use this velocity delta to create these clusters; we are testing the correlation between the velocity delta and the already found clusters. The 3-dimensional plots have a C(B,N) and A(B,N) on the x-axis and y-axis to mean the same as before, but now there is a C(B,V) z-axis. This new axis represents the correlation between the parallel velocity delta and the magnitude of the magnetic field delta. Therefore, if there is an anti-correlation between the velocity delta and the magnetic field delta, the clusters representing slow mode waves appear lower on the C(B,V) axis.

Looking at the KMeans graph in figure 9, the blue cluster is the slow mode wave cluster at C(B,N)¡0. The diamond shape formed in the higher density points in figure 4 transfers into the 3-dimensional space to form an octahedron shape. This blue cluster in the 3-dimensional space tends towards the C(B,V) axis as the C(B,N) gets more negative, with the data points getting closer to C(B,V) = 0. These results, therefore, indicate no correlation between the parallel velocity delta and the magnetic field

when there is a slow mode wave as defined by C(B,N).

Figure 10 investigates the correlation between the parallel velocity delta and the magnetic field delta over a 3600 second time window. In figure 10, it is again the blue cluster which contains slow mode waves. As the C(B,N) axis gets more negative, the data does not trend towards the plot's negative C(B,V) area. This shape again indicates no correlation between the magnetic field delta and the parallel velocity delta. The graphs were generated to investigate if the trend is visible at the 900-second or 2700-second time intervals; as with Figures 9 and 10, we see no correlation, so these graphs were not included in this report. With figure 10, however, the green cluster named cluster 3 in the previous section, has a large distribution along the C(B,V) axis. The points within this green cluster have a minimal correlation between the magnetic field delta and the density delta, but strong correlations exist between the magnetic field delta and the parallel velocity delta.
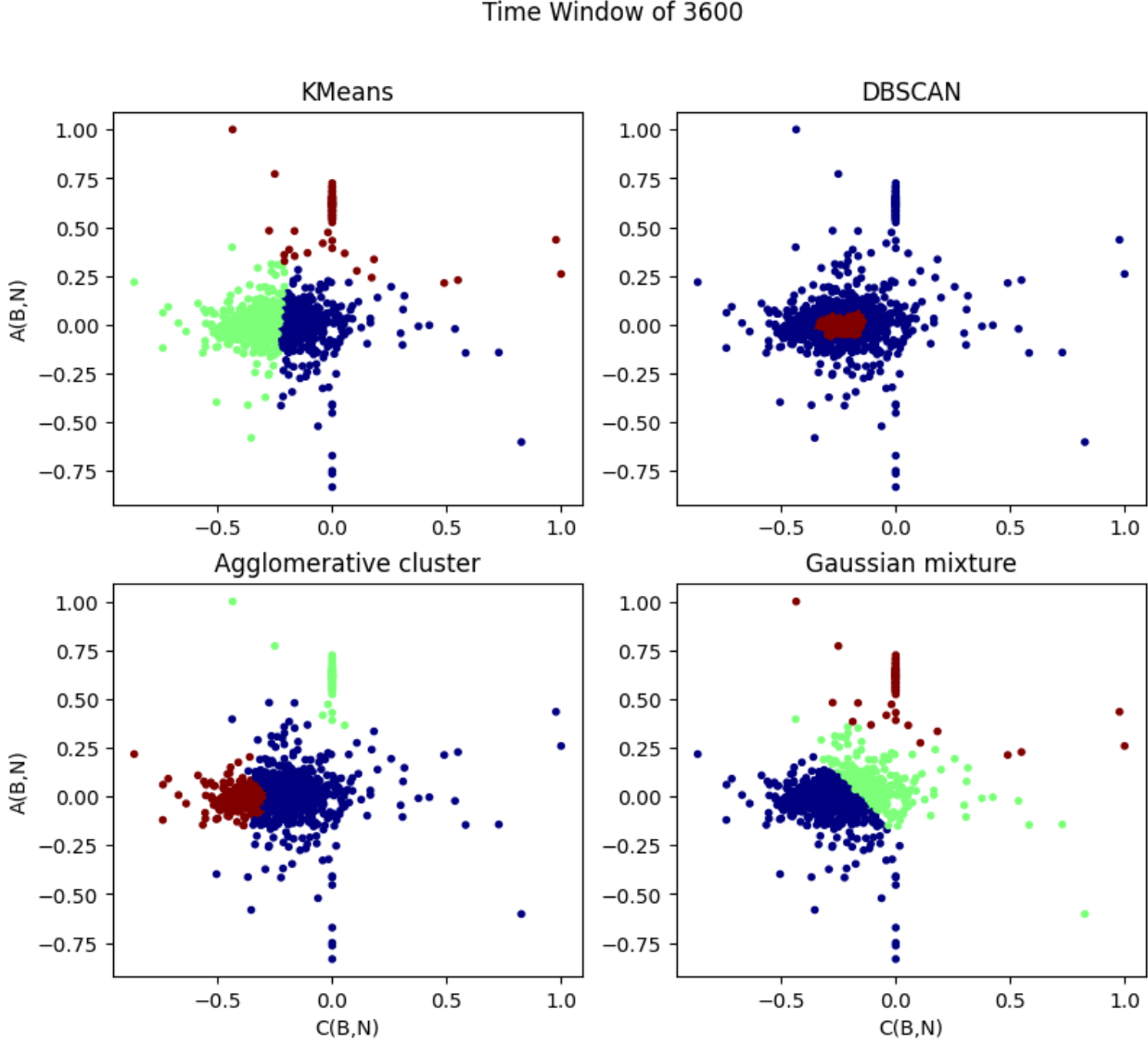
**Figure 7.** A 2-dimensional comparing A(B,N) and C(B,N). The time window used for this data was 900 seconds. Clusters to the left side of this parameter space have an anti-correlation between the proton density delta and magnetic field delta, and so have characteristics of a slow mode wave.

As these results were unexpected, we investigated if these same trends could be seen with data taken from a different time. The new dates that were selected were between 20 August 2021 and 20 November 2021, when Solar Orbiter was at 0.66 AU and 0.75 AU, respectively. The plots over a time window of 1800 seconds are shown in figure 10. The general distribution of data remains similar to figure 5, with a diamond shape of high-density data points residing primarily within the left-hand side of the parameter space. As well as a clear line of high density points extending up the C(B,N) = 0 axis. The performance in finding clusters on the left-hand side of this parameter space was very similar to figure 4, with DBSCAN performing poorly and the remaining algorithms succeeding. The number of slow mode waves categorised remains the same, with Gaussian Mixture finding the most slow mode waves. KMeans found the second most of these modes waves, and the Agglomerative cluster found the least. Inspecting the KMeans plot extended onto the 3-dimensional plot with C(B,V) in figure 11, as the C(B,N) axis gets more negative, the data close in on the C(B,V) = 0

plane. This shape was again an unexpected result as no correlation is demonstrated with the magnetic field delta and the parallel velocity delta when there is a slow mode wave. Although not shown as a distinct cluster in figure 11, the mentioned high-density line running up the C(B,N) = 0 axis shows considerable variation along the C(B,V) dimension. All other generated 3-dimensional graphs with axes C(B,N), A(B,N) and C(B,N) for the other time windows all showed the same trend. This trend was no visible correlation between the magnetic field delta and the parallel velocity delta.

### 7.2 Wind Velocity and Slow Mode Waves

In figure 13, the solar wind velocity has been plotted against the time series we investigated. It can be seen from this graph that the velocity readings can be seen to fluctuate significantly over the two months. With the velocity reaching above 700 $km\,s^{-1}$, and decreasing to below 300 $km\,s^{-1}$. The KMeans results, printed over the top of these velocity data, show clear gaps in certain regions, as referenced when we described the
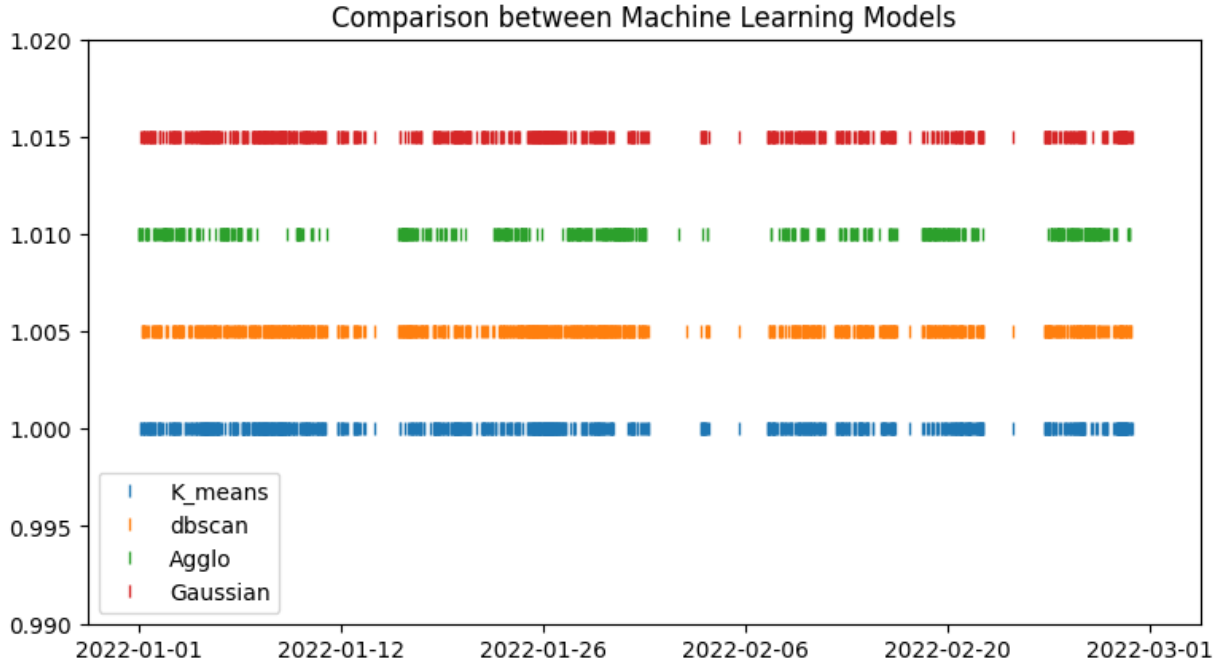
**Figure 8.** Comparison between ML models, each coloured vertical line means the algorithm has found a slow mode wave at this given time. Data are spread along the two months time series. Lines in blue corresponding to 1 on the y-axis are the results of the KMeans algorithm. Lines in orange corresponding to 1.005 on the y-axis are the DBSCAN algorithm's results. Lines in green corresponding to 1.010 on the y-axis are from the Agglomerative Cluster Algorithm. Lines in red corresponding to 1.015 on the y-axis are from the Gaussian Mixture algorithm.
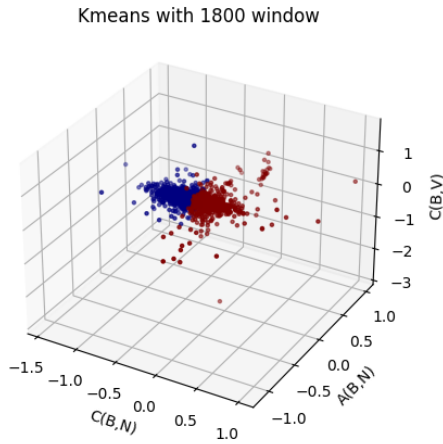


**Figure 9.** KMeans algorithm extended into 3-dimensions to test correlation over a 1800 second time window. The z-axis is the correlation between the magnetic field delta and the parallel velocity delta.
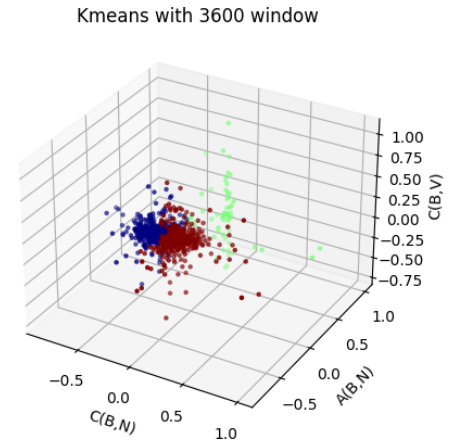


**Figure 10.** KMeans algorithm extended into 3-dimensions to test correlation over a 3600 second time window. The z-axis is the correlation between the magnetic field delta and the parallel velocity delta.

findings in figure 7. Figure 12 demonstrates a correlation between the solar wind velocity and the KMeans results. The sections where there is a gap in the slow mode waves correlate strongly to when there are peaks in the solar wind's velocity. In figure 12[a], there are five wave velocity spikes registering a velocity greater than 450 $km\,s^{-1}$.

The first of these spikes coincide with a gap in the KMeans results. The wind velocity then decreases again to approximately 400 $km\,s^{-1}$,

and the KMeans results can be seen to display infrequent slow mode waves. Then the velocity drops further, down to 300 $km\,s^{-1}$, at which point the KMeans result shows a high density of slow mode waves over this 1800-second time window. There is then an increase in wind velocity to approximately 400 $km\,s^{-1}$, and there is another gap in the KMeans results. This gap is followed by KMeans finding intermittent slow mode waves as the velocity drops again. The velocity then stays relatively low, apart from a spike in velocity after the 2021-09-24, which coincides
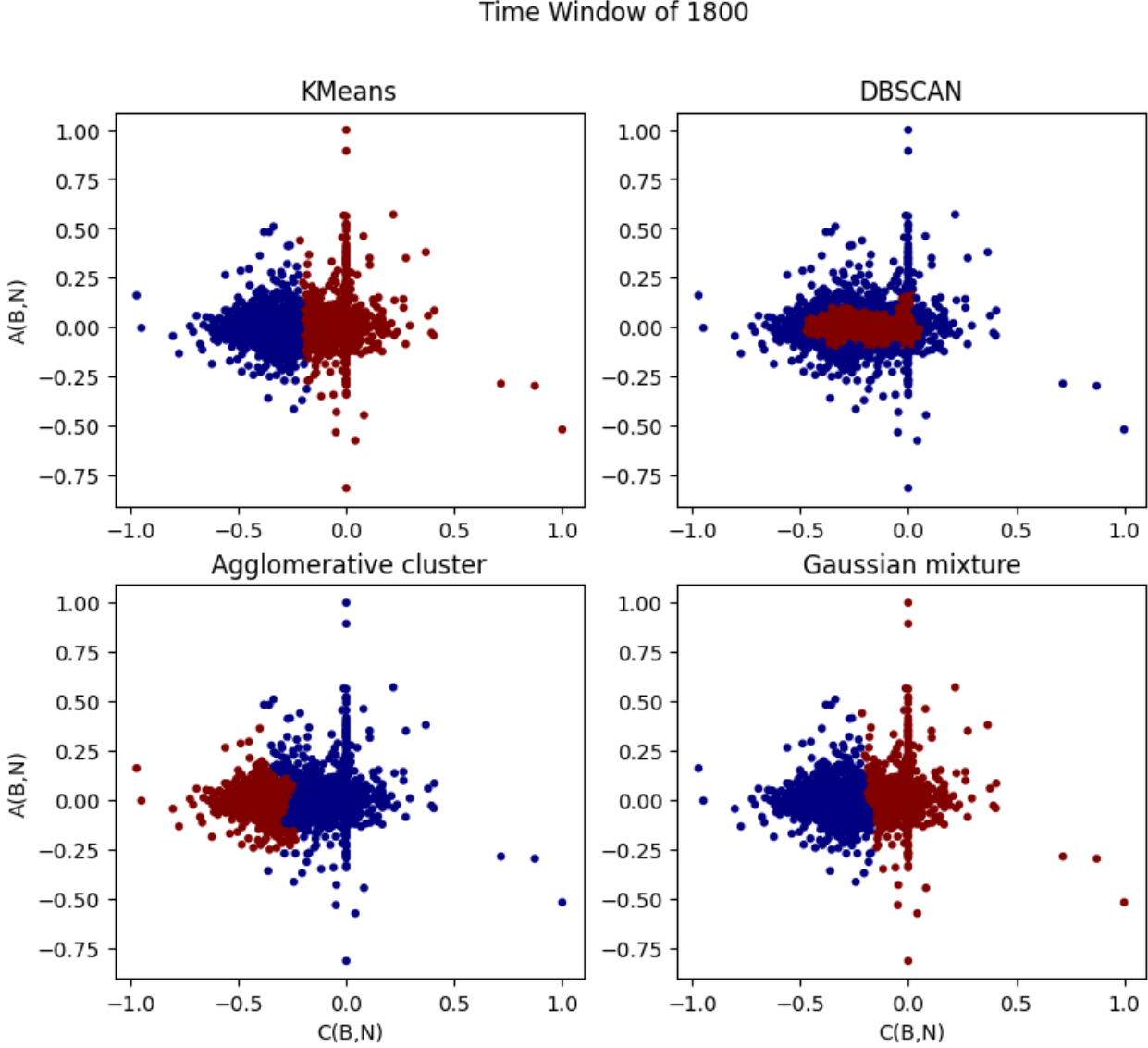
**Figure 11.** Visualisation of clusters found by four algorithms using a 1800 second time window from 20 August 2021 to 20 September 2021. Clusters found to the left side of the parameter space have an anti-correlation between the magnetic field delta and the proton density delta, and so demonstrate the characteristics of slow mode waves.

with a small decrease in the number of slow mode waves. Again, the velocity stays low, and the KMeans results display a continuous number of slow mode waves. This high number of slow mode waves continues until the wind velocity increases again, where the number of slow mode waves decreases again until a decrease follows a final velocity spike in the number of recorded slow mode waves. Figure 12[b] contains a more varied velocity profile, with regular spikes and dips throughout the two-month time series. The range of the velocity profile remains between 250 $km\,s^{-1}$ and 700 $km\,s^{-1}$. The same trends can be seen in this figure as in figure 12[a], with increases in velocity coinciding with gaps within the recorded slow mode wave data.

### 7.3 Wind Temperature and Slow Mode Waves

From inspecting the temperature distributions in figure 14, it is evident that the spikes in temperature observed happen at the same time as spikes in the velocity plots in figure 7. Furthermore, the KMeans results are

the same for figure 13 as they are for figure 12; therefore, figure 13 also shows a general correlation trend between the solar wind temperature and the number of slow mode waves predicted by KMeans.

## 8 DISCUSSION

The results indicate that the unsupervised machine learning algorithms have found clusters that exist only in the anti-correlation side of the parameter space. Therefore, all data points within these clusters display a negative correlation between the magnetic field delta and the proton density delta over the chosen time window. As mentioned in section 6, DBSCAN was inappropriate for finding slow mode waves. The remaining algorithms followed the general trend of Agglomerative Cluster being the most conservative, KMeans being the second most and then Gaussian Mixture being the least, finding the largest cluster for potential slow mode waves. The algorithms succeeded in categorising
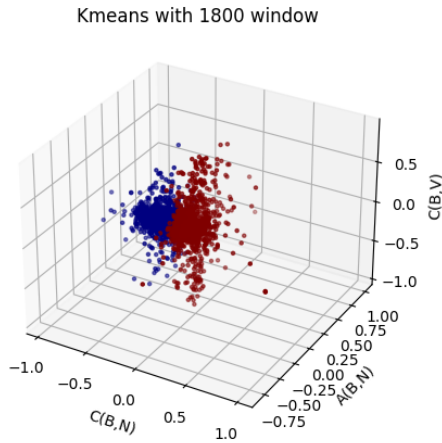
Kmeans with 1800 window

**Figure 12.** A 3-dimensional visualisation of the data in figure 10 over a time window of 1800 seconds. The z-axis is the correlation between the magnetic field delta and the parallel velocity delta.

slow mode waves, but we saw no correlation between the parallel velocity deltas and the magnetic field deltas in figures 9, 10 and 12. A likely reason is that Alfvén waves exist in the solar wind flow simultaneously with the slow mode waves. As stated in section 2.1.1, these Alfvén waves are non-compressive; thus, no correlation between the magnetic field deltas and the proton density deltas would be expected.

Alfvén waves do not have a correlation between the parallel velocity delta and magnetic field delta. However, if strong Alfvén waves are present in the data there is a lot of variation in the magnetic field direction, this brings uncertainty to the direction of $\mathbf{B_0}$. The parallel velocity is found using equation (29), if the direction of $\mathbf{B_0}$ is uncertain this introduces uncertainty to the direction of the parallel velocity fluctuations. The parallel fluctuation deltas relative to $\mathbf{B_0}$, therefore, will not show a correlation with the magnetic field deltas, as seen in figures 9, 10 and 12. Testing smaller time windows may decrease this uncertainty in $\mathbf{B_0}$ and this could produce a correlation between the parallel velocity delta and the magnetic field delta.

The relationship between the distribution of slow mode waves and the wind velocity could also be explained by the presence of Alfvén waves. This relationship exists between the solar wind bulk velocity and the magnetic field deltas in the presence of Alfvén waves; when the velocity rises, this could cause more rapid changes in the proton density Cramer (2001). These changes could ruin the correlation between this magnetic field delta and the proton density delta.

Figure 7 illustrates the distribution of slow mode waves over varying distances from the Sun. As the data is taken from a part of Solar Orbiter's orbit when it was reaching perihelion, it had a considerable velocity. Therefore in the two months of data collection, the solar orbiter travelled from 0.99Au to 0.57Au. A total difference of 0.42Au, or moving $6.2810^9 km$ closer to the Sun over the time series. Therefore, considering how efficient slow mode waves are at Landau damping, it is anticipated that there would be an increase in the total number of the wave mode as the solar orbiter moves closer to the Sun, assuming they are created at the Sun. This increase in slow mode wave numbers is not what the data shows.

## 9 CONCLUSION

We found in this project that unsupervised machine learning models can cluster the solar wind based on parameters consistent with slow mode waves. We found these clusters within similar parts of the parameter space, suggesting that the slow mode waves have consistent characteristics over all investigated time windows. We concluded that DBSCAN is not a good algorithm for detecting slow mode waves because it works based on the parameters' density in the space. KMeans, Gaussian Mixture and Agglomerative Cluster can all successfully find slow mode waves within the solar wind. Based on the further analysis completed in this project, KMeans has been considered the most successful algorithm, finding the second most slow mode waves within the flow. This algorithm ensured that all of the data has a large enough anti-correlation to fall within the typical characteristics of a slow mode wave. Using this KMeans algorithm, we conducted further analysis to find a correlation between the slow mode waves and the parallel wind velocity. No correlation was found. This unexpected non-correlation suggests that other wind fluctuations, such as Alfvén waves, affect the parallel velocity. Therefore we do not recommend using parallel velocity for future work categorising slow mode waves. We then looked at how the distribution of slow mode waves was affected by other wind parameters, namely the bulk velocity and temperature. We saw the number of slow mode waves in the solar wind decrease when the temperature and velocity spiked. Critically, this analysis shows that the number of slow mode waves recovers within the wind flow after being disturbed - suggesting either that the mechanism that produces slow mode waves happens within the solar wind rather than at the Corona - where the wind is formed, or that when the velocity or the temperature is high, the correlation between the magnetic field delta and proton density delta temporarily breaks down. The former solution implies that there must be an external mechanism creating these waves providing energy to the system for slow mode waves to be a contender to be the hidden heating mechanism in the solar wind.

Building upon the analysis within this work is necessary to calculate how much thermal energy the dampening of slow mode waves gives to the solar wind. A next step in this research is to analyse more data - with a comparison of the number of slow mode waves and other parameters within the solar wind over longer time scales at varying distances and latitudes from the Sun. This project has demonstrated that slow mode waves can be found using only the winds magnetic field deltas and proton density deltas, so we recommend only using these to train a KMeans algorithm - ensuring more parameters are available to test correlations with. These correlations would be useful because mapping these wave modes and the increased understanding of how they behave around other wind parameters will aid in finding the rate of dissipation of the wave mode in the wind. This dissipation rate will then give an idea of how much thermal energy is given to the wind. Furthermore, as unsupervised machine learning can successfully find this wave mode, this data can be used to train a supervised algorithm. This supervised algorithm would increase the speed at which the algorithm works to find slow mode waves. Finally, it would be useful to re-analyse the correlation between slow mode waves and the parallel velocity of the wind. For this, we would recommend using shorter time windows for the results. Even though Alfvén waves remain stronger than slow mode waves at all scales. On smaller scales, $\mathbf{B_0}$ may remain have less uncertainty in its direction - producing less uncertain parallel velocity delta figures.

This project has shown that the number of slow mode waves in the solar wind depends on the solar wind velocity. If the periods of high wind velocity were removed from the time series data, the number of slow mode waves would remain constant throughout the time series. This consistency indicates that a mechanism must produce these slow
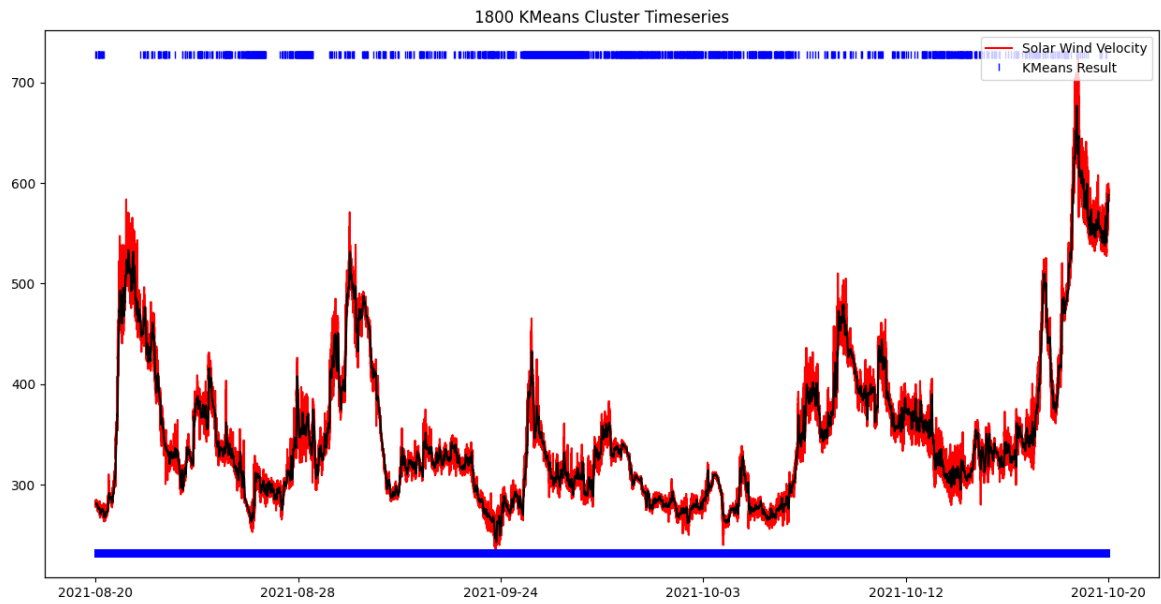
[a]
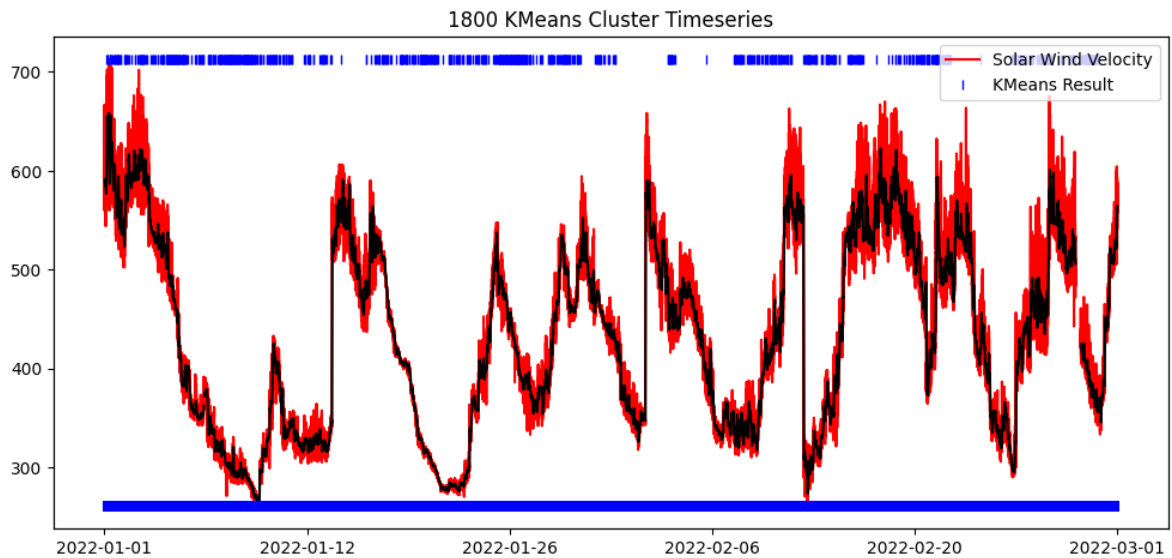


[b]

**Figure 13.** [a] is a time series plot of the solar wind velocity with units $km^{-1}$. The dates are from 20 August 2021 to 20 September 2021. [b] is also a time series of solar wind velocity [$km^{-1}$] between the dates 1 January 2022 to 1 March 2022. For both plots, the black line is a rolling average. Each blue marker at the top of the plot, the KMeans algorithm had detected a slow mode wave at that time.

mode waves within the solar wind rather than these waves forming with the solar wind at the Corona. This does not necessarily suggest that the dissipation of slow mode waves is not a primary heater of the solar wind. It does suggest, however, that if the dissipation of slow mode waves is heating the solar wind, the mechanism must provide energy to the solar wind system.

## REFERENCES

[Alfven 1963]    ALFVEN, H.: *Cosmical Electrodynamics*. Clarendon P., 1963 (International series of monographs on physics pt. 1). – URL https://books.google.co.uk/books?id=qVwVAwAAQBAJ. – ISBN 9785882322501

[Arthur und Vassilvitskii 2006]    ARTHUR, David ; VASSILVITSKII, Sergei: k-means++: The advantages of careful seeding / Stanford. 2006. – Forschungsbericht
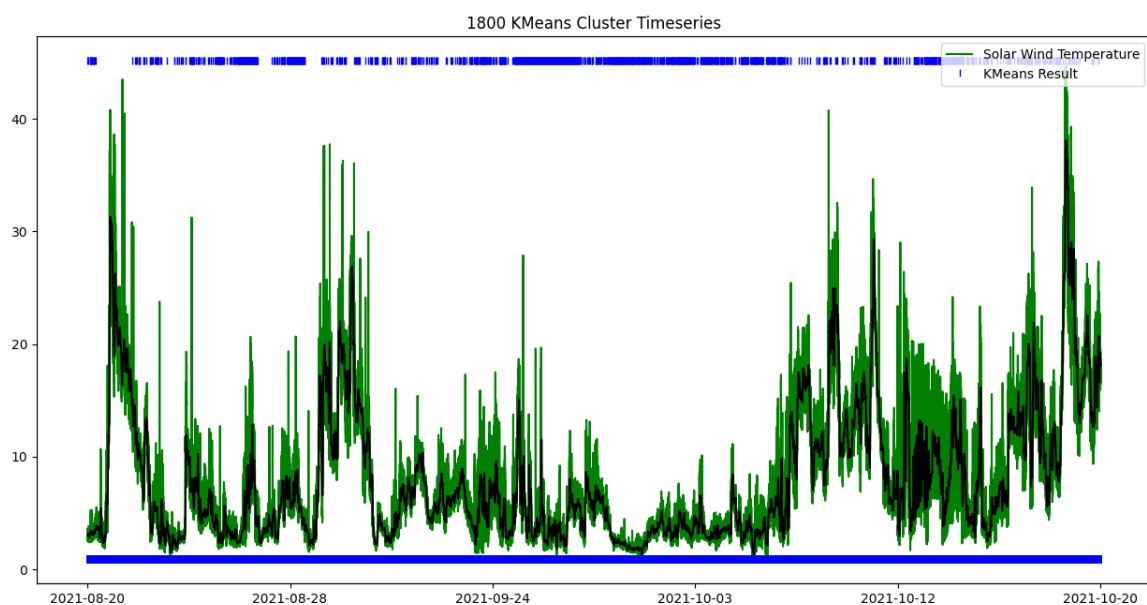
[a]
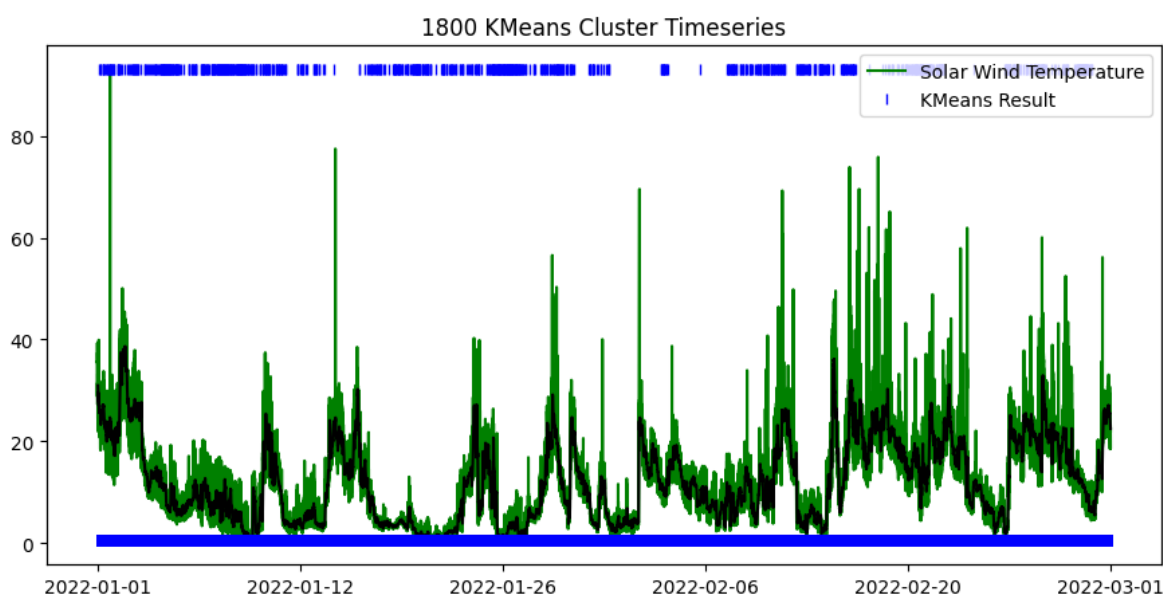


[b]

**Figure 14.** [a] is a time series of solar wind temperature with units eV. The green line is the temperature data plotted with no prepossessing from the solar orbiter archive. The dates are from 1 January 2022 to 1 March 2022. [b] is the same solar wind temperature (eV) time series but from 20 August 2021 to 20 September 2021. For both plots, the black line is a rolling average. Each blue marker at the top of the plot, the KMeans algorithm had detected a slow mode wave at that time

[Baranov und Kartalev 1972] BARANOV, V. B. ; KARTALEV, M. D.: Evolutionality of shock waves in the Chew-Goldberger-Low approximation. In: *Fluid Dynamics 1974 7:6* 7 (1972), 11, S. 1028–1030. – URL https://link.springer.com/article/10.1007/BF01176127. – ISSN 1573-8507

[Bouman 1995] BOUMAN, Charles A.: CLUSTER: An Unsupervised Algorithm for Modeling Gaussian Mixtures. In: *Purdue University* (1995)

[Camporeale u. a. 2017] CAMPOREALE, Enrico ; CARÈ, Algo ; BOROVSKY, Joseph E.: Classification of Solar Wind With Machine Learning. In: *Journal of Geophysical Research: Space Physics* 122 (2017), Nr. 11, S. 10,910–10,920. – URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JA024383

[Cao u. a. 1979] CAO, Jianan ; ZHAO, Juan ; CHENG, Shuai ; PRIMDAHL, F: Journal of Physics E: Scientific Instruments The fluxgate magnetometer. In: *J. Phys. E: Sci. Instrum* 12 (1979)

[Colak und Qahwaji 2009] COLAK, T. ; QAHWAJI, R.: Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. In: *Space Weather* 7 (2009), Nr. 6. – URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008SW000401

[Cramer 2001] CRAMER, N.F.: *The Physics of Alfven Waves*. Wiley, 2001. – URL https://books.google.co.uk/books?id=3L_vAAAAMAAJ. – ISBN 9783527402939

[D. Aaron Roberts und Lepri 2020] D. AARON ROBERTS, Tamara Sipes Yuan-Kuen K. ; LEPRI, Susan: Objectively Determining States of the Solar Wind Using Machine Learning. In: *The American Astronomical Society* (2020)

[D'amicis u. a. 2021] D'AMICIS, R ; BRUNO, R ; PANASENCO, O ; TELLONI, D ; PERRONE, D ; MARCUCCI, M F. ; WOODHAM, L ; VELLI, M ; MARCO, R D. ; JAGARLAMUDI, V ; COCO, I ; OWEN, C ; LOUARN, P ; LIVI, S ; HORBURY, T ; ANDRÉ, N ; ANGELINI, V ; EVANS, V ; FEDOROV, A ; GENOT, V ; LAVRAUD, B ; MATTEINI, L ; MÜLLER, D ; O'BRIEN, H ; PEZZI, O ; ROUILLARD, A P. ; SORRISO-VALVO, L ; TENERANI, A ; VERSCHAREN, D ; ZOUGANELIS, I: First Solar Orbiter observation of the Alfvénic slow wind and identification of its solar source. (2021)

[Elkan 2003] ELKAN, Charles: Using the Triangle Inequality to Accelerate k-Means. (2003)

[Géron 2019] GÉRON, Aurélien: *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc., 2019. – https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

[Horbury u. a. 2020] HORBURY, T S. ; O'BRIEN, H ; BLAZQUEZ, I C. ; BENDYK, M ; BROWN, P ; HUDSON, R ; EVANS, V ; ODDY, T M. ; CARR, C M. ; BEEK, T J. ; CUPIDO, E ; BHATTACHARYA, S ; DOMINGUEZ, J.-A ; MATTHEWS, L ; MYKLEBUST, V R. ; WHITESIDE, B ; BALE, S D. ; BAUMJOHANN, W ; BURGESS, D ; CARBONE, V ; CARGILL, P ; EASTWOOD, J ; ERDÖS, G ; FLETCHER, L ; FORSYTH, R ; GIACALONE, J ; GLASSMEIER, K.-H ; GOLDSTEIN, M L. ; HOEKSEMA, T ; LOCKWOOD, M ; MAGNES, W ; MAKSIMOVIC, M ; MARSCH, E ; MATTHAEUS, W H. ; MURPHY, N ; NAKARIAKOV, V M. ; OWEN, C J. ; OWENS, M ; RODRIGUEZ-PACHECO, J ; RICHTER, I ; RILEY, P ; RUSSELL, C T. ; SCHWARTZ, S ; VAINIO, R ; VELLI, M ; VENNERSTROM, S ; WALSH, R ; WIMMER-SCHWEINGRUBER, R F. ; ZANK, G ; MÜLLER, D ; ZOUGANELIS, I ; WALSH, A P.: Special issue The Solar Orbiter magnetometer. In: *AA* 642 (2020), S. 9. – URL https://doi.org/10.1051/0004-6361/201937257

[Hubert und Arabie 1985] HUBERT, Lawrence ; ARABIE, Phipps: Comparing partitions. In: *Journal of Classification* 2 (1985), Dec, Nr. 1, S. 193–218. – URL https://doi.org/10.1007/BF01908075. – ISSN 1432-1343

[Khan u. a. 2014] KHAN, Kamran ; REHMAN, Saif U. ; AZIZ, Kamran ; FONG, Simon ; SARASVADY, Sababady: DBSCAN: Past, present and future. In: *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)* IEEE (Veranst.), 2014, S. 232–238

[Li u. a. 2020] LI, Hui ; WANG, Chi ; TU, Cui ; XU, Fei: Machine Learning Approach for Solar Wind Categorization. In: *Earth and Space Science* 7 (2020), Nr. 5, S. e2019EA000997. – URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019EA000997. – e2019EA000997 10.1029/2019EA000997

[Marirrodriga u. a. 2021] MARIRRODRIGA, C. G. ; PACROS, A. ; STRANDMOE, S. ; ARCIONI, M. ; ARTS, A. ; ASHCROFT, C. ; AYACHE, L. ; BONNEFOUS, Y. ; BRAHIMI, N. ; CIPRIANI, F. ; DAMASIO, C. ; JONG, P. D. ; DÉPREZ, G. ; FAHMY, S. ; FELS, R. ; FIEBRICH, J. ; HASS, C. ; HERNÁNDEZ, C. ; ICARDI, L. ; JUNGE, A. ; KLETZKINE, P. ; LAGET, P. ; DEUFF, Y. L. ; LIEBOLD, F. ; LODIOT, S. ; MARLIANI, F. ; MASCARELLO, M. ; MÜLLER, D. ; OGANESSIAN, A. ; OLIVIER, P. ; PALOMBO, E. ; PHILIPPE, C. ; RAGNIT, U. ; RAMACHANDRAN, J. ; PÉREZ, J. M. S. ; STIENSTRA, M. M. ; THÜREY, S. ; URWIN, A. ; WIRTH, K. ; ZOUGANELIS, I.: Solar Orbiter: Mission and spacecraft design. In: *Astronomy Astrophysics* 646 (2021), 2, S. A121. – URL https://www.aanda.org/articles/aa/full_html/2021/02/aa38519-20/aa38519-20.htmlhttps://www.aanda.org/articles/aa/abs/2021/02/aa38519-20/aa38519-20.html. – ISSN 0004-6361

[Müller u. a. 2020] MÜLLER, D. ; CYR, O. C. S. ; ZOUGANELIS, I. ; GILBERT, H. R. ; MARSDEN, R. ; NIEVES-CHINCHILLA, T. ; ANTONUCCI, E. ; AUCHÈRE, F. ; BERGHMANS, D. ; HORBURY, T. S. ; HOWARD, R. A. ; KRUCKER, S. ; MAKSIMOVIC, M. ; OWEN, C. J. ; ROCHUS, P. ; RODRIGUEZ-PACHECO, J. ; ROMOLI, M. ; SOLANKI, S. K. ; BRUNO, R. ; CARLSSON, M. ; FLUDRA, A. ; HARRA, L. ; HASSLER, D. M. ; LIVI, S. ; LOUARN, P. ; PETER, H. ; SCHÜHLE, U. ; TERIACA, L. ; INIESTA, J. C. Del T. ; WIMMER-SCHWEINGRUBER, R. F. ; MARSCH, E. ; VELLI, M. ; GROOF, A. D. ; WALSH, A. ; WILLIAMS, D.: The Solar Orbiter mission - Science overview. In: *Astronomy Astrophysics* 642 (2020)

[Owen u. a. 2020] OWEN, C. J. ; BRUNO, R. ; LIVI, S. ; LOUARN, P. ; JANABI, K. A. ; ALLEGRINI, F. ; AMOROS, C. ; BARUAH, R. ; BARTHE, A. ; BERTHOMIER, M. ; BORDON, S. ; BROCKLEY-BLATT, C. ; BRYSBAERT, C. ; CAPUANO, G. ; COLLIER, M. ; DEMARCO, R. ; FEDOROV, A. ; FORD, J. ; FORTUNATO, V. ; FRATTER, I. ; GALVIN, A. B. ; HANCOCK, B. ; HEIRTZLER, D. ; KATARIA, D. ; KISTLER, L. ; LEPRI, S. T. ; LEWIS, G. ; LOEFFLER, C. ; MARTY, W. ; MATHON, R. ; MAYALL, A. ; MELE, G. ; OGASAWARA, K. ; ORLANDI, M. ; PACROS, A. ; PENOU, E. ; PERSYN, S. ; PETIOT, M. ; PHILLIPS, M. ; PŘECH, L. ; RAINES, J. M. ; REDEN, M. ; ROUILLARD, A. P. ; ROUSSEAU, A. ; RUBIELLA, J. ; SERAN, H. ; SPENCER, A. ; THOMAS, J. W. ; TREVINO, J. ; VERSCHAREN, D. ; WURZ, P. ; ALAPIDE, A. ; AMORUSO, L. ; ANDRÉ, N. ; ANEKALLU, C. ; ARCIULI, V. ; ARNETT, K. L. ; ASCOLESE, R. ; BANCROFT, C. ; BLAND, P. ; BRYSCH, M. ; CALVANESE, R. ; CASTRONUOVO, M. ; ERMÁK, I. ; CHORNAY, D. ; CLEMENS, S. ; COKER, J. ; COLLINSON, G. ; D'AMICIS, R. ; DANDOURAS, I. ; DARNLEY, R. ; DAVIES, D. ; DAVISON, G. ; SANTOS, A. De L. ; DEVOTO, P. ; DIRKS, G. ; EDLUND, E. ; FAZAKERLEY, A. ; FERRIS, M. ; FROST, C. ; FRUIT, G. ; GARAT, C. ; GÉNOT, V. ; GIBSON, W. ; GILBERT, J. A. ; GIOSA, V. D. ; GRADONE, S. ; HAILEY, M. ; HORBURY, T. S. ; HUNT, T. ; JACQUEY, C. ; JOHNSON, M. ; LAVRAUD, B. ; LAWRENSON, A. ; LEBLANC, F. ; LOCKHART, W. ; MAKSIMOVIC, M. ; MALPUS, A. ; MARCUCCI, F. ; MAZELLE, C. ; MONTI, F. ; MYERS, S. ; NGUYEN, T. ; RODRIGUEZ-PACHECO, J. ; PHILLIPS, I. ; POPECKI, M. ; REES, K. ; ROGACKI, S. A. ; RUANE, K. ; RUST, D. ; SALATTI, M. ; SAUVAUD, J. A. ; STAKHIV, M. O. ; STANGE, J. ; STUBBS, T. ; TAYLOR, T. ; TECHER, J. D. ; TERRIER, G. ; THIBODEAUX, R. ; URDIALES, C. ; VARSANI, A. ; WALSH, A. P. ; WATSON, G. ; WHEELER, P. ; WILLIS, G. ; WIMMER-SCHWEINGRUBER, R. F. ; WINTER, B. ; YARDLEY, J. ;

ZOUGANELIS, I.: The Solar Orbiter Solar Wind Analyser (SWA) suite. In: *Astronomy and Astrophysics* 642 (2020), 10. – ISSN 14320746

[Parks 2003]   PARKS, George K.: *Physics Of Space Plasmas*. CRC Press, 2019, 2003

[Priest 1982]   PRIEST, Eric R.: *Solar Magnetohydrodynamics*. Springer Dordrecht, 1982

[Reiss u. a. 2021]   REISS, M. A. ; MÖSTL, C. ; BAILEY, R. L. ; RÜDISSER, H. T. ; AMERSTORFER, U. V. ; AMERSTORFER, T. ; WEISS, A. J. ; HINTERREITER, J. ; WINDISCH, A.: Machine Learning for Predicting the Bz Magnetic Field Component From Upstream in Situ Observations of Solar Coronal Mass Ejections. In: *Space Weather* 19 (2021), Nr. 12, S. e2021SW002859. – URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021SW002859. – e2021SW002859 2021SW002859

[Roberts 2007]   ROBERTS, Paul H.: Alfvén's Theorem and the Frozen Flux Approximation. In: *Encyclopedia of Geomagnetism and Paleomagnetism* (2007), 7, S. 7–11. – URL https://link.springer.com/referenceworkentry/10.1007/978-1-4020-4423-6_5

[Ryutov 1999]   RYUTOV, D. D.: Landau damping: half a century with the great discovery. In: *Plasma Physics and Controlled Fusion* 41 (1999), 3, S. A1. – URL https://iopscience.iop.org/article/10.1088/0741-3335/41/3A/001https://iopscience.iop.org/article/10.1088/0741-3335/41/3A/001/meta. – ISSN 0741-3335

[Sakurai und Sato 2017]   SAKURAI, Takashi ; SATO, Katsuhiko: Heating mechanisms of the solar corona. In: *Proceedings of the Japan Academy, Series B* 93 (2017), 2, S. 87–97. – ISSN 0386-2208

[Sculley 2010]   SCULLEY, David: Web-scale k-means clustering. In: *Proceedings of the 19th international conference on World wide web*, 2010, S. 1177–1178

[Stocker 2022]   STOCKER, Adam: Interim Report. (2022)

[Verscharen u. a. 2016]   VERSCHAREN, Daniel ; CHANDRAN, Benjamin D. G. ; KLEIN, Kristopher G. ; QUATAERT, Eliot: COLLISIONLESS ISOTROPIZATION OF THE SOLAR-WIND PROTONS BY COMPRESSIVE FLUCTUATIONS AND PLASMA INSTABILITIES. In: *The Astrophysical Journal* 831 (2016), 11, S. 128. – URL https://iopscience.iop.org/article/10.3847/0004-637X/831/2/128https://iopscience.iop.org/article/10.3847/0004-637X/831/2/128/meta. – ISSN 0004-637X

[Verscharen u. a. 2019]   VERSCHAREN, Daniel ; KLEIN, Kristopher G. ; MARUCA, Bennett A.: The multi-scale nature of the solar wind. In: *Living Reviews in Solar Physics 2019 16:1* 16 (2019), 12, S. 1–136. – URL https://link.springer.com/article/10.1007/s41116-019-0021-0. – ISBN 4111601900210

[Vlasov 1968]   VLASOV, A A.: THE VIBRATIONAL PROPERTIES OF AN ELECTRON GAS. In: *Soviet Physics Uspekhi* 10 (1968), 6, S. 721. – URL https://iopscience.iop.org/article/10.1070/PU1968v010n06ABEH003709https://iopscience.iop.org/article/10.1070/PU1968v010n06ABEH003709/meta. – ISSN 0038-5670