

**Imperial College
London**

Multi-Spacecraft Analysis: Forecasting Geomagnetic Indices with Dynamic Solar Wind Parameters

IMPERIAL COLLEGE LONDON

DEPARTMENT OF PHYSICS

Logan Pyle



Supervised by

Tim Horbury, Science Lead of NASA's IMAP Magnetometer

Assessed by

Lorenzo Matteini, Senior Researcher for Solar Orbiter (ESA) and
Parker Solar Probe (NASA)

Layperson's Summary: Multi-Spacecraft Space Weather Forecasting

What if the internet was wiped out completely? Yes, that's right. The internet. Gone. This isn't a plot from a dystopian novel—it's a very real possibility, and it comes from the threat of severe space weather. We have already seen consequences of not being prepared, with satellites being destroyed, power grids failing, and communications systems going offline. Space weather has potentially drastic consequences, and it all comes from an energetic stream of particles emanating from the Sun: the solar wind.

'Space weather' refers to the solar wind and its influence on the Earth. The solar wind carries with it not just particles, but also a magnetic field: the interplanetary magnetic field (IMF). As it sweeps through space, it interacts with the Earth's magnetic field, the magnetosphere. A powerful demonstration of this interaction is the aurora borealis, commonly known as the 'northern lights'. This captivating light display results from solar wind ions transferring energy to particles trapped in the magnetosphere. Unfortunately, what causes these beautiful scenes can also pose significant risks. These events can induce currents strong enough to overload electrical grids and damage critical infrastructure. By understanding and monitoring the solar wind, scientists aim to predict these events, providing crucial warnings that help protect our modern world from unexpected and severe space weather impacts.

Predicting space weather relies on measurements of the IMF. This is measured by spacecraft upstream of the Earth, stationed at the Earth-Sun L1 Lagrange point, a gravitationally stable location where they orbit the Sun synchronously with the Earth. This allows continuous measurement of the Earth-directed solar wind. Scientists have devised techniques to predict the IMF at near-Earth locations, using a single spacecraft. But recently, additional spacecraft have been added to L1, and in 2025 two more spacecraft will be launched. By harnessing data from multiple spacecraft, scientists aim to improve forecasting methods, providing a more detailed picture of the approaching solar wind and increasing the accuracy of predictions.

The main setback in these methods comes from testing the predictions. Due to the Earth's own magnetic field and its complex interaction with the solar wind, we cannot simply measure the IMF on the Earth and compare it to our predictions. Instead, previous methods have used spacecraft in lunar orbit. Unfortunately, this has major limitations as the solar wind's path to the spacecraft is often blocked by the Moon itself. But what if there were a way to quantify the interaction between the Earth's magnetic field and the solar wind? Then, we would be able to test our predictions at all times of the year!

Fortunately, we can do exactly that. There are a number of ground-based observatories that measure the disturbance in Earth's magnetic field and quantify it with something called a geomagnetic index. Various models have been developed that predict these indices, using the IMF, density and velocity of the solar wind. In this work, we utilise one such model and apply it to both a single spacecraft and a multi-spacecraft method used previously to predict the IMF. We find that a multi-spacecraft approach produces the most accurate predictions of geomagnetic indices. As a result, we hope that future space weather forecasting methods can be tested with a model such as ours. This would remove any constraints on testing, while also providing more insights into space weather impacts on the Earth.

Abstract

Predicting the SYM-H geomagnetic index is more beneficial in space weather forecasting than predicting the magnetic field alone, as it quantifies disturbances within the Earth's magnetosphere. In this study, we propose and evaluate a two-stage model for predicting SYM-H. Utilising measurements of solar wind density, magnetic field z-component, and velocity from multiple spacecraft at the L1 Lagrange point, we adopt a multi-spacecraft approach and compare it with the predictions made by three individual spacecrafsts. The multi-spacecraft forecasts produce a root-mean-square error of $5.16 \pm 0.08\text{nT}$ with respect to the observed SYM-H distributions, compared to a mean root-mean-square error of $5.34 \pm 0.08\text{nT}$ in the single spacecraft forecasts. However, we find that the multi-spacecraft approach successfully predicts significant events $\text{SYM-H} < -15\text{nT}$ $70.4 \pm 0.7\%$ of the time, falling short of the $73.1 \pm 0.5\%$ success rate made by the single spacecraft predictions.

Contents

1	Introduction	4
1.1	Space Weather	4
1.2	Current IMF Forecasting Methods	5
1.3	Predicting at the Earth: Geomagnetic Indices	6
1.4	Data Usage and Limitations	7
2	A Unified Geomagnetic Index Prediction Model	8
2.1	Solar Wind-Magnetosphere Coupling	8
2.2	Empirical SYM-H Prediction Model	9
2.3	Solar Wind Dynamics and Assumptions	10
2.4	Solar Wind Parameter Propagation	12
2.5	Implementing a Two-Stage Forecasting Model	14
3	Single vs Multi-Spacecraft Evaluation	18
3.1	Timescales and Event Considerations	18
3.2	Cross-Correlation as a Performance Metric	20
3.3	Fit Performance Metrics	22
3.4	Event Detection Metrics	27
4	Improving the Model	29
4.1	Predicting Small Disturbances	30
4.2	Improving the Empirical SYM-H Model	31
4.3	Limitations in Multi-Spacecraft Predictions	31
4.4	Future Missions	33
5	Conclusions	35

1 Introduction

1.1 Space Weather

Since the first recorded space weather event in 1903 [1], global organisations have sought to predict these occurrences, recognising their dangers. Space weather encompasses heliospheric conditions in our solar system, and in particular, the potential implications of these conditions on the Earth. It is driven by the solar wind, a continuous outflow of charged particles in a plasma state, originating from the Sun's corona.

The solar wind carries with it the interplanetary magnetic field (IMF). When it reaches the Earth, it interacts with the Earth's magnetic field, the magnetosphere. This results in a transfer of energy from the solar wind into the magnetosphere, which can significantly alter conditions on the Earth. In fact, this interaction often manifests itself as the aurora borealis, or 'northern lights', when atmospheric particles trapped in the magnetosphere become excited by collisions with solar wind ions.

As well as captivating natural displays, the solar wind can also pose significant risks. When the IMF interacts with the magnetosphere, it causes a fluctuation in the Earth's magnetic field. As a result, electrical currents may be induced in segments of critical infrastructure, such as gas pipelines and transmission lines. This could necessitate an emergency shutdown of these components, which play pivotal roles in the world's economy [2]. Moreover, satellites can suffer from the increase in charged particles present in their environment, potentially disrupting navigation and communication services [3].

The potential economic impacts of such events are immense. For instance, in 1989, the largest geomagnetic storm in the modern technological era caused a 9-hour blackout in the city of Quebec, shutting down its power grids. If such an event were to occur again, the estimated economic impact would range from \$2.4 trillion to \$3.4 trillion over the course of a year [4].

The ability to forecast events such as these is made capable by a number of spacecraft that have been launched in recent decades. Currently, three spacecraft at the Earth-Sun L1 Lagrange point act as upstream solar wind monitors, providing real-time data on different parameters of the solar wind, such as its velocity and magnetic field strength. These spacecraft, ACE, DSCOVR and Wind, are the three used in this research. The launch of two additional spacecraft, IMAP and SWFO-L1, in 2025 will aid further research, improving our ability to forecast space weather using a multi-spacecraft approach.

Fig. 1 illustrates the process of solar wind propagation from the Sun to the Earth, followed by a process known as magnetic reconnection. This is the rearrangement of magnetic field lines at the boundary between two electrically conducting plasmas: the solar wind and the magnetosphere plasma. It is the dominant process that causes the energy injection from the solar wind into the magnetosphere, and is intensified when oppositely polarised field lines interact. The IMF component of interest is therefore southward, as this is anti-aligned with the direction of Earth's magnetic field lines.

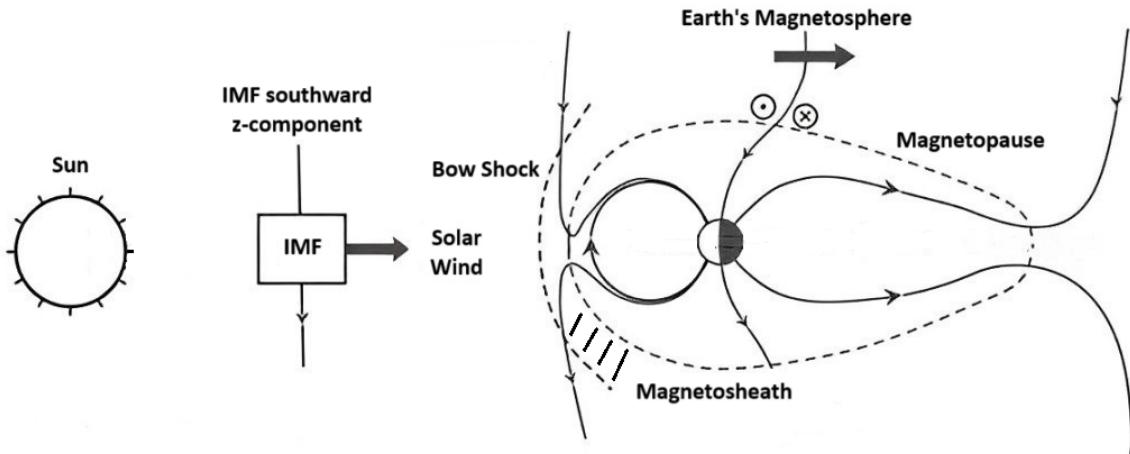


Figure 1: A schematic adapted from [5], illustrating the motion of a solar wind fluid parcel travelling towards the Earth, followed by magnetic reconnection at the magnetopause. The southward component of the IMF is oppositely polarised with respect to the magnetosphere, which intensifies magnetic reconnection. The bow shock is the shock wave created by the solar wind’s interaction with the magnetosphere; the Bow Shock Nose (BSN) is the point on the bow shock closest to the Sun; the magnetopause is the boundary where the pressure from the solar wind and the magnetosphere are equal; and the magnetosheath describes the region between the bow shock and magnetopause.

1.2 Current IMF Forecasting Methods

Magnetic reconnection demonstrates the significance of the IMF in affecting space weather conditions on the Earth. Given reconnection favours the southward component of the IMF, previous methods have concentrated on predicting this. In this work, we use the GSE coordinate system, as illustrated in Fig. 2. The southward IMF component is therefore the negative z -component in our coordinate system.

Forecasting the IMF downstream of L1 requires predicting the arrival time of a phase front of the solar wind. These are illustrated by the red lines on Fig 2. A phase front is a planar surface assumed to have identical IMF values across it, up to a certain length, called the coherence length. The orientation of this phase front has been investigated in various studies; however, a statistical analysis in [6] demonstrated that there is no significant difference between assuming an angled phase front, compared with a flat one.

In this work, we therefore assume a flat phase front. This aligns with our objective, which is to extend the multi-spacecraft study in [7], which also employed a flat phase front and demonstrated its effectiveness in yielding accurate results.

For a particular spacecraft, estimating the propagation time for a phase front to travel from L1 to a chosen location is a simple process. The x -axis separation between the spacecraft and a desired prediction plane, Q , is divided by the spacecraft’s measured x -component of velocity, yielding a propagation time, Δt . The IMF distribution

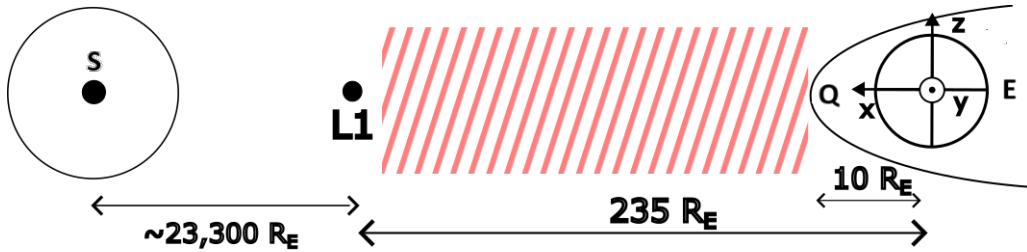


Figure 2: A schematic of the key locations involved in this work. The distance between L1 and the Earth is roughly 0.01AU. The red lines can be considered as phase fronts of the solar wind, which we assume to be flat. In this work, Geocentric Solar Ecliptic (GSE) coordinates are used with the origin located at the Earth’s centre, the positive x -axis extending towards the Sun in the ecliptic plane, positive y -axis opposing the direction of Earth’s motion in orbit around the Sun, and the z -axis completing the orthogonal set in a right-handed sense. The propagation region for our work is between L1 and the edge of the magnetosphere, shown at a GSE x -coordinate of $10R_E$, consistent with reported values [8].

measured at L1 is then tested against the corresponding measurements at Q , at a time Δt after the L1 measurements were made. In [7], ARTEMIS, a spacecraft in a lunar orbit, was used. In our work, Q will correspond to the edge of the magnetosphere, as shown in Fig. 2.

Overall, the space weather community is actively pursuing methods that utilise multiple spacecraft for predictions, aiming to capture the broader spatial content of the solar wind. Currently, different spacecraft can produce contrasting predictions, highlighting the issues with relying on a single data source [9]. We are interested in [7] because of their work in developing a multi-spacecraft method that outperformed the single spacecraft predictions. Their study used a weighted average function based on where each spacecraft predicted the solar wind to propagate to in the y - z GSE plane, assigning greater weight to predictions closer to the Earth-Sun line. This prioritises measurements of the solar wind that are more likely to influence the magnetosphere.

In this work, we use the weighted average method as the designated multi-spacecraft technique. Since it was shown to outperform the single spacecraft methods for predicting the IMF, we have extended it to show how future multi-spacecraft methods can be tested using an Earth-based model, which can be assessed with multiple performance metrics.

1.3 Predicting at the Earth: Geomagnetic Indices

Current multi-spacecraft approaches, including [7] and an earlier approach in [10], focus on predicting the IMF z -component. However, this has a number of limitations. To test their predictions they employ ARTEMIS, a spacecraft in lunar orbit. This puts constraints on the available times in which it is possible to test predictions, since the spacecraft is either blocked by the Moon, or within the Earth’s bow shock, both of which would interfere with measurements of the IMF.

In our study, we aimed to explore Earth-based predictions. We therefore required a method of quantifying the complex interaction between the solar wind and the magnetosphere. Subsequently, geomagnetic indices were chosen for this purpose. These are quantities derived from ground-based observations of the disturbance in the magnetosphere.

While numerous geomagnetic indices exist, our objective was to obtain a comprehensive measure of solar wind-induced disturbances. The Dst (Disturbance Storm Time) index is commonly used, quantifying the variation in the north/south component of the Earth's magnetic field close to the equator. However, Dst is only calculated at a 1-hour cadence, which is not sufficient for our space weather prediction requirements. Fortunately, an additional geomagnetic index exists, the SYM-H (Symmetric Horizontal Component) index, which is calculated at a 1-minute cadence and measures disturbances in Earth's magnetic field, similar to Dst. In fact, it has been shown that SYM-H can be used as a high-resolution version of Dst [11]. The 'Horizontal' component is a projection of the Earth's magnetic field onto the plane of Earth's surface at a particular observatory. SYM-H was therefore selected as it enables us to evaluate predictions at a 1-minute frequency.

1.4 Data Usage and Limitations

A number of sources were used to access the solar wind data measured by ACE, DSCOVR and Wind. The Python module pySPEDAS [12] was utilised to obtain the ACE data, while the DSCOVR [13] and Wind [14] data were acquired from sources within NASA's OMNIWeb Service. Geomagnetic index data was obtained directly from OMNIWeb [15]. All OMNIWeb-acquired data was of a 1-minute resolution, while pySPEDAS data was downloaded at a 16-second resolution, before being linearly interpolated to 1-minute intervals. The data acquired was as follows: spacecraft positions, velocity components, density and magnetic field z -component.

The predictions we could test were constrained by the availability of spacecraft data. The main limitation in raw data came from DSCOVR, which only has a short period of accessible data, from 2016 to 2019. Data from 2016 was further unavailable due to extensive gaps in ACE's density record. To address any remaining missing values within the dataset, a linear interpolation technique was applied. This method assumed a linear relationship between consecutive known data points and filled in the gaps accordingly.

Linear interpolation is applied up to a maximum length, determined by the method outlined in [16]. This method uses the correlation function,

$$C(j) = \frac{\sum_s \sum_{i=1}^{n-j} d(i) \cdot d(i+j)}{\sum_s \sum_{i=1}^{n-j} d(i) \cdot d(i)}, \quad (1)$$

where $d(i)$ is the difference between the parameter's value and the 20-day average of the parameter, at a data point i . The correlation function determines a correlation C between consecutive data values from point i to point $i + j$.

The maximum length allowed for interpolation, j_{int} , is determined by calculating the maximum value of j before $C(j)$ drops below a threshold of 0.8. This threshold, as recommended in [16], ensures that the interpolated values maintain a sufficiently strong correlation with the existing data points.

In our datasets, j_{int} is calculated separately for velocity, density, and IMF for each spacecraft. The minimum value obtained for j_{int} is 9, therefore interpolation is performed up to a maximum of 9 minutes within a data gap. All other data is discarded.

2 A Unified Geomagnetic Index Prediction Model

2.1 Solar Wind-Magnetosphere Coupling

The coupling of the solar wind to the magnetosphere is a complex interaction. When a phase front of the solar wind encounters the magnetosphere, it can induce disturbances within it. These disturbances are known as geomagnetic storms. In this work, we use a simplified model that characterises three different phases of geomagnetic storms, which are illustrated in Fig. 3.

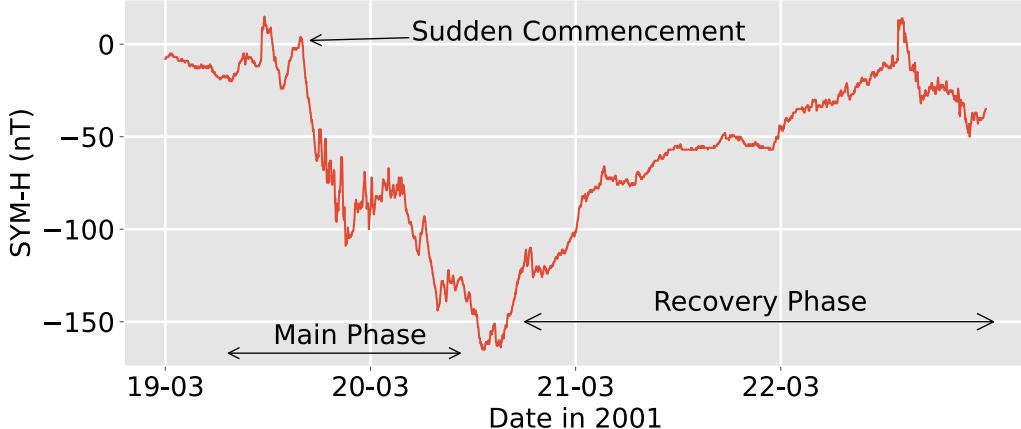


Figure 3: An example of a geomagnetic storm from 2001, measured with the SYM-H index. It has three distinct phases: a sudden commencement, associated with enhancements in solar wind dynamic pressure; a main phase, carried by the magnetic field z -component exceeding a threshold value; and a recovery phase, the gradual restoration of geomagnetic conditions to their pre-storm state. A more negative value of SYM-H corresponds to a greater disturbance in the magnetosphere.

To develop our model into one for predicting geomagnetic indices, we need to forecast not just the IMF z -component, but two other parameters of the solar wind: density n and velocity \vec{v} . Along with the IMF z -component, B_z , these define two important quantities involved in the solar wind-magnetosphere coupling. Eq. (2a) quantifies the solar wind dynamic pressure, which controls compressions of the magnetosphere, known as magnetopause currents [17]. Equation (2b) represents the in-

terplanetary electric field, responsible for driving the rate of energy injection into the magnetosphere [18].

$$P = nv^2 \quad (2a)$$

$$E = -vB_z \quad (2b)$$

The energy injection by the electric field manifests primarily through the ring current, a region of enhanced particle density encircling the Earth within the magnetosphere. The magnetic field induced by the ring current reduces the intensity of the geomagnetic field measured at ground-based observatories, and it is this disturbance that SYM-H measures.

Geomagnetic storms have been classified in various ways. Originally, the classification relied on the Dst index, as described in [19]. However, recent studies, such as the one in [20], have provided an identical characterisation using the SYM-H index:

$$\text{Small Storm: } -30 \text{ nT} > \text{SYM-H} > -50 \text{ nT} \quad (3)$$

$$\text{Moderate Storm: } -50 \text{ nT} > \text{SYM-H} > -100 \text{ nT} \quad (4)$$

$$\text{Intense Storm: } \text{SYM-H} < -100 \text{ nT}. \quad (5)$$

We are interested in all geomagnetic activity levels, not just storms, therefore the classification in [21], which defines the main phase as when SYM-H reaches -15nT, is most useful to our work. This labels the transition from a quiet period to an active one, and is crucial for our model's performance evaluation.

2.2 Empirical SYM-H Prediction Model

Various models have attempted to forecast SYM-H using measurements from space-craft within the magnetosheath. Significantly, the vast majority of these models, such as the recent work in [22], utilise an empirical algorithm, commonly referred to as the ‘Burton’ equation, named after its author [23]. While initially developed for Dst, [22] and others have extended its application to SYM-H, a practice justified by [11].

The model separately characterises the three phases of a geomagnetic storm into distinct terms:

$$\frac{d}{dt} \text{SYM-H} = -a\{\text{SYM-H} - b[P(t)]^{1/2} + c\} + F[E(t)] + \frac{d}{dt}\{(bP(t)^{1/2})\} \quad (6)$$

The first term on the right-hand side quantifies the decay of the ring current, indicating the decrease in its intensity and the restoration of the magnetosphere to its undisturbed state.

The electric field-dependent term, $F(E)$, is given by

$$F(E) = \begin{cases} 0 & \text{if } E \leq 0.5 \text{ mVm}^{-1} \\ d(E - 0.5) & \text{if } E > 0.5 \text{ mVm}^{-1}, \end{cases} \quad (7)$$

demonstrating an ability to ‘switch on’ if the electric field reaches a threshold value of $0.5mVm^{-1}$. This characteristic was found in [24], who noted that the main phase of a geomagnetic storm was triggered only when the southward B_z exceeded a certain threshold value.

The final derivative term identifies the magnetopause currents; compressions of the magnetosphere driven by the solar wind dynamic pressure. This term controls transient variations in SYM-H.

To utilise Eq. (6), we established a finite difference scheme for the differential equation,

$$S_{i+1} = S_i + (-a[S_i - bP_i^{1/2} + c] + F[E_i] + \frac{b(P_{i+1}^{1/2} - P_i^{1/2})}{\Delta t})\Delta t, \quad (8)$$

where Δt is the cadence of SYM-H, 60 seconds, and S_i represents the SYM-H value at a particular minute. This demonstrates the dependence of the model on the previous value of SYM-H as well as the solar wind parameters.

The model requires that each measurement of E_i and P_i be made at the nose of the magnetosphere. Hence, this is the location that the measured solar wind parameters at L1 should be propagated to.

This model is by no means perfect, but a significant number of studies have used it as a foundational model for the interaction between the solar wind and magnetosphere. In this work, we use it in its initial form, which was created to capture the fundamental physics of this interaction: energy conservation. The model approximates the non-linear injection of energy into the magnetosphere with the linear injection term, $F(E)$. There is potential to adapt this part of our model in the future to characterise the non-linearities more effectively, but for now we are interested in initiating an Earth-based approach for testing multi-spacecraft predictions, and the Burton equation produces sufficiently good results for our purposes.

2.3 Solar Wind Dynamics and Assumptions

In developing our model, we aimed to establish a foundational framework for using SYM-H to test multi-spacecraft predictions. However, it is important to acknowledge that due to time constraints on the project, the model is simplified with various assumptions. We acknowledge these assumptions here in the hope that future work may develop the model to reduce the uncertainties involved.

Firstly, an approximation that is used in IMF forecasting models such as [7] is known as the ‘frozen-in’ approximation of the solar wind. In order to predict the IMF configuration downstream of L1, we assume that the IMF is carried along, unchanged, with the solar wind plasma. This assumption is made using ideal magnetohydrodynamics (MHD), the study of electrically conducting plasmas, to characterise the dynamics we are interested in.

The magnetic induction equation,

$$\frac{\partial \vec{B}}{\partial t} = \nabla \times (\vec{u} \times \vec{B}) + \eta \nabla^2 \vec{B} \quad (9)$$

describes how the solar wind plasma evolves in time. \vec{B} quantifies the IMF vector, \vec{u} the bulk flow velocity of the plasma, while η is the magnetic diffusivity, characterising the viscosity of the magnetic field. This quantity η therefore controls the contribution from the second term in Eq. 9, the diffusion term. The first term accounts for the effects from magnetic induction, therefore the ratio between these two terms specifies the relative contributions from induction versus diffusion.

By assuming that both terms share a characteristic scale length, a common assumption in MHD, the differential operators ∇ and ∇^2 give rise to a $1/L$ factor in the first term and a $1/L^2$ factor in the second. The ratio is therefore approximately given by the magnetic Reynold's number,

$$R_M = \frac{uL}{\eta}, \quad (10)$$

where L is this common scale length and u is the scale velocity.

In ideal MHD, it is assumed that $R_M \gg 1$, such that the diffusion term vanishes, leaving

$$\frac{\partial \vec{B}}{\partial t} = \nabla \times (\vec{u} \times \vec{B}). \quad (11)$$

In this approximation, the magnetic field is ‘frozen-in’ to the plasma, such that magnetic flux is carried along, unchanged in a discrete solar wind fluid parcel of velocity \vec{u} . This approximation has been essential in forecasts of the solar wind, as it means the configuration of the IMF remains constant; the downstream distribution is simply a time-shifted version of the configuration measured at L1.

Although this is a common assumption in space weather forecasting, the magnetic field does actually move relative to the solar wind plasma. This results in an error of around 10% in our propagation time predictions, which is generally accepted in forecasting models [25].

Since our model includes not just the IMF, but solar wind density and velocity as well, we must apply similar approximations to both of these quantities. For the density, we assume negligible expansion of the solar wind between L1 and the Earth. This is justified by various studies, such as [26], which demonstrates that the short distance between L1 and the Earth, relative to the Earth-Sun distance, means the expansion is small enough for density to be considered constant. Similarly, in many studies the velocity of a particular fluid parcel is assumed constant; this is recognised as an effective method in [27] and used for the multi-spacecraft approach in [7].

The empirical model we use requires propagating solar wind parameters to the magnetosphere nose, therefore the solar wind will have travelled past the BSN, and into the magnetosheath. A constant dynamic pressure and electric field within the magnetosheath are both assumed. This simplification is used with the same uncertainties that were present in the initial model [23], which did not consider disturbances outside of the magnetosphere. We justify this as our goal is to initiate a geomagnetic index-based testing model for multi-spacecraft predictions, and we decided early on that the model was successful enough to be used for this purpose. There is evidence that these assumptions can be improved to reduce the uncertainties present in our model, and this is discussed in Section 4.2.

2.4 Solar Wind Parameter Propagation

In section 1.2, the method for forecasting B_z using a single spacecraft was established. We now have a model in which the desired prediction location Q is known: the edge of the magnetosphere, which we defined in section 1.1 as $x = 10R_E$ in GSE coordinates.

Fig. 4 illustrates how the single spacecraft predictions can be developed into a multi-spacecraft approach. A single spacecraft method focuses solely on the x -component of velocity. However, our interest lies in utilising all three spacecraft to capture the broader spatial content of the solar wind, therefore we also use the total measured velocity, \vec{v} .

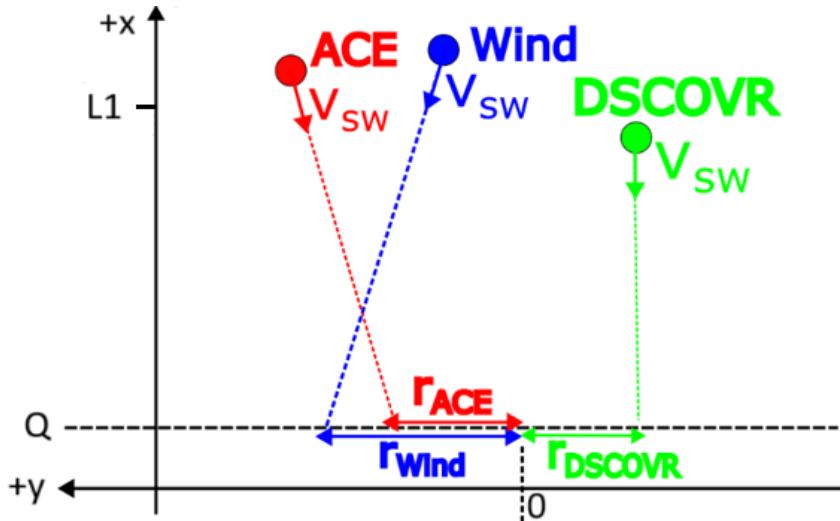


Figure 4: An illustration of the procedure for multi-spacecraft predictions in [7]. The solar wind parameters are shifted downstream to a desired prediction location Q using \vec{v} and Δt . Each propagated parameter has an associated offset r calculated, which are applied to Eq. (12) and Eq. (13), producing a multi-spacecraft prediction.

To do this, [7] calculated an additional parameter, the offset magnitude in the y - z plane, r , at the prediction location Q . This was calculated by shifting the measurements from all three spacecraft to Q , using \vec{v} and the calculated propagation time, Δt (as per the single spacecraft method). The result is a time series of both r and the parameter values, one for each spacecraft. These values are interpolated to a common time series so that quantities from different spacecraft predicted to arrive at similar times can be combined. The values of r_i , where i labels the spacecraft, are then applied to a weighted average method, where spacecraft with larger offsets contribute a smaller weight. The weighted average function is given by

$$w_i = e^{-r_i/r_0}, \quad (12)$$

where r_0 is the coherence length of the parameter you are forecasting, a number that quantifies the spatial variation of solar wind parameters in the transverse y - z plane; we use the values calculated in [28] for our model. The result is a single parameter distribution accounting for all three spacecraft's measurements,

$$X = \frac{\sum_{i=1}^3 w_i X_i}{\sum_{i=1}^3 w_i}, \quad (13)$$

where X_i is the propagated parameter value, with $i \in (\text{ACE}, \text{DSCOVR}, \text{Wind})$.

The coherence lengths r_0 are important quantities in multi-spacecraft predictions, restricting the transverse distance over which accurate magnetosphere predictions can be made. Fig. 5 visualises the coherence lengths with respect to the spacecraft's orbits in the transverse plane. This is necessary to justify the use of all three spacecraft, while emphasising the limitations of a larger orbit such as Wind's, which occasionally drifts outside of all three coherence lengths.

An important aspect to consider when making predictions of solar wind parameters at the magnetosphere is Earth's orbital motion. Considering Fig. 4, the line of $r = 0$ from which the offsets are calculated is constantly in motion. The GSE coordinate system moves around the Sun along with the Earth, therefore a prediction at a time Δt after a measurement needs to add an aberration term $u\Delta t$, where u is the orbital speed of the Earth, 30km s^{-1} . Since the Earth is moving in the negative y direction in GSE coordinates, our predictions are shifted in the positive y -axis by $u\Delta t$.

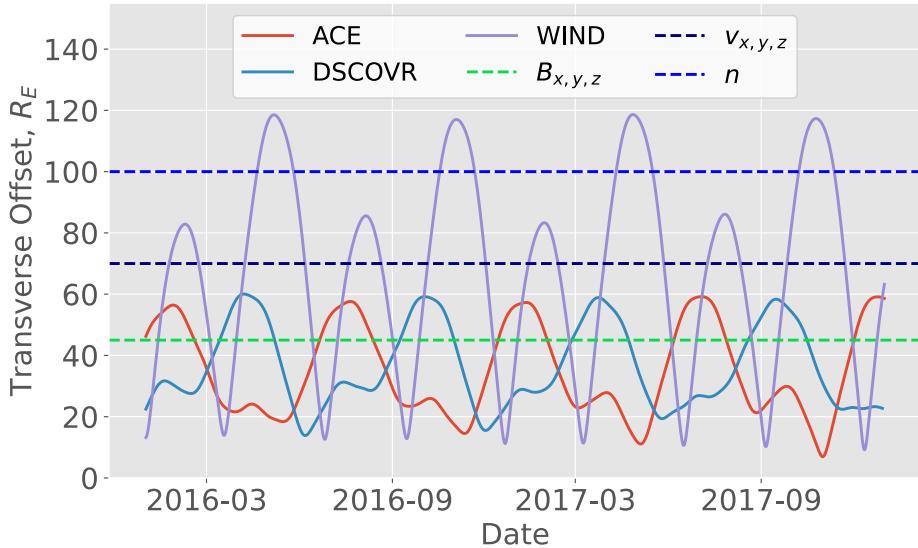


Figure 5: Coherence lengths of key parameters [28], along with the spacecrafts' y - z positions, accounting for Earth's orbital motion. Our goal is to monitor variations across the transverse scale of the magnetosphere; approximately $20R_E$ [29]. While optimal spacecraft positioning targets this region, effective measurements are feasible beyond it, provided the spacecraft remains within the parameter's coherence length. ACE and DSCOVR are well-positioned for precise magnetosphere monitoring. Concerns arise with Wind's positioning, which may not always capture magnetosphere-directed conditions, particularly in B_z .

Before implementing the weighted average method into our two-stage model, it was important to get an idea of the typical r values we were predicting with our dataset,

and the weights that these values were producing. We illustrate each forecast of our dataset’s r values in Fig. 6, where the red contours represent the weights assigned to B_z predictions at specific offsets in the y - z plane. As the smallest coherence length among our parameters, variations in B_z are the most significant, hence why it is displayed in this figure, as it is the limiting parameter in the transverse scale of multi-spacecraft predictions.

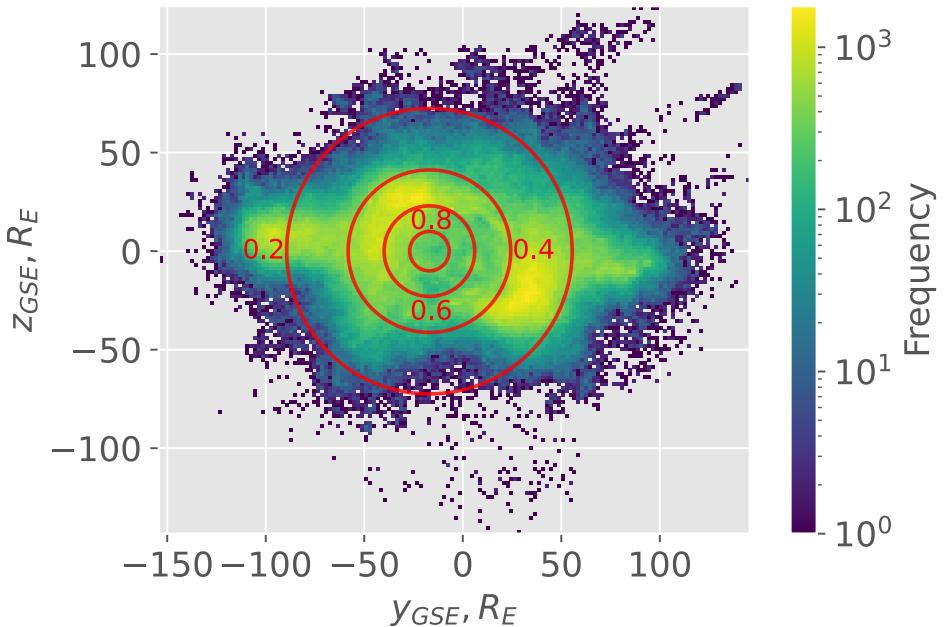


Figure 6: Colour density of predicted y - z offsets, r , from our dataset. The weights assigned at particular values for the offsets are shown by the red contours. The contour weights are centred off-zero as they are shifted to account for Earth’s orbital motion. This is chosen instead of shifting the r values to visually demonstrate the distance that Earth will have travelled in the time the solar wind has propagated from L1 to the magnetopause. In this diagram, we can imagine the Earth located at the center of the red contours, moving with respect to a set of predictions that have been made, altering the weights that will be assigned to those predictions.

2.5 Implementing a Two-Stage Forecasting Model

Now that we have established how solar wind parameters are propagated to the edge of the magnetosphere, and an empirical model that can be used to forecast SYM-H values, we can combine these into a two-stage forecasting model. Fig. 7 summarises the key steps in this model, with the final two stages constituting the core components of this ‘two-stage model’. To implement these stages seamlessly into a full-scale model, I developed a three-class object-oriented software package, consisting of various code modules, available on GitHub [30]. This includes three core classes: one giving users the flexibility to select between single or multi-spacecraft propagation; an algorithm

to generate a SYM-H distribution based on input parameters E and P ; and a final class responsible for implementing the full-scale model, allowing users to specify which methods to compare and which statistics to generate.

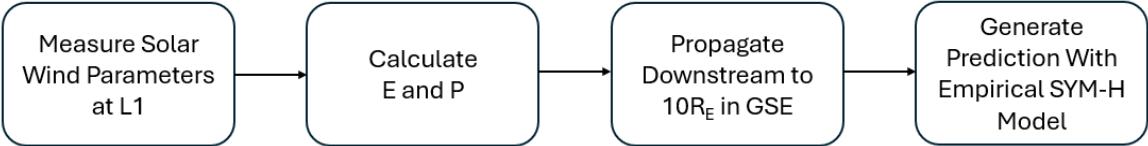


Figure 7: A summary flowchart of our SYM-H prediction model, combining solar wind parameter propagation with an empirical prediction algorithm at the edge of the magnetosphere.

To assess the reliability of our results, I developed methods for testing both the initial data and the quantities obtained at various stages of our model. The class structure of my software enabled outputs to be generated after each step. The initial B_z , n and \vec{v} data were assessed in the main class with the techniques described in section 1.4, before executing any phase of the model.

To assess the model, two forecast events are initially considered: a sudden commencement and main phase in Fig. 8, and a recovery phase in Fig. 9. The predicted distributions of E and P that were used as input parameters into both forecasts are shown alongside the SYM-H predictions, illustrating their effect on our ability to predict changes in SYM-H. Only the multi-spacecraft predictions are depicted here to provide a clear demonstration of the model’s predictions without cluttering the figures. They demonstrate some key characteristics of our model, before the different methods are compared in Section 3.

The two examples presented in Fig. 8 and Fig. 9 illustrate our model’s capabilities over two contrasting events within our dataset. They were chosen because they have an activity close to the mean value of SYM-H in our measured dataset, approximately -10nT, therefore represent the typical geomagnetic conditions that we observe.

Fig. 8 indicates our model’s ability to predict the onset of an event with a sudden drop in dynamic pressure and a sharp increase in the electric field. However, it also shows its limitations in predicting the evolution of that event. Conversely, Fig. 9 demonstrates that our model can reliably predict smaller-scale events caused by a spike in dynamic pressure. By comparing these contrasting periods, we gain some understanding about the inner workings of the model, and how it responds to varying solar wind conditions.

Overall, Fig. 8 and Fig. 9 demonstrate qualitatively that our model can make successful predictions of SYM-H using only calculations of E and P as input parameters. The next section will assess the model quantitatively, comparing predictions made by the single and multi-spacecraft propagation methods across a range of performance metrics.

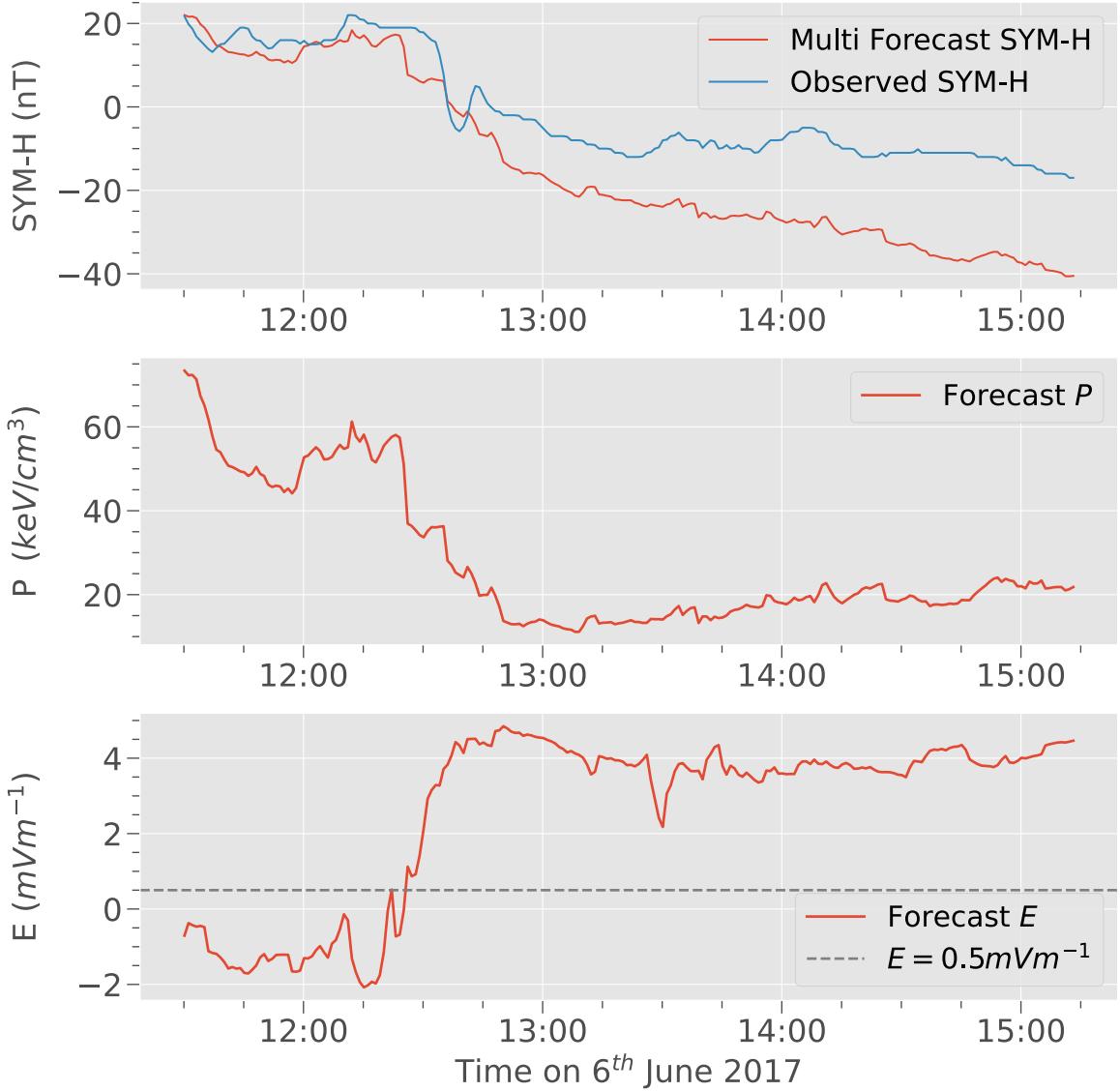


Figure 8: On top: example of a sudden commencement and main phase SYM-H prediction. The sudden commencement is predicted well, but we overpredict the intensity of the main phase. Input parameters for the predicted SYM-H are shown; middle: solar wind dynamic pressure, bottom: interplanetary electric field. The threshold value shown by the dashed line is given by Eq. (7). The sudden commencement is dominated by the dynamic pressure, while the main phase is sustained by the electric field, which remains above the threshold value.

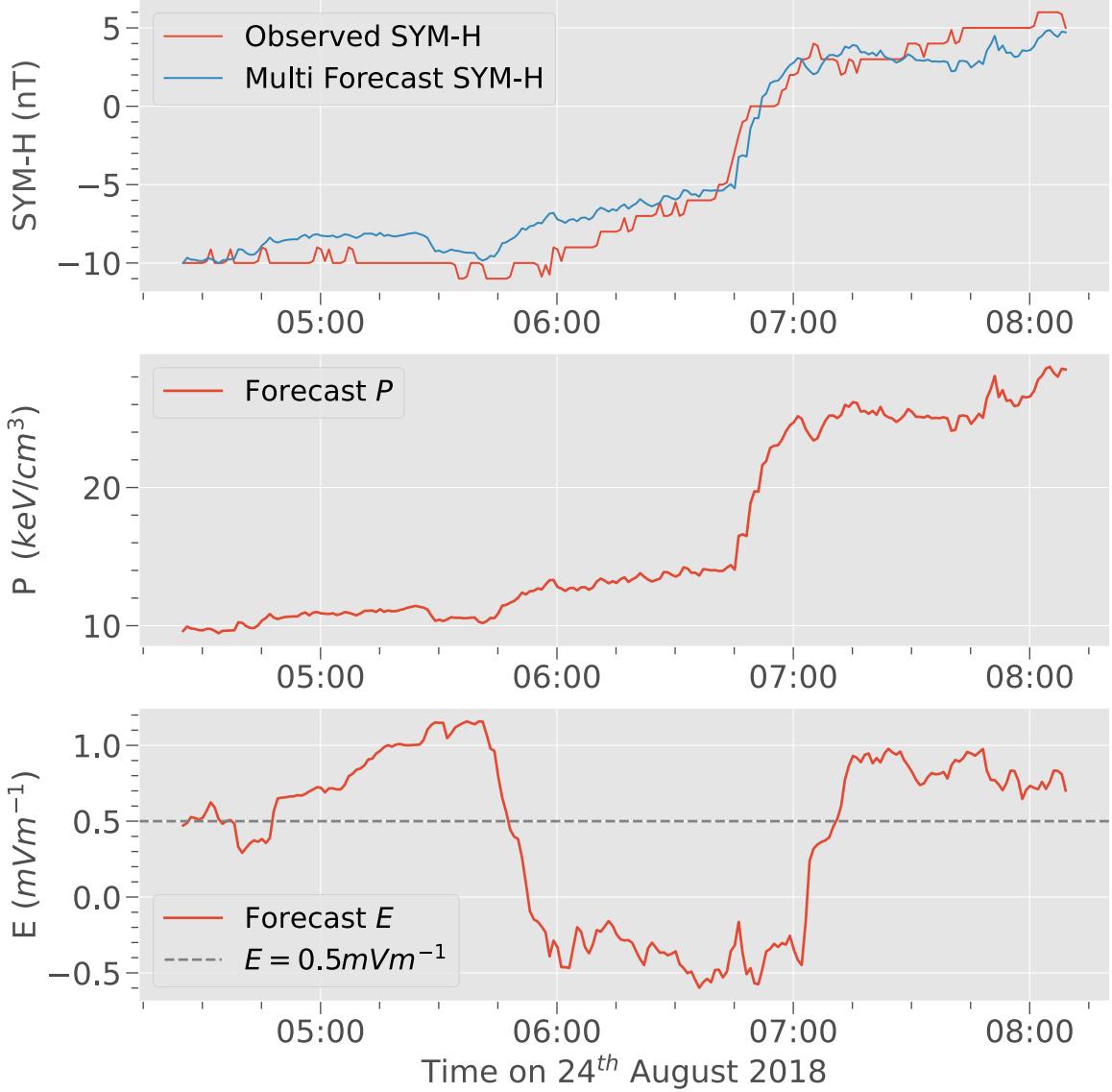


Figure 9: A good recovery phase prediction, occurring after a small disturbance to around -10nT. This plot is evidence that we do not just detect large geomagnetic storms in our model, we can also predict quiet events well. It also reveals the dominance of the solar wind dynamic pressure in recovery phases, as the sudden increase in SYM-H at around 06:45 is driven by a sharp increase in P . Despite the electric field reaching above the threshold value for around an hour at the start of this period, and an hour at the end, it does not contribute any significant decrease in SYM-H, demonstrating that a more powerful and consistent electric field needs to be maintained for events such as the one in Fig. 8 to occur.

3 Single vs Multi-Spacecraft Evaluation

3.1 Timescales and Event Considerations

Previous methods for evaluating IMF predictions have used the cross-correlation, c , between the predicted distribution $x(t)$ and an observed distribution $y(t)$, as a metric to assess accuracy. This has the form

$$c(\Delta t) = \frac{\sum_0^{N-1} (x'(t)y'(t + \Delta t))}{\sqrt{\sum_0^{N-1} (x'(t)^2) \sum_0^{N-1} (y'(t + \Delta t)^2)}}, \quad (14)$$

where N is the number of samples in the distributions, $x' = x - \bar{x}$, $y' = y - \bar{y}$, where \bar{x} and \bar{y} are the sample means, and the denominator is the normalisation factor. It quantifies the similarity between two distributions over a range of time delays between the two signals, with $\Delta t = 0$ a comparison at the exact time of prediction.

We use cross-correlation with $\Delta t = 0$ to initially compare results across different timescales and consider the strengths and weaknesses of these for multi-spacecraft predictions. As outlined in [31], choosing the correct timescale is critical when making predictions based on interplanetary conditions such as the solar wind. The typical timescales at which magnetospheric conditions show a strong correlation with interplanetary conditions are above 200 minutes. The dynamics involved in timescales less than 200 minutes have such a high level of complexity that a correlation between interplanetary conditions and geomagnetic processes cannot be established, since the dynamics are dominated by internal magnetospheric activity.

The chosen period could also not be too long that it would interfere with the assessment of multi-spacecraft predictions. For multi-spacecraft predictions to be effective, we need to detect finer details between individual spacecrafts, which become apparent at smaller timescales. This is demonstrated in Fig. 10(a), where the predictive performance of multi-spacecraft forecasts does not exceed the single-spacecraft forecasts over 24-hour periods. We see an averaging effect, where each spacecraft has picked up large variations in the solar wind and so predict a broader range of activities in SYM-H, resulting in diluted predictions that becomes less effective at longer timescales.

Conversely, shorter timescales enhance the sensitivity of predictions to variations in the solar wind, allowing multi-spacecraft methods to capture finer-scale structures which vary between different spacecraft. Hence we are more likely to see a higher proportion of accurate multi-spacecraft predictions for shorter timescales. However, with shorter timescales, more noisy predictions will be observed, since more inactive periods will be considered. Both of these effects are demonstrated by Fig. 10(b). With these limitations, we choose to evaluate our model across 2313 4-hour periods.

As well as timescales, it was important to consider the spectrum of SYM-H activities available in our dataset. Fig. 11 illustrates some key summary statistics of the observed SYM-H data, which serves as the benchmark against which our predictions were assessed. The 99th percentile value of -56.8nT demonstrates that high activity events are limited, although we have a large number of small to moderate events to consider, which are useful to see how well our model picks up smaller variations.

3 SINGLE VS MULTI-SPACECRAFT EVALUATION

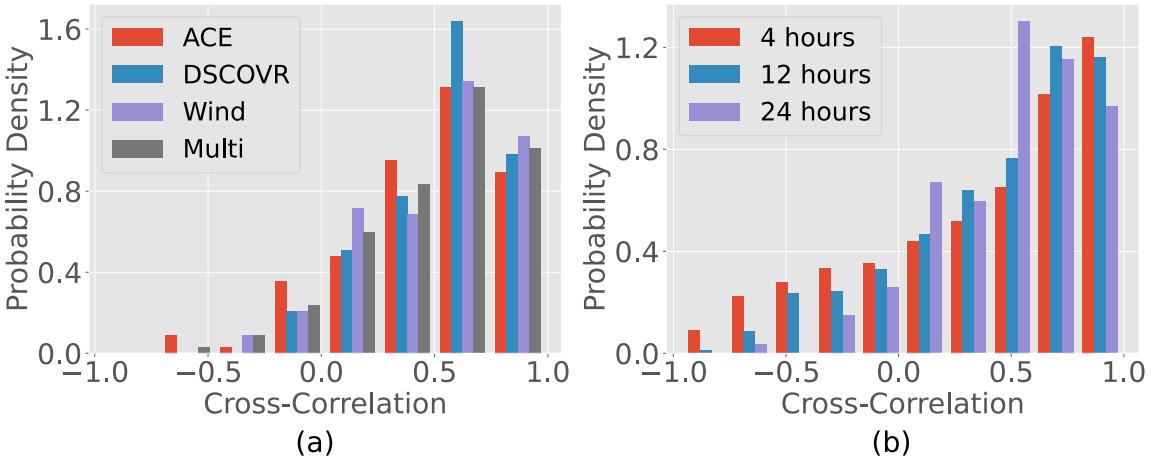


Figure 10: (a) Cross-correlation results over 24-hour periods, demonstrating the averaging effect which harms the multi-spacecraft predictions. (b) Timescale results comparison for multi-spacecraft predictions, demonstrating the effect of larger timescales on reducing the number of high-accuracy predictions. However, more noise is seen with 4 hours as there is a higher probability of observing an inactive period.

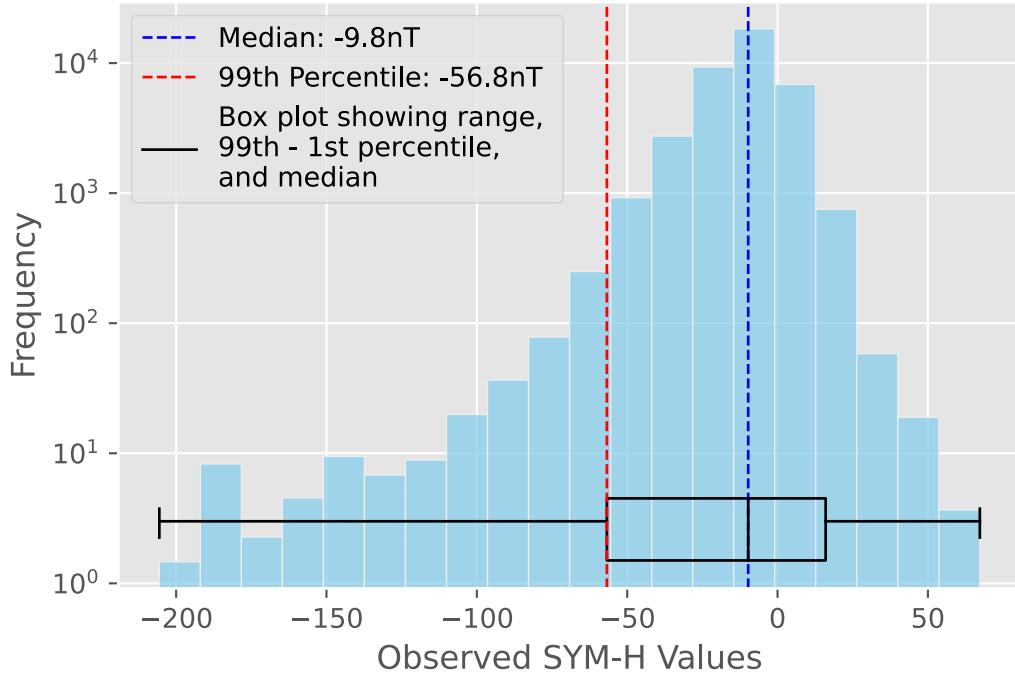


Figure 11: Histogram of the observed SYM-H dataset used in this analysis. The frequency axis is logarithmic due to the large disparity between the number of low activity versus high activity events. The box plot illustrates some key summary statistics, utilising 99th to 1st percentiles as the central box's range instead of the IQR, to convey the minimal number of high activity events we observe. The 99th percentile is shown with a red line to emphasise that 99% of our dataset has a value greater than -56.8nT.

3.2 Cross-Correlation as a Performance Metric

In this work, we utilise $\Delta t = 0$, since we are predicting a complex configuration that is non-linearly shaped by the magnetosphere, therefore the time lag is unpredictable and yields no meaningful results. Results of the predictions across the whole dataset are shown in Fig. 12. The results indicate that multi-spacecraft predictions produce the largest number of high-correlation predictions between 0.8 and 1.0, but do not produce the greatest accuracy on average.

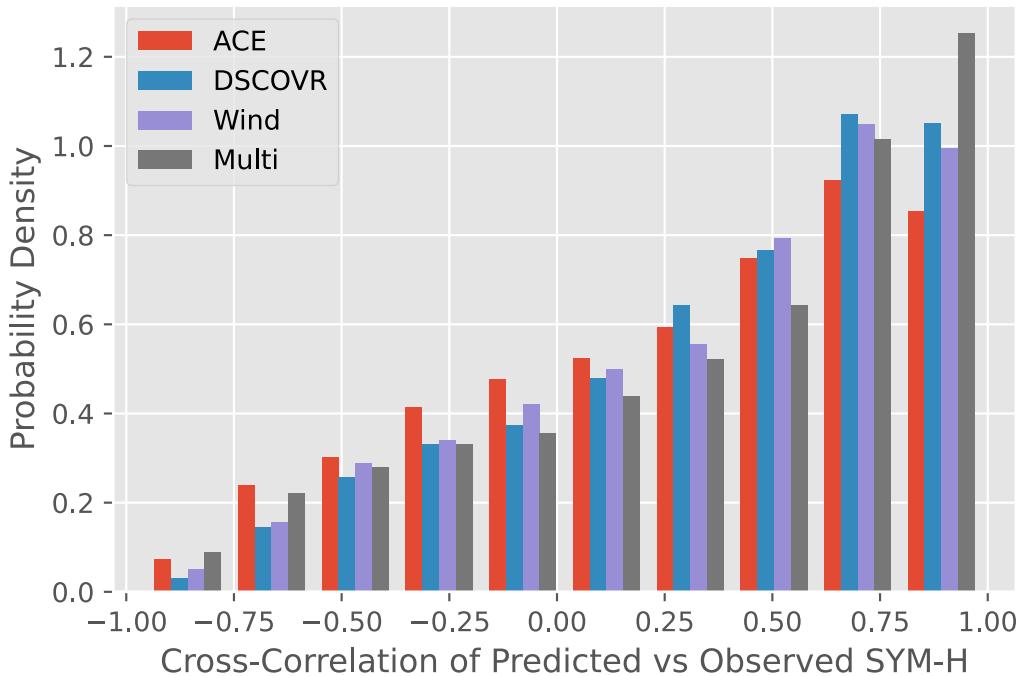


Figure 12: Cross-correlation with zero time lag. The mean correlations are as follows: 0.29 for ACE, 0.38 for DSCOVR, 0.37 for Wind, and 0.36 for Multi. The maximum mean comes from DSCOVR, however the multi-spacecraft method produces a noticeable peak in the highest correlation bin, indicating that it generates the majority of highly successful predictions.

We want to identify periods when each method makes the best predictions. Fig. 13 achieves this by applying four filtering processes, isolating the periods when a specific method produces the highest correlation value. These results demonstrate how often a particular method made the most accurate prediction, compared to how frequently it did not. There is a clear distinction between the success of multi-spacecraft compared with all three single spacecraft predictions. When the three single spacecraft predictions perform most accurately, you could choose any other method and it would give you a similar performance. Whereas if you choose a multi-spacecraft prediction that performs the best, you are almost guaranteed to have produced an excellent prediction with a high correlation value.

Significantly, when a multi-spacecraft prediction makes the most accurate forecast, it is rarely skewed by noisy predictions in the low and negative correlations. This result is emphasised in Table 1, which indicates that the mean correlations calculated in Fig. 12 are unreliable. The final column in the table shows the results for the highest two bins, with correlation values between 0.6 and 1.0. The multi-spacecraft method performs significantly more accurately than the others for a reliable prediction that is not just noise.

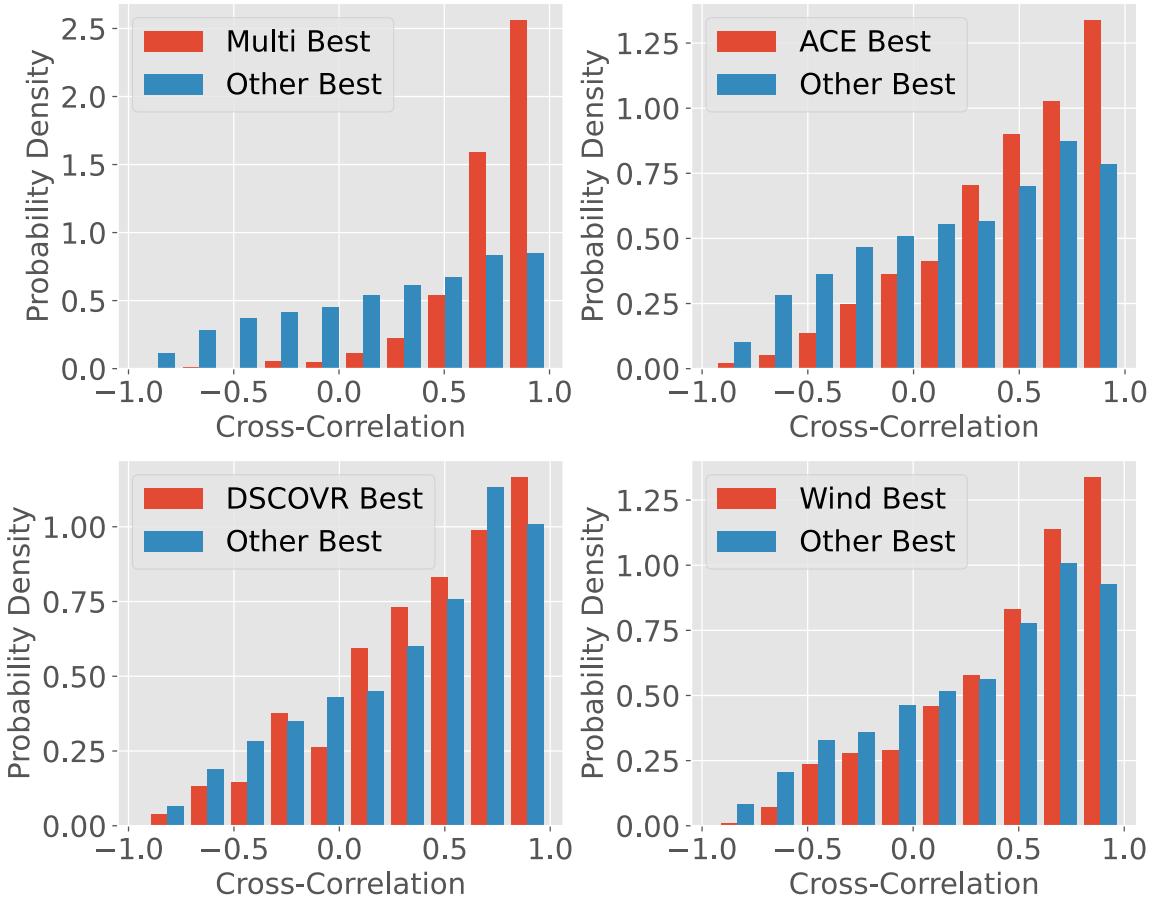


Figure 13: Cross-correlation histograms with the red bars showing only predictions where that particular method performed best. Probability density is used again, representing the cross-correlations produced if you take 2 equally sized samples, one where each method performs best, and one where it does not. The plots demonstrate that Multi will produce a significantly more accurate prediction than any single spacecraft. These produce similar correlations when they make the best prediction, compared to when they do not, while Multi produces a skewed distribution towards the highest accuracy values. Also, the multi-spacecraft forecasts are less influenced by ‘noisy’ predictions: ones which produce unphysical correlations (negative or close to zero).

Overall, cross-correlation is particularly effective when dealing with a static configuration, as the frozen-in approximation allows. However, in the context of our study, the interaction between the solar wind and the magnetosphere introduces non-

Method	Number of Most Accurate Predictions	Number With Correlation > 0.6
ACE	502	228
DSCOVR	683	286
Wind	579	276
Multi	549	444

Table 1: Frequency of most accurate predictions by method. The ordering is consistent with the mean values seen in Fig. 12, with DSCOVR producing the highest number of most accurate predictions and ACE the lowest. However, Fig. 13 proved that many of these are skewed by noisy predictions. The third column demonstrates that a multi-spacecraft approach outperforms the other three methods when you consider the most reliable predictions.

linearities to the solar wind propagation, meaning that solar wind conditions incident on the Earth are not always reflected linearly by SYM-H ground-based observations. Cross-correlation is still a useful metric for assessing our model’s sensitivity to variations; however, additional metrics are required to effectively quantify our results.

We adopt the model evaluation guidelines outlined in [32] as a basis for assessing the performance of our predictive model. It establishes two categories of metrics: fit performance metrics and event detection metrics. A combination of cross-correlation with these additional metrics provides a comprehensive analysis of our model.

3.3 Fit Performance Metrics

Eq. (15) and Eq. (16) are the first two fit performance metrics, the root-mean-square error (RMSE) and the mean absolute error (MAE). These quantities represent how accurately the model predicts each value of the observed SYM-H distribution.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

RMSE is more sensitive to outliers because of its squared residual term, therefore it highlights the largest deviations in our predictions. Consequently, it carries greater significance during active periods when the index’s magnitude is highest, leading to larger residuals $y_i - \hat{y}_i$. As a result, this quantity indicates how well our model performs at the most extreme values of SYM-H.

MAE, conversely, is described in [32] as ‘emphasising the usual state of the index’ because it sums the magnitude of all residuals without the squared term. It will be influenced more by the quiet periods of activity since there are significantly more of these in the dataset, as was illustrated in Fig. 11.

The last performance metric is the gradient from a linear fit between the predicted and observed values of SYM-H. This metric offers an alternative means of evaluating

3 SINGLE VS MULTI-SPACECRAFT EVALUATION

systematic biases in predictions, with 1 being the desired value indicating perfect match between predictions and observations. The results of these three metrics across all four methods are summarised in Table 2.

Method	Mean RMSE (nT)	Mean MAE (nT)	Mean Fit Gradient
Multi	5.16 ± 0.08	4.39 ± 0.07	0.43 ± 0.02
ACE	5.33 ± 0.08	4.51 ± 0.07	0.35 ± 0.01
DSCOVR	5.35 ± 0.08	4.54 ± 0.07	0.39 ± 0.01
Wind	5.33 ± 0.08	4.52 ± 0.07	0.41 ± 0.01

Table 2: Results of the fit performance metrics as specified in [32]. RMSE is more sensitive to outliers due to its squared residual, indicating that multi-spacecraft predictions make fewer poor predictions and perform well at highly active periods. The MAE indicates that multi-spacecraft predictions are generally closer to the true values of SYM-H. Finally, the linear fit produces a gradient below 1 in all 4 methods, suggesting overpredictions at the lowest magnitude values of SYM-H and underpredictions at the highest magnitudes. Multi-spacecraft forecasts still outperform the other three methods, with a gradient fit value closer to 1. Uncertainties on RMSE and MAE are calculated using bootstrapping over 100 thousand samples. The standard error on the mean is used for the fit gradients, σ/\sqrt{N} , where σ is the standard deviation and N is the number of samples, 2313.

Multi-spacecraft predictions demonstrate superior performance across all three metrics. The RMSE and MAE results indicate that we generally predict closer to the observed SYM-H value with a multi-spacecraft approach, while the linear fit values indicate that multi-spacecraft predictions track changes in SYM-H most accurately over the 4-hour period.

The RMSE and MAE results are analysed using bootstrapping, which generates uncertainties in our results. We generate 100,000 samples of our RMSEs and MAEs, with each value assigned an equal probability of selection. Replacement is allowed, such that repeat values may be generated in a random sample. This generates 100,000 distributions of 2313 values, which are representative of the typical results in our dataset. The mean of the bootstrapped RMSE values are plotted in Fig. 14. The Gaussian fits allow us to quantify the confidence levels associated with our predictions. The standard deviation on the mean RMSE and MAE distributions is used as the uncertainty in these performance metrics. The mean values of the bootstrapped samples were identical to the original sample means, to two decimal places, indicating a high level of consistency and reliability in bootstrapping.

The bootstrapped samples also represent the confidence with which we can say multi-spacecraft predictions produce the most accurate forecasts. The multi-spacecraft predictions are offset from each single spacecraft prediction in Fig. 14, and the means are separated by more than the sum of both σ s. Importantly, bootstrapping requires no underlying assumptions about the dataset, making it an effective method for estimating the statistical reliability of these results. Overall, this demonstrates the greater accuracy of the multi-spacecraft predictions.

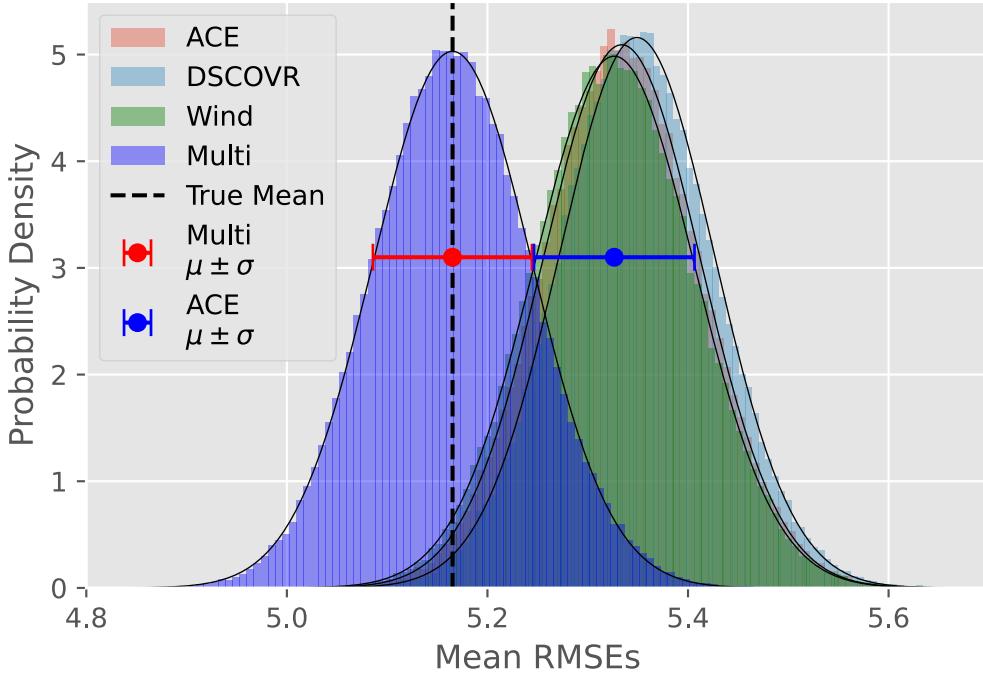


Figure 14: Histogram illustrating the distribution of mean RMSE values obtained through bootstrapping. Notably, the multi-spacecraft method produces the lowest error in its predictions. The significance of this result is shown by the comparison of standard deviations on the mean, depicted by the red and blue bars. Wind is selected as it has the smallest $\mu - \sigma$ value, therefore none of the single spacecraft methods produce standard deviations within 1σ of Multi’s mean, evident from the two error bars not overlapping.

The linear fit data is assessed in a different way, shown in Fig. 15, which fits Gaussian distributions to histograms of the fit gradients in each method. This method is preferred because the underlying data is normally distributed, evident from the χ^2 values calculated for the Gaussian fits in Table 3. The fits are shown to be statistically significant at the 5% confidence level, therefore they are good representations of the data.

These figures help visualise the variability in our predictions. The multi-spacecraft approach has a gradient value closest to 1, but the fit reveals it has the largest standard deviation among the four methods. This highlights a possible weakness in multi-spacecraft predictions, as there is a greater inconsistency in predicted gradients. This is likely because multi-spacecraft predictions are composed of all three single spacecraft measurements, therefore it will capture the outliers in all three. Although on average we will produce more accurate results, we should expect to produce more outliers in total with a multi-spacecraft prediction.

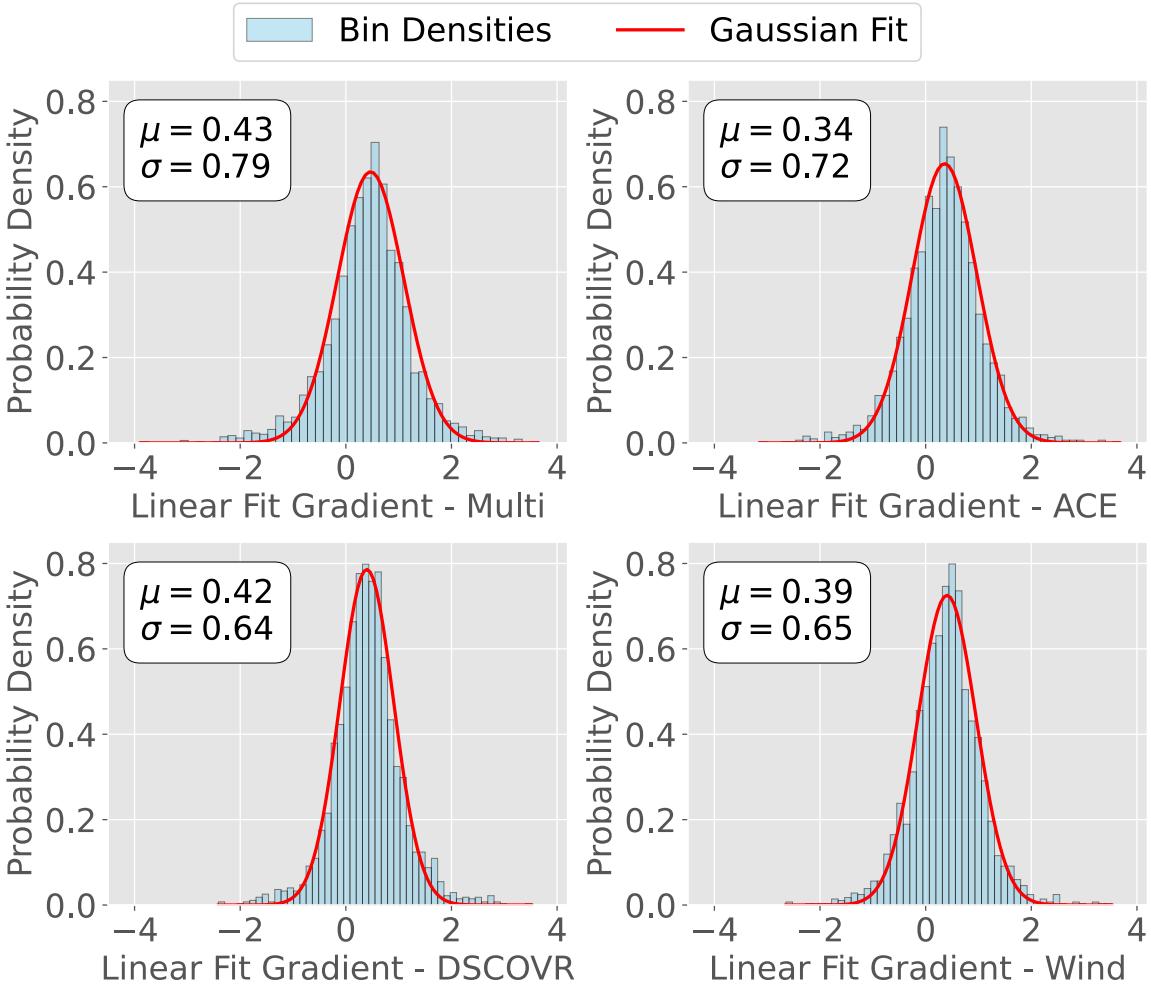


Figure 15: Histograms of the gradient distributions derived from linear fits to predicted versus observed SYM-H values, with 50 bins. Gaussian fits reveal the mean μ and standard deviation σ for each method. A greater consistency is observed in DSCOVR and Wind, which have smaller σ values of 0.64 and 0.65. In contrast, multi-spacecraft predictions have a larger mean, but a greater σ value of 0.79. While on average variations in SYM-H magnitude are predicted more successfully by the multi-spacecraft method, there is a greater spread in its predictions.

We saw in Fig. 11 that the activity levels of SYM-H vary greatly in our dataset. The large standard deviations seen in the linear fit gradient distributions prompt us to consider whether changes in SYM-H in each period affect the values of the gradients. Fig. 16 examines whether there is any correlation between particular phases of geomagnetic storms and the linear fit gradients, by plotting the change in SYM-H, Δ SYM-H, across each 4-hour period against the periods' linear fit gradients. The figure is split into two plots, one for negative Δ SYM-H to highlight a main phase with increasing activity, and one for positive Δ SYM-H to contrast recovery phases.

Method	χ^2 Statistic	p-values
Multi	32.7	0.96
ACE	31.3	0.97
DSCOVR	27.0	0.99
Wind	29.2	0.99

Table 3: χ^2 values obtained from fitting Gaussian distributions to the binned gradient histograms in Fig. 15. There are 48 degrees of freedom (50 bins - 2 independent variables μ , σ). The test is performed between $\mu \pm 3\sigma$ which ensures a minimum of 5 values per bin, therefore an effective χ^2 goodness of fit test can be carried out, while ensuring 99.7% of the data is incorporated. The p-values corresponding to each χ^2 demonstrate that all four results hold statistical significance at the 5% confidence level, therefore each Gaussian effectively approximates the underlying data.

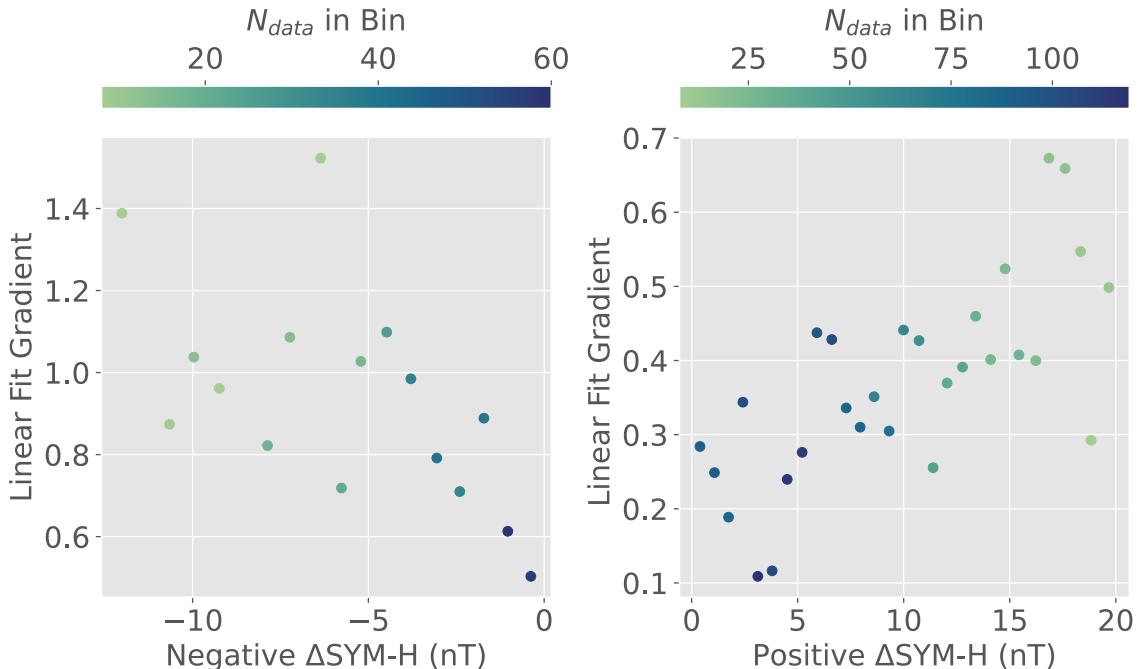


Figure 16: The relationship between the linear fit gradient and the signed spread of SYM-H, $\Delta \text{SYM-H}$. The spread, calculated as the difference between the last and first SYM-H values in a given period, is used to distinguish whether the period being considered is a main phase, recovery phase, or inactive period. The results are split into increasing and decreasing SYM-H to demonstrate the contrast between recovery and main phases. N_{data} is the number of points in a bin. Only bins with more than 10 values are displayed, reducing the effect of outliers. The gradients get closer to 1 with larger $\Delta \text{SYM-H}$ magnitudes, suggesting improved prediction accuracy during periods of larger geomagnetic activity.

The results show a clear trend of increasing gradient with a greater magnitude $\Delta \text{SYM-H}$. There are significantly more data points below a gradient of 1 than there are above it, resulting in the low mean values seen in Table 2. Notably, there are

more data points in the plot on the right, indicating that we have a larger number of recovery phases than main phases in our dataset.

The colour scheme for N_{data} confirms that the majority of data points have $\Delta\text{SYM-H}$ magnitudes $< 5\text{nT}$, which correspond to low activity events. This skews our linear fit gradients to lower values, in particular for small increases in SYM-H. This explains why we see a mean value much lower than 1 in Table 2. Accurately predicting small changes in geomagnetic activity is clearly a challenge. This shows the importance of considering the spread of SYM-H values in our analysis, as it directly influences the reliability of our predictions.

3.4 Event Detection Metrics

The second category employed in [32] assesses our model’s ability to detect significant geomagnetic events; a vital aspect of any space weather forecasting system. In this work, we are not interested in large storms as the capabilities to predict those are well-established. Using a single or multi-spacecraft method matters very little as they will be detected by all instruments. To assess multi-spacecraft predictions, we care about forecasting the finer details of the solar wind. Hence the minimum value of -15nT , defined in Section 2.1 as characterising the transition into a main phase, is used as a threshold value for event detection.

One approach is to classify events based on whether any SYM-H value in a period drops below the threshold, and determine whether any of our predictions follow and ‘successfully’ predict this event. Another approach is to classify events based on individual measurements, determining if a measurement falls below -15nT and whether our prediction at that time did the same.

The second approach is chosen because we are establishing this Earth-based testing model to quantify the success of solar wind propagation techniques, not to simply detect geomagnetic events. We want to know, at a high cadence, how our SYM-H predictions respond to changes in the solar wind, and using this metric for each individual prediction is useful for that purpose.

A confusion matrix quantifies the proportions of successes and failures in predictions above and below the threshold value. Fig. 17 compares confusion matrices for the multi-spacecraft predictions with the mean results of the single spacecraft predictions. The following elements make up the confusion matrices:

1. **True Positives (TP):** The forecast correctly predicts $\text{SYM-H} < -15\text{nT}$.
2. **True Negatives (TN):** The forecast correctly predicts $\text{SYM-H} > -15\text{nT}$.
3. **False Positives (FP):** The forecast fails to predict $\text{SYM-H} < -15\text{nT}$ (e.g. by overpredicting the magnitude of B_z).
4. **False Negatives (FN):** The forecast incorrectly predicts $\text{SYM-H} > -15\text{nT}$ (e.g. by wrong arrival time prediction).

We can define two other important quantities derived from these measures. Firstly, the sensitivity,

$$Sensitivity = \frac{TP}{TP + FN}, \quad (17)$$

quantifies the fraction of events ($SYM-H < -15nT$) that we predict successfully. On the other hand, the specificity, defined as

$$Specificity = \frac{TN}{TN + FP}, \quad (18)$$

tells us the fraction of non-events ($SYM-H > -15nT$) that we successfully predict.

The confusion matrix produced by our results is shown in Fig. 17, with the bootstrapping process to quantify the errors shown in Fig. 18. By propagating these errors, the sensitivity of our multi-spacecraft predictions was quantified as $70.4 \pm 0.7\%$, with a specificity of $97.0 \pm 0.1\%$. Single spacecraft predictions are combined into a mean sensitivity of $73.1 \pm 0.5\%$ and a mean specificity of $95.6 \pm 0.1\%$, indicating that we predict active events more accurately with a single spacecraft approach, and quiet periods more accurately with a multi-spacecraft one.

These results indicate that our model predicts non-events more accurately than it predicts events, with both a single and multi-spacecraft method. At low levels of activity around $-15nT$, this is not a major limitation of the model, but it means that we are struggling to pick up decreases in the values of $SYM-H$, which are driven by either the electric field term, $F(E)$, or the transient variations in pressure, P . Despite this observation, no correlation could be established between the periods in which we fail to predict events and either quantity, E or P . It is likely that we are seeing a consequence of the non-linear interaction between the solar wind and the magnetosphere, which is not accounted for in the empirical $SYM-H$ model. A future model should first aim to improve these event detection numbers as it is important for any $SYM-H$ model to detect the beginning of significant changes in solar wind conditions.

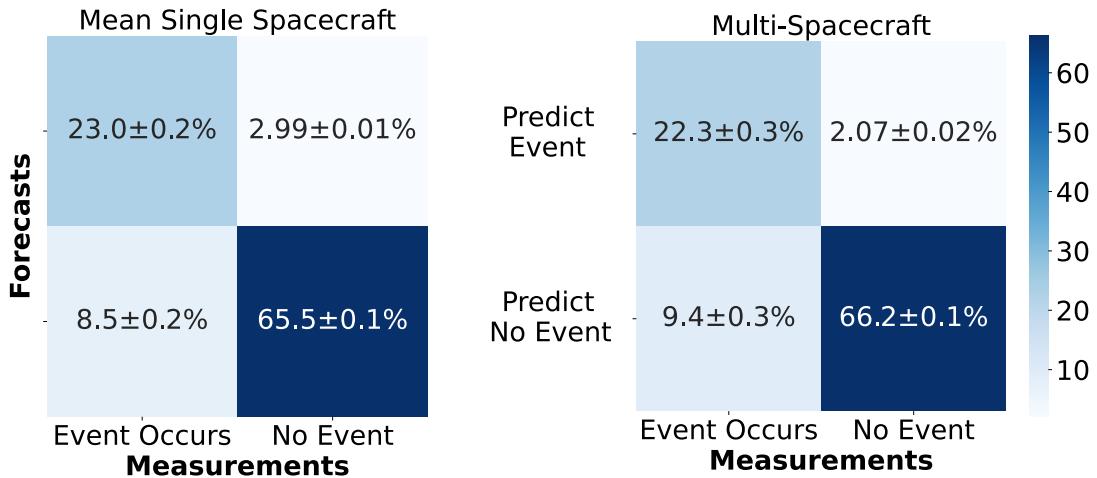


Figure 17: A confusion matrix representation of how the model predicts both significant and low activity events around a threshold value of $-15nT$. The multi-spacecraft model is compared with the mean of the quantities predicted by the single spacecrafts. Single spacecraft performs better in sensitivity (left column) while the multi-spacecraft predictions perform better in specificity (right column).

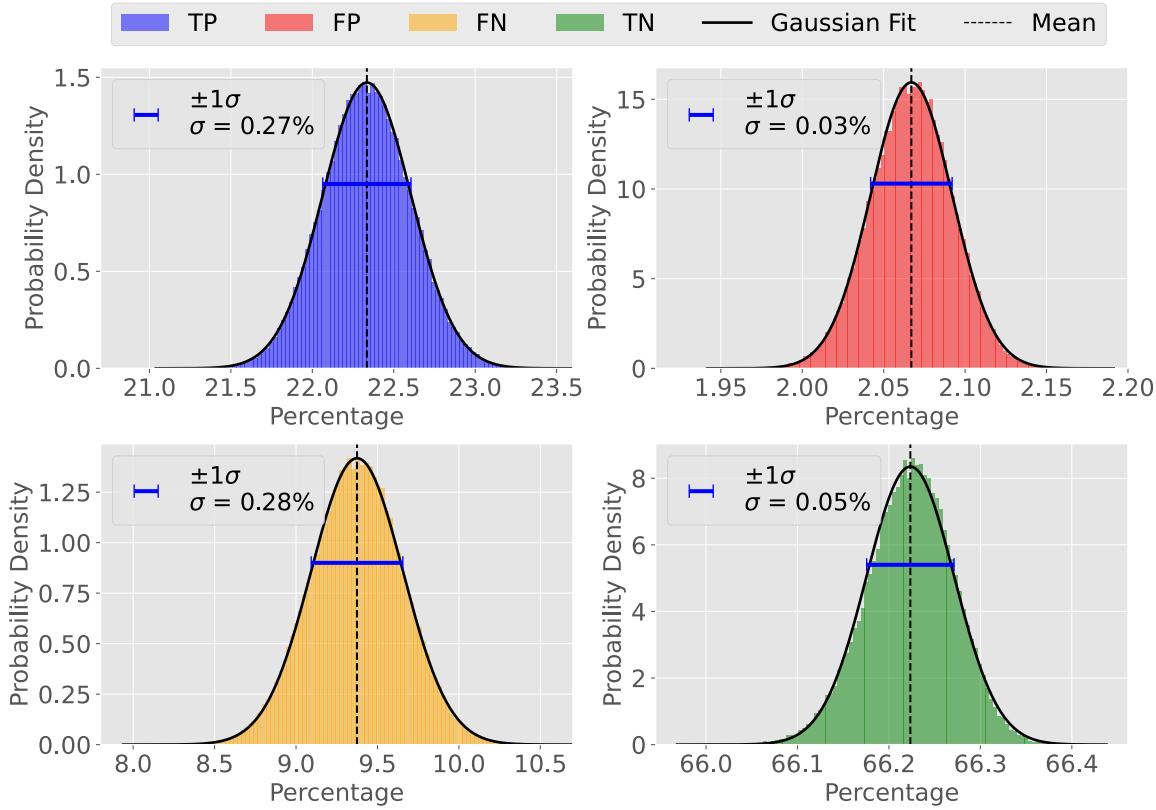


Figure 18: Bootstrapping of the multi-spacecraft event detection metrics over 100,000 samples, which generates standard deviations on the means in each sample, used as uncertainties on our results in Fig. 17. A similar bootstrapping analysis is applied to the single spacecraft predictions, which are combined to consider the mean event detection capabilities of single spacecraft methods. The uncertainties on these turn out lower than multi-spacecraft predictions despite including three measurements, indicating the greater impact of outliers on multi-spacecraft predictions that we saw with the linear fit gradients.

4 Improving the Model

The main objective of this work is to demonstrate that multi-spacecraft prediction methods can be tested effectively using geomagnetic indices. The weighted average method has been used because it was shown in [7] that it predicts the IMF z -component more successfully than single spacecraft methods. We needed a method that had shown such success, so that if we achieved the same success with our Earth-based model, we could then justify testing future multi-spacecraft approaches in this way. This would allow a greater freedom with testing data as the limitations discussed in section 1.3 would be removed.

Since using a SYM-H prediction model to test multi-spacecraft measurements is a new approach, there are lots of areas for improvement. This section assesses the weaknesses of the model's performance, while looking ahead to future spacecraft launches, in the hope that future models can consider these aspects in more detail.

4.1 Predicting Small Disturbances

In Fig. 16 we considered how the quantity $\Delta\text{SYM-H}$ impacted prediction accuracy with the linear fit gradients. Fig. 19 illustrates how cross-correlation varies with respect to $\Delta\text{SYM-H}$, which gives an insight into our predictive capabilities at different activity levels.

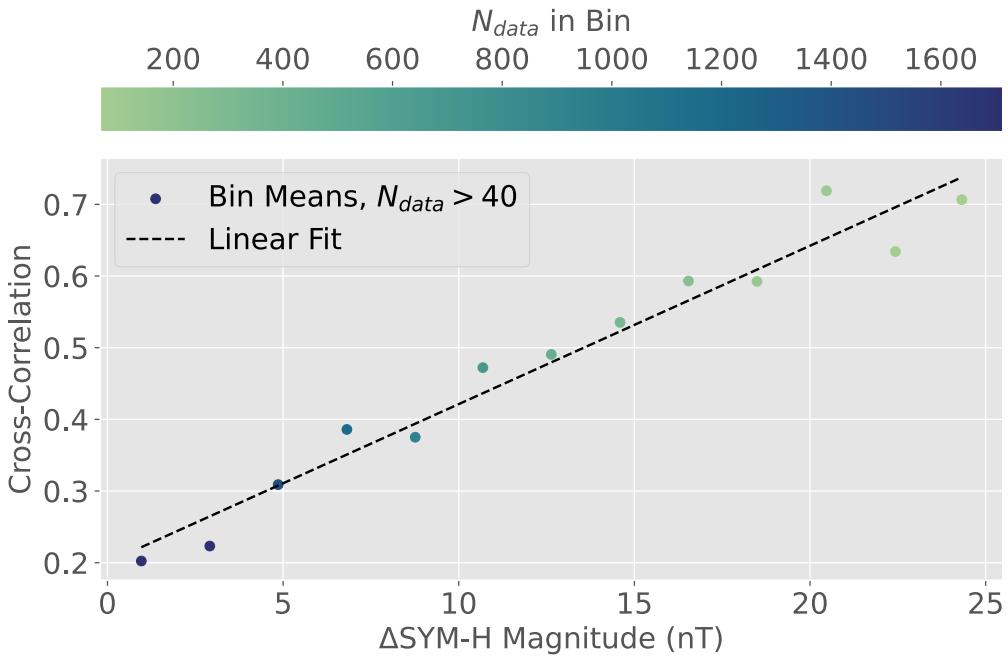


Figure 19: $\Delta\text{SYM-H}$ across a period versus the zero time lag cross-correlation produced by that period’s prediction for both single and multi-spacecraft predictions. $\Delta\text{SYM-H}$ values are placed into bins of width 2nT and the mean cross-correlation produced in all four prediction methods is plotted. A linear regression fit is applied to the data to demonstrate the relationship between our prediction accuracy and $\Delta\text{SYM-H}$. The choice of a linear fit, which produces an R^2 value of 0.96, demonstrates the systematic improvement in our predictions as $\Delta\text{SYM-H}$ increases. In Fig. 16, a minimum of 10 values per bin was used to remove outliers. Since we are considering all 4 methods in this plot, a minimum of 40 values is used.

Cross-correlation is used because larger $\Delta\text{SYM-H}$ values will naturally produce greater RMSE and MAE values, whereas cross-correlation is not directly impacted by the magnitudes of our results in this way. The number of data points in each bin is illustrated by the colour scheme, which demonstrates the large number of data points we have at low $\Delta\text{SYM-H}$ values, as was seen in Fig. 16. These results show the limitations in predicting the smallest disturbances in SYM-H, which we have a large sample size for. This likely arises from the non-linear interaction between the solar wind and the magnetosphere, as smaller changes in solar wind conditions will be more affected by this unpredictable dynamics. Larger, more distinct changes in the solar wind are more easily detected by SYM-H measurements and there is less uncertainty on these measures because of the magnetosphere interaction.

This demonstrates the limitations in the empirical SYM-H model that we use, which approximates the non-linear energy transfer by the linear term $F(E)$. It is therefore important to consider how we might go about improving this part of our model, to reduce the uncertainties present in testing both single and multi-spacecraft predictions.

4.2 Improving the Empirical SYM-H Model

We have provided a foundation for using geomagnetic indices to test predictions by using the Burton equation. Although the Burton equation captures the fundamental physics (energy conservation) of the solar wind-magnetosphere interaction, it does not deal with the complex non-linear processes involved. The model can be improved using a more advanced algorithm, which can be inserted into stage two of our model. Various machine-learning algorithms have been employed with a large amount of success, extending the Burton equation to ‘black box’ models using neural networks. This is an optimal approach, as phase 2 is simply a means to evaluate phase 1 of our model, the multi-spacecraft propagation, therefore we would like to minimise uncertainties in phase two as much as possible.

As well as improving our model, which has input parameters at the edge of the magnetosphere, a future model should also consider the propagation of the solar wind through the magnetosheath. Many of the ML algorithms developed, such as the neural network model [22] mentioned in section 2.2, consider the whole interaction, initiated at Earth’s bow shock nose. A future model should utilise these algorithms by shifting the solar wind parameters from L1 to the BSN, instead of the edge of the magnetosphere.

Concerning our model, a study in [33] was able to quantify uncertainties for the dynamic pressure related to processes within the magnetosheath. It recognised that the total pressure, comprising dynamic, magnetic, and thermal components, is what remains constant, rather than just the dynamic pressure that we assumed in our model. To address this limitation, future work could explore methods to calculate these additional components, and propagate the total pressure downstream, before calculating a dynamic pressure at the magnetopause, obeying this conservation principle.

4.3 Limitations in Multi-Spacecraft Predictions

In Section 3.3, we considered bootstrapping of the RMSE to compare single and multi-spacecraft predictions. This metric was chosen because it directly quantifies the average magnitude of errors between the predicted and observed values, providing a clear measure of prediction accuracy. However, we can also apply bootstrapping to the cross-correlation results discussed in Section 3.2, to assess the differences between each method in more detail. This metric assesses the model’s response to variations in the solar wind, with less concern over the magnitudes of predictions. The results are significant, as shown by Fig. 20. As before, bootstrapping generates standard deviations on the mean values which we use to quantify uncertainties in our results, shown in Table 4.

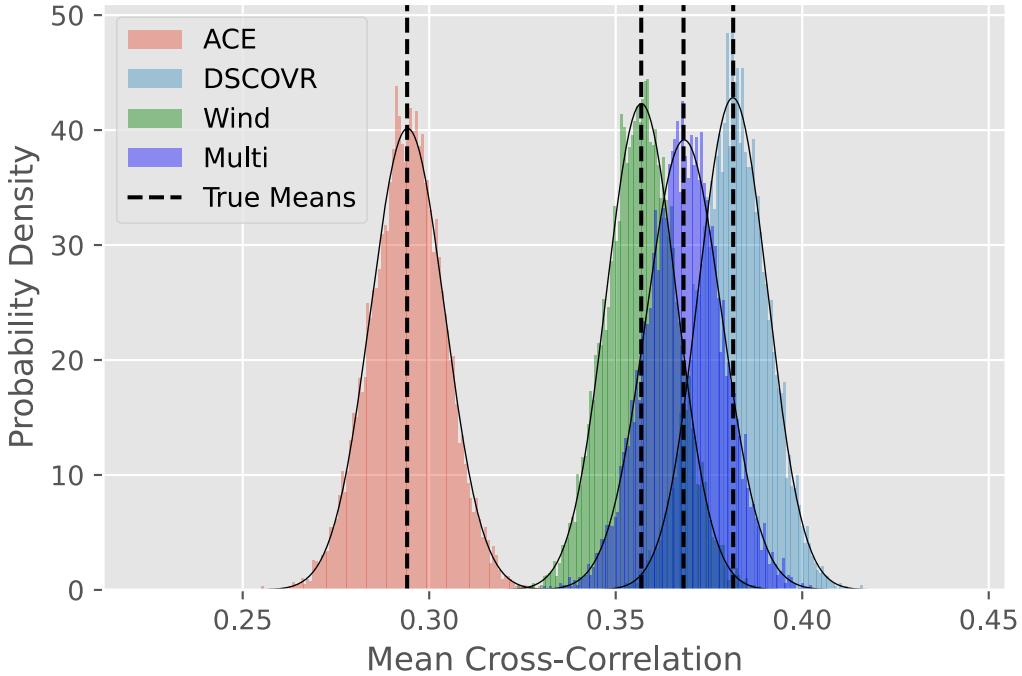


Figure 20: Mean cross-correlation values with zero time lag produced from bootstrapping the data in 100 thousand samples. ACE is a significant outlier, with a particularly low correlation mean of 0.29. Multi-spacecraft predictions still produce good predictions with a mean of 0.37, only beaten by DSCOVR at 0.38, indicating the strength of a multi-spacecraft approach to forecasting, which is hindered by ACE’s predictions. Standard deviations are 0.01 in each of the 4 methods.

Method	Mean Cross-Correlation
Multi	0.37 ± 0.01
ACE	0.29 ± 0.01
DSCOVR	0.38 ± 0.01
Wind	0.36 ± 0.01

Table 4: Zero lag cross-correlation mean values with uncertainties quantified by bootstrapping the data with 100 thousand samples. ACE produces a significantly lower value than the other three methods, therefore likely limits the performance of the multi-spacecraft method.

Fig. 20 highlights a serious limitation in using ACE to make predictions. In this metric, ACE produces significantly worse results than the other three methods. In Fig. 14, ACE was observed to have very similar RMSE values to the other two spacecraft, therefore this result highlights some key aspects about ACE. One possible cause of this result is that ACE systematically predicts SYM-H values at an offset time with

respect to the other two spacecraft. This could cause the magnitudes in predicted SYM-H to be close to the real values, leading to low RMSE values, but there would be a delay between the predictions and the measured values of SYM-H, causing a lower cross-correlation at zero time lag. However, this is an unlikely scenario as there was no clear offset between the datasets, which were thoroughly analysed before being used.

Alternatively, a more likely scenario is that ACE is less sensitive to the data. It can predict large events relatively successfully, and is good at capturing the average trends in the dataset (main phases, recovery phases), but it cannot predict the higher frequency disturbances in SYM-H, caused by the magnetopause currents. This consists of density data, which has already been significantly reduced due to the presence of substantial gaps mentioned in Section 1.4. Consequently, we may tentatively infer that the density data from ACE is unreliable.

Overall, the effectiveness of a multi-spacecraft method depends fundamentally on the quality of its individual components. The evidence suggests that ACE produces poor predictions compared with the other two spacecraft, therefore will limit the accuracy of the multi-spacecraft predictions. The fact that they still outperform single spacecraft predictions in a number of metrics indicates the effectiveness of a multi-spacecraft approach.

Testing the model with reliable data is a major goal in the future. One would assume that we will see the performance of multi-spacecraft predictions increase without the outlier seen in Fig. 20. Fortunately, two new spacecraft are planned for launch in 2025 and will aid the space weather forecasting community in improving the current methods.

4.4 Future Missions

With the launch of IMAP and SWFO-L1 in 2025, as well as the recent launch of Aditya-L1, new data will be available for testing our model. With access to IMAP’s orbital trajectory data, we can begin to consider the possibilities of using additional spacecraft for space weather forecasts. The orbital trajectories of the three spacecraft used in this work, along with IMAP, are plotted on Fig. 21, demonstrating the similarity in IMAP’s orbit to that of ACE and DSCOVR’s.

Coherence lengths, and the predicted y - z offsets of solar wind parameters, are critical aspects of the weighted average function used in our model. Fig. 22 demonstrates the proportion of time that we currently have spacecraft within the coherence length of B_z , and compares it to how frequently it will occur when IMAP is involved.

Overall, with more spacecraft within the coherence length of B_z , the predictions at the magnetosphere should become more reliable. There will be a maximum of 6 spacecraft in orbit at L1 within the next few years. In our research, we’ve demonstrated that an empirical SYM-H model, when used with a multi-spacecraft method, delivers more accurate predictions of solar wind parameters compared to single spacecraft methods. With this foundation, future methods that integrate data from IMAP and the other missions can apply our SYM-H model to test their predictions.

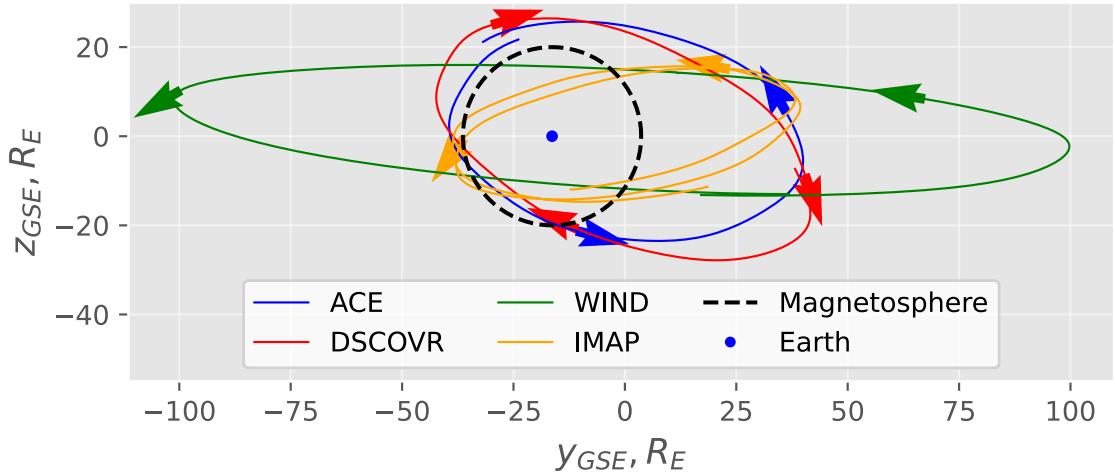


Figure 21: Orbital trajectories for ACE, DSCOVR, Wind and IMAP shown with respect to the magnetosphere, shifted to account for Earth’s orbital motion. IMAP will have a very similar orbit to ACE and DSCOVR’s, therefore will spend roughly the same amount of time within the coherence lengths of our parameters as these two.

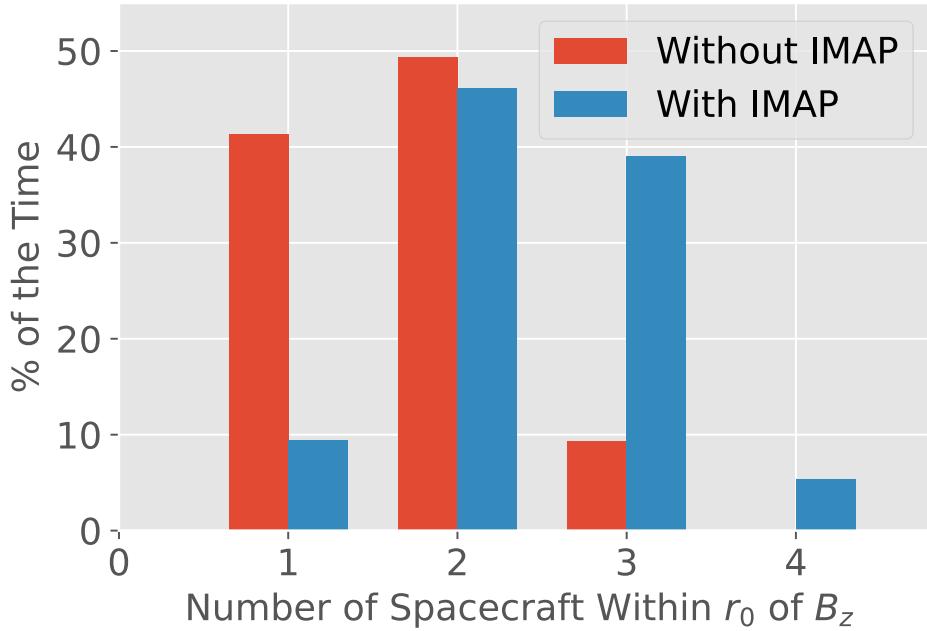


Figure 22: The number of spacecraft that will remain within the coherence length, r_0 , of B_z at L1, shifted to account for the effect of Earth’s orbital motion on predictions. IMAP significantly increases the fraction of time that three spacecraft lie within the coherence length, therefore we should expect more accurate multi-spacecraft predictions with IMAP included.

5 Conclusions

This work demonstrates the potential to test space weather predictions using an Earth-based geomagnetic index model. This would allow predictions to be tested without constraints on the availability of reference data, while producing a forecast that quantifies the impact of the solar wind on the Earth’s magnetosphere.

By testing our predictions of SYM-H with a range of performance metrics, we have seen that a multi-spacecraft approach produces predictions that are on average closer to the measured value of the index. This result was significant, with the nearest single spacecraft prediction producing a RMSE more than 1σ away from $\mu + \sigma$ of the multi-spacecraft prediction. However, the limitations in multi-spacecraft predictions were also apparent, specifically in detecting significant events $\text{SYM-H} < -15\text{nT}$. Possible causes of this, related to our approximation of the non-linear interaction between the solar wind and the magnetosphere, have been discussed and potential improvements to the empirical SYM-H model suggested. In particular, applying one of the neural network models developed, such as that in [22], would help reduce the uncertainties present in characterising the solar wind-magnetosphere coupling.

In the next few years, we will hopefully see huge advancements in multi-spacecraft predictions, as our ability to use more spacecraft within the coherence lengths of solar wind parameters increases. IMAP will help reduce this particular category of error, while it will also give us more important data for multi-spacecraft testing. With a significantly longer dataset, there will be an opportunity to test our predictions across a wider range of geomagnetic events.

At this present stage, there are a lot of uncertainties present in this model. However, there is evidence that these uncertainties can be reduced significantly. The model has shown promise in its ability to compare single and multi-spacecraft predictions, and in the future it would be desirable to continue using geomagnetic indices rather than spacecraft in lunar orbit to test predictions.

Acknowledgements

Thank you to our supervisor, Professor Tim Horbury, for his invaluable guidance, expertise, and support throughout this project. I would also like to thank my project partner, Ned, for his support and contributions during this project. I acknowledge the use of NASA/GSFC’s Space Physics Data Facility’s OMNIWeb service, and the use of OMNI data, and extend my thanks to Drew Turner of the APL for the IMAP orbital trajectory data. The data and code files that I developed can be found in [30].

References

- [1] P. Ribeiro, J. M. Vaquero, M. C. Gallego, and R. M. Trigo. The first documented space weather event that perturbed the communication networks in iberia. *Space Weather*, 14(7):464–468, 2016.

REFERENCES

- [2] Gabor Facsko, Gergely Koban, Nikolett Biro, and Munkhjargal Lkhagvadorj. Space weather effects on critical infrastructure, 2023.
- [3] Rositsa Miteva, Susan W. Samwel, and Stela Tkatchova. Space weather effects on satellites. *Astronomy*, 2(3):165–179, 2023.
- [4] H. Schulte in den Bäumen, D. Moran, M. Lenzen, I. Cairns, and A. Steenge. How severe space weather can disrupt global supply chains. *Natural Hazards and Earth System Sciences*, 14(10):2749–2759, 2014.
- [5] Nathani Basavaiah. *Upper Atmosphere Studies*, page 387–422. Springer Netherlands, 2011.
- [6] N. A. Case and J. A. Wild. A statistical comparison of solar wind propagation delays derived from multispacecraft techniques. *Journal of Geophysical Research: Space Physics*, 117(A2), 2012.
- [7] B. Boyde. Forecasting the interplanetary magnetic field at the earth using measurements from the l1 lagrange point. Master’s thesis, Imperial College London, 2021.
- [8] Donald H. Fairfield. Average and unusual locations of the earth’s magnetopause and bow shock. *Journal of Geophysical Research*, 76(28):6700–6716, 1971.
- [9] J. H. King and N. E. Papitashvili. Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data. *Journal of Geophysical Research: Space Physics*, 110(A2), 2005.
- [10] B. L. Burkholder, K. Nykyri, and X. Ma. Use of the l1 constellation as a multi-spacecraft solar wind monitor. *Journal of Geophysical Research: Space Physics*, 125(7), 2020.
- [11] James A. Wanliss and Kristin M. Showalter. High-resolution global storm index: Dst versus sym-h. *Journal of Geophysical Research: Space Physics*, 111(A2), 2006.
- [12] E. W. Grimes, N. Hatzigeorgiu, J. W. Lewis, C. Russell, J. M. McTiernan, A. Drozdov, and V. Angelopoulos. pySPEDAS: Space Physics Environment Data Analysis Software in Python. In *AGU Fall Meeting Abstracts*, volume 2019, pages SH41C–3313, 2019.
- [13] NASA Goddard Space Flight Center (GSFC)/Space Physics Data Facility (SPDF). DSCOVR data set. Created at GSFC/SPDF, 2019.
- [14] R. P. Lepping, M. Acuna, L. Burlaga, W. Farrell, et al. The wind magnetic field investigation. *Space Sci. Rev.*, 71:207, 1995.
- [15] Natalia E. Papitashvili and Joseph H. King. “OMNI 1-min Data” [Data set]. NASA Space Physics Data Facility, 2020.

REFERENCES

- [16] Z. Qin, R. E. Denton, N. A. Tsyganenko, and S. Wolf. Solar wind parameters for magnetospheric magnetic field modeling. *Space Weather*, 5(11), 2007.
- [17] Tohru Araki and Atsuki Shinbori. Relationship between solar wind dynamic pressure and amplitude of geomagnetic sudden commencement (sc). *Earth, Planets and Space*, 68(1), 2016.
- [18] I. I. ALEXEEV. The penetration of interplanetary magnetic and electric fields into the magnetosphere. *Journal of geomagnetism and geoelectricity*, 38(11):1199–1221, 1986.
- [19] W. D. Gonzalez, J. A. Joselyn, Y. Kamide, H. W. Kroehl, G. Rostoker, B. T. Tsurutani, and V. M. Vasyliunas. What is a geomagnetic storm? *Journal of Geophysical Research: Space Physics*, 99(A4):5771–5792, 1994.
- [20] Armando Collado-Villaverde, Pablo Muñoz, and Consuelo Cid. Classifying and bounding geomagnetic storms based on the sym-h and asy-h indices. *Natural Hazards*, 120(2):1141–1162, 2023.
- [21] S. J. Wharton, I. J. Rae, J. K. Sandhu, M.-T. Walach, D. M. Wright, and T. K. Yeoman. The changing eigenfrequency continuum during geomagnetic storms: Implications for plasma mass dynamics and ulf wave coupling. *Journal of Geophysical Research: Space Physics*, 125(6), 2020.
- [22] Yong Ji, Lan Ma, Chao Shen, Gang Zeng, YanYan Yang, and Shuo Ti. Composite model for predicting sym-h index. *Earth and Space Science*, 10(10), 2023.
- [23] R. K. Burton, R. L. McPherron, and C. T. Russell. An empirical relationship between interplanetary conditions and dst. *Journal of Geophysical Research*, 80(31):4204–4214, 1975.
- [24] Christopher T. Russell, Robert L. McPherron, and Rande K. Burton. On the cause of geomagnetic storms. *Journal of Geophysical Research*, 79(7):1105–1109, 1974.
- [25] Joseph E. Borovsky. On the motion of the heliospheric magnetic structure through the solar wind plasma. *Journal of Geophysical Research: Space Physics*, 125(2), 2020.
- [26] Mathew J. Owens, Mike Lockwood, Luke A. Barnard, Stephanie L. Yardley, Heli Hietala, Adrian T. LaMoury, and Laura Vuorinen. Annual variations in the near-earth solar wind. *Solar Physics*, 298(9), 2023.
- [27] B. Mailyan, C. Munteanu, and S. Haaland. What is the best method to calculate the solar wind propagation delay? *Annales Geophysicae*, 26(8):2383–2394, 2008.
- [28] John D. Richardson and Karolen I. Paularena. Plasma and magnetic field correlations in the solar wind. *Journal of Geophysical Research: Space Physics*, 106(A1):239–251, 2001.

REFERENCES

- [29] G.K. Parks. Magnetosphere. In James R. Holton, editor, *Encyclopedia of Atmospheric Sciences*, pages 1229–1237. Academic Press, Oxford, 2003.
- [30] L. Pyle. MSci-Project, GitHub repository. <https://github.com/loganp1/MSci-Project>, 2024.
- [31] T. Alberti, G. Consolini, F. Lepreti, M. Laurenza, A. Vecchio, and V. Carbone. Timescale separation in the solar wind-magnetosphere coupling during St. Patrick’s Day storms in 2013 and 2015. *Journal of Geophysical Research (Space Physics)*, 122(4):4266–4283, 2017.
- [32] Michael W. Liemohn, James P. McCollough, Vania K. Jordanova, et al. Model evaluation guidelines for geomagnetic index predictions. *Space Weather*, 16(12):2079–2102, 2018.
- [33] B. M. Walsh, T. Bhakyaibul, and Y. Zou. Quantifying the uncertainty of using solar wind measurements for geospace inputs. *Journal of Geophysical Research: Space Physics*, 124(5):3291–3302, 2019.