Draft version April 6, 2020
Typeset using LaTeX **manuscript** style in AASTeX63

SOLAR WIND RECONNECTION EXHAUSTS IN THE INNER HELIOSPHERE OBSERVED BY HELIOS AND DETECTED VIA MACHINE LEARNING

H. Tilquin,[1] J. P. Eastwood,[1] and T. D. Phan[2]

[1] *The Blackett Laboratory, Imperial College London, London SW7 2AZ, UK*

[2] *Space Sciences Laboratory, University of California, Berkeley, CA 94720, USA*

(Received April 6, 2020)

Submitted to the Astrophysical Journal

ABSTRACT

Reconnecting current sheets in the solar wind play an important role in the dynamics of the heliosphere and offer an opportunity to study magnetic reconnection exhausts under a wide variety of inflow and magnetic shear conditions. However, progress in understanding reconnection can be frustrated by the difficulty of finding events in very long time series of data. Here we describe a new method to detect magnetic reconnection events in the solar wind based on machine learning, and apply it to Helios spacecraft data in the inner heliosphere. The method searches for known solar wind reconnection exhaust features, and parameters in the algorithm are optimized to maximize the Matthews Correlation Coefficient using a training set of events and non-events. Applied to the whole Helios dataset, the trained algorithm generated a candidate set of events which were subsequently verified by hand, resulting in a database of 88 events. This approach offers a significant reduction in time needed to construct event databases

Corresponding author: J. P. Eastwood
jonathan.eastwood@imperial.ac.uk

compared to purely manual approaches. The database contains events covering a range of heliospheric distances from ~ 0.3 to ~ 1 au, and a wide variety of magnetic shear angles, but is limited by the relatively coarse time resolution of the Helios data. Analysis of these events suggests that proton heating by reconnection in the inner heliosphere depends on the available magnetic energy in a manner consistent with observations in other regimes such as at the Earth's magnetopause, suggesting this may be a universal feature of reconnection.

Keywords: magnetic reconnection — Sun: solar wind — machine learning — Helios

1. INTRODUCTION

Solar wind reconnection is crucial for many aspects of heliospheric physics such as modification of the global heliospheric magnetic field topology, erosion of large-scale structures including coronal mass ejections, solar wind heating and energetic particle acceleration (e.g. Gosling 2012; Bale et al. 2016; Kasper et al. 2016); it is therefore a science target for both Parker Solar Probe (Fox et al. 2016) and Solar Orbiter (Müller et al. 2013). Reconnection exhausts in the solar wind can grow to very large sizes, and exhausts have been encountered at a range of distances to the X-line (extending to 100s of Earth radii) (Phan et al. 2006, 2009; Eriksson et al. 2009; Gosling et al. 2007; Mistry et al. 2015, etc.), exceeding sizes typically observed in e.g. planetary magnetospheres. Furthermore, magnetic reconnection occurring across current sheets in the solar wind provides an excellent opportunity to study the physics of exhaust structure and plasma heating for a wide variety of inflow conditions and magnetic shear angles (Gosling et al. 2005; Drake et al. 2009; Phan et al. 2010; Eastwood et al. 2013). Reconnection in the inner heliosphere is likely to be characterized by lower plasma beta, which may affect exhaust structure, heating, and particle acceleration (Kontar et al. 2017; Drake & Swisdak 2014).

Whilst solar wind reconnection has been observed throughout the heliosphere, it has mainly been studied using data from spacecraft in the vicinity of 1 au. The majority of studies use direct inspection as the primary method of discovering magnetic reconnection events (e.g. Gosling et al. 2005; Phan

et al. 2010; Mistry et al. 2016, 2017b). Whilst most events can be discovered this way for small time frames (e.g., days or weeks), this approach becomes impractical for the analysis of long data sets, extending to many years or decades. Furthermore, an automatized detection algorithm will not suffer from bias or other human factors and will apply selection criteria consistently and uniformly.

Here we describe a new method to facilitate the detection of magnetic reconnection in the solar wind. The approach characterizes standard and robust features of previously identified solar wind reconnection events in a series of parameters, and uses machine learning to optimize thresholds for new event identification within these parameters. In particular, the method described below uses a training dataset to determine the best possible parameters for the automation by grid-searching and/or random-searching parameters and finding their associated Matthews Correlation Coefficient (Matthews 1975). Although verification of the algorithm output is subsequently required to identify and remove false positives, the method enables a considerably faster analysis of the data.

The method is demonstrated by detecting new events in data from the Helios spacecraft (e.g. Marsch & Schwenn 1990, and references therein), which measured the solar wind in the inner heliosphere from 0.3 - 1 au. Previous inner heliosphere reconnection studies using Helios have been based on a set of 28 events identified by eye (Gosling et al. 2006). Applied to 14 years of Helios 1 and 2 data, 77 new events were were detected, based on an initial training set derived from Gosling et al. (2006). However, we note that the time resolution of the Helios data is relatively coarse compared to other missions, and the true occurrence rate of solar wind reconnection in the inner heliosphere is likely to be higher since solar wind reconnection events lasting only a few seconds have been detected by spacecraft at 1 au (Gosling 2012). This has implications for novel high-time resolution observations by Parker Solar Probe and Solar Orbiter where the methodology presented here may find further application. This expanded database is used to examine some general properties of inner heliosphere magnetic reconnection, including the nature of ion heating which is found to be consistent with observations in other regimes, implying that heating by reconnection may follow a similar scaling law in a variety of environments.

The manuscript is structured as follows. In section 2 we describe the data sources and the analysis method. In section 3 we show the results of the machine learning, example events found by the method, and statistics relating to the occurrence of solar wind reconnection in the inner heliosphere. In section 4 we discuss the properties of solar wind reconnection exhaust heating in the inner heliosphere, and conclusions are presented in section 5.

2. METHODOLOGY

An overview of the event detection algorithm is summarized in Figure 1 as a flowchart. Following the initial data acquisition and processing, it is divided into three stages: correlation analysis; rotation into a local current sheet coordinate system; and further tests including the Walén test (Paschmann et al. 1979; Sonnerup et al. 1981).

Each stage of the event detection algorithm requires parameters, such as the time intervals to be used for the correlation analysis, thresholds on the magnetic field variations, etc. For a study of individual events, these parameters would typically be chosen on an event by event basis. However, selecting parameters for every candidate event would be impractical for our statistical study, because we sought to analyse the entire Helios data set. Instead, a fixed set of parameters is chosen before searching the Helios data. We adopt a machine learning approach to fix these parameters, whereby we first consider a training set of known reconnection events and non-reconnection events in order to best 'tune' the parameters of the event detection algorithm. We vary the parameter values and their combinations, and then test the event detection algorithm on the training events for each variation, checking to see which choice of parameters most accurately classifies the training event set. To quantify the accuracy of the event detection algorithm during its training process, we utilize a metric called the Matthews Correlation Coefficient (MCC), which is explained in more detail in Section 3. Essentially, the MCC approaches a perfect score of +1.0 as more of the training events are correctly identified; thus, optimizing the MCC will yield an optimum set of parameters to use for searching the Helios[1] data set. In particular we have used the reprocessed 'corefit' magnetic field and thermal

[1] available from http://helios-data.ssl.berkeley.edu

plasma dataset (Stansby et al. 2017, 2018) which is provided in the $[R, T, N]$ coordinate system. Helios 1 data is available from 1974-1984, and Helios 2 data from 1976-1980.

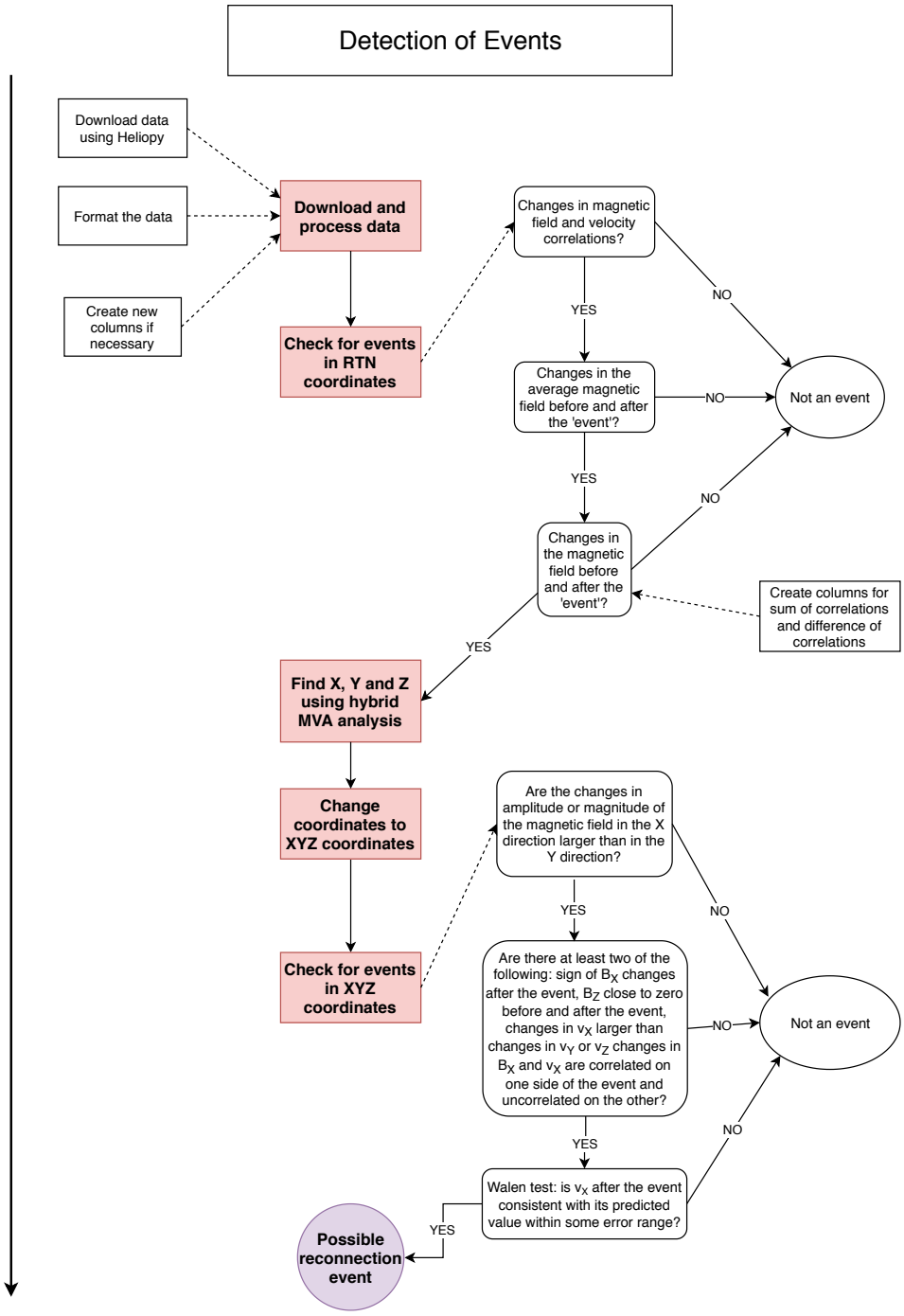

Figure 1. Flowchart summarizing the reconnection detection algorithm.

2.1. *Correlation Analysis*

Given the large extent of solar wind reconnection exhausts and the small size of the diffusion region, spacecraft typically observe crossings of the reconnection exhaust as it is convected past in the solar wind (Gosling et al. 2005). Reconnection events are characterized by a change in correlation of the magnetic field and velocity at the two exhaust edges, due to the Alfvénic disturbances of the exhaust being parallel to the magnetic field on one side of the exhaust, and anti-parallel on the other side (Gosling et al. 2005). The reconnecting component of the magnetic field changes sign across the exhaust; the duration depends both on the thickness of the exhaust (which varies with distance from the X-line) and the angle between the current sheet normal and the solar wind flow.

The first part of the detection process therefore examines whether there are significant changes in correlation between the magnetic field **B** and the proton velocity **v**. To capture this, we construct a correlation metric C where

$$C = \pm\frac{\sqrt{\Delta v_R \Delta B_R/\sigma_{v,R}\sigma_{B,R}}}{\Delta t} \pm \frac{\sqrt{\Delta v_T \Delta B_T/\sigma_{v,T}\sigma_{B,T}}}{\Delta t} \pm \frac{\sqrt{\Delta v_N \Delta B_N/\sigma_{v,N}\sigma_{B,N}}}{\Delta t}. \quad (1)$$

Here Δt is the time step and ΔB_i and Δv_i $(i = R, T, N)$ are the changes during one time step of the i^{th} component of **B** and **v** respectively. The parameters $\sigma_{v,i}$ and $\sigma_{B,i}$ are obtained by finding $\langle B_i \rangle$ and $\langle v_i \rangle$ over a sliding window of ten minutes duration, subtracting it from the corresponding B_i and v_i, and then taking the standard deviation of the new time series. Each term in the sum of equation 1 is treated as positive if the product under the square root is positive, and negative otherwise. Given the generally Alfvénic nature of the solar wind, we expect the sign of each term to be the same, and consequently, if **B** and **v** are correlated then C is positive, whereas if C is negative, this indicates an anti-correlation.

Large changes in correlation were then defined as a function of the surrounding values of C and $|dC/dt|$. Each point p in the data set was compared to the values surrounding it over a window of $\pm$6.75 minutes (optimised by maximising the MCC) on each side. Reference points were then defined. For C outliers, the reference point was defined to be 0. (Note that in this case, three minutes on each side of p were removed. The window length was still 13.5 minutes, but with a gap close to

the potential event, allowing the algorithm to detect events even when a series of points are outliers compared to their surroundings.) For $|dC/dt|$ outliers, the reference point was defined to be the median of the values in the window. If the values of C and $|dC/dt|$ at p compared to the reference point exceeded σ_C or $\sigma_{|dC/dt|}$ by some threshold (again optimised by maximising the MCC), then the point was considered an outlier. This approach was found to satisfactorily capture only large changes in the correlation that correspond to magnetic reconnection events and avoids small changes corresponding to fluctuations in the ambient solar wind, as illustrated in Figure 2.

Panels (a-c) show the R, T, and N components of $\mathbf{B}$ and $\mathbf{v}$ in pink and blue respectively. The reconnection exhaust was observed at just before 07:00 UT, and is primarily evident in the R and T components of the velocity. Figure 2(d-e) shows C and $|dC/dt|$, where the absolute value of the derivative is taken for convenience. C is large immediately adjacent to the reconnection exhaust, and changes from being strongly positive to strongly negative. Figure 2(f) shows the corresponding signatures in one component of the current sheet coordinate system, discussed in more detail below.

The algorithm then retains only events where there is a 'significant' change in correlation associated with 'strong' anti-correlations and correlations (as opposed to e.g., a strong correlation followed by a weaker one, which would still lead to a change in correlation). The meanings of 'significant' and 'strong' were defined by maximizing the MCC (see section 3 below). If these conditions were fulfilled, the time at which this occurred was retained for further analysis.

A further requirement is that the magnetic field changes from one stable orientation to another stable orientation either side of the change in correlation. The duration over which the average field was calculated was also optimized as part of the procedure described in section 3.

2.2. *Rotation to current sheet coordinate system*

Subsequent analysis of candidate solar wind reconnection exhausts is best performed in a local current sheet coordinate system $[X, Y, Z]$, where X is aligned to the reconnecting field, Y is in the out-of-plane direction, and Z is normal to the current sheet. This is because the reconnection exhaust should be seen in the X direction, and the reversal in the magnetic field across the current sheet should also be confined to the X direction, with the Z component being negligible. This can be seen in

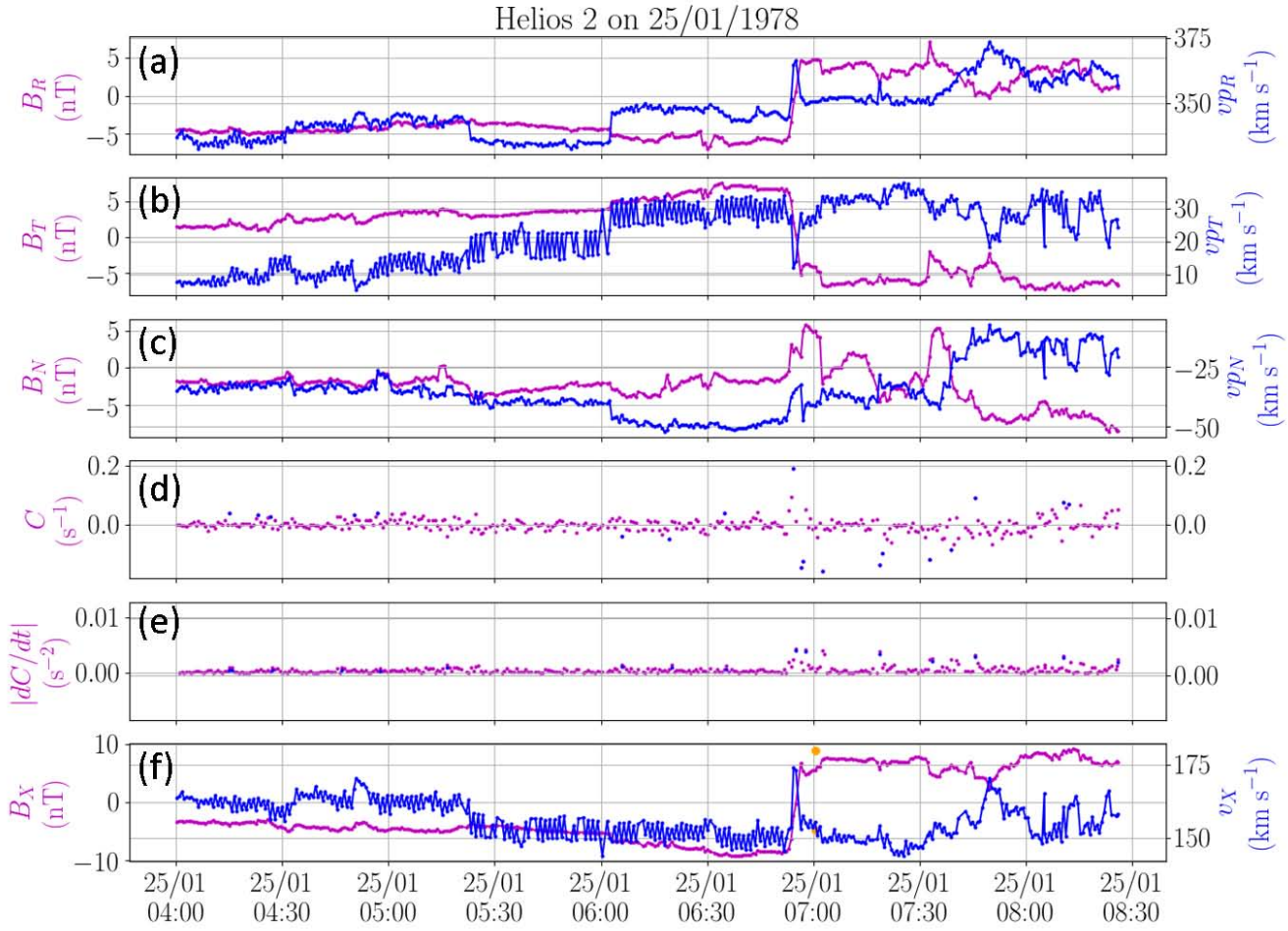

Figure 2. Solar wind reconnection exhaust observed by Helios 2 on 25 January 1978 at 06:56 UT. (a-c) R, T, and N components of the magnetic field (pink) and proton velocity (blue - N.B. variability in v_T is assumed to be due to instrumental effects); (d-e) correlation metric C and its derivative; (f) local current sheet X component of the magnetic field (pink) and proton velocity (blue). The blue dots in (d) and (e) represent outliers in the data, while the gold dots in panel (f) represent the thresholds for v_X suggested by the Walén test.

Figure 2(f). Here we identify the $[X, Y, Z]$ coordinate system using hybrid minimum variance analysis (hybrid MVA) applied to a ±30 minute interval centered on the exhaust (Sonnerup & Scheible 1998; Mistry et al. 2017a). X is obtained in the same way as in MVA, but Z is obtained by taking the

cross product of the magnetic fields of both incoming flows (therefore B_Z is zero at the two edge points); Y is then the cross product of Z and X (Mistry et al. 2017a).

2.3. *Further tests including the Walén test*

Other tests are then applied to candidate events to increase the likelihood that a positively identification is made. The following procedure is applied (Mistry 2016):

- Require that the variation in B_X is larger than in B_Y, both in amplitude and magnitude: we require either $|B_{1,X}-B_{2,X}| > |B_{1,Y}-B_{2,Y}|$ or $B_{1,X}^2 + B_{2,X}^2 > B_{1,Y}^2 + B_{2,Y}^2$,
- Check for a change in correlation between B_X and v_X,
- Ensure that the changes in v_X are larger than the changes in v_Y and v_Z,
- Require that B_X changes sign across the exhaust,
- Verify that $B_{1,Z} = B_{2,Z} = 0$.

The first requirement ensures that the expected current sheet magnetic field geometry is present. It was found that to recover a sufficiently large number of events for visual verification, only two of the other four tests had to be satisfied for the candidate to be retained as a possible reconnection event. These checks were automated, and natural variability in the solar wind meant that the correct intervals were not always chosen for computing these values, which in turn meant that some tests ended up not being satisfied even though the event was a reconnection exhaust. A more optimal method would be to determine the intervals for each event, but such a method has not been implemented at this stage.

Finally, we applied the Walén test (Paschmann et al. 1979; Sonnerup et al. 1981) to each candidate event. This provides a check as to whether the velocity of the exhaust is Alfvénic (Blagau et al. 2015). In particular, one can write that (Blagau et al. 2015; Hudson 1970):

$$\Delta\mathbf{v} = \pm\Delta\mathbf{v}_A = \pm\left(\sqrt{\frac{1}{\mu_0\rho_2}}\mathbf{B}_2 - \sqrt{\frac{1}{\mu_0\rho_1}}\mathbf{B}_1\right) \tag{2}$$

where symbols have their usual meaning. The subscripts 1 and 2 are used to denote the two sides of the exhaust. Furthermore, since the reconnection exhaust is expected to be aligned to the X direction, equation 2 can be further simplified by only taking the X components of $\mathbf{B}$ and $\mathbf{v}$ into account.

The Walén test is theoretically satisfied if equation 2 is fulfilled exactly. However, most of the time, the observed speed is typically within 30% of the predicted value (Sonnerup et al. 1981; Paschmann et al. 1986). The MCC procedure described below is used to optimize the threshold values for the Walén test to be satisfied. All candidate events with an exhaust speed deemed consistent with the Walén test are retained.

As a final check, the events produced by the detection algorithm were verified by hand. This leads to a final list of reconnection events that can be used for further analysis.

3. APPLICATION OF MACHINE LEARNING TO IDENTIFYING SOLAR WIND RECONNECTION EXHAUSTS

As noted above, several of the parameters in the algorithm described above were found by maximizing the Matthews Correlation Coefficient (MCC) for a given set of events and non-events. The MCC is a way to determine how good a set of parameters is at giving true positives and true negatives (for a binary data set), by using the following equation (Matthews 1975):

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}} \tag{3}$$

where TN, TP, FN and FP are the numbers of true negatives, true positives, false negatives and false positives respectively. The MCC can range from -1 to 1, with 1 being the optimal result, 0 corresponding to random outcomes and -1 corresponding to a situation where the opposite results are given all the time (Hu & Pan 2007).

Maximization of the MCC requires a training set; here we use 21 events previously identified by Gosling et al. (2006), and 20 non-reconnecting current sheets found by hand as the non-event set. The MCC was calculated for several sets and combinations of parameters, which form a grid

in n dimensions (where n is the number of parameters tested). Optimization of the parameters to maximize the MCC was first implemented with a grid search. Although relatively insensitive to grid resolution, this was found to be time consuming, taking several days to complete on a desktop machine when analyzing an extensive set of parameters. A random-search algorithm was therefore also implemented. While this method is less resistant to local maxima, it is much faster (taking a few hours) and does not restrict the search to the given parameters specified by the grid approach.

The best overall MCC obtained with this method was +0.40 (TN = 20, TP = 6, FN = 15, FP = 0), with the following parameter set and values found via random search:

- An interval of ±6.75 minutes to analyze C and $|dC/dt|$ on each side of the potential event,
- A standard deviation of C, $\sigma_C = 2.39$,
- A standard deviation of $|dC/dt|$, $\sigma_{(|dC/dt|)} = 2.91$,
- A magnetic field interval of ±4.91 minutes on each side to search for positive and negative $\mathbf{B}$ and the average of $\mathbf{B}$ on each side of the potential event,
- Minimum and maximum Walén fractions of 0.99 and 1.14.

The trained algorithm was then applied to the Helios 1 and 2 dataset. This resulted in a large database of several hundred candidate events which were then inspected by hand. This led to a database of 88 reconnection events which are listed in Table 1. Of these 88 events, 11 were previously found by Gosling et al. (2006), and formed part of the training set. This immediately illustrates an important limitation of the machine learning approach. Only if perfect performance can be achieved (i.e. MCC = 1) will the algorithm recover all the reconnection events in the training set, and by extension all similar events in the data. Nevertheless, this approach produced a roughly order-of-magnitude increase in the speed with which events could be identified, since it was necessary only to inspect the pre-selected candidates rather than the entire dataset during the parameter optimisation process.

4. RESULTS

4.1. *Statistical properties*

Figure 3(a-b) shows the number of events (panel a) and the dwell time (panel b) at different heliocentric distances. The occurrence rate of events is found to increase with distance from the Sun, except for 0.3-0.4 au, which does not follow the trend (see panel e). Figure 3(c-d) shows the occurrence of the events observed by Helios 1 (panel c) and Helios 2 (panel d) as a function of both time and distance from the Sun. In each case, the distance of the spacecraft from the Sun is shown in blue, and vertical blue bars mark data gaps. The occurrence of reconnection is shown in histogram format as grey bars, and also as black circles plotted on top of the heliospheric distance trace. This figure shows that although reconnection exhausts were observed throughout the range of heliocentric distances covered by the Helios orbit, they did not occur uniformly in time.

Statistical analysis shows that most reconnection events at 1 au last less than 2 minutes (Mistry 2016), and reconnection exhausts may last only a few seconds. It is possible that reconnection nearer the Sun may tend to be shorter in duration, and would not be as easily detected in the Helios data where the time resolution is of the order of 40 seconds. Further studies with higher time resolution data should resolve this question.

Finally, in the solar wind, it has been demonstrated that the magnetic shear across the reconnection exhaust can range from 180° (i.e., purely antiparallel) down to almost 0° (strong guide field reconnection) (Gosling 2012). Here we find that the database contains events with a wide variety of magnetic shear, as shown in Figure 4. This demonstrates the ability of the algorithm to detect even events with small magnetic shear where the reconnection jet is likely to be weaker than for antiparallel reconnection.

4.2. *Other events found as a by-product of the analysis procedure*

In the process of selecting candidate events, the algorithm may also produce events which are of interest but not directly the goal of the selection process. For example, Figure 5 shows a candidate event returned by the algorithm which is not included in the final list of exhaust crossings but is nevertheless of interest. Located at 0.96 au from the Sun, the current sheet is associated with what

Helios 1				Helios 2			
DD/MM/YYYY	HH:MM	DD/MM/YYYY	HH:MM	DD/MM/YYYY	HH:MM	DD/MM/YYYY	HH:MM
15/12/1974	**08:32**	19/01/1977	13:20	**27/01/1976**	**07:04**	22/02/1977	02:58
15/12/1974	**20:01**	29/01/1977	02:34	28/01/1976	05:11	27/02/1977	03:12
27/12/1974	00:32	03/02/1977	05:20	29/01/1976	13:40	31/03/1977	12:52
30/12/1974	23:03	20/02/1977	14:37	**30/01/1976**	**01:46**	18/04/1977	08:38
18/01/1975	**13:44**	15/04/1977	20:44	30/01/1976	03:32	23/11/1977	09:27
18/01/1975	15:26	30/10/1977	21:04	03/02/1976	08:55	04/12/1977	07:13
20/01/1975	23:15	25/11/1977	04:48	04/02/1976	00:09	04/12/1977	10:34
07/02/1975	**01:22**	03/03/1978	10:56	24/02/1976	15:47	05/12/1977	18:13
07/02/1975	12:06	10/12/1978	10:00	25/02/1976	19:50	09/12/1977	14:07
26/03/1975	00:54	23/12/1978	01:53	03/03/1976	21:27	16/12/1977	22:33
22/09/1975	**03:46**	19/01/1979	21:27	**04/03/1976**	**09:41**	31/12/1977	03:18
31/10/1975	14:42	10/03/1979	02:45	13/03/1976	11:56	31/12/1977	07:41
09/11/1975	21:17	28/01/1980	02:52	29/11/1976	23:10	21/01/1978	08:19
17/11/1975	00:52	09/02/1980	00:51	02/12/1976	19:21	**25/01/1978**	**06:56**
08/12/1975	02:08	17/02/1980	17:45	02/12/1976	23:43	27/01/1978	07:04
16/12/1975	22:56	27/04/1980	19:42	06/12/1976	06:03	28/01/1978	07:19
19/12/1975	**20:59**	29/05/1980	15:39	06/12/1976	06:37	23/02/1978	17:34
23/12/1975	03:33	30/06/1980	10:55	**15/12/1976**	**01:36**	22/04/1978	00:24
23/12/1975	05:42	02/11/1981	12:10	04/01/1977	01:45	22/04/1978	10:31
05/04/1976	01:58			16/01/1977	06:08	03/05/1978	14:56
01/11/1976	00:07			22/01/1977	04:29	26/08/1978	16:20
01/12/1976	05:49			31/01/1977	17:59	05/09/1978	01:53
15/01/1977	14:54			22/02/1977	02:01	28/11/1978	13:31

Table 1. Dates and times of reconnection events in the Helios 1 and 2 datasets, as found in this survey. Events previously identified by Gosling et al. (2006) are shown in bold.

appears to be a bi-directional jet which as expected is most prominent in Figure 5(g) showing B_X and v_x. A significant increase in the temperature is also observed during the current sheet crossing. There is also a bipolar variation in B_Z during the crossing, which could be suggestive of an X-line

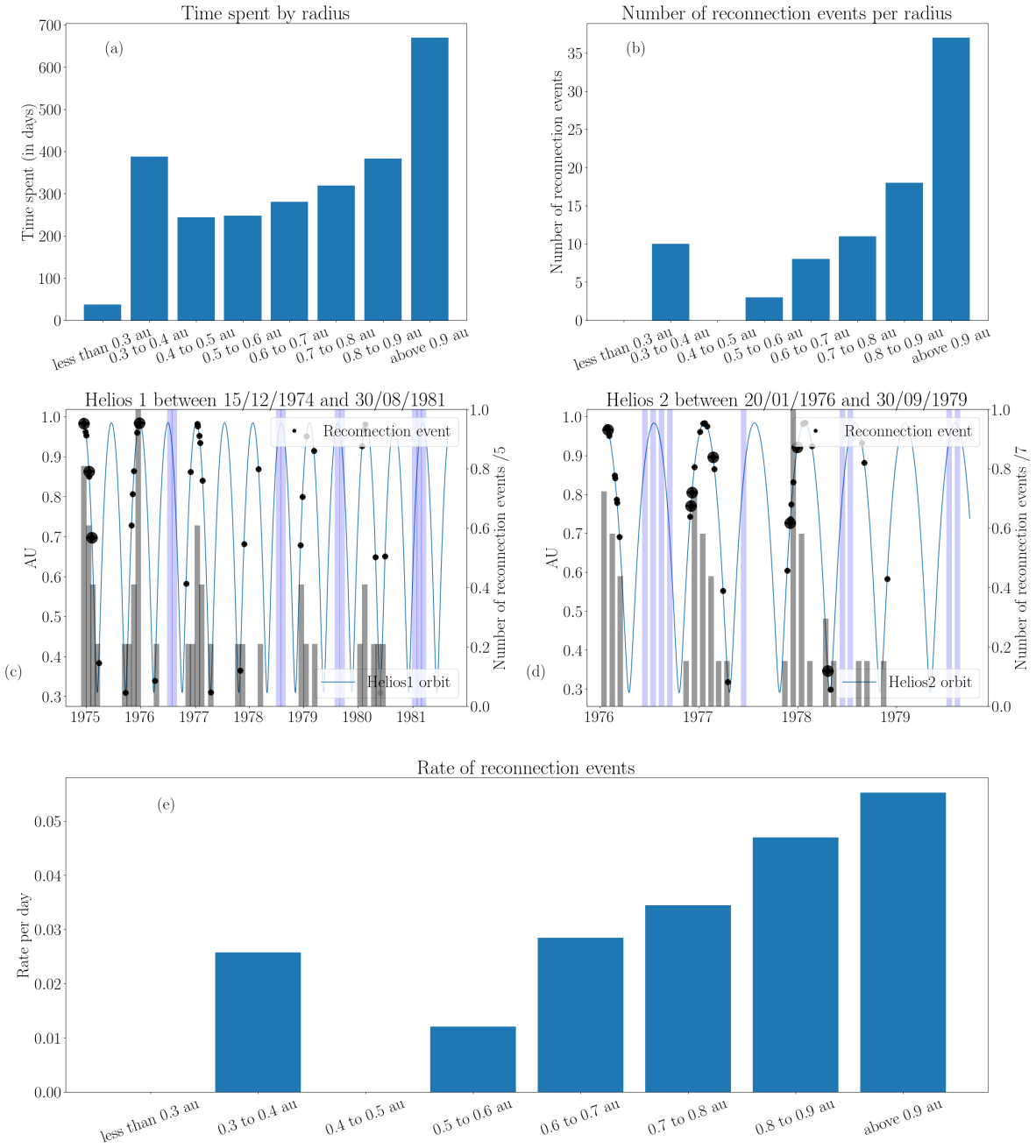

Figure 3. Occurrence of reconnection events. (a) Time spent by Helios 1 and 2 at different heliocentric distances. (b) Number of reconnection events as a function of distance from the Sun. (c-d) Events as a function of distance from the Sun and time for Helios 1 and 2. The grey lines represent the fraction of events normalised by the maximum number of events, and the blue shaded regions represent missing data. The area of the circles is proportional to the number of events per day for each month. Note that the Helios 1 data is shown only through 1981 as no events were detected after this point. (e) Number of reconnection events per day at different distances from the Sun.

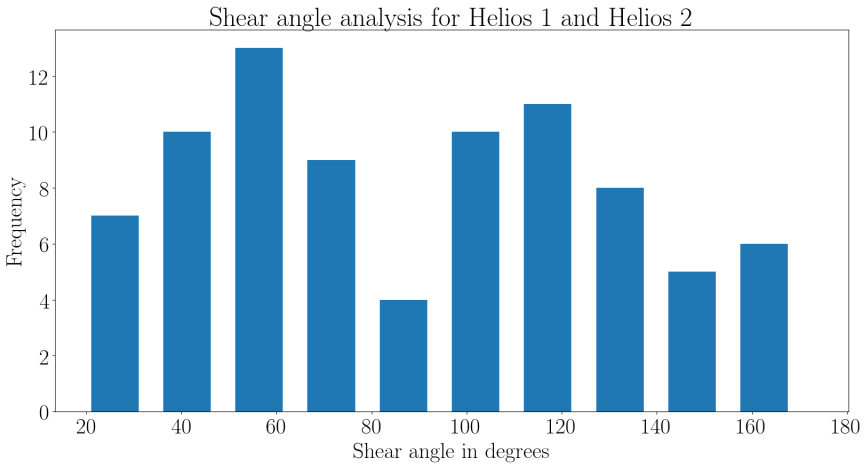

Figure 4. Histogram of the frequency of the magnetic shear angle for each reconnection event listed in Table 1

crossing, although the change in sign of B_Z is less clear. B_Z is the most difficult parameter to measure in reconnection because it is usually small, and therefore sensitive to small changes in the coordinate system (unlike B_X or B_Y).

The strength of the machine learning approach is that it enables a much more rapid survey of the data, whereas human skill allows for new, interesting and potentially unanticipated events to be identified. This event illustrates the strength of an approach which combines machine learning with human intervention.

4.3. *Reconnection heating*

According to theory, the proton heating due to reconnection, ΔT_p, should correspond to $\frac{1}{3}m_p v_A^2$, where m_p is the proton mass and v_A is the Alfvén speed (Drake et al. 2009). The $m_p v_A^2$ scaling of the change in temperature is related to the magnetic energy per particle that is available in the inflow. However, several studies show that the temperature increase inside the exhaust is much lower than expected (Drake et al. 2009; Phan et al. 2014; Mistry 2016). Drake et al. (2009) found a scaling coefficient of 0.13 when examining 22 solar wind events at 1 au. Mistry et al. (2017a) subsequently found a similar number (0.12) based on 188 solar wind events at 1 au, while Phan et al. (2014) found a coefficient of 0.13 at Earth's magnetopause.

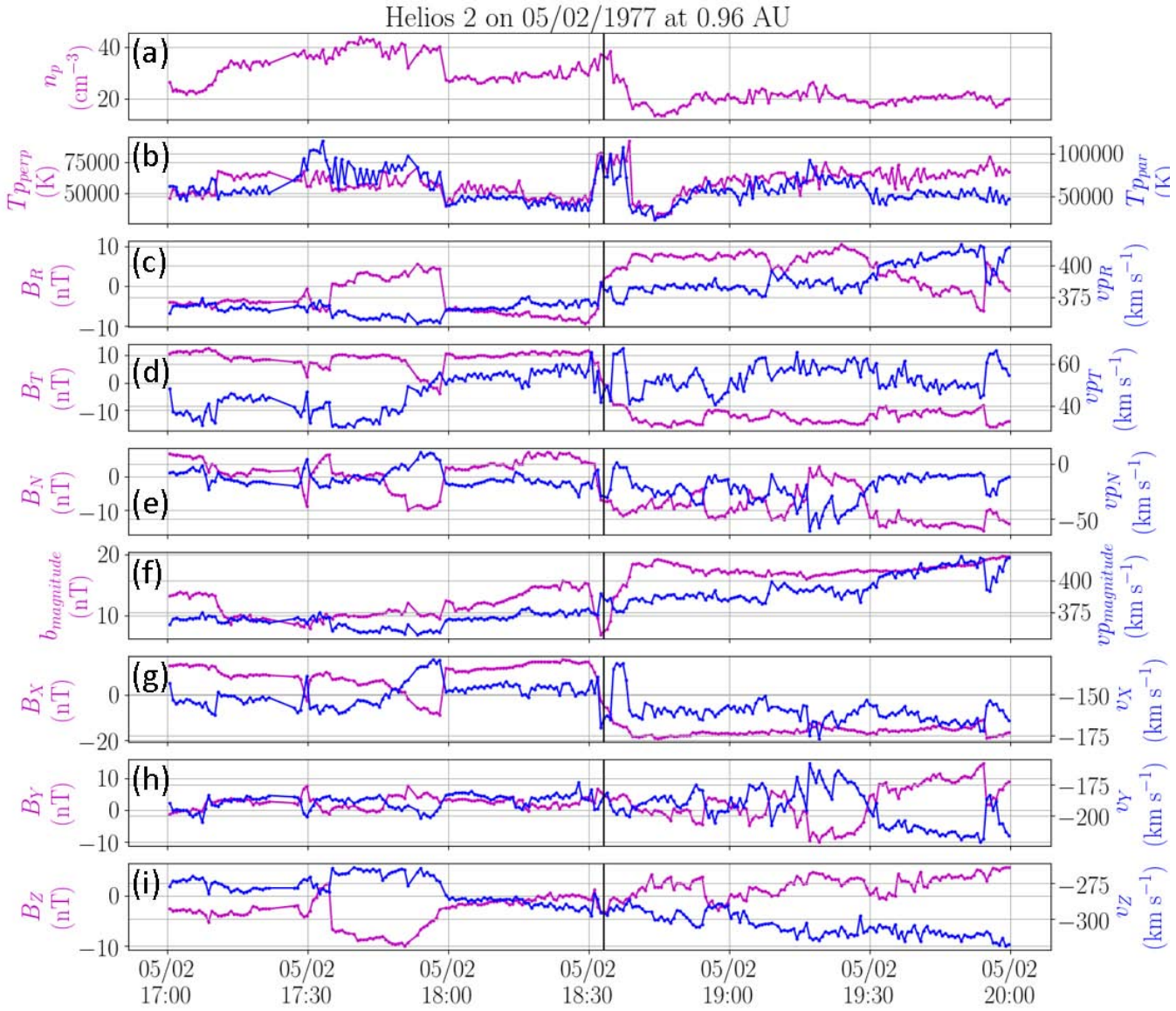

Figure 5. A possible X-line crossing detected by Helios 2 on 5 February 1977. (a) Proton number density; (b) proton perpendicular (pink) and parallel (blue) temperature (c-f) R, T, and N components, and magnitude, of the magnetic field (pink) and proton velocity (blue); (c-f) X, Y, and Z components of the magnetic field (pink) and proton velocity (blue) in the local current sheet coordinate system.

These lower than expected scaling coefficients are not yet fully understood. It has been suggested that this could be due to a slowing potential in the exhaust where an electric potential can develop and limit the energy gain of the particles (Haggerty et al. 2015). This potential decreases the speed of ions injected in the exhaust to below the Alfvén speed, which could then reduce the gradient to below $\frac{1}{3}$ (Haggerty et al. 2015). However, the study by Phan et al. (2014) suggests that both sub-Alfvénic and Alfvénic jets show the same kind of gradient when their speed is compared to ΔT_p.

The Helios database presented here offers an opportunity to extend this investigation into the inner heliosphere. Figure 6 shows ΔT_p plotted as a function of $m_p v_A^2$. Both the Alfvén speed and the inflow temperature were obtained as in the Phan et al. (2014) study, while the exhaust temperature was obtained by taking the 90^{th} percentile of the temperature within the exhaust. This characterizes the amount of heating without being unduly influenced by outliers as the Helios data can be noisy, and is consistent with previous approaches (Mistry 2016). We find that $\Delta T_p = 0.13 m_p v_A^2$ (see Figure 6), with a correlation coefficient of 0.63, in remarkable agreement with previous studies.

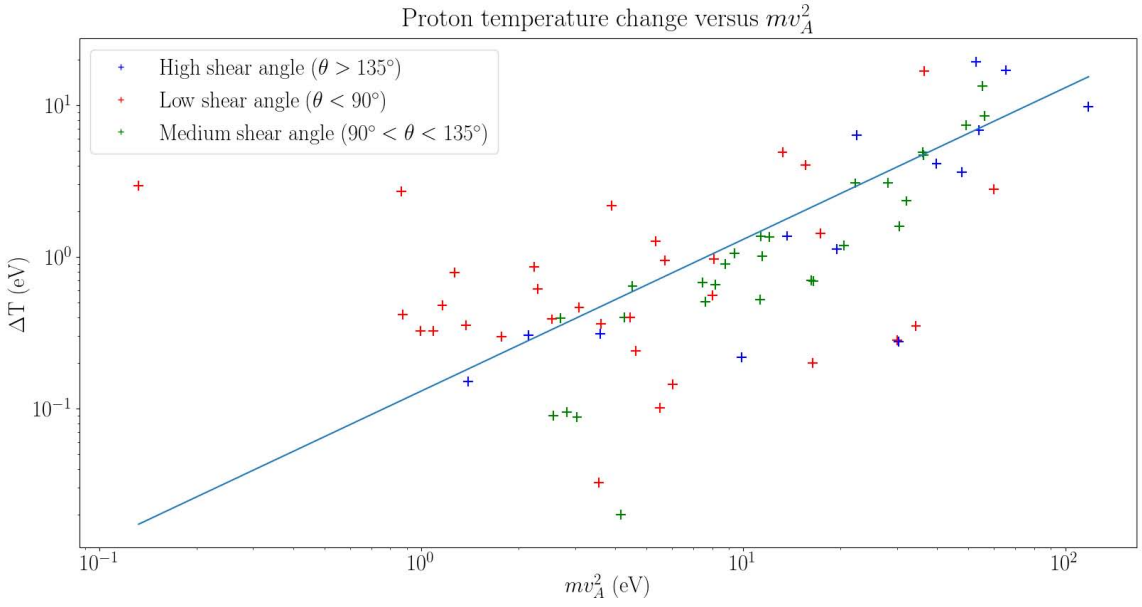

Figure 6. Proton temperature change, ΔT_i, as a function of inflow magnetic energy per particle, $m_p v_A^2$, in reconnection exhausts observed by Helios.

To explore the role of magnetic shear, the data have been further subdivided into low ($\theta < 90°$), medium ($90° < \theta < 135°$), and high ($\theta > 135°$) shear events. For high shear events the scaling coefficient stayed the same, at 0.131, but decreased to 0.009 when only low shear angles were taken into account. The worst correlation coefficient was obtained with small shear angles (0.40), while medium shear angles (between 90° and 135°) showed the best correlation coefficient (0.69). In contrast, at the magnetopause, Phan et al. (2014) did not observe as strong a sensitivity to shear angle.

The Helios data can also be used to investigate heating parallel and perpendicular to the magnetic field. Here we find the scaling coefficients between $m_p v_A^2$ and $\Delta T_{\rm perp}$ and $\Delta T_{\rm par}$ to be 0.10 and 0.22 respectively. These values are again close to those found previously in the Phan et al. (2014) study of magnetopause reconnection, who found scaling coefficients of 0.10 and 0.20. However, the coefficient of proportionality between $\Delta T_{\rm par}$ and $\Delta T_{\rm perp}$ was found to be 0.28 (with a correlation coefficient of 0.70). This is lower than the value (~ 0.5) one may expect based on the scaling of the parallel and perpendicular temperatures, as well as different from Phan et al. (2014) who found a value of 0.51, and appears to be due to the larger scatter found in the Helios data. We note the the main difference between the two studies is the typical change in temperature during the events, which are higher in the Phan et al. (2014) study. This may indicate that $\Delta T_{\rm perp}$ is less enhanced by reconnection if the overall temperature change is lower, or that heating in the inner heliosphere is more variable, but this requires further investigation.

5. CONCLUSIONS

In this work we have used machine learning to optimize an algorithm to detect reconnection exhausts in the solar wind. The algorithm is based on well-established generic features of solar wind reconnection such as changes in the correlation of the magnetic field and the solar wind velocity. It is trained by adjusting a variety of parameters so as to maximize the Matthews Correlation Coefficient. Here the training uses previously published reconnection events identified by hand. The output of the algorithm requires human intervention to finalize the list of events; in this way a database of 88 events was constructed from 14 years of Helios observations in the inner heliosphere. Running on

a standard desktop computer, the time taken for both training the algorithm and surveying the 14 years of Helios data is of the order of hours. Even taking account of the need to manually verify the algorithm output, this offers a very significant improvement compared to standard approaches requiring all the data to be inspected by hand.

More specifically, in this application a blended approach using machine learning and human intervention is found to be optimal. The machine learning aspect significantly reduces the amount of data under consideration. Even with moderate performance, it is a very effective tool for rapidly identifying enough events to draw meaningful conclusions about the properties of reconnection in the solar wind. The human intervention also allows for serendipitous discovery and identification of other events that might otherwise be missed, as illustrated in Figure 5. Changes to the algorithm such as using different methodologies, optimization techniques and parameter choices could all improve its efficiency, but it is unclear as to whether an algorithm capable of detecting every solar wind reconnection event is fundamentally feasible. This underlines the fact that such surveys should be used with care when trying to establish the absolute prevalence of reconnection.

Applied to the Helios dataset, events in the database occur over the full range of heliocentric distances covered by Helios, and a full range of magnetic shear angles. Analysis of the temperature increase shows that in the inner heliosphere a remarkably similar scaling coefficient is found as in other environments such as at the Earth's magnetopause and at 1 au. However, the time resolution of the Helios data is relatively coarse, and this limits the number of reconnection events that can be observed. The analysis has also now been successfully applied to the Ulysses dataset, and future application of this and similar techniques to other datasets, including those from Solar Orbiter and Parker Solar Probe, should provide further insight.

ACKNOWLEDGMENTS

The work of HT was performed under the auspices of the Undergraduate Research Opportunities Programme at Imperial College London. JPE was funded by STFC (UK) grant ST/N000692/1. TDP was supported by NASA grants NNX17AI25G, 80NSSC18K0157, and NASA contract NNG04EB99C. HT and JPE acknowledge the help of David Stansby in accessing the Helios data. The data used in this work are available from the Helios data archive http://helios-data.ssl.berkeley.edu and routines implementing the algorithms described in this work are available at https://github.com/THanae/MagRec. The authors thank the referee for their very thoughtful and constructive input during the review process.

REFERENCES

Bale, S. D., Goetz, K., Harvey, P. R., et al. 2016, Space Science Reviews, 204, 49, doi: 10.1007/s11214-016-0244-5

Blagau, A., Paschmann, G., Klecker, B., & Marghitu, O. 2015, Annales Geophysicae (09927689), 33

Drake, J., & Swisdak, M. 2014, PhysPlas, 21, 072903, doi: 10.1063/1.4889871

Drake, J., Swisdak, M., Phan, T., et al. 2009, J. Geophys. Res.: Space Physics, 114

Eastwood, J. P., Phan, T. D., Øieroset, M., et al. 2013, Plasma Physics and Controlled Fusion, 55, 124001, doi: 10.1088/0741-3335/55/12/124001

Eriksson, S., Gosling, J. T., Phan, T. D., et al. 2009, Journal of Geophysical Research: Space Physics, 114, doi: 10.1029/2008JA013990

Fox, N. J., Velli, M. C., Bale, S. D., et al. 2016, Space Science Reviews, 204, 7, doi: 10.1007/s11214-015-0211-6

Gosling, J., Eriksson, S., Phan, T., et al. 2007, Geophys. Res. Lett., 34

Gosling, J., Eriksson, S., & Schwenn, R. 2006, J. Geophys. Res.: Space Physics, 111

Gosling, J., Skoug, R., McComas, D., & Smith, C. 2005, J. Geophys. Res.: Space Physics, 110

Gosling, J. T. 2012, SSRv, 172, 187

Haggerty, C., Shay, M., Drake, J., Phan, T., & McHugh, C. 2015, Geophys. Res. Lett., 42, 9657

Hu, X., & Pan, Y. 2007, Knowledge discovery in bioinformatics: techniques, methods, and applications, Vol. 5 (John Wiley & Sons)

Hudson, P. 1970, Planetary and Space Science, 18, 1611

Kasper, J. C., Abiad, R., Austin, G., et al. 2016, Space Science Reviews, 204, 131, doi: 10.1007/s11214-015-0206-3

Kontar, E. P., Perez, J. E., Harra, L. K., et al. 2017, PhRvL, 118, 155101

Marsch, E., & Schwenn, R. 1990, in Physics of the Inner Heliosphere I, Large-Scale Phenomena, ed. R. Schwenn & E. Marsch (New York: Springer), 1

Matthews, B. W. 1975, Biochimica et Biophysica Acta (BBA)-Protein Structure, 405, 442

Mistry, R. 2016, PhD thesis, Imperial College London

Mistry, R., Eastwood, J., Phan, T., & Hietala, H. 2017a, J. Geophys. Res.: Space Physics, 122, 5895

Mistry, R., Eastwood, J. P., Haggerty, C. C., et al. 2016, PhRvL, 117, 185102, doi: 10.1103/PhysRevLett.117.185102

Mistry, R., Eastwood, J. P., Phan, T. D., & Hietala, H. 2015, Geophys. Res. Lett., 42, 10,513, doi: 10.1002/2015GL066820

—. 2017b, J. Geophys. Res., 122, 5895, doi: 10.1002/2017JA024032

Müller, D., Marsden, R. G., St. Cyr, O. C., Gilbert, H. R., & The Solar Orbiter Team. 2013, Solar Physics, 285, 25, doi: 10.1007/s11207-012-0085-7

Paschmann, G., Papamastorakis, I., Baumjohann, W., et al. 1986, J. Geophys. Res.: Space Physics, 91, 11099

Paschmann, G., Sonnerup, B. Ö., Papamastorakis, I., et al. 1979, Nature, 282, 243

Phan, T., Gosling, J., & Davis, M. 2009, Geophys. Res. Lett., 36

Phan, T., Gosling, J., Davis, M., et al. 2006, Nature, 439, 175

Phan, T., Drake, J., Shay, M., et al. 2014, Geophys. Res. Lett., 41, 7002

Phan, T. D., Gosling, J. T., Paschmann, G., et al. 2010, ApJ, 719, L199, doi: 10.1088/2041-8205/719/2/l199

Sonnerup, B. Ö., Paschmann, G., Papamastorakis, I., et al. 1981, Geophys. Res. Lett.: Space Physics, 86, 10049

Sonnerup, B. U., & Scheible, M. 1998, Analysis methods for multi-spacecraft data, 185

Stansby, D., Rai, Y., & Shaw, S. 2017, HelioPy, Zenodo, doi: 10.5281/zenodo.891405

Stansby, D., Salem, C., Matteini, L., & Horbury, T. 2018, SoPh, 293, 155, doi: 10.1007/s11207-018-1377-3