

## Classifying Magnetic Reconnection Regions Using $k$ -Means Clustering: Applications to Energy Partition



**Key Points:**

- $k$ -means clustering identifies statistically different regions of a symmetric reconnection site in a 2.5-D PIC simulation
- We confirm that outgoing particle energy fluxes decrease due to increased guide field
- Clustering may be applied to large quantities of simulation and in situ data to carry out statistical analysis of plasma phenomena

**Supporting Information:**

Supporting Information may be found in the online version of this article.





**Correspondence to:**

C. L. Waters,  
[cara.waters18@imperial.ac.uk](mailto:cara.waters18@imperial.ac.uk)

**Citation:**

Waters, C. L., Eastwood, J. P., Fargette, N., Newman, D. L., & Goldman, M. V. (2024). Classifying magnetic reconnection regions using  $k$ -means clustering: Applications to energy partition. *Journal of Geophysical Research: Space Physics*, 129, e2024JA033010. <https://doi.org/10.1029/2024JA033010>

Received 28 JUN 2024  
 Accepted 14 OCT 2024

Cara L. Waters<sup>1</sup> , Jonathan P. Eastwood<sup>1</sup> , Naïs Fargette<sup>1</sup>, David L. Newman<sup>2</sup> , and Martin V. Goldman<sup>2</sup> 

<sup>1</sup>The Blackett Laboratory, Imperial College London, London, UK, <sup>2</sup>Center for Integrated Plasma Studies, University of Colorado, Boulder, CO, USA

**Abstract** Magnetic reconnection is a fundamental plasma process which facilitates the conversion of magnetic energy to particle energies. This local process both contributes to and is affected by a larger system, being dependent on plasma conditions and transporting energy around the system, such as Earth's magnetosphere. When studying the reconnection process with in situ spacecraft data, it can be difficult to determine where spacecraft are in relation to the reconnection structure. In this work, we use  $k$ -means clustering, an unsupervised machine learning technique, to identify regions in a 2.5-D PIC simulation of symmetric magnetic reconnection with conditions comparable to those observed in Earth's magnetotail. This allows energy flux densities to be attributed to these regions. The ion enthalpy flux density is the most dominant form of energy flux density in the outflows, agreeing with previous studies. Poynting flux density may be dominant at some points in the outflows and is only half that of the Poynting flux density in the separatrixes. The proportion of outflowing particle energy flux decreases as guide field increases. We find that  $k$ -means is beneficial for analyzing data and comparing between simulations and in situ data. This demonstrates an approach which may be applied to large volumes of data to determine statistically different regions within phenomena in simulations and could be extended to in situ observations, applicable to future multi-point missions.

**Plain Language Summary** Magnetic reconnection is the change in connectivity of magnetic field lines. The small-scale physics at the center of reconnection causes conversion of magnetic energy to particle energies. This is important for the impacts on the large-scale system in which the reconnection takes place, such as Earth's magnetosphere. We use machine learning techniques to identify regions of the structure of the reconnection site. This can be done using only the plasma and field variables, without providing any information on the location. We can look at how energy is shared between the fields and particles in these regions to determine the behavior of the energy conversion across the reconnection site. We find ion energy is the largest output energy, agreeing with previous studies. As the magnetic field perpendicular to the reconnection structure increases, less energy is converted to the particles. This demonstrates that these machine learning techniques may be used to identify regions in plasma phenomena and could be applied to large amounts of data to carry out statistical studies.

### 1. Introduction

Magnetic reconnection is a universal process which facilitates the repartition of magnetic energy to particle energies. While it is a local scale process, it impacts a global system and so it is important to understand the structure of reconnection regions, and how it converts and partitions energy (Zweibel & Yamada, 2016). In the most basic sense, it has been described as a process where plasma flows across a surface separating regions of topologically different field lines (Vasyliunas, 1975). We can also consider it to be the local breakdown of flux conservation which results in a changing connectivity of field lines (Axford, 1984). Both of these definitions provide us with a simple structure of a reconnection site—inflowing regions of plasma which has not yet reconnected, outflowing regions of plasma which has been reconnected, and the surfaces separating these regions. At the center of the reconnection site, diffusion regions form where the particle flows are no longer frozen in to the magnetic field lines (Umeda et al., 2010; Yamada et al., 1997).

The separating surfaces are known as separatrixes, and in a 2-D projection take the form of lines. In reality, they are regions of complex plasma dynamics, including particle acceleration, instabilities, and energy conversion (Lapenta et al., 2015). They exhibit a layered structure with distinct sub-regions where ion and electron frozen-in conditions break down (Zhou et al., 2012), and are typically of the order of ion scales but with smaller structure on

the order of electron scales (Lindstedt et al., 2009) with structured and dynamic electric fields present even far from the central X-line (Retinò et al., 2006). Strong flow shears at the separatrix can lead to counter-streaming electron distributions, triggering beam-type instabilities and electrostatic turbulence, which contributes significantly to electron heating during the transition from inflow to outflow (Hesse et al., 2018). The presence of these various structures causing complex interactions results in significant energy conversion, with electrons reaching energies of up to 100 keV before entering the electron diffusion region (Lapenta et al., 2016).

Magnetic reconnection efficiently converts energy in an astrophysical plasma, and energy partition around diffusion regions and separatrices has been studied both in situ and in simulations (Aunai et al., 2011; Pucci et al., 2018; Tyler et al., 2016). We describe the energy partition in a space plasma using a continuity equation for energy,

$$\frac{\partial}{\partial t}(U_e + U_i + U_{EM}) + \nabla \cdot (\mathbf{Q}_e + \mathbf{Q}_i + \mathbf{S}) = 0, \quad (1)$$

where  $e$  and  $i$  subscripts represent electrons and ions respectively, the  $EM$  subscript is for electromagnetic energy,  $U$  is the energy density,  $\mathbf{Q}$  is the particle energy flux density, and  $\mathbf{S} = (\mathbf{E} \times \mathbf{B})/\mu_0$  is the Poynting flux density vector, where  $\mathbf{E}$  and  $\mathbf{B}$  are the electric and magnetic fields respectively (Birn & Hesse, 2005).

The particle energy flux densities  $\mathbf{Q}_s$  can be broken down further,

$$\mathbf{Q}_s = \mathbf{K}_s + \mathbf{H}_s + \mathbf{q}_s, \quad (2)$$

where  $\mathbf{K}_s$  is the species bulk kinetic energy flux density,  $\mathbf{H}_s$  is the species enthalpy flux density, and  $\mathbf{q}_s$  is the species heat flux density. The heat flux density  $\mathbf{q}_s$  is the trace of the third moment of the velocity distribution function of the species, and accurate values of this quantity may not always be obtained. The other components are calculated from the bulk velocity  $\mathbf{v}_s$ ,

$$\mathbf{K}_s = \frac{1}{2} m_s n_s v_s^2 \mathbf{v}_s, \quad (3)$$

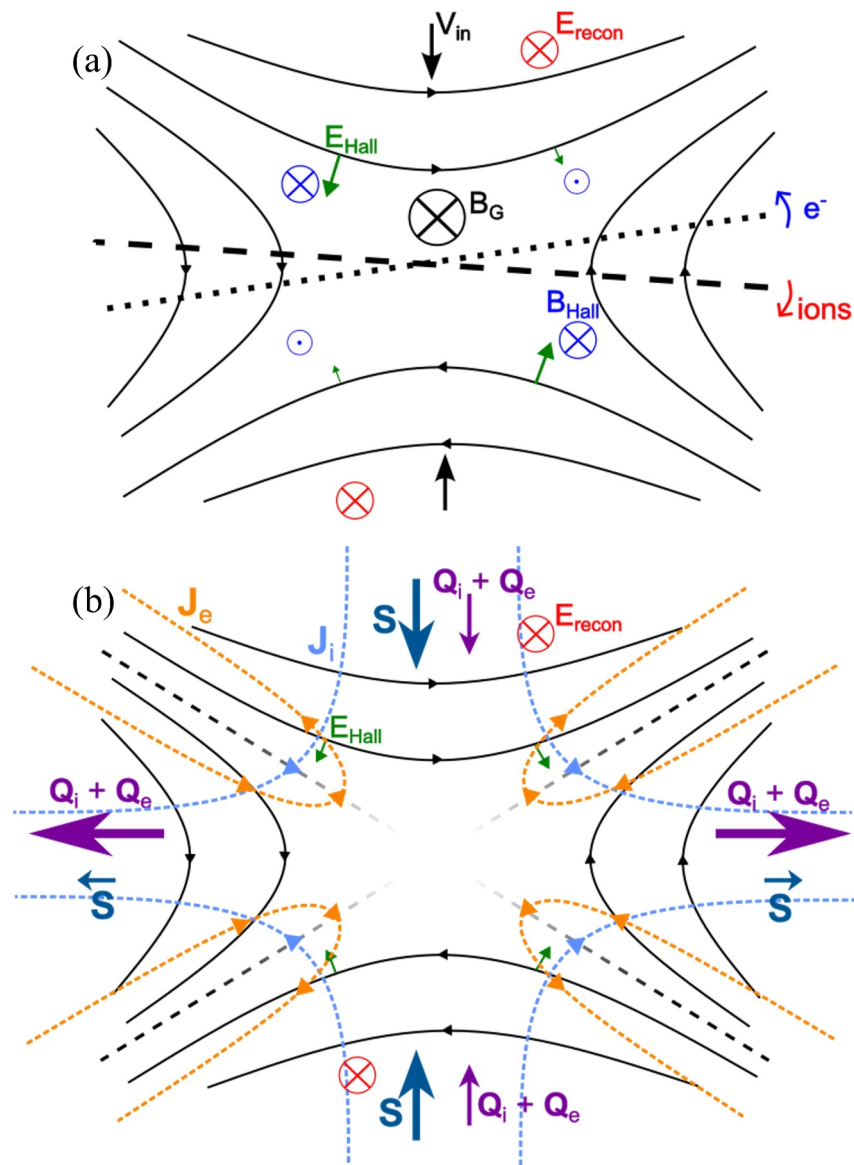
$$\mathbf{H}_s = \frac{\mathbf{v}_s \text{Tr}(\underline{\underline{\mathbf{P}}}_s)}{2} + \mathbf{v}_s \cdot \underline{\underline{\mathbf{P}}}_s, \quad (4)$$

where  $\underline{\underline{\mathbf{P}}}_s$  is the pressure tensor for the species,  $m_s$  is the species mass, and  $n_s$  is the species number density.

The energy flux density  $\mathbf{Q}$  or  $\mathbf{S}$  is the flow of each form of energy through each unit area, including the direction as it is a vector quantity. We expect a large, inward-oriented Poynting flux  $\mathbf{S}$  in the inflows, and a small outward-oriented one in the outflows. This change in Poynting flux results in a transfer to the particle energy flux densities  $\mathbf{Q}_i$  and  $\mathbf{Q}_e$  for ions and electrons respectively (Figure 1b).

Previous spacecraft observations of reconnection energy fluxes have mainly focused on the outflow, in part because they are a distinctive and largely unambiguous feature of reconnection in situ. It is found that for symmetric, collisionless reconnection in the Earth's magnetotail, ion enthalpy flux is the dominant output, although Poynting flux may dominate at certain regions within the ion diffusion region (IDR) (Eastwood et al., 2013). More recently, the Magnetospheric Multiscale (MMS) mission has given the first direct observations of the electron diffusion region (EDR), showing crescent-shaped electron velocity distributions and intense currents and electric fields (Burch et al., 2016). These crescent-shaped distributions are also observed in magnetotail reconnection, alongside super-Alfvénic electron jets (Torbert et al., 2018). On average, electrons have higher energies when local electric fields have high magnitudes in all reconnection environments (Oka et al., 2022), which relates to significant electron energy flux densities being found in the vicinity of the EDR which are orthogonal to the ion outflow (Eastwood et al., 2020).

However, to study how the energy partition varies through the reconnection structure, knowledge of the spacecraft location within said structure is necessary. Visual identification of such a structure can be challenging and time consuming. For specific events, field reconstruction has been carried out to identify the relative location



**Figure 1.** (a) The effect of a guide field  $B_G$ , into the page, on the structure of a magnetic reconnection site. The magnetic field lines are shown as solid oriented black lines, the reconnection electric field  $E_{recon}$  is in red and directed into the page, while the Hall electric fields  $E_{Hall}$  are shown as green arrows. The Hall magnetic field  $B_{Hall}$  is shown in blue, with direction in relation to the plane of the page indicated. The inflow velocity  $V_{in}$  is shown by black arrows. The plane of rotation of the ions is shown by a black dashed line, and that of the electrons by a black dotted line. (b) The flow direction of energy flux densities and the quantities responsible for their repartition as expected in a typical case of 2D symmetric, antiparallel magnetic reconnection. The separatrices are shown by black dashed lines. Electron and ion currents  $J_e$  and  $J_i$  are represented respectively as orange and light blue dashed arrows. The Poynting flux density vector  $S$  is shown by dark blue arrows, and total particle energy flux density  $Q_i + Q_e$  is shown by purple arrows.

of spacecraft (Denton et al., 2020; Hasegawa et al., 2019; Nakamura et al., 2019). Although these methods are accurate and produce very detailed descriptions of the surrounding field topology, they are computationally intensive and the results cannot be generalized to multiple events.

Furthermore, the reconnection process is influenced by the conditions of the surrounding plasma environment. One example of this is a guide field, which is an out-of-plane component of the magnetic field resulting in the field being non-zero at the X-line. This may distort the structure of the diffusion region and tilt the current sheet (Eastwood et al., 2010). This also causes the electron outflow to tilt toward the separatrices (Wilder et al., 2017), as observed in Figure 1a. The overall rate of reconnection decreases with guide field in the symmetric case (Hesse

**Table 1**  
Initial PIC Simulation Parameters

Input parameters	Value
$L_x \times L_y$ (units of $d_{i0}$ )	$200 \times 40$
$\Delta/d_{i0}$	0.5
$m_i/m_e$	256
$v_e/c$	0.045
$T_e/T_i$	5
$n_b/n_0$	0.1
$B_G/B_0$	0, 0.1, 0.2

et al., 2013; Ricci et al., 2004), and the energy partition is affected as the Poynting flux associated with the guide field is carried from inflow to outflow with little alteration (Birn & Hesse, 2010).

Recent applications of machine learning techniques have started to facilitate identification of regions for statistical studies of phenomena, including identifying EDRs in MMS data (Lenouvel et al., 2021), classifying the magnetospheric regions in which spacecraft observations are taken at Earth (Breuillard et al., 2020; Nguyen et al., 2021; Olshevsky et al., 2021; Toy-Edens et al., 2024) and at Saturn (Yeakel et al., 2022), and classifying solar wind (Camporeale et al., 2017). These techniques have also been applied to smaller-scale structures, including identifying regions within sunspots (Romano et al., 2023) and identifying regions of instabilities in particle-in-cell (PIC) simulations which correspond to

physically known properties (Köhne et al., 2023). These applications allow for large volumes of data to be labeled in relation to physically known regions or classified as statistically significant populations, which would be repetitive, time consuming, and include biases which cannot be understood if this labeling were carried out with human input.

In this work, we use  $k$ -means clustering, an unsupervised machine learning technique, to identify regions in a 2.5-D PIC simulation of symmetric magnetic reconnection with conditions comparable to those observed in Earth's magnetotail. We implement this technique such that direct comparison can be made with results from spacecraft observations and verify and extend understanding of energy flux densities in the labeled regions. Using three simulation runs, we investigate the effects of guide field varying between 0, 0.1 and 0.2 of the background field. Simulation facilitates studying the effects of this level of variation on energy repartition, as it may be difficult to determine the strength of the guide field to sufficient accuracy due to limitations in finding a suitable coordinate system for an event (Denton et al., 2024).

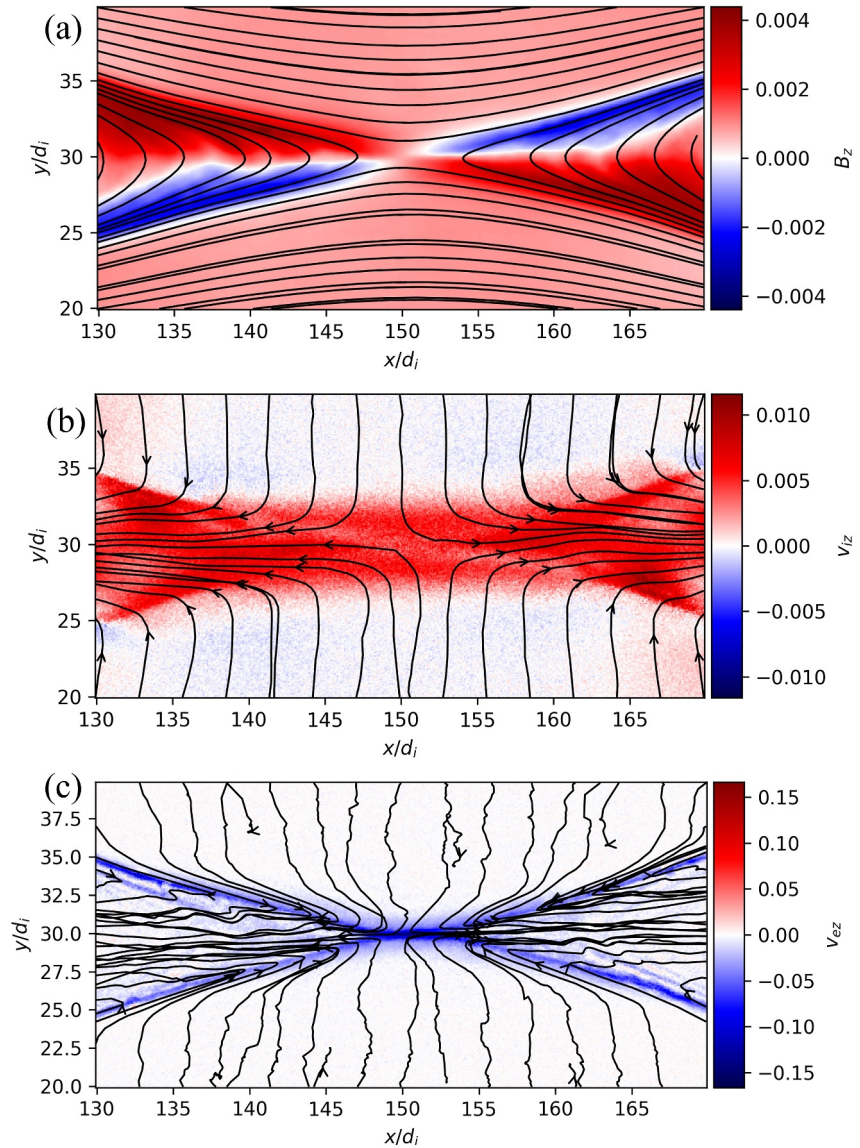
In Section 2, we discuss the parameters of the simulations used. Section 3 discusses the methods for clustering, including limitations in choice of variables for comparison with in situ data. Section 4 shows the results of the clustering including how robust these classifications are, and Section 5 evaluates the energy flux densities within the classes of regions to verify the methods compared to previous spacecraft observations. Section 6 discusses these results and ideas for application to spacecraft data. Section 7 concludes and looks to the benefits of these methods for future work.

## 2. Simulation

We use a 2.5-D PIC simulation using the iPic3D implicit PIC code, starting from a Harris equilibrium, similar to that described in Goldman et al. (2016). The boundary conditions are periodic in both  $x$  and  $y$  with two oppositely directed current sheets with initial perturbations to seed reconnection at  $(x, y) = (L_x/4, L_y/4)$  and  $(x, y) = (3L_x/4, 3L_y/4)$ , where  $L_{x,y}$  are the dimensions in the  $x$  and  $y$  directions respectively. Following these perturbations, the Harris current sheet 'tears' and the reconnecting current sheet forms.

The input parameters are the Harris-sheet thickness  $\Delta$ , the mass ratio  $m_i/m_e$ , the electron thermal velocity  $v_e/c$ , the electron-to-ion temperature ratio  $T_e/T_i$ , the background to Harris sheet density ratio  $n_b/n_0$ , and the ratio of guide field to asymptotic reconnecting field  $B_G/B_0$ , as given in Table 1. We use a grid spacing of  $1/12.8 d_{i0} = c/\omega_{i0}$  in both the  $x$  and  $y$  directions. We analyze simulation runs at a timestep of  $24,800 \omega_{ci} t$ , where  $\omega_{ci}$  is the ion cyclotron frequency. There are an average of 225 particles per cell of each species, and each particle has the same weight such that there are more particles per cell in the center of the Harris sheet and fewer in the exterior.

The simulation is cropped to contain a region of one reconnection site close to the EDR between 130 and 170  $d_{i0}$  in the  $x$  direction and 20 and 40  $d_{i0}$  in the  $y$  direction, as shown in Figure 2. Comparison to spacecraft data in the Earth's magnetotail using GSM coordinates requires  $x_{sim} = x_{GSM}$ ,  $y_{sim} = -z_{GSM}$ , and  $z_{sim} = y_{GSM}$ .



**Figure 2.** Overview of the 2.5-D PIC simulation used for this study. Here, we show the run with a guide field  $B_G = 0.1$ . The  $x$  and  $y$  components of quantities are given by black field/streamlines and the  $z$  component is shown in color for (a) the magnetic field  $\mathbf{B}$ , (b) the ion bulk velocity  $\mathbf{v}_i$ , and (c) the electron bulk velocity  $\mathbf{v}_e$ . Positive  $z$  components point out of the plane of the page. The central X-point is located close to  $(150, 30) d_i$ .

### 3. Methods

#### 3.1. $k$ -Means Clustering

A  $k$ -means algorithm clusters data by separating the samples given into  $k$  groups by minimizing the inertia of each group,

$$\sum_{i=0}^n \min_{j \leq k} (|x_i - \mu_j|^2), \quad (5)$$

where  $\mu_j$  are the means of the  $k$  clusters, often referred to as the centers of the clusters, and  $n$  is the total number of samples  $x_i$ . The inertia is a measure of how coherent each cluster is and how well the samples are described by

their associated cluster (Lloyd, 1982). To avoid the algorithm converging to a local (rather than global) minimum, the  $k$ -means++ initialization is used (Arthur & Vassilvitskii, 2006). This selects initial cluster centers using sampling based on a probability distribution relating to the contribution of the individual points to the overall inertia. It takes several initialisations of centers to check for convergence to local minima.

$k$ -means clustering methods have been applied elsewhere in space physics. These include classifying different types of solar wind (Heidrich-Meisner & Wimmer-Schweingruber, 2018; Roberts et al., 2020) and magnetopause crossings at Earth and Jupiter (Collier et al., 2020). The methods lend themselves to applications involving a distinct, known number of groups of points, and are often used due to their speed.

### 3.2. Silhouette Score

The performance of the clustering for a given value of  $k$  is evaluated using silhouette score  $s(x_i)$  for each data point  $x_i$  located in a cluster  $C_l$  (Rousseeuw, 1987),

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max(a(x_i), b(x_i))}, \quad (6)$$

where

$$a(x_i) = \frac{1}{N_{C_l} - 1} \sum_{j=0, j \neq i}^{N_{C_l}} |x_j - x_i|, \quad (7)$$

$$b(x_i) = \min_{C_j \neq C_l} \left( \frac{1}{N_{C_j}} \sum_{j=0, j \neq i}^{N_{C_j}} |x_j - x_i| \right), \quad (8)$$

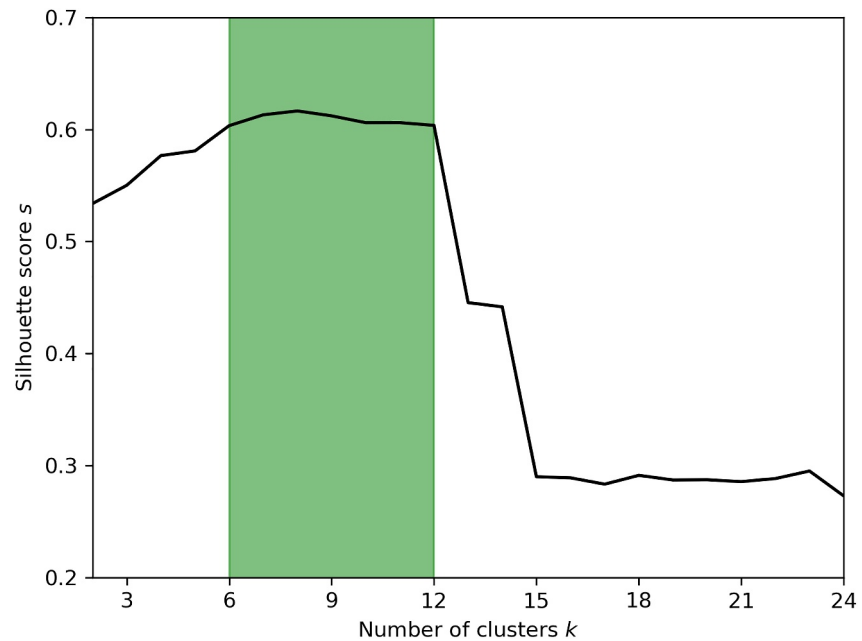
where  $C$  are clusters containing  $N_C$  points.  $a(x_i)$  represents the mean distance between a point and all of the other points in the cluster, and  $b(x_i)$  is the smallest mean distance to a point of all the points in any other cluster. Overall, this gives how similar each point is to its cluster compared to the others. To generalize this for an entire dataset, we take the mean over all data points  $s = \overline{s(x_i)}$  (Kaufman & Rousseeuw, 1990). A score  $s$  of close to one indicates appropriate clustering,  $s = 0$  indicates uncertainty in the clustering, and  $s = -1$  indicates incorrect clustering.

### 3.3. Data Normalization

Although we could classically normalize each variable relative to its distribution, here we seek to compare our simulation to spacecraft data. We normalize based on physical quantities such that all variables have comparable magnitudes.

We use the asymptotic reversing field  $B_0$  and the current sheet density  $n_0$  in the torn current sheet where reconnection is already occurring, as the Harris current sheet in the simulation has an artificially high density compared to the reconnecting current sheet. We calculate the Alfvén speed for both ions and electrons from  $B_0$  and  $n_0$  and use them to normalize quantities, with  $\mathbf{B}' = \mathbf{B}/B_0$ ,  $n'_s = n_s/n_0$ ,  $\mathbf{E}' = \mathbf{E}/B_0 V_{Ai0}$ ,  $\mathbf{V}'_i = \mathbf{V}_i/V_{Ai0}$ , and  $\mathbf{V}'_e = \mathbf{V}_e/V_{Ae0}$ . For the electric field, the ion Alfvén speed  $V_{Ai0}$  is used. The relevant species Alfvén speed is used to normalize the velocities to ensure the electron velocity components are of the same order as the ion velocity components.

For this study, we choose not to use any variables requiring higher moments of the velocity distribution function than the velocity for the clustering. This is due to the limitations of the distribution functions taken in situ by spacecraft. In regions of low density such as the magnetotail, very few particle counts are obtained, leading to high uncertainties in the corresponding moments of the distributions which become more problematic the higher the order of the moment becomes (Gershman et al., 2015).



**Figure 3.** The silhouette coefficient  $s$  for  $k$ -means clustering results carried out with each of  $k$  clusters between 2 and 24 for the simulation run with  $B_G = 0.1$ . The higher the value of the coefficient, the better the quality of the clusters. The highlighted region covers the peak values, between six and 12 clusters. The range of the coefficients for these results is 0.01, indicating that the clustering results are of similar quality.

## 4. Clustering Results

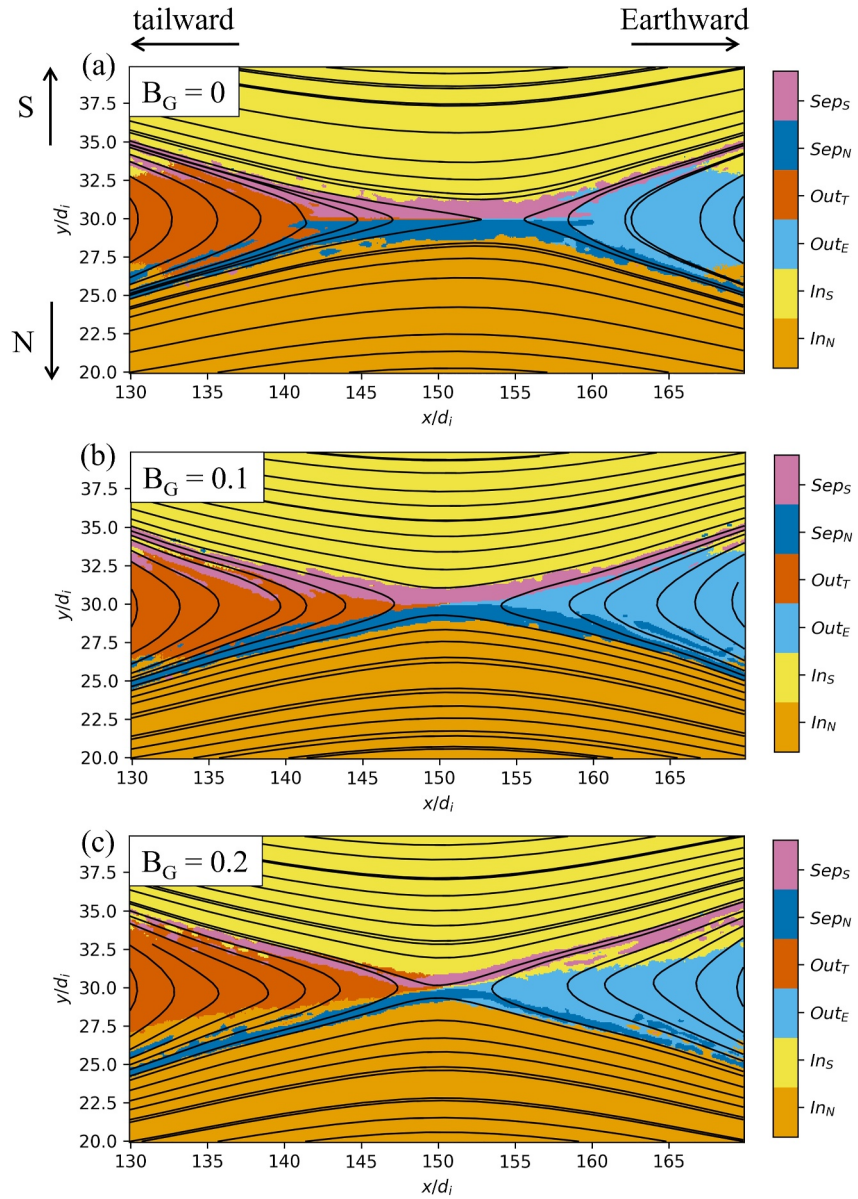
### 4.1. Silhouette Score

By carrying out the  $k$ -means clustering algorithm on the scaled simulation variables  $\mathbf{B}'$ ,  $\mathbf{E}'$ ,  $n'_i$ ,  $n'_e$ ,  $\mathbf{V}'_i$ , and  $\mathbf{V}'_e$  at each grid point for a range of numbers of clusters, we obtain the silhouette coefficient for each of these numbers (Figure 3).

The silhouette coefficient increases, reaching a maximum before decreasing rapidly for cluster numbers greater than 12, meaning that any number of clusters beyond this gives a much poorer clustering result. The coefficients differ very little between six and 12 clusters. As the number of clusters increases beyond six, we observe the separatrix-associated regions splitting into two, followed by splitting of the outflow regions into three regions each at 12 clusters. Beyond this, the inflows begin to split into smaller regions. Here we focus on the six cluster case and investigate how the structure identified changes between the simulation runs with differing guide field. However, the results with 12 clusters are provided in Figure S1 in Supporting Information S1.

### 4.2. Six Clusters

We begin with guide field  $B_G = 0.1$  due to this being most relevant to magnetotail reconnection due to the small but non-zero guide field (Figure 4b). We obtain three pairs of clusters - two inflow regions, two outflow regions, and two regions associated with pairs of separatrices. We define these regions relative to their equivalence in GSM coordinates to the field orientation in magnetotail reconnection; a northern inflow and separatrix-associated region,  $In_N$  and  $Sep_N$ , a southern inflow and separatrix-associated region,  $In_S$  and  $Sep_S$ , and Earthward and tailward outflows,  $Out_E$  and  $Out_T$ . Although there are only separatrix-associated regions due to the separatrices being a theoretical division in topology, we will still refer to these as ‘separatrix regions’ due to their spatial proximity to these expected divisions. The separatrix regions also incorporate other features, such as fast electron flows within low-density cavities near the separatrices (Cattell et al., 2005; Eastwood et al., 2018). The  $k$ -means clustering identifies these points as being close to each other within parameter space and belonging to the same cluster. Points may be identified as part of the ‘inflow’ regions past the separatrices, where we may expect them to be identified as ‘outflow’, as observed in Figure 2a. These points are close to each other in parameter space,



**Figure 4.** Results of carrying out  $k$ -means clustering with six clusters on comparable simulation runs with (a)  $B_G = 0$ , (b)  $B_G = 0.1$ , and (c)  $B_G = 0.2$ . Magnetic field lines are shown in black, with the color showing the regions identified by the  $k$ -means clustering. These regions are labeled in relation to the equivalent directions in GSM coordinates in the case of magnetotail reconnection. Each simulation run has  $k$ -means carried out independently with variables scaled in the same manner and subsequent clusters re-numbered for comparison between each case. The X-point may appear to be in a slightly different location in each due to the time-varying nature of the reconnection process causing it to drift as a result of plasmoids previously being produced.

indicating that the topological separation between regions by the separatrices does not cause a clear separation in the plasma and field parameters. We discuss this further in Section 6.

For the  $B_G = 0$  case (Figure 4a), X-point drift (due to previous plasmoids which have formed in this run of the simulation) causes some additional structuring of the separatrix regions on the tailward side of the reconnection site. Excluding these structures, the separatrix regions are overall symmetric and have a different shape to in the  $B_G = 0.1$  case. These regions are spatially much wider around the X-point, but thin with distance along the outflows. The outflow regions begin to narrow around  $x = 133$  and  $167 d_i$ , closer to the X-point than in the



$B_G = 0.1$  case. This leads to more of the ‘inflow’ region being identified within the area bounded by the separatrices.

Significant changes appear in the  $B_G = 0.2$  case (Figure 4c). A global feature is a tilt of the reconnection site; anti-clockwise for the separatrix regions and clockwise for the outflow regions, corresponding to the opposite charge of the particle populations which dominate in these clusters, although significantly smaller for the separatrix regions than outflows. This tilt is comparable to previous studies of the effect of a guide field on the outflows of reconnection (e.g., Eastwood et al., 2010; Hesse et al., 2002). There are again points labeled as ‘inflow’ bounded by the separatrices. In this run, this is due to the guide field tilt, not the shape of the outflows. The separatrix regions are not continuous, with ‘gaps’ present in the spatial structure formed by these points. This happens where the out-of-plane Hall magnetic field component is oppositely directed to the guide field and they partially ‘cancel out’. The separatrix regions are significantly thinner with higher guide field.

Overall, the six clusters correspond to regions which we can already identify in reconnection sites. These are comparable between the three simulation runs, and we can identify changes in the structure as a result of the varying guide field. However, we may expect to observe a region corresponding to the scales at which electron-scale physics occurs around the X-point, forming the electron diffusion region (EDR). This is not identified by the clustering. We would expect the electron inertial length  $d_e$  to be 16 times smaller than the ion inertial length  $d_i$  with the mass ratio of 256. This results in  $d_e$  being comparable to the simulation grid size. This may result in the moments used not capturing the physics occurring at these small scales. We may observe some characteristic features of this region if we were to use higher order moments or the particle velocity distributions themselves - however, as previously discussed, this is not possible in this study.

### 4.3. Robust Clustering

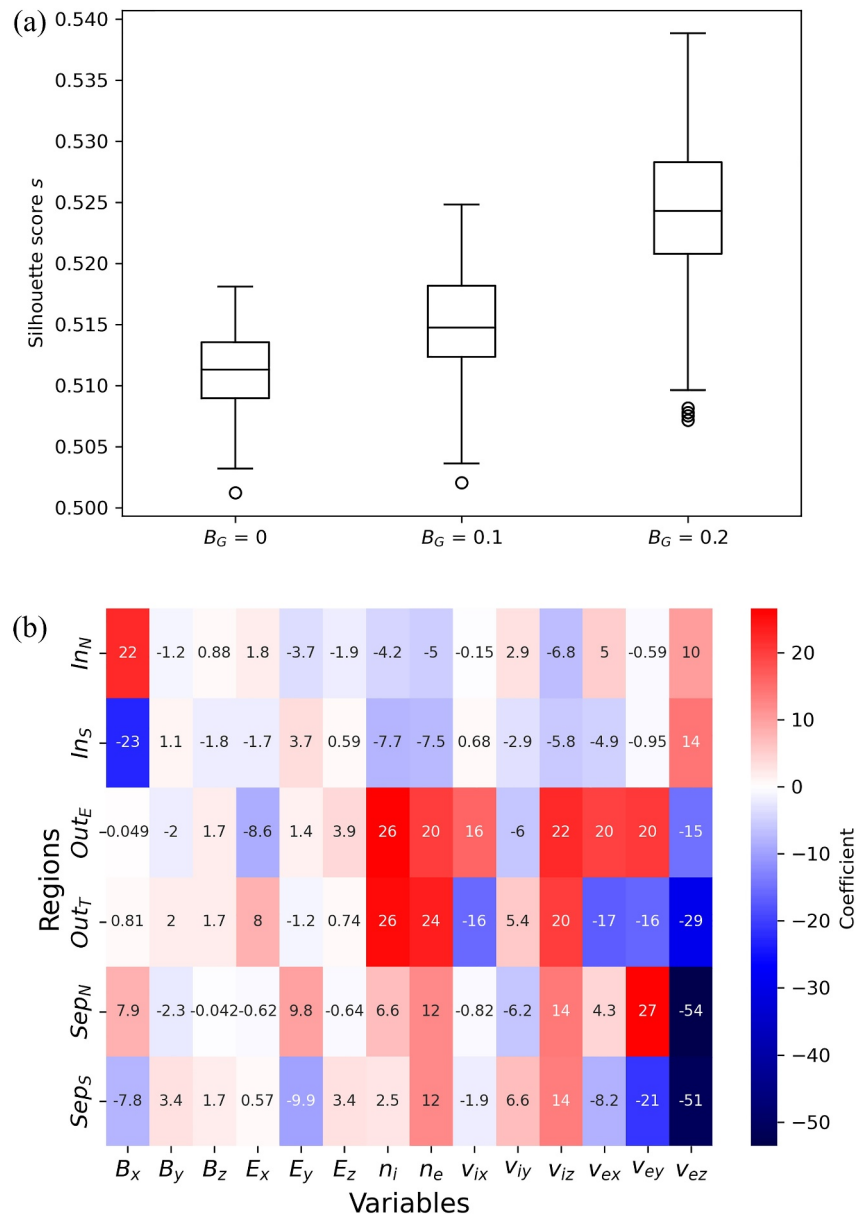
To evaluate how robust the  $k$ -means clustering is for all three of the simulation runs, we add random noise sampled from a normal distribution up to a magnitude of 0.1 in scaled units to all points in the simulation before carrying out the clustering and calculating the silhouette score for the resulting clusters. This is carried out for 100 iterations on each of the simulation runs, with the resulting distributions of silhouette score shown in Figure 5a.

We first notice that the value of the silhouette score is decreased from that shown in Figure 3a due to the effect of the added noise on the quality of the clusters. As silhouette score evaluates how close each of the points in a cluster is adding random noise of similar magnitude to all points will increase the variance of each individual cluster, reducing the score. Overall for each of the cases the spread in silhouette score is small - the interquartile range is 0.00461, 0.00582 and 0.00750 for  $B_G = 0, 0.1$  and  $0.2$  respectively. This indicates that the clustering is only affected by a small amount when the values of the variables provided vary, and the clusters vary very little in quality.

We can also investigate the contribution of each of the variables to distinguishing between regions using linear discriminant analysis (LDA). This is a supervised dimensionality reduction technique, as opposed to an unsupervised dimensionality reduction technique such as principal component analysis (PCA). It considers the classification of points and tells us how important the contribution of each of the variables is to distinguishing between classes (Hastie et al., 2009). Here we evaluate the contributions for the  $B_G = 0.1$  case. For the regions shown in Figure 4b, we obtain contributions as shown in Figure 5b.

Overall, variables either have a very large contribution, or little contribution at all to a region. The  $B_y$  and  $B_z$  components of the magnetic field make very little contribution to any of the regions. However,  $B_x$  is very important for identifying the inflow regions. The outflow regions are mostly defined by the number densities and velocities for both particle species, and out-of-plane electron velocity  $v_{ez}$  is dominant over any other variable for defining the separatrices. These clear contributions of the variables cause a high certainty in the clustering for all three simulation runs and indicate that the clusters are formed due to large differences in the variables provided.

However, it is important to remember that the  $k$ -means clustering is not physically informed. This is purely a statistical clustering associating points which are close to each other in parameter space, and there is no relation between the clusterings of each simulation run to maintain comparable structures. When analyzing properties of the resulting clusters we must take this into account and draw comparisons between statistically similar regions which relate to, but are not identical to, the structure we would expect to find in magnetic reconnection. However, the method may be applied for analysis of other comparable data.



**Figure 5.** (a) Box plots showing the median, upper and lower quartiles, extremes and outliers of silhouette score for 100 iterations of clustering with added noise for each of the simulation runs with varying guide field. (b) A heatmap showing the coefficients (weights) of the variables for the linear discriminants as a result of linear discriminant analysis for their contributions to distinguishing between each of the classes of region. The larger the magnitude, the higher the contribution of that variable. The sign indicates whether it is an increase or decrease in the variable which causes a point to be more likely to be classed in that region.

## 5. Energy Flux Densities

The identification of inflow, outflow, and separatrix regions by the  $k$ -means clustering provides an opportunity to study the properties of these regions in more detail. To explore this, we now use the clustering as a tool to explore the behavior of the energy flux densities within the different regions of these simulations. As the regions are comparable between the three varying guide field cases, we can look at how the distributions of the energy flux densities change. As we are comparing between simulation runs, we keep the energy flux densities in unscaled simulation units.

### 5.1. Energy Flux Magnitudes

We first analyze the magnitudes of the energy flux densities to give an overview of the contributions of each region. The median magnitudes of each of the energy flux densities in each region with associated guide field are given in Table 2. The overall contribution from these energy flux densities varies with guide field, and the shape of the distributions may vary, shown in Figure 6. We combine the pairs of regions to show the magnitudes as the distributions vary very little between the two in each case.

For the outflow regions, the median ion kinetic energy flux density  $|\mathbf{K}_i|$  decreases with guide field (Figure 6b). The median ion enthalpy flux density  $|\mathbf{H}_i|$  (Figure 6e) decreases more than the kinetic energy flux density. In both cases, the spread of the distribution increases with guide field. For the electrons, the kinetic energy flux density  $|\mathbf{K}_e|$  distribution (Figure 6h) narrows and becomes more positively skewed, although the median remains approximately constant. For the electron enthalpy flux density  $|\mathbf{H}_e|$  (Figure 6k), the median decreases and the distribution becomes more positively skewed. As we would expect due to an increased guide field providing an enhanced out-of-plane magnetic field component, the magnitude of the Poynting flux density vector  $|\mathbf{S}|$  (Figure 6n) increases in both outflows.

In the separatrix-associated regions, we see very little change in the magnitude of the ion kinetic energy flux density  $|\mathbf{K}_i|$  (Figure 6c). The magnitude of the ion enthalpy flux density  $|\mathbf{H}_i|$  (Figure 6f) decreases for the  $B_G = 0.2$  case, but there is little change between the 0 and 0.1 cases. The distribution of electron kinetic energy flux density  $|\mathbf{K}_e|$  (Figure 6i) narrows with increasing guide field, but the median remains approximately constant. There is also little change to  $|\mathbf{H}_e|$  (Figure 6l), and again the magnitude of the Poynting flux density vector  $|\mathbf{S}|$  (Figure 6o) increases, as expected.

When we compare these magnitudes between the two pairs of regions for the outflows and separatrices, we observe that the magnitude of the ion kinetic energy flux density  $|\mathbf{K}_i|$  is approximately four times larger in the outflow than in the separatrices, and that of the ion enthalpy flux density  $|\mathbf{H}_i|$  is approximately three times larger. However, the magnitude of the electron kinetic energy flux density  $|\mathbf{K}_e|$  is almost an order of magnitude smaller in the outflows than in the separatrices. Electron enthalpy flux density  $|\mathbf{H}_e|$  is of the same magnitude in both. Although we observe an increase in both sets of regions in the Poynting flux, the overall magnitude is twice as large in the separatrices than in the outflows.

Overall, this suggests that the guide field causes a shift in energy repartition to decrease the outgoing ion and electron energy. Comparison between the outflows and separatrix-associated regions shows the overall importance of the outflows for transport of ion energy and of the separatrices for transport of electron energy. Although we associate the separatrices with large fields, in the Poynting flux we only see this being twice as large in the separatrix-associated regions as we do in the outflows, suggesting the outflows are not insignificant for the flow of electromagnetic energy.

Through analyzing the medians and distributions of the energy fluxes in each of the clusters, we are able to identify changes due to the varying guide field. We now investigate where the different energy fluxes would be transported in a larger system by studying their orientation.

### 5.2. Energy Flux Orientations

We analyze the components of the energy flux densities to investigate the relative contributions of the separatrices and outflows. This is used to evaluate the energy output from symmetric reconnection and how this varies due to guide field. For completeness, the distributions of energy flux densities in the inflows are shown in Figure S2 in Supporting Information S1. We first look at the effect on the outflows, as given in Figure 7.

The  $x$  component of the ion kinetic energy flux density  $K_{ix}$  (Figure 7a) decreases in the respective direction to be carried away from the X-point (positive for the ‘Earthward’ region  $Out_E$ , and negative for the ‘tailward’ region  $Out_T$ ) as the guide field increases.  $K_{iz}$  also decreases with guide field. For the ion enthalpy flux density (Figure 7b), the  $x$  component  $H_{ix}$  also decreases in the respective outward direction, and the distribution becomes more bimodal.  $H_{iz}$  also decreases and becomes more bimodal. Turning to the electrons,  $K_{ex}$  (Figure 7c) retains a similar median with changing guide field, but the distribution narrows overall as guide field increases. This is similar for  $K_{ez}$ . For the electron enthalpy flux density (Figure 7d),  $H_{ex}$  decreases in the outward direction, but  $H_{ey}$  increases in opposite directions for each cluster. Finally, for the components of the Poynting flux density vector

**Table 2**  
Median Magnitudes of Energy Flux Densities

	$In_N$			$In_S$			$Out_E$			$Out_T$			$Sep_N$			$Sep_S$		
	0	0.1	0.2	0	0.1	0.2	0	0.1	0.2	0	0.1	0.2	0	0.1	0.2	0	0.1	0.2
$B_G$	0	0.1	0.2	0	0.1	0.2	0	0.1	0.2	0	0.1	0.2	0	0.1	0.2	0	0.1	0.2
$ K_i $	0.964	1.00	1.39	0.992	1.00	1.40	94.9	96.9	89.3	111	103	89.0	13.0	13.7	19.2	13.6	15.4	21.2
$ H_i $	9.08	9.44	10.4	9.14	9.37	10.4	230	221	198	262	234	198	23.6	15.7	45.4	24.7	28.2	52.0
$ K_e $	0.0114	0.0115	0.0132	0.0116	0.0115	0.0132	1.61	1.97	1.44	1.75	2.14	1.48	7.47	6.55	8.38	7.24	6.47	7.15
$ H_e $	2.04	2.05	2.14	2.07	2.05	2.16	113	110	72.3	125	114	74.9	40.0	40.2	29.7	37.5	47.4	28.2
$ S $	123	126	151	121	127	150	130	117	265	108	111	263	360	437	627	368	471	610

Note. All values are scaled relative to  $|K_i|$  for  $B_G = 0.1$  in the region  $In_N$ .

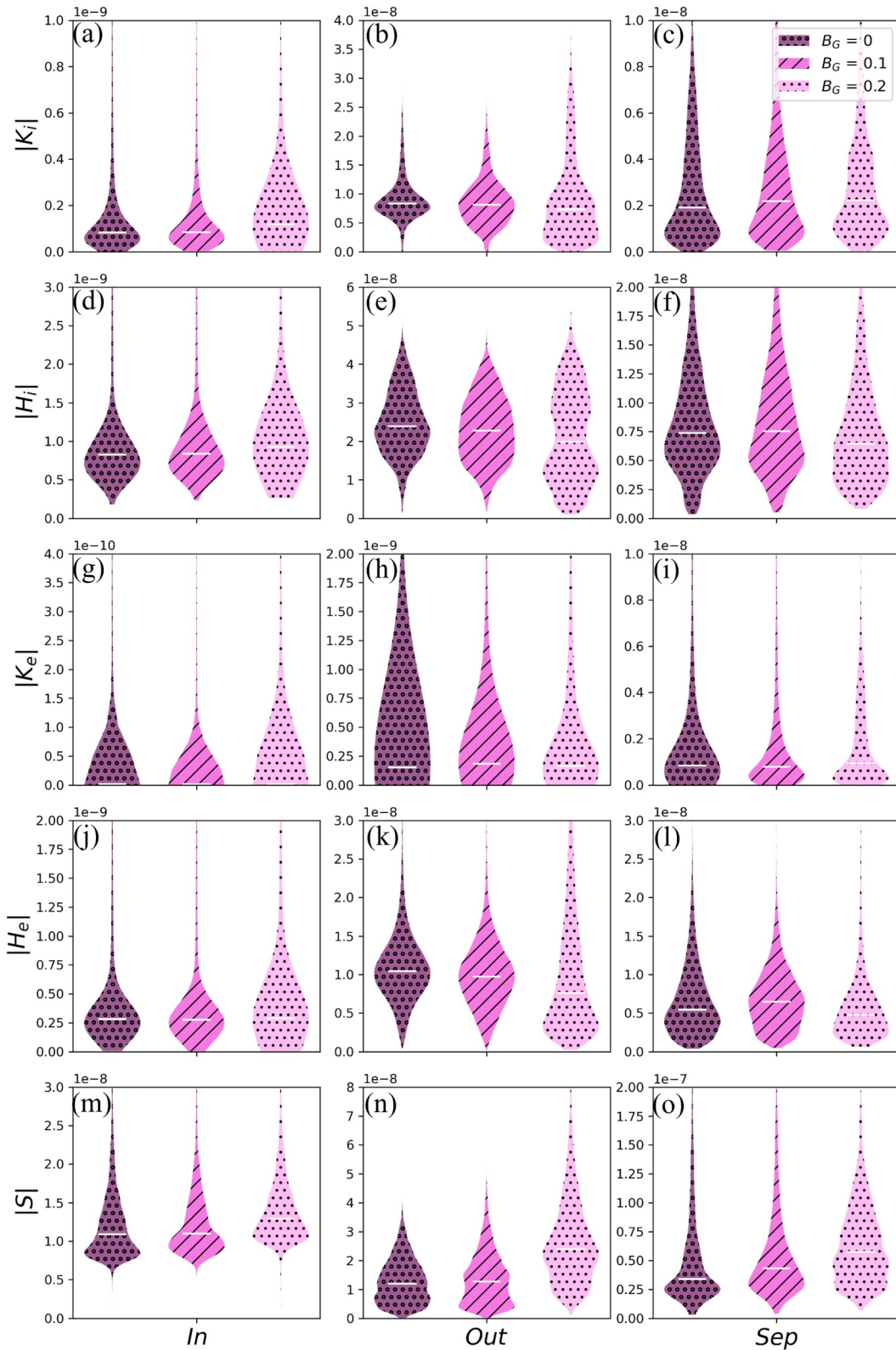


Figure 6.

(Figure 7e),  $S_x$  increases in the outward direction, and  $S_z$  becomes more negative, although this is slightly counteracted by the positive guide field in this direction.

The same quantities for the separatrices are given in Figure 8. Starting with the ions again, the ion kinetic energy flux density (Figure 8a) has a median in the  $K_{ix}$  component directed away from the X-point for the  $B_G = 0.2$  case. However, this is approximately zero for the  $B_G = 0$  and 0.1 cases, with the bulk of the distribution around zero for all cases. There is an additional tail to the distribution for  $B_G = 0.2$  causing this shift in the median.  $K_{iy}$  (Figure 8b) increases in the direction toward the X-point with guide field. For the ion enthalpy flux density,  $H_{ix}$  has the same trend in the median as for  $K_{ix}$ , although the distributions are more bimodal.  $H_{iy}$  decreases in the direction toward the X-point, and  $H_{iz}$  overall decreases - this is opposite in direction to the guide field. For the electron kinetic energy flux density (Figure 8c), the distributions of  $K_{ex}$  and  $K_{ez}$  narrow with increasing guide field, but the medians do not shift. For the electron enthalpy flux density (Figure 8d),  $H_{ez}$  becomes less negative, increasing in the direction of the guide field. The main changes are observed in the Poynting flux density vector (Figure 8e) -  $S_x$  becomes less bimodal as the guide field increases,  $S_y$  increases in the direction toward the X-point, and  $S_z$  becomes more negative, opposing the guide field imposed.

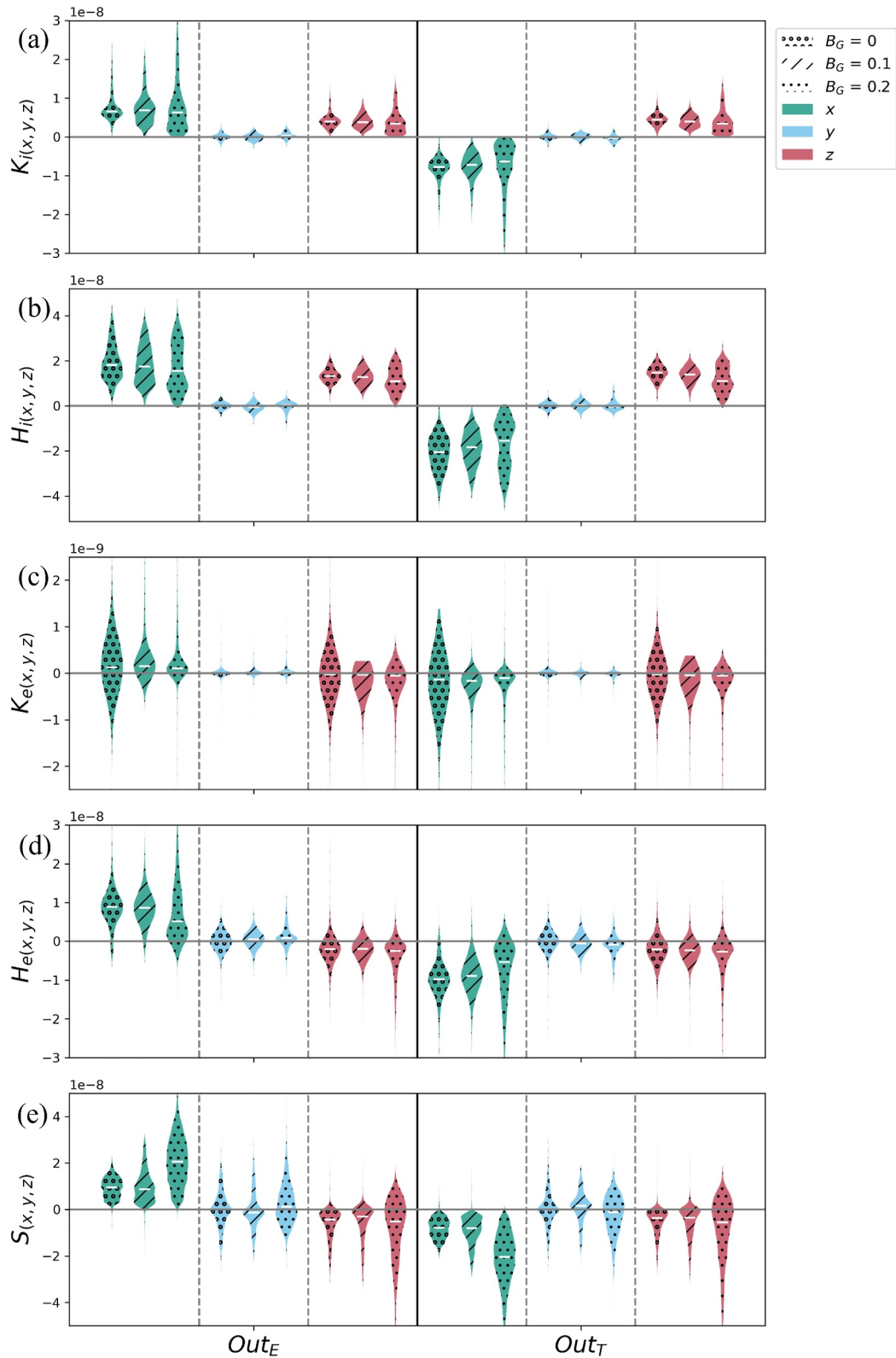
Comparing the behaviors,  $K_{ix}$  can be up the three times as large in the outflow as in the separatrices.  $K_{iz}$  is of similar magnitude in both, but very slightly larger in the outflows.  $K_{ex}$  is an order of magnitude larger in the separatrices than in the outflows, and  $K_{ez}$  is of a similar magnitude in both.  $H_{ex}$  is similar between both, but  $H_{ey}$  is slightly larger in the outflow, although the distribution is overall shifted further from zero in the separatrices. For the Poynting flux density vector,  $S_x$  is approximately twice as large in the separatrices as in the outflows.  $S_y$  is similar between both, and  $S_z$  is again approximately twice as large in the separatrices.

Overall, this indicates that particle energy flux densities directed away from the X-point in the outflows decrease as guide field increases. Out-of-plane particle energy flux density decreases as guide field increases. The outward Poynting flux density component increases with guide field. In the separatrices, the highest guide field case causes the ion kinetic energy flux density to be overall directed away from the X-point in the  $x$  direction, although this does not form a trend with the other guide field cases. In comparison, the electron kinetic energy flux density component distributions narrow with increasing guide field. For both particle populations, there is a change in the enthalpy flux density in the out of plane  $z$  direction as guide field decreases - an increase in the direction of the guide field for the electrons, and a decrease in the direction of the guide field for the ions. However, in the reconnection plane, there is little variation. The main change in the Poynting flux density in the separatrices is the out-of-plane component increasing oppositely to the direction of the guide field imposed. Comparing the roles of the two pairs of regions from this, the outflows are responsible mostly for carrying ion energy away from the X-point, whereas the separatrices are responsible for electron energy. This corresponds to what we would expect due to the small scale of field reversal associated with the separatrices being comparable to the electron scales. The separatrices also contribute to the transport of large amounts of Poynting flux in the direction of the exhausts and out-of-plane.

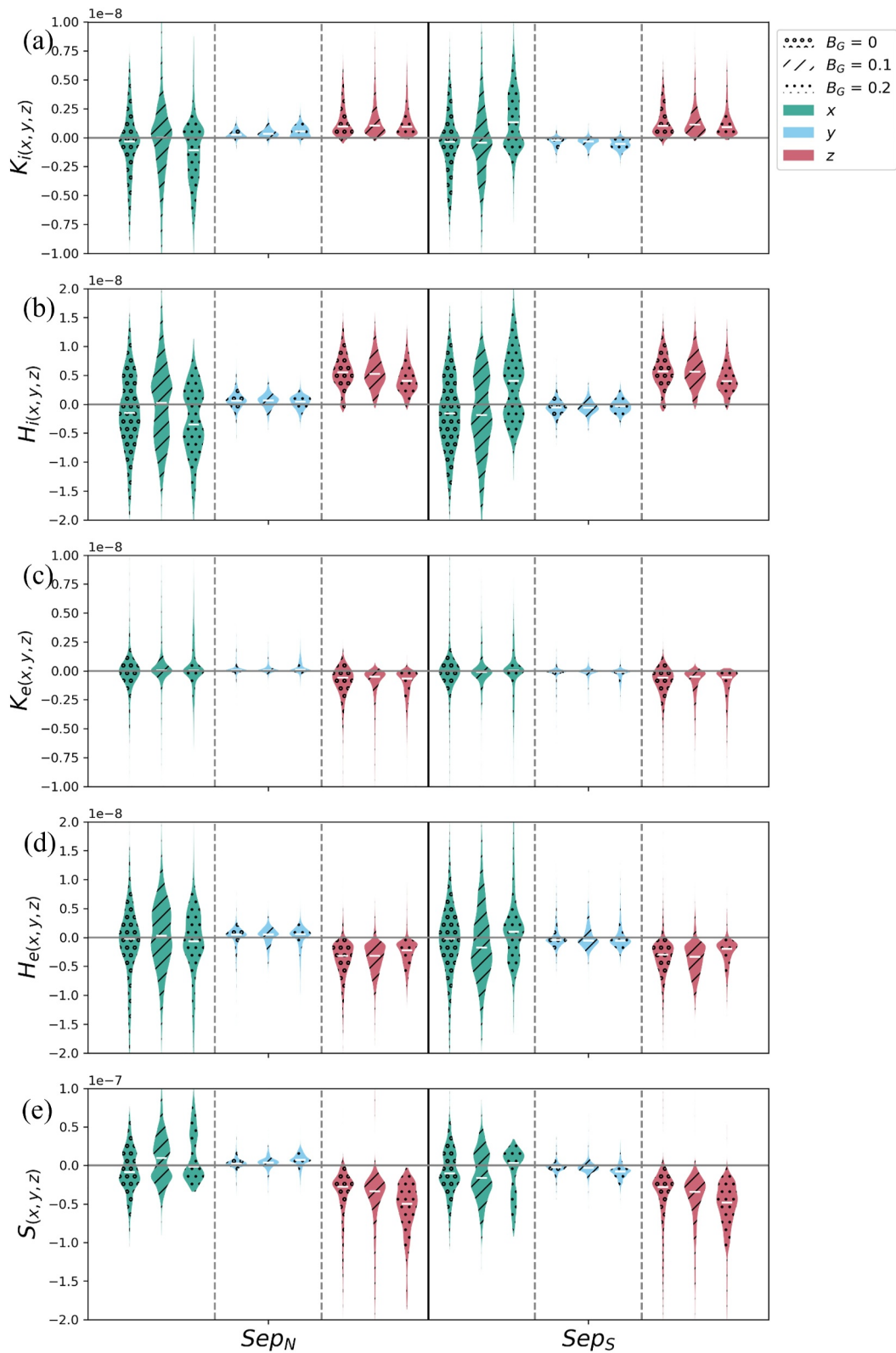
## 6. Discussion

In this work, we use  $k$ -means clustering to identify regions within a 2.5-D PIC simulation of magnetic reconnection. We verify this method by analyzing the energy flux densities within these clusters and find that the relationships between the energy flux densities are comparable to those as identified in studies using spacecraft data, with ion enthalpy flux being the overall dominant outflowing energy flux density, but with Poynting flux density dominant at some points (Eastwood et al., 2013). Many of our observations are consistent with existing literature, such as significant out-of-plane electron kinetic energy flux density close to the X-point (Eastwood et al., 2020) and in the separatrices (Lapenta et al., 2015), and the tilt of the outflows due to the guide field

**Figure 6.** The distributions of the magnitudes of each of the energy flux densities given for each guide field run in each of the clusters as violin plots, combining each pair of clusters. The median for each is given by a horizontal white bar. The scale for the  $y$  axis varies between panels. Panels a–c give the magnitude of the ion kinetic energy flux density  $|K_i|$ , panels d–f the magnitude of the ion enthalpy flux density  $|H_i|$ , panels g–i the magnitude of the electron kinetic energy flux density  $|K_e|$ , panels j–l the magnitude of the electron enthalpy flux density  $|H_e|$ , and panels m–o the magnitude of the Poynting flux density vector  $|S|$ . Panels a, d, g, j and m give these for the inflow regions, b, e, h, k and n for the outflow regions, and c, f, i, l and o for the separatrix-associated regions. These magnitudes are given in unscaled simulation units.



**Figure 7.** The distributions of the components of each of the energy flux densities given for each guide field run in each of the outflow clusters  $Out_E$  and  $Out_T$  as violin plots. The median for each is given by a horizontal white bar. Panel a shows the components of the ion kinetic energy flux density  $K_i$ , b the ion enthalpy flux density  $H_i$ , c the electron kinetic energy flux density  $K_e$ , d the electron enthalpy flux density  $H_e$ , and e the Poynting flux density vector  $S$ . Components here are given in simulation coordinates, with each of the energy flux densities in unscaled simulation units.



**Figure 8.** The distributions of the components of each of the energy flux densities given for each guide field run in each of the separatrix-associated clusters  $Sep_N$  and  $Sep_S$  as violin plots, presented in a similar manner to Figure 7.



(Eastwood et al., 2010; Wilder et al., 2017). Poynting flux due to the guide field flows through the reconnection site largely unchanged (Birn & Hesse, 2010).

Through our analysis of the clusters, we are also able to identify interesting differences to what we may expect. Although by definition the separatrices are the boundaries between the inflow and outflow regions of a reconnection site, the  $k$ -means identifies points statistically similar to the typical inflow region within the regions bounded by the separatrices that would traditionally be considered outflows. As further investigation shows that the probabilities of these points belonging to these regions are very high, we can infer that the separatrices do not fully prevent the two plasma populations merging. However, this could be investigated further by carrying out the same clustering techniques at different time steps of the simulation to see how the structure evolves and whether this is due to the reconnection not having reached a steady state.

As the outward Poynting flux density increases with guide field and the direction of the outward particle energy flux densities changes to reduce the component in the direction of the outflows, the guide field may change how the reconnection process affects other processes in a system such as a magnetosphere. The increase in  $y$  components of particle energy flux densities in simulation coordinates would correspond to an increase in  $z$  components in GSM, leading to acceleration of particles along field lines toward the poles. As  $z$  components of particle energy flux densities in the simulation decrease with guide field, this reduces the out-of-plane energy transport by particles, which would be across the tail in the  $y$  direction in GSM.

The main benefit of a machine learning approach to identifying regions is that we can classify many datapoints in a short amount of time using these methods. This facilitates statistical studies of phenomena, such as magnetic reconnection, where we have large quantities of spacecraft data obtained (Baker et al., 2016), much beyond those a human can easily survey themselves. With classification by human input, we also face the issue of biases which cannot be easily quantified. Although machine learning techniques are not unbiased, it is possible to quantify and understand the biases and the impacts that they may have on the results obtained.

We have shown that the  $k$ -means clustering can identify regions analogous to those we may already identify by eye in magnetic reconnection. Through further analysis, we show that these clusters have distinct differences in the plasma and field variables used to describe them and that each of the points in the clusters are identified with high probability. We study how properties of the clusters change and compare clusters between simulation runs. The clustering is also robust to noise being added to the variables, making it suitable for application to spacecraft data as well as simulation data.

The benefits of  $k$ -means clustering include it being simple to implement, efficient, and capable of handling data with many dimensions. It can take any number of variables  $n$ , suitably scaled, and determine  $k$  statistically different clusters based on distance in the  $n$ -dimensional parameter space. Although here we show a specific application to PIC simulations of magnetic reconnection, this is a simple and widely used machine learning method which could be applied to many simulations or to data from various spacecraft. As we do not provide any spatial variables, only those related to the plasma or field, the only necessary criterion is that there are enough data points to form a significant population for the  $k$ -means.

## 7. Conclusions

In this work, we carried out  $k$ -means clustering on a 2.5-D PIC simulation of symmetric magnetic reconnection with varying guide field. We identify three pairs of regions - inflows, outflows, and separatrices. These clusters are analogous to the regions identified in previous studies of reconnection. The clusters are identified with high certainty and each may be defined by different combinations of field and plasma parameters. We find that the overall behavior of the energy flux densities within these regions agrees with previous studies. Ion enthalpy flux is overall dominant in the outflows, with Poynting flux also being significant at some points. We observe significant out-of-plane electron kinetic energy flux close to the X-point and in the separatrices, and the Poynting flux due to the guide field flows through the reconnection site largely unchanged. The guide field increases the outward Poynting flux density, and the direction of outward particle energy flux densities change to reduce the component in the direction of the outflows, impacting further phenomena as a result of the energy flow around a system such as Earth's magnetosphere.

As  $k$ -means clustering is both simple to implement, efficient, and capable of handling data with high dimensionality, the method of analysis presented here is not limited to only reconnection or to solely simulations.

Although it could be argued that comparison is possible through human observation, biases introduced by clustering algorithms can be well understood whereas those introduced by human input present greater difficulty. The simplicity of the implementation would allow for comparison of applications in different cases, including direct comparison between simulation and spacecraft data. This also facilitates identification of regions in multiple simulations, or in the case of spacecraft data, multiple events, leading to statistical studies of phenomena. This issue will only grow as we look toward further multi-spacecraft missions such as Helioswarm (Klein et al., 2023), planned to launch in 2029, and Plasma Observatory (Retinò et al., 2022), currently under consideration by ESA. Although  $k$ -means is a simple method, the ability to classify large volumes of data will only become more important as we head toward further multi-point missions and increases the possible science output of such missions. Here, we apply the technique to only three simulation runs for magnetic reconnection, but  $k$ -means may be applied to any amount of simulation or spacecraft data with any number of variables given as input to identify statistically different regions.

### Data Availability Statement

Simulation data used to obtain figures and tables is available in the Zenodo data repository Waters et al. (2024), <https://doi.org/10.5281/zenodo.12570878>.

### References

- Arthur, D., & Vassilvitskii, S. (2006).  $k$ -means++: The advantages of careful seeding (Technical Report). In *Stanford infolab*. Retrieved from <http://ilpubs.stanford.edu:8090/778/>
- Aunai, N., Belmont, G., & Smets, R. (2011). Energy budgets in collisionless magnetic reconnection: Ion heating and bulk acceleration. *Physics of Plasmas*, 18(12), 122901. <https://doi.org/10.1063/1.3664320>
- Axford, W. I. (1984). Magnetic field reconnection. In E. W. Hones Jr (Ed.), *Magnetic reconnection in space and laboratory plasmas* (pp. 1–8). American Geophysical Union.
- Baker, D. N., Riesberg, L., Pankratz, C. K., Panneton, R. S., Giles, B. L., Wilder, F. D., & Ergun, R. E. (2016). Magnetospheric multiscale instrument suite operations and data system. *Space Science Reviews*, 199(1–4), 545–575. <https://doi.org/10.1007/s11214-014-0128-5>
- Birn, J., & Hesse, M. (2005). Energy release and conversion by reconnection in the magnetotail. *Annals of Geophysics*, 23(10), 3365–3373. <https://doi.org/10.5194/angeo-23-3365-2005>
- Birn, J., & Hesse, M. (2010). Energy release and transfer in guide field reconnection. *Physics of Plasmas*, 17(1), 012109. <https://doi.org/10.1063/1.3299388>
- Breuilleard, H., Dupuis, R., Retino, A., Le Contel, O., Amaya, J., & Lapenta, G. (2020). Automatic classification of plasma regions in near-earth space with supervised machine learning: Application to magnetospheric multi scale 2016–2019 observations. *Frontiers Astronomy Space Science*, 7. <https://doi.org/10.3389/fspas.2020.00055>
- Burch, J. L., Torbert, R. B., Phan, T. D., Chen, L.-J., Moore, T. E., Ergun, R. E., et al. (2016). Electron-scale measurements of magnetic reconnection in space. *Science*, 352(6290), aaf2939. <https://doi.org/10.1126/science.aaf2939>
- Camporeale, E., Carè, A., & Borovsky, J. E. (2017). Classification of solar wind with machine learning. *Journal of Geophysical Research: Space Physics*, 122(11). <https://doi.org/10.1002/2017JA024383>
- Cattell, C., Dombeck, J., Wygant, J., Drake, J. F., Swisdak, M., Goldstein, M. L., et al. (2005). Cluster observations of electron holes in association with magnetotail reconnection and comparison to simulations. *Journal of Geophysical Research: Space Physics*, 110(A01211). <https://doi.org/10.1029/2004JA010519>
- Collier, M. R., Gruesbeck, J. R., Connerney, J. E. P., Joy, S. P., Hospodarsky, G. B., Roberts, A., et al. (2020). A  $k$ -means clustering analysis of the jovian and terrestrial magnetopauses: A technique to classify global magnetospheric behavior. *Journal of Geophysical Research: Planets*, 125(9), e2019JE006366. <https://doi.org/10.1029/2019JE006366>
- Denton, R. E., Liu, Y., Agudelo Rueda, J. A., Genestreti, J. K., Hasegawa, H., Hosner, M., et al. (2024). Determining the orientation of a magnetic reconnection  $x$  line and implications for a 2d coordinate system. *Journal of Geophysical Research: Space Physics*, 129(1), e2023JA032167. <https://doi.org/10.1029/2023JA032167>
- Denton, R. E., Torbert, R. B., Hasegawa, H., Dors, I., Genestreti, K. J., Argall, M. R., et al. (2020). Polynomial reconstruction of the reconnection magnetic field observed by multiple spacecraft. *Journal of Geophysical Research: Space Physics*, 125(2), e2019JA027481. <https://doi.org/10.1029/2019JA027481>
- Eastwood, J. P., Goldman, M. V., Phan, T. D., Stawarz, J. E., Cassak, P. A., Drake, J. F., et al. (2020). Energy flux densities near the electron dissipation region in asymmetric magnetopause reconnection. *Physical Review Letters*, 125(26), 265102. <https://doi.org/10.1103/PhysRevLett.125.265102>
- Eastwood, J. P., Mistry, R., Phan, T. D., Schwartz, S. J., Ergun, R. E., Drake, J. F., et al. (2018). Guide field reconnection: Exhaust structure and heating. *Geophysical Research Letters*, 45(10), 4569–4577. <https://doi.org/10.1029/2018GL077670>
- Eastwood, J. P., Phan, T. D., Drake, J. F., Shay, M. A., Borg, A. L., Lavraud, B., & Taylor, M. G. G. T. (2013). Energy partition in magnetic reconnection in earth's magnetotail. *Physical Review Letters*, 110(22), 225001. <https://doi.org/10.1103/PhysRevLett.110.225001>
- Eastwood, J. P., Shay, M. A., Phan, T. D., & Øieroset, M. (2010). Asymmetry of the ion diffusion region hall electric and magnetic fields during guide field reconnection: Observations and comparison with simulations. *Physical Review Letters*, 104(20), 205001. <https://doi.org/10.1103/PhysRevLett.104.205001>
- Gershman, D. J., Dorelli, J. C., F.-Viñas, A., & Pollock, J. C. (2015). The calculation of moment uncertainties from velocity distribution functions with random errors. *Journal of Geophysical Research: Space Physics*, 120, 6633–6645. <https://doi.org/10.1002/2014JA020775>
- Goldman, M. V., Newman, D. L., & Lapenta, G. (2016). What can we learn about magnetotail reconnection from 2d PIC harris-sheet simulations? *Space Science Reviews*, 199(1), 651–688. <https://doi.org/10.1007/s11214-015-0154-y>

### Acknowledgments

We would like to acknowledge the contributions of Giovanni Lapenta and thank him for his insightful discussions throughout the duration of this work prior to his untimely passing in the spring of 2024. Gianni was a dear friend and mentor, offering wisdom and kindness to students and colleagues alike. CLW is supported by UKRI/STFC studentship ST/X508433/1. JPE and NF are supported by UKRI/STFC grant ST/W001071/1.

- Hasegawa, H., Denton, R. E., Nakamura, R., Genestreti, K. J., Nakamura, T. K. M., Hwang, K.-J., et al. (2019). Reconstruction of the electron diffusion region of magnetotail reconnection seen by the mms spacecraft on 11 July 2017. *Journal of Geophysical Research: Space Physics*, 124(1), 122–138. <https://doi.org/10.1029/2018JA026051>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Heidrich-Meisner, V., & Wimmer-Schweingruber, R. F. (2018). Solar wind classification via k-means clustering algorithm. In E. Camporeale, S. Wing, & J. R. Johnson (Eds.), *Machine learning techniques for space weather* (pp. 397–424). Elsevier. <https://doi.org/10.1016/B978-0-12-811788-0.00016-0>
- Hesse, M., Aunai, N., Zenitani, S., Kuznetsova, M., & Birn, J. (2013). Aspects of collisionless magnetic reconnection in asymmetric systems. *Physics of Plasmas*, 20(6), 061210. <https://doi.org/10.1063/1.4811467>
- Hesse, M., Kuznetsova, M., & Hoshino, M. (2002). The structure of the dissipation region for component reconnection: Particle simulations. *Geophysical Research Letters*, 29(12). <https://doi.org/10.1029/2001GL014714>
- Hesse, M., Norgren, C., Tenfjord, P., Burch, J. L., Liu, Y.-H., Chen, L.-J., et al. (2018). On the role of separatrix instabilities in heating the reconnection outflow region. *Physics of Plasmas*, 25(12), 122902. <https://doi.org/10.1063/1.5054100>
- Kaufman, L., & Rousseeuw, P. J. (1990). Partitioning around medoids (program pam). In L. Kaufman & P. J. Rousseeuw (Eds.), *Finding groups in data (chap. 2)*. John Wiley and Sons, Inc. <https://doi.org/10.1002/9780470316801.ch2>
- Klein, K. G., Spence, H., Alexandrova, O., Argall, M., Arzamasskiy, L., Bookbinder, J., et al. (2023). Helioswarm: A multipoint, multiscale mission to characterize turbulence. *Space Science Reviews*, 219(74), 74. <https://doi.org/10.1007/s11214-023-01019-0>
- Köhne, S., Boella, E., & Innocenti, M. E. (2023). Unsupervised classification of fully kinetic simulations of plasmoid instability using self-organizing maps (soms). *Journal of Plasma Physics*, 89(3), 895890301. <https://doi.org/10.1017/S0022377823000454>
- Lapenta, G., Markidis, S., Divin, A., Newman, D., & Goldman, M. (2015). Separatrices: The crux of reconnection. *Journal of Plasma Physics*, 81(1), 325810109. <https://doi.org/10.1017/S0022377814000944>
- Lapenta, G., Wang, R., & Cazzola, E. (2016). Reconnection separatrix: Simulations and spacecraft measurements. In W. Gonzalez & E. Parker (Eds.), *Magnetic reconnection* (pp. 315–344). Springer International Publishing. [https://doi.org/10.1007/978-3-319-26432-5\\_8](https://doi.org/10.1007/978-3-319-26432-5_8)
- Lenouvel, Q., Génot, V., Garnier, P., Toledo-Redondo, S., Lavraud, B., Aunai, G., et al. (2021). Identification of electron diffusion regions with a machine learning approach on mms data at the earth's magnetopause. *Earth and Space Science*, 8(5), e2020EA001530. <https://doi.org/10.1029/2020EA001530>
- Lindstedt, T., Khotyaintsev, Y. V., Vaivads, A., André, M., Fear, R. C., Lavraud, B., et al. (2009). Separatrix regions of magnetic reconnection at the magnetopause. *Annals of Geophysics*, 27(10), 4039–4056. <https://doi.org/10.5194/angeo-27-4039-2009>
- Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- Nakamura, R., Genestreti, K. J., Nakamura, T., Baumjohann, W., Varsani, A., Nagai, T., et al. (2019). Structure of the current sheet in the 11 July 2017 electron diffusion region event. *Journal of Geophysical Research: Space Physics*, 124(2), 1173–1186. <https://doi.org/10.1029/2018JA026028>
- Nguyen, G., Aunai, N., de Welle, B. M., Jeandet, A., Lavraud, B., & Fontaine, D. (2021). Massive multi-mission statistical study and analytical modeling of the earth's magnetopause: 1. A gradient boosting based automatic detection of near-earth regions. *Journal of Geophysical Research: Space Physics*, 127. <https://doi.org/10.1029/2021JA029773>
- Oka, M., Phan, T. D., Øieroset, M., Turner, D. L., Drake, J. F., Li, X., et al. (2022). Electron energization and thermal to non-thermal energy partition during earth's magnetotail reconnection. *Physics of Plasmas*, 29(5), 052904. <https://doi.org/10.1063/5.0085647>
- Olshesky, V., Khotyaintsev, Y. V., Lalti, A., Divin, A., Delzanno, G. L., Anderzén, S., et al. (2021). Automated classification of plasma regions using 3d particle energy distributions. *Journal of Geophysical Research: Space Physics*, 126(10). <https://doi.org/10.1029/2021JA029620>
- Pucci, F., Usami, S., Ji, H., Guo, X., Horiuchi, R., Okamura, S., et al. (2018). Energy transfer and electron energization in collisionless magnetic reconnection for different guide-field intensities. *Physics of Plasmas*, 25(12), 122111. <https://doi.org/10.1063/1.5050992>
- Retinò, A., Khotyaintsev, Y., Le Contel, O., Marcucci, M. F., Plaschke, F., Vaivads, A., et al. (2022). Particle energization in space plasmas: Towards a multi-point, multi-scale plasma observatory. *Experimental Astronomy*, 54(2), 427–471. <https://doi.org/10.1007/s10686-021-09797-7>
- Retinò, A., Vaivads, A., André, M., Sahraoui, F., Khotyaintsev, Y., Pickett, J. S., et al. (2006). Structure of the separatrix region close to a magnetic reconnection x-line: Cluster observations. *Geophysical Research Letters*, 33(6), 2005GL024650. <https://doi.org/10.1029/2005GL024650>
- Ricci, P., Brackbill, J. U., Daughton, W., & Lapenta, G. (2004). Collisionless magnetic reconnection in the presence of a guide field. *Physics of Plasmas*, 11(8), 4102–4114. <https://doi.org/10.1063/1.1768552>
- Roberts, D. A., Homa, K., Sipes, T., Ko, Y.-K., & Lepri, S. (2020). Objectively determining states of the solar wind using machine learning. *The Astrophysical Journal*, 889(2), 153. <https://doi.org/10.3847/1538-4357/ab5a7a>
- Romano, P., Schillirò, F., & Falco, M. (2023). Observations of the chromospheric evershed flow of sunspot penumbra with the application of the self-organizing map technique. *The Astrophysical Journal*, 958(1), 56. <https://doi.org/10.3847/1538-4357/acfc20>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Torbert, R. B., Burch, J. L., Phan, T. D., Hesse, M., Argall, M. R., Shuster, J., et al. (2018). Electron-scale dynamics of the diffusion region during symmetric magnetic reconnection in space. *Science*, 362(6421), 1391–1395. <https://doi.org/10.1126/science.aat2998>
- Toy-Edens, V., Mo, W., Raptis, S., & Turner, D. L. (2024). Classifying 8 years of mms dayside plasma regions via unsupervised machine learning. *Journal of Geophysical Research: Space Physics*, 129(6), e2024JA032431. <https://doi.org/10.1029/2024JA032431>
- Tyler, E., Cattell, C., Thaller, S., Wygant, J., Gurgiolo, C., Goldstein, M., & Mouikis, C. (2016). Partitioning of integrated energy fluxes in four tail reconnection events observed by cluster. *Journal of Geophysical Research: Space Physics*, 121(12). <https://doi.org/10.1002/2016JA023330>
- Umeda, T., Togano, K., & Ogino, T. (2010). Structure of diffusion regions in collisionless magnetic reconnection. *Physics of Plasmas*, 17(5), 052103. <https://doi.org/10.1063/1.3403345>
- Vasyliunas, V. M. (1975). Theoretical models of magnetic field line merging. *Reviews of Geophysics*, 13(1), 303–336. <https://doi.org/10.1029/RG013i001p00303>
- Waters, C. L., Eastwood, J. P., Fargette, N., Newman, D. L., & Goldman, M. V. (2024). 2.5-D PIC simulation of symmetric magnetic reconnection with varying guide field [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.12570878>
- Wilder, F. D., Ergun, R. E., Eriksson, S., Phan, T. D., Burch, J. L., Ahmadi, N., et al. (2017). Multipoint measurements of the electron jet of symmetric magnetic reconnection with a moderate guide field. *Physical Review Letters*, 118(26), 265101. <https://doi.org/10.1103/PhysRevLett.118.265101>

- Yamada, M., Ji, H., Hsu, S., Carter, T., Kulsrud, R., Ono, Y., & Perkins, F. (1997). Identification of y-shaped and o-shaped diffusion regions during magnetic reconnection in a laboratory plasma. *Physical Review Letters*, 78(16), 3117–3120. <https://doi.org/10.1103/PhysRevLett.78.3117>
- Yeakel, K. L., Vandegriff, J. D., Garton, T. M., Jackman, C. M., Clark, G., Vines, S. K., et al. (2022). Classification of cassini's orbit regions as magnetosphere, magnetosheath, and solar wind via machine learning. *Frontiers Astronomy Space Science*, 9, 875985. <https://doi.org/10.3389/fspas.2022.875985>
- Zhou, M., Deng, X. H., Pang, Y., Huang, S. Y., Yuan, Z. G., Li, H. M., et al. (2012). Revealing the sub-structures of the magnetic reconnection separatrix via particle-in-cell simulation. *Physics of Plasmas*, 19(7), 072907. <https://doi.org/10.1063/1.4739283>
- Zweibel, E. G., & Yamada, M. (2016). Perspectives on magnetic reconnection. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences A*, 472(2196), 20160479. <https://doi.org/10.1098/rspa.2016.0479>