CURRENT STRUCTURES AND RECONNECTION EVENTS ANALYSIS IN HYBRID-KINETIC TURBULENCE SIMULATIONS USING UNSUPERVISED MACHINE LEARNING

Manuela Sisti[1,2], Giorgio Pedrazzi[3], Francesco Finelli[2], Matteo Faganello[1], Francesco Califano[2], Francesca Delli Ponti[3], Andong Hu[4], and Jannis Teunissen[4]

Corresponding authors: manuela.sisti@univ-amu.fr

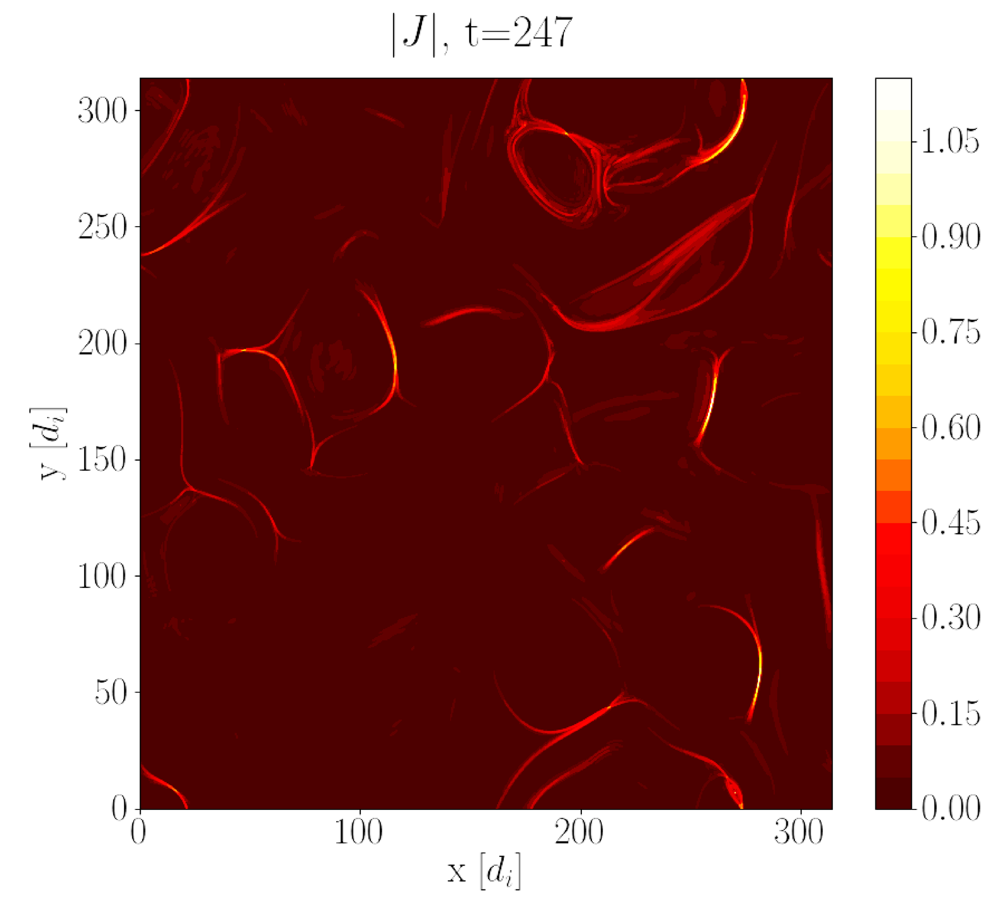
[1] Aix-Marseille University, CNRS, PIIM UMR 7345 - Marseille, France, [2] Physics Department, University of Pisa - Pisa, Italy,
[3] HPC Business Unit, Cineca, Bologna, Italia, EU, [4] Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands, EU

The AIDA project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 776262.

Introduction

In the context of the study of magnetized turbulent space plasmas, a keypoint concerns the formation of coherent structures and their disruption through reconnection. Up to now no well verified technique to automatically detect such structures has been developed, and reconnection can be only identified by human analysis looking at possible sites one by one. Lots of data are produced by simulations (MHDs, Hall-MHDs, hybrid ones, PICs) and by satellite measurements. Thus there is a need to find a way to rapidly and efficiently identify reconnection events among these big data. Our goal is to set-up an algorithm using unsupervised Machine Learning aimed at **automatically detecting** the presence of magnetic structures where reconnection is occurring (2D). The final objective will be to adapt these algorithms to satellite data.

This research is developed in the framework of the european project AIDA (Artificial Intelligence Data Analysis).



Model and simulation setup

Hybrid Vlasov-Maxwell code[a]
Normalized[b] equations used:

- Vlasov equation for the ion distribution function:

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla f + (\mathbf{E} + \mathbf{v}\wedge\mathbf{B})\cdot\frac{\partial f}{\partial \mathbf{v}} = 0 \qquad (1)$$

- The Ohm's law for the electric field:

$$\mathbf{E} - d_e^2\nabla^2\mathbf{E} = -(\mathbf{u}\wedge\mathbf{B}) + \frac{1}{n}(\mathbf{J}\wedge\mathbf{B}) + -\frac{1}{n}\nabla P_e + \frac{d_e^2}{n}\nabla\cdot[\mathbf{uJ}+\mathbf{Ju}] - \frac{1}{n}d_e^2\nabla\cdot\left(\frac{\mathbf{JJ}}{n}\right) \qquad (2)$$

- Faraday's equation:

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla\wedge\mathbf{E} \qquad \mathbf{J} = \nabla\wedge\mathbf{B} \qquad (3)$$

- $n_i \simeq n_e$

Box: 2D3V, $L_x = L_y = 2\pi * 50 d_i$, 3072*3072 grid points, resolution $\sim 0.1 d_i$.
Initial set-up:

- $B_0 = 1$, $m_i/m_e = 100$, $\beta_i \doteq 8\pi n T_{i0}/B_0^2 = 1$
- initial distribution function: Maxwellian ($T_{0i} = T_{0e}$)
- Turbulence is initialized with random, isotropic magnetic-field perturbations, with $k \in [0.02, 0.12]$ and $dB_{rms} \sim 0.28$.

[a] F. Valentini et al., Journal of Computational Physics 225 (2007)
[b] to ion mass, ion cyclotron frequency Ω_{ci}, Alfvèn velocity and ion skin depth d_i

Results 1: Overview

Variables used as reconnection proxies (normalized):

- current density $|\mathbf{J}|$
- in-plane electron velocity
- electron vorticity $\mathbf{\Omega_e} = \nabla\wedge\mathbf{V_e}$
- in-plane magnetic field
- electron decoupling: $\mathbf{E}' = \mathbf{E} + \mathbf{V_e}\wedge\mathbf{B}$, z-component
- $\mathbf{J}\cdot\mathbf{E}'$

Algorithm's steps:

1. Parallel tuning k for the k-means model:
   - to avoid overfitting or underfitting
   - $k = 11$ for t 247 (tuning with Davis-Bouldin index [1])
2. K-means[a] (Lloyd's algorithm)
   - find clusters in the variables space
   - we choose the cluster in variables space where the mean value of $|J|$ is the highest
3. DBscan[a] algorithm
   - the cluster chosen using K-means is made up by different structures in physical space, we use DBscan to distinguish them
   - $\epsilon = 50$ grid points $\sim 5d_i$ (search radius for DBscan)
   - Minimum points number to form a cluster (DBscan): 100
4. Threshold on structures' aspect ratio

[a] K-means and DBscan are techniques of unsupervised ML with the aim to learn a grouping structure in a dataset (clustering techniques) [1].

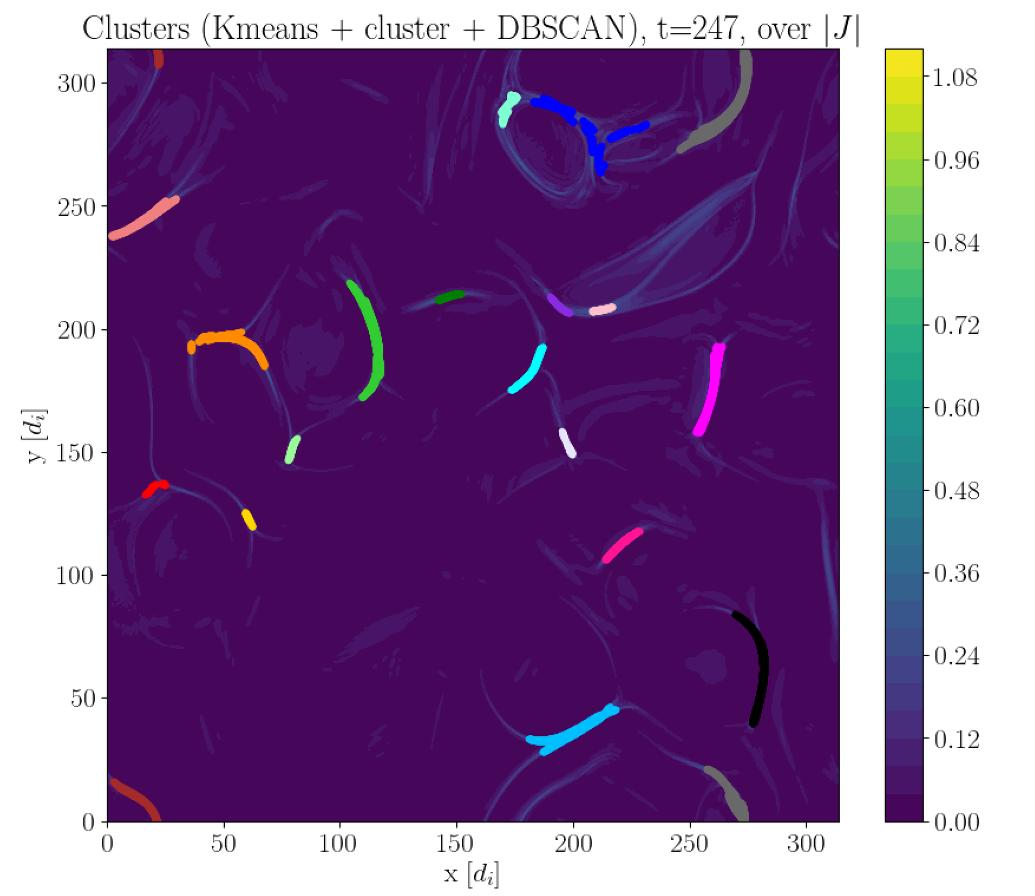
Results 1: Overview

- The aspect ratio of the reconnection structure is linked to the reconnection rate R, in particular $R \sim 0.1$ for fast reconnection model [2], which gives $AR = \frac{lenght}{width} \sim 10$

Quantities to evaluate performance:

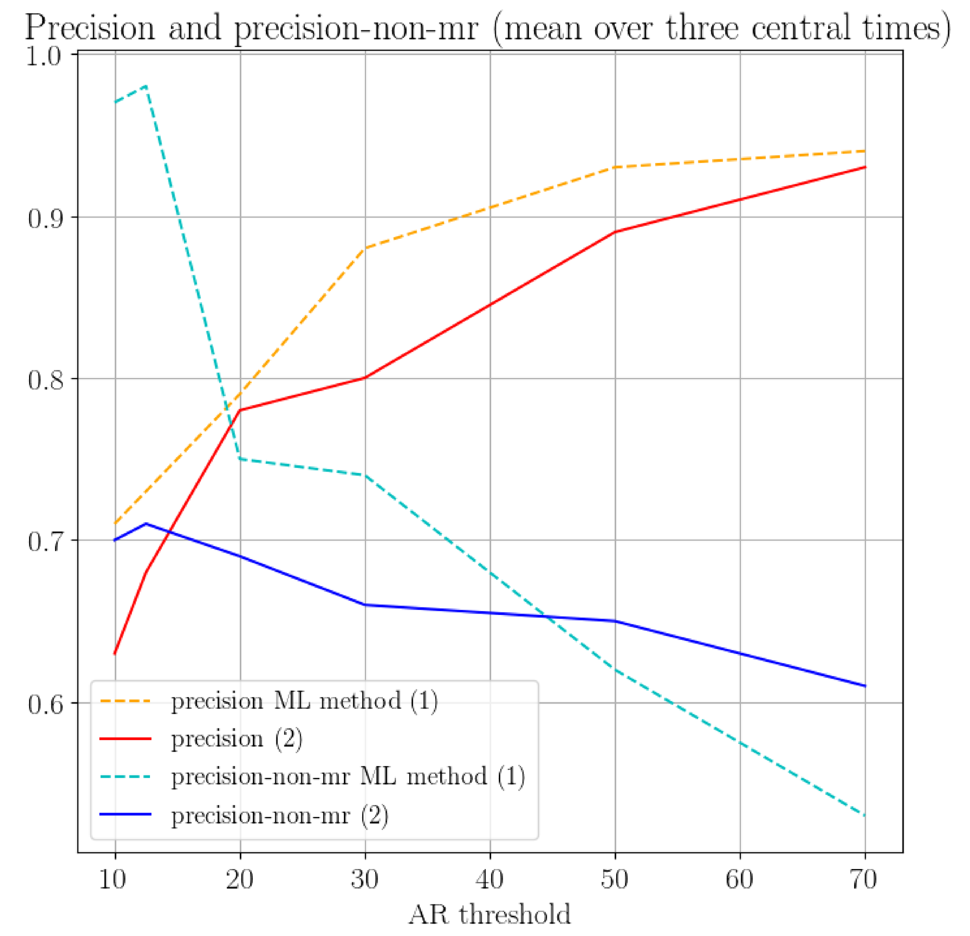
- Precision $= \frac{\text{\# reconnection sites among sites selected}}{\text{\# sites selected}}$
- Precision non-mr $= \frac{\text{\# non-reconnections among sites excluded}}{\text{\# sites excluded}}$

In Figure we show the final clusters (in different colors) which we obtain by applying the first three steps. We compute the aspect ratio of each of these structures.



Results 2: Performance

Comparison between (1) method: kmeans + DBscan + AR th,
(2) method: threshold over J + threshold over AR



t $[1/\Omega_{ci}]$	247	494	Av (3t[a])	Av (4t[b])
N. clusters	**19**	49		
N. clusters				
$AR > 10$	**17**	41		
$AR > 12.5$	17	35		
$AR > 20$	15	30		
$AR > 30$	10	22		
precision				
$AR > 10$	**0.82**	0.49	0.71	0.66
$AR > 12.5$	0.82	0.51	0.73	0.67
$AR > 20$	0.80	0.53	0.79	0.73
$AR > 30$	1	0.5	0.88	0.78
precision NON-mr				
$AR < 10$	**1**	0.87	0.97	0.95
$AR < 12.5$	1	0.79	0.98	0.93
$AR < 20$	0.5	0.74	0.75	0.75
$AR < 30$	0.55	0.63	0.74	0.71

[a] Average over three times (230,247,282)
[b] Average over four times (230,247,282,494)

Conclusion

In particular:

- Method which combines unsupervised machine learning and a threshold on the aspect ratio of the structures
- Precision: $\sim$ 80%; precision non-mr among excluded sites: $\sim$ 80% (AR th $\sim$ 18)
- Better than using only a threshold over J + threshold over AR

In general:

- We are working to obtain a quite accurate method to automatically find reconnecting current sheets in turbulence simulations
- In the framework of the AIDA project, we are creating an utility (in Python) which will become free-to-use and available in AIDA repository to all space plasma physicists

References

[1] Hefin I. Rhys, "Machine Learning with R, the tidyverse, and mlr", Manning Publications Co (2020)

[2] P. A. Cassak, Y.-H. Liu and M. A. Shay, "A review of the 0.1 reconnection rate problem", *J. Plasma Phys.* (2017), **vol. 83**, 715830501, doi:10.1017/S0022377817000666