Unsupervised classification of the solar wind using Self-Organising Maps

Jorge AMAYA

Research Expert at the CmPA

Mathematics Department, KU Leuven, Belgium

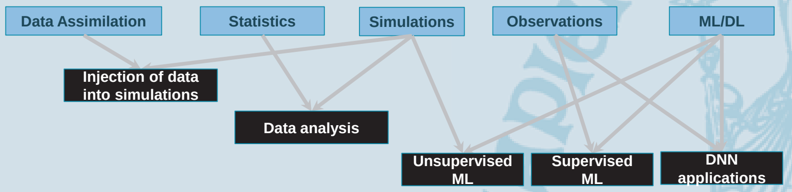
Artificial Intelligence and Data Analysis (AIDA) in Heliophysics:
Developing the AIDApy python package that centralizes and simplifies access to:

- Spacecraft data
- Space physics simulations
- Advanced statistical tools
- Machine Learning and Deep Learning algorithms and applications

This talk has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776262 (AIDA, www.aida-space.eu)

Artificial Intelligence and Data Analysis (AIDA) in Heliophysics:
Developing the AIDApy python package that centralizes and simplifies access to:

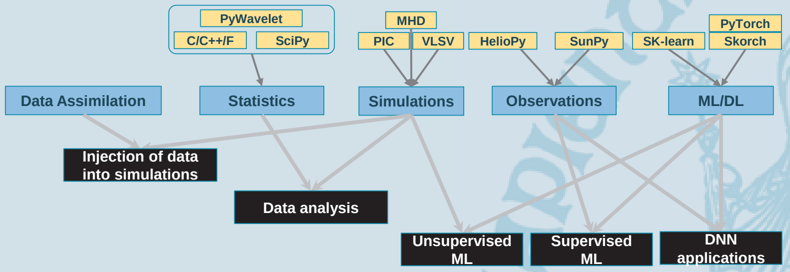
- Spacecraft data
- Space physics simulations
- Advanced statistical tools
- Machine Learning and Deep Learning algorithms and applications



This talk has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776262 (AIDA, www.aida-space.eu)

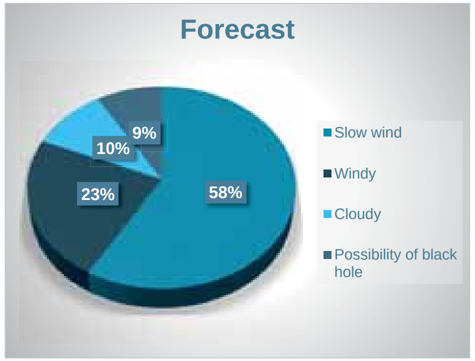
Artificial Intelligence and Data Analysis (AIDA) in Heliophysics:
Developing the AIDApy python package that centralizes and simplifies access to:

- Spacecraft data
- Space physics simulations
- Advanced statistical tools
- Machine Learning and Deep Learning algorithms and applications



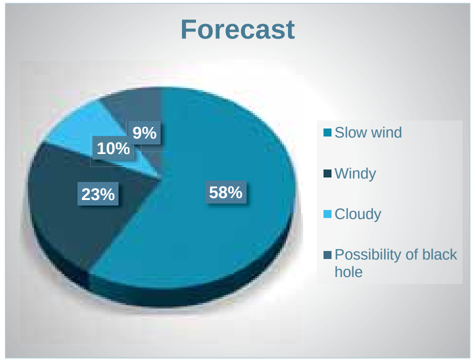
This talk has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776262 (AIDA, www.aida-space.eu)

Objective of solar wind classification



Xu, F., & Borovsky, J. E. (2015). JGR: Space Physics
Richardson, I. G., & Cane, H. V. (2012). JSWSC
Zastenker, G. N., et al. (2014). Cosmic Research

Objective of solar wind classification

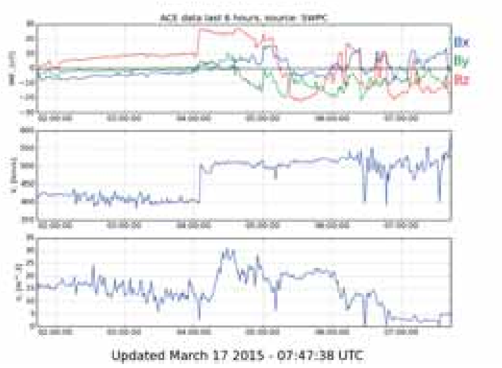


Reasons to use classification of the instantaneous solar wind

- Statistical characterisation of different plasma flows
- Study fluctuations in the plasma properties depending on the solar cycle
- Diagnose physical processes in the Sun based on the observations of plasma at 1AU

Xu, F., & Borovsky, J. E. (2015). JGR: Space Physics
Richardson, I. G., & Cane, H. V. (2012). JSWSC
Zastenker, G. N., et al. (2014). Cosmic Research

How to classify the solar wind



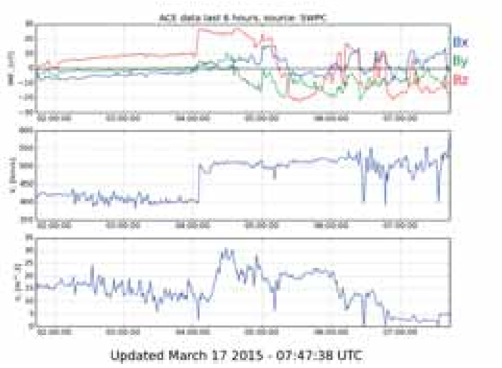
Arya, S., & Freeman, J. W. (1991). JGR: Space Physics, 96
Feldman, U., Landi, E., & Schwadron, N. A. (2005).JGR: Space Physics, 110

How to classify the solar wind

Use of basic solar wind properties: speed, magnetic field components, density.

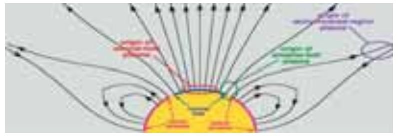
Historically classified as “fast” and “slow” solar wind

Mainly detects the occurrence of coronal holes



Arya, S., & Freeman, J. W. (1991). JGR: Space Physics, 96
Feldman, U., Landi, E., & Schwadron, N. A. (2005).JGR: Space Physics, 110

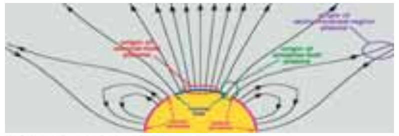
Advanced algebraic rules

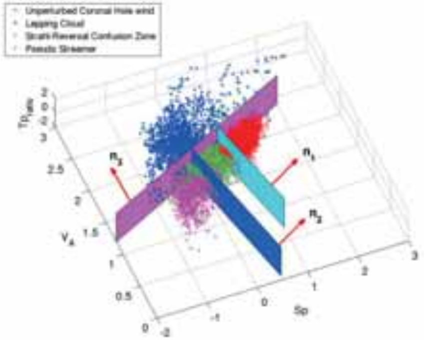


- Three/Four category-based classification on wind origin
- Based on more complex plasma properties, in particular heavy ion content

Zhao, L., Zurbuchen, T. H., & Fisk, L. A. (2009). GRL, 36
Xu, F., & Borovsky, J. E. (2015). JGR: Space Physics

Advanced algebraic rules





- Three/Four category-based classification on wind origin
- Based on more complex plasma properties, in particular heavy ion content
- Measurement of heavy ions, not available in all missions
- Translation to measured properties
- Proton Specific Entropy $S_p = T_p/n_p^{2/3}$
- Expected temperature ratio $T_{exp}/T_p = (V_{SW}/258)^{3.113}/T_p$
- Alfvén speed

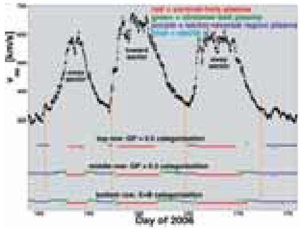
Zhao, L., Zurbuchen, T. H., & Fisk, L. A. (2009). GRL, 36
Xu, F., & Borovsky, J. E. (2015). JGR: Space Physics

Probabilistic classification

Probabilistic classification

Attribute	Symbol
Solar wind speed	V_{sw}
Proton temperature standard deviation	σ_T
Sunspot number	R
Solar radio flux (10.7 cm)	$f_{10.7}$
Alfven speed	v_A
Proton specific entropy	S_p
Temperature ratio	T_{exp}/T_p

Gaussian Process

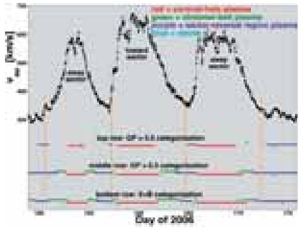


Camporeale, E., Carè, A., & Borovsky, J. E. (2017). JGR: *Space Physics*, *122*

Probabilistic classification

Attribute	Symbol
Solar wind speed	V_{sw}
Proton temperature standard deviation	σ_T
Sunspot number	R
Solar radio flux (10.7 cm)	$f_{10.7}$
Alfven speed	v_A
Proton specific entropy	S_p
Temperature ratio	T_{exp}/T_p

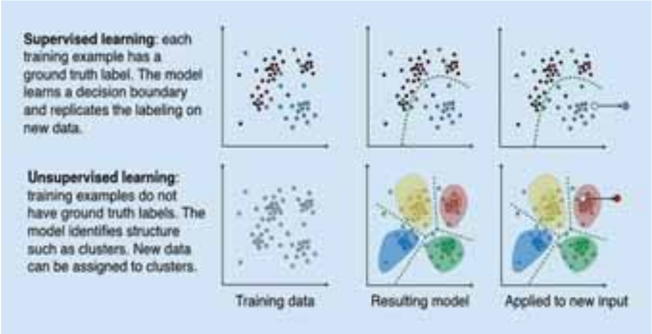
Gaussian Process



- Re-analysis of algebraic laws using Gaussian Process to include uncertainties
- Transformation of classification rules into probabilistic rules
- Allowing for ‘undefined’ classifications
- Better suited for forecasting and operational tools
- Input: OMNI data
- Output: Four-class probabilities

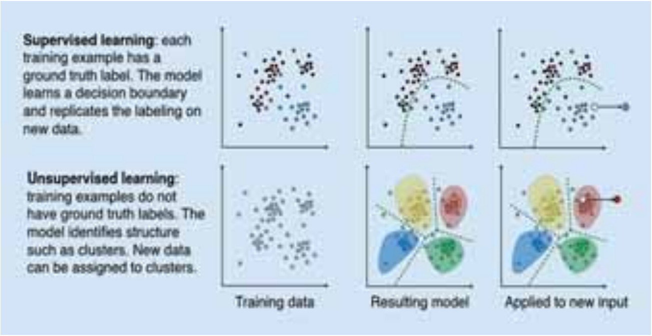
Camporeale, E., Carè, A., & Borovsky, J. E. (2017). JGR: *Space Physics*, *122*

ML alternative methods



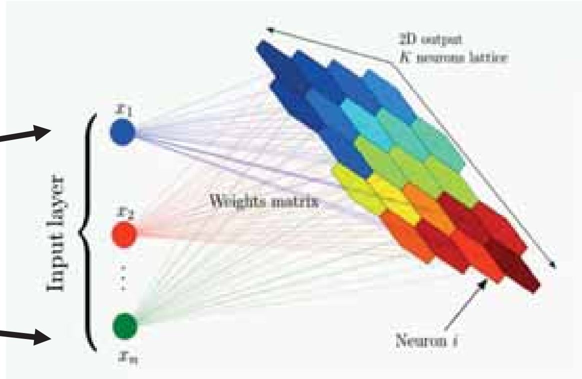
ML alternative methods

- There are limitations with the supervised methods: unavailable labelled data and small number of data points
- Look for unsupervised techniques to uncover hidden information
- Unsupervised learning is based on data, not on human perception



Self-Organizing Maps (SOM)

Attribute	Symbol
Solar wind speed	V_{sw}
Proton temperature standard deviation	σ_T
Sunspot number	R
Solar radio flux (10.7 cm)	$f_{10.7}$
Alfven speed	v_A
Proton specific entropy	S_p
Temperature ratio	T_{exp}/T_p



Kohonen, T. (1982). *Biological cybernetics*, *43*(1), 59-69.

SOM: how they learn?

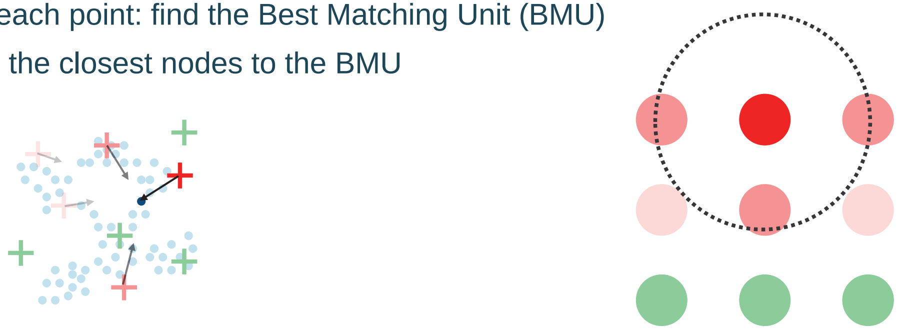
- Random initialization of ‘representative nodes’ in the features space
- The nodes belong to a ‘map’

Feature ND-space

SOM

SOM: how they learn?

- For each point: find the Best Matching Unit (BMU)
- Find the closest nodes to the BMU

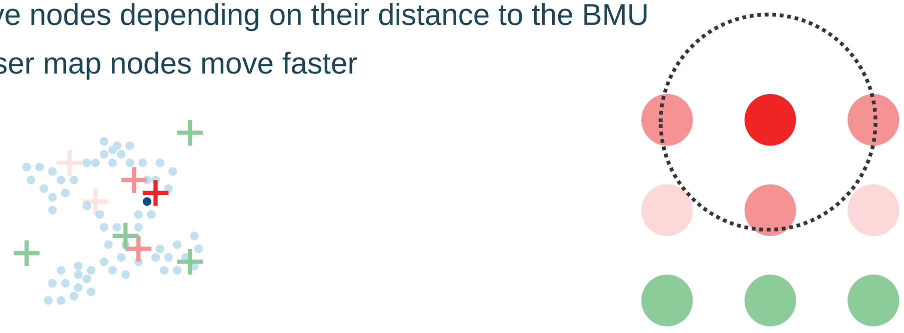


Feature ND-space

SOM

SOM: how they learn?

- Move nodes depending on their distance to the BMU
- Closer map nodes move faster



Feature ND-space

SOM

SOM: how they learn?

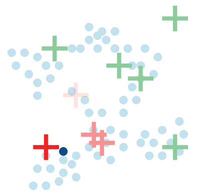
- The next data point will activate a different BMU
- Follow the same procedure as before

Feature ND-space

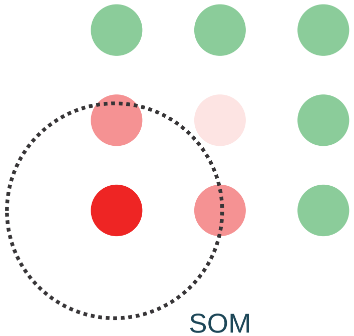
SOM

SOM: how they learn?

- Moving the nodes at different speeds allows to cover the ND-space, maintaining the similarity between neighbour map nodes.



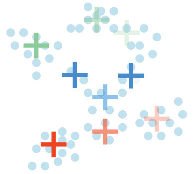
Feature ND-space



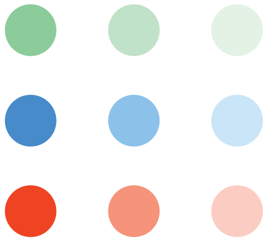
SOM

SOM: how they learn?

- Multiple epochs (full data iterations) allows the map to cover the full ND-space with nodes representing particularly dense regions, and maintaining similarity among neighbour nodes.



Feature ND-space



SOM

SOM example

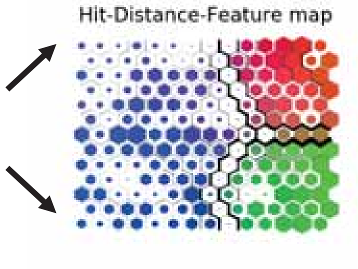
- RGB(0.2, 0.5, 0.77) = 
- 6000 points randomly distributed around the three colors

$$X=\begin{pmatrix} r,g,b \\ r,g,b \\ r,g,b \\ \cdot \\ \cdot \\ \cdot \\ r,g,b \end{pmatrix}$$

SOM example

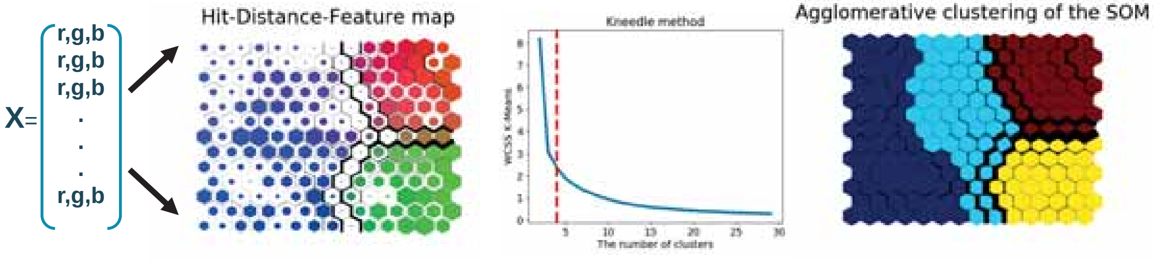
- RGB(0.2, 0.5, 0.77) = 
- 6000 points randomly distributed around the three colors

$$X=\begin{pmatrix} r,g,b \\ r,g,b \\ r,g,b \\ \cdot \\ \cdot \\ \cdot \\ r,g,b \end{pmatrix}$$



SOM example

- RGB(0.2, 0.5, 0.77) = 
- 6000 points randomly distributed around the three colors



Using SOM on solar wind data

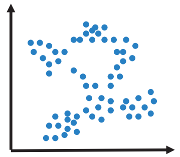
- OMNI data from 2000 to 2013:
- 100K points with around 40 features and derived properties

x_1 =

Attribute	Symbol
Solar wind speed	V_{sw}
Proton temperature standard deviation	σ_T
Sunspot number	R
Solar radio flux (10.7 cm)	$f_{10.7}$
Alfven speed	v_A
Proton specific entropy	S_p
Temperature ratio	T_{exp}/T_p

OMNI data pre-processing pipeline

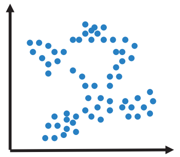
Read OMNI Data

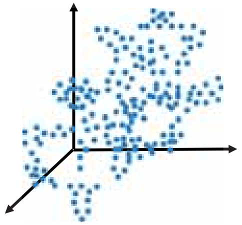


OMNI data pre-processing pipeline

Read OMNI Data

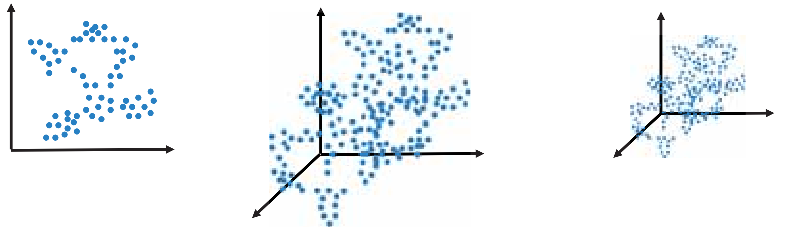
Enhanced Derived Data





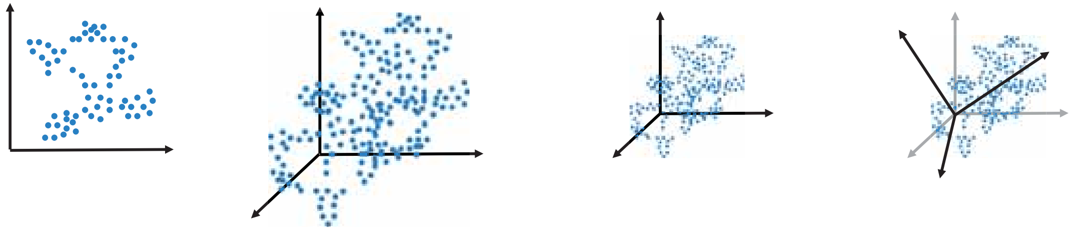
OMNI data pre-processing pipeline

Read OMNI Data | Enhanced Derived Data | Data Normalization



OMNI data pre-processing pipeline

Read OMNI Data | Enhanced Derived Data | Data Normalization | PCA



Hyper-parameters

- 13 x 14 map
- 60000 epochs
- 10 features

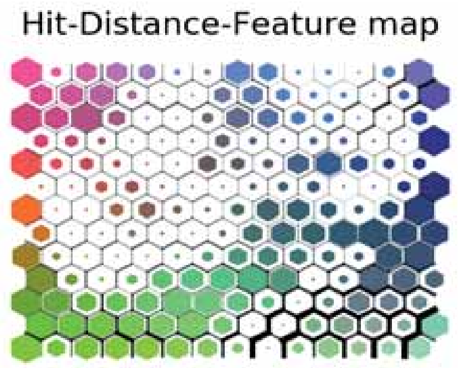
Na/Np	Newell	Boyle	Bz_GSM	By_GSM
Sp	Va	Tratio	f10.7	Ma

$$N = v^{4/3} B^{2/3} \sin^{8/3}(\theta/2)$$

$$BI = 10^{-4} v^2 + 11.7 B \sin^3(\theta/2)$$

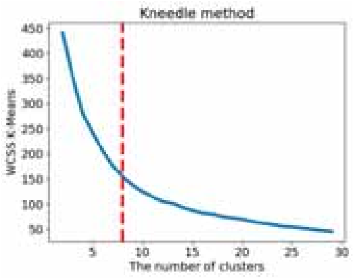
Boyle, C. B., Reiff, P. H., & Hairston, M. R. (1997). JGR: Space Physics, 102
Newell, P. T., Sotirelis, T., Liou, K., & Rich, F. J. (2008). JGR: Space Physics, 113

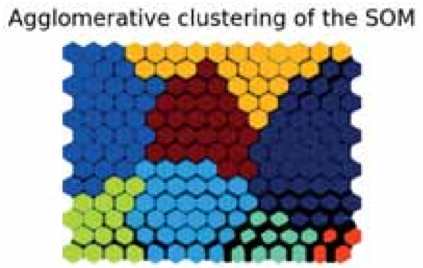
SOM hit and distance map



- Topology selection and training hyper-parameters selected using a grid search that minimizes the Quantization Error
- Hexagon size represents number of hits
- Line thickness represents the distance to the neighbouring node
- Colour is selected using the three most significant principal components

Node clustering

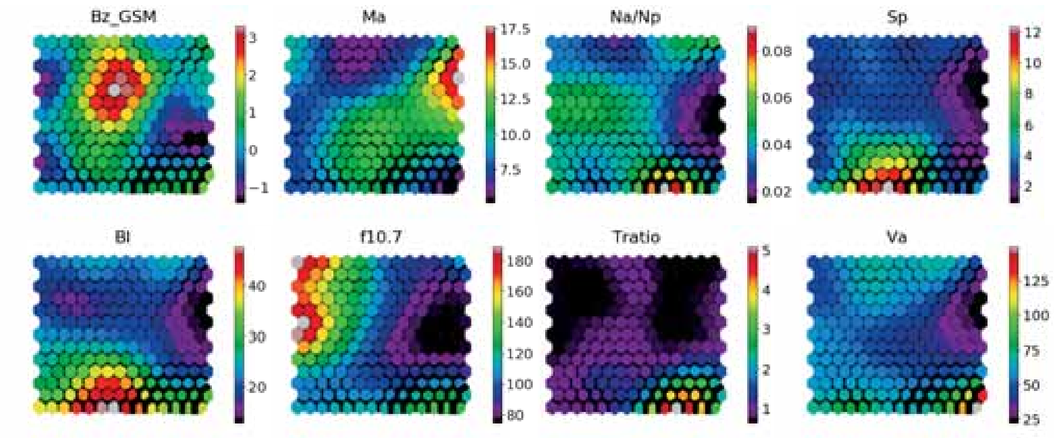




Within Cluster Sum of Squares: $\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x}\in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \arg\min_{\mathbf{S}} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i$

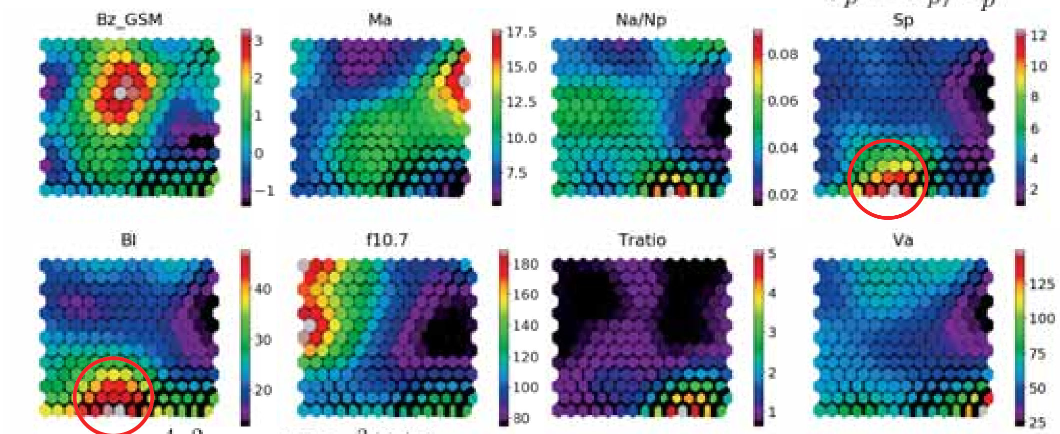
Kneedle Method to select 'k'

Feature maps



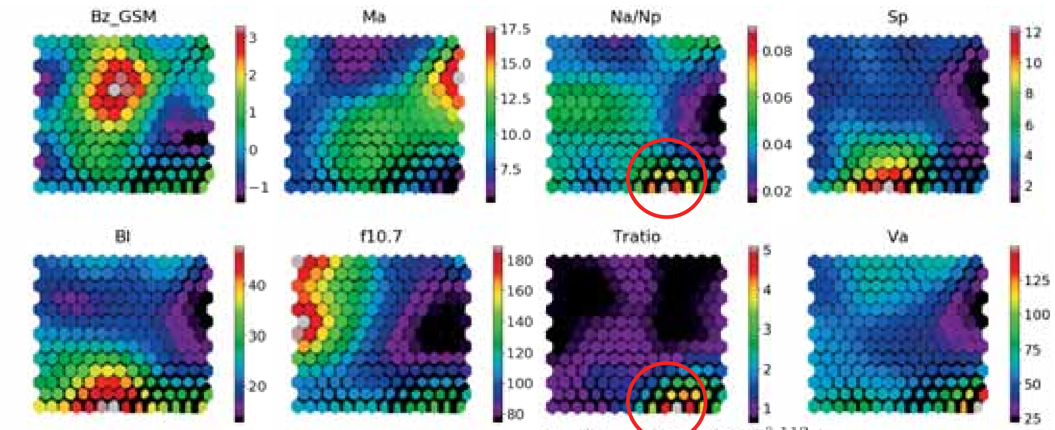
Feature maps

$$S_p = T_p / n_p^{2/3}$$



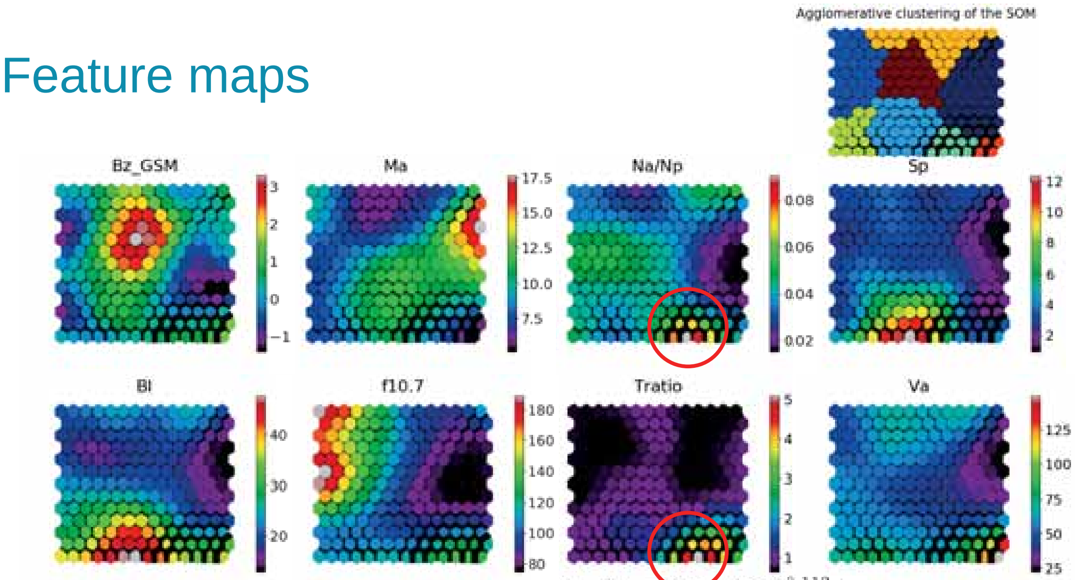
$$BI = 10^{-4} v^2 + 11.7 B \sin^3(\theta/2)$$

Feature maps



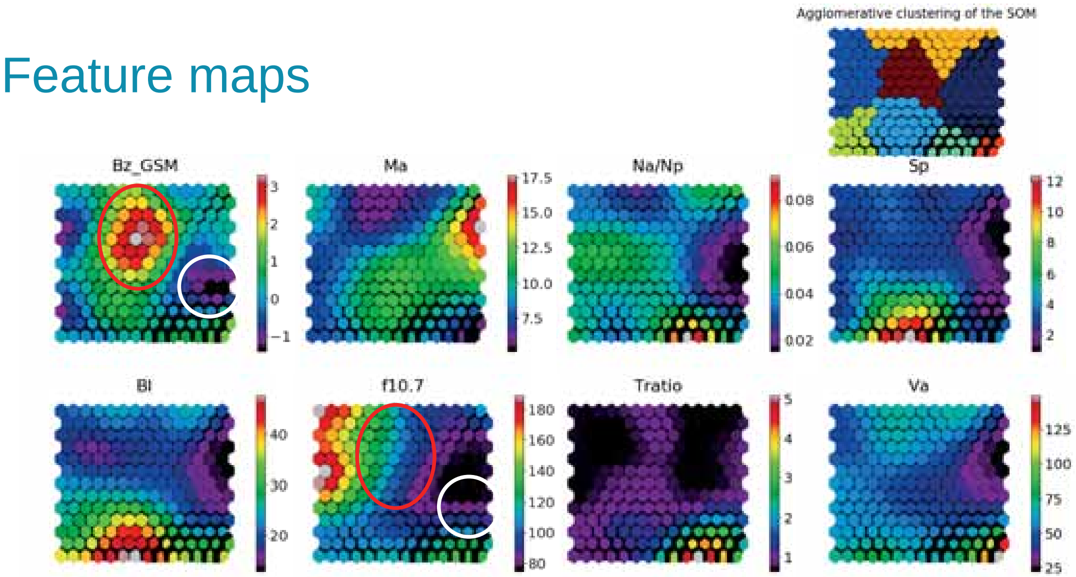
$$T_{exp}/T_p = (V_{SW}/258)^{3.113}/T_p$$

Feature maps

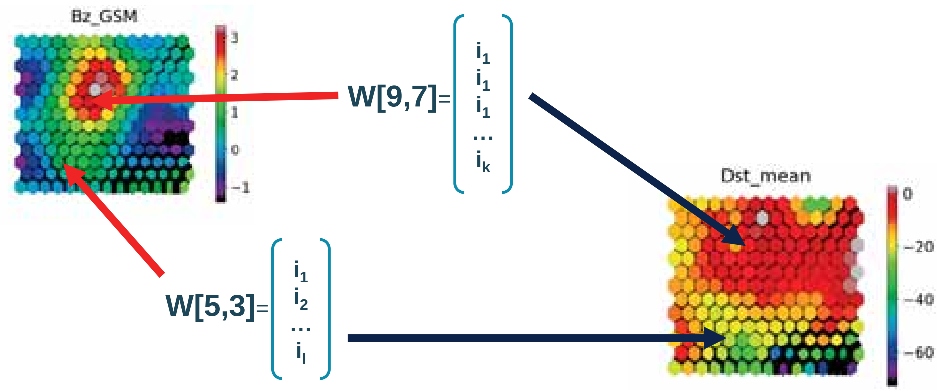


$$T_{exp}/T_p = (V_{SW}/258)^{3.113}/T_p$$

Feature maps

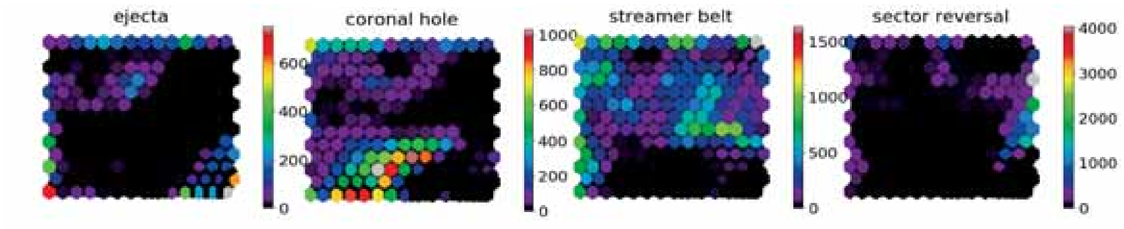


Classification and origin

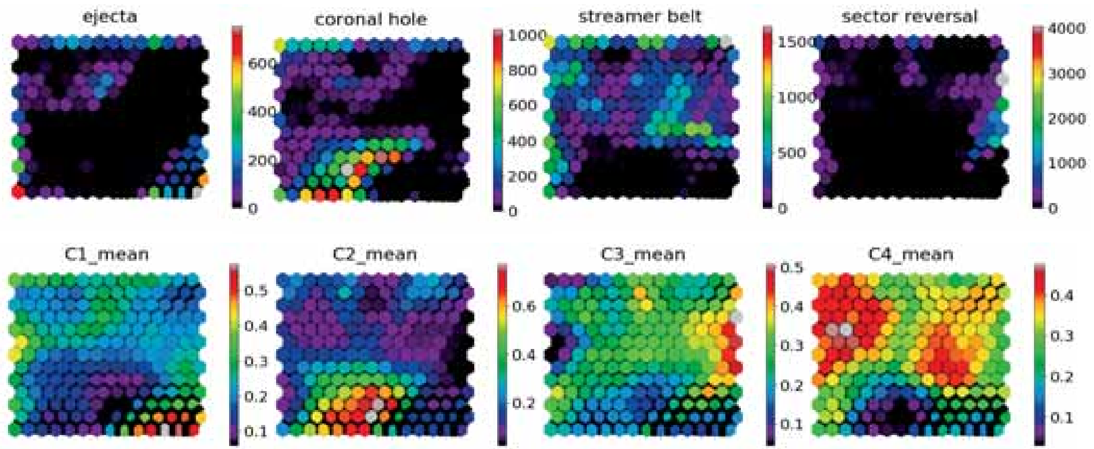


Classification and origin

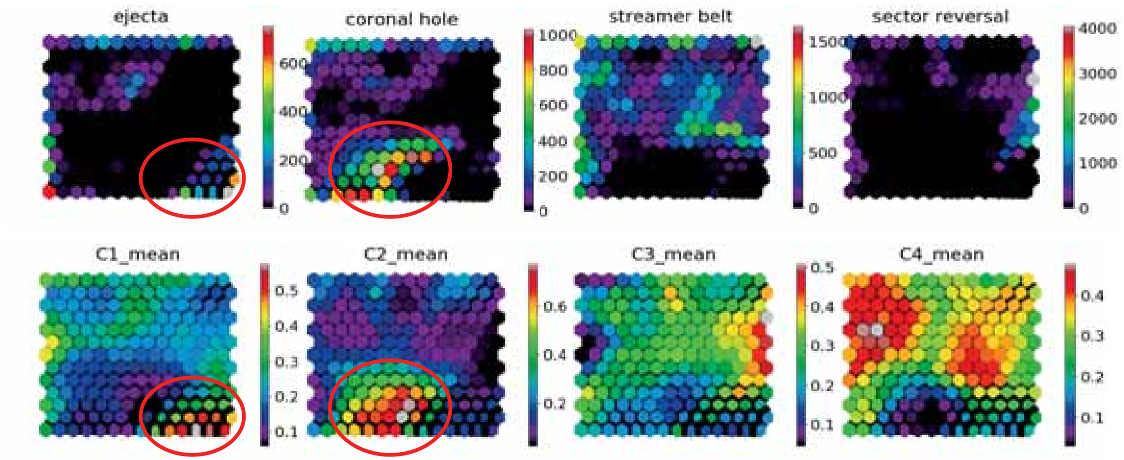
Classification and origin



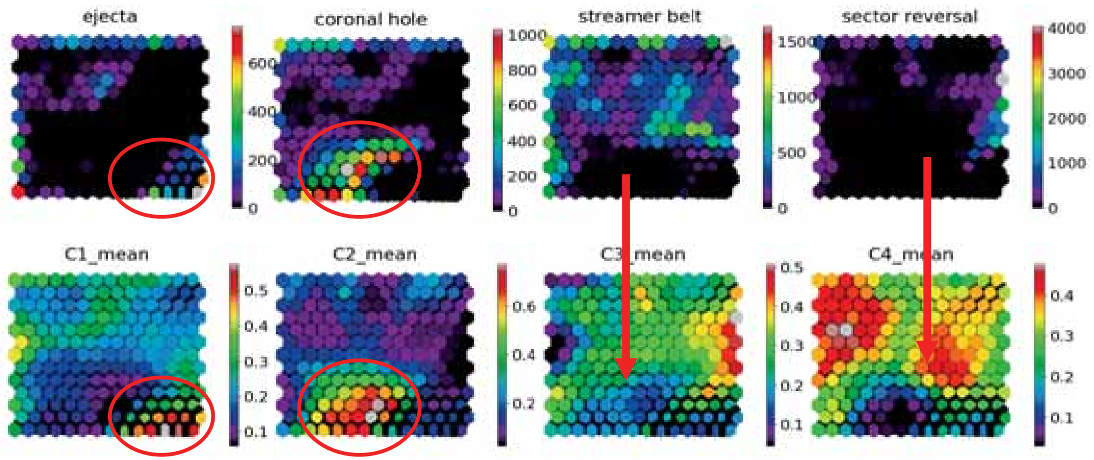
Classification and origin



Classification and origin

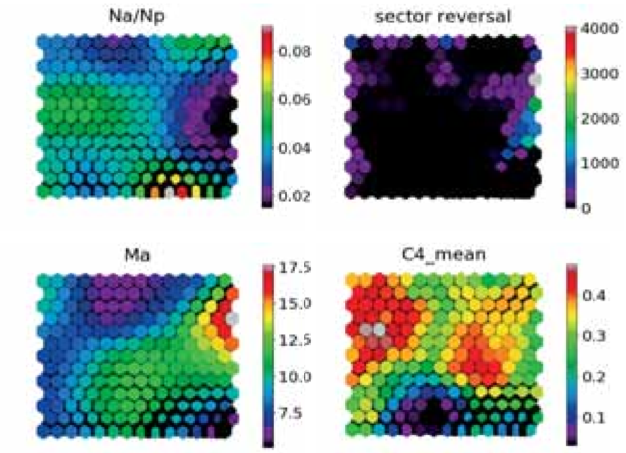


Classification and origin



Confirmation of previous results

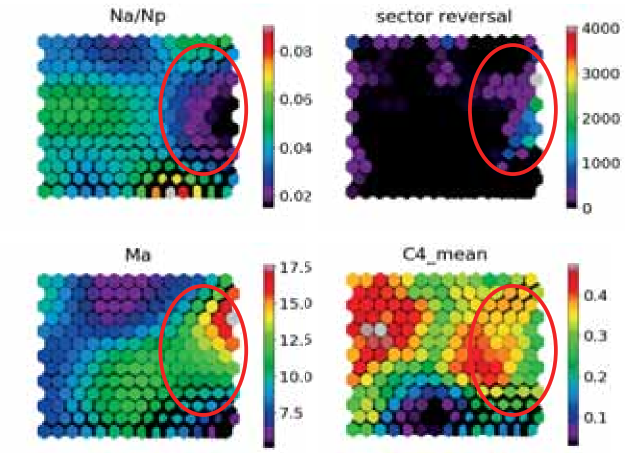
- The **sector-reversal**-region plasma appears to be a definite, different type of plasma from the other three types. It is found to have **low alpha-to-proton density** ratios and **higher-than-average Alfvén Mach** numbers.



Xu, F., & Borovsky, J. E. (2015). JGR: Space Physics

Confirmation of previous results

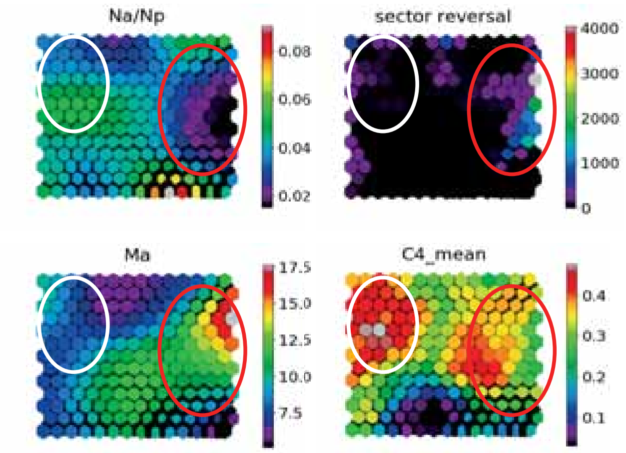
- The **sector-reversal**-region plasma appears to be a definite, different type of plasma from the other three types. It is found to have **low alpha-to-proton density** ratios and **higher-than-average Alfvén Mach** numbers.



Xu, F., & Borovsky, J. E. (2015). JGR: Space Physics

Confirmation of previous results

- The **sector-reversal**-region plasma appears to be a definite, different type of plasma from the other three types. It is found to have **low alpha-to-proton density** ratios and **higher-than-average Alfvén Mach** numbers.

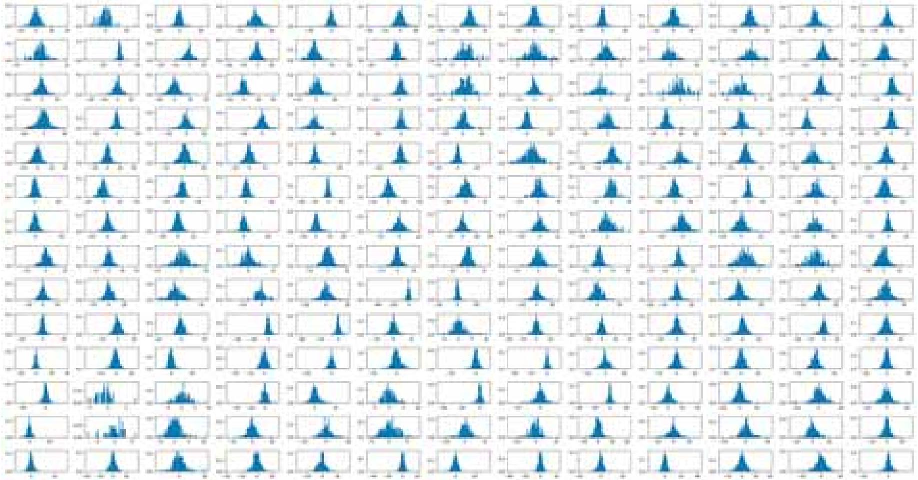


Xu, F., & Borovsky, J. E. (2015). JGR: Space Physics

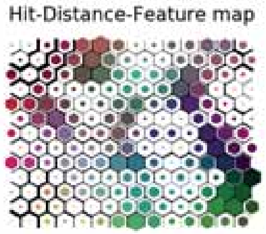
Node by node statistical analysis of the SW

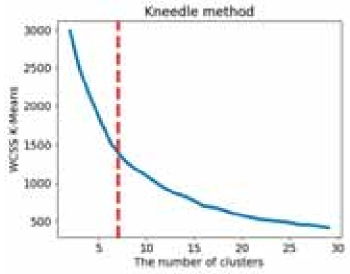
Node by node statistical analysis of the SW

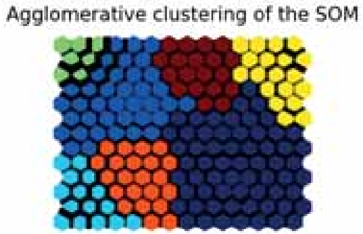
Bz



SOM of time series





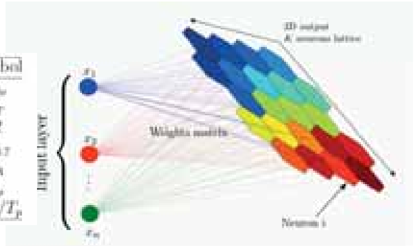


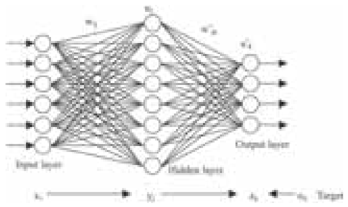
Time series SOM: feature map

Bz

Forecasting pipeline

Attribute	Symbol
Solar wind speed	V_{sw}
Proton temperature standard deviation	σ_T
Sunspot number	R
Solar radio flux (10.7 cm)	$f_{10.7}$
Alfven speed	v_A
Proton specific entropy	S_p
Temperature ratio	T_{exp}/T_p





Feature selection → Data reduction and transformation → Forecasting

Gathering | Discovery | Operations

On the works

- Coupling with ML techniques for forecasting
- Applying SOM to analyse other satellite data
- Including this tool in the AIDApy python package

Unsupervised classification of the solar wind using Self-Organising Maps

Jorge AMAYA

This talk has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776262 (AIDA, www.aida-space.eu)