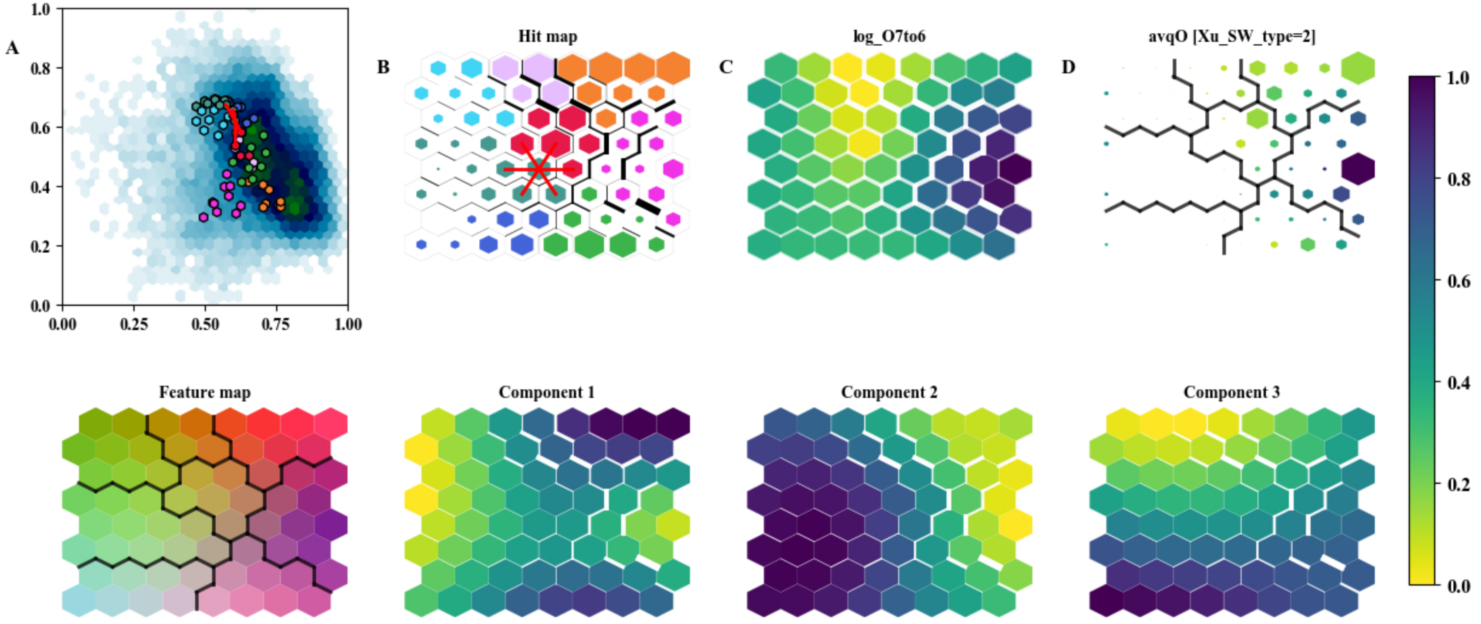
KU LEUVEN

Unsupervised classification of the solar wind using Self-Organizing Maps

Jorge AMAYA, *Romain Dupuis, Maria Elena Innocenti, Giovanni Lapenta*
Centre for mathematical Plasma-Astrophysics, Mathematics Department,
KU Leuven, Belgium
and
Jet Propulsion Laboratory, IHP Division, USA

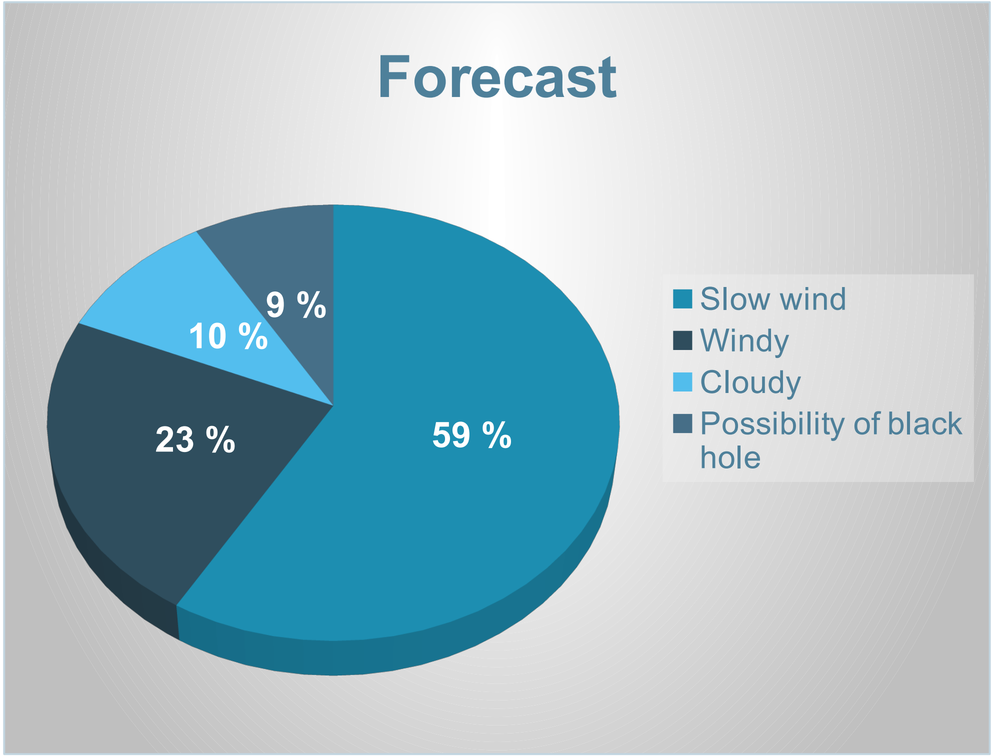
Reference



This work has been submitted for publication on the Research Topic Machine Learning in Heliophysics, Frontiers in Space Physics.

The preprint can be found in arXiv: **arXiv:2004.13430**

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

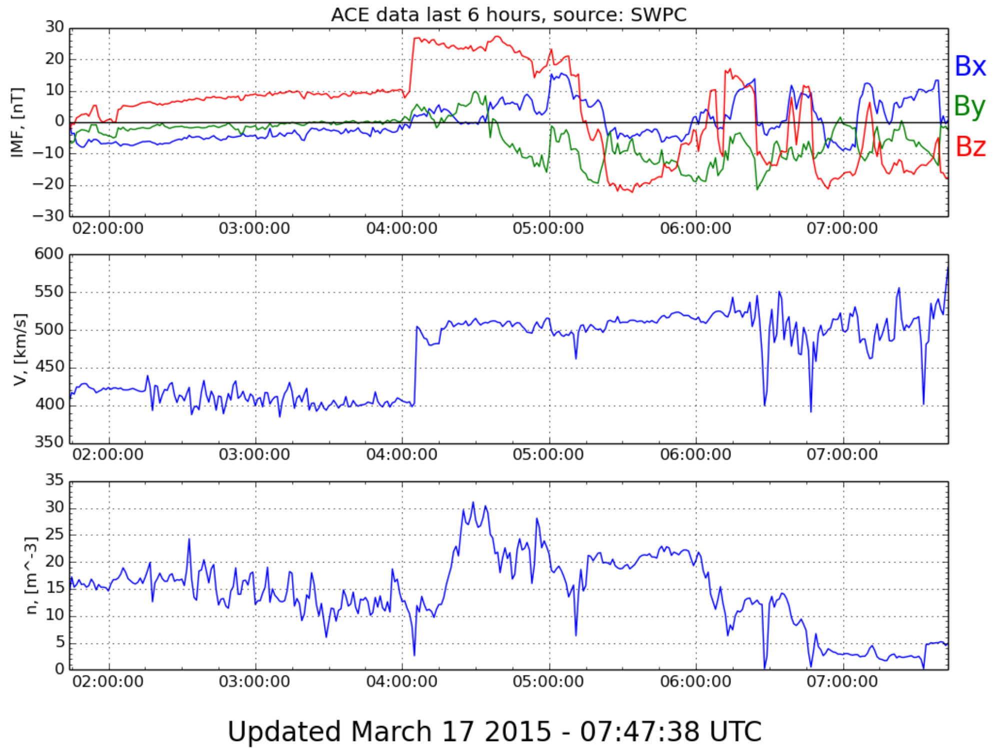
How to classify the solar wind

At the beginning basic solar wind properties were used to separate different classes: speed, magnetic field components, density.

Initially the wind was classified as “fast” and “slow”.

This basic classification mainly detects the occurrence of coronal holes

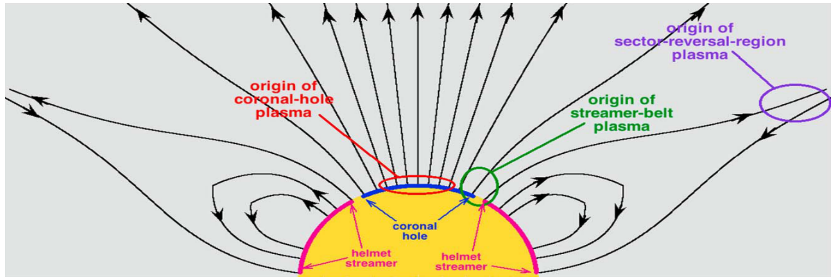
More advance algebraic empirical rules have been developed over time.

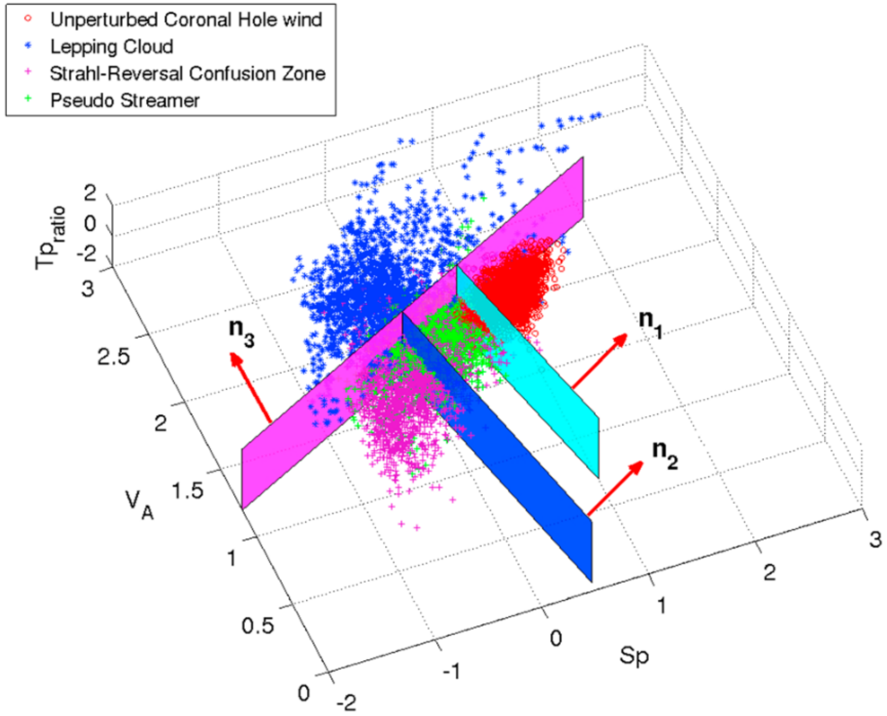


Arya, S., & Freeman, J. W. (1991). JGR: Space Physics, 96
Feldman, U., Landi, E., & Schwadron, N. A. (2005).JGR: Space Physics, 110

Advanced algebraic rules

- Three/Four category-based classification on wind origin [Zhao, 2009] [Xu & Borovsky, 2015]
- Xu classification is based on complex plasma properties, in particular heavy ion content
- Measurement of heavy ions: not available in all missions
- Properties based on moments used by Xu & Borovsky
  - Proton Specific Entropy $S_p = T_p/n_p^{2/3}$
  - Expected temperature ratio
  - Alfvén speed $T_{exp}/T_p = (V_{SW}/258)^{3.113}/T_p$

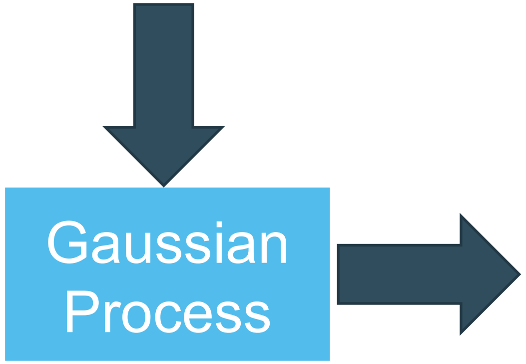


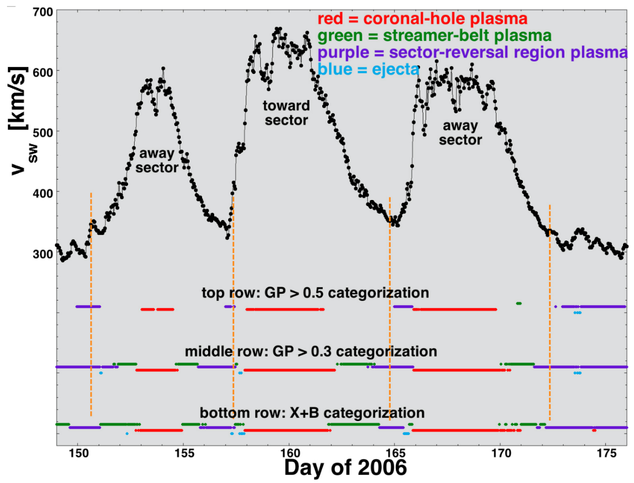


Zhao, L., Zurbuchen, T. H., & Fisk, L. A. (2009). GRL, 36
Xu, F., & Borovsky, J. E. (2015). JGR: Space Physics

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





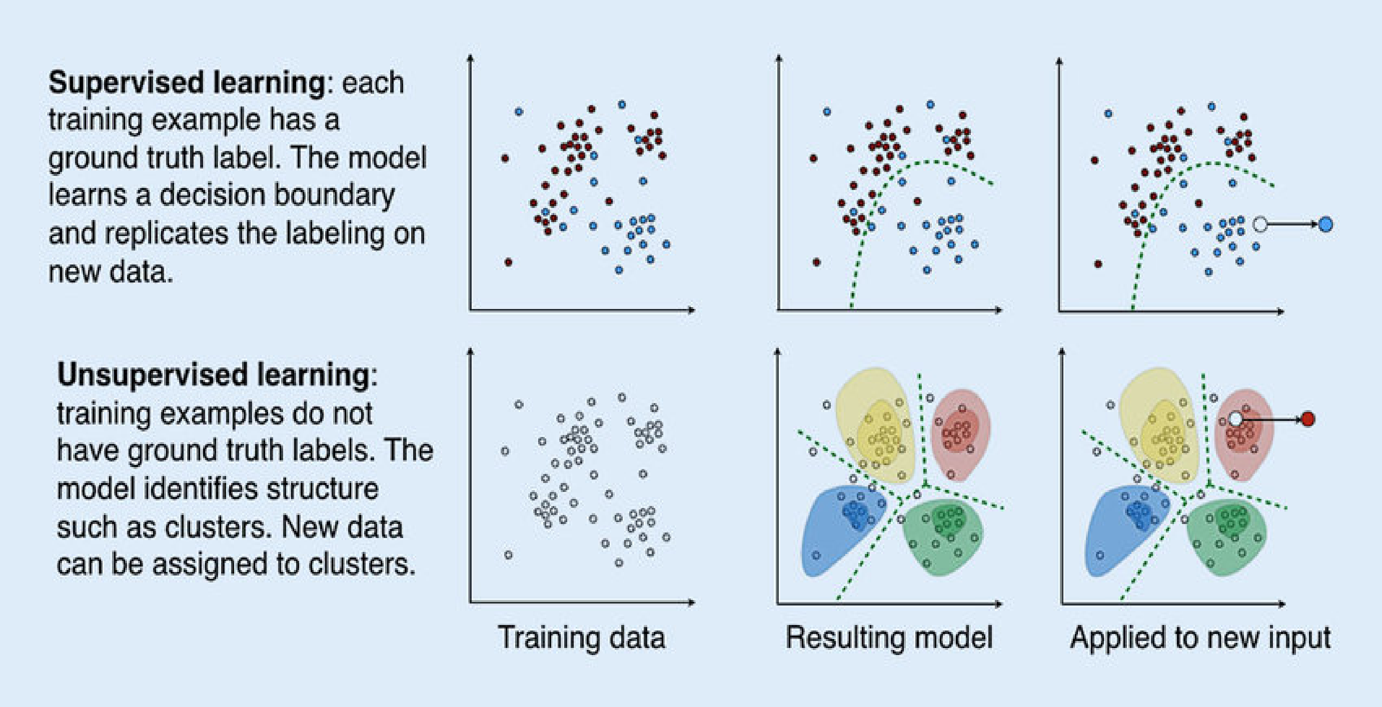
Camporeale, E., Carè, A., & Borovsky, J. E. (2017). JGR: *Space Physics*, *122*

A new apporach: probabilistic classification and machine learning

- 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

Machine Learning alternative methods

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



Self-Organizing Maps (SOM)

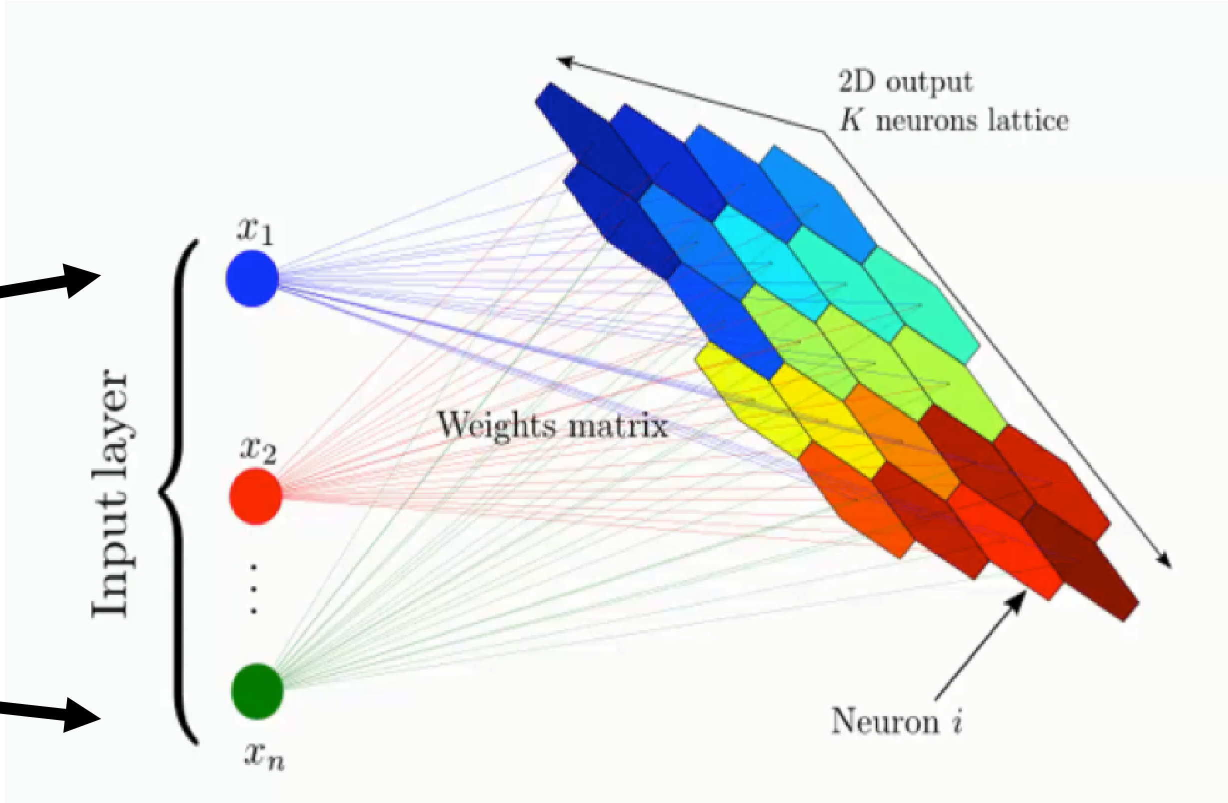
- SOM is a **clustering** technique
- It is classified as **unsupervised learning**: there is no “labeled” data
- In clustering we try to find a few points that represent groups of points in a N-dimensional space
- Other clustering methods include: *k*-means, GMM, DBCAN, agglomerative methods, etc.
- The difference of SOM is that in addition to cluster, **the Maps of the SOM contain topological information**
- Neighbor nodes in the SOM are also neighbors in the N-dimensional space
- ***We can then discover patterns in the groups of data by visual inspection***

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

Self-Organizing Maps (SOM) general overview

Solar Wind properties

- V_{sw}
- Density
- O^{7+}/O^{6+}
- C^{6+}/C^{5+}
- Fe/O
- $v_{He2+} - v_H$
- Plasma β
- Temperature
- σ_c : cross helicity
- σ_r : residual energy
- S_p: proton specific entropy
- V_A: Alfvén velocity
- T_{exp}/Temperature ratio
- $B_r - B_t$
- Density range
- B_t range
- B_n range

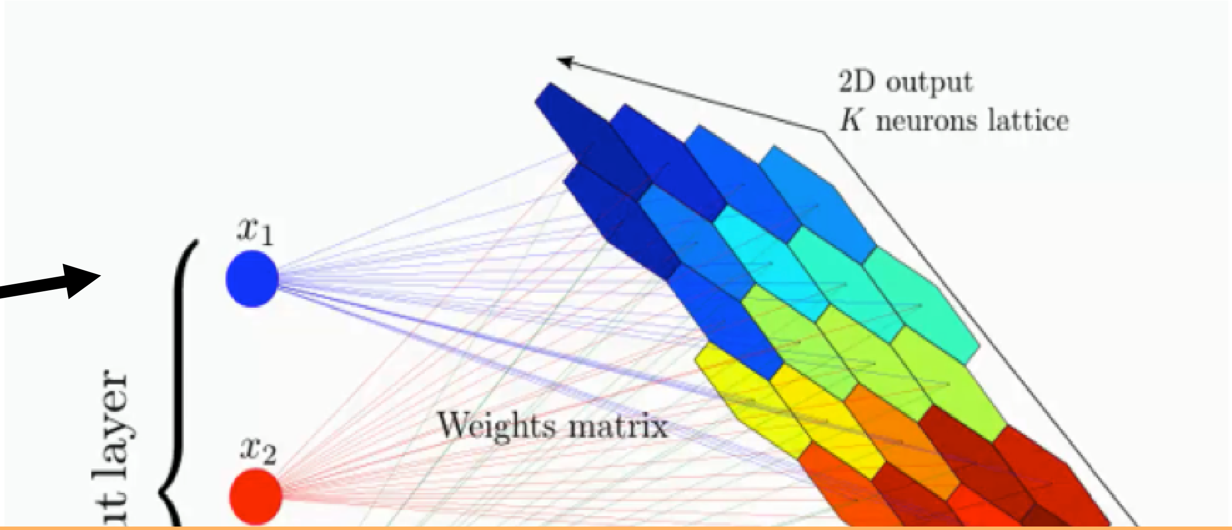


Kohonen, T. (1982). *Biological cybernetics*, *43*(1), 59-69.

Self-Organizing Maps (SOM)

nd properties

- V_{sw}
- Density
- O^{7+}/O^{6+}
- C^{6+}/C^{5+}
- Fe/O
- $v_{He2+} - v_H$
- Plasma β
- Temperature
- σ_c : cross helicity
- σ_r : residual energy



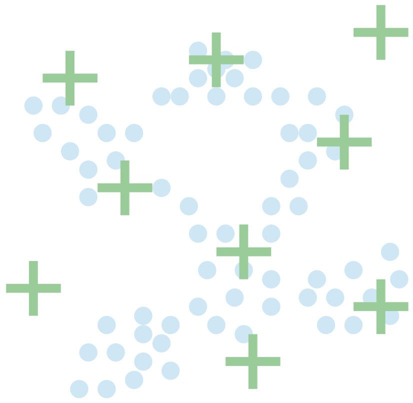
The goal and advantages of Self-Organazing Maps (SOM):

- **Reduce the data**: project N-dimensional points onto a 2D map
- **Group the data**: cluster data points around representative nodes
- **Visualize and understand the data**: maintaining topology information on the reduced space of the map

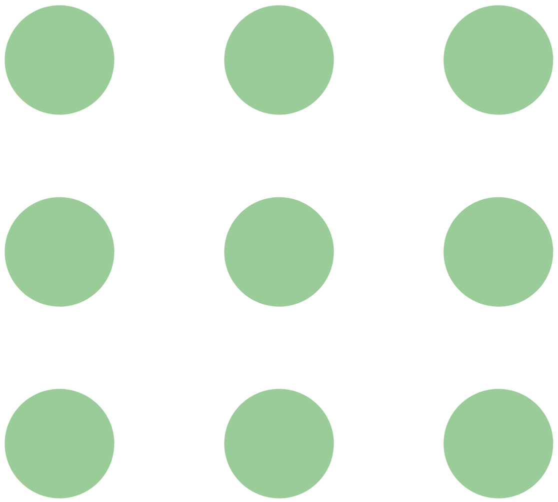
Kohonen, T. (1982). *Biological cybernetics*, *43*(1), 59-69.

SOM: how they learn?

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



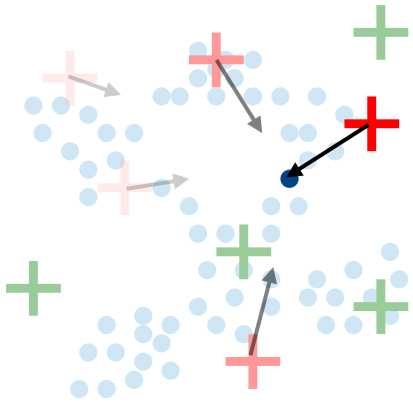
Feature ND-space



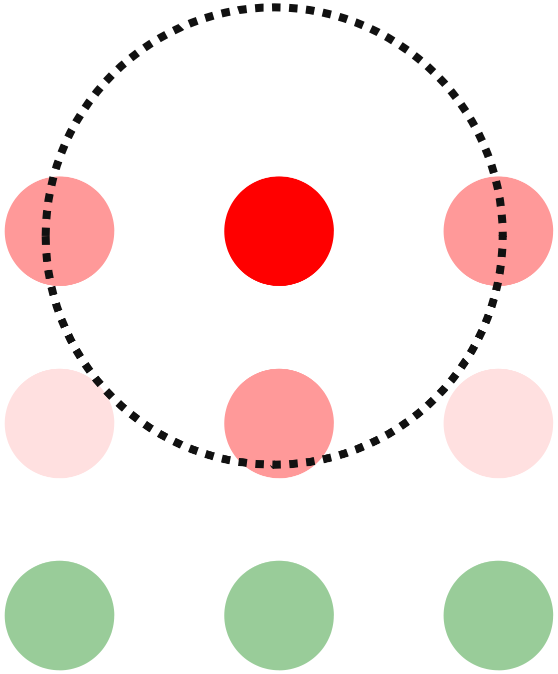
SOM

SOM: how they learn?

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



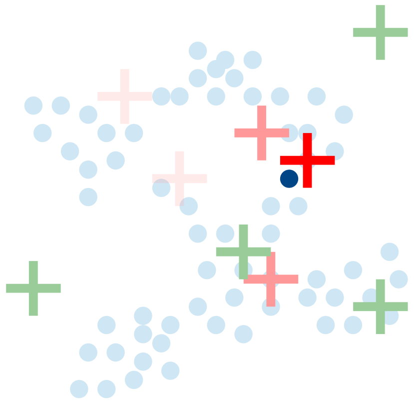
Feature ND-space



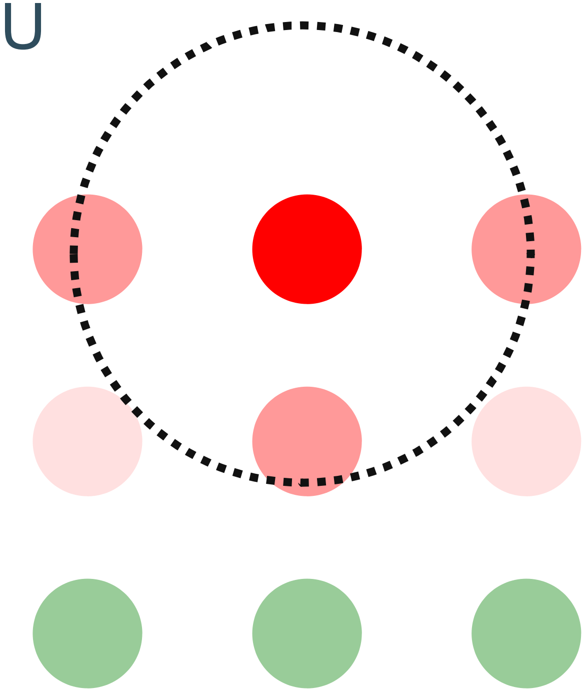
SOM

SOM: how they learn?

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



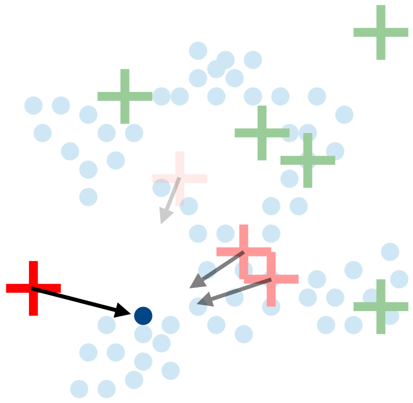
Feature ND-space



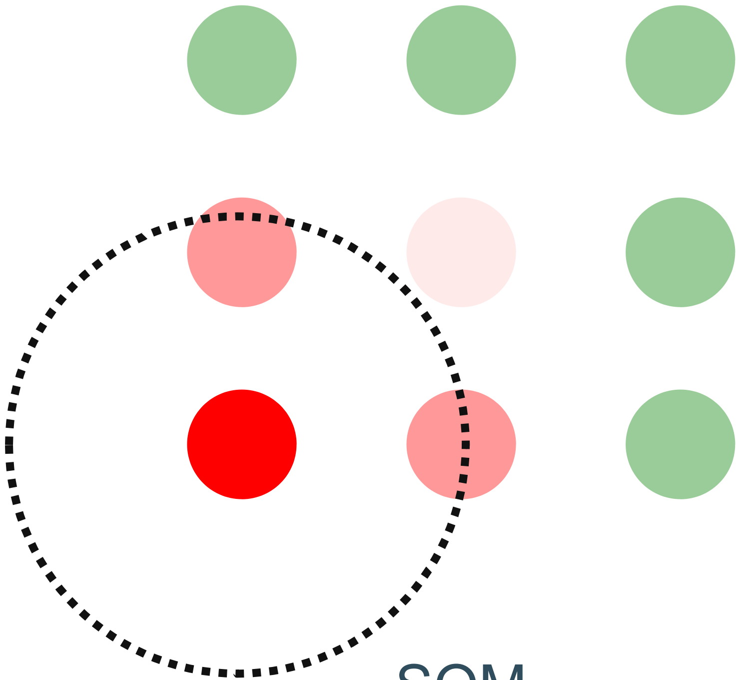
SOM

SOM: how they learn?

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



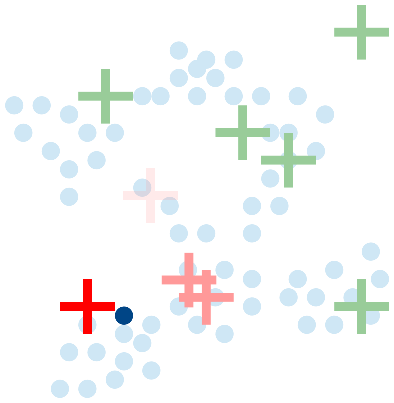
Feature ND-space



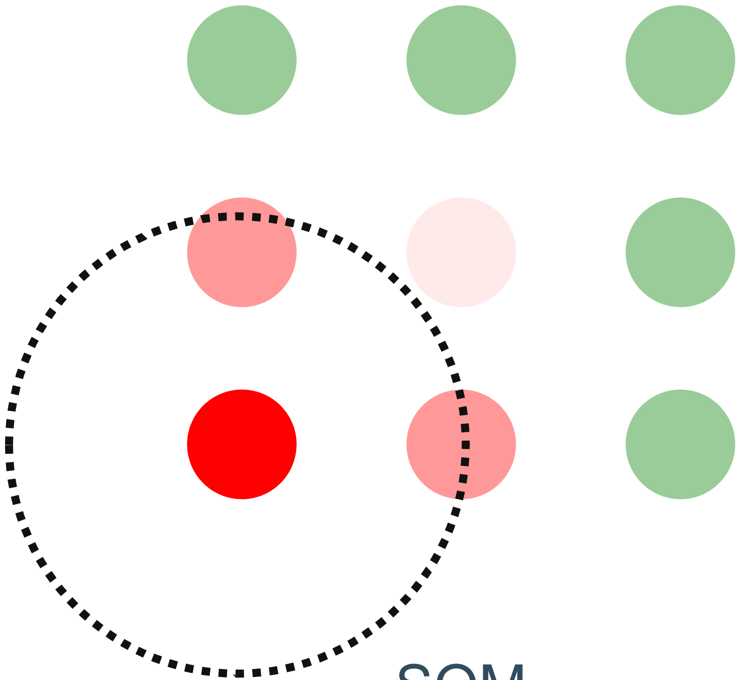
SOM

SOM: how they learn?

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



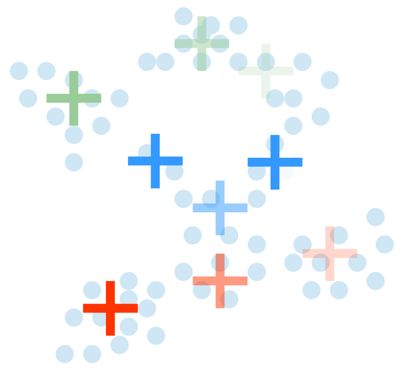
Feature ND-space



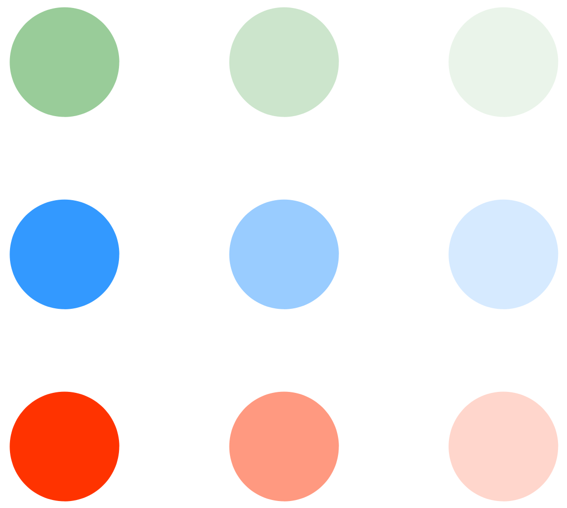
SOM

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.



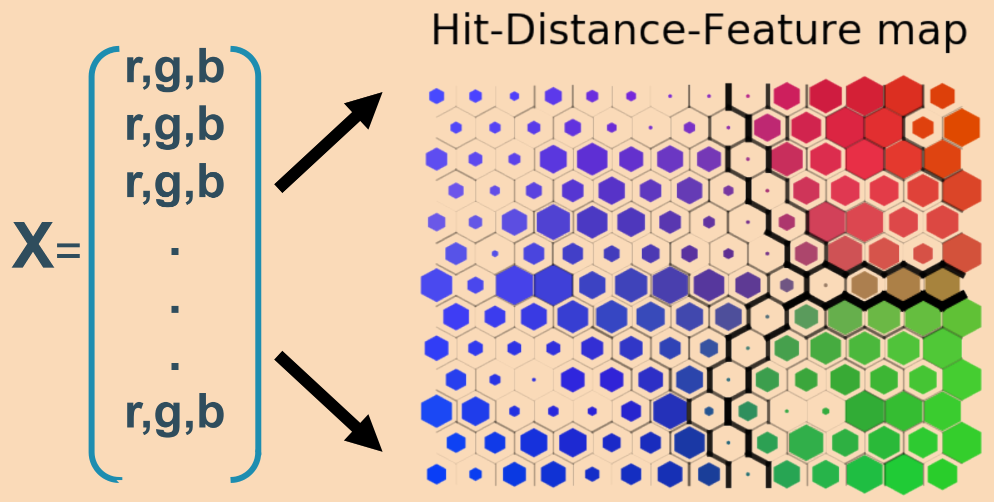
Feature ND-space



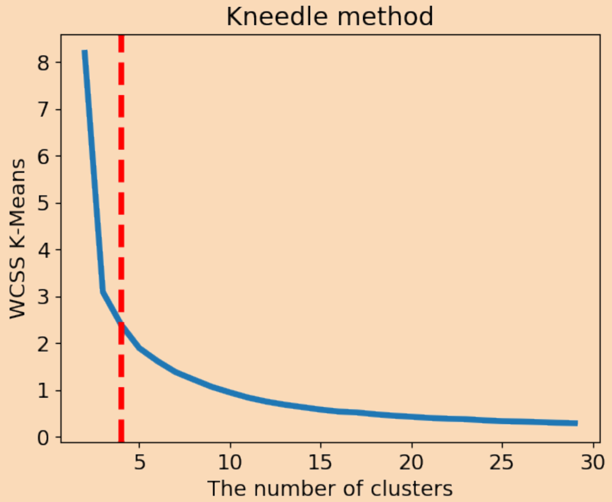
SOM

SOM example: random list of RGB colors

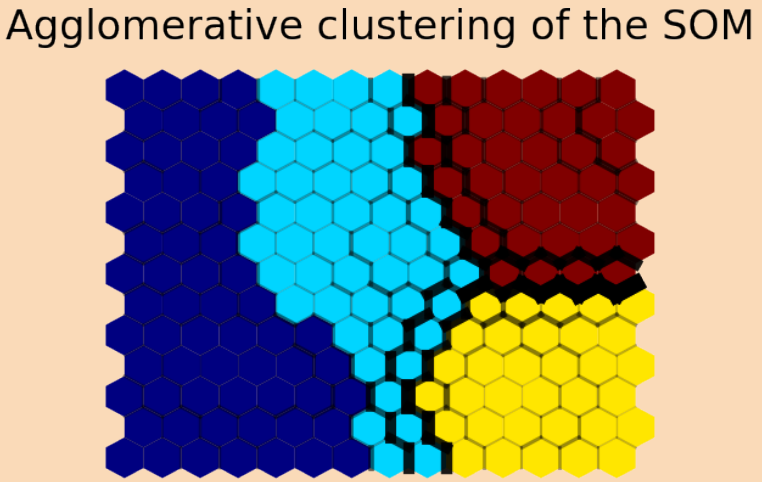
- RGB(0.2, 0.5, 0.77) =
- 6000 points randomly distributed around three selected colors.



step 1



step 2



step 3

Organize in SOM nodes → **Automatically select the # of classes** → **Group nodes that are similar**

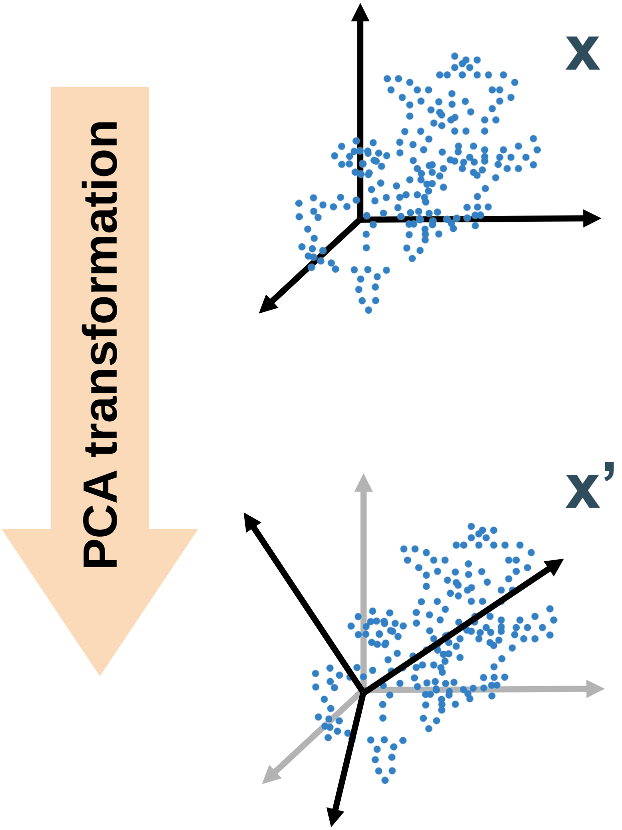
Using SOM on solar wind data

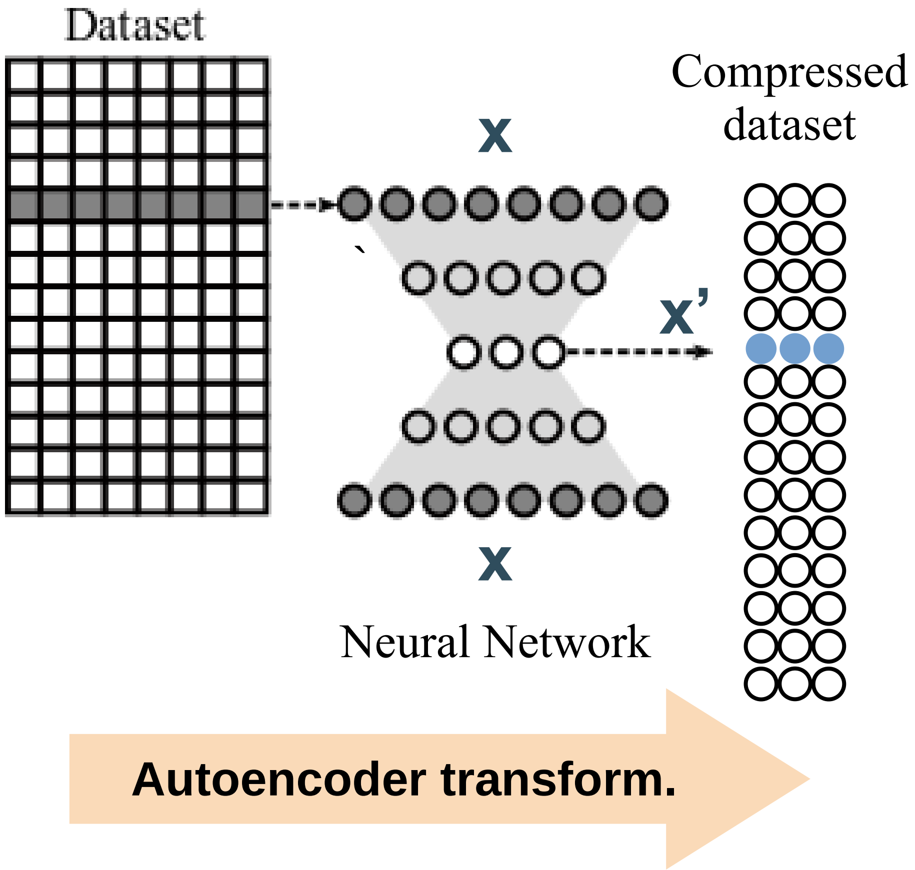
- ACE data from 1998 to 2011:
- 72K points with 17 features and derived properties

$$
\mathbf{x} = \begin{pmatrix}
V_{sw} \\
\text{Density} \\
O^{7+}/O^{6+} \\
C^{6+}/C^{5+} \\
\text{Fe/O} \\
v_{He2+} - v_{H} \\
\text{Plasma } \beta \\
\text{Temperature} \\
\sigma_c \text{ : cross helicity} \\
\sigma_r \text{ : residual energy} \\
S_p\text{: proton specific entropy} \\
V_A\text{: Alfvén velocity} \\
T_{exp}\text{/Temperature ratio} \\
B_r - B_t \\
\text{Density range} \\
B_t \text{ range} \\
B_n \text{ range}
\end{pmatrix}
$$

Data transformation: using better data

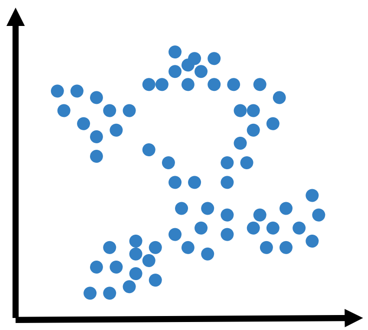
- We can transform the data into a more “useful” space.
- Two possible methods: PCA and Autoencoders

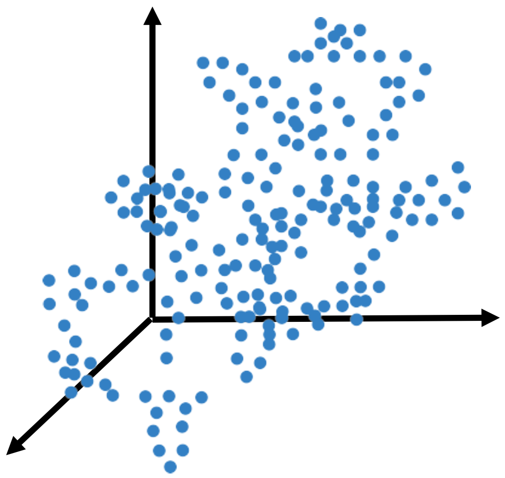


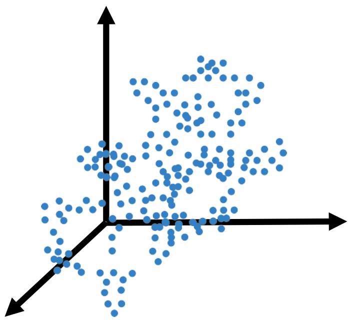


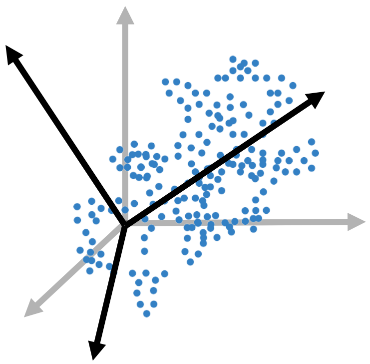
ACE data pre-processing pipeline

Read ACE Data	Enhanced Derived Data	Data Normalization	Transformation

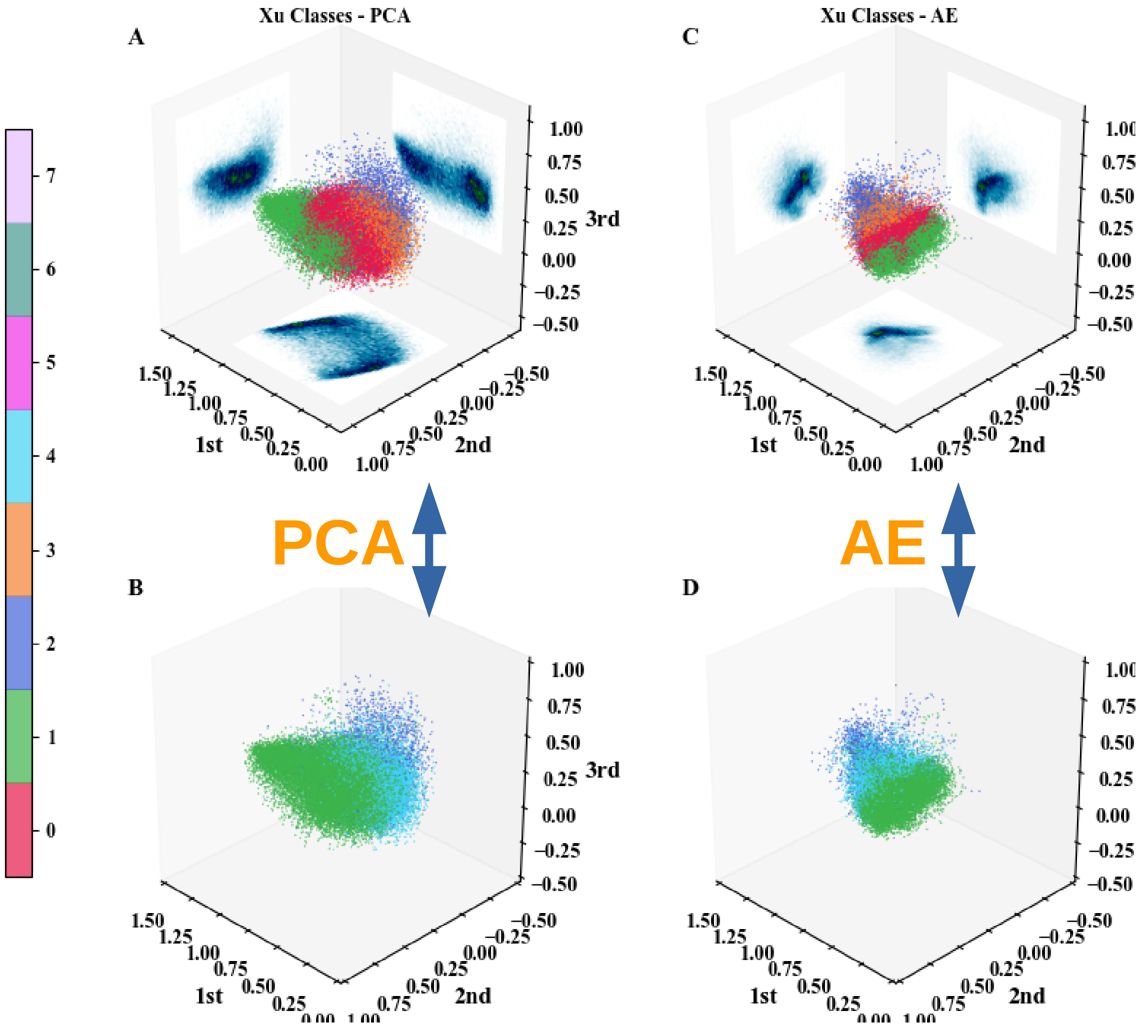








Clouds of points: dimenision reduction



Xu Classification

Zhao Classification

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

Hyper-parameters (automatically optimized)

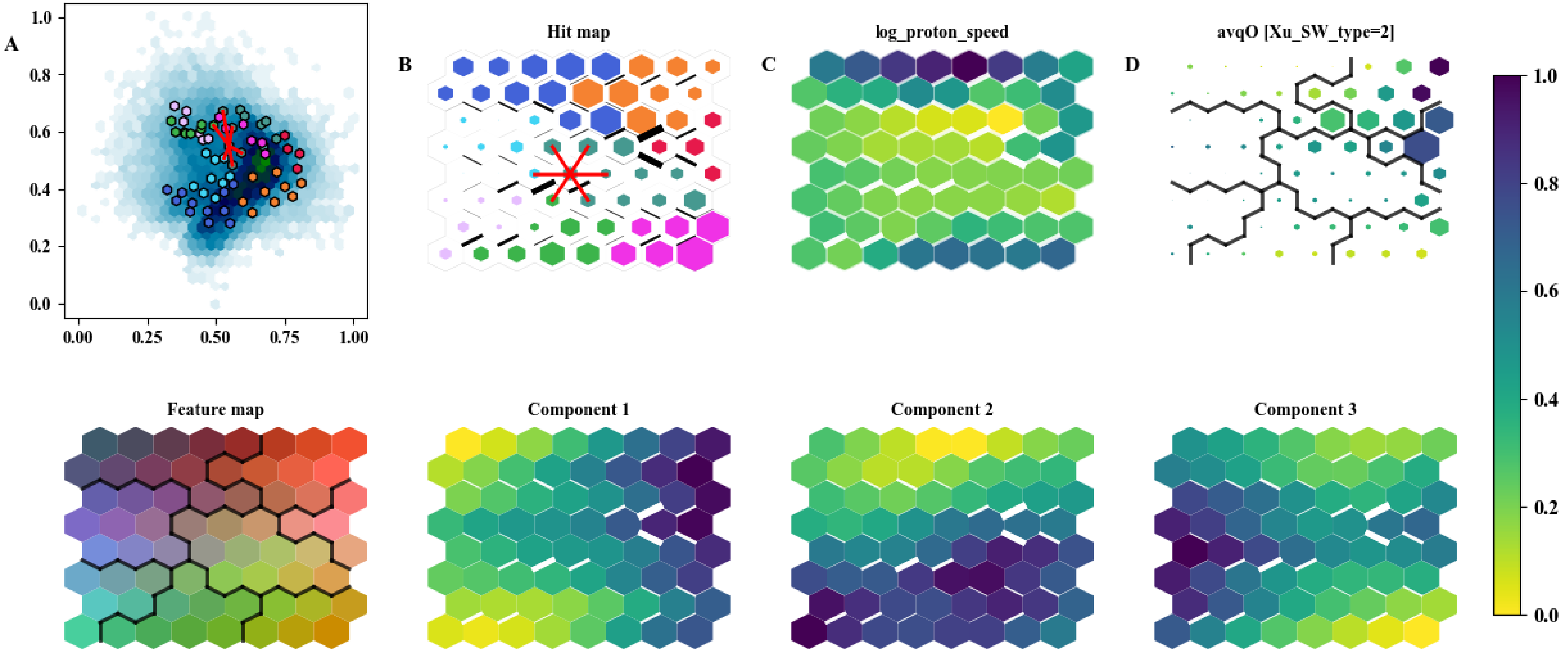
We used bayesian optimization to determine automatically the following parameters:

- Map size: 12 x 12
- Learning rate: 0.13
- Elasticity: 4.42

The values bellow were set-up manually:

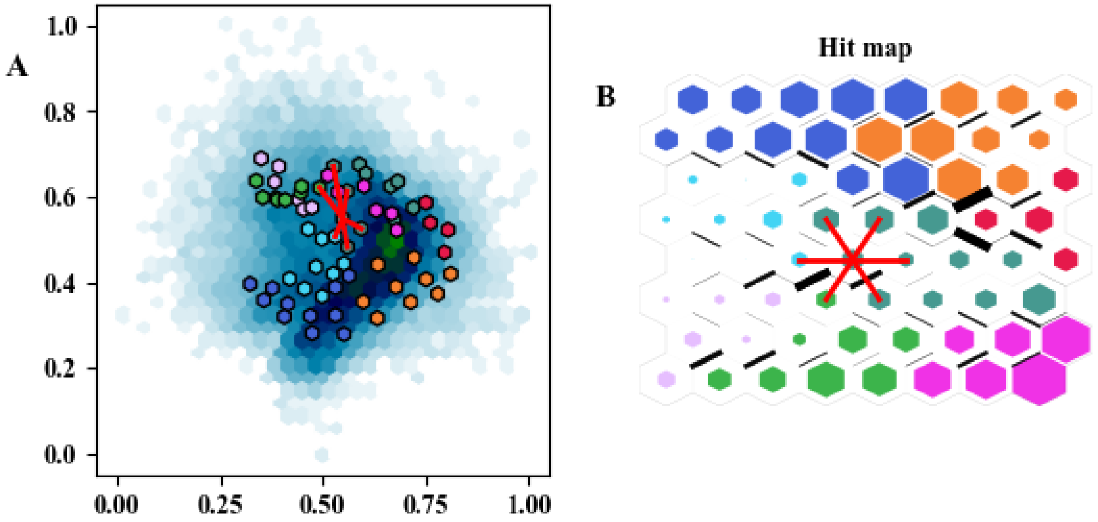
- 20000 epochs
- 17 features
- Compression bottleneck: 3

The maps: some examples of visualization



KU LEUVEN

The maps: feature space and map nodes

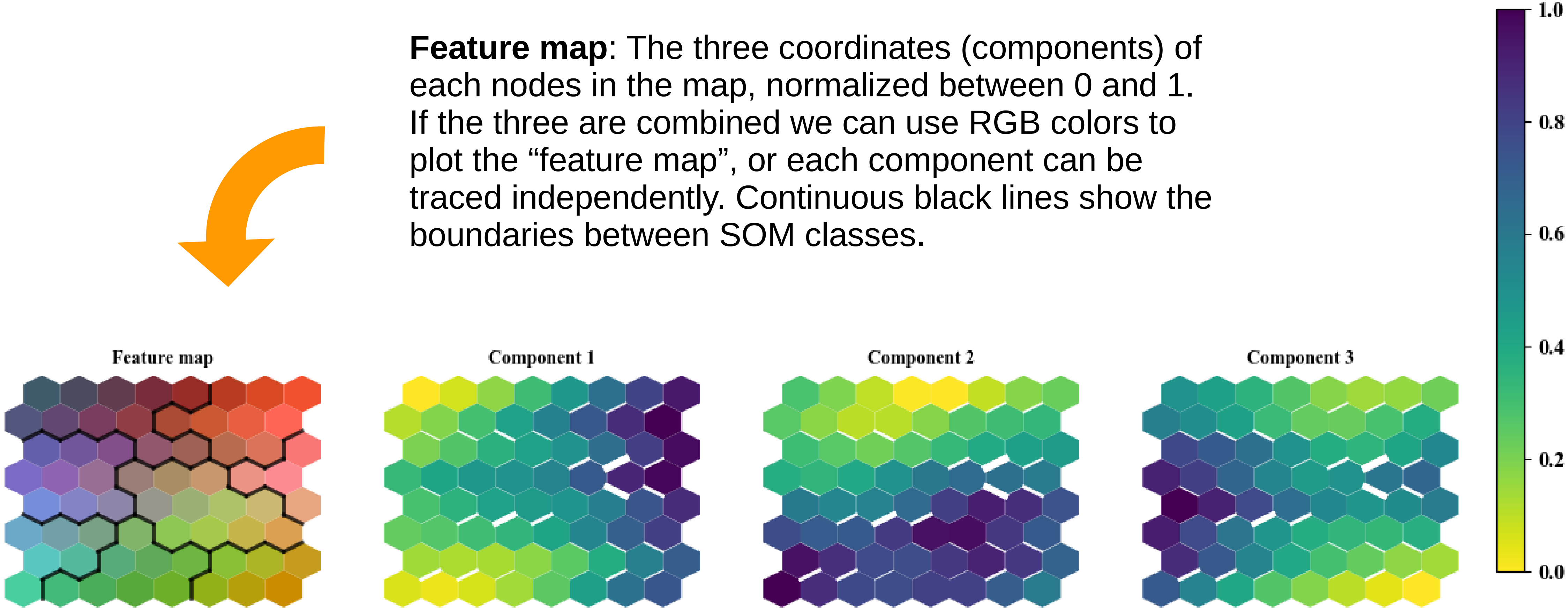


Hit map: each map “node” corresponds to a “code word” in the data space. The size of the node represents the number of solar wind points that belong to the node. The color represents the SOM class automatically detected. The thickness of the black lines represent the relative distance between neighbouring nodes.

Data space and “code words”: histogram of all solar wind points projected on the first two components of the transformed (compressed, reduced, latent) space. Distribution of the SOM nodes among all the solar wind points.

The maps: some examples of visualization

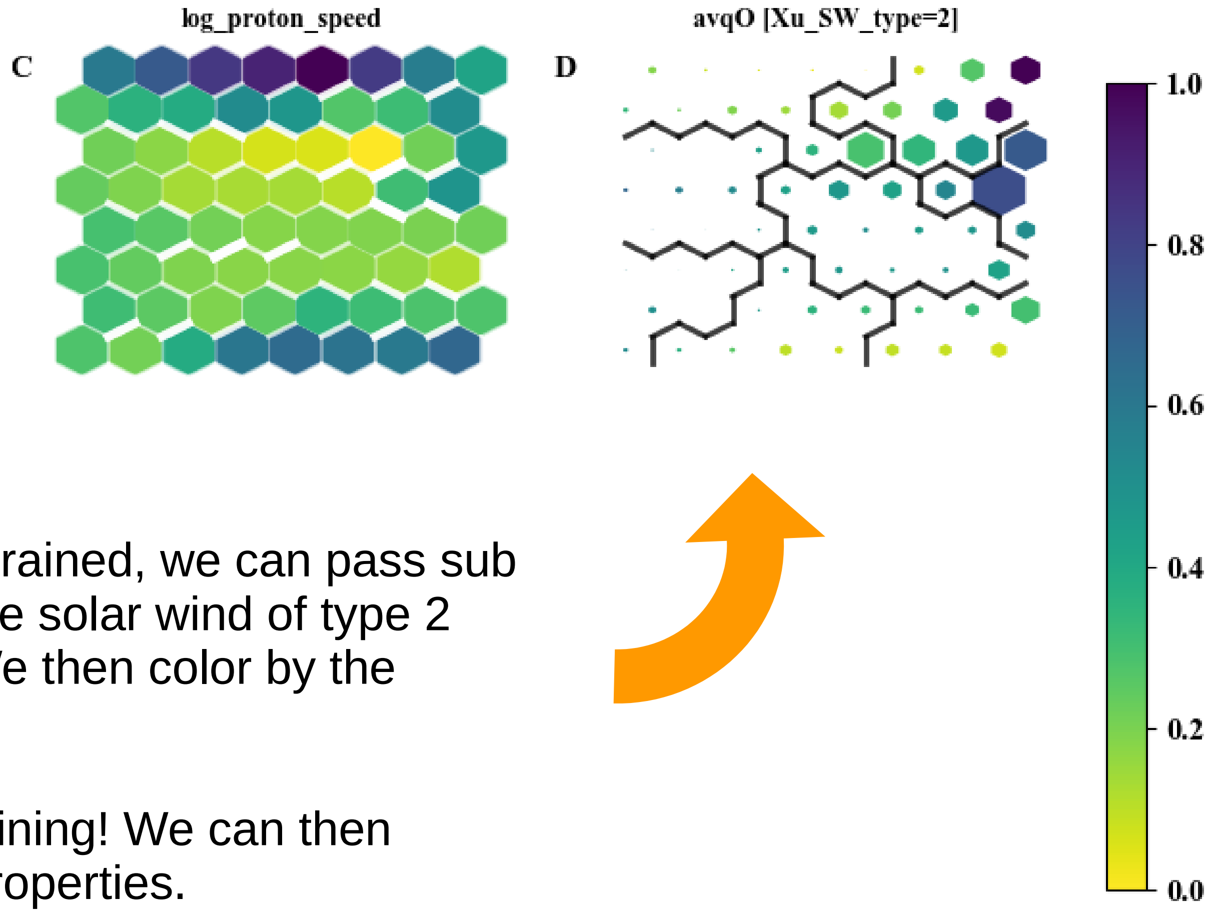
Feature map: The three coordinates (components) of each nodes in the map, normalized between 0 and 1. If the three are combined we can use RGB colors to plot the “feature map”, or each component can be traced independently. Continuous black lines show the boundaries between SOM classes.



The maps: some examples of visualization

Properties map: if we de-code the components we can obtain the original solar wind features, and plot each one of them.

Each node is then characterized by a particular set of solar wind properties



Sub set data and alternative visualizations: once trained, we can pass sub sets of the data to study. In this example we select the solar wind of type 2 from Xu. This sub set activates certain nodes only. We then color by the property 'avqO' (average oxygen charge).

Notice that none of these values was used for the training! We can then visually identify class activations and their physical properties.

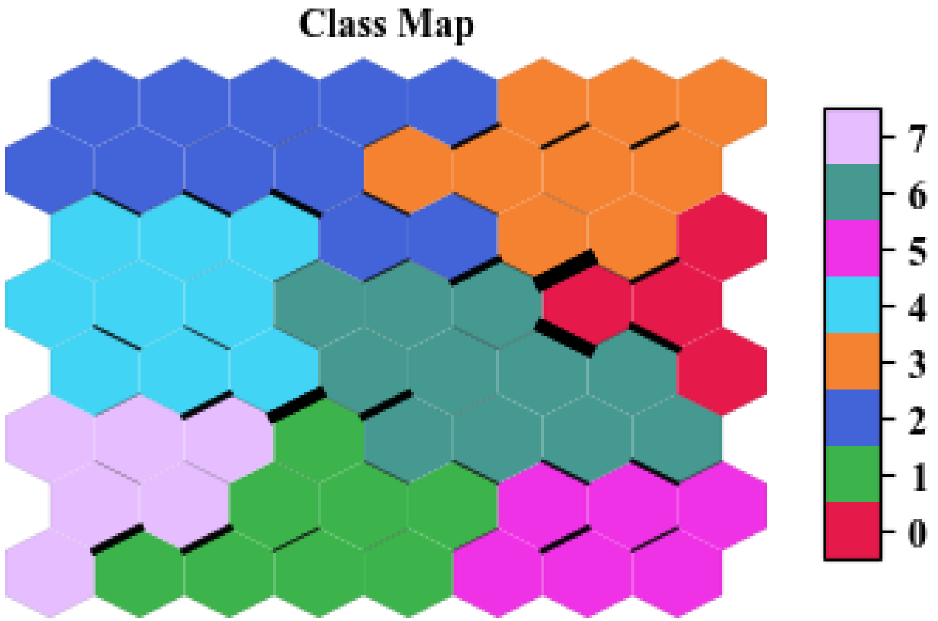
Xu, F., & Borovsky, J. E. (2015). JGR: Space Physics

Node clustering

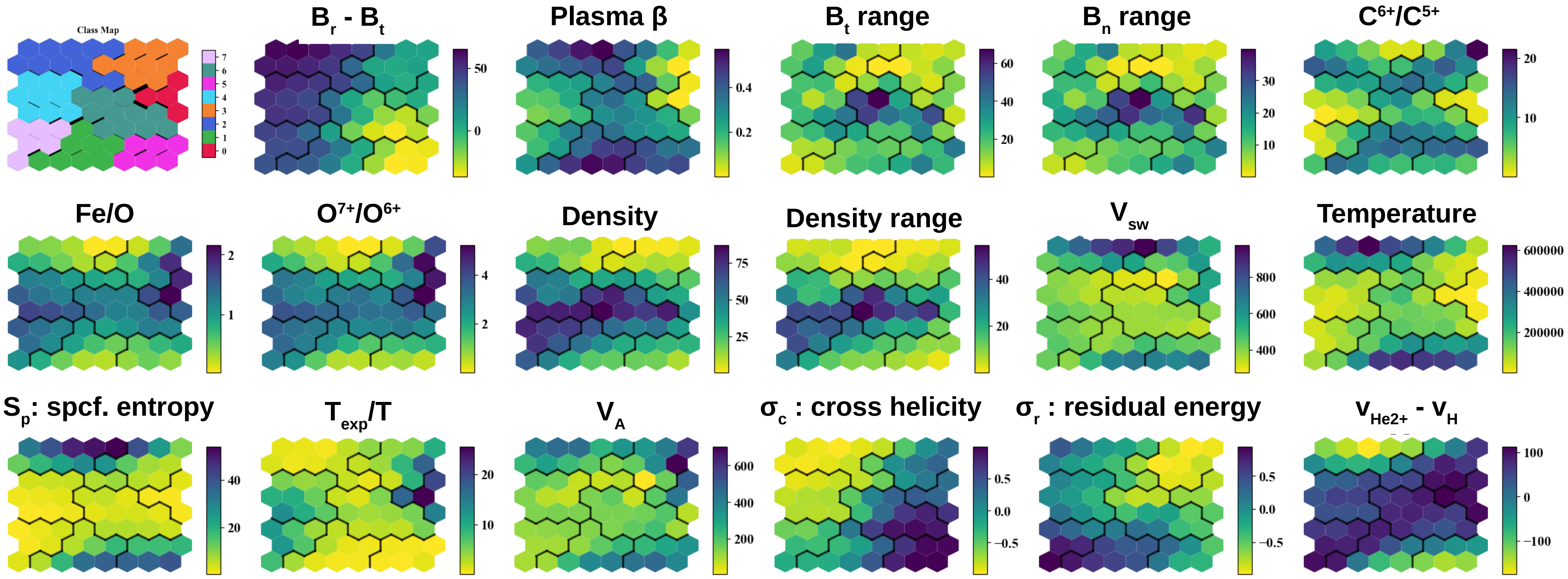
SOM classes: using k-means we cluster the nodes to obtain 8 different classes.

The number of classes has been given, but automatic techniques exist to make a selection of the correct number.

Notice that each class is contiguous in the map, and that node distances give a good indication of the boundaries.

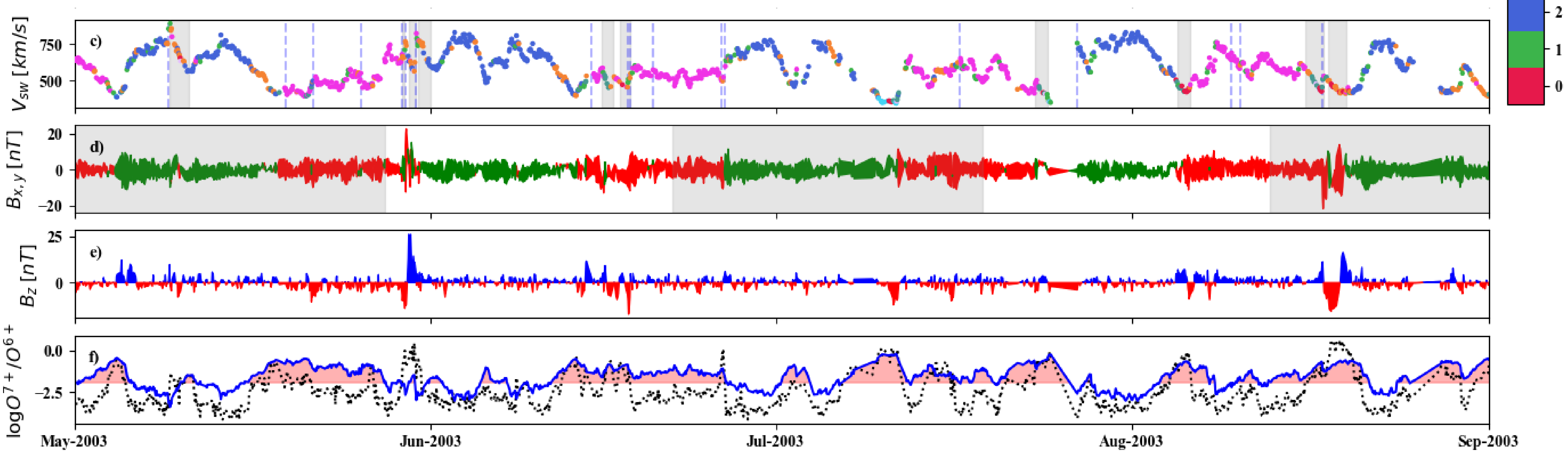


Feature maps



Time series

Window of 4 months were the classified data is plotted, including the solar wind speed, the IMF polarity and the O7+/O6+ ratio. This last one contains the limits of the Zhao classification for ICME (dots inside the red zone), coronal hole (dots above the red zone) and non-coronal hole origins (dots bellow the red zone). ICME and shock manual catalogues are used in the top panel to compare with SOM classes.



Zhao, L., Zurbuchen, T. H., & Fisk, L. A. (2009). GRL, 36

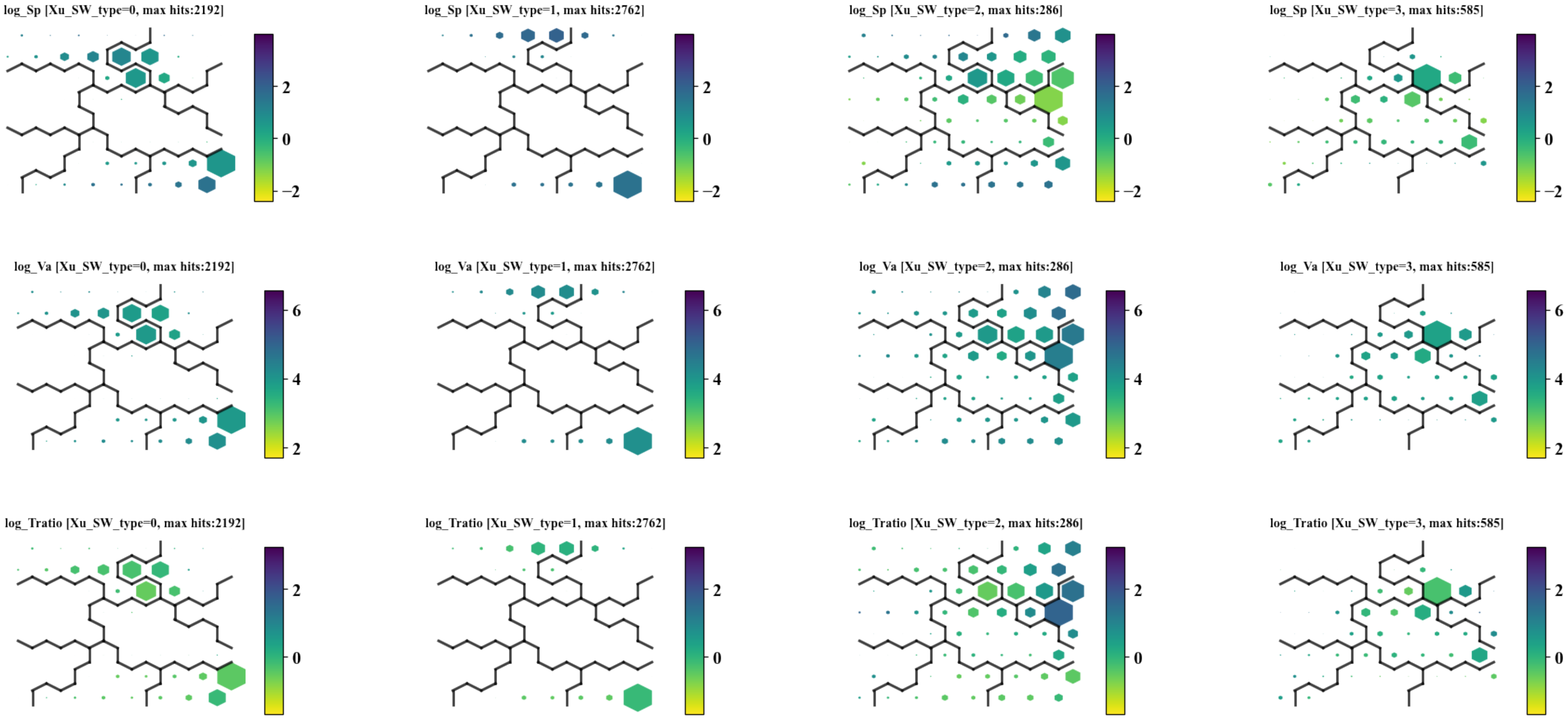
Comparison with Xu classification

Streamer belt | **Cor. Hole** | **Ejecta** | **Sector revers.**

$\log(S_p)$

$\log(V_A)$

$\log(T_{exp}/T)$



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

On the works

- Automatic selection of the number of classes
- Precisse analysis of each one of the detected classes
- Verification on multiple time series
- Precise comparison with different classification methods
- Publication of a physical interpretation of each one of the classes automatically detected using the procedure presented here.
- Application of the SOM technique to other heliophysics data.

Unsupervised classification of the solar wind using Self-Organising Maps

Jorge AMAYA
Romain DUPUIS
Maria Elena INNOCENTI
Giovanni LAPENTA

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