Influence of solar wind parameters on unsupervised solar wind classification

S. Teichmann [1], V. Heidrich-Meisner [2] , R. Wimmer-Schweingruber [2]
[1] *Georg-August-Universität Göttingen,* [2] *Christian-Albrechts Universität zu Kiel*

ML Helio
23.03.2022

Solar wind and solar wind classification

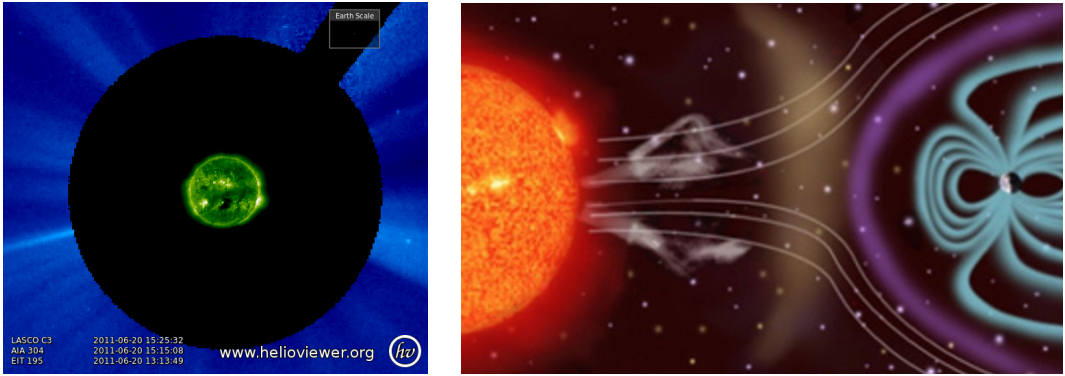


Figure: Left: Instruments on SOHO (Solar and Heliospheric Observatory): LASCO (Large Angle and Spectrometric Coronagraph) and EIT (Extreme ultraviolet Imaging Telescope), Right: Schematic display of the Sun, solar wind and Earth by ESA [2]. Visualisation of the 3 motivation for solar wind classification: understanding the origin, transport effects and space weather.

Solar wind classification - transport effects

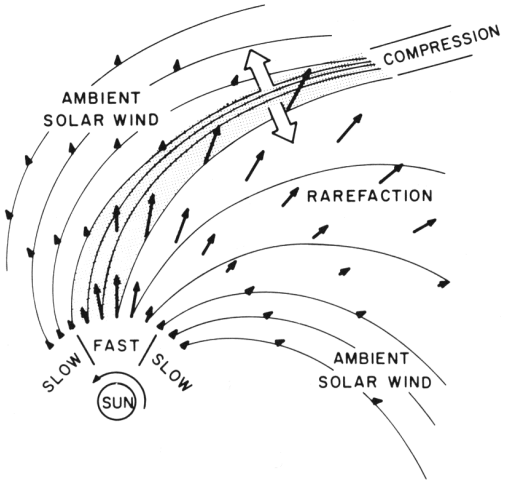


Figure: Visualisation of Stream interaction regions by Jian [3]

Parameters

density (n_{sw}) mean density of protons in the solar wind

velocity (v_{sw}) mean velocity of protons in the solar wind

temperature(T_{sw}) mean temperature of protons in the solar wind

magnetic field (B) absolute value of the magnetic field

collisional age (a_{col}) number of collisions in the plasma

oxygen ions ($n_{O^{7+}}/n_{O^{6+}}$) ratio between the densities O^{7+} and O^{6+}

Iron ions (q_{Fe}) mean charge state of iron

instruments on ACE:

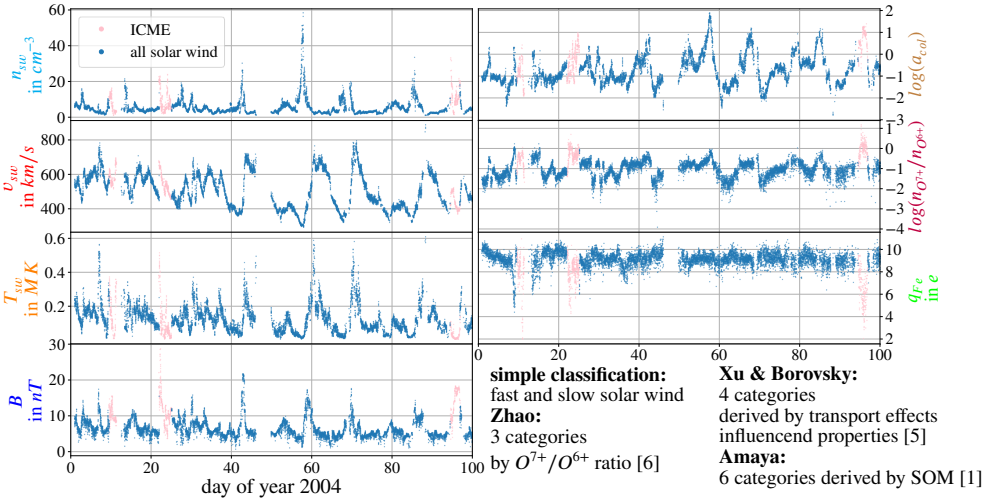
- Solar Wind Electron Proton and Alpha Monitor (SWEPAM)
- Solar Wind Ion Composition Spectrometer (SWICS)
- Magnetometer (MAG)

timeframe : 2001-2011

number of points in dataframe: 258574 (with ICME: 282231)

Figure: Logo of the Ace Mission [4]

Existing classifications



simple classification:
fast and slow solar wind
Zhao:
3 categories
by O^{7+}/O^{6+} ratio [6]

Xu & Borovsky:
4 categories
derived by transport effects
influencend properties [5]
Amaya:
6 categories derived by SOM [1]

⇒ transport effect vs. origin based classifications

How it's done: k-means clustering

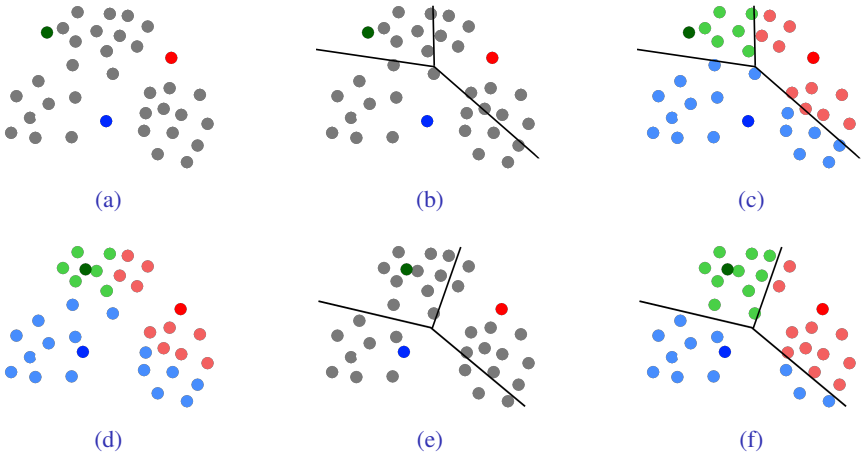
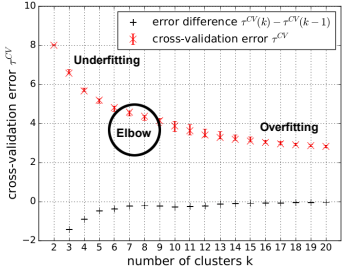


Figure: Visualisation of k-means clustering by Heidrich-Meisner. Implementation: sklearn version 0.23.2 in python version 3.9.2

How to choose k or on the number of solar wind types



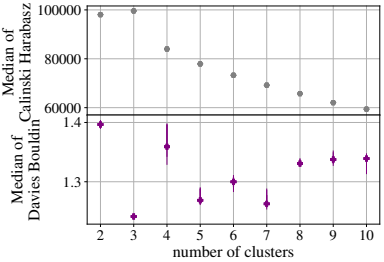


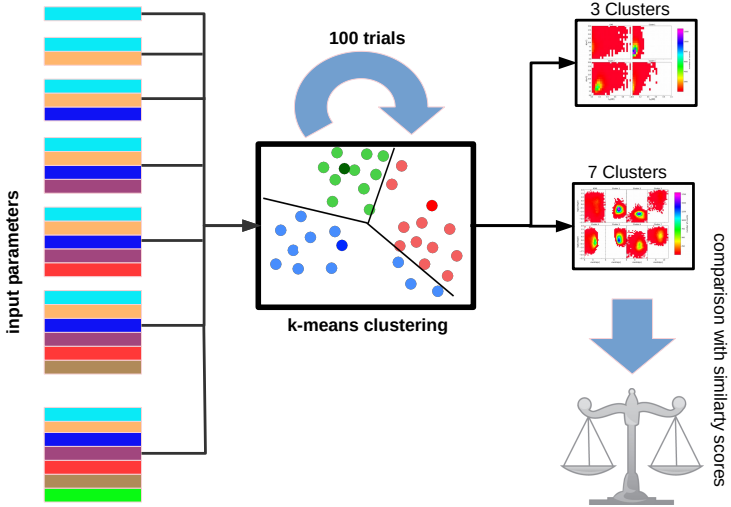
Figure: Left: Elbow plot by Heidrich-Meisner on the same dataset. Right: Elbow plot based on the experimental settings

⇒ resulting cluster are not convex

⇒ **k=3** and **k=7**

- **Cross validation error:** inner cluster distance
- **Calinski Harabasz score:** sum of between-clusters dispersion and of within-cluster dispersion
- **Davies Bouldin score:** average similarity between clusters

The experiment: variation of input parameters



How to evaluate the results

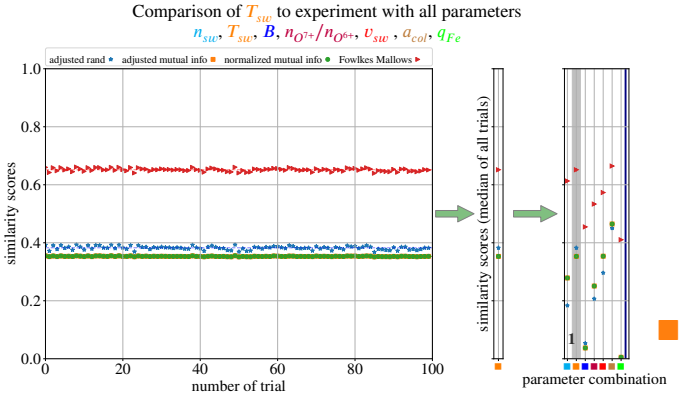
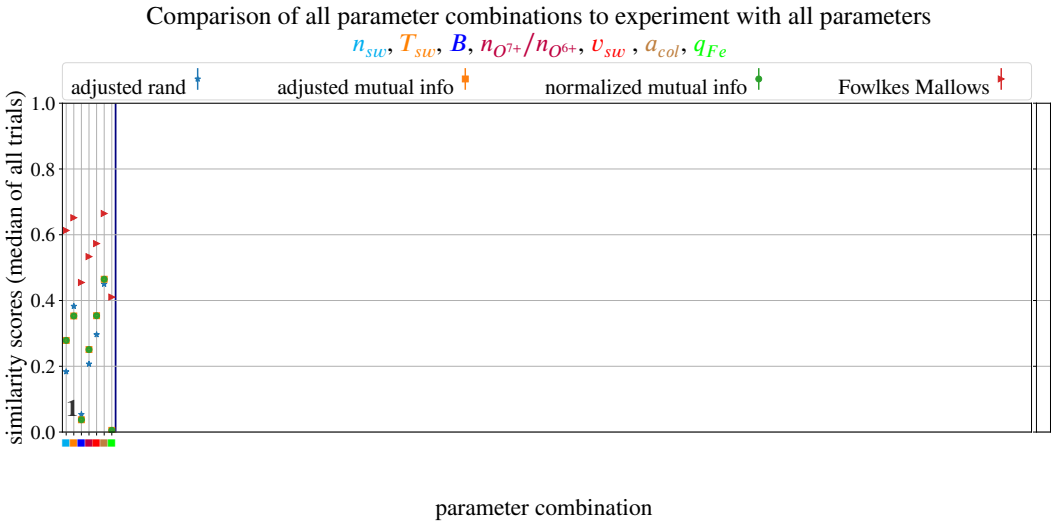


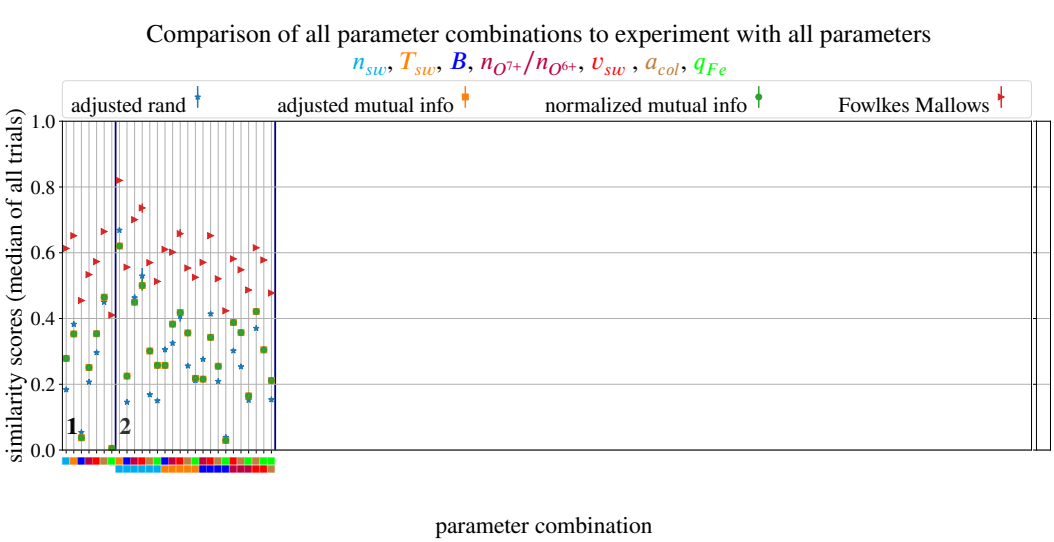
Figure: Schematic demonstration how the results for each parameter combination is evaluated on the example of T_{sw} .

- **Fowlkes Mallows score:** ranges from 0 to 1 based on statistical errors (true positive ...)
- **Adjusted rand score:** ranges from -1 to 1 counting pairs that are the same and differently labelled
- **Mutal information score:** **adjusted** and **normalized**, ranges from 0 to 1, shared information of two clusterings & if adjusted for chance effects

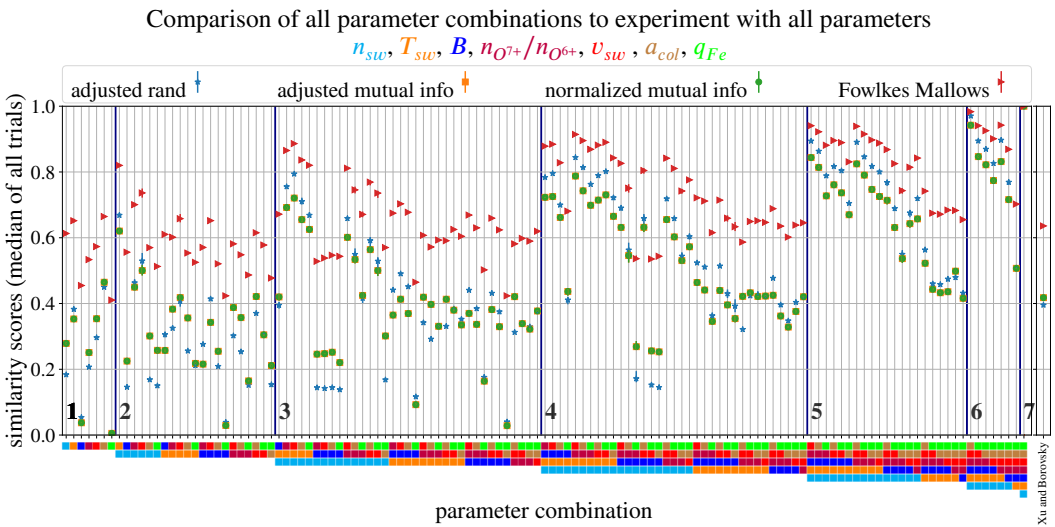
3 Clusters - overview of the results



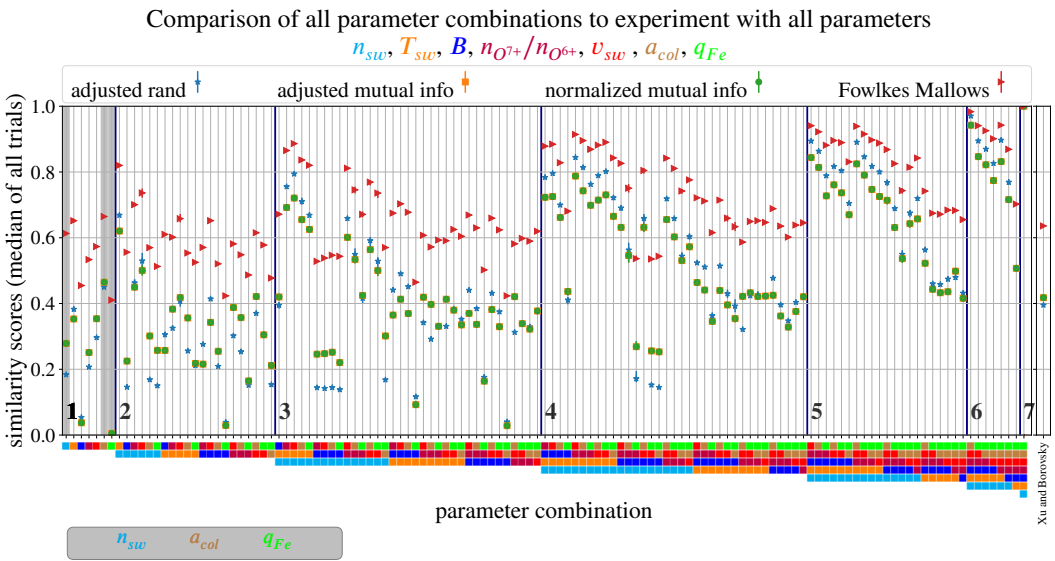
3 Clusters - overview of the results



3 Clusters - overview of the results



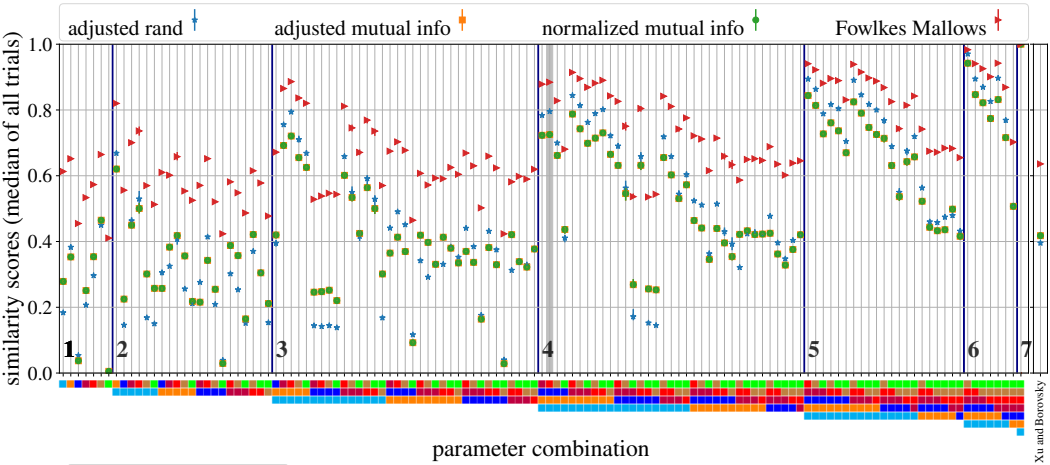
3 Clusters - overview of the results



3 Clusters - overview of the results

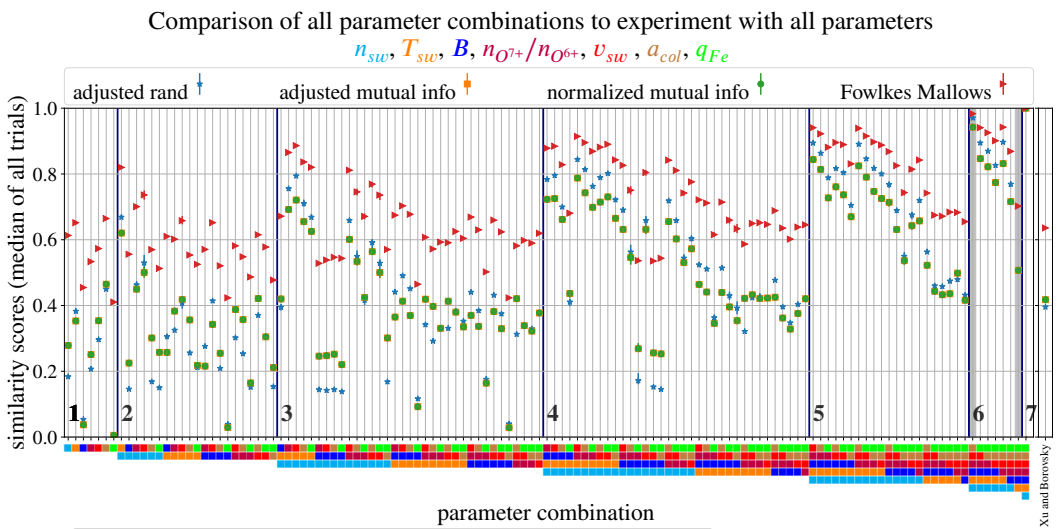
Comparison of all parameter combinations to experiment with all parameters

n_{sw}, T_{sw}, $\boldsymbol{B}$, $n_{O^{7+}}/n_{O^{6+}}$, v_{sw} , a_{col}, q_{Fe}

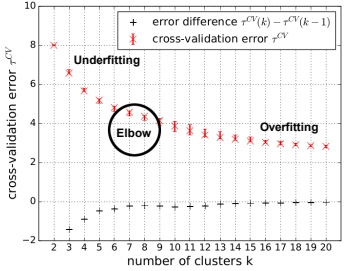


n_{sw},T_{sw},$\boldsymbol{B}$,v_{sw}

3 Clusters - overview of the results



How to choose k or on the number of solar wind types



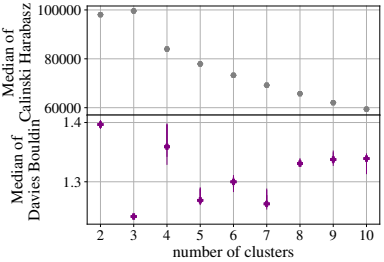


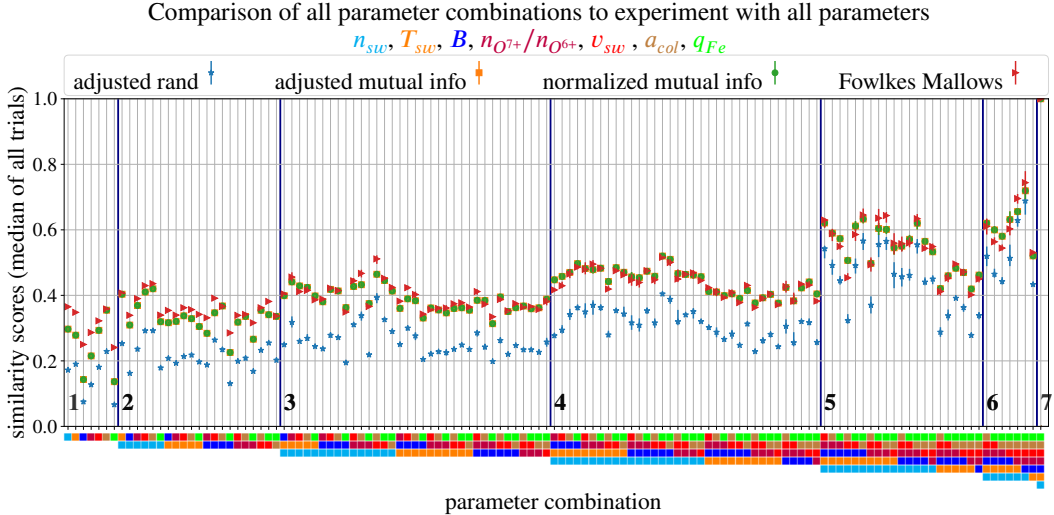
Figure: Left: Elbow plot by Heidrich-Meisner on the same dataset. Right: Elbow plot based on the experimental settings

⇒ resulting cluster are not convex

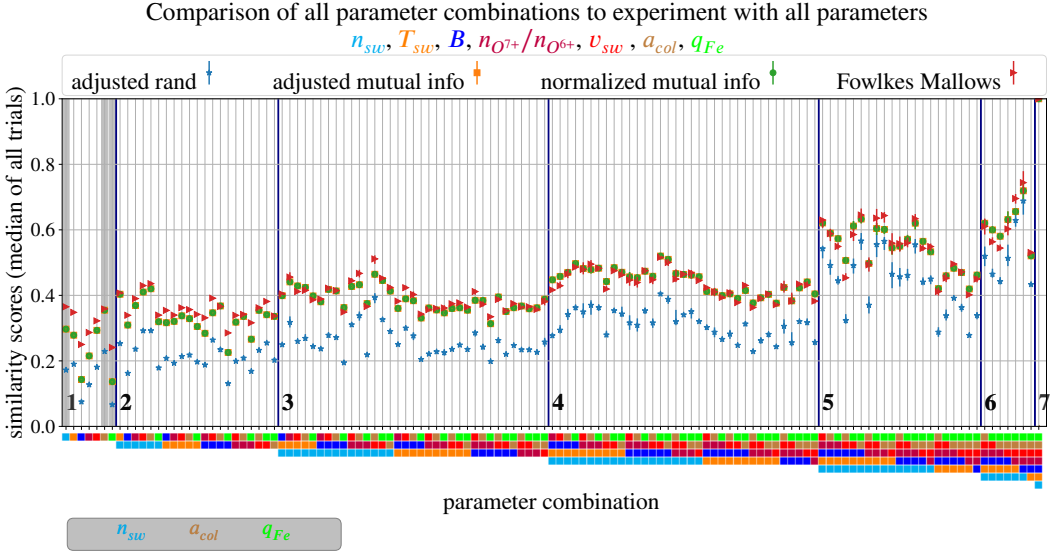
⇒ k=3 and **k=7**

- **Cross validation error:** inner cluster distance
- **Calinski Harabasz score:** sum of between-clusters dispersion and of within-cluster dispersion
- **Davies Bouldin score:** average similarity between clusters

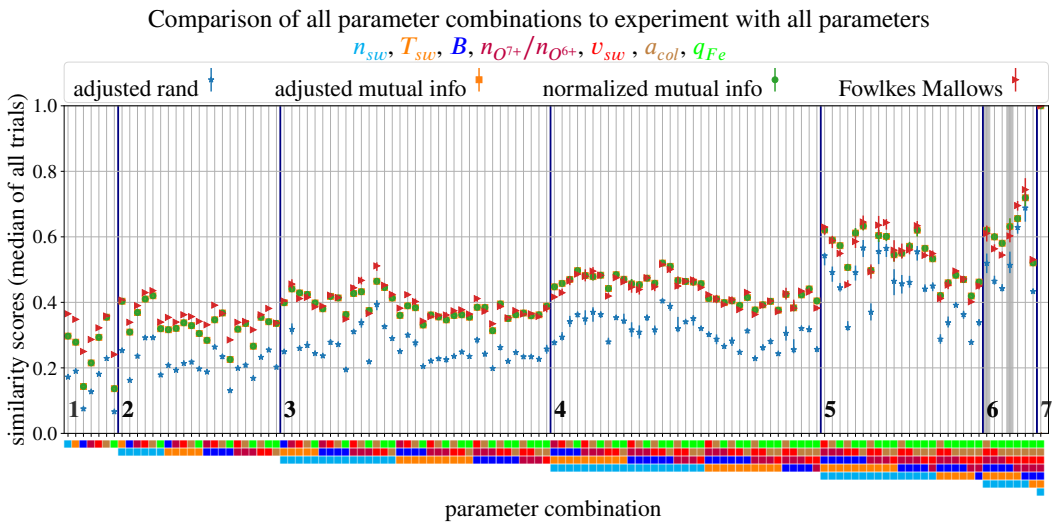
7 Clusters - overview of the results



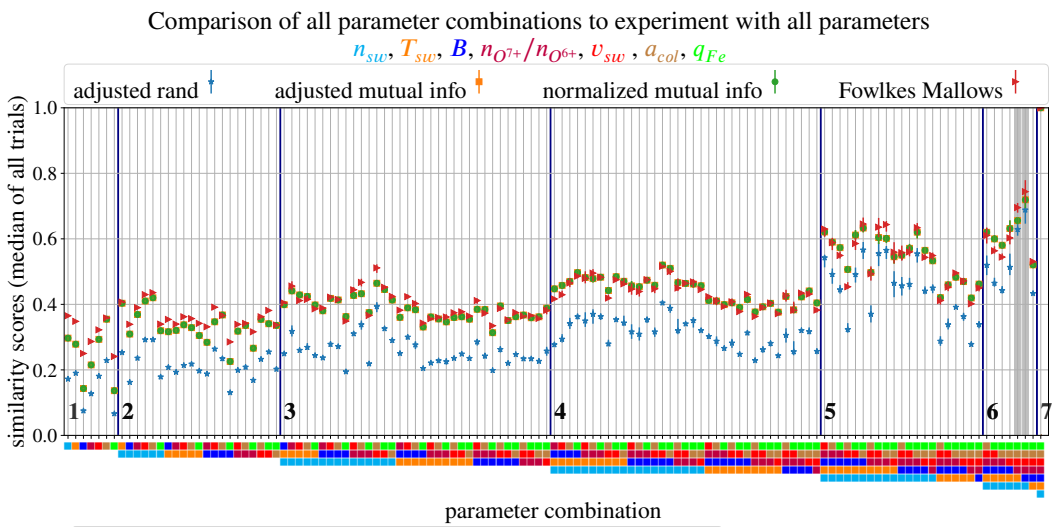
7 Clusters - overview of the results



7 Clusters - overview of the results



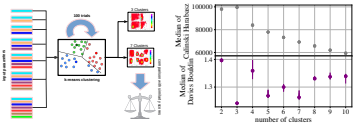
7 Clusters - overview of the results



Summary of the results

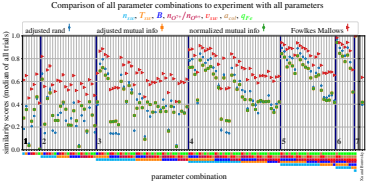
General results from k-means

- more than 3 different types are needed to differentiate between origin based and transport effects
- for 3 Clusters slow solar wind, fast solar wind and compression regions are identified



Conclusion

- n_{sw} is the most important parameter for classification
- charge states (especially q_{Fe}) are needed for detailed classification

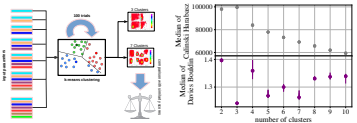


⇒ transport effects should be considered for detailed clustering!

Thank you for your attention & please ask questions !

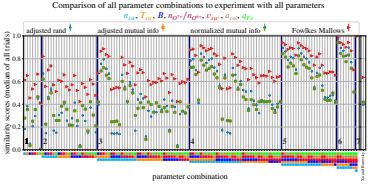
General results from k-means

- more than 3 different types are needed to differentiate between origin based and transport effects
- for 3 Clusters slow solar wind, fast solar wind and compression regions are identified



Conclusion

- n_{sw} is the most important parameter for classification
- charge states (especially q_{Fe}) are needed for detailed classification



⇒ transport effects should be considered for detailed clustering!

References I

J. Amaya, R. Dupuis, M. E. Innocenti, and G. Lapenta.
Visualizing and interpreting unsupervised solar wind classifications.
Frontiers in Astronomy and Space Sciences, 7:66, 2020.

ESA.
Der sonnenwind erzeugt direkte und indirekte effekte auf der erde, 2002.

L. Jian, C. Russell, J. Luhmann, and R. Skoug.
Properties of stream interactions at one au during 1995–2004.
Solar Physics, 239(1):337–392, 2006.

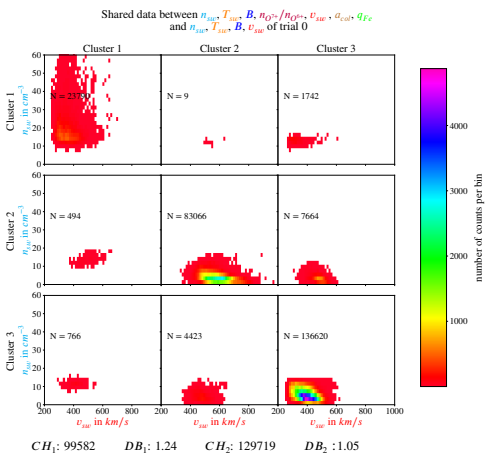
space.
Ace logo, 2022.

References II

F. Xu and J. E. Borovsky.
A new four-plasma categorization scheme for the solar wind.
Journal of Geophysical Research: Space Physics, 120(1):70–100, 2015.

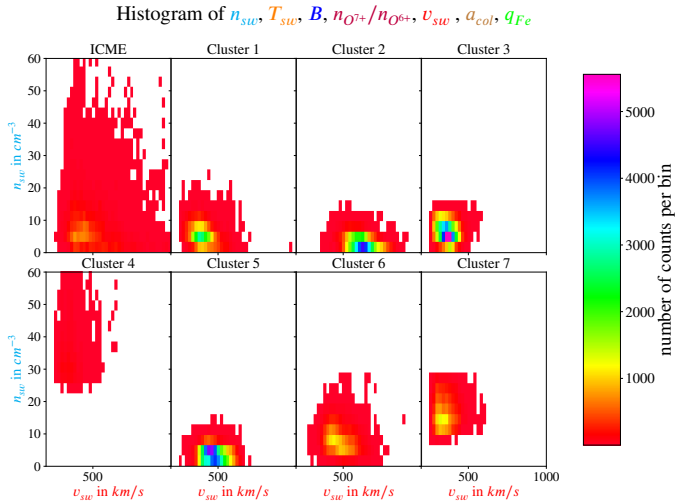
L. Zhao, T. Zurbuchen, and L. Fisk.
Global distribution of the solar wind during solar cycle 23: Ace observations.
Geophysical Research Letters, 36(14), 2009.

3 Cluster - First results

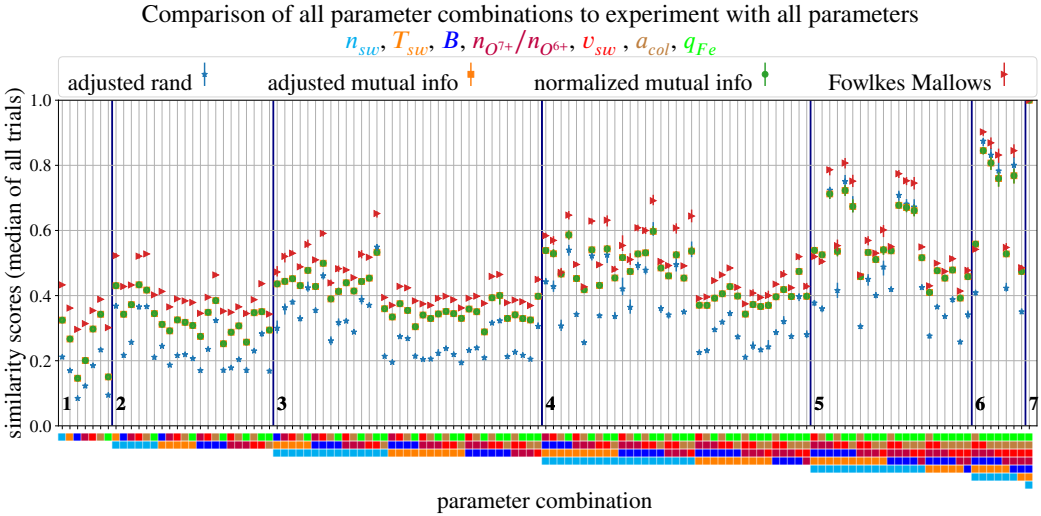


CH_1: 99582 DB_1: 1.24 CH_2: 129719 DB_2 : 1.05

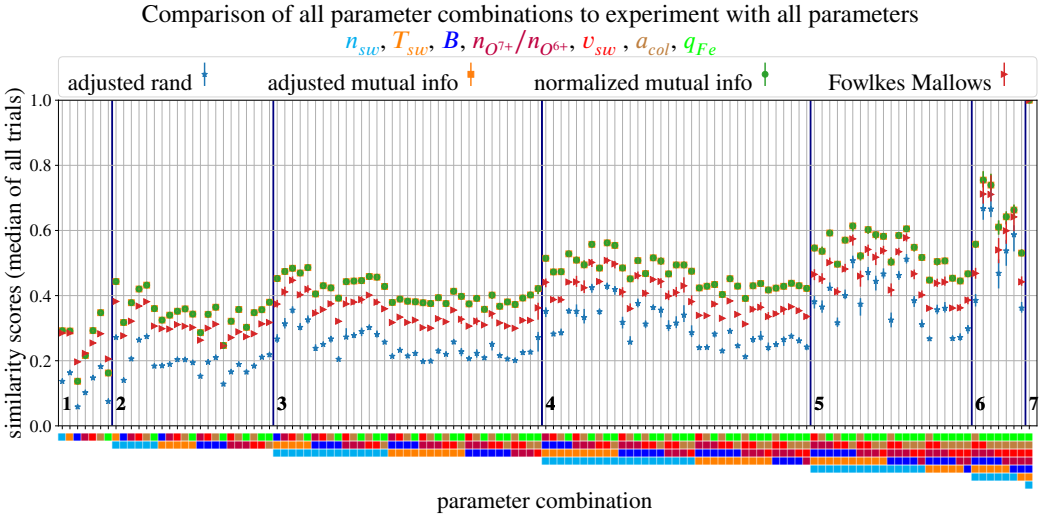
7 Cluster


Histogram of n_{sw}, T_{sw}, B, $n_{O^{7+}}/n_{O^{6+}}$, v_{sw} , a_{col}, q_{Fe}
ICME
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
n_{sw} in cm^{-3}
v_{sw} in km/s
number of counts per bin
CH: 69196
DB: 1.26

6 Cluster


Comparison of all parameter combinations to experiment with all parameters
n_{sw}, T_{sw}, B, $n_{O^{7+}}/n_{O^{6+}}$, v_{sw} , a_{col}, q_{Fe}
adjusted rand
adjusted mutual info
normalized mutual info
Fowlkes Mallows
similarity scores (median of all trials)
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
parameter combination

10 Cluster


Comparison of all parameter combinations to experiment with all parameters
n_{sw}, T_{sw}, B, $n_{O^{7+}}/n_{O^{6+}}$, v_{sw} , a_{col}, q_{Fe}
adjusted rand
adjusted mutual info
normalized mutual info
Fowlkes Mallows
similarity scores (median of all trials)
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
parameter combination