Using machine learning to find a mapping between states of polarization and atmospheric model variables

Mehdi Nourelahi

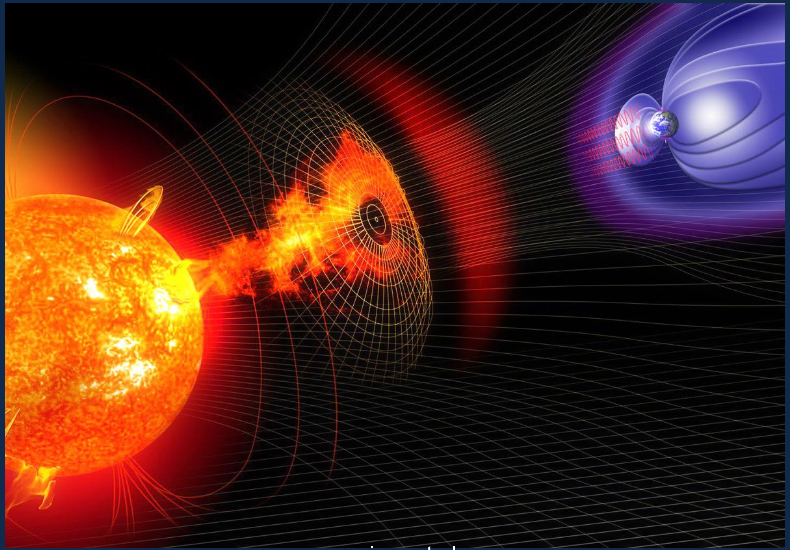
The University of Wyoming, NCAR-High Altitude Observatory intern

Mentors: Ricky Egeland, Rebecca Centeno, and Roberto Casini

NCAR
UCAR

Solar Influences on Earth's Space Environment

- Space weather originated by sun:
  - Flare
  - Coronal Mass Ejections



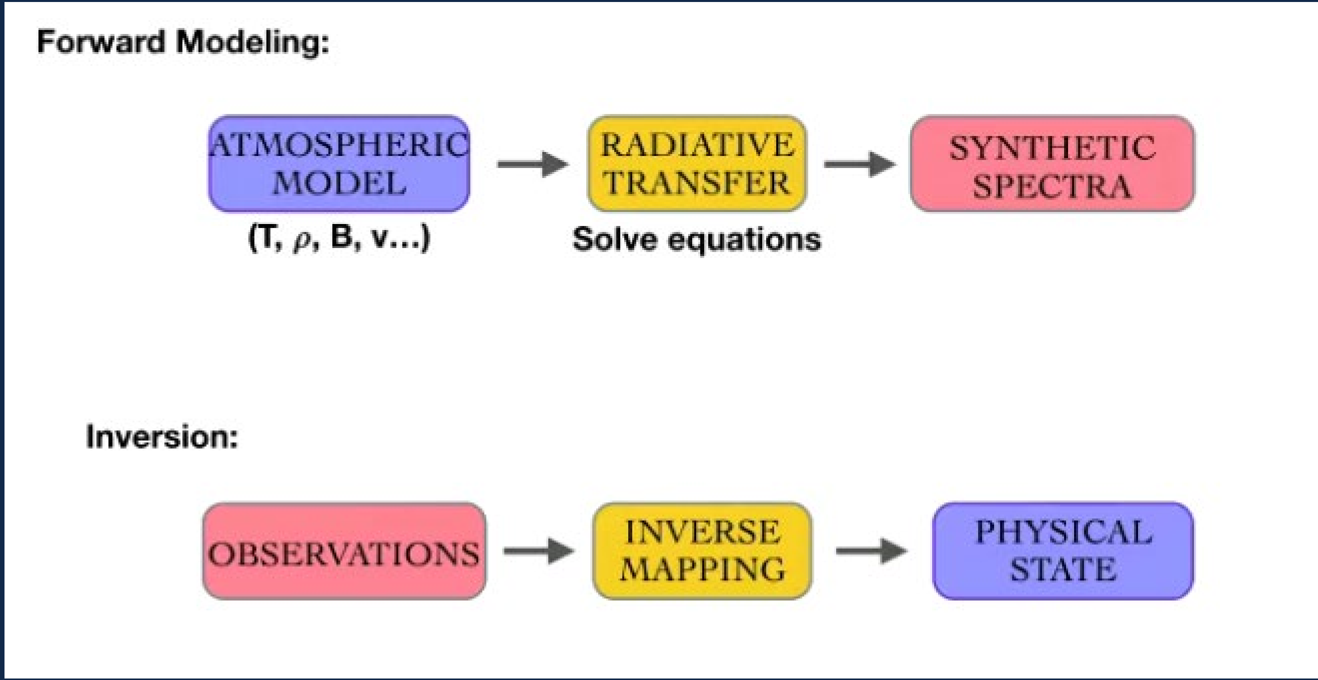
www.universetoday.com

- These phenomena triggered by magnetic field

Measuring magnetic field

- Using Polarizations:
  - Produced by symmetry-breaking processes
- State of Polarization: The Stokes vector$\{I, Q, U, V\}^T$

Forward problem and Inversion problem



Goals

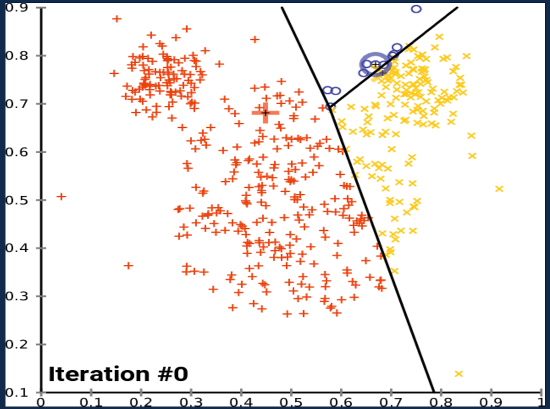
- Validating notable structures in Atmospheric model and polarized spectra
- How much diversity exist in solar atmosphere and how this varies by time
- Detecting any outlier or new structure can be found
- Ultimate goal: Finding a mapping

Methods: Clustering

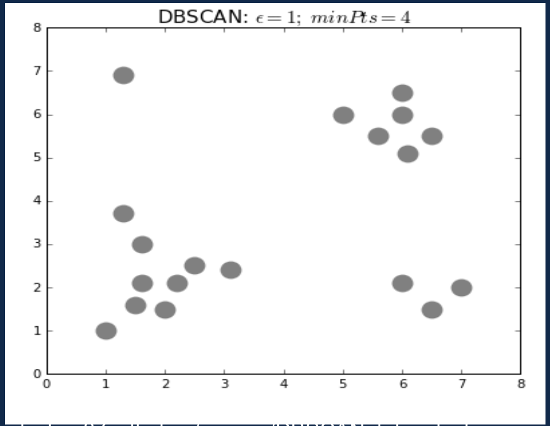
Clustering:

- K-means:
  - Number of clusters
  - Tolerance

- DBSCAN:
  - Epsilon
  - Minimum points



wikimedia.org/wiki/File:K-means_convergence.gif



dashee87.github.io/images/DBSCAN_tutorial.gif

Methods: Evaluation

- Using domain experts knowledge

  - finding familiar structures like magnetic regions and granulations

- Inertia

- Silhouette

$$s(i) = \frac{b(i) - a(i)}{Max\,(b(i), a(i))}$$

Hinode data-set

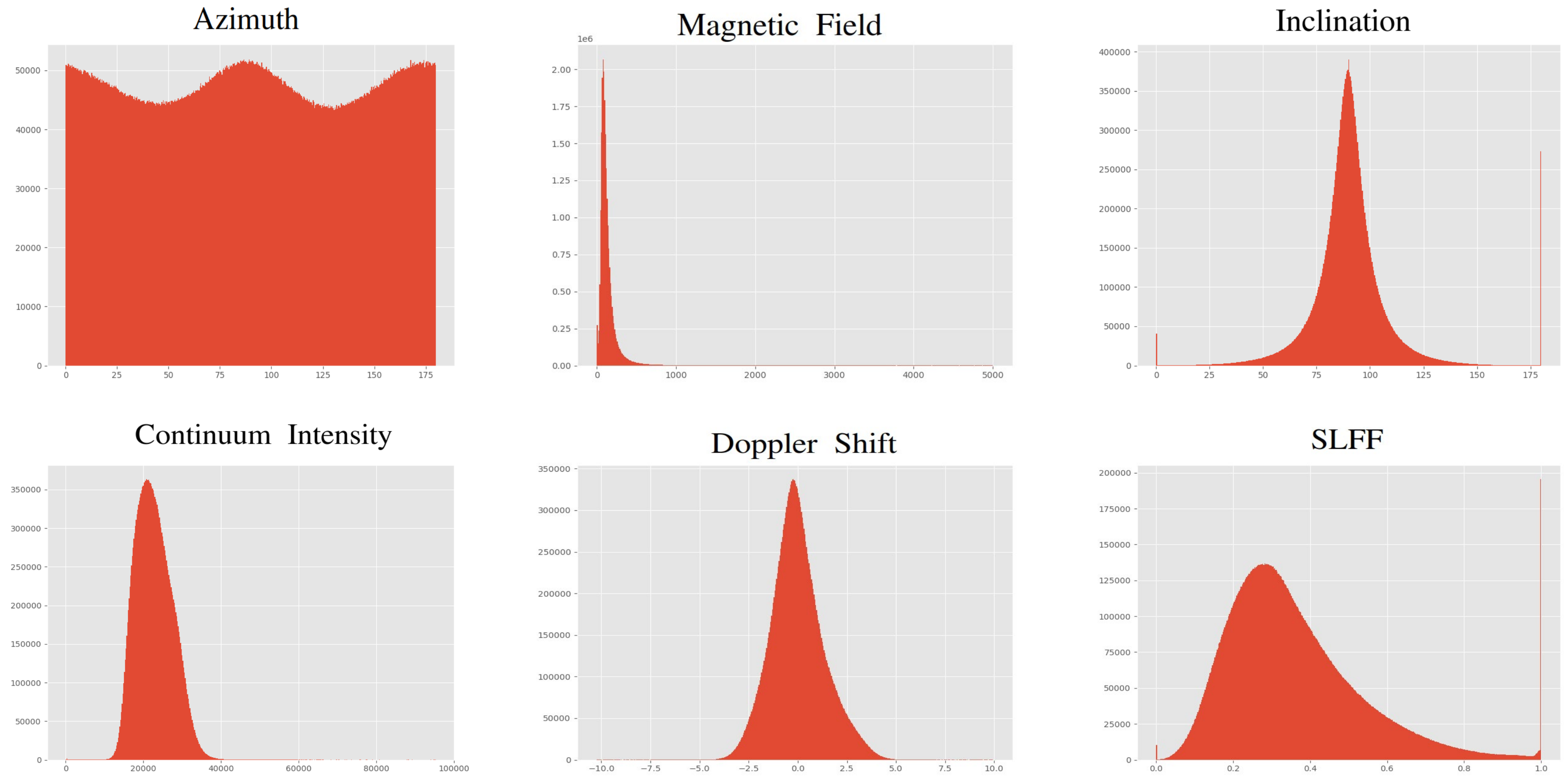
- HOP 336 dataset

- 3 sets of samples: North(+37º), South(-37º), and Equator

- Features: Magnetic field (strength, inclination, azimuth, fill factor), Doppler shift, Continuum intensity

- 300 maps in total, 153 maps in North, and almost 132 million samples

Hinode data-set: Distributions



Preprocessing: First Setting

- Obtaining Cos of Azimuth
- Using B*cos(inc) instead of B and inclination separately.
- Subtracting a regression line from each row for Doppler shift (for each map individually).
- Tried 3, 4, 5, and 10 clusters using k-means.
- Result: Centers were really close to each other and no structure found! Maybe due to local minima problem in k-means.

First Setting: Results

For K = 10 clusters

Preprocessing: Second Setting

- Using B and cosine of inclination Separately

- Removing Azimuth due to noisy characteristic.

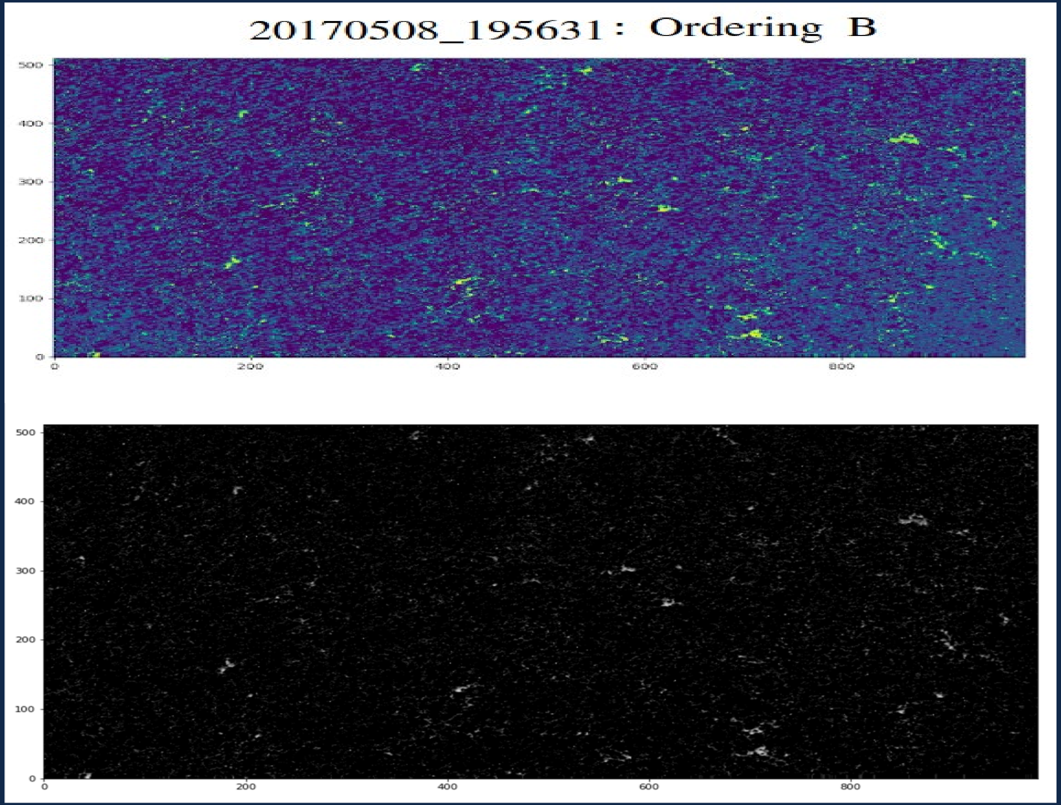
- Putting constraint over magnetic field. | B | < 200 G should be treated as noise:
  SLFF = 0, B = 0, and Inclination = 90.

- Try higher number of clusters: 25, 50, 100, 150, 200

- For better visualization, we ordered the same clustering results in 3 different order: based on B and CI centers

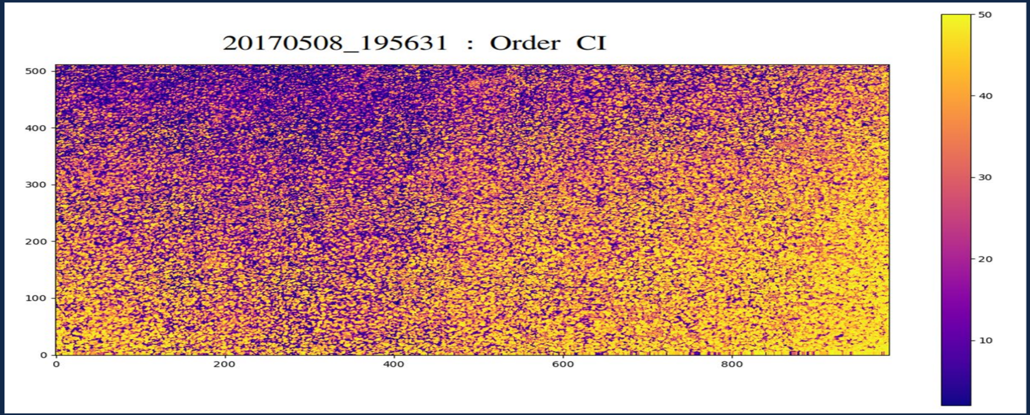
Second Setting: Results

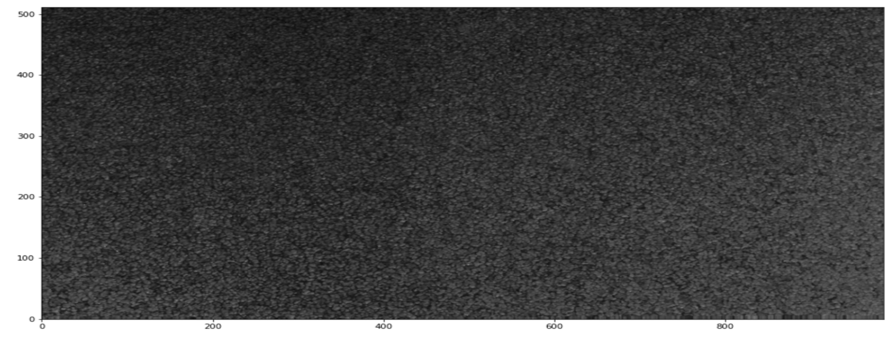
For K = 50 clusters



Second Setting: Results

For K = 50 clusters





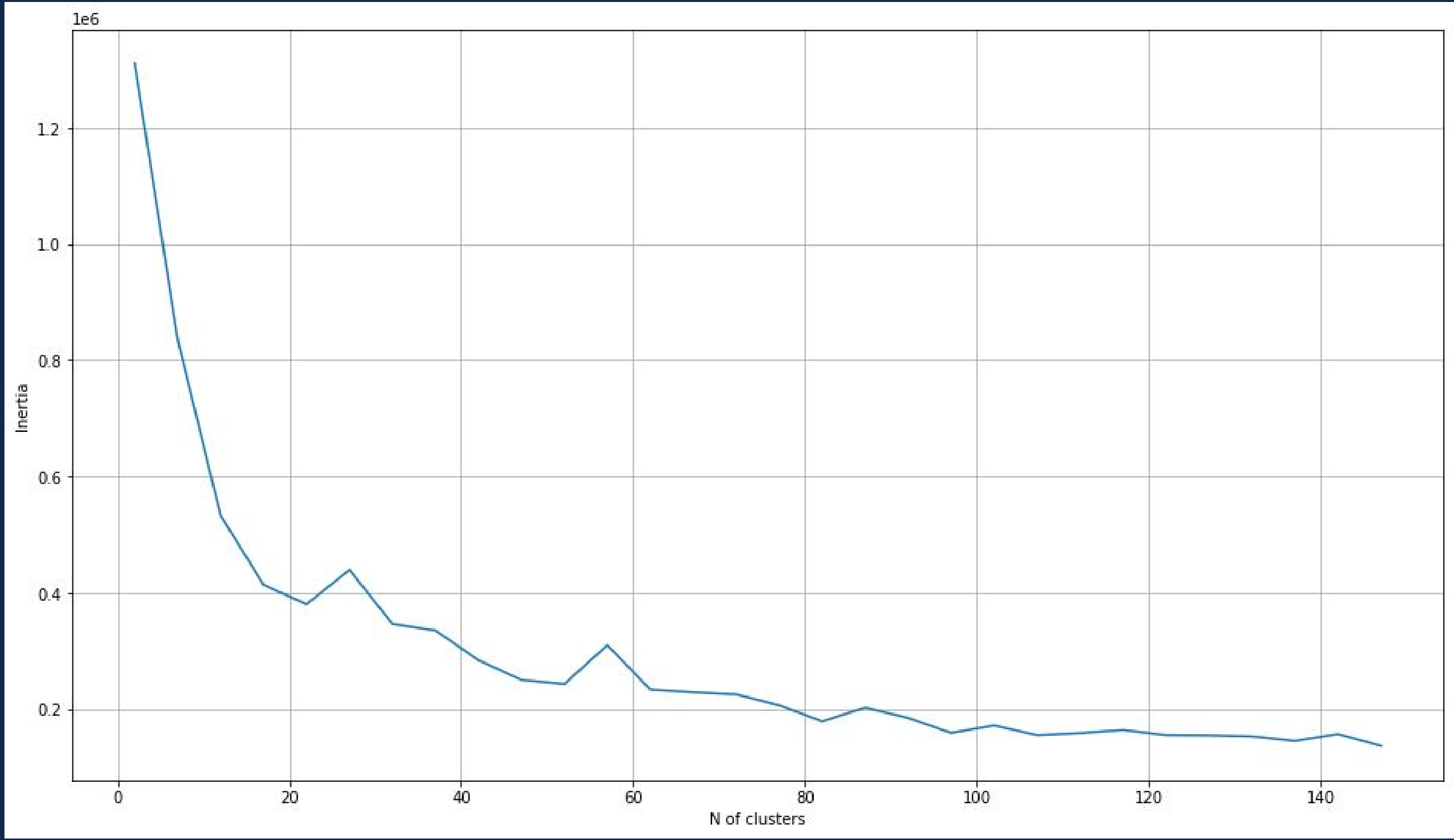
Second Setting: Results

- A strange artifact found when k = 50 clusters with a very low continuum intensity(450 compared to average of 21000)

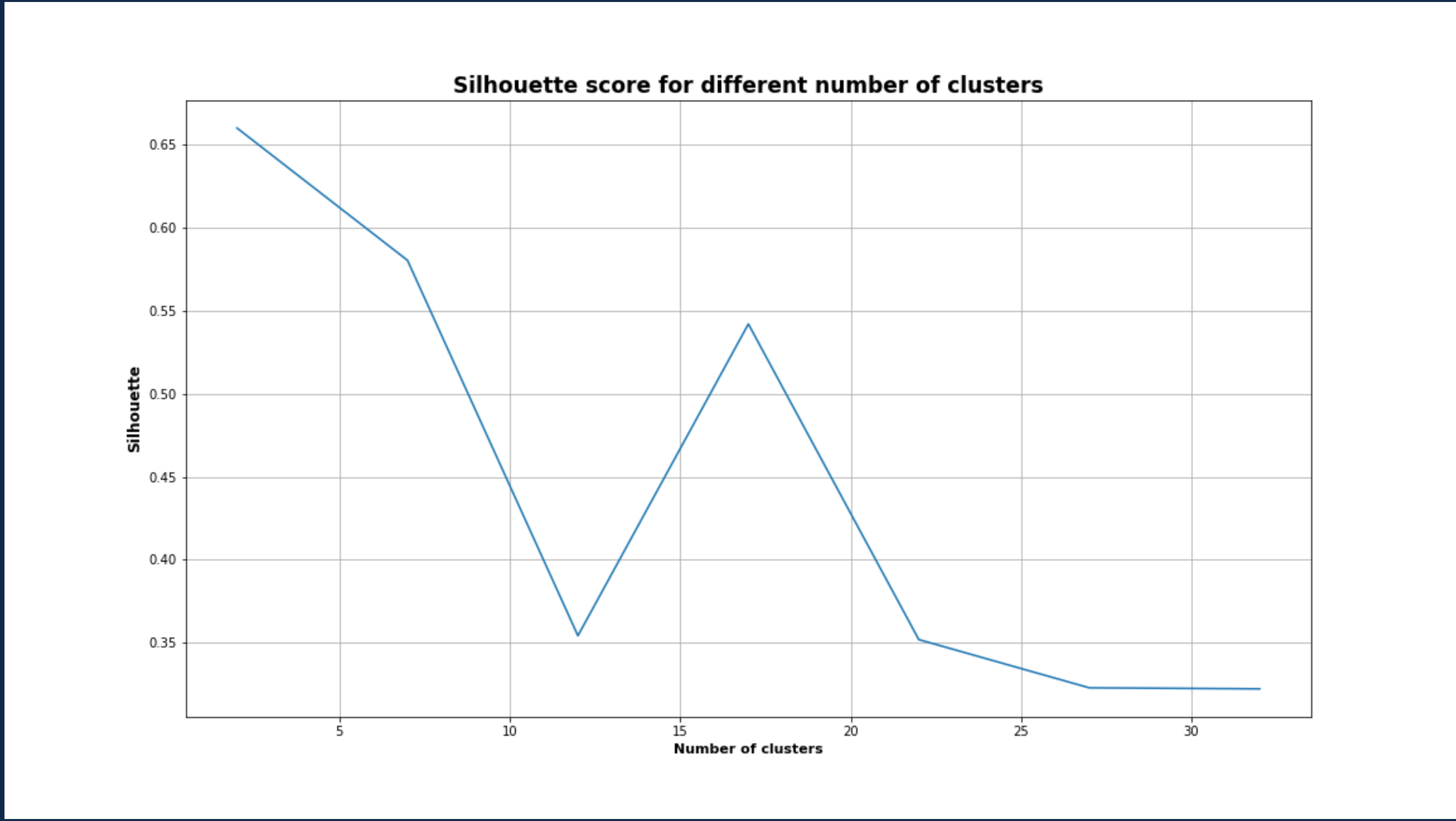
- We used Inertia as a metric.

- This setting, improved inertia by 50x, compared to first setting:
  It was expected since we removed Azimuth and zeroing features related to B < 200 G.

Optimal number of clusters: Inertia



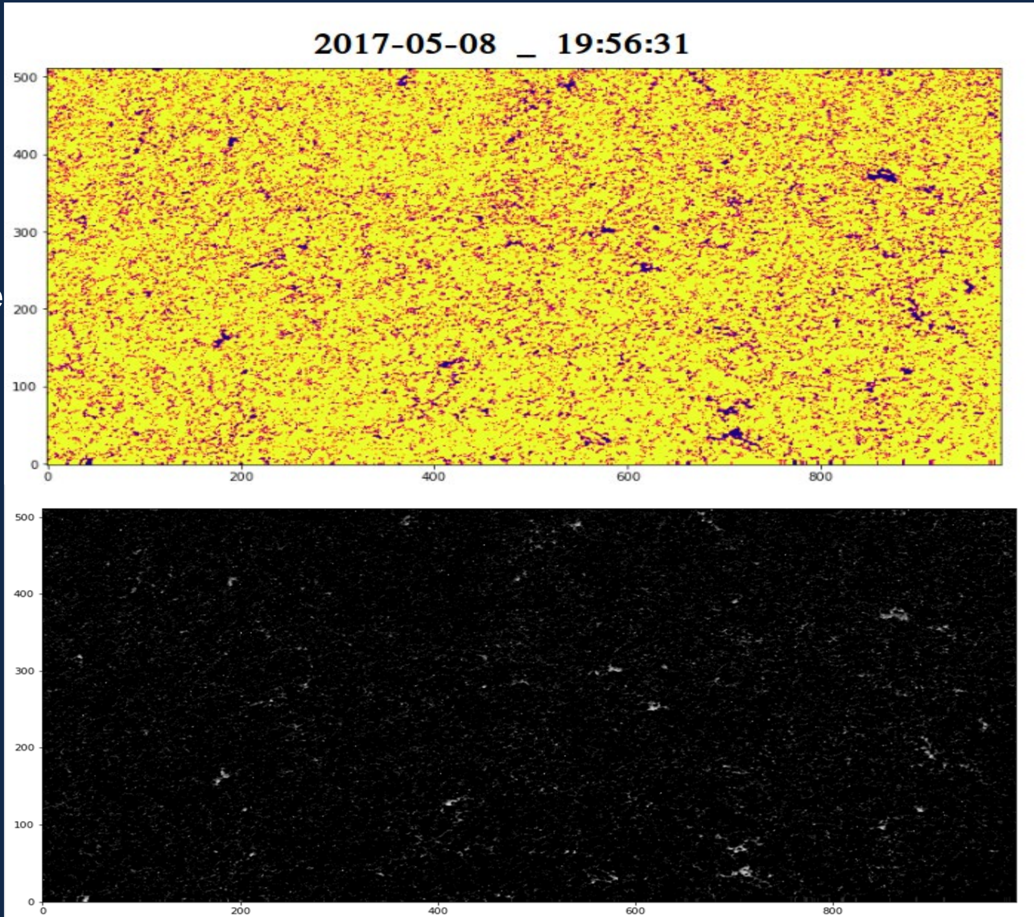
Optimal number of clusters: Silhouette



DBSCAN result

- Reaching to 2 clusters!

- 10% improvement in silhouette Compared to K-means



Future directions

- Using a grid search to find optimal parameters for DBSCAN

- Categorize our pixels/samples to two different data-set and apply clustering to them separately

- Using other clustering methods like Hierarchical methods

Thanks for your time

Mehdi Nourelahi
mnourelahi@protonmail.com