Leibniz-Institut für
Astrophysik Potsdam

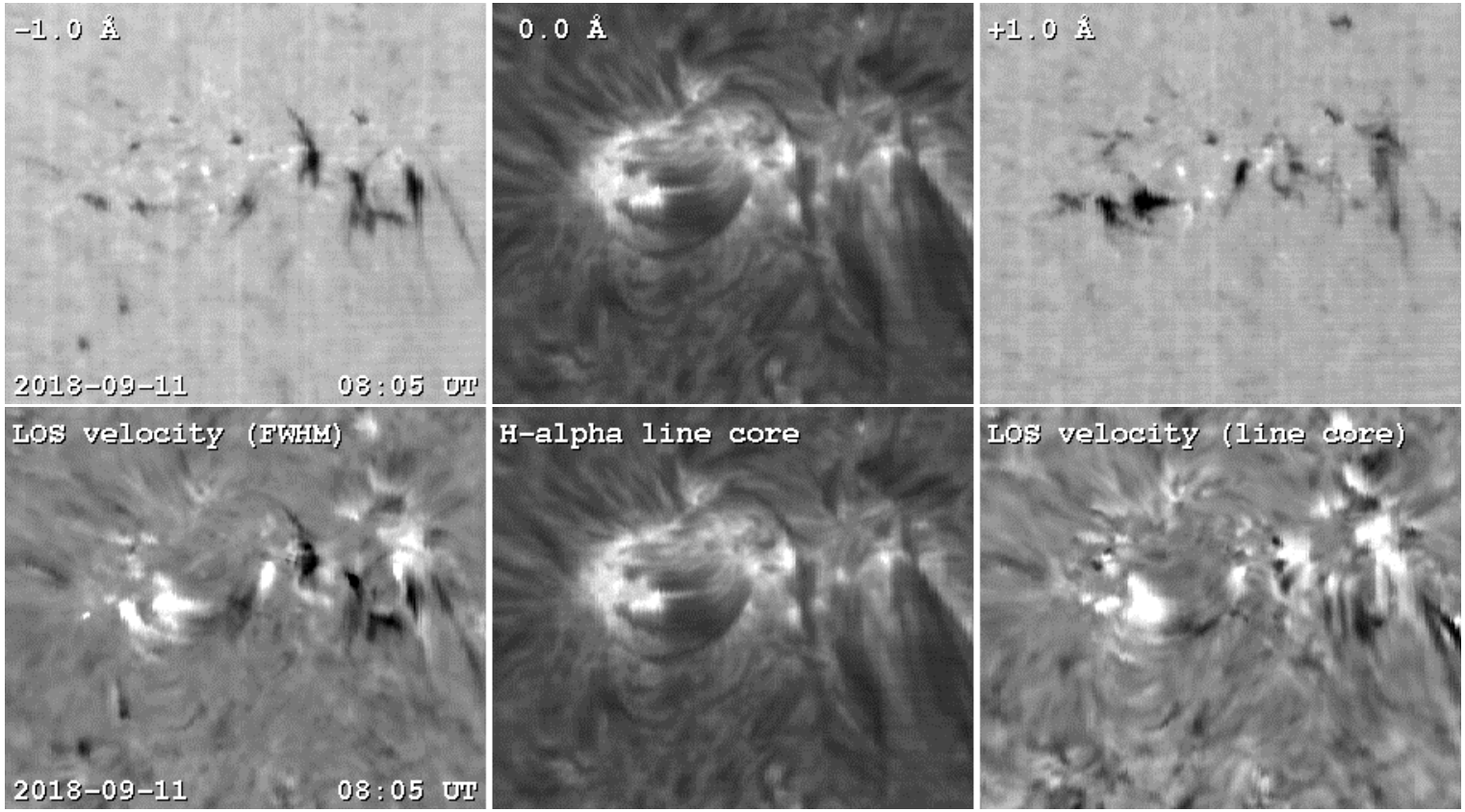
Classification of High-resolution Solar Hα Spectra using t-SNE

Meetu Verma

Gal Matijevič, Carsten Denker, Andrea Diercke, Christoph Kuckein, Ekaterina Dineva, Horst Balthasar, Ioannis Kontogiannis, and Partha S. Pal

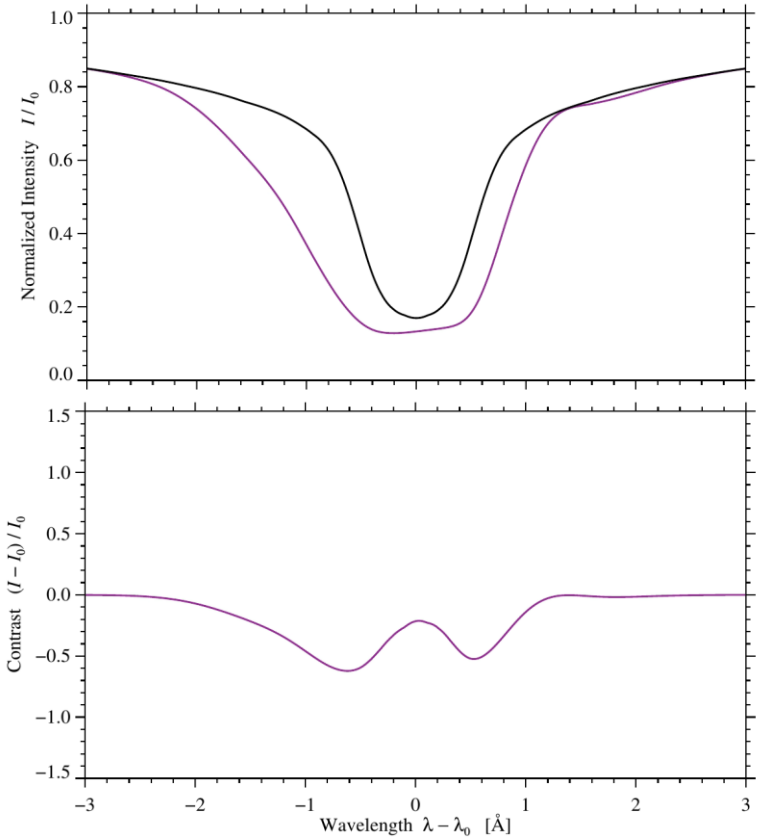
Why do we need classification?

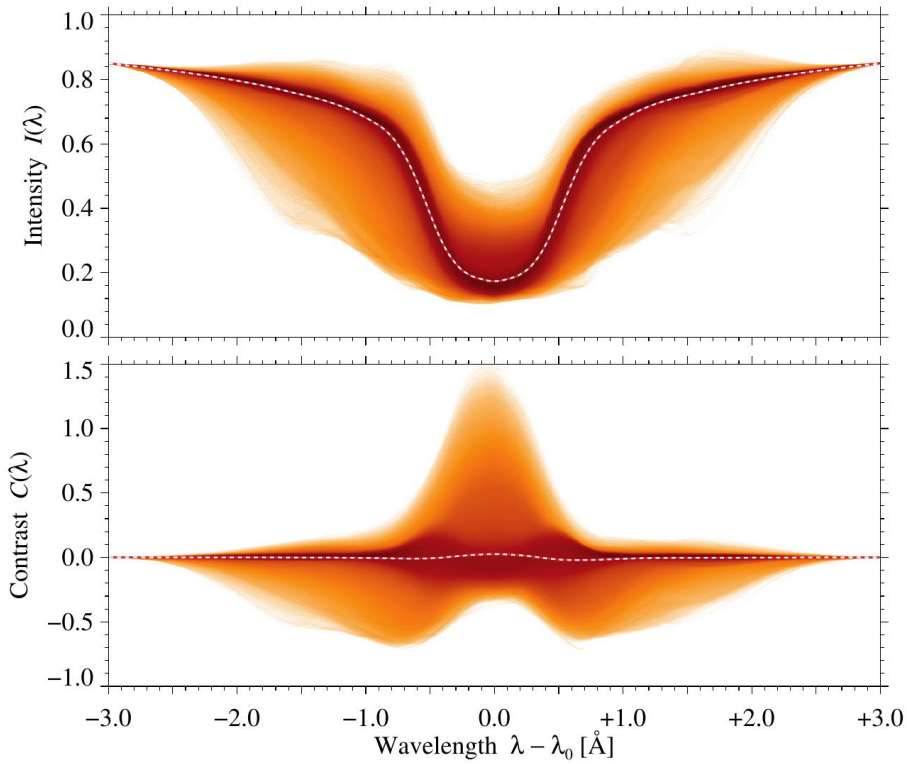
- On one observing day → time-series of 21 Hα spatio-spectral data cubes.



3D data → Science in every pixel

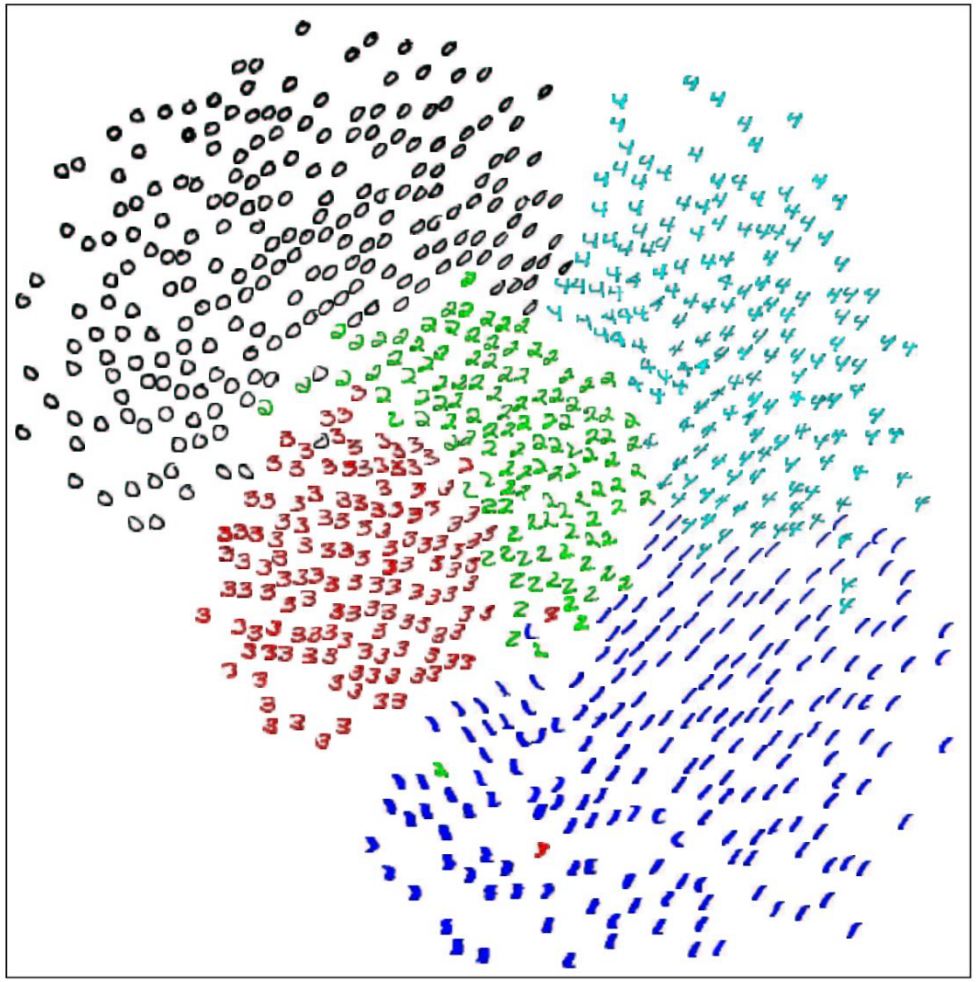
- Contains about 8.7 million intensity and contrast profiles.





t-SNE → Appropriate tool to classify spectra

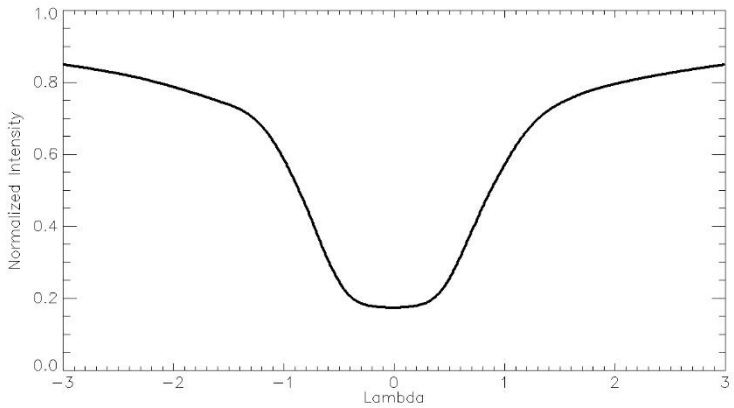
- Probabilistic approach
- Dimensionality reduction
- t-SNE result of classifying on 3000 256-dimensional grayscale images of handwritten digits.
- Classes are quite well separated even though t-SNE had no information about class labels.
- Within each class, properties like orientation, skew and stroke thickness tend to vary smoothly across the space.



L.J.P. van der Maaten and G.E. Hinton. **Visualizing High-Dimensional Data Using t-SNE**. *Journal of Machine Learning Research* 9(Nov):2579-2605, 2008

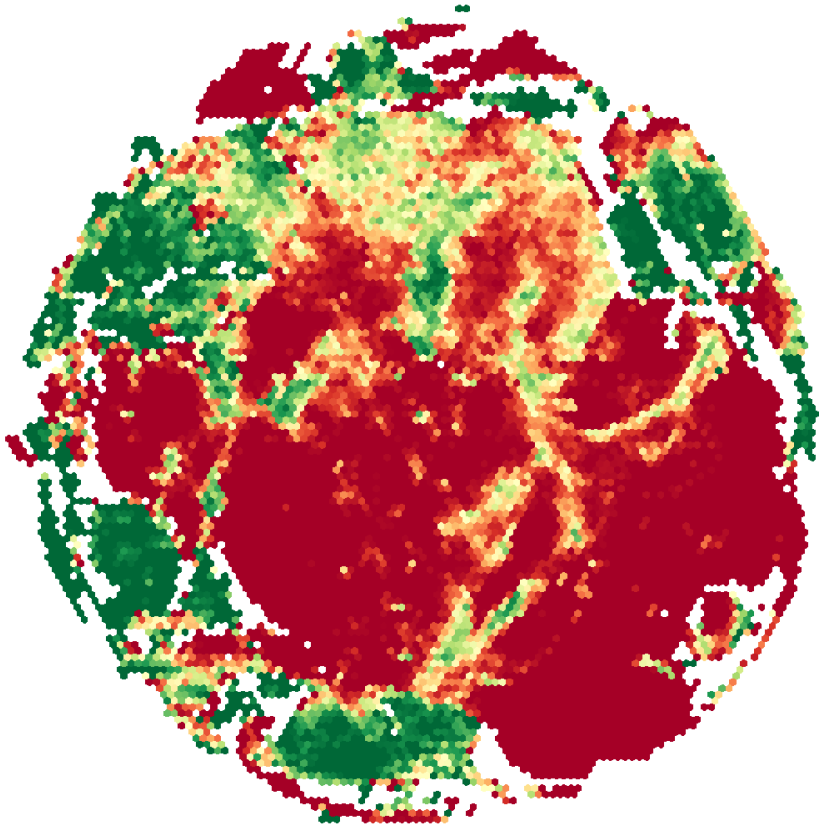
L.J.P. van der Maaten. **Accelerating t-SNE using Tree-Based Algorithms**. *Journal of Machine Learning Research* 15(Oct):3221-3245, 2014.

t-SNE → From profiles to classification



t-SNE projection of 630 x 660 spectral profiles with 601 wavelength points.

The choice of parameters perplexity = 50, theta = 0.5, number of iterations = 1000



Q1 Is the default choice ok?

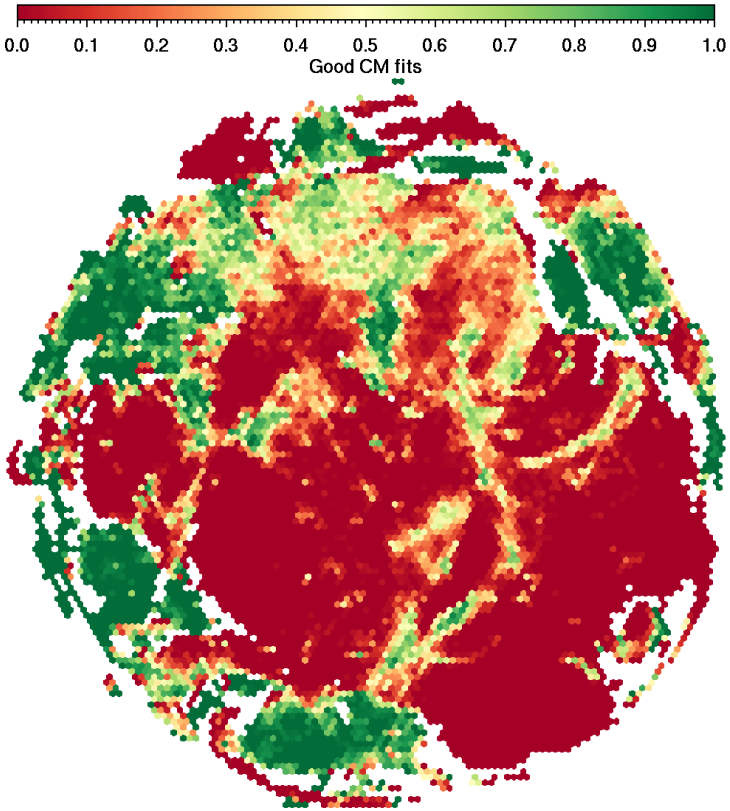
Q2 Is the projection different for profiles and PCA coefficients?

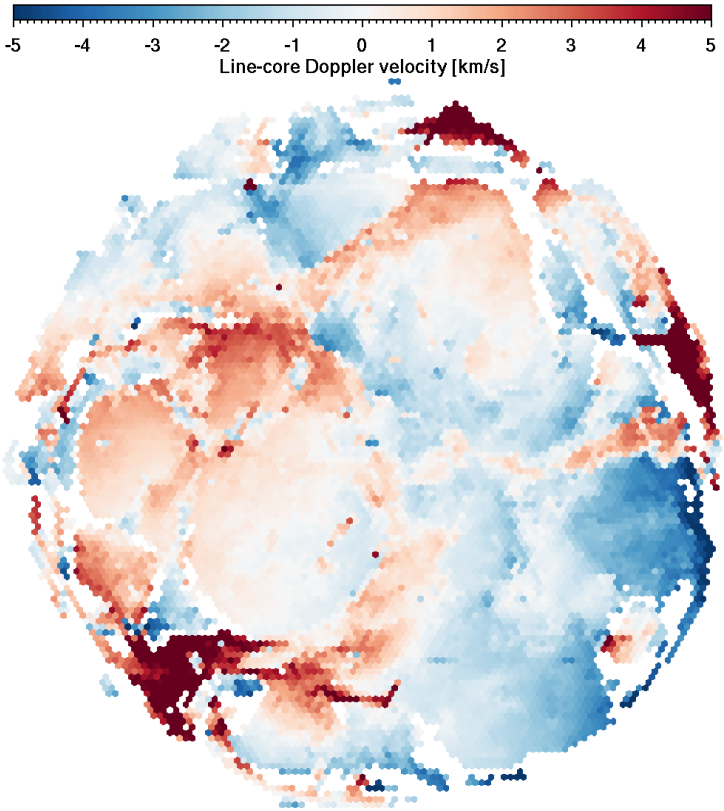
Q3 Is the projection affected by seeing conditions?

https://distill.pub/2016/misread-tsne/

Parameter study – Careful selection

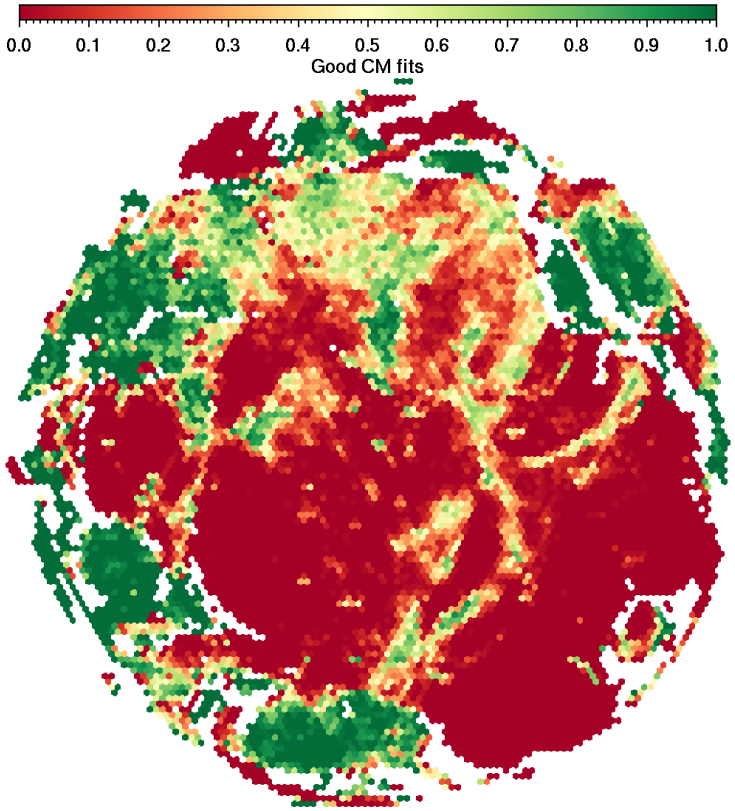
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.3, P = 50, n = 1000

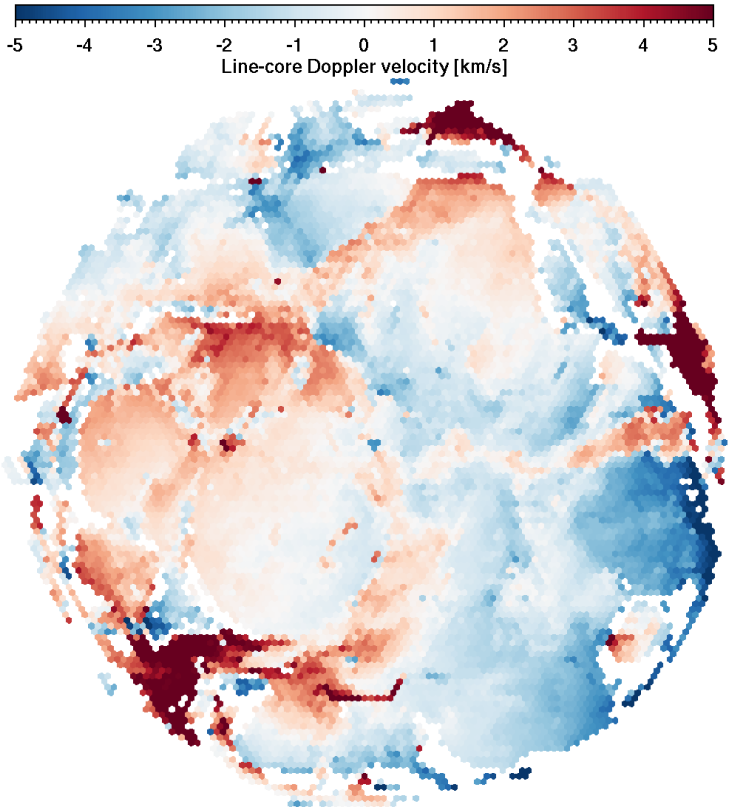




Parameter study – Careful selection

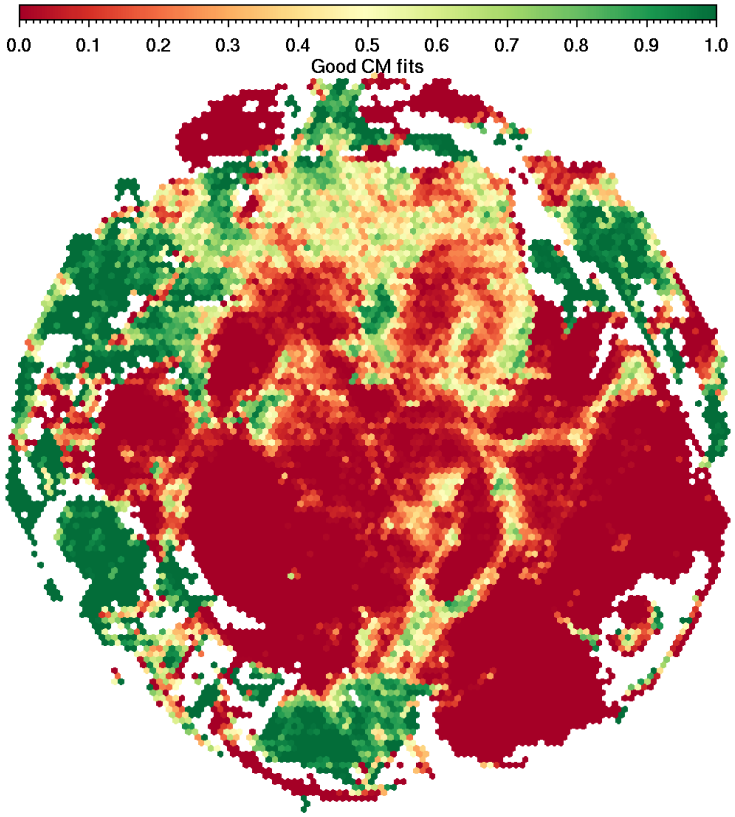
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.4, P = 50, n = 1000

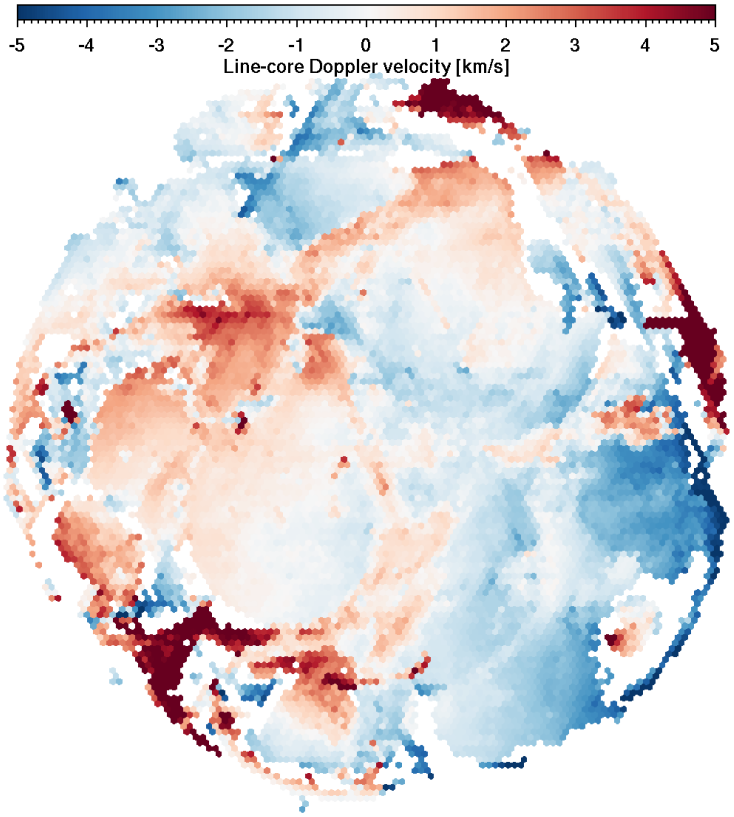




Parameter study – Careful selection

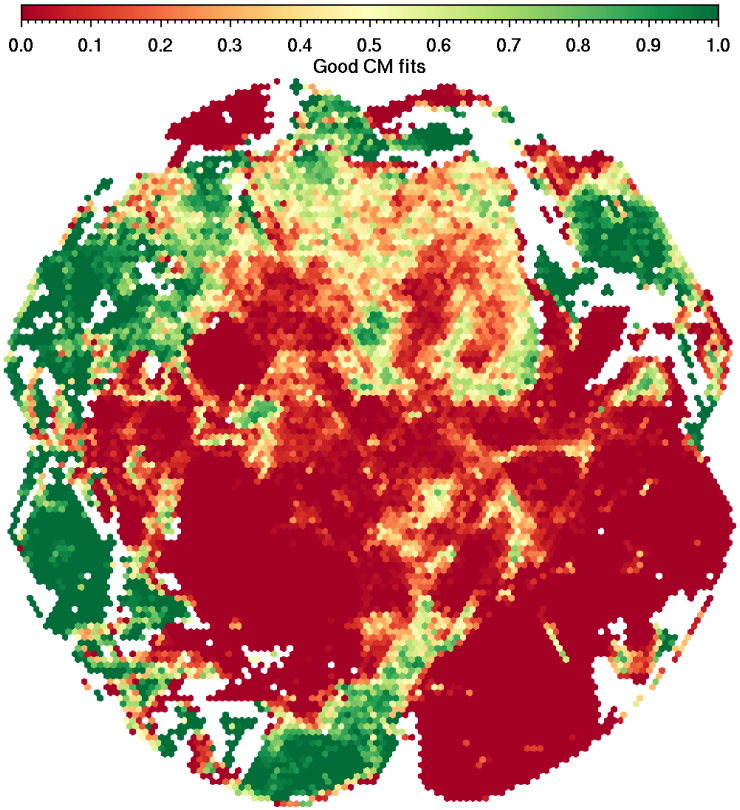
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.7, P = 50, n = 1000

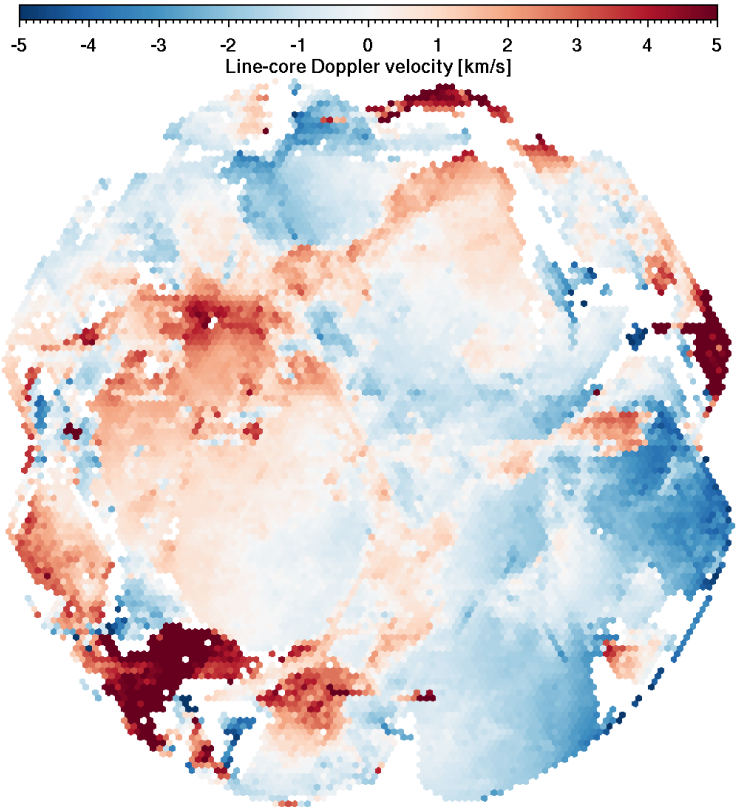




Parameter study – Careful selection

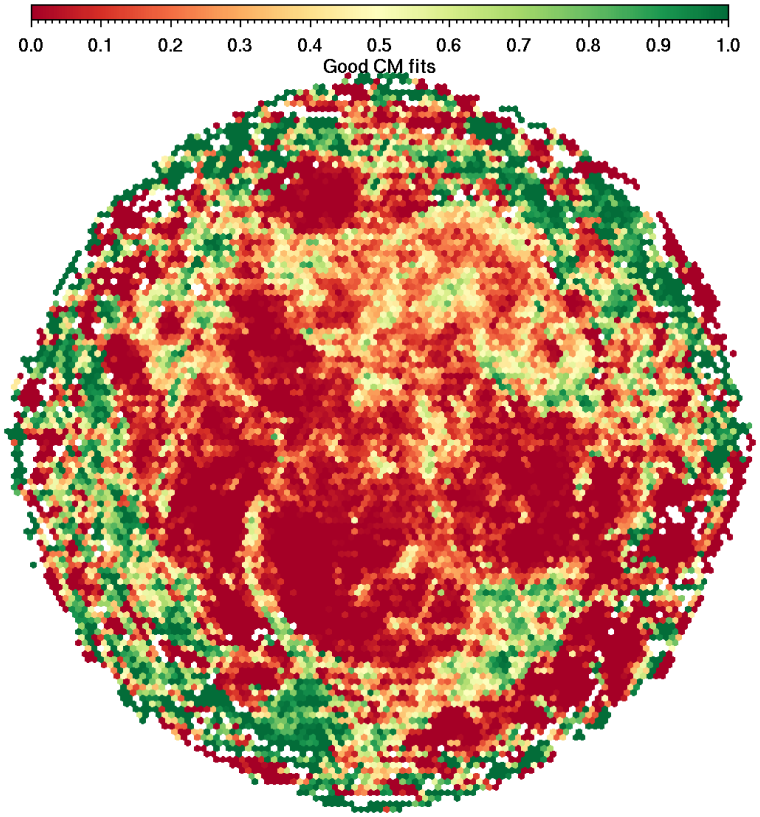
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.8, P = 50, n = 1000

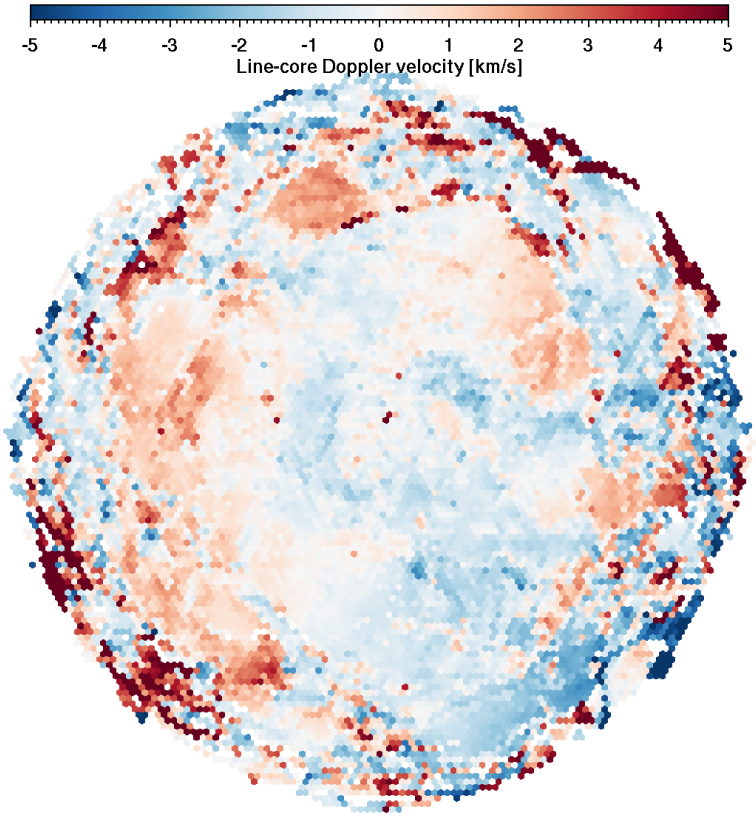




Parameter study – Careful selection

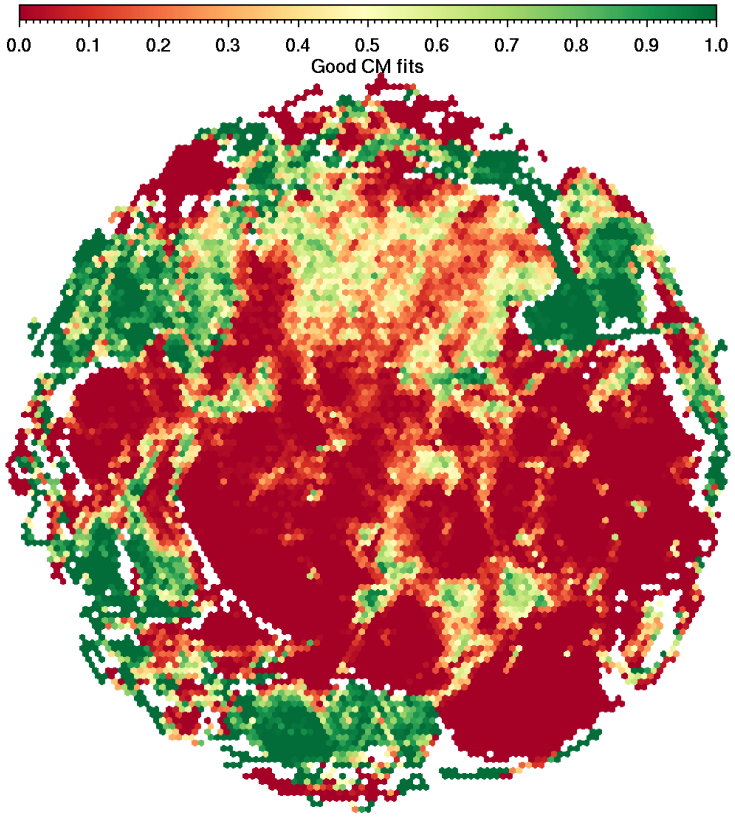
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.5, P = 10, n = 1000

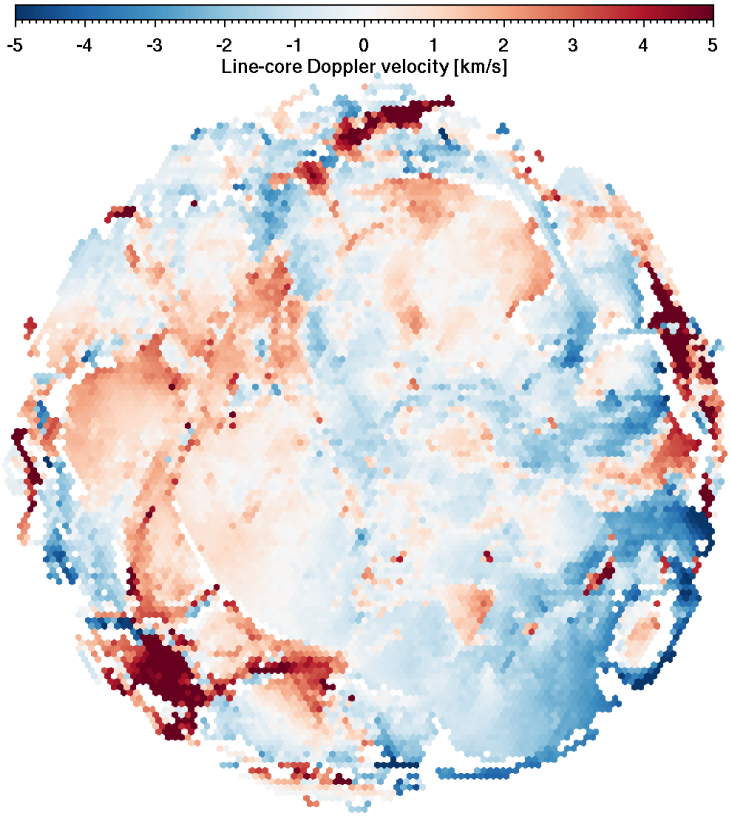




Parameter study – Careful selection

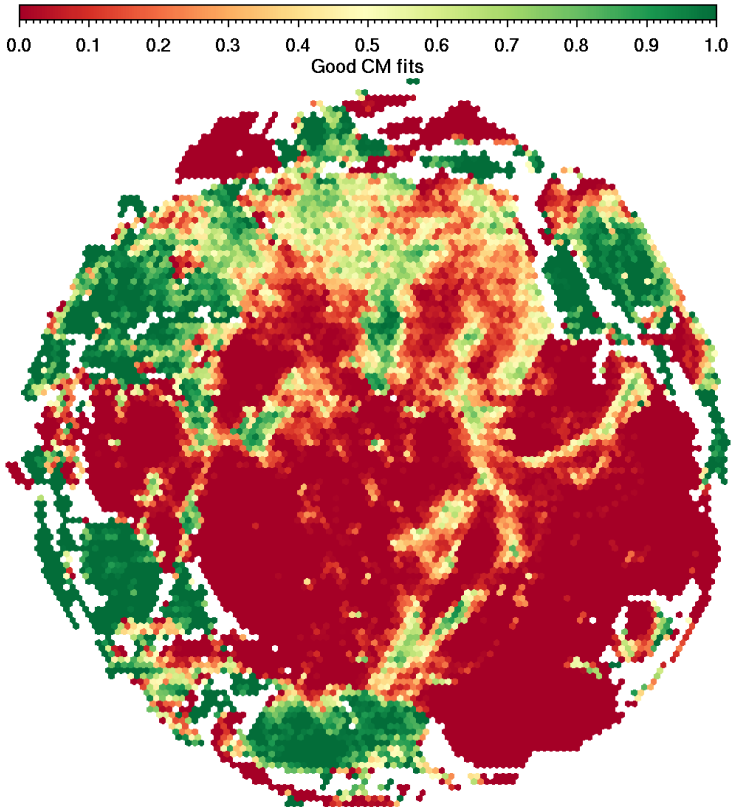
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.5, P = 30, n = 1000

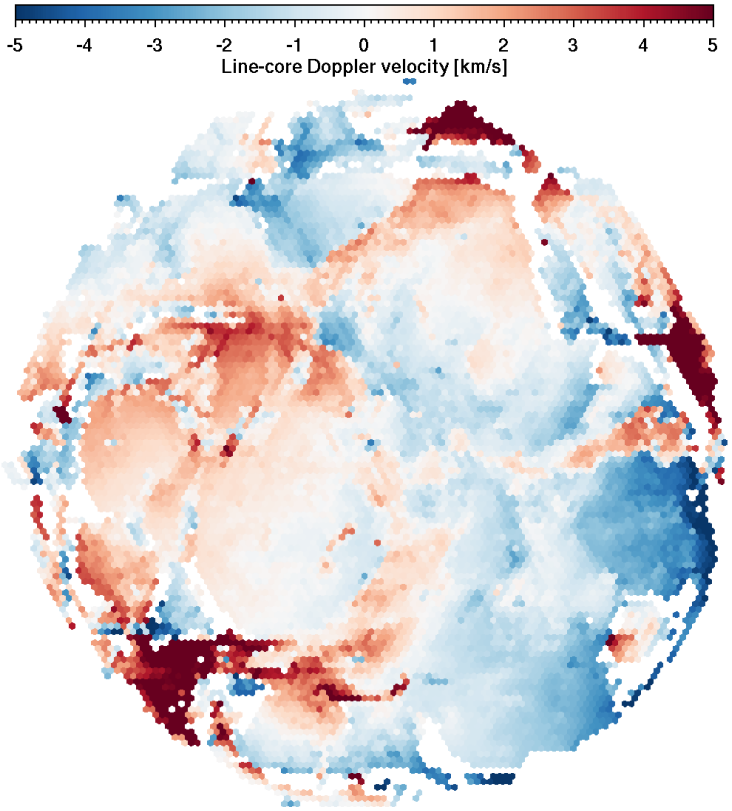




Parameter study – Careful selection

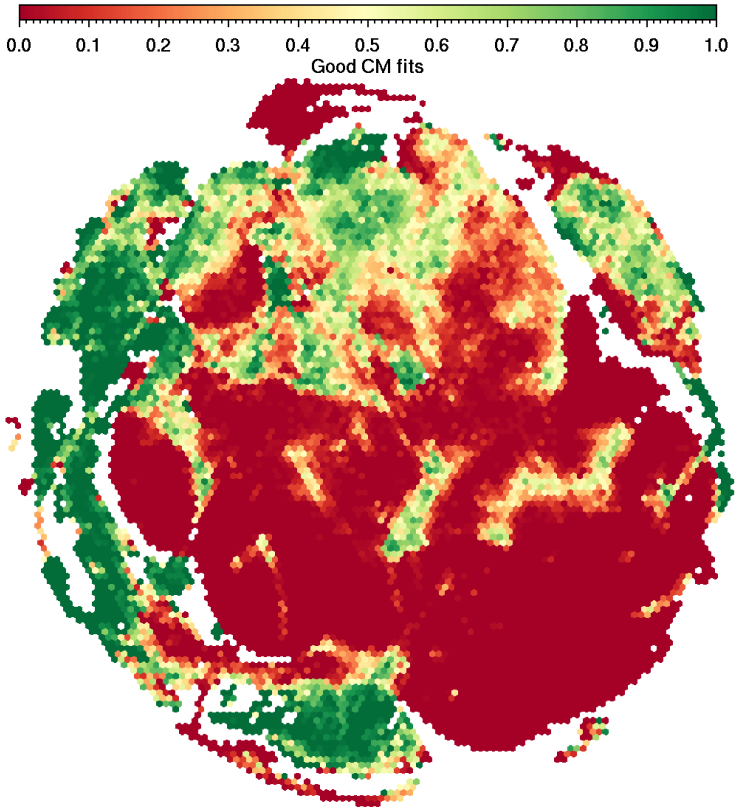
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.5, P = 50, n = 1000

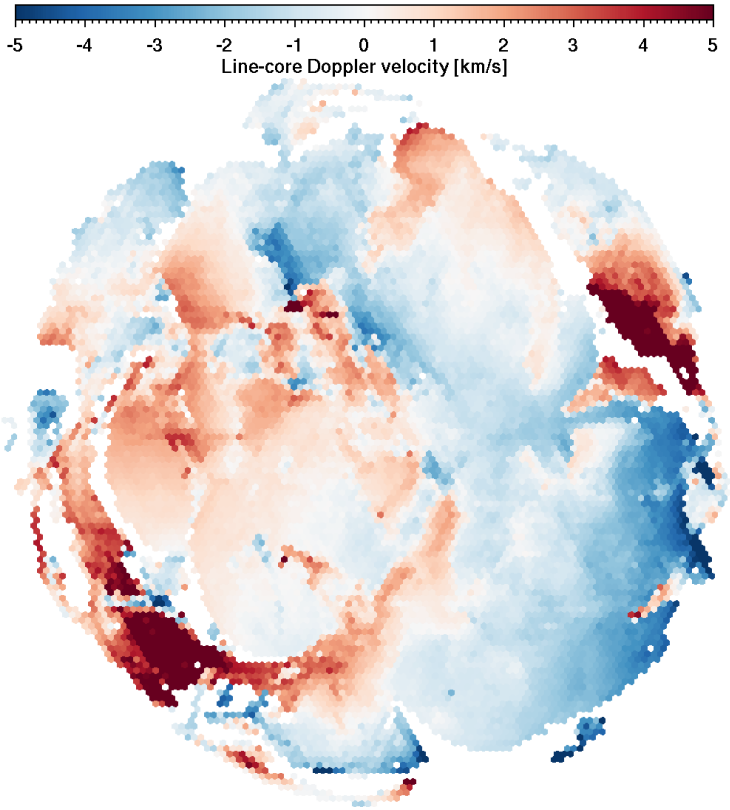




Parameter study – Careful selection

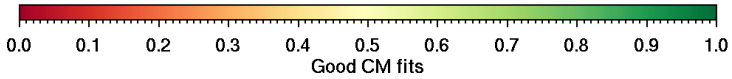
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.5, P = 80, n = 1000

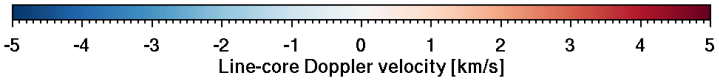


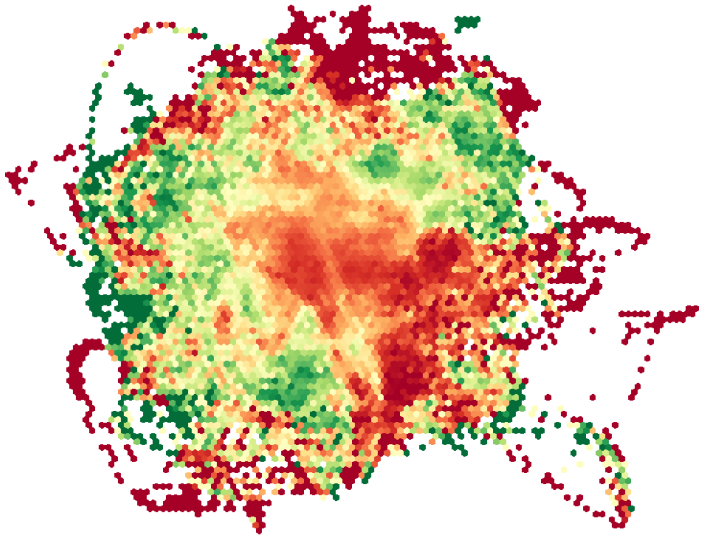


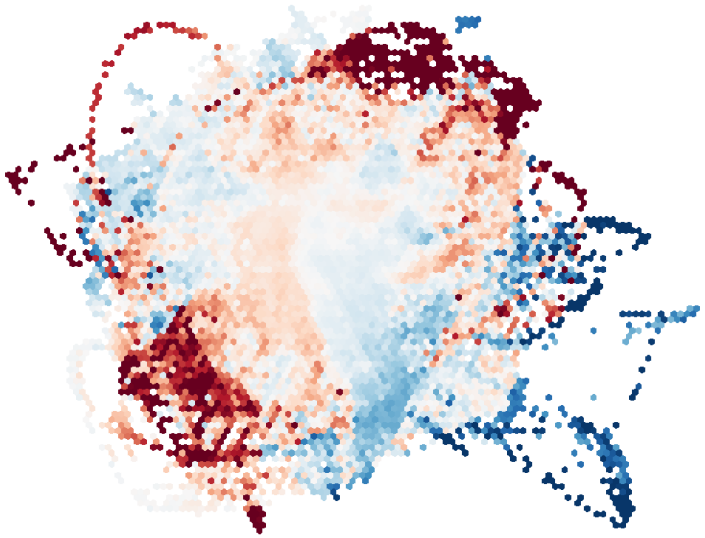
Parameter study – Careful selection

- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.5, P = 50, n = 200



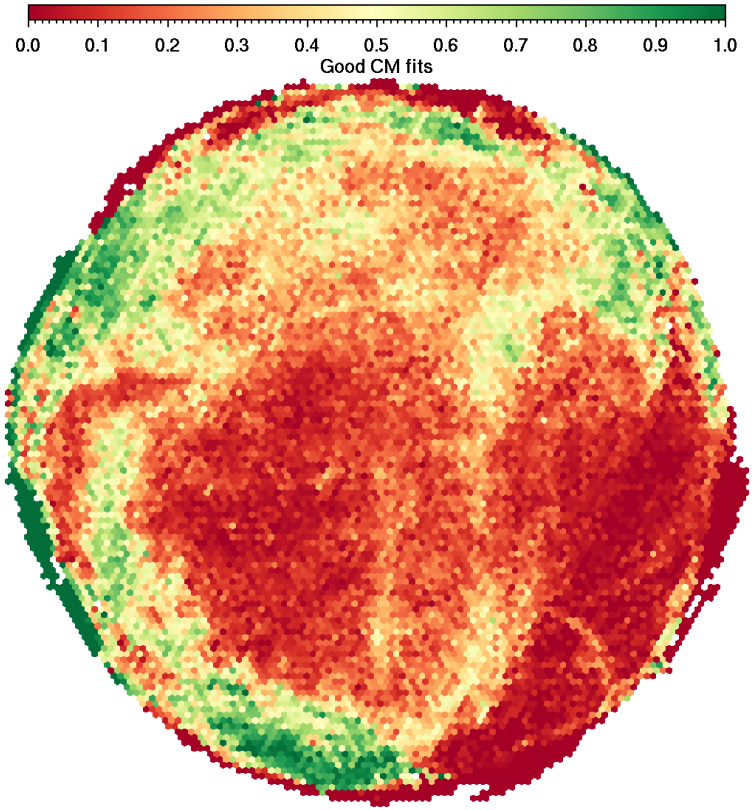


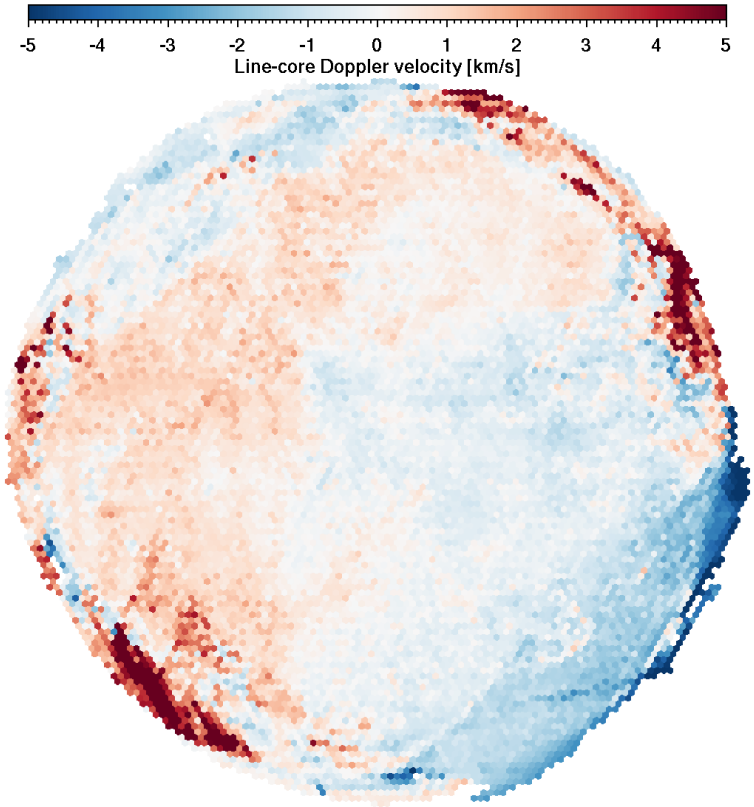




Parameter study – Careful selection

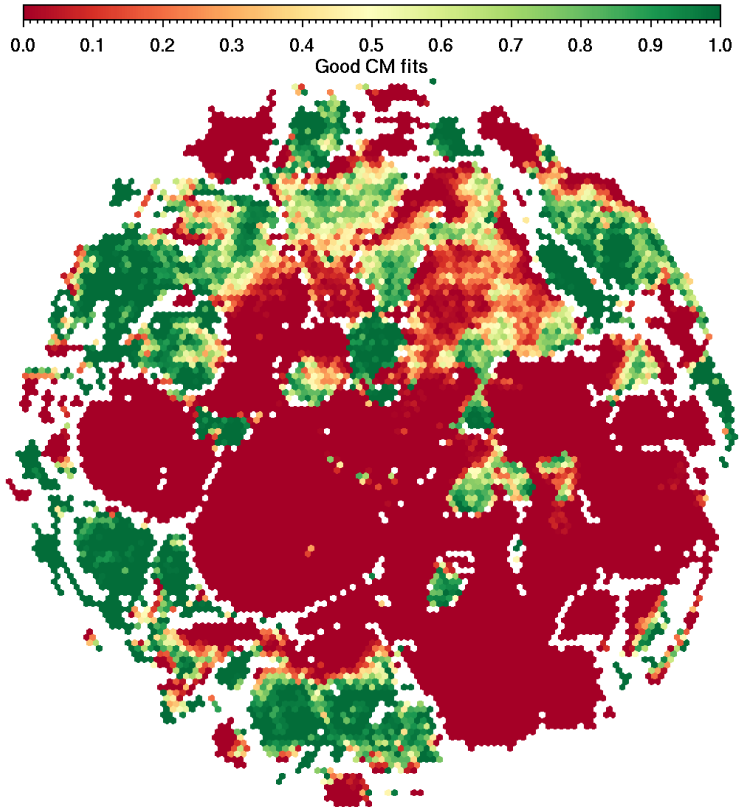
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.5, P = 50, n = 400

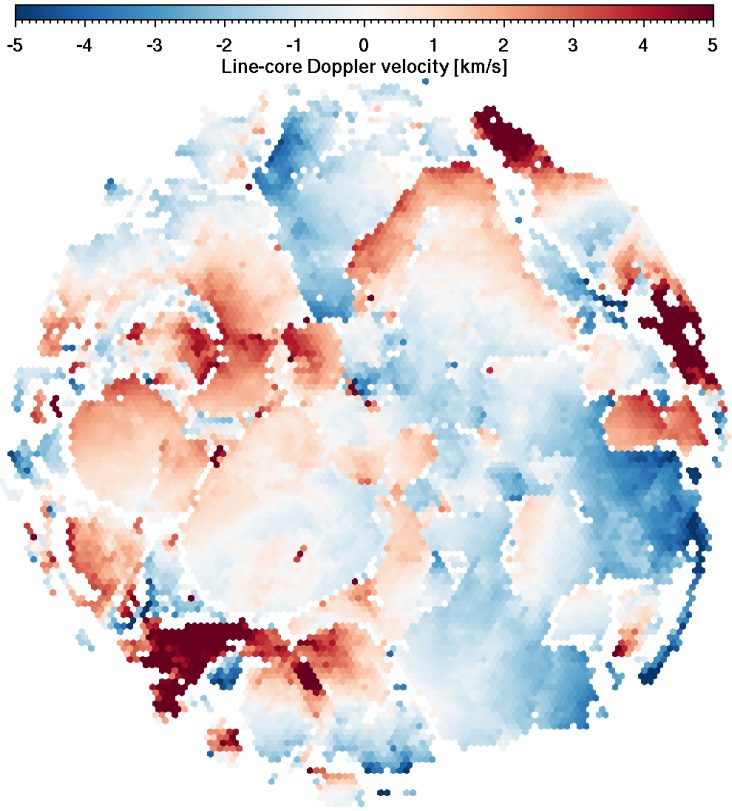




Parameter study – Careful selection

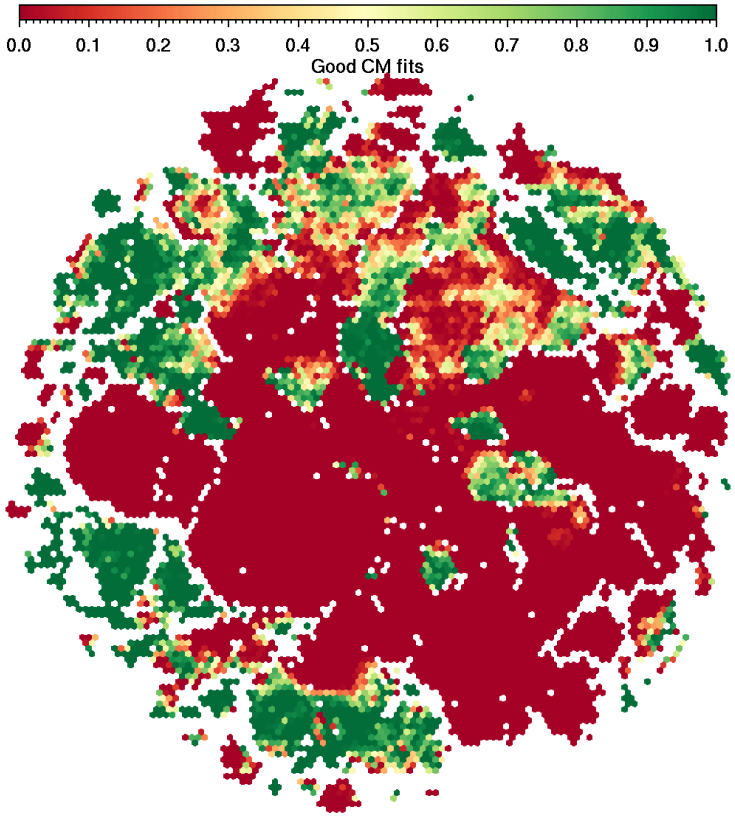
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.5, P = 50, n = 2000

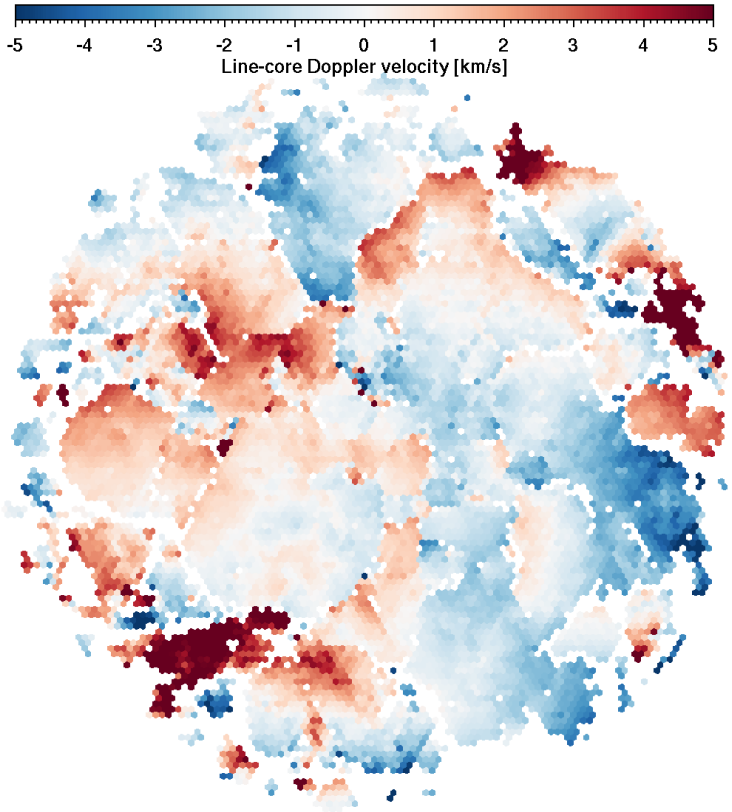




Parameter study – Careful selection

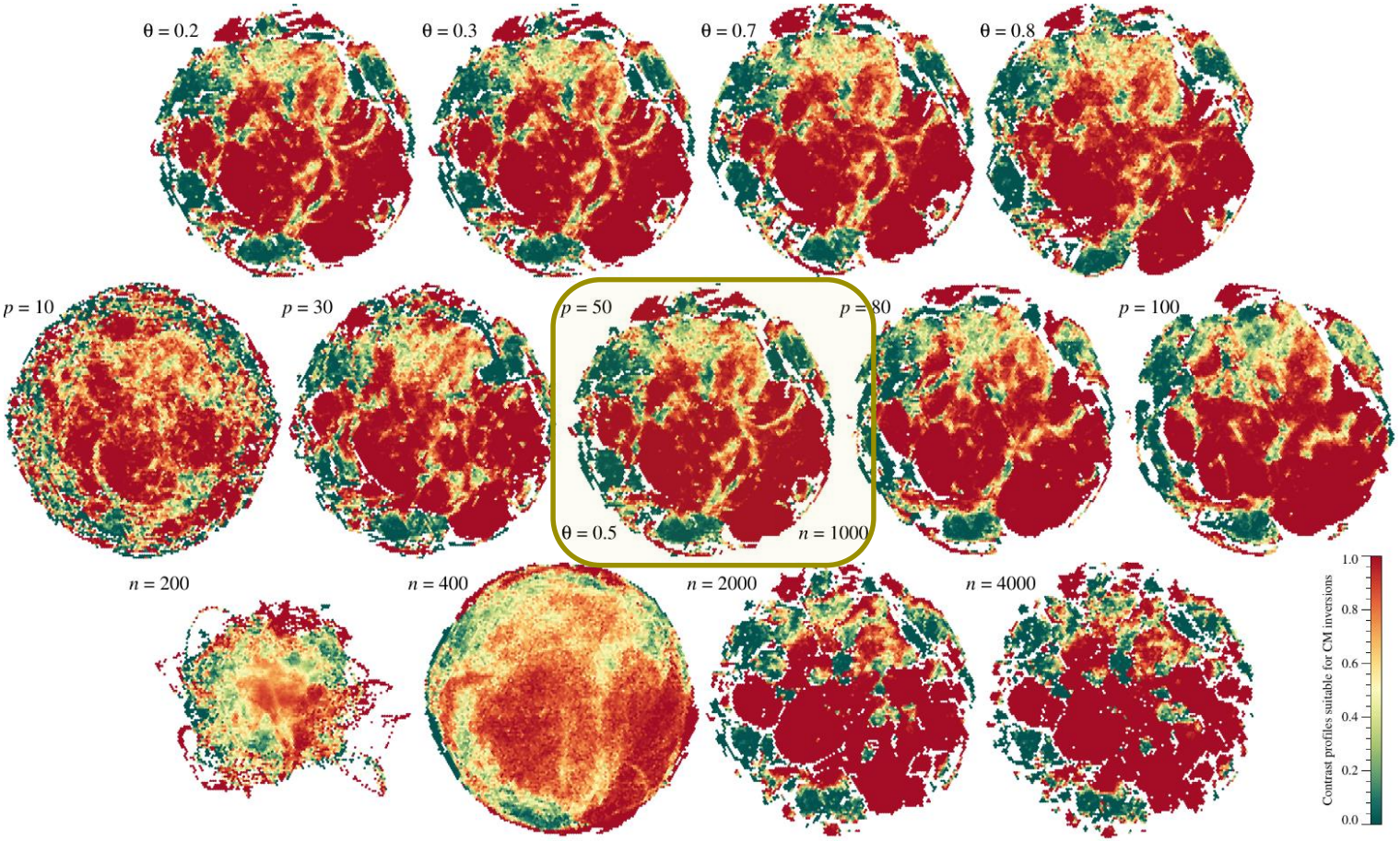
- Three parameters we can change.
  - Theta, Perplexity, Number of Iterations → θ = 0.5, P = 50, n = 4000





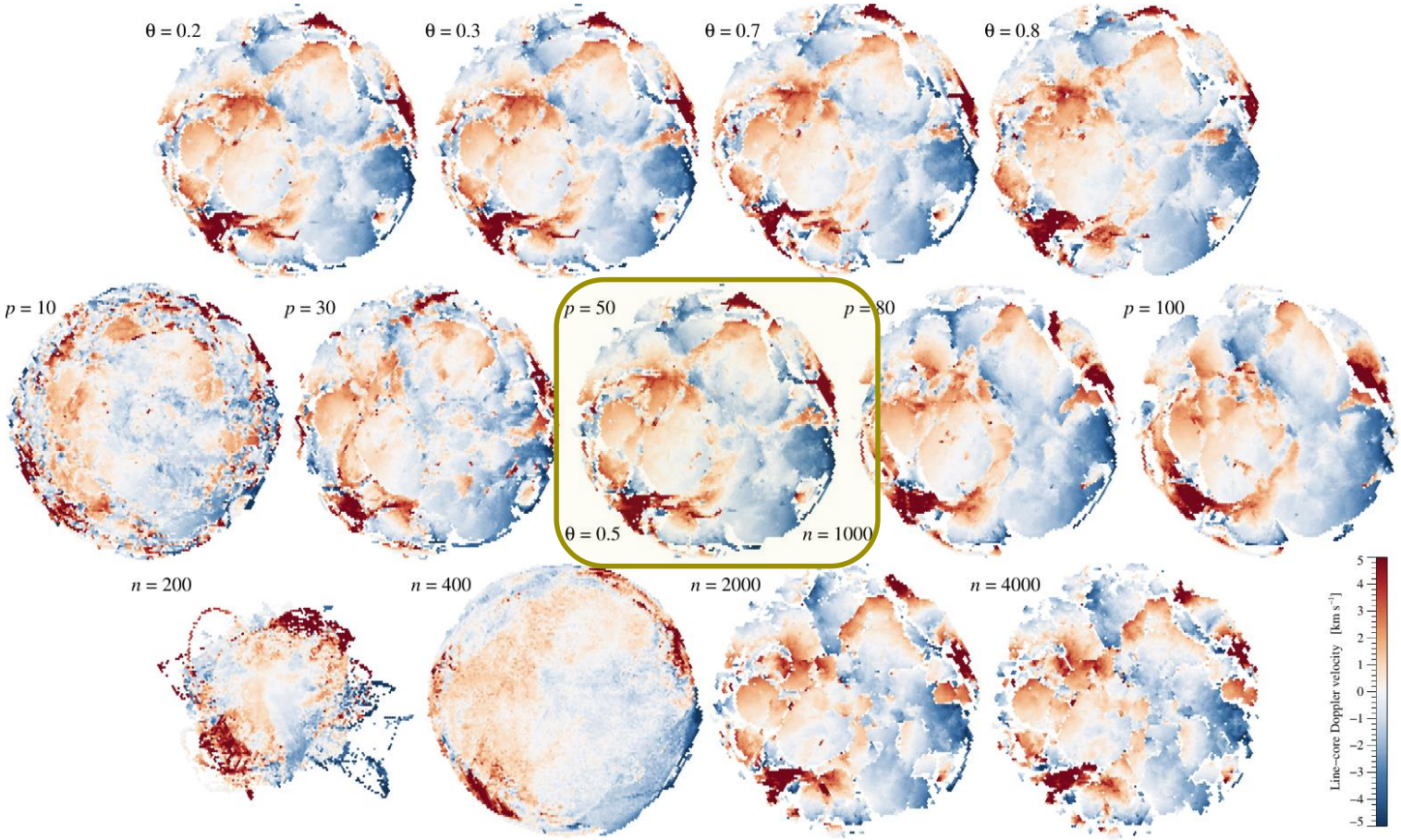
Parameter study – Careful selection

A1 The default parameters are fine, maybe the number of iteration has to be improved for large datasets

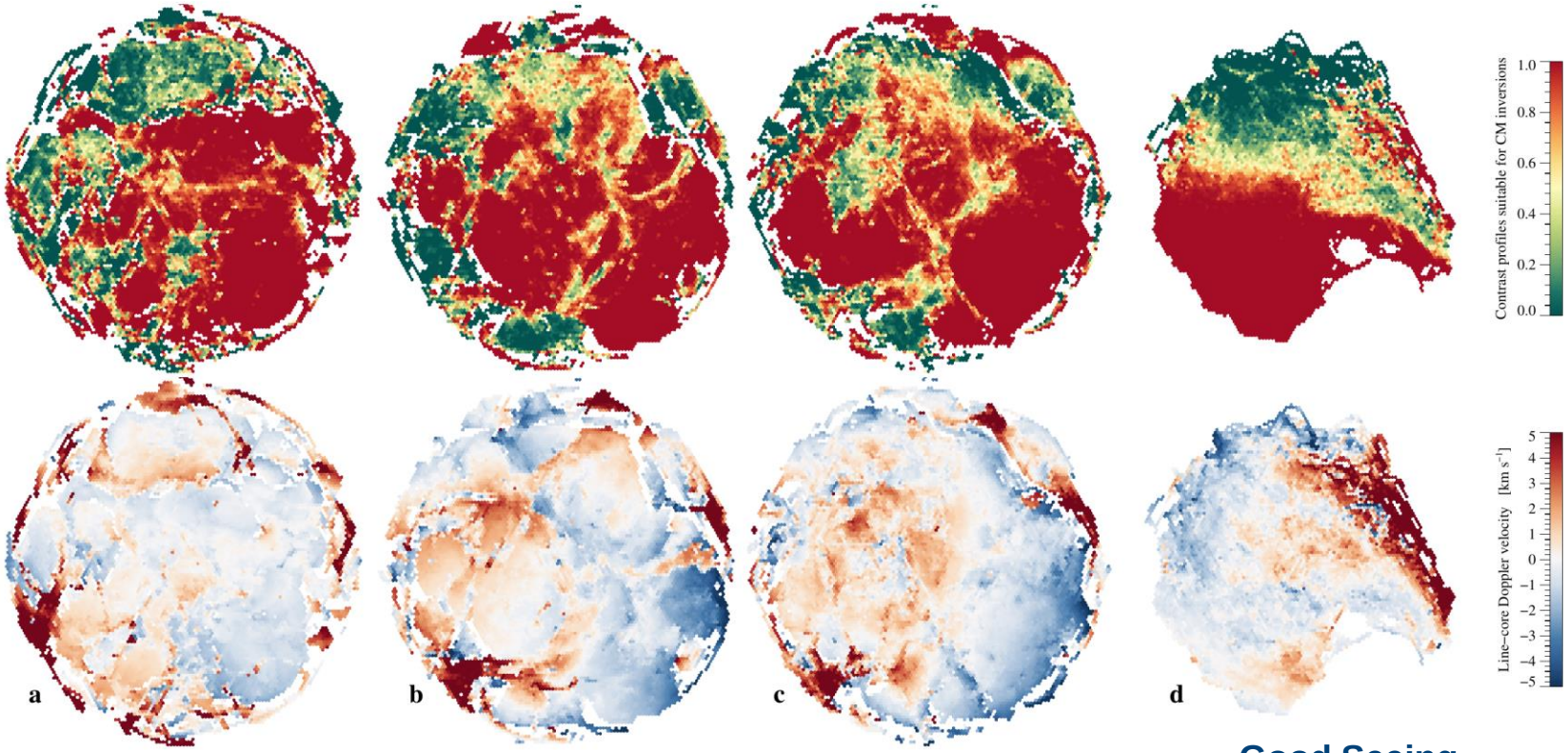


Parameter study – Careful selection

A1 The default parameters are fine, maybe the number of iteration has to be improved for large datasets



Good or Bad Seeing
PCA or Observed



Bad Seeing
Contrast profiles

Good Seeing
PCA coefficients

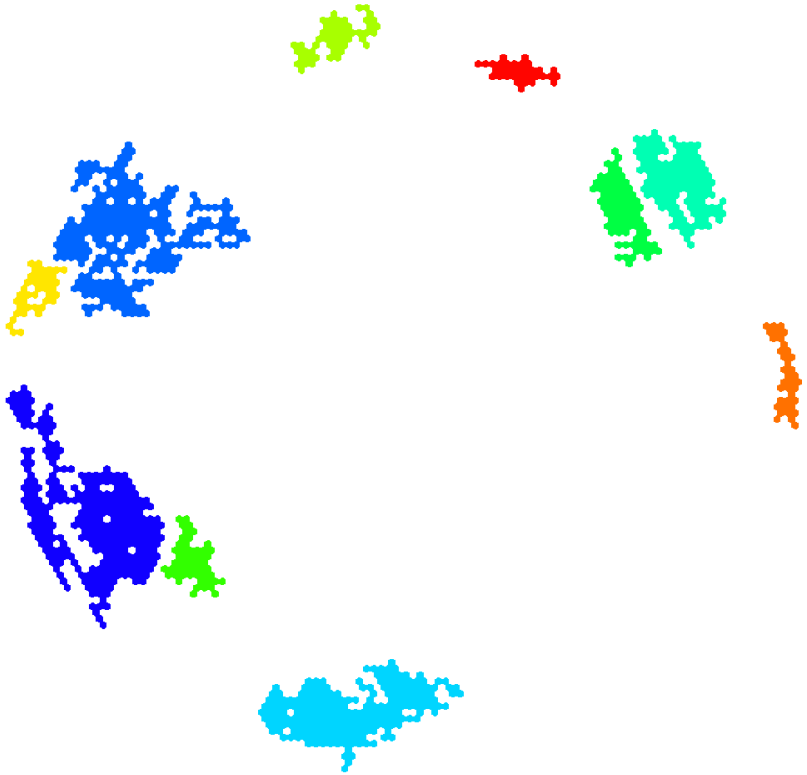
Good Seeing
Observed profiles

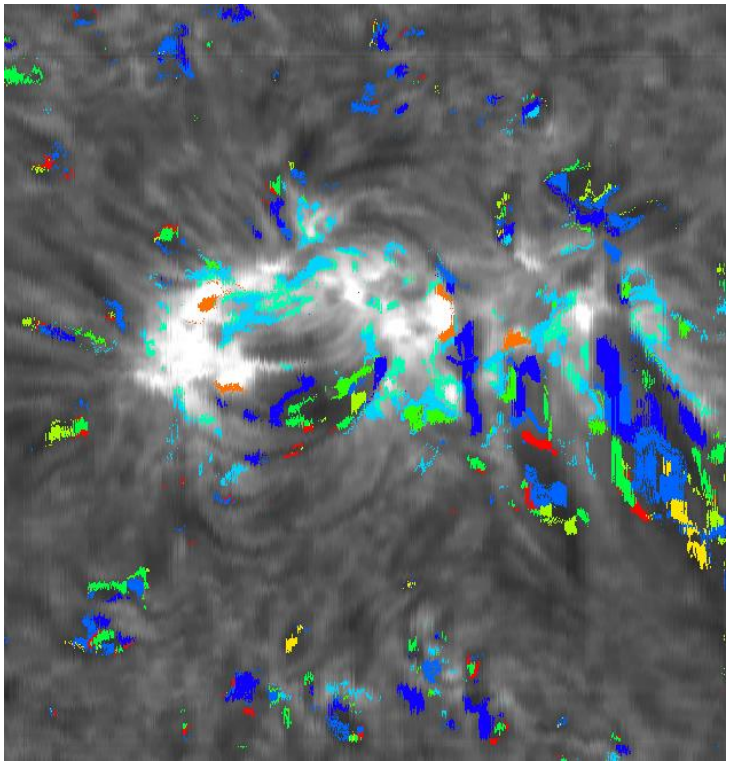
Good Seeing
Contrast profiles
UMAP

A2 Seeing does affect the projection but not much

A3 PCA coefficients and observed contrast profiles lead to similar results

Back mapping





- Selected the regions after using threshold of 0.9 and better.
- These are the regions where the profiles can surely be inverted using cloud model.

Conclusion

- t-SNE is a powerful tool to classify spectra.
- No prior information is needed.
- It classify good vs. bad profiles for inversion.
- Best settings perplexity = 50, theta = 0.5 and number of iterations = 1000 → based on time for computation and discerning power.
- Contrast as well as line profiles, PCA coefficients, PCA denoised or observed profiles lead to similar projection.
- Does show some differences for good and bad seeing.
- The regions which can surely be inverted using cloud models are discernible.
- OUTLOOK
- Performing cloud model inversions of selected regions in t-SNE projection.
- Projecting more data points on the already projected map.
- Comparison with UMAP.