

Finding star clusters in Gaia data

Tristan Cantat-Gaudin, Alfred Castro-Ginard, Carme Jordi, **Friedrich Anders**
(ICCUB)



gaia

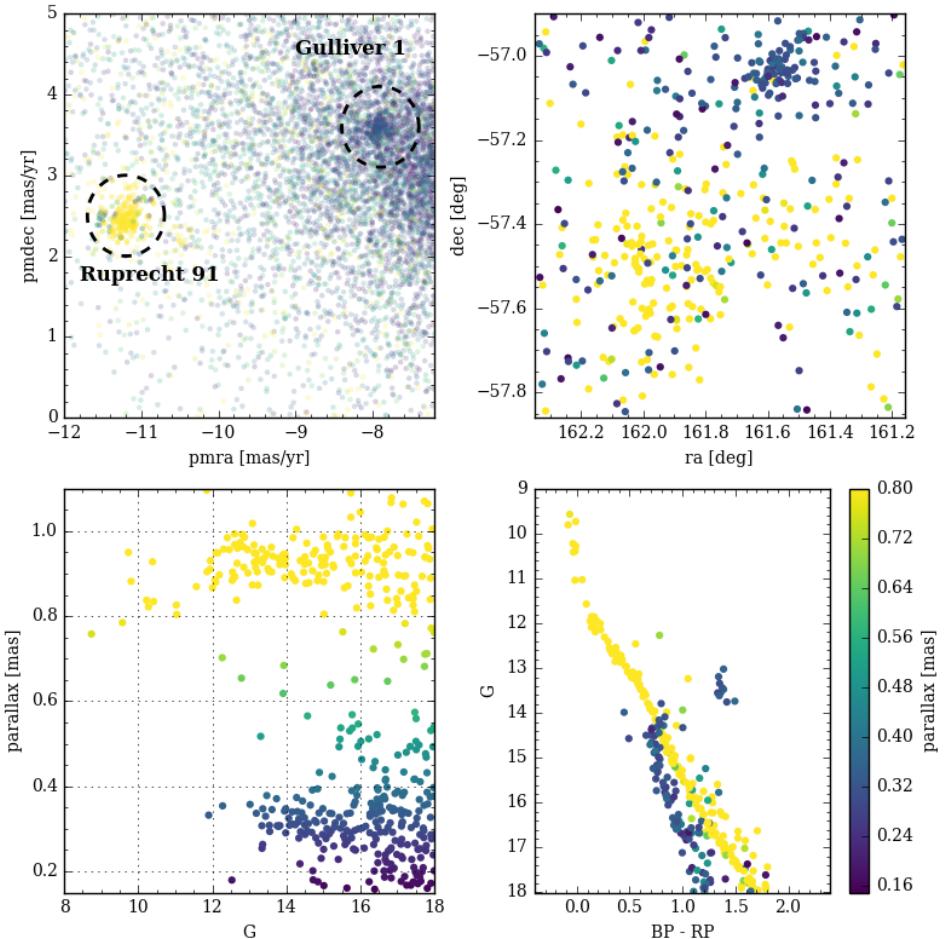


 **ICCUB**
Institute of Cosmos Sciences

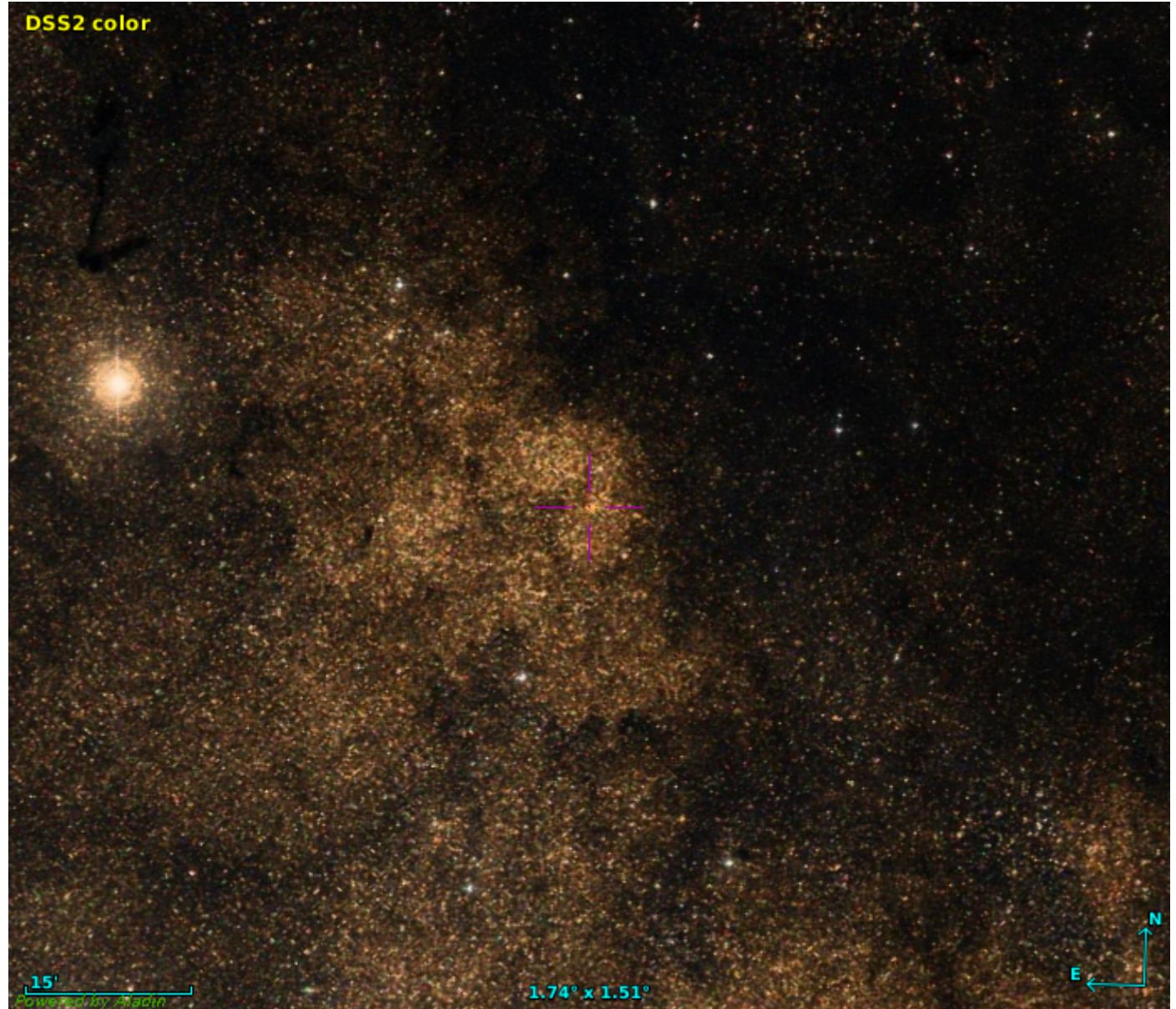
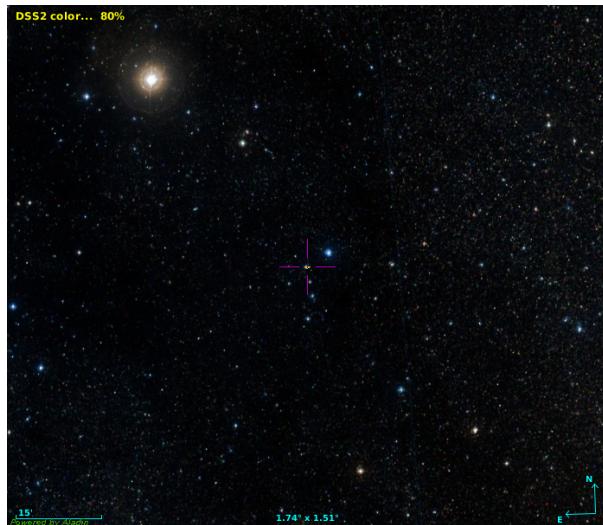
 EXCELENCIA
MARÍA
DE MAEZTU



UNIVERSITAT DE
BARCELONA



What is an open cluster?



Gaia: cluster finding in phase space

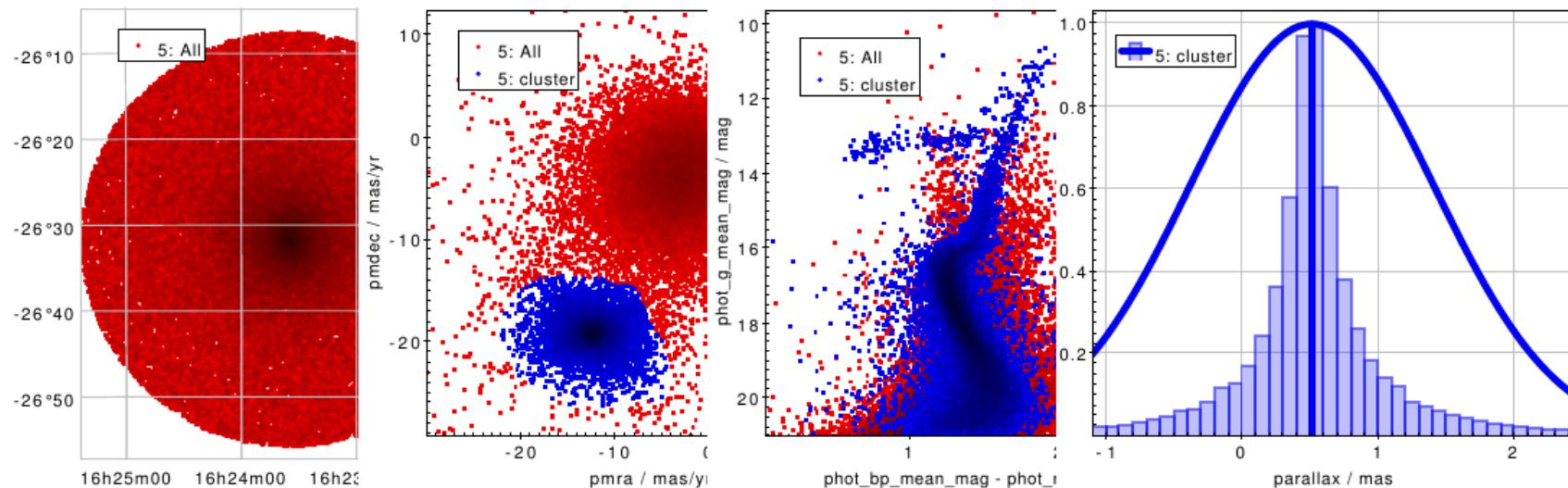
Example: Cluster Identification in 2d

Identify M4 members by proper motion clustering

- Cone Search M4, $r=0.3^\circ$; returns $\sim 62,000$ rows
- Sky Plot — cluster visible
- Proper Motion space plot `pmra` vs. `pmdec`
- Make graphical selection of comoving objects
- Colour-Magnitude diagram `bp_rp` vs. `phot_g_mean_mag`, view selection (**Subsets** tab)
- Plot cluster parallax histogram & Gaussian fit

Slide stolen from Mark Taylor
(yesterday's talk in Gaia session)

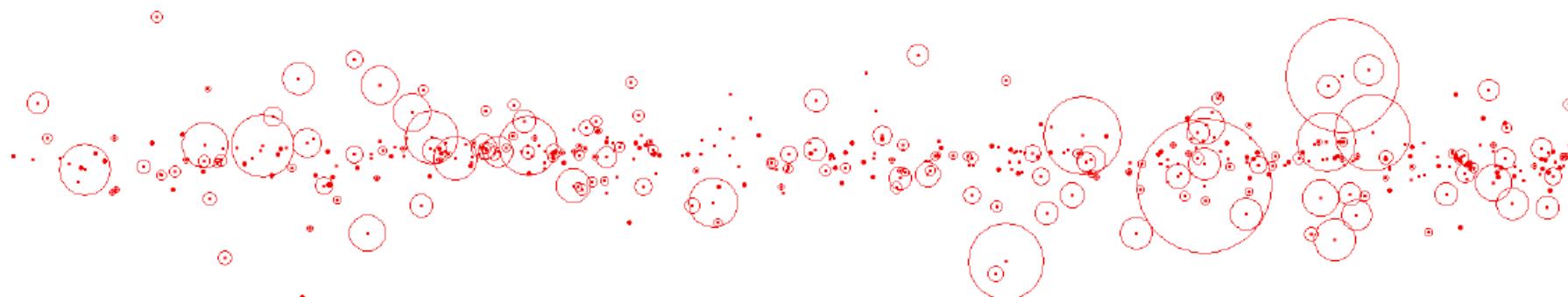
Before DR2, this only worked for the Pleiades!



The Gaia DR2 view of the Milky Way's open cluster population

April 25th 2018, 12:00:01

Started with an automatic query of the Gaia DR2 archive for every single open cluster listed in the catalogues of [Dias et al. \(2002\)](#) and [Kharchenko et al. \(2013\)](#) a total of **3383 cone searches**.



Only used Gaia DR2 sources brighter than G=18.

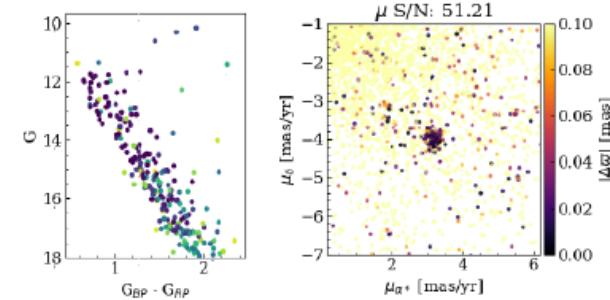
Analysed each investigated field with UPMASK ([Krone-Martins & Moitinho 2014](#)). The code makes groups in astrometric space (`pmra`,`pmdec`,`parallax`) and checks if they are more concentrated on the sky than a random distribution. It also takes into account the uncertainties on the astrometric parameters.

Clearly identified 1169 known and 60 previously unknown clusters!

A lot are missing because they are too distant or reddened to be identified on Gaia DR2 alone.

Some clusters recently discarded as “non-existent” are real!

e.g. NGC 886, sparse, high extinction,
but very compact in astrometric space



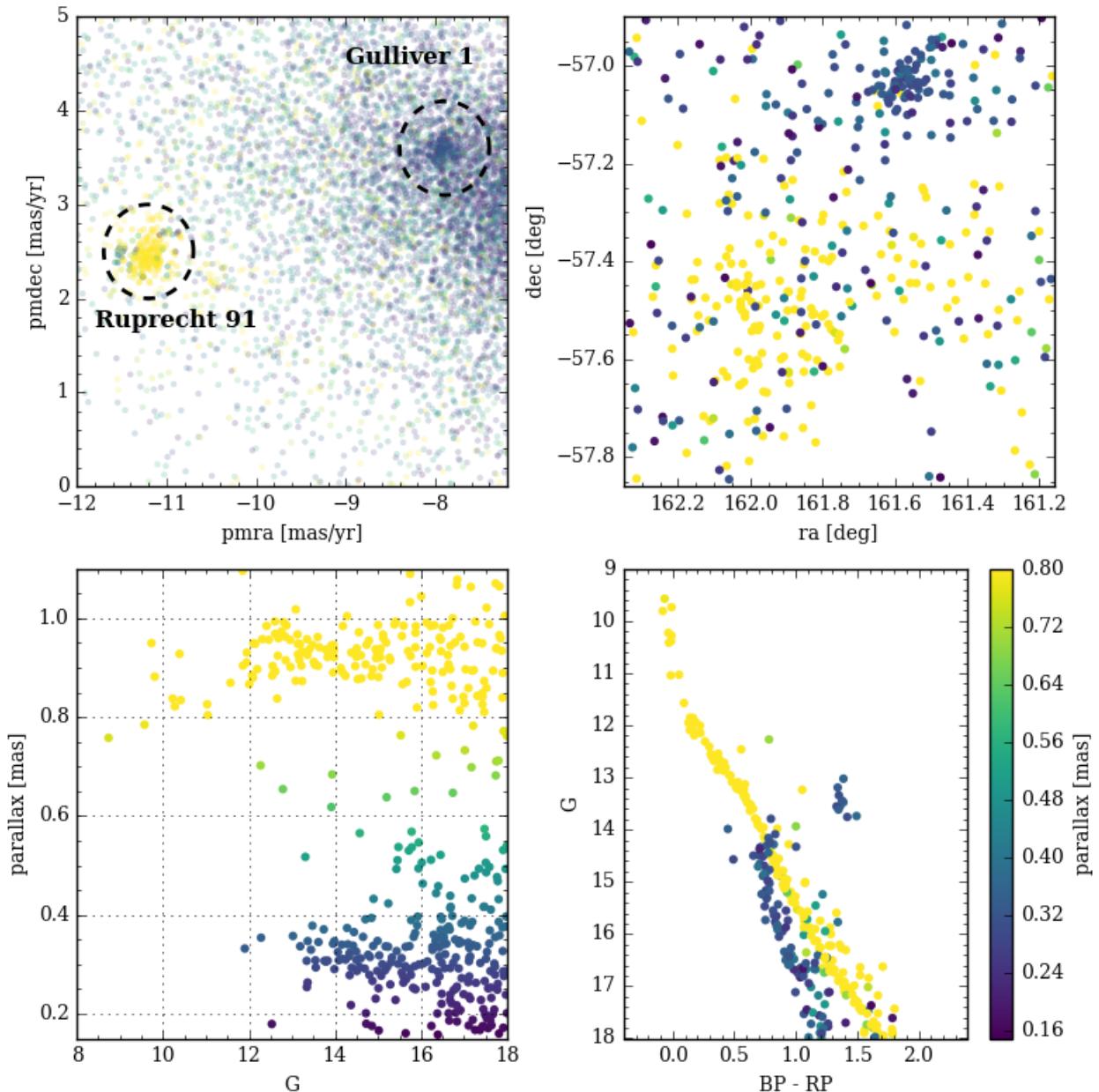
Cantat-Gaudin+2018: serendipitous Gaia DR2 discoveries

Basically:

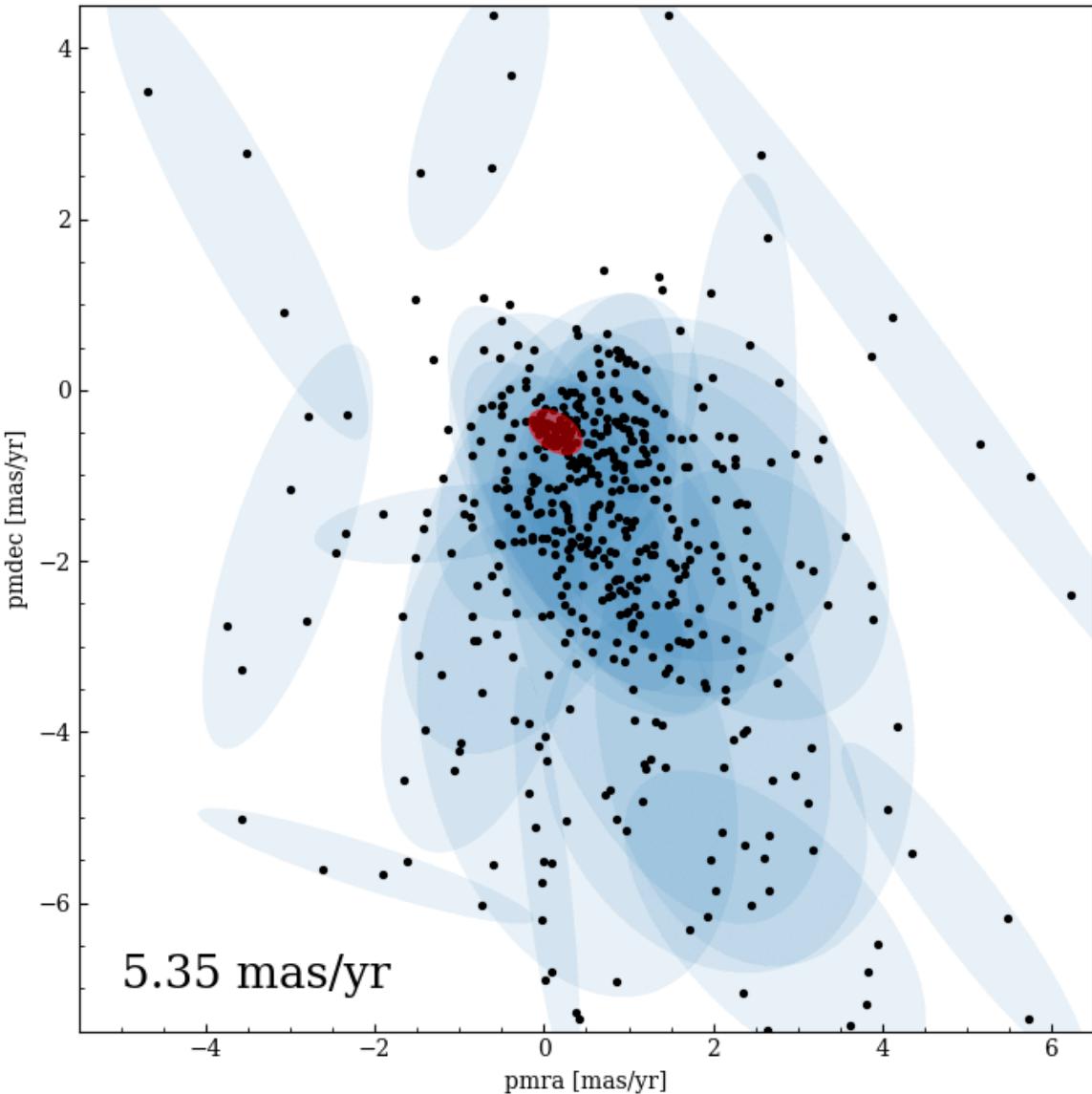
No clustering was needed at all.
New discoveries could be made **by eye**.

The Gaia DR2 data are often so good that they look like textbook examples for clustering algorithms.

It was clear that much more could be found...

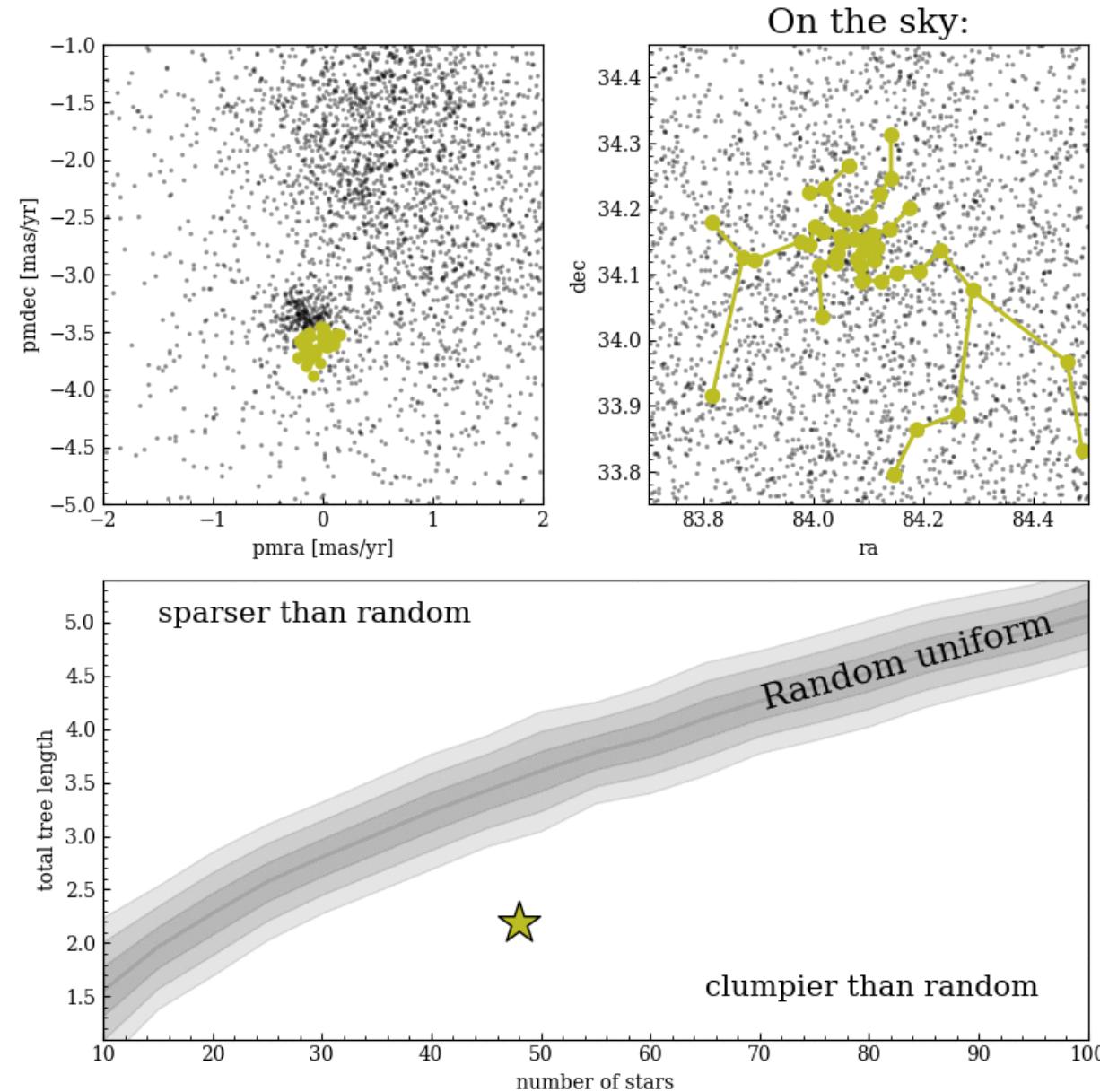


Cantat-Gaudin+2019: Applying GMM in the Perseus region



Step 1:
Use Gaussian mixture model to look
for sharp peaks in the velocity
distribution
→ cluster candidates

Cantat-Gaudin+2019: Applying GMM in the Perseus region



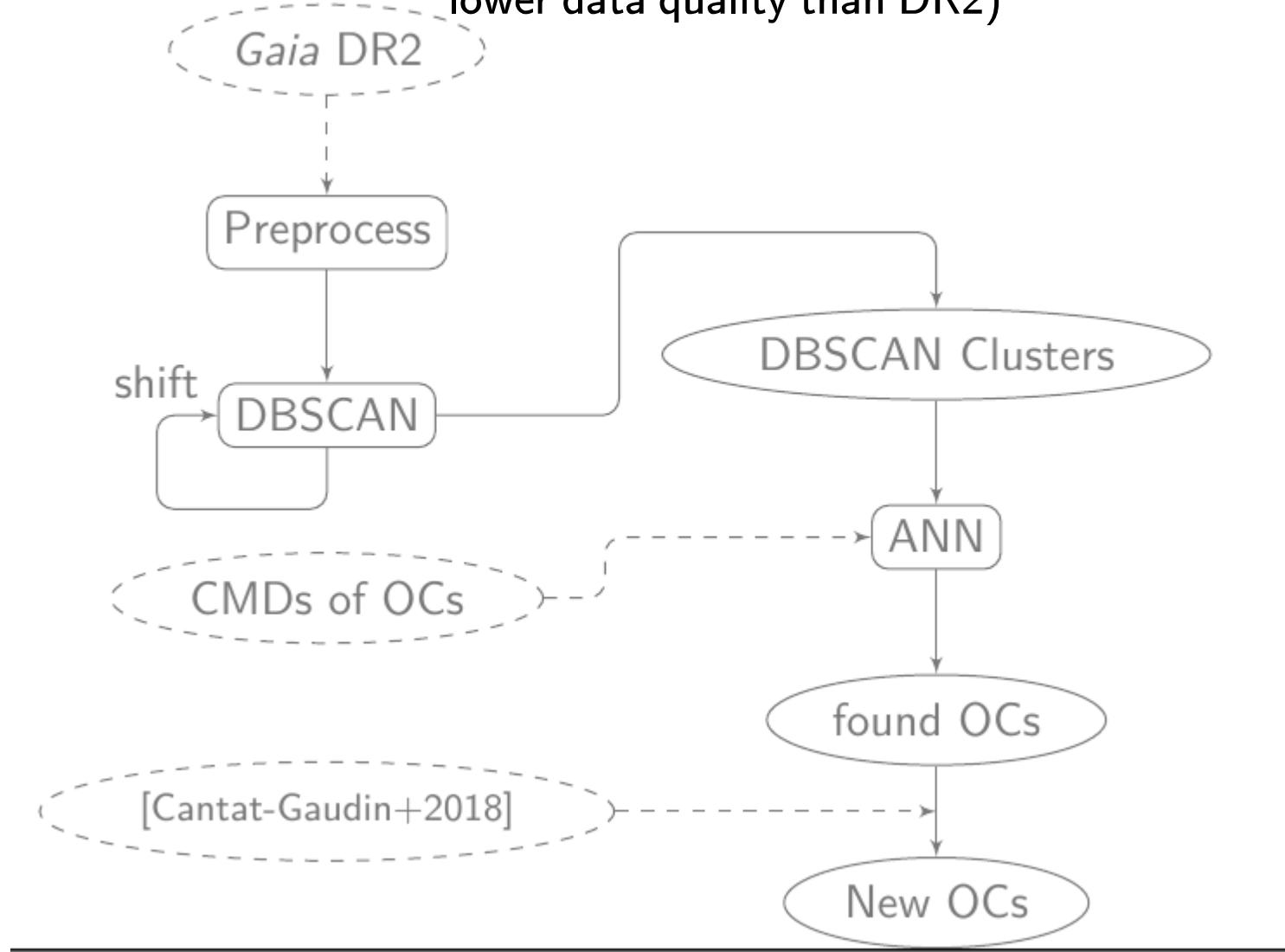
Step 2 (UPMASK):
K-means clustering in 5D astrometric space

Step 3 (UPMASK):
Compare minimum spanning tree distance with random distribution

→ confirm cluster candidates

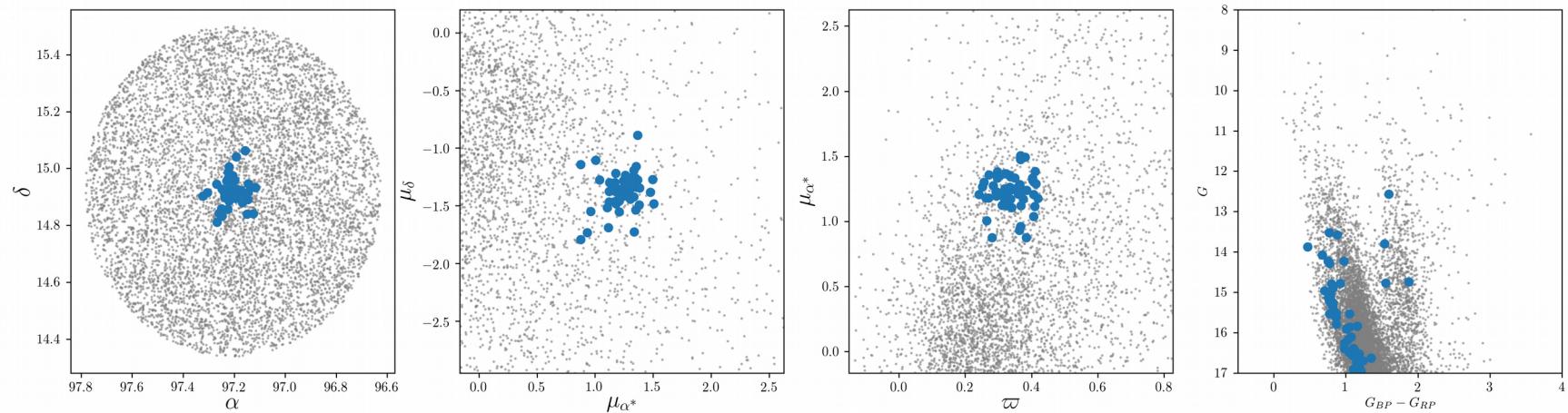
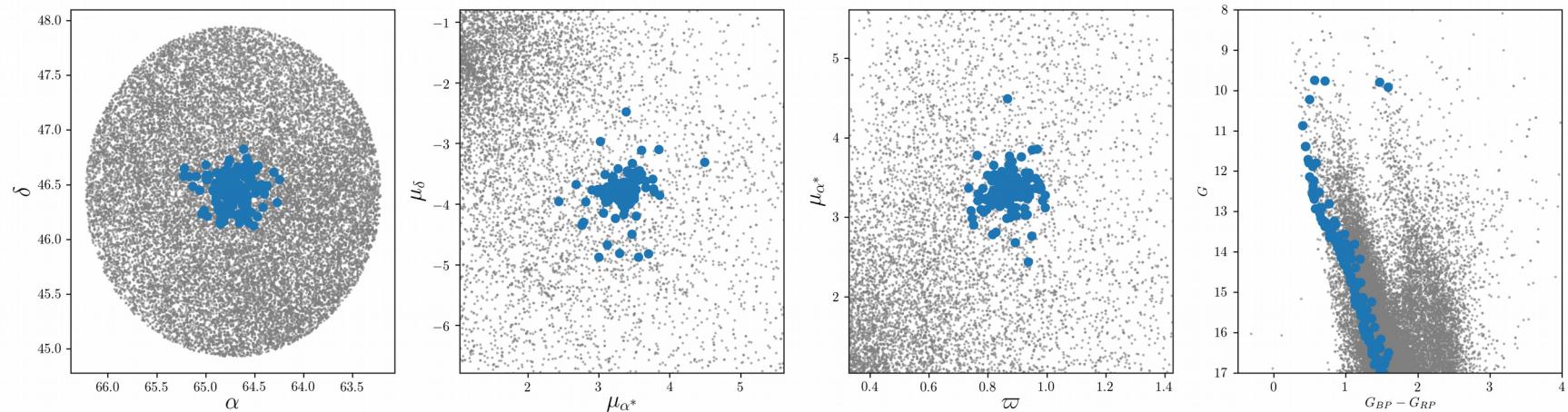
Castro-Ginard+2019: Applying DBSCAN in the Perseus region

Castro-Ginard+2018 already found 41 new clusters in Gaia DR1 (2M stars with much lower data quality than DR2)



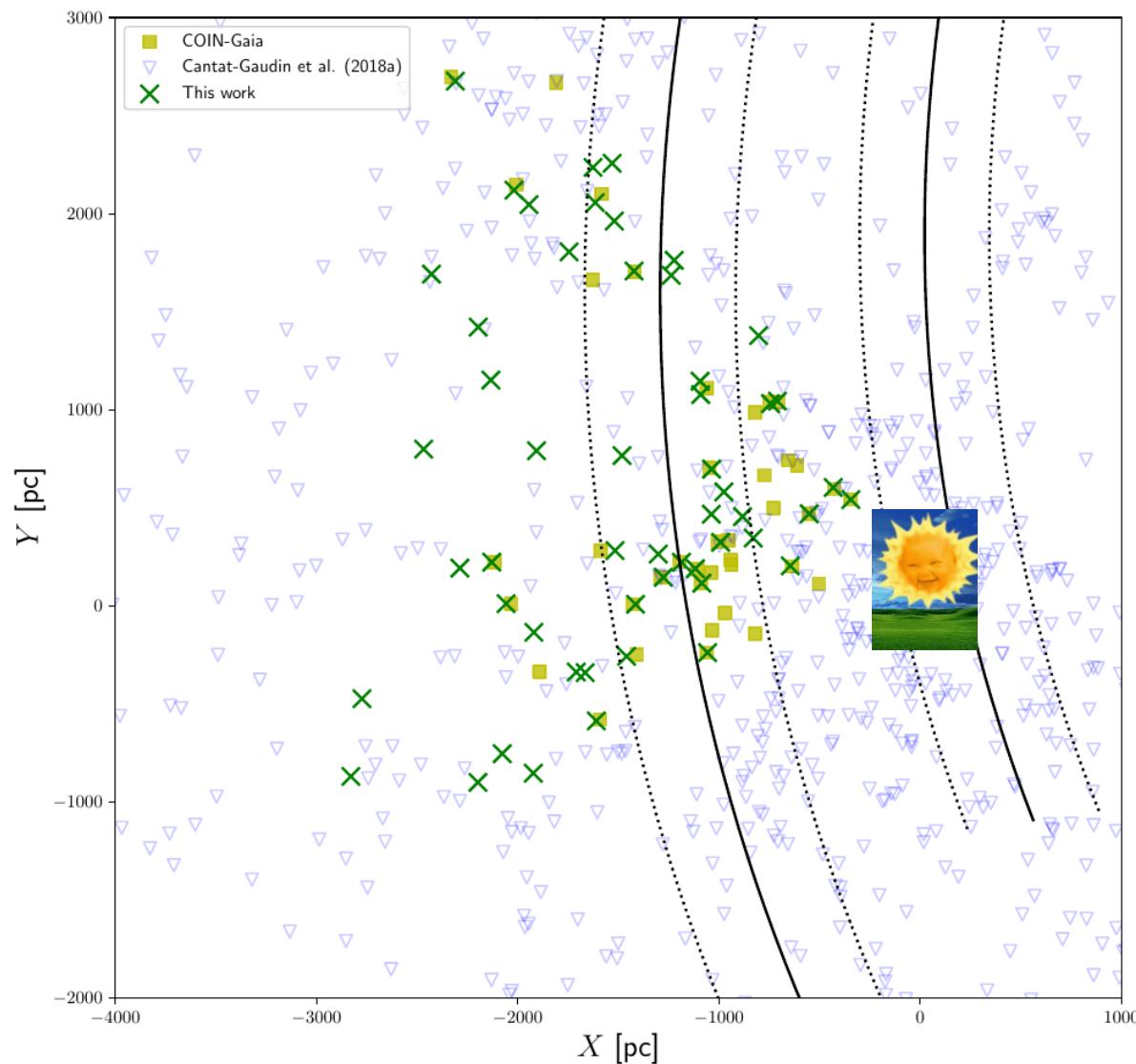
Castro-Ginard+2019: Applying DBSCAN in the Perseus region

5D clustering + simple ANN to distinguish cluster and asterism CMDs (90% accuracy)

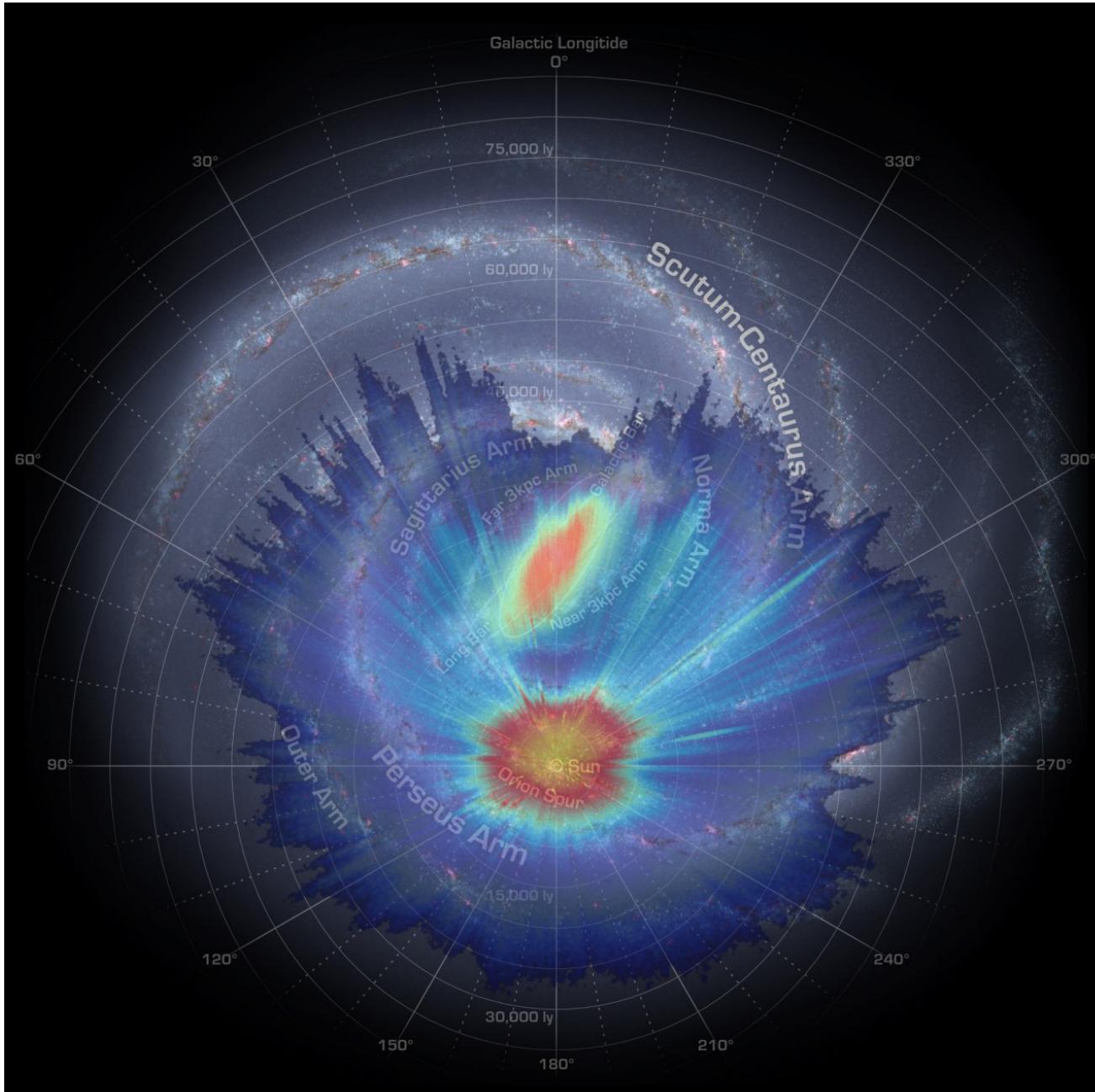


Castro-Ginard+2019: Applying DBSCAN in the Perseus arm

Some in common with Cantat-Gaudin+2019, but also ~40 new ones



Gaia DR2: What's still out there to discover



Most new discoveries are
still within ~2 kpc.

Gaia DR2 covers a much
larger volume than the one
probed by current open-
cluster works!

So more sophisticated
searches are promising.

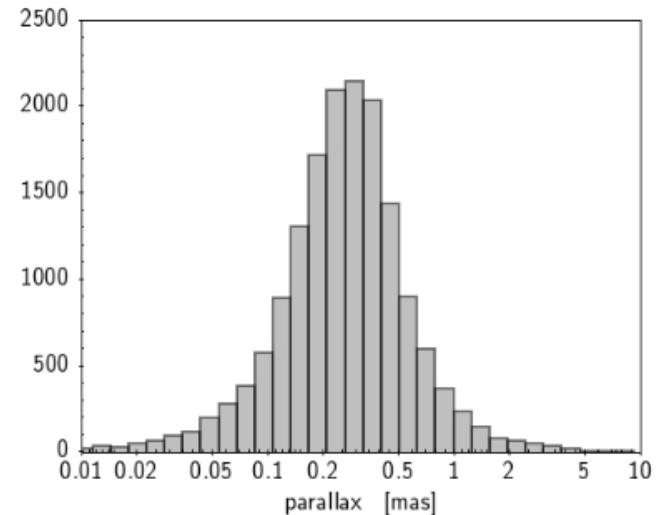
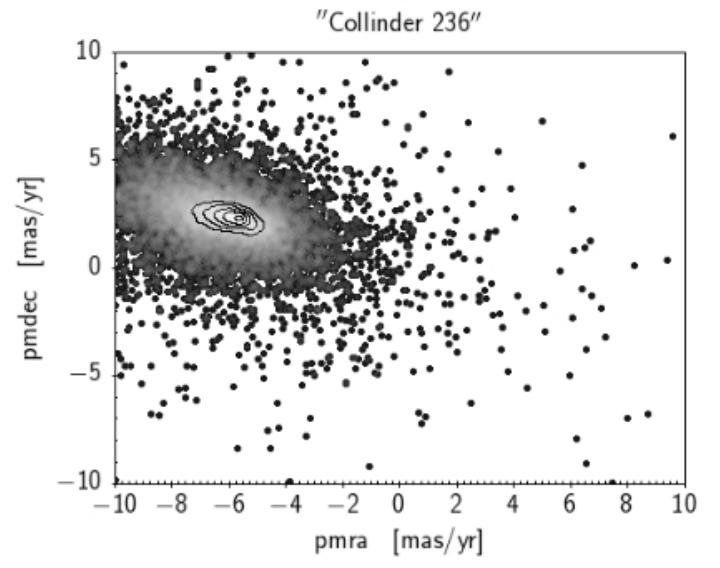
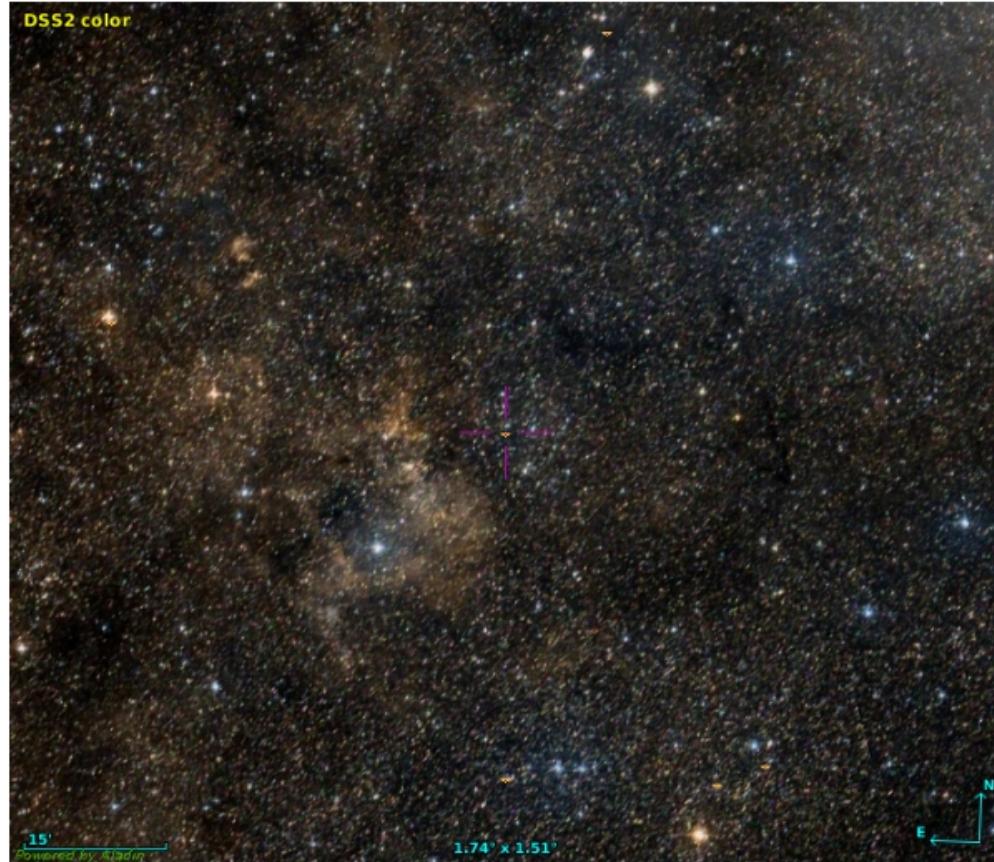
Expect a new ~1000
clusters from Castro-
Ginard et al. soon!

Why it isn't sooo easy in detail

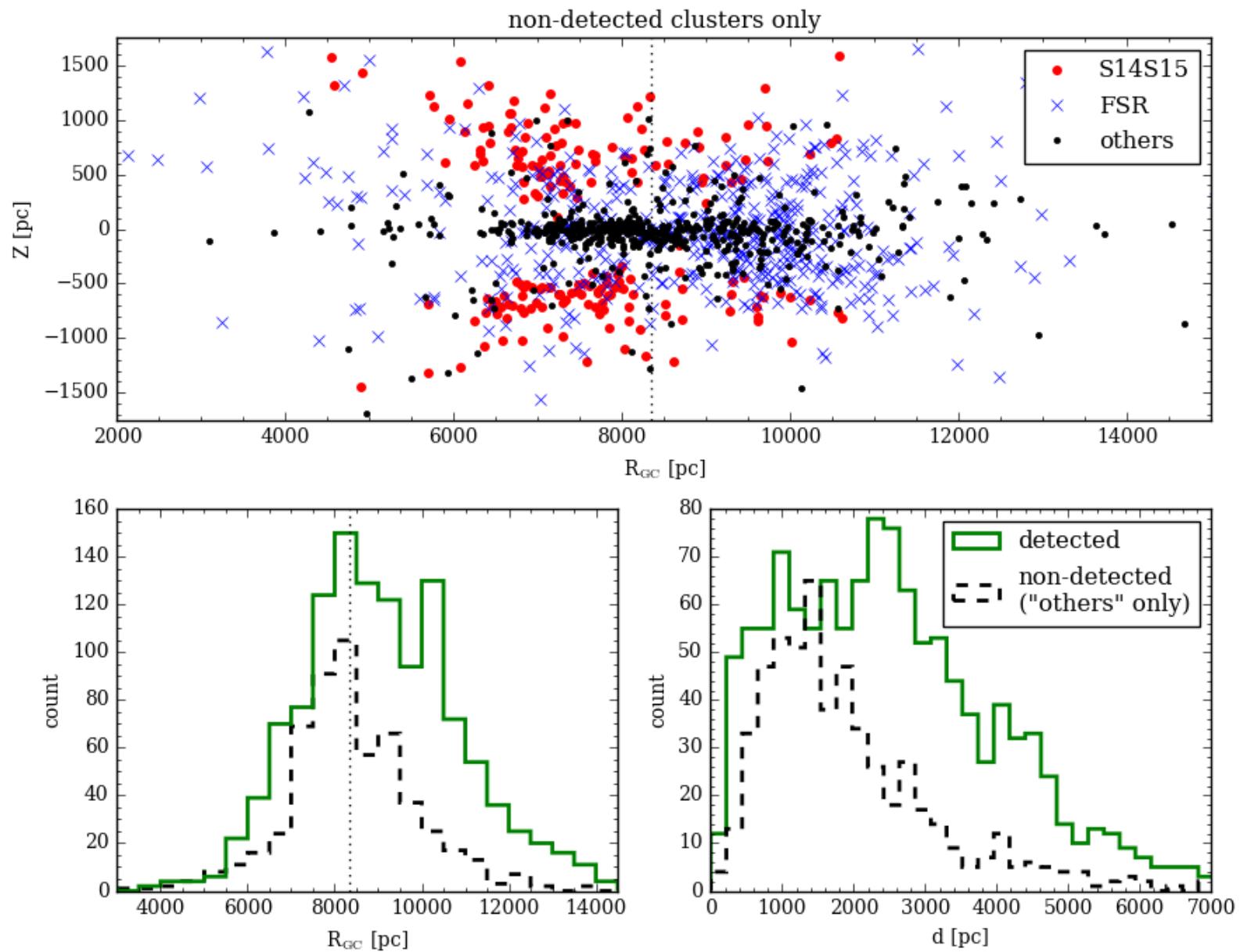


Look at the abstract for this talk
and compare to its content:)

Cantat-Gaudin+2018: Heaps of asterisms in the OC literature



Cantat-Gaudin+2018: Heaps of asterisms in the OC literature



Asterisms: Why we get fooled



In the discovery process:

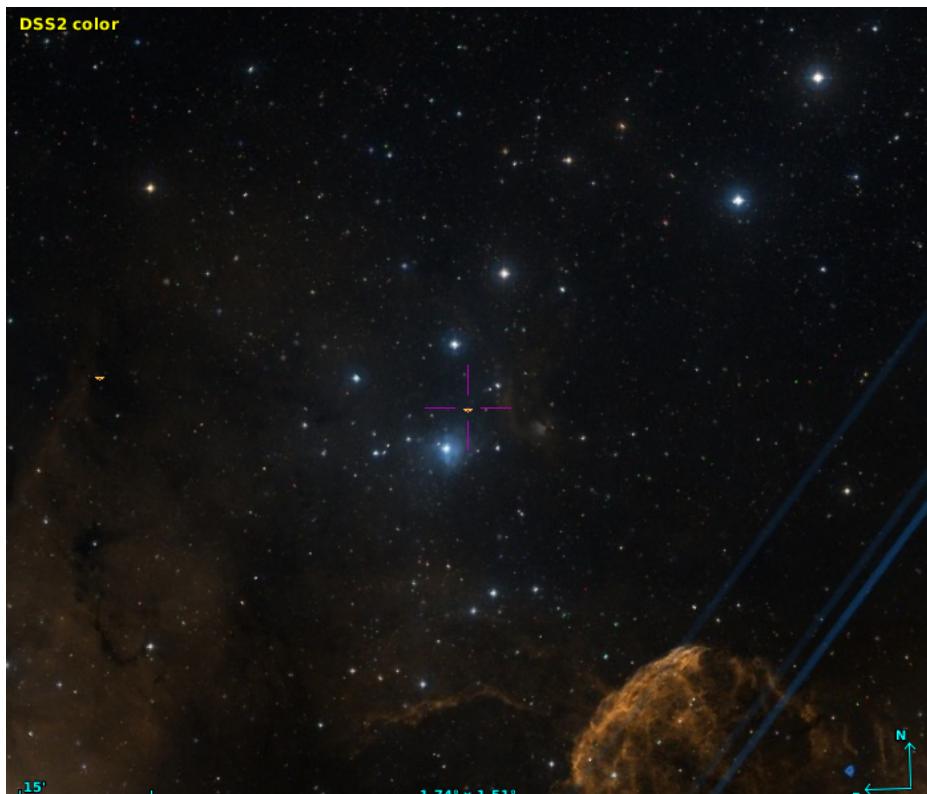
- because we want to discover *something*
- because 1-2 bright stars dominate the light profile and fool the eye
- because of holes in the dust distribution
- because of zonal effects in ancillary Schmidt plates

In literature work:

- because of tradition
- to avoid conflicts with referees and established researchers

Conclusions

- 1) Gaia DR2 is an optimal playground for clustering! We are not even doing the most fancy things. Almost every time we try a new algorithm, we find something new.
- 2) The hard work is not in *finding* candidates - but in vetting and interpreting them.
What is a good metric in 5D astrometric space?
- 3) Don't use catalogues without thought.
Nobody will ever write a paper about the non-existence of Basel 5.
That doesn't mean it actually exists.



Tristan Cantat-Gaudin @CantatGaudin · 25. Juni

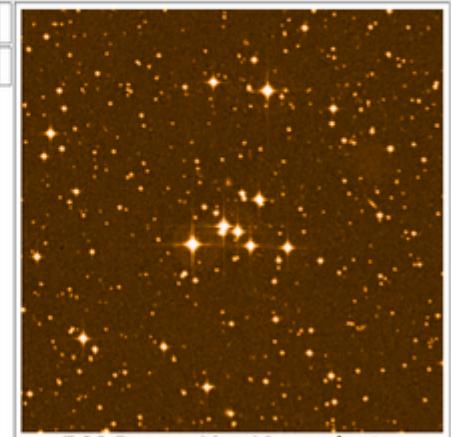
Antwort an @frediferente

My favourite asterism at the moment is Ruprecht 3. So deceitful!

[Tweet übersetzen](#)

[WEBDA page for open cluster Ruprecht 3](#)

Basic Parameters	
Right Ascension (2000)	06 42 07
Declination (2000)	-29 27 00
Galactic longitude	238.776
Galactic latitude	-14.815
Distance [pc]	
Reddening [mag]	
Distance modulus [mag]	
Log Age	
Metallicity	
Notes	



References

Discoveries in Gaia DR1:

- Castro-Ginard et al. (2018), A&A, 618, A59, arXiv:1805.03045 (based on DR1)

Gaia DR2 census of open star clusters (including 60 serendipitous new discoveries):

- Cantat-Gaudin et al. (2018), A&A, 618, A93, arXiv:1805.08726

Subsequent Gaia discoveries:

- Cantat-Gaudin et al. (2019), A&A, 624, A126, arXiv:1810.05494
- Castro-Ginard et al. (2019), A&A, accepted, arXiv:1905.06161

Mark Taylor's talk in the Gaia session:

<https://www.gaia.ac.uk/science/workshops/gaia-dr2-NAM2019-jul19>

Tristan's talk at ESLAB53: <https://zenodo.org/record/2672400#.XRzJskN7IE4>

Alfred's talk at ESLAB53: <https://zenodo.org/record/2784065#.XR0kFkN7IE4>

Pre-Gaia literature containing many likely false positives:

- Dias catalogue: Dias et al. (2002-2015), <http://cdsarc.u-strasbg.fr/viz-bin/cat/B/ocl>
- Milky Way Star Cluster Catalogue (MWSC): Kharchenko et al. (2013), Schmeja et al. (2014), Scholz et al. (2015)
- Froebrich et al. (2007), MNRAS, 374, 399

...