DR3 cross-match and the HPM sources

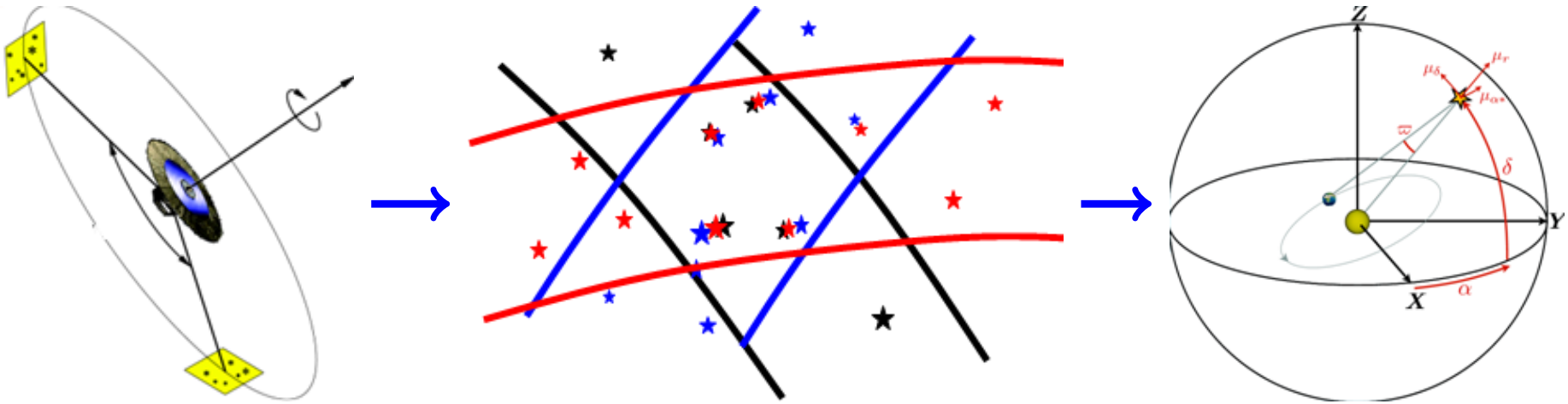
Ferran Torra Clotet
on behalf of the Gaia Barcelona Team

Institut de Ciències del Cosmos - Universitat de Barcelona
DAPCOM

Gaia RIA Workshop, Barcelona, 17th February 2020

XM in the Gaia Data Reduction

- Gaia scanning law ensures that each sky region is observed several times with nearly isotropic orientations.
- Source Identification to populate the Gaia catalogue and link the observations: **XM of Gaia observations**.
- Source parameters are then updated based on the observations linked through the cross-match solution.



Cross-matching Gaia observations

1. **XM based on clustering analysis**: group all the observations from each individual source.
   - We introduce a dissimilarity in accordance with the source model.
   - We agglomerate the observations using the Nearest Neighbor Chain algorithm, with an upper limit on the internal variance of a cluster.

2. **Source identification**: the identified clusters are linked to the existing entries in the working source list, or new entries are created from scratch.

Cross-matching Gaia observations

1. **XM based on clustering analysis**: group all the observations from each individual source.
   - We introduce a dissimilarity in accordance with the source model.
   - We agglomerate the observations using the Nearest Neighbor Chain algorithm, with an upper limit on the internal variance of a cluster.

2. **Source identification**: the identified clusters are linked to the existing entries in the working source list, or new entries are created from scratch.

Cluster analysis - Source model

Zeroth-order model (DR2)

The model is based on positional coincidence.

$$\Delta(C_i, C_j) = \frac{n_i n_j}{n_i + n_j} \|\boldsymbol{x}(C_i) - \boldsymbol{x}(C_j)\|^2$$

where $\boldsymbol{x}(C_i)$ is the center of the cluster C_i and n_i is the number of observations in the cluster C_i.

Proper motion model (DR3)

Let $u(t)$ be any of the coordinate functions, we define the linear model

$$u(t) = u_0 + u_1 t$$

where u_0 is the mean position and u_1 is the proper motion.

$$\Delta_u(C_i, C_j) = (\widehat{\boldsymbol{u}}_i - \widehat{\boldsymbol{u}}_j)^T \boldsymbol{N}_i(\boldsymbol{N}_i + \boldsymbol{N}_j)^{-1}\boldsymbol{N}_j(\widehat{\boldsymbol{u}}_i - \widehat{\boldsymbol{u}}_j)$$

where $\boldsymbol{N}_i$ is the normal matrix of the cluster C_i.

Cluster analysis - Source model

Zeroth-order model (DR2)

The model is based on positional coincidence.

$$\Delta(C_i, C_j) = \frac{n_i n_j}{n_i + n_j} \|\boldsymbol{x}(C_i) - \boldsymbol{x}(C_j)\|^2$$

where $\boldsymbol{x}(C_i)$ is the center of the cluster C_i and n_i is the number of observations in the cluster C_i.

Proper motion model (DR3)

Let $u(t)$ be any of the coordinate functions, we define the linear model

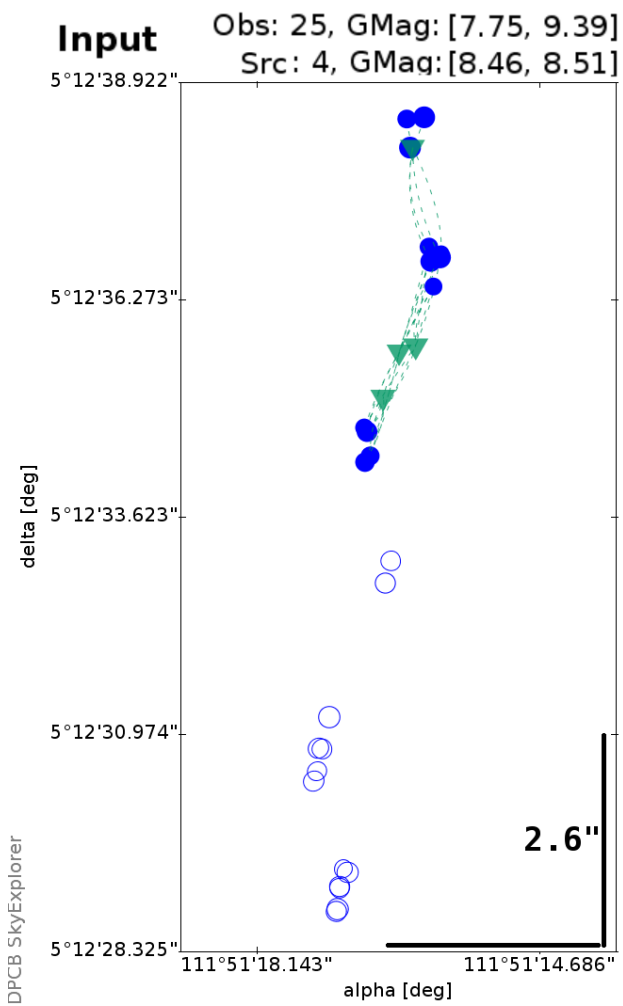
$$u(t) = u_0 + u_1 t$$

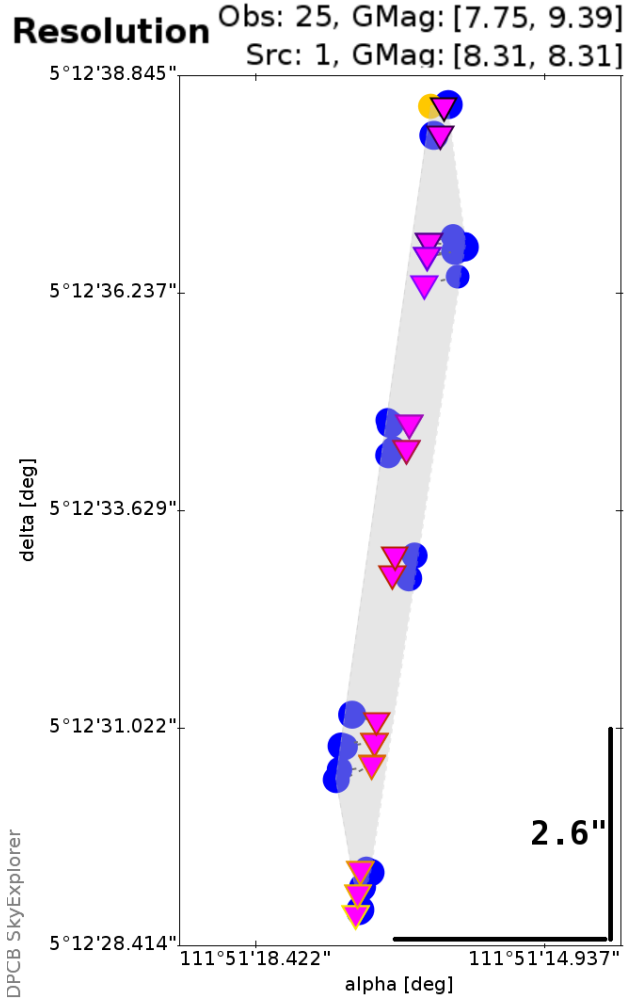
where u_0 is the mean position and u_1 is the proper motion.

$$\Delta_u(C_i, C_j) = (\widehat{\boldsymbol{u}}_i - \widehat{\boldsymbol{u}}_j)^T \boldsymbol{N}_i (\boldsymbol{N}_i + \boldsymbol{N}_j)^{-1} \boldsymbol{N}_j (\widehat{\boldsymbol{u}}_i - \widehat{\boldsymbol{u}}_j)$$

where $\boldsymbol{N}_i$ is the normal matrix of the cluster C_i.

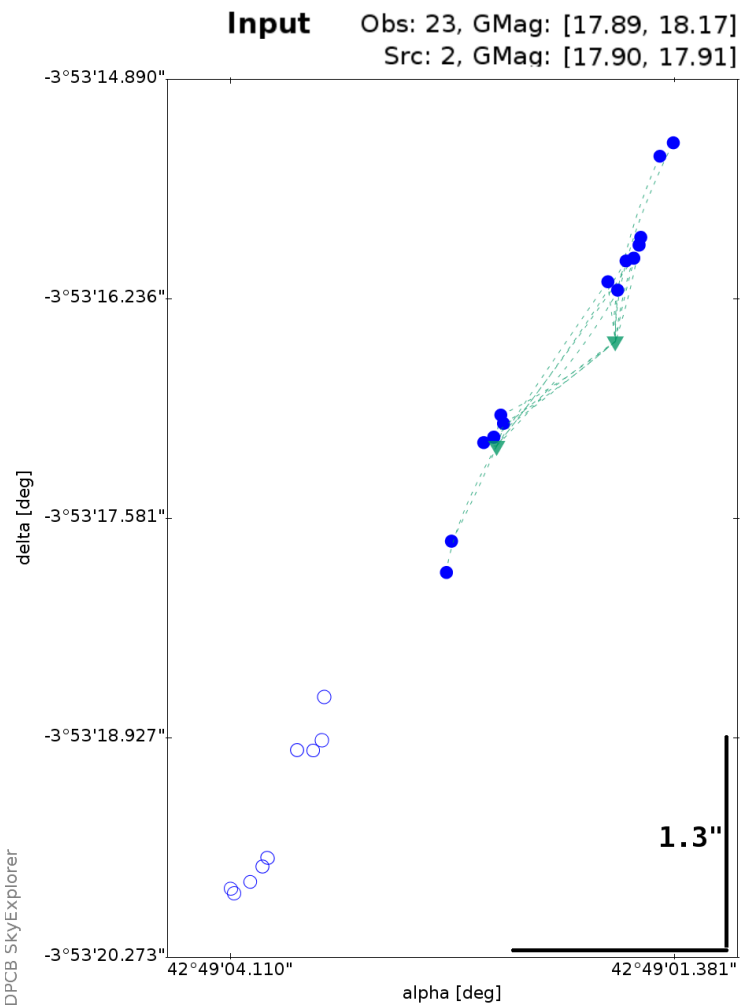
HPM source in DR3

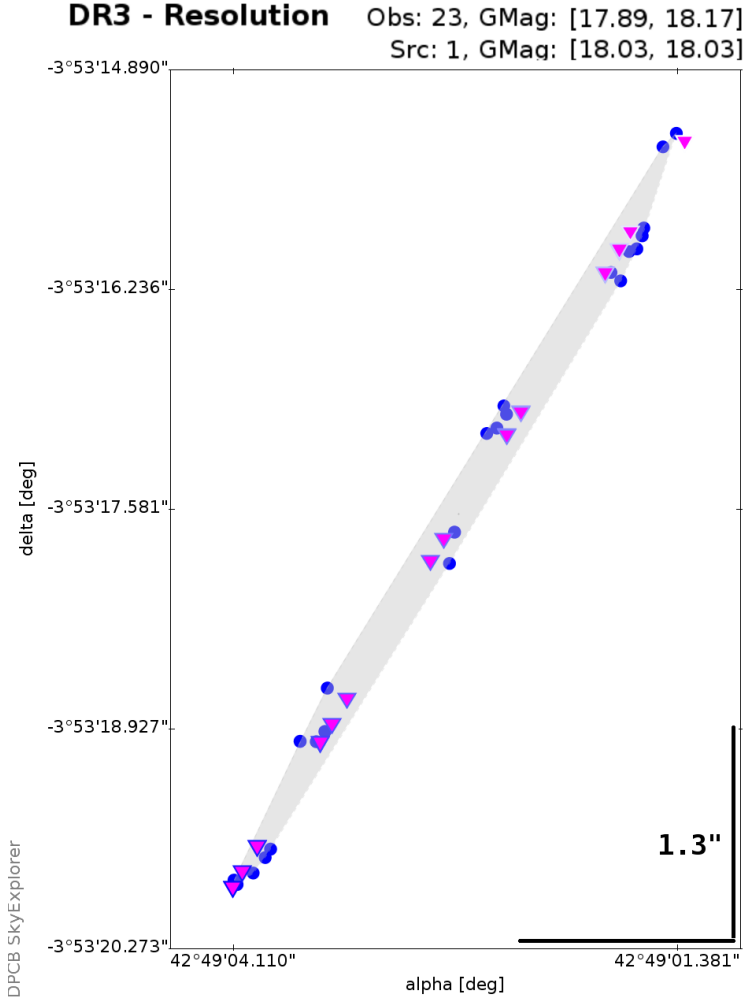




Observation DR2 source New source propagated at obs. epoch

HPM source in DR3





Observation DR2 source New source propagated at obs. epoch

Source identification

Each new cluster analysis starts from scratch ignoring any previous match solution.

Then the clusters have to be linked to the sources of the previous catalogue.

The possible resolutions of the sources of the previous catalogue are:

- **persisting source**: linked to a unique cluster,
- **merged source**: two (or more) sources linked to a unique cluster,
- **split source**: linked to more than one cluster,
- **source w/o matches**: deleted source.

Source identification

Each new cluster analysis starts from scratch ignoring any previous match solution.

Then the clusters have to be linked to the sources of the previous catalogue.

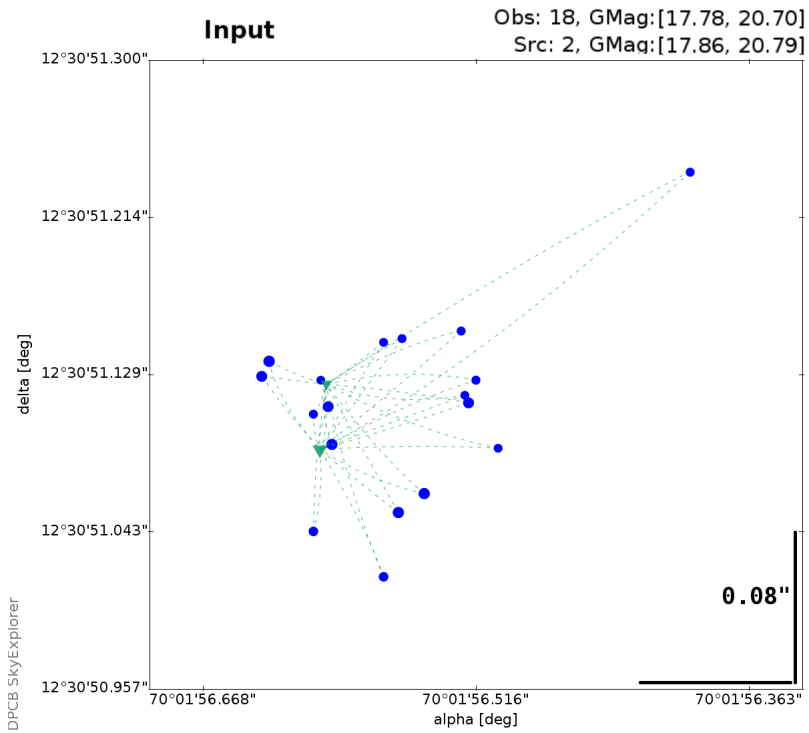
The possible resolutions of the sources of the previous catalogue are:

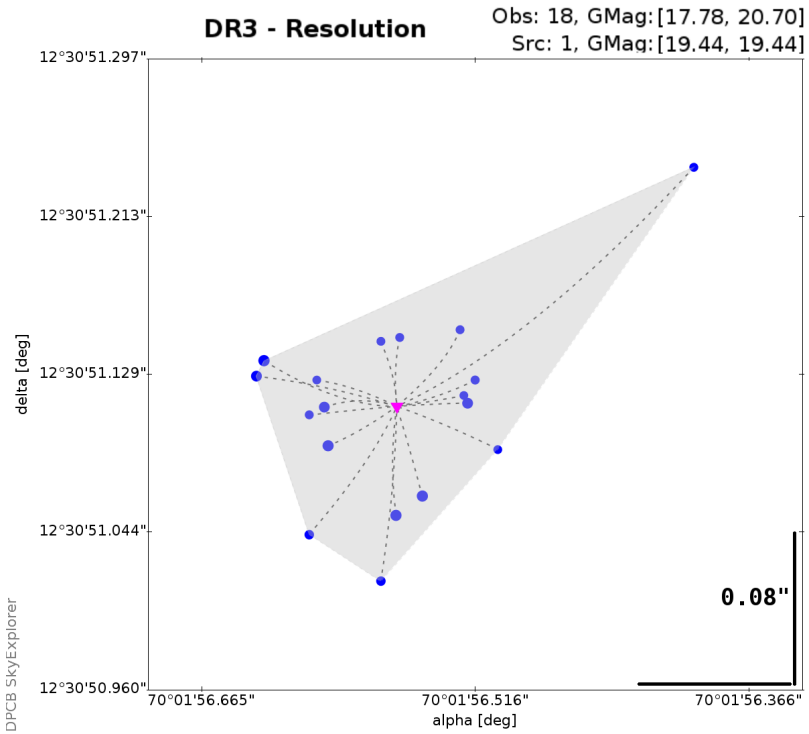
- **persisting source**: linked to a unique cluster,
- **merged source**: two (or more) sources linked to a unique cluster,
- **split source**: linked to more than one cluster,
- **source w/o matches**: deleted source.

Sources merged

The new HPM sources are usually created by merge.

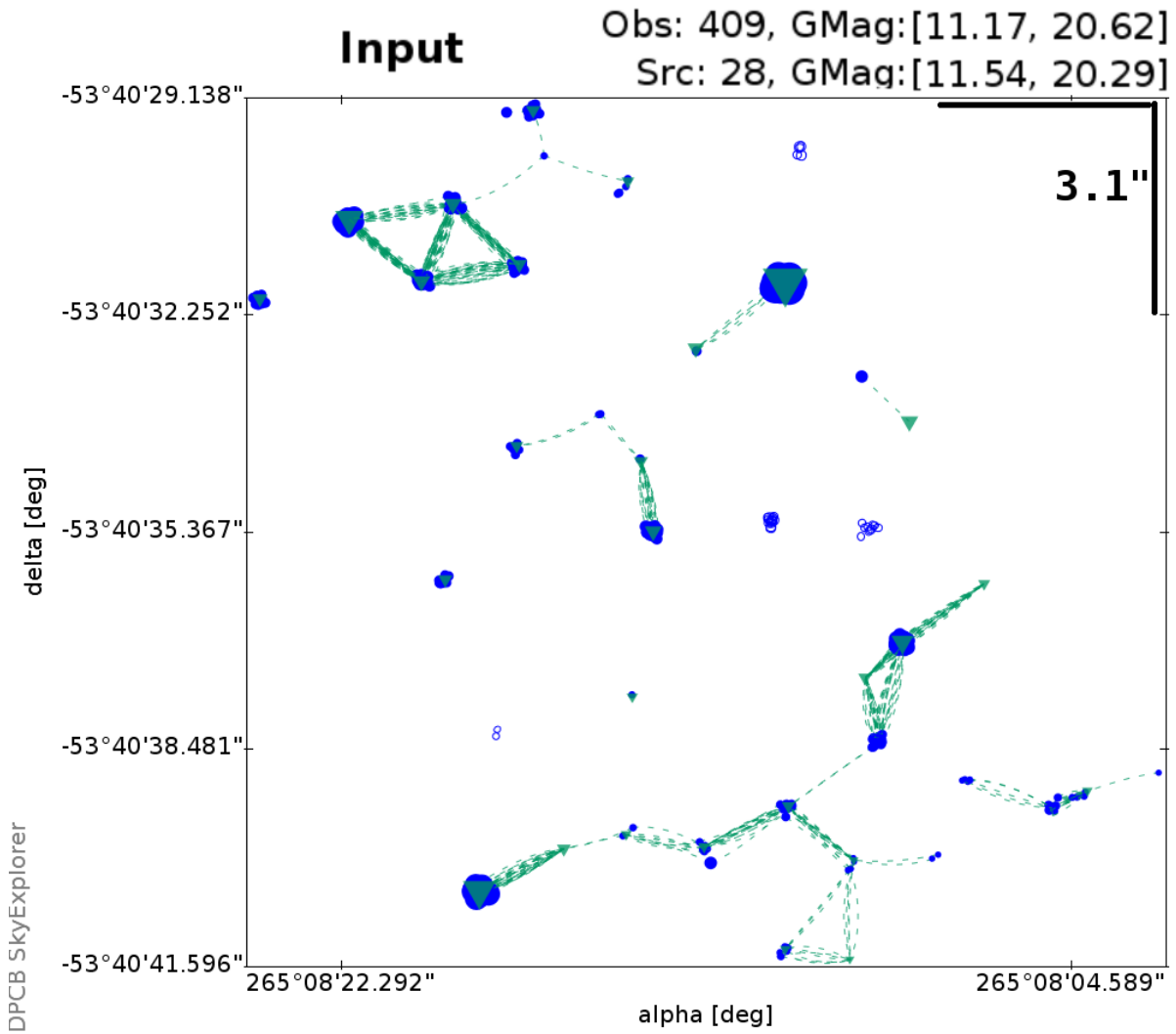
A variable star was split into multiple SourceIds in DR2: the brighter transits being grouped with one SourceId, and the fainter with the other. In DR3, these sources will be merged.





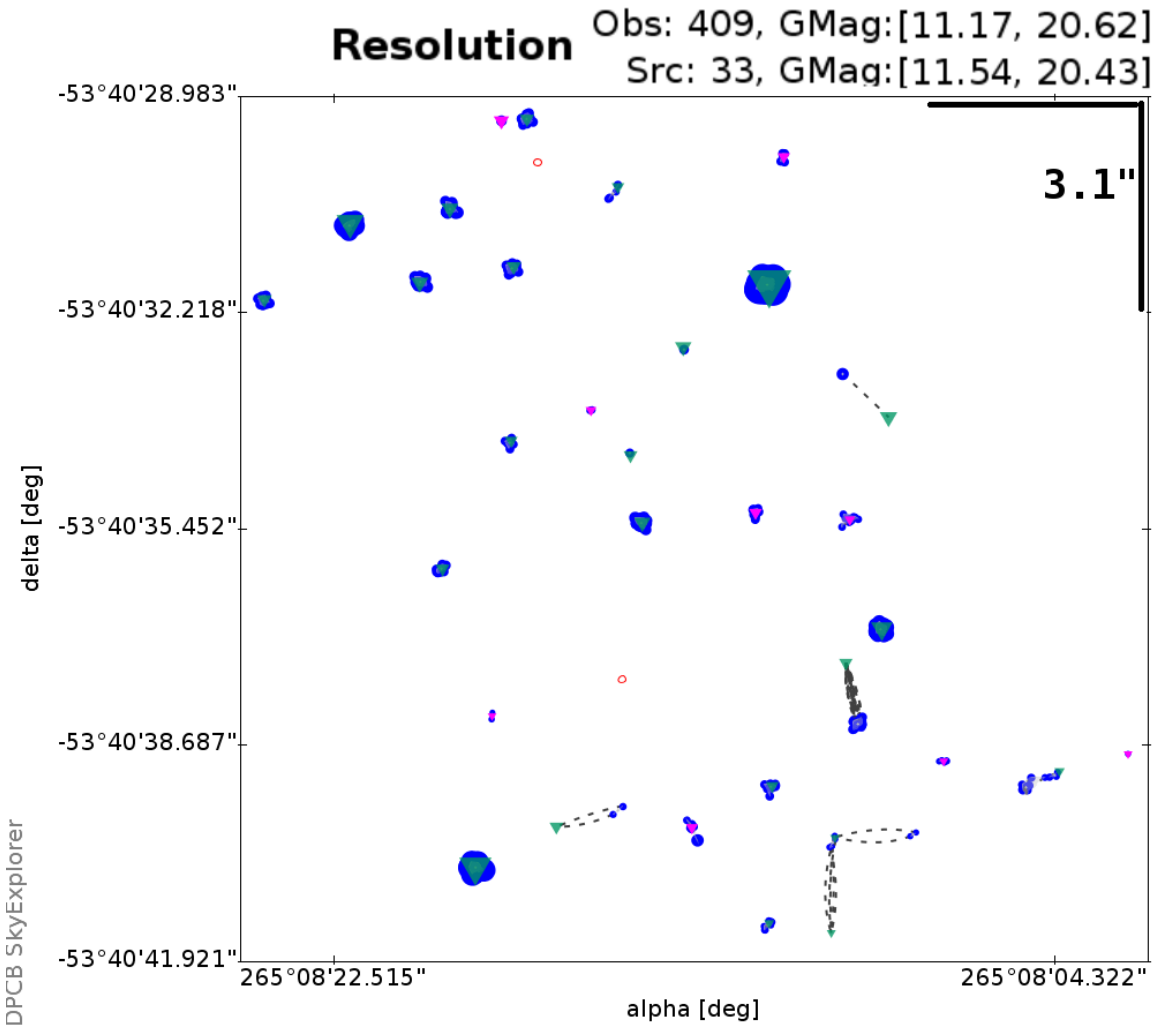
Observation DR2 source New source by merge

Conflict resolution example



Observation DR2 source

Conflict resolution example



DR3 resolution:

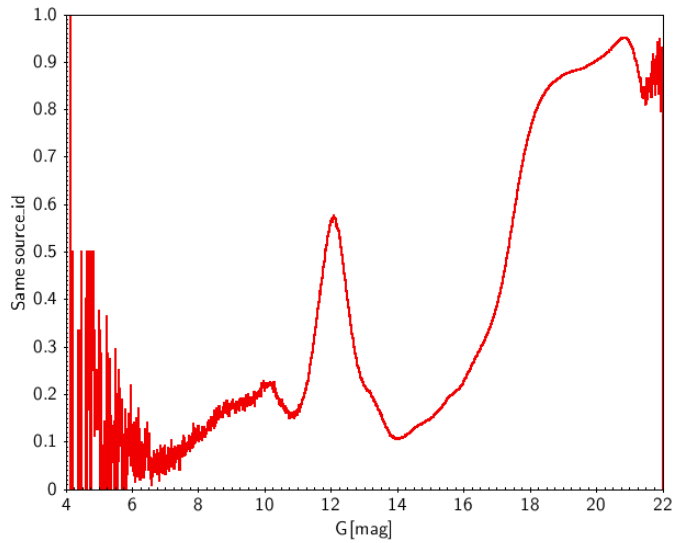
- 24 of 28 persisting sources.
- 8 new sources from scratch.
- 1 new source by merge.
- 2 sources w/o matches.

Observation DR2 source New source Blacklisted obs.

Source evolution

Source list changes between DR2 and DR3 are expected to be much less (respect to DR1 and DR2) in terms of deleted/superseded sources:

Fraction of DR1 sourceId also found in DR2



Ref: GDR2 Doc., Fig.10.5

Fraction of DR2 sourceId that will persist in DR3

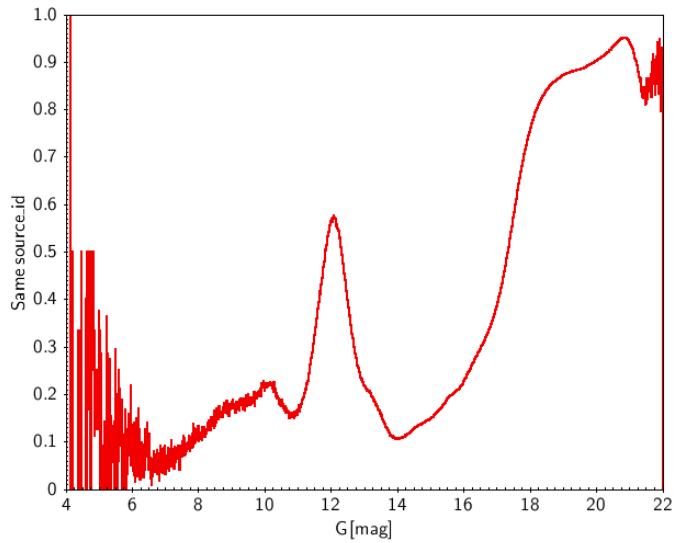
Magnitude	Persisting
$G < 10$	$\gtrsim 94\%$
$G < 12$	$\gtrsim 94\%$
$G < 14$	$\gtrsim 97\%$
Total	$\gtrsim 97\%$

Several improvements have been implemented in the clustering solution and consequently the sources parameters may show major updates: **do not rely on a direct sourceIds match**.

Source evolution

Source list changes between DR2 and DR3 are expected to be much less (respect to DR1 and DR2) in terms of deleted/superseded sources:

Fraction of DR1 sourceId also found in DR2



Ref: GDR2 Doc., Fig.10.5

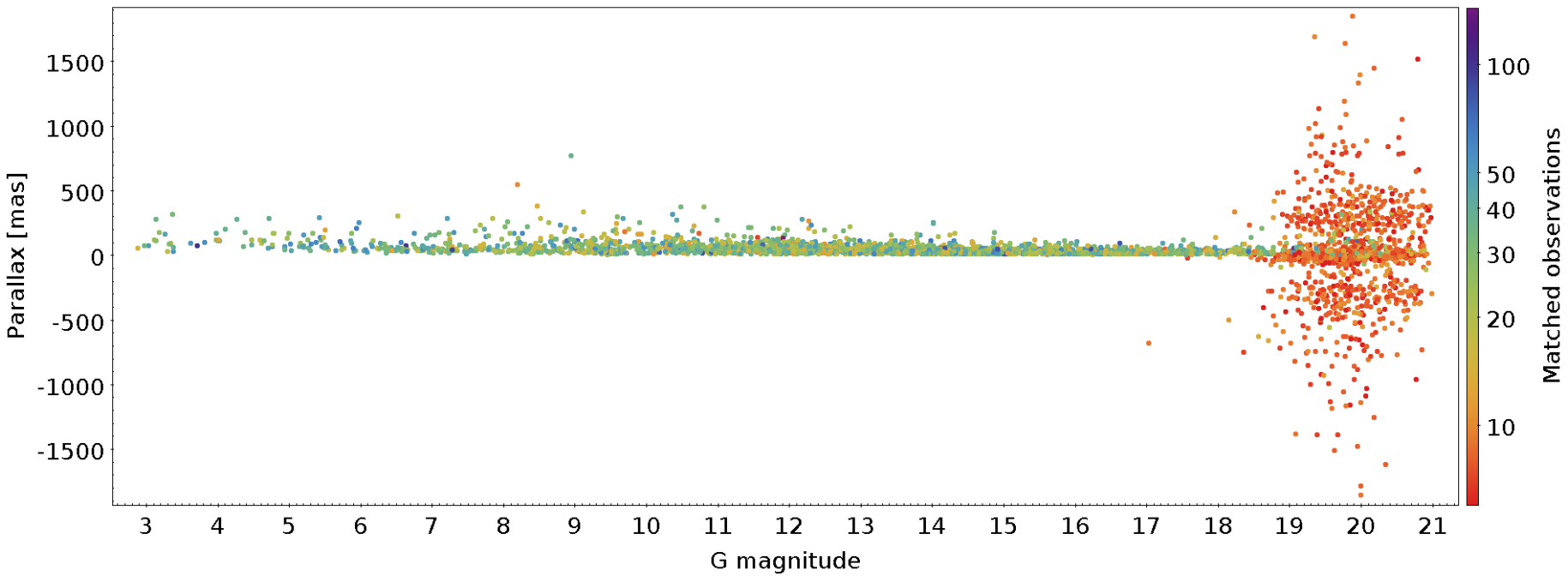
Fraction of DR2 sourceId that will persist in DR3

Magnitude	Persisting
$G < 10$	$\gtrsim 94\%$
$G < 12$	$\gtrsim 94\%$
$G < 14$	$\gtrsim 97\%$
Total	$\gtrsim 97\%$

Several improvements have been implemented in the clustering solution and consequently the sources parameters may show major updates: **do not rely on a direct sourceIds match**.

HPM sources in DR2

- Selection of HPM sources: $\mu > 600$ mas $\cdot$ yr^{-1}
- DR2 includes a large number of fainter HPM sources with weird parallaxes and a few number of matched observations: 14% of HPM sources with a negative parallax.



HPM sources in DR3

- Most of these HPM sources with weird parallaxes will not persist in DR3. No HPM source with a negative parallax.
- $\sim 11\%$ of HPM sources in DR3 are new sources by merge.
- The number of HPM sources with $G < 17$ will be incremented.
- The total number of HPM sources might be reduced but the quality will be significantly improved.

Thank you!

Ferran Torra Clotet

Institut de Ciències del Cosmos - Universitat de Barcelona

DAPCOM

ftorra@fqa.ub.edu