# Who is in and who is not?
# Determining the *Gaia* survey selection function



Sterrewacht Leiden
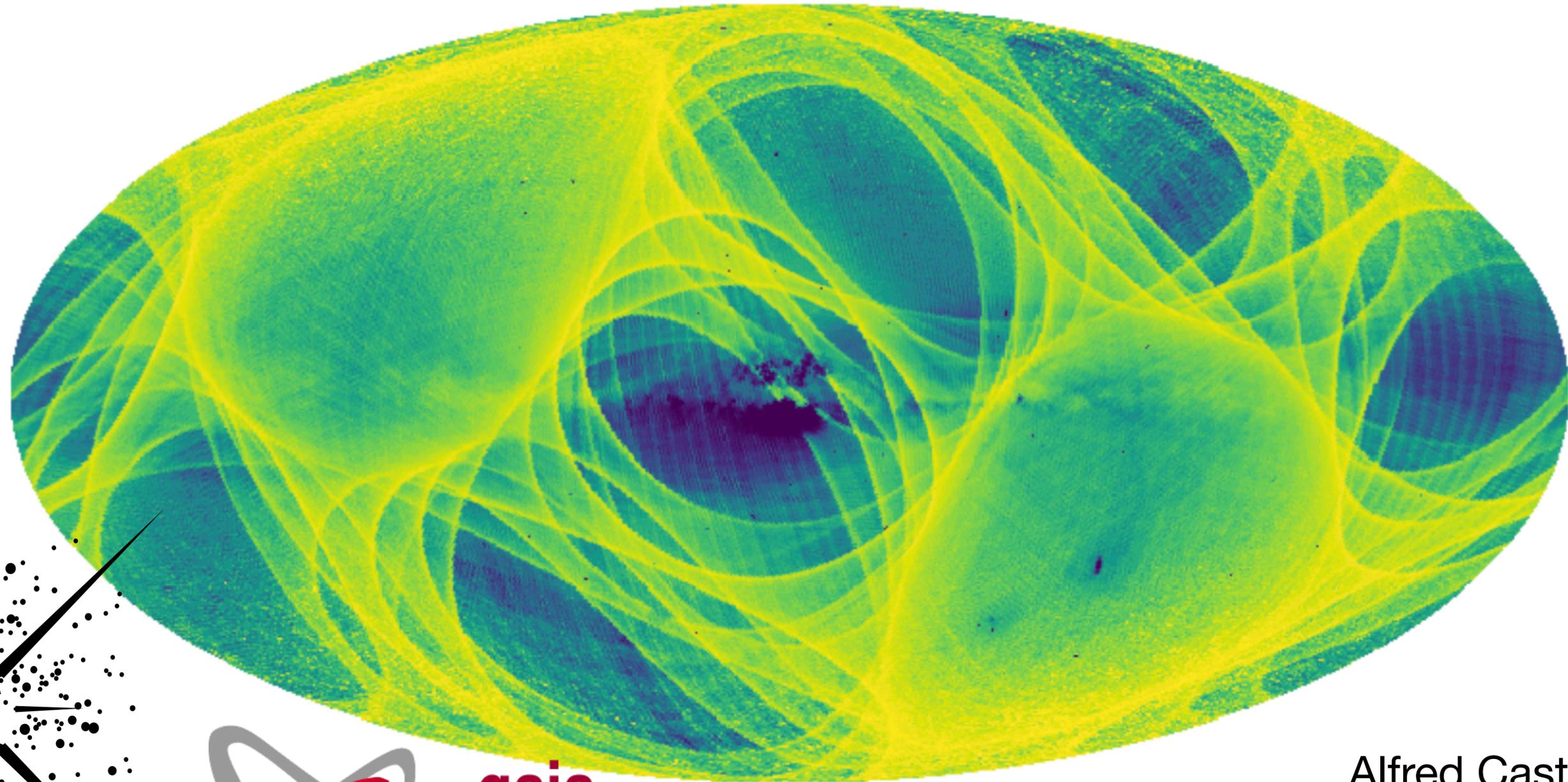
gaia unlimited

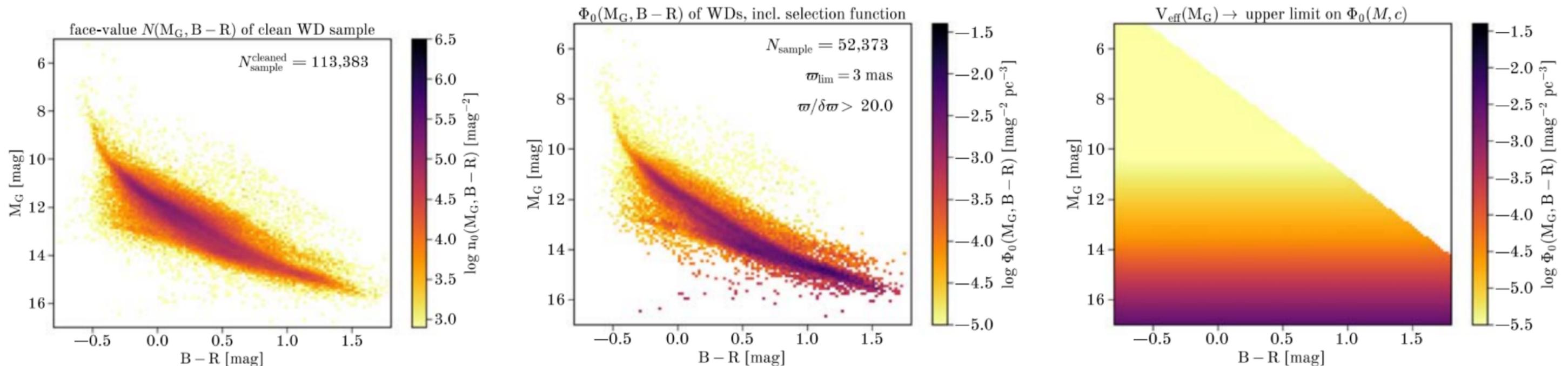Alfred Castro Ginard
Barcelona
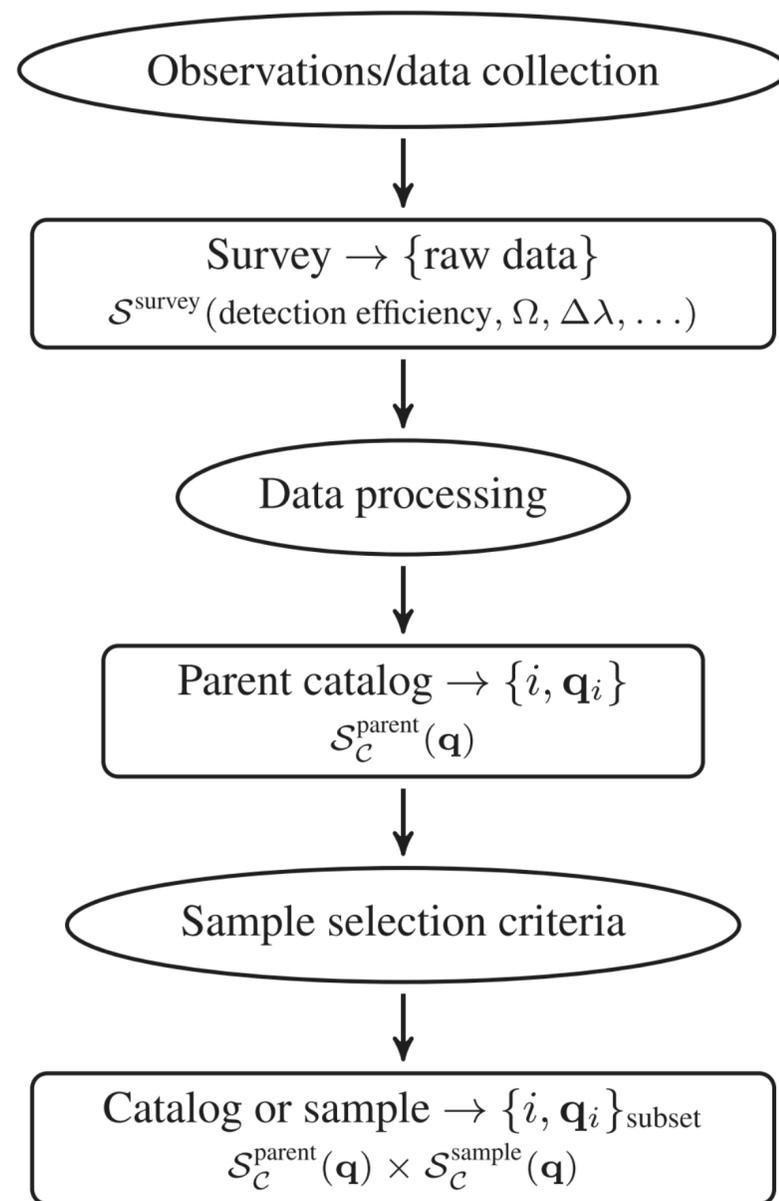7 Sept. 2023
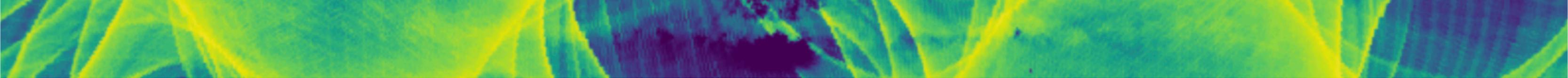
# GaiaUnlimited: What stars are we missing?

To reach reliable scientific conclusions when working with Gaia data, we need to account for the processes in which stars make it or not into the Gaia catalogue.

Understand under what circumstances an object is included into a catalogue, or a subsample drawn from it.

- Space density of white dwarfs as a function of magnitude and colour [Rix+21]

Observations/data collection

Survey $\rightarrow$ {raw data}
$\mathcal{S}^{\text{survey}}(\text{detection efficiency}, \Omega, \Delta\lambda, \ldots)$

Data processing

Parent catalog $\rightarrow \{i, \mathbf{q}_i\}$
$\mathcal{S}_\mathcal{C}^{\text{parent}}(\mathbf{q})$

Sample selection criteria

Catalog or sample $\rightarrow \{i, \mathbf{q}_i\}_{\text{subset}}$
$\mathcal{S}_\mathcal{C}^{\text{parent}}(\mathbf{q}) \times \mathcal{S}_\mathcal{C}^{\text{sample}}(\mathbf{q})$

$S_\mathcal{C}(\boldsymbol{q}_i)$: probability that a source $i$ with attributes $\boldsymbol{q}_i$ is contained in catalogue or sample $\mathcal{C}$
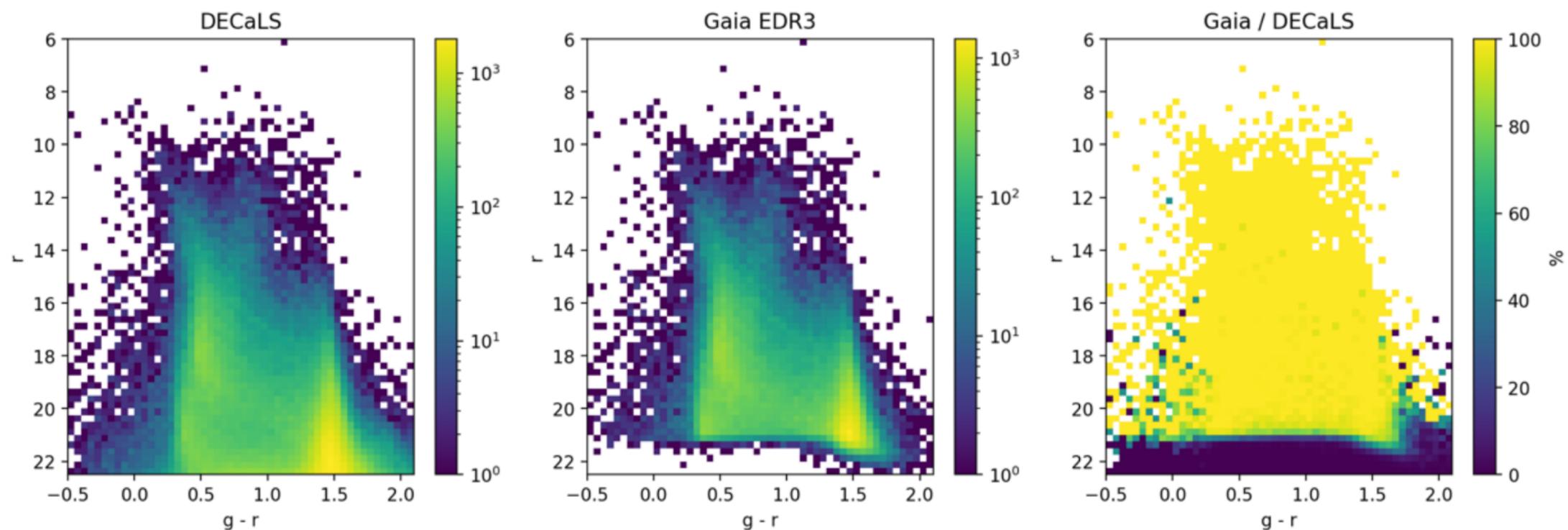
What we aim to provide:

- Selection function of **parent *Gaia* catalogue**
  - All sources with $\boldsymbol{q}_i : \{\alpha_i, \delta_i, G_i, \ldots\}$
- Selection function for **subsets**
  - 5-parameter astrometry, full photometry, radial velocity, RUWE $< x$, …
- Selection functions for combinations of *Gaia* and other surveys
  - Photometric and spectroscopic surveys
- Examples of specific *Gaia* selection functions
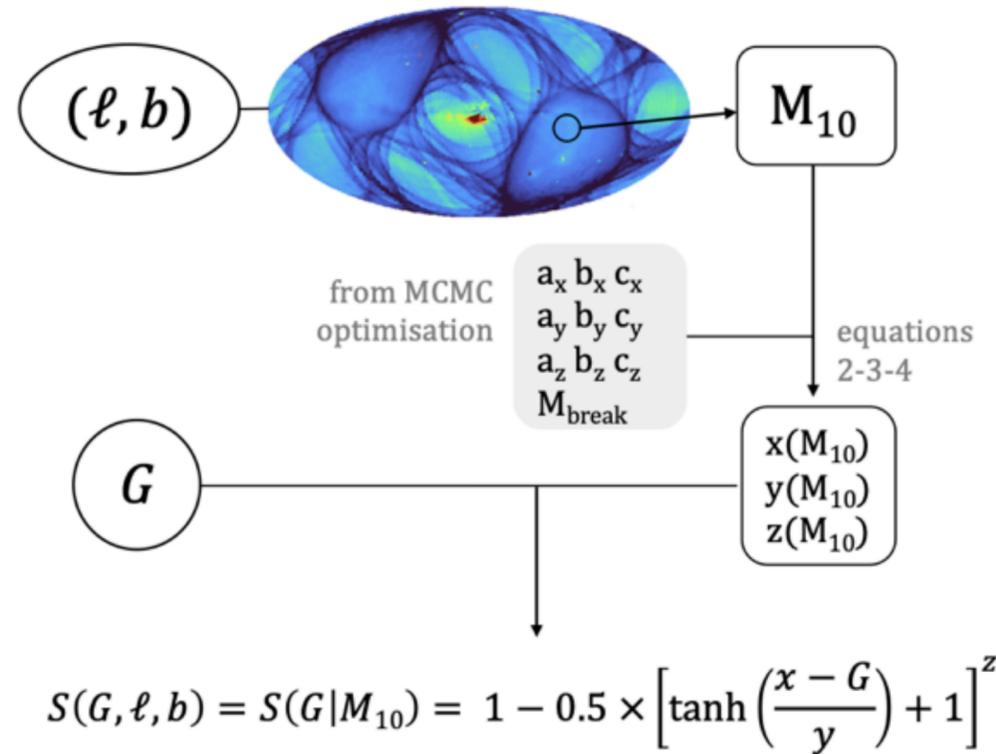  - **Binaries**, Cepheid variables, …

# *Gaia* catalogue selection function

Currently developing two approximations to estimate the *Gaia* Selection Function:

- Forward modelling (based on [Boubert+20,21])

  - From *Gaia* data only: Scanning law, ability to convert observations to detections, constraints in the used detections to get measurements, …

- Empirical model [Cantat-Gaudin+22]

  - Comparing to deeper photometric catalogues (here, DECaLS), considered as 'ground truth'

# *Gaia* Selection Function: empirical model



$$S(G, \ell, b) = S(G|M_{10}) = 1 - 0.5 \times \left[\tanh\left(\frac{x-G}{y}\right) + 1\right]^z$$
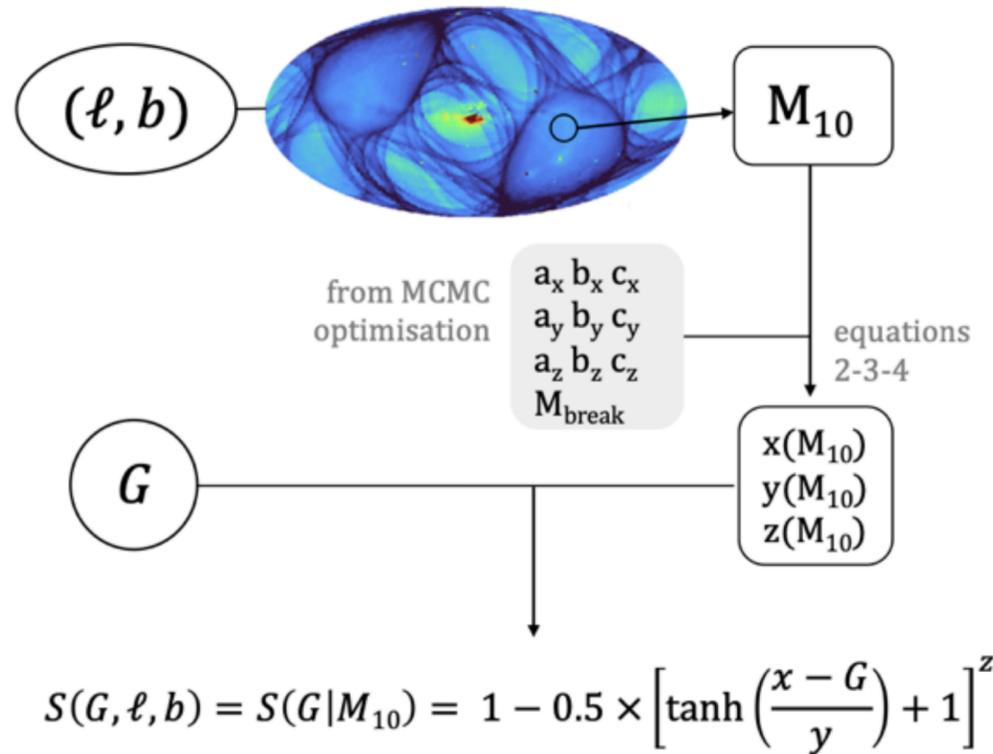
- Query $M_{10}$ as a function of sky position (healpix)
- Estimate completeness as a function of G from $M_{10}$

- Installable python package to query the Gaia Selection Function from a pre-computed $M_{10}$ map at HEALPix level 7
  - Github: https://github.com/gaia-unlimited/gaiaunlimited
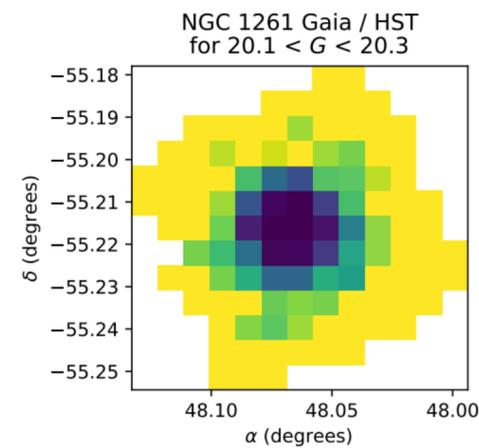  - PyPi: `pip install gaiaunlimited`
  - Documentation: https://gaiaunlimited.readthedocs.io/en/latest/

# *Gaia* Selection Function: empirical model



$$S(G, \ell, b) = S(G|M_{10}) = 1 - 0.5 \times \left[\tanh\left(\frac{x - G}{y}\right) + 1\right]^z$$

- Query $M_{10}$ as a function of sky position (healpix)
- Estimate completeness as a function of G from $M_{10}$
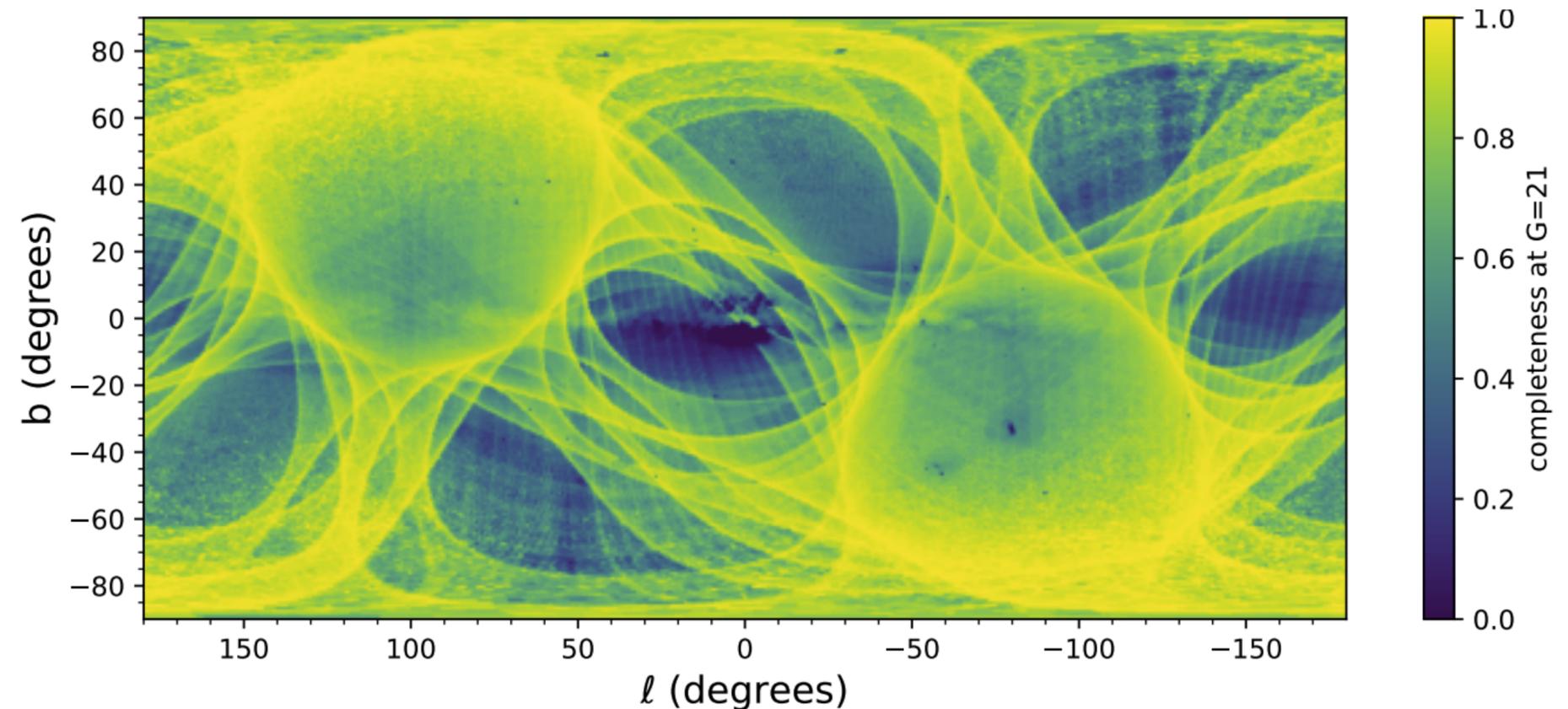
- Installable python package to query the Gaia Selection Function from a pre-computed $M_{10}$ map at HEALPix level 7
  - Github: https://github.com/gaia-unlimited/gaiaunlimited
  - PyPi: `pip install gaiaunlimited`
  - Documentation: https://gaiaunlimited.readthedocs.io/en/latest/

# *Gaia* Selection Function: subsamples

Studies on *Gaia* data often use only certain kind of objects (white dwarfs, red clump stars, stars with radial velocities…) or apply cuts on the data (based on colour, quality flags). All these cuts introduce additional selection effects.
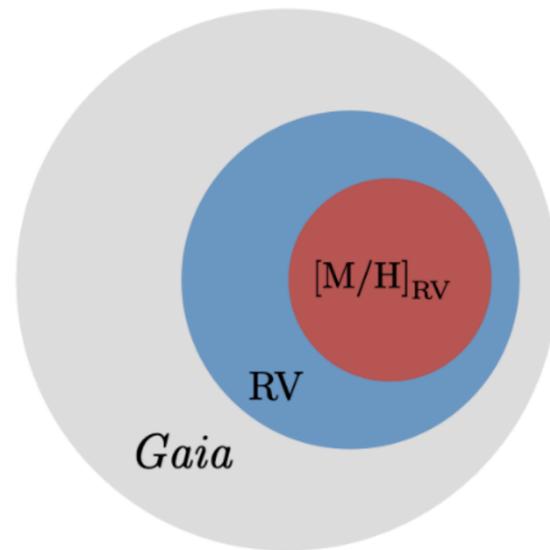
Estimate the probability of a star being in the subsample (fully Bayesian method):
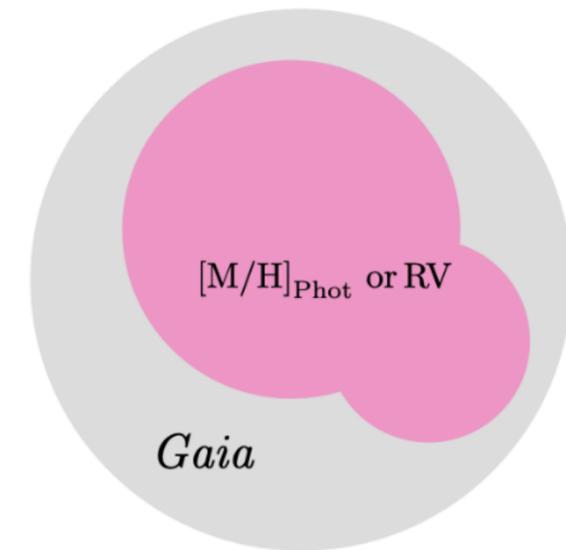
$$\mathrm{E}(p) = \frac{k+1}{n+2}$$

$$\mathrm{var}(p) = \frac{(k+1)(n-k+1)}{(n+2)^2(n+3)}$$

k: stars in sample

n: stars in Gaia



$\mathcal{S}([\mathrm{M/H}]_{\mathrm{RV}}) = \mathcal{S}(Gaia) \cdot \mathcal{S}(\mathrm{RV}|Gaia) \cdot \mathcal{S}([\mathrm{M/H}]_{\mathrm{RV}}|\mathrm{RV}, Gaia)$

$\mathcal{S}([\mathrm{M/H}]_{\mathrm{Phot}} \text{ or } \mathrm{RV}) = \mathcal{S}(Gaia) \cdot \mathcal{S}([\mathrm{M/H}]_{\mathrm{Phot}} \text{ or } \mathrm{RV}|Gaia)$

[Castro-Ginard+23]

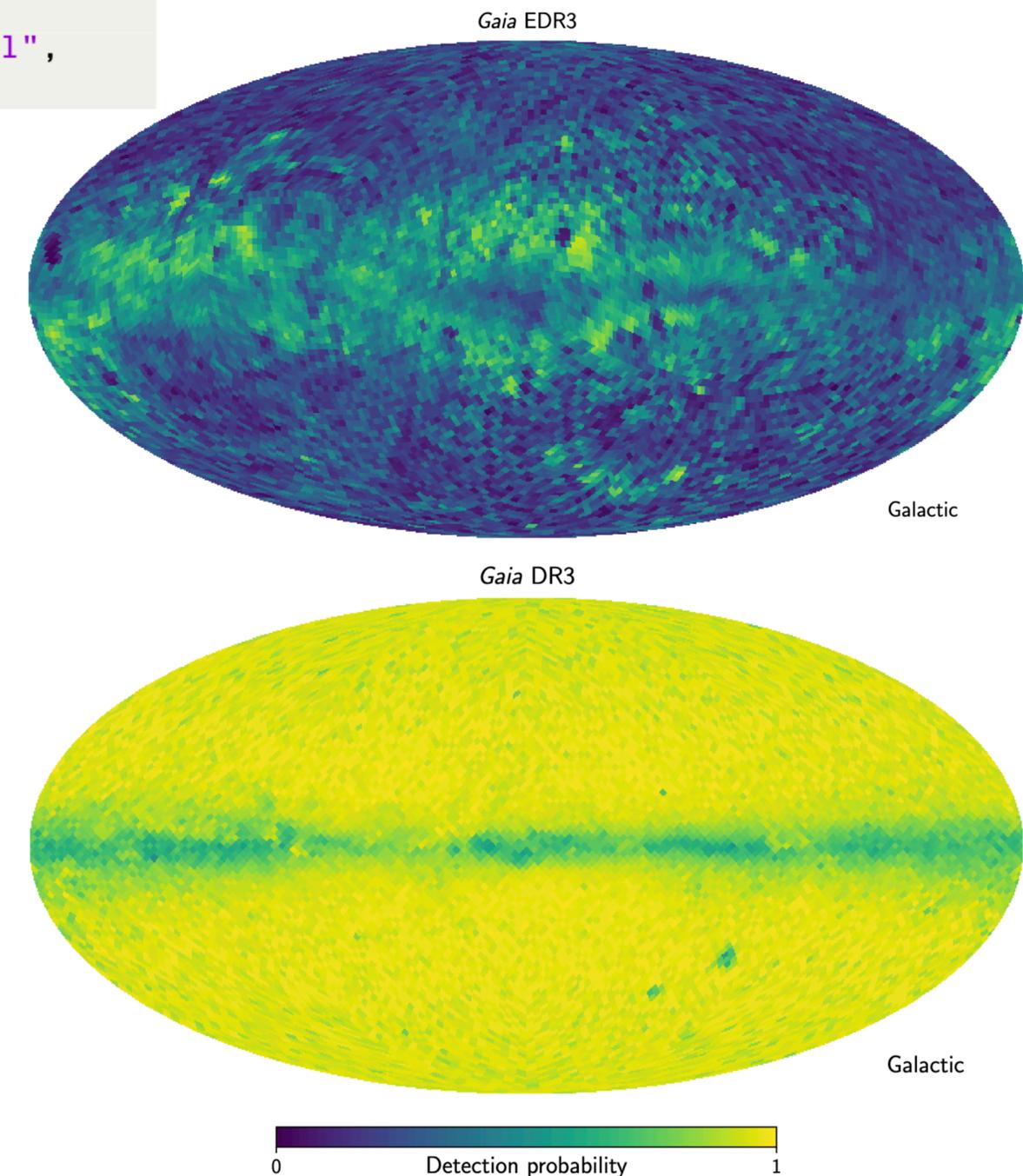# *Gaia* Selection Function: RV sample

```
10  #Define the dependencies and resolutions of the selection function
11  inDict = {'healpix': 5, 'phot_g_mean_mag': [3,20,0.2], 'g_rp': [-2.5,5.1,0.4]}
12
13  #Initiate the SubsampleSelectionFunction class
14  dr3SubsampleSF = SubsampleSelectionFunction(subsample_query = "radial_velocity is not null",
       file_name = "radial_velocity", hplevel_and_binning = inDict)
```

Compute subsample Selection Function considering:
- It is a function of sky position, magnitude and colour
- Pixels in this three-dimensional space are independent (no smoothing/correlation)
- The source counts follow a Beta-Binomial distribution function
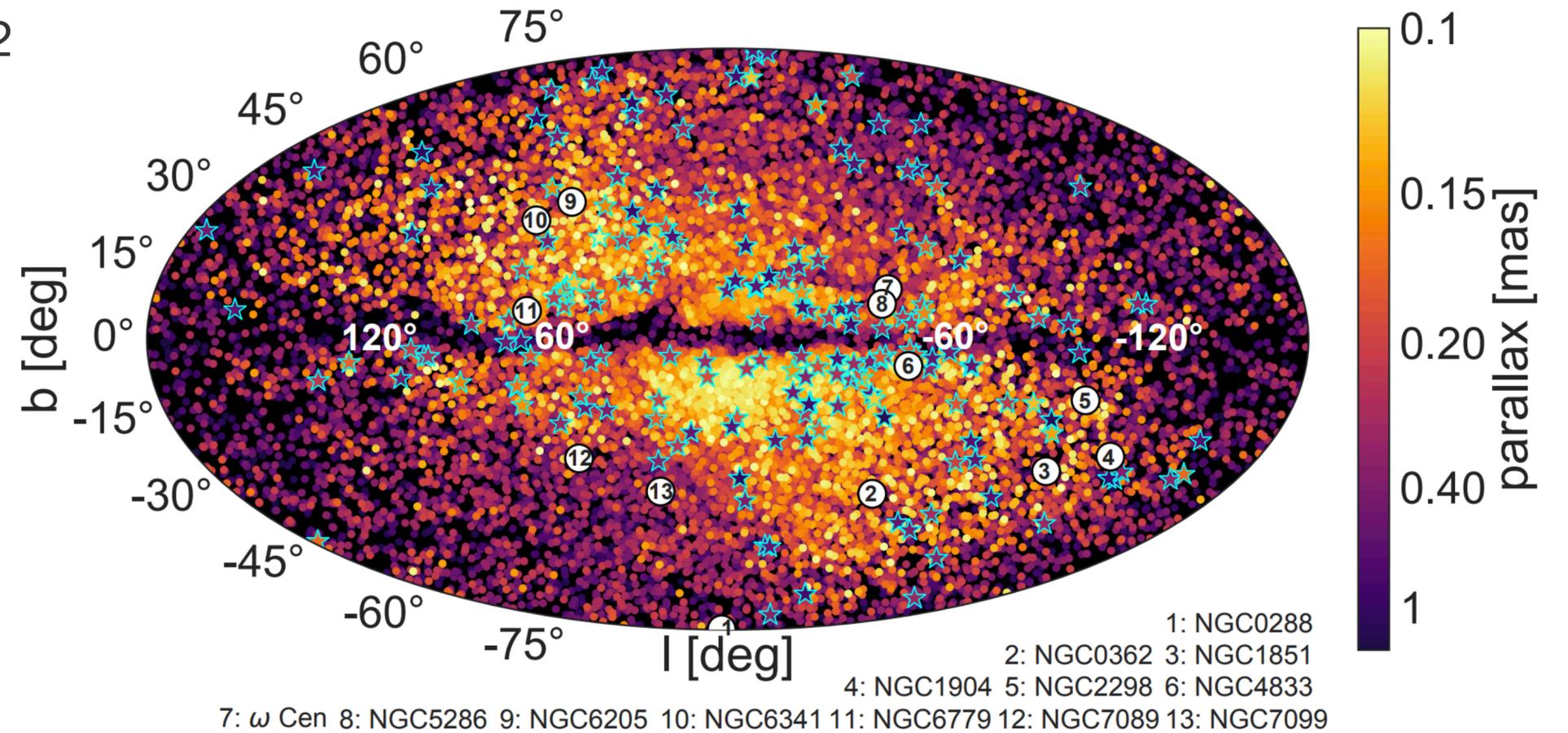
Caveats:
- Bias in the Beta-Binomial estimation
- Because we bin and count:
  - No estimates where there are no sources
  - Noisy estimates where there are few sources

*Gaia* EDR3

Galactic

*Gaia* DR3

Galactic

0    Detection probability    1

# *Gaia* Selection Function: GES sample

Helmi et al. (2018) used the 7 million

sources with RV measurements in *Gaia* DR2
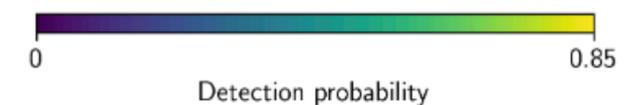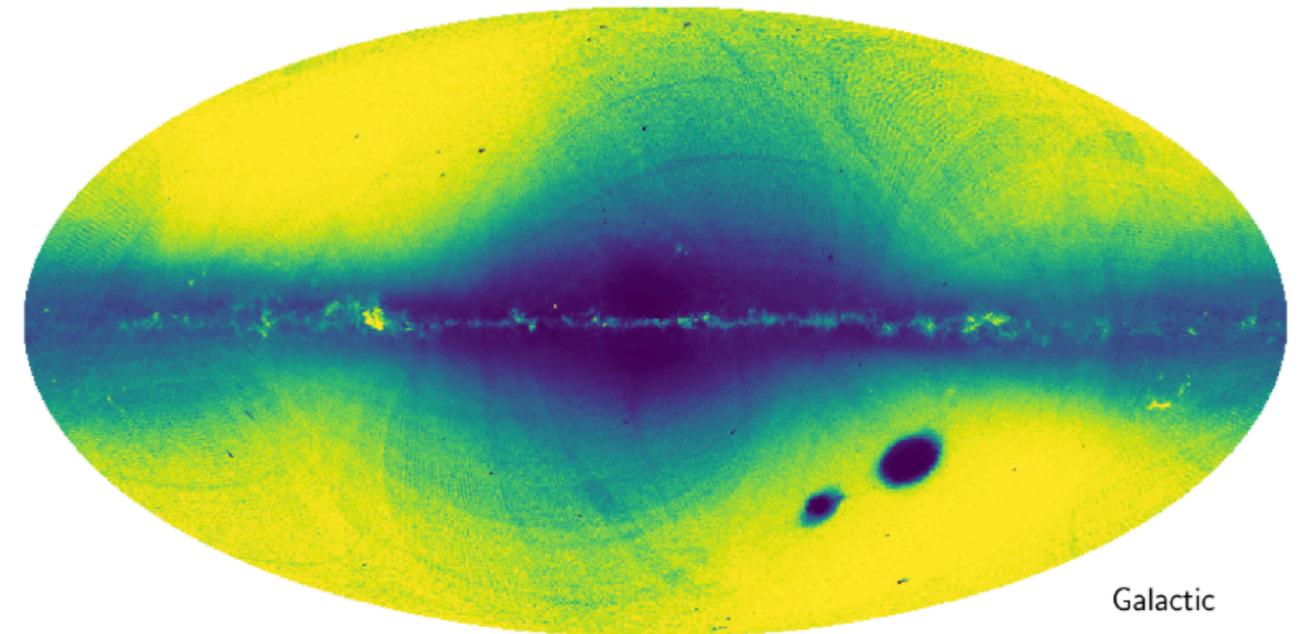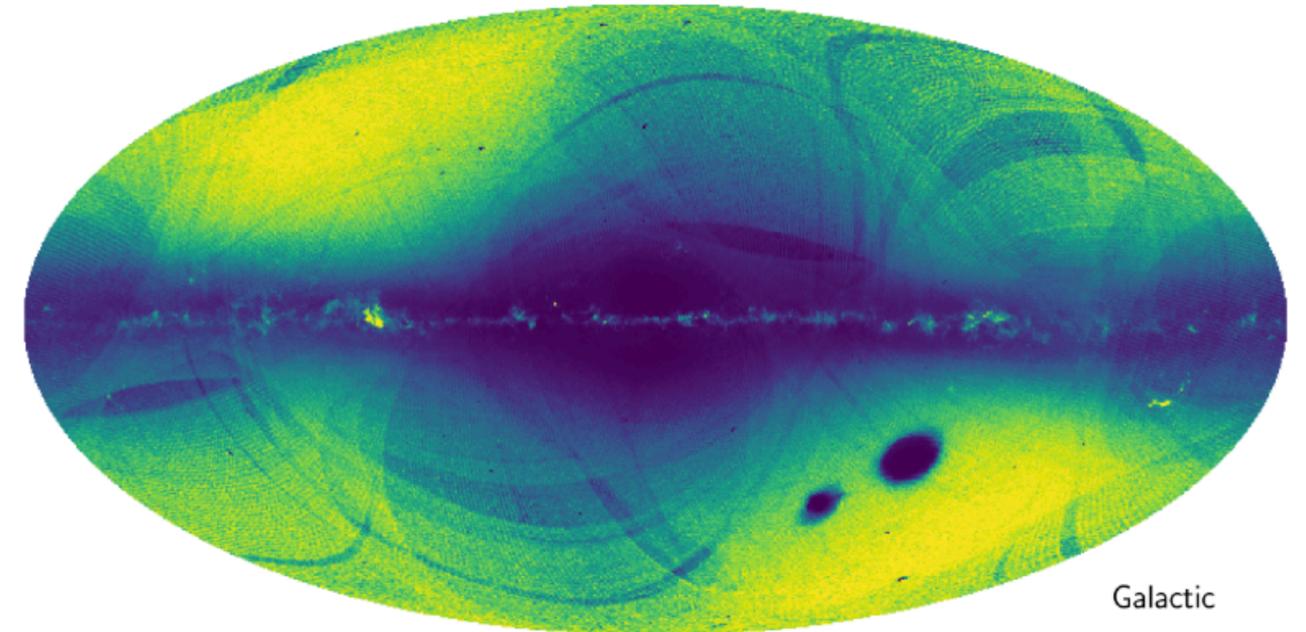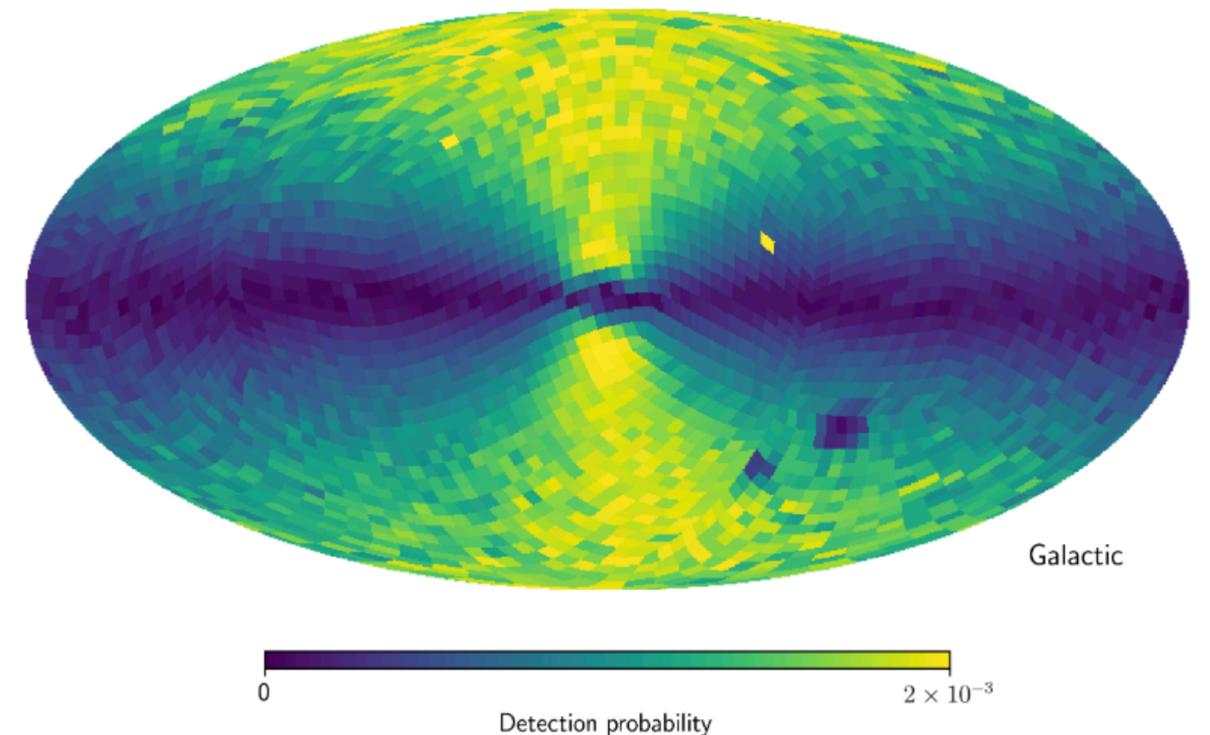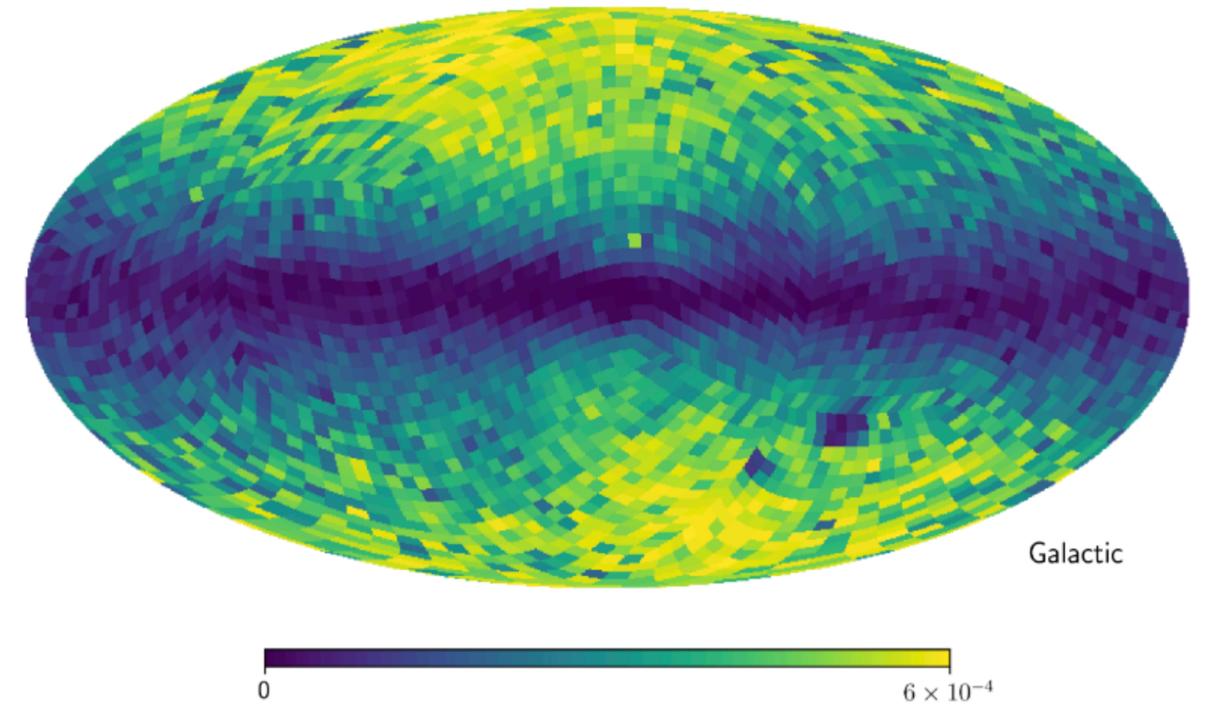
to report the discovery of *Gaia*-Enceladus

# *Gaia* Selection Function: GES sample

Helmi et al. (2018) used the 7 million

sources with RV measurements in *Gaia* DR2

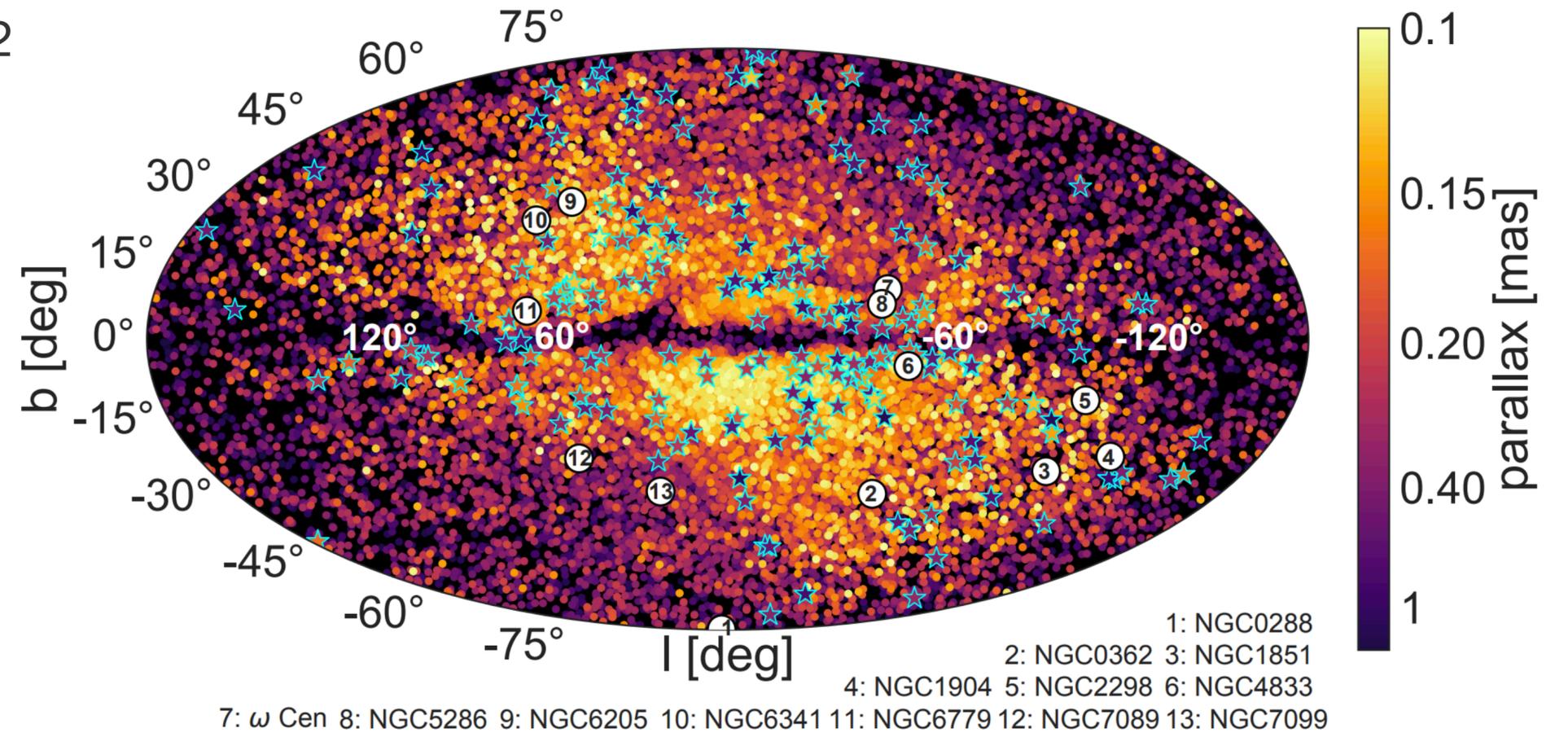to report the discovery of *Gaia*-Enceladus

Cuts on:

- Stars with RV
- $\varpi > 0.1$ mas
- $\varpi/\sigma_\varpi > 5$ mas



Galactic

Galactic

0                                         0.85
Detection probability

# *Gaia* Selection Function: GES sample

Helmi et al. (2018) used the 7 million

sources with RV measurements in *Gaia* DR2

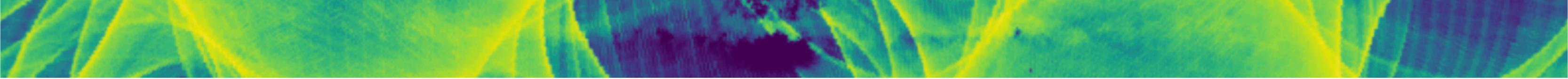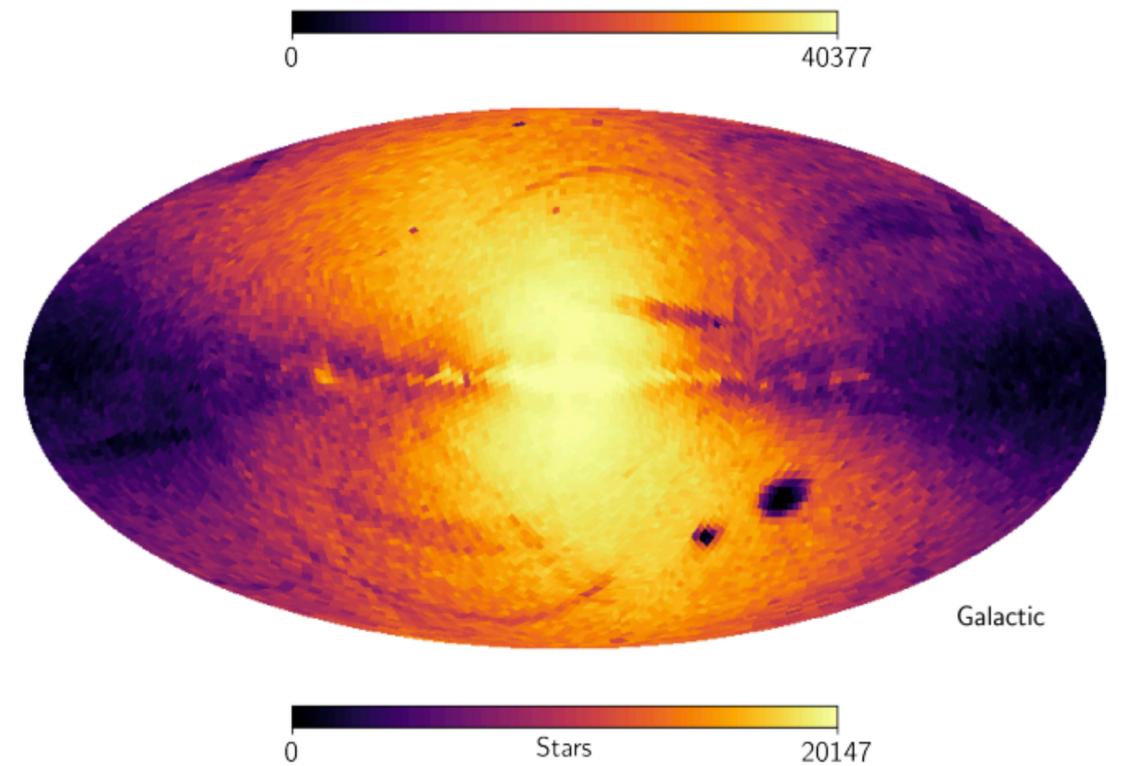to report the discovery of *Gaia*-Enceladus

Cuts on:

- Stars with RV

- $\varpi > 0.1$ mas

- $\varpi/\sigma_\varpi > 5$ mas

- $-1\,500 < L_z < 150$ kpc km/s

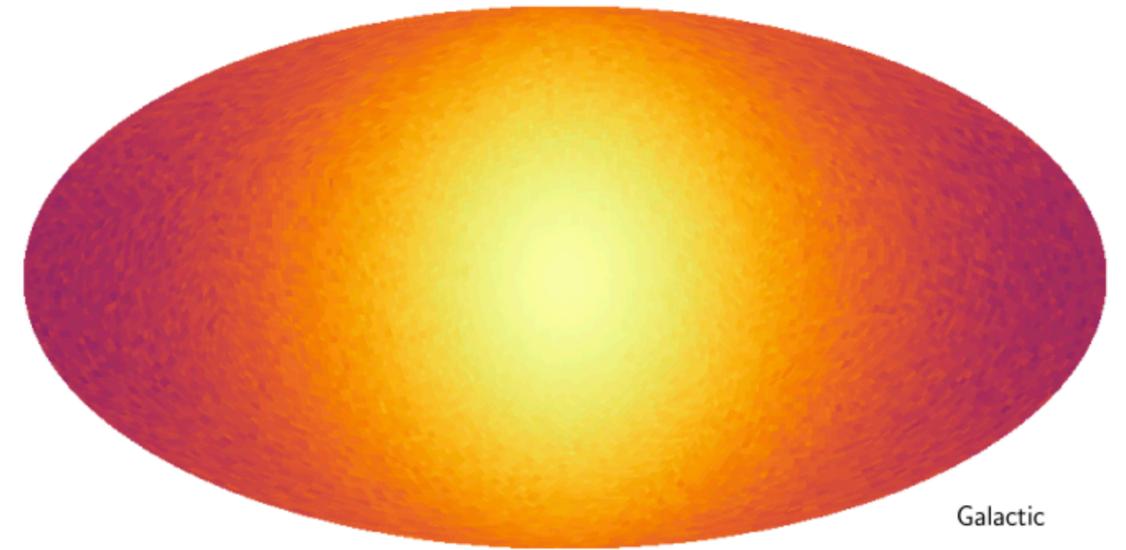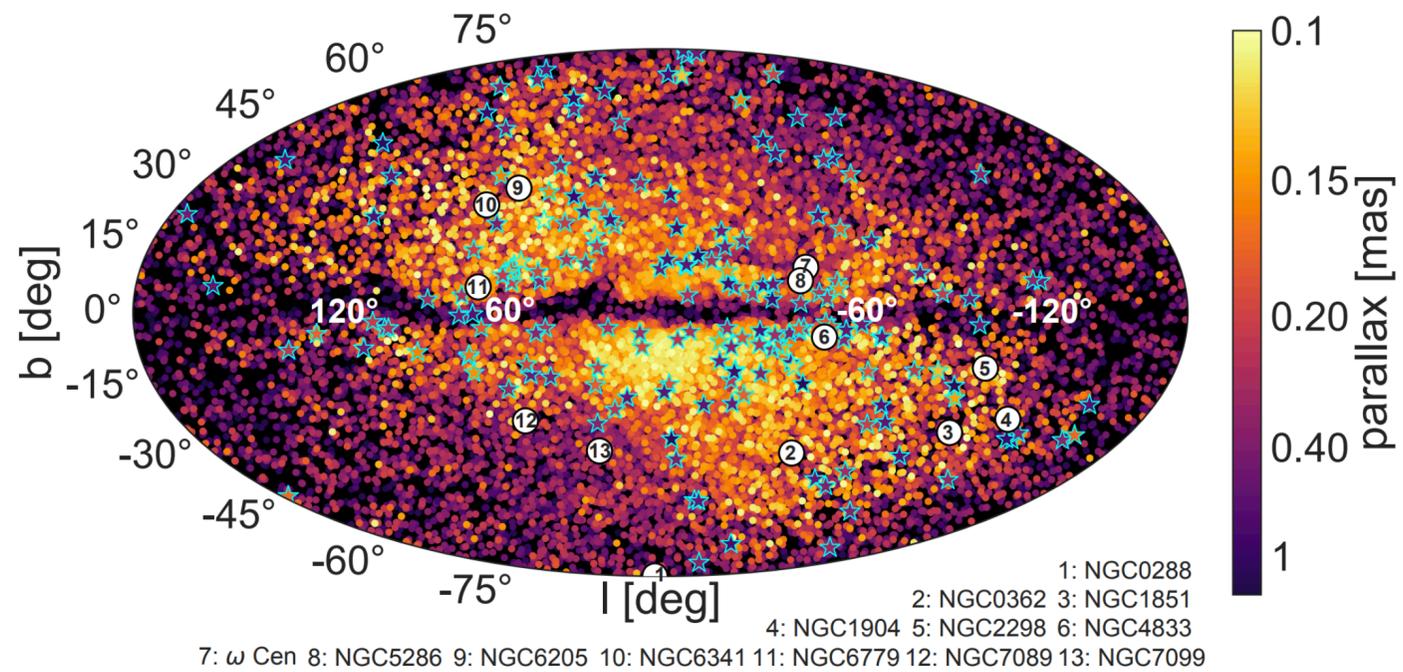# *Gaia* Selection Function: GES sample

Helmi et al. (2018) used the 7 million

sources with RV measurements in *Gaia* DR2

to report the discovery of *Gaia*-Enceladus

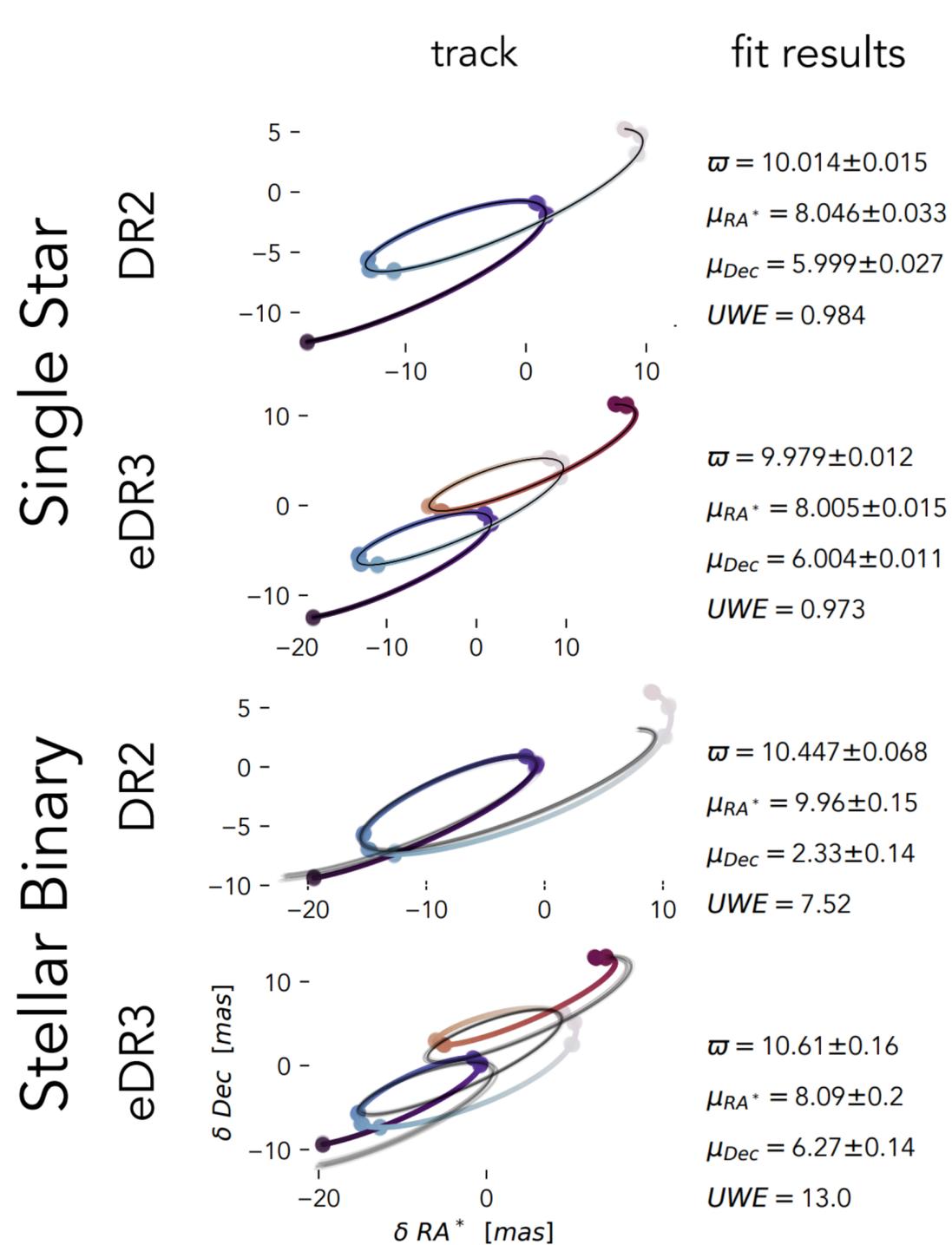# *Gaia* Selection Function: GES sample

Helmi et al. (2018) used the 7 million

sources with RV measurements in *Gaia* DR2

to report the discovery of *Gaia*-Enceladus

# *Gaia* Selection Function: binary systems



track          fit results

**Single Star**

DR2
$\varpi = 10.014 \pm 0.015$
$\mu_{RA^*} = 8.046 \pm 0.033$
$\mu_{Dec} = 5.999 \pm 0.027$
$UWE = 0.984$

eDR3
$\varpi = 9.979 \pm 0.012$
$\mu_{RA^*} = 8.005 \pm 0.015$
$\mu_{Dec} = 6.004 \pm 0.011$
$UWE = 0.973$

**Stellar Binary**

DR2
$\varpi = 10.447 \pm 0.068$
$\mu_{RA^*} = 9.96 \pm 0.15$
$\mu_{Dec} = 2.33 \pm 0.14$
$UWE = 7.52$

eDR3
$\varpi = 10.61 \pm 0.16$
$\mu_{RA^*} = 8.09 \pm 0.2$
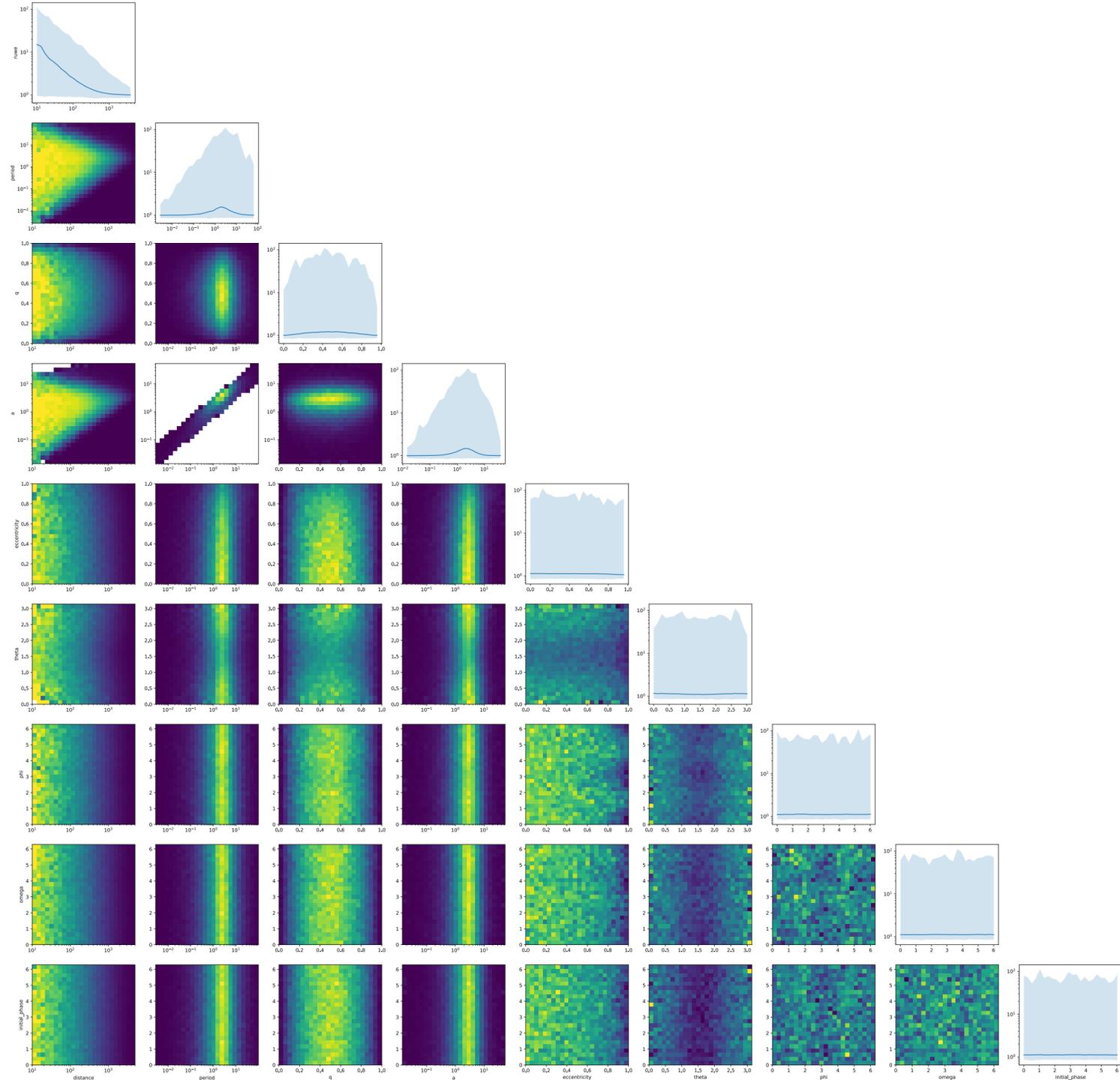$\mu_{Dec} = 6.27 \pm 0.14$
$UWE = 13.0$

[Penoyre+21]

- The presence of an unresolved companion has an effect in the astrometric solution by Gaia [Penoyre+21]
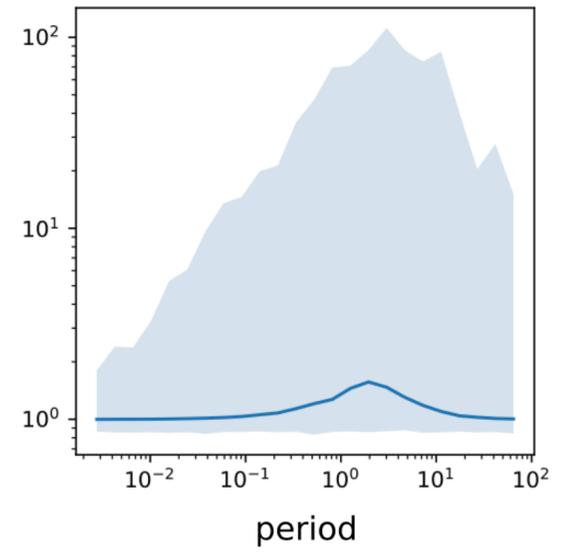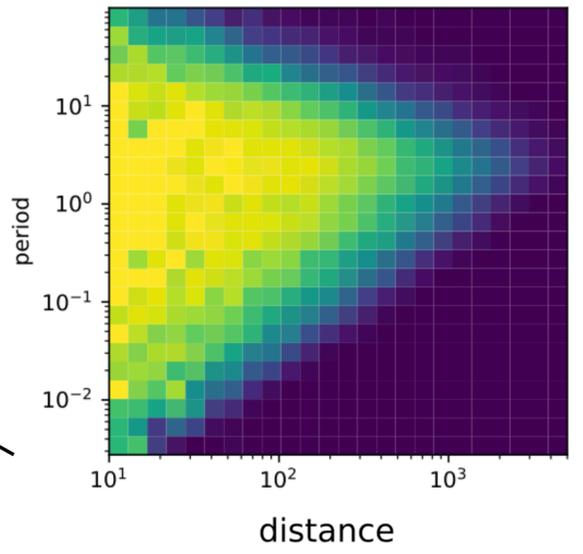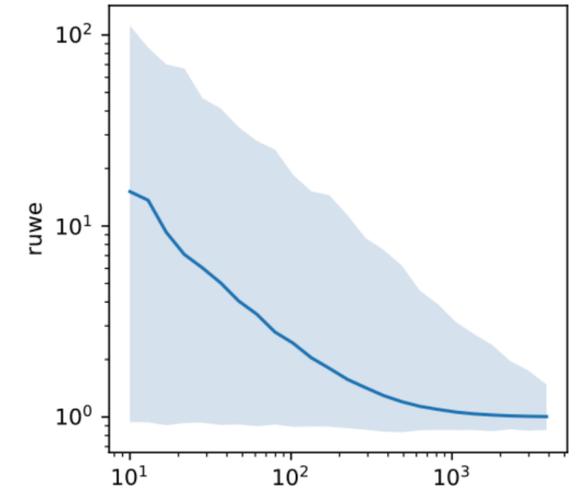- RUWE values higher than 1.25 (1.4) may indicate we are tracking a binary system
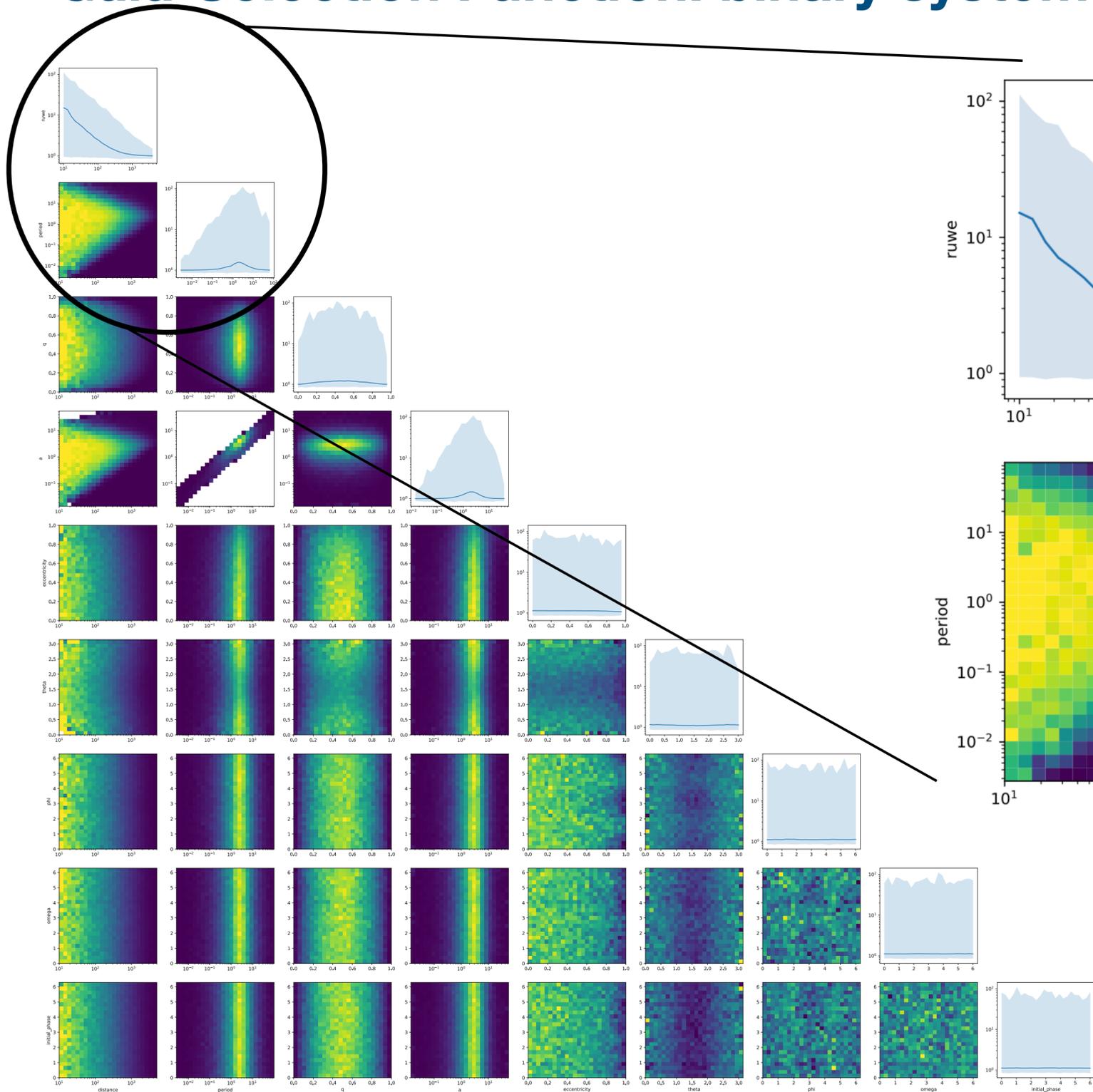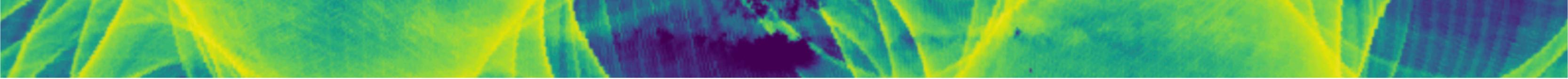


WORK IN PROGRESS
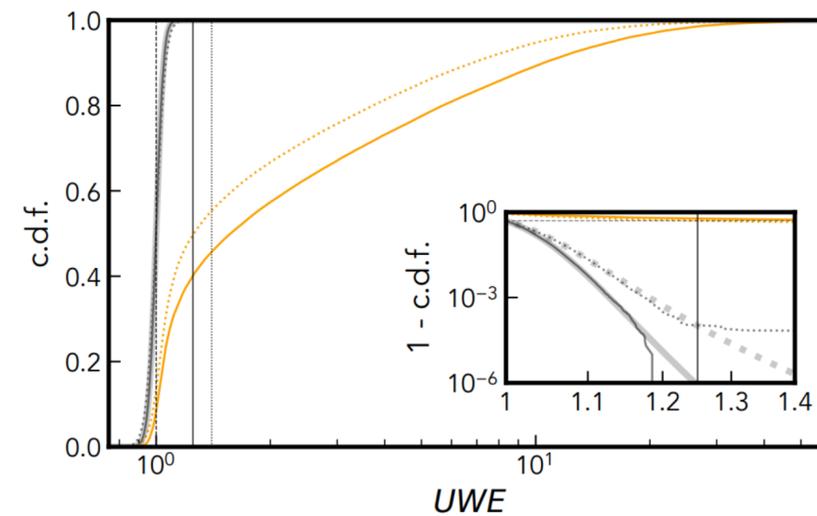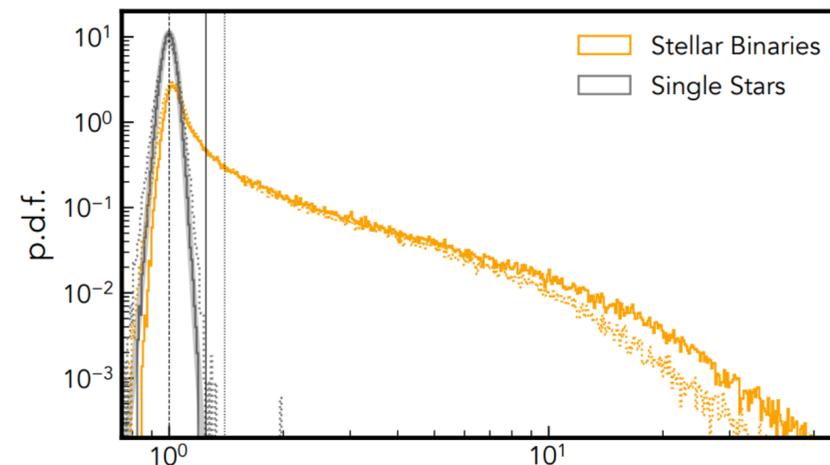
# *Gaia* Selection Function: binary systems

# *Gaia* Selection Function: binary systems

# Summary and conclusions

- Develop a methodology to estimate the selection function of the *Gaia* catalogue and their products
  - Need to correct for sample completeness and extract meaningful scientific conclusions: Asymmetry in the Gaia-Sausage/Enceladus debris is due to selection effects

- Dedicated and easy-to-use python package, with extensive documentation and tutorials

- Particular data products (beyond cuts and filters in the main catalogue) need dedicated solutions to estimate their selection function: binaries, cepheids, clusters, etc.

- GaiaUnlimited workshop in Torino (Italy), 4-6 October 2023: https://gaia-unlimited.github.io/community-workshop2
  - Good chance to suggest functionalities to python package (also now!)