

# Gaia DR3: The Extragalactic content

*Gaia* Collaboration

Coryn Bailer-Jones

Max Planck Institute for Astronomy, Heidelberg

Gaia Symposium, IIA Bengaluru

15 July 2022

# Multiple modules deal with extragalactic sources

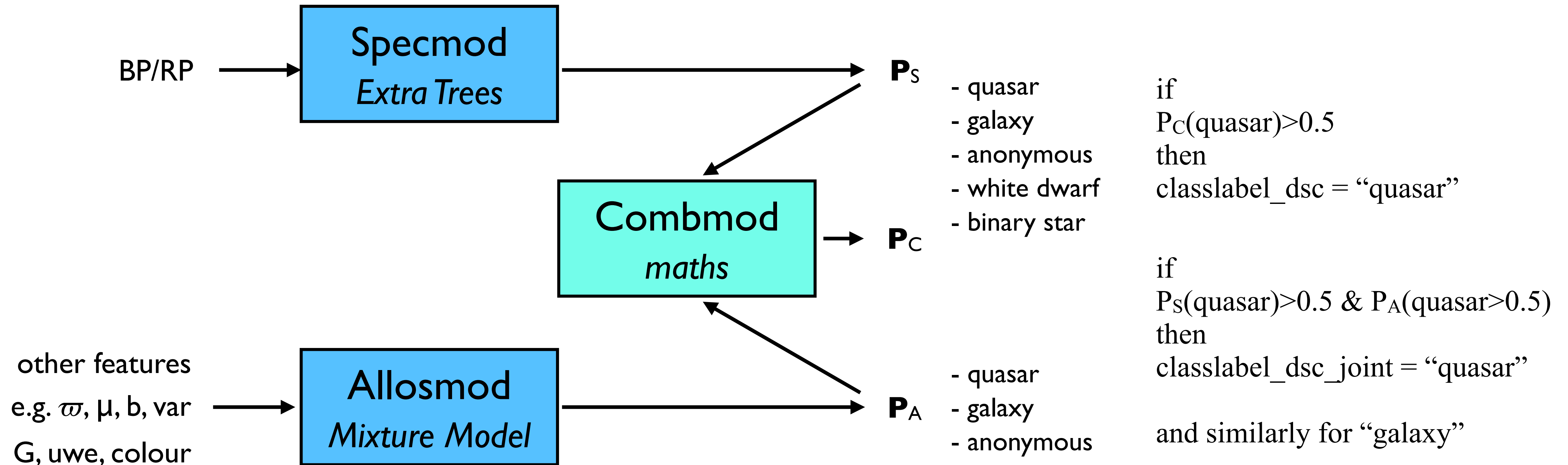
- Classification of Gaia objects
  - ▶ using BP/RP spectrum and astrometry (*Discrete Source Classifier, DSC*)
  - ▶ using photometric light curves (*Vari*)
- Input lists
  - ▶ fit 2D brightness profiles to extended objects (*Extended Objects, EO*)
  - ▶ identify objects from astrometry, similar to Gaia-CRF3
- Redshift estimates
  - ▶ quasars (*QSOC*) and galaxies (*UGC*)
- [Link: overview paper](#)

# The Discrete Source Classifier (DSC)

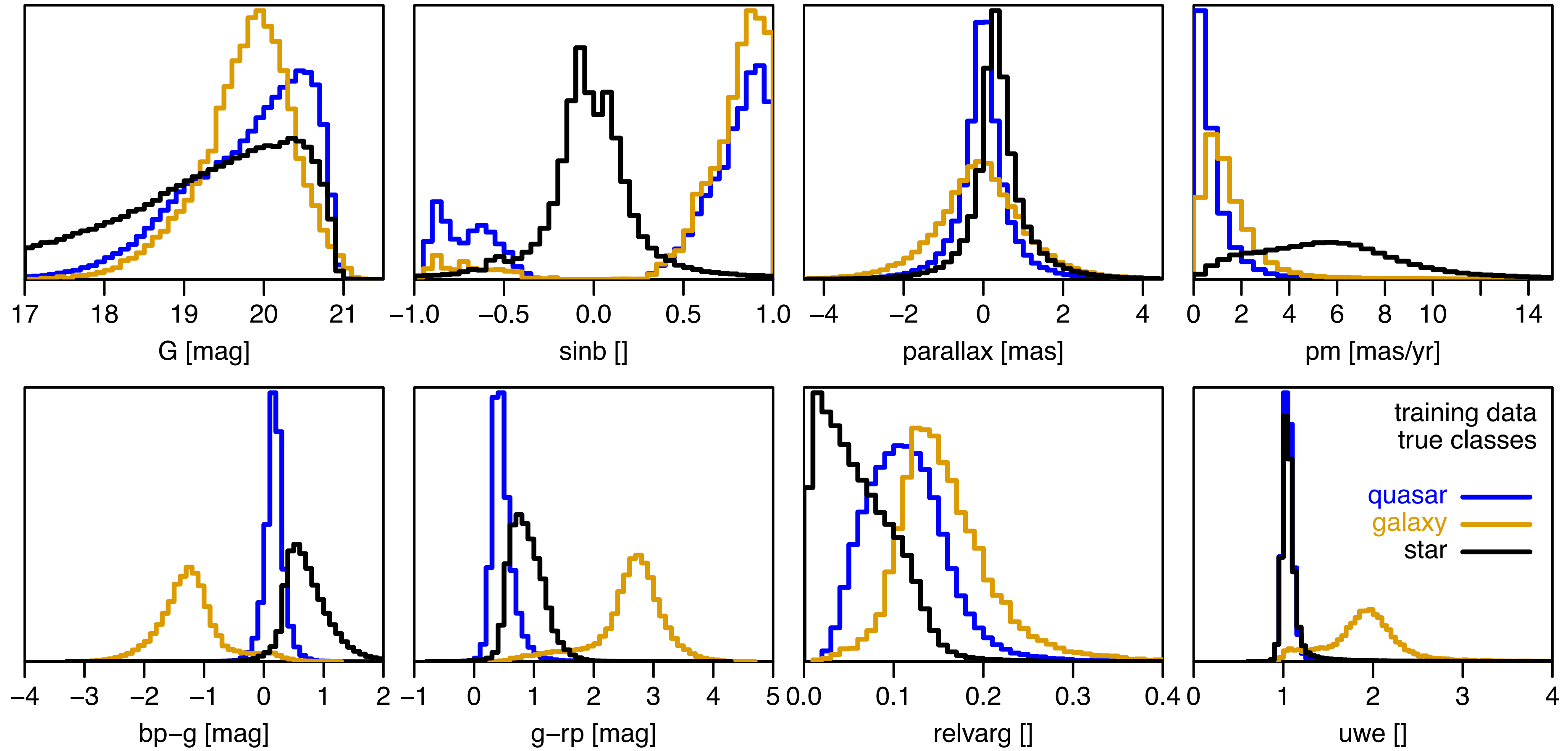
[Link: DSC overview](#)

[Link: DSC details](#)

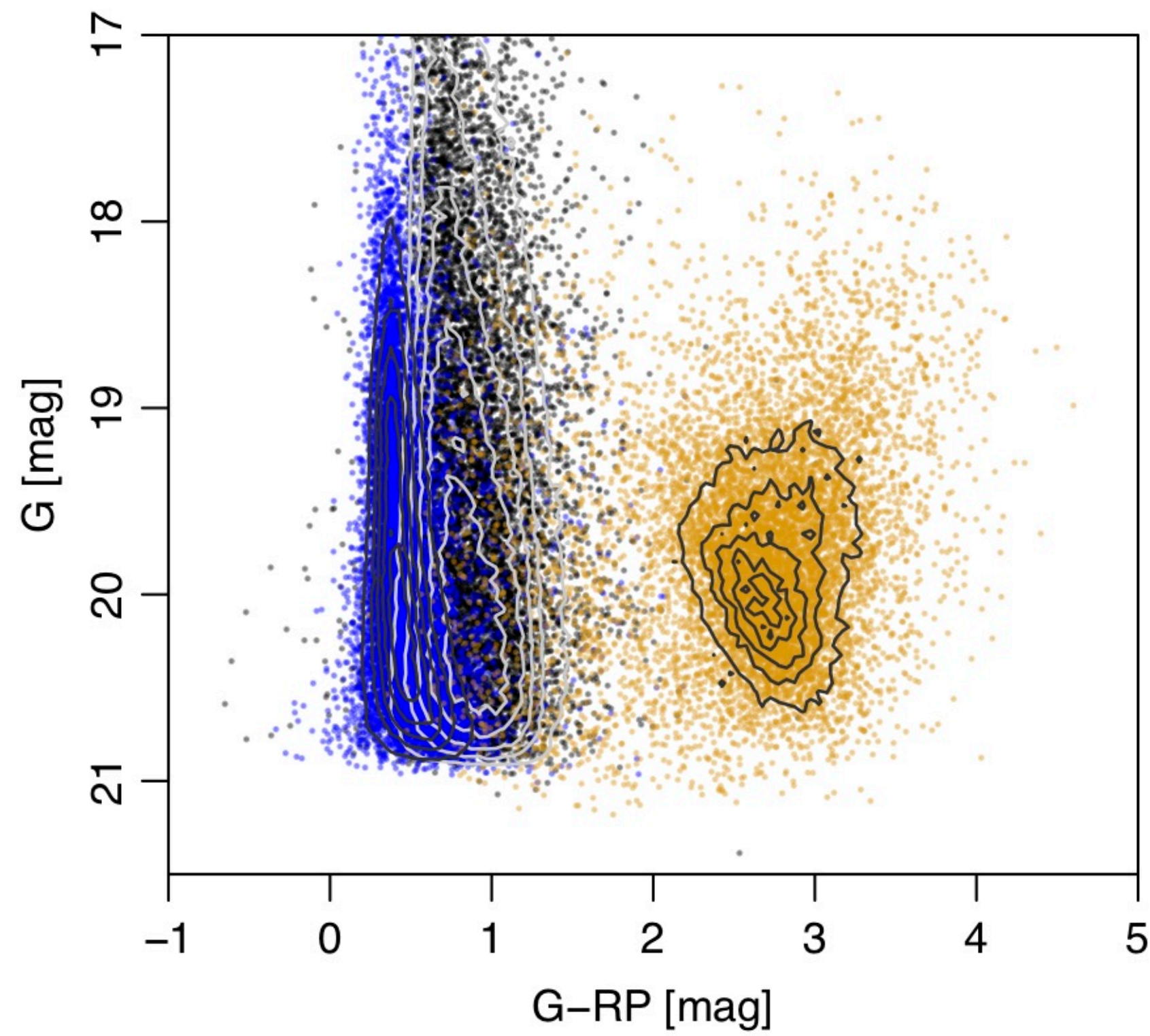
A probabilistic classifier



# DSC:Allosmod training features

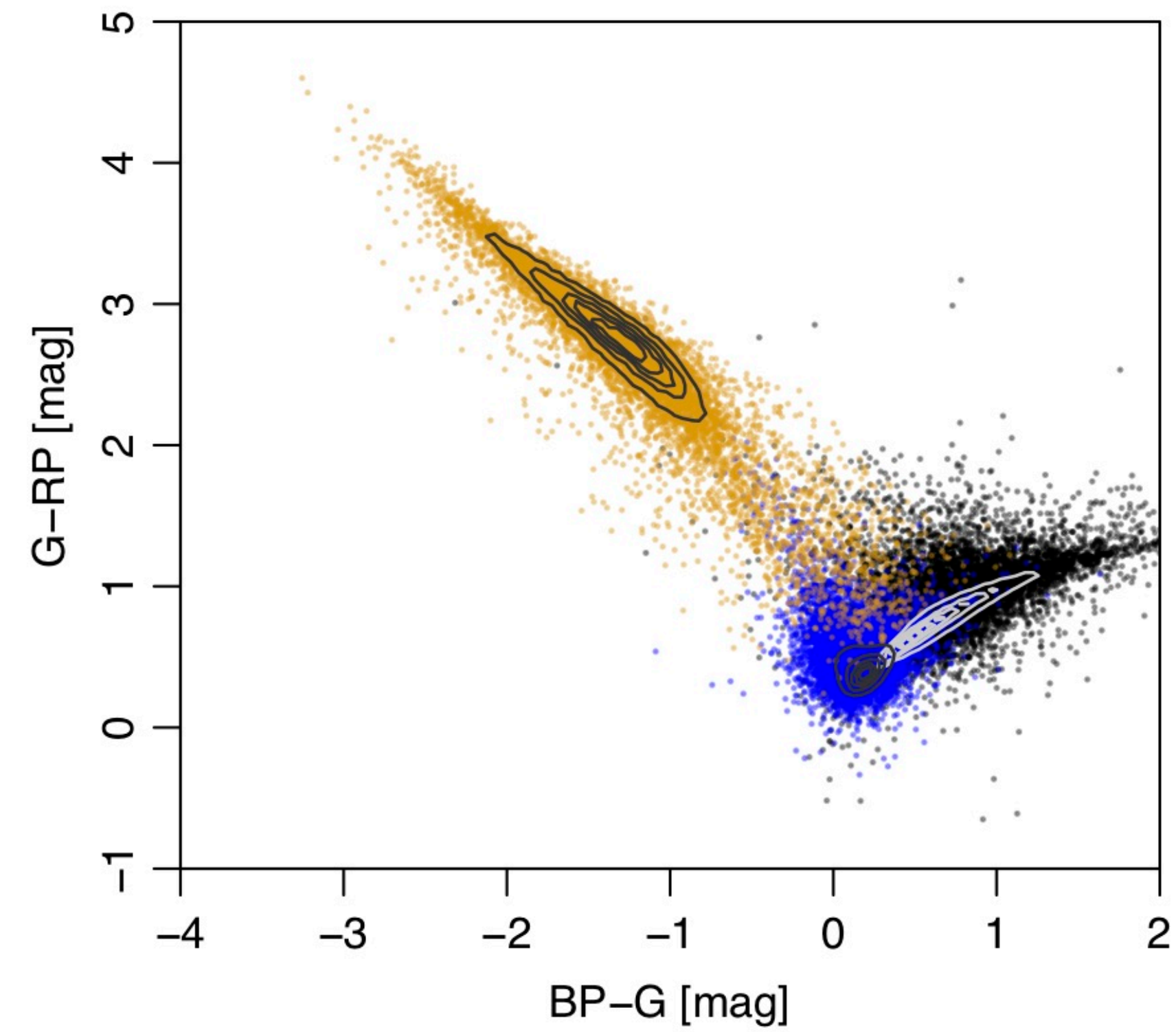


# DSC: CMD and CCD of training data



blue = quasars  
orange = galaxies  
black = stars

relative  
proportions  
not  
representative





# DSC: Class prior

	quasar	galaxy	star	white dwarf	physical binary star
$\propto$	1/1000	1/5000	1	1/5000	1/100
=	0.000989	0.000198	0.988728	0.000198	0.009887

## Important

The prior must also be taken into account when estimating the purity on validation data because validation data usually does not have enough stellar contaminants.

See [Bailer-Jones et al. \(2019; MNRAS 490, 5615\)](#) section 3.4

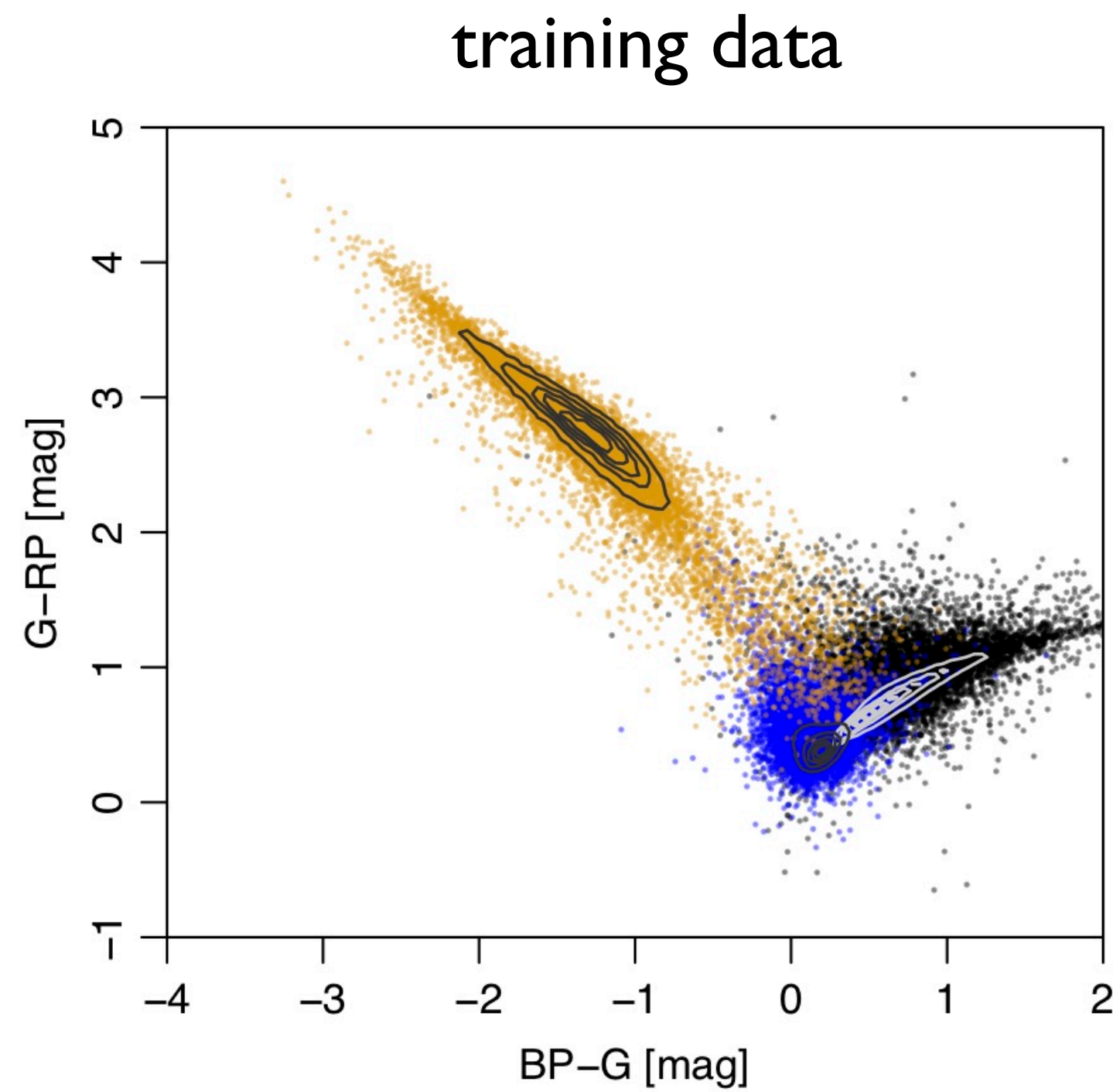
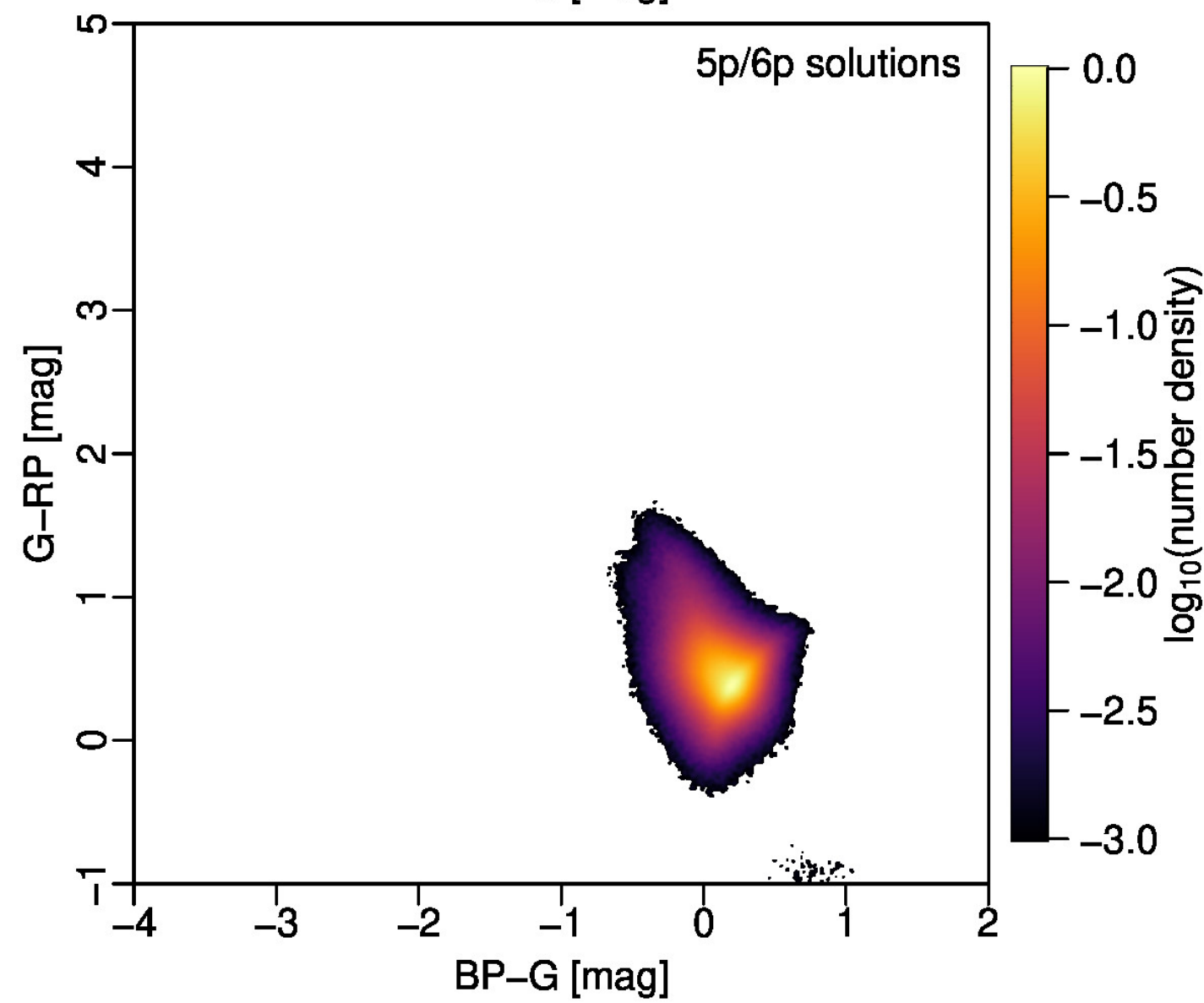
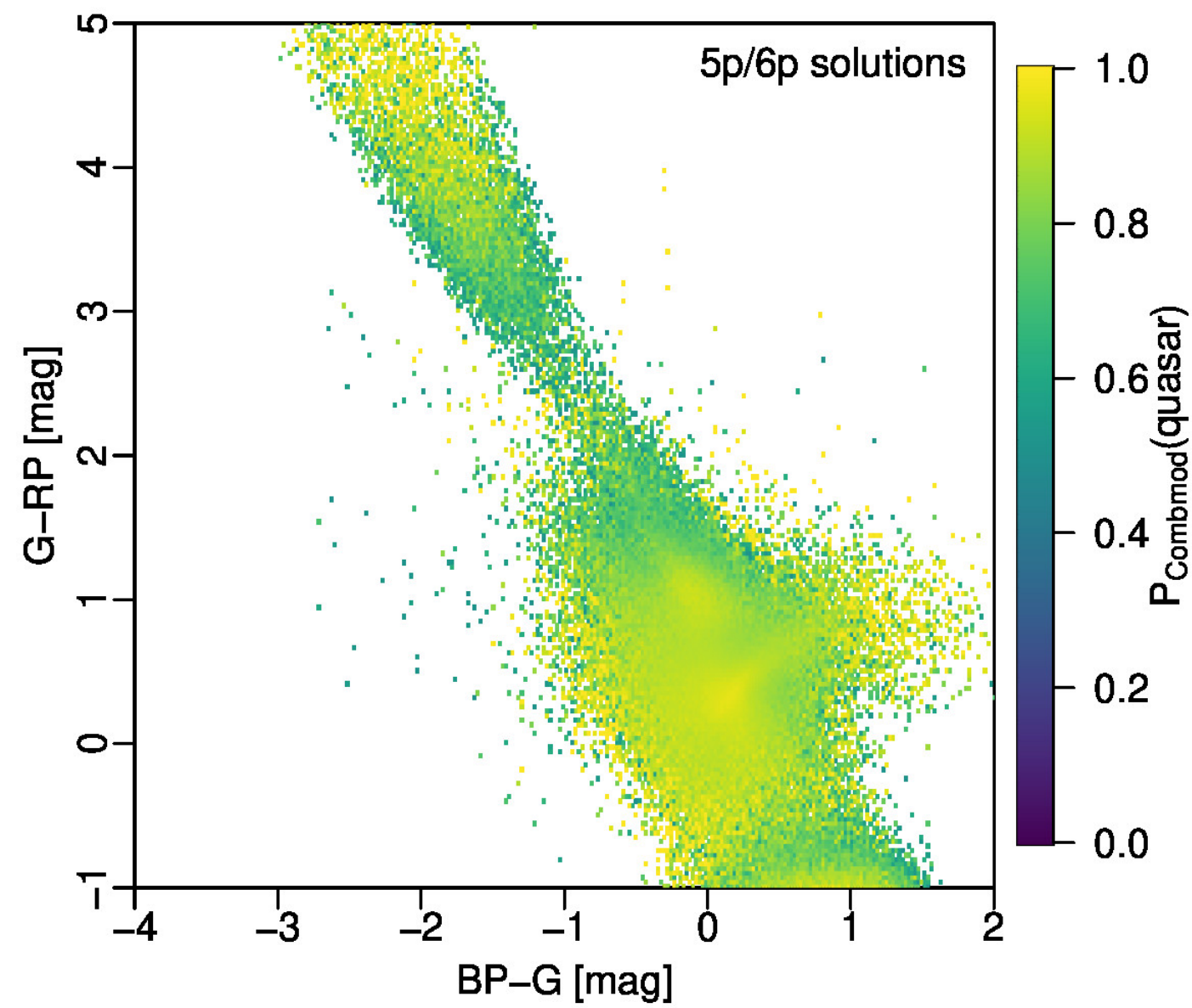
# DSC: Performance

classification by  $P > 0.5$

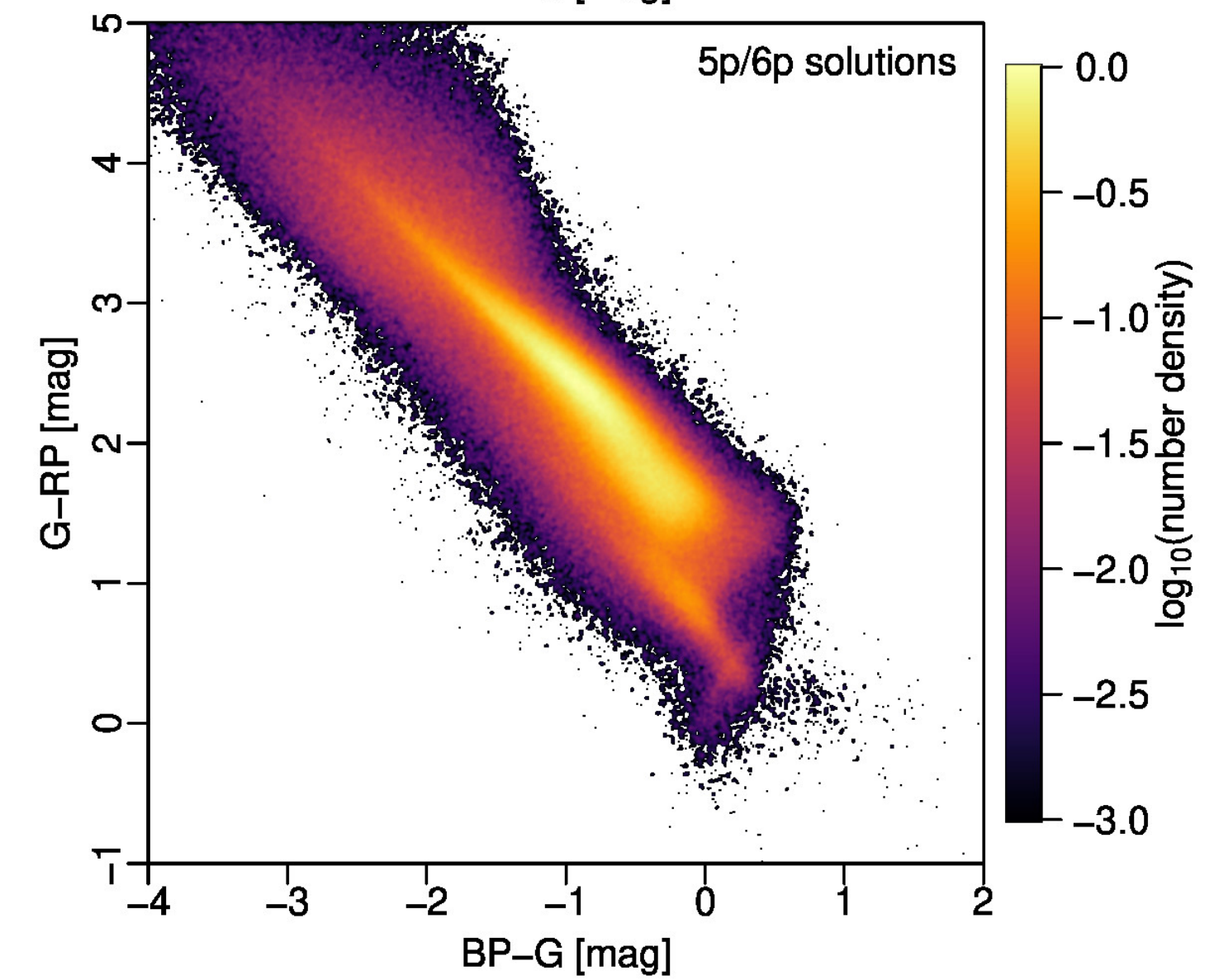
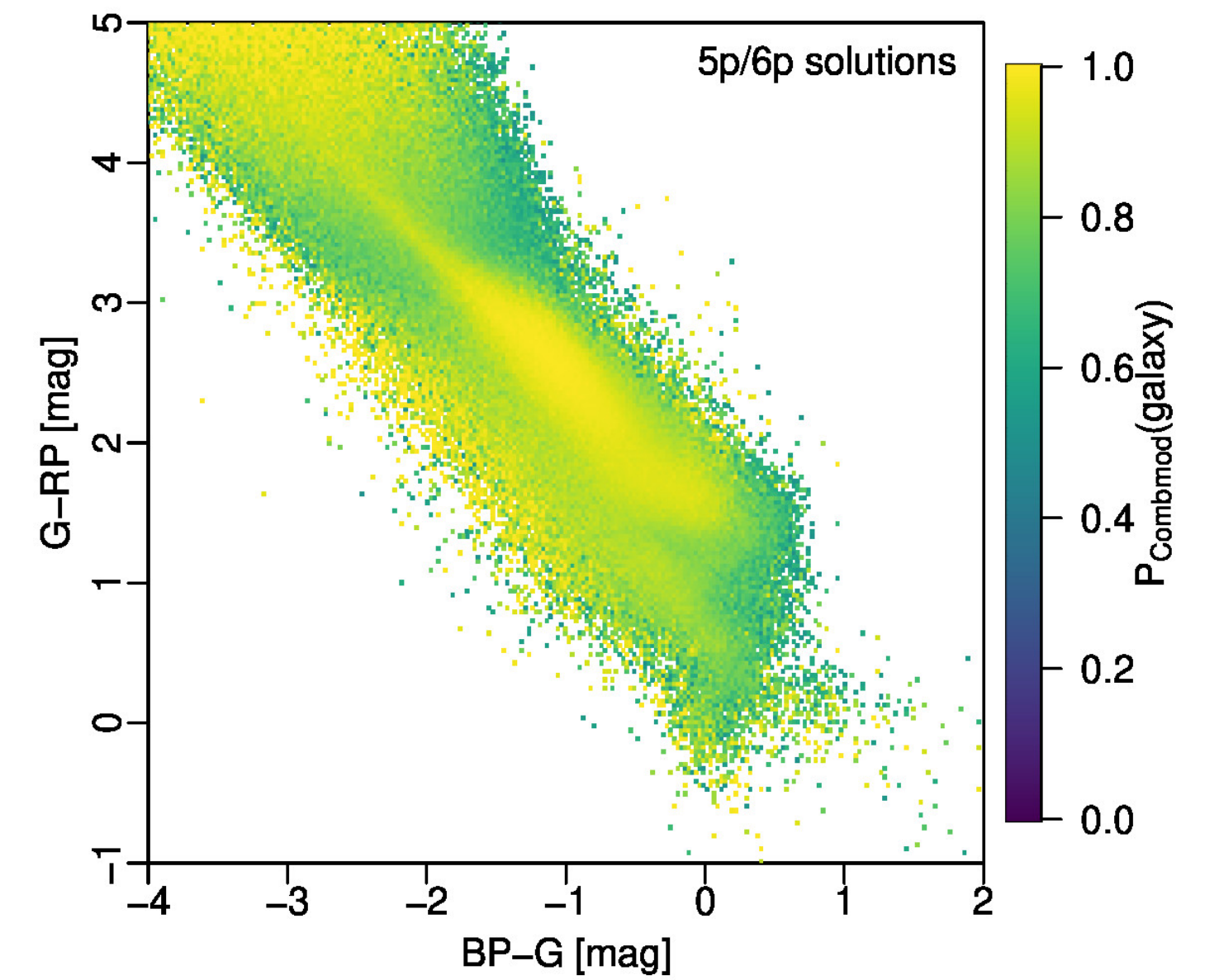
	Specmod		Allosmod		Combmod		Spec&Allos	
	compl.	purity	compl.	purity	compl.	purity	compl.	purity
quasar	0.409	0.248	0.838	0.408	0.916	0.240	0.384	0.621
galaxy	0.831	0.402	0.924	0.298	0.936	0.219	0.826	0.638
star	0.998	0.989	0.998	1.000	0.996	0.990	–	–
quasar, $ \sin b  > 0.2$	0.409	0.442	0.881	0.603	0.935	0.412	0.393	0.786
galaxy, $ \sin b  > 0.2$	0.830	0.648	0.928	0.461	0.938	0.409	0.827	0.817



# DSC: $P_{\text{Combmod}} > 0.5$



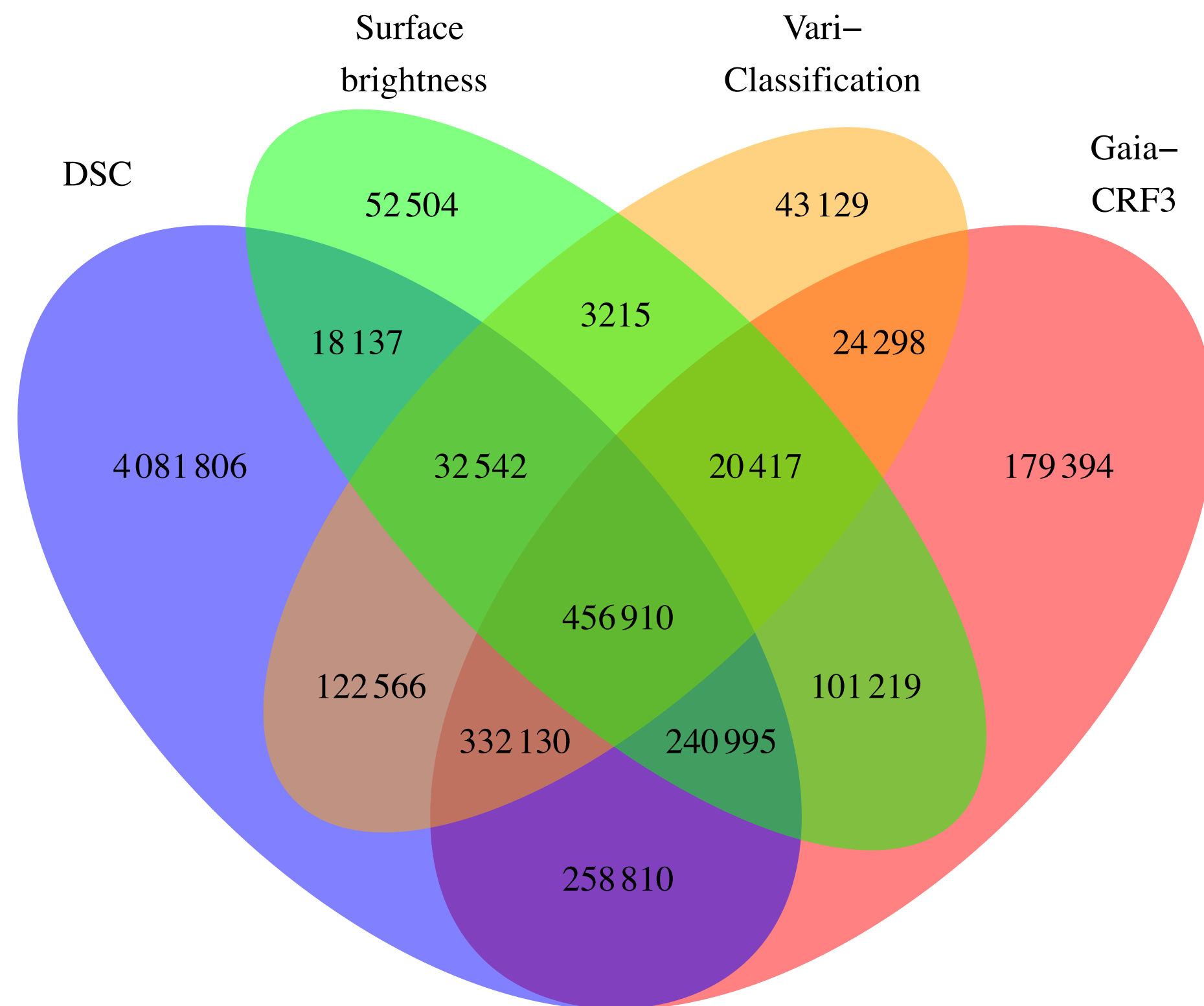
blue = quasars  
orange = galaxies  
black = stars



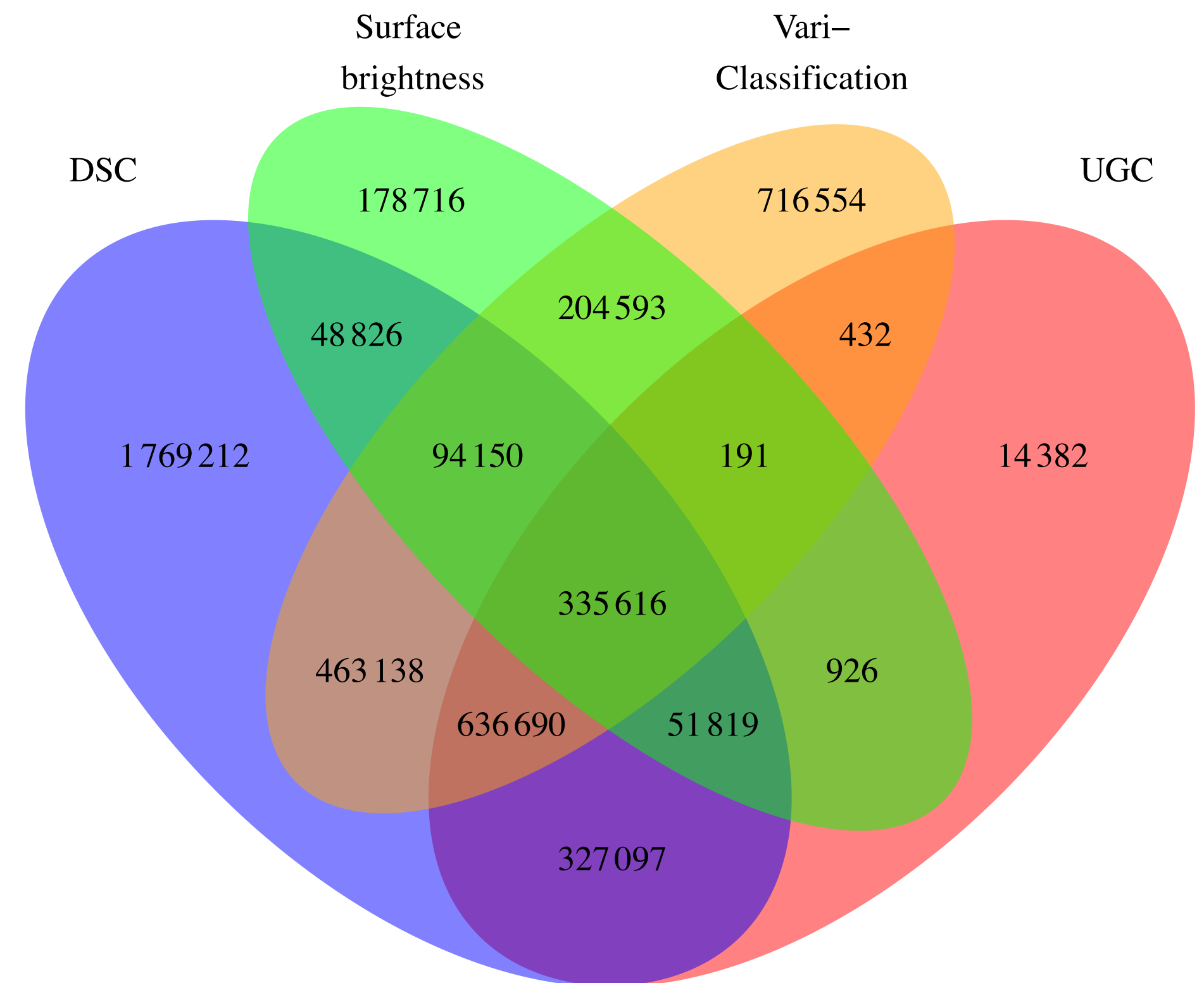


# Contributions to the extragalactic candidate tables

Quasar candidates  
6.6 million, 52% pure



Galaxy candidates  
4.8 million, 69% pure



# Comments on extragalactic candidate tables

- Heterogeneous sample
  - ▶ no common definition of “quasar” or “galaxy” across the modules
- Purity and completeness varies among subsets contributed by the modules
  - ▶ input lists and Vari driven by purity; DSC driven by completeness
  - ▶ higher purity subset achieved with further selections (at cost of completeness)
- If relative contamination is constant, absolute contamination follows source density
- Classification uses only Gaia data (higher purity expected with additional data)

# Purer subsets

**Table 11.** ADQL query to select the purer quasar sub-sample.

---

```
SELECT * FROM gaiadr3.qso_candidates
WHERE (gaia_crf_source='true' OR
       host_galaxy_flag<6 OR
       classlabel_dsc_joint='quasar' OR
       vari_best_class_name='AGN')
```

---

**Table 12.** ADQL query to select the purer galaxy sub-sample.

---

```
SELECT * FROM gaiadr3.galaxy_candidates
WHERE (radius_sersic IS NOT NULL OR
       classlabel_dsc_joint='galaxy' OR
       vari_best_class_name='GALAXY')
```

---

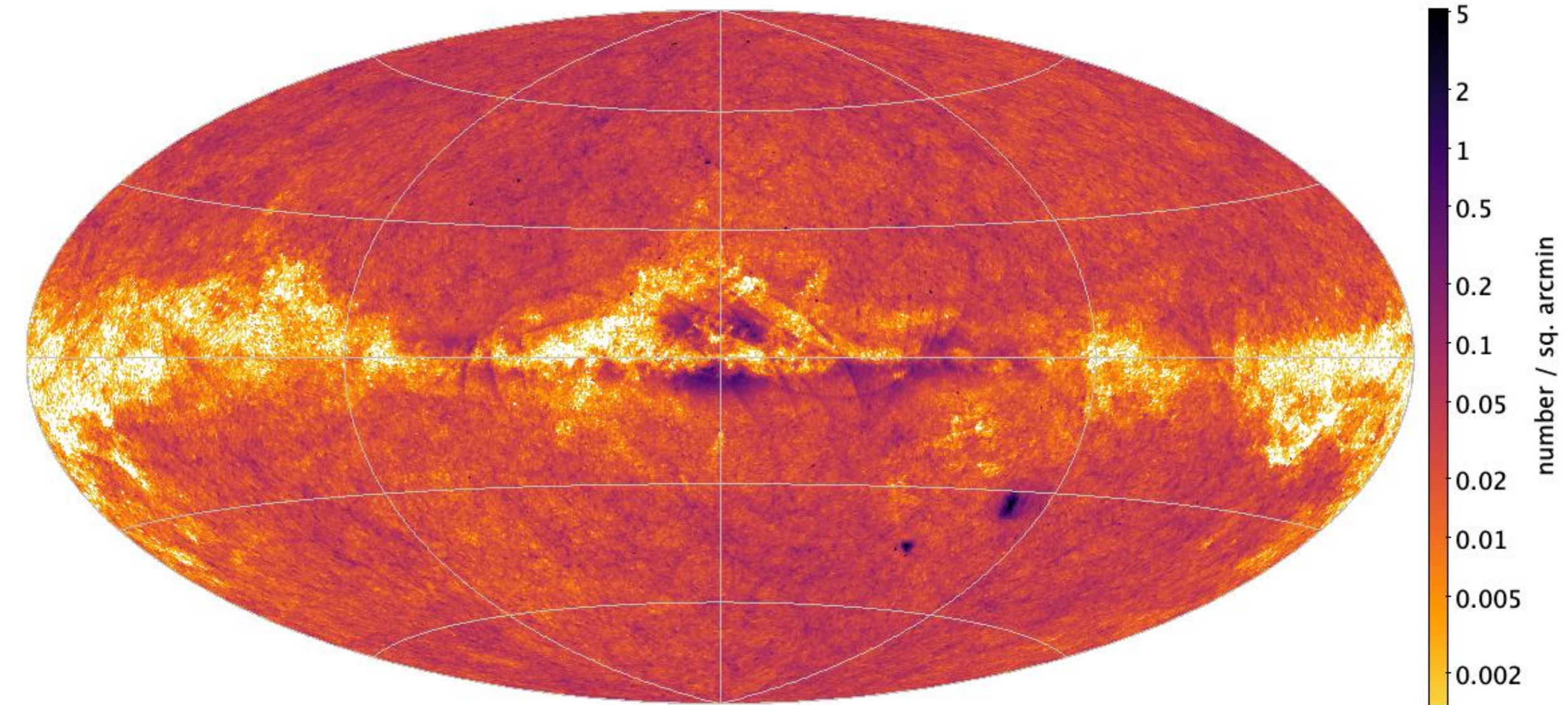
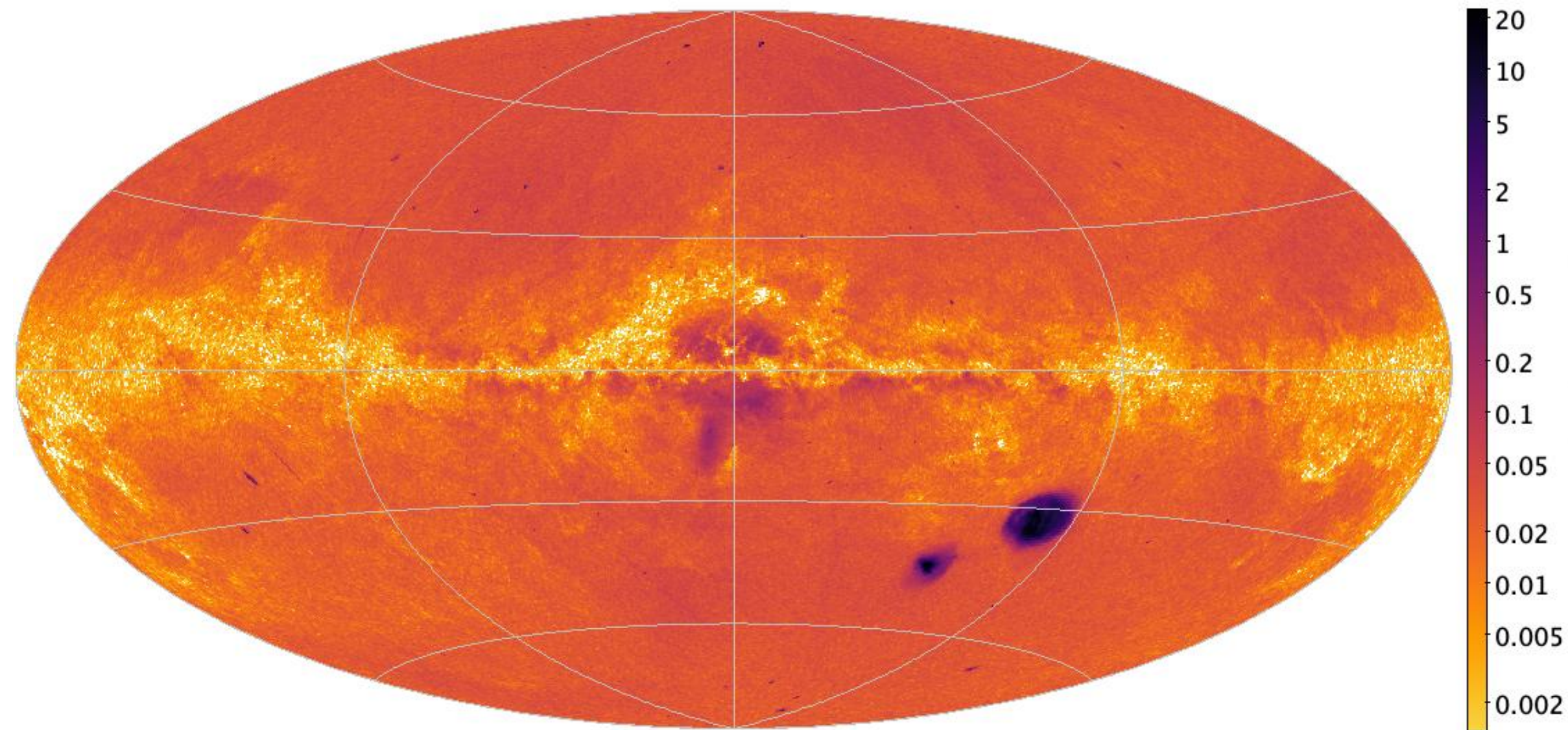


# Sky distribution of extragalactic candidates

## Full set

Quasar candidates  
6.6 million, 52% pure

Galaxy candidates  
4.8 million, 69% pure



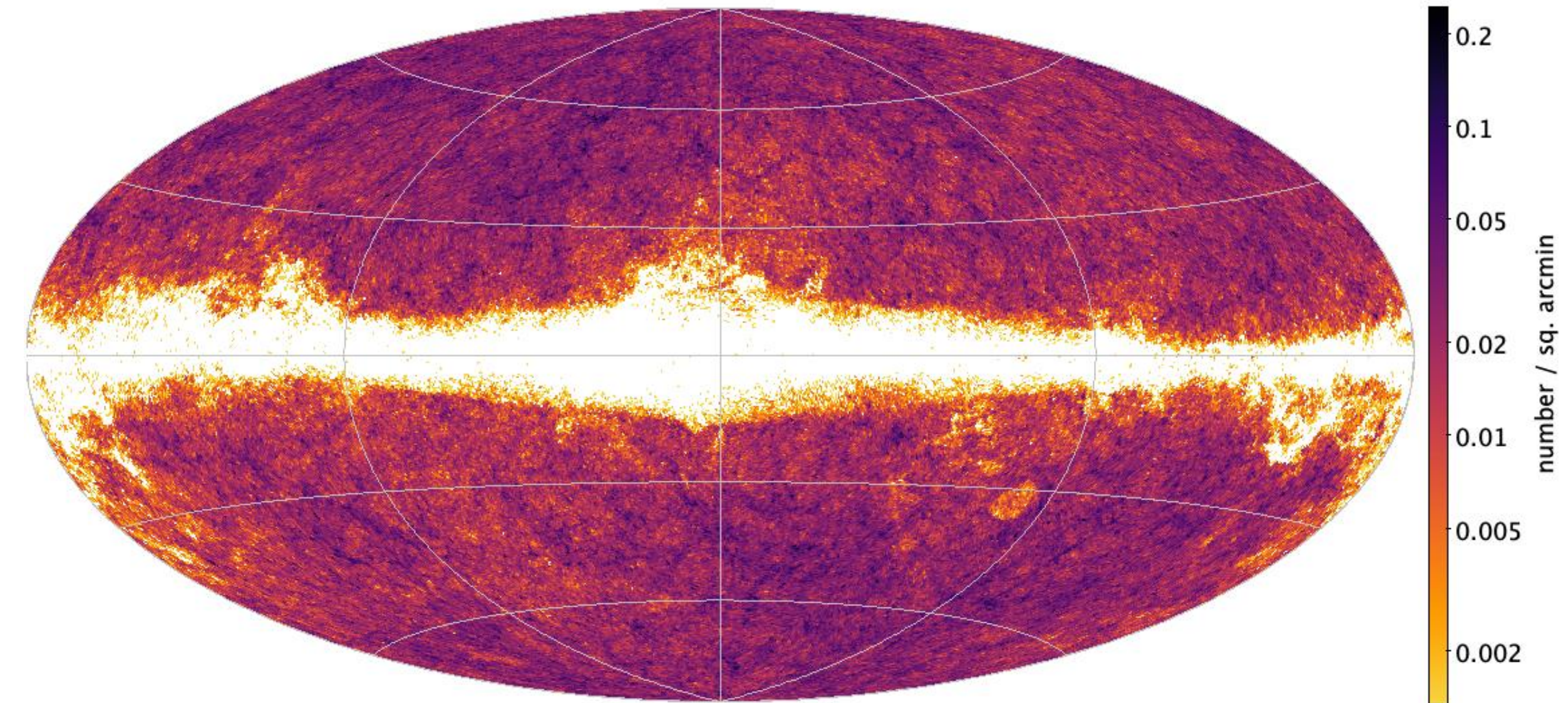
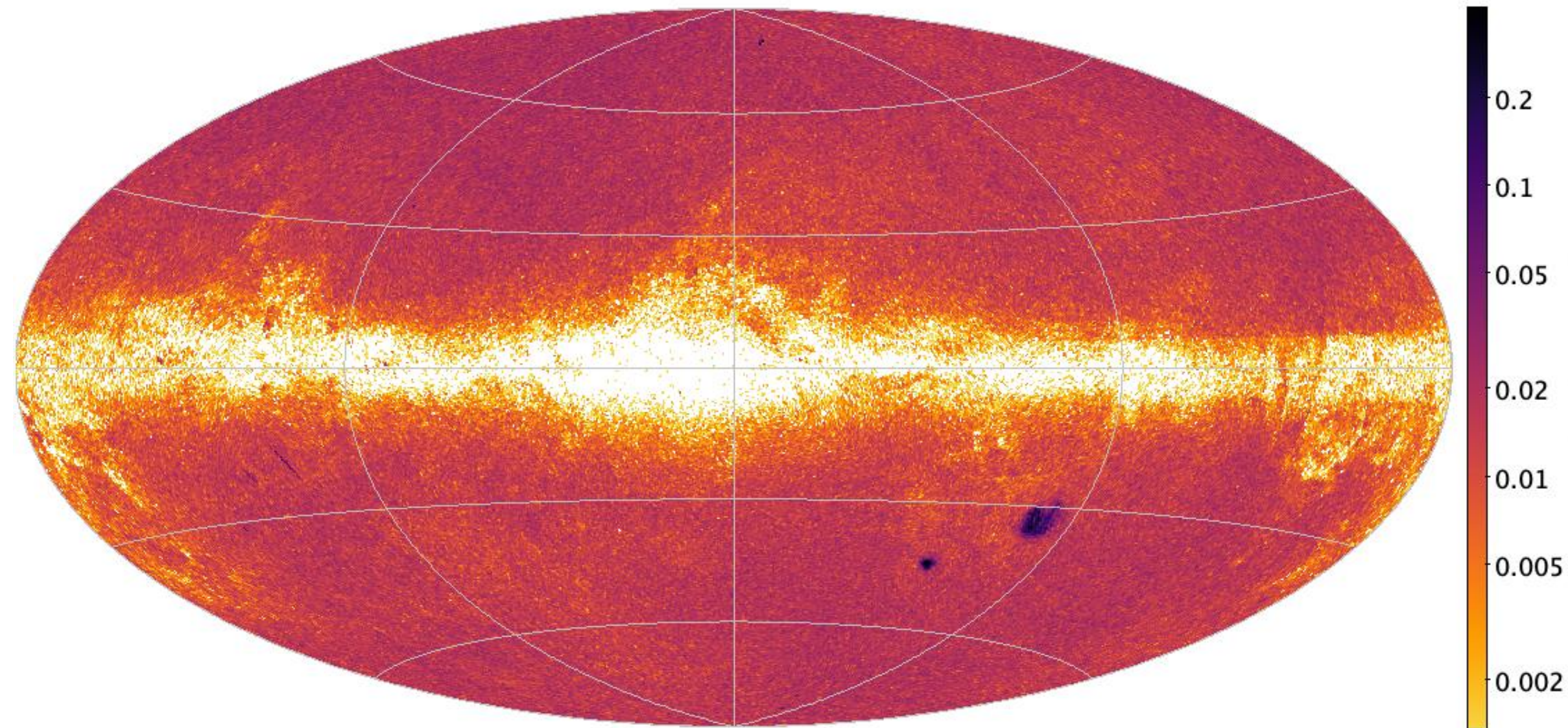


# Sky distribution of extragalactic candidates

## Purer subset

Quasar candidates  
1.9 million, 95% pure

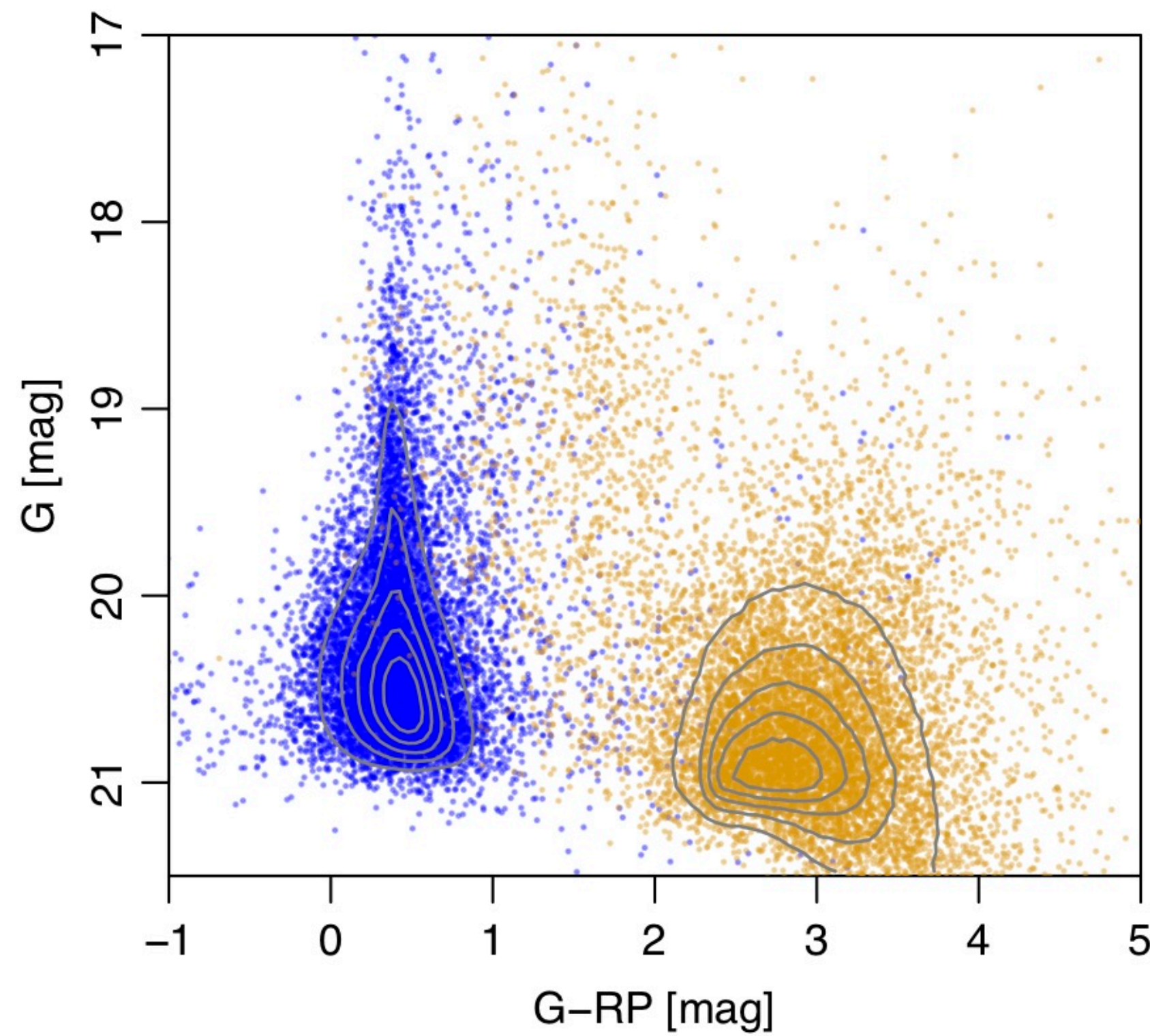
Galaxy candidates  
2.9 million, 94% pure



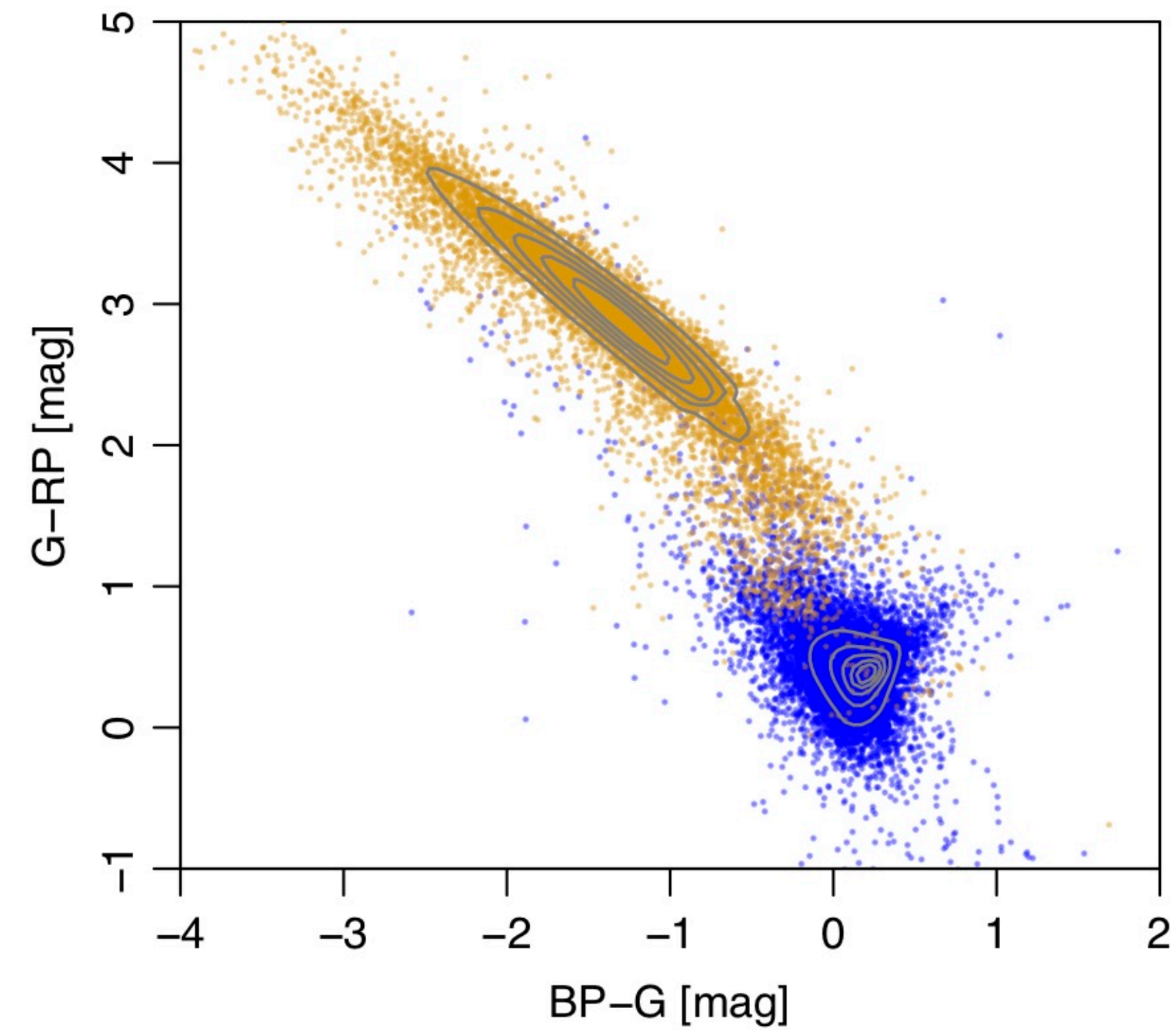


# Colour, magnitude distributions of extragalactic candidates

## Full set



blue = quasars  
orange = galaxies

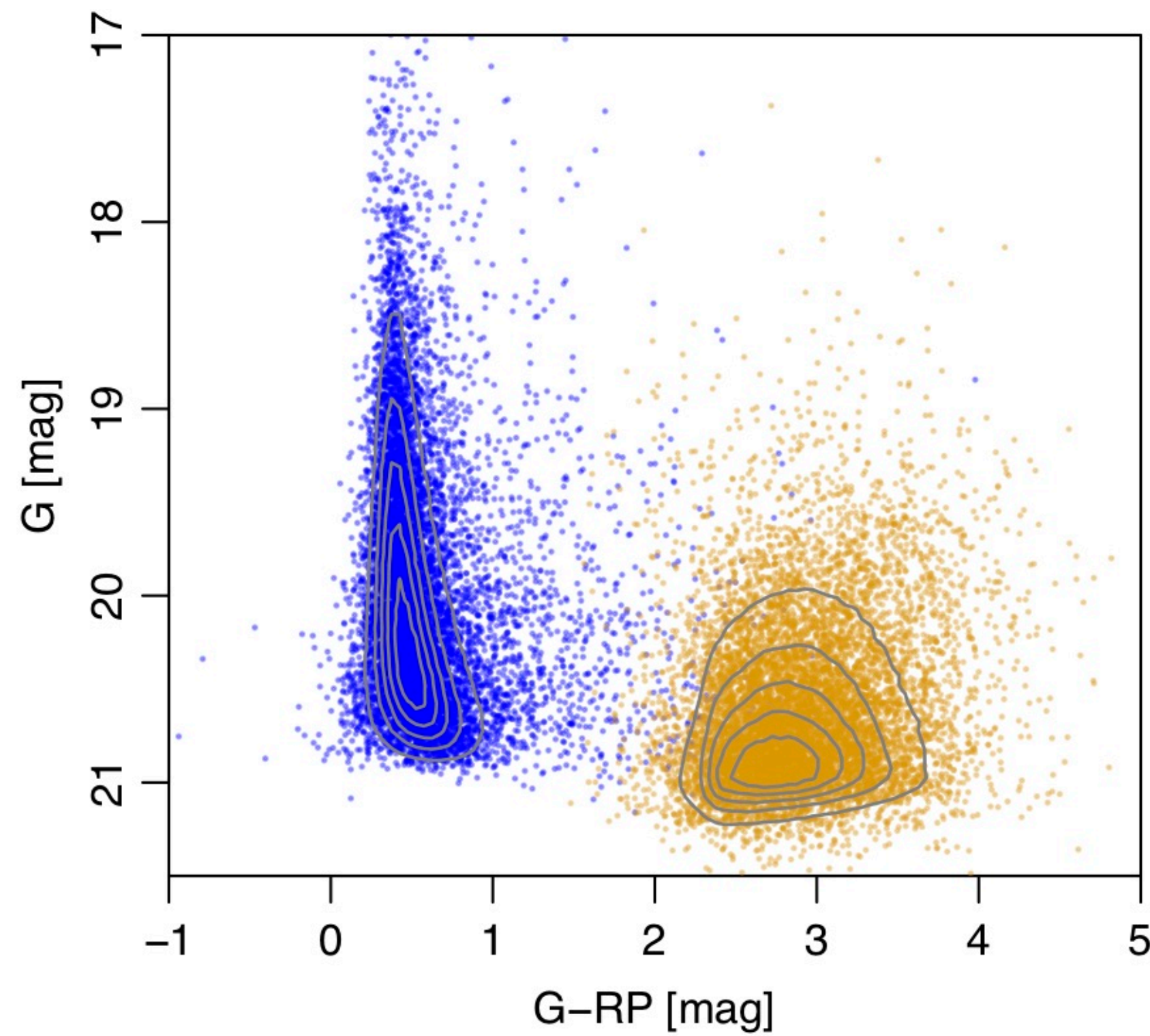


10 000 random sources of each class, linear density contours of full set

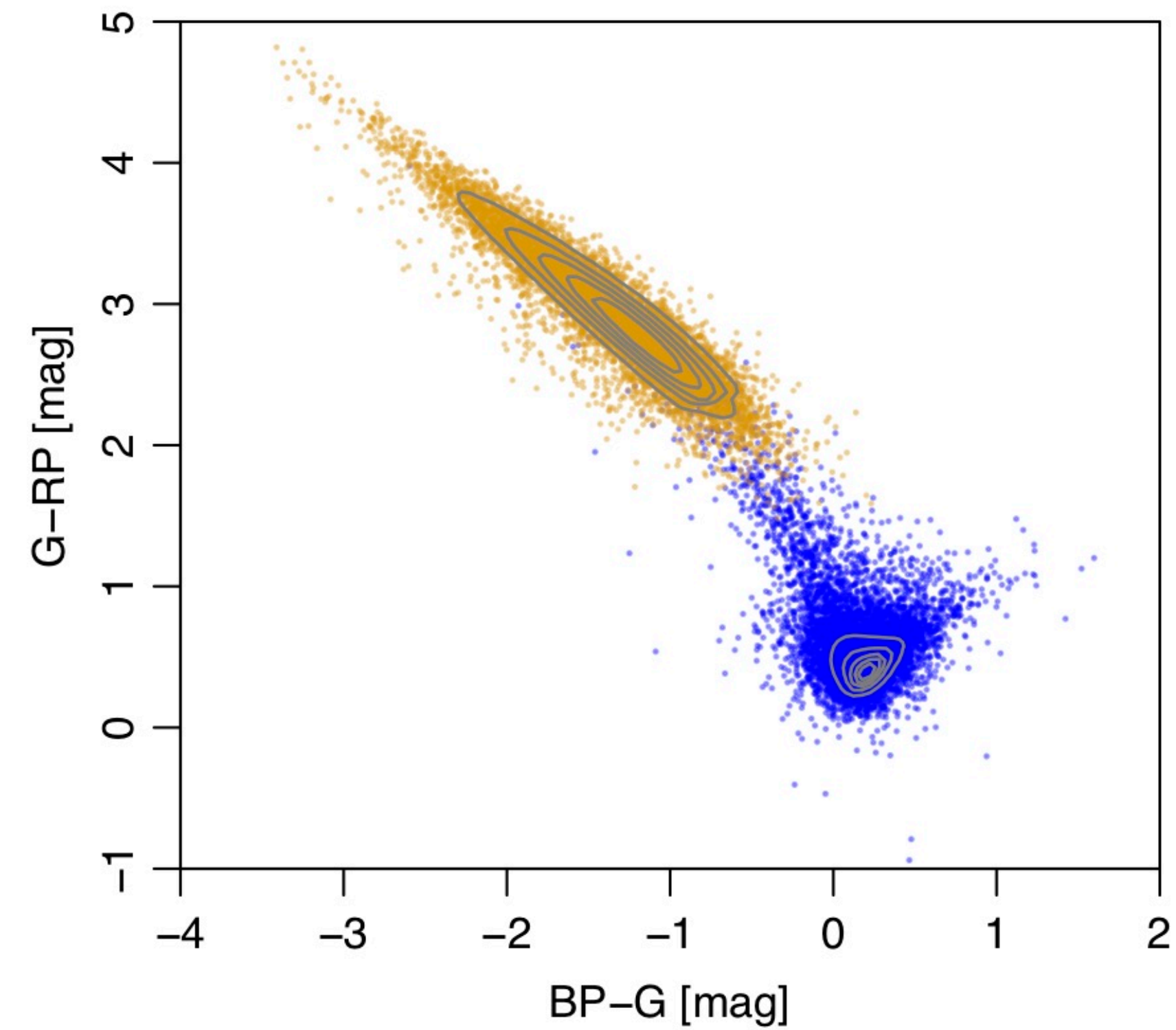


# Colour, magnitude distributions of extragalactic candidates

## Purer subset

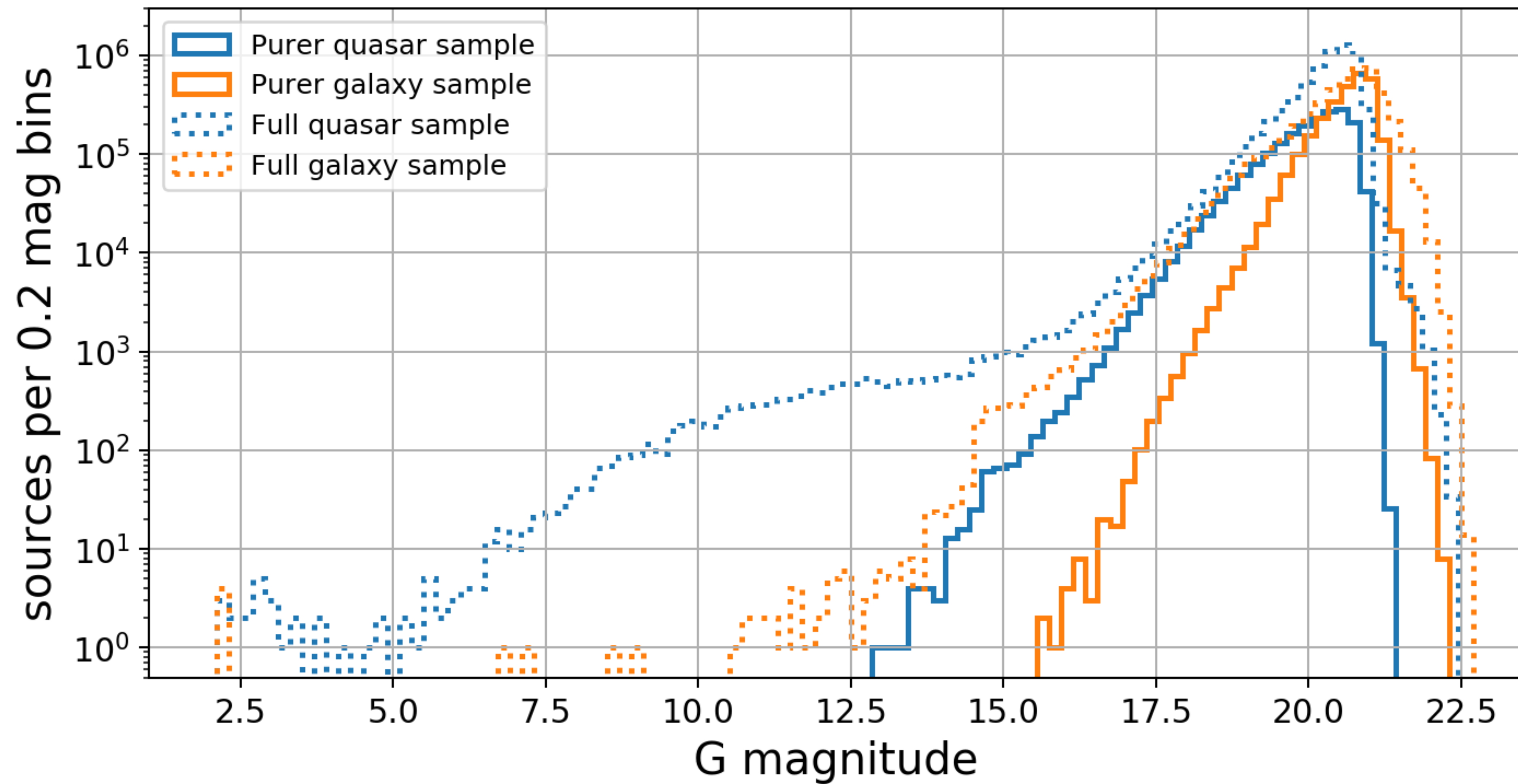


blue = quasars  
orange = galaxies



10 000 random sources of each class, linear density contours of full set

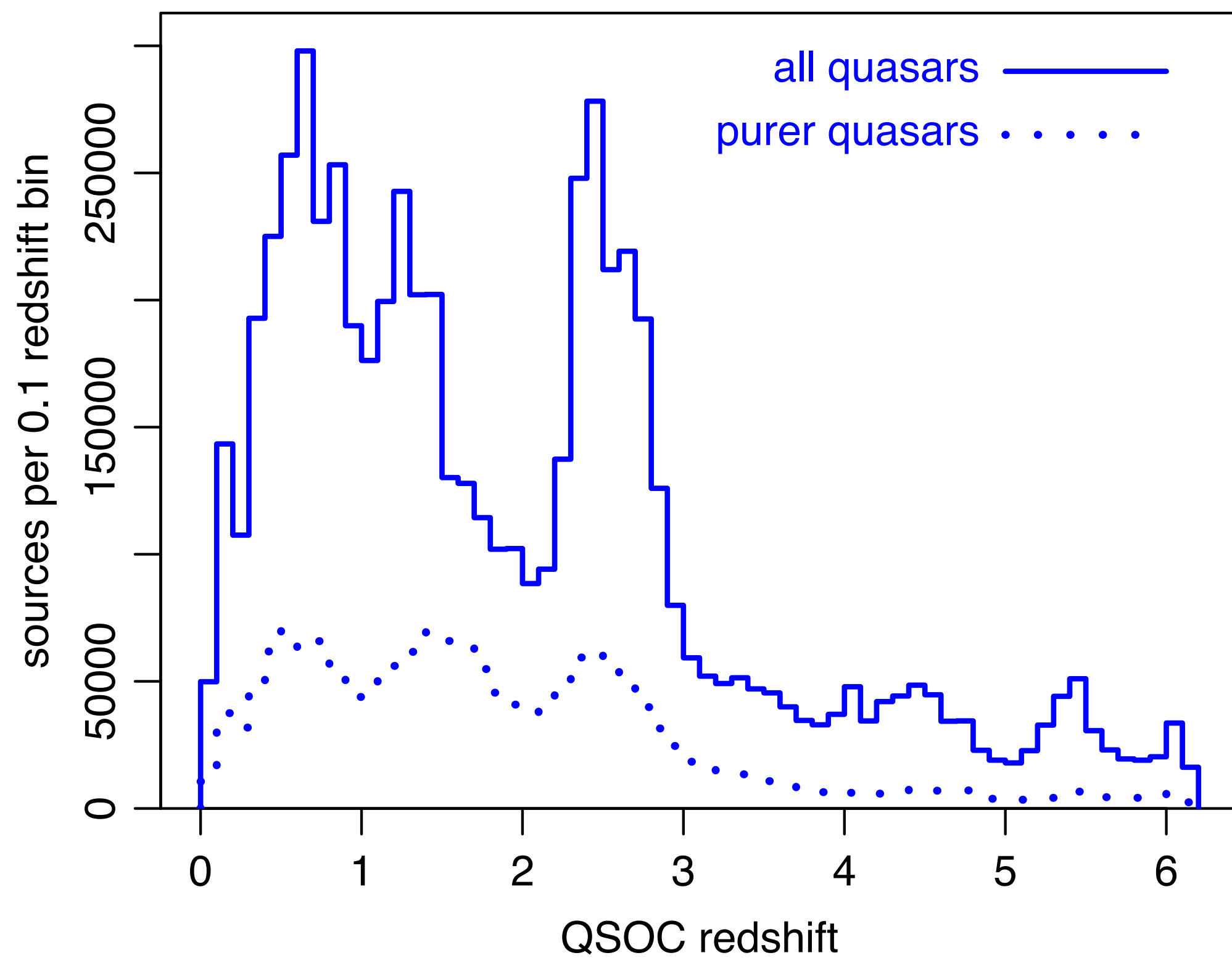
# Magnitude distributions



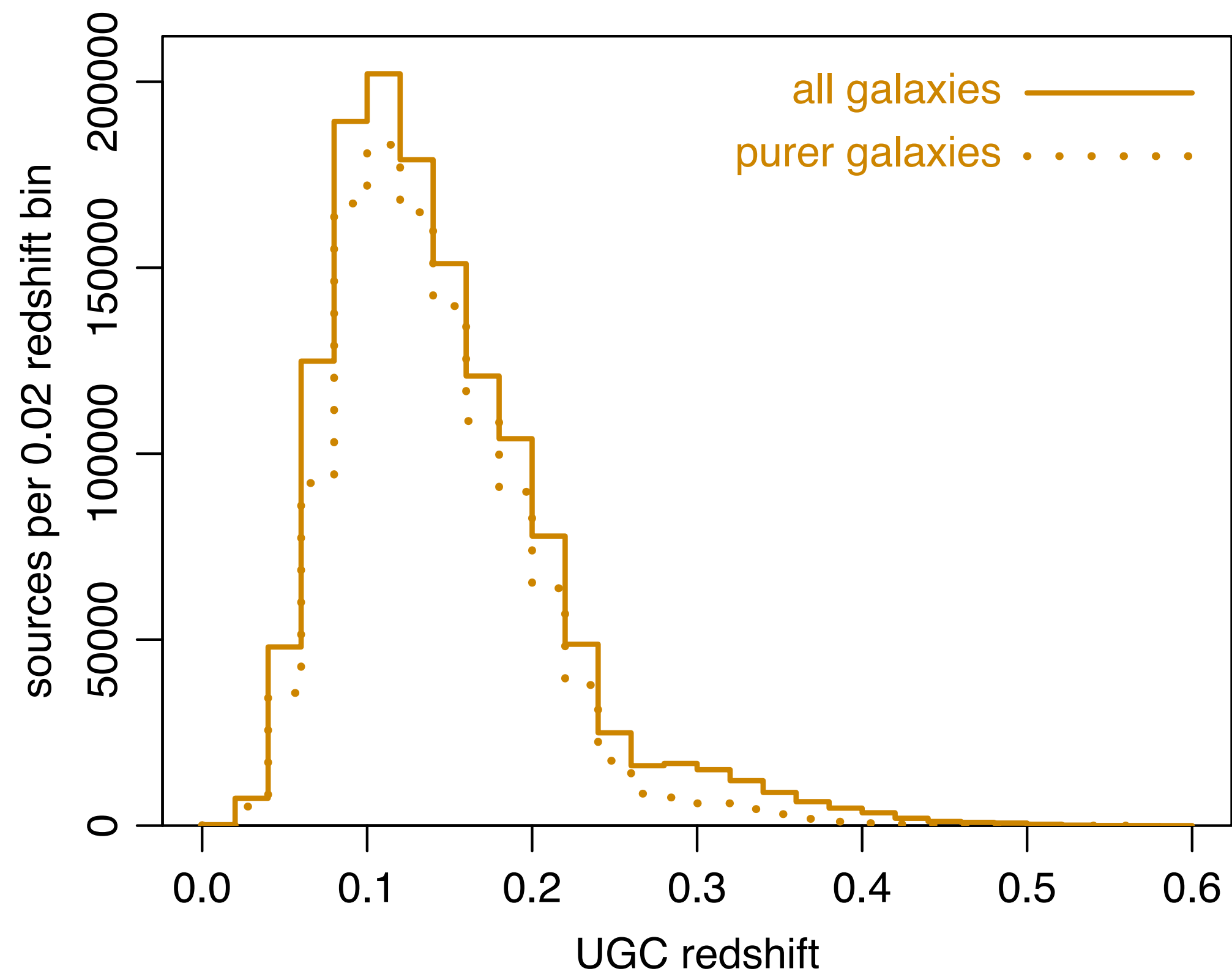


# Redshift distributions

6.4 million and 1.7 million sources



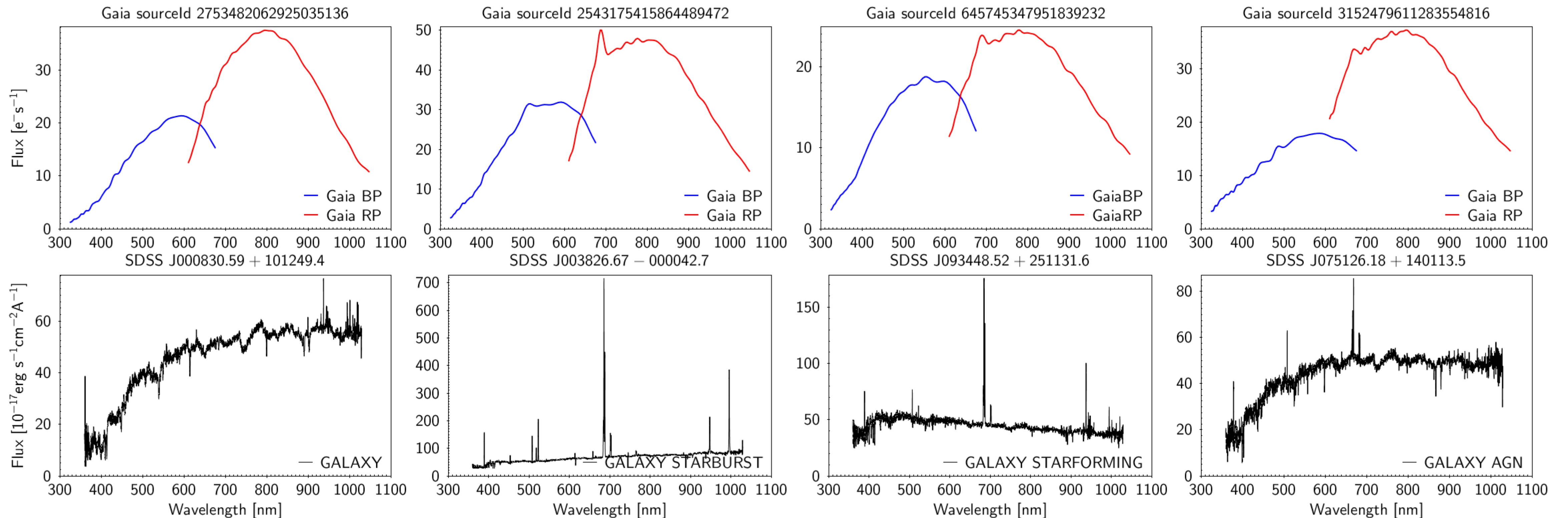
1.4 million and 1.1 million sources



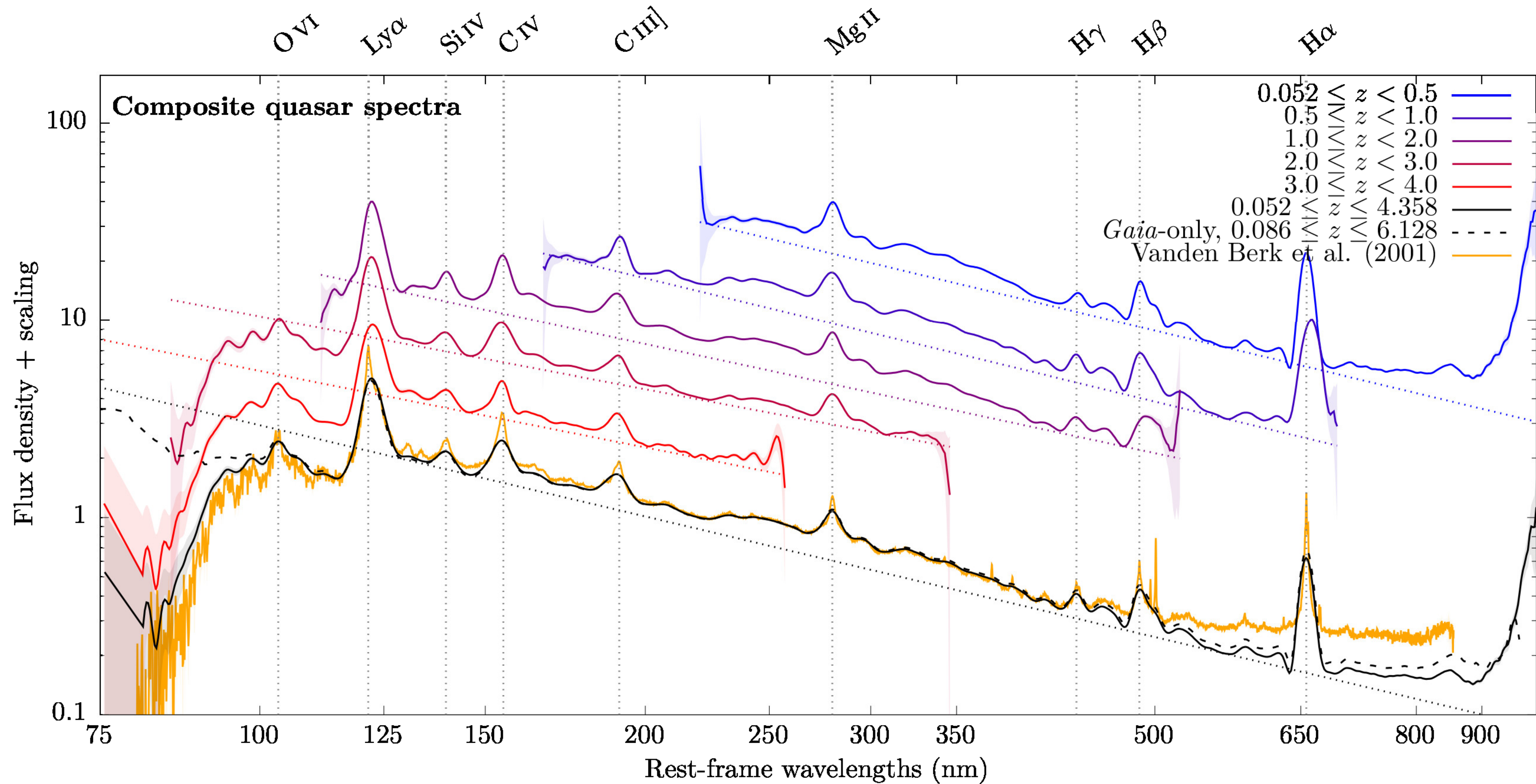
# BP/RP spectra

163 000 quasar candidate spectra published (119 000 in purer subset)  
26 500 galaxy candidate spectra published (12 600 in purer subset)

BP/RP compared to SDSS for some galaxy candidates



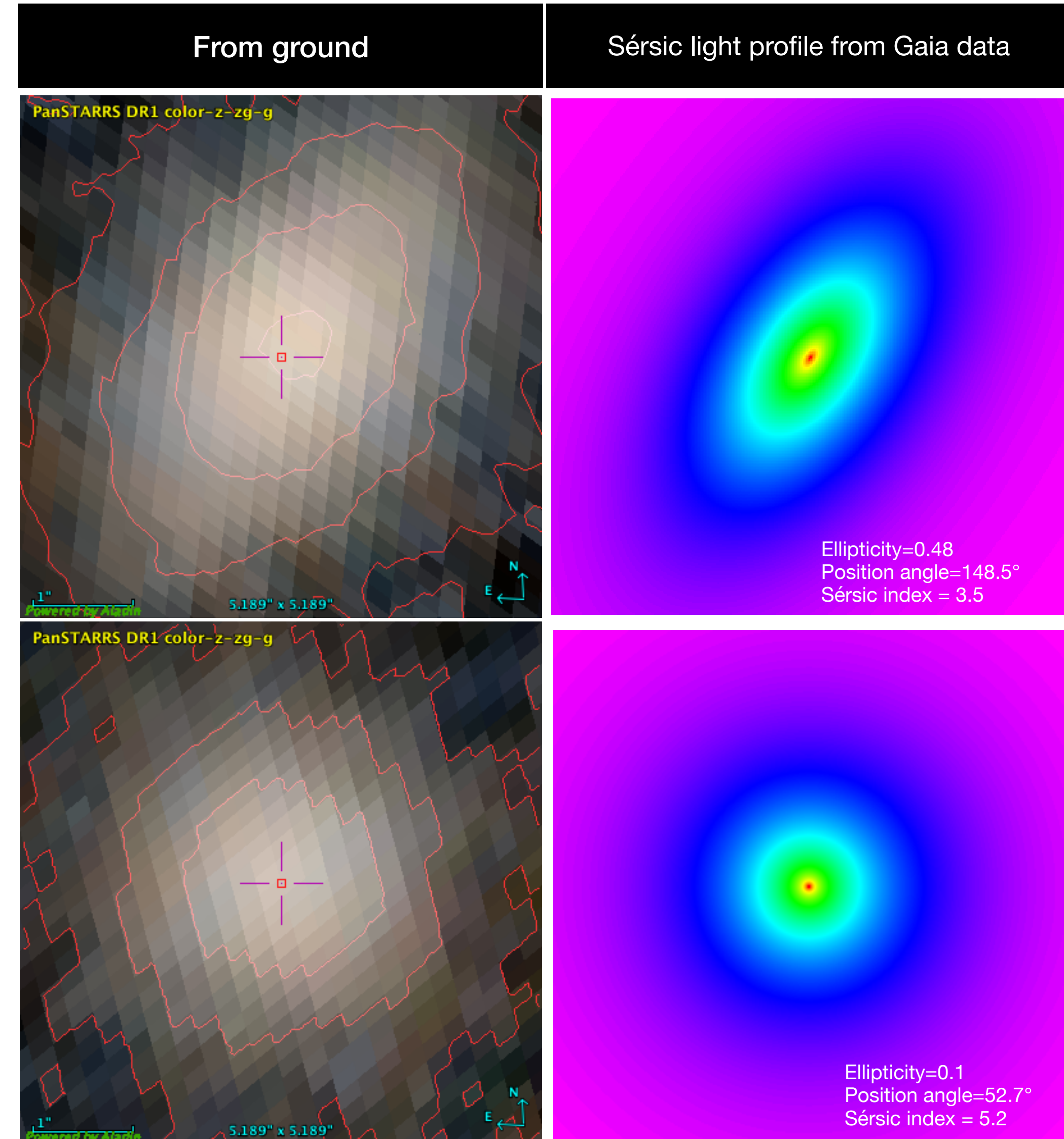
# Quasar composite spectra





# Surface brightness profiles

- 65 000 host galaxies of quasars detected
- 16 000 host galaxies of quasars with a fitted Sérsic profile
- 915 000 galaxies with fitted Sérsic and de Vaucouleurs profiles





# Summary

- Candidate tables are a mixture of input lists and Gaia data classification
- Complete tables
  - ▶ 6.6 million quasar candidates, 4.8 million galaxy candidates
  - ▶ 50 - 70% pure
- Purer subset (using a simple ADQL query)
  - ▶ 1.9 million quasar candidates, 2.9 million galaxy candidates
  - ▶ 95% pure
- Classifications, BP/RP spectra, redshifts, and 2D spatial profile fits