A scientific guide to Gaia Astrophysical Parameters

Day 3

organized by Coordination Unit 8

The Discrete Source Classifier in Gaia DR3

Sara Jamal, Coryn Bailer-Jones
Max Planck Institute for Astronomy, Heidelberg

CU8 webinar, March 2023

Overview

1. Extragalactic results in GDR3
2. The Discrete Source Classifier (DSC) in GDR3: design and performance
3. The extragalactic candidate tables in GDR3
4. Improving DSC in GDR4

Extragalactic results in GDR3

Extragalactic results in GDR3

- 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: Gaia Collaboration (Bailer-Jones et al.) 2023

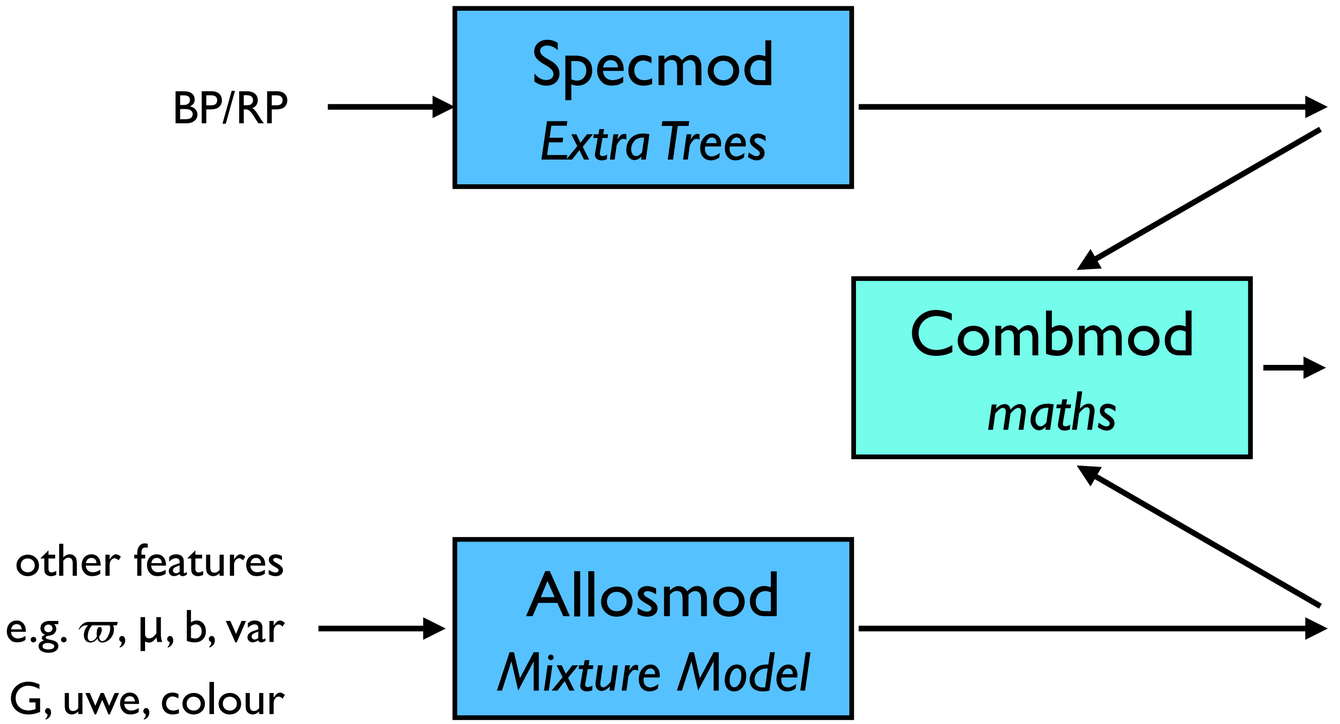
The Discrete Source Classifier (DSC)

The Discrete Source Classifier (DSC)

Link: DSC overview

Link: DSC details

A probabilistic classifier

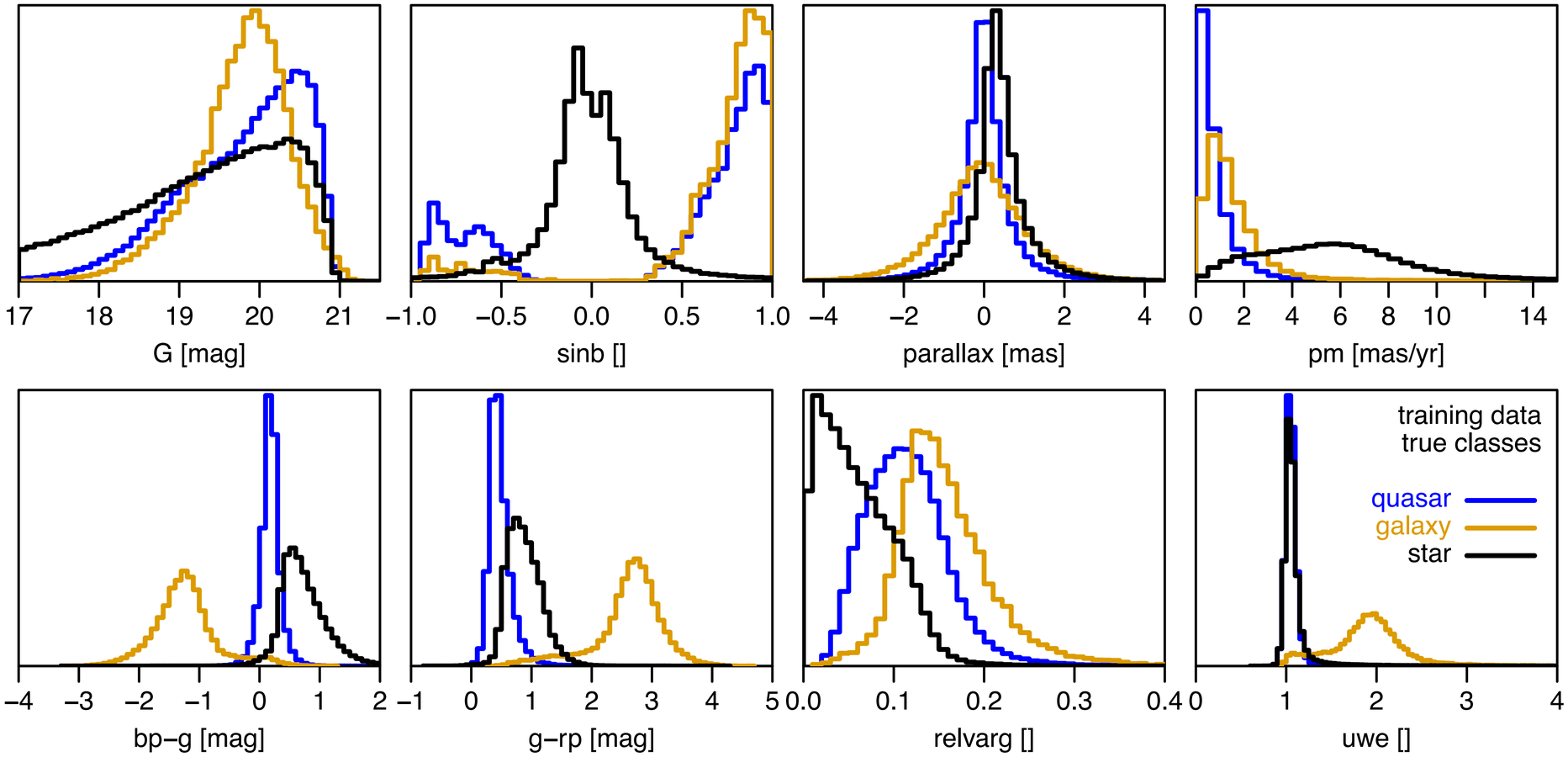


if
P_C(quasar)>0.5
then
classlabel_dsc = "quasar"

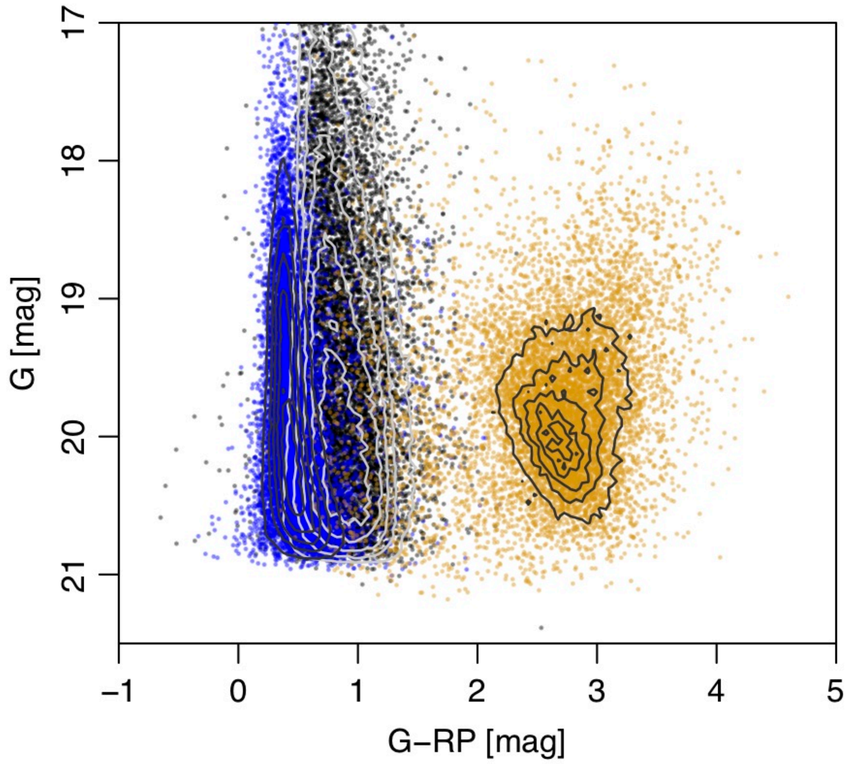
if
P_S(quasar)>0.5 & P_A(quasar>0.5)
then
classlabel_dsc_joint = "quasar"

and similarly for "galaxy"

DSC: Allosmod training features

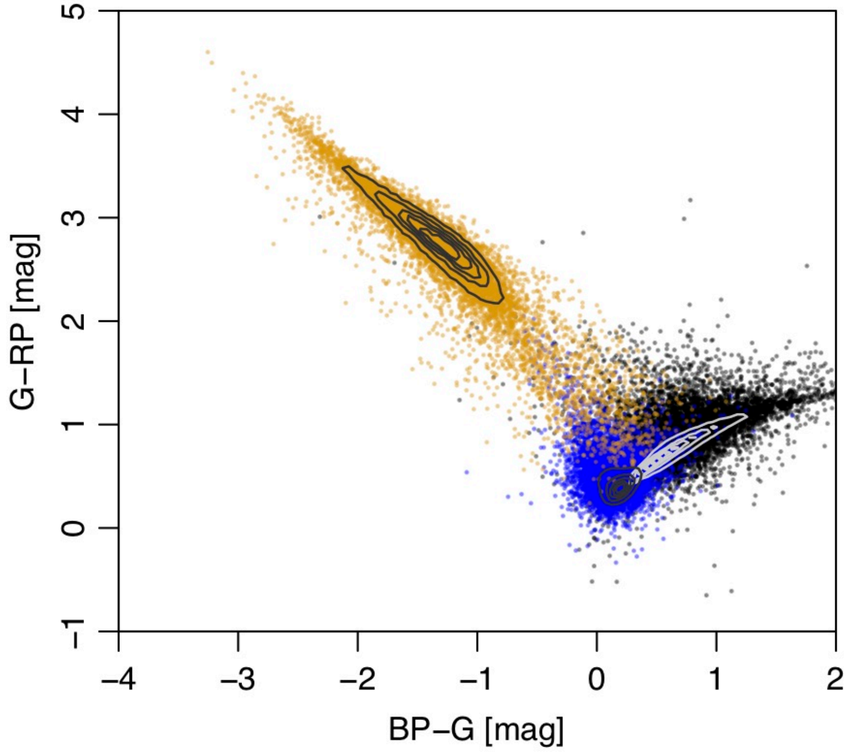


DSC: CMD and CCD of training data



blue = quasars
orange = galaxies
black = stars

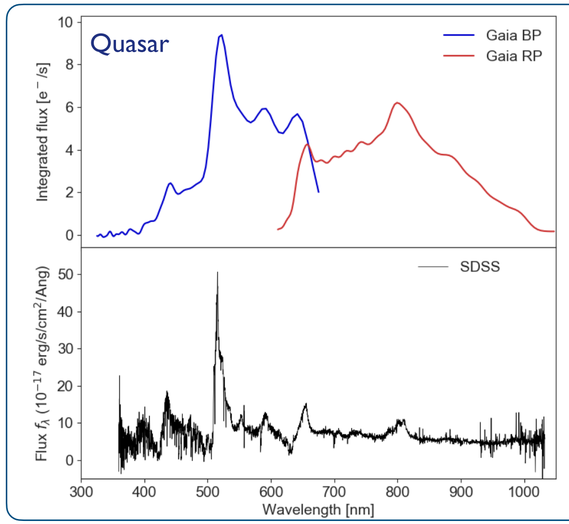
relative proportions not representative



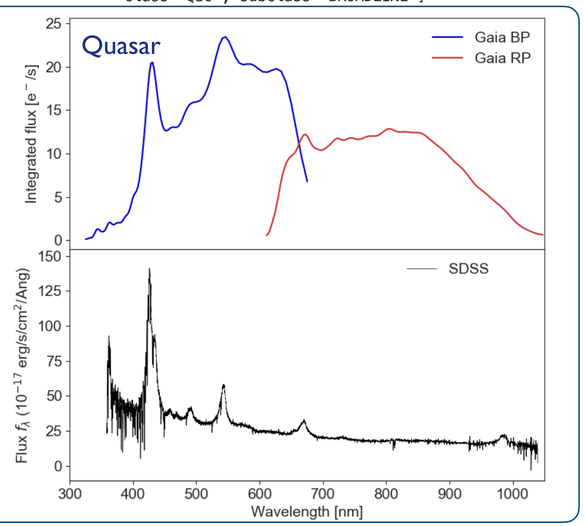
DSC: XP spectra of training data

- Selection of quasar and galaxy sources from SDSS, crossmatched with Gaia data.
- Sources from the star class drawn from a random selection on Gaia data.

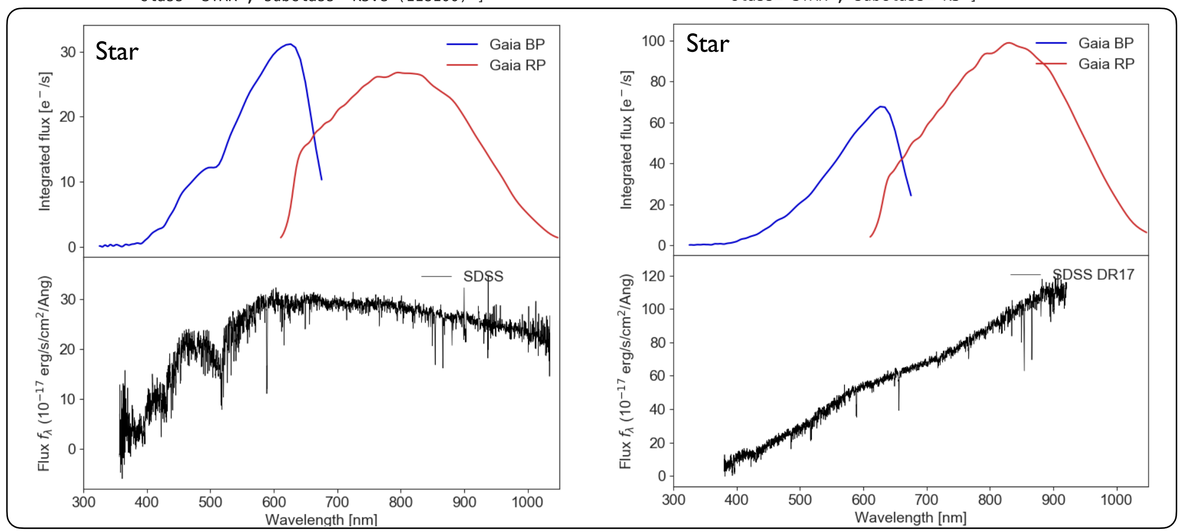
#QUASAR CLASS
Gaia DR3 666228626806595840
SDSS [specObjID 10808710709942441984, z=3.234,
class='QSO', subClass='BROADLINE']



#QUASAR CLASS
Gaia DR3 657558226003596288
SDSS [specObjID 5066737140521785344, z=2.510,
class='QSO', subClass='BROADLINE']

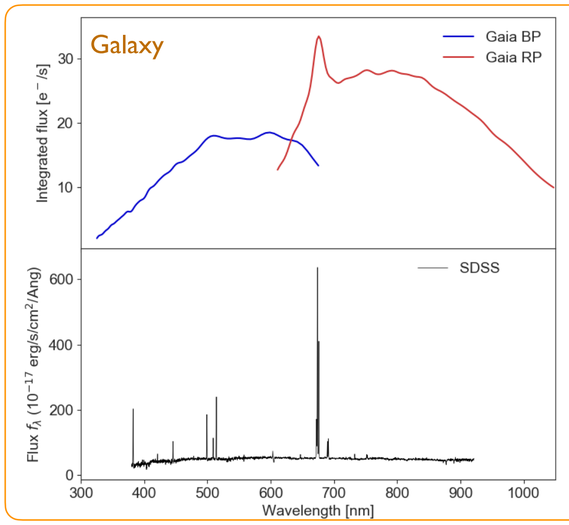


#STAR CLASS
Gaia DR3 362178958390810624
SDSS [specObjID 7739556466051536896, z=0.0,
class='STAR', subClass='K5Ve (118100)']

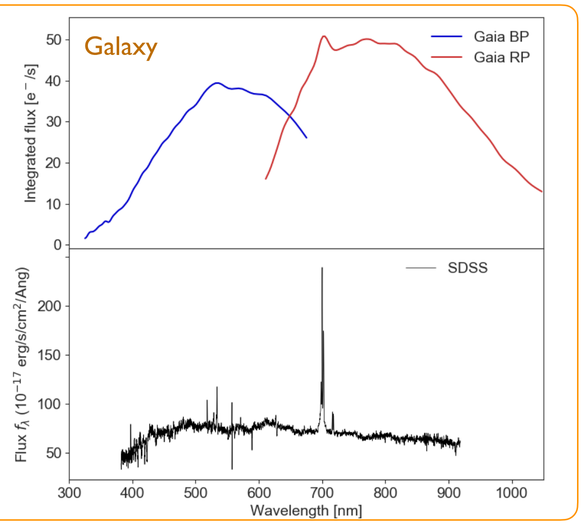


#STAR CLASS
Gaia DR3 148041342236265344
SDSS [specObjID 1414251003082991616, z=0.0,
class='STAR', subClass='K3']

#GALAXY CLASS
Gaia DR3 1251255712857474304
SDSS [specObjID 2988283284401711104, z=0.0267,
class='GALAXY', subClass='STARBURST']



#GALAXY CLASS
Gaia DR3 908924829134335232
SDSS [specObjID 854645203405924352, z=0.0666,
class='GALAXY', subClass='STARFORMING']



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 in GDR3

classification by P > 0.5

	Specmod		Allosmod		Combmod		Spec&Allos	
	compl.	purity	compl.	purity	compl.	purity	compl.	purity
quasar	0.409	0.248						
galaxy	0.831	0.402						
star	0.998	0.989						

DSC: Performance in GDR3

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				
galaxy	0.831	0.402	0.924	0.298				
star	0.998	0.989	0.998	1.000				

DSC: Performance in GDR3

classification by P > 0.5

classlabel_dsc

	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		
galaxy	0.831	0.402	0.924	0.298	0.936	0.219		
star	0.998	0.989	0.998	1.000	0.996	0.990		

DSC: Performance in GDR3

classification by P > 0.5

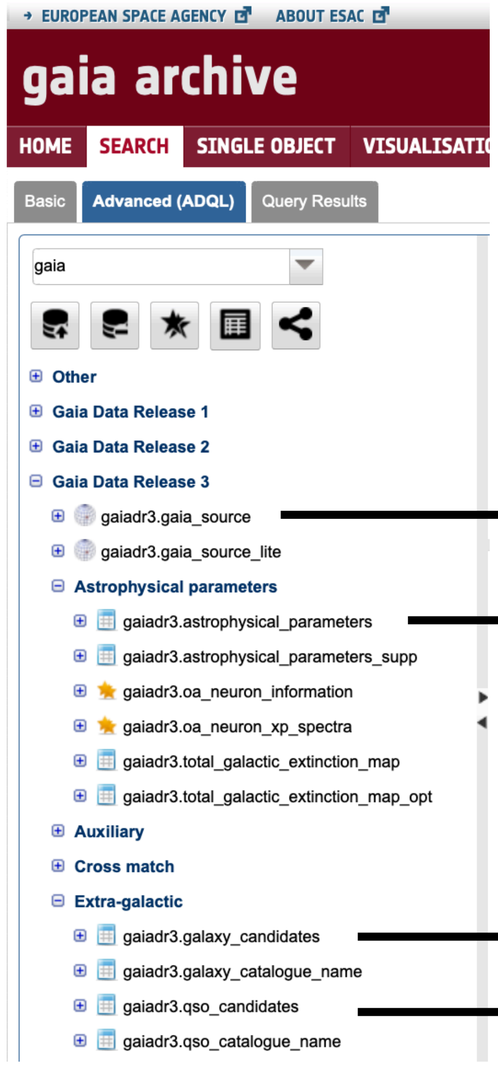
	classlabel_dsc				classlabel_dsc_joint			
	Specmod		Allosmod		Combmod		Spec&Allos	
	compl.	purity	compl.	purity	compl.	purity	compl.	purity
quasar	0.409	0.248	0.838	0.408	0.916	0.240	0.384	0.621
galaxy	0.831	0.402	0.924	0.298	0.936	0.219	0.826	0.638
star	0.998	0.989	0.998	1.000	0.996	0.990	–	–

DSC: Performance in GDR3

classification by P > 0.5

	classlabel_dsc						classlabel_dsc_joint	
	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

Where are the DSC results in the Gaia archive?



https://gea.esac.esa.int/archive/

Combmod probabilities; all sources

All probabilities and class labels; all sources

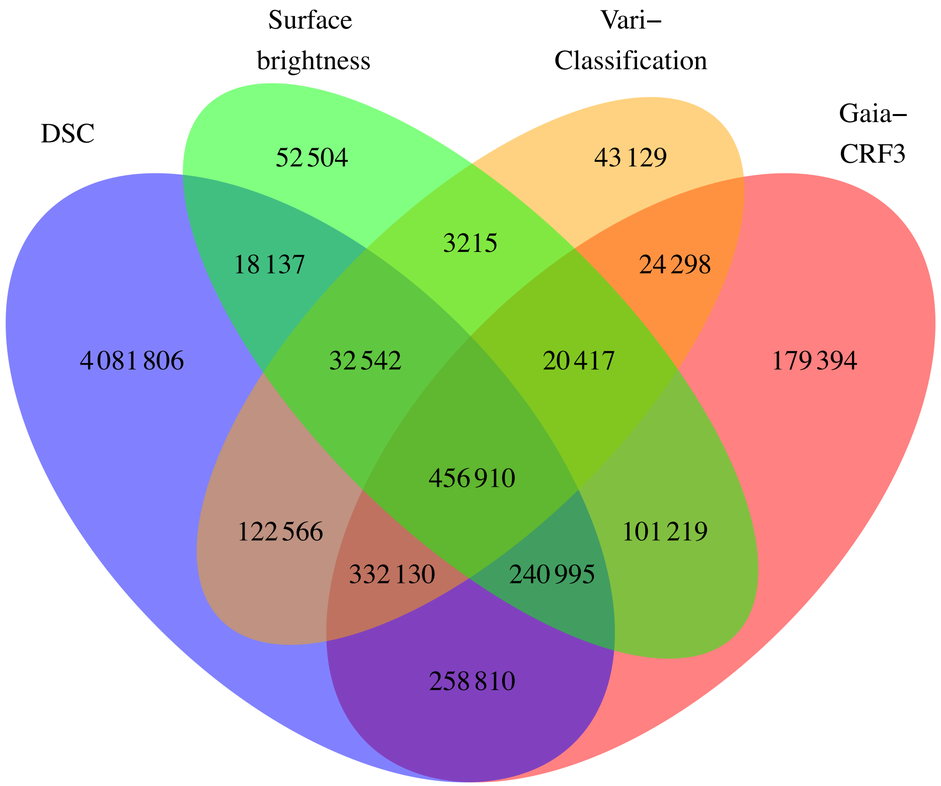
Combmod probabilities and class labels; galaxy candidates

Combmod probabilities and class labels; quasar candidates

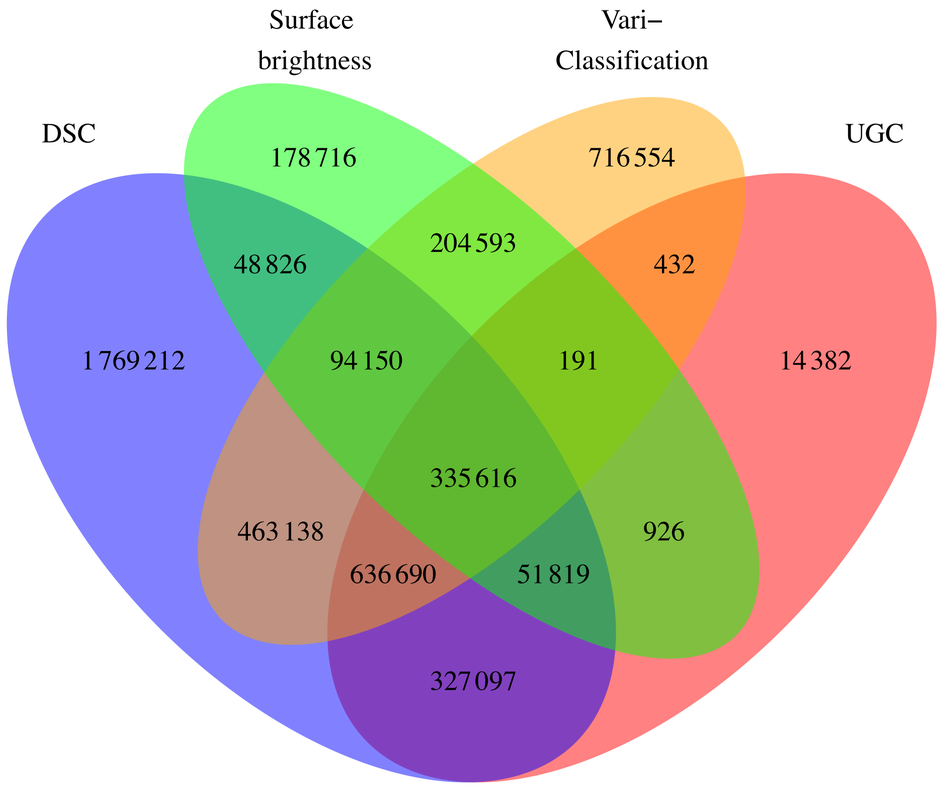
Extragalactic results in Gaia DR3

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)

How to select purer subsets in the extragalactic tables

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')
```

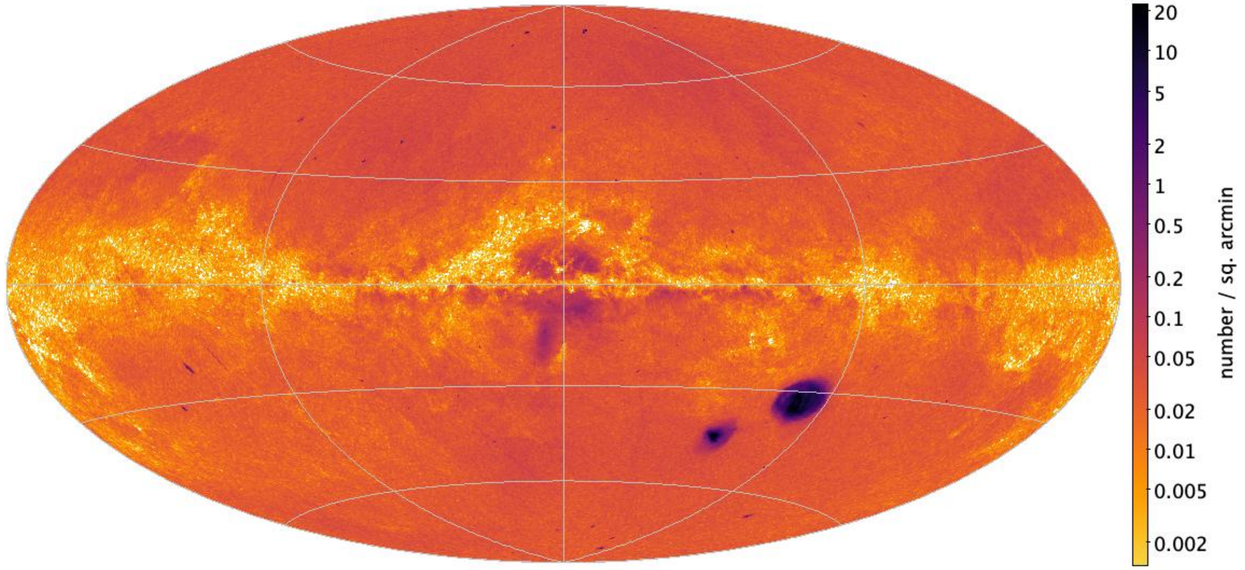
Table 12. ADQL query to select the purer galaxy sub-sample.

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

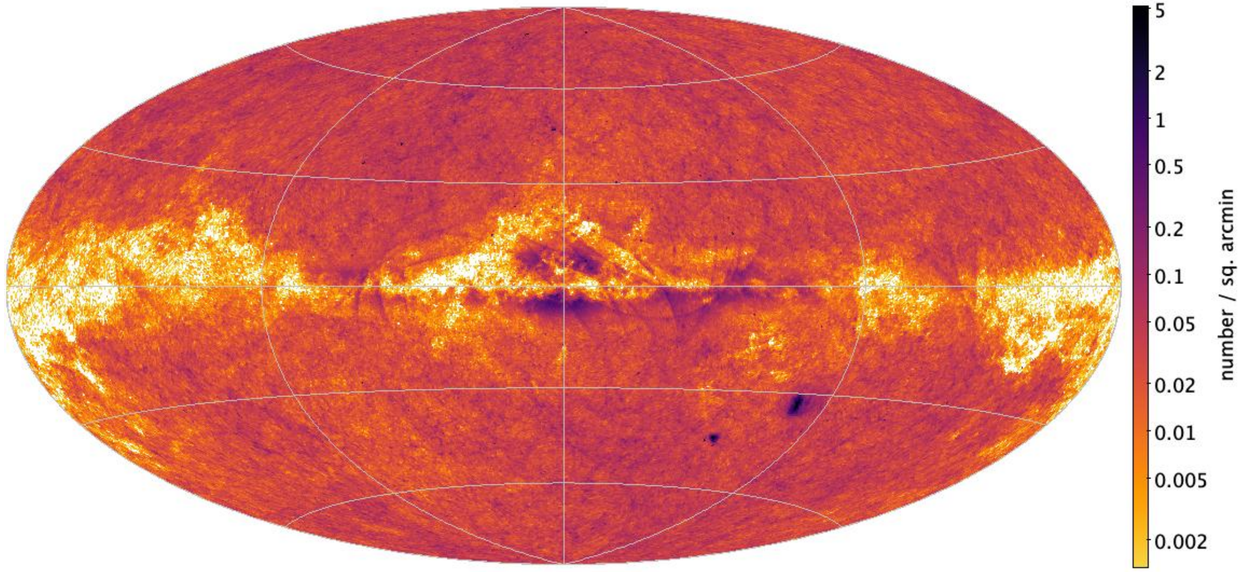
Sky distribution of extragalactic candidates

Full set

Quasar candidates
6.6 million, 52% pure



Galaxy candidates
4.8 million, 69% pure

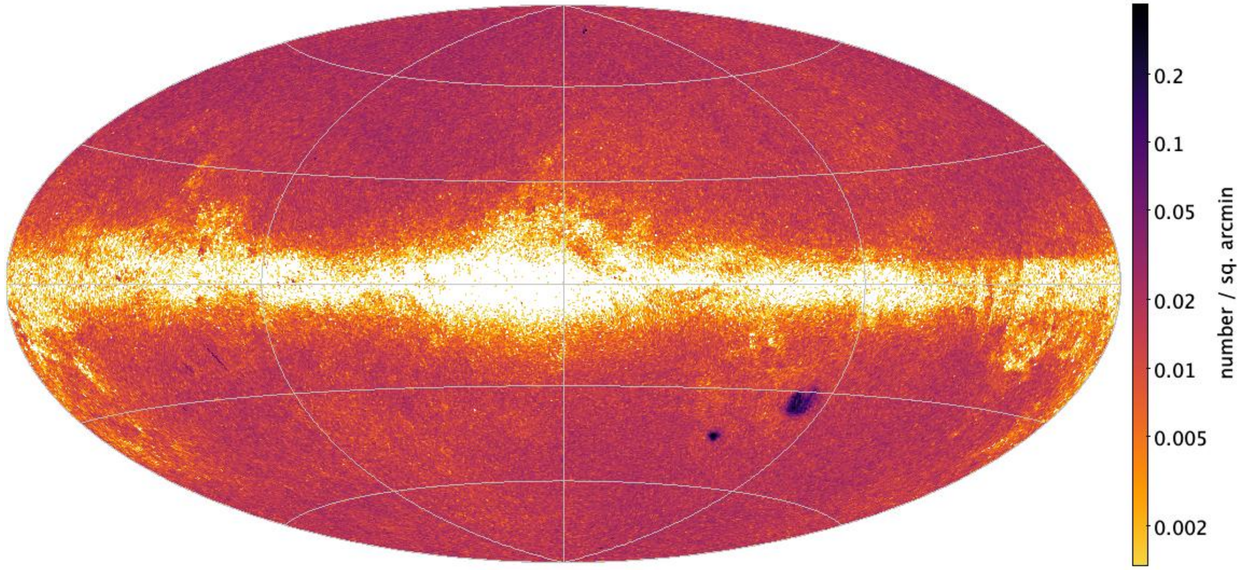


Hammer-Aitoff, Galactic coordinates, HEALpixel level 7 (0.21 sq. deg.)

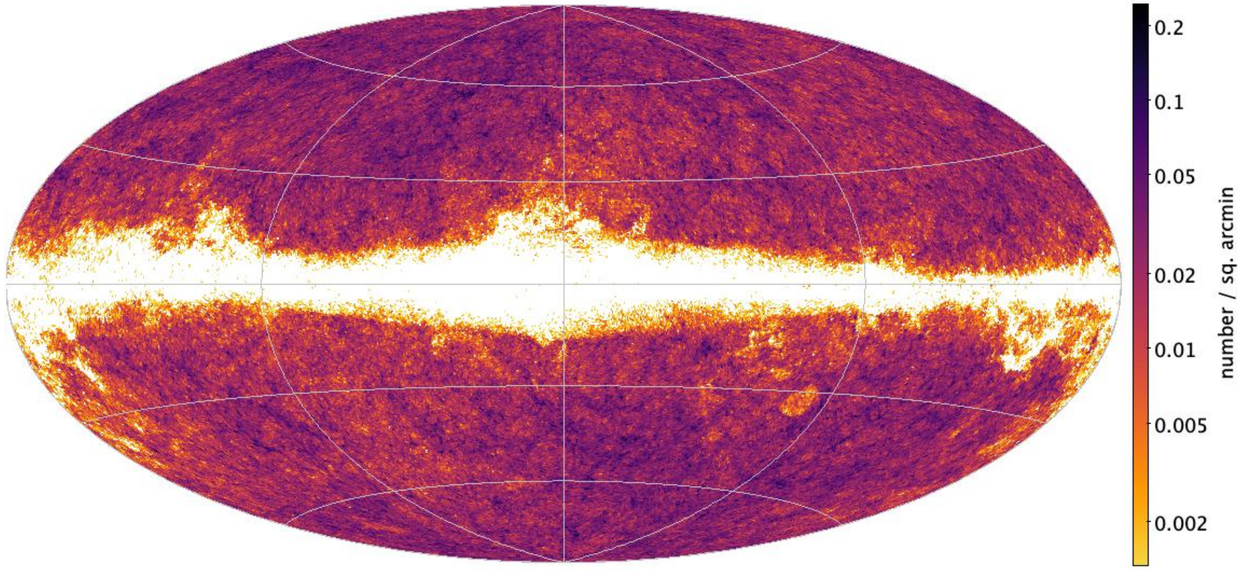
Sky distribution of extragalactic candidates

Purer subset

Quasar candidates
1.9 million, 95% pure



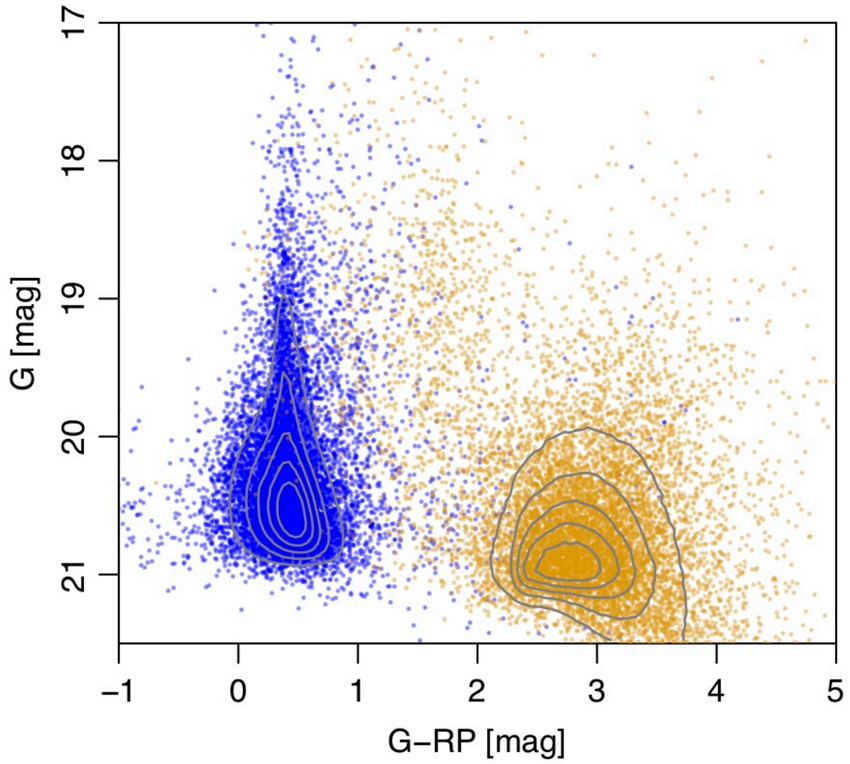
Galaxy candidates
2.9 million, 94% pure



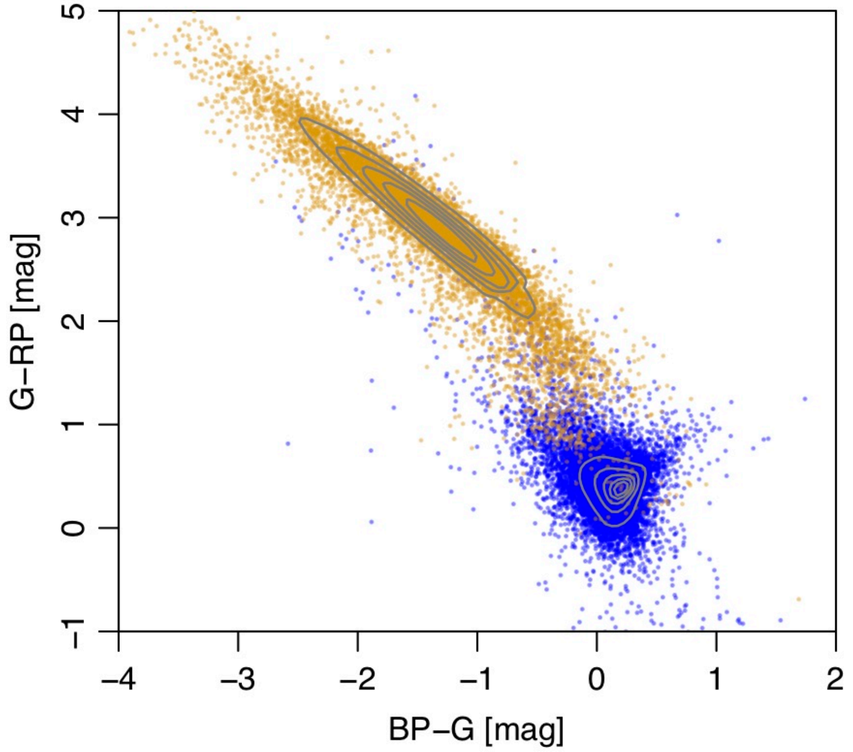
Hammer-Aitoff, Galactic coordinates, HEALpixel level 7 (0.21 sq. deg.)

Colour, magnitude distributions of extragalactic candidates

Full set



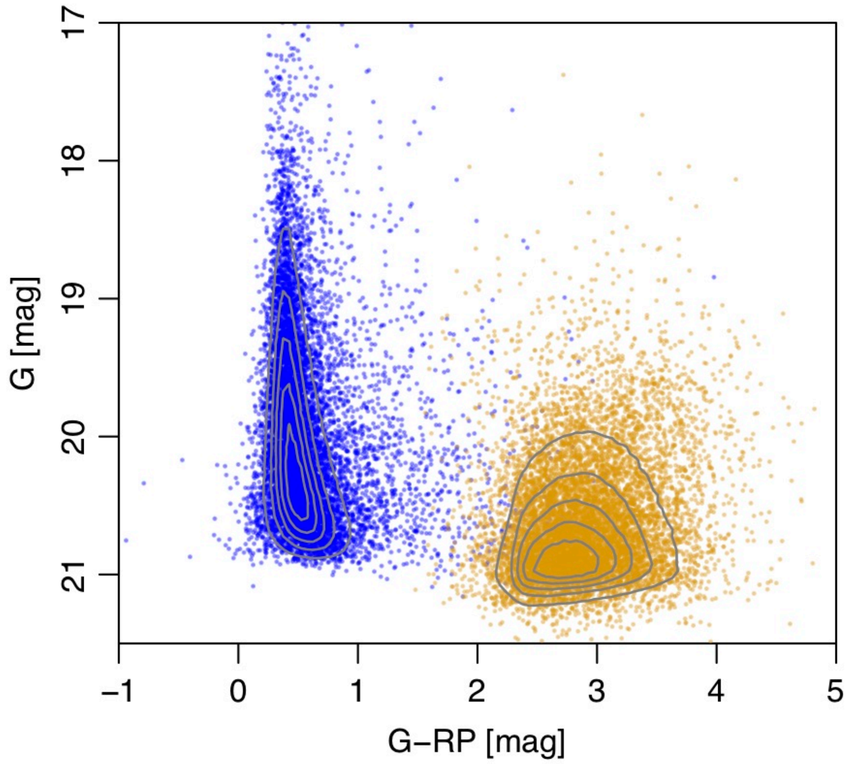
blue = quasars
orange = galaxies



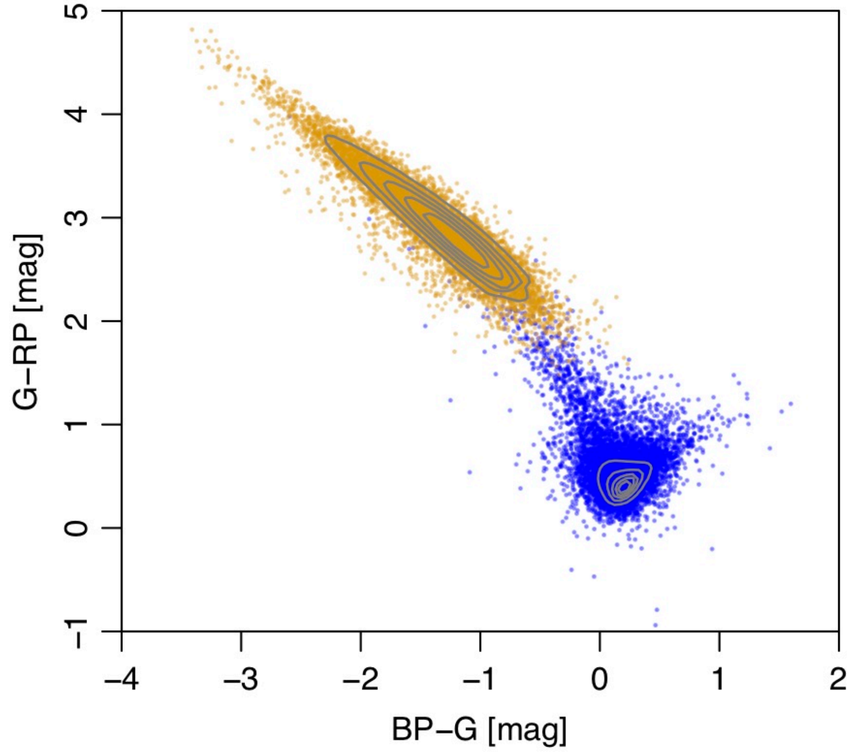
10 000 random sources of each class, linear density contours of full set

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

Planned improvements to DSC in GDR4

- Calibration of the posterior class probabilities.
- Algorithms, hyperparameters optimisation.
- Variable priors as function of brightness and galactic latitude

  [Hughes, Bailer-Jones and Jamal, A&A, 668, A99 (2022)]

- Rejection of contaminants in the training set.
- Evaluation of the classifier performances as function of brightness and galactic latitude.

Summary

- DSC classifies sources into three classes (star, quasar, galaxy) using BP/RP spectra, astrometry, colours, and variability information
- DSC extragalactic classes are determined by the SDSS training set
- DSC provides prior-adjusted probabilities
- DSC purity estimates take into account the rareness of extragalactic sources
- The GDR3 extragalactic candidate tables come from several modules, not just DSC
  - Can select a purer subset (~95% pure) using a simple ADQL query with:
  - 1.9 million quasar candidates, 2.9 million galaxy candidates