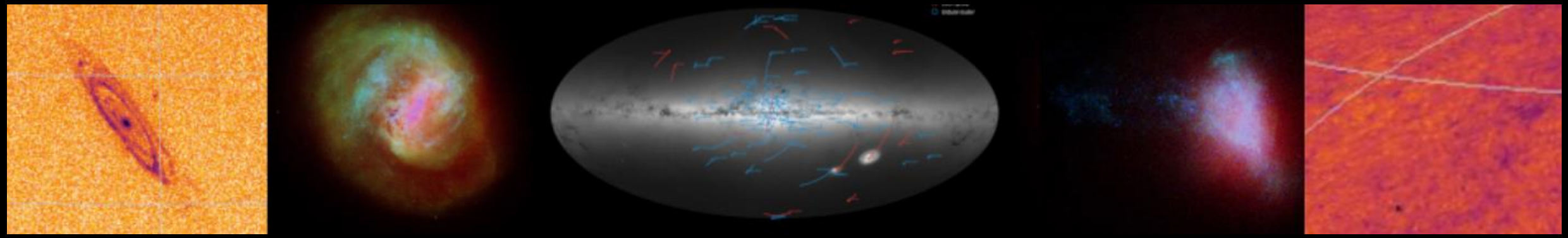


Redshifts of unresolved galaxies with Gaia

I. Bellas-Velidis (NOA)

D. Hatzidimitriou (UoA, NOA)

E. Livanou (UoA)

M. Kontizas (UoA)

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

--The photo of 2007

The Unresolved Galaxy Classifier - UGC

→ Extragalactic processing module of Coordination Unit CU8 (Astrophysical Parameters)

UGC
Input : BP/RP ***Output :*** Galaxy redshifts ***Number of sources :*** 1 367 153

Redshift estimator **redshift range $0 \leq z \leq 0.6$**

- Uses Support Vector Machines (SVM)
- Applies to sources with
  - DSC-Combmod galaxy probability ≥ 0.25
  - $13 \leq G \leq 21$ mag
  - BP/RP sampled mean spectra for which the flux is defined for all the samples (wavelength bins)

Main relevant DR3 publications in A&A

- **Gaia Collaboration, C.A.L. Bailer-Jones et al. 2022**, Gaia DR3 3: The extragalactic content
- **Delchambre et al. 2022**, Gaia DR3: Apsis III - Non-stellar content and source classification
- **Creevey et al. 2022**, Gaia DR33: Astrophysical parameters inference system (Apsis) I – methods and content overview
- **Online documentation** https://gea.esac.esa.int/archive/documentation/GDR3/index.html

DSC probabilities

- **The sources processed by UGC** are determined according to **classification probabilities** provided by the **Discrete Source Classifier (DSC)**

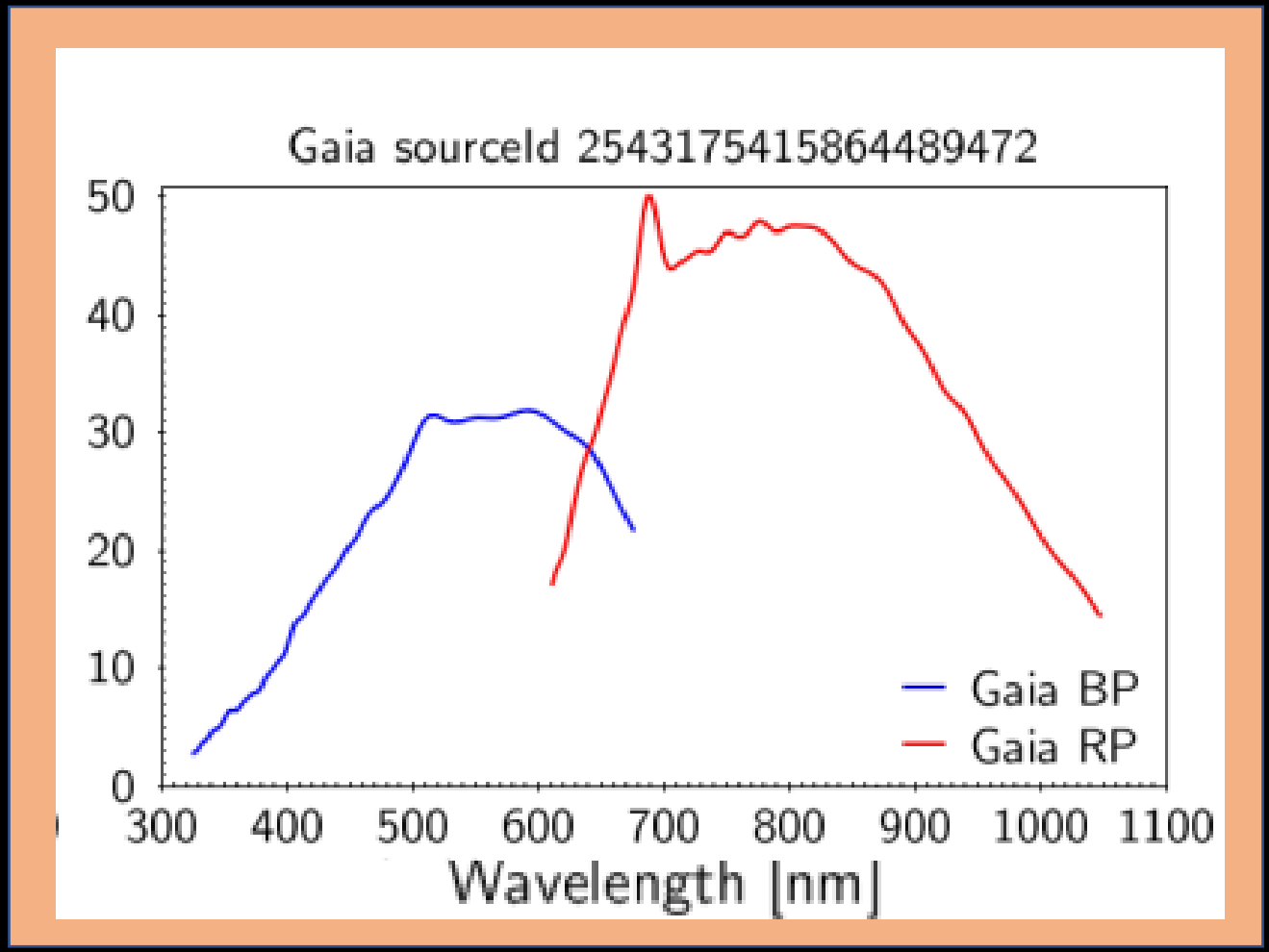
- **DSC** classifies sources probabilistically into **five classes** (quasar, galaxy, star, white dwarf, and physical binary star).

- DSC uses different three different classification methods, **Specmod**, **Allosmod** and **Combmod**, the latter combining the first two into a new posterior probability over all five classes.

- UGC processes sources that have a **DSC Combmod probability to be a galaxy ≥0.25**

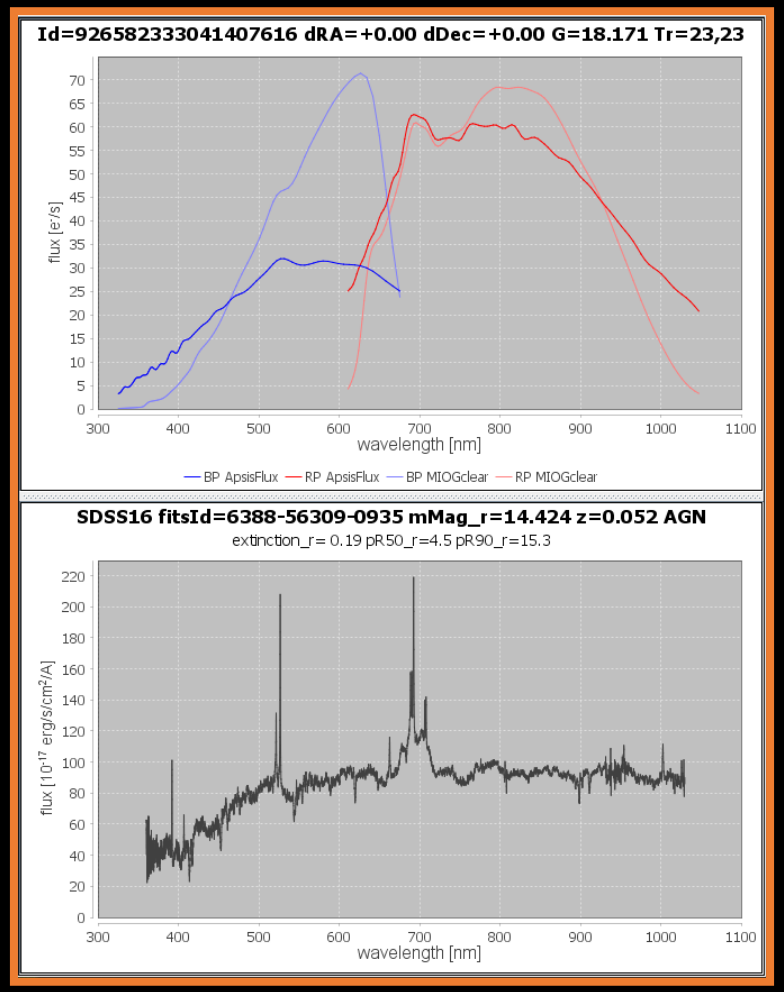
BP/RP Spectra of Galaxies

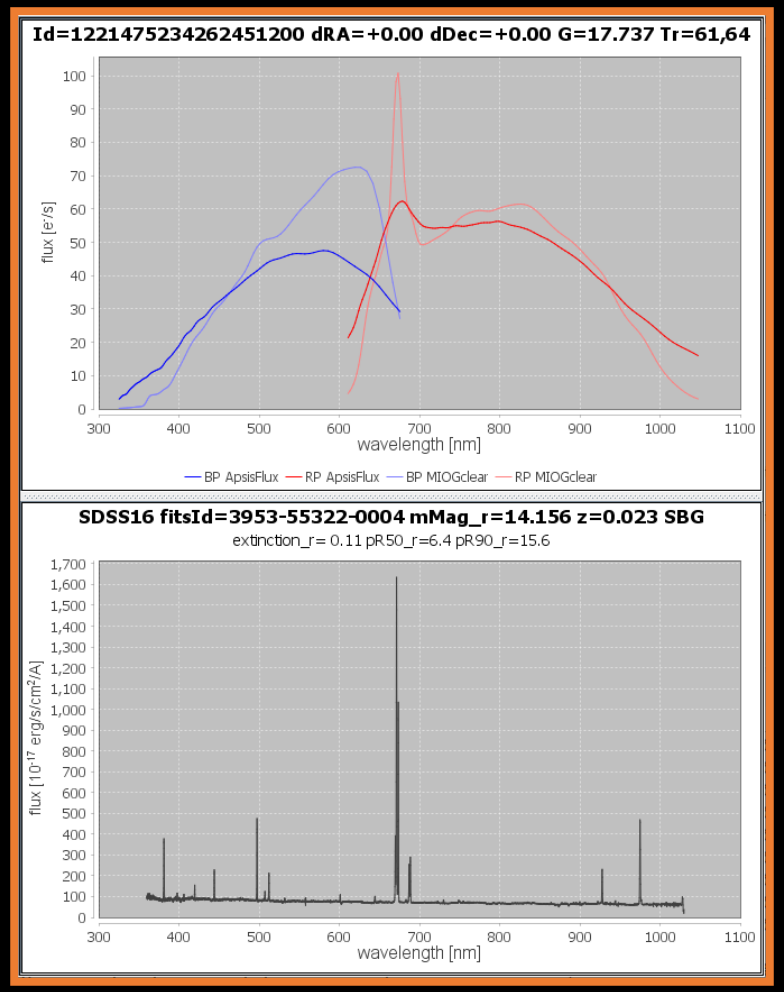
- The input of the SVM models are BP/RP mean spectra as sampled by SMSgen
- **Clipped edges of spectra (low S/N) → BP:366-627nm RP: 620-996nm**
- Each pair of truncated spectra is then concatenated to form the SVM input vector of **186 fluxes**.
- Total BP and RP fluxes recalculated for truncated wavelength ranges

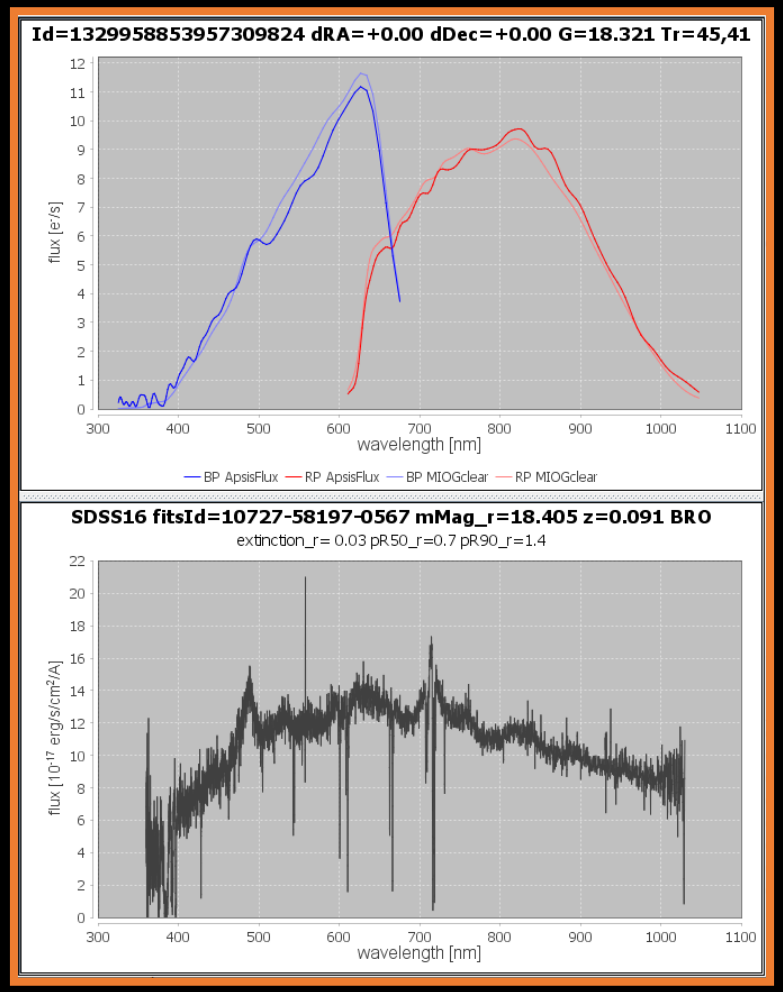


Examples of BP/RP Spectra of Galaxies

Comparison with SDSS DR16 spectra and MIOG simulated data of SDSS galaxy spectra





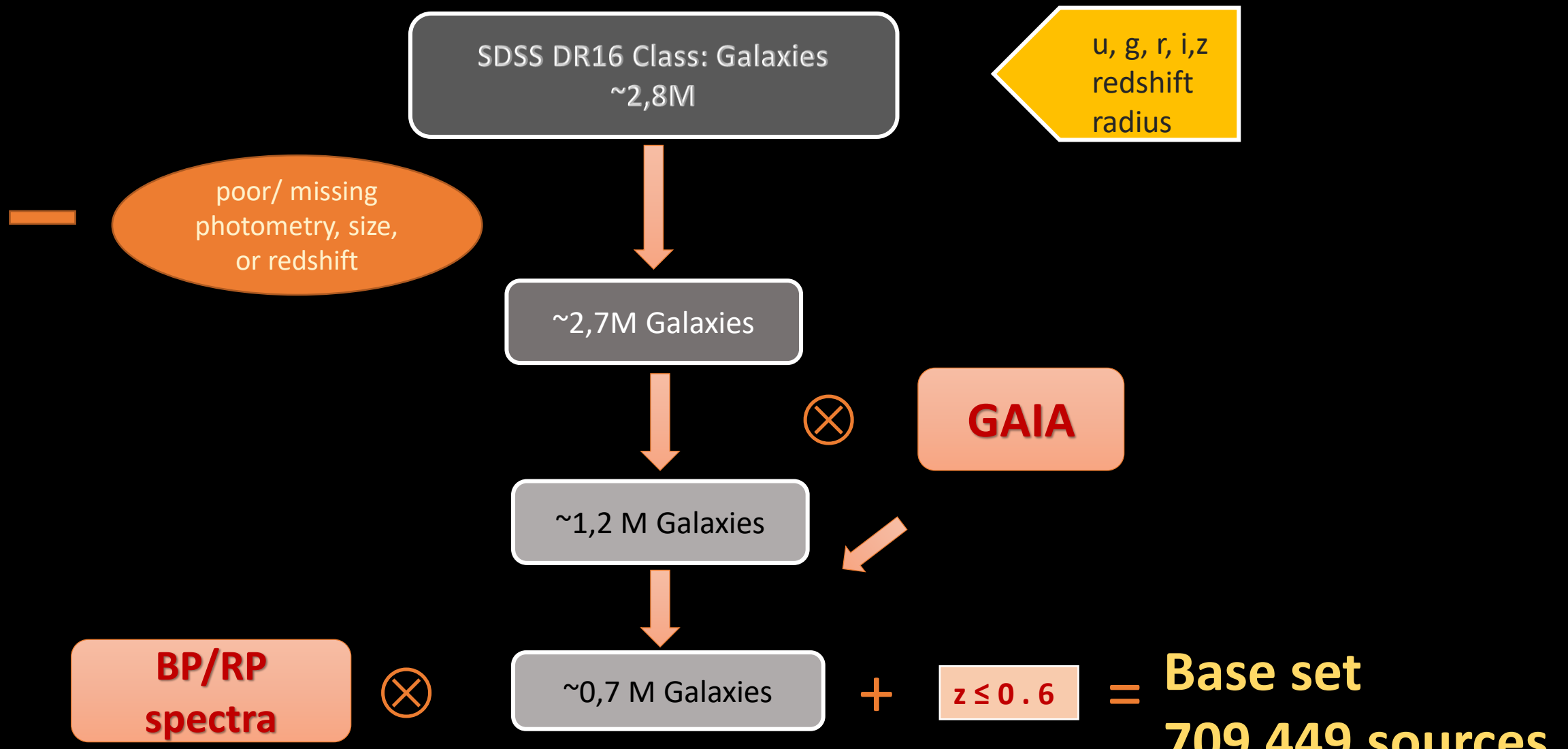


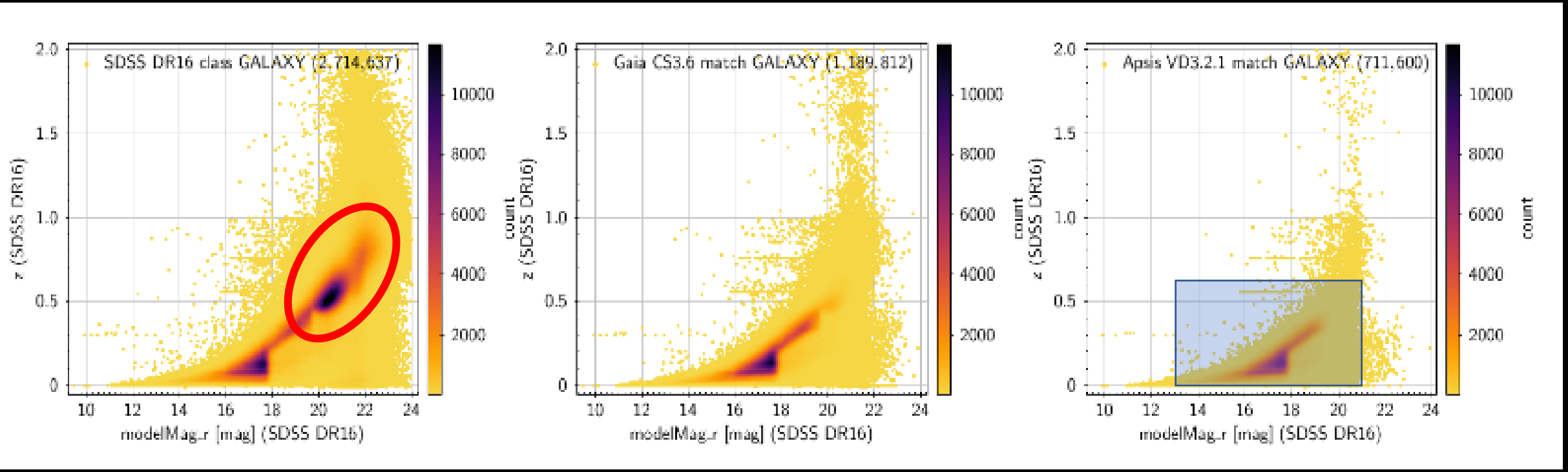
Support Vector Machines

- **based on the LIBSVM library of Chang & Lin (2011)**
- **three SVM models are built**:
  - **t-SVM**, the total-redshift range SVM model
    - → **computes the published redshift and associated prediction intervals**
  - **r-SVM**, a regression SVM model
  - **c-SVM**, a classification SVM model
    - → **applied to discretized redshifts**
    - → **used exclusively for internal validation of the redshift produced by the t-SVM model**

SVM Training and testing sets → The base set

SDSS DR16 archive (Ahumada et al. 2020)





SDSS DR16 redshift vs magnitude r

Subset cross-matched with Gaia

Subset with Apsis validated BP/RP spectra

Base set → Clean set → Training set

Base set → Clean set **380 000 galaxies**

- G ≤ 21 . 0 mag
- BP / RP spectra must be composed of a minimum of 6 epochs of observations
- 0 . 3 ≤ bpSpecFlux ≤ 100 and 0 . 5 ≤ rpSpecFlux ≤ 200
- imagesize (Petrosian radius) 0 . 5≤ petroRad50_r ≤ 5
- the interstellar extinction in the r-band below 0 . 5 mag to avoid highly reddened sources
- redshift must be larger than 0 . 01 in order to exclude nearby extended galaxies

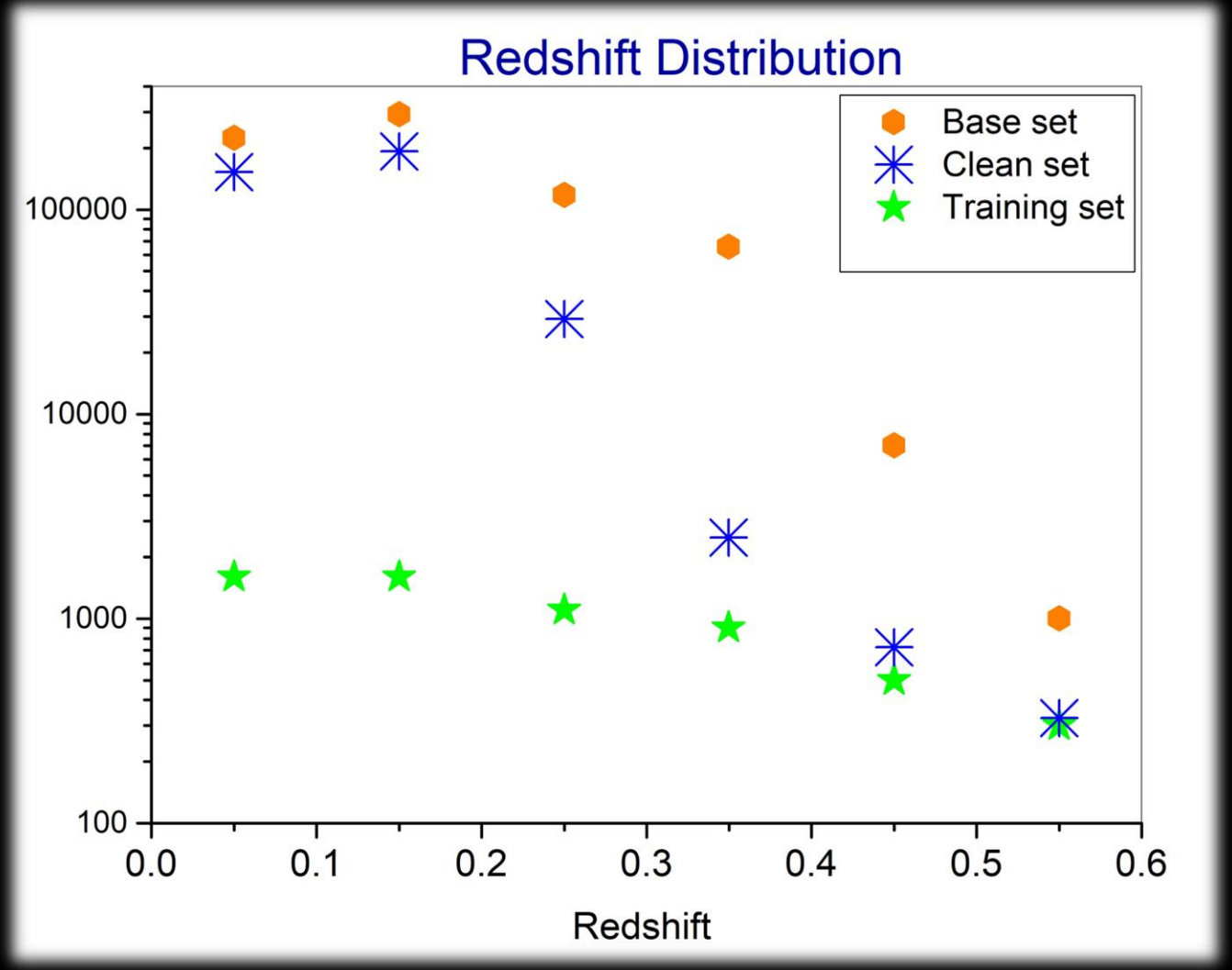
Clean set → Training set **6 000 galaxies**

randomly selected from clean set

Testing sets

- **Base test set = base set – Training set (703 449 galaxies)**
- **Clean test set = clean set – Training set (371 875 galaxies)**

Training set redshift distribution inbalance



UGC Output

Extragalactic Table: galaxy_candidates

- **The output is a Table which includes three fields**
  - ✓ the redshift value
  - ✓ the redshift upper prediction limit
  - ✓ the redshift lower prediction limit

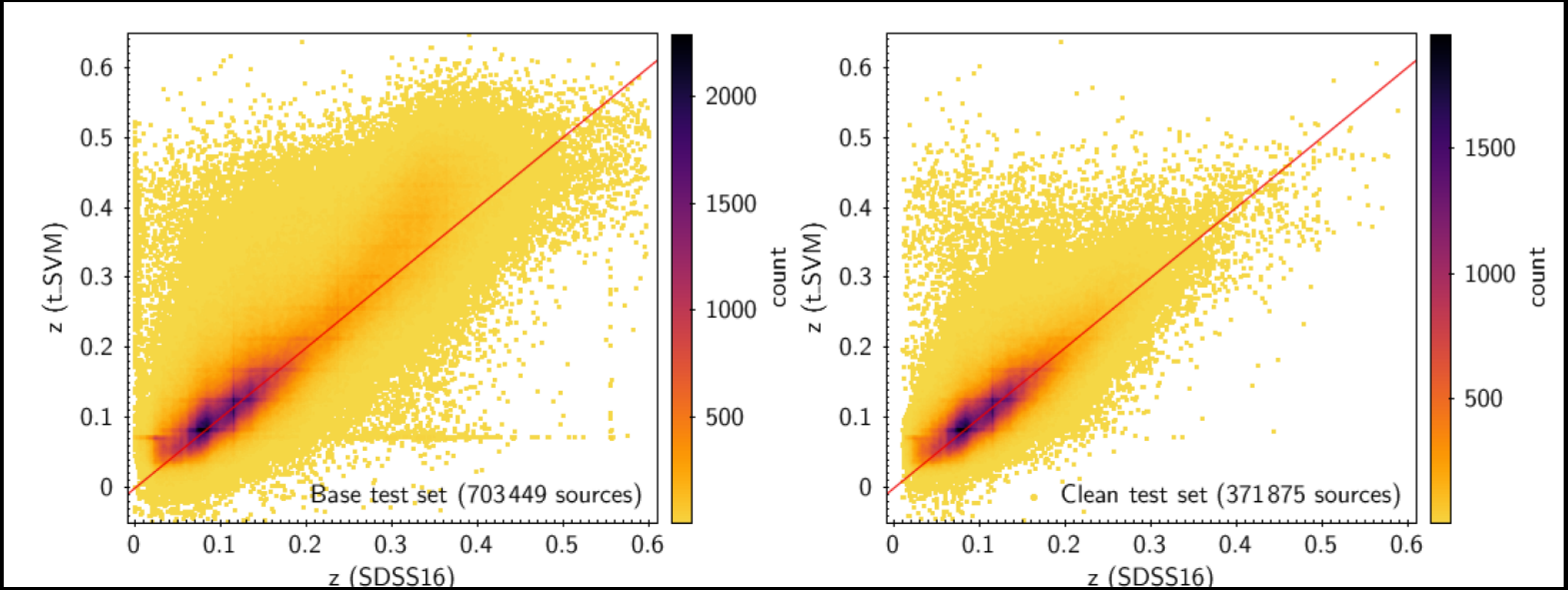
- **How are the prediction limits calculated**

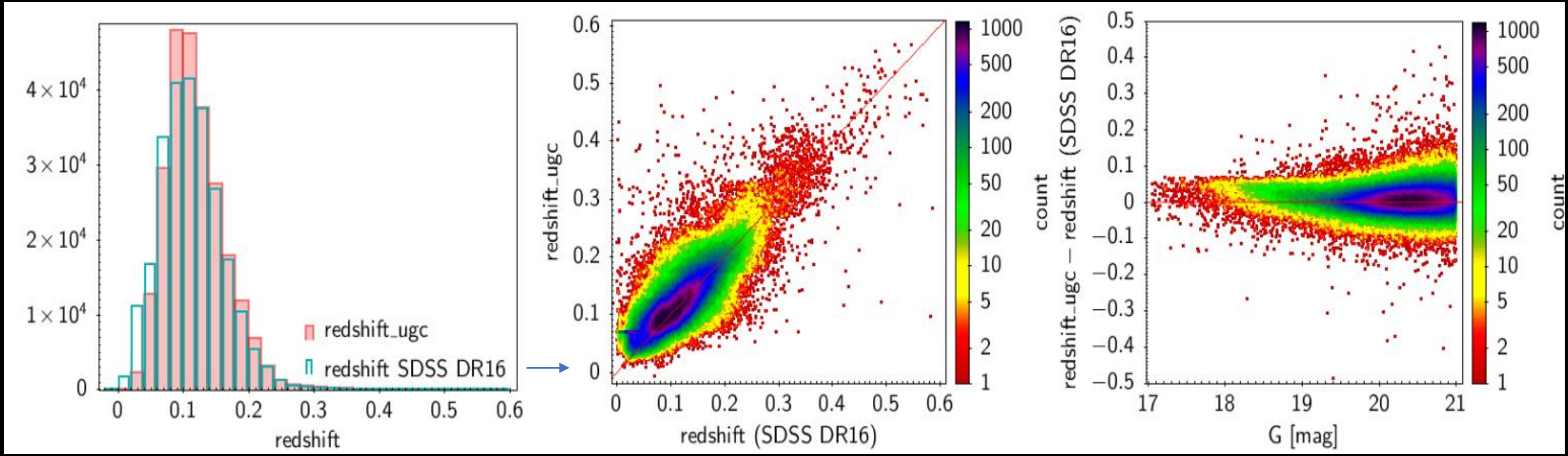
  Using the Test Set to calculate, in redshift bins of 0.02 size, the difference (bias) and rms of the difference between the SDSS and UGC redshifts.

  - ✓ $r_i = int($ redshift_ugc/0.02)
  - ✓ *redshift_ugc_lower* = redshift_ugc − predictDiffMean $[r_i]$ − predictDiffStDev $[r_i]$
  - ✓ *redshift_ugc_upper* = redshift_ugc − predictDiffMean $[r_i]$ + predictDiffStDev $[r_i]$

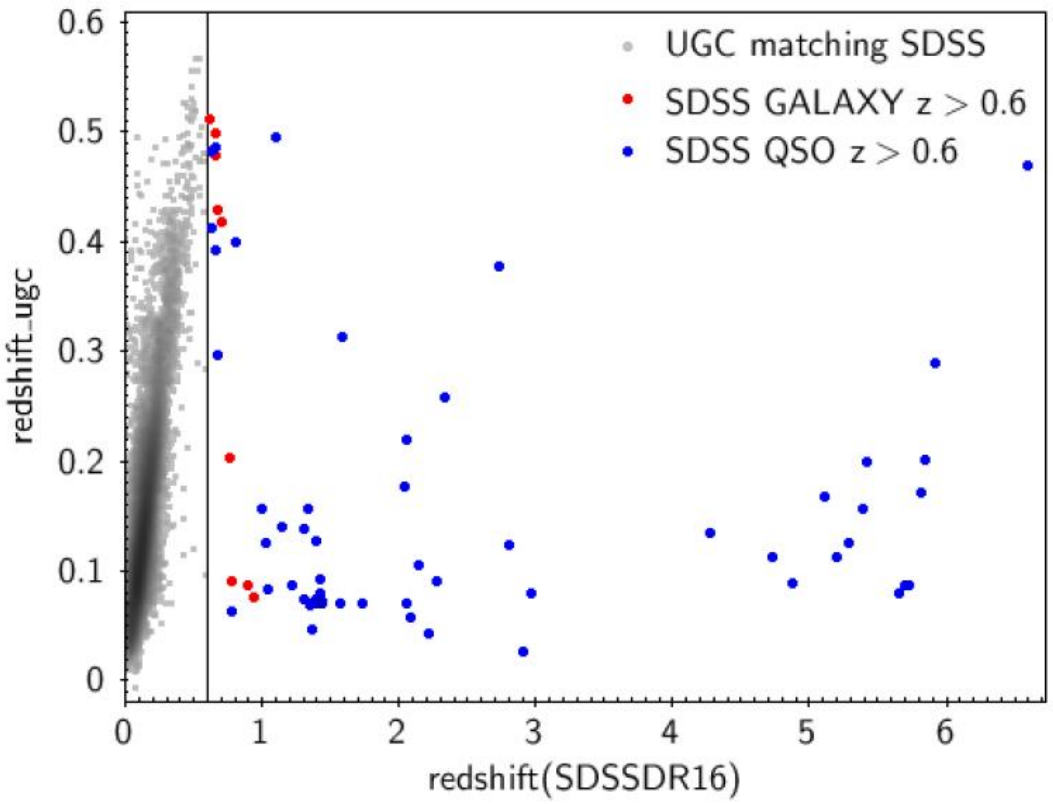
✓ The value of (redshift_ugc_upper−redshift_ugc_lower)/2 can be used as an estimate of the uncertainty in redshift_ugc.

UGC Performance
Comparison with SDSS DR16

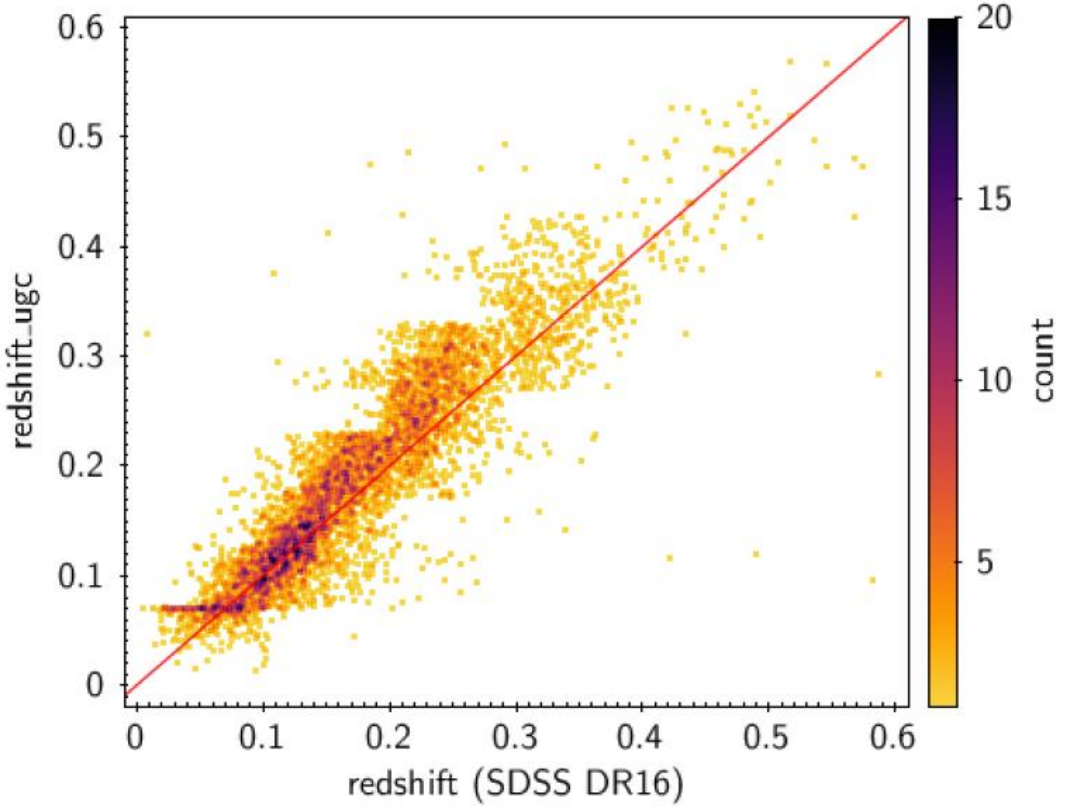




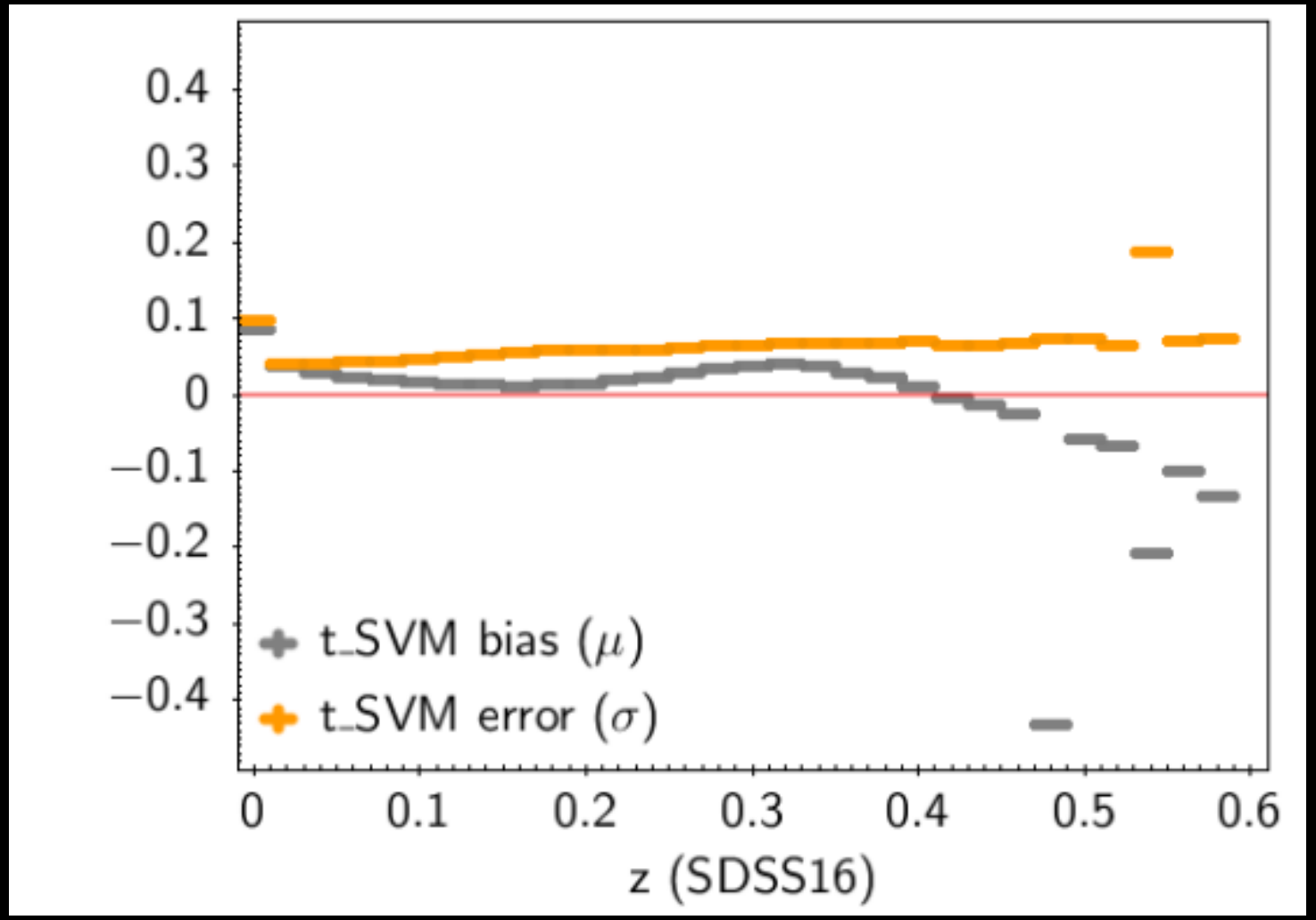
- UGC tends to overestimate low redshifts
- Artificial feature at z=0.07
  - ~17 000 sources in the bin 0 . 070 < redshift_ugc < 0 . 071
  - Most of these sources are bright and have SDSS redshifts below 0.04



UGC sources with high redshift from the SDSS DR16. Blue and red points are sources that are spectroscopically classified as 'QSO' and 'GALAXY' in the SDSS DR16, respectively.



. Comparison of the UGC redshifts for sources classified as 'QSO' in the SDSS DR16, with actual redshift lower than 0.6.

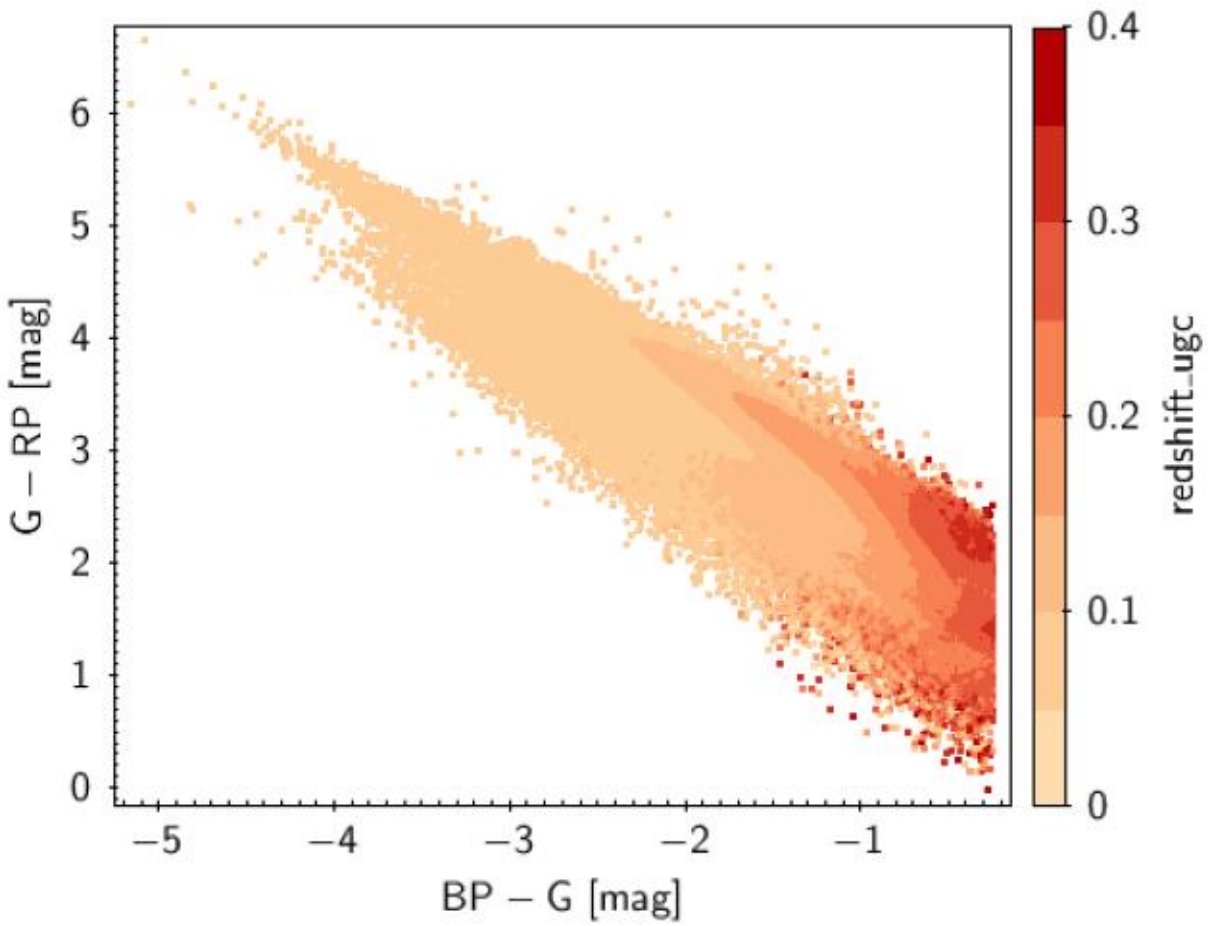


For the entire redshift range $0.0 \leq z \leq 0.6$
Bias in redshift -0.006 with rms 0.039

Validation - acceptability filters for "publishable" output

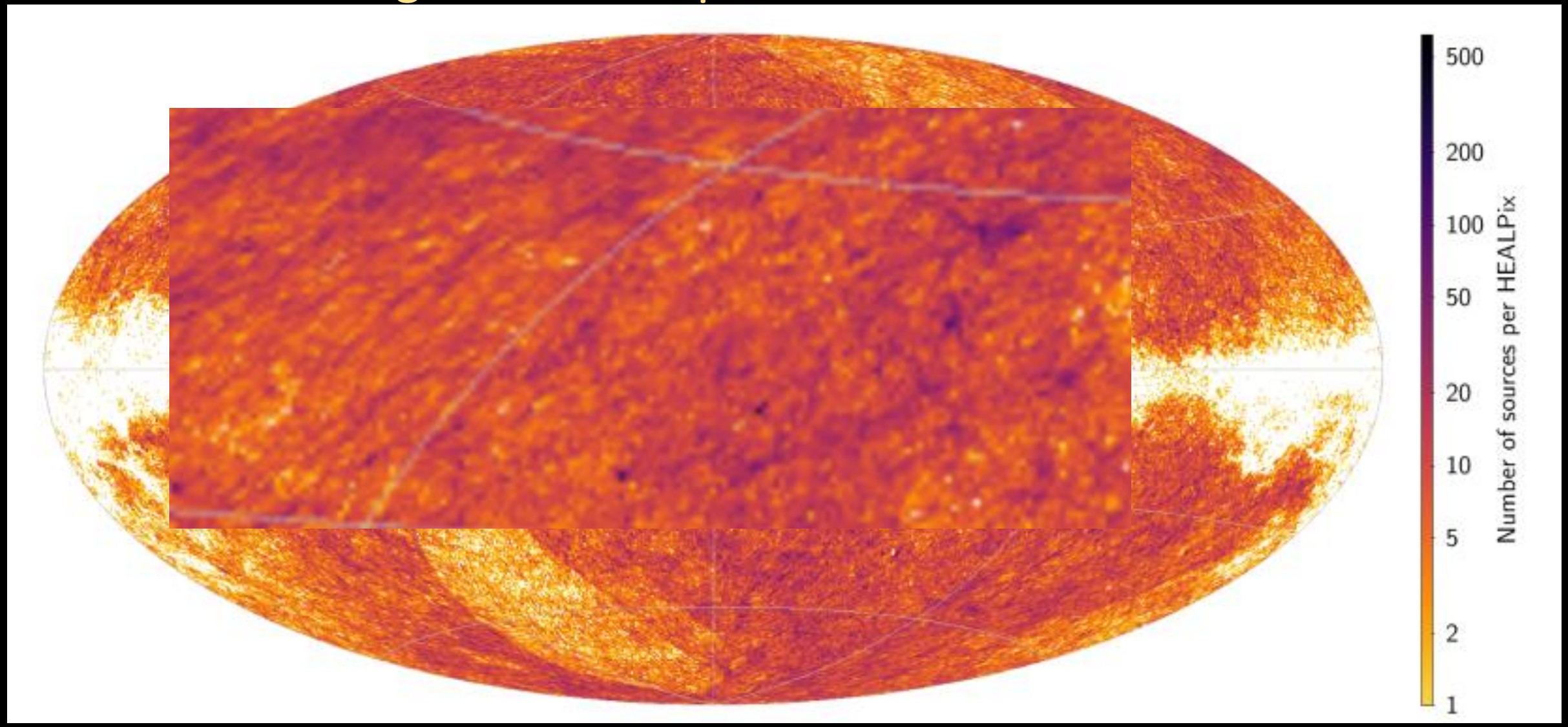
- Validation – Colour filtering
- Spatial filtering (crowded regions)

Area	Galactic coordinates range longitude [°]	latitude [°]	Colour-colour box A [mag]	Colour-colour box B [mag]
CNT	0.0 ± 15.0	-5.0 ± 5.0	$-0.5 < G - G_{BP} < 0.5$ $0.4 < G_{BP} - G_{RP} < 1.3$	$-0.5 < G - G_{BP} < 3.0$ $-0.2 < G_{BP} - G_{RP} < 1.4$
LMC	279.5 ± 4.0	-33.25 ± 3.25	$-3.0 < G - G_{BP} < -1.5$ $-0.4 < G_{BP} - G_{RP} < 1.0$	$-0.7 < G - G_{BP} < 2.0$ $-0.8 < G_{BP} - G_{RP} < 1.4$
SMC	303.0 ± 1.0	-44.0 ± 1.0	$-3.0 < G - G_{BP} < -1.5$ $-0.4 < G_{BP} - G_{RP} < 1.0$	$-0.7 < G - G_{BP} < 2.0$ $-0.8 < G_{BP} - G_{RP} < 1.4$



. Colour-colour diagram for the 1 367 153 galaxies for which redshifts are provided by UGC, colour-coded by redshift. A small number of sources have redshifts extending up to 0.6

Distribution of galaxies with published redshifts from UGC



Gaia DR3: Apsis III - Non-stellar content and source classification

Future data releases

- The original UGC module was designed to provide several parameters in addition to redshift (extinction, star formation rate etc).
- The basis for this was a synthetic library of galaxy spectra, simulated with the updated Gaia data model.
- For Gaia DR4 the effort is to use SDSS classes (probably only 2 or 3) to train the SVMs and classify the observed BP/RP spectra
- For Gaia DR5 the effort is to revisit the pproach of the synthetic spectra, with the aim to produce an improved semi-empirical library.

(see next presentation by P. Patsonis)