Object classification for SDSS DR12

F. Habibi, M. Moniez & R. Ansari

LSST-France
LPNHE Mars 2017

Aim

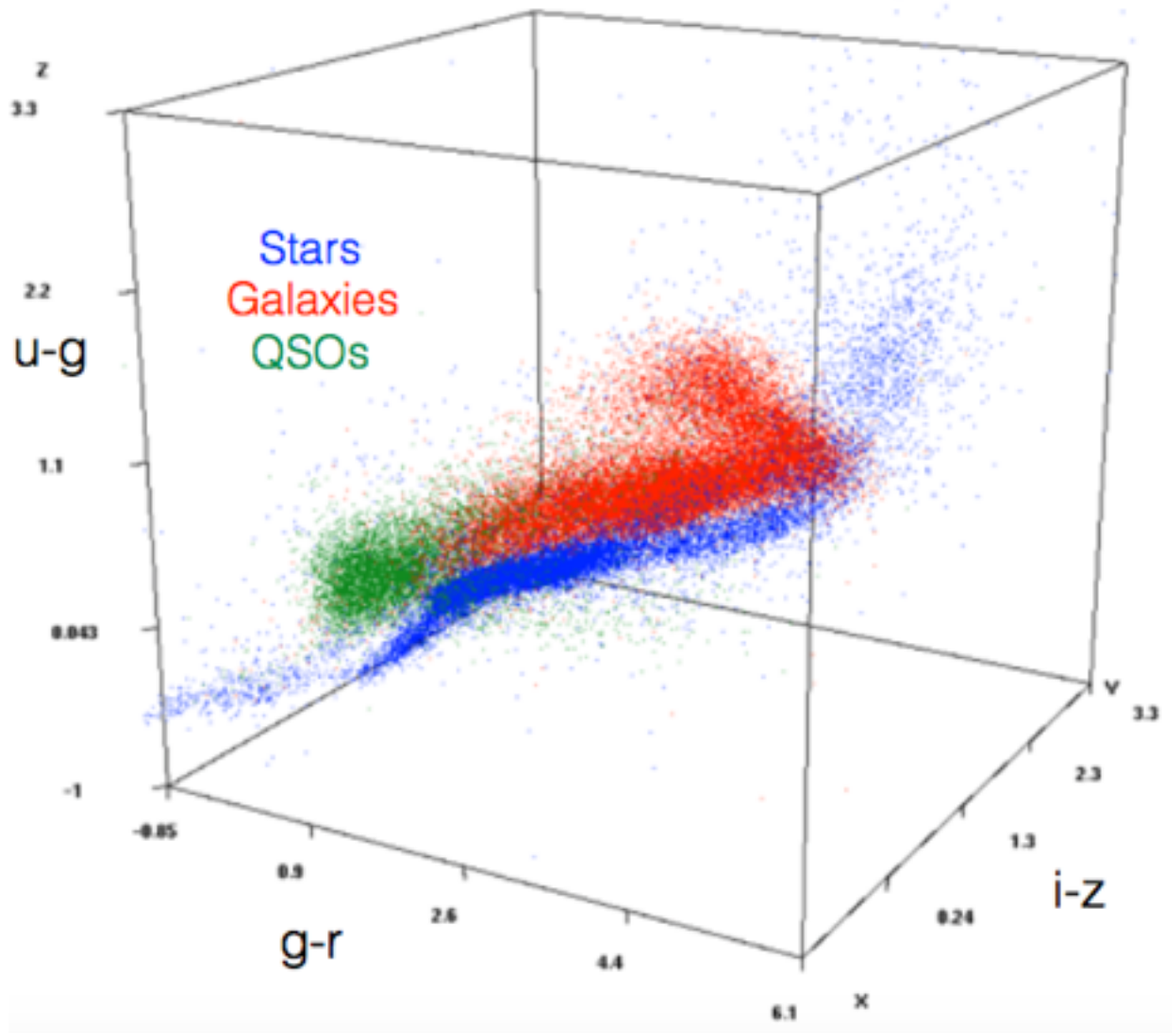
In the lack of spectroscopic data,
to separate galaxies from stars and QSOs

How?

Including all possible photometric information

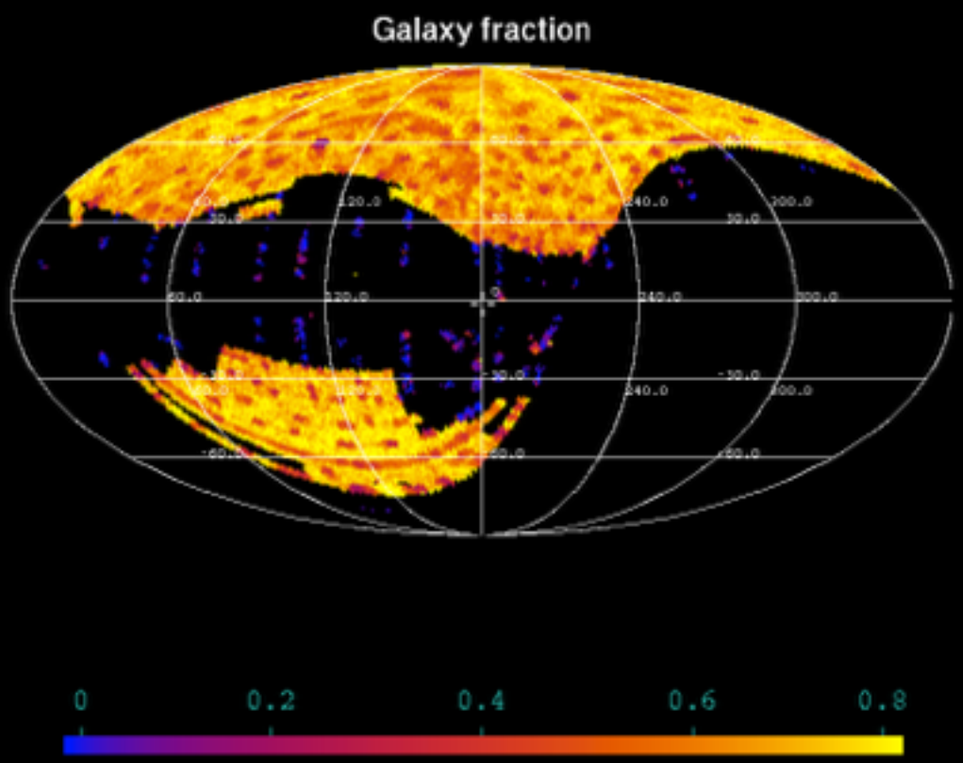
Colour indices:

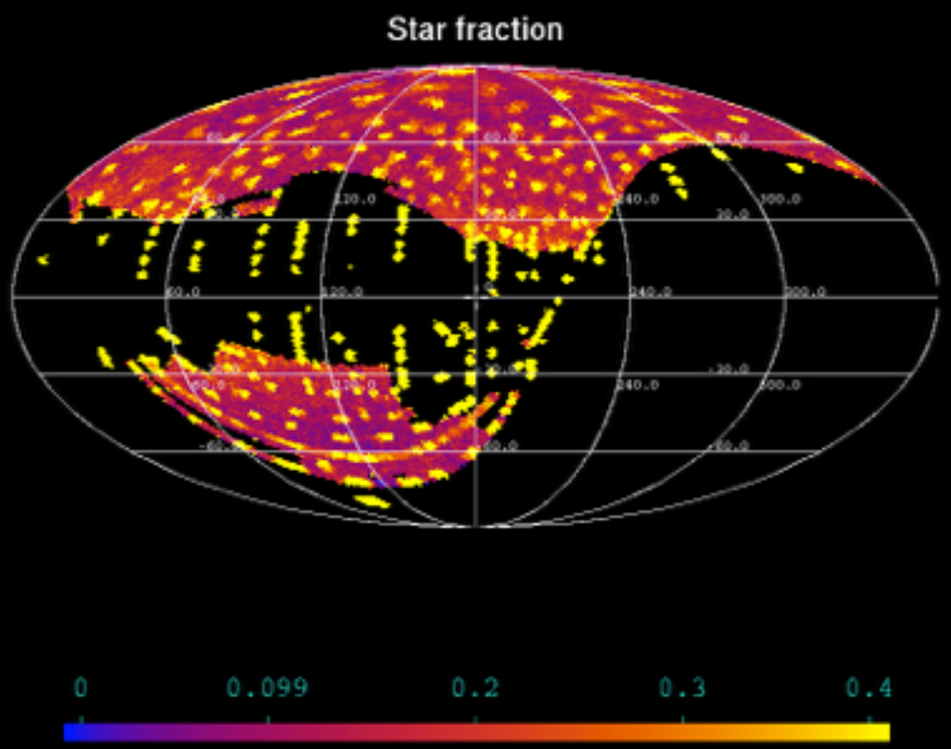
proper "features" for supervised perceptrons

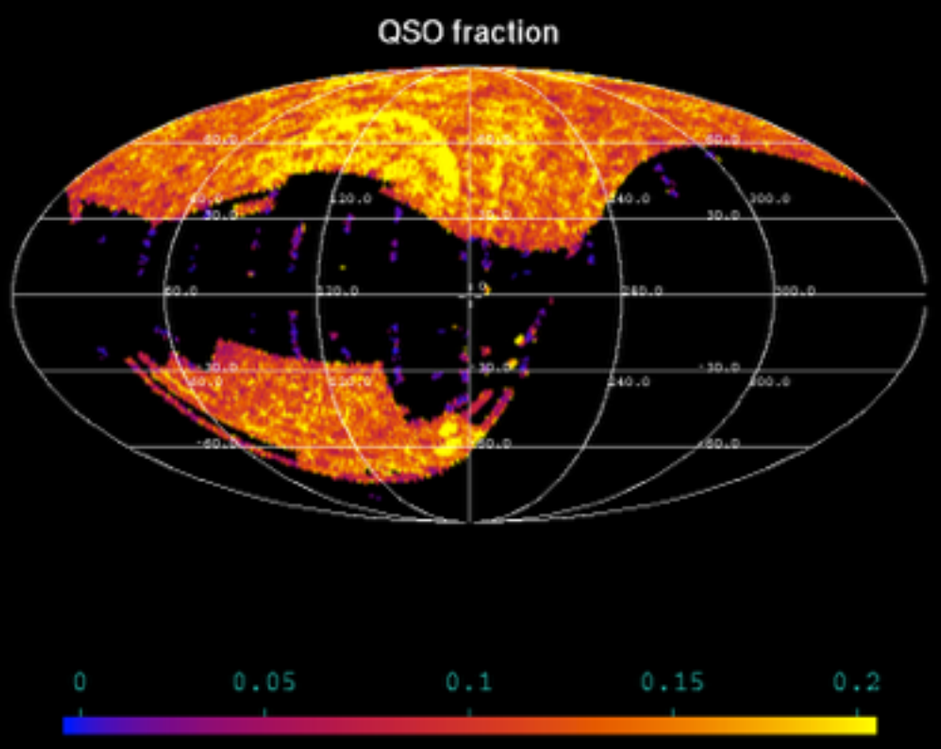


Using automatic classification for big number of data

SDSS DR12 data







Different sub-surveys

Objects selected with both photometric and spectroscopic data available

Stars	Galaxies	QSOs	Total
928,464	2,484,161	566,475	3,979,100
23%	62%	15%	

SDSS DR12 photometry

- PSF magnitude
- Model magnitude: de Vaucouleurs / exponentiel profile

$$I(r) = I_0 \exp\{-7.67\,[(r/r_e)^{1/4}]\}$$

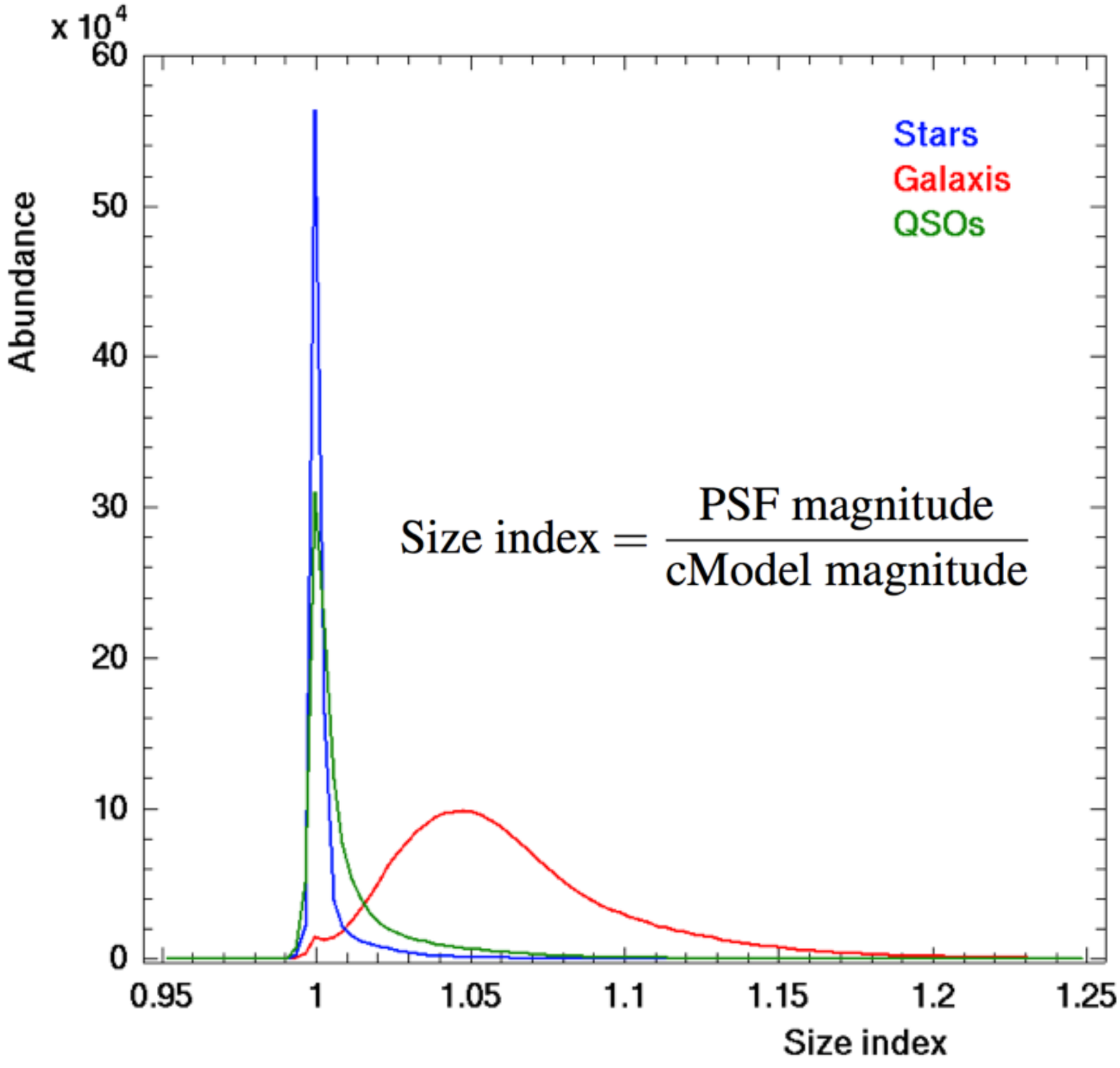
$$I(r) = I_0 \exp(-1.68 r/r_e)$$

- Composite model magnitude:

$$F_{\mathtt{composite}} = \mathtt{fracDeV}\ F_{\mathtt{dev}} + (1 - \mathtt{fracDeV})\ F_{\mathtt{exp}}$$

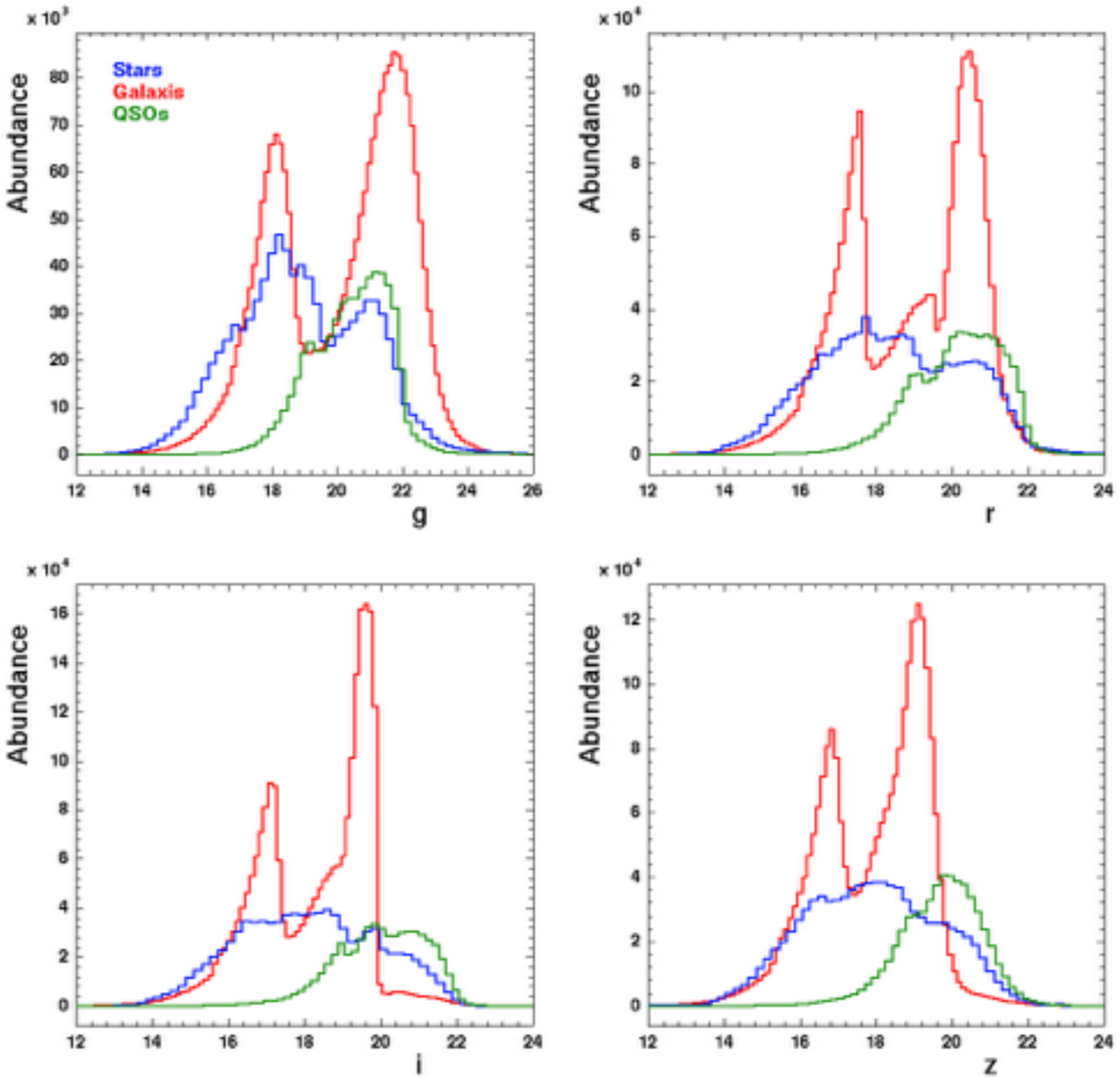
$$\text{Size index} = \frac{\text{PSF magnitude}}{\text{cModel magnitude}}$$

SDSS DR12 data



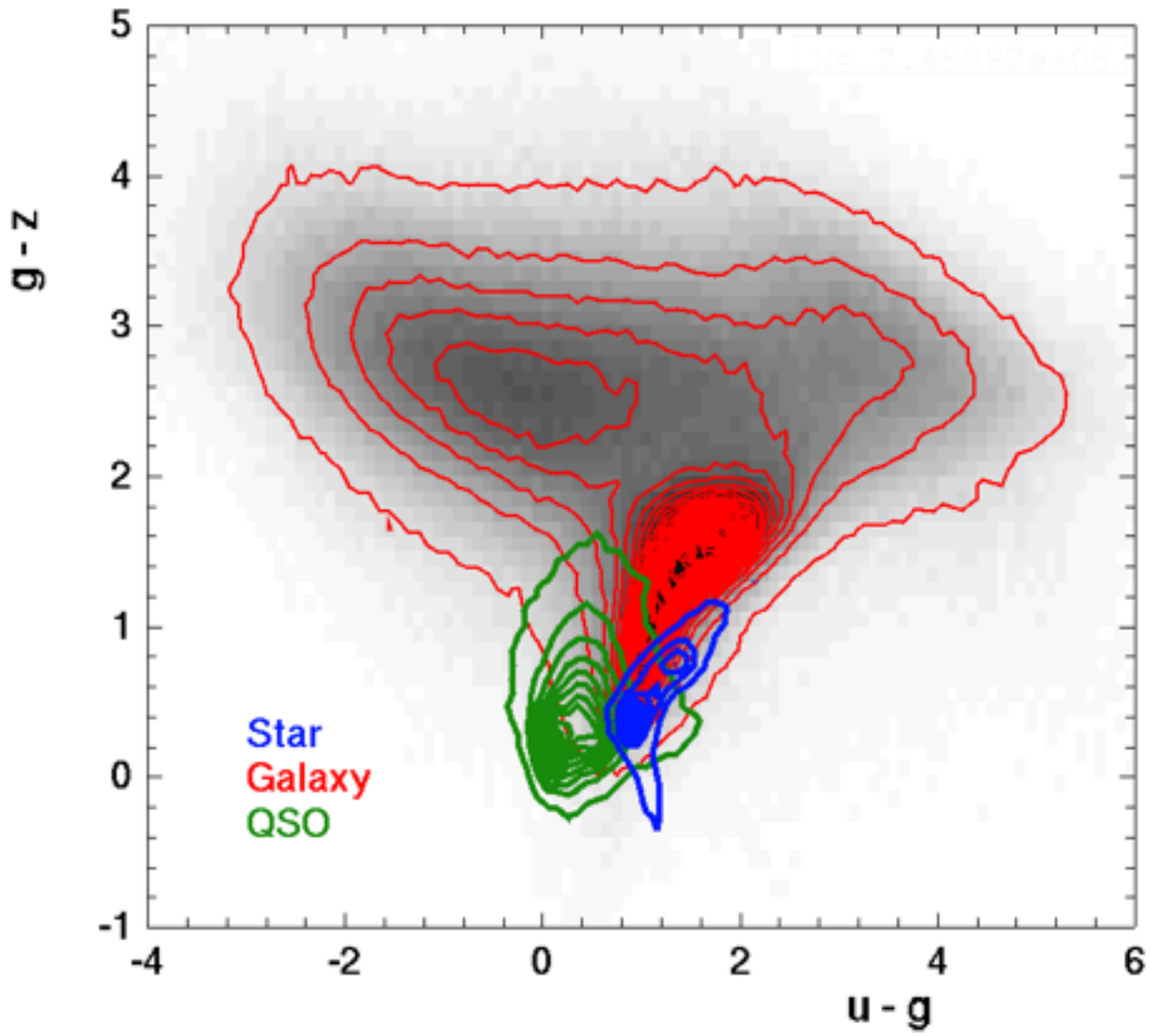
SDSS DR12 data

Magnitude distributions



Magnitudes:
insufficient to separate the objects

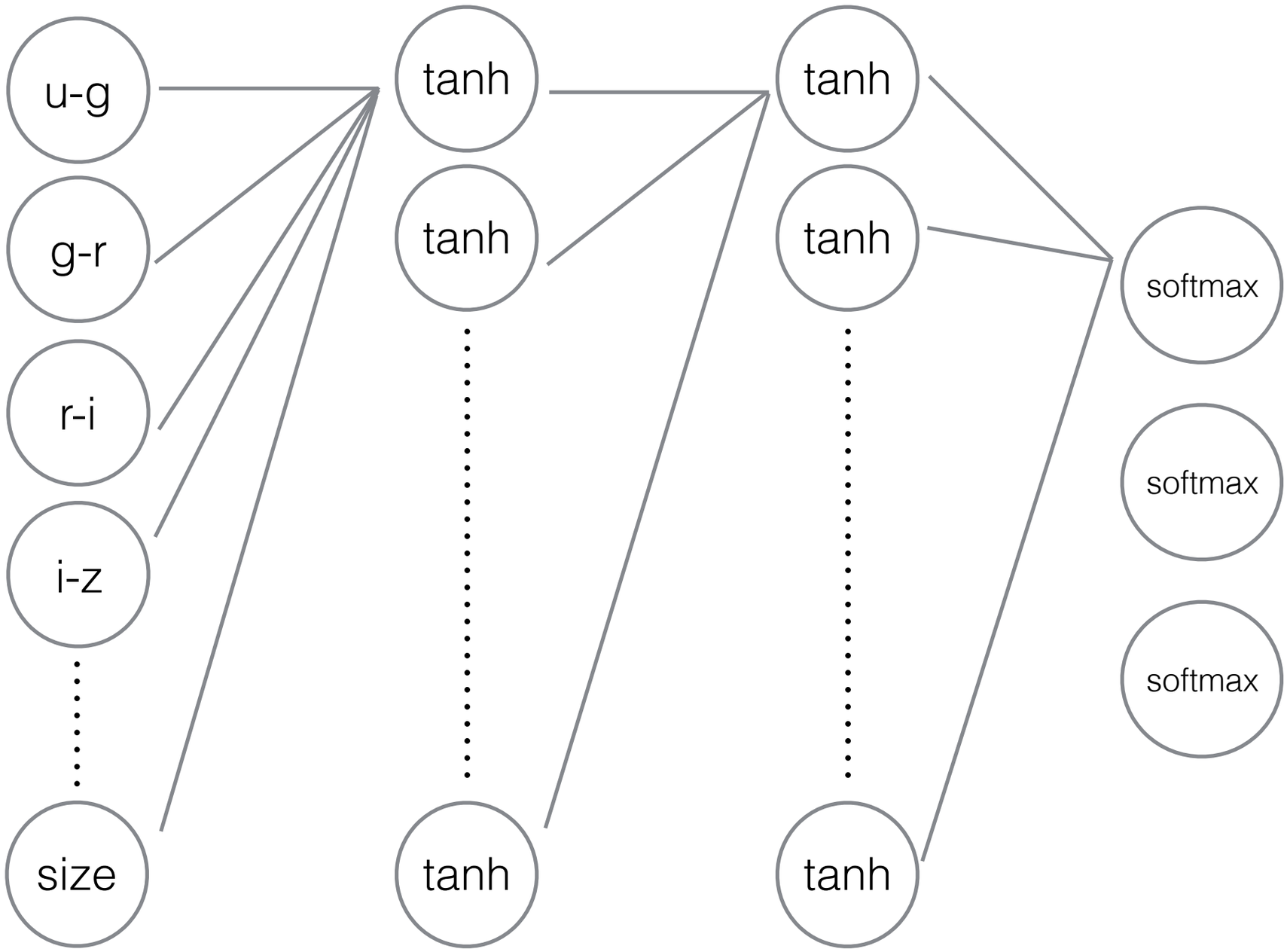
Colour-colour diagram



galaxies are redder
in average

The perceptron

Dense Neural Net with 2 hidden layers

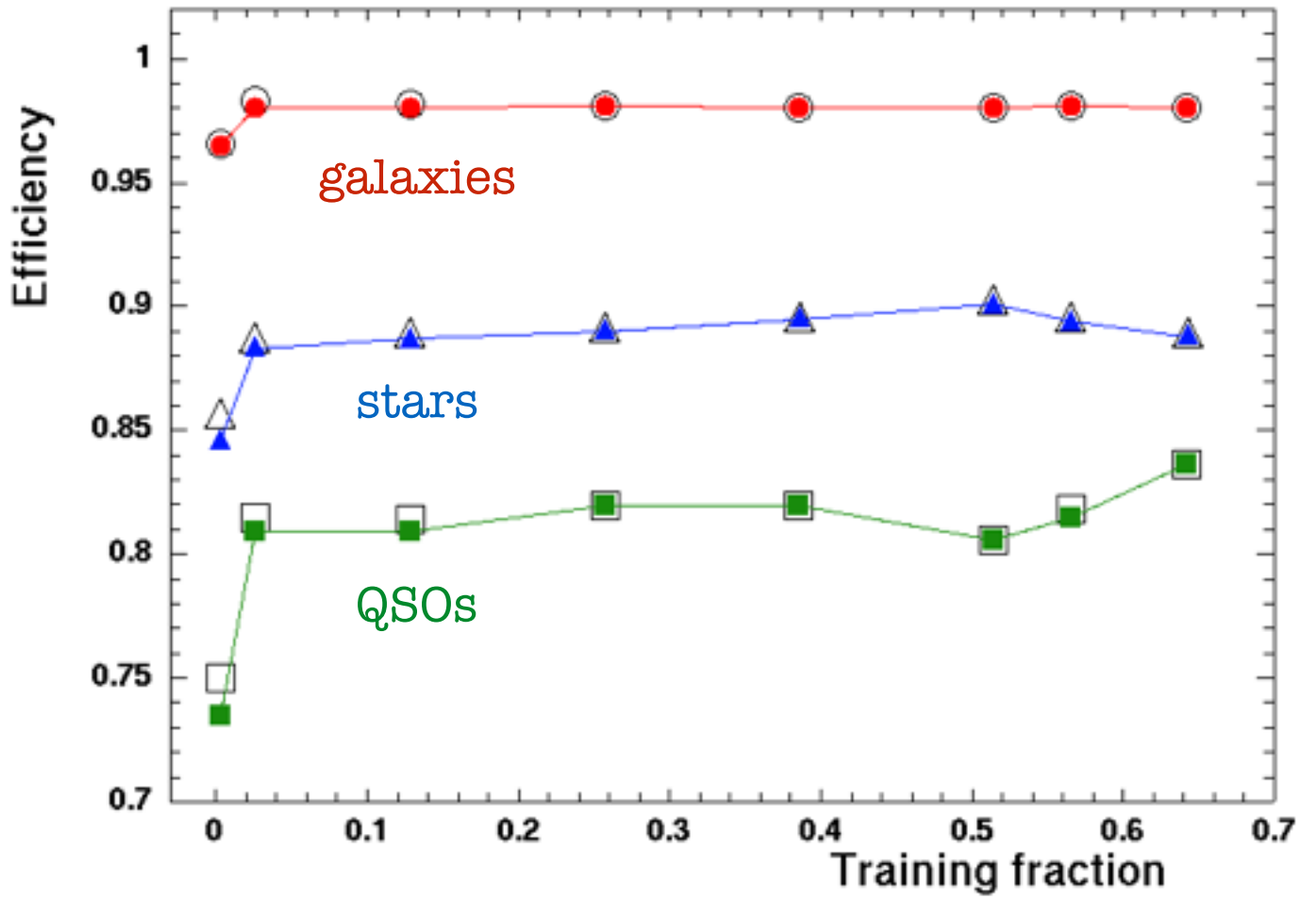


14 input features

50 activations

20 activations

3 output classes

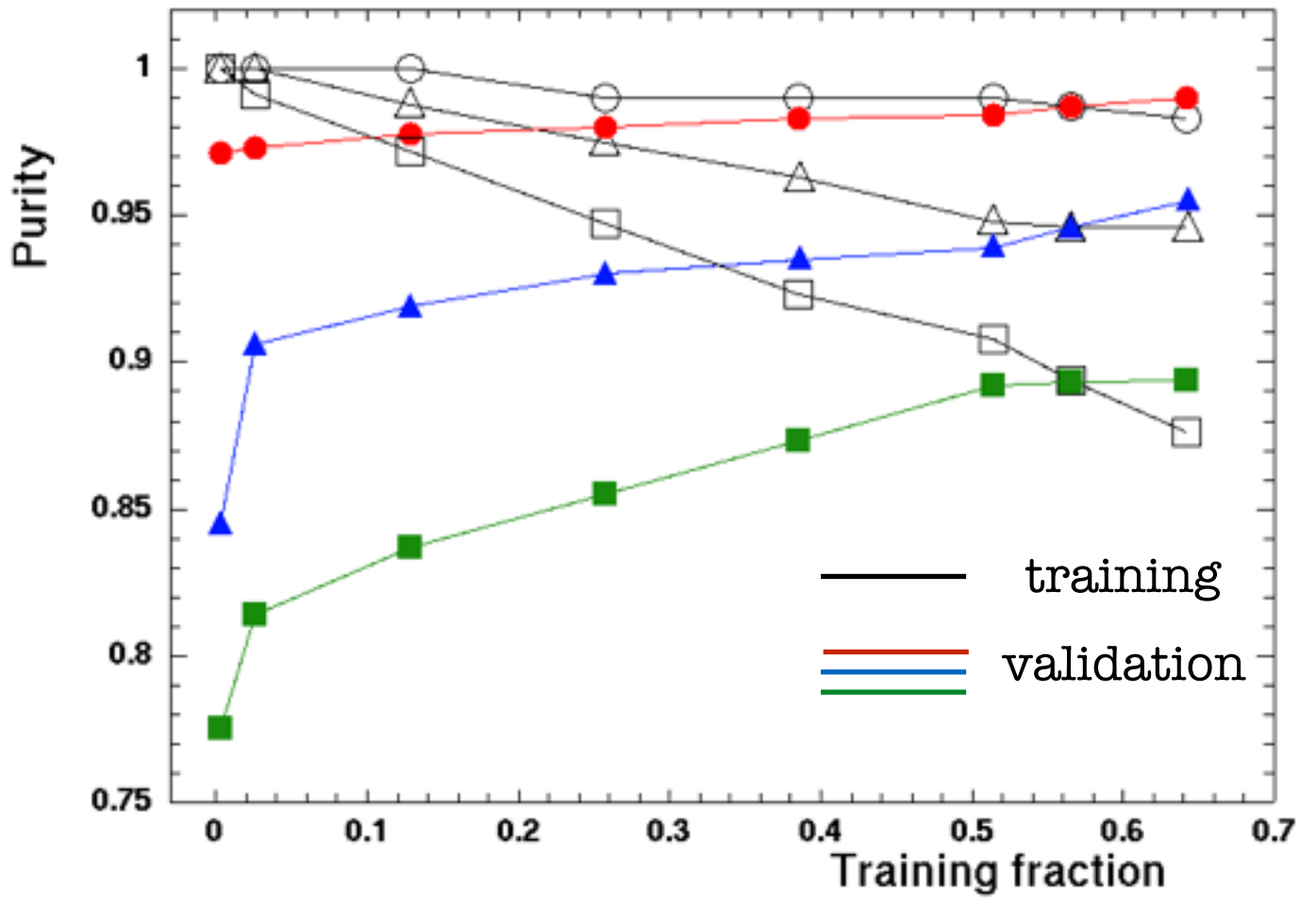


Training the perceptron

$$\text{efficiency}_i = \frac{n_{i\to i}}{n_i}$$

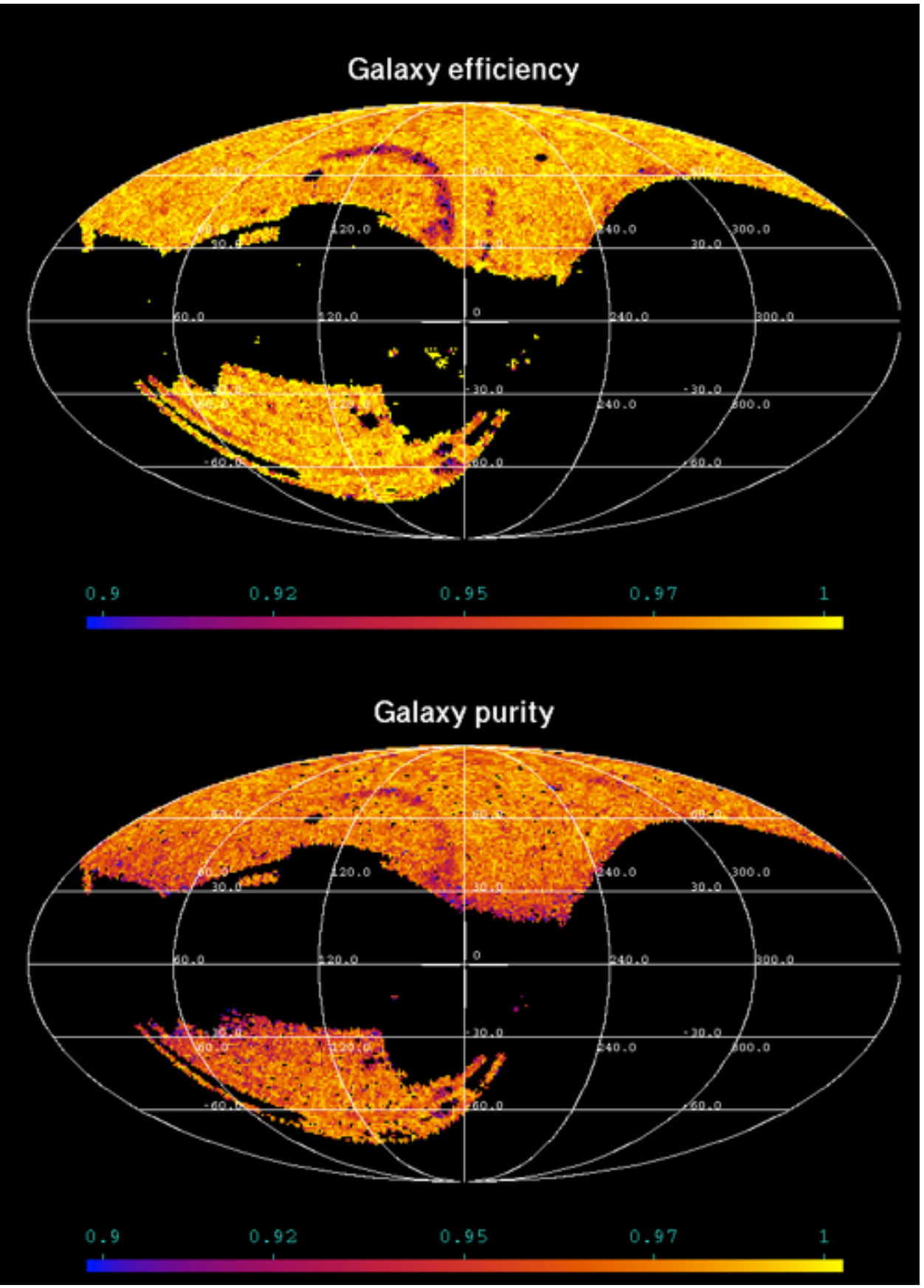
$$\text{purity}_i = \frac{n_{i\to i}}{n_{i\to i} + \sum_{j\neq i} n_{j\to i}}$$

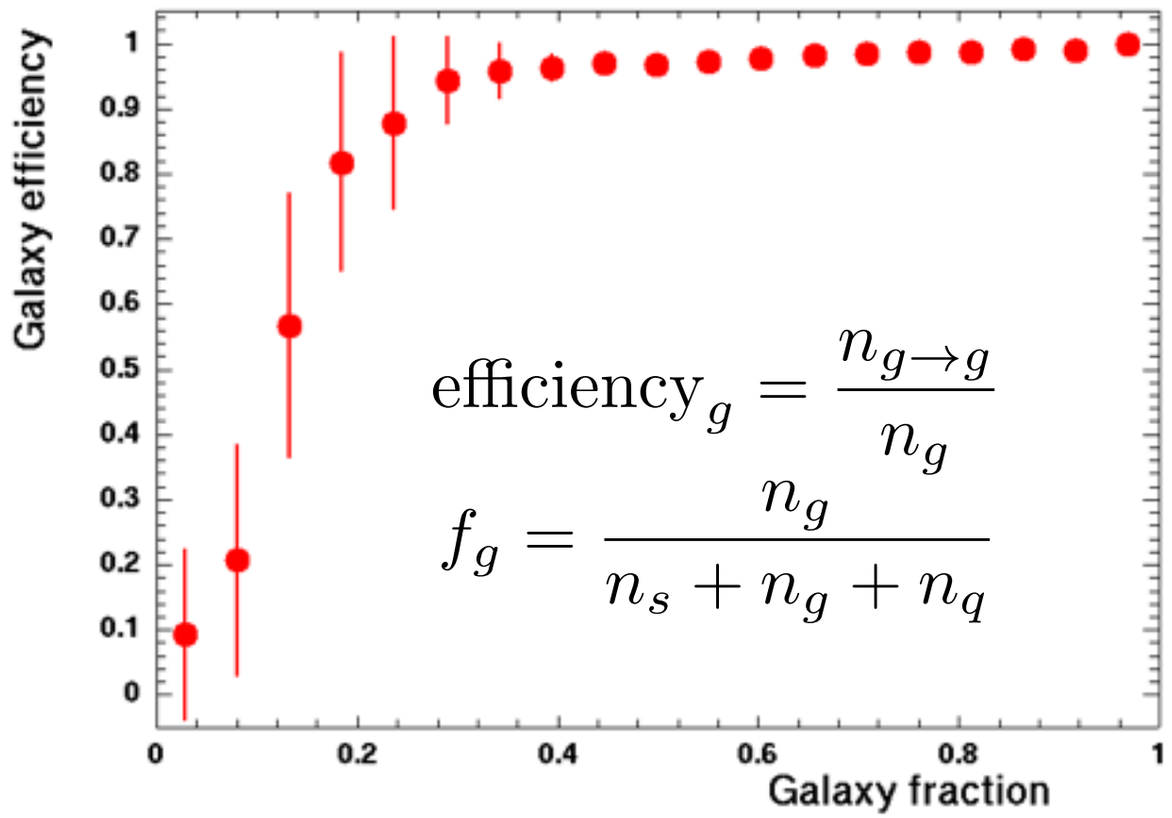
i : galaxy, star or QSO

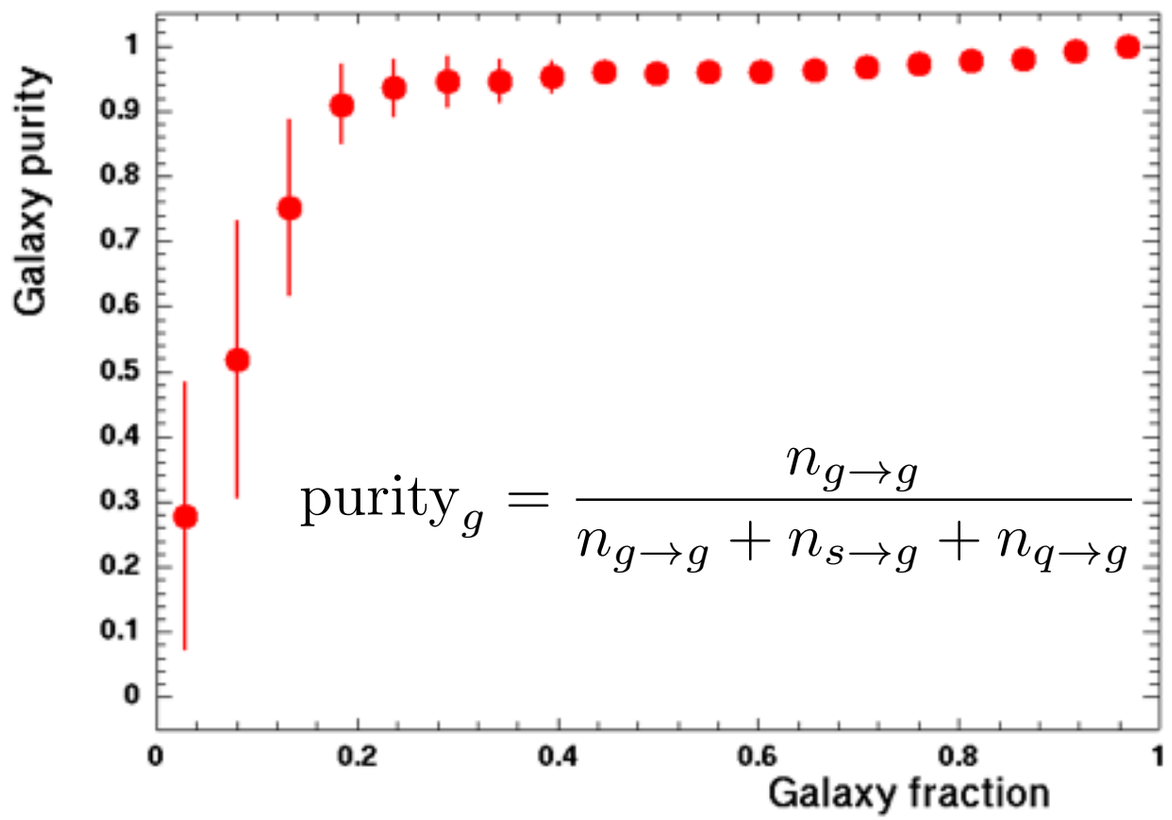


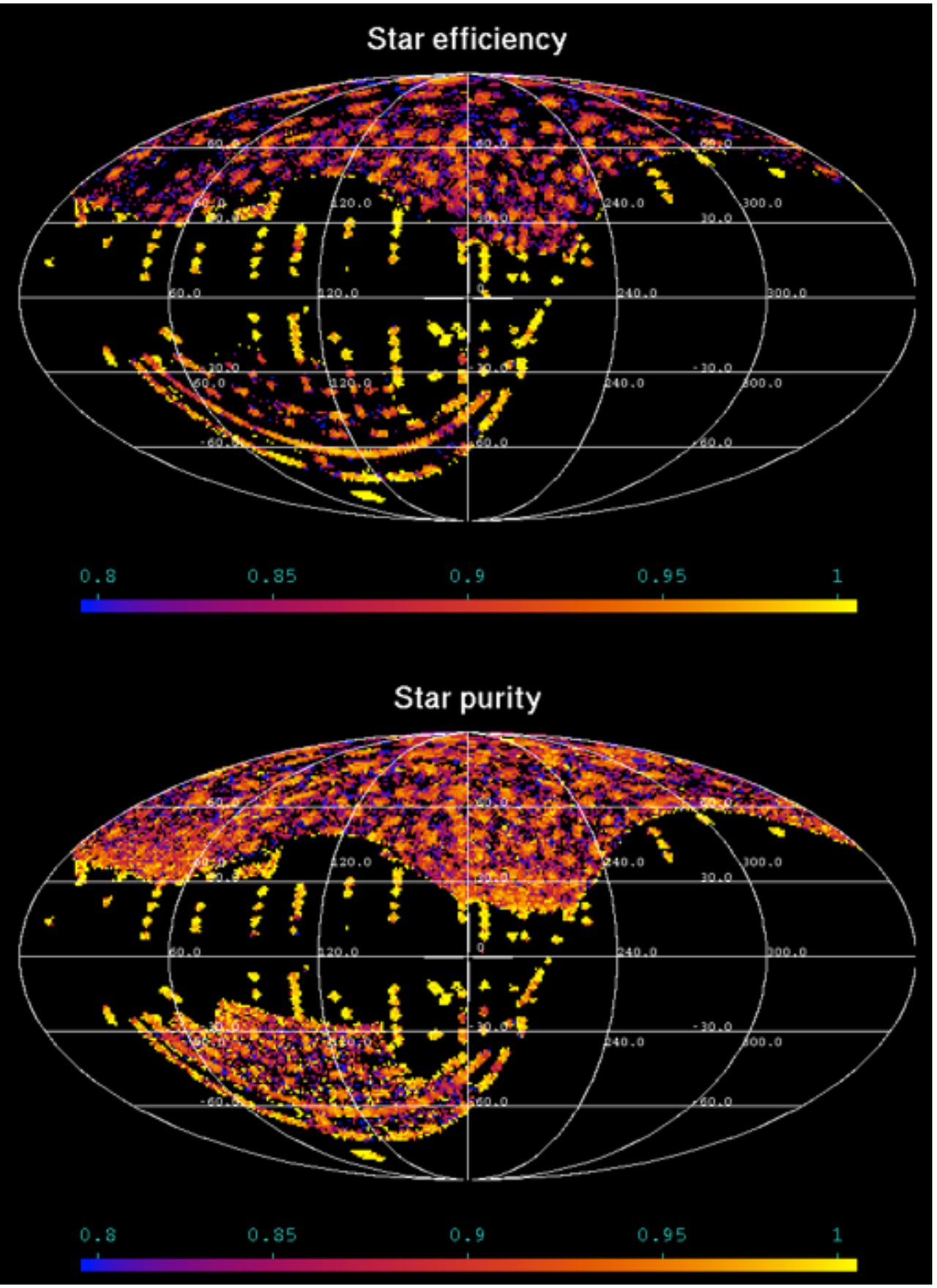
Comparing training and validation sets to find optimum number of objects for training

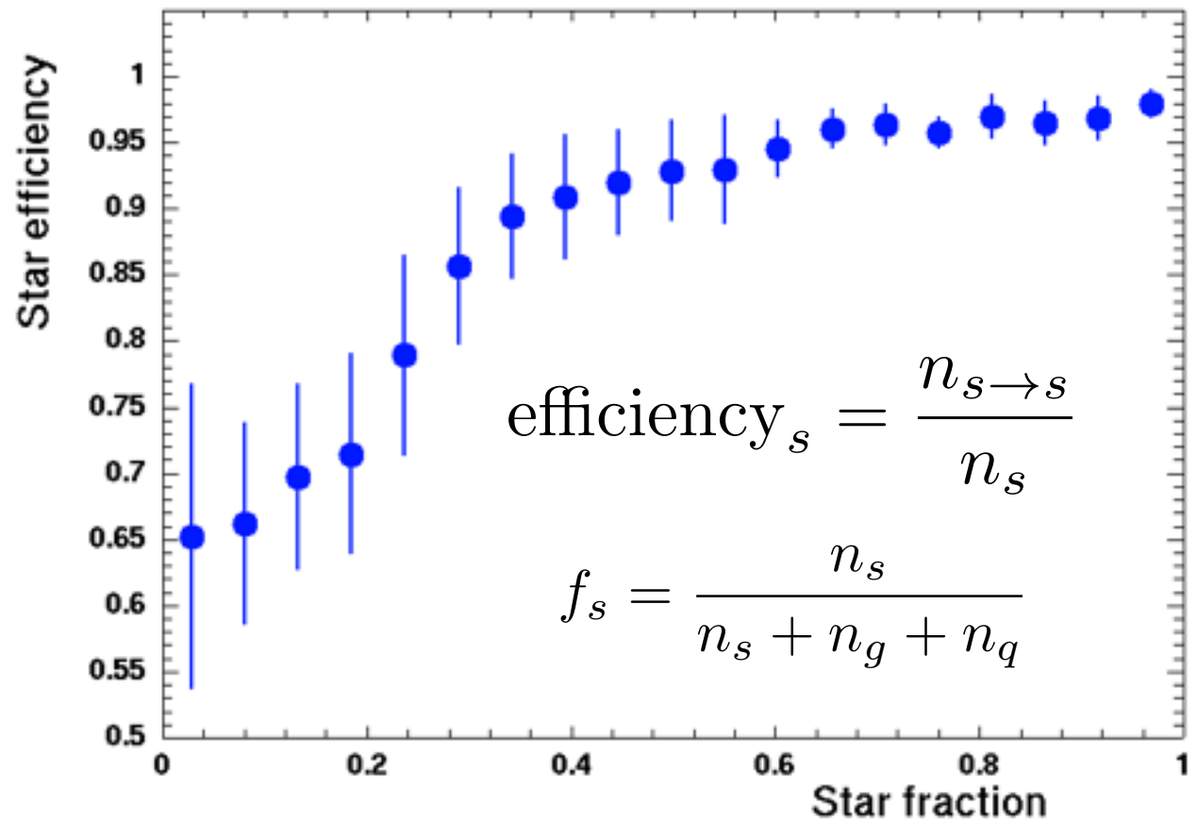
	Stars	Galaxies	QSOs	Total
Efficiency	89.4%	98.1%	81.5%	94%
Purity	94.6%	98.7%	89.3%	-

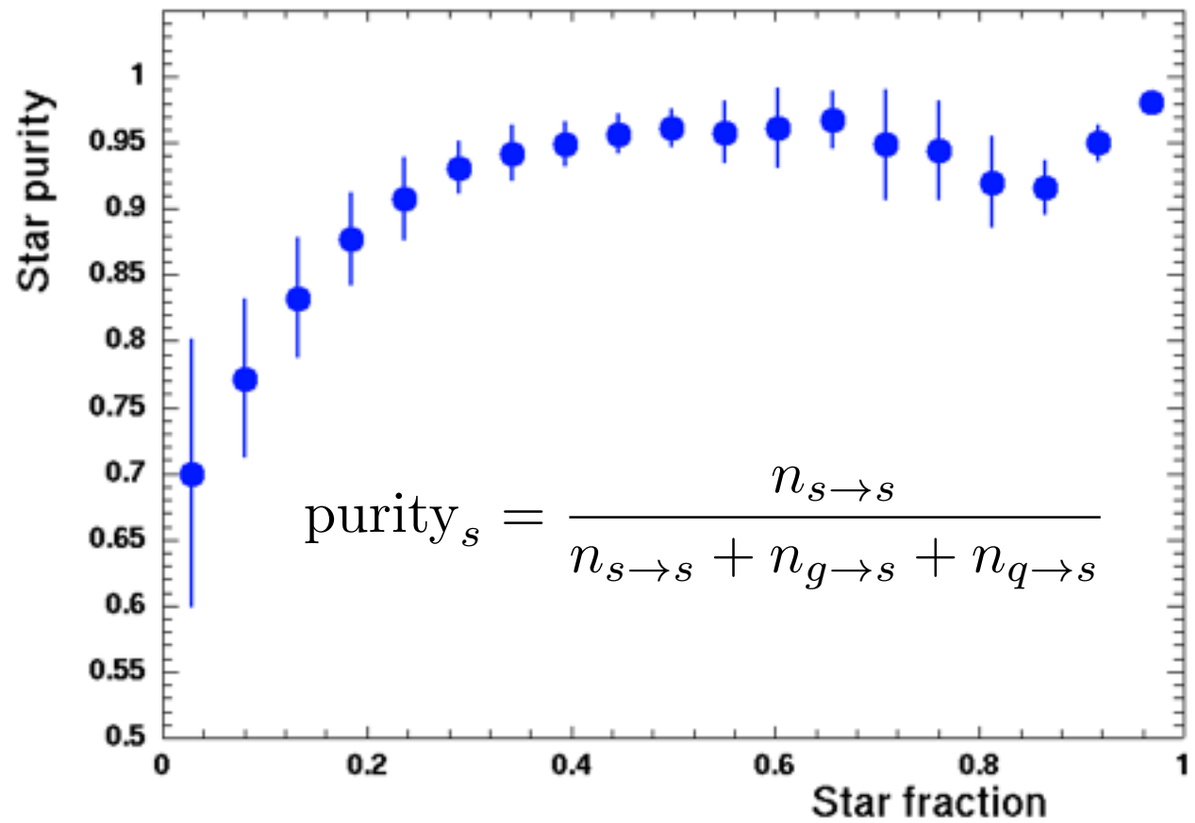

Galaxy efficiency
0.9
0.92
0.95
0.97
1
Galaxy purity
0.9
0.92
0.95
0.97
1

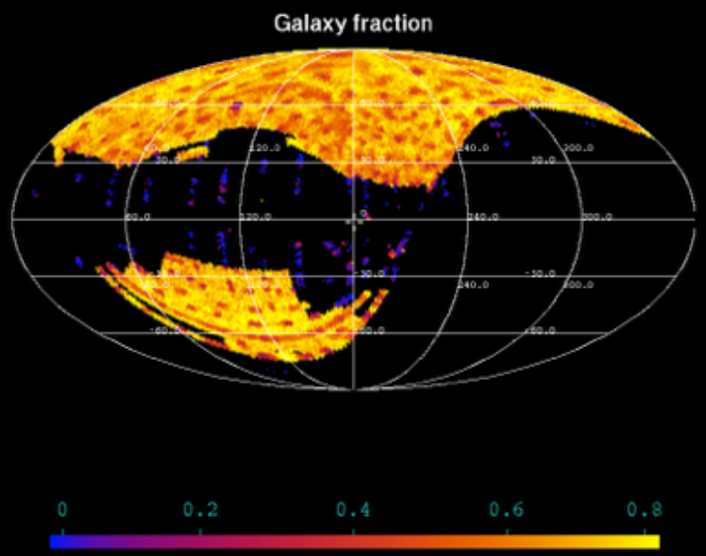

Galaxy efficiency
Galaxy fraction
$\text{efficiency}_g = \frac{n_{g\to g}}{n_g}$
$f_g = \frac{n_g}{n_s + n_g + n_q}$


Galaxy purity
Galaxy fraction
$\text{purity}_g = \frac{n_{g\to g}}{n_{g\to g} + n_{s\to g} + n_{q\to g}}$

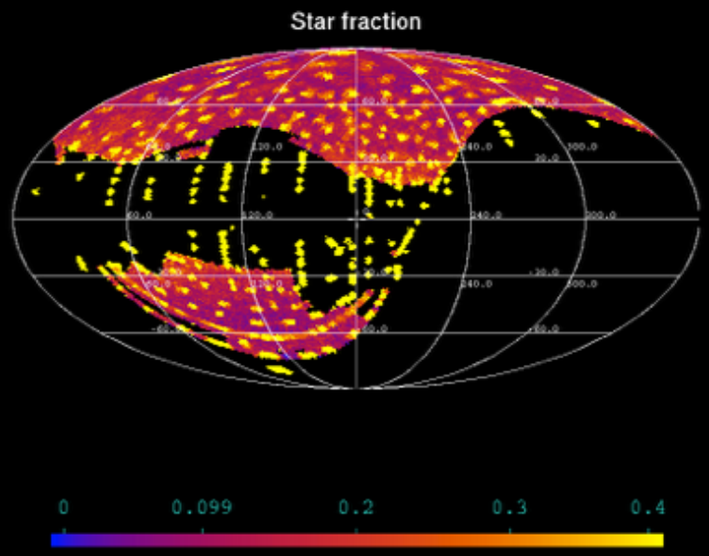

Star efficiency
0.8
0.85
0.9
0.95
1
Star purity
0.8
0.85
0.9
0.95
1

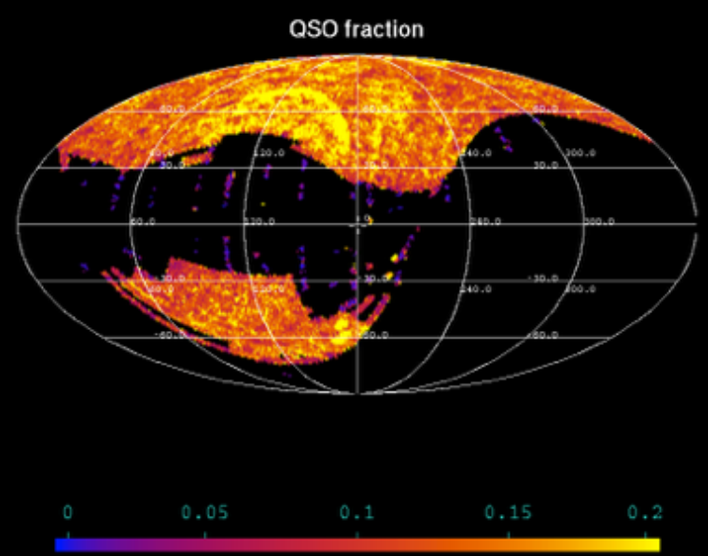

Star efficiency
Star fraction
$\text{efficiency}_s = \frac{n_{s\to s}}{n_s}$
$f_s = \frac{n_s}{n_s + n_g + n_q}$


Star purity
Star fraction
$\text{purity}_s = \frac{n_{s\to s}}{n_{s\to s} + n_{g\to s} + n_{q\to s}}$



DR 12



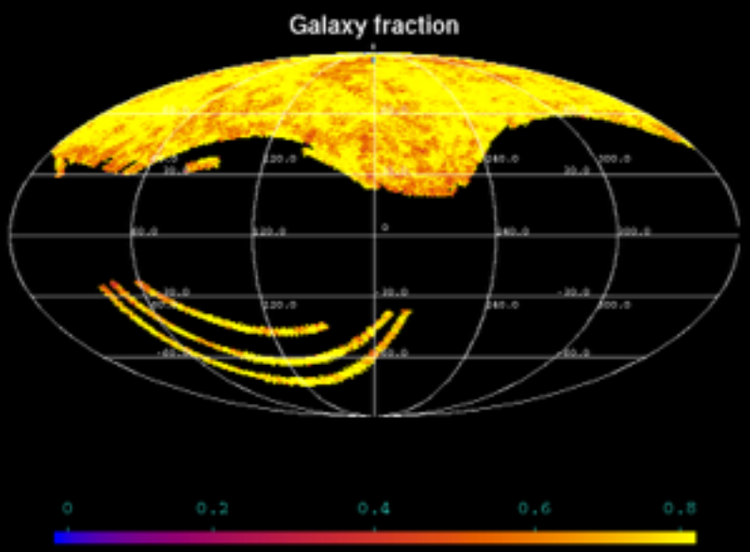


	Stars	Galaxies	QSOs	Total
Num	160,040	879,792	120,425	1,160,257
fraction	14%	76%	10%	
efficiency	95%	99%	90%	98%
purity	94%	99%	94%	

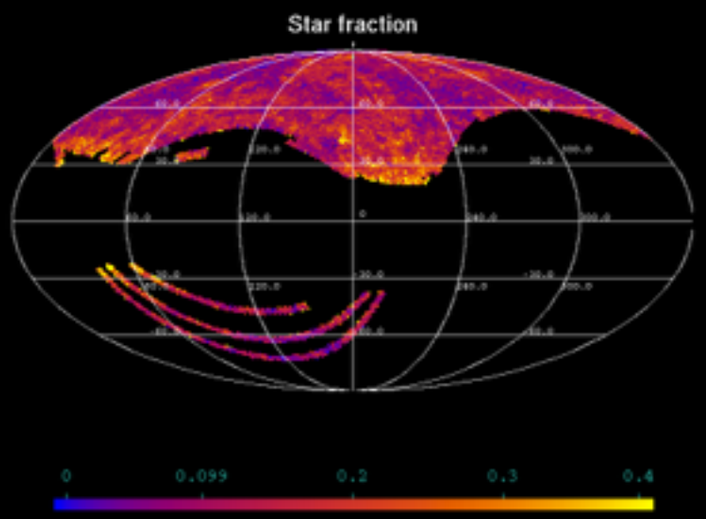
DR12

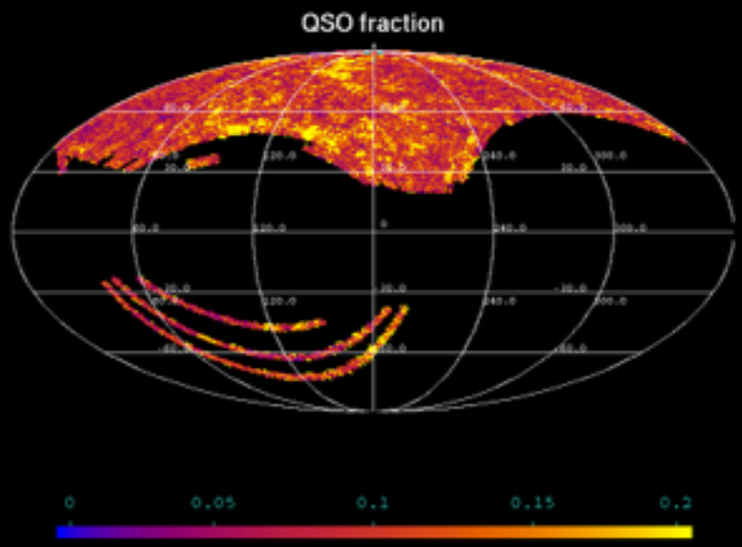
VS

Legacy

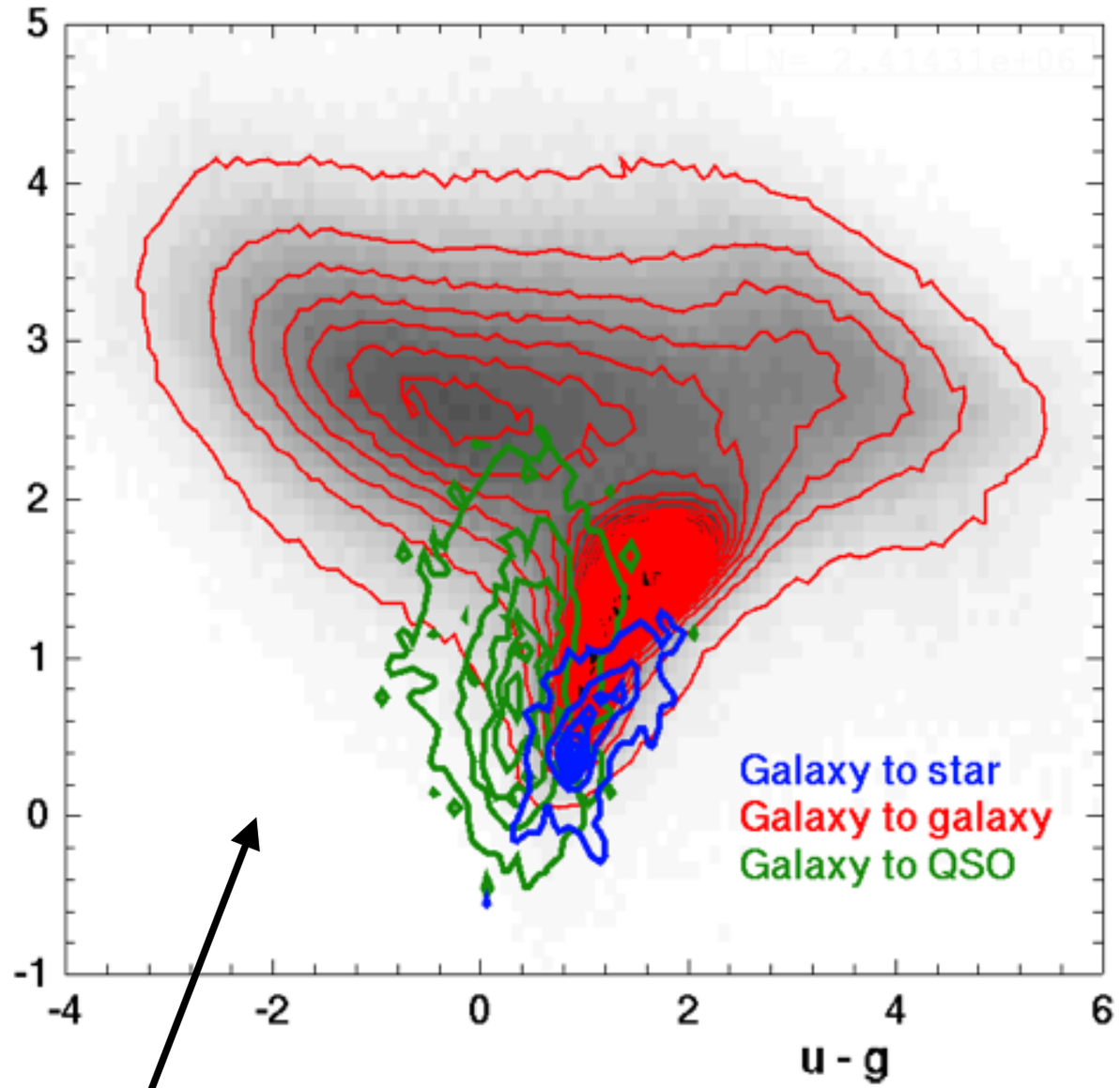


Legacy

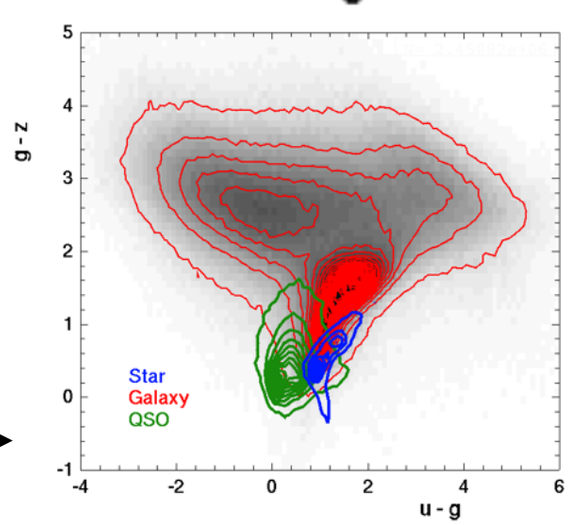




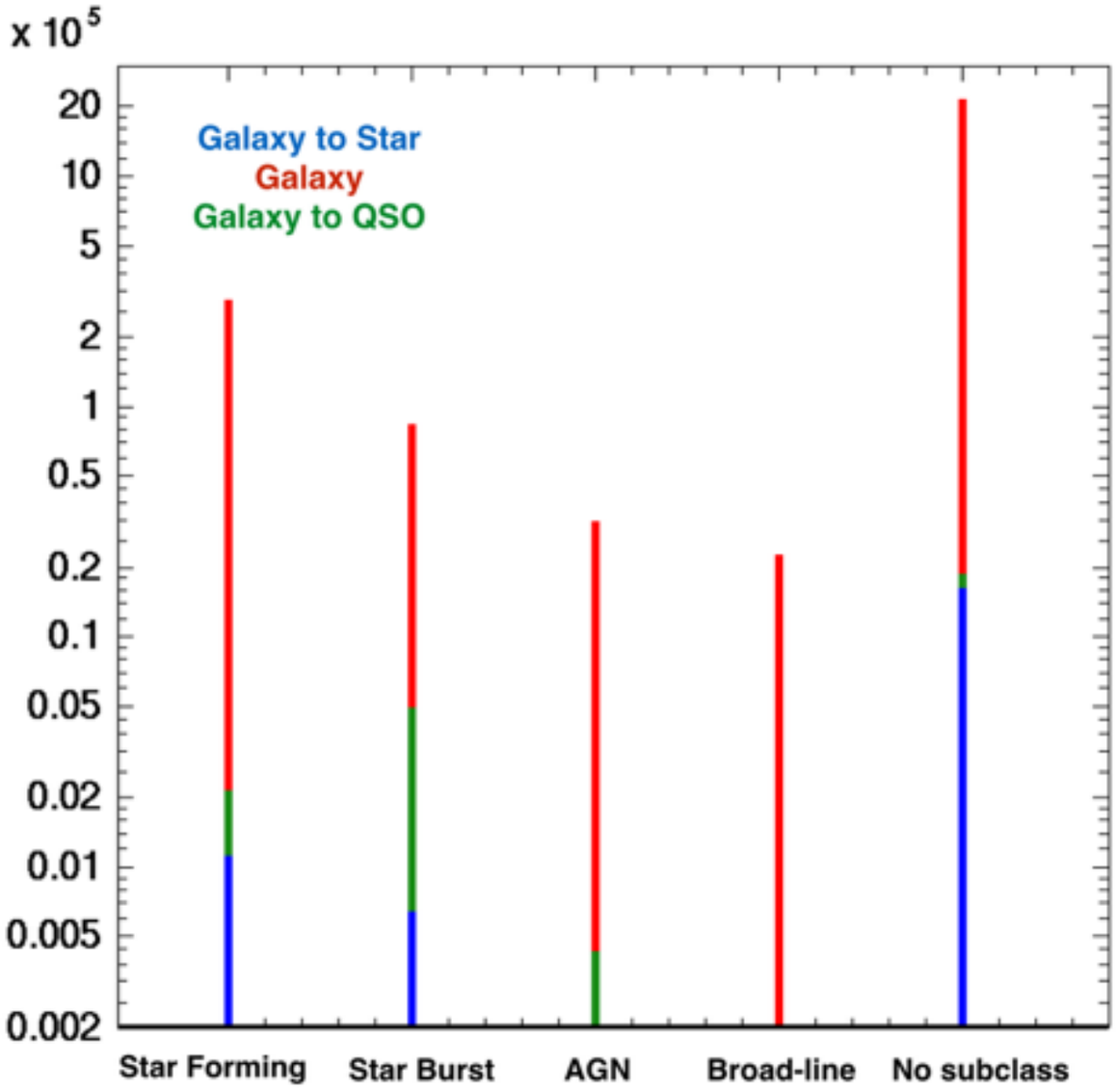
Galaxy misclassifications



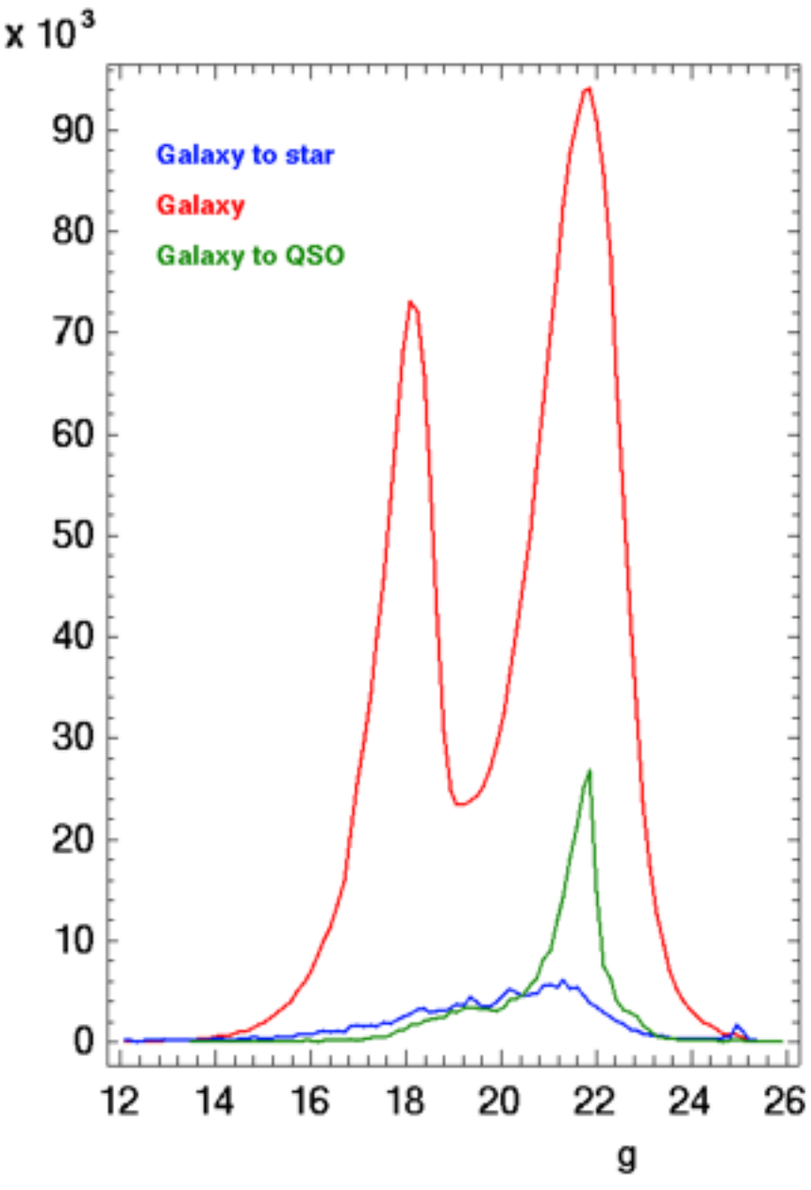
Classified types
vs
spectroscopic types

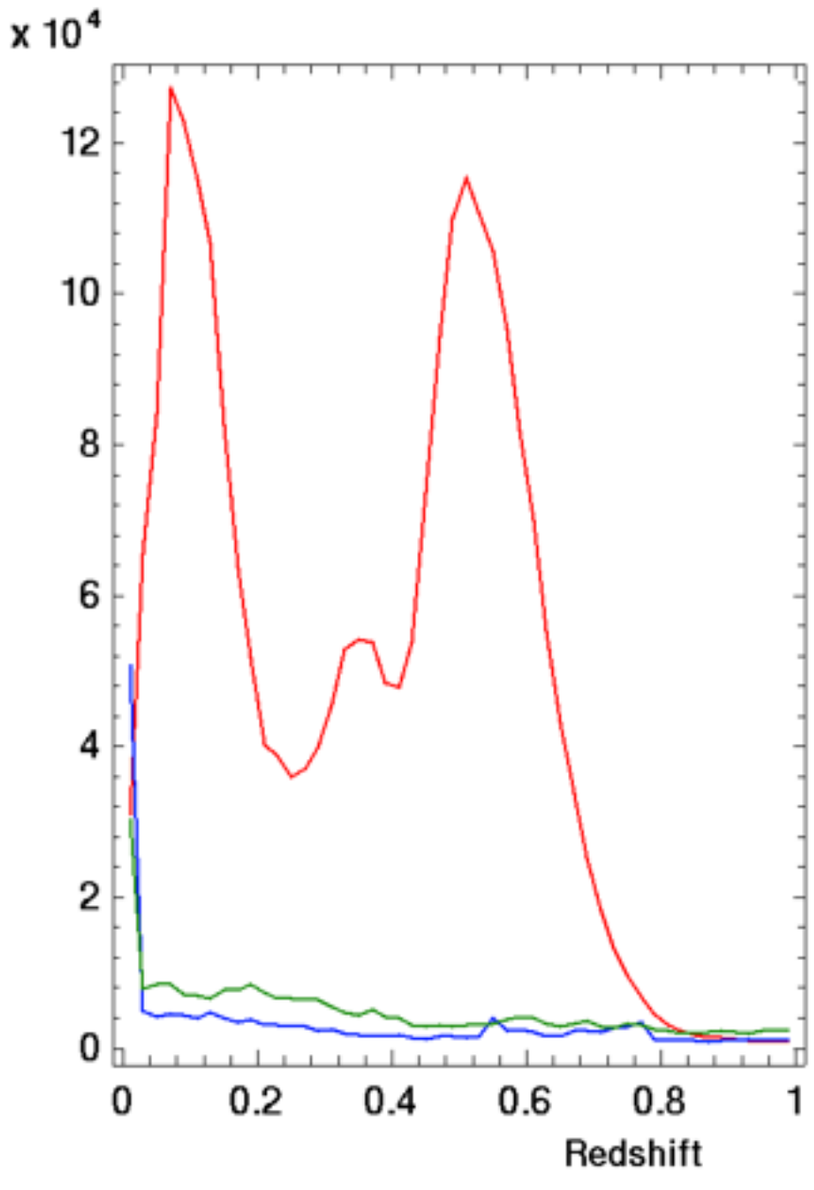


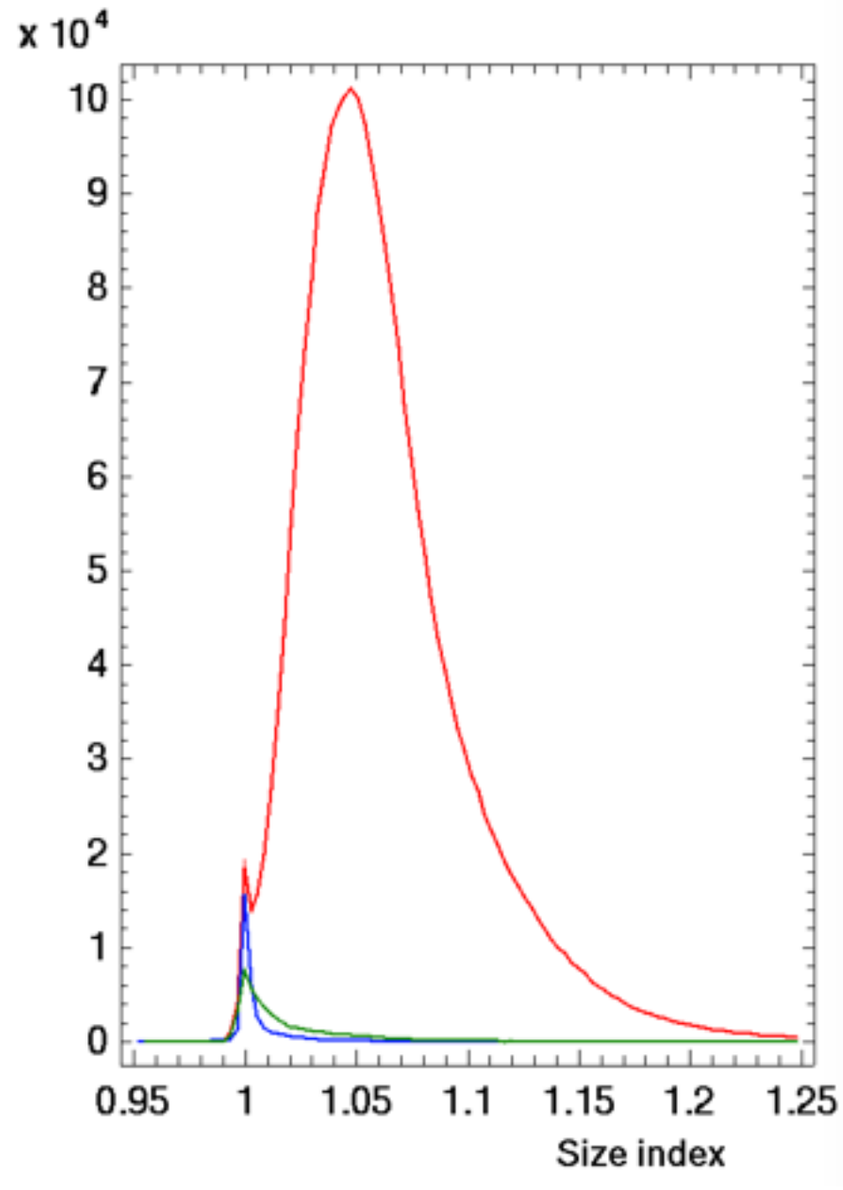
Misclassified sub-classes



Galaxy misclassifications

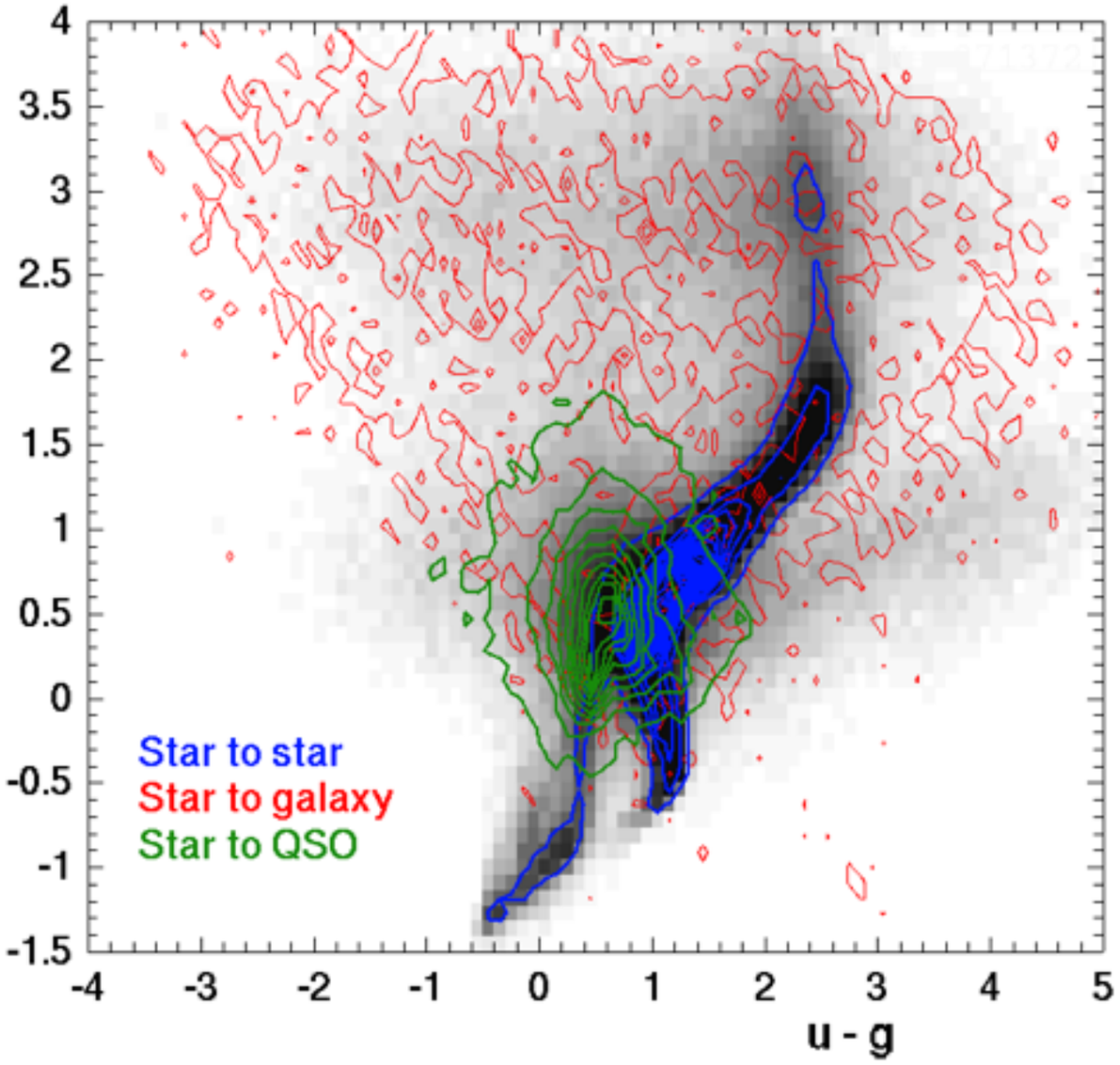






Faint nearby galaxies ==> point-like sources

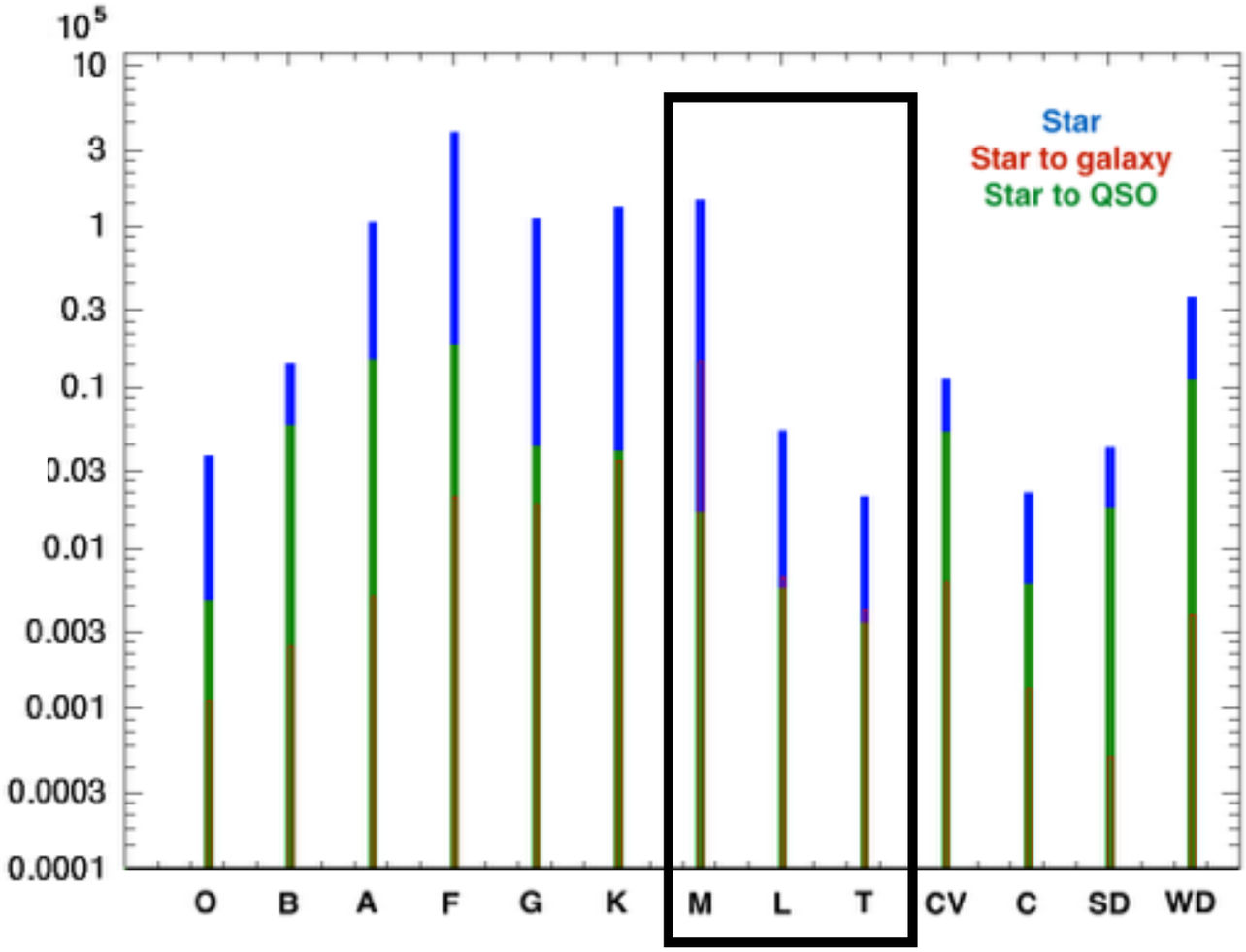
Star misclassifications



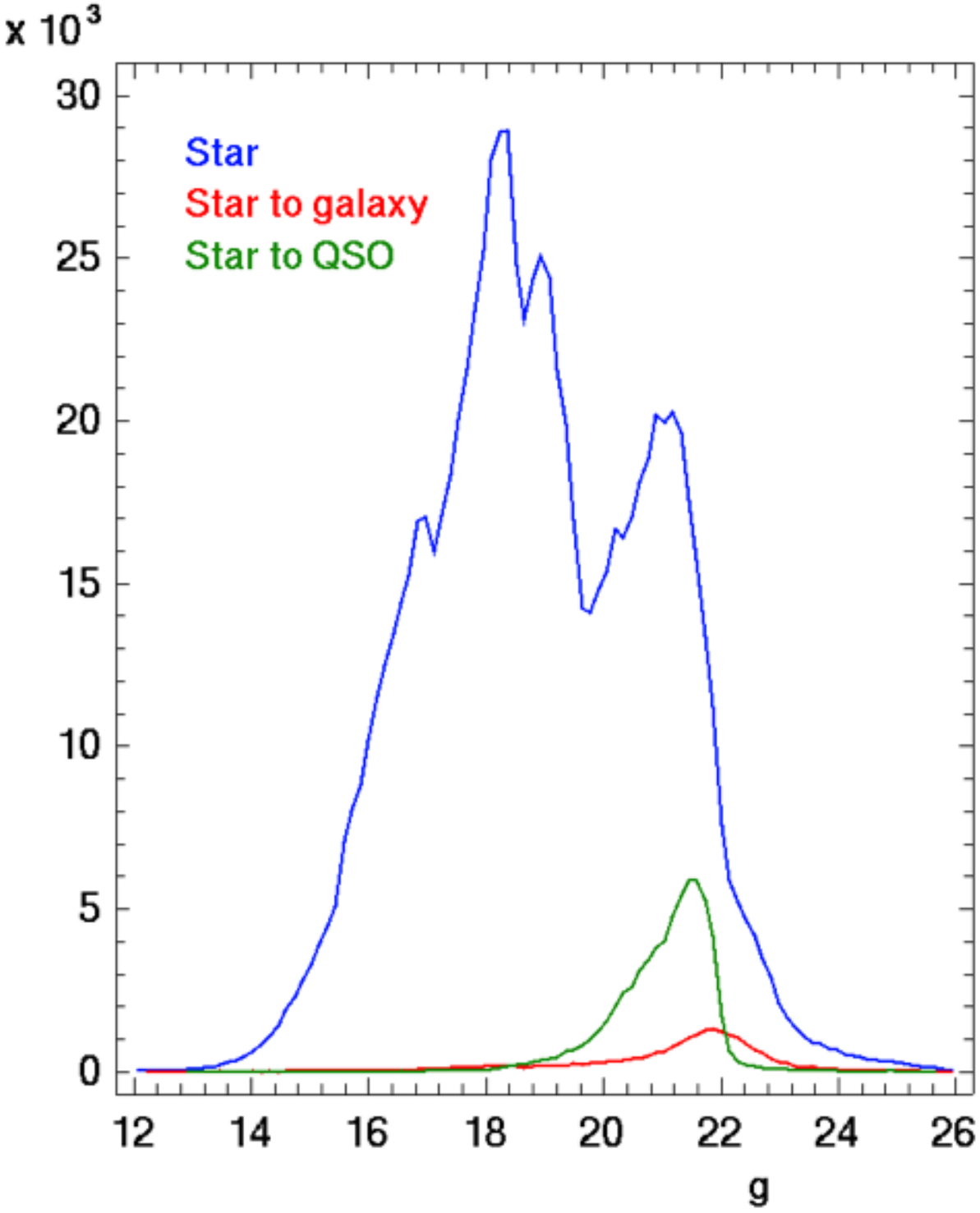
stars with
scattered colours
contaminates galaxy sample

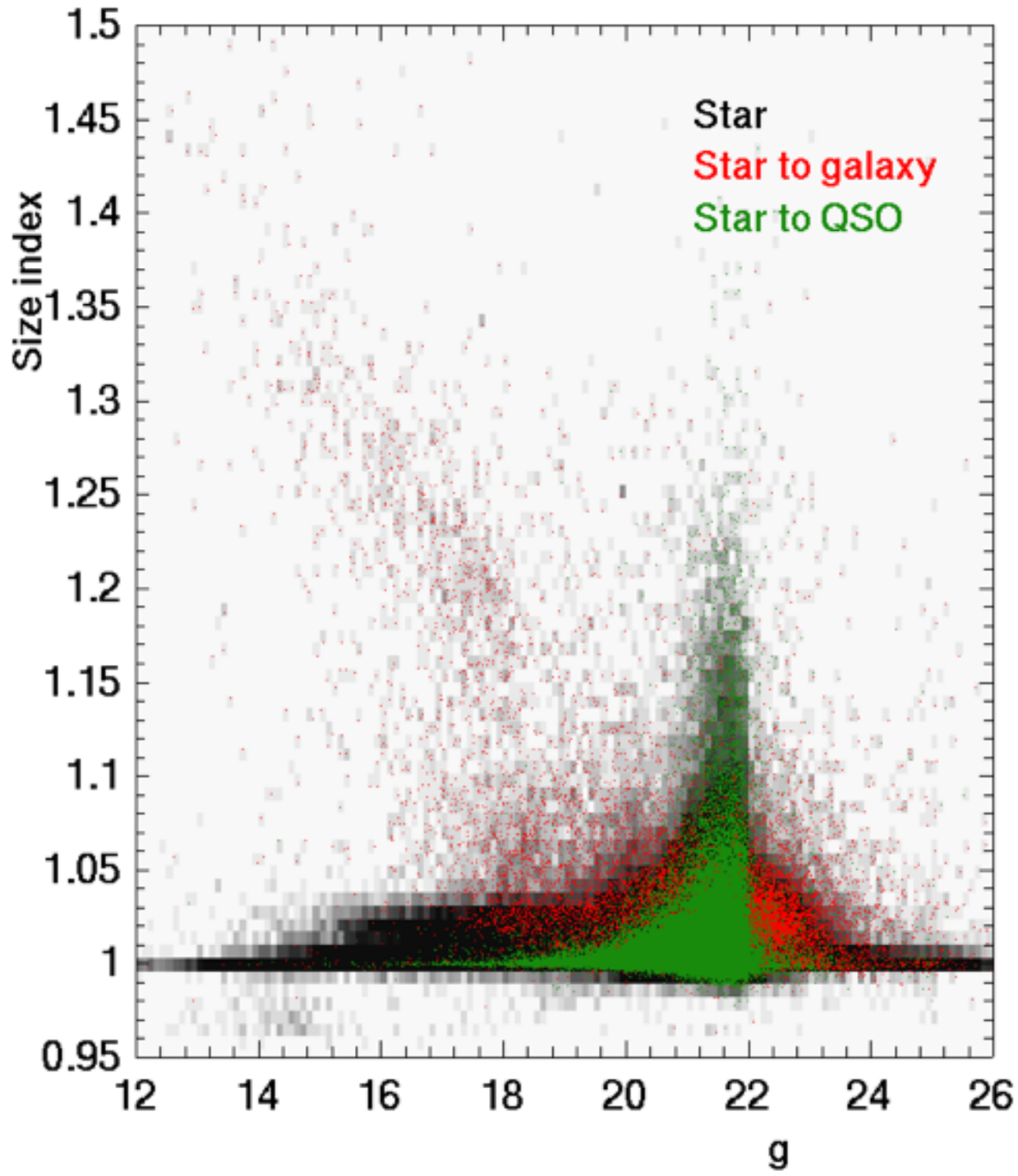
Misclassified sub-classes

> 10% of M, L, T stars ==> galaxy



Star misclassifications





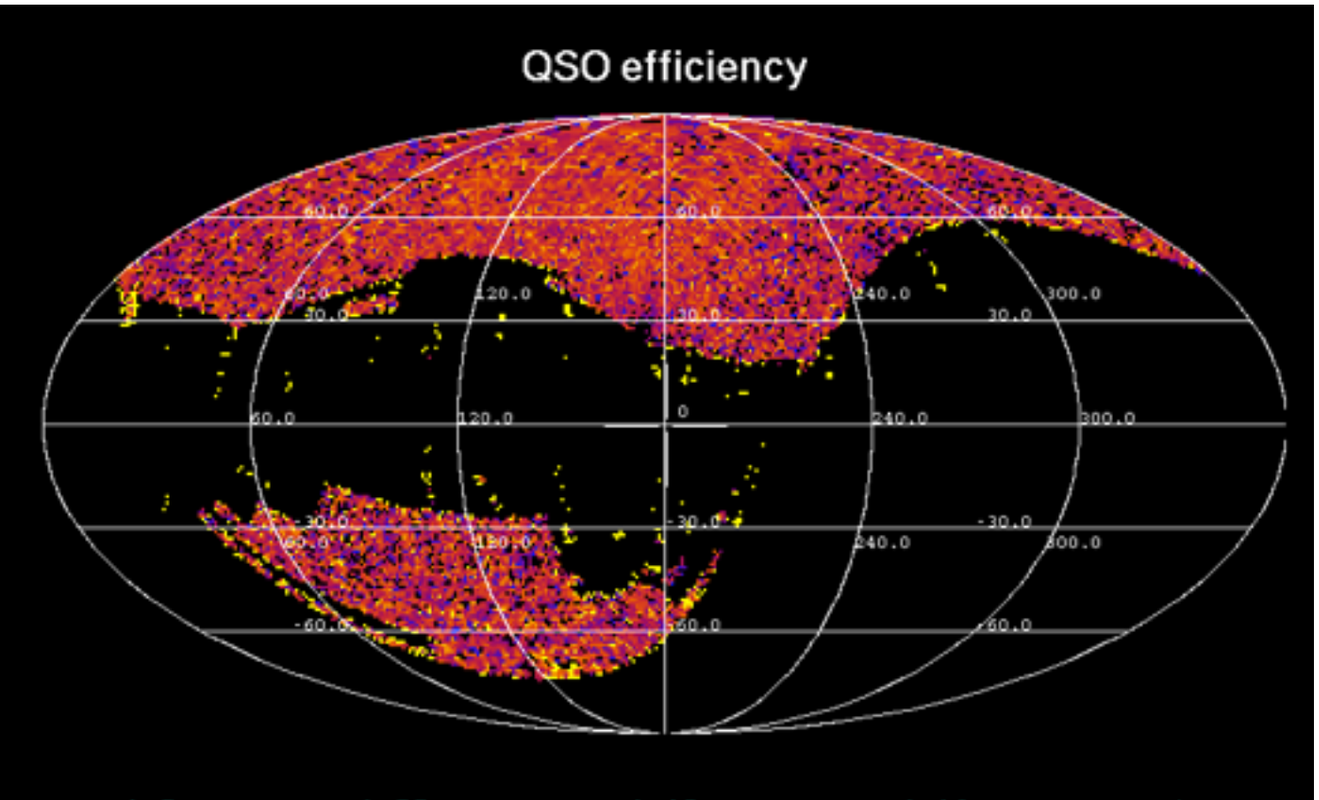
deviation from point-like source for faint stars

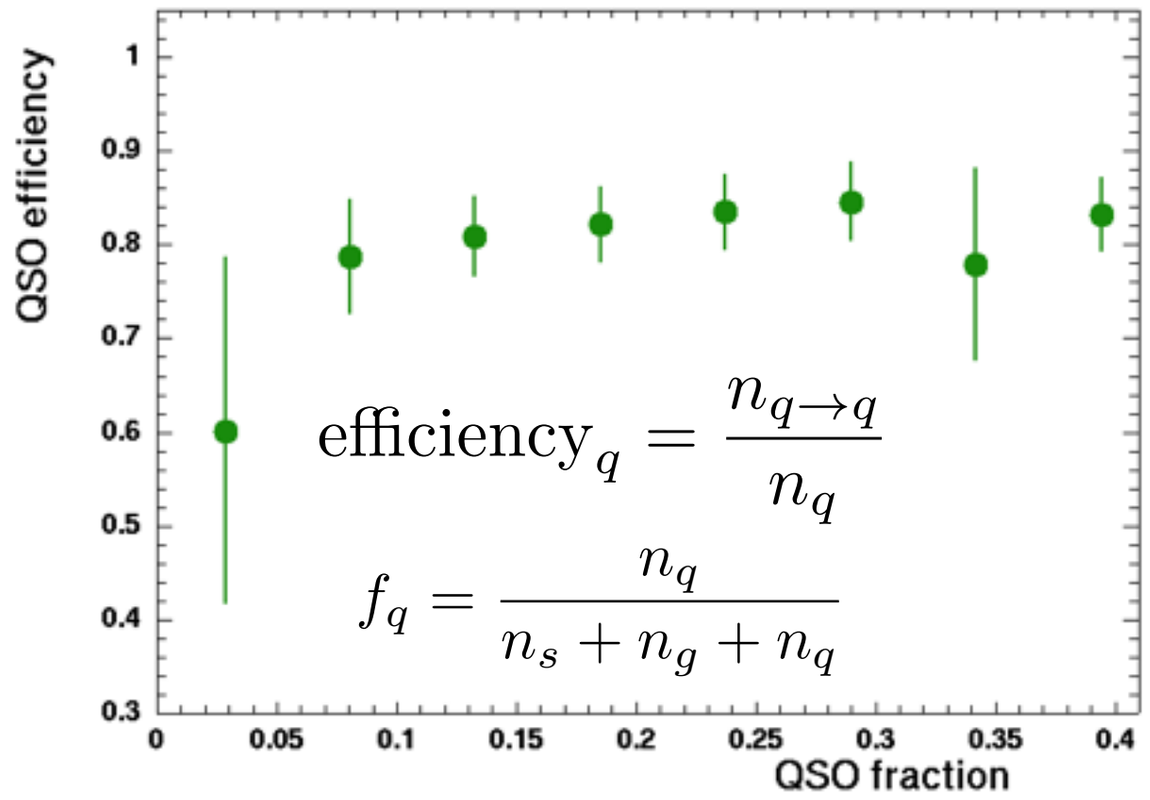
Classification for the LSST

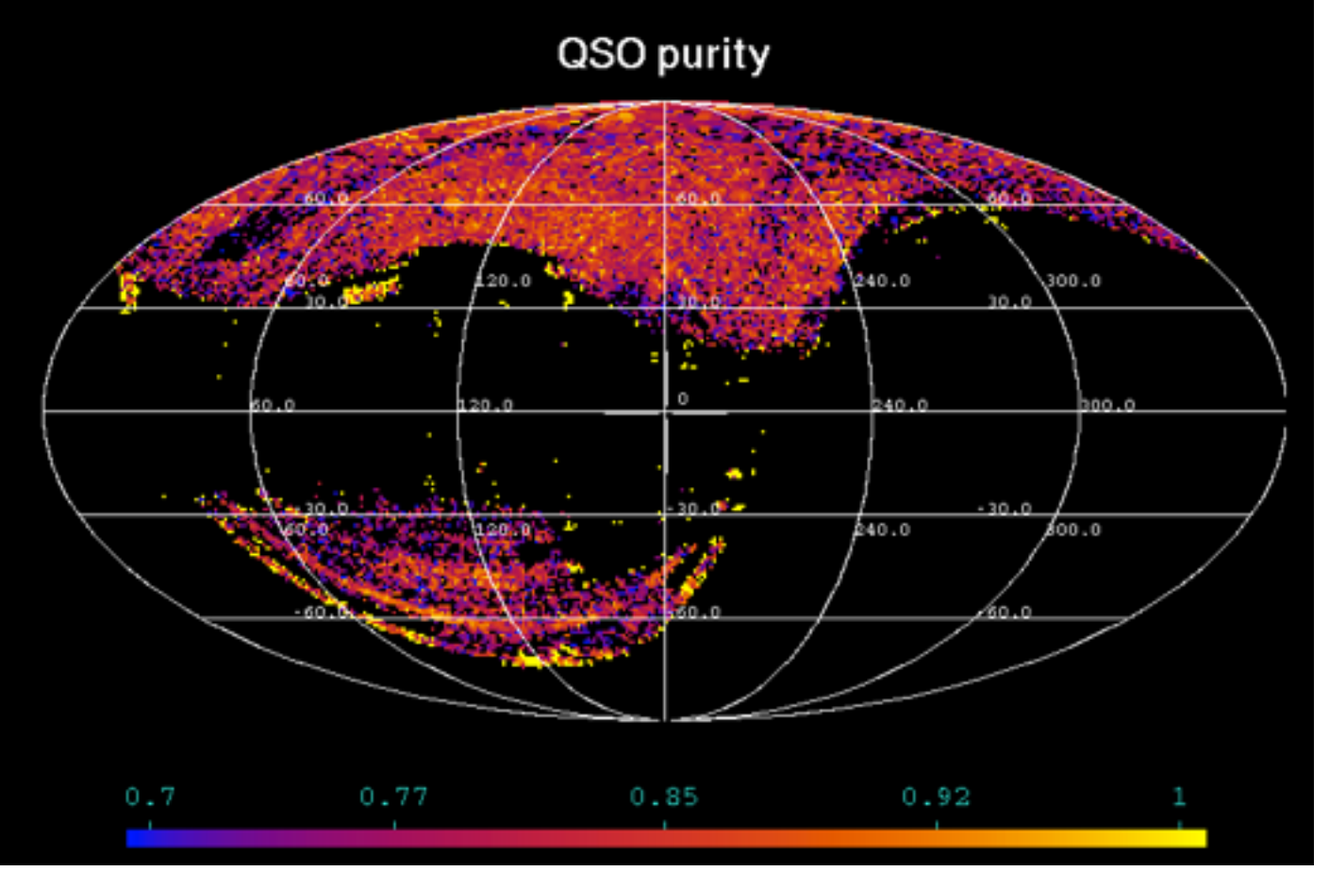
- Considering a redshift interval for the galaxies according to the LSST apparent magnitude limits
- Simulating the colour indices from galaxy SEDs
- Assigning luminosity profile to the galaxies
- Including the stars

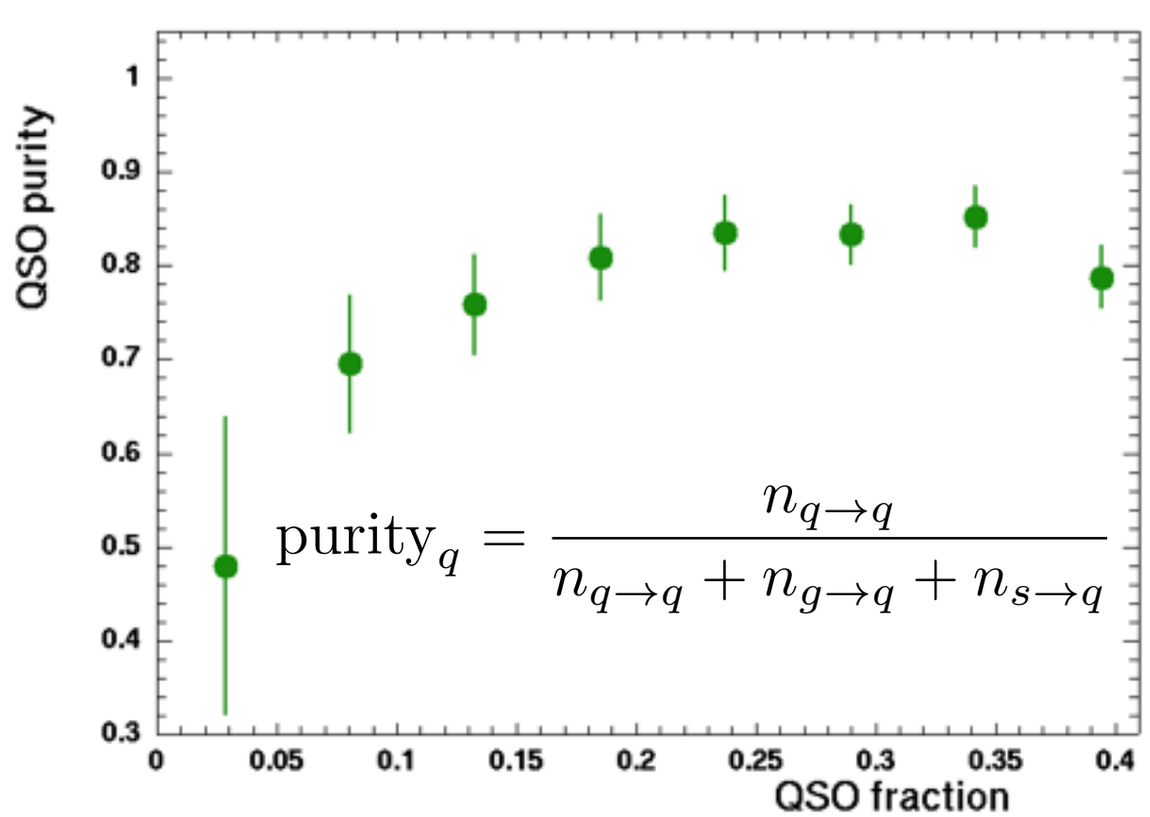
Joyeux Norouz (-:

Backups

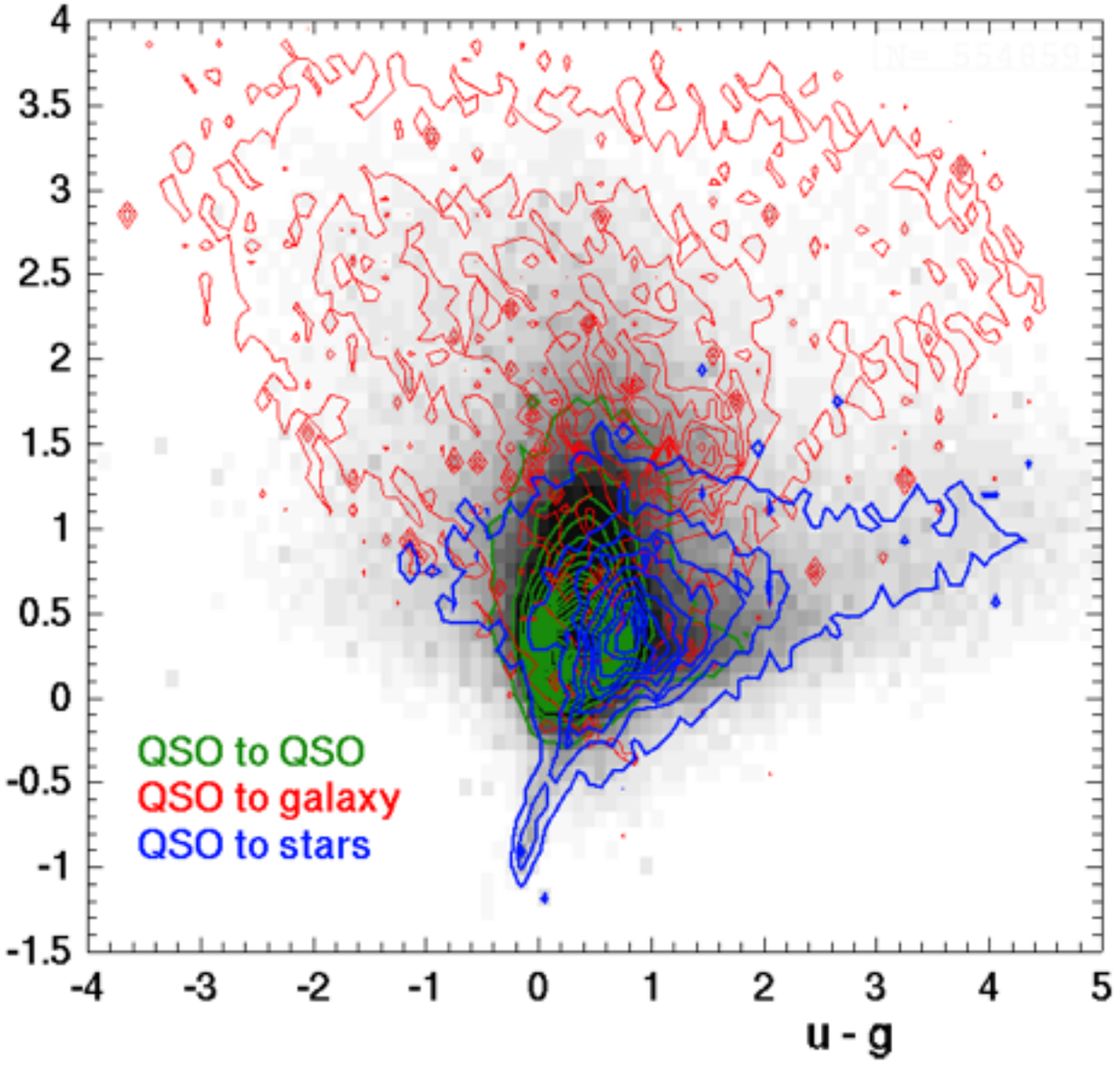

QSO efficiency
0.7
0.77
0.85
0.92
1

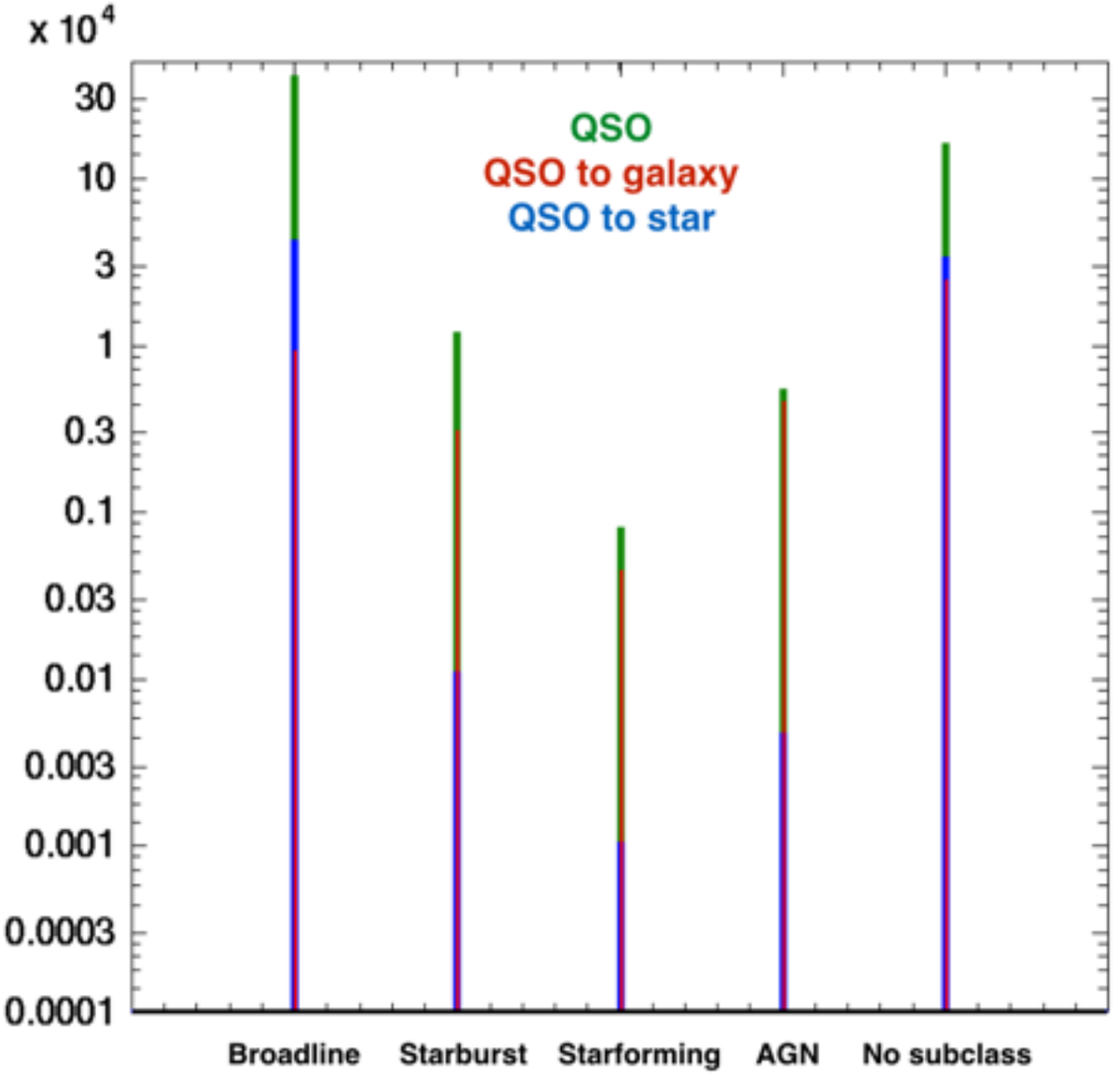

QSO efficiency
QSO fraction
$\text{efficiency}_q = \frac{n_{q\to q}}{n_q}$
$f_q = \frac{n_q}{n_s + n_g + n_q}$


QSO purity
0.7
0.77
0.85
0.92
1

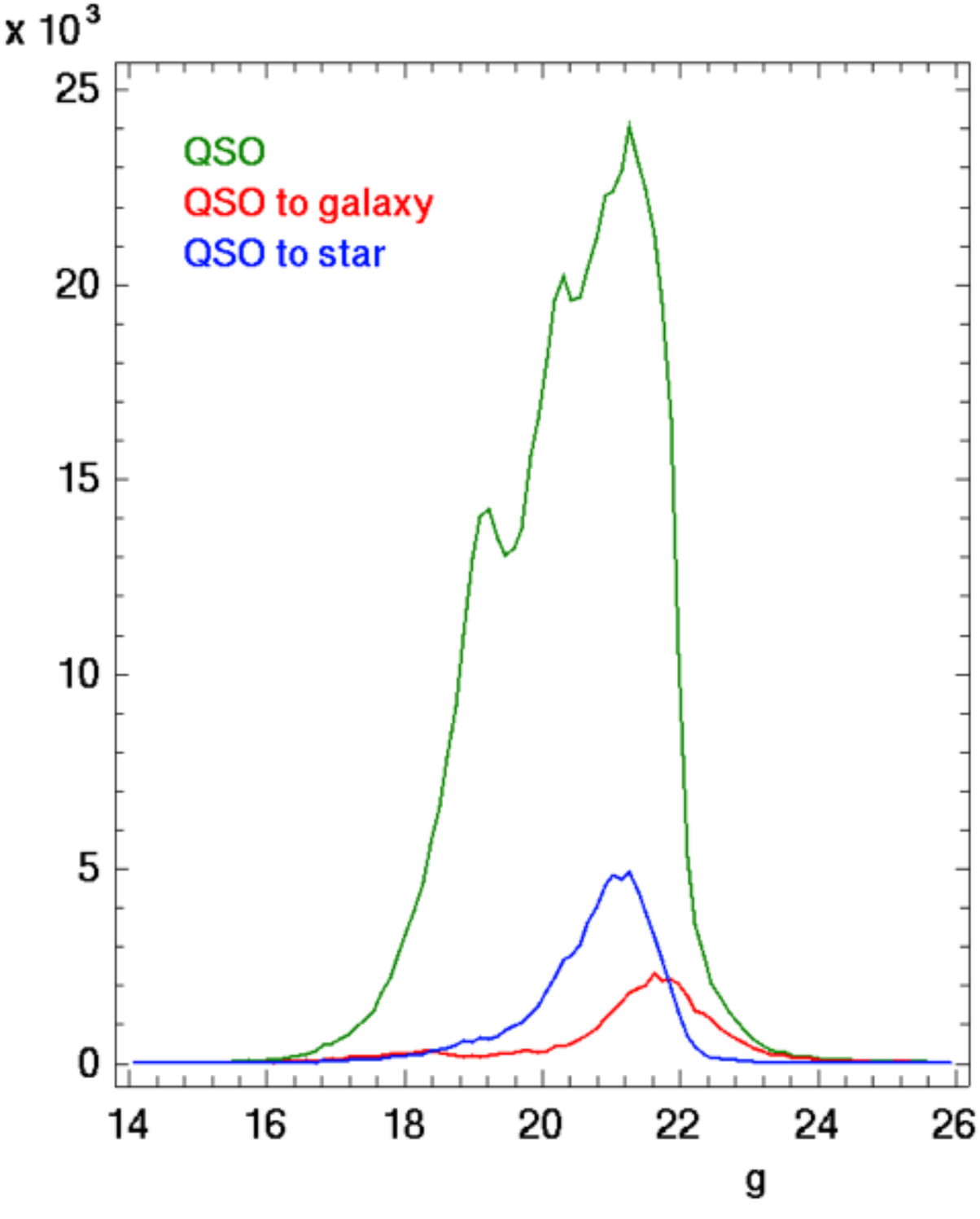

QSO purity
QSO fraction
$\text{purity}_q = \frac{n_{q\to q}}{n_{q\to q} + n_{g\to q} + n_{s\to q}}$

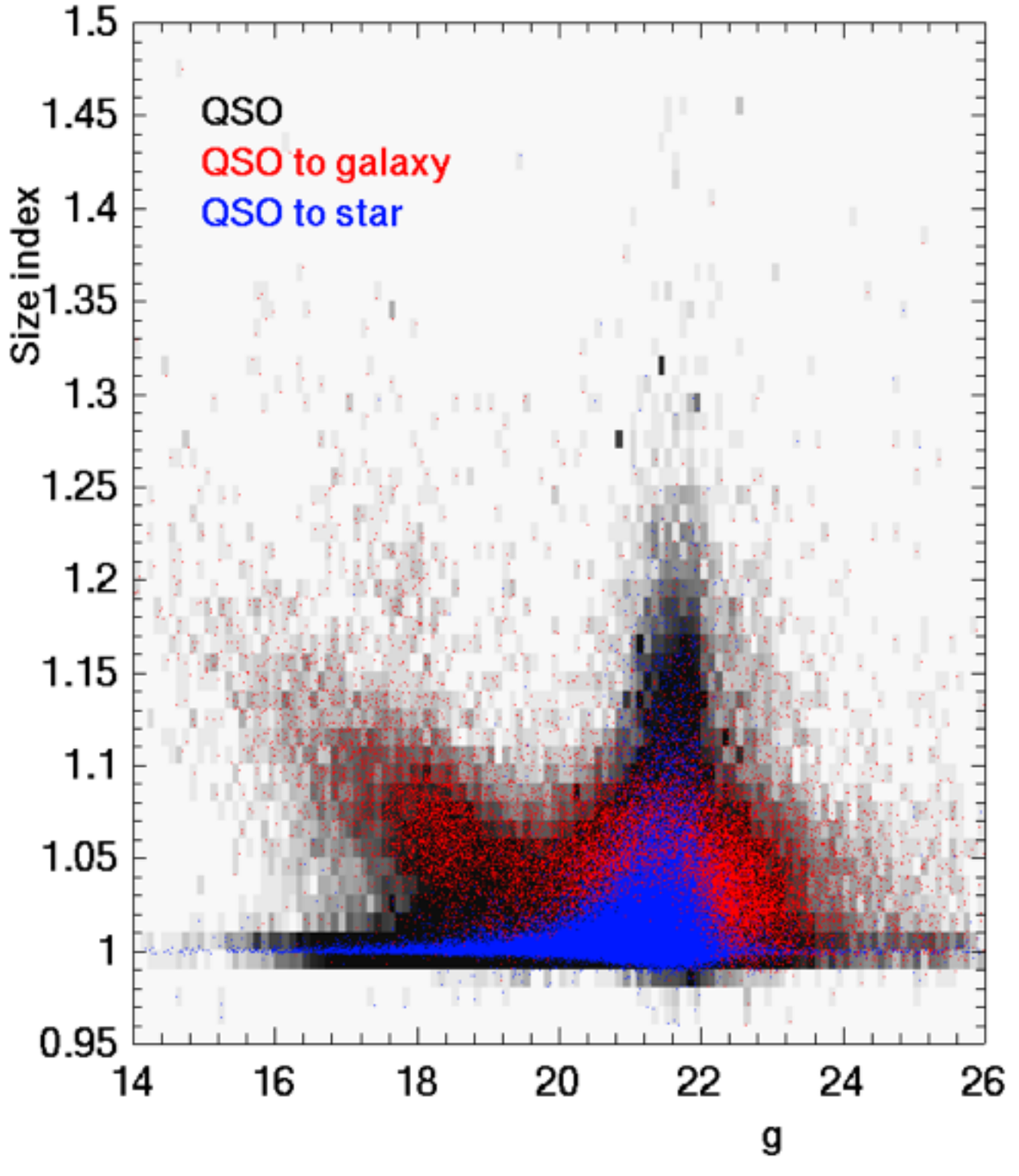
QSOs misclassifications





QSOs misclassifications





QSOs misclassifications