Photometric classification of QSOs from RCS2 using Random Forest

Felipe Barrientos (PUC)

Daniela Carrasco (PUC, U Merlbourne)
Karim Pichara (PUC)

Timo Anguita (UNAB)
Howard Yee (U Toronto)
Mike Gladders (U Chicago)
David Gilbank (SAAO)
+

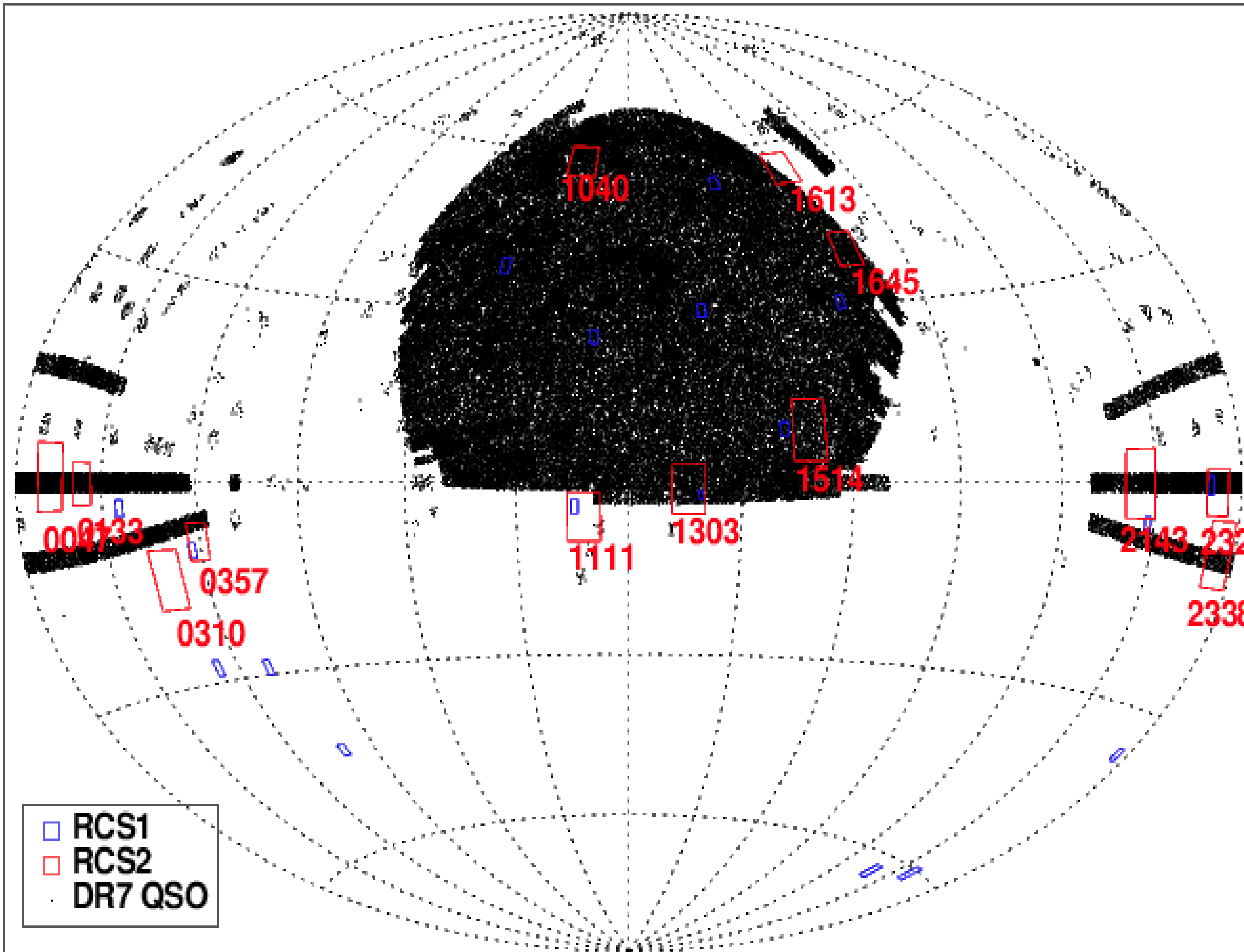
Clusters of galaxies are the most massive objects and are sensitive to the cosmology we live in

We have carried on two large imaging surveys

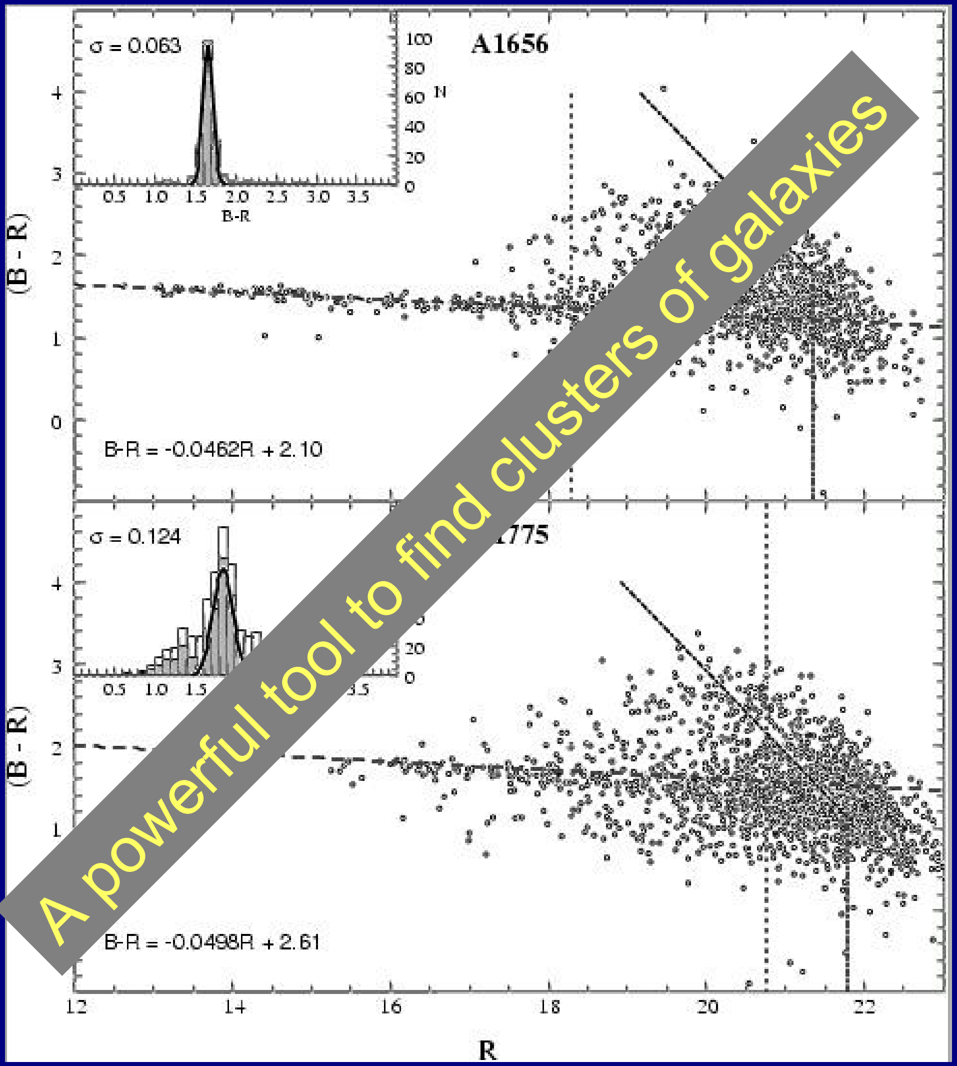
- The Red-sequence Cluster Survey 1 (RCS1)
  - 100 sq deg
  - 2 bands, R and z
  - 1.4 x 10^7 objects
  - ~ 10^3 clusters

- The Red-sequence Cluster Survey 2 (RCS2)
  - 1000 sq deg
  - 3+1 bands, grz+i
  - 1.1 x 10^8 objects
  - N x 10^4 clusters

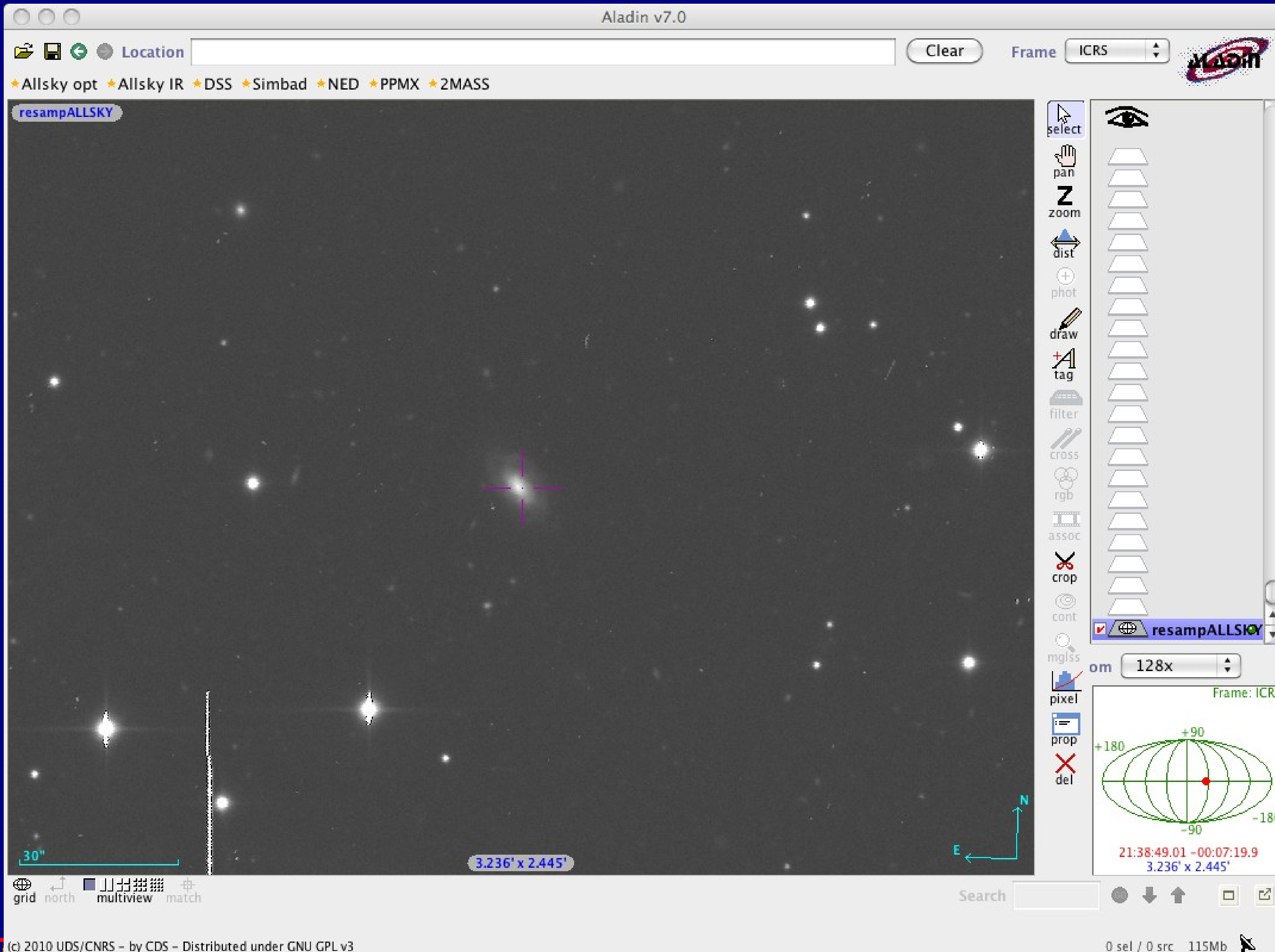
RCS 1 and 2 footprints

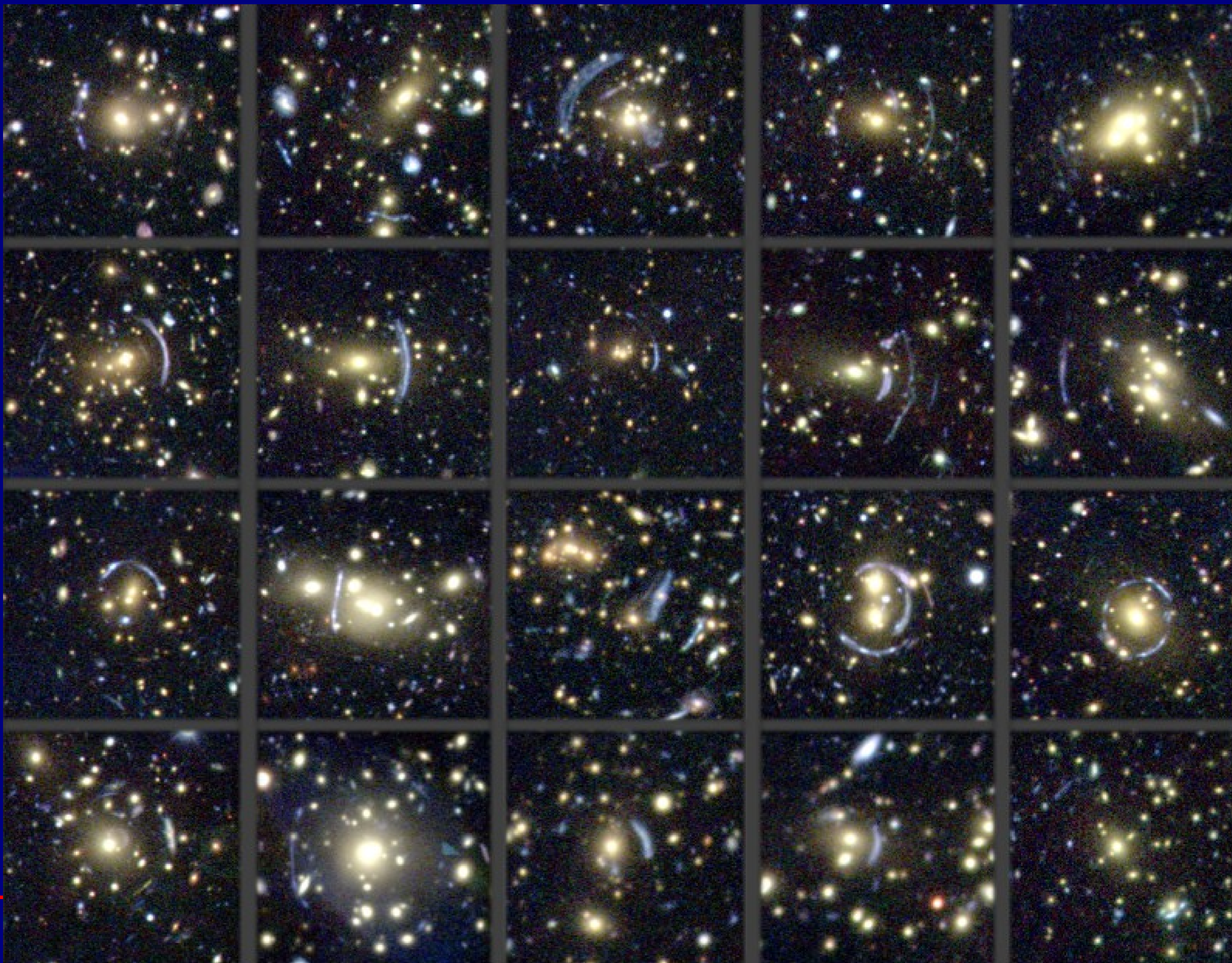


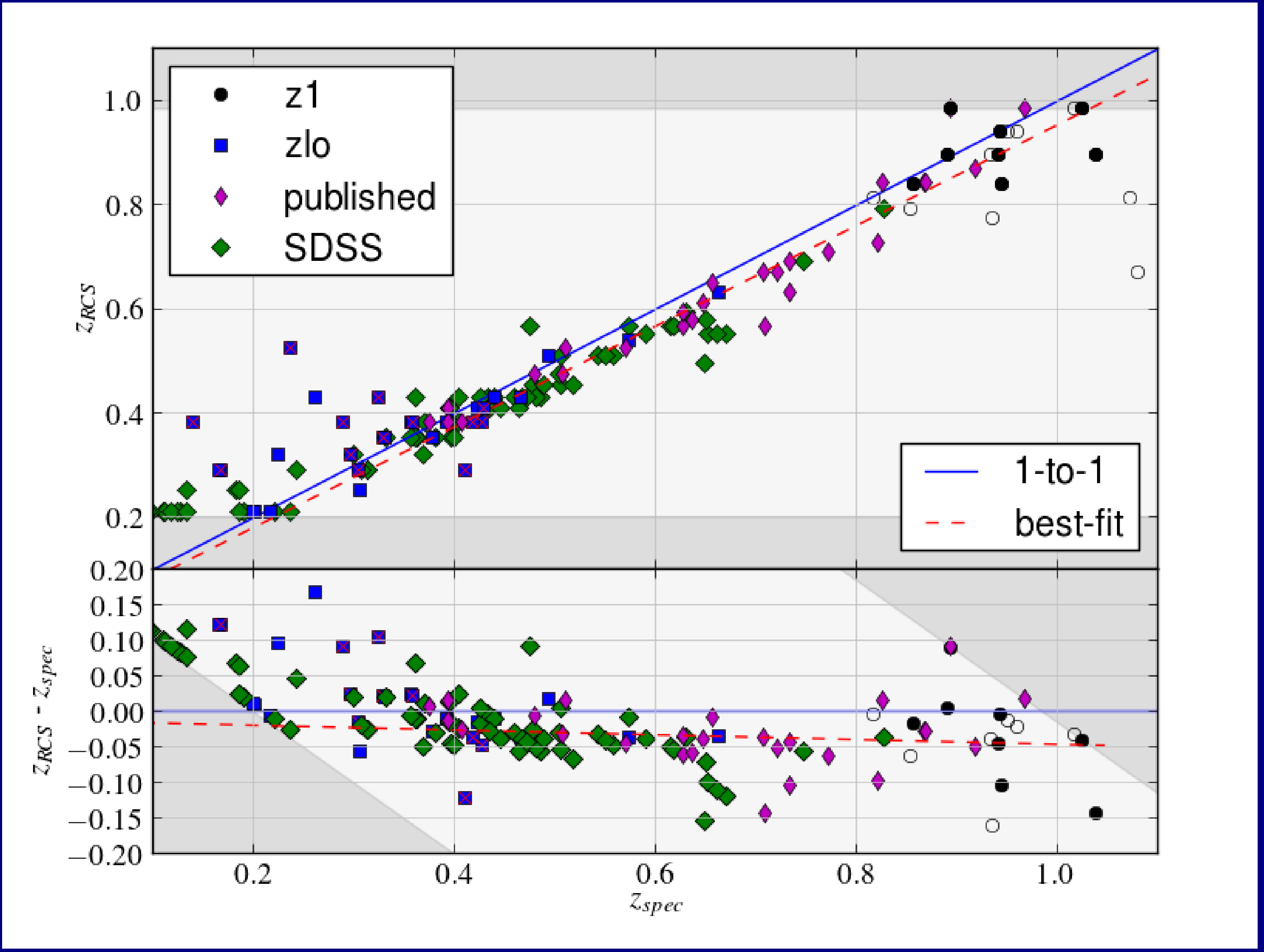
Homogeneity of the colors of the early-type (E/S0) population seen at low and intermediate redshifts, and present in all the clusters found to date.



López-Cruz 1998


Aladin v7.0
Location
Clear
Frame
ICRS
Allsky opt
Allsky IR
DSS
Simbad
NED
PPMX
2MASS
resampALLSKY
select
pan
zoom
dist
phot
draw
tag
filter
cross
rgb
assoc
crop
cont
mglss
pixel
prop
del
resampALLSKY
128x
Frame: ICRS
+90
+180
-90
-180
21:38:49.01 -00:07:19.9
3.236' x 2.445'
30"
3.236' x 2.445'
N
E
grid
north
multiview
match
Search
(c) 2010 UDS/CNRS – by CDS – Distributed under GNU GPL v3
0 sel / 0 src
115Mb



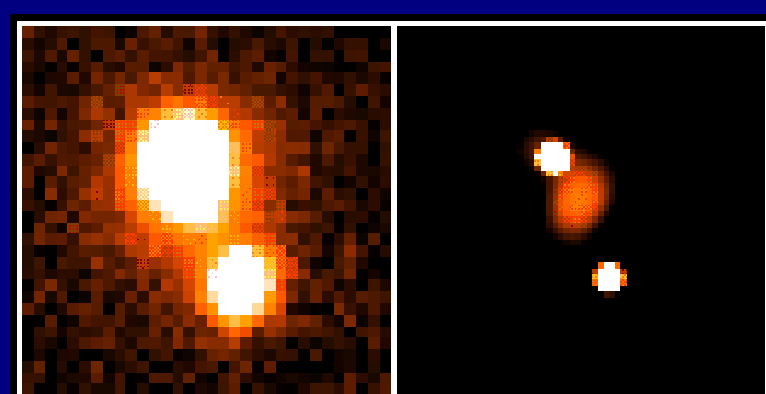


Very good photo-z

Samples of AGNs have a wide range of applications

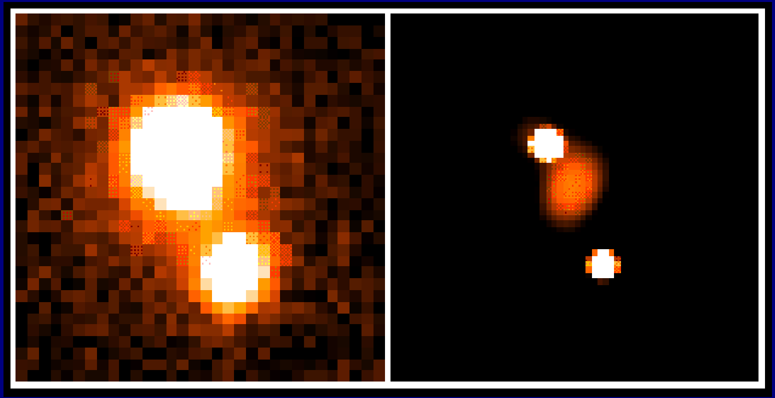
Galaxy evolution
Intergalactic medium
Local physics around black holes
Large scale structure
Astrometric references for local studies
+ ...

One particular application: multiple lensed quasars



Walsh et al (1979)

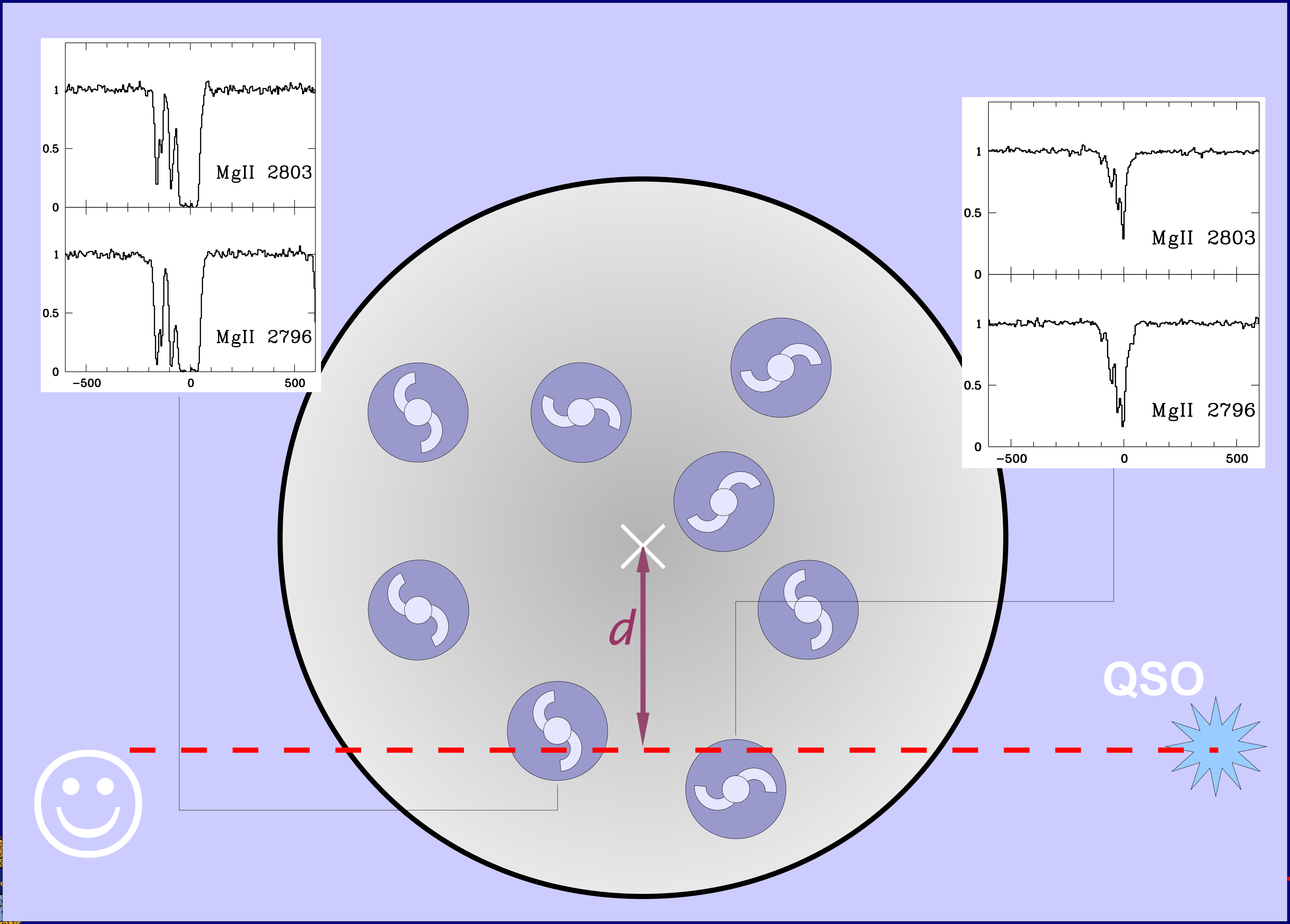
Multiple lensed quasars



Many applications:
- Mass determination of the lens
- Environmental effects
- Determination of the Hubble Constant
- The interstellar medium in lens galaxies
- Cosmography with lens statistics (pairs)
- Natural telescopes

Our goal is to find lensed quasars from our quasar candidates and quasars behind galaxy clusters

QbC: The Quasars behind Clusters Survey

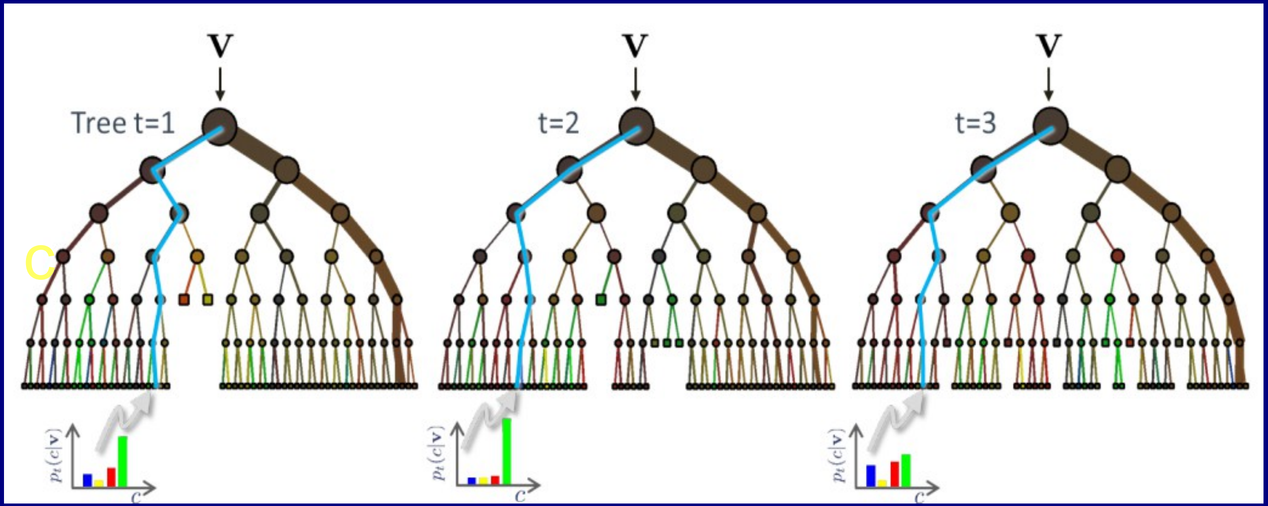


Selection of candidates: Random Forest

is a tree-based classification method (Breiman 2001)

Random Forest uses a training set (objects for which the class is known), and the resulting mapping is applied to objects whose class is not available

The process of building a Random Forest for some number of T trees given training data is:



Random Forest

Let N be the number of objects in the training set. The classifier samples N objects at random to produce S. This sample S will be the training set for the growing tree

- M is the number of attributes (i.e. colors & mags) for each object, and k (<< M) variables are selected randomly during the forest growing

- The predictor variable that provides the best split is used to split the node.

- Each tree is grown to the largest extent possible

- When a new input is entered into the system, it is run down all of the trees. The result is a voting majority.

- 10-fold cross-validation over the training set to estimate the performance.

- Quantify performance:

Predicted Label

Known Label

	Positive	Negative
Positive	True positive	False Negative
Negative	False Positive	True Negative

Predicted Label

Known Label

	Positive	Negative
Positive	True positive	False Negative
Negative	False Positive	True Negative

- Precision: The percentage of positive predictions that are correct

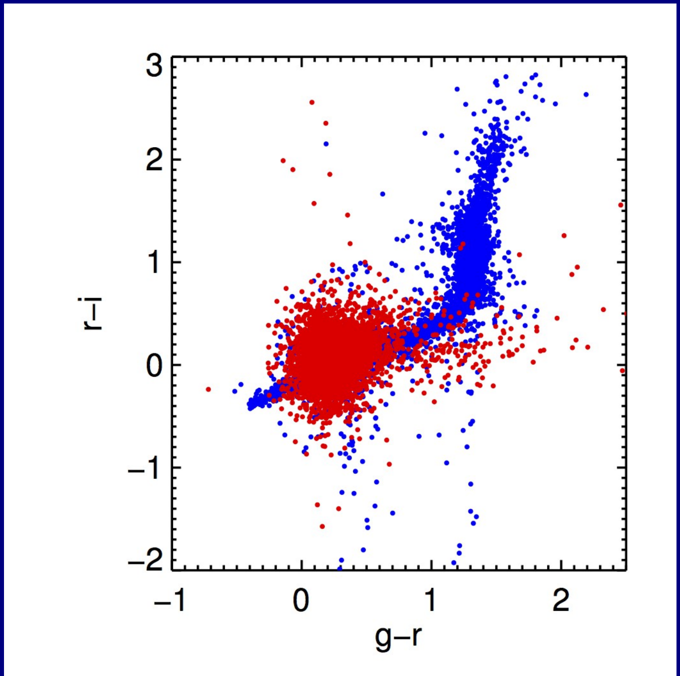
$$\frac{TP}{(TP+FP)}$$

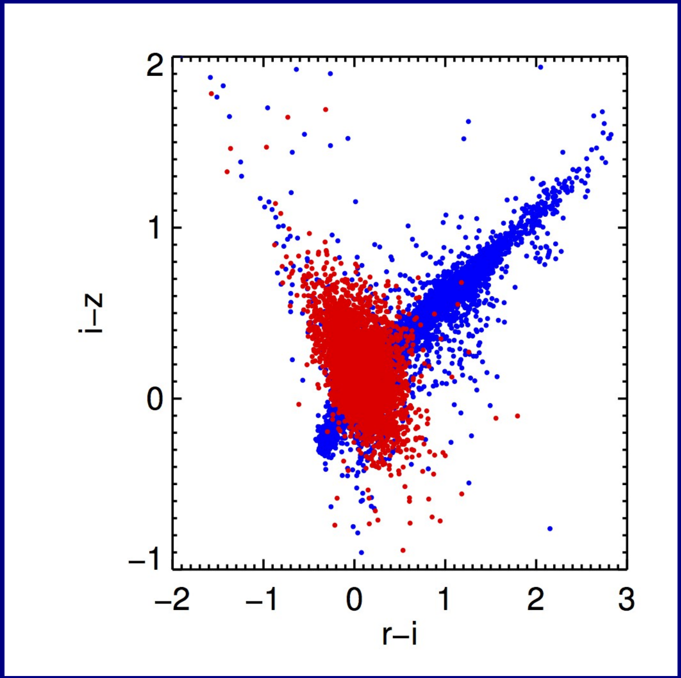
- Recall: The percentage of positive labeled instances that were predicted as positive.

$$\frac{TP}{(TP+FN)}$$

Training Set 1

- Spectroscopically confirmed quasars and stars from SDSS cross-identified in RCS-2.
- 4,916 quasars and 10,595 stars
- Matched with point sources in RCS2 (r < 0.5")
- Features: g, r, i, z, (g-r), (g-i), (g-z), (r-i), (r-z), (i-z)





Stars
QSOs

Test Set 1 (TS1)

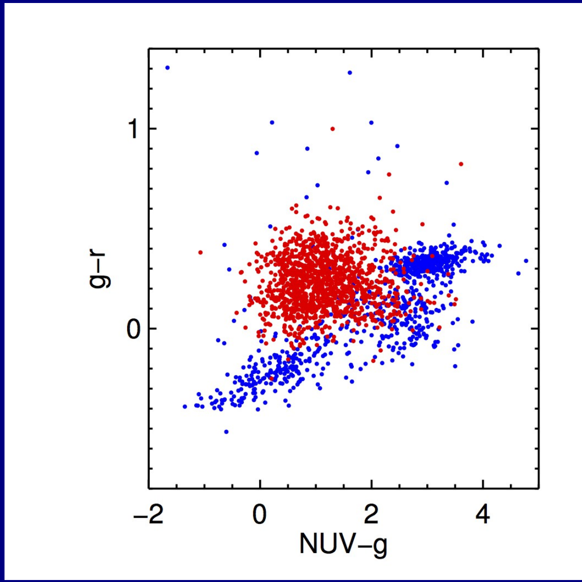
- A total of 1,863,970 objects
- Using 3 features and 70 trees
- 85,085 candidates (<5%)
- Consistent with predictions for LSST

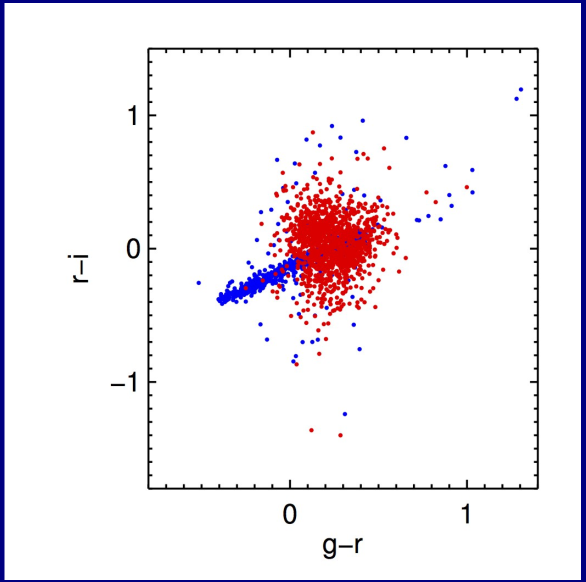
TS1	Recall	Precision
QSO	0.875	0.900
STAR	0.967	0.958

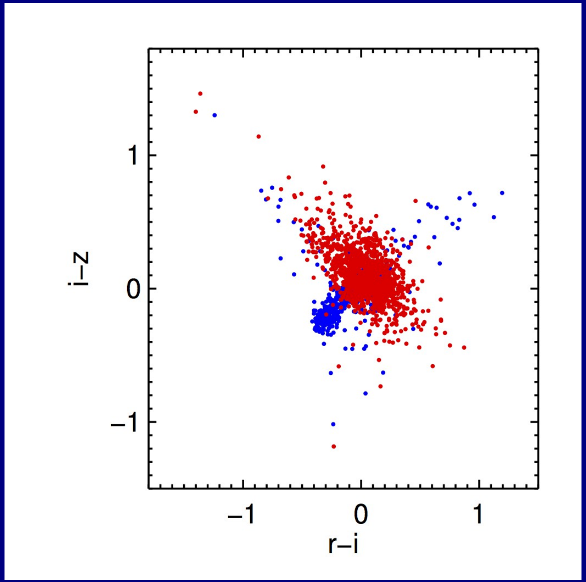
To improve classification add more data: GALEX (Training Set 2)

- NUV (λ 2267 A) photometry
- 1,228 quasars and 815 stars
- Matched with point sources in RCS2 (r < 2.0")
- Features: NUV, g, r, i, z, (g-r), (g-i), (g-z), (r-i), (r-z), (i-z)+ NUV-g, NUV-r, NUV-i, NUV-z
- About 1.5 mag brighter than Training Set 1

Stars
QSOs







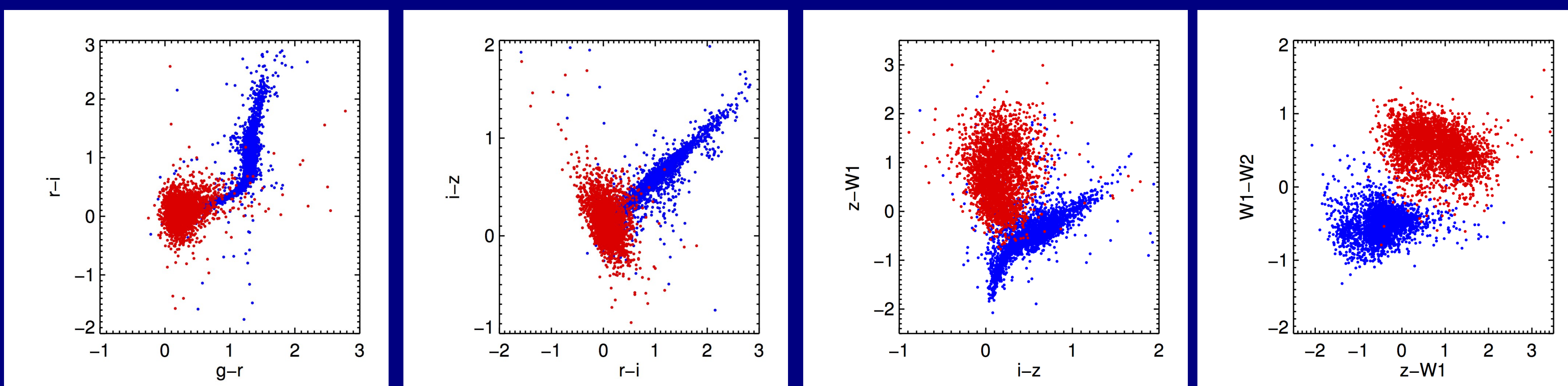
Test Set 2 (TS2)

- A total of 16,898 objects
- Using 9 features and 100 trees
- 6556 candidates (38.8%)

TS2	Recall	Precision
QSO	0.969	0.976
STAR	0.963	0.953

... And more data: WISE (Training Set 3)

- W1 (3.4 μm) and W2 (4.6 μm) photometry
- 2,748 quasars and 2,749 stars
- Matched with point sources in RCS2 (r < 2.0")
- Features: W1, W2, g, r, i, z, (g-r), (g-i), (g-z), (r-i), (r-z), (i-z)+ g-W1, g-W2, r-W1, r-W2, i-W1, i-W2, z-W1, z-W2, W1-W2



Stars
QSOs

Test Set 3 (TS3)

- A total of 242,902 objects
- Using 7 features and 60 trees
- 21,713 candidates (8.94%)

TS3	Recall	Precision
QSO	0.993	0.992
STAR	0.992	0.993

Summary

From a total set of 1,863,970 objects to classify:

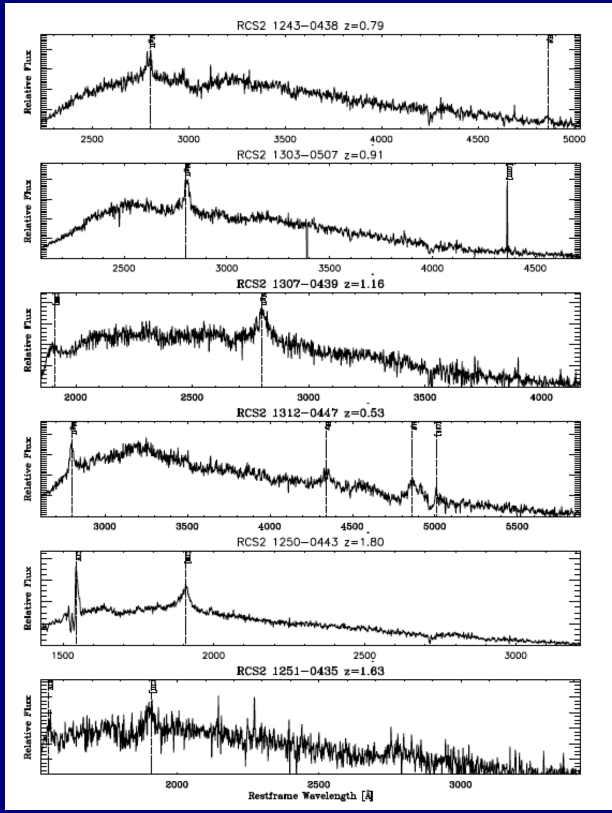
- 85,085 quasar candidates from TeS1
- 6,556 quasar candidates from TeS2
- 21,713 quasar candidates from TeS3

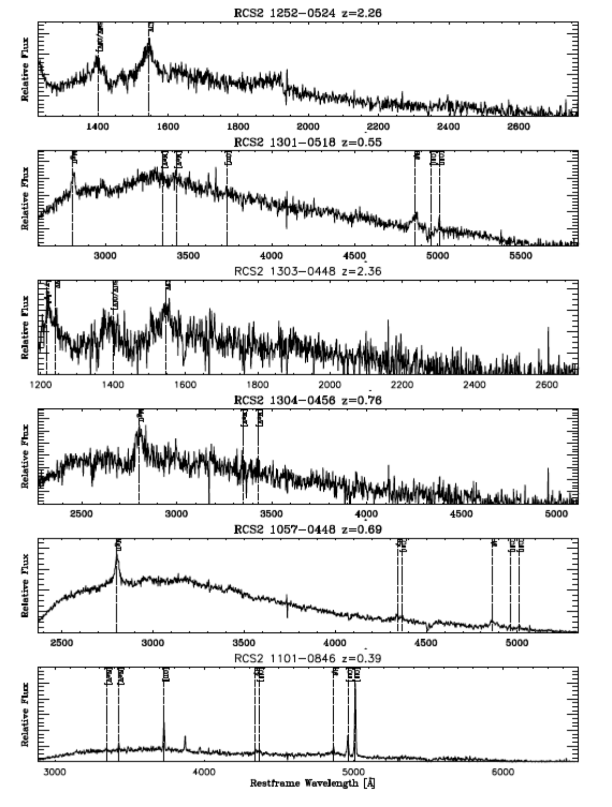
91,842 new quasar candidates from RCS-2

- 3,600 objects classified as quasars from the three test sets
- 13,691 objects classified as quasars from the two test sets
- 74,551 objects classified as quasars from only one test set

Spectroscopic confirmation

du Pont Telescope + Boller & Chivens spectrograph
Mag lim r ~ 19





Object	r mag	C1	C2	C3	Priority	Spec	Redshift
RCS2 1243-0438	17.45	Q	Q	Q	1	Quasar	0.79
RCS2 1303-0507	17.90	Q	Q	Q	1	Quasar	0.91
RCS2 1307-0439	18.01	Q	-	Q	2	Quasar	1.16
RCS2 1312-0447	17.75	Q	Q	Q	1	Quasar	0.53
RCS2 1250-0443	18.15	Q	Q	Q	1	Quasar	1.80
RCS2 1251-0435	18.20	Q	Q	Q	1	Quasar	1.63
RCS2 1252-0524	18.00	Q	-	Q	2	Quasar	2.26
RCS2 1301-0518	18.51	Q	Q	Q	1	Quasar	0.55
RCS2 1303-0448	18.53	Q	-	Q	2	Quasar	2.36
RCS2 1304-0456	18.34	Q	Q	Q	1	Quasar	0.76
RCS2 1057-0448	17.62	Q	Q	Q	1	Quasar	0.69
RCS2 1100-0313	18.18	Q	Q	Q	1	Star	-
RCS2 1101-0846	18.18	Q	Q	Q	1	Quasar	0.39
RCS2 1106-0821	18.19	Q	Q	Q	1	Quasar	1.43
RCS2 1310-0458	18.40	Q	-	Q	2	Quasar	2.65
RCS2 1303-0505	18.64	Q	-	Q	2	Quasar	0.48
RCS2 1305-0435	18.48	Q	Q	Q	1	Galaxy	-

15-16 out of 17 are confirmed QSOs

The selection does work!!

Results in Carrasco et al (submitted)

This process can be applied to other similar datasets

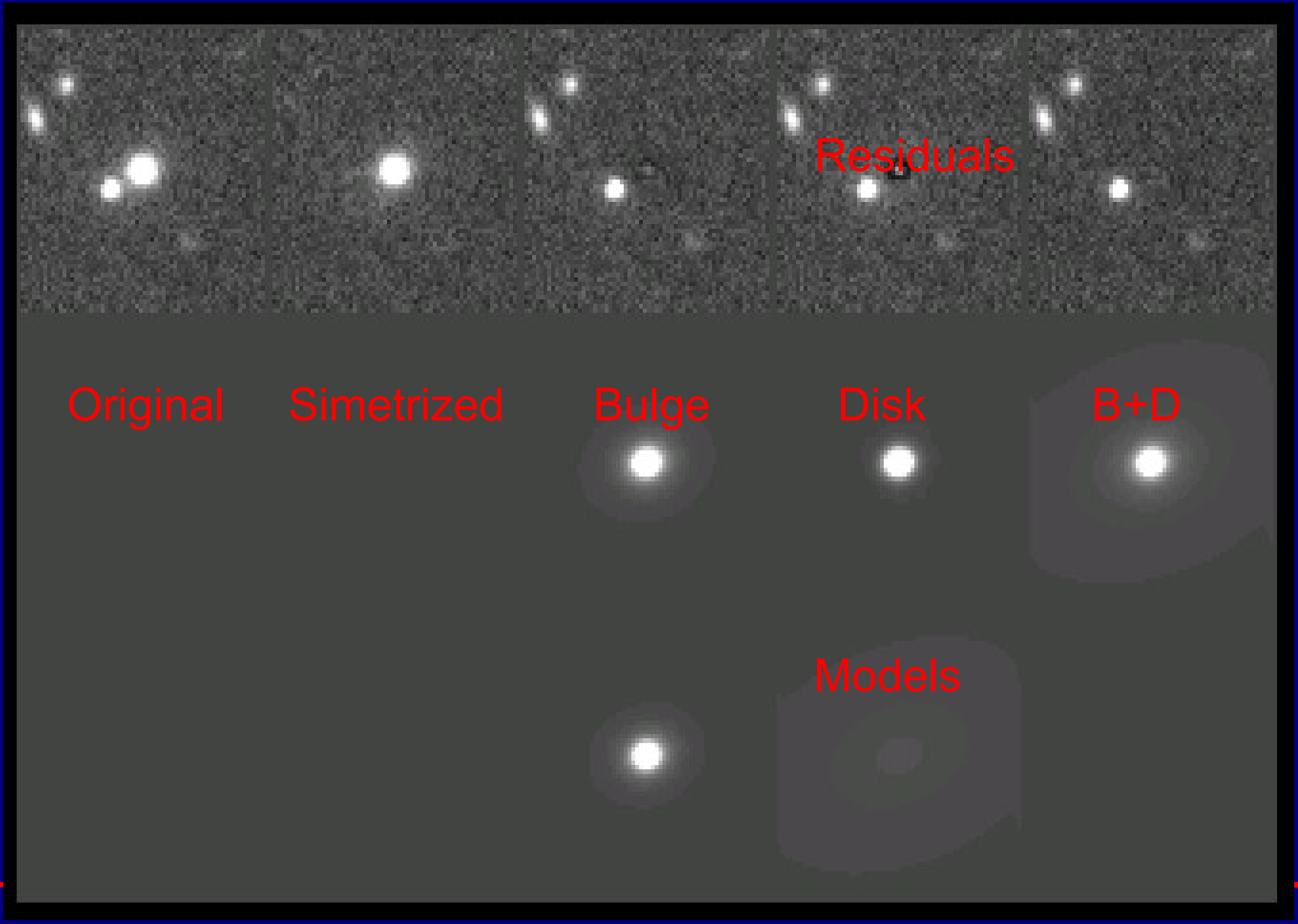
(Galaxy morphology)

Thanks

Morphology using a 2-D galaxy light profile fitting algorithm (L-M chi sq minimization – 5 parameters – own GALFIT)

This accounts for seeing, providing a galaxy size (i.e

- r_e for an $r^{1/4}$-law or h for an exponential disk)
- + VISUAL CLASSIFICATION



This approach is not longer possible with large datasets

LM minimization is too expensive (~ 1 min / galaxy)

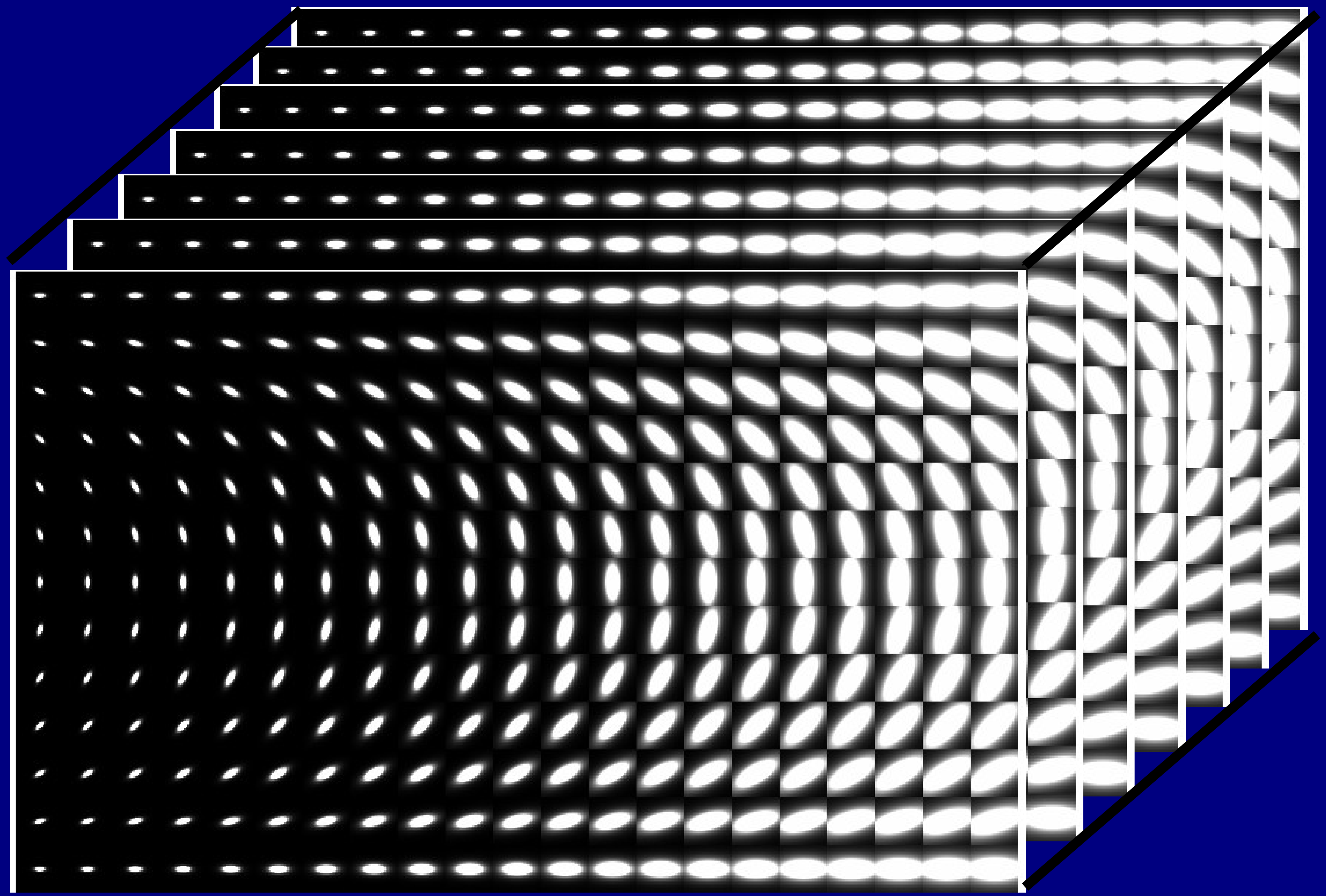
New algorithm

- Populate the parameter space with a discrete number of models

- move convolution out of the fitting process

Model cube
- position angle
- size
- axial ratio

~ 2000 models



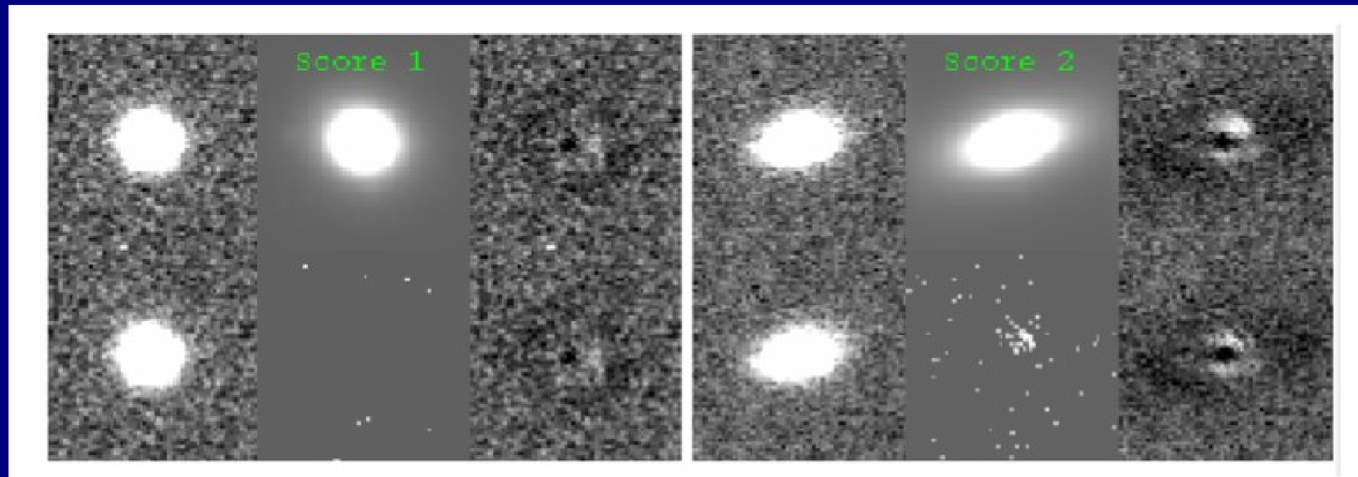
Convolved with the PSF of each CCD

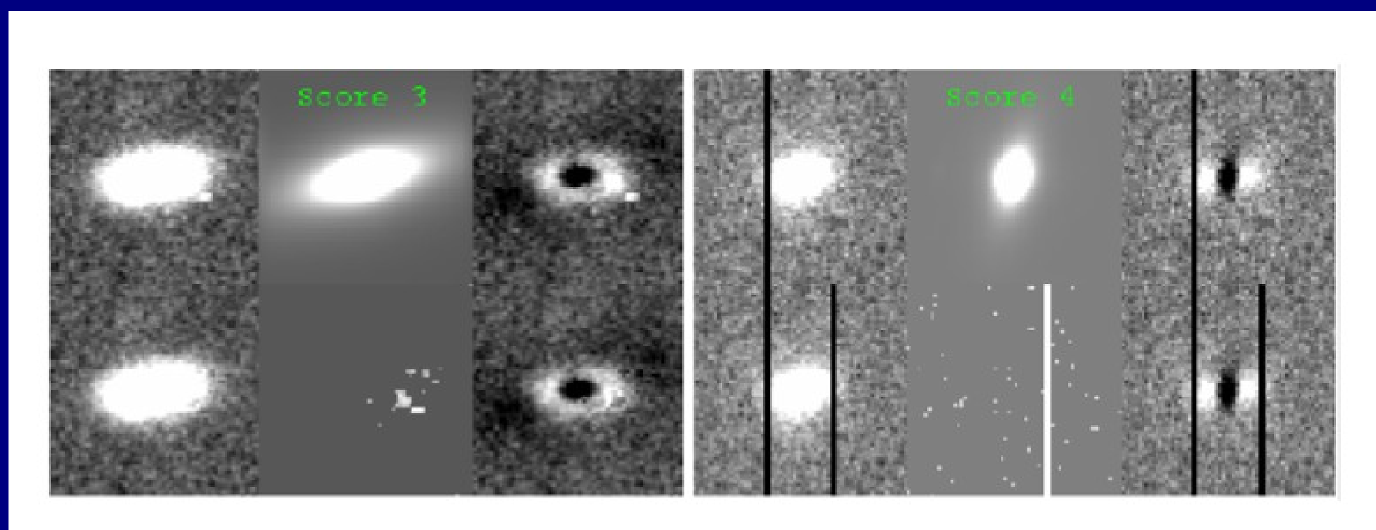
Code implemented in Python

First large test.

5 million objects fitted
r-band only
Using 200 cores
In 36 hours

~ 4000 galaxies visually inspected


Score 1
Score 2


Score 3
Score 4