Data mining for High redshift Radio loud Quasars

D. Tuccillo[1], R. McMahon[2], I.J. Gonzalez-Serrano[1], R. Carballo[3], M. Banerji[2]

1 - Istituto de Fisica de Cantabria (Spain) 2- Institute of Astronomy, Cambridge (UK),3-Dpto. de Matematica Aplicada, Univ. de Cantabria

15–18.10 2012 ESO HQ, Garching, Germany

50 YEARS 1962–2012

>>**Introduction**: Interest in Radio Loud quasar selection and Surveys used in our investigations

1) FIRST-SDSS [D. Tuccillo, I.J. Gonzalez-Serrano, R. Carballo]: Selection of a list of Radio Loud quasars candidates with 3.6<z<4.4. Results of their observation

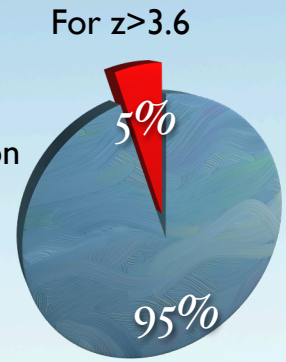
2) FIRST-UKIDSS- SDSS [D. Tuccillo, R. McMahon, I.J. Gonzalez-Serrano, M. Banerji]: Selection of a list of Radio Loud quasars at higher redshift or very red.

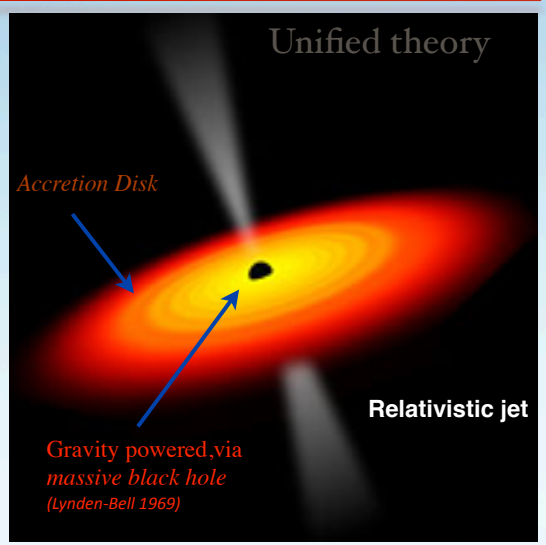
>>**Conclusions**

AGNs — Seyfert, QSOs

- Radio Quiet QSOs
- Radio Loud QSOs

The fraction of RL QSOs is a strong function of magnitude, redshift and radio Flux limit (Jiang, L., and al. 2007)

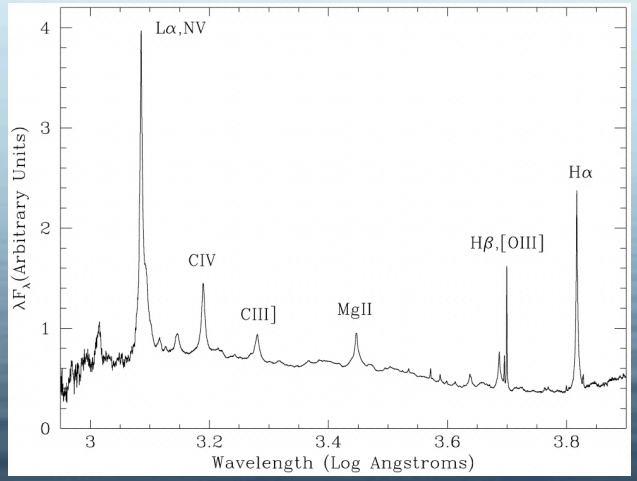




To identify a QSO we need their..Spectra

....but

as long as we can't have spectra of the whole sky we need candidates **Selection** criteria



Composite spectra of a RL QSOs from FIRST Bright Quasar Survey (Brotherton, M. and al. 2001)

Image from Hubble deep field telescope, covering an area of the sky 1.5 arc-minutes wide and 1.125 high

Classic criteria of selection: Color-Color plots → More efficient selections involves use of **multiwavelength surveys** and **data mining**

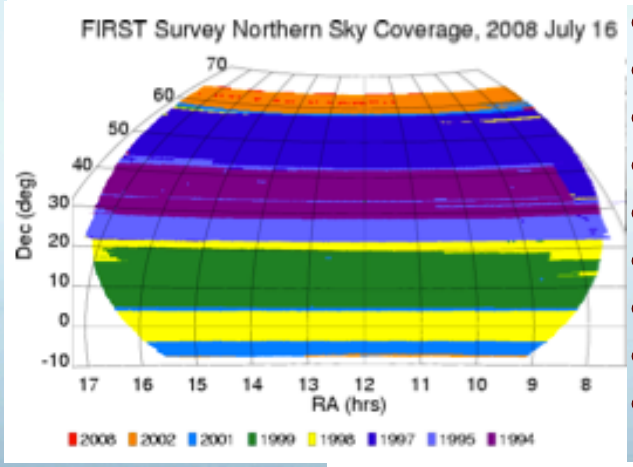
"Data mining is the process of automatically or semi-automatically discover useful information in large data repositories that might otherwise remain unknown"
[Data Mining- P.N. Tan et al.]

Surveys Used in our investigation

FIRST -- **F**aint **I**mages of the **R**adio **S**ky at **T**wenty-cm --

Selection based on radio survey-> less contamination stars, useful in finding highly reddened QSOs

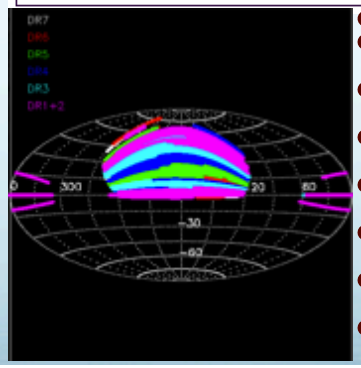
FIRST (2008 release) — Area covered: 9,033 sq. deg; 811,117 sources



FIRST

- *RADIO Survey, area chosen to coincide with SDSS
- *Use the Very Large Array (VLA)[Socorro,NewMexico]
- *Survey at 20 cm
- *detection limit ~1mJy
- *90% sources confidence error circles of radius <1" at survey threshold

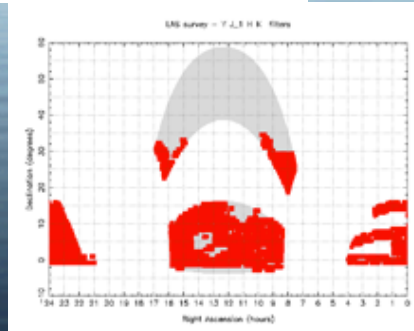
SDSS DR7 — Area covered: 11,663 sq. deg; 357 million distinct objects



The Sloan Digital Sky Survey

- *The major OPTICAL multi filter imaging and spectroscopic survey
- *5 wavebands : u,g,r,i,z
- *Apache Point Observatory [NewMexico]

UKIRT infrared Deep Sky Survey



UKIDSS DR8 — Area covered: about 2,500 sq. deg; 69 million objects

- *Near infrared, follows 2MASS and anticipates VISTA
- *4 wavebands :Y,J,H,K,
- *The UK Infrared Telescope [Mauna Kea, Hawaii]

1) investigation : FIRST- SDSS for RL QSOs selection at 3.6 <z < 4.4

Matching SDSS and FIRST Surveys, 1,5 arcsec

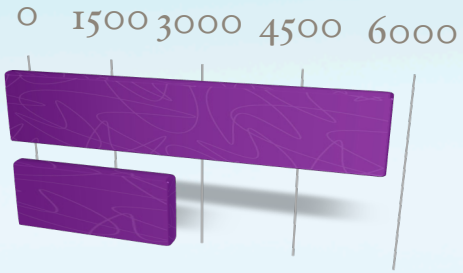
OVERLAPPING area: 8,073 sq. deg.

→ **222,517 sources** →

Pre-selection criteria

- 15<psfMag_r<20.2 (AB mag)
- Morphological Type =Star
- Exclusion photometrical FATAL ERRORS: Bright, Saturated, Edge,Blended
- psfMagError<0.2 at least in one band

Clean sample(9,140)



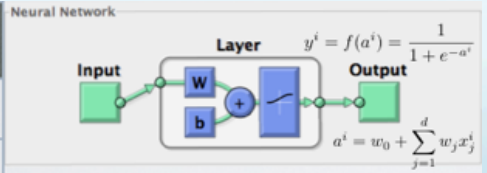
labelled (5784) **[with spectra]**

unlabelled (3,356) **[without spectra]**

- To train the Neural Network (NN) we use the labelled sample, divided in the two subsamples as "**target**" and "**not target**".

Supervised Feed Forward NN used as classifier

separation variables
Absolute r mag
(u-g)color
(g-r)
(r-i)
(i-z)
radio-opt sep
log10(FluxRadio)



- The output y gives the degree of similarity with the class $0 \leq y \leq 1$

Class assignment

$x \in C_1$ if $y(x) \geq y_c$

- C_1 is the class of the candidates.
- We test efficiency and completeness on the testing group as function of y_c

Training and test method: "leave one out method"

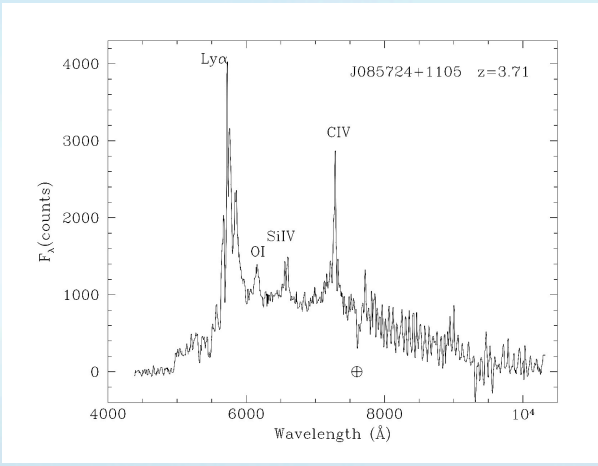
The target sample is composed by the ***3.6<z<4.4 RL QSOs*** *of the V quasar catalog (Schneider et al. 2010) matching our pre-selection criteria.* ***Just 73*** *sources out of a total of 5,784 <--> ~1.3 %*

1) Results : FIRST- SDSS for RL QSOs selection at 3.6 <z < 4.4

Between 2008 and 2012
*we observed 58 candidates and **30** of them resulted to be actual QSOs with 3.6<z<4.4*

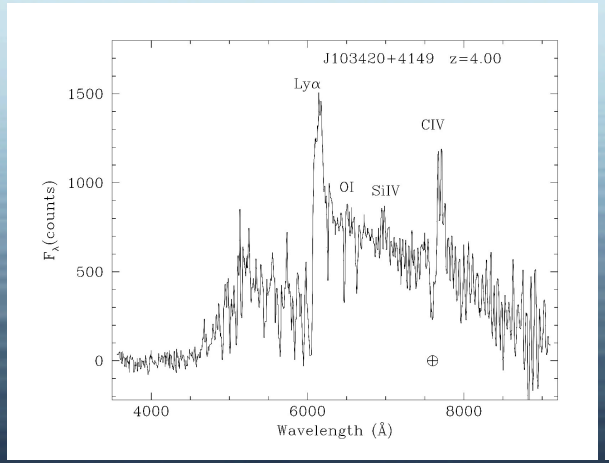
~52%
success rate

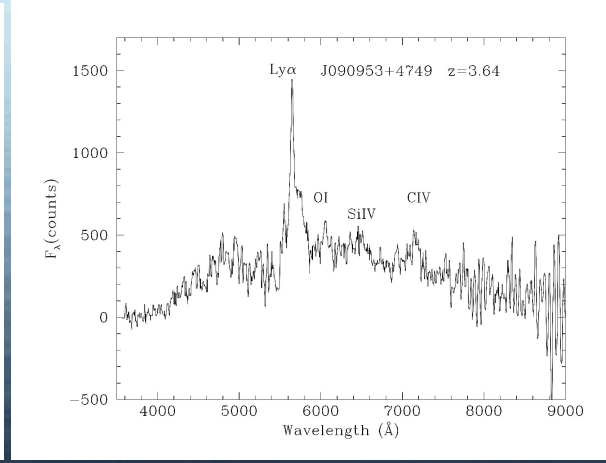
- Spectra of **3** (out of a total of **7**) of the RL QSO observed at the NOT telescope in March 2012



***NB** the observed candidate were selected with different NNs. The last version has*

Completeness ~ 97%
Efficiency ~ 66%



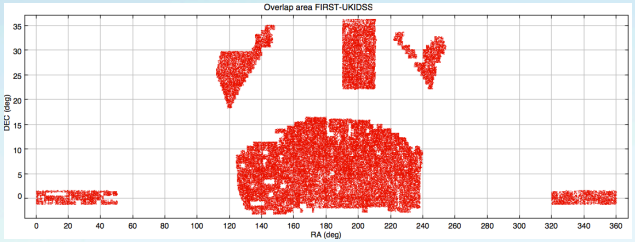


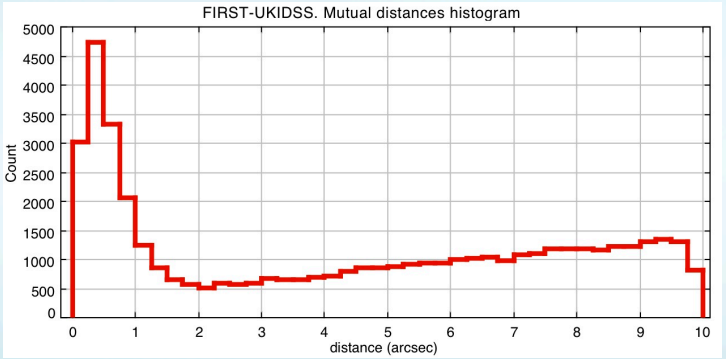
II) investigation : FIRST- UKIDSS-SDSS, for higher z QSOs or very red

FIRST 2008 — (1.5 arcsec) — UKIDSS DR9 (LAS)

76,541 matched sources

detected at least in one band (Y,J,H,K)





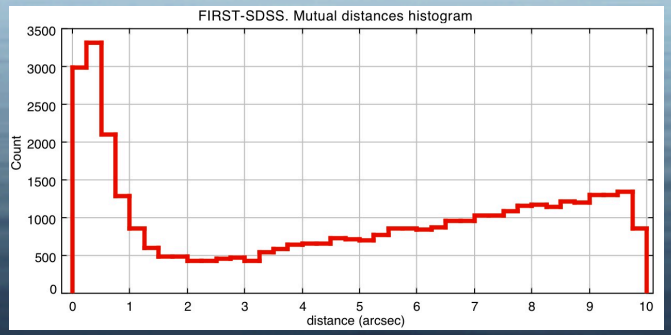
(10 arcsec)

SDSS DR7

Considering instead **1.5 arcsec, other 8,273 undetected**

68,186 detected on SDSS DR7

8,355 undetected on SDSS DR7



Removing the false negatives from subsample detected in FIRST and UKIDSS but not in SDSS

In order to have a subsample of sources real and authentically missed in SDSS

8,355 sources

List of techniques applied for cleaning up the sample:

1) excluding all the sources masked in SDSS

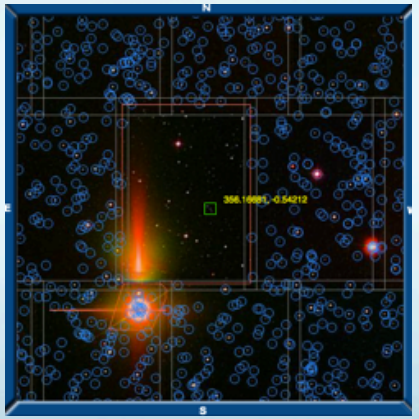
BLEEDING and BRIGHT mask
images areas covered by large saturated stars (approximate radius >=1 arcmin), with significant bleeding columns, which are expected to cause problems in photometry

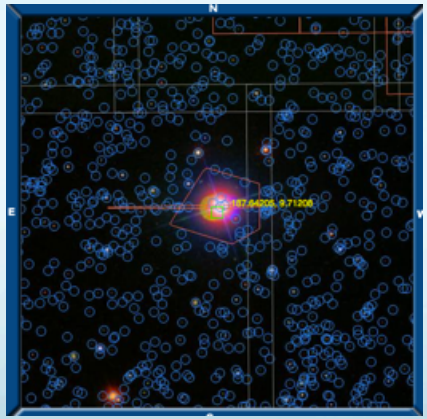
TRAIL masks
imaging areas covered by trails caused by meteors, satellites and the like.

HOLE masks
holes in the surveys and fields where the data quality is unacceptable

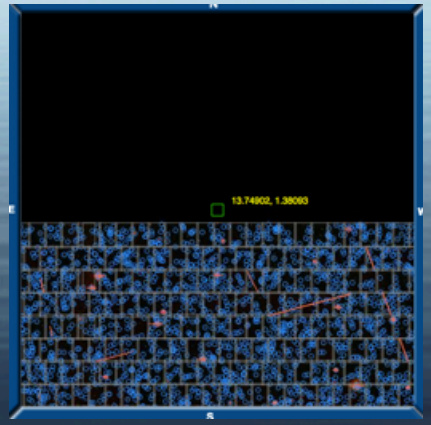
~400 masked

- *Two examples of sources masked. The first is also out field: sometimes the mask bound the entire field*





2) excluding all the sources outside of the SDSS footprint, that we call "out of field"



- The Masks polygons in red, the squares correspond to the fields and the blue circles are the photometric objects

3) excluding all the sources detected in SDSS DR8

~800

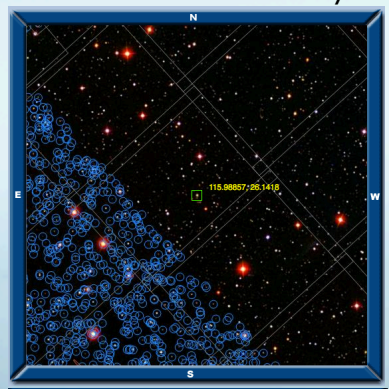
- *On the left an example of source "out of field" in SDSS DR7*

List of techiques applied for cleaning up the sample:

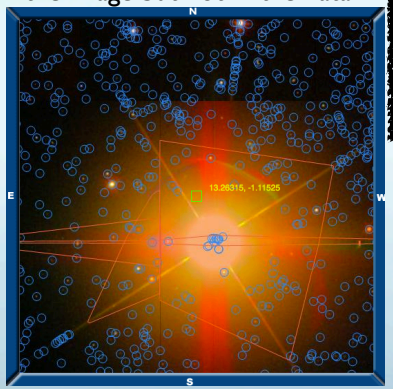
4) we found a sample of sources with secondary SDSS counterpart but NOT correspondent primary object. We exluded also this sample

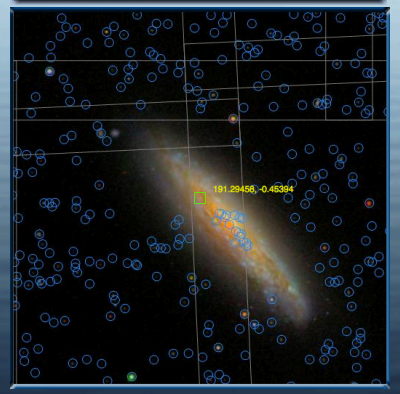
...BUT
besides all these criteria, visual inspection indicates some problems with the SDSS data:

- objects with "image" in SDSS but not Photometry



- objects that results masked in the image but not in the data

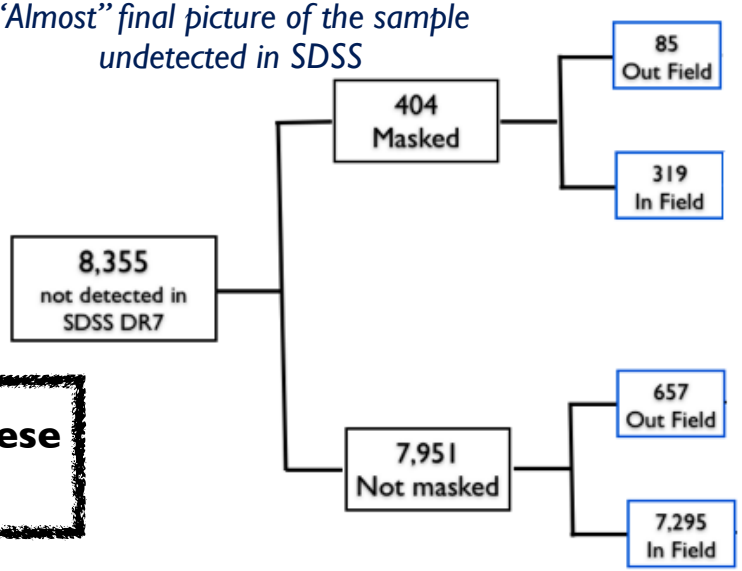




- Objects undetected but in "huge" galaxies

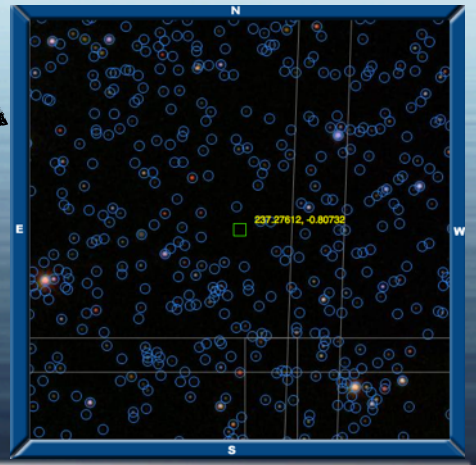
To exclude all these sources..

"Almost" final picture of the sample undetected in SDSS



1) we downloaded ALL the images from the SDSS image tool

2) we developed and used a JAVA program combined with a **decision tree** method to finally select REAL...



Ending with..

~6,200 sources

~75 %

Removing the false negatives from the remaining ~6000 sources

~6,200 sources

~ 45% detected at least in TWO UKIDSS band

~ 55% detected just in ONE UKIDSS band

- **~6% Y-only**
- **~4% J-only**
- **~10% H-only**
- **~80% K-only**

List of techniques applied to determine how many of them were real and not junk in UKIDSS

1) distribution of the cross match distances
2) Using UKIDSS flags: ErrBits and AperErr
3) Use of other Y apertures
4) Visual inspection of the UKIDSS images
5) Match of the sample with WISE

..we concluded that

most of the Y-only are spurious

~80% of the ONE band only are real

most of the K-only are real

~90% of sources detected at least in TWO bands are real

...But what are we searching for?

1

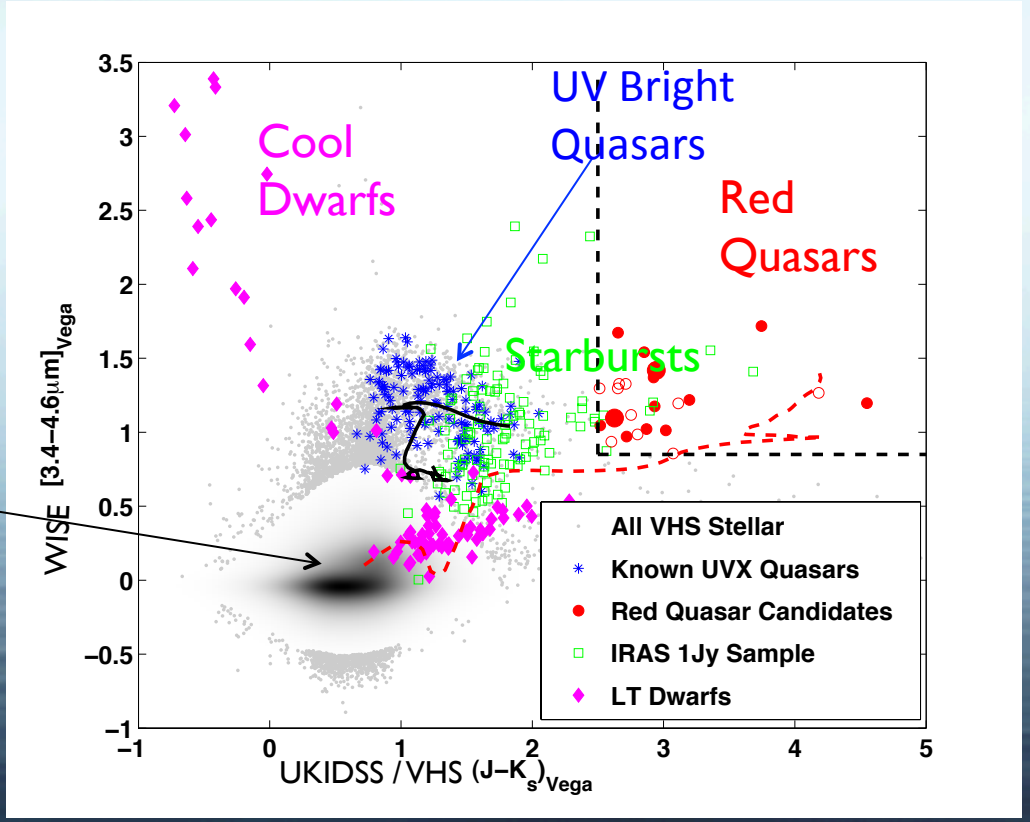
Very High-z red RL QSOs

2

Dust-reddened broad line quasars

(Banerji M. et al. 2012

Heavily Reddened Quasars at z ~2: A Transitional Phase in AGN Evolution

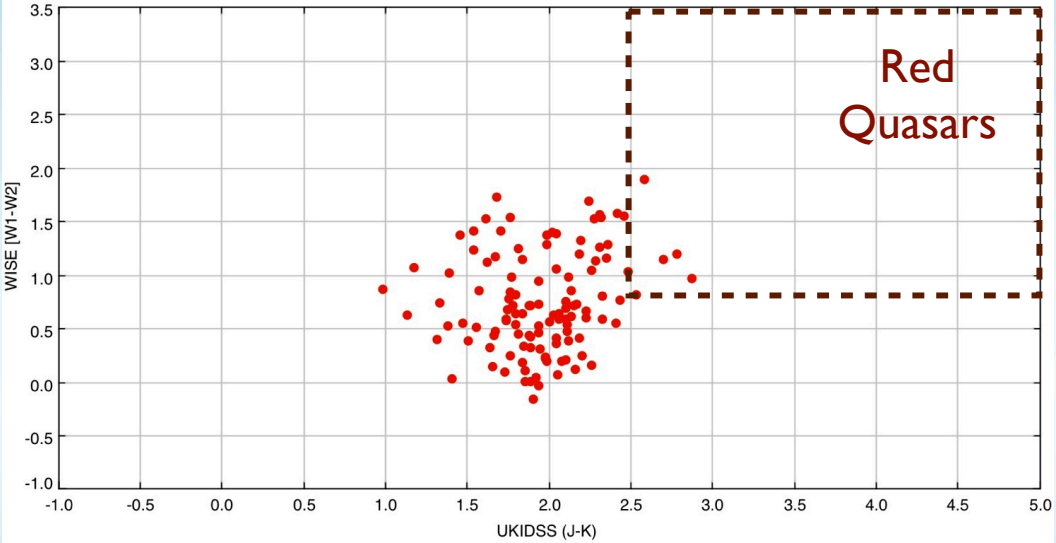


Banerji, McMahon et al. 2012b

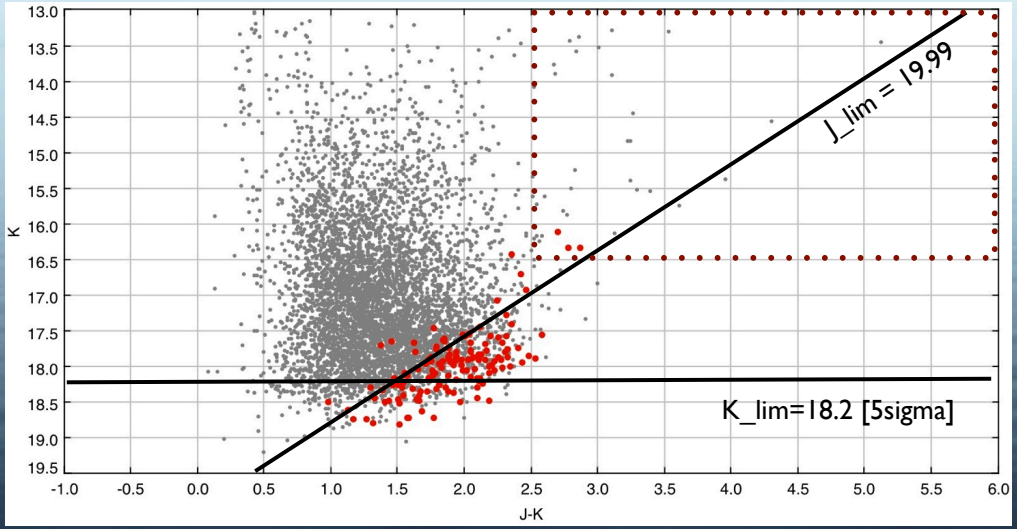
our interesting subsamples

sources with $(J-K>2.5)$	Galaxy	Star
Detected in SDSS	125 (79)	50 (13)
Undetected in SDSS	16(0)	5 (3)

In parenthesis the number of sources with K<= 16.5



Star-like sources. In gray sample detected in SDSS (~5,300), in red undetected in SDSS (136)



Star-like sources undetected in SDSS (117 on 136) detected in WISE

1) FIRST-SDSS [D. Tuccillo, I.J. Gonzalez-Serrano, R. Carballo]: we developed a very efficient multi-wavelength Machine Learning technique for Radio Loud quasars selection. The completeness of the selection allows tests for the completeness of SDSS and precise computation of Luminosity Function. Technique also interesting in itself because transferable in other contexts

2) FIRST-UKIDSS- SDSS [D. Tuccillo, R. McMahon, I.J. Gonzalez-Serrano, M. Banerji]: we carefully selected a sample of Radio-Infrared objects undetected in the Optical. A sample of objects potentially very interesting. We want now to understand their nature, like for example for the real bright sources detected just in one UKIDSS band.