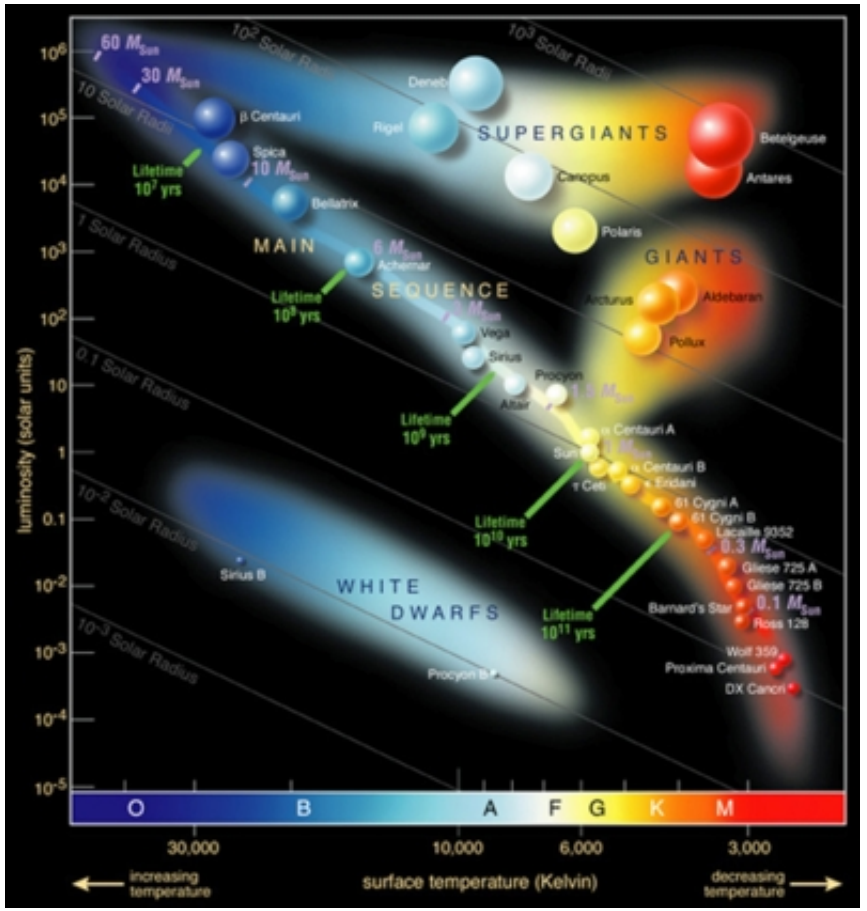
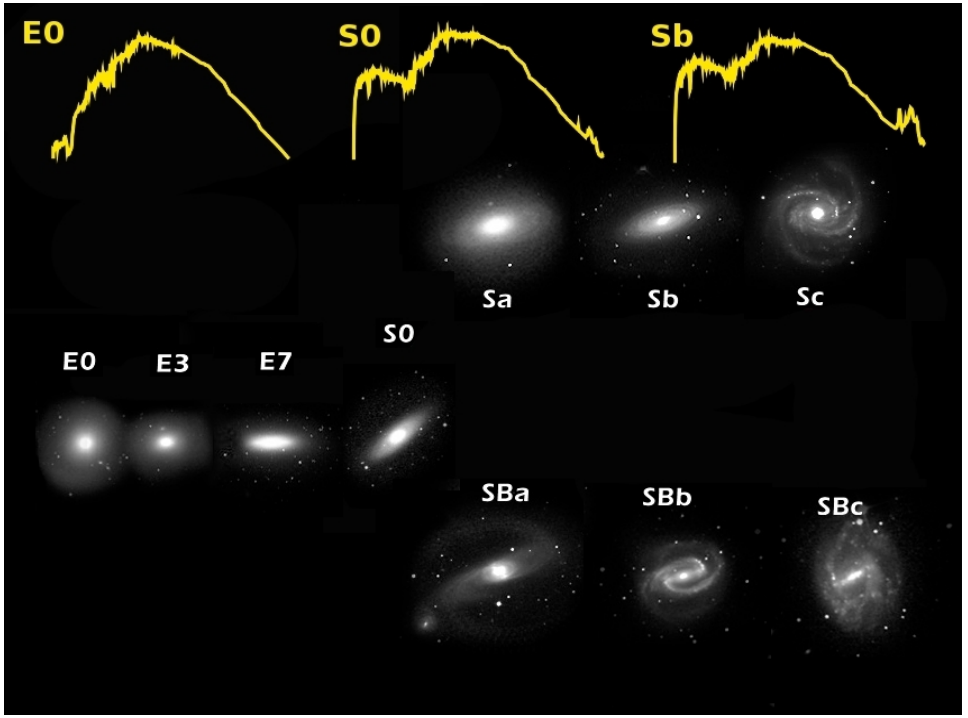
UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'astronomie

AGN identification: what lies ahead

Sotiria Fotopoulou
Sotiria.fotopoulou@unige.ch

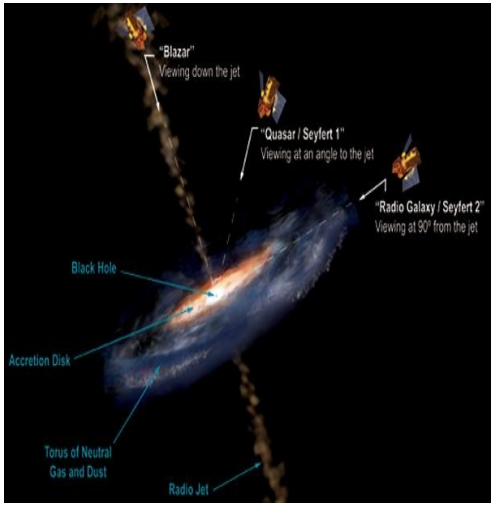
From stars to galaxies ...





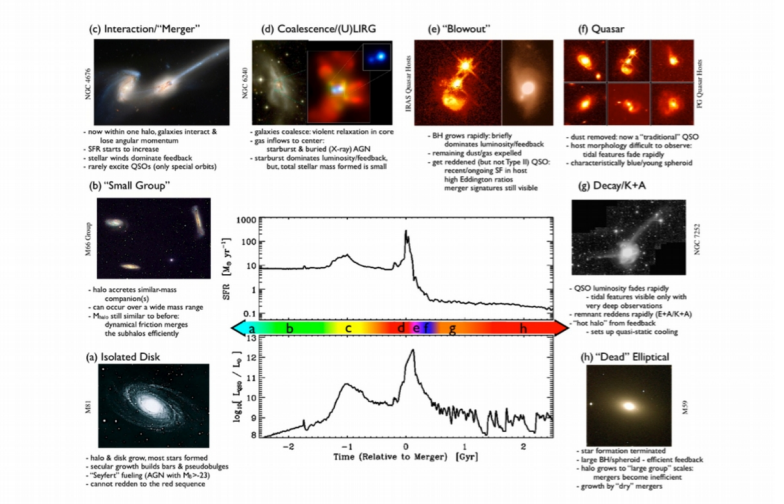
… to AGN

orientation: sets stage



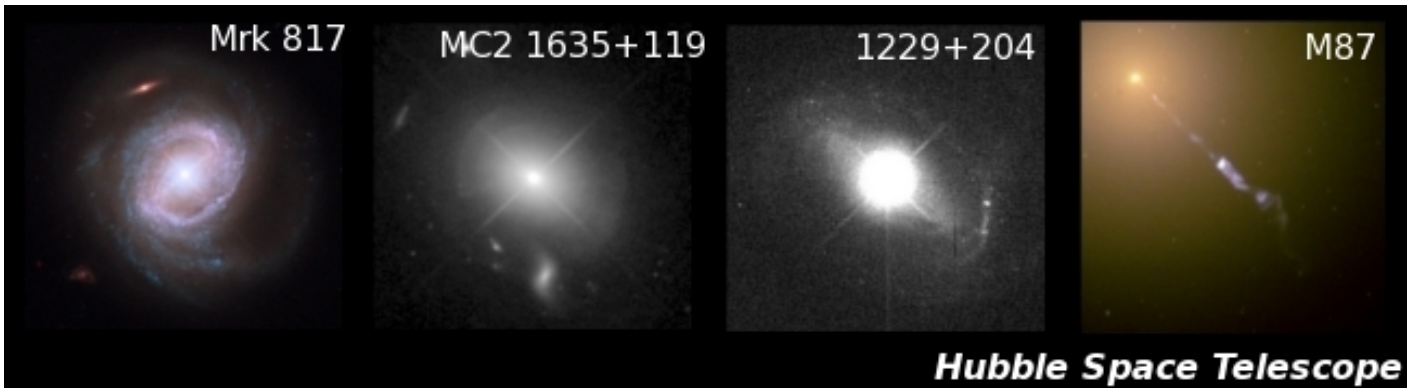
+

galaxy mergers, winds, etc: evolution



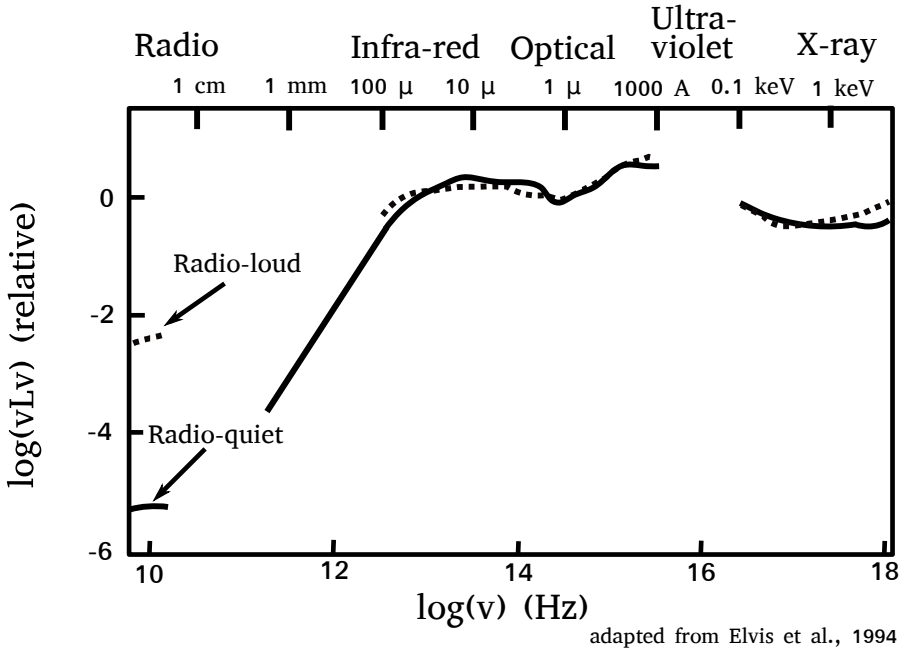
+

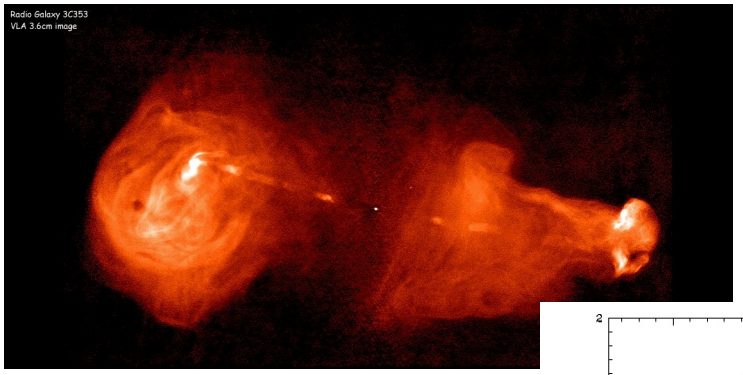
host galaxy: coevolution



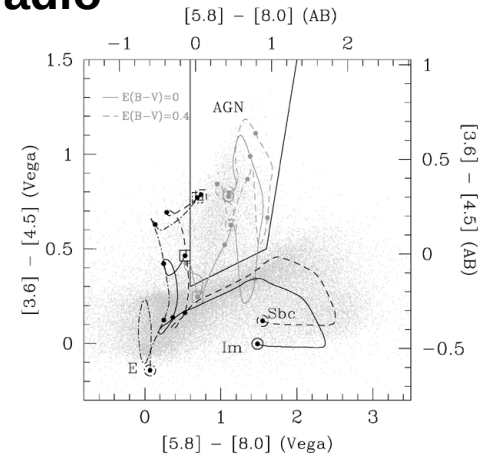
AGN

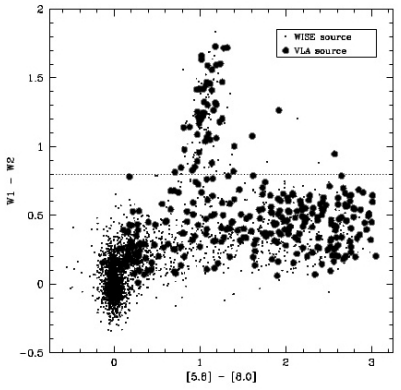
- **Radiation across the EM spectrum.**
- **myriads selection criteria**
- **classification depends on observations**



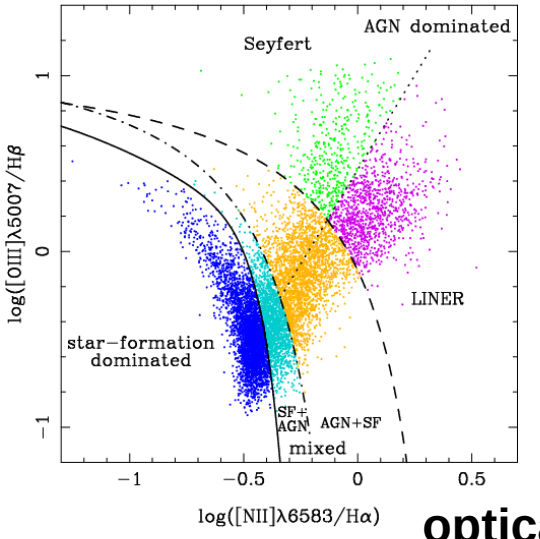


radio

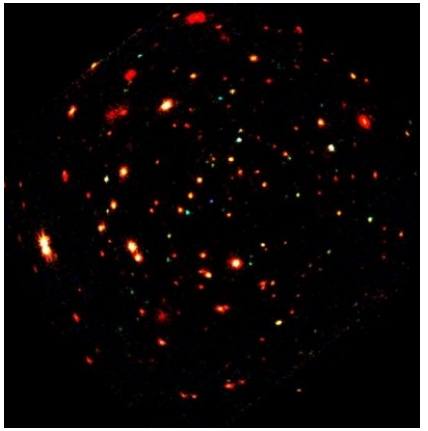




mid-IR



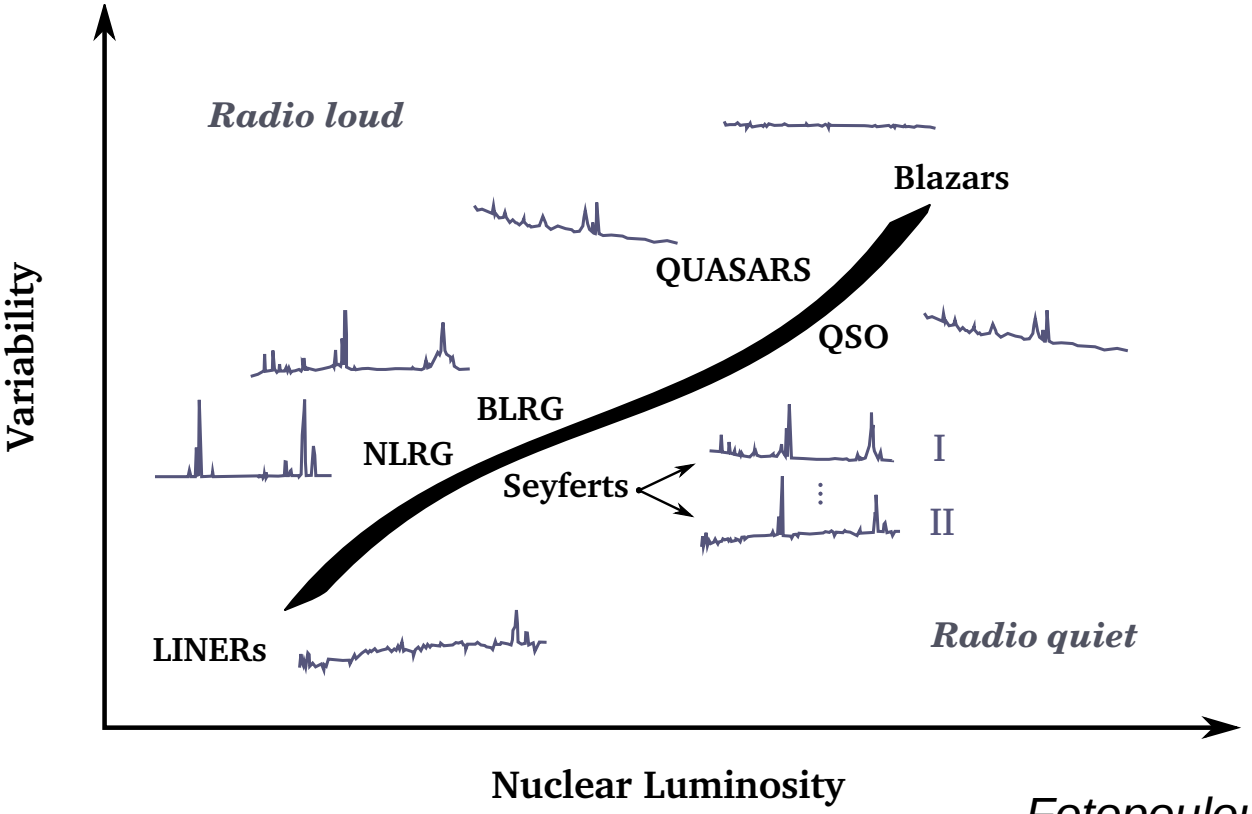
optical



X-rays

Question of the week:

- Is my source an AGN?

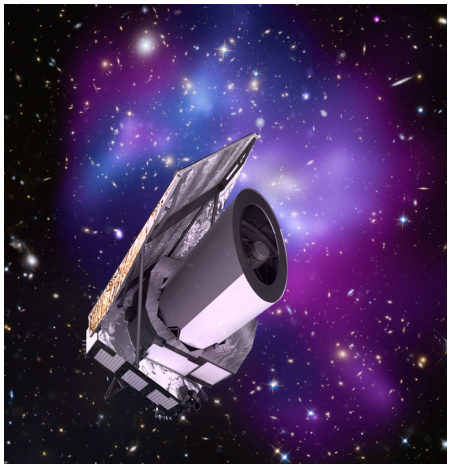


Fotopoulou, PhD Thesis


Find the AGN!
Quite unassuming in a single filter*

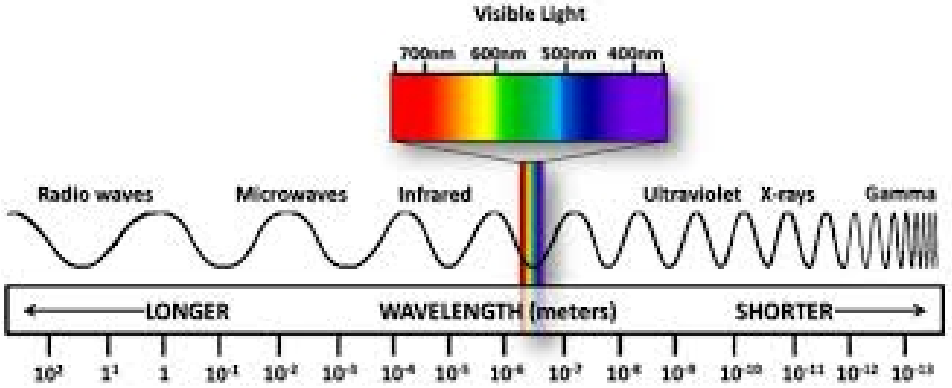
Planned surveys are massive!

Euclid: 400 sq. deg./month
10 billion sources



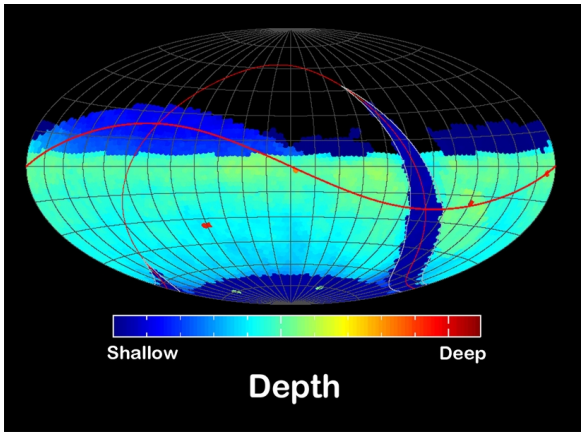
Powerful, fast, automated classifcation is needed.

SKA: 160TB/sec



eROSITA millions AGN

LSST: 30 TB/night



DES: 400GB/night

Question of the week:

- Is my source an AGN?

Question of the week:

- What is the probability that my source has a percentage of AGN emission?

Machine learning identification

1. Define question

pre-processing

- data gathering
- data cleaning
- class definition
- labeling

2. Solve problem

model

- select algorithm
- build neural network/trees
- test
- verification

3. Use new machinery

application

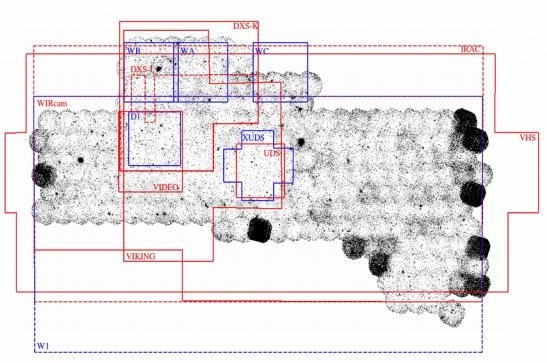
- apply solution to other data
- very fast!

XXL

"The XXL survey: VI. The 1000 brightest X-ray point sources"
Fotopoulou et al., A&A, 592 (2016) A5

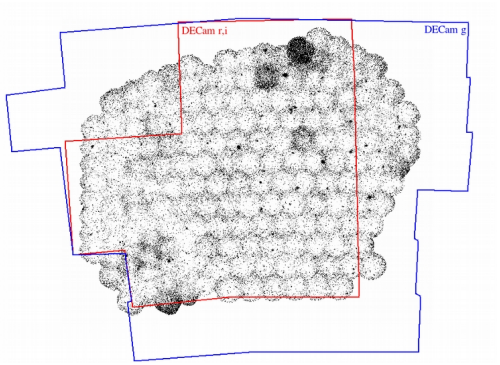
- **2 x 25 sq. deg.**
- **6Ms XMM Newton – avg 10 ks**
- **F[2-10 keV] ~ 1e-15 erg/s/cm^2**
- **22 000 detections**

XXL-N

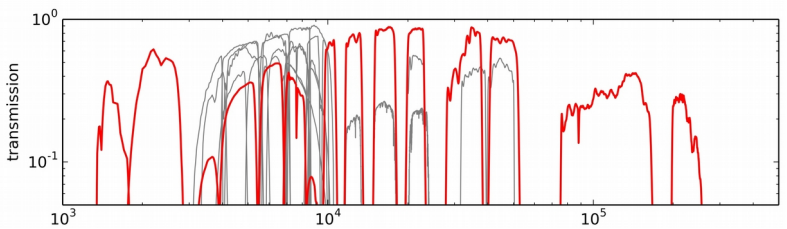


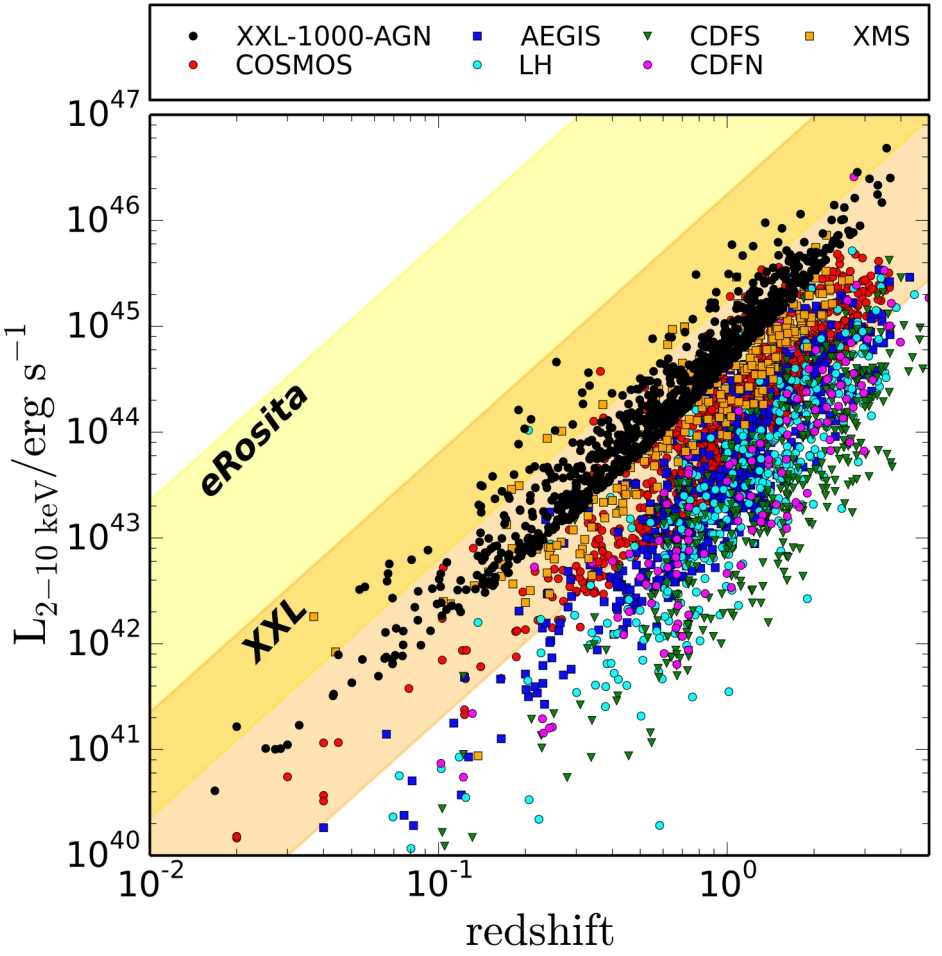
4.5M sources
14 000 X-ray

XXL-S



3.4M sources
12 000 X-ray

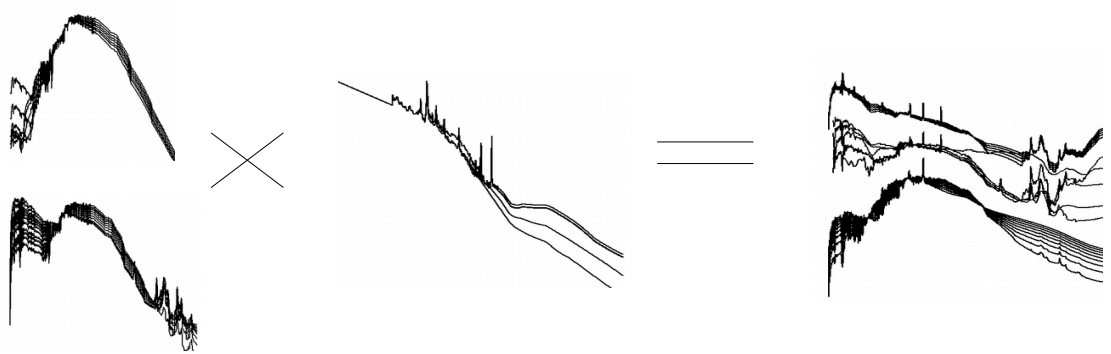


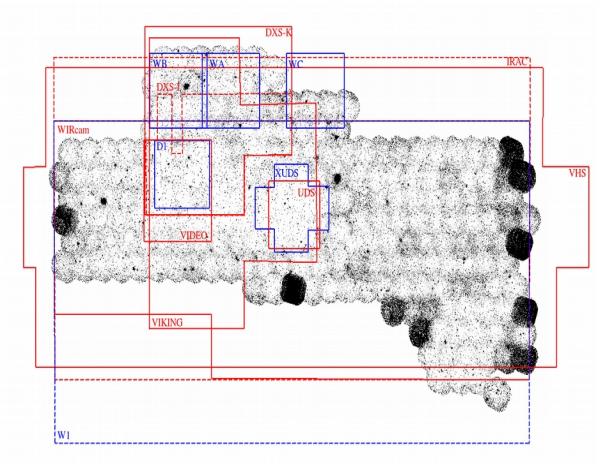


Task:
Predict fraction of AGN emission

XXL-N

1. **48 000 sources with secure spec-z**
2. **1830 hybrid SED models: gal - qso**

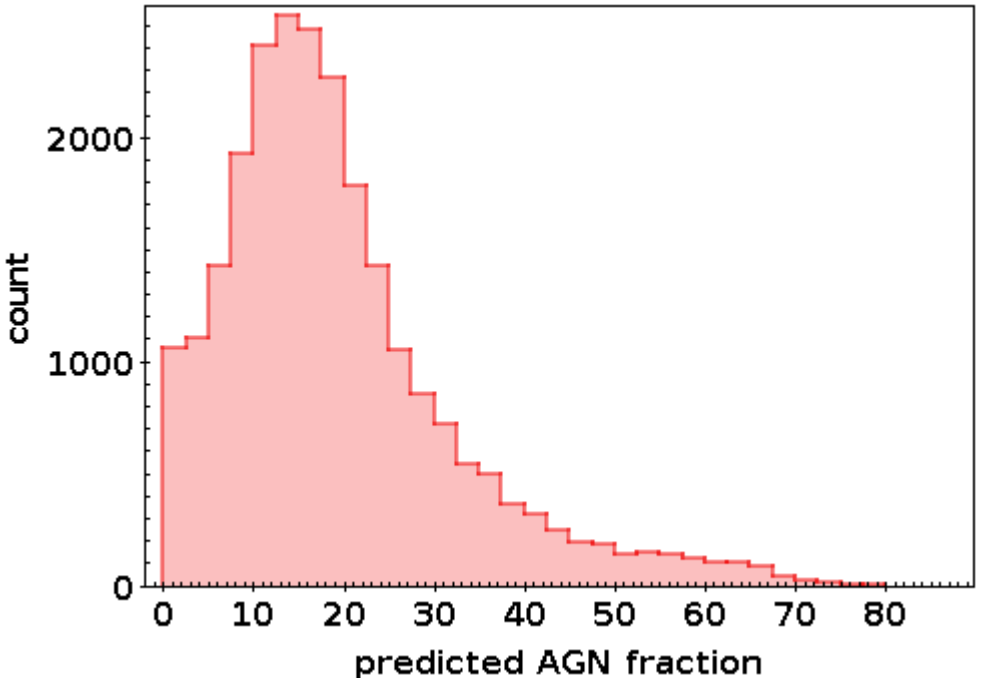


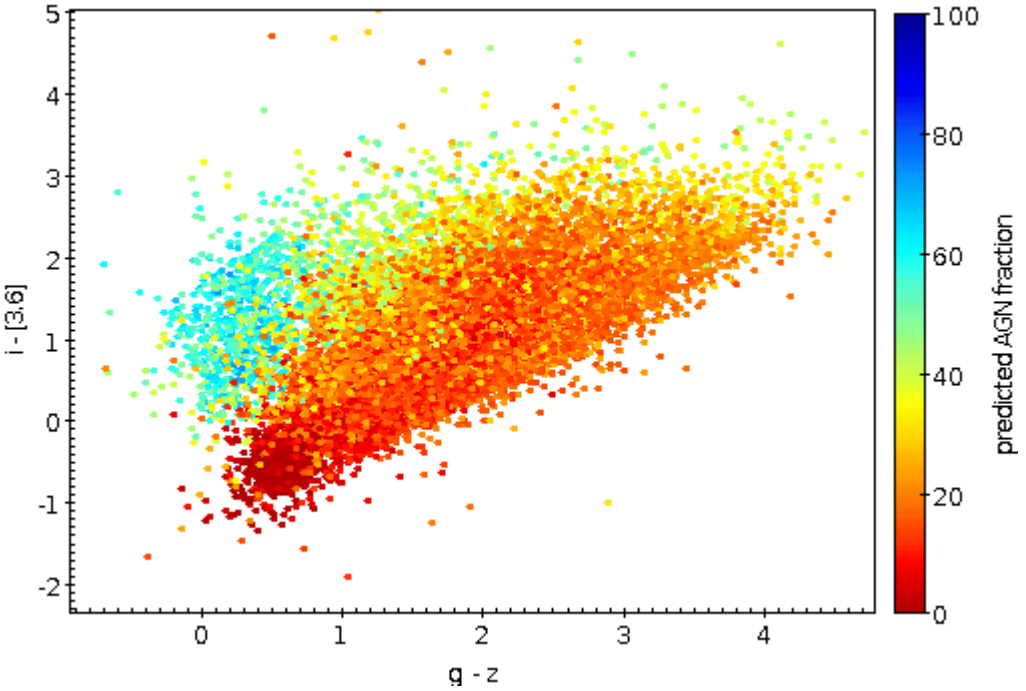


3. **find best fit template for training set**
4. **train random forest to identify**
   AGN fraction in SED
5. **50%train – 50%test**

Global results

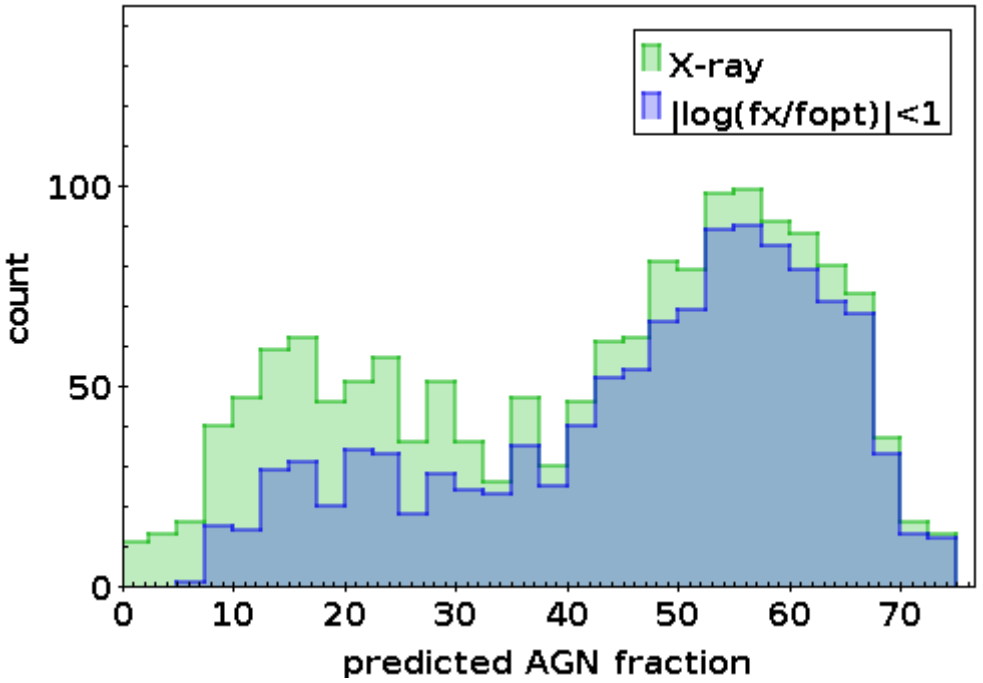
24 000 sources

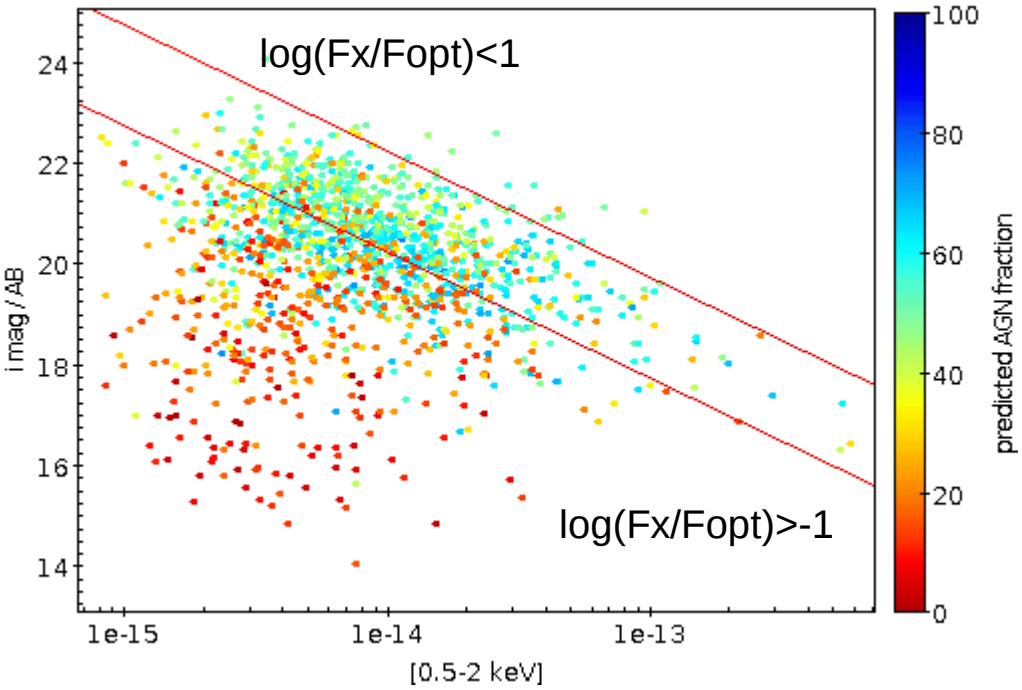




Comparison to X-ray selection

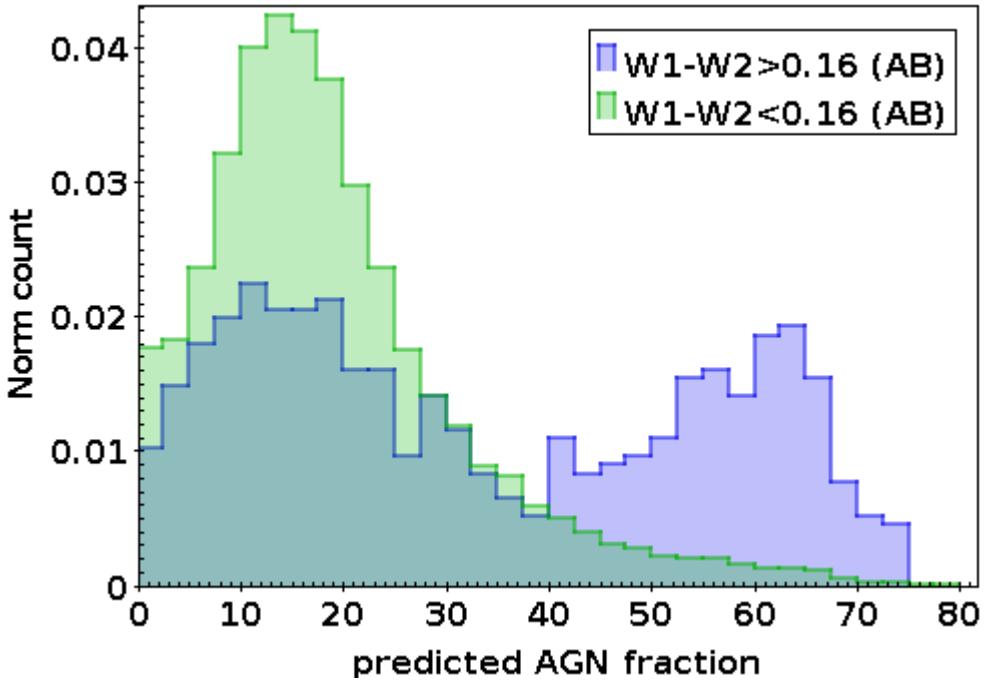
1552 F[0.5-2 keV]>0
676 w. AGN fraction >=50%

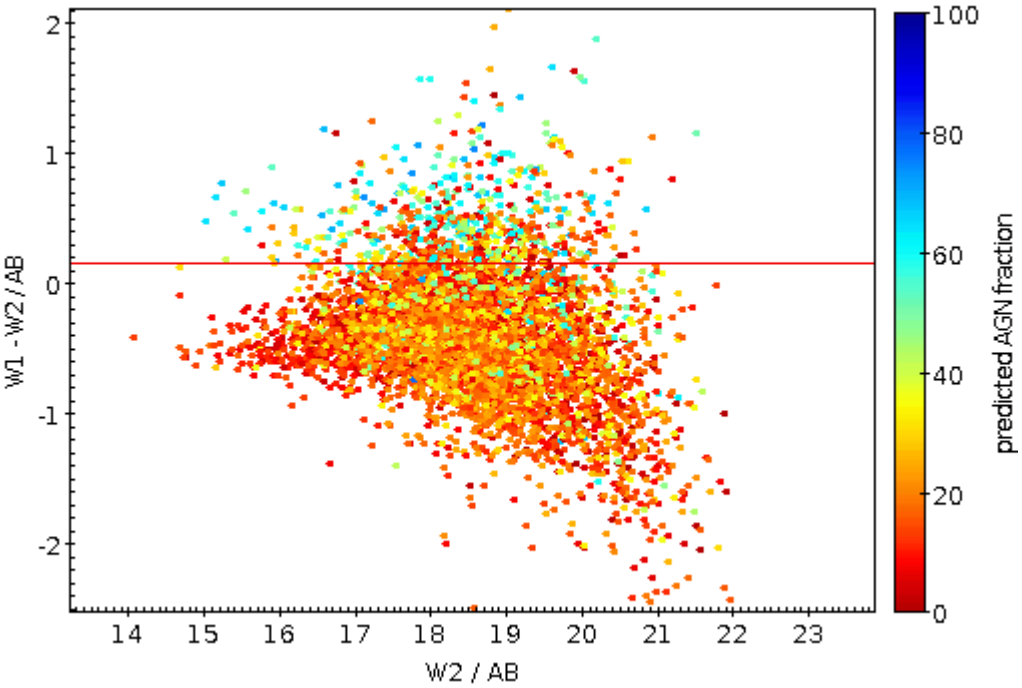




Comparison to WISE selection

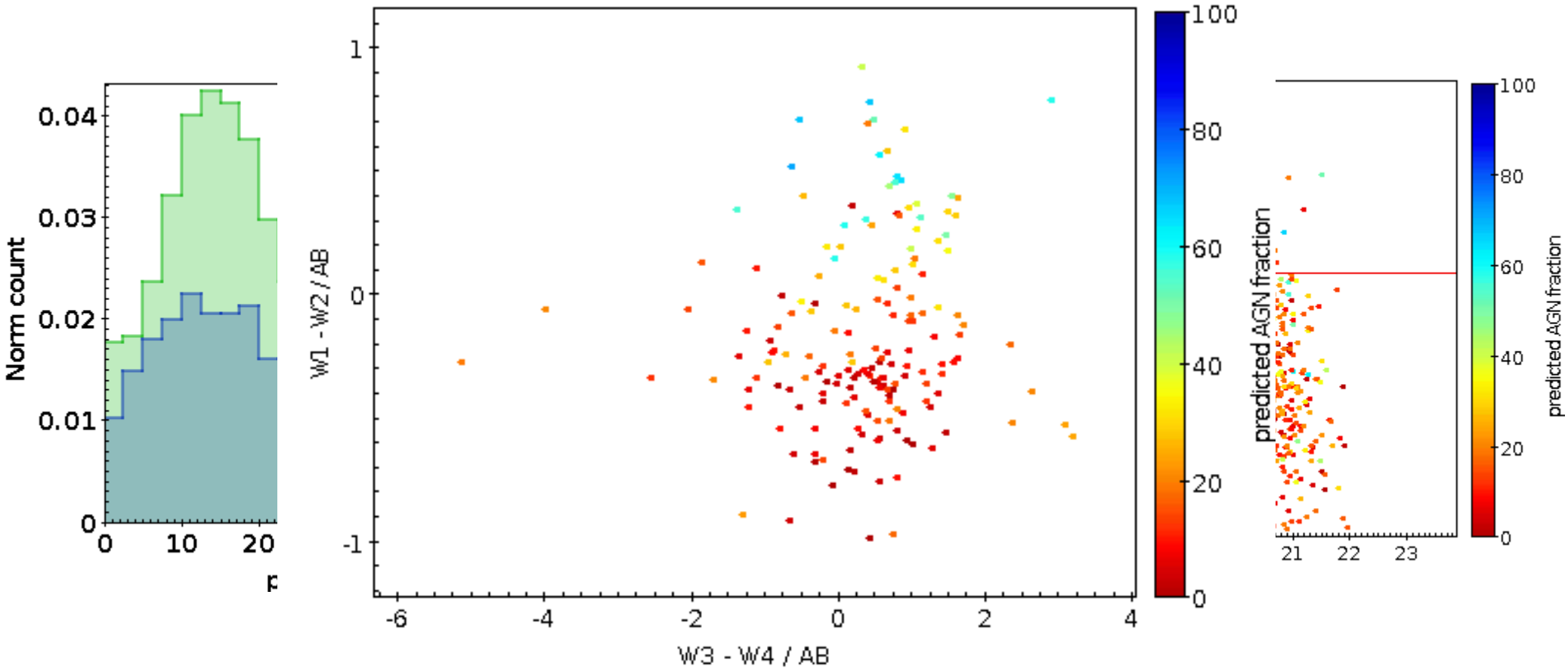
674 W1-W2>0.16
198 w. AGN fraction >=50%





Comparison to WISE selection

674 [0.5-2 keV]>0
198 w. AGN fraction >=50%



Summary

- **AGN SEDs contain a lot of information – differ from galaxies**
- **Machine learning and data mining can provide valuable samples based on feature similarities.**
- **Careful what you wish for: a classification is as good as the training sample: representative sample**