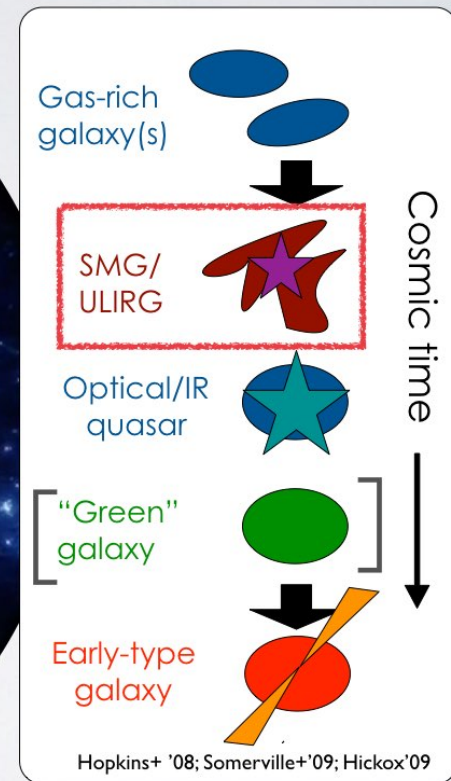
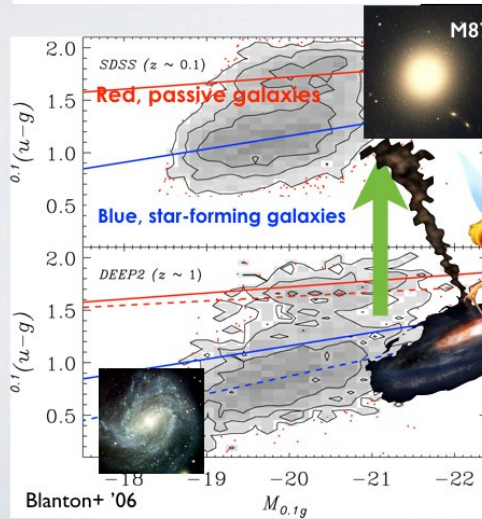
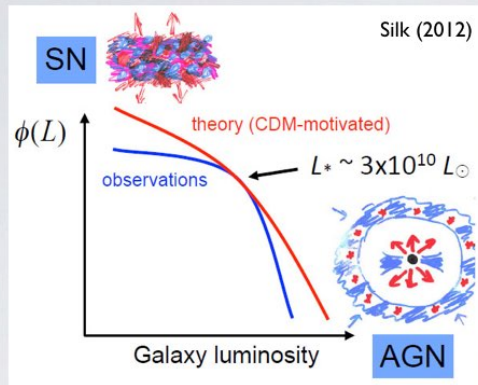


IDENTIFYING THE INTERESTING NEEDLES IN ASTRONOMICAL HAYSTACKS

Andy D. Goulding — Princeton University
ML@Ringberg Dec 2019

goulding@astro.princeton.edu

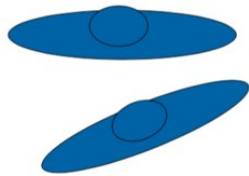
FORMATION OF STRUCTURE & SUPPRESSION OF STAR-FORMATION



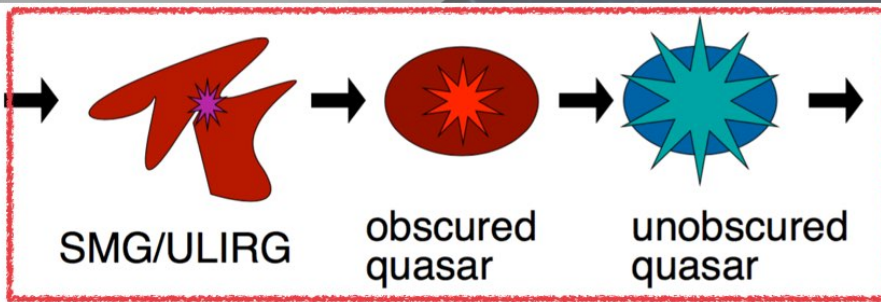
FORMATION OF STRUCTURE & SUPPRESSION OF STAR-FORMATION

SN

Silk (2012)



gas-rich galaxy(s)

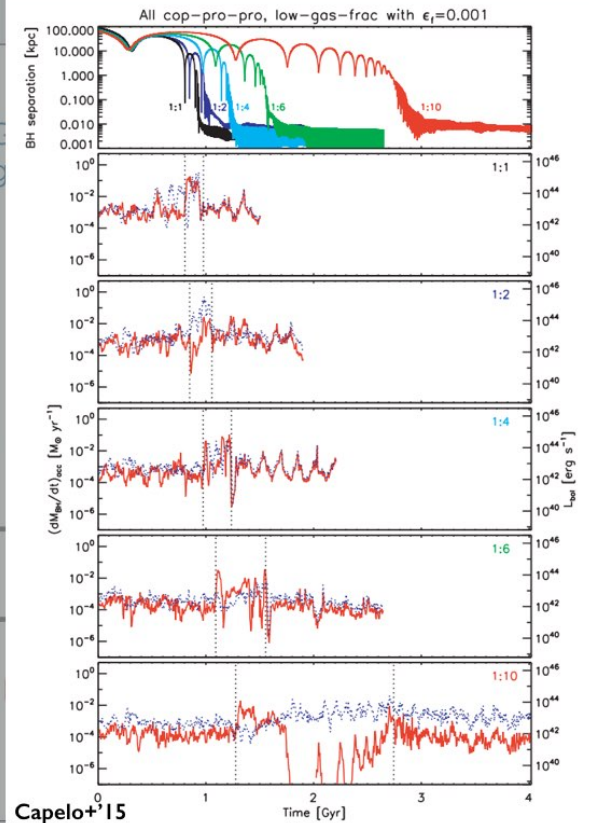
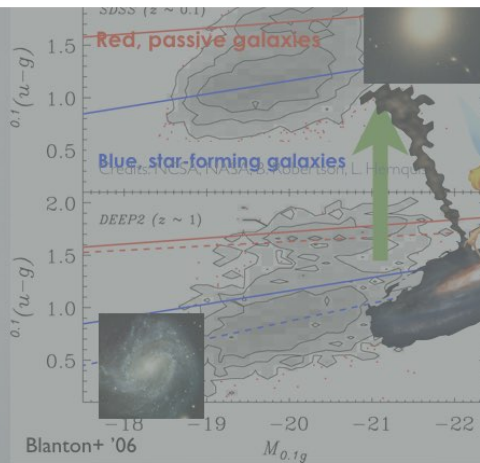


SMG/ULIRG

obscured quasar

unobscured quasar

early-type galaxy



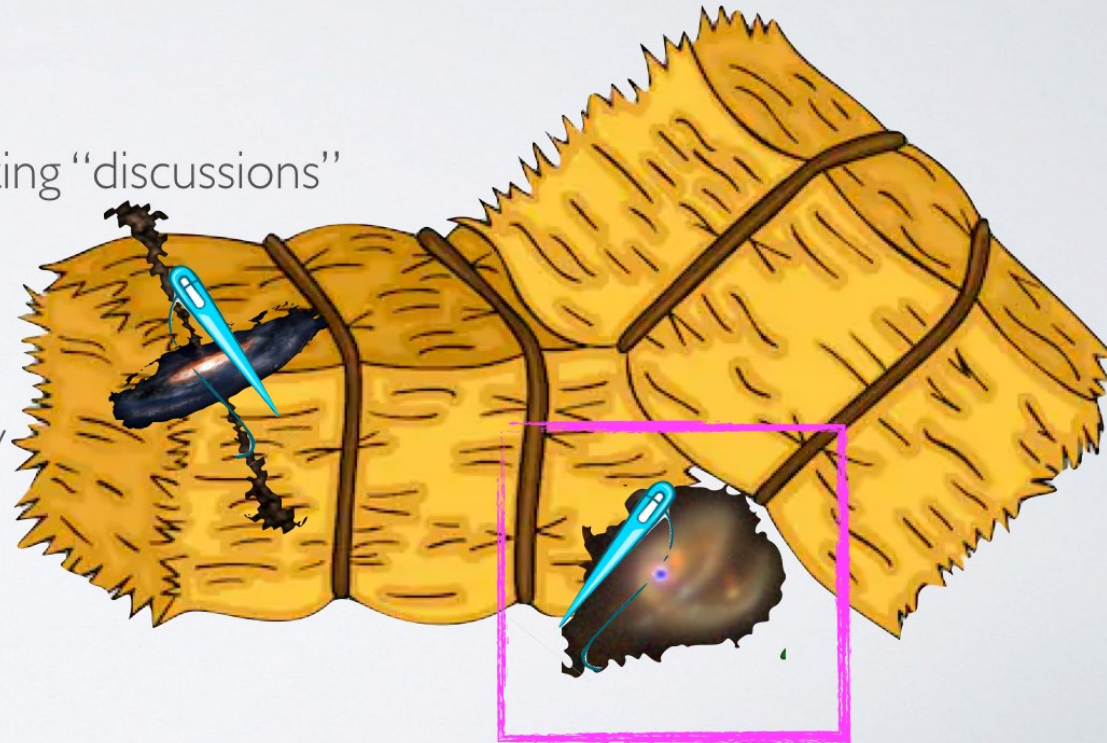
IDENTIFYING THE NEEDLES

Propaganda::

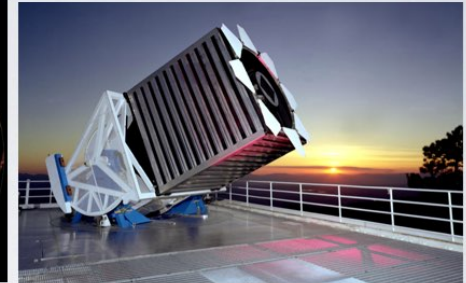
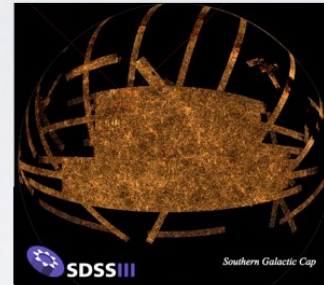
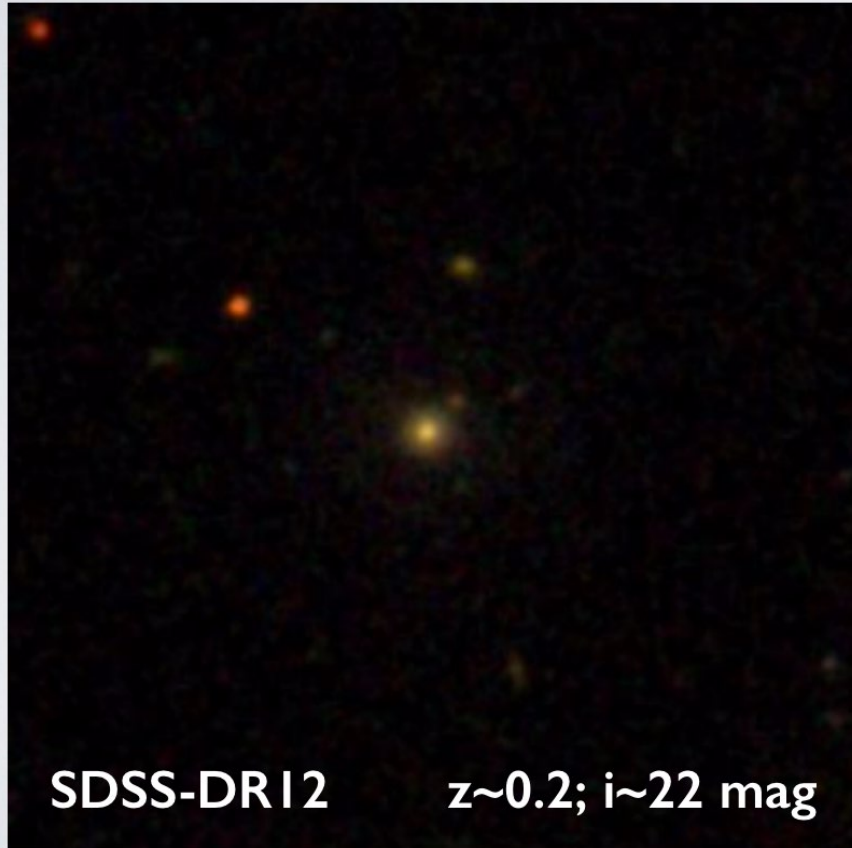
- AGN are interesting!
- Galaxy mergers are interesting!
- AGN in mergers lead to very interesting “discussions”

Problem::

- AGN are rare & difficult to identify
- Mergers are rare & difficult to identify



GALAXY MORPHOLOGIES TO $z \sim 1$ IN WIDE-FIELD OPTICAL SURVEYS



ENGLISH

GALAXY ZOO.org

Hi starstryder | Home | The Science | How to Take Part | Galaxy Analysis | Forum | Press | Blog | FAQ | Links | Contact Us | Logout

- Galaxy Tutorial
- Galaxy Analysis
- Galaxy Zoo - Thank You
- Show My Galaxies

Galaxy Analysis

Welcome to Galaxy Zoo's view of the Universe. If you're here you should already have seen the [Tutorial](#), but feel free to go and remind yourself. There's no need to agonise for too long over any one image, just make your best guess in each case.

Galaxy Ref: 587729387677679742

Choose the Galaxy Profile by clicking the buttons below

CLOCK PRINT COPY ONLY

SPIRAL GALAXY

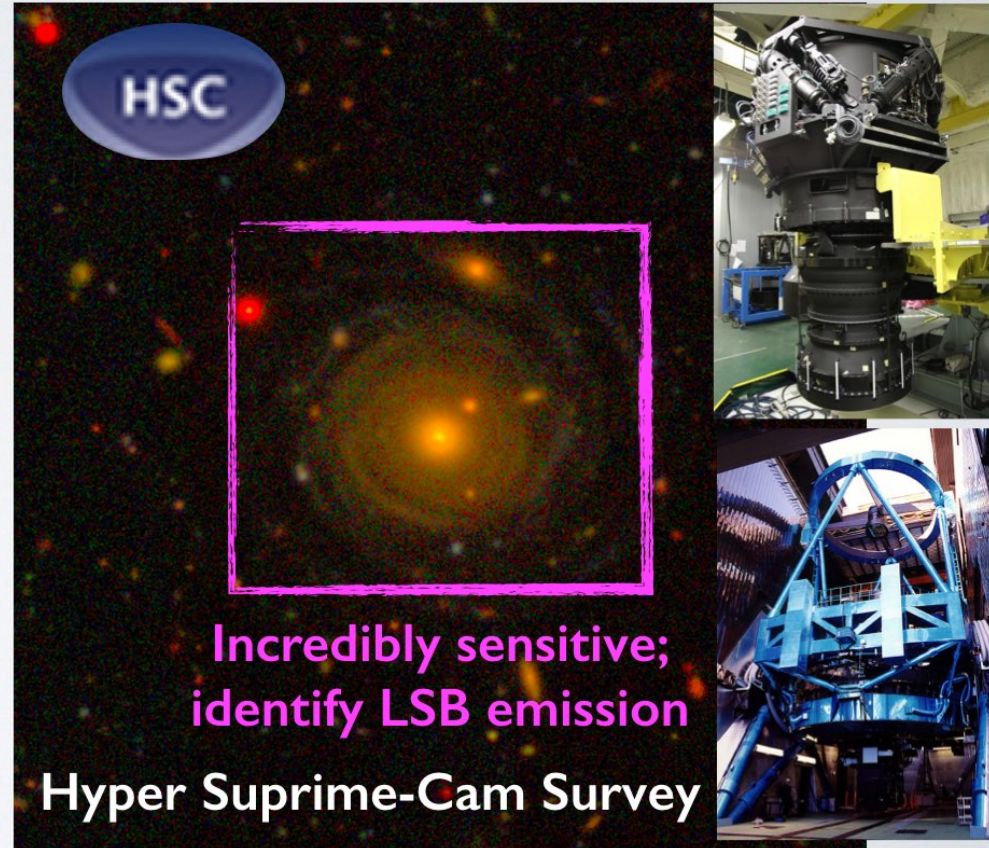
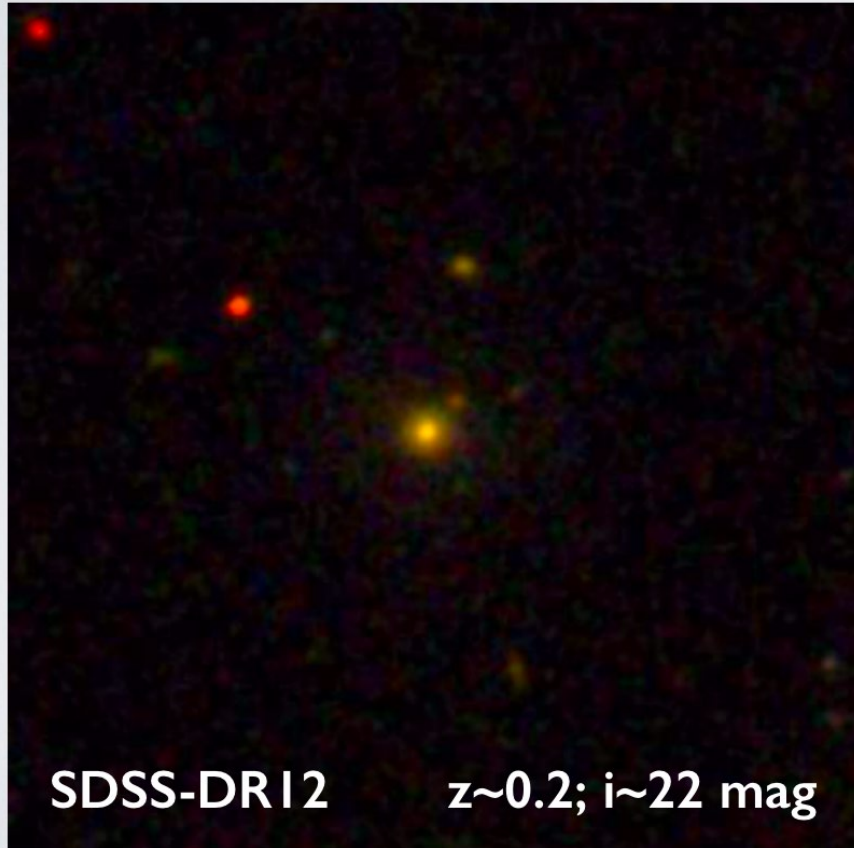
ELLIPTICAL GALAXY

STAR FORMING GALAXY

HERSHEL

Show Grid Overlay on the next Image

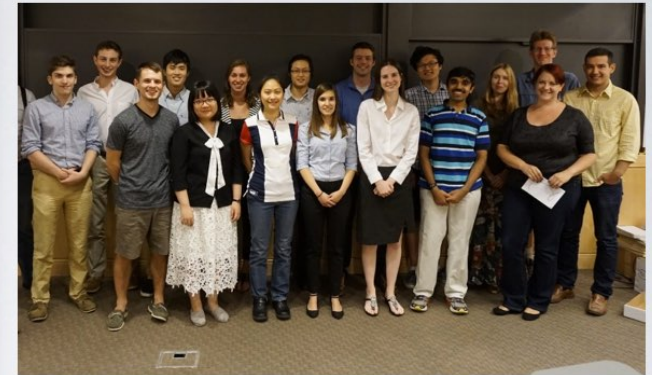
GALAXY MORPHOLOGIES TO $z \sim 1$ IN WIDE-FIELD OPTICAL SURVEYS



HSC

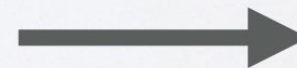


VISUAL CLASSIFICATIONS — THE GOLD STANDARD



~6000 visually classified systems

~400 major mergers



Too few to train a classifier

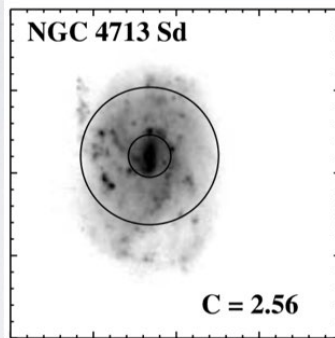


...and if I already had 6000 major mergers, I wouldn't need a classifier!

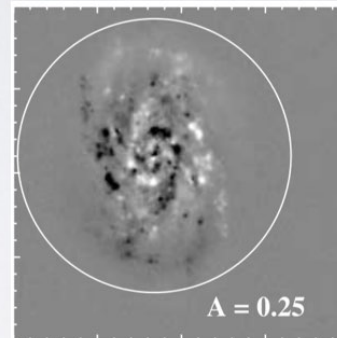
AUTOMATED INTERACTION-STAGE CLASSIFICATIONS



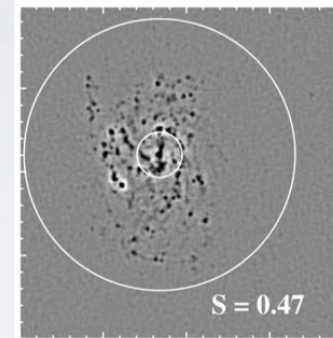
Concentration



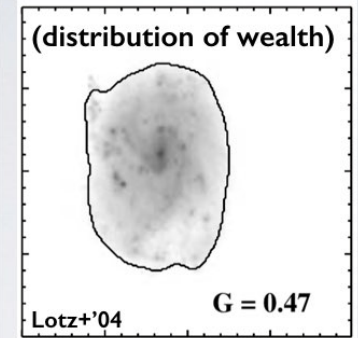
Asymmetry



Smoothness



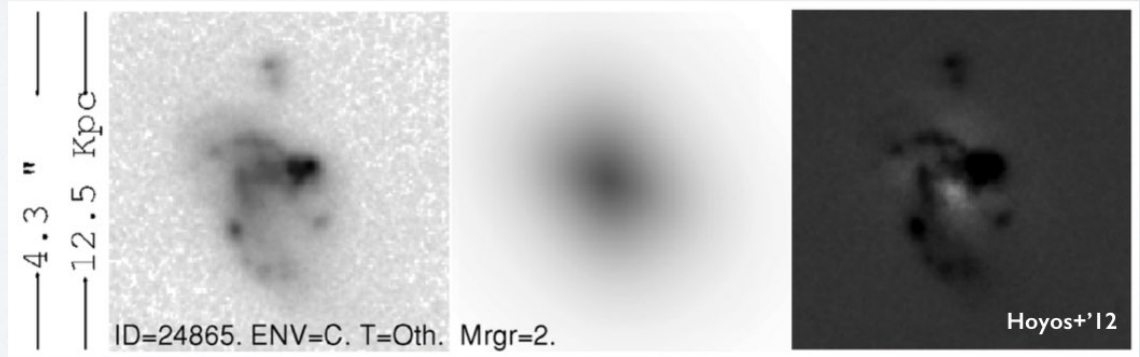
Gini



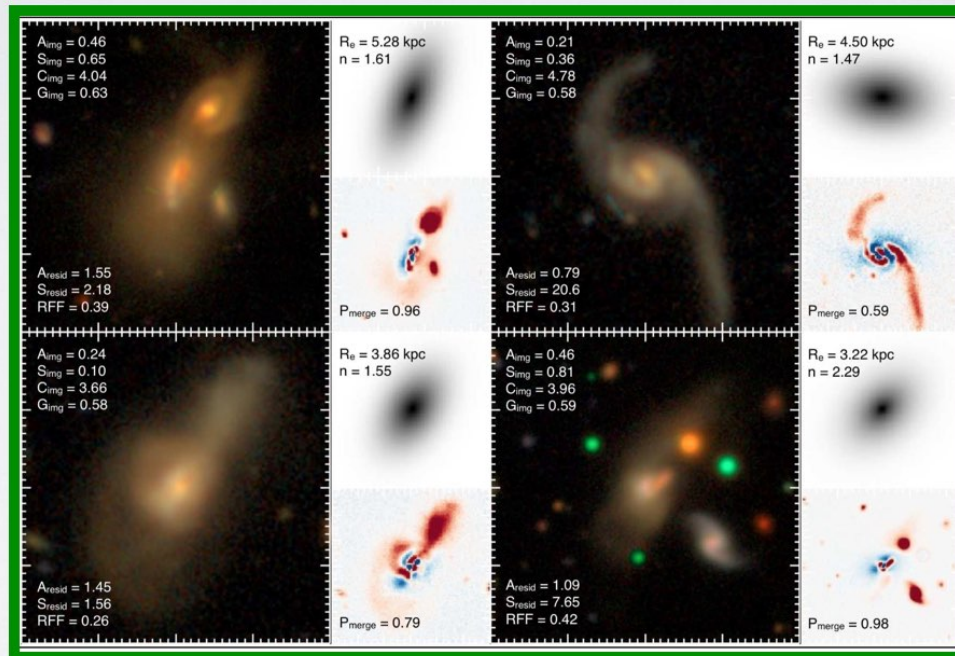
see Conselice '03; Lavery+'04; Lotz+'04; Cassata+'05; Conselice+'08; Jogee+'08,'09; Lotz+'08; Wen+'09; Hoyos+'12



**Residual
Flux
Fluctuations**



USING SUMMARY MEASURES



Smoothness
Gini
Asymmetry
Concentration
Residual
Flux
Fluctuations
Sersic Index
Effective Radius

Use of CNN is beginning to be explored (Martin+19; Weigel+19; Nevin+19)

MORE NEEDLES! — REBALANCING THE TRAINING



$$N(\text{merger, train}) \ll N(\text{isolated, train})$$

$$X \pm \sigma_X$$

→ Upsampling the small number of mergers

$$N(\text{merger, train}) \sim N(\text{isolated, train})$$

Smoothness

Gini

Asymmetry

Concentration

Residual

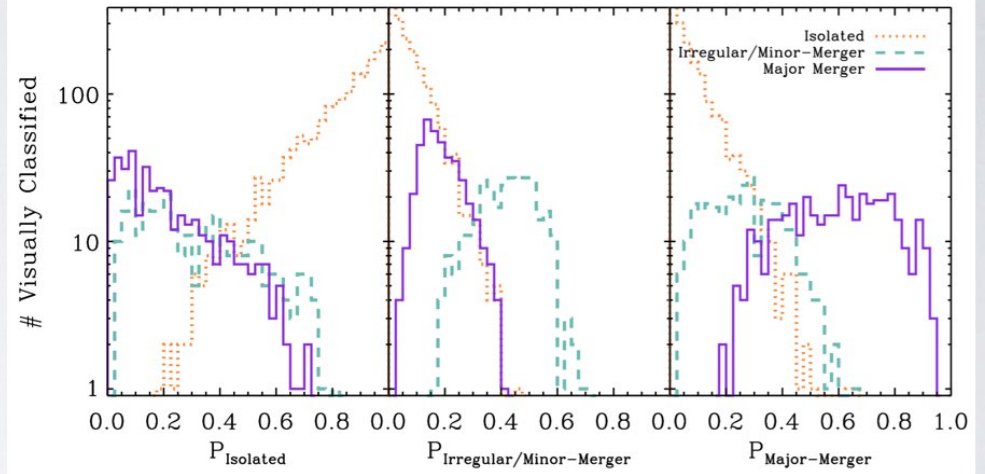
Flux

Fluctuations

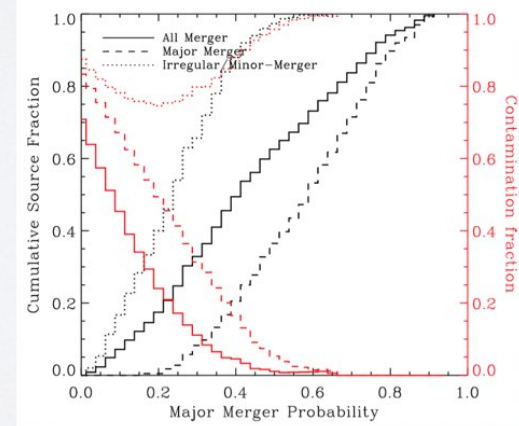
Sersic Index

Effective Radius

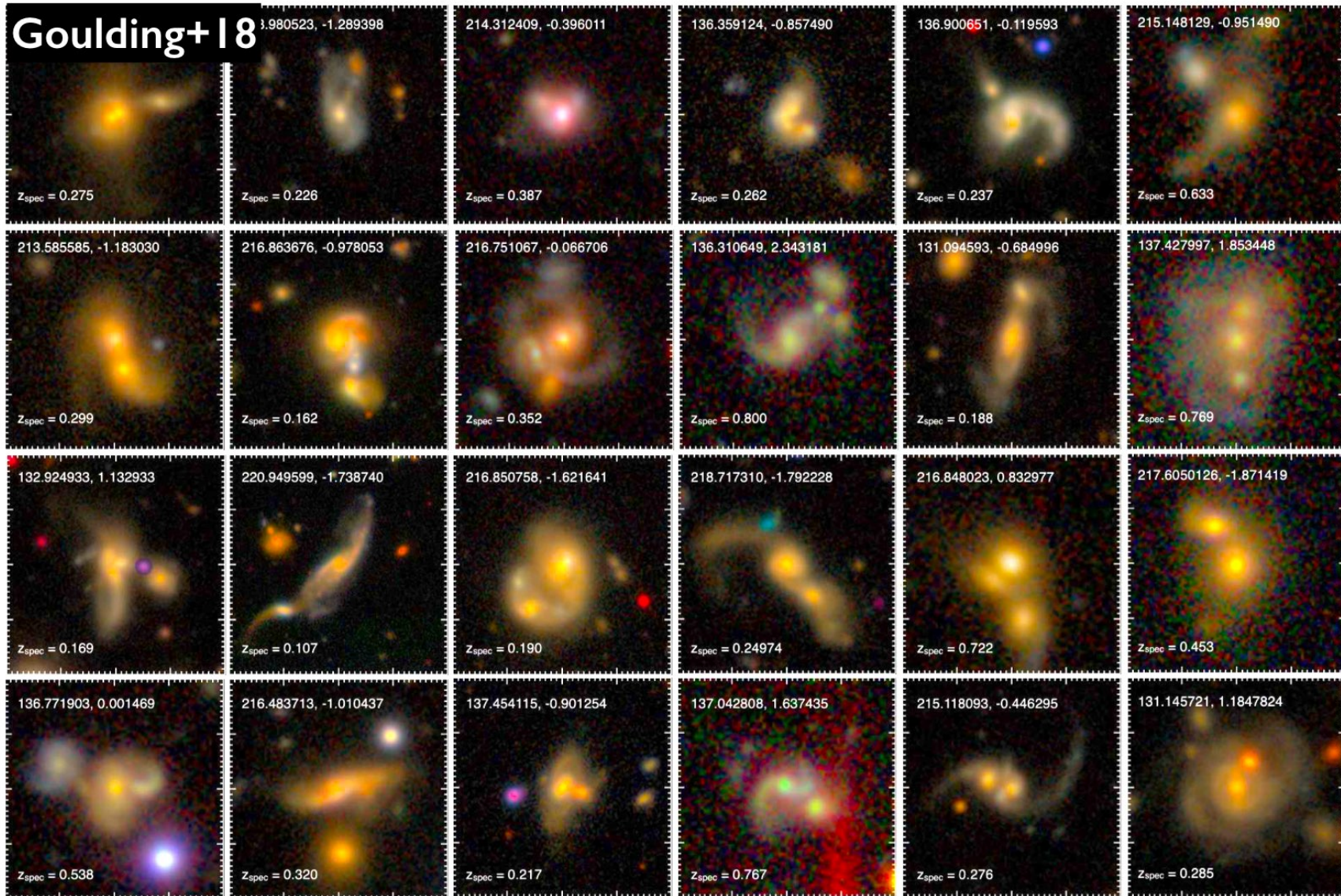
A (RANDOM DRAW) RANDOM FOREST CLASSIFIER



Interesting point: boosting failed miserably — not capturing the measurement uncertainties??



Goulding+18



IDENTIFYING THE NEEDLES

Propaganda::

- AGN are cool!
- Galaxy mergers are cool!
- AGN in mergers are very cool!

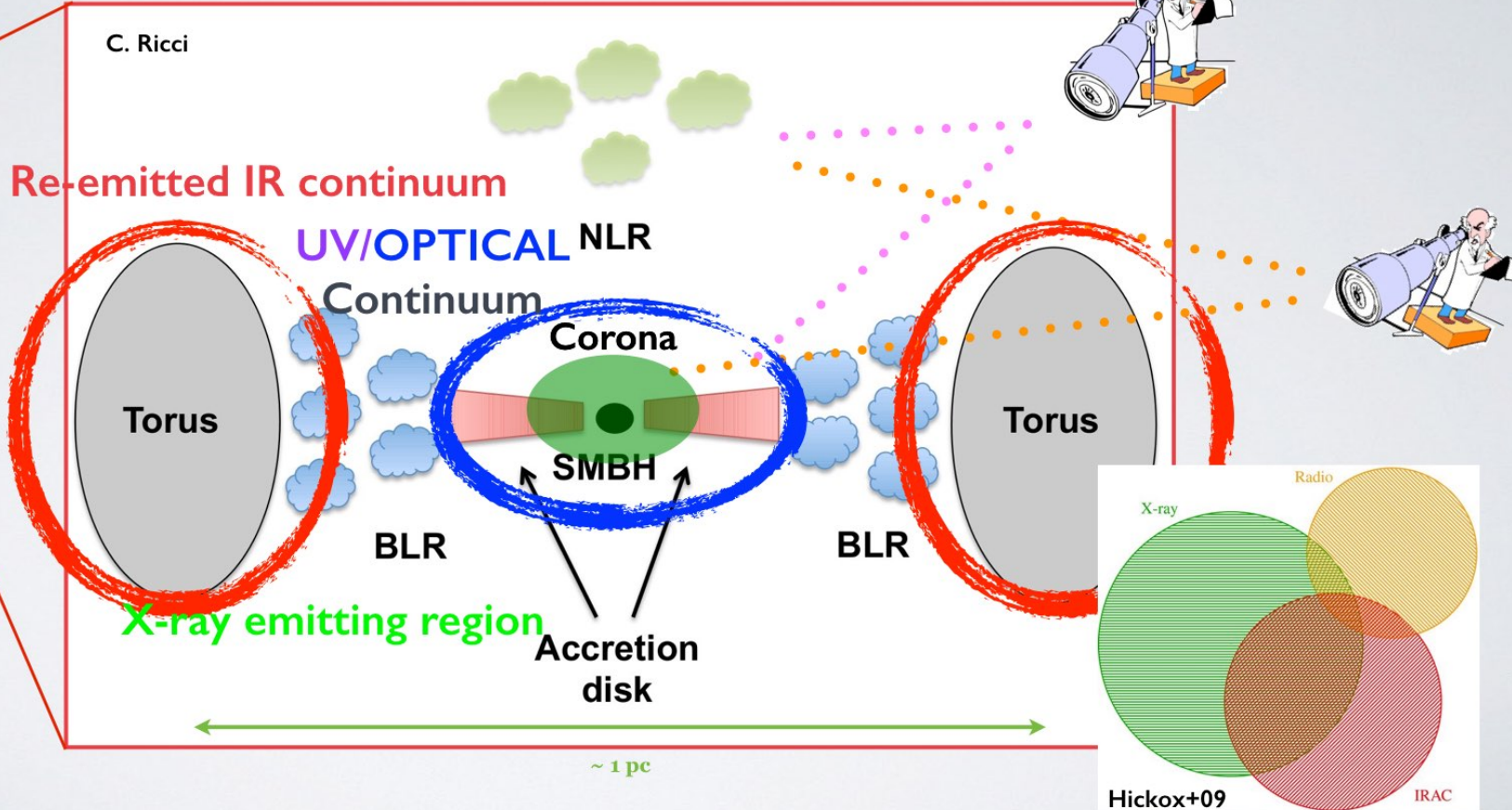
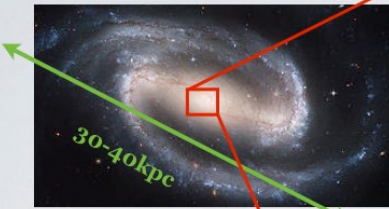
Problem::

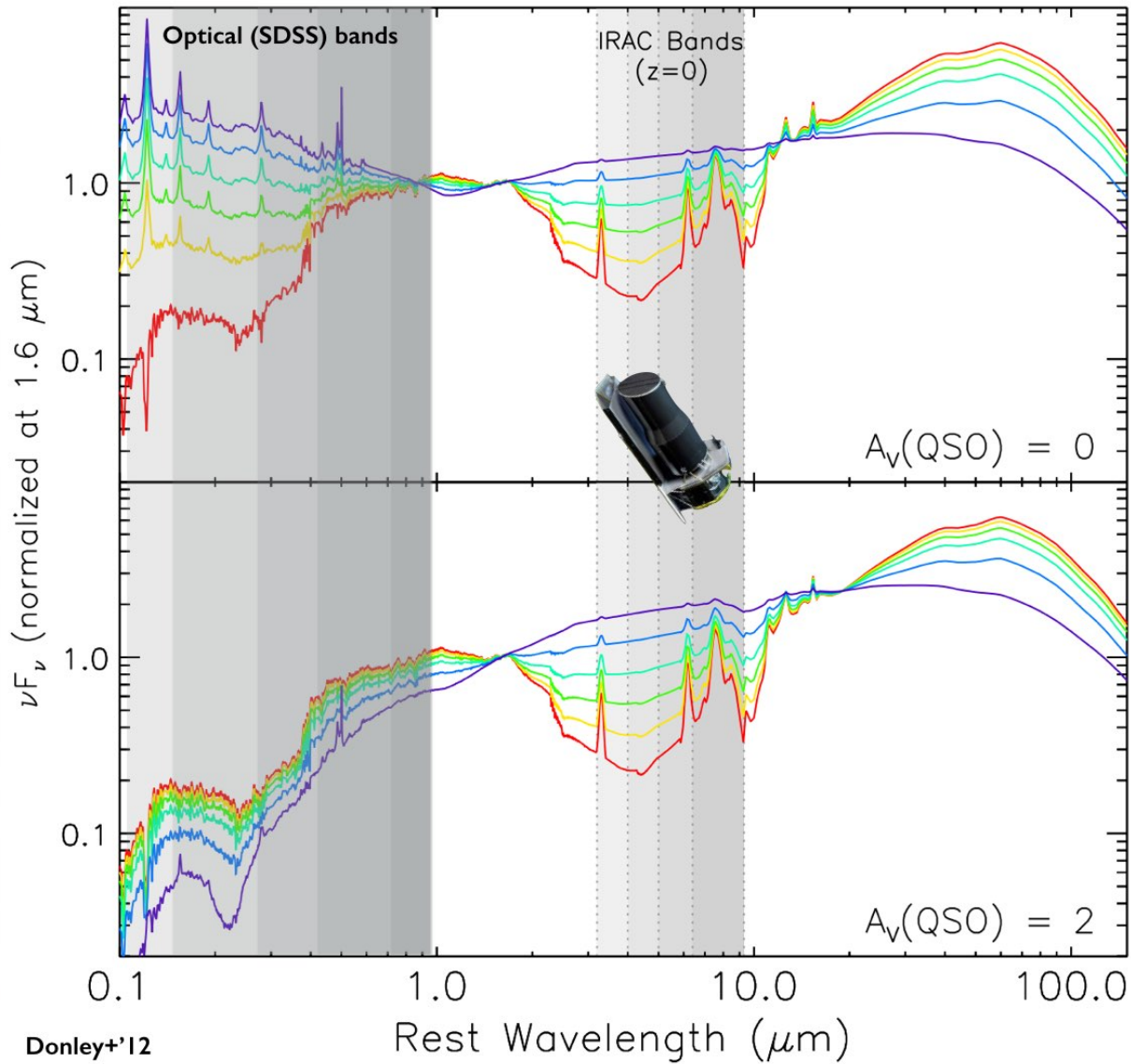
- AGN are rare & difficult to identify
- Mergers are rare & difficult to identify



NAIVE AGN SELECTION

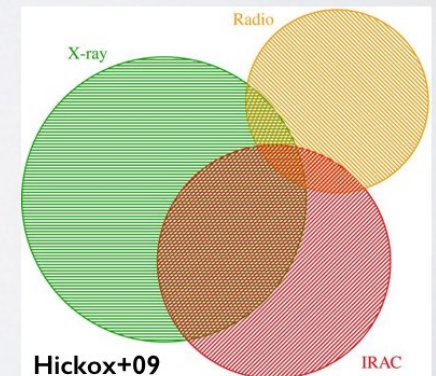
C. Ricci



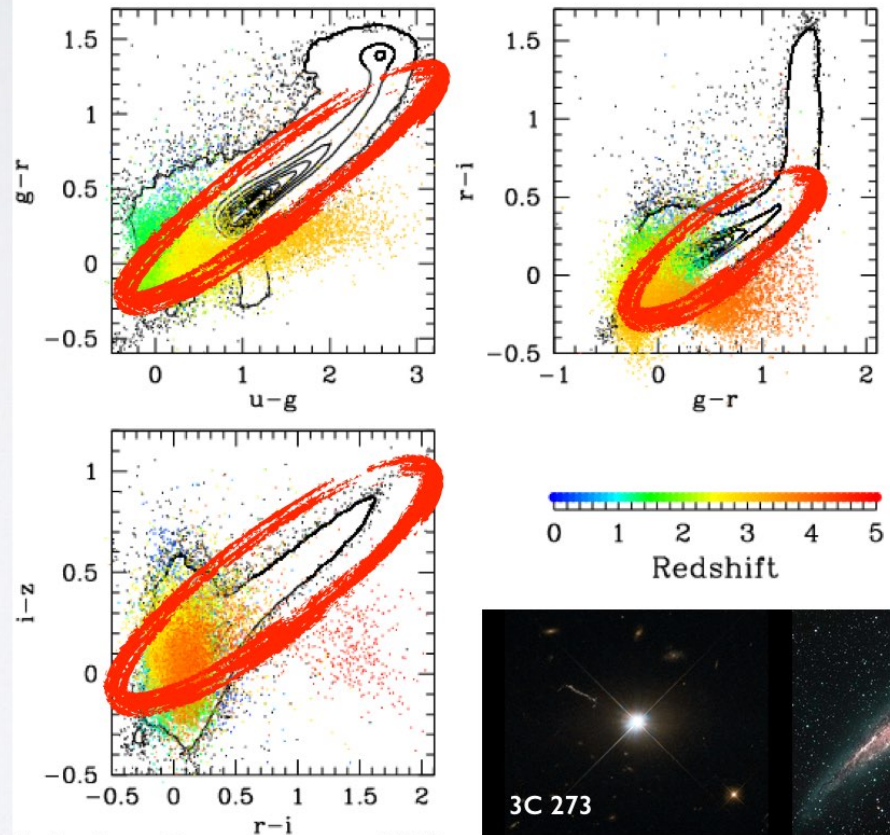
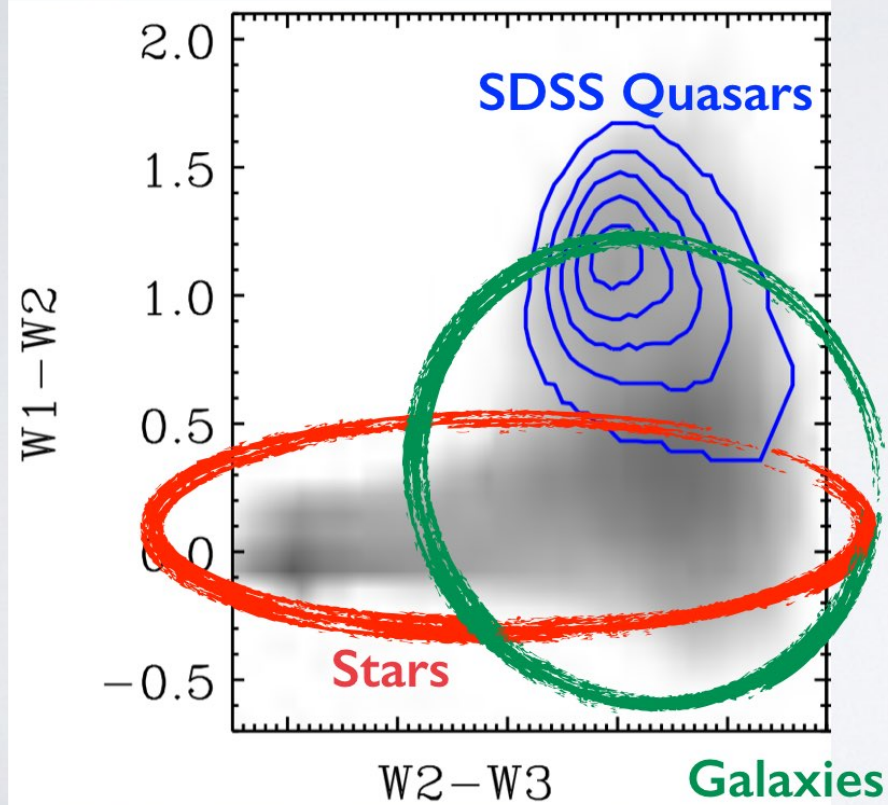


A tough problem to solve with spectra

...and don't even mention the word **'variability'** !

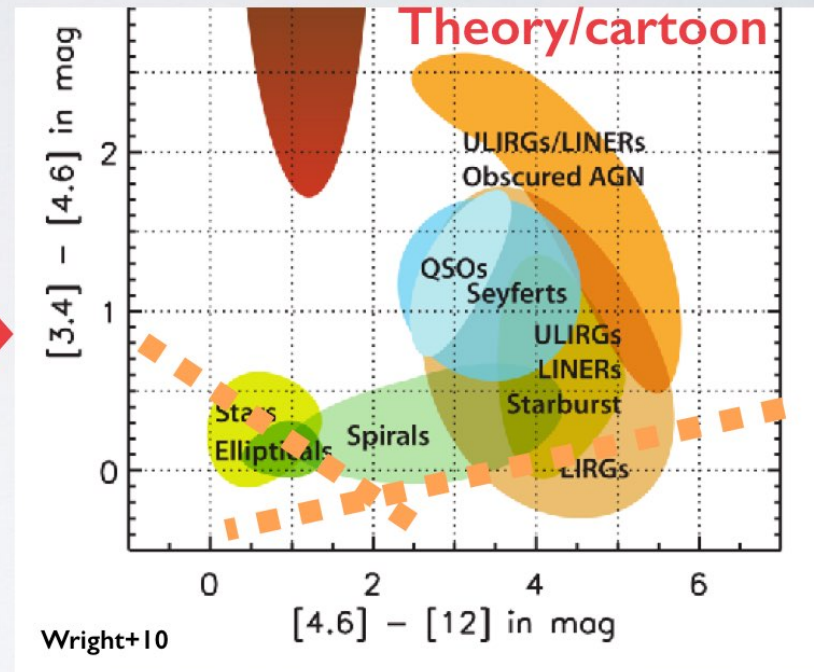
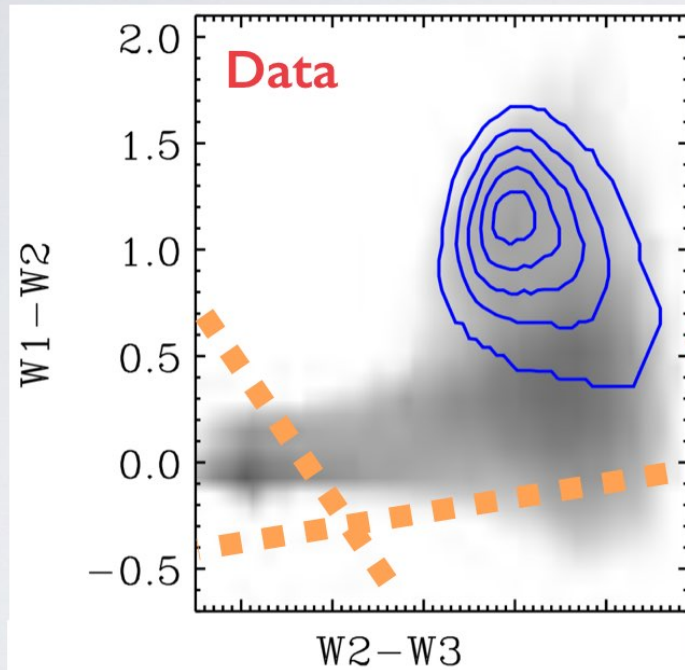


NAIVE AGN SELECTION



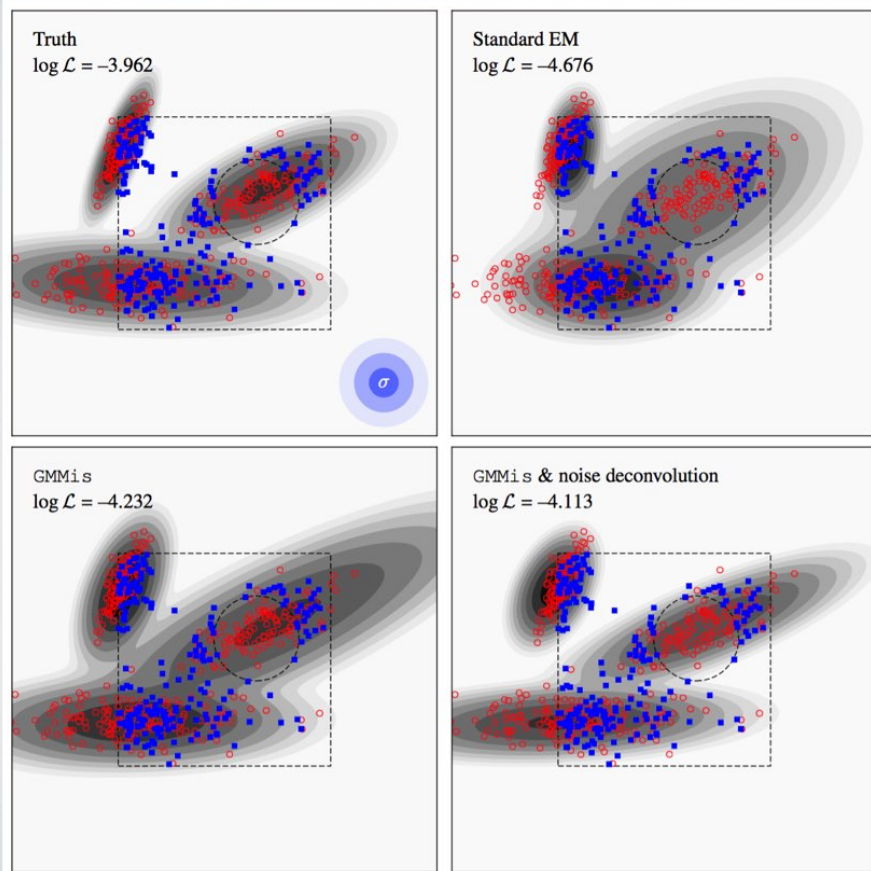
Can we identify the rest of the AGN & Quasars???

LOOKS SUSPICIOUSLY FAMILIAR...GMM?

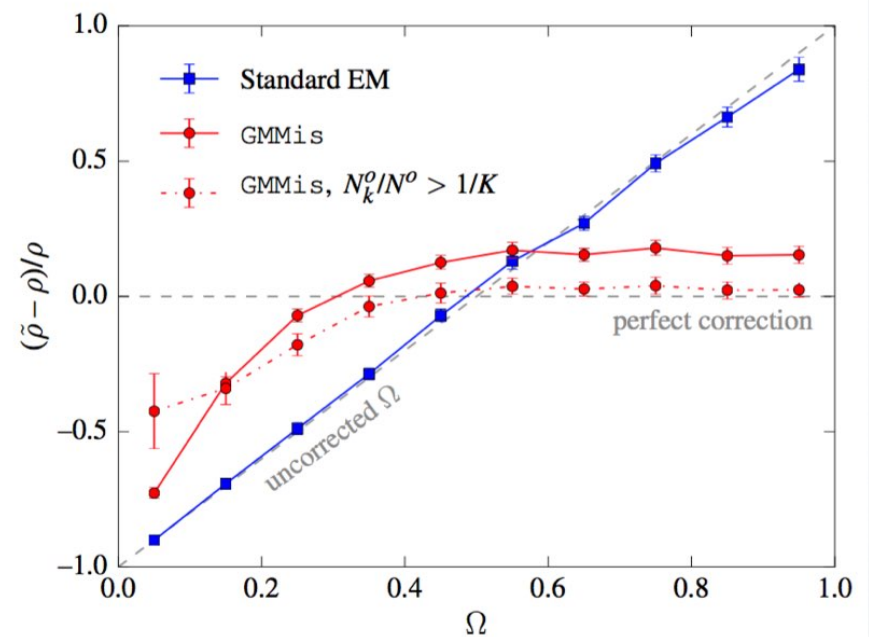


Proposal: BE SELFISH! Cut everything that's not your science ...but then can you still describe this with a GMM?

PYGMMIS (MELCHIOR & GOULDING 2018)



Can explicitly include noise & incompleteness into a GMM



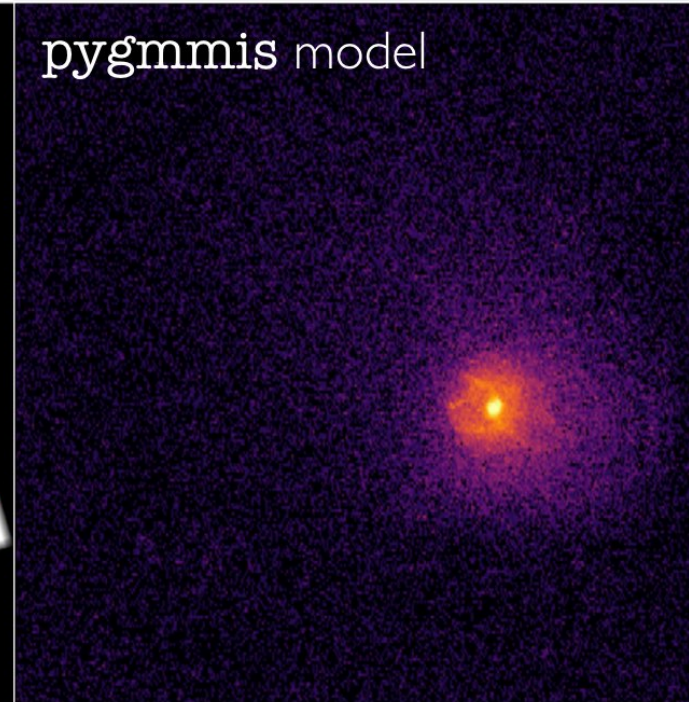
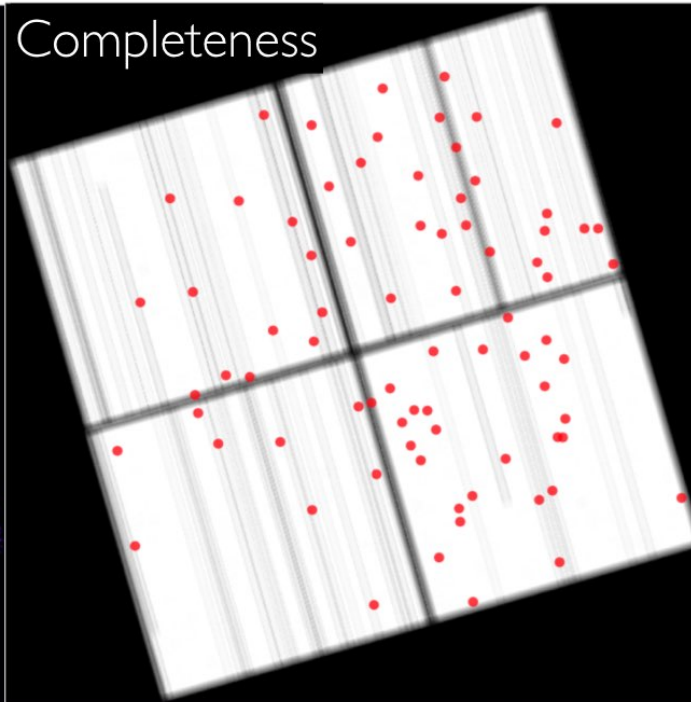
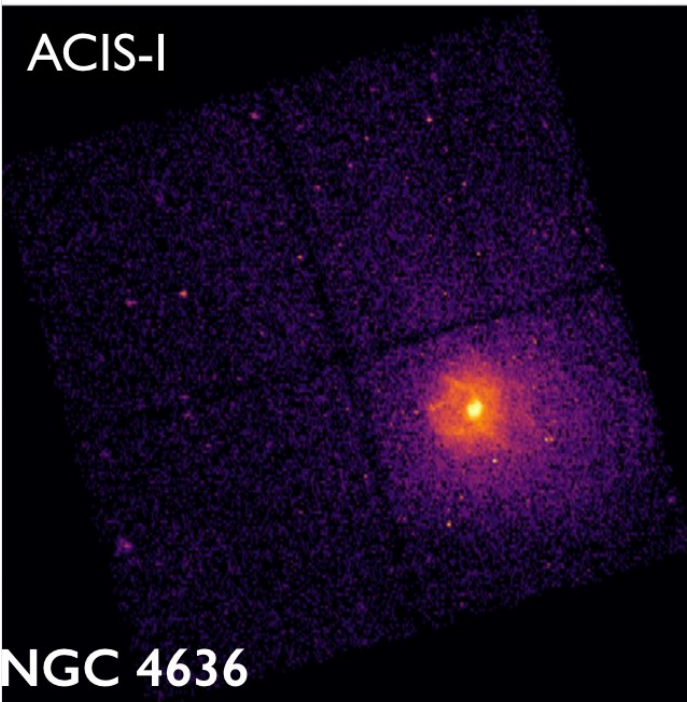
PYGMMIS (MELCHIOR & GOULDING 2018)

ACIS-I

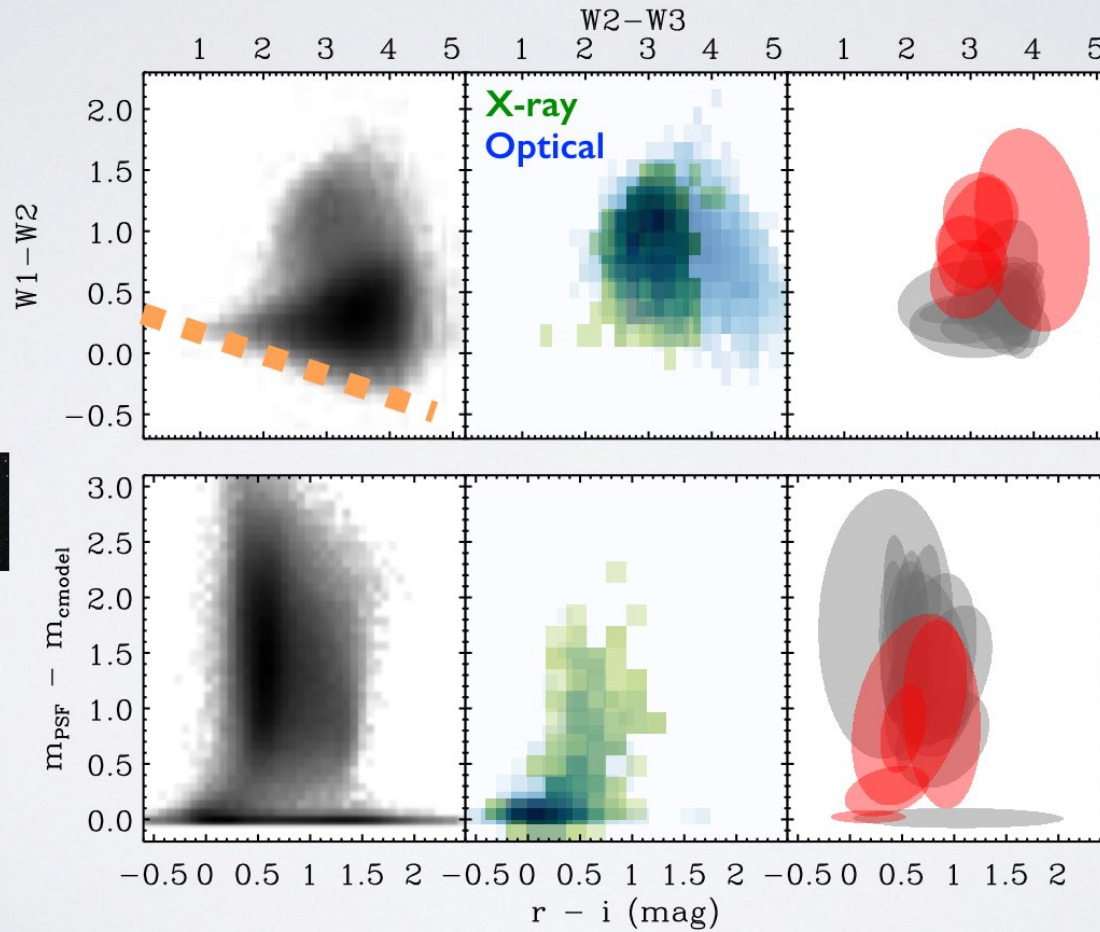
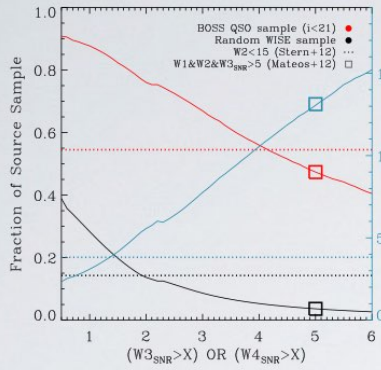
Completeness

pygmmis model

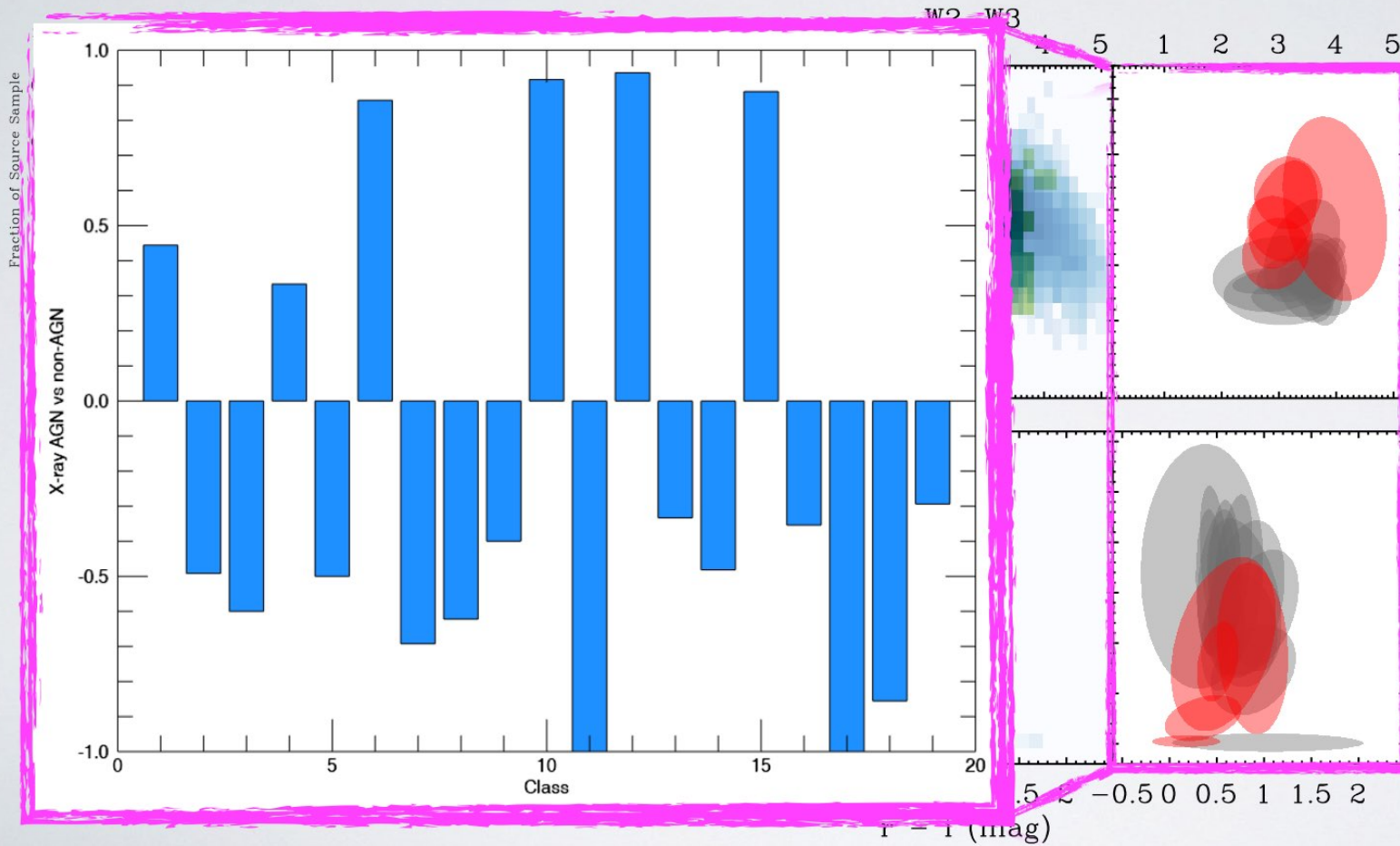
NGC 4636



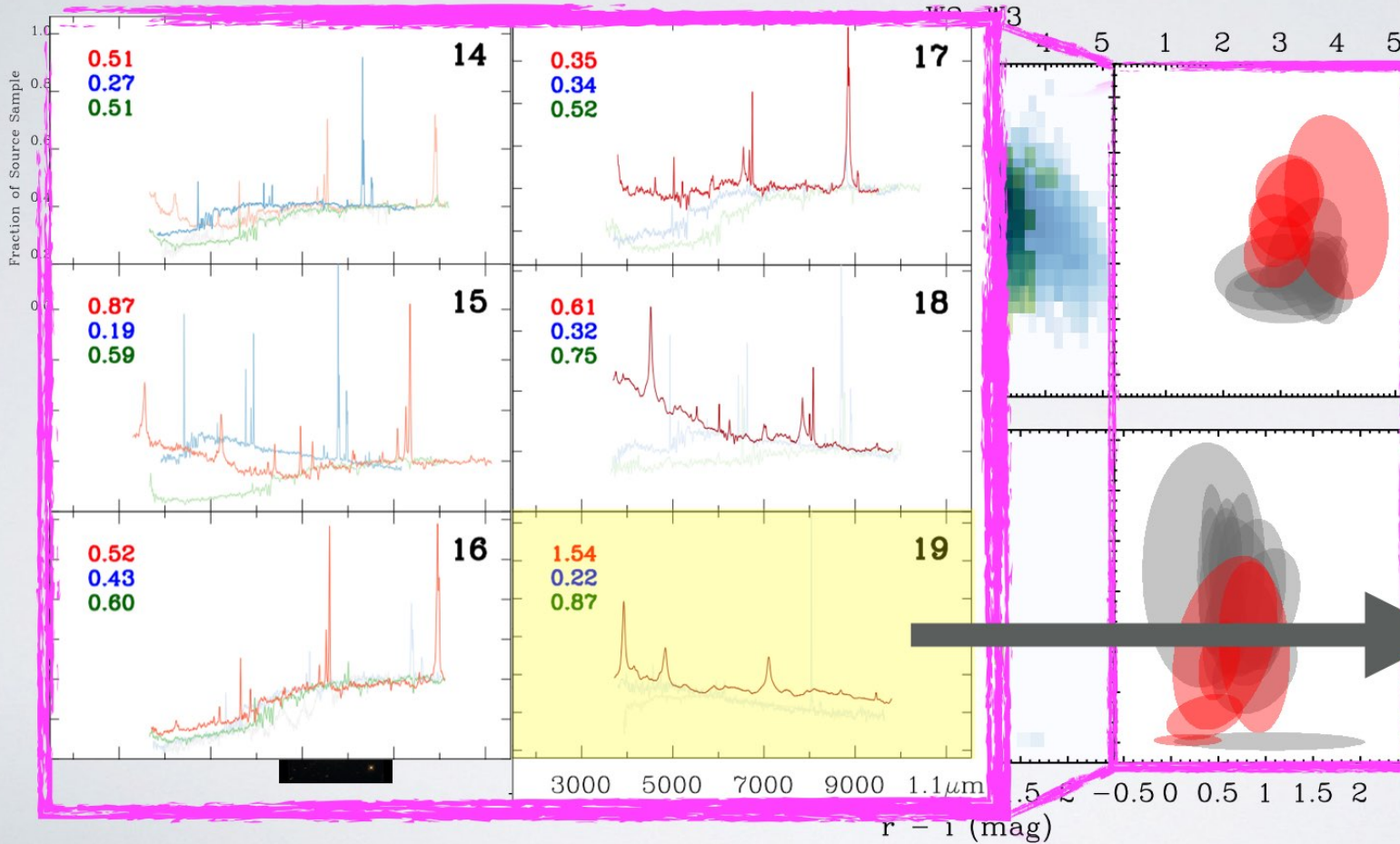
QUASARS IN HSC & WISE — INTERPRETING THE MODEL



QUASARS IN HSC & WISE — INTERPRETING THE MODEL



QUASARS IN HSC & WISE — INTERPRETING THE MODEL

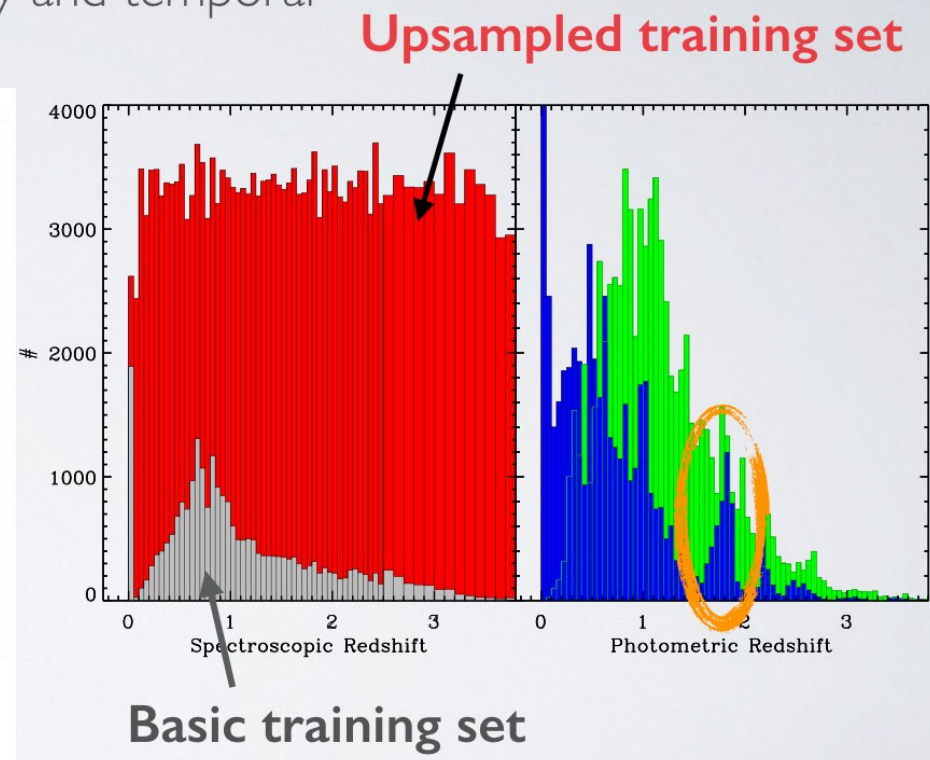
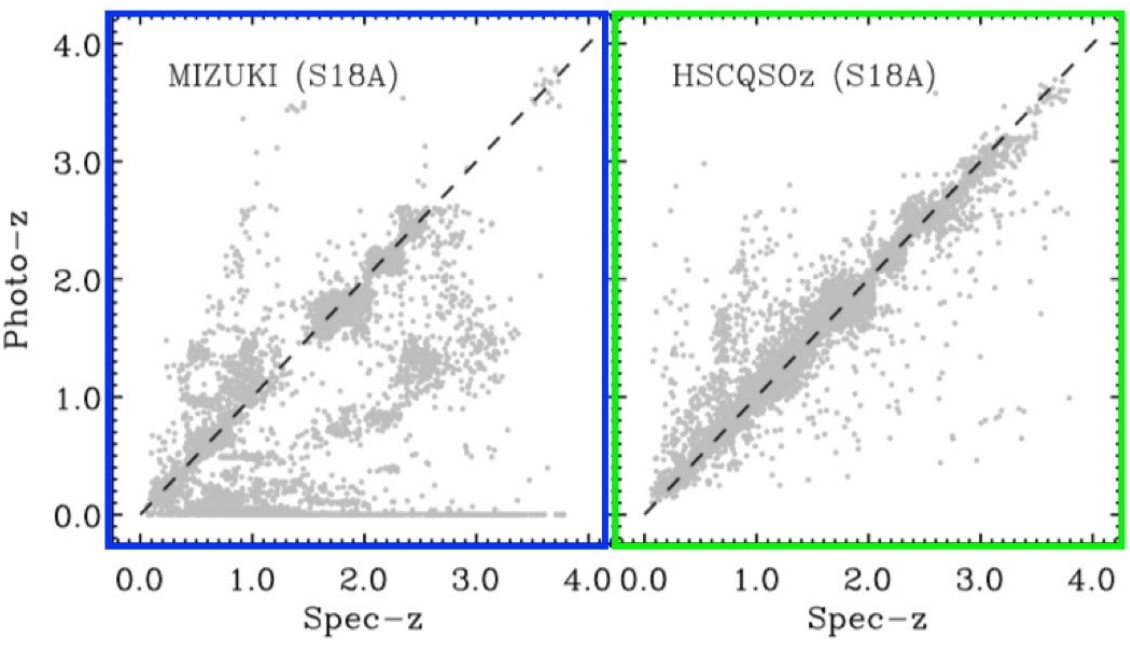


~3x more quasars
to $z \sim 3$ than
previously
identified

~95% of SDSS
spectroscopic
Quasars

BACK TO THE TREES & THE SAME METHODS AS THE MERGERS

Advantage: naturally builds measurement uncertainty and temporal variations into the model



Moral of the story: try to solve YOUR problem not ALL the problems...

LET'S GET CONTROVERSIAL...

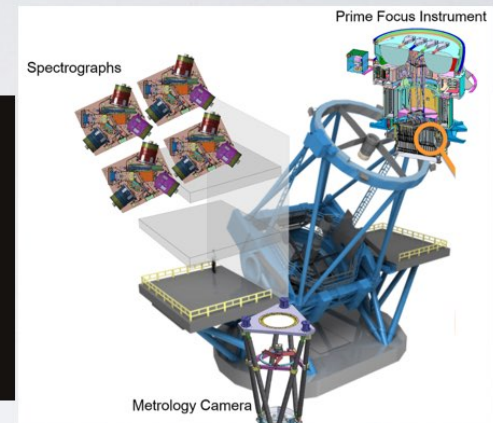
Stage I



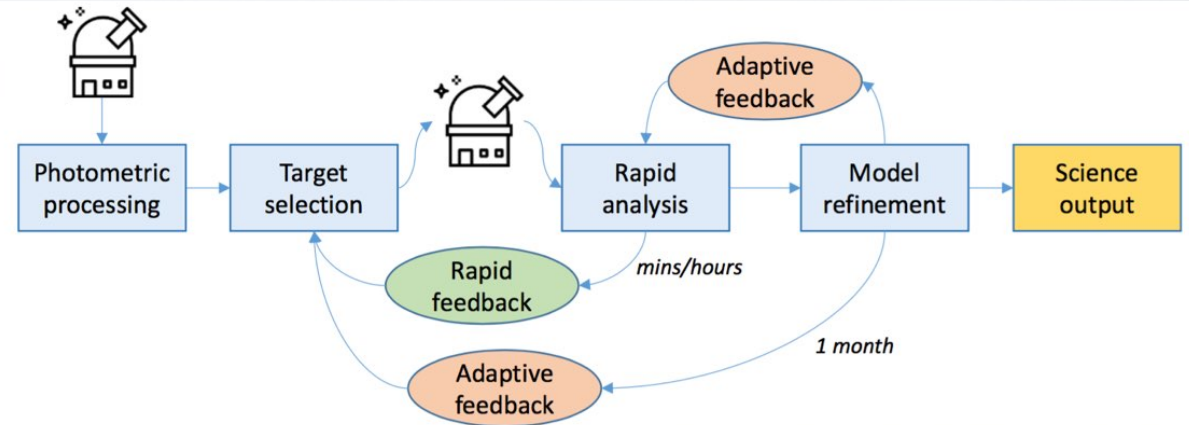
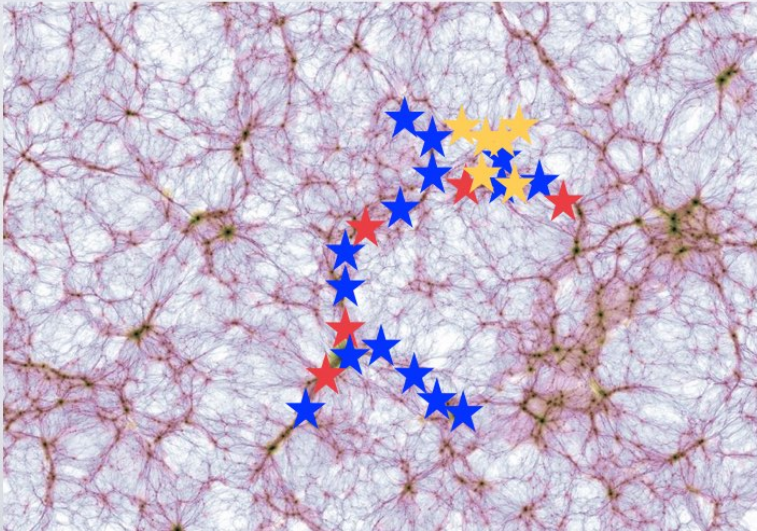
Stage II



Prime Focus Spectrograph



LET'S GET CONTROVERSIAL...



Can we optimize the spectroscopic targeting using a data driven selection function?

What are the primary factors that would drive such a utility function?

Can this still be reproducible for science?

How do we target the full distribution of objects... how do we ensure that we catch all the needles?