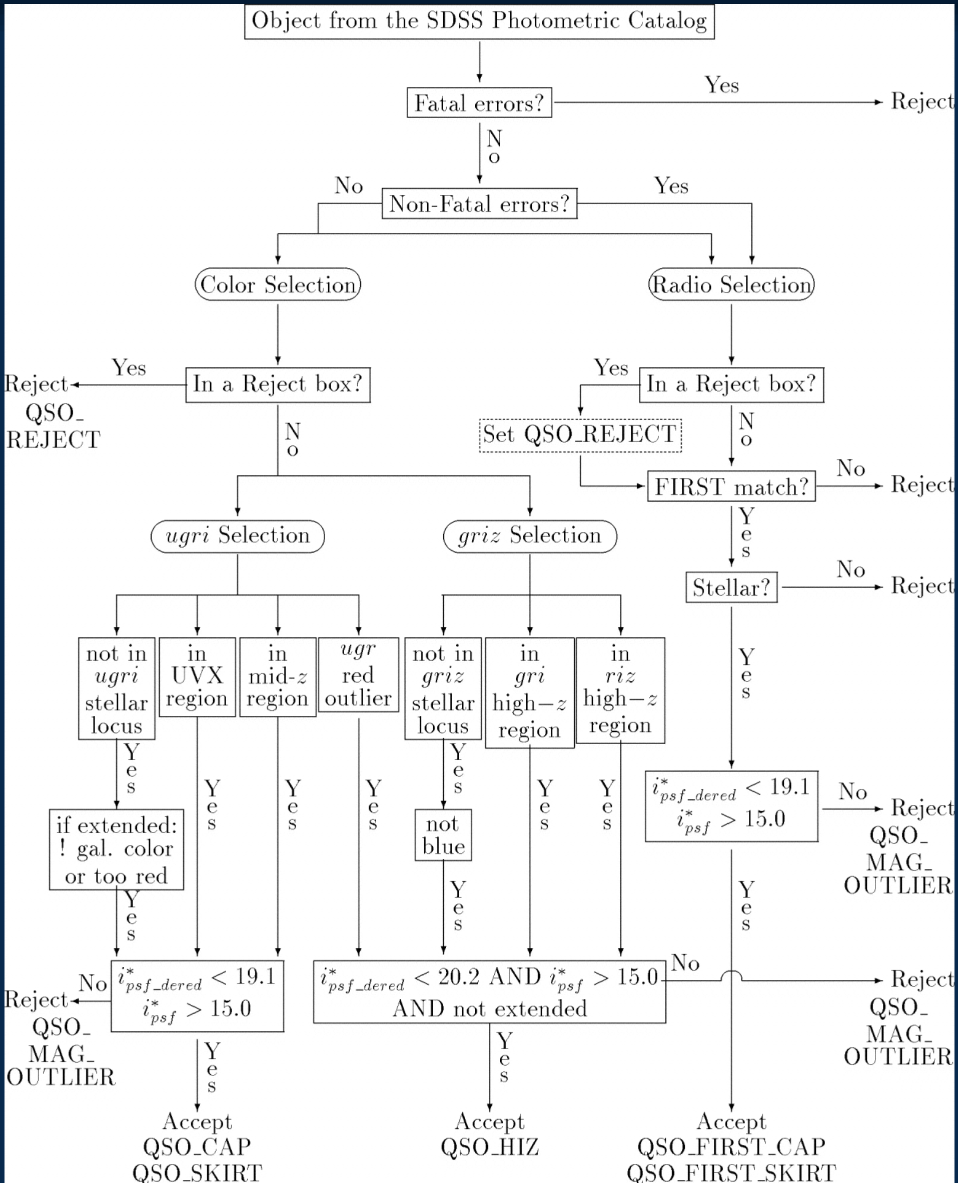
Classification of Quasars via Non-parametric Bayesian Analysis using Optical/IR Colors and Variability

Gordon Richards
Drexel University

Largely based on the work of
Drexel graduate students
Tina Peters and John Timlin
Machine learning code by Alex Gray (Skytree)

Finding Quasars: The Old Way

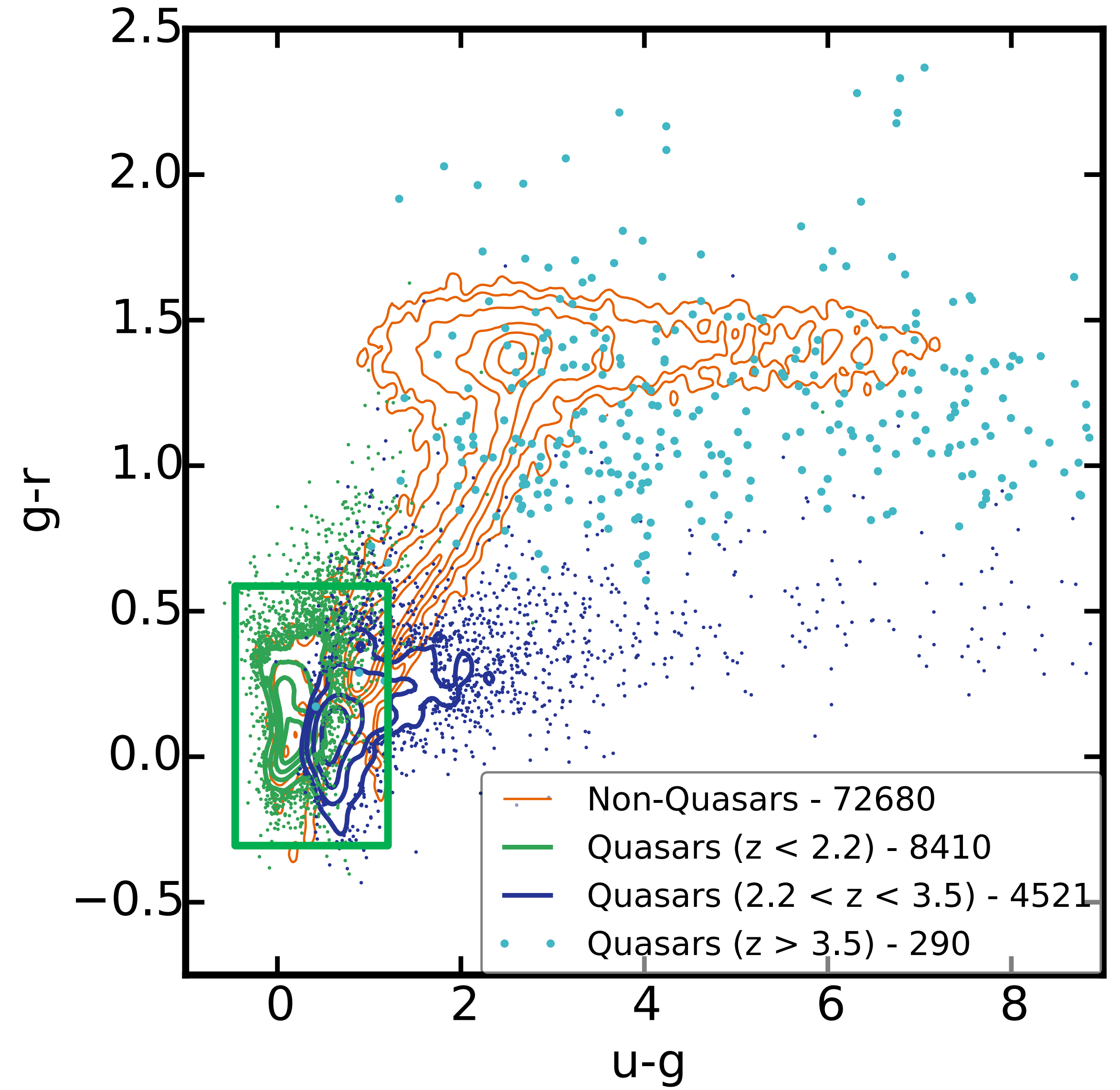
Complex, Glorified Color Cuts



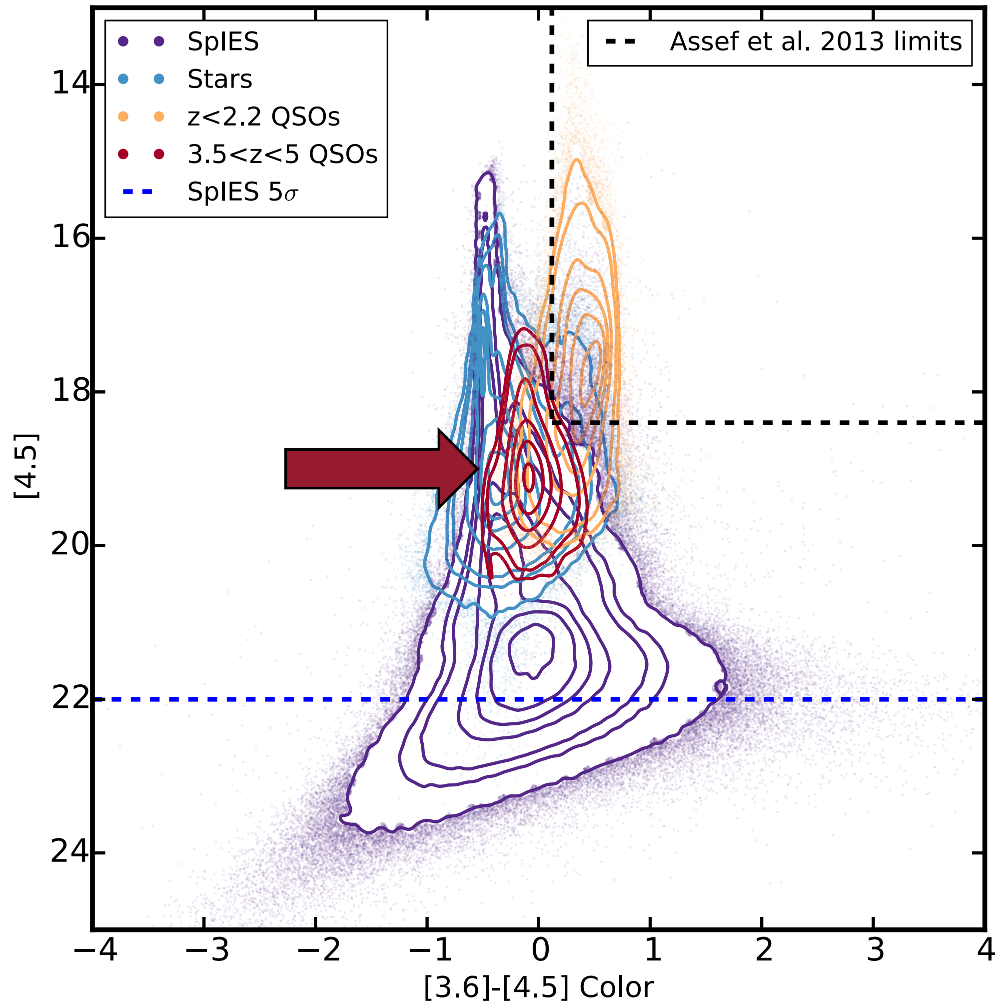
Richards et al. 2002

Naïve Color Selection

Color cuts can be arbitrarily complete, but suffer in efficiency (or purity).



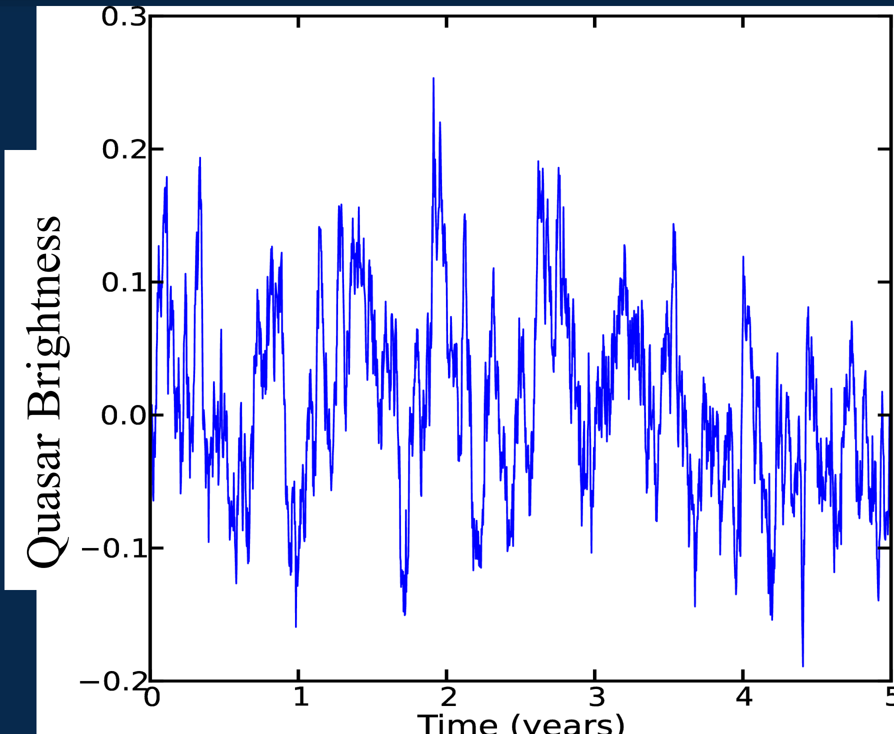
MIR Selection of High-z Quasars

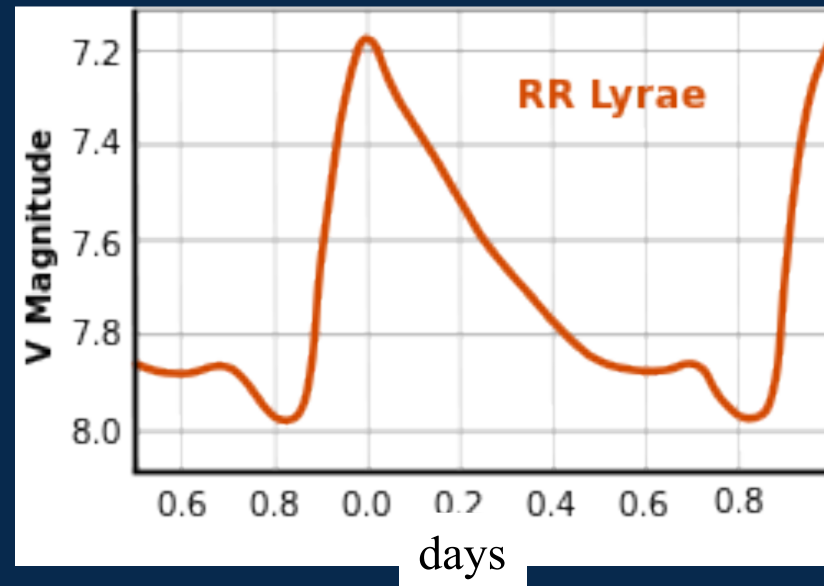


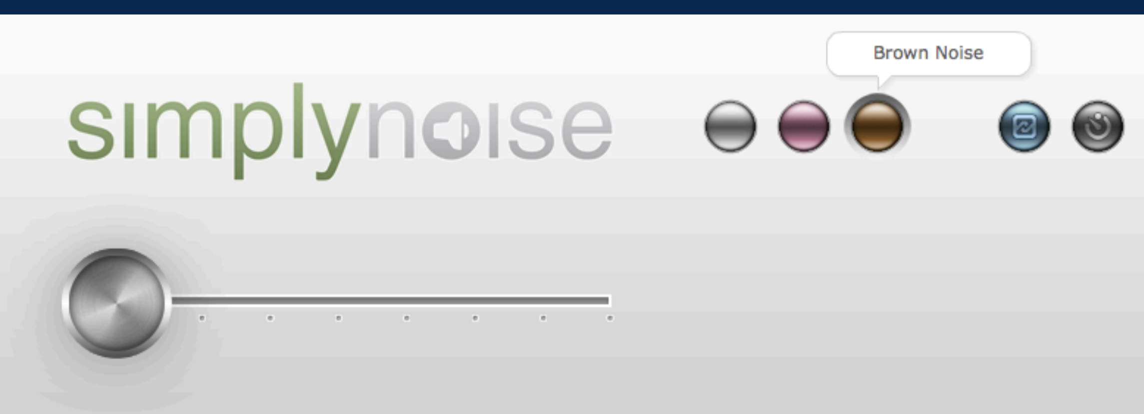
Standard MIR color selection works well for low-z, but is completely blind to high-z

Timlin et al. 2016

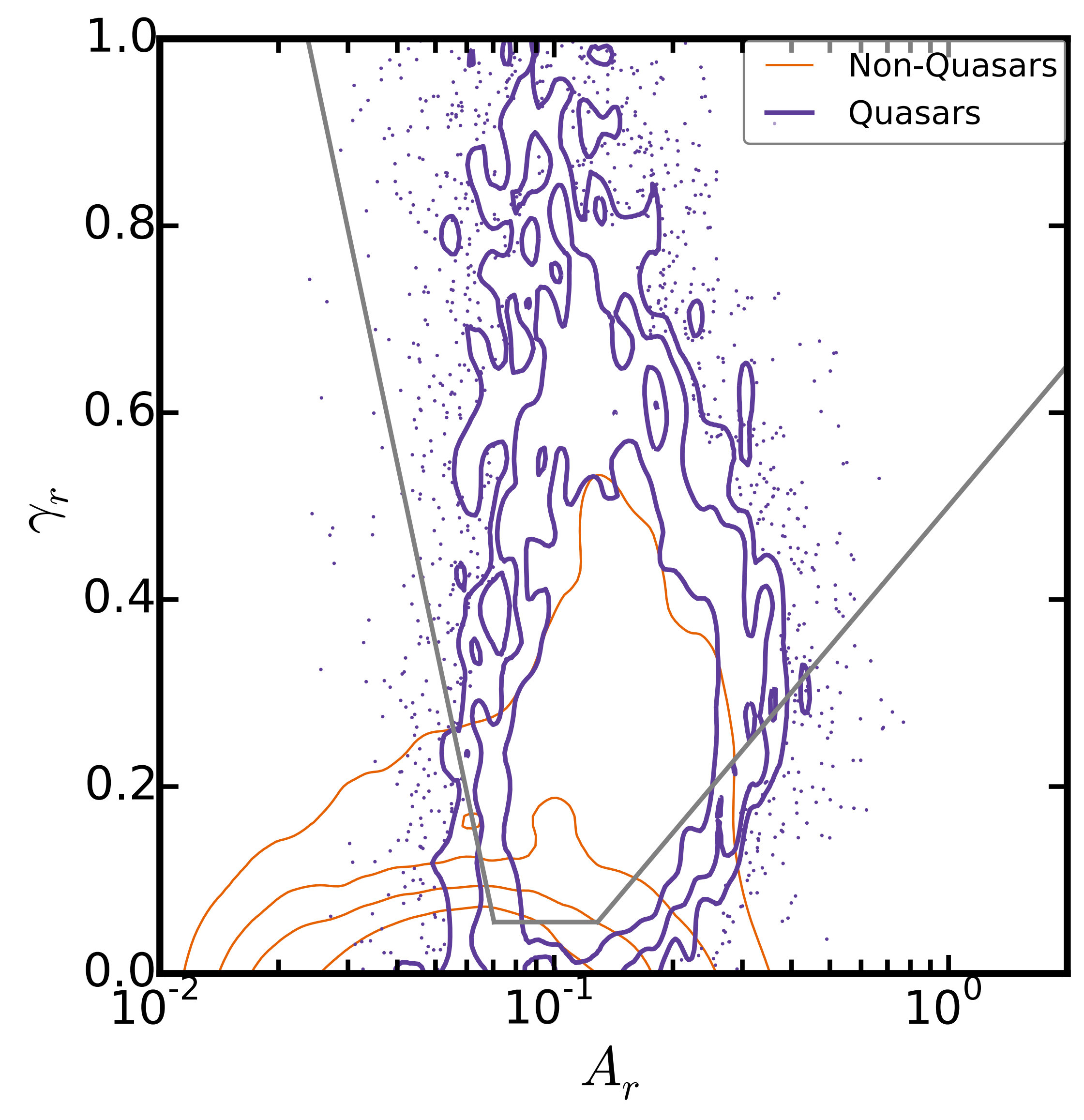
Quasars = Brown Noise







Characterized by Structure Function



$$SF \propto A(\Delta t)^{\gamma}$$

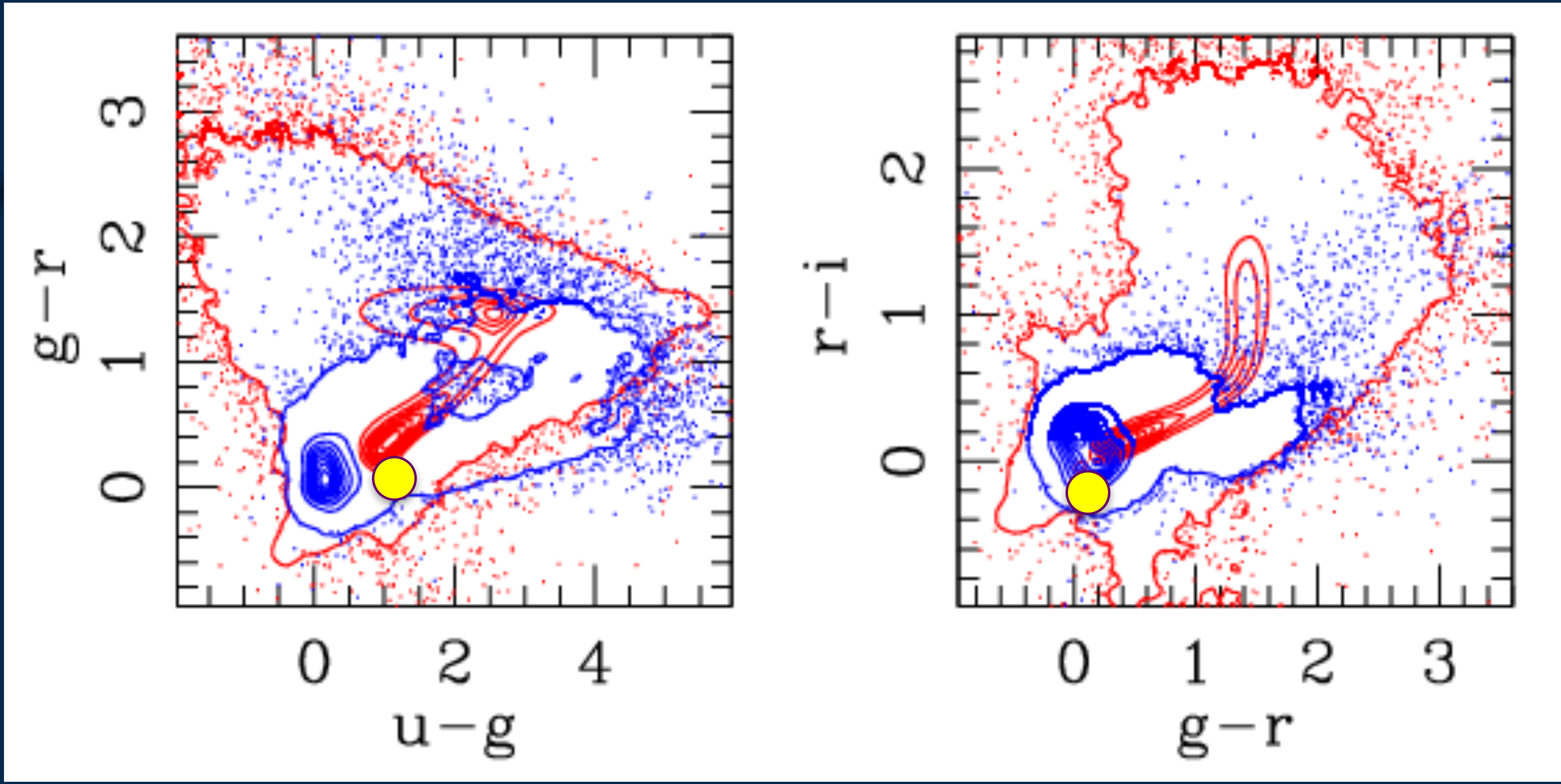
Quasars are variable, but not always distinguishable from stars

Peters et al. 2015

Bayesian Quasar Selection Solutions

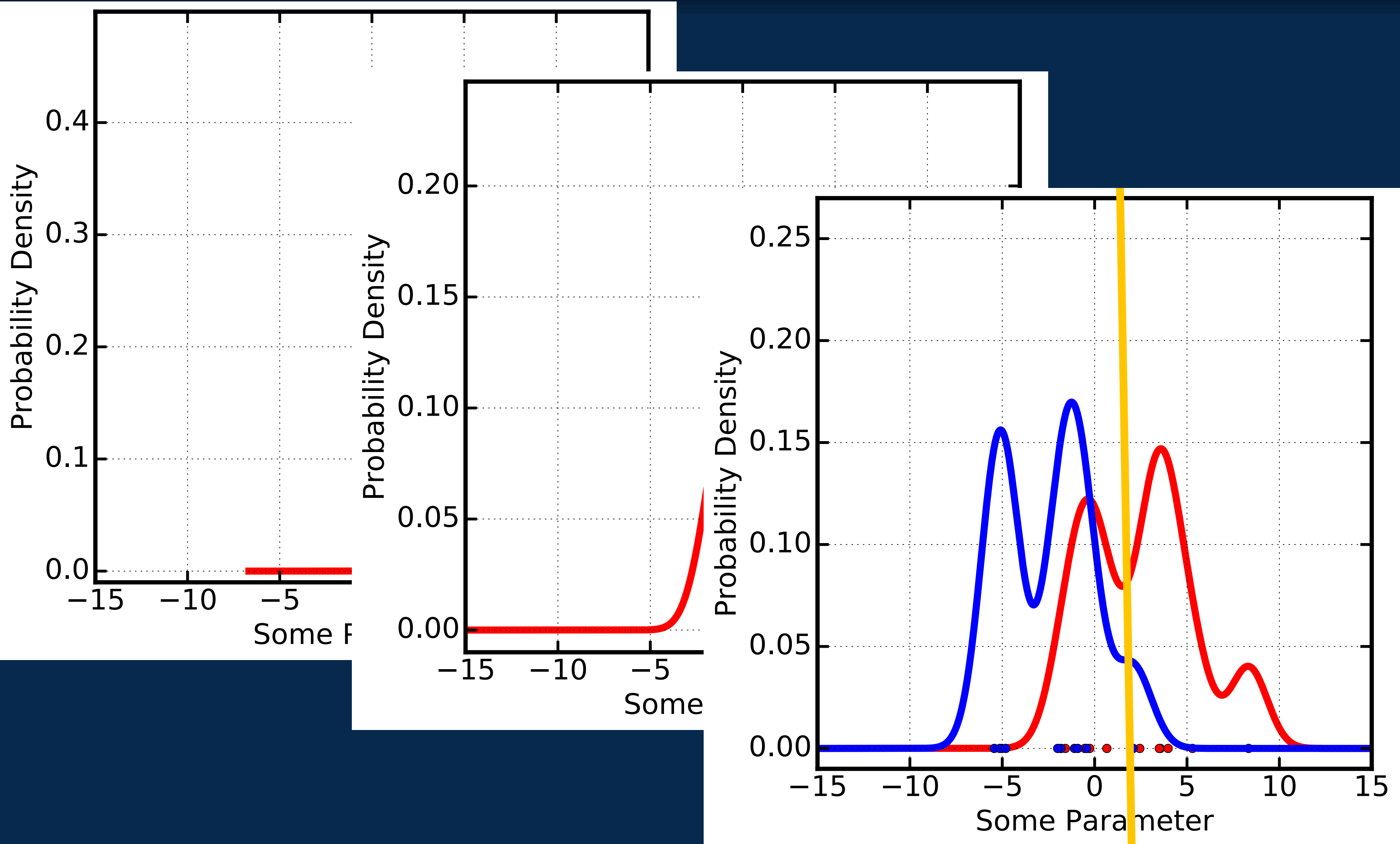
Using ALL of the data with modern statistical techniques produces much better results (completeness and efficiency).

Also allows us to probe deeper.



Given two training sets, Here quasars and stars (non-quasars), and an unknown object, which class is more likely?

KDE Methodology



“NBC”: Bayes’ (1763) Rule

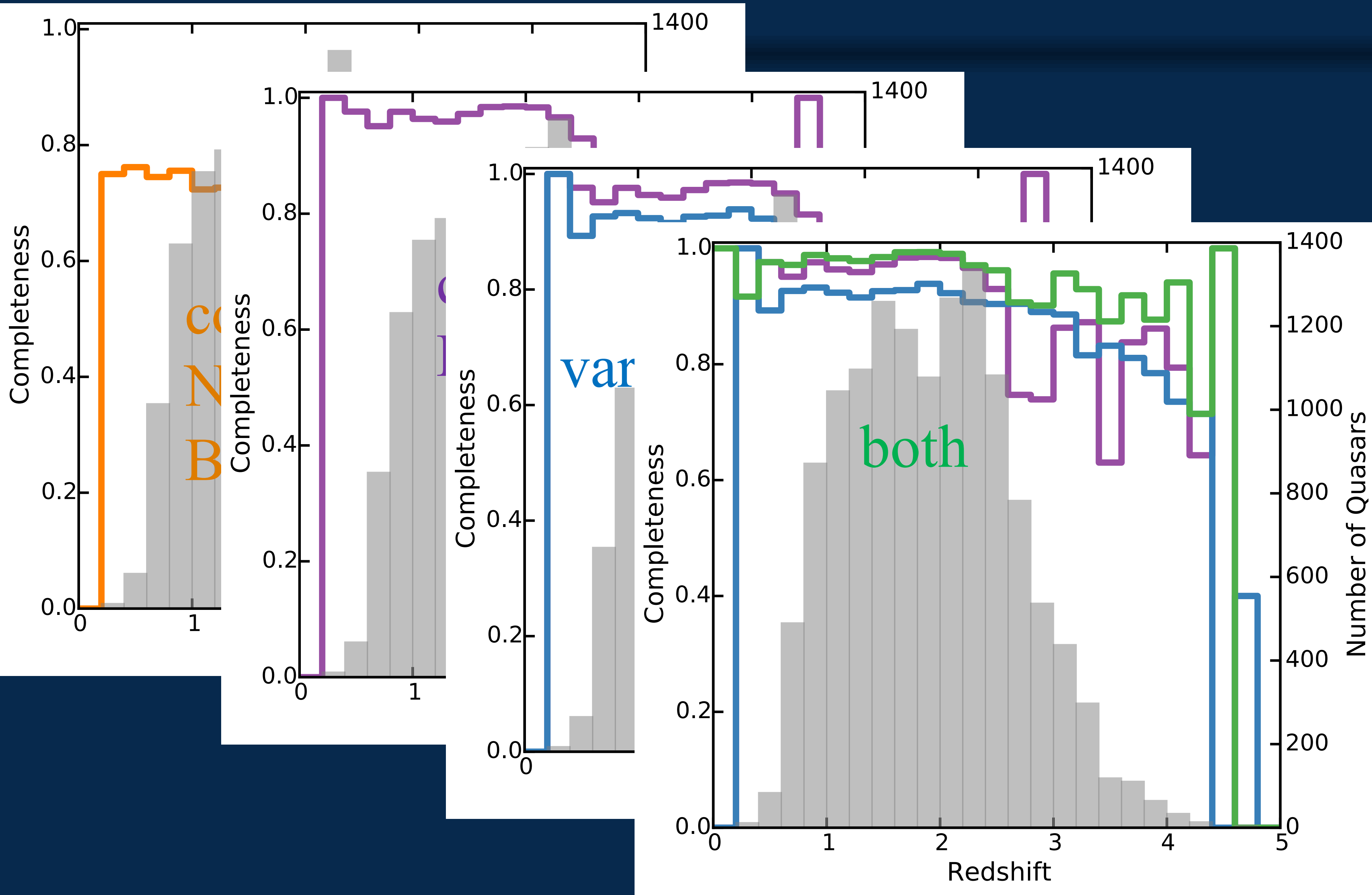
$$P(Star \mid x) = \frac{P(x \mid Star)P(Star)}{P(x \mid Star)P(Star) + P(x \mid QSO)P(QSO)}$$

Where

- x = N-D colors
- P(Star|x) = probability of being a star, given x
- P(x|Star) = probability of x, drawing from stars training set
- P(x|QSO) = probability of x, drawing from QSO training set
- P(Star) = stellar prior
- P(QSO) = quasar prior
- P(Star) + P(QSO) = 1
- Star if P(Star|x)>0.5, QSO if P(Star|x)<0.5

With help from Alex Gray (Georgia Tech)

Results Colors vs. Var vs. Both



Our Recent Catalogs

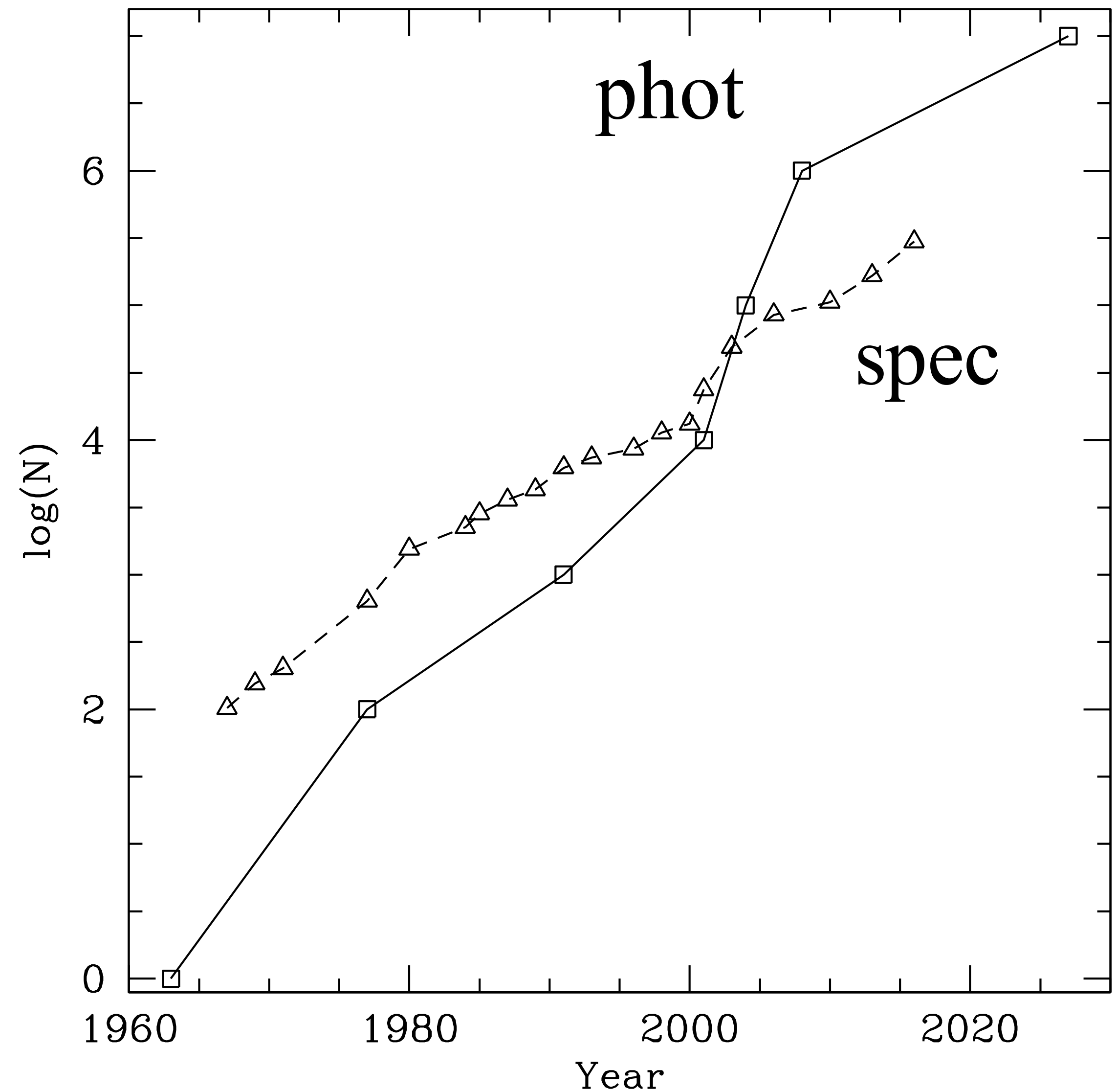
Optical + IR over full sky with WISE:
890k quasars (from 160k training objects) including 7800 at $z>3.5$ (Richards et al. 2015)

Optical + IR in SDSS Stripe 82 with Spitzer:
6500 at $z>3.5$ (in ~1/400th of sky)

Color + Variability in SDSS Stripe 82:
36000 at $0<z<5$ (Peters et al. 2015)

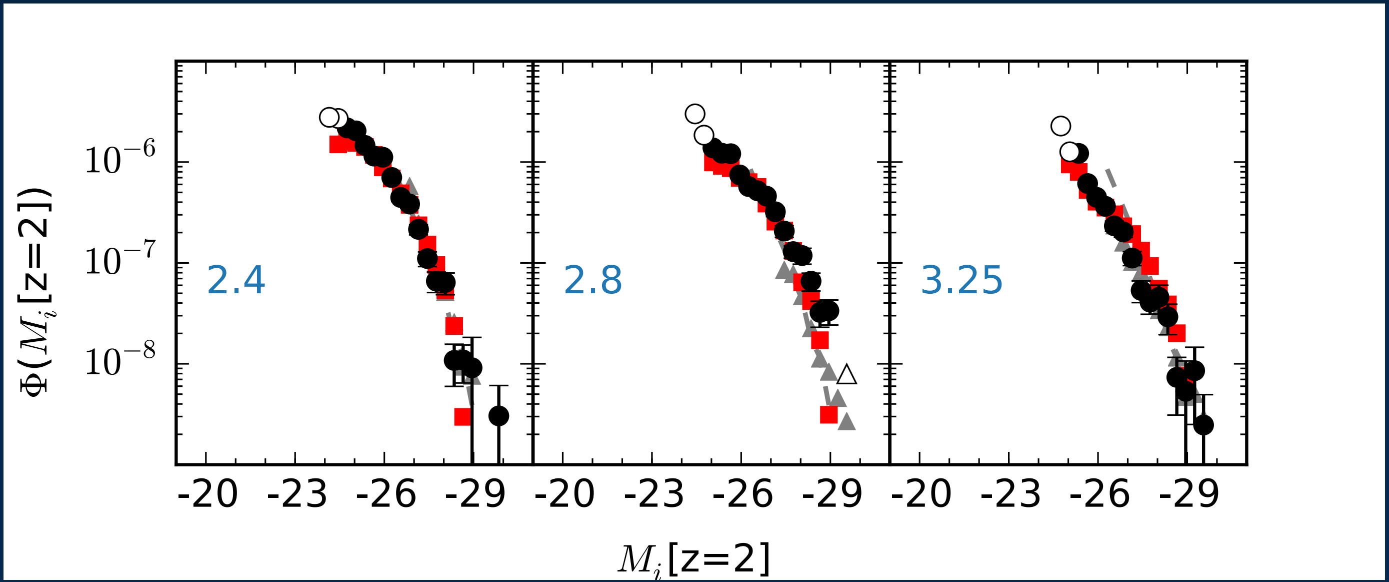
Photometric > Spectroscopic

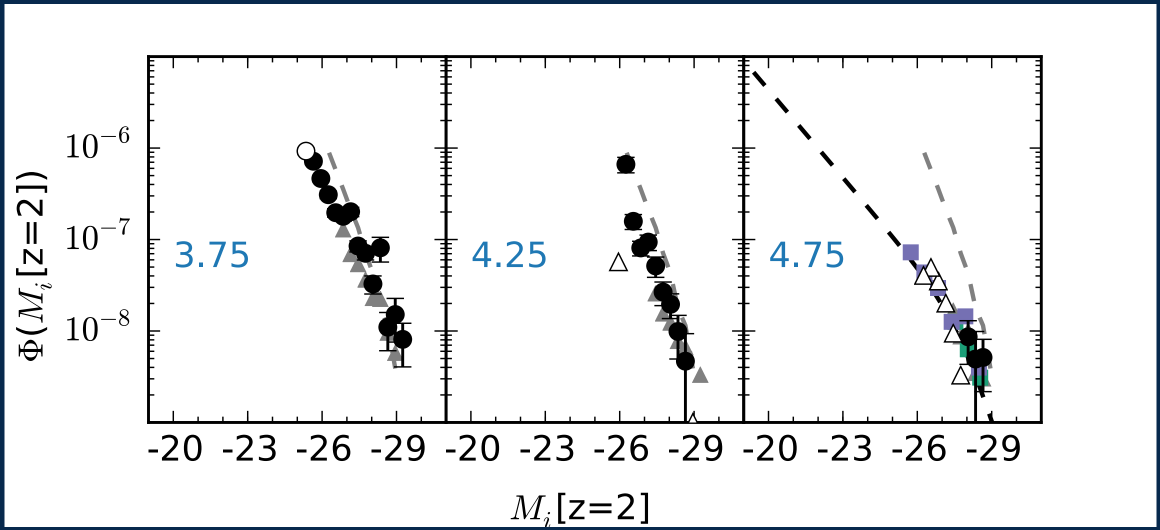
Increase in spectroscopic (triangles) and photometric (squares) quasar samples.



Updated from Richards et al. 2004

Doing Real Science: Quasar Luminosity Function

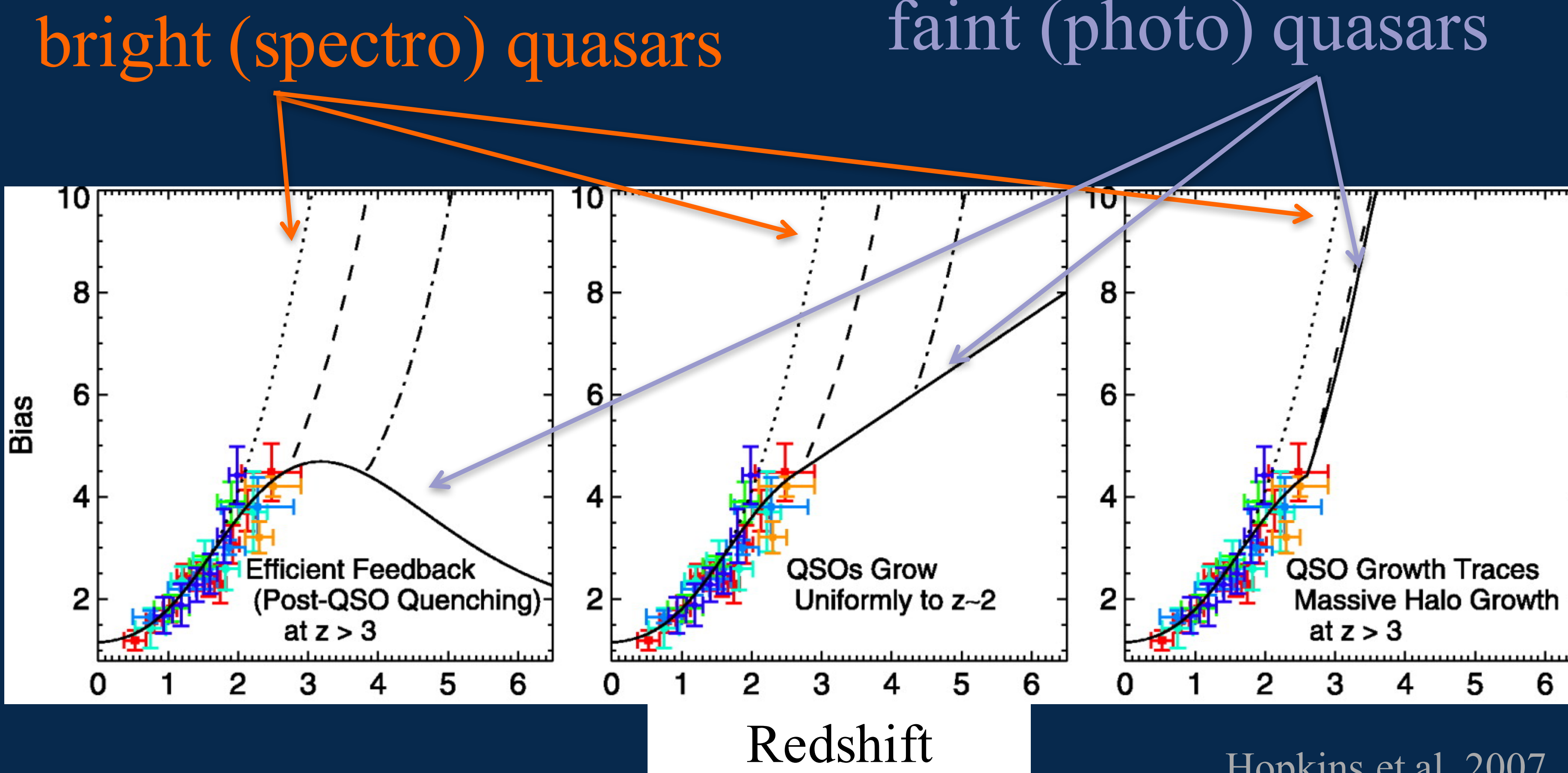




Peters et al. 2015

Doing Real Science: Clustering

Measuring bias of faint high-z quasars will break degeneracies between feedback models.



Hopkins et al. 2007

A Word of Caution

It is important to know the algorithms well and to try new ones.

But it is even more important to know the data well and realize when your results are garbage.

E.g., ~half of all the entries in the SDSS database are spurious

Advertisement

See Tina Peter's poster
for more details on:

- Color+Variability Analysis
- Photometric (and Astrometric!) Redshift determination
- Independent Component Analysis

Summary

- Color cuts are still norm in object classification.
- Significant gains to be had from using modern statistical techniques.
- Quasar samples are >10x larger as a result.
- Combining unrelated data sets is particularly powerful.
- Work should be driven by science and not algorithms.
- Ask me if interested in using our catalogs and please suggest other algorithms for us to try.