

Clustering-based redshift estimation



Mubdi Rahman



Alex Mendez

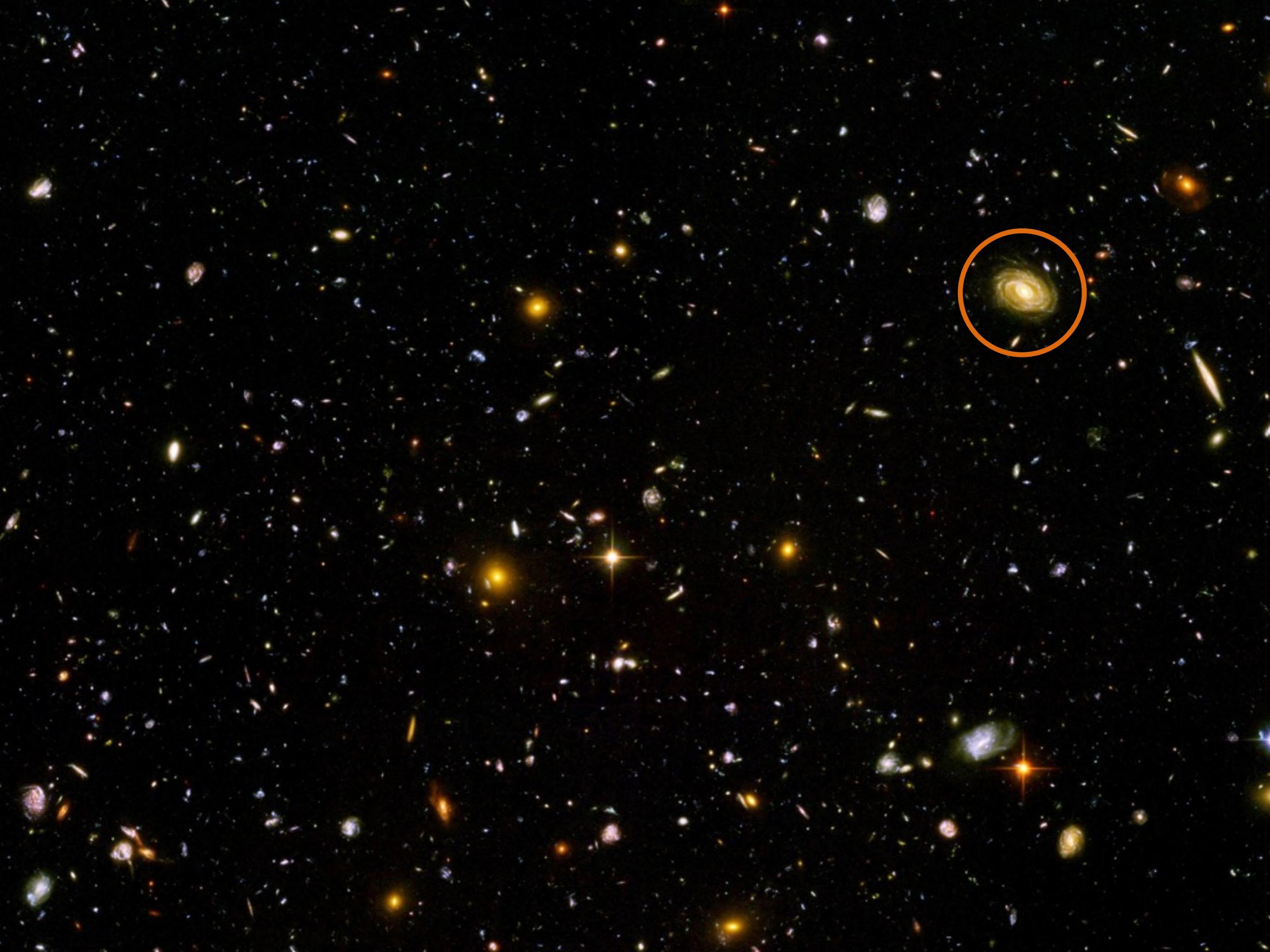


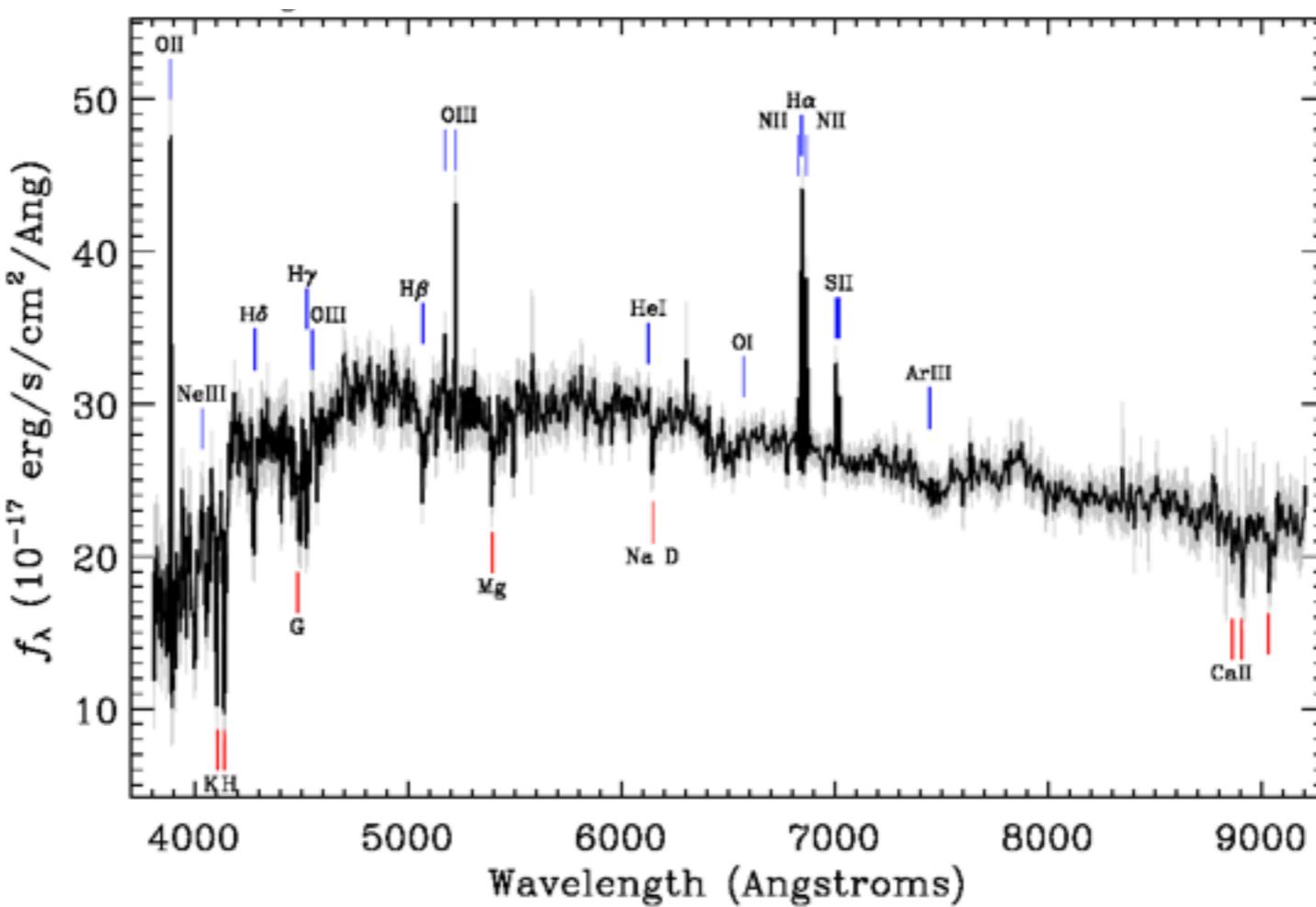
Ting-Wen Lan

+ Ryan Scranton, Sam Schmidt, Chris Morrison

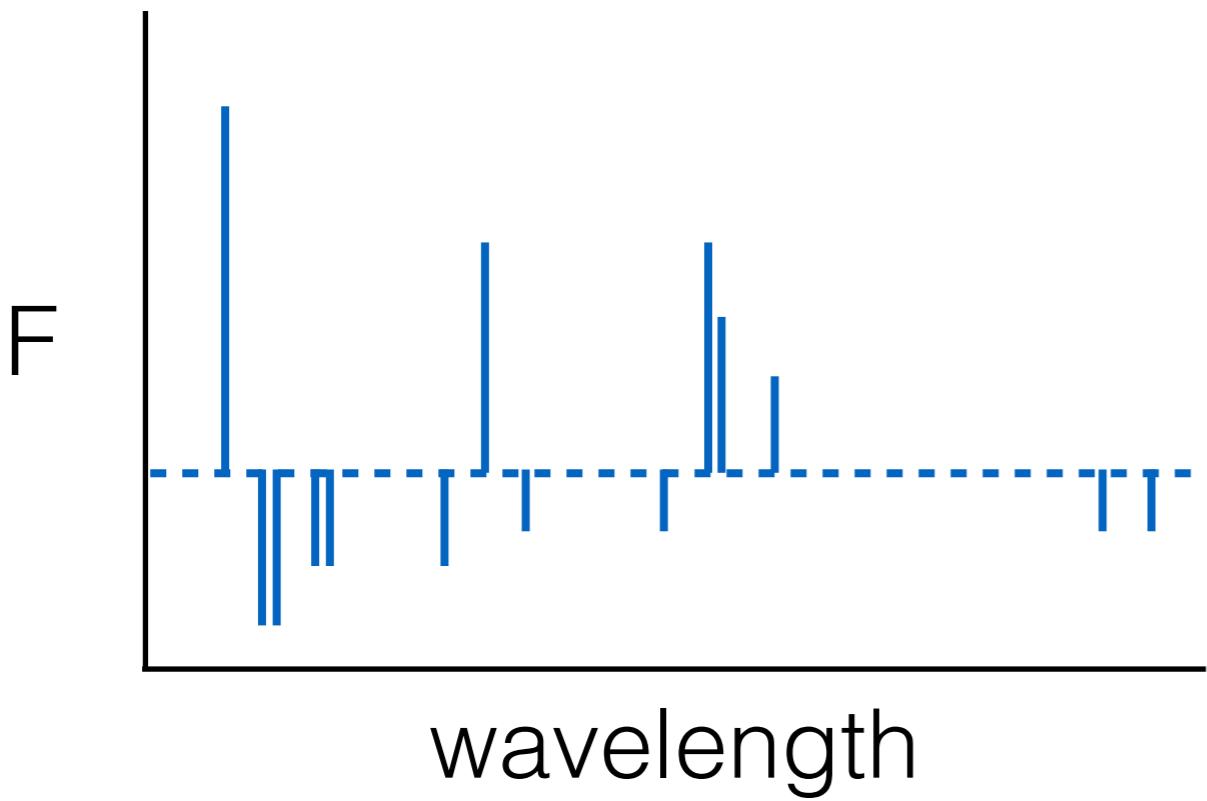
Brice Ménard

Johns Hopkins University
Kavli IPMU, Tokyo University

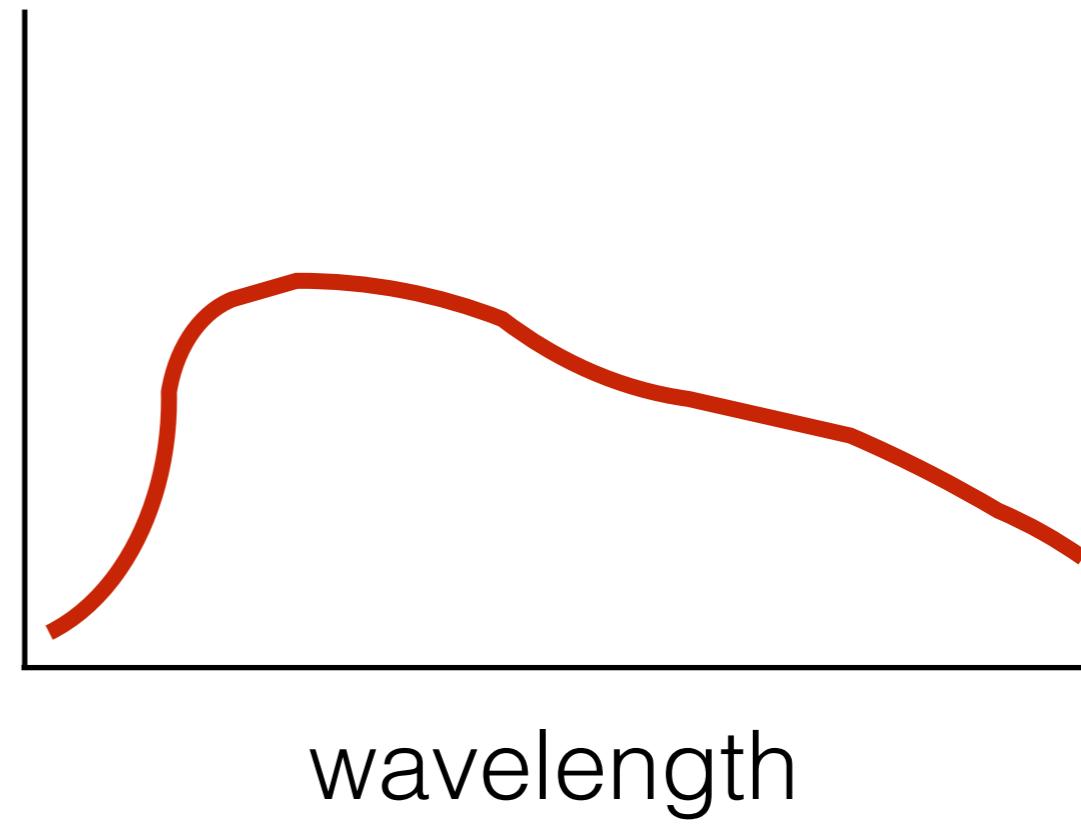


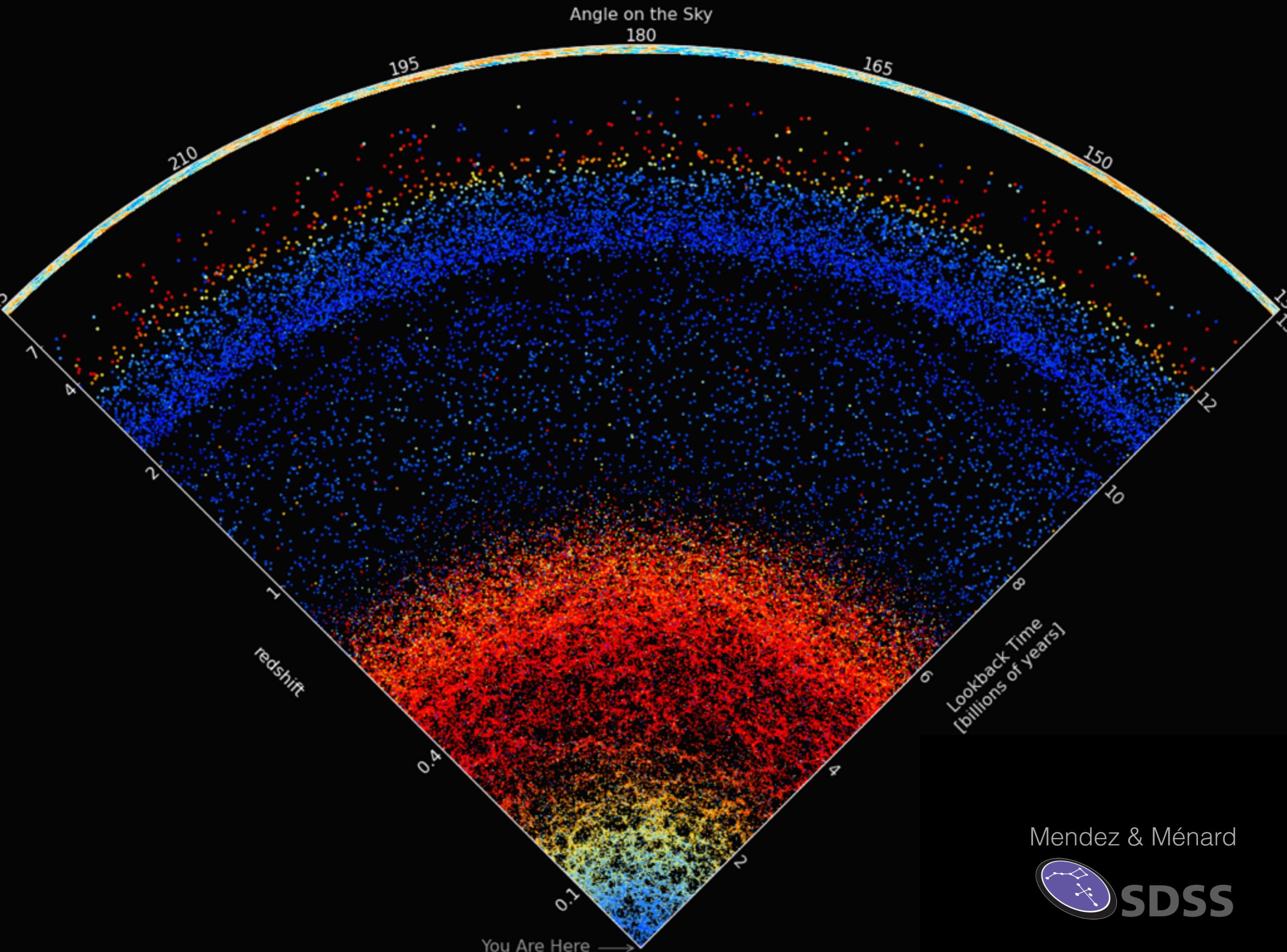


Spectroscopic redshift

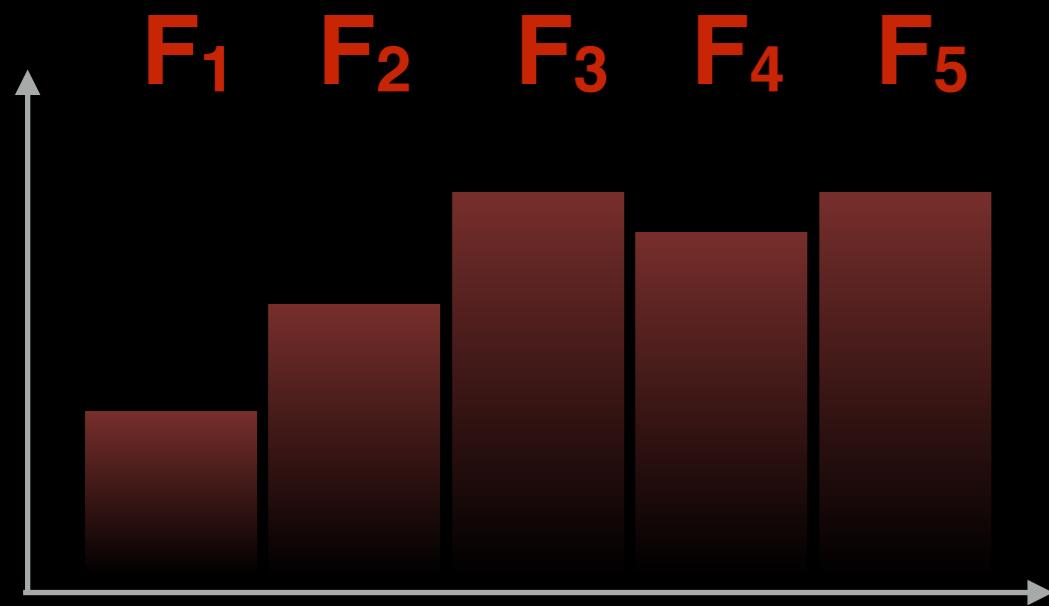


Photometric redshift

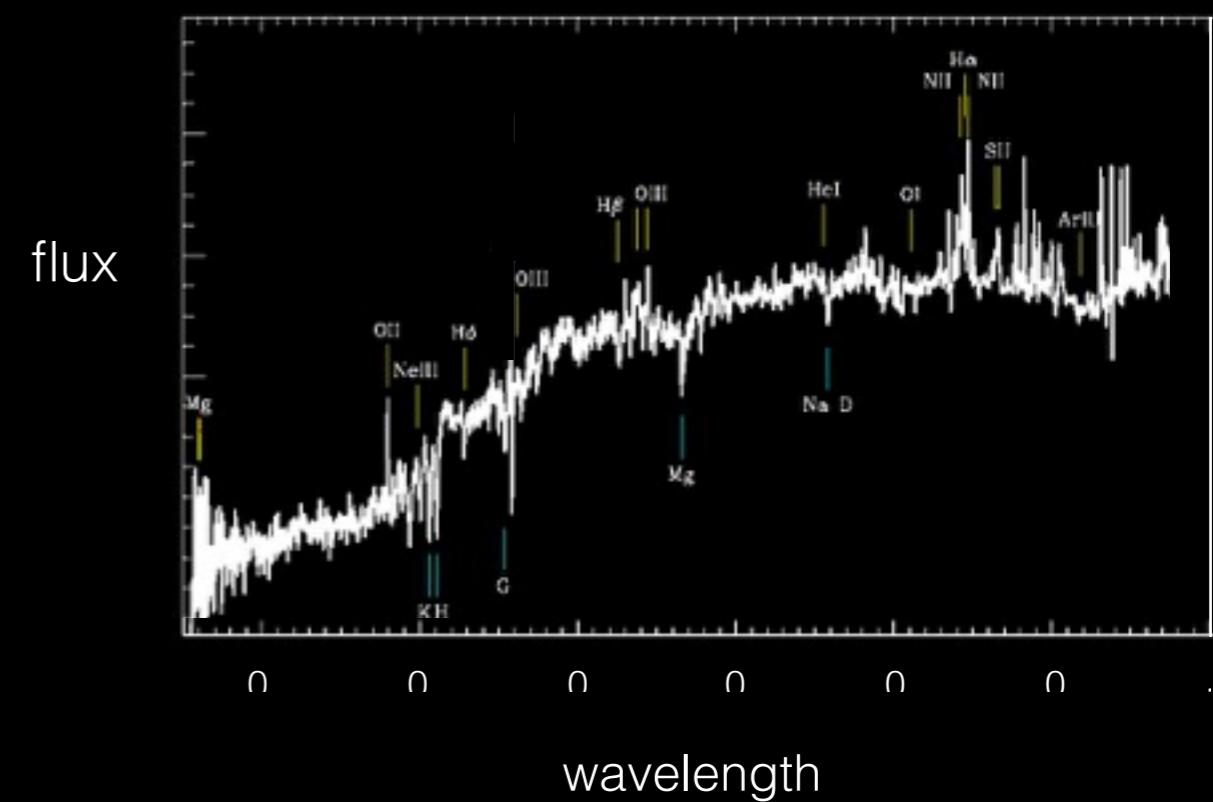




imaging

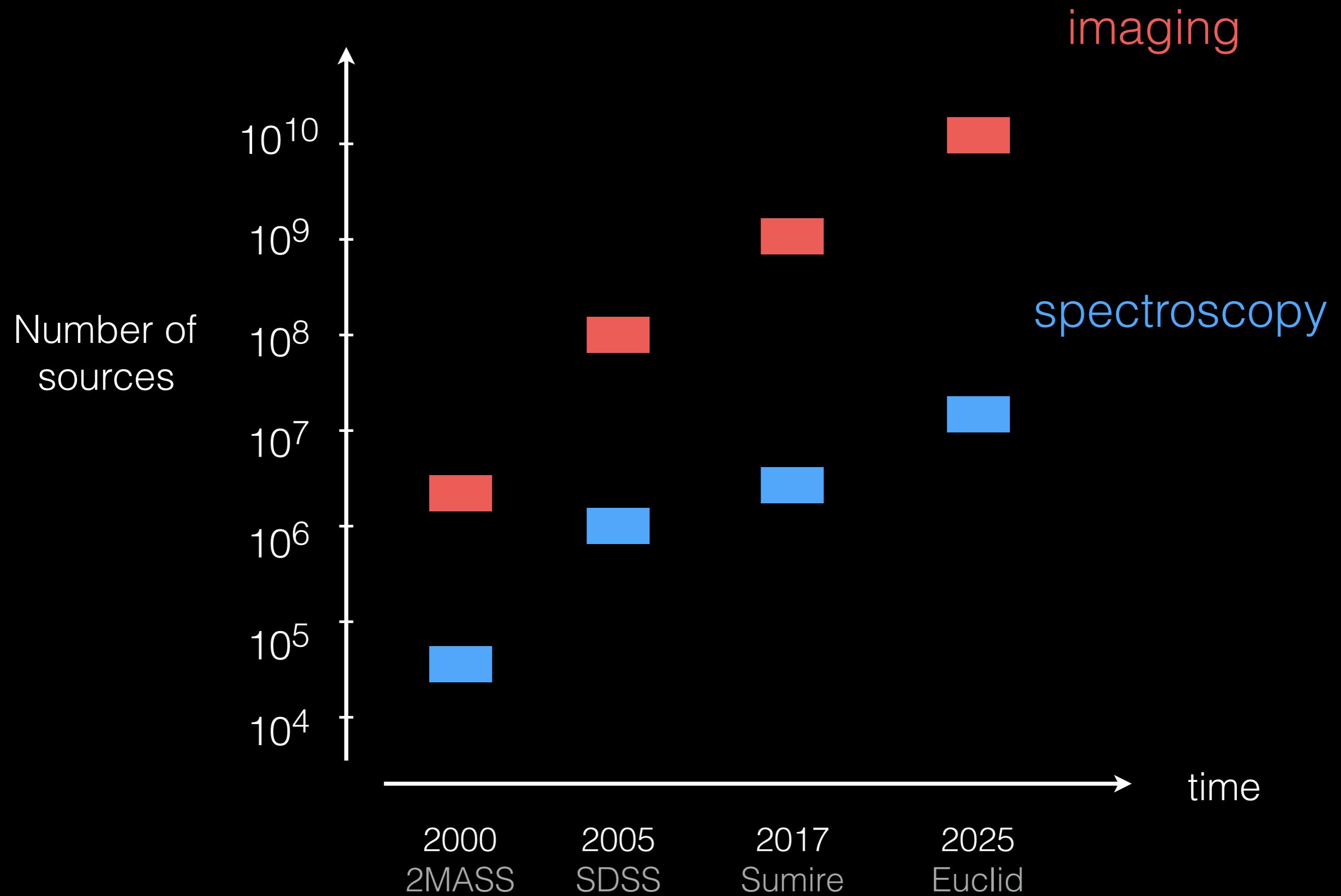


spectroscopy



99%

1%

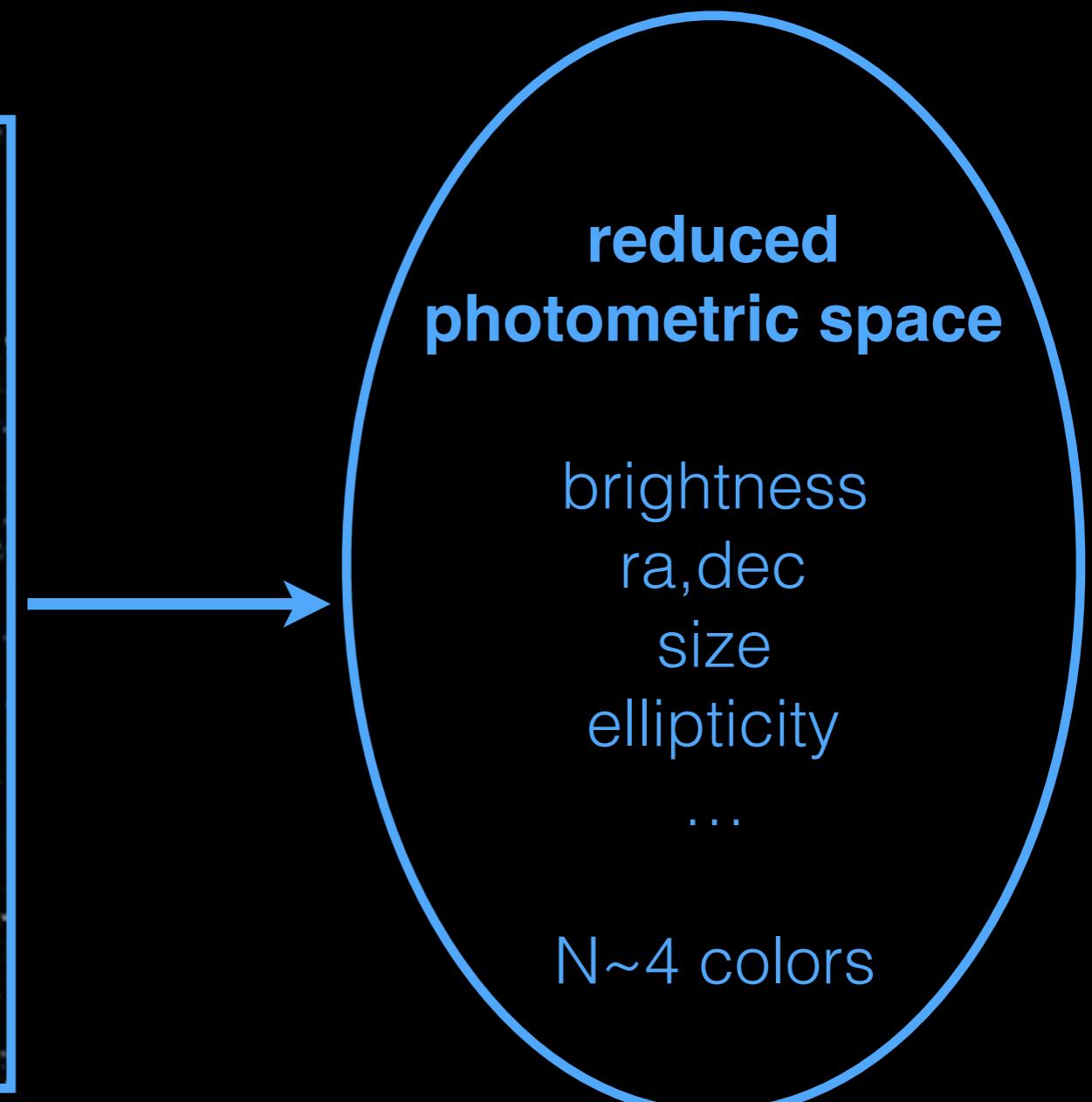


How much information goes into the catalogs?

main data product:
pixel based



working environment:
object based

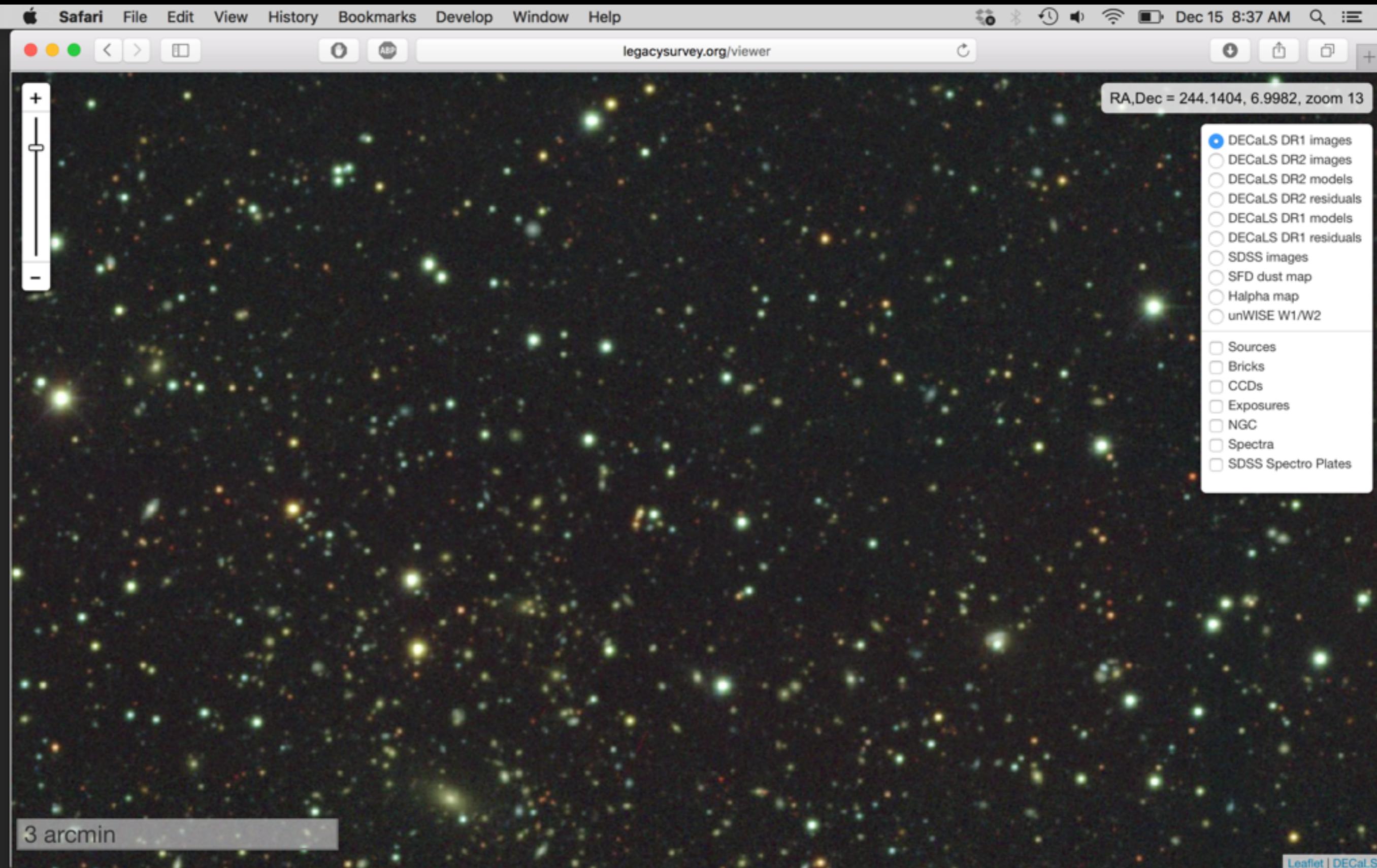


Dimensionality ~ 10

DECaLS survey

PIs: Dey & Schlegel

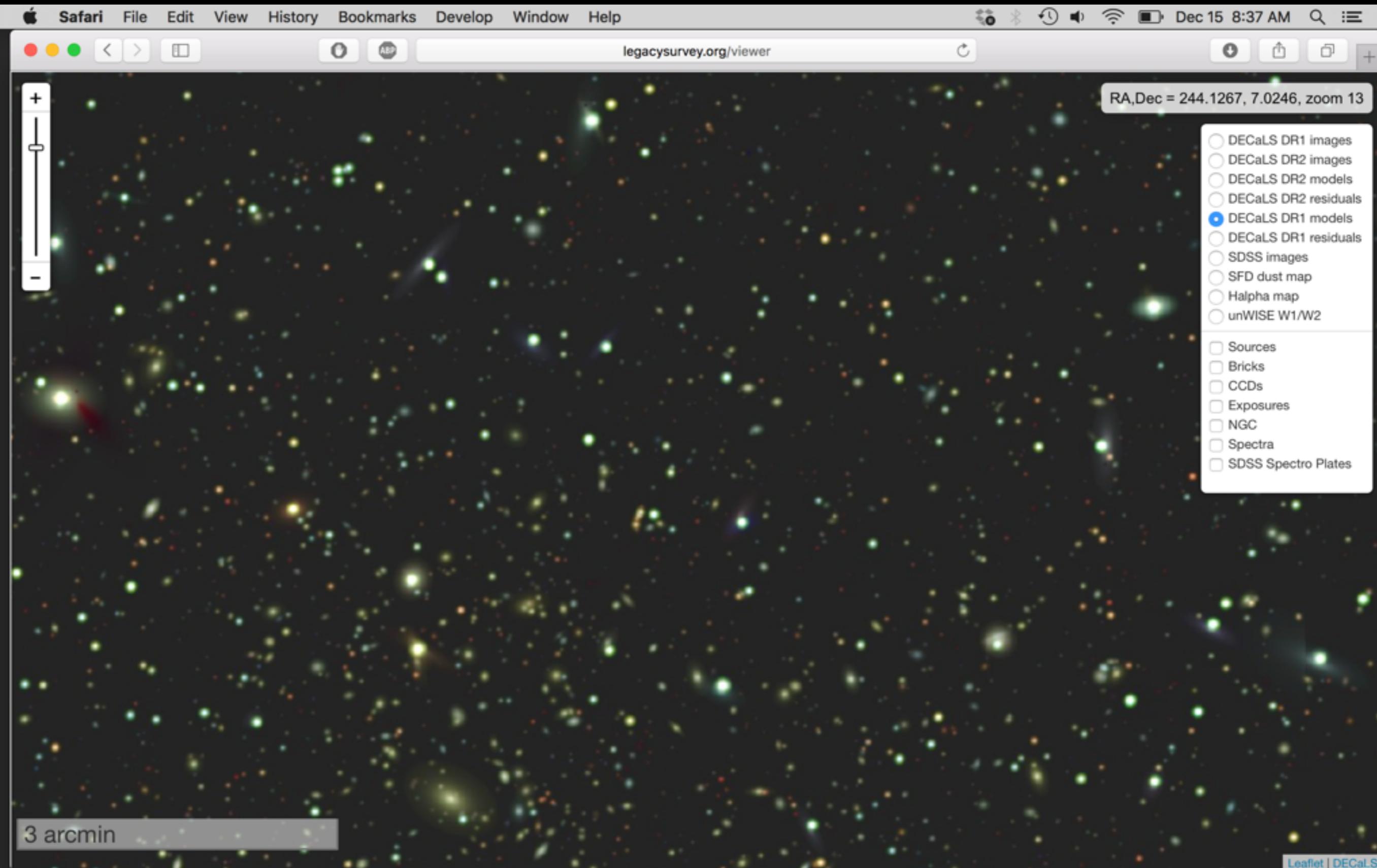
Visualization: D. Lang



DECaLS survey

PIs: Dey & Schlegel

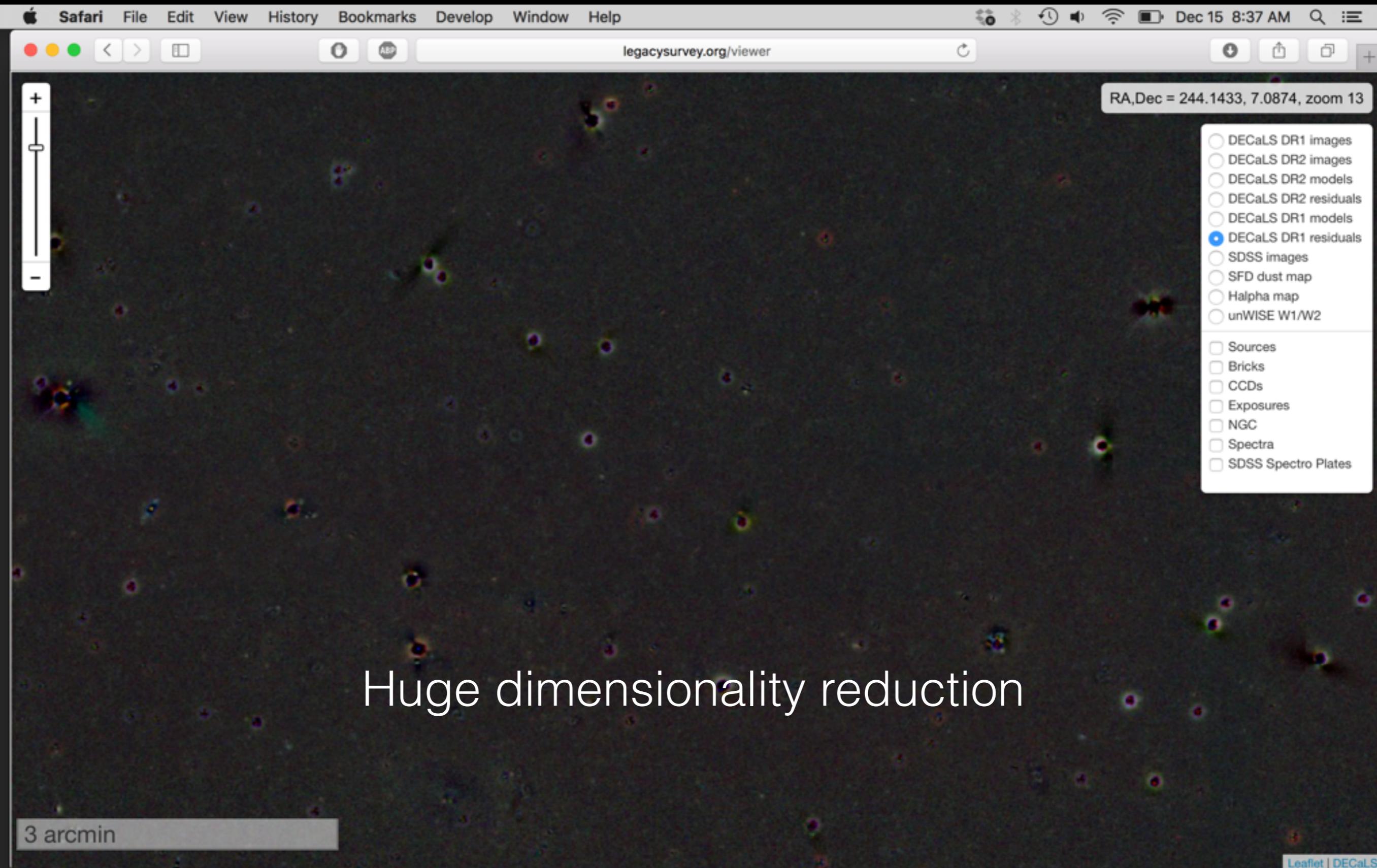
Visualization: D. Lang



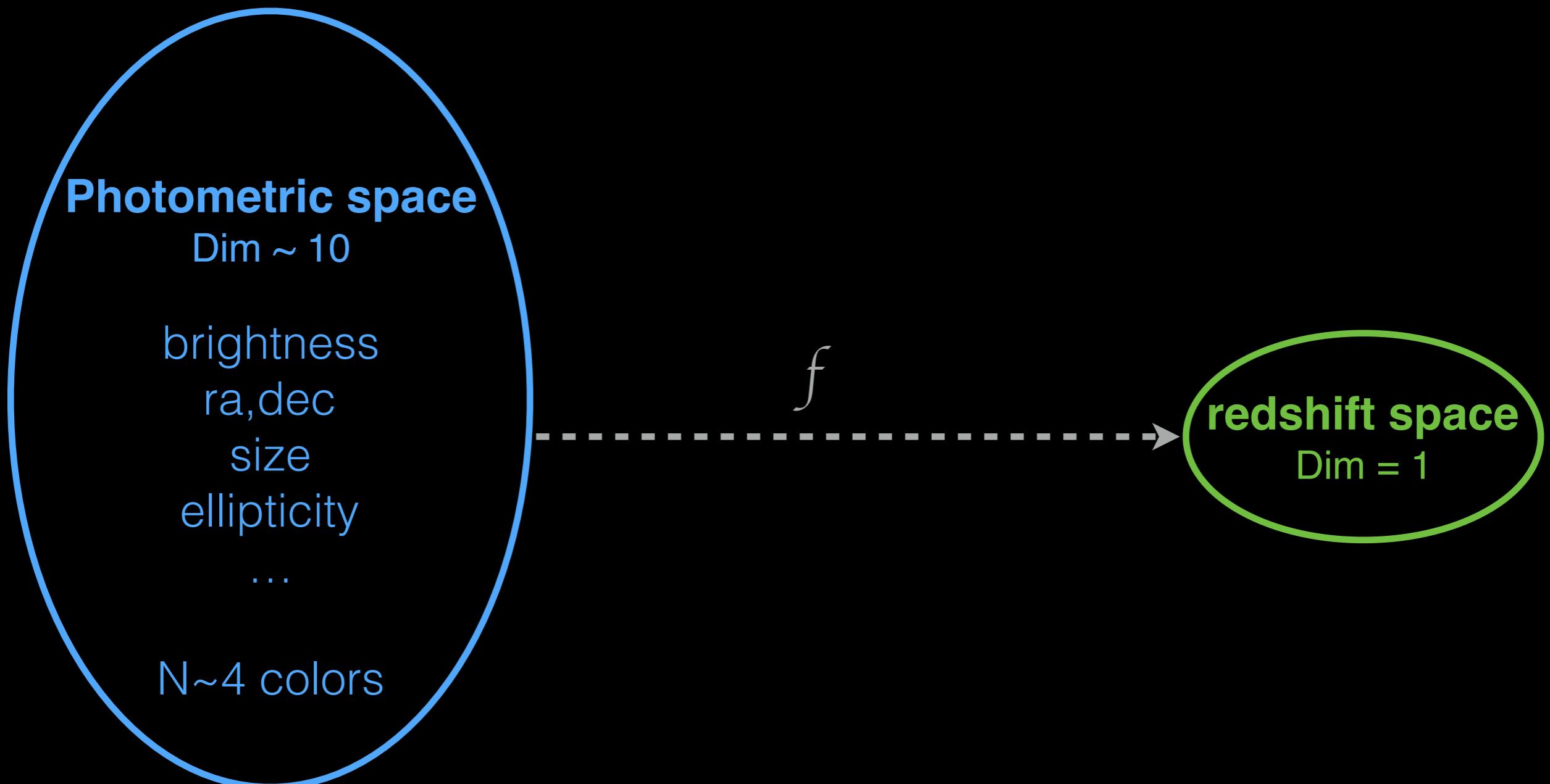
DECaLS survey

PIs: Dey & Schlegel

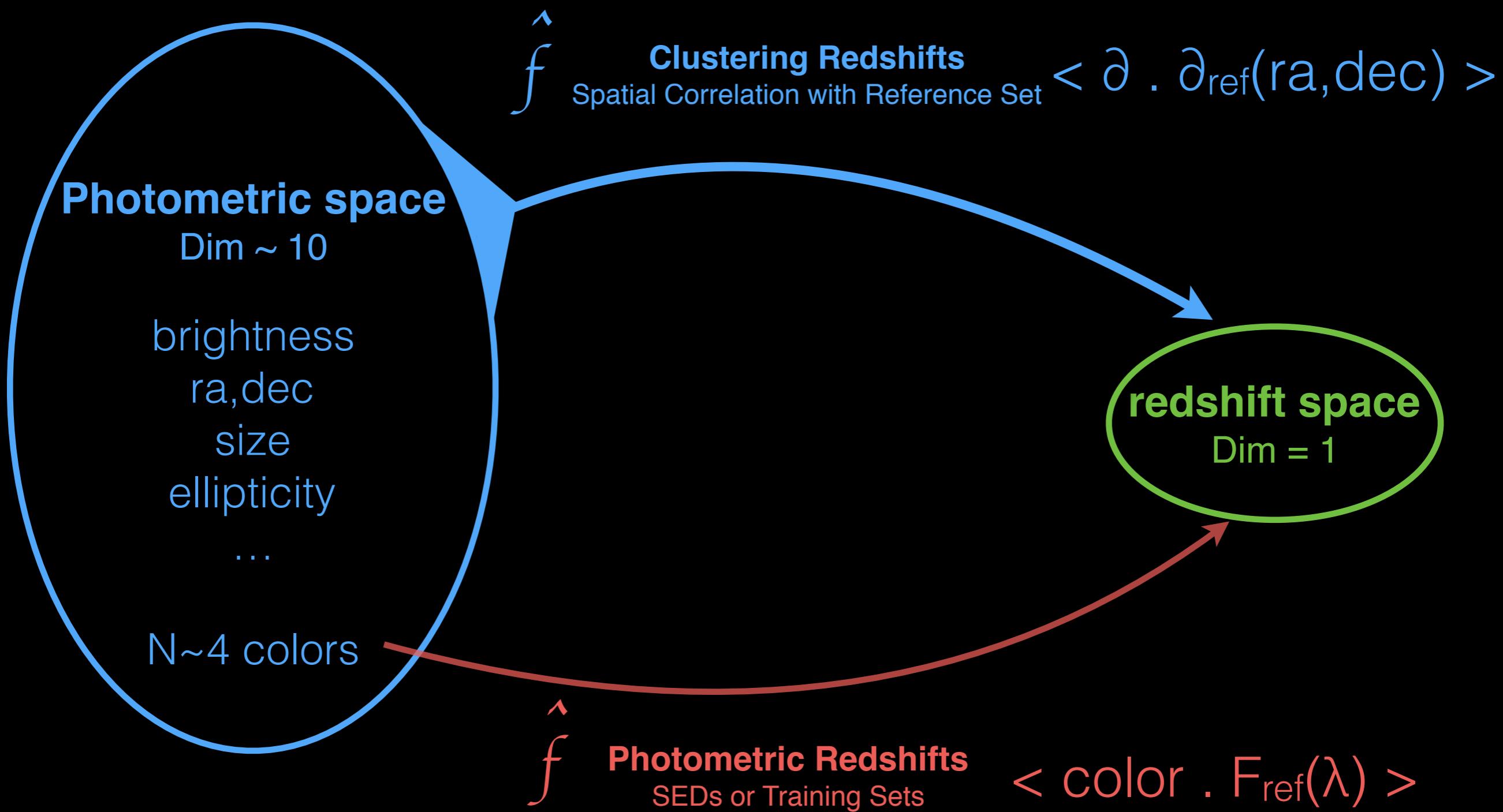
Visualization: D. Lang



Mapping the photometric space to redshift space



Mapping the photometric space to redshift space



Estimating redshifts using clustering

an old idea...

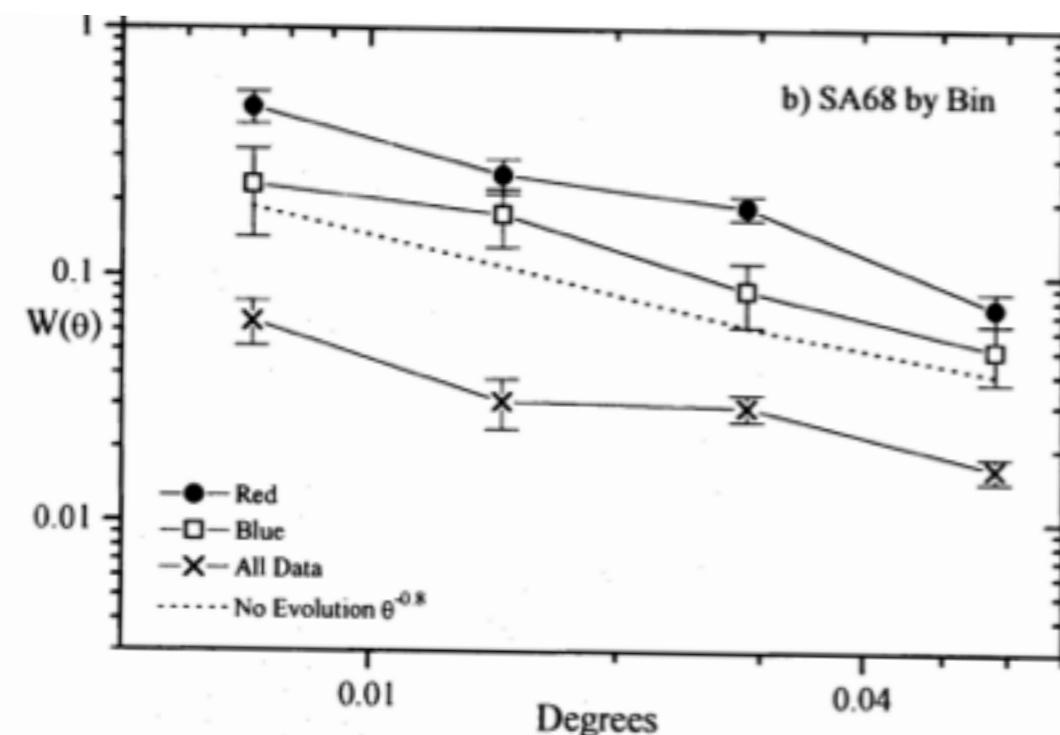
Seldner & Peebles (1979), Roberts & Odell (1979)

Landy, Szalay, Koo (1996)

STRONG ANGULAR CLUSTERING OF VERY BLUE GALAXIES: EVIDENCE OF A LOW-REDSHIFT POPULATION

STEPHEN D. LANDY,¹ ALEXANDER S. SZALAY,^{2,3} AND DAVID C. KOO⁴

Received 1995 February 17; accepted 1995 September 22



AUTO- AND CROSS-CORRELATIONS BY COLOR BAND

	Blue (20%)	Mid (60%)	Red (20%)
Blue	68 ± 20	9 ± 7	-18 ± 10
Mid	14 ± 1	20 ± 9
Red	130 ± 30

NOTES.—Cross- and autocorrelations between the 20% most extreme blue and red galaxies and the mid 60%, averaged over the two fields. The mean A_w for all galaxies in the two fields SA 57 and SA 68 was 18.0 ± 1.8 . Given the enhanced amplitude of the autocorrelations for the red and blue subsets, together with the weak negative cross-correlation between them, we conclude that these subsets are disparate populations.

Estimating redshifts using clustering

an old idea...

Seldner & Peebles (1979), Roberts & Odell (1979) -----

Landy, Szalay, Koo (1996) -----

[data analysis, $\langle z \rangle$]

Ho et al. (2008) -----

Newman (2008), Matthews & Newman (2010, 2012),
Schultz (2010), McQuinn & White (2013), de Putter et al. (2013)

[theory, simulations]

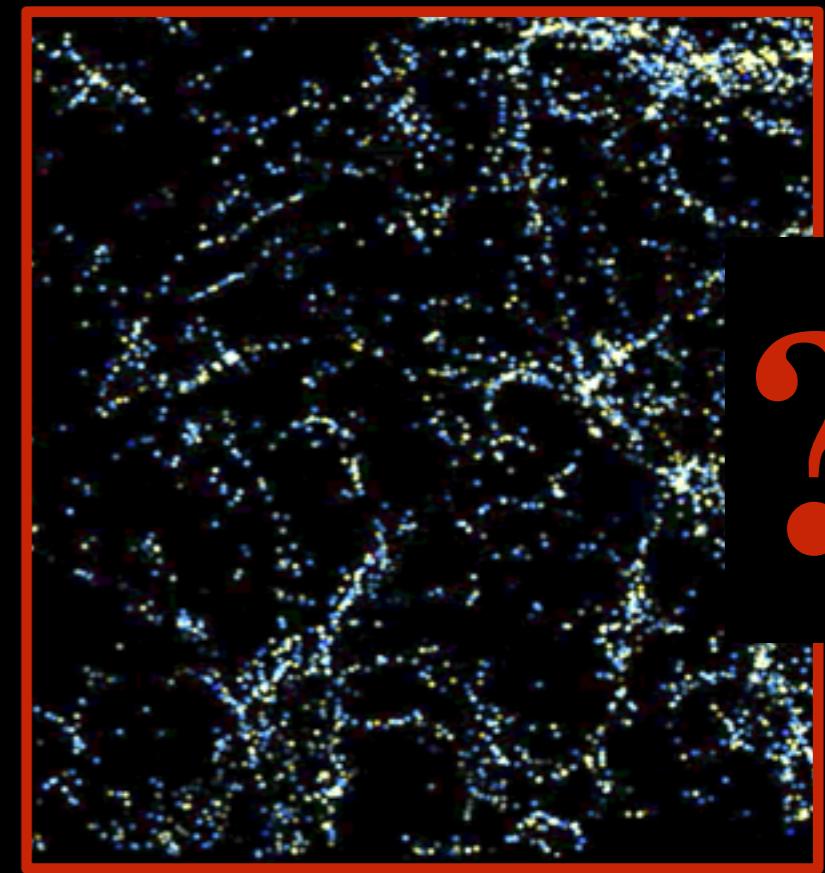
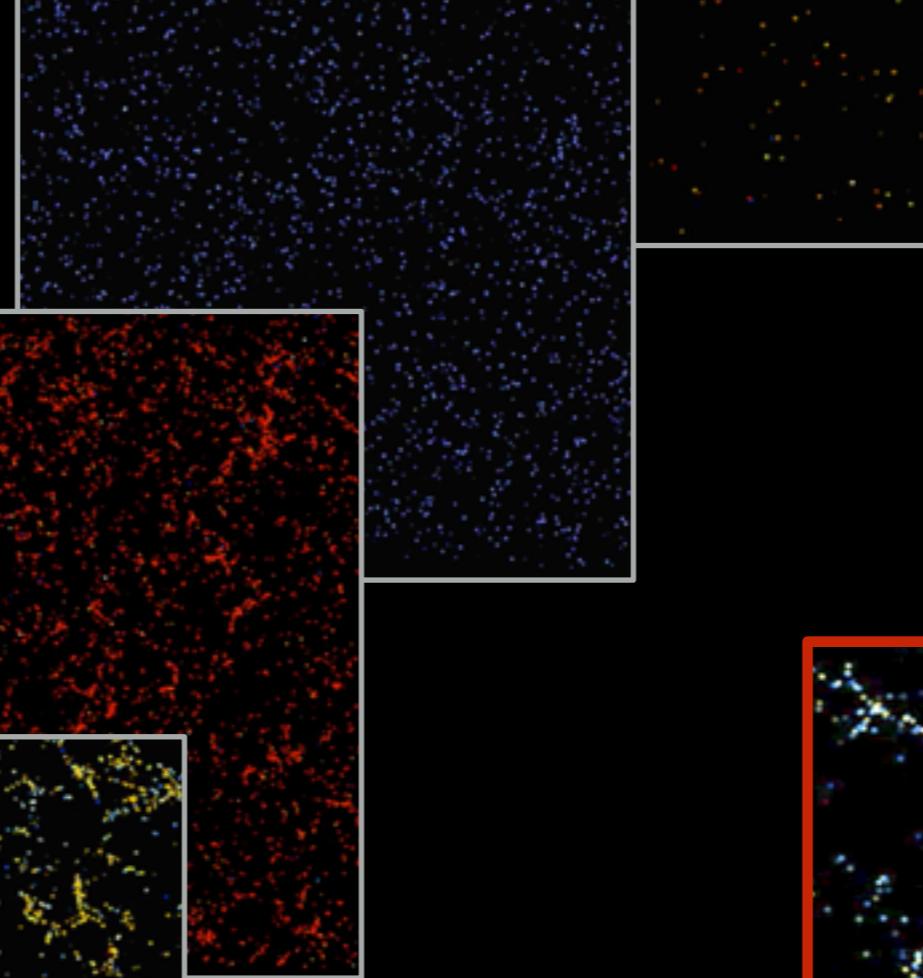
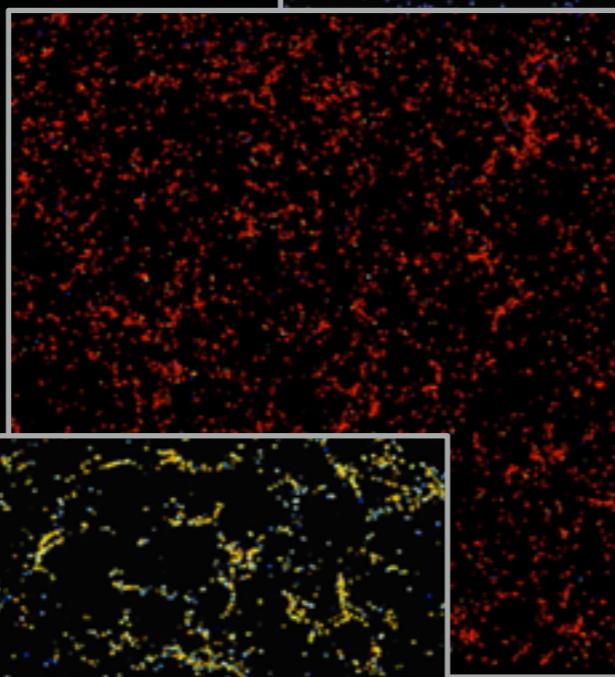
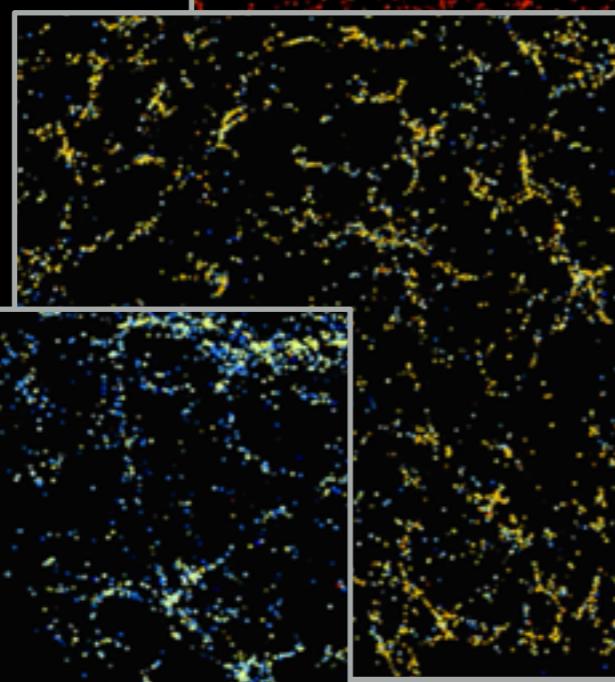
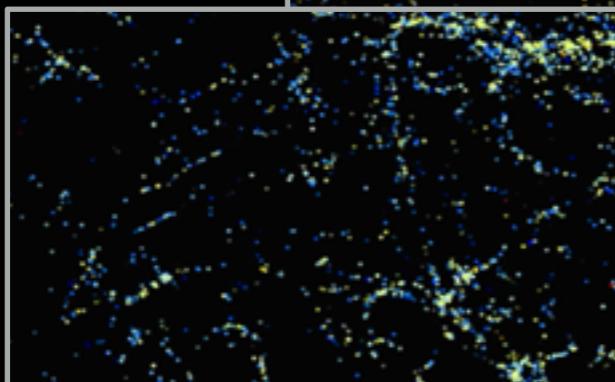
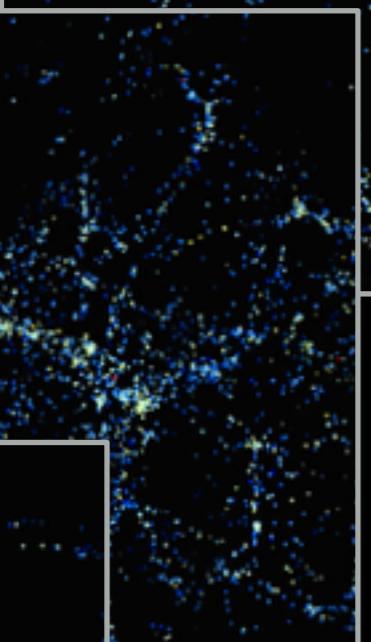
Ménard et al. (2013) - pilot study
Schmidt et al. (2014) - Planck
Rahman et al. (2014) - SDSS spectro
Rahman et al. (2015) - 2MASS
Rahman et al. (2016) - SDSS photo
Mendez et al. (in prep) - WISE



[data analysis, dN/dz]

Aragon-Calvo et al. (2014) “photoweb” -----

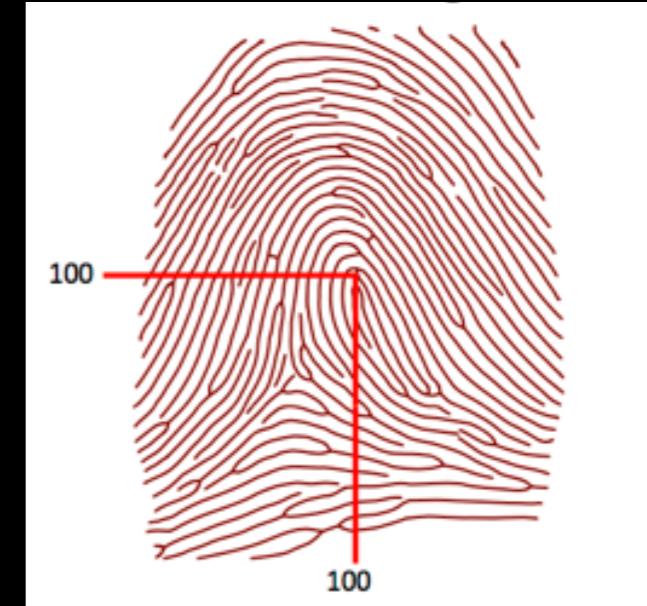
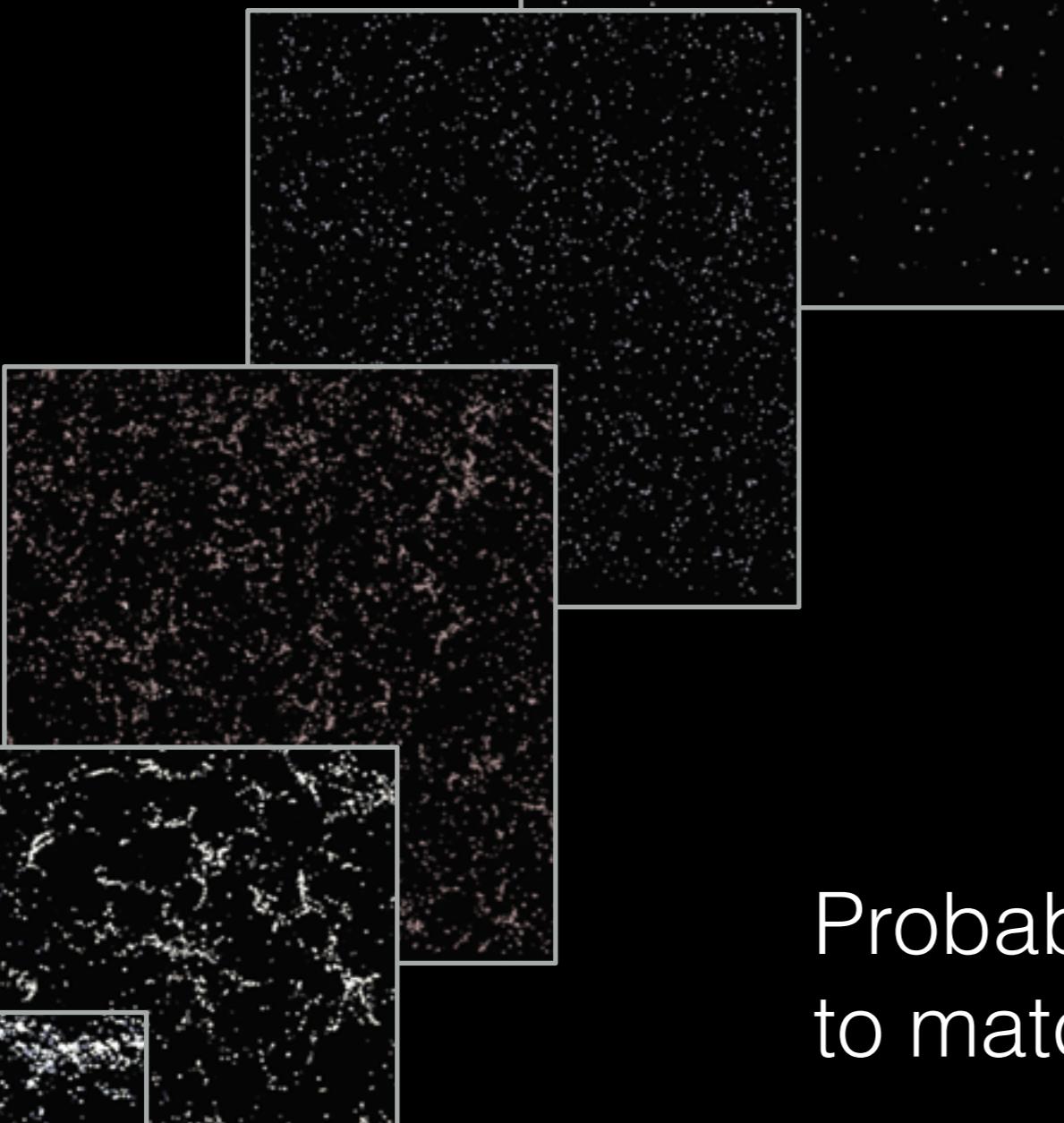
[data analysis,
 dN/dz at $z \sim 0$]



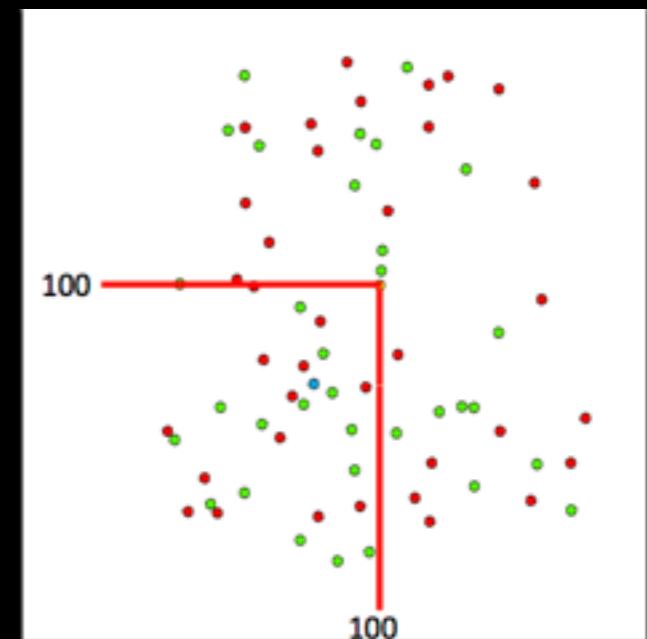
?

$\langle \partial_i \cdot \partial_{\text{unknown}} \rangle$

∂_i



fingerprint minutiae



Probability for two different fingerprints
to match $\sim 1/68$ billion

Galton (1892)

The clustering redshift technique

expected data:

one “unknown” sample $\{N_u, dN_u/dz, b_u(z)\}$
one reference sample $\{N_r, dN_r/dz, b_r(z)\}$

1. Ideal case:

$$\frac{dN_u}{dz} = N_u \delta_D(z - z_0)$$

locate in redshift:

$$dN_u/dz \propto w_{ur}(z_i)$$

normalize:

$$\int dz \frac{dN_u}{dz} = N_u$$

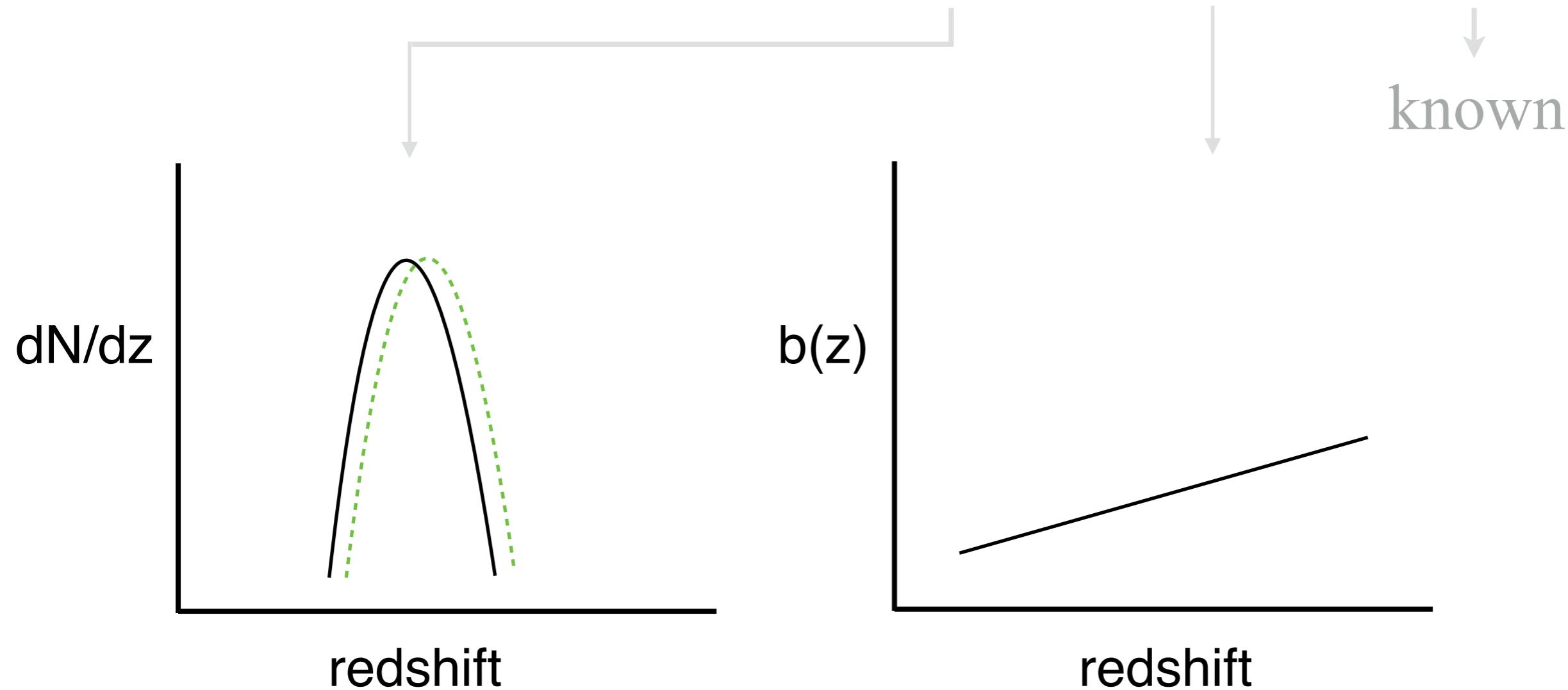
Note: clustering can be measured on any scale

The clustering redshift technique

2. Departure from the ideal case:

$$\frac{dN_u}{dz} = N_u G(z_0, \sigma_z)$$

The observable $\bar{w}_{ur}(z_i) \propto \frac{dN_u}{dz}(z_i) \bar{b}_u(z_i) \bar{b}_r(z_i)$



The clustering redshift technique

2. Departure from the ideal case:

$$\frac{dN_u}{dz} = N_u G(z_0, \sigma_z)$$

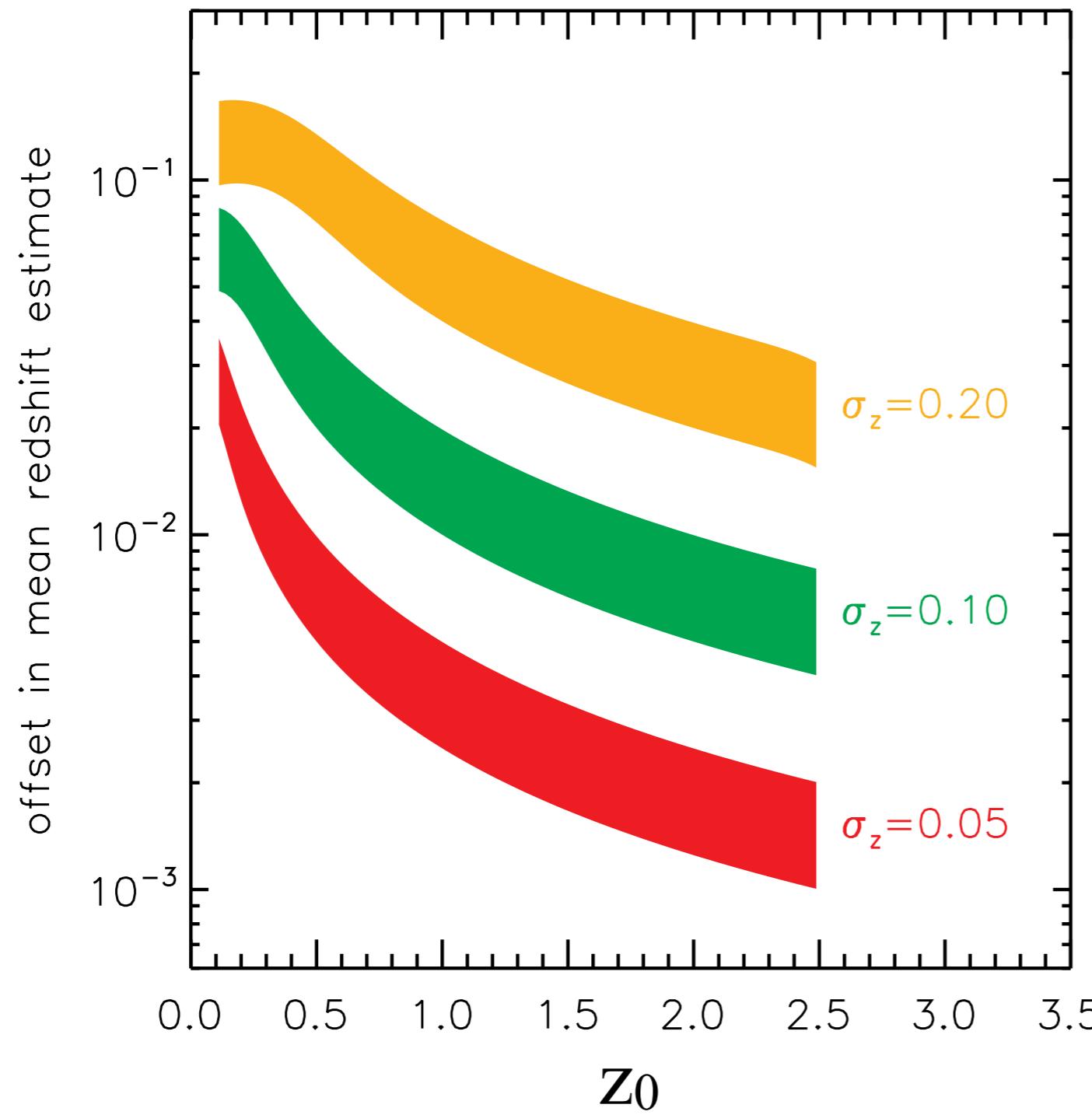
The observable $\bar{w}_{ur}(z_i) \propto \frac{dN_u}{dz}(z_i) \bar{b}_u(z_i) \bar{b}_r(z_i)$

But if the relative change in $b(z)$ is much smaller than in dN/dz

$$\frac{d \log dN_u/dz}{dz} \gg \frac{d \log \bar{b}_u}{dz} \quad \text{then we are in a quasi ideal case}$$

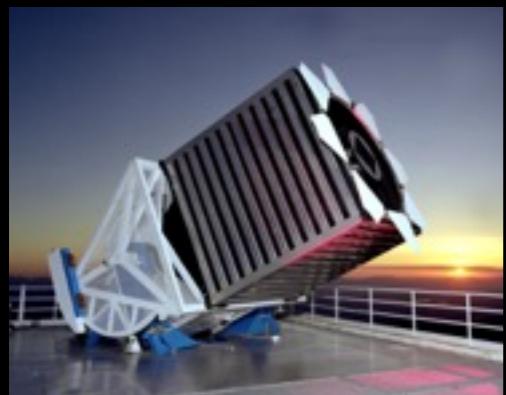
What error are we making if $b_u(z)$ changes?

example: $\bar{b}_u(z) \propto z^\alpha$ but we assume $b_u(z) = \text{constant}$



$$\frac{dN_u}{dz} = N_u G(z_0, \sigma_z)$$

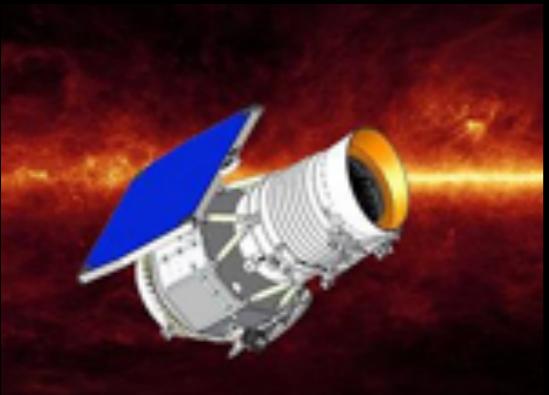
Applications of clustering redshifts



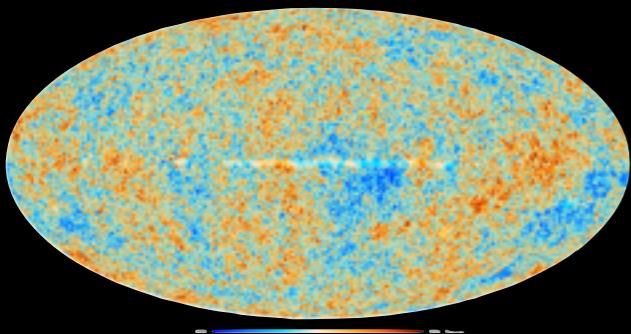
SDSS
optical



2MASS
near infrared

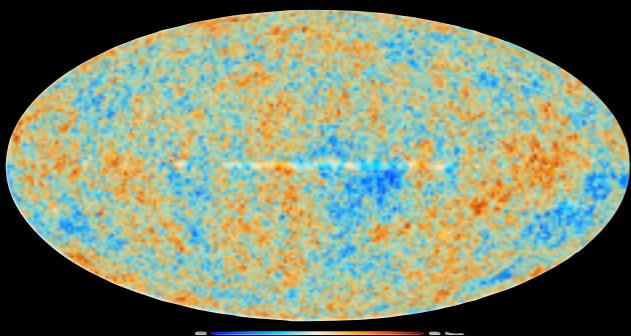
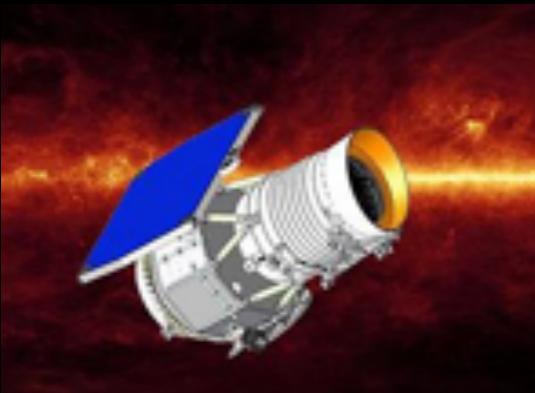
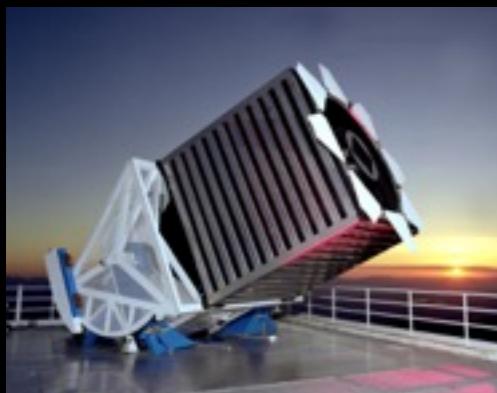


WISE
infrared



Planck
millimetric

Applications of clustering redshifts

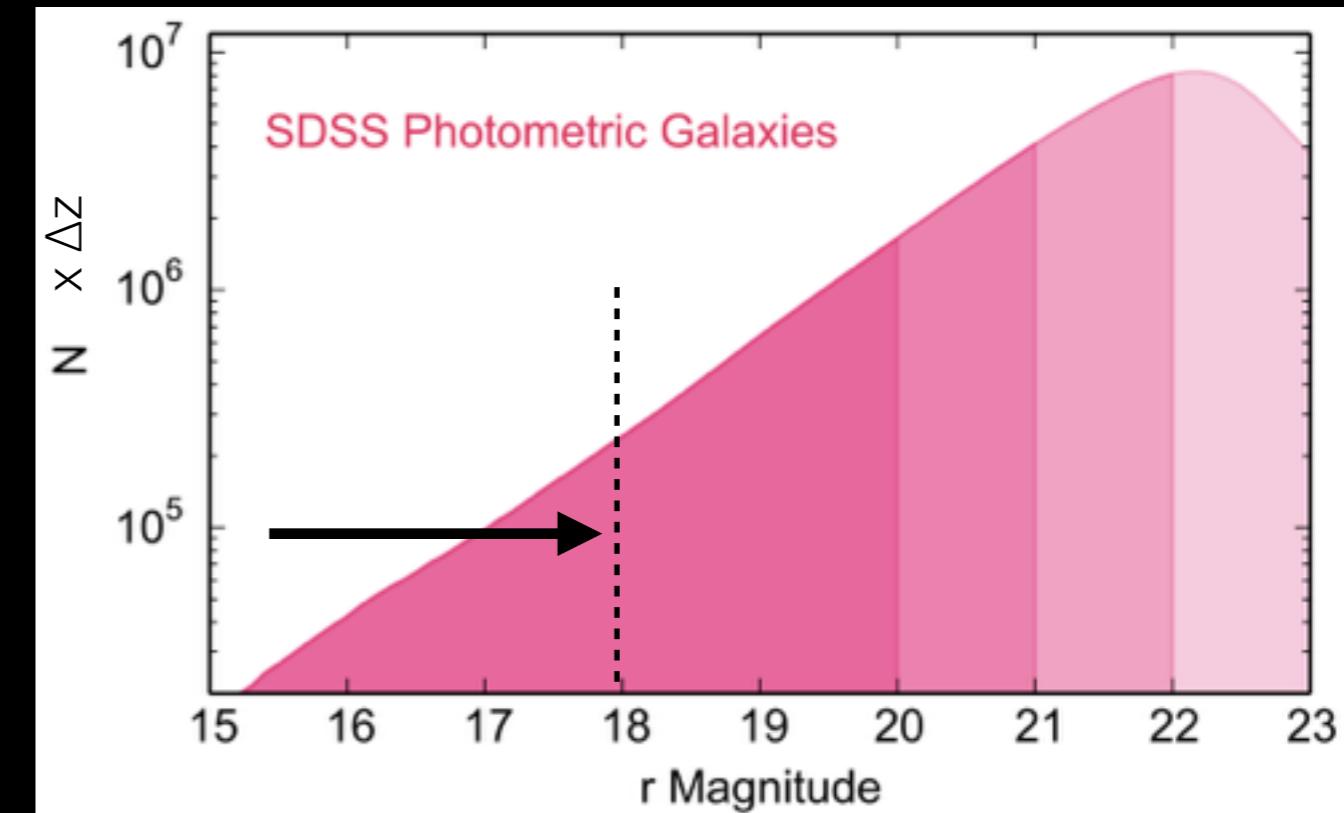


SDSS
optical

2MASS
near infrared

WISE
infrared

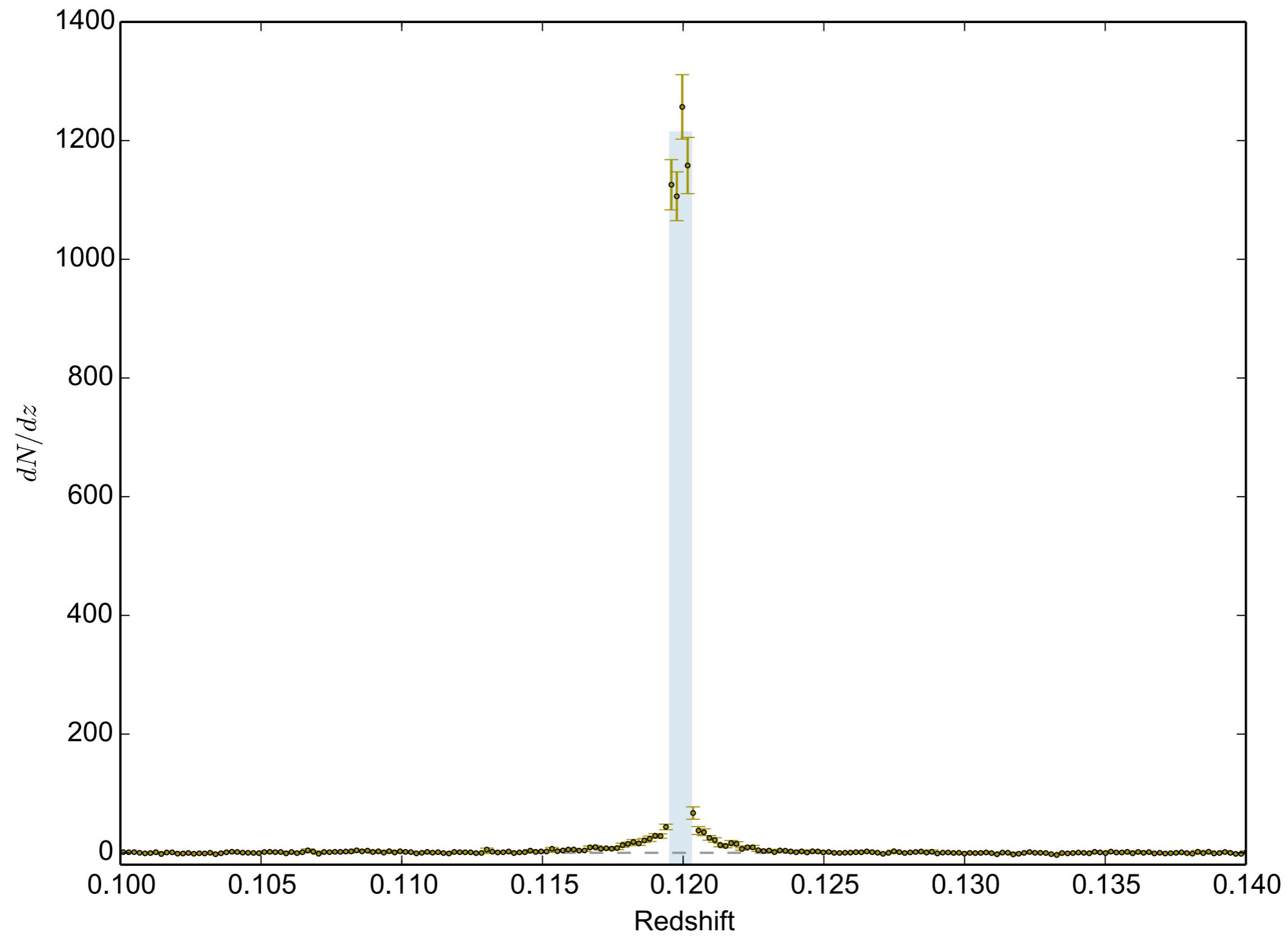
Planck
millimetric

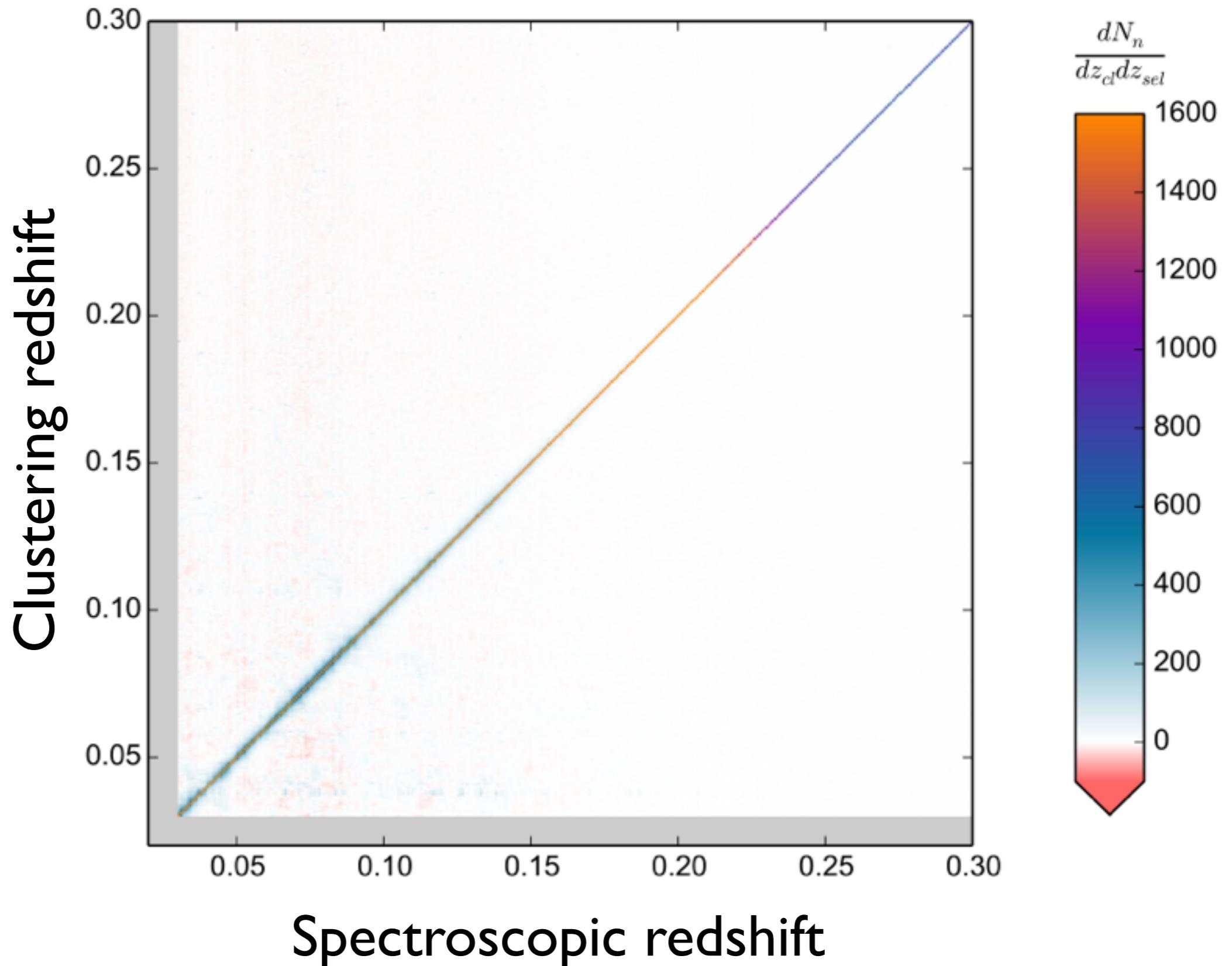


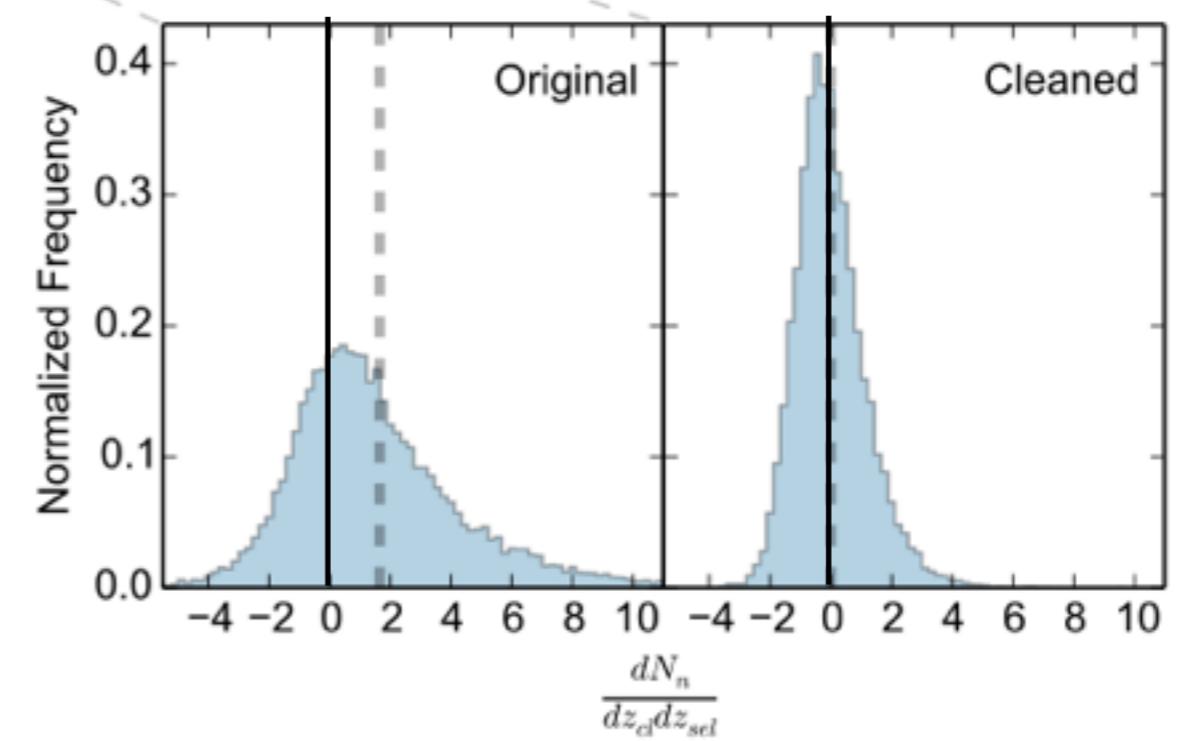
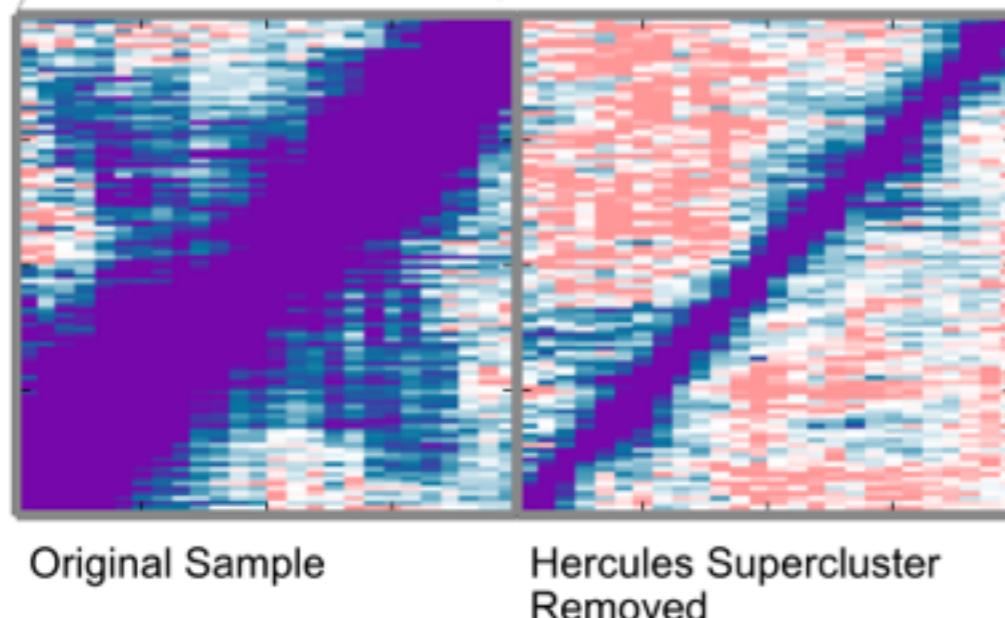
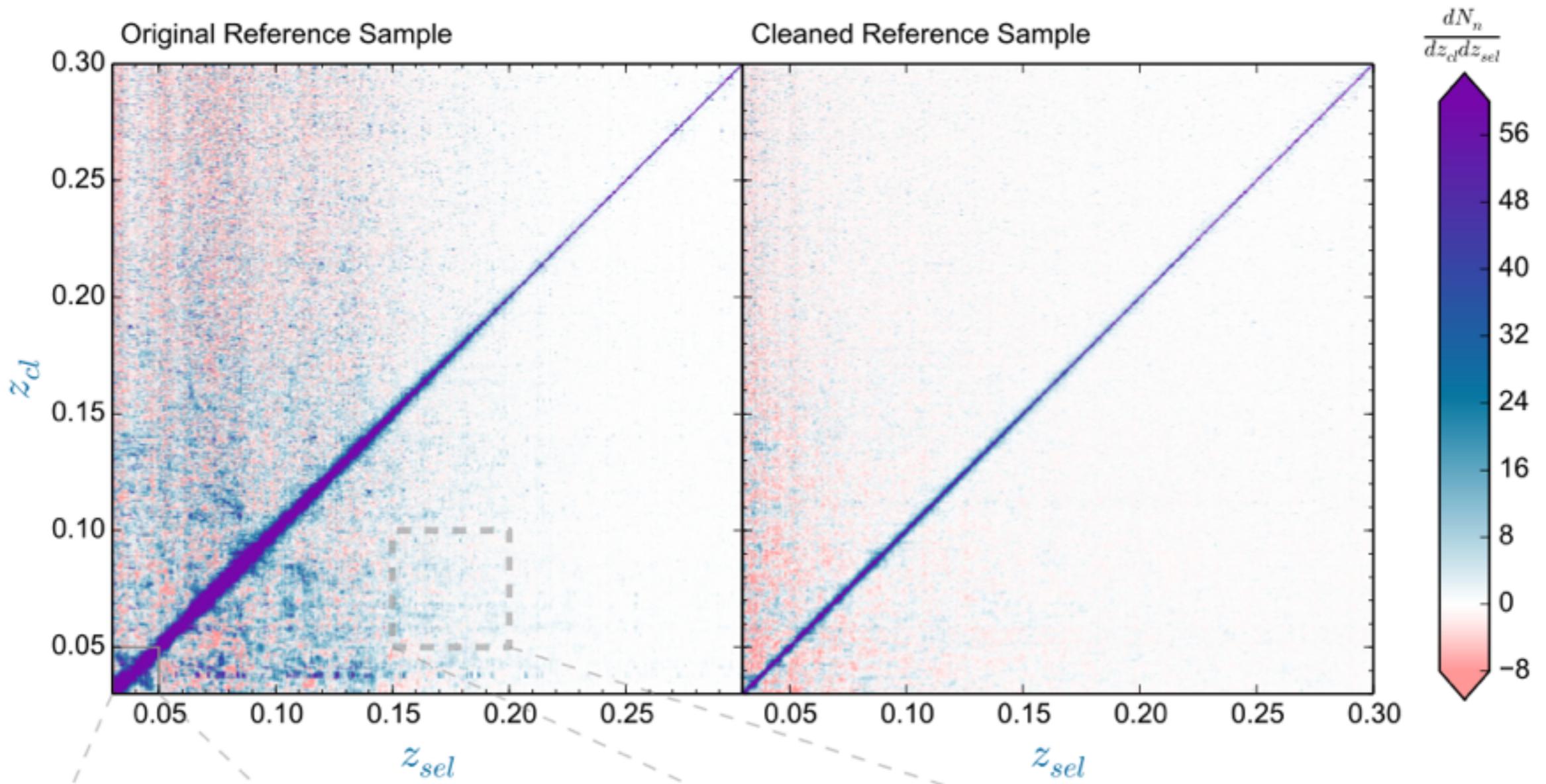
spectroscopic galaxy sample

$r < 18$ mag

1 million objects







Photometrically-selected galaxies

sample 1: $0.5 < g-r < 0.6$

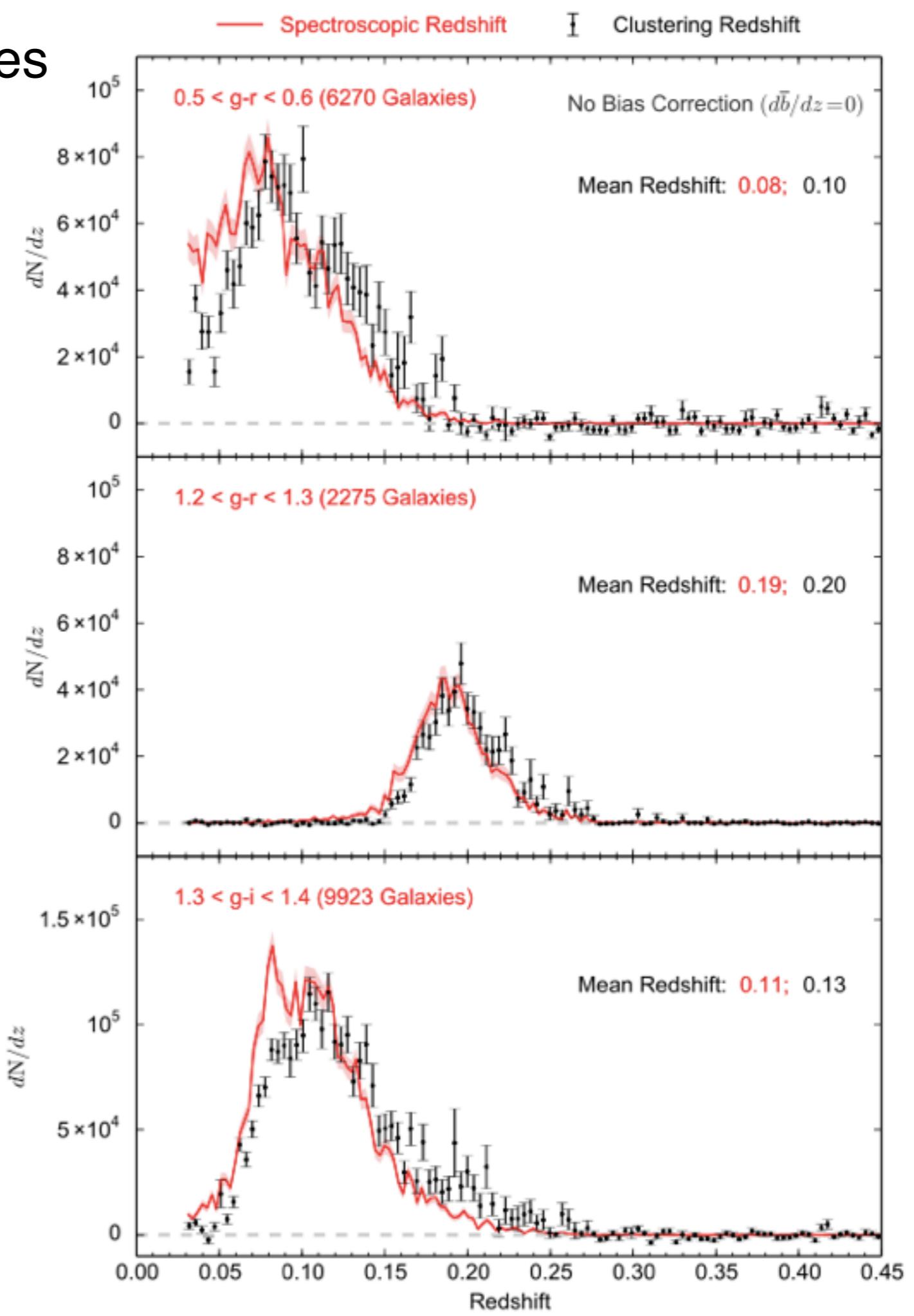
~ 6,300 galaxies

sample 2: $1.3 < g-i < 1.4$

~ 10,000 galaxies

sample 3: $1.2 < g-r < 1.3$

~ 2,500 galaxies



Photometrically-selected galaxies

sample 1: $0.5 < g-r < 0.6$

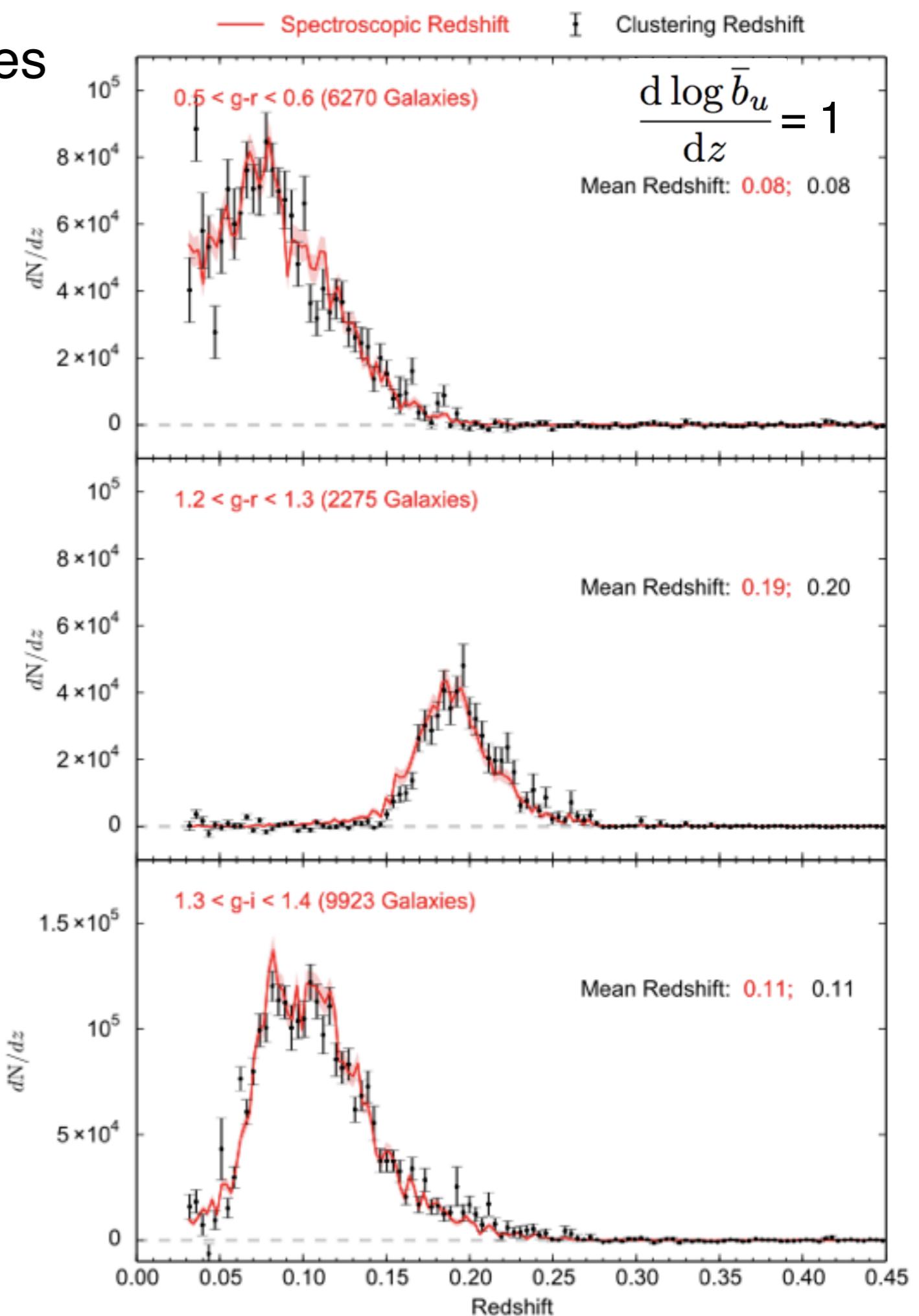
~ 6,300 galaxies

sample 2: $1.3 < g-i < 1.4$

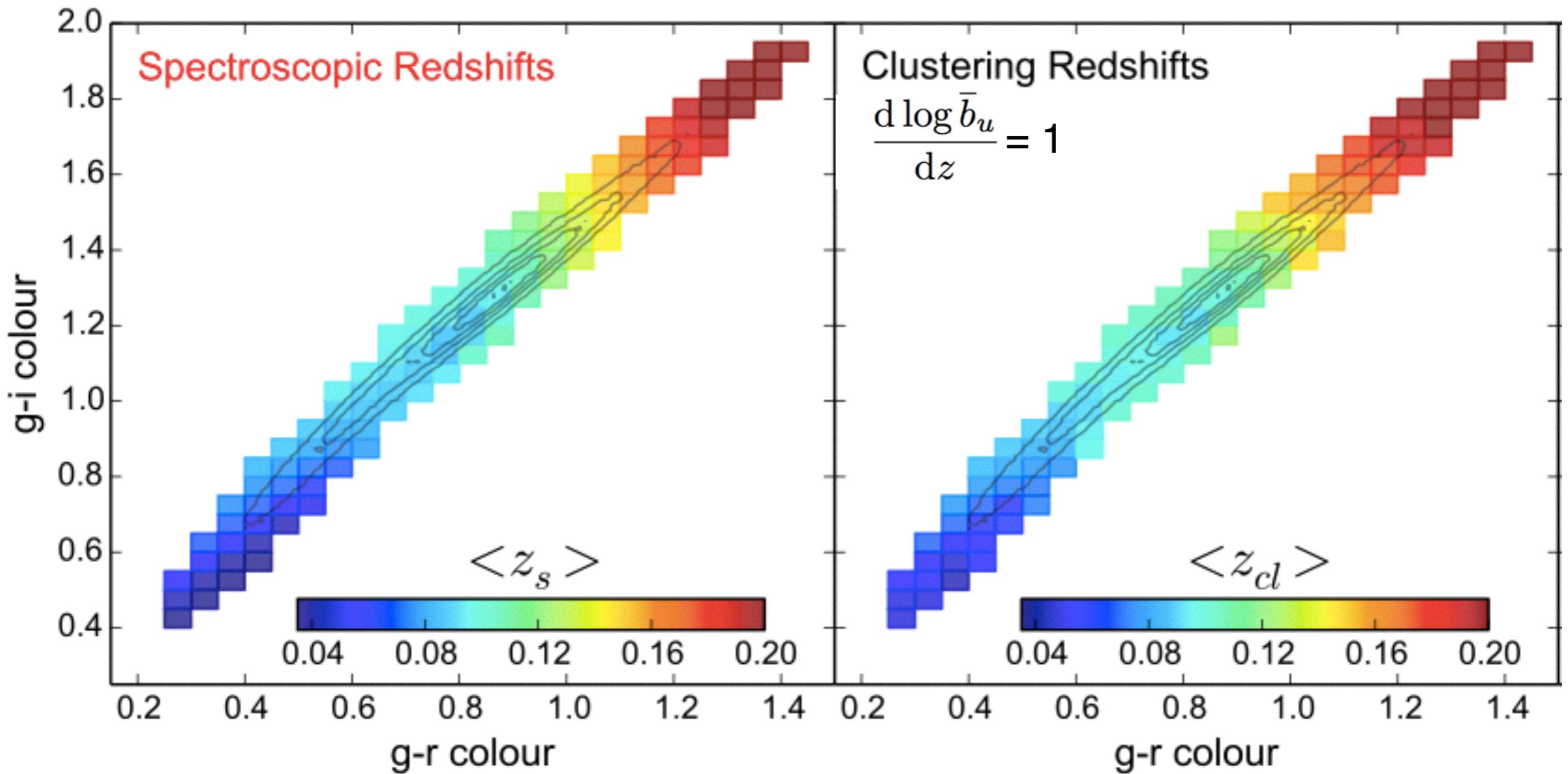
~ 10,000 galaxies

sample 3: $1.2 < g-r < 1.3$

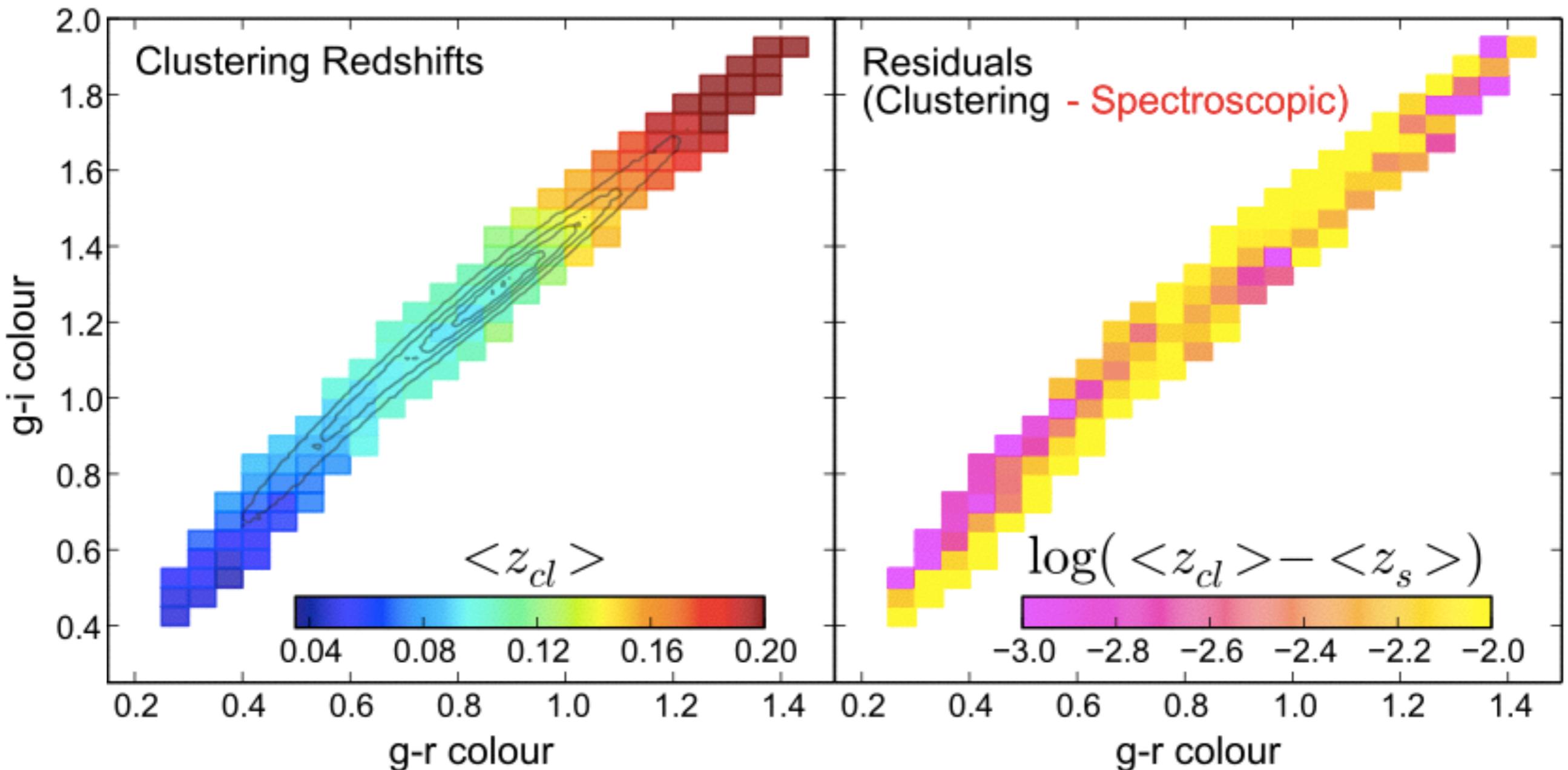
~ 2,500 galaxies



Generalization to one million galaxies



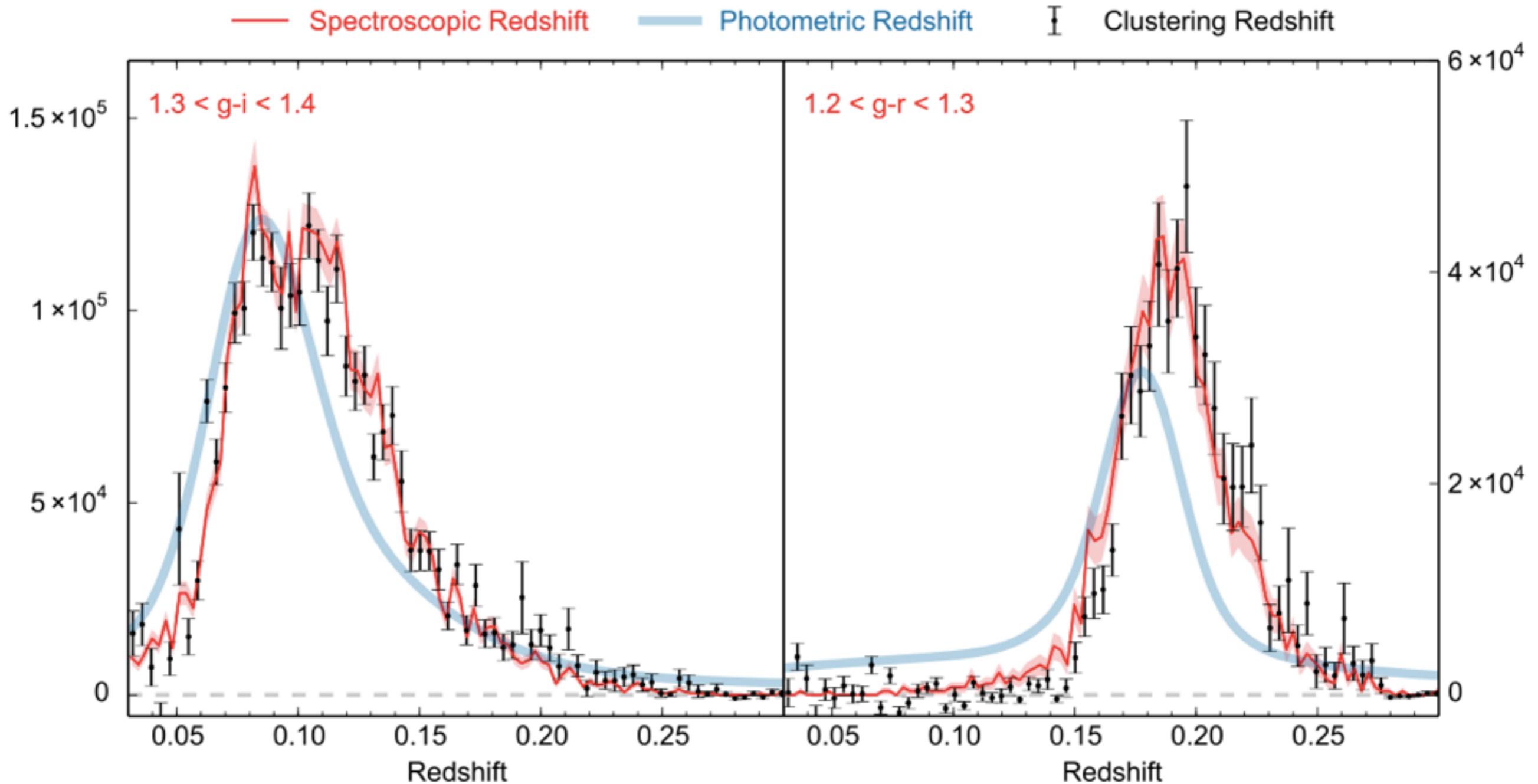
Generalization to one million galaxies



Comparison to photometric redshifts

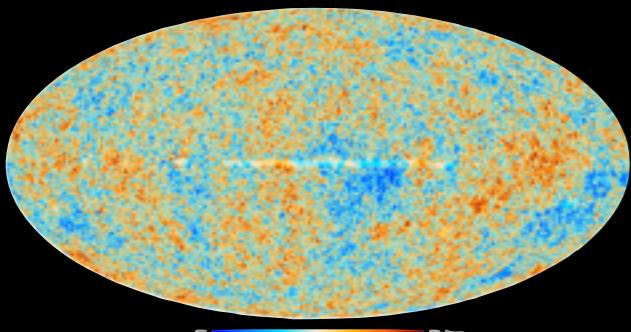
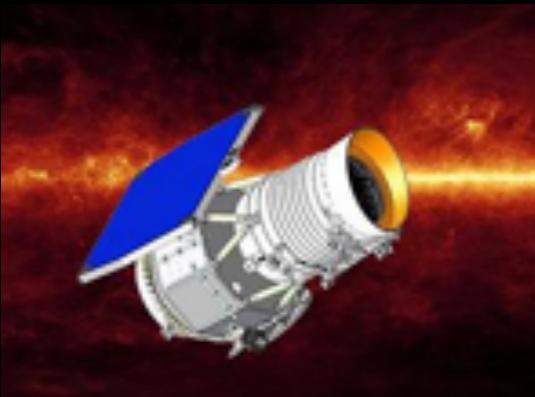
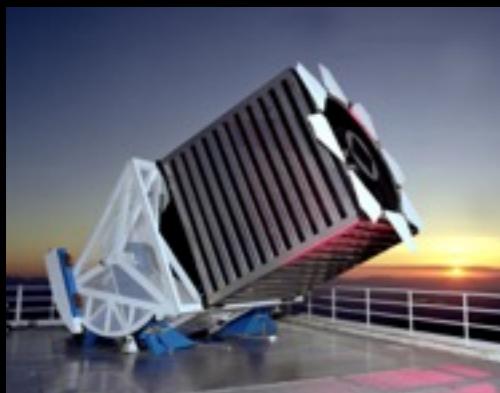
sample 2

sample 3



SDSS KD-tree photometric redshifts

Applications of clustering redshifts

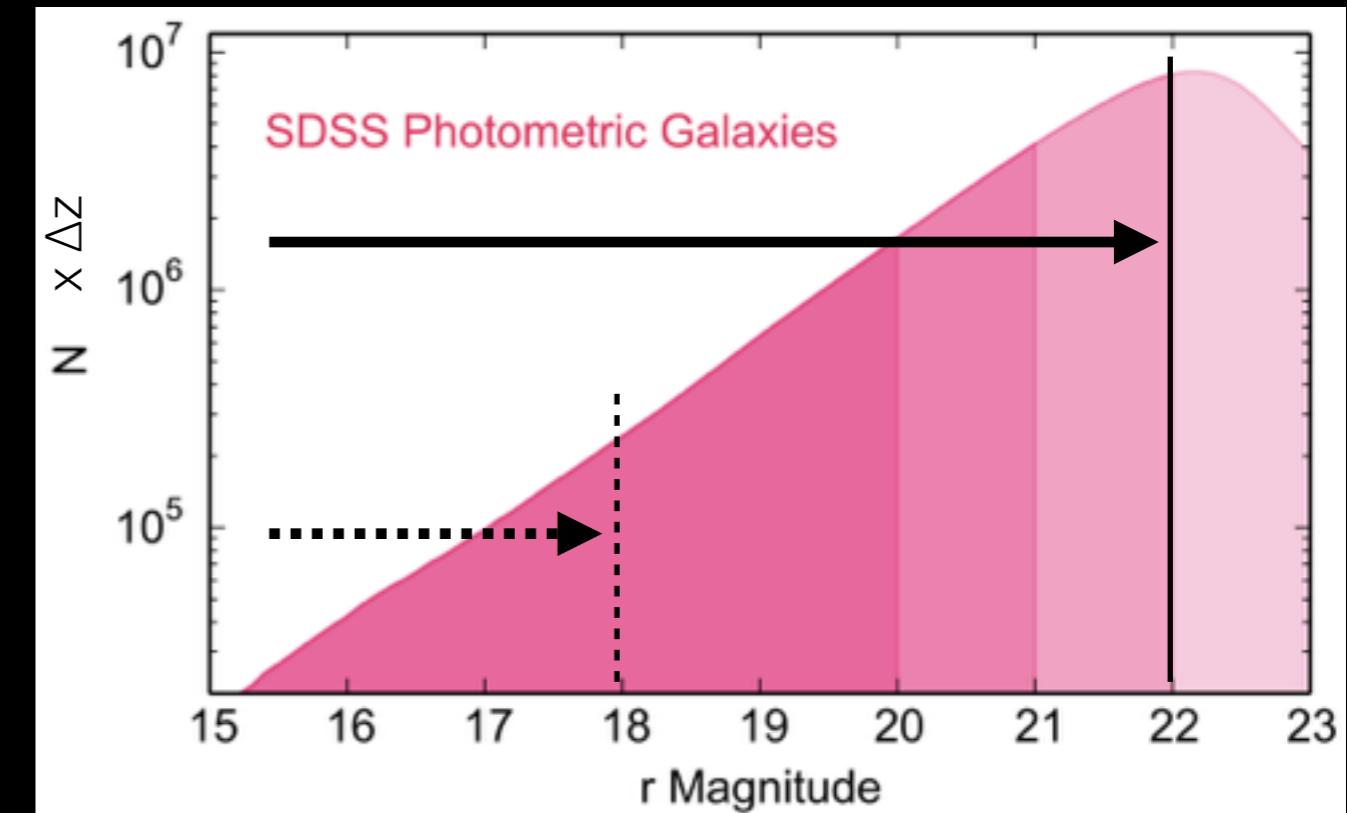


SDSS
optical

2MASS
near infrared

WISE
infrared

Planck
millimetric



Entire photometric sample

$r < 22$ mag

100 million objects

Photometrically-selected galaxies

sample 1: $0.53 < r-i < 0.54$

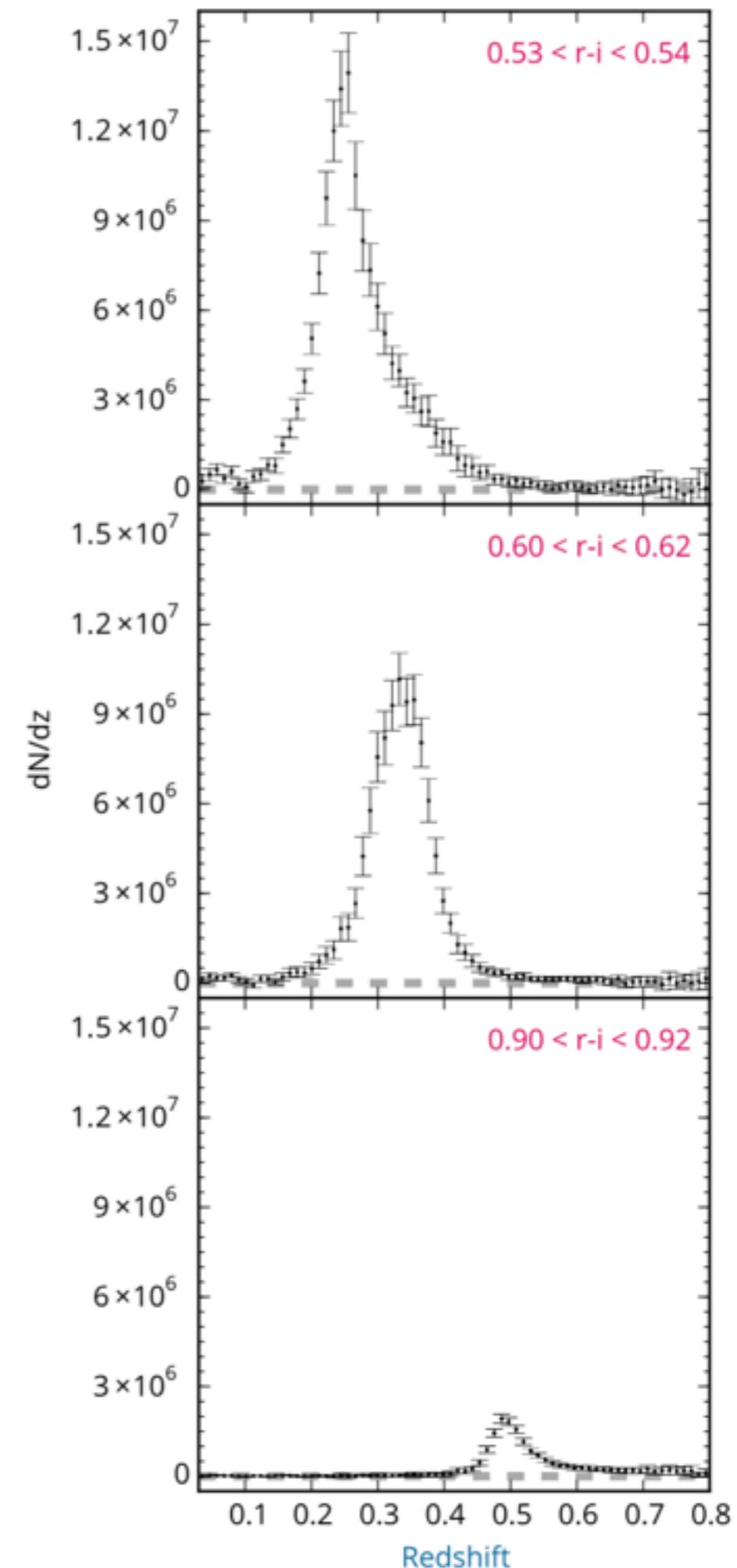
1.6 million galaxies

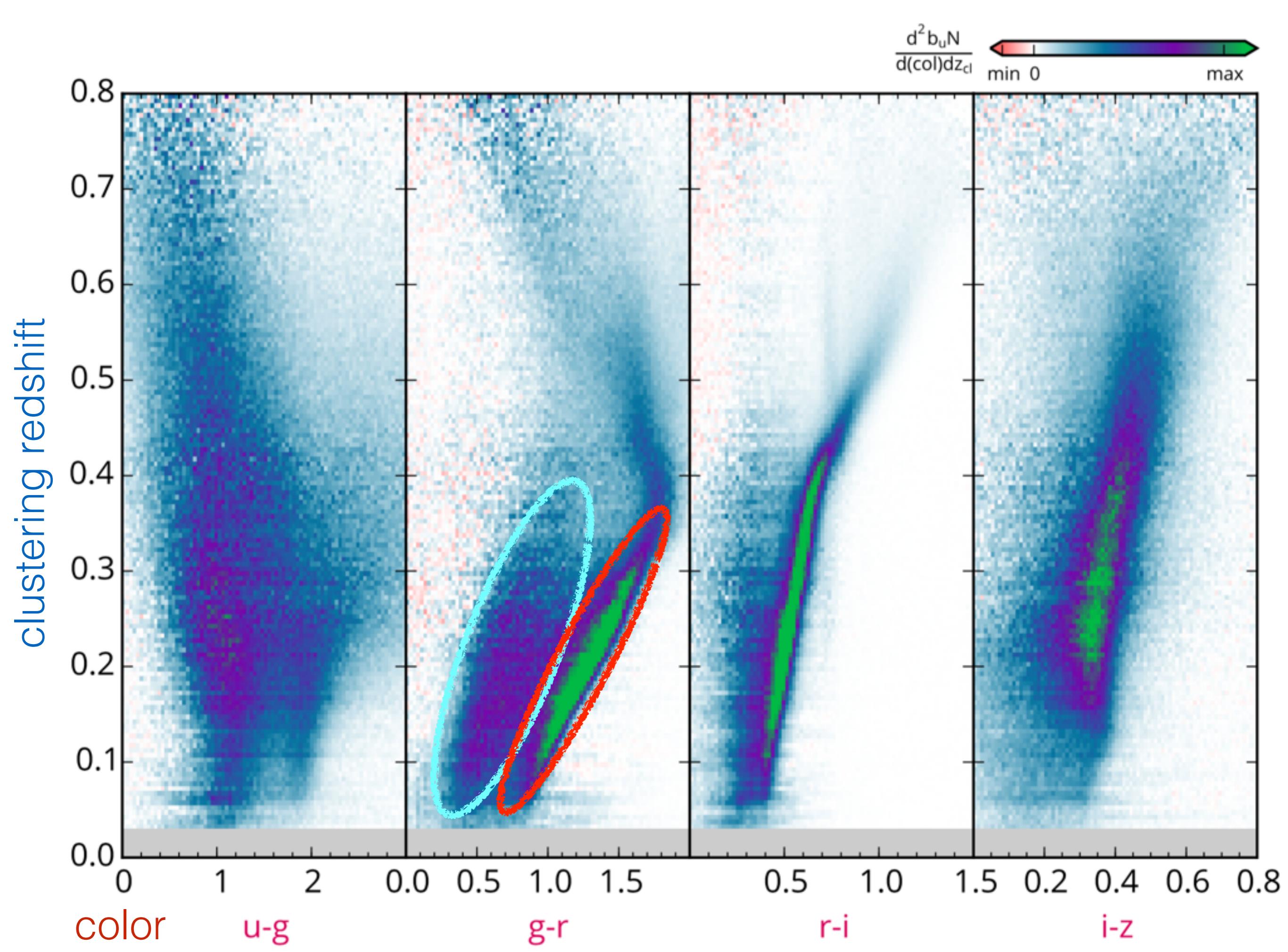
sample 3: $0.60 < r-i < 0.62$

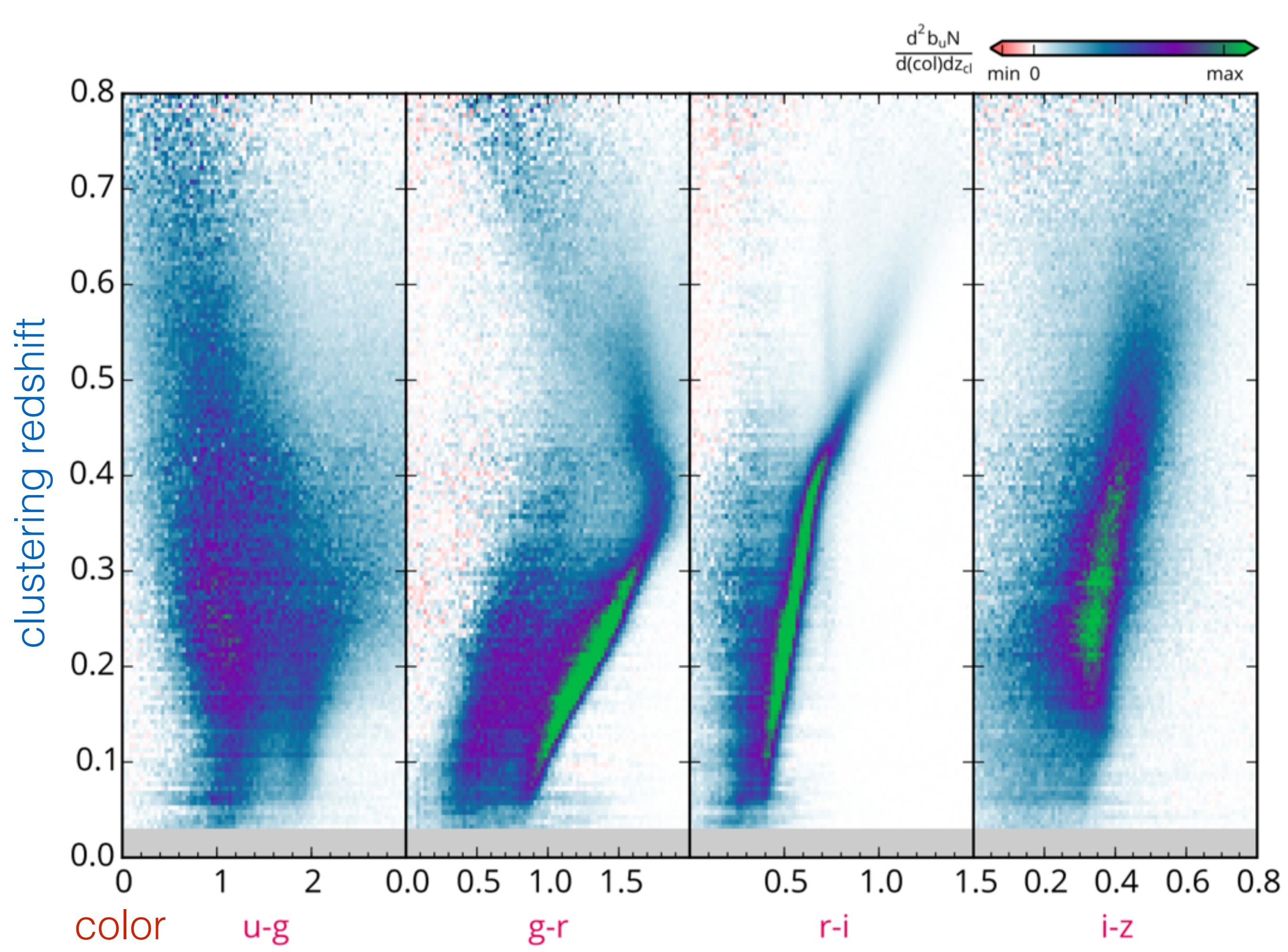
1.2 million galaxies

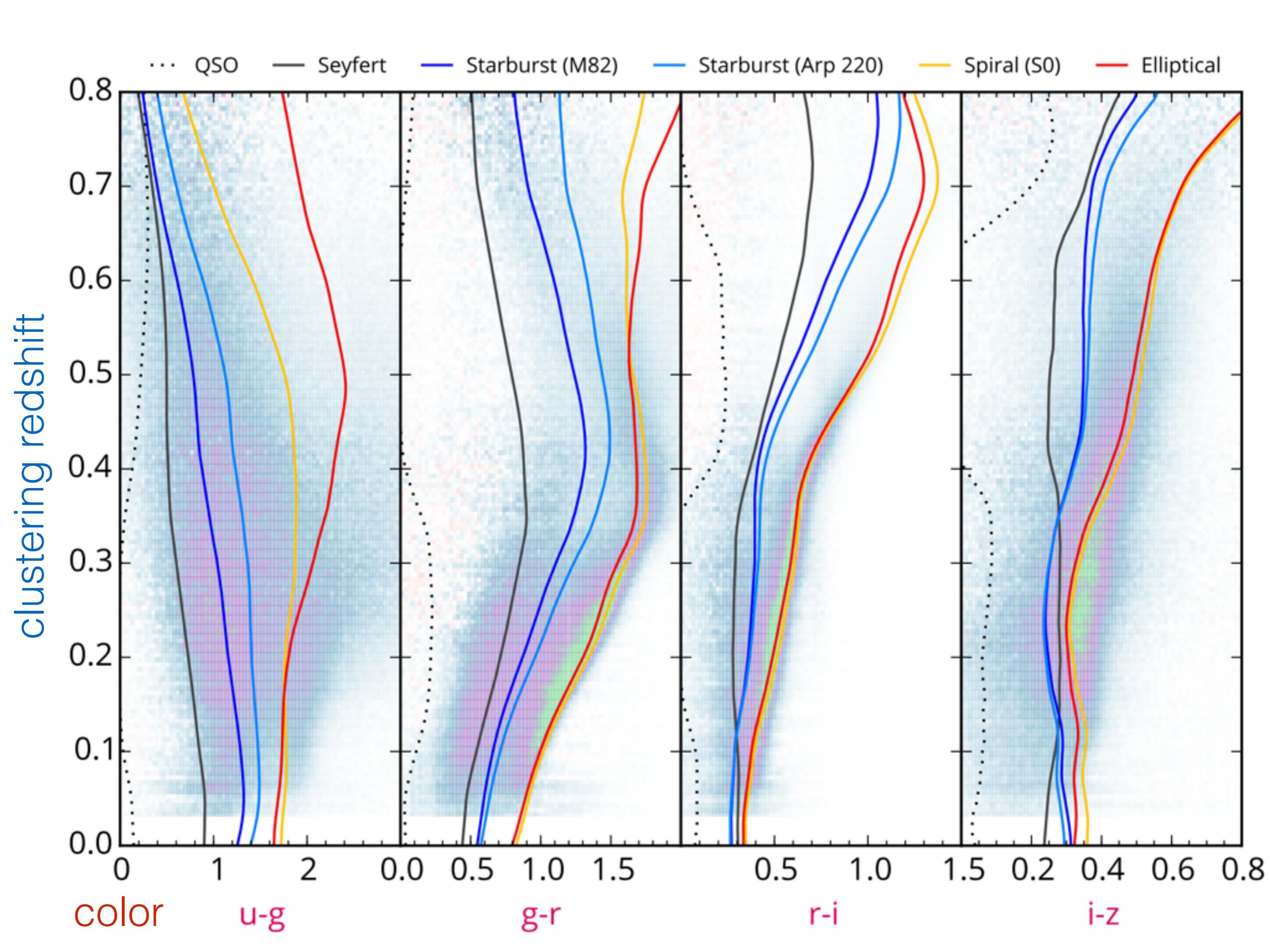
sample 3: $0.90 < r-i < 0.92$

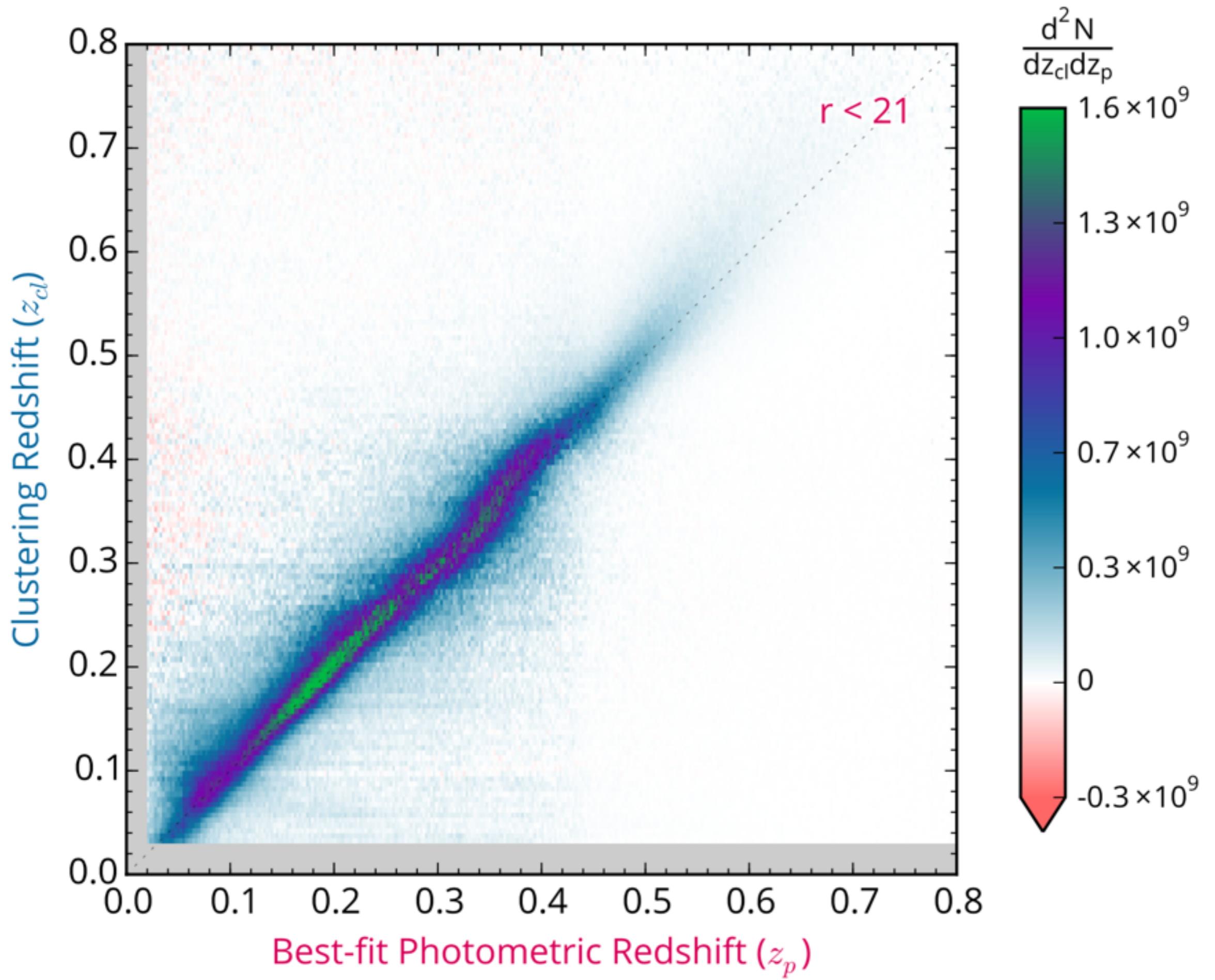
0.2 million galaxies



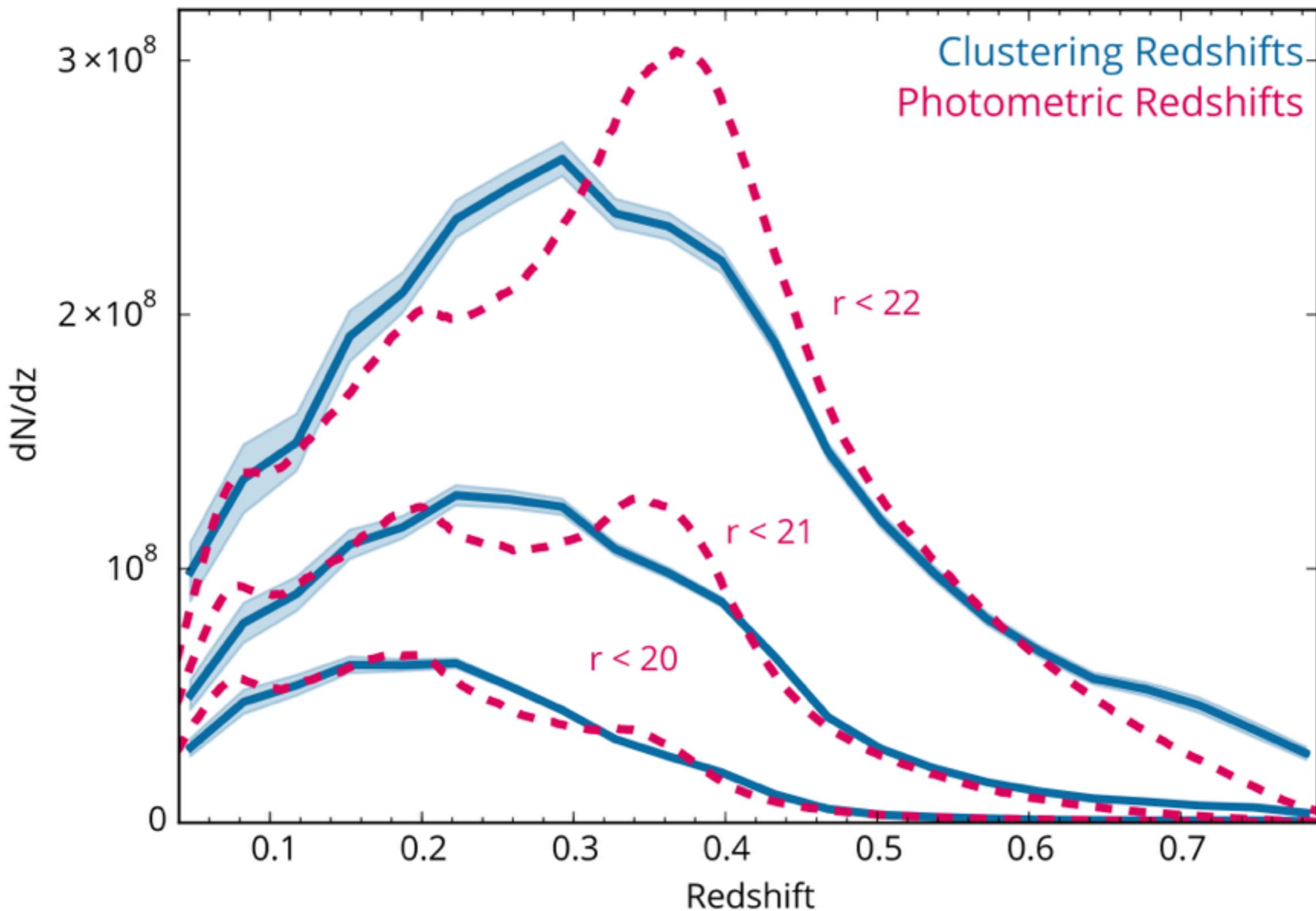




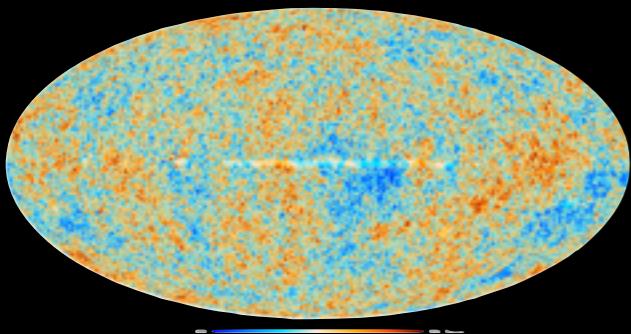
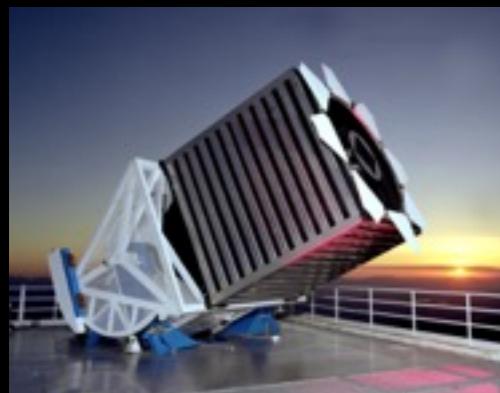




Redshift distribution of 100 million SDSS galaxies



Applications of clustering redshifts

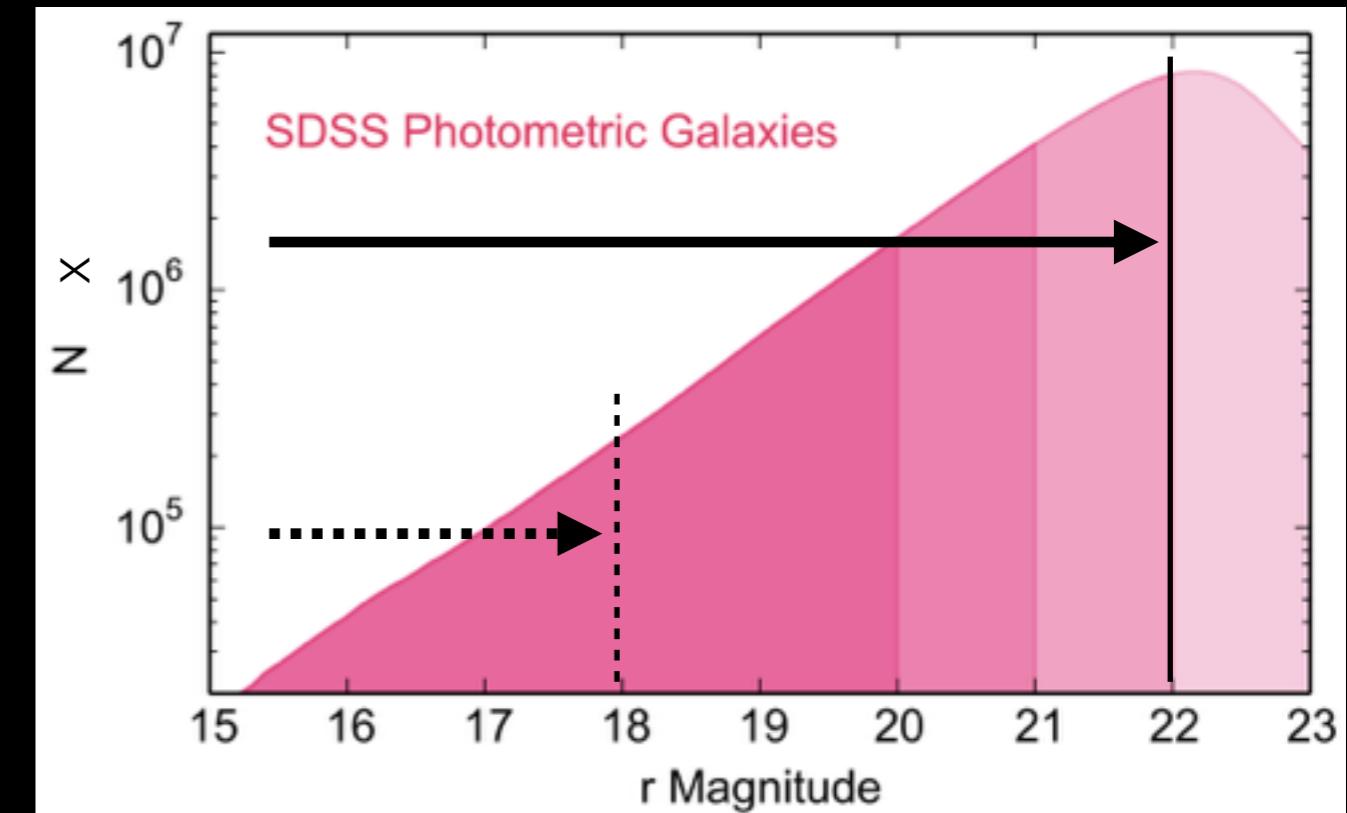


SDSS
optical

2MASS
near infrared

WISE
infrared

Planck
millimetric

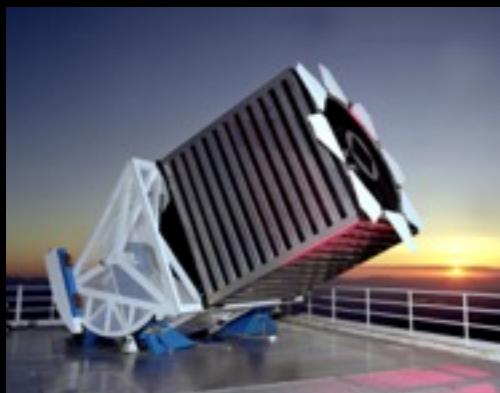


Entire photometric sample

$r < 22$ mag

100 million objects

Applications of clustering redshifts



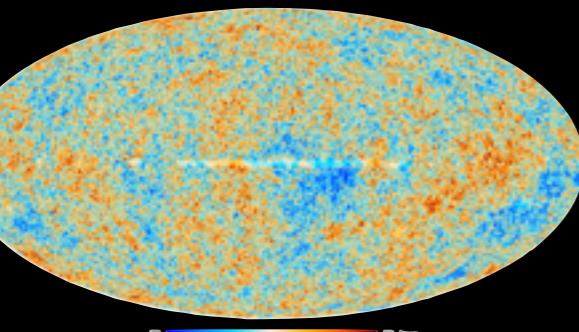
SDSS
optical



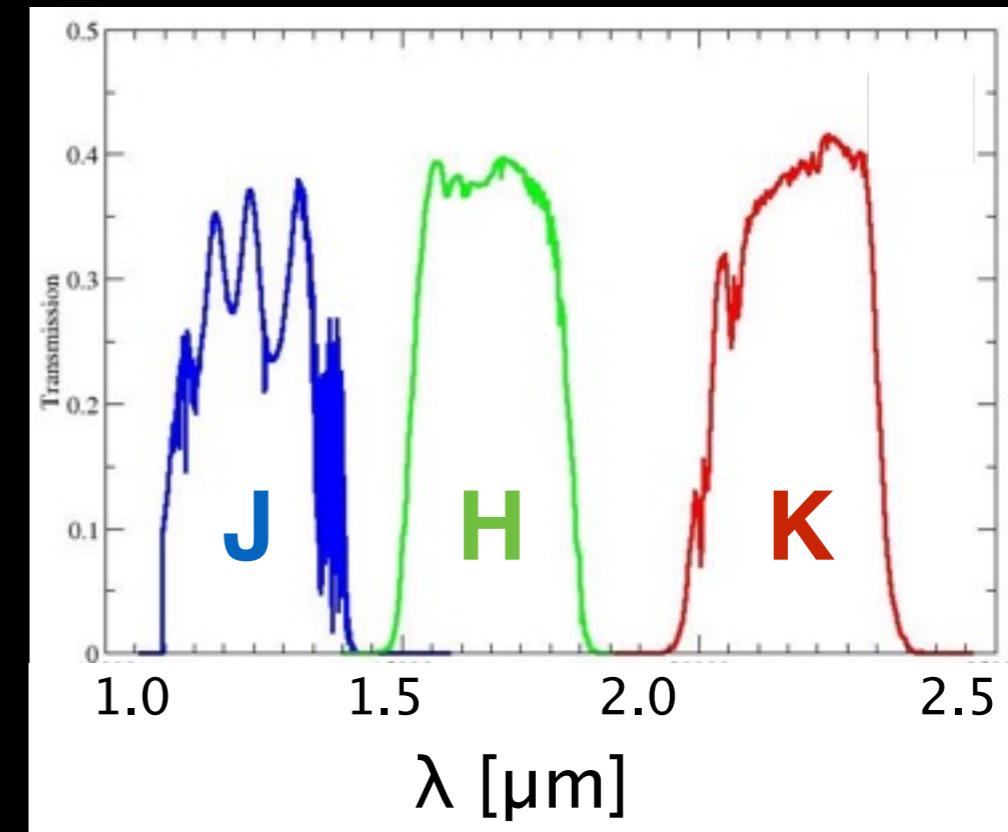
2MASS
near infrared



WISE
infrared

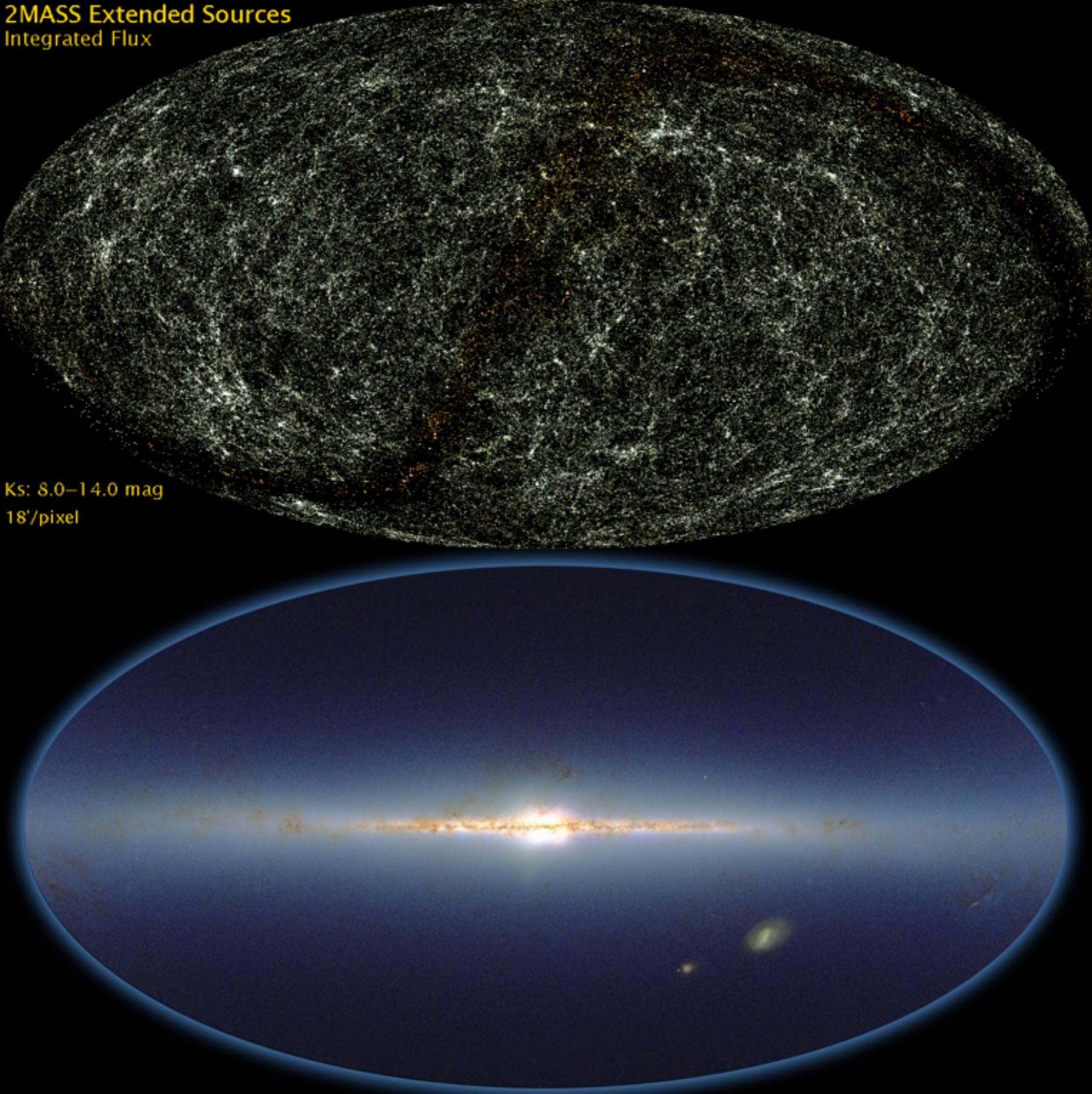


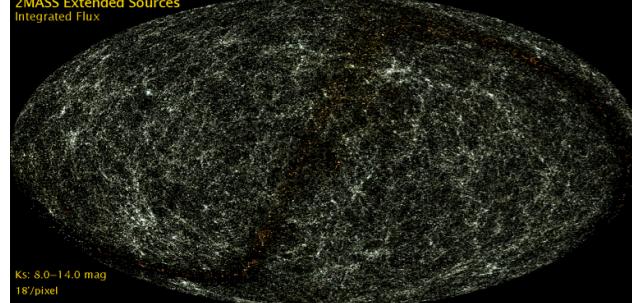
Planck
millimetric



2MASS $K < 14$ mag
1.5 million extended sources
470 million point sources

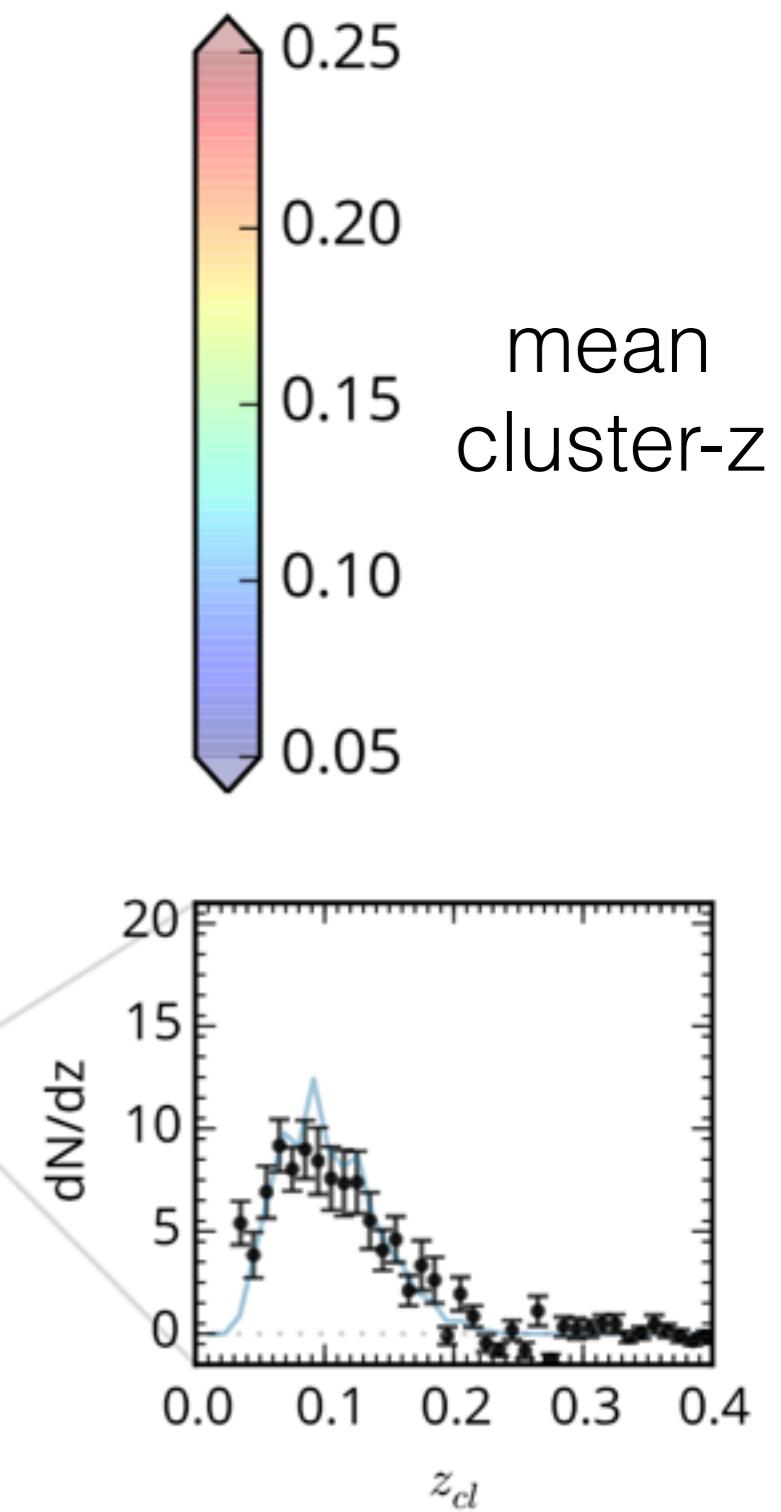
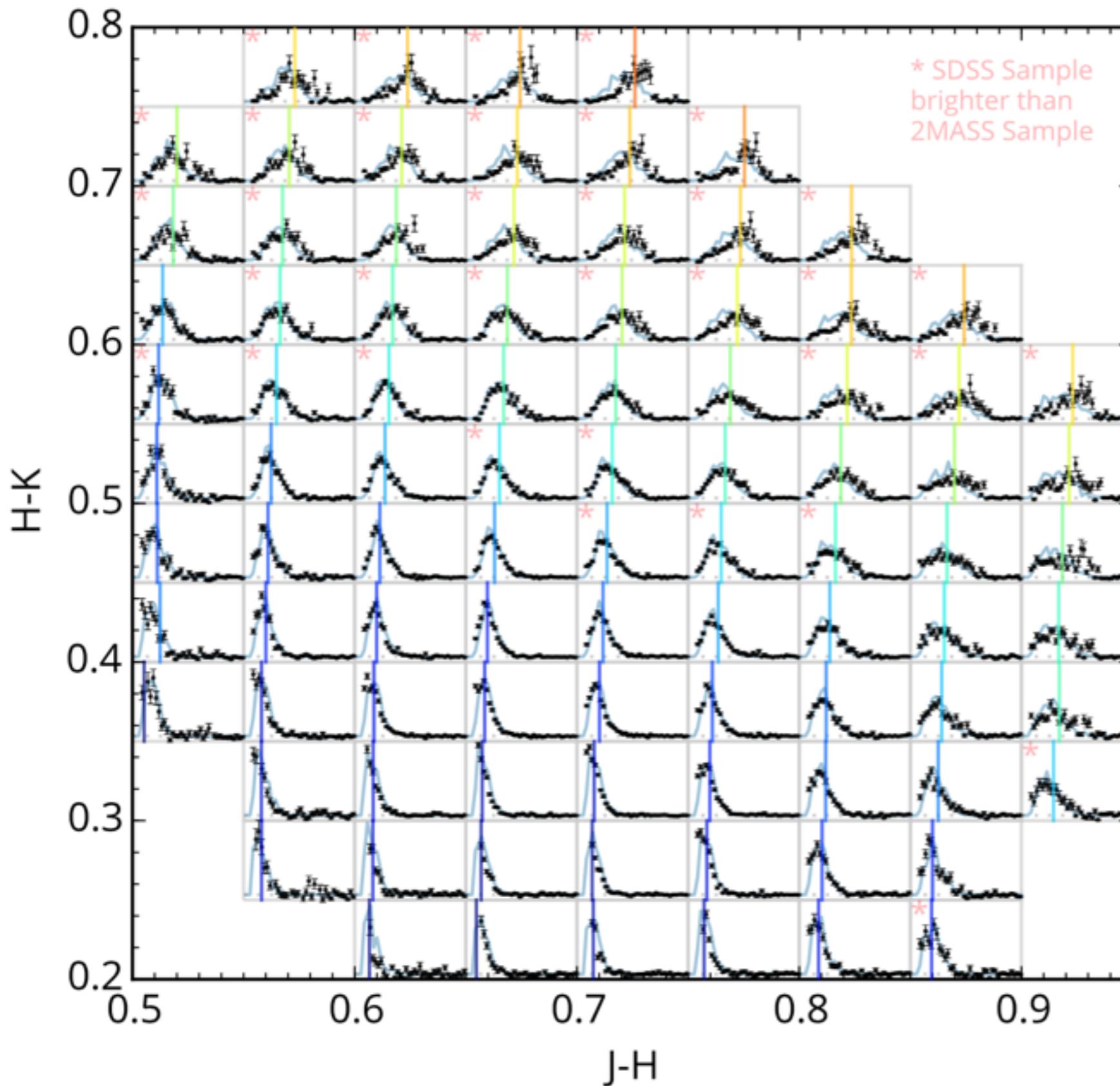
2MASS Extended Sources
Integrated Flux

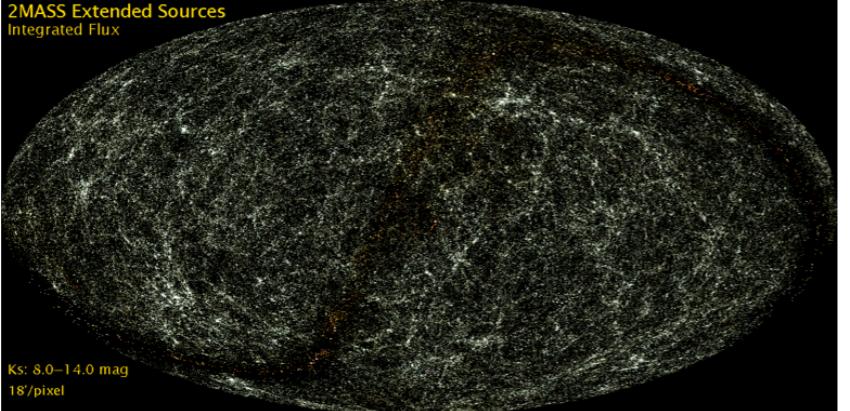




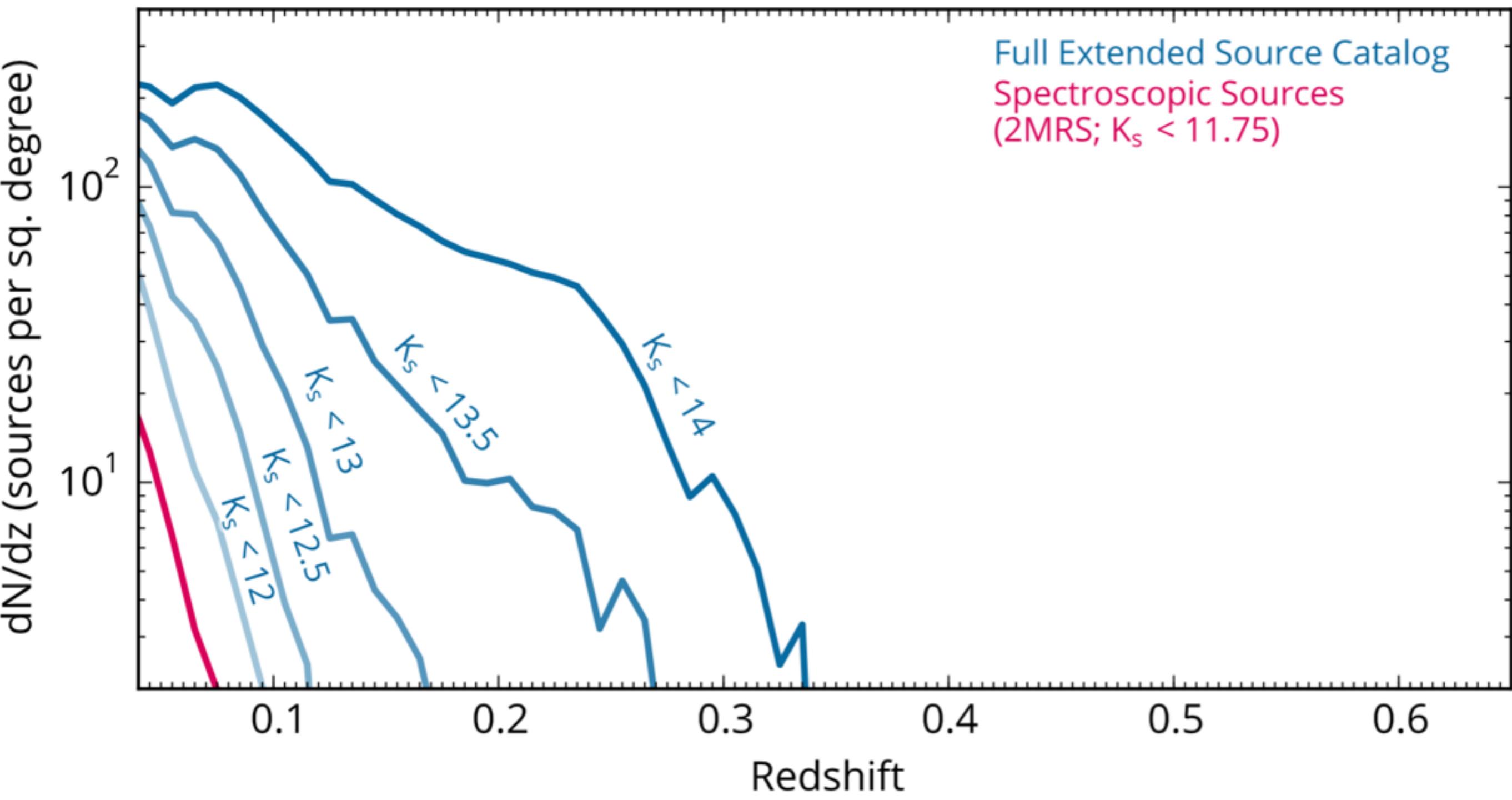
Skrutskie et al. (2006)

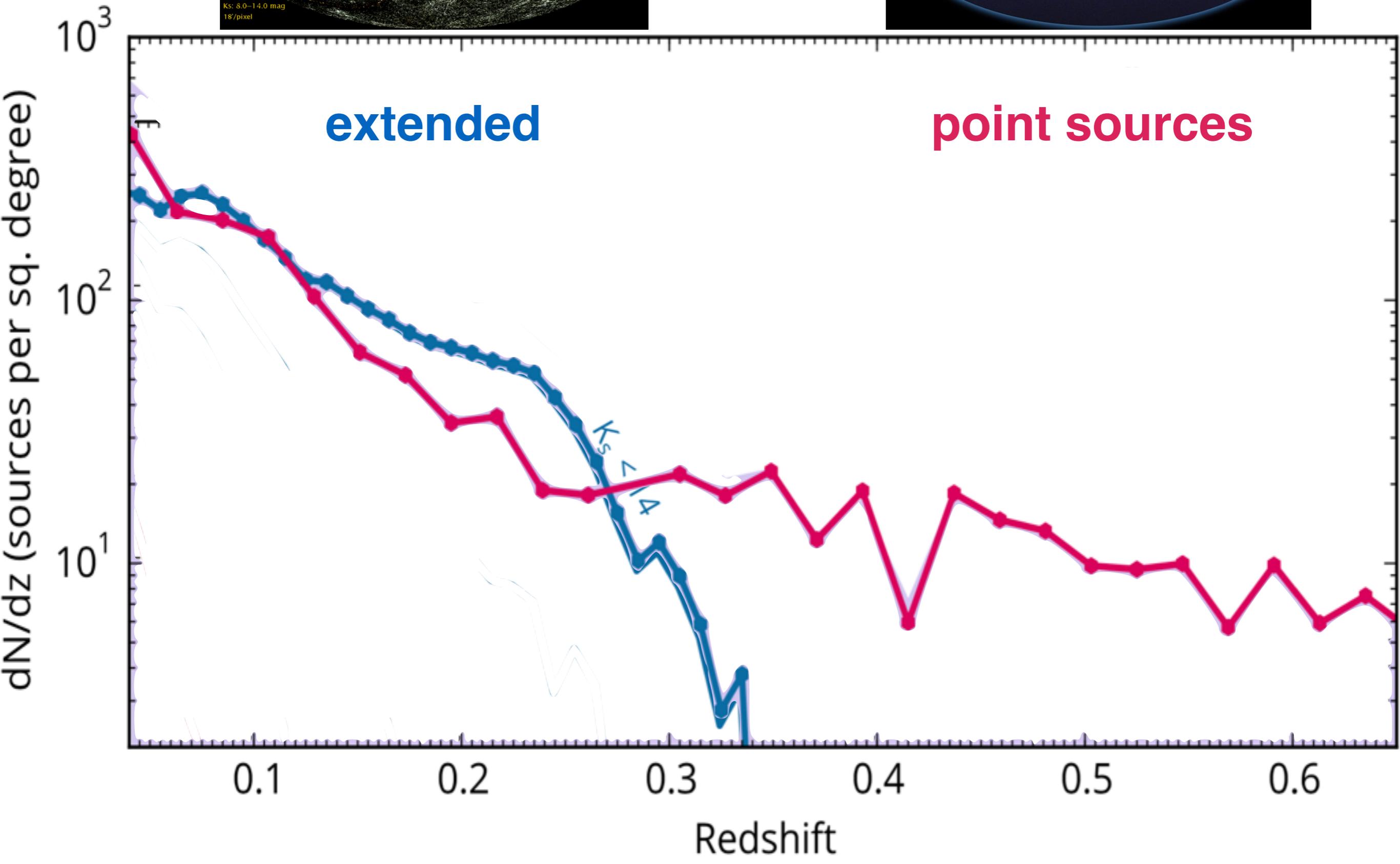
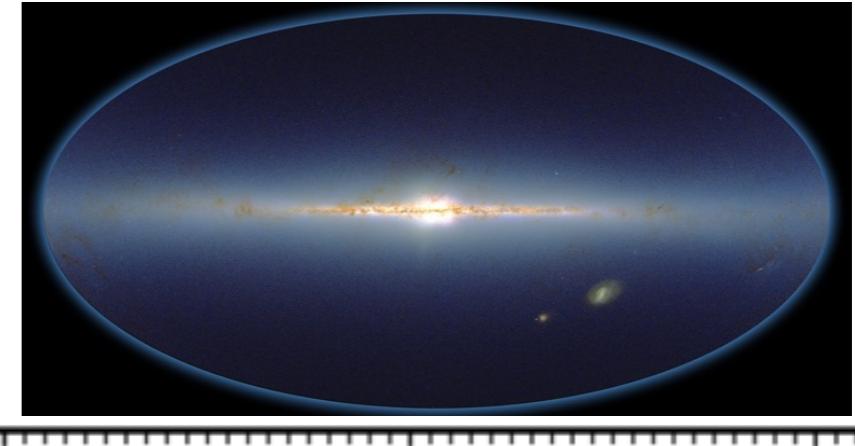
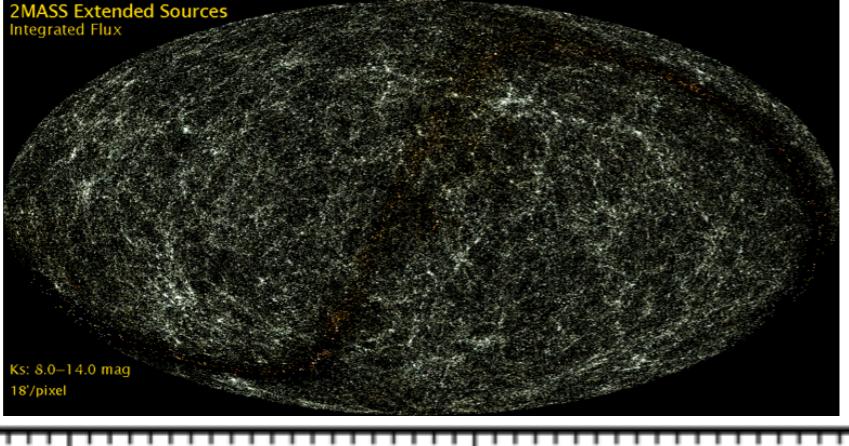
Observations: 1997-2001, J, H & K bands



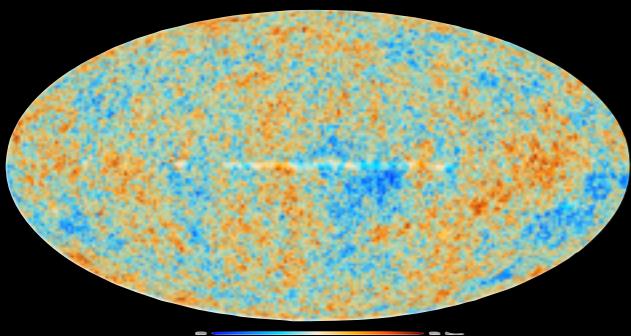
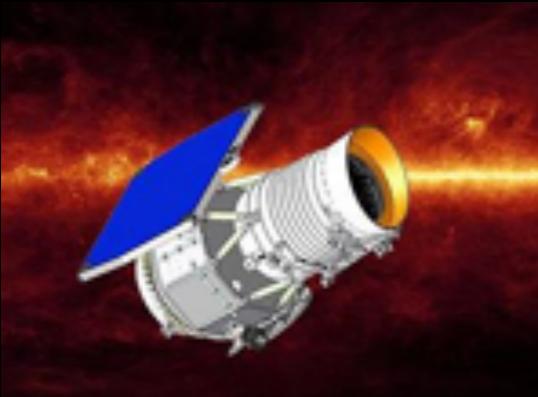
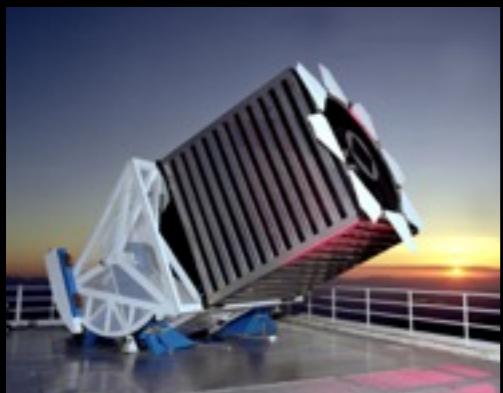


extended sources





Applications of clustering redshifts

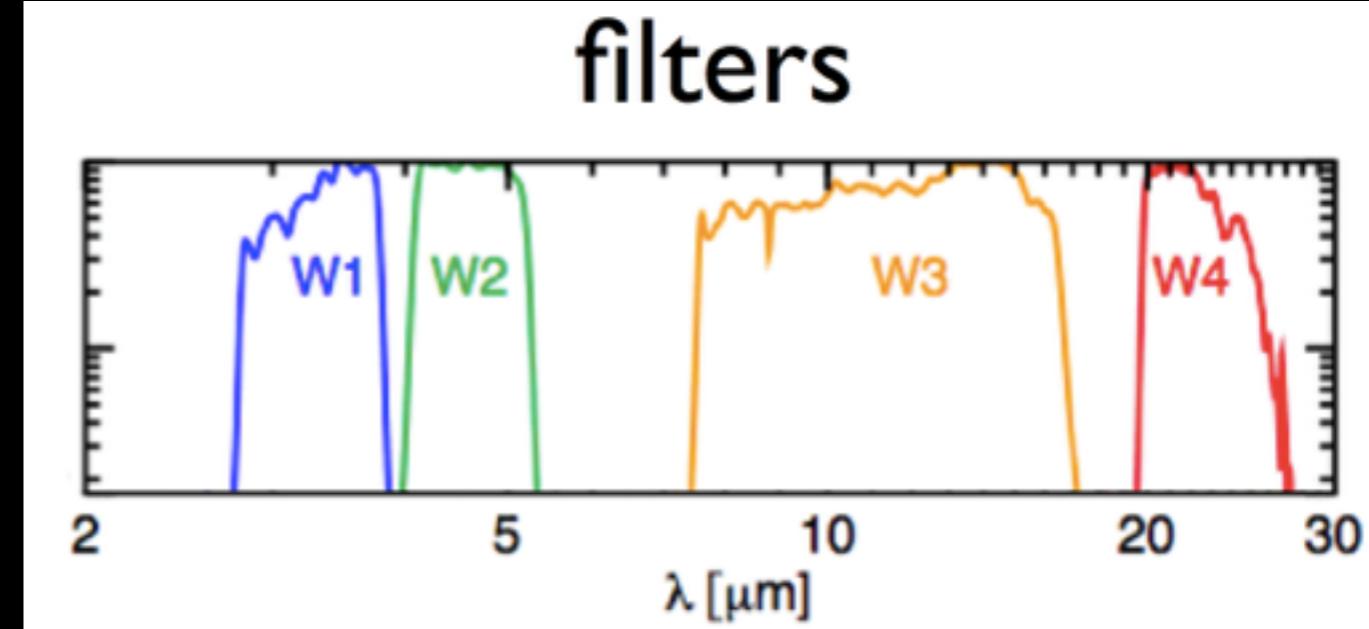


SDSS
optical

2MASS
near infrared

WISE
infrared

Planck
millimetric

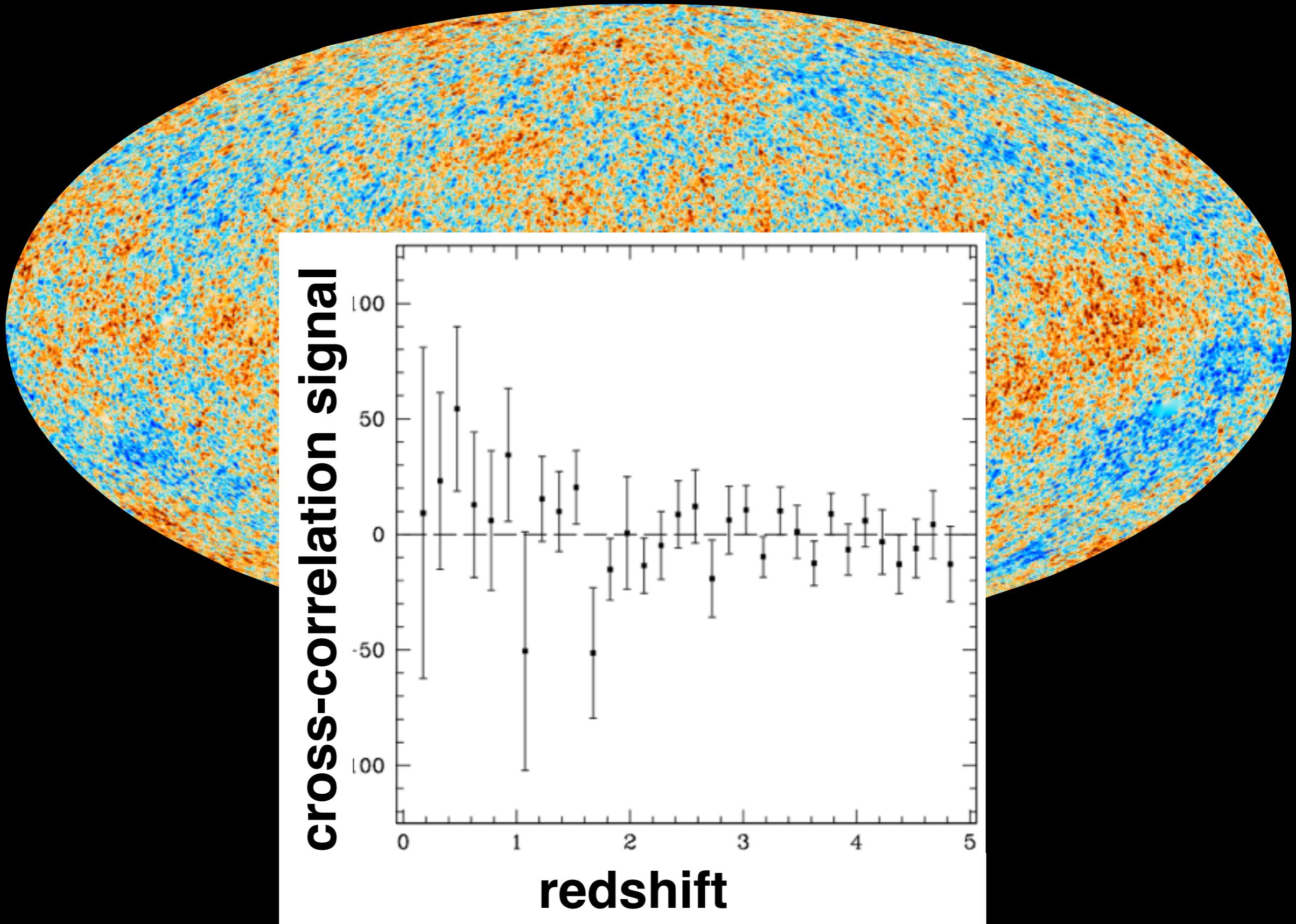


PSF ~ 6 arcsec
mostly point sources

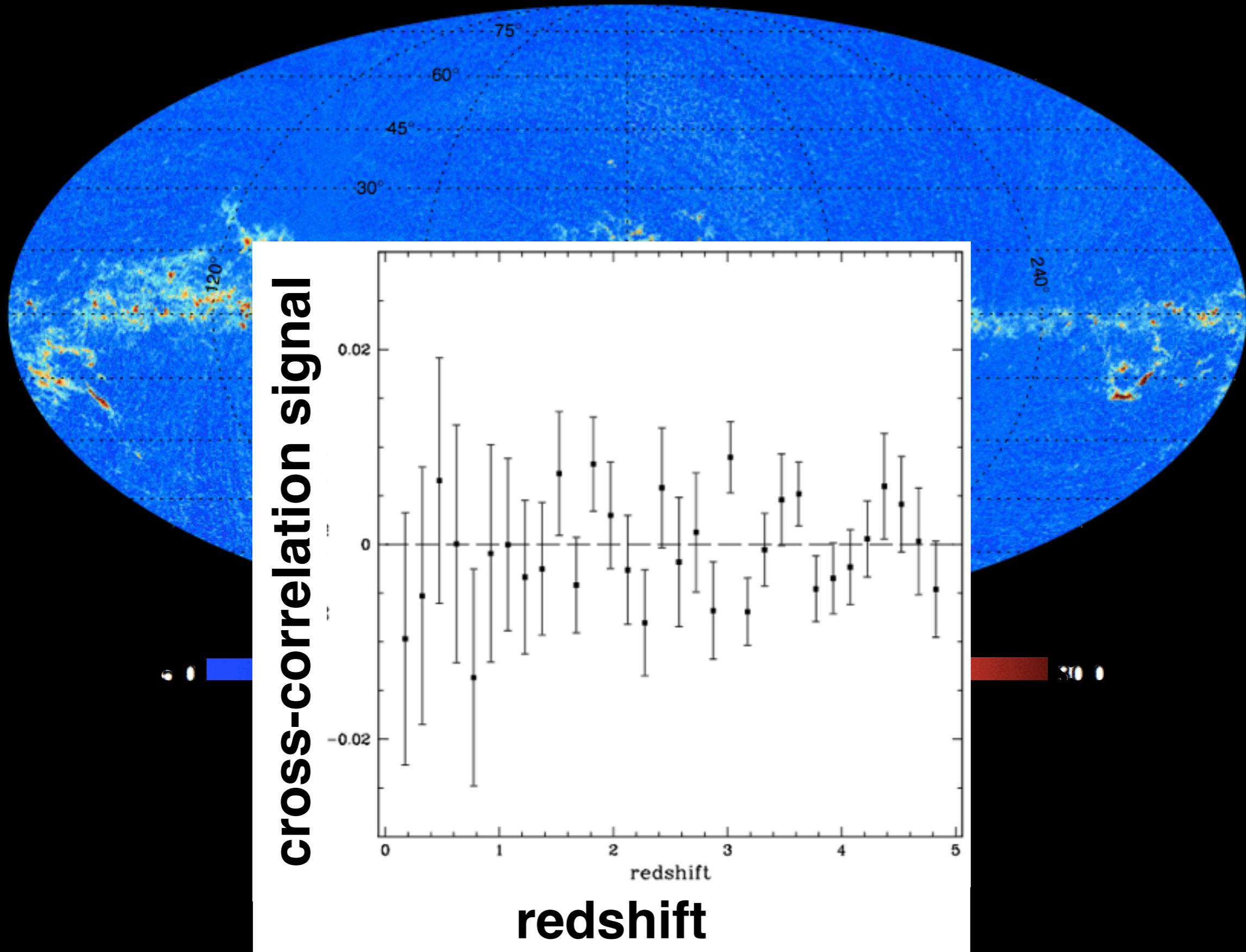
$W1(3 \mu\text{m}) < 16$ mag

millions of objects

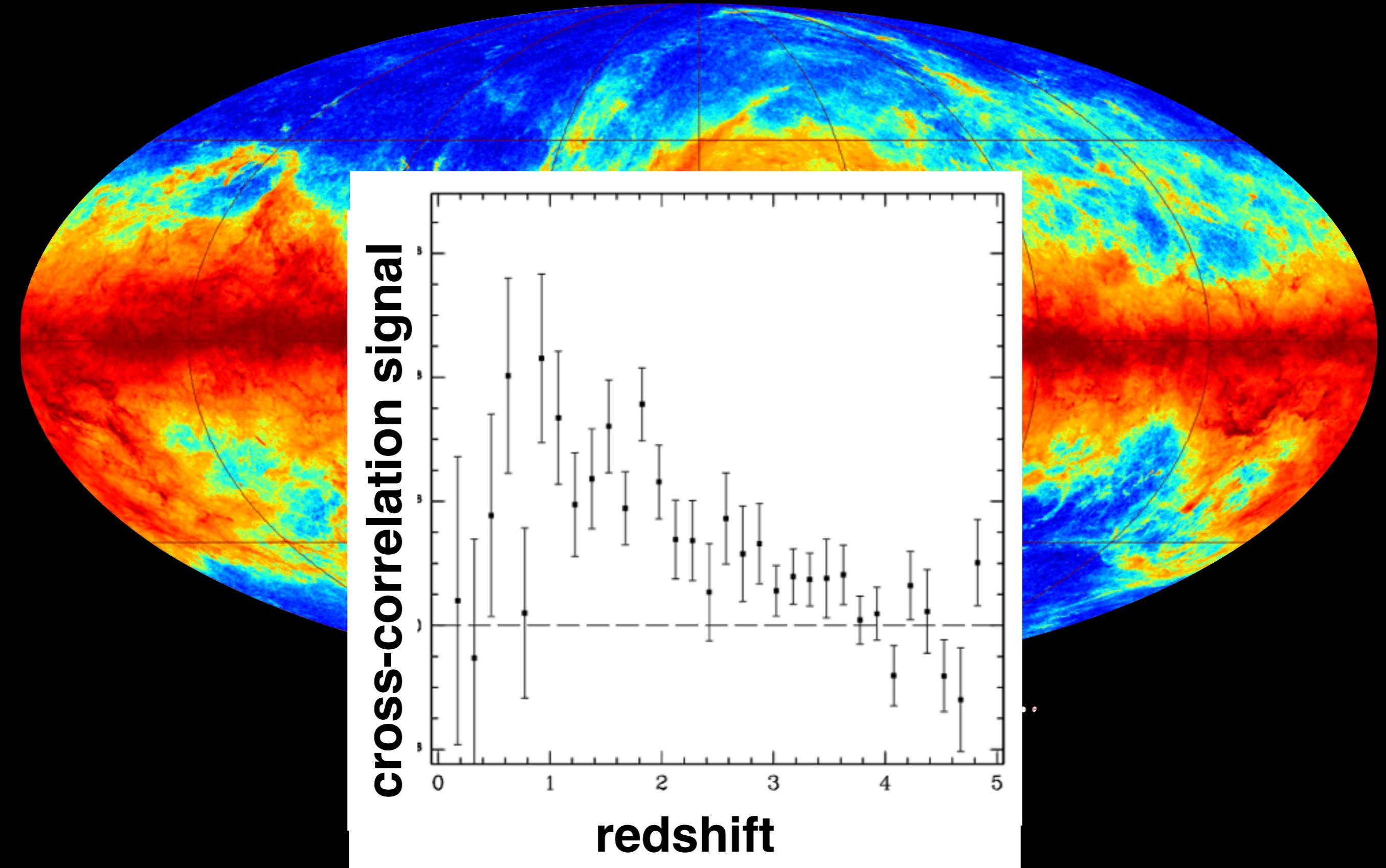
Planck CMB - SMICA map



Planck CO map J=0-1



Planck dust opacity map

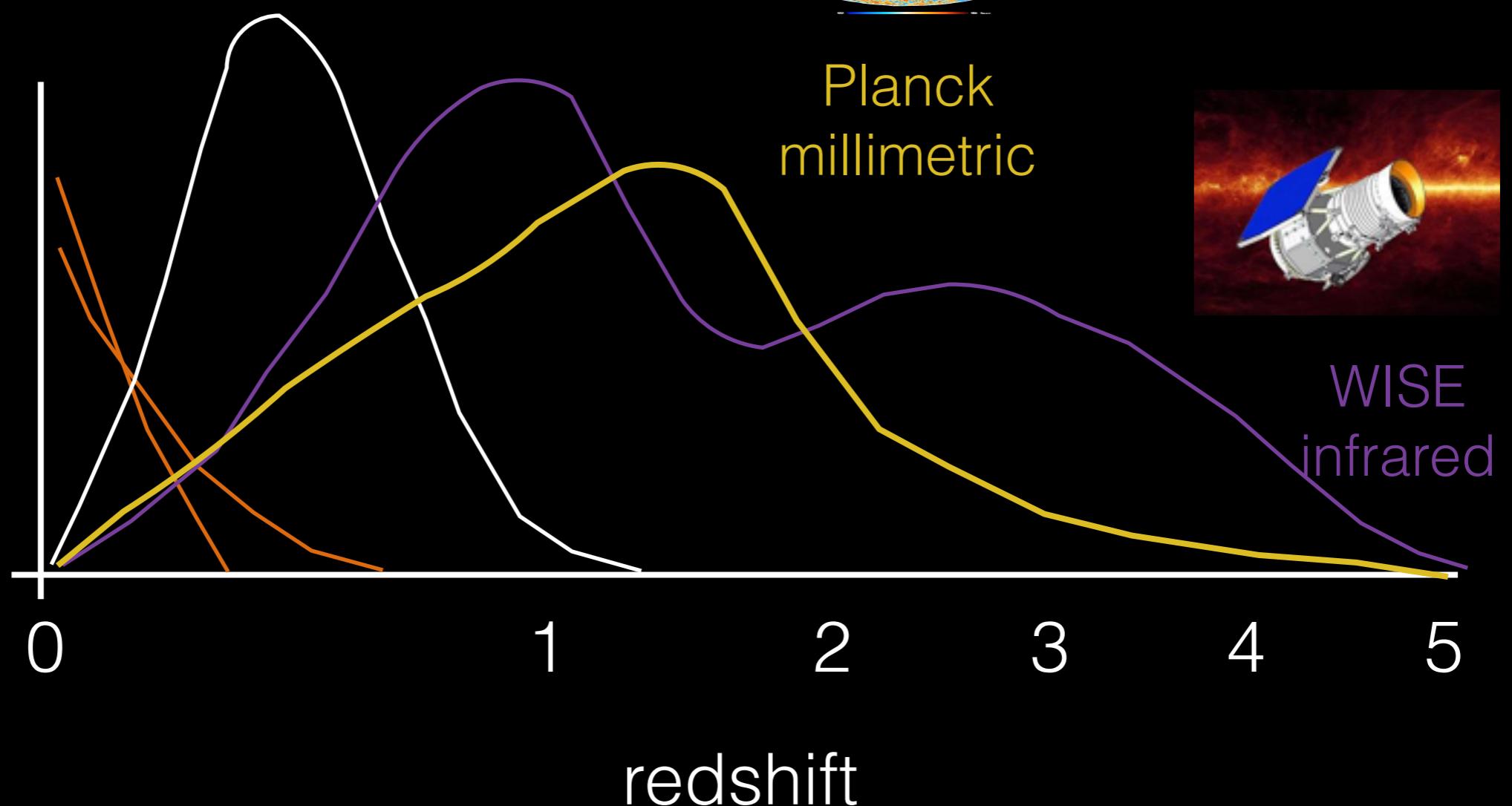




SDSS
optical



2MASS
near infrared

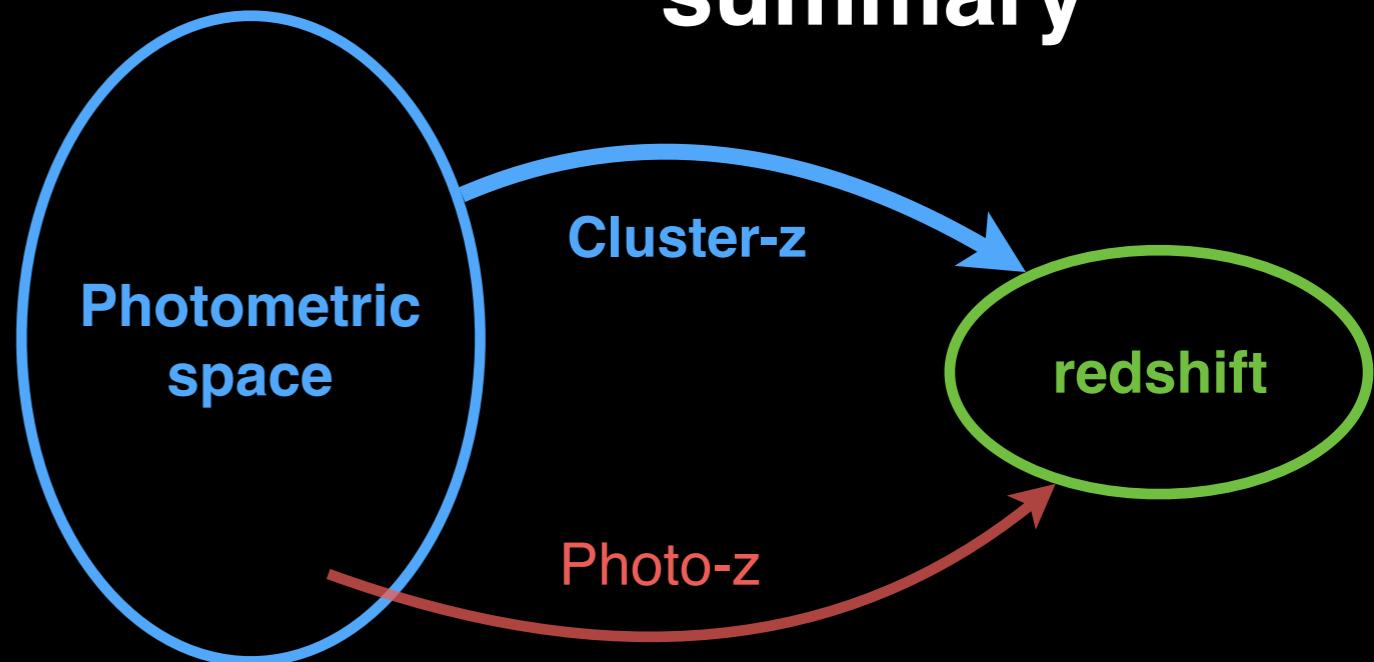


UV (GALEX), radio (FIRST, NVSS, ...), Gamma rays (Fermi), ...
as well as *combinations* of datasets

Clustering redshifts

summary

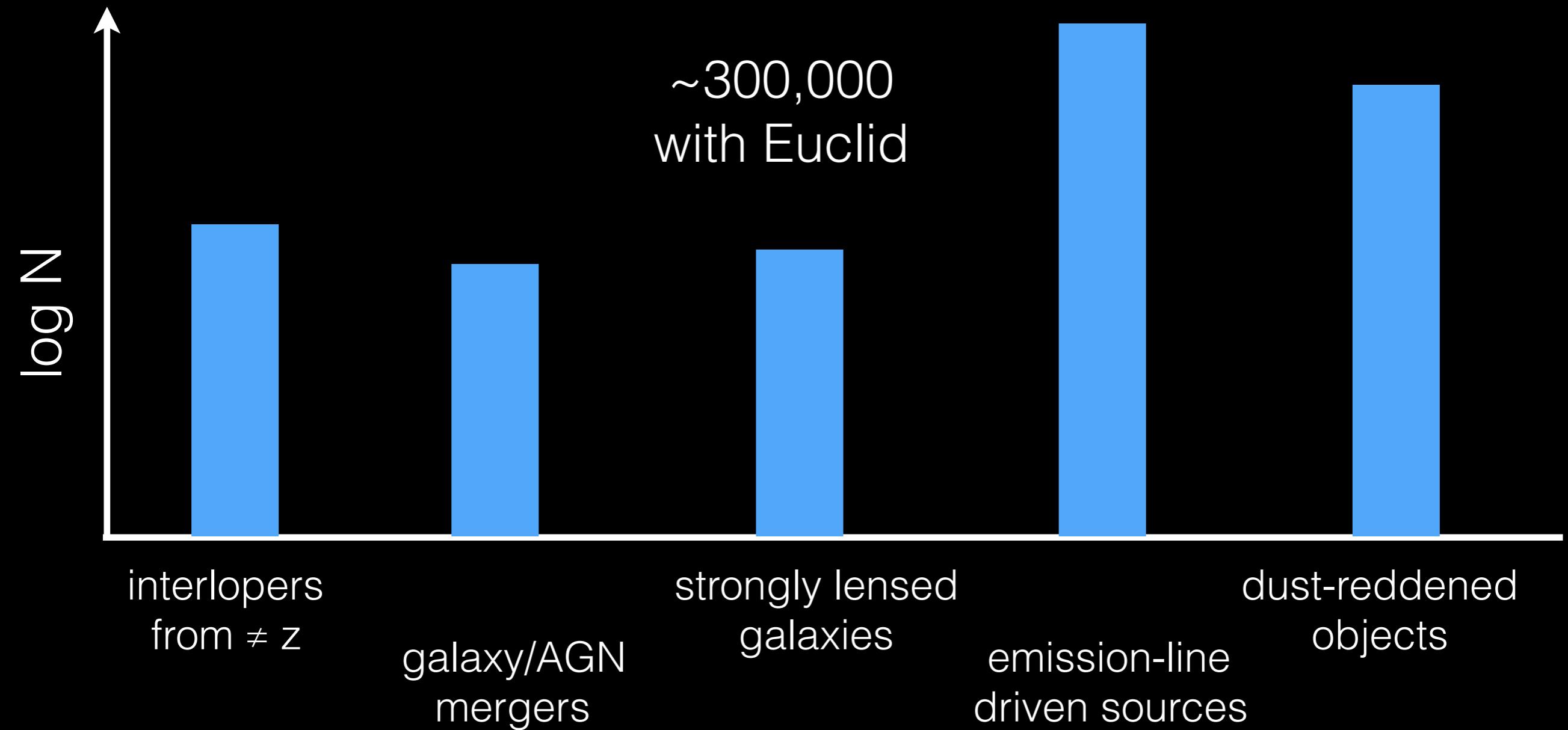
We have a new tool in hand to estimate the redshifts of photometric sources



We do not have to rely on source colors to estimate redshifts.

We now have two independent estimation techniques.

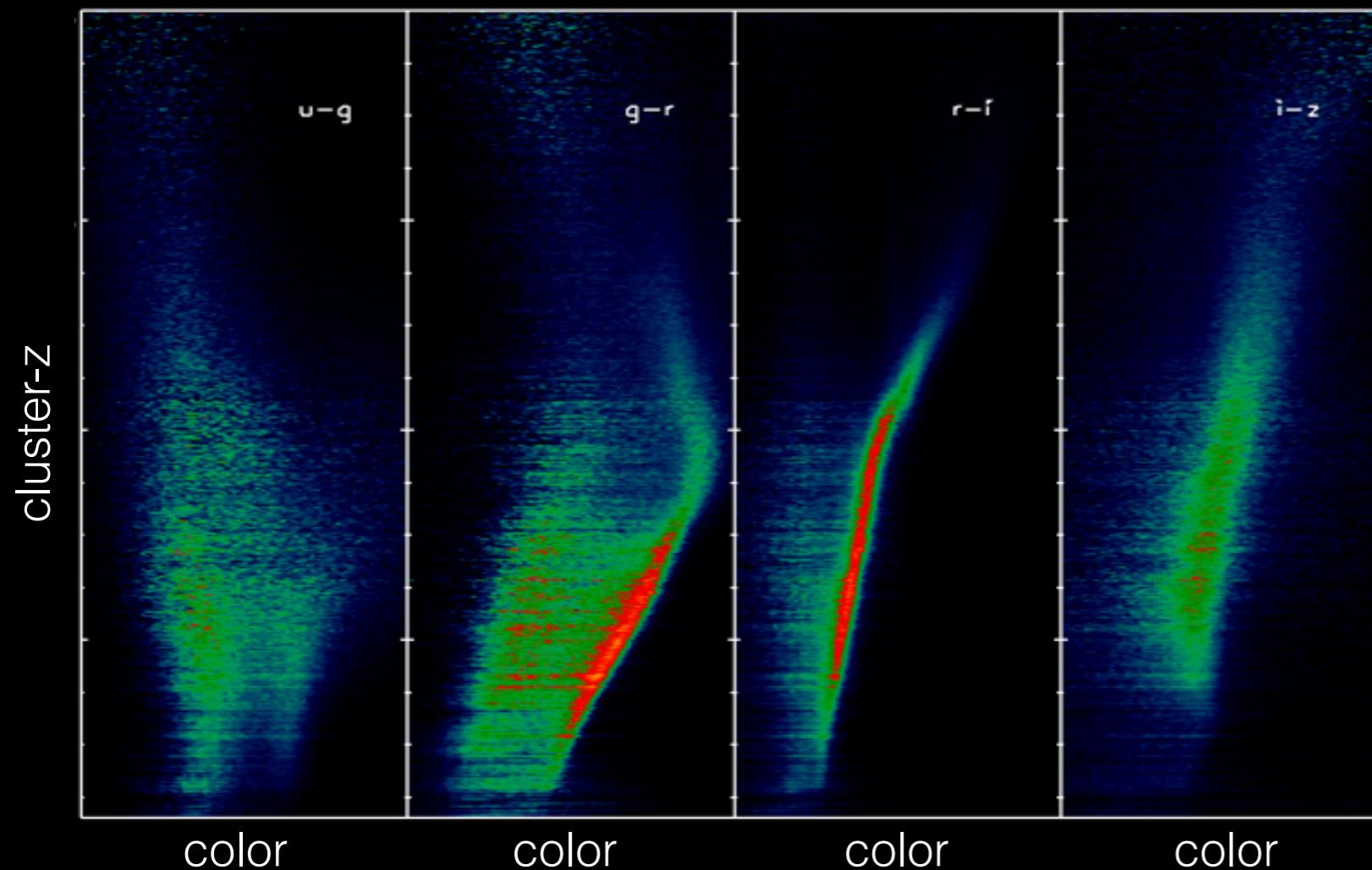
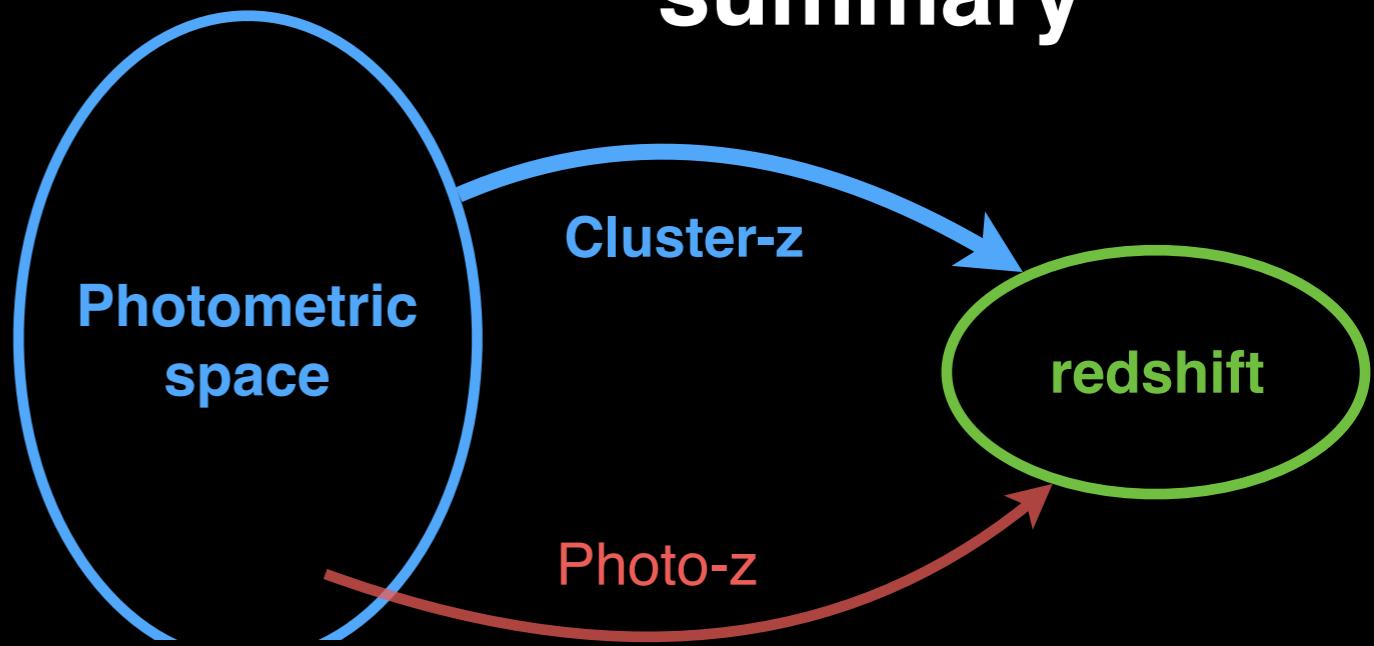
Difficult sources for photometric redshifts



Clustering redshifts

summary

We have a new tool in hand to estimate the redshifts of photometric sources



We can now “deproject” any photometric dataset, at any wavelength.