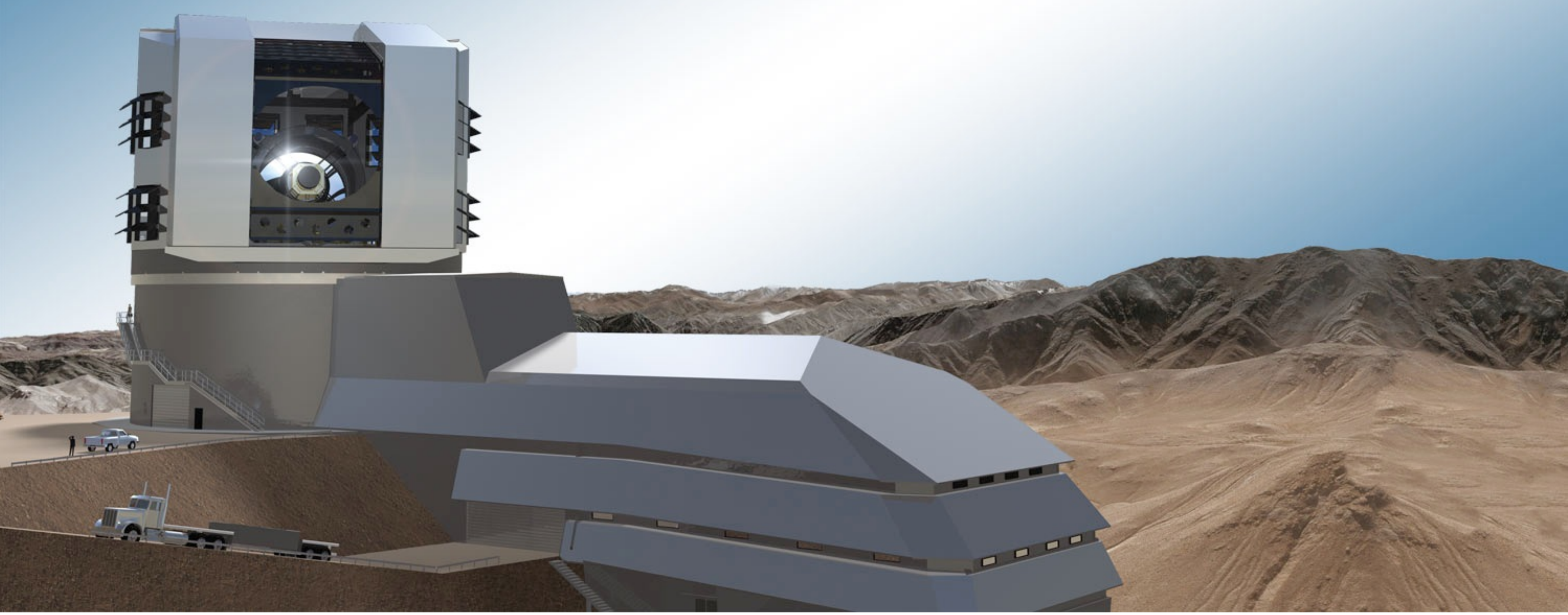
Statistical Challenges for Photometric Redshifts

Jeffrey Newman, U. Pittsburgh / PITT-PACC

Deputy Spokesperson, LSST Dark Energy Science Collaboration



Outline

- Overview of photometric redshifts
  - Template methods
  - Training-based methods
- Some open issues
  - Spectroscopic incompleteness
  - Robust training
  - p(z) coverage
  - Combining results from multiple codes
  - p(z,α) storage
  - Optimizing spectroscopic samples
  - Defining ideal LSST algorithm
- Some examples of issues with current codes

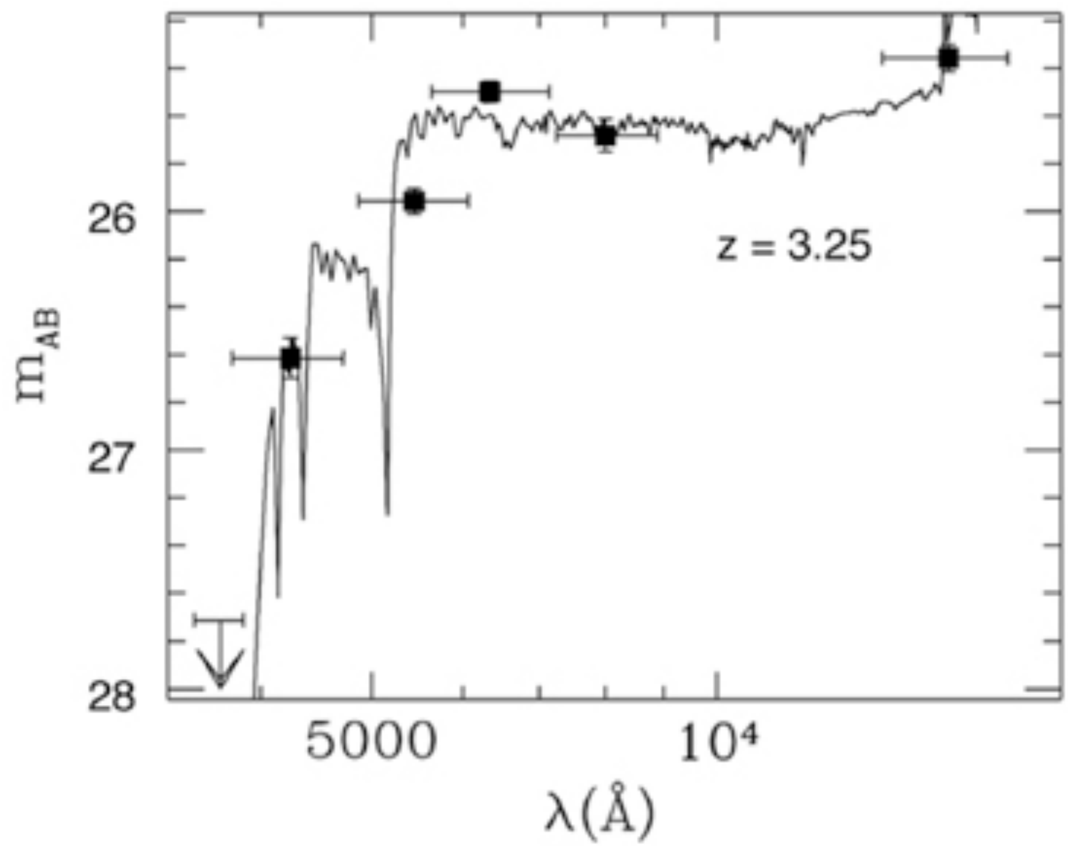
Spectroscopy provides ideal redshift measurements – but is infeasible for large samples

- Redshift ('z') measurements allow us to determine how far back in Universe's history we are looking for an object
- Study galaxy evolution, cosmology, etc. by measuring properties as a function of redshift
- To determine: measure spectrum of light from object with spectrograph; compare observed wavelengths of spectral features to rest frame values to get z
- At LSST "gold sample" (i<25.3) depths, ~100 hours on a 10m telescope to determine a redshift (75% of time) spectroscopically
- With a next-generation, 5000-fiber spectrograph on a 10m telescope, still >50,000 telescope-years to measure redshifts for LSST "gold" weak lensing sample (4 billion galaxies)!

Credit: ESO

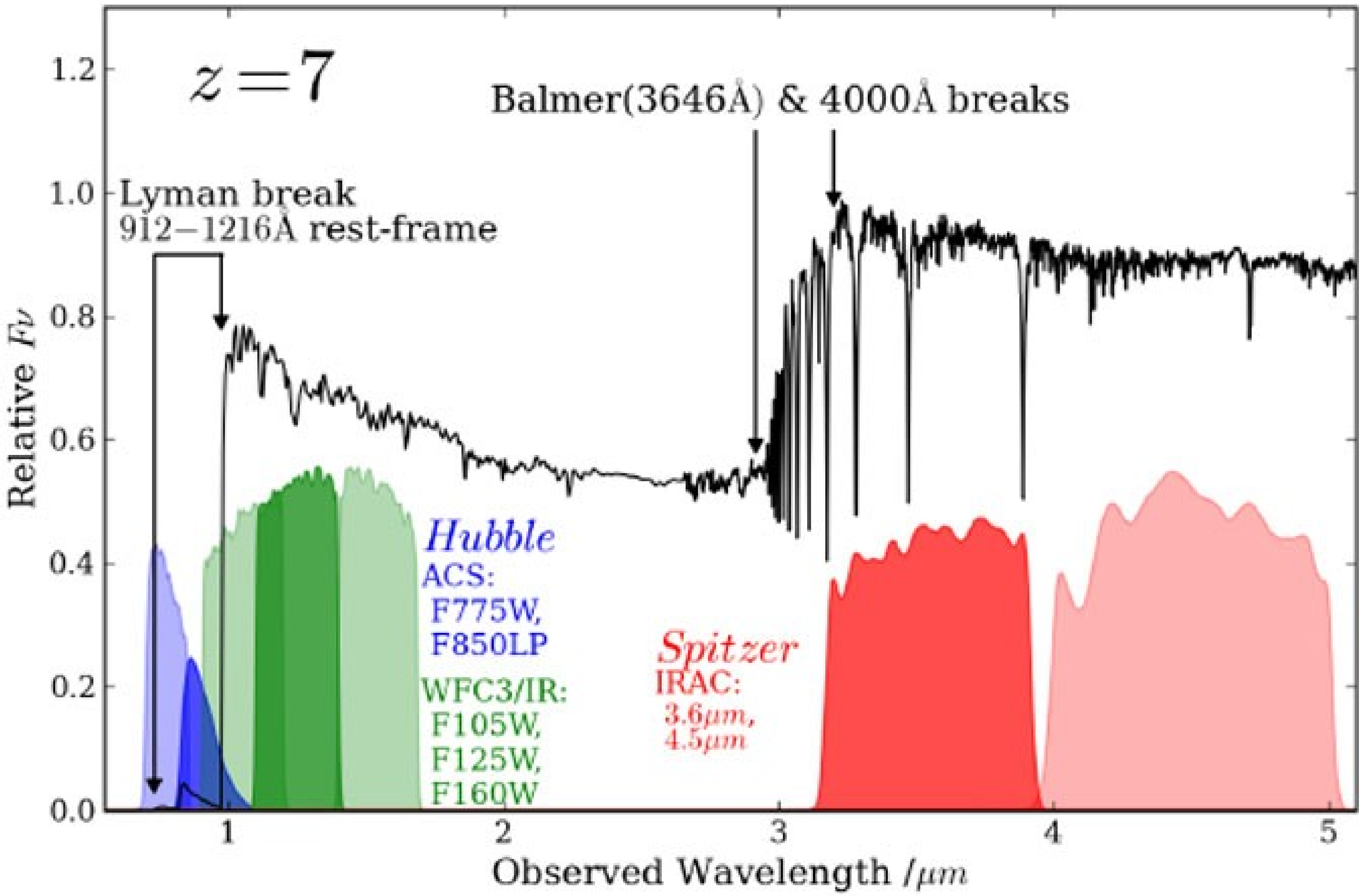
Spectroscopy provides ideal redshift measurements – but is infeasible for large samples

- **Alternative: use broad spectral features to determine *z* : a photometric redshift or photo-z**
- **Advantage: high multiplexing**
- **Disadvantages: lower precision, calibration uncertainties**



Credit: ESO

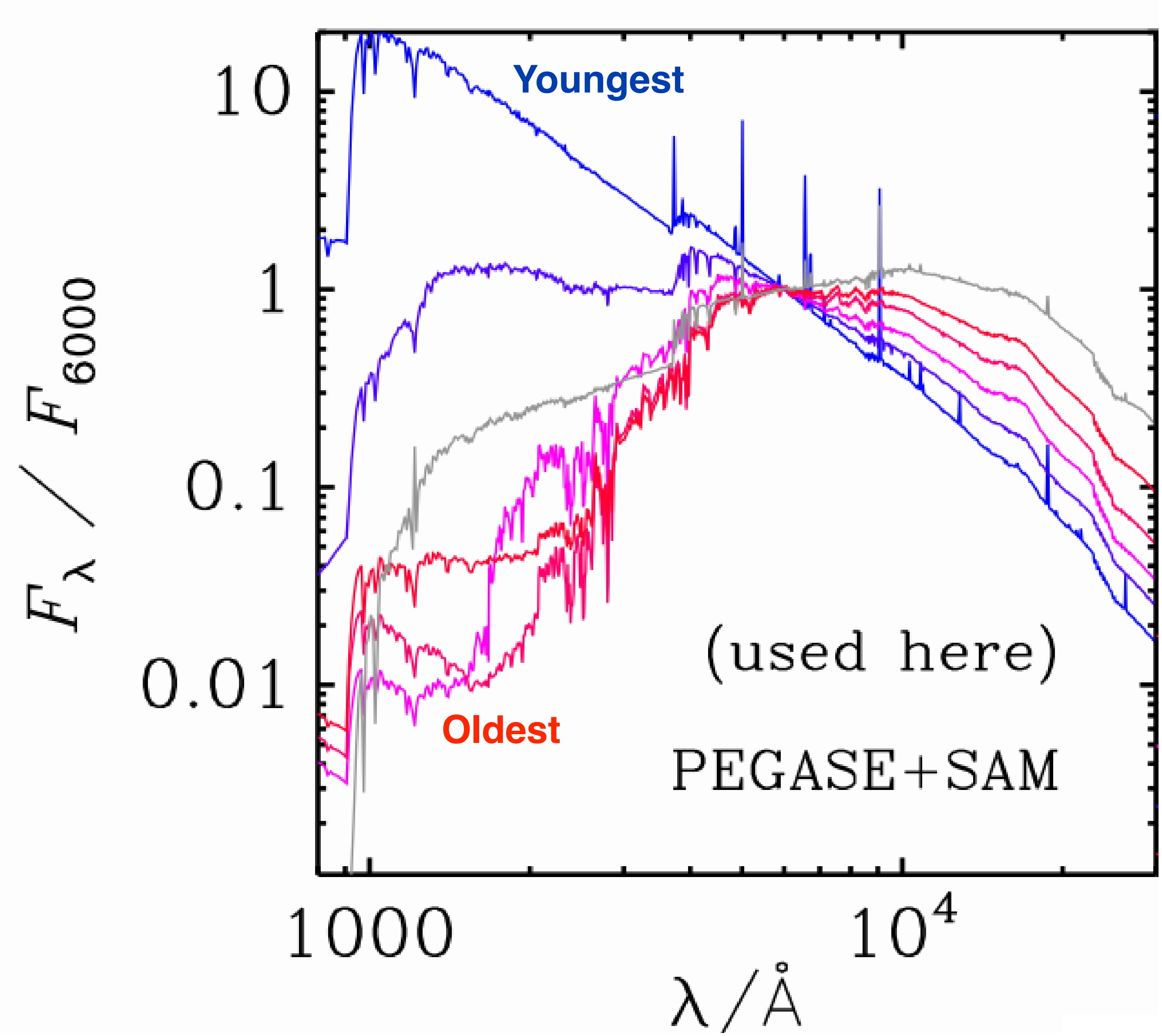
Photometric redshifts rely on the existence of broad spectral features in galaxy spectra...



Dunlop 2012

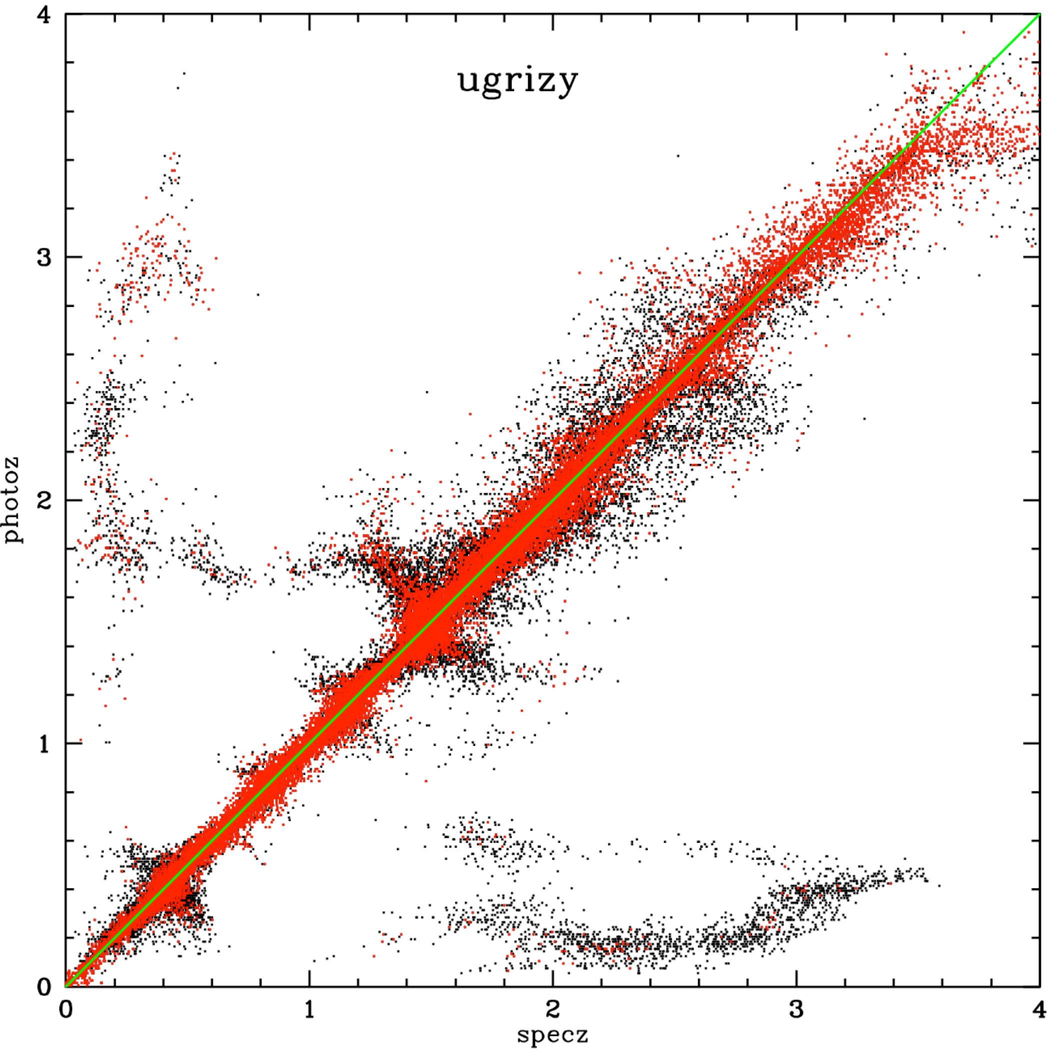
but those features are stronger in some galaxies than others

- Galaxies with older stellar populations exhibit stronger 'breaks'
- As a result, photo-z's can be more precise for redder galaxies
- At higher redshifts, blue galaxies with young stellar populations dominate - photo-z problem gets harder

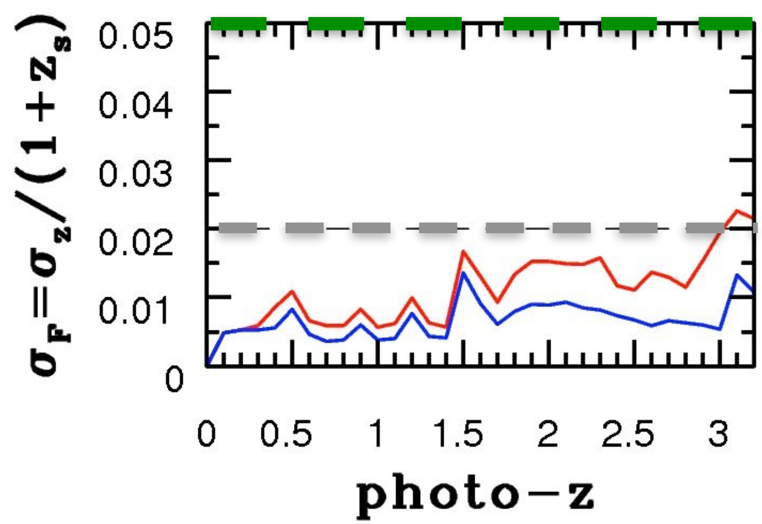


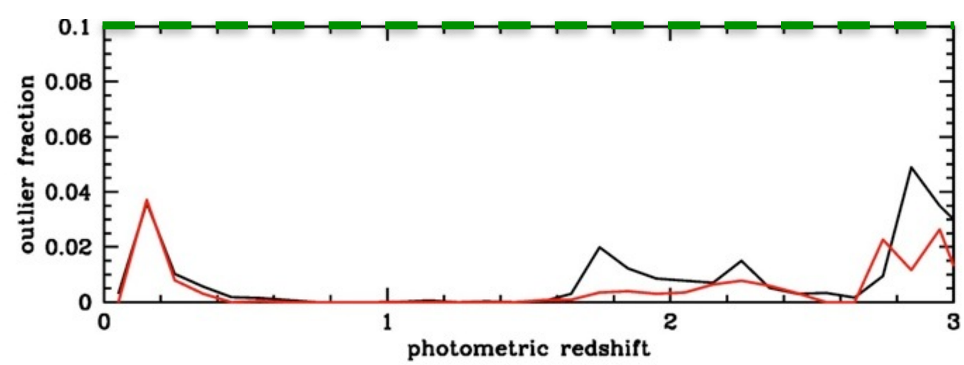
Brammer et al. 2008

Example: expected photo-z performance for LSST *ugrizy*



Green: Requirements on actual performance; grey: requirements on performance <u>with perfect template knowledge</u> (as in these sims)

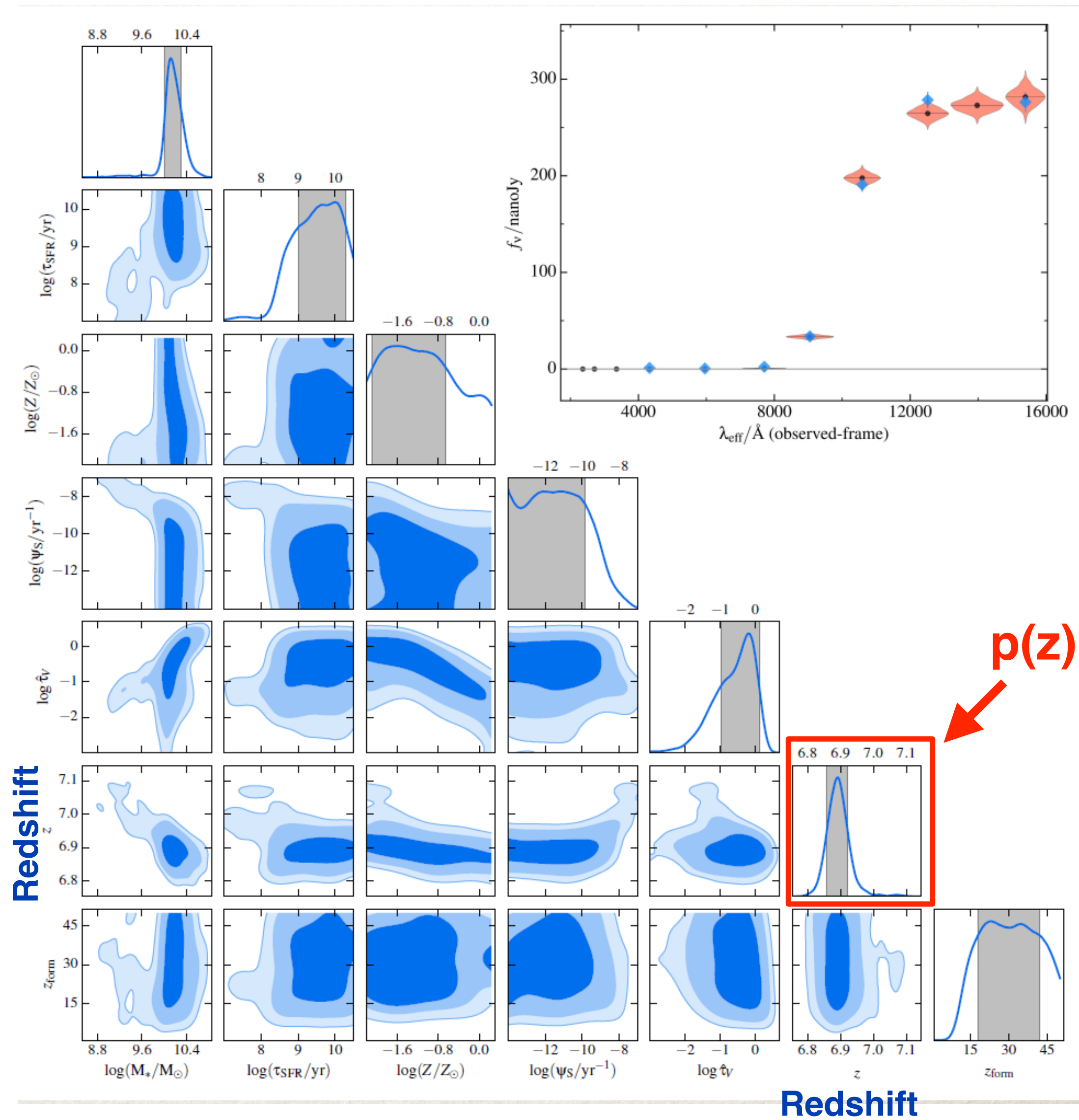




S. Schmidt

Basic methods: Template fitting photo-z's

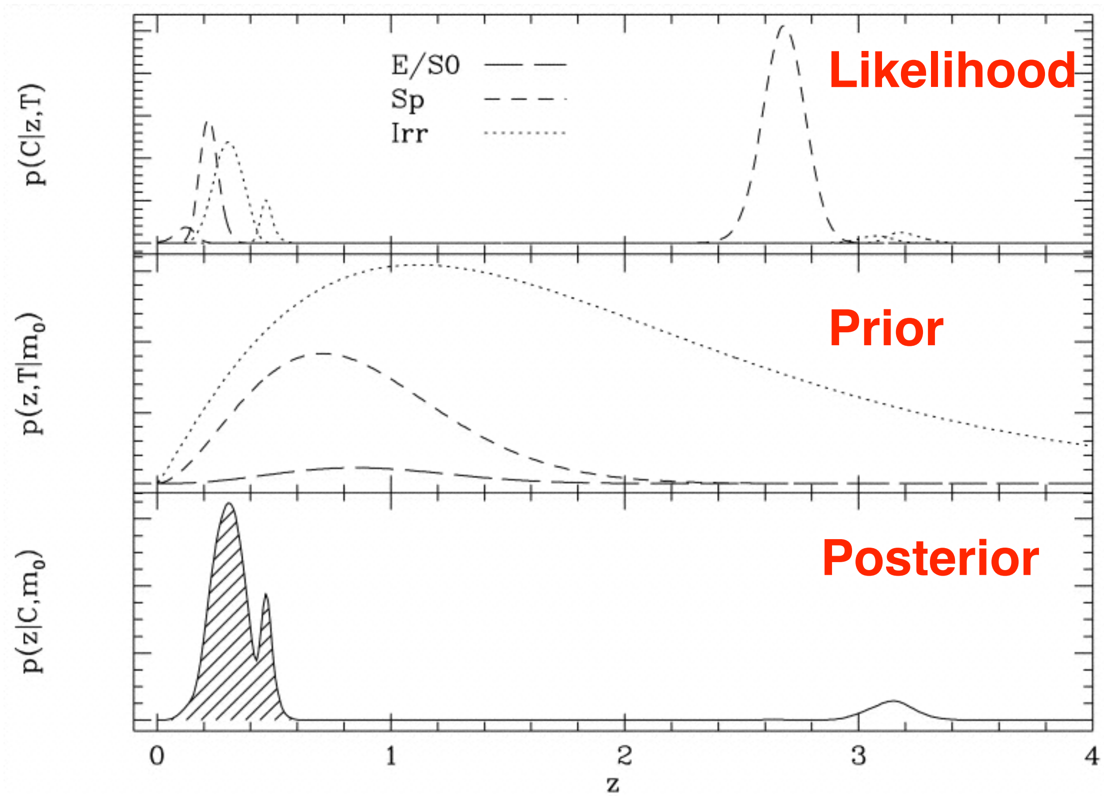
- **Generally determine posterior probability distribution for *z* | fluxes: p(z)**
- **Can also provide info on galaxy properties from template fit**
- **E.g., template index *T* or galaxy parameters α_i such as stellar mass, star formation rate, etc.): p(z,α) (cf. López-Sanjuan talk)**



Chevallard & Charlot 2016

Basic methods: Template fitting photo-z's

- **Typical algorithms:**
  - Determine likelihood of colors (=ratios of fluxes between bands) as a function of *z* and template
    - Often via $\chi^2(z,T)$ or min($\{\chi^2(z|T)\}$); some algorithms use linear combinations of templates



Benitez 2000

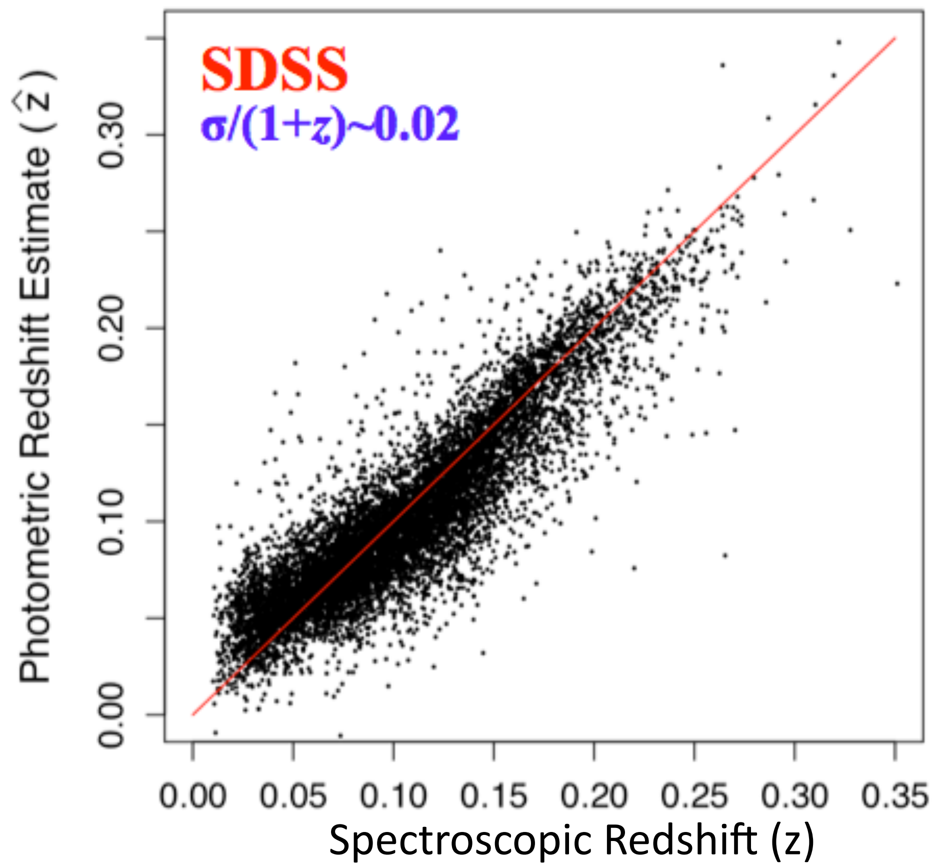
  - Typically utilize prior for redshift or redshift & type based on magnitude (sometimes size/morphology as well)
  - Then multiply to get posterior. . .

Can use spectra of galaxies spanning full range of possible properties to tune templates/filter systems, establish priors, etc.

Basic methods: Training-based photo-z's

- Use galaxies with known redshift and uniform/well-understood sampling to determine relationship between *z* and colors/fluxes

- Can take advantage of progress in machine learning & stats, but generally extrapolate poorly; Training set MUST span full range of properties & z of galaxies

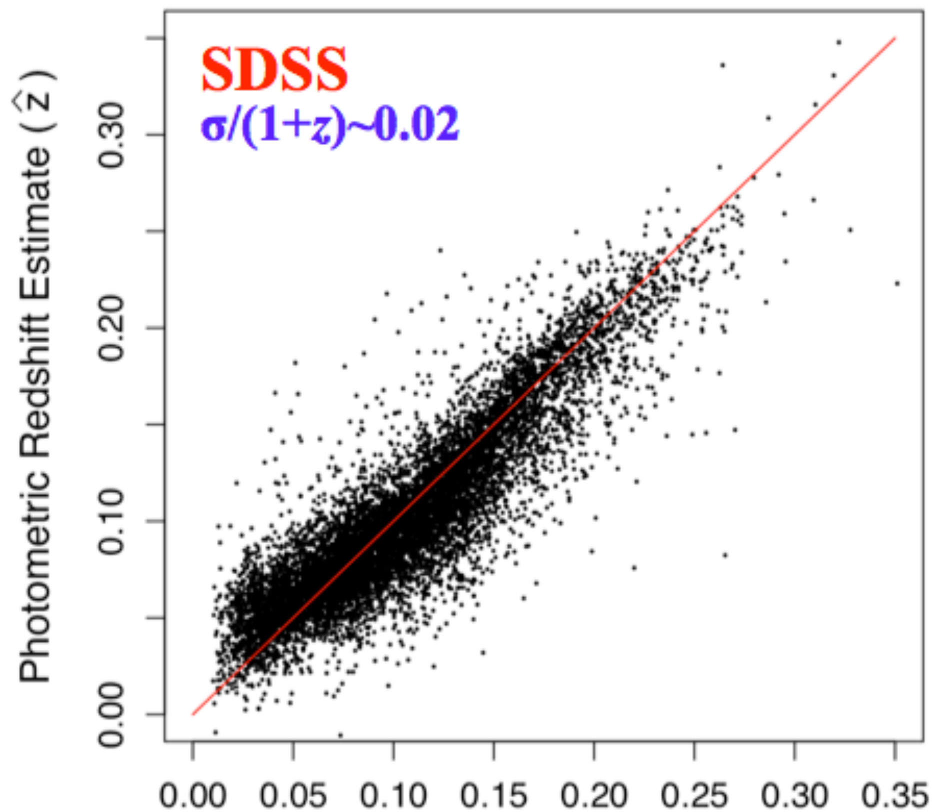


Freeman, JN et al. 2009

Basic methods: Training-based photo-z's

- Many algorithms: e.g.
  - Neural networks
  - Boosted Decision Trees
  - Random Forest regression
  - k-Nearest Neighbor
  - Diffusion map + regression

- For bright, nearby galaxies, training sets are ~complete and both template-based & training-set-based algorithms perform similarly

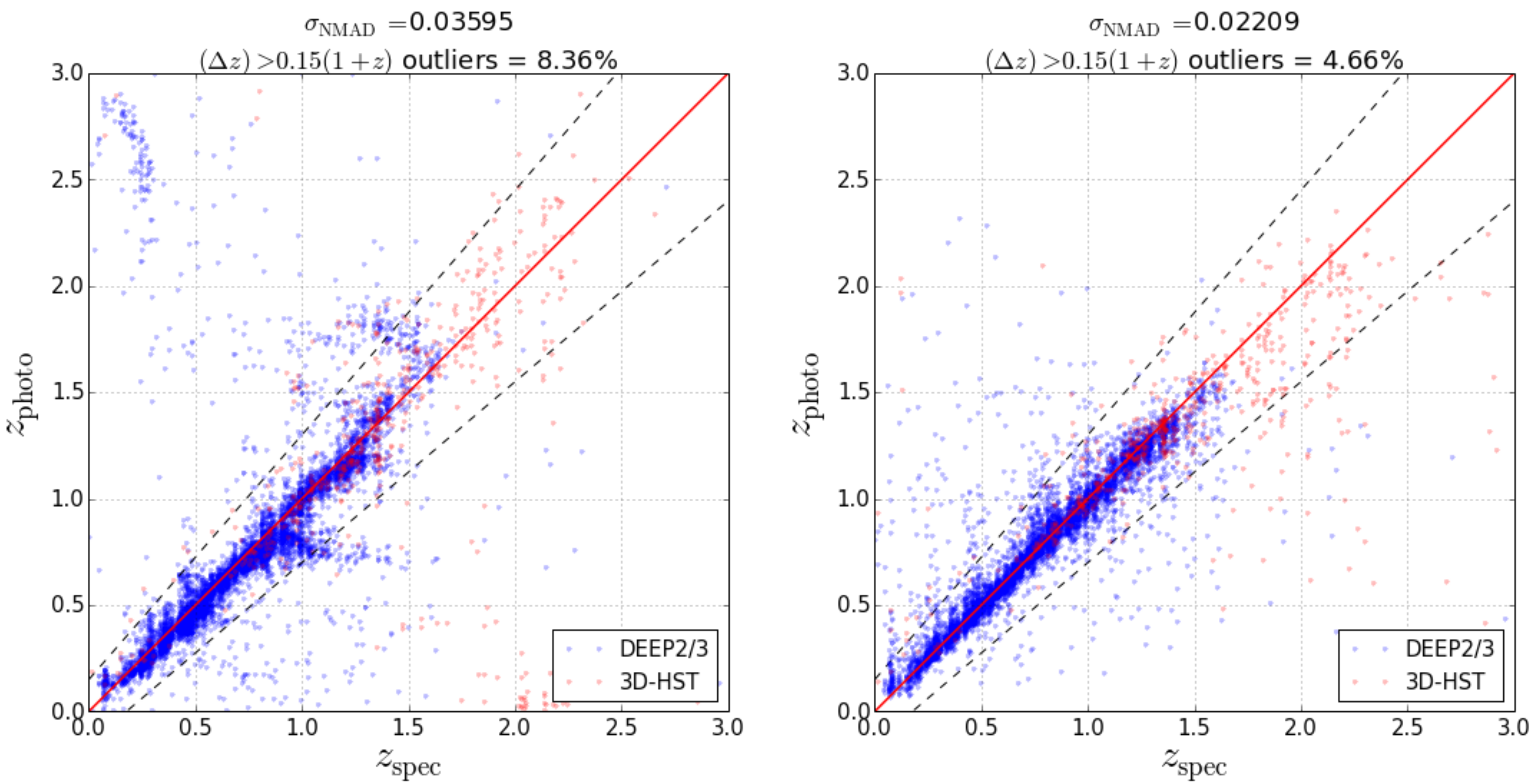


Freeman, JN et al. 2009

At higher redshifts, the photo-z problem is more difficult

- **Zhou et al. 2016 (in prep.): empirical, LSST-like dataset: CFHT LS *ugriz* + Subaru *y* + DEEP2/DEEP3/3D-HST redshifts**

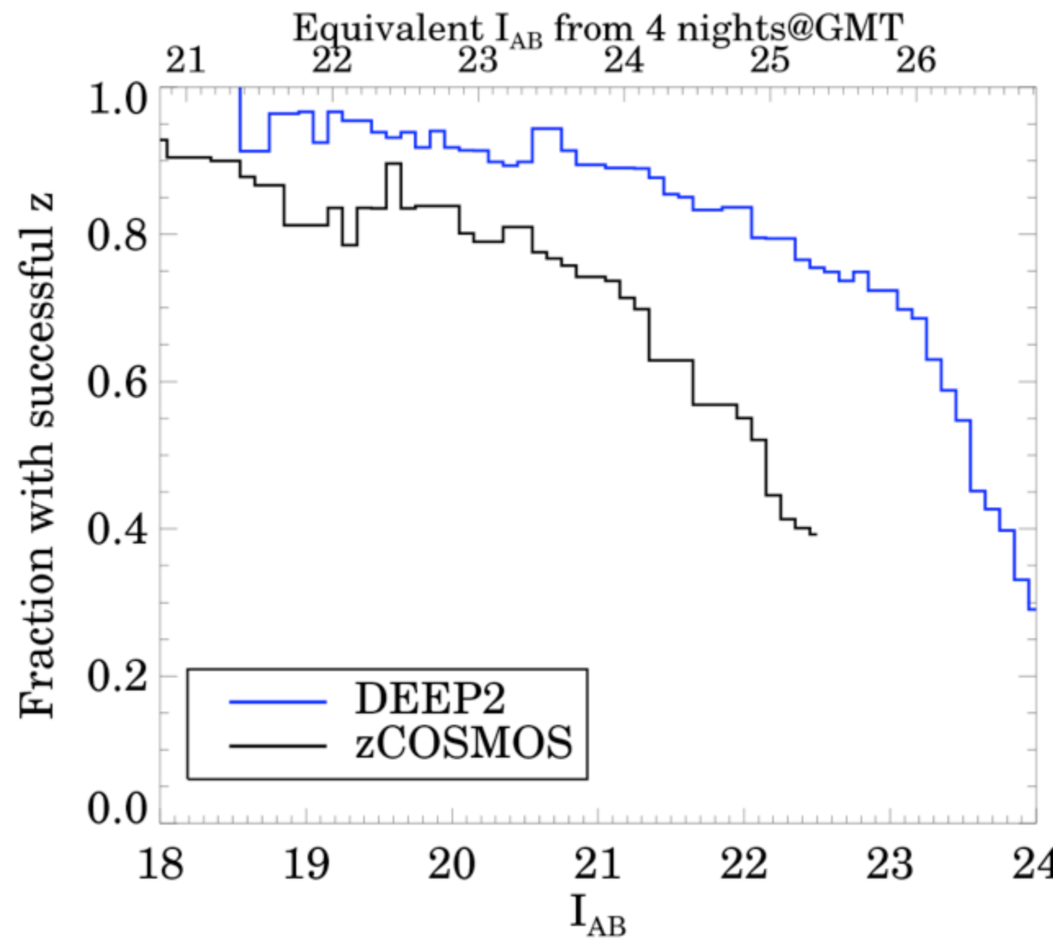
EAZY (template code, untuned) **Random Forest Regression**



Zhou, JN et al. 2016, in prep.

Open issues: dealing with incompleteness in training/ calibration datasets

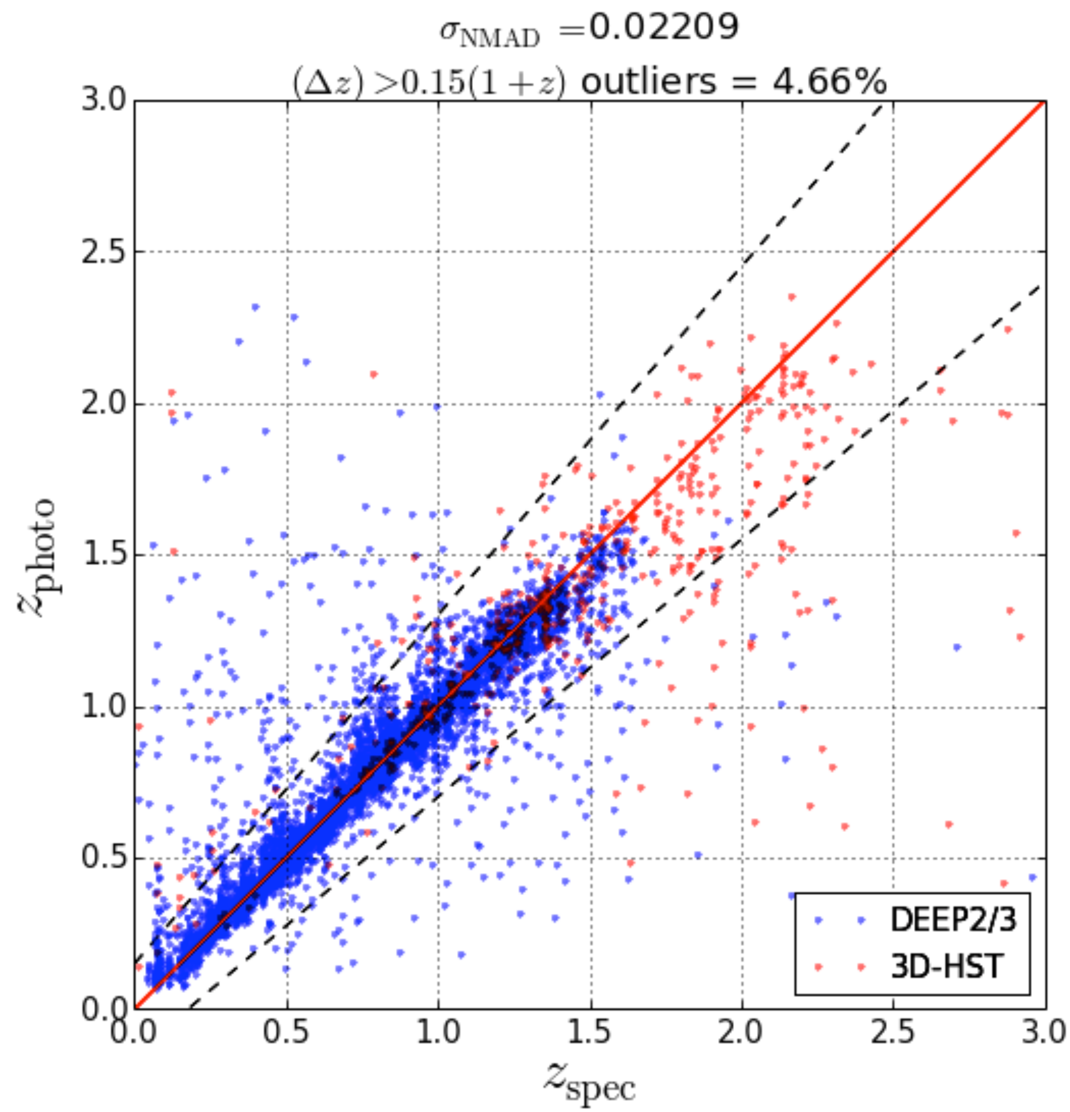
- In current deep spectroscopic surveys, 25-60% of targets fail to yield secure redshifts
- z success rate depends on galaxy properties
- Estimated need 99-99.9% completeness to prevent systematic errors in calibration, unless apply other methods (e.g., cross-correlations)
- Major issue for training-set techniques



Data from DEEP2 (Newman et al. 2013) and zCOSMOS (Lilly et al. 2009)

Open issues: Robust training methods

- **1% incorrect-redshift rate is sufficient to bias photo-z's beyond tolerances**
- **Depending on survey, up to 5% of 'secure' redshifts are incorrect**
- **If can train algorithms in a manner robust to outlier/ wrong redshifts, could use the broader set of less-secure spectroscopic redshifts**
- **ML methods that extrapolate well would also be interesting**



Zhou, JN et al. 2016, in prep.

Open issues: Making posteriors great again

- **CANDELS code comparison: Dahlen et al. 2013**
- **11 code/template combinations were tested using ~600 redshifts in GOODS-S (trained with a separate set of 600 redshifts)**
- **Generally χ^2 minimization, generally with some sort of prior.**
- **Codes with p(z)'s available are marked by ★**

Code		Code ID	Template set	bias$_z^a$	OLF^b	σ_F^c	σ_O^d	
Rainbow		A	PEGASE^b	-0.010	0.092	0.167	0.041	
GOODZ		B	CWW^c, Kinney^d	-0.007	0.036	0.099	0.035	
EAZY	★	C	EAZY^e+BX418^f	-0.009	0.051	0.114	0.044	
SPOC		D	BC03^g	-0.030	0.147	0.197	0.073	
zphot	★	E	PEGASEv2.0^b	-0.007	0.041	0.104	0.037	
EAZY		C	EAZY^e	-0.009	0.053	0.121	0.037	
SATMC		F	BC03^g	-0.008	0.093	0.272	0.064	
HyperZ		G	Maraston05^h	0.013	0.078	0.189	0.050	
LePhare	★	H	BC03^g+Polletta07^i	-0.008	0.048	0.132	0.038	
WikZ	★	I	BC03^g	-0.023	0.046	0.153	0.049	
EAZY	★	C	EAZY^e	-0.005	0.039	0.127	0.034	
				-0.008	0.029	0.088	0.031	median(all)
				-0.009	0.031	0.079	0.029	median(5)

Dahlen et al. 2013

Open issues: Making posteriors great again

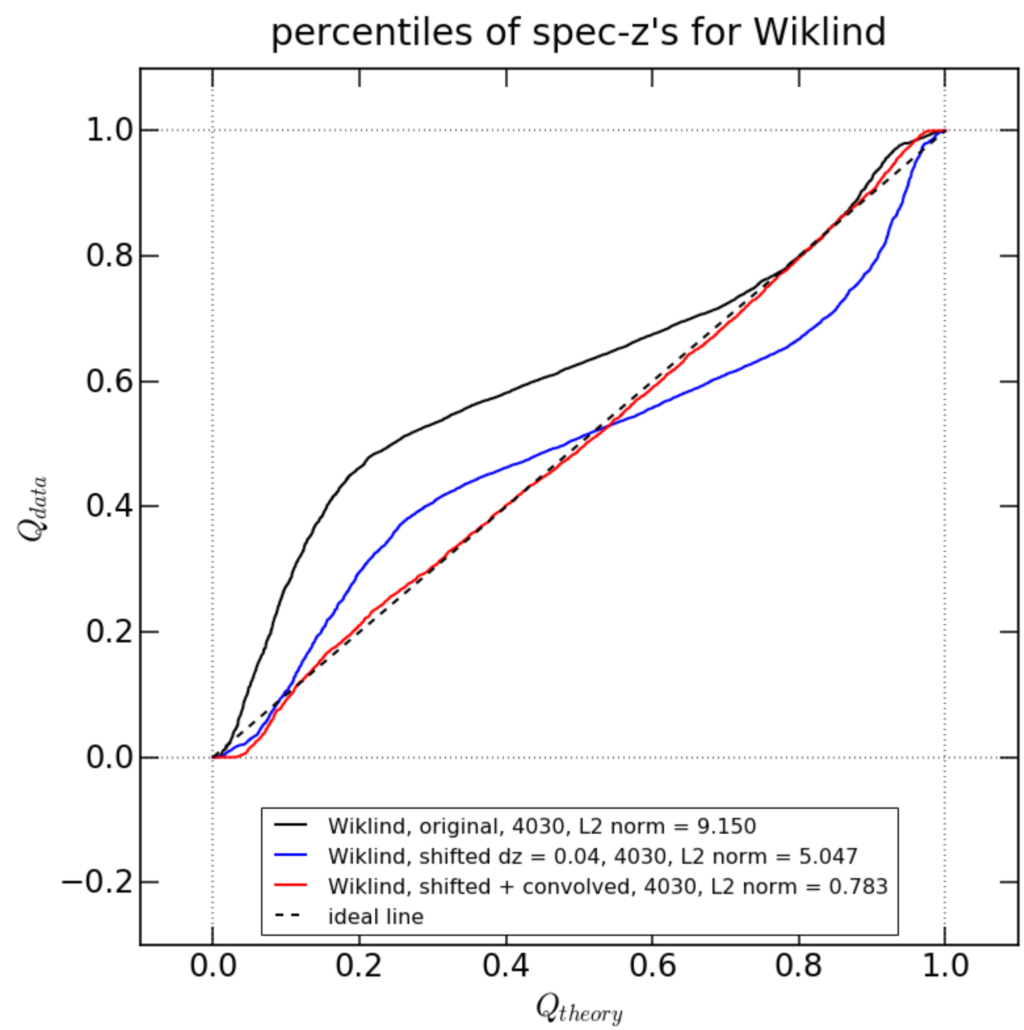
- Many analyses assume that photo-z codes are providing posterior PDFs with proper coverage (and assuming that they can add PDFs to get N(z); cf. Alex Malz's talk)
- Dahlen et al. 2013 tested the fraction of spectroscopic redshifts that are in the inner 68% or inner 95% of their PDFs
- Coverage is all over the place; no codes were good at both 68% and 95% points

Code		WFC3 *H*-selected	
conf. int:		68.3%	95.4%
2A		46.1	
3B		81.6	92.8
4C	★	64.0	88.2
5D		2.5	4.2
6E	★	52.0	84.7
7C		65.0	87.3
8F		15.3	15.6
9G		16.3	44.1
11H	★	35.2	54.0^a
12I	★	88.7	96.7
13C	★	52.0	72.7

Dahlen et al. 2013

Open issues: Making posteriors great again

- **LSST Dark Energy Science Collaboration is setting up controlled tests of the problem**
- **Meanwhile, kludge in Kodra et al. 2016: modify *p(z)*'s**
  - **Shift by constant in *z* direction; convolve with Gaussian kernel; and take to a power (equivalent to rescaling errors in χ^2 calculation)**
- **Optimize parameters by minimizing total L2 norm of deviation in Q-Q plot from expected line**

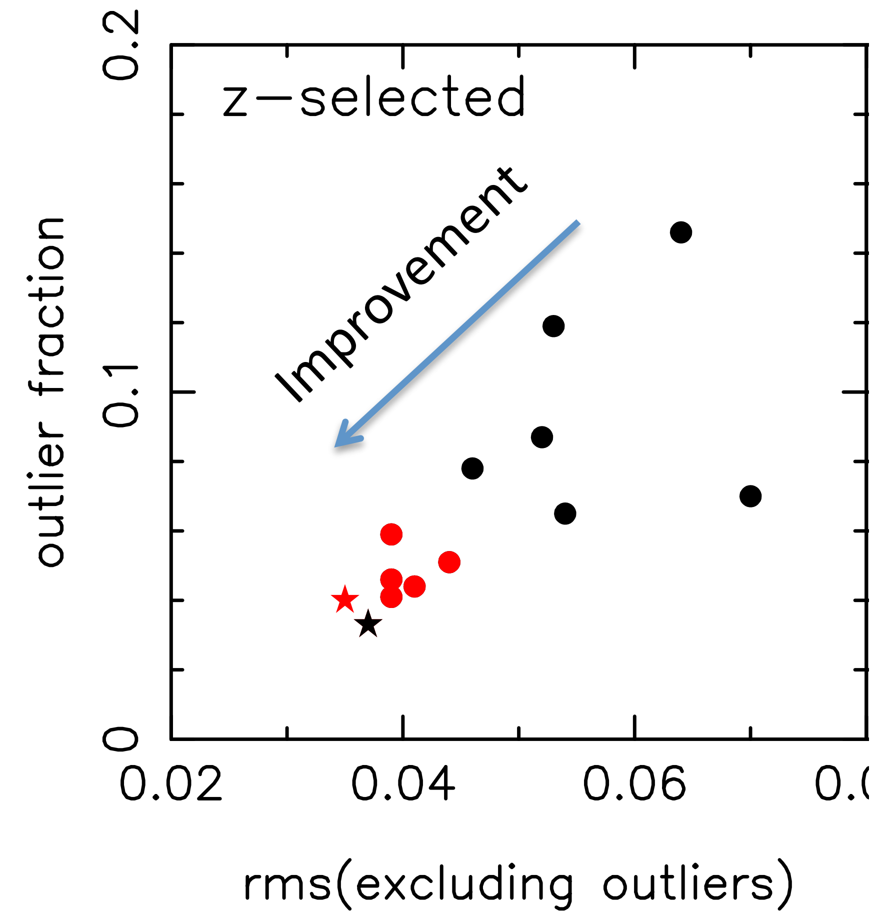


Kodra, JN et al. 2016, in prep.

Open issues: Combining PDF results from multiple codes

- **Dahlen et al. found that medians of point estimates from multiple codes (★'s) have smaller scatter (relative to spec-z) than any individual code**

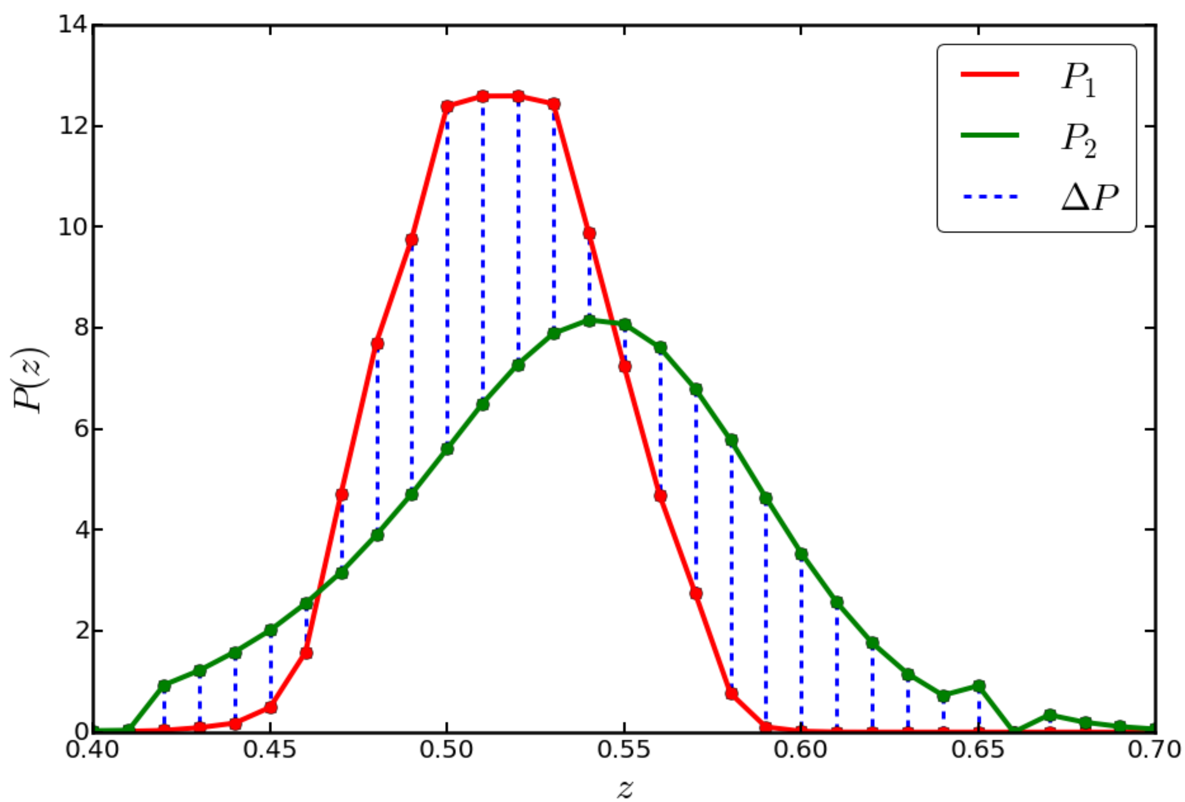
- **All codes are run on the same data! Current codes do not make optimal use of available information...**



Dahlen et al. 2013

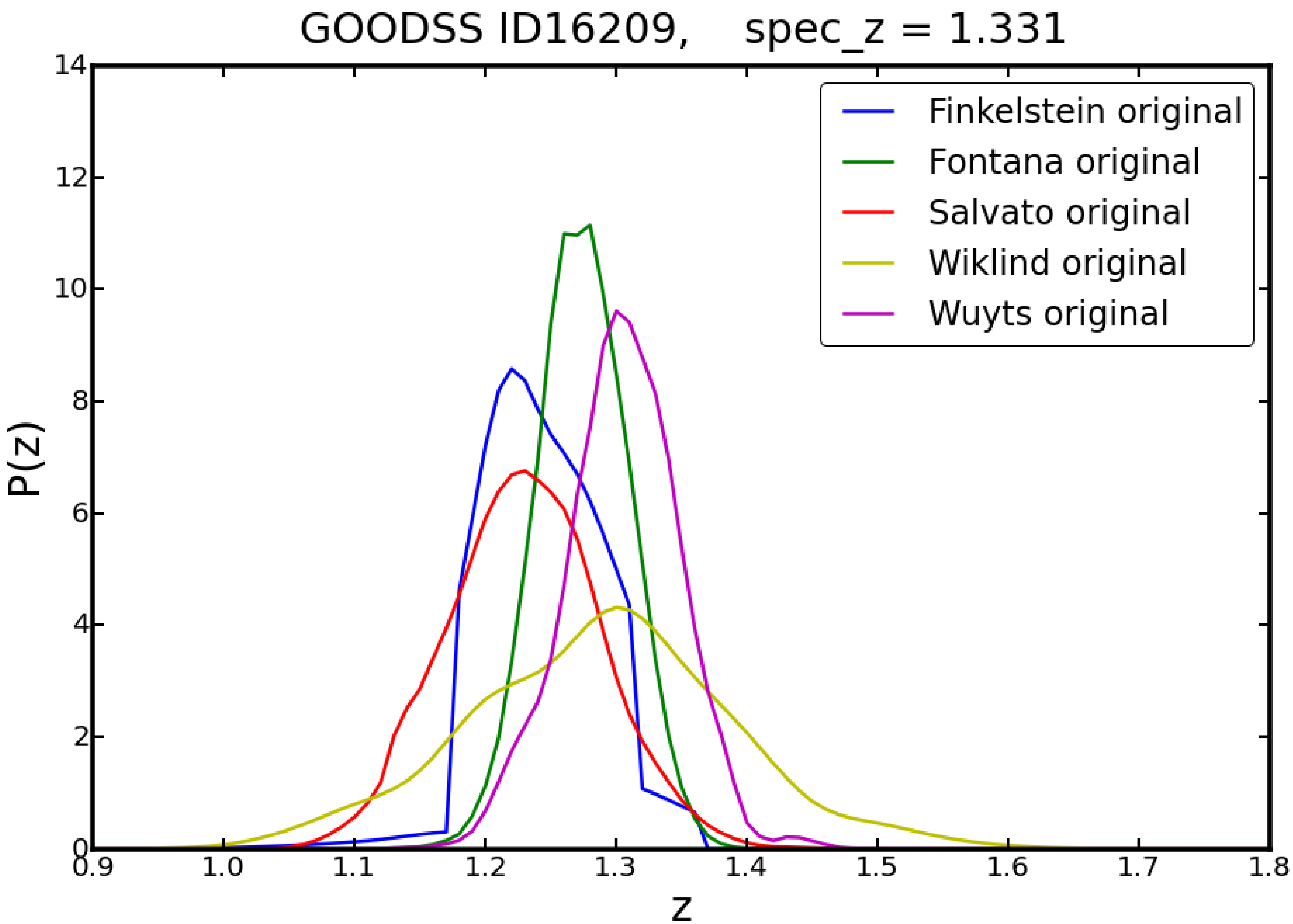
Open issues: Combining PDF results from multiple codes

- Dahlen et al. presented a hierarchical Bayesian combination method (cf. Press & Kochanek, Lang & Hogg, etc.)
- Izbicki & Lee 2016 use weighted combinations of codes
- Kodra et al. (in prep) investigates using PDF that minimizes total Fréchet distance to remaining PDFs: analogous to median



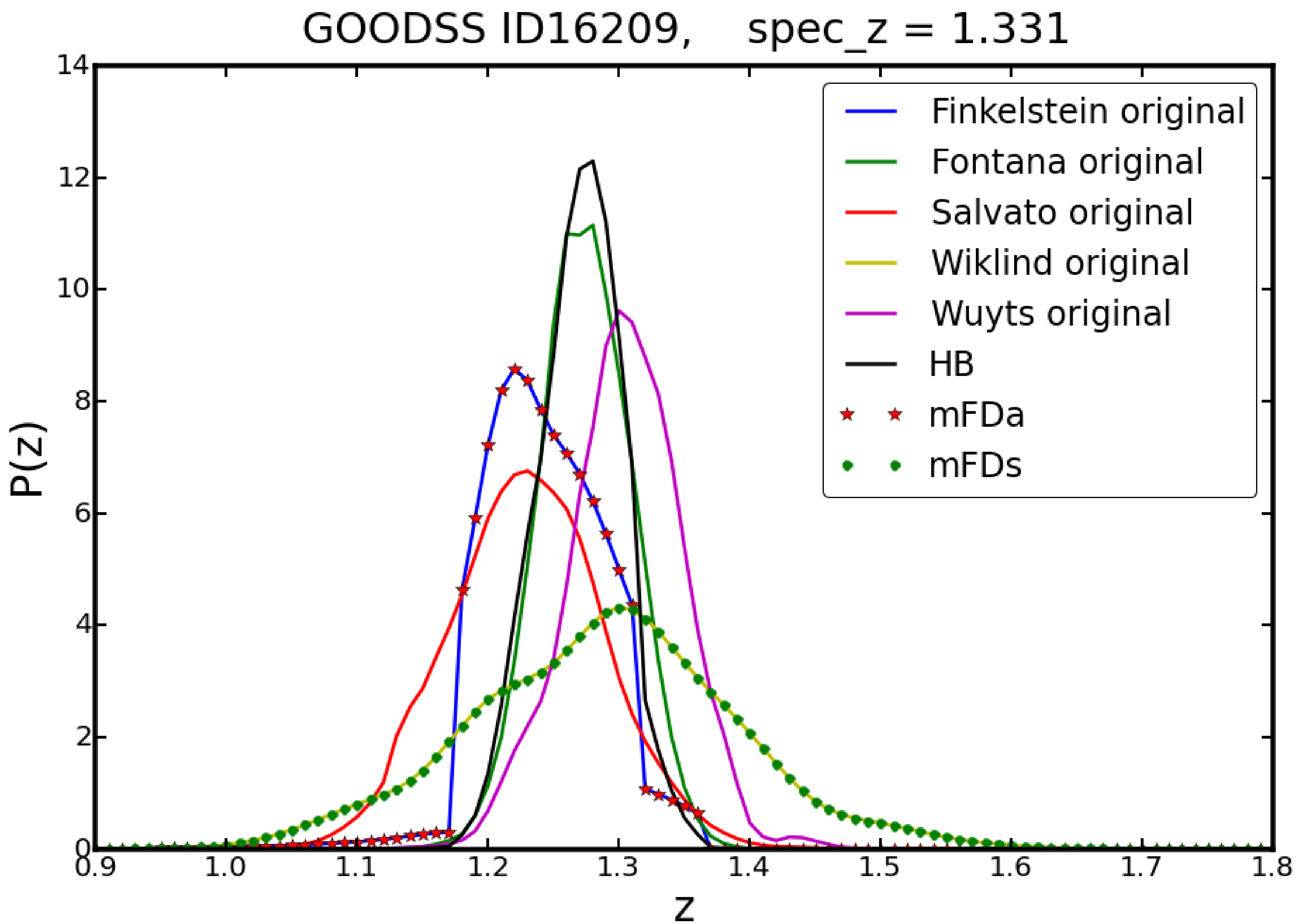
D. Kodra

Open issues: Combining PDF results from multiple codes



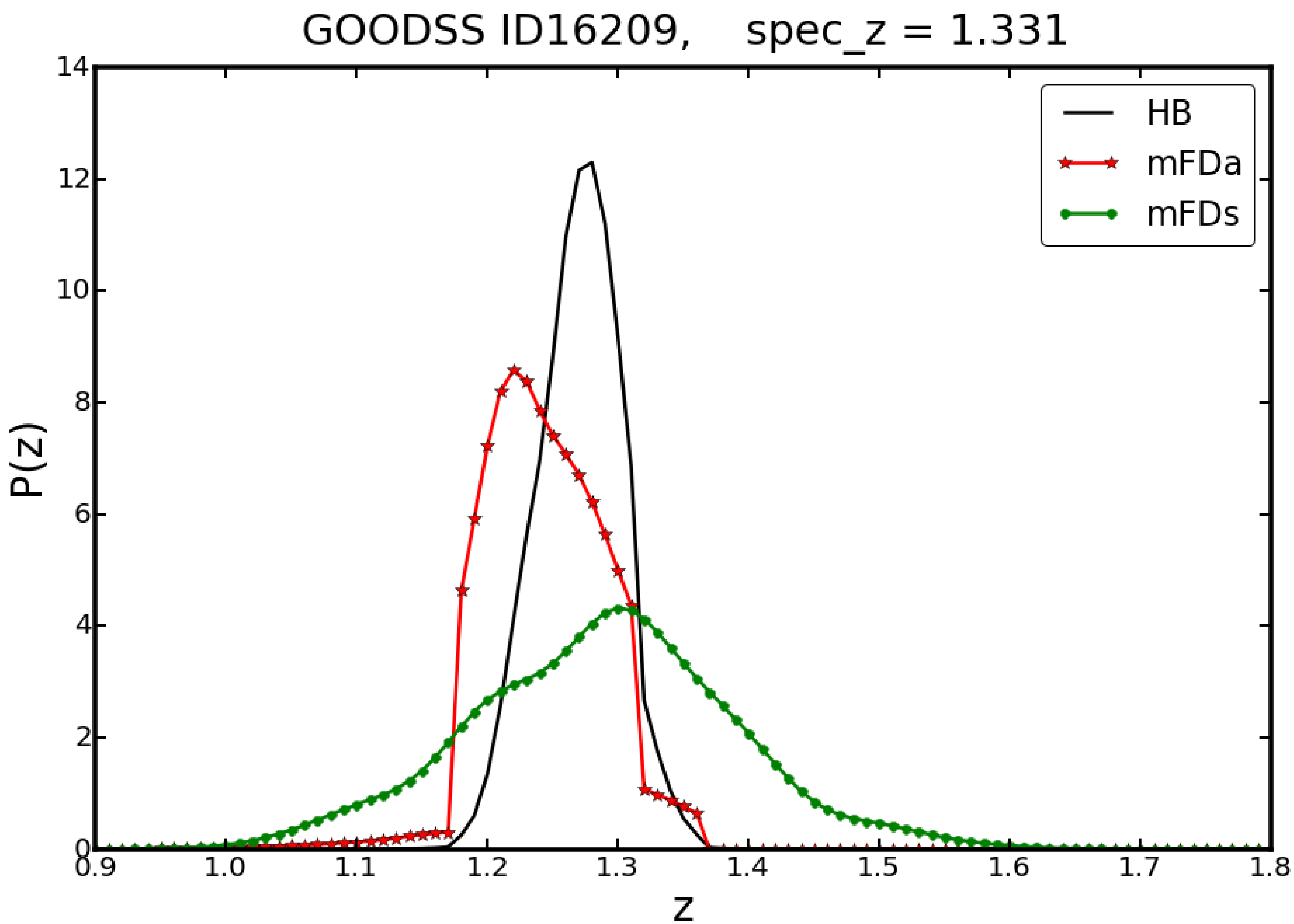
D. Kodra

Open issues: Combining PDF results from multiple codes



D. Kodra

Open issues: Combining PDF results from multiple codes

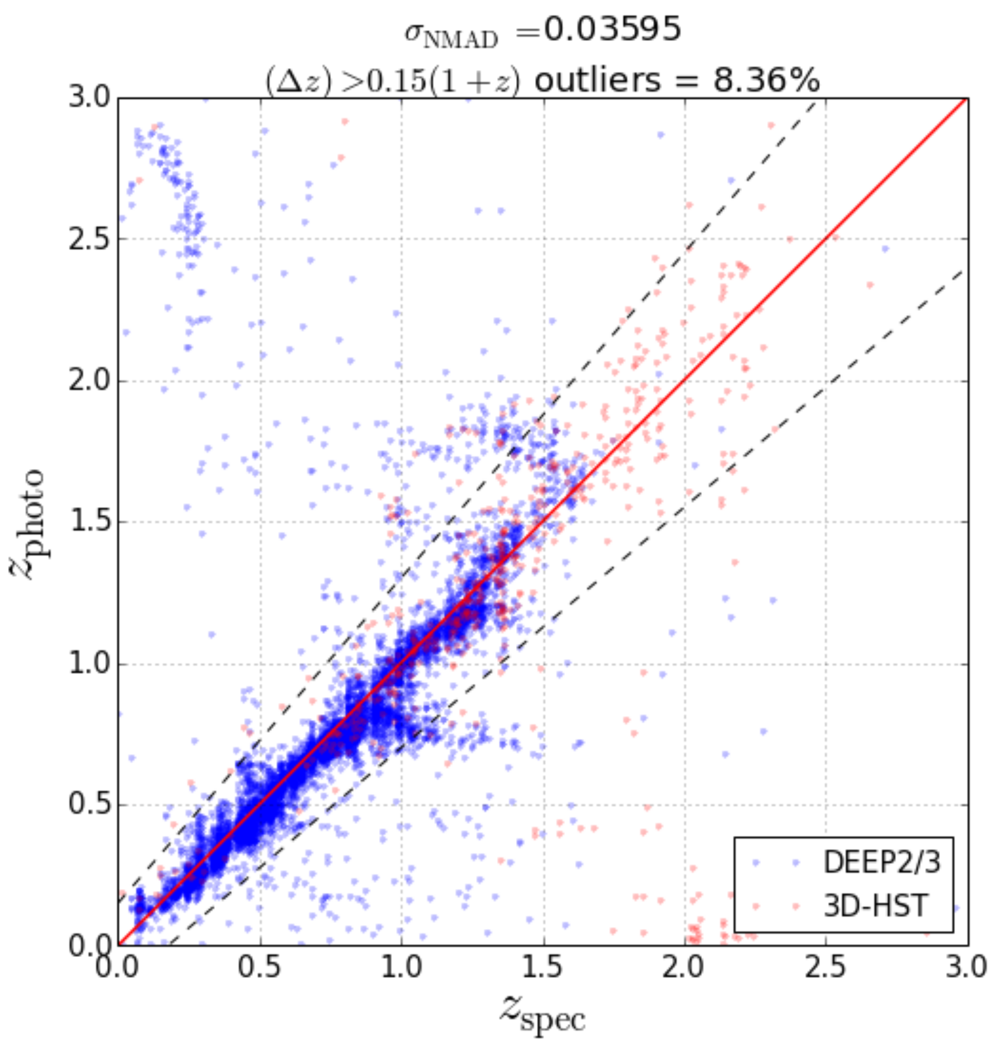


D. Kodra

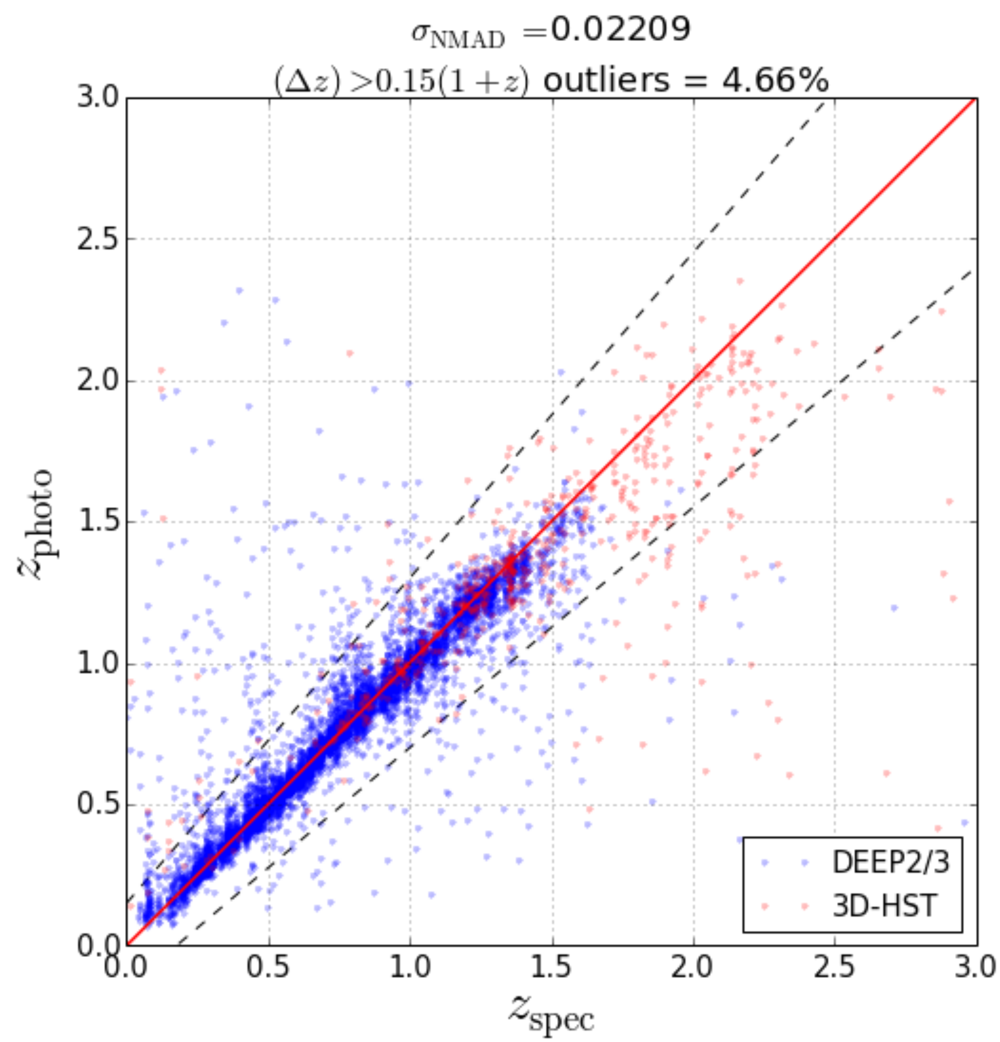
Another possible use case: template-based and training-based methods have different failure modes

- **Identify potential outliers from discrepant results?**

EAZY (template code, untuned)



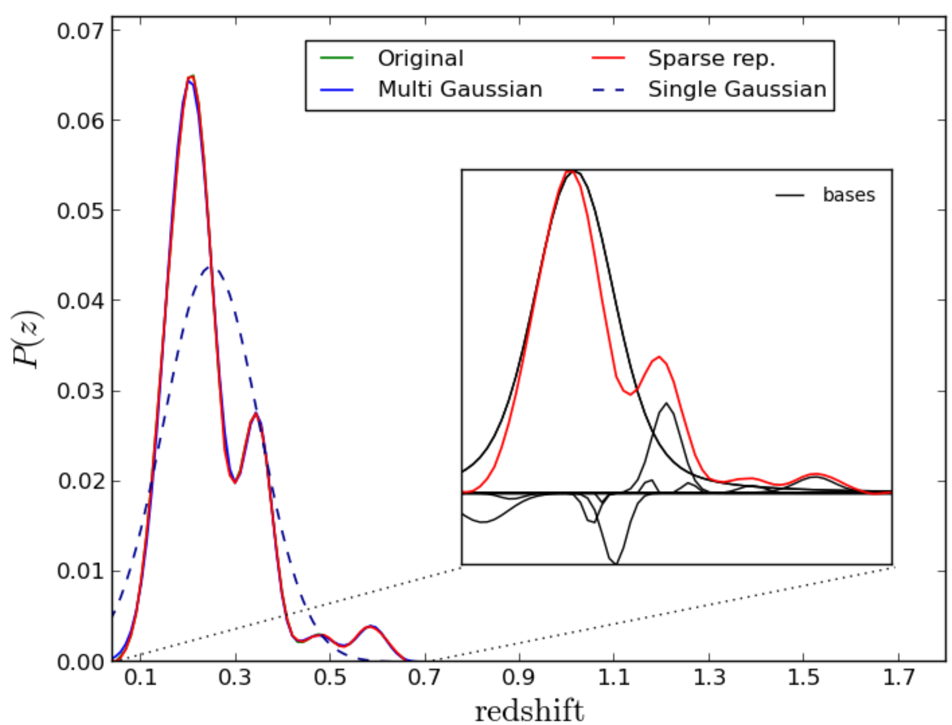
Random Forest Regression

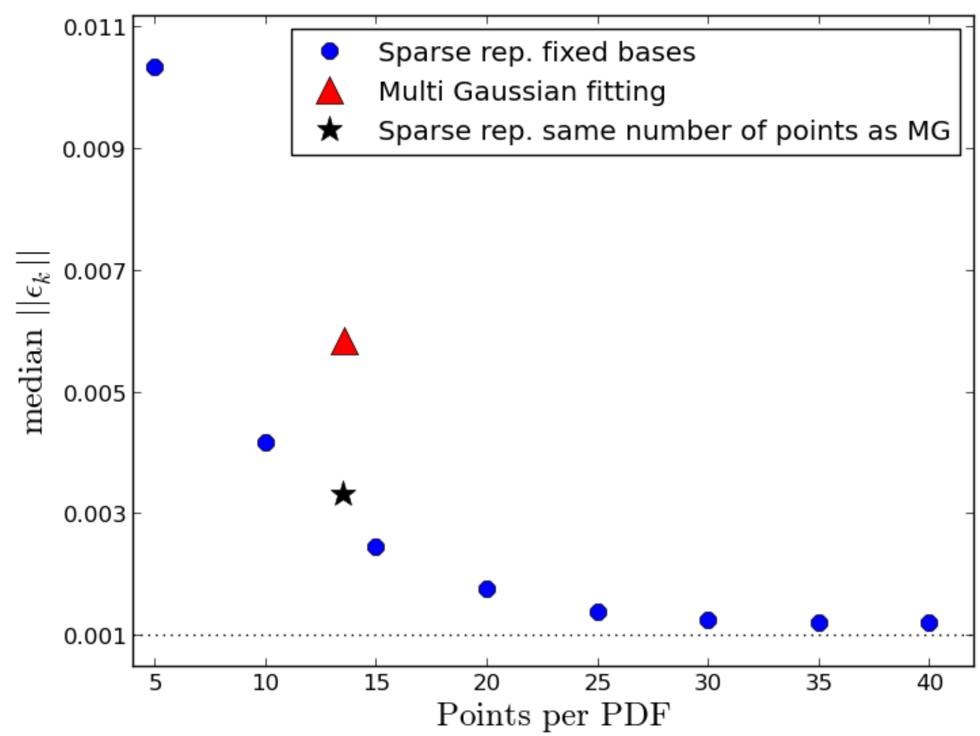


Zhou, JN et al. 2016, in prep.

Open issues: Storing p(z,α)

- Carrasco-Kind & Brunner 2014 achieved strong compression of photo-z PDFs using sparse representation and well-chosen basis set
- For many LSST applications, want 2+-dimensional PDFs
- Can suitably sparse (<few hundred #s) representations be achieved?
- Are samples from PDFs OK for all science cases?

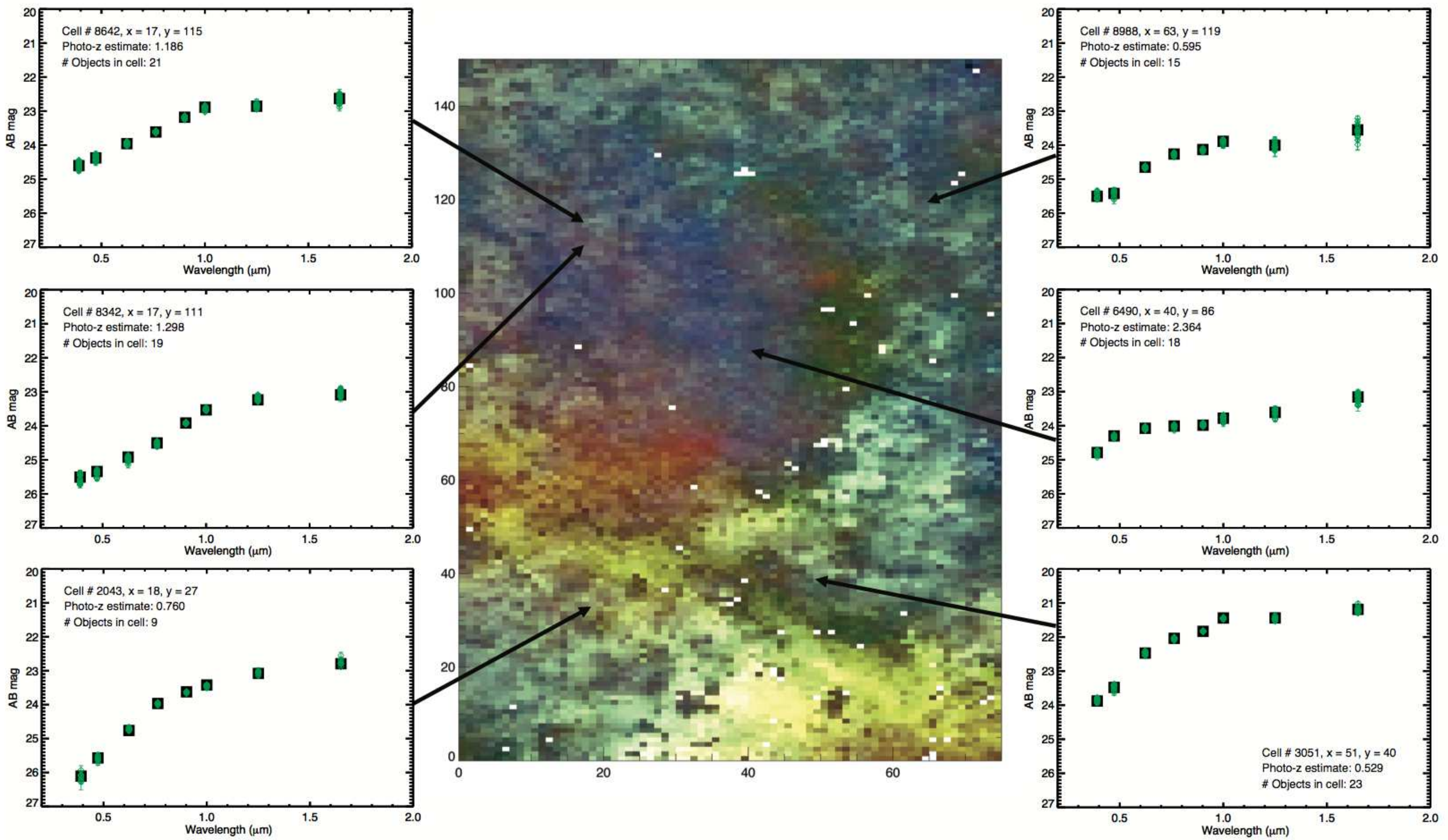




Carrasco-Kind & Brunner 2014

Open issues: Optimizing spectroscopic targeting

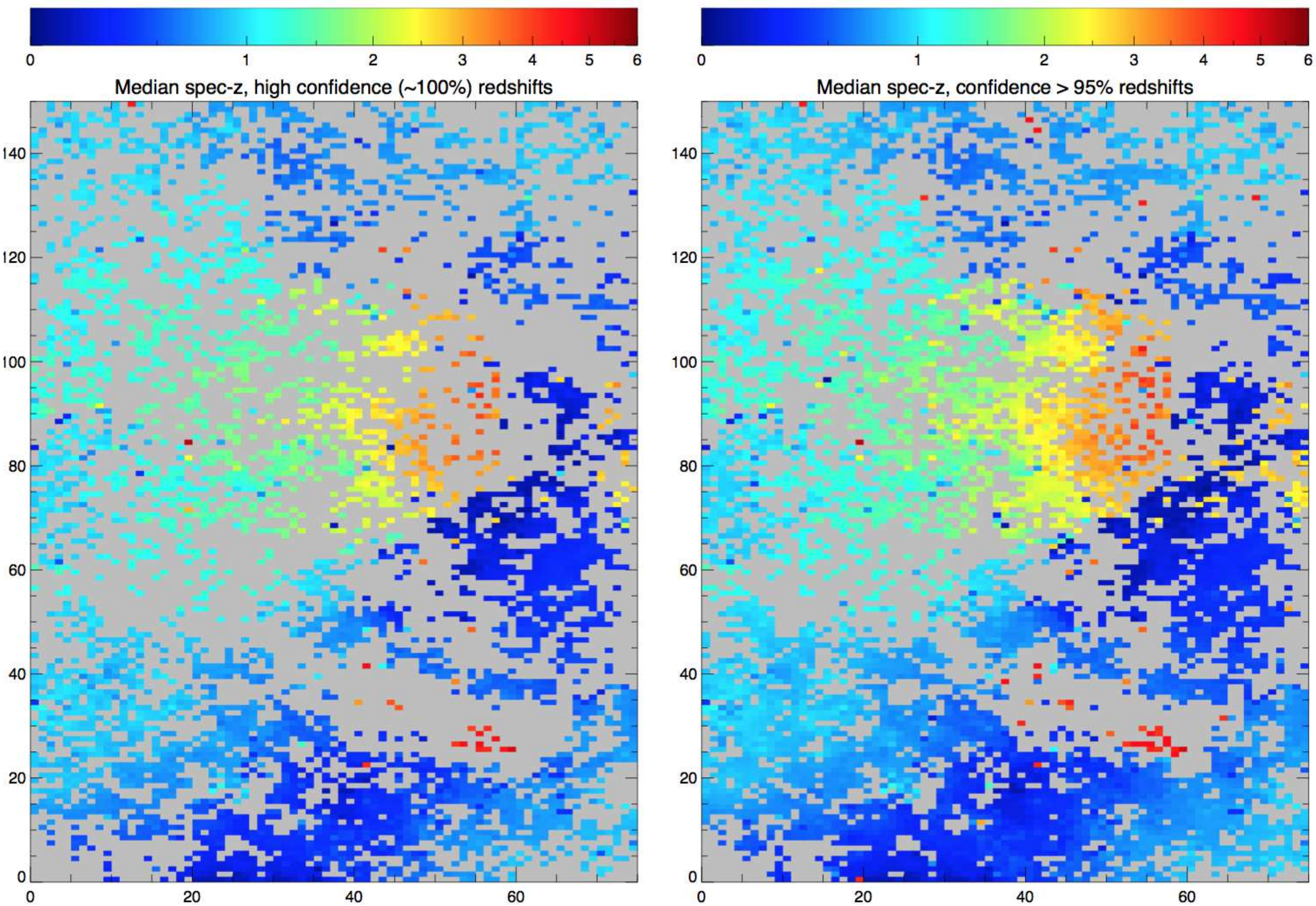
- **Current state of the art: Masters et al. 2015**
- **Self-organized map of galaxy colors**



Masters et al. 2015

Open issues: Optimizing spectroscopic targeting

- **Prioritize cells with few redshifts for spectroscopic follow-up**
- **Are there better ways to do this?**



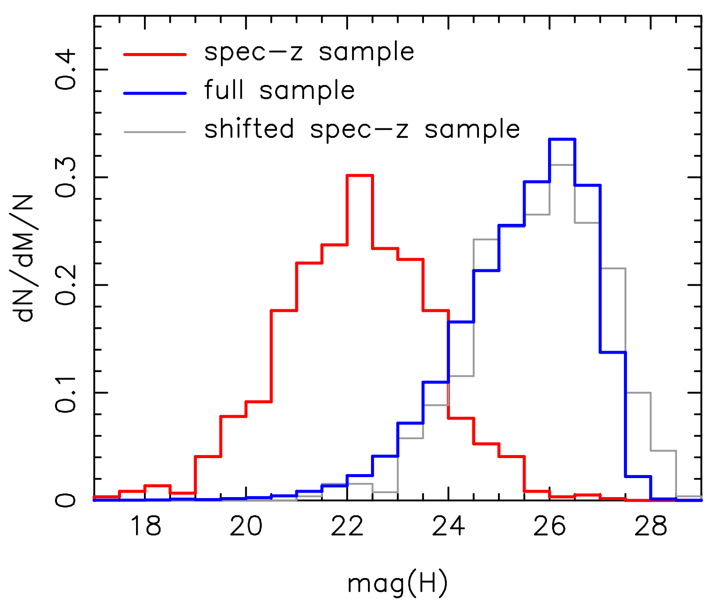
Masters et al. 2015

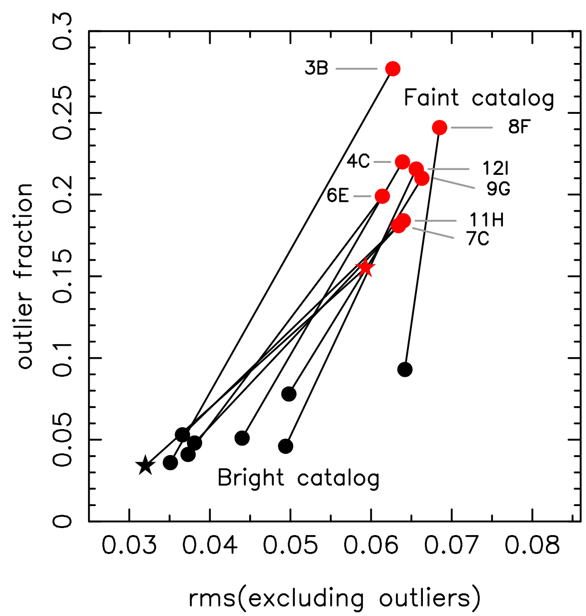
Open issues: Ideal photo-z code?

- What might an ideal LSST photo-z algorithm look like?
  - Trained with >30,000 spectra spanning range of spectra
  - Develops priors & tweaks templates via hierarchical Bayesian hyperparameters
  - Incorporates variations in effective filter wavelengths due to observational conditions: requires applying algorithm to O(1000) measurements instead of O(6)
  - Incorporates AGN classification and AGN photo-z determination: colors are not constant with time for many objects!
  - Want algorithms to be fast: create ML-based emulators for template photo-z's?
  - For bright objects, may also be useful to compare to ML techniques to identify potential outliers

CANDELS code comparison: Dahlen et al. 2013

- **Many tests of photo-z algorithms with deep, high-redshift dataset. Examples:**
  - **Test photo-z performance as degrade photometry (using same test spectroscopic data)**
  - **Dependence of errors on redshift, magnitude, & color**
  - **Investigation of (lack of) consistency between photometric zero point shifts from different codes**
  - **Empirical test of photo-z errors using Δz between close pairs**



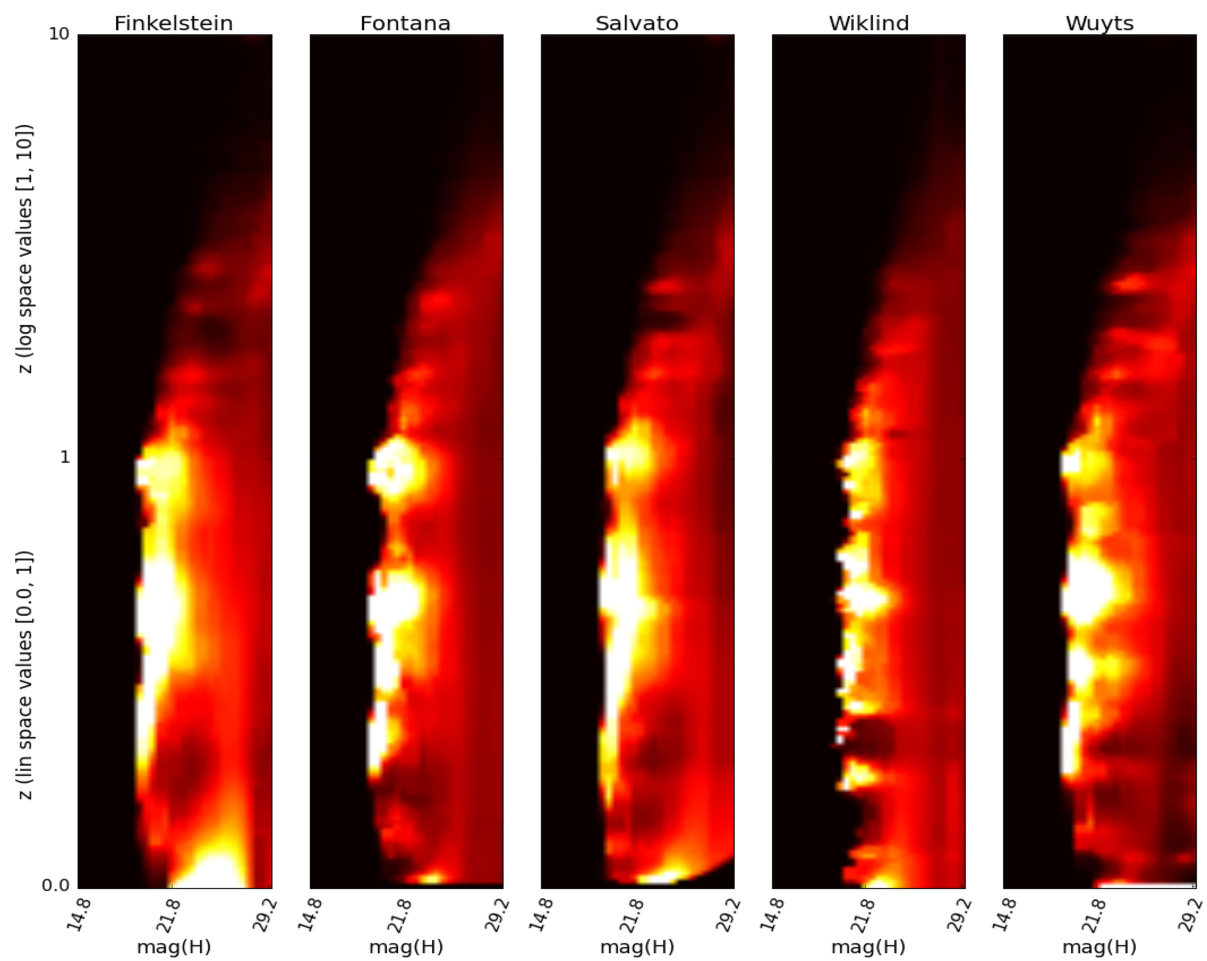


Dahlen et al. 2013

New work: Kodra et al. 2016

- Compare predictions of codes in space of p(z | H)

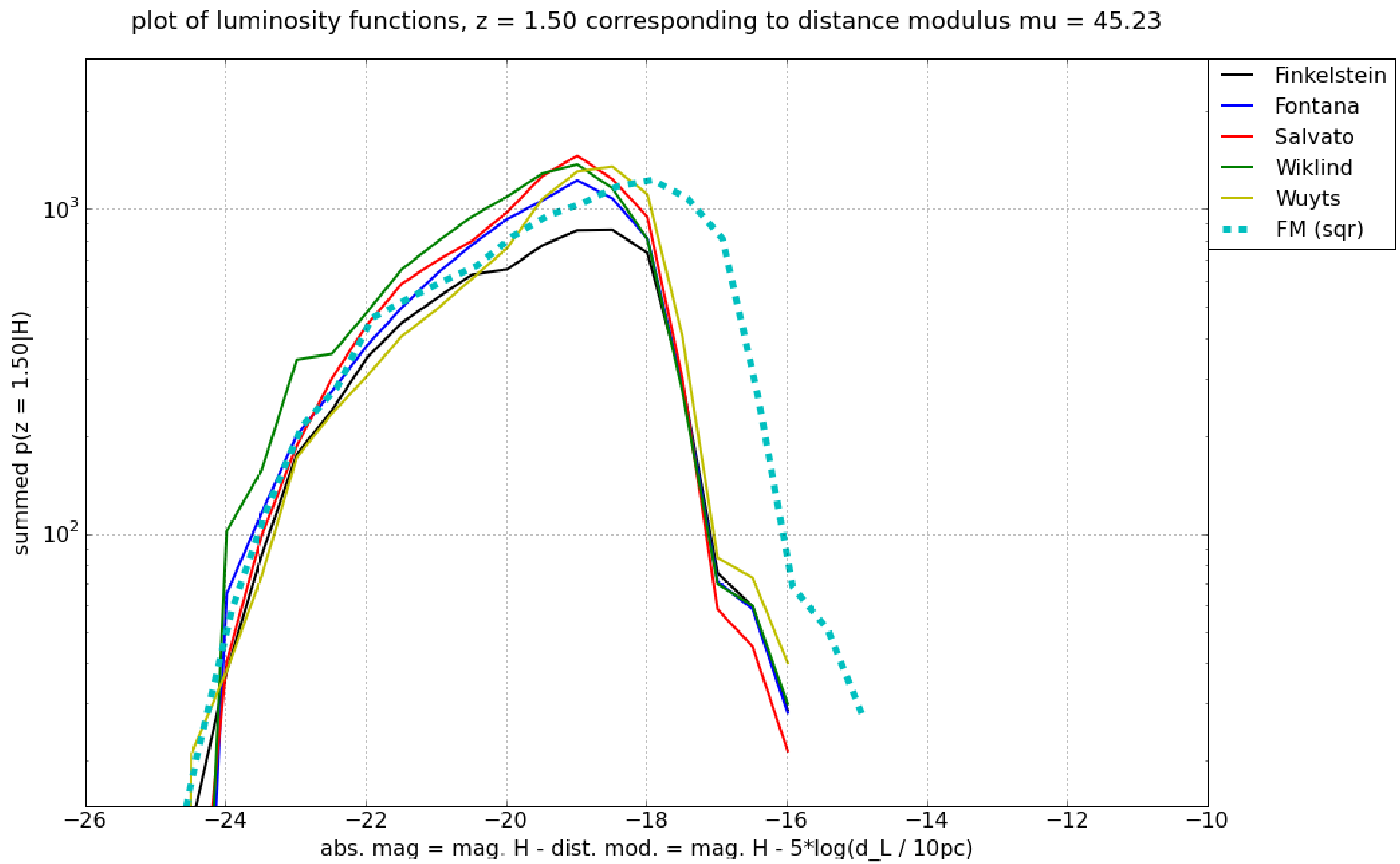
- Disagreement on where there are redshift spikes
- Priors have huge effect at low z (non-monotonic behavior)
- Different effective smoothings
- The performance of these codes for z_{peak} isn't all that different. . .



D. Kodra

New work: Kodra et al. 2016

- **This can have large (factor of few) effects on the inferred number of objects at a given redshift**



D. Kodra

Conclusions

- Training-based methods are easier to get good results from than template-based methods, but don't extrapolate well
- Key issue for LSST is inability to get complete training sets
- Other interesting issues for the next few years:
  - Training algorithms in the presence of false redshifts
  - Making sure p(z)'s have proper coverage
  - Combining results from multiple algorithms
  - Storing multi-dimensional PDFs compactly
  - Optimizing spectroscopic follow-up
  - Defining parameters of ideal LSST algorithm
- Current codes appear sufficient to meet LSST requirements, but are clearly suboptimal. Better photo-z's will greatly increase the value of LSST - e.g. 30% increase in Dark Energy Figure of Merit