Photometric redshift statistics

or how to avoid biases in cosmological contraints

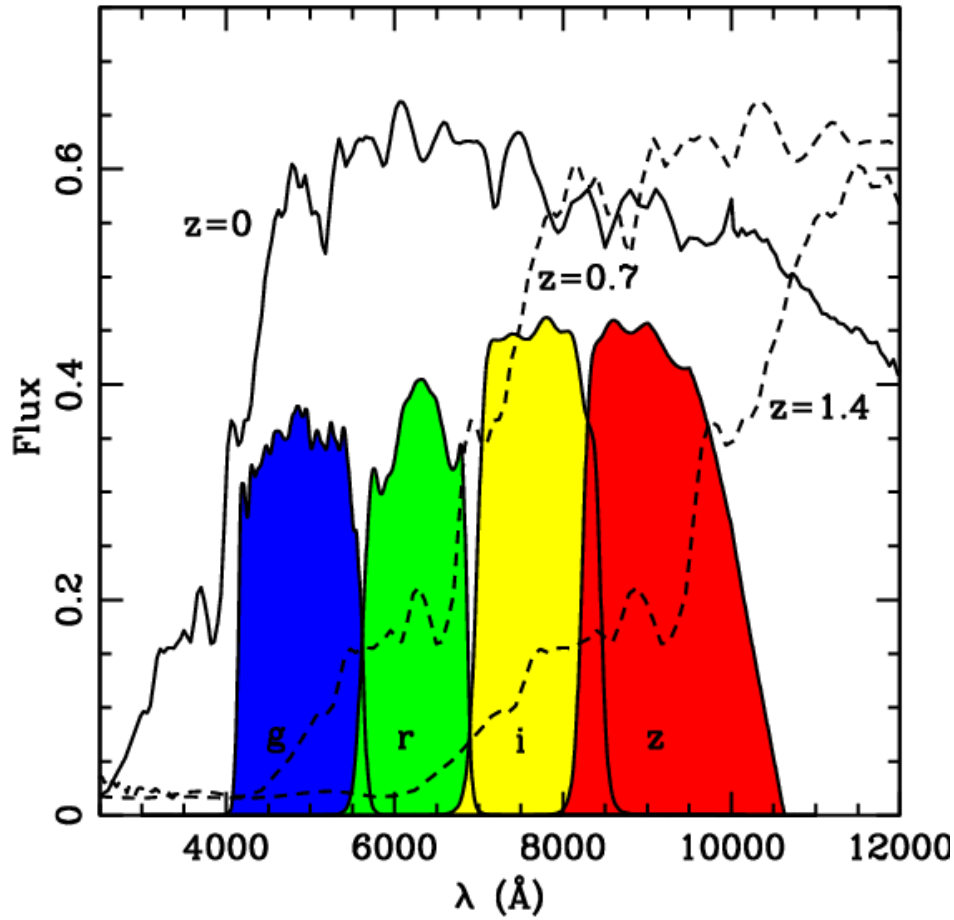
Carlos Cunha

University of Michigan

Great Lakes Cosmology Workshop
University of Chicago, June 16th, 2010

Basics of photo-z's

- Probe strong spectral features (4000 Å break)
- Flux in each filter depends on galaxy's type and redshift.

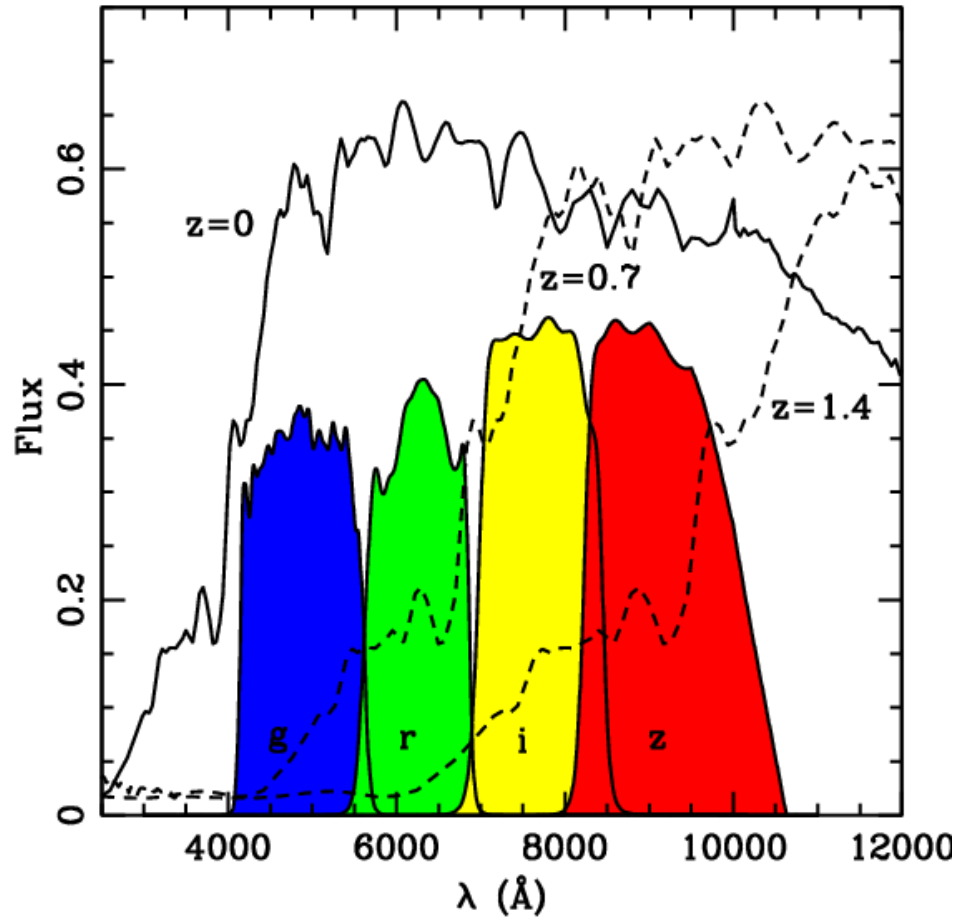


Basics of photo-z's

Two classes of methods:

- **Template-fitting:** compare observed fluxes with predicted fluxes from library of galaxy spectra.

- **Training set:** use subsample with known redshifts to "train" flux-redshift relation.



What **not** to do (but is often done)

- **Training set methods:** determine functional relation between photometric observables, m, and z_{phot} using a training set

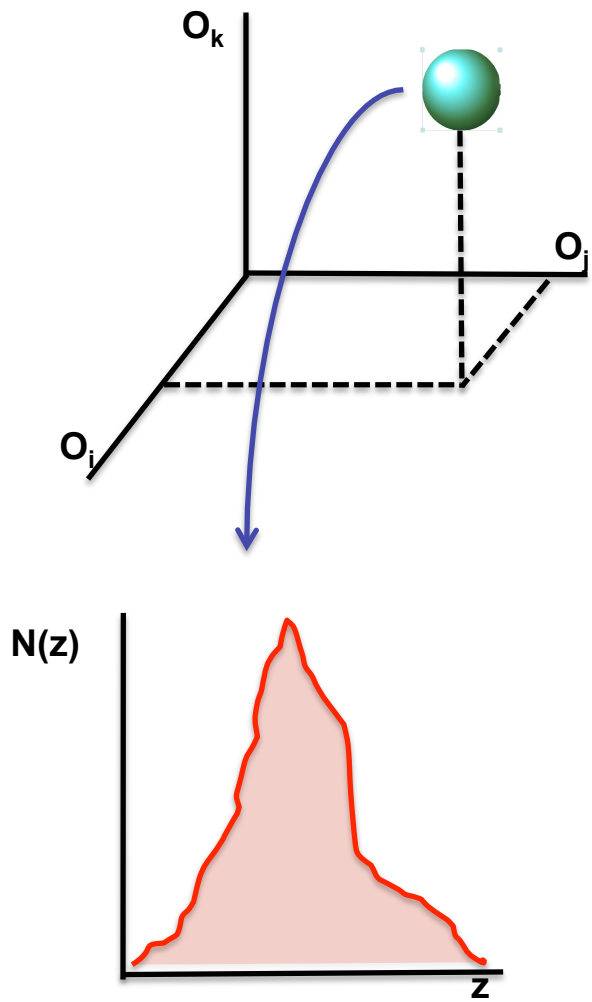
$$z_{phot} = z_{phot}(m)$$

- **Template-fitting methods:** choose the photo-z to be the peak of the likelihood (or posterior) distribution.

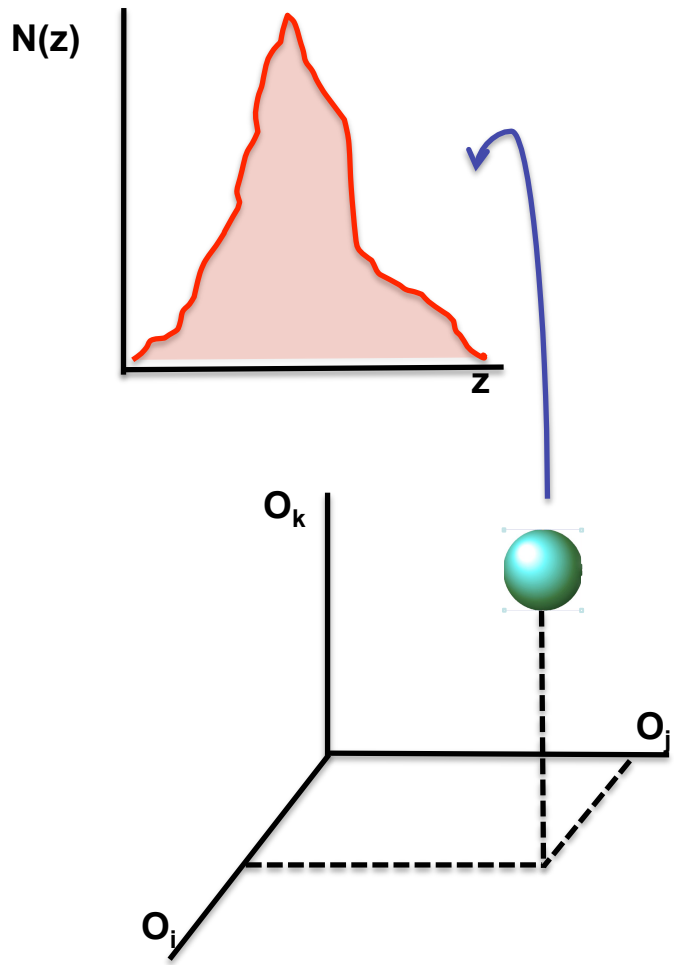
Why not?

Each region of observable space (fluxes, colors) is occupied by galaxies at a broad range of redshifts.

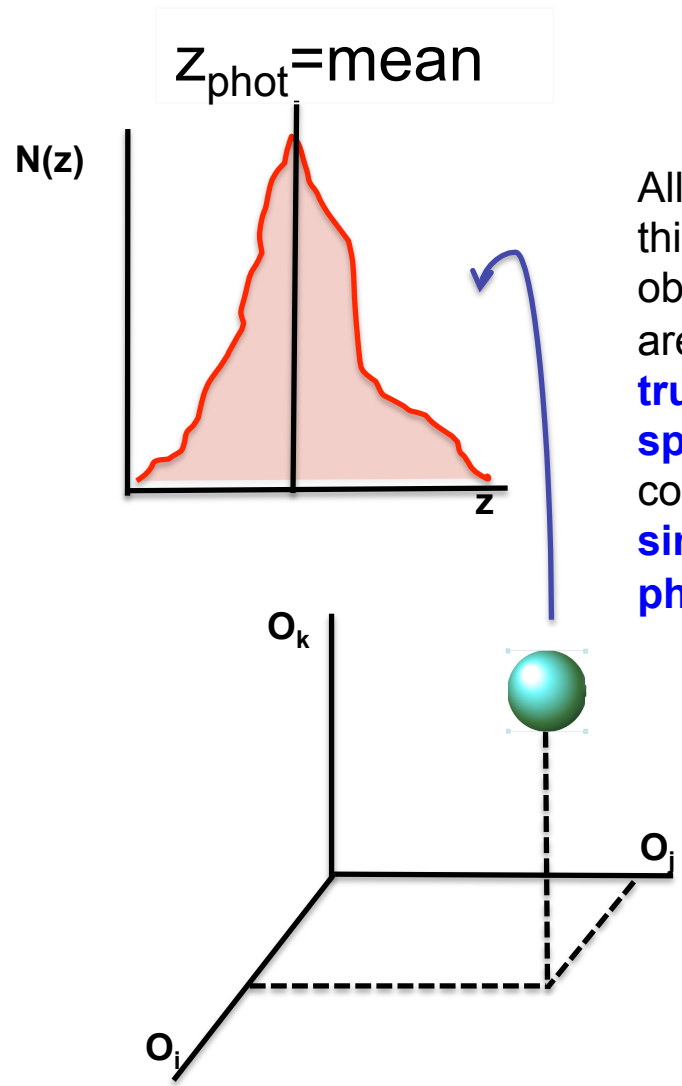
Representing these broad **redshift distributions** by a single number (the peak, mean, …) is the main source of **bias** in photometric redshifts.



The origin of the photo-z bias

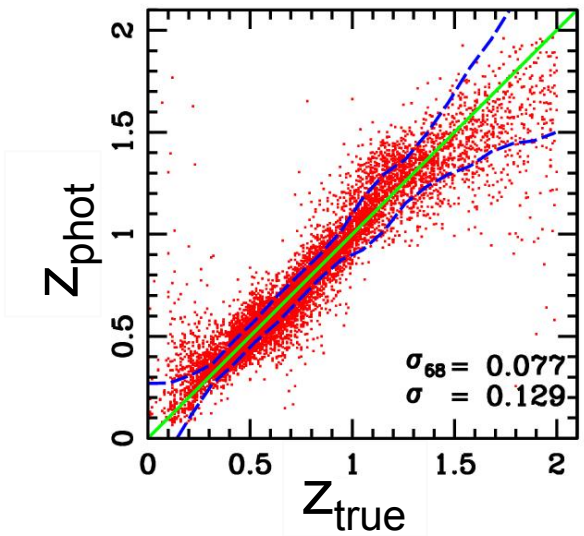


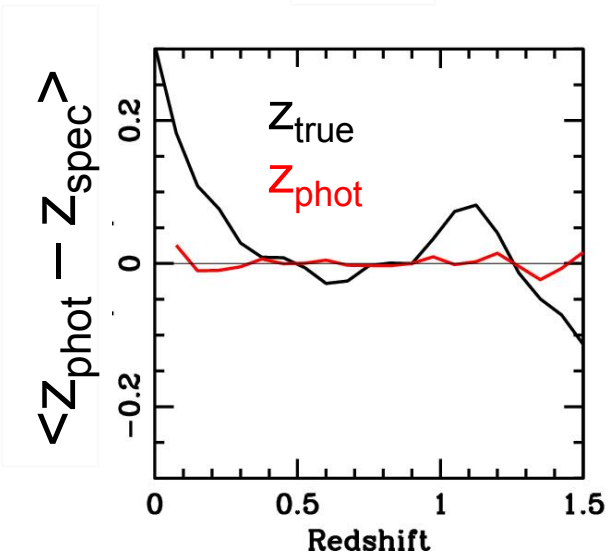
The origin of the photo-z bias

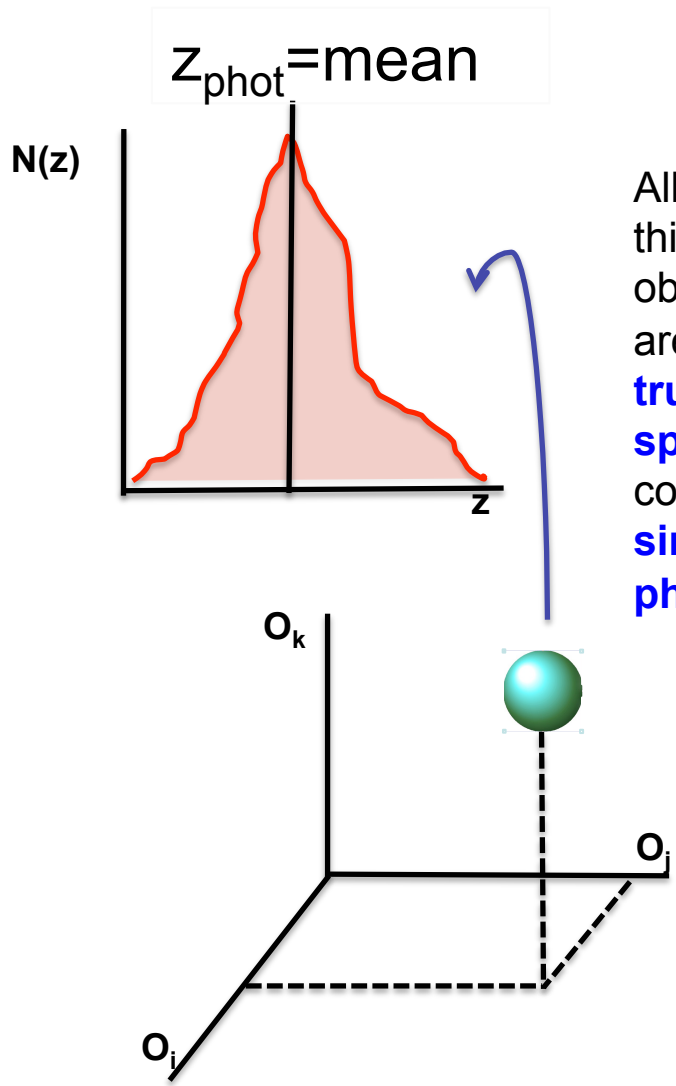


All objects in this region of observable space are **spread out in true redshift space**, but are concentrated in a **single point in photo-z space**.

The origin of the photo-z bias

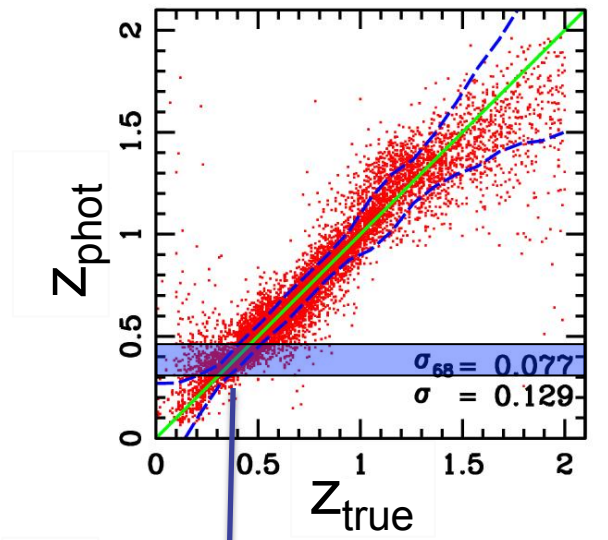


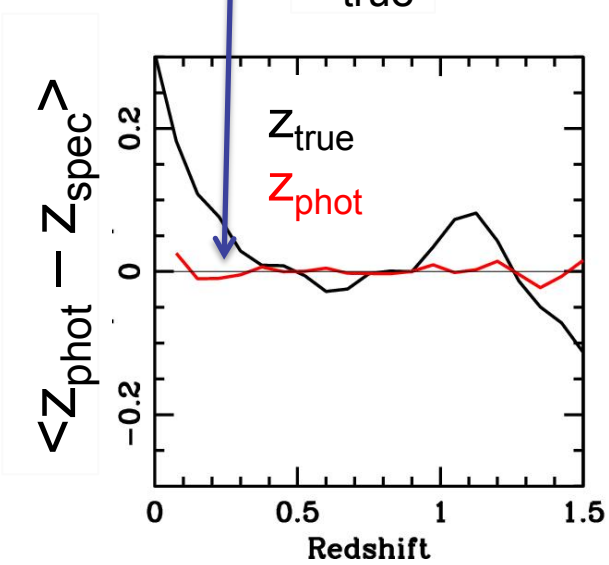


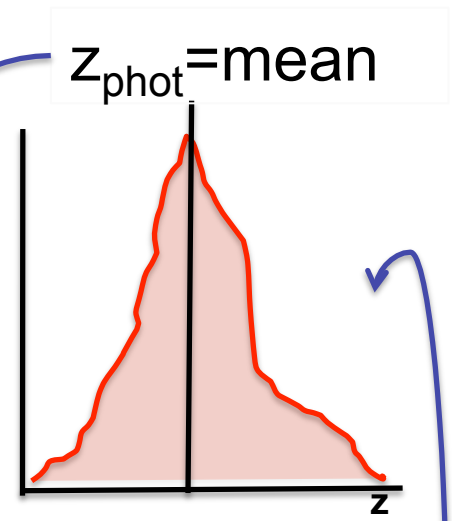


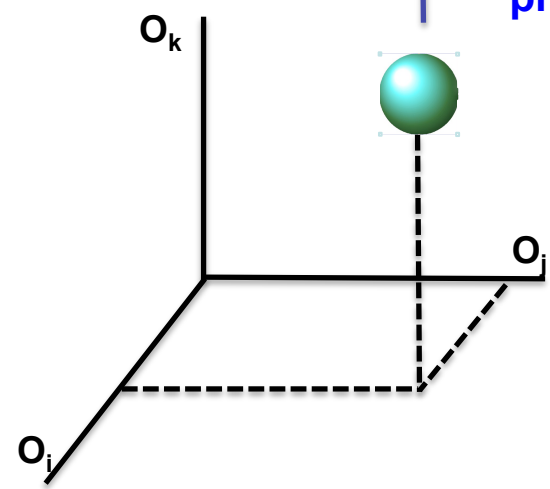
All objects in this region of observable space are **spread out in true redshift space**, but are concentrated in a **single point in photo-z space**.

The origin of the photo-z bias



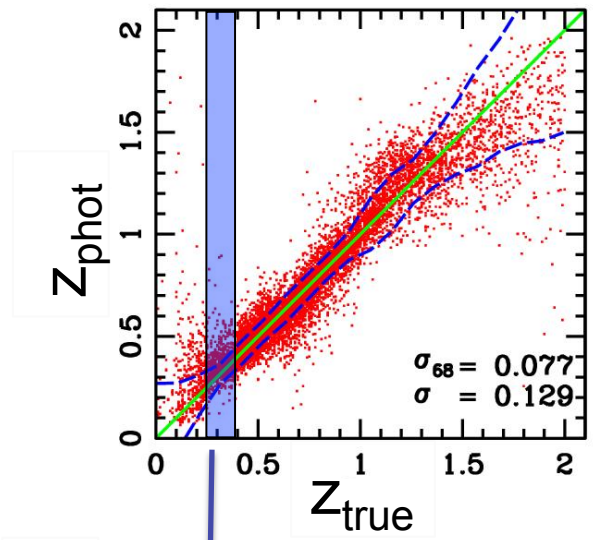


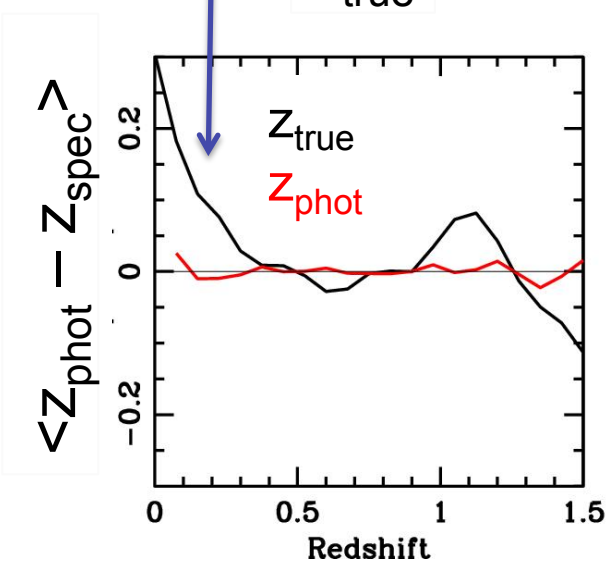


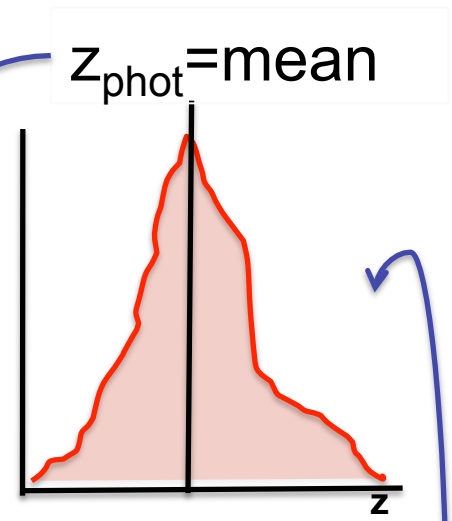


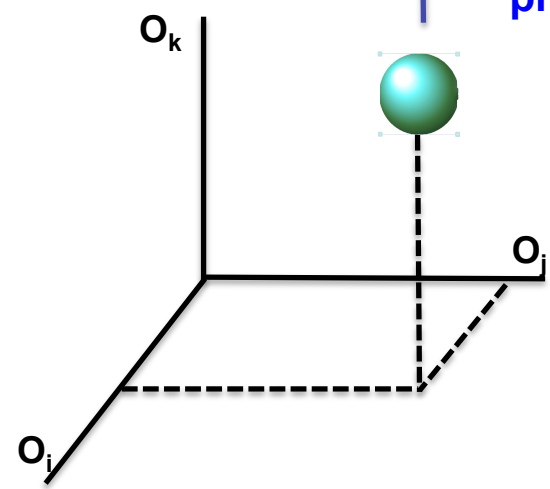
All objects in this region of observable space are **spread out in true redshift** space, but are concentrated in a **single point in photo-z space**.

The origin of the photo-z bias









All objects in this region of observable space are **spread out in true redshift** space, but are concentrated in a **single point in photo-z space**.

What to do

Use full redshift distribution, **p(z)**, for every galaxy.

How do we get p(z)'s?

Nearest neighbors are a great tool to estimate p(z)'s and have many other useful applications.

The Team:

Me (Michigan), M. Lima (UPenn), H. Oyaizu (a Big Bank in Japan), H.Lin (Fermilab), J. Frieman (Chicago, Fermilab).

Weights

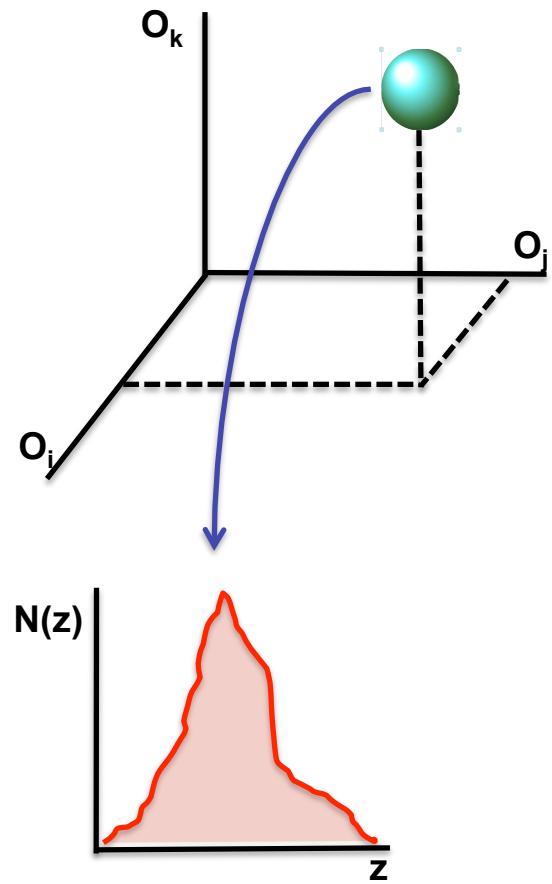
Match distributions of observables in **training** (spectroscopic or simulated) sample and **photometric** sample by assigning **weights** to training set galaxies.

$$Weight \propto \frac{\rho_{photo}}{\rho_{train}} \quad \text{where} \quad \rho_i = \frac{N_i}{V}$$

N_i : number of galaxies within ball of volume V.

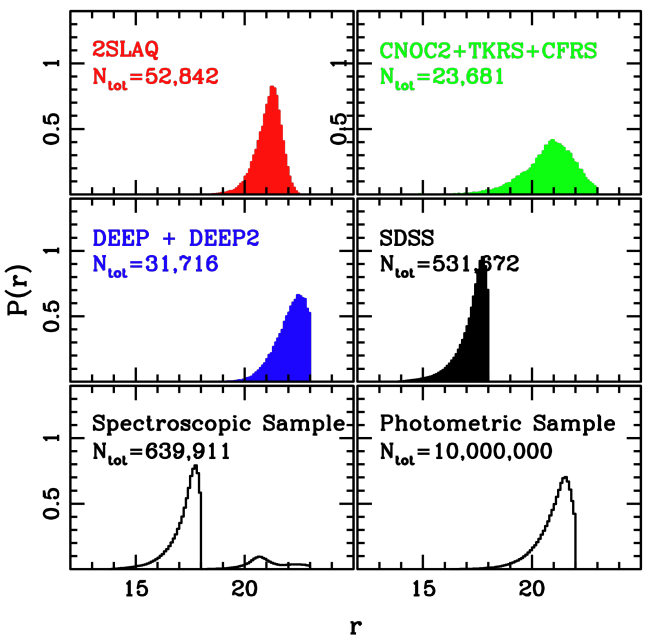
The radius of the ball is determined by the distance to 100th nearest neighbor in the training set in space of observables (colors and magnitudes).

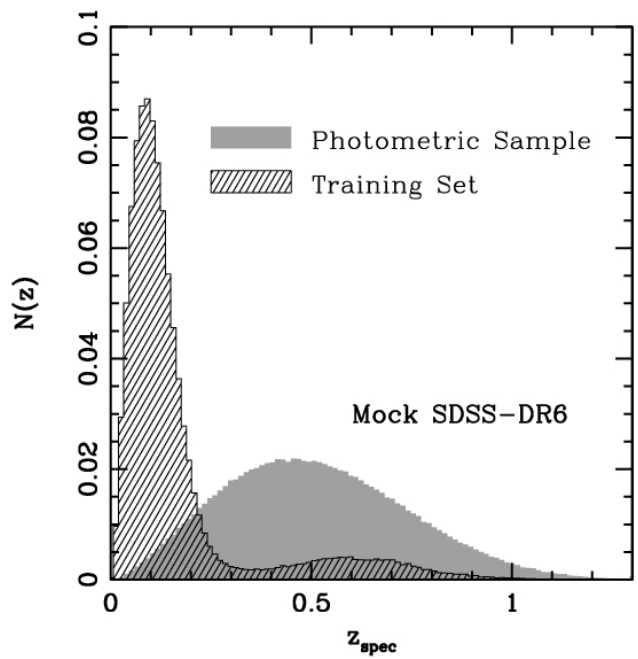
Assumption: Training set is **locally** representative of photometric set.
Is that true? Yes, **if** differences in selection are only in observable space.



SDSS Mock Catalog

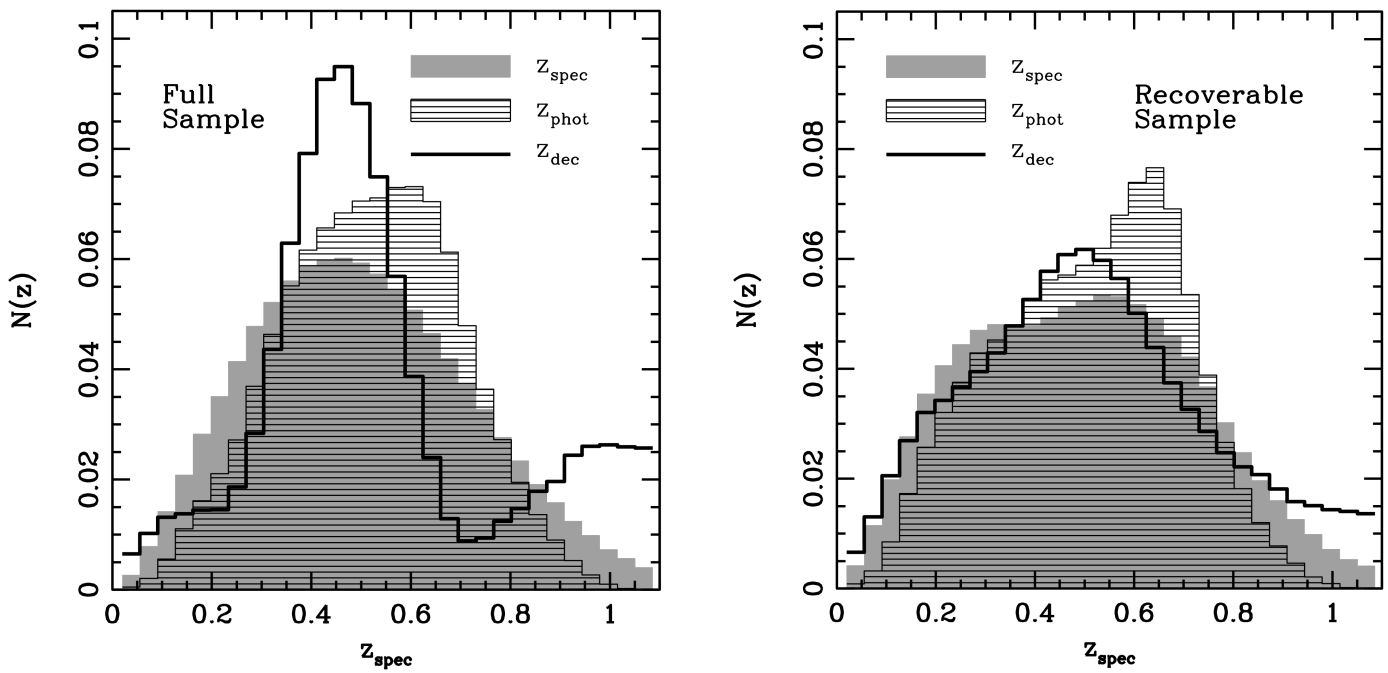
- Training composed of sets of galaxies with different color and magnitude cuts.
- Very different distributions from the photometric sample (sample for which there are no spectroscopic redshifts).





Weights applications I

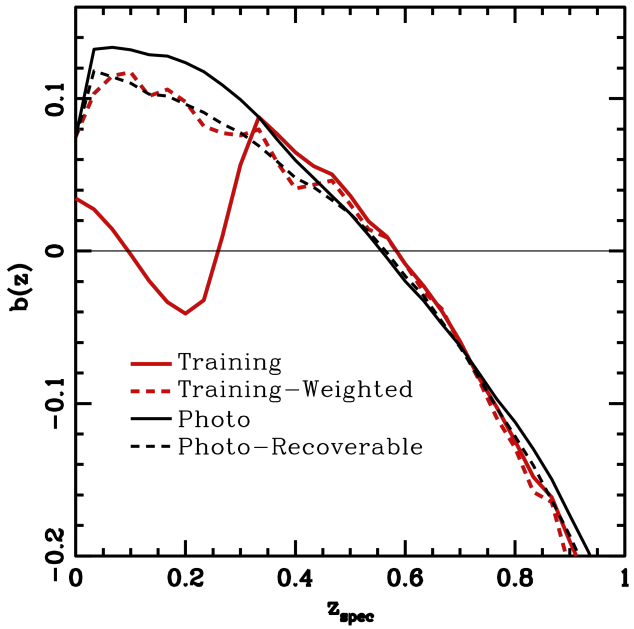
The recoverable sample:

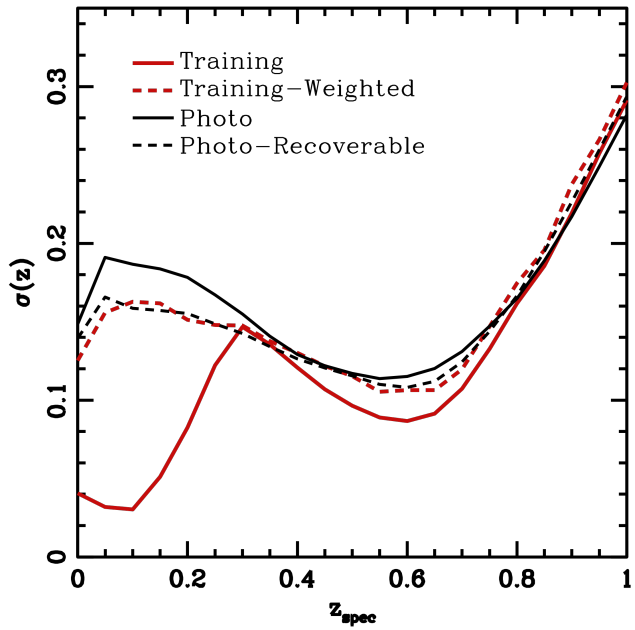


• Can only estimate photo-z's (or the redshift distribution) in regions of observable space covered by the training set, i.e. the **recoverable sample**.
• The recoverable sample is a by-product of the weights calculation.

Weights applications II

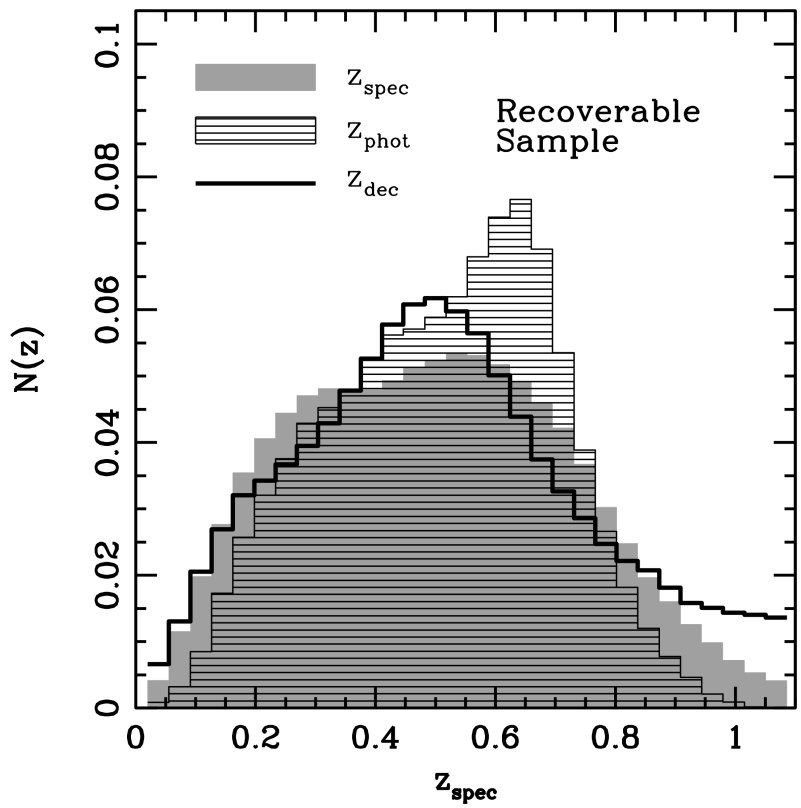
Estimating bias and scatter:

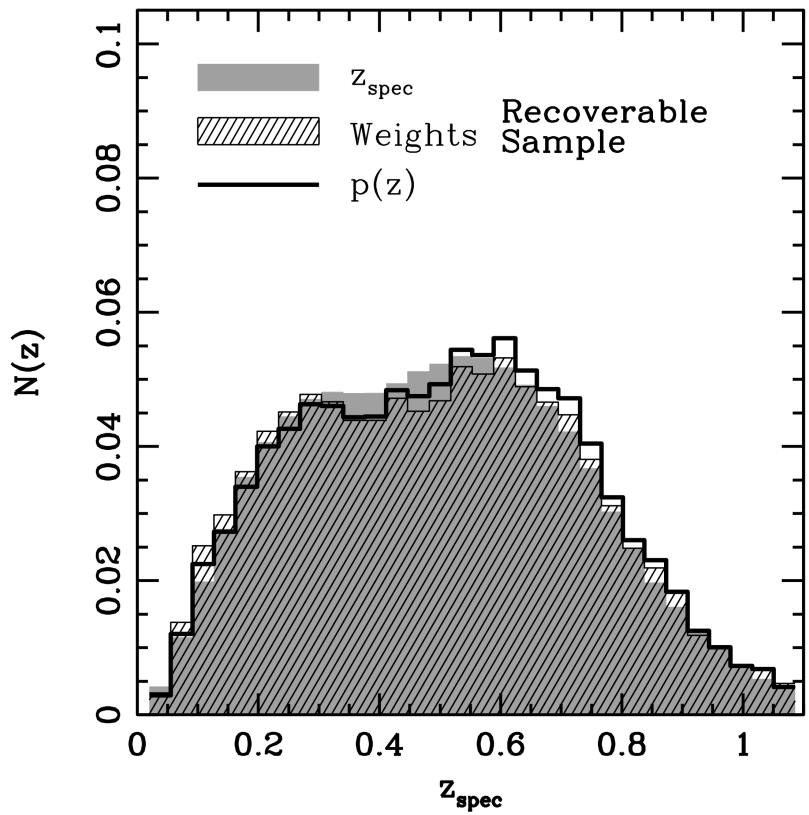




• Photo-z properties as a function of z_{spec} (or other non-observables) **must** use the weights and **only apply** to the recoverable sample.

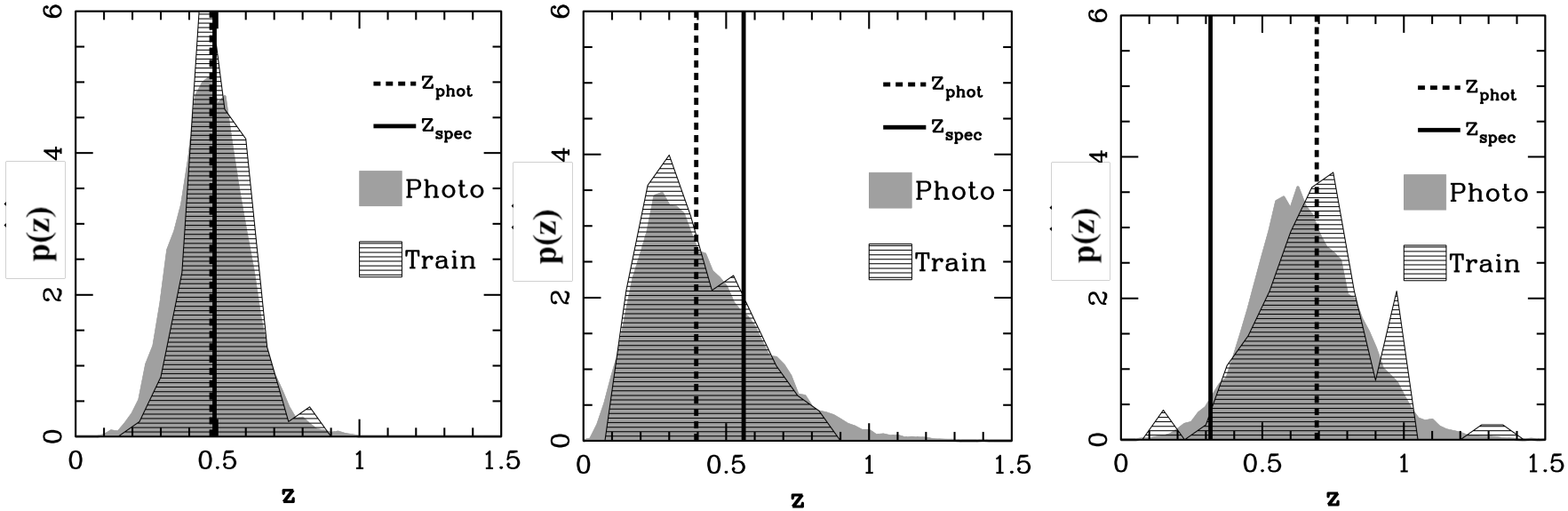
The weighted redshift distribution





Galaxy redshift probability

Examples of p(z)'s:

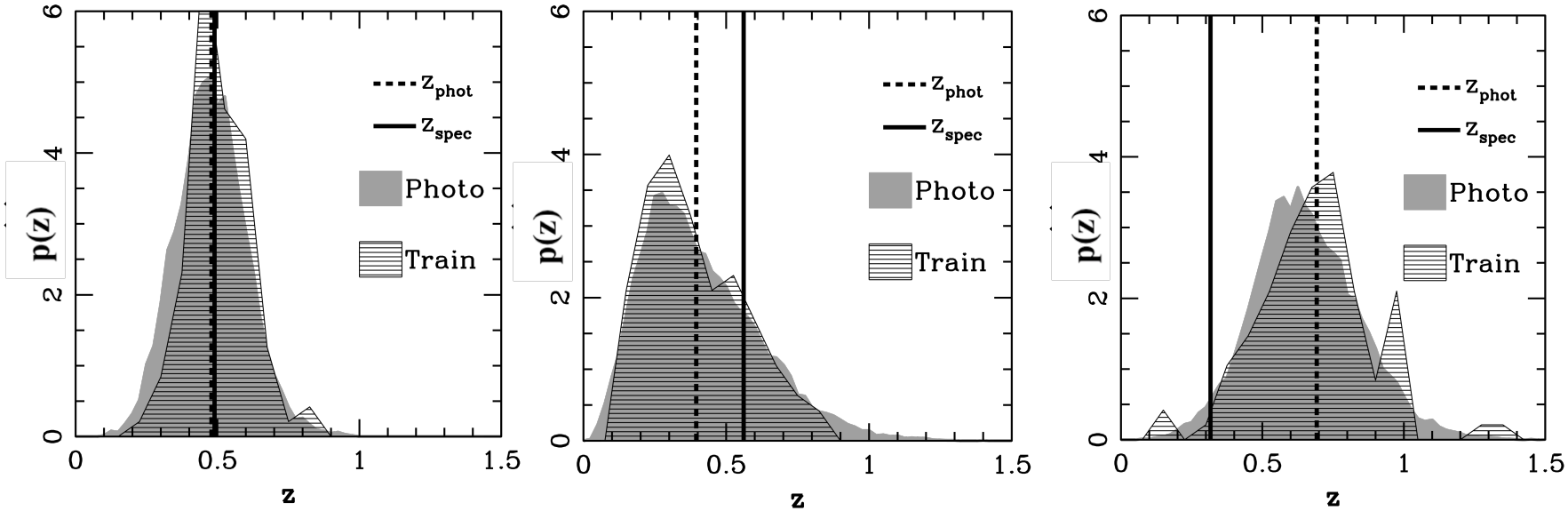


•The training does a decent job in estimating the probability for individual galaxies

• Fainter galaxies ⟶ broader probability distributions.

Galaxy redshift probability

Examples of p(z)'s:



•The training does a decent job in estimating the probability for individual galaxies

• Fainter galaxies ⟶ broader probability distributions.

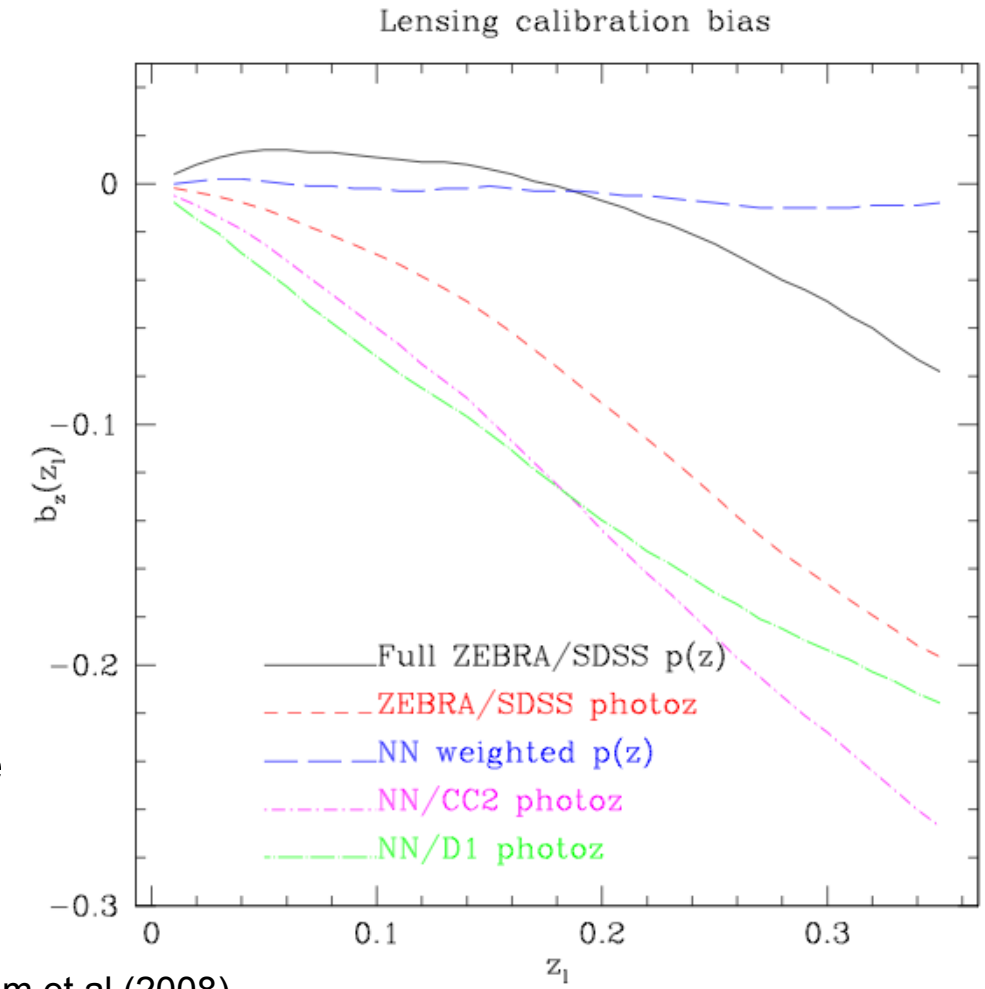
Representing p(z) by a single number generates photo-z biases!

p(z), a success story

Using p(z) eliminates lensing calibration bias, b_z.

$$b_z + 1 = \frac{\overline{\Delta\Sigma}}{\Delta\Sigma}$$

b_z: bias in differential surface density due to bias in Σ_c due to photo-z errors.



Mandelbaum et al (2008)

Conclusions

- Use p(z) not photo-z.
  - With smaller biases, degradations in constraints from uncertainties in photo-z errors are much smaller ➔ weaker requirements on calibration sets.
- Weights work well but,
- Need to worry about selection in non-observables:
  - Spectroscopic failures
  - Large-scale structure fluctuations (for pencil-beam training sets).
  - Effects are amplified by photometry errors.

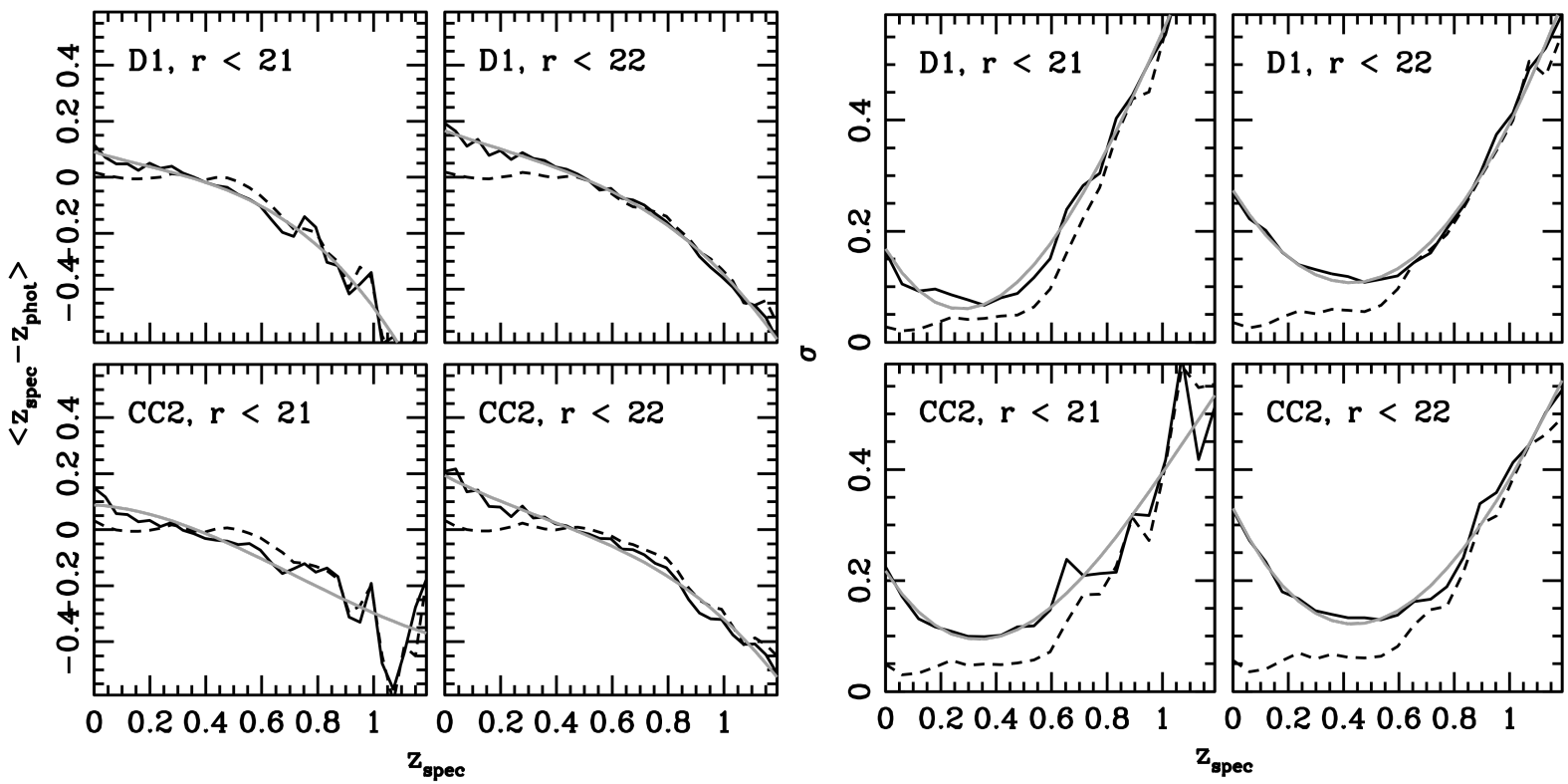
For more on weights and p(z), see Cunha et al. 2009 (arXiv:0810.2991) and Lima et al. 2009 (arXiv:0801.3822).

Public codes available at: http://kobayashi.physics.lsa.umich.edu/~ccunha/nearest/

p(z)'s for SDSS available at: http://www.sdss.org/DR7/products/value_added/index.html

Bonus slide

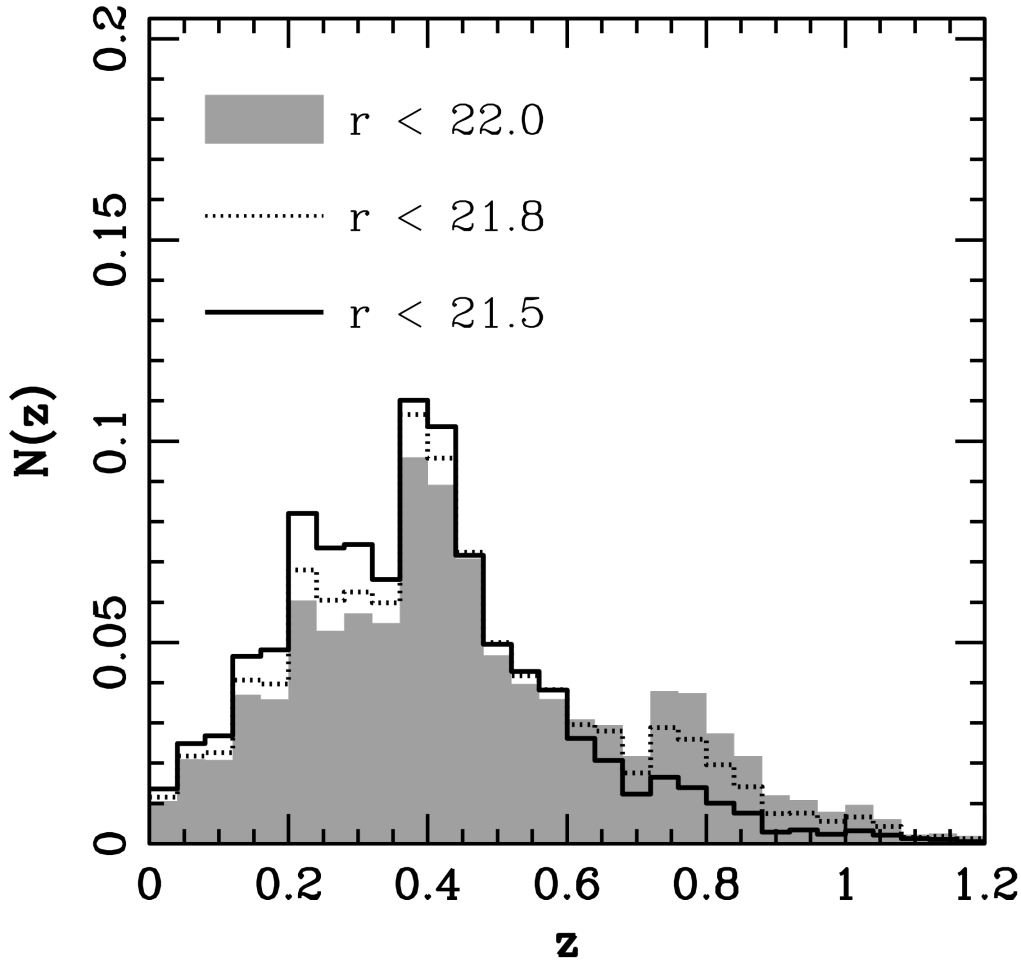
Weighted scatter and bias estimates for real SDSS photo-z's.



Cunha et al (2009)

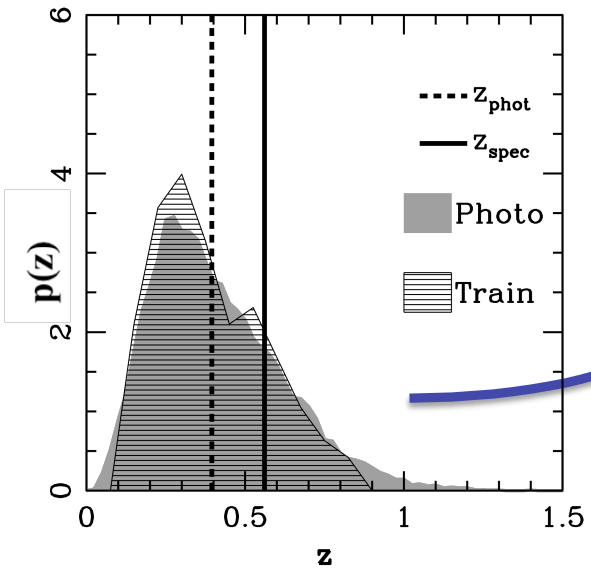
Bonus slide

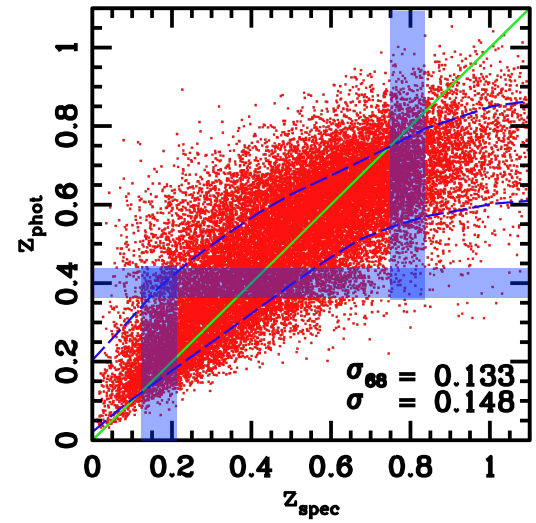
Real SDSS

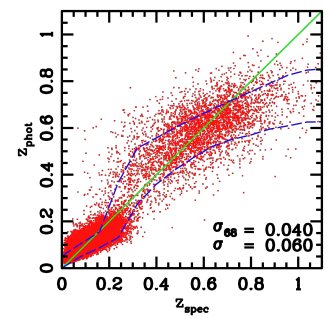


Cunha et al (2009)

Origin of the bias







All objects in this region of observable space are spread out in true redshift space, but are concentrated in a single point in photo-z space.