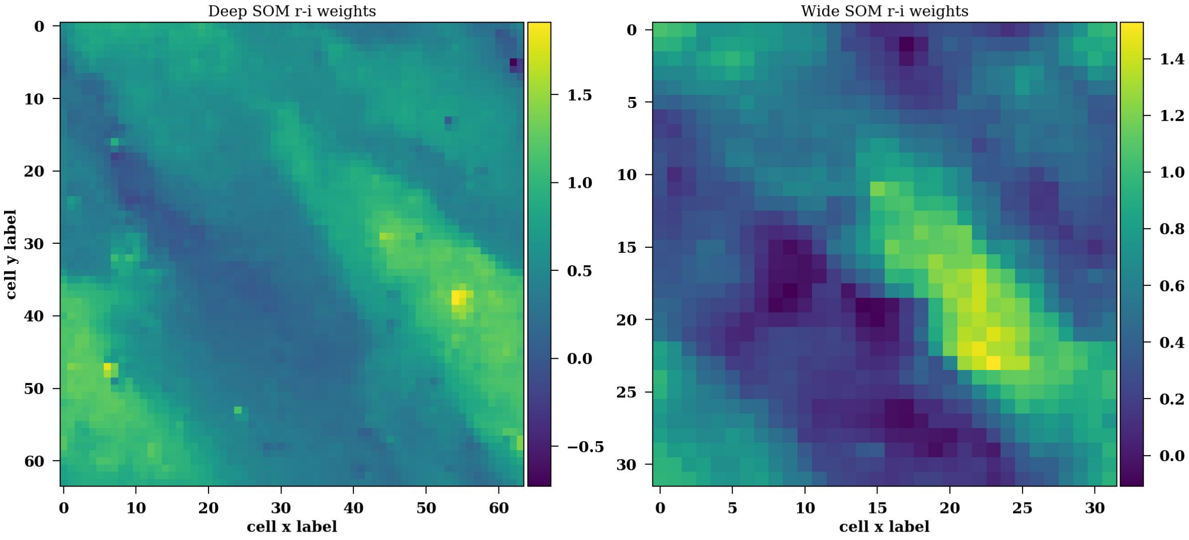
Measuring Lensing Survey Redshift Distributions with Self-Organizing Maps



Justin Myles

Advised by: Steve Allen, Alex Amon, and Daniel Gruen

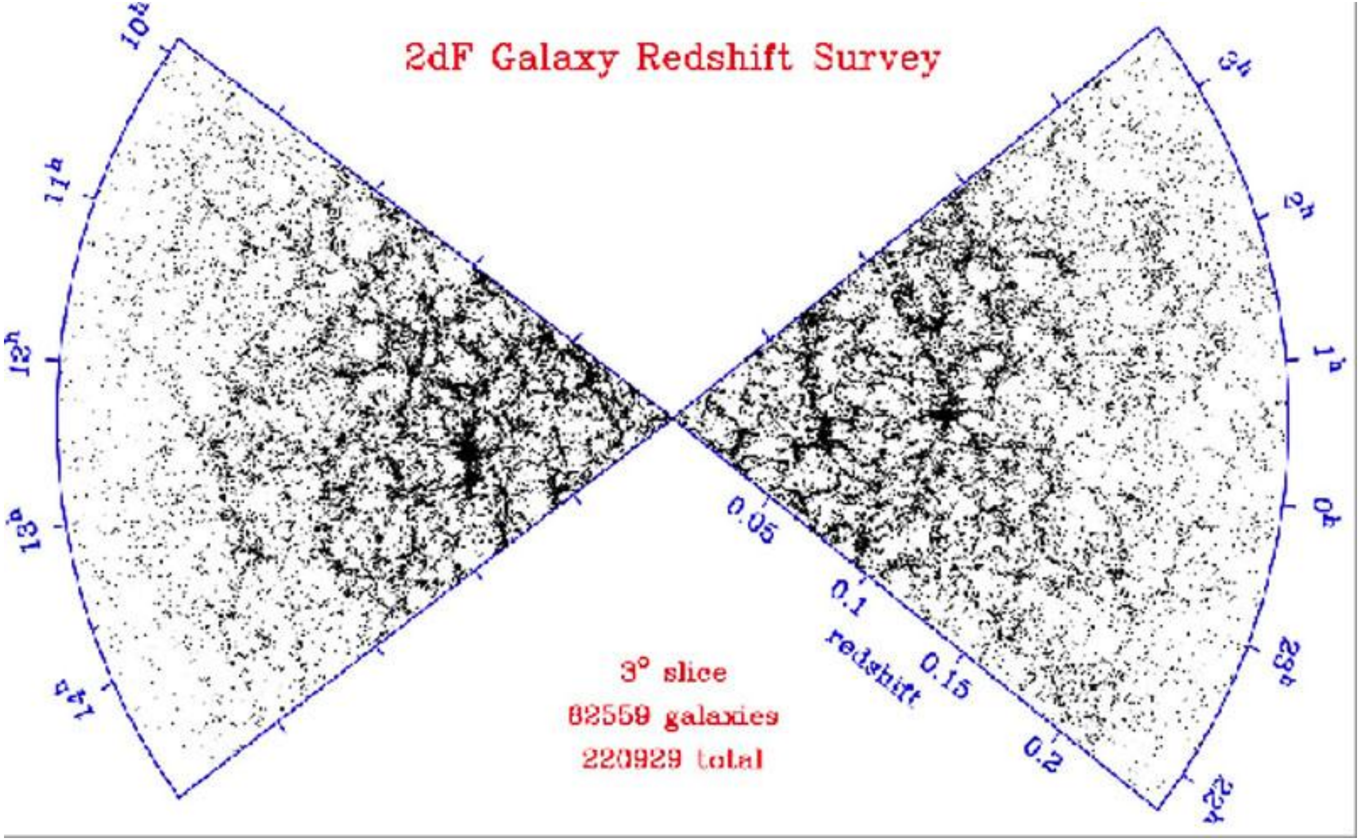
DES Collaboration, Àlex Alarcón, Gary Bernstein, Andresa Campos, Ami Choi, Juan Pablo Cordero, Joe DeRose, Scott Dodelson, Spencer Everett, Marco Gatti, Giulia Giannini, Ian Harrison, Will Hartley, Huan Lin, Jamie McCullough, Aaron Roodman, Carles Sánchez, Michael Troxel, Boyan Yin *et al.*

The Redshift Problem for Lensing Surveys

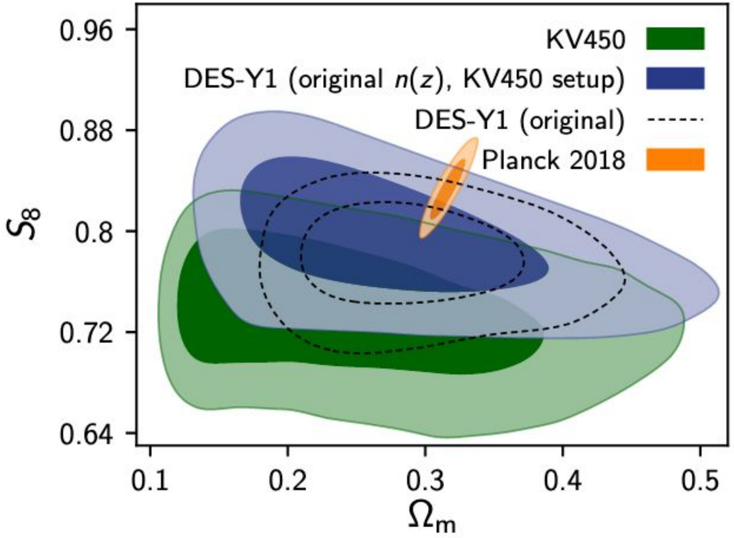
sompz Formalism and Methodology

DES Y3 Preliminary n(z) Results

Large scale structure is a valuable probe of cosmology.

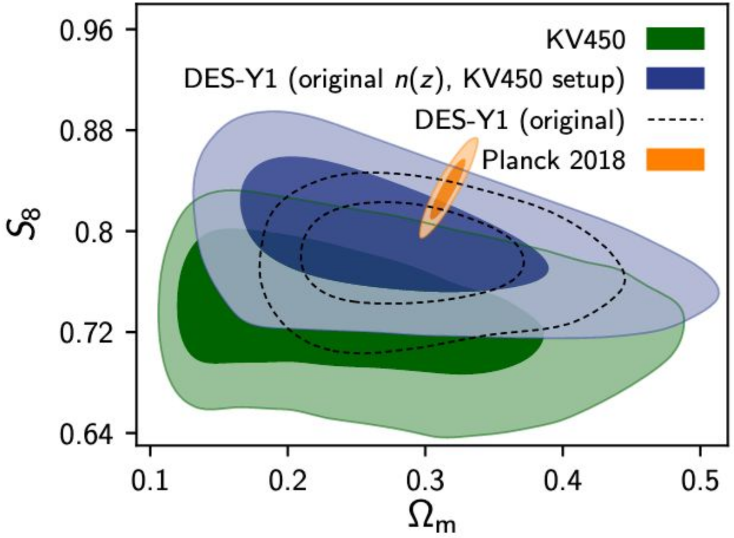


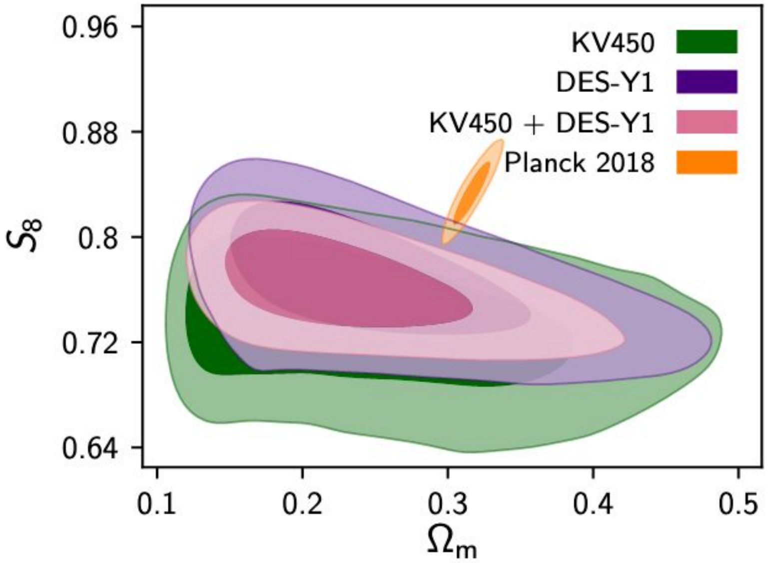
Cosmic shear is a valuable probe of cosmology.



Joudaki *et al*. 2019

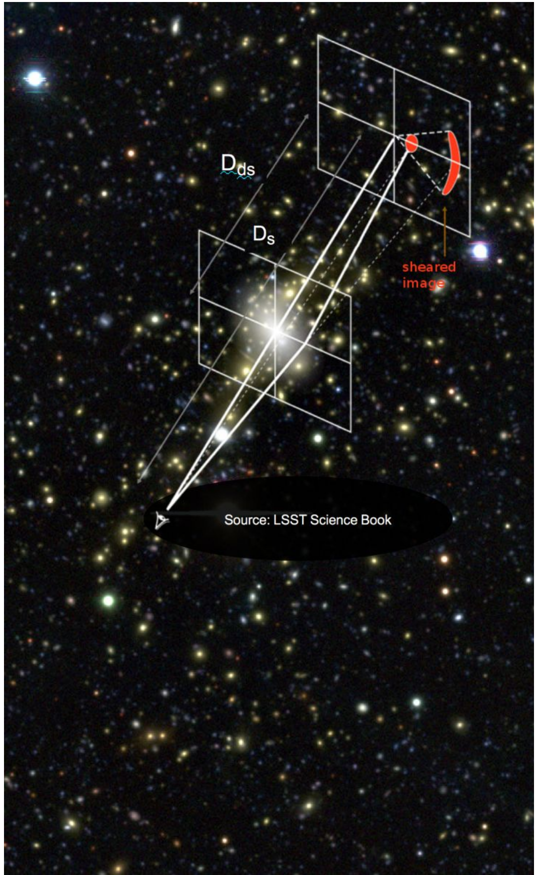
Cosmic shear is host to an interesting debate about redshifts.





Joudaki *et al.* 2019

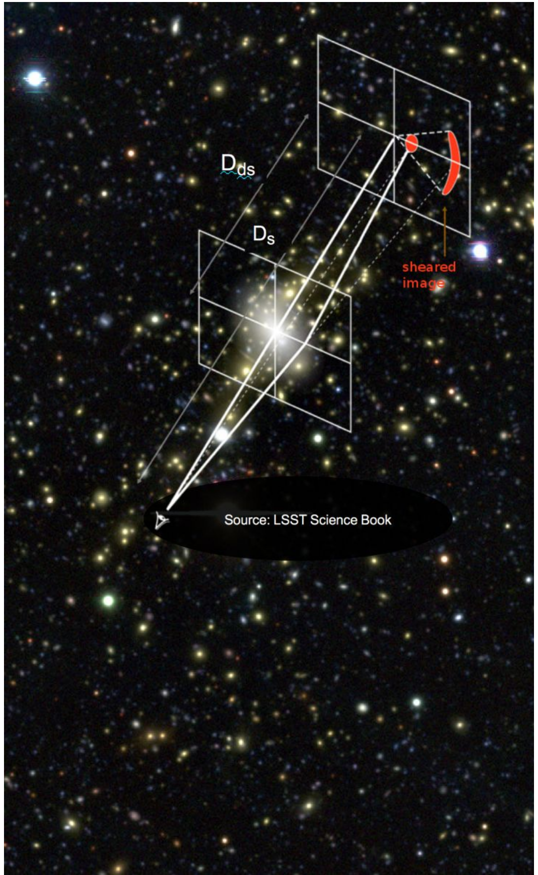
As light passes massive structures, it is bent due to gravity. This effect shifts, magnifies, and shears the galaxy image.



As light passes massive structures, it is bent due to gravity. This effect shifts, magnifies, and shears the galaxy image.

$$\gamma_t(\theta) = \langle \kappa(\theta') \rangle_{\theta' < \theta} - \kappa(\theta)$$

$$\kappa = \Sigma / \left[\frac{c^2}{4\pi G} \frac{D_s}{D_d D_{ds}} \right]$$

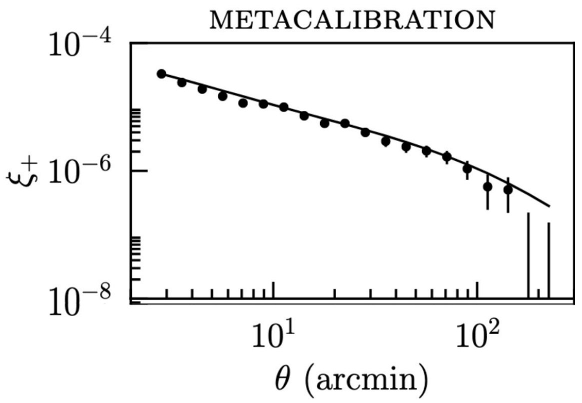


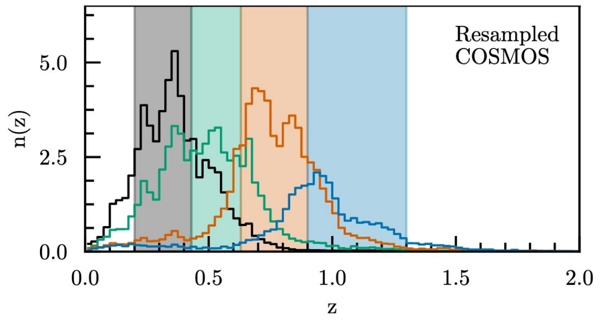
Testing a cosmological model with cosmic shear depends on a statistical ensemble of two basic measurements: galaxy shapes and redshifts.

$$P^{ij}_{\kappa}(\ell) = \int_0^{\chi_H} d\chi \frac{q^i(\chi) q^j(\chi)}{\chi^2} P_{\mathrm{NL}}\left(\frac{\ell + 1/2}{\chi}, \chi\right)$$

$$\hat{\xi}^{ij}_{\pm}(\theta) = \frac{1}{2\pi} \int d\ell \ell J_{0/4}(\theta\ell) P^{ij}_{\kappa}(\ell)$$

$$q^i(\chi) = \frac{3}{2}\Omega_m \left(\frac{H_0}{c}\right)^2 \frac{\chi}{a(\chi)} \int_{\chi}^{\chi_H} d\chi' n^i(\chi') \frac{\chi' - \chi}{\chi'}$$





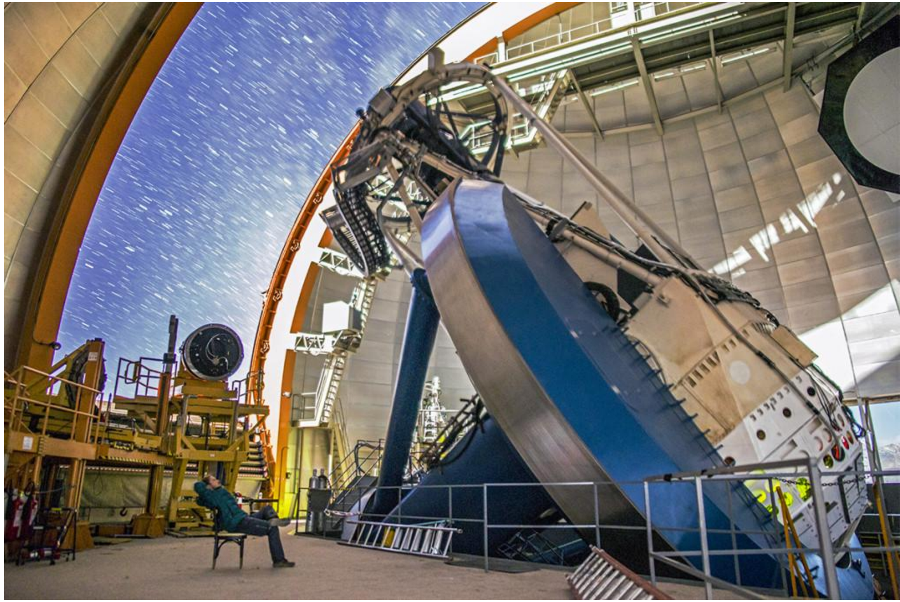
Troxel *et al.* 2018

The 3x2pt analysis is the cosmological workhorse of DES.

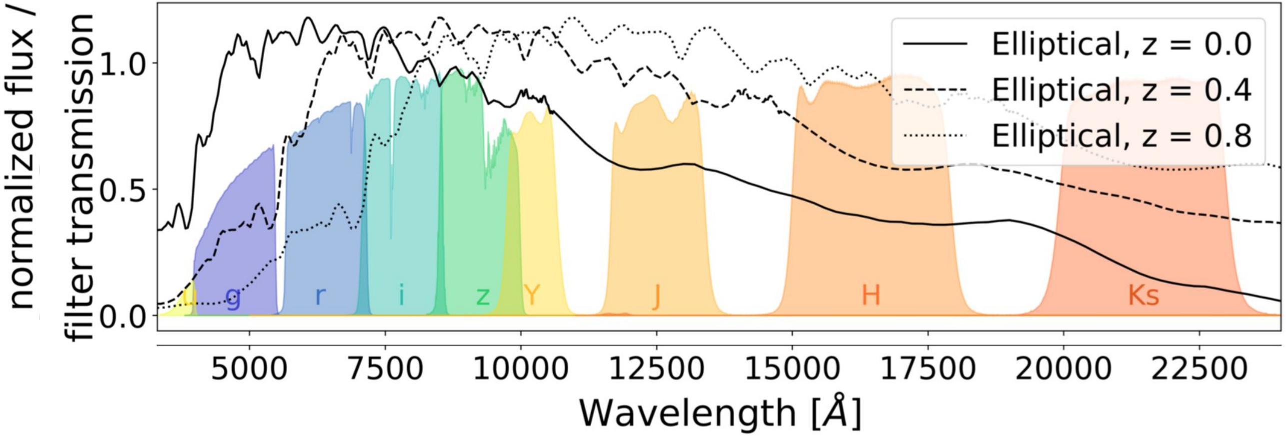
- Cosmic Shear
    - The galaxy shape two point correlation functions.
- Galaxy Clustering
    - The galaxy position two point correlation function.
- Galaxy-Galaxy Lensing
    - The correlation function of the shape of background galaxies with the positions of foreground galaxies.

The Dark Energy Survey is among a group of leading lensing experiments.

- 4m Blanco Telescope
- ~5000 square degree wide field survey in *g r i z Y* over 6 years
- >100M Y3 Galaxy Source Catalog
- 27 square degree deep time-domain survey in *u g r i z Y* overlapping with archival *Y J H K*
    - 8 square degree *ugrizJHK* used for DES Y3

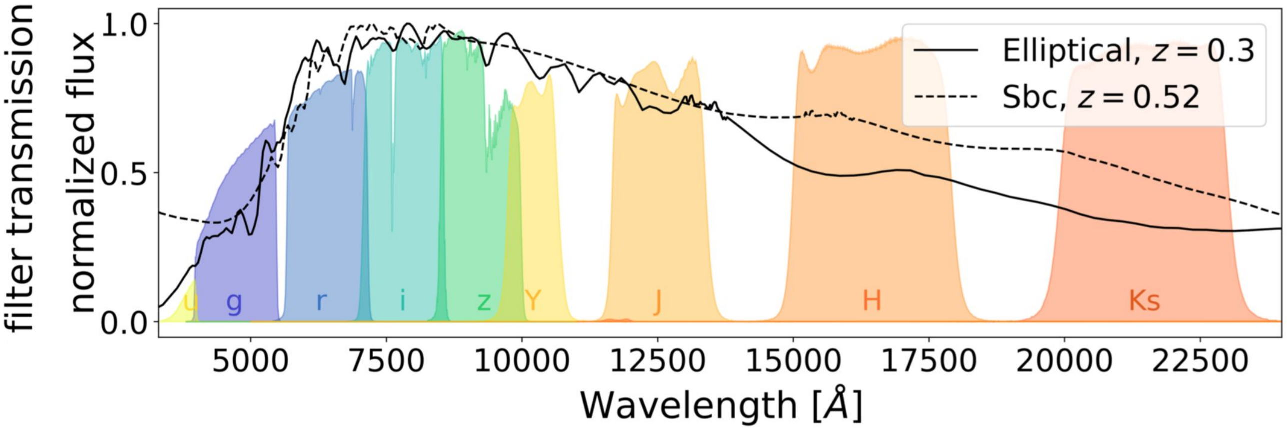


Photometric redshifts rely on understanding the color-redshift relation.



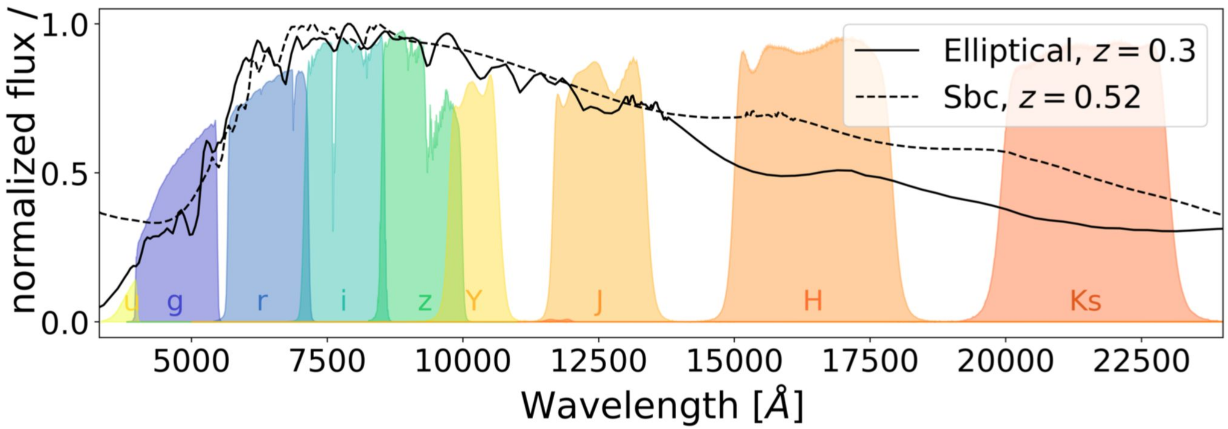
Buchs & Davis *et al.* 2019

The statistical color-redshift relation is fraught with degeneracies that are unavoidable in wide field surveys.



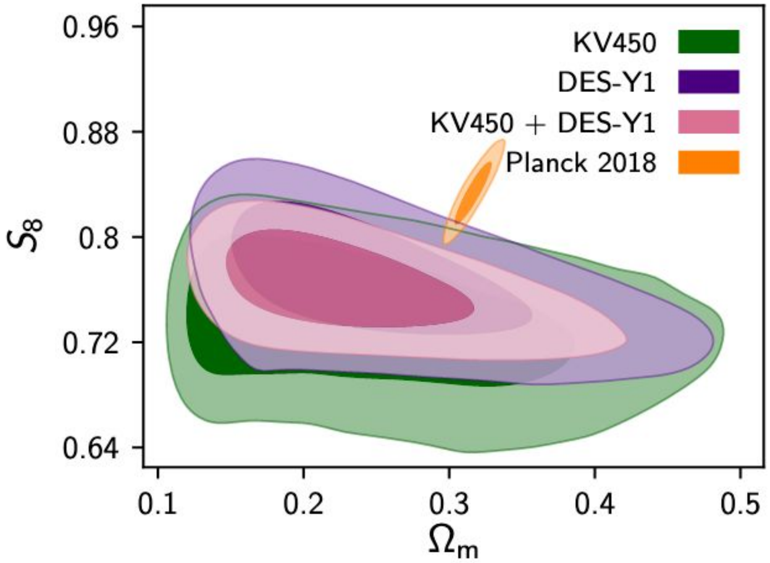
Buchs & Davis *et al.* 2019

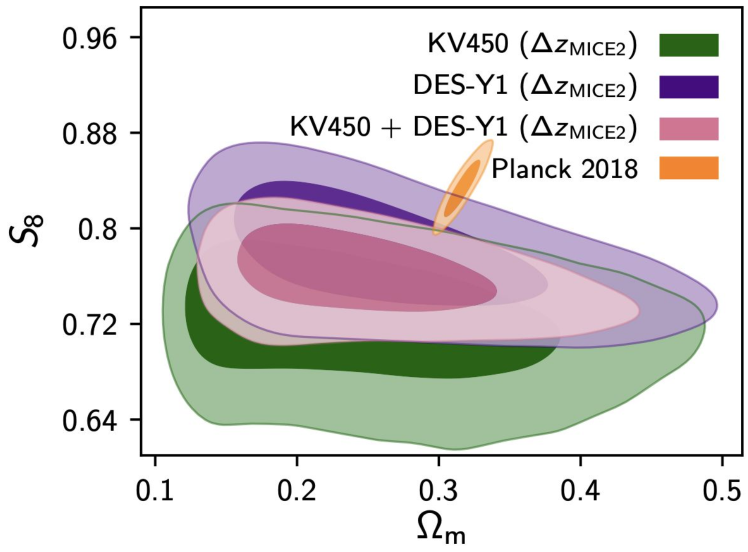
Neither of the two prevailing photo-z solution paradigms solve the real problem: degeneracies in the statistical color-redshift relation.



1. Machine learning the spectroscopic redshift to wide field flux relationship is flawed because they are trained on biased or incomplete spectroscopic samples whose selection function is necessarily different than the wide field sample.
2. Template fitting codes rely on analytical recipes that can be insufficient to accurately describe the observed color- redshift relation of galaxies

Selection biases introduced by using spectroscopic redshift samples are key to the redshift debate.





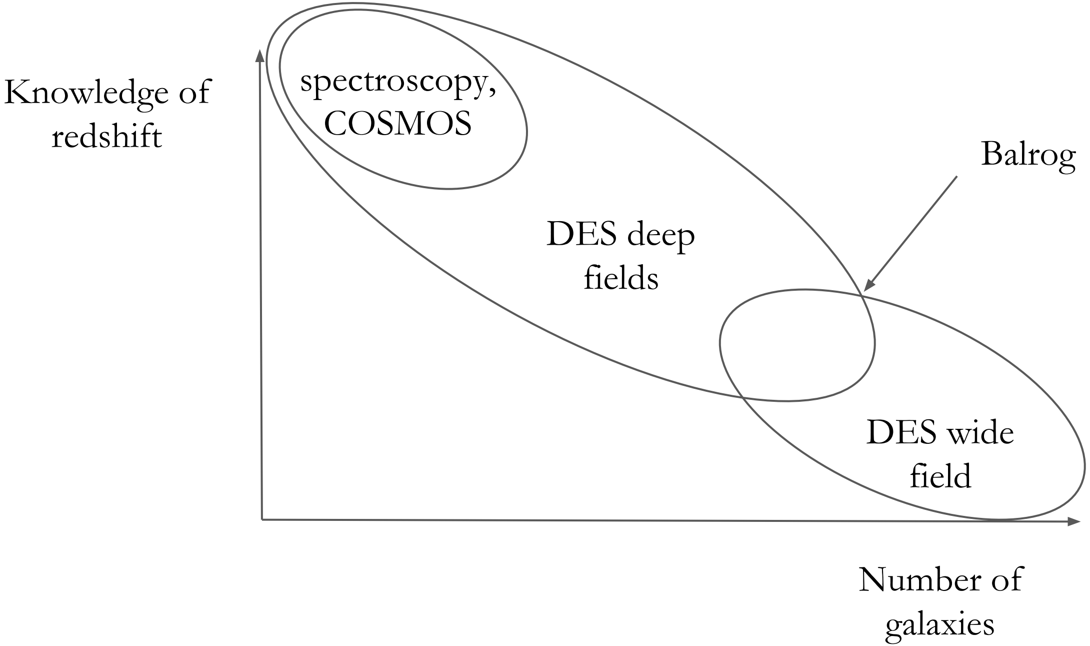
Joudaki *et al.* 2019

The Redshift Problem for Lensing Surveys

sompz Formalism and Methodology

DES Y3 Preliminary n(z) Results

sompz is one part of a larger DES Y3 redshift effort. Our work focuses on how to leverage the deep fields to break the key degeneracies in the statistical color-redshift relation.



We can leverage overlap of DES deep fields with archival NIR data

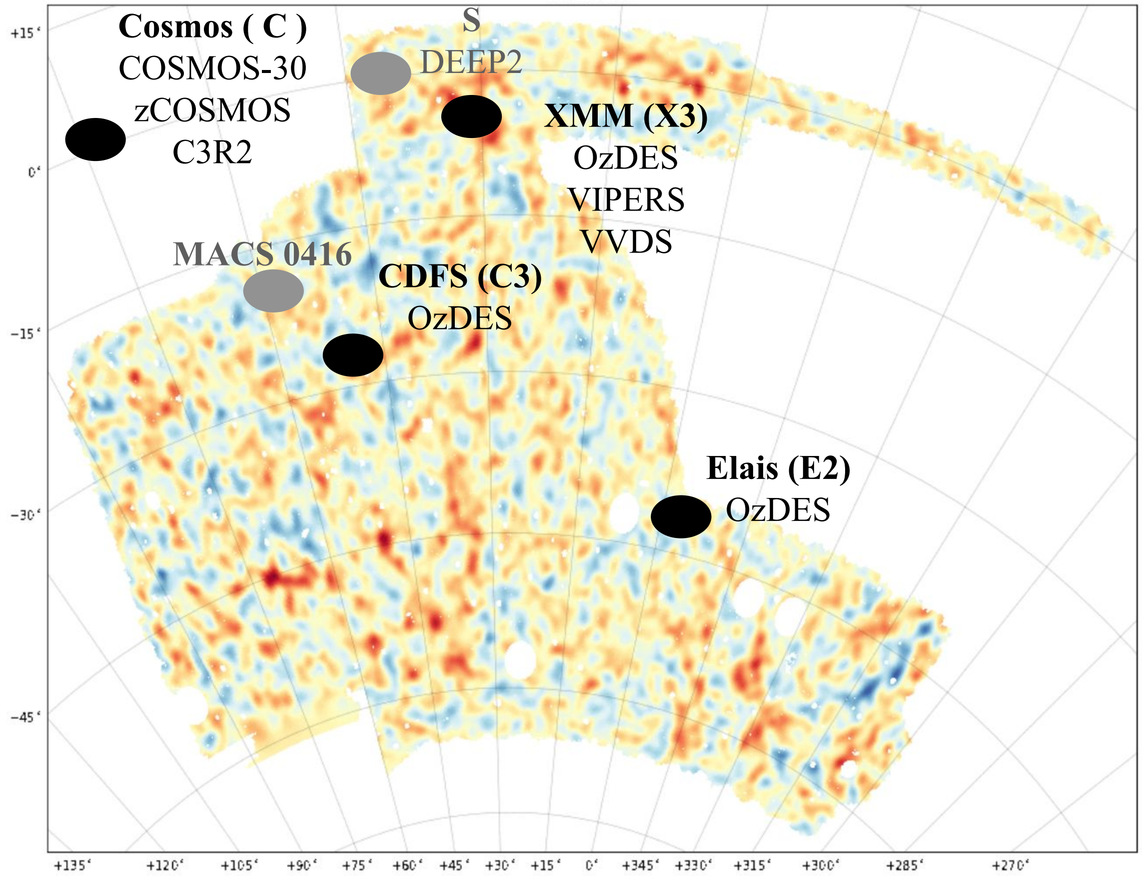


Figure Credit: Alex Amon and Deep Fields Team

Leveraging the deep fields to improve our knowledge of the statistical color-redshift relation amounts to marginalizing over deep photometric information.

$$p(\mathrm{z}|r,i,z) = \int p(\mathrm{z}|u,g,r,i,z,J,H,K)\,p(u,g,r,i,z,J,H,K|r,i,z)\,d[u,g,r,i,z,J,H,K]$$

p(z) at a given deep photometric color

Probabilistic mapping between deep and wide photometric colors

In order to marginalize over deep photometric information, we must replace regions of color space with discrete categories c and $\hat{c}$.

$$p(z|\hat{c}, \hat{s}) = \sum_{c} p(z|c, \hat{c}, \hat{s})p(c|\hat{c}, \hat{s})$$

p(z) at a given deep photometric color

Probabilistic mapping between deep and wide photometric colors

$$c = (u, g, r, i, z, J, H, K)$$

$$\hat{c} = (r, i, z)$$

$\hat{s}$ is the sample selection function

Our strategy is to leverage the deep fields in DES Y3 to improve our knowledge of the statistical color-redshift relation.

$$p(z|\hat{b},\hat{s}) = \sum_{\hat{c}\in\hat{b}} p(z|\hat{c},\hat{s})p(\hat{c}|\hat{s})$$

$$= \sum_{\hat{c}\in\hat{b}} \sum_{c} p(z|c,\hat{c},\hat{s})p(c|\hat{c},\hat{s})p(\hat{c}|\hat{s})$$

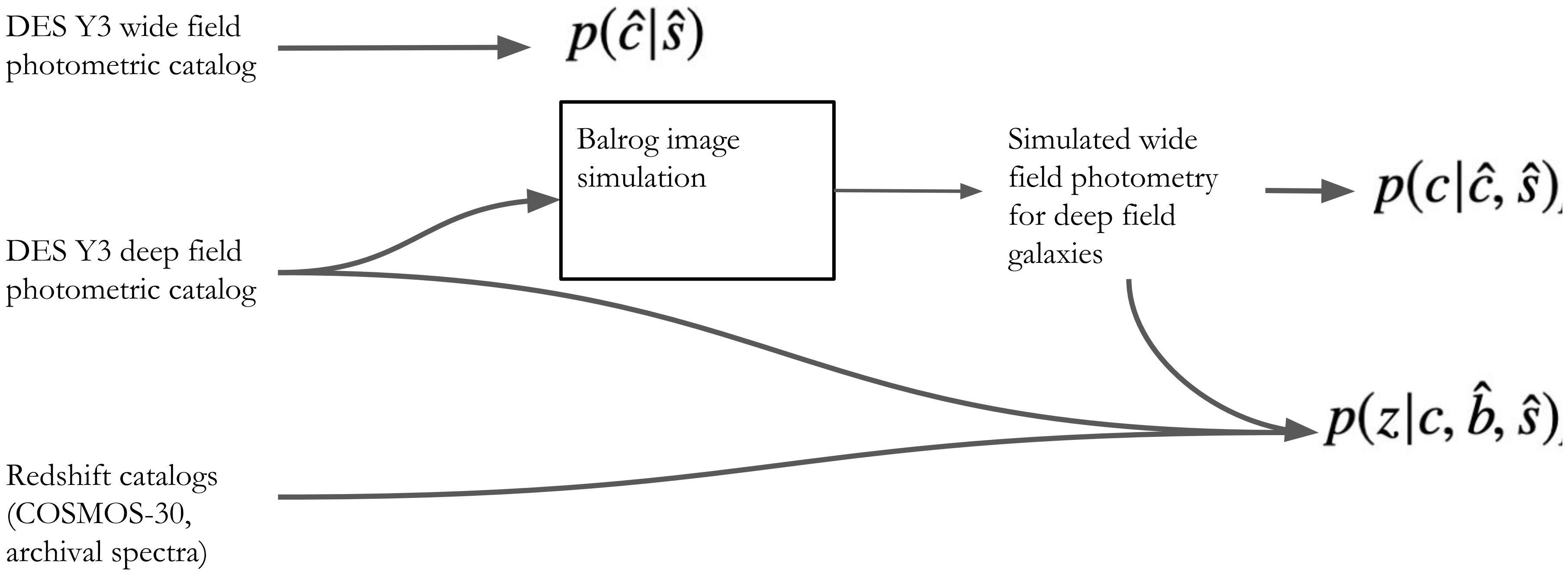
$$p(z|\hat{b},\hat{s}) \approx \sum_{\hat{c}\in\hat{b}} \sum_{c} p(z|c,\hat{b},\hat{s})p(c|\hat{c},\hat{s})p(\hat{c}|\hat{s})$$

$$c = (u, g, r, i, z, J, H, K)$$

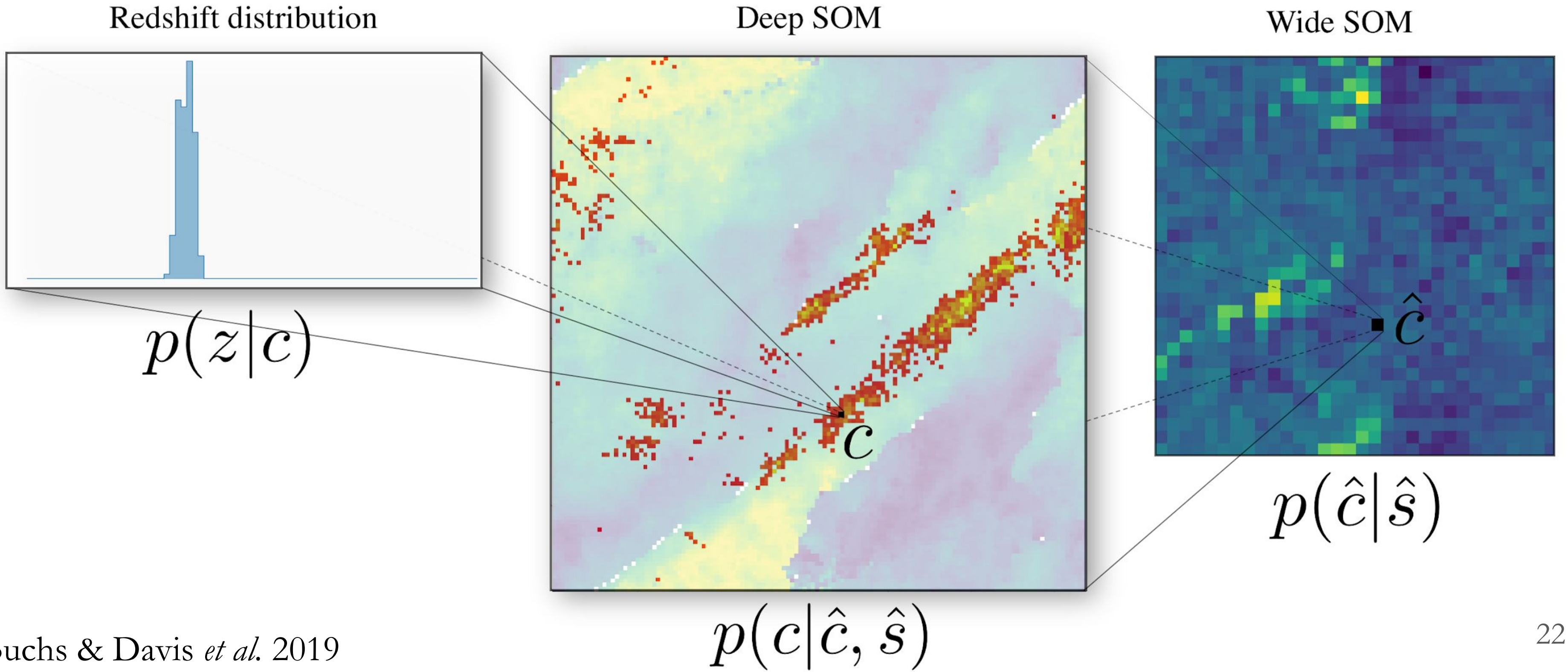
$$\hat{c} = (r, i, z)$$

$\hat{s}$ is the sample selection function

$$p(z|\hat{b}, \hat{s}) \approx \sum_{\hat{c} \in \hat{b}} \sum_{c} p(z|c, \hat{b}, \hat{s}) p(c|\hat{c}, \hat{s}) p(\hat{c}|\hat{s})$$

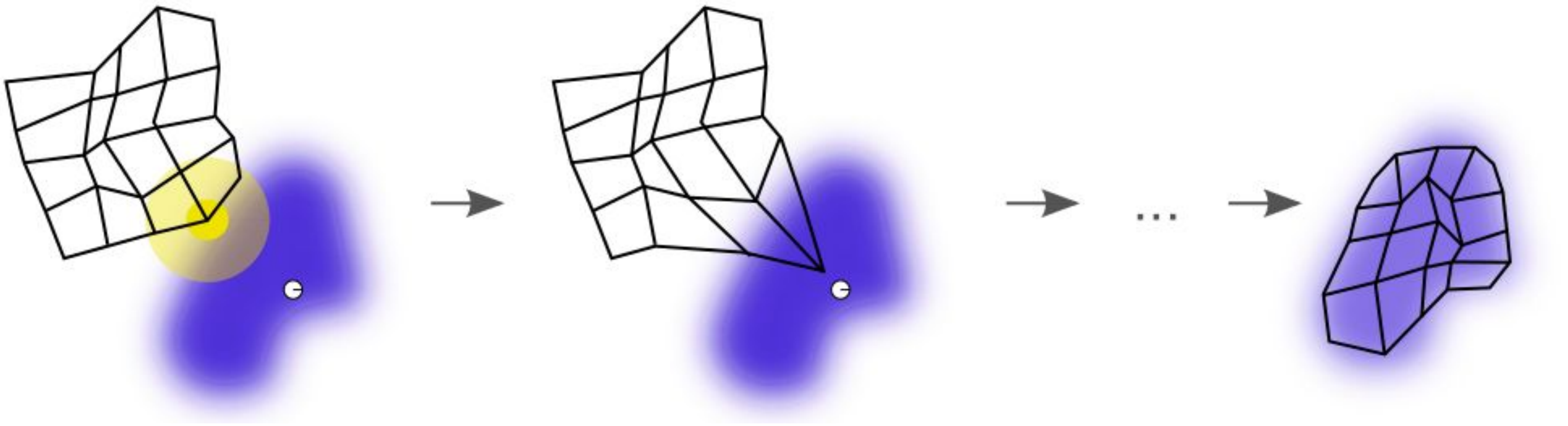


The self-organizing map classifies galaxies of similar colors into categories called cells.



Buchs & Davis *et al.* 2019

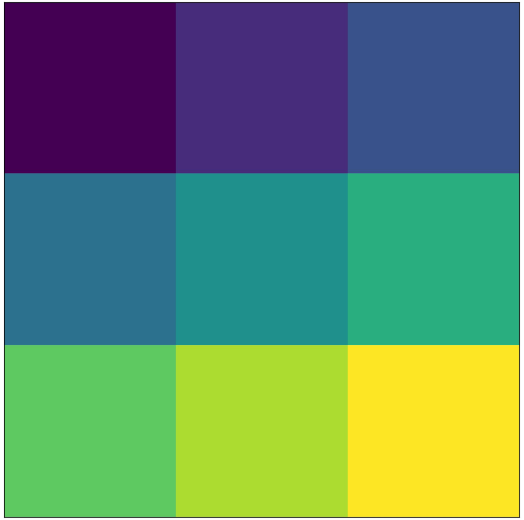
The self-organizing map learns to smoothly cover color space, which facilitates interpolation of cells for which we have less than average counts.



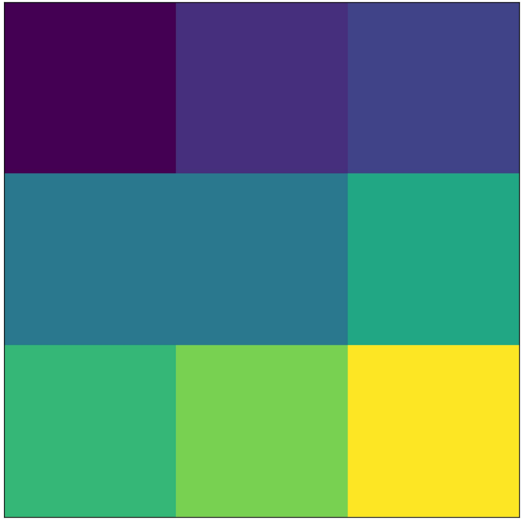
The self-organizing map groups together galaxies of similar colors in cells.

Training Input i: 

step i

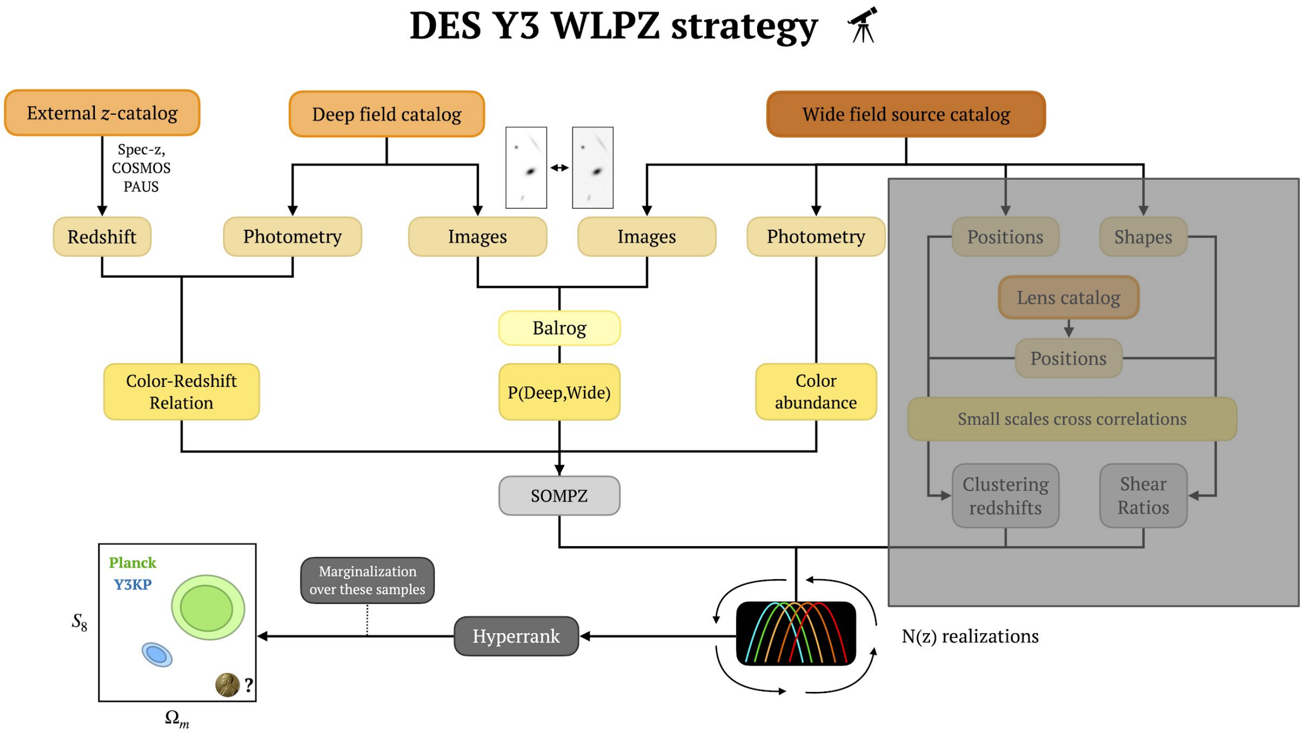


step i+1



A modern cosmological analysis necessarily depends on using numerous independent constraints and sophisticated sampling procedures over relevant uncertainties to verify the robustness of the overall measurement.

DES Y3 WLPZ strategy



Credit: Àlex Alarcón and Judit Prat

The Redshift Problem for Lensing Surveys

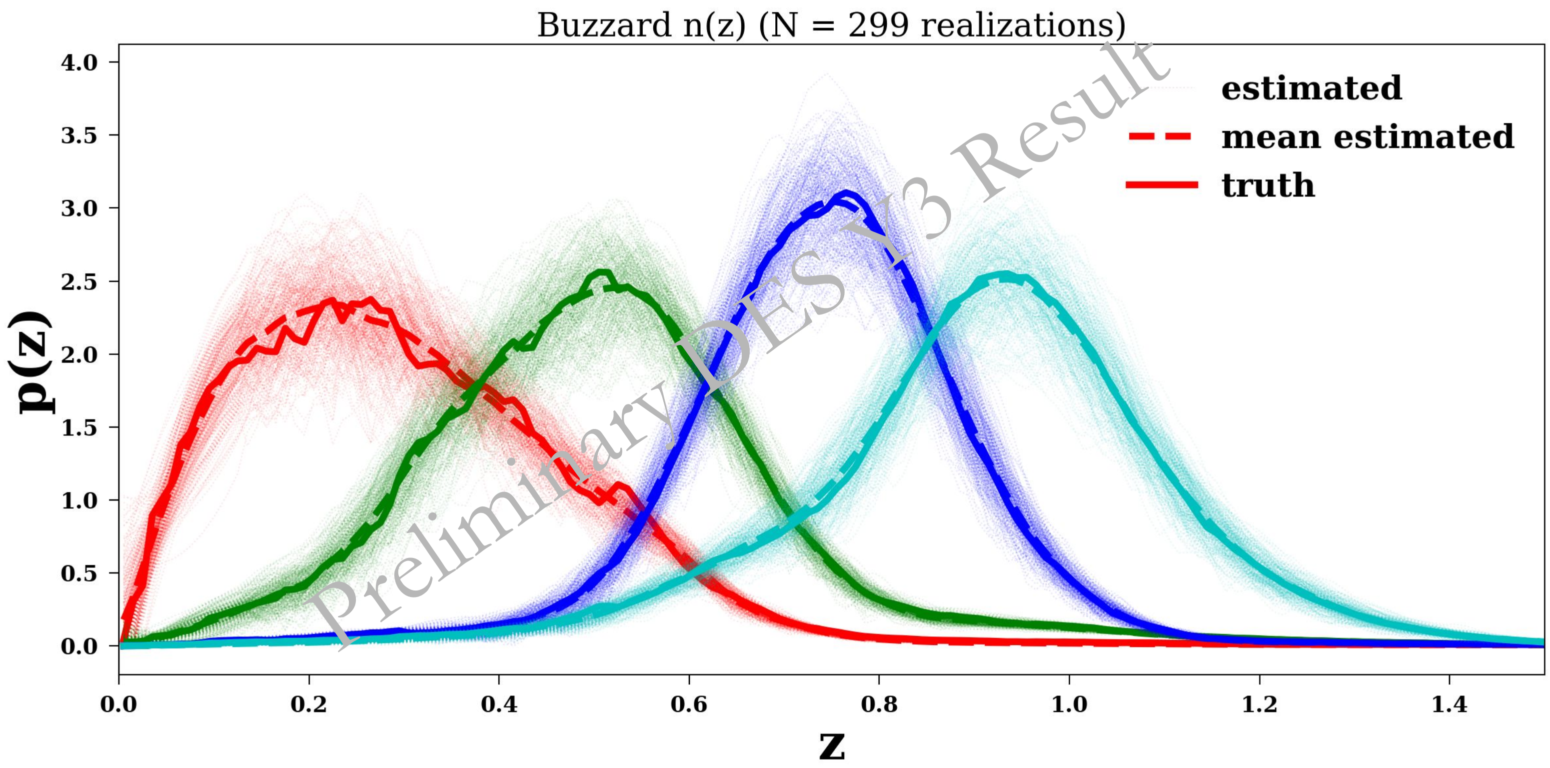
sompz Formalism and Methodology

DES Y3 Preliminary n(z) Results

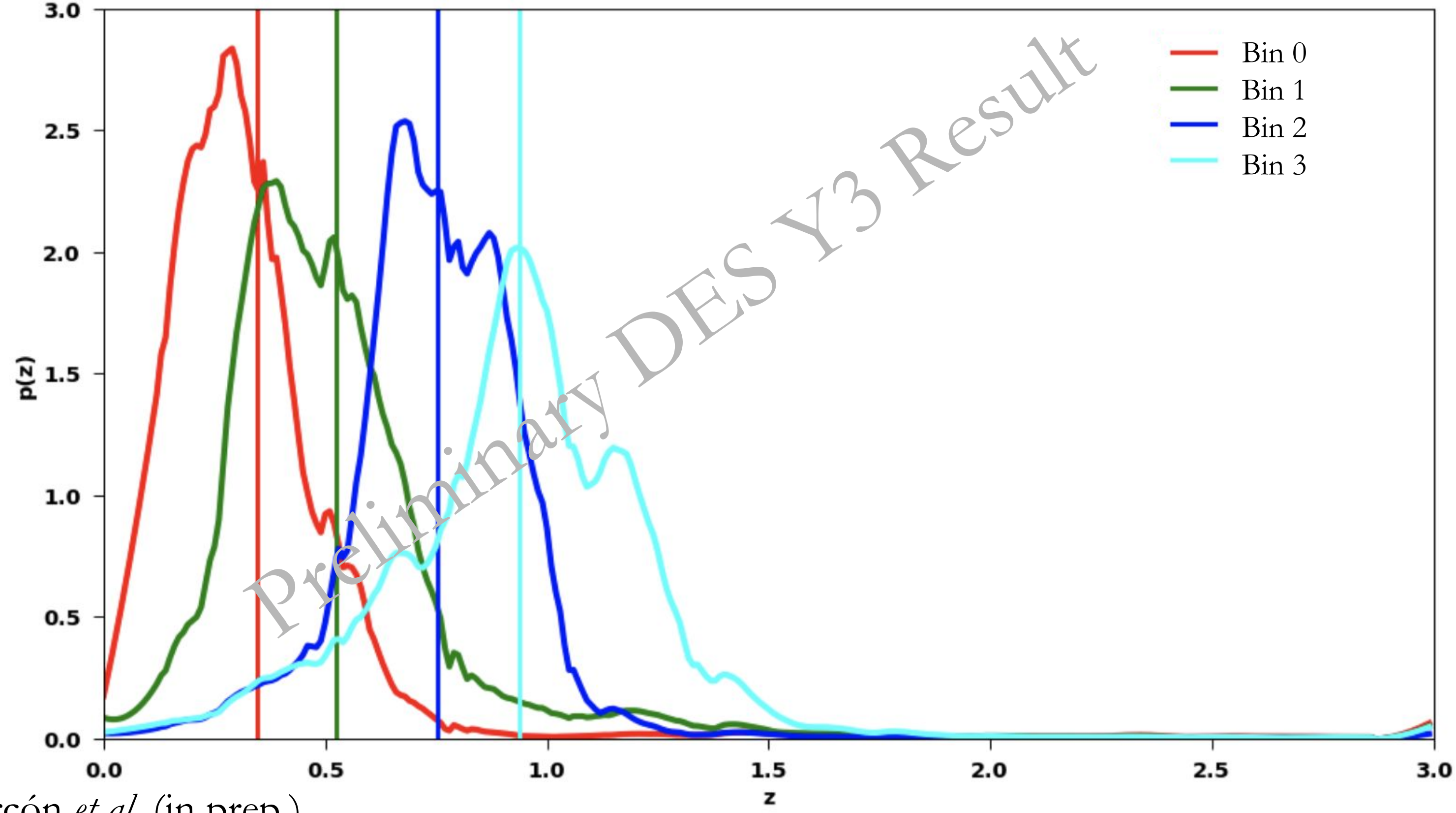
DES Y3 Preliminary n(z) Results

1. Validation of method on simulations
2. Result of method applied to data
3. Estimate of key sources of uncertainty
    - Sample Variance in the deep fields
    - Deep field photometric calibration error
    - Redshift sample biases
4. Probability Integral Transform (PIT) samples

On average, the method recovers the true n(z) in tests on simulations.



We have fiducial estimates of the redshift distributions of our weak lensing source sample in four tomographic bins.



Myles & Alarcón *et al.* (in prep.)

Which of the following sources of uncertainty in our measurement is most important?

- Sample Variance in the Deep Fields **A**
- Biases in the redshift sample **B**
- Photometric calibration error in the Deep Fields **C**

Respond at **PollEv.com/justinmyles663**

Text **JUSTINMYLES663** to **22333** once to join, then **A, B, or C**

Which of the following sources of uncertainty in our measurement is most important?

Sample Variance in the Deep Fields **A**

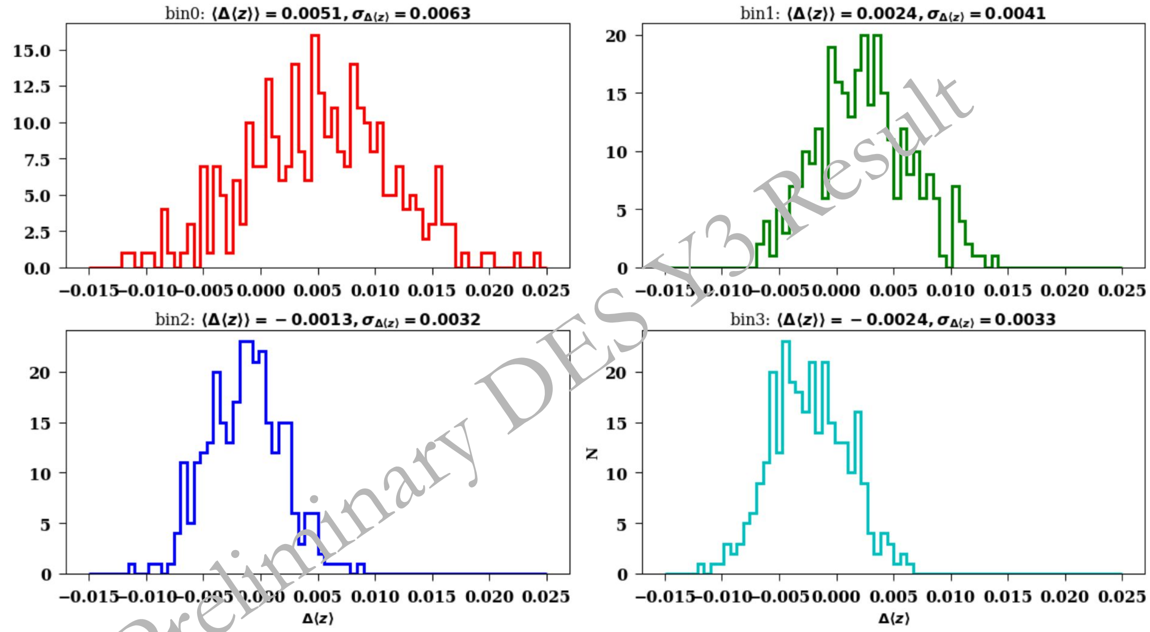
Biases in the redshift sample **B**

Photometric calibration error in the Deep Fields **C**

Which of the following sources of uncertainty in our measurement is most important?

- Sample Variance in the Deep Fields **A**
- Biases in the redshift sample **B**
- Photometric calibration error in the Deep Fields **C**

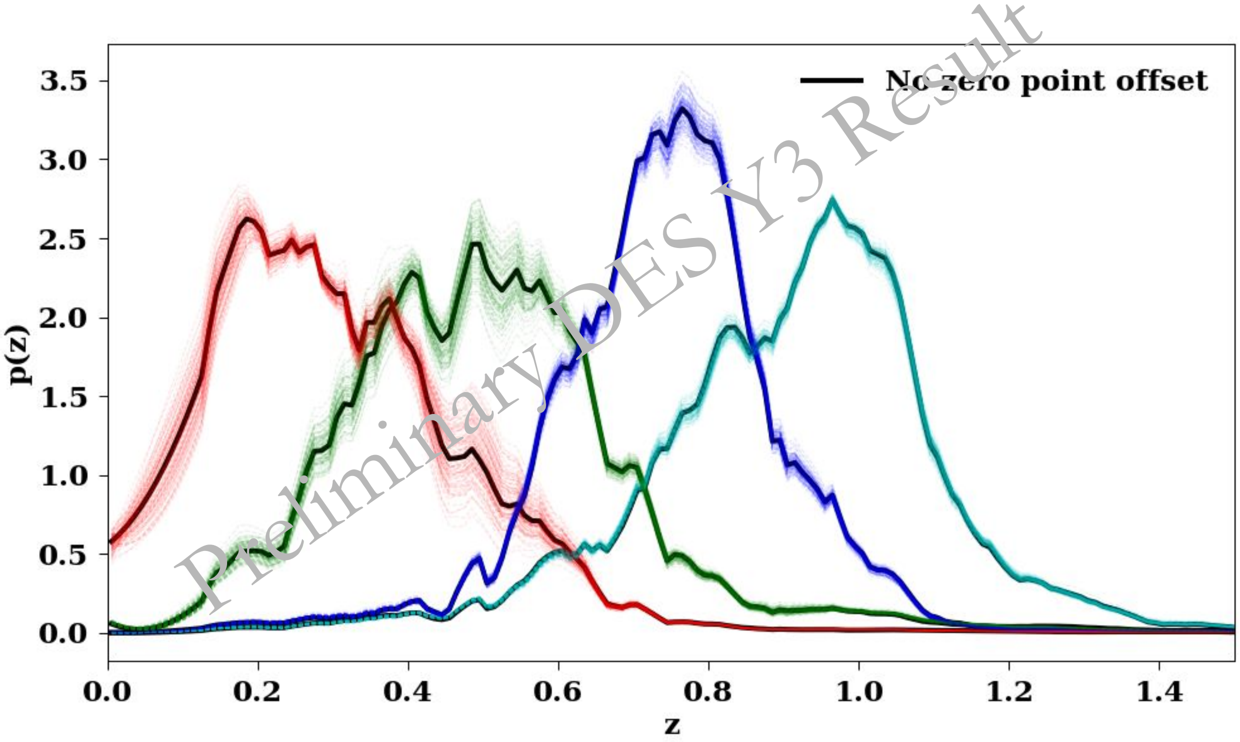
We can derive a simple estimate on the uncertainty in mean redshift per bin due to **sample variance** by running with many different underlying deep and redshift fields in simulations.



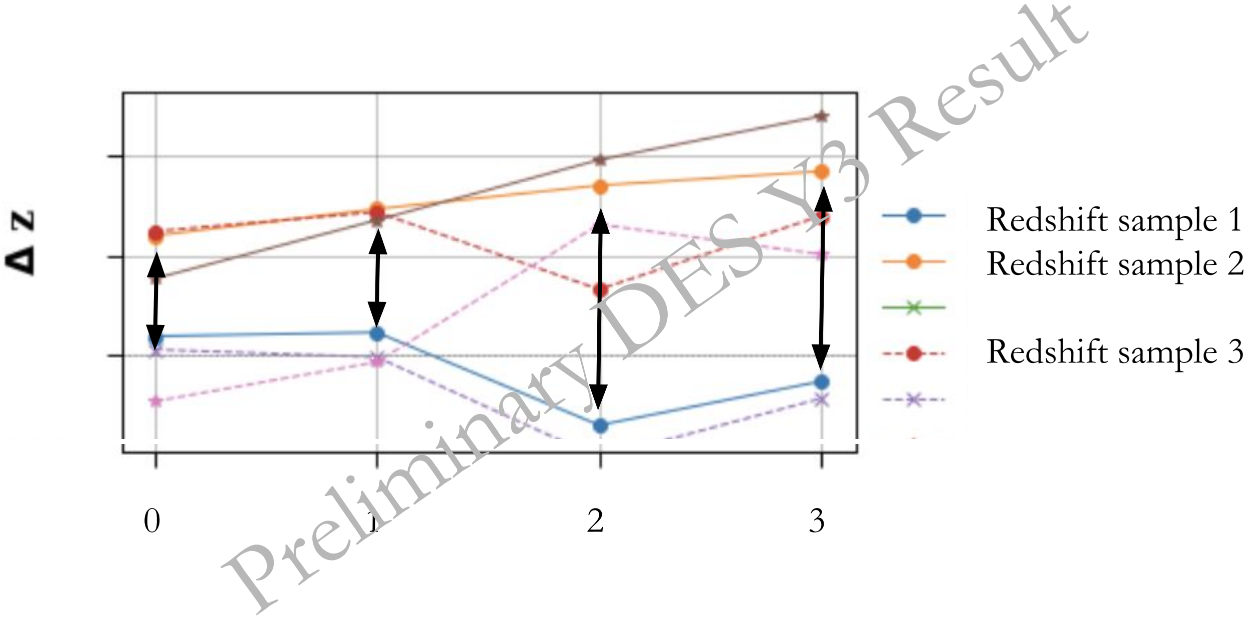
Y1: σ_z~0.012 from these effects

Value	Bin 1	Bin 2	Bin 3	Bin 4
z^{PZ} range	0.20–0.43	0.43–0.63	0.63–0.90	0.90-1.30
COSMOS footprint sampling	±0.0073	±0.0077	±0.0039	±0.0070
COSMOS limited sample size	±0.0009	±0.0017	±0.0018	±0.0030
COSMOS photometric calibration errors	±0.0030	±0.0040	±0.0039	±0.0059
COSMOS hidden variables	±0.0066	±0.0066	±0.0066	±0.0066
COSMOS errors in matching	±0.0073	±0.0073	±0.0073	±0.0073

Our cosmology result will robustly account for the uncertainty in our estimate due to the **photometric calibration error** of the deep fields.

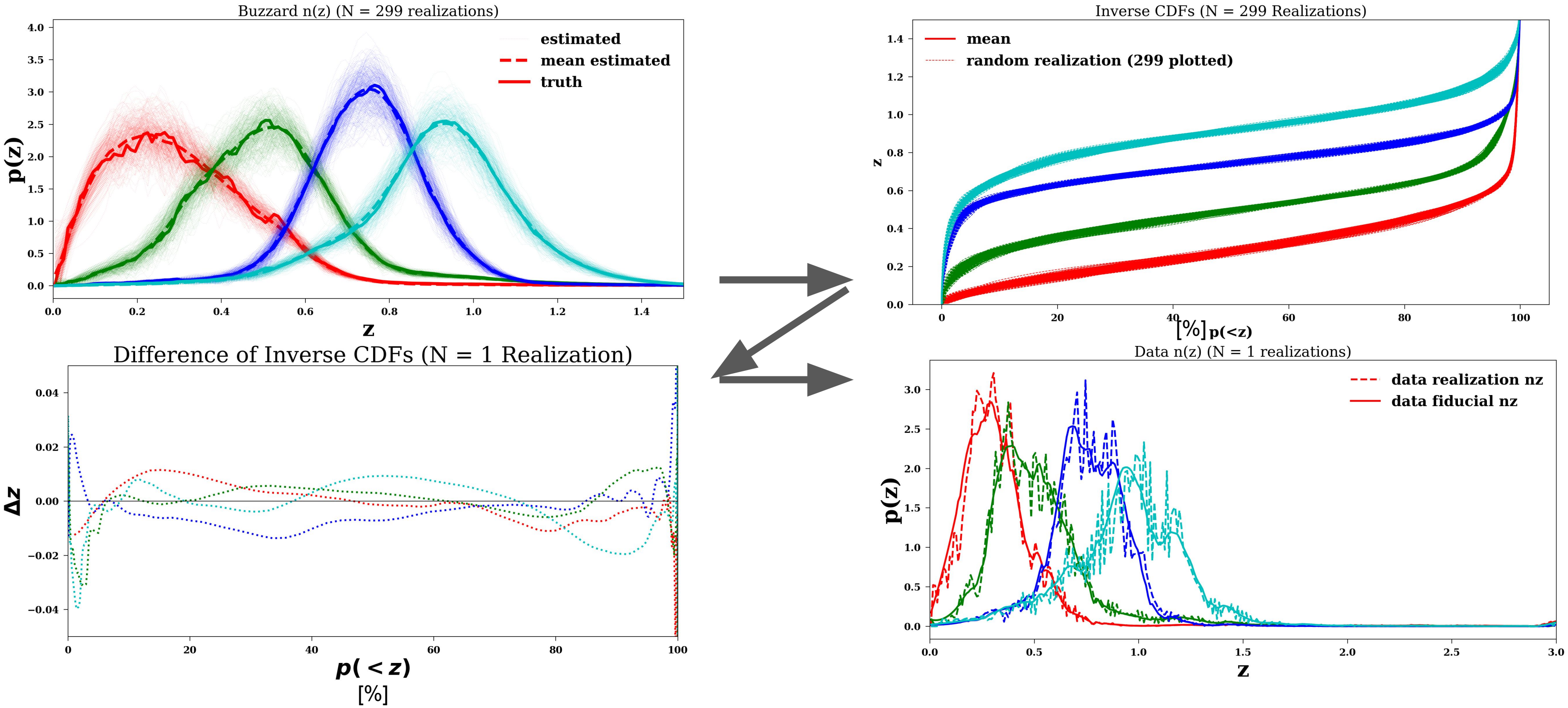


Our cosmology result will robustly account for the uncertainty in our estimate due to **systematic biases in the redshifts** of the sample.



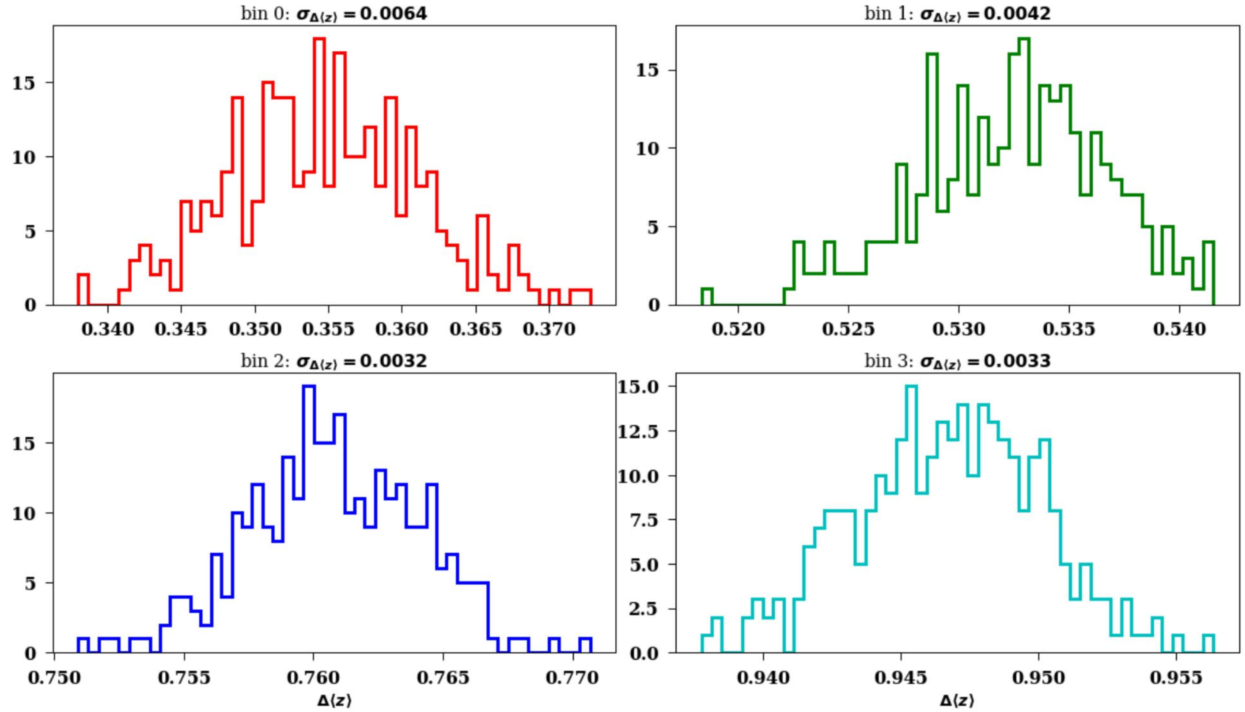
Credit: Alex Amon

We have developed a sampling procedure which preserves the full variation in an ensemble of n(z), enabling us to test how our cosmology depends on specific moments of the n(z).



Myles & Alarcón *et al.* (in prep.)

This PIT preserves the variance on the shift in the mean redshift per bin from the underlying Buzzard realizations. Importantly, the PIT transfers not only the shift in the mean, but the full information on the variation in n(z) from each sample



Summary

1. sompz aims to directly address the fundamental photo-z problem -- degeneracies in the color-redshift relation -- by leveraging deep, many-band photometry

2. We have characterized the key uncertainties in our result:
   - We need to maximize the overlap of NIR and optical wide field surveys
   - We need to collect more spectra
   - We need to improve the photometric calibration of our deep fields

3. We have a new method of generating an ensemble of n(z) reflecting our uncertainty in the underlying distribution which preserves not just the mean, but higher order moments

Credits

SOMPZ Team: Àlex Alarcón, Gary Bernstein, Andresa Campos, Ami Choi, Alex Amon, Juan Pablo Cordero, Joe DeRose, Scott Dodelson, Spencer Everett, Marco Gatti, Giulia Giannini, Daniel Gruen, Ian Harrison, Will Hartley, Huan Lin, Jamie McCullough, Justin Myles, Aaron Roodman, Carles Sánchez, Michael Troxel, Boyan Yin, *et al.*

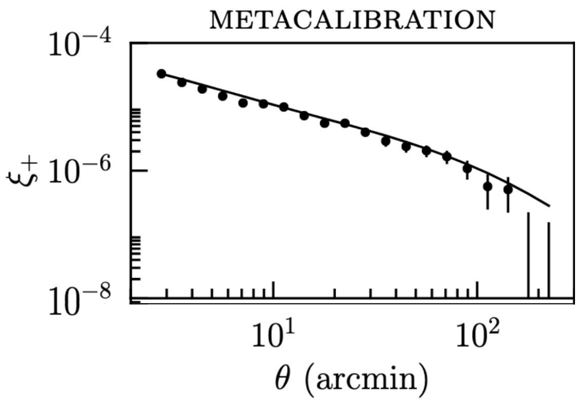
Deep Fields Team: Alex Amon, Gary Bernstein, Ami Choi, Katie Eckert, Daniel Gruen, Ian Harrison, Will Hartley, Robert Gruendl, Mike Jarvis, Nacho Sevilla, Erin Sheldon, *et al.*

Balrog Team: Spencer Everett, Eric Huff, Nikolay Kuropatkin, Brian Yanny, *et al.*

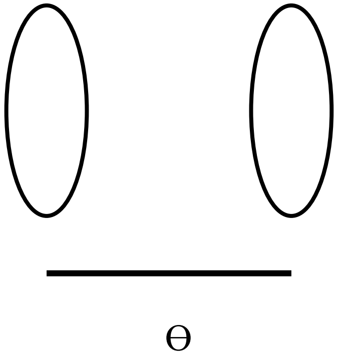
Extra Slides

Testing a cosmological model with cosmic shear depends on a statistical ensemble of two basic measurements: galaxy shapes and redshifts.

$$\hat{\xi}^{ij}_{\pm}(\theta) = \frac{1}{2\pi}\int d\ell\ell J_{0/4}(\theta\ell)P^{ij}_{\kappa}(\ell)$$

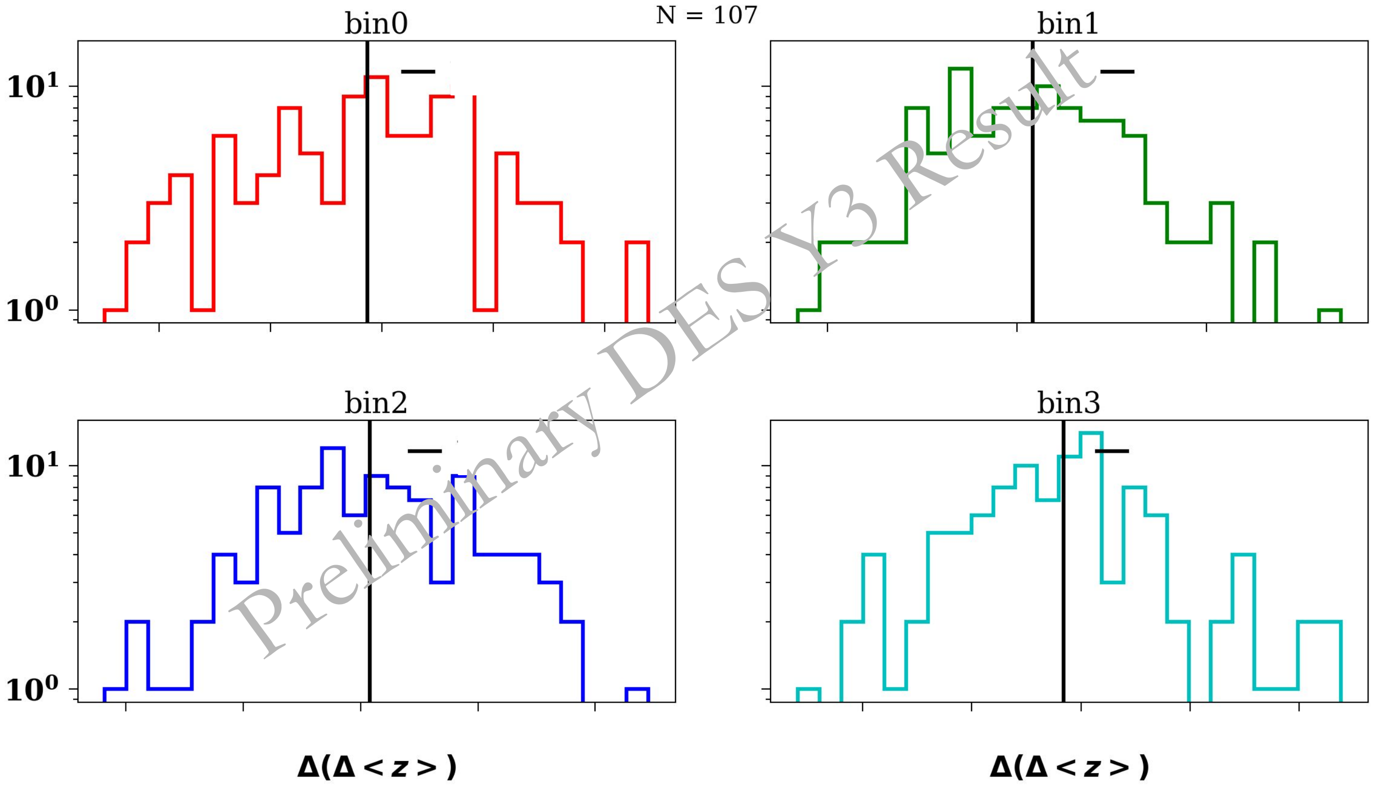


Troxel *et al.* 2018



The similar alignment of these two galaxies leads to a positive contribution to ξ^+

Our cosmology result will robustly account for the uncertainty in our estimate due to the **photometric calibration error** of the deep fields.



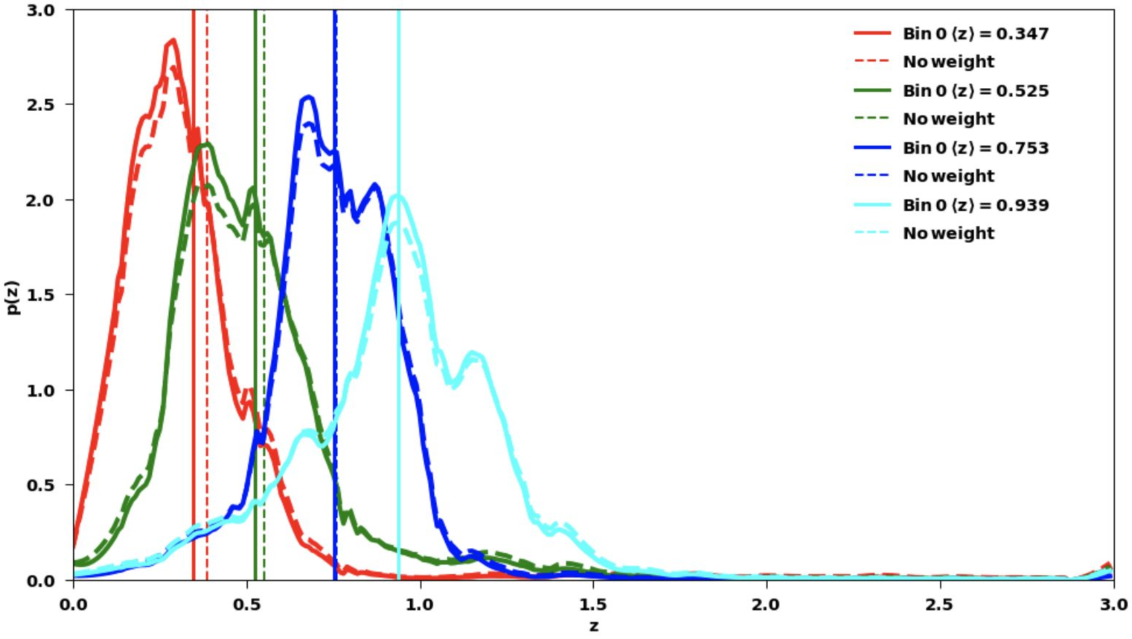
Results on Data

To get lensing weighted n(z), we use each galaxy's shear response R, an explicit lensing weight w, and, for Balrog galaxies, normalize by the number of injections.

$$p(\hat{c}|\hat{s}) = \sum_{i \in \hat{c}} w_i R_i$$

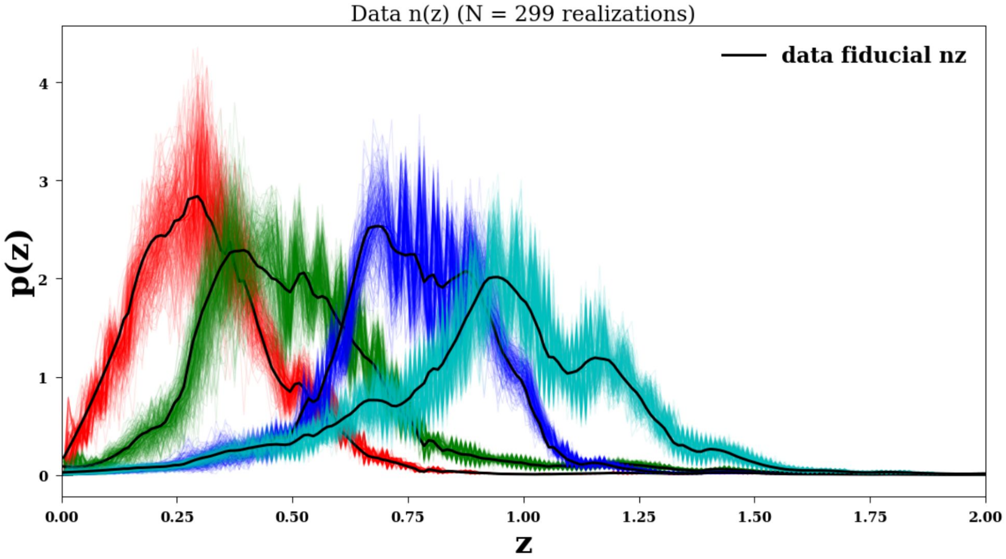
$$p(z|c, \hat{b}, \hat{s}) = \sum_{i \in (c, \hat{b})} \frac{w_i R_i p_i(z)}{N_{i,\text{inj.}}}$$

$$p(c|\hat{c}, \hat{s}) = \frac{p(c, \hat{c}|\hat{s})}{p(\hat{c}|\hat{s})} = \frac{\sum_{i \in \hat{s}} \delta_{c,c_i} \delta_{\hat{c},\hat{c}_i} w_i R_i / N_{i,\text{inj.}}}{\sum_{i \in \hat{s}} \delta_{\hat{c},\hat{c}_i} w_i R_i / N_{i,\text{inj.}}}$$



Probability Integral Transform Samples

1. Sample mean Buzzard n(z)
2. Sample realization Buzzard n(z)
3. Sample data n(z)
4. Order samples in 1 and 2 and compute difference between each pair of samples
5. Apply difference to data n(z) samples
6. Make histogram of data n(z) samples



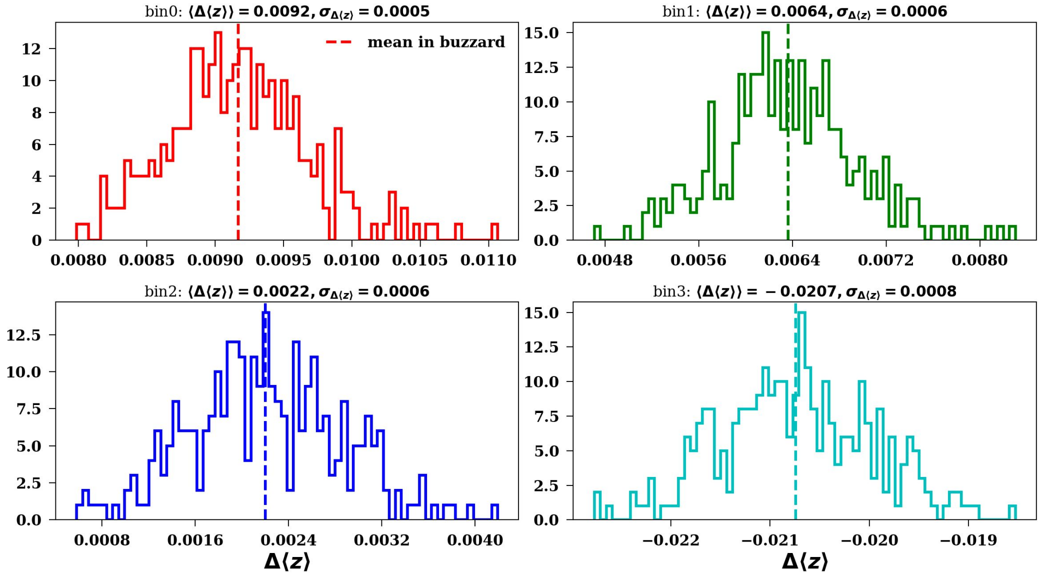
Results on Buzzard

We can use these samples to assess the impact of bin conditionalization on our result.

$$= \sum_{\hat{c}\in\hat{b}} \sum_{c} p(z|c, \hat{c}, \hat{s}) p(c|\hat{c}, \hat{s}) p(\hat{c}|\hat{s})$$

$$\approx \sum_{\hat{c}\in\hat{b}} \sum_{c} p(z|c, \hat{b}, \hat{s}) p(c|\hat{c}, \hat{s}) p(\hat{c}|\hat{s}) \quad \text{DES Y3}$$

$$\approx \sum_{\hat{c}\in\hat{b}} \sum_{c} p(z|c, \hat{s}) p(c|\hat{c}, \hat{s}) p(\hat{c}|\hat{s}) \quad \text{Buchs } et\ al.$$



Construction of bins is done by assigning each galaxy in our Balrog sample to a tomographic bin according to some arbitrary bin edges such that each bin have a similar number of galaxies, and assigning each wide som cell to the bin to which a plurality of its constituent Balrog galaxies are assigned.

The analysis on Buzzard (Buchs et al. 2018) suggests we are limited by deep fields.

- The scatter in the bias of the mean inferred redshift is dominated by limited deep fields, not limited redshift samples. This motivates follow-up observations overlapping with existing infrared surveys.
- The pheno-z method can reduce cosmic variance significantly.

The Deep Fields : deep DECam photometry + NIR

Field	COSMOS (C)	CDFS (C3)	XMM (X3)	Elais (E2)
NIR data	Ultravista	VIDEO	VIDEO	VIDEO
bands	*ugrizY* *YJHKs*	*ugrizY* *JHKs*	*ugrizY* *YJHKs*	*ugriz* *YJHKs*
Exposure time (sec.)	200x90	200x90	200x90	80x90
Overlap area (sq. deg.)	1.38	(1.94)	3.29	3.32

Total: 7.99 sq. deg. (9.93 no Y)

Figure Credit: Alex Amon, Will Hartley, Deep Fields Team