REDSHIFT ESTIMATION TECHNIQUES FOR THE DARK ENERGY SURVEY Y3 DATA

Giulia Giannini

Cosmology from Home, 24 August – 04 September 2020

OUTLINE

- Motivation
- Dark Energy Survey
- Application of redshift estimation methods to DES Y3:
  - SOMPZ (Self-Organizing Maps Photo-Z): Machine learning neural network based technique
  - Clustering redshifts: Cross correlation of the target sample with a high quality redshift sample
- Summary

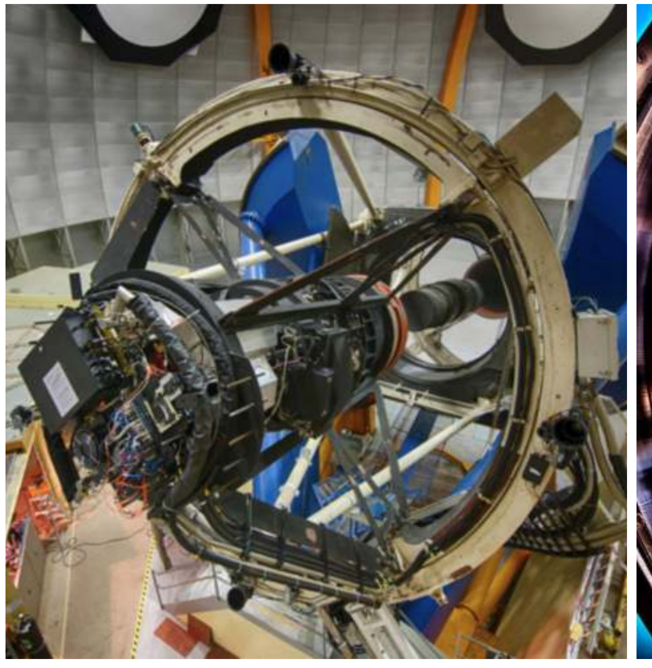
THE PHOTOMETRIC REDSHIFT PROBLEM

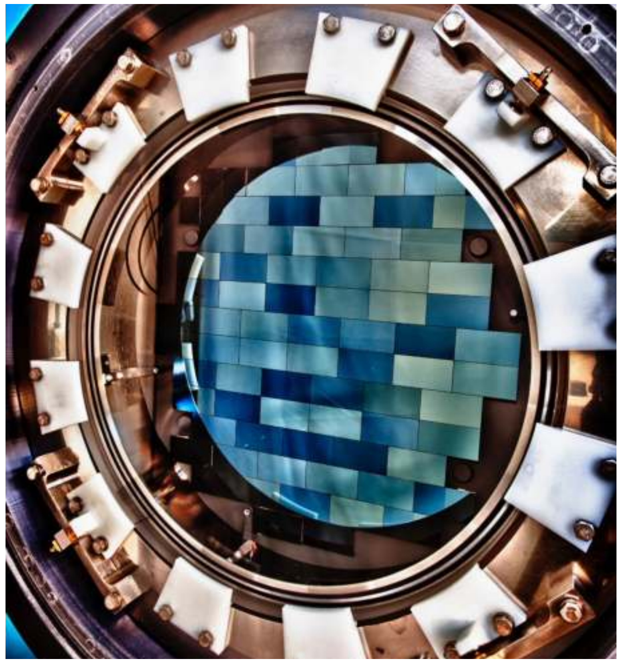
- Redshifts estimation is currently one of the limiting factor in wide field photometric galaxy surveys
- Photometric surveys have been relying on different methodologies but suffer from several issues (spectroscopic samples used for both training and calibration are incomplete, SED templates being inadequate, etc..)
- The correct cosmological interpretation of the measurements relies on an accurate estimate of the redshift distributions of the galaxy samples, which can otherwise lead to biases in the inferred cosmological parameters
- Novel techniques are being developed in order to improve the redshift distributions accuracy

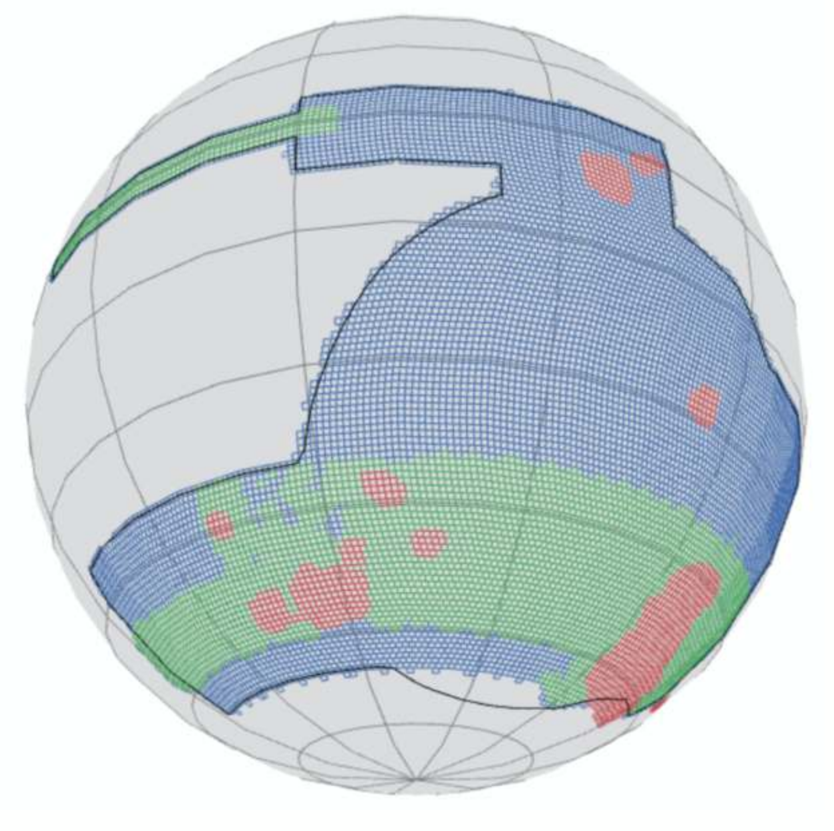
DARK ENERGY SURVEY

DARK ENERGY SURVEY

- Photometric wide-field survey
- 570 megapixel camera DECam
- Four-meter Victor M. Blanco telescope at the Cerro Tololo Inter-American Observatory (CTIO) in the Chilean Andes







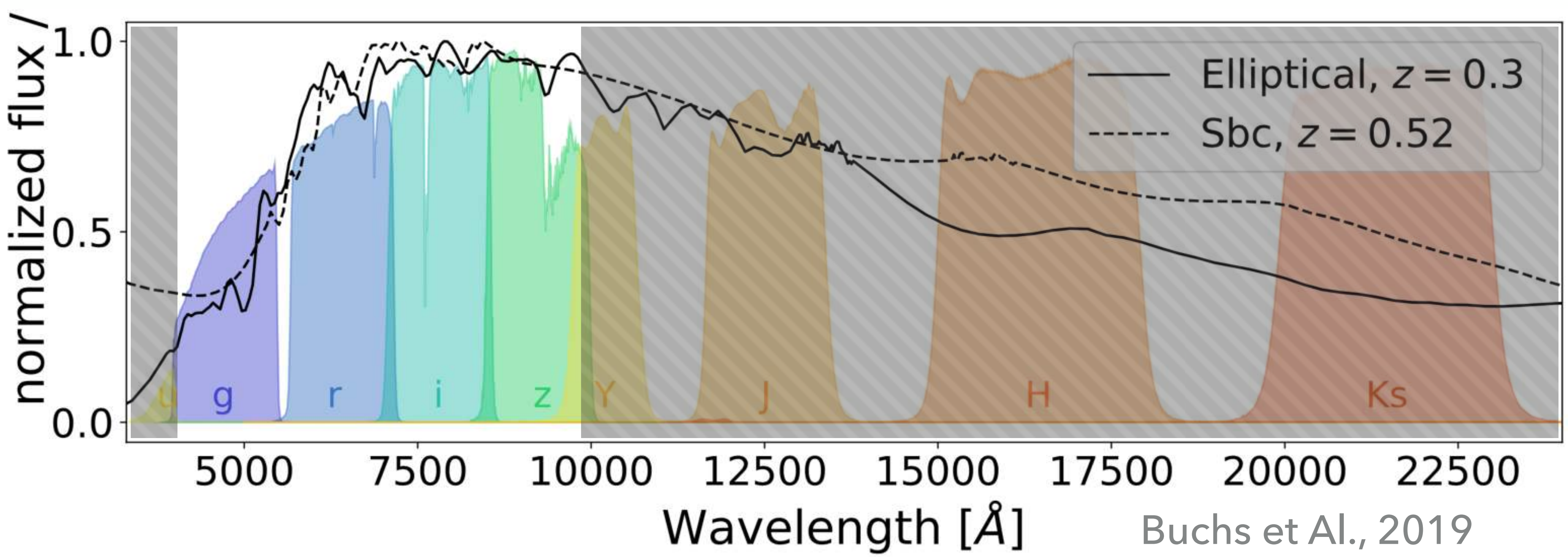
- 5000 deg^2 in the southern hemisphere, 1/8 of sky
- 300 million galaxies and a catalogue of thousands of supernovae
- Probes: Type I supernovae, baryon acoustic oscillations, galaxy clusters and weak gravitational lensing
- DES is preparing the publication of the results from the cosmological analysis of the first three years of data (DES-Y3)

OUTLINE

- Motivation
- Dark Energy Survey
- Application of redshift estimation methods to DES Y3:
  - **SOMPZ (Self-Organizing Maps Photo-Z):** Machine learning neural network based technique
  - Clustering redshifts: Cross correlation of the target sample with a high quality redshift sample
- Summary

TYPE-REDSHIFT DEGENERACY

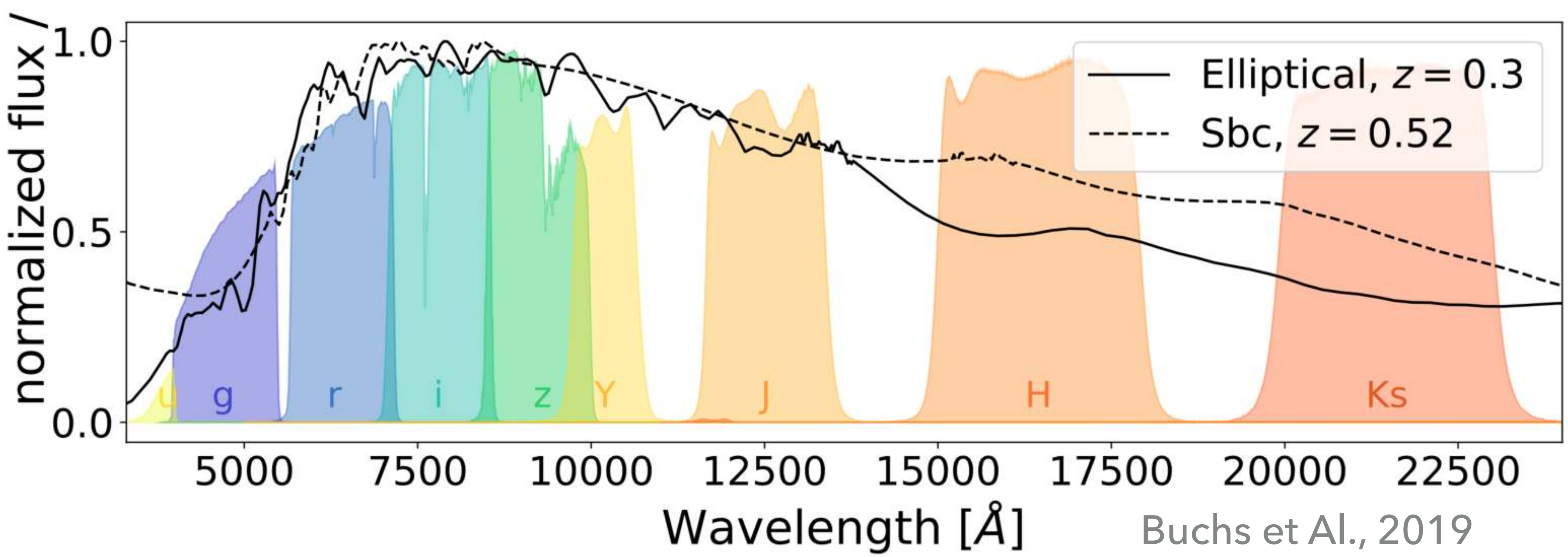
- In the griz bands used to target the whole wide field, redshifted SEDs of two types of galaxies are indistinguishable
- Type-redshift degeneracy is the fundamental cause of uncertainty in redshift calibration



Buchs et Al., 2019

TYPE-REDSHIFT DEGENERACY

- In the griz bands used to target the whole wide field, redshifted SEDs of two types of galaxies are indistinguishable
- Type-redshift degeneracy is the fundamental cause of uncertainty in redshift calibration



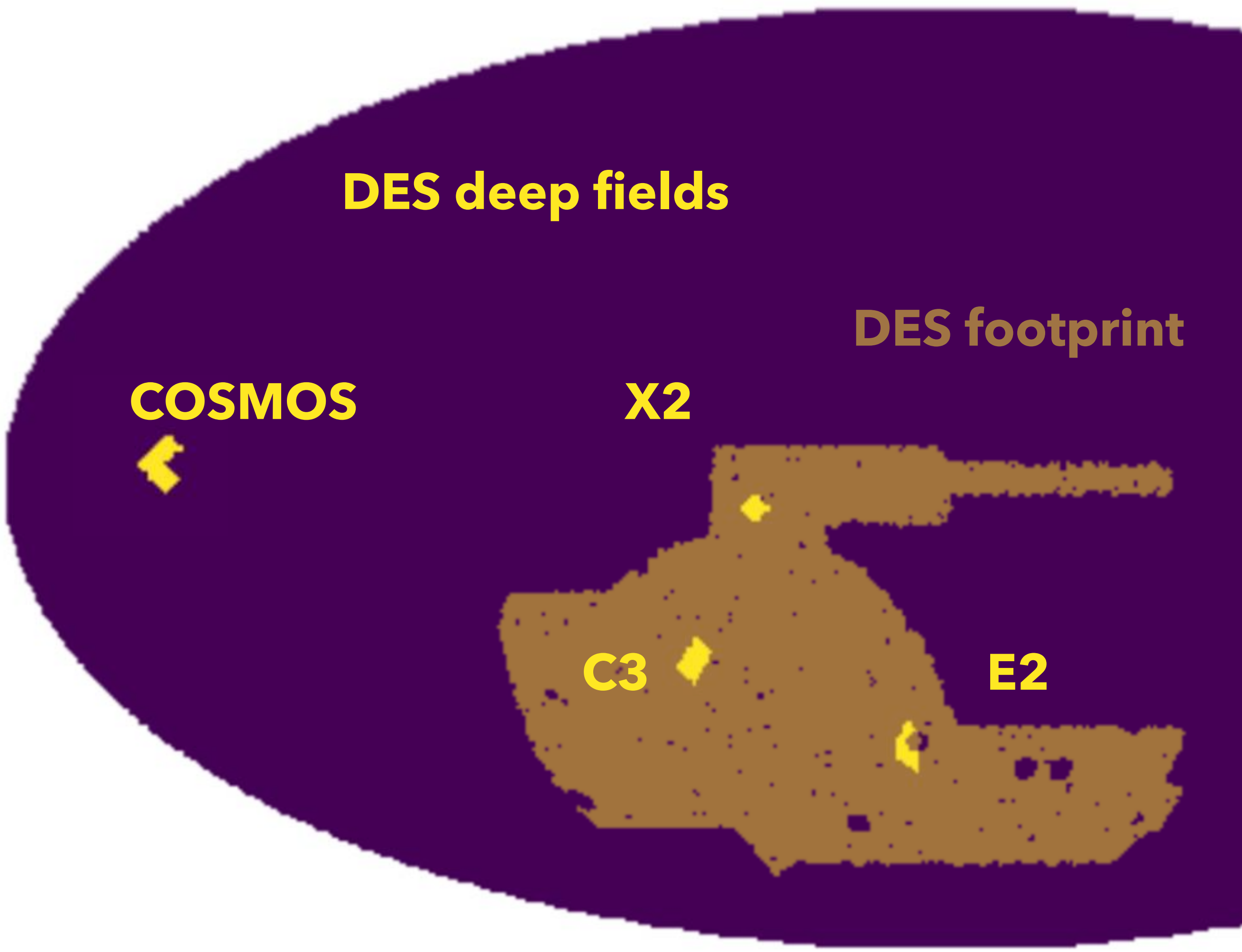
Buchs et Al., 2019

- Adding extra bands from DES or other surveys helps to break said degeneracy
- Extra bands are available only for a subset of the wide field galaxies

DEEP FIELDS

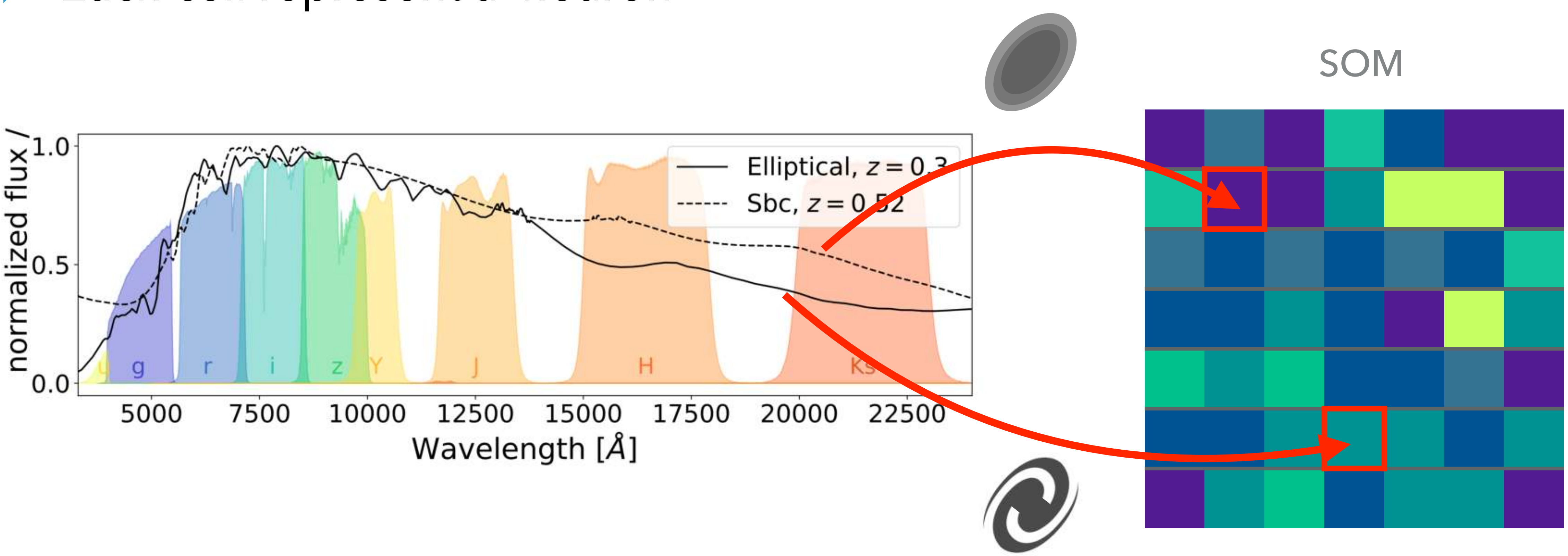
- Any galaxy that is detectable in the wide survey must be observed at high *S/N (at least 10× lower noise)* in the deep fields
    - In DES deep photometry is obtained in 10 supernova fields and in the COSMOS field
    - DES also conducted single-night observations in the *u* band and *Y* band

- Partial overlap from other surveys:
    - **COSMOS** (1.38 deg^2): **YJHKs** bands from deep VISTA measurements from the **UltraVista** survey
    - **Supernova fields C3, E2, X3** (3.32, 3.29, 1.94 deg^2): **YJHKs** bands from VISTA Deep Extragalactic Observations (**VIDEO**) survey
        - Y band not available in C3 field



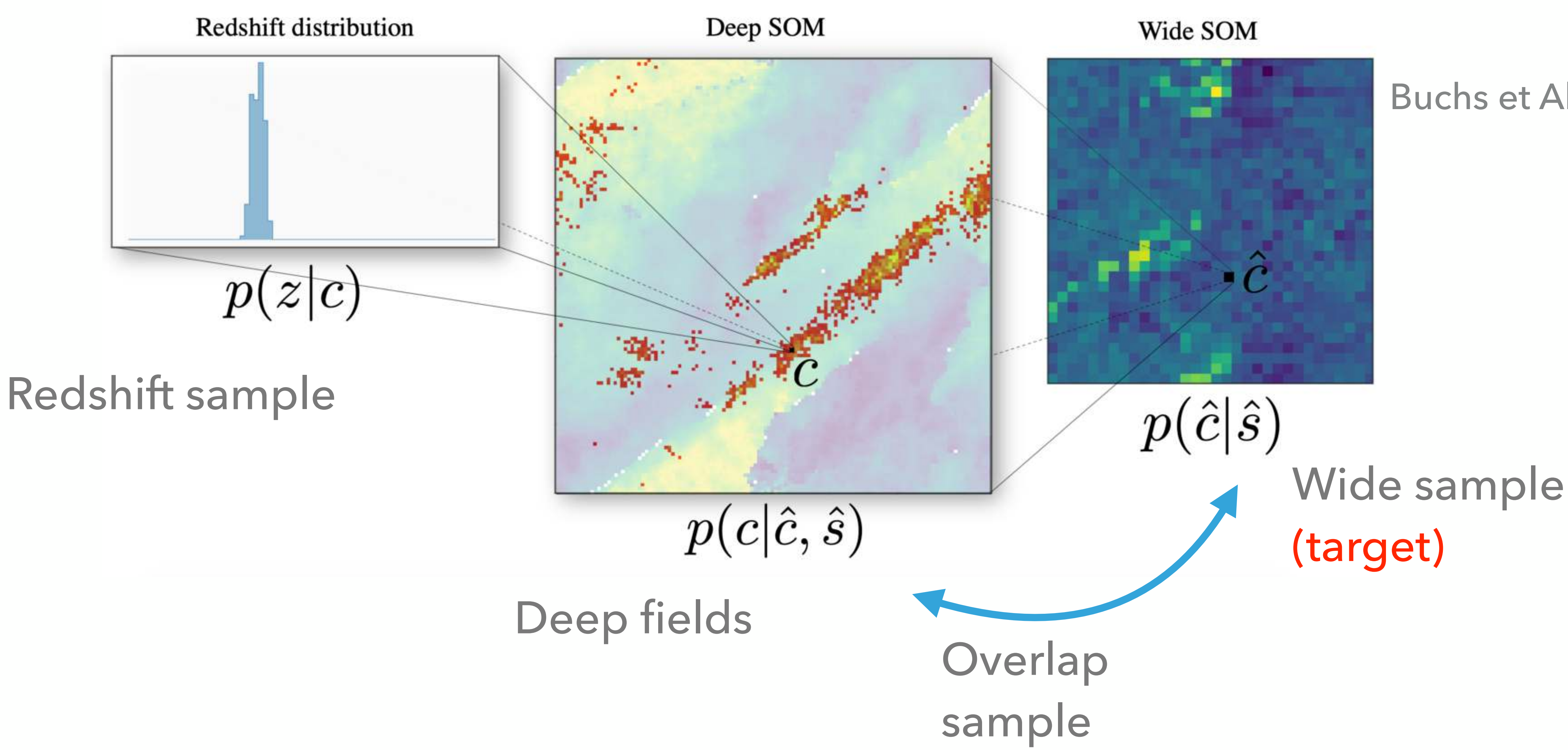
WHAT'S SOMPZ

- **Self Organising Map (SOM)** is a unsupervised artificial neural network that produces a discretized and low-dimensional representation of the input space
- In the SOMPZ method SOMs are used to classify galaxies in phenotypes according to their properties
    - Accurate phenotype classification is possible including the extra bands available in the deep fields
    - The topology of the grid was chosen to be square and period boundaries
    - Each cell represent a 'neuron'



SOMPZ PIPELINE

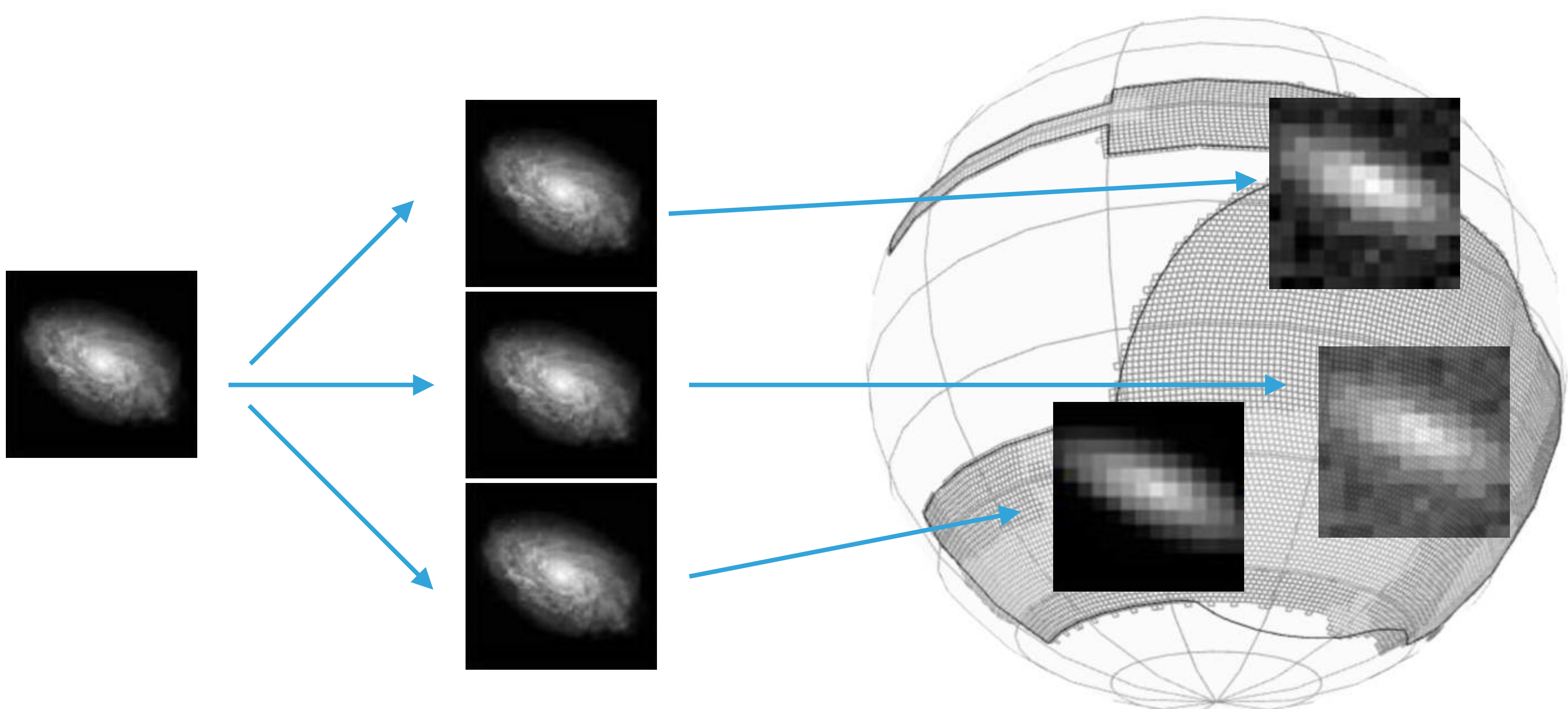
- Galaxies from wide and deep fields are grouped into phenotypes using SOMs
- The two SOMs are mapped through a transfer function
- Redshifts in the deep fields are validated through a high precision redshift sample
- Redshift information is finally transferred to the wide field



Buchs et Al., 2019

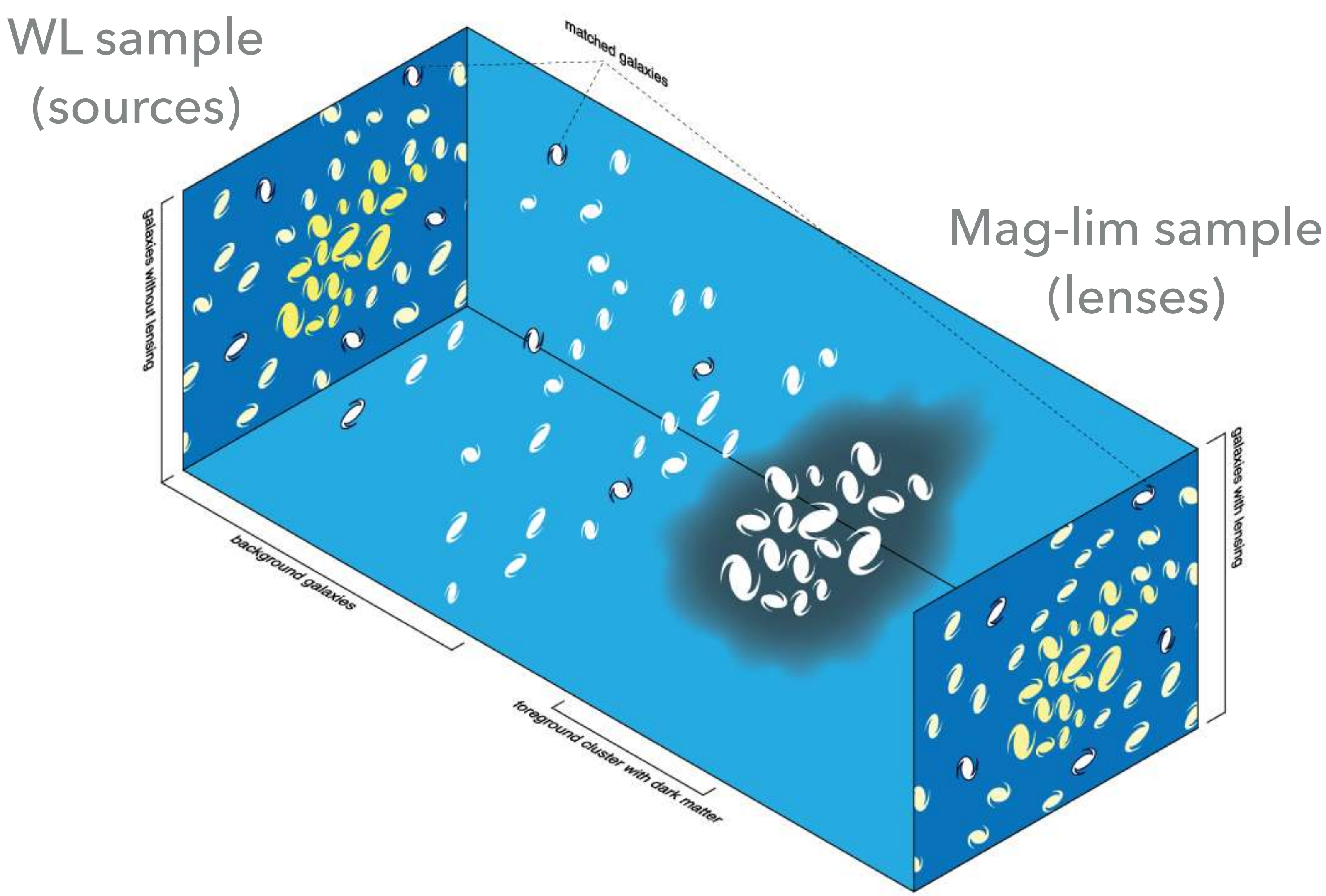
OVERLAP SAMPLE

- Necessary to transfer the redshift information from the deep to the wide field
- **Balrog** is a code creating noisy replicas of deep field galaxies by injecting them into real DES wide field images and processing them through the whole pipeline
  - Replicas of all deep field galaxies are pasted several times in random locations over the full wide footprint
  - An appropriate photometric error is assigned to each replica's fluxes depending on the location



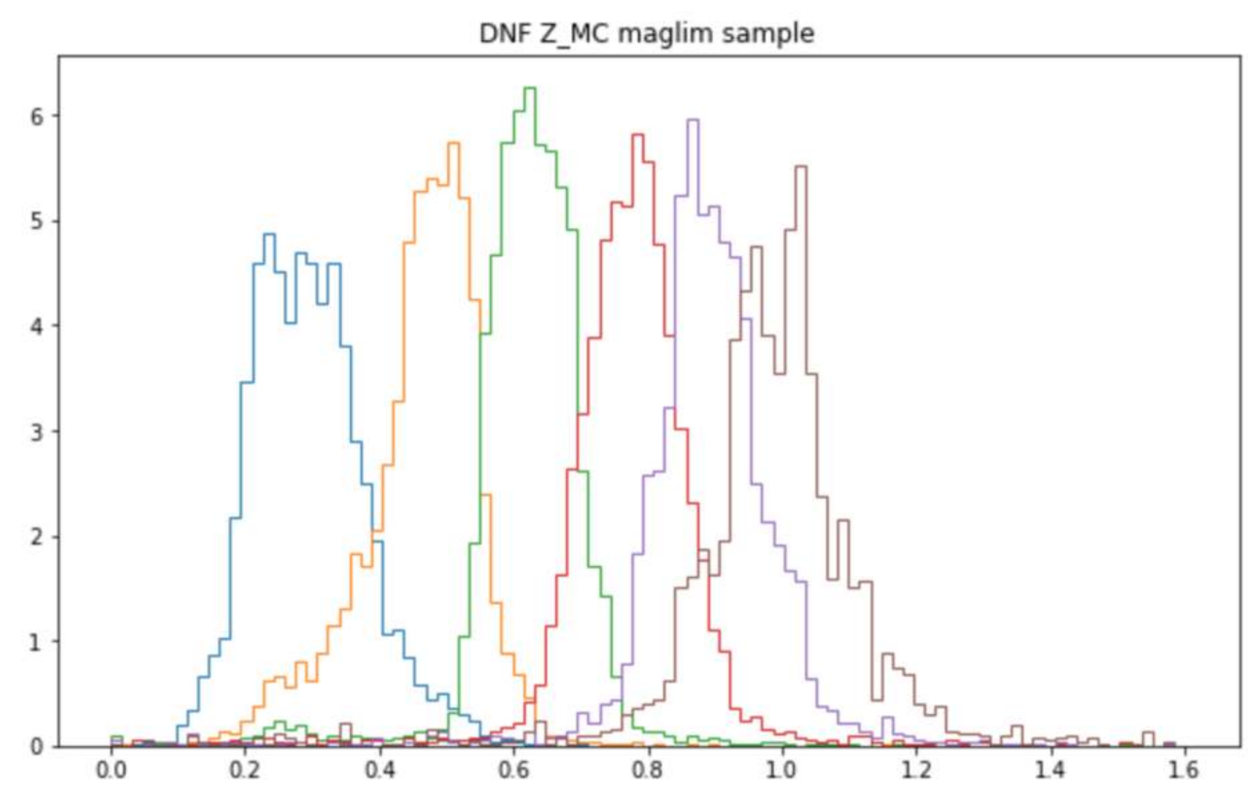
SOMPZ IN DES

- SOMPZ is among the techniques being used for the DES Y3 for 3x2 pt analysis:
    - Weak lensing (WL) sources catalogue
    - Magnitude-limited (mag-lim) lens catalogue



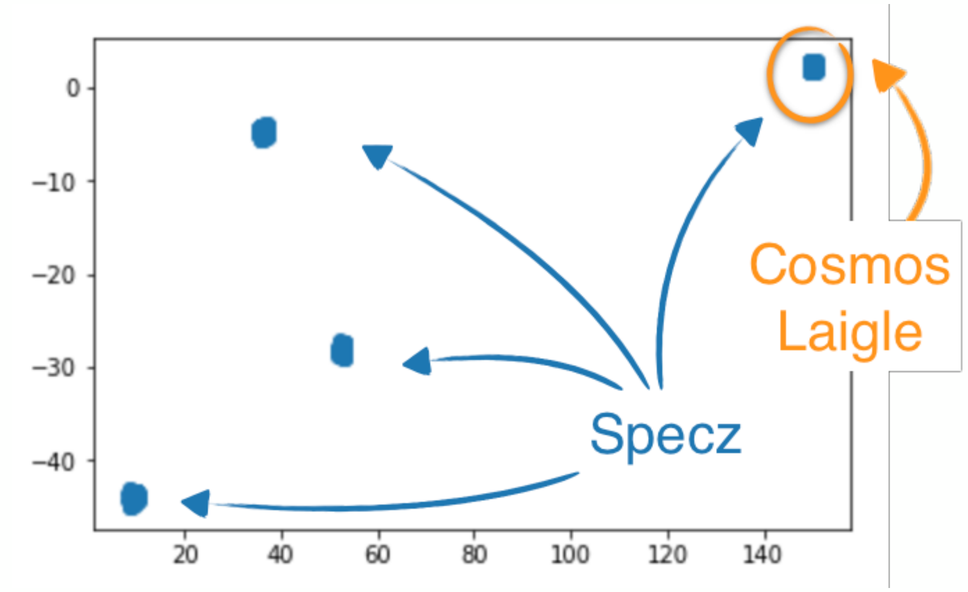
MAG-LIM SAMPLE

- One of the two lens samples to be used for 3x2pt analysis (as opposed to red bright galaxies redMaGiC)
  - Despite being noisier than redMaGiC, the mag-lim sample has a larger N density
- Bright magnitude limits at each tomographic bin imposed to reduce photo-z error
  - Sample selection is function of redshifts from the photometric code DNF:
    - mag_i < 4 * z + 18
    - mag_i > 17.5
- 6 tomographic bins:
  - Bin edges optimised for the cosmological analysis
  - Each bin treated independently (6 different SOMs)

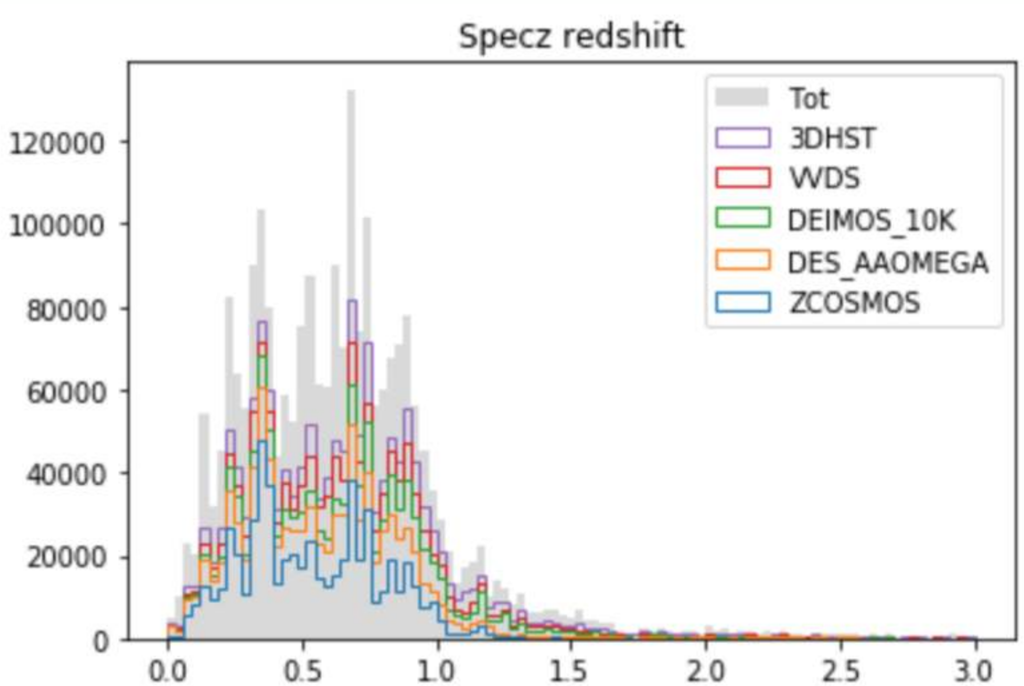


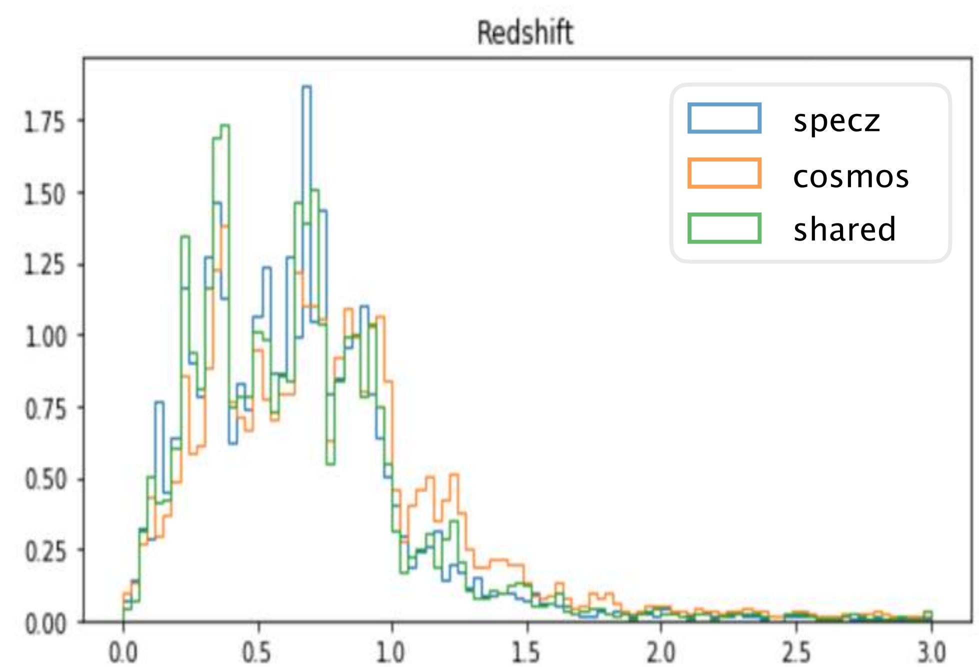
REDSHIFT SAMPLES

1) Combination of many **spectroscopic** surveys matched to deep fields galaxies (DES_AAOMEGA, ZCOSMOS, SDSS_DR14, DEIMOS_10K, etc.)
   - N after mag-lim sel: 116641



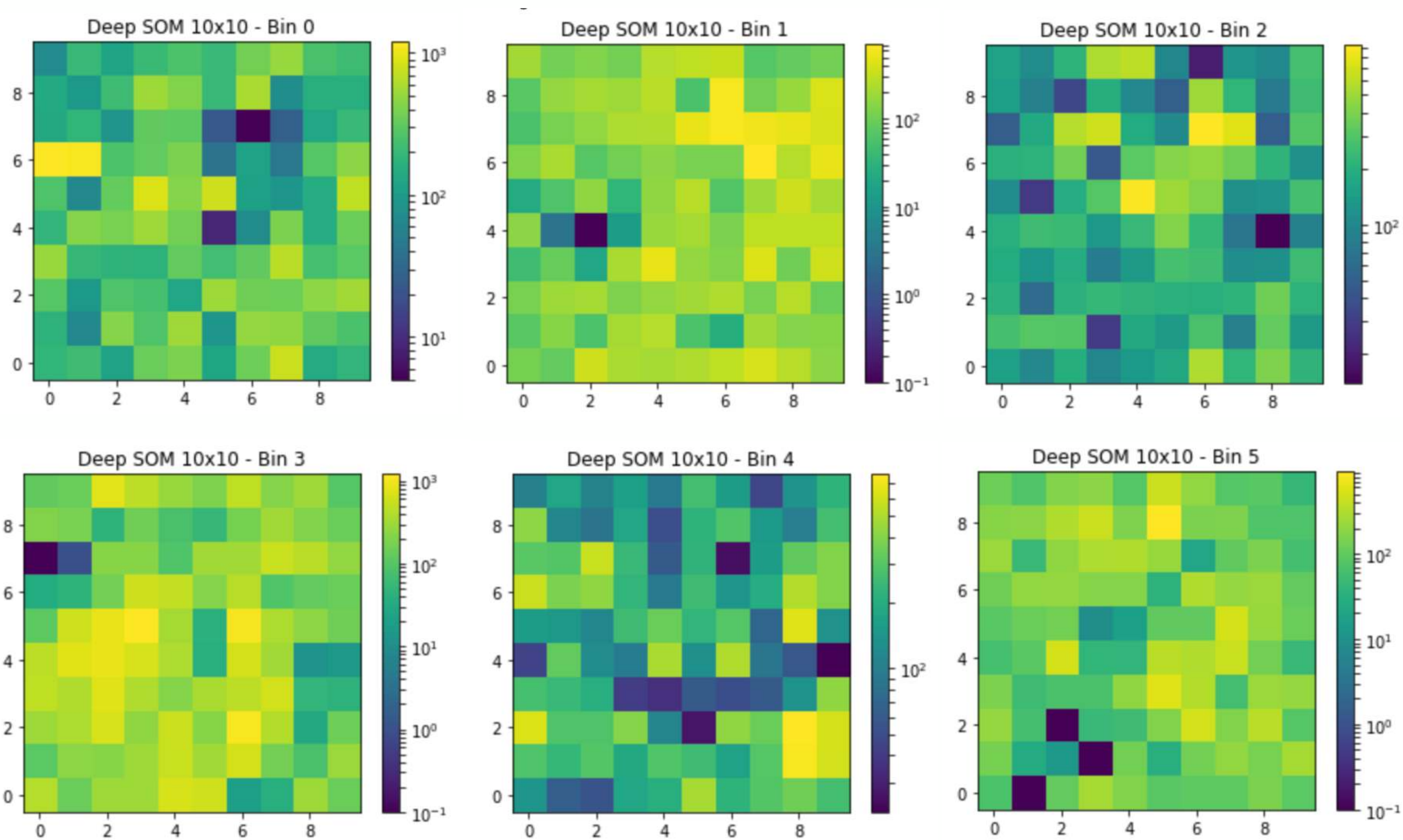
2) **COSMOS2015** 30-band photometric redshift catalog
   - N after mag-lim sel: 62362



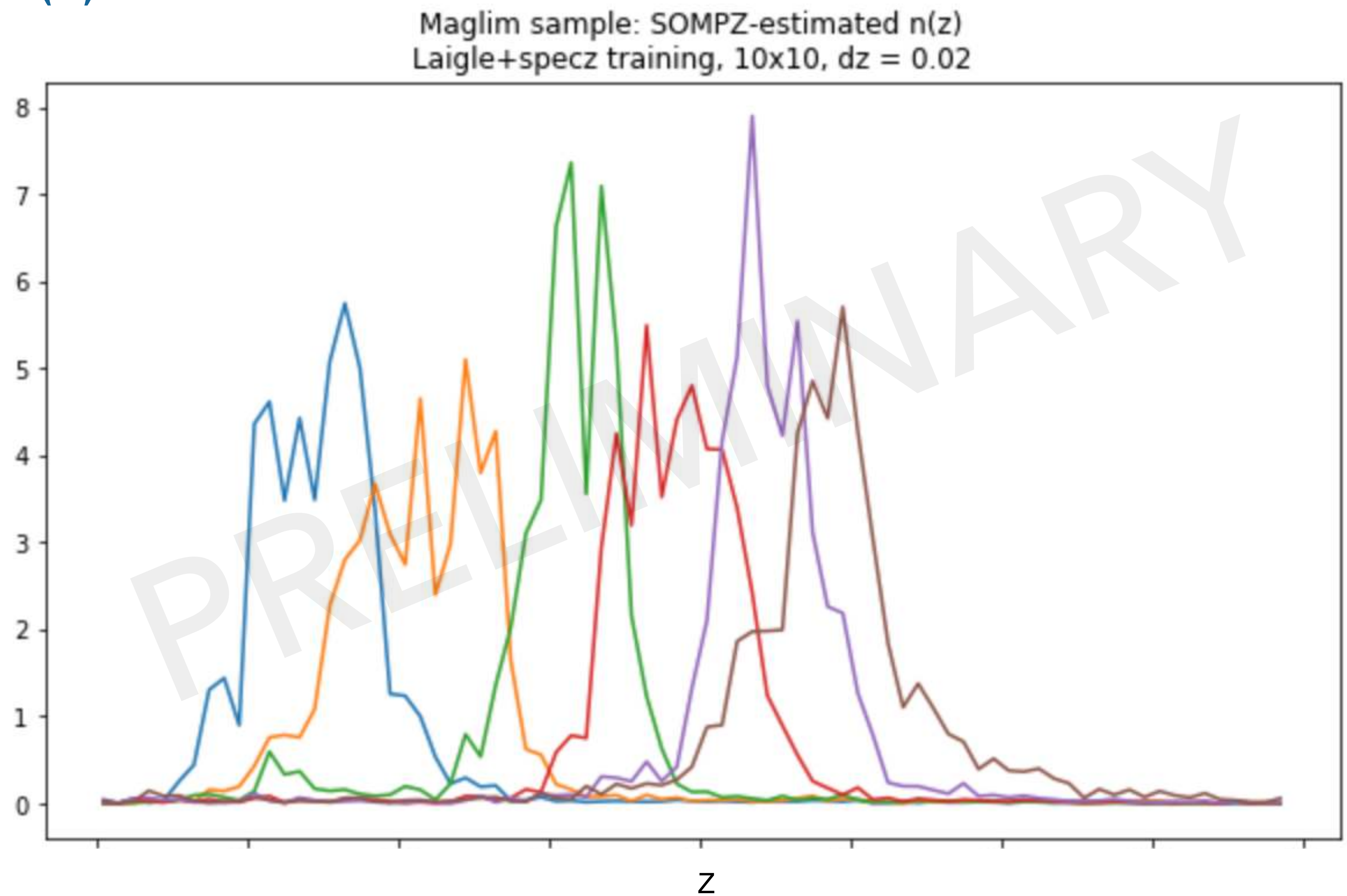


DEEP SOMS

- Combined redshift sample (specz + cosmos) for each tomographic bin assigned to deep SOMs of dimension 10x10 cells



SOMPZ N(Z)'S



- Preliminary n(z)'s in data obtained using the combined redshift sample
- First comparison with other methods (clustering-z, spectroscopic validation) show agreement within dz = 0.01
- Currently finalizing the pipeline in simulations for validation

UNCERTAINTIES

We identify the following statistical and systematic uncertainties:

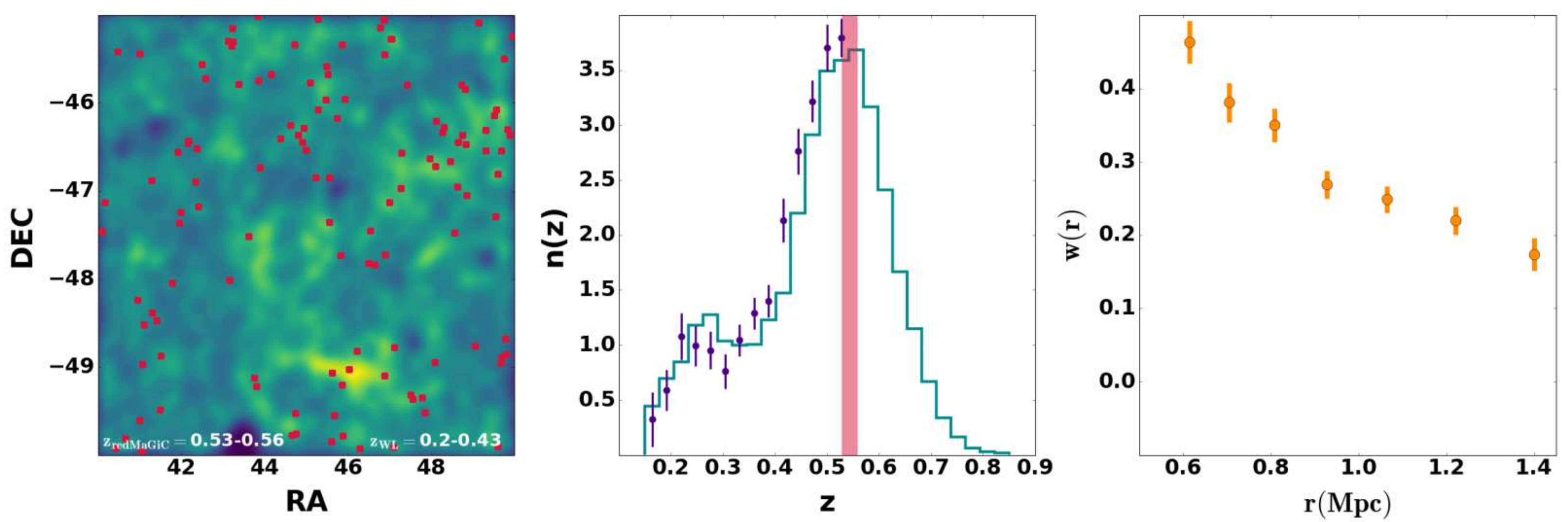
- **Cosmic variance systematics:** the deep fields have a small size (~10 deg^2) and might not be good representations of the Universe
- **Shot noise:** the redshift sample used have a limited number of galaxies, therefore in some cells the number of galaxies will be small
- **Redshift sample systematics:** inaccuracies propagated by the uncertainties of the redshift sample used

OUTLINE

- Motivation
- Dark Energy Survey
- Application of redshift estimation methods to DES Y3:
  - SOMPZ (Self-Organizing Maps Photo-Z): Machine learning neural network based technique
  - **Clustering redshifts:** Cross correlation of the target sample with a high quality redshift sample
- Summary

WHAT'S CLUSTERING-Z?

- Clustering redshifts method allows to estimate the redshift distribution of a target sample ("unknown") by exploiting the cross-correlation signal with a spatially overlapping "reference" sample with good redshifts
    - Divide reference sample with spec-z or high quality and accurate photo-z's into small redshift bins
    - Measure the cross-correlation signal with the given science sample



credit: P.Vielzeuf

REFERENCE SAMPLES

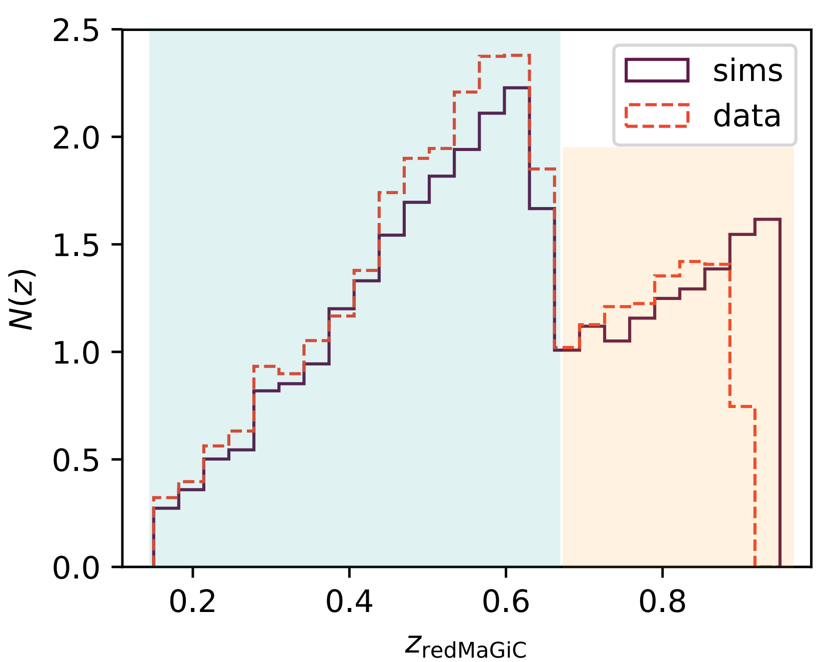
1) REDMAGIC

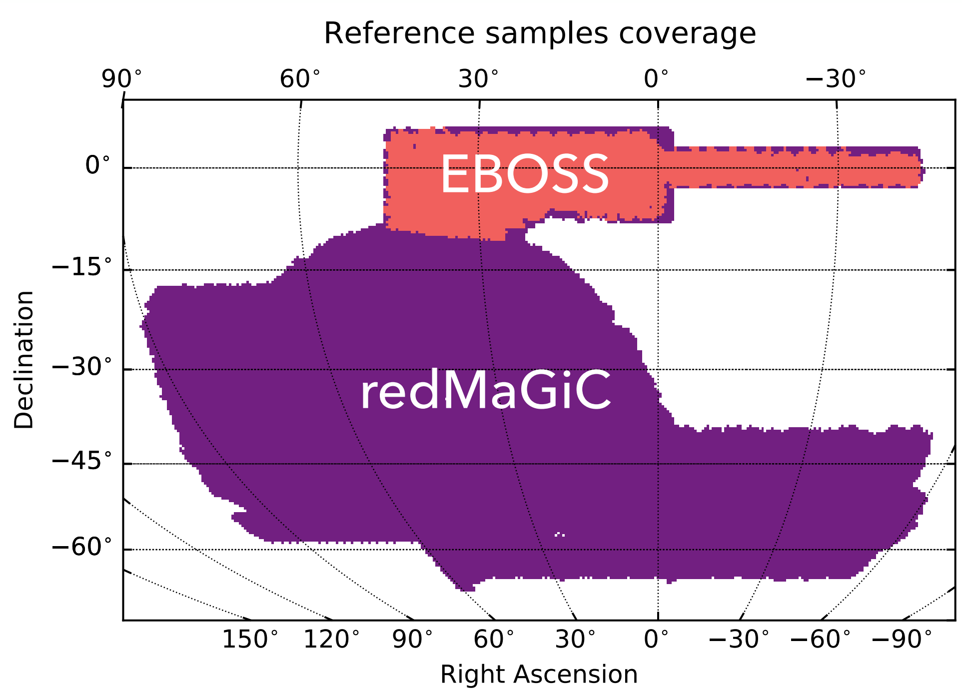
- Red luminous galaxies with high quality photometric redshift estimates
- 25 bins, 0.15 < z < 0.95
- Combination of two samples selected with different properties:
  - ✦ high density (L/L* > 0.5)
  - ✦ high luminosity (L/L* > 1)

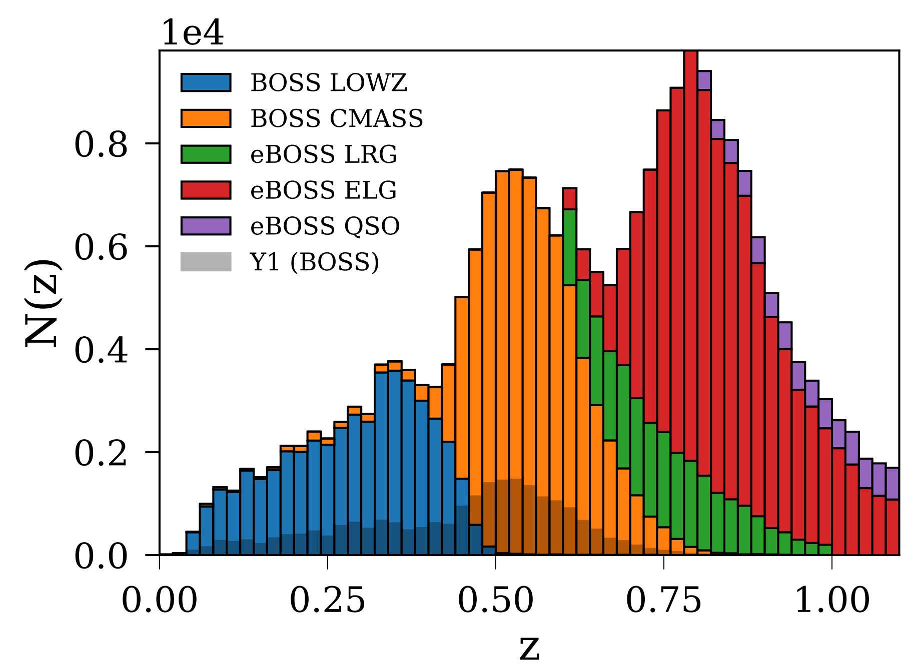
2) BOSS/EBOSS

- Spectroscopic redshifts from BOSS and eBOSS galaxies
- 15 bins, 0.1 < z < 1.1
  - Larger statistical error (250k galaxies in z range 0.1-1.1)
  - Less overlap with DES
  - Higher z coverage!
  - Spectroscopic!

redMaGiC spread (bias)







CLUSTERING-Z MODELLING

- Clustering-z expression can be approximated as

$$\omega_{ur}(\theta) = \int dz \; n(z)_u \; n(z)_r \; b(z)_u \; b(z)_r \; \omega(\theta)_{DM} - M(z, \theta)$$

- We can approximate $n(z)_r$ as a delta since the reference sample is binned in thin redshift bins

$$\omega_{ur}(\theta) \propto \; n(z)_u \; b(z)_u \; b(z)_r \; \omega(\theta)_{DM} - M(z, \theta)$$

goal: WL redshift distributions!

magnification effects

galaxy-matter bias of WL sample

dark matter clustering

galaxy-matter bias of reference sample

SYSTEMATIC UNCERTAINTIES

We identify the following systematics:

- **WL bias evolution systematic:** the clustering-based estimator ignores the redshift evolution of the galaxy-matter bias of the WL sample
- **red magic photo-z systematic:** photo-z as opposed to true redshifts are used to bin one of the reference sample
- **methodology systematic:** others ,e.g. inaccuracies in the modelling at small scales or lack of modelling (e.g. lensing magnification)

WZ LIKELIHOOD

- We forward model the full clustering signal, using as n(z) each SOMPZ realisation
- We compute the likelihood with the measured cross-correlation signal $w_{ur}(z_i)$ in data

$$\mathscr{L} = -\frac{1}{2}\chi^2\left(\omega_{ur};\ Sys(z, s_i)\ \hat{\omega}_{ur}(n^{pz,k}, p_i);\ \hat{\Sigma}^{-1}_{wz}\right)$$

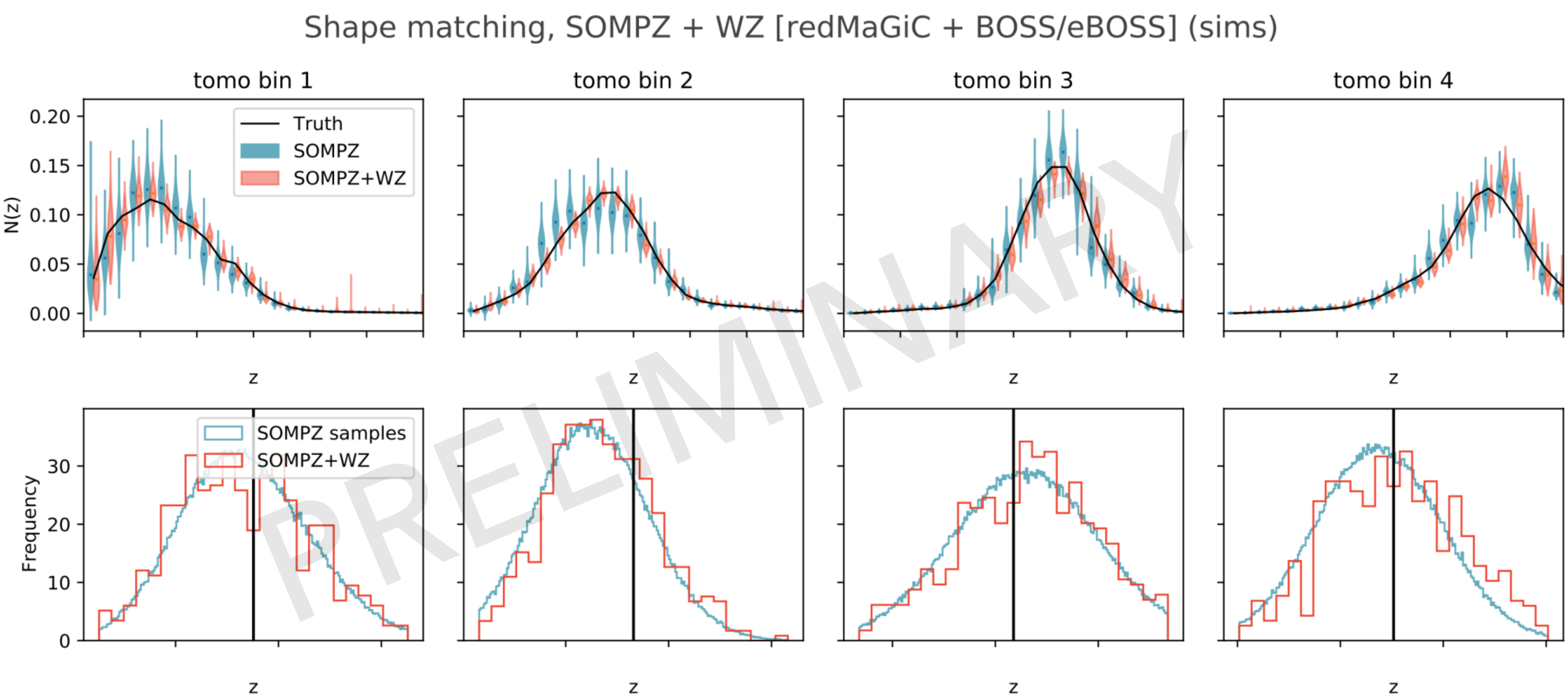
$$log[\mathrm{Sys}\left(\mathrm{z_i}, \{\mathrm{s_k}\}\right)] = \sum_{k<M} s_k P_k(z_i)$$

$P_k(z)$ Legendre Polynomials

- The model $\hat{\omega}$ is multiplied by a smooth function Sys(z, $\{s_i\}$) that accounts for all the systematic uncertainties of the method
- Sys(z, $\{s_i\}$) is flexible enough to fully describe all the systematic uncertainties found in the simulation
- The nuisance parameters $\{s_i\}$ are the parameters of the systematic function

COMBINATION WITH SOMPZ

- SOMPZ produces N realisation of the n(z)'s, encompassing all the uncertainties
- Those realisations are filtered using clustering-z, selecting those that maximise a joint likelihood of SOMPZ and clustering redshift estimates
- WZ helps to tighten the scatter on the shape of the n(z)'s while it's not able to reduce the scatter around the mean



SUMMARY

- Redshifts estimation is currently one of the limiting factor in wide field photometric galaxy surveys and novel techniques are being developed for the DES Y3 cosmological analysis
- SOMPZ (Self-Organizing Maps Photo-Z):
  - SOMs are used to classify galaxies in phenotypes depending on their properties
  - High-quality z and data from the DES deep fields (where additional photometry is available) are used to break type-redshift degeneracies for the wide field galaxies
- Clustering redshifts (WZ):
  - Cross correlation of a science sample with two high quality redshift samples binned in thin redshift bins
  - WZ likelihood is built by forward modelling the clustering signal and estimating the systematic uncertainties through a smooth function
- For the WL sample, SOMPZ and WZ's likelihoods are combined

THANK YOU FOR YOUR ATTENTION AND SEE YOU AT THE DISCUSSION!

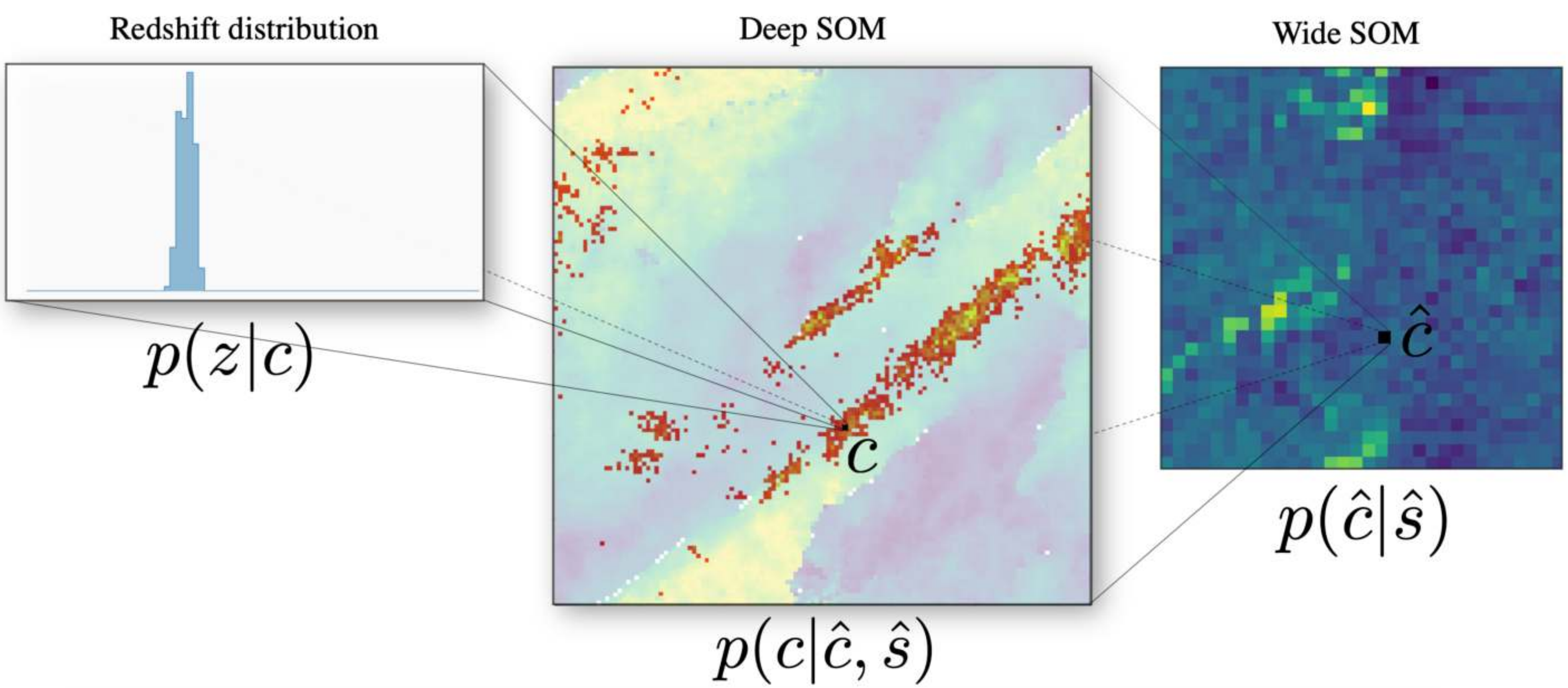
The project leading to these results have received funding from “la Caixa” Foundation (ID 100010434), under the agreement LCF/BQ/DI17/11620053

"la Caixa" Foundation

BACK UP SLIDES

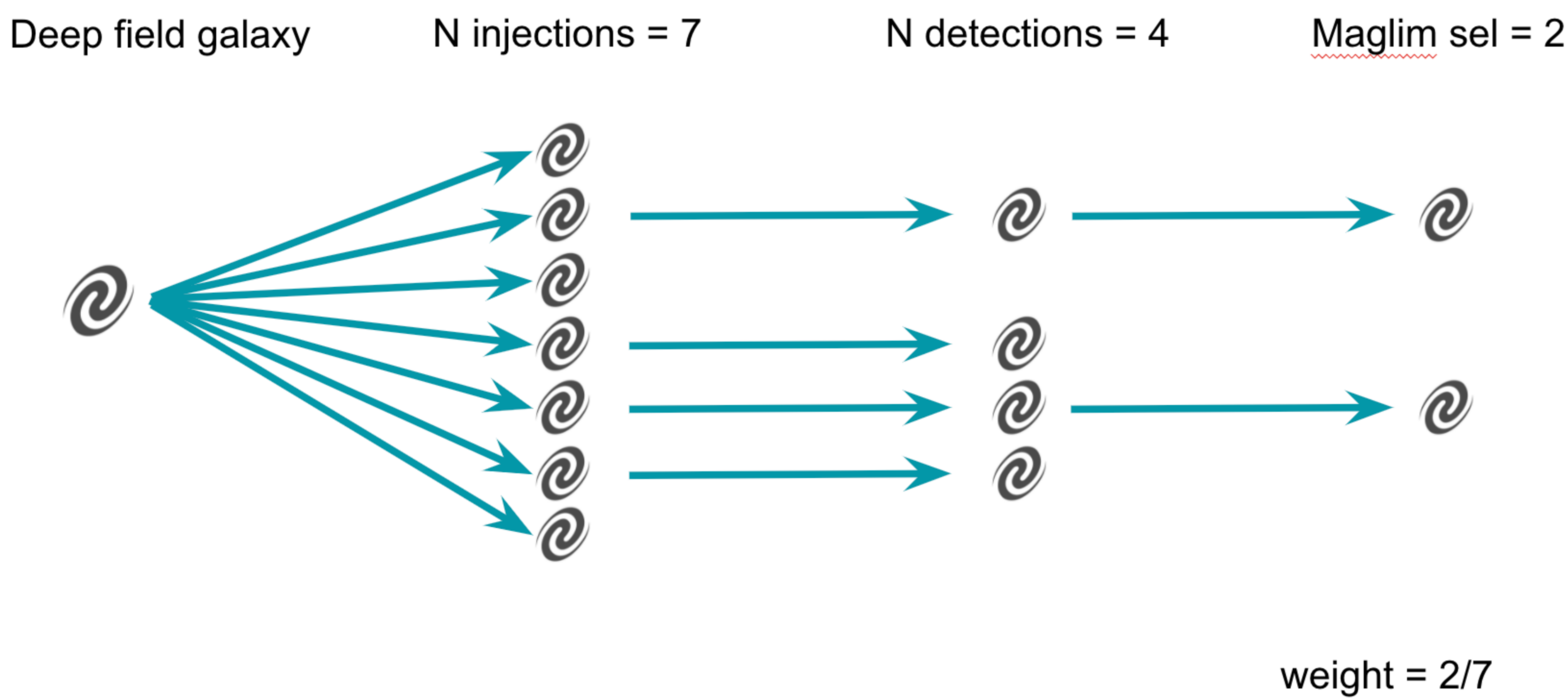
SOMPZ PIPELINE

- Galaxies from wide and deep fields are grouped into phenotypes using SOMs
- The two SOMs are mapped through a transfer function
- Redshifts in the deep fields are validated through a high precision redshift sample.
- Redshift information is finally transferred to the wide field



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

SOMPZ OVERLAP SAMPLE WEIGHTS



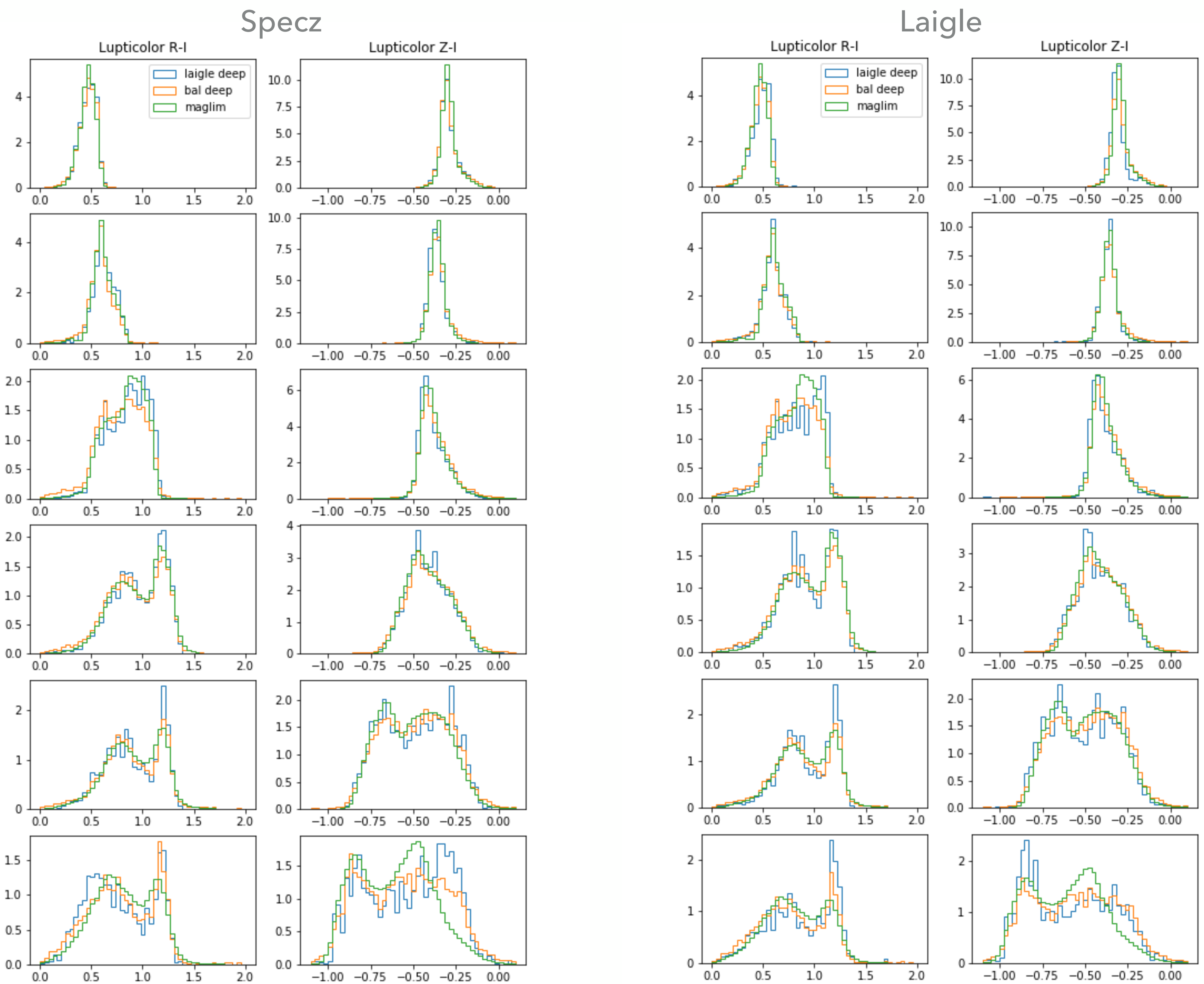
LUPTICOLORS

- SOM trainings and assignments are done using lupticolors, which are difference of luptitudes μ:

$$\mu_x = \mu_0 - a \sinh^{-1}(\frac{f_x}{2b})$$

- The zeropoint is μ0 = m0 − 2.5logb,
- a = 2.5log e and b is a softening parameter that sets the scale at which luptitudes transition between logarithmic and linear behaviour

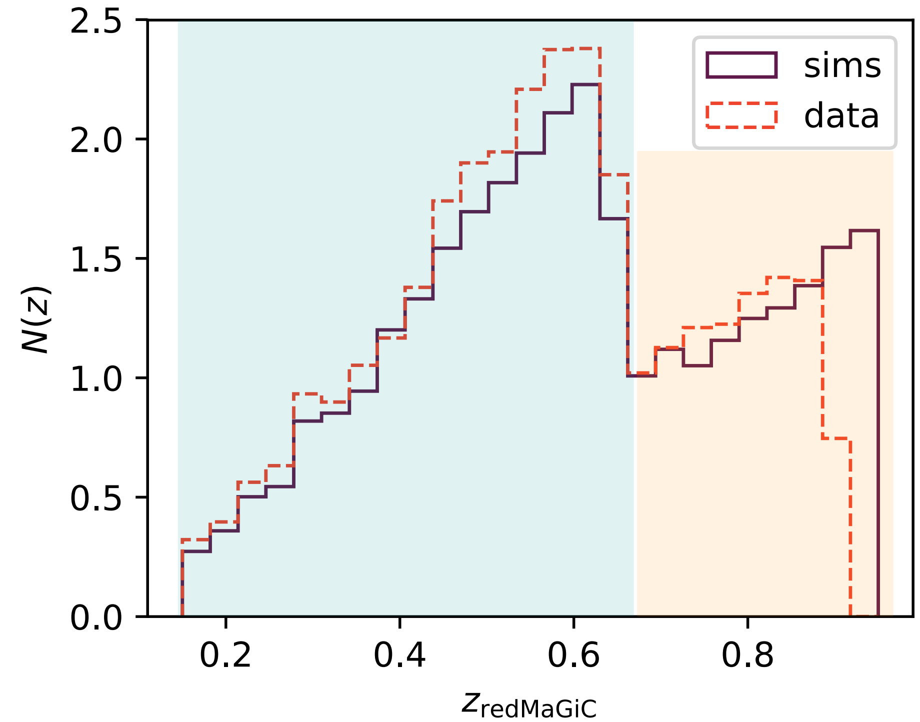
LUPTICOLORS OF DEEP, WIDE AND Z SAMPLE

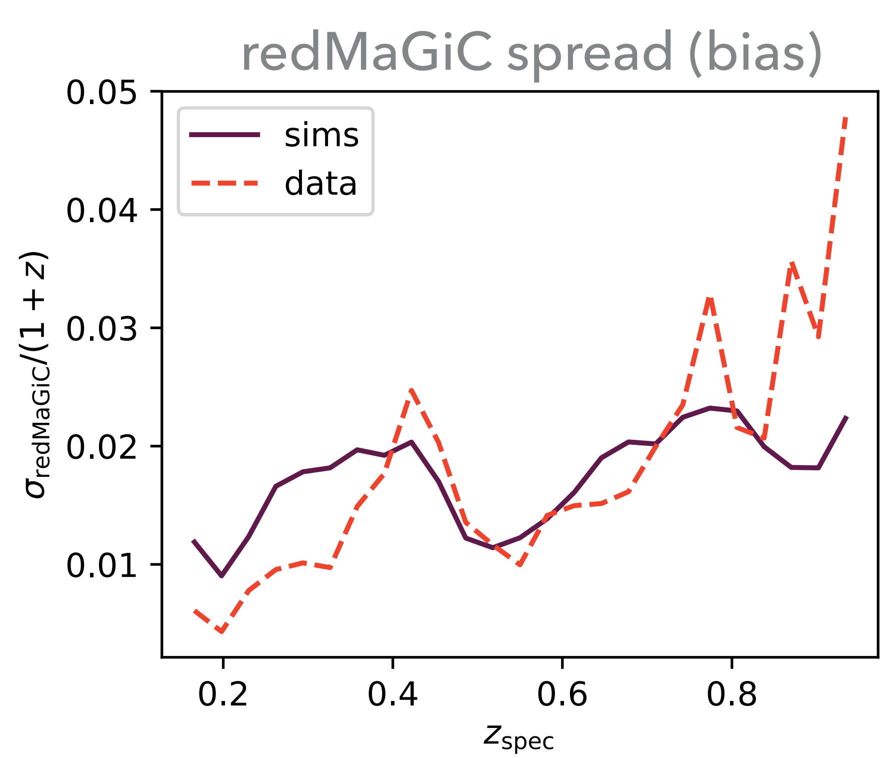


REFERENCE SAMPLES

1) REDMAGIC

- Red luminous galaxies with high quality photometric redshift estimates
- 25 bins, 0.15 < z < 0.95
- Three samples are selected with different:
    - high density (L/L* > 0.5)
    - high luminosity (L/L* > 1)





MAGNIFICATION

- Total overdensity: $\delta_{obs} = \delta_g + \delta_\mu$ (g = gravitational, μ = magnification)
- Cross-correlation has contribution from 4 terms:

Gravitational term only
(dominant, what we call clustering-z)

Magnification in reference sample
x clustering in unknown sample

$$w_{ur,tot}(\theta) =< \delta_{g,u}, \delta_{g,r} > (\theta)+ < \delta_{g,u}, \delta_{\mu,r} > (\theta)+ < \delta_{g,r}, \delta_{\mu,u} > (\theta)+ < \delta_{\mu,u}, \delta_{\mu,r} > (\theta).$$

Magnification in unknown sample
x clustering in reference sample

Magnification in both samples
(can be neglected)