# *Euclid*: constraining ensemble photometric redshift distributions with stacked spectroscopy

Università degli Studi di Milano - Dipartimento di Fisica

Marina S. Cagliari
(marina.cagliari@unimi.it)
Ben R. Granett

Barolo PRIN Meeting
September 2021

# Objectives

- We built a machine implementation of the *ensemble photometric redshifts* method proposed by Padmanabhan et al. (2019).

- Redshift distributions are needed in cosmology to analyse photometric samples of galaxies:
  - *to extract information from large-scale structure and galaxy clustering*,
  - to make tomographic studies with weak lensing.

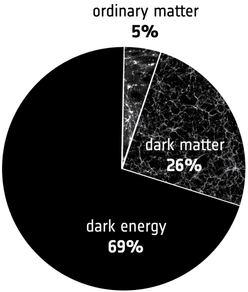



Figure credits: ESA

# Correlation Function

- The galaxy two-point correlation function is used to study large-scale structures.
- The angular correlation function is related to the 3D correlation function through the redshift distribution.
- It has a power law form, $\xi(s) \sim s^{-\gamma}$, and a peak due to Baryon Acoustic Oscillations (BAO).

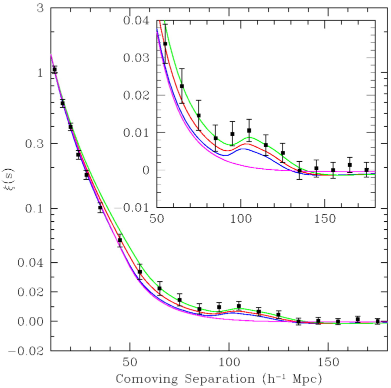
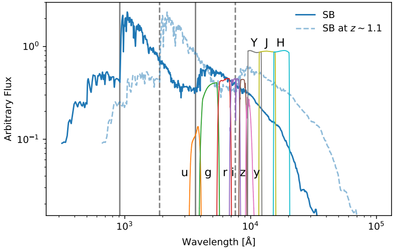
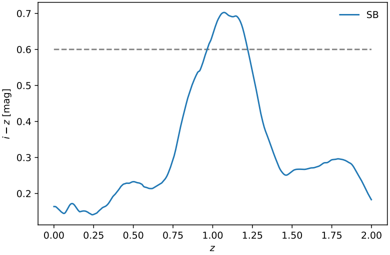



Figure from Eisenstein et al. (2005)

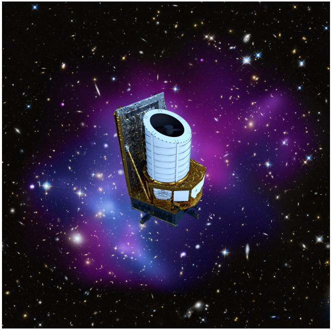
# Redshift Measurements




- Redshifts can be measured either with spectroscopy or photometry.
- Photemtric measurements are deeper and faster than spectroscopic ones.
- Photometric redshifts is calibrated with spectroscopic redshift in order to measure redshift distributions.

# Euclid



Credits: ESA/ATG medialab; NASA, ESA, CXC, Ma et al. and STScI

- Euclid is an upcoming ESA space mission, it will combine imaging and slitless spectroscopy surveys.
- Redshift distributions are used both in weak lensing analyses and BAO measurements.

# Euclid data

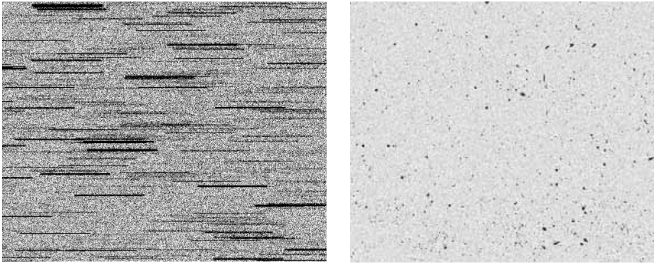
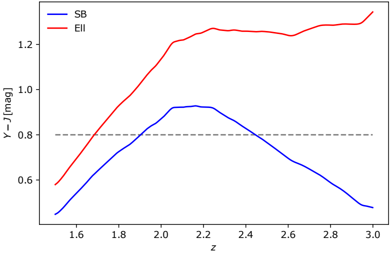
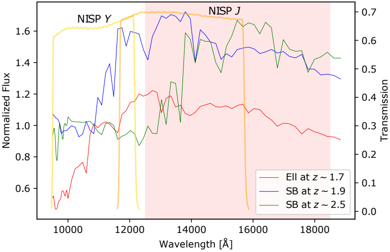


Figure from Laureijs et al. (2011)

- It will obtain NIR photometry and a shallow spectrum for every galaxy.
- It will not be able to measure the spectroscopic redshift of every galaxy.

# Ensemble photometric redshifts




- The ensemble photometric redshifts method aims to constrain the redshift distribution of a *photometrically-selected galaxy sample* (colour group) by using the *stacked spectrum* built from the average of many low signal-to-noise spectra.

# Stacked spectrum

- The stacked spectrum of a colour group is:

$$f_{\rm stack}^{\rm obs}(\lambda) = \frac{1}{N_{\rm gal}} \sum_{i=1}^{N_{\rm gal}} \frac{1}{\bar{f}_i} f_i(\lambda)\,, \quad \text{where } f(\lambda) = a\, T(\lambda|z)\,.$$

- Modelling the galaxy SEDs with a finite set of parameters:

$$f_{\rm stack}^{\rm model}(\lambda) = \sum_{\alpha=1}^{N_{\rm SED}} \sum_{z=z_{\rm min}}^{z_{\rm max}} p_{\alpha,z}\, T_{\alpha,z}(\lambda)\,.$$

- The redshift distribution is:

$$p_z \propto \sum_{\alpha=1}^{N_{\rm SED}} p_{\alpha,z}\,.$$

# Machine implementation

$$\begin{pmatrix} T_1^{z_1}(\lambda_1) & T_2^{z_1}(\lambda_1) & T_1^{z_2}(\lambda_1) & T_2^{z_2}(\lambda_1) \\ T_1^{z_1}(\lambda_2) & T_2^{z_1}(\lambda_2) & T_1^{z_2}(\lambda_2) & T_2^{z_2}(\lambda_2) \\ T_1^{z_1}(\lambda_3) & T_2^{z_1}(\lambda_3) & T_1^{z_2}(\lambda_3) & T_2^{z_2}(\lambda_3) \end{pmatrix} \begin{pmatrix} p_1^{z_1} \\ p_2^{z_1} \\ p_1^{z_2} \\ p_2^{z_2} \end{pmatrix} = \begin{pmatrix} f_s(\lambda_1) \\ f_s(\lambda_2) \\ f_s(\lambda_3) \end{pmatrix}$$

- Operatively we build a template matrix, however the problem is under-constrained and do not have a unique solution.
- We want to find a minimum set of templates that fits the stacked spectrum $\longrightarrow$ linear regression problem with non-negative constrain.

# Method's performance

- We compare the BAO angular position computed with the real redshift distribution and the fitted one:

$$w(\vartheta) = \int \frac{\mathrm{d}\ell\, \ell}{2\pi} J_0(\ell\vartheta) \int \mathrm{d}r \frac{\left[p(z)\frac{\mathrm{d}z}{\mathrm{d}r}\right]^2}{r^2} P_\mathrm{g}\left(\frac{\ell+\frac{1}{2}}{r}, z|\Omega\right) .$$

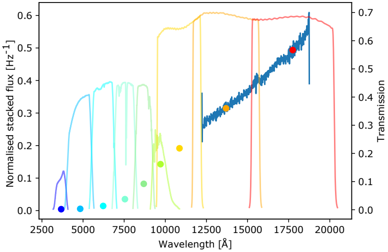
- The method error is quantified by:

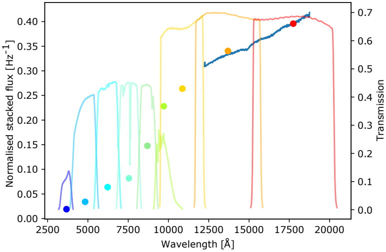
$$\left\langle\left|\frac{\Delta\vartheta_\mathrm{BAO}}{\vartheta_\mathrm{BAO}}\right|\right\rangle = \left\langle\left|\frac{\vartheta^\mathrm{real}_\mathrm{BAO} - \vartheta^\mathrm{fit}_\mathrm{BAO}}{\vartheta^\mathrm{real}_\mathrm{BAO}}\right|\right\rangle_\text{colour groups} .$$

# Euclid Flagship

- We base the synthetic photometric and spectroscopic data on the Euclid Flagship mock galaxy catalogue.
- We built three catalogues with different internal galaxy attenuation models:
  - *non-attenuated*: without galaxy attenuation,
  - *fixed* attenuation: galaxy attenuation is fixed and known,
  - *real* attenuation: galaxy attenuation is free and unknown.
- Each catalogue needs a specific template set in order to compute the redshift distributions.

# Data & Errors

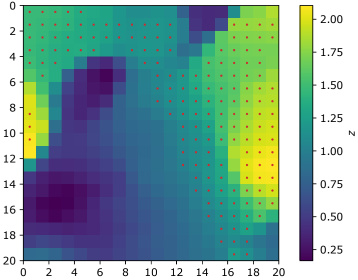



- Spectroscopy in the Euclid range ($1.25 \div 1.85\,\mu\mathrm{m}$), and photometry in the three Euclid NISP filters (*YJH*) and the six ones of the Vera C. Rubin observatory (*ugrizy*) are known.
- Observational error is generated with Gaussian extraction. Two different error models are assumed for photometry and slitless spectroscopy.

# Colour group division

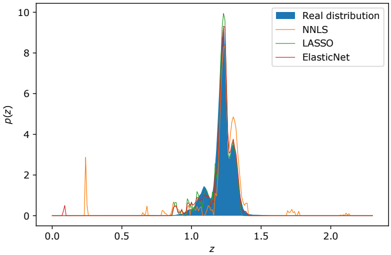
- Self-organising maps (SOMs) are an unsupervised machine-learning algorithm, which projects high-dimensional data onto a lower-dimensional grid.
- We use a SOM to divide the galaxies by their colours. The cells of the SOM are the colour groups.



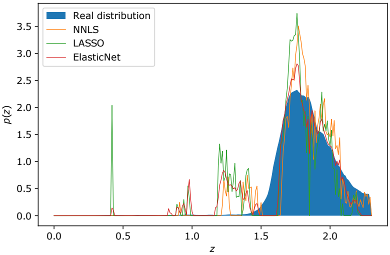

Red spots:
$z_{\text{mean}} > 1 \wedge \sigma_z < 0.2$

# Analyses without spectroscopic noise

## Non-attenuated catalogue

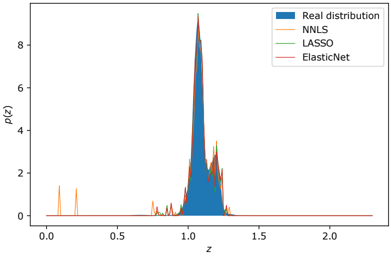



## Real attenuation catalogue

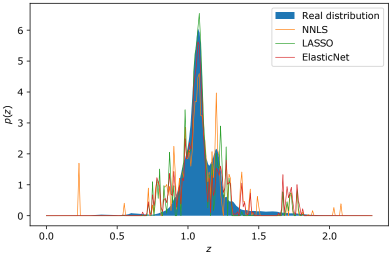


# Analyses with spectroscopic noise

## Non-attenuated catalogue




## Real attenuation catalogue



# Results

| Analysis without spectroscopic noise | | | |
|---|---|---|---|
| | NNLS | LASSO | ElasticNet |
| *non-attenuated* | 0.0069 | 0.0062 | 0.0053 |
| *fixed* | 0.0086 | 0.0050 | 0.0040 |
| *real* | 0.019 | 0.021 | 0.014 |
| **Analysis with spectroscopic noise** | | | |
| | NNLS | LASSO | ElasticNet |
| *non-attenuated* | 0.0073 | 0.0071 | 0.0063 |
| *fixed* | 0.0099 | 0.0066 | 0.0057 |
| *real* | 0.016 | 0.023 | 0.015 |

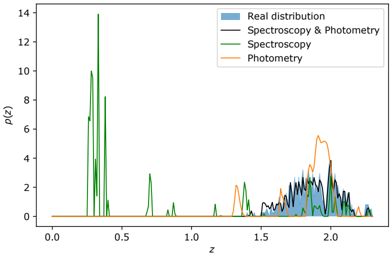
- Attenuation is the main limitation to the fit performance.
- Noise induces a 10% loss in the method precision introducing an additional degeneracy.

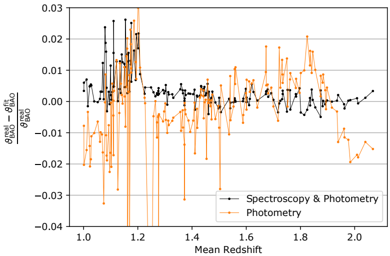
# Conclusions

- In ideal condition the method is able to reconstruct with great detail the redshift distributions.
- The accuracy of the redshift distribution estimation is *limited primarily by internal galaxy attenuation* and its modeling.
- We expect a further loss in precision when analysing real data.
- We hope to *improve the performance* by adding physical priors and optimizing the template set.

# Thanks for your Attention!

# Spectro-Photometry vs. Photrometry




$$\left\langle \left| \frac{\Delta\vartheta_{\mathrm{BAO}}}{\vartheta_{\mathrm{BAO}}} \right| \right\rangle = \begin{cases} 0.0044, & \text{for} \quad \text{spectro-photometry} \\ 0.010, & \text{for} \quad \text{photometry} \end{cases}$$