From DARK LIGHT to dark matter: understanding the galaxy-matter connection to measure the Universe

A view from Milan

Luigi Guzzo

ROMA TRE UNIVERSITÀ DEGLI STUDI

UNIVERSITÀ DEGLI STUDI DI MILANO LA STATALE

UNIVERSITÀ DEGLI STUDI DI TORINO

Cutting the story short, the goals of this PRIN boil down to:

Cutting the story short, the goals of this PRIN boil down to:

A) Probe cosmology and dark energy from large-scale structure measurements

Cutting the story short, the goals of this PRIN boil down to:

A) Probe cosmology and dark energy from large-scale structure measurements

B) Connect dark matter, neutrinos and “dark baryons” to large-scale structure (cross-correlations, simulations)

Cutting the story short, the goals of this PRIN boil down to:

A) Probe cosmology and dark energy from large-scale structure measurements

B) Connect dark matter, neutrinos and "dark baryons" to large-scale structure (cross-correlations, simulations)

>>> For both themes, the emphasis was on developing innovative techniques in preparation for new survey data, and first applications

RU1 (UniMI & INAF MI) original PRIN 2017 goals

- Develop advanced 2pt clustering analysis of galaxy surveys:
    - advanced RSD modelling (Bianchi, LG)
    - forward-modelling cosmology, halo-galaxy match (Granett, Carbone, LG)

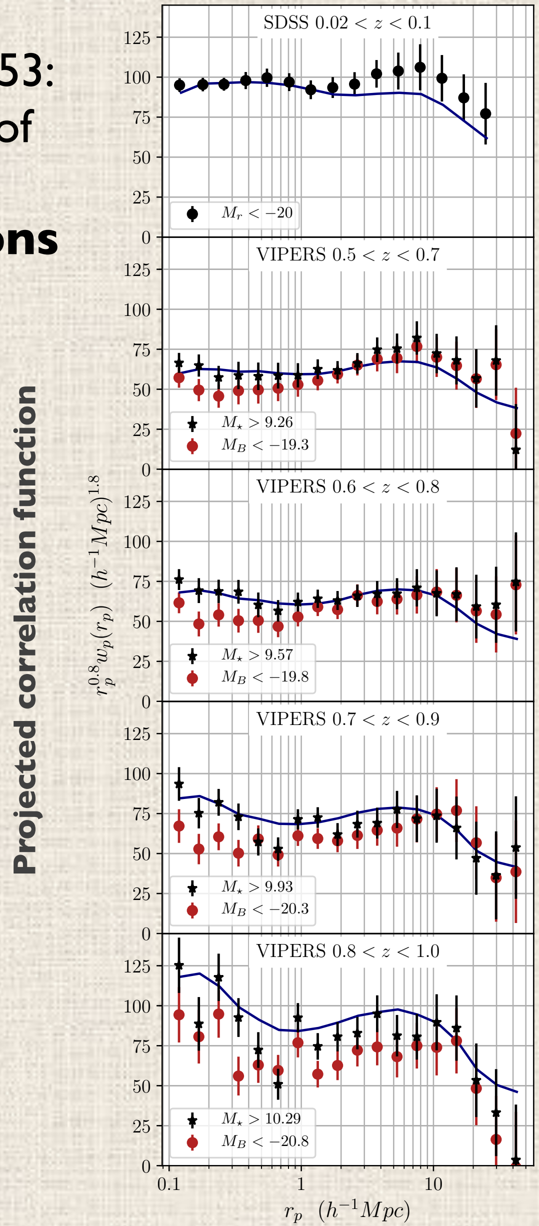
- Link to DM and particles:
    - n-body simulations in neutrino cosmologies (Carbone)

In fact, it enabled brand new directions to be explored…

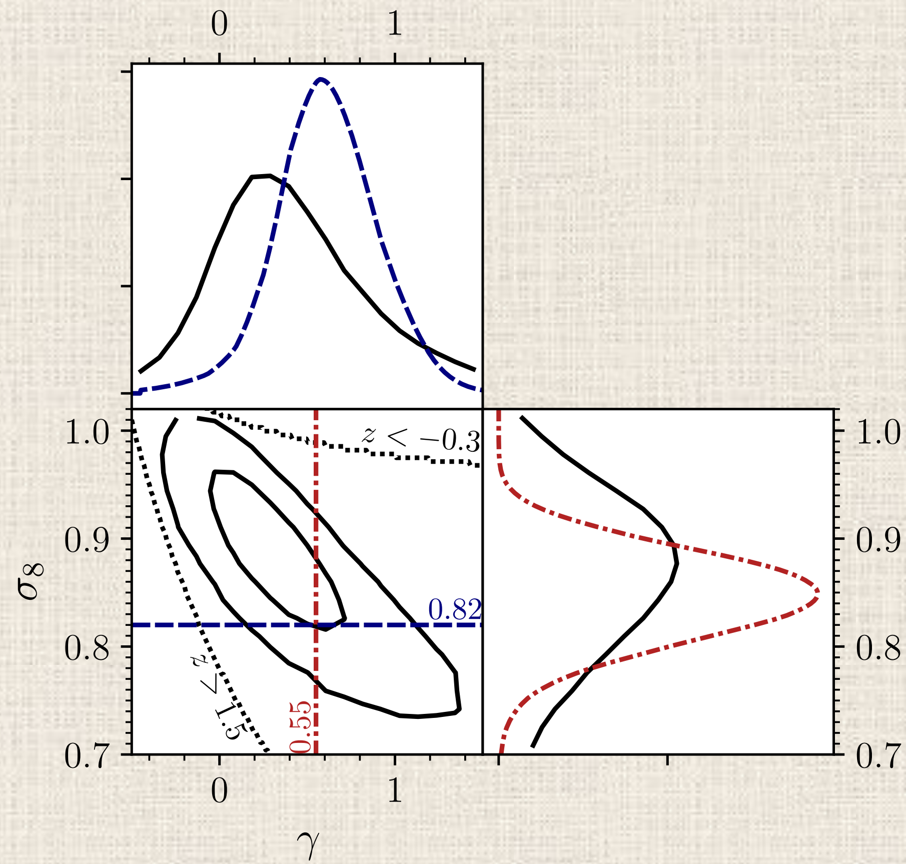
- Develop advanced 2pt clustering analysis of galaxy surveys:
  - advanced RSD modelling (Bianchi, LG)
  - forward-modelling cosmology, halo-galaxy match (Granett, Carbone, Tosone, LG)
  - Void cosmology (Carbone, Verza, Bonici,…)
  - Machine-Learning cosmological applications to data analysis and modelling (Tosone, Bonici, Cagliari, Granett, Bianchi, LG)
- Link to DM and particles:
  - n-body simulations in neutrino cosmologies (Carbone, Carella,…)
  - improved neutrino masses / "dark force" constraints from LSS (Archidiacono, Carbone, Castorina)

Connect real galaxies and DM halos using SHAM

Granett+ 2019, A&A 489, 653: simultaneous model fitting of simulation prediction to measured **2pt correlations between 0<z<1**



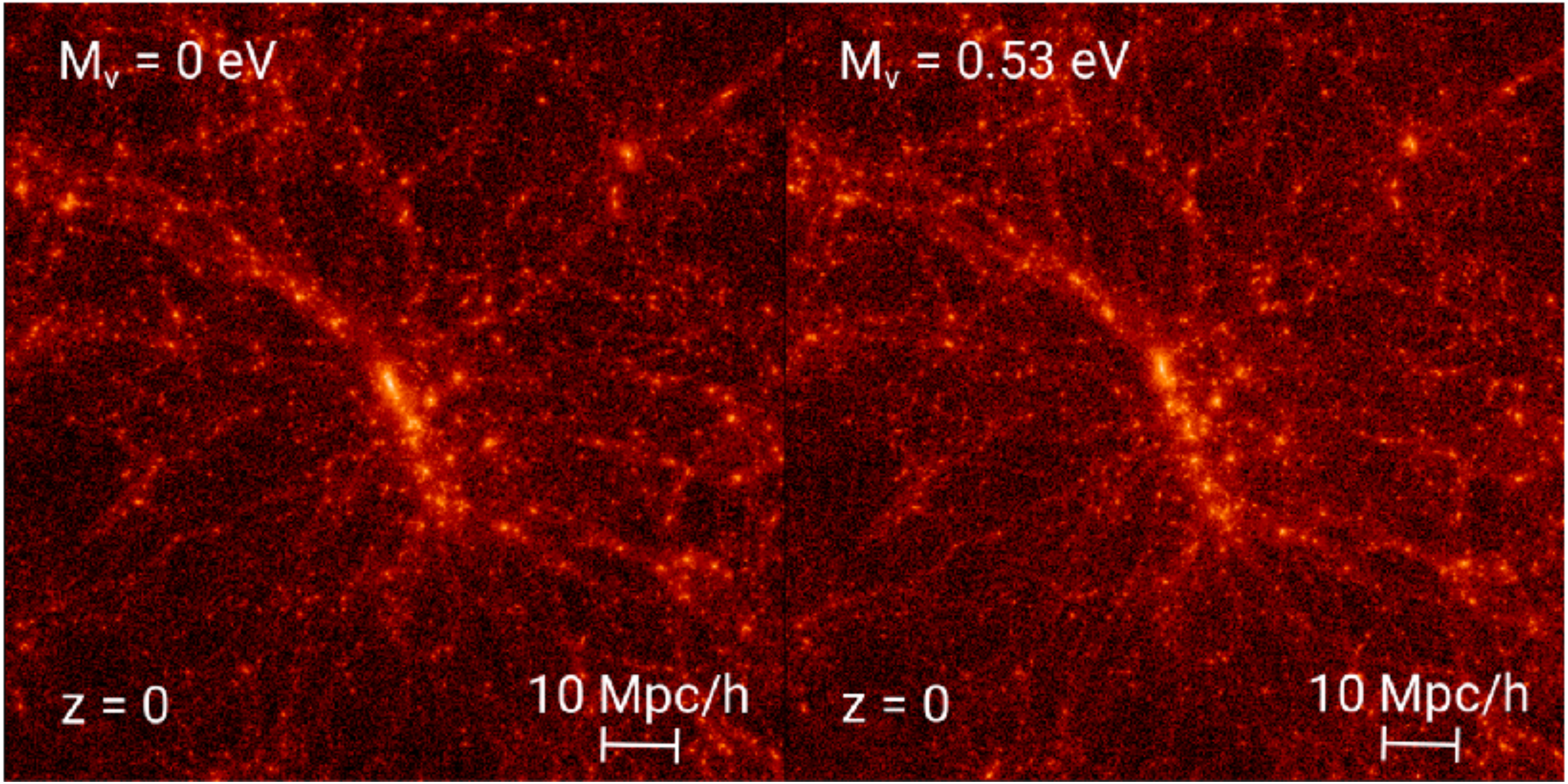
Joint analysis of SDSS and VIPERS



Predict amplitude and growth rate of clustering

Model the effect of massive neutrinos on large-scale structure

C. Carbone - DEMNUni N-body simulations in massive neutrinos and DE cosmologies: many applications and papers, including SHAM modelling vs survey data (Carella)

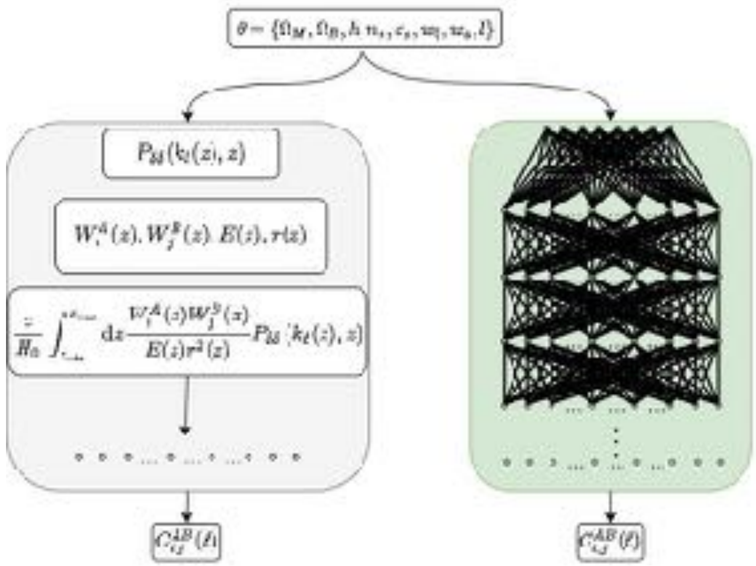


The new directions

I) Accelerate likelihood computation for summary statistics via ML emulation

(**Bonici**, Biggio, Carbone & Guzzo, 2024, MNRAS 531, 4203)

Replace expensive computations of model power spectra for the Likelihood with a trained neural network

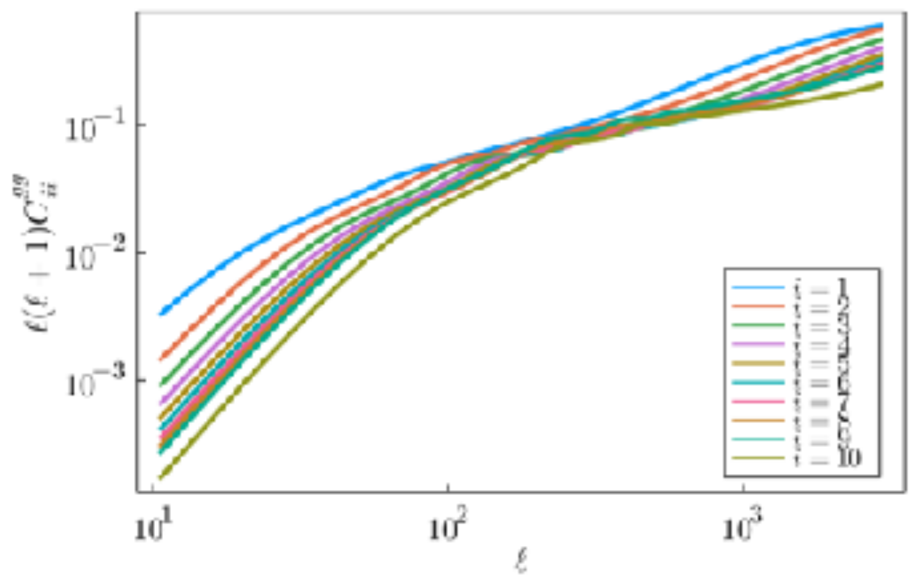


$$P(\theta \mid D) = \frac{\mathcal{L}(D|\theta)P(\theta)}{P(D)}$$

Our goal: **posterior** on data

Likelihood function, connecting **theory** and **data**

Prior on model parameters



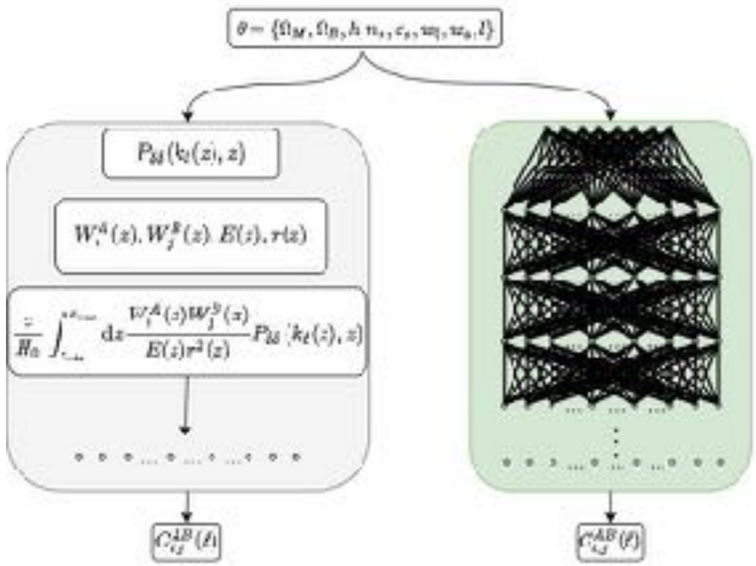
$$\mathcal{L}(\boldsymbol{D} \mid \boldsymbol{\theta}) \propto \exp\left(-\tfrac{1}{2}\left[(\boldsymbol{D} - \boldsymbol{C}(\boldsymbol{\theta}))^{\mathrm{T}} \boldsymbol{Cov}^{-1} (\boldsymbol{D} - \boldsymbol{C}(\boldsymbol{\theta}))\right]\right)$$

- **Speed-up ~10^3 x**
- **Accuracy <0.2% on all scales**

I) Accelerate likelihood computation for summary statistics via ML emulation

(**Bonici**, Biggio, Carbone & Guzzo, 2024, MNRAS 531, 4203)

Replace expensive computations of model power spectra for the Likelihood with a trained neural network

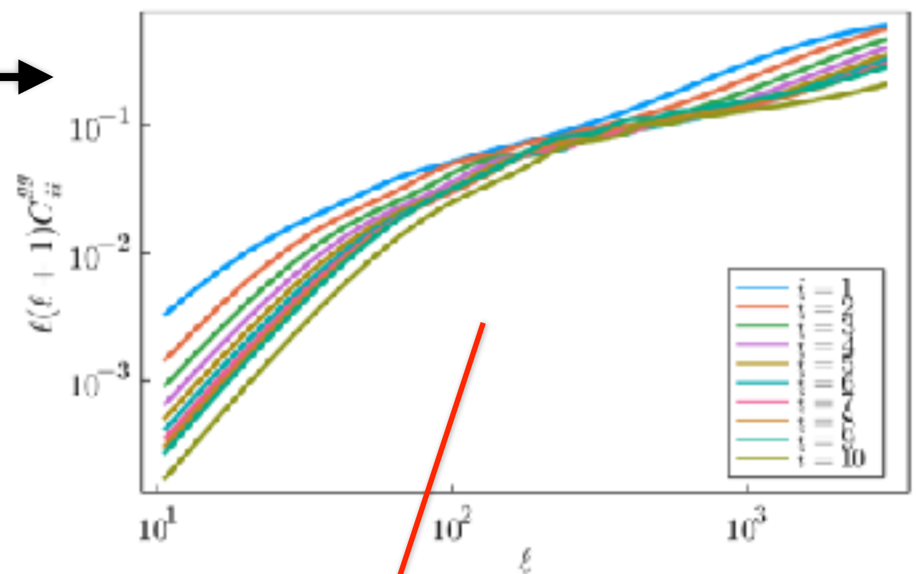


$$P(\theta \mid D) = \frac{\mathcal{L}(D|\theta)P(\theta)}{P(D)}$$

Our goal: **posterior** on data

Likelihood function, connecting **theory** and **data**

Prior on model parameters



$$\mathcal{L}(\boldsymbol{D} \mid \boldsymbol{\theta}) \propto \exp\left(-\tfrac{1}{2}\left[(\boldsymbol{D} - \boldsymbol{C}(\boldsymbol{\theta}))^{\mathrm{T}} \boldsymbol{Cov}^{-1} (\boldsymbol{D} - \boldsymbol{C}(\boldsymbol{\theta}))\right]\right)$$

- **Speed-up ~10^3 x**
- **Accuracy <0.2% on all scales**

I) Accelerate likelihood computation for summary statistics via ML emulation

(**Bonici**, Biggio, Carbone & Guzzo, 2024, MNRAS 531, 4203)

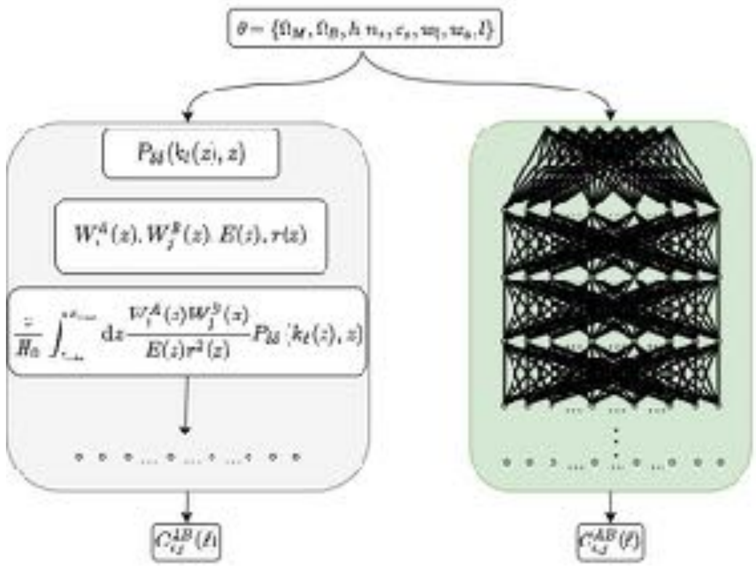
$$P(\theta \mid D) = \frac{\mathcal{L}(D|\theta)P(\theta)}{P(D)}$$

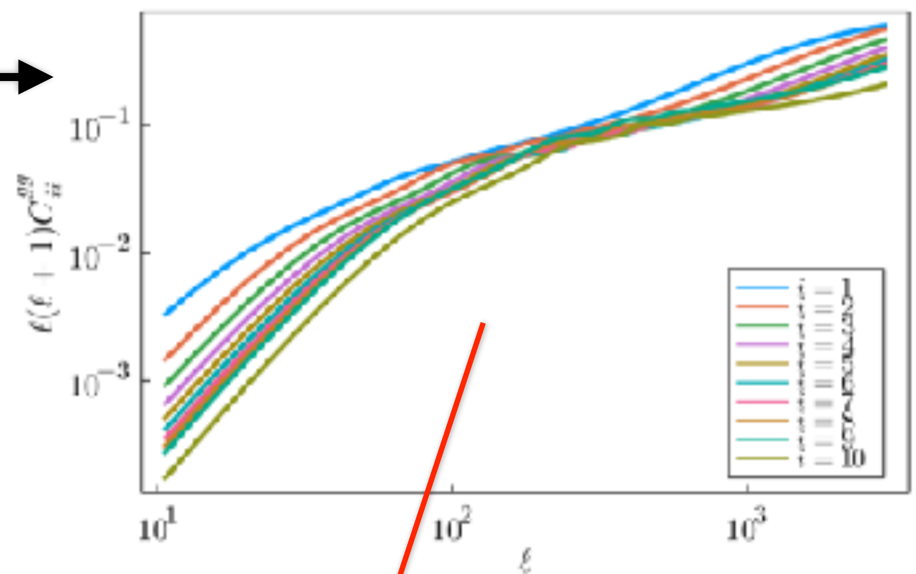
Our goal: **posterior** on data

Likelihood function, connecting **theory** and **data**

Prior on model parameters

Replace expensive computations of model power spectra for the Likelihood with a trained neural network





>>> Alternative route: ***symbolic regression***

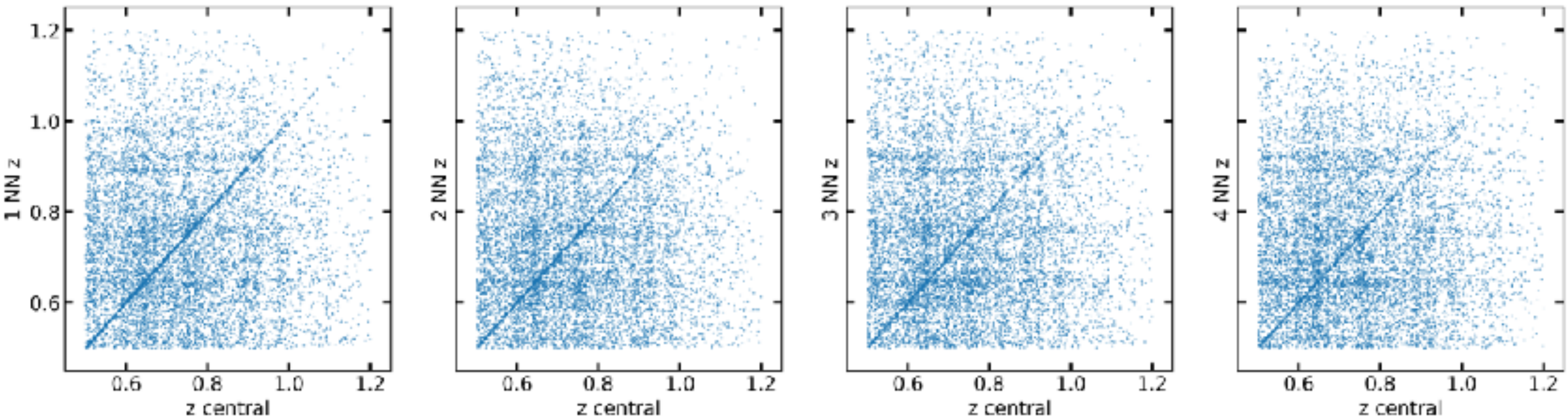
BSc theses of F. Farinelli and S. Carturan (2023, 2/ Bonici)

$$\mathcal{L}(\boldsymbol{D} \mid \boldsymbol{\theta}) \propto \exp\left(-\tfrac{1}{2}\left[(\boldsymbol{D} - \boldsymbol{C}(\boldsymbol{\theta}))^{\mathrm{T}} \boldsymbol{Cov}^{-1} (\boldsymbol{D} - \boldsymbol{C}(\boldsymbol{\theta}))\right]\right)$$

- **Speed-up ~10^3 x**
- **Accuracy <0.2% on all scales**

2) Improving photo-z accuracy by exploiting galaxy spatial correlations via Graph Neural Networks (GNN)

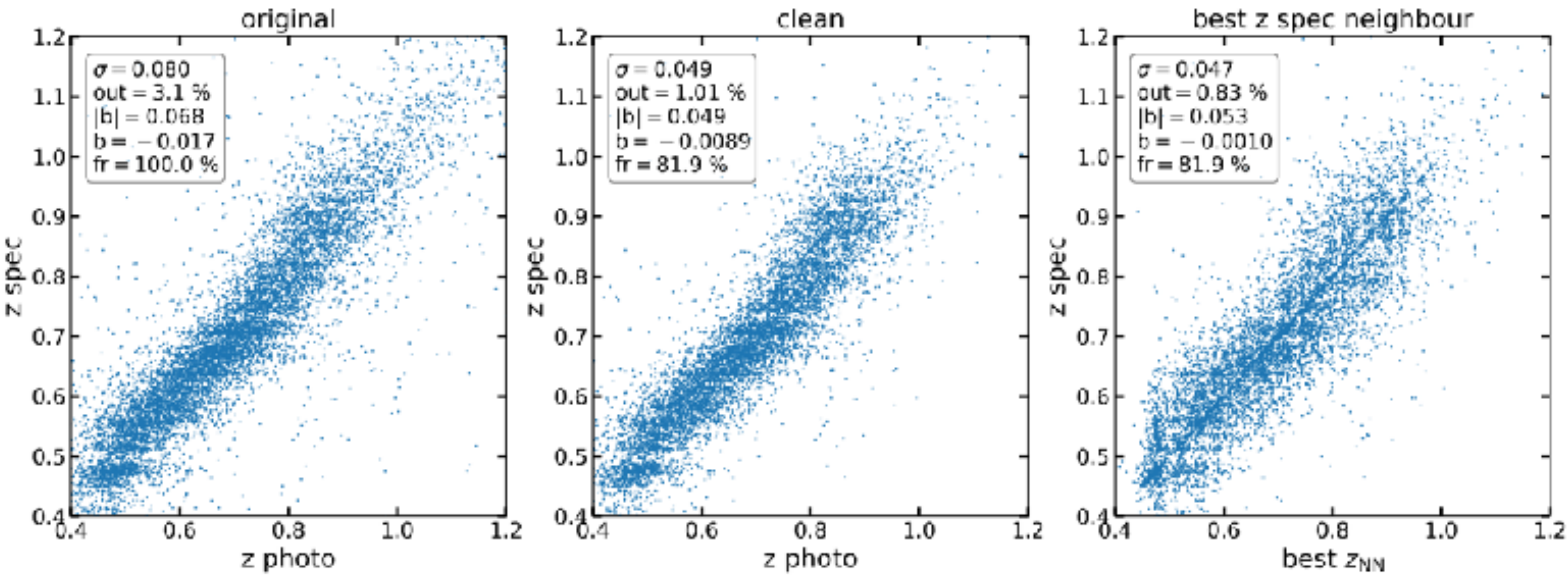
(**Tosone**, Cagliari, LG, + 2023)



- Angular pairs on the sky have a **higher-than-random probability to be also at similar redshift**: galaxies are correlated!
- **We can train a GNN to learn the link of colours and redshifts among neighbouring galaxies**

2) Improving photo-z accuracy by exploiting galaxy spatial correlations via Graph Neural Networks (GNN)

(**Tosone**, Cagliari, LG, + 2023)



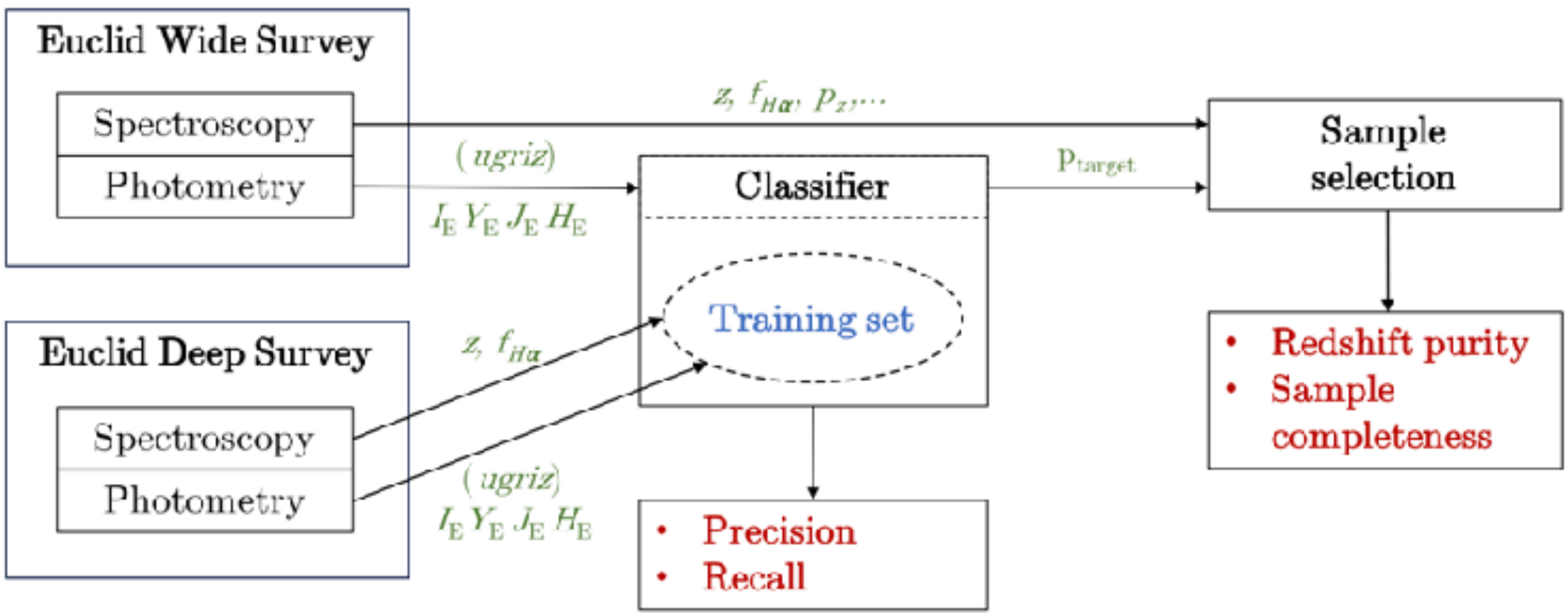
- Probability threshold for a galaxy to be classified as neighbour can be tuned to select (smaller and smaller) samples with increasingly precise redshifts
- **Redshift of a highly confident neighbour is typically more precise than original photo-z** (3d panel)

3) Improving purity & completeness of (Euclid) spectroscopic samples

Photometric selection for Euclid

Cagliari et al. (2024)

Due to **false detection** and **line mis-identification** the redshift purity of Euclid galaxy clustering sample is low.



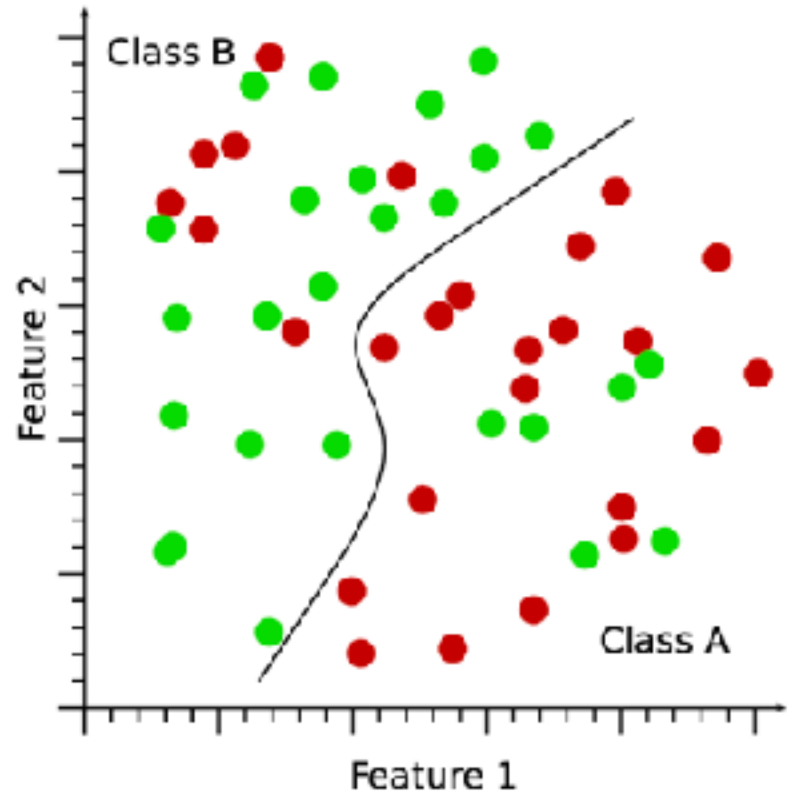
3) Improving purity & completeness of (Euclid) spectroscopic samples

Classifiers

- A classifier output is the probability of an object of being part of a class.
- Metrics of interest:

$$\text{precision} = \frac{TP}{TP + FP},$$
$$\text{recall} = \frac{TP}{TP + FN}.$$

- We tested six different ML-based classifiers.



3) Improving purity & completeness of Euclid spectroscopic samples

(**Cagliari**, Granett, LG, + 2024)

Mock data

$$\text{Targets:} \begin{cases} 0.9 < z < 1.8 \\ f_{\mathrm{H}\alpha}^{\mathrm{gal}} > 2 \times 10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}} \end{cases}$$

We test the algorithms with **EL-COSMOS** and **Flagship2** data.

Euclid

EL-COSMOS & Flagship

- $I_{\mathrm{E}} - Y_{\mathrm{E}}$,
- $Y_{\mathrm{E}} - J_{\mathrm{E}}$,
- $J_{\mathrm{E}} - H_{\mathrm{E}}$,
- H_{E}.

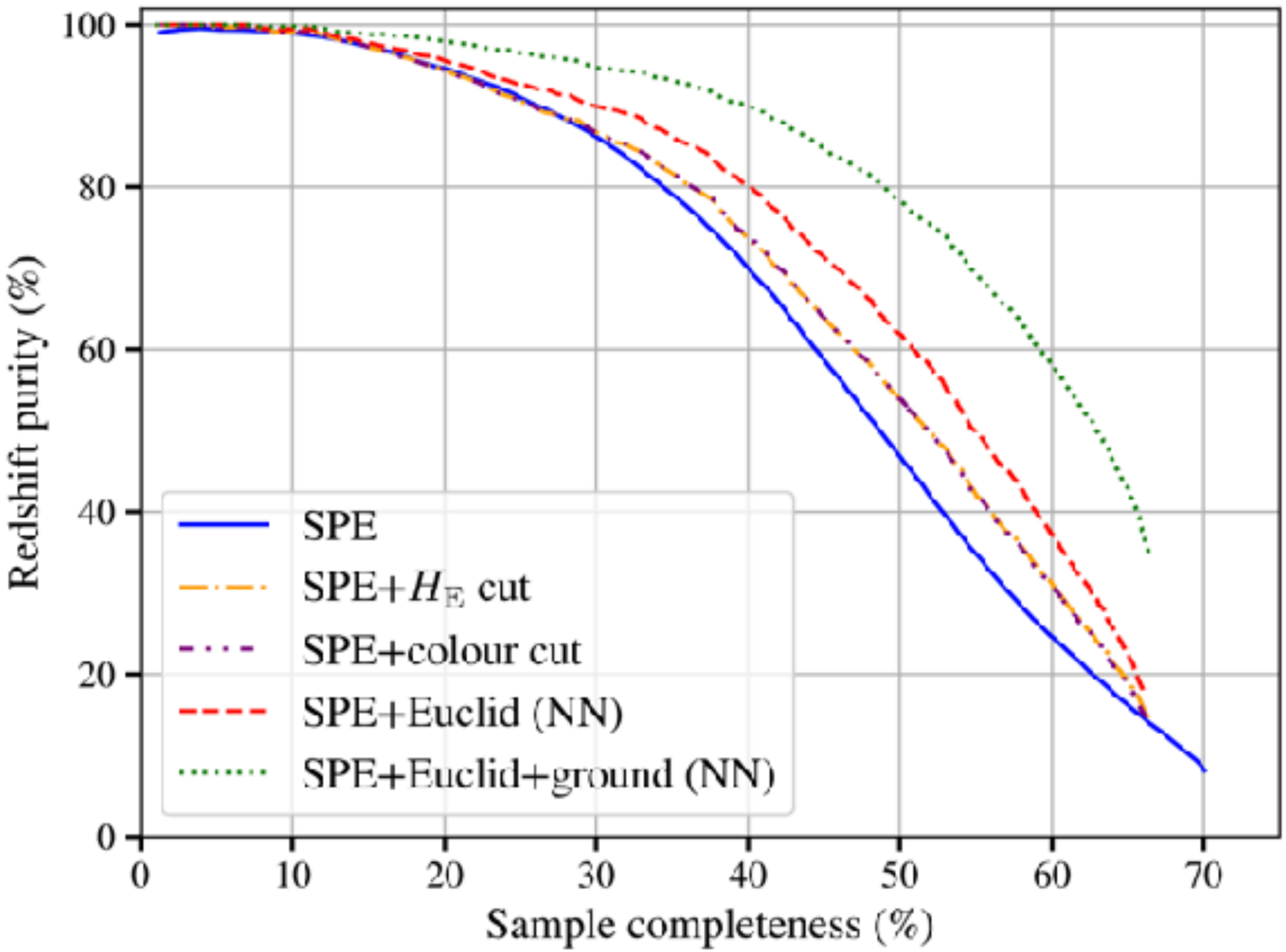
Euclid + ground

EL-COSMOS & Flagship

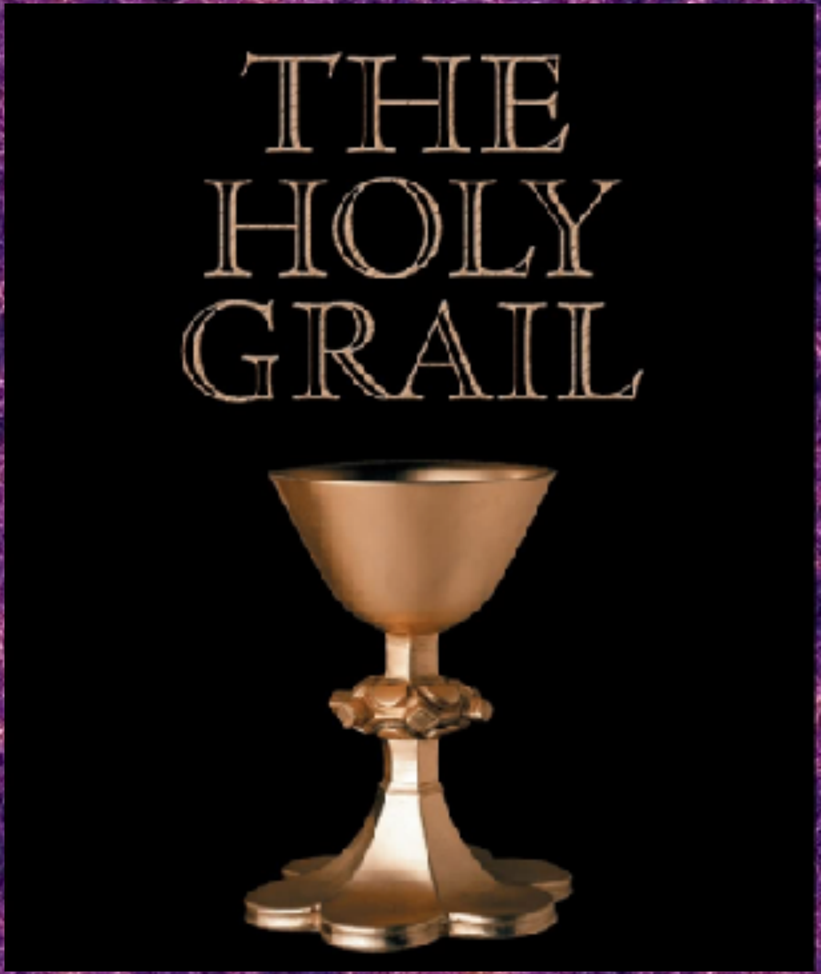
- $u - g$,
- $g - r$,
- $r - i$,
- $i - z$,
- $z - Y_{\mathrm{E}}$,
- $Y_{\mathrm{E}} - J_{\mathrm{E}}$,
- $J_{\mathrm{E}} - H_{\mathrm{E}}$,
- H_{E}.

Improving purity & completeness of Euclid spectroscopic samples

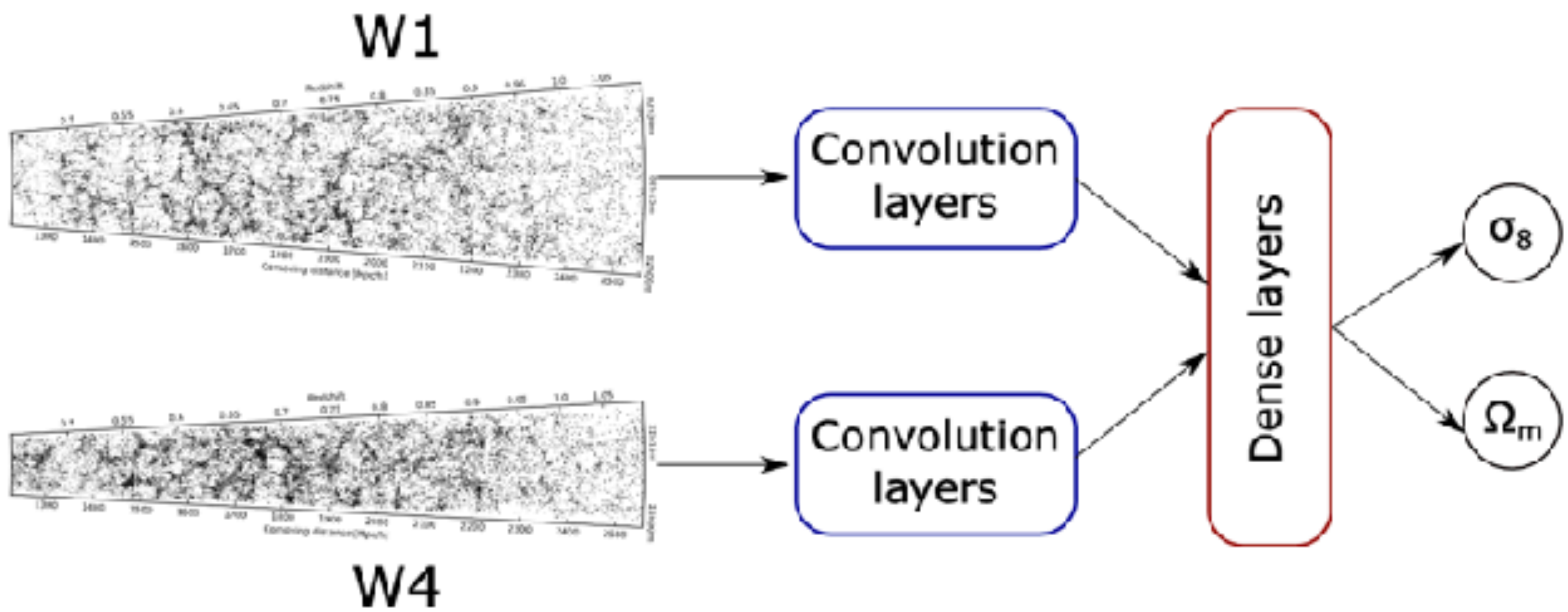
(**Cagliari**, Granett, LG, + 2024)



The future: bypass summary statistics and measure parameters directly from the galaxy field?



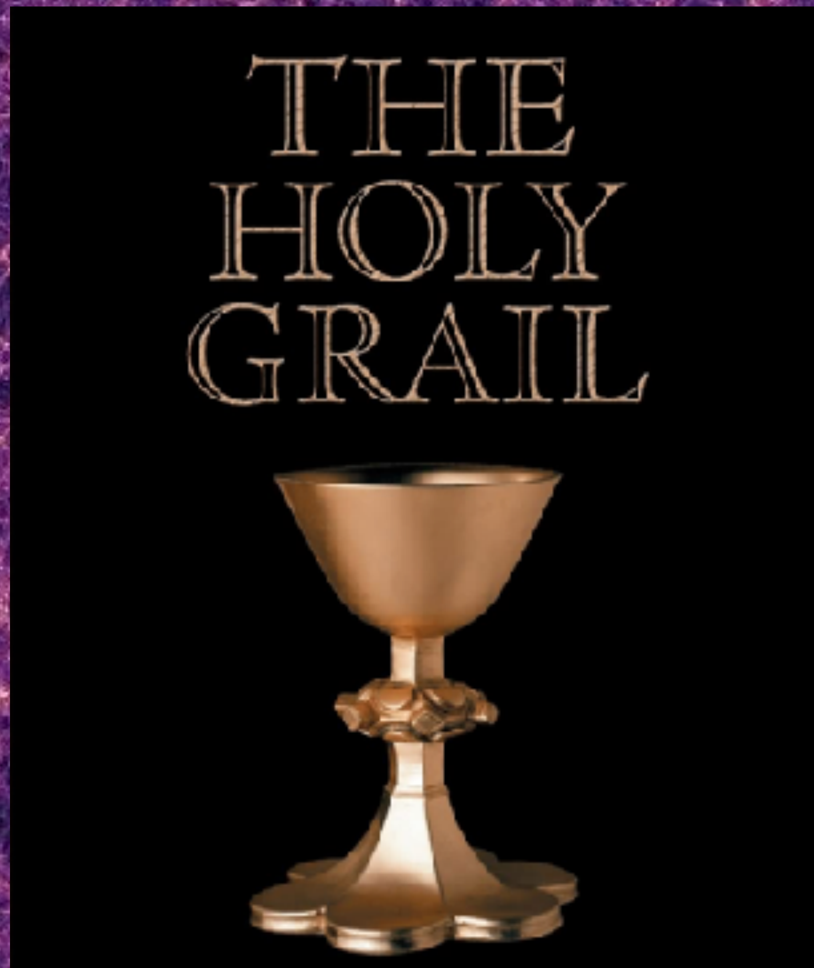
A field-level analysis with observational information



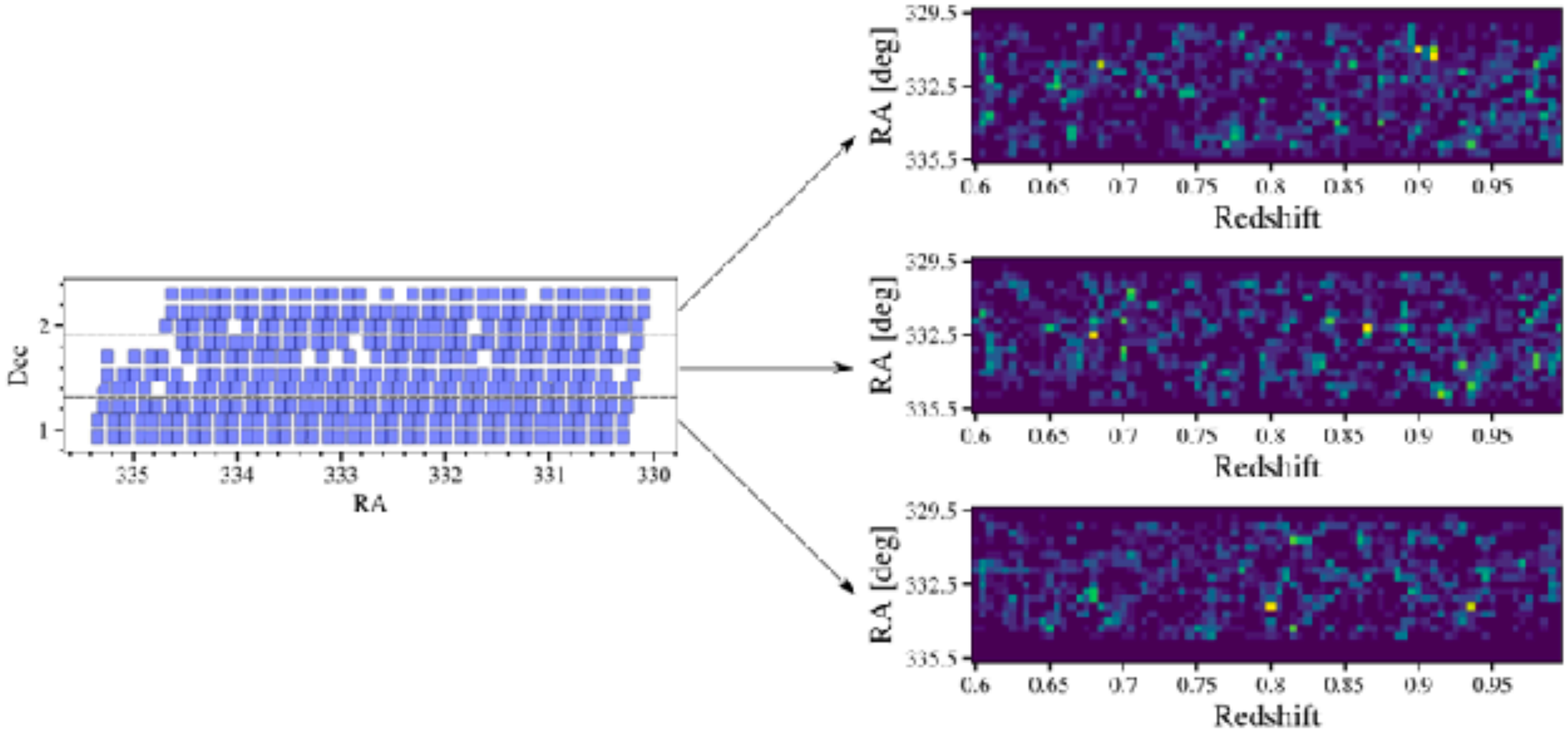
- The ambitious plan is a field-level analysis of VIPERS with a convolutional neural network.
- We use only observational information: the angular position and the redshift of the objects.

Cagliari, PhD thesis, 2024 - project idea by B. Granett)

Central problem: training samples.
We do not have “example Universes”



My current setup

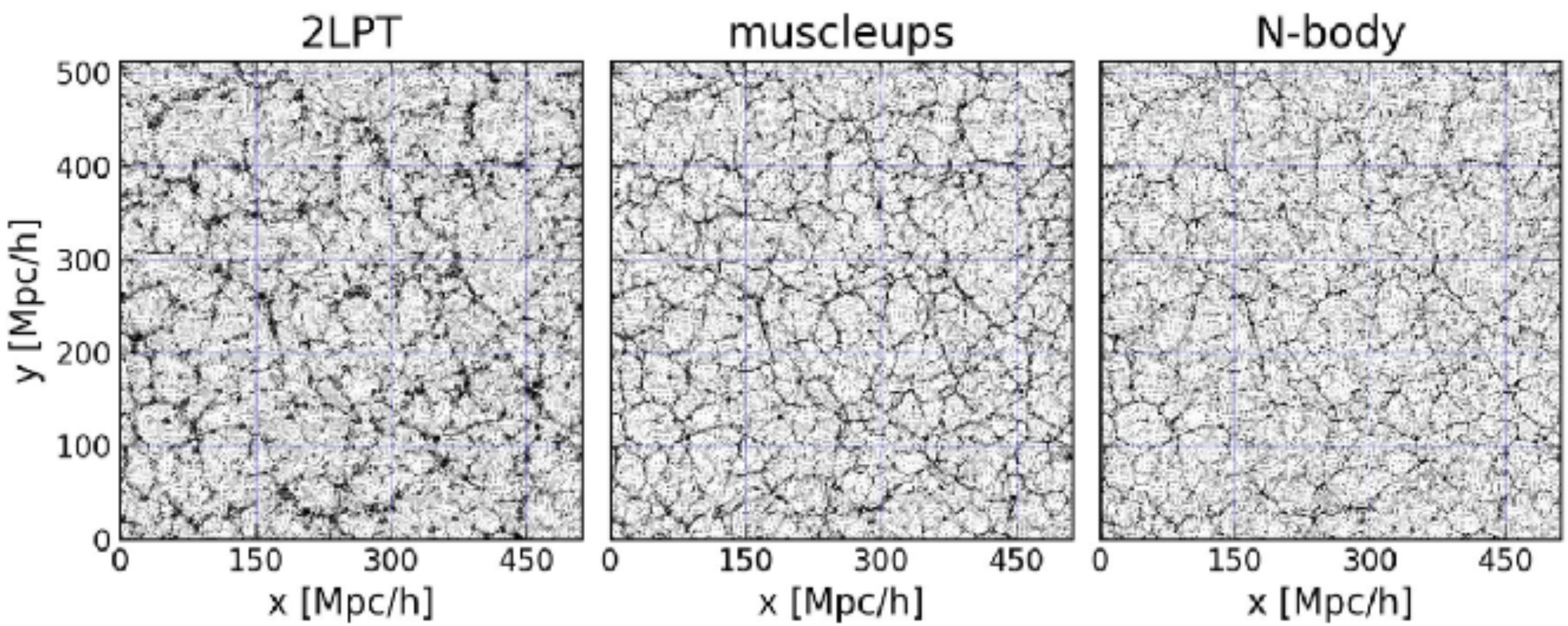


- I am working with DM halos in redshift space and real space.
- I am training the CNN with **3LPT** simulations and testing with an **N-body**.

Pinocchio (Monaco+ 2002)

Training sets from Lagrangian PT:

- State of the art: **Pinocchio** (Monaco+ 2002; Monaco 2016), extensively used in Euclid.
- **PRIN2017** contribution: **MUSCLE-UPS** (2021, w/ Neyrinck, Granett, LG - publicly available),
- Builds upon and improves prev. ALPT (Neyrinck 2013), PATCHY (Kitaura+ 2016) and MUSCLE (Neyrinck 2016)



Training sets from n-body ML Emulators? 2D tests

(**Sofia Chiarenza**, BSc 2021; **Marco Chiarenza**, BSc 2022 — supervisors **F. Tosone, LG**)

- **Emulate Lagrangian displacement field against N-body output**
- **Use CNN-UNET (Sofia C.) or Gen. Adversarial Networks (GAN, Marco C.)**

- q –> initial positions
- x –> final positions
- Ψ –> displacement

$$x = q + \Psi \tag{2}$$

- provide to the network $\nabla \cdot \Psi_{\text{approx}}$
- the Unet outputs $\nabla \cdot \Psi_{\text{unet}}$
- pixel-wise MSE loss

$$\mathcal{L} = \frac{1}{N} \sum_{\text{pixels}} (\nabla \cdot \Psi_{\text{N-body}} - \nabla \cdot \Psi_{\text{unet}})^2 \tag{3}$$

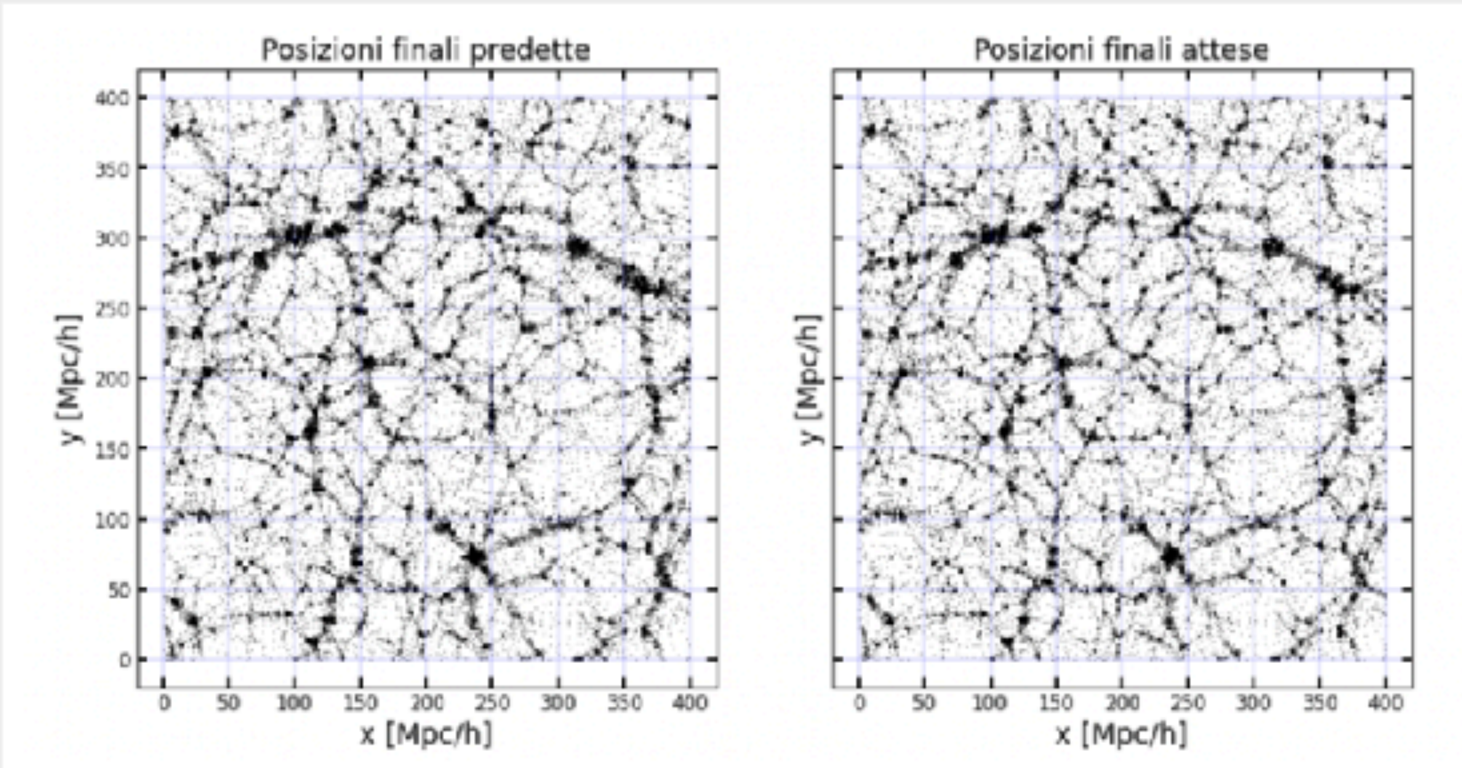
References: (He et al., 2019; Alves de Oliveira et al., 2020)

(courtesy F. Tosone)

Training sets from n-body ML Emulators? 2D tests

(**Sofia Chiarenza**, BSc 2021; **Marco Chiarenza**, BSc 2022 — supervisors **F. Tosone, LG**)

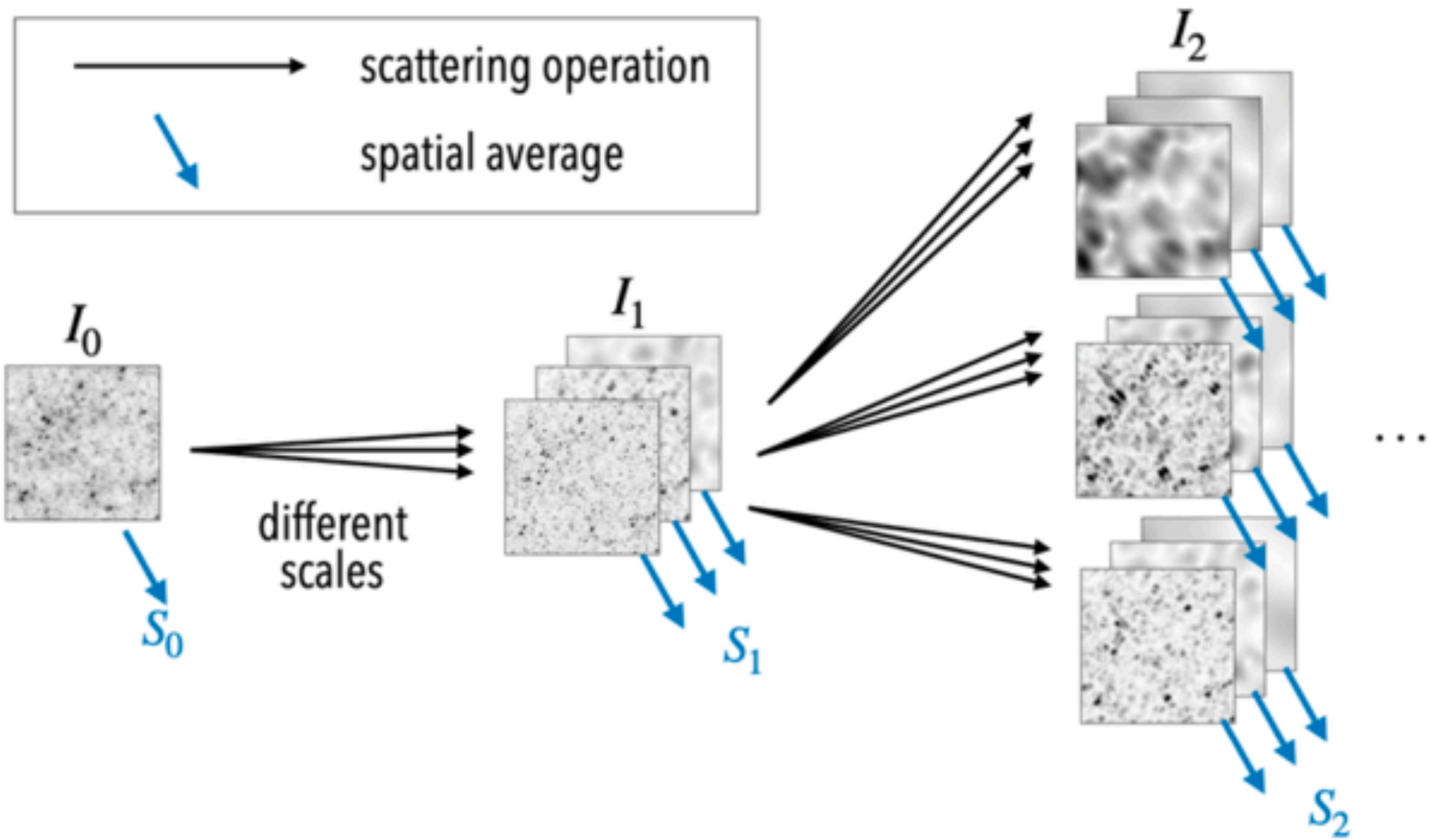
- **Emulate Lagrangian displacement field against N-body output**
- **Use CNN-UNET (Sofia C.) or Gen. Adversarial Networks (GAN)**



from Sofia Chiarenza bachelor thesis

Pseudo-CNN: the Wavelet Scattering Trasform

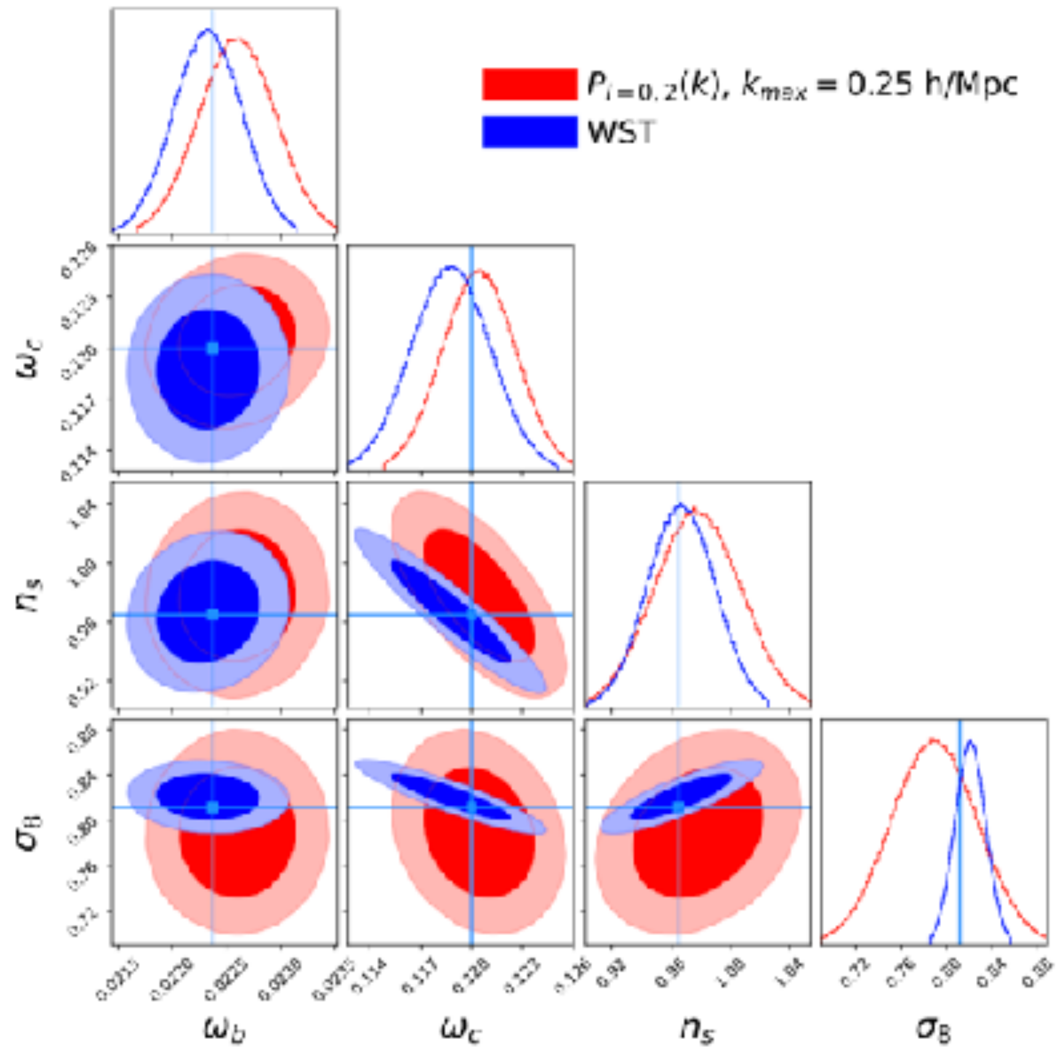
(**S. Chiarenza, MSc thesis**, /w D. Bianchi)



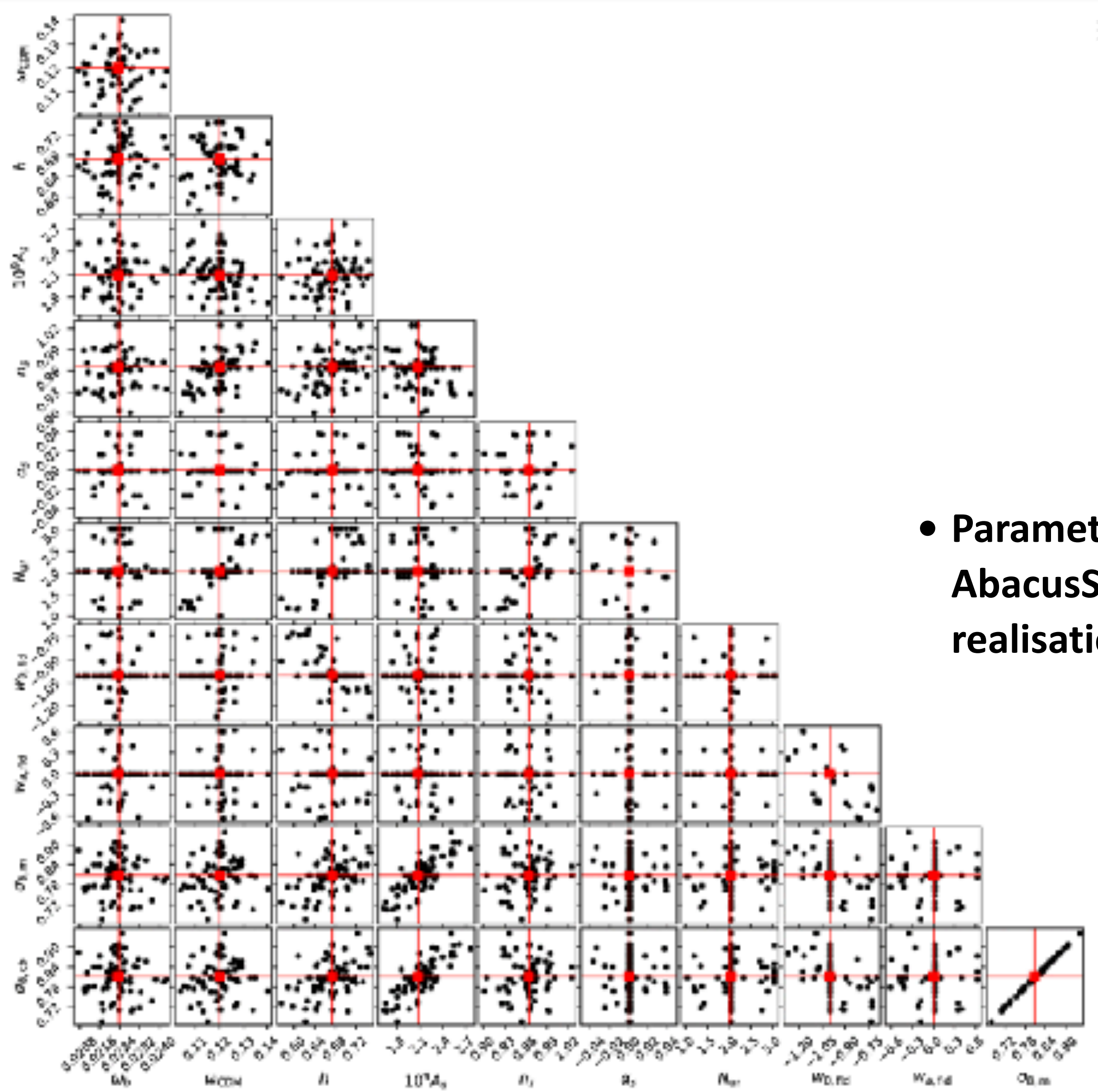
Pseudo-CNN: the Wavelet Scattering Trasform

(**S. Chiarenza, MSc thesis**, /w D. Bianchi)

- **Re-analysis of BOSS survey data: comparison of WST vs. P(k) monopole+quadrupole fit**



Pseudo-CNN: the Wavelet Scattering Trasform



arenza, **MSc thesis**, /w D. Bianchi)

- **Parameter space for the 96 AbacusSummit n-body realisations**

THE END