

Precision cosmology and the need of (many!) digital twins of the Universe

Pierluigi Monaco

(Trieste University, INAF, INFN, IFPU, ICSC)

Astrophysics of LSS

8 September 2025, IFPU, Trieste



Euclid (2023 - 2029+)

Satellite mission led by ESA,
Launched July 1, 2023



Wide survey:

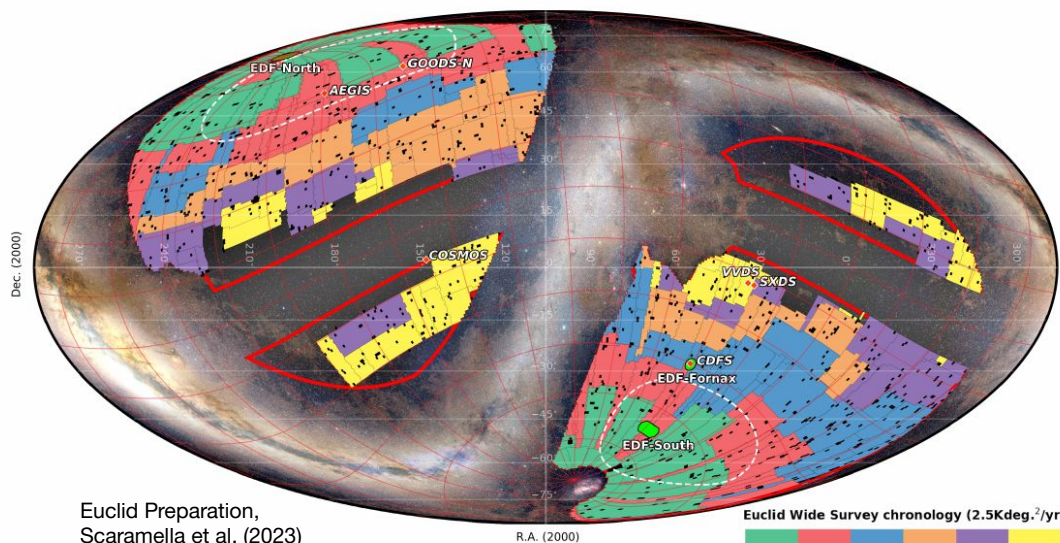
- 14,000 deg²
- VIS and NIR photometry
- slitless spectroscopy

Deep survey:

- 53 deg²
- 15 wide-like passes (red grism)
- 25 with blue grism

Releases:

- Q1 (no cosmology) in March 2025
- DR1 to the EC in March 2026
- DR1 to the world in October 2026

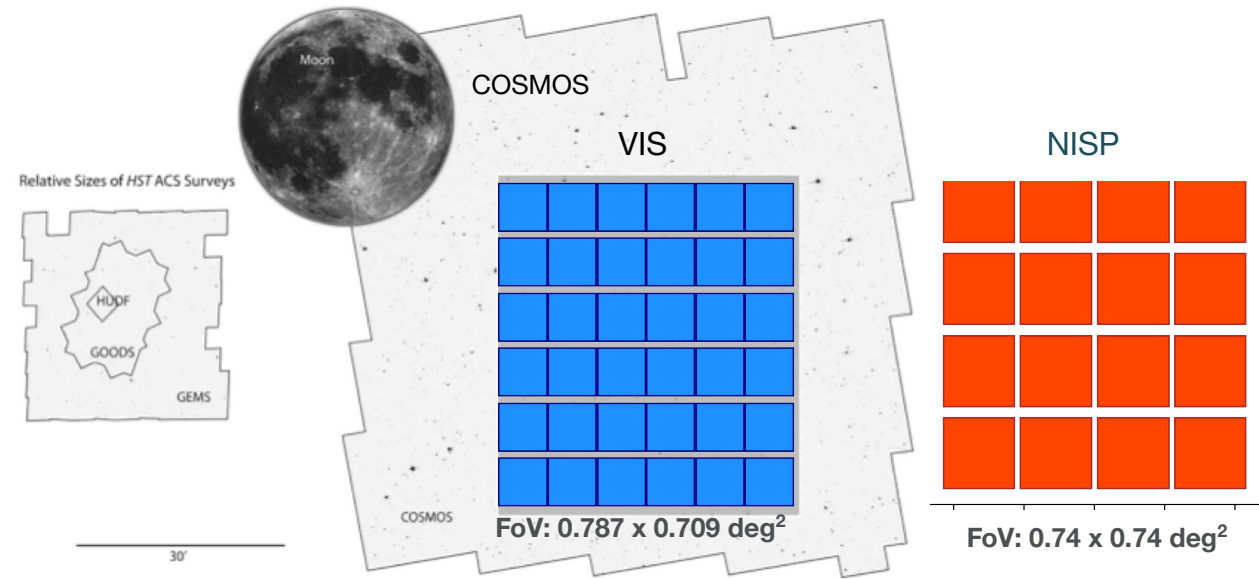
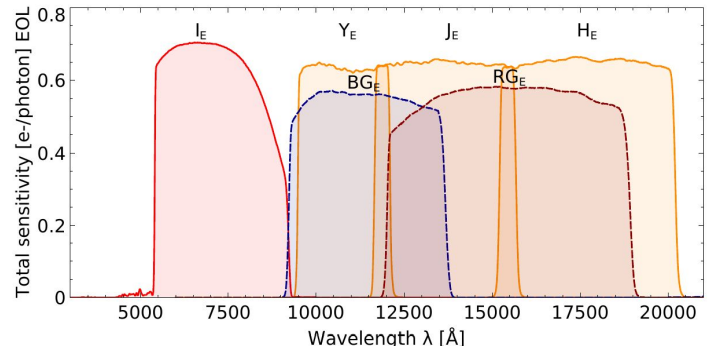
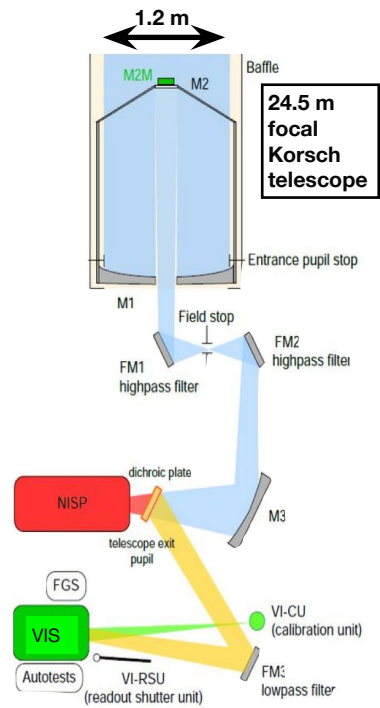


Euclid Preparation,
Scaramella et al. (2023)

Euclid Wide Survey chronology (2.5Kdeg.²/yr)

Year1	Year2	Year3	Year4	Year5	Year6
Green	Red	Blue	Orange	Purple	Yellow

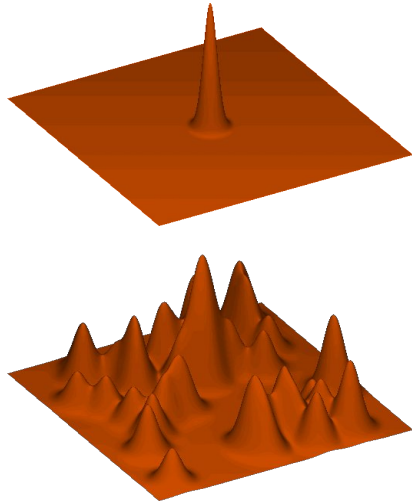
A dual wide-field imager



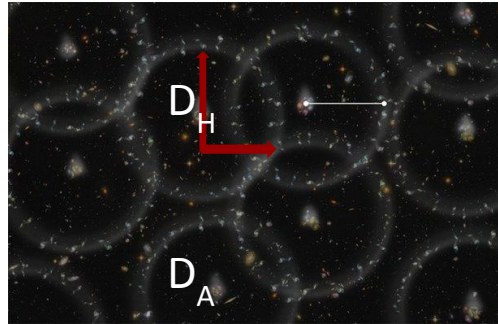
Overview of the PLM sub-systems – Courtesy Airbus Defence and Space.

Credit: Space Telescope Science Institute/Nick Scoville (Caltech)

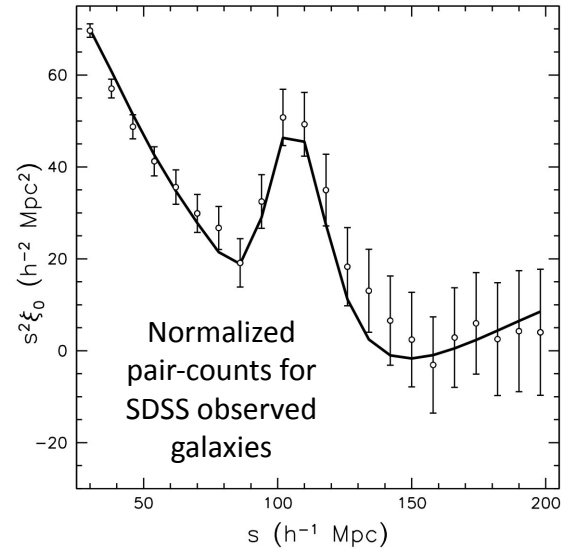
BAO as a standard ruler



Acoustic waves move material in the early (<300,000 years) universe



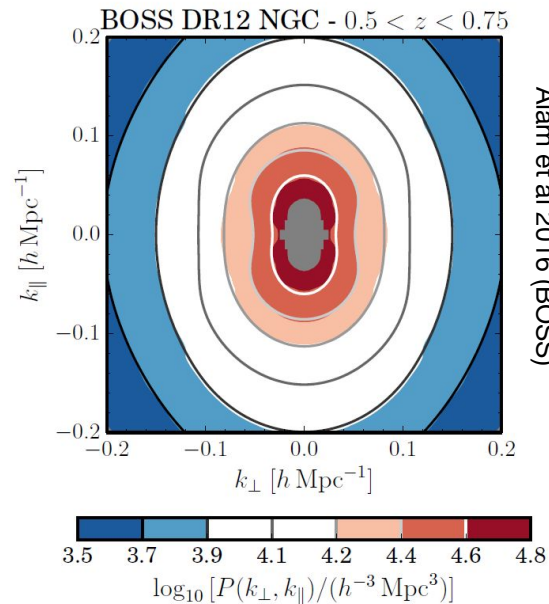
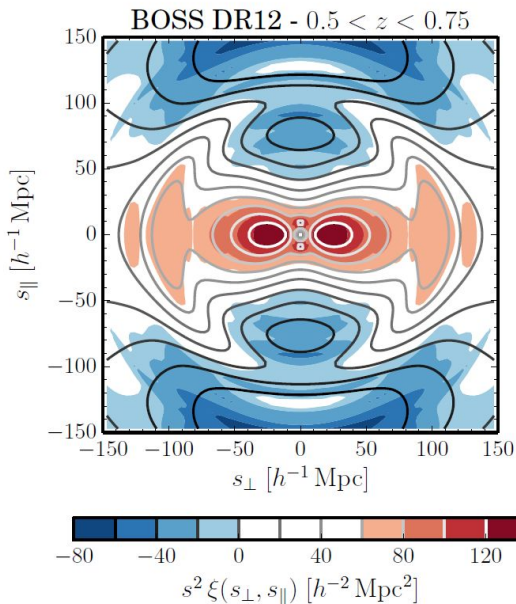
Galaxies form concentrations of mass; we see the imprint of the acoustic waves in the galaxy distribution



The projected scale of the acoustic waves gives a standard ruler with which to measure the universe

RSD in 2-point clustering

$\xi(\mathbf{r}_p, \pi)$

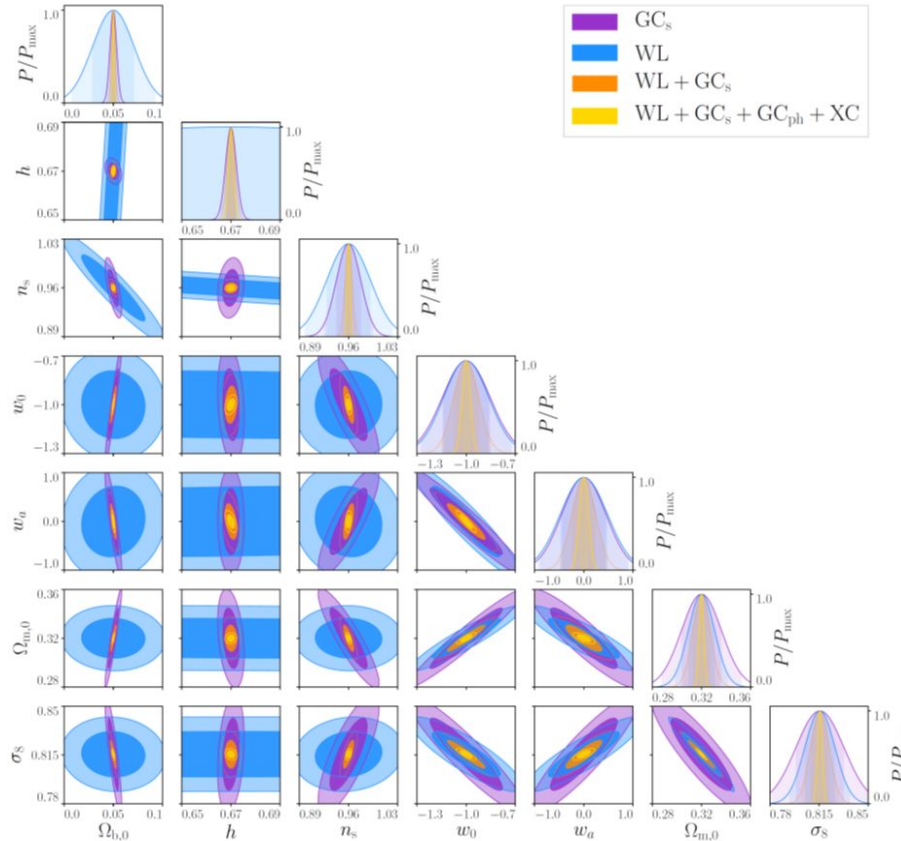


Alam et al 2016 (BOSS)

$P(\mathbf{k}, \mu)$

The measured pre-reconstruction correlation function (left) and power spectrum (middle) in the directions perpendicular and parallel to the line of sight

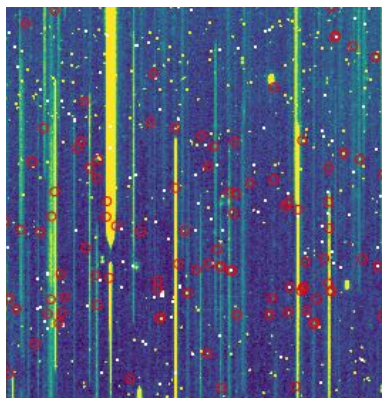
Precision cosmology from probe combination



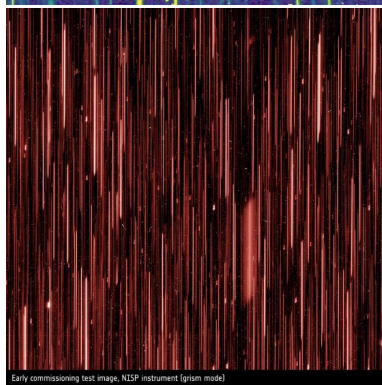
EC, Blanchard+ 2020

Slitless spectroscopy

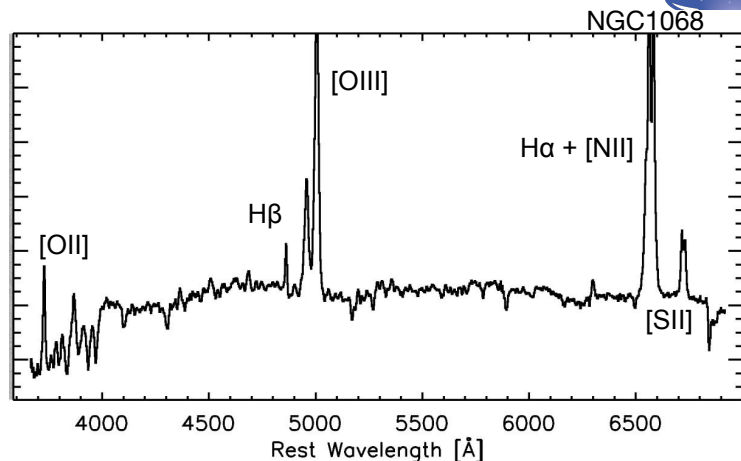
Credits: B. Granett & e2e group



simulated

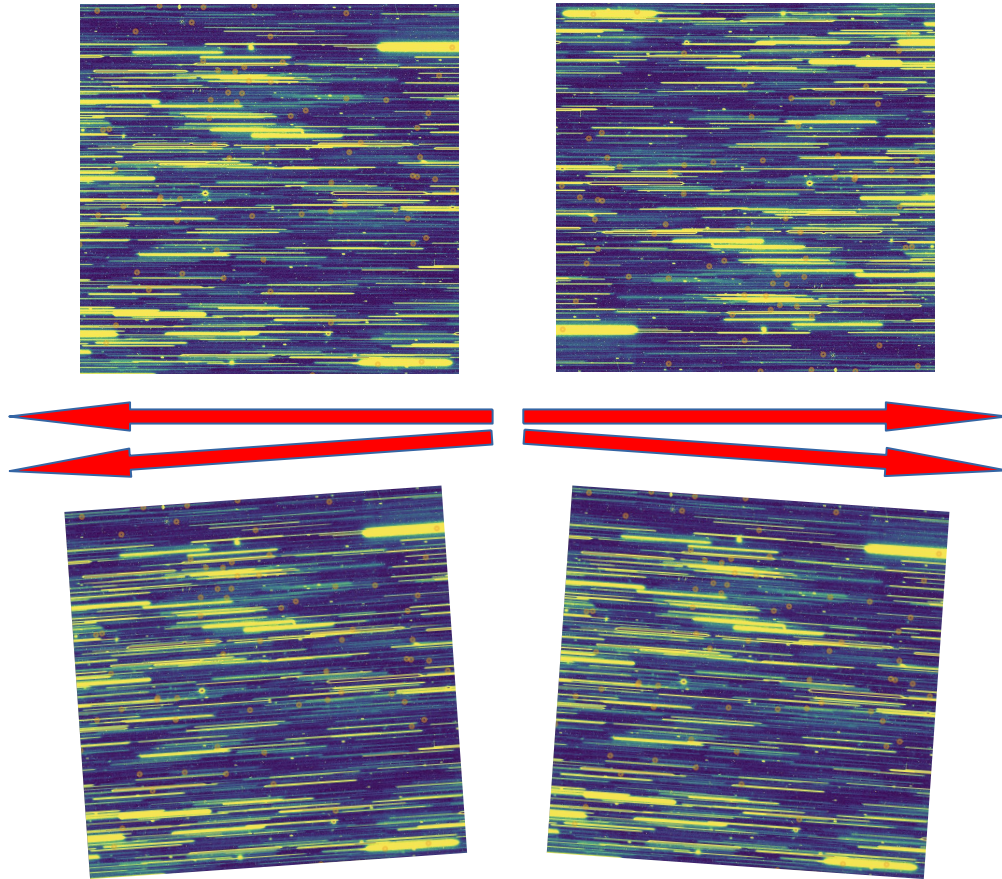


real



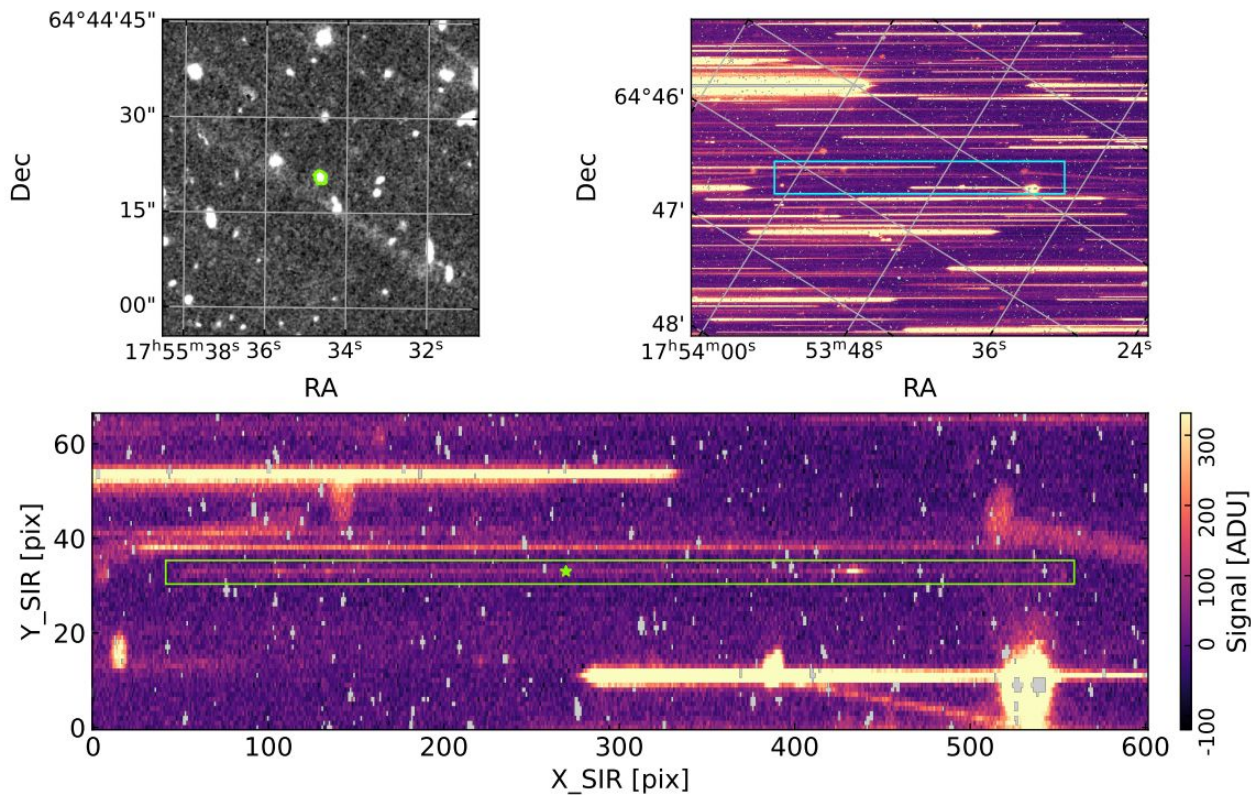
- All photons pass the grism (no slits or fibers)
 - Emission line galaxies are main targets
 - **No targeting required**
 - Efficiency loss due to higher background
- Euclid is the **first large-scale application** of this technique
- Slitless spectroscopy is technically simpler (no fibers), but the resulting selection function is complex
- Low signal-to-noise needed to have a dense sample
-> **high level of redshift interlopers**

Slitless spectroscopy



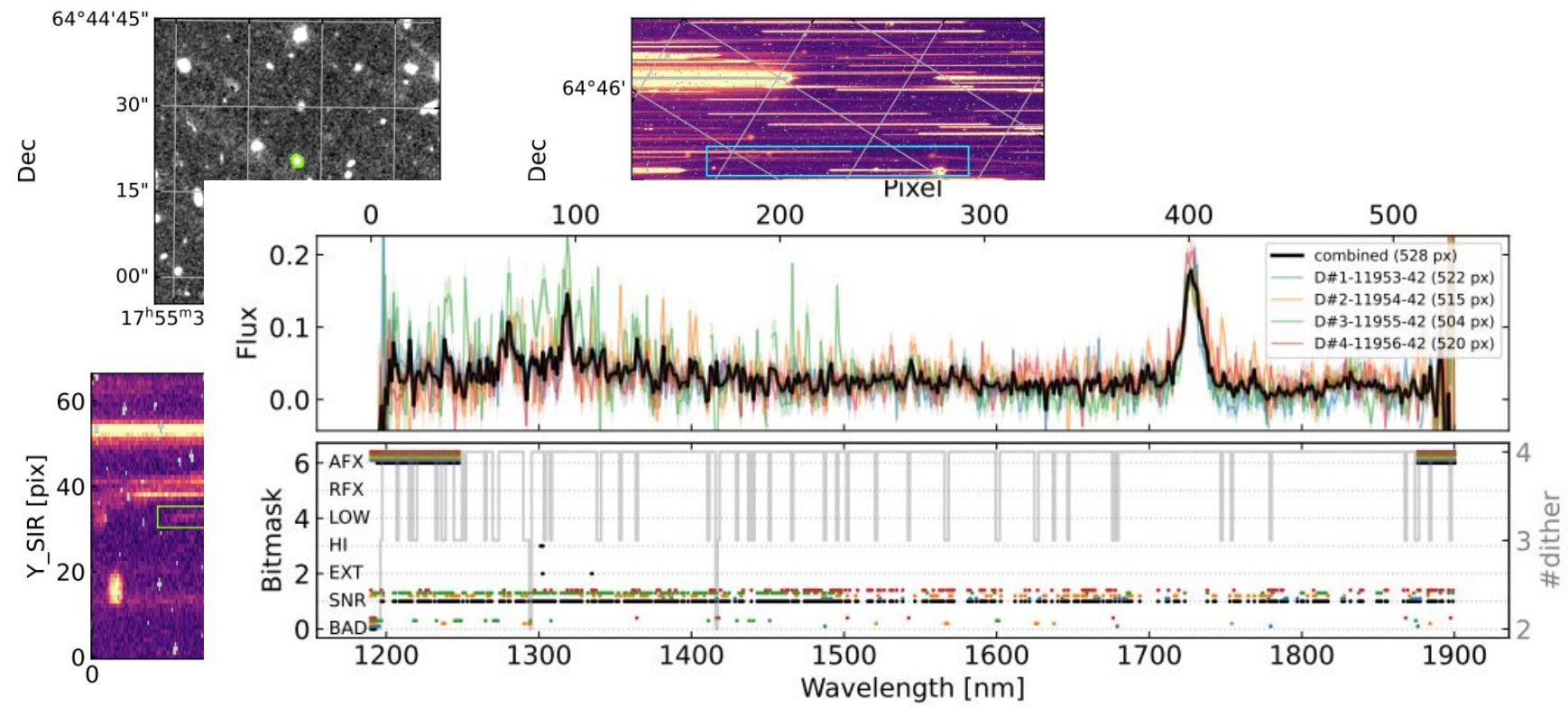
To limit confusion each field is observed with four dithers, with a **K-shaped pattern** in the orientation of the dispersion

Slitless spectroscopy



Slitless spectroscopy

EC-Q1, Copin et al., arXiv:2503.15307



Purity and completeness

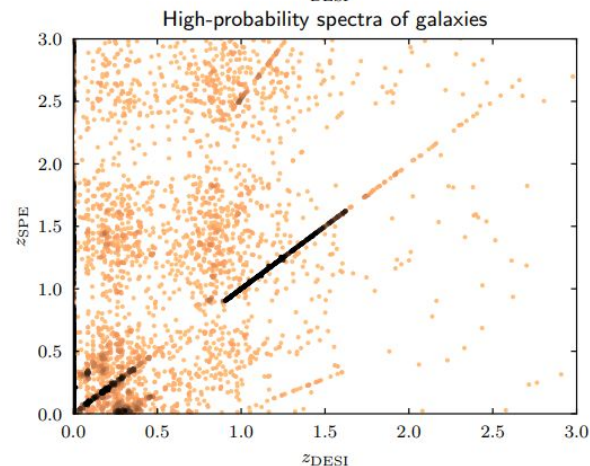
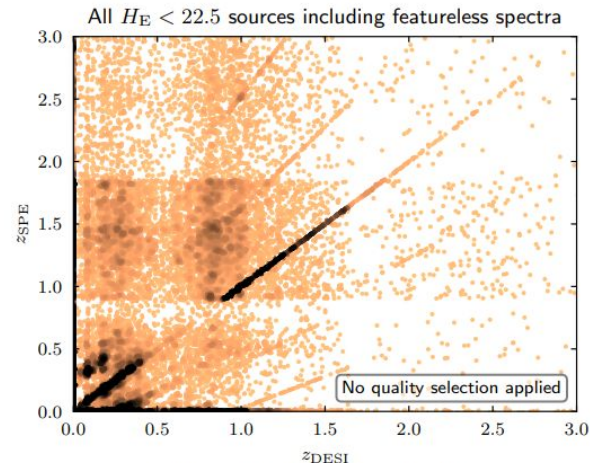
Due to the high background, the fraction of galaxies with a correct redshift is not expected to be very high.

In Q1 data the fraction of correct redshifts (purity) was still low, due to a large number of bad redshifts.

We expect:

- **line interlopers**: galaxies with a detected line that is misinterpreted as an H α line
- **noise interlopers**: galaxies without prominent lines, where a noise fluctuation or a detector feature not perfectly removed is interpreted as a line.

We aim at a purity of 80% for a sample with a galaxy density of at least 1700 deg $^{-2}$





A&A, 697, A5 (2025)
<https://doi.org/10.1051/0004-6361/202450853>
© The Authors 2025






























**Astronomy
&
Astrophysics**

Special issue

Euclid on Sky

Euclid

V. The Flagship galaxy mock catalogue: A comprehensive simulation for the *Euclid* mission

Euclid Collaboration: F. J. Castander^{1,2,*} , P. Fosalba^{1,2} , J. Stadel³ , D. Potter³ , J. Carretero^{4,5} , P. Tallada-Crespi^{4,5} , L. Pozzetti⁶ , M. Bolzonella⁶ , G. A. Mamon^{7,8} , L. Blot^{9,10} , K. Hoffmann¹ , M. Huertas-Company^{11,12,13,14} , **P. Monaco**^{15,16,17,18} , E. J. Gonzalez^{19,5,20} , G. De Lucia¹⁶ , C. Scarlata²¹ , M.-A. Breton^{1,22,10} , L. Linke²⁵ , C. Viglione^{2,1} , S.-S. Li²⁴ , Z. Zhai^{25,26,27,28} , Z. Baghkhani¹ , **K. Pardede**^{29,30,31,18} , C. Neissner^{19,5} , R. Teyssier³² , M. Crocce^{1,2} , I. Tutusaus³³ , L. Miller³⁴ , G. Congedo³⁵ , A. Biviano^{16,18} , M. Hirschmann^{36,16} , A. Pezzotta³⁷ , H. Aussel³⁸ , H. Hoekstra²⁴ , T. Kitching³⁹ , W. J. Percival^{27,28,40} , L. Guzzo^{41,42} , Y. Mellier^{8,7} , P. A. Oesch^{43,44,45} , R. A. A. Bowler⁴⁶ , S. Bruton²¹ , V. Allevalo⁴⁷ , V. Gonzalez-Perez⁴⁸ , M. Manera^{49,19} , S. Avila¹⁹ , A. Kovács^{50,51}

Flagship mock galaxy catalog

Simulation box of **3600 Mpc/h**

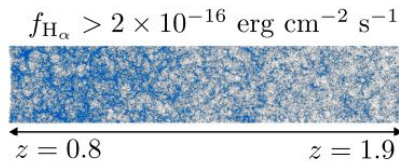
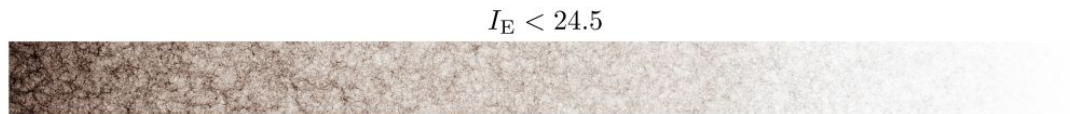
16000³ particles

particle mass of 10^9 Msun/h

output on the past light cone to $z=3$

~1 000 000 node hours on 4000 node of Piz Daint, ~68 000 000 core hours

Euclid Collaboration: Castander, F. J., et al.: A&A, 697, A5 (2025)



Flagship mock galaxy catalog

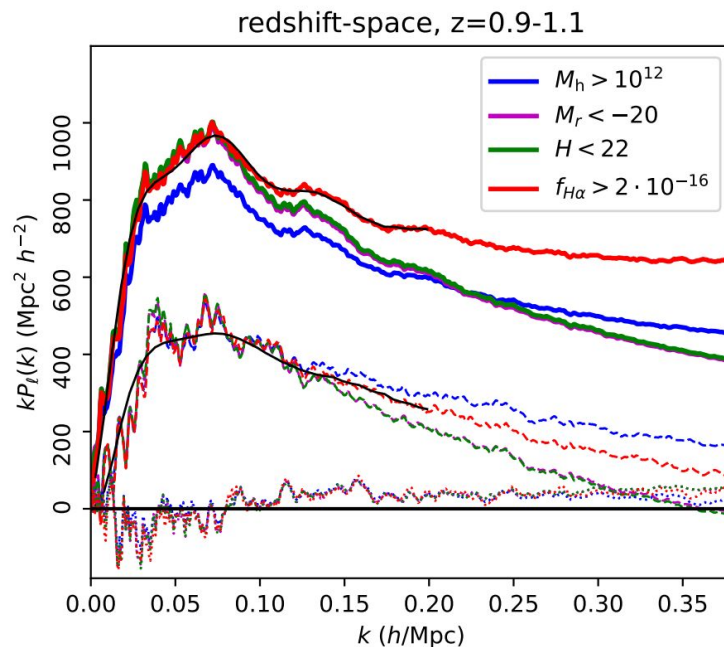
Halos populated with galaxies using an **HOD**

The HOD was **calibrated** to reproduce:

- galaxy magnitudes in all Euclid bands, including external data
- sizes
- colors
- main emission lines, including H α

The mock was **validated** checking

- weak lensing
- galaxy clustering
- properties of galaxies in clusters
- all galaxy properties used by legacy science, including
 - metallicity
 - morphology



Clustering analysis from K. Pardede

Flagship mock galaxy catalog

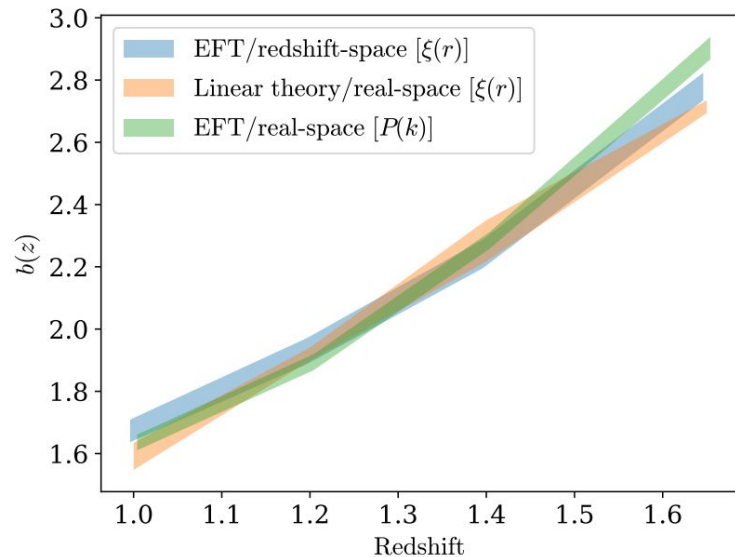
Halos populated with galaxies using an HOD

The HOD was calibrated to reproduce:

- galaxy magnitudes in all Euclid bands, including external data
- sizes
- colors
- main emission lines, including H α

The mock was validated checking

- weak lensing
- galaxy clustering
- properties of galaxies in clusters
- all galaxy properties used by legacy science, including
 - metallicity
 - morphology



Clustering analysis from K. Pardede

How many mocks do we need?

Noise in the **covariance matrix** due to sampling with a finite number of mocks propagates to the parameter covariance (errorbars) as a function of:

$$\langle \Delta\theta_i \Delta\theta_j \rangle = f \langle \Delta\theta_i \Delta\theta_j \rangle_{\text{ideal}}$$

$$f = 1 + B(N_b - N_p)$$

- N_b , length of the data vector (number of bins),
- N_p , number of parameters
- N_s , number of simulations (mocks)

$$B = \frac{(N_s - N_b - 2)}{(N_s - N_b - 1)(N_s - N_b - 4)}$$

N_b = Number of bins in **D**

N_p = number of parameters

(credits: A. Sanchez, L. Blot)

In Euclid we first aimed at **3500 simulations**, good for two cases:

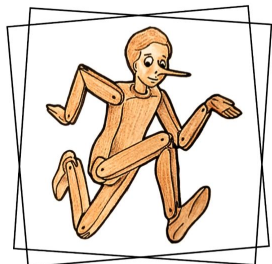
- $N_b = 600$ with required 10% precision in the errorbar,
- $N_b = 300$ with required 5% precision in the errorbar.

Advanced techniques (tapering, eigenvalue decomposition, shrinkage, fitting a model for the CM, denoising with ML...) weaken this constraint, so we aim at **a minimum of 1000 simulations**.

Euclid preparation

Simulating thousands of *Euclid* spectroscopic skies

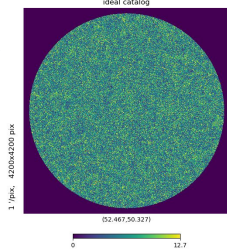
Euclid Collaboration: P. Monaco^{★1,2,3,4}, G. Parimbelli^{5,6,7}, M. Y. Elkhatab^{2,3,1,8}, J. Salvalaggio^{2,8,1,3},
T. Castro^{2,3,8,4}, M. D. Lepinza^{1,2}, E. Sarpa^{7,4,3}, E. Sefusatti^{2,8,3}, L. Stanco⁹, L. Tornatore², G. E. Addison¹⁰,
S. Bruton¹¹, C. Carbone¹², F. J. Castander^{5,13}, J. Carretero^{14,15}, S. de la Torre¹⁶, P. Fosalba^{13,5}, G. Lavaux¹⁷, S. Lee¹⁸,
K. Markovic¹⁸, K. S. McCarthy^{18,19}, F. Passalacqua^{20,9}, W. J. Percival^{21,22,23}, I. Risso^{24,25}, C. Scarlata²⁶,
P. Tallada-Crespí^{14,15}, M. Viel^{8,2,7,3,4}, Y. Wang²⁷, B. Altieri²⁸, S. Andreon²⁴, N. Auricchio²⁹, C. Baccigalupi^{8,2,3,7},
M. Baldi^{30,29,31}, S. D'Adda²⁹, D. DeGale²⁹, F. DeMatteis^{32,17}, A. DiMichele^{2,8}, F. Donati^{33,25,24}, M. Duvvuri^{34,35}



Pinocchio

PINOCCHIO github.com/pigimonaco/Pinocchio

halo catalogs
on the past light cone

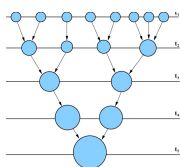


v5.1 just released (ICSC/Spoke3)

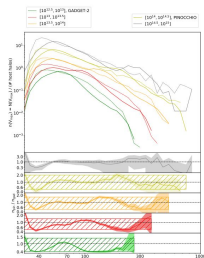
- Scaling up to 6144^3 particles using ~ 120 nodes
- + porting on GPUs (Lepinzan+ in prep)
- + displacing with Particle-Mesh...
- + Toward ML...



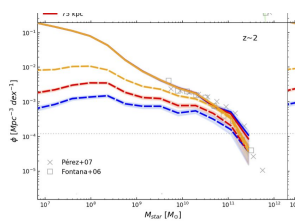
merger trees



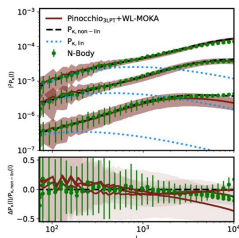
substructure
Berner+ 22



SAM (GAEA) Cammelli+ 24'



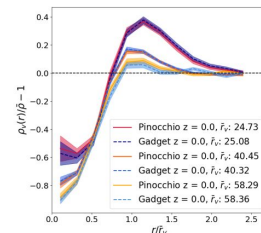
Lensing



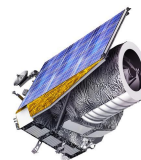
Giocoli+ 20
Castro, Lepinzan,
Elkhashab in prep

Voids

Lepinzan+
in prep

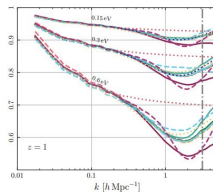


Euclid

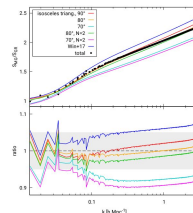


Beyond LCDM

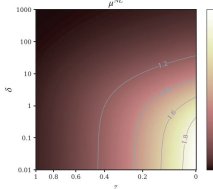
massive neutrinos
Rizzo+ 17
Adamek+ 23



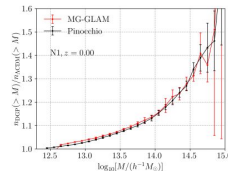
f(R) Moretti+ 20



cubic galileon
Song+ 21



nDGP Song+ 24



Mocks for numerical covariance

https://adlibitum.oats.inaf.it/pierluigi.monaco/euclid_mocks.html

Pinocchio lightcones

Simulations performed with PINOCCHIO v5.0 (and v4.1.3):

- Λ CDM cosmology similar to Flagship 1
- $M_p = 1.5 \cdot 10^{10} M_\odot/h$, smallest halo has 10 particles
- outputs at $z=1, 0$ + lightcone + histories
- periodic boxes available on request

CREDITS:

- computing time provided by **INFN**, CINECA (ISCRA-B), INAF (Pleiadi)
- post-processing time provided by SGS
- storage provided by SGS and INAF IA2 archives

GeppettoFC

- 3500 realizations
- 1200 Mpc/h box, 2160^3 particles
- lightcone ($z < 2$): circle of radius 30 deg
- 200 core-h per run, 700,000 core-h
- 23 GB each (8 GB for the lightcone), total ~80 TB

EuclidLargeBox

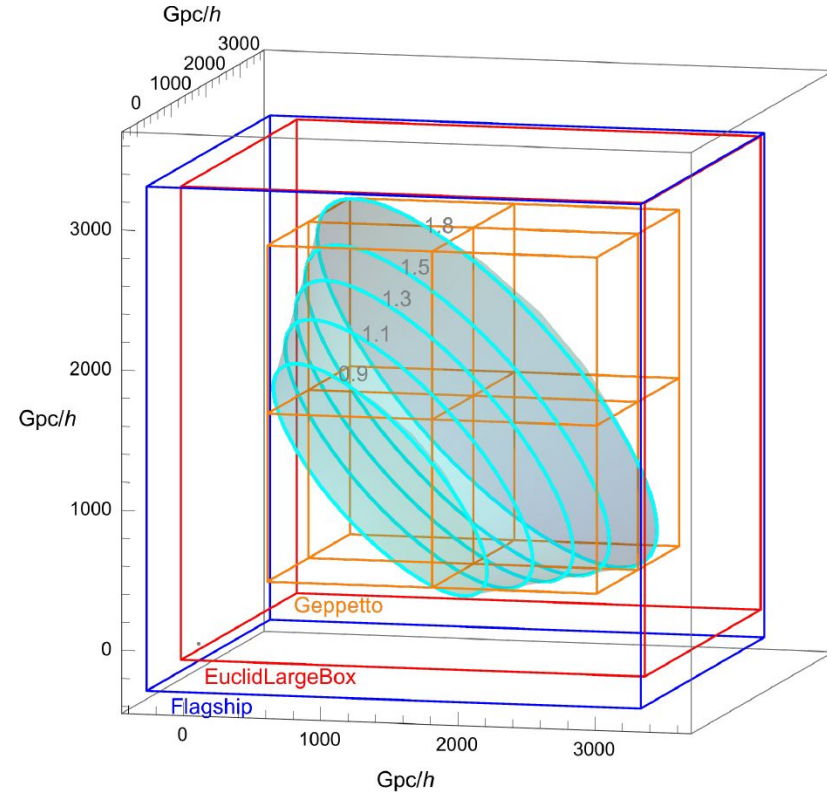
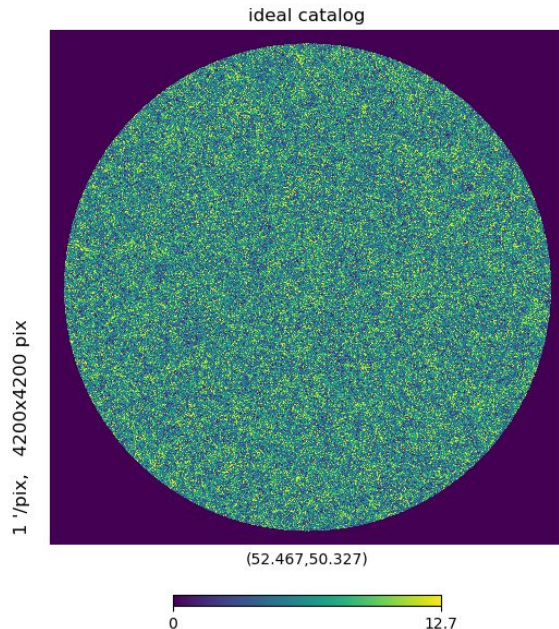
- 1000 realizations
- 3380 Mpc/h box, 6144^3 particles
- lightcone ($z < 4$): **half of the sky** (no MW)
- 4200 core-h per run, 4,200,000 core-h
- 220 GB each (80 GB for the lightcone), total 220 TB

30-degree-cone galaxy mocks

30 degree cone (circle of 30° of radius) survey footprint

1 mock: **2763 deg²**, larger than DR1 (~ 2000 deg²)

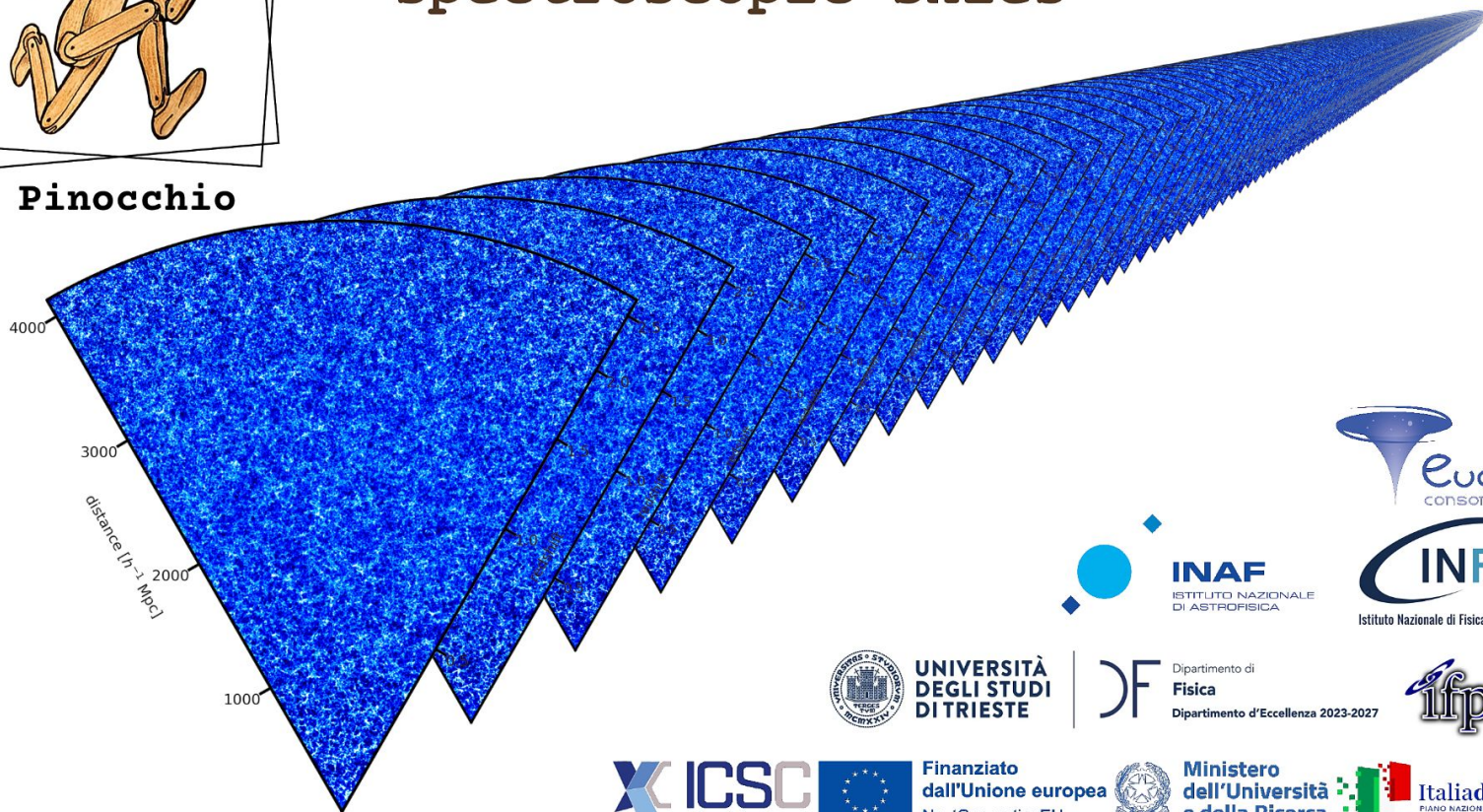
5 mocks: **13800 deg²**, similar to DR3





Pinocchio

Thousands of Euclid spectroscopic skies



**UNIVERSITÀ
DEGLI STUDI
DI TRIESTE**



Dipartimento di
Fisica
Dipartimento d'Eccellenza 2023-2027



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



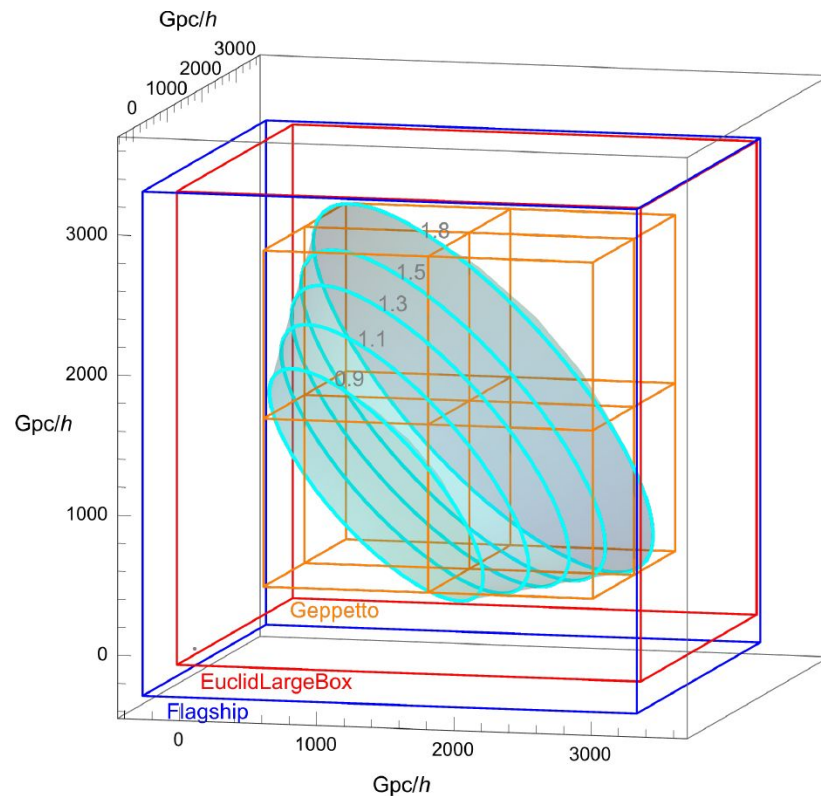
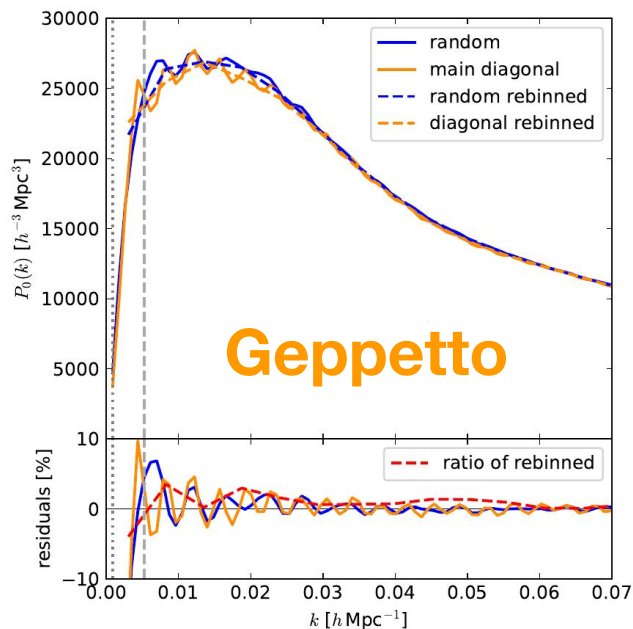
Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

30-degree-cone galaxy mocks

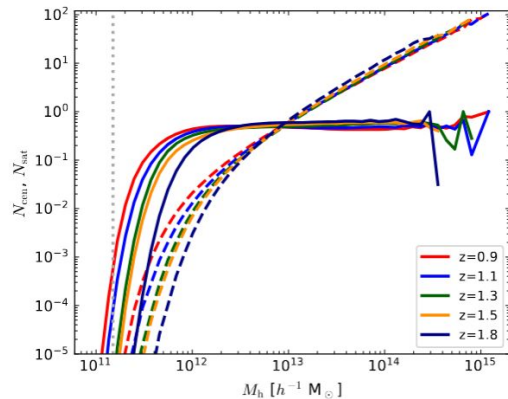
30 degree cone (circle of 30° of radius) survey footprint

1 mock: **2763 deg²**, larger than DR1 (~ 2000 deg²)

5 mocks: **13800 deg²**, similar to DR3

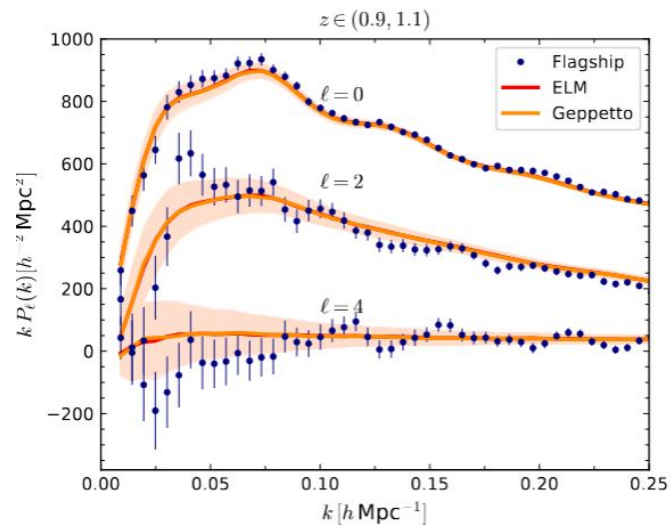
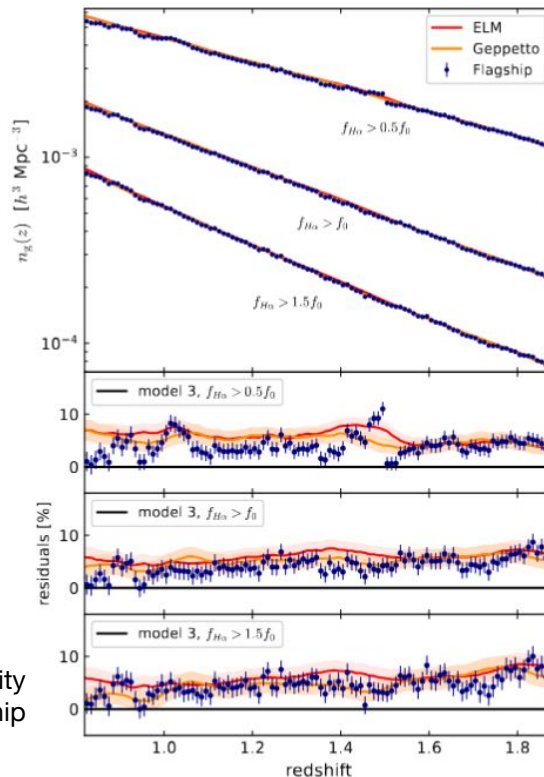


Calibrated on Flagship mock



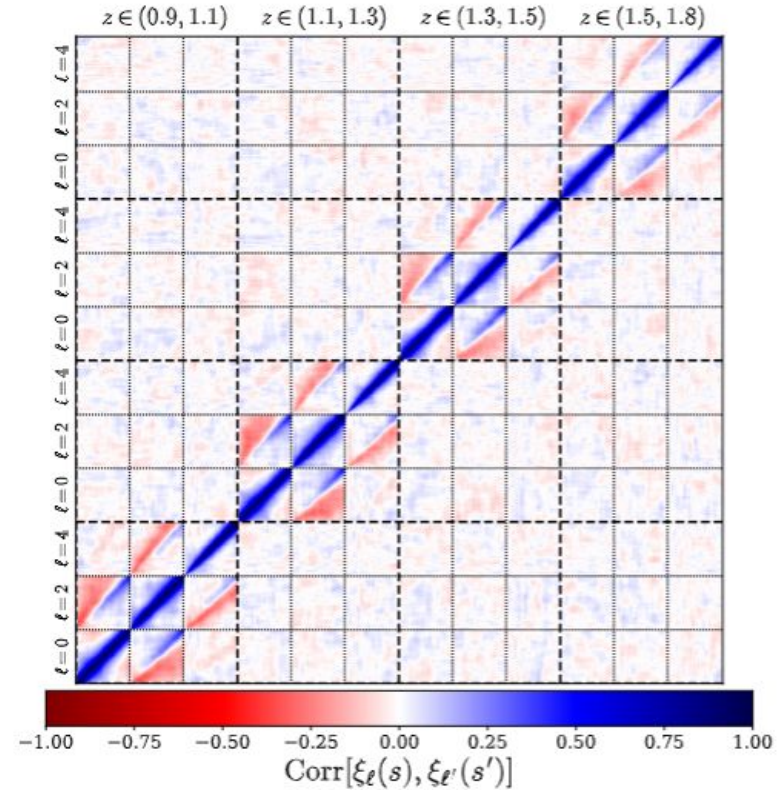
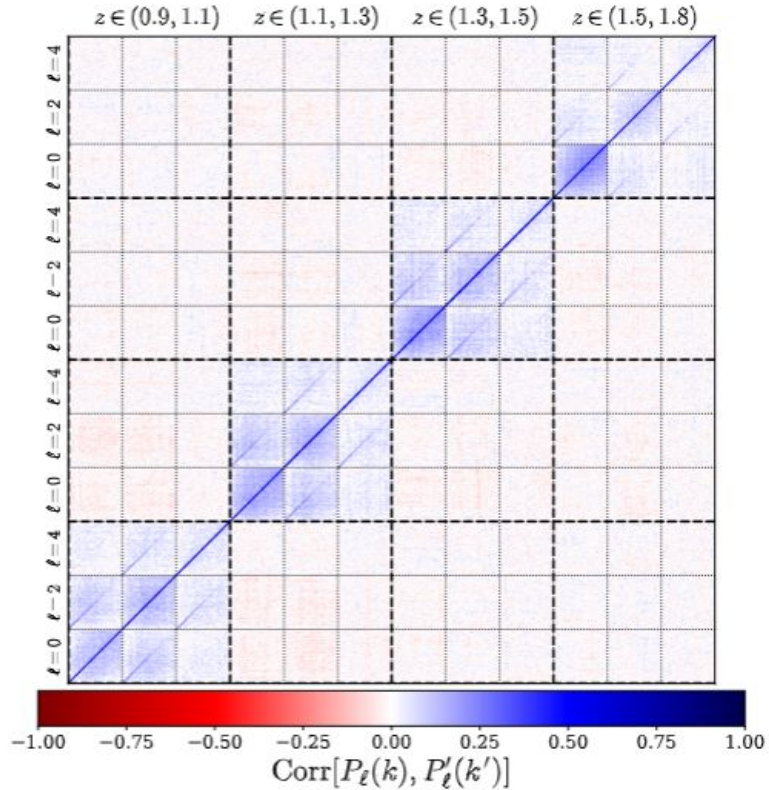
HOD curves $N_{\text{cen}}(M_h, z)$ and $N_{\text{sat}}(M_h, z)$

number density vs Flagship



galaxy power spectrum vs Flagship

Numerical covariance



Euclid Preparation.

Controlling data systematics in the Euclid spectroscopic galaxy sample: fluctuations of survey depth

Euclid Collaboration: P. Monaco^{1,2,3,4*}, M. Y. Elkhatab, B. Granett, J. Salvalaggio, E. Sefusatti, C. Scarlata, B. Zabelle, G. Addison, M. Bethermin, S. Bruton, C. Carbone, S. de la Torre, S. Dusini, A. Eggemeier, G. Lavaux, S. Lee, K. Markovich, K. McCarthy, M. Moresco, F. Passalacqua, W.J. Percival, I. Risso, A. Sanchez, C. Sirignano, Y. Wang, et al.

(Affiliations can be found after the references)

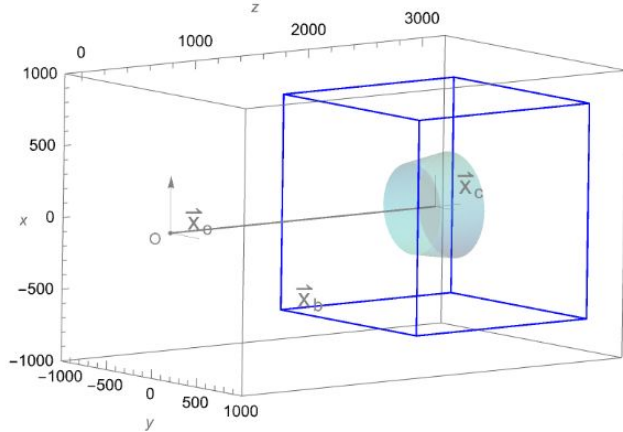
April 13, 2025

ABSTRACT

We present the strategy of *Euclid*'s galaxy clustering science working group to identify all potential sources of data systematics in the spectroscopic galaxy sample and quantify their impact on galaxy clustering measurements, on their covariance and on cosmological parameter estimation. We survey the processing pipeline to identify all possible sources of systematics, and classify them according to their impact, finding two broad classes: angular systematics, that modulate the galaxy number density across the sky, and redshift errors, that give origin to line and noise interlopers. We then test on simulated spectroscopic surveys our ability to mitigate realistic angular systematics by constructing a random catalogue that represents a dense set of unclustered objects subject to the same selection effects as the data catalogue; this random represents the visibility mask of the survey, and relies on a detection model that gives the probability to reliably measure the redshift of a galaxy as a function of the signal-to-noise ratio (SNR) of its emission lines. In the ideal case of perfect knowledge of the visibility mask, the galaxy power spectrum in presence of systematics is recovered, to within sub-percent accuracy, by convolving a theory power spectrum with a window function obtained from the random itself. In the case of approximate knowledge, we test the stability of power spectrum measurements and parameter posteriors by fitting the data using perturbed versions of the random. Appreciable effects are limited to very large scales, and parameter estimation results robust, with the most impacting effects being connected to the calibration of the detection model.

Key words. Surveys; Cosmology: observations; large-scale structure of Universe; cosmological parameters

Measuring the galaxy power spectrum



J. Salvalaggio, E. Sefusatti, A. Eggemeier,
K. Pardede, F. Rizzo

See talk by **Jacopo Salvalaggio**

With a survey geometry you are measuring the **convolution** of the galaxy power spectrum with a window function

$$\tilde{P}_\ell(k) = \sum_{\ell'} \int d^3 k' \mathcal{W}_{\ell\ell'}(k, k') P_{\ell'}(k')$$

$$W(\mathbf{x}) = W_{\text{fp}}(\hat{n}) w_{\text{fkp}}(z) \alpha n_r(\mathbf{x})$$

survey geometry FKP weights average density

Nuisance maps

tiling of the survey

Milky Way extinction

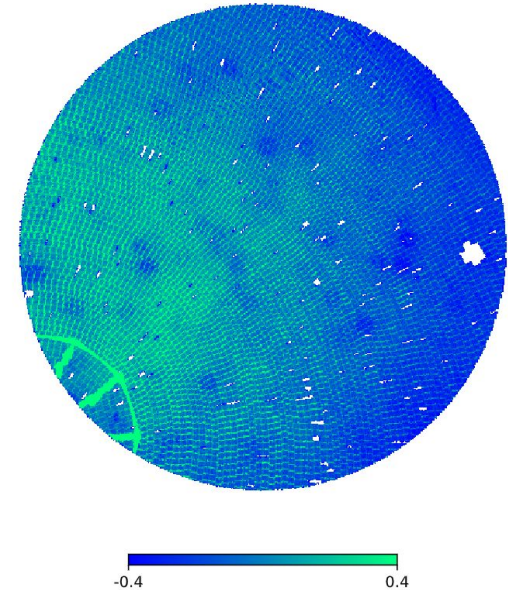
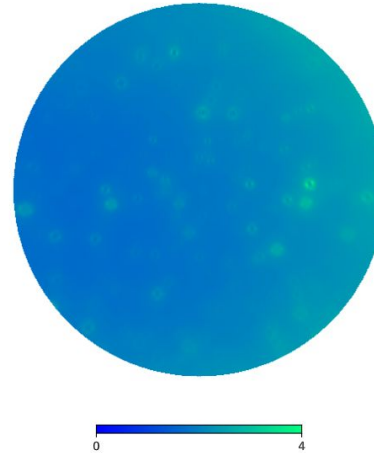
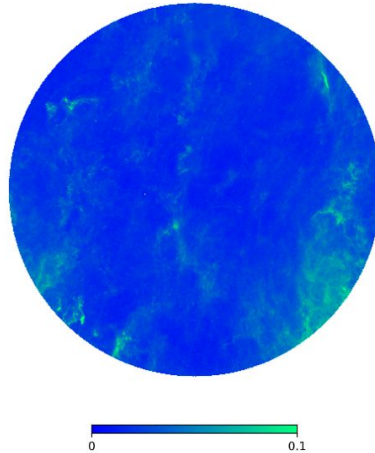
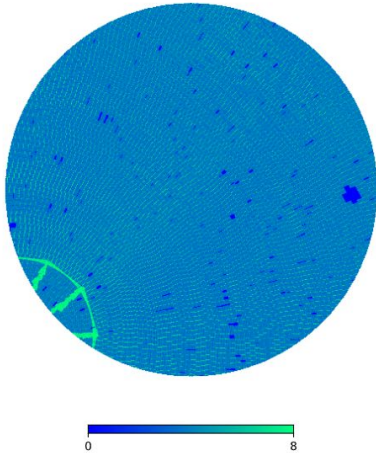
(simple) background noise

$$S = \kappa_1 f(\mathcal{E}) \sqrt{\frac{\mathcal{T}}{\mathcal{N}}}$$

exposure time (# of dithers)

reddening E(B-V)

noise (e⁻/s/pix)



The **random** catalog represents angular systematics

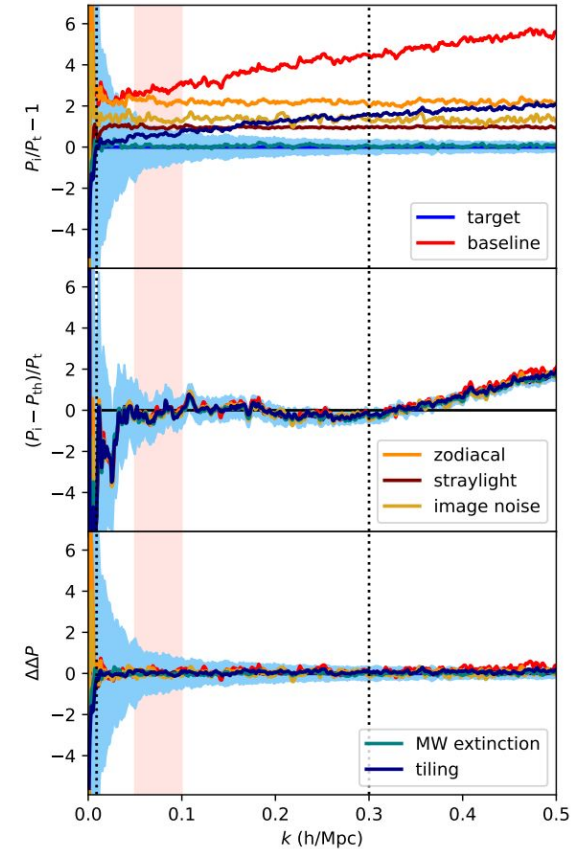
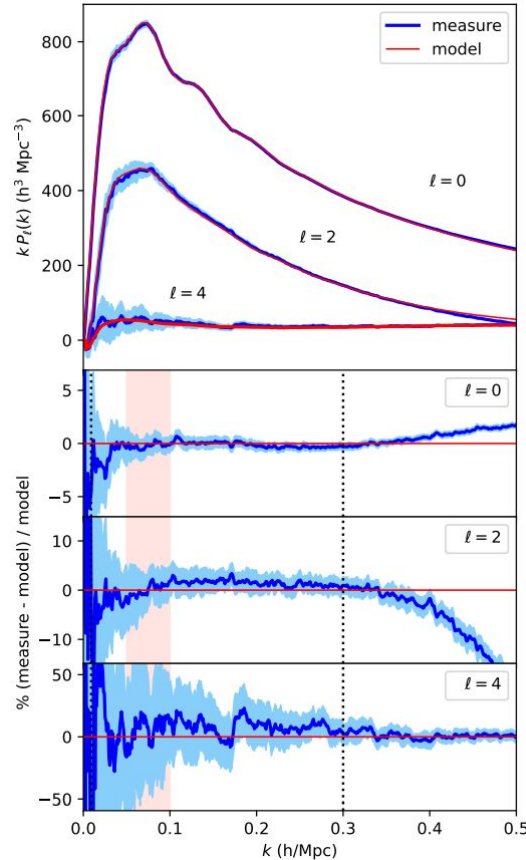
No bias when the window is known

we use **50 mocks**, averaging over groups of 5 - so we have 10 DR3-sized realizations

we use four **redshift bins** [0.9,1.1,1.3,1.5,1.8]

we fit the no-systematics case with **VDG** model

using the same model, we **convolve** it with different window functions and compare it with measurements



Euclid preparation

The impact of redshift interlopers on the two-point correlation function analysis

Euclid Collaboration: I. Risso^{*1,2}, A. Veropalumbo^{1,2,3}, E. Branchini^{3,2,1}, E. Maragliano^{3,2}, S. de la Torre⁴, E. Sarpa^{5,6,7}, P. Monaco^{8,9,7,10}, B. R. Granett¹, S. Lee¹¹, G. E. Addison¹², S. Bruton¹³, C. Carbone¹⁴, G. Lavaux¹⁵, K. Markovic¹¹, K. McCarthy¹¹, G. Parimbelli^{16,17,5}, F. Passalacqua^{18,19}, W. J. Percival^{20,21,22}, C. Scarlata²³, E. Sefusatti^{9,10,7}, Y. Wang²⁴, M. Bonici^{20,14}, F. Oppizzi², N. Aghanim²⁵, B. Altieri²⁶, A. Amara²⁷, S. Andreon¹, N. Auricchio²⁸, C. Raccialuni^{10,9,7,5}, M. Raldi^{29,28,30}, A. Balestra³¹, S. Bardelli²⁸, P. Rattalia²⁸, A. Rivano^{9,10}

May 9, 2025

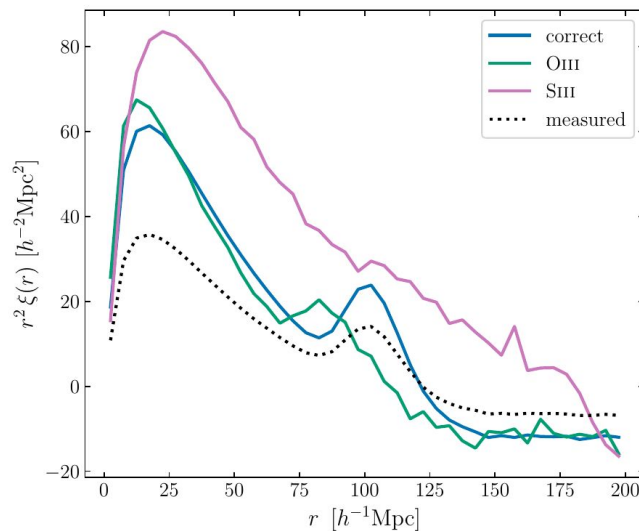
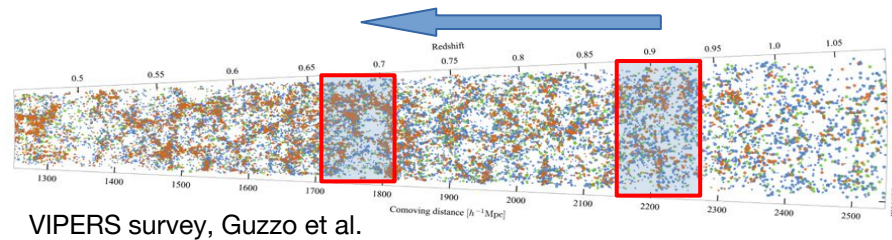
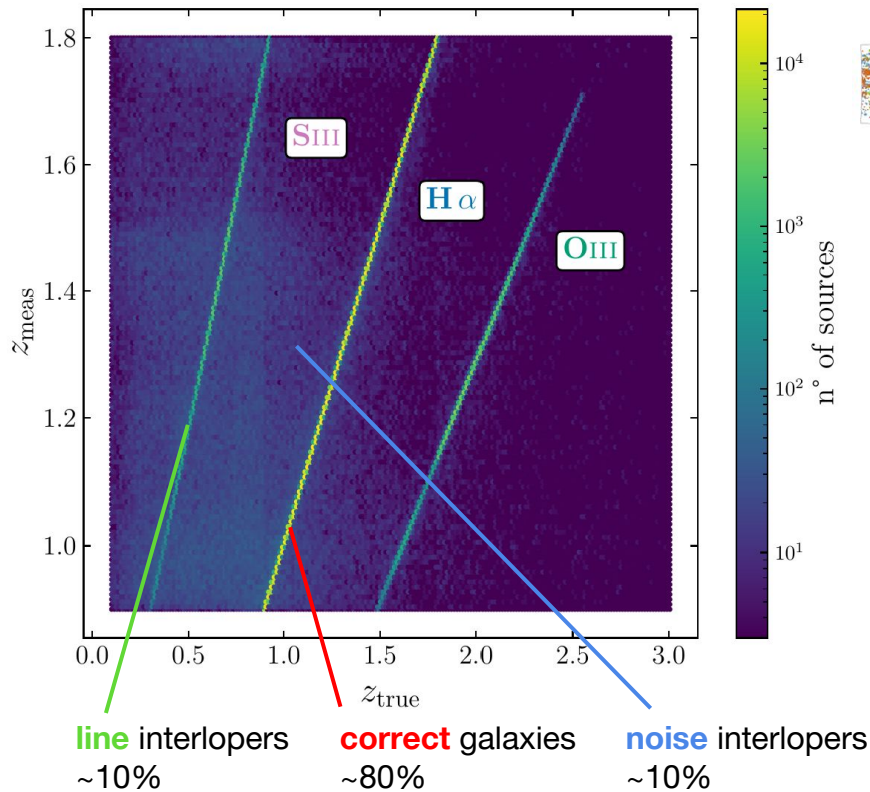
ABSTRACT

Context. The *Euclid* galaxy survey is designed to measure the spectroscopic redshift of emission-line galaxies (ELG) by identifying the H α emission line in their slitless spectra. The efficacy of this approach depends crucially on the signal-to-noise ratio (SNR) of the line, as sometimes noise fluctuations in the spectrum continuum can be misidentified as H α . In addition, other genuine strong emission lines can be mistaken for H α , depending on the redshift of the source. Both effects lead to ambiguities in the redshift measurement, potentially resulting in catastrophic redshift errors and the inclusion of ‘interloper galaxies in the sample.

Aims. This paper forecasts the impact on the galaxy clustering analysis of the expected redshift errors in the *Euclid* spectroscopic sample. Specifically, it investigates the effect of the redshift interloper contamination on the galaxy two-point correlation function (2PCF) and, in turn, on the inferred growth rate of structure $f\sigma_8$ and Alcock–Paczynski (AP) parameters, α_{\parallel} and α_{\perp} .

Methods. This work is based on the analysis of 1000 synthetic spectroscopic catalogues, the *EuclidLargeMocks*, which mimic the area and

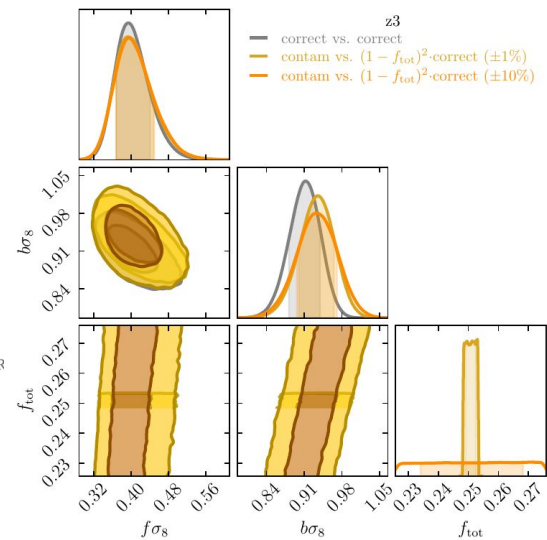
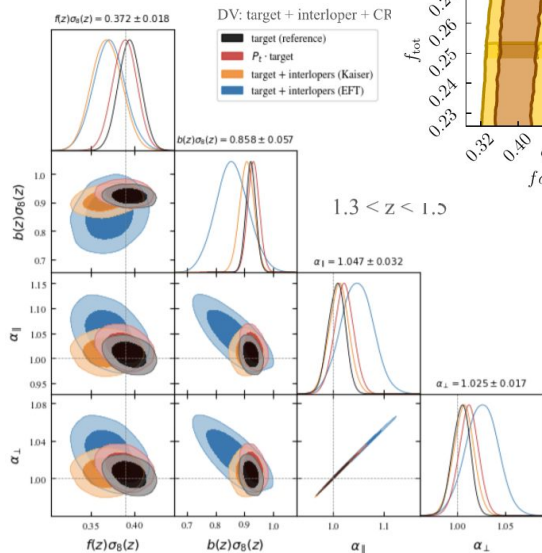
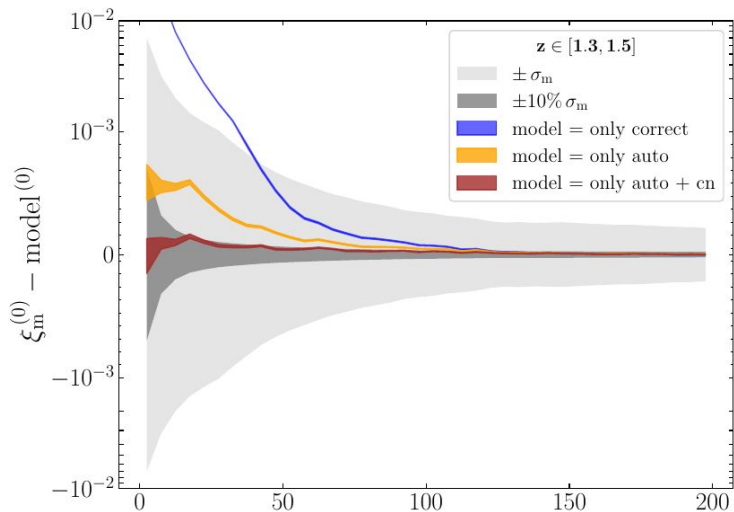
Line and noise interlopers





The role of interlopers

$$\xi_m(\mathbf{r}) = (1 - f_{\text{tot}})^2 \xi_{\text{cc}}(\mathbf{r}) + f_\ell^2 \xi_{\ell\ell}(\mathbf{r}) + f_n^2 \xi_{\text{nn}}(\mathbf{r}) + 2f_\ell(1 - f_{\text{tot}}) \xi_{\text{cl}}(\mathbf{r}) + 2f_n(1 - f_{\text{tot}}) \xi_{\text{cn}}(\mathbf{r}) + 2f_\ell f_n \xi_{\ell n}(\mathbf{r})$$



configuration space

BAO in Fourier space

At this level of contamination a simple $(1 - f_{\text{tot}})^2$ correction seems to be adequate

What do cosmologists care about galaxies?

Astrophysics is a nuisance to cosmologists!

Two approaches to take control of galaxy formation uncertainties:

- **Effective Field Theory**, where galaxy bias is expanded at a part of theory
 - pro: no assumption on how galaxy formation behaves
 - con: many parameters to be fitted, and counterterms can go wild! => projection effects
- example: **Velocity Difference Generator** makes assumption on the behaviour of fingers of God, gaining much better stability at high k

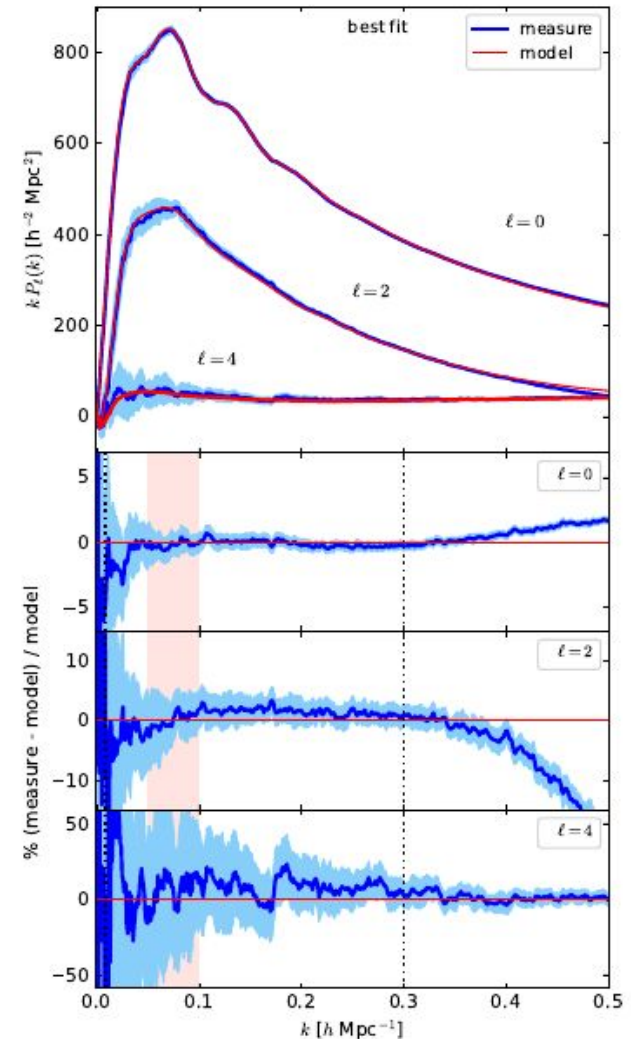
Parameter	EFT	VDG
h	$\mathcal{U}(0.5, 1.0)$	★
ω_c	$\mathcal{U}(0.085, 0.155)$	★
A_s	$\mathcal{U}(1.4, 2.6)$	★
n_s	$\mathcal{N}(0.96, 0.0041)$	★
ω_b	$\mathcal{N}(0.02218, 0.00055)$	★
b_1	$\mathcal{U}(0.25, 4.0)$	★
b_2	$\mathcal{U}(-5, 5)$	★
γ_2	$\mathcal{U}(-5, 5)$	★
γ_{21}	Coevolution	★
c_0 [Mpc ²]	$\mathcal{N}(0, 200)$	$\mathcal{U}(-400, 400)$
c_2 [Mpc ²]	$\mathcal{N}(0, 200)$	$\mathcal{U}(-400, 400)$
c_4 [Mpc ²]	$\mathcal{N}(0, 200)$	$\mathcal{U}(-400, 400)$
c_{nlo} [Mpc ⁴]	$\mathcal{N}(0, 200)$	–
a_{vir} [Mpc]	–	$\mathcal{U}(0, 20)$
N_0^P	$\mathcal{N}(1, 3)$	$\mathcal{U}(0, 3)$
N_{20}^P [Mpc ²]	$\mathcal{N}(0, 3)$	$\mathcal{U}(-3, 3)$
N_{22}^P [Mpc ²]	–	–

What do cosmologists care

Astrophysics is a nuisance to cosmologists!

Two approaches to take control of galaxy formation uncertainties:

- **Effective Field Theory**, where galaxy bias is expanded at a part of theory
 - pro: no assumption on how galaxy formation behaves
 - con: many parameters to be fitted, and counterterms can go wild! => projection effects
- example: **Velocity Difference Generator** makes assumption on the behaviour of fingers of God, gaining much better stability at high k

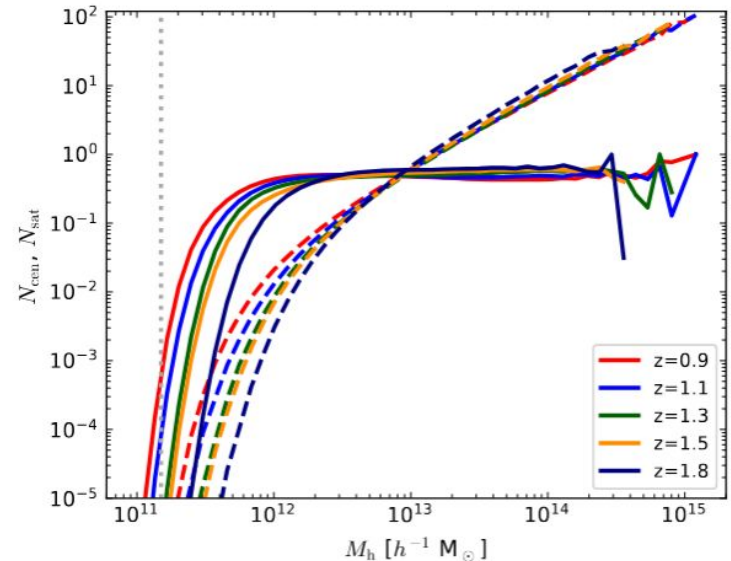


What do cosmologists care about galaxies?

Astrophysics is a nuisance to cosmologists!

Two approaches to take control of galaxy formation uncertainties:

- **Halo Occupation Distribution**, where the probability of having observed galaxies in the halo is assumed to be a function of halo mass and redshift
 - pro: simple approximation that can be used with N-body simulations
 - con: it neglects **secondary bias**, the dependence of galaxy properties on other halo properties (spin? environment?) and on its merger history (**assembly bias**)



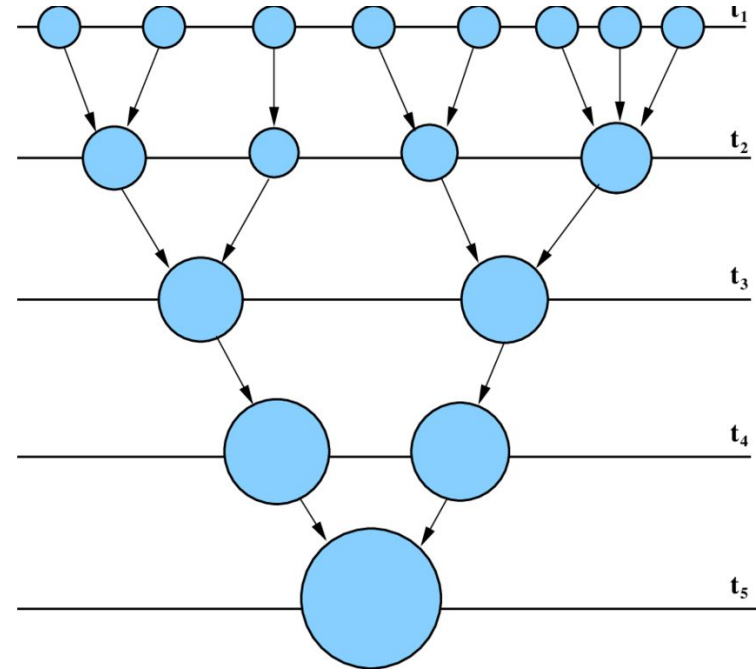
What do cosmologists care about galaxies?

Astrophysics is a nuisance to cosmologists!

Two approaches to take control of galaxy formation uncertainties:

- **Halo Occupation Distribution**, where the probability of having observed galaxies in the halo is assumed to be a function of halo mass and redshift
 - pro: simple approximation that can be used with N-body simulations
 - con: it neglects **secondary bias**, the dependence of galaxy properties on other halo properties (spin? environment?) and on its merger history (**assembly bias**)

Is it true that galaxy properties are a function of the galaxy merger tree + halo spin?



What do astrophysicists care?

What can we learn on galaxy formation from cosmology?

- **The method!** can we have “precision galaxy formation”?
- Shouldn't we add basic **galaxy observables to our data vector** when we fit for cosmology?
- Shouldn't we use **bias parameters** as a further constraint to galaxy formation?

an example of an astrophysics-aware SBI

