PhD in Physics and Nanoscience – XXXVIII cycle third year report
Antonio Farina

Tutors: Enzo Franco Branchini
Benjamin Rudolph Granett
Alfonso Veropalumbo

Research activity:

During my third PhD year, I advanced my research in Cosmology as part of the European Space Agency's Euclid mission. I continued to build on my previous work while actively branching out into new areas of study.
Most of my efforts have focused on Euclid's systematic effects, particularly on accurately characterizing the spectroscopic selection function. This is typically handled with random catalogs generated via the spectroscopic visibility mask (VMSP), which corrects for observational systematics using the Euclid Deep Field—a small but exceptionally pure and complete sample of Euclid galaxies. However, the Deep Field will not be available in time for the first data releases, making alternative approaches essential. To address this, I developed a method to generate random catalogs directly from Euclid Wide Survey data, avoiding reliance on the Deep Field. The approach corrects for systematic variations in galaxy number density by regressing the observed density against potential sources of systematics. The resulting trends are then applied to a *master random catalog*, a synthetic, spatially uniform catalog over the survey footprint—by assigning weights that suppress contaminated modes, thereby encoding the selection effects into the randoms. To perform this regression in a flexible, non-linear way, I employ Self-Organizing Maps (SOMs). The SOM clusters galaxies in the multi-dimensional space of observational parameters (e.g., noise properties, extinction, photometric calibration), capturing complex dependencies of galaxy density on systematics. Comparing the observed density across SOM cells with the expectation from the master random catalog allows us to derive corrective weights, which are then applied to the randoms. To validate the method, I generated 100 Euclid-like mock catalogs using the *Pypelid* survey simulator, an observation emulator capable of reproducing on simulations the main survey strategy and systematics. The starting point are 100 *Pinocchio* RR2 mocks, covering a nearly 500 deg^2 area of the DR1 footprint used to test the pipeline. Galaxy properties required by Pypelid are drawn from the Euclid Flagship II simulation. Measuring the two-point correlation function with both VMSP and SOM-based randoms, we find that the latter reproduces the former with discrepancies of only a few percent of the statistical uncertainty, demonstrating the accuracy and reliability of the approach. The validation and benchmarking of the SOM-based method is still ongoing; however, I have already exploited it to deliver the very first measurements of the galaxy two-point correlation function from Euclid data.

In addition to my work on systematic effects and selection functions, I contributed to the activities of the Euclid Science Working Groups (SWGs) and the Ground Segment, focusing on incorporating higher-order statistics into the official data analysis pipelines. In particular, I helped develop and validate robust 3PCF measurement pipelines capable of operating under realistic observational conditions. These tools were designed to scale efficiently with the large data volumes expected from Euclid while remaining resilient to observational systematics. In parallel, I contributed to efforts evaluating the regime of validity of perturbative models for both two- and three-point statistics. Within this framework, and leveraging a custom-built emulator, we carried out the first joint full-shape analysis of the 2PCF and 3PCF ever attempted.

In a related but independent line of research, I extended the theoretical and computational framework developed in Farina et al. (2024) to compute, for the first time, the velocity divergence-density-density 3-point correlation function. This novel extension provides a consistent and unified treatment of density and velocity statistics within the same multipole-based redshift-space formalism. As the velocity field is particularly sensitive to the growth of structure and potential deviations from General Relativity, this framework offers new opportunities for constraining fundamental physics with upcoming spectroscopic surveys. This work is still in a preliminary stage, but it lays the foundation for future analyses of higher-order velocity statistics.

Finally, I focused on one of key challenge in cosmological parameter inference: the accurate estimation of covariance matrices, which encode both statistical uncertainties and correlations in the data. Standard analyses rely on the sample covariance from simulations, but as the data vector length D approaches the number of realizations N, the sample precision matrix becomes increasingly noisy (and can even become singular if $N \leq D$). Generating the large suites of simulations required to avoid this is often computationally prohibitive. To overcome this limitation, we apply for the first time on a cosmological dataset a modified covariance estimator, the Rotational Invariant Estimator, which uses Random Matrix Theory to de-noise high-dimensional sample covariances. We benchmark its performance against the non-linear shrinkage estimator NERCOME in the context of galaxy clustering analyses. The first, preliminary results seem to indicate that both methods yield consistent, reliable precision matrices even with a very limited number of simulations.

Attended courses

I gave a report on the PhD course I attended last accademic year:

1. 5th Azores PhD school on Observational Cosmology, 02-08 September 2024, Angra do Heròismo, Açores, Portugal

And, during the present accademic year, I attended and passed the following courses:

2. Astrofisica e Cosmologia Computazionale – Master degree course by Prof. Alfonso Veropalumbo and Prof. Marco Raveri

Publications:

1. *Euclid Collaboration: M. Guidi et al., "Euclid preparation. Full-shape modelling of 2-point and 3-point correlation functions in real space", submitted for pubblication to A&A (arXiv: 2506.22257)*
2. *Euclid Collaboration: Y. Mellier et al., "Euclid. I. Overview of the mission", A&A 697, A1 (2025)*
3. A. Farina, A. Veropalumbo, E. Branchini, M. Guidi, "*Modeling and measuring the anisotropic halo 3-point correlation function: a coordinated study",* submitted for publication to JCAP (arXiv:2408.03036)
4. *A. Farina, M. Guidi, A. Veropalumbo, E. Branchini, "Denoising cosmological covariance matrices with rotational invariant estimators", in preparation.*
5. M. Guidi, A. Farina, A. Nusser, E. Branchini, "Modelling the galaxy-galaxy-velocity three-point correlation function", in preparation.
6. Euclid Collaboration: A. Veropalumbo et al., "*Euclid preparation: TBD. Three-dimensional galaxy clustering in configuration space. Part II Three-point correlation function estimation*", submitted for internal review of the Euclid Collaboration.

Conference presentations:

1. Speaker at the Euclid Galaxy Clustering meeting, 20th-24th January 2025, Garching, Germany – **Random catalogs for early spectroscopic data**
2. Speaker at the "*Optimizing the Extraction of Cosmological Information from the Latest Spectroscopic Redshift Surveys*" meeting, 14th-18th July 2025, Sexten (Bz), Italy - **Characterizing selection effects in the Euclid spectroscopic survey**

Teaching activity:

1. Tutor of General Physics for 1st year Management Engineering students
2. Tutor of General Physics for 1st year Biomedical Engineering students