Cross-Correlation Studies for Detecting Interlopers in the Euclid Survey

Alessandro Di Gregorio[1, a)] and Benjamin Granett[2, b)]

[1]*Physics and Astronomy Department, Washburn University, 1700 SW College Avenue, Topeka, Kansas 66621, USA*
[2]*Osservatorio Astronomico di Brera-INAF, Via Brera, 28, 20121 Milano MI, Italy*

[a)]Corresponding author: alessandro.digregorio@washburn.edu
[b)]benjamin.granett@inaf.it

Abstract. To probe the nature of dark energy, Euclid is carrying out imaging and spectroscopy over 14,000 square degrees of extragalactic sky for galaxy clustering and weak lensing measurements. Euclid's near infrared spectrometer and photometer (NISP) is a slitless grism spectrograph that records spectra of all sources in the field of view. The principal targets of the spectroscopic campaign are α emission line galaxies from redshift 0.9 to 1.8. However, slitless spectroscopy poses a challenge due to the superposition of spectra on the detector, which can lead to source confusion and spurious detections. In addition, only a single strong line (Hα) is in the wavelength range of the spectrograph from redshift 0.9 to 1.5. Therefore, if a different line is identified as α, a wrong redshift will be assigned. This leads to galaxy interlopers in the catalog, which originate from different redshifts. Our method aims to infer the fraction of galaxy interlopers in redshift bins using angular cross-correlation statistics between the Euclid catalog and a reference galaxy catalog that has secure redshift determinations. We present preliminary results on low redshift interlopers based on the Euclid Flagship mock galaxy catalog. We find that using a DESI-like catalog covering 5,000 square degrees, we can infer the interloper fraction originating from galaxies at redshift 0 to 0.5 with a typical error of 1%, with a systematic uncertainty related to the galaxy bias. Thus, if spectroscopic surveys are available that overlap on the sky, the angular cross-correlations are a promising tool to assess the interloper fractions.

INTRODUCTION

The discovery of the accelerated expansion of the universe has called attention to different hypotheses formulated in the last century. Among the possible explanations, the lambda cold dark matter (ΛCDM) model introduces a form of energy with negative pressure, also referred to as dark energy. This form of energy does not emit, absorb, or reflect light and is responsible for the rate of expansion of the universe. The Euclid mission is an example of an ongoing survey from space that aims to shed light on the mystery. One of the main observational probes used to test the cosmological model is galaxy clustering. By collecting cosmological redshift values for distant galaxies and measuring their clustering pattern, we can identify the dominant components of the universe that drive the dynamics. We can constrain the mass energy density and fundamental properties, such as the equation of state of dark energy [1].

At the moment, two spectroscopic surveys are underway to achieve this goal, the Dark Energy Spectroscopic Instrument (DESI) and the ESA Euclid space telescope. They are complementary experiments that survey different redshift ranges and galaxy populations. DESI measures the expansion history or distance-redshift relationship from the local universe to redshift 3.5 [2]. In the Euclid mission, the near infrared spectrometer and photometer (NISP) uses a grism to disperse and record spectra of all sources in the field of view. However, it does not make use of slits or fibers to target galaxies as in multiobject spectrographs. The wavelength range of its red grism was chosen to sample the $0.9 < z < 1.8$ redshift range and measure the redshift of an expected 20–30 million emission line galaxies, while DESI will target the lower-redshift side. Grism slitless spectroscopy poses a challenge because of the superposition of spectra on the detector, which increases the measurement noise. It is expected that for a subset of the measurements, a wrong redshift will be assigned in the range of 0.9–1.5 when a different line is identified as the strong line instead of the main target emission line (α). This event is referred to as interloper contamination, which has been of recent interest for low-resolution spectroscopic surveys like Euclid [1].

PROCEDURE

The aim of this work is to measure the fraction of interloper galaxies in the Euclid spectroscopic sample. Previous works have improved the use of cross-correlation between mock catalogs for the search of interlopers, providing a more accurate correction for contamination in spectroscopic surveys like the Hobby-Eberly Telescope Dark Energy

Experiment (HETDEX) and Euclid than assuming a uniform interloper rate [3]. Also, cross-correlation between mock Euclid photometric samples and spectroscopic samples from surveys like the Baryon Oscillation Spectroscopic Survey (BOSS) and DESI has been used to measure small-scale clustering redshifts, enabling calibration of the true redshift distributions and determination of the mean redshift with the precision required by Euclid's cosmological goals [4]. Our findings confirm the utility of cross-correlation methods, in this case, applied for the detection of interlopers in future catalogs. We followed a procedure described by Scottez et al. in [5], using the mathematical formalism of the angular cross-correlation function. The main idea relies on the fact that the positions of galaxies in space are correlated through the underlying density field. Thus, given a galaxy sample with unknown redshifts, we can infer the redshift distribution by measuring the cross-correlation signal with a reference sample for which we know the redshifts.

We are interested in estimating the fraction of interlopers in the Euclid sample. We aimed to infer this by measuring the cross-correlation signal with a sample of galaxies with known redshift from the DESI survey. We selected two mock galaxy samples from the Euclid Flagship mock galaxy catalog [6]. The α sample (with a range of $0.9 < z_{\text{true}} < 1.1$) represented Euclid targets. The low redshift sample with a limit in the magnitude *r* band (indicatively set to $18 < r < 19$), which we called the reference catalog, represented the DESI sample with secure redshift determinations. Then, we constructed an observed Euclid catalog (target catalog) that included interlopers by mixing a fraction of low redshift galaxies with the α sample. In a flux bin $r > 19$ we selected 10% low-z interlopers and a 90% α sample. Most of the galaxies were between the range $z < 0.5$, as shown in Fig 1. The cross-correlation measurement also required a random catalog that traced the same survey area. We constructed a random catalog 10 times larger than the reference catalog.

Formalism

We used the Landy-Szalay estimator for our correlation function in the form used in [7], in which pair-count normalizations are written explicitly:

$$\omega(\theta) = \frac{N_R(N_R-1)}{N_D(N_D-1)}\frac{DD(\theta)}{RR(\theta)} - \frac{N_R-1}{N_D}\frac{DR(\theta)}{RR(\theta)} + 1, \tag{1}$$

where DD, DR, and RR are the data-data, data-random, and random-random pair counts, and N_{D} and N_{R} are the number of objects in the data and random catalogs, respectively. We model the angular autocorrelation function of the reference sample as

$$\omega_{\text{ref}}(\theta) = \int dz \left(\frac{dN_{\text{ref}}}{dz} b_{\text{ref}}\right)^2 \omega_m(\theta), \tag{2}$$

where $\omega_m(\theta)$ is the angular correlation of the underlying matter density field, b_{ref} is the galaxy bias, z is the redshift, and $\frac{dN_{\text{ref}}}{dz}$ is the normalized redshift distribution. The cross-correlation between the Euclid and reference samples, which includes the interloper contribution, is modeled as

$$\omega_x(\theta) = \int dz \left(\frac{dN_{\text{ref}}}{dz} b_{\text{ref}}\right)\left(\frac{dN_{\text{Euclid}}}{dz} b_{\text{interloper}}\right)\omega_m(\theta). \tag{3}$$

The interloper fraction is defined as

$$f_{\text{interloper}} = \frac{N_{\text{interloper}}}{N_{\text{interloper}} + N_{\text{true}}}. \tag{4}$$

The normalized redshift distribution $\frac{dN_{\text{Euclid}}}{dz}$ is a function of the true redshift of the galaxies and includes the low redshift galaxies that enter as interlopers. Since the distribution integrates to 1, it is exactly equal to the interloper fraction. We use the average correlation function computed over angular scales with a window to compress information and express the correlation as a function of redshift. Following the standard power-law behavior of the angular galaxy correlation function, we introduce the term $\theta^{-0.8}$ [8]. The exponent -0.8 was chosen to match the shape of the galaxy angular correlation function.

$$\varpi(z) = \frac{\int \theta^{-0.8}\omega(\theta)d\theta}{\int \theta^{-0.8}d\theta}. \tag{5}$$

Given this parametrization, we can constrain the interloper fraction from the ratio of the cross-correlation to the reference autocorrelation. We consider a narrow redshift bin of the reference sample such that the integrals in Eqs. (2) and (3) can be dropped. This is valid as long as the bias factors are slowly varying. This results in the relation

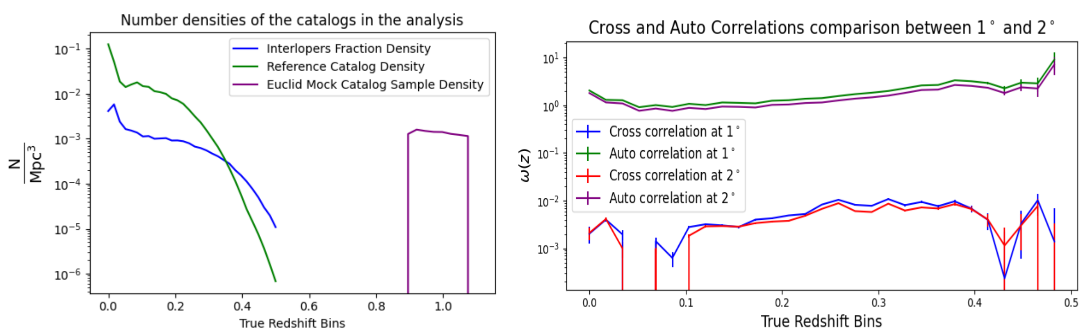
$$f_{\text{interloper}}(z) = \frac{\omega_x(z)}{\omega_{\text{ref}}(z)} \frac{b_{\text{ref}}(z)^2}{b_{\text{ref}}(z) b_{\text{interloper}}(z)}. \tag{6}$$

The ratio of the galaxy biases $\frac{b_{\text{ref}}}{b_{\text{interloper}}}$ of the reference and interloper samples cannot be directly measured in observations but is expected to be close to 1. We use the fiducial value of unity in our analysis. To validate this assumption, we measure the ratio in the mock catalogs. In the mock catalog, the bias ratio can be inferred from the ratio of the autocorrelation functions of the reference and interloper samples as

$$\frac{b_{\text{ref}}}{b_{\text{interloper}}} = \sqrt{\frac{\omega_{\text{ref}}}{\omega_{\text{interloper}}}}. \tag{7}$$

DATA ANALYSIS

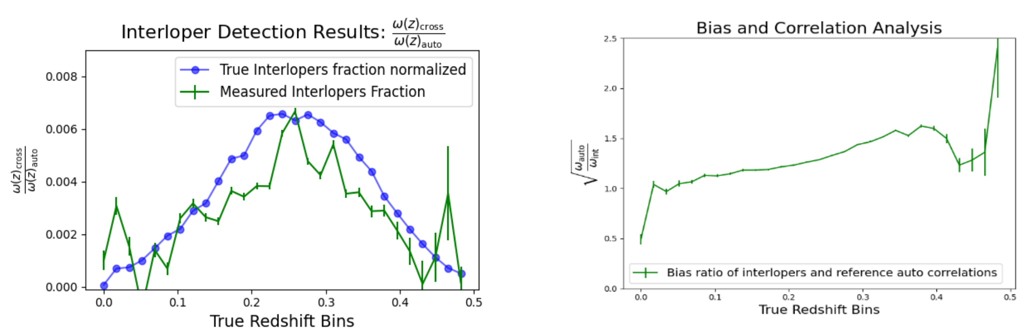
We began our analysis by measuring the angular correlation function. We developed a program that could perform cross-correlations and autocorrelations between a target, reference, and random catalog. Figure 1 shows the number densities of the different components of our analysis. The reference and interloper samples are located at low redshift $z < 0.4$, while the Euclid target sample that we consider is in the range $0.9 < z < 1.1$.



(a) (b)

Figure 1. (a) The number densities in units of Mpc^3. (b) The mean correlations as a function of redshift of the catalogs viewed in (a) for two configurations of the maximum angular separation.

We computed the correlation functions in 29 redshift bins between redshift 0 and 0.5. The maximum angular separation used in the correlation function is a parameter that must be fixed in the analysis. We carried out two tests with the maximum scale set at 1° and 2° to test consistency and evaluate if 1° could be chosen for faster computations. Figure 1(b) shows the measured correlation functions for the two configurations. The error bars represent absolute uncertainties calculated through error propagation after considering the Poisson error of the couples of the correlation in the Landy-Szalay estimator. A shift between the two results computed at 1° and 2° maximum separation is noticeable in Fig. 1(b). Indeed, the signal is concentrated at small separations, and the autocorrelation function integrated to 1° is higher than when integrated to 2°. We thus performed our analysis setting the maximum scale to 1°, which is also faster to compute.



(a) (b)

Figure 2. Plot (a) represents the actual interloper fraction compared to the one inferred from the correlations. Plot (b) represents the study of the bias ratio, which is the primary systematic uncertainty in the analysis.

RESULTS

The inference of the interloper fraction results is presented in Fig. 2. We compare the interloper fraction inferred from the ratio of the cross-correlation to the reference autocorrelation functions to the truth in the mock catalogs. To measure the interloper fraction, we fix the bias ratio to unity in Eq. (5). We find that the interloper fraction varies with redshift bin from 0 to 0.7%. The inferred interloper fractions follow the redshift dependence with an average error of 0.11%. The total inferred interloper fraction was 8.4%. Compared with the true value of 10%, this gives an error of 1.6%.

The error bars shown in Fig. 2(a) were derived from the Poisson error on the pair counts. As an alternative method to assess the measurement error, we performed 10 runs with different realizations of the random catalog. By measurement error we refer to the actual difference in the performances of multiple iterations with diverse random catalogs. We found that the results are robust, with an uncertainty of 0.9%. Our error analysis neglects the contribution from cosmic variance. The reference and interloper samples cover the same volume, and thus sample variance fluctuations will cancel in the ratio of the correlation functions. However, the Euclid sample, which is at much higher redshift, is uncorrelated and contributes noise to the cross-correlation function. We will assess the sample variance uncertainty in the next stage of this work. We expect the cosmic variance error to be larger than what we found.

The primary systematic uncertainty in the analysis arises from the galaxy bias, which enters in the ratio of correlation functions. Figure 2(b) shows the evolution of the bias ratio [Eq. (6)]. As can be seen, the bias ratio varies slowly between approximately 1 and 1.5 over the redshift bins. This confirms that the shape of the interloper redshift distribution we derived is not strongly affected by the bias factors. Considering a bias ratio of 1.2, we can understand the low value of the interloper fraction that we derived (8.4% compared with the true value of 10%). Correcting the measured interloper fraction by the bias ratio factor of 1.2 gives a result in agreement with the expectation of 10%.

CONCLUSION

Cross-correlation analyses have proven to be a helpful tool when studying interloper contamination for future catalogs. We test the approach using mock catalogs representing Euclid and DESI spectroscopic measurements in which the Euclid catalog contains interlopers from low redshift $z < 0.5$, which overlaps with DESI. The cross-correlation signal shows the expected behavior compared to the autocorrelation, with a shape influenced by interlopers in the target catalog. However, more work is required to assess the accuracy of the analysis when applied to actual survey data from Euclid and DESI. Improved modeling of the galaxy bias beyond the simple linear bias assumption may improve the results further. We must also take into account the uncertainty due to cosmic variance in the Euclid and DESI samples. The effort will allow us to introduce proper error bars in the interloper function.

In conclusion, cross-correlation methods will be a valuable resource for detecting interloper contamination in the upcoming Euclid catalogs. As part of our ongoing work, we are also considering how to detect interlopers originating from higher redshift using photometric redshift catalogs.

REFERENCES

1. Y. Mellier, J. A. Barroso, A. Achúcarro, J. Adamek, R. Adam, G. E. Addison, N. Aghanim, M. Aguena, V. Ajani, Y. Akrami, et al., [preprint] arXiv:2405.13491 (2024).
2. B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, D. M. Alexander, R. Alfarsy, L. Allen, C. A. Prieto, O. Alves, J. Ameel, et al., Astrophys. J. **164**(5), 207 (2022).
3. D. J. Farrow, A. G. Sánchez, R. Ciardullo, E. M. Cooper, D. Davis, M. Fabricius, E. Gawiser, H. S. Grasshorn Gebhardt, K. Gebhardt, et al., Mon. Not. R. Astron. Soc. **507**(3), 3187–3206 (2021).
4. K. Naidoo, H. Johnston, B. Joachimi, J. L. Van Den Busch, H. Hildebrandt, O. Ilbert, O. Lahav, N. Aghanim, B. Altieri, A. Amara, et al., Astron. Astrophys., **670**, A149 (2023).
5. V. Scottez, Y. Mellier, B. R. Granett, T. Moutard, M. Kilbinger, M. Scodeggio, B. Garilli, M. Bolzonella., S. de la Torre, L. Guzzo, et al., Mon. Not. R. Astron. Soc., **462**(2), 1683–1696 (2016).
6. F. J. Castander, P. Fosalba, J. Stadel, D. Potter, J. Carretero, P. Tallada-Crespí, L. Pozzetti, M. Bolzonella, G. A. Mamon, L. Blot, et al., "Euclid. V. The Flagship galaxy mock catalogue: A comprehensive simulation for the Euclid mission," [preprint] arXiv:2405.13495 (2024).
7. H. Kurki-Suonio, *Galaxy Survey Cosmology* (University of Helsinki, lecture notes, 31.5.2023), pp. 85–87.
8. A. J. Connolly, R. Scranton, D. Johnston, S. Dodelson, D. J. Eisenstein, J. A. Frieman, J. E. Gunn, L. Hui, B. Jain, S. Kent, et al., Astrophys. J. **579**(1), 42 (2002).