Euclid preparation. XII. Optimizing the photometric sample of the Euclid survey for galaxy clustering and galaxy-galaxy lensing analyses

Euclid Collaboration: A. Pocino1*, I. Tutusaus2,3, F.J. Castander2,3, F. Posalba2,3, M. Crocce1,2, A. Porredon2,3,4,5, S. Camera6,7,8, V. Cardone9, S. Casas10, T. Kitching11, F. Lacasa12, M. Martinelli13, A. Pourtisidou14, Z. Sakr15,16, S. Andreoni17, N. Auricchio18, C. Baccigalupi19,20,21,22, A. Balaguera-Antonio13,23,24, M. Baldi25,26,27, A. Balestra28, S. Bardelli29, R. Bender29,30, A. Biviano19,22, C. Bodendorf30, D. Bonino8, A. Boucaud31, E. Bozzo32, E. Branchini33,34, M. Brescia35, J. Brinchmann36,37, C. Burigana38,39,40, R. Cabanac41, V. Capobianco8, A. Cappi25,41, C.S. Carvalho42, M. Castellano9, G. Castignani43, S. Cavuoti35,44,45, A. Cimatti26,46, R. Cledassou47, C. Colodro-Conde24, G. Congedo48, C.J. Conselice49, L. Conversi50,51, Y. Copin52, L. Corcione8, A. Costille53, J. Coupon54, H.M. Courtois55, M. Cropper56, S. de la Torre57, D. Di Ferdinando38, F. Dubath22, C. Duncan27, X. Dupac51, S. Dusini58, S. Farrens10, P.G. Ferreira57, I. Ferrero59, F. Finelli35,38, S. Fotopoulou69, M. Frailis22, E. Franceschini25, S. Galeotta22, B. Garilli61, W. Gillard62, B. Gillis48, C. Giocoli25,27, G. Gozaliasl63, J. Graciá-Carpio30, F. Grupp59,30, L. Guzzo64,65, W. Hoggles66, F. Hormuth67, K. Jahnke68, E. Keihanen63, S. Kermiche62, A. Kiessling66, C.C. Kirkpatrick69, M. Kunz70, H. Kurki-Suonio69, S. Ligori8, P.B. Lilje59, I. Lloro71, D. Maino61,64,65, E. Maiorano25, O. Mansutti22, O. Marggraf72, N. Martinet53, F. Marulli25,26,27, R. Massey73, S. Maurogordato41, E. Medinaceli18, S. Mei74, M. Meneghetti35,38,75, R.Benton Metcalf26,76, G. Meylan77, M. Moresco25,26, B. Morin10, L. Moscardini25,26,27, E. Munari22, R. Nakajima72, C. Neissner77, R.C. Nichol78, S. Niemi79, J. Nightingale80, C. Padilla77, S. Paltani32, F. Pasian22, L. Patrizii27, K. Pedersen81, W.J. Percival82,83,84, V. Pettorino10, S. Pisani10, G. Polenta65, M. Poncet17, L. Popa86, D. Potter87, L. Pozzetti25, F. Raison20, A. Renz58,88, J. Rhodes66, G. Riccio35, E. Romelli22, M. Roncarelli25,26, E. Rossetti26, R. Saglia29,30, A.G. Sánchez30, D. Saponas89, R. Scaramella90, P. Schneider72, V. Scottet91, A. Secroun62, G. Seidel68, S. Serrano2,3, C. Sirignano58,88, G. Sirri27, L. Stanco38, F. Sureau10, A.N. Taylor48, M. Tenti27, I. Tereno42,55, R. Teyssier87, R. Toledo-Moreo92, A. Tramacere32, E.A. Valentijn93, L. Valenzano25,27, J. Valiviita69,94, T. Vassallo29, M. Vie19,20,21,22, Y. Wang55, N. Welikala48, L. Whitaker49,96, A. Zacchei12, G. Zamorani25, J. Zoubian62, E. Zucca25

(Affiliations can be found after the references)

ABSTRACT

Photometric redshifts (photo-zs) are one of the main ingredients in the analysis of cosmological probes. Their accuracy particularly affects the results of the analyses of galaxy clustering with photometrically-selected galaxies (GCph) and weak lensing. In the next decade, space missions like Euclid will collect precise and accurate photometric measurements for millions of galaxies. These data should be complemented with upcoming ground-based observations to derive precise and accurate photo-zs. In this article we explore how the tomographic redshift binning and depth of ground-based observations will affect the cosmological constraints expected from the Euclid mission. We focus on GCph and extend the study to include galaxy-galaxy lensing (GGL). We add a layer of complexity to the analysis by simulating several realistic photo-z distributions based on the Euclid Consortium Flagship simulation and using a machine learning photo-z algorithm. We then use the Fisher matrix formalism presented in Euclid Collaboration: Blanchard et al. (2020) together with these galaxy samples to study the cosmological constraining power as a function of redshift binning, survey depth, and photo-z accuracy. We find that bins with equal width in redshift provide a higher Figure of Merit (FoM) than those with equal number of redshift bins from 10 to 13 improves the FoM by 35% and 15% for GCph and its combination with GGL, respectively. For GCph, an increase of the survey depth provides a higher FoM. However, when we include faint galaxies beyond the limit of the spectroscopic training sample, the resulting FoM decreases because of the spurious photo-zs. When combining GCph and GGL, the number density of the sample, which is set by the survey depth, is the main factor driving the variations in the FoM. Adding galaxies at faint magnitudes and high redshift increases the FoM even when they are beyond the spectroscopic limit, since the number density increase compensates the photo-z degradation in this case. We conclude that there is more information that can be extracted beyond the nominal 10 tomographic redshift bins of Euclid and that we should be cautious when adding faint galaxies into our sample, since they can degrade the cosmological constraints.

1. Introduction

The goal of Stage-IV dark energy surveys (Albrecht et al. 2006), such as Euclid (Laureijs et al. 2011) and the Vera C. Rubin Observatory Legacy Survey of Space and Time (Rubin-LSST: LSST Science Collaboration: Abell, P.A. et al. 2009) is to measure both the expansion rate of the Universe and the growth of structures up to redshift z ∼ 2 and beyond. These surveys will allow us to constrain a large variety of cosmological models using
cosmological probes like weak gravitational lensing (WL) and galaxy clustering. Stage-IV surveys can be classified into spectroscopic and photometric surveys, depending on whether the redshift of the observed objects is estimated with spectroscopy or using photometric techniques. The latter can provide measurements for many more objects than the former but at the expense of a degraded precision on the redshift estimates, given that photometric surveys observe through multi-band filters instead of observing the full spectral energy distribution that requires more observational time. Because of this, galaxy clustering analyses are usually performed with data coming from spectroscopic surveys, while the data obtained from photometric surveys are generally used for WL analyses. However, given the current (and future) precision of our measurements, the signal we can extract from galaxy clustering analyses using photometric surveys is far from being negligible (see e.g. [Abbott et al. 2018; van Uitert et al. 2018] Euclid Collaboration: Blanchard et al. 2020; Tatusaus et al. 2020). Therefore, upcoming surveys can increase their constraining power if they optimize their photometric samples to include galaxy clustering studies in addition to WL analyses. The main aim of this work is to perform such an optimization study for the Euclid photometric sample.

The Euclid satellite will observe over a billion galaxies through an optical and three near-infrared broad bands. Given the specifications of the satellite, the combination of Euclid and ground-based surveys can enrich the science exploitation of both. On one hand, the WL analysis of Euclid data requires an accurate knowledge of the redshift distributions of the samples used for the analysis. Euclid photometric data alone cannot reach the necessary photometric redshift (photo-$z$) performance and additional ground-based data are required. On the other hand, Euclid will provide additional information to ground-based surveys such as very precise shape measurements – thanks to the high spatial resolution achieved being in space and avoiding atmospheric distortions – and near-infrared spectroscopy. Euclid’s data will help ground-based surveys improve their deblending of faint objects and improve their photo-$z$ estimates, which will definitely boost their scientific outcome. Surveys where these synergies can be established include the Panoramic Survey Telescope and Rapid Response System (PanSTARRS; Chambers et al. 2016), the Canada-France Imaging Survey (CFIS; Ibata et al. 2017), the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP;ARTHURA et al. 2017), the Javalambre-Euclid Deep Imaging Survey (JEDIS), the Dark Energy Survey (DES, Dark Energy Survey Collaboration 2005) or Rubin-LST (Ivezic et al. 2019). The latter is a Stage IV experiment with a strong complementarity with Euclid since it greatly overlaps in area, covers two Euclid Deep Fields and reaches a faint photometric depth that will lead to better photo-$z$ estimation (Rhodes et al. 2017; Capak et al. 2019). In this article we consider the addition of ground-based optical photometry to Euclid in order to assess the optimal photometric sample for galaxy clustering and galaxy-galaxy lensing (GGL) analyses.

The optimization of the sample of photometrically-selected galaxies for galaxy clustering analyses has been already studied in the literature. In [Tanoglidis et al. 2019], the authors focus their analysis on galaxy clustering for the first three years of DES data. Also for DES but including galaxy-galaxy lensing, [Porredon et al. 2021] studies lens galaxy sample selections based on magnitude cuts as a function of photo-$z$, balancing density and photo-$z$ accuracy to optimise cosmological constrains in the wCDM space. Another example is the recent analysis of [Eifler et al. 2020] on the Nancy Grace Roman Space Telescope (Spergel et al. 2015) High Latitude Survey (HLS), where the authors simulate and explore multi-cosmological probes strategies on dark energy and modified gravity to study observational systematics, such as photo-$z$. These studies show the importance of optimizing the galaxy sample for galaxy clustering analysis.

We aim to perform a similar optimization for the Euclid mission. Note that there have also been several studies optimizing the spectroscopic sample for galaxy clustering analysis with Euclid (Samushia et al. 2011; Wang et al. 2010).

We want to optimize the Euclid sample of galaxies detected with photometric techniques by performing realistic forecasts of its cosmological performance and observing the improvement on the cosmological constraining power of different galaxy samples. When performing galaxy clustering analyses with a photometric sample there are several effects that need to be taken into account such as galaxy bias, photo-$z$ uncertainties or shot noise, among other effects. Here, we try to follow the procedures one would perform in a real data analysis when selecting the samples for the analysis. For that purpose, we use the Euclid Flagship simulation (Euclid Collaboration, in prep; Potter et al. 2017). For a given expected limit of the photometric depth, we select the galaxies included within that magnitude limit and use a machine learning photo-$z$ method to study the optimal way to split the catalogue into subsamples for the analysis. We generate realistic redshift distributions, $n(z)$, for the chosen subsamples and estimate their galaxy bias, $b(z)$. We study the constraining power of these samples when we modify the number and width of the tomographic bins, and when we reduce the sample size by performing a series of cuts in magnitude.

The article is organized as follows. We present Euclid and ground-based surveys in Sect. 2 and Sect. 3. In Sect. 4 we introduce the Flagship simulation and describe how we create photometric samples with different selection criteria. We define the set of galaxy samples that will be used throughout the article, and explain how we estimate the photometric redshifts. In Sect. 5 we detail the forecast formalism and we describe the cosmological model in Sect. 6. In Sect. 7 we present the results of the optimization when changing the number and type of tomographic bins, and we study the dependency of the cosmological constraints on photo-$z$ quality and sample size. Finally, we present our conclusions in Sect. 8.

2. The Euclid survey

Euclid is an European Space Agency (ESA) M-class space mission due for launch in 2022. In the wide survey, it will cover over 15 000 deg$^2$ of the extra-galactic sky with the main aim of measuring the geometry of the Universe and the growth of structures up to redshift $z \sim 2$ and beyond. Euclid will have two instruments on-board: a near-infrared spectro-photometer (Costille et al. 2018), and an imager at visible wavelengths (Cupper et al. 2018). The imager of Euclid, called VIS, will observe galaxies through an optical broad band, $m_{\text{vis}}$, covering a wavelength range between 540 and 900 nm, with a magnitude depth of 24.5 at 10σ for extended sources. The spectro-photometric instrument, called NISP, has three near-infrared bands, $YJH$, covering a wavelength range between 920 and 2000 nm (Racca et al. 2016; 2018). The nominal survey exposure is expected to reach a magnitude depth of 24 at 5σ for point sources. If we convert this depth to 10σ level detections for extended sources we obtain

3 https://panstarrs.stsci.edu
4 http://www.cfht.hawaii.edu/Science/CFIS/
5 https://hsc.mtk.nao.ac.jp/ssp/
6 https://www.darkenergysurvey.org
a magnitude depth of about 23, which is the value we consider in Table 1. The deep survey will cover 40 deg$^2$ divided in three different fields: the Euclid Deep Field North and the Euclid Deep Field Fornax of 10 deg$^2$ each, and the Euclid Deep Field South of 20 deg$^2$ (Euclid Collaboration, in prep.). In these fields the magnitude depth will be two magnitudes deeper than in the wide survey. With its two instruments, Euclid will perform both a spectroscopic and a photometric galaxy survey that will allow us to determine cosmological parameters using its three main cosmological probes: galaxy clustering with the spectroscopic sample (GC$_*$), galaxy clustering with the photometric sample (GC$_{ph}$), and WL. We will study how the selection of the galaxy sample that enters into the analysis can be optimised to provide the tightest cosmological constraints focusing on the GC$_{ph}$ analysis and its cross-correlation with WL – also called GGL.

3. Ground-based surveys
The single broad band VIS of Euclid cannot sample the spectral energy distribution in the optical range. Euclid will require complementary observations in the optical from ground-based surveys to provide the photometry to estimate accurate photometric redshifts and achieve the scientific goals of Euclid. Several ground-based surveys will be needed to cover all the observed area of Euclid, as Euclid covers both celestial hemispheres and those cannot be reached from a single observatory on Earth. The ground-based complementary data will not cover uniformly the Euclid footprint. It is very likely that there will be at least three distinct areas in terms of photometric data available. The Southern hemisphere is expected to be covered with Rubin-LSST data, while the Northern hemisphere will be covered with a combination of surveys such as CFIS, PanSTARRS, JEDIS and HSC-SPP. In addition, some area North of the equator may also be covered by Rubin-LSST at a shallower depth than in the Southern hemisphere. In this work we include simulated ground-based photometry that try to encompass the range of possible ground-based depths that the Euclid analysis will have from the deepest Rubin-LSST data to the shallower data from other surveys.

Rubin-LSST is expected to start operations in 2022 and over 10 years it will observe over 20 000 deg$^2$ in the Southern hemisphere with 6 optical bands, ugrizy, covering a wavelength range from 320 to 1050 nm. The idealized final magnitude depth for coadded images for 5σ point sources are 26.1, 27.4, 27.5, 26.8, 26.1, 24.9, for ugrizy, respectively, based on the Rubin-LSST design specifications [Ivezic et al.2019]. Among other scientific themes, Rubin-LSST has been designed to study dark matter and dark energy using WL, GC$_{ph}$, and supernovae as cosmological probes. The Rubin-LSST survey will provide the best photometry for Euclid-detected galaxies at the time that Euclid data become available.

Another suitable ground-based candidate to cover the optical and near-infrared range in the Southern sky is the DES photometric survey. DES completed observations in 2019 after a 6-years program. It covered 5000 deg$^2$ around the Southern Galactic cap through 5 broad band filters, grizy, with wavelength ranging from 400 to 1065 nm, and redshift up to 1.4 [Dark Energy Survey Collaboration: Abbott et al.2016]. The median coadded magnitude limit depths for 10σ and 20σ diameter aperture are 24.3, 24.0, 23.3, 22.6, for grizy, respectively. These depths correspond to the published values of the first three years of observations (Sevilla-Noarbe et al.2020).

4. Generating realistic photometric galaxy samples
The cosmological constraining power of Euclid will depend on the external data available as it will dictate the photo-z performance of the samples to be studied. In order to study the impact of the available photometry, we create six samples selected with different photometric depths. For each sample, we compute the photo-z estimates using machine learning techniques taking into account the expected spectroscopic redshift distribution of the training sample. We use these photo-z estimates to split each sample into tomographic bins for which we can compute their photo-z distributions and galaxy bias from the simulation. These $n(z)$ and $b(z)$ are then used to forecast the cosmological performance. In this section we provide a detailed description of how we obtain the realistic photo-z estimates of the Euclid galaxies that are later used in the forecast. We first present the cosmological simulation used to extract the photometry and the galaxy distributions. We then explain how we generate realisations of the photometry for the simulated galaxies taking into account the expected depth of the Euclid and ground-based data. We finally present the method used to estimate the photo-z.

4.1. The Flagship simulation
We consider the Flagship galaxy mock catalogue of the Euclid Consortium (Euclid Collaboration, in preparation) to create the different samples. The catalogue uses the Flagship N-body dark matter simulation [Potter et al.2017]. Dark matter halos are identified using ROCKSTAR [Behroozi et al.2013] and are retained down to a mass of $2.4 \times 10^{10} h^{-1} M_{\odot}$, which corresponds to ten particles. Galaxies are assigned to dark matter halos using Halo Abundance Matching (HAM) and Halo Occupation Distribution (HOD) techniques. The cosmological model assumed in the simulation is a flat ΛCDM model with fiducial values Ω_m = 0.319, Ω_b = 0.049, Ω_Λ = 0.681, σ_8 = 0.83, n_s = 0.96, h = 0.67. The N-body simulation ran in a 3.78 h$^{-1}$ Gpc box with particle mass $m_p = 2.398 \times 10^8 h^{-1} M_{\odot}$. The galaxy mock generated has been calibrated using local observational constraints, such as the luminosity function from [Blanton et al.2003] and [Blanton et al.2005a] for the faintest galaxies, the galaxy clustering measurements as a function of luminosity and colour from [Zehavi et al.2011], and the colour-magnitude diagram as observed in the New York University Value Added Galaxy Catalog [Blanton et al.2005b]. The catalogue contains about 3.4 billion galaxies over 5000 deg$^2$ and extends up to redshift $z \approx 2.3$.

For this study we select an area of 402 deg$^2$, which corresponds to galaxies within the range of right ascension $15^\circ < \alpha < 75^\circ$ and declination $62^\circ < \delta < 90^\circ$. All the photometric galaxy distributions obtained in this patch are extrapolated to the 15 000 deg$^2$ of sky that Euclid is expected to observe. Note that the selected area is large enough to minimize the impact of sample variance, but small enough to allow for the production of several galaxy samples in a reasonable amount of time. After the photometric uncertainty is added to the photometry of each galaxy, we perform a magnitude cut in $m_{\text{vir}} < 25$ that leads to a number density of about 41.5 galaxies per arcmin$^2$.

4.2. Photometric depth
Each galaxy observation will lead to a measured value of its magnitude and its associated error. The magnitude depth is usually given as the magnitude at which the median relative error has a particular value. In galaxy surveys it is customary to express the depth at a signal-to-noise of 10 for extended objects,
that is, when the value of the noise is one tenth of its signal. As explained in detail below, we generate realisations of the photometric errors for a given survey taking into account its magnitude depth and scaling the values of the errors at other magnitudes assuming background limited observations, that is, that the background signal dominates the contribution to the error.

We simulate four different photometric survey depths. Table 1 shows their magnitude limits. The first column corresponds to a combination of Euclid and ground-based photometric depth expected to be achieved in the Southern hemisphere. We label this case as optimistic and it is the deepest case we will study. The magnitude limits for the optical bands are for extended sources at 10σ, similar to those expected from Rubin-LSST (LSST Science Collaboration et al. 2009). The values for Euclid correspond to a 10σ detection level for extended sources. In addition to the magnitude limits expected in the South, we also want to investigate how the cosmological constraints degrade as the depth is reduced. We investigate three other cases. First, a case were the depth in optical bands are reduced by a factor of two in signal-to-noise ratio. The second column shows the magnitude limits for this case where the optical bands are reduced by 0.75 magnitude. This column represents a possible case where the Rubin-LSST data have a reduced depth in areas outside its main footprint. We also study a case were the limiting fluxes of Euclid are brightened by 0.75 magnitudes, shown in the third column. Lastly, we explore a case where the ground-based data is degraded by a factor of five in signal-to-noise but the Euclid space data remains at their nominal depth values. This broadly represents the depth that can be achieved from other ground-based data in the Northern hemisphere.

For each survey case, we generate a galaxy catalogue drawn from the Flagship simulation. We assign observed magnitudes and errors with the following procedure. First, we compute the expected error for each galaxy, taking into account its magnitude in the Flagship catalogue and the magnitude limit of the survey as given in Table 1. We assume that the observations are sky limited (the noise is dominated by the shot noise of the sky), and therefore we scale the ratio of the signal to noise between two galaxies i and j as the ratio of their fluxes

\[
\frac{S}{N}_i = \left( \frac{S}{N}_j \right) \frac{f_i}{N_j}.
\]

where \(f_i\) is the observed flux of galaxy i detected at signal-to-noise ratio \((S/N)_i\). The magnitude (flux) limits in Table 1 give us the fluxes corresponding to a signal-to-noise ratio of 10, \(f_{10\sigma}\), and therefore we can compute the expected signal-to-noise at which a galaxy of a given magnitude is detected as

\[
\frac{S}{N}_i = 10 \frac{f_i}{f_{10\sigma}}.
\]

Using the definition of signal-to-noise, \((S/N)_i = f_i/\Delta f_i\), we can compute the expected flux error for each galaxy as

\[
\Delta f_i = \frac{f_{10\sigma}}{10}.
\]

The fluxes in the Flagship catalogue correspond to the real fluxes of each galaxy. Whenever we observe these galaxies in a given survey, we detect a realization of the real flux. For our study, we generate realisations of the observed fluxes \(f'_i\) for each survey as

\[
f'_i = f_i + N (\mu = 0, \sigma = f_{10\sigma}/10),
\]

where \(N\) is a random number from a normal distribution. We then assign errors to the resulting fluxes according to Eq. (3). Finally, the new fluxes and their assigned errors are converted into magnitudes and their respective magnitude errors.

4.3. Samples

We estimate the expected cosmological constraints using the galaxy clustering analysis of tomographic bins defined with photo-z (see Sect. 5). The magnitude limit of a given sample will give us the galaxies that form the overall sample, while the photo-z algorithm will split that sample into tomographic bins and will provide an estimate of the redshift distributions within these tomographic bins. We can better understand the uncertainties in the method using simulations where we know the true redshift distributions. So far, we have defined four different samples based on the available photometry representing the four cases defined in Table 1. The photo-z performance depends on the photometric depth and the spectroscopic data available to train the method. Now, we will generate study cases depending on the spectroscopic data available to train the photo-z technique we use. We will use three different spectroscopic samples with different completeness profiles as a function of magnitude. First, we consider an idealised case where the spectroscopic training sample is a random subsample of the whole sample and thus it is fully representative (blue line in Fig. 1). We consider a second case where the spectroscopic sample completeness as a function of magnitude follows the expectations from spectrographs on 8-m class telescopes (Newman et al. 2015). This case is shown in black in Fig. 1. This is intended to mimic the spectroscopic incompleteness as a function of magnitude of surveys like zCOSMOS (Lilly et al. 2007), VVDS (Le Fèvre et al. 2013) and DEEP2 (Newman et al. 2013) at least in its shape, although maybe optimistic in its normalisation. Finally, we consider a last case where the spectroscopic completeness is similar to the current available spectroscopic surveys, as those listed in Gschwend et al. (2018). We compute how the completeness in spectroscopic data as a function of redshift translates into completeness in \(m_{\text{vn}}\) (orange line in Fig. 1). These cases are explained in more detail later in this section. It is worth mentioning that we only consider galaxies and not stars in the samples under study. With the high spatial resolution of Euclid, the contamination in the sample due to stars is expected to be minimal. We have also assumed that the effects of Galactic extinction are corrected in the data reduction pipelines and therefore ignore Galactic extinction. These factors can be include in the future to add another layer of realism to the analysis.

We combine the four cases of photometric limits with the three cases of different spectroscopic data available to train the photo-z techniques to generate six galaxy samples for our study. With these six samples we try to encompass a wide range of scenarios to try to understand how the cosmological constraints vary depending on the sample available. We detail these six cases in the following subsections. Table 2 summarises all the cases we consider. All our samples have galaxies down to a magnitude limit of \(m_{\text{vn}} = 25\). Note that for our shallower survey (column four in Table 1), galaxies near this \(m_{\text{vn}}\) selection limit will have larger errors. It is also important to mention that in all cases we assume the magnitude limit in each band to be isotropic – homogeneous on the sky. This will definitely not be the case for Euclid, since ground-based data will consist on a compilation of different surveys pointing at different regions of the sky, with different depths and systematic uncertainties. For instance, Rubin-LSST focuses on the Southern hemisphere, while Euclid will...
Table 1. Limiting coadded depth magnitudes for extended sources at 10σ used in each sample.

| Band | Optimistic | Ground based | All | Ground based |
|------|------------|---------------|-----|--------------|
|      |            | degraded −0.75 |     | degraded −1.75 |
| u    | 25.55      | 24.8          | 24.8| 23.8         |
| g    | 26.75      | 26.0          | 26.0| 25.0         |
| r    | 26.95      | 26.2          | 26.2| 25.2         |
| i    | 26.25      | 25.5          | 25.5| 24.5         |
| y    | 24.15      | 23.4          | 23.4| 22.4         |

Table 2. Cases under study. The photometric limit value corresponds to the column number of Table 1 whose magnitude limit depths are used to define each photometric sample. The spectroscopic training sample used to determine the photo-z can be a representative subsample, a sample with a completeness drop in m\text{vis} or a sample with an inhomogeneous spectroscopic redshift distribution as shown in Fig. 1.

| Case name                  | Photometric limit | Spectroscopic training |
|----------------------------|-------------------|------------------------|
| Case 1: Optimistic         | 1                 | Subsample              |
| Case 2: Fiducial           | 1                 | Compl. drop            |
| Case 3: Mid-depth          | 2                 | Compl. drop            |
| Case 4: Mid-depth Euclid   | 3                 | Compl. drop            |
| Case 5: Shallow depth      | 4                 | Compl. drop            |
| Case 6: Inhomogeneous spec | 4                 | Inho. spec-z           |

also observe the Northern one. A more detailed analysis taking into account the depth anisotropy of the ground-based data is left for future work. A possible approach would be to generate several sets of ground-based photometry according to the specific limitations of each ground-based instrument and region of the sky covered, in order to reproduce the expected anisotropy of the photometry. Then we would mix the different sets of ground-based photometry and add them to the Euclid photometry in order to determine the photometric redshifts and redo the optimization analysis as performed in this article.

4.3.1. Case 1: Optimistic

This case uses the deepest magnitude limit and a highly idealised spectroscopic training sample. The sample has magnitudes and errors generated as described in Sect. 4.2 with the Euclid and ground-based photometric depth limits shown in the first column of Table 1. The photo-z are estimated using a training set that is a complete and representative subsample in both redshift and magnitude of the whole sample.

4.3.2. Case 2: Fiducial

We take this case to be our fiducial sample. We use the deepest photometry as in the optimistic case 1 but the photo-z estimation now makes use of a training sample that has a completeness drop at faint magnitudes that resembles the incompleteness of spectroscopic surveys carried out with spectrographs in 8m-class telescopes. We show the completeness drop in the spectroscopic training sample in Fig. 1 (black line). While the completeness as a function of magnitude intends to be realistic of current spectroscopic capabilities, we make the simplifying assumption that this incompleteness does not depend on any galaxy property except its magnitude and therefore we randomly subsample the whole distribution only taking into account the probability of being selected based on the galaxy magnitude.

4.3.3. Case 3: Ground-based mid-depth photometry

We define another sample trained with the same spectroscopic training sample completeness as in the fiducial case but with shallower ground-based magnitude limits in the photometry. The
ground-based magnitude limit is a factor of two shallower in signal-to-noise ratio than in cases 1–2. This corresponds to the second column in Table 1. This case is intended to represent areas on the sky between the celestial equator and low Northern declinations where Rubin-LSST data at shallower depth may be available.

4.3.4. Case 4: Euclid mid-depth photometry

To explore the possibilities of available photometry, especially the importance of deep near-infrared photometry, we define a case in which both the Euclid and ground-based photometric depth is reduced by 0.75 magnitudes (third column in Table 1). The spectroscopic training sample completeness is the same as in cases 2 and 3.

4.3.5. Case 5: Ground-based shallow depth photometry

The complementary ground-based photometry expected to be available in the Northern hemisphere is shallower than the magnitude limits used in our previous cases. We define a sample to roughly represent and cover this option by considering a ground-based flux limit 1.75 magnitudes brighter compared to our optimistic case (fourth column in Table 1). To compute the photometric redshift training set with the same completeness in m$_{VIS}$ as in cases 2, 3, and 4.

4.3.6. Case 6: Inhomogeneous spectroscopic sample

In this last sample, we want to study the case in which the spectroscopic training sample is very heterogeneous and composed of the combination of many surveys targeting galaxies with different selection criteria and with different spectroscopic facilities. We choose a spectroscopic training set that tries to model the n(z) of current available spectroscopic data coming from surveys as those listed in Gschwend et al. (2018). Given that some of these surveys have different colour selection cuts and magnitude limit depths, the combined redshift distribution is not homogeneous presenting peaks and troughs, which cause strong biases in the photo-z estimation due to over and under-represented galaxies at different redshift ranges (see e.g. Zhou et al., 2020). We want to remark that we only try to reproduce the n(z) of the overall spectroscopic sample. We do not try to gather this spectroscopic sample applying the same selection criteria of the different surveys used. We consider that this is not necessary for our purposes as we are only interested in the overall trend induced by using an inhomogeneous spectroscopic training sample. We create the spectroscopic training sample by randomly selecting galaxies based on their redshift to reproduce the overall targeted redshift distribution. Given that the Flagship simulation area we are using (see Sect. 4.1) is smaller than the surveys sampling the nearby Universe, our simulated spectroscopic training does not exactly reproduce our overall redshift distribution at low redshifts. The resulting completeness as a function of the m$_{VIS}$ of this spectroscopic sample can be seen in Fig. 1 (orange line). The modeled n(z) is shown in Fig. 2 (orange line). With this case, which intends to represent the currently available data, we can draw a lower bound on the photo-$z$ accuracy that can be expected for Euclid. In this case, we use the same photometric magnitude limits as in case 5.

The realism of our training samples is limited in the sense that we only try to reproduce the completeness in m$_{VIS}$ or the shape of the n(z) distribution. We do not take into account any dependence of the training samples on other characteristics such as galaxy type or the presence of emission lines, which would have an impact on the determination of the photo-$z$.

4.4. Photometric redshifts

The cosmological tomographic analysis of a photometric survey divides the whole sample into redshift bins selected with a photo-

Fig. 2. True redshift distributions of the training samples used to run DNF in all 6 cases. The training samples include magnitudes brighter than m$_{VIS}$ = 24.5. The true redshift comes from the Flagship simulation. The four training samples with almost identical true redshift distributions have the same completeness drop in m$_{VIS}$ and only differ in the photometric quality.

Fig. 3. Top: z$_{\text{mean}}$ photometric redshift distributions obtained with DNF for the 6 photometric samples up to m$_{VIS}$ = 25. The z$_{\text{mean}}$ photo-$z$ estimate returned by DNF is the value resulting from the mean of the nearest neighbours redshifts. Lower: Photometric redshift distributions obtained with DNF for the z$_{\text{rr}}$ statistic, which for each galaxy is a one-point sampling of the redshift probability distribution estimated from the nearest neighbour (see text for details).
z technique. In our study, we want to follow as close as possible the methodological steps that one would carry out in real surveys. For that purpose, we compute the photo-zs of all our study cases described in Table 2. We use the Directional Neighborhood Fitting (DNF; De Vicente et al. 2016) training-based algorithm to estimate realistic photo-z estimates of our simulated galaxies. The exact choice of the machine learning training set method is not important for our analysis as most methods of this type perform similarly to the precision levels we are interested in (see e.g. Euclid Collaboration: Desprez et al. 2020; Sánchez et al. 2014).

DNF estimates the photo-z of a galaxy based on its closeness in observable space to a set of training galaxies whose redshifts are known. The main feature of DNF is that the metric that defines the distance or closeness between objects is given by a directional neighbourhood metric, which is the product of a Euclidean and an angular neighbourhood metrics. This metric ensures that neighbouring objects are close in colour and magnitude space. The algorithm fits a linear adjustment, a hyperplane, to the directional neighbourhood of a galaxy to get an estimation of the photo-z. This photo-z estimate is called $z_{\text{mean}}$, which is the average of the redshifts from the neighbourhood. The residual of the fit is considered as the estimation of the photo-z error. In addition, DNF also produces another photometric redshift estimate, $z_{\text{mc}}$ that is a Monte Carlo draw from the nearest neighbour in the DNF metric for each object. Therefore, it can be considered as a one-point sampling of the photo-z probability density distribution. As such, it is not a good individual photo-z estimate of the object, but when all the estimates in a galaxy sample are stacked it can recover the overall probability density distribution of the sample (Rau et al. 2017). When working with tomographic bins, we will classify the galaxies into different bins using their $z_{\text{mean}}$ and we will obtain the photometric distribution, $n(z)$, within each bin by stacking their $z_{\text{mc}}$. This is an approach used by DES in analysing their First Year Data results (e.g., Hoyle et al. 2018, Crocce et al. 2019, Camacho et al. 2019) providing redshift distributions that were validated with other independent assessment methods. Therefore, we define the $n(z)$ by stacking the $z_{\text{mc}}$ estimator instead of the true redshift of the simulation to make the photo-z distribution close to what would be obtained in a real data analysis with the assurance that the method has been validated.
We select a patch of sky of 3.35 deg$^2$ to create the samples to train DNF. These training samples have the magnitudes and errors computed with the same magnitude limits as the sample whose photo-z we want to compute (see Table 1). We generate three types of spectroscopic training samples. For all of them we limit the spectroscopic training sample to galaxies brighter than $m_{\text{vis}} = 24.5$ as there are few objects whose redshift has been reliably determined beyond that magnitude limit. The spectroscopic training samples are described in Section 4.3.

The true redshift distributions of the spectroscopic training set used to train DNF for each of the sample cases considered here are shown in Fig. 2. In blue, we present the redshift distribution of case 1 with the first spectroscopic training sample that it is fully complete as a function of magnitude. We show in black the resulting $N(z)$ of case 2. Cases 3–5 (olive, red and orange colours in Figs. 2 and 3) have the same training sample completeness as a function of magnitude. The drop in completeness at faint magnitudes translates into a decrease of objects at high redshift. Last, we present the resulting redshift distribution with the third spectroscopic training set in orange. Gathering multiple selection criteria from different surveys leads to an inhomogeneous redshift distribution for the spectroscopic training set. In Fig. 3, we show the overall photo-z distributions of $z_{\text{true}}$ (top panel) and $z_{\text{phot}}$ (bottom panel) values obtained for the full sample for each of the six cases. We see how an inhomogeneous $N(z)$ in the training sample leads to an inhomogeneous distribution of the photo-z.

In Fig. 4, we show the photo-z obtained with DNF as a function of true redshift for the six samples up to $m_{\text{vis}} = 24.5$. This figure gives us an indication of how the photo-z scatter decreases with deeper photometry. Photometric samples go up to $m_{\text{vis}} = 25$. However, we cut the spectroscopic training sample at $m_{\text{vis}} = 24.5$ to be more realistic. The lack of objects between 24.5 and 25.0 in the training sample forces the algorithm to extrapolate beyond that magnitude, and thus noisier photometric redshifts are obtained.

In Table 3, we show the values obtained for these metrics for each photometric sample.

5. Building forecasts for Euclid

So far, we have seen how the photometric depth and the spectroscopic training sample determine the overall redshift distributions of the resulting samples. We have selected six cases to cover a range of possible scenarios that we may encounter in the analysis of Euclid data complemented with ground-based surveys. Once the galaxy distributions for the photometric cases under study have been obtained, we want to propagate the photo-z accuracy in determining tomographic subsamples to the final constraints on the cosmological parameters in order to understand how to optimize the photometric sample for galaxy clustering analyses.

We follow the forecasting prescription presented in Euclid Collaboration: Blanchard et al. (2020) hereafter EC20. We consider the same Fisher matrix formalism and make use of the CosmoSIS code validated for Euclid specifications therein. Our observable is the tomographically binned projected angular power spectrum, $C_{\ell}(\ell)$, where $\ell$ denotes the angular multipole, and $i, j$ stand for pairs of tomographic redshift bins. This formalism is the same for WL, galaxy clustering (with the photometric sample), and GGL, with the only difference being the kernels used in the projection from the power spectrum of matter perturbations to the spherical harmonic-space observable. We focus on the Gaussian, cosmological probe, as well as its combination with GGL. The projection to $C_{\ell}(\ell)$ is performed under the Limber, flat-sky and spatially flat approximations (Kitching et al. 2017; Kilbinger et al. 2017; Taylor et al. 2018). We also ignore redshift-space distortions, magnification, and other relativistic effects (Deshpande et al. 2020). To minimise the impact of neglecting relativistic effects, more relevant at large scales, in our analysis we consider multipole scales from $\ell \geq 10$ to $\ell \leq 750$, which corresponds to the more conservative scenario in EC20.

When considering GGL, its power spectrum contains contributions from galaxy clustering and cosmic shear, but also from intrinsic galaxy alignments (IA). We assume the latter is caused by a change in galaxy ellipticity that is linear in the density field. Note that such modelling is appropriate for large scales (Troxel et al. 2018), like the ones considered in this analysis, but more complex models should be used for the very small scales (see e.g. Blazek et al. 2019; Fortuna et al. 2020). Under this linear assumption we can define the density-intrinsic and intrinsic-intrinsic three-dimensional power spectra, $P_{\delta\delta}$ and $P_{\eta\eta}$, respectively. They can be related to the density power spectrum $P_{\delta\delta}$ with $P_{\delta\delta} = -A(z)P_{\delta\delta}$ and $P_{\eta\eta} = A(z)^2P_{\delta\delta}$. We follow EC20 in parameterising $A$ as

$$A(z) = \frac{\mathcal{A}_{\text{IA}}(\delta_{\text{IA}})D(z)}{C_{\text{IA}}},$$

where $C_{\text{IA}}$ is a normalisation parameter that we set as $C_{\text{IA}} = 0.0134$, $D(z)$ is the growth factor, and $\mathcal{A}_{\text{IA}}$ is a nuisance parameter fixing the amplitude of the IA contribution.

We model the redshift dependence of the IA contribution as

$$\mathcal{F}_{\text{IA}} = (1 + z)^{\beta_{\text{IA}}} \left\{ \frac{(L)(z)}{L_{\text{spec}}(z)} \right\}^{\eta_{\text{IA}}},$$

with $(L)(z)/L_{\text{spec}}(z)$ being the redshift-dependent ratio between the average source luminosity and the characteristic scale of the luminosity function (Hirata et al. 2007; Bridle & King 2007). For a detailed explanation on IA modeling see Samuroff et al. (2019).

We use the same ratio of luminosities for every galaxy sample. However, this ratio should in principle depend on the specific galaxy population. Since we select galaxies according to a $m_{\text{vis}}$ cut and not according to a particular galaxy type, we expect that the luminosity ratio does not change significantly between galaxy samples, and therefore use the same ratio for simplicity. We set the fiducial values for the intrinsic alignments nuisance parameters to

$$\left\{ \mathcal{A}_{\text{IA}}, \eta_{\text{IA}}, \beta_{\text{IA}} \right\} = (1.72, -0.41, 2.17),$$

https://bitbucket.org/joezuntz/cosmosis/wiki/Home
in agreement with the recent fit to the IA contribution in the Horizon-AGN simulation (Chisari et al. 2015), although the amplitude $A_{ IA}$ might be smaller in practice (Fortuna et al. 2020).

When considering GC$_{ph}$ and GGL one of the primary sources of uncertainty is the relation between the galaxy distribution and the underlying total matter distribution, i.e. the galaxy bias (Kaiser 1987). We consider a linear galaxy bias relating the galaxy density fluctuation to the matter density fluctuation with a simple linear relation

$$\delta_g(x, z) = b(z)\delta_m(x, z),$$

where we neglect any possible scale dependence. Note that a linear bias approximation is sufficiently accurate for large scales (Abbott et al. 2018). However, when adding very small scales into the analysis, a more detailed modeling of the galaxy bias is required (see e.g., Sánchez et al. 2016). One of the approaches to this modeling is through perturbation theory, which introduces a nonlinear and nonlocal galaxy bias (Desjacques et al. 2018).

We consider a constant galaxy bias in each tomographic bin. We get their fiducial values by fitting the directly measured bias in Flagship to the function

$$b(z) = \frac{A - B}{1 + z} + C,$$

where $A$, $B$ and $C$ are nuisance parameters. We select five subsamples with $m_{\text{lim}}$ limiting magnitudes: 25, 24.5, 24, 23.5, and 23 from the Flagship galaxy sample. We compute the bias values as a function of redshift for each of these magnitude-limited subsamples using directly the true redshift of Flagship at redshifts 0.5, 1, 1.5 and 2. As an approximation, we use the same galaxy bias for each of the six photometric samples and change the fiducial according to the magnitude limit cut. The obtained bias and fitted functions are shown in the left panel of Fig. 5.

To fit the bias-redshift relation we choose to use all galaxy bias values computed with the Flagship simulation, although values at $z = 2$ were less reliable. The value of the bias at $z = 1.5$ falls outside the bias-redshift fit for the $m_{\text{lim}} < 23$ sample. However, we recomputed the bias fit neglecting the value at $z = 2$ and including the value at $z = 1.5$, but no significant changes were appreciated, therefore we keep the bias computed using the fits shown in Fig. 5.

To validate the bias obtained with Flagship, we compare our bias values to the ones obtained from the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) data release 1 (DR1) by Nicola et al. (2020) N20 hereafter. The HSC survey has comparable survey depth and uses similar ground-based bands to the ones considered in this work. N20 fit galaxy bias on magnitude-limited galaxy samples down to $i < 24.5$. We compare their values to ours in the right panel of Fig. 5. We extrapolate their bias to $i < 25$ for our faintest magnitude bin. Strictly speaking, we are comparing $i$-band magnitude-selected samples from N20 to our $m_{\text{lim}}$-band magnitude-selected samples. We have checked in Flagship that the bias values for both $i$-band and $m_{\text{lim}}$-band selected samples cut at the same magnitude limit do not change by more than 10% and therefore our comparison is meaningful. N20 assume that bias can be split into two separated terms of redshift and limiting magnitude, and define it as

$$b_{\text{HSC}}(z, m_{\text{lim}}) = \bar{b}(m_{\text{lim}})D'(z),$$

where $\alpha$ is a variable that takes into account the inverse relation between the growth factor and galaxy bias. By fitting $\alpha$ and $\bar{b}(m_{\text{lim}})$ in a multi-step weighted process they find

$$\alpha = -1.30 \pm 0.19,$$

$$\bar{b}(m_{\text{lim}}) = b_1(m_{\text{lim}} - 24) + b_0,$$

where $b_1 = -0.0624 \pm 0.0070$ and $b_0 = 0.8346 \pm 0.161$. For a detailed explanation see Sect. 4.6 in N20. We compute $D(z)$ for our sample and use our $m_{\text{lim}}$ magnitude cuts as $m_{\text{lim}}$ along with their fitted parameters to get a bias to compare. The ratio between the HSC bias, $b_{\text{HSC}}$, and ours, $b(z)$, is shown in the right panel of Fig. 5.

Fig. 5. Left panel: Galaxy bias as a function of redshift. Dots correspond to the measured values in the Flagship simulation for different magnitude cuts and the solid lines are a fit (Eq. [11]). We plot with squares the bias values obtained for $z = 2$ to indicate that at that redshift there are few objects and thus the values are slightly less reliable. At $m_{\text{vis}} < 23$ there were not enough objects at $z = 2$ to compute the bias in Flagship. Right panel: Ratio between the HSC bias, $b_{\text{HSC}}$, from N20 and the Flagship bias for each magnitude-limited sample. To assess the $1\sigma$ uncertainty of $b_{\text{HSC}}$ along the redshift range, we generate a set of Gaussian random numbers for the free parameter $\alpha$, $b_1$, and $b_0$ of $b_{\text{HSC}}$ with their values as mean and their errors as standard deviation. Then we evaluate $b_{\text{HSC}}$ in the redshift range for all the set of free parameters previously generated. We pick the maximum and minimum $b_{\text{HSC}}$ at each redshift. This corresponds to the shaded regions.
panel of Fig. 5. Note that N2O compute their bias up to redshift 1.25 and that we have extrapolated their behaviour to higher redshifts for the comparison at \( z > 1.25 \). The values of the bias in Flagship stay within 1σ of the HSC values, \( bi_{\text{HSC}} \) (shaded area in the right panel of Fig. 5), confirming that the bias values we use are consistent with the HSC observations.

We consider the same redshift distributions for both GC\(_{\text{ph}}\) and GGL. In practice, this is an over-simplification, since these two probes will probably apply different selection criteria when determining their samples. GGL for instance will give some importance to the shape measurements of the galaxies. But for the present Fisher matrix analysis we limit ourselves to use the same sample for both probes.

### 6. Cosmological model

We optimize the photometric sample of Euclid considering the baseline cosmological model presented in [EC20] a spatially flat Universe filled with cold dark matter and dark energy. We approximate the dark energy equation of state parameter with the CPL (Chevallier & Polarski 2001; Linder 2005) parameterisation

\[
w(z) = w_0 + w_a \frac{z}{1 + z}.
\]

Therefore, the cosmological model is fully specified by the dark energy parameters, \( w_0 \) and \( w_a \), the total matter and baryon density today, \( \Omega_m \) and \( \Omega_b \), the dimensionless Hubble constant, \( h \), the spectral index, \( n_s \), and the RMS of matter fluctuations on spheres of 8 h\(^{-1}\)Mpc radius, \( \sigma_8 \). We assume a dynamically evolving, minimally-coupled scalar field, with sound speed equal to the speed of light and vanishing anisotropic stress as dark energy. Therefore, we neglect any dark energy perturbations in our analysis. We also allow the equation of state of dark energy to cross \( w(z) = -1 \) using the [Hu & Sawicki (2007)](Hu & Sawicki (2007)) prescription.

The fiducial values of the cosmological parameters are given by

\[
\{ \Omega_m, \Omega_b, w_0, w_a, h, n_s, \sigma_8 \} = \{0.32, 0.05, -1, 0, 0.67, 0.96, 0.816\}.
\]

Moreover, we fix the sum of neutrino masses to \( \sum m_\nu = 0.06 \text{ eV} \). Note that the linear growth factor depends on both redshift and scale when neutrinos are massive, but we follow [EC20] in neglecting this effect, given the small fiducial value considered. Therefore, we compute the growth factor accounting for massive neutrinos, but neglect any scale dependence. Note that the fiducial values used in this analysis are compatible with the fiducial cosmology of the Flagship simulation presented in Sect. 4.1 except for \( \sigma_8 \). This can be explained by the fact that the Flagship simulation does not account for massive neutrinos and therefore considers a slightly larger value for \( \sigma_8 \). However, since we are only extracting the galaxy bias and the galaxy distributions from Flagship, and we are computing Fisher forecasts, this difference in the fiducial \( \sigma_8 \) value does not have any impact on our results.

We quantify the performance of photometric galaxy samples in constraining cosmological parameters through the metric figure of merit (FoM), as defined in [Albrecht et al. (2006)](Albrecht et al. (2006)) but with the parameterisation defined in [EC20]. Our FoM is proportional to the inverse of the area of the error ellipse in the parameter plane of \( w_0 \) and \( w_a \) defined by the marginalised Fisher submatrix, \( F_{w_0w_a} \),

\[
\text{FoM}_{w_0w_a} = \sqrt{\det(F_{w_0w_a})}.
\]

We will use the FoM defined above throughout this article. The higher the FoM value, the higher the cosmological constraining power.

### 7. Results

In this section we carry out a series of tests to optimize the sample selection for GC\(_{\text{ph}}\) analyses. We want to determine the best number and type of tomographic bins to constrain cosmological parameters. We explore the influence of the accuracy in the photo-z estimation and sample size in providing cosmological constraints. We split the data in tomographic redshift bins in order to have more control in the variations of sample size and photo-z accuracy to better understand their impact in constraining cosmological parameters. We use the FoM defined in Eq. (16) to quantify the constraining power on the cosmological parameters. In addition, we also compute the FoM when combining GC\(_{\text{ph}}\) with GGL, assuming the same photo-z sample, which implies the same photo-z binning and number density. When computing the cosmological constraining power for GC\(_{\text{ph}}\) + GGL, we marginalize over the galaxy bias of each tomographic bin and intrinsic alignment parameters, whereas for GC\(_{\text{ph}}\) alone the galaxy bias parameters are fixed to their fiducial values. The main reason for this choice is that, under the linear galaxy bias approximation, there is a large degeneracy between the galaxy bias and \( \sigma_8 \). In this case, the Gaussianity assumption of the Fisher matrix approach breaks down and its constraints on the cosmological parameters are not reliable. Therefore, we fix the galaxy bias to break this degeneracy when considering GC\(_{\text{ph}}\) alone. Note that when we combine GC\(_{\text{ph}}\) with GGL, the additional information brought by the latter is enough to break such degeneracy and constrain \( \sigma_8 \) and the galaxy bias at the same time.

#### 7.1. Optimizing the type and number of tomographic bins

We bin galaxies into different numbers of redshift bins to study the impact of the number of redshift bins on the cosmological parameter inference. When we define redshift bins, we choose galaxies within the redshift range [0, 2] since the maximum lightcone outputs generated in Flagship are at \( z = 2.3 \) and we prefer to avoid working at the limit of the simulation. We check the effect of using bins with the same redshift width and bins with the same number of objects (equipopulated). We also see the difference when using only GC\(_{\text{ph}}\) or both GC\(_{\text{ph}}\) and GGL probes. This analysis is performed using our fiducial sample (case 2) up to \( m_w < 24.5 \). We compute the FoM for all the cases mentioned and show the results in Fig. 6. The FoM are normalised to ten bins since this is the default number used to compute the forecasts in [EC20].

As seen in Fig. 6 the general tendency of the FoM is to increase with the number of bins. [EC20] used ten tomographic bins as their fiducial value. For bins with equal width in redshift, the FoM increase when moving from ten to thirteen bins is 35.4% and 15.4%, for GC\(_{\text{ph}}\) only and for GC\(_{\text{ph}}\) + GGL, respectively. The FoM improvement we get from going to even more bins does not compensate the increase in computational time needed for the analysis. This is especially true when using both probes, where we notice that the curve flattens while in GC\(_{\text{ph}}\) the FoM continues to increase since the bias is fixed and thus the amount of information that can be extracted is larger than expected in practice. Moreover, our photo-z treatment may start to be too simplistic to realistically deal with too many photometric redshift bins. The FoM saturates with the increasing number of bins.
because it is not possible to extract more information on radial clustering when the width of the bins is smaller than the photo-z precision. At this limit, the uncertainty at which bin a particular galaxy belongs is greatly increased. For GC\textsubscript{ph} + GGL the curves flatten at lower number of bins since systematic effects in the marginalisation of galaxy bias and intrinsic alignment free parameters also affect the cosmological information that can be extracted. Therefore, we choose thirteen to be our fiducial number of bins as a conservative choice.

In addition, we choose bins with equal width in redshift as the optimal way of partitioning the sample since we observe that, overall, for GC\textsubscript{ph} the FoM is larger in this case than in the equipopulated one. For thirteen bins with equal width the FoM is 713 while it is 547 for equipopulated bins, which is an increase of 30\%. For the GC\textsubscript{ph} + GGL combined analysis, the FoM does not appreciably change between the use of bins with equal width and equipopulated ones. At thirteen bins, which is the fiducial choice, the FoM difference of using bins with equal width or equipopulated ones is negligible.

We will use these bin choices to analyze the dependency of cosmological constraints on the photo-z quality and size of the sample. In Fig. 7 we show the redshift distributions of the thirteen bins with equal width for our fiducial case 2 sample.

### 7.2. FoM dependency on photometric redshift quality and number density

Another aspect we want to study is the effect of the trade-off between photo-z accuracy and number density on the constraining power of cosmological parameters. For that purpose, we take the six photometric samples defined in Sect. 4.3 and apply five cuts (25, 24.5, 24, 23.5, 23) in \( m_{\text{vis}} \) to modify the sample size (leading to a number density of about 41, 29, 18, 12 and 9 galaxies per arcmin\(^2\) respectively). Besides reducing the number density of the photometric samples, the cut in \( m_{\text{vis}} \) also affects the photo-z distribution and accuracy of the overall sample. A bright magnitude cut, that eliminates the fainter galaxies, mostly removes galaxies with higher and thus less reliable redshifts. We compute the FoM for all the cases mentioned before and normalise them to the FoM of our fiducial (case 2) sample at \( m_{\text{vis}} < 24.5 \), for both GC\textsubscript{ph} only and GC\textsubscript{ph} + GGL. To help visualise the results, we present the resulting FoM in a grid format in Fig. 8 and the values themselves in Table 4.

Let us first discuss the case of GC\textsubscript{ph} alone. As seen in Fig. 8, in general, the FoM for GC\textsubscript{ph} increases with deeper photometric data, which improves the photo-z performance (increasing along the x-axis in the figure). The FoM also increases with number density, determined by the magnitude limit imposed (increasing along the y-axis). We notice a larger increase in the FoM with sample size in those samples where the photo-z quality is better (e.g., the optimistic, fiducial and mid depth ground-based photometry cases). In these cases, increasing the sample size from a magnitude cut, that eliminates the fainter galaxies, mostly removes galaxies with higher and thus less reliable redshifts. We compute the FoM for all the cases mentioned before and normalise them to the FoM of our fiducial (case 2) sample at \( m_{\text{vis}} < 24.5 \), for both GC\textsubscript{ph} only and GC\textsubscript{ph} + GGL. To help visualise the results, we present the resulting FoM in a grid format in Fig. 8 and the values themselves in Table 4.

The trend of increasing FoM as we take fainter magnitude limit cuts and increase the number density continues as long as the photo-z performance is not degraded. Once we push to faint magnitudes where there are no objects to train the photometric redshift algorithms, their performance degrades and the photometric redshift bins start to be wider. There are many objects that do not belong to the bins and spurious cross-correlations between different bins appear. As a result, the strength of the cosmological signal is diminished and the FoM decreases. This effect can be seen in Fig. 8 for the GC\textsubscript{ph} case (left panel), where we can appreciate a reduction in the FoM when we move from a magnitude-limited sample cut at \( m_{\text{vis}} < 24.5 \) (second row from the top) to a magnitude-limited sample cut at \( m_{\text{vis}} < 25.0 \) (top row). With this change, we are increasing the sample, but with galaxies that cannot be located in redshift as their photo-z cannot be calibrated. As a result, the clustering strength is diluted and some spurious cross-correlation signal appears resulting in a decreased FoM compared to a shallower sample with better photometric redshifts.

---

Footnote:
---

8 Recall that galaxy bias is fixed when considering GC\textsubscript{ph} alone, which provides these large absolute values for the FoMs.
galaxies that cannot be located in redshift. Pushing to faint magnitudes, there is no improvement including terms when performing cosmological inference with GC performance and the number density are important contributing factors to the FoM than the photo-z. We conclude that the GC probe is sensitive to the actual location of their tracer galaxies inside their tomographic bins. Both the photometric redshift performance and the number density are important contributing factors when performing cosmological inference with GC. When pushing to faint magnitudes, there is no improvement including galaxies that cannot be located in redshift.

To illustrate this effect, in Fig. 9 we show the redshift distribution of three tomographic bins for three samples with galaxies down to $m_{\text{VIS}} < 24.5$, 25, and with galaxies only between 24.5 and 25. Galaxies with $m_{\text{VIS}}$ between 24.5 and 25 are mostly outside their tomographic bin increasing the width of the distribution and diluting the signal. We conclude that the GC probe is sensitive to the actual location of their tracer galaxies inside their tomographic bins. Both the photometric redshift performance and the number density are important contributing factors when performing cosmological inference with GC. When pushing to faint magnitudes, there is no improvement including galaxies that cannot be located in redshift.

Let us now discuss the case where we add GGL to GC (right panel in Fig. 8). We observe that increasing the sample size (moving along the y-axis) has a more significant impact on the improvements of the FoM than the photo-z quality (changes along the x-axis). The greatest improvement, of about 50% for the best photo-z quality samples, takes place going from $m_{\text{VIS}} < 24$ to 24.5. The second largest improvement is of about 25 – 30% when adding objects from $m_{\text{VIS}} < 24.5$ to 25. In the GGL case, source galaxies outside the tomographic bin of the lens galaxy contribute to the signal. The lensing kernel is quite extended in redshift and galaxies beyond the lens contribute to the signal with only a mild dependence on their precise redshift, making the photometric redshift performance less important compared to the GC only case. On the other hand, the statistical nature of detecting the lensing signal makes the number density (and therefore the magnitude limit cut) a more important factor in determining the GGL cosmological inference power.

In the FoM grid, we find a counter intuitive behaviour for some samples when combining GC and GGL (Fig. 8 right panel). If we compare the mid depth and mid depth Euclid samples to the fiducial and optimistic samples at the same number density (along the x-axis), we find that the former pair gives better FoM constraints despite having larger photo-z scatter. This is counter-intuitive as fewer galaxies are properly located in redshift and still the FoM cosmological constraints are slightly better. As we mentioned before, whenever the photo-z performance

![Table 4](https://example.com/table4.png)

Table 4. Values of the FoM for samples defined in Sect. 4.3 with different photo-z accuracy and sample size (same cases as in Fig. 8). The results are normalised to the FoM of the fiducial sample with $m_{\text{VIS}} < 24.5$. For reference, the unnormalised value of our fiducial sample is 713 for GCph and 411 for GCph + GGL. Note that galaxy bias and intrinsic alignments nuisance parameters are free in the latter, which provides a lower FoM than in GCph alone.

![Fig. 8](https://example.com/fig8.png)

Fig. 8. FoM for the samples defined in Sect. 4.3 with different photo-z accuracy and sample size. The size has been reduced by performing a series of cuts in $m_{\text{VIS}}$. The results are normalised to the FoM of the fiducial sample with $m_{\text{VIS}} < 24.5$ (highlighted cell). The figures correspond to the results for using only GCph (left) and for combining it with GGL (right).
degrades, more galaxies supposedly being in our tomographic bin belong to other bins. This effect can increase the effective number of sources for our lenses and thus boost the GGL signal. However, this is at the expense of reducing the cosmological constraining power of the GC\textsubscript{ph} probe. The interplay between these two effects is difficult to gauge. The GGL increase appears slightly more prominent when pushing to fainter magnitude limits that produce a sizeable increase in number density.

The representativeness of the training sample also determines the photo-z performance and thus the cosmological constraining power. For GC\textsubscript{ph}, if we check the difference in FoM between our fiducial sample, trained with a spectroscopic sample that has a completeness drop at faint $m_{\text{vir}}$, and the same photometric sample trained with a fully representative training sample (optimistic sample) we see a gain of about 1–2% in the FoM. Note that the spectroscopic incompleteness in this case is small and only affecting faint magnitudes, so the effect on the FoM is also small. This difference greatly increases when we compare the FoM performance of shallower samples and higher incompleteness in the spectroscopic training sample. If we compare the shallow depth sample that was trained with a sample that has a completeness drop in faint $m_{\text{vir}}$ magnitude to the shallow depth inhomogeneous sample that was trained with a sample that is incomplete in the spectroscopic $n(z)$, the difference between FoMs can be up to 25% for GC\textsubscript{ph} and 39% for both probes combined.

Finally, we look at the difference due to the ground-based photometric depth. The difference between our fiducial and shallow depth cases may represent the change in depth to be achieved in the Southern and Northern hemispheres. For these cases the difference in cosmological constraint power is about 10% at $m_{\text{vir}} < 24.5$ for GC\textsubscript{ph}. This percentage reduces to 2% if we also consider GGL.

### 7.3. Impact on the cosmological parameters constraints

We further investigate the forecasts of the constraints on the cosmological parameters by looking at the parameter uncertainties, $\sigma_i = ((F^{-1})_{ii})^{1/2}$, given by the square root of the diagonal elements of the inverse of the Fisher matrix. The uncertainties are computed for all the photometric samples defined in Sect. 4.3 and for the different sample sizes. For visual clarity, we present the results in grid form in Figs. 10 and 11.

In Fig. 10 we show the uncertainties for the GC\textsubscript{ph} probe. We can appreciate that, in general, the uncertainties have a similar behaviour to the FoM, where the sample down to $m_{\text{vir}} < 24.5$ gives the higher FoM. However, there are parameters, such as $\Omega_m$, $w_a$, and $h$, that do not degrade as much their performance when going to the deeper $m_{\text{vir}} < 25$ sample.

In Fig. 11 we show the uncertainties of the cosmological parameters when we combine the GC\textsubscript{ph} and GGL cosmological probes. Again, we see similar trends compared to the FoM case, but with minor changes in the behaviour of how the uncertainties in some of the parameters vary. The addition of galaxies, increasing the survey depth, and the improvement of the photo-z performance produce lower uncertainties in the $\Omega_m$ and $h$ parameters. The reduction of the uncertainty obtained when considering the deepest $m_{\text{vir}} < 25$ case compared to the $m_{\text{vir}} < 24.5$ is minimal, though.

In addition to the values of the FoM and the uncertainties in the parameters, it is also informative to study the distribution of those uncertainties and the error contours in the determination of pairs of parameters. In Fig. 12 we present the confidence contour plots for our fiducial sample at $m_{\text{vir}} < 24.5$ and 23.5, to check how the number density affects the constraining power, and compare them to our shallow sample at $m_{\text{vir}} < 24.5$, to see the impact of having a sample with shallower ground-based photometry. The contours for the GC\textsubscript{ph} case are shown in the upper panel and the GC\textsubscript{ph} and GGL case in the lower panel. For both probes we see that the fiducial sample gives the best constraints and the largest improvement is gained when the sample size increases. The increase in constraining power with sample size is more prominent in the GC\textsubscript{ph} and GGL combined case in general, and for the parameters that characterise dark energy, $w_0$ and $w_a$, in particular.

### 7.4. Redshift distribution of the photometric redshift bins

To better understand the behaviour of the FoM in Sect. 7.2 and the constraints in Sect. 7.3 we take a closer look at the $n(z)$ of some of the samples used to perform the study. In the top panel of Fig. 13 we compare our fiducial photometric sample for $m_{\text{vir}}$ cuts at $< 25$, $< 24.5$, and $< 23.5$ to see the effects in the $n(z)$ when changing the magnitude limit and therefore the sample size. A shallower cut in magnitude removes objects at higher redshift. In the bottom panel of the figure we compare the $n(z)$ for the fiducial, mid depth, and shallow depth samples at $m_{\text{vir}} < 25$ to see how the behaviour of the $n(z)$ changes with the depth of the ground-based photometry and therefore with the photo-z performance. Overall, the shallower the photometry, the larger the width of the $n(z)$ distributions, especially at higher redshift. This effect spuriously dilutes the correlation signal inside bins and increases the cross-correlation signal between bins, bringing down the GC\textsubscript{ph} constraining power. On the contrary, for the GGL case the widening of the redshift distributions is less important given the width of the lensing kernel. In addition, the effect of an increase in the number density dominates the performance of the FoM that in general increases with depth.
8. Summary and conclusions

Our primary goal is to study the cosmological constraints that can be derived from galaxy clustering studies of photometrically-selected samples using the combination of Euclid and ground-based surveys. For that purpose we use the figure of merit, FoM, defined in Eq. (16) as our performance metric. We want to explore the impact of the ground-based photometry depth as well as the photometric redshift performance on the FoM constraints. To explore the photometric redshift performance, we vary both the survey depth and the spectroscopic specifications used in EC20. For this analysis we use the fiducial photometric sample defined in Sect. 4.3. Figure 6 shows the results to the case of ten bins with equal width since this is the number and type of tomographic bins. We normalise our results to the case of ten bins still provides an increase in the FoM for the GC and for GC only case. However, the photometric redshift scatter starts to be comparable to the bin width for such a large number of bins and our assumptions on how we train and compute the photo-z may start to be too simplistic.

First, we study the optimisation of the FoM with respect to the number and type of tomographic bins. We normalise our results to the case of ten bins with equal width since this is the specifications used in EC20. For this analysis we use the fiducial photometric sample defined in Sect. 4.3. Figure 6 shows the variations in the normalised FoM as a function of the number and type of bins. We find the best compromise for an optimal configuration to be:

- Number of bins: A number slightly larger than ten is preferred. We adopt a default value of thirteen bins for our study. For bins with equal width, the FoM increases when moving from ten to thirteen bins by 35% and 15% for GCph only and for GCph + GGL, respectively. We find that a larger number of bins still provides an increase in the FoM for the GCph only case. However, the photometric redshift scatter starts to be comparable to the bin width for such a large number of bins and our assumptions on how we train and compute the photo-z may start to be too simplistic.
Euclid Collaboration: A. Pocino et al.: Optimizing the photometric sample of the Euclid survey for GC\textsubscript{ph} and GGL analyses

- Type of bins: equal width. For the GC\textsubscript{ph} case the FoM increases by 30% for thirteen bins with equal width compared to equipopulated bins. When combining with GGL the difference in the FoM as a function of bin type is almost negligible.

These results are in nice agreement with Kitching et al. (2019) where they find similar conclusions of the optimum type of bins when optimizing the binning of photometric galaxy samples for cosmic shear analysis. The need of a larger number of bins, especially with good photometric redshift accuracy and the inclusion of intrinsic alignment parameters, to extract all the necessary information for cosmic shear is also found in Bridle & King (2007). In this latter study, they also conclude that the model and freedom of the intrinsic alignment parameters greatly impact the FoM of dark energy.

We further study the dependence of the FoM on the quality of the photo-\textit{z} and the size of the sample. We study possible scenarios of complementary ground-based data for Euclid that could be available in the Southern and Northern hemispheres and in the region in between. We take several magnitude limit cuts and generate realisations of the survey using the Flagship simulation. We also explore different possibilities of spectroscopy data available to train the photometric redshift techniques. We end up with a variety of samples with different number densities and photo-\textit{z} performance properties that try to encompass the possible samples that will be available for Euclid analyses. We compute the dark energy FoM for all these samples to study its variation. Our results are summarised in Fig. 8 and Table 4. For the GC\textsubscript{ph} case, we find a FoM of 713 for our fiducial sample with \( m_{\text{VIS}} < 24.5 \) (remember that galaxy bias is fixed for the GC\textsubscript{ph} case, providing larger absolute values for the FoM than in the combination of GC\textsubscript{ph} + GGL). The FoM improves with photo-\textit{z} quality and sample size. The trend with sample size or magnitude depth reverses when adding galaxies in a magnitude range (between 24.5 and 25 in our case) where photometric redshifts cannot be calibrated and are therefore of poor quality. There is a faster increase of the FoM with sample size in those samples where the photo-\textit{z} performance is better. For example, in the optimistic, fiducial and mid depth cases increasing the sample size from \( m_{\text{VIS}} < 23.5 \) to 24 and from 24 to 24.5 leads to an increase in the FoM of about 20%. When combining GC\textsubscript{ph} and GGL the FoM for the fiducial sample at \( m_{\text{VIS}} < 24.5 \) is 411. The FoM depends more strongly on the sample size (or survey depth) than on the photo-\textit{z} performance properties that try to encompass the possible samples that will be available for Euclid analyses. We compute the dark energy FoM for all these samples to study its variation. Our results are summarised in Fig. 8 and Table 4. For the GC\textsubscript{ph} case, we find a FoM of 713 for our fiducial sample with \( m_{\text{VIS}} < 24.5 \) (remember that galaxy bias is fixed for the GC\textsubscript{ph} case, providing larger absolute values for the FoM than in the combination of GC\textsubscript{ph} + GGL). The FoM improves with photo-\textit{z} quality and sample size. The trend with sample size or magnitude depth reverses when adding galaxies in a magnitude range (between 24.5 and 25 in our case) where photometric redshifts cannot be calibrated and are therefore of poor quality. There is a faster increase of the FoM with sample size in those samples where the photo-\textit{z} performance is better. For example, in the optimistic, fiducial and mid depth cases increasing the sample size from \( m_{\text{VIS}} < 23.5 \) to 24 and from 24 to 24.5 leads to an increase in the FoM of about 20%. When combining GC\textsubscript{ph} and GGL the FoM for the fiducial sample at \( m_{\text{VIS}} < 24.5 \) is 411. The FoM depends more strongly on the sample size (or survey depth) than on the photo-\textit{z} performance. The greatest FoM increase, of about 50%, takes

Fig. 11. Uncertainties of the cosmological parameters for all the cases considered in Sect. 7.2 for combined GC\textsubscript{ph} and GGL.
Fig. 12. Fisher matrix contours for our fiducial sample down to $m_{\text{VIS}} < 24.5$ (blue) and 23.5 (yellow), and the sample with ground-based photometry degraded by 1.75 magnitudes (red). **Top panel:** For GC_{ph}. **Bottom panel:** For GC_{ph} and GGL.
place when adding galaxies from $m_{\text{VIS}} < 24$ to 24.5. The FoM has a weak dependence on the photo-z performance. Generally, it improves with better photo-z accuracy.

The photometric redshift performance depends on the signal-to-noise of the photometry available and on the spectroscopic sample used in the photometric redshift algorithm. In our study, we use a machine learning technique, DNF. The representativeness of the training sample has a significant influence on the photo-z quality. The impact on the FoM is larger when the photometry is shallower. For the optimistic photometry, the improvement in the FoM is minimal, 1–2%, when we train the photometric redshifts with a representative subsample or with a subsample with a completeness drop at faint $m_{\text{VIS}}$. This minimum variation is because the spectroscopic sample incompleteness in the second case only affects the very faintest galaxies. In the cases where the spectroscopy incompleteness is representative of a larger fraction of the galaxy sample the FoM variation is larger. For example, for our shallowest photometric sample, the relative variation in FoM when trained with an incomplete $m(z)$ and with just a completeness drop only at the faintest $m_{\text{VIS}}$, can be of around 30%.

We also investigate the uncertainties in the constrains on our cosmological parameter set across the photo-z quality and sample density space. Cosmological parameters present similar trends to those of the FoM. But there are small differences between the different parameters. For GC$_{\text{ph}}$, in general the smallest uncertainty is achieved when we get the highest FoM, which is the optimistic sample at $m_{\text{VIS}} < 24.5$. However, $\Omega_b$, $w_a$ and $h$ get the smallest uncertainties for the same optimistic sample but for $m_{\text{VIS}} < 25$. The balance between the degradation of the photo-z and the increase in number density affects these parameters slightly differently. For GC$_{\text{ph}}$ combined with GGL, the uncertainty in the cosmological parameters presents a similar behaviour to the FoM trends. The lowest uncertainty in the parameters is achieved when the number density is largest, at $m_{\text{VIS}} < 25$. The trend with photo-z performance does not influence the level of uncertainty. In general the parameters are better constrained when the accuracy on the photo-z determination is higher. However, for some parameters this trend is different in the deepest sample.

To conclude, there is significant gain in the FoM when using a larger number of redshift bins than the nominal ten bins choice of Euclid, especially for GC$_{\text{ph}}$. We study the effect that the accuracy of the photo-z and the survey depth have on the FoM. When using the GC$_{\text{ph}}$ probe, the FoM increases with survey depth and with the reduction in photo-z uncertainties. We study the influence of the training sample in the photo-z performance and its implications on the FoM. We find that adding faint galaxies whose redshifts cannot be properly determined because there are no galaxies of those magnitudes in the training sample decreases the FoM. For the combination of the GC$_{\text{ph}}$ and GGL probes, there is even more gain on the cosmological constraining power when using larger samples than for GC$_{\text{ph}}$ alone. The photo-z quality has slightly less impact on the FoM than for GC$_{\text{ph}}$ alone. In general for the combination of probes, the num-

**Fig. 13.** Top panels: Redshift distribution ($z_{\text{ph}}$) of each tomographic bin for the fiducial sample at $m_{\text{VIS}} < 25$ compared to the fiducial at $m_{\text{VIS}} < 24.5$ (left) and 23.5 (right). Bottom panels: Redshift distribution ($z_{\text{ph}}$) of each redshift bin for the fiducial sample compared to the mid depth (left) and shallow depth samples at $m_{\text{VIS}} < 25$ (right).
ber density has a stronger influence on the FoM than the photo-z accuracy.

**Appendix A: Additional table of \( n(z) \) of photometric redshift bins**

In Section 7.3, we show the \( n(z) \) for the fiducial, mid depth, and shallow samples at \( m_0 < 25 \). In this appendix we present a detailed table containing the number of galaxies and the dispersion of \( n(z) \) at each bin for the fiducial and shallow samples for all magnitude cuts.

**Acknowledgements.** A. Pocino acknowledges financial support from the Secretaria d’Universitats i Recerca del Departament d’Empresa i Coneixement de la Generalitat de Catalunya with additional funding from the European FEDER/ERF funds, L’FSE invertix en el teu futur. I. Tutusaus acknowledges support from the Spanish Ministry of Science, Innovation and Universities through grant ESP2017-89838, and the H2020 programme of the European Commission through grant 776247. S. Camera acknowledges support from the ‘Departments of Excellence 2018-2022’ Grant awarded by the Italian Ministry of Education, University and Research (miur) L. 232/2016. S. Camera is supported by miur through Rita Levi Montalcini project ‘romantico – Probing and Relating Observables with Multi-wavelength Experiments To Help Enlightening the Universe’s Structure’. A. Poursíoudou is a UK Research and Innovation Future Leaders Fellow, grant MR/S016066/1. The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of Euclid, in particular the Academy of Finland, the Agenzia Spaziale Italiana, the Belgian Science Policy, the Canadian Euclid Consortium, the Centre National d’Études Spatiales, the Deutsches Zentrum für Luft- und Raumfahrt, the Danish Space Research Institute, the Fundação para a Ciência e a Tecnologia, the Ministerio de Economía y Competitividad, the National Aeronautics and Space Administration, the Netherlands Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Romanian Space Agency, the State Secretariat for Education, Research and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the Euclid web site (http://www.euclid-ec.org).

**References**

Abbott, T., Abdalla, F., Alarcon, A., et al. 2018, Phys. Rev. D, 98
Ahara, H., Arimoto, N., Armstrong, R., et al. 2017, PASJ, 70
Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, arXiv e-prints, astro-ph/0609591
Behroozi, P.,莴ung, C., Rossin, C., et al. 2017, ApJ, 848, 128
Ivezic, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Kaiser, N. 1987, MNRAS, 227, 1
Kilbinger, M., Heymans, C., Asgari, M., et al. 2017, MNRAS, 472, 2126
Kitching, T. D., Al Ngam, J., Heavens, A. F., et al. 2017, MNRAS, 469, 2737
Kitching, T., Taylor, P. L., Capak, P., Masters, D., & Hoekstra, H. 2019, Phys. Rev. D, 99
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193
Le Févre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14
Lilly, S. J., Faye, O. L. R., Carollo, M. D., et al. 2007, ApJS, 172, 70
Linder, E. V. 2005, Phys. Rev. D, 72
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201
LSST Science Collaboration: Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201
Newman, J. A., Abate, A., Abdalla, F. B., Allam, S., et al. 2015, Astropart. Phys., 63, 81
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5
Nicola, A., Alonso, D., Sánchez, J., et al. 2020, J. Cosmology Astropart. Phys., 2020, 44
Porredon, A., Croce, M., Fosalba, P., et al. 2021, Phys. Rev. D, 103, 043503
Potter, D., Stadel, J., & Teyssier, R. 2017, Comput. Astrophys. Cosmol., 4, 2
Rau, M., Hoyle, B., Parvis, K., & Saito, S. 2017, MNRAS, 466, 2927
Rhodes, J., Nichol, R. C., Aubourg, E., et al. 2017, ApJS, 233, 21
Samuroff, S., Blazek, J., Troxel, M. A., et al. 2019, MNRAS, 489, 5453
Samushia, L., Percival, W. J., Guzzo, L., et al. 2011, MNRAS, 410, 1993
Sánchez, C., Carrasco Kind, M., Lin, H., et al. 2014, MNRAS, 448, 1482
Savini, G., Halkola, K., Carrasco Kind, M., et al. 2020, arXiv e-prints, arXiv:2011.03407
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757
Sánchez, C., Sossi, A., Croccle, M., et al. 2016, MNRAS, 464, 1640
Tanoglidis, D., Chang, C., & Frieman, J. 2019, MNRAS, 491, 3535
Taylor, P. L., Kitching, T. D., McEwen, J. D., & Tram, T. 2018, Phys. Rev. D, 98
Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, Phys. Rev. D, 98, 043528
Tutusaus, I., Martinelli, M., Cardone, V. F., et al. 2020, arXiv e-prints, arXiv:2005.00055
van Uitert, E., Joehmich, J., Joudaki, S., et al. 2018, MNRAS, 476, 4662
Wang, Y., Percival, W., Cimatti, A., et al. 2010, MNRAS, 409, 737
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zhou, R., Newman, J. A., Mao, Y.-Y., et al. 2020, arXiv e-prints, arXiv:2001.06018

1 Institut de Ciencies de l’Espai (IEEC-CSIC), Campus UAB, Carrer de Can Magrans, s/n Cerdanyola del Vallés, 08193, Barcelona, Spain
2 Instituto de Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
3 Institut d’Estudis Espacials de Catalunya (IEEC), Carrer Gran Capità 2-4, 08034 Barcelona, Spain
4 Department of Physics, The Ohio State University, Columbus, OH 43210, USA
5 Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA
6 INFN-Section of Torino, Via P. Giuria 1, I-10125 Torino, Italy
7 Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, I-10125 Torino, Italy
8 INFN-Osservatorio Astrofisico di Torino, Via Osservatorio 10, I-10025 Pino Torinese (TO), Italy
9 INFN-Osservatorio Astronomico di Roma, Via Frascati 33, I-00078 Monteporzio Catone, Italy
10 AIM, CEA, CNRS, Université Paris-Saclay, Université de Paris, F-91191 Gif-sur-Yvette, France
11 Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
12 Université Paris-Saclay, CNRS, Institut d’astrophysique spatiale, 91405, Orsay, France
Table A.1. Number of objects and dispersion, $\sigma$, for each redshift distribution within our fiducial redshift binning and for each $m_{\text{VIS}}$ cut. The table lists results which correspond to the fiducial sample and the shallow sample whose ground-based photometry is degraded by 1.75. The photo-$z$ of both samples are trained with a spectroscopic sample that has a completeness drop in $m_{\text{VIS}}$.

**Fiducial sample**

| $m_{\text{VIS}}$ | z bin | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ |
|------------------|-------|------|---------|------|---------|------|---------|------|---------|------|---------|------|---------|------|---------|------|---------|
| 25               |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
|                  |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
| 24.5             |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
|                  |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
| 24               |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
|                  |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
| 23.5             |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
|                  |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
| 23               |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |

**Shallow sample**

| $m_{\text{VIS}}$ | z bin | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ | Num. | $\sigma$ |
|------------------|-------|------|---------|------|---------|------|---------|------|---------|------|---------|------|---------|------|---------|------|---------|
| 25               |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
|                  |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
| 24.5             |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
|                  |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
| 24               |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
|                  |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
| 23.5             |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
|                  |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
| 23               |       |      |         |      |         |      |         |      |         |      |         |      |         |      |         |      |         |
