Euclid

Euclid Collaboration; Valentijn, E. A.

Published in:
Astronomy and astrophysics

DOI:
10.1051/0004-6361/202244795

Publication date:
2023

A&A 670, A149 (2023)
https://doi.org/10.1051/0004-6361/202244795

Astronomy & Astrophysics

Euclid: Calibrating photometric redshifts with spectroscopic cross-correlations*

K. Naidoo[1,2], H. Johnston[3,2], B. Joachimi[2], J. L. van den Busch[4], H. Hildebrandt[5], O. Ilbert[6,7,8], O. Lahav[2], N. Aghanim[9], B. Altieri[10], A. Amara[11], M. Baldi[12,13,14], R. Bender[15,16], C. Bodendorf[16], E. Branchini[17,18], M. Brescia[19,20], J. Brinchmann[21], S. Camera[22,23,24], V. Capobianco[24], C. Carbone[25], J. Carretero[26,27], F. J. Castander[28,29], M. Castellano[30], S. Cavuoti[20,31,19], A. Cimatti[32,33], R. Cledassou[34,35], G. Congedo[36], C. J. Conselice[37], L. Conversi[38,10], Y. Copin[39], L. Corcione[24], F. Courbin[40], M. Cropper[41], A. Da Silva[42,43], H. Degaudenzi[44], J. Dinis[43,42], F. Dubath[44], X. Dupac[10], S. Dusini[45], S. Farrens[46], S. Ferriol[39], P. Fosalba[28,29], M. Frailis[47], E. Franceschi[13], P. Franzetti[25], M. Fumana[25], S. Galeotta[47], B. Garilli[25], W. Gillard[48], B. Gillis[36], C. Giocoli[49,50], A. Grazian[51], F. Grupp[16,15], S. V. H. Haugan[52], W. Holmes[8], F. Hormuth[53], A. Hornstrup[54], K. Jahnke[55], M. Kümmel[15], A. Kiessling[8], M. Kilbinger[46], T. Kitching[41], R. Kohley[10], H. Kurki-Suonio[56], S. Ligori[24], P. B. Lilje[52], I. Lloro[57], E. Maiorano[13], O. Mansutti[47], O. Marggraf[58], K. Markovic[59], F. Marulli[60,13,14], R. Massey[61], S. Maurogordato[62], M. Meneghetti[13,63], E. Merlin[30], G. Meylan[40], M. Moresco[60,13], L. Moscardini[60,13,14], E. Munari[47], R. Nakajima[58], S. M. Niemi[64], C. Padilla[26], S. Paltani[44], F. Pasian[47], K. Pedersen[65], W. J. Percival[66,67,68], V. Pettorino[46], S. Pires[69], G. Polenta[70], M. Poncet[35], L. Popa[71], L. Pozzetti[13], F. Raison[16], R. Rebolo[72,73], A. Renzi[74,45], J. Rhodes[8], G. Riccio[20], E. Romelli[47], C. Rosset[75], E. Rossetti[60], R. Saglia[16,15], D. Sapone[76], B. Sartoris[15,47], P. Schneider[58], A. Secroun[48], G. Seidel[55], C. Sirignano[74,45], G. Sirri[14], J.-L. Starck[69], C. Surace[6], P. Tallada-Crespí[77,27], A. N. Taylor[36], I. Tereno[42,78], R. Toledo-Moreo[79], F. Torradeflot[77,27], I. Tutusaus[80], E. A. Valentijn[81], L. Valenziano[13,63], T. Vassallo[47], Y. Wang[82], J. Weller[16,15], M. Wetzstein[16], A. Zacchei[47,83], G. Zamorani[13], J. Zoubian[48], S. Andreon[84], D. Maino[85,25,86], V. Scottez[87,88], and A. H. Wright[4]

(Affiliations can be found after the references)

Received 23 August 2022 / Accepted 29 December 2022

ABSTRACT

Cosmological constraints from key probes of the *Euclid* imaging survey rely critically on the accurate determination of the true redshift distributions, $n(z)$, of tomographic redshift bins. We determine whether the mean redshift, $\langle z \rangle$, of ten *Euclid* tomographic redshift bins can be calibrated to the *Euclid* target uncertainties of $\sigma(\langle z \rangle) < 0.002\,(1+z)$ via cross-correlation, with spectroscopic samples akin to those from the Baryon Oscillation Spectroscopic Survey (BOSS), Dark Energy Spectroscopic Instrument (DESI), and *Euclid*'s NISP spectroscopic survey. We construct mock *Euclid* and spectroscopic galaxy samples from the Flagship simulation and measure small-scale clustering redshifts up to redshift $z < 1.8$ with an algorithm that performs well on current galaxy survey data. The clustering measurements are then fitted to two $n(z)$ models: one is the true $n(z)$ with a free mean; the other a Gaussian process modified to be restricted to non-negative values. We show that $\langle z \rangle$ is measured in each tomographic redshift bin to an accuracy of order 0.01 or better. By measuring the clustering redshifts on subsets of the full Flagship area, we construct scaling relations that allow us to extrapolate the method performance to larger sky areas than are currently available in the mock. For the full expected *Euclid*, BOSS, and DESI overlap region of approximately 6000 deg^2, the uncertainties attainable by clustering redshifts exceeds the *Euclid* requirement by at least a factor of three for both $n(z)$ models considered, although systematic biases limit the accuracy. Clustering redshifts are an extremely effective method for redshift calibration for *Euclid* if the sources of systematic biases can be determined and removed, or calibrated out with sufficiently realistic simulations. We outline possible future work, in particular an extension to higher redshifts with quasar reference samples.

Key words. methods: data analysis – techniques: photometric – large-scale structure of Universe

1. Introduction

The European Space Agency's *Euclid* space mission[1] (Laureijs et al. 2011) will map out the positions of billions of galaxies in the Universe. The imaging survey is designed to measure the flux of galaxies in broadband photometric filters over visual and infrared wavelengths, and spectroscopic redshifts will be determined for a subsample of these galaxies using slitless spectroscopy of emission line galaxies (ELGs). Cosmological parameters and models will be constrained via clustering and weak-lensing measurements, made on galaxy samples split into photometric redshift (photo-z) bins. Critical to the accuracy and precision of cosmological parameter constraints is the determination of the true redshift distributions of these tomographic bins (see e.g., Huterer et al. 2006; van den Busch et al. 2020).

A variety of techniques have been developed for this purpose, including the aggregation, or 'stacking', of individual

* This paper is published on behalf of the Euclid Consortium.
[1] https://www.euclid-ec.org/

galaxy photometric redshift probability distribution functions (PDFs; Tanaka et al. 2018; Hadzhiyska et al. 2020; Euclid Collaboration 2021a); 'direct calibration' of the photometric sample $n(z)$ through the re-weighting of spectroscopic colour-redshift spaces (Lima et al. 2008; Hildebrandt et al. 2017; Wright et al. 2020); and, the focus of this paper, cross-correlations with spectroscopic samples, or 'clustering redshifts' (Newman 2008).

Since the seminal work of Newman (2008), cross-correlation redshift calibration has seen a fair amount of development (Schmidt et al. 2013; Ménard et al. 2013; McQuinn & White 2013; Sánchez et al. 2014; Scottez et al. 2016; Morrison et al. 2017; Scottez et al. 2018; Alarcon et al. 2020) but has only recently made its way into state-of-the-art large-scale structure analyses (Hoyle et al. 2018; Gatti et al. 2018, 2022; Hildebrandt et al. 2020; Myles et al. 2021; Cawthon et al. 2022). Clustering redshifts use small angular scales where the signal-to-noise of galaxy clustering is highest, and thus the achievable precision scales directly with the area of photometric–spectroscopic overlap (see Cawthon et al. 2018). Next-generation photometric surveys such as *Euclid* and the *Rubin* Observatory's Legacy Survey of Space and Time (LSST; LSST Science Collaboration 2009) will be able to employ thousands of square degrees of such overlap, in collaboration with the new generation of spectroscopic surveys, such as the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2016), allowing for the most powerful application of clustering redshifts to date.

Clustering redshifts are estimated through the angular cross-correlations of galaxy samples. A tracer sample with secure redshift estimates (from spectroscopy or narrow-band photometry) is cross-correlated with a target sample (which typically has only broadband photometry), for which we wish to accurately and precisely determine the redshift distribution, $n(z)$. Underlying the method is the assertion that the on-sky positions of galaxies in the two samples will only be correlated where they inhabit similar ranges in redshift, that is, where their $n(z)$ overlap. Thus, the amplitude of the cross-correlation at a given redshift contains information about the amplitude of the (known) tracer, $n(z)$, and the unknown target, $n(z)$, but is degenerate with the galaxy biases of both the photometric and spectroscopic galaxy samples. For the spectroscopic sample, the bias can be estimated by measuring the auto-correlation function. However, the photometric sample bias is more difficult to constrain and is the primary source of systematic errors for this mode of study (van den Busch et al. 2020).

Clustering redshift calibration is complementary to the more standard methods of photo-z calibration frequently employed in the literature. Such approaches rely upon real or simulated source photometry, either for characterising colour-redshift relations or for the use of photo-z estimation algorithms. Photometric uncertainties and biases then propagate directly into the $n(z)$ determination. The potential for biased $n(z)$ determination from these techniques can be significant due to degeneracies in the colour-redshift distribution and to the incompleteness of spectroscopic samples in colour-redshift space (Hartley et al. 2020; Wright et al. 2020). Clustering redshifts suffer none of these limitations, as they rely instead upon the accuracy and sufficient coverage of point-estimated redshifts. Being susceptible to independent sources of bias, each method is well suited to cross-check the others; Alarcon et al. (2020) combined the photometry- and clustering-based approaches in a hierarchical Bayesian framework, leveraging all available information to recover mean redshifts with uncertainties of $\sim 3 \times 10^{-3}$, even in scenarios of poor or biased spectroscopic completeness.

For *Euclid* to achieve one of its primary science goals, specifically a figure of merit (FoM) on dark energy of >400 (Laureijs et al. 2011), uncertainties on the mean redshift, $\sigma(\langle z \rangle)$, for each tomographic redshift bin must be less than $0.002(1 + z)$ (68% confidence limits; Laureijs et al. 2011). In this paper we assess the potential for photometric-spectroscopic cross-correlations to form a major component of the redshift calibration in *Euclid*. We aim to use approximately $402\,\mathrm{deg}^2$ of the *Euclid* Flagship simulation (Potter et al. 2017), with realistic photo-z point estimates based upon LSST-like photometry, to forecast the expected uncertainties of clustering redshift calibration as the *Euclid* data volume grows. We define realistic spectroscopic samples in the simulation, based upon DESI and Baryon Oscillation Spectroscopic Survey (BOSS) selection criteria (DESI Collaboration 2016; Dawson et al. 2013), and make use of the clustering redshift method (Schmidt et al. 2013; Ménard et al. 2013) as implemented in the `yet_another_wizz (YAW)` software package (van den Busch et al. 2020). We explore the dependence of $\sigma(\langle z \rangle)$ upon the area of spectroscopic overlap and extrapolate from $402\,\mathrm{deg}^2$ to the full projected overlap region for *Euclid*, BOSS, and DESI. We also explore various methods of mitigation for systematic biases associated with the unknown redshift evolution of the photometric galaxy bias, including the use of spectroscopic auto-correlations and internal consistency checks available only to simulations.

This paper is organised as follows. In Sect. 2 we describe the Flagship simulated data, the realistic photo-z employed in our analysis, and our definitions of spectroscopic sub-samples. Section 3 details our cross-correlation methodology, galaxy bias correction, and the extrapolation of uncertainties to the full *Euclid*, BOSS, and DESI overlap. In Sect. 4 we present the results of clustering redshift calibration for these simulations and in Sect. 5 discuss the implications for *Euclid* clustering redshifts and avenues to pursue for future research.

2. Data: Simulated photometric and spectroscopic samples

Below we explain our process for constructing simulated *Euclid* photometric 'target' samples and spectroscopic 'tracer' samples used to study clustering redshifts for *Euclid*.

2.1. Summary of photometric and spectroscopic samples

In this study, we cross-correlate simulated photometric target galaxies with *Euclid*-like mock photo-z and mock spectroscopic tracer samples from the BOSS-like LOWZ and CMASS; the DESI-like Bright Galaxy Survey (BGS); luminous red galaxies (LRGs) and ELGs; and the *Euclid*-like NISP-S galaxies. We defined our BOSS- and DESI-like galaxy samples on simulated, noiseless photometry, wherein randomly distributed redshift 'failures' are modelled by sparse sampling to appropriate number densities. However, systematic failures arising from noisy photometry, and potentially correlating spatially, or with variable survey depth, were not implemented – the impact of these should be explored in future work.

In Fig. 1 we display the true redshift distributions for the ten photometric bins, defined using one of the photo-z samples defined in Sect. 2.2, and compare them with our mock spectroscopic sample redshift distributions. The lack of spectroscopic samples above redshift 1.8 sets a hard upper-limit for our inferred target $n(z)$'s. Future studies will benefit greatly from

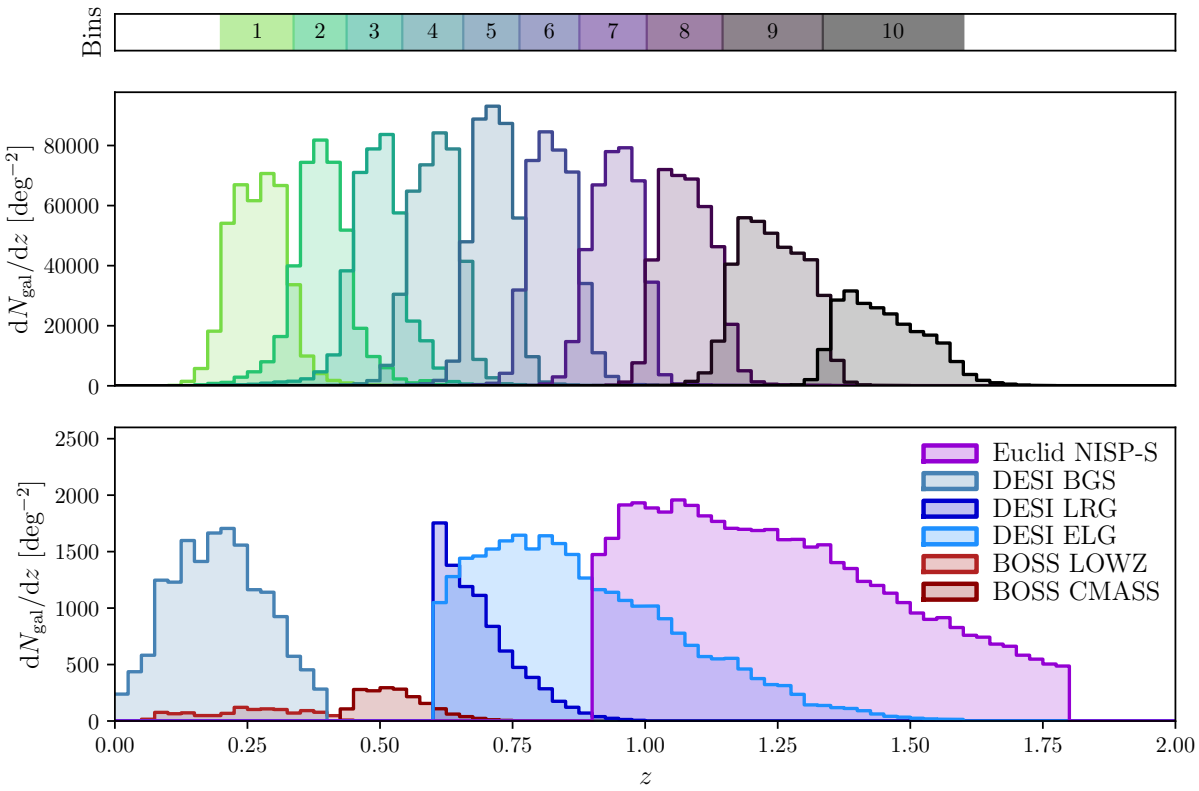

Fig. 1. Photometric redshift distribution in comparison to reference spectroscopic samples. *Top*: redshift ranges for each of the ten tomographic bins. *Middle*: true redshift distribution of each tomographic bin. *Bottom*: redshift distribution of the reference spectroscopic samples: BOSS-like LOWZ and CMASS; DESI-like BGS, LRG, and ELG; and *Euclid* NISP-S. The reference samples only support redshifts up to 1.8, and dictate the maximum redshift of bin 10.

understanding the role that quasar galaxy samples from BOSS, eBOSS (Dawson et al. 2016), or DESI (Yèche et al. 2020) can play in better constraining the *Euclid* photometric sample at higher redshifts. We also note the relatively small tracer–target redshift overlap for galaxies in the range 0.4–0.6, where only mock BOSS galaxies are available for cross-correlation; this has the potential to adversely affect the $n(z)$ constraints for this redshift range.

The footprints for the BOSS[2], DESI[3], and *Euclid* surveys (Amiaux et al. 2012; Euclid Collaboration 2022a) are shown on the sky in Fig. 2. Light grey areas indicate regions of the *Euclid* survey with no overlap, darker grey indicates regions overlapping with at least one of BOSS or DESI, and the darkest grey indicates regions overlapping with both BOSS and DESI. For comparison we also indicate the footprint of the Flagship galaxy simulation used in this study, which is a subset of the full Flagship octant. *Euclid* is projected to overlap with DESI over 9015 deg^2, and with both BOSS and DESI over 6005 deg^2. Unfortunately most of this overlap takes place across the Northern Hemisphere, where photometry from LSST will be limited to the southern-most regions, and therefore the accuracy of the photo-z is likely be worse than that for the ones used in this study. Since these photo-z catalogues were not produced without LSST-like photometry, we could not isolate this study to purely *Euclid* photometry. We have instead chosen to assume that either the *ugrizy* bands of LSST will be supplemented in the Northern Hemisphere by other photometric surveys of equivalent depth (as assumed by Euclid Collaboration 2021b) or that the spectroscopic overlaps of BOSS and DESI are supplied by an equivalent Southern Hemisphere survey such as those that will be observed with the 4-metre Multi-Object Spectroscopic Telescope[4] (4MOST). This means the quality of the *Euclid* photo-z will depends on the angular footprint of the supplementary photometry used from other surveys such as LSST and needs to be factored into future studies as this may also result in $n(z)$ that are angular dependent – which, if this is the case, will drastically complicate cosmological inference and analysis.

2.2. *Euclid* photometric sample

The Flagship simulation (Potter et al. 2017) is a large N-body simulation computed with over two trillion dark matter particles in a box of length 3.78 h^{-1}Gpc using the PKDGRAV3 N-body code. The simulation is constructed assuming a flat Λ cold dark matter model with cosmological parameters $\Omega_{\mathrm{m}} = 0.319$, $\Omega_{\mathrm{b}} = 0.049$, $\Omega_{\Lambda} = 0.681$, $\sigma_8 = 0.83$, $n_{\mathrm{s}} = 0.96$ and $h = 0.67$. Dark matter haloes were identified using ROCKSTAR (Behroozi et al. 2013), and galaxies assigned using halo abundance matching and halo occupation distribution techniques. Simulated *Euclid* photo-z were constructed for galaxies in the Flagship simulation over the RA range 15°–75°, and Dec range 62°–90° (Euclid Collaboration 2021b). We briefly describe the construction of this catalogue below but refer the reader to Euclid Collaboration (2021b) for a complete description.

[2] http://www.sdss3.org/dr9/algorithms/boss_tiling.php#footprint
[3] https://www.legacysurvey.org
[4] https://www.4most.eu/cms/

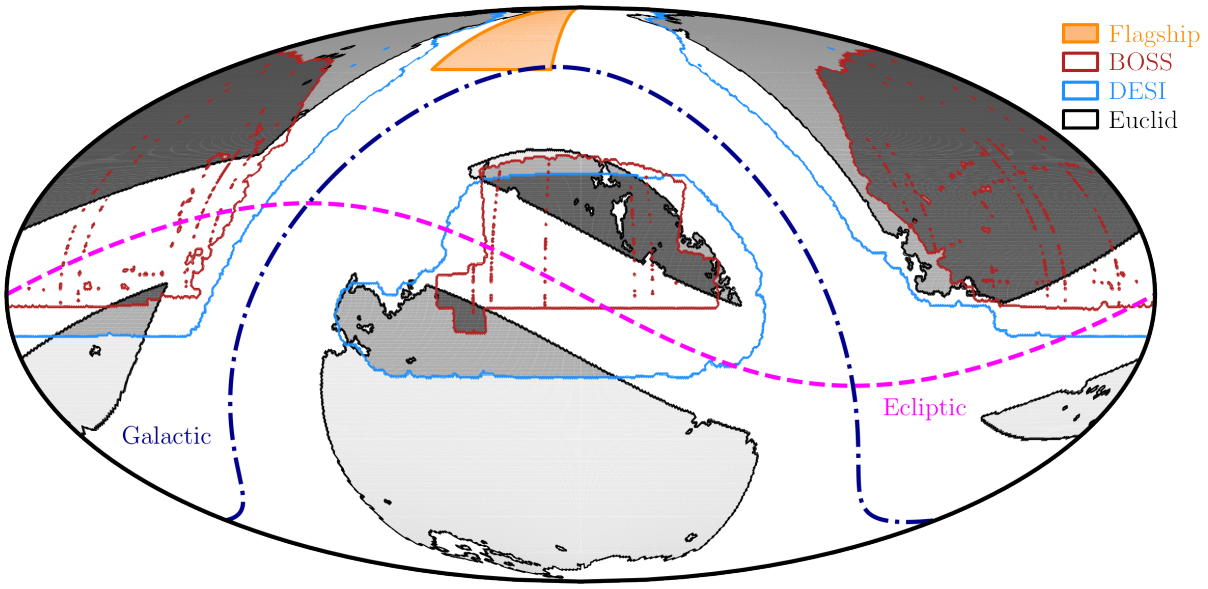

Fig. 2. Survey footprints for BOSS, DESI, and *Euclid* shown on the sky, with darker greys indicating regions of tracer–target sample overlap. The footprint perimeter for BOSS is indicated in red, for DESI in blue, and for *Euclid* in black. Light grey indicates regions of the *Euclid* survey alone, mid-grey indicates regions with *Euclid* overlapping with either BOSS or DESI, and dark grey indicates regions with *Euclid* overlapping both BOSS and DESI. For comparison, the on-sky subregion of the Flagship octant for which we have simulated photo-z is shown in orange. The *Euclid* survey region is defined to avoid both the Ecliptic (dashed pink line) and Galactic (dash-dotted blue line) planes.

To generate realistic photo-z catalogues, realistic fluxes were created by adding Gaussian noise to Flagship galaxy fluxes. The Gaussian noise is defined with a standard deviation of $f_{5\sigma}/5$ (the flux error associated with each galaxy) where $f_{5\sigma}$ is the limiting flux depth at a signal-to-noise ratio (S/N) of 5. The limiting magnitudes at 5σ depth considered to compute flux observations were 26.3, 27.5, 27.7, 27, 26.2, and 24.9 for the *ugrizy* bands of *Rubin* LSST, and 24.6, 23, 23, and 23 for the $I_{\rm E}$ and $Y_{\rm E}$, $J_{\rm E}$, $H_{\rm E}$ bands of *Euclid* at 10σ depth assuming both surveys are at end-of-survey depth.

The noisy $I_{\rm E}$ magnitudes were cut to less than 24.5 (Laureijs et al. 2011) to simulate the *Euclid* $I_{\rm E}$ galaxy sample. Realistic estimates of the photometric redshift of simulated galaxies were made using the directional neighbourhood fitting (DNF; De Vicente et al. 2016) training-based algorithm. The DNF photo-z are estimated according to the proximity in colour-magnitude space of target and training galaxies, where redshifts are known for the training set. The DNF algorithm produces two photo-z estimates; '$z_{\rm mean}$' takes the mean of neighbouring galaxies in colour and magnitude space, and '$z_{\rm mc}$' draws a random redshift from the neighbouring galaxies. A 3.35 deg^2 patch of Flagship was used to create training samples for the DNF. The first training sample was defined to be fully representative in redshift and magnitude space, and produced the photo-z catalogues $z_{\rm mean}$ and $z_{\rm mc}$. A second training sample featured a completeness drop in $I_{\rm E}$ magnitude (emulating the expected spectroscopic completeness fraction versus photometric depth function for surveys such as *Rubin*; see Newman et al. 2015), and produced the photo-z catalogues $z_{\rm mean\,\mathit{Rubin}}$ and $z_{\rm mc\,\mathit{Rubin}}$. We repeated our analysis on each of these four photo-z catalogues.

We produced ten tomographic redshift bins for each photo-z catalogue. These were constructed by selecting galaxies with photo-z between 0.2 and 1.6, and then dividing them into ten bins with approximately equal numbers of galaxies. The definition for the tomographic bins used in this study differ in two important ways to the definitions currently planned for *Euclid*; in the real survey the maximum photo-z will extend up to $z = 2.6$ rather than $z = 1.6$ and secondly 13 rather than 10 tomographic bins will be used (Euclid Collaboration 2020b). The justification for these difference is motivated by two factors, (1) a limit to the galaxies in the Flagship mocks of $z < 2.2$ and (2) a sharp cut off in simulated spectroscopic tracers to $z < 1.8$. Including galaxies with photo-z of greater than $z > 1.6$ will introduce a substantial set of galaxies with true redshifts beyond $z > 1.8$ where clustering redshift measurement cannot be made. Since this limitation is purely based on our setup we wish to limit any effect it may pose on the results and implications for *Euclid*. Quasars (e.g., as observed by DESI; DESI Collaboration 2016) will eventually provide reference samples for calibration at these high redshifts. We note that the clustering redshifts method will need to be carefully optimised for quasar samples, where bias modelling and angular scale utilisation are likely to pose challenges for the much sparser quasar population.

We note that *Euclid*'s ground-based photometry complement in the Northern Hemisphere will be somewhat shallower and less homogeneous than the LSST. Nevertheless, the photometry will meet the stringent *Euclid* requirements and hence yield high-quality photometric redshifts, so that our conclusions remain valid.

2.3. Reference spectroscopic samples

We detail below the various spectroscopic tracer samples that we define in the Flagship simulation, each designed to mimic a current or future galaxy sample observed by *Euclid*, BOSS, or DESI.

2.3.1. *Euclid* NISP-S samples

The *Euclid* near-infrared spectrometer (Schirmer 2022b) is designed to measure spectroscopic redshifts for over 50 million galaxies (referred to as the NISP-S sample). The redshift

determination is to be made with slitless spectroscopy dependent on the detection of emission lines, in particular the Hα line, in the near-infrared. We defined mock NISP-S samples in Flagship by selecting galaxies with Hα fluxes greater than 3×10^{-16} erg cm^{-2} s^{-1} (Laureijs et al. 2011), and cutting to a redshift range between 0.9 and 1.8. Comparisons of the mock NISP-S redshift distribution to the optimistic and pessimistic predictions made by Euclid Collaboration (2020a), we found consistency above redshifts of 1.35, but an overproduction of galaxies at lower redshifts. To correct for this we carry out sparse sampling for mock NISP-S galaxies with $z < 1.35$ to approximately match the optimistic expectation from Euclid Collaboration (2020a).

2.3.2. Baryon Oscillation Spectroscopic Survey LOWZ and CMASS samples

We defined BOSS-like samples according to the colour-magnitude selections specified by Dawson et al. (2013). BOSS produced two spectroscopic LRG samples, LOWZ and CMASS, targeting adjacent redshift intervals $z \in [0.15, 0.43]$ and $z \in (0.43, 0.7]$, respectively, and with true densities of about 30 deg^{-2} and 120 deg^{-2}, respectively. We replicated the BOSS colour-magnitude selections as follows, first defining

$$c_{\parallel} = 0.7\,(g - r) + 1.2\,(r - i - 0.18), \tag{1}$$
$$c_{\perp} = (r - i) - (g - r)/4 - 0.18, \tag{2}$$
$$d_{\perp} = (r - i) - (g - r)/8, \tag{3}$$

where *gri* denote apparent magnitudes (at our simulated LSST depth). We then defined mock LOWZ samples with the following criteria:

$$|c_{\perp}| < 0.2, \tag{4}$$
$$r < 13.6 + c_{\parallel}/0.3, \tag{5}$$
$$16 < r < 19.5, \tag{6}$$

and mock CMASS samples with

$$17.5 < i < 19.9, \tag{7}$$
$$d_{\perp} > 0.55, \tag{8}$$
$$i < 19.86 + 1.6\,(d_{\perp} - 0.8). \tag{9}$$

These cuts provided LOWZ- and CMASS-like objects in the Flagship simulation at rates of 67 deg^{-2} and 139 deg^{-2}, respectively. We sparse sample our selections to the desired densities, with sparse sampling fractions of 0.44 and 0.86.

2.3.3. Dark Energy Spectroscopic Instrument BGS, LRG, and ELG samples

We based our DESI-like sample selections on those detailed in Sect. 3 of DESI Collaboration (2016). DESI will measure the spectra of galaxies at relatively low redshifts with the BGS (designed to use spectra obtained during brighter sky conditions), and will also target two deeper surveys of LRGs and ELGs.

We define our mock BGS sample with the following criteria,

$$r < 19.5, \tag{10}$$
$$z < 0.4, \tag{11}$$

where z is the true object redshift. We define the mock LRG sample with

$$z < 20.46, \tag{12}$$
$$r < 23, \tag{13}$$
$$0.6 < z < 1.0, \tag{14}$$
$$\texttt{color\_kind} = \text{red sequence (0)}, \tag{15}$$

where `color_kind` is a flag labelling red sequence (0), green valley (1) and blue cloud (2) galaxies. Lastly, we defined the mock ELG sample of galaxies with [O II] emission line strengths greater than 8×10^{-17} erg s^{-1} cm^{-2}, and with the following cuts,

$$r < 23.4, \tag{16}$$
$$g - r < 0.7, \tag{17}$$
$$r - z > 0.3, \tag{18}$$
$$0.6 < z < 1.6, \tag{19}$$
$$\texttt{color\_kind} \neq \text{red sequence (0)}, \tag{20}$$

which act to isolate ELGs in colour-colour space, minimising contamination of the sample by lower-redshift objects or by stars. We also applied a hard cut to the expected redshift range, and required that the `color_kind` not specify a red sequence object, though these have comparatively small impacts upon the selection.

These cuts provided BGS-, LRG-, and ELG-like objects in the Flagship simulation at rates of 1174 deg^{-2}, 392 deg^{-2}, and 2021 deg^{-2}, respectively. We achieved the expected number densities (of 700 deg^{-2} for BGS, 285 deg^{-2} for LRG, and 1220 deg^{-2} for ELG) by sparse sampling with sparse sampling fractions of 0.6, 0.73, and 0.6, respectively.

We omitted DESI LRGs with $z < 0.6$, which have selection criteria similar to BOSS. Therefore, the dearth of reference galaxies in the range $0.4 < z < 0.6$ will be much less pronounced for real data, which makes our analysis conservative.

3. Method

In this section we briefly outline the clustering redshift method used in this study, the subsequent method for modelling the redshift distributions, and lastly the area rescaling approach used to predict the uncertainty on the mean redshift for the full *Euclid* survey.

3.1. Clustering redshifts

The angular correlation function $\omega_{12}(\theta, z)$ for two samples on the sky (denoted by 1 and 2), where sample 1 is at a fixed redshift z, is given by the Limber (1953) relation,

$$\omega_{12}(\theta, z) = b_1(z) \int_0^{\infty} \mathrm{d}z'\, n_2(z')\, b_2(z')\, \xi\left[R(\theta, z, z'), z\right], \tag{21}$$

where $b_1(z)$ and $b_2(z)$ are the redshift-dependent biases for samples 1 and 2, respectively, $n_2(z)$ is the normalised redshift selection function for sample 2, $\xi(R, z)$ is the matter auto-correlation function at redshift z and comoving three-dimensional separation

$$R(\theta, z, z') = \sqrt{\left[\chi(z) - \chi(z')\right]^2 + \left[f_{\mathrm{K}}(z')\,\theta\right]^2}, \tag{22}$$

where $\chi(z)$ is the radial comoving distance at redshift z and $f_{\mathrm{K}}(z)$ is the angular diameter distance at redshift z.

For the computation of clustering redshifts, the following assumptions are made. We first consider the angular correlation function for a spectroscopic tracer sample (denoted with an index s) and photometric target sample (denoted with an index p)

$$\omega_{\rm sp}(\theta, z) = b_{\rm s}(z) \int_0^\infty {\rm d}z'\, n_{\rm p}(z')\, b_{\rm p}(z')\, \xi\Big[R(\theta, z, z'), z\Big]. \tag{23}$$

Angular correlation functions are computed for narrow redshift slices of the spectroscopic tracer with a mean redshift z_i (where i denotes the redshift bin) and bin width of Δz. Furthermore, assuming measurements are only made for small values of θ we can can make the approximation that $\xi \neq 0$ only if the integral is calculated within the bin's redshift range $z_i \pm \Delta z/2$. The narrow sizes of the bins means n_p and b_p can be approximated as constants and therefore Eq. (23) simplifies to

$$\omega_{\rm sp}(\theta, z_i) = b_{\rm s}(z_i)\, n_{\rm p}(z_i)\, b_{\rm p}(z_i)\, \omega_{ii}(\theta, z_i), \tag{24}$$

where

$$\omega_{ii}(\theta, z_i) = \int_{z_i-\Delta z/2}^{z_i+\Delta z/2} {\rm d}z'\, \xi\Big[R(\theta, z_i, z'), z_i\Big]. \tag{25}$$

In our study these expressions were evaluated for the transverse physical scale $r = \theta\chi/(1+z)$, which led to the final definition of the angular correlation function for clustering redshifts,

$$\omega_{\rm sp}(r, z_i) = b_{\rm s}(z_i)\, n_{\rm p}(z_i)\, b_{\rm p}(z_i)\, \omega_{ii}(r, z_i). \tag{26}$$

Thus we can evaluate the target sample's redshift selection function $n_{\rm p}$ by rearranging Eq. (26) for

$$n_{\rm p}(z_i) = \frac{\omega_{\rm sp}(r, z_i)}{b_{\rm s}(z_i)\, b_{\rm p}(z_i)\, \omega_{ii}(r, z_i)}. \tag{27}$$

The above equation requires knowledge of the bias-redshift relation for both the spectroscopic tracer and photometric target sample. By evaluating Eq. (26) for the angular auto-correlation functions for the spectroscopic and photometric samples we can define the bias functions as

$$b_{\rm x} = \sqrt{\frac{\Delta z\, \omega_{\rm xx}(r, z_i)}{\omega_{ii}(r, z_i)}}, \tag{28}$$

where x can denote either the spectroscopic tracer or photometric target sample. Since the auto-correlation function is evaluated in a single, narrow redshift bin, the normalised redshift selection function $n_{\rm x} \to 1/\Delta z$. Using these relations, Eq. (27) can be expressed as

$$n_{\rm p}(z_i) = \frac{\omega_{\rm sp}(r, z_i)}{\Delta z\, \sqrt{\omega_{\rm ss}(r, z_i)\, \omega_{\rm pp}(r, z_i)}}. \tag{29}$$

However, in practice $\omega_{\rm pp}(r, z)$ is difficult to obtain, since we do not know the photometric sample redshifts a priori, and thus cannot bin target galaxies to measure the correct auto-correlations and anchor the photometric sample galaxy bias. The result is that the $n_{\rm p}(z)$ is biased and requires a correction scheme to account for any redshift evolution of the photometric sample galaxy bias. Several bias mitigation strategies can be employed.

Method 1: No corrections are applied. This is useful to test the success of the following bias correction methods. This means we assume $n_{\rm p}(z) \propto \omega_{\rm sp}(r, z)$.

Method 2: The spectroscopic biases are computed from the auto-correlation functions and are incorporated in the $n_{\rm p}(z)$ computation. This means we use Eq. (29) but $\omega_{\rm pp}(r, z)$ is evaluated for the entire tomographic bin instead of within thin redshift slices.

Method 3: 'Self-consistent bias mitigation' uses a one-parameter redshift power law, $\mathcal{B}_\alpha(z)$, fitted to the observed auto- and cross-correlations to account for redshift evolution in $b_{\rm p}$ (Davis et al. 2018; van den Busch et al. 2020)

$$\mathcal{B}_\alpha(z) = (1+z)^\alpha \propto \sqrt{\Delta z_i\, \bar{\omega}_{\rm pp}(z)}, \tag{30}$$

where α is a free model parameter, and barred correlations $\bar{\omega}(z)$ correspond to integrals over $\omega(r, z)$ between chosen limits $r_{\rm min}$ and $r_{\rm max}$. This means $\omega_{\rm pp}$ in Eq. (29) is replaced with $\mathcal{B}_\alpha(z)^2/\Delta z$.

Method 4: The bias for the photometric sample is computed from the auto-correlation function for only the photometric galaxies with true redshifts within a given redshift slice. Equation (29) is therefore applied exactly. We note that this method can only be applied to simulated data, and is useful for assessing the accuracy of other bias correction schemes.

3.2. Cross-correlating redshifts

We compute clustering correlations in the Flagship data using the single-bin (defined by $r_{\rm min}, r_{\rm max}$) method of Schmidt et al. (2013) and Ménard et al. (2013), implemented with the Davis-Peebles clustering estimator (Davis & Peebles 1983) as follows:

$$\bar{\omega} = \frac{N_{\rm R} \int_{r_{\rm min}}^{r_{\rm max}} {\rm d}r\, W(r)\, {\rm DD}(r)}{N_{\rm D} \int_{r_{\rm min}}^{r_{\rm max}} {\rm d}r\, W(r)\, {\rm DR}(r)} - 1, \tag{31}$$

where $W(r) \propto r^\beta$ (e.g., $\beta = -1$; Schmidt et al. 2013; Ménard et al. 2013), ${\rm DD}(r)$ and ${\rm DR}(r)$ are galaxy-galaxy and galaxy-random pair counts (in the comoving bin centred on r), respectively, and $N_{\rm R}/N_{\rm D}$ re-normalises pair counts to account for over-sampling of random points relative to the data. The method follows van den Busch et al. (2020) and uses the software package `YAW`[5] described in van den Busch et al. (2020) with errors and covariance matrices estimated using bootstrap resampling.

3.3. Redshift distribution fitting

Since the clustering redshift measurements have Gaussian error distributions, they allow for negative $n(z)$ values. Of course, no true PDF can exhibit this property and as a result the raw clustering redshift measurements will artificially inflate the uncertainties on the mean redshift $\langle z \rangle$. Therefore, we did not determine $\langle z \rangle$ directly from the clustering redshift measurements but rather by fitting model $n(z)$ distributions. Two model $n(z)$ distributions were considered. The first model, the shifted-true model, is simply the true $n(z)$ measured from simulations with the addition of two free parameters: A and δz, which are the amplitude and shift parameters, respectively (explained in greater detail in Sect. 3.3.1). The second model is the suppressed Gaussian process (suppressed-GP) model discussed in Sect. 3.3.2, which, unlike the first model, makes no assumptions about the shape of the $n(z)$ distribution. In both cases, the models are not normalised, since normalisation cannot be enforced on the clustering redshift measurements. Normalisation of the model is therefore only conducted when we are calculating the mean of the PDF.

[5] https://github.com/jlvdb/yet_another_wizz

3.3.1. Shifted-true

In the shifted-true model the clustering redshift measurements are fitted to the true $n(z)$. If these were real data, this true $n(z)$ would not be known and would have to be estimated through forward modelling on simulations such as Flagship, or from direct calibration (van den Busch et al. 2020). The measured $n(z)$ is fitted with two parameters, a shift along the redshift direction δz and an amplitude A,

$$n_{\rm model}(z;\delta z, A) = A\, n_{\rm true}(z+\delta z). \tag{32}$$

Fitting with the true distribution allows us to test the robustness of the output clustering redshift measurements, since the resultant best-fit values should be consistent with $\delta z = 0$ in the absence of systematic errors. The parameters are determined using a Gaussian likelihood,

$$\ln \mathcal{L}\left[\delta z, A \,|\, \hat{n}(z), \sigma_{\hat{n}}(z)\right] = -\frac{1}{2}\left[\frac{\hat{n}(z) - n_{\rm model}(z;\delta z, A)}{\sigma_{\hat{n}}(z)}\right]^2, \tag{33}$$

where $\hat{n}(z)$ and $\sigma_{\hat{n}}(z)$ are the clustering redshift measurements and 68.3% confidence intervals, respectively.

3.3.2. Suppressed Gaussian process

Whilst the shifted-true model is a valuable test for systematic errors, it is important to note that we will not have access to such a model for real data. In previous studies, clustering redshifts were fitted to models based on simulated photometric redshift samples while Gatti et al. (2022) used direct calibration of the data. Although the simulated photometric redshift $n(z)$ may work as a good proxy for the real model, it may be better to have a model that is more flexible and could be relied upon in a more general setting. With this in mind, we present a new, non-parametric approach to fitting clustering redshift distributions based upon Gaussian processes (GPs, and featuring a suppression function that damps signals in regions where the clustering redshift measurements are consistent with zero.

A GP is first fitted to the clustering redshift measurements. This is carried out by the `Python` package `George`[6] using a Matern-3/2 kernel. One benefit of such a model is that we can now draw random samples from the distribution to measure the mean redshift and its uncertainty. However, as with the clustering redshift measurements there is nothing limiting the GP model from drawing samples that are negative. This issue is further compounded by the fact that uncertainties crossing zero will create spurious fluctuations in the GP samples. To ensure that the GP is positive, and to remove spurious signals where the clustering redshift measurement is consistent with zero, we apply a suppression function to the GP realisations. The suppression function is defined by the following expression,

$$S(x,k) = \begin{cases} 0, & x \le 0, \\ 1-(1-x)^k, & 0 < x < 1, \\ 1, & x \ge 1, \end{cases} \tag{34}$$

where k is a damping factor taken to be 0.3 in this study and

$$x = \frac{n^{\rm i}_{\rm GP}(z)}{\Sigma_{\rm T}\left(\mathcal{G} * \Sigma\right)}, \tag{35}$$

$n^{\rm i}_{\rm GP}(z)$ is a random GP realisation, $\Sigma_{\rm T} = 3$ is the S/N threshold, $\Sigma = n_{\rm GP}(z)/\sigma_{\rm GP}(z)$ is the S/N function (where $n_{\rm GP}$ and $\sigma_{\rm GP}$ are the GP mean and 68.3% confidence interval), $\mathcal{G}$ is a Gaussian with standard deviation of 0.05 and $*$ is the convolution operator. The Gaussian convolution with the S/N prevents spurious random fluctuations in the S/N from impacting the activation of the suppression function and ensures the tails of the clustering redshift $n(z)$ are not too harshly damped. A standard deviation drastically smaller than 0.05 will be smaller than the bin size of the clustering redshift measurements and therefore will be equivalent to no smoothing. On the other hand, larger values will more strongly correlate bins that may be problematic since, in some cases, the $n(z)$ profile spans across only ~5 bins and may be completely washed out.

An example of this procedure is shown in Fig. 3. The suppressed-GP has the desired effect, that is to say, noisy PDFs are not drawn in regions of the clustering redshifts that are consistent with zero. A potential shortcoming is that tails in the distribution will be suppressed if they have low S/N. Steps to mitigate this effect could be studied in future work by exploring alternative suppression functions and techniques.

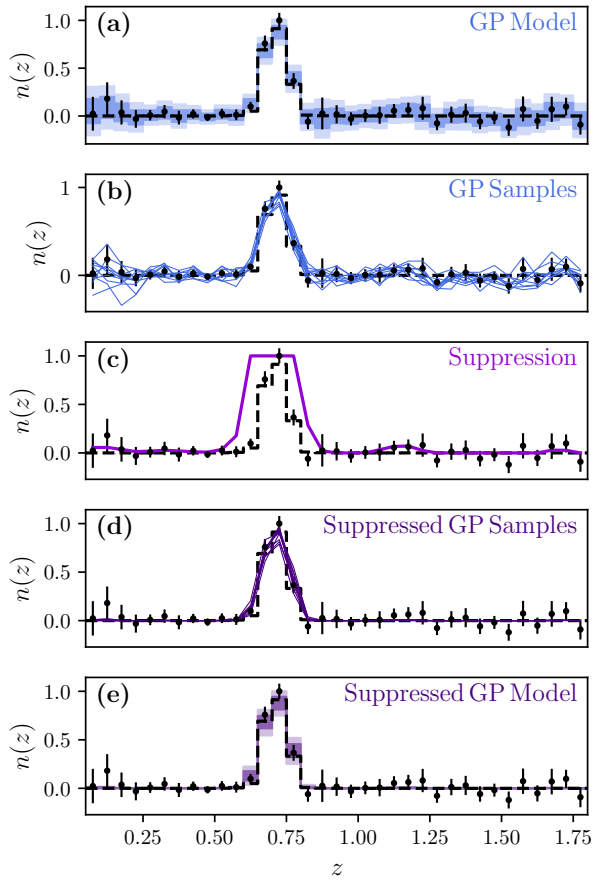

Fig. 3. Method for fitting suppressed-GP models to clustering redshift, $n(z)$, measurements. The true $n(z)$ for one tomographic bin is shown with dashed black lines. The measured clustering redshift, $n(z)$, is shown with black markers and error bars. *(a)* We fit a GP model to the clustering redshift, $n(z)$, measurements. *(b)* We draw random realisations from the GP. *(c)* We construct a suppression function, taking as input the GP draw and the smoothed S/N of the clustering redshift measurement. *(d)* We multiply the random realisations from the GP by the suppression function. This ensures that the suppressed-GP model is always positive and suppresses low-S/N fluctuations in the GP. *(e)* We construct 68.3% and 95.5% confidence envelopes from samples of the suppressed-GP model.

[6] https://george.readthedocs.io/en/latest/

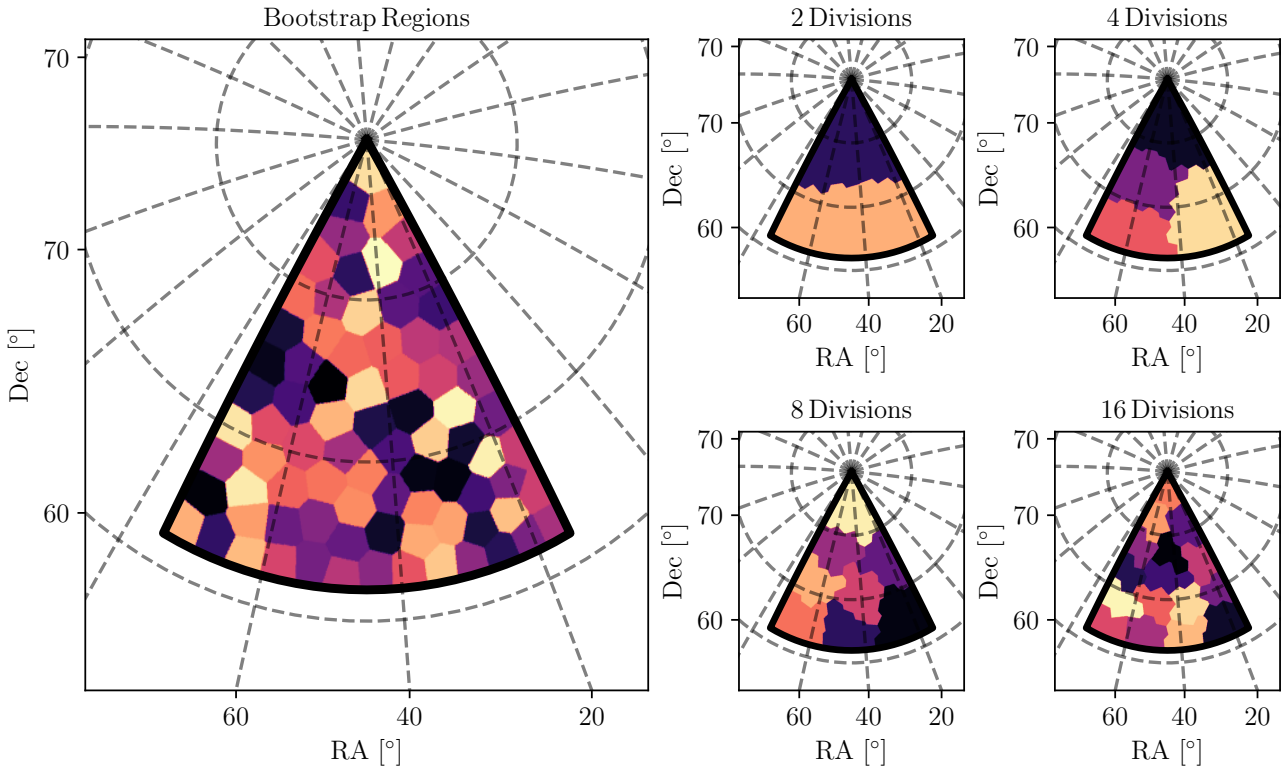

Fig. 4. Flagship footprint and area rescaling subregions. *Left*: bootstrap subregions, indicated with different colours, that are used to calculate errors in `YAW`. *Right*: different subregions of the Flagship footprint, shown with different colours, used to establish the relation between the tracer–target overlap area and the uncertainty on the mean redshifts of tomographic bins. The four subpanels show the Flagship footprint divided into 2 (*top left*), 4 (*top right*), 8 (*bottom left*), and 16 (*bottom right*) subregions. These are plotted on an orthographic projection of a subset of the sky focused on the Flagship footprint. The RA and Dec grid is shown with dashed grey lines. The values of the grid lines are indicated where they intercept the *x*-axis for the RA grid and the *y*-axis for the Dec grid.

3.4. Area rescaling

We seek to provide estimates for the uncertainties of the $n_\mathrm{p}(z)$ determination as a function of the projected *Euclid* data volume; thus we must extrapolate from the 402 deg^2 of the Flagship area to the larger spectroscopic overlap areas, and also attempt to characterise the impact of sample variance. *Euclid* is projected to overlap with DESI over 9015 deg^2, and with both DESI and BOSS over 6005 deg^2. To rescale the determined uncertainties on mean redshifts, we divided the Flagship lightcone into multiple subdivisions (see Fig. 4) and re-ran our analysis on each of these subdivisions independently. This allowed us to approximate the relationship between the tracer–target overlap area and the uncertainties on $\langle z \rangle$. We then fitted these relationships with power laws, and extrapolated to estimate the uncertainties from the full *Euclid*, BOSS, and DESI overlap area. We note we do not redefine the bootstrap regions when we run the analysis on each subdivision, this means estimates of the error are less accurate for smaller regions since they are based on fewer defined bootstrap regions. An alternative approach would be to use purely shot-noise covariance, but we leave this for future work. Furthermore, since these regions overlap, measurements on the subregions will be strongly correlated.

4. Results

This section details the results of our clustering redshift calibration, focusing on the uncertainties of mean redshift determination across our ten mock tomographic bins.

4.1. Clustering redshift measurements

We measured clustering redshifts using the `YAW` software package between scales $r_\mathrm{min} = 100$ kpc and $r_\mathrm{max} = 1000$ kpc (following the scales used by KiDS; van den Busch et al. 2020), considering separately each of the four simulated *Euclid* photometric redshift catalogues presented by Euclid Collaboration (2021b). We performed the analysis independently for the full Flagship region (with photo-z), and for each subregion (shown in Fig. 4).

In Fig. 5 we compare the clustering redshift measurements obtained by `YAW` (shown with purple error bars) to the true redshift distribution of the 'mean *Rubin*' photometric catalogue on Flagship. The measurements clearly trace the true distribution, and showcase the capability of clustering redshifts to retrieve redshift distributions for photometric catalogues. There are, however, some significant fluctuations, for example at the peak of bin 2. The measurements in Fig. 5 are corrected for the spectroscopic sample galaxy bias (method 2, discussed in Sect. 3.1).

4.2. Mean redshifts

Having measured the clustering redshifts, we wished to constrain and compare each bin's $n(z)$ distribution by measuring the mean redshift $\langle z \rangle$ and comparing to the truth. The mean redshift is determined by calculating

$$\langle z \rangle = \sum_{i=1}^{N_\mathrm{CC}} z_i \, n(z_i) \, \Delta z, \tag{36}$$

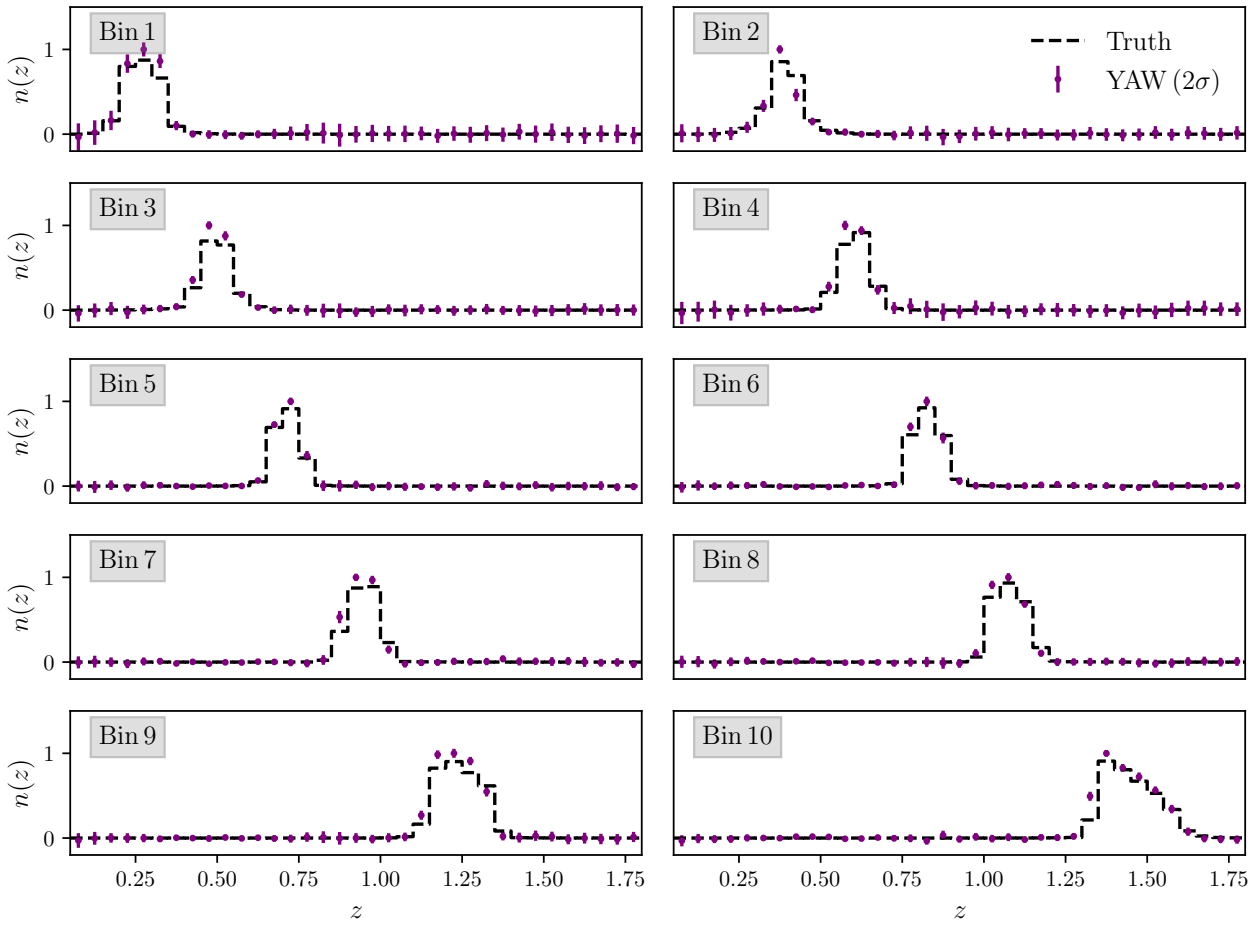

Fig. 5. Clustering redshift measurements in comparison to the true $n(z)$ for ten simulated *Euclid* Flagship tomographic photo-z bins. Measurements from `YAW` are shown in purple, with error bars indicating 95.5% confidence intervals. The measurements clearly trace the true $n(z)$, albeit with some spurious local fluctuations. These fluctuations probably arise due to an incomplete galaxy bias correction methodology. The bias correction method used in this figure is method 2 (Sect. 3.1).

where $n(z)$ is the normalised model redshift distribution within the redshift ranges $z_{\rm min}$ and $z_{\rm max}$, z_i is the redshift centre of a clustering redshift bin and $n(z_i)$ a clustering redshift measurement. The i is used to denote a specific measurement and $N_{\rm CC}$ the total number of measurements. We computed $\langle z\rangle$ for the $n(z)$ model only within the redshift range where clustering redshift measurements were made (therefore $z_{\rm min} = 0.05$ and $z_{\rm max} = 1.8$) while true values $\langle z\rangle_{\rm truth}$ were computed from the true $n(z)$ measured across the entire Flagship range $z < 2.2$. For both the shifted-true and suppressed-GP models, we drew 1000 realisations of the redshift distribution (for the shifted-true model, this means drawing random shift parameters from the posterior likelihood). We determine the mean redshift $\langle z\rangle$ from the sample $n(z)$, and the standard deviation on the mean $\sigma(\langle z\rangle)$ from the variance of the sample $\langle z\rangle$. In Fig. 6 we display the mean redshift error $\langle z\rangle - \langle z\rangle_{\rm truth}$ for both models, in each bin and for the different photometric catalogues, constrained by the full Flagship footprint. The constraints show significant biases, in particular for bins 7–10, which are biased to lower redshifts for both models. For the shifted-true model we see some significant biases in bins 2 and 4, independent of photo-z method. The suppressed-GP is biased for bins 1 and 4, but only for the photometric catalogue 'mc'. The broadly similar performance of the suppressed-GP and shifted-true models shows that the suppressed-GP model is performing well, and that it is a robust alternative for $n(z)$ fitting. We also note that the directions and amplitudes of biases appear to be consistent between models, suggesting that systematic errors are independent of the fitting method for this analysis. Rather surprisingly, bin 3, which is placed in the region ($z = 0.4$–0.6) with fewer spectroscopic tracers, is relatively unbiased. One interesting property of this bin is that the tails on both sides are well constrained, due to the high number of spectroscopic tracers on either of the tails of the distribution. This suggests the clustering redshift performance on the tails may be driving biases in the other results. We note that we also find the errors on $z_{\rm mean}$ and $z_{\rm mean\,\mathit{Rubin}}$ to be smaller than the $z_{\rm mc}$ and $z_{\rm mc\,\mathit{Rubin}}$ for the suppressed-GP owing to the larger tails in the 'mc' methods.

In Fig. 7 we compare the bias correction methods outlined in Sect. 3.1 by calculating the reduced $\chi^2_{\rm r}$ for each method,

$$\chi^2_{\rm r} = \frac{1}{N_{\rm Bins}} \sum_{i=1}^{N_{\rm Bins}} \left[\frac{\langle z\rangle - \langle z\rangle_{\rm truth}}{0.002\,(1 + \langle z\rangle_{\rm truth})} \right]^2, \tag{37}$$

on the photometric catalogue 'mean *Rubin*', where $N_{\rm Bins}$ is the number of tomographic bins. We see that method 1 (no galaxy bias correction) performs worst with a $\chi^2_{\rm r} = 7.8$, while method 2 (spectroscopic sample bias correction; the baseline method for all other figures) performs slightly better with a $\chi^2_{\rm r} = 4.2$. Method 3, which fits a power law for photometric sample bias correction, performs very similarly to method 2 – with a $\chi^2_{\rm r} = 4.2$. Method 4, which performs the best with a $\chi^2_{\rm r} = 3.7$, measures the true bias for photometric galaxies in each of the

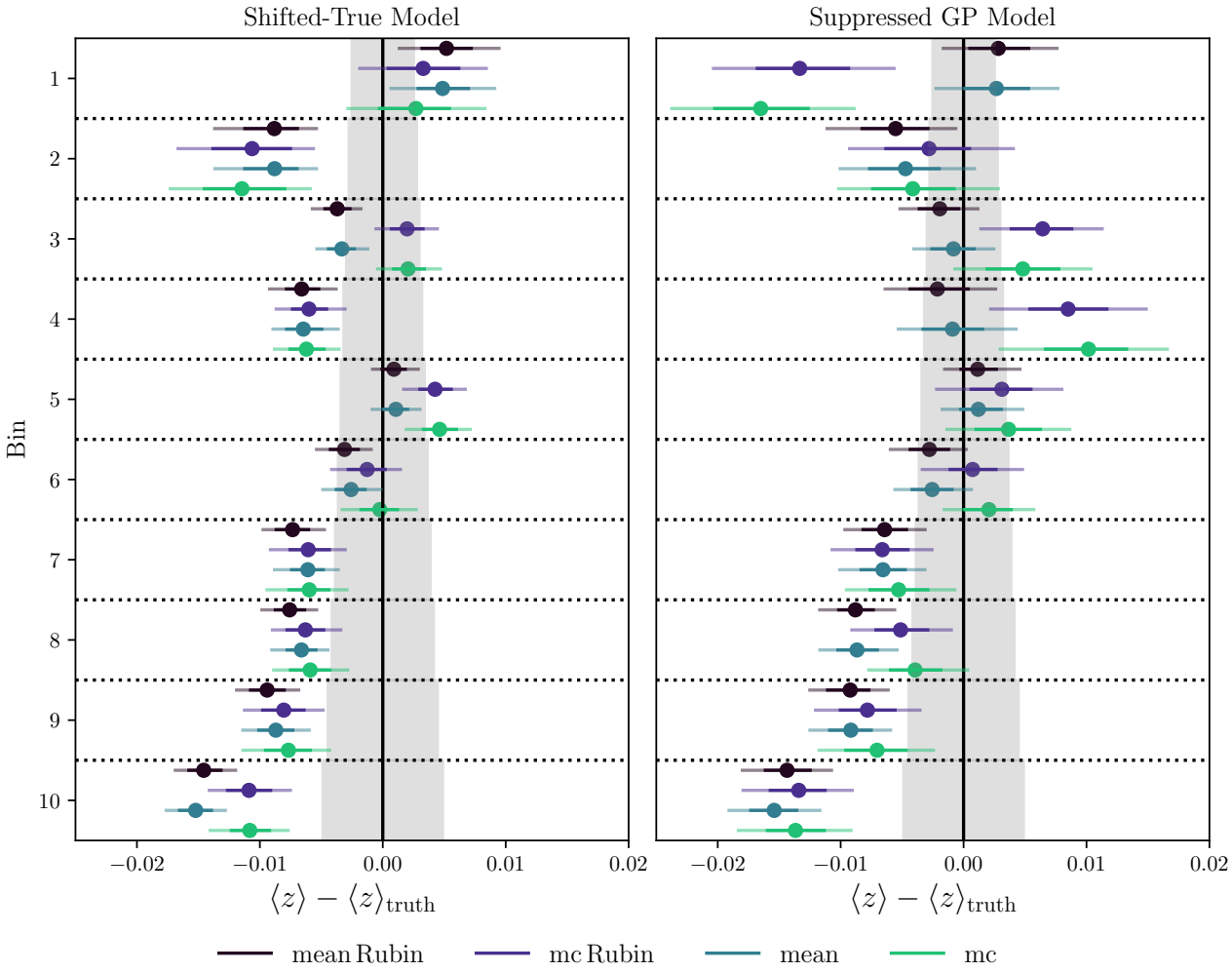

Fig. 6. Error on the mean redshift determination $\langle z \rangle - \langle z \rangle_{\rm truth}$ for ten *Euclid* Flagship tomographic bins, for each of our four photometric redshift catalogues ('mean *Rubin*' in dark purple, 'mc *Rubin*' in purple, 'mean' in dark green, and 'mc' in light green, with dark and light shades indicating 68.3% and 95.5% uncertainties). The grey bands show the target uncertainties for *Euclid*. Two $n(z)$ fitting methods are shown: the shifted-true model (*left*, Sect. 3.3.1) and the suppressed-GP model (*right*, Sect. 3.3.2), each fit to clustering redshift measurements made with spectroscopic sample bias corrections (i.e. method 2, Sect. 3.1).

spectroscopic slices, and cannot be measured on real data. However, in simulations like Flagship, it allows us to test the significance and implications of residual biases. Method 4 shows that correcting for photometric galaxy biases in bins 1–6 removes any systematic biases on the determination of the mean redshift, and indicates that the mean redshift uncertainties may be underestimated by other methods. It should be possible to calibrate for such underestimations of $\sigma(\langle z \rangle)$ with realistic simulations like Flagship. We assess whether or not the extra noise component revealed by method 4 is likely to affect our conclusions, as we forecast uncertainties on $\langle z \rangle$ versus tracer–target overlap area in Sect. 4.3. Bins 7–10 remain significantly biased in $\langle z \rangle$, even for method 4, which is why the $\chi^2_{\rm r}$ is still significantly larger than 1. The causes of these persistent biases are currently unknown, and require further investigation. They may be affected by the cross-correlation estimator; a move to the superior Landy & Szalay (1993) estimator should be explored in future work. Alternatively, the definition of target sample tomographic bins may be sub-optimal; wider equi-populated bins at higher redshifts will offer lower signal-to-noise of clustering redshift measurements, since those galaxies span a larger volume in a flux-limited survey where the number density falls with increasing redshift. Optimisation of the target tomography is another promising avenue for future work.

4.3. Area rescaling

A determination of whether or not clustering redshifts can constrain the $\sigma(\langle z \rangle)$ of *Euclid* tomographic redshift bins to less than $0.002\,(1 + z)$ requires extrapolation. We estimated the relation between $\sigma(\langle z \rangle)$ and tracer–target overlap area by measuring $\langle z \rangle$ within subregions of the Flagship region (see Fig. 4). We fitted the $\sigma(\langle z \rangle)$ data with power laws and then extrapolated to the total projected overlap. The relations can be seen in Fig. 8 for the shifted-true model, calibrated by method 2 (Sect. 3.1) clustering redshifts measured on the 'mean *Rubin*' photometric catalogue. The extrapolation projects the uncertainties to be significantly smaller than the required uncertainties for *Euclid*. We can be confident of this statement, as even the 402 deg^2 Flagship area yields uncertainties approaching or surpassing the *Euclid* requirement. Similar results are seen for the other three photometric catalogues. A possible cause for concern is the underestimation of the errors in the four highest-redshift bins, shown in Fig. 7. This underestimation can be calibrated with comparisons

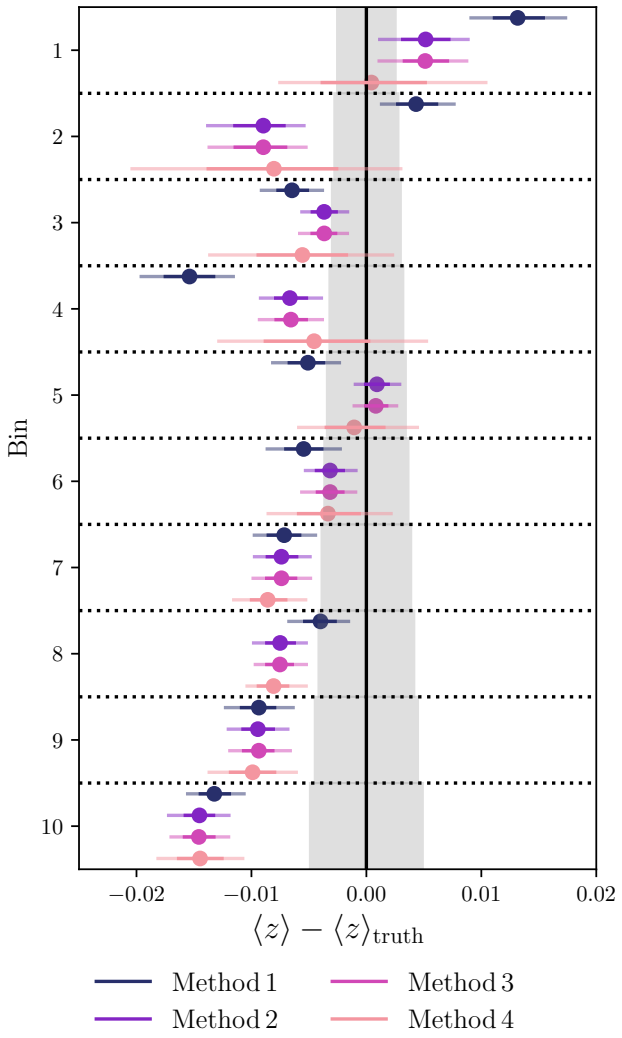

Fig. 7. Error on the mean redshift determination, $\langle z \rangle - \langle z \rangle_{\rm truth}$, in ten *Euclid* tomographic redshift bins, for each of the four galaxy bias correction strategies (Sect. 3.1), compared via the shifted-true model (Sect. 3.3.1). In dark blue is method 1, where no corrections are applied; in purple is method 2, where the spectroscopic galaxy bias is calibrated by spectroscopic auto-correlations; in magenta is method 3, where the photometric galaxy bias is calibrated by a power-law fit to noisy, photo-z-binned photometric auto-correlations; and in light pink is method 4, where the photometric galaxy bias is calibrated using auto-correlations binned by true redshifts (not applicable to real data). Dark and light shades indicate 68.3% and 95.5% confidence intervals. The grey band shows the target uncertainties for *Euclid*. Method 4 shows that systematic biases in bins 1–6 are caused by incomplete bias corrections, whilst the biases persist for higher-redshift bins, suggesting some other source of systematic error.

to simulations (such as Flagship), and accommodated by widening the error bars. We find that these uncertainties are underestimated by a factor of about two. Since Fig. 8 shows the projected errors for the full overlap region to be around an order of magnitude smaller than the *Euclid*'s required uncertainties, we are confident that these underestimated errors can be accommodated, and that the *Euclid* target remains achievable.

5. Conclusion

We measured clustering redshifts in the Flagship simulation to test their uncertainties in determining the mean redshifts, $\langle z \rangle$, for *Euclid* tomographic bins. The method uses cross-correlations with mock spectroscopic samples, modelled after BOSS, DESI, and *Euclid* NISP-S. Clustering redshifts were determined using the `YAW` software, for transverse pair separations between 100 kpc and 1000 kpc, using simple galaxy bias correction schemes. Simulated photometric samples were constructed using the DNF photometric redshift code (Euclid Collaboration 2021b). The redshift determinations were constructed from two sets of training samples, one that is fully representative in redshift and magnitude and a second that has a completeness drop off in $I_{\rm E}$ magnitudes similar to surveys such as *Rubin* LSST (denoted with '*Rubin*').

The clustering redshift distributions were fitted with two models: The first modifies the true $n(z)$ with an amplitude and a shift parameter (the 'shifted-true' model). The second fits a 'suppressed-GP' model, taking advantage of the non-parametric fitting ability of GPs and suppressing low signal-to-noise and negative fluctuations with a S/N-dependent suppression function.

These two models were fitted to Flagship measurements over an area of 402 deg^2. By making measurements of the clustering redshifts on subregions of the Flagship footprint, we established power-law relations between the uncertainty on the mean redshift and the spectroscopic–photometric overlapping area. We used these relations to extrapolate the uncertainty to the full expected overlap area for BOSS, DESI, and *Euclid* (approximately 6000 deg^2) and showed that both models achieve uncertainties on the mean redshifts of less than $0.002\,(1+z)$ – well within the required uncertainties for the *Euclid* FoM on dark energy of >400.

However, systematic biases currently dominate the mean redshift determination and are independent of the redshift distribution model. Determining and mitigating these sources of systematic biases will be critical for the usage of cross-correlation redshifts by *Euclid*. The most difficult step for clustering redshift calibration is determining the galaxy bias of photometric samples. This is the primary source of systematic error and was shown to be the cause of systematic biases in the low-redshift bins. However, these biases persist in the high-redshift bins even when taking advantage of simulation information to correct exactly for photometric galaxy bias, meaning some other source of error is responsible.

Further studies should seek to characterise and mitigate these additional sources of systematic error, by exploring different scales, biasing models, or cross-correlation estimators. Photo-z are limited to $z < 1.6$ to ensure that the true redshifts of galaxies rarely exceed $z > 1.8$ in the tenth tomographic bin, since the analysis was limited to the redshift regime where spectroscopic tracers were defined (i.e. $z < 1.8$). In a future analysis, quasar large-scale structure tracers from BOSS, eBOSS, and DESI should allow for a measurement of the $n(z)$ for high-redshift tomographic bins, though the sparsity of quasars is likely to pose challenges. As simulations of quasar samples are more difficult to implement, since they are highly biased tracers with more complex selection functions, we have not included them in this analysis. Future studies should consider creating such samples so that the biases and uncertainties on these higher-redshift bins can be better determined.

This study is based on idealised assumptions where target and tracer samples suffer no observational complications, such as the spatial incompleteness of spectroscopic samples, tracer–target density correlations with Galactic foregrounds, or systematic depth variations. As such, future studies, and *Euclid* photo-z mocks in particular, should attempt to model these sources of

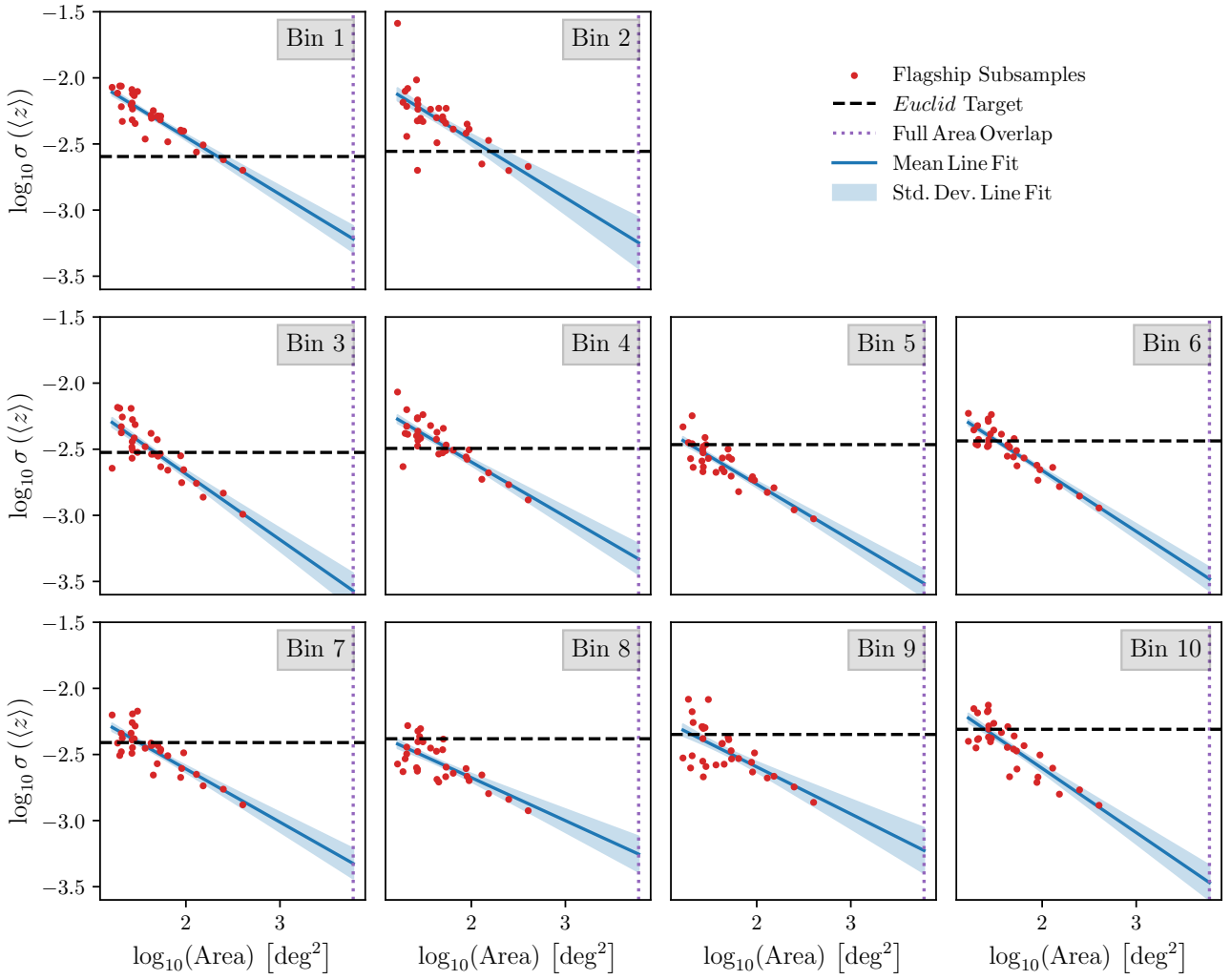

Fig. 8. Uncertainty on the mean redshift for each *Euclid* tomographic redshift bin shown as a function of tracer–target overlap area. The red points show the uncertainty on the mean redshift for each subregion, determined by fitting the shifted-true model (Sect. 3.3.1) to measured clustering redshifts (bias correction method 2, Sect. 3.1), with the largest area points representing constraints from the full 402 deg^2 Flagship footprint. These points are used to fit a power-law relationship between overlap area and mean redshift uncertainty, shown in blue (the line represents the mean fit and the envelope the uncertainty). Extrapolated to the full area overlap of BOSS, DESI, and *Euclid* (the dotted purple line), the projected uncertainties on the mean are shown to be much smaller than the required uncertainties for *Euclid*, itself indicated by the dashed black line.

systematic error. Furthermore, the photometric catalogues used in this study assume supplementary photometry from a *Rubin*-like survey; in reality, only the southern sky will be supplemented with *Rubin*, and the northern sky will use photometry from an array of imaging surveys such as CFHT, Pan-STARRS, and J-PAS. This is likely to result in systematic differences in the quality of photo-z based on positions on the sky. To study these effects, realistic *Euclid* mock catalogues that attempt to simulate these systematic errors across the full sky are required.

Acknowledgements. K. Naidoo acknowledges support from the Science and Technology Facilities Council grant ST/N50449X and from the (Polish) National Science Centre grant #2018/31/G/ST9/03388. H. Johnston acknowledges support from the Delta ITP consortium, a program of the Netherlands Organisation for Scientific Research (NWO) that is funded by the Dutch Ministry of Education, Culture and Science (OCW). O. Lahav acknowledges support from an STFC Consolidated Grant ST/R000476/1. K. Naidoo, H. Johnston, B. Joachimi, and O. Lahav were supported by the UK Space Agency. H. Hildebrandt is supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1) as well as an ERC Consolidator Grant (No. 770935). J. L. van den Busch acknowledges support from the European Research Council under grant numbers 770935. 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 French Centre National d'Etudes Spatiales, the Deutsches Zentrum für Luft- und Raumfahrt, the Danish Space Research Institute, the Fundação para a Ciência e a Tecnologia, the Ministerio de Ciencia e Innovación, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Netherlandse 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

Alarcon, A., Sánchez, C., Bernstein, G. M., & Gaztañaga, E. 2020, MNRAS, 498, 2614
Amiaux, J., Scaramella, R., Mellier, Y., et al. 2012, in Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, eds. M. C. Clampin, G. G. Fazio, H. A. MacEwen, J. Oschmann, & M. Jacobus, SPIE Conf. Ser., 8442, 84420Z
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109
Cawthon, R., Davis, C., Gatti, M., et al. 2018, MNRAS, 481, 2427
Cawthon, R., Elvin-Poole, J., Porredon, A., et al. 2022, MNRAS, 513, 5517

Davis, M., & Peebles, P. J. E. 1983, ApJ, 267, 465
Davis, C., Rozo, E., Roodman, A., et al. 2018, MNRAS, 477, 2196
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44
DESI Collaboration (Aghamousa, A., et al.) 2016, ArXiv e-prints [arXiv:1611.00037]
De Vicente, J., Sánchez, E., & Sevilla-Noarbe, I. 2016, MNRAS, 459, 3078
Euclid Collaboration (Blanchard, A., et al.) 2020a, A&A, 642, A191
Euclid Collaboration (Desprez, G., et al.) 2020b, A&A, 644, A31
Euclid Collaboration (Ilbert, O., et al.) 2021a, A&A, 647, A117
Euclid Collaboration (Pocino, A., et al.) 2021b, A&A, 655, A44
Euclid Collaboration (Scaramella, R., et al.) 2022a, A&A, 662, A112
Euclid Collaboration (Schirmer, M., et al.) 2022b, A&A, 662, A92
Gatti, M., Vielzeuf, P., Davis, C., et al. 2018, MNRAS, 477, 1664
Gatti, M., Giannini, G., Bernstein, G. M., et al. 2022, MNRAS, 510, 1223
Hadzhiyska, B., Alonso, D., Nicola, A., & Slosar, A. 2020, JCAP, 2020, 056
Hartley, W. G., Chang, C., Samani, S., et al. 2020, MNRAS, 496, 4769
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69
Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592
Huterer, D., Takada, M., Bernstein, G., & Jain, B. 2006, MNRAS, 366, 101
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118
Limber, D. N. 1953, ApJ, 117, 134
LSST Science Collaboration (Abell, P. A., et al.) 2009, ArXiv e-prints [arXiv:0912.0201]
McQuinn, M., & White, M. 2013, MNRAS, 433, 2857
Ménard, B., Scranton, R., Schmidt, S., et al. 2013, ArXiv e-prints [arXiv:1303.4722]
Morrison, C. B., Hildebrandt, H., Schmidt, S. J., et al. 2017, MNRAS, 467, 3576
Myles, J., Alarcon, A., Amon, A., et al. 2021, MNRAS, 505, 4249
Newman, J. A. 2008, ApJ, 684, 88
Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015, Astropart. Phys., 63, 81
Potter, D., Stadel, J., & Teyssier, R. 2017, Comput. Astrophys. Cosmol., 4, 2
Sánchez, C., Carrasco Kind, M., Lin, H., et al. 2014, MNRAS, 445, 1482
Schmidt, S. J., Ménard, B., Scranton, R., Morrison, C., & McBride, C. K. 2013, MNRAS, 431, 3307
Scottez, V., Mellier, Y., Granett, B. R., et al. 2016, MNRAS, 462, 1683
Scottez, V., Benoit-Lévy, A., Coupon, J., Ilbert, O., & Mellier, Y. 2018, MNRAS, 474, 3921
Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, PASJ, 70, S9
van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, A&A, 642, A200
Wright, A. H., Hildebrandt, H., van den Busch, J. L., & Heymans, C. 2020, A&A, 637, A100
Yèche, C., Palanque-Delabrouille, N., Claveau, C.-A., et al. 2020, Res. Notes Am. Astron. Soc., 4, 179

[1] Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668 Warsaw, Poland
[2] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
e-mail: krishna.naidoo.11@ucl.ac.uk
[3] Institute for Theoretical Physics, Utrecht University, Princetonplein 5, 3584 CE Utrecht, The Netherlands
[4] Ruhr-Universität Bochum, Astronomisches Institut, German Centre for Cosmological Lensing (GCCL), Universitätsstr. 150, 44801 Bochum, Germany
[5] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
[6] Aix-Marseille Univ, CNRS, CNES, LAM, Marseille, France
[7] California institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA
[8] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[9] Université Paris-Saclay, CNRS, Institut d'astrophysique spatiale, 91405, Orsay, France
[10] ESAC/ESA, Camino Bajo del Castillo, s/n., Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
[11] Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
[12] Dipartimento di Fisica e Astronomia, Universitá di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
[13] INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
[14] INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[15] Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
[16] Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
[17] Dipartimento di Fisica, Universitá degli studi di Genova, and INFN-Sezione di Genova, via Dodecaneso 33, 16146 Genova, Italy
[18] INFN-Sezione di Roma Tre, Via della Vasca Navale 84, 00146 Roma, Italy
[19] Department of Physics "E. Pancini", University Federico II, Via Cinthia 6, 80126 Napoli, Italy
[20] INAF-Osservatorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli, Italy
[21] Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal
[22] Dipartimento di Fisica, Universitá degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
[23] INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
[24] INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese, TO, Italy
[25] INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
[26] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra, Barcelona, Spain
[27] Port d'Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra, Barcelona, Spain
[28] Institut d'Estudis Espacials de Catalunya (IEEC), Carrer Gran Capitá 2-4, 08034 Barcelona, Spain
[29] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
[30] INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
[31] INFN section of Naples, Via Cinthia 6, 80126 Napoli, Italy
[32] Dipartimento di Fisica e Astronomia "Augusto Righi" – Alma Mater Studiorum Universitá di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[33] INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy
[34] Institut national de physique nucléaire et de physique des particules, 3 rue Michel-Ange, 75794 Paris Cédex 16, France
[35] Centre National d'Etudes Spatiales, Toulouse, France
[36] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[37] Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
[38] European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
[39] Univ Lyon, Univ Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, 69622 Villeurbanne, France
[40] Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
[41] Mullard Space Science Laboratory, University College London, Dorking, Surrey RH5 6NT, UK
[42] Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, 1749-016 Lisboa, Portugal
[43] Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
[44] Department of Astronomy, University of Geneva, ch. dÉcogia 16, 1290 Versoix, Switzerland
[45] INFN-Padova, Via Marzolo 8, 35131 Padova, Italy

[46] Université Paris-Saclay, Université Paris Cité, CEA, CNRS, Astrophysique, Instrumentation et Modélisation Paris-Saclay, 91191 Gif-sur-Yvette, France
[47] INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
[48] Aix-Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France
[49] Istituto Nazionale di Astrofisica (INAF) – Osservatorio di Astrofisica e Scienza dello Spazio (OAS), Via Gobetti 93/3, 40127 Bologna, Italy
[50] Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Via Irnerio 46, 40126 Bologna, Italy
[51] INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, 35122 Padova, Italy
[52] Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029, Blindern 0315, Oslo, Norway
[53] von Hoerner & Sulger GmbH, SchloßPlatz 8, 68723 Schwetzingen, Germany
[54] Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
[55] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[56] Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Finland
[57] NOVA optical infrared instrumentation group at ASTRON, Oude Hoogeveensedijk 4, 7991PD Dwingeloo, The Netherlands
[58] Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
[59] School of Physics and Astronomy, Faculty of Science, Monash University, Clayton, Victoria 3800, Australia
[60] Dipartimento di Fisica e Astronomia "Augusto Righi" – Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy
[61] Department of Physics, Institute for Computational Cosmology, Durham University, South Road, DH1 3LE, UK
[62] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France
[63] INFN-Bologna, Via Irnerio 46, 40126 Bologna, Italy
[64] European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
[65] Department of Physics and Astronomy, University of Aarhus, Ny Munkegade 120, 8000 Aarhus C, Denmark
[66] Centre for Astrophysics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
[67] Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
[68] Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
[69] AIM, CEA, CNRS, Université Paris-Saclay, Université de Paris, 91191 Gif-sur-Yvette, France
[70] Space Science Data Center, Italian Space Agency, via del Politecnico snc, 00133 Roma, Italy
[71] Institute of Space Science, Bucharest 077125, Romania
[72] Instituto de Astrofísica de Canarias, Calle Vía Láctea s/n, 38204 San Cristóbal de La Laguna, Tenerife, Spain
[73] Departamento de Astrofísica, Universidad de La Laguna, 38206 La Laguna, Tenerife, Spain
[74] Dipartimento di Fisica e Astronomia "G.Galilei", Universitá di Padova, Via Marzolo 8, 35131 Padova, Italy
[75] Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
[76] Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile
[77] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
[78] Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
[79] Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, 30202 Cartagena, Spain
[80] Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, 1211 Genève 4, Switzerland
[81] Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
[82] Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
[83] IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, 34151 Trieste, Italy
[84] INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
[85] Dipartimento di Fisica "Aldo Pontremoli", Universitá degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy
[86] INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
[87] Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014 Paris, France
[88] Junia, EPA department, 59000 Lille, France