

# ***Euclid*: Photometric redshift calibration performance with the clustering-redshifts technique in the Flagship 2 simulation<sup>★</sup>**

W. d'Assignies<sup>1,★,✉</sup>, M. Manera<sup>2,1</sup>, C. Padilla<sup>1</sup>, O. Ilbert<sup>3</sup>, H. Hildebrandt<sup>4</sup>, L. Reynolds<sup>1,5</sup>, J. Chaves-Montero<sup>1</sup>, A. H. Wright<sup>4</sup>, P. Tallada-Crespi<sup>6,7</sup>, M. Eriksen<sup>1,7</sup>, J. Carretero<sup>6,7</sup>, W. Roster<sup>8</sup>, Y. Kang<sup>9</sup>, K. Naidoo<sup>10</sup>, R. Miquel<sup>1,11</sup>, B. Altieri<sup>12</sup>, A. Amara<sup>13</sup>, S. Andreon<sup>14</sup>, N. Auricchio<sup>15</sup>, C. Baccigalupi<sup>16,17,18,19</sup>, D. Bagot<sup>20</sup>, M. Baldi<sup>21,15,22</sup>, A. Balestra<sup>23</sup>, S. Bardelli<sup>15</sup>, P. Battaglia<sup>15</sup>, A. Biviano<sup>17,16</sup>, E. Branchini<sup>24,25,14</sup>, M. Brescia<sup>26,27</sup>, S. Camera<sup>28,29,30</sup>, V. Capobianco<sup>30</sup>, C. Carbone<sup>31</sup>, V. F. Cardone<sup>32,33</sup>, S. Casas<sup>34</sup>, F. J. Castander<sup>35,36</sup>, M. Castellano<sup>32</sup>, G. Castignani<sup>15</sup>, S. Cavuoti<sup>27,37</sup>, K. C. Chambers<sup>38</sup>, A. Cimatti<sup>39</sup>, C. Colodro-Conde<sup>40</sup>, G. Congedo<sup>41</sup>, C. J. Conselice<sup>42</sup>, L. Conversi<sup>43,12</sup>, Y. Copin<sup>44</sup>, F. Courbin<sup>45,11</sup>, H. M. Courtois<sup>46</sup>, M. Crocce<sup>35,36</sup>, A. Da Silva<sup>47,48</sup>, H. Degaudenzi<sup>9</sup>, S. de la Torre<sup>3</sup>, G. De Lucia<sup>17</sup>, M. Douspis<sup>49</sup>, X. Dupac<sup>12</sup>, A. Ealet<sup>44</sup>, S. Escoffier<sup>50</sup>, M. Farina<sup>51</sup>, F. Faustini<sup>32,52</sup>, S. Ferriol<sup>44</sup>, F. Finelli<sup>15,53</sup>, P. Fosalba<sup>36,35</sup>, S. Fotopoulou<sup>54</sup>, M. Frailis<sup>17</sup>, E. Franceschi<sup>15</sup>, M. Fumana<sup>31</sup>, S. Galeotta<sup>17</sup>, K. George<sup>55</sup>, B. Gillis<sup>41</sup>, C. Giocoli<sup>15,22</sup>, P. Gómez-Alvarez<sup>56,12</sup>, J. Gracia-Carpio<sup>8</sup>, A. Grazian<sup>23</sup>, F. Grupp<sup>8,55</sup>, W. Holmes<sup>57</sup>, I. M. Hook<sup>58</sup>, A. Hornstrup<sup>59,60</sup>, K. Jahnke<sup>61</sup>, M. Jhabvala<sup>62</sup>, B. Joachimi<sup>63</sup>, E. Keihänen<sup>64</sup>, S. Kermiche<sup>50</sup>, A. Kiessling<sup>57</sup>, B. Kubik<sup>44</sup>, M. Kümmel<sup>55</sup>, M. Kunz<sup>65</sup>, H. Kurki-Suonio<sup>66,67</sup>, O. Lahav<sup>63</sup>, A. M. C. Le Brun<sup>68</sup>, S. Ligi<sup>30</sup>, P. B. Lilje<sup>69</sup>, V. Lindholm<sup>66,67</sup>, I. Lloro<sup>70</sup>, G. Mainetti<sup>71</sup>, D. Maino<sup>72,31,73</sup>, E. Maiorano<sup>15</sup>, O. Mansutti<sup>17</sup>, S. Marcin<sup>74</sup>, O. Marggraf<sup>75</sup>, K. Markovic<sup>57</sup>, M. Martinelli<sup>32,33</sup>, N. Martinet<sup>3</sup>, F. Marulli<sup>76,15,22</sup>, R. Massey<sup>77</sup>, D. C. Masters<sup>78</sup>, E. Medinaceli<sup>15</sup>, S. Mei<sup>79,80</sup>, M. Melchior<sup>81</sup>, Y. Mellier<sup>82,83</sup>, M. Meneghetti<sup>15,22</sup>, E. Merlin<sup>32</sup>, G. Meylan<sup>84</sup>, A. Mora<sup>85</sup>, M. Moresco<sup>76,15</sup>, L. Moscardini<sup>76,15,22</sup>, C. Neissner<sup>1,7</sup>, S.-M. Niemi<sup>86</sup>, S. Paltani<sup>9</sup>, F. Pasian<sup>17</sup>, K. Pedersen<sup>87</sup>, V. Pettorino<sup>86</sup>, S. Pires<sup>88</sup>, G. Polenta<sup>52</sup>, M. Poncet<sup>20</sup>, L. A. Popa<sup>89</sup>, L. Pozzetti<sup>15</sup>, F. Raison<sup>8</sup>, R. Rebolo<sup>40,90,91</sup>, A. Renzi<sup>92,93</sup>, J. Rhodes<sup>57</sup>, G. Riccio<sup>27</sup>, E. Romelli<sup>17</sup>, M. Roncarelli<sup>15</sup>, E. Rossetti<sup>21</sup>, R. Saglia<sup>55,8</sup>, Z. Sakr<sup>94,95,96</sup>, D. Sapone<sup>97</sup>, B. Sartoris<sup>55,17</sup>, J. A. Schewtschenko<sup>41</sup>, P. Schneider<sup>75</sup>, T. Schrabback<sup>98</sup>, A. Secroun<sup>50</sup>, E. Sefusatti<sup>17,16,18</sup>, G. Seidel<sup>61</sup>, M. Seiffert<sup>57</sup>, S. Serrano<sup>36,99,35</sup>, P. Simon<sup>75</sup>, C. Sirignano<sup>92,93</sup>, G. Sirri<sup>22</sup>, A. Spurio Mancini<sup>100</sup>, L. Stanco<sup>93</sup>, J. Steinwagner<sup>8</sup>, D. Tavagnacco<sup>17</sup>, A. N. Taylor<sup>41</sup>, H. I. Teplitz<sup>78</sup>, I. Tereno<sup>47,101</sup>, N. Tessore<sup>63</sup>, S. Toft<sup>102,103</sup>, R. Toledo-Moreo<sup>104</sup>, F. Torradeflot<sup>7,6</sup>, A. Tsyganov<sup>105</sup>, I. Tutusaus<sup>95</sup>, L. Valenziano<sup>15,53</sup>, J. Valiviita<sup>66,67</sup>, T. Vassallo<sup>55,17</sup>, G. Verdoes Kleijn<sup>106</sup>, Y. Wang<sup>78</sup>, J. Weller<sup>55,8</sup>, G. Zamorani<sup>15</sup>, E. Zucca<sup>15</sup>, M. Bolzonella<sup>15</sup>, C. Burigana<sup>107,53</sup>, L. Gabarra<sup>108</sup>, J. Martín-Fleitas<sup>109</sup>, I. Risso<sup>110</sup>, V. Scottez<sup>82,111</sup>, and M. Viel<sup>16,17,19,18,112</sup>

(Affiliations can be found after the references)

Received 16 May 2025 / Accepted 28 July 2025

## **ABSTRACT**

**Aims.** The precision of cosmological constraints from imaging surveys hinges on an accurately estimated redshift distribution  $n(z)$  of the tomographic bins, especially their mean redshifts. We assess the effectiveness of the clustering-redshifts technique in constraining *Euclid* tomographic redshift bins to meet the target uncertainty of  $\sigma(z) < 0.002(1+z)$ . We inferred these mean redshifts from the small-scale angular clustering of *Euclid* galaxies, which were distributed into bins with spectroscopic samples localised in narrow redshift slices.

**Methods.** We generated spectroscopic mocks from the Flagship2 simulation for the Baryon Oscillation Spectroscopic Survey (BOSS), the Dark Energy Spectroscopic Instrument (DESI), and the *Euclid* Near-Infrared Spectrometer and Photometer (NISIP) spectroscopic survey. We evaluated and optimised the clustering-redshifts pipeline, and we introduced a new method for measuring the photometric galaxy bias (clustering), which is the primary limitation of this technique.

**Results.** We have successfully constrained the means and standard deviations of the redshift distributions for all of the tomographic bins (with a maximum photometric redshift of 1.6). We achieved precision beyond the required thresholds. We have identified the main sources of bias, particularly the impact of the one-halo galaxy distribution, which imposed the minimal separation scale to be larger than 1.5 Mpc for evaluating cross-correlations. These results demonstrate that clustering-redshifts can meet the precision requirements for *Euclid*, and we highlighted several avenues for future improvements.

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

<sup>★</sup> This paper is published on behalf of the Euclid Consortium.

<sup>\*\*</sup> Corresponding author: [wdoumerg@iaf.ee](mailto:wdoumerg@iaf.ee)

## 1. Introduction

The *Euclid* space telescope is currently mapping the positions and shapes of billions of galaxies. This provides data that are critical to understand the large-scale structure of the Universe, in particular, the mysterious nature of dark matter and dark energy (Euclid Collaboration: Mellier et al. 2025). Through its imaging survey, *Euclid* measures the flux of galaxies in the visible and near-infrared wavelengths using broadband filters. Additionally, it determines spectroscopic redshifts for a subset of galaxies via slitless spectroscopy, targeting emission-line galaxies (ELGs). These data will enable the precise determination of cosmological parameters and models, primarily through the study of galaxy clustering, galaxy-galaxy lensing, and cosmic shear (Euclid Collaboration: Blanchard et al. 2020), with galaxy samples divided into approximate line-of-sight tomographic bins.

It is essential to accurately estimate the redshift distributions of these galaxy samples to interpret cosmological measurements correctly (see e.g. Huterer et al. 2006; Bordoloi et al. 2010; Hildebrandt et al. 2021; Stölzner et al. 2021). The vast amount of forthcoming data means that it is not feasible to obtain spectroscopic redshifts for every individual galaxy because spectroscopy of large samples is time-consuming and costly. Photometric surveys provide redshift estimates for each galaxy based on multi-band photometry of that galaxy, a technique called photometric redshift, or photo- $z$ . A large variety of photo- $z$  methods exists (e.g. Tanaka et al. 2018; Salvato et al. 2019; Euclid Collaboration: Ilbert et al. 2021; Euclid Collaboration: Desprez et al. 2020). The photometric information can also be used to produce redshift estimates using scheme based on a self-organising map (hereafter, SOM), which allows a more general control over all the known potential sources of uncertainties that affect the estimates (Wright et al. 2020a,b; Myles et al. 2021; Giannini et al. 2024; Campos et al. 2024; Wright et al. 2025; Roster et al. 2025). The degeneracies between colours and redshift, however, and unrepresentative spectroscopic samples for training and calibration ultimately limit the performance of photometric methods (Wright et al. 2020a; Hartley et al. 2020).

Clustering-based redshift estimation methods offer another alternative to infer redshift distributions (see e.g. Newman 2008; Ménard et al. 2013; McQuinn & White 2013; Morrison et al. 2017; Scottez et al. 2018; Gatti et al. 2018; van den Busch et al. 2020; Hildebrandt et al. 2021; Gatti et al. 2022; Cawthon et al. 2022; Rau et al. 2023; Wright et al. 2025). Unlike photo- $z$ , clustering-redshifts techniques statistically infer the redshift properties for entire bins instead of individual galaxies. Furthermore, this approach is independent of photometric uncertainties, observations over smaller deep fields, or representative spectroscopic samples.

The clustering-redshifts method relies on angular cross-correlations between spectroscopic samples, with secure redshifts, and photometric samples. Samples that overlap in redshift trace the same underlying dark matter field and are consequently correlated in their position. The amplitude of this angular correlation provides insight into their redshift overlap, which results in the measurement of the redshift distribution of the photometric samples. This amplitude is degenerate with galaxy biases (of both samples), however, which makes the measurement of these biases, and their redshift evolution, a critical step (McQuinn & White 2013; Gatti et al. 2018; Naidoo et al. 2023). While the galaxy bias for the spectroscopic sample can be directly constrained through auto-correlation functions, it is more challenging to estimate the galaxy bias for the photometric

sample (van den Busch et al. 2020; Cawthon et al. 2022), and is sometimes evaluated using simulations (Gatti et al. 2022).

The precision of the clustering-redshifts method depends on the redshift and sky overlap of the photometric and spectroscopic data, but also on range of the clustering angular scale. To achieve the primary science goal of *Euclid*, the uncertainty on the mean redshift,  $\sigma(\langle z \rangle)$ , for each tomographic bin must remain below  $0.002(1+z)$  at 68% confidence (Laureijs et al. 2011). Naidoo et al. (2023) found that clustering-redshifts would be able to reach the statistical uncertainties required by *Euclid* for a sufficiently large sky overlap, typically several hundreds of  $\text{deg}^2$ , when analysing scales from 100 kpc to 1 Mpc, although systematic biases limit the accuracy. In particular, there were some unknown residual biases for high-redshift bins with their methods or simulated data.

This paper aims to re-evaluate the potential of photometric-spectroscopic cross-correlations as a core component of the redshift calibration for *Euclid*. Our method may be seen as a continuation, and a substantial improvement from Naidoo et al. (2023). We used 5000  $\text{deg}^2$  of simulated data from the *Euclid* Flagship2 simulation (Euclid Collaboration: Castander et al. 2025). Our goal is to evaluate the uncertainties associated with the clustering-redshifts calibration through a realistic framework, and to determine the limitation of the method. We created spectroscopic mock samples for the Dark Energy Spectroscopic Instruments (DESI), or alternatively, for the 4-metre Multi-Object Spectroscopic Telescope (4MOST), the Baryon Oscillation Spectroscopic Survey (BOSS), and *Euclid*. We assessed the potential systematic effects on the measurement, introduced a new method to measure photometric galaxy bias, optimised the clustering- $z$  pipeline, and derived methods for inferring the redshift constraints from the amplitude of two-point correlations. We assumed the same flat  $\Lambda$ CDM cosmology as was used for the simulation, with  $\Omega_m = 0.319$ ,  $\Omega_b = 0.049$ ,  $\Omega_\Lambda = 0.681$ ,  $A_s = 2.1 \times 10^{-9}$ ,  $n_s = 0.96$ ,  $\sigma_8 = 0.813$ , and  $h = 0.67$ . Our code is publicly available<sup>1</sup>.

This paper is organised as follows. In Sect. 2 we describe the Flagship2 simulation and the mocks we generated to mimic *Euclid* and spectroscopic data. In Sect. 3 we give a detailed overview of angular clustering, its application to clustering redshifts, and potential sources of errors. In Sect. 4 we present the different tests we ran to optimise the pipeline and address potential sources of bias. Finally in Sect. 5 we apply the results pipeline to realistic *Euclid* tomographic bins and compare our results with previous work and the requirements.

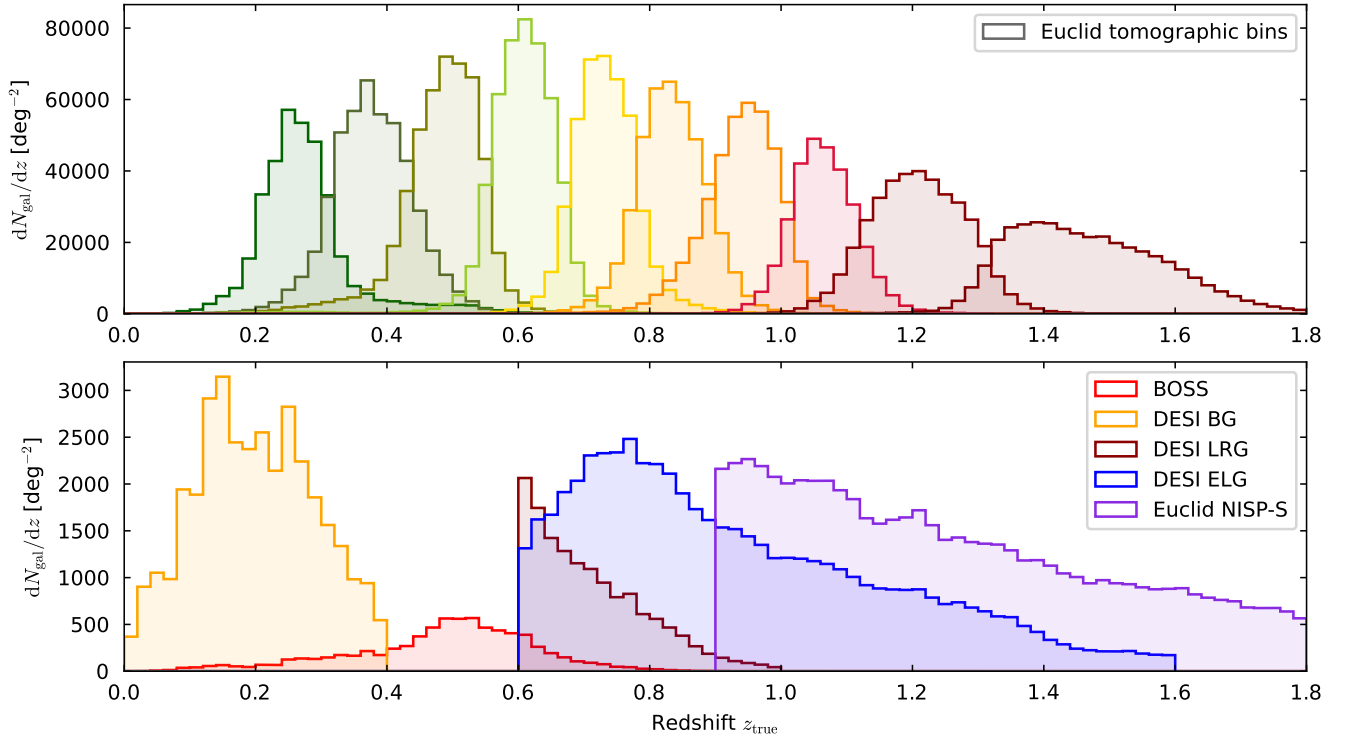
## 2. Simulated data

To assess the performance of the *Euclid* clustering-redshifts method, we constructed simulated datasets that were as similar as possible to the observational datasets. In this section, we describe how these samples can be generated from the Flagship simulation.

### 2.1. The Flagship simulation and survey mocks

For this study, we used the second edition of the Flagship catalogue (version 2.1.10), which was presented in detail in Euclid Collaboration: Castander et al. (2025). The *Euclid* Flagship  $N$ -body dark matter simulation comprises a simulation box measuring 3600  $h^{-1}$  Mpc on each side, containing 4 trillion particles. This simulation is the largest  $N$ -body simulation

<sup>1</sup> [https://github.com/wdassignies/Clustering\\_z.git](https://github.com/wdassignies/Clustering_z.git)



**Fig. 1.** *Top:* Simulated true-redshift distribution of each of the first ten tomographic bins ( $z_{\text{photo}} < 1.6$ ). *Bottom:* Simulated redshift distribution of the spectroscopic samples from the BOSS, DESI, and *Euclid* surveys. The lack of spectroscopic samples for  $z_{\text{spec}} > 1.8$  imposes the  $z_{\text{photo}} < 1.6$  condition for clustering-redshifts calibration.

performed to date and encompasses a cosmological volume comparable to what the telescope will survey. It enables precise resolution of 16 billion dark matter haloes, which host even the faintest galaxies that *Euclid* intends to observe. They are identified with the ROCKSTAR halo finder (Behroozi et al. 2013). From this dark matter simulation, a synthetic galaxy catalogue was generated with a prescription including the halo occupation distribution (HOD) and abundance matching (AM) techniques, as well as empirical relations between galaxy properties (Euclid Collaboration: Castander et al. 2025).

Each galaxy entry in the catalogue is associated with observed fluxes for multiple survey bands, observed spectroscopic redshift ( $z_{\text{spec}}$ , including peculiar velocity), photometric redshift estimates ( $z_p$ ), true redshift ( $z_{\text{true}}$ ), true and magnified coordinates, lensing convergence  $\kappa$  and shear  $\gamma = \gamma_1 + i\gamma_2$ . Source photometry is provided in terms of fluxes  $F$  (in  $\text{erg s}^{-1} \text{cm}^{-2} \text{Hz}^{-1}$ ) rather than in magnitude  $m$ . We generate magnified fluxes as in Lepori et al. (2025), using the relation

$$F_{\text{magn}} = \frac{F}{(1 - \kappa)^2 - \gamma_1^2 - \gamma_2^2}. \quad (1)$$

We do not include the Doppler effect, which is usually a very small correction, and note that the effect of magnification is achromatic: that is, it is the same for the different bands and it cancels for colours.

We need mock samples *Euclid*-like photometric galaxies, *Euclid*-like Near-Infrared Spectrometer and Photometer Survey (NISP-S) galaxies, and other spectroscopic tracers (Dawson et al. 2013; Adame et al. 2024): BOSS-like low-redshift (LOWZ) and constant mass (CMASS) galaxies; DESI-like bright galaxies (BGs), luminous red galaxies (LRGs), and ELGs. The selection of each sample is described in the following sections. Differences in the redshift distributions or colours

of observational and simulated galaxy samples may stem from the incompleteness of the target sample or limitations in the simulation's ability to replicate accurate colour distributions. As in Naidoo et al. (2023), systematic issues such as variable photometry, variable survey depth, and complex masks are not addressed. We generated one version of these samples with unmagnified fluxes, which will be our default samples, and one version with the same cuts on magnified photometry. We produced numerous fast generations of galaxy catalogues, varying sample definitions and areas, using the web application CosmoHub (Carretero et al. 2017; Tallada et al. 2020). Our photometric and spectroscopic samples are plotted in Fig. 1.

## 2.2. *Euclid* photometric sample

The photometric sample was constructed by imposing an upper limit on the  $I_E$  magnitude (Euclid Collaboration: Cropper et al. 2025), set at 24.5, to emulate the expected performance of *Euclid* (Laureijs et al. 2011). Several algorithms exist for evaluating the photometric redshift. We used the Deep- $z$  estimate which was run for the full catalogue. The Deep- $z$  photo- $z$  method, introduced for the Physics of the Accelerating Universe (PAU) survey in Eriksen et al. (2020), uses a linear neural network with a mixture density network for predicting the redshift distributions. The network has not been pretrained on simulations, avoiding potential problems with over-fitting by using too similar simulations. The network was trained with 20 000 simulated galaxies, magnitude-limited to  $i_{\text{LSST}} < 23$  without colour selection. The training used the Legacy Survey of Space and Time (LSST; LSST Science Collaboration 2009) bands plus *Euclid* near-infrared (NIR) bands, a batch size of 200 and 1000 epochs with varying learning rates. The photometric noise added to noiseless fluxes corresponds to the depth forecast in the southern

hemisphere for the *Euclid* DR3 time. The limiting magnitudes at  $10\sigma$  are 24.4, 25.6, 25.7, 25.0, 24.3, 24.3 for the *ugrizy* bands of LSST, and 25.0, 23.5, 23.5, 23.5 for the  $I_E$ ,  $Y_E$ ,  $J_E$ ,  $H_E$  bands of *Euclid* (*Euclid* Collaboration: Cropper et al. 2025; *Euclid* Collaboration: Jahnke et al. 2025).

LSST photometry will not be available over the whole *Euclid* footprint, and therefore the photometry will be different in the northern region where photometry will be provided by UNIONS (Gwyn et al. 2025). However, for clustering redshifts, we do not depend on photo- $z$  quality except for the tomographic bin definitions and photometric bias correction. Spectroscopy from DESI and BOSS-eBOSS will be available in the north, and 4MOST will provide DESI-like samples in the south. Thus, photometric angular anisotropies will be tested in practice.

We selected galaxies with photo- $z$  between 0.2 and 1.6. We did not consider galaxies with a photo- $z$  higher than 1.6 because we do not have large spectroscopic samples at redshifts greater than 1.8. Future studies could test the possibility of using quasar galaxy samples from BOSS, eBOSS, DESI, and 4MOST (Palanque-Delabrouille et al. 2016; Yèche et al. 2020) for high- $z$  calibration. BOSS and eBOSS quasars were used for the clustering-redshifts DES calibration, but as a complement to ELGs and LRGs data sets for  $z < 1.1$  (Gatti et al. 2022; Giannini et al. 2024). The low density of this sample, and clustering properties at high- $z$ , require a dedicated study. Nonetheless, eBOSS QSOs will be used for the high- $z$  calibration of the final DES data for  $0.8 < z < 2.2$  (d'Assignies et al., in prep.); thus one expects that DESI QSOs could be used up to  $z = 3.5$  for *Euclid*. Generating DESI QSO mocks and testing clustering-redshifts with them is left for future work.

We produced ten tomographic bins with approximately equal numbers of galaxies, through photo- $z$  cuts. The fiducial *Euclid* survey at data release 3 plans to use 13 bins instead of ten, and we can assume that three of these bins will contain galaxies with  $z_p > 1.6$ . The true redshift distributions (that we want to measure) of our simulated tomographic bins are reported in Fig. 1.

### 2.3. *Euclid* NISP sample

We defined mock NISP-S samples in Flagship2 by selecting galaxies with  $H\alpha$  fluxes greater than  $2 \times 10^{-16} \text{erg cm}^{-2} \text{s}^{-1}$  (Laureijs et al. 2011), in the expected redshift range, with

$$\log f_{\text{H}\alpha} \text{ model3\_ext} > -15.7, \quad (2)$$

$$0.9 < z_{\text{spec}} < 1.8. \quad (3)$$

We obtained a galaxy density of  $1990 \text{ deg}^{-2}$ . Unlike the other spectroscopic samples, this dataset is expected to suffer from a significant level of contamination by interlopers (*Euclid* Collaboration: Le Brun et al. 2025, Risso et al., in prep.). These interlopers are galaxies whose true redshifts differ substantially from their estimated values, leading to incorrect associations in redshift space. Although our analysis did not employ realistic mocks that include such contaminants, we adopted the approach of Contarini et al. (2022): we uniformly down-sampled the galaxy catalogue, retaining only 60% of the originally selected galaxies. This down-sampled catalogue is then assumed to have 100% purity, effectively removing the impact of interlopers in our modelling. Thus, our fiducial density is  $1200 \text{ deg}^{-2}$  as the sample from Naidoo et al. (2023), which was generated with a brighter cut on the  $H\alpha$  flux. This is a pessimistic sample from the precision point of view (i.e. fewer spectra means larger uncertainty on the redshift distribution reconstruction; cf. McQuinn & White 2013), but opti-

mistic in terms of accuracy, because our sample is interloper free (Addison et al. 2019). Interlopers may significantly degrade the performance of clustering-redshifts if not taken into consideration, as they are not necessarily distributed randomly in redshift. Indeed, clustering-redshifts might be helpful to quantify the interlopers' properties such as their fraction and redshift distribution, by cross-correlating them with spectroscopic samples which do not suffer from the same misidentification. We will explore these two aspects in future work.

### 2.4. BOSS LOWZ and CMASS samples

We replicated the BOSS colour-magnitude selections specified in Dawson et al. (2013) in Flagship, as in Naidoo et al. (2023). We first define

$$c_{\parallel} = 0.7(g - r) + 1.2(r - i - 0.18), \quad (4)$$

$$c_{\perp} = (r - i) - (g - r)/4 - 0.18, \quad (5)$$

$$d_{\perp} = (r - i) - (g - r)/8. \quad (6)$$

There are two spectroscopic BOSS LRG samples: LOWZ and CMASS. They correspond to adjacent redshift intervals  $0.15 \leq z_{\text{spec}} \leq 0.43$  and  $0.43 < z_{\text{spec}} \leq 0.7$ , respectively, with true densities of about  $30 \text{ deg}^{-2}$  and  $120 \text{ deg}^{-2}$ , respectively.

The LOWZ sample is defined by

$$|c_{\perp}| < 0.2, \quad (7)$$

$$16 < r < 19.5, \quad (8)$$

$$r < 13.6 + c_{\parallel}/0.3, \quad (9)$$

and the CMASS sample by

$$d_{\perp} > 0.55, \quad (10)$$

$$17.5 < i < 19.9, \quad (11)$$

$$i < 19.86 + 1.6(d_{\perp} - 0.8). \quad (12)$$

We got LOWZ- and CMASS- like objects, with densities of  $48.5 \text{ deg}^{-2}$  and  $164.3 \text{ deg}^{-2}$ . We then applied sparse sampling to get the desired densities. We refer to the joint LOWZ+CMASS sample as the BOSS sample in our study.

### 2.5. DESI samples

We generated DESI BG, LRG, and ELG mocks that mimic the sample characteristics reported in Adame et al. (2024). We associated these galaxy samples to the DESI survey, but similar samples are expected to be observed by 4MOST (Verdier et al. 2025). For BGs, and ELGs, we used the same cuts as Naidoo et al. (2023), which were based on DESI Collaboration (2016). For LRGs, we used a different selection to generate a sample covering the entire  $z < 1.1$  redshift range. As we previously mentioned in Sect. 2.1, there is a mismatch in the Flagship2 colour distribution and data expectation, and using the same colour cuts can lead to some significant differences. We find this issue to be particularly problematic for LRGs. Therefore we did not apply the same selection as Adame et al. (2024), but rather we tried to closely reproduce the redshift distribution and galaxy properties with similar but modified cuts.

We selected the BG-like galaxies with

$$r < 19.5, \quad (13)$$

$$z_{\text{spec}} < 0.4. \quad (14)$$

We adopted for LRGs the tailored cuts

$$18 < z < 21, \quad (15)$$

$$-6.5 < r - z - 0.4z < -5.8, \quad (16)$$

$$\text{color\_kind} = 0 \text{ (i.e. red sequence)}. \quad (17)$$

Finally, we selected ELGs with

$$r < 23.4, \quad (18)$$

$$r - z > 0.3, \quad (19)$$

$$g - r < 0.7, \quad (20)$$

$$0.6 < z_{\text{spec}} < 1.6, \quad (21)$$

$$\text{color\_kind} \neq 0, \quad (22)$$

$$\log f_{\text{o2\_model1\_ext}} > -16. \quad (23)$$

These cuts provided BG-, LRG-, and ELG-like objects with densities of 850, 909, and 4158  $\text{deg}^{-2}$ . We achieved the fiducial densities 700, 550, and 1140  $\text{deg}^{-2}$  with sparse sampling.

### 3. Theory and method

In this section, we provide a detailed description of the clustering-redshifts method. Our approach begins with an overview of galaxy clustering in Sect. 3.1. In Sect. 3.2 we explain how the clustering equations are simplified for some particular redshift distributions. We then apply this formalism to the clustering-redshifts context in Sect. 3.3. In Sect. 3.4, we explain how clustering-redshifts are evaluated with data in practice. Finally, in Sect. 3.5 we explain our strategy to deduce from the data-vectors constraints on the redshift distribution moments. In particular, in Sect. 3.5.1, we state the *Euclid* requirements for the  $n_p$  measurements, which must be fulfilled in order to provide robust constraints on cosmological parameters.

#### 3.1. Angular galaxy clustering

In this section we provide a general introduction to photometric angular clustering, as a basis for the development of our clustering-redshifts method. It draws inspiration from many previous weak-lensing and clustering- $z$  works such as Davis et al. (2018), Krause et al. (2021), Gatti et al. (2018, 2022), Cawthon et al. (2022), and Pandey et al. (2025).

##### 3.1.1. Galaxy density contrast

The observed angular galaxy count of a sample  $a$  (photometric or spectroscopic) is

$$N_a^{\text{obs}}(\theta) = \bar{N}_a (1 + \Delta_a^{\text{obs}}(\theta)), \quad (24)$$

where  $\bar{N}_a$  is the mean (observed) galaxy count, and  $\Delta_a^{\text{obs}}$  is the observed projected galaxy density contrast. It can be expressed as (Krause et al. 2021)

$$\Delta_a^{\text{obs}}(\theta) \approx \Delta_a^{\text{D}}(\theta) + \Delta_a^{\text{RSD}}(\theta) + \Delta_a^{\mu}(\theta), \quad (25)$$

where  $\mu$  refers to magnification, RSD to redshift space distortion (in the linear regime<sup>2</sup>), and  $\Delta_a^{\text{D}}$  to the line of sight projection of

<sup>2</sup> In the linear regime, RSD contribute to the projected galaxy density contrast through the apparent large-scale flow, of galaxies across the redshift boundaries of the samples. The redshift  $z$  we use to parametrise the  $D$  and  $\mu$  terms includes the peculiar velocity, and is referred to as observed redshift.

the three-dimensional (3D) galaxy density contrast  $\delta_a(z, \theta)$ :

$$\Delta_a^{\text{D}}(\theta) = \int dz n_a(z) \delta_a(z, \theta), \quad (26)$$

where  $n_a(z)$  is the redshift distribution of the observed sample, normalised to unity, and function of the observed redshift. All these quantities are associated with observations, and they depend on survey properties (depth, masks etc.), but for clarity, we drop the ‘‘obs’’ index for the rest of our work, as we do not consider observational systematics. We also do not model relativistic corrections.

##### 3.1.2. Angular two-point correlation function

The angular two-point correlation function of samples  $a$  and  $b$  (with possibly  $a = b$ ) from the observed projected density contrast can be written as

$$w_{ab}^{\text{full}}(\theta) = \langle \Delta_a^{\text{D}} \Delta_b^{\text{D}} \rangle(\theta) + \langle \Delta_a^{\text{RSD}} \Delta_b^{\text{D}} \rangle(\theta) + \langle \Delta_a^{\text{D}} \Delta_b^{\text{RSD}} \rangle(\theta) + \langle \Delta_a^{\text{D}} \Delta_b^{\mu} \rangle(\theta) + \langle \Delta_a^{\mu} \Delta_b^{\text{D}} \rangle(\theta). \quad (27)$$

where  $\langle \dots \rangle$  indicates an average over angular scales, and  $\theta$  is now a scalar<sup>3</sup>. The dominant corrections (RSD $\times$ D;  $\mu \times$ D) are already small with respect to the dominant term (D $\times$ D), and we decided to omit even smaller corrections (Krause et al. 2021; Gatti et al. 2022).

We will mainly focus on the D $\times$ D correlation. Since it corresponds to the leading order, we omit its D index in the rest of the article. The clustering two-point correlation can be expressed as

$$w_{ab}(\theta) = \langle \Delta_a^{\text{D}} \Delta_b^{\text{D}} \rangle(\theta) \quad (28)$$

$$= \iint dz_1 dz_2 n_a(z_1) n_b(z_2) \langle \delta_a \delta_b \rangle(z_1, z_2, \theta) \quad (29)$$

$$= \iint dz_1 dz_2 n_a(z_1) n_b(z_2) \xi_{ab}(z_1, z_2, \theta), \quad (30)$$

where we have defined the 3D cross-correlation function  $\xi_{ab}(z_a, z_b, \theta)$ , and used  $\langle \delta_x \rangle = 0$ . While we have introduced the correlations as a function of sky separation  $\theta$ , we can also express them as functions of the perpendicular comoving separation  $r_p$ , defined as

$$r_p = \theta \chi(z), \quad (31)$$

in the flat sky approximation, and small angle limit, where  $\chi$  is the comoving radial distance (equal to the comoving angular-diameter distance with no spatial curvature). Working with  $r_p$  instead of  $\theta$  is relevant when making particular scale choices for evaluating the correlation, but the Python package we use, TreeCorr (Jarvis 2015) has angular separation as the default convention.

<sup>3</sup> Our Universe and its observation is assumed to be isotropic. Although the real isotropy of our Universe seems a reasonable hypothesis, observation may not always support it. For example, closer to the Galactic plane, there is more star contamination, leading to angular anisotropy. The *Euclid* observation strategy might also induce some angular anisotropy. This could potentially introduce bias on the cosmological parameters, especially for experiments such as *Euclid* or LSST (Baleato Lizancos & White 2023). Indeed, clustering-redshifts may be a valuable tool to investigate this issue by splitting the samples into independent sky patches.

### 3.1.3. Moving to matter statistics with galaxy bias

For large-scale correlations, the linear galaxy bias provides a relation between the galaxy overdensity and the matter density field  $\delta_m$

$$\delta_g(\boldsymbol{\theta}, z) = b_g(z) \delta_m(\boldsymbol{\theta}, z), \quad (32)$$

where  $b_g$  is the galaxy sample bias, and  $\delta_m$  is the matter density field. Under this approximation, the angular correlation function becomes,

$$w_{ab}(\theta) = \iint dz_1 dz_2 b_a(z_1) b_b(z_2) n_a(z_1) n_b(z_2) \xi_m(z_1, z_2, \theta), \quad (33)$$

where  $\xi_m$  is the 3D matter two-point correlation.

Following the literature, we used the linear bias expression of Eq. (32) by default. It is worth noting, however, that the linear bias expression is expected to break at scales around  $6 h^{-1}$  Mpc. In addition, the next-order expression would extend this scale range to around  $4 h^{-1}$  Mpc (Pandey et al. 2020). Thus, none of the equations (linear or non-linear e.g. Appendix A) are properly adapted to the standard clustering-redshifts scale ranges,  $(0.1-1) h^{-1}$  Mpc,  $(0.5-1.5) h^{-1}$  Mpc or  $(1-5) h^{-1}$  Mpc. However, linear bias has been tested to give accurate enough modelling for clustering-redshifts analysis (Gatti et al. 2018, 2022; van den Busch et al. 2020; Cawthon et al. 2022; Naidoo et al. 2023). As a consequence, it is crucial to quantify and address potential deviations introduced in the reconstructed  $n(z)$ . We will pay particular attention to the galaxy bias issue, nonlinear contributions, and the scale range choice.

### 3.1.4. Magnification

In this section, we describe the two magnification terms in Eq. (27),

$$M_{ab}(r_p) = \langle \Delta_a^D \Delta_b^\mu \rangle(r_p) + \langle \Delta_a^\mu \Delta_b^D \rangle(r_p). \quad (34)$$

Magnification is a lensing effect due to the gravitational bending of distant source light by the matter between those sources and the observer (Krause et al. 2021; Elvin-Poole et al. 2023). It introduces an angular position correlation for galaxies distant in redshift because of the lensing of a high- $z$  galaxy sample by the matter traced by a lower- $z$  sample<sup>4</sup>. Fortunately, this effect, which breaks the naive hypothesis that in clustering-redshifts only galaxies close in redshift are correlated, can be modelled.

Magnification impacts the apparent position of galaxies and their density. In regions of positive convergence<sup>5</sup>, the apparent separation between any two points on a source plane is enlarged, leading the telescope to capture a larger apparent solid angle, consequently reducing the apparent galaxy density locally. Another effect of positive convergence is an increase in distant source flux received by the telescope, potentially resulting in a change in the number of galaxies passing through the selection function.

Magnification was taken into account for clustering- $z$  in DES, through direct modelling of its effect, and simulation-based estimation of magnification coefficients (Gatti et al. 2022; Cawthon et al. 2022). Complete inclusion of magnification

<sup>4</sup> The magnification of the two samples by the lower- $z$  matter is negligible as it is a smaller correction ( $\mu \times \mu$  cf. Sect. 3.1.2).

<sup>5</sup> Positive convergence is induced by galaxy over-density. Correlation with voids would probe the negative convergence regime.

should include position magnification, flux magnification, and its impact on photo- $z$ , which can be significant (Legnani et al., in prep.), but we restricted this work to position and flux magnification. We detail how magnification can be modelled in the Appendix B. Since the tomographic bins are expected to be narrower for *Euclid* than for DES, the effect of magnification is expected to be smaller, as illustrated by the toy examples in Appendix B. Therefore, we do not include magnification in our default modelling. We will reassess the need to incorporate magnification in future work, based on the properties of the data bins.

### 3.1.5. RSD

In the linear regime, RSD contributes to the projected galaxy density contrast through the apparent large-scale flow of galaxies across the redshift boundaries of the bins (Krause et al. 2021). Equivalently, the redshift distribution of galaxies as a function of the observed redshift is slightly different from the distribution as a function of the Hubble redshift. The latter should be used, however, as the modelling assumes isotropy through the correlation function  $\xi$ . Since the photometric tomographic bins are quite large, it would mainly affect the spectroscopic sample which is divided into narrow redshift slices; the narrower the slices, the larger the effect is. Therefore, only the term  $\langle \Delta_s^{\text{RSD}} \Delta_p^D \rangle(\theta)$  may contribute to the two-point correlation of Eq. (27). We tested in Sect. 4.7 the bias induced by the use of the spec- $z$  instead of true- $z$  for the spectroscopic slicing, in the Flagship simulation, since peculiar velocities are included, cf. Sect. 2.1.

## 3.2. Impact of the galaxy redshift distribution and its modelling

A tomographic bin will be referred to as a sample  $p$ , without referring to the tomographic ID, as the latter is always explicit. Different spectroscopic samples are considered in this article from different simulated surveys and tracers, such as DESI ELGs. For a given survey tracer, we will additionally distribute the sample into redshift slices, with spectroscopic redshift cuts. For a given redshift slice, the galaxies will be referred to as  $s_j$ , or  $\{s, z_j\}$  with  $z_j$  the mean of the redshift slice. We explore two cases for the modelling of the distribution and two-point correlations with a sliced spectroscopic sample.

### 3.2.1. Ideal case of a Dirac galaxy distribution

The situation often presented in clustering-redshifts literature (e.g. Gatti et al. 2022; Giannini et al. 2024) is the one of a sample localised in such a small redshift slice that it can be approximated by a Dirac distribution

$$n_a(z) \approx \delta_D(z - z_i). \quad (35)$$

The cross-correlation with a second sample is then simplified to

$$w_{ab}^{\text{Dirac}}(\theta) = b_a(z_i) \int dz b_b(z) n_b(z) \xi_m(z_i, z, \theta), \quad (36)$$

$$\approx b_a(z_i) b_b(z_i) n_b(z_i) \xi_m(z_i, \theta), \quad (37)$$

with

$$\xi_m(z_i, \theta) = \int dz \xi_m(z_i, z, \theta). \quad (38)$$

To go from Eq. (36) to Eq. (37), we assume the redshift evolution of the bias and distribution of the second sample is small in the support of  $\xi_m$  (i.e. in a small redshift interval centred in  $z_i$ ).

### 3.2.2. Realistic case: Spectroscopic sample into small redshift slices

Using the Limber approximation, Eq. (33) reduces to,

$$w_{ab}^{\text{Limber-full-z}}(\theta) = \int dz b_a(z) b_b(z) n_a(z) n_b(z) \xi_m(z, \theta), \quad (39)$$

where  $\xi_m$  is the same as in Eq. (38). In the standard scenario of clustering redshifts, the spectroscopic galaxy sample is localised within small redshift slices  $z_i \pm \Delta z_i/2$ , so that the distribution can be approximated as uniform

$$n_a(z) = \begin{cases} 1/\Delta z_i & \text{if } |z - z_i| < \Delta z_i/2, \\ 0 & \text{else,} \end{cases} \quad (40)$$

where  $\Delta z_i$  is typically between 0.01 and 0.05. This modelling is of course only approximated true in practice, and would lead to some deviations which we will evaluate. In particular, under the assumption of constant galaxy bias across the redshift bin, the spectroscopic auto-correlation ( $b = a$ ) becomes

$$w_{aa}^{\text{Limber-1-bin}}(\theta) \approx \frac{b_a^2(z_i)}{\Delta z_i^2} \int_{-\Delta z_i/2}^{\Delta z_i/2} dz \xi_m(z_i + z, \theta). \quad (41)$$

Thus, given a cosmology,  $w_{aa}$  provides a straightforward measurement of the spectroscopic galaxy bias. We note that this equation diverges as  $\Delta z \rightarrow 0$ , and that, more generally, both our modelling and the Limber approximation are expected to become less accurate for narrower redshift bins and on larger angular scales (Simon 2007).

For  $b \neq a$ , when the redshift evolution of the biases  $b_a$  and  $b_b$  and the redshift distribution  $n_b$  is neglected, we obtain

$$w_{ab}^{\text{Limber-1-bin}}(\theta) = \frac{b_a(z_i)}{\Delta z_i} b_b(z_i) n_b(z_i) \int_{-\Delta z_i/2}^{\Delta z_i/2} dz \xi_m(z_i + z, \theta). \quad (42)$$

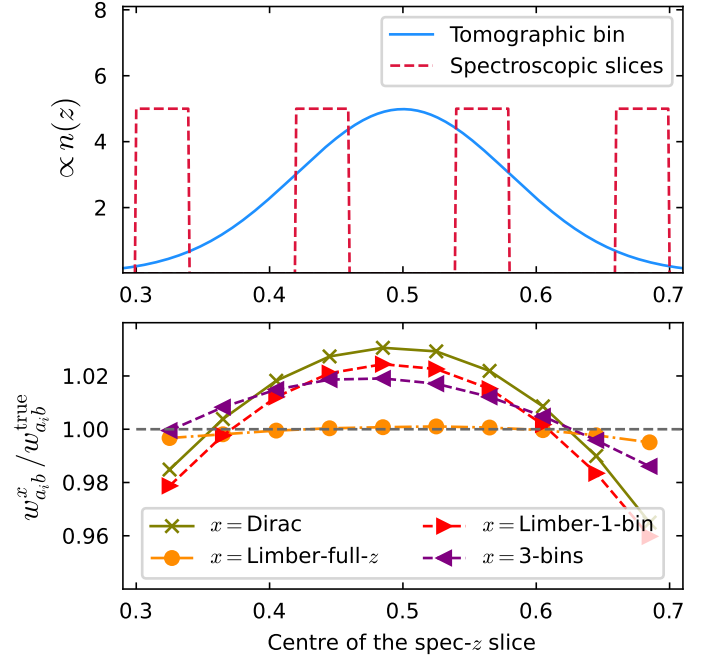
The superscript Limber-1-bin refers to the integration of the matter distribution over a redshift bin equal to the spectroscopic slice; we use the Limber approximation too. If one further neglects the matter correlation variation over  $\Delta z_i$ , one recovers the Dirac approximation Eq. (37); this is the approach of van den Busch et al. (2020), for example.

A limitation of Eq. (42) is the assumption that the redshift distribution  $n_b$  is constant across the bin. In practice,  $n_b$  could be larger near the bin edges and smaller near the centre (or the opposite), leading to an underestimation (resp. overestimation) of correlations with galaxies localised near  $z_i \pm \Delta z_i$ . To address these issues without resorting to Eq. (39), we propose a correction by explicitly incorporating contributions from galaxies outside the spectroscopic redshift slice (three-bins):

$$w_{ab}^{\text{3-bins}}(\theta) = \frac{b_a(z_i)}{\Delta z_i} \sum_{\delta z \in \{0, -\Delta z_i, \Delta z_i\}} b_b(z_i + \delta z) n_b(z_i + \delta z) \times \int_{-\Delta z_i/2}^{\Delta z_i/2} dz \xi_m^{\delta z}(z_i + z, \theta), \quad (43)$$

We introduce the phased correlations

$$\xi_m^{\delta z}(z, \theta) = \int_{-\Delta z_i/2}^{\Delta z_i/2} dz' \xi_m(z, z + z' + \delta z, \theta), \quad (44)$$



**Fig. 2.** *Top:* Redshift distribution of sample  $b$  in blue, and four out of the ten samples  $a_i$  in red. *Bottom:* Deviation of approximated  $w_{a_i b}$  with respect to the exact one for the ten cross-correlations.

which account for the correlation of two bins of widths  $\Delta z_i$ , separated by  $\delta z$ . Here, the three-bins modelling does not use the Limber approximation. One can write Eq. (43) as

$$w_{ab}(\theta) \approx b_a(z_i) \int_{-\Delta z_i/2}^{\Delta z_i/2} dz \xi_m^0(z_i + z, \theta) \times \sum_{j \in \{-1, 0, 1\}} b_b(z_i + j \Delta z) n_b(z_i + j \Delta z) \eta_j(z_i, \theta), \quad (45)$$

with  $\eta_0 = 1$  and

$$\eta_{\pm}(z_i, \theta) = \frac{\int_{-\Delta z_i/2}^{\Delta z_i/2} dz \xi_m^{\pm \Delta z_i}(z_i + z, \theta)}{\int_{-\Delta z_i/2}^{\Delta z_i/2} dz \xi_m^0(z_i + z, \theta)}, \quad (46)$$

where  $\xi_m^{0, \pm \Delta z_i}$  is defined in Eq. (44). This formulation allows for a linear inference of  $\mathbf{n}$  from the two-point correlation, which can be implemented in practice, unlike the Limber-full- $z$  case where  $n_b(z)$  is integrated over, complicating the process.

### 3.2.3. Comparison of the correlation modelling

In the previous sections, we introduced four different approximations of the two-point correlation function described by Eq. (33):  $w_{ab}^{\text{Dirac}}$ ,  $w_{ab}^{\text{Limber}}$ ,  $w_{ab}^{\text{1-bin}}$ , and  $w_{ab}^{\text{3-bins}}$ . In this section, we investigate the differences between them with a toy model. We consider 10 samples  $a_i$ , each with a step- $n(z)$  distribution as given by Eq. (40), centred at  $z_i = 0.325 + 0.05i$ , with  $i = 0, 1, \dots, 9$ , and  $\Delta z_i = 0.05$ . For the redshift distribution of sample  $b$  we take a Gaussian  $(\mu, \sigma) = (0.5, 0.08)$ . This reproduces an idealistic case scenario of clustering- $z$  calibration, with sample  $b$  representing the photometric data, and samples  $a_i$  representing the spectroscopic data. Several of these  $n(z)$ s are plotted in the upper panel of Fig. 2. We assume a constant galaxy bias  $b_a(z) = b_b(z) = 1$  for all samples, and evaluate correlation functions for  $r_p = 2$  Mpc

(a scale standard for clustering- $z$  analysis). We numerically evaluated the functions with the CCL library<sup>6</sup> (Chisari et al. 2019). We used the default configuration, which is based on predictions from CLASS (Blas et al. 2011) and halofit (Takahashi et al. 2012) for the nonlinear power spectrum. We evaluated the functions at scales where we know the predictions are not completely precise because of the galaxy-halo connection, but as will be explicit in Sect. 3.3, this is expected to have little impact on real data.

In the lower panel of Fig. 2, we show the deviation of the four approximations to the “exact” correlation Eq. (33), for the samples  $a_i, b$ . The level of the deviations depends on the  $n_p(z)$  where the correlation is evaluated: we observe a larger deviation near the peak and for the tails. We evaluated the root mean square (RMS) of the deviation to 1 and got 0.022 (Dirac), 0.021 (Limber-1-bin), 0.013 (3-bins), and 0.002 (Limber-full- $z$ ). We can see that the Dirac and one-bin approximations are very close. Taking into consideration the correlations from bin neighbours with three-bins, we get an improvement of 40%. Finally, the best approximation is the Limber-full- $z$  case [i.e. integrating  $n^2(z)$  instead of assuming some effective values], with a deviation of 0.2% for individual points, and an RMS smaller by an order of magnitude to the Dirac and one-bin approximations.

In clustering- $z$  studies, however, the redshift distribution  $n_p(z)$  is an unknown quantity that we aim to constrain from the measurement of  $w_{ab}$ . As a result, the Limber-full- $z$  case is challenging to apply in practice, as the unknown  $n_p(z)$  is integrated over. In contrast, for the Dirac and one-bin cases,  $n_p(z_i)$  enters the calculation linearly, making it easier to extract information. The same applies to the three-bins approximation, which provides a better approximation, as we have demonstrated.

### 3.3. Application to clustering redshifts

In the context of clustering redshifts, a photometric sample has an unknown redshift distribution  $n_p(z)$ . To reconstruct it, we measure the cross-correlations with spectroscopic samples, distributed into fine redshift slices (usually  $0.02 \leq \Delta z \leq 0.05$ ). The distribution of these spectroscopic samples is assumed to be constant in each of these smaller slices (see Eq. 40). The amplitude of the photometric distribution at redshift  $z_i$  can then be inferred linearly from the cross-correlation function between a spectroscopic bin centred in  $z_i$  and the photometric sample [noted  $w_{sp}(z_i, \theta)$  instead of  $w_{sp}(\theta)$ ], assuming the Dirac modelling Eq. (37):

$$n_p(z_i) = \frac{w_{sp}(z_i, \theta)}{b_s(z_i) b_p(z_i) \xi_m(z_i, \theta)}. \quad (47)$$

One can write a similar expression for the one-bin and three-bins modelling, as described in Sect. 3.2.3.

The spectroscopic bias can be measured with the spectroscopic auto-correlation, following Eq. (41),

$$n_p(z_i) = \frac{w_{sp}(z_i, \theta)}{\sqrt{\Delta z} w_{ss}(z_i, \theta) \xi_m(z_i, \theta) b_p(z_i)}. \quad (48)$$

The  $n_p(z)$  is still degenerate with the photometric galaxy bias. In the next section, we describe different possibilities for correcting this degeneracy.

#### 3.3.1. Correcting the photometric galaxy bias

It is more complicated to measure the photometric bias than the spectroscopic bias since the sample cannot be split into small redshift slices and Eq. (41) cannot be applied. Here we introduce different methods.

- M0: No correction at all. We assume there is no  $b_r$  and  $b_p$  redshift evolutions and renormalise the measured redshift distribution.
- M1: Spectroscopic bias only. We assume there is no  $b_p$  redshift evolution and renormalise the measured redshift distribution.
- M2: The photometric bias is measured for the whole tomographic bin. Hence we do not take into account the redshift evolution within the tomographic bin, but we are increasing the uncertainty on the  $n_p(z_i)$ , and thus we are more conservative.
- M3: The photometric bias is measured for relatively large photometric bins ( $\Delta z_{\text{photo}} = 0.1$ ), and the redshift evolution is deduced from a polynomial fit. The photo- $z$  spreading of the bin needs to be corrected before the fit (more details in Sect. 3.3.2).
- M4: The photometric bias is measured for small photometric bins ( $\Delta z_{\text{photo}} = 0.02$ ), and the redshift evolution is deduced from a polynomial fit. The photo- $z$  spreading of the bin needs to be corrected before the fit (cf. Sect. 3.3.2).
- M5: Full correction. The two biases are measured using the true redshift (simulation only).

Methods 0-1-2-5 were already tested in Naidoo et al. (2023). Methods 3 and 4 are new in that context, but similar to methods from Cawthon et al. (2022) and van den Busch et al. (2020). Ideally, we would like to measure  $w_{ss}$  and  $w_{pp}$  over the same redshift slices:  $[z_i \pm \Delta z/2]$ . In doing so, we would get

$$n_p(z_i) = \frac{w_{sp}(z_i, \theta)}{\Delta z \sqrt{w_{ss}(z_i, \theta) w_{pp}(z_i, \theta)}}. \quad (49)$$

We note that we derive this equation with the linear bias assumption. Indeed we show that it remains valid at the next-order of the bias expansion in Appendix A. This ideal situation corresponds to M5, the photometric sample being distributed into redshift slices with the true redshift variable. This method can only be evaluated via simulation. The aim of methods 3 and 4 is to estimate  $w_{pp}$  but with a different binning, and taking into account the photo- $z$  spreading. With M2, we don't correct for the redshift evolution, but we aim to take into account the variance induced by the photometric galaxy bias (Naidoo et al. 2023).

#### 3.3.2. More details about the M3 and M4 corrections

With methods 3 and 4, the photometric sample is divided into bins (denoted  $j$ ) based on photo- $z$  cuts. These differ from the spectroscopic distribution described in Eq. (40), therefore requiring correction for the photo- $z$  spreading. In other words, a photometric bin cannot be approximated by a redshift slice.

In practice, we measure the  $n(z)$  of these bins with the clustering-redshifts M1 method (cf. Sect. 3.3.1). This method assumes negligible photometric redshift evolution within each small bin. Then, we want to convert our imperfect-bin measurement to what we would measure for an ideal case with step  $n(z)$ . Using the Limber case (cf. Sect. 3.2.3), we introduce the correction to the two-point auto-correlation function

$$w_{pp}^{\text{corr}}(z_j, \theta) = w_{pp}^{\text{meas}}(z_j, \theta) \frac{(\Delta z)^{-2} \int_{-\Delta z/2}^{\Delta z/2} dz \xi_m(z_j + z, \theta)}{\int_0^\infty dz n_p^2(z) \xi_m(z, \theta)}, \quad (50)$$

<sup>6</sup> <https://github.com/LSSTDESC/CCL>

where the  $\Delta z$  corresponds to the spectroscopic redshift bin of the  $w_{\text{sp}}$  correlation. We refer to this as the “full” correction, as it accounts for the redshift evolution of the matter correlation function across the bin. We use CCL to evaluate  $\xi_m$  at small scales.

A simpler approximation has been implemented for the DES redMaGiC and Maglim samples (Cawthon et al. 2022; Giannini et al. 2024). The correction was modelled with

$$w_{\text{pp}}^{\text{corr}}(z_j, \theta) = w_{\text{pp}}^{\text{meas}}(z_j, \theta) \frac{(\Delta z)^{-1}}{\int_0^\infty dz n_i^2(z)}, \quad (51)$$

where  $n_i(z)$  was estimated for the redMaGiC subsample having a spec- $z$  and a photo- $z$  (Gatti et al. 2022), or with clustering-redshifts (Cawthon et al. 2022). This is referred to as the “partial” correction, as it makes additional simplifying assumptions compared to the “full” correction in Eq. (50).

The “width-only” correction applies when using different bin widths for the auto- and cross-correlation functions, while still neglecting the effects of photo- $z$  spreading across the bin. This correction simplifies to

$$w_{\text{pp}}^{\text{corr}}(z_j, \theta) = w_{\text{pp}}^{\text{meas}}(z_j, \theta) \frac{(\Delta z)_{\text{meas}}}{(\Delta z)_{w_{\text{sp}}}}. \quad (52)$$

In summary, the “full” correction accounts for the redshift evolution of the matter correlation function, the “partial” correction introduces additional approximations by assuming a simplified  $n(z)$ , and the “width-only” correction applies when focusing solely on differences in bin widths while neglecting photo- $z$  spreading. We illustrate these methods and apply them to clustering-redshifts in Sect. 4.8.

These two methods for correcting photometric galaxy bias share similarities with the method in van den Busch et al. (2020). We do not assume a bias model of the form  $B_p(z) \propto (1+z)^\alpha$ , however. Instead, we directly measure and correct  $w_{\text{pp}}(z)$  by constraining the photo- $z$  spreading using clustering- $z$  for smaller bins. Theoretically, the two approaches could be combined, which would eliminate any systematic effects continuous on the redshift, without assuming a particular model.

### 3.3.3. Data vectors reduced to scalars with some weighting

Using Eq. (48), one can extract a measurement of  $n_p$  for different  $r_p$  (or equivalently  $\theta$ ). To optimise the sensitivity, it is standard to reduce all the data vectors to scalars with a scale weighting  $W_w(r_p)$ ,

$$\bar{w}_{ab}(z_i) = \int_{r_p^{\text{min}}}^{r_p^{\text{max}}} dr_p W_w(r_p) w_{ab}(r_p, z_i). \quad (53)$$

We introduce the scale-weighted matter correlation

$$\bar{\xi}_m(z) = \int_{r_p^{\text{min}}}^{r_p^{\text{max}}} dr_p W_w(r_p) \xi_m(r_p, z). \quad (54)$$

It is easy to show that Eq. (48) or Eq. (49) are still valid with integrated quantities. The scale weighting is usually chosen as a power law  $W_w(r_p) \propto r_p^\gamma$ , where the function is normalised to unity (e.g. Gatti et al. 2018; van den Busch et al. 2020; Naidoo et al. 2023). Choosing  $\gamma < 0$  increases the impact of small scales. They are expected to be less noisy but also more sensitive to non-linearities. Inversely, the measurement is not biased at large scales (i.e. the linear bias is a good model), but uncertainties are usually larger. Another reason to use small

scales instead of large ones is to keep the calibration independent from the two-point weak-lensing analysis: using smaller and different scales, one can safely neglect the covariance between calibration and the observables. Nonetheless, these small scales are not used for the cosmological analysis for physically motivated reasons (non-linearities), which in theory should apply to clustering-redshifts as well (Pandey et al. 2020). Hence, the scale range is an important parameter and we will investigate this scale-weighting procedure in detail in Sect. 4.

Another possibility of weighting is to combine the different  $n_p(z_i|r_p)$ , with a particular weighting  $W_n(r_p)$ :

$$\bar{n}_p(z_i) = \int_{r_p^{\text{min}}}^{r_p^{\text{max}}} dr_p W_n(r_p) n_p(z_i|r_p). \quad (55)$$

In other words, we integrate directly  $n(z|r_p)$ , which is  $w_{\text{sp}}$  for M0,  $w_{\text{sp}}/\sqrt{w_{\text{ss}}}$  for M1 or  $w_{\text{sp}}/\sqrt{w_{\text{ss}}w_{\text{pp}}}$  for M2-3-4-5.

### 3.4. Estimating the correlations with pair counts

We evaluated position-position correlations by counting pairs separated by a certain distance divided by the expectation of uncorrelated samples (random) to properly normalise the correlation. We chose the Landy–Szalay (LS) estimator (Landy & Szalay 1993),

$$w_{ab}^{\text{data}}(r_p) = \frac{D_a D_b(r_p) - R_a D_b(r_p) - D_a R_b(r_p) + R_a R_b(r_p)}{R_a R_b(r_p)}, \quad (56)$$

where  $D_a D_b$ ,  $D_a R_b$ ,  $R_a D_b$ , and  $R_a R_b$  are the standard data-data, data-random, random-data, random-random pair counts of samples  $a$  and  $b$ . We note that we omitted the galaxy number normalisation factors which correct for the randoms being ten times more numerous than data. Further details are given in Appendix C. As we were working with a simulation omitting observational systematics and masks, we generated random catalogues homogeneously over the footprint, taking 50 times more randoms than data.

#### 3.4.1. Estimator weighting scheme

Given the weighting procedure Eq. (53), it is natural to weight the estimator as

$$\tilde{w}_{ab}^{(1)}(z_i) = \int_{r_p^{\text{min}}}^{r_p^{\text{max}}} dr_p W(r_p) \frac{(D_a - R_a)(D_b - R_b)(r_p)}{R_a R_b(r_p)}, \quad (57)$$

where the numerator is a concise notation for the numerator of Eq. (56). In the context of clustering redshifts, this weighted estimator was first introduced by Ménard et al. (2013), and later used by Cawthon et al. (2022).

Another weighting scheme was introduced by Schmidt et al. (2013), and was used by Davis et al. (2018), Gatti et al. (2018), van den Busch et al. (2020), and Naidoo et al. (2023). The principle is to scale-average data-data and random-random pair counts:

$$\tilde{w}_{ab}^{(2)}(z_i) = \frac{\int_{r_p^{\text{min}}}^{r_p^{\text{max}}} dr_p W(r_p) (D_a - R_a)(D_b - R_b)(r_p)}{\int_{r_p^{\text{min}}}^{r_p^{\text{max}}} dr_p W(r_p) R_a R_b(r_p)}. \quad (58)$$

Schmidt’s motivation to do so was to assume that the ratio of the averaged quantities would lead to higher signal-to-noise (S/N) than the averaged ratio (which is expected for a constant ratio,

which is not the case here). With this second estimator what we have is a re-weighting of the correlation functions with a new effective weighting, as

$$\tilde{w}_{ab}^{(2)}(z_i) = \iint dz_1 dz_2 n_a(z_1) n_b(z_2) \quad (59)$$

$$\times \frac{\int_{r_p^{\min}}^{r_p^{\max}} dr_p W(r_p) r_p \Delta r_p \eta_{\text{mask}}(r_p) \langle \delta_a \delta_b \rangle_{r_p, z_1, z_2}}{\int_{r_p^{\min}}^{r_p^{\max}} dr_p W(r_p) r_p \Delta r_p \eta_{\text{mask}}(r_p)},$$

where  $\eta_{\text{mask}}(r_p)$  is a geometrical factor due to the masking of some sky area (cf. Appendix C). In particular, the effective scale weighting is

$$W_{\text{eff}}(r_p) \propto W(r_p) r_p \Delta r_p \eta_{\text{mask}}(r_p) \quad (60)$$

$$\propto W(r_p) r_p^2 \eta_{\text{mask}} \quad \text{for a logarithmic binning,} \quad (61)$$

$$\propto r_p^{\gamma+2} \eta_{\text{mask}} \quad \text{for a power law weighting.} \quad (62)$$

This effective weighting is still re-normalised to unity as required by the denominator of Eq. (59). The mask incompleteness is not rigorously corrected (cf. Appendix C), except if  $\eta_{\text{mask}}$  is scale-independent<sup>7</sup>. By default, we consider the first estimator (Eq. 57), since understanding the weighting is straightforward, and the mask is better corrected, but we will test and compare the two in the results section<sup>8</sup>.

### 3.4.2. Covariance matrix

We used the jackknife covariance matrix as our fiducial covariance. For a measured data vector  $w$ , the covariance matrix is

$$C_w = \frac{N_{\text{Jkk}} - 1}{N_{\text{Jkk}}} \sum_{k=1}^{N_{\text{Jkk}}} (w^k - \tilde{w})(w^k - \tilde{w})^T, \quad (64)$$

where  $N_{\text{Jkk}}$  is the number of jackknife regions,  $w^k$  is the data vector excluding the data from the  $k^{\text{th}}$  region, and  $\tilde{w}$  is its mean value. Here,  $w$  can be, for example, the vector  $w_{\text{sp}}(r_p)$ , the ratio of vectors  $w_{\text{sp}}(r_p)/\sqrt{w_{\text{ss}}(r_p)}$ , the scalars  $\bar{w}_{\text{sp}}$ , or the ratio of scalars  $\bar{w}_{\text{sp}}/\sqrt{\bar{w}_{\text{ss}}}$ . In this article, we use a sufficiently large number of regions (200), with a small enough number of points (around a dozen), so that Hartlap (Hartlap et al. 2007) or Percival (Percival et al. 2022) corrections to the inverse of the covariance are at the percent level and can be safely neglected.

## 3.5. Extracting the unknown distribution moments

### 3.5.1. clustering-redshifts requirements for cosmic shear

The success of the *Euclid* mission requires achieving a particular precision in the characterisation of  $n(z)$  of the tomographic bins  $i$ . The requirements usually focus on the mean redshift, expressed as

$$\langle z \rangle_i = \int dz z n_{p,i}(z), \quad (65)$$

<sup>7</sup> For scales that are neither too small nor too large, it should be a valid hypothesis, at least to first order.

<sup>8</sup> Of course, in practice, all the previous integrals are discrete, with

$$\int dr_p \rightarrow \sum_i \Delta r_p^i. \quad (63)$$

as illustrated in works such as Gatti et al. (2018), van den Busch et al. (2020), and Gatti et al. (2022). The desired precision for the *Euclid* mission requires knowledge of the redshift standard deviation, however, which is denoted as

$$\sigma_{n_i} = \sqrt{\int dz (z - \langle z \rangle)^2 n_{p,i}(z)}, \quad (66)$$

as outlined in Reischke (2024). The criteria are as follows: the mean redshift bias (Laureijs et al. 2011; Reischke 2024) and the redshift standard deviation (Reischke 2024) for every bin should be measured with precision

$$\sigma(\langle z \rangle_i) \leq 0.002 (1 + \langle z \rangle_i), \quad (67)$$

$$\sigma(\sigma_{n_i}) \leq 0.1 \sigma_{n_i}. \quad (68)$$

### 3.5.2. Discretisation of a distribution

The last step of a clustering-redshifts pipeline is to derive the  $n_p(z)$  moments from the discrete measurements  $n_p(z_i)$ . The mean redshift of a bin is evaluated with

$$\langle z \rangle \approx \sum_j \Delta z z_j n_p^{\text{meas}}(z_j). \quad (69)$$

We are not directly measuring  $n_p(z_j)$  but rather its averaged value over ranges  $z_j \pm \Delta z/2$ ,

$$n_p^{\text{meas}}(z_j) \approx \frac{1}{\Delta z} \int_{-\Delta z/2}^{\Delta z/2} dz n_p(z_j + z). \quad (70)$$

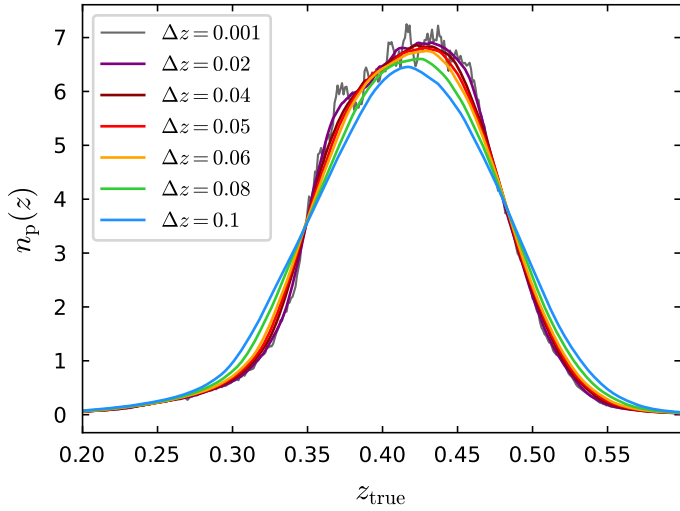
As a consequence, Eq. (69) is an accurate approximation of the exact integral form as long as  $\Delta z$  is small compared to the tomographic bin width. As an example, the numerical difference between Eq. (69) and Eq. (65) is  $\mathcal{O}(10^{-4})$  for our *Euclid* tomographic bins (cf. Sect. 2.2), and  $\Delta z = 0.02\text{--}0.5$ . Nonetheless, the recovered shape of the distribution is affected by Eq. (70) and for large  $\Delta z$  the measured distribution would be flattened. We illustrate this in Fig. 3, plotting a galaxy distribution (which in practice corresponds to a very thin slicing), and the same distribution with a  $\Delta z$  averaging as Eq. (70). We see that the flattening associated with the bin increases. These effects will also be tested in Sect. 4.

### 3.5.3. Overall strategy

The uncertainty on the mean- $z$  directly inferred with Eq. (69) would not fulfil the *Euclid* requirement Eq. (67). In doing this, we would be very conservative, allowing for any tomographic bin shape, whereas in reality, we know their shape a priori. Thus, we can define some priors on the  $n_p(z)$  properties to decrease the uncertainty. We describe two possibilities:

1. we assume a prior on the shape of the distribution, through a  $n(z)$  model;
2. we use Gaussian process formalism, to generate coherent realisations, assuming a kernel, which is related to how ‘‘fast’’ the distribution is evolving in redshift.

Thus in one case, we put a prior on the shape of the tomographic bin, and in the other, on the redshift variation.



**Fig. 3.** Galaxy distribution observed through different  $\Delta z$  averaging as Eq. (70). The redshift convolution modifies the shape of the distribution for larger slicing compared to the real slicing as estimated with a very thin slicing (grey). For this case, all slicings  $\Delta z \leq 0.06$  produce a good estimate.

### 3.5.4. Shifted-stretched model

For this approach, we evaluated a given model  $n^{\text{model}}(z)$  with some parameters  $\{p_1, p_2 \dots\}$ , by minimising the likelihood,

$$\ln \mathcal{L}(\{p_1, p_2 \dots\} | n, C_n) = -\frac{1}{2} \left[ n(z_i) - n^{\text{model}}(z_i) \right]^T C_n^{-1} \left[ n(z_i) - n^{\text{model}}(z_i) \right], \quad (71)$$

where  $\{n_p(z_i)\}_i$ , and the associated covariance  $C_n$  are the measured quantities. Then, using a Markov chain Monte Carlo<sup>9</sup>, the mean redshift and standard deviation were extracted from one thousand realisations of  $\{n^{\text{model}}\}$ .

For the  $n^{\text{model}}$ , we use the shifted-stretched model (SSM), which is commonly used for clustering redshifts; given a particular distribution  $n(z)$  with a mean redshift  $\langle z \rangle$ , two free parameters are used describing a redshift shift  $\delta z$  and a stretching  $s$ ,

$$n^{\text{model}}(z | s, \delta z) = \frac{1}{s} n \left( \frac{z - \langle z \rangle - \delta z}{s} + \langle z \rangle \right). \quad (72)$$

This model is particularly convenient because for a given  $\delta z$ ,  $s$ , the bias in mean redshift and standard deviation with respect to the model are  $\delta z$  and  $s$ . In our case, as the model is the true redshift distribution, we have

$$\delta z = \langle z \rangle - \langle z \rangle_{\text{true}}, \quad (73)$$

$$s = \sigma_n / \sigma_n^{\text{true}}. \quad (74)$$

As we use the true redshift distribution of the bin as a model (simulation only), the performance may be overestimated compared to a realistic scenario, where the model is taken from a Self-Organised-Map estimate (e.g. Hildebrandt et al. 2021), or a candidate from a simulation (e.g. Gatti et al. 2018, 2022; Cawthon et al. 2022). That is why we also consider a more direct method.

<sup>9</sup> We use the python package emcee (Foreman-Mackey et al. 2013).

### 3.5.5. Suppressed Gaussian process

We introduce the suppressed-Gaussian-process model (SGP) developed by Naidoo et al. (2023); GP reconstruction from clustering-redshifts was also performed in Johnson et al. (2017). The main benefit of this approach is that it is non-parametric. The procedure is as follows:

1. we generate realisations of Gaussian Process (GP) where the multivariate normal distributions are the measurements, and model the correlation between the points through a specific kernel<sup>10</sup>;
2. we apply to the realisations a suppression function that damps signals in regions where the measurements are consistent with zero;
3. the realisations are renormalised to unity;
4. the moments of the true distribution and their uncertainties are estimated from this large sample of normalised SGP realisations.

For the GP covariance, we chose a 3/2 Matern kernel function, and assume the length-scale  $l$  to be  $l = \Delta z/2$  where  $\Delta z$  corresponds to the redshift slicing<sup>11</sup>. To validate these choices, we tested different kernels and length-scale, by comparing generated GP from discrete  $n(z_i)$  and covariance to true  $n(z)$ ; see, for example, the right panel of Fig. 4.

For the second step, given a GP realisation  $i$ , the associated SPG is

$$n_{\text{SGP}}^i(z) = n_{\text{GP}}^i(z) S(x_i, k), \quad (75)$$

where  $S$  is a suppression factor that removes the noisy part of the distribution. We chose the parametrisation

$$S(x, k) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - (1 - x)^k & \text{if } 0 < x < 1 \\ 1 & \text{if } x > 1. \end{cases} \quad (76)$$

In the above, the  $x$  variable is a smoothing of the continuous signal-to-noise ratio (SNR)

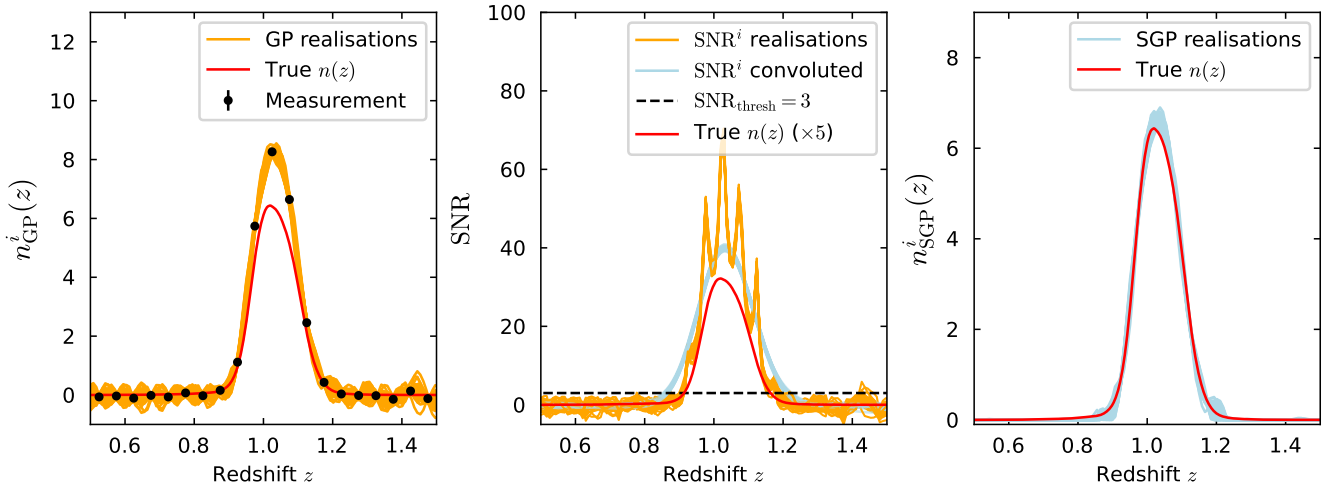
$$x_i = \frac{1}{\text{SNR}_{\text{thresh}}} \mathcal{G} * \frac{n_{\text{GP}}^i(z)}{\sigma_i(z)}. \quad (77)$$

Here  $\mathcal{G}$  is a Gaussian with a standard deviation  $\Delta z$ ,  $*$  represents the convolution operation, which smooths the SNR between data points (and avoids non-physical cuts), and  $\text{SNR}_{\text{thresh}}$  is the SNR threshold below which the suppressed factor is applied. Thus, if  $\text{SNR} < \text{SNR}_{\text{thresh}}$  over a redshift range larger than  $\Delta z$ , the amplitude of the realisation  $n_{\text{GP}}^i(z)$  would be reduced by a factor  $k/3 \times \text{SNR}$ . We fix  $k = 0.3$  in Eq. (76), so that the suppressed factor rapidly drops for  $x < 1$ , but we ensured that the impact of this parameter was small.

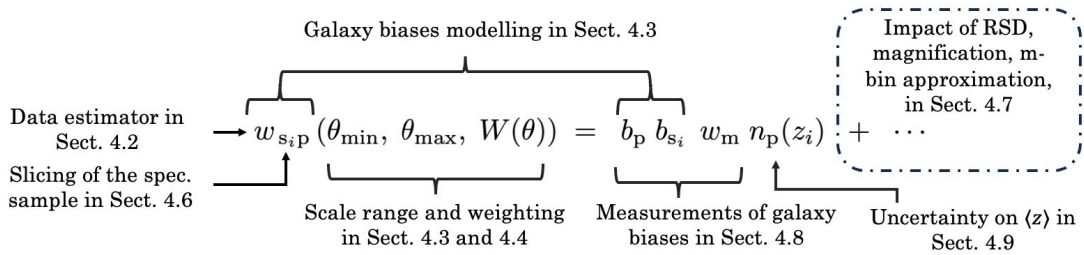
In Fig. 4 we illustrate the whole procedure. We show for a measurement  $\{n_p(z_i)\}_i$ ,  $C_n$  (not normalised to unity), GP realisations, their SNR  $n_{\text{GP}}/\sigma(z)$ , and the associated normalised SGP realisations compared with the true distribution. It is clear that going from the left to the right panel, the overall procedure greatly improves the realisations. Nonetheless, the procedure could be optimised further in the future, since we did not rigorously explore all of the SGP parametrisation.

<sup>10</sup> We use the python package Georges (Ambikasaran et al. 2015).

<sup>11</sup> The length-scale represents the characteristic distance over which the function values are correlated.



**Fig. 4.** *Left:*  $n(z)$  distribution measurement, with 50 associated GP realisations (orange), and the true distribution (red). *Middle:* SNR of the GP realisation (orange), and their convolution (blue). The SNR threshold ( $=3$ ) is reported with the dashed line. *Right:* Associated SGP realisations, renormalised to unity (blue), compared to the true distribution (red).



**Fig. 5.** Modelling the cross-correlation between spectroscopic samples in redshift slices and photometric bins, and reference to the corresponding tests we performed to optimise the pipeline.

## 4. Results

In this section, we present the clustering-redshifts pipeline optimisation. In Fig. 5 we summarise the different ingredients entering the modelling of clustering redshifts, and the corresponding validation tests presented in this section. In Sect. 4.1, we describe the simulated data used. In Sect. 4.2, we compare the performance of the clustering estimator. In Sect. 4.3, we evaluate the hypothesis that the galaxy biases of cross-correlations and auto-correlations are the same at small scales. In Sect. 4.4, we determine the optimal scale range, and in Sect. 4.5, the optimal scale weighting. In Sect. 4.6, we test different spectroscopic slicing  $\Delta z$  for measurements affected by systematics, with the effects of these systematics tested individually in Sect. 4.7. In Sect. 4.8, we compare different galaxy bias correction methods. Finally, in Sect. 4.9, we investigate the impact of the tomographic bin definition on the redshift constraints.

### 4.1. Investigation strategy

Our investigation strategy consists of varying a set of analysis parameters to check the impact on clustering redshifts. We chose to use three tomographic bins covering the full redshift range of interest and corresponding to different types of galaxies.

The three tomographic bins: “low- $z$ ”, “mid- $z$ ”, and “high- $z$ ”, were generated within the Flagship simulation with the photo- $z$  cuts:

$$0.35 < z_p < 0.47, \quad (78)$$

$$0.86 < z_p < 1, \quad (79)$$

$$1.30 < z_p < 1.6. \quad (80)$$

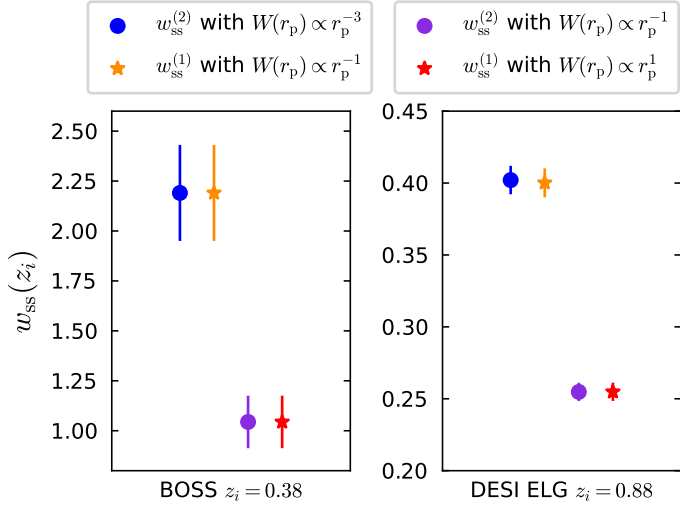
They corresponds to bin 2, 7, and 10 in Fig. 1. For reference samples, we used BOSS LOWZ and CMASS, DESI ELG, and *Euclid* NISP-S like samples. We report the redshift distribution of the tomographic bins and spectroscopic samples in Fig. 1.

To characterise the uncertainty from sample variance in our analysis, we split each tomographic bin into five independent Flagship2 sky patches<sup>12</sup> of 1000 deg<sup>2</sup>. As the simulation is limited to 5000 deg<sup>2</sup>, this is the maximum number of large independent patches that we can use. When investigating the optimal spectroscopic slicing we do not want our results to be biased because the points luckily cover the optimal part of the distribution. Thus, for every patch  $j = 1, \dots, 5$ , we varied the redshift slicing of the spectroscopic samples  $z_i$ ;  $\Delta z$ , by shifting the slice centres by  $j \times \Delta z/5$ . To summarise, we tested over five independent 1000 deg<sup>2</sup> sky patches for three tomographic bins, each cross-correlated with spectroscopic samples distributed into different slices for each patch (but with the same  $\Delta z$ ).

### 4.2. Choice of the estimator weighting scheme

In this section, we compare the two different estimators of the angular correlation function, described in Sect. 3.4.1. We measured the auto-correlations of spectroscopic samples for a 1000 deg<sup>2</sup> sky patch, a single redshift bin of width  $\Delta z = 0.05$ , and the two estimators:

<sup>12</sup> We applied cuts on the RA coordinates, with a width of  $\Delta RA = 17.5$  deg.



**Fig. 6.** Auto-correlations of two spectroscopic samples for a single redshift bin ( $\Delta z = 0.05$ ). The yellow and red points are obtained using the first estimator  $w_{ss}^{(1)}$  with  $\gamma \in \{-1, +1\}$ . The blue and purple points are obtained using the second one  $w_{ss}^{(2)}$  with  $\gamma \in \{-3, -1\}$ . The consistency of the points with different weights illustrates the effective weighting due to the estimator definition.

- $w_{ss}^{(1)}$  defined by Eq. (57), where weighting is applied to the ratio of pair counts;
- $w_{ss}^{(2)}$  defined by Eq. (58), where weighting is applied directly the pair counts.

Correlations were evaluated for the scale range 0.5–5 Mpc, and with different weighting functions  $W \propto r_p^\gamma$ . The results are reported in Fig. 6. We used logarithmic binning, and expect the measurements to be weighted by an effective function  $W_{\text{eff}} \propto r_p^{\gamma+2}$  (cf. Eq. (60)), which is confirmed by the matched amplitudes of blue-yellow and purple-red points. Importantly, we do not observe significant differences in the associated uncertainties between the two estimators with the corresponding weights. Given that the weighting variation does not offer any practical advantage, and leads to unnecessary complexity in the weighting process (cf. Appendix C), we adopted the first estimator, Eq. (57), as our fiducial choice for this analysis, and we will evaluate the optimal value of  $\gamma$ .

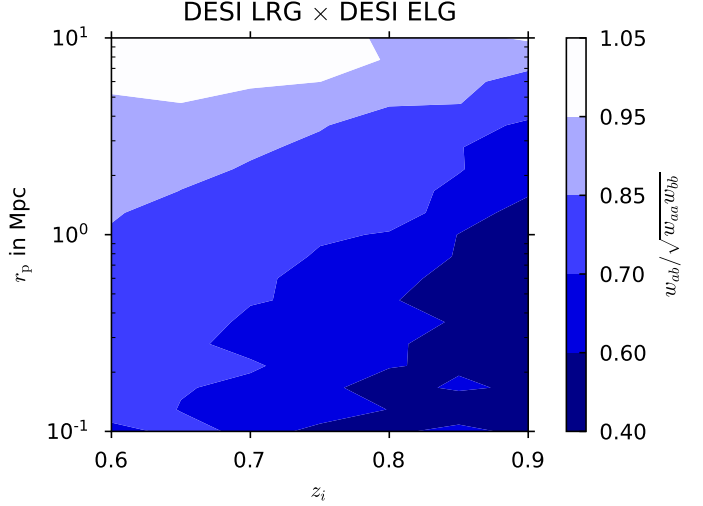
#### 4.3. Testing the bias-cancellation hypothesis

The scales used for clustering-redshifts calibration are typically very small [e.g. (0.1–5) Mpc] in order to minimise correlations between the calibration and the cosmological analysis, and to maximise the SNR. The associated nonlinear effects are often overlooked, however. Most authors assume that for two galaxy samples located within the same redshift bin, the two-point angular correlation function follows the relation:

$$w_{ab}(r_p) \approx \sqrt{w_{aa}(r_p) w_{bb}(r_p)}. \quad (81)$$

If we were using scalar product, instead of scale correlation function, this would be equivalent to the Cauchy–Schwarz equality, which implies that the density contrasts  $\Delta_a$  and  $\Delta_b$  are linearly dependent, and the Pearson correlation coefficient  $r_{ab}$  is 1, with

$$r_{ab} := \frac{\langle \Delta_a \Delta_b \rangle_{r_p}}{\sqrt{\langle \Delta_a \Delta_a \rangle_{r_p} \langle \Delta_b \Delta_b \rangle_{r_p}}} = 1. \quad (82)$$



**Fig. 7.** Variation in the Pearson coefficients  $r_{ab}$  for Flagship LRGs with ELGs. This coefficient informs on the small-scale under-correlation of blue and red galaxies.

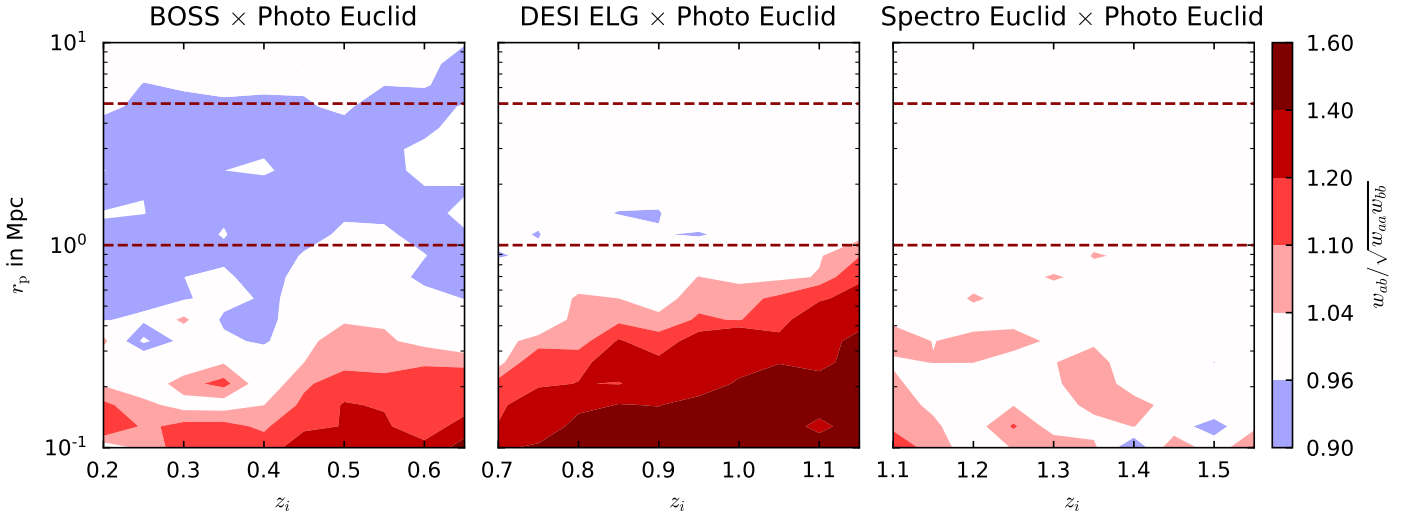
At large scales (>10 Mpc) galaxy samples act as linear tracers of the underlying dark matter density field. In this case,  $\delta_a = b_b/b_a \times \delta_b$ , which means that galaxy fields are linearly dependent. At smaller scales, the linearity between the dark matter field and galaxy statistics begins to break down. For a more complex biasing between matter field and galaxy field, however, we would still expect  $|r_{ab}| \leq 1$ <sup>13</sup>.

At even smaller scales (within the one-halo regime), the modelling of galaxy fields as random variables of the dark matter field breaks down (e.g. in Swanson et al. 2008, clustering properties between galaxy subpopulations was investigated), potentially impacting cosmological inference (e.g. Chaves-Montero et al. 2023, in the context of weak lensing). We illustrate this issue with an extreme example: one sample consists only of central galaxies, whilst the other consists only of satellite galaxies. If we assume there is at most one satellite galaxy per halo, the auto-correlation of central and satellite galaxies would be very small on scales smaller than the halo size, but still non-zero as we are not evaluating 3d clustering. The positive cross-correlation between the two would be strong, however. This would result in  $r_{ab} \gg 1$ , which is not possible for random variables. The reason for this discrepancy is that the statistics of the satellite distribution depend on those of the central galaxy distribution. In typical halo occupation distribution models, the probability of hosting a satellite galaxy is conditional on the presence of a central galaxy. Alternatively, if two galaxy samples populate different haloes, the cross-correlation would be very small on scales smaller than the halo size, while the auto-correlations would remain significant, leading to  $r_{ab} \ll 1$ .

The scale range mentioned earlier (referred to as “small” and “smaller”), and the behaviour of the  $r_{ab}$  coefficients are not clearly known. We decided to improve this aspect for our *Euclid* calibration method.

The small-scale effects that we have described are not a hypothetical framework, as this type of behaviour has already been observed in spectroscopic surveys. For instance, the first DESI data provides clear evidence that LRGs and ELGs exhibit

<sup>13</sup> As shown in Appendix A for quadratic non-linearities (second-order effects), we expect  $r_{ab} \approx 1$ , with the correction arising from third-order terms. Pushing further the perturbative development, we can show that indeed  $r_{ab} \leq 1$ , analogously to the Cauchy–Schwarz inequality.



**Fig. 8.** Pearson coefficient for our three simulated bins: low- $z$ , mid- $z$ , and high- $z$ . This term is assumed to be one for clustering-redshifts calibration. We see different behaviours,  $r = 1$ ,  $r < 1$ , and  $r > 1$  for different scales and redshifts, except for the high- $z$  bin, for which  $r = 1$  (almost) everywhere.

low cross-correlation signals at small scales (a phenomenon so-called ‘‘conformity’’; [Rocher et al. 2023](#); [Yuan et al. 2025](#); [Gao et al. 2024](#)). This occurs because these galaxies reside in different haloes, with their properties influenced by the history of the halo.

Given that the under-correlation between red and blue galaxies is well established, we aimed to test whether the Flagship simulation could reproduce this small-scale effect. Our strategy was to split the two samples into redshift slices ( $\Delta z = 0.05$ ) using true-redshift, and to measure the ratio with angular correlation,

$$r_{ab}(r_p, z_i) = \frac{w_{a_i b_i}(r_p)}{\sqrt{w_{a_i a_i}(r_p) w_{b_i b_i}(r_p)}}, \quad (83)$$

which is expected to be 1 under the linear approximation.

In [Fig. 7](#) we show  $r_{ab}(r_p, z)$  for the Flagship LRG and ELG-like samples. We recover the under-correlation at small scales, through  $r_{ab}(r_p, z) < 1$ . This ratio not only decreases with scales, but also with redshift. Since the recovered  $n(z)$  is degenerate with  $r$ , an evolving  $r$  with redshift would introduce a bias in the reconstructed  $n(z)$ . This confirms the presence of one-halo behaviour in the Flagship simulation. However, since these behaviours were only recently identified and are still being actively studied, it is likely that their implementation in the Flagship simulation is incomplete or not fully accurate. A thorough study of this aspect of the simulation and one-halo physics is beyond the scope of this article.

In [Fig. 8](#) we report  $r_{ab}(r_p, z)$  for the three simulated photometric  $\times$  spectroscopic bins (defined in [Sect. 4.1](#)). We observe significant scale and redshift evolution of the ratio, at least for the BOSS and DESI correlations, at scales smaller than 1 Mpc. Notably, we find  $r > 1$ , which we attribute to the physics of galaxy-halo connection and galaxy formation rather than non-linearities in the matter field (cf. the beginning of this subsection). Proving whether this is due to limitations in the HOD modelling or is a reliable prediction of the simulation is beyond the scope of our work. There is also a weak under-correlation for the low- $z$  bin (and the mid- $z$ ) at scales (1–8) Mpc, but it does not appear to evolve with redshift, and is therefore not expected to introduce a bias into the  $n(z)$ . This may correspond to the correction for non-linearities described above ( $r \lesssim 1$ ). For the high- $z$  bin, we do not measure deviation from unity at any scale.

Given these results, we are confident that using projected scales larger than 1.5 Mpc would avoid this one-halo effect<sup>14</sup>. As simulations improve in their modelling of galaxy halo properties, this test will need to be repeated. The ratio amplitude, scale, and redshift evolution are different depending on the tracers: under-correlation for DESI LRG and ELG in [Fig. 7](#), and over-correlation for *Euclid* photo and DESI ELG in [Fig. 8](#). We recommend to use only one type of spectroscopic tracer for the sample  $s$  when evaluating cross-correlations between photometric and spectroscopic samples at a specific redshift:  $w_{sp}(z_i)$ , avoiding mixing ELGs and LRGs for example. If different spectroscopic tracers  $\{s_j\}_j$  are available for the same redshift  $z_i$ , we can then predict different  $n_p(z_i | s_j)$  and combine them (e.g. [Hildebrandt et al. 2021](#)). Small-scale effect could lead to incompatibility as we show in this section, so that this procedure is more likely to detect systematics, and more robust.

#### 4.4. Choice of the scale range

In this section, we investigate the best scale range that minimises the deviation

$$\Delta n(z_i | r_p) = n_p(z_i | r_p) - n_{\text{true}}(z_i), \quad (84)$$

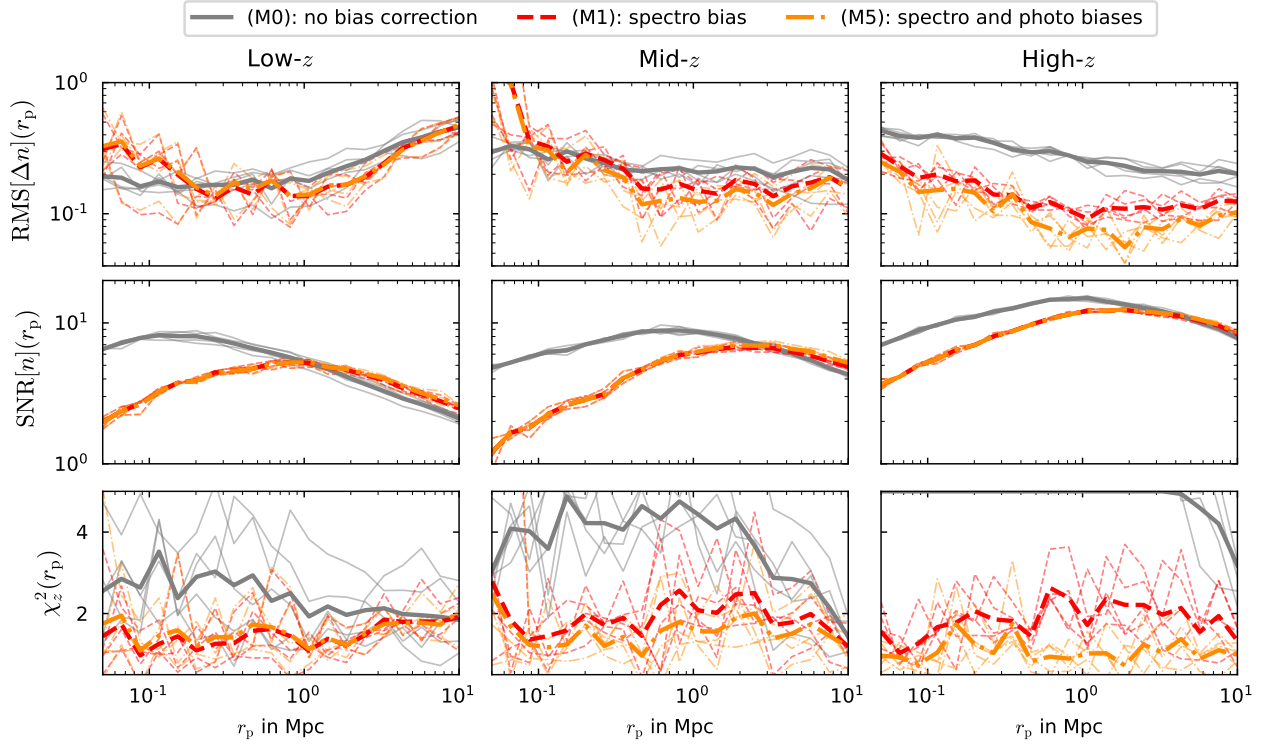
while maximising the SNR. We explored a wide range of scales, including those previously discarded in [Sect. 4.3](#). Unlike the estimator of  $w$  in [Eq. \(57\)](#), which averages over scales, our estimator was evaluated for a single scale<sup>15</sup> as in [Eq. \(48\)](#). Consequently, we refer to the distribution measured as  $n_p(z | r_p)$ . We always normalised the measured distributions to unity. We employed three methods for correcting the bias, covering all the bias correction cases: no correction (i.e. M0), spectroscopic correction (i.e. M1), and full correction (i.e. M5)<sup>16</sup>. This approach allows us to assess how bias correction influences the optimal scale range.

We used three statistical indicators, as a function of the projected scale

<sup>14</sup> We note that we are using angular correlation, so most of the 3-D separation between galaxies is much larger than this 1.5 Mpc.

<sup>15</sup> In practice, it is evaluated for a small scale bin.

<sup>16</sup> We expect the results with M2-3-4 to be similar to that of M5.



**Fig. 9.** RMS of  $\Delta n$ , SNR of  $n$ , and reduced  $\chi^2$  of  $\Delta n$  (top to bottom) as a function of the projected scale  $r_p$  for the three redshift bins (left to right) for three bias corrections (colours). The fine lines show five independent realisations of  $1000 \text{ deg}^2$ , and the thick lines show their means.

- the root mean square

$$\text{RMS}[\Delta n](r_p) = \sqrt{\frac{1}{N_z} \sum_{z_i} \Delta n(z_i | r_p)^2}; \quad (85)$$

- the (mean) signal-to-noise-ratio

$$\text{SNR}[n](r_p) = \frac{1}{N_z} \sum_{z_i} \frac{|n_p(z_i)|}{\sigma_n(z_i)}; \quad (86)$$

- the reduced  $\chi^2$

$$\chi_z^2(r_p) = \frac{1}{N_z} \left[ \Delta n(z_i | r_p) \right]_{z_i}^T C_{n(r_p)} \left[ \Delta n(z_i | r_p) \right]_{z_i}, \quad (87)$$

with  $N_z$  the number of points (equivalently the number of redshift slices of the spectroscopic sample). These indicators, evaluated at scales ranging from (0.05–10) Mpc for the different bins and bias corrections, are reported in Fig. 9. The thin lines represent individual realisations of  $1000 \text{ deg}^2$  patches, while the thicker lines indicate their means.

We first discuss the impact of the bias correction. Whilst the no-bias correction model M0 shows a higher SNR for scales lower than 1 Mpc, its corresponding  $\chi_z^2$  value is very high, indicating that it is not accurate. Therefore, we focus exclusively on M1 and M5 for the remainder of this section. Both methods perform similarly for the SNR, but for mid- $z$  and high- $z$ , M5 provides a  $\chi_z^2$  closer to unity.

For the three tomographic bins, the RMS is increasing for smaller scales. This increase is associated with a lower SNR, and the  $\chi_z^2$  remains relatively constant. This suggests an overall consistency, and we can choose the scale range based on the SNR only. For the low- $z$  bin, it peaks at 1 Mpc and then decreases. For the other two bins, it appears to peak around 2 Mpc, with a small decrease for larger scales.

In conclusion, the scale range (1.5–5) Mpc appears to be well suited for clustering redshifts; it is nearly optimal regarding the SNR; the  $\chi_z^2$  is almost minimal for the whole range; it avoids the  $r_p < 1.5 \text{ Mpc}$  one-halo regime (excluded in Sect. 4.3); these scales are not used for cosmological analysis.

This scale range is the same as in Gatti et al. (2022) and Giannini et al. (2024), but significantly larger than in other works such as van den Busch et al. (2020), Cawthon et al. (2022), and Naidoo et al. (2023), which used scales of  $0.1\text{--}1 \text{ Mpc } h^{-1}$  or  $0.5\text{--}1.5 \text{ Mpc } h^{-1}$ . Early clustering-redshifts studies even advocated for using smaller scales, down to the kpc range (Ménard et al. 2013; Schmidt et al. 2013). We note that we have chosen scales that avoid the one-halo term and the range of scales used for cosmology. One way to be conservative on the small-scales bias effects would be to use only large scales, overlapping those used in the  $3 \times 2$  point analysis. In previous calibrations such as HSC, KIDS, and DES, scale choice was largely driven by the lack of spectroscopic sample (Hildebrandt et al. 2021), and the range was chosen to be at the peak of the SNR (Gatti et al. 2018). For *Euclid*, this will no longer be the case, and from Fig. 9 it is clear that larger scales can be used for calibration. However, it requires the generation of a more complex covariance to accurately propagate the correlation. This was proved to be informative through a Fisher formalism in Johnston et al. (2025), but never implemented in practice with simulated or real data, and is beyond the scope of this article.

#### 4.5. Choice of the scale weighting

In this section, we varied the weighting  $W(r_p)$  in Eq. 53 and evaluated the RMS and SNR defined in Eqs. (85) and (86). The scale range was fixed to (1.5–5) Mpc following the last section's conclusion. We tested three power laws

$$W(r_p) = \frac{r_p^\gamma}{\int_{r_p^{\min}}^{r_p^{\max}} dr_p r_p^\gamma} \text{ with } \gamma \in \{-1, 0, 1\}, \quad (88)$$

and two different weighting schemes:

- we scale-average the different correlation functions and compute their ratio, as in Eq. (53),

$$\tilde{w}_{ab}(z_i) = \int_{r_p^{\min}}^{r_p^{\max}} dr_p W_w(r_p) w_{ab}(r_p, z_i); \quad (89)$$

- we scale-average the estimated  $n(z_i|r_p)$  as in Eq. (55),

$$\tilde{n}(z_i) = \int_{r_p^{\min}}^{r_p^{\max}} dr_p W_n(r_p) n(z_i|r_p). \quad (90)$$

In Fig. 10 we show the comparison between the SNR and the RMS for the three tomographic bins, with three power laws, two weighting schemes, and two bias correction methods (M1 and M5). The small markers are the realisations for the five individual 1000 deg<sup>2</sup> patches, while the larger markers indicate their means. The region of interest covers the lower right part of the plots, as it corresponds to minimum deviation and high SNR. Notably, the RMS/SNR plot displays a linear trend, indicating that points with a smaller deviation are associated with a better SNR, which is consistent with the expectations for an unbiased measurement. For all three redshift bins, the optimal weighting is the inverse scale weighting (darker red marker), a similar result to Schmidt et al. (2013) and Ménard et al. (2013). The reason is that SNR is higher, while RMS is lower, for scales around 1.5 Mpc than 5 Mpc (cf. Fig. 9).  $W_w$  and  $W_n$  are similar for the RMS, but the first performs slightly better for the SNR. These findings hold for M1 and M5. Additionally, we confirm that correcting the photometric bias for the first redshift bin is not strictly necessary in terms of RMS, but it does lead to a better SNR.

Thus, the best weighting scheme appears to be the inverse scale weighting applied to correlation functions, combined with a full bias correction approach.

#### 4.6. Finding the optimal spectroscopic slicing

In the previous sections, we did not include systematics (e.g. magnification) in the measurement because we were questioning the optimal statistical choices on the two-point clustering itself, hence the use of statistical quantities such as RMS and SNR. In the rest of the article, we consider realistic measurements, including systematics, and evaluate their impact on the moments of the redshift distribution we aim to calibrate.

In this section, we varied the slicing of the spectroscopic sample  $\Delta z$ , cf. Sect. 3.2.2. On the one hand, narrow slices are more sensitive to corrections associated with bin modelling (cf. Sect. 3.2.3), magnification (cf. Sect. 3.1.4), or RSD (cf. Sect. 3.1.5). On the other hand, if the slices are too wide, there may not be enough points covering the distribution to properly measure the  $n(z)$ , the shape would be flattened (cf. Sect. 3.5.2 and Fig. 3), and the assumption that galaxy biases and redshift distributions are constant may break (cf. Sect. 3.2.2). Therefore, we aim to find a compromise between narrow slices affected by systematic errors and large slices that are not statistically optimal, and for which the modelling might not be precise enough. The impacts of the systematic errors are tested one by one in the next Sect. 4.7.

We introduced the systematics using magnified sky coordinates and fluxes for the photometric and spectroscopic samples, and spectroscopic redshift (including peculiar velocity) for the spectroscopic sample. We used M5 as the bias correction method. For extracting the distribution moments from the  $\{n(z_i)\}$  we employed the SGP  $n(z)$  reconstruction introduced in Sect. 3.5.5. We did so because, even with few points, fitting the true distribution (as with the SSM, cf. Sect. 3.5.4) is likely to yield good yet unrealistic results, whereas the SGP provides weaker yet realistic constraints.

We measured the mean (absolute) biases on the mean  $|\langle z \rangle - \langle z \rangle_{\text{true}}|$  and standard deviation  $|\sigma_z/\sigma_z^{\text{true}} - 1|$  of the  $n_p(z)$  for different slicing, where the mean is evaluated over the five independent patches. We reported the mean values and their associated mean uncertainties in Fig. 11. We also created a similar plot using the SSM and obtain comparable results except for the uncertainties, which remain constant across the different slicing. This confirms our previous statement and justifies the choice of the SGP.

First of all, the uncertainties ( $y$  axis) on mean redshift and shape increase with the slicing  $\Delta z$ , and we also observe larger deviation for wider slices. Regarding the shape reconstruction (related to  $\sigma_z$ ), narrow slices are not associated with larger deviations, except for the mid- $z$  bin. For the mean redshift, small slicing is associated with similar or larger deviations than  $\Delta z \sim 0.05$ , but not larger uncertainties. The optimum slicing appears to correspond to  $0.04 \leq \Delta z \leq 0.06$ .

Consequently, we decided to use a redshift slicing of  $\Delta z = 0.05$  for every bin, for the rest of the study. This is larger than the slicing used for DES (Gatti et al. 2022; Giannini et al. 2024), but the same as the previous *Euclid* clustering-redshifts paper (Naidoo et al. 2023). As illustrated by the sandy region in Fig. 11, we predict that the systematic errors for this redshift slicing will not shift the data points outside the *Euclid* requirement. Since the  $\Delta z = 0.05$  points are predominantly on the diagonal (except for the mid- $z$  mean- $z$ ), this suggests that the corrections are fairly small for this slicing: the parameters are not biased more than  $1\sigma$  away from the truth. The mid- $z$  is the only case where we might fail to fulfil the requirement if we do not correct for the systematic errors. We investigated the origin of this deviation in detail in the next section.

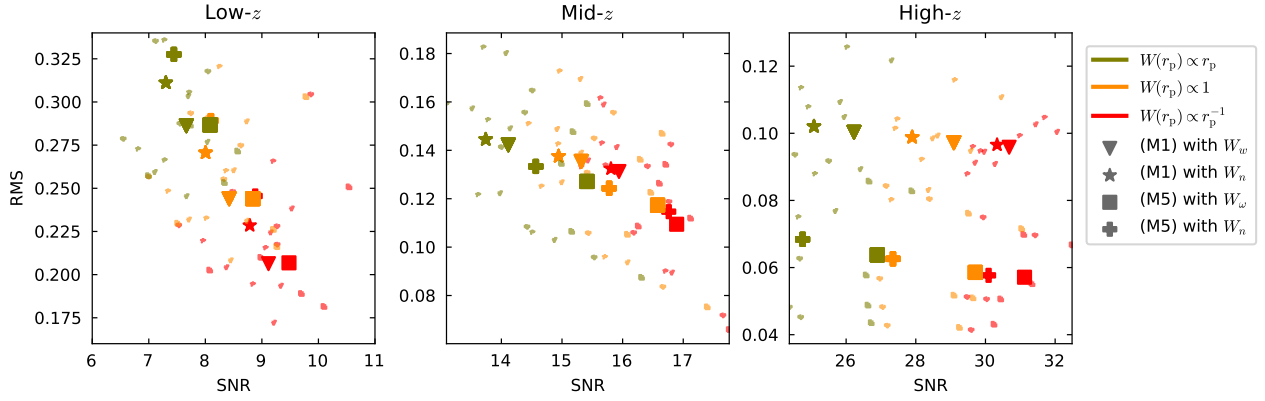
#### 4.7. Evaluating the (mis-)modelling of the clustering, magnification, and RSD

For clustering redshifts, we measure the cross-correlation of a wide photometric tomographic sample with several spectroscopic slices, and we assume that only the photometric galaxies within the redshift range of the spec- $z$  cuts contribute to a non-zero correlation. There can be three corrections:

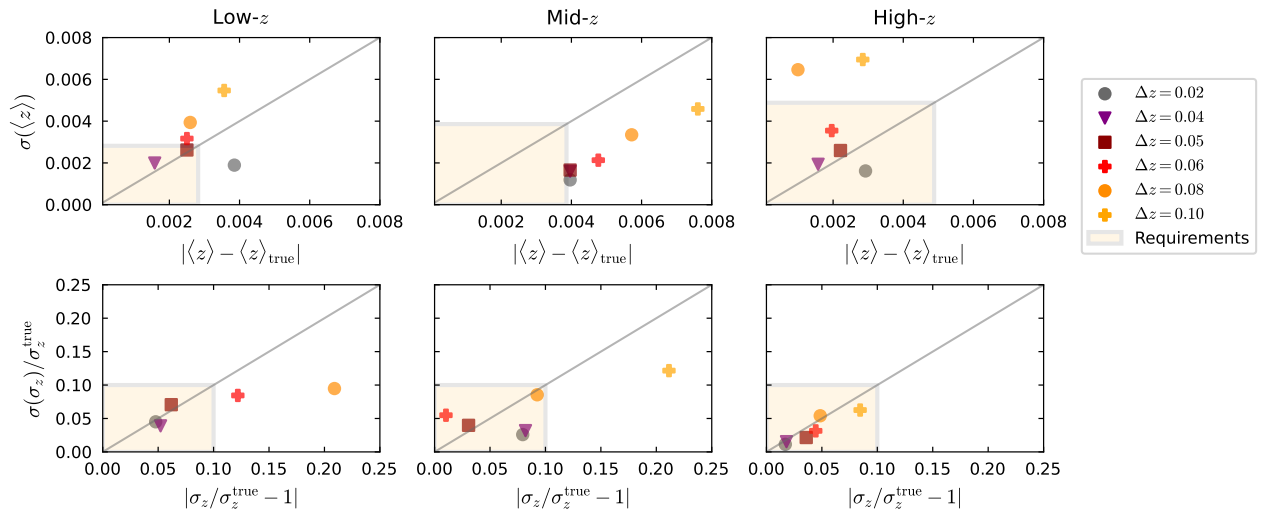
- galaxies in the nearby redshift (nearby bins) can correlate as well (we refer to this as the one-bin approximation);
- galaxies not covering the same redshift range can be correlated through magnification processes;
- spectroscopic redshifts can be different than true redshifts, leading to correction on the  $n_s$  at the boundary of the bin.

Two of these effects can be understood as bin border effects, and for larger slices they are expected to decrease in proportion. For wide spectroscopic slices, the assumption that  $n(z)$  and  $b(z)$  are constant within the bin might break down. We also note that spectroscopic interlopers were not included in the *Euclid* spectroscopic samples.

In the previous Sect. 4.6, we demonstrated that the compromise between having enough points  $n_p(z_i)$  and limiting these



**Fig. 10.** SNR and RMS for the three tomographic bins (left to right), with three power laws as weighting functions (colours), two weighting schemes, and two bias-correction methods (M1 and M5).



**Fig. 11.** Mean redshift and shape reconstruction with SGP, for the three tomographic bins. The different markers and colours are associated with different slicings  $\Delta z$ . The sandy region corresponds to the *Euclid* requirements, and the diagonal line shows a coherent deviation of  $1\sigma$ .

effects, corresponds to slicings  $\Delta z = 0.04$ – $0.06$ , regardless of the tomographic bin. To assess the influence of each effect, we used the idealised bias correction M5 to avoid the impact of galaxy bias, and proceed as follows:

- To test the impact of the one-bin approximation, we compared the cross-correlation  $w_{\text{sp}}(z_i)$  with spectroscopic slices and the tomographic bin for two cases. In the first one, referred to as the “full bin”, the photometric sample is the fiducial one used in this paper. For the second case, referred to as the “independent bins”, we used idealised tomographic bins. To generate these idealised tomographic bins, for every  $z_i$ , we keep only the photometric galaxies within the same redshift range as the spectroscopic galaxies (using  $\text{true-}z$ ), and then complete this sample with randoms to maintain the total number of galaxies as in the full tomographic bin. This ensures that only the galaxies within the redshift slice of the spectroscopic sample contribute to the correlation, for each spectroscopic slice.
- To test the magnification effect, we compared the results using the magnified variables for position and fluxes (over which we apply the samples cuts), or the unmagnified ones.
- To test the RSD effect, for the spectroscopic samples we used either the Flagship true redshifts, or the observed ones, the

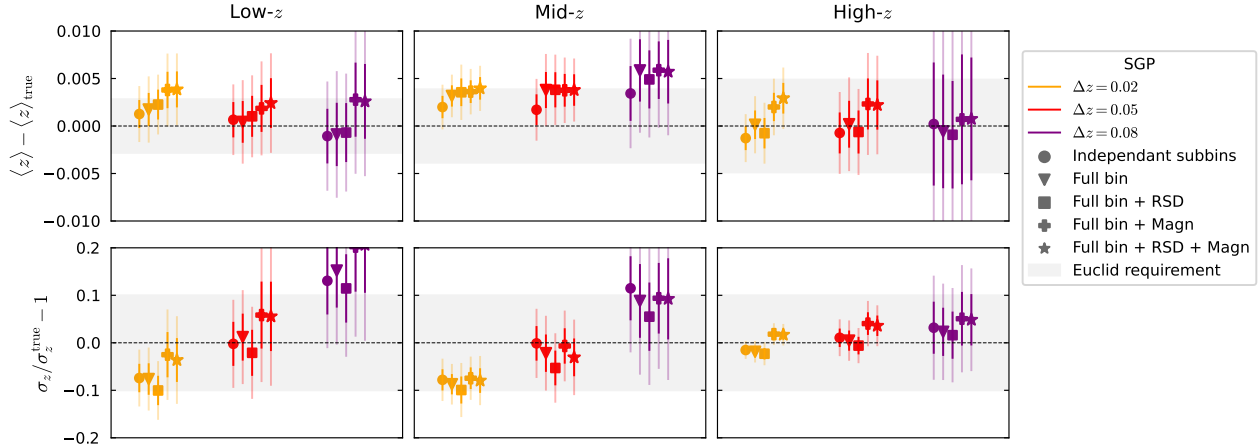
latter corresponding to a spectroscopic-like redshift estimate (including velocities).

- We also tested all these effects together, referred to as “Total”.

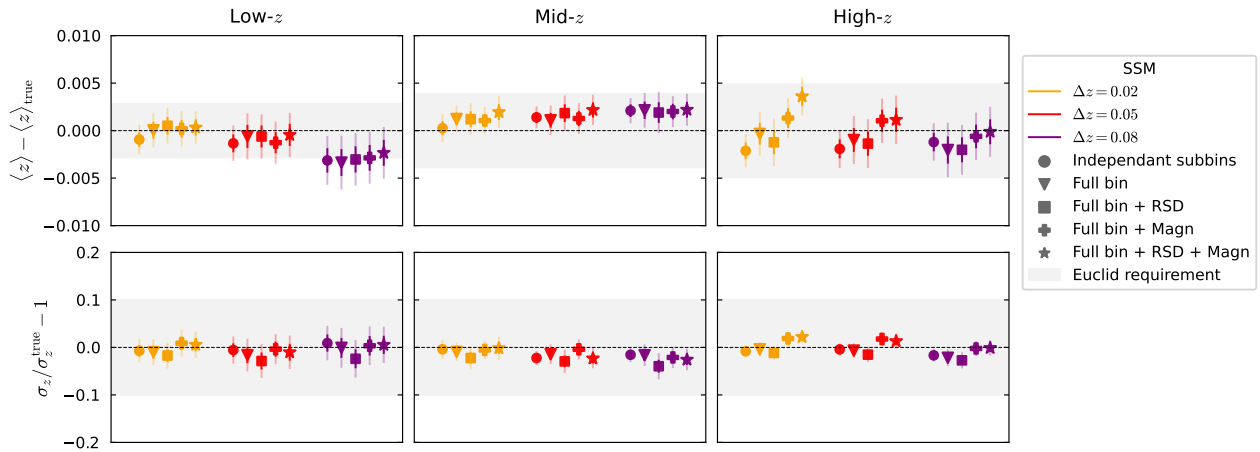
In Fig. 12 we show the deviations associated with the different systematic effects for different slicing, for the SGP. The same test for SSM methods is presented in Fig. 13. The corresponding values for the two procedures are listed in Table D.1 for  $\Delta z = 0.05$ , in Appendix D. For RSD and magnification, we have to consider the full tomographic bin, which by default comes with the one-bin approximation. Therefore, the full bin assumption has to be compared with “independent bins”, and the RSD or magnification has to be compared with the full bin.

For the SSM and SGP, with  $\Delta z = 0.05$ , the constraints on the redshift standard deviation  $\sigma$  are within the requirements independently from the effects. Only a large slicing of  $\Delta z = 0.08$  leads to some measurements out of the requirements because not enough points cover the distribution, and with larger slicing, the distribution appears flattened ( $\sigma_z/\sigma_z^{\text{true}} > 1$ ) as shown in Fig. 3.

Regarding the mean redshift bias, all the “independent bin” measurements are consistent with no deviation at the  $1$  to  $2\sigma$  level with the SGP and SSM. This suggests that our choice of scale and weighting is sufficient to limit the small-scale effects (cf. Sect. 4.4). In general, for the three tomographic bins, all the



**Fig. 12.** Deviations on the mean- $z$  (*top*) and standard deviation (*bottom*), with SGP fitting, for the three tomographic bins (*left, middle, and right*), with three  $\Delta z$  slicings (colours) and different implemented systematic effects (marker). The shadowed bands correspond to the *Euclid* requirement on both parameters. The dark and light error bars indicate 68.3% and 95.5% uncertainties



**Fig. 13.** Same as Fig. 12 for the two parameters of the SSM model, which are equivalent to  $\langle z \rangle - \langle z \rangle_{\text{true}}$  and  $\sigma_z / \sigma_{z,\text{true}} - 1$ .

systematics lead to some deviations compared to the independent bins, but distinguishing these variations from statistical fluctuations is complicated. It is clear that the SGP procedure is more sensitive to these effects. The reason for this is that with SSM we compare the measurements with a model free from these effects. As this model is, in our case, the truth, its results may be optimistic. Nevertheless, with a SOM estimate as a model, we might expect a similar low sensitivity to these effects. For smaller bin sizes, the error bars are smaller, but the deviation due to systematic effects is at least as large as for other sizes. Wider slices are associated with greater uncertainty, and the balance between accuracy and precision is achieved for  $\Delta z = 0.05$ . For this bin size, the systematic effects do not produce deviations that exceed the requirements. However, for the mid- $z$  bin, with SGP these systematic effects result in a deviation right at the limit of the budget on the mean- $z$ , as the bias is  $0.0038 \pm 0.0017$  (the requirements being 0.0038). For the optimistic case (SSM), the deviation remains within the requirements: the bias in mean redshift being  $0.0022 \pm 0.008$ . The effect causing this large deviation for SGP seems to be the one-bin approximation.

In Fig. 14, we report the true redshift distribution of three bins are shown alongside clustering-redshifts measurements for three slicing values:  $\Delta z = 0.02, 0.05, \text{ and } 0.08$ , excluding the effects of RSD and magnification. We observe increasing deviations with finer slicing, whilst the shape reconstruction becomes

less precise for coarser slicing (cf. Fig. 3). The small slicing deviations are higher for the mid- $z$  and high- $z$  bin, as one can expect since  $\eta_{\pm}$  increase with redshift. We also see that large slices lead to worst shape estimate for the low- $z$  bin, as its width is narrower.

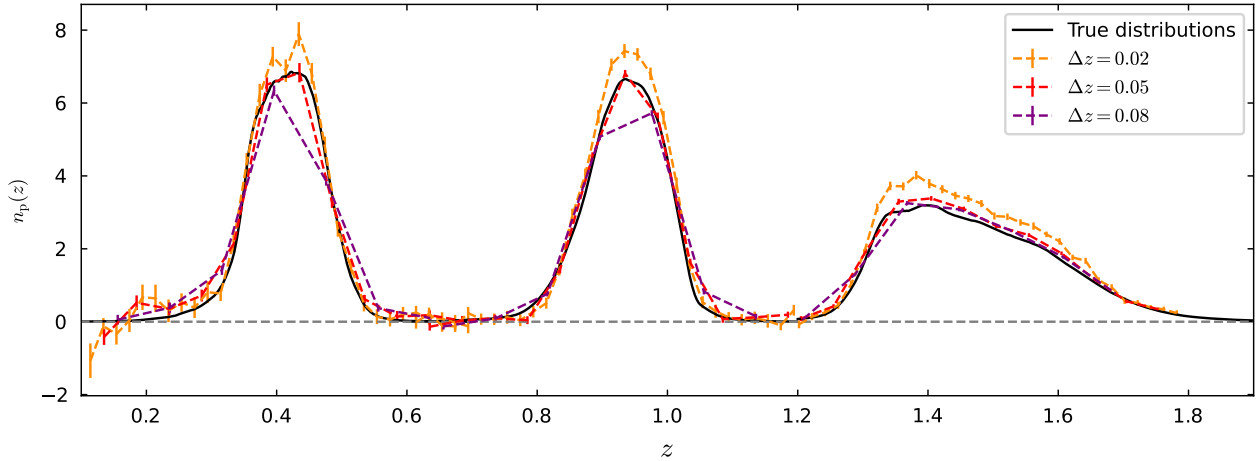
In Appendix E we explore an alternative modelling including correlation with neighbours slices (cf. Sect. 3.2.3). In short, despite our effort we did not manage to improve the accuracy of the modelling with extra-terms.

#### 4.8. Comparison of the different bias-correction methods

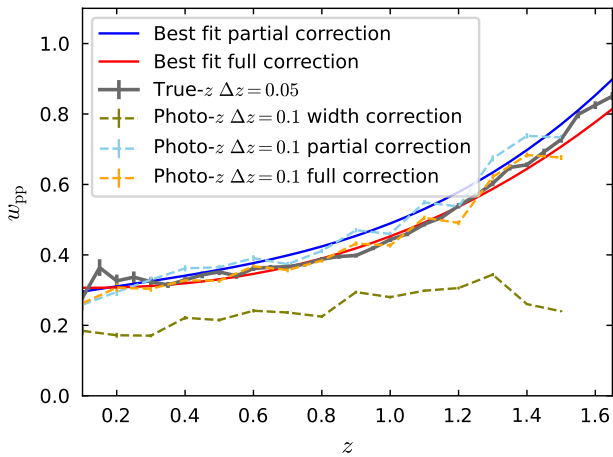
In this section, we explore the different possibilities for breaking the degeneracy between  $n_p(z)$  and  $b_p b_s$  from  $w_{\text{sp}}$  (cf. Eq. (47)). In the first subsection, we focus on measuring the photometric bias alone, since this is the challenging aspect.

##### 4.8.1. Measuring the photometric galaxy bias evolution

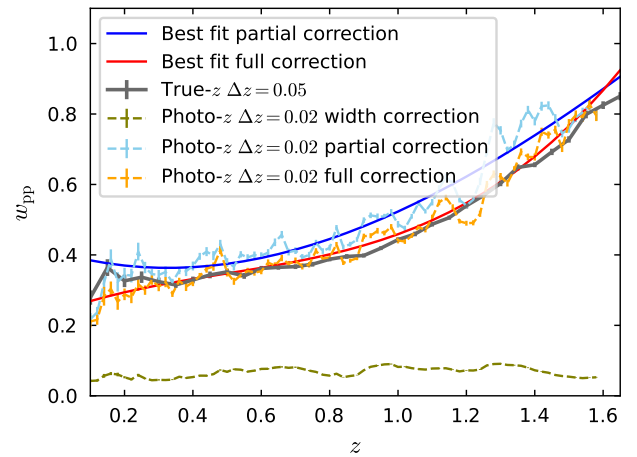
In Sect. 3.3.2, we proposed to improve existing photometric bias corrections (van den Busch et al. 2020; Cawthon et al. 2022) with two methods to measure  $b_p(z)$ , applicable to real data, specifically for photometric galaxies with  $z_{\text{photo}} < 1.6$  (i.e. fully covered by spectroscopic samples). The larger binning method M3 is applied in Fig. 15 and the smaller binning method M4 is applied in Fig. 16. They are represented by orange dashed lines.



**Fig. 14.** True-redshift distributions of tomographic bins in black, and the clustering-redshifts measurements with M5 for three spectroscopic slicings:  $\Delta z = 0.02, 0.05, 0.08$  (orange, red, and purple).



**Fig. 15.** True  $w_{pp}$  for  $\Delta z = 0.05$  in grey and measurement with corrections factors including M3 as dashed green, blue, and orange lines for photo- $z$  binning  $\Delta z = 0.1$ . The best fits from degree-3 polynomials are reported with solid lines. The  $w_{pp}$  used for M3 is the best-fit model of the full correction (solid red line).



**Fig. 16.** Same as Fig. 15, but for M4, with a smaller photo- $z$  binning  $\Delta z = 0.02$ . The  $w_{pp}$  used for M4 is the best-fit model of the full correction (solid red line).

For both figures we also include the results using either the partial correction (blue dashed) or width correction (green dashed), for comparisons with previous analyses (van den Busch et al. 2020; Cawthon et al. 2022; Giannini et al. 2024). Additionally, we present the best fit degree-3 polynomial for the partial and the full corrections (solid lines) from a standard  $\chi^2$  minimisation. The grey line shows the measurement using true- $z$ , which we want to recover after the corrections.

We find that assuming the photo- $z$  as the true- $z$  (width-only correction), neither the amplitude nor the redshift evolution is well measured. For the partial correction, whilst the redshift evolution appears well-recovered, differences in amplitude persist. Finally, our full correction method significantly outperforms the others: in both figures, we recover the amplitude and redshift evolution of the bias more accurately. Notably, the bias evolves weakly for  $z < 1$ , suggesting that correcting the photometric bias at low redshift may not be necessary.

Since the partial and full corrections both give a good trend for the redshift evolution of the galaxy bias, one could argue their performance are identical in the context of clustering redshifts, as the measured  $n_p(z)$  can always be latterly renormalised. With

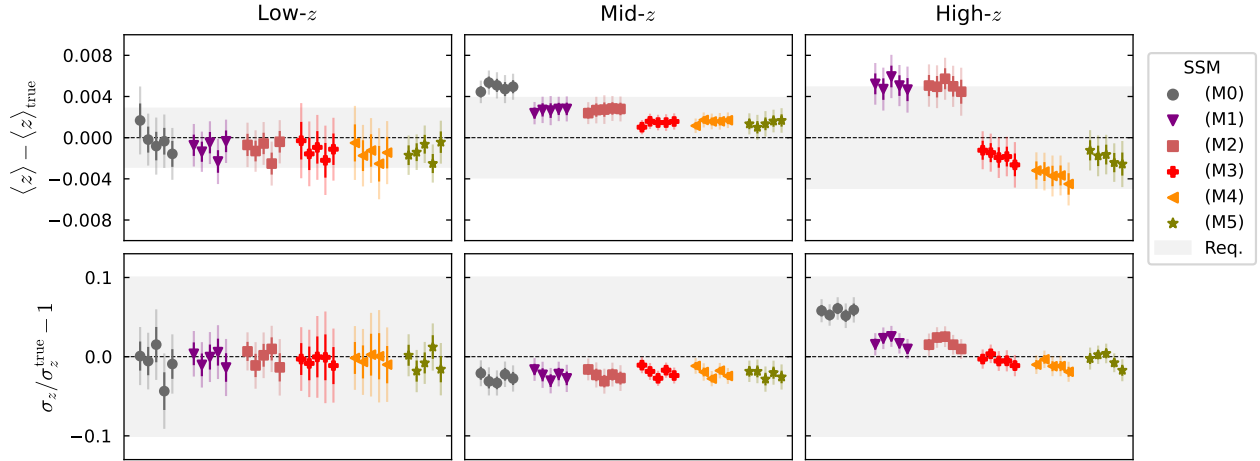
the correct amplitude for the galaxy bias, evaluating whether the  $n_p(z)$  measured is directly normalised to unity can provide valuable information. Deviations from unity could indicate

- a contamination from systematics, such as the small scale effect illustrated in Sect. 4.3;
- the photometric sample is not fully covered by spec (and what is the fraction);
- a problem with the spectroscopic redshifts, such as interlopers.

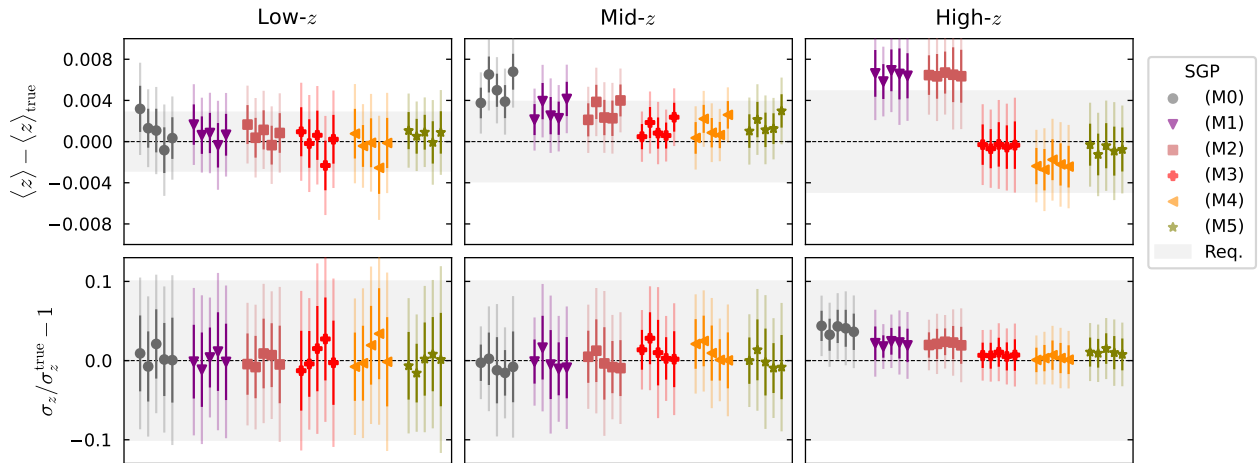
In that sense, the full correction performs significantly better than the partial one. The performance of our bias measurement may depend on the specific photo- $z$  code and the quality of the available photometry, and data implementation will have to address this question.

#### 4.8.2. Comparison of the methods

In this section, we compare the different bias correction methods introduced in Sect. 3.3.1, including M3 and M4. We aim to evaluate the bias correction performance, thus we work in an idealised framework with independent bins, no RSD, and no magnification (i.e. the “independent bin” test of Sect. 4.7). We employ



**Fig. 17.** Deviations with the SSM fitting on the mean redshift (upper panel) and standard deviation (lower panel) for the six methods M0–M5 in colour, for five sky patches (same colour). The requirements are represented by the shaded area. The dark and light error bars indicate 68.3% and 95.5% uncertainties.



**Fig. 18.** Same as Fig. 17 with the SGP fitting.

a redshift slicing  $\Delta z = 0.05$ . We report in colours the deviations on the mean redshift and standard deviation for the six methods M0–M5, in Figs. 17 (for SSM) and 18 (for SGP). For each method, we have five realisations by varying the sky patch. We also give the values for the biases on mean- $z$  and standard deviation in Table D.1, in Appendix D.

We can see that at low- $z$  it is not necessary to correct for the photometric bias, and spectroscopic correction M1 produces good estimates. At mid- $z$  it is necessary to correct at least for the spectroscopic bias, and ideally for both with M3 or M4. At high- $z$ , it is essential to correct both biases. We observe that M3 performs significantly better than M4 for SSM and SGP; indeed, the redshift evolution is better captured in Fig. 15 than in Fig. 16. This difference may be attributed to the fact that smaller slices are more sensitive to systematic effects, as discussed in Sect. 4.6. Overall, M3 performs similarly to the idealistic M5 and is consistently the best bias correction method, exhibiting performance akin to M1 for low- $z$  (though with slightly larger error bars).

In conclusion, using M3 for every bin yields unbiased estimates of the mean redshift and standard deviation across all bins, with uncertainties that fall within the requirements for both methods. The values presented in these two figures are reported in Table D.2.

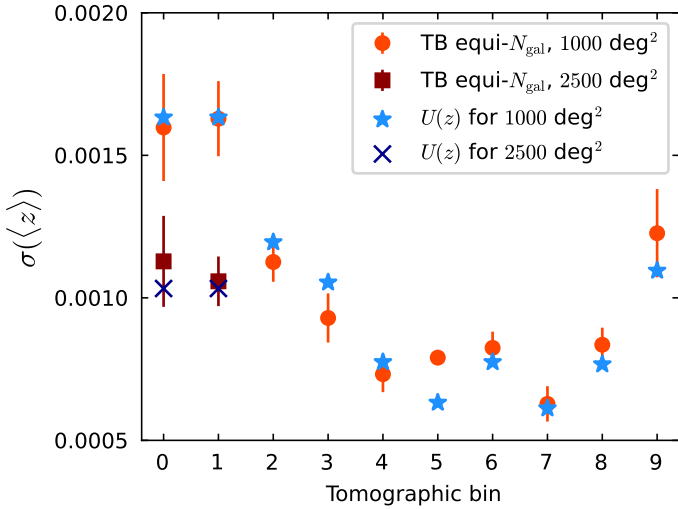
#### 4.9. Tomographic bin uncertainties

We defined the tomographic bins as equi-populated, based on the assumption that the uncertainty scales with the total number of galaxies in each bin. For the 10 spectroscopic samples, we used 1-DESI-BGs, 2-BOSS, 3-DESI-LRGs, 4-DESI-LRGs, 5-DESI-LRGs, 6-DESI-ELGs, 7-DESI-ELGs, 8-*Euclid*-NISP, 9-*Euclid*-NISP, 10-*Euclid*-NISP.

In Fig. 19 we plot in red the uncertainty on the mean redshift  $\sigma(\langle z \rangle)$  for every tomographic sample, and its standard deviation  $\sigma(\sigma(\langle z \rangle))$ , evaluated for the five 1000 deg<sup>2</sup> sky-patches with jackknife covariance, with SSM fitting. Interestingly, we find that the uncertainty is not constant. It remains the same for the first two samples, then decreases across the next six samples, before increasing again for the final ones. We also report for the first two samples, the case of 2500 deg<sup>2</sup> sky patches in dark red, and observe a mean- $z$  uncertainty reduction of 40%.

McQuinn & White (2013) predicted that “the linear bias times number of objects in a redshift bin generally can be constrained with cross-correlations to fractional error  $\approx \sqrt{10^2 N_{\Delta z} (dN_s/dz)^{-1}}$ ”, where  $N_{\Delta z}$  is the number of spectroscopic redshift slices, and  $N_s$  the number of spectroscopic galaxies<sup>17</sup>.

<sup>17</sup> We have slightly adjusted the convention to remain consistent with the rest of our article.



**Fig. 19.** Uncertainties on mean- $z$  with SSM fitting for five sky realisations in red. We also plot the best-fit uncertainty model in blue.

This expression describes how well the redshift distribution can be measured, whereas we are primarily interested in its moments. We assume that, at first order, the scaling should behave similarly, but with a different coefficient, which is complicated to predict because of the complexity of the mean- $z$  fitting procedure. As a consequence, the uncertainty should depend on the number of spectroscopic galaxies, rather than on the volume or the number of photometric galaxies. However, for the low- $z$  bins, the spectroscopic density is sufficiently high, especially with DESI-BG-like sample, that we might enter a cosmic variance (CV) regime. In this case, uncertainties are no longer sample-dominated (i.e. the shot noise  $1/n_s$  is subdominant). Thus, we introduce a maximum spectroscopic density to reach the cosmic variance limit  $n_{cv}$ , which in principle depends on  $z$ , but that we assume to be constant. Our uncertainty model is

$$U(z) = A \sqrt{\frac{N_{\Delta z}}{V_{TB} \min(n_{spec}, n_{cv})}}, \quad (91)$$

where  $A$  is a constant independent of the tomographic bin, and  $V_{TB}$  is the volume of the tomographic bin.

The equation above also indicates that the uncertainty scales with the inverse of the square root of the tomographic volume. This explains why in Fig. 8 of Naidoo et al. (2023), the uncertainty scales as a power law of the survey area  $A^\gamma$ , and the best-fit power law is close to  $\gamma = -1/2$ .

We constrained the  $A$  and  $n_{cv}$  coefficients with the set of tomographic bin measurements using a  $\chi^2$  procedure, and report in Fig. 19 the best-fit model in blue for 1000 and 2500  $\text{deg}^2$ . We observe a good agreement, particularly given that we neglected complications arising from galaxy biases and the mean-redshift fitting procedure. The uncertainty is largest for the low- $z$  tomographic bins where the survey volume is much smaller and precision is limited by the CV regime. For the high- $z$  samples, the uncertainty increases due to the reduced density of spectra, which is not sufficiently offset by the larger volume. The best-fit  $n_{cv}$  roughly corresponds to the DESI-BGS density divided by a factor of four (and the BOSS density by a factor of 1.8), and it only plays a role for the first two bins.

As a result, to achieve consistent uncertainty across bins with clustering- $z$ , the tomographic bin definition would

need to be adjusted according to an uncertainty model like Eq. (91). However, the tomographic bin definition is primarily influenced by cosmological constraining power (e.g. Euclid Collaboration: Pocino et al. 2021), and our results alone do not provide sufficient grounds to recommend a change in binning strategy. However, the precision of the mean redshift is one of the key limiting factors in cosmic shear analyses (see, e.g., Reischke 2024), and the impact of bin definitions on the accuracy of redshift calibration is often overlooked.

## 5. Application to the ten *Euclid* tomographic bins

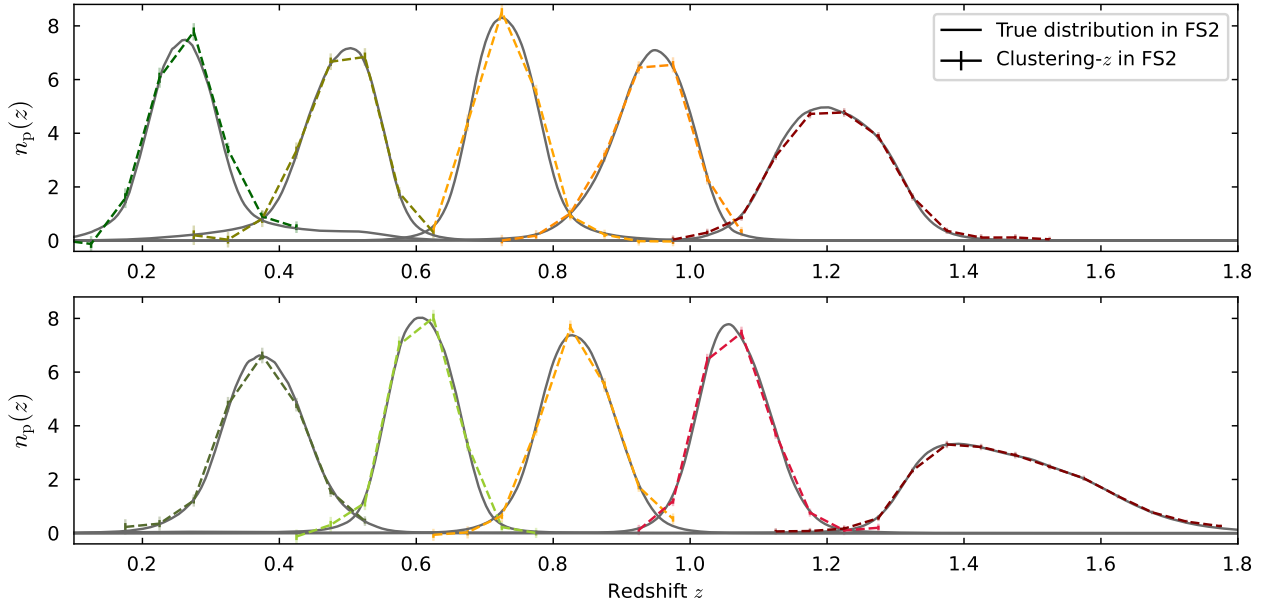
In this section, we applied our pipeline to calibrate the ten tomographic bins described in Sect. 2.2. For the spectroscopic samples, following our recommendation (see Sect. 4.3) we only correlated a tomographic bin with one spectroscopic tracer at a time. We used the following tracers for the different bins: 1-DESI-BGs, 2-BOSS, 3-DESI-LRGs, 4-DESI-LRGs, 5-DESI-LRGs, 6-DESI-ELGs, 7-DESI-ELGs, 8-*Euclid*-NISP, 9-*Euclid*-NISP, 10-*Euclid*-NISP. For simplicity, we did not consider the joint constraint of several spectroscopic tracers covering the same bin, which would provide additional information. For each bin, we produced several independent realisations by selecting distinct patches in Flagship 2. For the first two tomographic bins, we generated two 2500  $\text{deg}^{-2}$  patches, while for the remaining eight bins, we used five 1000  $\text{deg}^{-2}$  patches. This approach ensures similar uncertainties across bins, as explained in Sect. 4.9.

The true tomographic bin distribution, and the clustering-redshifts measurements are reported in Fig. 20. The measurements are visually indistinguishable from the truth. To quantify how good they are, we show the (optimistic) SSM calibration of the mean- $z$  and standard deviations in Fig. 21. The results with the conservative SGP calibration are shown in Fig. 22. For comparison, we also plot the previous results of Naidoo et al. (2023) for the mean redshift (with Flagship1 samples). Our results using the SSM method are fully successful; for every bin and patch realisation, we meet the *Euclid* requirements for the mean redshift and shape, unlike prior results. However, as expected, the "conservative" SGP method yields larger uncertainties and deviations. In particular, for some bins (especially bins 1, 3, and 7), we observe slightly larger deviations and uncertainties than allowed by the mean redshift requirement. We tested the effect of increasing the patch sizes by a factor of two and found that the mean- $z$  bias was reduced on average by 40%, which was sufficient to meet the requirement for all bins.

Part of the improvement to Naidoo et al. (2023) is due to the use of a larger cosmic volume with Flagship2 instead of Flagship1. Nevertheless, with our pipeline we also got significantly lower mean redshift biases than Naidoo et al. (2023). The main differences between our pipeline and the previous one are our estimator (cf. Sect. 4.2), the scale range (cf. Sect. 4.4), the measurements being evaluated for each spectroscopic tracer (cf. Sect. 4.3), the bias correction method (cf. Sect. 4.8), and some improvement in the implementation of SSM and SGP (cf. Sect. 3.5). We note that we did include the effect of magnification contrary to Naidoo et al. (2023). In conclusion, we predict that clustering-redshifts will be able to satisfy the *Euclid* requirements.

## 6. Conclusion

We optimised and evaluated the performance of the clustering-redshifts method for calibrating the redshift distributions of *Euclid* tomographic bins.



**Fig. 20.** True-redshift distribution of the simulated Flagship2 ten tomographic bins with  $z_p < 1.6$  (solid lines) and their clustering-redshifts measurements with our pipeline, including all systematics and realistic bias correction M3. The first two bin measurements are realised with 2500  $\text{deg}^2$  sky patches, and the other eight are realised with 1000  $\text{deg}^2$  sky patches. The dark and light error bars indicate 68.3% and 95.5% uncertainties, but are barely visible by eye.

Using the Flagship2 simulation, we generated realistic spectroscopic and photometric mock samples based on BOSS, DESI, and *Euclid*. The photometric sample was divided into ten equally populated tomographic bins, with a maximum photo- $z$  of 1.6. We used up to five 1000  $\text{deg}^2$  patches to fully exploit the volume of the simulation.

We measured the angular auto- and cross-correlations of these samples under various methodological choices and investigated alternative approximations to model the clustering.

From these clustering measurements, we calibrated the redshift distributions via two distinct methods. One method involved fitting the data points using a model  $n(z|p_1, \dots)$  with free parameters. In practice, a complementary calibration, such as the self-organising map can be used to define the model (e.g., Hildebrandt et al. 2021; Gatti et al. 2022). The other method relies on Gaussian processes to directly fit the data without assuming a specific model. With the latter, the SOM and clustering-redshifts calibrations can both be run independently, or be combined as in the case of DES (Myles et al. 2021; Gatti et al. 2022).

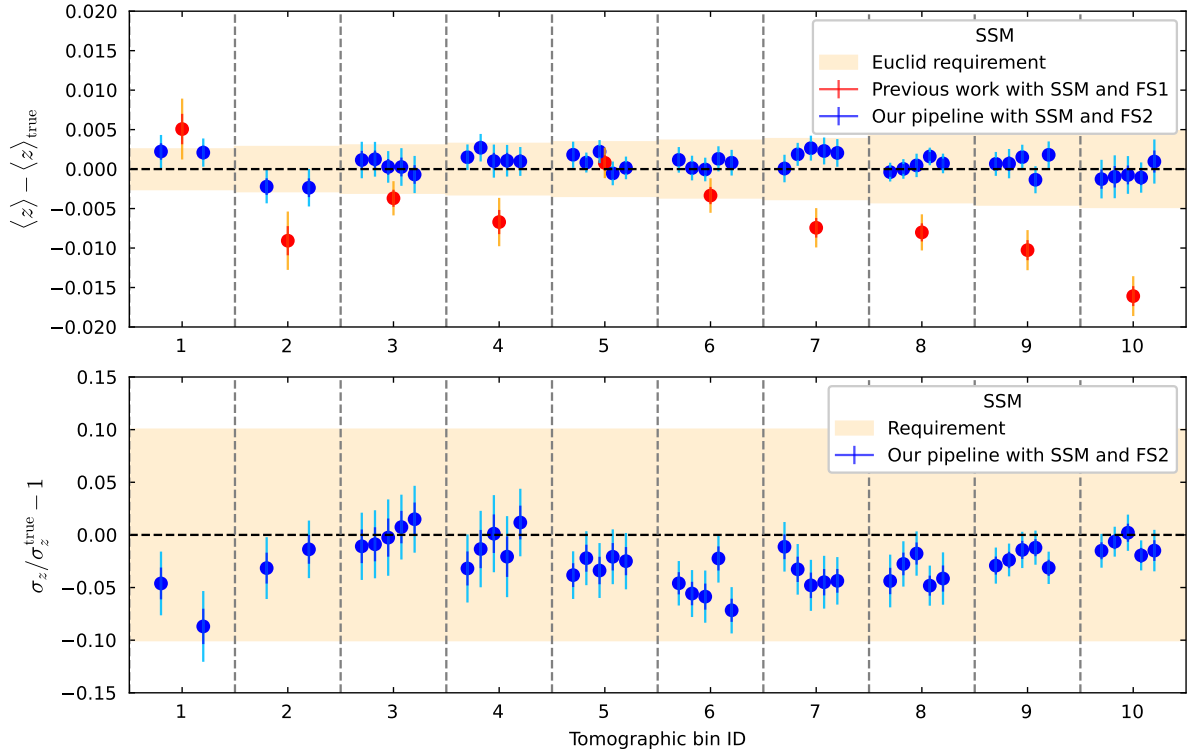
In Sect. 4 we investigated and concluded the following about the clustering-redshifts technique and angular clustering.

- In Sect. 4.2 we argued that the integrated pair-count estimator  $w^{(1)}$  is a better choice than  $w^{(2)}$  for measuring the integrated clustering (i.e. integrating the ratio DD/RR instead of taking the ratio of the integrated DD and RR).
- In Sect. 4.3 we explained a potential issue with very small scales that arises from the one-halo properties of the clustering. We concluded that scales should always be larger than 1.5 Mpc to limit this problem. We also recommended separating the clustering-redshifts measurements for the different spectroscopic tracers.
- In Sects. 4.4 and 4.5, we ran an optimisation of the scale range choice and weighting. With the Flagship simulation, we achieved a good compromise with  $1.5 < r_p < 5$  Mpc with an inverse scale weighting.

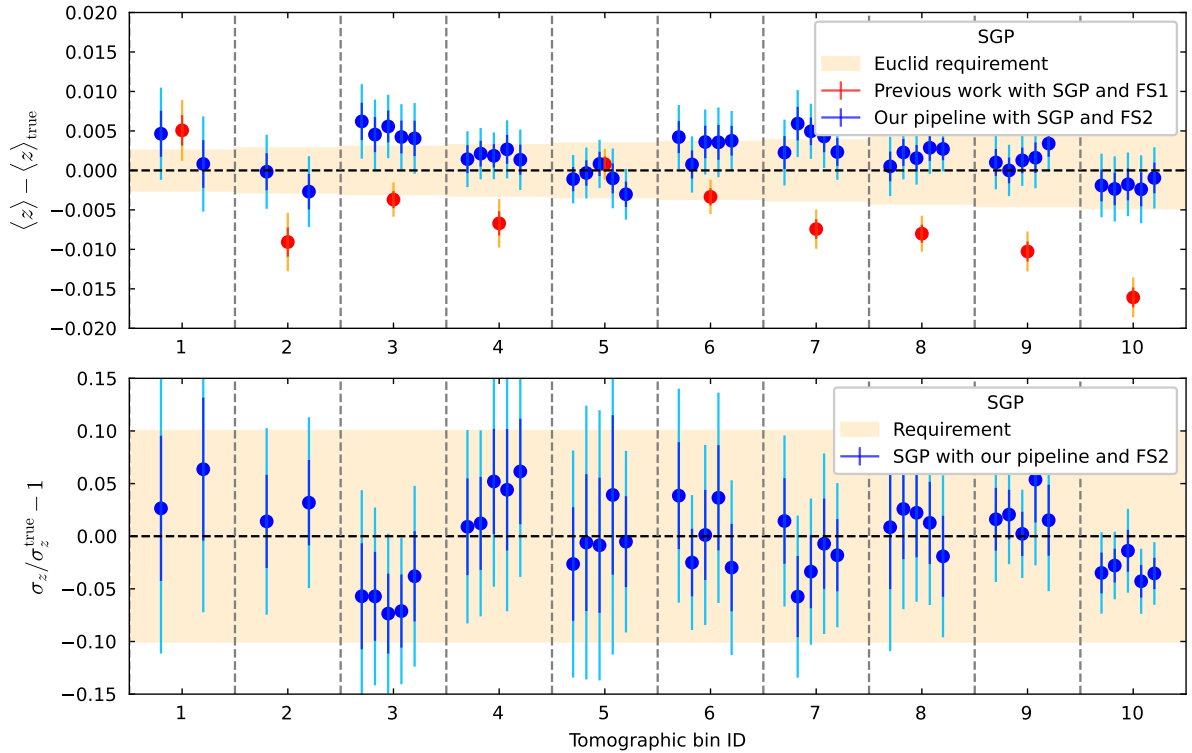
- In Sect. 4.6 we showed that the optimal spectroscopic slicing given our tomographic bins and the different systematics was  $\Delta z \approx 0.05$ .
- In Sect. 4.7 we showed that approximating  $\int dz n^2(z) \dots$  (Limber-full- $z$ ) with some  $n^2(z_i) \int dz \dots$  expressions (Limber-1-bin) is the main systematic effect and that magnification has some limited impact on the  $n_p(z)$  inference.
- In Sect. 4.8 we compared different bias corrections and introduced a way to measure the photometric galaxy bias by splitting the TB into smaller bins and interpolating these effective biases over redshift (i.e. M3). The results of this method were comparable to the ideal case M5, in which photometric bias is measured using true redshifts. It should be used as the fiducial correction scheme.
- Finally, we investigated in Sect. 4.9 the dependence of the mean- $z$  error bars on the clustering-redshifts parameters, and confirmed that they are mainly driven by the number of spectroscopic galaxies that is available in the footprint.

In Sect. 5 we applied the optimised pipeline to calibrate the *Euclid* simulated tomographic bins. The calibration performance using the SSM model met or exceeded the required accuracy, while the SGP method yielded comparable results, but they were slightly below the requirements in certain cases. It is important to note that these results were derived from limited sky patches with correction methods that can still be improved in future work. These results represent a significant improvement over the previous *Euclid* clustering-redshifts study (Naidoo et al. 2023) overall and demonstrated that the clustering-redshifts technique is robust.

Additionally, as discussed in Sect. 2, we assumed that the  $H\alpha$  sample had a purity of 100%. In reality, however, when we bin the sample into small redshift intervals, we expect that some of these galaxies may be interlopers, that is, they might be located at different redshifts due to noise or a misidentification of a spectral lines. This situation might significantly impact clustering-redshifts measurements because photometric



**Fig. 21.** Deviations in the mean redshift (*top*) and the redshift standard deviation (*bottom*) for the ten Flagship2 tomographic bins relative to the true-redshift distributions using the SSM method. The deviations are shown for each bin and sky patch. We consider patches of  $2500 \text{ deg}^2$  for the first two bins, and  $1000 \text{ deg}^2$  for the others. The *Euclid* requirements on the uncertainty are highlighted by the shaded area centred around zero. The uncertainties are consistent across all the bins for the mean redshift and we achieve a clear improvement from previous work shown in red. For the redshift standard deviations, the uncertainties are slightly underestimated, but this is within an acceptable range given the requirements.



**Fig. 22.** Deviations in the mean redshift (*top*) and the redshift standard deviation (*bottom*) using the SGP method for the same bin configuration as in Fig. 21. The *Euclid* requirements on the uncertainty are highlighted by the shaded area centred around zero. For some bins (particularly bins 1, 3, and 7), we observe slightly larger deviations and uncertainties than permitted on the mean redshift. However, a clear improvement compared to previous work (shown in red) is still evident. The redshift standard deviations are unbiased and fall within the required range (shaded sandy area).

galaxies at a similar redshifts as the interlopers would contribute to the clustering signal, which would bias the inferred  $n(z)$ . It is therefore crucial to determine the fraction of these interlopers and their redshift distribution to correct for this effect. The clustering-redshifts technique can indeed be used to constrain the properties of interlopers by cross-correlating the NISP sample with other spectroscopic data or with photometric data with priors on their  $n(z)$ .

As already mentioned in Sect. 4.3, the galaxy-halo connection implemented in the Flagship simulation might not be accurate, and our statistical estimator might be optimistic. It might therefore still be beneficial to use larger scales even though they are not statistically optimal. These larger scales will be used for the  $3 \times 2$  analysis, which necessitates the generation of a more complex covariance to accurately propagate the correlation with the calibration (e.g. the work of Johnston et al. 2025). Although this approach is more complicated, the advantage is that the distribution would be free of small-scale potential bias, which may prove worthwhile.

Finally, we restricted the calibration to galaxies with a maximum photo- $z$  of 1.6 because we lack dense spectroscopic samples above redshift 1.8, which sets a strict upper limit. Clustering- $z$  for HSC, KIDS, and DES was even more limited than *Euclid*, with a maximum redshift of  $z \sim 1$ . In the far future numerous spectroscopic samples from Lyman-break galaxies are expected to be observed by spectroscopic experiments such as DESI-II, MUST, and WST (if funded). In a much shorter time frame, QSOs might be used one could utilise QSOs to cover the redshift range  $1.5 < z < 3$ . These constraints alone would most likely not fulfil the *Euclid* requirements because QSOs have a low-density, however, but they would still provide valuable information.

**Acknowledgements.** The authors thank Alex Alarcon, Carles Sánchez, Elisa Chisari, and Santi Ávila for enlightening discussions, as well as Krishna Naidoo for his help, code, and results sharing. The authors also acknowledge Mathias Urbano for his code support. This work was supported by the MICINN projects PID2019-111317GB-C32, PID2021-123012NB-C41, PID2022-141079NB-C32, CNS2023-144328 (Carteu) as well as predoctoral program AGAUR-FI ajuts (2024 FI-1 00692) Joan Oró. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. This work has made use of CosmoHub, developed by PIC (maintained by IFAE and CIEMAT) in collaboration with ICE-CSIC. It received funding from the Spanish government (grant EQC2021-007479-P funded by MCIN/AEI/10.13039/501100011033), the EU NextGeneration/PRTR (PRTR-C17.I1), and the Generalitat de Catalunya. 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 Agenzia Spaziale Italiana, the Austrian Forschungsförderungsgesellschaft funded through BMIML, the Belgian Science Policy, the Canadian Euclid Consortium, the Deutsches Zentrum für Luft- und Raumfahrt, the DTU Space and the Niels Bohr Institute in Denmark, the French Centre National d'Etudes Spatiales, the Fundação para a Ciência e a Tecnologia, the Hungarian Academy of Sciences, the Ministerio de Ciencia, Innovación y Universidades, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Nederlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Research Council of Finland, 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 ([www.euclid-ec.org](http://www.euclid-ec.org)).

## References

Adame, A. G., Aguilar, J., Ahlen, S., et al. 2024, *AJ*, 167, 62  
 Addison, G. E., Bennett, C. L., Jeong, D., Komatsu, E., & Weiland, J. L. 2019, *ApJ*, 879, 15  
 Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2015, *IEEE Trans. Pattern Anal. Mach. Intell.*, 38, 252  
 Baleato Lizancos, A., & White, M. 2023, *JCAP*, 07, 044  
 Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, *ApJ*, 762, 109  
 Blas, D., Lesgourgues, J., & Tram, T. 2011, *JCAP*, 07, 034  
 Bordoloi, R., Lilly, S. J., & Amara, A. 2010, *MNRAS*, 406, 881

Campos, A., Yin, B., Dodelson, S., et al. 2024, ArXiv e-prints [arXiv:2408.00922]  
 Carretero, J., Tallada, P., Casals, J., et al. 2017, in *Proceedings of the European Physical Society Conference on High Energy Physics. 5-12 July*, 488  
 Cawthon, R., Elvin-Poole, J., Porredon, A., et al. 2022, *MNRAS*, 513, 5517  
 Chaves-Montero, J., Angulo, R. E., & Contreras, S. 2023, *MNRAS*, 521, 937  
 Chisari, N. E., Alonso, D., Krause, E., et al. 2019, *ApJS*, 242, 2  
 Contarini, S., Verza, G., Pisani, A., et al. 2022, *A&A*, 667, A162  
 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  
 DESI Collaboration (Aghamousa, A., et al.) 2016, ArXiv e-prints [arXiv:1611.00037]  
 Elvin-Poole, J., MacCrann, N., Everett, S., et al. 2023, *MNRAS*, 523, 3649  
 Eriksen, M., Alarcon, A., Cabayol, L., et al. 2020, *MNRAS*, 497, 4565  
 Euclid Collaboration (Blanchard, A., et al.) 2020, *A&A*, 642, A191  
 Euclid Collaboration (Desprez, G., et al.) 2020, *A&A*, 644, A31  
 Euclid Collaboration (Ilbert, O., et al.) 2021, *A&A*, 647, A117  
 Euclid Collaboration (Pocino, A., et al.) 2021, *A&A*, 655, A44  
 Euclid Collaboration (Castander, F., et al.) 2025, *A&A*, 697, A5  
 Euclid Collaboration (Cropper, M., et al.) 2025, *A&A*, 697, A2  
 Euclid Collaboration (Jahnke, K., et al.) 2025, *A&A*, 697, A3  
 Euclid Collaboration (Le Brun, V., et al.) 2025, *A&A*, submitted  
 Euclid Collaboration (Mellier, Y., et al.) 2025, *A&A*, 697, A1  
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306  
 Gao, H., Jing, Y. P., Xu, K., et al. 2024, *ApJ*, 961, 74  
 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  
 Giannini, G., Alarcon, A., Gatti, M., et al. 2024, *MNRAS*, 527, 2010  
 Gwyn, S., McConnachie, A. W., Cuillandre, J. C., et al. 2025, ArXiv e-prints [arXiv:2503.13783]  
 Hartlap, J., Simon, P., & Schneider, P. 2007, *A&A*, 464, 399  
 Hartley, W. G., Chang, C., Samani, S., et al. 2020, *MNRAS*, 496, 4769  
 Hildebrandt, H. 2016, *MNRAS*, 455, 3943  
 Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, *A&A*, 647, A124  
 Huterer, D., Takada, M., Bernstein, G., & Jain, B. 2006, *MNRAS*, 366, 101  
 Jarvis, M. 2015, *Astrophysics Source Code Library* [record ascl:1508.007]  
 Johnson, A., Blake, C., Amon, A., et al. 2017, *MNRAS*, 465, 4118  
 Johnston, H., Elisa Chisari, N., Joudaki, S., et al. 2025, *A&A*, 699, A127  
 Krause, E., Fang, X., Pandey, S., et al. 2021, ArXiv e-prints [arXiv:2105.13548]  
 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]  
 Lepori, F., Schulz, S., Tutusaus, I., et al. 2025, *A&A*, 694, A321  
 Pandey, S., Sánchez, C., Jain, B., & LSST Dark Energy Science Collaboration 2025, *Phys. Rev. D*, 111, 023527  
 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  
 Naidoo, K., Johnston, H., Joachimi, B., et al. 2023, *A&A*, 670, A149  
 Newman, J. A. 2008, *ApJ*, 684, 88  
 Nicola, A., Hadzhiyska, B., Findlay, N., et al. 2024, *JCAP*, 02, 015  
 Palanque-Delabrouille, N., Magneville, C., Yèche, C., et al. 2016, *A&A*, 587, A41  
 Pandey, S., Krause, E., Jain, B., et al. 2020, *Phys. Rev. D*, 102, 123522  
 Percival, W. J., Friedrich, O., Sellentin, E., & Heavens, A. 2022, *MNRAS*, 510, 3207  
 Rau, M. M., Dalal, R., Zhang, T., et al. 2023, *MNRAS*, 524, 5109  
 Reischke, R. 2024, *MNRAS*, 530, 4412  
 Rocher, A., Ruhlmann-Kleider, V., Burtin, E., et al. 2023, *JCAP*, 10, 016  
 Roster, W., Wright, A. H., Hildebrandt, H., et al. 2025, *ApJ* submitted [arXiv:2508.02779]  
 Salvato, M., Ilbert, O., & Hoyle, B. 2019, *Nat. Astron.*, 3, 212  
 Schmidt, S. J., Ménard, B., Scranton, R., Morrison, C., & McBride, C. K. 2013, *MNRAS*, 431, 3307  
 Scottze, V., Benoit-Lévy, A., Coupon, J., Ilbert, O., & Mellier, Y. 2018, *MNRAS*, 474, 3921  
 Simon, P. 2007, *A&A*, 473, 711  
 Stözlner, B., Joachimi, B., Korn, A., Hildebrandt, H., & Wright, A. H. 2021, *A&A*, 650, A148

- Swanson, M. E. C., Tegmark, M., Blanton, M., & Zehavi, I. 2008, *MNRAS*, **385**, 1635
- Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, *ApJ*, **761**, 152
- Tallada, P., Carretero, J., Casals, J., et al. 2020, *Astron. Comput.*, **32**, 100391
- 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
- Verdier, A., Rocher, A., Bandi, B., et al. 2025, ArXiv e-prints [arXiv:2508.07311]
- Wright, A. H., Hildebrandt, H., van den Busch, J. L., & Heymans, C. 2020a, *A&A*, **637**, A100
- Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020b, *A&A*, **640**, L14
- Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2025, *A&A*, submitted [arXiv:2503.19440]
- Yèche, C., Palanque-Delabrouille, N., Claveau, C.-A., et al. 2020, *Res. Notes Am. Astron. Soc.*, **4**, 179
- Yuan, S., Wechsler, R. H., Wang, Y., et al. 2025, *MNRAS*, **538**, 1216
- <sup>1</sup> Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
- <sup>2</sup> Serra Hünter Fellow, Departament de Física, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain
- <sup>3</sup> Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France
- <sup>4</sup> Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
- <sup>5</sup> Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
- <sup>6</sup> Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
- <sup>7</sup> Port d'Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain
- <sup>8</sup> Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
- <sup>9</sup> Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290 Versoix, Switzerland
- <sup>10</sup> Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
- <sup>11</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), Pas-seig de Luís Companys 23, 08010 Barcelona, Spain
- <sup>12</sup> ESAC/ESA, Camino Bajo del Castillo, s/n., Urb. Villafraña del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
- <sup>13</sup> School of Mathematics and Physics, University of Surrey, Guildford, Surrey, GU2 7XH, UK
- <sup>14</sup> INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
- <sup>15</sup> INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
- <sup>16</sup> IFPU, Institute for Fundamental Physics of the Universe, Via Beirut 2, 34151 Trieste, Italy
- <sup>17</sup> INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
- <sup>18</sup> INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste TS, Italy
- <sup>19</sup> SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste TS, Italy
- <sup>20</sup> Centre National d'Etudes Spatiales – Centre spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
- <sup>21</sup> Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
- <sup>22</sup> INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- <sup>23</sup> INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, 35122 Padova, Italy
- <sup>24</sup> Dipartimento di Fisica, Università di Genova, Via Dodecaneso 33, 16146, Genova, Italy
- <sup>25</sup> INFN-Sezione di Genova, Via Dodecaneso 33, 16146, Genova, Italy
- <sup>26</sup> Department of Physics "E. Pancini", University Federico II, Via Cinthia 6, 80126, Napoli, Italy
- <sup>27</sup> INAF-Osservatorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli, Italy
- <sup>28</sup> Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
- <sup>29</sup> INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
- <sup>30</sup> INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
- <sup>31</sup> INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
- <sup>32</sup> INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
- <sup>33</sup> INFN-Sezione di Roma, Piazzale Aldo Moro, 2 – c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy
- <sup>34</sup> Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, 52056 Aachen, Germany
- <sup>35</sup> Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
- <sup>36</sup> Institut d'Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels, Barcelona, Spain
- <sup>37</sup> INFN section of Naples, Via Cinthia 6, 80126, Napoli, Italy
- <sup>38</sup> Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
- <sup>39</sup> Dipartimento di Fisica e Astronomia "Augusto Righi" – Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- <sup>40</sup> Instituto de Astrofísica de Canarias, Vía Láctea, 38205 La Laguna, Tenerife, Spain
- <sup>41</sup> Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
- <sup>42</sup> Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
- <sup>43</sup> European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
- <sup>44</sup> Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, Villeurbanne, F-69100, France
- <sup>45</sup> Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
- <sup>46</sup> UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, 4 rue Enrico Fermi, 69622 Villeurbanne, France
- <sup>47</sup> Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal
- <sup>48</sup> Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- <sup>49</sup> Université Paris-Saclay, CNRS, Institut d'astrophysique spatiale, 91405, Orsay, France
- <sup>50</sup> Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
- <sup>51</sup> INAF-Istituto di Astrofisica e Planetologia Spaziali, via del Fosso del Cavaliere, 100, 00100 Roma, Italy
- <sup>52</sup> Space Science Data Center, Italian Space Agency, via del Politecnico snc, 00133 Roma, Italy
- <sup>53</sup> INFN-Bologna, Via Irnerio 46, 40126 Bologna, Italy
- <sup>54</sup> School of Physics, HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK
- <sup>55</sup> Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
- <sup>56</sup> FRACTAL S.L.N.E., calle Tulipán 2, Portal 13 1A, 28231, Las Rozas de Madrid, Spain
- <sup>57</sup> Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109, USA
- <sup>58</sup> Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK
- <sup>59</sup> Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
- <sup>60</sup> Cosmic Dawn Center (DAWN), Denmark
- <sup>61</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

- <sup>62</sup> NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
- <sup>63</sup> Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
- <sup>64</sup> Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Finland
- <sup>65</sup> Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
- <sup>66</sup> Department of Physics, P.O. Box 64, 00014 University of Helsinki, Finland
- <sup>67</sup> Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
- <sup>68</sup> Laboratoire d'étude de l'Univers et des phénomènes eXtremes, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 92190 Meudon, France
- <sup>69</sup> Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, 0315 Oslo, Norway
- <sup>70</sup> SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, Cheshire SK11 9FT, UK
- <sup>71</sup> Centre de Calcul de l'IN2P3/CNRS, 21 avenue Pierre de Coubertin 69627 Villeurbanne Cedex, France
- <sup>72</sup> Dipartimento di Fisica "Aldo Pontremoli", Università degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy
- <sup>73</sup> INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
- <sup>74</sup> University of Applied Sciences and Arts of Northwestern Switzerland, School of Computer Science, 5210 Windisch, Switzerland
- <sup>75</sup> Universität Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
- <sup>76</sup> Dipartimento di Fisica e Astronomia "Augusto Righi" – Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy
- <sup>77</sup> Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham, DH1 3LE, UK
- <sup>78</sup> Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
- <sup>79</sup> Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
- <sup>80</sup> CNRS-UCB International Research Laboratory, Centre Pierre Binétruy, IRL2007, CPB-IN2P3, Berkeley, USA
- <sup>81</sup> University of Applied Sciences and Arts of Northwestern Switzerland, School of Engineering, 5210 Windisch, Switzerland
- <sup>82</sup> Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014, Paris, France
- <sup>83</sup> Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
- <sup>84</sup> Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
- <sup>85</sup> Telespazio UK S.L. for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- <sup>86</sup> European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
- <sup>87</sup> DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155, 2200 Copenhagen, Denmark
- <sup>88</sup> Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France
- <sup>89</sup> Institute of Space Science, Str. Atomistilor, nr. 409 Măgurele, Ilfov, 077125, Romania
- <sup>90</sup> Consejo Superior de Investigaciones Científicas, Calle Serrano 117, 28006 Madrid, Spain
- <sup>91</sup> Universidad de La Laguna, Departamento de Astrofísica, 38206 La Laguna, Tenerife, Spain
- <sup>92</sup> Dipartimento di Fisica e Astronomia "G. Galilei", Università di Padova, Via Marzolo 8, 35131 Padova, Italy
- <sup>93</sup> INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
- <sup>94</sup> Institut für Theoretische Physik, University of Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany
- <sup>95</sup> Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
- <sup>96</sup> Université St Joseph; Faculty of Sciences, Beirut, Lebanon
- <sup>97</sup> Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile
- <sup>98</sup> Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
- <sup>99</sup> Atlantis, University Science Park, Sede Bld 48940, Leioa-Bilbao, Spain
- <sup>100</sup> Department of Physics, Royal Holloway, University of London, TW20 0EX, UK
- <sup>101</sup> Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
- <sup>102</sup> Cosmic Dawn Center (DAWN)
- <sup>103</sup> Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark
- <sup>104</sup> Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, Plaza del Hospital 1, 30202 Cartagena, Spain
- <sup>105</sup> Centre for Information Technology, University of Groningen, PO Box 11044, 9700 CA Groningen, The Netherlands
- <sup>106</sup> Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- <sup>107</sup> INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
- <sup>108</sup> Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
- <sup>109</sup> Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- <sup>110</sup> INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy, and INFN-Sezione di Genova, Via Dodecaneso 33, 16146, Genova, Italy
- <sup>111</sup> ICL, Junia, Université Catholique de Lille, LITL, 59000 Lille, France
- <sup>112</sup> ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing, Via Magnanelli 2, Bologna, Italy

## Appendix A: Higher-order development of the galaxy bias

We introduced a first-order bias expansion in Sect. 3.1.3

$$\delta_x = b_x^{(1)} \delta_m + o(\delta_m). \quad (\text{A.1})$$

The two-point correlation between two samples  $a$  and  $b$  localised at the same redshift is

$$\xi_{ab}(\theta) := \langle \delta_a \delta_b \rangle_\theta = b_a b_b \langle \delta_m \delta_m \rangle_\theta + o(\langle \delta_m \delta_m \rangle). \quad (\text{A.2})$$

We are using the ratio of correlations to infer  $n(z)$ , as it is equal to unity at main order (cf. Sect. 3.3),

$$\frac{\xi_{ab}(\theta)}{\sqrt{\xi_{aa}(\theta) \xi_{bb}(\theta)}} = 1 + o(1). \quad (\text{A.3})$$

In this appendix we investigate the impact of second-order terms in the bias expansion on the ratio Eq. (A.3). We can express the galaxy overdensity  $\delta_g$  perturbatively up to second-order with

$$\delta_g = b_g^{(1)} \delta_m + \sum_j b_{j,g}^{(2)} O_j + o(\sum_j O_j), \quad (\text{A.4})$$

where  $O_j$  are a set of second-order fields, with  $\langle O_j \rangle = 0$ . We note that we do not consider the stochastic contribution, as we are working with real-space correlations. For example, in Pandey et al. (2020) and Nicola et al. (2024) they used

$$\sum_j b_{j,g}^{(2)} O_j = \frac{b_g^{(2)}}{2} (\delta_m^2 - \langle \delta_m^2 \rangle) + \frac{s_g}{2} (s^2 - \langle s^2 \rangle) + d_g \nabla^2 \delta_m, \quad (\text{A.5})$$

where  $b_g^{(2)}$  is the quadratic bias,  $s_g$  is the tidal bias,  $s^2$  is the trace squared of the tidal tensor, and  $d_{\text{gal}}$  is the non-local bias. The correlation between  $a$  and  $b$  (with possibly  $a = b$ ) at second order is:

$$\xi_{ab} = b_a^{(1)} b_b^{(1)} \langle \delta_m \delta_m \rangle + \sum_j (b_{j,a}^{(2)} b_b^{(1)} + b_a^{(1)} b_{j,b}^{(2)}) \langle \delta_m O_j \rangle + o(\sum \langle \delta_m O_j \rangle) \quad (\text{A.6})$$

$$= b_a^{(1)} b_b^{(1)} \langle \delta_m \delta_m \rangle \left( 1 + \sum_j \left( \frac{b_{j,a}^{(2)}}{b_a^{(1)}} + \frac{b_{j,b}^{(2)}}{b_b^{(1)}} \right) \frac{\langle \delta_m O_j \rangle}{\langle \delta_m \delta_m \rangle} + o\left( \sum \frac{\langle \delta_m O_j \rangle}{\langle \delta_m \delta_m \rangle} \right) \right). \quad (\text{A.7})$$

We note we did drop the  $\theta$  index for clarity. Using this relation for the product of the auto-correlation, we get at second order,

$$(\xi_{aa} \xi_{bb})^{-1/2} = (b_a^{(1)} b_b^{(1)} \langle \delta_m \delta_m \rangle)^{-1} \left( 1 + \sum_j \left( 2 \frac{b_{j,a}^{(2)}}{b_a^{(1)}} + 2 \frac{b_{j,b}^{(2)}}{b_b^{(1)}} \right) \frac{\langle \delta_m O_j \rangle}{\langle \delta_m \delta_m \rangle} + o\left( \sum \frac{\langle \delta_m O_j \rangle}{\langle \delta_m \delta_m \rangle} \right) \right)^{-1/2} \quad (\text{A.8})$$

$$= (b_a^{(1)} b_b^{(1)} \langle \delta_m \delta_m \rangle)^{-1} \left( 1 - \sum_j \left( \frac{b_{j,a}^{(2)}}{b_a^{(1)}} + \frac{b_{j,b}^{(2)}}{b_b^{(1)}} \right) \frac{\langle \delta_m O_j \rangle}{\langle \delta_m \delta_m \rangle} + o\left( \sum \frac{\langle \delta_m O_j \rangle}{\langle \delta_m \delta_m \rangle} \right) \right). \quad (\text{A.9})$$

$$(\text{A.10})$$

Thus, there is no next-to-leading order correction due to non-linearities:

$$\frac{\xi_{ab}}{\sqrt{\xi_{aa} \xi_{bb}}} = 1 + o\left( \sum \frac{\langle \delta_m O_j \rangle}{\langle \delta_m \delta_m \rangle} \right). \quad (\text{A.11})$$

## Appendix B: Magnification modelling

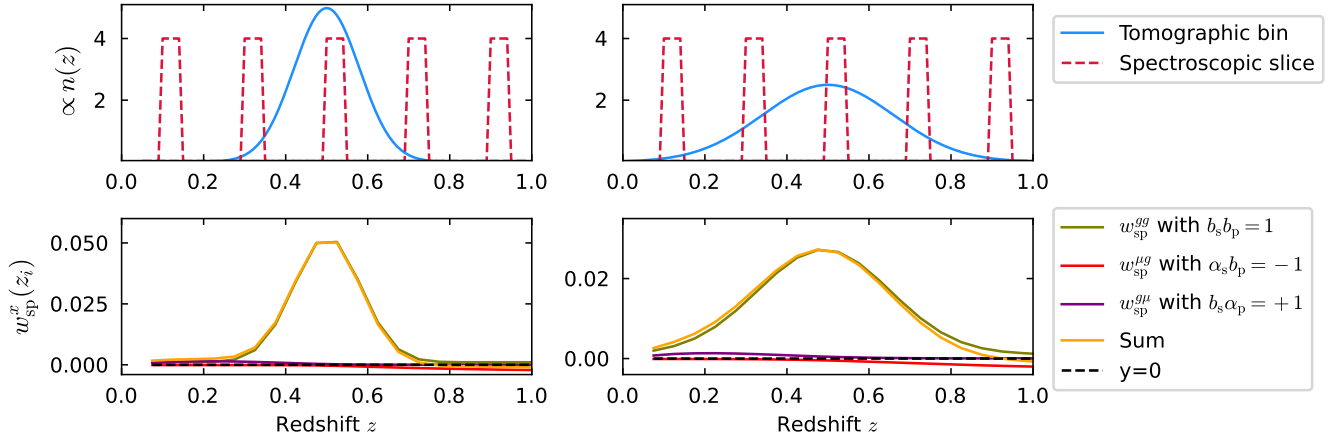
Here we describe the modelling of the magnification terms cf. Eq. (34). The impact of the magnification change in flux depends on a parameter  $s_\mu$  related to the selection function applied to the sample (Elvin-Poole et al. 2023). In the case of a magnitude threshold, it is the well-known derivative of the cumulative number count  $N_\mu^c$ , evaluated at the maximum magnitude,

$$s_\mu = \frac{d}{dm} [\ln N_\mu^c]_{m_{\text{max}}}. \quad (\text{B.1})$$

Usually, changes in flux and solid angle are concatenated through the coefficient<sup>18</sup>  $\alpha = 2.5s_\mu - 1$ .

Evaluating the  $\alpha$ s is in practice very challenging, since we are rarely considering only one magnitude cut as in Eq. (B.1), and it requires dedicated works (e.g. cf. the case of two magnitude cuts in Lepori et al. 2025), but is necessary for weak lensing (Hildebrandt 2016; Elvin-Poole et al. 2023).

<sup>18</sup> There are different and confusing conventions in the literature. For instance, sometimes  $\alpha$  is given as  $s$ ,  $2.5s$ , or  $5s - 2$ .



**Fig. B.1.** *Top panel:* The redshift distributions of the photometric sample (blue) and five out of the 20 spectroscopic samples (red) are shown. The photometric distribution on the left follows a Gaussian with parameters  $(\mu, \sigma) = (0.5, 0.08)$ , while on the right, it follows  $(\mu, \sigma) = (0.5, 0.16)$ . *Bottom panel:* The angular correlation is displayed for clustering only (green), magnification only (red and purple), and their sum (orange). The values of the coefficients  $b$  and  $\alpha$  are provided in the legend. Magnification impacts the shape of the angular correlation (sum vs. clustering only) and may influence the mean redshift, particularly for larger tomographic bins (right vs. left).

The spectroscopic sample is magnified by the matter traced by low- $z$  photometric galaxies, and the matter traced by spectroscopic galaxies is magnifying high- $z$  photometric galaxies. Neglecting RSD, using the conventions from Krause et al. (2021) and the Limber approximation, we have

$$w_{\text{sp}}^{\text{full}}(\theta) = \sum_{ij \in \{\text{DD}, \text{D}\mu, \mu\text{D}\}} c_s^i c_p^j \int d\chi \frac{\mathcal{W}_i^s \mathcal{W}_j^p}{\chi^2} \sum_{\ell} \frac{2\ell+1}{4\pi} \mathcal{L}_{\ell}(\cos \theta) P_m(k(\chi, \ell), z(\chi)), \quad (\text{B.2})$$

where  $c_a^{\text{D}} = b_a$  is the galaxy bias,  $c_a^{\mu} = \alpha_a$  is the magnification coefficient,  $\mathcal{L}_{\ell}$  is the Legendre polynomials of order  $\ell$ ,  $P_m$  is the matter power spectrum,  $k(\chi, \ell) = \frac{\ell+0.5}{\chi}$ , and

$$\mathcal{W}_{\text{D}}^a = n_a(z) \frac{dz}{d\chi} \quad \text{and} \quad \mathcal{W}_{\mu}^a = \frac{3\Omega_m H_0^2}{2c^2} \int_{\chi}^{\infty} d\chi' n_a(\chi') \frac{\chi}{a(\chi)} \frac{\chi' - \chi}{\chi'}. \quad (\text{B.3})$$

These expressions are simplified if the spectroscopic sample is localised in a small redshift slice ( $\Delta z$  small enough to neglect redshift variations) centred in  $z_s$ . In that case, we can model the distribution of the photometric sample by a step function

$$n_p(z) = \sum_i n_p(z_i) H(\Delta z/2 - |z - z_i|), \quad (\text{B.4})$$

where  $H$  is the Heaviside function, equal to unity for  $\Delta z/2 - |z - z_i| > 0$  and 0 otherwise. We then have

$$w_{\text{sp}}^{\text{DD}}(\theta) \approx b_s(z_s) b_p(z_s) n_p(z_s) \xi_m(\chi_s), \quad (\text{B.5})$$

$$w_{\text{sp}}^{\text{D}\mu}(\theta) = \frac{3\Omega_m H_0}{c} b_s(z_i) n_s(z_i) \frac{H_0(1+z_i)}{H(z_i)} \xi_m(z_i, \theta) \sum_{j=i+1, \dots} \Delta z \alpha_p(z_j) n_p(z_j) \frac{\chi(z_j) - \chi(z_i)}{\chi(z_j)} \chi(z_i), \quad (\text{B.6})$$

$$w_{\text{sp}}^{\mu\text{D}}(\theta) = \frac{3\Omega_m H_0}{c} \sum_{j=0, \dots, i-1} \Delta z b_p(z_j) n_p(z_j) \frac{H_0(1+z_j)}{H(z_j)} \xi_m(z_j, \theta) \alpha_s(z_j) \frac{\chi(z_i) - \chi(z_j)}{\chi(z_i)} \chi(z_j). \quad (\text{B.7})$$

We can rewrite our magnification term as

$$M(\theta) = \alpha_s(z_s) \sum_{j=\dots, i-1} b_p(z_j) n_p(z_j) D(z_j, z_i, \theta) + b_s(z_i) \sum_{j=i+1, \dots} \alpha_p(z_j) n_p(z_j) D(z_i, z_j, \theta), \quad (\text{B.8})$$

with

$$D(z_k, z_l, \theta) = \frac{3H_0\Omega_m}{c} \Delta z \frac{H_0}{H(z_k)} (1+z_k) \xi_m(z_k, \theta) \frac{\chi(z_l) - \chi(z_k)}{\chi(z_l)} \chi(z_k). \quad (\text{B.9})$$

Our final expression contains several differences with Gatti et al. (2022). We are taking into account the redshift variation of spectroscopic and photometric biases, and  $\alpha$ . Our conventions and the sum indices are slightly different (we try to be more explicit).

We illustrate the impact of magnification on angular correlation in Fig. B.1 using toy models. Two tomographic bins are introduced, with Gaussian redshift distributions centred at  $z = 0.5$  with widths  $\sigma = 0.08$  and  $\sigma = 0.16$ , respectively. The clustering and

the two magnification terms are evaluated using spectroscopic samples divided into 20 redshift slices with  $\Delta z = 0.05$ . We adopt  $b_s = b_p = 1$ ,  $\alpha_s = 1$ , and  $\alpha_p = -1$ .

A noticeable difference is observed between clustering alone and the sum of the three effects, indicating that magnification is not entirely negligible. Since different  $\alpha$  values are used for the spectroscopic and photometric samples, the combined effect appears to be shifted towards lower redshifts, potentially introducing a bias in the mean redshift. Finally, the impact is very limited for the narrower tomographic bin (left) but becomes more pronounced for the wider bin (right).

This highlights the importance of accounting for magnification effects when using broad tomographic bins. With UNION and LSST photometry, our bins are projected to be narrower and more localised than DES ones, and thus magnification may be negligible for *Euclid* clustering redshifts, even if it was not for DES.

## Appendix C: Pair counts

The pair-count between two samples is theoretically given by

$$D_a D_b(r_p) = \iint dz_1 dz_2 \langle N_a N_b \rangle_{r_p} = \iint dz_1 dz_2 N_a(z_1) N_b(z_2) \langle (1 + \delta_a)(1 + \delta_b) \rangle_{r_p}^{\text{data}}, \quad (\text{C.1})$$

where  $N(z) = N_{\text{tot}} n(z)$  is the number galaxy distribution. We use galaxy weights to correct selection effects, anisotropies due to observation conditions or background etc.

### C.1. From continuous to discrete

In practice, the pair count is evaluated for a certain distance range  $r_p \pm \Delta r_p$ . It means that for a galaxy of sample  $a$ , we are counting all the sample  $b$  galaxies within an annulus with radius  $r_p^i$  and width  $\Delta r_p^i$ . For instance, we use logarithmic binning, for which  $\Delta r_p^i \propto r_p^i$ . Thus, the correlation (including galaxy weights) of Eq. (C.1) is related to the theoretical one through,

$$\langle (1 + \delta_a)(1 + \delta_b) \rangle_{r_p^i}^{\text{data}} = \int_0^{2\pi} d\theta \int_{r_p^i \pm \Delta r_p^i / 2} dr_p \langle (1 + \delta_a)(1 + \delta_b) \rangle_{r_p, \theta} \approx 2\pi r_p^i \Delta r_p^i \langle (1 + \delta_a)(1 + \delta_b) \rangle_{r_p^i}. \quad (\text{C.2})$$

### C.2. Mask and estimator

There is an additional geometrical effect to Eq. (C.2) which is the mask: we do not observe the full sky. The annuli are not perfect and there is a geometrical correction

$$\langle (1 + \delta_a)(1 + \delta_b) \rangle_{r_p^i}^{\text{data}} \approx 2\pi r_p^i \Delta r_p^i \eta_{\text{mask}}(r_p) \langle (1 + \delta_a)(1 + \delta_b) \rangle_{r_p^i}. \quad (\text{C.3})$$

To correct it, we compare the pair count of the correlated samples, with the ones of uncorrelated samples, submitted to the same mask, the so-called ‘‘randoms’’,

$$R_a R_b(r_p) = N_a^{\text{rand}} N_b^{\text{rand}} 2\pi r_p^i \Delta r_p^i \eta_{\text{mask}}(r_p). \quad (\text{C.4})$$

Comparing different pair counts, we always correct the amplitude, which scales as the number of objects. We usually use 10 times more randoms than objects, so we multiply DD/RR by 100. This step is done by default in *Treecorr*.

Computing random-galaxy pair counts, DR [which scales as  $2\pi r_p^i \Delta r_p^i \eta_{\text{mask}}(r_p)$  as well] is optimal for variance reduction (Landy & Szalay 1993). That is why our fiducial analysis involves the full Landy–Szalay estimator,

$$w_{ab}^{\text{data}}(r_p) = \frac{D_a D_b - R_a D_b - D_a R_b + R_a R_b}{R_a R_b}(r_p) = \langle (1 + \delta_a)(1 + \delta_b) - (1 + \delta_a) - (1 + \delta_b) + 1 \rangle_{r_p^i} = \langle \delta_a \delta_b \rangle_{r_p^i}, \quad (\text{C.5})$$

where we omitted the galaxy number normalisation factors introduced below. A common choice for clustering-redshifts is to use the Davis and Peebles estimator

$$w_{ab}(r_p) \approx \frac{D_a D_b - R_a D_b}{R_a D_b}(r_p), \quad (\text{C.6})$$

mainly because there is no need for a random catalogue for the second sample (Davis & Peebles 1983; Gatti et al. 2018; van den Busch et al. 2020). For *Euclid*, a random catalogue would be created for the different analyses, so there is no motivation for using Eq. (C.6) rather than Eq. (C.5).

## Appendix D: Systematic errors and bias correction

In this appendix, we provide additional tables to Sects. 4.7 and 4.8. In Table D.1, we report the values of the deviations on the mean redshift and shape due to the systematic effects. The values are the means over the five sky patches, and correspond to  $\Delta z = 0.05$  and the M5 bias correction. We found similar results with other bias corrections. Finally in Table D.2, we report the values of the deviations on the mean redshift and standard deviation associated with the bias correction, in the absence of systematic effects.

**Table D.1.** Mean redshift and standard deviation bias for the different systematic effects (including none, and all) are reported; Req. refers to the requirement that needs to be fulfilled.

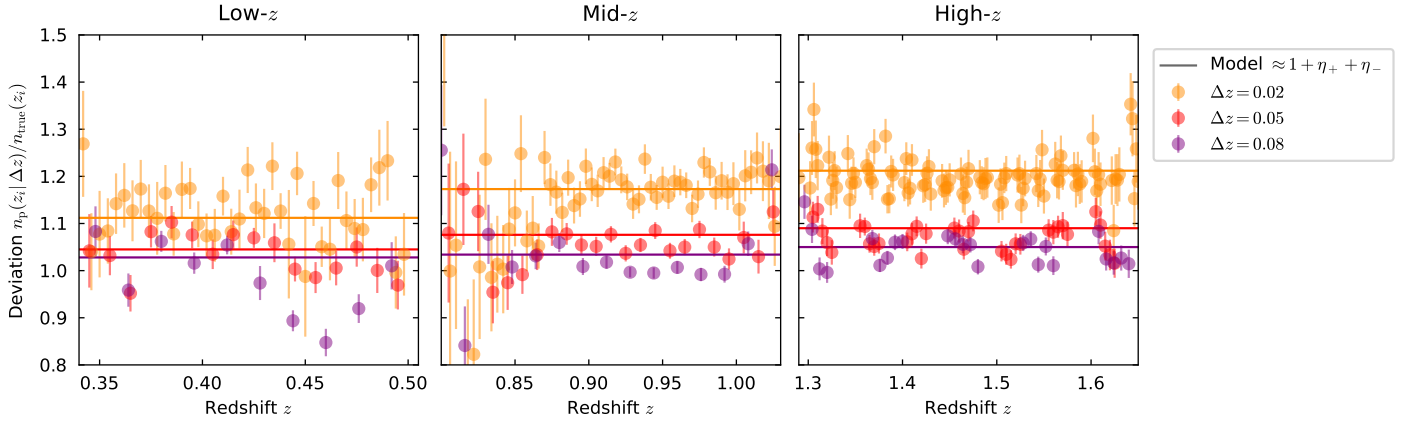
Bin	Configuration	Shifted-Stretched Model		Suppressed-GP	
		$\langle z \rangle - \langle z \rangle_{\text{true}} (\times 1000)$	$\sigma_z / \sigma_z^{\text{true}} - 1 (\times 100)$	$\langle z \rangle - \langle z \rangle_{\text{true}} (\times 1000)$	$\sigma_z / \sigma_z^{\text{true}} - 1 (\times 100)$
Low- $z$	Independent	$-1.3 \pm 0.9$	$0.6 \pm 1.4$	$0.7 \pm 1.9$	$-0.2 \pm 4.6$
	Full bin	$-0.6 \pm 1.2$	$1.6 \pm 1.7$	$0.4 \pm 2.2$	$1.2 \pm 4.9$
	Full bin +RSD	$-0.6 \pm 1.2$	$2.9 \pm 1.8$	$1.0 \pm 2.2$	$-2.1 \pm 4.8$
	Full bin +Magn	$-1.3 \pm 1.1$	$0.4 \pm 1.6$	$1.9 \pm 2.5$	$5.8 \pm 7.1$
	Total	$-0.4 \pm 1.2$	$1.0 \pm 1.8$	$2.4 \pm 2.6$	$5.6 \pm 7.3$
	<b>Req.</b>	2.8	10	2.8	10
Mid- $z$	Independent	$1.4 \pm 0.6$	$2.2 \pm 0.8$	$1.7 \pm 1.6$	$-0.1 \pm 3.6$
	Full bin	$1.1 \pm 0.8$	$1.4 \pm 1.0$	$3.8 \pm 1.9$	$-2.2 \pm 3.9$
	Full bin +RSD	$1.8 \pm 0.9$	$2.9 \pm 1.2$	$3.8 \pm 1.9$	$-5.3 \pm 3.7$
	Full bin +Magn	$1.3 \pm 0.8$	$0.4 \pm 1.1$	$3.8 \pm 1.7$	$-0.7 \pm 3.7$
	Total	$2.2 \pm 0.8$	$2.3 \pm 1.0$	$3.8 \pm 1.7$	$-3.1 \pm 4.0$
	<b>Req.</b>	3.8	10	3.8	10
High- $z$	Independent	$-1.9 \pm 1.0$	$0.4 \pm 0.6$	$-0.7 \pm 2.2$	$1.1 \pm 1.9$
	Full bin	$-1.0 \pm 1.3$	$0.7 \pm 0.8$	$0.2 \pm 2.5$	$0.5 \pm 2.1$
	Full bin +RSD	$-1.4 \pm 1.3$	$1.5 \pm 0.8$	$-0.6 \pm 2.3$	$-0.6 \pm 1.9$
	Full bin +Magn	$1.0 \pm 1.2$	$-1.8 \pm 0.7$	$2.3 \pm 2.7$	$4.0 \pm 2.4$
	Total	$1.1 \pm 1.3$	$-1.3 \pm 0.8$	$2.2 \pm 2.6$	$3.6 \pm 2.2$
	<b>Req.</b>	4.4	10	4.4	10

**Table D.2.** Mean redshift and standard deviation for the different bias corrections M0–M5, for SSM and SGP fitting, without systematics (independent case in Sect. 4.7); Req. refers to the requirement to fulfil.

Bin	Configuration	Shifted-Stretched Model		Suppressed-GP	
		$\langle z \rangle - \langle z \rangle_{\text{true}} (\times 1000)$	$\sigma_z / \sigma_z^{\text{true}} - 1 (\times 100)$	$\langle z \rangle - \langle z \rangle_{\text{true}} (\times 1000)$	$\sigma_z / \sigma_z^{\text{true}} - 1 (\times 100)$
Low- $z$	M0	$-0.3 \pm 1.4$	$0.8 \pm 2.0$	$1.0 \pm 2.1$	$0.5 \pm 4.7$
	M1	$-1.1 \pm 1.0$	$0.3 \pm 1.6$	$0.7 \pm 2.0$	$0.0 \pm 4.6$
	M2	$-1.1 \pm 1.0$	$0.1 \pm 1.5$	$0.7 \pm 1.9$	$-0.0 \pm 4.2$
	M3	$-1.2 \pm 1.6$	$0.5 \pm 2.4$	$-0.1 \pm 2.4$	$0.5 \pm 5.0$
	M4	$-1.5 \pm 1.6$	$0.3 \pm 2.4$	$-0.5 \pm 2.4$	$0.8 \pm 5.1$
	M5	$-1.3 \pm 0.9$	$0.6 \pm 1.4$	$0.7 \pm 1.9$	$-0.2 \pm 4.6$
	<b>Req.</b>	2.8	10	2.8	10
Mid- $z$	M0	$4.9 \pm 0.6$	$2.7 \pm 0.8$	$5.2 \pm 1.6$	$-0.7 \pm 3.4$
	M1	$2.7 \pm 0.6$	$2.4 \pm 0.8$	$3.0 \pm 1.7$	$-0.2 \pm 3.7$
	M2	$2.7 \pm 0.6$	$2.4 \pm 0.8$		$-0.1 \pm 3.7$
	M3	$1.4 \pm 0.4$	$2.0 \pm 0.5$		$1.1 \pm 3.2$
	M4	$1.5 \pm 0.4$	$2.0 \pm 0.5$		$1.1 \pm 3.2$
	M5	$1.4 \pm 0.6$	$2.2 \pm 0.8$		$-0.1 \pm 3.6$
	<b>Req.</b>	3.8	10	3.8	10
High- $z$	M0	$17.0 \pm 1.2$	$-5.6 \pm 0.7$	$17.8 \pm 2.6$	$3.9 \pm 2.1$
	M1	$5.1 \pm 1.0$	$-1.8 \pm 0.7$	$6.5 \pm 2.1$	$2.1 \pm 1.9$
	M2	$5.0 \pm 1.1$	$-1.8 \pm 0.7$	$6.5 \pm 2.3$	$2.1 \pm 1.9$
	M3	$-1.8 \pm 1.0$	$0.4 \pm 0.6$	$-0.4 \pm 2.1$	$0.7 \pm 1.8$
	M4	$-3.7 \pm 1.0$	$1.1 \pm 0.6$	$-2.3 \pm 2.0$	$0.3 \pm 1.8$
	M5	$-1.9 \pm 1.0$	$0.4 \pm 0.6$	$-0.7 \pm 2.2$	$1.1 \pm 1.9$
	<b>Req.</b>	4.4	10	4.4	10

## Appendix E: Mis-modelling of clustering-redshifts with the one-bin approximation

In Sect. 3.2.3, we introduced an improvement to the modelling to expand beyond this one-bin approximation, which in the context of M5 bias correction is,



**Fig. E.1.** Deviation of the measured  $n_p(z_i)$  to the true distribution, for different slicing  $\Delta z$  (points in colours), and the global offset predicted by our model, which includes correlation at the edge of the slices (solid lines in colours). The points correspond to the measurements for the five patches, hence the difference in redshift between them is not  $\Delta z$ .

$$\frac{w_{\text{sp}}}{\Delta z \sqrt{w_{\text{ss}} w_{\text{pp}}}}(z_i) \approx n_b(z_i) \times \left( 1 + \frac{n_b(z_i + \Delta z)}{n_b(z_i)} \eta_+(z_i) + \frac{n_b(z_i - \Delta z)}{n_b(z_i)} \eta_-(z_i) \right).$$

Since  $\eta_{\pm} > 0$ , we expect to measure  $n(z)$  with a norm higher than unity. The global offset should be approximately

$$\sum_{z_i} \frac{w_{\text{sp}}}{\Delta z \sqrt{w_{\text{ss}} w_{\text{pp}}}}(z_i) \approx 1 + \eta_+(\langle z \rangle) + \eta_-(\langle z \rangle). \quad (\text{E.1})$$

In Fig. E.1 we show the ratio of the measured  $n_p(z_i)$  to the true distribution, for three slicing  $\Delta z$ , and the five sky patches. We also report the offset predicted from Eq. (E.1) for each  $\Delta z$ . We observe consistency between the points and the rough prediction. Concatenating the measurements from the different redshift slices, we can express the correlation as

$$\mathbf{w}_{\text{sp}}(\theta) = B_s B_p \Lambda \mathbf{n}, \quad (\text{E.2})$$

where  $\mathbf{n} = \{n(z_i)\}_{i=1,\dots,n}$ ,  $B_x$  are diagonal matrices representing the galaxy biases, and  $\Lambda$  is a tridiagonal matrix, easily invertible, with unity on its diagonal and smaller terms (corrections) on the first off-diagonals. We attempted to implement Eq. (E.2) for the inference of  $n_p(z)$  by replacing the one-bin modelling with three-bins. We did not observe a significant improvement in the bias on the mean redshift compared to the standard renormalisation procedure, however. In Fig. E.1, we note that points from different patches at similar redshifts (visible as groups of 5 points corresponding to the 5 patches in the high- $z$  bin) are spread over a wide range, despite the true distributions being very similar across the patches. This behaviour could be attributed to substantial sample variance, which might explain why three-bins perform similarly to one-bin. Additionally, we observe that three-bins still exhibit significant deviations in Fig. 2, highlighting the challenges of improving one-bin modelling despite our efforts. We note that in Fig. 14, the measurements for  $0.75 < z < 0.85$  are systematically under the true  $n(z)$  for  $\Delta z = 0.02, 0.05$ , which might cause the positive shift in mean redshift. This under-correlation is also visible in Fig. E.1. We did not identify its cause.