

Euclid: Photometric redshift calibration with self-organising maps[★]

W. Roster^{1,★★}, A. H. Wright², H. Hildebrandt², R. Reischke³, O. Ilbert⁴, W. d'Assignies D.⁵, M. Manera^{6,5}, M. Bolzonella⁷, D. C. Masters⁸, S. Paltani⁹, W. G. Hartley⁹, Y. Kang⁹, H. Hoekstra¹⁰, B. Altieri¹¹, A. Amara¹², S. Andreon¹³, N. Auricchio⁷, C. Baccigalupi^{14,15,16,17}, M. Baldi^{18,7,19}, A. Balestra²⁰, S. Bardelli⁷, P. Battaglia⁷, R. Bender^{1,21}, A. Biviano^{15,14}, E. Branchini^{22,23,13}, M. Brescia^{24,25}, S. Camera^{26,27,28}, G. Cañas-Herrera^{29,30,10}, V. Capobianco²⁸, C. Carbone³¹, V. F. Cardone^{32,33}, J. Carretero^{34,35}, R. Casas^{36,37}, S. Casas³⁸, F. J. Castander^{37,36}, M. Castellano³², G. Castignani⁷, S. Cavuoti^{25,39}, K. C. Chambers⁴⁰, A. Cimatti⁴¹, C. Colodro-Conde⁴², G. Congedo⁴³, C. J. Conselice⁴⁴, L. Conversi^{45,11}, Y. Copin⁴⁶, A. Costille⁴, F. Courbin^{47,48}, H. M. Courtois⁴⁹, M. Cropper⁵⁰, A. Da Silva^{51,52}, H. Degaudenzi⁹, S. de la Torre⁴, G. De Lucia¹⁵, F. Dubath⁹, C. A. J. Duncan^{43,44}, X. Dupac¹¹, S. Dusini⁵³, S. Escoffier⁵⁴, M. Farina⁵⁵, R. Farinelli⁷, S. Farrens⁵⁶, F. Faustini^{32,57}, S. Ferriol⁴⁶, F. Finelli^{7,58}, P. Fosalba^{36,37}, N. Fourmanoit⁵⁴, M. Frailis¹⁵, E. Franceschi⁷, M. Fumana³¹, S. Galeotta¹⁵, K. George⁵⁹, W. Gillard⁵⁴, B. Gillis⁴³, C. Giocoli^{7,19}, J. Gracia-Carpio¹, A. Grazian²⁰, F. Grupp^{1,21}, S. V. H. Haugan⁶⁰, W. Holmes⁶¹, F. Hormuth⁶², A. Hornstrup^{63,64}, P. Hudelot⁶⁵, K. Jahnke⁶⁶, M. Jhabvala⁶⁷, B. Joachimi⁶⁸, E. Keihänen⁶⁹, S. Kermiche⁵⁴, B. Kubik⁴⁶, H. Kurki-Suonio^{70,71}, A. M. C. Le Brun⁷², D. Le Mignant⁴, S. Ligori²⁸, P. B. Lilje⁶⁰, V. Lindholm^{70,71}, I. Lloro⁷³, D. Maino^{74,31,75}, E. Maiorano⁷, O. Mansutti¹⁵, O. Marggraf³, M. Martinelli^{32,33}, N. Martinet⁴, F. Marulli^{76,7,19}, R. J. Massey⁷⁷, E. Medinaceli⁷, S. Mei^{78,79}, M. Melchior⁸⁰, Y. Mellier^{81,65}, M. Meneghetti^{7,19}, E. Merlin³², G. Meylan⁸², A. Mora⁸³, M. Moresco^{76,7}, L. Moscardini^{76,7,19}, R. Nakajima³, C. Neissner^{5,35}, S.-M. Niemi²⁹, C. Padilla⁵, F. Pasian¹⁵, K. Pedersen⁸⁴, V. Pettorino²⁹, S. Pires⁵⁶, G. Polenta⁵⁷, M. Poncet⁸⁵, L. A. Popa⁸⁶, L. Pozzetti⁷, F. Raison¹, R. Rebolo^{42,87,88}, A. Renzi^{89,53}, J. Rhodes⁶¹, G. Riccio²⁵, E. Romelli¹⁵, M. Roncarelli⁷, C. Rosset⁷⁸, E. Rossetti¹⁸, R. Saglia^{21,1}, Z. Sakr^{90,91,92}, D. Sapone⁹³, B. Sartoris^{21,15}, M. Schirmer⁶⁶, P. Schneider³, T. Schrabback⁹⁴, M. Scodreggio³¹, A. Secroun⁵⁴, E. Sefusatti^{15,14,16}, G. Seidel⁶⁶, S. Serrano^{36,95,37}, P. Simon³, C. Sirignano^{89,53}, G. Sirri¹⁹, J. Skottfelt⁹⁶, L. Stanco⁵³, J. Steinwagner¹, P. Tallada-Crespi^{34,35}, A. N. Taylor⁴³, H. I. Teplitz⁸, I. Tereno^{51,97}, N. Tessore⁶⁸, S. Toft^{98,64}, R. Toledo-Moreo⁹⁹, F. Torradeflot^{35,34}, I. Tutusaus⁹¹, L. Valenziano^{7,58}, J. Valiviita^{70,71}, T. Vassallo^{21,15}, G. Verdoes Kleijn¹⁰⁰, A. Veropalumbo^{13,23,22}, Y. Wang⁸, J. Weller^{21,1}, G. Zamorani⁷, F. M. Zerbi¹³, E. Zucca⁷, C. Burigana^{101,58}, L. Gabarra¹⁰², C. Porciani³, V. Scottez^{81,103}, and M. Sereno^{7,19}

(Affiliations can be found after the references)

Received 13 August 2025 / Accepted 2 December 2025

ABSTRACT

The *Euclid* large-scale weak-lensing survey aims to trace the evolution of cosmic structures up to redshift $z \sim 3$ and beyond. Its success depends critically on obtaining highly accurate mean redshifts for ensembles of galaxies $n(z)$ in all tomographic bins, essential for deriving robust cosmological constraints. However, photometric redshifts (photo- z s) are affected by systematic biases, arising from various sources of uncertainty and dominated by selection effects of the spectroscopic sample used for calibration. To address these challenges, we utilised self-organising maps (SOMs) with mock samples resembling the Euclid Wide Survey (EWS) from the Flagship2 simulation, to validate *Euclid*'s uncertainty requirement of $|\Delta(z)| = \langle z_{\text{est}} \rangle - \langle z \rangle \leq 0.002(1+z)$ per tomographic bin, assuming DR3-level data. Consequently, we identify the most effective galaxy selection for our tomographic bins, while systematically examining the implementation of quality control cuts to reduce sources of uncertainty. In particular, we observe that defining the redshift tomography using the mean spectroscopic redshift (spec- z) per SOM cell, results in none of the ten tomographic redshift bins satisfying the requirement. In contrast, the redshift tomography on the photo- z s of the EWS-like sample yields superior results, with eight out of ten bins [$0 < z \leq 2.5$] meeting the *Euclid* requirement. To enhance the realism of our study, we morph our calibration sample to mimic the C3R2 survey in incremental steps. In this context, a maximum of six out of ten bins meet the requirement, strongly advocating the adoption of a redshift tomography defined by the photo- z s of individual galaxies rather than the commonly used mean spec- z of SOM cells. To examine the impact on the expected biases for Ω_m , σ_8 , and Δw_0 measured by *Euclid*, we perform a Fisher forecast for cosmic shear only, based on our redshift uncertainties. Here, we find that even under an evaluation of the uncertainty where the impact of the redshift bias is substantial, most absolute biases remain below 0.1σ in the idealised scenario and below 0.3σ in the more realistic case.

Key words. methods: data analysis – techniques: photometric – galaxies: photometry – cosmological parameters – dark energy

[★] This paper is published on behalf of the Euclid Consortium.

^{★★} Corresponding author: wroster@mpe.mpg.de

1. Introduction

Our current understanding of the Universe is primarily guided by the widely accepted Λ cold dark matter (Λ CDM) model, recognised for its accuracy and reliability in replicating observations (Frieman et al. 2008; Thomas et al. 2016; Planck Collaboration I 2020). Despite the success thus far, our knowledge of its implicated components, such as dark energy (DE), remains limited. The primary impact of DE lies in inducing late-time acceleration and influencing the growth rate of cosmic structures. In fact, exploring alternatives to Λ with an equation-of-state (EOS) parameter $w(z) = -1$ of a dynamical DE model evolving as $w(z) = w_0 + w_a \frac{z}{1+z}$, where w_0 denotes the EOS at $z = 0$ and w_a describes its evolution with respect to the scale factor a , could show incompatibility, and thus invalidate the Λ CDM model (Caldwell et al. 1998; Euclid Collaboration: Ilbert et al. 2021). Among other approaches, this theory can be probed by weak gravitational lensing (WL) through cosmic shear surveys (e.g. Hildebrandt et al. 2017; Troxel et al. 2018; Mandelbaum 2018; Hikage et al. 2019; Li et al. 2023; Wright et al. 2025a). These measure the coherent distortion of galaxy shapes in images by large-scale structures, allowing us to trace the apparent underlying mass distribution along the line of sight (Kilbinger 2015). Past wide sky precision cosmology surveys, such as the ground-based Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS, Erben et al. 2013), the Kilo-Degree Survey (KiDS, Hildebrandt et al. 2017; Wright et al. 2025a), the Hyper-Suprime Camera Survey (HSC, Aihara et al. 2018; Hikage et al. 2019), or the Dark Energy Survey (DES, Abbott et al. 2018; Troxel et al. 2018) have successfully utilised cosmic shear tomography to reconstruct a three-dimensional layout from two-dimensional observations by dividing the observations into a number of redshift bins, depending on the respective probes (Wong et al. 2026). Following in their footsteps, we hope to gain improved insight with the *Nancy Grace Roman* Space Telescope (Spergel et al. 2015), the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezić & Kahn 2019), and the recently launched optical to near-infrared European Space Agency (ESA) *Euclid* satellite (Laureijs et al. 2011; Euclid Collaboration: Mellier et al. 2025). Central to these efforts is the sufficiently accurate measurement of galaxy redshifts, which is critical for tomographic binning and, in turn, for interpreting cosmological observations (Bordoloi et al. 2012; Stözlner et al. 2021). Correspondingly, *Euclid* will cover over 14 000 deg² (Euclid Collaboration: Scaramella et al. 2022) during its six-year mission, charting the extragalactic distance-redshift relation for 1.5 billion galaxies out to $z \sim 2$, thereby covering the period in which the Universe passed from DM to DE domination (Weinberg et al. 2013).

The rapid expansion of wide-field imaging campaigns has intensified efforts to achieve precise redshift calibration across large datasets. Given the impracticality of obtaining spectroscopic redshifts (spec- z) for the vast numbers of distant and often faint galaxies essential to WL studies, redshift estimates, ideally in the form of full probability distributions $p(z)$, are instead inferred from photometric observations using intermediate- or broad-band filters (Newman et al. 2015). Advantageously, imaging surveys provide more galaxies with redshift estimates per unit of telescope time compared to spectroscopic surveys, resulting in significant growth of such photometric redshift (photo- z) methods over the past two decades (Salvato et al. 2019). Several approaches have been employed to infer, among other things, the galaxy redshift dis-

tribution $n(z)$ of photometric samples more efficiently, albeit with significantly reduced precision. These include clustering redshifts, which utilise the spatial distribution of overlapping spectroscopic samples (Newman 2008; McQuinn & White 2013; Morrison et al. 2017; Gatti et al. 2021; d’Assignies et al. 2025), template fitting methods (Benitez 2000; Brammer et al. 2008; Arnouts & Ilbert 2011; Euclid Collaboration: Ilbert et al. 2021), and machine learning algorithms such as fully connected neural networks, decision trees, Gaussian processes, and support vector machines (Kind & Brunner 2013; Sadeh et al. 2016; Euclid Collaboration: Desprez et al. 2020).

Since tomographic cosmic shear studies are primarily sensitive to the average distance of sources, they require accurate knowledge of the $n(z)$ distribution and strongly rely on the ability to bin galaxies by redshift coarsely (Huterer et al. 2006; Newman et al. 2015; Hildebrandt et al. 2021). To ensure this, a robust redshift calibration is necessary. However, correctly fine-tuning photo- z is particularly challenging (Hoyle et al. 2018; Merz et al. 2024), as systematic uncertainties commonly induce undesired bias and scatter. A bias in the estimated $n(z)$, such as introduced by systematically underestimated redshift values, would unequivocally lead to the incorrect model of an overall denser, more highly clustered gravitational landscape than exists in reality (Wright et al. 2020a). Therefore, an accurate determination of the mean redshift per bin is necessary to derive the cross-correlation of the source distortions between tomographic redshift planes, which in turn provide estimates on the distribution of mass along the line of sight. This poses an arduous task, since the photo- z calibration efforts, similar to clustering redshifts, are heavily dependent on the available reference samples of spec- z (Stanford et al. 2021; d’Assignies et al. 2025). Minimising bias and its uncertainty is crucial in order to obtain robust WL constraints in future measurements that, once combined with cosmic microwave background (CMB) measurements from the *Planck* satellite (Planck Collaboration I 2020), can test discrepancies with other cosmological constraints on the amount and clustering of (predominantly dark) matter across cosmic time (Asghari et al. 2019; Wright et al. 2020b).

In this paper we focus on analysing the current state-of-the-art efforts in redshift calibration to validate whether the residual systematic bias for the *Euclid* mission can be minimised to satisfy its stringent requirement of $\Delta\langle z \rangle \leq 0.002(1+z)$ per tomographic redshift bin, which is comparable to the precision expected for stage-IV surveys (Bordoloi et al. 2012). Comparable to the approaches taken in Buchs et al. (2019), Zuntz et al. (2021), Myles et al. (2021), or Gatti et al. (2021), for example, the indicated objective is subsequently realised by training an unsupervised, competitive machine learning (ML) algorithm known as a self-organising map (SOM, Kohonen 1982) with a large¹ mock catalogue to map the galaxies’ multi-dimensional magnitude-colour space ($u - g, g - r, \dots, J - H$) onto a grid of representative cells (Masters et al. 2015). The SOM performs dimensionality reduction in colour space, preserving the topological relationships of the input features. Since galaxy colours correlate strongly, though not uniquely, with redshift, the resulting SOM tends to place galaxies with similar redshifts in nearby cells. However, because the colour-redshift relation is non-linear and can be degenerate in specific regions of colour space, neighbouring cells may not necessarily have small proximities in redshift. The baseline of this method is that a deep spectroscopic calibration sample can be reweighted such

¹ Compared to the sample size used in previous photometric WL studies, such as KiDS.

that it is representative of a wide photometric validation sample of unknown $n(z)$ (Lima et al. 2008; van den Busch et al. 2022; Campos et al. 2024). Furthermore, SOMs allow for the identification and systematic removal of sources in the validation sample that are missing representation in the calibration sample, which, if included in the distribution, would bias the estimates. Under suitable conditions, SOMs have previously been shown to reach residual biases in mean redshifts of ≈ 0.01 or better (Hildebrandt et al. 2021; Myles et al. 2021; Giannini et al. 2024). As *Euclid* datasets will be complemented by ground-based optical data, this paper tests various tomographic redshift calibration approaches and establishes a proof of concept under an idealised framework for *Euclid* and LSST Year 3 (Y3)-like data, not representative of *Euclid* DR1 capabilities. By focusing on this controlled set-up, this study serves as an essential test of robustness, demonstrating the viability and effectiveness of our approach, while setting the stage for future applications.

This work is structured as follows. In Sect. 2, we introduce the SOM and the implementation of the calibration methodology. Section 3 gives an overview of *Euclid* and the complementary mocks. Section 4 investigates the alteration in performance for various bias calibration reduction techniques attributed to different redshift tomographies. Section 5 covers the scientific outlook and the levels of accuracy to expect from *Euclid* based on our findings. The main results are summarised in Sect. 6.

2. Methodology

Empirical reconstructions of the colour-redshift space can be used in direct calibration, which aims at estimating the unknown redshift distribution $n(z)$ of a photometric galaxy sample. This method involves matching a deep spectroscopic sample S to a photometric sample P , only requiring that S covers, even sparsely, the range of shear observations spanned by P (Wright et al. 2020b). However, in practice S often shows different selection functions compared to P , caused by spectroscopic targeting strategies and success rates (Beck et al. 2017; Gruen & Brimiouille 2017; Hartley et al. 2020). Work by Wright et al. (2020b) has shown that this issue can be addressed by additional cleaning of P , resulting in a so-called gold sample. This is implemented by training a SOM on a set of values consisting of magnitudes and/or colours drawn from S , before parsing P into the same cells. Cells that are not occupied by objects in both samples are rejected, effectively removing critical parts of the colour space that are not adequately represented by S (for more details on recalibration, see Hildebrandt et al. 2021). In principle, additional clustering of cells with similar photometric properties can improve the balance between cell removal due to colour-space exclusion as well as bias introduced by a flawed representation of the gold sample selection (van den Busch et al. 2022).

Once trained, a higher-dimensional feature space of multiple attributes is projected onto the SOM. In other words, a SOM trained on input data attributes a , such as colours, can be painted by an attribute b , such as redshift (Davidzon et al. 2019). This unsupervised technique does not rely on labels, such as redshifts from a spectroscopic sample, during the training process (Kind & Brunner 2013). Since the representation of each data point is fixed by a particular cell after training, redshift can be mapped as a function of colour. However, it should be mentioned that this is not true in the case of colour degeneracy.

The finite binning of the SOM manifold inherently discretises the data space, which can distort the true balance between calibration and validation samples. Since the SOM is not pop-

ulated evenly, this hurdle can be leveraged to identify and filter out poorly represented photometric data by excluding cells lacking spectroscopic information. Much like for clustering algorithms like k -nearest neighbour (k -NN) or probabilistic principal component analysis (PCA), objects with similar characteristics are grouped together on the SOM. Cells unoccupied by either photometric or spectroscopic sources are removed, as they indicate an imbalance in representation likely caused by selection effects or intrinsically underpopulated regions of colour space. This trade-off involves sacrificing a portion of the photometric sample but aligns with the overarching objective of mitigating redshift bias, as k -NN matching can extend to sources well beyond what one might consider the local region of the n -dimensional manifold (Wright et al. 2020a). Therefore, by painting the canvas of the colour space with the slice of well-represented photometric data, which we refer to as gold sample, one can also recognise regions that are difficult to access by direct spectroscopy and specifically target in spectroscopic surveys to widen the coverage of the colour-space to redshift mapping. The most notable implementation of this type of approach can be found in the Complete Calibration of the Color-Redshift Relation (C3R2) project (Masters et al. 2015, 2017, 2019; Stanford et al. 2021). Here, the C3R2 team endeavours to utilise SOMs to identify unexplored parts of the n -dimensional colour-redshift hyper-volume due to complex selection functions in the ability to measure distinct spectra (Hoyle et al. 2018; Wright et al. 2020a). They subsequently observe the spectra and thus, as the name suggests, extend the calibration of the colour-redshift relation for use in weak lensing surveys (Amendola et al. 2013; Wright et al. 2020a; Myles et al. 2021; Campos et al. 2024; *Euclid* Collaboration: Mellier et al. 2025).

Unlike other calibration efforts, such as geometry-dependent angular cross-correlation (Moessner & Jain 1998; d’Assignies et al. 2025), the two samples used for calibration and validation need not overlap in the projected sky plane. On the contrary, the spec- z required in the calibration sample can be accumulated from a multitude of surveys observing galaxies unrelated to those used to gather the photometry for both the calibration and validation samples. However, as with any ML algorithm, SOMs lack the ability to reliably extrapolate beyond the feature space covered by the calibration data, for example going to ever fainter magnitudes. As such, the calibration sample must cover the parameter range occupied by the validation sample passed to the SOM. As SOMs focus on the intrinsic properties of galaxies instead of positional parameters used in galaxy clustering, these methods are complementary (Newman 2008).

3. Data

The payload on *Euclid* feeds two instruments operating in parallel. Firstly, the panoramic very broad-band (r, i, z) visible-light 600 megapixel imager VIS with a single band wavelength range, I_E [$\lambda = 5300\text{--}9000 \text{ \AA}$] (*Euclid* Collaboration: Cropper et al. 2025). Secondly, the Near-Infrared 3-band ($Y_E, J_E,$ and H_E) Photometer (NIS-P), with wavelength ranges [$\lambda = 9500\text{--}12\,123 \text{ \AA}$], [$\lambda = 11\,676\text{--}15\,670 \text{ \AA}$], [$\lambda = 15\,215\text{--}20\,214 \text{ \AA}$] at a pixel size of 0.3 arcseconds, as well as a slitless Spectrograph (NIS-S) comprising two grisms jointly covering $\lambda = 9260\text{--}18\,920 \text{ \AA}$ (*Euclid* Collaboration: Jahnke et al. 2025). The *Euclid* wide survey (EWS), reaching (median) depths of $I_E = 24.5$, represents the core of the dark energy mission from which weak lensing,

baryon acoustic oscillation, and redshift space distortion signals will be measured (see [Euclid Collaboration: Mellier et al. 2025](#), for more details).

3.1. Simulated data: Flagship catalogue

In this section, we describe the development of synthetic observations resembling *Euclid*-like datasets, used here to validate the performance of the SOM redshift calibration. The *Euclid* Flagship mock galaxy catalogue v.2.1.10 (Flagship2, [Euclid Collaboration: Castander et al. 2025](#)) is the largest cosmologically simulated galaxy catalogue to date. Flagship is based on the $\sim 4 \times 10^{12}$ dark-matter particle N -body simulation containing 2.6 billion galaxies over 1/8th of the sky while extending up to redshift $z \leq 3$ ([Potter et al. 2017](#)). It reproduces the temporal evolution of large-scale structures with billions of dark matter halos hosting galaxies and WL observables. From this dark-matter cosmic web simulation ([Potter et al. 2017](#)), a synthetic galaxy catalogue, implementing the halo occupation distribution and halo abundance matching technique ([Berlind & Weinberg 2002](#)), has been generated ([Euclid Collaboration: Castander et al. 2025](#)). Synthetic galaxies in this simulation mimic, with great detail, the complex properties that observed sources display: ranging from their shapes, luminosities ([Blanton et al. 2003](#)), spectral energy distributions (SEDs, [Ilbert et al. 2006](#); [Polletta et al. 2007](#)), and emission lines in their spectra, to the gravitational lensing distortions that affect the light observed by distant galaxies and noise realisations for observed fluxes. An example of the Flagship2 galaxy colour-colour space is shown in Fig. 1. Each galaxy in the catalogue includes entries for observed spectroscopic redshifts (z_{obs}), photometric redshifts (z_p), and true redshifts (z_t) down to apparent magnitudes of $I_E < 25.75$ or $H_E < 24.25$. Here, z_{obs} accounts for both cosmic expansion and the peculiar velocities of the galaxies, including those of the halo and the galaxy within the halo. The estimation of z_p is achieved using either the Deep Neural Mixture Density Network (Deepz, [Eriksen et al. 2020](#)) or the Nearest Neighbours Photometric Redshifts (NNPZ, [Tanaka et al. 2018](#)) pipeline. Deepz is trained on 20 000 simulated noisy galaxies without any colour selection limited to $i_{\text{LSST}} < 23$, providing the full $p(z)$ and point estimates, which are defined as the first mode, respectively. Although the property distributions of its simulated training set are considered realistic, it is somewhat optimistic by construction. As a result, the achieved data quality exceeds the expected capabilities of *Euclid*. Therefore, to avoid artificially improving our results, we utilise NNPZ in its standard configuration (using 30 neighbours) throughout this study. NNPZ is a ML approach that estimates $p(z)$ using a k -nearest-neighbour search in flux or colour–magnitude space ([Cunha et al. 2009](#)). For each target galaxy, the method identifies nearby objects from a photometric reference sample that is deeper than the sample under consideration. The $p(z)$ distributions associated with these neighbours, which themselves are obtained from template-fitting photo- z codes applied to the reference sample, are then combined. Each neighbour is assigned a weight based on its distance in colour space, accounting for photometric uncertainties. The final $p(z)$ of the target galaxy is obtained as the weighted redshift histogram of these neighbours. Point estimates are available either as the first mode or the median of the $p(z)$, where the latter was used in this work. However, all analyses in this paper were also performed with Deepz, yielding similar or slightly improved results. The exclusive use of ML photo- z estimators, rather than a combination of ML and template-fitting methods, is primarily a practical con-

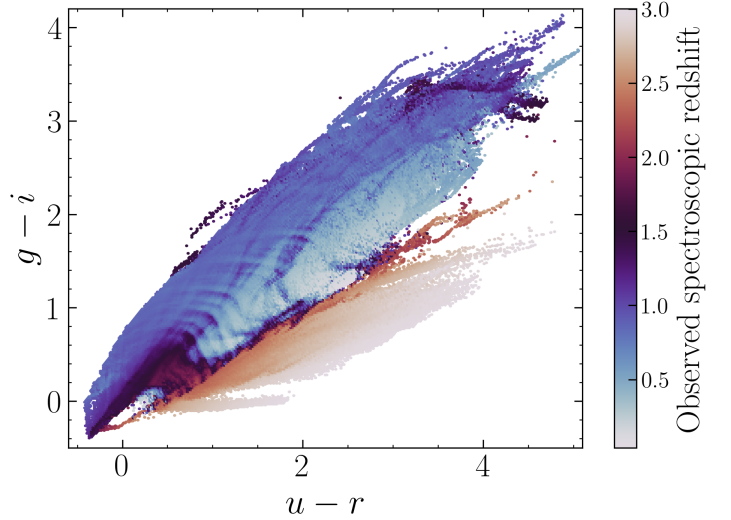


Fig. 1. Colour-colour diagram of Flagship2 galaxies free of flux errors plotted according to the corresponding true redshift. Blue represents the lower-redshift sources and red the higher-redshift sources [$0 \leq z_{\text{obs}} \leq 3$], exemplifying the non-linearity and degeneracies of the colour-redshift relation.

sequence of computational constraints. Running template-fitting codes to derive redshifts or subsequent physical properties for many millions of sources is prohibitively time-consuming, as these methods remain computationally expensive. Moreover, in Flagship2 the galaxy colours are generated directly from SED templates, suggesting such algorithms would perform artificially well in the simulation compared to real observations, where the diversity of galaxy SEDs is significantly larger.

3.2. Sample requirements

We start from a set of two isolated and compatible catalogues selected on fundamentally different specifications. The calibration sample used for training the SOM is designed to mimic non-trivial, real-life spectroscopic observations, including all current magnitude restrictions, realistic number counts, and quality indicators. The validation sample consists of a large enough number of data points or area on the sky to reduce sample noise to a negligible influence. The two distributions are not required to share an overlapping region on the sky but merely need to feature a similar parameter space regarding the variables used to train the SOM. In this manner, both samples (with $N_{\text{calib}} \ll N_{\text{valid}}$) exhibit comparable galaxy attributes, although they were selected using distinct criteria. Flagship2 does not inherently offer two samples of this nature; hence, they must initially be generated.

3.3. Flagship mocks

The Flagship2 simulation provides flux measurements of the observed photometry including contributions from both the continuum and emission lines, accounting for internal attenuation and Milky Way extinction, provided in cgs units of $\text{erg cm}^{-2} \text{s}^{-1} \text{Hz}^{-1}$. To ensure accurate representation of observational constraints, the photometric noise, driven by systematic effects (aperture size, detector filters, size of the PSF, or natural effects such as wavelength-dependent extinction or atmospheric seeing), has been calibrated to match the expected depths of the EWS in the southern hemisphere provided by the complementary ground-based surveys

at the DR3 time frame (Euclid Collaboration: Scaramella et al. 2022; Euclid Collaboration: Mellier et al. 2025). Given the flux descriptions above, the observed flux F_{obs} for each source can be derived as

$$F_{\text{obs}} = F_{\text{intrinsic}} + F_{\text{error}}, \quad (1)$$

with AB magnitudes

$$m_{\text{obs}} = -2.5 \log_{10}(F_{\text{obs}}) - 48.6. \quad (2)$$

The depth limits are defined at a 10σ level for a 2 arcsec diameter aperture for extended sources. Intrinsic (noiseless) and observed (noisy) colour and magnitude distributions for LSST ($ugriz$), as well as for *Euclid* (I_E, Y_E, J_E, H_E) at the depths expected to be reached by the third year, are displayed in Fig. 2, with limiting magnitudes up to 25.7 AB for extended sources. Note that we do not account for flux magnification, as the effect cancels for colours (Euclid Collaboration: Lepori et al. 2022), or the influence of variable survey depth. Due to computational load, we restrict our base sample to a 1/256th slice in the Flagship2 mock, corresponding to roughly 19 million sources from CosmoHub (Carretero et al. 2017; Tallada et al. 2020). After downselecting sources with MNPZ $z_p \leq 3$, we are left with roughly 6 million sources, while being largely unaffected by sample variance. This effect is distinct from shot noise given by the non-homogeneity of galaxies and cosmic variance induced by differing realisations of the observable Universe (Wright et al. 2020b).

In addition, we are particularly interested in how large-scale structures in the field affect the observed $n(z)$ and, in turn, the estimation of SOM redshift bias across tomographic bins. For this reason, the sample is further dissected into a grid of 10×10 equi-sized, completely independent line-of-sight sub-field samples (see Fig. 3). This allows us to probe the impact of sample variance in the training samples exclusively. Even though every sub-field sample features different galaxies, the influence of photometric noise is not investigated, as all samples use the same noise level. For each sub-field sample, the same training and calibration pipeline, as outlined in the following sections, is executed independently. Subsequently, we derive the mean bias and its uncertainty per tomographic bin, utilising 100 separately analysed samples.

4. Calibration efforts

All SOM analysis in this work relies exclusively on the implementation of the widely used kohonen package (Wehrens & Kruisselbrink 2018), as a branched version² within the programming framework R. To emulate *Euclid* observations, the Flagship2 sub-field samples are preprocessed using dynamic flux cuts, retaining only galaxies with signal-to-noise ratios (S/N) of $S/N > 5$ in I_E . This avoids imposing a fixed magnitude limit, such as the EWS threshold as mentioned in Sect. 3, and preserves fainter galaxies present in Flagship2, which is not constrained by this magnitude limit. We end up with a EWS-like catalogue of ~ 1.7 million galaxies. Nonetheless, this is a slightly pessimistic approach, as we assume all I_E detections to make it into the shear sample, which worsens the calibration. Lastly, all galaxies with negative fluxes after error inclusion in any band are removed from the reduced EWS-like sample.

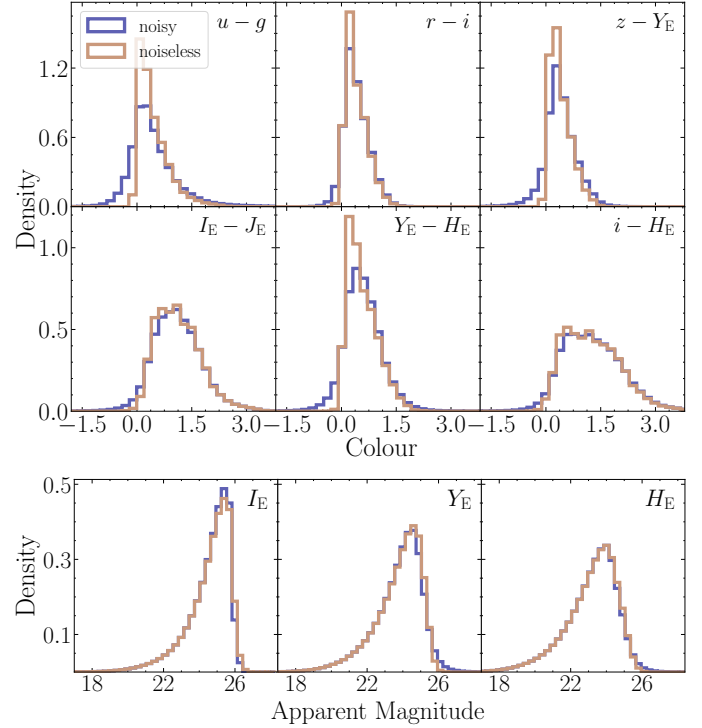


Fig. 2. Comparison of some exemplary noisy (blue) and noiseless (brown) Flagship2 EWS-like colours (top) and magnitudes (bottom). The bands comprise LSST u, g, r, i, z and *Euclid* I_E, Y_E, J_E, H_E at year 3 depth.

4.1. SOM architecture

Masters et al. (2015) implement a 75×150 SOM of rectangular nature for C3R2, allowing for asymmetrical multi-dimensional parameter space, arguing the manifold gives preferential direction to the principal geometry and therefore improves convergence. In addition to the raw number of cells in each SOM dimension, a different cell shape and surface topology can be chosen. The prior becomes particularly important when dealing with denser areas of the manifold, where it can directly affect the local data distribution. Conversely, SOM topology becomes most important in areas of the manifold which are sparsely populated. Generally, the choice of topology is either flat or toroidal, meaning the edges of the SOM manifold either act as boundaries or reconnect to form a continuous surface (Wright et al. 2020b). Finally, the fidelity of the SOM in terms of its spectroscopic representation is driven by what information is relevant to parse during training. Leveraging the extensive nine-band photometry available with Flagship2, we harness a rich array of colour information encompassing 36 magnitude combinations to focus on the colour-redshift relation. Additionally, we introduce the I_E magnitude as an extra parameter axis to enhance stability without introducing excessive redundancy. Nonetheless, there are undoubtedly more modifications that one can make to the SOM which are not considered in this work (Kind & Brunner 2013). In line with Wright et al. (2020b), the SOM shows considerably lower spectroscopic representation when trained solely on magnitudes, as in this scenario the SOM must first learn that the magnitude combinations, such as colours, are indicative features of redshift estimates. Consequently, adding further magnitudes as training parameters has little to no effect. The argumentation of training parameters, however, is to some extent arbitrary, as

² <https://github.com/AngusWright/kohonen.git>

Table 1. Comparison of the characteristics of the noiseless and noisy SOMs.

SOM parameter	Noiseless SOM	Noisy SOM
Training expression	36 colours & I_E magnitude	36 colours & I_E magnitude
Dimension	75×150	101×101
Topology	Flat and edge limited	toroidal and continuous
Cell type	Hexagon	Hexagon
Data threshold	[0, 30] mag	[0, 30] mag; $S/N(I_E) > 5$
Training iterations	200	200

Notes. The noiseless SOM Masters et al. (2015) layout, trained on noiseless Flagship2 photometry. Opposed to the noiseless SOM, the noisy SOM is not trained on a wide field sub-sample, but on the respective calibration sample, generated from the noiseless SOM.

human reasoning becomes increasingly imperfect for visualisations beyond 3 dimensions.

4.2. Calibration sample creation

Figure 3 provides an overview of the steps involved in the analysis described in this section. Following the approach taken in C3R2 (Masters et al. 2015), we utilise Flagship2 with roughly 150 000 calibration sources per sub-sample to train 100 Masters-like SOMs as characterised by the central column in Table 1. By sampling the SOMs, we create artificial spec- z calibration samples that closely resemble real-life availability, due to their non-trivial selection functions. Since flux errors are correlated, the introduction of noise to the photometry results in multi-directional shifts across the SOM. However, in the idealised Flagship2 mock catalogue, these higher-dimensional flux error cross-correlations do not exist, making it a more optimistic scenario compared to real *Euclid* observations (see Kang et al., in prep.). To prevent this restructuring of the original feature space from happening, and thus securing a uniformly populated and unbiased (Wright et al. 2020b) multi-dimensional parameter volume, the SOMs are instead trained on colours calculated using not entirely unrealistic noiseless photometry, by considering the greater depth of the *Euclid* auxiliary fields used for calibration.

Once training commences, the input data is whitened before being presented to the SOM. This involves zero centering, decorrelating, such that the covariance matrix becomes diagonal, and rescaling the data. By standardising the variances in each direction of feature space, whitening typically speeds up convergence and causes models to better capture contributions from low-variance feature directions, as training is not overshadowed by properties which tend to have larger values (Lecun et al. 1998). The training process is stopped after several hundred iterations, as the mean distance to the closest unit $d_j(t)$ drops below 0.005, as introduced in Appendix A. Training could, in principle, be extended to a larger number of iterations; however, continuous training can lead to over-fitting, leading to a drop in model performance. After training, the Masters-like SOM (noiseless SOM, hereafter) cells can be painted by the spatial distribution of numerous variables, such as the inhomogeneous occupation count or training sample-specific attributes, such as colours, as shown in Fig. 4. These plots show the noiseless SOM coloured according to its cell properties with red referring to lower values and blue to higher values. Empty cells holding no galaxies, are painted white. To now generate a C3R2-like successful spec- z calibration sample, individual galaxies are randomly drawn from the noiseless SOM. Here we address every cell individually and select ≤ 3 sources, respectively. These drawn samples

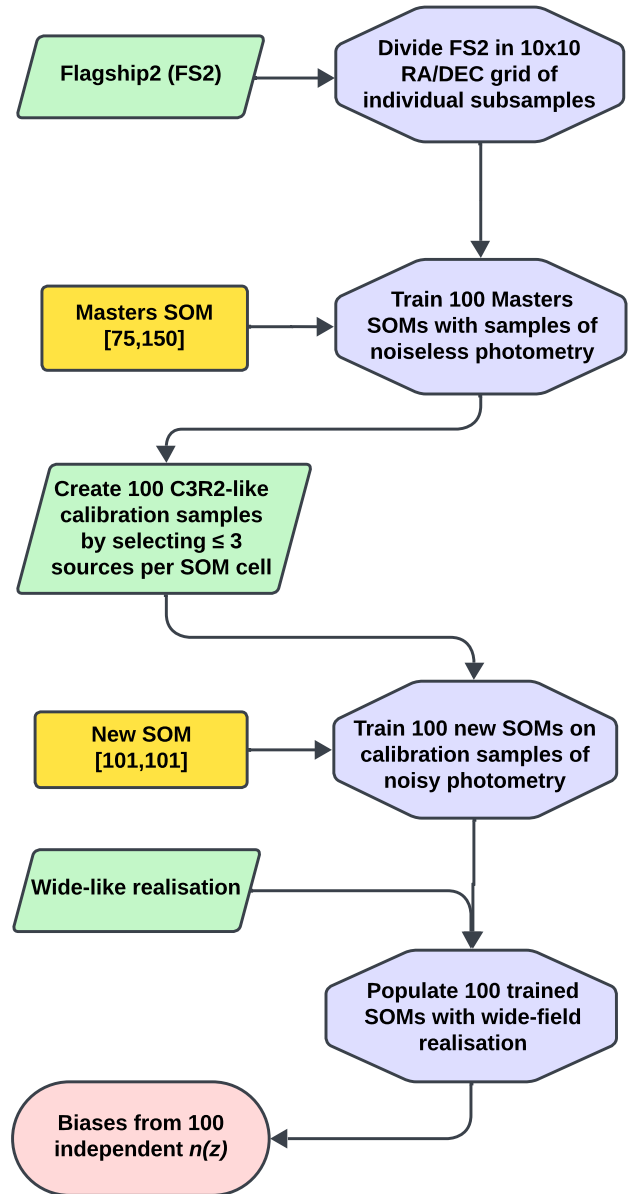


Fig. 3. Flowchart outlining the sequential process of utilising noiseless SOMs to create a set of calibration samples, which are then used to train a secondary set of SOMs using noisy photometric data. Lastly, the trained noisy SOMs are populated by the EWS-like data.

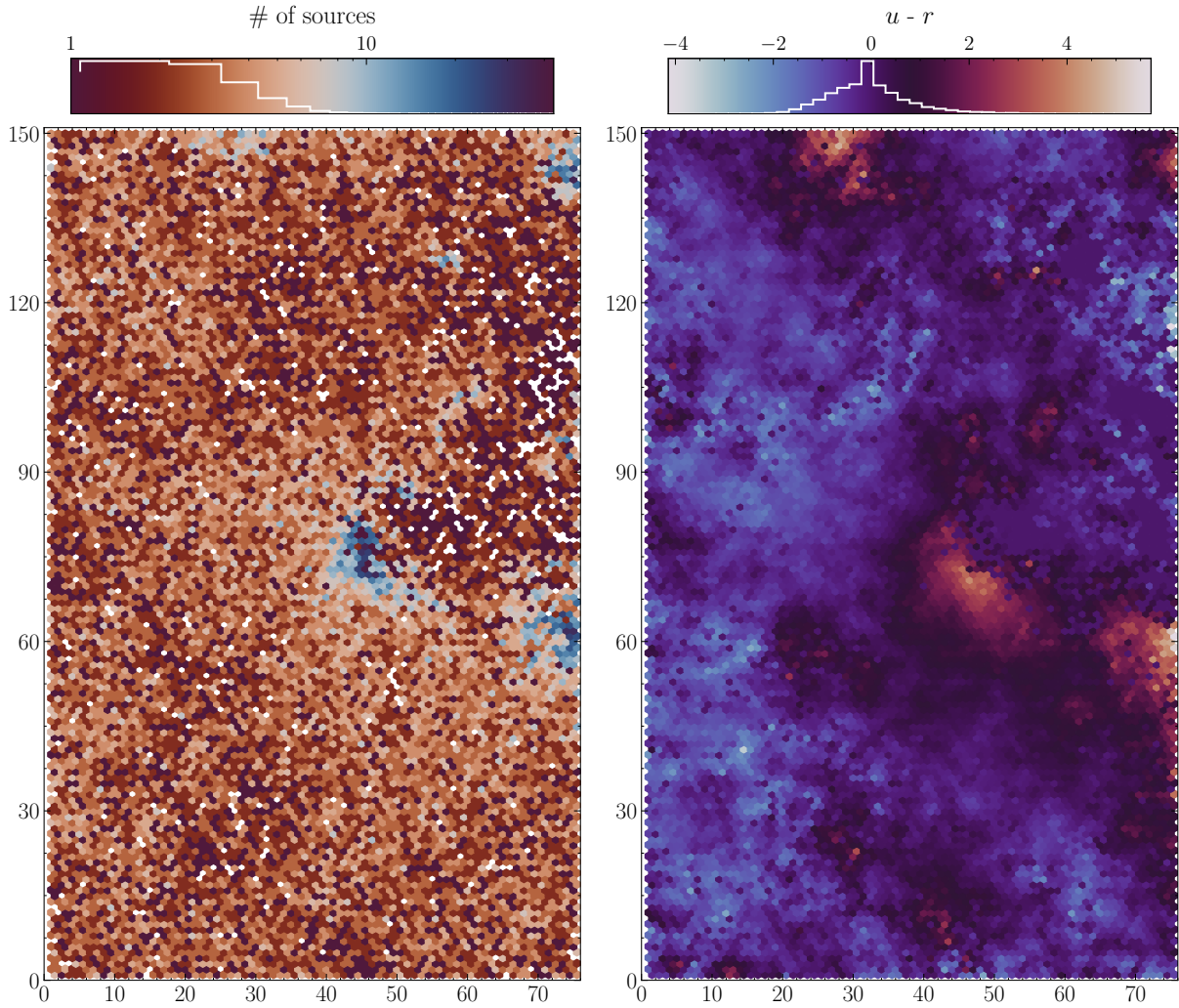


Fig. 4. Masters-like SOM coloured according to different properties of the calibration sample. These include the occupation count (left) and the cells’ average $u - r$ colour (right). Furthermore, cells lacking a representation of physical quantities are painted white. The white histograms overlaid within the upper colour bars represent the normalised linear distributions of the respective SOM statistics.

have an average size of $\sim 33\,000$ sources and will be referred to as calibration samples from here on. By excluding mean magnitude from the selection criteria, we avoid biasing the calibration sample towards brighter sources. Instead, the selection is driven solely by the colour–redshift space of the SOM.

4.3. SOM spec- z based tomography

In contrast to training the noiseless SOM on the photometric EWS-like sample, we train a second SOM on the noisy photometry of the calibration sample drawn from the noiseless SOM (see Table 1). Training a SOM on noiseless data before populating it with noisy data produces misleading results, since this difference in samples introduces covariate shift as introduced in Appendix B, rendering this exercise invalid. In contrast to the Masters’ SOM, this second, noisy SOM, features a symmetrical 101×101 grid of toroidal topology, meaning there are no hard boundaries limiting the SOM surface plane (see Table 1). This has the beneficial effect that data parsed to the SOM experiences no prioritised axis. Moreover, the SOM can more effectively encapsulate or wrap around the multi-dimensional data distribution. As shown in Fig. 5, the distribution of spec- z found in the validation sample shown on the right, tightly follows the

distribution forged during training shown on the left. In addition, areas of spectroscopic incompleteness (i.e. white cells in the left map), which cannot be used for training and areas of unrepresented colour-colour space following the S/N cut in the validation sample (i.e. white cells in the right plot), become detectable. Once the noisy SOM is trained, we propagate the entire EWS-like sample (~ 1.7 million sources), from now on referred to as validation sample, to the noisy SOM, producing like-for-like groupings.

We quantify the spectroscopic representation of photometric sources by way of example for a single independent calibration sub-sample in Table 2. The noisy SOM coverage statistics (Wright et al. 2020b) feature the overall size of the spectroscopic calibration sample (“Sample (full)”; 33 602 galaxies) and the fraction of cells these sources occupy (“ f_{pix} ”; 96.8%). The latter statistic indicates that $\approx 3.2\%$ of the noisy SOM are unoccupied by the calibration sample. However, the counts in the validation sample exhibit significant variation across the SOM, rendering this value less indicative of the absolute fraction of unrepresented sources in the wide sample. Therefore, Table 2 presents the fraction of cells f_{gs} , for each of the tomographic bins that contain both calibration and wide sample sources (see Sect. 4.3.2). It should also be mentioned that not all photometric

Table 2. Calibration sample distribution on noisy SOMs.

Sample (full) = 33 602										
f_{pix} (all, %) = 96.8										
Bins	bin 1	bin 2	bin 3	bin 4	bin 5	bin 6	bin 7	bin 8	bin 9	bin 10
$z \in$	(0.05, 0.295]	(-, 0.54]	(-, 0.785]	(-, 1.03]	(-, 1.275]	(-, 1.52]	(-, 1.765]	(-, 2.01]	(-, 2.255]	(-, 2.50]
f_{gs}	98.3	97.7	93.6	90.9	91.1	93.9	95.5	96.4	97.8	96.1
N_{spec}	2041	7404	6508	6334	4177	3430	1834	962	603	222

Notes. The table presents the sample size and fraction of noisy SOM cells occupied by the calibration sample, f_{pix} (all, %), along with the fractions of cells representative of the wide sample f_{gs} and spectroscopic number counts N_{spec} across 10 equi-sized tomographic bins (Wright et al. 2020b). The data reveals a notable scarcity of sources for calibration in bins 1, 7, 8, 9, and 10 compared to the other bins.

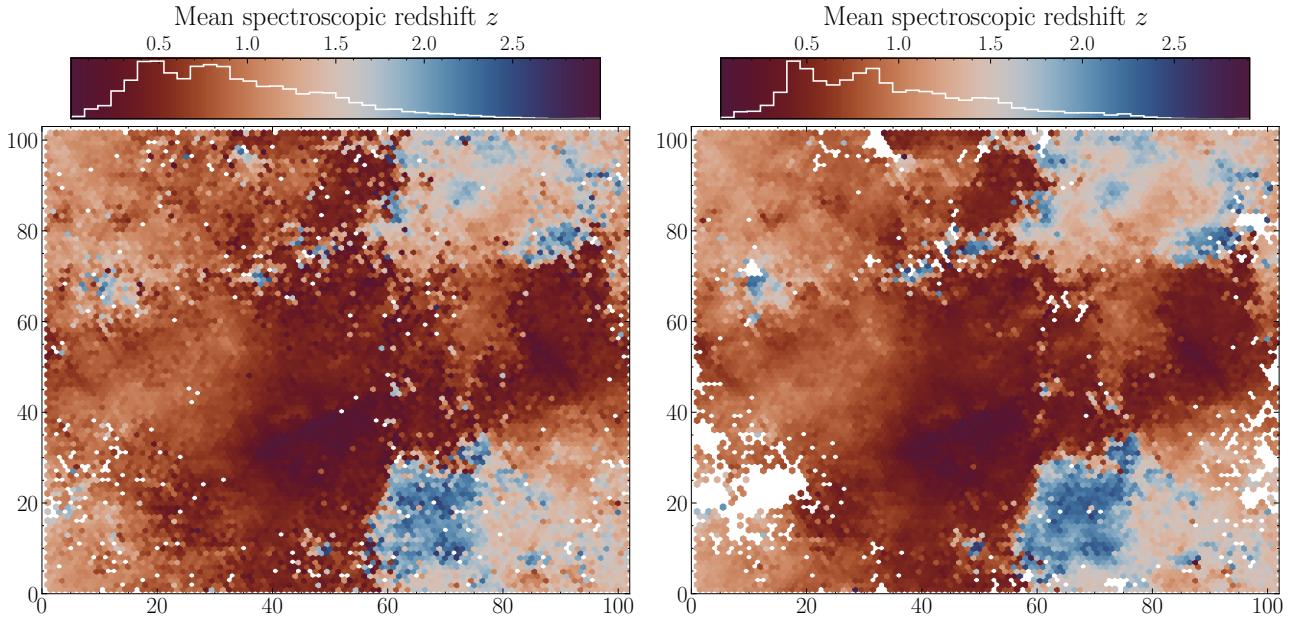


Fig. 5. Noisy SOM painted by the mean spec- z for both the calibration sample used in training (left panel) and the validation sample (right panel). A similar underlying pattern is visible. Areas of spectroscopic incompleteness, coloured in white, cannot be used for calibration purposes as per the gold sample selection.

sources hold the same lensing weight in cosmic shear estimates. Given the effective number of weighted photometric sources, merely $\approx 1.2\%$ of the validation sample sources located in cells unoccupied by the calibration sample, lie in relevant parts of the noisy SOM (Wright et al. 2020b).

4.3.1. Bias calculation

The mean spec- z of the calibration sample are compared to the mean spec- z of the validation sample per cell. The measurement within each cell is weighted by the number of galaxies from the validation sample allocated to that specific cell. The mean of these weighted deviations yields the true redshift bias,

$$B_s = \langle z_{\text{spec, cal}} - z_{\text{spec, val}} \rangle. \quad (3)$$

It is important to clarify that we do not know B_s , but we desire to constrain it since, in reality, we lack information on the spec- z of the EWS-like validation sample. In a similar fashion, the mean spec- z values from the calibration sample are compared to the mean photo- z of the validation sample per cell. The weighted mean of this quantity gives the photo- z bias

$$B_p = \langle z_{\text{spec, cal}} - z_{\text{photo, val}} \rangle. \quad (4)$$

Detached from the extensively investigated disparity between spectroscopic and photometric redshifts, we utilise Eq. (4) to get a handle on the entwined bias denoted by Eq. (3). This is motivated by the near-linear relationship between the biases, given that photo- z estimates typically provide an approximation to spec- z s with scatter of order $\sigma_{\Delta z/(1+z)} \sim 0.01\text{--}0.03$ and outlier fractions of a few percent for wide-field surveys (Coupon et al. 2009; Tanaka et al. 2018; Duncan 2022; Euclid Collaboration: Tucci et al. 2025). For this work, the redshift dimension in the range of $0 \leq z \leq 2.5$, is sliced into ten equi-sized tomographic redshift bins. The methodology used to allocate galaxies to their respective bins significantly affects the bias estimation. In an initial approach, the redshift tomography is established by assigning each cell to its corresponding redshift bin, based on the cells' mean spec- z derived from training, $z_{\text{spec, cal}}$.

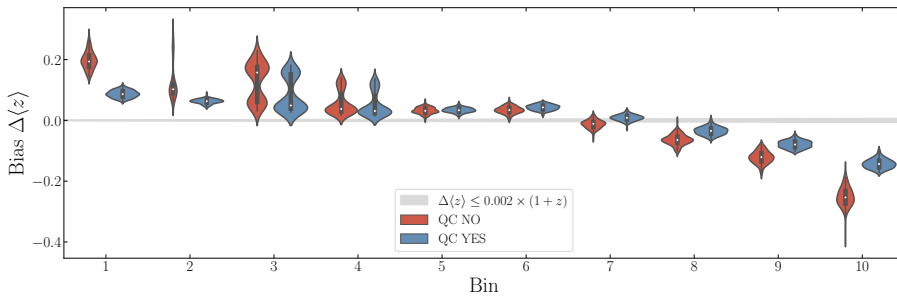
4.3.2. Gold sample

To enhance the statistical robustness of this approach, we employ gold sample selection. This selection is based on the identification cells without counterparts, whether spectroscopic or photometric, after projecting the EWS-like sample into the noisy SOM. Consequently, we exclusively focus on cells providing

Table 3. Biases per bin for redshift tomography defined on the mean spec- z of individual SOM cells.

Bias ($\Delta\langle z \rangle$)	bin 1	bin 2	bin 3	bin 4	bin 5
QC no	0.197 ± 0.029	0.122 ± 0.057	0.127 ± 0.061	0.056 ± 0.038	0.032 ± 0.010
QC yes	0.087 ± 0.012	0.062 ± 0.008	0.084 ± 0.056	0.050 ± 0.039	0.034 ± 0.008
Bias ($\Delta\langle z \rangle$)	bin 6	bin 7	bin 8	bin 9	bin 10
QC no	0.034 ± 0.012	-0.013 ± 0.014	-0.064 ± 0.018	-0.122 ± 0.020	-0.253 ± 0.034
QC yes	0.042 ± 0.009	0.008 ± 0.010	-0.035 ± 0.013	-0.079 ± 0.013	-0.143 ± 0.015

Notes. The values listed are the mean biases over 100 different line of sight sub-samples, including the respective standard deviation. Under the recognition of QC, merely bin 7 comes close to sufficing the dynamic *Euclid* bias requirement of $\Delta\langle z \rangle < 0.002(1+z)$ when excluding standard deviations and considering the upper redshift bin limits from Table 2. The expression ‘QC’ refers to the consideration of quality control (see Sect. 4.3.3).

**Fig. 6.** Violin plot of biases per bin using tomography defined by SOM spec- z before (red) and after (blue) applying QC. These distributions also include box-and-whisker plots. Lastly, the dynamic $\Delta\langle z \rangle < 0.002(1+z)$ *Euclid* requirement are given by the grey shaded area.

a guaranteed representation. This criterion effectively identifies and eliminates empty cells, along with the regions of the colour-colour space where the sample distributions significantly diverge (Wright et al. 2020b). Given that $N_{\text{calib}} \ll N_{\text{valid}}$, cells are more likely to lack calibration rather than validation sample sources, making this selection primarily responsible for removing validation sources from each bin. This may not be confused with the larger number of white, empty cells on the right hand side of Fig. 5. Although the validation sample is larger than the calibration sample, it spans a smaller feature space, following the one-sided S/N cut and use of noisy as opposed to noiseless photometry, as mentioned at the beginning of Sect. 4.

4.3.3. Quality control

To further reduce B_s , we apply an additional, stricter cell selection which we refer to as quality control (QC). It is applied with the goal to remove SOM cells likely to produce biased redshift estimates. Fainter sources generally have larger photometric uncertainties and noisier colours, which can broaden photo- z PDFs and introduce multimodality in the aggregated SOM-cell $p(z)$. However, since SOM cells are defined primarily in colour space, both bright and faint sources can occupy the same cell. Differences between cells passing and failing QC thus reflect a combination of photometric quality, colour-space structure, intrinsic source properties, and sample composition. QC down-selects cells based on their photo- z bias, such that cells are selected if

$$|B_p| < 0.5. \quad (5)$$

As elaborated earlier, we can exploit the correlation of the two biases to further constrain B_s by restricting the parameter space of B_p . On average, QC removes roughly 10% of galaxies from the wide-like sample, with the fraction increasing from $\sim 5\%$ at low redshift to $\sim 20\%$ at high redshift. Given the decreasing fraction of gold sample-selected cells for higher-redshift bins, the QC condition could alternatively be defined dynamically

as a function of redshift. Therefore, the implemented QC steps serve merely as a demonstration of the refinement possible using SOMs (Wright et al. 2020b). The results of the spec- z biases per bin as defined by the mean spec- z of SOM cells, subsequent to gold sample selection and QC consideration, are depicted in Table 3 and Fig. 6, respectively. We note that, in line with the issues raised in Appendix B, none of the bins in this configuration satisfy the *Euclid* requirement outlined in Sect. 1. All biases but those in bins 5, 6, 7, and 8 are more than a factor of 10 too large. Furthermore, the biases estimated by this method change sign from positive for the first six bins to negative for the remaining four bins, limiting the impact of the redshift bias on cosmological conclusions somewhat (Wright et al. 2020b). This trend can be explained by the offset in redshift distributions between the two SOM cell samples, causing the reconstructed $n(z)$ of the wide sample to be compressed towards the calibration mean, effectively pulling the distribution tails into the middle, due to the lower number of calibration sources in these bins. Table 3 demonstrates that, except for bins 5 and 6, the implementation of QC reduces the redshift bias in the majority of cases. We note that the above-mentioned results do not account for a division of the northern and southern hemispheres. This should be kept in mind, as each hemisphere observes sources with different instruments and therefore bands that may influence the SOM.

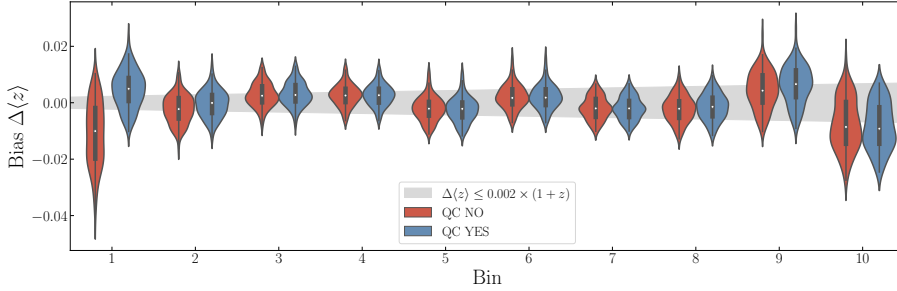
4.4. Photo- z based tomography

An alternative approach to defining redshift tomography is based on the bin assignment from individual NNPZ PDF medians as photo- z point estimates of the calibration sample, as briefly discussed in Appendix B. SOM cells can contain heterogeneous galaxy populations, which may lead to non-Gaussian or even bimodal $p(z)$ within a cell, since in such cases the mean of the SOM-cell $p(z)$ is not necessarily representative of the dominant population. If, instead, redshift binning is performed using per-galaxy photo- z estimates, then a multimodal SOM cell can naturally be sub-sampled as most individual photo- z PDFs are

Table 4. Biases per bin for redshift tomography defined on the mean photo- z of individual sources.

Bias ($\Delta\langle z \rangle$)	bin 1	bin 2	bin 3	bin 4	bin 5
QC no	-0.011 ± 0.011	-0.002 ± 0.005	0.003 ± 0.004	0.003 ± 0.004	-0.002 ± 0.004
QC yes	0.005 ± 0.007	0 ± 0.005	0.003 ± 0.004	0.003 ± 0.004	-0.002 ± 0.004
Bias ($\Delta\langle z \rangle$)	bin 6	bin 7	bin 8	bin 9	bin 10
QC no	0.002 ± 0.004	-0.002 ± 0.004	-0.002 ± 0.005	0.004 ± 0.007	-0.008 ± 0.009
QC yes	0.002 ± 0.004	-0.002 ± 0.004	-0.002 ± 0.005	0.006 ± 0.007	-0.009 ± 0.008

Notes. Same as Table 3, where all but bins 1 and 10 lie within the dynamic *Euclid* bias requirement of $(\Delta\langle z \rangle) < 0.002(1+z)$ when excluding standard deviations. Biases with values $(\Delta\langle z \rangle) < 0.001$ are marked with a null symbol “0”.


Fig. 7. Same as Fig. 6, but using tomography defined by the photo- z of individual galaxies.

unimodal, making their medians effectively equivalent to their modes. This is particularly relevant for cells, where the calibration and wide-like redshift distributions differ significantly. While bimodality in individual photo- z PDFs does occur, we find it to be considerably less common than multimodality in the cell-aggregated SOM $p(z)$; therefore, this issue is substantially reduced when using photo- z point estimates.

In line with the pipeline used for tomography defined on the mean spec- z of the SOM, gold sampling and QC cuts are also applied to the photo- z based tomography. As in Fig. 6, the biases estimated with this tomography (see Fig. 7) change sign. In this case, bin 1 as well as bins 2, 5, 7, 8 and 10 have a mean negative sign, whereas the remaining bins have a mean positive sign. However, in contrast to Fig. 6, there is no degrading bias trend. Figure 7 and Table 4 demonstrate that defining the redshift tomography on photo- z rather than on the mean spec- z of every SOM cell (see Table 3), provides much improved bias results for noisy photometry. In fact, considering QC, all but two bin means lie within the *Euclid* requirement for $\Delta\langle z \rangle$. This enables the use of nearly the entire spectroscopic sample for calibration, as its selection is no longer tied to the mean spec- z of SOM cells (see Fig. D.1). To evaluate the effectiveness of galaxy selection strategies in the SOM, we illustrated the two approaches in Fig. C.1. Galaxies selected via photo- z binning in the horizontal band exhibit significantly more compact and localised contours, well aligned with the selected redshift slice. In contrast, the SOM cell-based selection leads to a broader distribution that extends well beyond the target bin, as reflected in the larger extent of its contours. This highlights that photo- z binning within the SOM offers greater fidelity in isolating galaxies truly belonging to the desired tomographic redshift range.

5. Cosmological parameter biases

Given the bias uncertainties expected for redshift measurements performed by *Euclid*, one can propagate those into biases of cosmological parameters. The obtained level of accuracy to quantify this task is provided via the margin of uncertainty regarding the

cosmological parameters of interest. Consequently, said uncertainties are calculated based on the results obtained in Sect. 4.4 and Appendix C.

To estimate the effect on cosmological parameter inference, we sample the shifts of the mean redshift measurements from their covariance matrix. Next, we follow the prescription denoted in Reischke (2024) and convert those samples into shifts $n^{(m)}(z_n)$ in the functional space of the $n(z)$. Here the redshift support was already discretised, i.e. $\Delta n^m(z_n)$ is the shift (uncertainty) of the m -th tomographic bin, evaluated at the n -th redshift. These shifts can be rearranged into a single vector Δn^α and we use Greek indices to denote indices running over the $n(z)$ shifts. By estimating the response of the lensing angular power spectrum, C_ℓ , to changes in the $n(z)$ via functional derivatives (Reischke 2024), one can propagate these uncertainties into the shifts of cosmological parameters, $\Delta\theta^i$, to be

$$\Delta\theta^i = -(\mathbf{F})^i_k \mathcal{F}_\beta^k \Delta n^\beta. \quad (6)$$

Here we assume the sum convention, so that repeated indices appearing both as vector and dual vector index are summed over. Latin indices are used as labels for cosmological parameters, while \mathcal{F}_β^k is the mixed pseudo-Fisher matrix

$$\mathcal{F}_\beta^k := -\mathbb{E} \left[\frac{\partial \ln L}{\partial \theta_k} \frac{\delta \ln L}{\delta n^\beta} \mathcal{D}\chi_\beta \right]. \quad (7)$$

The functional derivative is indicated with δ with $\mathcal{D}\chi_\alpha$ as integration measure that arises from the discretisation (we refer to Reischke 2024, for more details). Note that the index β is not summed over in the above equation, as it appears as a dual vector index twice. \mathbf{F} is the Fisher matrix of the cosmological parameters. Lastly, L is the likelihood, which we assume to be Gaussian for the forecasting done here, by this assumption, all cosmological information is contained in the angular power spectrum, C_ℓ , where ℓ is the multipole. Given a set of tomographic bins, we collect all spectra in a matrix \mathbf{C}_ℓ , where the Fisher matrix in this case is given by (Tegmark et al. 1997)

$$F^i_k = f_{\text{sky}} \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \frac{2\ell+1}{2} \text{tr} \left[\mathbf{C}_\ell^{-1} \partial^i \mathbf{C}_\ell \mathbf{C}_\ell^{-1} \partial_k \mathbf{C}_\ell \right], \quad (8)$$

Table 5. Biases on cosmological parameters due to redshift uncertainties.

Case	$\Delta\Omega_m/\sigma_{\Omega_m}$	$\Delta\sigma_8/\sigma_{\sigma_8}$	$\Delta w_0/\sigma_{w_0}$	$\Delta w_a/\sigma_{w_a}$
Realistic and QC yes	0.27 ± 0.16	0.11 ± 0.06	0.46 ± 0.26	0.24 ± 0.14
Realistic and QC no	0.28 ± 0.15	0.11 ± 0.06	0.46 ± 0.26	0.24 ± 0.13
Best and QC yes	0.09 ± 0.02	0.04 ± 0.01	0.16 ± 0.04	0.08 ± 0.02
Best and QC no	0.08 ± 0.02	0.03 ± 0.01	0.14 ± 0.04	0.07 ± 0.02

Notes. All values are given in units of σ considering the four different cases of redshift calibration (refer to Sect. 4.4 and Appendix C). The errors are given as standard deviations of the distribution displayed in Fig. 8. Note that they are given as symmetric only for reference.

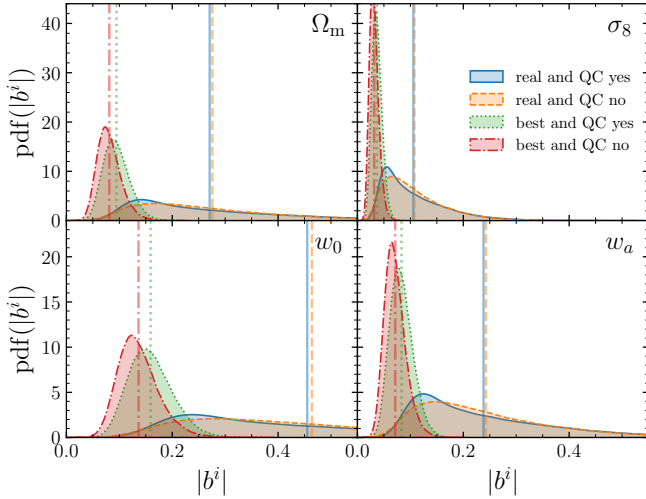


Fig. 8. Forecasted absolute biases in units of standard deviations for cosmological parameters. Each panel shows a different cosmological parameter; the different line styles and colours correspond to the different cases presented in Sect. 4.4 and Appendix C. The vertical lines indicate the expected bias corresponding to each case. The kernel density estimate was taken from 10^6 realisations of the $n(z)$.

with the sky fraction, f_{sky} . Deviating from the tomographic bin definition used for calibration purposes in this work, we instead follow the boundaries of the 10 equi-populated bins defined in [Euclid Collaboration: Blanchard et al. \(2020\)](#). The final cosmic shear only forecast biases are reported as

$$b^i = \frac{\Delta\theta^i}{[(\mathbf{F}^{-1})_{ii}]^{1/2}}, \quad (9)$$

in units of the marginalised standard deviation of a forecasted *Euclid* setting (labelled as optimistic in [Euclid Collaboration: Blanchard et al. 2020](#)). In particular, we focus on the matter density, Ω_m , the power spectrum amplitude σ_8 , and two equations of state parameters w_0 and w_a . The following scenarios are assumed: the calibration settings which achieve the lowest biases (best), see Sect. 4.4 and those which achieve the most realistic biases (real), see Appendix C.3.

Table 5 shows mean relative biases across realisations, expressed in units of the statistical uncertainties. Since our forecast is optimistic by design, we expect the real-data scenario to be more favourable due to conservative assumptions in our modelling. In fact, we find that the lower idealised bias outperforms the realistic biases for all cases and each parameter. Furthermore, QC does not necessarily reduce the biases for the cosmological parameters. The lowest bias is induced on σ_8 which is to be expected, as completely coherent shifts (required for an overall amplitude change leading to a significant σ_8 bias) dur-

ing sampling from the $n(z)$ are very unlikely. While both w_0 and w_a are affected by uncertainties in the $n(z)$, we find that w_0 , the present-day value of the DE EOS, is more susceptible to systematic biases caused by redshift errors than w_a , which characterises its evolution with redshift. The biases are displayed in Fig. 8 and appear generally asymmetric with a long tail to larger biases, in excess of 1σ . The vertical lines indicate the expected bias, showing again that w_0 is most affected, while all biases are smaller than 1σ on average. In general, one should keep in mind that the uncertainties are derived from an optimistic setting without the inclusion of any additional systematic effects. Hence a fair assessment of the performance of the redshift calibration is to treat it as idealised, since a more realistic analysis would produce weaker constraints on cosmological parameters, decreasing the significance of their biases due to the shifts in the $n(z)$.

6. Discussion and outlook

In this paper, we validated various approaches regarding the calibration of photometric redshift distributions for *Euclid*. The obtained results are indicative as to whether the current aims and goals concerning the implementation of Stage IV precision cosmology are achievable assuming DR3-level data or require updated methodologies.

Similar to previous work performed for KiDS, such as by [Wright et al. \(2020b, 2025b\)](#), this work focused on implementing a SOM to reduce $n(z)$ biases within 10 tomographic bins [$0 < z \leq 2.5$] to satisfy the *Euclid* precision requirement. In extension of this, two types of fundamental tomography layouts were considered, with additional steps undertaken to increase the level of realism. With mocks from the Flagship2 catalogue, we trained a SOM on a calibration sample subject to a non-trivial selection function, before populating it with a large EWS-like photometric sample (see Sect. 4.2). The obtained biases were investigated under the implementation of gold-sample selection and QC control cuts. In the set-up, where the redshift tomography was defined on the mean spec- z per SOM cell, we find that, considering QC, none of the ten biases lie within the *Euclid* requirement (see Sect. 4.3). In contrast, where the redshift tomography was defined on photo- z , we find that eight out of ten biases satisfy the *Euclid* requirement (see Sect. 4.4). This strongly implies that redshift tomography should generally be defined using photo- z s of individual galaxies, rather than the mean spec- z of SOM cells. The level of realism of the calibration sample used in this approach was subsequently increased by adjusting it to mimic the C3R2 survey sample. For this, three different set-ups were implemented, for which we demonstrate the necessity of spectra in typically under-sampled colour-colour space and the performance increase achievable with photo- z tomography. As a result, in Appendix C.3, six out of ten bins satisfy the *Euclid* requirement. In addition, ideas and suggestions to further raise the level of realism were discussed.

Lastly, we utilised a cosmic shear only Fisher forecast to estimate the biases induced into the measurement of cosmological parameters considering the different calibration approaches. By sampling from the covariance of shifts in the mean of the photo- z distribution, we deduced parameter bias distributions for Ω_m , σ_8 , w_0 , and w_a . We showed, that while the relative bias of w_0 displays an increased mean due to the long tail of its distribution, all other biases are observed to be lower than 0.1σ in the idealised case and 0.3σ for increased realism (see Sect. 5). Generally, biases in the photo- z distribution lead to larger biases in parameters controlling the temporal evolution of the background cosmology and perturbations.

Despite these promising results, it should be in the interest of the precision cosmology community to assemble an extensive spec- z calibration sample. Significant challenges related to survey limitations persist, and preparing such a sample appears crucial to pave the way for successful redshift calibration. As our analysis was conducted under idealised conditions, it serves primarily as a proof of concept to guide the optimal definition of redshift tomography. While we strongly recommend the adoption of our findings, they will need to be adapted and re-evaluated in the context of the data available at the time of the first cosmological analysis with *Euclid* DR3.

Acknowledgements. WR acknowledges DLR support (Foerderkennzeichen 50002207). 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 BMK, 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 Onderzoeksschool 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). This work has made use of CosmoHub, developed by PIC (maintained by IFAE and CIEMAT) in collaboration with ICE-CSIC. CosmoHub received funding from the Spanish government (MCIN/AEI/10.13039/501100011033), the EU NextGeneration/PRTR (PRTR-C17.11), and the Generalitat de Catalunya.

References

- Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, *Phys. Rev. D*, **98**, 043526
- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, *PASJ*, **70**, S4
- Amendola, L., Appleby, S., Bacon, D., et al. 2013, *Liv. Rev. Relat.*, **16**, 6
- Arnouts, S., & Ilbert, O. 2011, *Astrophysics Source Code Library* [record ascl:[1108.0097](https://arxiv.org/abs/1108.0097)]
- Asghari, M., Jiménez, J. B., Khosravi, S., & Mota, D. F. 2019, *JCAP*, **04**, 042
- Autenrieth, M., van Dyk, D. A., Trotta, R., & Stenning, D. C. 2021, *ArXiv e-prints* [arXiv:[2106.11211](https://arxiv.org/abs/2106.11211)]
- Beck, R., Lin, C.-A., Ishida, E. E. O., et al. 2017, *MNRAS*, **468**, 4323
- Benitez, N. 2000, *ApJ*, **536**, 571
- Berlind, A. A., & Weinberg, D. H. 2002, *ApJ*, **575**, 587
- Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, *ApJ*, **592**, 819
- Bordoloi, R., Lilly, S. J., Amara, A., et al. 2012, *MNRAS*, **421**, 1671
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, **686**, 1503
- Buchs, R., Davis, C., Gruen, D., et al. 2019, *MNRAS*, **489**, 820
- Caldwell, R. R., Dave, R., & Steinhardt, P. J. 1998, *Phys. Rev. Lett.*, **80**, 1582
- Campos, A., Yin, B., Dodelson, S., et al. 2024, *ArXiv e-prints* [arXiv:[2408.00922](https://arxiv.org/abs/2408.00922)]
- Carretero, J., Tallada, P., Casals, J., et al. 2017, *Proceedings of the European Physical Society Conference on High Energy Physics*, 5–12 July, 488
- Coupon, J., Ilbert, O., Kilbinger, M., et al. 2009, *A&A*, **500**, 981
- Cunha, C. E., Lima, M., Oyaizu, H., Frieman, J., & Lin, H. 2009, *MNRAS*, **396**, 2379
- d'Assisgnies, W., Manera, M., Padilla, C., et al. 2025, *A&A*, **702**, A155
- Davidzon, I., Laigle, C., Capak, P. L., et al. 2019, *MNRAS*, **489**, 4817
- Duncan, K. J. 2022, *MNRAS*, **512**, 3662
- Erben, T., Hildebrandt, H., Miller, L., et al. 2013, *MNRAS*, **433**, 2545
- 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 (Lepori, F., et al.) 2022, *A&A*, **662**, A93
- Euclid Collaboration (Scaramella, R., et al.) 2022, *A&A*, **662**, A112
- Euclid Collaboration (Castander, F. J., et al.) 2025, *A&A*, **697**, A5
- Euclid Collaboration (Cropper, M. S., et al.) 2025, *A&A*, **697**, A2
- Euclid Collaboration (Jahnke, K., et al.) 2025, *A&A*, **697**, A3
- Euclid Collaboration (Mellier, Y., et al.) 2025, *A&A*, **697**, A1
- Euclid Collaboration (Tucci, M., et al.) 2025, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202554588> (*Euclid* Q1 SI)
- Frieman, J. A., Turner, M. S., & Huterer, D. 2008, *ARA&A*, **46**, 385
- Gatti, M., Giannini, G., Bernstein, G. M., et al. 2021, *MNRAS*, **510**, 1223
- Geach, J. E. 2011, *MNRAS*, **419**, 2633
- Giannini, G., Alarcon, A., Gatti, M., et al. 2024, *MNRAS*, **527**, 2010
- Gruen, D., & Brimiouille, F. 2017, *MNRAS*, **468**, 769
- Hartley, W. G., Chang, C., Samani, S., et al. 2020, *MNRAS*, **496**, 4769
- Hikage, C., Oguri, M., Hamana, T., et al. 2019, *PASJ*, **71**, 43
- Hildebrandt, H., Viola, M., Heymans, C., Joudaki, S., et al. 2017, *MNRAS*, **465**, 1454
- Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, *A&A*, **647**, A124
- Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, *MNRAS*, **478**, 592
- Huterer, D., Takada, M., Bernstein, G., & Jain, B. 2006, *MNRAS*, **366**, 101
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, **457**, 841
- Ivezić, Ž., & Kahn, S. M. 2019, *ApJ*, **873**, 111
- Johnston, H., Wright, A. H., Joachimi, B., et al. 2021, *A&A*, **648**, A98
- Kilbinger, M. 2015, *Rep. Prog. Phys.*, **78**, 086901
- Kind, M. C., & Brunner, R. J. 2013, *MNRAS*, **432**, 1483
- Kohonen, T. 1982, *Biol. Cybern.*, **43**, 59
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *ArXiv e-prints* [arXiv:[1110.3193](https://arxiv.org/abs/1110.3193)]
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, *Proc. IEEE*, **86**, 2278
- Li, S.-S., Hoekstra, H., Kuijken, K., et al. 2023, *A&A*, **679**, A133
- Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, *MNRAS*, **390**, 118
- Luo, S., Luo, J., Chen, Y., et al. 2022, *Pattern Recognition and Computer Vision: 5th Chinese Conference, PRCV 2022, Shenzhen, China, November 4–7, 2022, Proceedings, Part II*, Springer, 706
- Mandelbaum, R. 2018, *ARA&A*, **56**, 393
- Masters, D., Capak, P., Stern, D., et al. 2015, *ApJ*, **813**, 53
- Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, *ApJ*, **841**, 111
- Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, *ApJ*, **877**, 81
- McQuinn, M., & White, M. 2013, *MNRAS*, **433**, 2857
- Merz, G., Liu, X., Schmidt, S., et al. 2024, *ArXiv e-prints* [arXiv:[2411.18769](https://arxiv.org/abs/2411.18769)]
- Moessner, R., & Jain, B. 1998, *MNRAS*, **294**, L18
- Morrison, C. B., Hildebrandt, H., Schmidt, S. J., et al. 2017, *MNRAS*, **467**, 3576
- Myles, J., Alarcon, A., Amon, A., et al. 2021, *MNRAS*, **505**, 4249
- Newman, J. A. 2008, *ApJ*, **684**, 88
- Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015, *Astropart. Phys.*, **63**, 81
- Planck Collaboration I. 2020, *A&A*, **641**, A1
- Polletta, M., Tajer, M., Maraschi, L., et al. 2007, *ApJ*, **663**, 81
- Potter, D., Stadel, J., & Teyssier, R. 2017, *Comput. Astrophys. Cosmol.*, **4**, 2
- Reischke, R. 2024, *MNRAS*, **530**, 4412
- Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, *PASP*, **128**, 104502
- Salvato, M., Ilbert, O., & Hoyle, B. 2019, *Nat. Astron.*, **3**, 212
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, *ArXiv e-prints* [arXiv:[1503.03757](https://arxiv.org/abs/1503.03757)]
- Stanford, S. A., Masters, D., Darvish, B., et al. 2021, *ApJS*, **256**, 9
- Stölzner, B., Joachimi, B., Korn, A., Hildebrandt, H., & Wright, A. H. 2021, *A&A*, **650**, A148
- 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
- Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, *ApJ*, **480**, 22
- Thomas, D. B., Kopp, M., & Skordis, C. 2016, *ApJ*, **830**, 155
- Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, *Phys. Rev. D*, **98**, 043528
- van den Busch, J. 2021, Ph.D. Thesis, University of Bochum
- van den Busch, J. L., Wright, A. H., Hildebrandt, H., et al. 2022, *A&A*, **664**, A170
- Wehrens, R., & Krusselbrink, J. 2018, *J. Stat. Softw.*, **87**
- Weinberg, D. H., Mortonson, M. J., & Eisenstein, D. J. 2013, *Phys. Rep.*, **530**, 87
- Wong, J. H. W., Brown, M. L., Duncan, C. A. J., et al. 2026, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202553742>
- 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., Stözlner, B., Asgari, M., et al. 2025a, *A&A*, 703, A158
- Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2025b, *A&A*, 703, A144
- Zuntz, J., Lanusse, F., Malz, A. I., et al. 2021, *Open J. Astrophys.*, 4, 13
-
- ¹ Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
 - ² Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
 - ³ Universität Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
 - ⁴ Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France
 - ⁵ Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra, (Barcelona), Spain
 - ⁶ Serra Hünter Fellow, Departament de Física, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain
 - ⁷ INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
 - ⁸ Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
 - ⁹ Department of Astronomy, University of Geneva, Ch. d'Ecogia 16, 1290 Versoix, Switzerland
 - ¹⁰ Leiden Observatory, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands
 - ¹¹ ESAC/ESA, Camino Bajo del Castillo, s/n, Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
 - ¹² School of Mathematics and Physics, University of Surrey, Guildford, Surrey GU2 7XH, UK
 - ¹³ INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
 - ¹⁴ IFPU, Institute for Fundamental Physics of the Universe, Via Beirut 2, 34151 Trieste, Italy
 - ¹⁵ INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
 - ¹⁶ INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste, TS, Italy
 - ¹⁷ SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, TS, Italy
 - ¹⁸ Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
 - ¹⁹ INFN-Sezione di Bologna, Viale Bertini Pichat 6/2, 40127 Bologna, Italy
 - ²⁰ INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, 35122 Padova, Italy
 - ²¹ Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
 - ²² Dipartimento di Fisica, Università di Genova, Via Dodecaneso 33, 16146 Genova, Italy
 - ²³ INFN-Sezione di Genova, Via Dodecaneso 33, 16146 Genova, Italy
 - ²⁴ Department of Physics "E. Pancini", University Federico II, Via Cinthia 6, 80126 Napoli, Italy
 - ²⁵ INAF-Osservatorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli, Italy
 - ²⁶ Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
 - ²⁷ INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
 - ²⁸ INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese, (TO), Italy
 - ²⁹ European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
 - ³⁰ Institute Lorentz, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
 - ³¹ INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
 - ³² INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
 - ³³ INFN-Sezione di Roma, Piazzale Aldo Moro, 2 – c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy
 - ³⁴ Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
 - ³⁵ Port d'Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra, (Barcelona), Spain
 - ³⁶ Institut d'Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels, Barcelona, Spain
 - ³⁷ Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
 - ³⁸ Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, 52056 Aachen, Germany
 - ³⁹ INFN Section of Naples, Via Cinthia 6, 80126 Napoli, Italy
 - ⁴⁰ Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
 - ⁴¹ Dipartimento di Fisica e Astronomia "Augusto Righi" – Alma Mater Studiorum Università di Bologna, Viale Bertini Pichat 6/2, 40127 Bologna, Italy
 - ⁴² Instituto de Astrofísica de Canarias, Vía Láctea, 38205 La Laguna, Tenerife, Spain
 - ⁴³ Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
 - ⁴⁴ Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
 - ⁴⁵ European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
 - ⁴⁶ Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, Villeurbanne F-69100, France
 - ⁴⁷ Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
 - ⁴⁸ Institució Catalana de Recerca i Estudis Avançats (ICREA), Pas-seig de Luíis Companys 23, 08010 Barcelona, Spain
 - ⁴⁹ UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, 4 Rue Enrico Fermi, 69622 Villeurbanne, France
 - ⁵⁰ Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
 - ⁵¹ Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal
 - ⁵² Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
 - ⁵³ INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
 - ⁵⁴ Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
 - ⁵⁵ INAF-Istituto di Astrofisica e Planetologia Spaziali, Via del Fosso del Cavaliere, 100, 00100 Roma, Italy
 - ⁵⁶ Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France
 - ⁵⁷ Space Science Data Center, Italian Space Agency, Via del Politecnico snc, 00133 Roma, Italy
 - ⁵⁸ INFN-Bologna, Via Irnerio 46, 40126 Bologna, Italy
 - ⁵⁹ University Observatory, LMU Faculty of Physics, Scheinerstrasse 1, 81679 Munich, Germany
 - ⁶⁰ Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029, Blindern 0315, Oslo, Norway
 - ⁶¹ Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
 - ⁶² Felix Hormuth Engineering, Goethestr. 17, 69181 Leimen, Germany
 - ⁶³ Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
 - ⁶⁴ Cosmic Dawn Center (DAWN), Denmark
 - ⁶⁵ Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis Boulevard Arago, 75014 Paris, France
 - ⁶⁶ Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
 - ⁶⁷ NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

- ⁶⁸ Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
- ⁶⁹ Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Finland
- ⁷⁰ Department of Physics, P.O. Box 64, 00014 University of Helsinki, Finland
- ⁷¹ Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
- ⁷² Laboratoire d'étude de l'Univers et des phénomènes eXtremes, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 92190 Meudon, France
- ⁷³ SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, Cheshire SK11 9FT, UK
- ⁷⁴ Dipartimento di Fisica "Aldo Pontremoli", Università degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy
- ⁷⁵ INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
- ⁷⁶ Dipartimento di Fisica e Astronomia "Augusto Righi" – Alma Mater Studiorum Università di Bologna, Via Piero Gobetti 93/2, 40129 Bologna, Italy
- ⁷⁷ Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK
- ⁷⁸ Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
- ⁷⁹ CNRS-UCB International Research Laboratory, Centre Pierre Binétruy, IRL2007, CPB-IN2P3, Berkeley, USA
- ⁸⁰ University of Applied Sciences and Arts of Northwestern Switzerland, School of Engineering, 5210 Windisch, Switzerland
- ⁸¹ Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014 Paris, France
- ⁸² Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
- ⁸³ 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
- ⁸⁴ DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155, 2200 Copenhagen, Denmark
- ⁸⁵ Centre National d'Études Spatiales – Centre spatial de Toulouse, 18 Avenue Edouard Belin, 31401 Toulouse Cedex 9, France
- ⁸⁶ Institute of Space Science, Str. Atomistilor, Nr. 409 Măgurele, Ilfov 077125, Romania
- ⁸⁷ Consejo Superior de Investigaciones Científicas, Calle Serrano 117, 28006 Madrid, Spain
- ⁸⁸ Universidad de La Laguna, Departamento de Astrofísica, 38206 La Laguna, Tenerife, Spain
- ⁸⁹ Dipartimento di Fisica e Astronomia "G. Galilei", Università di Padova, Via Marzolo 8, 35131 Padova, Italy
- ⁹⁰ Institut für Theoretische Physik, University of Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany
- ⁹¹ Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
- ⁹² Université St Joseph; Faculty of Sciences, Beirut, Lebanon
- ⁹³ Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile
- ⁹⁴ Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
- ⁹⁵ Satlantis, University Science Park, Sede Bld 48940, Leioa-Bilbao, Spain
- ⁹⁶ Centre for Electronic Imaging, Open University, Walton Hall, Milton Keynes MK7 6AA, UK
- ⁹⁷ Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
- ⁹⁸ Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark
- ⁹⁹ Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, Plaza del Hospital 1, 30202 Cartagena, Spain
- ¹⁰⁰ Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- ¹⁰¹ INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
- ¹⁰² Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
- ¹⁰³ ICL, Junia, Université Catholique de Lille, LITL, 59000 Lille, France

Appendix A: Self-organising maps

The self-organising map can be used as a highly effective tool for calibrating redshift distributions. Its mechanism, initially proposed by Lima et al. (2008), is based on the assumption that galaxy samples with similar colour-space distributions resemble alike underlying redshift distributions. In this section, we introduce the working principle of the SOM.

A SOM consists of a fixed number of cells arranged on a grid of arbitrary, but commonly two, dimensions. At initialisation, the nodes, neurons, or cells of the (2D) map, which typically are arranged as a grid of hexagonal units, are appointed a (random) weight vector w of the same dimension as an input training vector x (Johnston et al. 2021). In the case of a galaxy survey for example, x could comprise five measurements of *ugriz* photometry (Geach 2011). This vector denotes the relative position of each node in the multi-dimensional parameter space occupied by the data. Training commences by presenting the map with a set of data points which the cells compete for. Distance metrics determine the winner neuron for any given training object (Masters et al. 2015). The distance metric $d_j(t)$ of the cell weight vector w to an input vector x , often referred to as Euclidean distance (Geach 2011), is defined as

$$d_j(t) := \sqrt{\sum_{k=1}^n [x_k(t) - w_{j,k}(t)]^2} \quad (\text{A.1})$$

at a given time step t , with j referring to the neuron in question, k being the data attribute, and n the total number of attributes for any input vector $x \in \mathbb{R}^n$. The neuron whose vector best resembles any given data point is henceforth referred to as best matching unit (BMU). A visual display in form of a sketch is given by Fig. A.1.

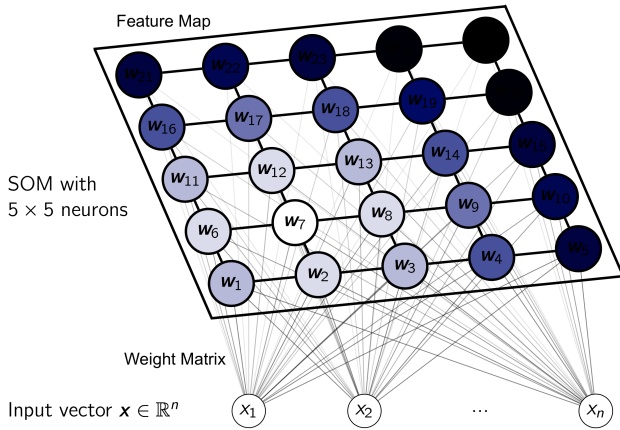


Fig. A.1. Architecture of a 5×5 neuron SOM with a Kohonen layer, depicted here as ‘Feature Map’. The input vector x is mapped onto the neurons, represented by their weight vectors w_j . The best matching unit (here w_7) and the surrounding weights are adjusted according to the current learning rate and neighbourhood function. The neurons in this example are coloured based on the magnitude of the adjustment in the current training step, where white are high magnitudes of adjustment and black are low magnitudes. Adapted from van den Busch (2021).

Because this metric treats all input dimensions equally, features with inherently larger uncertainties end up disproportionately influencing the results. To address this, the algorithm reduces the high-dimensional input space, regardless of its original dimensionality, into a single scalar value. This dimensionality reduction ensures that relative distances between neurons

meaningfully reflect the structure of the data, enabling an effective 2D mapping of the populated SOM (Johnston et al. 2021). Even so, the fundamental property which distinctly separates SOMs from other unsupervised learning techniques, is the fact that at each iteration, the weights of the BMU are adjusted, along with smaller updates applied to the neighboring neurons’ weights (Kind & Brunner 2013; Davidzon et al. 2019). By reweighing the importance of every neuron per data point, neurons of better representation receive greater updates than negligible neurons. This evaluation of the relative influence between neurons is expressed via the topological neighbourhood function

$$T_{a,B(x)}(e) = \exp\left(-\frac{S_{a,B(x)}^2}{2\sigma(t)^2}\right) \quad (\text{A.2})$$

which, depending on the epoch e , can be encoded as a normalised Gaussian kernel centred on the BMU (Masters et al. 2015; Campos et al. 2024). The parameter S denotes the lateral distance between cell a and the winning neuron $B(x)$,

$$S_{a,B(x)} = |w_a - w_{B(x)}|. \quad (\text{A.3})$$

The width of the Gaussian neighbourhood $\sigma(t)$ is fixed to

$$\sigma(t) = \sigma_{\text{init}} \left(\frac{1}{\sigma_{\text{init}}}\right)^{\frac{t}{N}}, \quad (\text{A.4})$$

where N refers to the number of training iterations. Initially, σ_{init} is chosen large enough to cover most of the map. The width itself shrinks during training so that by the end of this process, only the BMU and directly adjacent cells are (significantly) affected by new data (Masters et al. 2015). After identifying the BMU per iteration, the weight vector is updated by equations

$$\Delta w_a(t) = \eta(e) T_{a,B(x)}(e) [x(t) - w_a(t)] \quad (\text{A.5})$$

and

$$w_a(t+1) = w_a(t) + \Delta w_a(t), \quad (\text{A.6})$$

with a resembling a neuron, η the monotonically time decreasing learning rate, and $B(x)$ the winning neuron. As such, the SOM becomes progressively less responsive to new training data per iteration. The topological neighbourhood T is then a normalised measure for every neurons weight adaptation. This becomes clearer if one assumes two neurons, a and b , to both be the winner neuron. In this case, the lateral distance is $S = 0$ and therefore $T(e) = 1$. After training, the map can be used to categorise previously unseen data, by appointing the closest weight vector to a given data point, based on the learned representation (Davidzon et al. 2019).

Appendix B: Covariate shift

One of the most frequent sources of error encountered in statistical analysis with ML is caused by covariate shift (Luo et al. 2022). This shift refers to the difference in feature space between the calibration and validation samples. In redshift calibration, this difference arises from the fact that the methods with which the samples are selected are fundamentally different (Beck et al. 2017). It is important to note that the definition of the redshift tomography and subsequent selection of the samples introduce additional bias. Using Flagship2 mock data (see Sect. 3.1), the upper panel in Fig. B.1 features the true $n(z)$ of a wide-like validation sample represented by the orange histogram. In contrast,

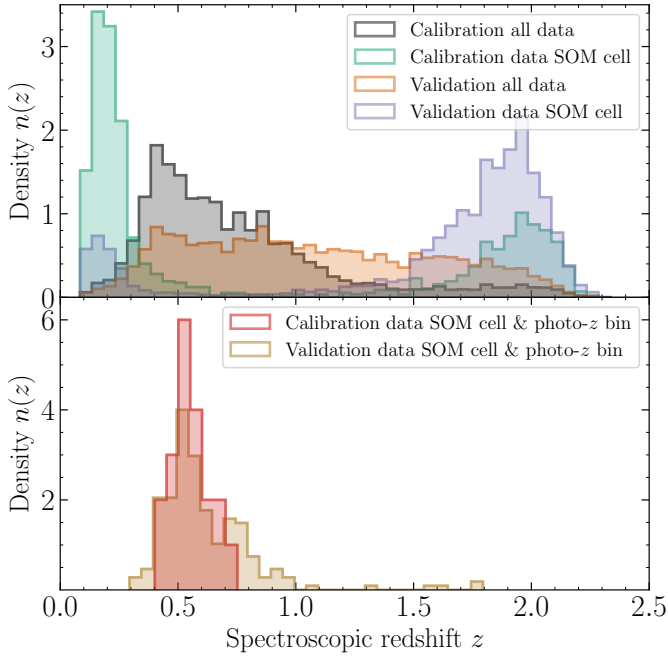


Fig. B.1. *Top:* Mean spec- z distributions for the validation and calibration sample subject to a non-trivial selection function. The respective distributions are also depicted for a single SOM cell chosen at random. *Bottom:* Display of the distributions found for the same randomly selected cell after applying photo- z binning.

the grey histogram depicts the $n(z)$ of a spectroscopic calibration sample, taking into account a non-trivial selection function, specifically cuts made in colour-colour space. The purple and green histograms in the upper panel respectively portray the $n(z)$ of galaxies of a random cell that belong to the validation sample and to the calibration sample, respectively. Clearly, the green distribution in Fig. B.1 is inadequate for calibrating the purple distribution attributed to the selected cell. This highlights the need for caution when considering the mean spec- z s of SOM cells as metric to assign galaxies to their respective tomographic redshift bins, given that redshift bias is driven by the discrepancy between the green and purple histograms. However, if an identical photo- z binning (here $0.5 \leq z_p \leq 0.7$) is applied to both samples, as done for the brown and red distributions in the lower panel of Fig. B.1, only representative sources in redshift are used for calibration purposes. In doing so, the discrepancy in distributions is drastically reduced. Applying a one-sided (spectroscopic-only) cut would change the feature space of the training sample (Autenrieth et al. 2021). Because the spectroscopic $n(z)$ aligns with the colour-based structure of the SOM grid, cells with low mean photo- z s may inadvertently be used to calibrate high-redshift galaxies if photo- z binning is not enforced. Instead, when both cell distributions are divided into photo- z bins first, every cell can, in principle, be used for calibration.

Appendix C: Improved realism

While the majority of bins listed in Table 4 meet the requirement set by *Euclid*, their biases are somewhat optimistic. The calibration sample used to train the SOM undergoes a less complex selection compared to the selection boundaries faced by real observations, as addressed by Masters et al. (2015, 2017, 2019). This pertains to the calibration sample generation process, where

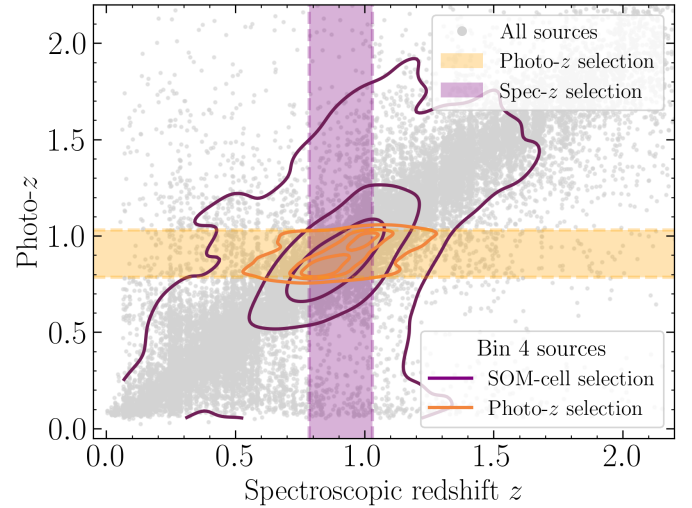


Fig. C.1. Distribution of photo- z over spec- z , where the full sample is shown in grey, and the shaded areas represent the fourth tomographic bin listed in Table 2 for photometric and spectroscopic redshift (orange and purple bands, respectively). Overlaid are 1, 2, and 3σ completeness contours for the two selection strategies: photo- z binning within the SOM (orange) and mean spec- z binning per SOM cell (purple).

sources are selected from a SOM trained on noiseless and therefore unrealistic photometry, resulting in a more uniform distribution in colour-colour space. Consequently, the current calibration sample used for training includes distribution counts for certain areas in the colour-colour space with limited or no reliably observed sources. Hence, it would be imprudent to assume that *Euclid* will achieve the same level of accuracy as indicated by these predictions. It is therefore important to enhance the realism of the calibration data to a point where the predictions made within this work become reliably applicable to *Euclid*. Given that C3R2 (Masters et al. 2017) currently stands as the most extensive spectroscopic sample used for calibration efforts, we test our pipeline on various varied calibration samples by changing the sampling process itself in order to compare subsequent distributions to those found with C3R2. To achieve this, we utilise SOM statistics extracted from Masters et al. (2019), including cell ID, median spec- z , associated colours and the count of reliable spectra where the quality flag (Q) is set to $Q = 4$ in each cell. It is worth noting that this initial approach is somewhat conservative, as it does not yet incorporate spectra from other surveys beyond C3R2, which could contribute to the calibration, nor does it include less reliable spectra currently classified as $Q = 3$. By leveraging the available information, we can simulate a set of more realistic spectroscopic calibration samples. In the pursuit of robust calibration, we propose four distinct approaches, based in part on the distributions found in C3R2, to alternate samples for calibration purposes.

C.1. Pathological spec- z calibration sample & mean spec- z SOM cell tomography

The first among these samples is what we colloquially refer to as the pathological calibration sample. This sample intentionally comprises catastrophic instances that deviate markedly from the characteristics observed in both the established Masters SOM calibration sample and our prior calibration sample (see Sect. 4.3). Notably, it introduces challenges by incorporating a substantial redshift desert where no data points exist. By

examining the calibration performance on this extreme sample, we aim to assess the SOMs robustness and its ability to handle unconventional data distributions. The redshift distributions are depicted in Fig. C.2. This new calibration sample is then used to train a noisy SOM with bias calculations based on tomography defined on the mean spec- z of SOM cells as introduced in Sect. 4.3. As with Fig. 6, we observe that none of the bins satisfy the *Euclid* requirement (see Fig. C.3 and Table C.1). Notably, we observe larger mean biases accompanied by even larger standard deviations with distributions at times characterised by extreme outliers.

C.2. Masters-like photo- z calibration sample & mean spec- z SOM cell tomography

Compared to Appendix C.1, we adjust the $n(z)$ of our calibration sample to mimic the mean photo- z distribution of the Masters calibration sample. This requires binning the redshift range $[0.07 \leq z_{\text{obs}} \leq 1.41]$ spanned by the sample into 245 equidistant slices, referred to as redshift blocks. This ensures that each block contains ≥ 1 cell, aiming for the highest level of fidelity. Reshaping the calibration sample $n(z)$ by updating the way in which we sample the trained noiseless SOM, involves examining the ratio of sources found for each redshift block,

$$f_{\text{rb}} = \frac{\text{Number of galaxies noiseless SOM}}{\text{Number of galaxies Masters SOM}}, \quad (\text{C.1})$$

for both, the Masters and noiseless SOM. Based on f_{rb} , we proceed to resample the noiseless SOM per redshift block. In the cases where $f_{\text{rb}} < 1$, up to four adjacent blocks are included, increasing the number of spectra considered for sampling, provided their combined respective fraction is $f_{\text{rb}} \geq 1$. Finally, any duplicates are removed from the modified distribution of the calibration sample. The new training sample (see Fig. C.2) is used to train a noisy SOM with bias calculation based on tomography following Sect. 4.3. The bias results are shown in Fig. C.4 and Table C.2. We drastically reduce the biases per tomographic redshift bin compared to Appendix C.1, showcasing the importance of both a representative training sample in ML applications and spectroscopic coverage in the commonly less accessible colour-colour space.

C.3. Masters-like photo- z calibration sample & photo- z tomography

The calibration sample remains the same as defined in Appendix C.2. However, in this instance, the calibration sample (consistent with that of Appendix C.2) is used to train a noisy SOM with bias calculation determined by tomography defined on photo- z rather than mean spec- z of SOM cells, as elucidated in Sect. 4.4. The obtained biases are presented in Fig. C.5 and Table C.3. Considering QC, we achieve sufficient biases in six bins, with results comparable to those obtained in Table 4. Contrary to before, however, we observe smaller fluctuations, potentially due to the calibration sample distribution. We attribute the notable increase in performance compared to Appendix C.2 to the implementation of photo- z tomography, which in all scenarios presented delivers much reduced biases compared to the tomography most commonly defined on the SOM cell mean spec- z .

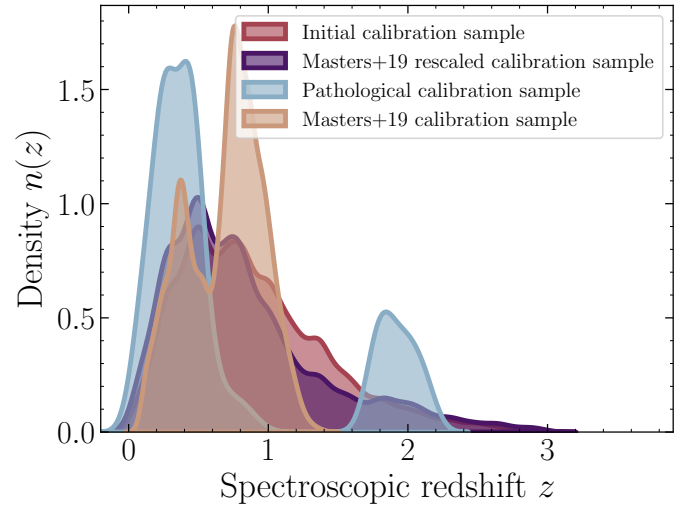


Fig. C.2. Display of the $n(z)$ distributions of the calibration sample introduced in Sect. 4.3 (red), the Masters SOM cell mean spec- z (brown) as well as Appendix C.2 (blue) and Appendix C.3 (purple) calibration samples.

C.4. Colour-colour sampling & photo- z tomography

While redshift biases achieved in Appendix C.3 outperform respective biases found in Appendix C.1 and Appendix C.2, the selection and approach of simulating a “very realistic” spec- z calibration sample is based exclusively on the Masters SOM statistics. Since SOM cells can provide the mean or median of the underlying data distributions they represent, deviations and rounding are introduced, which prevent the new calibration sample from being a perfect representation of the Masters SOM distribution. In other words, the available Masters data provides feature-averaged statistics only. An improved method would therefore require the entire calibration sample catalogue, containing all galaxy attributes, used to train the Masters SOM in order to use multi-dimensional colour-colour space information. However, this poses a challenging task, as it entails aligning two multi-dimensional colour-colour spaces between SOMs and subsequently translating the colour-redshift relation.

C.5. Future optimisation

A strategy to mitigate data loss due to QC involves the clustering of cells. This method avoids the exclusion of EWS-like data located in cells with no spectroscopic calibration data available, by merging them with cells carrying contributions from both samples in nearby colour-colour space (van den Busch et al. 2022). As demonstrated in the Appendix of Wright et al. (2020b), biases per bin can be computed as a function of clusters (N_{clust}), selecting N_{clust} to maximise data completeness while minimising the bias. One promising approach involves reconstructing the true $n(z)$ distribution for each SOM cell. This two-step process involves initially distributing sources based on the maps topology, followed by refining each cell by uncovering the true $n(z)$ of its sources. Lastly, it seems that the relationship between biases may contain additional information that could be valuable in constraining and mitigating spectroscopic bias. Finally, it may be useful to examine redshifted SEDs of individual low- Q galaxies, as they impose an extra selection on the true $n(z)$. Removing or correcting them via SED-based training selection could potentially improve bias estimates.

Table C.1. Biases in the mean redshift estimation per bin according to Appendix C.1.

Bias ($\Delta\langle z \rangle$)	bin 1	bin 2	bin 3	bin 4	bin 5
QC no	0.219 ± 0.093	0.182 ± 0.069	0.173 ± 0.043	0.211 ± 0.205	0.171 ± 0.177
QC yes	0.083 ± 0.014	0.103 ± 0.033	0.152 ± 0.028	0.137 ± 0.076	0.113 ± 0.142
Bias ($\Delta\langle z \rangle$)	bin 6	bin 7	bin 8	bin 9	
QC no	0.006 ± 0.223	-0.139 ± 0.163	-0.308 ± 0.199	-0.275 ± 0.199	
QC yes	-0.025 ± 0.176	-0.097 ± 0.160	-0.205 ± 0.153	-0.096 ± 0.141	

Notes. The values shown are the mean biases over 60 different lines of sight as well as standard deviations for the lowest nine bins. Under the recognition of QC, none of the bins lie within the *Euclid* bias requirement.

Table C.2. Biases in the mean redshift estimation per bin according to Appendix C.2.

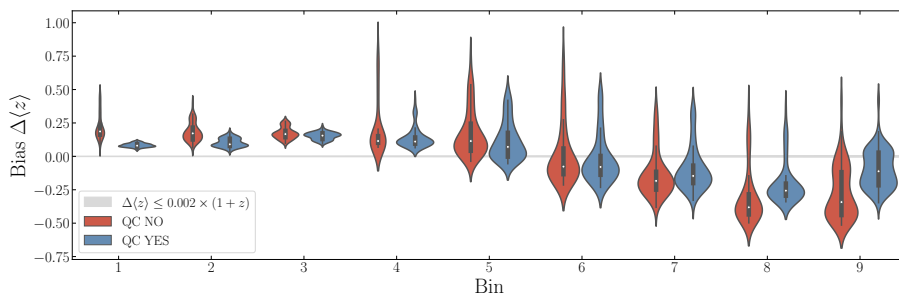
Bias ($\Delta\langle z \rangle$)	bin 1	bin 2	bin 3	bin 4	bin 5
QC no	0.077 ± 0.023	0.035 ± 0.028	0.033 ± 0.028	0.006 ± 0.017	0.003 ± 0.014
QC yes	0.059 ± 0.012	0.030 ± 0.025	0.031 ± 0.028	0.006 ± 0.016	0.007 ± 0.018
Bias ($\Delta\langle z \rangle$)	bin 6	bin 7	bin 8	bin 9	
QC no	0.009 ± 0.018	-0.034 ± 0.060	-0.044 ± 0.048	-0.063 ± 0.024	
QC yes	0.020 ± 0.014	-0.009 ± 0.035	-0.018 ± 0.032	-0.033 ± 0.021	

Notes. The values shown are the mean biases over 100 different lines of sight as well as standard deviations for the lowest nine bins. Under the recognition of QC, bins 2, 3, and 5 lie within the *Euclid* bias requirement.

Table C.3. Biases in the mean redshift estimation per bin according to Appendix C.3.

Bias ($\Delta\langle z \rangle$)	bin 1	bin 2	bin 3	bin 4	bin 5
QC no	0.003 ± 0.012	0.003 ± 0.011	0 ± 0.008	0.001 ± 0.014	0.001 ± 0.011
QC yes	0.017 ± 0.008	0.007 ± 0.008	0 ± 0.007	0.002 ± 0.014	0 ± 0.010
Bias ($\Delta\langle z \rangle$)	bin 6	bin 7	bin 8	bin 9	
QC no	0 ± 0.013	0.009 ± 0.038	-0.007 ± 0.054	-0.007 ± 0.067	0.008 ± 0.015
QC yes	0.001 ± 0.013	0.009 ± 0.036	0.012 ± 0.053	-0.006 ± 0.007	0.009 ± 0.015

Notes. The values shown are the mean biases over 50 different lines of sight as well as standard deviations for all ten bins. Under the recognition of QC, six bins lie within the *Euclid* bias requirement, while bins 1, 8, and 10 fail to satisfy the requirement by errors on the order of ~ 0.001 only, showcasing the improvement achievable with photo- z based tomography.

**Fig. C.3.** Violin plot of biases per bin according to Appendix C.1.

Appendix D: Calibration coverage per tomography

Depending on the chosen tomography, the fractional use of the spectroscopic sample for calibration varies significantly. As displayed in Fig. D.1, photo- z based tomography consistently displays fractional coverage above $\sim 98\%$ per bin, as its selection is no longer linked to the mean spec- z of SOM cells. This cannot be said for SOM-based tomography, where underrepresented bins at low and high redshift face up to $\sim 25\%$ loss of calibration galaxies.

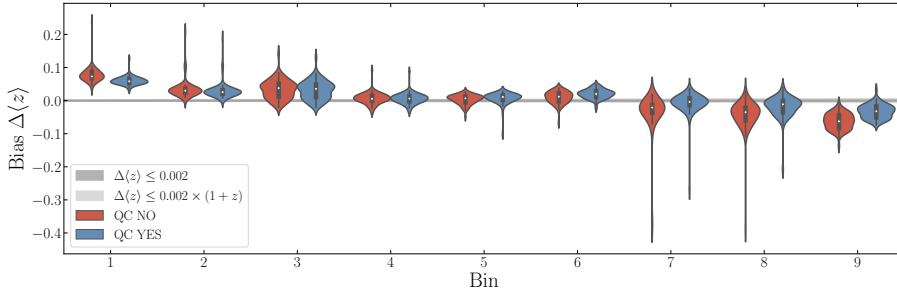


Fig. C.4. Violin plot of biases per bin according to Appendix C.2.

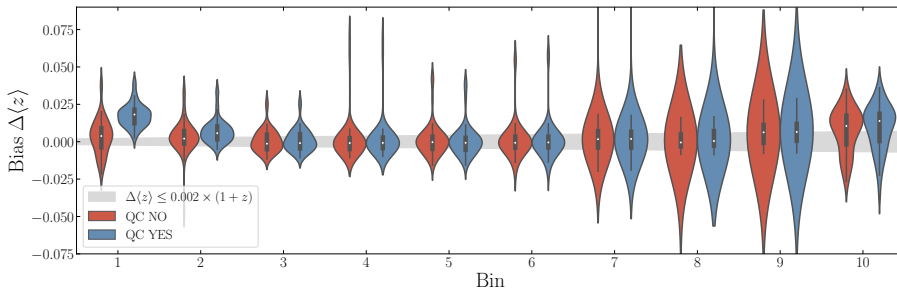
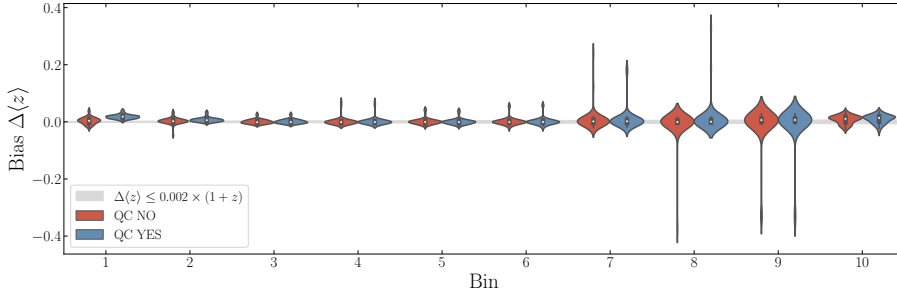


Fig. C.5. Violin plot of biases per bin according to Appendix C.3. *Top:* Includes the entire distributions. *Bottom:* *y*-axis limited version, effectively removing 2% of extreme outliers from the sample to uncover the mean biases with respect to the requirement.

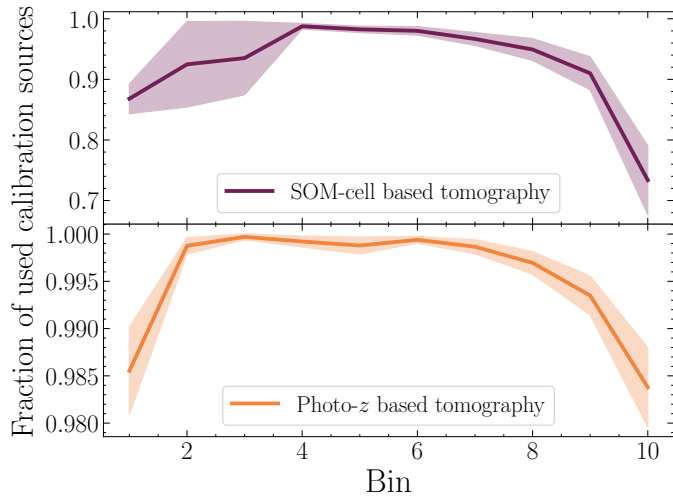


Fig. D.1. Visualisation of the effective fraction of calibration data considered after applying QC and gold sample selection for the tomography defined on the mean spec-*z* of SOM cells (purple) and photo-*z* (orange).