

# KiDS-1000: cosmic shear with enhanced redshift calibration

J. L. van den Busch<sup>1,2</sup>, A. H. Wright<sup>1</sup>, H. Hildebrandt<sup>1</sup>, M. Bilicki<sup>3</sup>, M. Asgari<sup>4,5</sup>, S. Joudaki<sup>6,7</sup>, C. Blake<sup>8</sup>,  
C. Heymans<sup>4,1</sup>, A. Kannawadi<sup>9</sup>, H. Y. Shan<sup>10,11</sup>, and T. Tröster<sup>4</sup>

<sup>1</sup> Ruhr-University Bochum, Astronomical Institute, German Centre for Cosmological Lensing, Universitätsstr. 150, 44801, Bochum, Germany, e-mail: jlvdb@astro.rub.de

<sup>2</sup> Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

<sup>3</sup> Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668, Warsaw, Poland

<sup>4</sup> Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, U.K.

<sup>5</sup> E. A. Milne Centre, University of Hull, Cottingham Road, Hull, HU6 7RX, UK

<sup>6</sup> Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

<sup>7</sup> Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

<sup>8</sup> Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia

<sup>9</sup> Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA

<sup>10</sup> Shanghai Astronomical Observatory (SHAO), Nandan Road 80, Shanghai 200030, China

<sup>11</sup> University of Chinese Academy of Sciences, Beijing 100049, China

Received 25 August 2021 / Accepted 4 April 2022

## ABSTRACT

We present a cosmic shear analysis with an improved redshift calibration for the fourth data release of the Kilo-Degree Survey (KiDS-1000) using self-organising maps (SOMs). Compared to the previous analysis of the KiDS-1000 data, we expand the redshift calibration sample to more than twice its size, now consisting of data of 17 spectroscopic redshift campaigns, and significantly extending the fraction of KiDS galaxies we are able to calibrate with our SOM redshift methodology. We then enhance the calibration sample with precision photometric redshifts from COSMOS2015 and the Physics of the Accelerated Universe Survey (PAUS), allowing us to fill gaps in the spectroscopic coverage of the KiDS data. Finally we perform a Complete Orthogonal Sets of E/B-Integrals (COSEBIs) cosmic shear analysis of the newly calibrated KiDS sample. We find  $S_8 = 0.748^{+0.021}_{-0.025}$ , which is in good agreement with previous KiDS studies and increases the tension with measurements of the cosmic microwave background to  $3.4\sigma$ . We repeat the redshift calibration with different subsets of the full calibration sample and obtain, in all cases, agreement within at most  $0.5\sigma$  in  $S_8$  compared to our fiducial analysis. Including additional photometric redshifts allows us to calibrate an additional 6% of the source galaxy sample. Even though further systematic testing with simulated data is necessary to quantify the impact of redshift outliers, precision photometric redshifts can be beneficial at high redshifts and to mitigate selection effects commonly found in spectroscopically selected calibration samples.

**Key words.** cosmology: observations – gravitational lensing: weak – galaxies: distances and redshifts – surveys

## 1. Introduction

Over the past decade, gravitational lensing (Bartelmann & Schneider 2001) has emerged as one of the most powerful tools to study gravity and the dark sectors of the Universe, dark matter and dark energy, through the impact of these components on the density fluctuations of matter and their evolution with cosmic time (Peacock et al. 2006). In the limit of weak lensing, massive structures along the line of sight imprint a subtle shearing on the shapes of distant galaxies. This signal can be extracted by statistically analysing the ellipticity of galaxy images in large surveys (Refregier 2003). These cosmic shear surveys (Kilbinger 2015) face the challenge that they must accurately reconstruct the galaxy redshift distribution in order to interpret the cosmological signal correctly. Even small biases in the first moment of the redshift distribution may introduce significant biases in the recovered cosmological parameters (e.g. Huterer et al. 2006; Ma et al. 2006). There exists tension between constraints from cosmic shear and the cosmic microwave background, first seen between the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS, Heymans et al. 2013; MacCrann et al. 2015; Joudaki et al. 2017) and Planck (Planck Col-

laboration et al. 2014), but for example also between the Kilo-Degree Survey (KiDS, Kuijken et al. 2015) and Planck legacy (Planck Collaboration et al. 2020), recently for KiDS-1000 (Asgari et al. 2021). The most recent cosmic shear results from the Dark Energy Survey (DES, Flaugher et al. 2015) are very similar to KiDS-1000 (Amon et al. 2022; Secco et al. 2022) albeit at lower statistical tension with Planck<sup>1</sup>. In the light of these repeatedly reported tensions, redshift calibration has come under scrutiny as one of the systematics for cosmic shear experiments (e.g. Joudaki et al. 2020).

Due to the statistical nature of the shear measurements, current generation (stage-III) cosmic shear surveys, such as KiDS, DES, and the Hyper Suprime-Cam Subaru Strategic Program (HSC, Aihara et al. 2018), rely on imaging of tens of millions of galaxies for which spectroscopic redshifts cannot be measured directly. Instead, galaxy redshifts are determined with secondary redshift estimates, the most notable ones are direct calibration with spectroscopic training samples (e.g. Lima et al.

<sup>1</sup> Conclusions over the degree of tension differ primarily owing to the way tension is quantified and different prior choices for the neutrino mass in the Planck re-analysis.

2008; Hildebrandt et al. 2017, 2020; Buchs et al. 2019; Wright et al. 2020a), clustering redshifts (which infer redshift distributions by exploiting the gravitational clustering of galaxies at similar redshifts, e.g. Newman 2008; Matthews & Newman 2010; Schmidt et al. 2013; Ménard et al. 2013; van den Busch et al. 2020; Hildebrandt et al. 2021; Gatti et al. 2022) and methods that make use of a combination of both these approaches (Sánchez & Bernstein 2019; Alarcon et al. 2020; Myles et al. 2021).

The redshift calibration of the fourth data-release of KiDS (Kuijken et al. 2019; Hildebrandt et al. 2021) relies on an implementation of the direct calibration that utilises a self-organising map (SOM, Kohonen 1982; Wright et al. 2020a) based on work by Masters et al. (2016). The fundamental principle of this method is to re-weight a spectroscopic reference sample such that it is representative of a photometric dataset with an unknown redshift distribution. The weighted redshift distribution of the reference sample is then a direct estimate of the unknown distribution. Additionally, the SOM method allows for the removal of galaxies from the photometric dataset for which no similar galaxies exist in the reference sample. Their inclusion would otherwise bias the estimated redshift distribution. We call the subset of the remaining, well represented galaxies the ‘gold sample’. In this work we explore the redshift calibration of the KiDS data with a significantly enhanced reference sample that is composed of a variety of spectroscopic redshift campaigns and precision photometric redshifts. This allows us to expand the KiDS gold sample and calibrate redshifts in regions of the colour space that are difficult to access by direct spectroscopy. We then study how these additional calibration data influence our ability to calibrate the redshifts of the source sample and how selection effects and changes in the calibration propagate to cosmological constraints. We compare our results to the original KiDS-1000 cosmic shear analysis by Asgari et al. (2021).

This paper is structured as follows: in Sect. 2 we describe the KiDS data and the redshift calibration sample (further details on this compilation in App. A), and in Sect. 3 and 4 we present the SOM redshift calibration and our cosmic shear analysis methods. We present and discuss the newly calibrated gold samples and cosmological constraints in Sect. 5 and Sect. 6. Finally we conclude and summarise in Sect. 7.

## 2. Data

This paper explores redshift calibration of the KiDS cosmic shear weak lensing sample (Sect. 2.1) with increasingly deep redshift calibration catalogues. Our fiducial analysis relies exclusively on spectroscopic data which is compiled from a variety of different spectroscopic surveys (Sect. 2.2 and App. A). We then add less accurate redshift estimates derived from narrow-band photometry from the Physics of the Accelerated Universe Survey (PAUS, Padilla et al. 2019) and finally from medium- and broadband photometry from COSMOS2015 (Sect. 2.3). This approach is similar to the construction of the calibration sample for the DES Y3 redshift calibration (Myles et al. 2021), however the KiDS lensing and redshift calibration samples are both covered by the same nine bands which avoids mapping different photometries via a transfer function.

### 2.1. KiDS-1000 photometric data

The Kilo-Degree Survey (KiDS, Kuijken et al. 2015; de Jong et al. 2015, 2017; Kuijken et al. 2019) is a public European Southern Observatory (ESO) survey that has been designed particularly with weak gravitational lensing applications in mind.

The complete survey will deliver about  $1350 \text{ deg}^2$  of *ugri* imaging split into an equatorial and a southern field. Combined with *ZYJHK<sub>s</sub>* imaging from its companion infra-red survey, the VISTA Kilo-Degree Infrared Galaxy Survey (VIKING Edge et al. 2013; Venemans et al. 2015), this constitutes a nine-band, matched-depth data-set with primary imaging in the *r*-band, observed at a mean seeing of  $0.7''$ . This work is based on the fourth data release of KiDS which covers  $1006 \text{ deg}^2$ . The weak lensing source catalogue (KiDS-1000, Giblin et al. 2021) is divided into five tomographic redshift bins, based on nine-band photometric redshifts (four bins with  $\Delta Z_B = 0.2$ , starting from  $Z_B = 0.1$  and a fifth bin at  $0.9 < Z_B \leq 1.2$ ) computed with BPZ (Bayesian Photometric Redshift, Benítez 2000). It contains all objects with non-zero shear measurement weights obtained from *lensfit* (Miller et al. 2007; Fenech Conti et al. 2017), which effectively selects objects with *r*-band magnitudes between  $20 \leq r \leq 25$ .

In addition to the main survey imaging there are observations of six fields dedicated primarily to redshift calibration. These ‘KiDZ’ fields cover approximately  $1 \text{ deg}^2$  each and target areas of the sky also observed by different spectroscopic campaigns, which are summarised in Sect. 2.2. Just like the main survey, KiDZ provides KiDS+VIKING 9-band imaging<sup>2</sup> which reaches or exceeds the depth of the main survey. In the latter case we homogenise the data depth by applying Gaussian noise to obtain matched photometry.

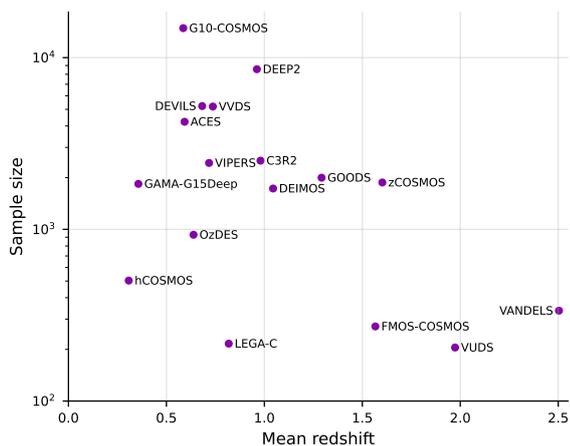
### 2.2. Spectroscopic data for calibration

The most important spectroscopic campaigns that overlap with the six KiDZ fields are zCOSMOS (Lilly et al. 2009), VVDS (VIMOS VLT Deep Survey, Le Fèvre et al. 2005, 2013), DEEP2 (Newman et al. 2013), the GAMA (Galaxy And Mass Assembly, Driver et al. 2011) deep field G15deep (Kafle et al. 2018; Driver et al. 2022), and a compilation of spectra covering the Chandra Deep Field South (CDF-S, App. A). This data has been used in previous KiDS redshift calibration works (Wright et al. 2020a; Hildebrandt et al. 2021).

We extend this compilation by adding data from C3R2 (Complete Calibration of the Colour-Redshift Relation, Masters et al. 2017, 2019; Euclid Collaboration et al. 2020; Stanford et al. 2021), DEVILS (Deep Extragalactic Visible Legacy Survey, Davies et al. 2018), VIPERS (VIMOS<sup>3</sup> Public Extragalactic Redshift Survey, Scodreggio et al. 2018) and a variety of spectroscopic campaigns that target the CDF-S and COSMOS fields which are detailed in App. A. We also revise the selection of sources included for calibration by removing duplicates, both from spatial overlap as well as within the datasets, and by homogenising redshift quality flags based on the original information in the input samples. If, for a given source, there are redshifts from different surveys available, we assign the most reliable measurement based on a specific ‘hierarchy’ of surveys (see App. A for details). For objects with multiple spectroscopic measurements within a particular survey, we either take the redshift with the highest quality flag or, if various entries for the same source have the same quality flag and the reported redshifts differ by no more than 0.005, we take the average. However, if the reported redshift differences exceed this threshold, we ex-

<sup>2</sup> In case of the COSMOS field we instead use existing data in the CFHT (Canada France Hawaii Telescope) *z*-band which is sufficiently similar to the VISTA InfraRed CAMera (VIRCAM) *Z*-band at the Visible and Infrared Survey Telescope for Astronomy (VISTA) given the photometric and redshift calibration uncertainties.

<sup>3</sup> Visible Multi-Object Spectrograph



**Fig. 1.** Sample size and mean redshift of the different surveys that are part of our spectroscopic compilation after matching to their counterparts in KiDS imaging. Objects with redshifts from multiple sources are assigned to the survey with the most reliable redshift estimate.

clude such a source from the compilation. We restrict the selection to objects with high quality spectroscopic redshifts (approximately corresponding to  $\geq 95\%$  confidence or redshift quality code  $nQ \geq 3$ ). Fig. 1 compares the number of galaxies and their mean redshift for all samples that enter the spectroscopic compilation. These values apply after removing duplicates between overlapping catalogues and only for those objects with photometric coverage in KiDZ (Sect. 2.1).

### 2.3. Photometric data for calibration

The success rate of spectroscopically determined redshifts is very different from (typically flux-limited) imaging data. Therefore, it is very difficult to obtain a spectroscopic calibration sample that is representative of photometric data in magnitude-/colour-space, especially at faint magnitudes. Instead, we additionally include galaxy samples with high quality photometric redshifts to achieve a greater overall coverage of the KiDS data by the calibration sample which is beneficial for our redshift calibration technique of choice (Sect. 3.1).

#### 2.3.1. COSMOS2015

The COSMOS2015 catalogue (Laigle et al. 2016) constitutes a sample of about half a million galaxies in the COSMOS field with precision photometric redshifts<sup>4</sup> derived from up to 30 photometric bands, ranging from near ultra-violet to mid infrared, including 14 medium and narrow band filters. This sample extends to higher redshifts ( $z_{\max} \approx 6$ ) and fainter magnitudes than our spectroscopic compilation, but at the cost of less secure redshift estimates with an outlier fraction ranging from 0.5% at low redshifts to 13.2% for  $3 < z < 6$  (Laigle et al. 2016).

#### 2.3.2. PAUS

The PAUS photometric redshift sample<sup>5</sup> (Alarcon et al. 2021) is a combination of 26 optical and near-infrared bands from COSMOS2015 that are matched against observations of the COS-

MOS field in 40 narrow band filters by the PAU survey. These PAU filters sample the optical regime between 450 nm to 850 nm at a bandwidth of  $\Delta\lambda = 12.5$  nm (Padilla et al. 2019) and the combined photometric catalogue is limited to  $i_{AB} < 23$ . Due to the relatively high spectral resolution of the dataset a new Bayesian spectral energy distribution (SED) fitting technique is required that accounts for individual emission lines (Alarcon et al. 2021). This allows the PAUS photo- $z$  to achieve a  $3\times$  ( $1.7\times$ ) lower photo- $z$  scatter at the bright (faint) end of the magnitude distribution and marginally smaller outlier fractions compared to the original COSMOS2015 photo- $z$  at  $i_{AB} < 23$ . The scaled photo- $z$  bias is very low and has a  $|\text{median}(\Delta_z)| < 0.001$  over the whole redshift range of the PAUS sample. Therefore this sample positions itself right between the spectroscopic data and COSMOS2015 in terms of completeness and redshift precision.

### 2.4. Combined calibration sample

In this work we select data from the full COSMOS2015 photometric catalogue and combine this data hierarchically with PAUS and the spectroscopic data. Finally we match this unified catalogue to the KiDZ imaging to form our redshift calibration sample.

We prepare the full COSMOS2015 photometric catalogue similar to Laigle et al. (2016), that is, we select only those sources which fall into the intersection of the footprint of the COSMOS field ( $f_{\text{flag\_cosmos}} = 1$ ) and the UltraVISTA observations ( $f_{\text{flag\_hjmcc}} = 0$ ), which provides essential IR spectral coverage. We exclude data from saturated areas ( $f_{\text{flag\_peter}} = 0$ ) and additionally remove objects that are classified to be most likely stars ( $\text{type} \neq 1$ ) or have no photo- $z$  estimate. This selection yields about half a million objects.

About 40,000 sources of the PAUS sample are, by design, matched against COSMOS2015 and therefore require no further preparation. Therefore we can directly combine the spectroscopic compilation, PAUS, and the subset of COSMOS2015 by matching objects within  $1''$ . We maintain a hierarchy to ensure that we always choose the most reliable redshift estimate available: spec- $z$  supersede PAUS photo- $z$  which supersede the COSMOS2015 photo- $z$ . Finally, we assign 9-band KiDS magnitudes to this compilation by matching against the KiDZ data, again within  $1''$ . This combination of spec- $z$  (in all KiDZ fields), photo- $z$  (only in COSMOS) and KiDS imaging represents our full redshift calibration sample.

The method to combine the two photometric redshift samples with our spectroscopic compilation in an hierarchical manner is very similar to the approach taken for the redshift calibration of the DES Y3 data (Myles et al. 2021). There are, however, two key differences to their approach. First, our compilation of spectroscopic redshifts covers a much wider range of the colour-redshift space than their selection of spectra. This allows us to construct more representative calibration samples that consist purely of spectroscopic redshifts, photometric redshifts, or a combination thereof. Second, the primary KiDS imaging data and the calibration data are observed in all nine photometric bands at a comparable depth, which simplifies the mapping from galaxy colour to redshift significantly (see Sect. 3).

#### 2.4.1. Primary compilations

From this heterogeneous sample with redshift estimates from very different sources we select three subsets, each with a higher redshift precision but lower completeness:

<sup>4</sup> We use the median of the photo- $z$  likelihood distribution (PHOTOZ column in the catalogue).

<sup>5</sup> available at [cosmohub.pic.es](https://cosmohub.pic.es)

**Table 1.** Number counts and mean redshifts of the original KiDS-1000 redshift calibration sample and different subsets of the new redshift compilation.

Compilation	Count	$\langle z \rangle$
Hildebrandt et al. (2021)	25 373	0.796
spec-z fiducial	52 911	0.788
spec-z+PAUS	61 163	0.776
spec-z+PAUS+COS15	112 400	1.002
spec-z $nQ \geq 4$	24 117	0.832
only PAUS	24 229	0.640
only COS15	80 632	1.081
only PAUS+COS15	80 635	1.084

**Notes.** The full compilation is represented as *spec-z+PAUS+COS15*. Values apply after matching with the KiDZ imaging and removing duplicates in the spectroscopic data.

- the full compilation (to which we refer as *spec-z+PAUS+COS15*),
- objects with either spectroscopic redshifts and/or PAUS photo- $z$  (*spec-z+PAUS*), and
- our fiducial sample containing only those objects that have spectroscopic redshifts (*spec-z fiducial*).

The main properties and redshift distributions of these three primary compilations are summarised in Table 1 and Fig. 2.

The fiducial sample is already about twice as large as the calibration sample used previously by Hildebrandt et al. (2021) to calibrate the KiDS-1000 redshifts. Of the additional spectra we consider DEVILS and C3R2, the latter of which is designed to target regions of the galaxy colour-space with currently little spectroscopic coverage, to be the most important contributions. Similarly *spec-z fiducial* already contains about 66% of the matched PAUS sources, of which the majority has redshift  $z < 1$ . Due to its limited depth, the *spec-z+PAUS* sample presents only a small improvement over the fiducial case. The COSMOS2015 data, on the contrary, nearly doubles the compilation to its final size of about 112,000 objects. Due to the significantly higher depth of the COSMOS2015 photo- $z$ , the fraction of sources with  $z > 1$  nearly triples, pushing the mean redshift to  $\langle z \rangle \approx 1.0$ . Nevertheless this comes at the cost of a lower redshift accuracy compared to the rest of the sample.

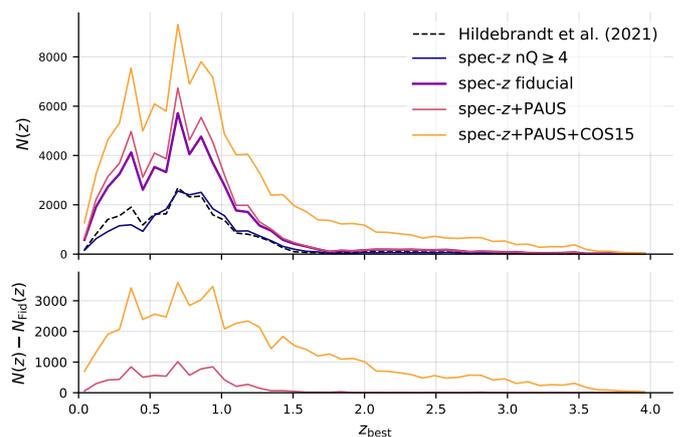
#### 2.4.2. Secondary compilations

In addition to the three primary compilations we also consider a subset that is restricted to only the most secure spectroscopic redshifts. This *spec-z  $nQ \geq 4$*  sample is, due to the large fraction of shared spectra, closest to the one of Hildebrandt et al. (2021) except that it lacks some low redshift sources (Fig. 2)

Finally, we create three subsets of the full redshift compilation that rely purely on photometric redshift estimates. We achieve this by recompiling the redshift compilation according to Sect. 2.4 but omitting all spectroscopic redshifts, therefore maintaining the usual hierarchy of PAUS and COSMOS2015 photo- $z$ . These are

- only objects from the PAU survey (*only-PAUS*),
- all objects with photo- $z$  (*only-PAUS+COS15*), and
- the pure COSMOS2015 subset (*only-COS15*, also discarding PAUS photo- $z$  from the stack).

Since the PAUS sample is essentially a subset of the COSMOS2015 catalogue, the latter two samples are almost identical



**Fig. 2.** Redshift distribution of different calibration sample subsets (coloured, solid lines) and the one used by Hildebrandt et al. (2021) for KiDS-1000 (dashed black line). The bottom panel shows the excess of sources with photometric redshifts contributed by PAUS and PAUS + COSMOS2015 compared to the fiducial spectroscopic sample.

except that for 30% of the sources the photo- $z$  are augmented by the PAUS data (see Table 1). The PAUS sample is about half the size of the fiducial spectroscopic compilation and, while achieving a higher completeness at  $i_{AB} < 23$ , lacks many important faint, high redshift objects.

### 3. Redshift calibration with self-organising maps

A self-organising map (SOM, Kohonen 1982) is a very powerful tool that allows us to calibrate the redshift distribution of the KiDS-1000 lensing sample using the redshift compilations defined in the previous section. We adopt the SOM methodology of Wright et al. (2020a) which additionally provides a metric to select only those parts of the KiDS colour space in which we can reliably map out the colour-redshift relation.

#### 3.1. SOM redshift calibration methodology

The basic idea of the SOM methodology dates back to Lima et al. (2008) who introduced a redshift calibration strategy built on the assumption that two galaxy samples with the same colour-space distribution follow the same redshift distribution. Therefore they suggest to derive the unknown redshift distribution  $N(z)$  of a photometric galaxy sample from a calibration sample with accurate, preferentially spectroscopic redshifts  $N^{cal}(z)$  that is constructed such that it is representative of the photometric sample. This method is called ‘direct calibration’ (DIR). In practice, however, such a calibration sample has typically a substantially different selection function. Therefore Lima et al. (2008) propose a re-weighting scheme to match the calibration to the photometric sample by computing the ratio of the local galaxy density of both samples in the high-dimensional colour-space spanned by the photometric observations. This can be achieved for example by counting neighbours in a fixed volume around a point in the colour-space or by computing the volume occupied by a fixed number of nearest neighbours. Provided that both samples initially cover the same volume of the colour-space this method should recover the true redshift distribution, even in the presence of colour-redshift degeneracies.

This method is still susceptible in particular to selection biases and incompleteness introduced by spectroscopic targeting

strategies and success rates (e.g. Gruen & Brimiouille 2017; Hartley et al. 2020). Recent work by Wright et al. (2020a) has shown that this can be alleviated by performing additional cleaning and selections (quality control, see Sect. 3.2) on the unknown sample, creating a ‘gold sample’ containing only galaxies of the photometric sample that are sufficiently represented by the calibration sample. They implement this by training a SOM on the colour-space of the calibration sample and then parse the photometric sample into the same cells. Cells that are not occupied by objects from both samples are rejected, effectively removing those critical parts of the colour-space. They improve the cleaning procedure by applying hierarchical clustering on the SOM to find groups of cells with similar photometric properties instead of filtering individual cells. This allows a more fine-grained trade-off between the number of photometric sources rejected due to partitioning of the high-dimensional colour-space and the bias introduced by misrepresentation of the gold sample.

Finally, they compute the DIR weight for each of the  $n$  SOM groupings  $\mathcal{G} = \{g_1, \dots, g_n\}$  which is the ratio of calibration-to-gold sample objects. They obtain the redshift distribution of the gold sample

$$N(z) = \sum_{g \in \mathcal{G}} N_g^{\text{cal}}(z) \frac{N_{g,\text{tot}}^{\text{gold}}}{N_{g,\text{tot}}^{\text{cal}}} \quad (1)$$

by calculating the DIR-weighted sum of the redshift distributions  $N_g^{\text{cal}}(z)$  of the calibration sample in each SOM grouping.  $N_{g,\text{tot}}^{\text{cal}}$  and  $N_{g,\text{tot}}^{\text{gold}}$  are the total number of calibration sample and gold sample objects of group  $g$ , respectively.

### 3.2. Application to KiDS-1000

For our analysis we largely follow Wright et al. (2020a) and train a SOM with  $101 \times 101$  hexagonal cells and periodic boundaries on the full calibration sample (*spec-z+PAUS+COS15*, see Sect. 2.4.1). The input features are the matched KiDS  $r$ -band magnitudes and all 36 possible KiDS-colours that can be formed from the *ugriZYZJK<sub>s</sub>* imaging. Next, we divide the calibration and the KiDS-1000 source sample into the five tomographic bins and parse both samples into the SOM cells. We then run the hierarchical clustering for which we use the same number of clusters per bin (4000, 2200, 2800, 4200, and 2000) as Wright et al. (2020a) since these numbers were calibrated using simulations<sup>6</sup> (van den Busch et al. 2020). Even though each gold sample has a different optimal number of clusters, simulating the new redshift compilation and including realistic photo- $z$  errors is beyond the scope of this work.

We are using the same SOM for the remaining calibration samples defined in Sect. 2.4 and simply parse the corresponding subset of the full calibration sample back into the SOM before running the hierarchical clustering. For each of these calibration samples we apply a final cleaning step to the SOM groupings by defining a quality cut

$$|\langle z_{\text{cal}} \rangle - \langle z_{\text{B}} \rangle| > 5\sigma_{\text{mad}}, \quad (2)$$

where  $\sigma_{\text{mad}} = \text{nMAD}(\langle z_{\text{cal}} \rangle - \langle z_{\text{B}} \rangle)$  is the normalised median absolute deviation from the median, where the normalisation ensures that the nMAD reproduces the traditional standard deviation in the limit of Gaussian noise. This selection rejects clusters of SOM cells in which the mean calibration sample redshift

<sup>6</sup> These simulations are tailored to fit the KiDS imaging data and are based on the MICE2 simulation (Fosalba et al. 2015a,b; Crocce et al. 2015; Carretero et al. 2015; Hoffmann et al. 2015).

$\langle z_{\text{cal}} \rangle$  and the mean KiDS photometric redshifts  $\langle z_{\text{B}} \rangle$  catastrophically disagree. Wright et al. (2020a) found that this additional cleaning significantly reduces the SOM redshift bias while the impact on the number density is small and does not exceed a few percent. The rejection threshold of  $\sigma_{\text{mad}} \approx 0.12$  is calculated for the *spec-z fiducial* case and is applied to all other samples. This choice is motivated by the fact that this value is very close to the one calibrated with mock data for KiDS-1000 by (Hildebrandt et al. 2021), whereas  $\sigma_{\text{mad}}$  would be twice as large if we were calculating this threshold based on *spec-z+PAUS+COS15*. One reason for this difference in  $\sigma_{\text{mad}}$  is that the COSMOS2015 data allow the inclusion of additional populations of faint galaxies for which the calibration sample reference redshifts and the KiDS photo- $z$  are more likely discrepant, increasing the spread of the distribution of  $\langle z_{\text{cal}} \rangle - \langle z_{\text{B}} \rangle$ . We discuss this effect further in Sect. 6.1 and App. B.

This final selection step defines our gold sample for which we compute the redshift distributions according to Eq. (1). Since we require weighted redshift distributions for our cosmological analysis we substitute  $N_{g,\text{tot}}^{\text{gold}}$  by  $W_{g,\text{tot}}^{\text{gold}} = \sum_{i \in g} w_i$  which is the sum over the individual galaxy weights  $w_i$  from shape measurements in the SOM group  $g$ .

### 3.3. Clustering redshifts

There is one key difference to the calibration methodology of Hildebrandt et al. (2021) which is that we choose to omit the clustering redshift analysis in this work. While this choice limits our ability to validate the redshift distributions of our new gold samples, the SOM method is our fiducial calibration method and is therefore the focus of this work. In addition to that, the newly included calibration data does not increase the spatial overlap with KiDS significantly and is, due to its inhomogeneity, difficult to administer in a cross-correlation analysis. We leave this validation and joint analysis with the clustering redshifts to future work.

## 4. Cosmological analysis

In this section we summarise our cosmic shear analysis pipeline which we adopt from Asgari et al. (2021, A21 hereafter).

### 4.1. Cosmic shear

The primary observable of cosmic shear are the shear two-point correlation functions (2PCFs, Kaiser 1992)

$$\xi_{\pm}(\theta) = \langle \gamma_t \gamma_t \rangle(\theta) \pm \langle \gamma_{\times} \gamma_{\times} \rangle(\theta), \quad (3)$$

where  $\gamma_t$  and  $\gamma_{\times}$  are the tangential and the cross component of the shear, defined with respect to the line connecting a pair of galaxies (see e.g. Bartelmann & Schneider 2001). We use a weighted estimator for the shear correlations  $\xi_{\pm}$  as a function of the separation angle  $\theta$  between two tomographic redshift bins  $i$  and  $j$ :

$$\hat{\xi}_{\pm}^{(ij)}(\bar{\theta}) = \frac{\sum_{ab} w_a w_b [\epsilon_{t,a}^{\text{obs}} \epsilon_{t,b}^{\text{obs}} \pm \epsilon_{\times,a}^{\text{obs}} \epsilon_{\times,b}^{\text{obs}}] \Delta_{ab}^{(ij)}(\bar{\theta})}{\sum_{ab} w_a w_b (1 + \bar{m}_a)(1 + \bar{m}_b) \Delta_{ab}^{(ij)}(\bar{\theta})} \quad (4)$$

Here,  $\Delta_{ab}^{(ij)}(\bar{\theta})$  is a function that expresses whether a pair of galaxies,  $a$  and  $b$ , falls into an angular bin labelled by  $\bar{\theta}$ . Each galaxy

has a weight  $w$  and measured ellipticities  $\epsilon_{\ell}^{\text{obs}}$  and  $\epsilon_{\times}^{\text{obs}}$ . The denominator applies the multiplicative shear bias  $m$ , which corrects the measured shear to match the true galaxy shear.<sup>7</sup>

We extract the cosmological information from the shear correlation signal using COSEBIs (complete orthogonal sets of E/B-integrals, Schneider et al. 2010). These present a method to cleanly decompose the shear 2PCFs into  $E$ - and  $B$ -modes by applying a set of oscillatory filter functions defined over a finite angular range between  $\theta_{\min}$  and  $\theta_{\max}$ . The filter functions  $T_{\pm n}(\theta)$  for the  $n$ -th COSEBI mode ( $E_n$  and  $B_n$ ) have exactly  $n + 1$  roots:

$$E_n = \frac{1}{2} \int_{\theta_{\min}}^{\theta_{\max}} d\theta \theta [T_{+n}(\theta) \xi_+(\theta) + T_{-n}(\theta) \xi_-(\theta)] \quad (5)$$

$$B_n = \frac{1}{2} \int_{\theta_{\min}}^{\theta_{\max}} d\theta \theta [T_{+n}(\theta) \xi_+(\theta) - T_{-n}(\theta) \xi_-(\theta)] \quad (6)$$

One of the advantages of this formalism compared to the classical 2PCFs is that COSEBIs are less sensitive to small scales, where the complex physics of baryon feedback plays an important role, if a subset of the modes is chosen accordingly (Asgari et al. 2020).

## 4.2. Analysis pipeline

Our analysis pipeline is an upgraded version of CosmoPipe<sup>8</sup> (Wright et al. 2020b) which is a wrapper for Cat\_to\_Obs<sup>9</sup> (Giblin et al. 2021) and the KiDS Cosmology Analysis Pipeline<sup>10</sup> (KCAP, Joachimi et al. 2021; Asgari et al. 2021; Heymans et al. 2021; Tröster et al. 2021) that have both been used previously to analyse the KiDS-1000 data. The pipeline measures the shear 2PCFs using TREECORR (Jarvis et al. 2004; Jarvis 2015) on angular scales between  $0.5'$  and  $300'$  from which we compute the first five COSEBI modes using the logarithmic versions of the filter functions  $T_{\pm n}(\theta)$ . The logarithmic versions achieve a better compression of the cosmological signal onto fewer COSEBI modes.

We use the CosmoSIS framework (Zuntz et al. 2015) to compute theoretical predictions with the KCAP COSEBI module (Asgari et al. 2012). The linear matter power spectrum is modelled with CAMB (Lewis et al. 2000; Howlett et al. 2012) and its non-linear evolution with HMCode (Mead et al. 2015, 2016), whereas intrinsic alignments are calculated based on the model of Hirata & Seljak (2004); Bridle & King (2007). We then compare these predictions to the measured COSEBIs by sampling a Gaussian likelihood with MULTINEST (Feroz et al. 2009) using the analytical covariance model and priors of Joachimi et al. (2021). From this we infer constraints on the cosmological parameters of a spatially flat  $\Lambda$ CDM model. We additionally marginalise over a set of sample-dependent nuisance parameters which capture uncertainties in the shear and redshift calibration. Since Monte-Carlo samplers like MULTINEST are not designed to find the best fitting model parameters, we additionally run a Nelder-Mead minimiser (Nelder & Mead 1965) starting from the maximum posterior point of all chains.

Based on this we quote parameter constraints and their uncertainty as the fit parameter value and the projected joint highest posterior density (PJ-HPD) that we obtain from the MULTINEST

<sup>7</sup> In practice we do not apply the  $m$ -bias per galaxy in Eq. (4) but instead take the average value in each tomographic bin to avoid effects such as galaxy detection biases (Kannawadi et al. 2019).

<sup>8</sup> <https://github.com/AngusWright/CosmoPipe>

<sup>9</sup> [https://github.com/KiDS-WL/Cat\\_to\\_Obs\\_K1000\\_P1](https://github.com/KiDS-WL/Cat_to_Obs_K1000_P1)

<sup>10</sup> <https://github.com/KiDS-WL/kcap>

**Table 2.** The revised multiplicative shear bias  $m_{\text{new}}$  for the KiDS-1000 sample compared to the original values  $m_{\text{old}}$  of Asgari et al. (2021).

Bin	Photo- $z$ range	$m_{\text{old}}$	$m_{\text{new}}$	$\sigma_m$	$\Delta z = z_{\text{est}} - z_{\text{true}}$
1	$0.1 < Z_B \leq 0.3$	-0.009	-0.010	0.019	$0.000 \pm 0.0096$
2	$0.3 < Z_B \leq 0.5$	-0.011	-0.009	0.020	$0.002 \pm 0.0114$
3	$0.5 < Z_B \leq 0.7$	-0.015	-0.011	0.017	$0.013 \pm 0.0116$
4	$0.7 < Z_B \leq 0.9$	0.002	0.008	0.012	$0.011 \pm 0.0084$
5	$0.9 < Z_B \leq 1.2$	0.007	0.012	0.010	$-0.006 \pm 0.0097$

**Notes.** We note that the uncertainties  $\sigma_m$  remain unchanged. The last column summarises the parameters of the correlated Gaussian redshift priors ( $\mu_i \pm \sigma_i$ , where  $\sigma_i$  is the square root of the diagonal element of the covariance matrix).

chains. We note that both best fit parameters as well as the PJ-HPD have statistical uncertainties of about  $0.1\sigma$  or 10% on the  $1\sigma$  constraints due to the limited number of posterior samples (Joachimi et al. 2021).

### 4.2.1. Redshift uncertainty

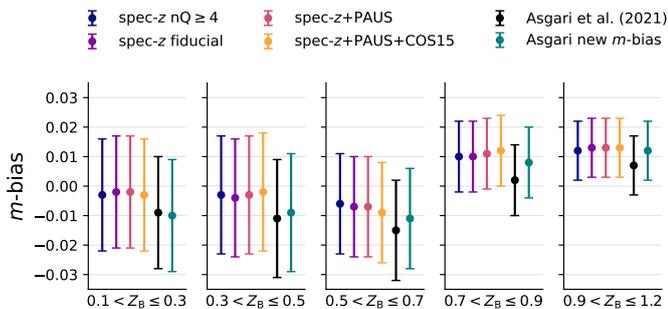
We propagate uncertainties in the redshift calibration to the cosmological constraints by allowing the redshift distribution of each tomographic bin  $i$  to vary by a shift  $\delta z_i$ . We use a set of correlated Gaussian priors  $\delta z_i \sim N(\mu_i, \sigma_i)$  which allows us to apply an empirical redshift bias correction by choosing offsets  $\mu_i \neq 0$ . These offsets and their correlations (Table 2) are calibrated from spectroscopic and KiDS-like mock data (van den Busch et al. 2020) in Hildebrandt et al. (2021). Since our analysis uses different calibration samples with altered sample selections, we would in principle need to perform a similar mock data analysis to recalibrate the priors for  $\delta z_i$ . However, these new samples contain many new spectroscopic datasets and the inclusion of photometric redshifts presents an additional challenge when attempting to model realistic photo- $z$  errors. Therefore, we assume that the variance of the KiDS-1000 priors is conservative enough to absorb potential changes of the redshift biases from KiDS-1000 to the new gold samples.

### 4.2.2. Multiplicative shear uncertainty

The second set of sample-dependent nuisance parameters is the average multiplicative shear bias ( $m$ -bias, see Eq. 4) in each tomographic bin. The effect of the  $m$ -bias and its uncertainty on the COSEBIs is captured in the covariance matrix and is calibrated by comparing the true galaxy ellipticities to those measured from a suite of image simulations generated by Kannawadi et al. (2019). The  $m$ -bias values vary little from sample to sample but are by up to  $0.5\sigma$  larger than those of the KiDS-1000 sample (Giblin et al. 2021). This led to the discovery of an issue with the way KiDS galaxies are assigned to galaxies in the COSMOS field which provides us with accurate shape information. We recompute the  $m$ -bias (Table 2, Fig. 3) and find that the revised values are in good agreement with those of the new gold samples. The updated values are also well within the uncertainty on  $m$  that was accounted for in A21.

## 5. Results

In this section we present the new KiDS gold samples and cosmological constraints from an analysis of COSEBIs.



**Fig. 3.** The  $m$ -bias values calculated for each gold sample per tomographic bin. Additionally, the values used by [Asgari et al. \(2021\)](#) are shown in black and their revised values in teal.

### 5.1. New KiDS gold samples

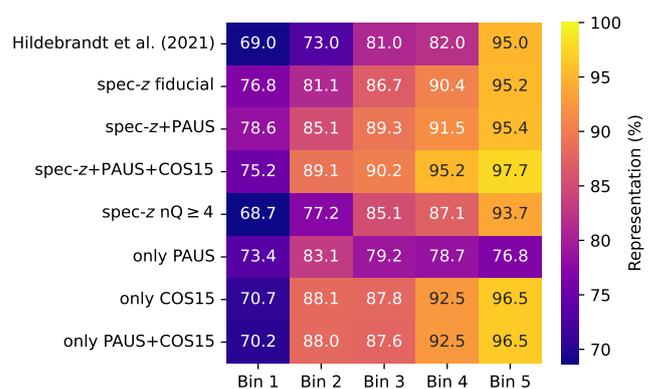
Similar to the calibration samples (Sect. 2.4) we divide the gold samples in two categories: primary, which are based on the full compilation of spectroscopic data plus optionally photo- $z$  (*spec- $z$  fiducial*, *spec- $z$ +PAUS*, and *spec- $z$ +PAUS+COS15*, see Sect. 2.4.1), and secondary samples, which are calibrated with subsets of the spectroscopic calibration sample (*spec- $z$   $nQ \geq 4$* ) or by using exclusively photo- $z$  (*only-PAUS*, *only-COS15*, and *only-PAUS+COS15*, see Sect. 2.4.2).

#### 5.1.1. Primary gold samples

We make a quantitative comparison of the selection of the three primary KiDS gold samples based on the representation fraction of each tomographic bin, the effective sample number density compared to the density of the full KiDS-1000 source sample. These are summarised for all gold samples in Fig. 4. The numbers show that our new spectroscopic redshift compilation provides a much greater coverage of the KiDS source sample since our fiducial gold sample has a 9% higher accumulated number density than the previous data set calibrated by [Hildebrandt et al. \(2021\)](#), increasing the total representation fraction from 80% to 89%. In comparison to the former, our *spec- $z$  fiducial* gold sample and those constructed by the addition of the PAUS and COSMOS2015 photo- $z$  steadily increase the coverage fraction of KiDS galaxies across all tomographic bins, rising from 73%, 81%, and 82% to 88%, 90%, and 95% in bins 2, 3, and 4 respectively. The fifth tomographic bin, which contributes most of the cosmological signal in the cosmic shear analysis, shows the least change in its representation fraction due to the already very high coverage of 95% reported by [Hildebrandt et al. \(2021\)](#).

The only exception to the steadily increasing representation fractions is the first bin of *spec- $z$ +PAUS+COS15*, where the number density is about 2.5% lower compared to the fiducial case. This is the result of the quality control (Eq. 2) removing some SOM groupings due to discrepancies in  $\langle z_{\text{cal}} \rangle$  and  $\langle z_B \rangle$ , which arise when adding calibration sources and/or changing the gold selection. Expanding the calibration sample may shift  $\langle z_{\text{cal}} \rangle$  significantly, in particular in sparsely occupied SOM groupings, such that  $\langle z_{\text{cal}} \rangle - \langle z_B \rangle$  exceeds the quality control threshold  $5\sigma_{\text{mad}}$ , which will flag and exclude the corresponding KiDS galaxies from the gold sample.

The SOM redshift distributions of the gold samples are shown in Fig. 5; note that these samples do not represent the same galaxies. A comparison reveals two effects when adding photo- $z$  to the calibration sample: First, the bulk of the redshift distributions is skewed to lower redshifts as we add more data to



**Fig. 4.** Representation fractions, the effective number density of the different gold samples relative to the full KiDS-1000 source sample, per tomographic bin. The effective number density factors in the lensing weight of each object and is calculated according to Eq. (C.12) in [Joachimi et al. \(2021\)](#).

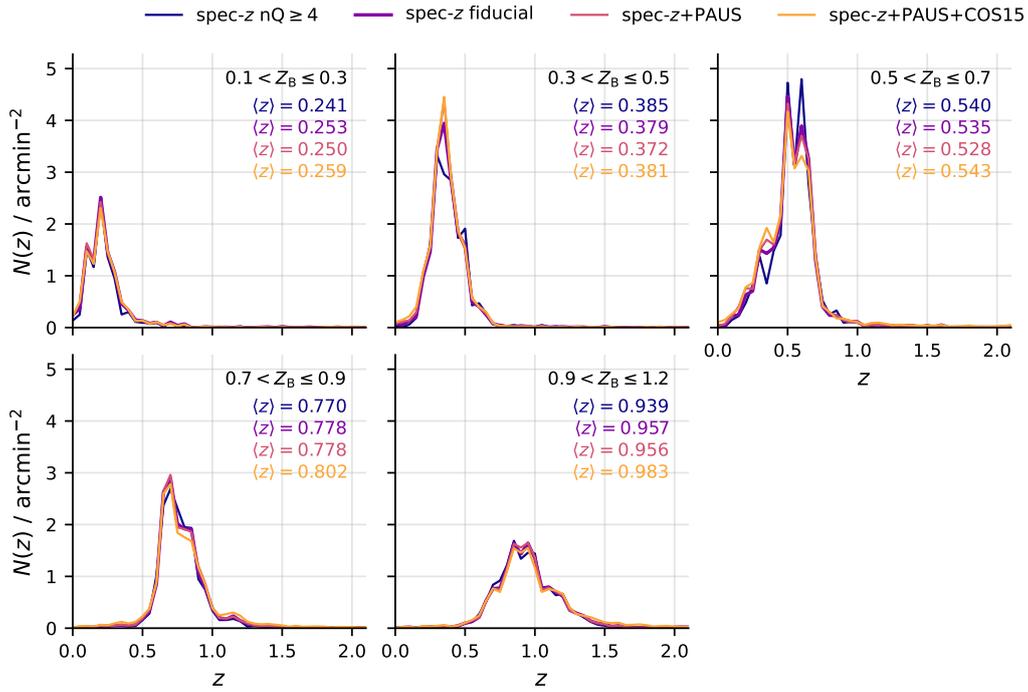
the calibration sample which is most evident in the third tomographic bin. Secondly, COSMOS2015 adds a significant portion of high redshift objects to the compilation that extends the coverage of KiDS galaxies to higher redshifts, enhancing the tails of the redshift distributions and significantly increasing the mean redshifts, in particular of bin 5.

Finally we compare the calibration sample redshift distributions in each tomographic bin to the resulting gold sample redshift distributions (Fig. 6). For most of the samples these distributions have very similar shapes and with mean redshifts agreeing within  $\pm 0.02$ , which indicates that the re-weighting (Eq. 1) of the SOM groupings is very small on average. This changes once the COSMOS2015 redshifts are added to the calibration sample. Due to their significantly higher depth and mean redshifts, the redshift tails must be down-weighted significantly (up-weighted in bin 1 and 2) to match the density of the KiDS source sample. The down-weighting of the low redshift tails in the upper three bins can be explained by the fact that COSMOS2015 adds faint galaxies at these redshifts. The corresponding KiDS galaxies have a low lensing weight, which must be compensated by the SOM cell weights.

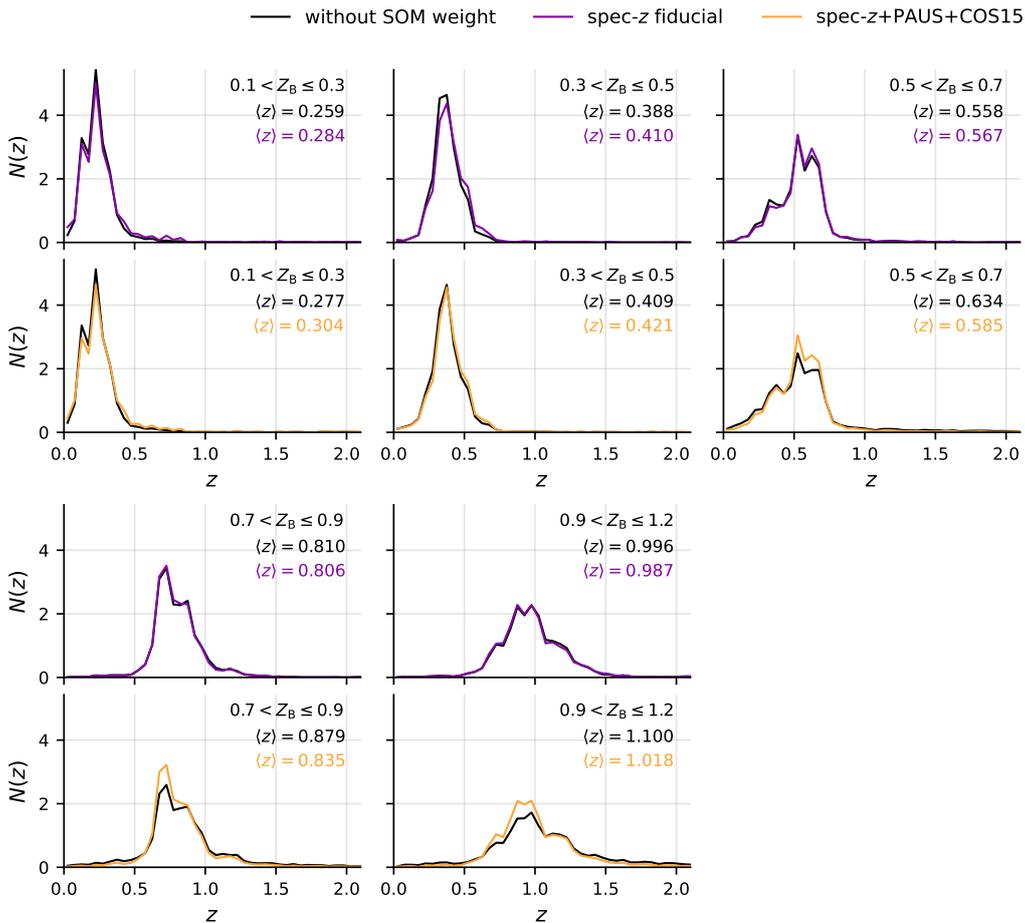
#### 5.1.2. Secondary gold samples

The first of our secondary gold samples we calibrate using only those galaxies of the spectroscopic compilation that have the most secure spectroscopic redshifts of at least 99% confidence. Due to similarities to the SOM calibration sample used by [Hildebrandt et al. \(2021, see Sect. 2.4\)](#), this *spec- $z$   $nQ \geq 4$*  gold sample positions itself in between the latter and the *spec- $z$  fiducial* sample in terms of representation fractions. In tomographic bin 4 and 5 however it is lacking some of the high redshift sources due to the more conservative spectroscopic flagging, reducing both the mean redshifts (Fig. 5) as well as the representation fractions (Fig. 4).

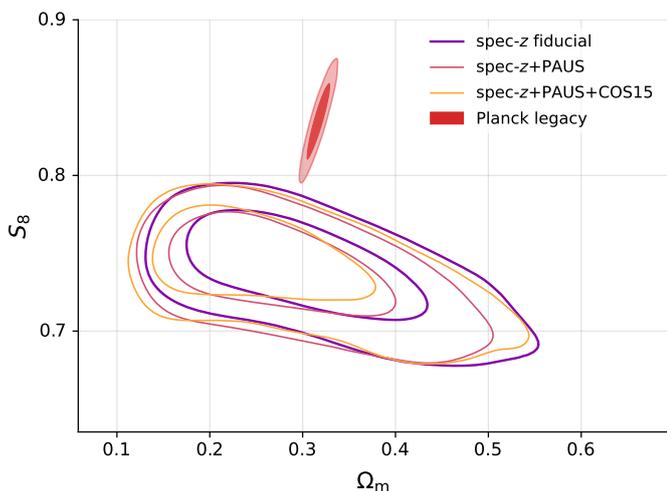
The remaining three gold samples exclusively rely on photometric redshifts from PAUS and COSMOS2015 (Sect. 2.4.2). Due to the great overlap of sources, both samples that contain the COSMOS2015 photo- $z$  (*only-COS15* and *only-PAUS+COS15*) achieve representation fractions which are just 2% smaller than those of the full redshift compilation, again due to the quality control (Eq. 2). As seen for *spec- $z$ +PAUS+COS15*, the effective density in the first bin is significantly reduced. The *only-PAUS*



**Fig. 5.** Comparison of the gold sample redshift distributions and their mean redshifts for all tomographic bins obtained from the different subsets of the calibration sample. We note that they all describe different subsets of the full KiDS-1000 source sample.



**Fig. 6.** Comparison of the tomographically binned calibration sample (black lines) to the gold sample redshift distributions (coloured lines). The greater the difference between the black and the coloured lines, the more weighting is applied by the SOM to match the calibration sample to the KiDS data.



**Fig. 7.** Marginalised constraints for the joint distributions of  $S_8$  and  $\Omega_m$  (68 % and 95 % credible regions) obtained for different gold samples and Planck legacy (TT, TE, EE + lowE). Since the contours represent different galaxy samples, some deviation is expected.

gold sample exhibits significantly suppressed tails in bin 4 and 5 due to a lack of high redshift sources that only the much deeper COSMOS2015 provides. This also imprints on the representation fractions which are similar to the fiducial gold sample in bin 1 and 2 but are about 10 % lower in bin 3 and 4 and 18 % lower in the fifth tomographic bin. This example highlights the abilities of the SOM to flag and remove sources that cannot be calibrated by the DIR approach with the particular calibration sample.

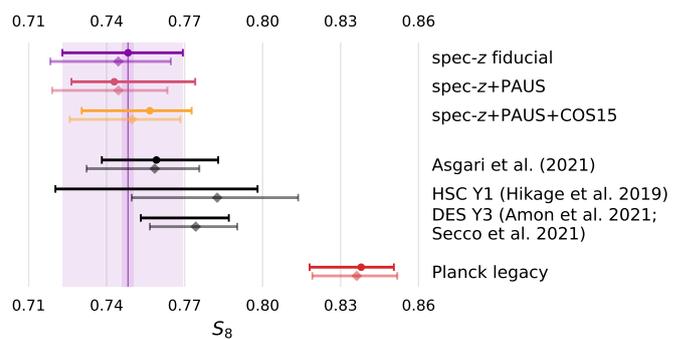
## 5.2. Cosmological constraints

We present cosmological results for our primary KiDS gold samples, focusing on a relative comparison to A21 and other literature values. We summarise the numerical values of the most relevant cosmological parameters in Table 3. In Fig. 7 and 8 we highlight comparisons of the derived parameter  $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$ , which is the primary measurable of weak lensing due to the degeneracy between  $\Omega_m$ , the dimensionless matter density parameter, and  $\sigma_8$ , parameterising the amplitude of the linear power spectrum.

First we ensure that our wrapper for the cosmological pipeline delivers results that are consistent with those of A21. We re-analyse the original KiDS-1000 data, running our pipeline with the same input parameters (‘Asgari reanalysis’). We find that the constraints on all cosmological parameters agree within the expected variance of the Monte-Carlo sampler. The best fit solution has a slightly larger  $\chi^2$  of 83.8 compared to 82.2 for KiDS-1000 which is driven primarily by the slight change in the data vector (see App. C). Similarly, the correction of the  $m$ -bias values (Table 2) has no significant impact on the cosmological constraints either (Table. C.1).

### 5.2.1. Primary gold samples

For our KiDS gold samples we find a tendency to lower  $S_8$ -values, albeit at low significance (Figs. 7, 8 and Table 3). In particular we obtain  $S_8 = 0.748^{+0.021}_{-0.025}$  for the *spec-z fiducial* and  $S_8 = 0.743^{+0.031}_{-0.016}$  for the *spec-z+PAUS* gold samples which is about  $0.4\sigma$  lower compared to A21 and exceeds the expected statistical variance of the sampler (about  $0.1\sigma$ ). Deviations to



**Fig. 8.** Comparison of the  $S_8$  constraints from our gold samples to other studies. We show the best fit (where available) and 68th-percentile PJ-HPD (circles, opaque data points) and the maximum of the marginal distribution and the associated 68th-percentile (diamonds, semi-transparent). We compare to Asgari et al. (2021), HSC-Y1 (Hikage et al. 2019), DES-Y3 (Amon et al. 2021; Secco et al. 2021), and Planck legacy. The coloured vertical line and outer bands indicate the constraints from the fiducial gold sample, the inner bands the expected variance of the sampler.

some degree between these gold samples are expected, since they all represent different galaxy selections. The goodness of fit improves for all gold samples but in particular for *spec-z fiducial*, where it reduces to  $\chi^2 \approx 63$  compared to 82 in KiDS-1000. The *spec-z+PAUS+COS15* sample on the other hand is in very good agreement with the original KiDS-1000 analysis.

In addition to  $S_8$  we consider the more general  $\Omega_m$ - $\sigma_8$ -degeneracy case of  $\Sigma_8 = \sigma_8(\Omega_m/0.3)^\alpha$ , where  $\alpha$  is a free parameter. We estimate the reference value for  $\alpha$  by fitting the posterior samples of the fiducial chain and obtain  $\alpha_{\text{fid}} = 0.55$  which is close to  $\alpha = 0.54$  for COSEBIs in A21. This projection optimises the signal-to-noise-ratio compared to  $S_8$  and we find  $\Sigma_8 = 0.744^{+0.017}_{-0.021}$  for *spec-z fiducial* compared to  $S_8 = 0.751^{+0.024}_{-0.016}$  which we obtain for A21 with  $\alpha = \alpha_{\text{fid}}$ . Furthermore, the scatter of  $\Sigma_8$  for the different gold samples is significantly smaller than the scatter in  $S_8$  (see Table 3).

If we compare the marginal errors of  $S_8$  for the different gold samples (Table D.1), which have a smaller statistical variance than the PJ-HPD, we find that the constraints improve by 5 % when including PAUS and another 4 % when including COSMOS2015. Since the constraints on  $\Sigma_8$  are almost constant, these changes in  $S_8$  are most probably related to small changes in the  $\Omega_m$ - $\sigma_8$ -degeneracy. We also compare  $A_{\text{IA}}$ , which is the dimensionless amplitude of the intrinsic alignment galaxy power spectrum, and find that its value is stable within the uncertainties in all our analyses.

## 6. Discussion

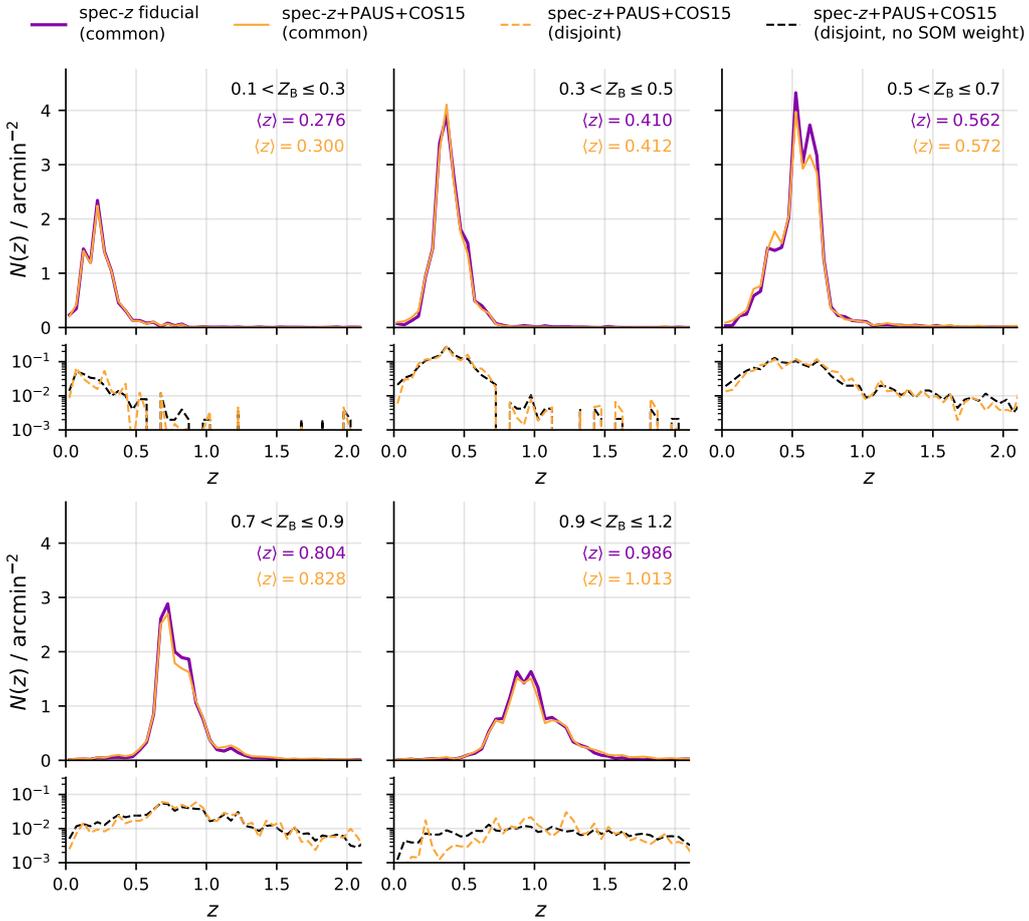
### 6.1. Gold sample selection and calibration

A side-by-side comparison of the different KiDS gold samples presented in Sect. 5.1 is non-trivial. On the one hand adding or removing galaxies from the calibration sample changes the redshift distribution of each of the SOM groupings that, according to Eq. (1), determine the sample’s redshift distribution. On the other hand two distinct gold samples are comprised of different galaxies since a modification of the calibration sample will also apply an implicit selection on the set of representative SOM groupings. Both of these effects combined determine the overall calibrated redshift distribution.

**Table 3.** Summary of the main cosmological parameter constraints (best fit and 68th-percentile PJ-HPD) from COSEBIs for all gold samples and their comparison to [Asgari et al. \(2021\)](#) and Planck legacy (TT, TE, EE + lowE).

Sample	$\chi^2$	$A_{IA}$	$\Omega_m$	$\sigma_8$	$S_8$	$\Sigma_8$
spec-z fiducial	63.2	$0.301^{+0.343}_{-0.377}$	$0.268^{+0.126}_{-0.055}$	$0.791^{+0.110}_{-0.155}$	$0.748^{+0.021}_{-0.025}$	$0.744^{+0.017}_{-0.021}$
spec-z+PAUS	64.0	$-0.097^{+0.422}_{-0.416}$	$0.226^{+0.108}_{-0.053}$	$0.857^{+0.130}_{-0.164}$	$0.743^{+0.031}_{-0.016}$	$0.732^{+0.028}_{-0.012}$
spec-z+PAUS+COS15	68.3	$0.164^{+0.450}_{-0.428}$	$0.194^{+0.109}_{-0.048}$	$0.940^{+0.144}_{-0.200}$	$0.757^{+0.016}_{-0.026}$	$0.740^{+0.013}_{-0.025}$
spec-z $nQ \geq 4$	61.6	$-0.033^{+0.408}_{-0.393}$	$0.195^{+0.079}_{-0.046}$	$0.951^{+0.113}_{-0.171}$	$0.766^{+0.023}_{-0.023}$	$0.750^{+0.021}_{-0.019}$
only PAUS	74.1	$-0.186^{+0.526}_{-0.345}$	$0.203^{+0.108}_{-0.053}$	$0.916^{+0.163}_{-0.183}$	$0.752^{+0.028}_{-0.019}$	$0.737^{+0.023}_{-0.017}$
only COS15	67.2	$0.164^{+0.339}_{-0.545}$	$0.253^{+0.132}_{-0.074}$	$0.814^{+0.157}_{-0.177}$	$0.747^{+0.013}_{-0.033}$	$0.741^{+0.017}_{-0.023}$
only PAUS+COS15	67.1	$0.078^{+0.515}_{-0.369}$	$0.216^{+0.176}_{-0.033}$	$0.876^{+0.085}_{-0.243}$	$0.744^{+0.020}_{-0.028}$	$0.731^{+0.024}_{-0.017}$
Asgari et al. (2021)	82.2	$0.264^{+0.424}_{-0.336}$	$0.246^{+0.101}_{-0.060}$	$0.838^{+0.140}_{-0.141}$	$0.759^{+0.024}_{-0.021}$	$0.751^{+0.024}_{-0.016}$
Planck legacy	—	—	$0.319^{+0.006}_{-0.010}$	$0.813^{+0.007}_{-0.008}$	$0.838^{+0.013}_{-0.020}$	$0.841^{+0.013}_{-0.021}$

**Notes.** Additionally shown are the  $\chi^2$  values (for 70.5 effective degrees of freedom) and  $\Sigma_8 = \sigma_8(\Omega_m/0.3)^\alpha$  calculated for  $\alpha_{fid} = 0.55$ , determined from a fit to the posterior samples of the fiducial chain.



**Fig. 9.** Comparison of the redshift distributions between the fiducial sample and the sample including the PAUS and COSMOS2015 photo-z. Top: The SOM redshifts derived from the subset of SOM groupings present in both samples. Bottom: The SOM redshift distributions of the groupings that are found only in the *spec-z+PAUS+COS15* sample (dashed yellow) and the underlying redshift distribution in the calibration sample of the same groupings, i.e. not applying the SOM weighting in Eq. (1), scaled to the amplitude of the former for comparison.

This is exemplified by the fact that, as we expand the calibration sample from the *spec-z  $nQ \geq 4$*  subset to *spec-z+PAUS+COS15*, we generally see that both the representation fractions and the mean redshifts increase across all bins. The galaxies added in each iteration are typically fainter (with the exception

of PAUS) at the cost of lower redshift accuracy, which allows us to calibrate additional KiDS galaxies, preferentially at the tails of the redshift distributions. At the same time we can observe in Fig. 5 that the redshift distributions are skewed to lower redshifts, which can be explained by the fact that there are dispro-

portionately many more galaxies with  $z < 0.5$  added in each iteration to the calibration sample (Fig. 2). These in turn increase the representation fraction of low and intermediate redshift galaxies in the gold sample (see also Fig. 4). This implies that we are changing the redshift calibration since the skewing applies to each individual SOM group. We can separate these two effects by splitting the *spec-z fiducial* and *spec-z+PAUS+COS15* gold samples in two subsets, one containing those SOM groupings that are common to both samples (i.e. containing the same KiDS galaxies) and groupings that can only be calibrated using the full redshift compilation. The subset of KiDS galaxies that is common to both gold samples shows the same redshift skewing as seen with all galaxies (top panels of Fig. 9), whereas the additional COSMOS2015 galaxies contribute significantly at the low and high redshift tails of the tomographic bins (bottom panels of Fig. 9).

The final ingredient to the redshift calibration is the quality control cut (Eq. 2) that we apply to remove potentially miscalibrated parts of the colour space. This becomes most obvious when comparing the representation fractions of the first tomographic bin (Fig. 4) which decrease whenever we add the COSMOS2015 data (compare *spec-z+PAUS* to *spec-z+PAUS+COS15* and *only-PAUS* to *only-PAUS+COS15*). The low redshift of the KiDS galaxies in this bin makes the sample particularly susceptible to the addition of high redshift galaxies. These can significantly change the mean redshift of the calibration sample ( $\langle z_{\text{cal}} \rangle$ ) compared to the KiDS photo- $z$  in the SOM groupings which then may fail to pass the quality control (see also App. B). The great depth of the COSMOS2015 data compared to the spectroscopic data also explains why the SOM needs to apply more weighting to match the *spec-z+PAUS+COS15* compilation to the KiDS colour-space (Fig 6).

## 6.2. Cosmological constraints

The gold sample selection effects are, due to their redshift dependence, directly propagated to the cosmological constraints (Sect. 5.2), causing shifts in  $S_8$  of up to  $0.5\sigma$  from sample to sample. One of the assumptions in our analysis is that we can adopt the same Gaussian priors for the  $\delta z_i$  nuisance parameters (Sect. 4.2) that are used by A21. We therefore re-analyse the fiducial gold sample assuming no knowledge of the empirical redshift bias by centring the priors on  $\mu_i = 0$ . For this run we find that the value of  $S_8 = 0.747^{+0.022}_{-0.022}$  is in good agreement with the fiducial analysis.

On the other hand it may seem that our choice for the widths  $\sigma_i$  of the  $\delta z_i$  priors may be insufficient to accommodate for the apparent variance in the mean redshifts of the different gold samples (see Fig. 5). This variance, however, is not only determined by potential systematic biases in the redshift calibration between any of these samples, but also by changes in the gold sample selection itself, as discussed above. Therefore, the question of the correct redshift prior can only be answered with realistic simulated data sets that are currently not available for our extended redshift calibration sample. Nevertheless, a comparison of the  $S_8$  values allows us to get an estimate of the variance induced by the selection effects in the calibration data and the resulting parameter constraints from the gold samples.

### 6.2.1. The fiducial gold sample

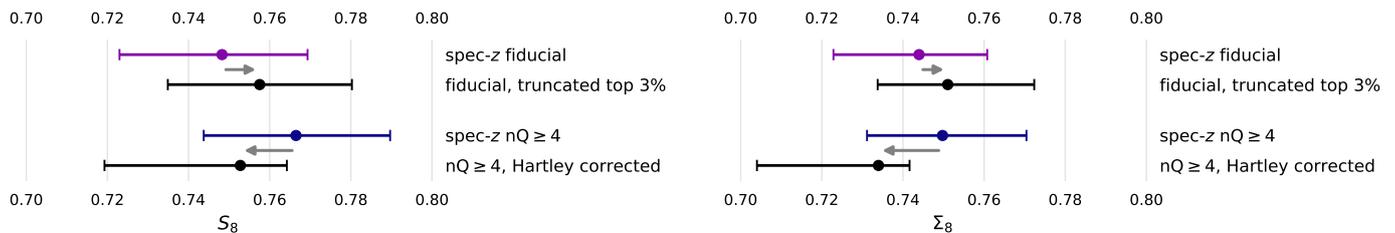
Next to the *spec-z fiducial* sample that is based on the full spectroscopic compilation we also define the *spec-z  $nQ \geq 4$*  sam-

ple that relies only on the most secure redshifts. The estimated  $S_8 = 0.766^{+0.023}_{-0.023}$  for the latter, arising from particularly low  $\Omega_m$  and high  $\sigma_8$  values, is about  $0.5\sigma$  larger than in the fiducial case and surpasses all other gold samples (Table 3). In the  $\Sigma_8$  projection this difference reduces to  $0.2\sigma$ .

The shift in  $S_8$  between gold samples that are both calibrated with spectroscopic data begs the question which of these estimates is more reliable. The primary difference between the two calibration datasets is the selection using redshift quality flags. Selecting spectra based on the redshift confidence is a trade-off between constructing a sample that is confined to regions of the colour-redshift-space in which galaxies have distinct spectral features that allow secure redshift determination and a sample with an increasing fraction of galaxies with catastrophically misidentified redshifts. In the latter case the redshifts of the calibration sample themselves cause a biasing of the gold sample redshifts and in turn  $S_8$ . In case of selecting only the highest quality redshifts, the biases arises from a misrepresentation of the imaging data by the calibration sample, as shown by Hartley et al. (2020). The redshift distribution of the calibration sample in each SOM grouping depends on the quality flag and thus the relative representation of different galaxy populations in the gold sample may change. We investigate the magnitude of both these effects by assuming two worst-case scenarios which shift down the  $S_8$  estimate obtained for *spec-z fiducial* and shift up  $S_8$  for *spec-z  $nQ \geq 4$* :

- A) We assume in case of the *spec-z fiducial* sample that 5 % of truly low redshift galaxies with  $3 \leq nQ < 4$  (nominal 95 % certainty), as well as 1 % in case of  $nQ \geq 4$  (99 % certainty), are catastrophically misidentified as high redshift galaxies. Since both redshift flags are equally common in the fiducial sample we expect a combined spectroscopic failure rate of about 3 %. We implement this worst-case scenario on *spec-z fiducial* by truncating the top 3 % of all redshift distributions which should increase the recovered  $S_8$  value. We calculate the redshift  $z_{97}$  corresponding to the 97-th percentile of the redshift distribution  $n(z)$ , set  $n(z) = 0$  at  $z > z_{97}$  and re-normalise to reproduce the original gold sample number density.
- B) We speculate that the calibration of the *spec-z  $nQ \geq 4$*  sample suffers from the same spectroscopic misrepresentation effects studied by Hartley et al. (2020, from Fig. 6 therein), who found redshift biases  $\langle z \rangle - \langle z_{\text{true}} \rangle$  of 0.008, 0.022,  $-0.003$ , and  $-0.058$  in the four tomographic redshifts bins of simulated DES and spectroscopic data, for the first time implementing a realistic, simulated  $nQ \geq 4$  sample selection. Since we currently do not have comparable spectroscopic mock data in KiDS, we assume in this scenario that the bias applies at the same magnitude to the *spec-z  $nQ \geq 4$*  gold sample. We therefore correct the assumed bias by interpolating the values from the four DES bins to the five tomographic bins of KiDS and shift the *spec-z  $nQ \geq 4$*  redshift distributions by  $-0.008$ ,  $-0.015$ ,  $-0.014$ ,  $0.003$ , and  $0.058$ . We consider this to be an even more conservative assumption than scenario A, since the bias should be significantly smaller in case of KiDS thanks to the 9-band imaging and the improvements of the SOM calibration over the classical DIR approach (Sect. 3.1) that is used by Hartley et al. (2020).

With these modifications to the redshift distributions, scenario A (*fiducial, truncated top 3 %*) yields a higher and scenario B ( *$nQ \geq 4$ , Hartley corrected*) a lower estimate for  $S_8$  (Fig. 10, left). These results indicate that the combination of these two effects may explain the observed differences between  $S_8$  in *spec-z*



**Fig. 10.** Comparison between constraints (best fit and 68th-percentile PJ-HPD) on  $S_8$  (left) and  $\Sigma_8$  (right) for *spec-z fiducial*, *spec-z  $nQ \geq 4$* , and the corresponding test scenario A and B. The grey arrows indicate shifts in  $S_8$  introduced in the tests by modifying the redshift distributions.

*fiducial* and *spec-z  $nQ \geq 4$* . When comparing the projection  $\Sigma_8$  instead (Fig. 10, right), which is less susceptible to shifts along the  $\Omega_m$ - $\sigma_8$ -degeneracy, the difference between *spec-z fiducial* and *spec-z  $nQ \geq 4$*  is much smaller. However the shift, introduced when correcting the redshift distributions in scenario B, is about twice as big compared to the  $S_8$  case, indicating that the selection effects studied in Hartley et al. (2020) can have a significant impact on cosmological constraints. Furthermore the completeness, which determines the ability to correctly map out colour-redshift degeneracies in each SOM grouping, and the quality of the calibration sample redshifts should be balanced carefully to minimise biases in the redshift calibration and the cosmological analysis.

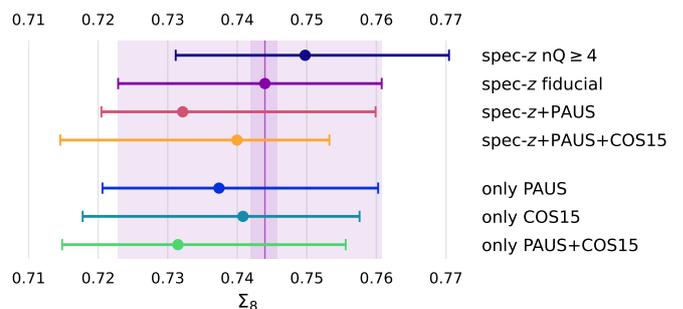
### 6.2.2. Other gold samples

Finally, we make a relative comparison of the  $\Sigma_8$  constraints from the remaining gold samples, since  $\Sigma_8$  typically exhibits a smaller scatter than the corresponding  $S_8$  values. The gold samples that are calibrated using only photo- $z$  from PAUS, COSMOS2015, or a combination of both, prefer smaller  $\Sigma_8$  values compared to *spec-z fiducial* (Fig. 11). While the cosmological constraints in a similar comparison between redshifts calibrated using spectroscopic data and COSMOS2015 (Hildebrandt et al. 2020) show a more pronounced shift in the opposite direction, this difference is caused by the gold sample selection (Wright et al. 2020b). In our analysis the shifts in  $\Sigma_8$  may be explained by the fact that COSMOS2015 tends to calibrate the KiDS galaxies to higher redshifts than the spectroscopic calibration data alone (see Fig. 9), translating to lower  $\Sigma_8$ . The same reasoning does not explain why  $\Sigma_8$  reduces further when the PAUS data is included, therefore this behaviour is most likely owed to the quality control (Eq. 2). When the photo- $z$  data is combined with the spectroscopic compilation, it may result in a significantly different distribution of  $|\langle z_{\text{cal}} \rangle - \langle z_B \rangle|$  for the SOM groupings, which has non-trivial implications for the gold sample selection and the derived cosmological constraints.

## 7. Conclusion

We apply the SOM redshift calibration technique (Wright et al. 2020a) to define and calibrate the redshift distributions of a new set of KiDS-1000 gold samples by adopting a new spectroscopic calibration sample. Compared to previous work by Hildebrandt et al. (2021) we double the size of this calibration sample by adding more than ten additional spectroscopic campaigns such as C3R2 and DEVILS, which allows us to calibrate an additional 9% of the KiDS galaxies.

We take this one step further by enhancing the calibration sample with precision photometric redshifts from the PAU survey and COSMOS2015, maintaining a hierarchy that prefers



**Fig. 11.** Constraints on  $\Sigma_8$  (best fit and 68th-percentile PJ-HPD) for all primary and secondary gold samples. The coloured vertical line and outer bands indicate the constraints from the fiducial gold sample, the inner bands the expected variance of the sampler.

spectroscopic over PAUS and COSMOS2015 redshifts to resolve duplicates in the three catalogues. The resulting KiDS gold sample increases by additional 6% and covers nearly 98% of all KiDS galaxies in the fifth tomographic bin.

When comparing these gold samples we find changes in the mean redshifts of up to  $|\Delta z| = 0.026$  which originate from selection effects in the calibration sample. First, there are residual modifications to the redshift calibration of those KiDS sources that are found in both gold samples. These modifications are a direct consequence of changing the redshift distribution of the calibration sample when adding the photo- $z$ . Second and most important, the selection of KiDS sources itself changes since the faint COSMOS2015 data allows us to calibrate additional galaxies at both low and high redshifts. These results highlight the importance of quantifying and calibrating potential method dependent redshift biases arising from selection effects, as has been shown in previous work. This requires sophisticated galaxy mock data with sufficient redshift coverage, realistic galaxy colours and accurate modelling of photometric and spectroscopic galaxy samples, in particular if one aims to study the impact of photo- $z$  outliers on the calibration sample.

In the second part of this study we perform a cosmic shear analysis using COSEBIs and find  $S_8 = 0.748^{+0.021}_{-0.025}$  for our fiducial and  $S_8 = 0.757^{+0.016}_{-0.026}$  for the photo- $z$ -enhanced gold sample which is slightly lower than, but still in excellent agreement with, previous work on KiDS-1000 by Asgari et al. (2021). As part of our additional systematic testing we create a third gold sample that we calibrate using only the most secure spectra of our spectroscopic compilation ( $nQ \geq 4$ ). We measure  $S_8 = 0.766^{+0.023}_{-0.023}$  which corresponds to an increment of  $0.5\sigma$  compared to the fiducial case. Our analysis speculates that this sample underestimates the true redshifts and overestimates  $S_8$  due to an implicit selection introduced when limiting the calibration to the most secure spectroscopic redshifts (Hartley et al. 2020). Therefore

completeness and quality of the calibration sample can be optimised, this however requires realistic simulations to quantify the impact of this trade-off on the recovered redshift distributions.

Finally we analyse four KiDS gold samples which are all calibrated from different subsets of our extended calibration sample or using only photo- $z$  from PAUS and COSMOS2015. No matter how we calibrate the KiDS source galaxies, we find that all seven gold samples studied in this work scatter in the range of  $S_8 = 0.743 \dots 0.766$  around our fiducial analysis. This further confirms previously reported tensions of KiDS with measurements of the CMB by [Planck Collaboration et al. \(2020\)](#) at  $3.0\sigma$  to  $3.6\sigma$  for a flat  $\Lambda$ CDM model.

In summary, there seems to be little benefit in using precision photometric redshifts for the SOM redshift calibration given the excellent coverage of the KiDS data by our spectroscopic compilation alone. The *spec- $z$ +PAUS+COS15* gold sample achieves, compared to our fiducial sample, a 6% improvement in terms of the number density, yet improvements on the cosmological constraints are marginal. Nevertheless, if the spectroscopic coverage is significantly lower or one wishes to target higher redshifts, photo- $z$  samples are a valuable source of complementary calibration data. However, the greater the dependence on photometric redshifts is, the more attention must be paid to redshift outliers to guarantee a good balance between statistical uncertainties and systematic biases in the redshift calibration. Given their challenging calibration requirements this is in particular true for the next generation, stage IV surveys such as Euclid ([Laureijs et al. 2011](#)) or the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST, [Ivezić et al. 2019](#)). Despite photo- $z$  outliers, photometric redshift samples have one significant advantage over spectroscopically selected samples: they achieve a much higher completeness, which can mitigate selection effects in the calibration sample and may improve the ability of the SOM to map out the full extent of colour-redshift-degeneracies.

*Acknowledgements.* We acknowledge support from the European Research Council under grant numbers 770935 (JvdB, AHW, HH) and 647112 (MA, CH, TT). HH is also supported by a Heisenberg grant (Hi1495/5-1) of the Deutsche Forschungsgemeinschaft. MB is supported by the Polish National Science Center through grants no. 2020/38/E/ST9/00395, 2018/30/E/ST9/00698 and 2018/31/G/ST9/03388, and by the Polish Ministry of Science and Higher Education through grant DIR/WK/2018/12. CH is additionally supported by the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck-Humboldt Research Award endowed by the Federal Ministry of Education and Research. HYS acknowledges the support from CMS-CSST-2021-A01, NSFC of China under grant 11973070, the Shanghai Committee of Science and Technology grant No.19ZR1466600 and Key Research Program of Frontier Sciences, CAS, Grant No. ZDBS-LY-7013. We are grateful to Mara Salvato and the zCOSMOS team to give us early access to additional deep spectroscopic redshifts that were not available in the public domain, as well as to Dan Masters for sharing pre-release C3R2 DR3 data. This work is based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 100.A-0613, 102.A-0047, 179.A-2004, 177.A-3016, 177.A-3017, 177.A-3018, 298.A-5015, and on data products produced by the KiDS consortium. This work has made use of CosmoHub ([Carretero et al. 2017](#)). CosmoHub has been developed by the Port d'Informació Científica (PIC), maintained through a collaboration of the Institut de Física d'Altes Energies (IFAE) and the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), and was partially funded by the "Plan Estatal de Investigación Científica y Técnica y de Innovación" program of the Spanish government. The figures in this work were produced with MATPLOTLIB ([Hunter 2007](#)) and GETDIST ([Lewis 2019](#)).

*Author Contributions.* All authors contributed to the development and writing of this paper. The authorship list is given in three groups: the lead authors (JLvdB, AHW, HH, MB, and MA), followed by two alphabetical groups. The first alphabetical group includes those who are key contributors to both the scientific analysis and the data products. The second group covers those who have either made a significant contribution to the data products or to the scientific analysis.

## References

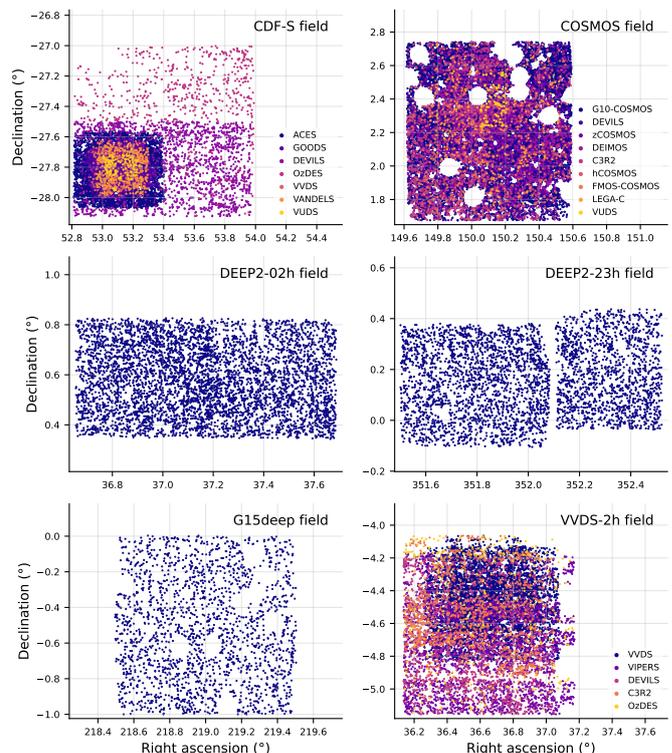
- Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, *PASJ*, 70, S8  
 Alarcon, A., Gaztanaga, E., Eriksen, M., et al. 2021, *MNRAS*, 501, 6103  
 Alarcon, A., Sánchez, C., Bernstein, G. M., & Gaztañaga, E. 2020, *MNRAS*, 498, 2614  
 Amon, A., Gruen, D., Troxel, M. A., et al. 2022, *Phys. Rev. D*, 105, 023514  
 Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, *A&A*, 645, A104  
 Asgari, M., Schneider, P., & Simon, P. 2012, *A&A*, 542, A122  
 Asgari, M., Tröster, T., Heymans, C., et al. 2020, *A&A*, 634, A127  
 Balestra, I., Mainieri, V., Popesso, P., et al. 2010, *A&A*, 512, A12  
 Bartelmann, M. & Schneider, P. 2001, *Phys. Rep.*, 340, 291  
 Benítez, N. 2000, *The Astrophysical Journal*, 536, 571  
 Bridle, S. & King, L. 2007, *New Journal of Physics*, 9, 444  
 Buchs, R., Davis, C., Gruen, D., et al. 2019, *MNRAS*, 489, 820  
 Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, *MNRAS*, 447, 646  
 Carretero, J. et al. 2017, *PoS, EPS-HEP2017*, 488  
 Cooper, M. C., Yan, R., Dickinson, M., et al. 2012, *MNRAS*, 425, 2116  
 Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, *MNRAS*, 453, 1513  
 Damjanov, I., Zahid, H. J., Geller, M. J., Fabricant, D. G., & Hwang, H. S. 2018, *ApJS*, 234, 21  
 Davies, L. J. M., Driver, S. P., Robotham, A. S. G., et al. 2015, *MNRAS*, 447, 1014  
 Davies, L. J. M., Robotham, A. S. G., Driver, S. P., et al. 2018, *MNRAS*, 480, 768  
 de Jong, J. T. A., Kleijn, G. A. V., Erben, T., et al. 2017, *A&A*, 604, A134  
 de Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., et al. 2015, *A&A*, 582, A62  
 Driver, S. P., Bellstedt, S., Robotham, A. S. G., et al. 2022, *MNRAS*  
 Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, *MNRAS*, 413, 971  
 Edge, A., Sutherland, W., Kuijken, K., et al. 2013, *The Messenger*, 154, 32  
 Euclid Collaboration, Guglielmo, V., Saglia, R., et al. 2020, *A&A*, 642, A192  
 Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, *MNRAS*, 467, 1627  
 Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, 398, 1601  
 Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *AJ*, 150, 150  
 Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015a, *MNRAS*, 448, 2987  
 Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2015b, *MNRAS*, 447, 1319  
 Garilli, B., McLure, R., Pentericci, L., et al. 2021, *A&A*, 647, A150  
 Gatti, M., Giannini, G., Bernstein, G. M., et al. 2022, *MNRAS*, 510, 1223  
 Giblin, B., Heymans, C., Asgari, M., et al. 2021, *A&A*, 645, A105  
 Gruen, D. & Brimiouille, F. 2017, *MNRAS*, 468, 769  
 Hartley, W. G., Chang, C., Samani, S., et al. 2020, *MNRAS*, 496, 4769  
 Hasinger, G., Capak, P., Salvato, M., et al. 2018, *ApJ*, 858, 77  
 Heymans, C., Grocutt, E., Heavens, A., et al. 2013, *MNRAS*, 432, 2433  
 Heymans, C., Tröster, T., Asgari, M., et al. 2021, *A&A*, 646, A140  
 Hikage, C., Oguri, M., Hamana, T., et al. 2019, *PASJ*, 71, 43  
 Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, *A&A*, 633, A69  
 Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, *A&A*, 647, A124  
 Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, *MNRAS*, 465, 1454  
 Hirata, C. M. & Seljak, U. 2004, *Phys. Rev. D*, 70, 063526  
 Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, *MNRAS*, 447, 1724  
 Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, *J. Cosmology Astropart. Phys.*, 2012, 027  
 Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90  
 Hunter, D., Takada, M., Bernstein, G., & Jain, B. 2006, *MNRAS*, 366, 101  
 Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111  
 Jarvis, M. 2015, *TreeCorr: Two-point correlation functions*  
 Jarvis, M., Bernstein, G., & Jain, B. 2004, *MNRAS*, 352, 338  
 Joachimi, B., Lin, C. A., Asgari, M., et al. 2021, *A&A*, 646, A129  
 Joudaki, S., Blake, C., Heymans, C., et al. 2017, *MNRAS*, 465, 2033  
 Joudaki, S., Hildebrandt, H., Traykova, D., et al. 2020, *A&A*, 638, L1  
 Kafle, P. R., Robotham, A. S. G., Driver, S. P., et al. 2018, *MNRAS*, 479, 3746  
 Kaiser, N. 1992, *ApJ*, 388, 272  
 Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, *A&A*, 624, A92  
 Kilbinger, M. 2015, *Reports on Progress in Physics*, 78, 086901  
 Kohonen, T. 1982, *Biological Cybernetics*, 43, 59  
 Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, *A&A*, 625, A2  
 Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, *MNRAS*, 454, 3500  
 Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24  
 Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *arXiv e-prints*, arXiv:1110.3193  
 Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, *A&A*, 559, A14  
 Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, *A&A*, 576, A79  
 Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, *A&A*, 439, 845  
 Lewis, A. 2019, *arXiv e-prints*, arXiv:1910.13970  
 Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, 538, 473  
 Lidman, C., Tucker, B. E., Davis, T. M., et al. 2020, *MNRAS*, 496, 19

- Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, *ApJS*, 184, 218
- Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, *MNRAS*, 390, 118
- Ma, Z., Hu, W., & Huterer, D. 2006, *ApJ*, 636, 21
- MacCrann, N., Zuntz, J., Bridle, S., Jain, B., & Becker, M. R. 2015, *MNRAS*, 451, 2877
- Masters, D. C., Capak, P. L., Stern, D., et al. 2016, in *American Astronomical Society Meeting Abstracts*, Vol. 227, American Astronomical Society Meeting Abstracts #227, 139.14
- 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
- Matthews, D. J. & Newman, J. A. 2010, *ApJ*, 721, 456
- Mead, A. J., Heymans, C., Lombriser, L., et al. 2016, *MNRAS*, 459, 1468
- Mead, A. J., Peacock, J. A., Heymans, C., Joudaki, S., & Heavens, A. F. 2015, *MNRAS*, 454, 1958
- Ménard, B., Scranton, R., Schmidt, S., et al. 2013, *arXiv e-prints*, arXiv:1303.4722
- Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, *MNRAS*, 382, 315
- Myles, J., Alarcon, A., Amon, A., et al. 2021, *MNRAS*, 505, 4249
- Nelder, J. A. & Mead, R. 1965, *The Computer Journal*, 7, 308
- Newman, J. A. 2008, *ApJ*, 684, 88
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, 208, 5
- Padilla, C., Castander, F. J., Alarcón, A., et al. 2019, *AJ*, 157, 246
- Peacock, J. A., Schneider, P., Efstathiou, G., et al. 2006, *ESA-ESO Working Group on "Fundamental Cosmology"*, *ESA-ESO Working Group on "Fundamental Cosmology"*
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *A&A*, 571, A16
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6
- Popesso, P., Dickinson, M., Nonino, M., et al. 2009, *A&A*, 494, 443
- Refregier, A. 2003, *ARA&A*, 41, 645
- Sánchez, C. & Bernstein, G. M. 2019, *MNRAS*, 483, 2801
- Schmidt, S. J., Ménard, B., Scranton, R., Morrison, C., & McBride, C. K. 2013, *MNRAS*, 431, 3307
- Schneider, P., Eifler, T., & Krause, E. 2010, *A&A*, 520, A116
- Scodreggio, M., Guzzo, L., Garilli, B., et al. 2018, *A&A*, 609, A84
- Secco, L. F., Samuroff, S., Krause, E., et al. 2022, *Phys. Rev. D*, 105, 023515
- Silverman, J. D., Kashino, D., Sanders, D., et al. 2015, *ApJS*, 220, 12
- Stanford, S. A., Masters, D., Darvish, B., et al. 2021, *ApJS*, 256, 9
- Tröster, T., Asgari, M., Blake, C., et al. 2021, *A&A*, 649, A88
- van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, *A&A*, 642, A200
- van der Wel, A., Noeske, K., Bezanson, R., et al. 2016, *ApJS*, 223, 29
- Venemans, B. P., Verdoes Kleijn, G. A., Mwebaze, J., et al. 2015, *MNRAS*, 453, 2259
- 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
- Zuntz, J., Paterno, M., Jennings, E., et al. 2015, *Astronomy and Computing*, 12, 45

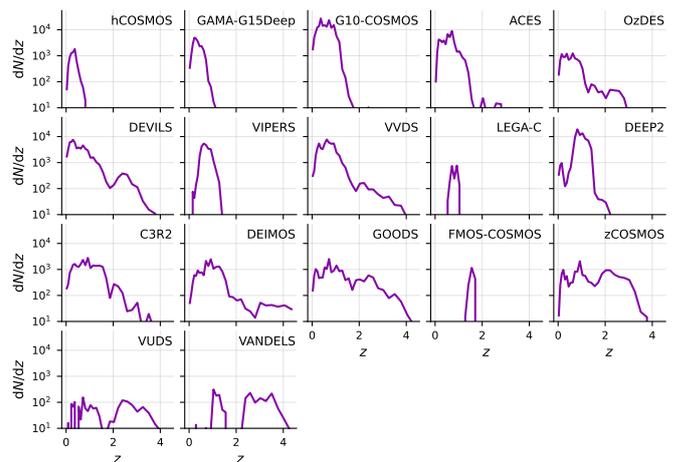
## Appendix A: Updated spectroscopic compilation

Here we present the details of our extended and revised spectroscopic compilation discussed in Sect. 2.2. It includes redshifts from surveys covering a number of deep extragalactic fields for which we have 9-band VST+VISTA photometry of similar quality as the main KiDS+VIKING data (‘KiDZ’). In this paper we use 6 such fields shown in Fig. A.1; three of them (COSMOS, CDF-S and VVDS-2h) are covered by a number of partly overlapping surveys and we merged them into our redshift calibration sample. In doing this, we translate the input redshift quality flags or assessments into our flag NQ, with  $NQ \geq 4$  indicating the most secure spectroscopic redshifts (confidence 99% or more), while  $3 \leq NQ < 4$  indicate secure, but lower-confidence redshifts. The particular surveys are listed in Table A.1, in the ascending order of their mean redshift determined after removing multiples and cross-matching with our imaging. The details of the selection and quality assignment for the particular samples are as follows:

- hCOSMOS (Damjanov et al. 2018): we use all the galaxies from the published dataset and assign them redshift quality  $NQ = 4$ ;
- GAMA-G15Deep (Kafle et al. 2018; Driver et al. 2022): we select galaxies with input redshift quality  $Z\_QUAL \geq 3$  and with redshifts  $z > 0.001$  to avoid stellar contamination. We assign  $NQ = Z\_QUAL$ ;
- G10-COSMOS (Davies et al. 2015): we use  $Z\_BEST$  as the redshift value and follow the recommended selection for galaxy spectroscopic redshifts, that is  $Z\_BEST > 0.0001$ ,  $Z\_USE < 3$  and  $STAR\_GALAXY\_CLASS = 0$ . As the redshifts provided in that compilation do not always have a quality flag that could be easily translated into our NQ, we assign all these galaxies a specific flag  $NQ = 3.5$ ;
- ACES (Cooper et al. 2012): we select redshifts with  $Z\_QUALITY \geq 3$  and with redshift errors below 1%. We assign  $NQ = Z\_QUALITY$ ;
- OzDES (Lidman et al. 2020): we use two patches partly overlapping with KiDZ imaging (around CDF-S and around VVDS-2h). We select redshifts  $z > 0.002$  (stellar contamination removal) and require quality  $qop \in \{3, 4\}$ . We assign  $NQ = qop$ ;
- DEVILS (Davies et al. 2018): we select spectroscopic redshifts only ( $zBestType = spec$ ) and require the flags  $starFlag = 0$ ,  $mask = 0$ , and  $artefactFlag = 0$ . As the DEVILS sample is a compilation including both own redshifts and external ones, we assign  $NQ = 4$  if  $zBestSource = DEVILS$  (redshift obtained from DEVILS observations) and  $NQ = 3$  (external redshifts);
- VIPERS (Scodreggio et al. 2018): we require sources with the flag  $2 \leq zflg < 10$  or  $22 \leq zflg < 30$ . We assign  $NQ = 4$  if  $3 \leq zflg < 5$  or  $23 \leq zflg < 25$ , and  $NQ = 3$  otherwise;
- VVDS (Le Fèvre et al. 2005, 2013): we join the WIDE, DEEP and UDEEP sub-samples, and use only sources with  $ZFLAGS \in \{3, 4, 23, 24\}$ . We assign  $NQ = 4$  to these redshifts;
- LEGA-C (van der Wel et al. 2016): we select sources with the flag  $f\_use = 1$  and assign  $NQ = 4$  to them;
- DEEP2 (Newman et al. 2013): as in the previous KiDS papers (Hildebrandt et al. 2017, 2020, 2021), we use sources from two equatorial fields (0226 & 2330), require sources with  $ZQUALITY \geq 3$ , and with redshift errors smaller than 1%. We assign  $NQ = Z\_QUALITY$ ;
- C3R2: we combine four public datasets, DR1 (Masters et al. 2017), DR2 (Masters et al. 2019), DR3 (Stanford et al. 2021), and KMOS (Euclid Collaboration et al. 2020) and exclude



**Fig. A.1.** On-sky distribution of spectroscopic sources matched against the six KiDZ fields.



**Fig. A.2.** Redshift distributions of the spectroscopic samples that enter the spectroscopic compilation, ordered by mean redshift (Table A.1). The distributions are calculated after removing duplicates between overlapping catalogues and matching with the KiDS imaging.

the Northern EGS field. We take sources with  $QFLAG \geq 3$  and assign  $NQ = QFLAG$ ;

- DEIMOS (Hasinger et al. 2018): we require the quality flag  $Q = 2$ , while for assigning NQ we use the other flag provided, namely  $NQ = 4$  for  $Q_f \in \{4, 14\}$ ,  $NQ = 3$  otherwise;
- GOODS/CDF-S: we use a public ESO compilation of spectroscopy in the CDF-S field<sup>11</sup> (Popesso et al. 2009; Balestra et al. 2010). From the compilation, we choose secure redshifts (assigning  $NQ = 4$  to them) and likely redshifts ( $NQ =$

<sup>11</sup> Available from <https://www.eso.org/sci/activities/garching/projects/goods/MasterSpectroscopy.html>

**Table A.1.** Listing of the spectroscopic samples that enter the spectroscopic compilation summarising the sample sizes and mean redshifts. The values are calculated after removing duplicates between overlapping catalogues and matching with the KiDS imaging.

Sample	Count	$\langle z \rangle$	Reference
hCOSMOS	503	0.308	<a href="#">Damjanov et al. (2018)</a>
GAMA-G15Deep	1 840	0.357	<a href="#">Kafle et al. (2018)</a> ; <a href="#">Driver et al. (2022)</a>
G10-COSMOS	14 849	0.586	<a href="#">Davies et al. (2015)</a>
ACES	4 233	0.593	<a href="#">Cooper et al. (2012)</a>
OzDES	930	0.638	<a href="#">Lidman et al. (2020)</a>
DEVILS	5 222	0.682	<a href="#">Davies et al. (2018)</a>
VIPERS	2 436	0.718	<a href="#">Scodreggio et al. (2018)</a>
VVDS	5 190	0.737	<a href="#">Le Fèvre et al. (2005, 2013)</a>
LEGA-C	216	0.818	<a href="#">van der Wel et al. (2016)</a>
DEEP2	8 564	0.962	<a href="#">Newman et al. (2013)</a>
C3R2	2 512	0.980	<a href="#">Masters et al. (2017, 2019)</a> ; <a href="#">Euclid Collaboration et al. (2020)</a> ; <a href="#">Stanford et al. (2021)</a>
DEIMOS	1 729	1.045	<a href="#">Hasinger et al. (2018)</a>
GOODS	1 999	1.292	<i>ESO compilation of GOODS/CDF-S spectroscopy</i>
FMOS-COSMOS	272	1.566	<a href="#">Silverman et al. (2015)</a>
zCOSMOS	1 875	1.602	<i>private comm. from M. Salvato</i>
VUDS	205	1.974	<a href="#">Le Fèvre et al. (2015)</a>
VANDELS	336	2.504	<a href="#">Garilli et al. (2021)</a>

3) following the recommendations in the dataset description<sup>12</sup>;

- FMOS-COSMOS ([Silverman et al. 2015](#)): we select sources with the quality flag  $q_z \geq 2$  and assign  $NQ = 4$  if  $q_z = 4$  or  $NQ = 3$  otherwise;
- zCOSMOS: we use this name for a proprietary compilation of various spectroscopic surveys in the COSMOS field, kindly provided to us by Mara Salvato, updated as of 01 September 2017. That dataset includes some of the surveys already included in our compilation, but it also provides redshifts from various other campaigns. We use the provided quality flag and select sources meeting the criteria  $3 \leq Q_f \leq 5$  or  $13 \leq Q_f \leq 15$  or  $23 \leq Q_f \leq 25$  or  $Q_f \in \{6, 10\}$ , do not use sources from lower-confidence determinations (e.g. grism), and limit the redshifts to  $z_{\text{spec}} > 0.002$  to avoid stellar contamination. We assign redshift quality as  $NQ = \min((Q_f \bmod 10), 4)$ ;
- VUDS ([Le Fèvre et al. 2015](#)): we use sources with redshift flag  $z\text{flags}$  ending with  $\{3, 4, 9\}$  (reliability  $\geq 80\%$ ) and assign  $NQ = 4$  if  $3 \leq z\text{flags} < 5$  or  $13 \leq z\text{flags} < 25$ , otherwise  $NQ = 3$ ;
- VANDELS ([Garilli et al. 2021](#)): we select sources for which  $(z\text{flg} \bmod 10) \in \{2, 3, 4\}$  and assign  $NQ = 4$  if  $(z\text{flg} \bmod 10) \in \{3, 4\}$  or  $NQ = 3$  otherwise. The reassignment of the quality flags is motivated by the reportedly high redshift confidence of objects with flag values of 2 and 3.

When joining the above samples into one dataset, we remove duplicates both from overlapping surveys as well as within the individual ones. In the former case, if for a given source there are redshifts from different surveys available, we assign the most reliable measurement based on a specific ‘hierarchy’. Namely, we join the catalogues by cross-matching objects within  $1''$  radius and apply the following order of preference:

- COSMOS field: G10-COSMOS > DEIMOS > hCOSMOS > VVDS > Lega-C > FMOS > VUDS > C3R2 > DEVILS > zCOSMOS;
- CDF-S field: ACES > VANDELS > VVDS > VUDS > GOODS/CDF-S > DEVILS > OzDES;

- VVDS-2h field: VIPERS > VVDS > C3R2 > DEVILS > OzDES.

For objects with multiple spectroscopic measurements within a particular survey, we either take the redshift with the highest quality flag or, if various entries for the same source have the same quality flag and the reported redshifts differ by no more than 0.005, we calculate the average. If the reported redshifts have the same quality flag but differ by more than 0.005, we exclude such a source from the compilation.

## Appendix B: SOM quality control

The original quality control (Eq. 2) was calibrated for the spectroscopic calibration data of [Hildebrandt et al. \(2021\)](#) with  $\sigma_{\text{mad}} = n\text{MAD}(\langle z_{\text{cal}} \rangle - \langle z_B \rangle) \approx 0.12$ . Since this criterion may not be optimal for our new calibration dataset, we would in principle need to recalibrate the SOM cell rejection threshold to obtain the optimal trade-off between redshift bias and the size of the gold sample. It is, however, very challenging to simulate the photo- $z$  errors and success rates of our new calibration dataset and therefore we choose to keep the threshold fixed for all gold sets.

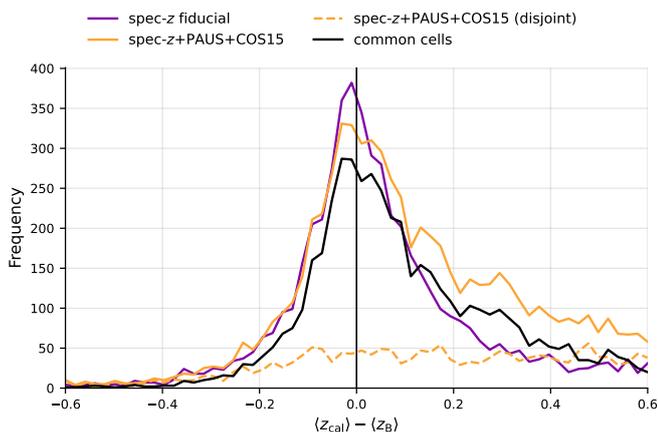
Nevertheless, the addition of COSMOS2015 to the calibration data has a significant impact on the distribution of  $\langle z_{\text{cal}} \rangle - \langle z_B \rangle$  for the fully trained cells of the SOM (Fig. B.1). While the distribution is almost symmetrical for the case of *spec- $z$  fiducial*, it is skewed to positive values for *spec- $z$ +PAUS+COS15*. The effect is more distinct for those cells that are occupied exclusively by the COSMOS2015 data than for those that exist in both gold samples (compare the orange dotted and black lines in Fig. B.1). This is a clear indication that much of the additional calibration data are faint objects for which the KiDS 9-band photo- $z$  catastrophically underestimates the redshift obtained from the much deeper COSMOS2015 data. This effect is responsible for the reduced representation fraction in the first tomographic bin, as discussed in Sect. 5.1.1.

<sup>12</sup> [https://www.eso.org/sci/activities/garching/projects/goods/MASTERCAT\\_v3.0.dat](https://www.eso.org/sci/activities/garching/projects/goods/MASTERCAT_v3.0.dat)

**Table C.1.** Summary of the main cosmological parameter constraints (best fit and 68th-percentile PJ-HPD) from COSEBIs for [Asgari et al. \(2021\)](#) and our reanalysis with and without the corrected  $m$ -bias values.

Sample	$\chi^2$	$A_{IA}$	$\Omega_m$	$\sigma_8$	$S_8$	$\Sigma_8$
Asgari et al. (2021)	82.2	$0.264^{+0.424}_{-0.336}$	$0.246^{+0.101}_{-0.060}$	$0.838^{+0.140}_{-0.141}$	$0.759^{+0.024}_{-0.021}$	$0.751^{+0.024}_{-0.016}$
Asgari reanalysis	83.8	$0.243^{+0.370}_{-0.424}$	$0.240^{+0.101}_{-0.059}$	$0.846^{+0.146}_{-0.146}$	$0.758^{+0.029}_{-0.018}$	$0.749^{+0.025}_{-0.013}$
Asgari new $m$ -bias	83.2	$0.233^{+0.406}_{-0.388}$	$0.242^{+0.104}_{-0.060}$	$0.844^{+0.114}_{-0.165}$	$0.757^{+0.022}_{-0.023}$	$0.749^{+0.024}_{-0.017}$

**Notes.** Additionally shown are the  $\chi^2$  values (for 70.5 effective degrees of freedom) and  $\Sigma_8 = \sigma_8(\Omega_m/0.3)^\alpha$  calculated for  $\alpha_{fid} = 0.55$ , determined from a fit to the posterior samples of the fiducial chain.



**Fig. B.1.** Comparison of the mean calibration sample redshift and KiDS 9-band photo- $z$  in each SOM cell within the fiducial  $\pm 5\sigma_{mad}$  interval. The coloured solid lines represent the fiducial (purple) and the full (orange) gold sample. The black line is the subset of SOM cells that belong to both gold samples, whereas the dashed orange line represents those cells that are occupied only by data from COSMOS2015.

### Appendix C: Differences from the original KiDS-1000 analysis

The major difference between the analysis of the original KiDS-1000 data and this work is that our redshift calibration strategy focuses exclusively on improving the SOM redshifts. We omit revisiting the KiDS clustering redshifts, which [Hildebrandt et al. \(2021\)](#) used as a validation for the fiducial SOM redshift distributions. We note, however, that the additional calibration data is only beneficial for the SOM calibration, since such inhomogeneous data, in particular photometric redshifts, cannot be applied easily in clustering redshifts.

Furthermore there are small differences in the cosmological analysis compared to [A21](#), which are mostly related to small differences in the COSEBI data vector. We find that these differences originate from using a newer version of TREECORR, however it has no impact on our cosmological constraints which agree within the expected statistical variance of the sampler. Differences in the measured COSEBIs also propagate to the covariance matrix, resulting in sub-percent differences between our covariance and the one used by [A21](#). Finally, we assume 99 % correlation between  $m$ -bias values per tomographic bin instead of 100 % (perfect correlation) as in [A21](#). The effect of this choice is negligible compared to those introduced by the variations in the COSEBI data vector.

We also test whether the correction of the  $m$ -bias (Table 2) has a significant impact on the cosmological constraints in [A21](#) by analysing the same data (‘Asgari new  $m$ -bias’) with other-

wise identical parameters (except for the differences summarised above). Our results show that the recovered primary cosmological parameters, in particular  $S_8 = 0.757^{+0.022}_{-0.023}$ , are almost identical with those of our reanalysis of the KiDS-1000 data, performed with the old  $m$ -bias values (Table C.1). Given that the uncertainties of the  $m$ -biases are unchanged and the central value changes by less than  $0.5\sigma$  in any of the tomographic bins, these results are as expected and have no further implication for the conclusions of [A21](#).

### Appendix D: Marginal parameter constraints

Table D.1 lists an alternative representation of the parameter constraints of Table 3, showing the maximum of the marginal distribution and the associated 68th-percentile instead of the best fit and the 68th-percentile PJ-HPD.

### Appendix E: Goodness-of-fit

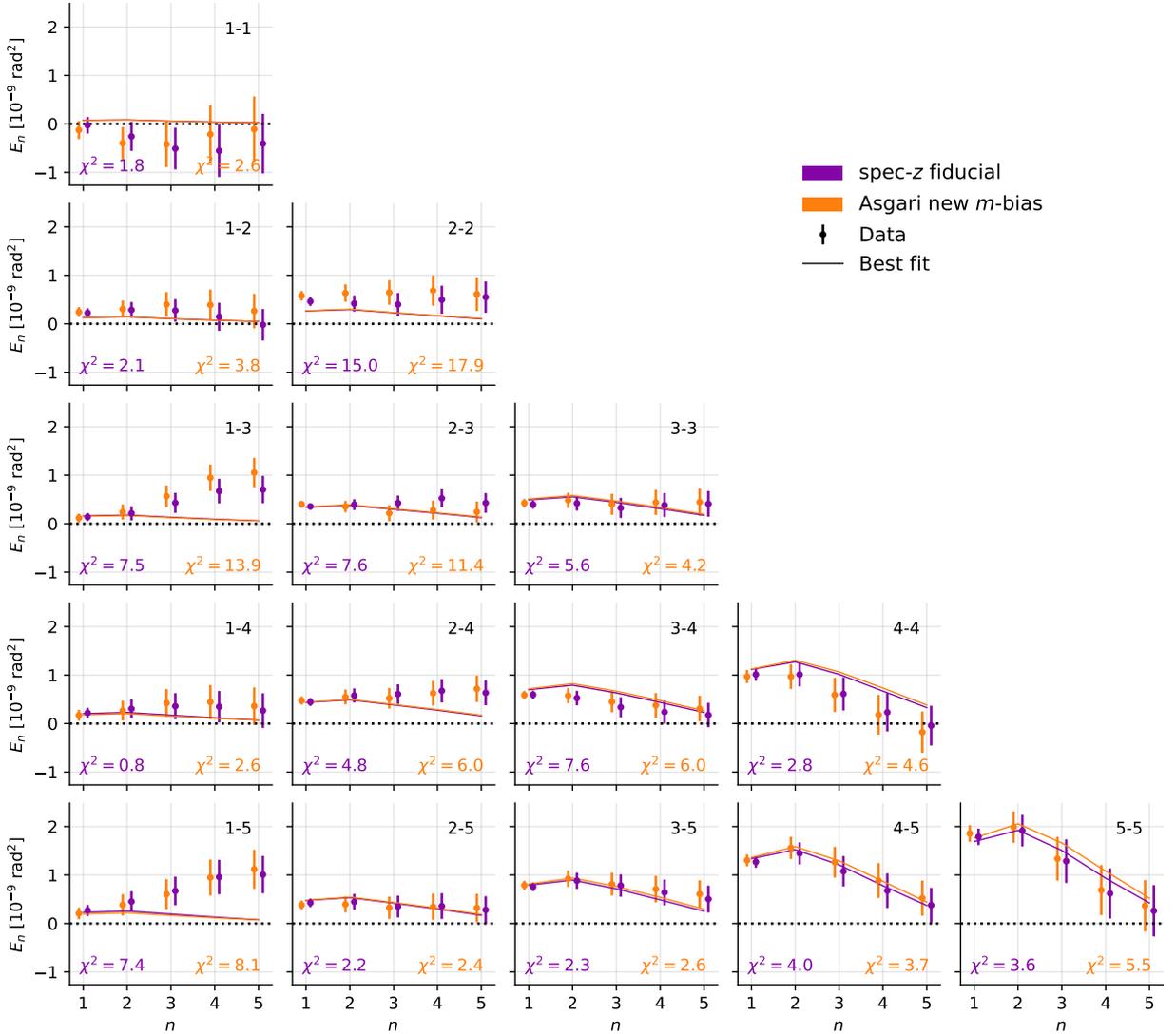
We find that the cosmological model fits to the new gold samples achieve a better goodness-of-fit ( $\chi^2$  between 62 and 74, Table 3) than those to the original KiDS-1000 gold sample with  $\chi^2 \approx 83$  as reference. Since our methodology is identical in both cases, these differences stem from changes of the data vector that propagate also into the covariance matrix. We compare in particular the COSEBI data vectors with their respective best fit models for *spec-z fiducial* and *Asgari new  $m$ -bias* (App. C) in Fig. E.1. Although a comparison by eye of model and data is difficult due to the strong correlation between the different COSEBI modes, the most significant changes in the data vector occur between modes that involve bin 1 and 2. In contrast to this the fitted models are very close in the first two bins and begin to deviate with increasing redshift.

These changes propagate to the  $\chi^2$ , which becomes apparent when summing the individual  $\chi^2$ -values in each panel of Fig. E.1 for the two left-most columns. While the sum over the remaining three columns (dominated by bin 3, 4, and 5) is almost identical for *spec-z fiducial* and *Asgari new  $m$ -bias*, the first two columns fully account for the smaller goodness-of-fit of  $\Delta\chi^2 \approx 20$ . Since changing the  $m$ -bias seems to have a rather small impact on the  $\chi^2$  (see Table C.1), this behaviour is most likely linked to the changes in the gold sample selection and redshift distributions, which are most significant for the first two bins. [A21](#) report tensions between parameter constraints obtained from using only the second bin (including its cross-correlations) and the remainder of the data vector. A reduction of the  $\chi^2$ -value alone does not allow any conclusions on whether the improved calibration sample has an impact on the internal consistency of the cosmological analysis.

**Table D.1.** Summary of the main cosmological parameter constraints (maximum of the marginal distribution and the associated 68th-percentile) from COSEBIs for all gold samples and the comparison to [Asgari et al. \(2021\)](#) and Planck legacy (TT, TE, EE + lowE).

Sample	$A_{IA}$	$\Omega_m$	$\sigma_8$	$S_8$	$\Sigma_8$
spec-z fiducial	$0.394^{+0.328}_{-0.387}$	$0.278^{+0.097}_{-0.084}$	$0.734^{+0.134}_{-0.129}$	$0.745^{+0.020}_{-0.026}$	$0.740^{+0.019}_{-0.019}$
spec-z+PAUS	$0.108^{+0.347}_{-0.441}$	$0.238^{+0.101}_{-0.066}$	$0.800^{+0.118}_{-0.159}$	$0.745^{+0.019}_{-0.025}$	$0.736^{+0.020}_{-0.019}$
spec-z+PAUS+COS15	$0.440^{+0.328}_{-0.465}$	$0.206^{+0.109}_{-0.057}$	$0.843^{+0.152}_{-0.163}$	$0.750^{+0.019}_{-0.024}$	$0.739^{+0.018}_{-0.021}$
spec-z $nQ \geq 4$	$0.088^{+0.395}_{-0.378}$	$0.203^{+0.080}_{-0.050}$	$0.877^{+0.134}_{-0.138}$	$0.758^{+0.019}_{-0.022}$	$0.746^{+0.019}_{-0.019}$
only PAUS	$0.078^{+0.367}_{-0.488}$	$0.210^{+0.112}_{-0.057}$	$0.879^{+0.112}_{-0.202}$	$0.753^{+0.021}_{-0.025}$	$0.740^{+0.023}_{-0.017}$
only COS15	$0.345^{+0.324}_{-0.452}$	$0.245^{+0.137}_{-0.072}$	$0.672^{+0.204}_{-0.100}$	$0.736^{+0.024}_{-0.023}$	$0.733^{+0.020}_{-0.019}$
only PAUS+COS15	$0.392^{+0.349}_{-0.476}$	$0.264^{+0.111}_{-0.099}$	$0.743^{+0.149}_{-0.166}$	$0.737^{+0.021}_{-0.026}$	$0.730^{+0.020}_{-0.020}$
Asgari et al. (2021)	$0.384^{+0.360}_{-0.408}$	$0.254^{+0.088}_{-0.076}$	$0.773^{+0.145}_{-0.125}$	$0.758^{+0.017}_{-0.026}$	$0.753^{+0.015}_{-0.023}$
Planck legacy	—	$0.316^{+0.009}_{-0.008}$	$0.813^{+0.007}_{-0.008}$	$0.836^{+0.016}_{-0.017}$	$0.838^{+0.017}_{-0.018}$

**Notes.** Additionally shown is  $\Sigma_8 = \sigma_8(\Omega_m/0.3)^\alpha$  calculated for  $\alpha_{fid} = 0.55$ , determined from a fit to the posterior samples of the fiducial chain.



**Fig. E.1.** Comparison of the COSEBI data vector and the best fit model for *spec-z fiducial* and the reanalysis of [Asgari et al. \(2021\)](#) with corrected *m*-bias. The triangular arrangement of all 15 combinations of tomographic bins are labelled with their individual  $\chi^2$ -values. The COSEBIs modes are significantly correlated, therefore their goodness-of-fit cannot be estimated by eye.