



UNIVERSIDADE ESTADUAL DE CAMPINAS  
SISTEMA DE BIBLIOTECAS DA UNICAMP  
REPOSITÓRIO DA PRODUÇÃO CIENTÍFICA E INTELLECTUAL DA UNICAMP

**Versão do arquivo anexado / Version of attached file:**

Versão do Editor / Published Version

**Mais informações no site da editora / Further information on publisher's website:**

[https://www.aanda.org/articles/aa/full\\_html/2024/07/aa48615-23/aa48615-23.html](https://www.aanda.org/articles/aa/full_html/2024/07/aa48615-23/aa48615-23.html)

**DOI: 10.1051/0004-6361/202348615**

**Direitos autorais / Publisher's copyright statement:**

©2024 by EDP Sciences. All rights reserved.

DIRETORIA DE TRATAMENTO DA INFORMAÇÃO

Cidade Universitária Zeferino Vaz Barão Geraldo

CEP 13083-970 – Campinas SP

Fone: (19) 3521-6493

<http://www.repositorio.unicamp.br>

# The SRG/eROSITA All-Sky Survey

## Dark Energy Survey year 3 weak gravitational lensing by eRASS1 selected galaxy clusters

S. Grandis<sup>1</sup>, V. Ghirardini<sup>2</sup>, S. Bocquet<sup>3</sup>, C. Garrel<sup>2</sup>, J. J. Mohr<sup>3,2</sup>, A. Liu<sup>2</sup>, M. Kluge<sup>2</sup>, L. Kimmig<sup>3</sup>, T. H. Reiprich<sup>4</sup>, A. Alarcon<sup>5</sup>, A. Amon<sup>6,7</sup>, E. Artis<sup>2</sup>, Y. E. Bahar<sup>2</sup>, F. Balzer<sup>2,3</sup>, K. Bechtol<sup>8</sup>, M. R. Becker<sup>9</sup>, G. Bernstein<sup>10</sup>, E. Bulbul<sup>2</sup>, A. Campos<sup>11</sup>, A. Carnero Rosell<sup>12,13,14</sup>, M. Carrasco Kind<sup>15,16</sup>, R. Cawthon<sup>17</sup>, C. Chang<sup>18,19</sup>, R. Chen<sup>20</sup>, I. Chiu<sup>21</sup>, A. Choi<sup>22</sup>, N. Clerc<sup>23</sup>, J. Comparat<sup>2</sup>, J. Cordero<sup>24</sup>, C. Davis<sup>25</sup>, J. Derose<sup>26</sup>, H. T. Diehl<sup>27</sup>, S. Dodelson<sup>11,28</sup>, C. Doux<sup>14,29</sup>, A. Drlica-Wagner<sup>18,19,27</sup>, K. Eckert<sup>10</sup>, J. Elvin-Poole<sup>30</sup>, S. Everett<sup>31</sup>, A. Ferte<sup>32</sup>, M. Gatti<sup>10</sup>, G. Giannini<sup>33</sup>, P. Giles<sup>34</sup>, D. Gruen<sup>3</sup>, R. A. Gruendl<sup>15,16</sup>, I. Harrison<sup>35</sup>, W. G. Hartley<sup>36</sup>, K. Herner<sup>27</sup>, E. M. Huff<sup>31</sup>, F. Kleinebreil<sup>1</sup>, N. Kuropatkin<sup>27</sup>, P. F. Leget<sup>37</sup>, N. Maccrann<sup>38</sup>, J. McCullough<sup>37</sup>, A. Merloni<sup>2</sup>, J. Myles<sup>3,37,39</sup>, K. Nandra<sup>2</sup>, A. Navarro-Alsina<sup>40</sup>, N. Okabe<sup>41,42,43</sup>, F. Pacaud<sup>4</sup>, S. Pandey<sup>10</sup>, J. Prat<sup>18,19</sup>, P. Predehl<sup>2</sup>, M. Ramos<sup>2</sup>, M. Raveri<sup>44</sup>, R. P. Rollins<sup>24</sup>, A. Roodman<sup>32,37</sup>, A. J. Ross<sup>45</sup>, E. S. Rykoff<sup>32,37</sup>, C. Sanchez<sup>46</sup>, J. Sanders<sup>2</sup>, T. Schrabback<sup>1</sup>, L. F. Secco<sup>18,19</sup>, R. Seppi<sup>2</sup>, I. Sevilla-Noarbe<sup>46</sup>, E. Sheldon<sup>47</sup>, T. Shin<sup>48</sup>, M. Troxel<sup>20</sup>, I. Tutusaus<sup>49,50</sup>, T. N. Varga<sup>2,3</sup>, H. Wu<sup>51</sup>, B. Yanny<sup>31</sup>, B. Yin<sup>52</sup>, X. Zhang<sup>2</sup>, Y. Zhang<sup>53</sup>, O. Alves<sup>53</sup>, S. Bhargava<sup>34</sup>, D. Brooks<sup>61</sup>, D. L. Burke<sup>32,37</sup>, J. Carretero<sup>33</sup>, M. Costanzi<sup>54,55,56</sup>, L. N. da Costa<sup>13</sup>, M. E. S. Pereira<sup>59</sup>, J. De Vicente<sup>46</sup>, S. Desai<sup>60</sup>, P. Doel<sup>61</sup>, I. Ferrero<sup>62</sup>, B. Flaugher<sup>27</sup>, D. Friedel<sup>15</sup>, J. Frieman<sup>19,27</sup>, J. García-Bellido<sup>43,63</sup>, G. Gutierrez<sup>27</sup>, S. R. Hinton<sup>25</sup>, D. L. Hollowood<sup>57</sup>, K. Honscheid<sup>45,64,65</sup>, D. J. James<sup>65</sup>, N. Jeffrey<sup>61</sup>, O. Lahav<sup>61</sup>, S. Lee<sup>31</sup>, J. L. Marshall<sup>66</sup>, F. Menanteau<sup>15,16</sup>, R. L. C. Ogando<sup>67</sup>, A. Pieres<sup>13,67</sup>, A. A. Plazas Malagón<sup>32,37</sup>, A. K. Romer<sup>34</sup>, E. Sanchez<sup>46</sup>, M. Schubnell<sup>53</sup>, M. Smith<sup>68</sup>, E. Suchyta<sup>58</sup>, M. E. C. Swanson<sup>15</sup>, G. Tarle<sup>53</sup>, N. Weaverdyck<sup>26,53</sup>, and J. Weller<sup>2,3</sup>

(The Dark Energy Survey Collaboration and the eROSITA-DE Consortium)

(Affiliations can be found after the references)

Received 15 November 2023 / Accepted 27 March 2024

### ABSTRACT

**Context.** Number counts of galaxy clusters across redshift are a powerful cosmological probe if a precise and accurate reconstruction of the underlying mass distribution is performed – a challenge called mass calibration. With the advent of wide and deep photometric surveys, weak gravitational lensing (WL) by clusters has become the method of choice for this measurement.

**Aims.** We measured and validated the WL signature in the shape of galaxies observed in the first three years of the Dark Energy Survey (DES Y3) caused by galaxy clusters and groups selected in the first all-sky survey performed by SRG (Spectrum Roentgen Gamma)/eROSITA (eRASS1). These data were then used to determine the scaling between the X-ray photon count rate of the clusters and their halo mass and redshift.

**Methods.** We empirically determined the degree of cluster member contamination in our background source sample. The individual cluster shear profiles were then analyzed with a Bayesian population model that self-consistently accounts for the lens sample selection and contamination and includes marginalization over a host of instrumental and astrophysical systematics. To quantify the accuracy of the mass extraction of that model, we performed mass measurements on mock cluster catalogs with realistic synthetic shear profiles. This allowed us to establish that hydrodynamical modeling uncertainties at low lens redshifts ( $z < 0.6$ ) are the dominant systematic limitation. At high lens redshift, the uncertainties of the sources' photometric redshift calibration dominate.

**Results.** With regard to the X-ray count rate to halo mass relation, we determined its amplitude, its mass trend, the redshift evolution of the mass trend, the deviation from self-similar redshift evolution, and the intrinsic scatter around this relation.

**Conclusions.** The mass calibration analysis performed here sets the stage for a joint analysis with the number counts of eRASS1 clusters to constrain a host of cosmological parameters. We demonstrate that WL mass calibration of galaxy clusters can be performed successfully with source galaxies whose calibration was performed primarily for cosmic shear experiments, opening the way for the cluster cosmological exploitation of future optical and NIR surveys like *Euclid* and LSST.

**Key words.** gravitational lensing: weak – large-scale structure of Universe – X-rays: galaxies: clusters

## 1. Introduction

General relativity describes the gravitational force as the curvature of space-time induced by mass and energy densities. This curvature determines the so-called null geodesics on which photons travel through space-time. As a consequence, the path of the

light from a distant source to the observer is deflected by intervening gravitational potentials, with a deflection angle that is proportional to the path integral of the transverse component of the negative gradient of the potential. Indeed, Dyson et al. (1920) measured the predicted displacement in the positions of distant stars when observed close to the Sun during a solar eclipse. In the

cosmological context, the original position of the sources is not known. In general, the gravitational potential of the intervening lenses is also unknown because the majority of cosmic matter is believed to be invisible, and its density is among the targets of our experiments.

On cosmic scales, gravitational lensing of distant galaxies thus relies on the fact that the differential deflection of neighboring light paths locally induces a magnification and a distortion of the source galaxy image (for a pedagogical introduction see [Schneider 2006](#)). In this regime, large-scale gravitational potentials create spatially coherent, small distortions in the observed shapes of background galaxies, a process called weak gravitational lensing (hereafter WL). A special case is the WL caused by massive gravitationally collapsed, virialized structures, referred to as halos, which dominate the line-of-sight-integrated gravitational potential (for a review see [Umetsu 2020](#)). These objects induce a tangential distortion in the shapes of the background galaxies, that traces the density contrast of the 2D projected density profiles. The 3D density profiles of halos are well understood in simulations and are found to tightly follow a parametric class of functions with two free parameters: the total halo mass and the typical halo scale ([Navarro et al. 1996](#)). Fitting predictions derived from this family of models to the tangential reduced shear profiles thus enables the direct measurement of the halo mass.

The most massive halos host galaxy clusters, whose observed properties display tight correlations among themselves ([Mohr & Evrard 1997](#)), and with the host halos mass ([Angulo et al. 2012](#)), enabling a clean halo selection. For such samples of galaxy clusters, the WL mass information is then used to reconstruct the differential number density of halos as a function of mass and redshift, the halo mass function. This function traces the growth of the structure in the Universe and is therefore a sensitive probe to the total matter density, the accelerated expansion, and gravity itself ([Haiman et al. 2001](#); [Majumdar & Mohr 2004](#); [Allen et al. 2011](#)). Examples of such analyses are [Mantz et al. \(2015\)](#), who used X-ray-selected galaxy clusters with pointed WL observations, and [Bocquet et al. \(2019\)](#), who used galaxy clusters selected at millimeter wavelengths. Both of these analyses resulted in competitive cosmological constraints, which are largely complementary and independent of other cosmological probes (for a recent review, see [Huterer 2023](#)).

Our ability to detect larger numbers of massive halos has recently been transformed by eROSITA, which performed its first all-sky survey from December 2019 to June 2020 ([Predehl et al. 2021](#); [Sunyaev et al. 2021](#)), detecting more than 1 million X-ray sources in the Western Galactic Hemisphere ([Merloni et al. 2024](#)). Among these X-ray sources,  $\sim 12$  k are confirmed as clusters of galaxies through significant extended X-ray emission and associated red galaxy members in the  $12791 \text{ deg}^2$  footprint covered by the DESI Legacy Survey DR9 and DR10 data ([Bulbul et al. 2024](#); [Kluge et al. 2024](#)). In this work, we restricted ourselves to the 5263 clusters selected for the cosmology sample via a higher extent likelihood cut. These clusters have percent-accurate photometric redshifts, in the range of 0.1–0.8, and a well calibrated selection function ([Clerc et al. 2024](#)).

We complement the eROSITA data with the wide photometric Dark Energy Survey year 3 (DES Y3) data ([Sevilla-Noarbe et al. 2021](#)). From these data, we used more than  $10^8$  galaxy shape measurements measured in the *riz* bands by [Gatti et al. \(2021\)](#) in an effective survey area of  $4143 \text{ deg}^2$  of the southern sky. The strength of the WL signal around galaxy

clusters depends not only on their mass but also on the geometrical configuration of the lens and source. To statistically estimate the latter’s distance, we leveraged the exquisite accuracy of the photometric redshift calibration of the DES Y3 sources by [Myles et al. \(2021\)](#).

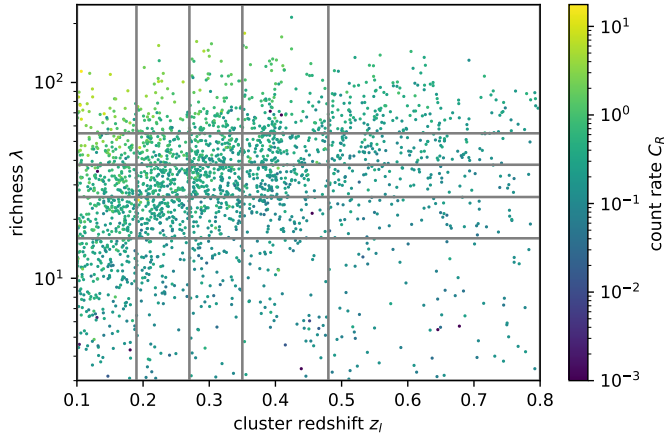
Besides the uncertainty on the shape and photometric redshift measurements for the source galaxies, percent-accurate cluster WL measurements need to account for the cluster mis-centering errors and contamination of the source sample by cluster galaxies (for a complete analysis on DES year 1 data, see [McClintock et al. 2019](#)). Even if all observational systematics are well calibrated and accounted for, the modeling uncertainties of baryonic feedback processes alter the mass distribution of galaxy clusters. This poses a theoretical limit to our ability to calibrate cluster masses via WL ([Grandis et al. 2021a](#)). As argued in that work, setting up many realizations of realistic, synthetic cluster WL mock datasets and analyzing them with the same mass extraction method as the real data allows one to calibrate both the absolute values and – crucially – the uncertainty of the WL mass bias. This approach has already been successfully demonstrated in work by [Chiu et al. \(2022\)](#) in the context of Hyper Supreme Camera (HSC) observations of the eROSITA Final Equatorial-Depth Survey (eFEDS) used for science verification.

This paper is organized as follows: In Sect. 2, we describe the lens and source catalogs, as well as the calibration products used in this analysis. Section 3 then describes the measurement process for the tangential reduced shear and the suite of associated validation and calibration steps. In Sect. 4, we then describe the shear profile model used for the mass measurement, as well as its calibration on realistic synthetic data. These two sections follow in significant part the analysis performed by [Bocquet et al. \(2023\)](#) of the DES Y3 WL signal around South Pole Telescope (SPT) selected clusters. The actual mass calibration is then performed and validated in Sect. 5, and we discuss our results in Sect. 6. Halo masses in this work are reported as spherical over density masses  $M_{\Delta c}$ , with  $\Delta = 500, 200$ . This means that they are defined via the radius  $R_{\Delta c}$  enclosing a sphere of average density  $\Delta$  times the critical density of the Universe at that redshift, that is,  $M_{\Delta c} = \frac{4\pi}{3} \Delta \rho_{\text{crit}}(z) R_{\Delta c}^3$ . We used a flat  $\Lambda$ CDM cosmology with the parameters  $\Omega_M = 0.3$ , and  $h = 0.7$  as the reference cosmology, where not specified otherwise. Throughout this work, we use lens redshift  $z_l$ , and cluster redshift  $z_{cl}$  interchangeably, depending on which is more applicable to the context of the discussion.

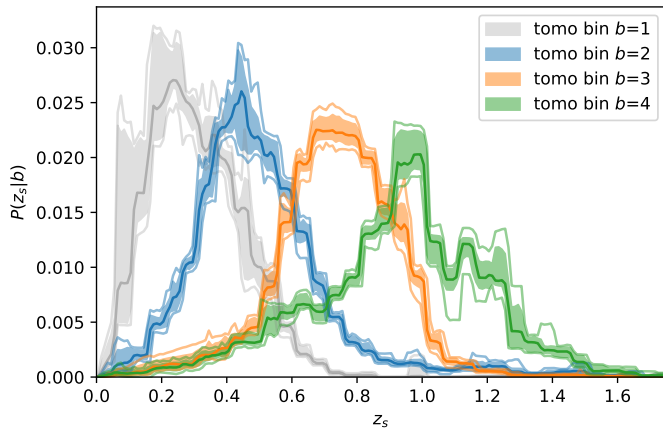
## 2. Data

### 2.1. eRASS1 cluster and group sample

As a lens catalog, we use the cluster catalog acquired from the first eROSITA All-Sky Survey on the Western Galactic Hemisphere, completed on June 11, 2020. The detection properties of the X-ray sources detected in this survey, including extended sources, are discussed in detail in [Merloni et al. \(2024\)](#), [Bulbul et al. \(2024\)](#). We use the cosmology catalog described in detail in [Bulbul et al. \(2024\)](#), comprised of X-ray sources detected as significantly extended, with reliable photometric confirmation and redshift estimation using the DESI Legacy Survey DR10 ([Dey et al. 2019](#)). Specifically, [Kluge et al. \(2024\)](#) run an adapted version of the redMaPPer algorithm ([Rykoff et al. 2014, 2016](#); [Rozo et al. 2015](#)) on the X-ray cluster candidate position, thus measuring their richness and photometric redshift. Of the 5263 clusters and groups in that catalog, 2201



**Fig. 1.** Lens sample composed of the eRASS1 cosmology cluster and group sample with DES WL data, plotted in redshift against richness and with the X-ray count rate color coded. The gray lines mark the edges of the richness–redshift bins used for the calibration and validation of the WL measurement. Cosmology ready WL data products are, however, produced for each cluster individually.



**Fig. 2.** Source redshift distributions of the tomographic redshift bin of the DES Y3 data. We plot the mean of the 1000 realizations (full line), together with the 16–84 percentile (filled region) and the 2.5–97.5 percentile (faded lines). The first tomographic bin is not used, while the others are weighted to ensure a lens-dependent background selection.

have DES Y3 shape and photo- $z$  information (described below), and are therefore used for the WL analysis in this work (see Ghirardini et al. 2024, Fig. 1 for a comparison of the two survey footprints). Their distribution in redshift, richness and X-ray photon count rate is shown in Fig. 1. The bulk of the lens sample lies at low redshift, ideally placed for WL studies. Some objects at low richness have redshifts much larger than expected for an X-ray-selected sample. We account for these contaminants in our mass calibration (see Sect. 5.1.1).

## 2.2. DES WL data

The Dark Energy Survey is an approximately  $5000 \text{ deg}^2$  photometric survey in the optical bands *grizY*, carried out at the 4 m Blanco telescope at the Cerro Tololo Inter-American Observatory (CTIO), Chile, with the Dark Energy Camera (DECam Flaugher et al. 2015). In this analysis, we utilize data from the first three years of observations (DES Y3), covering the full survey footprint.

### 2.2.1. The shape catalog

The DES Y3 shape catalog (Gatti et al. 2021) is built from the  $r, i, z$ -bands using the METACALIBRATION pipeline (Huff & Mandelbaum 2017; Sheldon & Huff 2017). Other DES Y3 publications contain more detailed information about the photometric dataset (Sevilla-Noarbe et al. 2021), the Point-Spread Function modeling (Jarvis et al. 2021), and image simulations (MacCrann et al. 2022). We refer the reader to these works for more information. After applying all source selection cuts, the DES Y3 shear catalog contains about 100 million galaxies over an area of 4143 square degrees. Its effective source density is  $5\text{--}6 \text{ arcmin}^{-2}$ , depending on the exact definition. In Sect. 3, we will derive the exact values of effective source density.

### 2.2.2. Source redshift distributions and shear calibration

Our analysis uses the same selection of lensing source galaxies in tomographic bins as the DES  $3 \times 2$ pt analysis (Dark Energy Survey Collaboration 2022). This selection is defined in Myles et al. (2021). In that work, source redshifts are estimated with Self-Organizing Maps, and the method is thus referred to as SOMp. The final calibration accounts for the (potentially correlated) systematic uncertainties in source redshifts and shear measurements, as determined in the image simulations by MacCrann et al. (2022). For each tomographic source bin  $b$ , the estimated redshift distribution  $P_{\mathcal{H}}(z_s|b)$  is provided, and the systematic uncertainties on this estimate are captured through 1000 realizations (see Fig. 2). These realizations are indexed via the hyperparameter  $\mathcal{H}$ , which takes integer values between 0 and 999 (Cordero et al. 2022). For convenience, these redshift distributions are normalized to varying values  $(1 + m)$ , where the varying  $m$  spans the range of the multiplicative shear bias values. Note that the DES Y3 survey only has minor depth variations over the survey area. The survey averaged quantities, like the source redshift distributions, thus apply to the joint eROSITA-DE – DES Y3 footprint.

We will also use DNF (De Vicente et al. 2016) and BPZ (Benítez 2000) source redshift measurements to constrain the amount of cluster member contamination in our source sample (see Sect. 3.3.1). While BPZ is a Bayesian template fitting code, DNF is based on a nearest-neighbor interpolation on the color-magnitude space of spectroscopic reference samples. As such, BPZ is more robust against the incompleteness of the spectroscopic reference samples, while DNF is more data-driven and avoids the biases that occur from the template choice (see Sevilla-Noarbe et al. 2021, Sect. 6.3, for more details).

## 3. Measurement

In this section, we shall outline (1) the methods used to measure the WL signal around eROSITA selected clusters, (2) the steps undertaken to calibrate this measurement correctly, and (3) the use of this measurement to determine the mass scale of the eROSITA selected clusters and groups. Note that the methods presented in this work draw significantly upon those presented in Bocquet et al. (2023).

### 3.1. WL by massive halos

Assuming that background galaxies are isotropically oriented intrinsically, the coherent tangential distortion induced by the gravitational potential on background galaxy images at a radius



$R$  from the halo center can be obtained, to linear order, by averaging the tangential component of their ellipticities,

$$g_t(R) = \mathcal{R}^{-1} \langle e_t \rangle_R, \quad (1)$$

where  $g_t(R)$  is the reduced tangential reduced shear, and  $\mathcal{R}$  is the average response of the measured ellipticity to a shear (shear response). In practical applications, instrumental effects and noise make it different from 1. It is determined through galaxy shape measurement and validated through image simulations. Below, we discuss how we include this effect in our measurements and how we account for nonlinear responses. While the ellipticity field itself is not Gaussian, our reduced shear estimator is constructed by averaging many sources. The central limit theorem thus ensures that the statistical error of the shear estimate can be directly derived from the effective dispersion  $\sigma_{\text{eff}}$  in the ellipticity of the source galaxies and their effective number  $N_{\text{eff}}$ . Below, we will discuss how we estimate these two quantities.

Given a source galaxy at redshift  $z_s$ , the lens redshift  $z_l$ , and the surface mass density  $\Sigma(\mathbf{x})$  in the lens plane, the tangential reduced shear in any position on the sky  $\boldsymbol{\theta} = \mathbf{x}/D_l$  is given by

$$g_t(\boldsymbol{\theta}) = \frac{\gamma_t(\mathbf{x})}{1 - \kappa(\mathbf{x})} \text{ with } \kappa(\mathbf{x}) = \Sigma_{\text{crit}}^{-1} \Sigma(\mathbf{x}), \quad (2)$$

where  $\gamma_t(\mathbf{x})$  is the tangential component of the shear,  $\kappa(\mathbf{x})$  the convergence,  $D_l$  the angular diameter distance to the lens, and  $\Sigma_{\text{crit}}^{-1}$  the inverse of the critical lensing surface density. The latter expresses the geometrical configuration of the source and lens and is given by

$$\Sigma_{\text{crit,ls}}^{-1} = \frac{4\pi G}{c^2} \frac{D_l}{D_s} \max[0, D_{ls}], \quad (3)$$

where  $G$  is the gravitational constant,  $c$  the speed of light. Furthermore, we use the angular diameter distance between observer and source ( $D_s$ ), and lens and source ( $D_{ls}$ ). Note that for sources in front of the lens,  $D_{ls}$  becomes negative. These sources are not lensed, which we account for by setting their inverse critical lensing surface density to zero. Hence the term  $\max[0, D_{ls}]$ .

In general relativity, the azimuthally averaged tangential shear at projected radius  $R$  is proportional to the density contrast,

$$\gamma_t(R) = \Sigma_{\text{crit}}^{-1} [\langle \Sigma(<R) \rangle - \Sigma(R)], \quad (4)$$

while the orthogonal,  $B$ -mode-like component  $\gamma_x$ , called cross-shear (see below for the exact definition of the decomposition), averages to zero when integrated over closed paths such as radial bins,

$$\gamma_x(R) = 0. \quad (5)$$

We will use the latter as a validation test for our measurement.

### 3.2. Measurement

To perform the measurement, for each lens  $l$ , we query the shape catalog for all sources at a projected physical distance of  $<15 h^{-1}$  Mpc in the reference cosmology ( $h = 0.7$ ,  $\Omega_M = 0.3$ ). For each source lens pair  $(i, l)$ , we compute the position angle  $\varphi$  at which the source is seen from the lens' position.

We then project the ellipticity components  $e_{1,2}$  onto the tangential and cross components as

$$\begin{aligned} e_t &= -e_1 \cos 2\varphi + e_2 \sin 2\varphi \\ e_x &= e_1 \sin 2\varphi + e_2 \cos 2\varphi. \end{aligned} \quad (6)$$

We use the tangential/cross decomposition for the case where the  $e_2$ -component is defined with respect to right ascension, which is the case in the DES Y3 shape catalog<sup>1</sup>. We also record the source weight  $w_i^s$ , and the photometric redshift estimates in the form of the SOMpZ cell  $c_{\text{SOM},i}$  of the source, and the photometric redshift estimates  $\hat{z}_{\text{DNF},i}$  and  $\hat{z}_{\text{BPZ},i}$ . Furthermore, we compute the response  $\mathcal{R}_i$  of every single source by interpolating the response as a function of size with respect to the PSF and signal-to-noise, as presented in the upper right panel of Fig. 4 in Gatti et al. (2021), Sect. 4.3.

To select only sources in the background of each lens, we discard the first tomographic redshift bin and apply a weight to each source depending on the tomographic redshift bin  $b$  it resides in,

$$w^b = \begin{cases} \langle \Sigma_{\text{crit,ls}}^{-1} \rangle_{s \in b} & \text{for } z_l < z_{\text{med},b} \text{ and } b > 1 \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where the average is taken using the mean source redshift distribution of the tomographic bin, as reported by Myles et al. (2021). Our background selection is thus a weighted sum of the tomographic bin selections. This is convenient, as the shear and photo- $z$  calibration (Myles et al. 2021; Cordero et al. 2022; MacCrann et al. 2022) is only valid for the specific selection criteria of the tomographic redshift bins. It can be extended to our analysis by using the same weights as above. Our selection excludes the 2nd tomographic redshift bins for  $z_l > 0.47$ , and the third for  $z_l > 0.74$ .

To validate and calibrate the cluster WL measurement, we define lens richness bins (3, 16, 26, 38, 55, 250), and redshift bins (0.10, 0.19, 0.27, 0.35, 0.48, 0.8), shown as gray lines in Fig. 1, and tuned to result in approximately even occupations of the lenses. In these bins, we measure the following quantities for radial bins whose upper edges are log-equally spaced between 0.2–15  $h^{-1}$  Mpc at the lens redshift in the reference cosmology:

- the raw tangential and cross-shear, with the tangential components shown in Fig. 3,

$$g_{\alpha, \text{raw}} = \frac{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s e_{\alpha,i}}{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s \mathcal{R}_i} \text{ for } \alpha \in (t, x), \quad (8)$$

where  $i \in b$  stands for the source  $i$  in the tomographic bin  $b$  and in the respective lens richness, redshift, and distance bins.

- the effective number of sources

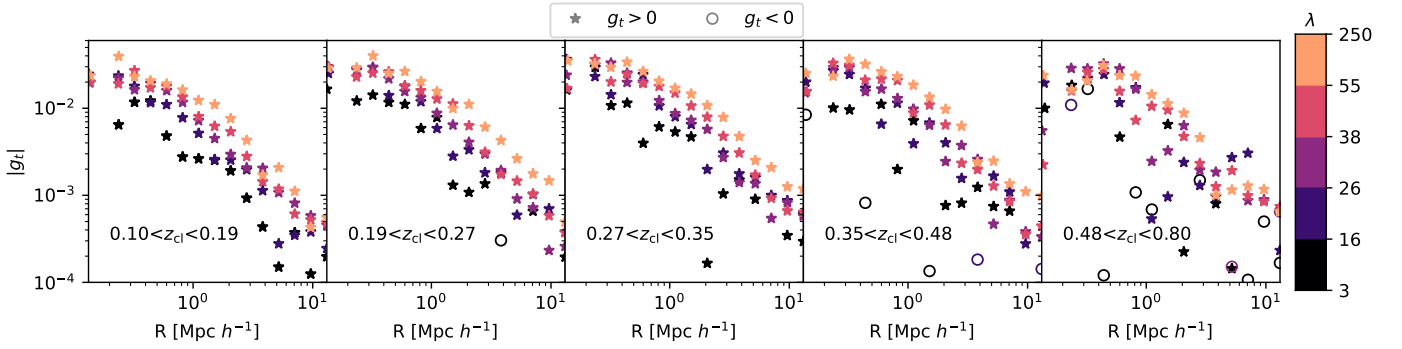
$$N_{\text{eff}} = \frac{\left( \sum_{b=2,3,4} \sum_{i \in b} w^b w_i^s \mathcal{R}_i \right)^2}{\sum_{b=2,3,4} \sum_{i \in b} \left( w^b w_i^s \mathcal{R}_i \right)^2}, \quad (9)$$

- the effective dispersion of source ellipticities

$$\sigma_{\alpha, \text{eff}}^2 = \frac{\sum_{b=2,3,4} \sum_{i \in b} \left( w^b w_i^s \right)^2 e_{\alpha,b,i}^2}{\sum_{b=2,3,4} \sum_{i \in b} \left( w^b w_i^s \mathcal{R}_i \right)^2} \text{ for } \alpha \in (t, x), \text{ and} \quad (10)$$

- the source redshift distribution for fine source redshift bins with edges  $(z_{s-}, z_{s+})$ , reading

<sup>1</sup> The prevalent notation (e.g. Umetsu 2020, Eq. (83)) is valid for  $e_2$  defined for a right-handed, local coordinate system ( $x \propto -RA$ ). It can be recovered by the transformation  $e_2 \mapsto -e_2$ .



**Fig. 3.** Absolute value of the raw tangential reduced shear profile, Eq. (8), in bins of cluster redshift (panels: lowest redshift left, higher redshift right) and of richness (color coding), for our background selection. Stars denote positive values, while empty circles denote negative values. The profiles show a clear increase toward the cluster centers, as well as a trend with richness, in line with our qualitative expectations.

$$\hat{P}_\beta(z_s) = \frac{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s \mathcal{R}_i \mathcal{I}_s(\hat{z}_{\beta,i})}{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s \mathcal{R}_i} \text{ for } \beta \in (\text{BPZ, DNF}), \quad (11)$$

where  $\mathcal{I}_s(\hat{z}_{\beta,i}) = 1$  when  $z_{s-} < \hat{z}_{\beta,i} < z_{s+}$ . Practically speaking, this is a weighted histogram of the photo- $z$  estimates  $\hat{z}_{\beta,i}$ . It is important to note that also the estimates of the effective source density, the shape dispersion and the redshift distribution need to take the shear response  $\mathcal{R}$  into account. For a derivation for Eqs. (9) and (10), we refer the reader to Appendix A, where we present the derivation of the effective number of sources and their effective shape dispersion in the presence of a shear response and un-normalized weights.

### 3.3. Calibration and validation

We shall now discuss several calibration and validation steps we take to ensure the quality of our WL measurements.

#### 3.3.1. Cluster member contamination

As suggested by their name, galaxy clusters are substantial overdensities not only in matter but also in the galaxy field. As a result, an a priori unknown fraction of the cluster members contaminates any source background selection. These cluster member contaminants carry no shear distortion signal from the cluster potential and therefore dilute the shear signal (Hoekstra et al. 2012; Gruen et al. 2014; Dietrich et al. 2019; Varga et al. 2019). This effect is equivalent to source-lens clustering in other WL applications (compare, for instance, with Prat et al. 2022, Sect. III.B). Current simulations are not accurate enough to reconstruct the colors and magnitudes of cluster member galaxies with sufficient fidelity to calibrate this effect directly in simulations. We therefore resorted to empirical calibration methods. We shall outline in the following paragraphs the model for the cluster member contamination, and the two methods we use to fit for it, as well as a comparison of the fit results.

**Cluster member contamination model.** In modeling the cluster member contamination, we adopt an approach that was developed previously in Paulus (2021). We parametrize the radial fraction of cluster contaminants as

$$f_{\text{cl}}(R|\lambda, z) = \frac{A(\lambda, z, R)}{1 + A(\lambda, z, R)} \text{ with} \\ A(\lambda, z_j, R) = e^{A_j} \left( \frac{\lambda}{25} \right)^{B_\lambda} \Sigma_{\text{NFW}}^{\text{norm}} \left( R \middle| r_s = c^{-1} \left( \frac{\lambda}{20} \right)^{1/3} \right), \quad (12)$$

where  $A_j$  is a free amplitude parameter for each cluster redshift bin  $j$  we consider,  $B_\lambda$  describes the richness trend of the cluster member contamination, and  $\Sigma_{\text{NFW}}^{\text{norm}}(R|r_s)$  is a 2d projected NFW profile, normalized to 1 at its scale radius  $r_s$ , which we parameterize via a concentration  $c$  and a scaling with richness.

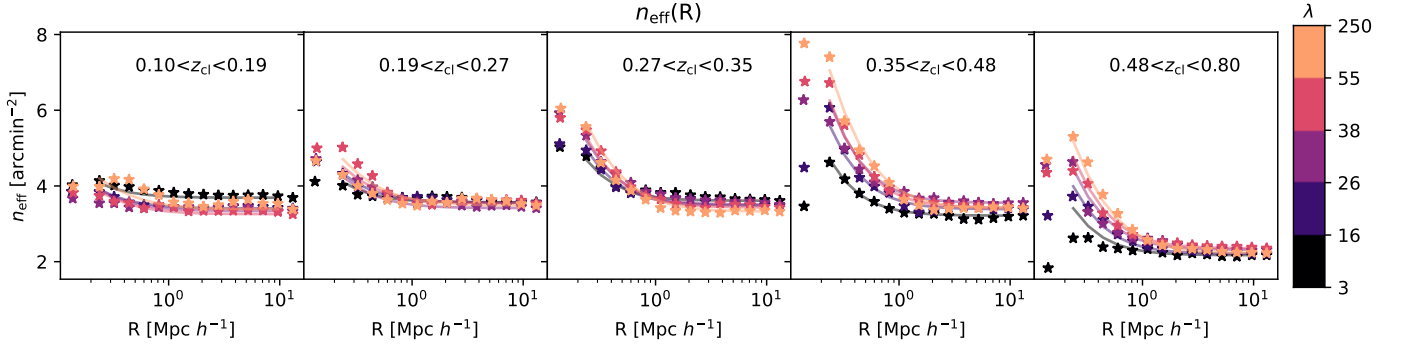
As shown in Appendix B,  $A(\lambda, z, R)$  is proportional to the radial number density profile of the cluster member contaminants, which we therefore modeled as a 2d projected NFW profile. Given that at this stage of the analysis, the cluster mass is still unknown, we resort to using the richness as a mass proxy. For ICM-selected cluster samples, the richness is known to scale approximately linearly with mass (Saro et al. 2015; Bleem et al. 2020; Grandis et al. 2020, 2021b). Furthermore, the richness provides a convenient proxy for the total number density of cluster member galaxies. It also has contributions from correlated structures along the line of sight that would also contribute almost unsheared contaminants to the background sample.

**Regularisation.** Given the complex interactions of cluster redshift, cluster member colors, photo- $z$  estimation, and background selection, we opt for a non-parametric fit for the redshift evolution by fitting independent amplitudes  $A_j$  for each cluster redshift bin. To impose a data-driven smoothness on the redshift evolution, we place a Gaussian prior with unknown variance  $\sigma_{\text{reg}}^2$  on the difference between the amplitudes of neighboring redshift bins, implemented via the Gaussian likelihood

$$\ln \mathcal{L}_{\text{reg}} = \frac{n_{z\text{-bins}} - 1}{2} \ln \sigma_{\text{reg}}^2 - \frac{1}{2\sigma_{\text{reg}}^2} \sum_j (A_j - A_{j+1})^2, \quad (13)$$

where  $n_{z\text{-bins}}$  is the number of redshift bins used. This likelihood adds an extra fit parameter,  $\sigma_{\text{reg}}$ , when added to the primary likelihood of the cluster member contamination.

**Source density fit.** The cluster member contamination can be fitted from the effective source density as a function of projected distance from the clusters (which are our lenses) stacked in bins of cluster redshift and richness, which we present in Fig. 4. This density is computed by dividing the effective number of sources, Eq. (9), by the geometric area of the radial bin. Given the presence of cluster member galaxies obstructing the detection of background galaxies and of detection masks, this results in an underestimation of the actual source density, as explored in detail by Kleinebreil et al. (2024) in the context of Kilo Degree Surveys (KiDS) WL around SPT selected clusters. This can be seen most clearly in the reported number density in the most central radial bin  $R < 0.2 h^{-1} \text{ Mpc}$ . On these radial scales, cluster lines of sight are typically dominated by the massive central



**Fig. 4.** Effective source density estimated from the effective number of source, Eq. (9), and the geometric area of the radial bins stacked in bins of cluster redshift (panels, lowest redshift left, higher redshift right) and of richness (color coding) as stars, for our background selection. Cluster members contaminating our background source sample leads to an increase of the effective source density toward the cluster center, which depends on both cluster richness and redshift. This signal needs to be determined, as cluster member contaminants are not sheared and dilute the shear estimator. Our model (lines) captures this trend well.

galaxy and its stellar envelope. We exclude this radial bin from the following analyses.

We fit the effective number density of cluster members as a function of projected radius  $R$  for a cluster of richness  $\lambda$  and redshift  $z$  as

$$n_{\text{eff}}(R|\lambda, z) = \frac{n_{\text{field}}(z)}{1 - f_{\text{cl}}(R|\lambda, z)}, \quad (14)$$

where the effective number density in the field  $n_{\text{field}}(z)$  is extracted around our clusters in the outer regions,  $11 h^{-1} \text{ Mpc} < R < 15 h^{-1} \text{ Mpc}$ . This choice of background is designed to measure the excess source lens clustering induced by the photometric misclassification of cluster members as background sources, which is the sole goal of the cluster member contamination correction.

For the fit of the parameters of the cluster member contamination,  $\mathbf{p}_{\text{cl}} = (A_j, B_\lambda, c)$ , we set up a Gaussian likelihood for the effective number density measured in the above mentioned richness-redshift bins. Our data vector, the effective source density, results from the weighted sum of all tomographic redshift bins (see Eq. (9)). We therefore fit one global cluster member contamination for our overall background selection. We assume a covariance matrix that combines Poisson noise from the effective number of sources and the variance we find in the outer radial bin as a proxy for the cosmic variance in the source density. Sampling this likelihood together with the regularization likelihood gives us constraints on the parameters of the cluster member contamination.

**Source redshift distribution decomposition.** The estimate from the effective source density around clusters needs to be validated independently, as in the current implementation, we did not consider source obstruction and masking. As demonstrated by Varga et al. (2019), and applied by McClintock et al. (2019), Paulus (2021), Shin et al. (2021), Bocquet et al. (2023), cluster member contamination can also be determined by decomposing the source redshift distribution as a function of the radius into a field component and a Gaussian component of cluster members, that is,

$$P(z_s|R, \lambda, z) = (1 - f_{\text{cl}}(R|\lambda, z))\hat{P}_{\text{field}}(z_s|z) + f_{\text{cl}}(R|\lambda, z)P_{\text{cl}}(z_s|z), \quad (15)$$

where the field source redshift distribution  $\hat{P}_{\text{field}}(z_s|z)$  is measured in the outer radial bin. The cluster member component is modeled as a Gaussian  $P_{\text{cl}}(z_s|z) = \mathcal{N}(z_s|z + \mu(z), \sigma^2(z))$  with mean offset  $\mu(z) = \mu_0 + \mu_z(z - z_0)$ , and standard deviation

$\sigma(z) = \sigma_0 + \sigma_z(z - z_0)$ , with pivot  $z_0 = 0.3$ , close to the cluster redshift median. This technique adapts the method proposed by Gruen et al. (2014) to wide photometric surveys with readily available photometric redshift estimates. Given that we fit for the shape of a normalized source redshift distribution, the amount of masking does not impact this inference. The likelihood is

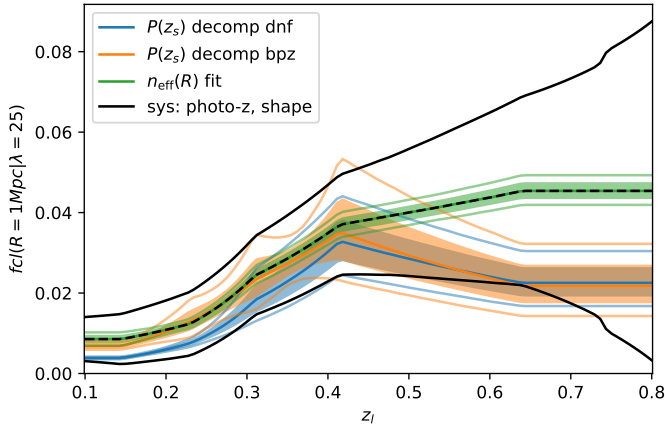
$$\ln \mathcal{L}_\beta = \sum N_{\text{eff}} \hat{P}_\beta(z_s) \ln P(z_s|R, \lambda, z) \text{ for } \beta \in (\text{BPZ}, \text{DNF}), \quad (16)$$

where the sum runs over richness, cluster redshift, cluster-centric distance and source redshift bins<sup>2</sup>. Sampled together with the regularization likelihood, this provides us with constraints on the parameters of the cluster member contamination. We perform this fit for redshift distributions constructed both for BPZ and DNF photo-zs.

**Results and comparison.** The best-fit results of the number density fit are over-plotted as lines on the effective source density data in Fig. 4, demonstrating that our model can capture the trends in the data well. Furthermore, we find a concentration for the contaminants profile of  $c = 2.5 \pm 0.1$ . This matches the concentration values found for the galaxy populations of massive clusters (Hennig et al. 2017). Our results also indicate a slight richness slope of the contamination fraction of  $B_\lambda = 0.47 \pm 0.1$ . The redshift evolution is shown in Fig. 5 in green (with the filled band corresponding to the 1-sigma uncertainty region and the green lines indicating the 2-sigma). For a cluster at richness pivot  $\lambda = 25$  at a cluster-centric distance of  $R = 1 \text{ Mpc}$ , we find that the contamination fraction increases from around  $f_{\text{cl}} \approx 0.01$  at  $z_{\text{cl}} = 0.1$  to  $f_{\text{cl}} \approx 0.05$  for  $z_{\text{cl}} > 0.5$ , with the strongest increase happening between redshifts  $0.25 < z < 0.4$ . We will use this setting to gauge the impact of the different cluster member contamination fits with respect to other systematics.

As our source selection is discontinuous as a function of lens redshift, we investigate the stability of our results with respect to the presence of the regularisation condition (Eq. (13)), which might impose an unphysical constraint on the cluster member contamination. As discussed above, we exclude the second tomographic redshift bin for lens redshifts  $z_l > 0.47$ . This could induce a sudden reduction of the cluster member contamination. We explore the impact of this by fitting the source redshift decomposition without the regularisation likelihood. We

<sup>2</sup> This follows directly from the likelihood of a sample  $x_i$  being drawn from a model distribution  $P(x)$ ,  $\ln \mathcal{L} = \sum_{i \in b} \ln P(x_i)$ . In our case, the model distribution is  $P(z_s|R, \lambda, z)$ , and the number of samples in each richness, cluster-centric distance, source redshift bin is  $N_{\text{eff}} \hat{P}_\beta(z_s)$ .



**Fig. 5.** Cluster redshift trend of the cluster member contamination of a richness  $\lambda = 25$  object at 1 Mpc from the cluster center. In green, the fit to the effective source number density, while in blue and orange, the constraints from the DNF and BPZ source redshift distribution decomposition are shown. The latter method is independent of masking. While we find marginal agreement between the contamination fractions determined by the two methods, the difference is smaller than the relative error induced by shape and photo- $z$  systematics – plotted in black extending from the source density fit for comparison.

forego a rebinning of our data, as the first four redshift bins up to  $z_1 > 0.48$  comprise mainly lenses for which we did not exclude the second tomographic redshift bin. Similarly, most lenses of the highest lens redshift bin,  $0.48 < z_1 < 0.8$ , utilize the third and fourth tomographic bins, as we exclude the third tomographic bin for lenses with  $z_1 > 0.74$ . We find shifts in the amplitude parameter no larger than  $|\Delta A_i| < 0.1$ , mostly in the last to lens redshift bins. This implies that the regularisation impacts the cluster member contamination  $f_{cl}$  no more than 10% in relative terms, and  $|\Delta f_{cl}| < 0.004$  in absolute term. These shifts are negligible compared to other sources of systematic uncertainty, as discussed below.

The source redshift distribution decomposition fits are validated by plotting differences between the measured source redshift distribution and the re-scaled field distribution, as shown in Fig. 6 for 1 Mpc from the center. This subtraction clearly highlights an approximately Gaussian component, that is well described by our model, plotted as full lines. The low amplitude negative dips at source redshifts slightly larger than the Gaussian component show that our field re-scaling, and thus our cluster member estimate, is imperfect. These residuals are clearly smooth in nature, making it very unlikely that they are of statistical origin. They might be systematically due to an alteration of the source redshift distributions in cluster lines of sight. Nonetheless, they are less than 1% in amplitude. This is within our systematic error budget, as discussed below (see Sect. 3.3.2). We therefore accepted the source redshift distribution decomposition as a valid fit.

Constraints on the concentration of the contaminants profile and the richness trend of the contamination fraction are in statistical agreement between the fit to the effective number density and the source redshift distributions. The contamination fraction resulting from the source redshift distribution is plotted in Fig. 5. It shows qualitative agreement with the contamination fraction constrained from the effective source density at low redshift. We will argue in the following, why the high redshift differences will not bias the mass calibration.

### 3.3.2. Shape measurement and photo- $z$ uncertainty

We compare the difference between the cluster member contamination estimates to the shape and photo- $z$  measurement uncertainties of the DES Y3 data as follows. The shape and photo- $z$  uncertainties of the tomographic redshift bin  $b$  are summarized by realizations of the source redshift distributions,  $P_{\mathcal{H}}(z_s|b)$ , running over a hyperparameter  $\mathcal{H}$ . The variation in the shape of these distributions expresses the uncertainty in the photo- $z$  calibration, while variation in the normalization expresses uncertainties in the multiplicative shear bias  $1 + m$ , as described in more detail in Cordero et al. (2022). Folding this together with our background selection, implemented via lens-redshift dependent weights  $w_b$  for the different tomographic redshift bins  $b$  (e.g., Eq. (7)), we find a fractional uncertainty of mean lensing efficiency

$$r_{g,s} = \frac{\sqrt{\text{Var}_{\mathcal{H}} \left[ \sum_b w_b \int dz_s P_{\mathcal{H}}(z_s|b) \Sigma_{\text{crit,ls}}^{-1} \right]}}{E_{\mathcal{H}} \left[ \sum_b w_b \int dz_s P_{\mathcal{H}}(z_s|b) \Sigma_{\text{crit,ls}}^{-1} \right]}, \quad (17)$$

which we take to express the impact of shear and photo- $z$  measurement uncertainties on the measured shear profiles. As the source redshift distributions for the tomographic redshift bins are already response-weighted, no further response-weighting is necessary for the expression above.

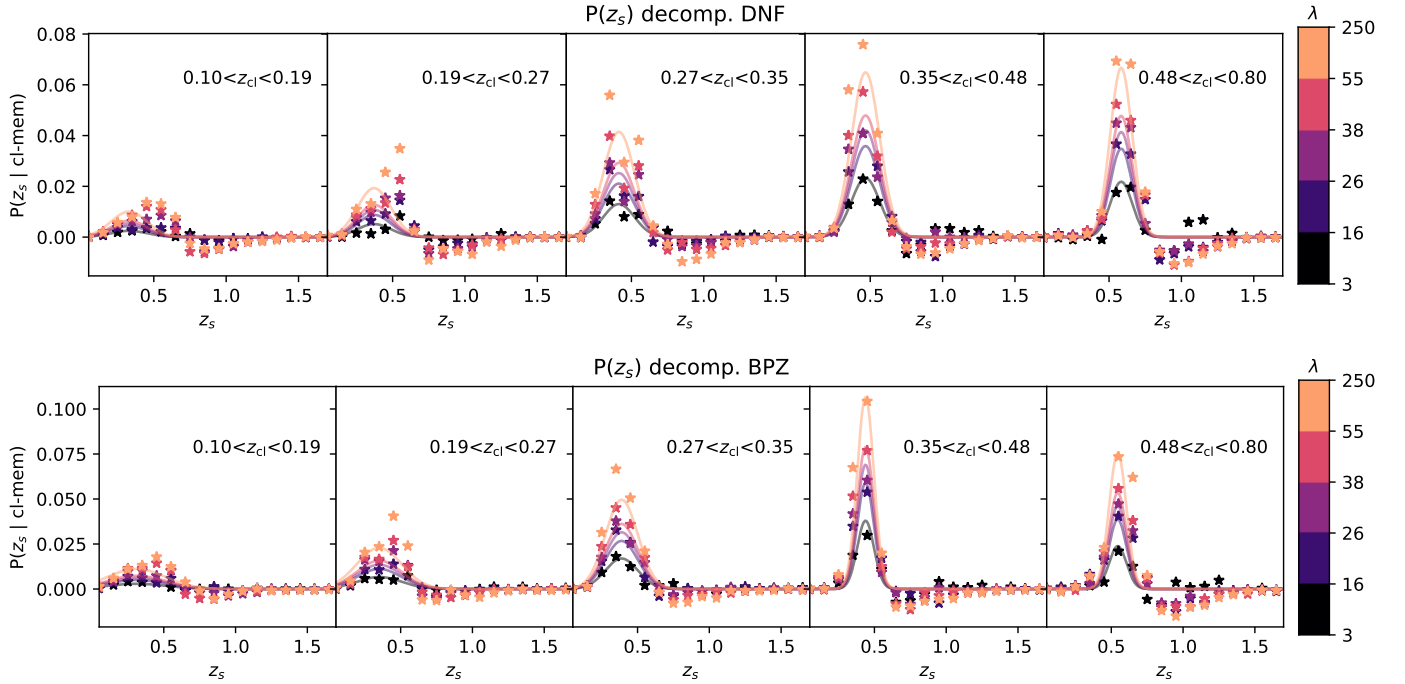
The fractional uncertainty increases from 0.55% at  $z_{cl} = 0.1$  to 4.3% at  $z_{cl} = 0.8$ . We plot these systematic uncertainties around the best-fit values for our cluster member contamination from the effective source density in Fig. 5. The difference between our different cluster member contamination estimates is smaller than this systematic uncertainty. We therefore concluded that we can determine the cluster member contamination to a higher accuracy than provided by the shape and photo- $z$  measurement uncertainties.

### 3.3.3. Shape noise

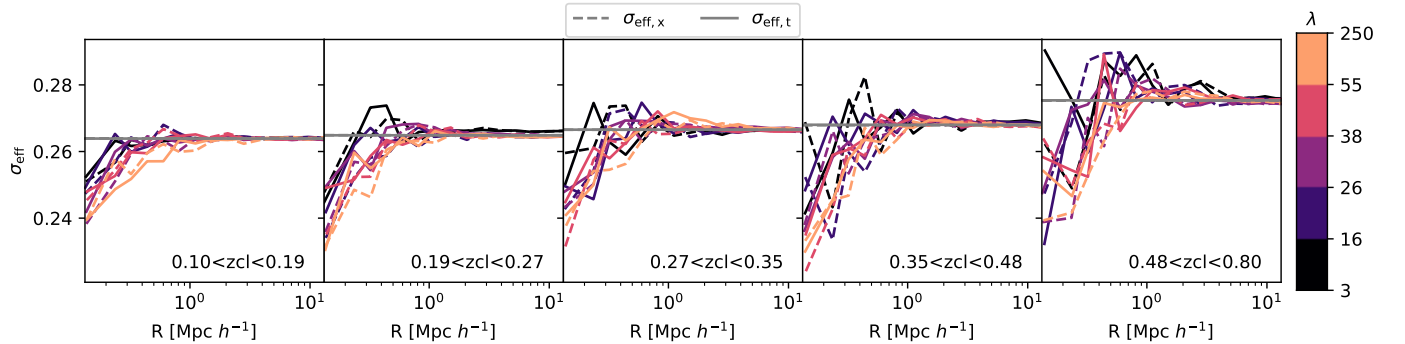
In Fig. 7, we show the measured effective shape noise as a function of cluster redshift and richness and as a function of cluster-centric projected distance. The effective shape noise increases with cluster redshift and richness and decreases toward the cluster center. In conjunction with the cluster member contamination estimate from the previous section, we can reconstruct the effective shape noise of cluster members from this decrement. To this end, we fit the measured effective shape noise with a squared sum of field shape noise determined in the outskirts of the clusters and an unknown cluster-member shape noise. The latter is found to be between 18% (for  $0.1 < z_{cl} < 0.19$ ) and 6% (for  $0.48 < z_{cl} < 0.80$ ) lower than the field shape noise, with the trend steadily decreasing with redshift. Given that most of the signal-to-noise in cluster mass measurements comes from scales larger than  $R > 1 h^{-1}$  Mpc, we ignore this effect. In future work, we aim to investigate if this is due to cluster member galaxies being inherently rounder, if the survey-averaged shear response is impacted by the stronger blending in the crowded cluster lines of sight, or if cluster members are just brighter, and therefore have a smaller shape measurement uncertainty contribution to the effective shape noise.

For our shape noise modeling, we instead focus on these larger scales, where we measure that the effective shape noise increases with source redshift, as shown in Fig. 8, both for the tangential and the cross component. This trend matches the survey averaged effective shape dispersion found by Friedrich et al. (2021) for the tomographic redshift bins. For later use in our





**Fig. 6.** Validation of the source redshift distribution decomposition for DNF (top) and BPZ (bottom) source redshift estimates. In different richness (color coded) and redshift bins (different panels), we show the difference between the source redshift distribution measured along the cluster line of sights and the field distribution extracted at a safe distance from the cluster center, shown as stars. Both are shown here for the projected cluster-centric distance of 1 Mpc in our reference cosmology. The resulting increment is well modeled by a Gaussian component (lines) caused by the cluster members contaminating the source sample. The amplitude of this component scales with cluster richness.



**Fig. 7.** Effective shape noise as a function of cluster-centric distances stacked in bins of cluster redshift (panels: lowest redshift left, higher redshift right) and of richness (color coding). Despite significant scattering, increasing toward the cluster center, a trend of lower-shaped noise toward the center is visible. As gray horizontal lines, the effective shape noise in the field, used for the semi-analytical covariance of the shear profiles, is shown.

semi-analytical covariance matrix modeling, we fit this trend with the quadratic expression

$$\sigma_{\text{eff}}(z_{\text{cl}}) = 0.2635 + 0.0300 z_{\text{cl}}^2 - 0.0008 z_{\text{cl}}. \quad (18)$$

This fit is shown as a dashed line in Fig. 8, excellently matching the trend we measure in the data. Extrapolated to  $z_{\text{cl}} \rightarrow 0$ , it matches closely the value reported by Gatti et al. (2021). Without a reliable estimate for the uncertainty of the effective shape noise, we forego reporting error bars on this fit, using it as an effective smoothing to avoid noise on the error estimate of our target quantity, the reduced shear profile.

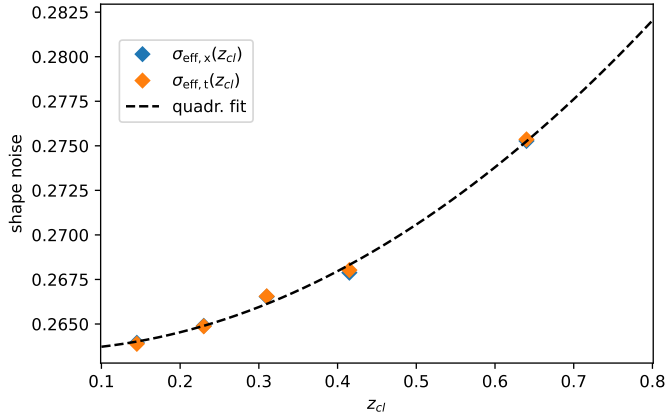
### 3.3.4. Cross-shear

Using our fitting formula for the effective shape noise, for each bin in richness and lens-redshift, we can com-

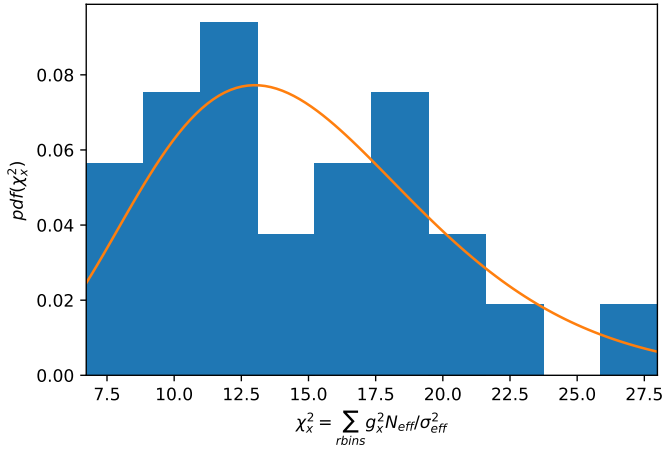
pute the consistency of the cross-shear signal with zero as

$$\chi_x^2 = \sum_{r\text{-bins}} \frac{g_x^2 N_{\text{eff}}}{\sigma_{\text{eff}}^2}. \quad (19)$$

These squared losses follow a chi-squared distribution with  $n_r$  degrees of freedom, where  $n_r$  is the number of radial bins we consider, as shown in Fig. 9. This demonstrates that our cross-shear component averages to zero, consistently with the expectation of well calibrated shape measurements distorted by gravity. We also visually inspect the cross-shear profiles in the above-defined richness and redshift bins and find no significant deviations from zero. Computing instead the global degree of consistency of the raw tangential



**Fig. 8.** Cluster redshift trend of the effective shape noise in the field (shown as blue and orange diamonds for the cross and tangential reduced shear, respectively). Given our cluster redshift-dependent background selection, the mix of source galaxies changes as a function of cluster redshift. The trend is well-fitted by a quadratic expression, shown as a black dashed line.



**Fig. 9.** Distribution of the uncertainty-scaled squared residuals of the cross-shear profile with respect to zero signal, as a blue histogram. In orange, the chi-squared distribution for 15 degrees of freedom matches the number of radial bins. The squared loss distribution is well described by the chi-squared distribution, indicating that the cross-shear profiles are consistent with zero given our shape noise modeling.

reduced shear profiles with zero, we find a signal-to-noise of 92<sup>3</sup>.

### 3.3.5. Shear response

As a further validation test of our WL measurement, we compute the mean shear responses in richness, redshift, and distance bins, given by

$$\mathcal{R} = \frac{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s \mathcal{R}_i}{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s} \text{ for } \alpha \in (t, x), \quad (20)$$

shown in Fig. 10.

Notably, the shear response increases toward the cluster center while also showing a gentle trend with richness. The shear

response used here is a smooth function of the source relative size with respect to the PSF and its signal-to-noise. Cluster member galaxies are both larger and brighter than the average field population, which could explain the trend we see. However, this signal seems to affect all lens redshift bins equally, independently of their cluster member contamination levels. Another contributing effect could be that field galaxies in cluster lines of sight are magnified by the cluster potential, altering their sizes and signal-to-noise ratios compared to the field galaxies. While interesting, at the scales that we will use for mass calibration,  $R > 0.5 h^{-1} \text{ Mpc}$ , the effect is smaller than 1%. We can thus safely ignore it in light of the other, larger systematic uncertainties. We plan to investigate how these properties are impacted by cluster lines of sight in future work.

### 3.4. Selection response

The quantities that we use for our source selection are impacted by the shear we try to measure. For instance, the flux of a sheared galaxy is different from the flux of the original galaxy if it was not sheared. This means that the shear we try to measure impacts our selection of source galaxies and biases our estimators. To correct this bias, we have to measure how our estimator fares on catalogs selected from artificially sheared versions of the actual DES Y3 images. Four such catalogs have been created, each with an artificial shear of  $\Delta\gamma = 0.01$  positive/negative (+/-) on the first/second Cartesian component. The selection response in richness, redshift, and distance bins, is given by

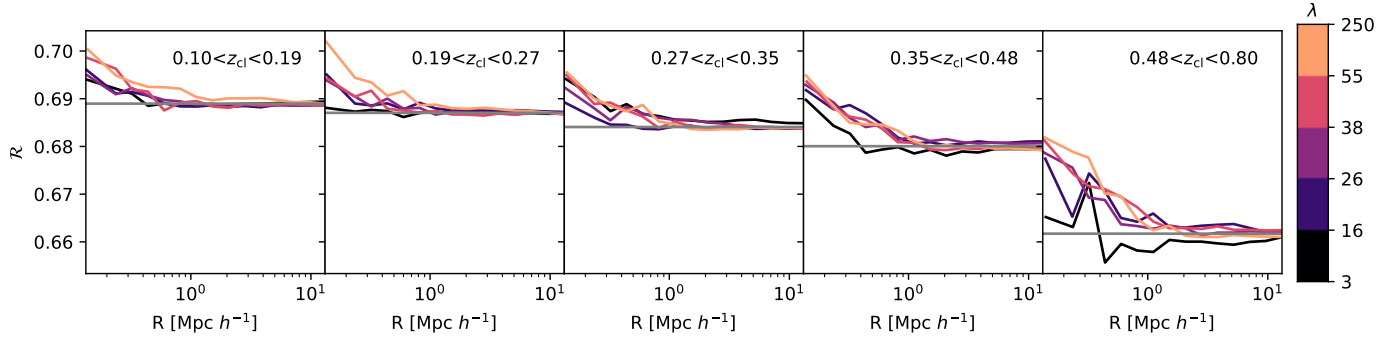
$$\begin{aligned} \mathcal{R}_{\alpha, \text{sel}} &= \frac{\mathcal{R}_{\alpha, \text{sel}, 1} + \mathcal{R}_{\alpha, \text{sel}, 2}}{2} \\ &= \frac{1}{2\Delta\gamma} \left( \langle g_{\alpha, \text{raw}} \rangle^{1+} - \langle g_{\alpha, \text{raw}} \rangle^{1-} + \langle g_{\alpha, \text{raw}} \rangle^{2+} - \langle g_{\alpha, \text{raw}} \rangle^{2-} \right), \end{aligned} \quad (21)$$

for  $\alpha \in (t, x)$ . Here  $\langle g_{\alpha, \text{raw}} \rangle^{1+}$  denotes the tangential reduced shear estimator on the source sample extracted from the images that have been artificially sheared in positive (+) direction along the first (1) Cartesian shear component, and similarly for the negatively sheared images (-), and the second Cartesian component (2), respectively. The radial, richness, and redshift trends of the selection response are shown in Fig. 11. Especially for high cluster redshift bins, the selection response is very noisy such that no evident trends can be made out.

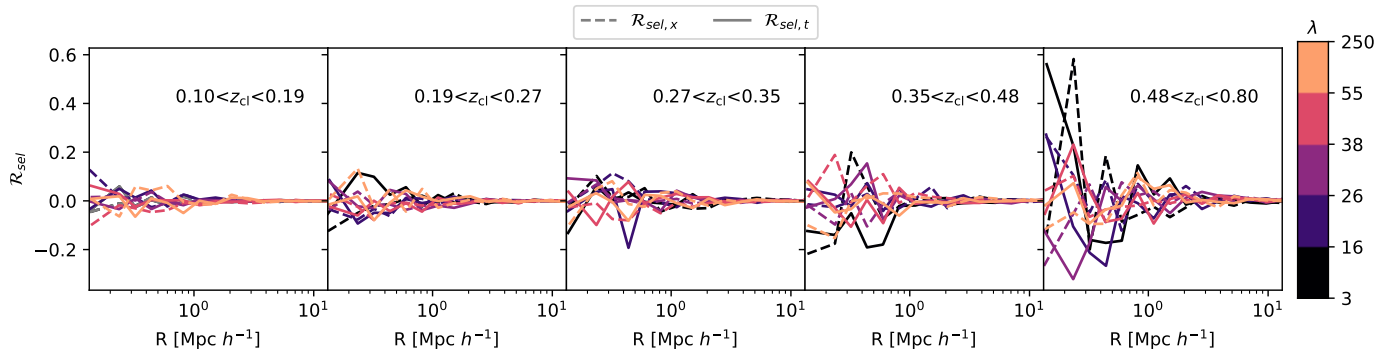
A reliable estimate can thus only be obtained after averaging over richness and radius. We find that the global selection response is positive and bounded by  $\mathcal{R}_{\alpha, \text{sel}} < 0.001$  for  $z_{cl} < 0.48$ , and  $\mathcal{R}_{\alpha, \text{sel}} \sim 0.004$  for  $0.48 < z_{cl} < 0.8$ .

Direct comparison between the mean shear responses and the selection response shows that the latter is of order 0.1% of the former for  $z_l < 0.4$ , increasing to 0.5% for the redshift bin  $0.48 < z_l < 0.8$ . In both cases, this is a factor of a few less than the uncertainty induced by shape and photo- $z$  measurements and, as such, can be ignored. This assessment contrasts the selection response's impact on galaxy-galaxy lensing studies (Prat et al. 2022). The main difference is that we do not consider each tomographic bin independently but rather make a selection based on the lens redshift. For low lens redshifts,  $z_l < 0.48$ , we thus include a large fraction of the total source sample, reducing the impact of selection effects compared to a tomographic redshift bin selection. At high lens redshift,  $0.48 < z_l < 0.8$ , the close proximity of the source sample to the lens dramatically increases the effects of photo- $z$  uncertainties on the lensing efficiency, dwarfing the effect of the selection response.

<sup>3</sup> Here we use the definition of signal-to-noise  $S/N = \sqrt{\sum \chi_i^2}$ , summed over all bins. This quantity does not enter our analysis and is simply used to give a rough estimate of the constraining power of our dataset.



**Fig. 10.** Average shear response, as a function of cluster-centric distance, stacked in bins of cluster richness (color coded), and cluster redshift panels: lowest redshift left, higher redshift right. Noticeably, the shear response increases toward the cluster center. Interestingly, this signal seems to affect all lens redshift bins equally, independently of their cluster member contamination levels. We account for this effect in the cosmology-ready data products.



**Fig. 11.** Selection response of our tangential (full lines) and cross (dashed lines) shear estimator, averaged over artificial shears in the two Cartesian coordinates, as a function of cluster-centric distance, stacked in bins of cluster richness (color coded), and cluster redshift panels: lowest redshift left, higher redshift right. The selection response is very noisy but shows no significant trends with richness or cluster-centric distance.

**Cosmology-ready data products.** Given the various distance dependencies of the WL signal, the mass calibration is cosmology-dependent. To marginalize this dependency in this work and to enable the WL mass calibration of our sample self-consistently with the cosmological number counts experiment in [Ghirardini et al. \(2024\)](#), we need to define cosmology-independent data products. Furthermore, given that we pursue a Bayesian population analysis in which each individual cluster likelihood is evaluated, we construct the following data products for each cluster in the cosmology sample that has DES Y3 lensing information:

- the measured reduced shear profile

$$\hat{g}_t = \frac{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s e_{t,b,i}}{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s \mathcal{R}_i}, \quad (22)$$

in radial bins with  $0.5 h^{-1} \text{ Mpc} < R < 3.2(1 + z_{\text{cl}})^{-1} h^{-1} \text{ Mpc}$ , in our reference cosmology. The outer bin is chosen following [Grandis et al. \(2021a\)](#) to limit our extraction to the 1-halo term region. We use only the shear response  $\mathcal{R}_i$ , and ignore the selection response, as argued in Sect. 3.4;

- the statistical uncertainty on the reduced shear,

$$\delta g_t = \frac{\sigma_{\text{eff}}(z_{\text{cl}})}{\sqrt{N_{\text{eff}}}} \text{ with } N_{\text{eff}} = \frac{\left( \sum_{b=2,3,4} \sum_{i \in b} w^b w_i^s \mathcal{R}_i \right)^2}{\sum_{b=2,3,4} \sum_{i \in b} \left( w^b w_i^s \mathcal{R}_i \right)^2}, \quad (23)$$

in the same bins. Note here that we use the quadratic fit to the global shape noise given in Eq. (18) to suppress the noise in the individual shear variance estimators;

- the average angular distance of the binned sources from the cluster position in the corresponding radial bins

$$\theta = \frac{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s \mathcal{R}_i \theta_i}{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s \mathcal{R}_i}. \quad (24)$$

We report the angular scale, because in the cosmological inference, the distance–redshift relation needs to be self-consistently re-calculated while the data binning is frozen, and

- the field source redshift distribution extracted in the outer regions of the cluster ( $11 h^{-1} \text{ Mpc} < R < 15 h^{-1} \text{ Mpc}$ ) based on the SOM-Pz redshift estimation method,

$$P(z_s) = \frac{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s \mathcal{R}_i P(z_s | \hat{c}_i)}{\sum_{b=2,3,4} w^b \sum_{i \in b} w_i^s \mathcal{R}_i}, \quad (25)$$

where  $\hat{c}_i$  is the SOM cell in which the respective source falls. This extraction ensures that our source redshift distribution is a representation of the local field sources.

Using these scale cuts, we reduce the signal-to-noise ratio in the tangential reduced shear to 65. These cuts are necessary to avoid the increased systematic uncertainty toward the cluster centers (see [Grandis et al. 2021a](#), Sect. 3.2). The upper limit was chosen to limit our WL observable to the 1-halo regime for high mass systems,  $M_{500c} > 2 \times 10^{14} h^{-1} M_{\odot}$ . As such, we avoid the 2-halo regime with its more complicated cosmology dependence and possible effects of assembly bias for a majority of our cluster sample and in the higher WL signal-to-noise regime.

**Table 1.** Overview of the simulation inputs used for the calibration of the WL bias and scatter.

Name	Use case	Section
TNG300	Projected surface mass density of massive halos at different redshift snapshots	4.1
MAGNETICUM PATHFINDER	Offsets between X-ray surface brightness peak and projected halo center	4.2.1
eROSITA all-sky survey twin	Offsets between X-ray surface brightness peak and measured X-ray position	4.2.2
	Mapping between halo mass and redshift, and detection likelihood and extent	4.4
Monte Carlo realizations	Propagation of systematic uncertainties to error budget on the WL bias	4.5

**Notes.** The respective references, as well as detailed explanations, are given in the respective sections.

#### 4. Calibration of mass extraction

Proper cosmological exploitation of the data products derived above must accurately map the WL signal to the halo mass used in the number counts experiment. In this work, we follow the approach of Grandis et al. (2021a) by simulating synthetic shear profiles starting from hydrodynamical simulations, WL survey specifications, and characteristics of the lens sample. The actual mass measurement is performed on the synthetic shear profiles, resulting in the so-called WL mass, which displays a bias and scatter with respect to the true halo mass. Priors on the WL bias and scatter are derived from Monte Carlo realizations of the synthetic shear profiles, which sample the range of systematic uncertainties on the input specifications. The different simulation inputs used for this calibration are summarized in Table 1.

##### 4.1. Hydrodynamical simulations input

We base the creation of our synthetic shear profiles on the surface mass maps of halos with  $M_{200c} > 3 \times 10^{13} h^{-1} M_{\odot}$ , extracted with the technique outlined in Grandis et al. (2021a) from the cosmological hydrodynamical TNG300 simulations (Pillepich et al. 2018; Marinacci et al. 2018; Springel et al. 2018; Nelson et al. 2018, 2019; Naiman et al. 2018) at redshifts  $z_{\text{snap}} = 0.24, 0.42, 0.64, 0.95$ . To boost the statistical sample size, for each halo, we created three maps by projecting them along each Cartesian coordinate. As described in Grandis et al. (2021a), these maps are processed into scaled tangential shear  $\Gamma_t(R, \phi|R_{\text{mis}})$  and scaled convergence  $\Sigma(R, \phi|R_{\text{mis}})$  maps, binned in polar coordinates  $(R, \phi)$  around isotropically mis-centered positions for a range of mis-centering distances  $R_{\text{mis}}$ .  $\Sigma(R, \phi|R_{\text{mis}})$  is the surface mass density in units of  $h M_{\odot} \text{Mpc}^2$ , which multiplied by the inverse critical surface density gives the convergence,  $\kappa = \Sigma_{\text{crit}}^{-1} \Sigma$ . The definition of  $\Gamma_t$  has been introduced in Grandis et al. (2021a) to signify the result of applying the inverse of Kaiser–Squires algorithm to the surface mass density  $\Sigma(x)$ , and then determining the tangential component with respect to the chosen center. As both these operations are linear, the resulting quantity is directly related to the tangential shear map as  $\gamma_t(R, \phi|R_{\text{mis}}) = \Sigma_{\text{crit}}^{-1} \Gamma_t(R, \phi|R_{\text{mis}})$ . Being independent from the specific source redshift distribution, this is a convenient simulation output.

This data format conserves the azimuthal anisotropy sourced by the halo triaxiality and by the mis-centering. Possible inaccuracies induced by the correlation between halo shape and mis-centering direction, recently highlighted by Sommer et al. (2023), have been shown to not significantly impact the cosmological results derived from the mass calibration presented in this work (see Ghirardini et al. 2024, Sect. 5.1), but likely need to be accounted for in future analyses with larger statistical constraining power.

Note also that we employ in this work the strategy to mitigate hydrodynamical modeling uncertainties proposed by Grandis et al. (2021a) and used by Chiu et al. (2022, 2023), Bocquet et al. (2023). The hydrodynamical simulations from which we source the surface density maps have gravity-only twin runs with the same initial conditions. It is well established that both runs form exactly the same halos (Castro et al. 2021; Grandis et al. 2021a), albeit with different masses. We purposefully choose the fictitious mass in the gravity-only simulation, as the halo mass functions used in the number counts experiment are calibrated on gravity-only simulations. We can thus use these calibrations to absorb the effects of hydrodynamical feedback on the mass distributions of massive halos in the mapping between the (gravity-only) halo mass and the WL mass, and account for the uncertainty in this mapping in the systematic error budget (see Sect. 4.5).

##### 4.2. Mis-centering

A faithful simulation of synthetic shear profiles requires us to account for the fact that the observed position of the lens is mis-centered from the projected position of the true halo center, which is defined consistently through this work as the position of the most bound particle. The surface mass maps are centered around this position (Grandis et al. 2021a).

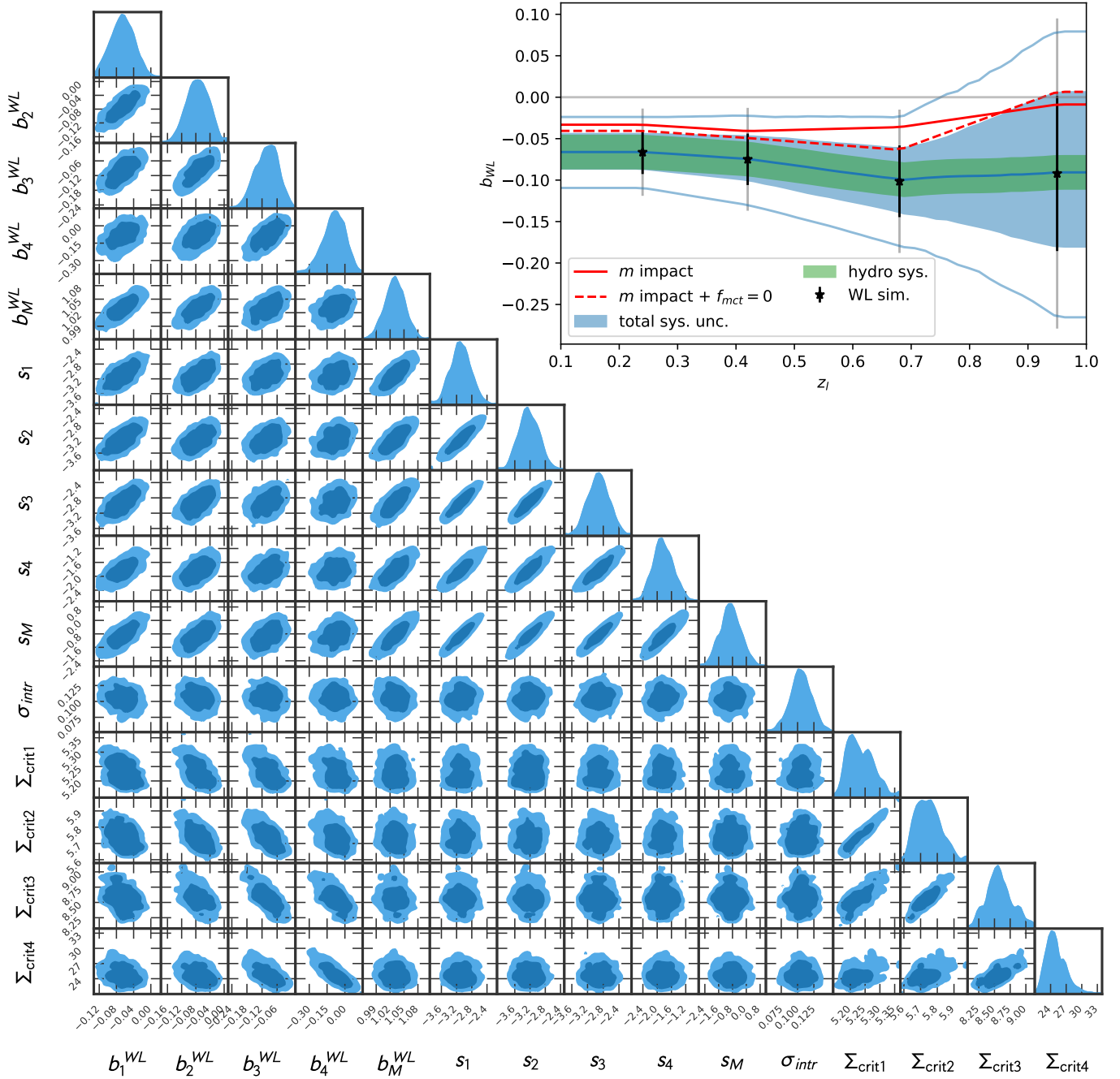
We distinguish two sources of mis-centering: (1) the offset between the noise-free peak of the 2d X-ray surface brightness profile and the 2d position of the halo center (hereafter called intrinsic), and (2) the offset between the noise-free peak of the 2d X-ray surface brightness profile and the measured X-ray position (hereafter called observational). The latter displacement is induced by the photon shot noise and the PSF of the eROSITA cameras and is a special case of the astrometric uncertainty discussed in Merloni et al. (2024).

###### 4.2.1. Intrinsic mis-centering

The intrinsic mis-centering is studied through the use of Box2b/hr of the hydrodynamical cosmological simulation suite MAGNETICUM PATHFINDER<sup>4</sup> (Dolag et al., in prep.), performed assuming the WMAP-7 cosmology as  $\Omega_0 = 0.272$  and  $h = 0.704$  as given by Komatsu et al. (2011). The volume of the Box2b/hr box is  $(909 \text{ cMpc})^3$ , allowing for a broad range of galaxy cluster masses up to virial masses  $M_{\text{vir}}$  of several times  $10^{15} M_{\odot}$ . The galaxy cluster sample, comprised of 116 clusters with  $M_{\text{vir}} \geq 10^{14} M_{\odot}$  at  $z = 0.252$ , is described in detail in Kimmig et al. (2023). We select an additional 75 galaxy clusters at  $z = 0.518$  to account for redshift trends. The second sample is taken from a different box of equal resolution, Box2/hr of size  $(500 \text{ cMpc})^3$ , to avoid double counting the halos at different redshifts.

<sup>4</sup> [www.magneticum.org](http://www.magneticum.org)





**Fig. 12.** Joint distribution on the WL bias and scatter parameters ( $b_l^{WL}$ ,  $s_l = \ln(\sigma_l^{WL})^2$ ), and their mass slopes ( $b_M^{WL}$ ,  $\sigma_M$ ), along with the principal sources of systematic uncertainty: the lensing efficiencies  $\Sigma_{crit,i}$  to the snapshots  $i$ , in units of  $10^{15} h M_\odot \text{Mpc}^{-2}$ , and the scale of the intrinsic mis-centering  $\sigma_{intr}$ . The high redshift biases anticorrelate with the respective lensing efficiencies, indicating that in this regime, we are limited by the DES Y3 photo- $z$  accuracy. Conversely, the low redshift biases and the scatter values correlate among themselves, as they are dominated by the hydro uncertainty floor we added. Insert: Redshift interpolation of the WL lensing bias in cyan (filled region 1 sigma, faded lines 2 sigma), together with the 1 sigma uncertainty due to hydro modeling (green), explorations of the mean impact of the multiplicative shear bias  $m$  modeling (red), and the additional omission of the high mis-centering tail  $f = 0$  (dashed red), and the posteriors of the WL bias in the four snapshots. Hydrodynamical modeling uncertainties are dominant at low redshift. When considering our multiplicative shear bias modeling choices, the bias is consistent with zero, while the presence of the highly mis-centered tail does not affect the WL bias.

Every cluster is projected from 100 random viewing angles. We then determine the projected center-of-mass via the shrinking sphere method (Power et al. 2003) and around this center, the peak of the X-ray emission between 0.5–2 keV. The intrinsic mis-centering is then the projected displacement between the X-ray peak and the center of the halo, given by the most-bound particle.

We find that the mis-centering is well described by a single Rayleigh distribution with mean  $\langle R_{mis}^{intr} \rangle = \sigma_{intr} R_{500c}$ , with  $\sigma_{intr} = 0.104 \pm 0.016$ , and with no further resolved mass or redshift trends. As shown below in Sect. 4.5 and Fig. 12, this uncertainty is not relevant when taken together with other systematic effects, especially photometric source redshift and hydrodynamical modeling uncertainties. Valid concerns about

the fidelity of the Magneticum prediction of the surface brightness in inner cluster regions are therefore unlikely to impact our results. In future analyses, we plan to validate the centering choice more carefully, for instance, through the blind comparison of the results based on different centers (as done for instance by Bocquet et al. 2023).

#### 4.2.2. Observational mis-centering

For the observational mis-centering, we use the offset  $\theta_{\text{sep}}$  between input and output positions of clusters in the “eROSITA all-sky survey twin” (Comparat et al. 2019, 2020; Seppi et al. 2022), which produces and analyses simulations of the X-ray signature from a realistic active galactic nucleus and cluster population in eRASS1 observational conditions that include exposure time variations and background. Furthermore, the X-ray cluster finding algorithm, as described in Merloni et al. (2024), Bulbul et al. (2024) is run on the event file generated from this simulation, ensuring that PSF and shot noise of the real observations are well mapped. We select the input clusters matched with simulated sources that have detection likelihood  $\mathcal{L}_{\text{det}} > 3$ , extent likelihood  $\mathcal{L}_{\text{EXT}} > 6$ , and extent  $4 \leq \text{EXT} \leq 60$ , to match the real cosmology sample. For each such object, we compute the separation  $\theta_{\text{sep}}$  between input and output position, using the entries described in Table A.1 by Seppi et al. (2022).

We model the distribution of mis-centering with a mixture of a Rayleigh distribution and a Gamma distribution with shape parameter 2 as

$$P(\theta_{\text{sep}} | \langle \theta_{\text{sep}} \rangle, f, \lambda) = \frac{1}{1+f} \frac{\theta_{\text{sep}}}{\langle \theta_{\text{sep}} \rangle^2} \exp\left(-\frac{1}{2} \frac{\theta_{\text{sep}}^2}{\langle \theta_{\text{sep}} \rangle^2}\right) + \frac{f}{1+f} \frac{\lambda^2 \theta_{\text{sep}}}{\langle \theta_{\text{sep}} \rangle^2} \exp\left(-\frac{\lambda \theta_{\text{sep}}}{\langle \theta_{\text{sep}} \rangle}\right), \quad (26)$$

with the mean observational mis-centering given by

$$\langle \theta_{\text{sep}} \rangle = \langle \theta_{\text{sep}} | \alpha, \sigma_0, k \rangle = \left( \frac{\mathcal{L}_{\text{det}}}{38.7} \right)^{-\alpha} \sqrt{\sigma_0^2 + k^2 \text{EXT}^2}, \quad (27)$$

where  $\mathcal{L}_{\text{det}}$  is the detection likelihood, EXT the extent, as defined by the best-fit detection beta model, and  $\mathbf{p}_{\text{mct}} = (\alpha, \sigma_0, k, \ln f, \ln \lambda)$  are free parameters.  $f$  is the fraction of mis-centered objects in the heavy-tailed component,  $\alpha$  the detection likelihood trend of the mean observational mis-centering,  $\sigma_0$  the extrapolated mean observational mis-centering for a point source (EXT = 0),  $k$  the slope of the mean mis-centering with respect to the sources extent, and  $\lambda$  the length of the mis-centering tail in units of the mean mis-centering.

We fit these by sampling the likelihood

$$\ln \mathcal{L} = \sum_v \ln P(\theta_{\text{sep}}^v | \langle \theta_{\text{sep}} \rangle, \alpha, \sigma_0, k, \ln f, \ln \lambda), \quad (28)$$

assuming flat priors on all parameters, with  $v$  running over the clusters selected from the digital twin simulation. All parameters are well constrained; Table 2 reports their means and standard deviations. Noticeably, the mean dependence on the detection likelihood  $\alpha = 0.43 \pm 0.02$  is quite close to the  $\alpha = 0.5$  trend one would expect for matched filter extraction in the presence of Gaussian noise (Story et al. 2011; Song et al. 2012). Our constraint on  $\ln f$  translates into a relative weight of  $0.318 \pm 0.043$  for the flatter Gamma distribution component, albeit only a part of

**Table 2.** Calibration parameters with their priors for the Monte Carlo marginalization of the WL bias and scatter determination.

Cluster member contamination (Sect. 3.3.1)	
$\log_{10} c = 4.83 \pm 0.22$	$B_{\text{cl}} = 0.467 \pm 0.017$
$\ln \sigma_{\text{reg}} = -0.61 \pm 0.43$	$A_0 = -4.75 \pm 0.11$
$A_1 = -4.38 \pm 0.05$	$A_2 = -3.68 \pm 0.04$
$A_3 = -3.25 \pm 0.04$	$A_4 = -3.04 \pm 0.04$
Mis-centering parameters (Sect. 4.2)	
$\sigma_0 [\text{arcsec}] = 4.83 \pm 0.22$	$k = 0.308 \pm 0.007$
$\alpha = 0.443 \pm 0.012$	$\ln f = -0.76 \pm 0.11$
$\ln \beta = -0.18 \pm 0.04$	$\sigma_{\text{intr}} = 0.104 \pm 0.016$
$\ln \lambda = -0.18 \pm 0.04$	
Richness mass relation (Chiu et al. 2022, Appendix A)	
$A_\lambda = 36.2 \pm 3.6$	$B_\lambda = 0.881 \pm 0.083$
$C_\lambda = -0.46 \pm 0.52$	$\sigma_\lambda = 0.274 \pm 0.066$
Photo- $z$ and shape calibrations	
$\mathcal{H} \sim \text{Int}(0, 999)^{(*)}$	$\alpha_{\text{NL}} = 0.6 \pm 0.4^{(\dagger)}$
Hydrodynamical modeling uncertainties	
$\Delta b = 0.02, \Delta b_{\text{M}} = 0.018$	$\Delta s = 0.25, \Delta s_{\text{M}} = 0.59$

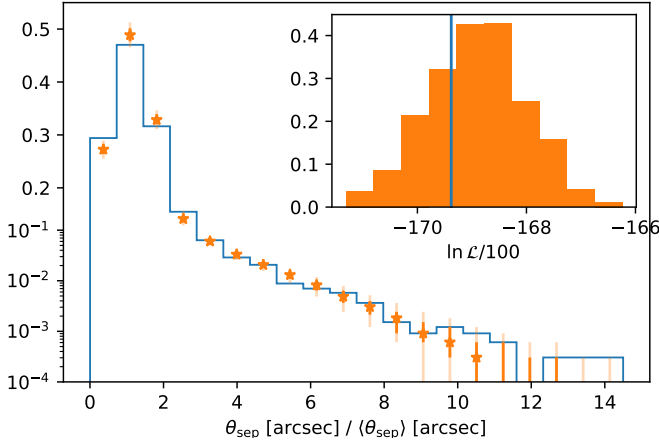
**Notes.** Note that the parameters for the cluster member contamination and the mis-centering are actually drawn from the respective posterior fits to conserve correlations among the parameters.  $^{(*)}\text{Int}(a, b)$  is a uniform distribution over the integers  $a \leq i \leq b$  (cf. Sect. 3.3.2);  $^{(\dagger)}$ nonlinear multiplicative shear bias (Grandis et al. 2021a, Sect. 2.1.7).

that distribution populates the high mis-centering tail. Our other fit parameters result in a typical mis-centering  $\langle \theta_{\text{sep}} \rangle \sim 11$  arcsec.

We also check that once the individual mis-centerings are corrected for the typical mis-centering of sources of that detection significance and extent, the residuals show no correlation with exposure time. This ensures that our parametrization captures the main trends of the observational mis-centering and that it can be safely used despite the exposure time varying over the survey footprint. To further validate our model, we draw mock catalogs by selecting a random posterior point  $\mathbf{p}_{\text{mct}} = (\alpha, \sigma_0, k, \ln f, \ln \lambda)$ . For each simulated cluster, we then compute the typical mis-centering  $\langle \theta_{\text{sep}} \rangle$  and draw a mock separation angle  $\theta_{\text{sep}}$  from Eq. (26). This procedure creates 1000 mock cluster samples that follow our model by drawing 1000 independent posterior points. In Fig. 13 we plot in blue the mis-centering distribution of the actual data, while in orange data points, the distribution of mock data, sampled from our posterior, are shown. The two are indistinguishable, indicating that the actual mis-centering distribution from the digital twin looks like data drawn from our mis-centering model. Comparisons of the maximum likelihood of the data and the mocks show that this also holds at the likelihood level (see insert of Fig. 13). Our observational mis-centering model is therefore a good fit.

#### 4.3. Extraction model

When performing the mass measurement, we need to specify a mass extraction model and coherently apply it to both the real and the synthetic data. In this analysis, we choose a simplified mis-centered model, corrected for the mean cluster member contamination, as suggested by Grandis et al. (2021a). This model faithfully captures the main biases induced by mis-centering and cluster member contamination while not adding numerical complexity compared to the plain 2d-projected NFW



**Fig. 13.** Observational mis-centering between the input X-ray surface brightness peaks and the output positions of clusters in the eROSITA digital twin,  $\theta_{\text{sep}}$ , scaled by the best-fit average offset,  $\langle \theta_{\text{sep}} \rangle$ , shown in blue as a histogram. Overlaid in orange are realizations of mock offsets drawn from the posterior of our model. The data is indistinguishable from the mock draws. Our model is also validated by comparing, in the insert, the likelihood of the data in blue with the likelihood of the mocks in orange.

model. This approach is well suited for the repeated calls in the Monte Carlo Markov chains used for parameter inference. Bocquet et al. (2023) make the same choices for the same reasons.

For the model, we assume that a cluster of radius  $R_{500c}$  and mean observational mis-centering  $\langle \theta_{\text{sep}} \rangle$  is effectively mis-centered by a physical distance

$$R_{\text{mis}}^{\text{extr}} = \sqrt{(\sigma_{\text{intr}} R_{500c})^2 + (D_l \langle \theta_{\text{sep}} \rangle)^2}, \quad (29)$$

evaluated at the mean of the observational mis-centering parameters determined in Sect. 4.2.2. The radius  $R_{500c}$  is derived from the input mass of the extraction model. The cluster surface mass distribution is then well approximated by

$$\Sigma(R|M) = \begin{cases} \Sigma_{\text{NFW}}(R_{\text{mis}}^{\text{extr}}|M) & \text{for } R < R_{\text{mis}}^{\text{extr}}, \text{ and} \\ \Sigma_{\text{NFW}}(R|M) & \text{otherwise,} \end{cases} \quad (30)$$

where  $\Sigma_{\text{NFW}}(R|M)$  is the 2d projected NFW profile, evaluated with the mean concentration mass relation reported by Ragagnin et al. (2021). Deviations of this concentration mass relation from the one in the simulations are absorbed in the WL bias, while scatter in concentration at fixed mass contributes to the WL scatter. As shown in Bocquet et al. (2023), this surface mass profile results in a radial density contrast  $\Delta\Sigma(R|M) = 0$  for  $R < R_{\text{mis}}^{\text{extr}}$ , and

$$\Delta\Sigma(R|M) = \Delta\Sigma_{\text{NFW}}(R|M) - \left( \frac{R_{\text{mis}}^{\text{extr}}}{R} \right)^2 \Delta\Sigma_{\text{NFW}}(R_{\text{mis}}^{\text{extr}}|M), \quad (31)$$

at a larger projected radius. This expression is numerical of the same complexity as evaluating the 2d density contrast of NFW profile  $\Delta\Sigma_{\text{NFW}}(R|M)$  but is significantly more accurate.

We then considered the average lensing efficiency

$$\Sigma_{\text{crit},l}^{-1} = \langle \Sigma_{\text{crit},ls}^{-1} \rangle_{P_l(z_s)} \quad (32)$$

of the lens  $l$ , obtained by averaging the lensing efficiency of the individual source,  $\Sigma_{\text{crit},ls}^{-1}$ , Eq. (3), with the source redshift distribution  $P_l(z_s)$ . Using also the cluster member contamination

fraction  $f_{\text{cl}}(R|\lambda, z)$ , evaluated at the mean cluster member contamination parameters, for the cluster richness  $\lambda$  and the cluster redshift  $z_{\text{cl}}$ , we defined the reduced shear model:

$$g_t^{\text{mod}}(R|M) = \frac{\Sigma_{\text{crit},l}^{-1} \Delta\Sigma(R|M)}{1 - \Sigma_{\text{crit},l}^{-1} \Sigma(R|M)} (1 - f_{\text{cl}}(R|\lambda, z_{\text{cl}})). \quad (33)$$

This extraction model also depends on the clusters detection likelihood  $\mathcal{L}_{\text{det}}$  and extent EXT, and the mean parameters of the mis-centering through the mean extraction mis-centering, Eq. (29). All these quantities are readily available for each cluster.

Any mismatches between the real profiles and the extraction model are captured by considering that the best-fit mass assuming this model, the WL mass  $M_{\text{WL}}$ , will be biased and scattered around the true halo mass.

#### 4.4. Synthetic shear profiles

The creation of the synthetic shear profiles proceeds as follows.

1. The redshift of the synthetic clusters is fixed to the redshift of the hydro-simulation outputs.

2. Assuming the mean concentration  $c_{200c}$ –mass relation by Child et al. (2018), we convert the masses  $M_{200c}$  from the simulation to  $M_{500c}$ , using the customary transformations (see for instance Ettori et al. 2011, Appendix A). Note that the concentration–mass relation employed here was calibrated on gravity-only simulations, which matches the fact that our halo masses are also gravity-only masses.

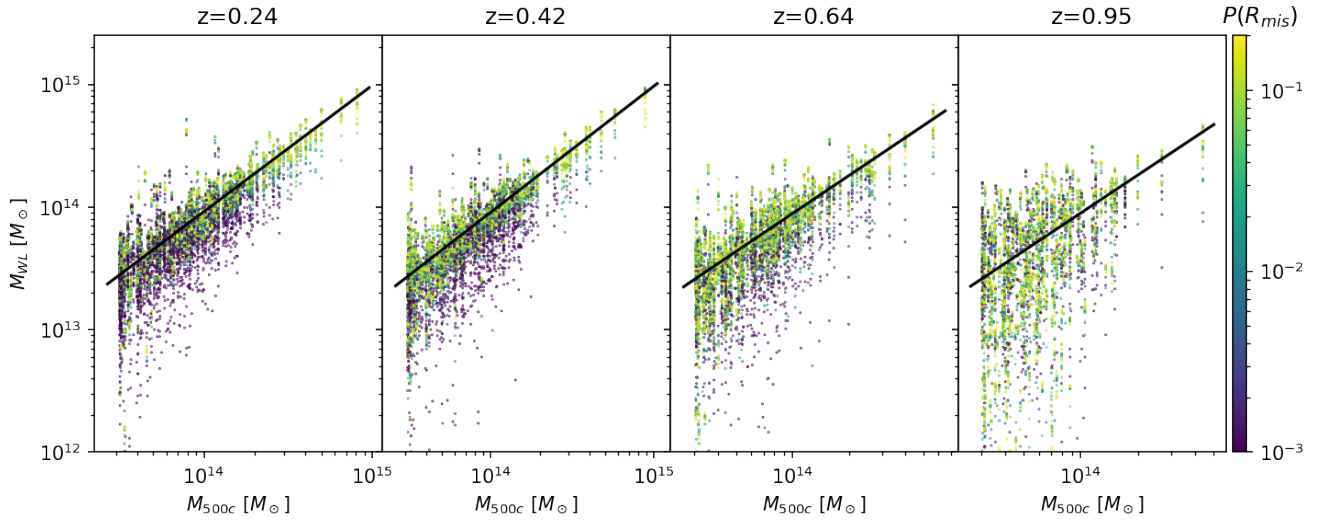
3. We assign a synthetic source redshift distribution  $P_l^{\text{synth}}(z_s)$  based on the tomographic bin weights  $w_b(z_i)$ , as defined in Eq. (7), and the realisations of their redshift distribution  $P_{\mathcal{H}}(z_s|b)$ , as provided by Myles et al. (2021), discussed in Sect. 3.3.2, and shown in Fig. 2. The normalization of these distributions takes account of the multiplicative shear bias.

4. Each cluster is assigned a richness  $\lambda$  based on the richness–mass relation calibrated in Chiu et al. (2022). For later use, we remind the reader that this relation is defined by the parameters of the richness–mass relation  $\mathbf{p}_{\text{LM}}$ . The richness is required both as an input for the extraction model, as well as to simulate the cluster member contamination,  $f_{\text{cl}}(R|\lambda, z_{\text{cl}}, \mathbf{p}_{\text{cl}})$ , on the synthetic shear profile, where  $\mathbf{p}_{\text{cl}}$  are the parameters of the cluster member contamination.

5. Querying the halos with  $|\Delta \log_{10} M_{500c}/M_{\odot}| < 0.1$  and  $|\Delta z| < 0.05$  from the eROSITA digital twin, we select the detection likelihood  $\mathcal{L}_{\text{det}}$  and extent EXT of a random pick. These two observables are needed to define the mean mis-centering needed by the extraction model and to assign a correct mis-centering distribution  $P(R_{\text{mis}}|M_{500c}, z, \mathcal{L}_{\text{det}}, \text{EXT}, \mathbf{p}_{\text{mct}})$ , resulting from the convolution of the observational and intrinsic mis-centering discussed in Sect. 4.2. Note that this step depends on the parameters of the mis-centering  $\mathbf{p}_{\text{mct}}$ , determined in the same section.

6. Improving on previous work, we now compute the reduced shear not only for each polar position in the map but also for each source redshift  $z_s$ , as presented also in Bocquet et al. (2023). The azimuthal and source redshift average are computed after the computation of the reduced shear as follows

$$g_t^{\text{synth}}(R|R_{\text{mis}}) = \int dz_s P^{\text{synth}}(z_s) \int \frac{d\phi}{2\pi} \frac{\Sigma_{\text{crit},ls}^{-1} \Gamma_t(R, \phi|R_{\text{mis}})}{1 - \Sigma_{\text{crit},ls}^{-1} \Sigma(R, \phi|R_{\text{mis}})}. \quad (34)$$



**Fig. 14.** Scatter plot of the input halo mass versus the output WL mass from one realization of the synthetic shear profiles. Color coded by the value of the mis-centering distribution for that specific halos' mis-centering, and as solid black lines, the mis-centering-distribution-weighted scaling between WL and halo mass, which results in an estimate for the WL bias. The scatter around this relation results in the WL scatter. Generally speaking, extreme outliers also have highly improbable mis-centering.

Deviating from this integration order leads to biases in the synthetic profiles of the order of 0.01, which is unacceptable given our accuracy requirements.

7. We add cluster member contamination following the fit we derived in Sect. 3.3.1, specifically the number density fit. We also add a nonlinear shear bias and extra noise from the uncorrelated large-scale structure along the line of sight, as discussed in Grandis et al. (2021a).

8. The extraction of the WL mass via the extraction model and the fit for the WL bias and scatter weighted by the mis-centering distribution follow exactly the method outlined in Grandis et al. (2021a).

The output data product of such synthetic shear profile production and subsequent fit with the extraction model is shown in Fig. 14. There, we plot the input halo masses versus the WL masses, color coded by the mis-centering distribution probability for the respective synthetic halo. Note that extreme outliers in WL mass are associated predominantly with highly mis-centered synthetic clusters and are thus quite rare, given the mis-centering distribution. As such, they contribute only marginally to the fit.

#### 4.5. WL bias and scatter

At the end of the synthetic shear profile production and WL extraction, we get a bias and scatter ( $b_l^{\text{WL}}$ ,  $\sigma_l^{\text{WL}}$ ) for each snapshot  $l$ , as well as a mass trend for both the bias ( $b_M^{\text{WL}}$ ), and the scatter ( $s_M$ ). They were obtained by fitting the relation

$$\left\langle \ln \frac{M_{\text{WL}}}{M_0} \middle| M, z_{\text{cl}} \right\rangle = b_l^{\text{WL}} + b_M^{\text{WL}} \ln \left( \frac{M}{M_0} \right) \\ \ln \sigma_{\text{WL}}^2 = s_l + s_M \ln \left( \frac{M}{M_0} \right), \quad (35)$$

with pivot mass  $M_0 = 2 \times 10^{14} M_\odot$ , and  $s_l = \ln(\sigma_l^{\text{WL}})^2$ . The fit of the relation includes cluster-by-cluster weights corresponding to the probability of the mis-centering of the individual synthetic halo, as discussed in Grandis et al. (2021a). The mean relations resulting from the fit at each redshift are shown in Fig. 14 as black lines.

**Systematic uncertainty estimation.** The uncertainty on these quantities is obtained by re-running the synthetic shear profile creation and WL mass extraction  $\mathcal{O}(1000)$  times with slightly perturbed input parameters. Specifically, we vary the hyperparameter of the DES Y3 source redshift distributions  $\mathcal{H} \in (0, 999)$  (Myles et al. 2021, and as outlined in Sect. 3.3.2), the parameters of the richness–mass relation  $\mathbf{p}_{\text{LM}}$  within the posterior reported by Chiu et al. (2022), the parameters of the cluster member contamination  $\mathbf{p}_{\text{cl}}$  within the posterior determine in Sect. 3.3.1, the parameters of the mis-centering distribution  $\mathbf{p}_{\text{mct}}$  within the posterior determine in Sect. 4.2, as well as a wide prior on the nonlinear shear biases. To these uncertainties, we add a hydrodynamical modeling uncertainty of 2% on the WL bias and corresponding values for the other parameters, as determined by Grandis et al. (2021a, Table 2).

The resulting parameter posterior distributions are presented in the contour plots in Fig. 12. We omit parameters that do not (anti)correlate with the WL bias and scatter parameters. These include, but are not limited to, the parameters of the richness mass relation  $\mathbf{p}_{\text{LM}}$ . This indicates that the weak lensing bias calibration is insensitive to the richness assignment in our synthetic simulations. Indeed, the extraction model (Eq. (33)) and the synthetic shear profiles depend in very similar ways on the richness via the cluster member contamination. To first order the assumed richness–mass relation thus cancels. Also, the nonlinear multiplicative shear bias  $\alpha_{\text{NL}}$  does not affect our systematics despite its large uncertainty. Despite the stronger shears expected in cluster lines of sight compared to the field, we thus do not require a dedicated nonlinear shear bias calibration.

We also report critical surface densities  $\Sigma_{\text{crit},l} = \langle \Sigma_{\text{crit},s,l}^{-1} \rangle_{P_s^{\text{snth}}(z_s)}^{-1}$  for each snapshot  $l$ , defined as the inverse lensing efficiency, reported in units of  $10^{15} h M_\odot \text{Mpc}^{-2}$ . This helps visualize the impact of the marginalization over the photometric redshift uncertainties, which is implemented via the hyperparameter  $\mathcal{H}$ . As can be seen in Fig. 12, the WL bias in the highest redshift snapshot  $b_4^{\text{WL}}$  anticorrelates tightly with the critical surface density  $\Sigma_{\text{crit},4} = \langle \Sigma_{\text{crit},s,l=z_4}^{-1} \rangle_{P_s^{\text{snth}}(z_s)}^{-1}$ . This indicates



that the uncertainty on this bias parameter is dominated by the uncertainty on the lensing efficiency, as induced by the photo- $z$  uncertainty. The other WL bias parameters have a more gentle correlation with the lensing efficiencies, indicating a smaller impact of photo- $z$  uncertainties in lower redshift clusters.

**Compression.** In principle, the full distribution of weak lensing biases  $b_l^{\text{WL}}$ , the logarithm weak lensing squared scatterers  $s_l$ , as well as their mass trends, should be used as priors for a mass calibration analysis of the actual data to account for the systematic uncertainties. This would, however, significantly bloat the parameter space of that likelihood analysis. As we sample the mass calibration likelihood presented in this work simultaneously with the cosmological inference from number counts in Ghirardini et al. (2024), reducing the number of parameters is of great benefit.

We compress the 4-dimensional posterior on the WL bias parameters  $b_l^{\text{WL}}$  into mean values  $\mu_{b,l}$ , and two principal components  $\delta_{b1,l}$  and  $\delta_{b2,l}$ , scaled by their variances. The other two principal components of the 4-dimensional random variable are found to not contribute significantly to the overall scatter in the WL bias posterior. This means that a realization of the WL bias at an arbitrary redshift  $z_{\text{cl}}$  can be constructed by interpolating the mean and principal components and drawing two random standard normal variates  $A_{\text{WL}}$ ,  $B_{\text{WL}}$ , such that

$$b(z_{\text{cl}}) = \mathcal{I}(z_{\text{cl}}|z_l, \mu_{b,l}) + A_{\text{WL}}\mathcal{I}(z_{\text{cl}}|z_l, \delta_{b1,l}) + B_{\text{WL}}\mathcal{I}(z_{\text{cl}}|z_l, \delta_{b2,l}), \quad (36)$$

where  $\mathcal{I}(z_{\text{cl}}|z_l, \mu_{b,l})$  is the interpolation of the data vectors  $(z_l, \mu_{b,l})$  at redshift  $z_{\text{cl}}$ . The resulting posterior on the WL bias as a function of cluster redshift is shown in the insert of Fig. 12 in blue bands for the  $1\sigma$  region, and faded lines for the  $2\sigma$  regions, together with the results for the 4 snapshots as black data points. For the 4 values of the natural logarithm of the variance, whose posterior is dominated by the uncertainty from hydrodynamical modeling, we employ a single principal component (see below, Eq. (45)). We do not compress the mass trends of the bias and the variance. These compressions reduced the number of free parameters from 10 to 5 while still capturing the relevant systematic uncertainties. As we consider 2 principal components for the weak lensing bias, we also represent the systematic uncertainty on its redshift evolution.

**Sensitivity study.** The numerical value of our WL bias is somewhat low, ranging from  $-0.05$  to  $-0.1$ . A contributing factor is the response of the WL bias to our treatment of the multiplicative shear bias. While we include the multiplicative shear bias in the synthetic shear profiles, we do not include this bias in the model to reduce the number of independent calibration products we need to communicate to the cosmological number counts analysis (Ghirardini et al. 2024). To gauge the impact of this modeling choice on the numerical value of the bias, we perform a synthetic shear and mass extraction run at the mean values of all parameters but artificially set the multiplicative shear bias  $m = 0$ , that is, we normalize the synthetic source redshift distributions to 1. This results in a shift, shown as the red line in the insert of Fig. 12 of the order of  $-0.03$ . It is two sigma away from our inferred bias values (black data points). Discounting for this effect, we would find bias values ranging from  $-0.03$  to  $-0.07$ , and thus consistent with zero bias at less than  $2\sigma$  of our systematic error budget. This qualitatively indicates that our simplified extraction model is a good representation of the much more complex synthetic data. Slightly negative WL biases can be explained by the fact that the density contrast of a halo is reduced

by the projection of nearby correlated structure, resulting in a WL mass that is lower than the halo mass (Oguri & Hamana 2011; Becker & Kravtsov 2011; Bahé et al. 2012).

We also test the sensitivity of the WL bias to the presence of the highly mis-centered tail in the eROSITA digital twin (cf. Sect. 4.2). A similar large mis-centering tail is found by Bulbul et al. (2024), Sect. 5 and Fig. 9, left panel, when comparing the centers obtained from the source detection algorithm and the X-ray post-processing. While the origin of this tail is still under investigation, we explore the hypothesis that the large mis-centering tail is completely suppressed by the X-ray post-processing. Setting the fraction  $f = 0$  fully suppresses the high exponential tail but only mildly alters the WL bias, as can be seen by the difference between the red solid and dashed line. Our WL bias numbers, and therefore the mass anchoring of the number counts experiments, are stable with respect to the presence of this highly mis-centered tail, which might be an artifact of the eROSITA digital twin processing, and be corrected by the X-ray post-processing.

**Summary.** The two main contributions to the uncertainty of the WL bias are the systematic floor resulting from the comparison of different hydrodynamical simulations established by Grandis et al. (2021a), and, at high redshift  $z_l > 0.6$  also the uncertainty of the DES Y3 photometric redshift distributions, shown in Fig. 2. Several other effects we studied, such as mis-centering and nonlinear multiplicative shear bias, do not contribute significantly to the systematic uncertainty of the mass extraction from our WL data. We compress the results of our synthetic simulations via principle component analysis, finding that 5 nuisance parameters are sufficient to represent all the systematic uncertainty trends our simulations revealed.

## 5. Mass constraints

WL measurements around galaxy clusters provide a highly accurate (see Sect. 4) mass proxy. They have, however, a quite low precision at the individual object basis, considering that the approximately 2.2 k clusters we consider collectively only reach a signal-to-noise of 65 (cf. Sect. 3.4). A natural approach to counter-act this is to perform the mass measurement on stacked lensing signals (e.g. McClintock et al. 2019; Bellagamba et al. 2019). While this simplifies the mass measurement task, it significantly complicates the modeling in a cosmological context by introducing scale-dependent selection bias effects (Sunayama et al. 2020; Dark Energy Survey Collaboration 2020; Wu et al. 2022; Sunayama 2023). In this work, we therefore followed the Bayesian Population modeling approach pioneered by Mantz et al. (2015), Bocquet et al. (2019), Grandis et al. (2019), which predicts the distribution of expected shear profiles given the selection observable as a function of the parameters governing the scaling between halo mass and selection observables. Complete derivations of this analysis framework can be found in Ghirardini et al. (2024), Bocquet et al. (2023) in the context of eROSITA-selected clusters and SPT-selected clusters, respectively. Our results here are based on the implementation described in Ghirardini et al. (2024), Sect. 4.

### 5.1. Constraints of the count rate–mass relation

In the following, we discuss the likelihood setup used for the mass calibration as well as the resulting mass calibration constraints. In general, WL mass calibration relies on statistically determining the scaling relation between the

observables and halo mass (Mantz et al. 2016; Dietrich et al. 2019; Schrabbach et al. 2021; Zohren et al. 2022, for instance). In our case, we expand this standard treatment to accommodate for the residual contamination of our sample by random superpositions of X-ray candidates and optical systems and, more importantly, misclassified active galactic nuclei (see Bulbul et al. 2024; Kluge et al. 2024; Ghirardini et al. 2024). Misclassified AGNs and random superpositions are expected to produce far less WL signal than clusters, which inhabit massive halos. We therefore employ a mixture model, a weighted sum of the different sub-populations, to perform the mass calibration in the presence of non-negligible contamination.

### 5.1.1. Mixture model

For each individual cluster, we construct the probability density function (PDF) of its shear profile  $g_t$  given the cluster observables count rate  $\hat{C}_R$ , redshift  $z_{cl}$  and sky position  $\hat{\mathcal{H}}$  as

$$P(\hat{g}_t|\hat{C}_R, z_{cl}, \hat{\mathcal{H}}) = f_C(\hat{C}_R, z_{cl}, \hat{\mathcal{H}}) P(\hat{g}_t|\hat{C}_R, z_{cl}, \hat{\mathcal{H}}, C) + f_{RS}(\hat{C}_R, z_{cl}, \hat{\mathcal{H}}) P(\hat{g}_t|RS) + f_{AGN}(\hat{C}_R, z_{cl}, \hat{\mathcal{H}}) P(\hat{g}_t|AGN, z_{cl}), \quad (37)$$

a weighted sum of the PDF of the shear profile for clusters (C), random superpositions (RS), and misclassified active galactic nuclei (AGN), with their respective frequencies  $f_\tau(\hat{C}_R, z_{cl}, \hat{\mathcal{H}})$  as a function of count rate, redshift and sky position, where  $\tau \in (C, RS, AGN)$ . Naturally,  $\sum_\tau f_\tau(\hat{C}_R, z_{cl}, \hat{\mathcal{H}}) = 1$ , where the sum runs over  $\tau \in (C, RS, AGN)$ . As outlined in Bulbul et al. (2024), Kluge et al. (2024), Ghirardini et al. (2024), the shape of the AGN and the random superpositions contributions are calibrated partially on X-ray image simulations, partially with optical follow-up of random positions and point sources, and have amplitudes ( $f_{RS}$ ,  $f_{AGN}$ ) that are fitted for using the richness distribution of the sample at a given count rate, redshift, and sky position. They find  $f_{RS} = 0.0061 \pm 0.0023$  and  $f_{AGN} = 0.0462 \pm 0.0038$ , which we adopt as priors.

For the WL modeling, we assumed that the density contrast of a random source is zero, resulting in their shear profile PDF being

$$\ln P(\hat{g}_t|RS) = -\frac{1}{2} \sum_k \frac{\hat{g}_{t,k}^2}{\delta g_{t,k}^2}, \quad (38)$$

where the sum runs over the radial bins  $k$ , with the measured shear profile  $\hat{g}_{t,k}$  defined in Eq. (22), and its shape noise error  $\delta g_{t,k}$  defined in Eq. (23). In practice, all objects in our lens sample have a richness  $\hat{\lambda} > 3$ . At the lens redshift, the random source will thus have a few red galaxies along their lines of sight. Their density contrast is, however, expected to be small compared to cluster density contrasts.

For AGNs we assume that their density contrast  $\Delta\Sigma_{AGN}(R)$  follows the profile measured by Comparat et al. (2023) for AGN in eFEDS using HSC WL data. We multiply this with the average lensing efficiency  $\Sigma_{crit,c}^{-1} = \langle \Sigma_{crit,cs}^{-1} \rangle_{P_c(z_s)}$ , evaluated as a function of cosmology for the source redshift distribution  $P_c(z_s)$  constructed in Eq. (25). The PDF thus reads

$$\ln P(\hat{g}_t|AGN, z_{cl}) = -\frac{1}{2} \sum_k \frac{1}{\delta g_{t,k}^2} \left( \hat{g}_{t,k} - \Sigma_{crit,c}^{-1} \Delta\Sigma_{AGN}(R = D_c \theta_k) \right)^2, \quad (39)$$

where  $D_c$  is the angular diameter distance to the cluster, and  $\theta_k$  the average angular distance to the sources in the respective

radial bin  $k$ , as computed in Eq. (24). In practice, the density contrast of AGN is much smaller than that of galaxy clusters, given the lower host halo mass.

For galaxy clusters, the PDF of the shear profile depends on our extraction model  $g_t^{\text{mod}}$ , and the WL mass  $M_{WL}$  defined by this model (cf. Sect. 4.3). It therefore reads

$$\ln P(\hat{g}_t|C, z_{cl}, M_{WL}) = -\frac{1}{2} \sum_k \frac{1}{\delta g_{t,k}^2} \left( \hat{g}_{t,k} - g_t^{\text{mod}}(R = D_c \theta_k | M_{WL}) \right)^2, \quad (40)$$

with the extraction model depending on a host of cluster-specific observables, as outlined above. Note that in all three PDFs, we omitted the normalization term, which is the same for all three, and just depends on  $\delta g_{t,k}$ , thus being constant with respect to the parameters of the inference.

The PDF of the shear profile for clusters needs to be marginalized over the distribution of WL masses  $M_{WL}$  compatible with the clusters count rate  $\hat{C}_R$ , redshift  $z_{cl}$ , and sky position  $\hat{\mathcal{H}}$ , that is  $P(M_{WL}|\hat{C}_R, z_{cl}, \hat{\mathcal{H}}, \mathbf{p}_{SR})$ , whose construction and dependence on model parameters  $\mathbf{p}_{SR}$  we shall discuss below. The PDF of the shear profile for the cluster component is thus

$$P(\hat{g}_t|\hat{C}_R, z_{cl}, \hat{\mathcal{H}}, C) = \int dM_{WL} P(\hat{g}_t|C, z_{cl}, M_{WL}) P(M_{WL}|\hat{C}_R, z_{cl}, \hat{\mathcal{H}}, \mathbf{p}_{SR}). \quad (41)$$

The distribution of WL masses given the count rate, redshift and sky position is proportional to

$$P(M_{WL}|\hat{C}_R, z_{cl}, \hat{\mathcal{H}}, \mathbf{p}_{SR}) \propto \iint dM dC_R P(M_{WL}, C_R|M, z_{cl}, \mathbf{p}_{SR}) P(I|C_R, z_{cl}, \hat{\mathcal{H}}_i) P(\hat{C}_R|C_R) P(M, z_{cl}), \quad (42)$$

with  $P(\hat{C}_R|C_R)$  modeling the observational uncertainty on the count rate,  $P(M, z_{cl})$  being proportional to the halo mass function, and  $P(I|C_R, z_{cl}, \hat{\mathcal{H}}_i)$  the X-ray incompleteness due to the extent selection. This ensures that we correctly model Eddington bias, which manifests as more low-mass objects scattering up to a given observable value than high-mass objects scattering down to the same observable values. This happens even when the observable–mass scatter is modeled as symmetric because there are just many more low-mass than high-mass objects. This fact is quantitatively expressed by the halo mass function. The expression given above is normalized to be a PDF in  $M_{WL}$  via integration in  $M_{WL}$  and subsequent re-scaling. See Ghirardini et al. (2024) for an alternative, but equivalent, derivation of the same likelihood.

### 5.1.2. Scaling relation

Given their deep potential wells, cluster observables scale tightly with the host halo mass. Leveraging this strong physical prior, we model the distribution of WL mass  $M_{WL}$  and intrinsic (that is noise-free) count rate  $C_R$  at given halo mass  $M$  and redshift  $z$  as a bi-variate log-normal, reading

$$P(M_{WL}, C_R|M, z_{cl}, \mathbf{p}_{SR}) = \mathcal{N}(\bar{\mu}, \bar{\Sigma}), \quad (43)$$

with mean and variance

$$\bar{\mu} = [\langle \ln C_R|M, z_{cl} \rangle, \langle \ln M_{WL}|M, z_{cl} \rangle]$$

**Table 3.** Calibration of WL bias and scatter for DES Y3 WL on eRASS1 clusters.

$z$	0.24	0.42	0.68	0.95
$\mu_b$	-0.066	-0.075	-0.101	-0.092
$\delta_{b1}$	-0.013	-0.018	-0.034	-0.092
$\delta_{b2}$	-0.019	-0.022	-0.023	0.015
$\mu_s$	-3.094	-3.232	-2.925	-1.833
$\delta_s$	0.272	0.270	0.275	0.283
$b_M = 1.030 \pm 0.021$		$s_M = -0.919 \pm 0.608$		

**Notes.** Numerical values for the WL bias ( $\mu_b$ ,  $\delta_{b1}$ ,  $\delta_{b2}$ ) and scatter ( $\mu_s$ ,  $\delta_s$ ) calibration at different redshift  $z$ , and their global mass trends  $b_M$ , and  $s_M$ , respectively.

$$\tilde{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho_{WL,X} \sigma_X \sigma_{WL} \\ \rho_{WL,X} \sigma_X \sigma_{WL} & \sigma_{WL}^2 \end{bmatrix}$$

where  $\langle \ln C_R | M, z_{cl} \rangle$  and  $\langle \ln M_{WL} | M, z_{cl} \rangle$  are the scaling relations between true mass and count rate and WL mass, respectively,  $\sigma_X$  and  $\sigma_{WL}$  their intrinsic scatters, and  $\rho_{WL,X}$  the correlation coefficient between the two intrinsic scatters. Inclusion of the latter is crucial to marginalize over possible selection effects like more concentrated halos simultaneously having a higher count rate and WL mass compared to other objects at their mass and redshift, or any other effect that might similarly correlate the intrinsic scatters.

As calibrated on dedicated simulations in Sect. 4, the mean WL mass – halo mass relation is given by

$$\left\langle \ln \frac{M_{WL}}{M_p} \middle| M, z_{cl} \right\rangle = b(z_{cl}) + b_M \ln \left( \frac{M}{M_p} \right), \quad (44)$$

with the pivot value for the mass  $M_p = 2 \times 10^{14} M_\odot$ , the redshift dependent amplitude of the WL bias given by Eq. (36), and the mass trend  $b_M = \mu_{bM} + C_{WL} \cdot \sigma_{bM}$ , with mean and standard deviation derived in Sect. 4.5 and reported in Table 3.

The scatter is similarly modeled as

$$\sigma_{WL}^2 = \exp \{ \mathcal{I}(z_{cl}|z_l, \mu_{s,l}) + D_{WL} \mathcal{I}(z_{cl}|z_l, \delta_{s,l}) \}, \quad (45)$$

where  $\mu_{s,l}$  and  $\delta_{s,l}$  are the mean and standard deviations of the WL scatter in the simulation snapshot  $l$ , derived in Sect. 4.5 and reported in Table 3, and  $\mathcal{I}(z_{cl}|z_l, \mu_{s,l})$  is the interpolation of the data vectors  $(z_l, \mu_{s,l})$  to the redshift  $z_{cl}$ . Given the tight correlations between the systematic uncertainties in the WL scatters of the different redshift seen in Fig. 12, we opt to model the WL scatter with a single principal component. We also set the mass trend of the scatter to  $s_M = 0$  as in this work, we are not interested in exploring mass trends in the intrinsic scatter of the other observables.

The count rate–mass relation is modeled as

$$\left\langle \ln \frac{C_R}{C_{R,p}} \middle| M, z_{cl} \right\rangle = \ln A_X + b_X(M, z_{cl}) \cdot \ln \frac{M}{M_p} + e_X(z_{cl}) \quad (46)$$

where  $C_{R,p} = 0.1 s^{-1}$  is the pivot value for the count rate,  $M_p = 2 \times 10^{14} M_\odot$  is the pivot value for the mass.  $b_X(M, z)$  expresses the mass-redshift dependent slope of the scaling relation, given by

$$b_X(M, z_{cl}) = \left( B_X + C_X \cdot \ln \frac{M}{M_p} + F_X \cdot \ln \frac{1 + z_{cl}}{1 + z_p} \right) \quad (47)$$

where  $B_X$  is the classic standard single value mass slope,  $C_X$  allows the slope to be mass dependent, and  $F_X$  allows the slope

to be redshift dependent, and  $z_p = 0.35$  is the pivot value for the redshift. As shown in Grandis et al. (2019), Appendix B,  $F_X$  needs to be sampled to retrieve unbiased cosmological results from the number counts. Chiu et al. (2023) explored the impact of allowing for a mass dependence in the slope in a cosmological analysis of the eFEDS sample, finding that giving the scaling relation that freedom is unnecessary. We thus set  $C_X = 0$ .

The redshift evolution  $e_X(z_{cl})$  of the X-ray scaling relation is

$$e_X(z_{cl}) = D_X \cdot \ln \frac{d_L(z_{cl})}{d_L(z_p)} + E_X \cdot \ln \frac{E(z_{cl})}{E(z_p)} + G_X \cdot \ln \frac{1 + z_{cl}}{1 + z_p} \quad (48)$$

where  $E_X = 2$  and  $D_X = -2$  are the values predicted by the self-similar model. The last term, with  $G_X$ , quantifies the deviation from the self-similar model.  $E(z_{cl}) = H(z_{cl})/H_0$  is the dimensionless Hubble factor. The intrinsic scatter of our scaling relation is represented by a single value,  $\sigma_X$ , independent of mass and redshift.

### 5.1.3. Likelihood setup and priors

As outlined in Ghirardini et al. (2024), the total likelihood of our shear profiles  $\hat{g}_{t,\kappa}$  of several clusters, given their count rate  $\hat{C}_{R,\kappa}$ , redshift  $z_{cl,\kappa}$  and sky position  $\hat{\mathcal{H}}_\kappa$ , where  $\kappa$  runs over the cluster catalog, is given by

$$\ln \mathcal{L}_{WL} = \sum_{\kappa} \ln P(\hat{g}_{t,\kappa} | \hat{C}_{R,\kappa}, z_{cl,\kappa}, \hat{\mathcal{H}}_\kappa). \quad (49)$$

In light of the derivations above, this is a function of the parameters of the WL scaling relation  $\mathbf{p}_{WL} = (A_{WL}, B_{WL}, C_{WL}, D_{WL})$ . To represent the WL anchoring performed in Sect. 4, we place independent Gaussian priors on these. The parameters of the count rate–mass relation  $\mathbf{p}_X = (A_X, B_X, F_X, G_X, \sigma_X)$ , instead, are the true target of our analysis. These are sampled within wide flat priors, reported in Table 4. Also, the correlation coefficient between the WL and count rate intrinsic scatters is left free, but within the bounds  $\rho_{WL,X} \in (-0.9, 0.9)$ . We purposefully do not sample all the way to  $\pm 1$ , as the covariance of intrinsic scatters becomes singular in that limit.

We place priors on the random source fractions and AGN ( $f_{RS}, f_{AGN}$ ) from our optical follow-up observations. We do not expect the WL to significantly inform these fractions, as the individual shear profiles are hardly significantly different from zero themselves. As such, on an individual object basis, WL just lacks the statistical power to tell clusters apart from AGN or random sources.

We fix all cosmological parameters to their Planck values, except  $\Omega_M = 0.331 \pm 0.038$  from the first Dark Energy Survey Supernovae Type Ia results (Dark Energy Survey Collaboration 2019). Several elements of our likelihood, like angular diameter distances and the extraction model, depend on cosmology via the background evolution. The exact cosmological dependence of the WL-calibrated number counts is accounted for by re-running the mass calibration presented here together with the number counts. In this work, we wish to provide mass calibration results marginalized over a reasonable range of cosmologies. Our  $\Omega_M$ -prior allowed us to do this in a physically motivated way, while remaining independent of CMB results, with which the cosmological number counts analysis might be in tension.



**Table 4.** Priors on the parameters used in the fitting of the count rate–mass relation.

Parameter	Units	Prior
Cosmology		
$\Omega_M$	–	$\mathcal{N}(0.331, 0.038)$
$\log_{10} A_S$	–	Fixed to $-8.635$
$H_0$	$\frac{\text{km}}{\text{Mpc}}$	Fixed to $68.06$
$\Omega_{b,0}$	–	Fixed to $0.048595$
$n_S$	–	Fixed to $0.9707$
$w_0$	–	Fixed to $-1$
$w_a$	–	Fixed to $0$
$\sum m_\nu$	eV	Fixed to $0$
$\Omega_{k,0}$	–	Fixed to $0$
X-ray scaling relation		
$A_X$	–	$\mathcal{U}(0.01, 3)$
$B_X$	–	$\mathcal{U}(0.1, 5)$
$C_X$	–	Fixed to $0$
$D_X$	–	Fixed to $-2$
$E_X$	–	Fixed to $2$
$F_X$	–	$\mathcal{U}(-5, 5)$
$G_X$	–	$\mathcal{U}(-5, 5)$
$\sigma_X$	–	$\mathcal{U}(0.05, 2)$
WL mass calibration		
$A_{\text{WL,DES}}$	–	$\mathcal{N}(0, 1)$
$B_{\text{WL,DES}}$	–	$\mathcal{N}(0, 1)$
$C_{\text{WL,DES}}$	–	$\mathcal{N}(0, 1)$
$D_{\text{WL,DES}}$	–	$\mathcal{N}(0, 1)$
$\rho_{\text{WL,X}}$	–	$\mathcal{U}(-0.9, 0.9)$
Contamination modeling		
$f_{\text{AGN}}$	–	$\mathcal{T}\mathcal{N}(0.0412, 0.0040, 0, \infty)$
$f_{\text{RS}}$	–	$\mathcal{T}\mathcal{N}(0.0064, 0.0024, 0, \infty)$

**Notes.** With  $\mathcal{U}(\text{min}, \text{max})$  we indicate a uniform distribution between “min” and “max”. With  $\mathcal{N}(\mu, \sigma)$  we indicate a normal distribution centered on  $\mu$  and with standard deviation  $\sigma$ . With  $\mathcal{T}\mathcal{N}(\mu, \sigma, \text{min}, \text{max})$  we indicate a truncated normal distribution centered on  $\mu$  and with standard deviation  $\sigma$  bounded between “min” and “max”.

#### 5.1.4. Scaling relation parameters constraints

We sample the total likelihood for the measured shear profiles with the priors using *ultranest* (Buchner 2021) to create a sample of the resulting posterior. We find that all parameters of interest, the parameters of the count rate–mass relation, are well constrained by our analysis (shown as blue contours in Fig. 15), while all the calibration and other nuisance parameters have posteriors consistent with their priors. We find a constraint for the amplitude of the scaling relation  $A_X = 0.88 \pm 0.20$ , when considering the mean and standard deviation computed on the posterior sample. The mass slope of the relation is constrained to  $B_X = 1.62 \pm 0.14$ , the redshift trend of the mass slope to  $F_X = -0.85 \pm 0.93$ . No significant deviation from the self-similar redshift evolution is detected, as  $G_X = -0.32 \pm 0.69$  is consistent with 0. For the intrinsic scatter, we find  $\sigma_X = 0.61 \pm 0.19$ . Noticeably, the intrinsic scatter can be constrained, despite the low precision of the individual lensing data, in line with the predictions by Grandis et al. (2019). While the intrinsic scatter values are somewhat large, they match the count rate mass scatter we find in the digital twin simulations (Comparat et al. 2020; Seppi et al. 2022). To better visualize our scaling relation results, we create a posterior predictive distribution for the mean count

rate as a function of mass and redshift. The 1 sigma region of the predicted relation at three redshifts is shown in Fig. 16 in filled bands.

#### 5.2. Goodness of fit

To assess the goodness of fit of our mass calibration, we resort to binning the cluster sample in redshift, count rate bins, with edges (0.10, 0.19, 0.27, 0.35, 0.48, 0.8), and (0.01, 0.16, 0.25, 0.30, 0.64, 20), in redshift and count rate respectively. For all clusters  $\kappa$  falling in each of the bins, we stack the tangential reduced shear profiles by defining the stacking weights  $W_\kappa = \sum_{b=2,3,4} w^b \sum_{i \in b_\kappa} w_i^s \mathcal{R}_i$ , computing

$$\hat{g}_t^{\text{stack}} = \frac{\sum_\kappa W_\kappa \hat{g}_{t,\kappa}}{\sum_\kappa W_\kappa}, \quad (50)$$

where  $i \in b_\kappa$  symbolizes the sources  $i$  associated to the cluster  $\kappa$  in the tomographic bin  $b$ . Combining this expression with Eq. (22) shows that it is equivalent to the raw tangential reduced shear estimator in Eq. (8). Using the same weighting, we also derive the error on the stacked reduced shear

$$(\delta g_t^{\text{stack}})^2 = \frac{\sum_\kappa W_\kappa^2 \delta g_{t,\kappa}^2}{\sum_\kappa W_\kappa^2}, \quad (51)$$

and the mean angular separation of the source galaxies

$$\theta^{\text{stack}} = \frac{\sum_\kappa W_\kappa \theta_\kappa}{\sum_\kappa W_\kappa}. \quad (52)$$

The resulting stacked tangential reduced shear profiles are shown as black points with their error bars in Fig. 17 given by the square root of the variance calculated above.

For the model prediction, we draw  $N_{\text{MC}} = 1000$  points  $\mathbf{p}_{\text{SR}}^v$  from the posterior sample. For each cluster  $\kappa$ , and for each posterior point  $v$ , we compute its best-guess WL mass

$$M_{\text{WL},\kappa}^v = \underset{M_{\text{WL}}}{\text{argmax}} P(M_{\text{WL}} | \hat{\mathcal{C}}_{R,\kappa}, z_\kappa, \hat{\mathcal{H}}_\kappa, \mathbf{p}_{\text{SR}}^v), \quad (53)$$

as the maximum of the PDF of WL masses, given the count rate, redshift, and sky position, evaluated for our specific posterior draw  $\mathbf{p}_{\text{SR}}^v$ . We report the median mass of the clusters in each stacking bin in Fig. 17. For each WL mass  $M_{\text{WL},\kappa}^v$  we evaluate our extraction model  $g_{t,\kappa}^{\text{mod},v}$ , given in Eq. (33) using the source redshift distribution computed on the real data, Eq. (25), and the mean positions of the sources  $\theta_\alpha$  from Eq. (24). For each cluster and posterior point, we thus construct the predicted reduced shear profile as

$$g_{t,\kappa}^{\text{pred},v} = (1 - f_{\text{RS},\kappa}^v - f_{\text{AGN},\kappa}^v) g_{t,\kappa}^{\text{mod},v} + f_{\text{AGN},\kappa}^v g_{t,\kappa}^{\text{AGN}}, \quad (54)$$

with  $f_{\text{RS},\kappa}^v$  and  $f_{\text{AGN},\kappa}^v$  the random source and AGN fractions at the count rate, redshift and sky position of the cluster  $\kappa$  given the model parameters  $\mathbf{p}_{\text{SR}}^v$ .

The average reduced shear model in the bin is then given by stacking the predicted reduced shear profile with the same weights as the data and taking a mean over the posterior points,

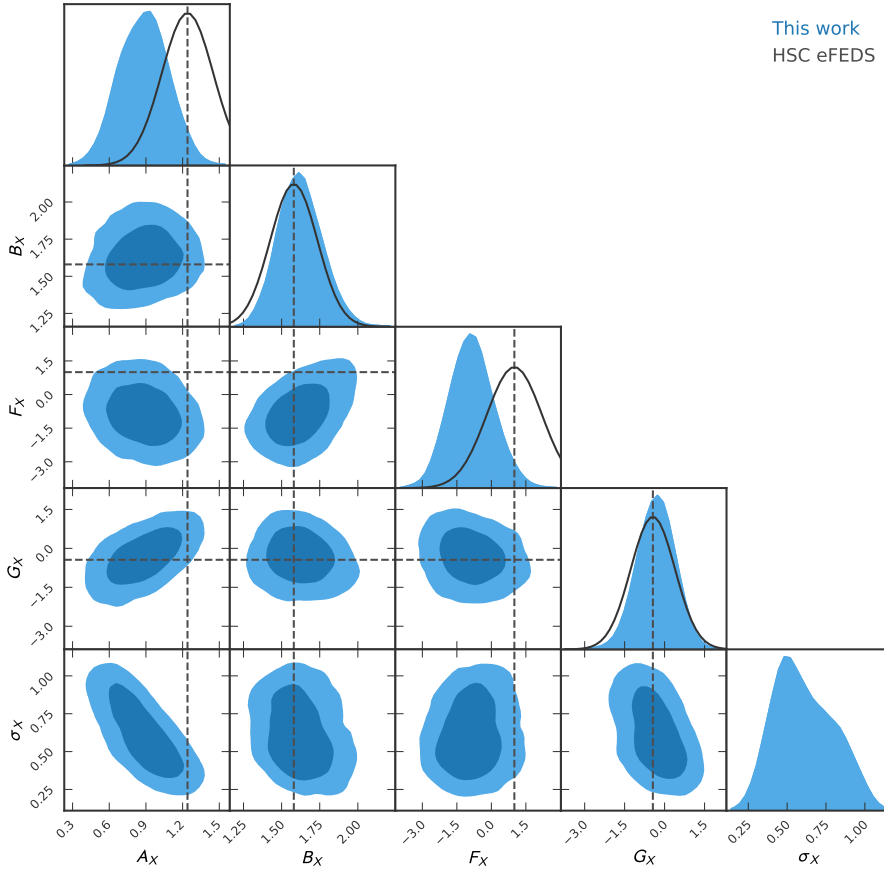
$$g_t^{\text{pred}} = \frac{1}{N_{\text{MC}}} \sum_v \frac{\sum_\kappa W_\kappa g_{t,\kappa}^{\text{pred},v}}{\sum_\kappa W_\kappa}, \quad (55)$$

and variance

$$(\delta g_t^{\text{pred}})^2 = \frac{1}{N_{\text{MC}}} \sum_v \frac{\sum_\kappa W_\kappa (g_{t,\kappa}^{\text{pred},v} - g_t^{\text{pred}})^2}{\sum_\kappa W_\kappa}, \quad (56)$$



This work  
HSC eFEDS



**Fig. 15.** One- and two-dimensional marginal posterior on the parameters of the count rate–mass scaling relations from this work in blue, compared to the 1-d marginals of the analysis of HSC WL around eFEDS clusters and groups (Chiu et al. 2022) in black. We find good agreement on the mass trend  $B_X$ , and the deviation from self-similar redshift scaling  $G_X$ , while constraints on the amplitude  $A_X$ , and the redshift evolution of the mass slope  $F_X$  are shifted by 1- to 2-sigma, consistent with statistical fluctuations. Contrary to the eFEDS-HSC analysis, we do not place an informative prior of the intrinsic scatter  $\sigma_X$ , and get only weak constraints.

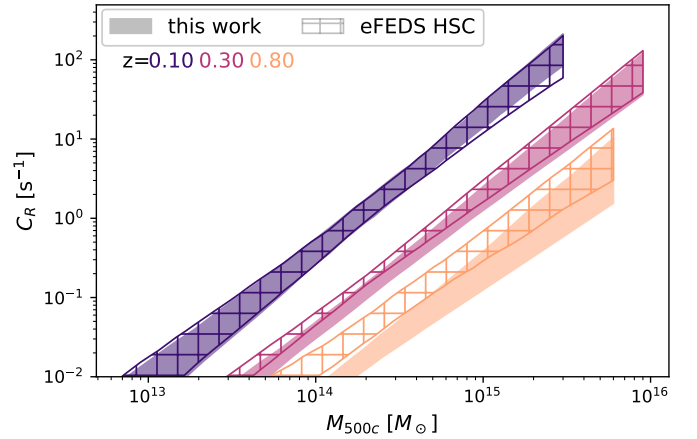
propagating the uncertainty in the model parameters expressed by their posterior to the reduced shear profile prediction. The mean shear profile and its systematic uncertainty are shown as green lines in Fig. 17. The visual impression of a good fit is corroborated by considering that the chi-squared between stacked data and predicted model is  $\chi^2 = 180.0^{+45.8}_{-30.4}$  for 150 data points and 5 parameters that are effectively constrained. The upper and lower errors for the  $\chi^2$  result from the difference induced when evaluating it with the model  $g_t^{\text{pred}} \pm \delta g_t^{\text{pred}}$ , instead of the mean model. Within the errors propagated from the posterior, we attain a very good chi-squared.

## 6. Discussion

### 6.1. Comparison to previous work

The only previous work that calibrated the eROSITA count rate – halo mass relation was performed by Chiu et al. (2022), measuring, calibrating, and analyzing the WL signal around 313 eFEDS selected clusters and groups with HSC data. Their results are shown as black lines in Fig. 15. We agree on all parameters within 2-sigma, with larger than 1-sigma deviations on the amplitude, the redshift trend of the slope and the scatter. When considering the predicted mean count rate as a function of mass and redshift, their results predict higher count rates for low-mass and high redshift objects (hatched bands in Fig. 16). The difference is, however, not statistically significant. The constraining power of this work is only marginally better than the calibration derived by Chiu et al. (2022).

The similar constraining power of eFEDS-HSC analysis to this work, despite the significantly smaller number of objects (313 compared to 2201), is explained by the higher source den-

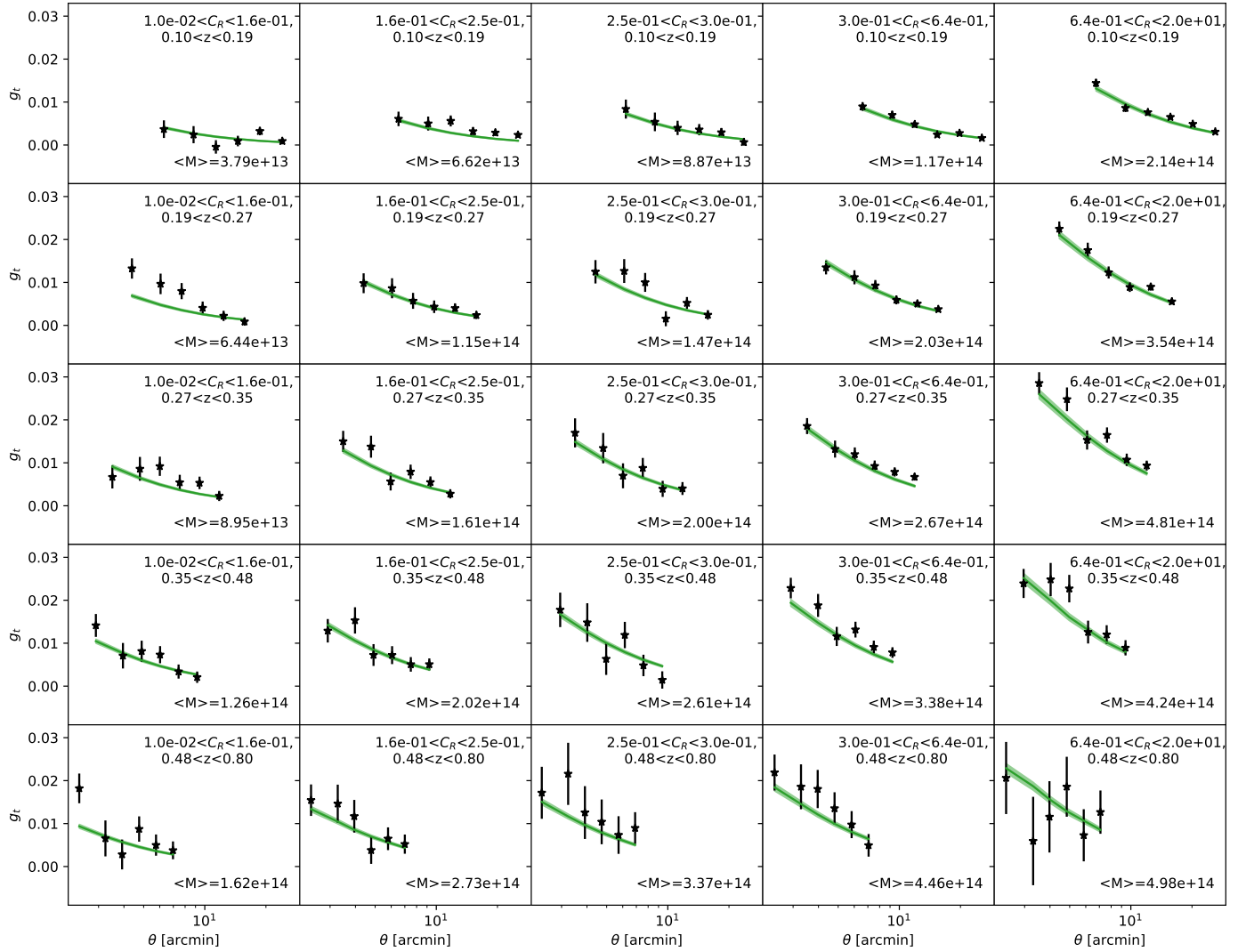


**Fig. 16.** Posterior predictive distribution of the mean count rate–mass relation (the band ranges from the 16th to the 84th percentile at each mass), for three different redshifts (color coded). Filled bands represent the analysis in this work, and hatched bands represent the eFEDS HSC result from Chiu et al. (2022). The predictions diverge for low-mass, high-redshift systems.

sity in HSC and the cleaner background selection. Indeed, the signal-to-noise ratio of an individual cluster is proportional to

$$S/N \propto \beta \sqrt{n_{\text{eff}}} \sqrt{1 - f_{\text{cl}}}, \quad \text{with } \beta = \left\langle \frac{\min(0, D_{\text{ls}})}{D_s} \right\rangle_{P(z_s)}, \quad (57)$$

where  $\beta$  is the lensing efficiency, which is 0 for sources at the lens redshift and converges to 1 for infinitely distant sources.  $n_{\text{eff}}$  is the effective source density, and  $f_{\text{cl}}$  the cluster member



**Fig. 17.** Stacked measured reduced tangential shear profiles with measurement uncertainty in black for different count rate (rows) and redshift (columns) bins. As a green line, the best-fit shear profile model and with filled green regions, the propagation of the posterior uncertainty on the mean model. The chosen model fits the data well at all count rates and redshifts. We also report the median WL mass of each bin in units of  $M_\odot$ .

contamination. When considering these elements, each cluster in the HSC analysis has a WL signal-to-noise approximately a factor of 2.3 larger than the clusters considered in this work, assuming they have the same mass and ignoring cluster member contamination. Considering the overall signal-to-noise gain stemming from the total number of clusters,  $\sqrt{2201/313} = 2.6$ , we find that it just about compensates the much lower individual clusters signal-to-noise. Note also that in the eFEDS-HSC analysis, a tighter  $\Omega_M$  and an informative  $\sigma_X$  prior were used. Taken together, this explains the similar constraining power of the two works.

## 6.2. Cluster line of sight anomalies

We find several cluster line of sight anomalies that we shall discuss here cohesively. Firstly, it is somewhat anomalous that the cluster member contamination estimates based on the effective number density of sources display a larger contaminant fraction than the estimate based on the decomposition of the source redshift distribution. The latter should be impervious to masking and obstruction effects, as argued by Gruen et al. (2014), Varga et al. (2019). Estimations of the cluster member contami-

nation via the effective number density of sources are biased low because of masking and obstruction by foreground and bright cluster sources, as shown by Kleinebreil et al. (2024). It is therefore surprising that we find the number density estimate to be larger than the source redshift distribution estimate despite it potentially being biased low. A possible source of error here is our very lenient background selection and the possible performance loss of the source redshift decomposition method if the cluster redshift lies too close to the peak of the source redshift distribution of the field. This might be hinted at by the imperfect subtraction signatures discussed in Sect. 3.3.1 and seen in Fig. 6.

Also interesting are the decrement of the effective shape noise and the increment of the smooth shear response toward the cluster center. The effect is rather constant in lens redshift, as opposed to the cluster member contamination. Indeed, we see this anomaly also in the lowest source redshift bin,  $0.1 < z_1 < 0.19$ , where we have practically no cluster member contamination. This makes it implausible that cluster member contaminants have significantly different responses or intrinsic shapes. It is worthwhile noting in Eq. (10) that a larger shear response leads to a smaller estimated effective shape noise. Also, the shear response used in this work is a smooth function of the source's

signal-to-noise and its size with respect to the PSF. Noticeably, both these quantities are magnified by the cluster’s potential – a secondary effect of WL. Such magnification effects might thus alter the inferred smooth shear response. In general, shape measurement and photo- $z$  calibration in cluster lines of sight might be affected to second order not only by magnification but also by increased blending (Hernández-Martín et al. 2020) and the intra-cluster light (Zhang et al. 2019). Exploiting future-wide optical and NIR surveys might necessitate dedicated assessments of the performance of shape and photo- $z$  measurement techniques in cluster lines of sight.

### 6.3. Impact of the mixture model

Some elements of the mixture model used to account for the residual contamination of our cluster sample (see Sect. 5.1.1) merit further discussion. Around 4.6% of our sample are found to be misclassified AGN, while another 0.6% are noise fluctuations (Bulbul et al. 2024; Kluge et al. 2024; Ghirardini et al. 2024). In both cases, we find an optical richness  $\hat{\lambda} > 3$  for these objects, suggesting that they coincide with a small overdensity of red galaxies along the line of sight. Focussing on the AGN contaminants, we assume that their WL signal is sourced by an NFW with halo mass  $M_{200c}^{\text{AGN}} = 4.95^{+2.63}_{-1.99} \times 10^{12} M_{\odot}$  as measured by Comparat et al. (2023) in HSC data around eFEDS selected AGNs.

This can be contrasted with the mean halo mass  $\log_{10} M_{200m}^{5 < \hat{\lambda} < 10} \sim 14.1 + \log_{10} M_{\odot}$  measured by McClintock et al. (2019) for redmapper selected objects with richness  $5 < \hat{\lambda} < 10$  in the DES year 1 data. The latter measurement provides a weak upper bound on the halo mass of the AGN contaminants if we assume that their optical properties are equivalent to random superposition with purely optically selected clusters of similar richness. This is, however, an overestimation, as purely optically selected clusters with richness  $5 < \hat{\lambda} < 10$  also comprise clusters and groups that we would have selected through X-rays and thus classified as “clusters” in our mixture model, especially at low redshifts.

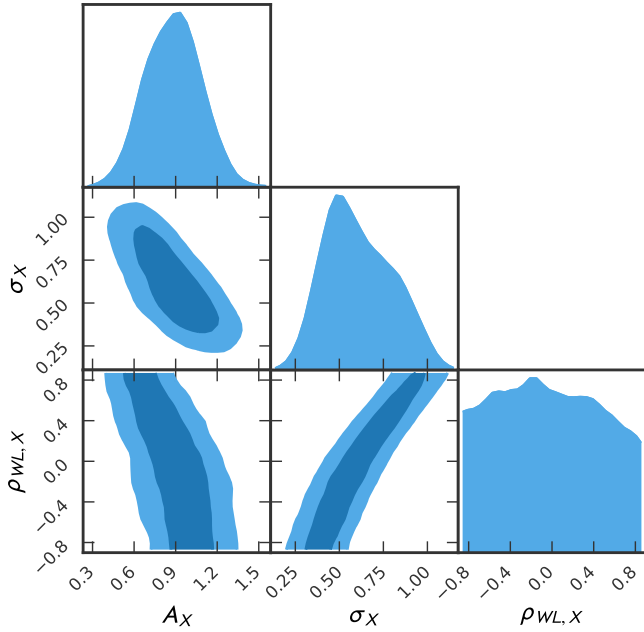
In the context of cosmological inference from WL calibrated cluster number counts, Ghirardini et al. (2024) have studied the impact of the mixture model in Appendix B.4. As a test, they treated the entire sample as “clusters” by setting the fraction of random source and misclassified AGN to zero. The resulting constraints on the cosmological parameters  $\Omega_M$  and  $\sigma_8$ , the present-day amplitude of matter fluctuations on a scale of  $8 h^{-1}$  Mpc, are only marginally affected by this modeling choice. Both these parameters are very sensitive to the mass calibration and, thus, to the implementation of the mixture model in the context of the WL signal, which we are discussing in this paper. We therefore concluded that the modeling choices for the WL signal from the contaminants do not impact the mass calibration at a level that would be noticeable in cosmological inference.

### 6.4. Accounting for cluster selection biases

Much of the astrophysically driven conversation about the viability of galaxy clusters as a cosmological probe and the limits of our ability to effectively calibrate their masses focus on so-called selection biases. When taking a forward modeling approach using Bayesian population models, these concerns can be naturally accounted for. We shall discuss in the following the most (in)famous “selection bias” as an example: the “cool core bias” (Rossetti et al. 2017). The cooling time in the center of mas-

sive galaxy clusters can fall below the Hubble time, especially if they have been undisturbed for a long time. Effective cooling boosts the central luminosity, resulting in very peaked X-ray surface brightness profiles. Conversely, the SZE selection, which is sensitive to pressure, is not affected by this. As discussed already above, taking a population modeling approach immediately ensures that any sample selected on an observable that scatters around the halo mass will preferentially contain up-scattered objects compared to down-scattered objects on account of the much higher frequency of the former. If a specific physical property drives the scatter in some observable at fixed mass, a sample selected in that observable will show more objects with that physical property compared to a sample selected on another observable. X-ray-selected samples will preferentially contain objects that are scattered up in X-rays for whatever astrophysical reason.

For cosmological inference, the complication arises if the scatter in the follow-up observable used for the mass calibration correlates with the scatter in the selection observable. A glaring case being investigated is clusters selected via the number of photometric member galaxies and the WL signal of these objects. The scatter of both depends on the concentration, orientation, and triaxiality of the halo, as well as the neighboring environment, resulting in strong selection biases hampering cosmological analyses on those samples (Dark Energy Survey Collaboration 2020; Wu et al. 2022). Our analysis is unaffected by optical selection effects. While we do impose a cut of richness  $\hat{\lambda} > 3$  (Kluge et al. 2024) in our analysis, this cut only leads to the exclusion of 1.1% of clusters at low redshift, low mass, and in high exposure time regions (Ghirardini et al. 2024, Appendix D). Our selection is thus almost exclusively based on the X-ray signatures. X-ray-selected clusters have been reported to live preferentially in nodes of the cosmic web, as compared to filaments or sheets, as found by Popesso et al. (2024) when comparing eFEDS-selected halos with spectroscopically selected ones. This means that at fixed halo mass, the cluster properties may depend on secondary halo properties like environment or concentration – an effect analogous to the assembly bias of galaxies (Wu et al. 2008; Wechsler & Tinker 2018). Our scale cuts on the tangential reduced shear are designed to ensure that we measure the WL signal only in the 1-halo regime (Grandis et al. 2021a), that is, scales dominated by the halo’s own density profile and not by neighboring structures. This already limits significantly the impact of assembly bias. Still, the eRASS1 selection might correlate with some properties that also impact the WL mass, such as concentration, triaxiality, or dynamical state. To account for any such possible effect, we introduce the correlation coefficient  $\rho_{\text{WL},X}$  among the scatters in count rate and WL mass. Figure 18 shows the posterior on the correlation coefficient, together with its degeneracies with the amplitude of the scaling relation and intrinsic scatter at fixed halo mass and redshift. The prior range for the correlation coefficient is naturally given by  $(-1, 1)$ , though we sample  $(-0.9, 0.9)$  to avoid numerical instabilities for a singular covariance matrix. Marginalizing over the correlation coefficient makes our analysis impervious to selection bias effects but deteriorates our ability to calibrate the observable–mass relation. Given that the numerical value of the amplitude of the scaling relation directly impacts our cosmological results and, crucially, their agreement with external experiments, foregoing the use of strong astrophysical priors on the selection effects at the cost of a less stringent mass calibration is undoubtedly the more prudent choice.



**Fig. 18.** One- and two-dimensional marginal contours of the amplitude of the count rate–mass scaling relation  $A_X$ , the intrinsic scatter around that relation  $\sigma_X$ , and the correlation coefficient between the WL and count rate scatter  $\rho_{M_{WL},C_R}$ , labeled as  $\rho_{WL,X}$  for brevity in the plot. While the correlation coefficient remains unconstrained, it leads to significant degeneracies with the amplitude and scatter. Astrophysical processes like cluster age, merger state, environment, orientation, triaxiality, etc., might impact both observables and lead to selection biases. We agnostically marginalize over the entire range of such effects, thereby weakening our amplitude and scatter measurements but making the mass calibration impervious to such astrophysical effects.

## 7. Conclusions

In this work, we measured and calibrated the WL signal around eRASS1 clusters and groups in the DES Y3 data. To this end, we defined a background sample of DES galaxies by weighting tomographic redshift bins defined and calibrated for the cosmic shear experiment (Myles et al. 2021; Gatti et al. 2021) as a function of the photometric redshift of our clusters. After querying sources around our lenses, out to an angular distance corresponding to  $15 h^{-1}$  Mpc, we validated and calibrated that measurement. We fit the lens richness, redshift, and source–lens distance dependence of the effective source density and the source redshift distribution to determine the contamination of our source sample. The different methods employed agree with each other within the systematic uncertainties of the source redshift and shape measurement.

We fit a lens redshift trend to the effective shape noise, allowing us to define a semi-analytical covariance metric for the shape noise. The profiles of the  $B$ -mode-like cross-component are consistent with zero, as expected for a well calibrated WL signal sourced by gravity. Investigation of the selection and shear response of our tangential reduced shear estimators show that the former can be ignored, as it induces a correction much smaller than the systematic uncertainty induced by source redshift and shape measurements. On small cluster-centric distances, we find interesting anomalies in the shear response and effective shape noise that might point at cluster line-of-sight-specific calibration issues, which arguably are too small to impact our subsequent analysis steps.

We derived cosmology-ready data products to be used as inputs for the mass calibration of eRASS1 selected clusters, both in this work and self-consistently as well as in the subsequent cosmological analysis by Ghirardini et al. (2024). These products are:

- The tangential reduced shear profile in angular bins, corrected for the shear response;
- A semi-analytical estimate of the shape noise on that profile; and
- An estimate of the field source redshift distribution from the local background

for each cluster with DES Y3 lensing data. This resulted in WL data for 2201 eRASS1 selected clusters, with a raw signal-to-noise ratio of 92, reduced to 65 after our radial scale cuts are applied.

Exploitation of these data for mass calibration and number counts cosmology requires the anchoring of the mapping between halo mass and WL signal via  $O(1000)$  realizations of synthetic shear profile simulations, following the scheme presented by Grandis et al. (2021a). These simulations include:

- Realistic 2D-projected surface mass density maps around massive halos, extracted from the cosmological hydrodynamical TNG300 simulations (Pillepich et al. 2018; Marinacci et al. 2018; Springel et al. 2018; Nelson et al. 2018, 2019; Naiman et al. 2018);
- A distribution of the off-set between the observational centers used for the tangential reduced shear profile measurements, constructed from X-ray images of the MAGNETICUM PATHFINDER simulations (Dolag et al., in prep.), and the “eROSITA digital sky twin” (Seppi et al. 2022);
- A painting of the richness using the richness–mass relation calibrated by Chiu et al. (2022);
- Dilution of the tangential reduced shear signal based on the degree of cluster member contamination we calibrated;
- A noise contribution from the uncorrelated large-scale structure; and
- Realistic DES Y3 source distributions reflecting our background selection based on weighted sums of the DES Y3 tomographic redshift bins (Myles et al. 2021), together with a prescription for nonlinear shear bias.

These synthetic shear profiles were analyzed using the same shear profile model that we used in the subsequent mass calibration. It features a simplified mis-centered NFW with a fixed concentration–mass relation and accounts for the mean cluster member contamination, as advocated by Grandis et al. (2021a). It thus has one free parameter, the WL mass. Extracting the WL mass on our synthetic shear profile simulations allowed us to establish the relation between WL and halo mass, summarized in the WL bias and scatter. Monte Carlo realizations of the synthetic shear profile simulation, obtained by varying all the inputs within our expectation, resulted in an uncertainty estimate on the WL bias and scatter that acts as a calibration prior to the subsequent mass calibration. We find that at low redshift, our mass extraction uncertainty is dominated by the effect of hydrodynamical modeling, leading to a systematics floor of 2%. At higher redshifts, DES photo- $z$  uncertainties become incrementally more important.

We calibrated the count rate to halo mass relation by using a Bayesian population model to forward model the probability density function of each cluster’s shear profile given its count rate and redshift. Improving on previous work (Mantz et al. 2016; Dietrich et al. 2019; Bocquet et al. 2019; Chiu et al. 2022), we took explicit account of contaminants via a mixture model. We find that we can measure five parameters



of the count rate–mass relation: the amplitude, the slope and its redshift dependence, the deviation from a self-similar redshift evolution, and the intrinsic scatter around the relation. We assessed the quality of this population model-based fit on stacked signals in count rate and redshift bins and found excellent agreement between the data and the model. Our analysis finds comparable and marginally tighter constraints than the calibration of the eROSITA count rate–mass relation performed by Chiu et al. (2022) with HSC data on eFEDS-selected clusters.

This analysis will serve, along with similar analyses on HSC and KiDS, as the main mass calibration for the number counts of eRASS1-selected clusters (Ghirardini et al. 2024), which constrain different cosmological parameters such as the present-day matter density, the amplitude of matter fluctuations, the equation of state of dark energy, and the sum of the neutrino masses. Methodologically, we have demonstrated how photometric data from wide surveys, calibrated specifically for cosmic shear experiments, can be effectively used to perform a WL mass calibration of cluster number counts. These two experiments are highly complementary. Cosmic shear probes the large-scale structure’s linear and mildly nonlinear scale, while WL-calibrated number counts probe the most nonlinear regions and massive galaxy clusters. These probes have very different sensitivities to astrophysical and observational systematics, only sharing a joint dependence on the source photo- $z$  and shape calibration. In light of upcoming deep and wide surveys, tailored mainly for cosmic shear, including *Euclid*<sup>5</sup>, LSST<sup>6</sup>, or *Roman*<sup>7</sup>, this work provides a demonstration of how to leverage these data to inform cluster number counts, enabling independent and competitive cosmological constraints.

**Acknowledgements.** This work is based on data from eROSITA, the soft X-ray instrument aboard SRG, a joint Russian-German science mission supported by the Russian Space Agency (Roskosmos), in the interests of the Russian Academy of Sciences represented by its Space Research Institute (IKI), and the Deutsches Zentrum für Luft und Raumfahrt (DLR). The SRG spacecraft was built by Lavochkin Association (NPOL) and its subcontractors and is operated by NPOL with support from the Max Planck Institute for Extraterrestrial Physics (MPE). The development and construction of the eROSITA X-ray instrument was led by MPE, with contributions from the Dr. Karl Remeis Observatory Bamberg & ECAP (FAU Erlangen-Nürnberg), the University of Hamburg Observatory, the Leibniz Institute for Astrophysics Potsdam (AIP), and the Institute for Astronomy and Astrophysics of the University of Tübingen, with the support of DLR and the Max Planck Society. The Argelander Institute for Astronomy of the University of Bonn and the Ludwig Maximilians Universität Munich also participated in the science preparation for eROSITA. The eROSITA data shown here were processed using the eSASS software system developed by the German eROSITA consortium. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities, the Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich,

Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the OzDES Membership Consortium, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. E.B., A.L., V.G., and X.Z. acknowledge financial support from the European Research Council (ERC) Consolidator Grant under the European Union’s Horizon 2020 research and innovation programme (grant agreement CoG Dark-Quest No 101002585). L.K. acknowledges support by the COMPLEX project from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program grant agreement ERC-2019-AdG 882679. N.O. acknowledges JSPS KAKENHI Grant Number JP19KK0076. The Innsbruck group acknowledges support from the German Research Foundation (DFG) under grant 415537506, as well as the Austrian Research Promotion Agency (FFG) and the Federal Ministry of the Republic of Austria for Climate Action, Environment, Mobility, Innovation and Technology (BMK) via grants 899537 and 900565.

## References

- Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, *ARA&A*, 49, 409  
 Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, *MNRAS*, 426, 2046  
 Bahé, Y. M., McCarthy, I. G., & King, L. J. 2012, *MNRAS*, 421, 1073  
 Becker, M. R., & Kravtsov, A. V. 2011, *ApJ*, 740, 25  
 Bellagamba, F., Sereno, M., Roncarelli, M., et al. 2019, *MNRAS*, 484, 1598  
 Benítez, N. 2000, *ApJ*, 536, 571  
 Bleem, L. E., Bocquet, S., Stalder, B., et al. 2020, *ApJS*, 247, 25  
 Bocquet, S., Dietrich, J. P., Schrabback, T., et al. 2019, *ApJ*, 878, 55  
 Bocquet, S., Grandis, S., Bleem, L. E., et al. 2023, *Phys. Rev. D*, submitted [arXiv:2310.12213]  
 Buchner, J. 2021, *J. Open Source Softw.*, 6, 3001  
 Bulbul, E., Liu, A., Kluge, M., et al. 2024, *A&A*, 685, A106  
 Castro, T., Borgani, S., Dolag, K., et al. 2021, *MNRAS*, 500, 2316  
 Child, H. L., Habib, S., Heitmann, K., et al. 2018, *ApJ*, 859, 55  
 Chiu, I. N., Ghirardini, V., Liu, A., et al. 2022, *A&A*, 661, A11  
 Chiu, I. N., Klein, M., Mohr, J., & Bocquet, S. 2023, *MNRAS*, 522, 1601  
 Clerc, N., Comparat, J., Seppi, R., et al. 2024, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202449447>  
 Comparat, J., Merloni, A., Salvato, M., et al. 2019, *MNRAS*, 487, 2005  
 Comparat, J., Eckert, D., Finoguenov, A., et al. 2020, *Open J. Astrophys.*, 3, 13  
 Comparat, J., Luo, W., Merloni, A., et al. 2023, *A&A*, 673, A122  
 Cordero, J. P., Harrison, I., Rollins, R. P., et al. 2022, *MNRAS*, 511, 2170  
 Dark Energy Survey Collaboration (Abbott, T. M. C., et al.) 2019, *ApJ*, 872, L30  
 Dark Energy Survey Collaboration (Abbott, T. M. C., et al.) 2020, *Phys. Rev. D*, 102, 023509  
 Dark Energy Survey Collaboration (Abbott, T. M. C., et al.) 2022, *Phys. Rev. D*, 105, 023520  
 De Vicente, J., Sánchez, E., & Sevilla-Noarbe, I. 2016, *MNRAS*, 459, 3078  
 Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, 157, 168  
 Dietrich, J. P., Bocquet, S., Schrabback, T., et al. 2019, *MNRAS*, 483, 2871  
 Dyson, F. W., Eddington, A. S., & Davidson, C. 1920, *Philos. Trans. R. Soc. Lond. Ser. A*, 220, 291  
 Ettori, S., Gastaldello, F., Leccardi, A., et al. 2011, *A&A*, 526, 1  
 Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *AJ*, 150, 150  
 Friedrich, O., Andrade-Oliveira, F., Camacho, H., et al. 2021, *MNRAS*, 508, 3125  
 Gatti, M., Sheldon, E., Amon, A., et al. 2021, *MNRAS*, 504, 4312  
 Ghirardini, V., Bulbul, E., Artis, E., et al. 2024, *A&A*, submitted [arXiv:2402.08458]  
 Grandis, S., Mohr, J. J., Dietrich, J. P., et al. 2019, *MNRAS*, 488, 2041  
 Grandis, S., Klein, M., Mohr, J. J., et al. 2020, *MNRAS*, 498, 771  
 Grandis, S., Bocquet, S., Mohr, J. J., Klein, M., & Dolag, K. 2021a, *MNRAS*, 507, 5671  
 Grandis, S., Mohr, J. J., Costanzi, M., et al. 2021b, *MNRAS*, 504, 1253  
 Gruen, D., Seitz, S., Brimiouille, F., et al. 2014, *MNRAS*, 442, 1507  
 Haiman, Z., Mohr, J. J., & Holder, G. P. 2001, *ApJ*, 553, 545  
 Hennig, C., Mohr, J. J., Zenteno, A., et al. 2017, *MNRAS*, 467, 4015  
 Hernández-Martín, B., Schrabback, T., Hoekstra, H., et al. 2020, *A&A*, 640, A117  
 Hoekstra, H., Mahdavi, A., Babul, A., & Bildfell, C. 2012, *MNRAS*, 427, 1298

<sup>5</sup> <http://www.euclid-ec.org/>

<sup>6</sup> <https://www.lsst.org/>

<sup>7</sup> <https://roman.gsfc.nasa.gov/>

- Huff, E., & Mandelbaum, R. 2017, ArXiv e-prints [arXiv:1702.02600]
- Huterer, D. 2023, *A&ARv*, 31, 2
- Jarvis, M., Bernstein, G. M., Amon, A., et al. 2021, *MNRAS*, 501, 1282
- Kimmig, L. C., Remus, R.-S., Dolag, K., & Biffi, V. 2023, *ApJ*, 949, 92
- Kleinebreil, F., Grandis, S., Schrabback, T., et al. 2024, *A&A*, submitted [arXiv:2402.08456]
- Kluge, M., Comparat, J., Liu, A., et al. 2024, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202349031>
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, *ApJS*, 192, 18
- MacCrann, N., Becker, M. R., McCullough, J., et al. 2022, *MNRAS*, 509, 3371
- Majumdar, S., & Mohr, J. J. 2004, *ApJ*, 613, 41
- Mantz, A. B., von der Linden, A., Allen, S. W., et al. 2015, *MNRAS*, 446, 2205
- Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2016, *MNRAS*, 463, 3582
- Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, *MNRAS*, 480, 5113
- McClintock, T., Varga, T. N., Gruen, D., et al. 2019, *MNRAS*, 482, 1352
- Merloni, A., Lamer, G., Liu, T., et al. 2024, *A&A*, 682, A34
- Mohr, J. J., & Evrard, A. E. 1997, *ApJ*, 491, 38
- Myles, J., Alarcon, A., Amon, A., et al. 2021, *MNRAS*, 505, 4249
- Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 477, 1206
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
- Nelson, D., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 475, 624
- Nelson, D., Springel, V., Pillepich, A., et al. 2019, *Comput. Astrophys. Cosmol.*, 6, 2
- Oguri, M., & Hamana, T. 2011, *MNRAS*, 414, 1851
- Paulus, M. 2021, Ph.D. Thesis, Ludwig-Maximilians University of Munich, Germany
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *MNRAS*, 475, 648
- Popesso, P., Biviano, A., Bulbul, E., et al. 2024, *MNRAS*, 527, 895
- Power, C., Navarro, J. F., Jenkins, A., et al. 2003, *MNRAS*, 338, 14
- Prat, J., Blazek, J., Sánchez, C., et al. 2022, *Phys. Rev. D*, 105, 083528
- Predehl, P., Andriatchke, R., Arefiev, V., et al. 2021, *A&A*, 647, A1
- Ragagnin, A., Saro, A., Singh, P., & Dolag, K. 2021, *MNRAS*, 500, 5056
- Rossetti, M., Gastaldello, F., Eckert, D., et al. 2017, *MNRAS*, 468, 1917
- Rozo, E., Rykoff, E. S., Becker, M., Reddick, R. M., & Wechsler, R. H. 2015, *MNRAS*, 453, 38
- Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, *ApJ*, 785, 104
- Rykoff, E. S., Rozo, E., Hollowood, D., et al. 2016, *ApJS*, 224, 1
- Saro, A., Bocquet, S., Rozo, E., et al. 2015, *MNRAS*, 454, 2305
- Schneider, P., 2006, in *Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro*, eds. G. Meylan, P. Jetzer, P. North, et al., 269
- Schrabback, T., Bocquet, S., Sommer, M., et al. 2021, *MNRAS*, 505, 3923
- Seppi, R., Comparat, J., Bulbul, E., et al. 2022, *A&A*, 665, A78
- Sevilla-Noarbe, I., Bechtol, K., Carrasco Kind, M., et al. 2021, *ApJS*, 254, 24
- Sheldon, E. S., & Huff, E. M. 2017, *ApJ*, 841, 24
- Shin, T., Jain, B., Adhikari, S., et al. 2021, *MNRAS*, 507, 5758
- Sommer, M. W., Schrabback, T., Ragagnin, A., & Rockenfelder, R. 2023, *MNRAS*, submitted [arXiv:2306.13187]
- Song, J., Zenteno, A., Stalder, B., et al. 2012, *ApJ*, 761, 22
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, 475, 676
- Story, K., Aird, K. A., Andersson, K., et al. 2011, *ApJ*, 735, L36
- Sunayama, T. 2023, *MNRAS*, 521, 5064
- Sunayama, T., Park, Y., Takada, M., et al. 2020, *MNRAS*, 496, 4468
- Sunyaev, R., Arefiev, V., Babushkin, V., et al. 2021, *A&A*, 656, A132
- Umetsu, K. 2020, *A&ARv*, 28, 7
- Varga, T. N., DeRose, J., Gruen, D., et al. 2019, *MNRAS*, 489, 2511
- Wechsler, R. H., & Tinker, J. L. 2018, *ARA&A*, 56, 435
- Wu, H.-Y., Rozo, E., & Wechsler, R. H. 2008, *ApJ*, 688, 729
- Wu, H.-Y., Costanzi, M., To, C.-H., et al. 2022, *MNRAS*, 515, 4471
- Zhang, Y., Yanny, B., Palmese, A., et al. 2019, *ApJ*, 874, 165
- Zohren, H., Schrabback, T., Bocquet, S., et al. 2022, *A&A*, 668, A18
- <sup>7</sup> Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- <sup>8</sup> Physics Department, University of Wisconsin-Madison, Madison, WI 53706, USA
- <sup>9</sup> Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA
- <sup>10</sup> Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
- <sup>11</sup> Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15312, USA
- <sup>12</sup> Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain
- <sup>13</sup> Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil
- <sup>14</sup> Universidad de La Laguna, Dpto. Astrofísica, 38206 La Laguna, Tenerife, Spain
- <sup>15</sup> Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA
- <sup>16</sup> Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA
- <sup>17</sup> Physics Department, William Jewell College, Liberty, MO 64068, USA
- <sup>18</sup> Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA
- <sup>19</sup> Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
- <sup>20</sup> Department of Physics, Duke University Durham, Durham, NC 27708, USA
- <sup>21</sup> Department of Physics, National Cheng Kung University, 70101 Tainan, Taiwan
- <sup>22</sup> NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA
- <sup>23</sup> IRAP, Université de Toulouse, CNRS, UPS, CNES, Toulouse, France
- <sup>24</sup> Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
- <sup>25</sup> School of Mathematics and Physics, University of Queensland, Brisbane QLD 4072, Australia
- <sup>26</sup> Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
- <sup>27</sup> Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA
- <sup>28</sup> NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA
- <sup>29</sup> Université Grenoble Alpes, CNRS, LPSC-IN2P3, 38000 Grenoble, France
- <sup>30</sup> Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
- <sup>31</sup> Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA
- <sup>32</sup> SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
- <sup>33</sup> Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
- <sup>34</sup> Department of Physics and Astronomy, University of Sussex, Pevensey Building, Brighton BN1 9QH, UK
- <sup>35</sup> School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA, UK
- <sup>36</sup> Department of Astronomy, University of Geneva, ch. d'Écogia 16, 1290 Versoix, Switzerland
- <sup>37</sup> Kavli Institute for Particle Astrophysics & Cosmology, Stanford University, PO Box 2450, Stanford, CA 94305, USA
- <sup>38</sup> Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK
- <sup>39</sup> Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
- <sup>40</sup> Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859 Campinas, SP, Brazil

<sup>1</sup> Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria  
e-mail: sebastian.grandis@uibk.ac.at

<sup>2</sup> Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse 1, 85748 Garching, Germany

<sup>3</sup> Faculty of Physics, Ludwig-Maximilians-Universität München, Scheinerstr. 1, 81679 Munich, Germany

<sup>4</sup> Argelander-Institut für Astronomie (AIfA), Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

<sup>5</sup> Linköpings Universitet, Institutionen för Systemteknik, Linköpings Universitet, Linköping 581 83, Sweden

<sup>6</sup> Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

- <sup>41</sup> Physics Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
- <sup>42</sup> Hiroshima Astrophysical Science Center, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
- <sup>43</sup> Core Research for Energetic Universe, Hiroshima University, 1-3-1, Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
- <sup>44</sup> Department of Physics, University of Genova and INFN, Via Dodecaneso 33, 16146 Genova, Italy
- <sup>45</sup> Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
- <sup>46</sup> Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
- <sup>47</sup> Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA
- <sup>48</sup> Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA
- <sup>49</sup> Département de Physique Théorique and Center for Astroparticle Physics (CAP), University of Geneva, 24 quai Ernest Ansermet, 1211 Genève, Switzerland
- <sup>50</sup> Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
- <sup>51</sup> Department of Physics, Boise State University, Boise, ID 83725, USA
- <sup>52</sup> McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
- <sup>53</sup> Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
- <sup>54</sup> Astronomy Unit, Department of Physics, University of Trieste, Via Tiepolo 11, 34131 Trieste, Italy
- <sup>55</sup> INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
- <sup>56</sup> Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy
- <sup>57</sup> Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA
- <sup>58</sup> Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
- <sup>59</sup> Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany
- <sup>60</sup> Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India
- <sup>61</sup> Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK
- <sup>62</sup> Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029 Blindern, 0315 Oslo, Norway
- <sup>63</sup> Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain
- <sup>64</sup> Department of Physics, The Ohio State University, Columbus, OH 43210, USA
- <sup>65</sup> Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
- <sup>66</sup> George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA
- <sup>67</sup> Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil
- <sup>68</sup> School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK



## Appendix A: Effective shape noise and number density in the presence of shear responses

The aim of this appendix is to present a derivation for the expression of the effective shape variance, Eq. 10 and the effective number density, Eq. 9, in the presence of a shear response  $\mathcal{R}$ .

Friedrich et al. (2021) demonstrate that the variance of a reduced shear profile  $\hat{g}_t$  is given by

$$\text{Var}[\hat{g}_t] = \frac{\sigma_{\text{eff}}^2}{N_{\text{pair}}} \text{ with } \sigma_{\text{eff}}^2 = \frac{1}{2N_s} \sum |e|^2 w_n^2, \quad (\text{A.1})$$

where  $N_{\text{pair}}$  is the number of source lens pairs,  $N_s$  the number of sources,  $|e|^2 = e_r^2 + e_\chi^2$  the squared modulus of the ellipticity when interpreted as a complex number, and  $w_n$  normalized weights.

To generalize this expression to un-normalized weight  $w$ , just consider that such un-normalized weights can always be rescaled to normalized weights with the ansatz  $w_n = Cw$ . The normalization condition then readily leads to an expression for  $C$ , reading

$$\frac{1}{N_s} \sum w_n = 1 = \frac{C}{N_s} \sum w \Leftrightarrow C^{-1} = \frac{1}{N_s} \sum w. \quad (\text{A.2})$$

Moving into the regime of individual cluster lensing, where  $N_{\text{pair}} = N_s$ , combining all the above equations, we find

$$\text{Var}[\hat{g}_t] = \frac{1}{2} \frac{1}{(\sum w)^2} \sum |e|^2 w^2, \quad (\text{A.3})$$

for generic, un-normalized weights.

When considering shear data with shear responses  $\mathcal{R}$ , Friedrich et al. (2021), Appendix D suggests transforming the above equations by re-scaling both the ellipticities  $e$ , as well as the weight  $w$ , as follows

$$\tilde{e} = \frac{e}{\mathcal{R}} \text{ and } \tilde{w} = \mathcal{R}w, \quad (\text{A.4})$$

thus correcting the ellipticity for the shear response and adjusting the weights for the necessary re-scaling. Note that the summand in the numerator of Eq. A.3 is invariant under this re-scaling,  $|\tilde{e}|^2 \tilde{w}^2 = |e|^2 w^2$ , and we thus only have to modify the summand in the denominator. Given that the above expression is valid for un-normalized weights, no further corrections are required.

Dropping the tildes, we can readily write the variance of a tangential reduced shear profile in the presence of shear response as

$$\text{Var}[\hat{g}_t] = \frac{1}{2} \frac{1}{(\sum w\mathcal{R})^2} \sum |e|^2 w^2. \quad (\text{A.5})$$

Similarly, this justifies also why the source redshift distribution needs to be shear response weighted.

Interpreting this expression as the ratio between an effective shape dispersion  $\sigma_{\text{eff},\alpha}$  for  $\alpha \in (t, x)$ , and an effective number of source  $N_{\text{eff}}$  we can recover the above expression by using

$$N_{\text{eff}} = \frac{(\sum w\mathcal{R})^2}{\sum w^2 \mathcal{R}^2} \text{ and } \sigma_{\text{eff},\alpha}^2 = \frac{\sum e_\alpha^2 w^2}{\sum w^2 \mathcal{R}^2} \text{ for } \alpha \in (t, x), \quad (\text{A.6})$$

which are the expressions that we also use for our shape noise modeling.

## Appendix B: Parametrisation of cluster member contamination

We shall justify in the following the parameterization used for the cluster member contamination in Section 3.3.1. As shown by Hennig et al. (2017), cluster member galaxies follow an NFW profile, with the total number of galaxies enclosed in the virial region correlating tightly with mass, while the concentration varies depending on the galaxy type. Red-sequence galaxies have a larger concentration, which is more similar to the concentration of dark matter than non-red-sequence galaxies. We therefore assumed that the cluster member galaxies that contaminated our background sample also follow an NFW profile.

The fraction of red-sequence galaxies in the cluster member contamination is unknown, though it stands to argue that their more precise photometric redshifts make it easier to exclude them with a background selection. We therefore let the concentration  $c$  of the cluster member contaminants profile as a free parameter.

Similarly, we choose an agnostic approach to the redshift trend of the cluster member contamination. We simply assume that the ratio between the number density of field galaxies  $n_{\text{field}}(z)$  and cluster member contaminants is a smooth function of redshift. We therefore employ a non-parametric fit of the amplitudes  $A_{z_i}$  for pivot redshifts  $z_i$ .

To account for possible mass trends, we use the richness  $\lambda$  as a mass proxy, given that we have to fit for the cluster member contamination prior to performing a mass calibration and therefore do not have WL-calibrated mass estimates at our disposal. For ICM-selected cluster samples, richness is known to scale approximately linearly with halo mass,  $\lambda \sim M$  (Saro et al. 2015; Bleem et al. 2020; Grandis et al. 2020, 2021b). We therefore used the richness to define an approximate virial radius for the NFW profile of the cluster member contaminants as  $R_{\text{vir}} \propto (\lambda/20)^{1/3}$ . We also allowed the amplitude of the cluster member contamination to vary with richness.

Taken together, the above considerations lead to the ansatz

$$n_{\text{icl}}(\lambda, z, R) = A(\lambda, z, R) n_{\text{field}}(z) \text{ with}$$

$$A(\lambda, z, R) = A_j \left( \frac{\lambda}{25} \right)^{B_\lambda} \Sigma_{\text{NFW}}^{\text{norm}} \left( R \middle| r_s = c^{-1} \left( \frac{\lambda}{20} \right)^{1/3} \right). \quad (\text{B.1})$$

The total number density results from the sum of the cluster member contaminants' number density and the field number density as

$$n_{\text{tot}} = n_{\text{icl}}(\lambda, z, R) + n_{\text{field}}(z) = [A(\lambda, z, R) + 1] n_{\text{field}}(z), \quad (\text{B.2})$$

while the fraction of cluster member contaminants is

$$f_{\text{icl}}(\lambda, z, R) = \frac{n_{\text{icl}}}{n_{\text{tot}}} = \frac{A(\lambda, z, R)}{A(\lambda, z, R) + 1}, \quad (\text{B.3})$$

thus justifying the parametrisation chosen in Eq. 12. This parametrization has the benefit that for  $A(\lambda, z, R) > 0$ ,  $0 < f_{\text{icl}}(\lambda, z, R) < 1$ , as already noted by Grandis et al. (2021b) in the context of the richness-dependent contamination of photometrically selected cluster samples.

Finally, one can show that  $1 + A = (1 - f_{\text{icl}})^{-1}$ , thus also motivating Eq. 14.