



Universiteit
Leiden
The Netherlands

Weak lensing constraints on the stellar-to-halo mass relation of galaxy groups with simulation-informed scatter

Li, S.-S.; Hoekstra, H.; Kuijken, K.; Schaller, M.; Schaye, J.

Citation

Li, S. -S., Hoekstra, H., Kuijken, K., Schaller, M., & Schaye, J. (2025). Weak lensing constraints on the stellar-to-halo mass relation of galaxy groups with simulation-informed scatter. *Astronomy And Astrophysics*, 700. doi:10.1051/0004-6361/202452892

Version: Accepted Manuscript

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4288620>

Note: To cite this publication please use the final published version (if applicable).

Weak lensing constraints on the stellar-to-halo mass relation of galaxy groups with simulation-informed scatter

Shun-Sheng Li^{1,2}, Henk Hoekstra¹, Konrad Kuijken¹, Matthieu Schaller^{1,3}, and Joop Schaye¹

¹ Leiden Observatory, Leiden University, Einsteinweg 55, 2333 CC Leiden, the Netherlands
e-mail: sqli@strw.leidenuniv.nl

² Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France

³ Lorentz Institute for Theoretical Physics, Leiden University, PO Box 9506, 2300 RA Leiden, the Netherlands

Received 5 November 2024 / Accepted 7 July 2025

ABSTRACT

Understanding the scaling relation between baryonic observables and dark matter halo properties is crucial not only for studying galaxy formation and evolution, but also for deriving accurate cosmological constraints from galaxy surveys. In this paper, we constrain the stellar-to-halo mass relation of galaxy groups identified by the Galaxy and Mass Assembly survey, using weak lensing signals measured by the Kilo-Degree Survey. We compare our measured scaling relation with predictions from the FLAMINGO hydrodynamical simulations and the L-GALAXIES semi-analytical model. We find a general agreement between our measurements and simulation predictions for halos with masses $\geq 10^{13.5} h_{70}^{-1} M_{\odot}$, but observe slight discrepancies with the FLAMINGO simulations at lower halo masses. We explore improvements to the current halo model framework by incorporating simulation-informed scatter in the group stellar mass distribution as a function of halo mass. We find that including a simulation-informed scatter model tightens the constraints on scaling relations, despite the current data statistics being insufficient to directly constrain the variable scatter. We also test the robustness of our results against different statistical models of miscentring effects from selected central galaxies. We find that accounting for miscentring is essential, but our current measurements do not distinguish among different miscentring models.

Key words. cosmology: observations – dark matter – galaxies: groups: general – gravitational lensing: weak – methods: statistical – surveys

1. Introduction

According to the current standard model of cosmology, galaxies form within cold dark matter halos, which originate from small initial density perturbations amplified by gravitational instability. This framework predicts a strong correlation between properties of galaxies and their host dark matter halos (see [Wechsler & Tinker 2018](#), for a review). Dark matter halos dominate the local gravitational potential and provide the environment for galaxy formation and evolution (e.g. [Blumenthal et al. 1984](#); [Davis et al. 1985](#)). Conversely, various baryonic processes associated with galaxy formation and evolution, especially the energetic feedback processes from supernovae and active galactic nuclei (AGNs), impact the matter distribution on small scales (e.g. [van Daalen et al. 2011](#); [Hellwing et al. 2016](#); [Chisari et al. 2018](#); [van Daalen et al. 2020](#)). Therefore, a comprehensive understanding of the galaxy-halo connection is crucial not only for studying galaxy formation and evolution, but also for ensuring the accuracy of cosmological constraints inferred from observations of large-scale structures (e.g. [Semboloni et al. 2011](#); [Schneider et al. 2020](#); [Castro et al. 2021](#); [Debackere et al. 2021](#)).

Given that dark matter halos typically host multiple galaxies, galaxy groups and clusters play a central role in studying the galaxy-halo connection. Although massive galaxy clusters serve as a powerful tool for constraining cosmological models, they are rare and represent extreme conditions in the Universe (see [Allen et al. 2011](#), for a review). In contrast, galaxy groups, which host the majority of present-day galaxies and a significant portion of baryons, are more representative (e.g. [Robotham et al.](#)

[2011](#)). Besides, the relatively low gravitational binding energy of galaxy groups makes them particularly valuable for studying the impact of baryonic feedback (e.g. [McCarthy et al. 2010](#); [Kettila et al. 2015](#)). They also contribute significantly to the cosmic shear signal, making it important to understand their properties in weak lensing studies (e.g. [Semboloni et al. 2011](#); [Debackere et al. 2020](#)).

However, robustly identifying galaxy groups is challenging. One approach is to use spectroscopic surveys with high spatial and redshift completeness. The Galaxy and Mass Assembly project (GAMA, [Driver et al. 2011](#)) represents one such effort. With a spectroscopic completeness of 95 percent for galaxies down to an r -band magnitude of 19.65, and sky coverage of approximately 250 square degrees, GAMA offers the highest redshift density available for such an extensive area to date ([Driver et al. 2022](#)). The galaxy group catalogue, as one of its key products, is an invaluable resource for studying group properties ([Robotham et al. 2011](#)).

The next challenge lies in measuring the dark matter properties of galaxy groups. This becomes evident even when inferring basic properties such as halo masses. Unlike massive galaxy clusters, where X-ray measurements of the intracluster medium provide good mass estimates (e.g. [Cavaliere & Fusco-Femiano 1976](#); [Evrard et al. 1996](#)), galaxy groups have much fainter X-ray signals, limiting the effectiveness of this technique (e.g. [Eckmiller et al. 2011](#); [Pop et al. 2022](#); [Bahar et al. 2022](#)). Additionally, baryonic processes such as cooling, star formation, and feedback cause deviations from hydrostatic equilibrium, which is a common assumption in X-ray mass measurements. These

deviations introduce biases in mass estimates that rely on this assumption (e.g. [Rasia et al. 2006](#); [Biffi et al. 2016](#); [Barnes et al. 2021](#); [Logan et al. 2022](#)).

Weak gravitational lensing provides an alternative approach to estimate halo masses. It measures the subtle yet coherent distortions in the shapes of background galaxies, caused by the gravitational field of a foreground lens (see [Bartelmann & Schneider 2001](#), for a review). These distortions directly trace the matter distribution of the foreground lens, enabling mass estimation without assumptions about its dynamical state (e.g. [Tyson et al. 1990](#); [Kaiser & Squires 1993](#); [von der Linden et al. 2014](#); [Hoekstra et al. 2015](#); [Robertson et al. 2024](#)).

However, the weak lensing signals induced by individual galaxy groups often have low signal-to-noise ratios (S/Ns), limiting the precision of individual mass estimates. Therefore, in practice, we select the lens sample based on certain observable properties, then measure the averaged signals to estimate the average masses. This approach boosts the measurement of the S/N and provides a statistical description of the scaling relation between halo mass and the corresponding observable properties (e.g. [Johnston et al. 2007](#); [Leauthaud et al. 2010](#); [Han et al. 2015](#); [Viola et al. 2015](#); [Rana et al. 2022](#)).

To interpret the averaged weak lensing signal, we need a statistical model to describe the galaxy and dark matter properties. The halo model, combined with halo occupation statistics, offers such a theoretical framework (e.g. [Seljak 2000](#); [Peacock & Smith 2000](#); [Cooray & Sheth 2002](#); [Berlind & Weinberg 2002](#); [Yang et al. 2003](#); [Vale & Ostriker 2004](#); [Cooray 2006](#); [van den Bosch et al. 2013](#)). It assumes that all dark matter exists within virialised halos and that galaxies are populated within these dark matter halos. After parameterising the desired scaling relation between galaxy observables and halo properties within the model, we can directly constrain it by fitting to the averaged weak lensing signals (e.g. [Guzik & Seljak 2002](#); [Mandelbaum et al. 2006](#); [van Uitert et al. 2011](#); [Cacciato et al. 2014](#)).

In practice, the halo model contains many theoretically motivated and empirically informed ingredients, each governed by a set of free parameters that may not be well constrained by the available data. Furthermore, the potential degeneracy among these parameters can complicate the interpretation of results derived from the halo model (e.g. [Viola et al. 2015](#)). To address these challenges, we can improve the halo model by refining certain components or providing informed priors for model parameters, using our knowledge of galaxy formation and dark matter properties (e.g. [Smith et al. 2003](#); [Mead et al. 2015](#); [Fortuna et al. 2021](#)).

This knowledge is often encoded in cosmological simulations, which numerically model various physical processes to track the non-linear evolution of matter in an expanding cosmological background, using initial conditions constrained by the well-measured cosmic microwave background. With decades of development and vastly improved computational power, modern cosmological simulations can solve gravitational N -body problems with high accuracy for volumes that are sufficient for interpreting observations from the next generation of galaxy surveys (see [Angulo & Hahn 2022](#), for a recent review). Nevertheless, there are still challenges in accounting for the more complicated baryonic effects in these simulations.

The two main strategies to address baryonic effects currently in active development are semi-analytical modelling and hydrodynamical simulations. The former still relies on gravity-only simulations but populates galaxies based on a semi-analytical model (SAM, e.g. [White & Rees 1978](#); [White & Frenk 1991](#); [Cole 1991](#)). This SAM consists of a set of simplified equations

to account for the key baryonic processes that affect the formation and evolution of galaxies. Because the ingredients in a SAM are theoretically simplified and often phenomenological, not all parameters can be rigidly determined from physical arguments. This results in a number of assumptions and free parameters that require calibration based on observational data.

The latter, hydrodynamical simulations, adopt a more sophisticated approach by self-consistently solving the co-evolution of dark matter and baryons from the outset (see [Vogelsberger et al. 2020](#); [Crain & van de Voort 2023](#), for some recent reviews). However, due to the vast range of physical processes involved and the limited resolution of simulations, some effective models are still necessary to capture subgrid baryonic processes that are not resolved by the numerical calculations. These subgrid models also contain free parameters that require calibration based on observations (see [Schaye et al. 2015](#), for a discussion).

We study this intriguing interplay between cosmological simulations and weak lensing analysis in this paper. First, we constrain the stellar-to-halo mass scaling relation of GAMA galaxy groups through a halo model-based analysis of weak lensing signals measured from the Kilo-Degree Survey (KiDS, [de Jong et al. 2013](#); [Kuijken et al. 2015](#)). The constrained scaling relation is then compared with predictions from the latest FLAMINGO cosmological hydrodynamical simulations ([Schaye et al. 2023](#); [Kugel et al. 2023](#)) and the L-GALAXIES SAM run on the IllustrisTNG gravity-only simulations ([Henriques et al. 2015](#); [Ayromlou et al. 2021](#)) to assess the reliability of the simulation predictions. Alternatively, one can directly compare the averaged weak lensing signal between observational data and simulations by including measurement effects into simulations (e.g. [Velliscig et al. 2017](#); [Jakobs et al. 2018](#); [Gouin et al. 2019](#)).

After validating the simulation predictions, we explore how insights from these recent simulations can be used to improve our current halo model. For this, we also use results from the IllustrisTNG hydrodynamical simulations ([Nelson et al. 2019](#)), which cover the low halo mass range not constrained by our data but necessary for our halo modelling. We focus on improving one of the key halo model ingredients: the scatter in the group stellar mass distribution as a function of halo mass. Besides, we assess the impact of different statistical models for the miscentring of identified central galaxies. These investigations guide the future development of more robust, simulation-informed halo models, which will be crucial for interpreting the significantly improved weak lensing measurements from upcoming surveys like the ESA *Euclid* space mission ([Euclid Collaboration et al. 2025](#)) and the *Rubin* Observatory Legacy Survey of Space and Time (LSST, [Ivezić et al. 2019](#)).

The rest of the paper is structured as follows. Sections 2 and 3 describe the data and simulations, respectively. Section 4 details the measurement of weak lensing signals and the covariance matrix. Our baseline model is outlined in Sect. 5, and its results are compared to previous studies and simulation predictions in Sect. 6. We introduce a simulation-informed scatter model and update the scaling relation constraints in Sect. 7. Section 8 tests different miscentring models, and we conclude in Sect. 9. Appendix A investigates the higher-order shear biases.

The overdensity threshold of a virialised dark matter halo is defined such that the average density within the virial radius is 200 times the mean matter density of the Universe at the redshift of the halo. When reporting values dependent on Hubble's constant, we use $h_{70} = H_0/70 \text{ km s}^{-1}\text{Mpc}^{-1}$ to facilitate comparison of results derived from observations and different simulations.

All measurements are presented in comoving units, and the logarithm base is 10.

2. Data

Our lens sample is from the GAMA survey¹, while the source sample is from the KiDS survey². In this section, we provide an overview of the catalogues used in our study. For technical details, we direct interested readers to the relevant data release papers.

2.1. Lenses: GAMA groups

GAMA is a high-density, high-completeness spectroscopic survey conducted using the AAOmega instrument on the Anglo-Australian Telescope (Driver et al. 2011). We used data from three equatorial fields of the GAMA II phase (G09, G12, G15), each covering a sky area of 60 square degrees (Liske et al. 2015). The GAMA data in these fields have a spectroscopic completeness of 98 percent for galaxies within the observed magnitude limit of $r < 19.58$ (Driver et al. 2022). In particular, we used three key GAMA products: the G³C group catalogue (version 10, Robotham et al. 2011), the StellarMassesLambdar catalogue (version 24, Taylor et al. 2011), and the random catalogue (version 2, Farrow et al. 2015).

The G³C group catalogue (version 10) includes 26,194 groups identified using an implementation of the friends-of-friends (FoF) algorithm based on galaxy separations. It separately treats the projected and radial separations to account for line-of-sight effects caused by peculiar velocities within groups. The two key linking parameters, the linking length (which defines the overdensity) and the radial expansion factor (which accounts for the peculiar motions of galaxies within groups), scale with the observed density contrast and depend on the galaxy positions and the magnitude limit of the survey. These parameters are optimised for quality of group finding through tests on GAMA lightcone mock data derived from semi-analytical simulations (see Robotham et al. 2011, for further details). To minimise the impact of interlopers, we used groups with at least five identified members, resulting in a final sample of 2752 groups.

We used the sky position of the brightest group (or cluster) galaxy (BCG) as the group centre for our lensing measurements. Another commonly used method for selecting the central galaxy is by iteratively removing group members that are furthest from the light centre of the group. However, Robotham et al. (2011) found that for groups with at least five members, this iterative procedure converges on the BCG 95% of the time. Given the current measurement uncertainties, the subtle difference between these two methods is even more negligible (see Appendix A of Viola et al. 2015). Therefore, we opted not to repeat measurements with alternative centres and instead modelled the miscentring statistically within our analysis (see Sect. 5).

The stellar masses of galaxies were obtained from the StellarMassesLambdar catalogue (version 24), which uses stellar population synthesis models from Bruzual & Charlot (2003), assuming a Chabrier (2003) initial mass function. The model fits were applied over a fixed rest-frame wavelength range (300–11000 Å) using matched aperture photometry from the Lambda Adaptive Multi-Band Deblending Algorithm in R (LAMBDA, Wright et al. 2016). The LAMBDA code is

designed to ensure consistent photometry and uncertainty estimation across a wide range of photometric imaging for calculating spectral energy distributions. It uses predefined elliptical apertures, initially estimated using SExtractor (Bertin & Arnouts 1996) runs on SDSS r -band and VIKING Z -band imaging, followed by visual inspection and manual adjustments for objects flagged with poor aperture determinations. We used the \log_{MSTAR} value, which represents the total mass of luminous material and remnants, excluding mass recycled into the interstellar medium. We opted not to adjust the aperture-based stellar mass to total stellar mass because not all galaxies in the GAMA survey have accurate total flux estimates. This also aligns with our use of aperture stellar mass from simulations (see Sect. 3).

To calculate the total stellar masses of galaxy groups, we applied the correction method from Robotham et al. (2011) used for estimating the r -band total luminosity. This method, based on the global GAMA galaxy luminosity function, accounts for missing flux due to the flux limit of the survey and includes a global optimisation factor to address biases from the luminosity function-based corrections, environmental effects, and extrapolation. The final corrections are redshift-dependent, ranging from a few percent at low redshift to a few factors at high redshift. We applied the same corrections to the observed group stellar masses, assuming a comparable stellar mass-to-light ratio between observed and intrinsic group properties. This assumption is valid given the depth of the GAMA survey, which recovers nearly all of the luminosity (or stellar mass) density (Loveday et al. 2012).

We used the GAMA random catalogue (version 2) to assess additive shear biases in measured weak lensing signals, as is detailed in Sect. 4.1. This catalogue consists of randomly distributed points, reflecting the same selection function as the main spectroscopic survey. For our analysis, we randomly selected one million points from this catalogue for each of the GAMA fields under consideration. We measured weak lensing signals around these random points to quantify and correct potential additive shear biases in our measurements, following Dvornik et al. (2017).

2.2. Sources: KiDS galaxies

KiDS is a wide-field imaging survey, with weak gravitational lensing analysis as its primary scientific objective (de Jong et al. 2013; Kuijken et al. 2015). The complete survey covers 1350 deg² of the sky, with optical images in the $ugri$ bands taken from the ESO VLT Survey Telescope. Among these, the r -band images, offering the highest imaging quality, are used for measuring galaxy shapes. In collaboration with the VISTA Kilo-degree Infrared Galaxy survey (VIKING, Edge et al. 2013) conducted with the nearby ESO VISTA telescope, the KiDS shear catalogue also contains photometry from five $ZYJHK_s$ near-infrared bands. This additional near-infrared dataset significantly enhances the accuracy of photometric redshift estimates.

For our analysis, we used the latest KiDS-1000 shear catalogue (v2) from Li et al. (2023a). This catalogue is based on the fourth data release of KiDS (Kuijken et al. 2019; Giblin et al. 2021) with enhanced redshift calibration from van den Busch et al. (2022) and updated shear measurement and calibration from Li et al. (2023b). It fully covers the three equatorial fields of GAMA, as illustrated in Fig. 1. Thanks to this complete coverage, we are now able to measure weak lensing signals around all 2752 selected GAMA groups, approximately doubling the number used in previous similar analyses by Viola et al. (2015) and Rana et al. (2022).

¹ <https://www.gama-survey.org/dr4>

² <https://kids.strw.leidenuniv.nl/DR4>

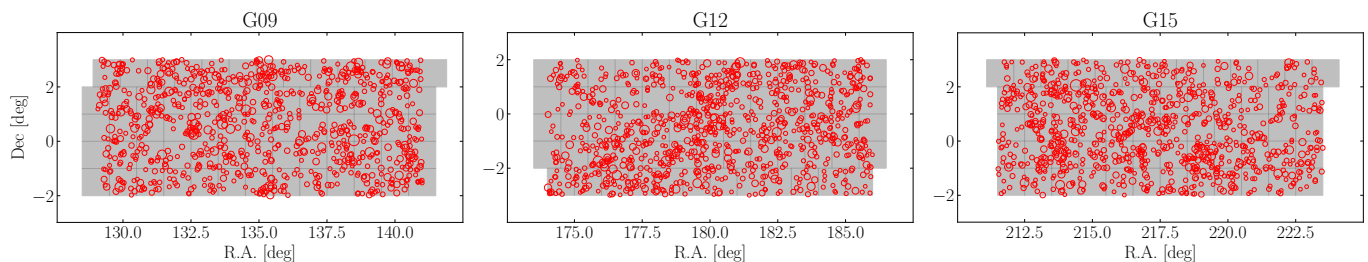


Fig. 1. Sky coverage of the KiDS-1000-v2 shear catalogue for the three equatorial GAMA fields (G09, G12, G15). The grey boxes represent KiDS tile images, each covering 1 deg^2 . The red circles indicate the selected GAMA groups, each consisting of at least five members. The size of these circles corresponds to the logarithm of the group richness.

3. Simulations

In this study, we used two state-of-the-art cosmological (magneto)hydrodynamical simulations, namely FLAMINGO³ and IllustrisTNG⁴, which offer complementary combinations of volume and resolution. Additionally, we used the L-GALAXIES SAM⁵ run on the IllustrisTNG gravity-only simulations as a representation of the other simulation technique, so we have a broad coverage of simulation uncertainties that we can account for in our halo model development. This section provides an overview of these simulations and describes how we construct mock group catalogues with the desired properties from them. Units in this section are relative to the cosmology adopted by each specific simulation, except when comparing properties across different simulations or with observational data. In such cases, we scaled the quantities to account for differences in the Hubble constant, while ignoring the impact of differences in other cosmological parameters.

3.1. FLAMINGO simulations

FLAMINGO is a suite of hydrodynamical cosmological simulations generated using the smoothed particle hydrodynamics-based code SWIFT (Schaller et al. 2024). The FLAMINGO fiducial cosmology is based on the Dark Energy Survey year three results ($3 \times 2\text{pt}$ plus external constraints, Abbott et al. 2022; $\Omega_\Lambda=0.694$, $\Omega_m=0.306$, $\Omega_b=0.0486$, $\sigma_8=0.807$, $n_s=0.967$, and $h=0.681$) and includes a single massive neutrino species with a mass of 0.06 eV and two massless species. Its galaxy formation model builds upon those developed for the OWLS (Schaye et al. 2010), cosmo-OWLS (Le Brun et al. 2014), EAGLE (Schaye et al. 2015), and BAHAMAS (McCarthy et al. 2017) projects, including radiative cooling and heating, star formation and evolution, stellar energy feedback, growth of supermassive black holes, and AGN feedback. A notable advancement of the FLAMINGO galaxy formation model is the calibration of its free parameters to the observed present-day galaxy stellar mass function (GSMF) and low-redshift cluster gas fraction using Gaussian process emulators, which explicitly accounts for observational uncertainties and biases (see Kugel et al. 2023 and Schaye et al. 2023, for a detailed description of the model).

The simulation snapshots are saved at various redshifts from 10 to 0. In each snapshot, halos and substructures are identified using the VELOCIRAPTOR subhalo finder (Elahi et al. 2019). In short, it first identifies halos using a standard 3D FoF algo-

rith with a linking length of 0.2 (Davis et al. 1985). Then, within each FoF halo, it iteratively searches for subhalos that are dynamically distinct from the mean background halo using a 6D FoF algorithm in position-velocity phase space. Finally, (sub)halo properties are computed for a range of apertures using SOAP (Spherical Overdensity Aperture Processor)⁶. These aperture property estimates are essential for building mock galaxy group catalogues, which we compare to observations.

In this work, we used two FLAMINGO simulations with volumes of $(1 \text{ Gpc})^3$, and initial mean baryonic particle masses of $1.34 \times 10^8 M_\odot$ (L1_m8) and $1.07 \times 10^9 M_\odot$ (L1_m9). Both simulations employ the fiducial galaxy formation model and assume the fiducial cosmology. However, despite the resolution-dependent calibration of the FLAMINGO model, which uses different mass ranges for calibration, the resulting predictions still exhibit some variation across different resolutions (Kugel et al. 2023). By using simulations from both resolutions, we can account for the impact of these resolution-related differences in our halo model development.

3.2. IllustrisTNG simulations

IllustrisTNG is a suite of magnetohydrodynamical cosmological simulations generated with the moving-mesh code AREPO (Springel 2010). It assumes a cosmology consistent with the Planck Collaboration et al. (2016) results ($\Omega_\Lambda=0.6911$, $\Omega_m=0.3089$, $\Omega_b=0.0486$, $\sigma_8=0.8159$, $n_s=0.9667$, and $h=0.6774$). Its galaxy formation model builds upon the framework used in the Illustris project (Vogelsberger et al. 2014; Genel et al. 2014), including all the main ingredients mentioned in Sect. 3.1 but with different implementations (see Vogelsberger et al. 2013 for details). The free parameters in the IllustrisTNG model are manually tuned to alleviate the most striking observational tensions identified in the Illustris simulations (Nelson et al. 2015). Key advancements in IllustrisTNG include the inclusion of a seed magnetic field, a revised galactic wind implementation, and a kinetic AGN feedback model for low-accretion states (see Pillepich et al. 2018 for details).

The simulation snapshots are saved at redshifts from 20 to 0. For each snapshot, halos are first identified using a standard 3D FoF group finder with a linking length of 0.2. Then, within each FoF halo, the gravitationally bound substructures are located and characterised hierarchically using the SUBFIND algorithm (Springel et al. 2001).

In this work, we used their flagship run with a box side length of 110.7 Mpc and an effective mass resolution of $1.39 \times 10^6 M_\odot$.

⁶ <https://github.com/SWIFTSIM/SOAP.git>

³ <https://flamingo.strw.leidenuniv.nl/>

⁴ <https://www.tng-project.org/>

⁵ <https://lgalaxiespublicrelease.github.io/>

for baryons and $7.46 \times 10^6 M_\odot$ for dark matter (TNG100-1, Nelson et al. 2019). The TNG100-1 simulation provides a different balance between volume and resolution, covering the low halo mass range missed by the FLAMINGO simulations. Most importantly, it includes aperture stellar mass estimates produced by Engler et al. (2021), which are essential for our study (see Sect. 3.4).

3.3. L-Galaxies semi-analytical model

L-GALAXIES is a SAM of galaxy formation designed to run on subhalo merger trees produced by N -body simulations. The model is based on seminal works of White & Frenk (1991), Kauffmann et al. (1993, 1999) and Springel et al. (2001), with its first relatively mature implementation in the Millennium Simulations by Springel et al. (2005). Since then, the model has undergone a series of updates driven by discrepancies identified between model predictions and observations, resulting in a series of public releases of mock catalogues.

In our study, we used the catalogues produced by Ayromlou et al. (2021), who ran the Henriques et al. (2015) version of the L-GALAXIES model on the companion gravity-only simulations of IllustrisTNG. The Henriques et al. (2015) version of L-GALAXIES builds on Guo et al. (2011), aiming for a better representation of the observed evolution of low-mass galaxies. The model contains a set of coupled differential equations to follow the evolution of baryonic components in each hierarchical merger tree. A complete description of the model treatment can be found in the supplementary material associated with the arXiv version of Henriques et al. (2015). The free parameters in the model are calibrated to match the observed GSMF at $z = 0, 1, 2$, and 3, as well as the fraction of red galaxies as a function of stellar mass at $z = 0, 0.4, 1, 2$, and 3.

We used two L-GALAXIES runs with box sizes of 110.7 Mpc (LGal100-1) and 302.6 Mpc (LGal300-1). The former shares the same mass coverage as TNG100-1, enabling a direct comparison between hydrodynamical simulations and SAM results, while the latter overlaps in mass range with FLAMINGO simulations and is better aligned with the mass range of our weak lensing measurements.

3.4. Construction of mock group catalogues

To perform a consistent comparison between simulations and observations, it is crucial to translate simulated properties into mock observables that are comparable to those from observations. In our study, these observables mainly concern the virial mass of halos and the total stellar mass of galaxy groups. The primary factor influencing halo properties is the particle resolution of the simulations. Therefore, we implemented a halo mass cut in the simulations based on their mean dark matter particle masses. The detailed mass cut selections are summarised in Table 1. These thresholds correspond to a minimum of ~ 5000 particles for IllustrisTNG and ~ 1000 particles for FLAMINGO, ensuring that only well-resolved halos are included in our mock catalogues.

To estimate the total stellar masses of galaxy groups, we considered both observational effects and the particle resolution of simulations. As is described in Sect. 2.1, the total stellar masses of GAMA galaxy groups are estimated using the stellar mass measurements of individual galaxies and the global galaxy luminosity function. Ideally, we would mimic these observational estimates by deriving stellar mass from measured aperture flux

Table 1. Sample selection for construction of mock group catalogues.

Simulations	Halo mass cut	Stellar mass cut
FLAMINGO (L1_m9)	$\geq 10^{12.7} M_\odot$	$\geq 10^{10} M_\odot$
FLAMINGO (L1_m8)	$\geq 10^{11.8} M_\odot$	$\geq 10^{8.5} M_\odot$
LGal300-1	$\geq 10^{11.5} M_\odot$	–
LGal100-1	$\geq 10^{10.6} M_\odot$	–
TNG100-1	$\geq 10^{10.6} M_\odot$	–

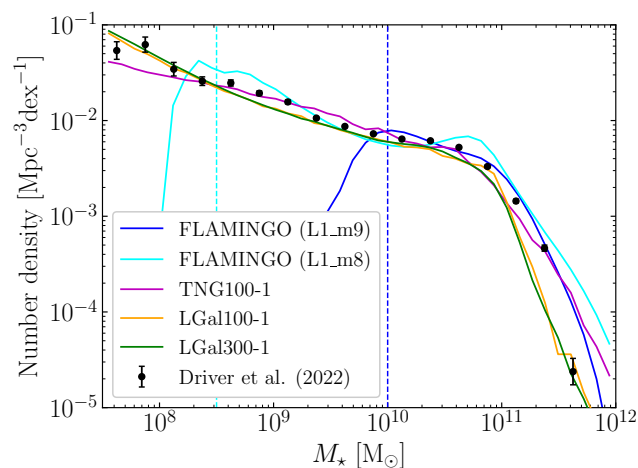


Fig. 2. Galaxy stellar mass function at redshift $z=0$ from simulations and GAMA observations (Driver et al. 2022). For comparison, all properties have been converted to a h_{70} cosmology, with masses from simulations scaled as h_{70}^{-1} and masses from observational data scaled as h_{70}^{-2} . The vertical dashed lines indicate the stellar mass cuts applied to the two FLAMINGO simulations, based on their respective resolutions. The overall agreement between simulations and observations supports our approach for estimating the total stellar masses of mock galaxy groups, while the resolution limits of the FLAMINGO simulations highlight the need for careful treatment, as is detailed in Sect. 3.4.

and galaxy profile fitting, but this is challenging, particularly given the limited resolution of simulations used in this study (de Graaff et al. 2022; Kugel et al. 2023).

Therefore, we adopted a simplified approach while maintaining the principle of using individual galaxy stellar masses to estimate the total stellar masses of galaxy groups. In the case of SAM, the galaxy stellar mass is relatively well defined as the sum of disc and bulge stellar masses. For hydrodynamical simulations, there are several different methods for estimating galaxy stellar masses. Following previous studies comparing simulation and observational results (Schaye et al. 2015; de Graaff et al. 2022), we opted to use the 3D physical aperture stellar mass estimation as the galaxy stellar mass, which is defined as the sum of gravitationally bound stellar particles within a given radius.

For TNG100-1, we used the 30 kpc aperture estimates provided in their supplementary data release (Engler et al. 2021). For FLAMINGO, we used the 50 kpc aperture estimates, following their model calibration choice. As shown by Kugel et al. (2023), the impact of this difference in aperture size is negligible for galaxies with stellar masses below $10^{11} M_\odot$. We confirmed that our results remain unchanged when using the 30 kpc aperture estimates for all simulations.

Figure 2 compares the present-day GSMF between the simulations and the latest GAMA observations from Driver et al. (2022). Overall, the simulations show good agreement with the observations, which is expected given that the present-day GSMF is one of the key observational constraints used to tune the free parameters in the simulations (Henriques et al. 2015; Pillepich et al. 2018; Kugel et al. 2023). However, the FLAMINGO simulations exhibit a drop at the low-mass end, attributed to their resolution limit. To account for this low-mass drop, we took extra care when calculating the total stellar mass of FLAMINGO galaxy groups. Specifically, we first implemented a stellar mass cut in the FLAMINGO galaxies, setting a minimum of $10^{8.5} M_{\odot}$ and $10^{10} M_{\odot}$ for L1_m8 and L1_m9, respectively, based on the visual inspection of the dropping points (vertical dashed lines in Fig. 2). These cuts correspond to at least three bounded particles in L1_m8 and ten bounded particles in L1_m9. We tested a more conservative stellar mass cut, with a minimum of $10^{9.3} M_{\odot}$ for the L1_m8 simulations and found consistent results. After this selection, we estimated the total stellar mass of each galaxy group using

$$M_{\star}^{\text{grp}} = \left(\sum_i M_{\star, i} \right) \frac{\int_0^{\infty} dM_{\star} \phi(M_{\star}) M_{\star}}{\int_{M_{\star, \text{min}}}^{\infty} dM_{\star} \phi(M_{\star}) M_{\star}}. \quad (1)$$

In practice, we approximated zero and infinity by using $1 M_{\odot}$ and $10^{13} M_{\odot}$, respectively, as the contribution to the stellar mass density from galaxies outside this mass range is negligible. For the GSMF, $\phi(M_{\star})$, we used the double Schechter function fit from the latest GAMA observations (Driver et al. 2022, Table 7).

Since TNG100-1 resolves galaxies down to $10^7 M_{\odot}$ (corresponding to ~ 10 stellar particles), these additional steps of stellar mass cut and boost factor are unnecessary. This also holds true for the L-GALAXIES SAM, where completeness is maintained down to $10^6 M_{\odot}$. Therefore, for these simulations, we simply sum the stellar masses of all member galaxies to obtain the total stellar mass of galaxy groups.

For all simulations, we constructed mock group catalogues only from their present-day snapshot ($z=0$), as the GAMA galaxy groups are local, with a mean redshift of ~ 0.2 . We examined the evolution of the desired galaxy group and halo properties from redshift 0 to 0.2 in all simulations and found negligible differences.

4. The weak lensing signals

The weak lensing effect introduces coherent tangential distortions in the observed shapes of background galaxies. These distortions, known as the tangential shear, γ_t , correlate with the projected mass density contrast of the foreground lens⁷ (e.g. Bartelmann & Schneider 2001):

$$\Delta\Sigma(R) \equiv \bar{\Sigma}(\leq R) - \Sigma(R) = \Sigma_{\text{cr}}\gamma_t(R), \quad (2)$$

where the mass density contrast, $\Delta\Sigma(R)$, is also commonly referred to as the excess surface density (ESD). The $\Sigma(R)$ represents the local surface mass density at a projected comoving separation, R , between the lens and source, while $\bar{\Sigma}(\leq R)$ denotes the mean surface density within this radius. The critical surface density, Σ_{cr} , serves as a measure of lensing efficiency and is defined

⁷ Throughout this work, we do not distinguish between the original shear γ and the reduced shear $g \equiv \gamma/(1 - \kappa)$, given that the convergence κ is much less than one in the weak lensing regime.

as

$$\Sigma_{\text{cr}} \equiv \frac{c^2}{4\pi G} \frac{r_s}{(1 + z_d) r_d r_{\text{ds}}}, \quad (3)$$

where G and c denote the gravitational constant and the speed of light, respectively. The r_d , r_s , and r_{ds} are the comoving distances to the lens, source and between these two, respectively, and z_d is the redshift of the lens. The factor $(1 + z_d)$ arises due to our use of comoving distances and co-ordinates (see Appendix C of Dvornik et al. 2018 for a detailed derivation).

Therefore, by assuming a certain density profile for an object, we can infer its mass by measuring the ESD signals around it. In this section, we detail how we estimate ESD for the selected GAMA galaxy groups using the KiDS shear measurements and the corresponding covariance matrix necessary for modelling.

4.1. ESD measurements

We estimated the tangential shear using the azimuthal average of the tangential projection, ϵ_t , of the *lensfit* measured ellipticities of the KiDS source galaxies. It is defined as

$$\begin{bmatrix} \epsilon_t \\ \epsilon_{\times} \end{bmatrix} \equiv \begin{bmatrix} -\cos(2\varphi) & -\sin(2\varphi) \\ \sin(2\varphi) & -\cos(2\varphi) \end{bmatrix} \cdot \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \quad (4)$$

where φ denotes the relative position angle of the source with respect to the lens. The azimuthal average of the cross projection, ϵ_{\times} , can serve as an indicator of potential systematic contamination, given that the lensing effect only introduces tangential shear to the leading order.

To account for both measurement uncertainties and lensing efficiency, a weight was assigned to each lens-source pair when calculating the azimuthal average. This weight is given by

$$w_{\text{ds}} \equiv w_s \tilde{\Sigma}_{\text{cr,d}}^{-2}, \quad (5)$$

where w_s is the *lensfit* weight, reflecting the individual galaxy shape measurement uncertainties, and $\tilde{\Sigma}_{\text{cr,d}}$ is the ‘effective critical surface density’, which down-weights lens-source pairs that are close in redshift and thus carry lower lensing signals.

Following Dvornik et al. (2017), we calculated $\tilde{\Sigma}_{\text{cr,d}}$ for each lens by integrating the redshift distribution of the whole source sample behind the given lens. This averaging approach aligns with the KiDS-1000 redshift calibration (Hildebrandt et al. 2021). An alternative way to determine source distances is by using the individual posterior redshift distributions of each source galaxy, as in Viola et al. (2015). Dvornik et al. (2017) verified that these two approaches yield consistent signals within the error budget for the lensing signal of the GAMA sample. Following Eq. (3), the calculation is formulated as

$$\tilde{\Sigma}_{\text{cr,d}}^{-1} = \frac{4\pi G}{c^2} (1 + z_d) r_d \int_{z_d + \delta_z}^{\infty} dz_s \frac{r_{\text{ds}}}{r_s} n(z_s), \quad (6)$$

where the source redshift distribution, $n(z_s)$, was determined from the redshift calibration reference sample of Li et al. (2023a), which is based on the fiducial spectroscopic sample of van den Busch et al. (2022). A redshift difference threshold, $\delta_z=0.2$, is introduced to mitigate contamination from group members to the source sample. This redshift cut-off, $z_s > z_d + \delta_z$, is applied to the source galaxies involved in the calculation as well as to the reference spectroscopic sample.

The median velocity dispersion of the GAMA galaxy groups used in our study is $\sim 300 \text{ km s}^{-1}$, which is not massive enough

to enable the lensing measurement for individual groups. Therefore, we use a stacking process to enhance the S/N, following the methodology of previous KiDS analyses (e.g. Viola et al. 2015; Dvornik et al. 2017). It estimates the stacked ESD profile for an ensemble of galaxy groups as

$$\Delta\Sigma(R) = \left[\frac{\sum_{ds} w_{ds} \epsilon_t \tilde{\Sigma}_{cr,d}}{\sum_{ds} w_{ds}} \right] \frac{1}{1+K}, \quad (7)$$

where the correction

$$K = \frac{\sum_{ds} w_{ds} m_s}{\sum_{ds} w_{ds}}, \quad (8)$$

accounts for the multiplicative bias, m_s , of our *lensfit* shear measurements.

We estimated the multiplicative bias, m_s , using the latest SKiLLS image simulations developed by Li et al. (2023b). To capture the variation in m_s , we first divided SKiLLS galaxies into small bins based on redshift, resolution (the ratio of galaxy size to PSF size), and S/N. Specifically, we used uniform redshift bins with a width of 0.1, and for each redshift bin, 20×20 weighted quantile bins in resolution and S/N. We then calculated the average m_s within each bin and assigned these values to KiDS galaxies based on their corresponding properties. This approach accounts for the strong dependence of m_s on redshift, resolution, and S/N. However, it implicitly assumes that the multiplicative bias does not depend on the shear, which may not always be correct, depending on the magnitude of the shear signal and the specific shear measurement algorithm (e.g. Kitching & Deshpande 2022; Jansen et al. 2024). In Appendix A, we investigate the potential violation of this linear shear bias assumption using dedicated simulations. We find that the multiplicative bias from *lensfit* remains fairly linear within the typical shear amplitude range ($|\gamma| \lesssim 0.1$). However, in higher-shear regimes, higher-order terms up to the third order are needed to capture the non-linear behaviour if sub-percent level calibration accuracy is required. We also detect a small shear bias ($\sim 10^{-4}$) responding to the shear in the other component. The overall correction factor, K , is at the sub-percent level, with small variations across angular separation and lens observable bins.

Besides the multiplicative shear bias, we accounted for the additive shear bias by measuring lensing signals around one million random points selected from the GAMA random catalogue (version 2, Farrow et al. 2015). The additive bias is scale-dependent, with larger biases at larger lens-source separations, and depend on the GAMA patch (see Appendix A of Dvornik et al. 2017). Thus, we corrected the three GAMA patches (G9, G12, and G15), separately. The overall correction is small, with values at the sub-percent level, attributable to the relatively small scales our study focused on and the complete coverage of the GAMA fields by the current KiDS observations.

In this study, we focus on the halo mass relation with the total stellar mass of galaxy groups, which can be directly compared to the properties extracted from the mock group catalogues as described in Sect. 3.4. Besides, we verify our results, which feature updated shear measurements and new halo model ingredients, against previous similar studies that measured the halo mass relation with the r -band total luminosity of galaxy groups (Viola et al. 2015; Rana et al. 2022). Therefore, we performed two sets of ESD measurements by dividing the GAMA groups into six bins based on either their group stellar masses or their r -band total luminosity. We set lower and upper limits to exclude the tails of the distribution, as shown in Fig. 3, to mitigate selection effects at both ends and facilitate the modelling. Table 2 provides detailed information about the defined bins.

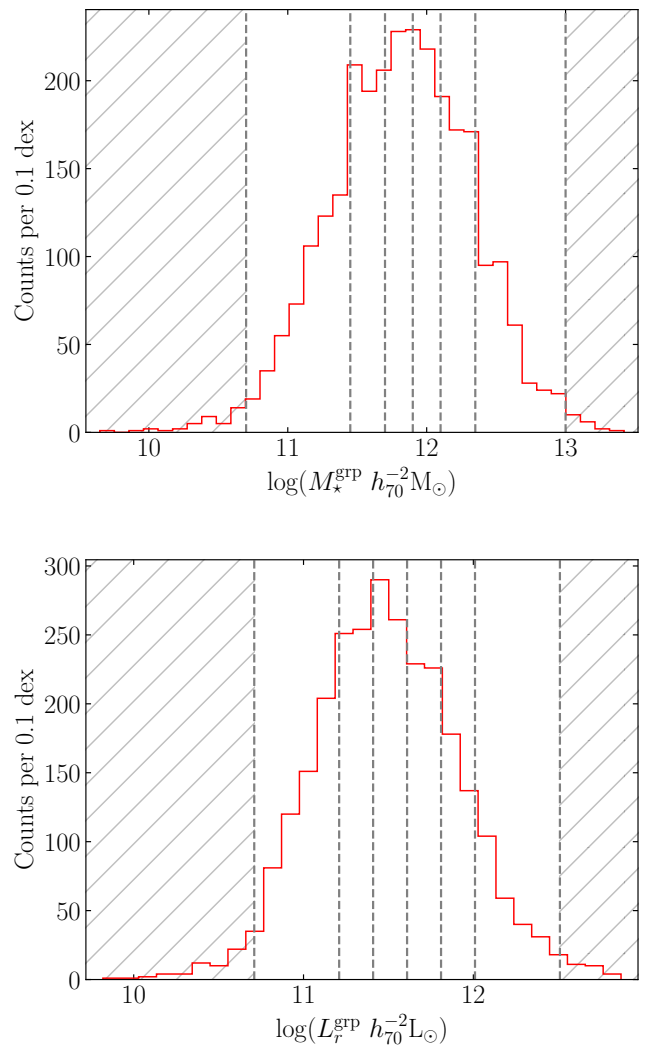


Fig. 3. Distributions of the group total stellar mass (upper panel) and the group r -band total luminosity (bottom panel). The vertical lines represent the boundaries of the bins for measuring stacked ESD profiles. The corresponding values are listed in Table 2. Objects falling within the hatched regions are excluded from our analysis.

We measured the ESD profiles in ten logarithmically spaced comoving radial bins, ranging from 0.04 to $2.86 h_{70}^{-1}$ Mpc. The centre of the measured ESD profile was defined as the location of the identified BCG for each group. The lower limit was chosen to balance the S/Ns and the impact of blending effects, while the upper limit is set to mitigate large-scale systematics (Viola et al. 2015). Figures 4 and 5 show the resulting ESD profiles for M_*^{grp} and L_r^{grp} binning, respectively. The signals have been corrected for additive and multiplicative shear biases as previously described. The overall S/N, accounting for the full correlation among data points (see Sect. 4.2), is 27.7 for M_*^{grp} binning and 27.6 for L_r^{grp} binning. The S/N for each bin is shown in the plots.

4.2. Covariance matrix estimation

The source galaxies can be used multiple times to estimate $\Delta\Sigma(R)$ across different radial and observable bins, leading to correlations between the stacked ESD measurements. To account for these correlations in our modelling, we adopted the covari-

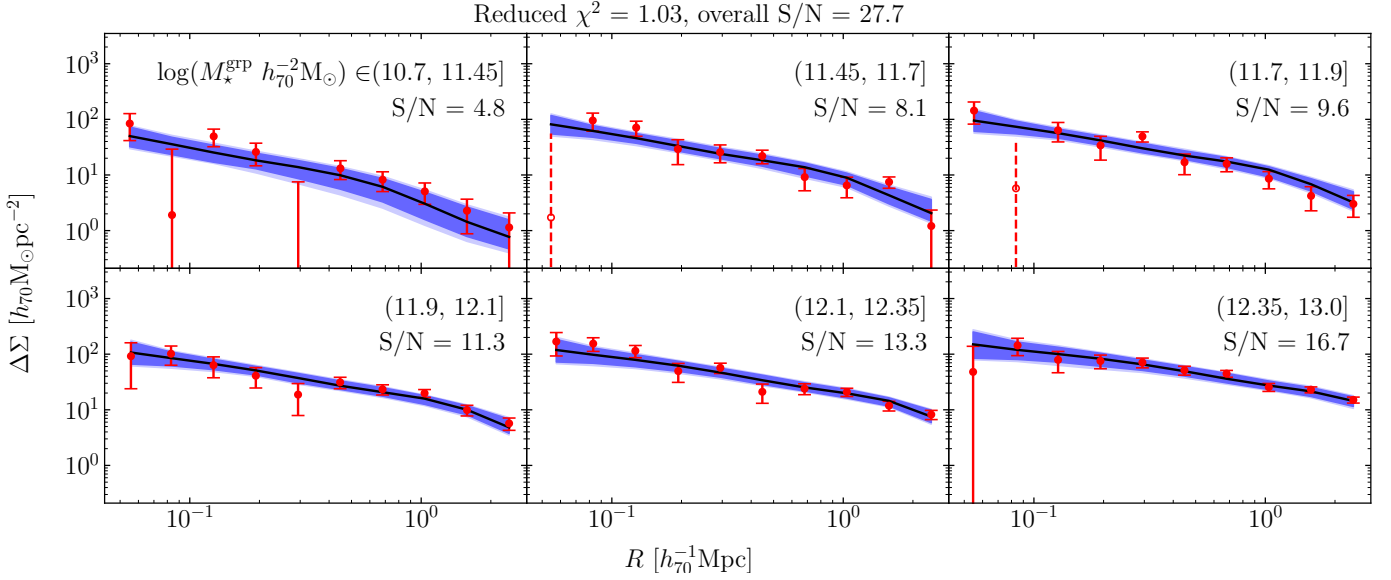


Fig. 4. Stacked ESD profiles in the six bins of group total stellar mass (M_{\star}^{grp}). The error bars correspond to the square root of the diagonal elements of the covariance matrix. We use open circles with dashed bars for negative values of the ESD. The black lines show the best-fit results from our baseline model (Sect. 5), with the shaded dark and light blue regions indicating the 68% and 95% credible intervals, respectively. The S/N for each M_{\star}^{grp} bin only accounts for correlations within that bin across different radial bins, while the overall S/N also accounts for correlations between different M_{\star}^{grp} bins. The reduced χ^2 value of 1.03 (considering 7 independent fitting parameters and 60 data points) for the best-fit results suggests an overall good fit to the data.

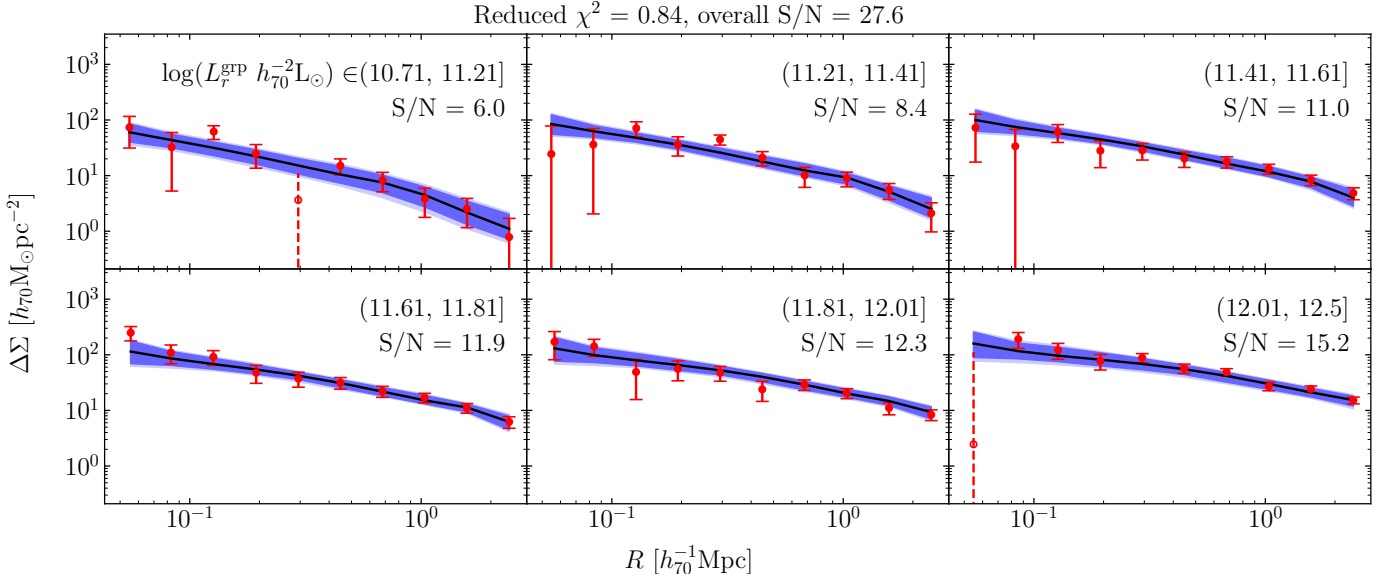


Fig. 5. Same as Fig. 4, but for the six bins of group r -band total luminosity (L_r^{grp}).

ance matrix estimation method developed by Viola et al. (2015). It analytically calculates the covariance directly from the data, accounting for correlations introduced by repeated entries of source galaxies and contributions from shape noise, while ignoring cosmic variance (see Section 3.4 of Viola et al. 2015 for details). This method has been used in other KiDS+GAMA analyses and shown to be sufficiently accurate for measurements within our adopted scale range of $R \leq 2.86 h_{70}^{-1}$ Mpc (e.g. Sifón et al. 2015; Brouwer et al. 2016).

5. Halo model and occupation statistics

From a statistical perspective, the ESD profile, $\Delta\Sigma(R)$, of an ensemble of lenses is related to the galaxy-matter power spectrum,

$P_{\text{gm}}(k)$, through (e.g. Murata et al. 2018)

$$\Delta\Sigma(R) = \frac{\bar{\rho}_m}{2\pi} \int_0^\infty dk k P_{\text{gm}}(k) J_2(kR), \quad (9)$$

where k represents the comoving wavenumber, $\bar{\rho}_m$ is the current mean matter density of the Universe, and $J_2(kR)$ is the second-order Bessel function of the first kind. The specific form of $P_{\text{gm}}(k)$ depends on the redshift of the lens and the types of galaxies contributing to the lens sample, such as central or satellite galaxies. Therefore, we can interpret the measured $\Delta\Sigma(R)$ signal if we have a model to describe $P_{\text{gm}}(k)$, tailored to the characteristics of the lens sample. The halo model, complemented by the halo occupation statistics, offers such a theoretical framework (e.g. Seljak 2000; Cooray & Sheth 2002; Peacock & Smith

Table 2. Summary of the binning boundaries, number of groups, mean redshift of the groups, and mean stellar mass of the BCGs for each bin used in the stacked ESD measurements.

Observable	Range	N_{groups}	z_{mean}	$\log(\overline{M}_{\star}^{\text{BCG}})$
$\log(M_{\star}^{\text{grp}})$	(10.70, 11.45]	589	0.12	10.83
	(11.45, 11.70]	470	0.16	11.07
	(11.70, 11.90]	429	0.20	11.19
	(11.90, 12.10]	416	0.24	11.28
	(12.10, 12.35]	419	0.28	11.36
	(12.35, 13.00]	368	0.32	11.49
$\log(L_r^{\text{grp}})$	(10.71, 11.21]	628	0.12	10.90
	(11.21, 11.41]	477	0.17	11.10
	(11.41, 11.61]	528	0.21	11.22
	(11.61, 11.81]	432	0.26	11.31
	(11.81, 12.01]	312	0.29	11.39
	(12.01, 12.50]	267	0.32	11.49

Notes. The unit of stellar mass is $h_{70}^{-2}M_{\odot}$, and the unit of luminosity is $h_{70}^{-2}L_{\odot}$.

2000; Berlind & Weinberg 2002; Yang et al. 2003; van den Bosch et al. 2013).

In this section, we detail our approach to interpreting the stacked ESD measurements using the halo model framework. First, we provide a concise overview of the halo model formalism, primarily following the notation of van den Bosch et al. (2013) and van Uitert et al. (2016). We then outline our baseline model ingredients, drawing primarily on previous KiDS studies (e.g. Viola et al. 2015; van Uitert et al. 2016; Dvornik et al. 2017), but incorporating some improvements motivated by recent progress.

5.1. Halo model with galaxy population statistics

The halo model framework is built on the assumption that all dark matter resides within virialised halos, whose sizes and masses are determined by a chosen overdensity threshold. By adopting models for the internal density profiles of these halos, we can use them to describe the matter-matter power spectrum of the Universe. Furthermore, by incorporating statistical models of how galaxies populate these dark matter halos, often referred to as halo occupation statistics or the halo occupation distribution (HOD), the halo model framework can also be extended to describe the galaxy-matter power spectrum (e.g. van den Bosch et al. 2013).

Following the notation of van den Bosch et al. (2013) and van Uitert et al. (2016), we formulate the galaxy-matter power spectrum as

$$P_{\text{gm}}(k) = P_{\text{gm}}^{\text{1h}}(k) + P_{\text{gm}}^{\text{2h}}(k), \quad (10)$$

where the one-halo term, describing correlations within a single halo, is defined as

$$P_{\text{gm}}^{\text{1h}}(k) \equiv \int dM_{\text{h}} \mathcal{H}_{\text{m}}(k, M_{\text{h}}) \mathcal{H}_{\text{g}}(k, M_{\text{h}}) n_{\text{h}}(M_{\text{h}}), \quad (11)$$

and the two-halo term, representing correlations between different halos, is defined as

$$P_{\text{gm}}^{\text{2h}}(k) \equiv P_{\text{m}}^{\text{lin}}(k) \int dM_{\text{h},1} \mathcal{H}_{\text{m}}(k, M_{\text{h},1}) n_{\text{h}}(M_{\text{h},1}) b_{\text{h}}(M_{\text{h},1}) \int dM_{\text{h},2} \mathcal{H}_{\text{g}}(k, M_{\text{h},2}) n_{\text{h}}(M_{\text{h},2}) b_{\text{h}}(M_{\text{h},2}). \quad (12)$$

Here, $P_{\text{m}}^{\text{lin}}(k)$ is the linear matter power spectrum, $n_{\text{h}}(M_{\text{h}})$ is the halo mass function, and $b_{\text{h}}(M_{\text{h}})$ is the associated large-scale halo bias, both with respect to the halo mass, M_{h} . These properties are redshift-dependent, and in our analysis, we used the mean redshift of the lens samples for each stacked bin, as is shown in Table 2. The subscript ‘g’ denotes galaxies, and ‘m’ denotes matter, corresponding to different forms of $\mathcal{H}_{\text{x}}(k, M_{\text{h}})$:

$$\mathcal{H}_{\text{m}}(k, M_{\text{h}}) \equiv \frac{M_{\text{h}}}{\bar{\rho}_{\text{m}}} \tilde{u}_{\text{m}}(k|M_{\text{h}}), \quad (13)$$

or

$$\mathcal{H}_{\text{g}}(k, M_{\text{h}}) \equiv \frac{\langle n_{\text{g}}|M_{\text{h}} \rangle}{\bar{n}_{\text{g}}} \tilde{u}_{\text{g}}(k|M_{\text{h}}). \quad (14)$$

The terms $\tilde{u}_{\text{m}}(k|M_{\text{h}})$ and $\tilde{u}_{\text{g}}(k|M_{\text{h}})$ describe the normalised density profiles of dark matter halos and galaxy distributions within a halo, respectively, in Fourier space. The term $\langle n_{\text{g}}|M_{\text{h}} \rangle$ is the average number of galaxies in a halo of mass M_{h} , and \bar{n}_{g} is the average galaxy number density, given by

$$\bar{n}_{\text{g}} = \int dM_{\text{h}} \langle n_{\text{g}}|M_{\text{h}} \rangle n_{\text{h}}(M_{\text{h}}). \quad (15)$$

5.2. Baseline model ingredients

We define the overdensity threshold of a virialised dark matter halo such that the average density within the virial radius is 200 times the mean matter density of the Universe. Consequently, the mass of a specific halo is formulated as

$$M_{\text{h}} = \frac{4\pi}{3} 200 \bar{\rho}_{\text{m}} r_{200\text{m}}^3. \quad (16)$$

For the internal density profile of these halos, we adopted the Navarro-Frenk-White (NFW, Navarro et al. 1997) profile truncated at the virial radius, following the Fourier transform form from Takada & Jain (2003). Alternatively, the profile can also be smoothly truncated in the manner proposed by Baltz et al. (2009), which has been shown to perform better in the transition region between one-halo and two-halo terms (Oguri & Hamana 2011). However, since our measurements are dominated by the one-halo term, we are not sensitive to the subtle differences between these truncation strategies. The mass-concentration relation of the NFW profile is based on Duffy et al. (2008):

$$c_{\text{m}} = f_{\text{c}} \times 10.14 \left(\frac{M_{\text{h}}}{2.86 \times 10^{12} h_{70}^{-1} M_{\odot}} \right)^{-0.081} (1 + z_{\text{d}})^{-1.01}, \quad (17)$$

where f_{c} is a free scaling parameter introduced to account for potential deviations of the weak lensing measurements from the original simulation-based fitting results. We do not vary the redshift or mass dependence in this equation, as more complex dependencies are predominantly found at redshifts greater than one (e.g. Muñoz-Cuartas et al. 2011; Wang et al. 2024), which exceed the highest lens redshift in our study.

To account for the mass contribution from central galaxies residing in the innermost region of the dark matter halo, we incorporated a point mass into the NFW density profile. This mass was set to be linearly scaled with the mean stellar mass of the BCGs for each stacked bin (see Table 2):

$$M_p \equiv A_p \bar{M}_\star^{\text{BCG}}, \quad (18)$$

where the scaling factor, A_p , is one of the free parameters we vary during the model fitting. Considering the scales of our ESD measurements, our analysis is insensitive to the detailed stellar mass distributions within the innermost part of the dark matter halo.

For the halo mass function and halo bias, we used the calibrated fitting functions from Tinker et al. (2010), which are derived from a series of cosmological N -body simulations within the Λ CDM framework. Other halo mass function estimates based on different cosmological simulations, including both N -body and hydrodynamical simulations, exist in the literature, and their predictions can vary significantly, often exceeding Poisson errors (Schaye et al. 2023). However, these discrepancies are likely dominated by the different halo definitions used by various halo finders (Euclid Collaboration et al. 2023). Therefore, to ensure consistency in halo definitions with our halo model, we chose the Tinker et al. (2010) halo mass function, as it employs the same mass definitions. Our analysis is also insensitive to the exact form of the halo bias function given the current measurement uncertainties. This is particularly true as we fit the ESD profiles only up to $2.86 h_{70}^{-1}$ Mpc, and the halo bias only affects our calculations through the two-halo term (Eq. 12).

When calculating the halo mass function, we adopted the Planck Collaboration et al. (2020) cosmological parameters ($\Omega_\Lambda=0.6842$, $\Omega_m=0.3158$, $\Omega_b=0.04939$, $\sigma_8=0.8120$, $n_s=0.96605$, and $h=0.6732$). To test the robustness of our results against uncertainties in these cosmological parameters, we also used the latest KiDS cosmic shear results (Li et al. 2023a), which feature a lower $\sigma_8(\Omega_m/0.3)^{0.5}$ value. We find consistent outcomes, confirming that our analysis is insensitive to the current level of cosmological uncertainties. This test also addresses some of the concerns about the uncertainty in the halo mass function used in our model, as the variation in cosmological parameters tested here is more extreme than the current differences in halo mass function estimates in the literature (Bocquet et al. 2016).

If the selected central galaxy (in our case, the BCG) resides exactly at the centre of its host halo, the $\tilde{u}_g(k|M_h)$ term shown in Eq. (14) would be unity. However, in reality, the BCGs are not always located at the true gravitational centre of the host halo due to physical factors such as galaxy evolution and interaction (see, e.g. Cui et al. 2016; Zhang et al. 2019 and references therein), as well as observational effects such as the misidentification of central galaxies or fragmentation and aggregation from group-finding algorithms (e.g. Jakobs et al. 2018; Ahad et al. 2023; Kelly et al. 2024). Thus, we need a proper model to account for the miscentring of the selected central galaxies.

In our baseline model, we adopt a two-component miscentring model defined as

$$\tilde{u}_g(k|M_h) = (1 - p_{\text{off}}) + p_{\text{off}} \times \tilde{P}_{\text{off}}(k|\mathcal{R}_{\text{off}}), \quad (19)$$

where

$$\tilde{P}_{\text{off}}(k|\mathcal{R}_{\text{off}}) = \left(\frac{1}{2x} - x \right) D_+(x) + \frac{1}{2}, \quad (20)$$

with $x \equiv (k \mathcal{R}_{\text{off}} r_s) / \sqrt{2}$, and $D_+(x)$ being the Dawson integral. This model assumes that a fraction p_{off} of BCGs is miscentred,

with the normalised radial distribution of these miscentred galaxies relative to the true halo centre following a Rayleigh distribution with a scatter of \mathcal{R}_{off} times the scale radius of the halo, r_s . The choice of the Rayleigh distribution follows Johnston et al. (2007). We explore other statistical distributions in Sect. 8. Both p_{off} and \mathcal{R}_{off} are free parameters.

The HOD term $\langle n_g|M_h \rangle$ in Eq. (14) is modelled using the conditional stellar mass function (CSMF, Yang et al. 2003). This choice is motivated by the direct connection of the CSMF to the relation between baryonic properties and halo mass, which is the focus of this study. Additionally, it facilitates the implementation of simulation predictions, which is the other key aspect of our study. Specifically, we assume a log-normal distribution for the group CSMF based on previous studies (Yang et al. 2008; Cacciato et al. 2009; van den Bosch et al. 2013; van Uitert et al. 2016):

$$\Phi(M_\star^{\text{grp}}|M_h) = \frac{1}{\ln(10) \sqrt{2\pi} M_\star^{\text{grp}} \sigma_{\log M_\star^{\text{grp}}}} \times \exp \left[-\frac{(\log M_\star^{\text{grp}} - \mu_{\log M_\star^{\text{grp}}})^2}{2\sigma_{\log M_\star^{\text{grp}}}^2} \right]. \quad (21)$$

We further validated the log-normal form using FLAMINGO simulations. Although these simulation-based tests do not account for stellar mass measurement uncertainties, previous studies have shown that the distribution of these measurement uncertainties for a stacked ensemble is also well approximated by a log-normal distribution (Yang et al. 2009; Behroozi et al. 2010). Therefore, the log-normal form remains appropriate even in the presence of measurement uncertainties, although the inferred scatter will reflect a combination of intrinsic scatter and measurement uncertainty (Moster et al. 2010; Leauthaud et al. 2012).

Equation (21) has two free parameters: the logarithmic scatter $\sigma_{\log M_\star^{\text{grp}}}$ and the logarithmic mean $\mu_{\log M_\star^{\text{grp}}}$ of the group stellar mass distribution for a given halo mass M_h . We model the logarithmic mean using a power-law scaling relation:

$$\mu_{\log M_\star^{\text{grp}}} = 11.8 + \log(A_s) + \alpha_s \log \left(\frac{M_h}{10^{14.15} h_{70}^{-1} M_\odot} \right), \quad (22)$$

where the amplitude A_s and index α_s are free parameters. The normalisation 11.8 corresponds to the logarithmic mean group stellar mass of the full GAMA sample, and the choice of $10^{14.15} h_{70}^{-1} M_\odot$ follows Viola et al. (2015). In our baseline model, we assume the logarithmic scatter to be a halo mass-independent free parameter. However, we explore a more realistic, simulation-informed scatter model in Sect. 7.

Under the assumption of sample completeness, which is valid given the high completeness of the GAMA survey and our exclusion of distribution tails (see Fig. 3), we can calculate the mean number of groups per specific observable bin as follows:

$$\langle n_g|M_h \rangle = \int_{M_\star^{\text{grp}, \text{min}}}^{M_\star^{\text{grp}, \text{max}}} dM_\star^{\text{grp}} \Phi(M_\star^{\text{grp}}|M_h), \quad (23)$$

where the integral limits, $M_\star^{\text{grp}, \text{min}}$ and $M_\star^{\text{grp}, \text{max}}$, correspond to the stacked bin boundaries, as detailed in Table 2.

We tested the impact of potential sample incompleteness by introducing an additional incompleteness function into Eq. (23), following van Uitert et al. (2016). This function, based on an error function, includes two additional free parameters. We find that these incompleteness parameters are not constrained by the

Table 3. Free parameters in our baseline model for fitting stacked ESD profiles binned by group total stellar mass.

Parameter	Prior	Constraints	Best-fit values
f_c	[0.2, 3]	$0.88^{+0.57}_{-0.24}$	1.38
A_p	[0.1, 5]	$2.19^{+1.44}_{-1.32}$	1.16
p_{off}	[0, 1]	$0.34^{+0.18}_{-0.19}$	0.50
\mathcal{R}_{off}	[0, 5]	$2.47^{+1.31}_{-0.92}$	2.26
$\sigma_{\log M_{\star}^{\text{grp}}}$	[0.01, 1]	$0.14^{+0.11}_{-0.09}$	0.09
A_s	[0, 5]	$2.36^{+0.60}_{-0.63}$	2.72
α_s	[0, 5]	$1.02^{+0.11}_{-0.11}$	1.04

Notes. Constraints are presented as the median of the marginalised posterior distributions, with 68% credible intervals. Best-fit values correspond to the parameter set that minimises the χ^2 value of the model fit.

current data, while constraints on the other parameters remain fully consistent with our baseline model. This confirms the assumption of high completeness in our sample and indicates that introducing additional incompleteness parameters into the model is unnecessary.

We used the same HOD form for the relation between halo mass and group luminosity, namely a log-normal distribution for the group conditional luminosity function (Eq. 21), with the logarithmic mean of the group luminosity scaling with halo mass through a power-law relation (Eq. 22). It is worth noting that this conditional luminosity function-based HOD model differs from those used by Viola et al. (2015) and Rana et al. (2022), who defined the HOD directly based on the average number of galaxies as a function of halo mass. Moreover, they constrained the scaling relation as the mean halo mass for a given luminosity, which differs from the logarithmic mean luminosity for a given halo mass due to the scatter. In order to compare our results with these previous results, we derived the mean halo mass for a given luminosity using Bayes theorem (e.g. Coupon et al. 2015):

$$\langle M_{\text{h}} | L_r^{\text{grp}} \rangle = \frac{\int dM_{\text{h}} n_{\text{h}}(M_{\text{h}}) \Phi(L_r^{\text{grp}} | M_{\text{h}}) M_{\text{h}}}{\int dM_{\text{h}} n_{\text{h}}(M_{\text{h}}) \Phi(L_r^{\text{grp}} | M_{\text{h}})}, \quad (24)$$

where $\Phi(L_r^{\text{grp}} | M_{\text{h}})$ is the group conditional luminosity function, following the same form as Eq. (21).

5.3. Model fitting

The baseline model outlined above contains seven free parameters, each with broad, uninformative priors, as detailed in Tables 3 and 4. We performed a joint fit to the stacked ESD measurements of the six observable bins using the specified halo model, accounting for the full correlations between the 60 data points (see Sect. 4.2). The posterior parameter space was sampled using the emcee code (Foreman-Mackey et al. 2013), an implementation of the affine-invariant Markov chain Monte Carlo (MCMC) ensemble sampler (Goodman & Weare 2010). The convergence of the MCMC chains was assessed using the integrated autocorrelation time (e.g. Goodman & Weare 2010).

6. Results from the baseline model

Figures 4 and 5 show the best-fit ESD profiles along with their 68% and 95% credible intervals for M_{\star}^{grp} and L_r^{grp} binning, re-

Table 4. Same as Table 3, but for fitting measurements binned by group r -band total luminosity.

Parameter	Prior	Constraints	Best-fit values
f_c	[0.2, 3]	$0.92^{+0.41}_{-0.24}$	0.93
A_p	[0.1, 5]	$2.15^{+1.42}_{-1.29}$	2.00
p_{off}	[0, 1]	$0.36^{+0.14}_{-0.17}$	0.38
\mathcal{R}_{off}	[0, 5]	$2.82^{+1.11}_{-0.95}$	2.66
$\sigma_{\log L_r^{\text{grp}}}$	[0.01, 1]	$0.13^{+0.10}_{-0.08}$	0.06
A_s	[0, 3]	$0.87^{+0.22}_{-0.23}$	0.94
α_s	[0, 5]	$0.88^{+0.10}_{-0.10}$	0.89

spectively. Assuming independence among the seven free parameters, our baseline model fitting achieves a reduced χ^2 of 1.03 for M_{\star}^{grp} binning and 0.84 for L_r^{grp} binning, indicating an overall good fit to the data. Tables 3 and 4 present the marginalised median constraints and best-fit values for these parameters, with uncertainties corresponding to 68% credible intervals.

Figure 6 shows the scaling relation between the mean halo mass and group r -band luminosity, compared to previous studies by Viola et al. (2015) and Rana et al. (2022). We re-emphasise that our halo model directly constrains the logarithmic mean of the group luminosity as a function of halo mass (Eq. 22), while Viola et al. (2015) and Rana et al. (2022) constrained the scaling relation as the mean halo mass for a given luminosity. To enable a comparison with these previous results, we converted our constraint using Eq. (24). Our new analysis features updated shear measurements and calibration relative to Viola et al. (2015) and employs a different modelling approach compared to these two studies (see Sect. 5.2). Despite these changes, we observe good agreement between our measurements and previous studies, verifying the reliability of our new modelling approach.

To be more quantitative, we fitted a power-law relation between the mean halo mass and the group r -band luminosity and found

$$\frac{\langle M_{\text{h}} | L_r^{\text{grp}} \rangle}{10^{14.15} h_{70}^{-1} M_{\odot}} = (0.96^{+0.24}_{-0.15}) \left(\frac{L_r^{\text{grp}}}{10^{11.8} h_{70}^{-1} L_{\odot}} \right)^{1.08^{+0.10}_{-0.09}}, \quad (25)$$

where the pivot values followed the choice of Viola et al. (2015). The linear regression is performed on all MCMC chains with the mean halo mass $\langle M_{\text{h}} | L_r^{\text{grp}} \rangle$ estimated using Eq. (24). The reported values correspond to the median of the marginalised distributions, with 68% credible intervals for the uncertainties. These results are fully consistent with those reported by Viola et al. (2015) and Rana et al. (2022), who yielded a normalisation and power-law index combination of 0.95 ± 0.14 and 1.16 ± 0.13 , and 0.81 ± 0.04 and 1.01 ± 0.07 , respectively.

While our lens sample size is larger than that used by Viola et al. (2015), the uncertainties of our final constraints are comparable to theirs. This is largely because we applied a more stringent scale cut to alleviate blending effects on small scales—we used a scale cut of $0.04 h_{70}^{-1} \text{Mpc}$ compared to their $0.03 h_{70}^{-1} \text{Mpc}$. Additionally, we excluded the tails of the L_r^{grp} distributions to mitigate potential group detection effects, as shown in Fig 3. In this sense, we consciously traded some statistical power for increased robustness.

Figure 7 compares the constrained scaling relation to the simulation predictions. Unlike in Fig. 6, here we present the di-

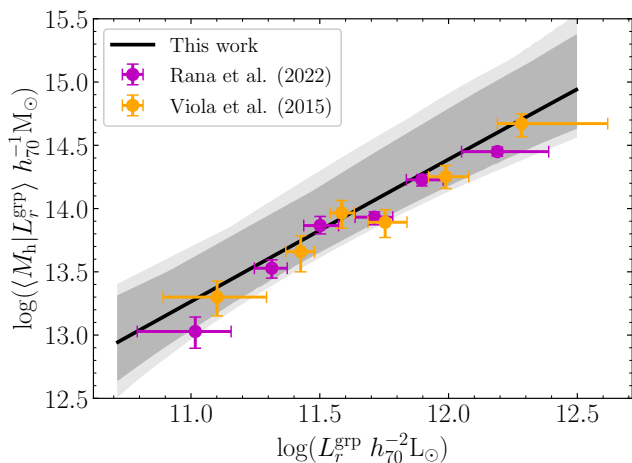


Fig. 6. Scaling relation between the halo mass and r -band total luminosity of galaxy groups from our baseline model. The black line shows the best-fit results, with the shaded regions illustrating the corresponding 68% and 95% credible intervals. The parameter values are provided in Table 4. The results are compared to previous measurements from Viola et al. (2015) (orange points) and Rana et al. (2022) (magenta points). All three measurements are based on the GAMA group catalogue, but with different shear measurements and modelling approaches. We note that the scaling relation is demonstrated as the mean halo mass for a given luminosity.

rect constraints on the logarithmic mean of the group stellar mass as a function of halo mass without further transformation, as this scaling relation can be easily measured from the mock group catalogues (Sect. 3.4). To delineate the mass regions covered by our data measurements from those inferred through model extrapolation, we also include data points for the six observable bins where we measured the stacked ESD signals (Sect. 4.1). These data points were estimated by running separate MCMC chains for each observable bin, fixing the baseline model parameters to their best-fit values from the joint fit, except for the A_s parameter. The new A_s constraints can provide halo mass estimates for each stacked bin using Eq. (22), given that the mean group stellar mass for each stacked bin is well measured. The uncertainties in the halo mass estimates were calculated using the 68% credible intervals of the new constrained A_s distributions. The results indicate that our current sample is sensitive to halos in the mass range of $\sim 10^{13.1}$ to $\sim 10^{14.6} h_{70}^{-1} M_{\odot}$. The relation outside this mass range is an extrapolation from our assumed power-law model of Eq. (22), thus requiring caution when interpreting these extrapolated regions, especially if the scaling relation deviates from the assumed power-law behaviour.

In general, we find a good agreement between the results inferred from our measurements using our baseline model and all simulation predictions at the high-mass end. At the low-mass end, the FLAMINGO simulations slightly over-predict the mean group stellar mass for a given halo mass, while the L-GALAXIES SAM predictions remain closer to our data constraints. The TNG100-1 simulations do not have sufficient volume to cover the halo mass regions measured by our sample, thus cannot be directly tested by our measurements, but they sit between the FLAMINGO and L-GALAXIES SAM predictions at the low-mass end.

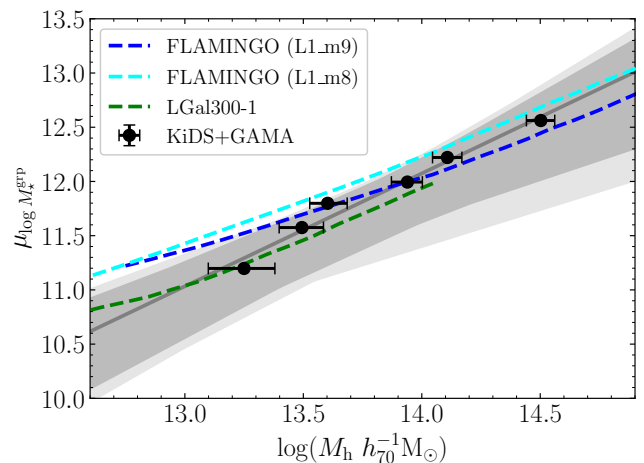


Fig. 7. Scaling relation between the total stellar mass of galaxy groups and their halo masses from our baseline model. The grey line shows the best-fit results, with the shaded regions illustrating the corresponding 68% and 95% credible intervals. The parameter values are provided in Table 3. The black points represent the halo masses calculated by allowing A_s to vary for each stacked bin while fixing other parameters to their best-fit values from the joint fit. The error bars correspond to the 68% credible intervals of the new constrained A_s distributions. The corresponding $\mu_{\log M_{\star}^{\text{grp}}}$ values for these points are the mean log-stellar mass of all groups in the given stacked bin. Predictions from simulations, represented by dashed lines, are estimated from the mock catalogue built in Sect. 3. All values of $\mu_{\log M_{\star}^{\text{grp}}}$ are converted to a h_{70} cosmology for comparison, with M_{\star}^{grp} from simulations scaled as h_{70}^{-1} and those from observations scaled as h_{70}^{-2} . We note that the scaling relation is demonstrated as the mean log-stellar mass at a fixed halo mass.

7. Simulation-informed scatter model

After validating the simulation predictions against the scaling relation constrained by our baseline model, we refine the halo model by incorporating insights from simulations. In this study, we focus on one of the key simplifications in the current halo model, where the scatter in the group stellar mass distribution is assumed to be mass-independent. This assumption contrasts recent findings from semi-empirical models (e.g. Bradshaw et al. 2020), as well as those from semi-analytical models and hydrodynamical simulations (e.g. Pei et al. 2024). With improved measurement statistics and wider coverage of the halo mass range, revisiting this simplification becomes important.

Figure 8 shows the scatter in the group stellar mass distribution as a function of halo mass, measured from the mock group catalogues built from simulations (Sect. 3.4). We observe a general decreasing trend in scatter with increasing halo mass, except for the L-GALAXIES SAM results, which show an increasing trend at the low halo mass end. However, when considering the uncertainties in current measurements, as indicated by the shaded region representing the 68% credible interval of the constant scatter constrained by our baseline model, this scatter trend is relatively small within the halo mass range covered by our measurements. Therefore, instead of attempting to directly constrain this scatter trend from our measurements, we opt to incorporate this simulation-informed scatter trend into our halo model and assess how it affects the constrained scaling relation.

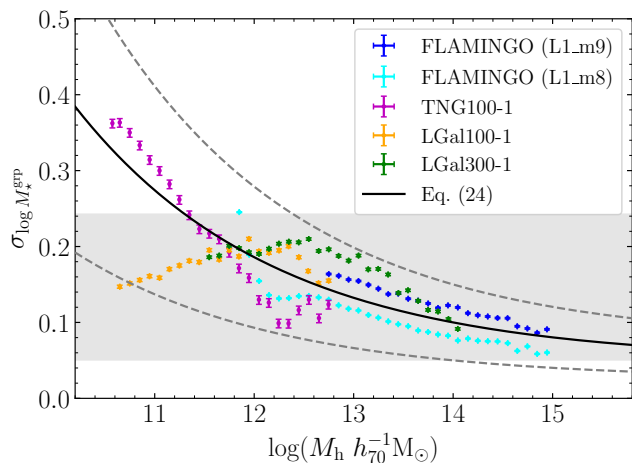


Fig. 8. Scatter in the group stellar mass distribution as a function of halo mass, measured from cosmological simulations. The shaded region indicates the 68% credible interval of the constant scatter constrained by our baseline model (Table 3). The solid and dashed lines represent the scatter model from Eq. (26) with A_σ set to 0.1 and 0.1 ± 0.05 , respectively.

Specifically, we model this scatter-halo mass relation with an exponential equation of the form:

$$\sigma_{\log M_*^{\text{grp}}} \equiv \frac{A_\sigma}{2} \left[\exp \left(-0.5 \log \left(\frac{M_h}{10^{14} h_{70}^{-1} M_\odot} \right) \right) + 1 \right], \quad (26)$$

where A_σ is the amplitude, allowed to vary to account for uncertainties among different simulation predictions. This equation captures the general behaviour of the scatter, as shown in Fig. 8: a decreasing trend with increasing halo mass at lower mass scales and flattens out at higher mass scales. We find that a Gaussian prior with a mean of 0.1 (solid line in the plot) and a standard deviation of 0.05 (dashed lines in the plot) for A_σ sufficiently covers the uncertainties among different simulation predictions. We tested using a flat prior and found consistent results, confirming that our current data statistics are insufficient to distinguish subtle differences in the scatter-halo mass relation.

Figure 9 compares the constraints on the scaling relation parameters between the new scatter model and the baseline constant scatter model. The new results remain consistent with the baseline model. However, we observe tighter and less degenerate constraints, demonstrating the benefits of including the simulation-informed scatter model even with the current data statistics. The consistency between the constant scatter model and the variable scatter model is largely due to the minor scatter variation within the halo mass range covered by our current measurements ($\sim 10^{13.1}$ to $\sim 10^{14.6} h_{70}^{-1} M_\odot$) relative to the measurement uncertainties. With future analyses extending to lower mass ranges and improved statistics, we anticipate a greater impact from the scatter model, warranting continued investigation of the scatter-halo mass relation.

8. Sensitivity to the miscentring models

To test the importance of miscentring modelling on our results, we focus on two aspects: first, the necessity of accounting for miscentring effects, and second, the sensitivity of the current analysis to different statistical miscentring models. Besides the

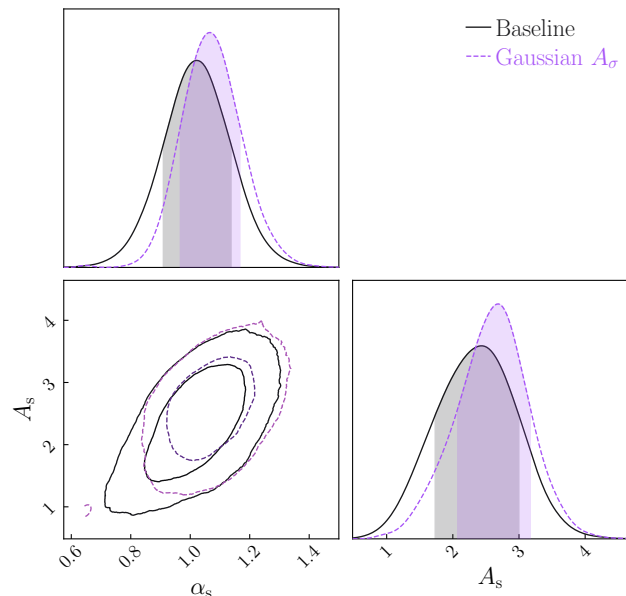


Fig. 9. Comparison of projected posterior distributions between the baseline constant scatter model and the simulation-derived scatter model for the two parameters of the scaling relation in Eq. (22). The ‘Gaussian A_σ ’ refers to the use of scatter-halo mass relation of Eq. (26) with a Gaussian prior for A_σ . The contours represent the 68% and 95% credible intervals, smoothed with a matched elliptical Gaussian kernel density estimator.

Rayleigh distribution, the two other commonly used statistical distributions are the Gaussian distribution, formulated as

$$\tilde{P}_G(k|\mathcal{R}_{\text{off}}) = \exp \left[-\frac{1}{2} k^2 (r_s \mathcal{R}_{\text{off}})^2 \right], \quad (27)$$

and the Gamma distribution, formulated as

$$\tilde{P}_\Gamma(k|\mathcal{R}_{\text{off}}) = \frac{1}{3} \frac{3 - k^2 (r_s \mathcal{R}_{\text{off}})^2}{(k^2 (r_s \mathcal{R}_{\text{off}})^2 + 1)^3}. \quad (28)$$

Figure 10 compares the constraints on the scaling relation parameters from different treatments of miscentring effects. Noticeable shifts in both the scaling relation amplitude A_s and scatter $\sigma_{\log M_*^{\text{grp}}}$ are observed when we ignore the miscentring effects in the modelling. However, shifts among different statistical models are negligible given the current uncertainties, implying that our current analysis is insensitive to the subtle differences in the assumed miscentring distributions.

These conclusions hold when we check the constraints on the parameters describing the halo inner mass distribution, as is shown in Fig. 11. Without a miscentring model, the scaling parameters for mass-concentration, f_c , and point mass contribution, A_p , have much narrower but potentially biased constraints due to the high degeneracy between these parameters and the miscentring parameters. This finding is consistent with the results of Viola et al. (2015), and it cautions against interpreting the halo mass concentration constrained by weak lensing analyses without a realistic miscentring model.

When comparing the reduced χ^2 values of these different models, none stands out as superior to the others. The model with the Gamma distribution has the best reduced χ^2 value of

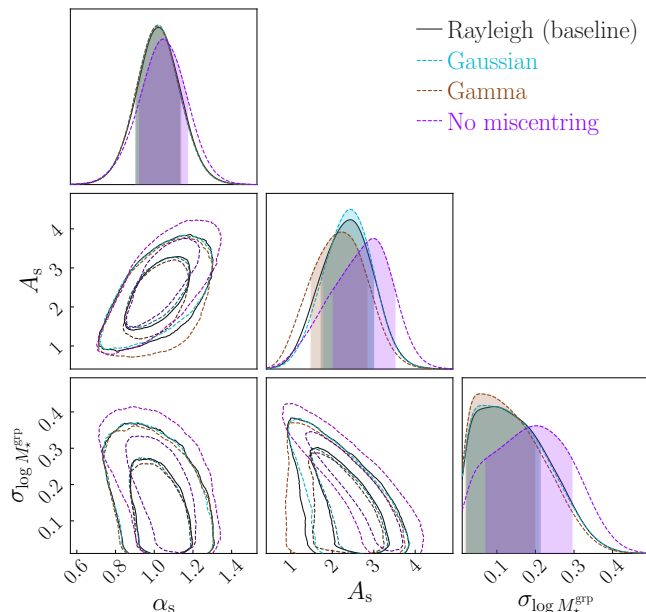


Fig. 10. Comparison of projected posterior distributions across different treatments of miscentring effects for the scaling relation parameters and scatter parameter. The contours represent the 68% and 95% credible intervals, smoothed with a matched elliptical Gaussian kernel density estimator.

1.02, while the model without miscentring shows the worst reduced χ^2 value of 1.05. However, it is important to note that the calculated reduced χ^2 value assumes independence among free parameters, even though some degeneracy between parameters is observed in the posterior distributions. Therefore, the reported reduced χ^2 values should be seen as indicative rather than definitive for ruling out models.

In practice, the cause of miscentring in a group sample is more intricate than what the adopted statistical distributions can account for. For example, [Ahad et al. \(2023\)](#) found that line-of-sight projections, which result in a discrepancy between the projected and intrinsic luminosity, account for approximately half of the identified miscentred groups in their simulations. Furthermore, the aggregation and fragmentation effects, referring to the phenomena where two small groups are identified as a single larger group, and a single large group is identified as several smaller groups, respectively, are common in real data group-finding algorithms (see Appendix A of [Jakobs et al. 2018](#)), which further complicate the distribution of miscentred BCGs. Developing a sophisticated miscentring model that accounts for these more complicated selection effects is still important but can only be addressed using simulations that include the sample selection from a specific survey, which is beyond the scope of our current study.

9. Conclusions

We conducted a weak lensing analysis using the latest KiDS-1000 shear catalogue (v2) from [Li et al. \(2023a\)](#) to constrain the scaling relation between baryonic observables and halo masses for galaxy groups identified by GAMA. Using a baseline halo model with seven free parameters, we achieved a good fit to the measured ESD signals, with a reduced χ^2 of 1.03 for stacks

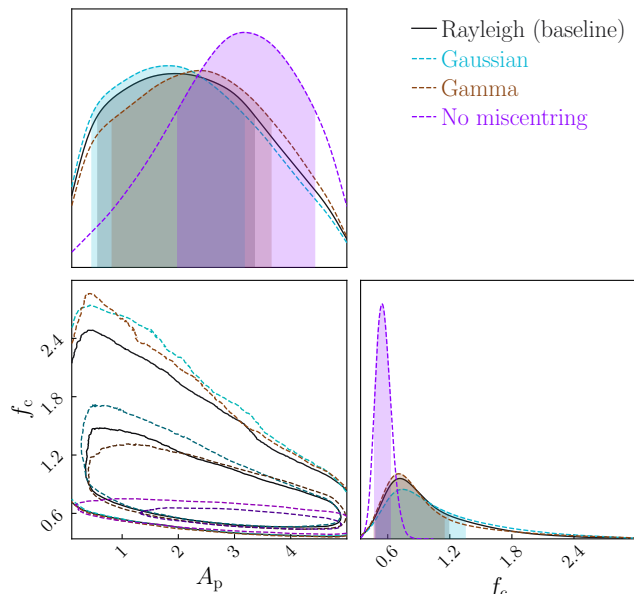


Fig. 11. Same comparison as Fig. 10 but for the parameters of mass concentration f_c and point mass contribution A_p .

based on group stellar masses and 0.84 for stacks based on group r -band luminosity.

Compared to previous studies by [Viola et al. \(2015\)](#) and [Rana et al. \(2022\)](#), we refined the lens sample selection, updated the shear measurements and calibration, and adopted a new modelling approach based on the conditional stellar mass (or luminosity) function within the halo model framework. Despite these changes, our constraints on the scaling relation between the mean halo mass and group r -band luminosity are fully consistent with the ones from previous studies. Specifically, our baseline model yields a normalisation and power-law index combination of $0.96^{+0.24}_{-0.15}$ and $1.08^{+0.10}_{-0.09}$ (Eq. 25), whereas [Viola et al. \(2015\)](#) and [Rana et al. \(2022\)](#) reported combinations of 0.95 ± 0.14 and 1.16 ± 0.13 , and 0.81 ± 0.04 and 1.01 ± 0.07 , respectively.

We further compared the constrained group stellar mass-halo mass relation to predictions from the latest FLAMINGO cosmological simulations, as well as the L-GALAXIES SAM implemented in IllustrisTNG gravity-only simulations. We find a general agreement between our measurements and simulation predictions for halos with masses $\geq 10^{13.5} h_{70}^{-1} M_{\odot}$. For halos with masses below this value, the FLAMINGO simulations slightly over-predict group stellar masses, while the L-GALAXIES SAM shows better agreement with our measurements. These findings are consistent with those of [Schaye et al. \(2023\)](#), who found that the FLAMINGO stellar-to-halo mass relation for central galaxies is higher at the low-halo-mass end compared to the semi-empirical UniverseMachine model results from [Behroozi et al. \(2019\)](#).

After validating the simulation predictions, we improved our baseline halo model by incorporating the simulation-informed scatter in the group stellar mass distribution as a function of halo mass. Using an exponential equation with a variable amplitude (Eq. 26), the improved halo model captures the general decreasing trend of scatter with increasing halo masses and accounted for uncertainties among different simulation predictions. Although the current measurement statistics are insufficient to

directly constrain the variable scatter, we find that the updated model yields tighter constraints on the scaling relation parameters, highlighting the advantages of simulation-informed halo modelling.

We tested the robustness of our scaling relation results against sensible changes to the miscentring modelling. We verified that including a statistical model to account for the potential miscentring of the selected central galaxies is necessary. Ignoring this miscentring effect would bias not only the estimation of the mass concentration in the inner region of the halo profile but also the scaling relation constraints. When testing different statistical models for miscentring, we observe minor shifts in the scaling relation parameters that are well within the current measurement uncertainties. This suggests that the current data statistics are insufficient to distinguish among different statistical models of miscentring. However, with improved statistics of lens samples from upcoming spectroscopic surveys such as the 4MOST Wide Area VISTA Extragalactic Survey (WAVES; Driver et al. 2019) and Hemisphere Survey of the Nearby Universe (4HS; Taylor et al. 2023), along with significantly enhanced weak lensing measurements from the ESA *Euclid* (Euclid Collaboration et al. 2023) and *Rubin* LSST (Ivezić et al. 2019) surveys, we will be able to measure galaxy-galaxy lensing signals down to much smaller scales. To accurately model these small-scale lensing signals, further development of realistic miscentring models that account for observational effects is warranted.

Acknowledgements. We thank Andrej Dvornik for his assistance with the KiDS galaxy-galaxy lensing pipeline, Roi Kugel for his help with the FLAMINGO products, and Maciej Bilicki and Giorgio Francesco Lesci for their valuable comments. We acknowledge support from: the Netherlands Research School for Astronomy (SSL); the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme with Grant agreement No. 101053992 (SSL, HHO). The KiDS data used in this paper are based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018 and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples). GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including GALEX MIS, VST KiDS, VISTA VIKING, WISE, Herschel-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is <https://www.gama-survey.org/>. Author contributions: All authors contributed to the development and writing of this paper. The authorship list is given in two groups: the lead authors (SSL, HHO, KK), followed by an alphabetical group, which includes those who are key contributors to both the scientific analysis and the data products.

References

Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, *Phys. Rev. D*, 105, 023520
 Ahad, S. L., Bahé, Y. M., & Hoekstra, H. 2023, *MNRAS*, 518, 3685
 Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, *ARA&A*, 49, 409
 Angulo, R. E. & Hahn, O. 2022, *Living Reviews in Computational Astrophysics*, 8, 1
 Ayromlou, M., Nelson, D., Yates, R. M., et al. 2021, *MNRAS*, 502, 1051
 Bahar, Y. E., Bulbul, E., Clerc, N., et al. 2022, *A&A*, 661, A7
 Baltz, E. A., Marshall, P., & Oguri, M. 2009, *J. Cosmology Astropart. Phys.*, 2009, 015
 Barnes, D. J., Vogelsberger, M., Pearce, F. A., et al. 2021, *MNRAS*, 506, 2533
 Bartelmann, M. & Schneider, P. 2001, *Phys. Rep.*, 340, 291
 Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, *MNRAS*, 488, 3143
 Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, *ApJ*, 717, 379

Berlind, A. A. & Weinberg, D. H. 2002, *ApJ*, 575, 587
 Bertin, E. & Arnouts, S. 1996, *A&AS*, 117, 393
 Biffi, V., Borgani, S., Murante, G., et al. 2016, *ApJ*, 827, 112
 Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, *Nature*, 311, 517
 Bocquet, S., Saro, A., Dolag, K., & Mohr, J. J. 2016, *MNRAS*, 456, 2361
 Bradshaw, C., Leauthaud, A., Hearin, A., Huang, S., & Behroozi, P. 2020, *MNRAS*, 493, 337
 Brouwer, M. M., Cacciato, M., Dvornik, A., et al. 2016, *MNRAS*, 462, 4451
 Bruzual, G. & Charlot, S. 2003, *MNRAS*, 344, 1000
 Cacciato, M., van den Bosch, F. C., More, S., et al. 2009, *MNRAS*, 394, 929
 Cacciato, M., van Uitert, E., & Hoekstra, H. 2014, *MNRAS*, 437, 377
 Castro, T., Borgani, S., Dolag, K., et al. 2021, *MNRAS*, 500, 2316
 Cavaliere, A. & Fusco-Femiano, R. 1976, *A&A*, 49, 137
 Chabrier, G. 2003, *PASP*, 115, 763
 Chisari, N. E., Richardson, M. L. A., Devriendt, J., et al. 2018, *MNRAS*, 480, 3962
 Cole, S. 1991, *ApJ*, 367, 45
 Cooray, A. 2006, *MNRAS*, 365, 842
 Cooray, A. & Sheth, R. 2002, *Phys. Rep.*, 372, 1
 Coupon, J., Arnouts, S., van Waerbeke, L., et al. 2015, *MNRAS*, 449, 1352
 Crain, R. A. & van de Voort, F. 2023, *ARA&A*, 61, 473
 Cui, W., Power, C., Biffi, V., et al. 2016, *MNRAS*, 456, 2566
 Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, 292, 371
 de Graaff, A., Trayford, J., Franx, M., et al. 2022, *MNRAS*, 511, 2544
 de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, *Experimental Astronomy*, 35, 25
 Debackere, S. N. B., Schaye, J., & Hoekstra, H. 2020, *MNRAS*, 492, 2285
 Debackere, S. N. B., Schaye, J., & Hoekstra, H. 2021, *MNRAS*, 505, 593
 Driver, S. P., Bellstedt, S., Robotham, A. S. G., et al. 2022, *MNRAS*, 513, 439
 Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, *MNRAS*, 413, 971
 Driver, S. P., Liske, J., Davies, L. J. M., et al. 2019, *The Messenger*, 175, 46
 Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, *MNRAS*, 390, L64
 Dvornik, A., Cacciato, M., Kuijken, K., et al. 2017, *MNRAS*, 468, 3251
 Dvornik, A., Hoekstra, H., Kuijken, K., et al. 2018, *MNRAS*, 479, 1240
 Eckmiller, H. J., Hudson, D. S., & Reiprich, T. H. 2011, *A&A*, 535, A105
 Edge, A., Sutherland, W., Kuijken, K., et al. 2013, *The Messenger*, 154, 32
 Elahi, P. J., Cañas, R., Poulton, R. J. J., et al. 2019, *PASA*, 36, e021
 Engler, C., Pillepich, A., Joshi, G. D., et al. 2021, *MNRAS*, 500, 3957
 Euclid Collaboration, Castro, T., Fumagalli, A., et al. 2023, *A&A*, 671, A100
 Euclid Collaboration, Mellier, Y., Abdurro’uf, et al. 2025, *A&A*, 697, A1
 Evrard, A. E., Metzler, C. A., & Navarro, J. F. 1996, *ApJ*, 469, 494
 Farrow, D. J., Cole, S., Norberg, P., et al. 2015, *MNRAS*, 454, 2120
 Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, *MNRAS*, 467, 1627
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
 Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2021, *MNRAS*, 501, 2983
 Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, 445, 175
 Giblin, B., Heymans, C., Asgari, M., et al. 2021, *A&A*, 645, A105
 Goodman, J. & Weare, J. 2010, *Communications in Applied Mathematics and Computational Science*, 5, 65
 Gouin, C., Gavazzi, R., Pichon, C., et al. 2019, *A&A*, 626, A72
 Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, *MNRAS*, 413, 101
 Guzik, J. & Seljak, U. 2002, *MNRAS*, 335, 311
 Han, J., Eke, V. R., Frenk, C. S., et al. 2015, *MNRAS*, 446, 1356
 Helling, W. A., Schaller, M., Frenk, C. S., et al. 2016, *MNRAS*, 461, L11
 Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, *MNRAS*, 451, 2663
 Hernández-Martín, B., Schrabback, T., Hoekstra, H., et al. 2020, *A&A*, 640, A117
 Heymans, C., Van Waerbeke, L., Bacon, D., et al. 2006, *MNRAS*, 368, 1323
 Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, *A&A*, 647, A124
 Hoekstra, H., Franx, M., Kuijken, K., & Squires, G. 1998, *ApJ*, 504, 636
 Hoekstra, H., Herbonnet, R., Muzzin, A., et al. 2015, *MNRAS*, 449, 685
 Hoekstra, H., Kannawadi, A., & Kitching, T. D. 2021, *A&A*, 646, A124
 Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
 Jakobs, A., Viola, M., McCarthy, I., et al. 2018, *MNRAS*, 480, 3338
 Jansen, H., Tewes, M., Schrabback, T., et al. 2024, *A&A*, 683, A240
 Johnston, D. E., Sheldon, E. S., Wechsler, R. H., et al. 2007, *arXiv e-prints*, arXiv:0709.1159
 Kaiser, N. & Squires, G. 1993, *ApJ*, 404, 441
 Kaiser, N., Squires, G., & Broadhurst, T. 1995, *ApJ*, 449, 460
 Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, *A&A*, 624, A92
 Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, *MNRAS*, 303, 188
 Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, *MNRAS*, 264, 201
 Kelly, P. M., Jobel, J., Eiger, O., et al. 2024, *MNRAS*, 533, 572
 Kettula, K., Giodini, S., van Uitert, E., et al. 2015, *MNRAS*, 451, 1460

- Kitching, T. D. & Deshpande, A. C. 2022, *The Open Journal of Astrophysics*, 5, 6
- Kugel, R., Schaye, J., Schaller, M., et al. 2023, *MNRAS*, 526, 6103
- Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, *A&A*, 625, A2
- Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, *MNRAS*, 454, 3500
- Le Brun, A. M. C., McCarthy, I. G., Schaye, J., & Ponman, T. J. 2014, *MNRAS*, 441, 1270
- Leauthaud, A., Finoguenov, A., Kneib, J.-P., et al. 2010, *ApJ*, 709, 97
- Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, *ApJ*, 744, 159
- Li, S.-S., Hoekstra, H., Kuijken, K., et al. 2023a, *A&A*, 679, A133
- Li, S.-S., Kuijken, K., Hoekstra, H., et al. 2023b, *A&A*, 670, A100
- Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, *MNRAS*, 452, 2087
- Logan, C. H. A., Maughan, B. J., Diaferio, A., et al. 2022, *A&A*, 665, A124
- Loveday, J., Norberg, P., Baldry, I. K., et al. 2012, *MNRAS*, 420, 1239
- Luppino, G. A. & Kaiser, N. 1997, *ApJ*, 475, 20
- Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, *MNRAS*, 368, 715
- McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, *MNRAS*, 465, 2936
- McCarthy, I. G., Schaye, J., Ponman, T. J., et al. 2010, *MNRAS*, 406, 822
- Mead, A. J., Peacock, J. A., Heymans, C., Joudaki, S., & Heavens, A. F. 2015, *MNRAS*, 454, 1958
- Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, *ApJ*, 710, 903
- Muñoz-Cuartas, J. C., Macciò, A. V., Gottlöber, S., & Dutton, A. A. 2011, *MNRAS*, 411, 584
- Murata, R., Nishimichi, T., Takada, M., et al. 2018, *ApJ*, 854, 120
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, 490, 493
- Nelson, D., Pillepich, A., Genel, S., et al. 2015, *Astronomy and Computing*, 13, 12
- Nelson, D., Springel, V., Pillepich, A., et al. 2019, *Computational Astrophysics and Cosmology*, 6, 2
- Oguri, M. & Hamana, T. 2011, *MNRAS*, 414, 1851
- Peacock, J. A. & Smith, R. E. 2000, *MNRAS*, 318, 1144
- Pei, W., Guo, Q., Shao, S., He, Y., & Gu, Q. 2024, *MNRAS*, 531, 2262
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, *MNRAS*, 473, 4077
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, 594, A13
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6
- Pop, A.-R., Hernquist, L., Nagai, D., et al. 2022, arXiv e-prints, arXiv:2205.11528
- Rana, D., More, S., Miyatake, H., et al. 2022, *MNRAS*, 510, 5408
- Rasia, E., Ettori, S., Moscardini, L., et al. 2006, *MNRAS*, 369, 2013
- Robertson, N. C., Sifón, C., Asgari, M., et al. 2024, *A&A*, 681, A87
- Robotham, A. S. G., Norberg, P., Driver, S. P., et al. 2011, *MNRAS*, 416, 2640
- Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, *Astronomy and Computing*, 10, 121
- Schaller, M., Borrow, J., Draper, P. W., et al. 2024, *MNRAS*, 530, 2378
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, 446, 521
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS*, 402, 1536
- Schaye, J., Kugel, R., Schaller, M., et al. 2023, *MNRAS*, 526, 4978
- Schneider, A., Stoira, N., Refregier, A., et al. 2020, *J. Cosmology Astropart. Phys.*, 2020, 019
- Seitz, C. & Schneider, P. 1997, *A&A*, 318, 687
- Seljak, U. 2000, *MNRAS*, 318, 203
- Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P., & McCarthy, I. G. 2011, *MNRAS*, 417, 2020
- Sifón, C., Cacciato, M., Hoekstra, H., et al. 2015, *MNRAS*, 454, 3938
- Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, *MNRAS*, 341, 1311
- Springel, V. 2010, *MNRAS*, 401, 791
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, *MNRAS*, 328, 726
- Takada, M. & Jain, B. 2003, *MNRAS*, 340, 580
- Taylor, E. N., Cluver, M., Bell, E., et al. 2023, *The Messenger*, 190, 46
- Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, *MNRAS*, 418, 1587
- Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, *ApJ*, 724, 878
- Tyson, J. A., Valdes, F., & Wenk, R. A. 1990, *ApJ*, 349, L1
- Vale, A. & Ostriker, J. P. 2004, *MNRAS*, 353, 189
- van Daalen, M. P., McCarthy, I. G., & Schaye, J. 2020, *MNRAS*, 491, 2424
- van Daalen, M. P., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2011, *MNRAS*, 415, 3649
- van den Bosch, F. C., More, S., Cacciato, M., Mo, H., & Yang, X. 2013, *MNRAS*, 430, 725
- van den Bosch, J. L., Wright, A. H., Hildebrandt, H., et al. 2022, *A&A*, 664, A170
- van Uitert, E., Cacciato, M., Hoekstra, H., et al. 2016, *MNRAS*, 459, 3251
- van Uitert, E., Hoekstra, H., Velander, M., et al. 2011, *A&A*, 534, A14
- Velliscig, M., Cacciato, M., Hoekstra, H., et al. 2017, *MNRAS*, 471, 2856
- Viola, M., Cacciato, M., Brouwer, M., et al. 2015, *MNRAS*, 452, 3529
- Vogelsberger, M., Genel, S., Sijacki, D., et al. 2013, *MNRAS*, 436, 3031
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, 444, 1518
- Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020, *Nature Reviews Physics*, 2, 42
- von der Linden, A., Mantz, A., Allen, S. W., et al. 2014, *MNRAS*, 443, 1973
- Wang, C., Li, R., Zhu, K., et al. 2024, *MNRAS*, 527, 1580
- Wechsler, R. H. & Tinker, J. L. 2018, *ARA&A*, 56, 435
- White, S. D. M. & Frenk, C. S. 1991, *ApJ*, 379, 52
- White, S. D. M. & Rees, M. J. 1978, *MNRAS*, 183, 341
- Wright, A. H., Robotham, A. S. G., Bourne, N., et al. 2016, *MNRAS*, 460, 765
- Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, *MNRAS*, 339, 1057
- Yang, X., Mo, H. J., & van den Bosch, F. C. 2008, *ApJ*, 676, 248
- Yang, X., Mo, H. J., & van den Bosch, F. C. 2009, *ApJ*, 695, 900
- Zhang, Y., Jeltama, T., Hollowood, D. L., et al. 2019, *MNRAS*, 487, 2578

Appendix A: Higher-order shear biases in *lensfit* measurements

In a weak lensing analysis, it is typically assumed that shear bias does not depend on the shear if the signal is small, and there is no cross-talk between the two shear components (e.g. Heymans et al. 2006). However, this assumption warrants reconsideration when studying galaxy clusters and groups, where the shear amplitude near the mass centre can be several times larger than that of cosmic shear. This appendix examines the potential higher-order biases in the KiDS *lensfit* shear measurements in these regions.

Before investigating the higher-order biases in high-shear regimes, it is instructive to first illustrate the typical shear amplitudes encountered in studies of galaxy clusters and groups. To do this, we constructed a toy model based on the Navarro-Frenk-White (NFW, Navarro et al. 1997) profile, with the mass-concentration relation from Duffy et al. (2008). We adopted the lens system geometry corresponding to the extreme cases in our analysis: a lens redshift of 0.3, a source redshift of 0.9, and a projected separation of $0.05 h_{70}^{-1} \text{Mpc}$. Figure A.1 shows the tangential shear amplitude as a function of lens halo mass from this toy model. The shear amplitude reaches 0.1 for a halo mass of $\sim 10^{14.9} h_{70}^{-1} \text{M}_{\odot}$. However, it is important to note that for galaxy group studies using KiDS-like ground-based data, measurements in this regime are already heavily impacted by blending effects, as indicated in Figs. 4 and 5, where the innermost point in the highest mass bin shows a drop with large uncertainties.

To study the higher-order shear biases, we generated a set of image simulations covering 37 different shear setups. These include one zero-shear simulation (0, 0) and 36 non-zero shear simulations. For the non-zero shear simulations, the input shear amplitudes range from 0.014 to 0.14 per component, with four combinations: $(\pm\gamma, \pm\gamma)$. Each shear setup contains 18 KiDS *r*-band tile images, paired with counterparts where galaxies are rotated by 90 degrees to reduce shape noise. All simulations were generated using the KiDS MultiBand_ImSim pipeline (Li et al. 2023b)⁸, built on the GalSim package (Rowe et al. 2015)⁹.

The zero-shear simulation was used to account for correlations across different shear setups, as they share the same galaxy population, observational conditions, and noise realisations. This approach significantly reduces uncertainties in the shear bias estimates and is valid as we are only concerned with the bias changes relative to the input shear amplitude rather than the absolute shear bias. We denote the zero-shear-subtracted shear estimates as $\tilde{\gamma}_i^{\text{obs}}$. To capture higher-order biases, we extend the linear shear bias model by including additional terms:

$$\tilde{\gamma}_i^{\text{obs}} - \gamma_i^{\text{input}} = \tilde{m}_i \gamma_i^{\text{input}} + \tilde{c}_i + \tilde{d}_i (\gamma_i^{\text{input}})^2 + \tilde{q}_i (\gamma_i^{\text{input}})^3 + \tilde{m}_{\perp,i} \gamma_j^{\text{input}}, \quad (\text{A.1})$$

where subscripts *i* and *j* represent different shear components. The equation includes higher-order terms up to the third order, and also has a linear cross-talk term $\tilde{m}_{\perp,i}$. We used tildes on the bias parameters to distinguish our estimates from the true shear bias.

Figure A.2 shows the measured biases along with the fitting results for both components. The corresponding shear bias values are presented in Table A.1. We observe clear non-linear behaviour for $|\gamma_i| \gtrsim 0.1$, which is dominated by the third-order

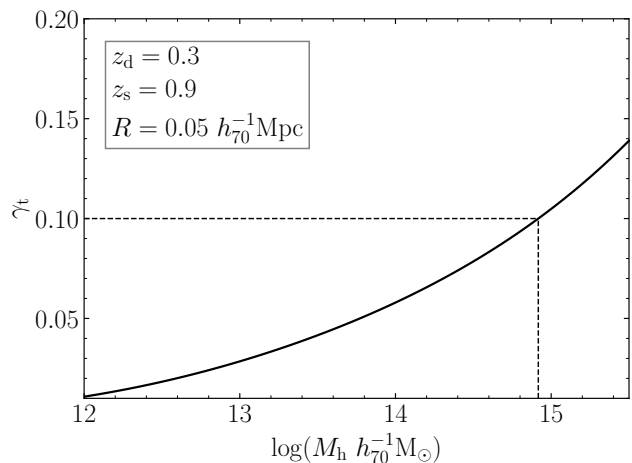


Fig. A.1. Amplitude of tangential shear as a function of lens halo mass for a toy model based on the NFW profile, with the lens system geometry listed in the figure. The geometry corresponds to the extreme cases from our analysis, where measurements are already heavily affected by blending effects. In typical cases, the majority of shear amplitudes for a given halo mass are much smaller than what is shown here.

term. However, the fitting results also reveal small but non-zero quadratic and cross-talk terms.

To further investigate where these higher-order terms arise, we traced back to the sample detection and selection processes, which are also known to introduce biases at the percent level (e.g. Fenech Conti et al. 2017; Kannawadi et al. 2019; Hernández-Martín et al. 2020; Hoekstra et al. 2021). Besides, we emphasise that, in higher-shear regimes, some fundamental assumptions based on the small shear signal may no longer hold. Specifically, the transformation between the intrinsic complex ellipticity, ϵ^s , and the shear distorted ellipticity,

$$\epsilon^{\text{obs}} = \frac{\epsilon^s + \gamma}{1 + \gamma^* \epsilon^s}, \quad (\text{A.2})$$

cannot be simplified as $\epsilon^{\text{obs}} \approx \epsilon^s + \gamma$ (Seitz & Schneider 1997; Bartelmann & Schneider 2001). The asterisk in the equation denotes the complex conjugate. In other words, the averaged observed ellipticity per component will no longer provide an unbiased estimate of the underlying shear per component, even the full sample has $\langle \epsilon^s \rangle = 0$.

Given these concerns, we measured the shear biases for three other samples, all based on a perfect galaxy shape measurement algorithm, meaning a direct use of Eq. (A.2) to recover the observed galaxy ellipticity. For the first sample, we used all galaxies in our input sample with shape noise cancellation, where $\langle \epsilon^s \rangle = 0$ holds by design. This helps identify biases introduced by the assumption of small shear. For the second sample, we used galaxies detected by SExtractor with observed magnitudes in the range of (20, 24.5). This magnitude range is close to those measurable by the *lensfit* algorithm in the KiDS data. For the third sample, we used galaxies selected by *lensfit*, which removes artefacts, identified stars, poorly resolved objects, blended objects, and so on (see Li et al. 2023b for the detailed selection criteria).

The shear biases estimated from these samples are also presented in Table A.1. Consistent with findings from Kannawadi et al. (2019) and Hoekstra et al. (2021), we detect percent-level \tilde{m}_i biases from SExtractor detection and *lensfit* selection. Moreover, we find that the quadratic and cross-talk terms are already

⁸ https://github.com/KiDS-WL/MultiBand_ImSim

⁹ <https://github.com/GalSim-developers/GalSim>

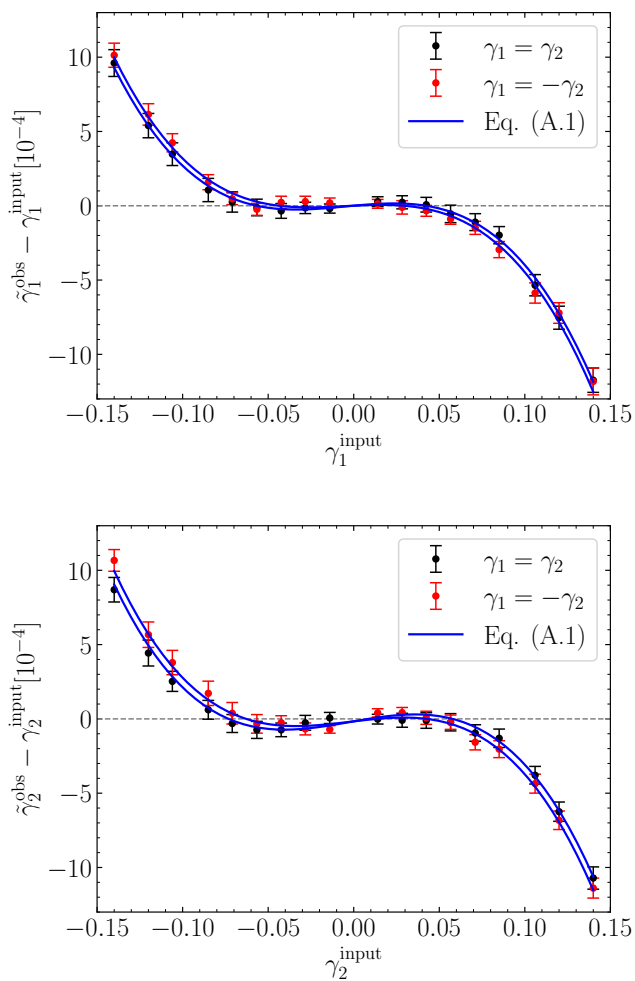


Fig. A.2. Difference between the estimated and input shear values as a function of the input shear. The estimated shears are corrected using the zero-shear simulations to account for correlations across different shear setups, as described in the main text. Error bars represent the uncertainties in the weighted mean estimates, calculated by bootstrapping the 18 independent shear estimates for each input shear value. The black and red points distinguish between simulations with the two shear components having the same or opposite signs. The blue lines show the fitting results of Eq. (A.1), with the constrained parameter values presented in Table A.1.

present at similar levels as in the *lensfit* measurement sample, while the third-order term remains relatively small. This implies that the higher-order biases have different origins, with the former primarily arising from the sample selection while the latter mainly arising from the shape measurement.

Interestingly, we observe small but non-zero linear and cross-talk biases in the complete sample, confirming the deviation from $\epsilon^{\text{obs}} \approx \epsilon^s + \gamma$ in high-shear regimes. Notably, about one-quarter of the final cross-talk bias is already present in the complete sample, cautioning against over-interpreting biases at the 10^{-4} level when the shear signal is no longer small.

To quantify the impact of these non-linear effects on the linear shear bias calibration, we compare the multiplicative biases, \tilde{m}_i , between the non-linear fit and the linear fit, where the higher-order and cross-talk terms in Eq. (A.1) are set to zero. Figure A.3 shows the difference in \tilde{m}_i as a function of the maximum shear amplitude used in the fit. Consistent with the results

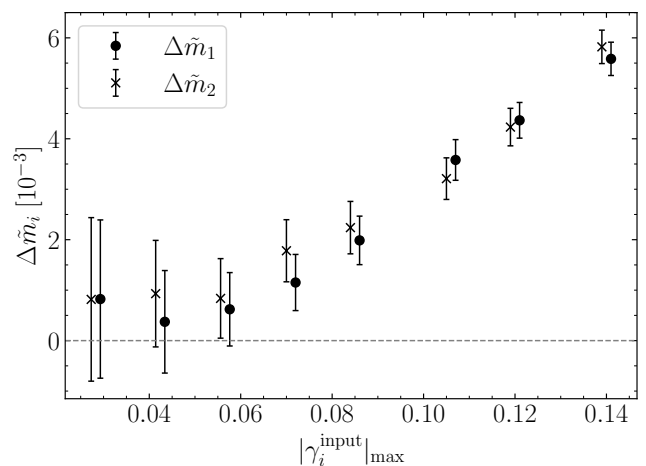


Fig. A.3. The difference in \tilde{m}_i between the non-linear and linear fits as a function of the maximum shear amplitude used in the fit. The difference is defined as $\Delta \tilde{m}_i \equiv \tilde{m}_i^{\text{non-linear}} - \tilde{m}_i^{\text{linear}}$. Error bars are calculated from bootstrapping the independent shear estimates used in each fit.

of Fig. A.2, we observe that the difference in \tilde{m}_i between the two fits increases as more high-shear amplitude simulations are included in the fit. For current KiDS-like analyses, where percent-level accuracy is required, the biases introduced by neglecting these non-linear shear effects are negligible, even in the most extreme cases from our setups. However, these effects become significant for future weak lensing surveys aiming for sub-percent level accuracy, particularly in the presence of high-shear signals ($|\gamma_i^{\text{input}}| \gtrsim 0.1$).

We note that our results, based on KiDS *lensfit* measurements, show more linear behaviour compared to the recent findings of Jansen et al. (2024), who used the KSB algorithm (Kaiser et al. 1995; Luppino & Kaiser 1997; Hoekstra et al. 1998) implemented in GalSim and observed significant non-linear effects at $|\gamma_i| \gtrsim 0.05$. However, it is important to recognise that different interpretations of the KSB algorithm can result in subtle variations across various implementations (e.g. Heymans et al. 2006). Therefore, the more pronounced non-linear effects reported by Jansen et al. (2024) only reflect the performance of the specific KSB implementation in GalSim, rather than the general KSB method. In contrast, Hoekstra et al. (in prep.) find much weaker non-linear effects with a different KSB implementation.

Table A.1. Shear biases constrained from Eq. (A.1) for different samples.

Sample	$\tilde{m}_i [10^{-3}]$	$\tilde{c}_i [10^{-5}]$	$\tilde{d}_i [10^{-3}]$	$\tilde{q}_i [10^{-1}]$	$\tilde{m}_{\perp,i} [10^{-4}]$
$i = 1$					
Complete	0.042 ± 0.000	0.000 ± 0.002	0.000 ± 0.002	0.000 ± 0.000	0.792 ± 0.002
SExtractor (20, 24.5)	-23.22 ± 0.069	-1.314 ± 0.250	-0.497 ± 0.451	0.262 ± 0.058	0.919 ± 0.379
<i>lensfit</i> selection	-13.63 ± 0.142	-0.568 ± 0.582	-2.514 ± 0.871	-0.041 ± 0.116	4.289 ± 0.693
<i>lensfit</i> measurement	0.808 ± 0.140	0.040 ± 0.578	-6.505 ± 0.936	-4.388 ± 0.119	2.431 ± 0.741
$i = 2$					
Complete	-0.042 ± 0.000	0.000 ± 0.002	0.000 ± 0.002	0.002 ± 0.000	0.795 ± 0.002
SExtractor (20, 24.5)	-23.32 ± 0.083	1.903 ± 0.305	1.216 ± 0.527	0.148 ± 0.070	2.853 ± 0.437
<i>lensfit</i> selection	-12.56 ± 0.173	-3.208 ± 0.644	-3.270 ± 1.006	-0.478 ± 0.137	1.459 ± 0.858
<i>lensfit</i> measurement	1.686 ± 0.179	-1.661 ± 0.698	-3.147 ± 1.044	-4.609 ± 0.139	3.234 ± 0.876

Notes. The first three samples are based on a perfect galaxy shape measurement of Eq. (A.2) but with different sample selections. The label ‘Complete’ refers to the full sample with perfect shape noise cancellation, ‘SExtractor (20, 24.5)’ indicates the sample detected by SExtractor with observed magnitudes in the range of (20, 24.5), and ‘*lensfit* selection’ denotes samples further selected by *lensfit*. The ‘*lensfit* measurement’ sample is the usual sample with galaxy shapes measured by *lensfit*.