



Universiteit  
Leiden  
The Netherlands

## The galaxy-dark matter connection: a KiDS study

Dvornik, A.

### Citation

Dvornik, A. (2019, November 13). *The galaxy-dark matter connection: a KiDS study*. Retrieved from <https://hdl.handle.net/1887/80327>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/80327>

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

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/80327> holds various files of this Leiden University dissertation.

**Author:** Dvornik, A.

**Title:** The galaxy–dark matter connection: a KiDS study

**Issue Date:** 2019-11-13

The Galaxy–Dark Matter Connection:  
A KiDS Study

*ISBN: 978-94-6323-865-6*

# The Galaxy–Dark Matter Connection: A KiDS Study

Proefschrift

ter verkrijging van  
de graad van Doctor aan de Universiteit Leiden,  
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,  
volgens besluit van het College voor Promoties  
te verdedigen op woensdag 13 november 2019  
klokke 13:45 uur

door

Andrej Dvornik

geboren te Maribor, Slovenië  
in 1990

Promotores: Prof. dr. Koenraad Kuijken  
Prof. dr. Henk Hoekstra

Promotiecommissie: Prof. dr. Huub Röttgering  
Prof. dr. Joop Schaye  
Prof. dr. Catherine Heymans (University of Edinburgh)  
Dr. Alexie Leauthaud (University of California, Santa Cruz)  
Dr. Rychard Bouwens

Cover: graphics designed by Gregor Dvornik

*Staršem*





---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Our view of the Universe . . . . .	1
1.2	Gravitational lensing . . . . .	5
1.3	The halo model . . . . .	8
1.4	The Kilo-Degree Survey . . . . .	11
1.5	This thesis . . . . .	12
<b>2</b>	<b>A KiDS weak lensing analysis of assembly bias in GAMA galaxy groups</b>	<b>15</b>
2.1	Introduction . . . . .	16
2.2	Weak galaxy-galaxy lensing theory . . . . .	18
2.3	Data and sample selection . . . . .	19
2.3.1	Lens galaxy selection . . . . .	19
2.3.2	Measurement of the ESD profile . . . . .	22
2.3.3	Covariance matrix estimation . . . . .	24
2.4	Halo model . . . . .	26
2.4.1	Model specifics . . . . .	26
2.4.2	Fitting procedure . . . . .	29
2.5	Results . . . . .	29
2.6	Discussion and conclusions . . . . .	34
2.A	Systematics tests . . . . .	36
2.A.1	Multiplicative bias . . . . .	36
2.A.2	Additive bias . . . . .	36
2.A.3	Group member contamination of the source galaxies . . . . .	37
2.A.4	Source redshift distribution . . . . .	38
2.B	Full posterior distributions . . . . .	41
<b>3</b>	<b>Unveiling Galaxy Bias via the Halo Model, KiDS and GAMA</b>	<b>43</b>
3.1	Introduction . . . . .	44
3.2	Biasing . . . . .	46
3.3	Halo model . . . . .	48
3.3.1	Halo model ingredients . . . . .	48
3.3.2	Conditional stellar mass function . . . . .	51
3.3.3	Projected functions . . . . .	53
3.4	Data and sample selection . . . . .	57
3.4.1	Lens galaxy selection . . . . .	57
3.4.2	Measurement of the $\Delta\Sigma_{\text{gm}}(r_p)$ signal . . . . .	58
3.4.3	Measurement of the $w_p(r_p)$ profile . . . . .	60
3.4.4	Covariance matrix estimation . . . . .	61
3.4.5	Fitting procedure . . . . .	61

3.5	Results . . . . .	62
3.5.1	KiDS and GAMA results . . . . .	62
3.5.2	Investigation of the possible bias in the results . . . . .	66
3.5.3	Comparison with EAGLE simulation . . . . .	66
3.6	Discussion and conclusions . . . . .	68
3.A	Analytical covariance matrix . . . . .	70
3.B	Full posterior distributions . . . . .	74
3.C	Relation between the lensing signal and the galaxy-matter cross-correlation function . . . . .	76
3.C.1	Derivation . . . . .	76
3.C.2	Relation to previous work . . . . .	79
<b>4</b>	<b>The case for two-dimensional galaxy-galaxy lensing</b>	<b>83</b>
4.1	Introduction . . . . .	84
4.2	2D galaxy-galaxy lensing formalism . . . . .	86
4.3	Lens model . . . . .	87
4.4	Proof of concept . . . . .	89
4.5	Evaluation of the two methods with the EAGLE simulation . . . . .	94
4.6	Discussion and conclusions . . . . .	98
<b>5</b>	<b>KiDS+GAMA: Inferring satellite halo masses using 2D shear maps</b>	<b>101</b>
5.1	Introduction . . . . .	102
5.2	Data and sample selection . . . . .	103
5.3	2D galaxy-galaxy lensing formalism . . . . .	104
5.4	Lens model . . . . .	106
5.5	Results . . . . .	108
5.6	Discussion and conclusions . . . . .	113
<b>6</b>	<b>The fourth data release of the Kilo-Degree Survey</b>	<b>115</b>
6.1	Introduction: the Kilo-Degree and VIKING Surveys . . . . .	116
6.2	The fourth KiDS data release . . . . .	120
6.3	Data processing . . . . .	122
6.3.1	Changes to the ASTRO-WISE image processing pipeline . . . . .	122
6.3.2	Changes to the THELI pipeline . . . . .	130
6.4	Single-band $u$ , $g$ , $r$ and $i$ catalogues, images and masks . . . . .	132
6.5	The joint KiDS-VIKING nine-band catalogue . . . . .	132
6.6	Data quality . . . . .	138
6.6.1	Image quality . . . . .	138
6.6.2	Astrometry . . . . .	139
6.6.3	Photometry . . . . .	140
6.6.4	Photometric redshifts . . . . .	144
6.6.5	Photometric depth and homogeneity . . . . .	146

6.7	Data access . . . . .	149
6.7.1	ESO science archive . . . . .	149
6.7.2	ASTRO-WISE archive . . . . .	149
6.7.3	KiDS DR4 website . . . . .	150
6.8	Summary and outlook . . . . .	150
<b>7</b>	<b>Samenvatting</b>	<b>153</b>
7.1	De connectie tussen sterrenstelsels en donkere materie . . . . .	153
7.2	Zwaartekrachtlenwerking . . . . .	156
7.3	Dit proefschrift . . . . .	156
<b>8</b>	<b>Povzetek</b>	<b>159</b>
8.1	Povezava med galaksijami in temno snovjo . . . . .	159
8.2	Gravitacijsko lečenje . . . . .	161
8.3	Ta disertacija . . . . .	162
	<b>Bibliography</b>	<b>164</b>
	<b>Publications</b>	<b>175</b>
	<b>Curriculum Vitae</b>	<b>177</b>
	<b>Acknowledgements</b>	<b>179</b>




# 1

---

## Introduction

---

### 1.1 OUR VIEW OF THE UNIVERSE

UMANS have always wondered about the origin and history of the Universe. For millennia we have thought that the Earth is in the centre of the Universe – a fact that was further ingrained in our minds due to Claudius Ptolemy. This world view started to change around 500 years ago with the works of Nicolaus Copernicus and Galileo Galilei who showed that actually, it was the Earth that moved around the Sun, together with the other planets, thus giving rise to the heliocentric system. With the publication of Isaac Newton's *Principia Mathematica* in 1687 the motion of the planets could be explained by gravitational interaction between them. Later on, it also became clear that the Sun is just one of many stars in the sky, all of which constitute the Galaxy. The beginning of the 20th century brought us images and studies of other faint objects in the sky, called nebulae. Nowadays, we know that these nebulae imaged by Hubble (1926) are galaxies in their own right, much like our own Milky Way. With the same nebulae, Lemaître (1927) and Hubble (1929) figured out that they are moving faster away from us, the further away they are. This again changed our perspective of the Universe: it is expanding and it is not static as it was thought before. If the Universe was to be expanding, that also means it had to have a start. The term for this point in time was coined – the Big Bang. The theory of the Big Bang is nowadays well established observationally (even though it still has problems in certain theoretical aspects) and we know to a great precision when the Universe began: 13.8 billion years ago, thanks to the recent observations of the cosmic microwave background by WMAP (Hinshaw et al. 2013) and Planck (Planck Collaboration et al. 2018) satellites. Another shift in understanding the Universe began to happen around the same time as the expansion of the Universe was first measured. In 1933 Fritz Zwicky realised that the mass resulting from the sum of the masses of individual galaxies in the Coma cluster was not sufficient to explain the mass of the cluster inferred from a dynamical study of these galaxies (Zwicky 1933). Later on, Vera Rubin (amongst others, for a review and timeline see Bertone & Hooper 2018) found that the visible light in galaxies could not explain the high stellar velocities (Rubin et al. 1980). Both studies provided evidence for the presence of invisible mass

(nowadays we use the term *dark matter*) or an incomplete understanding of gravity on those scales. Later on, evidence for *dark energy* began to emerge, most clearly in the expansion rate of distant galaxies as measured from Supernovae. If the Universe is composed only of matter (be it visible or dark), then this will cause the recession velocities of the galaxies to be slowed down, due to the attractive effect of gravity. It came as a surprise when two teams (Riess et al. 1998; Perlmutter et al. 1999), who looked at the fluxes and distances to the type Ia Supernovae, found that they are in fact accelerating. The observed acceleration might be caused by the cosmological constant or more generally the dark energy, an unknown energy component of the Universe. All these discoveries contributed to the emergence of a concordance model of the Universe that can successfully explain all the observed astrophysical phenomena and can provide accurate theoretical predictions. This model is called  $\Lambda$ CDM, reflecting the dominating contents of the Universe:  $\Lambda$  – the cosmological constant, one possible origin of dark energy, and CDM – cold dark matter, contributing to around 70% and 30% of the Universe energy budget, respectively.

In this thesis, the research focuses on the properties of dark matter and dark matter haloes and how they connect with the galaxies we can observe in the Universe. Because of the still unknown nature of dark matter, we tend to study it using the properties of its distribution and its properties on galactic scales and beyond. The galaxy–dark matter connection is important for three main reasons, and understanding it helps with answering the largest questions in astrophysics and cosmology today. First question includes the understanding of the physics of galaxy formation. Secondly, the inference of cosmological parameters – if we want to robustly measure the cosmological parameters, we have to understand, how the galaxies interplay with the dark matter, and thirdly, the inference of evolution of matter distribution and properties of dark matter (Wechsler & Tinker 2018).

In the standard cold dark matter and cosmological constant dominated ( $\Lambda$ CDM) cosmological framework, structure formation in the Universe is mainly driven by the dynamics of cold dark matter. The gravitational collapse of dark matter density fluctuations and their subsequent virialization leads to the formation of dark matter haloes from the highest density peaks in the initial Gaussian random density field. It is widely accepted that every galaxy resides in a parent dark matter halo. The galaxies that reside in the bottom of the dark matter halo’s potential well are referred to as central galaxies and galaxies that are orbiting the central galaxy within the halo are referred to as satellite galaxies. The exact way galaxies populate the dark matter haloes is still a topic of active research to which this thesis contributed some insights. Generally, the connection between the population of galaxies and the population of dark matter haloes is modelled statistically using an extended Press-Schechter formalism (Press & Schechter 1974). This formalism postulates that massive galaxies form in the highest density peaks of the underlying dark matter distribution and that the connection can be linked with a quantity called galaxy bias. While this formalism predicts the correct number of haloes as a function of mass and also predicts the clustering of galaxies we observe to high accuracy, it has some shortcomings. More explicitly, the galaxy bias itself can have non-trivial mass and scale dependence and is generally accepted to not be linear or deterministic. Moreover, the assembly his-

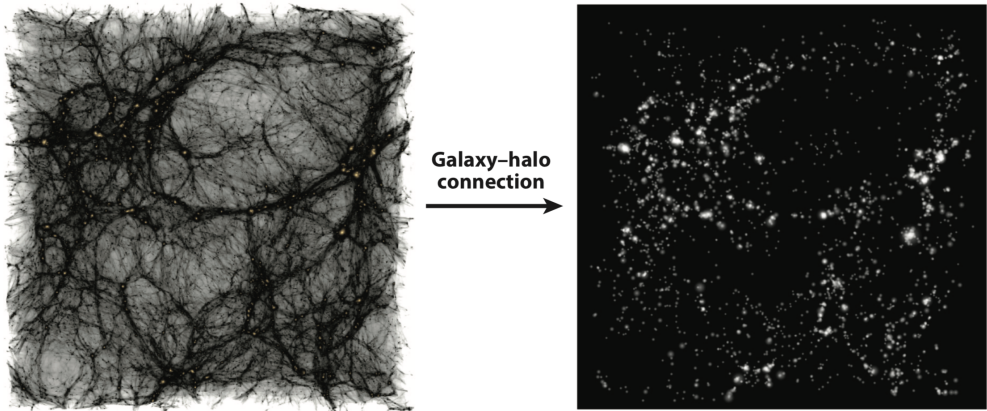


Figure 1.1: *Left panel:* a slice of a cosmological simulation showing the dark matter distribution. *Right panel:* The distribution of galaxies in the same simulation. The dark matter–galaxy connection tries to describe and study the details of dark matter distribution using the statistical properties of galaxies. Picture credit: Wechsler & Tinker (2018).

tory of dark matter haloes will leave signatures in the observed distributions that will no longer agree with the theoretical predictions. The assembly history, and mechanisms of satellite mass stripping and mergers also leave us with different properties of the dark matter connection for central and satellite galaxies. Because of all this, a large variety of different models exist all building on the statistical postulates of Press-Schechter formalism. Figure 1.1 shows a sketch of the galaxy–dark matter connection.

A popular and successful way to describe the galaxy–dark matter connection is through the halo occupation distributions (HOD), which specify the probability distributions for the number of galaxies with a certain property (luminosity or stellar mass) in a halo, given as a function of halo mass. The halo occupation distributions are quantified separately for the central galaxies and the satellite galaxies, due to their fundamental observational differences. Under these assumptions, the standard HOD is thus fully characterised by its mean occupation number of galaxies residing in a halo of a mass  $M$ . In principle, the HOD can be a function of properties other than halo mass, which can help us link the galaxies with the assembly history of dark matter haloes (Wechsler & Tinker 2018).

The HOD models can be further extended to better resemble galaxy observations and populations. The conditional luminosity (CLF) and conditional stellar mass functions (CSMF) describe the full distribution of galaxy stellar masses and luminosities as a function of the halo mass. They are, as well, usually separated into contributions from central galaxies and satellite galaxies and can be directly measured on a sample of galaxy groups and clusters (van den Bosch et al. 2013; Cacciato et al. 2013).

Both the CLF and the HOD models specify the number of galaxies per halo and the model predictions can be made in two ways. Most straightforwardly, one can populate the dark matter halos from an N-body simulation using a Monte Carlo ap-

proach and then measure the properties of galaxies from the assembled catalogue. Alternatively, both CLF and HOD can also be combined with an analytic halo model to predict the observables in a semi-analytic way (Seljak 2000; Cooray & Sheth 2002). The halo model approach assumes that all the matter in the Universe is in halos that can be thought of as gravitationally bound objects of matter that have decoupled from the expansion of the Universe and spherically collapsed, with mass  $M$  contained in a radius where the mean density is 200 times larger than the mean density of the Universe. The abundance of the dark matter haloes can then be characterised by the halo mass function, which yields the number of haloes given a mass  $M$ . If we take the results from the N-body simulations where it was found that the density profile follows a universal mass function (Navarro et al. 1997) and combine them with the halo mass function and the initial matter power spectrum, as well as the HOD/CLF models, we are able to predict a plethora of observables, through which we can then study the galaxy–dark matter connection in greater detail.

Under the assumption that the galaxy and the dark matter halo properties are closely connected, the most constraining observational measurement for any model is the abundance of galaxies. The model of the galaxy–dark matter connection, given the cosmological model, should be able to predict the abundance of galaxies as a function of their stellar mass or luminosity. Even though this observational property is the most constraining, it does not account for all the properties of the galaxy–dark matter connection and can lead to wrong interpretations. To overcome this, one can also use other probes (together with the galaxy abundance) to get a better perspective of the galaxy–dark matter connection. The next measurement one can make use of is the two-point galaxy clustering. As the abundance of dark matter haloes is strongly connected with their clustering properties, the stellar to halo mass relation will also predict the clustering properties of the galaxies residing in those haloes. The two-point galaxy clustering together with the galaxy abundance will thus fully characterise any model one wants to use to describe the connection between the dark matter and galaxies, and thus learn about the nature of dark matter.

With the halo model, one can also obtain predictions for the cross-correlation between galaxies and dark matter (the two-point galaxy clustering describes the correlation between the galaxies). The halo model predicts the galaxy-mass correlation function  $\xi_{\text{gm}}$ , related to the excess surface mass density  $\Delta\Sigma$ , which to the first order measures the projected dark matter halo profiles, that can in turn be directly measured using galaxy-galaxy lensing. Galaxy-galaxy lensing is the measurement of the gravitational lensing effect around single galaxies and the strength of it is directly proportional to the mass of the dark matter haloes around those galaxies. We will further discuss the gravitational lensing in the next Section. The galaxy-galaxy lensing signal is in most cases quite complicated to interpret as central and satellite galaxies reside in completely different haloes, thus a full model of the galaxy–dark matter connection is needed. Nevertheless, the galaxy-galaxy lensing measurements provide a complementary view to the measurements of abundances and two-point clustering of galaxies. Most importantly, they directly provide the mass estimates of the dark matter haloes and are also used to break the remaining degeneracies in the model, namely the galaxy bias.



## 1.2 GRAVITATIONAL LENSING

Einstein's hundred year old theory of General Relativity (Einstein 1916) describes gravity as a curvature of space-time around a massive object. As light travels along a straight path through flat space-time, whenever it passes through a curved space-time the path of a light ray will change. This means, that the light traveling from distant parts of the Universe, can be affected by the distribution of mass on its way. The relativistic description can be simplified to form a theory that can be completely developed in the the Newtonian framework. Because the effect is analogous to optical lensing, this effect is known as *gravitational lensing*.

Gravitational lensing can be used to probe the matter distribution of massive objects in the Universe. In the following few paragraphs we will follow the derivation of Bartelmann & Schneider (2001), a standard text known to everyone who studies the gravitational lensing, presenting the basics of this theory.

Figure 1.2 shows us a typical lensing system configuration. A point mass is positioned at angular diameter distance  $D_L$  (or redshift  $z_L$ ), which deflects light coming from a source at distance  $D_S$ . Angular diameter distance is defined as  $D = x/\theta$ , where  $x$  is the physical size of the object and  $\theta$  the angular size as viewed from Earth. The distance  $D_L$  is usually obtained from redshifts measured using galaxy spectra but  $D_S$  is obtained as an average over photometric redshift (photo- $z$ ) distances. The first approximation that we take into consideration here is that the size of the lensing object is very small compared to the distances  $D_S$ ,  $D_L$  and  $D_{LS}$ . We can then speak about a thin lens approximation, which gives us a description of the system quite similar to geometrical optics. The apparent position of the image of the source object on the sky can thus be described by the deflection angle  $\hat{\alpha}$ . Using this, we can write down the lens equation:

$$\vec{\beta} = \vec{\theta} - \frac{D_{LS}}{D_S} \hat{\alpha} \equiv \vec{\theta} - \vec{\alpha}, \quad (1.1)$$

where  $\vec{\theta}$  is an apparent direction of ray's arrival and  $\hat{\alpha}$  is the scaled deflection angle. The scaled deflection angle depends on the mass  $M$  of the lensing object and the impact vector  $\vec{\xi}$  of the light ray. At linear order (when  $\xi$  is large compared to Schwarzschild radius  $R_s = 2GM/c^2$ ), we can write the deflection angle as:

$$\vec{\alpha} = \frac{4GM}{c^2 \xi}, \quad (1.2)$$

where  $G$  is the gravitational constant and  $c$  the speed of light. If one has an extended lens, more appropriate for the case of galaxies, the deflection angle can be written as:

$$\hat{\alpha} = \frac{4G}{c^2} \int d^2\xi' \int dr'_3 \rho(\xi'_1, \xi'_2, r'_3) \frac{\vec{\xi} - \vec{\xi}'}{|\vec{\xi} - \vec{\xi}'|^2}, \quad (1.3)$$

where  $(\xi_1, \xi_2, r_3)$  describes a trajectory in space. Since the last factor in Equation 1.3 is independent of  $r'_3$ , the integration can be carried out by defining the surface mass density:

$$\Sigma(\vec{\xi}) = \int dr_3 \rho(\xi_1, \xi_2, r_3), \quad (1.4)$$

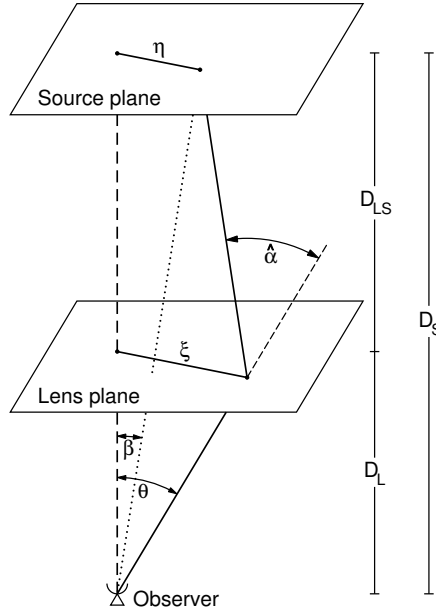


Figure 1.2: Sketch of a typical gravitational lens system. A light ray travels along from the source plane at the original angle  $\beta$  and is deflected at the lens plane by the angle  $\hat{\alpha}$ . All the distances, denoted  $D_i$  are angular diameter distances. Figure from Bartelmann & Schneider (2001).

which is the mass density projected onto a plane perpendicular to the incoming light ray.

After that, the deflection angle becomes:

$$\hat{\alpha} = \frac{4G}{c^2} \int d^2\xi' \Sigma(\vec{\xi}') \frac{\vec{\xi} - \vec{\xi}'}{|\vec{\xi} - \vec{\xi}'|^2}, \quad (1.5)$$

where  $\vec{\xi} = D_L \vec{\theta}$ . To quantify the strength of the deflection, one typically defines the dimensionless surface mass density,  $\kappa(\vec{\theta})$ , as:

$$\kappa(\vec{\theta}) \equiv \frac{\Sigma(D_L \vec{\theta})}{\Sigma_{\text{cr}}}, \quad (1.6)$$

where

$$\Sigma_{\text{cr}} \equiv \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}}. \quad (1.7)$$

$\Sigma_{\text{cr}}$  is called the critical surface density and also discriminates between the two different lensing regimes – strong and weak. The weak lensing regime is defined as the region where  $\Sigma \ll \Sigma_{\text{cr}}$  and the converse is true for the strong lensing regime – the strong lensing regime causes the source galaxies to be heavily distorted into arcs and rings, causing multiple images of the same object. Using these definitions we can finally write our scaled deflection angle as:

$$\vec{\alpha}(\vec{\theta}) = \frac{1}{\pi} \int_R d^2\theta' \kappa(\vec{\theta}') \frac{\vec{\theta} - \vec{\theta}'}{|\vec{\theta} - \vec{\theta}'|^2}. \quad (1.8)$$

This also suggests that we can write the deflection angle as a gradient of the deflection potential  $\alpha = \nabla\Psi$ , where  $\Psi$  is given the following form:

$$\Psi(\vec{\theta}) = \frac{1}{\pi} \int_{R^2} d^2\theta' \kappa(\vec{\theta}') \ln(|\vec{\theta} - \vec{\theta}'|), \quad (1.9)$$

and also satisfies the Poisson equation  $\kappa(\vec{\theta}) = \frac{1}{2}\nabla^2\Psi(\vec{\theta})$ . The surface brightness of a lensed object  $f^{\text{obs}}(\vec{\theta})$  can be related to the surface brightness of the unlensed object  $f^{\text{s}}(\vec{\theta})$  by the following mapping:

$$f^{\text{obs}}(\vec{\theta}) = f^{\text{s}}(\mathcal{A}\vec{\theta}). \quad (1.10)$$

The distortion matrix  $\mathcal{A}$  can be written as:

$$\mathcal{A} = \begin{bmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{bmatrix}, \quad (1.11)$$

where we introduced the complex shear  $\gamma \equiv \gamma_1 + i\gamma_2$ , which is related to the deflection potential through

$$\gamma_1 = \frac{1}{2} \left( \frac{\partial^2\Psi}{\partial x_1^2} - \frac{\partial^2\Psi}{\partial x_2^2} \right) \quad \text{and} \quad \gamma_2 = \frac{\partial^2\Psi}{\partial x_1\partial x_2}. \quad (1.12)$$

The observed effect of gravitational lensing on an image of a background galaxy is to magnify and tidally stretch the original shape. The tidal stretching of the images is directly proportional to the amount of mass present between such a galaxy and us as observers and it can be used to measure the masses of dark matter haloes using the galaxy-galaxy lensing (e.g. Leauthaud et al. 2011; van Uitert et al. 2011; Velander et al. 2014; Cacciato et al. 2014; Viola et al. 2015). Gravitational lensing can also be used to study the nature of the Universe with the lensing by the large scale structure itself, called cosmic shear (Bartelmann & Schneider 2001; Hildebrandt et al. 2017).

If we want to measure the galaxy-galaxy lensing signal and thus the mass estimates of the dark matter haloes, we can use the azimuthally averaged tangential shear. For any mass distribution, it measures the contrast in surface density

$$\langle\gamma_T\rangle(R) = \frac{\bar{\Sigma}(<R) - \bar{\Sigma}(R)}{\Sigma_{\text{cr}}} = \frac{\Delta\Sigma(R)}{\Sigma_{\text{cr}}}, \quad (1.13)$$

where we have defined the excess surface density  $\Delta\Sigma(R)$ , where  $R$  is the 2D projected radius with tangential shear defined as

$$\gamma_T = -[\gamma_1 \cos(2\phi) + \gamma_2 \sin(2\phi)]. \quad (1.14)$$

Here  $\phi$  is the azimuthal angle around the lens.

Excess surface density can also be computed from the halo model as:

$$\Delta\Sigma(R) = \frac{2}{R^2} \int_0^R \Sigma(R') R' dR' - \Sigma(R), \quad (1.15)$$

where the  $\Sigma(R)$  is the projected surface density, which is by definition related to galaxy-matter correlation function,  $\xi_{\text{gm}}$ , projected along the line of sight. In the distant observer approximation it takes the form of an Abel transform:

$$\Sigma(R) = \bar{\rho}_m \int_R^{R_{\text{max}}} \xi_{\text{gm}}(r) \frac{r dr}{\sqrt{r^2 - R^2}}. \quad (1.16)$$

### 1.3 THE HALO MODEL

In a hierarchical cosmological structure formation formalism, the dark matter particles are expected to reside in dark matter haloes. This suggests that the dark matter distribution can be described in the terms of its halo building blocks: on small scales, the density field is related to the density distribution of individual halos and on large scales, it reflects the spatial distribution of halos. Furthermore, those haloes can be observationally linked to the galaxies that reside in their centres.

To describe the clustered structure and matter distribution in the Universe, one can use the linear theory of structure formation, which describes the terms we will use in the description of the halo model. Let us consider a part of the Universe with a mean density  $\bar{\rho}$ . At any position  $\mathbf{x}$ , we can calculate a local density  $\rho(\mathbf{x})$ , which, in general, may be different than the mean density of the Universe. Using these two quantities, we can build a density contrast (fluctuation) field as:

$$\delta_m(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}. \quad (1.17)$$

Under gravity a local overdensity will grow (in linear theory) and attract more and more matter, thus forming massive structures in the Universe. The second moment of the density contrast field can be written as:

$$\xi_{\text{mm}}(\mathbf{x}) = \langle \delta_m(\mathbf{r}) \delta_m(\mathbf{r} + \mathbf{x}) \rangle, \quad (1.18)$$

where we have in this way defined a two-point correlation function  $\xi_{\text{mm}}(\mathbf{x})$ .

If we consider those fluctuations as superpositions of plane waves, they can be expressed in Fourier space as:

$$\delta_m(\mathbf{x}) = \frac{1}{(2\pi)^3} \int \delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}} d^3\mathbf{k}, \quad (1.19)$$

where  $\delta_{\mathbf{k}}$  denotes the density waves and  $\mathbf{k}$  is the wave vector, related to wavelength by  $\lambda = 2\pi/|\mathbf{k}|$ .

To quantify the amount of structure at each scale  $k$  ( $k = |\mathbf{k}|$ ), one usually defines a quantity called matter power spectrum. It is defined as the average of the squared Fourier transform of the density contrast  $\delta_{\mathbf{k}}$  over the same waves with scale  $k$ :

$$P_{\text{mm}}(k) \equiv \langle |\delta_{\mathbf{k}}|^2 \rangle_k. \quad (1.20)$$

The power spectrum is directly related to the two-point correlation function  $\xi_{\text{mm}}(\mathbf{x})$  as:

$$\xi_{\text{mm}}(\mathbf{x}) = \frac{1}{(2\pi)^3} \int P_{\text{mm}}(k, z) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3\mathbf{k}. \quad (1.21)$$

One can also write a similar density contrast for the population of galaxies using their number density, which can then be used to estimate the galaxy-galaxy two-point correlation function  $\xi_{\text{gg}}(\mathbf{x})$ , which describes the clustering properties of galaxies, similarly to how the  $\xi_{\text{mm}}(\mathbf{x})$  describes the clustering properties of all the matter. This also allows one to study the cross-correlation between the matter and galaxies, simply by considering the  $\xi_{\text{gm}}(\mathbf{x})$  two-point function. The latter will describe the distribution of matter around galaxies and can be measured using galaxy-galaxy lensing.

The halo model is built upon the statistical description of the properties of dark matter haloes (namely the average density profile, large scale bias and abundance) as well as on the statistical description of the galaxies residing in them. We assume that dark matter haloes are spherically symmetric, on average, and have density profiles

$$\rho(r|M) = M u_{\text{h}}(r|M), \quad (1.22)$$

that depends only on their mass  $M^1$ , and  $u_{\text{h}}(r|M)$  is the normalised density profile of a dark matter halo. The functional form of the power spectrum using the halo model assumption and accounting for all the possible correlations between dark matter haloes and galaxies residing in them (for detailed derivations one can consult Seljak 2000; Cooray & Sheth 2002; Mo et al. 2010; van den Bosch et al. 2013), can be summarised as

$$P_{\text{xy}}(k) = P_{\text{xy}}^{\text{1h}}(k) + P_{\text{xy}}^{\text{2h}}(k) \quad (1.23)$$

with

$$P_{\text{xy}}^{\text{1h}}(k) = \int_0^\infty \mathcal{H}_{\text{x}}(k, M) \mathcal{H}_{\text{y}}(k, M) n(M) dM, \quad (1.24)$$

for the 1-halo terms (correlations within one halo), and

$$\begin{aligned} P_{\text{xy}}^{\text{2h}}(k) &= P_{\text{lin}}(k) \int_0^\infty dM_1 \mathcal{H}_{\text{x}}(k, M_1) b_{\text{h}}(M_1) n(M_1) \\ &\quad \times \int_0^\infty dM_2 \mathcal{H}_{\text{y}}(k, M_2) b_{\text{h}}(M_2) n(M_2), \end{aligned} \quad (1.25)$$

---

<sup>1</sup>This assumption is generally used to simplify the halo model, in practice, it can depend on any other halo or galaxy quantities. Any dependence on a secondary properties is most commonly called assembly bias (Hearin et al. 2016), which is an important part of the galaxy–dark matter connection. We address it with a study in a later part of the thesis.

for the 2-halo terms (correlations between two different haloes), where  $P_{\text{lin}}(k)$  is the linear power spectrum, and  $b_{\text{h}}(M, z)$  is the halo bias function. Here we have defined

$$\mathcal{H}_{\text{m}}(k, M) = \frac{M}{\bar{\rho}_{\text{m}}} \tilde{u}_{\text{h}}(k|M), \quad (1.26)$$

$$\mathcal{H}_{\text{c}}(k, M) = \frac{\langle N_{\text{c}}|M \rangle}{\bar{n}_{\text{c}}}, \quad (1.27)$$

and

$$\mathcal{H}_{\text{s}}(k, M) = \frac{\langle N_{\text{s}}|M \rangle}{\bar{n}_{\text{s}}} \tilde{u}_{\text{s}}(k|M), \quad (1.28)$$

with  $\tilde{u}_{\text{h}}(k|M)$  and  $\tilde{u}_{\text{s}}(k|M)$  the Fourier transforms of the halo density profile and the satellite number density profile, respectively, both normalised to unity [ $\tilde{u}(k=0|M)=1$ ]. Above 'x' and 'y' are either 'c' (for central), 's' (for satellite), or 'm' (for matter), and  $n(M)$  is the halo mass function in the following form:

$$n(M) = \frac{\bar{\rho}_{\text{m}}}{M^2} \nu f(\nu) \frac{d \ln \nu}{d \ln M}, \quad (1.29)$$

with  $\nu = \delta_{\text{c}}/\sigma(M)$ , where  $\delta_{\text{c}}$  is the critical overdensity for spherical collapse at redshift  $z$ , and  $\sigma(M)$  is the mass variance. For  $f(\nu)$  one usually uses a fitting function obtained from numerical simulations. Furthermore, we assume that satellite galaxies in haloes of mass  $M$  follow a spherical number density distribution  $n_{\text{s}}(r|M) = N_{\text{s}} u_{\text{s}}(r|M)$ , where  $u_{\text{s}}(r|M)$  is the normalised density profile of satellite galaxies. Central galaxies always have  $r = 0$ . We assume that the density profile of dark matter haloes follows an NFW profile (Navarro et al. 1997). Since centrals and satellites are distributed differently, we can write the galaxy-galaxy 1-halo power spectrum as:

$$P_{\text{gg}}(k) = f_{\text{c}}^2 P_{\text{cc}}(k) + 2f_{\text{c}}f_{\text{s}}P_{\text{cs}}(k) + f_{\text{s}}^2 P_{\text{ss}}(k), \quad (1.30)$$

while the 1-halo galaxy-dark matter cross power spectrum is given by:

$$P_{\text{gm}}(k) = f_{\text{c}}P_{\text{cm}}(k) + f_{\text{s}}P_{\text{sm}}(k). \quad (1.31)$$

Here  $f_{\text{c}} = \bar{n}_{\text{c}}/\bar{n}_{\text{g}}$  and  $f_{\text{s}} = \bar{n}_{\text{s}}/\bar{n}_{\text{g}} = 1 - f_{\text{c}}$  are the central and satellite fractions, respectively, and the average number densities  $\bar{n}_{\text{g}}$ ,  $\bar{n}_{\text{c}}$  and  $\bar{n}_{\text{s}}$  follow from:

$$\bar{n}_{\text{x}} = \int_0^{\infty} \langle N_{\text{x}}|M \rangle n(M) dM, \quad (1.32)$$

and here 'x' stands for either 'g' (for galaxies), 'c' (for centrals) or 's' (for satellites).

In this formalism, the matter-matter power spectrum simply reads:

$$P_{\text{mm}}(k) = P_{\text{mm}}^{\text{1h}}(k) + P_{\text{mm}}^{\text{2h}}(k). \quad (1.33)$$

The two-point correlation functions corresponding to these power-spectra are obtained by simple Fourier transformation:

$$\xi_{\text{xy}}(r) = \frac{1}{2\pi^2} \int_0^{\infty} P_{\text{xy}}(k) \frac{\sin kr}{kr} k^2 dk. \quad (1.34)$$

The halo model can, through the two-point correlation functions  $\xi_{xy}$  predict the clustering of galaxies, cosmic shear signal and the galaxy-galaxy lensing signal. Together with the aforementioned HOD/CLF formalism, which enters the halo model through the occupational numbers  $\langle N|M \rangle$ , one can constrain the different aspects of the galaxy–dark matter connection, using a single model but with different observables. As part of this thesis, I have developed a halo model following the above formalism and the model is now part of the galaxy-galaxy lensing suite used by the KiDS collaboration<sup>2</sup>. Together with the model and following the same formalism, I have also developed a theoretical covariance matrix, that can provide the covariances for the galaxy-galaxy lensing signal, galaxy clustering, and cosmic shear and all the cross-terms between those observables. The halo model and parts of it were used in several of papers published by the KiDS collaboration and outside collaborators, including Brouwer et al. (2016, 2017); Dvornik et al. (2017, 2018); Grootes et al. (2018); Dvornik et al. (2019).

## 1.4 THE KILO-DEGREE SURVEY

To measure the above-mentioned galaxy-galaxy lensing signal and to constrain the different aspects of the galaxy–dark matter connection, we need some real observational data. High-fidelity images of the sky are invaluable for these kinds of studies, as they can provide the images of galaxies whose shapes can be measured with great precision and accuracy, and together with the distance information obtained from different images at different wavelengths one can constrain the effect of gravitational lensing of large scale structure, individual galaxies, and clusters.

The ESO VLT Survey Telescope (VST; Capaccioli & Schipani 2011) at ESO’s Paranal observatory was specifically designed for wide-field, optical imaging. Its focal plane contains a square 268-million pixel CCD mosaic camera OmegaCAM (Kuijken 2011) that covers a  $1 \text{ deg}^2$  area, and the observatory site and telescope optics provide for an image quality that is well suited for weak gravitational lensing studies. Since starting operations in October 2011, more than half of the available time on the telescope has been used for a set of three wide-area ‘Public Imaging Surveys’ for the ESO community. The Kilo-Degree Survey (KiDS; de Jong et al. 2013) is the deepest of these, and the one that exploits the best observing conditions (Kuijken et al. 2019).

KiDS was designed as a cosmology survey, to study the galaxy population out to large redshifts and in particular to measure the effect on galaxy shapes due to weak gravitational lensing by structure along the line of sight. Combining the imaging power of KiDS with the highly complete spectroscopic survey Galaxy And Mass Assembly (GAMA Driver et al. 2011), allows one to study the galaxy-galaxy lensing and galaxy clustering and thus tightly constrain the galaxy–dark matter connection using all the quantities predicted by the halo model, using a consistent set of lenses and sources for all observables.

---

<sup>2</sup>The code can be found at <https://github.com/KiDS-WL/KiDS-GGL>, currently still as a private repository.

To meet its primary science goal KiDS observes over 1350 square degrees of the sky in four bands:  $u$ ,  $g$ ,  $r$  and  $i$ . The  $r$  band data is taken in the best seeing conditions and are used to make deep images for the measurement of galaxy shapes. To provide colours for photometric redshift estimates of the same sources, the  $r$ -band data are supplemented with  $u$ -band,  $g$ -band, and  $i$ -band data (**Chapter 6**).

## 1.5 THIS THESIS

In this thesis we explore different aspects of the galaxy–dark matter connection that can be measured using gravitational lensing, more specifically, using galaxy-galaxy lensing as our primary probe. We use the halo model together with the halo occupation distributions to statistically describe the galaxy-halo connection and to constrain assembly bias in rich galaxy groups. The same theoretical framework is also used to constrain the nature of galaxy bias. Furthermore we re-evaluate the performance of the two-dimensional approach of galaxy-galaxy lensing and study the biases that might arise in this different study. Finally, we use the two-dimensional method to measure the satellite stellar-to-halo mass relation.

In **Chapter 2** we investigate possible signatures of halo assembly bias for spectroscopically selected galaxy groups from the Galaxy And Mass Assembly (GAMA) survey using weak lensing measurements from the spatially overlapping regions of the deeper, high-imaging-quality photometric Kilo-Degree Survey (KiDS). We use GAMA groups with an apparent richness larger than 4 to identify samples with comparable mean host halo masses but with a different radial distribution of satellite galaxies, which is a proxy for the formation time of the haloes. We measure the weak lensing signal for groups with a steeper than average and with a shallower than average satellite distribution and find no sign of halo assembly bias, with the bias ratio of  $0.85^{+0.37}_{-0.25}$ , which is consistent with the  $\Lambda$ CDM prediction. Our galaxy groups have typical masses of  $10^{13} M_{\odot}/h$ , naturally complementing previous studies of halo assembly bias on galaxy cluster scales.

In **Chapter 3** we measure the projected galaxy clustering and galaxy-galaxy lensing signals using the GAMA survey and KiDS survey to study galaxy bias. We use the concept of non-linear and stochastic galaxy biasing in the framework of halo occupation statistics to constrain the parameters of the halo occupation statistics and to unveil the origin of galaxy biasing. The bias function  $\Gamma_{\text{gm}}(r_{\text{p}})$  is evaluated using the analytical halo model from which the scale dependence of  $\Gamma_{\text{gm}}(r_{\text{p}})$ , and the origin of the non-linearity and stochasticity in halo occupation models can be inferred. Our observations unveil the physical reason for the non-linearity and stochasticity, further explored using hydrodynamical simulations, with the stochasticity mostly originating from the non-Poissonian behaviour of satellite galaxies in the dark matter haloes and their spatial distribution, which does not follow the spatial distribution of dark matter in the halo. The observed non-linearity is mostly due to the presence of the central galaxies, as was noted from previous theoretical work on the same topic. We also see that overall, more massive galaxies reveal a stronger scale dependence, and out to a larger radius. Our results show that a wealth of information about galaxy



bias is hidden in halo occupation models. These models should therefore be used to determine the influence of galaxy bias in cosmological studies.

In **Chapter 4** we revisit the performance and biases of the two-dimensional approach to galaxy-galaxy lensing. This method exploits the information for the actual positions and ellipticities of source galaxies, rather than using only the ensemble properties of statistically equivalent samples. We compare the performance of this method with the traditionally used one-dimensional tangential shear signal on a set of mock data that resemble the current state-of-the-art weak lensing surveys. We find that under idealised circumstances, the confidence regions of joint constraints for the amplitude and scale parameters of the NFW model in the two-dimensional analysis can be more than 3 times tighter than the one-dimensional results. Moreover, this improvement depends on the lens number density and it is larger for higher densities. We compare the method against the results from the hydrodynamical EAGLE simulation in order to test for possible biases that might arise due to missing lens galaxies, and find that the method is able to return unbiased estimates of halo masses when compared to the true properties of the EAGLE galaxies. Because of its advantage in high galaxy density areas, the method is especially suitable for studying the properties of satellite galaxies in clusters of galaxies.

In **Chapter 5** we use data from the Kilo-Degree Survey (KiDS) and the Galaxy And Mass Assembly (GAMA) surveys to simultaneously constrain the stellar-to-halo mass relations of both central and satellite galaxies of spectroscopically confirmed galaxies in galaxy groups using weak lensing. For the analysis we use the traditional one-dimensional method in the form of the stacked tangential shear measurements to determine the halo and subhalo masses of our galaxies and to constrain the stellar-to-halo mass relation, as well as a two-dimensional fit to the full shear field that uses all the available information about lens galaxies and exact source galaxies positions and ellipticities. We find that the two-dimensional method performs better than the one-dimensional method statistically. Both methods lead to similar parameters of the stellar-to-halo mass relation, which are consistent with previous results found in the literature, showing that the satellite galaxies have generally lower halo masses than the central galaxies, given the same stellar mass.

Finally in **Chapter 6** we present the fourth public data release of the Kilo Degree Survey which more than doubles the area of sky covered by data release 3, the data we were primarily using as our source of gravitational lensing measurements in the chapters above. My contribution to the paper on which this Chapter is based consisted of leading the ASTRO-WISE photometric data reduction, which was used to produce stacked images of 1006 pointings in the four bands, from which the photometry in the catalogues is obtained. The exact contribution is described in Chapter 6, specifically in the Section 6.3.1.




# 2

---

## A KiDS weak lensing analysis of assembly bias in GAMA galaxy groups

---

E investigate possible signatures of halo assembly bias for spectroscopically selected galaxy groups from the GAMA survey using weak lensing measurements from the spatially overlapping regions of the deeper, high-imaging-quality photometric KiDS survey. We use GAMA groups with an apparent richness larger than 4 to identify samples with comparable mean host halo masses but with a different radial distribution of satellite galaxies, which is a proxy for the formation time of the haloes. We measure the weak lensing signal for groups with a steeper than average and with a shallower than average satellite distribution and find no sign of halo assembly bias, with the bias ratio of  $0.85^{+0.37}_{-0.25}$ , which is consistent with the  $\Lambda$ CDM prediction. Our galaxy groups have typical masses of  $10^{13} M_{\odot}/h$ , naturally complementing previous studies of halo assembly bias on galaxy cluster scales.

A. Dvornik, M. Cacciato, K. Kuijken, M. Viola, et al.  
*MNRAS*, **Volume 468**, Issue 3, p. 3251-3265 (2017)

## 2.1 INTRODUCTION

In the standard cold dark matter and cosmological constant dominated ( $\Lambda$ CDM) cosmological framework, structure formation in the Universe is mainly driven by the dynamics of cold dark matter. The gravitational collapse of dark matter density fluctuations and their subsequent virialization leads to the formation of dark matter haloes from the highest density peaks in the initial Gaussian random density field (e.g. Mo et al. 2010, and the references therein). As dark matter haloes trace the underlying mass distribution, the halo bias (the relationship between the spatial distribution of dark matter haloes and the underlying dark matter density field) is naively expected to depend only on the halo mass, and can be used to predict the large-scale clustering of the dark matter haloes (Zentner et al. 2014; Hearin et al. 2016).

However, cosmological N-body simulations have shown that the abundance and clustering of the haloes depend on properties other than the halo mass alone. These for instance include formation time and concentration (Wechsler et al. 2006; Gao & White 2007; Dalal et al. 2008; Wang et al. 2009; Lacerna et al. 2014). The dependence of the spatial distribution of dark matter haloes on any of those properties, or on any property beside mass, it is commonly called *halo assembly bias* (Hearin et al. 2016).

Cosmological N-body simulations indicate that the origin of halo assembly bias is twofold. While for the high-mass haloes the assembly bias comes purely from the statistics of density peaks (related to the curvature of Lagrangian peaks in the initial Gaussian random density field; Dalal et al. 2008), the origin of halo assembly bias for low-mass haloes is rather a signature of cessation of mass accretion onto haloes (Wang et al. 2009; Zentner et al. 2014).

As galaxies are biased tracers of the underlying dark matter distribution, halo assembly bias, to some extent, violates the standard halo occupation models, which in most cases assume that the halo mass alone can completely describe the statistical properties of galaxies residing in such dark matter haloes at a given time (Leauthaud et al. 2011; van den Bosch et al. 2013; Cacciato et al. 2014), and are used to connect the galaxies with their parent haloes in which they are formed. The central quantity upon which halo occupation models are built, is the probability of a halo hosting a given number of galaxies, given its halo mass. Assembly bias will thus violate the mass-only assumption, and those models will introduce systematic errors when predicting the lensing signal and/or clustering measurements of galaxies, groups and clusters when split into subsamples of a different secondary observable (for instance, concentration) (Zentner et al. 2014). Because of that, there has been an increased effort in the last couple of years to accommodate models for assembly bias, by expanding them to allow for secondary properties to govern the occupational distributions (Hearin et al. 2016).

It has also been shown that assembly bias introduces a bimodality to the halo bias function – the function relating the clustering of matter with the observed clustering of haloes (i.e. one gets two functions, whose properties differ by the secondary observable) – but preserving the overall mass dependence (the more massive the halo, the larger the split and thus the assembly bias; Gao & White 2007). As halo assembly bias can be a signature of a multitude of secondary properties (formation time, con-

centration, host galaxy colour, amongst others), further study across multiple mass scales (from galaxies to galaxy clusters) using the same proxy is needed, as the mass dependence of halo assembly bias is not completely determined observationally.

Several studies have presented observational evidence of halo assembly bias. Yang et al. (2006) showed that at fixed halo mass, galaxy clustering increases with decreasing star formation rate (SFR) and that the reshuffling of observational quantities (dynamical mass and the total stellar mass) affects the clustering signal by up to 10%. Their results are in agreement with the findings from Gao et al. (2005), who used results from the Millennium simulation (Springel et al. 2005). Similar results were more recently obtained by Tinker et al. (2012) using observations of the COSMOS field. They find that the stellar mass of the star-forming galaxies, residing in galaxy groups, is a factor of 2 lower than for passive galaxies residing in halos with the same mass. Moreover, a similar trend is observed when they divide the population of galaxies by their morphology (for details see the definition therein), emphasising the significantly different clustering amplitudes of the two observed samples. On the other hand, Lin et al. (2016) investigated some of these claims on galaxy scales using SDSS DR7 data (Abazajian et al. 2009) and found no evidence for halo assembly bias, concluding that the observed differences in clustering were due to contamination from satellite galaxies.

More recently, Miyatake et al. (2016) used galaxy-galaxy lensing and clustering measurements of more than 8000 SDSS galaxy clusters with typical halo masses of  $\sim 2 \times 10^{14} M_{\odot}/h$ , found using the redMaPPer method (Rykoff et al. 2014). They divided the clusters into two subsamples according to the radial distribution of the photometrically selected satellite galaxies from the brightest cluster galaxy. They found that the halo bias of clusters of the same halo mass but with different spatial distributions of satellite galaxies, differs up to  $2.5\sigma$  in weak lensing, and up to  $4.6\sigma$  in clustering measurements. Zu et al. (2017) argue that the detection of halo assembly bias by Miyatake et al. (2016) is driven purely by projection effect, and they show that the effects is smaller and consistent with  $\Lambda$ CDM predictions.

We aim to investigate whether signatures of halo assembly bias are present in galaxy groups with typical masses of  $10^{13} M_{\odot}/h$ , using measurements of the weak gravitational lensing signal. Specifically we use *spectroscopically* selected galaxy groups from the GAMA survey (Driver et al. 2011) and measure the weak lensing signal from the spatially overlapping regions of the deeper, high imaging quality photometric KiDS survey (Kuijken et al. 2015; de Jong et al. 2015). As the GAMA survey provides us with spectroscopic information on the group membership, any potential projection effects are much more confined. In order to see if the two population of groups have the clustering properties consistent with what halo masses dictate, we need to know the halo masses of the two populations. Because of that we interpret the measured signal in the context of the halo model (Seljak 2000; Cooray & Sheth 2002; van den Bosch et al. 2013; Cacciato et al. 2014).

The outline of this paper is as follows. In Section 2.2 we describe the basics of the weak lensing theory, and we describe the data and sample selection in Section 2.3. The halo model is described in Section 2.4. In Section 2.5 we present the galaxy-galaxy lensing results. We conclude and discuss in Section 2.6. Throughout the paper we use the following cosmological parameters entering in the calculation of the distances and in the halo model (Planck Collaboration et al. 2013):  $\Omega_m = 0.315$ ,  $\Omega_\Lambda = 0.685$ ,  $\sigma_8 = 0.829$ ,  $n_s = 0.9603$  and  $\Omega_b h^2 = 0.02205$ . All the measurements presented in the paper are in comoving units.

## 2.2 WEAK GALAXY-GALAXY LENSING THEORY

Matter inhomogeneities deflect light rays of distant objects along their path. This effect is called gravitational lensing. As a consequence the images of distant objects (sources) appear to be tangentially distorted around foreground galaxies (lenses). The strength of the distortion is proportional to the amount of mass associated with the lenses and it is stronger in the proximity of the centre of the overdensity and becomes weaker at larger transverse distances (for a thorough review, see Bartelmann & Schneider 2001).

Under the assumption that source galaxies have an intrinsically random ellipticity, weak gravitational lensing then introduces a coherent tangential distortion. The typical change in ellipticity due to gravitational lensing is much smaller than the intrinsic ellipticity of the source, even in the case of clusters of galaxies, but this can be overcome by averaging the shapes of many background galaxies.

Weak gravitational lensing from a galaxy halo of a single galaxy is too weak to be detected. One therefore relies on a statistical approach in which one stacks the contributions from different lens galaxies, selected by similar observational properties (e.g. stellar masses, luminosities or in our case, the properties of the host of the satellite galaxies). Average halo properties, such as halo masses and large-scale halo biases, are then inferred from the resulting high signal-to-noise ratio measurements. This technique is commonly referred to as galaxy-galaxy lensing, and it is used as a method to measure statistical properties of dark matter halos around galaxies.

Given its statistical nature, galaxy-galaxy lensing can be considered as a measurement of the cross-correlation of galaxies and the matter density field:

$$\xi_{g,m}(|\mathbf{r}|) = \langle \delta_g(\mathbf{x}) \delta_m(\mathbf{x} + \mathbf{r}) \rangle_{\mathbf{x}}, \quad (2.1)$$

where  $\delta_g$  is the galaxy density contrast,  $\delta_m$  the matter density contrast,  $\mathbf{r}$  is the three-dimensional comoving separation and  $\mathbf{x}$  the position of the galaxy. From equation (2.1) one can obtain the projected surface mass density around galaxies which, in the distant observer approximation, takes the form of an Abel transform:

$$\Sigma(R) = 2\bar{\rho}_m \int_R^\infty \xi_{g,m}(r) \frac{r \, dr}{\sqrt{r^2 - R^2}}, \quad (2.2)$$

where  $R$  is the comoving projected separation from the galaxy,  $\bar{\rho}_m$  is the mean comoving density of the Universe and  $r$  is the 3D comoving separation.<sup>1</sup> Being sensitive to density contrasts, gravitational lensing is actually a measure of the excess surface mass density (ESD):

$$\Delta\Sigma(R) = \bar{\Sigma}(\leq R) - \Sigma(R), \quad (2.3)$$

where  $\bar{\Sigma}(\leq R)$  follows from:

$$\bar{\Sigma}(\leq R) = \frac{2}{R^2} \int_0^R \Sigma(R') R' dR'. \quad (2.4)$$

The ESD can finally be related to the tangential shear  $\gamma_t$  of background objects, which is the main lensing observable:

$$\Delta\Sigma(R) = \gamma_t(R)\Sigma_{\text{cr}}, \quad (2.5)$$

with

$$\Sigma_{\text{cr}} = \frac{c^2}{4\pi G(1+z_l)^2} \frac{D(z_s)}{D(z_l)D(z_l, z_s)}, \quad (2.6)$$

the critical surface mass density, a geometrical factor accounting for the lensing efficiency. In the above equation,  $D(z_l)$  is the angular diameter distance to the lens,  $D(z_l, z_s)$  the angular diameter distance between the lens and the source and  $D(z_s)$  the angular diameter distance to the source. In this equation  $c$  denotes the speed of light and  $G$  the gravitational constant. In this work, the distances are evaluated using spectroscopic redshifts for the lenses and photometric redshifts for the sources.

Predictions on ESD profiles can be obtained by using the halo model of structure formation (Seljak 2000; Peacock & Smith 2000; Cooray & Sheth 2002; van den Bosch et al. 2013; Mead et al. 2015) and we will base the interpretation of the measurements on this framework, which is presented in Section 2.4.

## 2.3 DATA AND SAMPLE SELECTION

### 2.3.1 LENS GALAXY SELECTION

The foreground galaxies used in this lensing analysis are taken from the Galaxy And Mass Assembly (hereafter GAMA) survey (Driver et al. 2011). GAMA is a spectroscopic survey carried out on the Anglo-Australian Telescope with the AAOmega spectrograph. Specifically, we use the information of GAMA galaxies from three equatorial regions, G9, G12 and G15 from the GAMA II data release (Liske et al. 2015). We do not use the G02 and G23 regions, due to the fact that the first one does not overlap with KiDS and the second one uses a different target selection compared to the one used in the equatorial regions. These equatorial regions encompass  $\sim 180 \text{ deg}^2$ ,

---

<sup>1</sup>Throughout the paper we assume that the averaged mass profile of haloes is spherically symmetric, since we measure the lensing signal from a stack of many different haloes with different orientations, which averages out any potential halo triaxiality.

containing 180 960 galaxies (with  $nQ > 3$ , where the  $nQ$  is a measure of redshift quality) and are highly complete down to a Petrosian  $r$ -band magnitude  $r = 19.8$ . For weak lensing measurements, we can use all the galaxies in the three equatorial regions as potential lenses.

We use the GAMA galaxy group catalogue version 7 (Robotham et al. 2011) to separate galaxies into centrals and satellites. The centrals are used as centre of the haloes in the lensing analysis, while the distribution of satellites is used to separate haloes with an early and late formation time. The group catalogue is constructed with a Friends-of-Friends (FoF) algorithm that takes into account the projected and line-of-sight separations, and has been carefully calibrated against mock catalogues (Robotham et al. 2011), which were produced using the Millennium simulation (Springel et al. 2005), populated with galaxies according to the semi-analytical model by Bower et al. (2006).

We select central galaxies residing in GAMA groups (the definition of the central galaxy used in this paper is the Brightest Cluster Galaxy<sup>2</sup> – BCG) to trace the centres of the groups. We select all groups with an apparent richness<sup>3</sup> ( $N_{\text{FoF}}$ ) larger than  $N_{\text{FoF}} = 4$ , covering a redshift range  $0.03 \leq z < 0.33$ . With this apparent richness cut we minimise the fraction of spurious groups and the redshift cut provides a more reliable group sample (above the redshift of  $z \sim 0.3$ , the linking length used in the FoF algorithm can become excessively large). This selection yields 2061 galaxy groups. If we include all the GAMA groups up to the redshift of  $z = 0.5$ , the final results do not change significantly, apart from having a higher signal-to-noise ratio in the lensing measurements, a result of having  $\sim 200$  more galaxies in that sample. We thus opt for a cleaner sample of galaxy groups, whose membership is better under control. There exists a possibility that the BCGs in the group sample are not centrals but in fact satellite galaxies. We have not studied the effect of this misidentification in this paper.

As a proxy for the halo assembly bias signatures of our galaxy groups we employ the average projected separation of satellite galaxies,  $\langle R \rangle$ , from the central. The radial distribution of satellite galaxies is connected to the halo concentration and thus with the halo formation time, as shown in simulations (Duffy et al. 2008; Bhattacharya et al. 2011). This measurement is naturally given by the FoF algorithm run on the GAMA survey.

Furthermore, we use this proxy to split our sample of central galaxies into two. We take 10 equally linearly spaced bins in  $z$  and 15 in  $N_{\text{FoF}}$  and perform a cubic spline fit for the median  $\langle R \rangle$  as a function of  $z$  and  $N_{\text{FoF}}$  (see Figure 2.1).

The spline fit gives us a limit between the central galaxies with satellites that are on average further apart from (upper half – hereafter  $\langle R \rangle^+$ ), or closer to (lower half – hereafter  $\langle R \rangle^-$ ) the BCG. The  $\langle R \rangle^+$  sample has 987 galaxy groups and the  $\langle R \rangle^-$  sample 1074 galaxy groups. This provides us, by construction, with two samples that have similar redshift, richness and stellar mass distributions, as can be seen in Figure 2.2.

<sup>2</sup>As shown in Robotham et al. (2011), the iterative centre is the most accurate tracer of the centre of group, but using BCG as a tracer is not very different from it.

<sup>3</sup> $N_{\text{FoF}}$  is defined by the number of GAMA galaxies associated with the group and it is dependent on the group selection function.



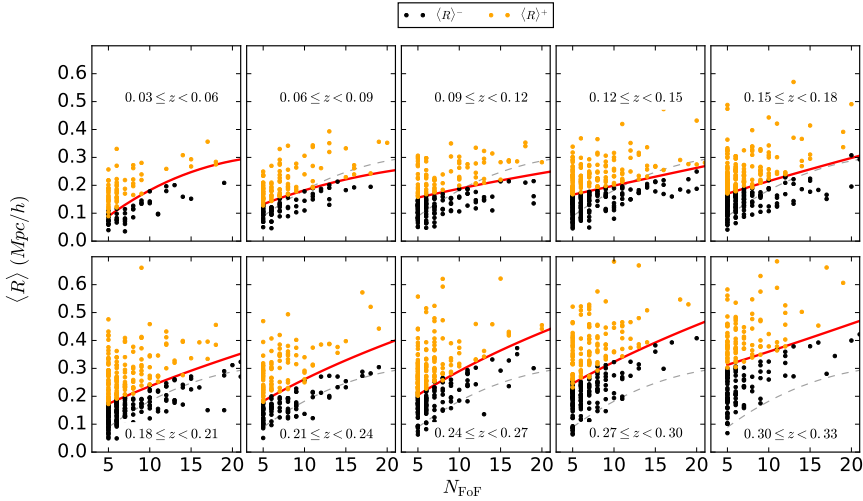


Figure 2.1: Selection of GAMA groups with apparent richness  $N_{\text{FoF}} \geq 5$  and redshift  $0.03 \leq z < 0.33$ . In each panel groups are further split by the average projected distance,  $\langle R \rangle$ , of their satellite galaxies using a spline fit for the median of  $\langle R \rangle$  (red curves). For brevity, we show only the apparent richnesses up to 20. We plot the spline fit from the first redshift bin in all other bins in grey dashed lines. They are used to guide one’s eye to see how spline changes from bin to bin.

Table 2.1: Overview of median stellar masses of central galaxies, median redshifts and number of lenses in each selected sample. Stellar masses are taken from version 16 of the stellar mass catalogue, an updated version of the catalogue created by Taylor et al. (2011).

Sample	$\log(\langle M_{\star} / [M_{\odot} h^{-1}] \rangle)$	$\langle z \rangle$	Number of lenses
<i>Full</i>	11.32	0.188	2061
$\langle R \rangle^+$	11.33	0.186	987
$\langle R \rangle^-$	11.30	0.190	1074

The median stellar masses and redshifts are listed in Table 2.1. As the dark matter haloes are located in different cosmic environments, we also want to check for the presence of apparent trends in our two samples with their environments.

Brouwer et al. (2016) presented a study of galaxies residing in different cosmic environments and they find a clear correlation of the halo bias with the cosmic environment of the haloes the galaxies are residing in. We check for the presence of apparent trends in our two samples, by comparing the distribution of the galaxies residing in voids, sheets, filaments and knots (for the exact definition of the environment classification see Eardley et al. 2015), and we do not see a large difference (see Figure 2.2). It should be noted that the classification of galaxies in Eardley et al. (2015) is only evaluated up to redshift  $z = 0.263$ , and because of that this test is only indicative.

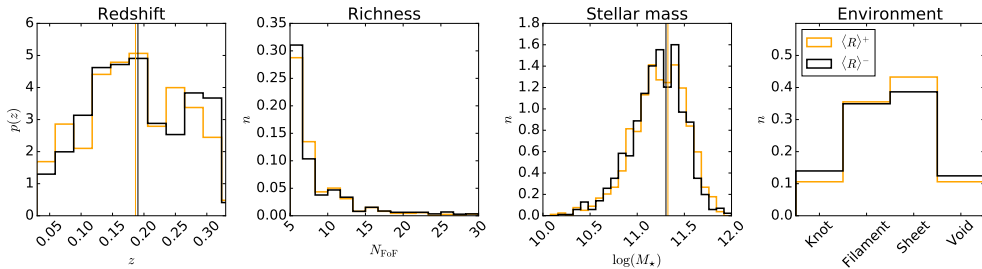


Figure 2.2: *Left panel:* Redshift distributions of the GAMA groups used in this paper for both the  $\langle R \rangle^+$  and the  $\langle R \rangle^-$  samples, shown as orange and black histograms. *Middle left panel:* Apparent richness distributions of the GAMA groups used in this paper for both the  $\langle R \rangle^+$  and the  $\langle R \rangle^-$  samples. *Middle right panel:* Stellar mass distributions of the GAMA groups used in this paper for both the  $\langle R \rangle^+$  and the  $\langle R \rangle^-$  samples. *Right panel:* Distribution of the galaxy groups in different cosmic environments. The solid orange and black vertical lines indicate the median of the redshift and stellar mass distributions for the  $\langle R \rangle^+$  and  $\langle R \rangle^-$  sample, respectively.

### 2.3.2 MEASUREMENT OF THE ESD PROFILE

We use imaging data from 180 deg<sup>2</sup> of the Kilo-Degree Survey (KiDS; Kuijken et al. 2015; de Jong et al. 2015) that overlaps with the GAMA survey (Driver et al. 2011), to obtain shape measurements of the galaxies. KiDS is a four-band imaging survey conducted with the OmegaCAM CCD mosaic camera mounted at the Cassegrain focus of the VLT Survey Telescope (VST); the camera and telescope combination provides us with a fairly uniform point spread function across the field-of-view.

From the KiDS data we use the  $r$ -band based shape measurements of galaxies, with an average seeing of 0.66 arcsec. The image reduction, photometric redshift calibration and shape measurement analysis is described in detail in Hildebrandt et al. (2017).

We measure galaxy shapes using *lensfit* (Miller et al. 2013; Fenech Conti et al. 2017, where the method calibration is described), which provides measurements of the galaxy ellipticities ( $\epsilon_1$ ,  $\epsilon_2$ ) with respect to an equatorial coordinate system. For each source-lens pair we compute the tangential  $\epsilon_t$  and cross component  $\epsilon_x$  of the source's ellipticity around the position of the lens:

$$\begin{bmatrix} \epsilon_t \\ \epsilon_x \end{bmatrix} = \begin{bmatrix} -\cos(2\phi) & -\sin(2\phi) \\ \sin(2\phi) & -\cos(2\phi) \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \quad (2.7)$$

where  $\phi$  is the angle between the  $x$ -axis and the lens-source separation vector.

The azimuthal average of the tangential ellipticity of a large number of galaxies in the same area of the sky is an unbiased estimate of the shear. On the other hand, the azimuthal average of the cross ellipticity over many sources should average to zero (Schneider 2003). Therefore, the cross ellipticity is commonly used as an estimator of possible systematics in the measurements such as non-perfect PSF deconvolution, centroid bias and pixel level detector effects.

Each lens-source pair is then assigned a weight

$$\tilde{w}_{\text{ls}} = w_s \left( \tilde{\Sigma}_{\text{cr,ls}}^{-1} \right)^2, \quad (2.8)$$

which is the product of the *lensfit* weight  $w_s$  assigned to the given source ellipticity and  $\tilde{\Sigma}_{\text{cr,ls}}^{-1}$  – the effective inverse critical surface mass density, which is a geometric term that downweights lens-source pairs that are close in redshift. We compute the effective inverse critical surface mass density for each lens using the spectroscopic redshift of the lens  $z_l$  and the full redshift probability distribution of the sources,  $n(z_s)$ , calculated using a direct calibration method presented in Hildebrandt et al. (2017). This is different from what was presented in Viola et al. (2015) and used in previous studies on KiDS DR1/2 data, where they used individual  $p(z_s)$  per source galaxy. The effective inverse critical surface density can be written as:

$$\tilde{\Sigma}_{\text{cr,ls}}^{-1} = \frac{4\pi G}{c^2} (1 + z_l)^2 D(z_l) \int_{z_l + \delta_z}^{\infty} \frac{D(z_l, z_s)}{D(z_s)} n(z_s) dz_s, \quad (2.9)$$

where  $\delta_z$  is an offset to mitigate the effects of contamination from the group galaxies (see Appendix 2.A). We determine the  $n(z_s)$  for every lens redshift separately, by selecting all galaxies in the spectroscopic sample with a  $z_s$  larger than  $z_l + \delta_z$ , with  $\delta_z = 0.2$ . The same cut is applied to the photometric redshifts  $z_s$  of the sources entering the calculation of the lensing signal. This condition was not necessary in Viola et al. (2015) as the individual  $p(z_s)$  accounted for the possible cases when the sources would be in front of the lens. Thus, the ESD can be directly computed [using equation (2.5)] in bins of projected distance  $R$  to the lenses as:

$$\Delta\Sigma(R) = \left[ \frac{\sum_{\text{ls}} \tilde{w}_{\text{ls}} \epsilon_{t,s} \Sigma'_{\text{cr,ls}}}{\sum_{\text{ls}} \tilde{w}_{\text{ls}}} \right] \frac{1}{1 + \mu}. \quad (2.10)$$

where  $\Sigma'_{\text{cr,ls}} \equiv 1/\tilde{\Sigma}_{\text{cr,ls}}^{-1}$  and the sum is over all source-lens pairs in the distance bin, and

$$\mu = \frac{\sum_i w'_i m_i}{\sum_i w'_i}, \quad (2.11)$$

is an average correction to the ESD profile that has to be applied to correct for the multiplicative bias  $m$  in the *lensfit* shear estimates. The sum goes over thin redshift slices for which  $m$  is obtained using the method presented in Fenech Conti et al. (2017), weighted by  $w' = w_s D(z_l, z_s)/D(z_s)$  for a given lens-source sample. The value of  $\mu$  is around  $-0.014$ , independent of the scale at which it is computed. Estimates of  $m$  for each redshift slice used in the calculation are presented in Figure 2.10.

It should be noted that the photometric redshift calibration and shape measurement steps differ significantly from the methods used in Viola et al. (2015) and thus we have to examine the possible systematic errors and biases. In order to do so, we devise a number of tests to see how the data behave in different observational limits, and the results are presented in Appendix 2.A. We test for the presence of additive bias as well as for the presence of cross shear over a wide range of scales.

Furthermore, we check how much the GAMA galaxy group members contaminate our source population, and what differences are introduced by the use of a global  $n(z_s)$  instead of individual  $p(z_s)$  per galaxy. We conclude that one should use comoving scales between 70 kpc/ $h$  and 10 Mpc/ $h$  (this range is motivated by the significant contamination by the GAMA group galaxies on the source population on small scales, and non-vanishing cross-term and additive biases present in the lensing signal calculated around random points on large scales), and use between 5 and 20 radial bins, depending on the choice of error estimation technique and the maximum scale, which is dictated by the number of independent regions one can use to estimate the bootstrap errors and the number of independent entries in the resulting covariance matrix (see further motivation in Section 2.3.3). Here, we use 8 radial bins between 70 kpc/ $h$  and 10 Mpc/ $h$ . For the sources we adopt the redshift range [0.1, 0.9], motivated by Hildebrandt et al. (2017).

### 2.3.3 COVARIANCE MATRIX ESTIMATION

Statistical error estimates on the lensing signal are obtained in two ways. First we follow the prescription used in Viola et al. (2015) which was shown to be valid in Sifón et al. (2015), van Uitert et al. (2016) and Brouwer et al. (2016), where we calculate the analytical covariance matrix from the contribution of each source in radial bins. This prescription accounts for shape noise of source galaxies and includes information about the survey geometry (including the masking of the lens and source galaxies). However, this method does not account for sample variance, but Viola et al. (2015) showed that this prescription works sufficiently well up to 2 Mpc/ $h$ . As we calculate the lensing signal up to 10 Mpc/ $h$ , we use the bootstrap method, as the analytical covariance tends to underestimate the errors on scales greater than 2 Mpc/ $h$  (see Figure 2.3, where we compare the different methods for estimating the errors). We first test the bootstrap method by bootstrapping the lensing signal measured around lenses in each of the 1 deg<sup>2</sup> KiDS tiles. We randomly select 180 of these tiles with replacement and stack the signals. We repeat this procedure  $10^5$  times. The covariance matrix is well constrained by the 180 KiDS tiles used in this analysis, as the number of independent entries in the covariance matrix is equal to 36.

As the physical size of the tile is comparable to the maximum separations we are considering (one degree at the median redshift of our sample corresponds to  $\sim 8$  Mpc/ $h$ ), there is a concern that the KiDS tiles might not well describe the errors on scales larger than 2 Mpc/ $h$ , because the tiles are not truly independent from each other. In fact, the sources in neighbouring tiles do contribute to the lensing signal of a group in a certain tile and the tiles are thus not independent on scales above 8 Mpc/ $h$ . We thus repeat the above exercise and calculate the bootstrapped covariance matrix using 4 deg<sup>2</sup> KiDS patches (by combining 4 adjacent KiDS tiles), which leaves us with 45 independent bootstrap regions (which is still enough to constrain the 36 independent entries in our covariance matrix). The square root of diagonal elements compared to the result of the analytical covariance can be seen in Figure 2.3 and the full bootstrap correlation matrix in Figure 2.4. For a shape noise dominated measurement one would expect that all three methods yield the same results on scales smaller

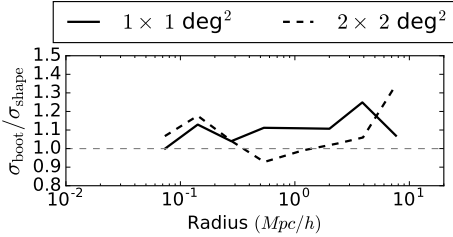


Figure 2.3: Ratios of the errors obtained using a bootstrap method and the errors obtained from the analytical covariance. Ratios for  $1 \text{ deg}^2$  KiDS tiles and  $4 \text{ deg}^2$  patches are shown in solid and dashed black lines. The errors are taken as the square root of the diagonal of the respective covariance matrices.

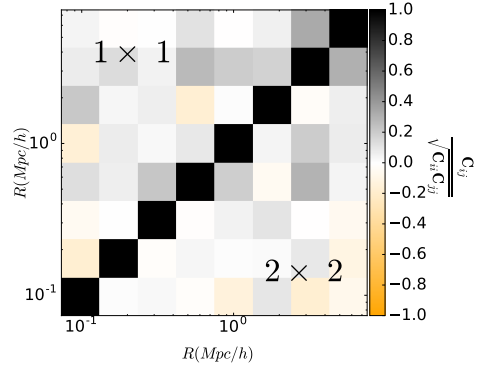


Figure 2.4: The ESD correlation matrix between different radial bins estimated using a bootstrap technique. Bootstrap covariance accounts both for shape noise and cosmic variance. In the upper triangle we show the correlation matrix when using  $1 \text{ deg}^2$  tiles, and in the lower triangle the correlation matrix when using  $4 \text{ deg}^2$  patches (as indicated).

than  $2 \text{ Mpc}/h$ . While this holds for all methods on small scales, it certainly does not hold at scales larger than  $2 \text{ Mpc}/h$  for the analytical and bootstrap covariances, when taking only  $1 \text{ deg}^2$  tiles. The main issue here is that one lacks large enough independent regions to properly sample the error distribution on large scales, and thus the resulting errors are highly biased. Taking all this considerations into account, we decide to use the bootstrapping over  $4 \text{ deg}^2$  patches as our preferred method of estimating the errors of our lensing measurements. Due to noise, the inverse covariance matrix calculated from the covariance matrix,  $\mathbf{C}_*^{-1}$ , is not an unbiased estimate of the true inverse covariance matrix  $\mathbf{C}^{-1}$  (Hartlap et al. 2007). In order to derive an unbiased estimate of the inverse covariance we need to apply a correction so that  $\mathbf{C}^{-1} = \alpha \mathbf{C}_*^{-1}$ . In the case of Gaussian errors and statistically independent data vectors, this correction factor is:

$$\alpha = \frac{n - p - 2}{n - 1}, \quad (2.12)$$

where  $n$  is the total number of independent bootstrap patches, i.e. 45 in our case, and  $p$  is the number of data points we use, i.e. in our case 8. Hartlap et al. (2007) also show that for  $p/n \lesssim 0.8$  (in our case we have  $p/n = 0.18$ ) this correction produces an unbiased estimate of the inverse covariance matrix  $\mathbf{C}^{-1}$  and we use this correction in our analysis.

When fitting the halo model to the data, we use the inverse covariance matrix from the bootstrap using  $4 \text{ deg}^2$  patches. One could use more sophisticated methods to precisely estimate the errors on very large scales. For instance, the analytical covariance method from Hildebrandt et al. (2017) can be adapted for galaxy-galaxy lensing or using galaxy-galaxy lensing specific mock catalogues to estimate the covariance matrix. Future studies using the KiDS data, expanding the analysis over greater separations or simply having more data points should employ methods like that one, but for the purposes of this study, the covariance matrix presented here is sufficient.

## 2.4 HALO MODEL

A successful analytic framework to describe the clustering of dark matter and its evolution in the Universe is the halo model (Seljak 2000; Peacock & Smith 2000; Cooray & Sheth 2002; van den Bosch et al. 2013; Mead et al. 2015). The halo model provides an ideal framework to describe the statistical weak lensing signal around a selection of galaxies. One of the assumptions of the halo model is that halo bias is only a function of halo mass, an assumption we want to test in this work. The halo model is built upon the statistical description of the properties of dark matter haloes (namely the average density profile, large scale bias and abundance) as well as on the statistical description of the galaxies residing in them.

The mass of a dark matter halo in the halo model framework is defined as:

$$M = \frac{4\pi}{3} r_{\Delta}^3 \Delta \bar{\rho}_m, \quad (2.13)$$

enclosed by the radius  $r_{\Delta}$  within which the mean density of the halo is  $\Delta$  times  $\bar{\rho}_m$ . Throughout the paper we use  $\bar{\rho}_m$  as the mean comoving matter density of the Universe ( $\bar{\rho}_m = \Omega_{m,0} \rho_{\text{crit}}$ , where  $\rho_{\text{crit}} = 3H_0^2/8\pi G$  and  $\Delta = 200$ ). We assume that the density profile of dark matter haloes follows an NFW profile (Navarro et al. 1997).

### 2.4.1 MODEL SPECIFICS

The ESD profile as defined in equation (2.3), which is related to the galaxy-matter cross-correlation function  $\xi_{g,m}(r, z)$ , can be obtained by Fourier transforming the galaxy-matter power spectrum  $P_{g,m}(k, z)$ :

$$\xi_{g,m}(r, z) = \frac{1}{2\pi^2} \int_0^{\infty} P_{g,m}(k, z) \frac{\sin kr}{kr} k^2 dk, \quad (2.14)$$

where  $k$  is the wavenumber and the subscripts m and g stand for *matter* and *galaxy*. Equation (2.14) can be expressed as a sum of a term that describes the small scales (one-halo, 1h), and one describing the large scales (two-halo, 2h) (see equation 2.15).

As we calculate the stacked ESD profile around the central galaxies of the GAMA groups, the only contribution to the one-halo term arises from central galaxies. The contribution of satellite galaxies is not modelled as it does not induce coherent distortions in our stacked measurements. As galaxies are not isolated at large scales, the signal there is dominated by the clustering of dark matter halos. This so-called two-halo term will play an important role in characterising halo assembly bias. Thus, we write the power spectrum as:

$$P_{g,m}(k, z) = P_{g,m}^{\text{1h},c}(k, z) + P_{g,m}^{\text{2h},c}(k, z), \quad (2.15)$$

where:

$$P_{g,m}^{\text{1h},c}(k, z) = \frac{1}{\bar{\rho}_m \bar{n}_g} \int dM \frac{dn(M, z)}{d \ln M} u_g(k|M) \langle N_g^c | M \rangle, \quad (2.16)$$

and  $\frac{dn(M, z)}{d \ln M}$  is the halo mass function (number density of haloes as a function of their mass),  $\langle N_g^c | M \rangle$  is an average number of central galaxies residing in a halo with given mass  $M$  and the  $u_g(k|M)$  is the normalised Fourier transform of the group density profile. For the halo mass function we use the analytical function presented in Tinker et al. (2010). Furthermore we define the comoving number density of groups  $\bar{n}_g$  as:

$$\bar{n}_g = \int \langle N_g^c | M \rangle \frac{dn(M, z)}{d \ln M} \frac{dM}{M}. \quad (2.17)$$

We require that the halo mass function obeys the following normalization relation:

$$\int_0^\infty dM \frac{dn(M, z)}{d \ln M} = \bar{\rho}_m, \quad (2.18)$$

which is satisfied in the case of using the halo mass function from Tinker et al. (2010). The two-halo term can be written as:

$$P_{g,m}^{\text{2h},c}(k, z) = b P_m(k, z), \quad (2.19)$$

where  $b = A_b b_g$  and  $b_g$  is given by:

$$b_g = \frac{1}{\bar{n}_g} \int \langle N_g^c | M \rangle b_h(M, z) \frac{dn(M, z)}{d \ln M} \frac{dM}{M}, \quad (2.20)$$

where  $A_b$  is a free parameter that we fit for,  $b_h(M, z)$  is the halo bias function and  $P_m(k, z)$  is the linear matter-matter power spectrum. For the halo bias function we use the fitting function from Tinker et al. (2010), as it was obtained using the same numerical simulation from which the halo mass function was calibrated. This form of the two-halo term is motivated by the fact that the halo density contrast and matter density contrast can be related with a halo bias function that can be linearised (van den Bosch et al. 2013). The extra free parameter  $A_b$  is introduced, because any signature of halo assembly bias will break the mass-only Ansatz of the halo model precisely at this point.

We have adopted the parametrization of the concentration-mass relation, given by Duffy et al. (2008):

$$c(M, z) = f_c \times 10.14 \left[ \frac{M}{(2 \times 10^{12} M_\odot / h)} \right]^{-0.081} (1+z)^{-1.01}, \quad (2.21)$$

with a free normalisation  $f_c$ .

The halo occupation statistics of central galaxies are defined via the function  $\langle N_g^c | M \rangle$ , the average number of galaxies as a function of halo mass  $M$ . We model  $\langle N_g^c | M \rangle$  as a error function characterised by a minimum mass,  $\log[M_1 / (h^{-1} M_\odot)]$ , and a scatter  $\sigma_c$ :

$$\langle N_g^c | M \rangle = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{\log M - \log M_1}{\sigma_c} \right) \right]. \quad (2.22)$$

We caution the reader against over-interpreting the physical meaning of this parametrization. This functional form mainly serves the purpose of assigning a distribution of halo masses around a mean halo mass value.

As in Viola et al. (2015) we assume that the degree of miscentering of the groups in three dimensions is proportional to the halo scale radius  $r_s$ , a function of halo mass and redshift, and we parametrise the probability that a central galaxy is miscentered as  $p_{\text{off}}$ . This gives

$$u_g(k|M) = u_m(k|M) \left( 1 - p_{\text{off}} + p_{\text{off}} e^{[-0.5k^2(r_s \mathcal{R}_{\text{off}})^2]} \right), \quad (2.23)$$

where  $u_m(k|M)$  is the Fourier transform of the normalised dark matter density profile, which is assumed to follow an NFW profile (Navarro et al. 1997), and  $\mathcal{R}_{\text{off}}$  the typical miscentering distance.

We include the contribution of the stellar mass of the BCGs to the lensing signal as a point mass approximation, which we can write as:

$$\Delta \Sigma_{\text{pm}} = \frac{\langle M_\star \rangle}{\pi R^2}, \quad (2.24)$$

where  $\langle M_\star \rangle$  is the average stellar mass of the selected galaxies obtained directly from the GAMA catalogue. Stellar masses are taken from version 16 of the stellar mass catalogue, an updated version of the catalogue created by Taylor et al. (2011), who fitted Bruzual & Charlot (2003) synthetic stellar spectra to the broadband SDSS photometry assuming a Chabrier (2003) IMF and a Calzetti et al. (2000) dust law. This stellar mass contribution is kept fixed for all of our samples.

The free model parameters for each sample are  $\lambda = [f_c, p_{\text{off}}, \mathcal{R}_{\text{off}}, \log(M_1), \sigma_c, b]$ , and when fitting we also store the derived parameter  $\log(M_h)$  – an effective mean halo mass:

$$M_h = \frac{1}{\bar{n}_g} \int \langle N_g^c | M \rangle \frac{dn(M, z)}{d \ln M} dM, \quad (2.25)$$

which accounts for weighting of the given fitted masses by the halo mass function. We use this mean halo mass when reporting our results.



### 2.4.2 FITTING PROCEDURE

We fit this model to each of our two samples ( $\langle R \rangle^+$  and  $\langle R \rangle^-$ ) with independent parameters and covariance matrices. This gives us a total of 12 free parameters. We use a Bayesian inference method in order to obtain full posterior probabilities using a Monte Carlo Markov Chain (MCMC) technique; more specifically we use the `emcee` Python package (Foreman-Mackey et al. 2013). The likelihood  $\mathcal{L}$  is given by

$$\mathcal{L} \propto \exp \left[ -\frac{1}{2} (\mathbf{O}_i - \mathbf{M}_i)^T \mathbf{C}_{ij}^{-1} (\mathbf{O}_j - \mathbf{M}_j) \right], \quad (2.26)$$

where  $\mathbf{O}_i$  and  $\mathbf{M}_i$  are the measurements and model predictions in radial bin  $i$ ,  $\mathbf{C}_{ij}^{-1}$  is the element of the inverse covariance matrix that accounts for the correlation between radial bins  $i$  and  $j$ . In the fitting procedure we use the inverse covariance matrix as described in Section 2.3.3. We use wide flat priors for all the parameters, and the ranges can be seen in Table 5.2. The halo model (halo mass function and the power spectrum) is evaluated at the median redshift for each sample. We run the sampler using 120 walkers, each with 2000 steps (for the combined number of 240 000 samples), out of which we discard the first 600 burn-in steps (72 000 samples). The resulting MCMC chains are well converged according to the integrated autocorrelation time test.

Figure 2.5 shows the stacked ESD profile for all 2061 galaxy groups (full sample). In comparison to Viola et al. (2015), this sample has around  $\sim 40\%$  more galaxy groups, given by the fact we are using the full equatorial KiDS and GAMA overlap. We calculate the lensing signal for all our samples according to the procedure described in Section 2.3.2. In the same figure, we also show the halo model fit to the data, as described in this section.

## 2.5 RESULTS

We fit the halo model as presented in Section 2.4.1 to the two subsamples ( $\langle R \rangle^+$  – sample with more dispersed satellite galaxies and  $\langle R \rangle^-$  – sample with more concentrated satellite galaxies). The fits have a reduced  $\chi^2_{\text{red}}$  ( $= \chi^2/\text{d.o.f}$ ) equal to 1.31 and 1.41 for the  $\langle R \rangle^+$  and  $\langle R \rangle^-$  sample, respectively, and the best fit models are presented in Figure 2.6, plotted with the 16 and 84 percentile confidence intervals. We also plot the stacked ESD profiles for both samples of galaxies, with  $1\sigma$  error bars, which are obtained by taking the square root of the diagonal elements of the bootstrap covariance matrix.

The measured parameters are summarised in Table 5.2, and their full posterior distributions are shown in Figure 2.18. The various parameters show similar results between the  $\langle R \rangle^+$  and  $\langle R \rangle^-$  subsamples. The normalisations of the concentration-halo mass relations  $f_c$  are  $f_c^+ = 1.08^{+0.99}_{-0.58}$  and  $f_c^- = 1.61^{+0.99}_{-0.53}$  for  $\langle R \rangle^+$  and  $\langle R \rangle^-$  respectively, in accordance with the results for the full sample (see Table 5.2). Furthermore the scatter in halo masses,  $\sigma_c$  is constrained to  $\sim 0.6$  for both samples and it is also consistent with the results for the full sample (see Table 5.2). We observe lower probabilities for miscentering of the central galaxy than reported in Viola et al. (2015), but with a

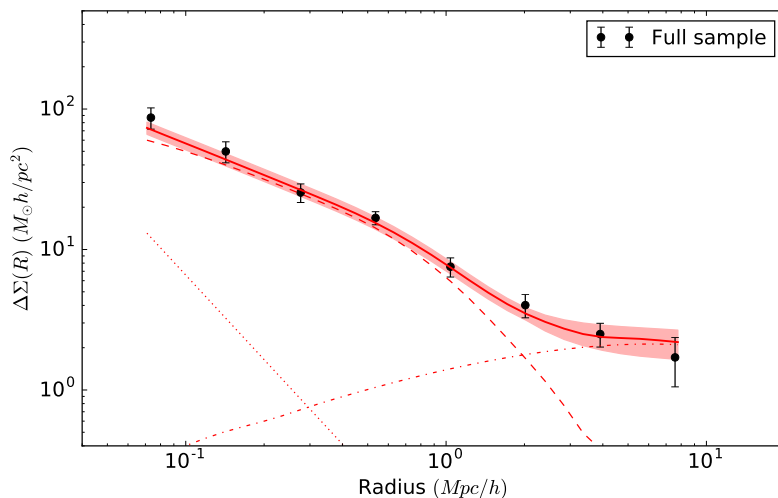


Figure 2.5: Stacked ESD profiles measured around the central galaxies of GAMA groups from the full sample of galaxies used in this study. The solid red lines represent the best-fitting halo model as obtained using a MCMC fit, with the 68% confidence interval indicated with a shaded region. Dashed, dash-dotted and dotted lines represent the one-halo term, two-halo term and stellar contribution, respectively (see Section 2.4.1).

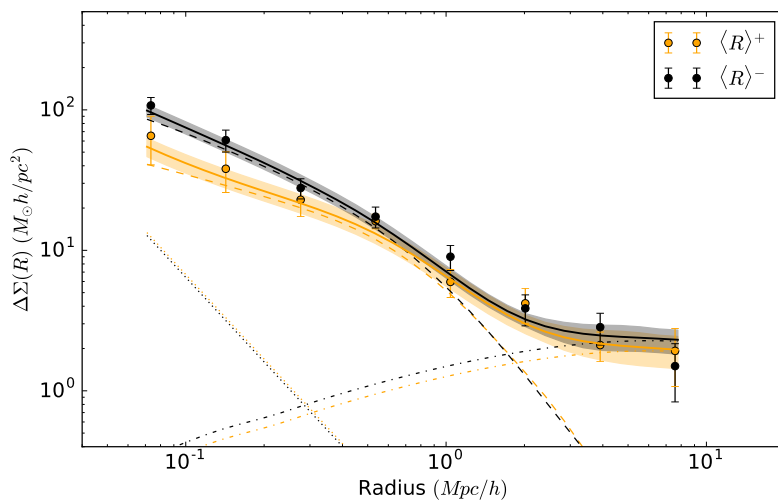


Figure 2.6: Stacked ESD profiles measured around the central galaxies of GAMA groups, selected according to the average separation of satellite galaxies (see Section 2.3.1). The solid orange and black lines represent the best-fitting halo model as obtained using a MCMC fit, with the 68% confidence interval indicated with a shaded region. Dashed, dash-dotted and dotted lines represent the one-halo term, two-halo term and stellar contribution, respectively.

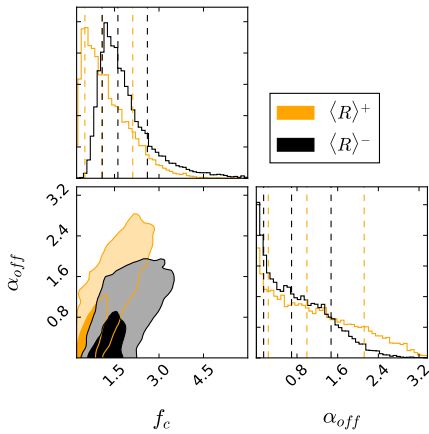


Figure 2.7: The posterior distributions of the average projected offset  $\alpha_{\text{off}}$  and the normalisation of the concentration-halo mass relation  $f_c$ . The contours indicate  $1\sigma$  and  $2\sigma$  confidence regions.

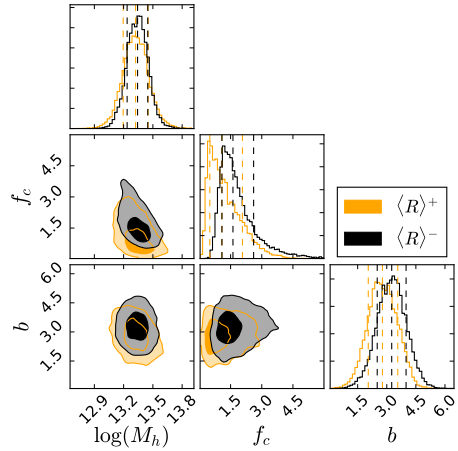


Figure 2.8: The posterior distributions of the halo model parameters  $M_h$ ,  $f_c$  and  $b$ . The posterior distributions clearly show a slight difference in the obtained halo masses as well as no difference in the obtained halo biases. The contours indicate  $1\sigma$  and  $2\sigma$  confidence regions.

larger miscentering distance. It should be noted, that the average projected offset  $\alpha_{\text{off}}$  ( $\alpha_{\text{off}} = p_{\text{off}} \times \mathcal{R}_{\text{off}}$ ) is highly degenerate with the concentration normalisation  $f_c$  and the posterior probability distribution is shown in Figure 2.7. The resulting degeneracy is similar to the one presented in Viola et al. (2015). Since we consider ESD profiles out to  $10 \text{ Mpc}/h$ , the halo masses are well constrained by the inner-most part of the same ESD profile ( $r_{200}$  associated with this mass scale is significantly smaller than  $10 \text{ Mpc}/h$ ). The contribution to the ESD profile beyond  $2 \text{ Mpc}/h$  can be associated purely with the two-halo term (see Figure 2.6). The ratio of the obtained halo biases is  $b^+/b^- = 0.85^{+0.37}_{-0.25}$ . The posterior probability distributions of the obtained halo masses and biases can be seen in Figures 2.8 and 2.18.

With the lensing measurements providing us the same halo masses for the two samples (within the errors), we report a null detection of halo assembly bias on galaxy groups scales. Our result is in accordance with what one would expect if halo bias is only a function of mass (see Figure 2.9). In Figure 2.9, we also compare our results with the biases obtained by Miyatake et al. (2016) and to the predictions for a concentration dependent halo bias from Wechsler et al. (2006). To account for the slightly different masses of our two samples one can also compare the difference arising purely from the normalisation of the bias  $A_b$  [as defined in equation (2.19)]. The ratio of obtained normalisations is still compatible with a null detection;  $A_b^+/A_b^- = 0.86^{+0.43}_{-0.28}$  ( $0.4\sigma$ ).

Table 2.2: Summary of the lensing results obtained using MCMC halo model fit to the data. All the parameters are defined in Section 2.4.1.  $f_c$  is the normalisation of the concentration-halo mass relation,  $p_{\text{off}}$  the miscentering probability,  $\mathcal{R}_{\text{off}}$  the miscentering distance,  $M_1$  central mass used to parametrise the HOD,  $\sigma_c$  scatter in HOD distribution and  $b$  bias.

Sample	$\log(M_h[M_\odot/h])$	$f_c$	$p_{\text{off}}$	$\mathcal{R}_{\text{off}}$	$\log(M_1[M_\odot/h])$	$\sigma_c$	$b$
Priors	–	[0.0, 6.0]	[0.0, 1.0]	[0.0, 3.5]	[11.0, 17.0]	[0.05, 1.5]	[0.0, 10.0]
$\langle R \rangle^+$	$13.32^{+0.13}_{-0.13}$	$1.08^{+0.99}_{-0.58}$	$0.58^{+0.27}_{-0.36}$	$2.10^{+0.99}_{-1.23}$	$13.07^{+0.19}_{-0.18}$	$0.60^{+0.05}_{-0.05}$	$2.77^{+0.78}_{-0.73}$
$\langle R \rangle^-$	$13.34^{+0.10}_{-0.11}$	$1.61^{+0.99}_{-0.53}$	$0.37^{+0.24}_{-0.23}$	$2.40^{+0.81}_{-1.50}$	$13.10^{+0.17}_{-0.16}$	$0.61^{+0.05}_{-0.05}$	$3.25^{+0.74}_{-0.74}$
Full	$13.42^{+0.09}_{-0.08}$	$1.03^{+0.63}_{-0.35}$	$0.42^{+0.21}_{-0.24}$	$2.46^{+0.73}_{-1.24}$	$13.22^{+0.14}_{-0.13}$	$0.60^{+0.05}_{-0.05}$	$3.05^{+0.72}_{-0.75}$

If the halo assembly bias due to different spatial distributions of satellite galaxies traces the halo bias due to different halo concentrations, then one would expect that the halo assembly bias would follow the predictions presented in Wechsler et al. (2006), and would also not be significant near the halo collapse mass  $M_c$ . The halo collapse masses for our two samples are  $M_c = 2.12 \times 10^{12} M_\odot/h$  and  $M_c = 2.02 \times 10^{12} M_\odot/h$  for the  $\langle R \rangle^+$  and  $\langle R \rangle^-$  subsamples, which are  $\sim 8\sigma$  below the obtained halo masses. The cancelation effect of the halo assembly bias due to the predicted sign change (clearly seen in Figure 2.9) of the concentration dependent halo bias near the  $M_c$  cannot be the cause of the null detection of halo assembly bias, as none of our lenses have halo masses that are below the  $M_c$ . We however acknowledge that the differences in predicted halo bias following Wechsler et al. (2006) for  $c'$  (as defined therein) of our two samples at the obtained halo masses are rather small (halo bias ratio of 1.06) and challenging to observe in the first place.

As the results can potentially depend on the choice of the concentration-mass relation, and to see if the choice of our fiducial Duffy et al. (2008) concentration-mass relation does not significantly influence our results, we perform a test where we change the fiducial concentration-mass relation to a parameter that is constant with mass and free to fit. The obtained concentrations for the  $\langle R \rangle^+$  and  $\langle R \rangle^-$  subsamples are  $c^+ = 5.64^{+3.64}_{-2.57}$  and  $c^- = 8.36^{+2.38}_{-2.14}$  – again highly degenerate with the average projected offset  $\alpha_{\text{off}}$ . The ratio of obtained halo biases in this case is  $b^+/b^- = 0.86^{+0.41}_{-0.28}$  and the ratio of obtained normalisations is  $A_b^+/A_b^- = 0.89^{+0.45}_{-0.31}$ . We further check if the method presented can detect a bias ratio different than unity using a sample which is known to have one. For this we split our full sample into two samples with different apparent richnesses by making a cut at  $N_{\text{FoF}} = 10$  (in order to have two samples with comparable  $S/N$ ). We fit the halo model as presented in Section 4.1 to obtain the posterior distributions of the halo biases.

As expected, the two samples have significantly different halo masses with the high richness sample having a halo mass of  $\log(M_h[M_\odot/h]) = 13.72^{+0.13}_{-0.11}$  and the low richness sample having a halo mass of  $\log(M_h[M_\odot/h]) = 13.24^{+0.09}_{-0.09}$ . The obtained halo bias ratio is, as expected, different than unity  $b^{\text{high}}/b^{\text{low}} = 2.84^{+1.75}_{-1.01}$ , which is also true when one accounts for the fact that the samples have different halo masses. In this case, the ratio of obtained normalisations is  $A_b^{\text{high}}/A_b^{\text{low}} = 2.14^{+1.42}_{-0.85}$ , which is  $1.3\sigma$  away from unity. The lensing signal and posterior distributions for this test can be seen in Figures 2.16 and 2.17.

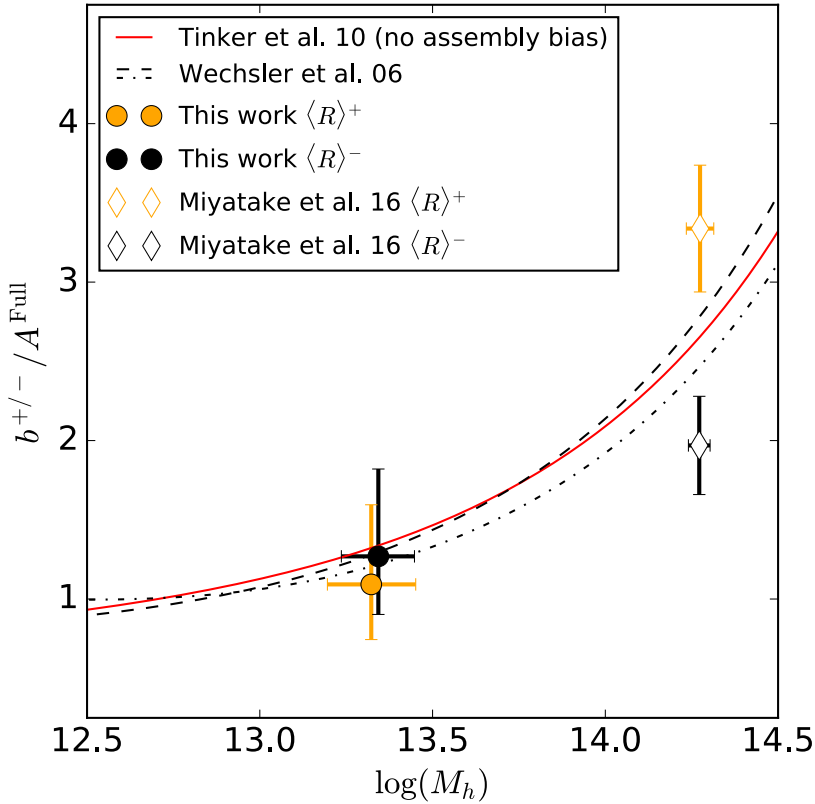


Figure 2.9: Comparison between the halo bias  $b$  and the predictions from the halo bias function from Tinker et al. (2010) and the concentration dependent halo bias from Wechsler et al. (2006), as a function of halo mass  $M_h$ . Here circles with error bars show the best fit value for  $b$  for each sample and diamonds show the results from Miyatake et al. (2016). The halo bias function from Tinker et al. (2010) is shown with a red line and the predictions from Wechsler et al. (2006) for different values of  $c'$  and a halo collapse mass  $M_c = 2.1 \times 10^{12} M_\odot/h$  (as defined therein). The dashed and dash-dotted lines are predictions for  $c'$  derived for our two samples,  $\langle R \rangle^+$  and  $\langle R \rangle^-$ , respectively. Note that the biases are normalised by the  $A^{\text{full}}$ .

## 2.6 DISCUSSION AND CONCLUSIONS

We have measured the galaxy-galaxy lensing signal of a selection of GAMA groups split into two samples according to the radial distribution of their satellite galaxies. We use the radial distribution of the satellite galaxies as a proxy for the halo assembly time, and report no evidence for halo assembly bias on galaxy group scales (typical masses of  $10^{13} M_{\odot}/h$ ). We use a halo model fit to constrain the halo masses and the large scale halo bias in order to see if the halo biases are consistent with those dictated solely by their halo masses. In this analysis, we used the KiDS data covering  $180 \text{ deg}^2$  of the sky (Hildebrandt et al. 2017), that fully overlaps with the three GAMA equatorial patches (G9, G12 and G15). As the photometric calibration and shape measurements analysis differ significantly from the previous KiDS data releases, we also perform additional tests for any possible systematic errors and biases that the new procedures might introduce (see Appendix 2.A).

Our findings are in agreement with the results from Zu et al. (2017), who re-analysed the SDSS redMaPPer clusters sample used in Miyatake et al. (2016) and found no evidence for halo assembly bias as previously claimed by Miyatake et al. (2016). They argue that that analysis suffered from misidentification of cluster members due to projection effects (Zu et al. 2017), which are minimised in the case when one uses spectroscopic information on cluster or group membership.

It is unlikely that our analysis suffers from the mis-identification of the GAMA galaxy groups members and/or contamination from background galaxies to the degree present in the SDSS case (up to 40%, Zu et al. 2017), and thus artificially changing the radial distribution of the satellite galaxies. The projection effects in our case come only from peculiar velocities (and mismatching from the FoF algorithm), whereas the projection effects in Miyatake et al. (2016) are dominated by photo- $z$  uncertainties and errors, which are much larger than peculiar velocities. If that would be the case, this would indeed have a larger effect on groups with a low number of member galaxies (and thus in the same regime we are using for our study). The GAMA groups are, due to available spectroscopic redshifts, highly pure and robust – for groups with  $N_{\text{FoF}} \geq 5$  the purity approaches 90% as assessed using a mock catalogue (Robotham et al. 2011). An issue that remains is the possible fragmentation of the GAMA galaxy groups by the FoF algorithm and a full assessment of this potential issue is beyond the scope of this paper and we defer these topics to a study in the future.

Additionally, the assumption of a NFW profile as our fiducial dark matter density profile can potentially affect the results. Exploration of different profiles is beyond the scope of this paper, but one would not expect that the different profiles would introduce differences in the obtained halo biases. The dark matter density profile does not enter into predictions for the two-halo term which carries all the biasing information. Moreover, any systematic effects due to the differences in profile would enter into both samples in the same way, and when taking the ratio of any quantities, they would to a large extent cancel out.

In order to reach a better precision in our lensing measurements, we could use the full KiDS-450 survey area. This is limited however by the lack of spectroscopy to create a group catalogue. The GAMA survey will be expanded into a newer and upcoming spectroscopic survey named WAVES (Driver et al. 2016)<sup>4</sup>, which is planned to cover the southern half of the KiDS survey (700 deg<sup>2</sup>) and provide redshifts for up to 2 million galaxies, which should provide us with enough statistical power not only to access the signatures of assembly bias in those galaxies but to extend the observational evidence also to galaxy scales.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for their very useful comments and suggestions. AD would like to thank to Keira J. Brooks and Christos Georgiou for proof reading the manuscript.

KK acknowledges support by the Alexander von Humboldt Foundation. HHo and RH acknowledges support from the European Research Council under FP7 grant number 279396. RN acknowledges support from the German Federal Ministry for Economic Affairs and Energy (BMWi) provided via DLR under project no. 50QE1103. MV acknowledges support from the European Research Council under FP7 grant number 279396 and the Netherlands Organisation for Scientific Research (NWO) through grants 614.001.103. IFC acknowledges the use of computational facilities procured through the European Regional Development Fund, Project ERDF-080 – A supercomputing laboratory for the University of Malta. CH acknowledges support from the European Research Council under grant number 647112. HHi is supported by an Emmy Noether grant (No. Hi 1495/2-1) of the Deutsche Forschungsgemeinschaft. This work is supported by the Deutsche Forschungsgemeinschaft in the framework of the TR33 ‘The Dark Universe’. EvU acknowledges support from an STFC Ernest Rutherford Research Grant, grant reference ST/L00285X/1. AC acknowledges support from the European Research Council under the FP7 grant number 240185.

This research is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017 and 177.A-3018, and on data products produced by Target/OmegaCEN, INAF-OACN, INAF-OAPD and the KiDS production team, on behalf of the KiDS consortium.

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 programs 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 <http://www.gama-survey.org>.

---

<sup>4</sup>Homepage: <http://www.wavesurvey.org>

This work has made use of Python (<http://www.python.org>), including the packages `numpy` (<http://www.numpy.org>) and `scipy` (<http://www.scipy.org>). The halo model is built upon `hmf` Python package by Murray et al. (2013). Plots have been produced with `matplotlib` (Hunter 2007) and `corner.py` (Foreman-Mackey 2016).

## 2.A SYSTEMATICS TESTS

We show here additional systematic tests performed as the image reduction procedure, photometric redshift calibration and shape measurement steps differ significantly from the methods used in Viola et al. (2015). We devise a number of tests to see how the obtained data behaves in different observational limits, and the results are presented in the following paragraphs.

### 2.A.1 MULTIPLICATIVE BIAS

The estimates of the average multiplicative bias  $m$  for each redshift slice used in the calculation are obtained using a method presented in Fenech Conti et al. (2017). They are further weighted by the weight  $w' = w_s D(z_l, z_s) / D(z_s)$  for a given lens-source sample. Typically, the value of the  $\mu$  correction is around  $-0.014$ , independent of the scale at which it is computed. Figure 2.10 shows the estimates of the average multiplicative bias  $m$  for each redshift slice used in the calculation.

### 2.A.2 ADDITIVE BIAS

Secondly, we test for the presence of the additive shear bias, by checking the tangential shear component measured around random points. This is calculated by performing lensing measurements around 10 million random points in RA and DEC (for all three GAMA patches), which have the same assigned redshift distribution as the GAMA galaxies. We use version 1 of the GAMA random catalogue, created as described in Farrow et al. (2015). Like the cross component of the measured ellipticities, also the azimuthally averaged tangential shear signal around random points should equal to zero. Figures 3.3 and 2.12 show significant systematic errors on scales larger than  $1 \text{ Mpc}/h$  as well as patch-dependent systematic errors. We perform the analysis on three patches separately (G9, G12 and G15). As discussed in Hildebrandt et al. (2017) and Fenech Conti et al. (2017), the correction for the additive bias obtained using image simulations should only be obtained for individual KiDS patches, due to specific systematics associated with each patch. We also check for the behaviour of the cross shear component. Any presence of the cross component signal points towards the presence of systematic errors and thus measurements on scales with significant cross component signal have to be corrected before using them for scientific purposes.



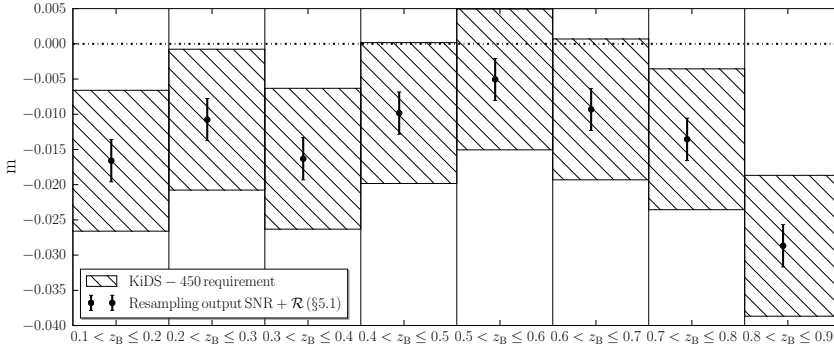


Figure 2.10: Multiplicative bias calculated using the resampling technique of Fenech Conti et al. (2017, chapter 5.1) in the redshift slices used in this analysis. The hatched area indicates the requirement on the knowledge of the multiplicative bias for KiDS-450 cosmic shear analysis (Hildebrandt et al. 2017).

One could estimate the additive bias using image simulations (using a method shown in Fenech Conti et al. 2017), but that will only account for the PSF effects. We correct for the additive bias using the results obtained from the random signal as the additive bias might arise because of spurious objects (including asteroids, stellar spikes, pixel defects, etc.) in our lensing data, apart from PSF effects. It is thus important to correct for it using the data. Correction of additive bias is performed by subtracting the random signal obtained for each patch from the true ESD measurement in the same patches. Doing so, that also gives better covariance matrix estimates (Singh et al. 2017). The final ESD profile is calculated by combining the random-subtracted signals from all three patches.

### 2.A.3 GROUP MEMBER CONTAMINATION OF THE SOURCE GALAXIES

The next important test we perform is to check how much the GAMA galaxy group members contaminate our source population (the so-called *boost factor*; Miyatake et al. 2015; van Uitert et al. 2017). Those galaxies will dilute the lensing signal (as they are not lensed). The resulting lensing signal will be biased (Figure 2.13) on small scales with the source over-density up to 30% at 75 kpc/h (Figure 2.13). We can impose a more stringent cut than the cut  $z_s > z_1$  used in previous studies on KiDS and GAMA data, by adding an offset  $\delta_z$  to the cut on the source population. As seen in Figure 2.13, using a conservative cut with  $\delta_z = 0.1$  still leaves a 10% over-density in the source sample. More conservative cuts lower the observed over-density, as expected. They also suppress the contamination, but this is not ideal as real source galaxies are removed as well, since it decreases the lensing signal-to-noise. On the small scales (below 75 kpc/h) the decrease of the source density is connected with the fact that the source galaxies become obscured by the host BCG of the GAMA group. The ESD

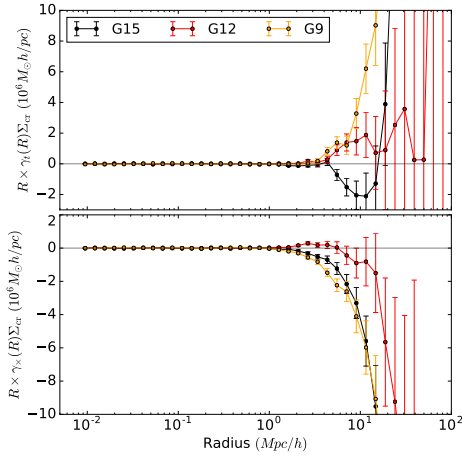


Figure 2.11: Shear signal around 10 million random points having the same redshift distribution as GAMA galaxies, split between the three GAMA patches. Shown are both tangential ( $\gamma_{t,}$  upper panel) and cross ( $\gamma_{\times}$  lower panel) components. We use these measurements to correct for the additive bias in our measured ESD signal.

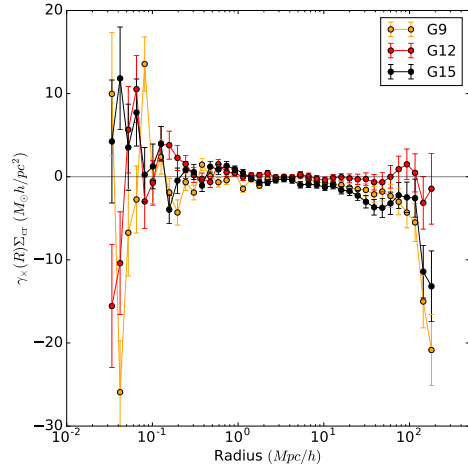


Figure 2.12: Lensing signal computed from the cross component of measured ellipticities, around all GAMA galaxies in the three equatorial patches (G9, G12 and G15). One can see, that the systematic errors significantly affect the signal below 70 kpc/h and above 10 Mpc/h, with the G12 patch being the least affected, even after subtracting the signal computed around random points.

signals in Figure 2.13 are corrected with the boost factor using the factors shown in the top panel of the same figure and have lensing efficiency calculated separately for each redshift cut. We find that for a redshift offset of  $\delta_z = 0.2$  the boost correction is not necessary.

#### 2.A.4 SOURCE REDSHIFT DISTRIBUTION

The significant difference between this analysis and previous method presented in Viola et al. (2015) is the usage of full redshift probability distribution of the sources,  $n(z_s)$ , compared to Viola et al. (2015) where each source is given its own posterior redshift distribution  $p(z_s)$  obtained from BPZ. With the following tests we want to see what the difference between having only the global  $n(z_s)$  has on the error budget and the resulting lensing signals. The observable lensing signal depends on the angular diameter distances to the lens and source galaxies (equation 2.9). The redshifts to the lens galaxies are known from the GAMA spectroscopic survey, while for the sources we need to resort to the photometric redshifts derived using multi-band images (in *ugri* photometric bands) of the KiDS survey. The colors obtained using those images are a basis for the photometric redshift estimates, which also provides us the

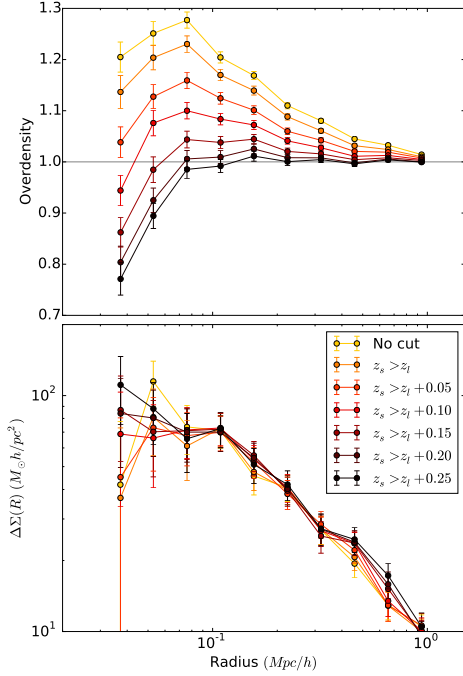


Figure 2.13: *Top panel:* The overdensity of KiDS source galaxies around GAMA galaxy groups with richness  $N_{\text{FOF}} \geq 5$ . The various lines correspond to different redshift cuts applied to the source sample. Even for a conservative cut of  $z_s > z_l + 0.1$ , we find a residual contamination of group members in the source sample of up to 10% at 75 kpc. *Bottom panel:* The ESD signal around GAMA galaxy groups with richness  $N_{\text{FOF}} \geq 5$  up to 2 Mpc/h. The various lines correspond to different redshift cuts applied to the source sample. The redshift cut does not significantly affect the lensing signal, but one removes any possible problems due to group contamination. The lensing signals are computed using different lensing efficiencies and are corrected with the boost factor using the factors shown in the top panel.

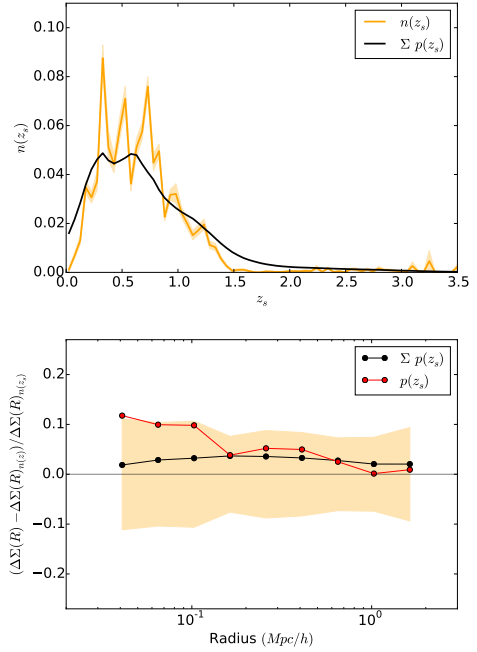


Figure 2.14: *Top panel:* Comparison of the  $n(z_s)$  as given by the direct calibration method (DIR) and the stacked  $p(z_s)$  obtained from BPZ (Hildebrandt et al. 2017). As already noted in Hildebrandt et al. (2017), the stacked  $p(z_s)$  does not accurately reproduce the features seen in the DIR method, and its usage is discouraged. *Bottom panel:* Difference between the lensing signal using three different source redshift distributions.  $p(z_s)$  represents the method as used in Viola et al. (2015), compared to the stacked  $p(z_s)$  and the  $n(z_s)$  obtained using DIR (for all 180,960 GAMA galaxies). Within the error budget, all the methods are in agreement (the orange area is the error on the lensing signal calculated using the  $n(z_s)$ ).

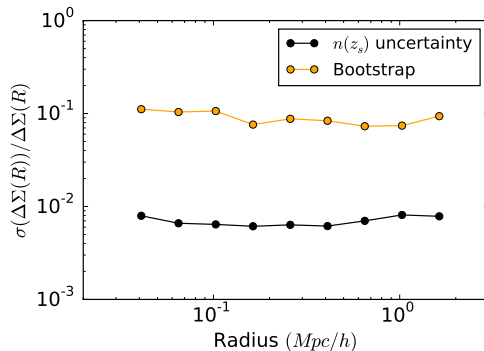


Figure 2.15: Relative error estimates of the  $n(z_s)$  uncertainty compared to the uncertainty as obtained using the bootstrap method on the lensing signal (including shape noise and cosmic variance contributions), calculated for the full sample of GAMA galaxies in the three equatorial patches (G9, G12 and G15). It can be seen that the contribution to the total error budget from the uncertainty of the redshift distribution is negligible.

full redshift probability distribution of the sources,  $n(z_s)$ , obtained using the direct calibration method (for more information and comparison with other techniques see Hildebrandt et al. 2017). Comparison between the final lensing signals using the individual  $p(z_s)$ , the stack of  $p(z_s)$  and the global  $n(z_s)$  can be seen in the bottom panel of Figure 2.14 and the difference between the stacked  $p(z_s)$  and  $n(z_s)$  probability distributions in the top panel of the same Figure. The resulting lensing signals do not change much, and are all in agreement within the error budget of the lensing signal of all the GAMA galaxies. Following Hildebrandt et al. (2017), we adopt the redshift range [0.1, 0.9], which is the same as the covered range by the 4 tomographic bins used in Hildebrandt et al. (2017).

The uncertainty on the  $n(z_s)$  contributes to the total error budget of the lensing signal. As the errors due to this uncertainty can affect the conclusions of the quantitative results, we look into how much the actual contribution is. We take 1000 bootstrap realisations of the weighted spectroscopic catalogue (Hildebrandt et al. 2017) giving us 1000 different realisations of  $n(z_s)$ , for which we calculate the lensing signal. This gives us enough samples to constrain the uncertainty on the lensing signal due to the uncertainty on the  $n(z_s)$ . We compare the given  $1\sigma$  errors with the total error on our lensing signal. The results can be seen in Figure 2.15, where it is clearly seen that the uncertainty on  $n(z_s)$  is sub-dominant to the whole error budget.

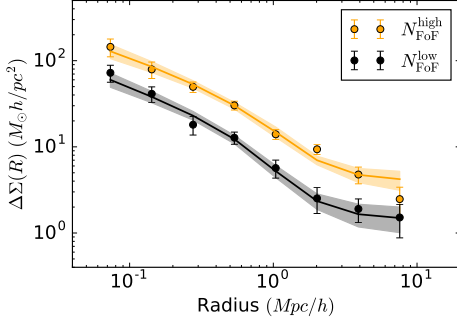


Figure 2.16: Stacked ESD profiles measured around the central galaxies of GAMA groups, selected according to the apparent richness of the groups. The solid orange and black lines represent the best-fitting halo model as obtained using a MCMC fit, with the 68% confidence interval indicated with a shaded region.

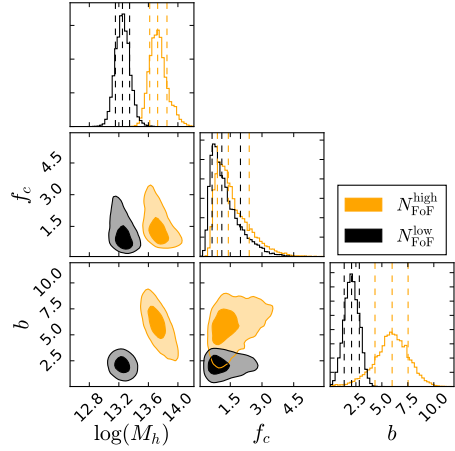


Figure 2.17: The posterior distributions of the halo model parameters  $M_h$ ,  $f_c$  and  $b$  for the sample of lenses split according to their apparent richness. The posterior distribution clearly shows a difference in the obtained halo masses as well as a significant difference in the obtained halo biases. The contours indicate  $1\sigma$  and  $2\sigma$  confidence regions.

## 2.B FULL POSTERIOR DISTRIBUTIONS

Figures 2.16 and 2.17 show lensing signal and posterior distributions of the additional test of splitting the full sample to two samples with high and low richnesses (as discussed in Section 2.5). In Figure 2.18 we show the full posterior probability distribution for all fitted parameters in our MCMC fit as discussed in Sections 2.4.1 and 2.5.

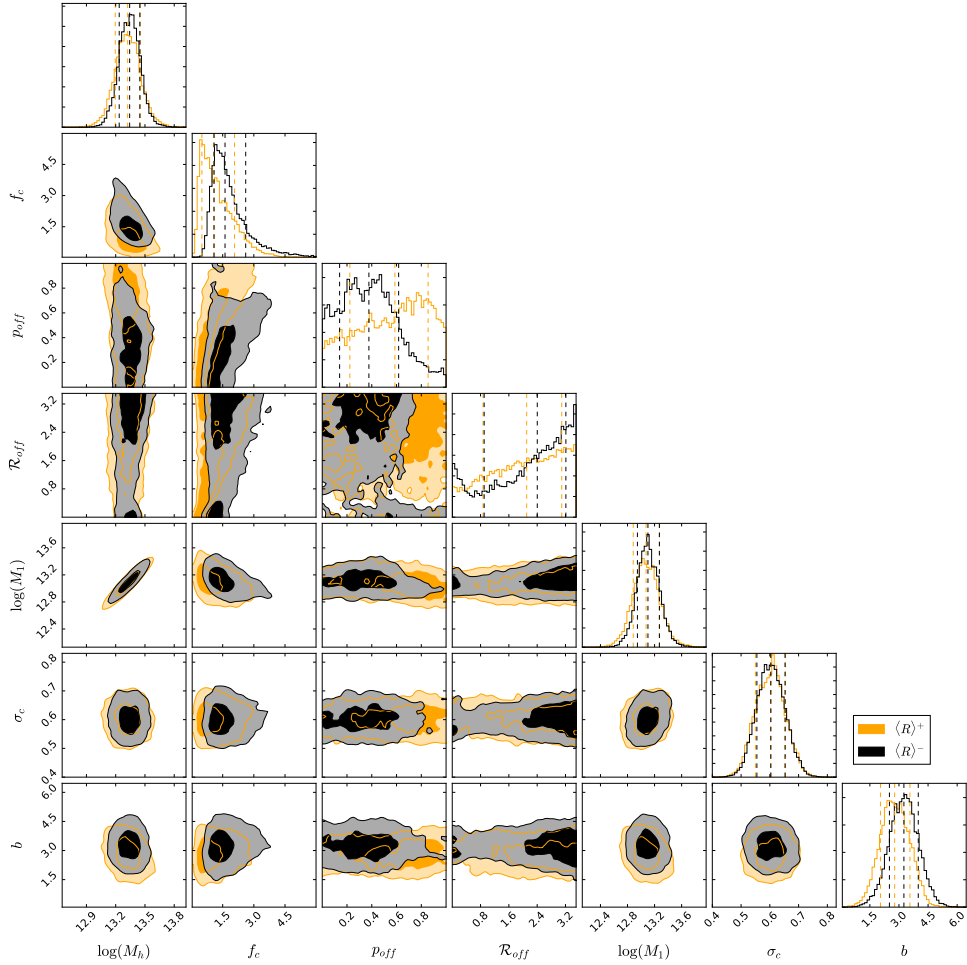



Figure 2.18: The full posterior distributions of the halo model parameters  $M_h$ ,  $f_c$ ,  $p_{\text{off}}$ ,  $\mathcal{R}_{\text{off}}$ ,  $M_1$ ,  $\sigma_c$  and  $b$ . The posterior distribution clearly shows a slight difference in the obtained halo masses as well as no difference in the obtained halo biases, the miscentering parameters and the normalisation of concentration-halo mass relation. The contours indicate  $1\sigma$  and  $2\sigma$  confidence regions. Priors used in the MCMC fit can be found in Section 2.4.1.

# 3

---

## Unveiling Galaxy Bias via the Halo Model, KiDS and GAMA

---

E measure the projected galaxy clustering and galaxy-galaxy lensing signals using the Galaxy And Mass Assembly (GAMA) survey and Kilo-Degree Survey (KiDS) to study galaxy bias. We use the concept of non-linear and stochastic galaxy biasing in the framework of halo occupation statistics to constrain the parameters of the halo occupation statistics and to unveil the origin of galaxy biasing. The bias function  $\Gamma_{\text{gm}}(r_p)$ , where  $r_p$  is the projected comoving separation, is evaluated using the analytical halo model from which the scale dependence of  $\Gamma_{\text{gm}}(r_p)$ , and the origin of the non-linearity and stochasticity in halo occupation models can be inferred. Our observations unveil the physical reason for the non-linearity and stochasticity, further explored using hydrodynamical simulations, with the stochasticity mostly originating from the non-Poissonian behaviour of satellite galaxies in the dark matter haloes and their spatial distribution, which does not follow the spatial distribution of dark matter in the halo. The observed non-linearity is mostly due to the presence of the central galaxies, as was noted from previous theoretical work on the same topic. We also see that overall, more massive galaxies reveal a stronger scale dependence, and out to a larger radius. Our results show that a wealth of information about galaxy bias is hidden in halo occupation models. These models should therefore be used to determine the influence of galaxy bias in cosmological studies.

A. Dvornik, H. Hoekstra, K. Kuijken, P. Schneider, et al.  
*MNRAS*, **Volume 479**, Issue 1, p. 1240-1259 (2018)

### 3.1 INTRODUCTION

In the standard cold dark matter and cosmological constant-dominated ( $\Lambda$ CDM) cosmological framework, galaxies form and reside within dark matter haloes, which themselves form from the highest density peaks in the initial Gaussian random density field (e.g. Mo et al. 2010, and references therein). In this case one expects that the spatial distribution of galaxies traces the spatial distribution of the underlying dark matter. Galaxies are however, biased tracers of the underlying dark matter distribution, because of the complexity of their evolution and formation (Davis et al. 1985; Dekel & Rees 1987; Cacciato et al. 2012). The relation between the distribution of galaxies and the underlying dark matter distribution, usually referred as *galaxy bias*, is thus important to understand in order to properly comprehend galaxy formation and interpret studies that use galaxies as tracers of the underlying dark matter, particularly for those trying to constrain cosmological parameters.

If such a relation can be described with a single number  $b$ , the galaxy bias is linear and deterministic. As galaxy formation is a complex process, it would be naive to assume that the relation between the dark matter density field and galaxies is a simple one, described only with a single number. Such a relation might be non-linear (the relation between a galaxy and matter density fields cannot be described with only a single number), scale dependent (the galaxy bias is different on the different scales studied) or stochastic (the biasing relation has an intrinsic scatter around the mean value). Numerous authors have presented various arguments for why simple linear and deterministic bias is highly questionable (Kaiser 1984; Davis et al. 1985; Dekel & Lahav 1999). Moreover, cosmological simulations and semi-analytical models suggest that galaxy bias takes a more complicated, non-trivial form (Wang et al. 2008; Zehavi et al. 2011).

Observationally, there have been many attempts to test if galaxy bias is linear and deterministic. There have been studies relying on clustering properties of different samples of galaxies (e.g. Wang et al. 2008; Zehavi et al. 2011), studies measuring high-order correlation statistics and ones directly comparing observed galaxy distribution fluctuations with the matter distribution fluctuations measured in numerical simulations (see Cacciato et al. 2012, and references therein). What is more, there have also been observations combining galaxy clustering with weak gravitational (galaxy-galaxy) lensing measurements (Hoekstra et al. 2002; Simon et al. 2007; Jullo et al. 2012; Buddendiek et al. 2016). The majority of the above observations have confirmed that galaxy bias is neither linear nor deterministic (Cacciato et al. 2012).

Even though the observational results are in broad agreement with theoretical predictions, until recently there was no direct connection between measurements and model predictions, mostly because the standard formalism used to define and predict the non-linearity and stochasticity of galaxy bias is hard to interpret in the framework of galaxy formation models. Cacciato et al. (2012) introduced a new approach that allows for intuitive interpretation of galaxy bias, that is directly linked to galaxy formation theory and various concepts therein. They reformulated the galaxy bias description (and the non-linearity and stochasticity of the relation between the galaxies and underlying dark matter distribution) presented by Dekel & Lahav (1999)



using the formalism of halo occupation statistics. As galaxies are thought to live in dark matter haloes, halo occupation distributions (a prescription on how galaxies populate dark matter haloes) are a natural way to describe the galaxy-dark matter connection, and consequently the nature of galaxy bias. Combining the halo occupation distributions with the halo model (Seljak 2000; Peacock & Smith 2000; Cooray & Sheth 2002; van den Bosch et al. 2013; Mead et al. 2015; Wibking et al. 2019), allows us to compare observations to predictions of those models, which has the potential to unveil the *hidden factors* – sources of deviations from the linear and deterministic biasing (Cacciato et al. 2012). Recently Simon & Hilbert (2017) also showed that the halo model contains important information about galaxy bias. In this paper, however, we demonstrate how the stochasticity of galaxy bias arises from two different sources; the first is the relation between dark matter haloes and the underlying dark matter field, and the second is the manner in which galaxies populate dark matter haloes. As in Cacciato et al. (2012), we will focus on the second source of stochasticity, which indeed can be addressed using a halo model combined with halo occupation distributions.

The aim of this paper is to measure the galaxy bias using state of the art galaxy surveys and constrain the nature of it using the halo occupation distribution formalism. The same formalism can provide us with insights on the sources of deviations from the linear and deterministic biasing and the results can be used in cosmological analyses using the combination of galaxy-galaxy lensing and galaxy clustering and those based on the cosmic shear measurements. In this paper we make use of the predictions of Cacciato et al. (2012) and apply them to the measurements provided by the imaging Kilo-Degree Survey (KiDS; Kuijken et al. 2015; de Jong et al. 2015), accompanied by the spectroscopic Galaxy And Mass Assembly (GAMA) survey (Driver et al. 2011) in order to get a grasp of the features of galaxy bias that can be measured using a combination of galaxy clustering and galaxy-galaxy lensing measurements with high precision.

The outline of this paper is as follows. In Section 3.2 we recap the galaxy biasing formulation of Cacciato et al. (2012). In Section 3.3 we introduce the halo model, its ingredients and introduce the main observable, which is a combination of galaxy clustering and galaxy-galaxy lensing. In Section 3.4 we present the data and measurement methods used in our analysis. We present our galaxy biasing results in Section 3.5, together with comparison with simulations and discuss and conclude in Section 3.6. In the Appendix, we detail the calculation of the analytical covariance matrix, and provide full pairwise posterior distributions of our derived halo model parameters. We also provide a detailed derivation of the connection between the galaxy-matter correlation and the galaxy-galaxy lensing signal, explaining the use of two different definitions of the critical surface mass density in the literature. We highlight the key differences between our expressions and those found in several recent papers.

Throughout the paper we use the following cosmological parameters entering in the calculation of the distances and in the halo model (Planck Collaboration et al. 2016):  $\Omega_m = 0.3089$ ,  $\Omega_\Lambda = 0.6911$ ,  $\sigma_8 = 0.8159$ ,  $n_s = 0.9667$  and  $\Omega_b = 0.0486$ . We also use  $\bar{\rho}_m$  as the present day mean matter density of the Universe ( $\bar{\rho}_m = \Omega_{m,0} \rho_{\text{crit}}$ , where  $\rho_{\text{crit}} = 3H_0^2/(8\pi G)$ ) and the halo masses are defined as  $M = 4\pi r_\Delta^3 \Delta \bar{\rho}_m / 3$  enclosed by the radius  $r_\Delta$  within which the mean density of the halo is  $\Delta$  times  $\bar{\rho}_m$ , with  $\Delta = 200$ ). All the measurements presented in the paper are in comoving units, and  $\log$  and  $\ln$  refer to the 10-based logarithm and the natural logarithm, respectively.

## 3.2 BIASING

This paper closely follows the biasing formalism presented in Cacciato et al. (2012), and we refer the reader to that paper for a thorough treatment of the topic. Here we shortly recap the galaxy biasing formalism of Cacciato et al. (2012) and correct a couple of typos that we discovered during the study of his work. In this formalism the mean biasing function  $b(M)$  (the equivalent of the mean biasing function  $b(\delta_m)$  as defined by Dekel & Lahav 1999) is, using new variables: the number of galaxies in a dark matter halo,  $N$ , and the mass of a dark matter halo,  $M$ :

$$b(M) \equiv \frac{\bar{\rho}_m}{\bar{n}_g} \frac{\langle N|M \rangle}{M}, \quad (3.1)$$

where  $\bar{n}_g$  is the average number density of galaxies and  $\langle N|M \rangle$  is the mean of the halo occupation distribution for a halo of mass  $M$ , defined as:

$$\langle N|M \rangle = \sum_{N=0}^{\infty} N P(N|M), \quad (3.2)$$

where  $P(N|M)$  is the halo occupation distribution. Note that in this case, the simple linear, deterministic biasing corresponds to:

$$N = \frac{\bar{n}_g}{\bar{\rho}_m} M, \quad (3.3)$$

which gives the expected value of  $b(M) = 1$ . As  $N$  is an integer and the quantities  $\bar{\rho}_m$ ,  $\bar{n}_g$  and  $M$  are in general non-integer, it is clear that in this formulation the linear, deterministic bias is unphysical. We define the moments of the bias function  $b(M)$  as

$$\hat{b} \equiv \frac{\langle b(M)M^2 \rangle}{\langle M^2 \rangle}, \quad (3.4)$$

and

$$\tilde{b}^2 \equiv \frac{\langle b^2(M)M^2 \rangle}{\langle M^2 \rangle}, \quad (3.5)$$

where  $\langle \dots \rangle^1$  indicates an effective average (an integral over dark matter haloes) defined in the following form:

$$\langle x \rangle \equiv \int_0^\infty x n(M) dM, \quad (3.6)$$

where  $n(M)$  is the halo mass function and  $x$  is a property of the halo or galaxy population. In the case of linear bias,  $b(M)$  is a constant and hence  $\tilde{b}/\hat{b} = 1$ . The same ratio,  $\tilde{b}/\hat{b}$ , is the relevant measure of the non-linearity of the biasing relation (Dekel & Lahav 1999). Its deviation from unity is a sign of a non-linear galaxy bias. From equation 3.1 we can see that linear bias corresponds to halo occupation statistics for which  $\langle N|M \rangle \propto M$ .

In the same manner Cacciato et al. (2012) also define the random halo bias of a single halo of mass  $M$ , that contains  $N$  galaxies, as:

$$\varepsilon_N \equiv N - \langle N|M \rangle, \quad (3.7)$$

which, by definition, will have a zero mean when averaged over all dark matter haloes, i.e.  $\langle \varepsilon_N|M \rangle = 0$ . This can be used to define the halo stochasticity function:

$$\sigma_b^2(M) \equiv \left( \frac{\bar{\rho}_m}{\bar{n}_g} \right)^2 \frac{\langle \varepsilon_N^2|M \rangle}{\langle M^2 \rangle}, \quad (3.8)$$

from which, after averaging over halo mass, one gets the stochasticity parameter:

$$\sigma_b^2 \equiv \left( \frac{\bar{\rho}_m}{\bar{n}_g} \right)^2 \frac{\langle \varepsilon_N^2 \rangle}{\langle M^2 \rangle}. \quad (3.9)$$

If the stochasticity parameter  $\sigma_b = 0$ , then the galaxy bias is deterministic. In addition to the two bias moments  $\tilde{b}$  and  $\hat{b}$ , one can also define some other bias parameters, particularly the ratio of the variances  $b_{\text{var}}^2 \equiv \langle \delta_g^2 \rangle / \langle \delta_m^2 \rangle$  (Dekel & Lahav 1999; Cacciato et al. 2012). Using this definition and an HOD-based formulation, Cacciato et al. (2012) show that:

$$b_{\text{var}}^2 = \left( \frac{\bar{\rho}_m}{\bar{n}_g} \right)^2 \frac{\langle N^2 \rangle}{\langle M^2 \rangle}, \quad (3.10)$$

where the averages are again calculated according to equation (3.6). As the bias parameter is sensitive to both non-linearity and stochasticity, the total variance of the bias  $b_{\text{var}}^2$  can also be written as:

$$b_{\text{var}}^2 = \tilde{b}^2 + \sigma_b^2. \quad (3.11)$$

Combining equation (3.10) and (3.11) we find a relation for  $\langle N^2 \rangle$

$$\langle N^2 \rangle = \left( \frac{\bar{n}_g}{\bar{\rho}_m} \right)^2 \left[ \tilde{b}^2 + \sigma_b^2 \right] \langle M^2 \rangle. \quad (3.12)$$

<sup>1</sup>Cacciato et al. (2012) used  $\sigma_M^2 \equiv \langle M^2 \rangle$  throughout the paper, and we decided to drop the  $\sigma_M^2$  for cleaner and more consistent equations.

We can compare this to the covariance, which is obtained directly from equations (3.1) and (3.3):

$$\langle NM \rangle = \frac{\bar{n}_g}{\rho_m} \hat{b} \langle M^2 \rangle. \quad (3.13)$$

From all the equations above, it also directly follows that one can define a linear correlation coefficient as:  $r \equiv \langle NM \rangle / [\langle N^2 \rangle \langle M^2 \rangle]$ , such that, combining equations (3.12) and (3.13),  $\hat{b}$  can be written as:  $\hat{b} = b_{\text{var}} r$ .

This enables us to consider some special cases. The discrete nature of galaxies does not allow us to have galaxy bias that is both linear and deterministic (Cacciato et al. 2012). Despite that, halo occupation statistics do allow bias that is linear and stochastic where;

$$\begin{aligned} \hat{b} = \tilde{b} = b(M) = 1 & & b_{\text{var}} = (1 + \sigma_b^2)^{1/2} \\ \sigma_b \neq 0 & & r = (1 + \sigma_b^2)^{-1/2}. \end{aligned} \quad (3.14)$$

or non-linear and deterministic;

$$\begin{aligned} \hat{b} \neq \tilde{b} \neq 1 & & b_{\text{var}} = \tilde{b} \\ \sigma_b = 0 & & r = \hat{b}/\tilde{b} \neq 1. \end{aligned} \quad (3.15)$$

### 3.3 HALO MODEL

To express the HOD, we use the halo model, a successful analytic framework used to describe the clustering of dark matter and its evolution in the Universe (Seljak 2000; Peacock & Smith 2000; Cooray & Sheth 2002; van den Bosch et al. 2013; Mead et al. 2015). The halo model provides an ideal framework to describe the statistical weak lensing signal around a selection of galaxies, their clustering and cosmic shear signal. The halo model is built upon the statistical description of the properties of dark matter haloes (namely the average density profile, large scale bias and abundance) as well as on the statistical description of the galaxies residing in them. The halo model allows us to unveil the *hidden* sources of bias stochasticity (Cacciato et al. 2012).

#### 3.3.1 HALO MODEL INGREDIENTS

We assume that dark matter haloes are spherically symmetric, on average, and have density profiles,  $\rho(r|M) = M u_h(r|M)$ , that depend only on their mass  $M$ , and  $u_h(r|M)$  is the normalised density profile of a dark matter halo. Similarly, we assume that satellite galaxies in haloes of mass  $M$  follow a spherical number density distribution  $n_s(r|M) = N_s u_s(r|M)$ , where  $u_s(r|M)$  is the normalised density profile of satellite galaxies. Central galaxies always have  $r = 0$ . We assume that the density profile of dark matter haloes follows an NFW profile (Navarro et al. 1997).

Since centrals and satellites are distributed differently, we write the galaxy-galaxy power spectrum as:

$$P_{\text{gg}}(k) = f_c^2 P_{\text{cc}}(k) + 2f_c f_s P_{\text{cs}}(k) + f_s^2 P_{\text{ss}}(k), \quad (3.16)$$

while the galaxy-dark matter cross power spectrum is given by:

$$P_{\text{gm}}(k) = f_c P_{\text{cm}}(k) + f_s P_{\text{sm}}(k). \quad (3.17)$$

Here  $f_c = \bar{n}_c/\bar{n}_g$  and  $f_s = \bar{n}_s/\bar{n}_g = 1 - f_c$  are the central and satellite fractions, respectively, and the average number densities  $\bar{n}_g$ ,  $\bar{n}_c$  and  $\bar{n}_s$  follow from:

$$\bar{n}_x = \int_0^\infty \langle N_x | M \rangle n(M) dM, \quad (3.18)$$

where 'x' stands for 'g' (for galaxies), 'c' (for centrals) or 's' (for satellites) and  $n(M)$  is the halo mass function in the following form:

$$n(M) = \frac{\bar{\rho}_m}{M^2} \nu f(\nu) \frac{d \ln \nu}{d \ln M}, \quad (3.19)$$

with  $\nu = \delta_c/\sigma(M)$ , where  $\delta_c$  is the critical overdensity for spherical collapse at redshift  $z$ , and  $\sigma(M)$  is the mass variance. For  $f(\nu)$  we use the form presented in Tinker et al. (2010). In addition, it is common practice to split two-point statistics into a 1-halo term (both points are located in the same halo) and a 2-halo term (the two points are located in different haloes). The 1-halo terms are:

$$P_{\text{cc}}^{\text{1h}}(k) = \frac{1}{\bar{n}_c}, \quad (3.20)$$

$$P_{\text{ss}}^{\text{1h}}(k) = \beta \int_0^\infty \mathcal{H}_s^2(k, M) n(M) dM, \quad (3.21)$$

and all other terms are given by:

$$P_{\text{xy}}^{\text{1h}}(k) = \int_0^\infty \mathcal{H}_x(k, M) \mathcal{H}_y(k, M) n(M) dM. \quad (3.22)$$

Here 'x' and 'y' are either 'c' (for central), 's' (for satellite), or 'm' (for matter),  $\beta$  is a Poisson parameter which arises from considering a scatter in the number of satellite galaxies at fixed halo mass [in this case a free parameter – we define the  $\beta$  in detail using equations (3.40), (3.41) and (3.42)] and we have defined

$$\mathcal{H}_m(k, M) = \frac{M}{\bar{\rho}_m} \tilde{u}_h(k|M), \quad (3.23)$$

$$\mathcal{H}_c(k, M) = \frac{\langle N_c | M \rangle}{\bar{n}_c}, \quad (3.24)$$

and

$$\mathcal{H}_s(k, M) = \frac{\langle N_s | M \rangle}{\bar{n}_s} \tilde{u}_s(k|M), \quad (3.25)$$

with  $\tilde{u}_h(k|M)$  and  $\tilde{u}_s(k|M)$  the Fourier transforms of the halo density profile and the satellite number density profile, respectively, both normalised to unity [ $\tilde{u}(k=0|M)=1$ ]. The various 2-halo terms are given by:

$$P_{xy}^{2h}(k) = P_{\text{lin}}(k) \int_0^\infty dM_1 \mathcal{H}_x(k, M_1) b_h(M_1) n(M_1) \times \int_0^\infty dM_2 \mathcal{H}_y(k, M_2) b_h(M_2) n(M_2), \quad (3.26)$$

where  $P_{\text{lin}}(k)$  is the linear power spectrum, obtained using the Eisenstein & Hu (1998) transfer function, and  $b_h(M, z)$  is the halo bias function. Note that in this formalism, the matter-matter power spectrum simply reads:

$$P_{\text{mm}}(k) = P_{\text{mm}}^{1h}(k) + P_{\text{mm}}^{2h}(k). \quad (3.27)$$

The two-point correlation functions corresponding to these power-spectra are obtained by simple Fourier transformation:

$$\xi_{xy}(r) = \frac{1}{2\pi^2} \int_0^\infty P_{xy}(k) \frac{\sin kr}{kr} k^2 dk, \quad (3.28)$$

For the halo bias function,  $b_h$ , we use the fitting function from Tinker et al. (2010), as it was obtained using the same numerical simulation from which the halo mass function was obtained. We have adopted the parametrization of the concentration-mass relation, given by Duffy et al. (2008):

$$c(M, z) = 10.14 A_c \left[ \frac{M}{(2 \times 10^{12} M_\odot / h)} \right]^{-0.081} (1+z)^{-1.01}, \quad (3.29)$$

with a free normalisation  $A_c$  that accounts for the theoretical uncertainties in the concentration-mass relation due to discrepancies in the numerical simulations (mostly resolution and cosmologies) from which this scaling is usually inferred (Viola et al. 2015). We allow for additional normalisation  $A_s$  for satellites, such that

$$c_s(M, z) = A_s c(M, z), \quad (3.30)$$

which governs how satellite galaxies are spatially distributed inside a dark matter halo and tests the assumption of satellite galaxies following the density distribution of the dark matter haloes. If  $A_s \neq 1$ , the galaxy bias will vary on small scales, as demonstrated by Cacciato et al. (2012).

### 3.3.2 CONDITIONAL STELLAR MASS FUNCTION

In order to constrain the cause for the stochasticity, non-linearity and scale dependence of galaxy bias, we model the halo occupation statistics using the Conditional Stellar Mass Function (CSMF, heavily motivated by Yang et al. 2008; Cacciato et al. 2009, 2013; Wang et al. 2013; van Uitert et al. 2016). The CSMF,  $\Phi(M_\star|M)$ , specifies the average number of galaxies of stellar mass  $M_\star$  that reside in a halo of mass  $M$ . In this formalism, the halo occupation statistics of central galaxies are defined via the function:

$$\Phi(M_\star|M) = \Phi_c(M_\star|M) + \Phi_s(M_\star|M). \quad (3.31)$$

In particular, the CSMF of central galaxies is modelled as a log-normal,

$$\Phi_c(M_\star|M) = \frac{1}{\sqrt{2\pi} \ln(10) \sigma_c M_\star} \exp\left[-\frac{\log(M_\star/M_c^*)^2}{2 \sigma_c^2}\right], \quad (3.32)$$

and the satellite term as a modified Schechter function,

$$\Phi_s(M_\star|M) = \frac{\phi_s^*}{M_s^*} \left(\frac{M_\star}{M_s^*}\right)^{\alpha_s} \exp\left[-\left(\frac{M_\star}{M_s^*}\right)^2\right], \quad (3.33)$$

where  $\sigma_c$  is the scatter between stellar mass and halo mass and  $\alpha_s$  governs the power law behaviour of satellite galaxies. Note that  $M_c^*$ ,  $\sigma_c$ ,  $\phi_s^*$ ,  $\alpha_s$  and  $M_s^*$  are, in principle, all functions of halo mass  $M$ . We assume that  $\sigma_c$  and  $\alpha_s$  are independent of the halo mass  $M$ . Inspired by Yang et al. (2008), we parametrise  $M_c^*$ ,  $M_s^*$  and  $\phi_s^*$  as:

$$M_c^*(M) = M_0 \frac{(M/M_1)^{\gamma_1}}{[1 + (M/M_1)]^{\gamma_1 - \gamma_2}}. \quad (3.34)$$

$$M_s^*(M) = 0.56 M_c^*(M), \quad (3.35)$$

and

$$\log[\phi_s^*(M)] = b_0 + b_1(\log m_{12}), \quad (3.36)$$

where  $m_{12} = M/(10^{12} M_\odot h^{-1})$ . The factor of 0.56 is also inspired by Yang et al. (2008) and further tests by van Uitert et al. (2016) showed that using this assumption does not significantly affect the results. We can see that the stellar to halo mass relation for  $M \ll M_1$  behaves as  $M_c^* \propto M^{\gamma_1}$  and for  $M \gg M_1$ ,  $M_c^* \propto M^{\gamma_2}$ , where  $M_1$  is a characteristic mass scale and  $M_0$  is a normalisation. Here  $\gamma_1$ ,  $\gamma_2$ ,  $b_0$  and  $b_1$  are all free parameters.

From the CSMF it is straightforward to compute the halo occupation numbers. For example, the average number of galaxies with stellar masses in the range  $M_{\star,1} \leq M_\star \leq M_{\star,2}$  is thus given by:

$$\langle N|M \rangle = \int_{M_{\star,1}}^{M_{\star,2}} \Phi(M_\star|M) dM_\star. \quad (3.37)$$

The distinction we have made here, by splitting galaxies into centrals or satellites, is required to illustrate the main source of non-linearity and scale dependence of galaxy bias (see results in Section 3.5). To explore this, we follow Cacciato et al. (2012), and define the random halo biases following similar procedure as in equation (3.7):

$$\varepsilon_c \equiv N_c - \langle N_c | M \rangle \quad \text{and} \quad \varepsilon_s \equiv N_s - \langle N_s | M \rangle, \quad (3.38)$$

and the halo stochasticity functions for centrals and satellites are given by:

$$\begin{aligned} \langle \varepsilon_c^2 | M \rangle &= \sum_{N_c=0}^{\infty} (N_c - \langle N_c | M \rangle)^2 P(N_c | M) \\ &= \langle N_c^2 | M \rangle - \langle N_c | M \rangle^2 \\ &= \langle N_c | M \rangle - \langle N_c | M \rangle^2, \end{aligned} \quad (3.39)$$

$$\begin{aligned} \langle \varepsilon_s^2 | M \rangle &= \sum_{N_s=0}^{\infty} (N_s - \langle N_s | M \rangle)^2 P(N_s | M) \\ &= \langle N_s^2 | M \rangle - \langle N_s | M \rangle^2, \end{aligned} \quad (3.40)$$

where we have used the fact that  $\langle N_c^2 | M \rangle = \langle N_c | M \rangle$ , which follows from the fact that  $N_c$  is either zero or unity. We can see that central galaxies only contribute to the stochasticity if  $\langle N_c | M \rangle < 1$ . If  $\langle N_c | M \rangle = 1$ , then the HOD is deterministic and the stochasticity function  $\langle \varepsilon_c^2 | M \rangle = 0$ . The CSME, however, only specifies the first moment of the halo occupation distribution  $P(N|M)$ . For central galaxies this is not a problem, as  $\langle N_c^2 | M \rangle = \langle N_c | M \rangle$ . For satellite galaxies, we use that

$$\langle N_s^2 | M \rangle = \beta(M) \langle N_s | M \rangle^2 + \langle N_s | M \rangle, \quad (3.41)$$

where  $\beta(M)$  is the mass dependent Poisson parameter defined as:

$$\beta(M) \equiv \frac{\langle N_s(N_s - 1) | M \rangle}{\langle N_s | M \rangle^2}, \quad (3.42)$$

which is unity if  $P(N_s|M)$  is given by a Poisson distribution, larger than unity if the distribution is wider than a Poisson distribution (also called super-Poissonian distribution) or smaller than unity if the distribution is narrower than a Poisson distribution (also called sub-Poissonian distribution). If  $\beta(M)$  is unity, then equation (3.40) takes a simple form  $\langle \varepsilon_s^2 | M \rangle = \langle N_s | M \rangle$ . In what follows we limit ourselves to cases in which  $\beta(M)$  is independent of halo mass, i.e.,  $\beta(M) = \beta$ , and we treat  $\beta$  as a free parameter.

Even without an application to the data, we can already learn a lot about the nature of galaxy bias from combining the HOD and halo model approaches to galaxy biasing as described in Section 3.2. As realistic HODs (as formulated above) differ strongly from the simple scaling  $\langle N | M \rangle \propto M$  (equation 3.3, which gives the linear and deterministic galaxy bias), they will inherently predict a galaxy bias that is strongly



non-linear. Moreover, this seems to be mostly the consequence of central galaxies for which  $\langle N_c|M \rangle$  never follows a power law. Even the satellite occupation distribution  $\langle N_s|M \rangle$  is never close to the power law form, due to a cut-off at the low mass end, as galaxies at certain stellar mass require a minimum mass for their host halo (Cacciato et al. 2012, see also Figure 2 therein). Given the behaviour of the halo model and the HOD, the stochasticity of the galaxy bias could most strongly arise from the non-zero  $\sigma_c$  in equation (3.32) and the possible non-Poissonian nature of the satellite galaxy distribution for less massive galaxies. For more massive galaxies the main source of stochasticity can be shot noise, which dominates the stochasticity function,  $\sigma_b$  in equation (3.9), when the number density of galaxies is small. We use those free parameters of the HOD in a fit to the data (see Section 3.4), to constrain the cause for the stochasticity, non-linearity and scale dependence of galaxy bias.

### 3.3.3 PROJECTED FUNCTIONS

We can project the 3D bias functions as defined by Dekel & Lahav (1999); Cacciato et al. (2012) into two-dimensional, projected analogues, which are more easily accessible observationally. We start by defining the matter-matter, galaxy-matter, and galaxy-galaxy projected surface densities as:

$$\Sigma_{xy}(r_p) = 2\bar{\rho}_m \int_{r_p}^{\infty} \xi_{xy}(r) \frac{r dr}{\sqrt{r^2 - r_p^2}}, \quad (3.43)$$

where ‘x’ and ‘y’ stand either for ‘g’ or ‘m’, and  $r_p$  is the *projected* separation, with the change from standard line-of-sight integration to the integration along the projected separation using an Abel transformation. We also define  $\bar{\Sigma}_{xy}(< r_p)$  as its average inside  $r_p$ :

$$\bar{\Sigma}_{xy}(< r_p) = \frac{2}{r_p^2} \int_0^{r_p} \Sigma_{xy}(R') R' dR', \quad (3.44)$$

which we use to define the excess surface densities (ESD)

$$\Delta\Sigma_{xy}(r_p) = \bar{\Sigma}_{xy}(< r_p) - \Sigma_{xy}(r_p). \quad (3.45)$$

We include the contribution of the stellar mass of galaxies to the lensing signal as a point mass approximation, which we can write as:

$$\Delta\Sigma_{gm}^{\text{pm}}(r_p) = \frac{M_{\star,\text{med}}}{\pi r_p^2}, \quad (3.46)$$

where  $M_{\star,\text{med}}$  is the median stellar mass of the selected galaxies obtained directly from the GAMA catalogue (Taylor et al. 2011, see Section 3.4.1 and Table 3.1 for more details). This stellar mass contribution is fixed by each of our samples. According to the checks performed, the inclusion of the stellar mass contribution to the lensing signal does not affect our conclusions.

The obtained projected surface densities can subsequently be used to define the projected, 2D analogues of the 3D bias functions ( $b_g^{3D}$ ,  $\mathcal{R}_{gm}^{3D}$  and  $\Gamma_{gm}^{3D}$ , Dekel & Lahav 1999; Cacciato et al. 2012) as:

$$b_g(r_p) \equiv \sqrt{\frac{\Delta\Sigma_{gg}(r_p)}{\Delta\Sigma_{mm}(r_p)}}, \quad (3.47)$$

$$\mathcal{R}_{gm}(r_p) \equiv \frac{\Delta\Sigma_{gm}(r_p)}{\sqrt{\Delta\Sigma_{gg}(r_p)\Delta\Sigma_{mm}(r_p)}}, \quad (3.48)$$

and

$$\Gamma_{gm}(r_p) \equiv \frac{b_g(r_p)}{\mathcal{R}_{gm}(r_p)} = \frac{\Delta\Sigma_{gg}(r_p)}{\Delta\Sigma_{gm}(r_p)}. \quad (3.49)$$

In what follows we shall refer to these as the ‘projected bias functions’.

In the case of the galaxy-dark matter cross correlation, the excess surface density  $\Delta\Sigma_{gm}(r_p) = \gamma_t(r_p) \Sigma_{cr,com}$ , where  $\gamma_t(r_p)$  is the tangential shear, which can be measured observationally using galaxy-galaxy lensing, and  $\Sigma_{cr,com}$  is the comoving critical surface mass density:<sup>2</sup>

$$\Sigma_{cr,com} = \frac{c^2}{4\pi G(1+z_l)^2} \frac{D(z_s)}{D(z_l)D(z_l, z_s)}, \quad (3.50)$$

where  $D(z_l)$  is the angular diameter distance to the lens,  $D(z_l, z_s)$  is the angular diameter distance between the lens and the source and  $D(z_s)$  is the angular diameter distance to the source. In Appendix 3.C we discuss the exact derivation of equation (3.50) and the implications of using different coordinates. In the case of the galaxy-galaxy autocorrelation we can write that

$$\Delta\Sigma_{gg}(r_p) = \bar{\rho}_m \left[ \frac{2}{r_p^2} \int_0^{r_p} w_p(R') R' dR' - w_p(r_p) \right], \quad (3.51)$$

where  $w_p(r_p)$  is the projected galaxy correlation function, and  $w_p(r_p) = \Sigma_{gg}(r_p)/\bar{\rho}_m$ . It is immediately clear that  $\Delta\Sigma_{gg}(r_p)$  can be obtained from the projected correlation function  $w_p(r_p)$ , which is routinely measured in large galaxy redshift surveys.

In terms of the classical 3D bias functions  $b_g^{3D}$ ,  $\mathcal{R}_{gm}^{3D}$  and  $\Gamma_{gm}^{3D}$  (Cacciato et al. 2012), the galaxies can be unbiased with respect to the underlying dark matter distribution, if and only if the following conditions are true: they are not central galaxies, the occupation number of satellite galaxies obeys Poisson statistics ( $\beta = 1$ ), the normalised number density profile of satellite galaxies is identical to the one of the dark matter, and the occupational number of satellites is directly proportional to halo mass as  $\langle N_s \rangle = M\bar{n}_s/\rho$ . When central galaxies are added to the above conditions, one expects a strong scale dependence on small scales, due to the fact that central galaxies are strongly biased with respect to dark matter haloes. In the case of a non-Poissonian

<sup>2</sup>In Chapter 2, the same definition was used in all the calculations and plots shown, but erroneously documented in the paper. The equations (6) and (9) of that paper should have the same form as equations (3.50) and (3.54), as discussed in Appendix 3.C.

satellite distribution, one still expects  $b_g^{3D} = 1$  on large scales, but with a transition from 1 to  $\beta$ , roughly at the virial radius when moving towards the centre of the halo (see also Figure 3 in Cacciato et al. 2012). The same also holds for the case where the density profile of satellites follows that of dark matter (Cacciato et al. 2012).

Given all these reasons, as already pointed out by Cacciato et al. (2012), one expects scale independence on large scales (at a value dependent on halo model ingredients), with the transition to scale dependence on small scales (due to the effects of central galaxies) around the 1-halo to 2-halo transition. The same holds for the projected bias functions ( $b_g$ ,  $\mathcal{R}_{gm}$  and  $\Gamma_{gm}$ ), which also carry a wealth of information regarding the non-linearity and stochasticity of halo occupation statistics, and consequently, galaxy formation.

This is demonstrated in Figure 3.1 where we show the influence of different values of  $\sigma_c$ ,  $A_s$ ,  $\alpha_s$  and  $\beta$  on the bias function  $\Gamma_{gm}$  as a function of stellar mass. From the predictions one can clearly see how the different halo model ingredients influence the bias function. The halo model predicts, as mentioned before, scale independence above  $10 \text{ Mpc}/h$  and a significant scale dependence of galaxy bias on smaller scales, with the parameters  $\alpha_s$ ,  $A_s$  and  $\beta$  having a significant influence at those scales. Any deviation from a pure Poissonian distribution of satellite galaxies will result in quite a significant feature at intermediate scales, therefore it would be a likely explanation for detected signs of stochasticity [as the deviation from unity will drive the stochasticity function  $\sigma_b$  or alternatively  $\varepsilon$  away from 0, as can be seen from equations (3.38) to (3.42)]. In Figure 3.1 we also test the influence of having different  $\Omega_m$  and  $\sigma_8$  on the  $\Gamma_{gm}$  bias function, as generally, any bias function is a strong function of those two parameters (Dekel & Lahav 1999; Sheldon et al. 2004). We test this by picking 4 combinations of  $\Omega_m$  and  $\sigma_8$  drawn from the  $1\sigma$  confidence contours of Planck Collaboration et al. (2016) measurements of the two parameters. Given the uncertainties of those parameters and their negligible influence on the  $\Gamma_{gm}$  bias function, the decision to fix the cosmology seems to be justified.

We would like to remind the reader, that our implementation of the halo model does not include the scale dependence of the halo bias and the halo-exclusion (mutual exclusiveness of the spatial distribution of the haloes). Not including those effects can introduce errors on the 1-halo to 2-halo transition region that can be as large as 50% (Cacciato et al. 2012; van den Bosch et al. 2013). However, the bias functions as defined using equations (3.47) to (3.49) are much more accurate and less susceptible to the uncertainties in the halo model, by being defined as ratios of the two-point correlation functions (Cacciato et al. 2012).

Despite of this, we decided to estimate the halo model parameters and the nature of galaxy bias using the fit to the  $\Delta\Sigma_{gm}(r_p)$  and  $w_p(r_p)$  signals separately, rather than the ratio of the two (using the  $\Gamma_{gm}$  bias function directly). This approach will still suffer from a possible bias due to the fact that we do not include the scale dependent halo bias or the halo-exclusion in our model. This choice is motivated purely by the fact that the covariance matrix that would account for the cross-correlations between the lensing and clustering measurements cannot be properly taken into account when fitting the  $\Gamma_{gm}$  bias function directly. We investigate the possible bias in our results in Section 3.5.2.

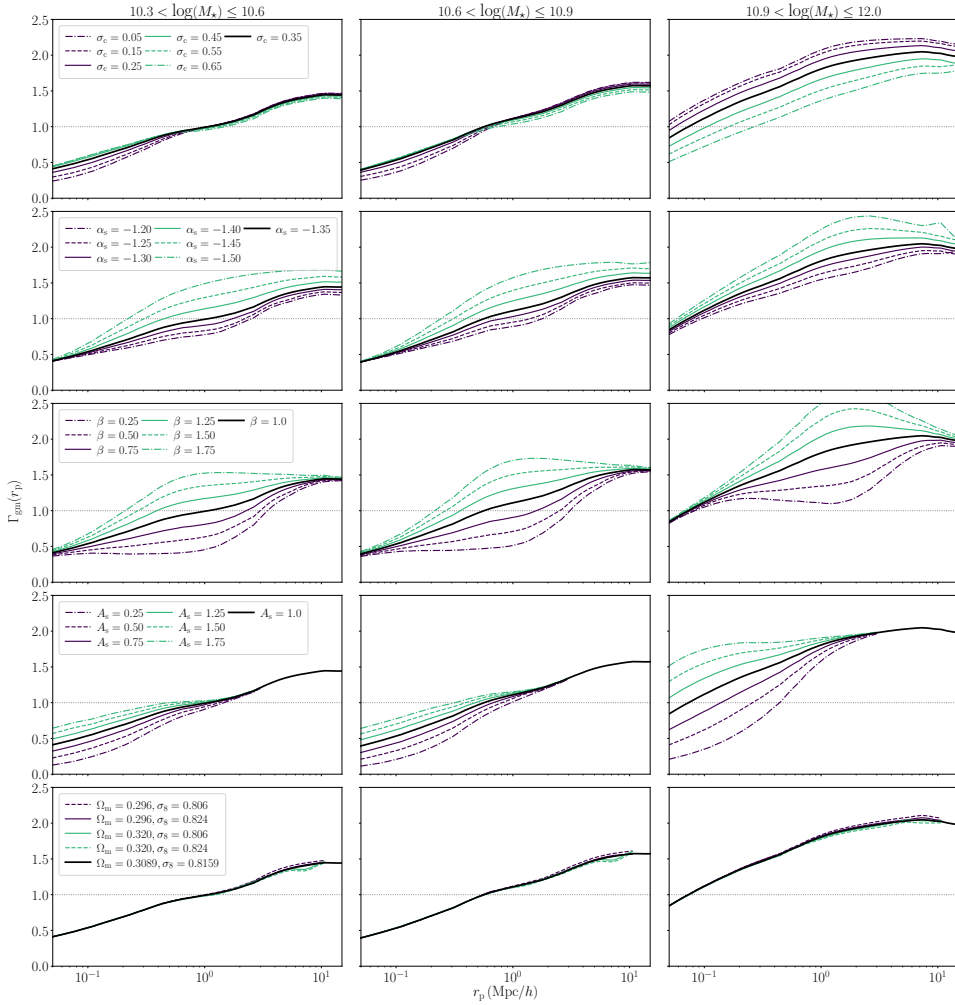


Figure 3.1: Model predictions of scale dependence of the galaxy bias function  $\Gamma_{\text{gm}}$  (equation 3.49) for three stellar mass bins (defined in Table 3.1), with stellar masses given in units of  $[\log(M_*/[M_\odot/h^2])]$ . With the black solid line we show our fiducial halo model (with other parameters adapted from Cacciato et al. 2013), and the different green and violet lines show different values of  $\sigma_c$ ,  $\alpha_s$ ,  $\beta$ ,  $A_s$  and combinations of  $\Omega_m$  and  $\sigma_8$ , row-wise, with values indicated in the legend. The full set of our fiducial parameters can be found in Table 3.2.

## 3.4 DATA AND SAMPLE SELECTION

### 3.4.1 LENS GALAXY SELECTION

The foreground galaxies used in this lensing analysis are taken from the Galaxy And Mass Assembly (hereafter GAMA) survey (Driver et al. 2011). GAMA is a spectroscopic survey carried out on the Anglo-Australian Telescope with the AAOmega spectrograph. Specifically, we use the information of GAMA galaxies from three equatorial regions, G9, G12 and G15 from GAMA II (Liske et al. 2015). We do not use the G02 and G23 regions, because the first one does not overlap with KiDS and the second one uses a different target selection compared to the one used in the equatorial regions. These equatorial regions encompass  $\sim 180 \text{ deg}^2$ , contain 180 960 galaxies (with  $nQ \geq 3$ , where the  $nQ$  is a measure of redshift quality) and are highly complete down to a Petrosian  $r$ -band magnitude  $r = 19.8$ . For the weak lensing measurements, we use all the galaxies in the three equatorial regions as potential lenses. To measure their average lensing and projected clustering signals, we group GAMA galaxies in stellar mass bins, following previous lensing measurements by van Uitert et al. (2016) and Velliscig et al. (2017). The bin ranges were chosen this way to achieve a good signal-to-noise ratio in all bins and to measure the galaxy bias as a function of different stellar mass. The selection of galaxies can be seen in Figure 5.1, and the properties we use in the halo model are shown in Table 3.1. Stellar masses are taken from version 19 of the stellar mass catalogue, an updated version of the catalogue created by Taylor et al. (2011), who fitted Bruzual & Charlot (2003) synthetic stellar population SEDs to the broadband SDSS photometry assuming a Chabrier (2003) IMF and a Calzetti et al. (2000) dust law. The stellar masses in Taylor et al. (2011) agree well with MagPhys derived estimates, as shown by Wright et al. (2017). Despite the differences in the range of filters, star formation histories, obscuration laws, the two estimates agree within 0.2 dex for 95 percent of the sample.

Table 3.1: Overview of the median stellar masses of galaxies, median redshifts and number of galaxies/lenses in each selected bin, which are indicated in the second column. Stellar masses are given in units of  $[\log(M_\star/[M_\odot/h^2])]$ .

Sample	Range	$M_{\star,\text{med}}$	$z_{\text{med}}$	# of lenses
Bin 1	(10.3, 10.6]	10.46	0.244	26224
Bin 2	(10.6, 10.9]	10.74	0.284	20452
Bin 3	(10.9, 12.0]	11.13	0.318	10178

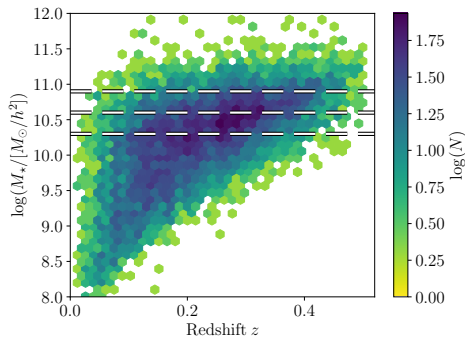


Figure 3.2: Stellar mass versus redshift of galaxies in the GAMA survey that overlap with KiDS. The full sample is shown with hexagonal density plot and the dashed lines show the cuts for the three stellar mass bins used in our analysis.

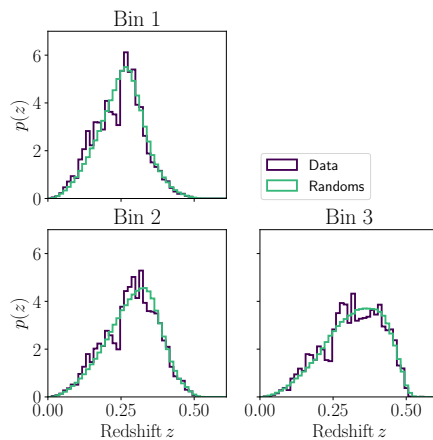


Figure 3.3: A comparison between the redshift distribution of galaxies in the data and the matched galaxies in GAMA random catalogue (Farrow et al. 2015) for our three stellar mass bins. We use the same set of randoms for both galaxy clustering and galaxy-galaxy lensing measurements.

### 3.4.2 MEASUREMENT OF THE $\Delta\Sigma_{\text{GM}}(r_p)$ SIGNAL

We use imaging data from 180 deg<sup>2</sup> of KiDS (Kuijken et al. 2015; de Jong et al. 2015) that overlaps with the GAMA survey (Driver et al. 2011) to obtain shape measurements of background galaxies. KiDS is a four-band imaging survey conducted with the OmegaCAM CCD mosaic camera mounted at the Cassegrain focus of the VLT Survey Telescope (VST); the camera and telescope combination provide us with a fairly uniform point spread function across the field-of-view.

We use shape measurements based on the  $r$ -band images, which have an average seeing of 0.66 arcsec. The image reduction, photometric redshift calibration and shape measurement analysis is described in detail in Hildebrandt et al. (2017).

We measure galaxy shapes using calibrated *lensfit* shape catalogs (Miller et al. 2013) (see also Fenech Conti et al. 2017, where the calibration methodology is described), which provides galaxy ellipticities ( $\epsilon_1$ ,  $\epsilon_2$ ) with respect to an equatorial coordinate system. For each source-lens pair we compute the tangential  $\epsilon_t$  and cross component  $\epsilon_x$  of the source's ellipticity around the position of the lens:

$$\begin{bmatrix} \epsilon_t \\ \epsilon_x \end{bmatrix} = \begin{bmatrix} -\cos(2\phi) & -\sin(2\phi) \\ \sin(2\phi) & -\cos(2\phi) \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \quad (3.52)$$

where  $\phi$  is the angle between the  $x$ -axis and the lens-source separation vector.

The azimuthal average of the tangential ellipticity of a large number of galaxies in the same area of the sky is an unbiased estimate of the shear. On the other hand, the azimuthal average of the cross ellipticity over many sources is unaffected by gravitational lensing and should average to zero (Schneider 2003). Therefore, the cross ellipticity is commonly used as an estimator of possible systematics in the measurements such as non-perfect PSF deconvolution, centroid bias and pixel level detector effects (Mandelbaum 2017). Each lens-source pair is then assigned a weight

$$\tilde{w}_{ls} = w_s \left( \tilde{\Sigma}_{cr,ls}^{-1} \right)^2, \quad (3.53)$$

which is the product of the *lensfit* weight  $w_s$  assigned to the given source ellipticity and the square of  $\tilde{\Sigma}_{cr,ls}^{-1}$  – the effective inverse critical surface mass density, which is a geometric term that downweights lens-source pairs that are close in redshift. We compute the effective inverse critical surface mass density for each lens using the spectroscopic redshift of the lens  $z_l$  and the full normalised redshift probability density of the sources,  $n(z_s)$ , calculated using the direct calibration method presented in Hildebrandt et al. (2017).

The effective inverse critical surface density can be written as:

$$\tilde{\Sigma}_{cr,ls}^{-1} = \frac{4\pi G}{c^2} (1 + z_l)^2 D(z_l) \int_{z_l}^{\infty} \frac{D(z_l, z_s)}{D(z_s)} n(z_s) dz_s. \quad (3.54)$$

The galaxy source sample is specific to each lens redshift with a minimum photometric redshift  $z_s = z_l + \delta_{z_l}$  with  $\delta_{z_l} = 0.2$ , where  $\delta_z$  is an offset to mitigate the effects of contamination from the group galaxies (for details see also the methods section and Appendix of **Chapter 2**). We determine the source redshift distribution  $n(z_s)$  for each sample, by applying the sample photometric redshift selection to a spectroscopic catalogue that has been weighted to reproduce the correct galaxy colour-distributions in KiDS (for details see Hildebrandt et al. 2017).

Thus, the ESD can be directly computed in bins of projected distance  $r_p$  to the lenses as:

$$\Delta \Sigma_{gm}(r_p) = \left[ \frac{\sum_{ls} \tilde{w}_{ls} \epsilon_{l,s} \Sigma'_{cr,ls}}{\sum_{ls} \tilde{w}_{ls}} \right] \frac{1}{1 + \bar{m}}. \quad (3.55)$$

where  $\Sigma'_{cr,ls} \equiv 1/\tilde{\Sigma}_{cr,ls}^{-1}$  and the sum is over all source-lens pairs in the distance bin, and

$$\bar{m} = \frac{\sum_i w'_i m_i}{\sum_i w'_i}, \quad (3.56)$$

is an average correction to the ESD profile that has to be applied to correct for the multiplicative bias  $m$  in the *lensfit* shear estimates. The sum goes over thin redshift slices for which  $m$  is obtained using the method presented in Fenech Conti et al. (2017), weighted by  $w' = w_s D(z_l, z_s)/D(z_s)$  for a given lens-source sample. The value of  $\bar{m}$  is around  $-0.014$ , independent of the scale at which it is computed. Furthermore, we subtract the signal around random points using the random catalogues from Farrow et al. (2015) (for details see analysis in the Appendix of **Chapter 2**).

### 3.4.3 MEASUREMENT OF THE $w_p(r_p)$ PROFILE

We compute the three-dimensional autocorrelation function of our three lens samples using the Landy & Szalay (1993) estimator. For this we use the same random catalogue and procedure as described in Farrow et al. (2015), applicable to the GAMA data. To minimise the effect of redshift-space distortions in our analysis, we project the three dimensional autocorrelation function along the line of sight:

$$w_p(r_p) = 2 \int_0^{\Pi_{\max}=100 \text{ Mpc}/h} \xi(r_p, \Pi) d\Pi. \quad (3.57)$$

For practical reasons, the above integral is evaluated numerically. This calls for consideration of our integration limits, particularly the choice of  $\Pi_{\max}$ . Theoretically one would like to integrate out to infinity in order to completely remove the effect of redshift space distortions and to encompass the full clustering signal on large scales. We settle for  $\Pi_{\max} = 100 \text{ Mpc}/h$ , in order to project the correlation function on the separations we are interested in (with a maximum  $r_p = 10 \text{ Mpc}/h$ ). We use the publicly available code `SWOT`<sup>3</sup> (Coupon et al. 2012) to compute  $\xi(r_p, \Pi)$  and  $w_p(r_p)$ , and to get bootstrap estimates of the covariance matrix on small scales. The code was tested against results from Farrow et al. (2015) using the same sample of galaxies and updated random catalogues (internal version 0.3), reproducing the results in detail. Randoms generated by Farrow et al. (2015) contain around 750 times more galaxies than those in GAMA samples. Figure 3.3 shows the good agreement between the redshift distributions of the GAMA galaxies and the random catalogues for the three stellar mass bins.

The clustering signal  $w_p(r_p)$  as well as the lensing signal  $\Delta\Sigma_{\text{gm}}(r_p)$  are shown in Figure 3.4, in the right and left panel, respectively. They are shown together with MCMC best-fit profiles as described in Section 3.4.5, using the halo model as described in Section 3.3. The best-fit is a single model used for all stellar masses and not independent for the three bins we are using. In order to obtain the galaxy bias function  $\Gamma_{\text{gm}}(r_p)$  (equation 3.49) we project the clustering signal according to the equation (3.57). The plot of this resulting function can be seen in Figure 3.5.

---

<sup>3</sup><http://jeancoupon.com/swot>



### 3.4.4 COVARIANCE MATRIX ESTIMATION

Statistical error estimates on the lensing signal and projected galaxy clustering signal are obtained using an analytical covariance matrix. As shown in **Chapter 2**, estimating the covariance matrix from data can become challenging given the small number of independent data patches in GAMA. This becomes even more challenging when one wants to include in the mixture the covariance for the projected galaxy clustering and all the possible cross terms between the two. The analytical covariance matrix we use is composed of three main parts: a Gaussian term, non-Gaussian term and the super-sample covariance (SSC) which accounts for all the modes outside of our KiDSxGAMA survey window. It is based on previous work by Takada & Jain (2009), Joachimi et al. (2008), Pielorz et al. (2010), Takada & Hu (2013), Li et al. (2014a), Marian et al. (2015), Singh et al. (2017) and Krause & Eifler (2017), and extended to support multiple lens bins and cross terms between lensing and projected galaxy clustering signals. The covariance matrix was tested against published results in these individual papers, as well as against real data estimates on small scales and mocks as used by van Uitert et al. (2018a). Further details and terms used can be found in Appendix 3.A. We first evaluate our covariance matrix for a set of fiducial model parameters and use this in our MCMC fit and then take the best-fit values and re-evaluate the covariance matrix for the new best-fit halo model parameters. After carrying out the re-fitting procedure, we find out that the updated covariance matrix and halo model parameters do not affect the results of our fit, and thus the original estimate of the covariance matrix is appropriate to use throughout the analysis.

### 3.4.5 FITTING PROCEDURE

The free parameters for our model are listed in Table 3.2, together with their fiducial values. We use a Bayesian inference method in order to obtain full posterior probabilities using a Monte Carlo Markov Chain (MCMC) technique; more specifically we use the emcee Python package (Foreman-Mackey et al. 2013). The likelihood  $\mathcal{L}$  is given by

$$\mathcal{L} \propto \exp \left[ -\frac{1}{2} (\mathbf{O}_i - \mathbf{M}_i)^T \mathbf{C}_{ij}^{-1} (\mathbf{O}_j - \mathbf{M}_j) \right], \quad (3.58)$$

where  $\mathbf{O}_i$  and  $\mathbf{M}_i$  are the measurements and model predictions in radial bin  $i$ , and  $\mathbf{C}_{ij}^{-1}$  is the element of the inverse covariance matrix that accounts for the correlation between radial bins  $i$  and  $j$ . In the fitting procedure we use the inverse covariance matrix as described in Section 3.4.4 and Appendix 3.A. We use wide flat priors for all the parameters (given in Table 3.2). The halo model (halo mass function and the power spectrum) is evaluated at the median redshift for each sample.

We run the sampler using 120 walkers, each with 12 000 steps (for a combined number of 14 400 000 samples), out of which we discard the first 1000 burn-in steps (120 000 samples). The resulting MCMC chains are well converged according to the integrated autocorrelation time test.

Table 3.2: Summary of the lensing results obtained using MCMC halo model fit to the data. Here  $M_0$  is the normalisation of the stellar to halo mass relation,  $M_1$  is the characteristic mass scale of the same stellar to halo mass relation,  $A_c$  is the normalisation of the concentration-mass relation,  $\sigma_c$  is the scatter between the stellar and halo mass,  $\gamma_1$  and  $\gamma_2$  are the low and high-mass slopes of the stellar to halo mass relation,  $A_s$  is the normalisation of the concentration-mass relation for satellite galaxies,  $\alpha_s$ ,  $b_0$  and  $b_1$  govern the behaviour of the CSMF of satellite galaxies, and  $\beta$  is the Poisson parameter. All parameters are defined in Section 3.3, using equations (3.29) to (3.42).

	$\log(M_0/[M_\odot/h^2])$	$\log(M_1/[M_\odot/h])$	$A_c$	$\sigma_c$	$\gamma_1$	$\gamma_2$
Fiducial	9.6	11.25	1.0	0.35	3.41	0.99
Priors	[7.0, 13.0]	[9.0, 14.0]	[0.0, 5.0]	[0.05, 2.0]	[0.0, 10.0]	[0.0, 10.0]
Posteriors	$8.75^{+1.62}_{-1.28}$	$11.13^{+1.10}_{-1.11}$	$1.33^{+0.20}_{-0.19}$	$0.25^{+0.24}_{-0.18}$	$2.16^{+4.43}_{-1.52}$	$1.32^{+0.51}_{-0.34}$
	$A_s$	$\alpha_s$	$b_0$	$b_1$	$\beta$	
Fiducial	1.0	-1.34	-1.15	0.59	1.0	
Priors	[0.0, 5.0]	[-5.0, 5.0]	[-5.0, 5.0]	[-5.0, 5.0]	[0.0, 2.0]	
Posteriors	$0.24^{+0.30}_{-0.14}$	$-1.36^{+0.19}_{-0.13}$	$-0.71^{+0.34}_{-0.55}$	$0.13^{+0.29}_{-0.30}$	$1.67^{+0.15}_{-0.16}$	

## 3.5 RESULTS

### 3.5.1 KIDS AND GAMA RESULTS

We fit the halo model as described in Section 3.4.5 to the measured projected galaxy clustering signal  $w_p(r_p)$  and the galaxy-galaxy lensing signal  $\Delta\Sigma_{\text{gm}}(r_p)$ , using the covariance matrix as described in Section 3.4.4. The resulting best fits are presented in Figure 3.4 (together with the measurements and their respective  $1\sigma$  errors obtained by taking the square root of the diagonal elements of the analytical covariance matrix). The measured halo model parameters, together with the  $1\sigma$  uncertainties are summarised in Table 3.2. Their full posterior distributions are shown in Figure 3.9. The fit of our halo model to both the galaxy-galaxy lensing signal and projected galaxy clustering signal, using the full covariance matrix accounting for all the possible cross-correlations, has a reduced  $\chi^2_{\text{red}}$  ( $\equiv \chi^2/\text{d.o.f.}$ ) equal to 1.15, which is an appropriate fit, given the 33 degrees of freedom (d.o.f.). We urge readers not to rely on the ‘‘chi-by-eye’’ in Figures 3.4 and 3.5 due to highly correlated data points (the correlations of which can be seen in Figure 3.8) and the joint fit of the halo model to the data.

Due to the fact that we are only using samples with relatively high stellar masses, we are unable to sample the low-mass portion of the stellar mass function, evident in our inability to properly constrain the  $\gamma_1$  parameter, which describes the behaviour of the stellar mass function at low halo mass. Mostly because of this, our results for the HOD parameters are different compared to those obtained by van Uitert et al. (2016), who analysed the full GAMA sample. There is also a possible difference arising due to the available overlap of KiDS and GAMA surveys used in van Uitert et al. (2016)

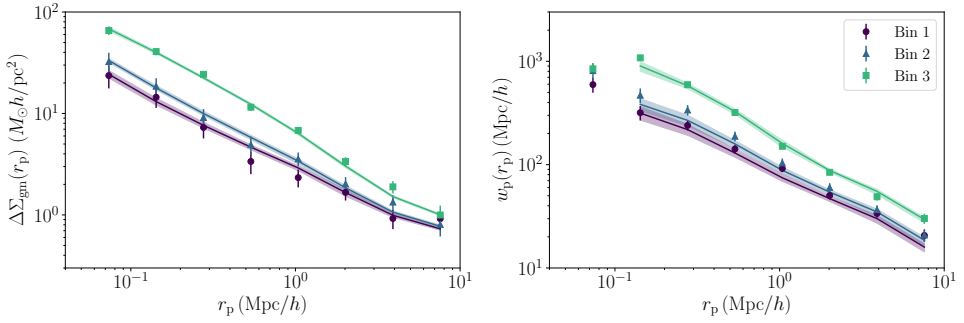


Figure 3.4: The stacked ESD profile (*left panel*) and projected galaxy clustering signal (*right panel*) of the 3 stellar mass bins in the GAMA galaxy sample defined in Table 3.1. The solid lines represent the best-fitting halo model as obtained using an MCMC fit, with the 68 percent confidence interval indicated with a shaded region. Using those two measurements we obtain the bias function  $\Gamma_{\text{gm}}(r_p)$ . We do not use the measurements in the grey band in our fit, as the clustering measurements are affected by blending in this region. The best-fit halo model parameters are listed in Table 3.2.

and our analysis, as van Uitert et al. (2016) used the lensing data from only  $100 \text{ deg}^2$  of the KiDS data, released before the shear catalogues used by Hildebrandt et al. (2017) and in **Chapter 2**, amongst others, became available. Our inferred HOD parameters are also in broad agreement with the ones obtained by Cacciato et al. (2014) for a sample of SDSS galaxies.

The main result of this work is the  $\Gamma_{\text{gm}}(r_p)$  bias function, presented in Figure 3.5, together with the best fit MCMC result – obtained by projecting the measured galaxy-galaxy clustering result according to equation (3.57) – and combining with the galaxy-galaxy lensing result according to equation (3.49). The obtained  $\Gamma_{\text{gm}}(r_p)$  bias function from the fit is scale dependent, showing a clear transition around  $2 \text{ Mpc}/h$ , in the 1-halo to 2-halo regime, where the function slowly transitions towards a constant value on even larger scales, beyond the range studied here (as predicted in Cacciato et al. 2012). Given the parameters obtained using the halo model fit to the data, the preferred value of  $\beta$  is larger than unity with  $\beta = 1.67^{+0.15}_{-0.16}$ , which indicates that the satellite galaxies follow a super-Poissonian distribution inside their host dark matter haloes, and are thus responsible for the deviations from constant in our  $\Gamma_{\text{gm}}(r_p)$  bias function at intermediate scales. Following the formulation by Cacciato et al. (2012), this also means that the galaxy bias, as measured, is highly non-deterministic. As seen by the predictions shown in Figure 3.1, the deviation of  $\beta$  from unity alone is not sufficient to explain the full observed scale dependence of the  $\Gamma_{\text{gm}}(r_p)$  bias function. Given the best-fit parameter values using the MCMC fit of the halo model, the non-unity of the mass-concentration relation normalisation  $A_s$  and other CSMF parameters (but most importantly the  $\alpha_s$  parameter, which governs the power law behaviour of the satellite CSMF) are also responsible for the total contribution to the observed scale dependence, and thus the stochastic behaviour of the galaxy bias on all scales observed.

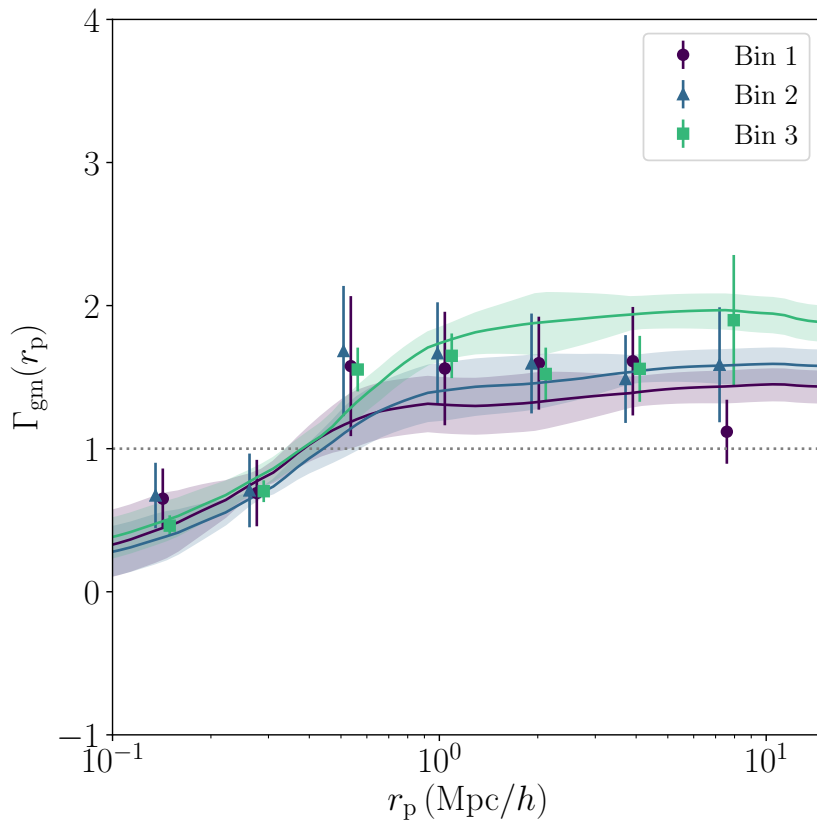


Figure 3.5: The  $\Gamma_{\text{gm}}(r_p)$  bias function as measured using a combination of projected galaxy clustering and galaxy-galaxy lensing signals, shown for the 3 stellar mass bins as used throughout this paper. The solid lines represent the best-fitting halo model as obtained using an MCMC fit to the projected galaxy clustering and galaxy-galaxy lensing signal, combined to obtain  $\Gamma_{\text{gm}}(r_p)$ , as described in Section 3.3. The colour bands show the 68 percent confidence interval propagated from the best-fit model. Error bars on the data are obtained by propagating the appropriate sub-diagonals of the covariance matrix and thus do not show the correct correlations between the data points and also overestimate the sample variance and super-sample covariance contributions.

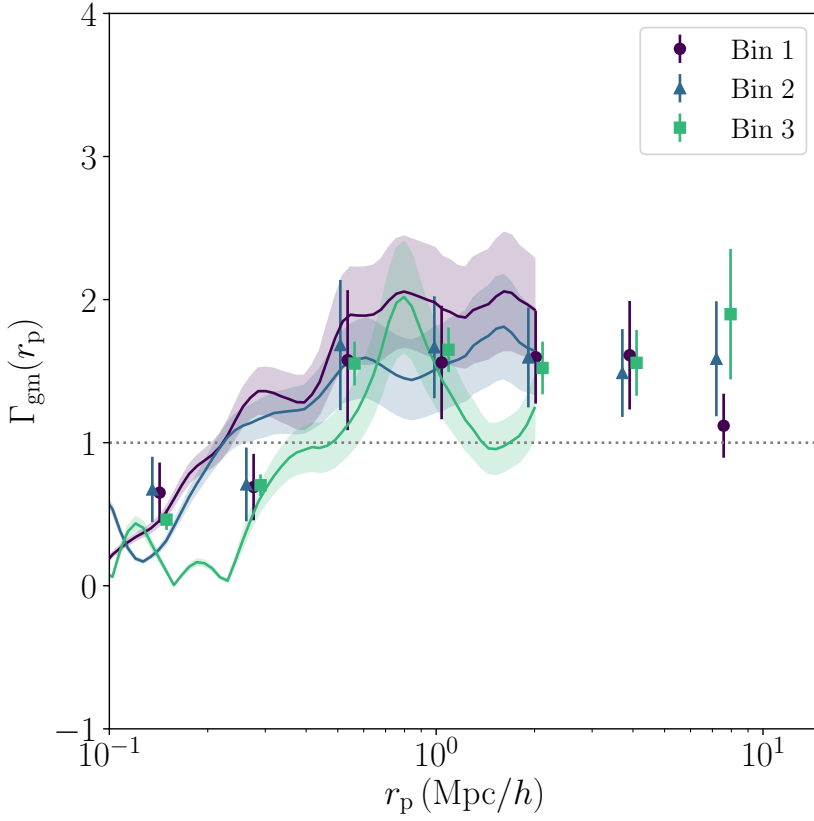


Figure 3.6: The  $\Gamma_{\text{gm}}(r_p)$  bias function as measured using the combination of projected galaxy clustering and galaxy-galaxy lensing signals, shown for the 3 stellar mass bins as used throughout this paper. The solid lines represent the same measurement repeated on the EAGLE simulation, with the colour bands showing the  $1\sigma$  errors. Note that those measurements are noisy due to the fact that the EAGLE simulation box is rather small, resulting in a relatively low number of galaxies in each bin (factor of around 26 lower, compared to the data). Due to the box size, we can also only show the measurement to about  $2 \text{ Mpc}/h$ .

### 3.5.2 INVESTIGATION OF THE POSSIBLE BIAS IN THE RESULTS

Due to the fact that we have decided to fit the model to the  $\Delta\Sigma_{\text{gm}}(r_p)$  and  $w_p(r_p)$  signals, we investigate how this choice might have biased our results. To check this we repeat our analysis using the  $\Gamma_{\text{gm}}(r_p)$  bias function directly. As our data vector we take the ratio of the projected signals as shown in Figure 3.5 and we use the appropriately propagated sub-diagonals of the covariance matrix as a rough estimate of the total covariance matrix. Such a covariance matrix does not show the correct correlations between the data points (and the bins) and also overestimates the sample variance and super-sample covariance contributions. Nevertheless the ratio of the diagonals as an estimate of the errors is somewhat representative of the errors on the measured  $\Gamma_{\text{gm}}(r_p)$  bias function. The fit procedure (except for a different data vector, covariance and output of the model) follows the method presented in Section 3.4.5. Using this, we obtain the best-fit values that are shown in Figure 3.9, marked with blue points and lines, together with the full posterior distributions from the initial fit. The resulting fit has a  $\chi^2_{\text{red}}$  equal to 1.29, with 9 degrees of freedom. As the results are consistent with the results that we obtain using a fit to the  $\Delta\Sigma_{\text{gm}}(r_p)$  and  $w_p(r_p)$  signals separately, it seems that, at least for this study, the halo model as described does not bias the overall conclusions of our analysis.

### 3.5.3 COMPARISON WITH EAGLE SIMULATION

In Figure 3.6 we compare our measurements of the GAMA and KiDS data to the same measurements made using the hydrodynamical EAGLE simulation (Schaye et al. 2015; McAlpine et al. 2016). EAGLE consists of state-of-the-art hydrodynamical simulations, including sub-grid interaction mechanisms between stellar and galactic energy sources. EAGLE is optimised such that the simulations reproduce a universe with the same stellar mass function as our own (Schaye et al. 2015). We follow the same procedure as with the data, by separately measuring the projected galaxy clustering signal and the galaxy-galaxy lensing signal and later combining the two accordingly. We measure the 3D galaxy clustering using the Landy & Szalay (1993) estimator, closely following the procedure outlined in Artale et al. (2017). We adopt the same  $\Pi_{\text{max}} = 34 \text{ Mpc}/h$  as used by Artale et al. (2017) in order to project the 3D galaxy clustering  $\xi(r_p, \Pi)$  to  $w_p(r_p)$ , which represents  $\sim L/2$  of the EAGLE box (Artale et al. 2017); see also equation (3.57). This limits the EAGLE measurements to a maximum scales of  $r_p < 2 \text{ Mpc}/h$ . As we do not require an accurate covariance matrix for the EAGLE results (we do not fit any model to it), we adopt a Jackknife covariance estimator using 8 equally sized sub-volumes. The measured EAGLE projected galaxy clustering signal is in good agreement with the GAMA measurements in detail, a result also found in Artale et al. (2017).

To estimate the galaxy-galaxy lensing signal of galaxies in EAGLE, we use the excess surface density (i.e., lensing signal) of galaxies in EAGLE calculated by Velliscig et al. (2017). We again select the galaxies in the three stellar mass bins, but in order to mimic the magnitude-limited sample we have adopted in our measurements of the galaxy-galaxy lensing signal on GAMA and KiDS, we have to weight our galaxies in the selection according to the satellite fraction as presented in Velliscig et al. (2017).

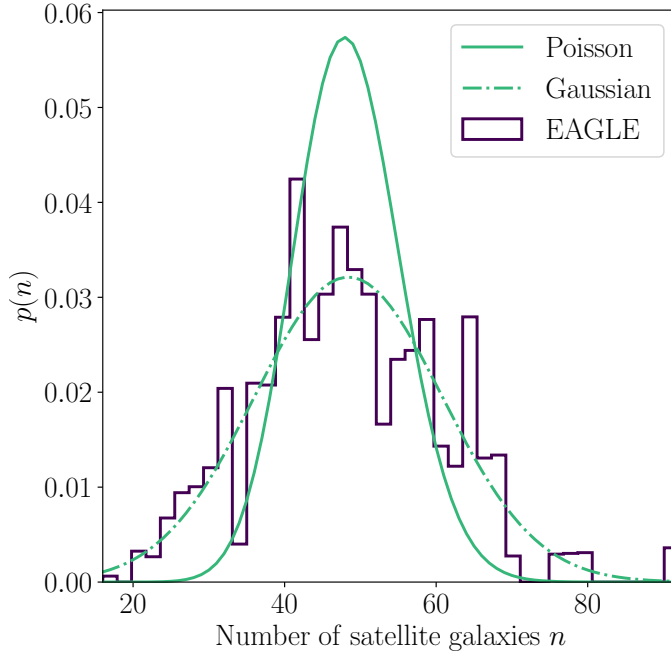


Figure 3.7: Distribution of satellite galaxies in a halo of fixed mass within  $12.0 < \log(M/M_\odot) < 12.2$  (histogram). This can be compared to a Poisson distribution with the same mean (solid curve) and a Gaussian distribution with the same mean and standard deviation as the data (dot-dashed curve).

Our two measurements (projected galaxy clustering and the galaxy-galaxy lensing) are then combined according to the definition of the  $\Gamma_{\text{gm}}(r_p)$  bias function, which is shown in Figure 3.6. There we directly compare the bias function as measured in the KiDS and GAMA data to the one obtained from the EAGLE hydrodynamical simulation (shown with full lines). The results from EAGLE are noisy, due to the fact that one is limited by the number of galaxies present in EAGLE.

Using the EAGLE simulations, we can directly access the properties of the satellite galaxies residing in the main halos present in the simulation. We select a narrow bin in halo masses of groups present in the simulation (between 12.0 and 12.2 in  $\log(M/M_\odot)$ ) and count the number of subhalos (galaxies). The resulting histogram, showing the relative abundance of satellite galaxies can be seen in Figure 3.7. We also show the Poisson distribution with the same mean as the EAGLE data, as well as the Gaussian distribution with the same mean and standard deviation as the distribution of the satellite galaxies in our sample. It can be immediately seen that the distribution of satellite galaxies at a fixed halo mass does not follow a Poisson distribution, and it is significantly wider (thus indeed being super-Poissonian).

The comparison nevertheless shows that the galaxy bias is intrinsically scale dependent and the shape of it suggests that it can be attributed to the non-Poissonian behaviour of satellite galaxies (and to lesser extent also to the precise distribution of satellites in the dark matter halo, governed by  $\alpha_s$  and  $A_s$  in the halo model).

### 3.6 DISCUSSION AND CONCLUSIONS

We have measured the projected galaxy clustering signal and galaxy-galaxy lensing signal for a sample of GAMA galaxies as a function of their stellar mass. In this analysis, we use the KiDS data covering  $180 \text{ deg}^2$  of the sky (Hildebrandt et al. 2017), that fully overlaps with the three equatorial patches from the GAMA survey that we use to determine three stellar mass selected lens galaxy samples. We have combined our results to obtain the  $\Gamma_{\text{gm}}(r_p)$  bias function in order to unveil the hidden factors and origin of galaxy biasing in light of halo occupation models and the halo model, as presented in the theoretical work of Cacciato et al. (2012). We have used that formalism to fit to the data to constrain the parameters that contribute to the observed scale dependence of the galaxy bias, and see which parameters exactly carry information about the stochasticity and non-linearity of the galaxy bias, as observed. Due to the limited area covered by the both surveys, the covariance matrix used in this analysis was estimated using an analytical prescription, for which details can be found in Appendix 3.A.

Our results show a clear trend that galaxy bias cannot be simply treated with a linear and/or deterministic approach. We find that the galaxy bias is inherently stochastic and non-linear due to the fact that satellite galaxies do not strictly follow a Poissonian distribution and that the spatial distribution of satellite galaxies also does not follow the NFW profile of the host dark matter halo. The main origin of the non-linearity of galaxy bias can be attributed to the fact that the central galaxy itself is heavily biased with respect to the dark matter halo in which it is residing. Those findings give additional support for the predictions presented by Cacciato et al. (2012), as their conclusions, based only on some fiducial model, are in line with our finding for a real subset of galaxies. We observe the same trends in the cosmological hydrodynamical simulation EAGLE, albeit out to smaller scales. We have also shown that the  $\Gamma_{\text{gm}}(r_p)$  bias function can, by itself, measure the properties of galaxy bias that would otherwise require the full knowledge of the  $b_g(r_p)$  and  $\mathcal{R}_{\text{gm}}(r_p)$  bias functions.

Our results are also in a broad agreement with recent findings of Gruen et al. (2018); Friedrich et al. (2018), who used the density split statistics to measure the cosmological parameters in SDSS (Rozo et al. 2015) and DES (Drlica-Wagner et al. 2018) data, and as a byproduct, also the  $b$  and  $r$  functions directly (at angular scales around  $20 \text{ arcmin}$ , which correspond to  $3.5 - 7 \text{ Mpc}/h$  at redshifts of  $0.2 - 4.5$ ). They find that the SDSS and DES data strongly prefer a stochastic bias with super-Poissonian behaviour. To obtain an independent measurement of galaxy bias and to further confirm our results we could use this method on our selection of galaxies, as well as the reconstruction method of Simon & Hilbert (2017). This work is, however, out of the scope of this paper.



Our findings show a remarkable wealth of information that halo occupation models are carrying in regard of understanding the nature of galaxy bias and its influence on cosmological analyses using the combination of galaxy-galaxy lensing and galaxy clustering. These results also show that the theoretical framework, as presented by Cacciato et al. (2012), is able to translate the constraints on galaxy biasing into constraints on galaxy formation and measurements of cosmological parameters. As an extension of this work, we could fold in the cosmic shear measurements of the same sample of galaxies, and thus constrain the galaxy bias and the sources of non-linearity and stochasticity further. This would allow a direct measurement of all three bias functions [ $\Gamma_{\text{gm}}(r_p)$ ,  $b_g(r_p)$  and  $\mathcal{R}_{\text{gm}}(r_p)$ ], which could then be used directly in cosmological analyses. On the other hand, for a more detailed study of the HOD beyond those parameters that influence the galaxy bias, we could include the stellar mass (or luminosity) function in the joint fit. We leave such exercises open for future studies.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for their very useful comments and suggestions. AD would like to thank Marcello Cacciato for all the useful discussions, support and the hand written notes provided on the finer aspects of the theory used in this paper.

KK acknowledges support by the Alexander von Humboldt Foundation. HHo acknowledges support from Vici grant 639.043.512, financed by the Netherlands Organisation for Scientific Research (NWO). This work is supported by the Deutsche Forschungsgemeinschaft in the framework of the TR33 ‘The Dark Universe’. CH acknowledges support from the European Research Council under grant number 647112. HHi is supported by an Emmy Noether grant (No. Hi 1495/2-1) of the Deutsche Forschungsgemeinschaft. AA is supported by a LSSTC Data Science Fellowship. RN acknowledges support from the German Federal Ministry for Economic Affairs and Energy (BMWi) provided via DLR under project no. 50QE1103.

This research is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017 and 177.A-3018, and on data products produced by Target/OmegaCEN, INAF-OACN, INAF-OAPD and the KiDS production team, on behalf of the KiDS consortium.

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 programs 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 <http://www.gama-survey.org>.

This work has made use of Python (<http://www.python.org>), including the packages `numpy` (<http://www.numpy.org>) and `scipy` (<http://www.scipy.org>). The halo model is built upon the `hmf` Python package by Murray et al. (2013). Plots have been produced with `matplotlib` (Hunter 2007) and `corner.py` (Foreman-Mackey 2016).

### 3.A ANALYTICAL COVARIANCE MATRIX

Here we present the expressions for the covariance of the auto-correlation and cross-correlation function of two observables, in our case specifically for the galaxy-galaxy, galaxy-mass auto-correlation functions and the cross-correlation function between the two. The expressions are an extension to the Gaussian part of the covariance as presented in Singh et al. (2017) and include the non-Gaussian terms and the super-sample covariance terms that are by themselves an extension (Krause & Eifler 2017) to the non-Gaussian terms as previously described for cosmic shear only by Takada & Hu (2013); Li et al. (2014a). We follow Singh et al. (2017); Marian et al. (2015) for the Gaussian terms, but excluding the additional contributions that arise due to not subtracting the signal obtained using random positions on the sky, as in our analysis this is performed during the signal extraction.

In general, the covariance matrix between two observables can be written as:

$$\text{Cov}(X, Y) = \text{Cov}^G(X, Y) + \text{Cov}^{\text{NG}}(X, Y) + \text{Cov}^{\text{SSC}}(X, Y), \quad (3.59)$$

where  $X$  and  $Y$  are either  $w_p(r_p)$  or  $\Delta\Sigma_{\text{gm}}^m(r_p)$ , and the  $G$  stands for the Gaussian term,  $\text{NG}$  for the non-Gaussian term and  $\text{SSC}$  stands for the contributions from the super-sample covariance. Furthermore, following Singh et al. (2017, starting with equation A18) and Marian et al. (2015, using the derivations in their Appendix), the Gaussian terms for each auto-correlation or cross-correlation can be written as (where indices  $i, j$  stand for individual projected radial bins and indices  $n, m$  stand for individual galaxy sample bins):

$$\text{Cov}^G[w_p^n, w_p^m](r_{p,i}, r_{p,j}) = 2 \frac{\mathcal{A}_{W2}(r_{p,i}, r_{p,j})}{\mathcal{A}_{W1}(r_{p,i}) \mathcal{A}_{W1}(r_{p,j})} \int \frac{dk k}{2\pi} J_0(kr_{p,i}) J_0(kr_{p,j}) \left( P_{\text{gg}}^n(k) + \delta_{nm} \frac{1}{n_g} \right) \left( P_{\text{gg}}^m(k) + \delta_{nm} \frac{1}{n_g} \right), \quad (3.60)$$

$$\begin{aligned} \text{Cov}^G[\Delta\Sigma_{\text{gm}}^n, \Delta\Sigma_{\text{gm}}^m](r_{p,i}, r_{p,j}) &= \bar{\rho}_m^2 \frac{\mathcal{A}_{W2}(r_{p,i}, r_{p,j})}{\mathcal{A}_{W1}(r_{p,i}) \mathcal{A}_{W1}(r_{p,j})} \int \frac{dk k}{2\pi} J_2(kr_{p,i}) J_2(kr_{p,j}) \left( P_{\text{gg}}^n(k) + \delta_{nm} \frac{1}{n_g} \right) \left( P_{\text{mm}}^n(k) + \delta_{nm} \frac{1}{n_\gamma} \right) \\ &+ \bar{\rho}_m^2 \frac{\mathcal{A}_{W2}(r_{p,i}, r_{p,j})}{\mathcal{A}_{W1}(r_{p,i}) \mathcal{A}_{W1}(r_{p,j})} \int \frac{dk k}{2\pi} J_2(kr_{p,i}) J_2(kr_{p,j}) P_{\text{gm}}^n(k) P_{\text{gm}}^m(k), \end{aligned} \quad (3.61)$$

$$\text{Cov}^G[w_p^n, \Delta\Sigma_{\text{gm}}^m](r_{p,i}, r_{p,j}) = \bar{\rho}_m \frac{\mathcal{A}_{W2}(r_{p,i}, r_{p,j})}{\mathcal{A}_{W1}(r_{p,i}) \mathcal{A}_{W1}(r_{p,j})} \int \frac{dk k}{2\pi} J_0(kr_{p,i}) J_2(kr_{p,j}) \left[ \left( P_{\text{gg}}^n(k) + \delta_{nm} \frac{1}{n_g} \right) P_{\text{gm}}^m + \left( P_{\text{gg}}^m(k) + \delta_{nm} \frac{1}{n_g} \right) P_{\text{gm}}^n \right], \quad (3.62)$$

where

$$\mathcal{A}_{W1}(r_p) = \int \frac{dk k}{2\pi} J_0(kr_p) \widetilde{W}^2(k) \quad (3.63)$$

and

$$\mathcal{A}_{W2}(r_{p,i}, r_{p,j}) = \int \frac{dk k}{2\pi} J_0(kr_{p,i}) J_0(kr_{p,j}) \widetilde{W}^2(k) \quad (3.64)$$

are the pre-factors arising from the survey geometry, with  $J_n$  being Bessel function of the  $n$ -th kind,  $\widetilde{W}(k)$  is the window function defined in equation (3.66),  $\delta_{nm}$  is the Kronecker delta symbol,  $\bar{n}_g$  the number density of galaxies in the bin  $n$ ,  $P_{\text{sy}}(k)$  are individual power spectra as defined in Section 3.3.1, and the  $\bar{n}_\gamma$  is the shape noise given by (see also Marian et al. 2015, equation C2):

$$\frac{1}{\bar{n}_\gamma} = \frac{\sigma_\gamma^2 \Sigma_\gamma^2}{\bar{n}_s} \frac{\Sigma_{\text{cr,com}}^2 D^2(z_l)}{\bar{\rho}_m^2 2\Pi_{\text{max}}}, \quad (3.65)$$

where  $\sigma_\gamma$  is the shape variance of the sources used in the analysis, the  $\bar{n}_s$  is the source density given by Hildebrandt et al. (2017) in gal/arcmin<sup>2</sup> (converted to radians), the  $D(z_l)$  is the angular diameter distance at  $z_l$  and  $\Pi_{\text{max}}$  is the projection length used throughout this work. The value  $\bar{n}_g$  and  $\sigma_\gamma$  are measured from the lens and source galaxies used in our 3 samples, respectively. We assume a circular survey geometry with a window function:

$$\widetilde{W}(k) = 2\pi R^2 \frac{J_1(kR)}{kR}, \quad (3.66)$$

where  $R$  is the radius of the circular window with area covering 180 deg<sup>2</sup>. The  $\Sigma_{\text{cr,com}}$  is calculated using the same prescription as defined in equation (3.54). To project our covariance matrices, we use the Limber approximation as demonstrated by Marian et al. (2015), using the slightly more accurate survey area normalisation by Singh et al. (2017).

The super-sample covariance terms are given by the following expressions:

$$\text{Cov}^{\text{SSC}}[w_p^n, w_p^m](r_{p,i}, r_{p,j}) = \frac{\mathcal{A}_{W2}(r_{p,i}, r_{p,j})}{\mathcal{A}_{W1}(r_{p,i}) \mathcal{A}_{W1}(r_{p,j})} \int \frac{dk k}{2\pi} J_0(kr_{p,i}) J_0(kr_{p,j}) \frac{\partial P_{\text{gg}}^m(k)}{\partial \delta_b} \frac{\partial P_{\text{gg}}^n(k)}{\partial \delta_b} \sigma_{\text{snm}}^2, \quad (3.67)$$

$$\text{Cov}^{\text{SSC}}[\Delta\Sigma_{\text{gm}}^n, \Delta\Sigma_{\text{gm}}^m](r_{p,i}, r_{p,j}) = \bar{\rho}_m^2 \frac{\mathcal{A}_{W2}(r_{p,i}, r_{p,j})}{\mathcal{A}_{W1}(r_{p,i}) \mathcal{A}_{W1}(r_{p,j})} \int \frac{dk k}{2\pi} J_2(kr_{p,i}) J_2(kr_{p,j}) \frac{\partial P_{\text{gm}}^n(k)}{\partial \delta_b} \frac{\partial P_{\text{gm}}^m(k)}{\partial \delta_b} \sigma_{\text{snm}}^2, \quad (3.68)$$

$$\text{Cov}^{\text{SSC}}[w_p^n, \Delta\Sigma_p^m](r_{p,i}, r_{p,j}) = \bar{\rho}_m \frac{\mathcal{A}_{W,2}(r_{p,i}, r_{p,j})}{\mathcal{A}_{W,1}(r_{p,i}) \mathcal{A}_{W,1}(r_{p,j})} \int \frac{dk k}{2\pi} J_0(kr_{p,i}) J_2(kr_{p,j}) \frac{\partial P_{\text{gg}}^n(k)}{\partial \delta_b} \frac{\partial P_{\text{gm}}^m(k)}{\partial \delta_b} \sigma_{\text{sum}}^2, \quad (3.69)$$

where the responses  $\frac{\partial P_{xy}(k)}{\partial \delta_b}$  are given by the following equations [following galaxy-matter and galaxy-galaxy extensions to the matter-matter only responses in Takada & Hu (2013); Li et al. (2014a) as derived by Krause & Eifler (2017)]:

$$\frac{\partial P_{\text{gg}}(k)}{\partial \delta_b} = \left( \frac{68}{21} - \frac{1}{3} \frac{d \ln k^3 P_{\text{lin}}(k)}{d \ln k} \right) I_{\text{g}}^1(k) I_{\text{g}}^1(k) P_{\text{lin}}(k) + I_{\text{gg}}^1(k, k) - 2b_{\text{g},1} P_{\text{gg}}(k), \quad (3.70)$$

$$\frac{\partial P_{\text{gm}}(k)}{\partial \delta_b} = \left( \frac{68}{21} - \frac{1}{3} \frac{d \ln k^3 P_{\text{lin}}(k)}{d \ln k} \right) I_{\text{g}}^1(k) I_{\text{m}}^1(k) P_{\text{lin}}(k) + I_{\text{gm}}^1(k, k) - b_{\text{g},1} P_{\text{gm}}(k), \quad (3.71)$$

$$\frac{\partial P_{\text{mm}}(k)}{\partial \delta_b} = \left( \frac{68}{21} - \frac{1}{3} \frac{d \ln k^3 P_{\text{lin}}(k)}{d \ln k} \right) I_{\text{m}}^1(k) I_{\text{m}}^1(k) P_{\text{lin}}(k) + I_{\text{mm}}^1(k, k), \quad (3.72)$$

where we have introduced the halo model integrals  $I_x^\alpha(k)$  and  $I_{xy}^\alpha(k, k')$  as:

$$I_x^\alpha(k) = \int \mathcal{H}_x(k, M) b_{i,\alpha} n(M) dM, \quad (3.73)$$

$$I_{xy}^\alpha(k, k') = \int \mathcal{H}_x(k, M) \mathcal{H}_y(k', M) b_{i,\alpha} n(M) dM. \quad (3.74)$$

In the case when 'x' equals 'g',  $\mathcal{H}_g(k, M)$  is the sum of the  $\mathcal{H}_c(k, M)$  and  $\mathcal{H}_s(k, M)$  functions. Bias functions  $b_{i,\alpha}$  are either 0 if  $\alpha = 0$ , 1 if  $i = m$  and  $\alpha = 1$ , and  $\int dM \mathcal{H}_g(k, M) b_{\text{h}}(M) n(M) / \bar{n}_g$  if  $i = g$  and  $\alpha = 1$ . Note that these subscripts are not related to the ones used in Section 3.3.1. In the equations (3.67), (3.68) and (3.69), we have also used the survey variance defined as:

$$\sigma_s^2 = \frac{1}{\mathcal{A}_{W,1}(R)} \int \frac{dk k^2}{2\pi^2} \tilde{W}^2(k) P_{\text{lin}}(k). \quad (3.75)$$

The connected, non-Gaussian terms of the covariance matrix can be written as (again following the extension to the matter-matter only derivation by Krause & Eifler 2017) (see also Cooray & Sheth 2002; Takada & Jain 2009):

$$\text{Cov}^{\text{NG}}[w_p^n, w_p^m](r_{p,i}, r_{p,j}) = \frac{\mathcal{A}_{W,2}(r_{p,i}, r_{p,j})}{\mathcal{A}_{W,1}(r_{p,i}) \mathcal{A}_{W,1}(r_{p,j})} \int \frac{dk_i k_j}{\sqrt{2\pi}} \int \frac{dk_j k_i}{\sqrt{2\pi}} J_0(k_i r_{p,i}) J_0(k_j r_{p,j}) \cdot T_{\text{gggg}}^{nm}(k_i, -k_i, k_j, -k_j), \quad (3.76)$$

$$\text{Cov}^{\text{NG}}[\Delta\Sigma_{\text{gm}}^n, \Delta\Sigma_{\text{gm}}^m](r_{\text{p},i}, r_{\text{p},j}) = \bar{\rho}_{\text{m}}^2 \frac{\mathcal{A}_{\text{W}2}(r_{\text{p},i}, r_{\text{p},j})}{\mathcal{A}_{\text{W}1}(r_{\text{p},i})\mathcal{A}_{\text{W}1}(r_{\text{p},j})} \int \frac{dk_i k_j}{\sqrt{2\pi}} \int \frac{dk_i k_j}{\sqrt{2\pi}} J_2(k_i r_{\text{p},i}) J_2(k_j r_{\text{p},j}) \cdot T_{\text{gm gm}}^{\text{nm}}(k_i, -k_i, k_j, -k_j), \quad (3.77)$$

$$\text{Cov}^{\text{NG}}[\mathcal{W}_{\text{p}}^n, \Delta\Sigma_{\text{p}}^m](r_{\text{p},i}, r_{\text{p},j}) = \bar{\rho}_{\text{m}} \frac{\mathcal{A}_{\text{W}2}(r_{\text{p},i}, r_{\text{p},j})}{\mathcal{A}_{\text{W}1}(r_{\text{p},i})\mathcal{A}_{\text{W}1}(r_{\text{p},j})} \int \frac{dk_i k_j}{\sqrt{2\pi}} \int \frac{dk_i k_j}{\sqrt{2\pi}} J_0(k_i r_{\text{p},i}) J_2(k_j r_{\text{p},j}) \cdot T_{\text{gg gm}}^{\text{nm}}(k_i, -k_i, k_j, -k_j), \quad (3.78)$$

where the individual  $T_{\text{xyzw}}$  terms are given by the combination of 1-halo matter trispectrum and (2 + 3 + 4)-halo matter trispectrum as:

$$T_{\text{xyzw}}^{\text{nm}}(k_i, -k_i, k_j, -k_j) = (b_x b_y b_z b_w)^{\text{nm}} T_{\text{m}}^{2\text{h}+3\text{h}+4\text{h}}(k_i, -k_i, k_j, -k_j) + T_{\text{xyzw}}^{\text{lh,nm}}(k_i, k_i, k_j, k_j), \quad (3.79)$$

where  $b_x$  is the same bias used in equations (3.73) and (3.74). The 1-halo matter trispectrum is given by the following integral over halo model building blocks:

$$T_{\text{xyzw}}^{\text{lh,nm}}(k_i, k_i, k_j, k_j) = \int dM n(M) \mathcal{H}_x^{\text{n}}(k_i, M) \mathcal{H}_y^{\text{n}}(k_i, M) \mathcal{H}_z^{\text{m}}(k_j, M) \mathcal{H}_w^{\text{m}}(k_j, M), \quad (3.80)$$

and the (2 + 3 + 4)-halo matter trispectrum  $T_{\text{m}}^{2\text{h}+3\text{h}+4\text{h}}$  is calculated according to the estimate presented in Pielorz et al. (2010). Additionally, the resulting covariance matrix is bin averaged according to the same radial binning scheme as used with our data. The resulting covariance matrix obtained using the analytical prescription presented here can be seen in Figure 3.8 (shown as a correlation matrix), which shows the auto-correlation between the 3 bins for the lensing signal, clustering signal and the cross-correlation between the two observables in the off-diagonal  $3 \times 3$  blocks. Individual combinations between all the bins are marked above the corresponding block matrices.

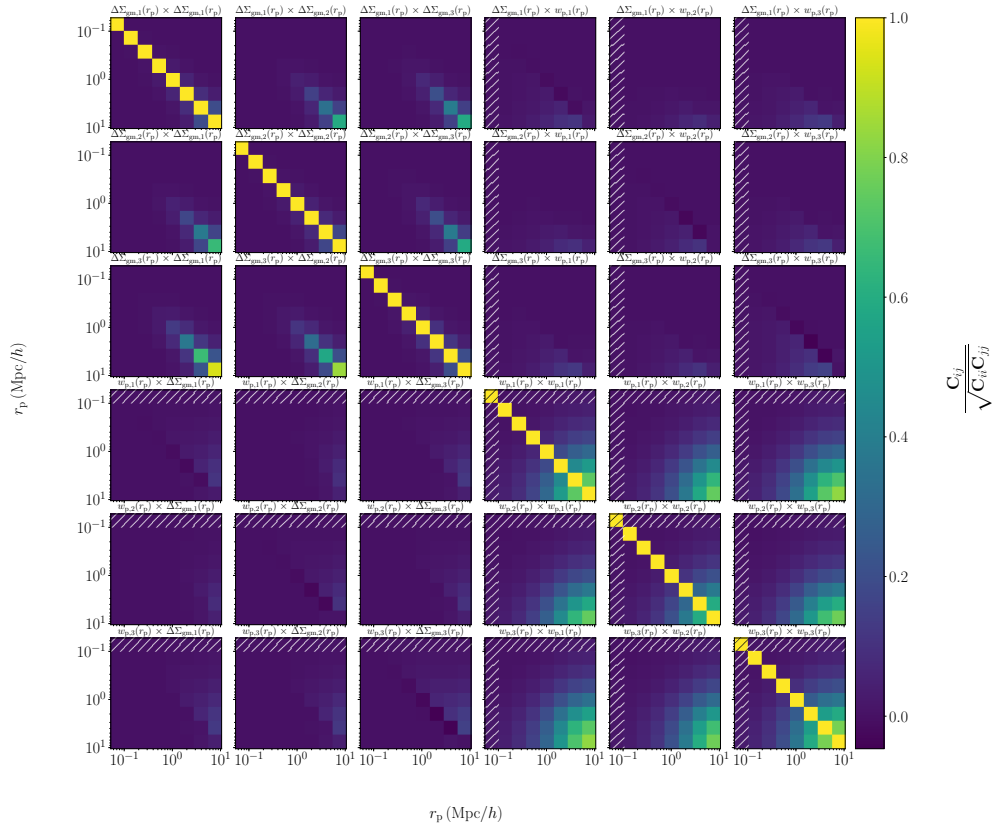


Figure 3.8: The full analytical correlation matrix for the lensing and clustering signals and their cross terms. Individual combinations between all the bins are marked above the corresponding block matrices, with indices 1,2 and 3 corresponding to the stellar mass bins as defined in Table 3.1. We do not use the covariance estimates in the hatched areas in our fit, as the clustering measurements are affected by the blending on these scales.

### 3.B FULL POSTERIOR DISTRIBUTIONS

In Figure 3.9 we show the full posterior probability distribution for all fitted parameters in our MCMC fit as discussed in Section 3.4.

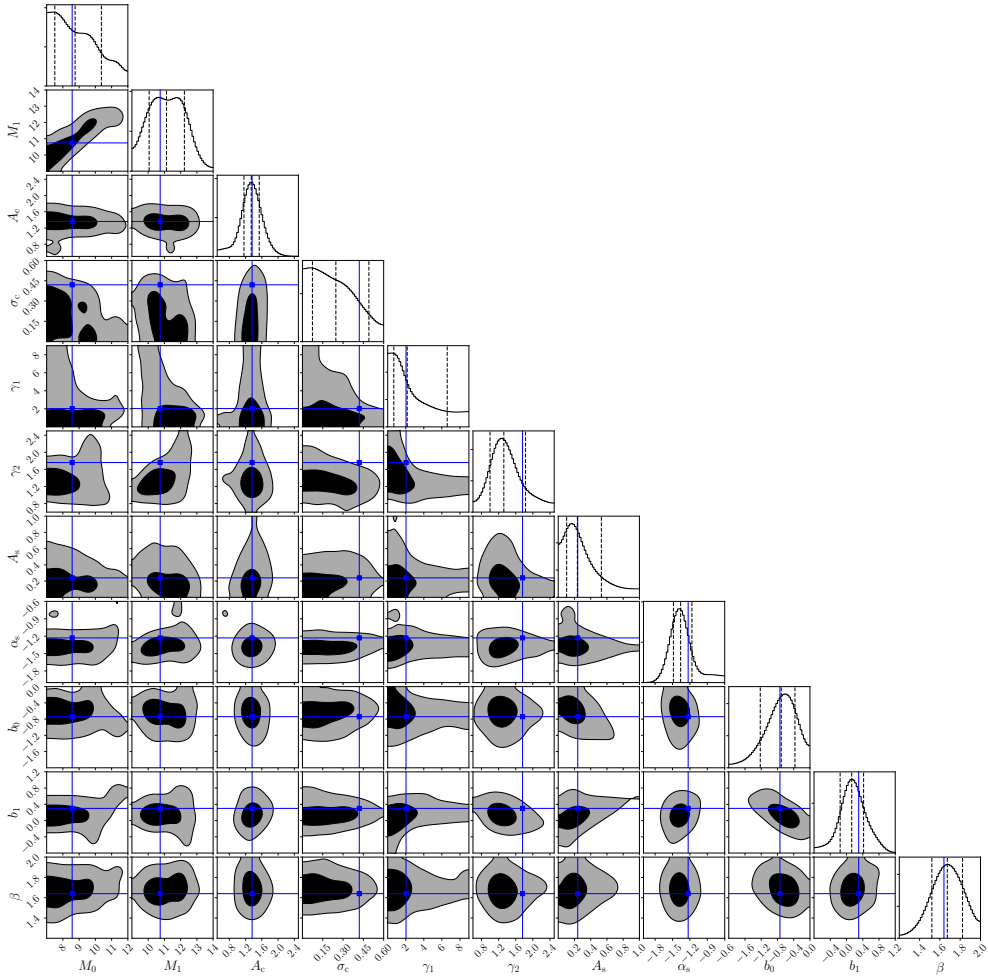


Figure 3.9: The full posterior distributions of the halo model parameters (where the priors are listed in Table 3.2). The contours indicate  $1\sigma$  and  $2\sigma$  confidence regions.

### 3.C RELATION BETWEEN THE LENSING SIGNAL AND THE GALAXY-MATTER CROSS-CORRELATION FUNCTION

In this appendix, we provide a step-by-step derivation of the relation between the galaxy-galaxy lensing signal and the galaxy-matter cross-correlation function. As a side-product, we motivate the two different definitions of the critical surface mass density that are used in this field. Finally, we compare our results with those in some recent papers, pointing out differences, and discussing their implications. Since the results of this appendix apply to several papers, we choose to use a slightly more explicit notation here in comparison to the rest of this paper.

#### 3.C.1 DERIVATION

The equivalent weak lensing convergence  $\kappa$  for a three-dimensional mass distribution characterized by the fractional density contrast  $\delta$ , for sources at comoving distance  $\chi_s$ , is given by (e.g., Bartelmann & Schneider 2001; Schneider 2006)

$$\kappa(\boldsymbol{\theta}) = \frac{3H_0^2\Omega_m}{2c^2} \int_0^{\chi_s} d\chi \frac{\chi(\chi_s - \chi)}{\chi_s a(\chi)} \delta(\chi\boldsymbol{\theta}, \chi) = \bar{\rho}_m \frac{4\pi G}{c^2} \int_0^{\chi_s} d\chi \frac{\chi(\chi_s - \chi)}{\chi_s a(\chi)} \delta(\chi\boldsymbol{\theta}, \chi), \quad (3.81)$$

where we assumed for notational simplicity a spatially flat cosmological model. Here,  $\bar{\rho}_m$  is the current mean matter density in the Universe, and we used the relation between mass density and density parameter in the second step, i.e.,  $\bar{\rho}_m = 3H_0^2\Omega_m/(8\pi G)$ . The relation (3.81) is valid in the framework of the Born approximation and by neglecting lens-lens coupling (see, e.g. Hilbert et al. 2009; Krause & Hirata 2010, for the impact of these effects).

Let  $\delta_g$  be the three-dimensional fractional density contrast of galaxies of a given type. Their fractional density contrast on the sky,  $\kappa_g(\boldsymbol{\theta}) = [n(\boldsymbol{\theta}) - \bar{n}]/\bar{n}$ , with  $\bar{n}$  being the mean number density, is related to  $\delta_g$  by

$$\kappa_g(\boldsymbol{\theta}) = \int d\chi p_f(\chi) \delta_g(\chi\boldsymbol{\theta}, \chi), \quad (3.82)$$

where  $p_f(\chi)$  is the probability distribution of the selected ‘foreground’ galaxy population in comoving distance, equivalent to a redshift probability distribution. For the following we will assume that this distribution is a very narrow one around redshift  $z_l$ , and thus approximate  $p_f(\chi) = \delta_D(\chi - \chi_l)$ . We assume throughout that  $\chi_l < \chi_s$ . The correlator between  $\kappa$  and  $\kappa_g$  then becomes

$$\begin{aligned} \langle \kappa_g(\boldsymbol{\theta}) \kappa(\boldsymbol{\theta} + \boldsymbol{\vartheta}) \rangle &= \bar{\rho}_m \frac{4\pi G}{c^2} \int_0^{\chi_s} d\chi \frac{\chi(\chi_s - \chi)}{\chi_s a(\chi)} \int d\chi' \delta_D(\chi' - \chi_l) \langle \delta_g(\chi'\boldsymbol{\theta}, \chi') \delta[\chi(\boldsymbol{\theta} + \boldsymbol{\vartheta}), \chi] \rangle \\ &= \bar{\rho}_m \frac{4\pi G}{c^2} \int_0^{\chi_s} d\chi \frac{\chi(\chi_s - \chi)}{\chi_s a(\chi)} \langle \delta_g(\chi_l\boldsymbol{\theta}, \chi_l) \delta[\chi(\boldsymbol{\theta} + \boldsymbol{\vartheta}), \chi] \rangle. \end{aligned} \quad (3.83)$$



Since the correlator is significantly non-zero only over a small interval in  $\chi$  around  $\chi_1$ , the prefactor in the integrand can be considered to be constant over this interval and taken out of the integral. This yields

$$\begin{aligned} \langle \kappa_{\mathbf{g}}(\boldsymbol{\theta}) \kappa(\boldsymbol{\theta} + \boldsymbol{\vartheta}) \rangle &= \frac{4\pi G}{c^2} \frac{\chi_1 (\chi_s - \chi_1)}{\chi_s a(\chi_1)} \bar{\rho}_m \int_0^{\chi_s} d\chi \xi_{\text{gm}} \left( \sqrt{\chi_1^2 |\boldsymbol{\vartheta}|^2 + (\chi - \chi_1)^2} \right) \\ &= \Sigma_{\text{cr,com}}^{-1} \Sigma_{\text{com}}(\chi_1 |\boldsymbol{\theta}|) , \end{aligned} \quad (3.84)$$

where the galaxy-matter cross-correlation function  $\xi_{\text{gm}}$  (at fixed redshift  $z_1$ ) is defined through

$$\langle \delta(\mathbf{x}) \delta_{\mathbf{g}}(\mathbf{x} + \mathbf{y}) \rangle = \xi_{\text{gm}}(|\mathbf{y}|) , \quad (3.85)$$

in which  $\mathbf{x}$  and  $\mathbf{y}$  are comoving spatial vectors, and the sole dependence on  $|\mathbf{y}|$  is due to the assumed homogeneity and isotropy of the density fields in the Universe. Furthermore, we have defined the comoving critical surface mass density  $\Sigma_{\text{cr,com}}$  through

$$\Sigma_{\text{cr,com}}^{-1} = \frac{4\pi G}{c^2} \frac{\chi_1 (\chi_s - \chi_1)}{\chi_s a(\chi_1)} H(\chi_s - \chi_1) , \quad (3.86)$$

with  $H(x)$  being the Heaviside unit step function,<sup>4</sup> and the comoving surface mass density as

$$\Sigma_{\text{com}}(R_{\text{com}}) = \bar{\rho}_m \int_0^{\chi_s} d\chi \xi_{\text{gm}} \left( \sqrt{R_{\text{com}}^2 + (\chi - \chi_1)^2} \right) , \quad (3.87)$$

as a function of the comoving projected separation  $R_{\text{com}}$ . In this paper,  $\Sigma_{\text{cr,com}}$  is termed  $\Sigma_{\text{crit}}$  – see equation (3.50), and  $R_{\text{com}}$  and  $\Sigma_{\text{com}}$  are called  $r_p$  and  $\Sigma_{\text{gm}}$  – see equation (3.43).

The interpretation of equation (3.84) is then that  $\bar{\rho}_m \xi_{\text{gm}}$  is the average comoving overdensity of matter around galaxies, caused by the correlation between them, and that the integral over comoving distance then yields the comoving surface mass density of this excess matter. The corresponding convergence is then obtained by scaling with the comoving critical surface mass density  $\Sigma_{\text{cr,com}}$ .

There is another form in which equation (3.84) can be written by rearranging factors of  $a(\chi_1)$ , namely

$$\begin{aligned} \langle \kappa_{\mathbf{g}}(\boldsymbol{\theta}) \kappa(\boldsymbol{\theta} + \boldsymbol{\vartheta}) \rangle &= \frac{4\pi G}{c^2} \frac{\chi_1 (\chi_s - \chi_1) a(\chi_1)}{\chi_s} \bar{\rho}_m a^{-2}(\chi_1) \int_0^{\chi_s} d\chi \xi_{\text{gm}} \left( \sqrt{\chi_1^2 |\boldsymbol{\vartheta}|^2 + (\chi - \chi_1)^2} \right) \\ &=: \Sigma_{\text{cr}}^{-1} \Sigma(\chi_1 |\boldsymbol{\theta}|) , \end{aligned} \quad (3.88)$$

where we defined the (proper) critical surface mass density  $\Sigma_{\text{cr}}$  through

$$\Sigma_{\text{cr}}^{-1} = \frac{4\pi G}{c^2} \frac{\chi_1 (\chi_s - \chi_1) a(\chi_1)}{\chi_s} H(\chi_s - \chi_d) = \frac{4\pi G}{c^2} \frac{D_1 D_{\text{ls}}}{D_s} , \quad (3.89)$$

<sup>4</sup>The corresponding expression for a general curvature parameter reads

$$\Sigma_{\text{cr,com}}^{-1} = \frac{4\pi G}{c^2} \frac{f_K(\chi_1) f_K(\chi_s - \chi_1)}{f_K(\chi_s) a(\chi_1)} H(\chi_s - \chi_1) ,$$

where  $f_K(\chi)$  is the comoving angular-diameter distance to a comoving distance  $\chi$ , and either the identity for spatially flat models, or a sin or sinh function for closed or open models, respectively.

and in the last step we introduced the angular-diameter distances  $D_l = D(0, z_l)$ ,  $D_s = D(0, z_s)$  and  $D_{ds} = D(z_l, z_s)$ , with  $D(z_1, z_2) = a(z_2) [\chi(z_2) - \chi(z_1)] H(z_2 - z_1)$  being the angular-diameter distance of a source at redshift  $z_2$  as seen from an observer at redshift  $z_1$ .<sup>5</sup> Furthermore,

$$\Sigma(R_{\text{com}}) = \rho_m(\chi_l) \int_0^{\chi_s} d\chi a(\chi) \xi_{\text{gm}} \left( \sqrt{R_{\text{com}}^2 + (\chi - \chi_l)^2} \right). \quad (3.90)$$

We note that, due to the assumed localized nature of the correlation function, we could write  $a(\chi)$  into the integrand in equation (3.90). The interpretation of equation (3.88) is now that the (proper) overdensity around galaxies caused by the galaxy-matter cross-correlation,  $\rho_m(\chi_l) \xi_{\text{gm}} = \bar{\rho}_m (1 + z_l)^3 \xi_{\text{gm}}$ , is integrated along the l.o.s. in proper coordinates,  $dr_{\text{prop}} = a d\chi$ , and the resulting (proper) surface mass density is scaled by the critical surface mass density  $\Sigma_{\text{cr}}$ . We note that the argument of the (proper) surface mass density  $\Sigma$  is a comoving transverse separation, since the correlation function is a function of comoving separation.

The relation between the two different equations (3.86) and (3.89) of the critical surface mass density is

$$\Sigma_{\text{cr,com}} = a^2(\chi_l) \Sigma_{\text{cr}}, \quad (3.91)$$

so that the comoving critical surface density is smaller by a factor  $a^2(\chi_l)$ . This makes sense: for a given lens, the comoving surface mass density (mass per unit comoving area) is smaller than the proper surface mass density,

$$\Sigma_{\text{com}}(R_{\text{com}}) = a^2(\chi_l) \Sigma(R_{\text{com}}), \quad (3.92)$$

since the comoving area is larger than the proper one by a factor  $a^{-2}$ . Correspondingly, since the convergence, or the correlation function in equation (3.84), is independent of whether proper or comoving measures are used, the comoving critical surface mass density is smaller by the same factor.

If  $N$  lensing galaxies at redshift  $z_l$  are located at positions  $\theta_i$  within a solid angle  $\omega$ , the corresponding fractional number density contrast reads

$$\kappa_{\text{g}}(\boldsymbol{\theta}) = \frac{1}{\bar{n}} \sum_{i=1}^N \delta_{\text{D}}(\boldsymbol{\theta} - \boldsymbol{\theta}_i) - 1, \quad (3.93)$$

where for large  $N$  and  $\omega$ ,  $\bar{n} = N/\omega$ . To evaluate the correlator of equation (3.83) in this case, we replace the ensemble average with an angular average, as is necessarily done in any practical estimation,

$$\langle \kappa_{\text{g}}(\boldsymbol{\theta}) \kappa(\boldsymbol{\theta} + \boldsymbol{\vartheta}) \rangle \approx \frac{1}{\omega} \int_{\omega} d^2\theta \kappa_{\text{g}}(\boldsymbol{\theta}) \kappa(\boldsymbol{\theta} + \boldsymbol{\vartheta}) = \frac{1}{\omega} \int_{\omega} d^2\theta \left[ \frac{1}{\bar{n}} \sum_{i=1}^N \delta_{\text{D}}(\boldsymbol{\theta} - \boldsymbol{\theta}_i) \right] \kappa(\boldsymbol{\theta} + \boldsymbol{\vartheta}) = \frac{1}{N} \sum_{i=1}^N \kappa(\boldsymbol{\theta}_i + \boldsymbol{\vartheta}), \quad (3.94)$$

valid for separations  $\vartheta$  which are much smaller than the linear angular extent  $\sqrt{\omega}$  of the region (to neglect boundary effects), and we employed the fact that the ensemble

<sup>5</sup>For a model with free curvature,  $D(z_1, z_2) = a(z_2) [f_K [\chi(z_2) - \chi(z_1)]] H(z_2 - z_1)$ .

average – and in the same approximation as above, the angular average – of  $\kappa(\boldsymbol{\theta})$  vanishes. We thus see that the correlator  $\langle \kappa_g \kappa \rangle$  can be obtained from the average convergence around the foreground galaxies, a quantity probed by the shear. Thus we find the relations

$$\gamma_t(\theta) = \Sigma_{\text{cr}}^{-1} \Delta\Sigma(\chi_1\theta) = \Sigma_{\text{cr,com}}^{-1} \Delta\Sigma_{\text{com}}(\chi_1\theta), \quad (3.95)$$

where

$$\Delta\Sigma(R_{\text{com}}) = \frac{2}{R_{\text{com}}^2} \int_0^{R_{\text{com}}} dR R \Sigma(R) - \Sigma(R_{\text{com}}), \quad (3.96)$$

and the analogous definition for  $\Delta\Sigma_{\text{com}}$ .

A further subtlety and potential source of confusion is that frequently, the surface mass density  $\Sigma$  is considered a function of *proper* transverse separation  $R = a(\chi_1)R_{\text{com}}$ . For the purpose of this appendix, we call this function  $\Sigma_p$ , which is related to  $\Sigma$  by

$$\Sigma_p(R) = \Sigma[R/a(\chi_1)], \quad \text{or} \quad \Sigma(R_{\text{com}}) = \Sigma_p[a(\chi_1)R_{\text{com}}], \quad (3.97)$$

yielding

$$\gamma_t(\theta) = \Sigma_{\text{cr}}^{-1} \Delta\Sigma(\chi_1\theta) = \Sigma_{\text{cr}}^{-1} \Delta\Sigma_p(D_1\theta). \quad (3.98)$$

We argue that the definition used should depend on the science case. For example, when considering the mean density profile of galaxies, it is more reasonable to use proper transverse separations – as that density profile is expected to be approximately stationary in proper coordinates. For larger-scale correlations between galaxies and matter, however, the use of comoving transverse separations is more meaningful, since the shape of the cross-correlation function on large scales is expected to be approximately preserved.

### 3.C.2 RELATION TO PREVIOUS WORK

In the literature on galaxy-galaxy lensing, one finds relations that differ from the ones derived above; we shall comment on some of these differences here.

The first aspect is that in several papers (e.g. Mandelbaum et al. 2010; Viola et al. 2015; de la Torre et al. 2017), the integrand in equation (3.84) is replaced by  $1 + \xi_{\text{gm}}$ , implying that the corresponding  $\Sigma_{\text{com}}$  contains the line-of-sight integrated mean density of the Universe, in addition to the correlated density. This constant term is, however, not justified by the derivation in Appendix 3.C.1. While such a constant drops out in the definition of  $\Delta\Sigma_{\text{com}}$ , and thus does not impact on quantitative results, it nevertheless causes a principal flaw: its inclusion would imply that the convergence  $\kappa = \Sigma_{\text{cr,com}}^{-1} \Sigma_{\text{com}}$  for all lines-of-sight to redshifts  $z_s \sim 1$  would be several tenths, causing a large difference between shear and reduced shear [ $g = \gamma/(1 + \kappa)$ ], which is the observable in weak lensing. That is, it would strongly modify the relation between  $\gamma$  and the observable image ellipticities, yielding significant biases in all weak lensing studies. Indeed, the mean density of the Universe is already taken into account by the Robertson–Walker metric: the fact that the angular-diameter distance is a non-monotonic function of redshift can be considered as being due to the gravitational

light deflection by the mean mass density of the Universe – the convergence part in the optical tidal equation (see, e.g. Seitz et al. 1994). However, this is usually not called ‘lensing’, but ‘curvature of the metric’. Lensing is usually ascribed solely to the effect caused by density inhomogeneities. But in any case: the mean cosmic density can not be accounted for twice, once for the metric [and thus the use of (comoving) angular-diameter distances in a FRW model], and a second time for the convergence on such a background model.

A second issue in some of the recent GGL papers is mixing the use of the comoving surface mass density,  $\Sigma_{\text{com}}$  (equation 3.87), with the proper critical surface mass density,  $\Sigma_{\text{cr}}$  (equation 3.89). For example, de la Torre et al. (2017) in their equation (18) (apart from the constant term discussed above) define  $\Sigma_{\text{com}}$  as in equation (3.89), but use in their equations (9) and (10) the proper critical surface mass density  $\Sigma_{\text{cr}}$  to relate the tangential shear to  $\Sigma_{\text{com}}$ . Hence, this relation would cause an offset by a factor  $(1 + z_1)^2$  from the correct result. However, the inconsistency appears only in the text of the paper and not in the code or calculations performed (de la Torre 2018, private communication).

The same issue occurs in the write-up in several earlier publications of our KiDS team. For example, the equations (2,5,6) in Viola et al. (2015) show this inconsistency (where we also point out a typo in the integration limits of equation 2), as well as equations (1,2,3) in van Uitert et al. (2016) and, as mentioned already in the main text, equations (1,2,3) in **Chapter 2**. We have checked the codes that were used to derive the quantitative results in these papers to see whether they employ the same inconsistent use of quantities. We found that there is an inconsistency present only in writing, namely in equation (2) of Viola et al. (2015) and not in the code that was used to produce the results. As pointed out above the correlation function is given in comoving coordinates, while the extraction of the galaxy-galaxy lensing signal is calculated in proper coordinates. Thus, the equation (2) of Viola et al. (2015) should correctly read as (using their notation):

$$\Sigma(R) = 2\bar{\rho}_m (1 + \langle z_1 \rangle)^2 \int_0^{\pi_s} \xi_{\text{gm}}(\sqrt{R^2 + \Pi^2}) d\Pi. \quad (3.99)$$

Put differently, while the data were indeed extracted in proper coordinates, the outputs of the halo model were not reported to be scaled to the coordinates used by the data (Cacciato 2016, private communication). Secondly, the equations (6) and (9) in **Chapter 2** should have the same form as equation (3.50) and (3.54) in this paper, again having a mistake only in writing. The exact same correction should also be applied to the equation (3) of Velliscig et al. (2017).

The erroneous formulation also occurred in a previous paper from van Uitert et al. (2016). Their equation (2), should read correctly as equation (3.50), which is (again, using their notation):

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G(1+z)^2} \frac{D_s}{D_L D_{LS}}. \quad (3.100)$$

Further KiDS analyses from Amon et al. (2018a) and Amon et al. (2018b) measure large-scale galaxy-mass correlations using the comoving critical surface mass density. The definitions, presented in these two papers are consistent with the data and the equations derived in Appendix 3.C.1.

On the other hand, the KiDS analyses from Sifón et al. (2015), Brouwer et al. (2016) and Brouwer et al. (2017) use a separate NFW stacking method that does not rely on the halo model, and all use proper coordinates that are consistent with the data.




# 4

---

## The case for two-dimensional galaxy-galaxy lensing

---

E revisit the performance and biases of the two-dimensional approach to galaxy-galaxy lensing. This method exploits the information for the actual positions and ellipticities of source galaxies, rather than using only the ensemble properties of statistically equivalent samples. We compare the performance of this method with the traditionally used one-dimensional tangential shear signal on a set of mock data that resemble the current state-of-the-art weak lensing surveys. We find that under idealised circumstances the confidence regions of joint constraints for the amplitude and scale parameters of the NFW model in the two-dimensional analysis can be more than three times tighter than the one-dimensional results. Moreover, this improvement depends on the lens number density and it is larger for higher densities. We compare the method against the results from the hydrodynamical EAGLE simulation in order to test for possible biases that might arise due to lens galaxies being missed, and find that the method is able to return unbiased estimates of halo masses when compared to the true properties of the EAGLE galaxies. Because of its advantage in high galaxy density areas, the method is especially suitable for studying the properties of satellite galaxies in clusters of galaxies.

A. Dvornik, S. L. Zoutendijk, H. Hoekstra, K. Kuijken  
*A&A*, **Volume 627**, Issue 1, p. A74 (2019)

## 4.1 INTRODUCTION

One of the fundamental ingredients needed to understand galaxy formation is the relation between stellar mass and the host halo mass (e.g. Courteau et al. 2014). However, inferring the total mass from a galaxy's emitted light is not feasible. We must instead rely on different probes to constrain the mass of dark matter haloes around galaxies one wants to study. A powerful mechanism that can be used for this is gravitational lensing, when matter inhomogeneities deflect light rays from distant objects along their path. As a consequence the images of distant objects (sources) appear to be tangentially distorted around foreground galaxies (lenses). The strength of the distortion is proportional to the amount of mass associated with the lenses (and consequently the dark matter haloes) and it is stronger in the proximity of the centre (for a thorough review, see Bartelmann & Schneider 2001).

Weak gravitational lensing induces a coherent tangential distortion. Under the assumption that galaxies are randomly oriented, the lensing signal can be inferred by simply averaging the ellipticities of the source galaxies. The typical change in ellipticity due to gravitational lensing is much smaller than the intrinsic ellipticity of the source, even in the case of clusters of galaxies. The weak gravitational lensing signal from a single galaxy halo is therefore too weak to be detected, and we must rely on a statistical approach in which the contributions from different lens galaxies are stacked, selected by similar observational properties (e.g. stellar masses, luminosities, size). The usual method used to analyse weak lensing data is to average the tangential component of the distortion in radial bins. As the signal from a single lens is purely tangential, this is a succinct way of showing the information contained in the distortions induced by one lens; there is no information lost in azimuthally averaging a radially symmetric signal and therefore the mass distribution of the lens can be perfectly determined from this radial profile. Average halo properties, such as halo masses, are then inferred from the resulting high signal-to-noise ratio measurements. This technique is commonly referred to as galaxy-galaxy lensing, and it is used as a method to measure statistical properties of dark matter haloes around galaxies (e.g. Leauthaud et al. 2011; van Uitert et al. 2011; Velander et al. 2014; Cacciato et al. 2014; Viola et al. 2015). The stacking mentioned here is not required per se, but it provides a convenient and unbiased data compression method that also allows for separate study of central and satellite galaxies. However, it does typically result in a loss of information about the halo properties.

For the lenses that do not exist in isolation the signal is not purely tangential. In this case the distortions around a lens are the sum of the tangential patterns of all the neighbouring lenses. An azimuthal average of these distortions will discard the azimuthal information that is present in this case. This non-optimal use of information will result in a less precise mass estimation than would be possible with a two-dimensional method.



Here we revisit a different method of analysing galaxy-galaxy lensing data, first proposed by Schneider & Rix (1997), and we make a case for why it should be considered again: it uses the unique signatures of overlapping regions of lenses to constrain the halo properties more precisely. Two-dimensional galaxy-galaxy lensing tries to fit a two-dimensional shear field directly to the galaxy ellipticity measurements. It was initially named ‘maximum-likelihood galaxy-galaxy lensing’ after the fitting method it was first studied with (e.g. Schneider & Rix 1997; Hudson et al. 1998; Geiger & Schneider 1999; Hoekstra et al. 2003, 2004; Han et al. 2015). Maximum-likelihood galaxy-galaxy lensing is thus a misnomer and in principle one could use any form of fitting method to infer the desired parameters of the two-dimensional weak lensing maps, ideally using a fully Bayesian model (Sonnenfeld & Leauthaud 2018).

This method went out of fashion due to the unavailability of the galaxy grouping information that would accurately classify the galaxies as centrals and satellites (Hoekstra 2014) as it was realised that these objects need to be modelled separately. Treating the galaxies as centrals and satellites in a statistical way when considering the stacked signal could be naturally accounted for with the halo model (Seljak 2000; Peacock & Smith 2000; Cooray & Sheth 2002), thus overcoming the observational shortcomings. In recent years the galaxy grouping information has become available due to the power of wide-field photometric surveys (e.g. KiDS; Kuijken et al. 2015; de Jong et al. 2015) complemented with spectroscopic group information (from spectroscopic surveys like GAMA; Driver et al. 2011; Robotham et al. 2011) that treat the central and satellite galaxies deterministically (e.g. Sifón et al. 2015; Brouwer et al. 2017). One important advantage of the two-dimensional method is that it exploits all the information of the actual image configuration (the model predicts the shear for each individual galaxy image) using various parameters, including the galaxies’ exact positions, ellipticities, magnitudes, luminosities, stellar masses and group membership information rather than using only the ensemble properties of statistically equivalent samples (Schneider & Rix 1997). Moreover, the clustering of the lenses is naturally taken into account, although it is more difficult to account for the expected diversity in density profiles (Hoekstra 2014).

The outline of this paper is as follows. In Sect. 5.3 we present the maximum likelihood formalism used for galaxy-galaxy lensing, for both one-dimensional and two-dimensional methods. In Sect. 5.4 we present the lens model used to construct the mock observations and investigation of EAGLE galaxies (Schaye et al. 2015; Crain et al. 2015; McAlpine et al. 2016). The mock observations are further described in Sect. 4.4 where we also test the two-dimensional method and examine the limitations in the case of masked data. In Sect. 4.5 we examine the EAGLE simulation (Schaye et al. 2015) using the two-dimensional galaxy-galaxy lensing methodology. We conclude and discuss in Sect. 5.6. Throughout the paper we use the following cosmological parameters in the calculation of the distances and other relevant properties (Planck Collaboration et al. 2013, as used in the EAGLE simulation):  $\Omega_m = 0.307$ ,  $\Omega_\Lambda = 0.693$ ,  $\sigma_8 = 0.8288$ ,  $n_s = 0.9611$ ,  $\Omega_b = 0.04825$ , and  $h = 0.6777$ . All the measurements presented in the paper are in comoving units.

## 4.2 2D GALAXY-GALAXY LENSING FORMALISM

The likelihood of a model with a set of parameters  $\theta$  given data  $\mathbf{d}$  is parametrised in the form

$$\mathcal{L}(\theta | \mathbf{d}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{C}|}} \exp \left[ -\frac{1}{2} (\mathbf{m}(\theta) - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{m}(\theta) - \mathbf{d}) \right], \quad (4.1)$$

where  $\mathbf{m}(\theta)$  is the value of  $\mathbf{d}$  predicted by the model with parameters  $\theta$ . We assume the measured data points  $\mathbf{d} = [d_1, \dots, d_n]$  are drawn from a normal distribution with a mean equal to the true values of the data. The likelihood function accounts for correlated data points through the covariance matrix  $\mathbf{C}$ . The covariance matrix  $\mathbf{C}$  consists of two parts, the first arising from the shape noise and the second from the presence of cosmic structure between the observer and the source (Hoekstra 2003):

$$\mathbf{C} = \mathbf{C}^{\text{shape}} + \mathbf{C}^{\text{LSS}}. \quad (4.2)$$

Using the Equation 4.1, the parameters of the best-fitting model can be determined with

$$\tilde{\theta}(\mathbf{d}) \equiv \underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta | \mathbf{d}) = \underset{\theta}{\operatorname{argmin}} \chi^2(\theta | \mathbf{d}). \quad (4.3)$$

For convenience we define

$$\chi_{\min}^2(\mathbf{d}) \equiv \chi^2(\tilde{\theta}(\mathbf{d}) | \mathbf{d}) \quad (4.4)$$

as the value of the chi-square statistic for the best-fitting model, which is also the minimal value of the chi-square statistic.

When fitting one-dimensional tangential shear profiles stacked over a sample of lenses, the likelihood function can be written as

$$\mathcal{L}(M_h, M_\star, c | \gamma_t^{\text{obs}}) = \prod_i \frac{1}{\sigma_{\gamma_t, i} \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\gamma_{t, i}(M_h, R, z) - \gamma_{t, i}^{\text{obs}}}{\sigma_{\gamma_t, i}} \right)^2 \right], \quad (4.5)$$

where we use  $m_i = \gamma_{t, i}(M_h, R, z)$  as the model prediction given halo mass  $M_h$ , radial bin  $R$ , and redshift of the lens  $z$ , and the  $d_i = \gamma_{t, i}^{\text{obs}}$  as the tangentially averaged shear of a sample of lenses measured from observations. Here we also use the (statistical) uncertainty on our measurement given by the  $\sigma_{\gamma_t, i}$  calculated from the intrinsic shape noise of sources in each radial bin. Moreover, we assume that the variance  $\sigma^2$  is the diagonal of the full covariance matrix

$$\sigma = \sqrt{|\mathbf{C}|}; \quad (4.6)$$

i.e., we only account for the error due to the shape noise. Similarly, the likelihood function can be defined for the case when fitting the two-dimensional shear field

$$\mathcal{L}(M_h, M_\star, c | \epsilon^{\text{obs}}) = \prod_i \frac{1}{\sigma_{\epsilon, i} \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{g_i(M_h, \theta, z) - \epsilon_i^{\text{obs}}}{\sigma_{\epsilon, i}} \right)^2 \right], \quad (4.7)$$

where  $g_i(M_h, \theta, z)$  are the reduced shears evaluated at each source position  $\theta$ ,  $\epsilon_i^{\text{obs}}$  are the observed ellipticities of real galaxies, and  $\sigma_{\epsilon,i}$  is the intrinsic shape noise of our galaxy sample per component, and is the same as the  $\sigma_{\gamma_i,i}$ . In practice, the two-dimensional fit to the ellipticities is carried out for each cartesian component of ellipticity  $\epsilon_1$  and  $\epsilon_2$  with respect to the equatorial coordinate system of the real data or mock catalogues used in our validation study.

### 4.3 LENS MODEL

The most widely assumed density profile for dark matter haloes is the Navarro-Frenk-White (NFW) profile (Navarro et al. 1996). Using simple scaling relations this profile can be matched to simulated dark matter haloes over a wide range of masses and was found to be consistent with observations (Navarro et al. 1996). It is defined as

$$\rho_{\text{NFW}}(r) = \frac{\delta_c \bar{\rho}_m}{(r/r_s)(1+r/r_s)^2}, \quad (4.8)$$

where the free parameters  $\delta_c$  and  $r_s$  are called the overdensity and the scale radius, respectively, and  $\bar{\rho}_m$  is the mean density of the universe, where  $\bar{\rho}_m = \Omega_m \rho_c$  and  $\rho_c$  is the critical density of the universe, defined by

$$\rho_c \equiv \frac{3H_0^2}{8\pi G}, \quad (4.9)$$

where  $H_0$  is the present day Hubble parameter.

The NFW profile in its usual parametrisation has two free parameters for each halo, halo mass  $M_h$ , and concentration  $c$ , and using these parameters is the conventional way of modelling halo profiles. However, having two free parameters for each halo is computationally very expensive. Instead, we would like to describe these parameters through relations that depend on halo properties, and then fit to a few free parameters in these global relations instead of hundreds or thousands of free, halo-specific parameters.

To this end, we adopt the halo mass–concentration relation of Duffy et al. (2008), which is also an adequate description of the measured halo mass–concentration relation of central and satellite galaxies in the EAGLE simulation (Schaye et al. 2015; Schaller et al. 2015)

$$c(M_h, z) = 10.14 \left[ \frac{M_h}{(2 \times 10^{12} M_\odot / h)} \right]^{-0.081} (1+z)^{-1.01}. \quad (4.10)$$

We also adopt the stellar mass-to-halo mass relation, as measured in the EAGLE simulation, using the functional form presented in Matthee et al. (2017),

$$\log(M_\star/M_\odot) = \alpha - e^\gamma (M_h/M_\odot)^{\beta \log(e)}, \quad (4.11)$$

where  $\alpha = 11.50$ ,  $\beta = -0.86$ , and  $\gamma = 10.58$ .

After removing all halo-specific degrees of freedom, we introduce two new, global degrees of freedom in order to avoid recalculating the shape of the profile in every single model evaluation. They are introduced in the form of the factors  $f$  and  $g$ , which scale the values of the scale radius  $r_s$  and product  $\delta_c r_s$  relative to the values  $\tilde{r}_s(M_\star)$  and  $\tilde{\delta}_c(M_\star) \tilde{r}_s(M_\star)$  expected from a lens with a stellar mass  $M_\star$  through the two scaling relations for  $c$  and  $M_h$ :

$$\begin{aligned} r_s &= f \tilde{r}_s(M_\star), \\ \delta_c r_s &= g \tilde{\delta}_c(M_\star) \tilde{r}_s(M_\star). \end{aligned} \quad (4.12)$$

Our two parameters thus correspond to a scaling of the amplitude and scale of the NFW profile. This makes the interpretation of results straightforward and is the most general parametrisation of the NFW profile. These parameters are expected to be of order unity. While the scaling relations were measured on the EAGLE simulation, which we use to validate the method, the slight differences on exact definitions of quantities as measured on the simulations and what weak gravitational lensing infers (and scatter around the mean of those distributions) might cause slight changes in the value of the fiducial parameters. We do not expect to see any in the case of simulated, toy model observations. These lens models can be generalised to account for scatter (e.g. in the stellar mass-to-halo mass relation or in the concentration–mass relation) by making it fully Bayesian, similar to the model presented in Sonnenfeld & Leauthaud (2018).

The gravitational shear and convergence profiles are then calculated using the equations presented by Wright & Brainerd (2000), from which the predicted ellipticities for all the lenses are calculated according to the weak lensing relations presented in Schneider (2003). We first calculate the reduced shear for our NFW profiles,

$$g(\theta, z_s) = \frac{\gamma(\theta, z_s)}{1 - \kappa(\theta, z_s)}, \quad (4.13)$$

from which the ellipticities are calculated according to

$$\epsilon = \begin{cases} g & |g| \leq 1 \\ 1/g^* & |g| > 1 \end{cases}, \quad (4.14)$$

where we assume that the intrinsic ellipticities of the sources average to 0, due to their random nature. In practice, we avoid the strong lensing regime by removing these sources from our catalogue.

## 4.4 PROOF OF CONCEPT

We created the mock catalogues in a semi-empirical manner. In order to test the method on a realistic dataset, the mock catalogues were made to closely resemble the Kilo Degree Survey (KiDS) properties (de Jong et al. 2015; Kuijken et al. 2015). We randomly placed 30 700 sources at a redshift of 0.7 in a  $1 \text{ deg}^2$  field. This corresponds to the size of one KiDS tile with the number of sources reflecting the observed number density (Hildebrandt et al. 2017) at the median redshift for the whole survey.

We did not assign any intrinsic orientation or ellipticity to our sources; this uncertainty can be accommodated for directly in our maximum likelihood fits by scaling the covariance matrix (or in this case the variance used in the likelihood functions) so that the intrinsic source ellipticity uncertainty is representative of the shape noise in the KiDS survey, considering the overlap with the Galaxy and Mass Assembly survey (GAMA; Liske et al. 2015).

The generated source field was then used to calculate the weak lensing effect of the foreground lenses that we placed in the same field. We calculated the effect of each lens according to the model presented in Sec. 5.4, using only one stellar mass for all the lens galaxies placed in the mock catalogue. We decided to assign a stellar mass of  $M_\star = 10^{12} M_\odot$  and  $f = g = 1$ . We positioned all the lenses at the same redshift of 0.2, which is around the median redshift of the GAMA survey commonly used in KiDS galaxy-galaxy lensing studies (Viola et al. 2015; Sifón et al. 2015; van Uitert et al. 2016; Brouwer et al. 2016; Dvornik et al. 2018, amongst others). The contributions from multiple lenses to the shear (and consequently ellipticity) of one source galaxy can be summed together linearly, i.e.

$$\gamma(\theta, z_s)_i = \sum_j \gamma(\theta, z_s)_{ij}, \quad (4.15)$$

where the sum goes over the  $j$  lenses in the catalogue, with shear evaluated at each source position  $i$ .<sup>1</sup> This means that we actually allow for contributions of neighbouring haloes, which will become evident later on in the paper. We also assume that each lens galaxy is exactly at the centre of its dark matter halo, ignoring the possibility of miscentring. When placing the lenses in our mock field, we draw their positions in the same way as for the sources, but we do not allow for exact spatial overlap of any lens. The number of lenses that we add to the KiDS-like field varies between 1 and 720 (the latter reflects the typical density of the GAMA galaxies) in order to study the performance of the method as a function of galaxy number density so we can test the effects of the neighbouring haloes.

When working with ground-based observations we would never have included strongly lensed sources in the analysis as in the majority of cases these sources are also blended with the lens galaxy (the typical Einstein radius for GAMA galaxies is smaller than 5 arcsec); the simulated mock observations can have sources that are strongly lensed because we distributed them randomly. To eliminate this problem, all the sources with  $|g| > 0.3$  were removed from the catalogue in order to limit our

---

<sup>1</sup>We first calculate the  $\gamma(\theta, z_s)$  and the  $\kappa(\theta, z_s)$ , then use Equation 5.10 to calculate the reduced shear.

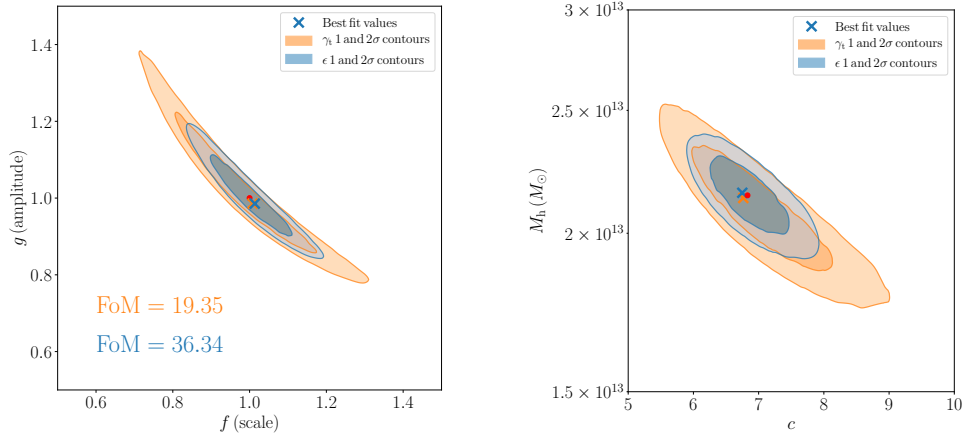


Figure 4.1: Confidence areas of the scale parameter  $f$  and amplitude parameter  $g$  (left panel) and the confidence areas of the halo mass  $M_h$  and halo concentration  $c$  jointly derived from the constraints on the  $f$  and  $g$  parameters (right panel) for an analysis of the mock KiDS+GAMA area. Orange contours show the maximum likelihood fit on the stacked tangential shear profiles, and the blue contours the maximum likelihood fit as it was performed on the ellipticities of sources used directly, using all the galaxies in the mock field simultaneously. Shown are the best-fitting values for each method (orange and blue crosses) and the fiducial lens model (red circle). The contours correspond to the case with 50 lenses per  $\text{deg}^2$  in the simulated field.

analysis to the weak lensing regime. This threshold is quite low, but it makes sure we always stay in the weak lensing regime that motivates the use of Equation 4.15.

Using these mock catalogues, we test our lens model and compare the results obtained using the one-dimensional stacked tangential shear method against the two-dimensional method that uses measured ellipticities directly. At the same time, this allows us to study the two methods under known conditions and makes the results easier to understand.

The main question we want to address here is how the effective lens galaxy density influences the performance of the two-dimensional galaxy-galaxy lensing method as the unique signatures caused by the spatial lens configuration on the shear field result in information gain for the inference of halo masses and halo concentrations.

The second question we want to address is the sensitivity of the two-dimensional method to incompleteness in the lens sample. This bias can be induced by lenses outside of the observed field (or masked from the data), and corrections to account for this effect were already studied in the past (Hudson et al. 1998). We applied a typical KiDS data mask to the generated mock catalogues and also studied the unmasked mock catalogues, but purposefully ignored a number of lenses that are present in the field.

We first applied both methods to the same source sample in which we varied the number of lenses from 1 to  $720 \text{ deg}^{-2}$ . We assigned the uncertainty of the measured shapes to  $\sigma_\epsilon = 0.3$ . For the one-dimensional method, each source's uncertainty was further weighted by  $\sqrt{N}$ , where  $N$  is the number of lenses that contribute to the total shear of that source, to account for proper covariance between the sources. This is naturally captured by the two-dimensional galaxy-galaxy lensing method. Furthermore, we limited ourselves to a subset of sources that we use in both methods. The subset of sources is selected by the smallest and largest annuli ( $R_{\text{max}}$ ) in which we calculate the tangentially averaged shear profiles of our galaxies. This allows us to directly compare the methods, as for the case of one lens. Given that we use the same source galaxies, the results from the two methods should be exactly the same. At the same time, we also select the lens galaxies that are at least  $R_{\text{max}}$  from the field edge to minimise the effects of missing source galaxies beyond our simulated field. We fit the data using Equations 5.3 and 5.5 for the tangential shears measured on the data and the ellipticities as created in our mock catalogues, respectively. In the fit we vary the parameters  $f$  and  $g$ , which scale the reference NFW profile for the typical scale and amplitude. We sample the values of  $f$  and  $g$  on a Latin hypercube grid (McKay et al. 1979) using 500 points. We compare the inferred best-fit values and the  $1\sigma$  and  $2\sigma$  contours obtained from a  $\chi^2$  surface, which is in turn computed from the aforementioned grid using an interpolation on a finer linearly spaced grid. Using this information we calculate a figure of merit (FoM) which is defined as an inverse of the 68% confidence level area and we study the ratio of the FoM between the one-dimensional stacked tangential shear method and the two-dimensional method.

The results using 50 lenses per  $\text{deg}^2$  can be seen in Fig. 4.1, where we show the fiducial value of the  $f$  and  $g$  parameters, the best-fit values, and the  $1\sigma$  and  $2\sigma$  uncertainty contours on the derived best-fit values. Similarly, in Fig. 4.1, we show the constraints on the halo mass  $M_h$  and halo concentration  $c$ , as derived from the constraints on parameters  $f$  and  $g$ . The best-fit values with the individual 68% confidence intervals are listed in Table 4.1, for the parameters  $f$  and  $g$ , and for the halo mass  $M_h$  and halo concentration  $c$ . Both methods are capable of recovering the input values. What is more, the contours for the two-dimensional method are noticeably smaller. This can be seen more clearly in Fig. 4.2 where the orange line shows the FoM as function of number of lens galaxies in our mock field. This figure shows that information is gained as the contributions of neighbouring dark matter haloes leave unique shear configuration signatures that can only be accounted for using a two-dimensional galaxy-galaxy lensing method. At low lens densities we expect the two methods to perform identically (with FoM ratio = 1) as the separation of the galaxies is large enough for us to assume that the lenses are isolated, such that  $\gamma_t$  contains all lensing information. The same effect (ratio of FoM = 1) should also be observed if two lenses are exactly on the same line of sight. We note that we consider here a noiseless mock dataset, with shape noise accounted through the covariance matrix, and this means that the signal-to-noise ratio at low densities does not influence our ability to constrain contribution of individual haloes, and consequently allows us to obtain the ideal case of FoM = 1 for the case of one lens. The figure of merit stays close to 1 as long as the separations are large enough for contributions of neighbour-

Table 4.1: Best-fit values for the  $f$  and  $g$  parameters and for the halo mass  $M_h$  and halo concentration  $c$ , together with their individual 68% confidence intervals for the model using 50 lenses per  $\text{deg}^2$  in mock KiDS and GAMA like data.

	$f$	$g$	$M_h[10^{13}M_\odot]$	$c$
1D	$1.0 \pm_{0.19}^{0.14}$	$0.98 \pm_{0.11}^{0.22}$	$2.13 \pm_{0.43}^{0.37}$	$6.76 \pm_{1.28}^{2.26}$
2D	$1.01 \pm_{0.11}^{0.08}$	$0.98 \pm_{0.07}^{0.11}$	$2.15 \pm_{0.25}^{0.24}$	$6.75 \pm_{0.86}^{1.18}$

ing lenses to remain sufficiently low. With large lens galaxy number densities the lenses start overlapping, we gain less information, and the figure of merit starts levelling off. This is caused by the source number density that stays the same for any number of lens galaxies we add to the field, which limits the available signal-to-noise ratio of the measured source ellipticities.

We now focus on the second question in our investigation, whether there is any bias introduced when not all lensing galaxies are accounted for. To this end, we use the same KiDS-like mock field with 720 lens galaxies, but now we remove one lens in each iteration, thus effectively accounting for the possible bias we might introduce in real observations by not accounting for galaxies just outside of our observed field or not accounting for lens galaxies that were masked out of the data. Figure 4.2 shows the shift of the best-fit parameters away from the fiducial model as a function of the field completeness, averaged over five different realisations of the lens distribution. What is immediately clear is that the NFW fit to the one-dimensional tangential shear profiles recovers the true input parameters (as it is essentially removing any configuration information from the sample by the tangential averaging), also for the cases of low completeness. The two-dimensional method can only do this successfully at high completeness values; any small deviation and unexpected features in the field caused by the presence of lenses not accounted for drives the recovered values of the input parameters away from the truth as the model tries to accommodate the missing lenses.

We also study the effect of the masking introduced by a realistic KiDS survey mask, shown in Fig. 4.3. We apply this mask to our mock catalogues and repeat the fitting of our model to the lenses and sources that remain in the mock catalogues. We again change the number of lenses in the field and the results of this exercise can be seen in Fig. 4.2 (blue line). What can be observed is that the two-dimensional method, even in the case of masking, is still more precise, and that the difference in precision is a direct result of the amount of masked area. The accuracy of the method due to masking behaves in a similar way to that shown in Fig. 4.2, but the observed bias is smaller because a larger number of lenses remain in the field. A typical KiDS survey mask reduces the number of lenses by about 20% (Kuijken et al. 2015; de Jong et al. 2015; Hildebrandt et al. 2017), which can bias the fitted parameters up to 10%, as shown in Fig. 4.2 (vertical grey line). This needs to be accounted for in an application to real data.



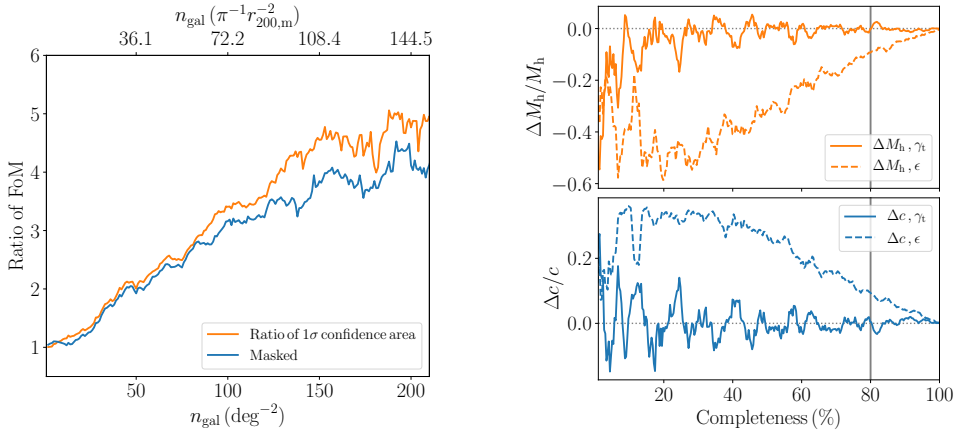


Figure 4.2: *Left panel*: Figure of merit as a function of lens number density in a simulated  $1 \text{ deg}^2$  field. The orange line shows the case where we consider all the galaxies in the field, and thus gives us an estimate of improvement in precision when using a two-dimensional method. The improvement levels off at a value of around 5, which indicates that in dense galaxy fields, the loss of signal-to-noise ratio due to the limited number of sources cannot be overcome. The blue line shows the case where we apply a typical KiDS survey mask to our mock catalogues. *Right panel*: Relative shift of the halo mass  $M_h$  (*top*) and halo concentration  $c$  (*bottom*) derived from the constraints on the  $f$  and  $g$  parameters from the fiducial model as a function of completeness. Shown are the shift of the recovered parameters for the one-dimensional method (solid lines) and the shift of the recovered parameters for the two-dimensional method (dashed lines). Also shown is a typical completeness due to a mask in a KiDS like survey (vertical grey line).

Thus far we have ignored systematic biases in the galaxy shape measurements. They can be split into a multiplicative bias, which leads to an overall scaling of the signal, and an additive bias that manifests as a preferred orientation of galaxies. As the former simply scales the signal, the impact on the one-dimensional and two-dimensional analyses is the same. The situation is different in the case of additive bias: a constant signal will simply vanish when we consider the azimuthally averaged tangential shear (in the limit of no edge effects). Even a spatially varying additive bias is expected to vanish because it typically does not align with the line connecting the lens and the source. In contrast, in the two-dimensional case we expect the  $\chi^2$  to become poor as the systematic signal contributes to it. To examine whether this has any impact on the recovered model parameters we mimic a systematic shape measurement error by adding a constant uniform shear to our mock dataset and repeat our analysis. We find that the overall  $\chi^2$  surface indeed becomes offset by a constant (positive) value; however, we are nonetheless able to recover the input parameters exactly as in our fiducial case.<sup>2</sup>

<sup>2</sup>This can be explicitly seen by writing out the  $\chi^2$  with the added constant shear,  $\chi^2 \propto (\gamma_t + c - m)^2$ , where  $c$  is the constant shear and  $m$  the model prediction. The cross terms in the expanded form average to 0 and

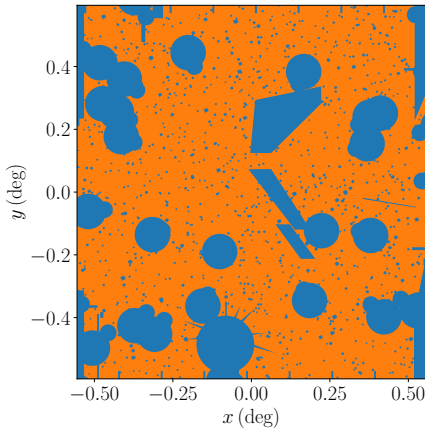


Figure 4.3: Typical KiDS survey  $r$ -band mask used to evaluate the effect of masking on the inference of best-fit parameters, and the possible bias masking might introduce.

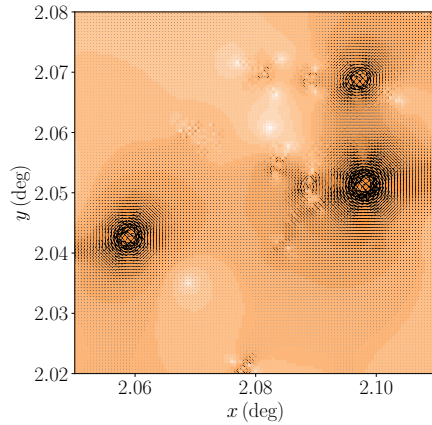


Figure 4.4: Segment of the shear map derived from the EAGLE particle data. A number of notable features of weak lensing are visible in this plot. The ellipticities are tangentially aligned with the lenses and the strength of gravitational lensing diminishes with distance from the lens. The lens configuration creates a unique pattern that contains information about mass distribution that is otherwise lost when tangentially averaging the observed shears.

## 4.5 EVALUATION OF THE TWO METHODS WITH THE EAGLE SIMULATION

Motivated by the success of the two-dimensional galaxy-galaxy lensing method from the previous section, we now focus on more realistic tests using the EAGLE hydrodynamical simulation (Schaye et al. 2015; Crain et al. 2015; McAlpine et al. 2016) as our input data. Studying a simulation gives us the ability to compare our two-dimensional galaxy-galaxy lensing results against the truth, properties as measured directly from particle properties in the simulation. We note that for the purpose of this study we do not use a lightcone generated from the EAGLE simulation. Although we include complexities of neighbouring galaxies, we do not capture projections along the line of sight or missing galaxies, for example. We use the AGN simulation AGNdT9L0050N0752, which has  $752^3$  dark matter particles and a box size of 50 comoving Mpc (Schaye et al. 2015), and it is calibrated in such a way that it

---

we gain a constant term  $c^2$ , which worsens the overall  $\chi^2$ , but does not inhibit the ability to minimise the  $(\gamma_t - m)^2$  difference.

reproduces global observables of our Universe. The EAGLE simulation was also shown to correctly predict the galaxy-galaxy lensing signal when compared to the KiDS+GAMA data (Velliscig et al. 2017), for both central and satellite galaxies. We take the full particle information in a box with a comoving size of 50 Mpc which is then binned to  $8195 \times 8195 \times 8195$  pixels. The box is then projected along the axes, yielding three different mass maps of the EAGLE simulation. To calculate the shear at each location, we first position the mass map at redshift of 0.2 (by scaling it comovingly), calculate the density map using the mean density of the Universe, and use the Kaiser & Squires (1993) prescription to calculate the shear, rolling the edges of the map. Because we want the sources to be positioned at a redshift that resembles the typical source redshift in the KiDS survey, we calculate the convergence map using the critical surface mass density at redshift of 0.7. A small portion of an EAGLE shear map is shown in Fig. 4.6. The whole EAGLE map corresponds to a  $60 \text{ deg}^2$  patch of sky.

Further information about the properties of the lens galaxies are queried from the public EAGLE database (McAlpine et al. 2016), such as the total halo mass, centre of mass, centre of potential, stellar mass, stellar mass within certain aperture, group memberships, and group properties. From the database we select galaxies with stellar masses ranging from  $10^{9.6} M_{\odot}$  to  $10^{11.2} M_{\odot}$ , which is a range that allows us to have enough galaxies in finer stellar mass bins (which we will use for our fiducial stacked tangential shear method) and ensures that the galaxies in EAGLE are well defined in terms of simulation particle mass. From this selection of galaxies we take both the centrals and satellite galaxies. Inclusion of satellite galaxies in the study is crucial for the two-dimensional method, as otherwise the results can be substantially biased (up to 10% for typical survey masks), as demonstrated in the previous section and in Fig. 4.2, and at the same time it allows their properties to be studied, as was previously done using the one-dimensional method, by Sifón et al. (2015), amongst others. In total, after applying all the selection criteria, we are left with 859 galaxies (520 centrals and 339 satellites).

We calculate the tangential shear signal for each galaxy in our sample using the tangential  $\epsilon_1$  component of the source's ellipticity around the position of the lens. The azimuthal average of the tangential ellipticity is then our unbiased estimate of the tangential shear. For the two-dimensional method, we use the  $\epsilon_1$  and  $\epsilon_2$  values directly. The tangential shear profiles and their averages for the  $10^{10.8}$  to  $10^{11.0} M_{\odot}$  stellar mass bin can be seen in Fig. 4.6. The noisiness of the the individual profiles can be directly attributed to the fact that we are azimuthally averaging the data on a square grid.

The central and satellite galaxies in the EAGLE simulation follow a different stellar-to-halo mass relation, and we also account for this in the model. For this we use the same relation used in Sec. 5.4 and we fit it to the halo masses derived from the NFW fits to the convergence field of individual galaxies, both centrals and satellites (this is done in order to use the same definition of the halo mass for both galaxy types). The two stellar-to-halo mass relations are shown in Fig. 4.5, and the parameters obtained are listed in Table 4.2. The different stellar-to-halo mass relations are then accounted for in the modified lens model that differentiates between the central

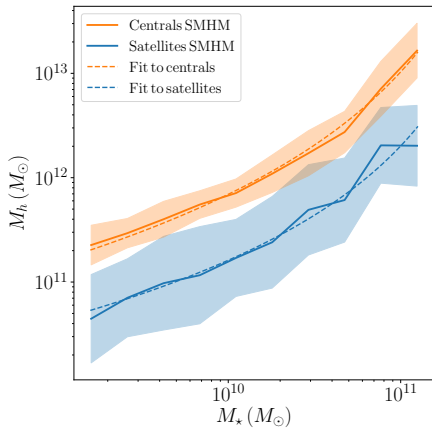


Figure 4.5: Stellar-to-halo mass relation for the central galaxies in the EAGLE simulation (orange) and for the satellites (blue). With full orange and blue lines we show the median stellar-to-halo mass relation with the corresponding scatter shown with the filled areas of the same colours. The dashed orange and blue lines are the models for the stellar-to-halo mass relation as used in the analysis.

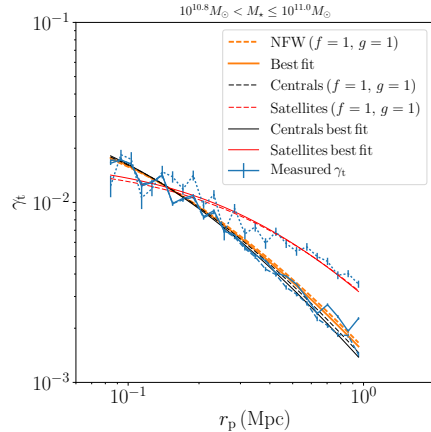


Figure 4.6: Stacked tangential shear profiles for the lenses selected from the EAGLE simulation (shown with blue lines) in the  $10^{10.8}$  to  $10^{11.0} M_{\odot}$  stellar mass bin. The dashed orange, red and black lines we show the signals as predicted by our fiducial lens model using  $f = 1$  and  $g = 1$ . The total best-fitting model is shown with the solid orange line and the best-fitting model for centrals and satellites in full black and red lines, respectively. The corresponding halo masses and concentrations of input models and the best fit results are listed in Table 4.3.

and satellite galaxies for the one-dimensional method and for the two-dimensional method. From the same fit we note that the concentrations of the haloes are generally lower than the prediction from Duffy et al. (2008). While they still follow the same trend, the normalisation of the relation is lower for both centrals and satellites, with a normalisation of 0.6 for centrals and 0.25 for satellites, with the lower values arising because we include all the haloes, not only the relaxed ones. This is consistent with what was found by Viola et al. (2015).

We fit the two lens models<sup>3</sup> to the mean tangential shear profiles per bin and to the full ellipticity data using Equations 5.3 and 5.5. To account for the uncertainty in our ellipticity measurements we again assign the standard deviation of 0.3, scaled to the typical number density of GAMA and KiDS set-up due to the size of the pixel in our mass maps. This results in an uncertainty of  $\sigma_{\epsilon} = 0.015$ . The gain in

<sup>3</sup>One for centrals and one for satellites, for a combined set of four parameters; each model has a set of  $(f, g)$  parameters.

Table 4.2: Parameters of the stellar-to-halo mass relation used in the analysis of the EAGLE simulation, used together with the functional form presented in Equation 5.9, and the normalisations of the concentration–mass relation.

	$\alpha$	$\beta$	$\gamma$	$c/c_{\text{Duffy}}$
Centrals	11.81	-0.68	8.69	0.6
Satellites	11.72	-0.79	9.44	0.25

Table 4.3: Central values of the input and best-fit halo masses and concentrations for the  $10^{10.8}$  to  $10^{11.0}M_{\odot}$  stellar mass bin. The input values are a median of the halo masses and concentrations of haloes in that bin measured from the convergence fit to the EAGLE data and the output values are the predictions from the best-fit one-dimensional model, which can be seen in Fig. 4.6.

	Input $M_{\text{h}}$	Best-fit $M_{\text{h}}$	Input $c$	Best-fit $c$
Centrals	7.17	6.62	5.62	6.08
Satellites	1.33	1.30	2.72	2.96

precision is 3.9, which is the ratio of the area of the 68% confidence level contours ( $\text{FoM}_{2\text{D}}/\text{FoM}_{1\text{D}} = 3.9$ ). We show the separate credibility contours for the halo mass and concentration in Fig. 4.7 for the one-dimensional and two-dimensional methods, separated into contributions from central and satellite galaxies; both values are scaled with the input stellar-to-halo mass relation and the concentration–mass relation, which then show the relative change of the halo masses and concentration from fiducial values obtained from the simulation.

At first sight it might seem that the results are somewhat biased with regard to the actual measured scaling relation of the EAGLE galaxies, but we do observe an almost equal effect on the  $f$  and  $g$  parameters for the two methods. This is not necessarily due to a bias in the analysis. After all, the intrinsic scatter in the concentration–mass relation and in the stellar mass-to-halo mass relation are not accounted for in the model and are the most likely cause of small shifts in the methods presented. Both methods also give robust estimates for the properties of central and satellite galaxies and given the results, the two-dimensional method is much more precise in constraining the two fitted parameters than the one-dimensional method. Given the large uncertainty on the recovered parameters of the one-dimensional method for satellite galaxies, the one-dimensional method is unable to robustly capture the contribution from satellites in the KiDS+GAMA data, as was also demonstrated by Sifón et al. (2015). Given the results of this exercise, we expect to capture the contribution from satellites when we apply the two-dimensional method to the data.

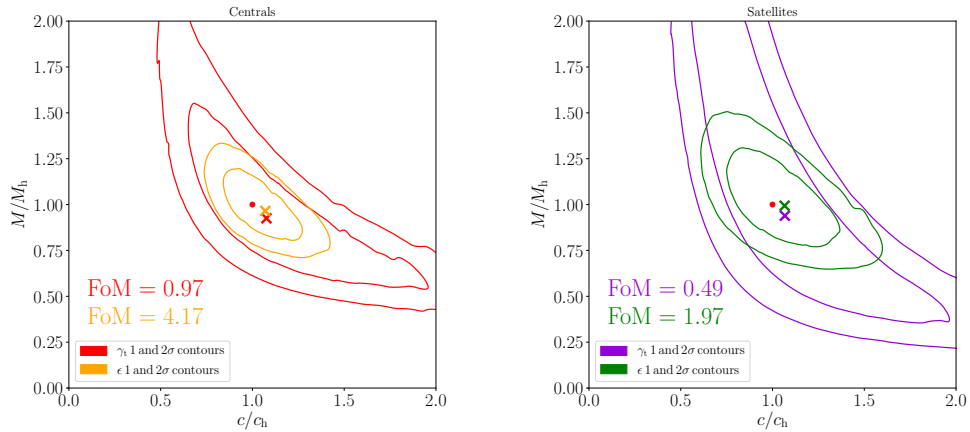


Figure 4.7: *Left panel:* Confidence areas of the halo mass  $M_h$  and concentration  $c$  of central galaxies for the analysis of the EAGLE simulation using the one-dimensional method and the two-dimensional method, scaled with the input stellar-to-halo mass relation and the concentration-mass relation. *Right panel:* Confidence areas of the halo mass  $M_h$  and concentration  $c$  of satellite galaxies for the analysis of the EAGLE simulation using the one-dimensional method and the two-dimensional method, scaled with the input stellar-to-halo mass relation and the concentration-mass relation. The contours show the results of the maximum likelihood fit on the central galaxies (red and purple) and satellite galaxies (orange and green). Crosses (in corresponding colours) show the best-fitting values for each method and galaxy sample, and the red circles show the fiducial models. The contours are calculated from the contours obtained as a fit of the  $f$  and  $g$  parameters.

## 4.6 DISCUSSION AND CONCLUSIONS

We have investigated the precision and bias of one and two-dimensional galaxy-galaxy lensing analyses of weak lensing data, using tangential averaged shear profiles and ellipticities, respectively, keeping in mind current and upcoming state-of-the-art large weak lensing galaxy surveys. The main difference between the two methods lies in the fact that the two-dimensional approach uses all the available information in an observed field. While the one-dimensional method uses only the ellipticities of source galaxies to infer the stacked tangential shear signal, the two-dimensional method uses actual relative positions of all the lens galaxies in a field and the ellipticities of all the sources in the field. Because the two-dimensional galaxy-galaxy lensing accounts for spatial configuration of the lens galaxies, the unique signatures in the shear field caused by overlapping regions of influence contain more information about the halo properties of the lenses we want to study and result in a significant improvement over the traditional one-dimensional stacking methods.

We tested the method on mock observations generated in a semi-empirical way where we assumed a model with the gravitational lenses represented by the NFW profiles with properties determined from observable quantities such as stellar mass, taking into account a typical configuration and properties of KiDS and GAMA surveys. We find that the two-dimensional method gives better constraints on those same parameters: the FoM is more than three times larger compared to the results from stacked tangential shear profiles. This suggests that there can be an equal amount of information hidden in the exact configuration of the lenses and their overlaps, which is lost when a one-dimensional method is used. The precision gain also depends on the lens density. In denser fields of gravitational lenses, the gain in precision from using the two-dimensional method is larger, as the signal becomes more heavily influenced by neighbouring gravitational lenses. We also studied the case where we removed a significant fraction of galaxies present in a mock field from our analysis, and while the two-dimensional method still gives us better constraints on the NFW parameters, the accuracy of these parameters starts to suffer because the modelling of the lenses does not account for the contributions of shears that are caused by the galaxies we left out of our analysis. While this indeed produces a noticeable bias, and thus needs to be corrected to properly recover the true values of the parameters we study, the case where such a large fraction of galaxies would be missed is rather severe. This effect of correlated structure—undetected galaxies that are clustered with the observed galaxies and the matter distribution on group scale—is in reality negligibly small (as discussed in detail already by Hudson et al. 1998).

We assumed a model where lenses are represented by the NFW profiles, up to constant pre-factors for the lensing signal amplitude and scale. We used the same lens model as well for the study on the EAGLE simulation (Schaye et al. 2015; McAlpine et al. 2016). As we used the concentration–mass relation that closely describes the one measured in the EAGLE simulation and a stellar-to-halo mass relation of the EAGLE central and satellite galaxies in our lens model, we expected both methods to recover the input parameters values. We find that the two methods are able to almost perfectly recover these values, and the small differences can be attributed to the non-ideal modelling of the galaxies in the EAGLE simulation. The two-dimensional method does indeed perform better.

Given that the two-dimensional galaxy-galaxy lensing method requires knowledge of group (and/or cluster) membership, preferentially inferred from spectroscopic data, we identify two cases where using the two-dimensional method could be preferred over the one-dimensional method. The most obvious one is studying the group properties as a function of halo mass, where using the two-dimensional method can give better constraints on scaling relations of group halo mass with luminosity of central galaxies, their stellar mass, size, X-ray gas emission, and the concentration of such haloes. As we have precise membership information of galaxies in clusters and because of the increased number density of these galaxies, the second case is to study the sub-halo mass function of galaxy clusters to a high precision. The two-dimensional galaxy-galaxy lensing, together with the group or cluster membership information that is available by using highly complete spectroscopic surveys is an obvious choice for galaxy-galaxy studies on dense galaxy fields in general.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for the very useful comments and suggestions. We thank Mike Hudson for the useful discussion on the topic. AD acknowledges support from grant number 614.001.541 and HH acknowledges support from Vici grant number 639.043.512, both financed by the Netherlands Organisation for Scientific Research (NWO). KK acknowledges support from the Alexander von Humboldt Foundation.

This work has made use of Python (<http://www.python.org>), including the packages `numpy` (<http://www.numpy.org>) and `scipy` (<http://www.scipy.org>). The plots were produced with `matplotlib` (Hunter 2007).




# 5

---

## KiDS+GAMA: Inferring satellite halo masses using two-dimensional shear maps

---

 E use data from the Kilo-Degree Survey (KiDS) and the Galaxy And Mass Assembly (GAMA) surveys to simultaneously constrain the stellar-to-halo mass relations of both central and satellite galaxies of spectroscopically confirmed galaxies in galaxy groups using weak lensing. For the analysis we use the traditional one-dimensional method in the form of the stacked tangential shear measurements to determine the halo and subhalo masses of our galaxies and to constrain the stellar-to-halo mass relation, as well as a two-dimensional fit to the full shear field that uses all the available information about lens galaxies and exact source galaxies positions and ellipticities. We find that the two-dimensional method performs better than the one-dimensional method statistically by a factor of  $\sim 2$ . Both methods lead to similar parameters of the stellar-to-halo mass relation, which are consistent with previous results found in the literature, showing that the satellite galaxies have generally lower halo masses than the central galaxies, given the same stellar mass.

A. Dvornik, K. Kuijken, H. Hoekstra, with KiDS and GAMA collaborations

## 5.1 INTRODUCTION

According to the hierarchical galaxy formation model, galaxy groups and clusters form by accretion of isolated galaxies and groups. Such an assembly process will tidally strip mass from the infalling satellite galaxies/haloes. Because the dark matter is dissipationless it will be more easily stripped from the subhalo than the baryons, which will dissipate some of their energy and sink to the centre of their potential well before forming stars (White & Rees 1978). This model thus predicts that the satellite galaxies will be preferentially stripped of their dark matter and the effect can be observed as higher stellar mass to halo mass ratios of satellite galaxies compared to their central counterparts of similar stellar mass.

While the stellar-to-halo mass relation of central galaxies has been successfully measured by many studies (for instance by Hoekstra et al. 2005; Mandelbaum et al. 2006; More et al. 2011; van Uitert et al. 2011; Leauthaud et al. 2012), this is not the case for satellite galaxies whose stellar-to-halo mass relation remains essentially unconstrained (Sifón et al. 2018). Recently, several weak gravitational lensing studies using galaxy groups and clusters have been undertaken (such as the ones by Limousin et al. 2007; Li et al. 2014b, 2016; Sifón et al. 2015, 2018), all finding that the satellite galaxies are heavily truncated with the respect to the central and field galaxies. All the previous simulation studies (Bower et al. 2006) show that the stellar-to-halo mass relation of satellite galaxies is significantly different from the stellar-to-halo mass relation of central galaxies.

In order to measure the stellar-to-halo mass ratio of satellite galaxies, one needs to estimate the total mass of satellite galaxies. Weak gravitational lensing, through the lensing of background sources by a sample of galaxies – commonly called galaxy-galaxy lensing, directly measures the total mass of lensing galaxies, without assuming their dynamical state (Bartelmann & Schneider 2001; Courteau et al. 2014), and it is currently the only method available to measure the total mass of samples of galaxies directly. Measuring the lensing signal around satellite galaxies, however, can be particularly challenging for several reasons: a small contribution to the lensing signal by the host galaxy group, source blending at small separations and sensitivity to field galaxy contamination (Sifón et al. 2018). As pointed out by Sifón et al. (2015), the latter point is quite important as the field galaxies will not be stripped and the contamination complicates the interpretation of the lensing signal.

To overcome the aforementioned challenges in measuring the satellite galaxies' lensing signal in this study we use the two-dimensional galaxy-galaxy lensing method, first proposed by Schneider & Rix (1997), to analyse galaxy-galaxy lensing data. The two-dimensional galaxy-galaxy lensing method tries to fit a two-dimensional shear field directly to the galaxy ellipticity measurements, and it was shown to perform significantly better for dense lens populations, compared to the traditional one-dimensional method in the form of the stacked tangential shear estimates or excess surface density (ESD) profiles (**Chapter 4**).

This method went out of fashion due to the unavailability of galaxy grouping information that would accurately classify galaxies as centrals and satellites (Hoekstra 2014), the same information needed to robustly study the stellar mass to halo mass

relation of satellite galaxies (Sifón et al. 2015). Treating the galaxies as centrals and satellites in a statistical way when considering the stacked signal could be naturally accounted for with the halo model (Seljak 2000; Peacock & Smith 2000; Cooray & Sheth 2002), thus overcoming the observational shortcomings. In recent years the galaxy grouping information has become available thanks to the power of wide-field photometric surveys (for instance KiDS; Kuijken et al. 2015; de Jong et al. 2015) complemented with spectroscopic group information (from spectroscopic surveys like Galaxy And Mass Assembly (hereafter GAMA) survey; Driver et al. 2011; Robotham et al. 2011) that allow one to treat the central and satellite galaxies deterministically. One important advantage of the two-dimensional method lies in the fact that it exploits all the information of the actual image configuration (the model predicts the shear for each individual background galaxy image) using the galaxies' exact positions, ellipticities, magnitudes, luminosities, stellar masses, group membership, information, etc., rather than using only the ensemble properties of statistically equivalent samples (Schneider & Rix 1997). Moreover, the clustering of the lenses is naturally taken into account, although it is more difficult to account for the expected diversity in density profiles (Hoekstra 2014).

In this paper we present a two-dimensional galaxy-galaxy lensing measurement of the stellar-to-halo mass relation for central and satellite galaxies, by combining a sample of spectroscopically confirmed galaxy groups from the Galaxy And Mass Assembly survey (Driver et al. 2011) and background galaxies from the Kilo-Degree Survey (**Chapter 6**). We use these measurements to constrain the stellar-to-halo mass relation using both one-dimensional stacked tangential shear profiles and the two-dimensional galaxy-galaxy lensing method.

The outline of this paper is as follows. In Sec. 5.2 we present the lens and source sample used in this analysis. In Sec. 5.3 we present the two-dimensional galaxy-galaxy lensing formalism and in Sec. 5.4 we present the specific lens model used in the paper. We present the results in Sec. 5.5 and conclude with Sec. 5.6. Throughout the paper we use the following cosmological parameters entering in the calculation of the distances and other relevant properties (Planck Collaboration et al. 2013):  $\Omega_m = 0.307$ ,  $\Omega_\Lambda = 0.693$ ,  $\sigma_8 = 0.8288$ ,  $n_s = 0.9611$ ,  $\Omega_b = 0.04825$  and  $h = 0.6777$ . The halo masses are defined as  $M = 4\pi r_\Delta^3 \Delta \bar{\rho}_m / 3$  enclosed by the radius  $r_\Delta$  within which the mean density of the halo is  $\Delta$  times the mean density of the Universe  $\bar{\rho}_m$ , with  $\Delta = 200$ . All the measurements presented in the paper are in comoving units.

## 5.2 DATA AND SAMPLE SELECTION

The foreground galaxies used in this lensing analysis are taken from the GAMA survey (Driver et al. 2011), a spectroscopic survey carried out on the Anglo-Australian Telescope with the AAOmega spectrograph. Specifically, we use the information of GAMA galaxies from three equatorial regions, G9, G12 and G15 from GAMA II (Liske et al. 2015). We do not use the G02 and G23 regions, because the first one does not overlap with KiDS and the second one uses a different target selection compared to the one used in the equatorial regions. These equatorial regions encom-

pass  $\sim 180 \text{ deg}^2$ , contain 180 960 galaxies (with  $nQ \geq 3$ , where the  $nQ$  is a measure of redshift quality) and are highly complete down to a Petrosian  $r$ -band magnitude  $r = 19.8$ . For this thesis Chapter we only use galaxies in the G9 field (a full KiDS and GAMA overlap study will be performed later on) and we use only the galaxies that reside in groups identified by Robotham et al. (2011). Inclusion of only the group galaxies might potentially bias our results, but we defer this analysis to the upcoming and complete study on the full GAMA area. The GAMA galaxy group catalogue was constructed using a 3-dimensional Friends-of-Friends (FoF) algorithm, linking galaxies in projected and line-of-sight separation. We use version 10 of the group catalogue (G3Cv10), which contains 26 194 (7481 in G9) groups with at least 2 members. Following (Viola et al. 2015), we restrict ourselves to galaxy groups with at least 5 members as low multiplicity groups are contaminated with interlopers (Robotham et al. 2011) and consider all the galaxies within those groups whose stellar mass is between  $10^8 M_\odot$  and  $10^{12} M_\odot$ . Stellar masses are taken from version 20 of the LAMBDA stellar mass catalogue, described in Wright et al. (2017). The final selection of galaxies can be seen in Fig. 5.1 and Fig. 5.2, and all the relevant properties we need in our analysis are presented in Table 5.1. The stellar mass binning is used only for the one-dimensional galaxy-galaxy lensing case in order to obtain stacks of tangential shear signal. In the two-dimensional case, we directly use the relevant galaxy quantities in the model.

We use imaging data from the  $60 \text{ deg}^2$  of KiDS (**Chapter 6**) that overlaps with the G9 patch of the GAMA survey (Driver et al. 2011) to obtain shape measurements of background galaxies. KiDS is a four-band imaging survey conducted with the OmegaCAM CCD mosaic camera mounted at the Cassegrain focus of the VLT Survey Telescope (VST); the camera and telescope combination provide us with a fairly uniform point spread function across the field-of-view.

We use shape measurements based on the  $r$ -band images, which have an average seeing of 0.66 arcsec. The image reduction, photometric redshift calibration and shape measurement analysis is described in detail in Hildebrandt et al. (2018) and in **Chapter 6**. We measure galaxy shapes using *lensfit* (Miller et al. 2013), which has been calibrated using image simulations described in Kannawadi et al. (2019). This provides galaxy ellipticities ( $\epsilon_1, \epsilon_2$ ) with respect to an equatorial coordinate system.

### 5.3 2D GALAXY-GALAXY LENSING FORMALISM

In this study of satellite galaxy-galaxy lensing we use the two-dimensional galaxy-galaxy lensing formalism as discussed in **Chapter 4**, following the model therein and adapting it to work with KiDS+GAMA data, taking into account the survey specific requirements. Generally, for both the one-dimensional and two-dimensional cases, the likelihood of a model with a set of parameters  $\theta$  given data  $\mathbf{d}$  can be parametrised in the following form:

$$\mathcal{L}(\theta | \mathbf{d}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{C}|}} \exp \left[ -\frac{1}{2} (\mathbf{m}(\theta) - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{m}(\theta) - \mathbf{d}) \right], \quad (5.1)$$

Table 5.1: Overview of the number of galaxies/lenses, median stellar masses of galaxies and median redshifts in each selected bin used for our one-dimensional stacked tangential shear analysis. Stellar masses are given in units of  $[\log(M_\star/[M_\odot])]$ .

Bin	Range	$N_{\text{tot}}$	$N_{\text{cen}}$	$N_{\text{sat}}$	$M_{\star,\text{med}}$	$z_{\text{med}}$
1	(8.0,10.0]	1004	3	1001	9.66	0.095
2	(10.0,10.2]	491	5	486	10.11	0.167
3	(10.2,10.4]	740	10	730	10.32	0.192
4	(10.4,10.6]	1034	29	1005	10.51	0.202
5	(10.6,10.8]	1295	53	1242	10.71	0.247
6	(10.8,11.0]	1298	108	1190	10.90	0.272
7	(11.0,11.2]	1016	235	781	11.09	0.282
8	(11.2,11.4]	513	229	284	11.28	0.282
9	(11.4,11.6]	239	171	68	11.47	0.294
10	(11.6,11.8]	40	36	4	11.66	0.314
11	(11.8,12.0]	5	3	2	11.88	0.271

where  $\mathbf{m}(\theta)$  is the value of  $\mathbf{d}$  predicted by the model with parameters  $\theta$ . We assume the measured data points  $\mathbf{d} = [d_1, \dots, d_n]$  are drawn from a normal distribution with a mean equal to the true values of the data. The likelihood function accounts for correlated data points through the covariance matrix  $\mathbf{C}$ . The covariance matrix  $\mathbf{C}$  consists of two parts, the first one arising from shape noise and the second part from the presence of cosmic structure between the observer and the source (Hoekstra 2003):

$$\mathbf{C} = \mathbf{C}^{\text{shape}} + \mathbf{C}^{\text{LSS}}. \quad (5.2)$$

In the case when one wants to fit one-dimensional tangential shear profiles, stacked over a sample of lenses, the likelihood function can be written as:

$$\begin{aligned} \mathcal{L}(M_h, M_\star, c | \gamma_t^{\text{obs}}) & \quad (5.3) \\ &= \prod_i \frac{1}{\sigma_{g_{t,i}} \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{g_{t,i}(M_h, R, z) - g_{t,i}^{\text{obs}}}{\sigma_{g_{t,i}}} \right)^2 \right], \end{aligned}$$

where we have used  $m_i = g_{t,i}(M_h, R, z)$  as the model prediction given halo mass  $M_h$ , radial bin  $R$  and redshift of the lens  $z$ , and the  $d_i = g_{t,i}^{\text{obs}}$  as the tangentially averaged shear of a sample of lenses measured from observations. Here we have also used the uncertainty of our measurement, given by the  $\sigma_{g_{t,i}}$  calculated from the intrinsic shape noise of sources in each radial bin. Moreover we assume that the variance  $\sigma^2$  is the diagonal of the full covariance matrix:

$$\sigma = \sqrt{|\mathbf{C}|}, \quad (5.4)$$

i.e., we only account for the error due to the shape noise. Similarly, the likelihood function can be defined for the case when one would like to fit the two-dimensional

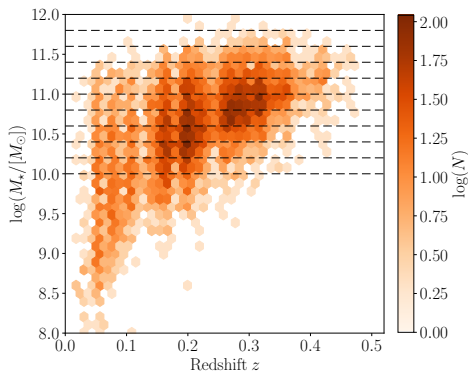


Figure 5.1: Stellar mass versus redshift of galaxies in rich groups in the G9 region of the GAMA survey that overlap with KiDS. The full sample is shown with hexagonal density plot and the dashed lines show the cuts for the stellar mass bins used in our analysis.

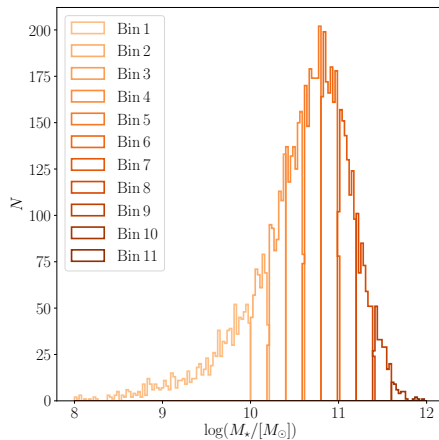


Figure 5.2: Stellar mass distributions in our 11 bins used for one-dimensional stacked tangential shear measurements. The exact bin values are presented in Table 5.1.

shear field:

$$\begin{aligned} \mathcal{L}(M_h, M_*, c | \epsilon^{\text{obs}}) & \quad (5.5) \\ &= \prod_i \frac{1}{\sigma_{\epsilon,i} \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{g_i(M_h, x_i, z) - \epsilon_i^{\text{obs}}}{\sigma_{\epsilon,i}} \right)^2 \right], \end{aligned}$$

where  $g_i(M_h, x_i, z)$  are the reduced shears evaluated at each source position  $x_i$ ,  $\epsilon_i^{\text{obs}}$  the observed ellipticities of real galaxies and  $\sigma_{\epsilon,i}$  is the intrinsic shape noise of our galaxy sample per component, calculated from the *lensfit* weights following the description by (Heymans et al. 2012) and it is the same as the  $\sigma_{g_i,i}$ . The same *lensfit* weights are used to weight the  $\epsilon_i^{\text{obs}}$  as well. In practice, the two-dimensional fit to the ellipticities is carried out for each cartesian component of ellipticity  $\epsilon_1$  and  $\epsilon_2$  with respect to the equatorial coordinate system.

## 5.4 LENS MODEL

The most widely assumed density profile for dark matter haloes is the Navarro–Frenk–White (NFW) profile (Navarro et al. 1996). Using simple scaling relations this profile can be matched to simulated dark matter haloes over a wide range of masses and was found to be consistent with observations (Navarro et al. 1996). It is defined

as:

$$\rho_{\text{NFW}}(r) = \frac{\delta_c \bar{\rho}_m}{(r/r_s)(1+r/r_s)^2}, \quad (5.6)$$

where the free parameters  $\delta_c$  and  $r_s$  are called the overdensity and the scale radius, respectively, and  $\bar{\rho}_m$  is the mean density of the universe, where  $\bar{\rho}_m = \Omega_m \rho_c$  and  $\rho_c$  is the critical density of the universe, defined by

$$\rho_c \equiv \frac{3H_0^2}{8\pi G}, \quad (5.7)$$

where  $H_0$  is the present day Hubble parameter.

The NFW profile in its usual parametrisation has two free parameters for each halo, halo mass  $M_h$  and concentration  $c$ , and using those is the conventional way of modelling halo profiles. However, having two free parameters for each halo is computationally very expensive and not supported by the data. Rather we would like to describe these parameters through relations that depend on halo properties, and then fit to a few free parameters in these global relations instead of hundreds or thousands of free, halo-specific parameters. To do so, we adopt the halo mass – concentration relation of Duffy et al. (2008), with a free concentration normalisation  $f_c$ :

$$c(M_h, z) = f_c 10.14 \left[ \frac{M_h}{(2 \times 10^{12} M_\odot/h)} \right]^{-0.081} (1+z)^{-1.01}, \quad (5.8)$$

We also adopt the stellar mass to halo mass relation in a shape of a single power law:

$$M_h = M_0 \left( \frac{M_\star}{10^{11} M_\odot} \right)^\alpha, \quad (5.9)$$

where  $\alpha$  and  $M_0$  are the free parameters we will be fitting. We use separate relations for the central and satellite galaxies, as we want to constrain the stellar-to-halo mass relation for those populations separately.

The gravitational shear and convergence profiles are then calculated using the equations presented by Wright & Brainerd (2000), from which the predicted ellipticities for all the lenses are calculated according to the weak lensing relations presented in Schneider (2003). We first calculate the reduced shear for our NFW profiles:

$$g(x_i, z_s) = \frac{\gamma(x_i, z_s)}{1 - \kappa(x_i, z_s)}, \quad (5.10)$$

from which the ellipticities are calculated according to the following equation:

$$\epsilon = \begin{cases} g & |g| \leq 1 \\ 1/g^* & |g| > 1 \end{cases}, \quad (5.11)$$

where we have assumed that the intrinsic ellipticities of the sources average to 0, due to their random nature.

We compute the effective critical surface mass density that we need in our lens model for each lens using the spectroscopic redshift of the lens  $z_1$  and the full normalised redshift probability density of the sources,  $n(z_s)$ , calculated using the direct calibration method presented in Hildebrandt et al. (2017, 2018).

The effective inverse critical surface density can be written as:

$$\Sigma_{\text{cr,ls}}^{-1} = \frac{4\pi G}{c^2} (1 + z_1)^2 D(z_1) \int_{z_1}^{\infty} \frac{D(z_1, z_s)}{D(z_s)} n(z_s) dz_s, \quad (5.12)$$

where  $D(z_1)$  is the angular diameter distance to the lens,  $D(z_1, z_s)$  is the angular diameter distance between the lens and the source and  $D(z_s)$  is the angular diameter distance to the source.

The galaxy source sample is specific to each lens redshift with a minimum photometric redshift  $z_s = z_1 + \delta_z$ , with  $\delta_z = 0.2$ , where  $\delta_z$  is an offset to mitigate the effects of contamination from the group galaxies (for details see also the methods section and Appendix of **Chapter 2**). We determine the source redshift distribution  $n(z_s)$  for each sample, by applying the sample photometric redshift selection to a spectroscopic catalogue that has been weighted to reproduce the correct galaxy colour-distributions in KiDS (for details see Hildebrandt et al. 2018). We correct the measured ellipticities for the multiplicative shear bias per source galaxy per redshift bin as defined in Hildebrandt et al. (2018) with correction values estimated from the image simulations (Kannawadi et al. 2019).

## 5.5 RESULTS

The free parameters for our model are listed in Table 5.2, together with their prior ranges. We use a Bayesian inference method in order to obtain full posterior probabilities using a Monte Carlo Markov Chain (MCMC) technique; more specifically we use the `emcee` Python package (Foreman-Mackey et al. 2013). The likelihoods we use are the same as given by Equations 5.3 and 5.5. We use wide flat priors for all the parameters (given in Table 5.2), which are the same for both our methods as well.

We run the sampler using 32 walkers, each with 50 000 steps (for a combined number of 1 600 000 samples), out of which we discard the first 5000 burn-in steps, 160 000 samples). The resulting MCMC chains are well converged according to the integrated autocorrelation time test.

We fit the lens model as described in Sect 5.4 to the measured stacked tangential shear measurements in our 11 stellar mass bins. A single stacked tangential shear profile for the GAMA lenses (blue points) in the  $10^{11.2}$  to  $10^{11.4} M_{\odot}$  stellar mass bin is shown in Fig. 5.3, with the measurements and their respective  $1\sigma$  errors (orange lines and bands). The measured lens model best-fit parameters, together with their 68% credibility intervals are presented in Table 5.2. Their full posterior distributions are shown in Fig. 5.6. The resulting fit has a reduced  $\chi^2_{\text{red}} (\equiv \chi^2/\text{d.o.f.})$  equal to 0.94, which is an appropriate fit, given the 93 degrees of freedom (d.o.f.).

The main results of this work are the stellar-to-halo mass relations for centrals and satellites, which for the one-dimensional study, we show in Fig. 5.4. The stellar-



Table 5.2: Priors and marginalised posterior estimates of the free parameters used in our lens model, for both the one-dimensional method and the two-dimensional method. All priors are uniform in linear space in the quoted range. The central values are calculated as median of the MCMC samples and the uncertainties are the 68% credibility interval.

	$\log(M_{0,\text{cen}}/M_{\odot})$	$\alpha_{\text{cen}}$	$f_{c,\text{cen}}$	$\log(M_{0,\text{sat}}/M_{\odot})$	$\alpha_{\text{sat}}$	$f_{c,\text{sat}}$
Priors	[5, 15]	[0, 5]	[0, 5]	[5, 15]	[0, 5]	[0, 5]
1D	$13.37^{+0.19}_{-0.25}$	$0.61^{+0.48}_{-0.40}$	$0.43^{+0.15}_{-0.12}$	$12.89^{+0.11}_{-0.13}$	$0.35^{+0.15}_{-0.15}$	$0.03^{+0.02}_{-0.01}$
2D	$13.44^{+0.11}_{-0.22}$	$0.85^{+0.09}_{-0.25}$	$0.49^{+0.17}_{-0.20}$	$12.91^{+0.05}_{-0.08}$	$0.47^{+0.14}_{-0.05}$	$0.09^{+0.03}_{-0.06}$

to-halo mass relations are completely described with two parameters each (two for centrals and two for satellites) – the normalisation  $M_0$  and slope  $\alpha$ , for which the obtained values for centrals and satellite are  $\log(M_{0,\text{cen}}/M_{\odot}) = 13.37^{+0.19}_{-0.25}$ ,  $\alpha_{\text{cen}} = 0.61^{+0.48}_{-0.40}$  and  $\log(M_{0,\text{sat}}/M_{\odot}) = 12.89^{+0.11}_{-0.13}$ ,  $\alpha_{\text{sat}} = 0.35^{+0.15}_{-0.15}$  respectively. As expected, the stellar-to-halo mass relations are significantly different for the central and satellite galaxies, showing that the stripping of the dark matter does indeed take place (the stellar-to-halo mass relation of satellite galaxies is lower than the one of the centrals). What is more, the width of the obtained stellar-to-halo mass relation is similar to the one that can be seen in simulations, for instance by the EAGLE hydrodynamical simulation (Schaye et al. 2015; Matthee et al. 2017).

The normalisations of the concentration-halo mass relation  $f_c$  are  $f_{c,\text{cen}} = 0.43^{+0.15}_{-0.12}$  and  $f_{c,\text{sat}} = 0.03^{+0.02}_{-0.01}$  for centrals and satellites, respectively, comparable to the values that are found in hydrodynamical simulations (**Chapter 4**). The concentration-halo mass normalisation is not significantly correlated with any of the other parameters of the stellar-to-halo mass relation, thus not significantly influencing our results. Those values are also consistent with the observational findings that prefer lower normalisations than expected in simulations, such as in the studies of Viola et al. (2015); Sifón et al. (2015); Dvornik et al. (2017).

For the two-dimensional results we can only present the obtained posterior distributions of our free parameters, as there is no direct representation of the results as in the case of one-dimensional galaxy-galaxy lensing. The measured lens model best-fit parameters, together with their 68% credibility intervals are presented in Table 5.2. Their full posterior distributions are shown in Fig 5.6. The resulting fit has a reduced  $\chi^2_{\text{red}}$  ( $\equiv \chi^2/\text{d.o.f.}$ ) equal to 0.92, with 2 235 297 degrees of freedom (d.o.f.), which means that the model is slightly over-fitting the data and a reduced set of parameters might be needed.

For the two-dimensional study we present the obtained stellar-to-halo mass relations in Fig. 5.5. The obtained values for centrals and satellite in the case when using the two-dimensional galaxy-galaxy lensing are  $\log(M_{0,\text{cen}}/M_{\odot}) = 13.44^{+0.11}_{-0.22}$ ,  $\alpha_{\text{cen}} = 0.85^{+0.09}_{-0.25}$  and  $\log(M_{0,\text{sat}}/M_{\odot}) = 12.91^{+0.05}_{-0.08}$ ,  $\alpha_{\text{sat}} = 0.47^{+0.14}_{-0.05}$  respectively, reported as well in Table 5.2. These results are comparable to the ones from the one-dimensional method and show that the two-dimensional method performs better statistically. Furthermore, we have not used all the galaxies in the G9 GAMA patch, leaving out the

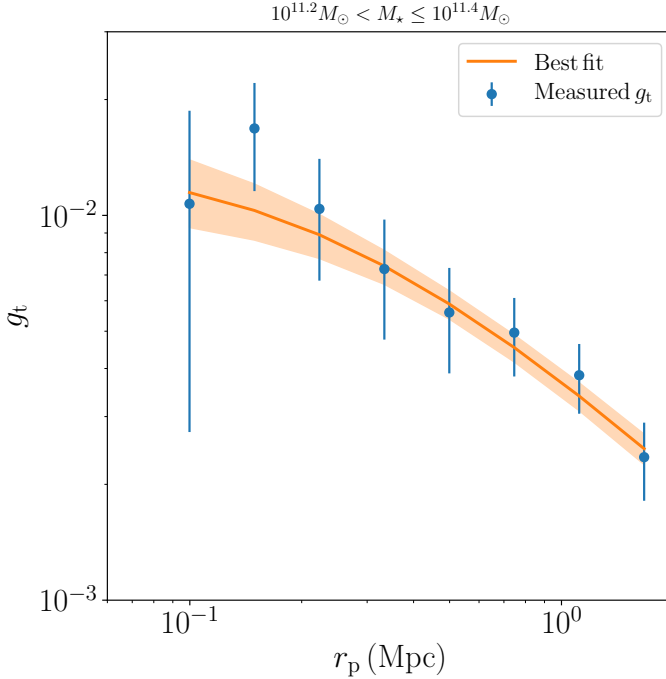


Figure 5.3: Stacked tangential shear profile for the GAMA lenses (blue points) in the  $10^{11.2}$  to  $10^{11.4}M_{\odot}$  stellar mass bin. The full orange line together with the orange band shows the best fitting lensing model for the 1D method, consisting of contribution from centrals and satellites and the 68% credibility interval, respectively.

groups with small member populations and all the isolated galaxies, thus possibly biasing our two-dimensional inference. Any galaxies left out of the analysis will bias the results up to 20% estimated for the KiDS and GAMA study (**Chapter 4**), mostly affecting the normalisation of the stellar-to-halo mass relation. To properly account for this, we need to repeat the analysis using all the available lens galaxies as well as test for the robustness of the central and satellite galaxy classification in the GAMA catalogue. The group catalogue is known to be contaminated by the misidentification of the central galaxy in a group, such that the true central galaxy would be included in the satellite sample, which can introduce roughly a 15% bias on the inferred masses (Sifón et al. 2015). What is more, the satellite stellar-to-halo mass relation at high stellar mass is possibly driven by the misidentification of satellite galaxies, which should actually be classified as centrals, given the high halo masses measured. This is a likely consequence of the observed problem with the Friends-of-Friend (FoF) algorithm used to identify galaxy groups in the GAMA survey, but it does not seem to largely affect the obtained results. The FoF algorithm will separate groups into

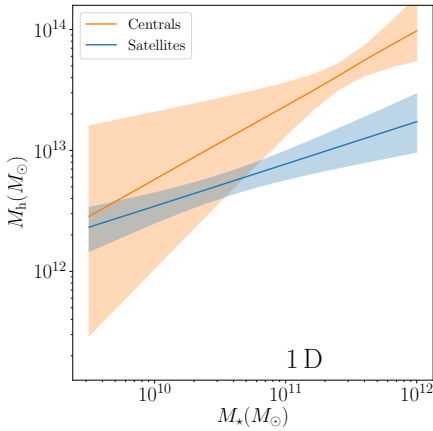


Figure 5.4: Stellar-to-halo mass relation for the central galaxies (orange) and satellite galaxies (blue) for the one-dimensional tangential shear analysis of GAMA groups. The solid lines show the median relation obtained from our MCMC fit and the orange and blue band show the 68% credibility interval for the inferred relation.

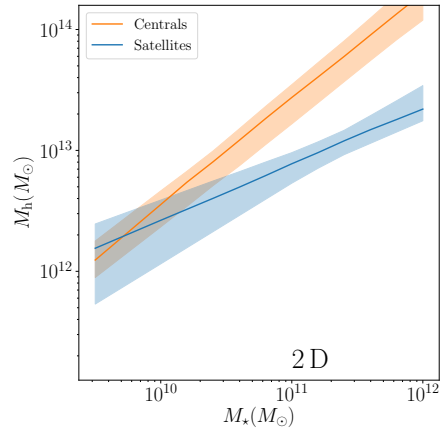


Figure 5.5: Stellar-to-halo mass relation for the central galaxies (orange) and satellite galaxies (blue) for the two-dimensional tangential shear analysis of GAMA groups. The solid lines show the median relation obtained from our MCMC fit and the orange and blue band show the 68% credibility interval for the inferred relation.

a number of smaller groups or aggregate smaller, unrelated groups into one large group, which would then host more than one central galaxy with them being classified as a satellite (Jakobs et al. 2018).

The normalisations of the concentration-halo mass relation  $f_{c, \text{cen}}$  and  $f_{c, \text{sat}}$  for centrals and satellites respectively, are also comparable to the ones for the one-dimensional method. They are less consistent with previous lensing measurements of GAMA galaxies (Viola et al. 2015; Sifón et al. 2015), especially for satellite galaxies, a fact that needs to be addressed in the future. The likely inconsistency might arise from the fact that Sifón et al. (2015) fixed the normalisation of the concentration-halo mass relation of satellite galaxies to 1. This might as well explain our over-fitting problem. The results show our ability to use the two-dimensional galaxy-galaxy lensing to constrain the stellar-to-halo mass relation of both central and satellite galaxies. The statistical performance of the two-dimensional method is better, as expected, but it is clearly noticeable that it might be significantly biased compared to the one-dimensional method.

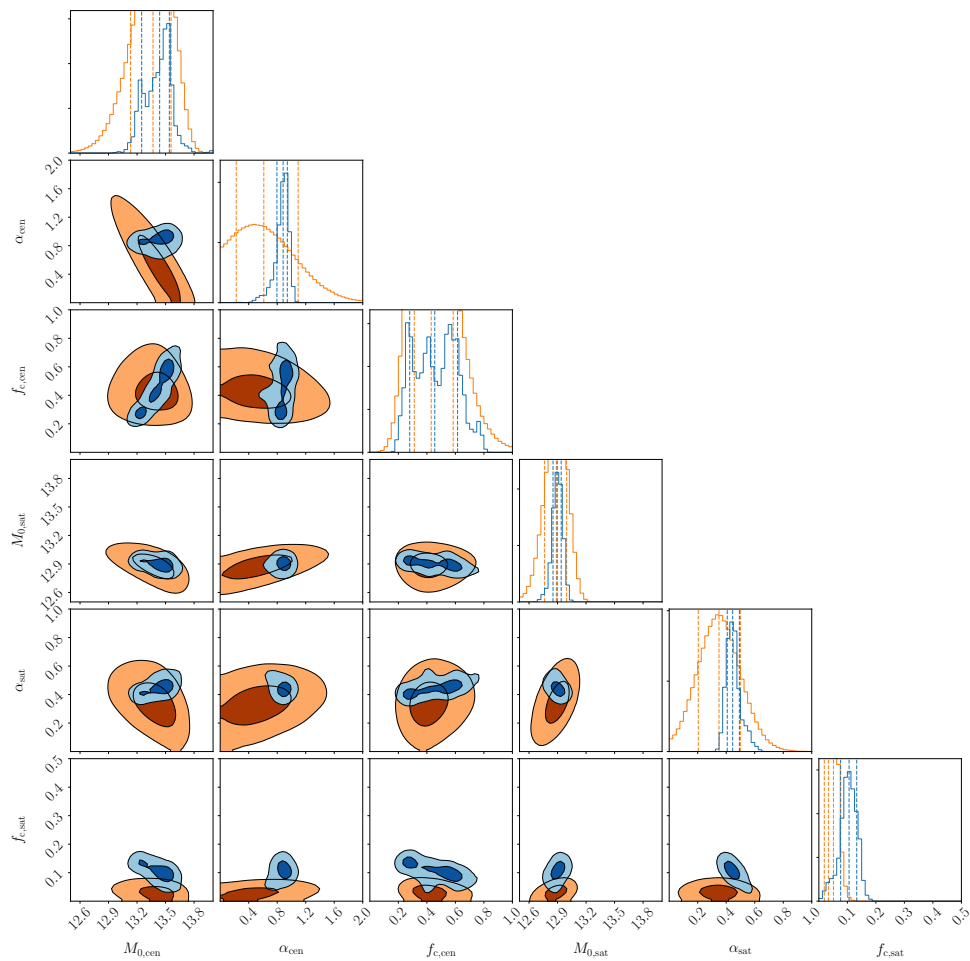


Figure 5.6: Full posterior distributions of the model parameters  $M_{0,cen}$ ,  $\alpha_{cen}$ ,  $f_{c,cen}$ ,  $M_{0,sat}$ ,  $\alpha_{sat}$  and  $f_{c,sat}$ , for both the one-dimensional stacked tangential shear measurements (in orange) as well as the two-dimensional galaxy-galaxy lensing method (in blue). The contours indicate the  $1\sigma$  and  $2\sigma$  credibility regions. Priors used in the MCMC fit can be seen in Table 5.2.

## 5.6 DISCUSSION AND CONCLUSIONS

We have made a preliminary measurement of the stellar-to-halo mass relation of central and satellite galaxies located in the GAMA groups. In this analysis we use the two-dimensional galaxy-galaxy lensing method to better constrain the stellar-to-halo mass relation, given the observed advantage of it over traditionally used stacked tangential shear method (also referred here as the one-dimensional galaxy-galaxy lensing, **Chapter 4**).

We use one of the equatorial GAMA patches (G9) that overlaps with the KiDS data in order to calculate both the tangential shear signal around the galaxies in rich groups with more than 5 members, and the two-dimensional galaxy-galaxy lensing constraints on the same lenses and sources. The tangential shear signal is then used to constrain the stellar-to-halo mass relation of central and satellite galaxies.

We model the lensing signal using an NFW profile together with the concentration-mass relation by Duffy et al. (2008), scaled by a normalisation factor that we fit for. We assume a functional form for the stellar-to-halo mass relation in the form of a single power-law and fold it through our model, thus directly fitting for the normalisation and slope of the stellar-to-halo mass relation. The lens model is used to calculate the tangential shear profile that is then fitted to the measured tangential shear profile from the GAMA and KiDS data as well as to directly predict the two cartesian components of the galaxies' ellipticities used in our two-dimensional method.

We find that the the stellar-to-halo mass relation can be successfully measured using the two-dimensional method, with a better statistical power than the traditional one-dimensional method using the stacked tangential shear measurements. Both methods give us similar results for the stellar-to-halo mass relations, showing that the two-dimensional method is indeed a robust way to measure properties of the galaxy-halo connection, without using statistically equivalent samples as in the case of the one-dimensional method, nor using more complicated halo models or relying on support from other probes. The obtained stellar-to-halo mass relations are broadly in agreement with the literature, although further study of remaining biases in our analysis is needed and the measurements will be improved shortly.

Following this pilot analysis we will refine the analysis by the inclusion of all the available overlapping KiDS and GAMA data, together with the inclusion of the remaining ungrouped galaxies will make the two-dimensional method more accurate, less biased and statistically more powerful than the one-dimensional one.

## ACKNOWLEDGEMENTS

AD and KK acknowledge support from grant number 614.001.541 and HH acknowledges support from Vici grant number 639.043.512, both financed by the Netherlands Organisation for Scientific Research (NWO). KK acknowledges support by the Alexander von Humboldt Foundation.


This research is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017 and 177.A-3018, and on data products produced by Target/OmegaCEN, INAF-OACN, INAF-OAPD and the KiDS production team, on behalf of the KiDS consortium.

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 programs 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 <http://www.gama-survey.org>.

This work has made use of Python (<http://www.python.org>), including the packages `numpy` (<http://www.numpy.org>) and `scipy` (<http://www.scipy.org>). Plots have been produced with `matplotlib` (Hunter 2007).

---

## The fourth data release of the Kilo-Degree Survey: *ugri* imaging and nine-band optical-IR photometry over 1000 square degrees

 THE Kilo-Degree Survey (KiDS) is an ongoing optical wide-field imaging survey with the OmegaCAM camera at the VLT Survey Telescope, specifically designed for measuring weak gravitational lensing by galaxies and large-scale structure. When completed it will consist of 1350 square degrees imaged in four filters (*ugri*). Here we present the fourth public data release which more than doubles the area of sky covered by data release 3. We also include aperture-matched *ZYJHK<sub>s</sub>* photometry from our partner VIKING survey on the VISTA telescope in the photometry catalogue. We illustrate the data quality and describe the catalogue content. Two dedicated pipelines are used for the production of the optical data. The ASTRO-WISE information system is used for the production of co-added images in the four survey bands, while a separate reduction of the *r*-band images using the THELI pipeline is used to provide a source catalogue suitable for the core weak lensing science case. All data have been re-reduced for this data release using the latest versions of the pipelines. The VIKING photometry is obtained as forced photometry on the THELI sources, using a re-reduction of the VIKING data that starts from the VISTA paw-prints. Modifications to the pipelines with respect to earlier releases are described in detail. The photometry is calibrated to the Gaia DR2 *G* band using stellar locus regression. In this data release a total of 1006 square-degree survey tiles with stacked *ugri* images are made available, accompanied by weight maps, masks, and single-band source lists. We also provide a multi-band catalogue based on *r*-band detections, including homogenized photometry and photometric redshifts, for the whole dataset. Mean limiting magnitudes ( $5\sigma$  in a  $2''$  aperture) and the tile-to-tile rms scatter are  $24.23 \pm 0.12$ ,  $25.12 \pm 0.14$ ,  $25.02 \pm 0.13$ ,  $23.68 \pm 0.27$  in *ugri*, respectively, and the mean *r*-band seeing is  $0''.70$ .

K. Kuijken, C. Heymans, A. Dvornik, H. Hildebrandt,  
J.T.A. de Jong, A.H. Wright, et al.  
*A&A*, Volume 625, Issue 1, p. A2 (2019)

## 6.1 INTRODUCTION: THE KILO-DEGREE AND VIKING SURVEYS

High-fidelity images of the sky are one of the most fundamental kinds of data for astronomy research. While for many decades photographic plates dominated optical sky surveys, the advent of large-format CCD detectors for astronomy opened up the era of digital, high resolution, high sensitivity, linear-response images.

The ESO VLT Survey Telescope (VST; Capaccioli & Schipani 2011) at ESO's Paranal observatory was specifically designed for wide-field, optical imaging. Its focal plane contains the square 268-million pixel CCD mosaic camera OmegaCAM (Kuijken 2011) that covers a  $1^{\circ}013 \times 1^{\circ}020$  area at  $0''.213$  pitch, and the site and telescope optics (with actively controlled primary and secondary mirrors) ensure an image quality that is sub-arcsecond most of the time, and that does not degrade towards the corners of the field. Since starting operations in October 2011, more than half of the available time on the telescope has been used for a set of three wide-area 'Public Imaging Surveys' for the ESO community. The Kilo-Degree Survey (KiDS; de Jong et al. 2013)<sup>1</sup> is the deepest of these, and the one that exploits the best observing conditions.

KiDS was designed as a cosmology survey, to study the galaxy population out to redshift  $\sim 1$  and in particular to measure the effect on galaxy shapes due to weak gravitational lensing by structure along the line of sight. By combining galaxy shapes with photometric redshift estimates it is possible to locate the redshift at which the gravitational lensing signal originates, and hence to map out the growth of large-scale structure, an important aspect of the evolution of the Universe and a key cosmology probe. Together with KiDS, two other major surveys are engaged in such measurements: the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005)<sup>2</sup> and the HyperSuprimeCam survey (HSC; Aihara et al. 2018)<sup>3</sup>, and all three have reported intermediate cosmology results (Hildebrandt et al. 2017, henceforth [KiDS450]; Troxel et al. 2018; Hikage et al. 2019). Their precision is already such that the measurements can constrain some parameters in the cosmological model to a level that is comparable to what is achieved from the cosmic microwave background anisotropies (Planck Collaboration et al. 2018). Since ground-based surveys are limited fundamentally by the atmospheric disturbance on galaxy shapes and photometry, space missions Euclid (Laureijs et al. 2011) and later WFIRST (Spergel et al. 2015) are planned to increase the fidelity of such studies further.

To meet its primary science goal KiDS observes the sky in four bands:  $u$ ,  $g$ ,  $r$  and  $i$ . The  $r$  band is used in dark time during the best seeing conditions ( $\text{FWHM} < 0''.8$ ), to make deep images for the measurement of galaxy shapes. In order to provide colours for photometric redshift estimates of the same sources, the  $r$ -band data are supplemented with  $g$ - and  $u$ -band data taken in dark time of progressively worse seeing conditions ( $< 0''.9$  and  $< 1''.1$ , respectively), and with  $i$ -band data taken in grey or bright moon time with a mild seeing constraint ( $< 1''.1$ ). All observations consist of

---

<sup>1</sup><http://kids.strw.leidenuniv.nl>

<sup>2</sup><http://darkenergysurvey.org>

<sup>3</sup><http://hsc.mtk.nao.ac.jp/ssp/>



Table 6.1: KiDS observing strategy: observing condition constraints and exposure times.

Filter	Max. lunar illumination	Min. moon distance [deg]	Max. seeing [arcsec]	Max. airmass	Sky transp.	Dithers	Total Exp. time [s]
<i>u</i>	0.4	90	1.1	1.2	CLEAR	4	1000
<i>g</i>	0.4	80	0.9	1.6	CLEAR	5	900
<i>r</i>	0.4	60	0.8	1.3	CLEAR	5	1800
<i>i</i>	1.0	60	1.1	2.0	CLEAR	5	1200

Table 6.2: Approximate boundaries of the KiDS fields (see also Fig. 6.1).

Field	RA range	Dec range
KiDS-S	[330°0, 52°5]	[−35°6, −26°6]
KiDS-N	[155°5, 225°5] [225°5, 238°5]	[−4°0, +4°0] [−2°0, +3°0]
KiDS-N-W2	[128°5, 141°5]	[−2°0, +3°0]
KiDS-N-D2	[149°5, 150°5]	[+1°7, +2°7]

multiple dithered exposures to minimize the effect of gaps between the CCD’s in the mosaic. Observing constraints and exposure times are summarized in Table 6.1.

KiDS is targeting around 1350 square degrees of extragalactic sky, in two patches to ensure year-round observability. The Northern patch, KiDS-N, contains two additional smaller areas: KiDS-N-W2, which coincides with the G9 patch of the GAMA survey (Driver et al. 2011), and KiDS-N-D2, a single pointing on the COSMOS field. In a coordinated effort over the same part of the sky, the VISTA Kilo-degree INfrared Galaxy survey (VIKING; Edge et al. 2013) on the nearby VISTA telescope added the five bands *Z*, *Y*, *J*, *H* and *K<sub>s</sub>*. VIKING observations are complete<sup>4</sup> and available in the ESO archive<sup>5</sup>. Table 6.2 and Fig. 6.1 show the full KiDS footprint on the sky, as well as the part that is covered by the data contained in this data release (KiDS-ESO-DR4, or DR4 for short). The fields that were previously released under DR1+2+3 are also indicated: this is the area that was used for the [KiDS450] cosmic shear analysis, with the corresponding shape/photometric redshift catalogue released as DR3.1.

In order to improve the fidelity of the photometric redshift-based tomography, and to enable inclusion of high-value sources in the redshift range 0.9–1.2, Hildebrandt et al. (2018) added VIKING photometry to the KiDS-450 data set, as described in Wright et al. (2018). The resulting cosmological parameter constraints of this new analysis, dubbed ‘KV450’, are fully consistent with [KiDS450]. DR4 incorporates the methodology developed for KV450 and includes VIKING photometry for all sources. Of all wide-area surveys, this makes it the one with by far the deepest near-IR data.

<sup>4</sup>The originally planned KiDS area was 1500 square degrees, but this was reduced to match the footprint of the VIKING area.

<sup>5</sup><http://archive.eso.org>

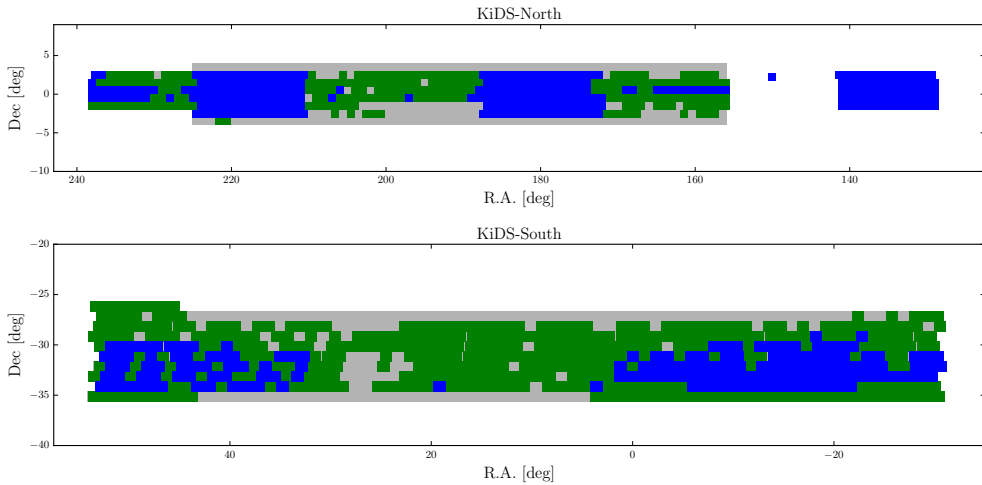


Figure 6.1: Sky distribution of survey tiles released in KiDS-ESO-DR4. Tiles shown in green are released for the first time; those in blue were included in the earlier data releases (DR1+2+3) but have been reprocessed for DR4. The full KiDS+VIKING area ( $\sim 1350 \text{ deg}^2$ ) is shown in grey. *Top*: KiDS-North. *Bottom*: KiDS-South. The single pointing at RA=150 deg is centred on the COSMOS/CFHTLS D2 field.

Though designed for the primary cosmology science case ([KiDS450]; Joudaki et al. 2017; van Uitert et al. 2018b; Köhlinger et al. 2017; Harnois-Déraps et al. 2017; Amon et al. 2018b; Shan et al. 2018; Martinet et al. 2018; Giblin et al. 2018; Asgari et al. 2019), KiDS data are also being used for a variety of other studies, including the galaxy-halo connection (van Uitert et al. 2016, 2018a), searches for strongly lensed galaxies (Petrillo et al. 2018) and quasars (Spiniello et al. 2018; Sergeev et al. 2018), solar system objects (Mahlke et al. 2018), photometric redshift machine learning method development (Amaro et al. 2019; Bilicki et al. 2018), studies of galaxy evolution (Tortora et al. 2018a,b; Roy et al. 2018), bias (**Chapter 3**), environment (Brouwer et al. 2016, 2018; Costa-Duarte et al. 2018) and morphology (Kelvin et al. 2018), galaxy group properties (Viola et al. 2015; Jakobs et al. 2018), galaxy cluster searches (Maturi et al. 2019; Bellagamba et al. 2019), intrinsic alignment of galaxies (Georgiou et al. 2019,?), satellite halo masses (Sifón et al. 2015), and searches for luminous red galaxies (Vakili et al. 2019) and quasars (Nakoneczny et al. 2019).

The outline of this paper is as follows. Sect. 6.2 is a discussion of the contents of KiDS-ESO-DR4. Sect. 6.3 summarises the differences in terms of processing and data products with respect to earlier releases. Sects. 6.4 and 6.5 describe the single-band data products and the KiDS+VIKING nine-band catalogue, respectively. Sect. 6.6 illustrates the data quality. Data access routes are summarised in Sect. 6.7 and a summary and outlook towards future data releases is provided in Sect. 6.8. The Appendix gives a full listing of the information included in the images and catalogues, and of the data structure.

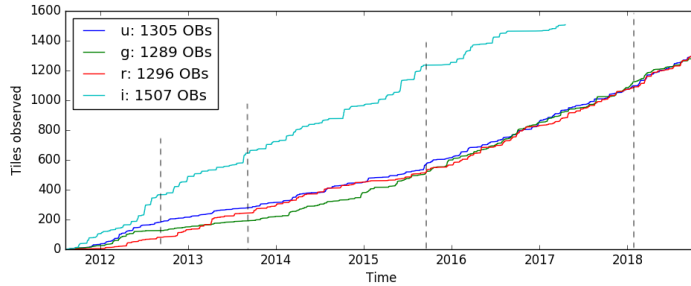


Figure 6.2: Progress of the KiDS observations at the VST, in the four survey bands. Each Observing Block (OB) produces a square-degree co-added image. The *i*-band data, for which data taking was significantly faster because of less competition for bright time on the telescope, had covered the originally planned 1500 square degree footprint by the time it was decided to limit the survey to the 1350 square degree area that comprise the completed VIKING area. The dashed lines indicate the cutoff dates for KiDS-ESO data releases 1 to 4.

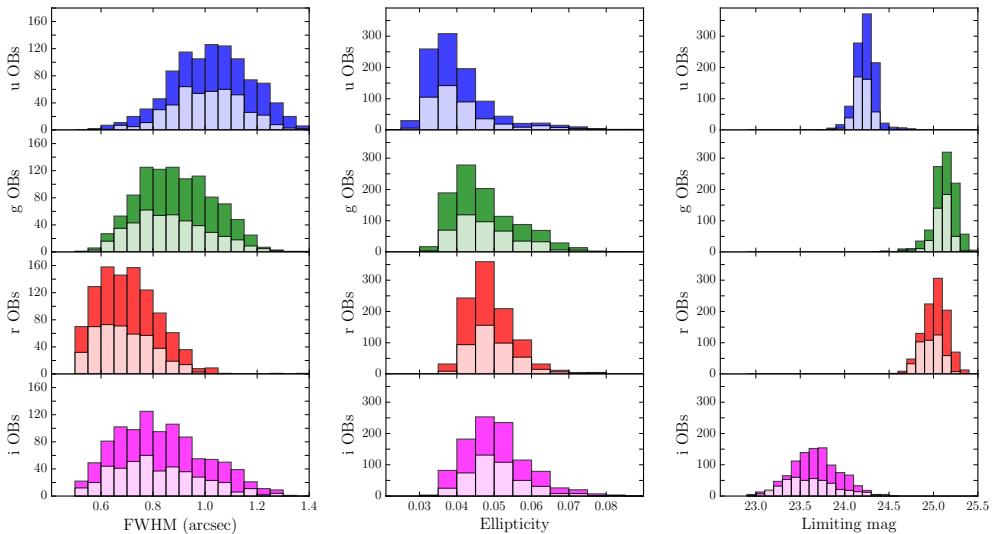


Figure 6.3: Distributions of tile-by-tile data quality parameters for the KiDS DR4 data, grouped by filter, from top to bottom *u*, *g*, *r* and *i*. The light-coloured histograms represent the subset of the data that was previously released in DR1+2+3. Left: seeing. The differences between the bands reflect the observing strategy of reserving the best-seeing dark time for *r*-band observations. Middle column: Average PSF ellipticity  $\langle |e_{\text{psf}}| \rangle$ , where  $e$  is defined as  $1 - b/a$  for major/minor axis lengths  $a$  and  $b$ . Right: Limiting AB magnitude ( $5\text{-}\sigma$  in a  $2''$  aperture). The wider distribution of the *i*-band observations is caused by variations in the moon illumination, since the *i*-band data were mostly taken in bright time.

## 6.2 THE FOURTH KIDS DATA RELEASE

Unlike the previous incremental KiDS data releases, KiDS-ESO-DR4 represents a complete re-reduction of all the data using improved pipeline recipes and procedures. The differences will be described below. In terms of content of the data release, the main changes with respect to the earlier releases (de Jong et al. 2015, 2017) are the more than doubling of the area, and the inclusion of photometry from the near-IR VIKING images into the multi-band catalogue. Whereas the sky coverage of the earlier data releases was still quite fragmented, DR4's greater homogeneity will for the first time enable wide-area studies over the full length of the survey patches.

KiDS observations consist of individual square-degree tiles. Each tile is covered by a set of five (four in  $u$ ) dithered exposures with OmegaCAM/VST, consisting of 32 individual CCD images each. The dither step sizes are matched to the gaps between CCD's ( $25''$  in RA,  $85''$  in declination), to ensure that each part of the tile is covered by at least three (two in  $u$ ) sub-exposures. Overlaps between adjacent tiles are small, of order 5%. All observations in a single band are taken in immediate succession (KiDS is not designed for variability measurements), but there is no constraint on the time between observations of any given tile in the different filters. Typically the shutter is closed for 35-60 seconds between the sub-exposures to allow for CCD readout, telescope repointing and active optics adjustments.

With little exception, the DR4 data comprise all KiDS tiles for which the 4-band observations had been taken by January 24th, 2018. Over half of the data is from after mid-2015, which is when the VST saw several improvements that affect the quality of the data (and improved operational efficiency as well, see Fig. 6.2). The two main improvements were (i) the baffling of the telescope was improved to the point that stray light from sources outside the field of view of the camera was drastically reduced; and (ii) the on-line image analysis system (based on simultaneous pre- and post-focus star images at the edge of the field of view) was modified to control also the tilt of the secondary mirror, improving pointing and especially off-axis image quality.

Figure 6.3 shows the distributions of key data quality parameters of the observations: point spread function (PSF) full width at half maximum (FWHM), average PSF ellipticity and limiting magnitude. It illustrates that the global quality of the DR4 data is very similar to the earlier KiDS data releases. Limiting AB magnitudes ( $5\text{-}\sigma$  in  $2''$  aperture) are  $24.23 \pm 0.12$ ,  $25.12 \pm 0.14$ ,  $25.02 \pm 0.13$ ,  $23.68 \pm 0.27$  in  $ugri$ , respectively, with the error bars representing the RMS scatter from tile to tile. The mean seeing in  $r$  band is  $0''.70$ .

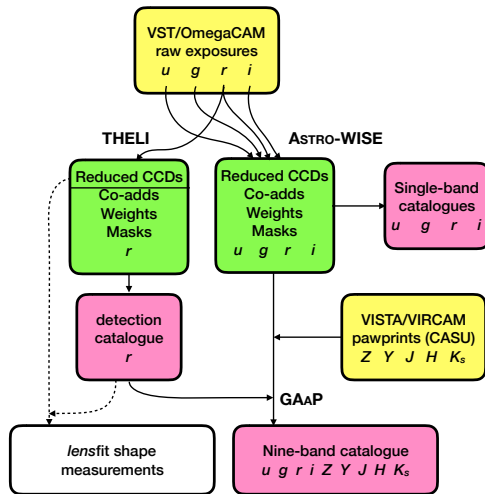


Figure 6.4: Schematic representation of the DR4 processing steps and content. Yellow boxes show the input data from VST and VISTA, green indicates image products, and source lists are shown in pink. The *lensfit*-based lensing measurements – initially not released in DR4 – are shown as dotted lines.

KiDS is targeting around 1350 square degrees of extragalactic sky, in two patches to ensure year-round observability. The Northern patch, KiDS-N, contains two additional smaller areas: KiDS-N-W2, which coincides with the G9 patch of the GAMA survey (Driver et al. 2011), and KiDS-N-D2, a single pointing on the COSMOS field. In a coordinated effort over the same part of the sky, the VISTA Kilo-degree INfrared Galaxy survey (VIKING; Edge et al. 2013) on the nearby VISTA telescope added the five bands  $Z$ ,  $Y$ ,  $J$ ,  $H$  and  $K_s$ . VIKING observations are complete<sup>6</sup> and available in the ESO archive<sup>7</sup>. Table 6.2 and Fig. 6.1 show the full KiDS footprint on the sky, as well as the part that is covered by the data contained in this data release (KiDS-ESO-DR4, or DR4 for short). The fields that were previously released under DR1+2+3 are also indicated: this is the area that was used for the [KiDS450] cosmic shear analysis, with the corresponding shape/photometric redshift catalogue released as DR3.1.

In order to improve the fidelity of the photometric redshift-based tomography, and to enable inclusion of high-value sources in the redshift range 0.9–1.2, Hildebrandt et al. (2018) added VIKING photometry to the KiDS-450 data set, as described in Wright et al. (2018). The resulting cosmological parameter constraints of this new analysis, dubbed ‘KV450’, are fully consistent with [KiDS450]. DR4 incorporates the methodology developed for KV450 and includes VIKING photometry for all sources. Of all wide-area surveys, this makes it the one with by far the deepest near-IR data.

<sup>6</sup>The originally planned KiDS area was 1500 square degrees, but this was reduced to match the footprint of the VIKING area.

<sup>7</sup><http://archive.eso.org>

The KiDS images were processed with two independent pipelines, as was the case for the KiDS-450 weak lensing analysis that was based on KiDS-ESO-DR3. The ASTRO-WISE pipeline and data reduction environment (McFarland et al. 2013)<sup>8</sup> was used to produce stacked images in the four bands, from which the photometry in the catalogues is obtained. The THELI pipeline (Erben et al. 2005)<sup>9</sup>, which is optimised for weak lensing measurements, was used for a separate reduction of the  $r$ -band data only. In order to have consistent source lists, the detection and astrometry of the  $r$ -band sources is performed on these THELI images (which are also the ones which are used for the weak lensing measurements). This detection catalogue is then used as the basis of list-driven ‘forced’ photometry on the  $u$ ,  $g$ ,  $r$  and  $i$  ASTRO-WISE stacked images and the VIKING  $Z$ ,  $Y$ ,  $J$ ,  $H$  and  $K_s$  images.<sup>10</sup> The data flow is summarized in Fig. 6.4.

The set of KiDS-ESO-DR4 data products includes 5030 separate co-added images (1006 square-degree tiles in the  $ugri$  filters, plus the separate  $r$ -band co-adds from THELI), with corresponding weight and mask flag images. The images are photometrically and astrometrically calibrated using a combination of nightly photometric calibration information, the Gaia DR2 photometry (Brown et al. 2018) with stellar locus regression, and the SDSS and 2MASS astrometry (Alam et al. 2015; Skrutskie et al. 2006). Each image has a corresponding source catalogue as well. In addition, catalogues of nine-band  $ugriZYJHK_s$  photometry are provided containing list-driven (i.e., forced), PSF- and aperture-matched photometry using the GAAP technique (Kuijken et al. 2015), applied to the KiDS tiles and the overlapping VIKING data.

## 6.3 DATA PROCESSING

For details of the image processing pipelines we refer to the description in the DR1/2 and DR3 release papers (de Jong et al. 2015, henceforth [DR1/2]) and (de Jong et al. 2017, henceforth [DR3]), noting the changes described in Sects. 6.3.1 and 6.3.2 below.

The DR4 catalogues are of two types: single-band catalogues for the four KiDS bands, and a combined nine-band catalogue that includes KiDS and VIKING photometry for the  $r$ -band detected sources.

### 6.3.1 CHANGES TO THE ASTRO-WISE IMAGE PROCESSING PIPELINE

#### Co-added image creation

The production of KiDS-ESO-DR4 includes  $4 \times 1006$  OmegaCAM tiles, which required some 611,000 individual CCD exposures, as well as associated calibration observations, to be processed. The automatic processing steps followed closely those described in [DR3]. Individual CCD exposures are corrected for electronic crosstalk,

<sup>8</sup><http://www.astro-wise.org>

<sup>9</sup><https://www.astro.uni-bonn.de/theli/>

<sup>10</sup>Note that for DR4 there is no separate multi-band catalogue based on  $r$ -band detections in the ASTRO-WISE data, as there was for [DR3].

bias corrected, flat-fielded, illumination corrected and (for *i* band only) fringe subtracted. New coefficients were determined for the electronic crosstalk correction between CCD's ES0\_CCD\_#95 and ES0\_CCD\_#96 (see [DR1/2]) with validity periods determined by maintenance or changes to the instrument; these are listed in Table 6.3. An automatic mask is also generated for each CCD, which marks the location of saturated, hot, and cold pixels, as well as satellite tracks identified through a Hough transform analysis.

These 'reduced science frames' are then astrometrically calibrated and regridded using the SCAMP and SWARP software (Bertin et al. 2002; Bertin 2006, 2010a,b), in two steps: first a 'local' step establishes a per-detector solution using the 2MASS stars in the frame<sup>11</sup>, and then these solutions are refined using SCAMP into a tile-wide 'global solution' for the full co-added stacked image that uses the information from fainter overlapping objects. For DR4 the astrometric solution was made more robust by starting from a model of the focal plane that accounts for detector array lay-out and instrument optics distortion, and using this as input to SCAMP. Using SWARP, the global astrometric solution is then used to resample each CCD exposure into a 'regridded science frame', with a uniform 0'20 pixel grid with tangent projection centred on the nominal tile centre. During this step the background, determined by interpolating a  $3 \times 3$  median-filtered map of background estimates in  $128 \times 128$  pixel blocks, is subtracted. Finally these regridded images are co-added, taking account of the weight maps generated by SWARP, and the masks. Each co-added image is about  $18,500 \times 19,500$  pixels in size, and takes up about 1.5 Gbyte of storage. For every co-added image a mask that flags reflection haloes of bright stars in the field is also produced, using the PULECENELLA code developed for [DR1/2] (see Sect. 6.4).

$12 \times 12$  binned versions of the co-added images (at two contrast settings) and of the weight image are then visually inspected, together with a set of diagnostic plots that show the PSF ellipticity and size as function of position on the field, astrometry solution residuals, and the PSF size before and after co-addition. The main issues that get flagged at this stage are (i) residual satellite tracks (ii) background features associated with stray light casting shadows of the baffles mounted above the CCD bond wires (iii) unstable CCDs (gain jumps) (iv) residual fringing in the background of the *i*-band images or (v) large-scale reflections. In DR3 any issues found at this stage were addressed by masking the co-added image, even though in many cases the problem only affected one sub-exposure. In DR4 issues (i)–(iii) were solved with new procedures, as described below. The other cases are still in the data: the residual fringes are not corrected for but will be fixed with re-observations of the full survey footprint in the *i* band, and large-scale reflections need to be masked manually or otherwise identified in the catalogues as groups of sources with unusual colours.

The new procedures for removing residual satellite tracks, bond wire baffle features, and unstable CCDs, involve a minimum of manual intervention. The satellite tracks are marked by clicking on their ends on a display of the inspection JPG images, after which an automatic procedure converts the pixel positions to sky coordinates, checks which of the sub-exposures that contribute to the co-added image contains

---

<sup>11</sup>We have found the 2MASS catalogue to be sufficient as astrometric reference, but intend to move to Gaia in the future.

Table 6.3: Applied cross-talk coefficients.

Period	CCD #95 to CCD #96 <sup>a</sup>		CCD #96 to CCD #95 <sup>a</sup>	
	$a$	$b (\times 10^{-3})$	$a$	$b (\times 10^{-3})$
2011-08-01 - 2011-09-17	-210.1	-2.504	59.44	0.274
2011-09-17 - 2011-12-23	-413.1	-6.879	234.8	2.728
2011-12-23 - 2012-01-05	-268.0	-5.153	154.3	1.225
2012-01-05 - 2012-07-14	-499.9	-7.836	248.9	3.110
2012-07-14 - 2012-11-24	-450.9	-6.932	220.7	2.534
2012-11-24 - 2013-01-09	-493.1	-7.231	230.3	2.722
2013-01-09 - 2013-01-31	-554.2	-7.520	211.9	2.609
2013-01-31 - 2013-05-10	-483.7	-7.074	224.7	2.628
2013-05-10 - 2013-06-24	-479.1	-6.979	221.1	2.638
2013-06-24 - 2013-07-14	-570.0	-7.711	228.9	2.839
2013-07-14 - 2014-01-01	-535.6	-7.498	218.9	2.701
2014-01-01 - 2014-03-08	-502.2	-7.119	211.6	2.429
2014-03-08 - 2014-04-12	-565.8	-7.518	215.1	2.578
2014-04-12 - 2014-08-12	-485.1	-6.887	201.6	2.237
2014-08-12 - 2014-01-09	-557.9	-7.508	204.2	2.304
2014-01-09 - 2015-05-01	-542.5	-7.581	219.9	2.535
2015-05-01 - 2015-07-25	-439.3	-6.954	221.5	2.395
2015-07-25 - 2015-08-25	-505.6	-7.535	229.7	2.605
2015-08-25 - 2015-11-10	-475.2	-7.399	218.0	2.445
2015-11-10 - 2016-06-17	-457.8	-6.831	201.6	2.212
2016-06-17 - 2016-06-25	-351.8	-4.973	165.3	1.168
2016-06-25 - 2016-09-08	-476.3	-6.920	200.4	2.202
2016-09-08 - 2017-08-01	-465.3	-6.594	184.7	1.980
2017-08-01 - 2018-02-15	-492.3	-6.480	169.9	1.802

<sup>a</sup> Correction factors  $a$  and  $b$  are applied to each pixel in the target CCD based on the pixel values in the source CCD:

$$I'_i = \begin{cases} I_i + a, & \text{if } I_j = I_{\text{sat.}}; \\ I_i + bI_j, & \text{if } I_j < I_{\text{sat.}}. \end{cases} \quad (6.1)$$

where  $I_i$  and  $I_j$  are the pixel values in CCDs  $i$  and  $j$ ,  $I'_i$  is the corrected pixel value in CCD  $i$  due to cross-talk from CCD  $j$ , and  $I_{\text{sat.}}$  is the saturation pixel value.



the track, measures the track's width, and updates the corresponding CCDs' masks before stacking anew. Similarly, the bond wire baffle features have a typical width and all the inspector needs to do is to indicate whether the shadow is visible on the upper, lower or both sides of the baffles, so that the corresponding lines in the CCD images can be masked. Unstable CCDs are simply removed from the list of exposures to be co-added.

### PSF Gaussianization and GAAP photometry

The Gaussian Aperture and PSF (GAAP) photometry method (Kuijken et al. 2015) developed for KiDS multi-band photometry was improved further. GAAP entails (i) convolving each image with a spatially variable kernel designed to render the PSF homogeneous and Gaussian, (ii) defining a pre-seeing Gaussian elliptical aperture function for every source, and (iii) for every band deconvolving this aperture by the corresponding Gaussianized PSF and performing aperture photometry. The method is superior to traditional techniques such as dual-image mode SExtractor measurements as it explicitly allows for PSF differences between exposures in multiple bands, and it reduces noise by measuring colours from the highest SNR part of the sources. A demonstration of the improvement provided by GAAP photometry is given in Hildebrandt et al. (2012). GAAP photometry gives colours that are corrected for PSF differences, but when the source is more extended than the aperture function the fluxes are underestimates of the total flux. For stars and other unresolved sources GAAP fluxes *are* total fluxes.

For DR4 we have modified the procedure for step (i). We still use the several thousand stars in each image as samples of the PSF, but rather than first modelling the PSF  $P$  as a spatially varying, truncated shapelet expansion, from which the convolution kernel is then constructed in shapelet coefficient space, we now directly solve for the kernel shapelet coefficients that give a Gaussian PSF in pixel space. Thus we obtain the kernel coefficients  $k_{abc}$  of the shapelet components  $S_{ab}^{\beta_c}$  as the least-squares solution of

$$\sum_{abc} k_{abc} [S_{ab}^{\beta_c} \otimes P](x, y) = \frac{\exp[-(x^2 + y^2)/2\beta_g^2]}{2\pi\beta_g^2}. \quad (6.2)$$

As in [DR3], the size  $\beta_g$  of the target Gaussian is set to 1.3 times the median dispersion of Gaussian fits to the stars in the image. The fit of Eq. (6.2) is performed over all pixels out to  $15\beta_g$  from the centre of each star. In our implementation we use terms with  $\beta_1 = \beta_g$ ,  $a + b \leq 8$  plus a set of wider shapelets with  $\beta_2 = 2.5\beta_g$ ,  $3 \leq a + b \leq 6$  specifically designed to model the wings of the kernel better. The large- $\beta$  shapelets with  $a + b < 3$  are not included in the series as they are not sufficiently orthogonal to the small- $\beta$  terms.

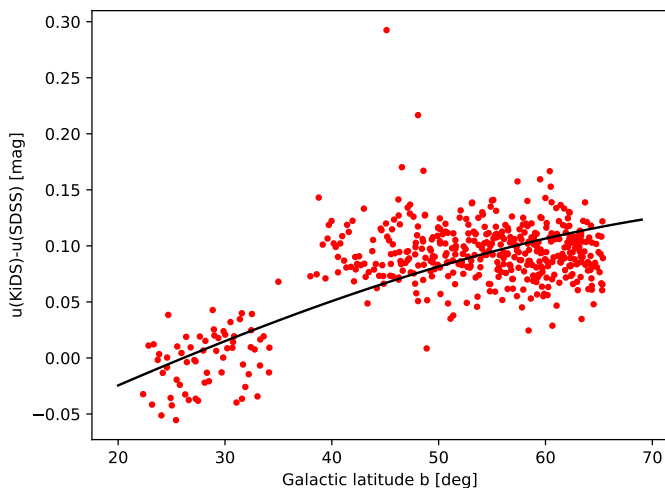


Figure 6.5: Residual  $u$ -band magnitude variation when Eq. (6.3) is used without a dependence on Galactic latitude  $b$  (i.e.,  $f(b) = 0$ ). Each point shows the average offset  $u_{\text{KiDS}} - u_{\text{SDSS}}$  for the calibration stars in a separate KiDS-N tile. The line represents the Galactic latitude correction of Eq. (6.4).

### Photometric calibration using Gaia and stellar locus regression

All KiDS ‘reduced science frame’ images are initially put on a photometric scale by using nightly zeropoints derived from standard star observations taken in the middle of the night. For DR4, these zeropoints are refined with a combination of Stellar Locus Regression (SLR) and calibration to Gaia photometry. We use GAAP photometry for the stars: this is appropriate since for unresolved sources the GAAP magnitudes correspond to the total flux of the source.

With the advent of Gaia Data Release 2 (Brown et al. 2018) a deep, homogeneously calibrated, optical all-sky catalogue is now available. Each KiDS tile contains several thousand Gaia stars, with broad-band photometry measurements that are individually accurate to better than 0.01 magnitudes<sup>12</sup>. KiDS DR4 photometry is calibrated to the Gaia DR2 catalogue in two steps. First, we calibrate the colours  $u - g$ ,  $g - r$  and  $r - i$  by comparing the stellar colour-colour diagrams to fiducial sequences, using the ‘stellar locus regression’ described in [DR3]. Based on the work of Ivezić et al. (2004), four principal colours  $P2s$ ,  $P2w$ ,  $P2x$  and  $P2k$  were initially used to derive colour offsets (see appendix B of [KiDS450]). After this procedure, while validating the results by comparing the magnitudes of stars in KiDS-N and SDSS, we found that the  $P2k$  principal colour, which is the most sensitive to the  $u$  band, gave unreliable re-

<sup>12</sup>Note that the Gaia DR2  $g$ -band calibration differs from what was used in Gaia DR1, through a new determination of the filter bandpass.

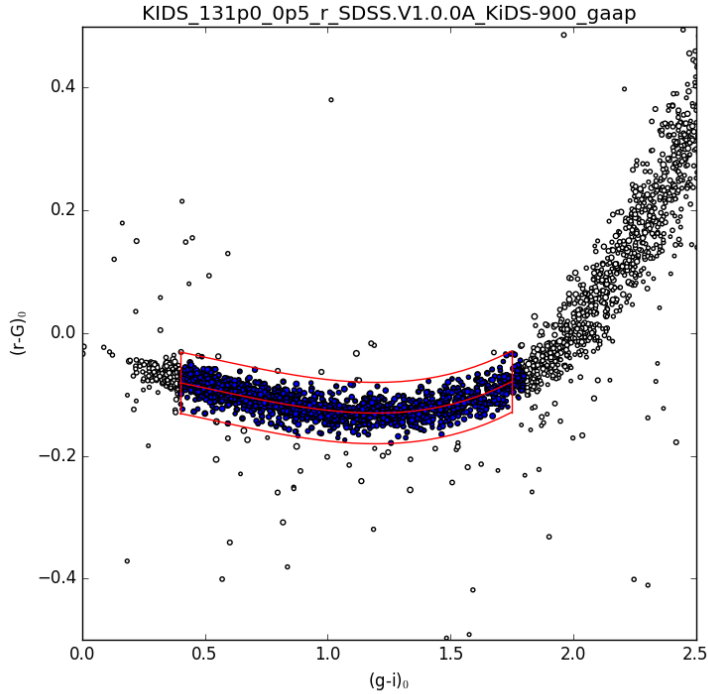


Figure 6.6: Colour-colour relation used to calibrate the KiDS  $r$ -band measurements to the Gaia  $G$ -band catalogue, for an example tile. The blue points and the box indicate the dereddened  $(g-i)$  colour range for the stars used, and the  $\pm 0.05$  magnitude iterative clipping width about the fiducial sequence. The shape of the sequence is determined from the overlap area between KiDS-N and SDSS.

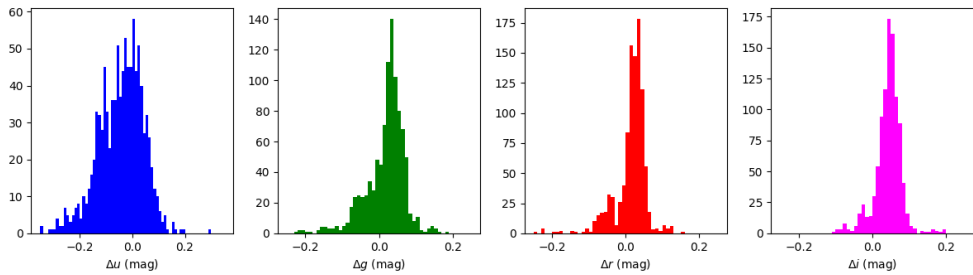


Figure 6.7: Distribution of the  $u$ ,  $g$ ,  $r$  and  $i$  zero-point corrections using the stellar locus regression plus Gaia calibration.

sults. The  $u$ -band zeropoints were therefore determined using a modified procedure, which was found to be more robust, as follows. For stars in Gaia with dereddened KiDS GAAP colour  $(g - r)_0$  between 0.15 and 0.8 and  $u < 21$ , the quantity

$$\mathcal{U} = g_0 + 2(g - r)_0 + 0.346 + \max\{0, 2[0.33 - (g - r)_0]\} - f(b) \quad (6.3)$$

is a good predictor of the SDSS  $u$  magnitude. The Galactic latitude dependence  $f(b)$ , shown in Fig. 6.5, can be fitted as

$$f(b) = 0.25|\sin b| - 0.11. \quad (6.4)$$

Stars closer to the Galactic plane are fainter in  $u$  than the mean  $(u - g, g - r)$  colour-colour relation predicts, as is expected if the high-latitude sample is dominated by halo stars of lower metallicity than the disk stars found at low  $b$ . Since KiDS-N spans latitudes from  $22^\circ$  to  $66^\circ$ , the correction is significant. KiDS-S covers latitudes from the South Galactic Pole to  $-53^\circ$ , and though we cannot test the latitude dependence for this part of the survey because of a lack of a suitable calibration data set, we have applied the same correction. We therefore adjust the KiDS  $u$ -band zeropoints in each tile until the average  $\mathcal{U} - u = 0$ , with an iterative clipping of outliers more than 0.1 mag from the mean relation. For applications where  $u$ -band photometry is critical, we caution that because of the uncertain functional form of the latitude dependence, currently the calibration of this band is subject to a residual uncertainty of up to 0.05 magnitudes.

The zeropoint of the  $r$ -band magnitude is then tied to Gaia by matching the dereddened  $(r - G, g - i)$  relation to the one followed by the stars in the SDSS-KiDS overlap region (Fig. 6.6). As reported in [DR3], there is a slight colour term between the SDSS and KiDS  $r$  filters: we have arbitrarily forced the KiDS and SDSS  $r$ -band zeropoints to agree for stars of  $(g - i)_0 = 0.8$ , adopting

$$r_{\text{KiDS}} - r_{\text{SDSS}} = -0.02[(g - i)_0 - 0.8]. \quad (6.5)$$

For DR4, extinction corrections were derived using the Schlegel et al. (1998)  $E(B - V)$  map in combination with the  $R_V = 3.1$  extinction coefficients from Schlafly & Finkbeiner (2011)<sup>13</sup>. For the  $ugri$  filters we adopt the corresponding SDSS filter values. Since the VISTA bands were not included in the Schlafly & Finkbeiner (2011) tables, as an approximation we have taken the values for SDSS  $z$ , LSST  $y$ , and UKIRT  $JHK$  filters. From a regression of  $r_{\text{SDSS}} - G$  vs.  $E(B - V)$  we derive the  $G$ -band extinction coefficient as  $A_G/A_r = 0.96$ . Given that  $E(B - V)$  values in the KiDS tiles are typically below 0.05 magnitudes, residual uncertainties in these coefficients are of little consequence. The adopted extinction coefficients are summarised in Table 6.4.

<sup>13</sup>The earlier data releases used the Schlegel et al. (1998) coefficients.

Table 6.4: Extinction coefficients  $R_f = A_f/E(B-V)$  used in this work, from Schlafly & Finkbeiner (2011) (SF11). These coefficients are used to scale the  $E(B-V)$  values in the Schlegel et al. (1998) map.

Filter $f$	$R_f$	Source
$u$	4.239	SF11 (SDSS)
$g$	3.303	SF11 (SDSS)
$r$	2.285	SF11 (SDSS)
$i$	1.698	SF11 (SDSS)
$G$	2.194	This work
$Z$	1.263	SF11 (SDSS $z$ )
$Y$	1.088	SF11 (LSST $y$ )
$J$	0.709	SF11 (UKIRT)
$H$	0.449	SF11 (UKIRT)
$K_s$	0.302	SF11 (UKIRT)

This direct, tile-by-tile calibration of the KiDS photometry to Gaia obviates the need for the overlap photometry that was used in [DR3]. Effectively, we are tying KiDS to the Gaia DR2  $G$ -band photometry, and anchoring it to the SDSS calibration. Specifically, calibrating each KiDS tile to Gaia DR2 involves the following steps:

1. Select Gaia stars in the tile area with  $16.5 < G < 20$ , and with unflagged photometric measurements.
2. Keep those stars with SLR-calibrated, dereddened  $(g-i)_0$  colours in the range  $[0.4, 1.8]$ .
3. Predict dereddened  $(r-G)_0$  values from these  $(g-i)_0$  colours using the following relation, obtained by fitting the difference between the predicted  $r_{\text{KiDS}}$  (from Eq. 6.5) and the measured  $G$  in the KiDS-SDSS overlap region:

$$(r_{\text{KiDS}} - G)_0 = -0.0618 - 0.0724y + 0.0516y^2 + 0.0665y^3 \quad (6.6)$$

where  $y = (g-i)_0 - 0.8$  (see Fig. 6.6).

4. Determine the median offset between this fiducial  $r-G$  and the measured value, using iterative clipping.
5. Apply this median offset to the  $ugri$  magnitudes for all sources in the tile.

Figure 6.7 shows the distribution of tile-by-tile zeropoint corrections that have been applied to the magnitudes in the catalogues. Typical values are of the order of 0.05–0.1 magnitudes, and about twice that in the  $u$  band.

Note that the SLR procedure aligns the dereddened stellar loci of all the tiles, assuming that all the dust is in the foreground and not mixed in with the stars (which is a reasonable assumption given the high Galactic latitude and bright-end limit of the calibration stars).

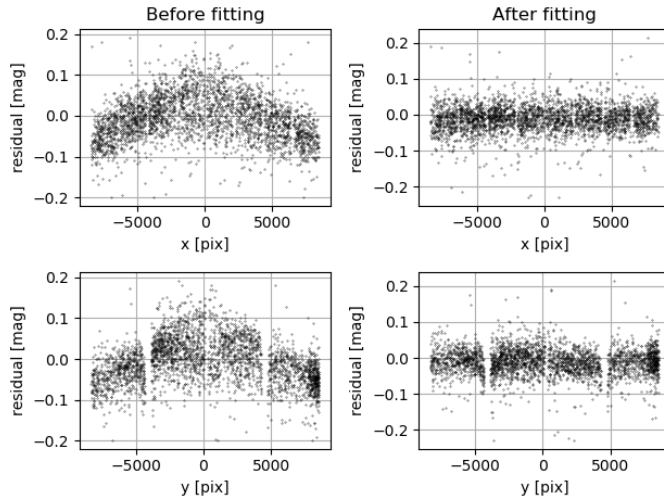


Figure 6.8: Photometric illumination correction of the weak-lensing THELI data: Flat-fielded OmegaCAM data show systematic zeropoint variations over the field-of-view if the complete mosaic is calibrated with a single photometric zeropoint (left panels). The residuals can well be fitted and corrected with a two-dimensional, second-order polynomial over the field-of-view (right panels)

In DR4 all GAAP photometry is performed twice, with minimum aperture settings  $\text{MIN\_APER}=0''.7$  and  $1''.0$  (see Sect. 6.5 below). The photometric zeropoint determinations are also performed with both settings, and the one based on the largest number of stars with valid measurements is recorded in the header of the images and single-band source catalogues with the DMAG keyword. CALMINAP gives the value of MIN\_APER that was used for the calibration, and CALSTARS is the number of Gaia stars with valid measurements.

### 6.3.2 CHANGES TO THE THELI PIPELINE

In [KiDS450], the THELI pipeline was used to process the  $r$ -band images for the weak lensing analysis, which resulted in two different  $r$ -band source catalogues with multi-band photometry (DR3 and the ‘lensing catalogue’ DR3.1). For DR4 we unify the analysis, using the  $r$ -band sources detected on the THELI co-added images as the basis of the multi-band photometry as well as the forthcoming lensing analyses.

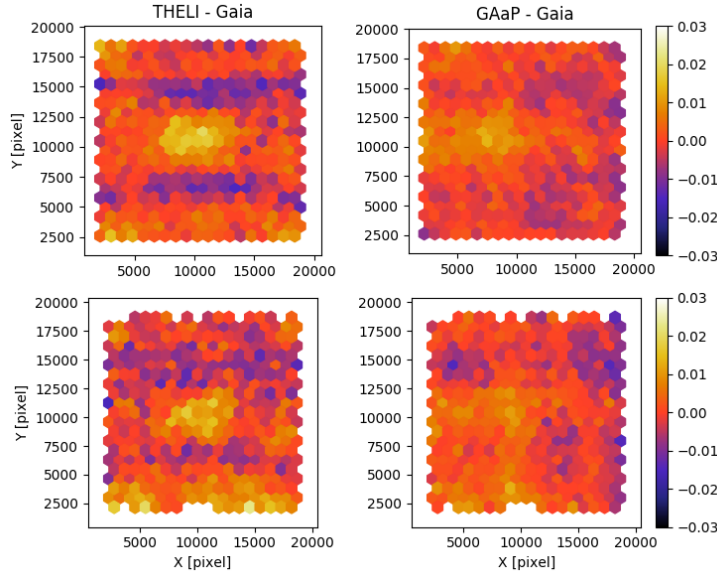


Figure 6.9: Check of the illumination corrections, through direct comparison of star magnitudes with Gaia. Each panel shows the median difference between a KiDS-DR4  $r$ -magnitude and Gaia-DR2  $G$ -band magnitudes, as a function of  $(X, Y)$  position in the focal plane. Stars in the range  $18 < r < 19$  with  $0.7 < (g - i)_0 < 1.5$  are used, on the flat part of the relation shown in Fig. 6.6. The top row shows the results for the tiles in KiDS-N, the bottom row for those in KiDS-S. THELI MAG\_AUTO is shown on the left, MAG\_GAAP\_r on the right.

Compared to [DR3], the main change for DR4 is the inclusion of the photometric illumination correction. This additional processing step is included in the photometric calibration procedure described in sect. 3.1 (item 4) of [DR3]. After obtaining a photometric zeropoint, extinction coefficient and colour term of OmegaCAM data overlapping with SDSS, we measure the residual systematic differences between OmegaCAM and SDSS-magnitudes over the OmegaCAM field-of-view. These differences are fitted and corrected with a second-order, two-dimensional polynomial over the OmegaCAM field-of-view – see also Fig. 6.8. The correction, which like the other calibration images is determined separately for each two-week ‘observing run’ that is processed by THELI, is directly applied to the single-frame pixel-data. Figure 6.9 compares stellar Gaia  $G$  magnitudes and the  $r$ -band magnitudes from the THELI and ASTRO-WISE reductions, as a function of position on the focal plane. The small residuals on the order of 0.02 mag are caused by different ways of treating the region of the images that are affected by scattered light shadows from the bond wire baffles above the CCD mosaic (see [DR3] for more details).

Another, more minor, change to the THELI workflow is the streamlining of the procedure for masking residual satellite trails on the single exposures. Whereas previously this was done per CCD, the new procedure allows the inspector to mask the track on the entire mosaic in one step.

## 6.4 SINGLE-BAND $u$ , $g$ , $r$ AND $i$ CATALOGUES, IMAGES AND MASKS

For every co-added image from the ASTRO-WISE pipeline a single-band catalogue was produced using SExtractor (Bertin & Arnouts 1996), using the same settings as in [DR3]<sup>14</sup>. These catalogues are calibrated photometrically using the nightly zeropoints. To convert the fluxes in these catalogues into SLR+Gaia calibrated magnitudes, the zeropoint given in the DMAG keyword should be used. This value should be added to any magnitude found in the catalogue, and any flux in the catalogue can be turned into a magnitude via

$$m = \text{DMAG} - 2.5 \log_{10} \text{FLUX}. \quad (6.7)$$

The SExtractor Kron-like MAG\_AUTO and isophotal magnitude MAG\_ISO are provided, as well as a range of circular-aperture fluxes, star-galaxy classification and shape parameters. App. 6.8 gives a full list of the parameters included in the catalogues. Note that the single-band catalogues are derived independently from each co-added KiDS observation, without cross-calibration or source matching across filter bands. In particular, the sources in the  $r$ -band catalogues are extracted from the ASTRO-WISE co-added images, and differ from those in the nine-band catalogue presented below, which are extracted from the THELI  $r$ -band images. Also, sources in the overlap region between adjacent tiles will appear in multiple single-band catalogues, with independent measurements for position, flux, etc. The single-band  $u$ ,  $g$ ,  $r$  and  $i$  catalogues contain an average of 22k, 79k, 125k, and 65k sources, respectively, per KiDS tile.

DR4 includes the co-added images with corresponding weight maps and masks, as well as the single-band catalogues described above with SExtractor output. The PULECENELLA masks identifying stellar reflection haloes are generated in the same way as for [DR3], and are described there.

Though it is possible to generate spectral energy distributions by matching the sources in the  $ugri$  catalogues for any given tile, it should be remembered that the PSF differs across the bands, and from tile to tile. A better way of obtaining reliable colours is to use the matched-seeing, matched-aperture catalogues described in the next section.

## 6.5 THE JOINT KiDS-VIKING NINE-BAND CATALOGUE

Besides more than doubling the area of sky covered with respect to DR3, the major new feature of KiDS-ESO-DR4 is the inclusion of near-infrared fluxes, from the VIKING survey (Edge et al. 2013). This survey was conceived together with KiDS, as a means to improve knowledge of the spectral energy distributions (SEDs) of the

---

<sup>14</sup>Note that the SExtractor settings are optimised for small sources; measurements for large objects such as extended galaxies should be used with care because of possible shredding or oversubtraction of the background (e.g., Kelvin et al. 2018).



sources, and in particular to enhance the quality of the photometric redshift estimates. Because VISTA entered operations before the VST, VIKING was completed first: full coverage of the  $1350 \text{ deg}^2$  area was reached in August of 2016, while repeat observations of low-quality data were completed in February of 2018<sup>15</sup>. DR4 includes photometry for all sources detected on the  $r$ -band co-added images as processed with the THELI pipeline:  $r$ -band Kron-like MAG\_AUTO, isophotal MAG\_ISO magnitudes and a range of circular-aperture fluxes, as well as nine-band optical/near-infrared GAAP fluxes for SED estimation. Note that only  $r$ -band total magnitudes are supplied.

Full details of the near-IR photometry are presented in Wright et al. (2018). Briefly summarized, the measurements start from the ‘pawprint’ images processed by the Cambridge Astronomical Survey Unit (CASU), which combine each set of jittered observations (offsets of a few arcseconds) into an astrometrically and photometrically calibrated set of sixteen detector-sized images. Each VIKING tile consists of six such pawprints, with observations offset by nearly a full detector width in right ascension, and half that in declination. Because the large gaps between the detectors can result in significant PSF quality jumps after co-addition, the VIKING photometry for KiDS-ESO-DR4 is performed by running the PSF Gaussianization and GAAP separately on each pawprint detector. The final flux in each VISTA band is the optimally weighted average of the individual flux measurements, using the individual flux errors that are derived by propagating the pixel errors (including covariance) through the GAAP procedure as described in appendix A of Kuijken et al. (2015). Due to the VIKING observing strategy sources typically appear on two VIKING pawprints (four in case of  $J$ -band), but a few percent of the sources appear six times (twelve in  $J$ ) or more. The number of exposures that contribute to each source’s VIKING fluxes is given in the catalogue. Note that the optical photometry is performed on the co-added images, which are much less sensitive to PSF changes between sub-exposures because of the very high pixel coverage fraction of the focal plane of the optical camera (which has minimal gaps between the three-edge buttable CCDs in the instrument).

Because of their respective cameras’ different footprints on the sky, KiDS and VIKING tile the sky differently. Data quality variations (depth, seeing, background...) follow a square-degree pattern for the KiDS data, and a  $1.5 \times 1$  degree pattern for VIKING; moreover within VIKING tiles the variations can be more complex because of the larger gaps between VIRCAM detectors.

The essence of the GAAP photometry (Kuijken et al. 2015) contained in the DR4 nine-band catalogue is to provide accurately aperture-matched fluxes across all wavebands, properly corrected for PSF differences. The aperture major and minor axis lengths,  $A_{\text{gaper}}$  and  $B_{\text{gaper}}$ , are set from the SExtractor size and shape parameters measured on the detection image, via

$$X_{\text{gaper}} = \left( X_{\text{WORLD}}^2 + \text{MIN\_APER}^2 \right)^{1/2} \quad \text{for } X = A, B \quad (6.8)$$

(see Fig. 6.10) with the position angle equal to the position angle THETA\_WORLD

<sup>15</sup>The VIKING repeat observations taken by September 26, 2016 are incorporated into the DR4 catalogues presented here.

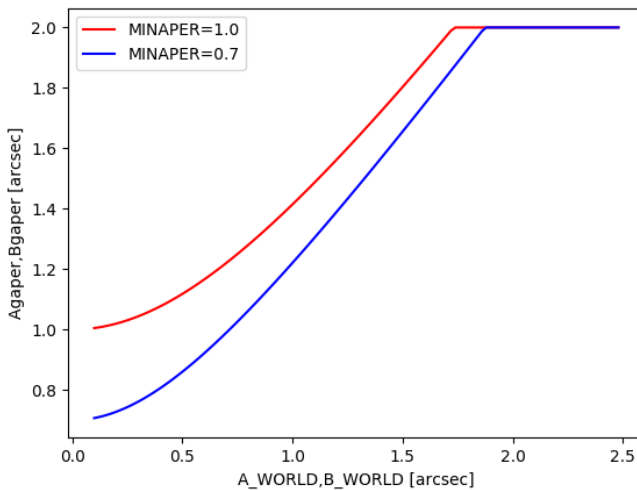


Figure 6.10: Relation between the SETRACTOR major and minor axis measurements ( $A$  and  $B$ ) and the adopted GAAP apertures, for the two minimum aperture (MIN\_APER) values that have been used.

from the detection image.<sup>16</sup> As it was in [DR3], the MIN\_APER parameter is set to  $0''.7$  for all sources, and in addition  $A_{\text{gaper}}$  and  $B_{\text{gaper}}$  are maximized at  $2''$ . Imposing a maximum aperture size helps to ensure that the GAAP colours are not contaminated by neighbouring sources, but we do not attempt here to deblend overlapping sources. We note that an explicit flagging of sources affected by neighbours is done by the SETRACTOR source detection step, and is also an important part of the forthcoming *lensfit* shape measurements. For further discussion of the choice of GAAP aperture size, see Kuijken et al. (2015), appendix A2.

In rare cases, no GAAP flux can be measured with this setup, because GAAP photometry can only be determined when the specified aperture size is larger than the Gaussianized PSF. To provide colours for these sources, DR4 includes a second run with larger apertures, obtained by setting MIN\_APER= $1''.0$ . The results from both GAAP runs are present in the catalogues, with keywords whose names contain `_0p7` and `_1p0` respectively. In general the `1p0` fluxes will have a larger error than those from the standard `0p7` setup, since the larger aperture includes more background noise. However, when the Gaussianized PSF size is close to that of the GAAP aperture, the error increases<sup>17</sup>; in such cases it may happen that the larger MIN\_APER= $1''.0$  leads to a smaller flux error (and when the PSF is too broad the flux error is formally infinite). In the DR4 catalogue a source-by-source decision is made which optimal

<sup>16</sup>The convention in the GAAP code is that the position angle is measured from East to North, so the catalogue contains the angle `PA_GAAP= 180-THETA_WORLD`.

<sup>17</sup>because the aperture that is used for the photometry is the GAAP aperture deconvolved by the PSF

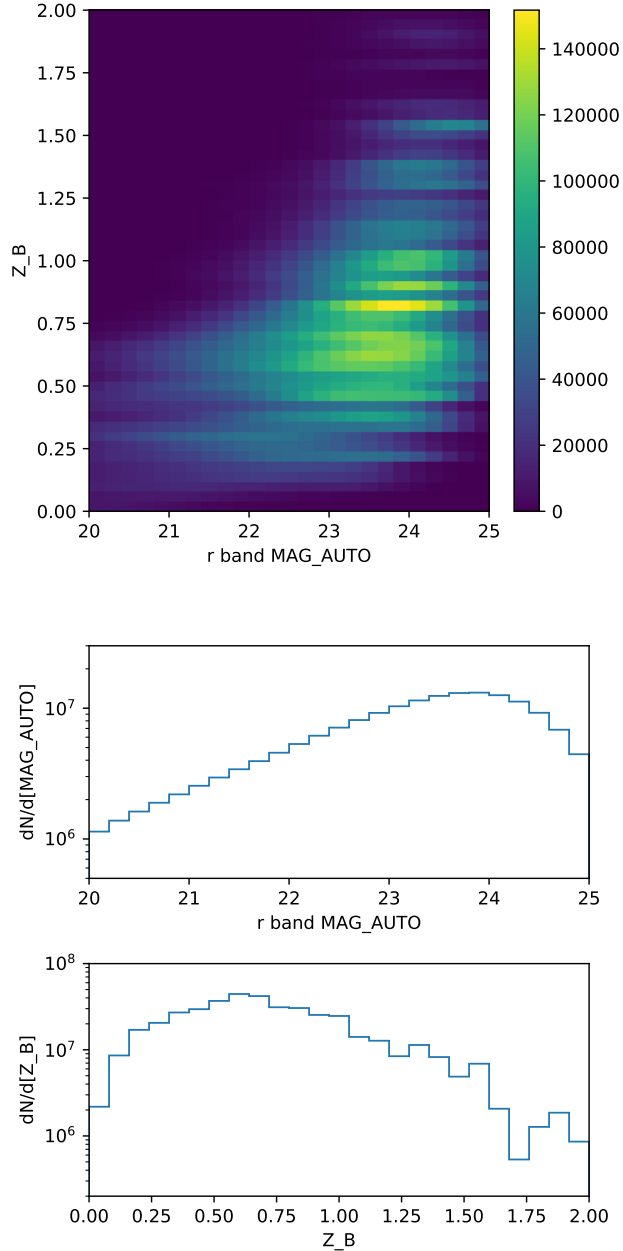


Figure 6.11: Joint and marginal histograms of the  $r$ -band MAG\_AUTO magnitude and the  $Z_B$  photometric redshift estimate for the sources in the nine-band catalogue.

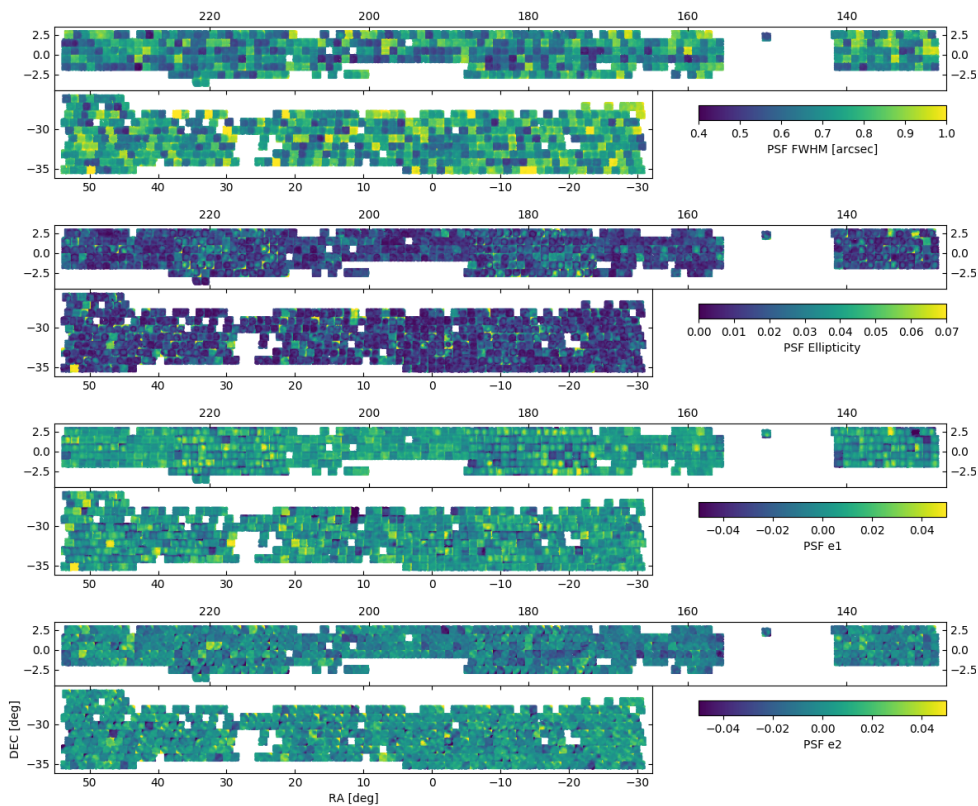


Figure 6.12:  $r$ -band PSF properties across KiDS-ESO-DR4. From top to bottom: FWHM, ellipticity modulus, and ellipticity components 1 and 2 (elongation along the pixel  $x$  axis and diagonal, respectively).

MIN\_APER choice to use as input for the photometric redshifts, as follows:

1. For all bands  $x$ , calculate the flux error ratios

$$R_x = \text{FLUXERR\_GAAP\_1p0\_x} / \text{FLUXERR\_GAAP\_0p7\_x} \quad (6.9)$$

2. If

$$\max_x(R_x) \times \min_x(R_x) < 1 \quad (6.10)$$

then use the 1p0 fluxes for this source, else adopt 0p7.

This choice ensures that the smaller 0p7 aperture is used, unless there is a band for which the larger aperture gives a smaller flux error, for the reasons indicated above: in that case, the 1p0 fluxes are used if the fractional reduction in the error in that band is greater than the fractional penalty suffered by the other bands.<sup>18</sup>

The larger MIN\_APER is preferred in some four percent of the cases. These optimal GAAP fluxes are reported in the catalogue with the FLUX\_GAAP\_ $x$  keywords. The flux and magnitude zeropoints in the catalogue are as follows.

- All GAAP fluxes are reported in the units of the image pixel ADU values. For the KiDS images these correspond approximately to a photometric AB magnitude zeropoint of 0; for the VISTA data this zeropoint is 30.
- The GAAP magnitudes MAG\_GAAP\_0p7\_ $x$  and MAG\_GAAP\_1p0\_ $x$  for the KiDS *ugri* bands are calculated from the corresponding fluxes using the zeropoint DMAG\_ $x$  or DMAG\_ $x$ \_1 (see Sect. 6.3.1) and recorded in the catalogue header.
- The GAAP magnitudes for the VIKING *ZYJHK<sub>s</sub>* bands are calculated with the zeropoint 30.
- The optimal GAAP fluxes FLUX\_GAAP\_ $x$  are equal to one of the 0p7 or 1p0 sets, as described above.
- The optimal GAAP magnitudes MAG\_GAAP\_ $x$  are calculated as above, but in addition are also corrected for Galactic extinction by subtracting the EXTINCTION\_ $x$  value (obtained using the data in Table 6.4).
- The colours COLOUR\_GAAP\_ $x$ \_ $y$  in the catalogue are obtained as differences between these extinction-corrected MAG\_GAAP\_ $x$  magnitudes, and are therefore corrected for Galactic reddening.

The nine-band catalogue also contains those KiDS sources that fall outside the VIKING footprint, as can be seen on the maps of limiting magnitude in Sect. 6.6.5 below. Note in particular that the COSMOS field near (RA,DEC)=(150°0, 2°5) (KiDS-N-D2 in Table 6.2) is not part of the VIKING survey, though other – in some cases much deeper – VISTA data do exist for this field. These near-infrared data do not

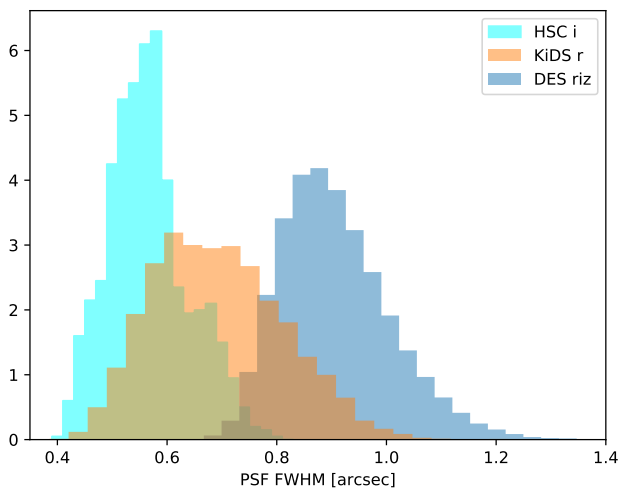


Figure 6.13: The KiDS  $r$ -band seeing FWHM distribution (red) compared to those used for shape measurements in the other major ongoing weak lensing surveys: HSC  $i$  band (cyan) and DES  $r+i+z$  (blue).

form part of KiDS-DR4, but they are incorporated in the calibration of the photometric redshifts for the KV450 analysis in Hildebrandt et al. (2018).

Fig. 6.11 illustrates the  $r$ -band depth and photometric redshift distribution of the sources in the nine-band catalogue. For a list of all columns, see App. 6.8. The catalogue contains just over 100 million objects.

## 6.6 DATA QUALITY

### 6.6.1 IMAGE QUALITY

For weak lensing, the most critical science case for KiDS, image quality is a crucial property of the data. KiDS scheduling is designed to take advantage of the periods of excellent seeing on Paranal, by prioritising the  $r$ -band exposures at those times. The resulting seeing distribution of the four KiDS bands was shown in Fig. 6.3.

In Fig. 6.12 we present maps of the KiDS-ESO-DR4  $r$ -band PSF size and ellipticity, obtained by running the *lensfit* PSF modelling code used for the KiDS-450 analysis [KiDS450] on the THELI images. The top row of the Figure shows the PSF FWHM. Individual KiDS tiles are clearly visible. The second row shows the PSF ellipticity, and the bottom two rows the ‘1’ and ‘2’ ellipticity components. The results of the

<sup>18</sup>This formulation also handles the convention that an unmeasured GAAP flux returns an error of  $-1$ , provided the rare cases that a  $1p0$  flux cannot be measured but a  $0p7$  flux can, are caught.

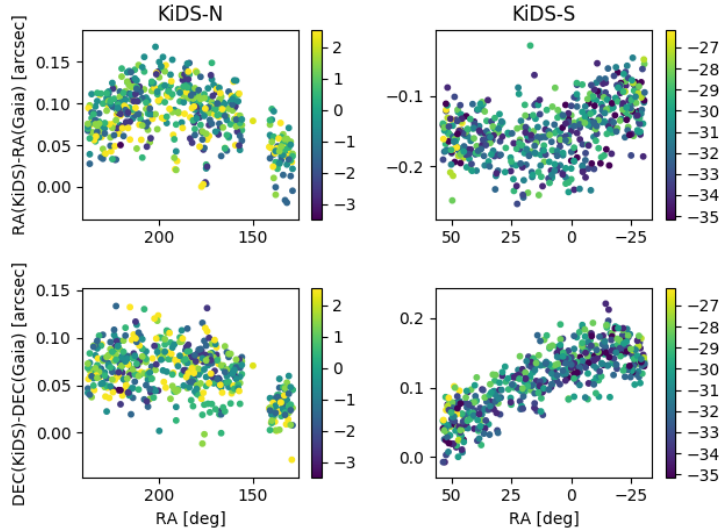


Figure 6.14: Median THELI astrometry residuals per KiDS tile in the nine-band catalogues, as measured from unsaturated Gaia stars, in arcseconds. Left/right plots show KiDS-N/S, and the top and bottom rows show the median offsets per tile in RA and DEC. The colours indicate the declination of the tiles, in degrees.

VST improvements that were implemented in 2015 are reflected in the DR4 data: PSF variations are significantly reduced in the newly added data compared to the data from DR1+2+3 (see the blue areas in Fig. 6.1).

In Fig. 6.13 we compare the seeing distribution of the KiDS-DR4  $r$ -band data with those of the images used for the lensing measurements in the other major ongoing surveys: the DES-Year 1  $riz$  data (Zuntz et al. 2018), which has a similar depth to KiDS, and the HSC DR1  $i$ -band lensing catalogue (Mandelbaum et al. 2018). The superior seeing of KiDS compared to DES explains why both surveys are providing cosmic shear constraints of comparable power, despite the larger DES area. The impact that can be expected from the complete HSC survey is also evident, even more so considering that it is significantly deeper than KiDS and DES.

### 6.6.2 ASTROMETRY

The astrometry in the THELI-processed  $r$ -band detection images and catalogues is tied to SDSS in the North, and to 2MASS in the South. The ASTRO-WISE images and single-band catalogues are all tied to 2MASS, as is VIKING. Slight differences exist between these two reference catalogues.

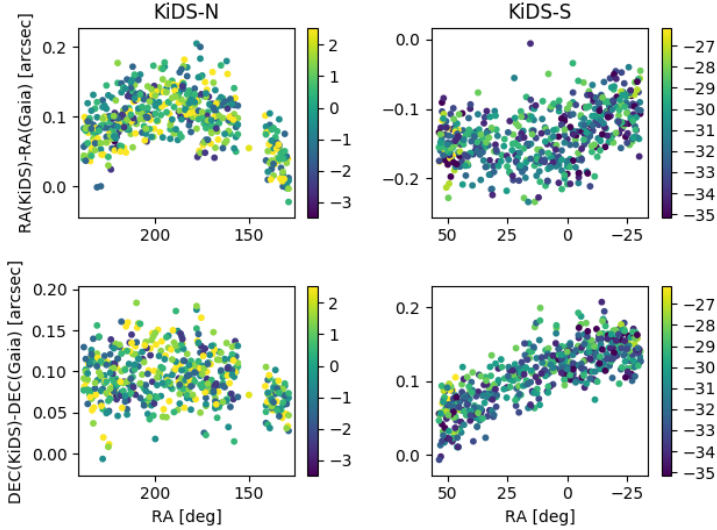


Figure 6.15: As Fig. 6.14, but for the ASTRO-WISE single-band  $r$  catalogues.

We compare the THELI and ASTRO-WISE astrometry to the Gaia DR2 data in Figs. 6.14 and 6.15. Systematic residuals between either catalogue and Gaia are at the level of 200 mas; between the THELI and ASTRO-WISE reductions the differences are at most 50 mas. These latter differences are sufficiently small that they will not affect the GAAP photometry (where the apertures are defined on the THELI images, but the fluxes measured at the corresponding positions in the ASTRO-WISE images).

### 6.6.3 PHOTOMETRY

We assess the accuracy of the KiDS photometric calibration by comparing with overlapping, shallower surveys. For galaxies, comparing the KiDS photometry with catalogues from other surveys is complex, since for extended sources the GAAP fluxes do not measure total fluxes (see Sect. 6.3.1). If desired, the  $r$ -band circular aperture fluxes in the catalogue can be used to generate a curve-of-growth, and total magnitudes can then be estimated from the GAAP colours. (Alternatively, the  $r$ -band MAG\_AUTO can be combined with the GAAP colours to generate estimated Kron-like magnitudes in the other bands. Such procedures assume that there are no colour gradients in the galaxy, an assumption that could be tested by comparing the 0p7 and 1p0 GAAP fluxes.)



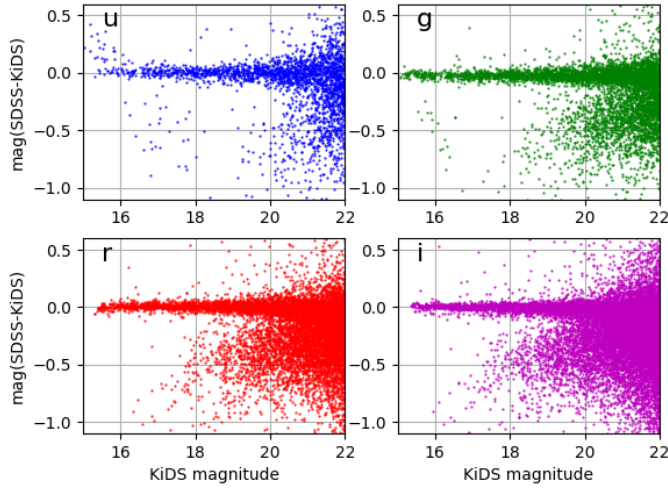


Figure 6.16: Comparison between the SDSS DR9 model magnitudes and KiDS-ESO-DR4 GAAP photometry. The comparisons are shown for  $u$ ,  $g$ ,  $r$ , and  $i$  bands, for an example tile (KIDS\_188.0\_-0.5).

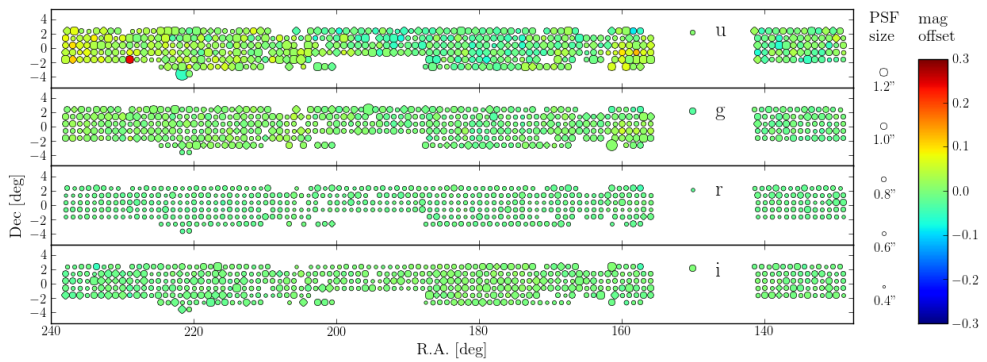


Figure 6.17: Photometric comparison between KiDS and SDSS photometry of stars in the KiDS-N area. The colourscale indicates the mean magnitude offset  $m_{\text{KiDS}} - m_{\text{SDSS}}$  of high-signal stars  $u < 20$ ,  $g < 22$ ,  $r < 22$ ,  $i < 20$  in every tile. The size of the symbol increases with the FWHM of the PSF. The one outlier in the  $u$ -band photometry, tile KIDS\_229.0\_-2.5\_u, is one of the highest-extinction fields in the survey and contains few  $u$ -band stars.

For stars and other unresolved objects, the situation is different, since their GAAP fluxes *are* total fluxes: the GAAP flux with Gaussian aperture function  $W = \exp[-\frac{1}{2}(x^2/A^2 + y^2/B^2)]$  (in coordinates centered on the source and rotated to align with the aperture major and minor axes) is the integral (on the pre-seeing sky)

$$\int dx dy I(x, y) W(x, y) \quad (6.11)$$

which evaluates to the flux  $F$  of the source when the intensity  $I(x, y)$  is  $F$  times a delta function.

Such a comparison is shown in Fig. 6.16, for an example tile in KiDS-N where SDSS and KiDS overlap. As expected, the stars form tight sequences close to the line of zero magnitude difference, demonstrating that the KiDS and SDSS zero points are consistent for this tile, while the KiDS GAAP magnitudes of galaxies trail towards fainter magnitudes than the corresponding SDSS magnitude. A similar comparison for the near-IR data is presented in Wright et al. (2018). Tile-by-tile consistency in the four bands, for the KiDS-SDSS overlap, is shown in Fig. 6.17. The larger scatter in the  $u$  band, already discussed in Sect. 6.3.1, is evident, particularly in the fields with higher extinction at the extremes of the RA range. In these tiles there are fewer stars with reliable  $u$  band photometry; in addition, because of their lower Galactic latitude the foreground extinction screen approximation is less well justified.

Figure 6.18 illustrates an internal consistency check of the GAAP magnitudes, that can also serve as a new star-galaxy classifier. The top panel shows the difference between the ‘0p7’ and ‘1p0’  $r$ -band GAAP magnitudes in an example tile. This comparison clearly reveals two populations which are well-separated at the bright end, down to magnitude 22.5. A conservative cut, shown in gold, identifies unresolved objects, for which the GAAP magnitude is independent of aperture size. (Most of the remaining objects are resolved, with higher fluxes for the larger aperture.) The bottom panel shows the location in the  $g - r, r - i$  colour-colour diagram of the two populations, confirming that the unresolved objects have mostly stellar colours, whereas the others show the colour distribution expected of a population of galaxies at a wide range of redshifts.

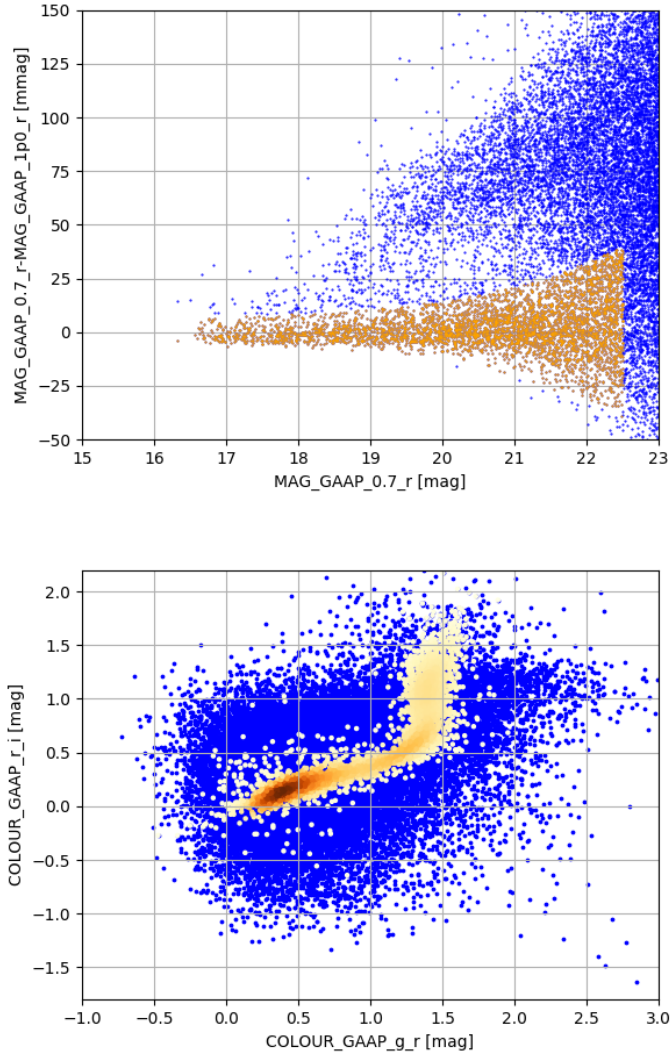


Figure 6.18: Top: star-galaxy separation using the 0p7 and 1p0 GAAP magnitudes. The objects shown in gold form a sequence along which both apertures yield consistent fluxes, indicating that they are unresolved. The bottom plot shows the  $g - r, r - i$  colour-colour diagram for the same sources, confirming that these sources are stars.

### 6.6.4 PHOTOMETRIC REDSHIFTS

The nine-band catalogue contains photometric redshift estimates, obtained with the BPZ code (Benitez 2000). It gives the most probable redshift values, as well as the ‘ $1\sigma$ ’ 32nd and 68th percentiles of the posterior probability distributions, and the best-fit SED type. The BPZ version and settings are given in Table 6.5.

Table 6.5: Settings for the BPZ photometric redshift calculations

Parameter	Value
BPZ version	1.99.3
ZMAX	7.0
INTERP	10
ODDS	0.68
MIN_RMS	0.067
PHOTO_ERRORS	yes

Since [DR3] we have implemented several changes to our photo- $z$  setup. We updated the prior redshift probability used in BPZ to the one given in Raichoor et al. (2014). This prior reduced uncertainties and catastrophic failures for faint galaxies at higher redshifts, but appears to generate a redshift bias for bright, low-redshift galaxies. We therefore caution users of the catalogue to calibrate the BPZ redshifts appropriately before using them. At bright magnitudes, where complete training data are available, it is advantageous to use an empirical photo- $z$  technique like the one presented for the [DR3] data set in Bilicki et al. (2018); specific selection and calibration of luminous red galaxies (LRG, Vakili et al. 2019) is also effective. Bright and LRG samples based on DR4 and taking advantage of its unique, deep, nine-band coverage are in preparation.

Another change is related to the photo- $z$  errors and the ODDS quality indicator. In previous releases we reported 95% confidence intervals for the Bayesian redshift estimate. With DR4 we switch to 68% confidence intervals as mentioned above, which requires some changes to the settings and also changes the values of the ODDS parameter. The changes in prior and BPZ settings, together with the fact that we are using full nine-band photometry in a KiDS data release for the first time mean that previous photo- $z$  results based on optical-only photometry (e.g., Kuijken et al. 2015) are no longer advocated.

Further discussion of the nine-band KiDS+VIKING photometric redshifts, as well as a comparison to the [KiDS450] photo- $z$ , is provided in Wright et al. (2018) where a similar setup<sup>19</sup> was used. There the BPZ photo- $z$  are tested against several deep spectroscopic surveys. At full depth ( $r \leq 24.5$ ) the photo- $z$  show a scatter (normalised-median-absolute-deviation) of  $\sigma_m = 0.072$  of the quantity  $\Delta_z/(1+z) = (z_B - z_{\text{spec}})/(1+z_{\text{spec}})$  and a fraction  $\zeta_{0.15} = 17.7\%$  of outliers with  $|\Delta_z/(1+z)| \geq 0.15$ . The magnitude

<sup>19</sup> The main difference of the DR4 setup is the implementation of two different minimum aperture sizes as discussed in Sect. 6.5. As this change only impacts data with seeing that greatly varies between bands in Wright et al. (2018), it does not affect the photo- $z$  statistics presented here.

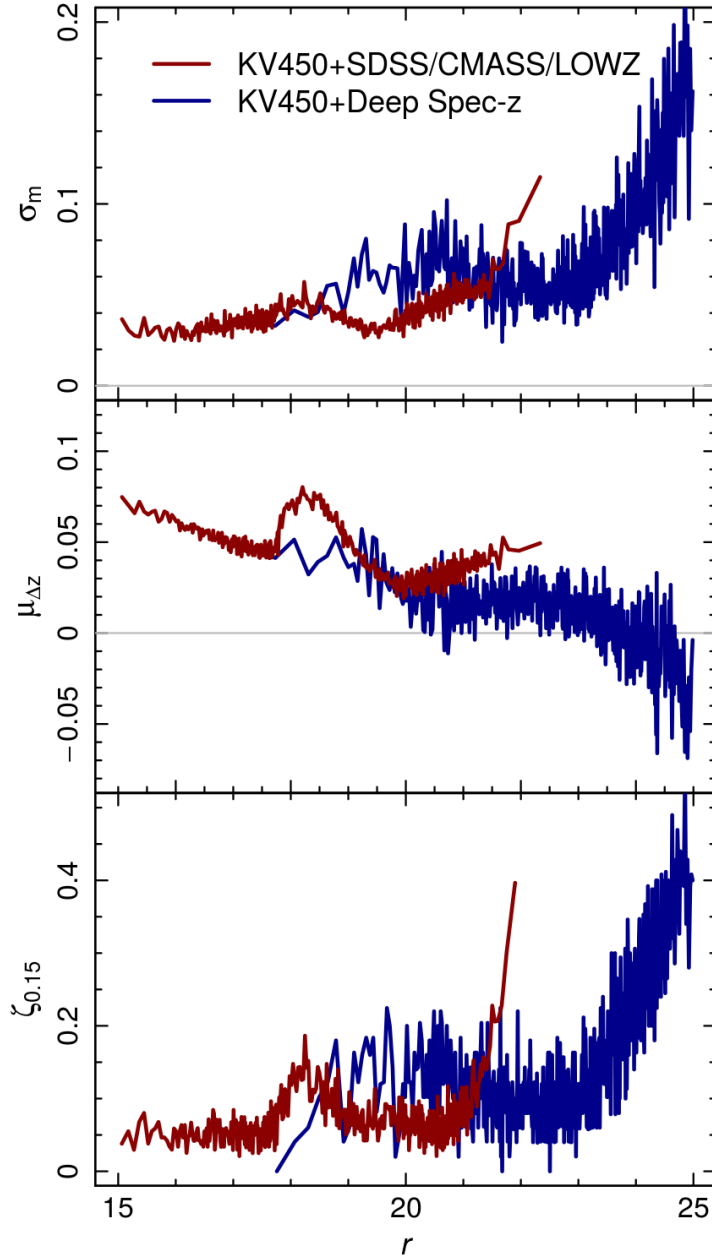


Figure 6.19:  $r$ -Magnitude dependence of KiDS DR4 photo- $z$  statistics based on the findings of Wright et al. (2018) for the deep spec- $z$  fields (blue lines) and a direct comparison of DR4 photo- $z$  and SDSS/2dFLenS spec- $z$  (red lines). The top panel shows the normalised-median-absolute-deviation of the quantity  $\Delta_z/(1+z)$ , the middle panel shows the mean  $\mu_{\Delta_z}$  of that quantity, and the lower panel shows the rate  $\zeta_{0.15}$  of outliers with  $|\Delta_z/(1+z)| \geq 0.15$ . (Note that this definition of  $\zeta_{0.15}$  exaggerates the outlier fraction when  $\sigma_m$  approaches 0.1.) The scatter between neighbouring points gives an indication of the error bars on these quantities.

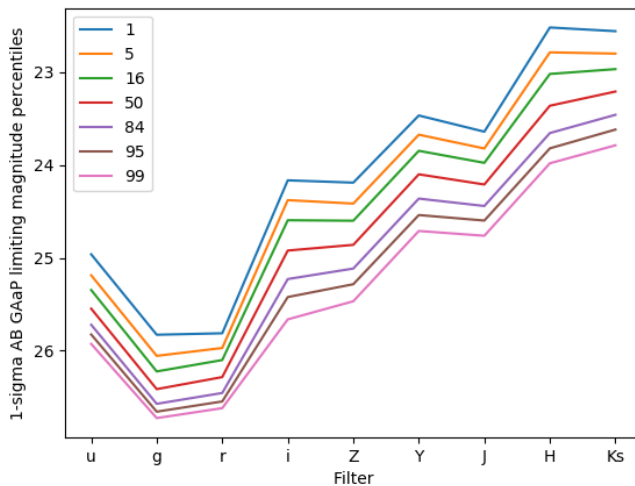


Figure 6.20: Various percentiles of the  $1\text{-}\sigma$  GAAP limiting magnitudes for the nine wavelength bands. The width of the distributions is driven by differences in seeing, air mass and sky brightness across the KiDS and VIKING surveys.

dependence of these quantities is shown in Fig. 6.19. The effect of the different selection criteria in the spectroscopic catalogues is evident (e.g., the bump in the redshift bias  $\mu_{\Delta z}$  near  $r = 18$ , which marks the transition from the BOSS LOWZ to CMASS samples), illustrating that calibrating the photometric redshift error distribution requires care (see Bilicki et al. 2018 and the extensive discussion of direct calibration techniques in [KiDS450] for further details). Note also that the photo-z setup was optimised for the fainter,  $r > 20$  galaxies that are of interest for the KiDS cosmology analysis, and not for the brighter galaxies.

### 6.6.5 PHOTOMETRIC DEPTH AND HOMOGENEITY

Figure 6.20 shows the distribution of the different bands'  $1\text{-}\sigma$  limiting magnitudes in the catalogue. Note the narrow range of the limiting magnitudes in the  $u$ ,  $g$ , and particularly the  $r$  band, a consequence of the KiDS observing strategy of choosing which dark-time band to observe in according to the seeing conditions. Maps of the median limiting magnitude in  $0^\circ 1 \times 0^\circ 1$  cells are presented in Fig. 6.21 (KiDS  $ugri$ ) and 6.22 (VIKING  $ZYJHK_s$ ).

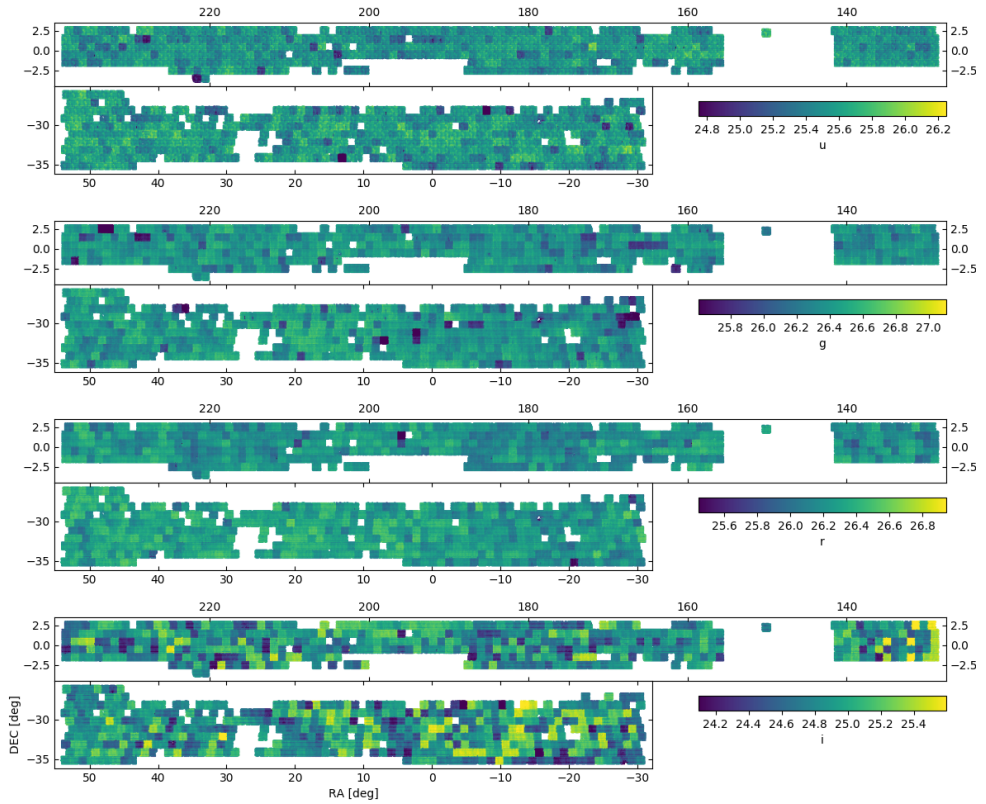


Figure 6.21: Maps of the median limiting GAAP magnitude, corresponding to the  $1\text{-}\sigma$  flux error, in  $0^\circ 1 \times 0^\circ 1$  cells, for the four KiDS filter bands. The colour scale in every map spans  $\pm 0.75$  magnitude about the median. Note the significantly greater inhomogeneity of the  $i$ -band data: this is expected to improve in the final data release after a second pass is completed.

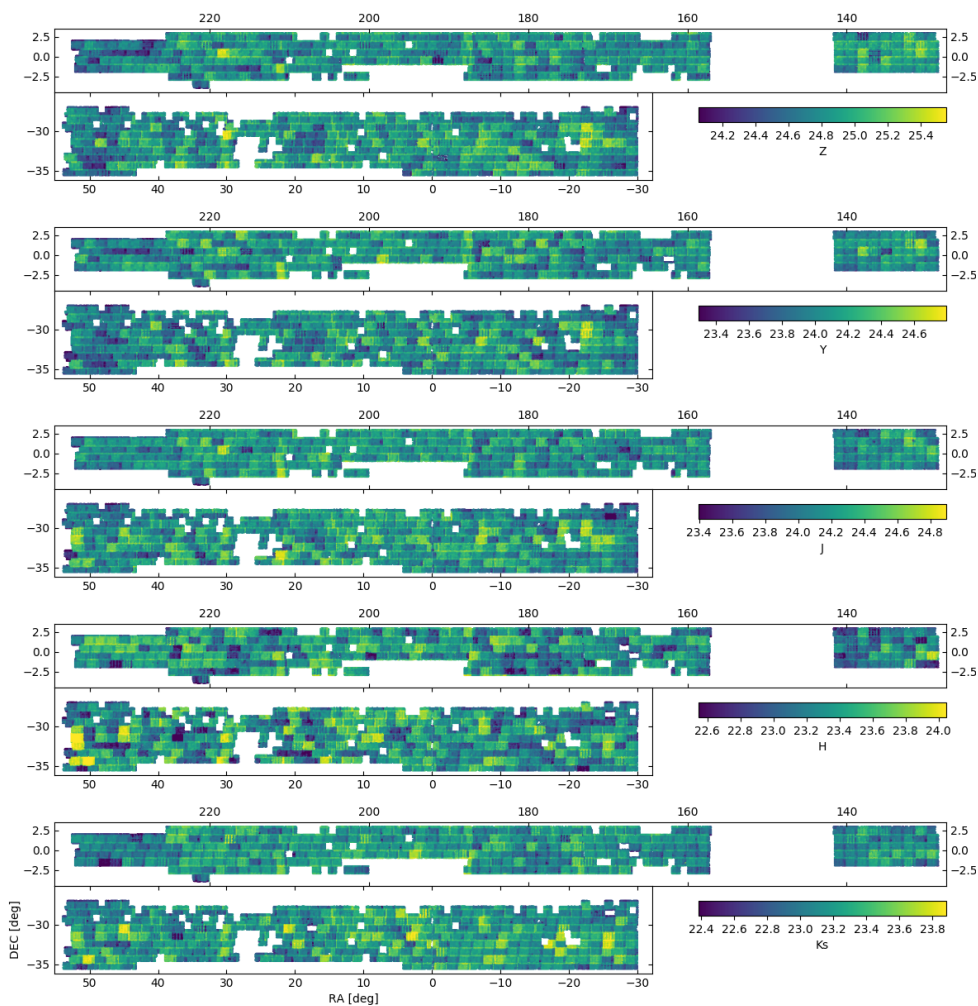


Figure 6.22: Maps of the  $r$ -band selected sources' median limiting GAAP magnitude, corresponding to the  $1\text{-}\sigma$  flux error, in  $0^\circ 1 \times 0^\circ 1$  cells, for the five VIKING bands. Note the rectangular  $1^\circ 5 \times 1^\circ 0$  patterns, due to the footprint of the VIRCAM instrument.



## 6.7 DATA ACCESS

There are several ways through which KiDS-ESO-DR4 data products can be accessed. An overview is presented in this section and up-to-date information is also available at the KiDS DR4 website.

The data products that constitute the DR4 release (stacked *ugri* images and their associated weight maps, flag maps, and single-band source lists for 1006 survey tiles, the *r*-band detection images and their weight maps, as well as the multi-band catalog combining KiDS and VIKING photometry together with flag maps combining mask information from all filters), are released via the ESO Science Archive, and also accessible via ASTRO-WISE and the KiDS website.

### 6.7.1 ESO SCIENCE ARCHIVE

All main release data products are disseminated through the ESO Science Archive Facility<sup>20</sup>, which provides several interfaces and query forms. All image stacks, weight maps, flag maps and single-band source lists are provided on a per tile basis via the ‘Phase 3 main query form’. This interface supports queries on several parameters, including position, object name, filter, observation date, etc. and allows download of the tile-based data files. Also the multi-band catalog, which is stored in per-tile data files, is available in this manner. A more advanced method to query the multi-band catalog is provided by the ‘Catalogue Facility query interface’, which enables users to perform queries on any of the catalog columns, for example facilitating selections based on area, magnitude, photo-*z* or shape information. Finally, data can be queried directly from a new graphical sky projection interface known as the ‘Science Portal’. Query results can subsequently be exported to various (single-file) formats.

### 6.7.2 ASTRO-WISE ARCHIVE

Most of the data products can also be retrieved from the ASTRO-WISE system (Begegan et al. 2012). This data processing and management system is used for the production of these data products and retains the full data lineage. For scientists interested in access to various quality controls, further analysis tools, or reprocessing of data this access route may be convenient. The DBviewer web service<sup>21</sup> allows querying for data products and supports file downloads, viewing of inspection plots, and data lineage browsing. Links with DBviewer queries to complete sets of data products are compiled on the KiDS DR4 website. Several data products that are not in the ESO archive may be retrieved through this route: most importantly, the PSF-Gaussianized images and the individual CCD sub-exposures after various stages of processing in the ASTRO-WISE pipeline are available here.

---

<sup>20</sup><http://archive.eso.org/cms.html>

<sup>21</sup><http://dbview.astro-wise.org>

### 6.7.3 KiDS DR4 WEBSITE

Apart from offering an up-to-date overview of all data access routes, the KiDS DR4 website<sup>22</sup> also provides alternative ways for data retrieval and quality control.

The synoptic table presents for each observation (tile/filter) a combination of inspection plots relating to the image and source extraction quality, as well as links for direct downloads of the various data FITS files. Furthermore, direct batch downloads of all DR4 FITS files are supported by supplying WGET scripts.

## 6.8 SUMMARY AND OUTLOOK

With the KiDS-ESO-DR4 data release, data for over 1000 square degrees, more than two thirds of the target KiDS footprint, is now publicly available. Co-added images with associated weights and masks, as well as single-band source catalogues, may now be accessed through the ESO archive or the KiDS project website.

Moreover, through a combined analysis of these KiDS images with data from the VIKING survey, a nine-band matched-aperture  $u-K_s$  catalogue containing some 100 million galaxies has been created, with limiting  $5-\sigma$  AB magnitudes ranging from ca. 25 in  $g$  and  $r$  bands to 23 in  $J$  and 22 in  $K_s$  (Fig. 6.20). This data set is by far the largest-area optical+near-IR data set to this depth. The galaxies in this catalogue have been detected using a reduction of the data that has been optimised for weak gravitational lensing measurements, to enable the primary science goal of KiDS. The GAAP photometry in the nine-band catalogue uses the positions and sizes of these galaxies to define the apertures. Analysis of the gravitational lensing information in the data set is in progress, and shear estimates for these sources will be released in due course. Photometric redshift estimates based on the nine-band photometry are already included in the DR4 catalogue (but see Sect. 6.6.4 for a discussion of redshift biases for bright sources).

Multiple other applications of this unique optical+near-infrared catalogue are foreseen. For example, stellar mass estimates for galaxies will benefit greatly from the inclusion of the near-IR fluxes (Wright et al. 2018), red-sequence cluster searches can be pushed to greater redshift, and star-galaxy separation and galaxy SED typing can be made more accurate as well (e.g., Daddi et al. 2004; Tortora et al. 2018b).

The data processing for DR4 largely followed the procedures established for the previous data release as described in [DR3], with a few improvements:

1. this is the first KiDS data release for which the photometry has been tied to the Gaia database;
2. satellite tracks and other artefacts are now masked at the sub-exposure level, increasing the usable area of the co-added images;
3. the PSF Gaussianization procedure now operates in pixel space, solving directly for a double-shapelet convolution kernel that renders the PSF Gaussian;

---

<sup>22</sup><http://kids.strw.leidenuniv.nl/DR4>

4. extinction corrections have been updated to the Schlafly & Finkbeiner (2011) extinction coefficients;
5. GAAP photometry is run twice, with the second run using larger apertures to be able to include occasional poor-seeing KiDS or VIKING data in the photometry catalogue;
6. the THELI processing of the images on which the sources for the nine-band catalogue are detected now includes an illumination correction;
7. the inclusion of the VIKING data involved a re-reduction of the VIKING paw-print level data (see Wright et al. 2018), and is the first time the PSF Gaussianization and GAAP photometry have been performed at sub-exposure level and combined.

The data are publicly available via the ESO archive, the ASTRO-WISE system, and the KiDS project website. A description of the data format may be found in the Appendix.

KiDS observations continue, and are expected to wind down by the middle of 2019, at which point some 1350 square degrees will have been mapped in 9 photometric bands by the combined KiDS+VIKING project.

A repeat pass of the whole survey area in the *i*-band is also close to completion. These data will enable variability studies on timescales of several years, as well as improving the overall quality of the *i*-band data, which has the greatest variation in observing conditions and cosmetic quality. In addition, a number of fields with deep spectroscopic redshifts are also being targeted with the VST and VISTA to provide KiDS+VIKING-like photometry for large samples of faint galaxies that can be used as redshift calibrators.

The next full data release, DR5, is expected to be the final one, containing data from the full KiDS/VIKING footprint shown in Fig. 6.1. Intermediate ‘value-added’ public releases based on DR4, including one with weak lensing shape measurements, will be made together with the corresponding scientific analyses.

## ACKNOWLEDGEMENTS

We are indebted to the staff at ESO-Garching and ESO-Paranal for managing the observations at VST and VISTA that yielded the data presented here. 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). Data processing for VIKING has been contributed by the VISTA Data Flow System at CASU, Cambridge and WFAU, Edinburgh. This work is supported by the Deutsche Forschungsgemeinschaft in the framework of the TR33 ‘The Dark Universe’. We acknowledge support from European Research Council grants 647112 (CH, BG) and 770935 (HHi), the Deutsche Forschungsgemeinschaft (HHi, Emmy Noether grant Hi 1495/2-1 and Heisenberg grant Hi 1495/5-1), the Polish Ministry of Science and Higher Education (MB, grant DIR/WK/2018/12), the INAF PRIN-SKA 2017 program 1.05.01.88.04 (CT), the Alexander von Humboldt Foundation (KK), the STFC (LM, grant ST/N000919/1), and NWO (KK, JdJ, MB, HHo, research grants 621.016.402, 614.001.451 and 639.043.512). We are very grateful to the Lorentz Centre and ESO-Garching for hosting several team meetings.

## 6.A DESCRIPTION OF THE DATA PRODUCTS

The appendices of this chapter can be found in the published journal article:  
K. Kuijken et al., *A&A*, **Volume 625**, Issue 1, p. A2 (2019)

---

## Samenvatting

---

### 7.1 DE CONNECTIE TUSSEN STERRENSTELSELS EN DONKERE MATERIE

**I**N dit proefschrift richt het onderzoek zich op de eigenschappen van donkere materie en donkeremateriehalo's en hoe ze verbonden zijn met de sterrenstelsels die we in het heelal kunnen waarnemen. Vanwege de nog steeds onbekende aard van donkere materie bestuderen we haar doorgaans door middel van de eigenschappen van haar verdeling en haar eigenschappen op de schaal van sterrenstelsels en groter. De connectie tussen sterrenstelsels en donkere materie is belangrijk vanwege drie hoofdredenen, en haar begrijpen helpt bij het beantwoorden van de grootste vragen in de astrofysica en kosmologie van vandaag. De eerste vraag behelst het begrijpen van de fysica van sterrenstelselvorming. Ten tweede, de gevolgtrekking van kosmologische parameters – als we de kosmologische parameters robuust willen meten, moeten we het samenspel van de sterrenstelsels en de donkere materie begrijpen, en ten derde, de gevolgtrekking van de evolutie van de materieverdeling en de eigenschappen van donkere materie (Wechsler & Tinker 2018).

In het standaard, door koude donkere materie en kosmologische constante gedomineerde ( $\Lambda$ KDM) kosmologische raamwerk wordt structuurvorming in het heelal voornamelijk gedreven door de dynamica van koude donkere materie. De gravitationele ineenstorting van donkeremateriefluctuaties en hun volgende virialisatie leidt tot de vorming van donkeremateriehalo's uit de hoogste dichtheidspieken in het aanvankelijke Gaussische willekeurige dichtheidsveld. Het is breed aangenomen dat elk sterrenstelsel zich bevindt in een moederhalo van donkere materie. De sterrenstelsels die zich op de bodem van de potentiaalput van de donkeremateriehalo bevinden, worden aangeduid als centrale sterrenstelsels en sterrenstelsels die zich in een baan om het centrale sterrenstelsel bevinden, worden aangeduid als satellietsterrenstelsels. De precieze manier waarop sterrenstelsels de donkeremateriehalo's bevolken is nog steeds een onderwerp waar actief onderzoek naar gedaan wordt, waaraan dit proefschrift enkele inzichten bijdraagt. In het algemeen wordt de connectie tussen de populatie van sterrenstelsels en de populatie van donkeremateriehalo's statistisch gemodelleerd met een uitgebreid Press-Schechterformalisme (Press & Schechter 1974).

Dit formalisme postuleert dat zware sterrenstelsels zich in de hoogste dichtheidspieken van de onderliggende donkerematerieverdeling vormen en dat de connectie gerelateerd kan worden aan een grootte genaamd sterrenstelselvertekening. Hoewel dit formalisme het juiste aantal halo's als functie van massa voorspelt en ook het clusteren van sterrenstelsels dat we waarnemen tot grote nauwkeurigheid voorspelt, heeft het enkele tekortkomingen. Expliciet kan de sterrenstelselvertekening zelf niet-triviaal afhangen van massa en schaal en is het algemeen geaccepteerd dat ze niet lineair of deterministisch is. Bovendien zal de assemblagegeschiedenis van donkeremateriehalo's kenmerken achterlaten in de waargenomen verdelingen die niet langer overeen zullen stemmen met de theoretische voorspellingen. De assemblagegeschiedenis en mechanismen van satellietmassastripping en samensmelting laten ons ook achter met verschillende eigenschappen van de donkerematerieconnectie voor centrale en satellietsterrenstelsels. Vanwege al dit bestaat een grote verscheidenheid aan verschillende modellen, alle bouwende op de statistische postulaten van het Press-Schechterformalisme.

Een populaire en succesvolle manier om de connectie tussen sterrenstelsels en donkere materie te beschrijven is met de halobezettingsverdelingen (HBV), welke de kansverdelingen voor het aantal sterrenstelsels met een bepaalde eigenschap (lichtkracht of stellaire massa) in een halo specificeren, gegeven als een functie van halomassa. De halobezettingsverdelingen worden apart gekwantificeerd voor de centrale sterrenstelsels en de satellietsterrenstelsels, vanwege hun fundamentele observationele verschillen. Onder deze aannames is de standaard HBV aldus volledig gekarakteriseerd door haar gemiddelde bezettingsgraad van sterrenstelsels die zich in een halo van massa  $M$  bevinden. In principe kan de HBV een functie van eigenschappen anders dan halomassa zijn, wat ons kan helpen het verband te leggen tussen de sterrenstelsels en de assemblagegeschiedenis van donkeremateriehalo's (Wechsler & Tinker 2018).

De HBV-modellen kunnen verder uitgebreid worden om beter te lijken op sterrenstelselwaarnemingen en -populaties. De voorwaardelijke lichtkracht- (VLF) en voorwaardelijke stellairemassafuncties (VSMF) beschrijven de volledige verdeling van stellaire massa's en lichtkrachten van sterrenstelsels als functie van de halomassa. Ze zijn daarnaast doorgaans gescheiden in bijdragen van centrale sterrenstelsels en satellietsterrenstelsels en kunnen direct gemeten worden van een monster van sterrenstelselgroepen en -clusters (van den Bosch et al. 2013; Cacciato et al. 2013).

Zowel de VLF als de HBV-modellen specificeren het aantal sterrenstelsels per halo en de modelvoorspellingen kunnen op twee manieren gemaakt worden. Het meest voor de hand liggend is het bevolken van de donkeremateriehalo's uit een meerlichamensimulatie gebruikmakend van een Monte Carloaanpak, gevolgd door het meten van de eigenschappen van sterrenstelsels uit de samengestelde catalogus. Als alternatief kunnen zowel VLF als HBV gecombineerd worden met een analytisch halomodel om de observabelen op een semi-analytische manier te voorspellen (Seljak 2000; Cooray & Sheth 2002). De halomodelaanpak neemt aan dat alle materie in het heelal zich bevindt in halo's die als gravitationeel gebonden objecten kunnen worden beschouwd die ontkoppeld zijn van de uitdijning van het heelal en sferisch ineengestort zijn, met een massa  $M$  besloten in een straal waar de gemiddelde dicht-

heid 200 keer groter is dan de gemiddelde dichtheid van het heelal. De abundantie van donkeremateriehalo's kan dan gekarakteriseerd worden door de halomassafunctie, welke het aantal halo's gegeven een massa  $M$  oplevert. Als we de resultaten nemen van de meerlichamensimulaties waaruit gevonden werd dat het dichtheidsprofiel een universele massafunctie volgt (Navarro et al. 1997) en deze combineren met de halomassafunctie en het aanvankelijke materievermogenspectrum alsook de HBV/VLF-modellen, zijn we in staat om een veelvoud aan observabelen te voorspellen, waardoor we dan de connectie tussen sterrenstelsels en donkere materie in meer detail kunnen bestuderen.

Onder de aanname dat de sterrenstelsel- en donkerematerie-eigenschappen nauw verbonden zijn, is de meest begrenzende observationele meting voor ieder model de abundantie van sterrenstelsels. Het model van de connectie tussen sterrenstelsels en donkere materie, gegeven het kosmologisch model, zou in staat moeten zijn om de abundantie van sterrenstelsels te voorspellen als functie van hun stellaire massa of lichtkracht. Hoewel deze observationele eigenschap de meest begrenzende is, neemt ze niet alle eigenschappen van de connectie tussen sterrenstelsels en donkere materie in ogenschouw en kan tot een verkeerde interpretatie leiden. Om dit te overkomen kan men ook andere meetmethoden gebruiken (samen met de sterrenstelselabundantie) om een beter perspectief te krijgen van de connectie tussen sterrenstelsels en donkere materie. De volgende meting die men kan gebruiken is de tweepuntsclustering van sterrenstelsels. Omdat de abundantie van donkeremateriehalo's sterk verbonden is aan hun clusteringeigenschappen, zal de relatie van stellaire tot halomassa ook de clusteringeigenschappen voorspellen van de sterrenstelsels die zich in die halo's bevinden. De tweepuntsclustering van sterrenstelsels samen met de sterrenstelselabundantie zal aldus ieder model dat men wil gebruiken om de connectie tussen de donkere materie en sterrenstelsels te beschrijven, en aldus om over de aard van donkere materie te leren, volledig karakteriseren.

Met het halomodel kan men ook voorspellingen verwerven voor de kruiscorrelatie tussen sterrenstelsels en donkere materie (de tweepuntsclustering van sterrenstelsels beschrijft de correlatie tussen de sterrenstelsels). Het halomodel voorspelt de correlatiefunctie  $\xi_{\text{gm}}$  van sterrenstelsels en massa, gerelateerd aan het overschot van oppervlaktemassadichtheid  $\Delta\Sigma$ , welke tot eerste orde de geprojecteerde profielen van donkeremateriehalo's meet, die op hun beurt direct gemeten kunnen worden gebruikmakend van de lenswerking tussen sterrenstelsels. Lenswerking tussen sterrenstelsels is de meting van het zwaartekrachtlenseffect rond individuele sterrenstelsels en haar sterkte is rechtevenredig met de massa van de donkeremateriehalo's rond die sterrenstelsels. Het signaal van de lenswerking tussen sterrenstelsels is in de meeste gevallen vrij ingewikkeld om te interpreteren omdat centrale en satellietsterrenstelsels zich in volledig verschillende halo's bevinden, aldus is een volledig model van de connectie tussen sterrenstelsels en donkere materie nodig. Desalniettemin leveren de metingen van lenswerking tussen sterrenstelsels een complementaire blik op de metingen van abundanties en tweepuntsclustering van sterrenstelsels. Het belangrijkste is dat ze direct massaschattingen van de donkeremateriehalo's leveren en dat ze ook gebruikt worden om de overgebleven degeneraties in het model, namelijk de sterrenstelselvertekening, te verbreken.

## 7.2 ZWAARTEKRACHTLENSWERKING

Einsteins honderd jaar oude algemene relativiteitstheorie (Einstein 1916) beschrijft zwaartekracht als een kromming van ruimtetijd rond een zwaar object. Terwijl licht een recht pad aflegt door vlakke ruimtetijd, zal het pad van een lichtstraal veranderen wanneer ze door een gekromde ruimtetijd beweegt. Dit betekent dat het licht dat uit de verre uithoeken van het heelal komt beïnvloed kan zijn door de verdeling van massa op zijn weg. De relativistische beschrijving kan vereenvoudigd worden om tot een theorie te komen die volledig ontwikkeld kan worden in het Newtoniaanse raamwerk. Omdat dit effect analoog is aan optische lenswerking, staat dit effect bekend als *zwaartekracht lenswerking*.

De zwaartekracht lenswerking kan gebruikt worden om de materieverdeling van zware objecten in het heelal te onderzoeken. Het waargenomen effect van zwaartekracht lenswerking op een beeld van een achtergrondsterrenstelsel is een vergroting en een getijde-achtige uitrekking van de oorspronkelijke vorm. De getijde-achtige uitrekking van de beelden is rechtevenredig met de hoeveelheid massa die zich tussen zo'n sterrenstelsel en ons als waarnemers bevindt en ze kan gebruikt worden om de massa's van donkeremateriehalo's te meten gebruikmakend van de lenswerking tussen sterrenstelsels (bijv. Leauthaud et al. 2011; van Uitert et al. 2011; Velandier et al. 2014; Cacciato et al. 2014; Viola et al. 2015). Zwaartekracht lenswerking kan ook gebruikt worden om de aard van het heelal te bestuderen met de lenswerking door de grootschalige structuur zelf, genaamd de kosmische afschuiving (Bartelmann & Schneider 2001; Hildebrandt et al. 2017).

## 7.3 DIT PROEFSCHRIFT

In dit proefschrift verkennen we verschillende aspecten van de connectie tussen sterrenstelsels en donkere materie die gemeten kunnen worden met zwaartekracht lenswerking, specifiek, gebruikmakend van lenswerking tussen sterrenstelsels als onze primaire meetmethode. We gebruiken het halomodel samen met de halobezettingsverdelingen om de connectie tussen sterrenstelsels en halo's statistisch te beschrijven en om de assemblagevertekening in rijke sterrenstelselgroepen te begrenzen. Hetzelfde theoretische raamwerk is ook gebruikt om de aard van sterrenstelselvertekening te begrenzen. Daarnaast herevalueren we de prestaties van de tweedimensionale aanpak van lenswerking tussen sterrenstelsels en bestuderen we de systematische fouten die zouden kunnen optreden in deze andere bestudering. Ten slotte gebruiken we de tweedimensionale methode om de relatie van stellaire tot halomassa van satellieten te meten.

In **Hoofdstuk 2** onderzoeken we mogelijke kenmerken van haloassemblagevertekening voor spectroscopisch geselecteerde sterrenstelselgroepen uit de Galaxy And Mass Assembly-opmeting (GAMA-opmeting) gebruikmakend van metingen van de zwakke zwaartekracht lenswerking uit ruimtelijk overlappende regionen van de diepere en hogere beeldkwaliteit hebbende fotometrische Kilo-Degree Survey (KiDS). We gebruiken GAMA-groepen met een schijnbare rijkheid groter dan 4 om monsters



te identificeren met vergelijkbare gemiddelde moederhalomassa's maar met een andere radiële verdeling van satellietsterrenstelsels, wat instaat voor de formatietijd van de halo's. We meten het zwakkelenwerkingssignaal voor groepen met een steilere en een vlakke satellietverdeling dan gemiddeld en vinden geen teken van haloassemblagevertekening, met de vertekeningverhouding van  $0.85^{+0.37}_{-0.25}$ , wat consistent is met de  $\Delta$ KDM-voorspelling. Onze sterrenstelselgroepen hebben typische massa's van  $10^{13} M_{\odot}/h$ , wat op een natuurlijke manier eerdere studies van haloassemblagevertekening op schalen van sterrenstelselclusters complementeert.

In **Hoofdstuk 3** meten we de geprojecteerde sterrenstelselclustering en de signalen van lenswerking tussen sterrenstelsels gebruikmakend van de GAMA-opmeting en KiDS-opmeting om sterrenstelselvertekening te bestuderen. We gebruiken het concept van niet-lineaire en stochastische sterrenstelselvertekening in het raamwerk van halobezettingsstatistieken om de parameters van de halobezettingsstatistieken te begrenzen en om de oorsprong van sterrenstelselvertekening te onthullen. De vertekeningfunctie  $\Gamma_{\text{gm}}(r_p)$  wordt geëvalueerd gebruikmakend van het analytische halomodel waaruit gevolgtrekkingen gemaakt kunnen worden over de schaalafhankelijkheid van  $\Gamma_{\text{gm}}(r_p)$  en de oorsprong van de niet-lineariteit en stochasticiteit in halobezettingsmodellen. Onze waarnemingen onthullen de fysieke reden voor de niet-lineariteit en stochasticiteit, verder verkend gebruikmakend van hydrodynamische simulaties, waar de stochasticiteit voornamelijk voortkomt uit het niet-Poissoniaanse gedrag van satellietsterrenstelsels in de donkeremateriehalo's en hun ruimtelijke verdeling, welke niet volgt uit de ruimtelijke verdeling van donkere materie in de halo. De waargenomen niet-lineariteit is voornamelijk te wijten aan de aanwezigheid van de centrale sterrenstelsels, zoals opgemerkt was uit eerder theoretisch werk aan hetzelfde onderwerp. We zien ook dat over het geheel genomen zwaardere sterrenstelsels een sterkere schaalafhankelijkheid tonen, en tot op grotere stralen. Onze resultaten laten zien dat een schat aan informatie over sterrenstelselvertekening verborgen is halobezettingsmodellen. Deze modellen zouden daarom gebruikt moeten worden om de invloed van sterrenstelselvertekening te bepalen in kosmologische studies.

In **Hoofdstuk 4** bekijken we opnieuw de prestaties en systematische fouten van de tweedimensionale aanpak van lenswerking tussen sterrenstelsels. Deze methode buit de informatie over de daadwerkelijke posities en afplattingen van bronsterrenstelsels uit, in plaats van alleen gebruik te maken van de algemene eigenschappen van statistisch equivalente monsters. We vergelijken de prestaties van deze methode met het traditioneel gebruikte eendimensionale tangentiële afschuivingsignaal op een set van gemodelleerde data die lijkt op de huidige top van zwakkelenwerking-opmetingen. We vinden dat onder geïdealiseerde omstandigheden, de betrouwbaarheidsregio's van simultane begrenzingen voor de amplitude- en schaalparameters van het NFW-model in de tweedimensionale analyse meer dan 3 keer nauwer kunnen zijn dan de eendimensionale resultaten. Bovendien is deze verbetering afhankelijk van de lensaantaldichtheid en is ze groter voor hogere dichtheden. We vergelijken de methoden met de resultaten van de hydrodynamische EAGLE-simulatie ten behoeve van het testen voor mogelijke systematische fouten die zouden kunnen optreden vanwege missende lenssterrenstelsels, en vinden dat de methode in staat is om onvertekende schattingen van halomassa's te geven wanneer ze wordt ver-

geleken met de echte eigenschappen van de EAGLE-sterrenstelsels. Vanwege haar voordelen in gebieden met een hoge sterrenstelseldichtheid, is de methode bijzonder geschikt voor het bestuderen van de eigenschappen van satellietsterrenstelsels in clusters van sterrenstelsels.

In **Hoofdstuk 5** gebruiken we data van de opmetingen Kilo-Degree Survey (KiDS) en Galaxy And Mass Assembly (GAMA) om tegelijkertijd de relaties van stellaire tot halomassa van zowel centrale als satellietsterrenstelsels te begrenzen, voor spectroscopisch bevestigde sterrenstelsels in sterrenstelselgroepen, gebruikmakend van zwakke lenswerking. Voor de analyse gebruiken we de traditionele eendimensionale methode in de vorm van de gestapelde tangetiële afschuivingsmetingen om de halo- en subhalomassa's van onze sterrenstelsels te bepalen en om de relatie van stellaire tot halomassa te begrenzen, en daarnaast een tweedimensionale fit aan het volledige afschuivingsveld die alle beschikbare informatie over lenssterrenstelsels en precieze bronsterrenstelselposities en -afplattingen gebruikt. We vinden dat de tweedimensionale methode statistisch beter presteert dan de eendimensionale methode. Beide methoden leiden tot vergelijkbare parameters van de relatie van stellaire tot halomassa, welke consistent zijn met eerdere resultaten die in de literatuur gevonden zijn, wat laat zien dat de satellietsterrenstelsels in het algemeen lagere halomassa's hebben dan de centrale sterrenstelsels, gegeven dezelfde stellaire massa.

Ten slotte presenteren we in **Hoofdstuk 6** de vierde publieke data-uitgifte van de Kilo Degree Survey welke het oppervlak van de hemel beslagen door data-uitgifte 3, de data die we hoofdzakelijk gebruikten als onze bron van metingen van zwaartekracht lenswerking in de bovenstaande hoofdstukken, meer dan verdubbelt. Mijn bijdrage aan het artikel waarop dit Hoofdstuk is gebaseerd bestond uit het leiden van de ASTRO-WISE fotometrische datareductie, welke gebruikt werd om gestapelde afbeeldingen van 1006 richtingen in de vier banden te produceren, waaruit de fotometrie in de catalogi verkregen is. De precieze bijdrage is beschreven in Hoofdstuk 6, specifiek in Paragraaf 6.3.1.

## 8.1 POVEZAVA MED GALAKSIJAMI IN TEMNO SNOVJO

**G**LAVNA tema te disertacije je osredotočena na lastnosti in haloje temne snovi in na njihovo povezavo z galaksijami, ki jih lahko opazujemo v vesolju. Zaradi še vedno neznane narave temne snovi se le ta ponavadi preučuje z uporabo njene porazdelitve in lastnosti na galaktičnih in večjih razdaljah. Povezava med galaksijami in temno snovjo je pomembna iz treh glavnih razlogov, njeno razumevanje pa pomaga pri odgovoru na največja vprašanja v astrofiziki in kozmologiji danes. Prvo vprašanje vključuje razumevanje fizike tvorbe galaksij. Drugič, pomaga pri merjenju kozmoloških parametrov – če hočemo natančno izmeriti kozmološke parametre, moramo razumeti kako se galaksije povežejo s temno snovjo, in tretjič, pomaga pri poznavanju razvoja porazdelitve snovi in lastnosti temne snovi (Wechsler & Tinker 2018).

V standardnem kozmološkem modelu, v katerem prevladuje hladna temna snov in kozmološka konstanta (tako imenovani  $\Lambda$ CDM model), formiranje struktur v vesolju vodi predvsem dinamika hladne temne snovi. Gravitacijski kolaps fluktuacij gostote temne snovi in njihova kasnejša virializacija privedeta do tvorbe halojev temne snovi iz najbolj gostih mest v začetnem naključnem Gaussovem polju. Splošno je sprejeto, da vsaka galaksija prebiva v matičnem haloju iz temne snovi. Galaksije, ki se nahajajo na dnu gravitacijskega potenciala haloja temne snovi, so označene kot osrednje galaksije in galaksije, ki krožijo okoli osrednje galaksije, se imenujejo satelitske galaksije. Natančen način, kako galaksije poseljujejo haloje temne snovi, je še vedno tema aktivnih raziskav, h katerim ta disertacija prispeva nekaj spoznanj. Na splošno se povezava med populacijo galaksij in populacijo halojev temne snovi statistično modelira s Press-Schechterjevim formalizmom (Press & Schechter 1974). Ta formalizem postulira, da se masivne galaksije tvorijo na najbolj gostih delih osnovne porazdelitve temne snovi in da je povezava med galaksijam in temno snovjo mogoče povezati s količino, imenovano galaktični bias. Medtem ko ta formalizem napove pravilno število halojev kot funkcijo njihove mase in tudi dobro napove kopičenje galaksi, ima nekaj pomanjkljivosti. Bolj natančno, galaktični bias ima lahko ne-trivialno odvisnost od mase in razdalje. Splošno je sprejeto, da ta povezava ni linearna ali

deterministična. Poleg tega, zgodovina sestavljanja halojev temne snovi pušča odtise v opazovanih porazdelitvah galaksij, ki se nato ne ujemajo več s teoretičnimi napovedmi. Zgodovina sestavljanja, mehanizmi odstranjevanja mase in združevanja satelitskih galaksij nam prav tako puščajo različne odtise v povezavi temne snovi med osrednjimi in satelitskimi galaksijami. Zaradi teh dejstev obstaja veliko različnih modelov, ki temeljijo na statističnih postulatih Press-Schechterjevega formalizma in omogočajo da se povezava med temno snovjo in galaksijami modelira bolj natančno.

Priljubljen in uspešen način opisovanja povezave med galaksijami in temno snovjo je skozi porazdelitve zasedenosti halojev (halo occupation distribution – HOD), ki opisujejo verjetnostno porazdelitev števila galaksij z določeno lastnostjo (izsev ali vidna masa) v haloju, kot funkcijo mase haloja. Porazdelitve zasedenosti halojev se zaradi temeljnih razlik v lastnostih osrednjih in satelitskih galaksij posebej izrazijo za osrednje galaksije in satelitske galaksije. V skladu s temi predpostavkami je standardni HOD v celoti opisan s povprečnim številom galaksij, ki prebivajo v haloju z maso  $M$ . Načeloma je HOD lahko odvisen tudi od drugih lastnosti in ne samo mase haloja, kar nam lahko pomaga povezati galaksije z zgodovino sestavljanja halojev temne snovi (Wechsler & Tinker 2018).

HOD modele je mogoče še bolj posplošiti na tak način da lahko bolje opišejo opazovane populacije galaksij. Pogojna izsevnna funkcija (conditional luminosity function – CLF) in pogojna funkcija vidne mase (conditional stellar mass function – CSMF) popolnoma opišeta porazdelitev vidne mase galaksij ali njihovega izseva kot funkcijo mase haloja. Običajno tudi ti dve funkciji razdelimo na dva dela: prispevek osrednjih galaksij in prispevek satelitskih galaksij. Te funkcije je mogoče neposredno določiti na vzorcu skupin in jat galaksij (van den Bosch et al. 2013; Cacciato et al. 2013).

Oba načina modeliranja povezave med galaksijami in temno snovjo, CLF in HOD, specificirata število galaksij na halo, napovedi iz modela pa lahko uporabimo na dva načina. Najbolj preprosto je, da s pomočjo računalniške simulacije  $N$ -teles ustvarimo haloje temne snovi in jih z metodo Monte Carlo naselimo z galaksijami, izbira galaksij pa sledi našemu izbranemu modelu. Nato izmerimo lastnosti galaksij iz sestavljenega kataloga in jih primerjamo z opazovanji. Drugi način uporabe vključuje združevanje CLF in HOD z analitičnim halo modelom s katerim lahko napovemo opazljive količine na analitičen način (Seljak 2000; Cooray & Sheth 2002). Pristop s halo modelom predpostavlja, da je vsa materija v Vesolju v halojih, ki jih je mogoče obravnavati kot gravitacijsko vezane objekte, ki so se odklopili od širitve Vesolja in se sferično sesedli pod lastnim gravitacijskim privlakom in imajo maso  $M$ , ki je vsebovana v sferi katere povprečna gostota je 200-krat večja od povprečne gostote vesolja. Število halojev temne snovi lahko nato opišemo z masno funkcijo, kar nam da število halojev z maso  $M$ . Če uporabimo rezultate simulacij  $N$ -teles, kjer je bilo ugotovljeno, da profil gostote halojev sledi univerzalni funkciji (Navarro et al. 1997) in ta profil združimo z masno funkcijo, začetnim spektrom moči snovi, in s HOD/CLF modelom, lahko napovemo množico opazljivk, s pomočjo katerih lahko podrobneje preučimo povezavo med galaksijami in temno snovjo.

Pod predpostavko, da so lastnosti galaksij in halojev temne snovi tesno povezane, je za vse modele najbolj omejujoča opazovalna meritev številčnost galaksij. Vsak model povezave med galaksijami in temno snovjo, v danem kozmološkem okvirju, bi

moral biti sposoben predvideti število galaksij kot funkcijo njihove vidne mase ali izseva. Čeprav je ta lastnost najbolj omejujoča, ne upošteva vseh lastnosti povezave med galaksijami in temno snovjo in lahko vodi do napačnih razlag. Da premagali to pomanjkljivost, lahko uporabimo tudi druge opazljivke (skupaj z številom galaksij), da dobimo boljši pogled na relacijo med temno snovjo in galaksijami. Naslednja meritev, ki jo lahko uporabimo, je dvotočkovna avtokorelacijska funkcija kopičenja galaksij. Ker je število halojev temne snovi močno povezano z njihovim kopičenjem, bo razmerje med vidno in temno snovjo opisalo tudi lastnosti kopičenja galaksij, ki prebivajo v teh halojih. Dvotočkovna avtokorelacijska funkcija, skupaj s številom galaksij, bo tako v celoti karakterizirala kateri koli model, ki ga želimo uporabiti za opis povezave med temno snovjo in galaksijami ter tako posredno spoznati naravo temne snovi.

S halo modelom je mogoče pridobiti tudi napovedi za dvotočkovno križnokorelacijsko funkcijo med galaksijami in temno snovjo (dvotočkovna avtokorelacijska funkcija opisuje korelacijo med galaksijami). Halo model nam da dvotočkovno križnokorelacijsko funkcijo med galaksijami in maso  $\xi_{gm}$ , ki je povezana s presežno površinsko gostoto  $\Delta\Sigma$ , ki v prvem približku opisuje projiciran radialni profil halojev temne snovi, ki ga je mogoče neposredno izmeriti z šibkim gravitacijskim lečenjem med posameznimi galaksijami (lečenje galaksija-galaksija). Lečenje galaksija-galaksija je meritev učinka gravitacijskega lečenja okoli posameznih galaksij, njegova moč pa je neposredno sorazmerna z maso halojev temne snovi okoli teh galaksij. Signal lečenja galaksija-galaksija je v večini primerov precej zapleteno opisati, saj osrednje in satelitske galaksije prebivajo v popolnoma različnih halojih, zato je potreben celoten model povezave med galaksijami in temno snovjo. Kljub temu meritve lečenja galaksija-galaksija dopolnjujejo pogled na meritve števila in dvotočkovne avtokorelacijske funkcije. Najpomembneje je, da lečenje galaksija-galaksija neposredno meri maso temne snovi in se tako tudi uporablja tudi za razvozlanje preostalih degeneriranosti v modelu, in sicer galaktičnega biasa.

## 8.2 GRAVITACIJSKO LEČENJE

Einsteinova sto let stara teorija splošne relativnosti (Einstein 1916) opisuje gravitacijo kot ukrivljenost prostora in časa okoli masivnih objektov. Ko svetloba potuje po ravni poti skozi raven prostor-čas, se bo vsakič, ko gre skozi ukrivljen prostor-čas, spremenila njena pot. To pomeni, da na svetlobo, ki potuje iz oddaljenih delov Vesolja, vpliva porazdelitev mase na njeni poti. Relativistični opis gravitacije na pot svetlobe je mogoče poenostaviti tako, da tvori teorijo, ki jo je mogoče popolnoma razviti v newtonskem okviru. Ker je učinek analogen optičnemu lečenju, je ta učinek znan kot *gravitacijsko lečenje*.

Gravitacijsko lečenje lahko uporabimo za ugotavljanje porazdelitve snovi v masivnih objektih v vesolju. Gravitacijsko lečenje nam bo sliko oddaljene galaksije povečalo in raztegnilo v smeri plimske sile. Plimsko raztezanje slik je neposredno sorazmerno s količino mase, ki je prisotna med takšno galaksijo in nami kot opazovalci, in jo lahko uporabimo za merjenje mase halojev temne snovi z lečenjem galaksija-

galaksija (npr. Leauthaud et al. 2011; van Uitert et al. 2011; Velander et al. 2014; Cacciato et al. 2014; Viola et al. 2015). Gravitacijsko lečenje lahko uporabimo tudi za preučevanje lastnosti Vesolja z lečenjem same strukture Vesolja, imenovano kozmično striženje (Bartelmann & Schneider 2001; Hildebrandt et al. 2017).

### 8.3 TA DISERTACIJA

V tej disertaciji smo raziskali različne vidike povezave galaksijami in temno snovjo, ki jih je mogoče meriti z gravitacijskim lečenjem, natančneje z uporabo lečenja galaksija-galaksija kot naše primarne metode. Za to smo uporabili halo model skupaj s porazdelitvami zasedenosti halojev za statistični opis povezave med galaksijami in haloji in tako izmerili "assembly bias" v skupinah galaksij. Isti teoretični model smo uporabili tudi za meritve lastnosti galaktičnega biasa. Poleg tega smo ponovno ocenili zmožnosti dvodimenzionalnega pristopa k lečenju galaksija-galaksija ter hkrati preučili napake, ki bi se lahko pojavile pri tej metodi. Na koncu smo uporabili dvodimenzionalno metodo za meritev razmerja vidne mase in temne snovi v satelitskih galaksijah.

V **2. poglavju** smo raziskovali možne odtise zgodovine sestavljanja halojev za spektroskopsko izbrane skupine galaksij iz pregleda neba Galaxy and Mass Assembly (GAMA) z uporabo šibkega gravitacijskega lečenja iz prekrivajočih se območij globljšega visokokakovostnega fotometričnega pregleda neba Kilo-Degree Survey (KiDS). Uporabili smo skupine galaksij iz GAMA pregleda neba s številom galaksij v skupini večjim od 4 iz katerih smo dobili vzorce skupin galaksij s primerljivimi povprečnimi masami halojev, vendar z drugačno porazdelitvijo satelitskih galaksij okoli osrednjih galaksij. Porazdelitev satelitskih galaksij je povezana s časom nastanka primarnih halojev. Izmerili smo signal gravitacijskega lečenja za skupine, ki imajo porazdelitev satelitskih galaksij bolj ali manj oddaljene od povprečja vseh skupin v vzorcu in ne najdemo nobenih znakov "assembly bias"-a. Razmerje biasa je  $0,85^{+0,37}_{-0,25}$ , kar je skladno z napovedjo  $\Lambda$ CDM modela. Naše skupine galaksij imajo značilne mase  $10^{13} M_{\odot}/h$ , kar dopolnjuje prejšnje raziskave zgodovine sestavljanja halojev na jatah galaksij.

V **3. poglavju** smo izmerili projicirano dvotočkovno avtokorelacijsko funkcijo galaksij in lečenje galaksija-galaksija s pomočjo GAMA in KiDS pregledov neba. Meritve smo nato uporabili za proučevanje lastnosti galaktičnega biasa. Koncept nelinearnega in stohastičnega galaktičnega biasa smo uporabili v HOD statističnem modelu s katerim lahko izmerimo proste parametre v HOD modelu in tako tudi razkrijemo naravo galaktičnega biasa. Funkcijo, s katero lahko opišemo galaktični bias  $\Gamma_{\text{gm}}(r_p)$ , smo ocenili s pomočjo analitičnega halo modela, s pomočjo katerega lahko ugotovimo kako se funkcija  $\Gamma_{\text{gm}}(r_p)$  obnaša na različnih razdaljah in kje je izvir nelinearnosti in stohastičnosti v modelih HOD. Naša opazovanja so razkrila fizični razlog za nelinearnost in stohastičnost, ki smo ga nadalje podkrepili z analizo hidrodinamičnih računalniških simulacij. Glavni razlog za stohastičnost modelov HOD večinoma izvira iz ne-Poissonovega obnašanja satelitskih galaksij v halojih temne snovi in njihove prostorske razporeditve, ki ne sledi prostorski porazdelitvi temne snovi v haloju.

Opazena nelinearnost je večinoma posledica prisotnosti osrednjih galaksij, kar je v skladu s predhodnimi teoretičnimi raziskavami. Prav tako smo opazili, da bolj masivne galaksije močneje variirajo na različnih razdaljah. Naši rezultati kažejo, da se v HOD modelih skriva veliko informacij o galaktičnem biasu in je zato te modele priporočljivo uporabiti za določitev vpliva galaktičnega biasa v kozmoloških študijah.

V **4. poglavju** smo ponovno preučili zmožnosti in napake dvodimenzionalne metode lečenja galaksija-galaksija. Ta metoda izkorišča vse informacije o lečenih galaksijah, kot so njihovi natančni položaji in eliptičnosti. Tako ne uporablja samo lastnosti vzorca statistično enakovrednih galaksij, kot je to običajno pri enodimenzionalnem lečenju. Učinkovitost te metode smo primerjali s tradicionalno uporabljenim enodimenzionalnim tangencialnim strižnim signalom na naboru umetno generiranih podatkov, ki so generirani tako, da upoštevajo vse lastnosti trenutno naj sodobnejših pregledov neba, ki se uporabljajo za raziskave gravitacijskega lečenja. Ugotovili smo, da so v idealiziranih okoliščinah območja zaupanja meritev parametrov modela NFW v dvodimenzionalni analizi več kot 3-krat strožja od enodimenzionalnih rezultatov. Poleg tega je to izboljšanje odvisno od gostote števila leč in je večje pri večjih gostotah. Metodo smo primerjali tudi s pomočjo meritev na hidrodinamični računalniški simulaciji EAGLE, da bi preverili morebitne napake, ki bi se lahko pojavile zaradi manjkajočih galaksij, in ugotovili, da je metoda sposobna vrniti nepristranske ocene mas halojev. Zaradi svoje prednosti na območjih z visoko številčno gostoto galaksij je metoda še posebej primerna za preučevanje lastnosti satelitskih galaksij v jatah.

V **5. poglavju** smo uporabili podatke iz KiDS in GAMA pregledov neba, s pomočjo katerih smo sočasno izmerili razmerje med vidno maso osrednjih in satelitskih galaksij in maso njenih halojev. Razmerje smo izmerili za galaksije ki so spektroskopsko potrjene v skupinah galaksij v pregledu neba GAMA. Uporabili smo tako tradicionalno enodimenzionalno metodo v obliki tangencialnih strižnih meritev za določitev mase halojev naših galaksij ter za meritev razmerja med vidno maso in maso temne snovi, kot tudi dvodimenzionalno metodo. Za dvodimenzionalno metodo uporabimo celotno gravitacijsko strižno polje, ki uporabi vse razpoložljive informacije o lečah in natančnih položajih oddaljenih galaksij in njih eliptičnosti. Ugotovili smo, da je dvodimenzionalna metoda statistično boljša kot enodimenzionalna metoda. Obe metodi vodita do podobnih parametrov izmerjenih razmerij, ki so skladni s prejšnjimi rezultati v literaturi. Rezultati kažejo, da imajo satelitske galaksije na splošno nižje mase halojev kot osrednje galaksije glede na isto vidno maso.

Na koncu, v **poglavju 6** predstavljamo četrto javno objavo podatkov pregleda neba KiDS, ki več kot podvoji območje opazovanega dela neba, ki je bilo zajeto v prejšnji javni objavi podatkov, ki smo jih uporabili za glavni vir meritev šibkega gravitacijskega lečenja v zgornjih poglavjih. Moj prispevek k objavljenemu članku, na katerem temelji to poglavje, je vseboval vodenje obdelave fotometričnih podatkov s pomočjo sistema ASTRO-WISE, ki je bil uporabljen za izdelavo mozaika iz 1006 posameznih fotografij v štirih vidnih pasovih, iz katerih je pridobljena fotometrija v objavljenih katalogih. Natančen prispevek je opisan v poglavju 6, natančneje v razdelku 6.3.1.





---

# Bibliography

---

- Abazajian K. N., et al., 2009, *ApJS*, 182, 543
- Aihara H., et al., 2018, *Publ. Astron. Soc. Japan*, 70, 1
- Alam S., et al., 2015, *ApJS*, 219, 12
- Amaro V., et al., 2019, *MNRAS*, 482, 3116
- Amon A., et al., 2018a, *MNRAS*, 477, 4285
- Amon A., et al., 2018b, *MNRAS*, 479, 3422
- Artale M. C., et al., 2017, *MNRAS*, 470, 1771
- Asgari M., et al., 2019, *A&A*, 624, A134
- Bartelmann M., Schneider P., 2001, *Phys. Rep.*, 340, 291
- Begeman K., Belikov A. N., Boxhoorn D. R., Valentijn E. A., 2012, *Exp. Astron.*, 35, 1
- Bellagamba F., et al., 2019, *MNRAS*, 484, 1598
- Benitez N., 2000, *ApJ*, 536, 571
- Bertin E., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, *Astronomical Society of the Pacific Conference Series Vol. 351, Astron. Data Anal. Softw. Syst. XV*. p. 112
- Bertin E., 2010a, *SCAMP: Automatic Astrometric and Photometric Calibration*
- Bertin E., 2010b, *SWarp: Resampling and Co-adding FITS Images Together*
- Bertin E., Arnouts S., 1996, *Astron. Astrophys. Suppl. Ser.*, 117, 393
- Bertin E., Mellier Y., Radovich M., Missonnier G., Didelon P., Morin B., 2002, in Bohlender D. A., Durand D., Handley T. H., eds, *Astronomical Society of the Pacific Conference Series Vol. 281, Astron. Data Anal. Softw. Syst. XI*. p. 228
- Bertone G., Hooper D., 2018, *Rev. Mod. Phys.*, 90, 045002
- Bhattacharya S., Heitmann K., White M., Lukić Z., Wagner C., Habib S., 2011, *ApJ*, 732, 122
- Bilicki M., et al., 2018, *A&A*, 616, A69
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645

- Brouwer M. M., et al., 2016, *MNRAS*, 462, 4451
- Brouwer M. M., et al., 2017, *MNRAS*, 466, 2547
- Brouwer M. M., et al., 2018, *MNRAS*, 481, 5189
- Brown A. G. A., et al., 2018, *A&A*, 616, A1
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Buddendiek A., et al., 2016, *MNRAS*, 456, 3886
- Cacciato M., 2016, Personal communication
- Cacciato M., van den Bosch F. C., More S., Li R., Mo H. J., Yang X., 2009, *MNRAS*, 394, 929
- Cacciato M., Lahav O., van den Bosch F. C., Hoekstra H., Dekel A., 2012, *MNRAS*, 426, 566
- Cacciato M., van den Bosch F. C., More S., Mo H., Yang X., 2013, *MNRAS*, 430, 767
- Cacciato M., van Uitert E., Hoekstra H., 2014, *MNRAS*, 437, 377
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
- Capaccioli M., Schipani P., 2011, *The Messenger*, 146, 2
- Chabrier G., 2003, *ApJ*, 586, L133
- Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
- Costa-Duarte M. V., et al., 2018, *MNRAS*, 478, 1968
- Coupon J., et al., 2012, *A&A*, 542, A5
- Courteau S., et al., 2014, *Rev. Mod. Phys.*, 86, 47
- Crain R. A., et al., 2015, *MNRAS*, 450, 1937
- Daddi E., Cimatti A., Renzini A., Fontana A., Mignoli M., Pozzetti L., Tozzi P., Zamorani G., 2004, *ApJ*, 617, 746
- Dalal N., White M., Bond J. R., Shirokov A., 2008, *ApJ*, 687, 12
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- Dekel A., Lahav O., 1999, *ApJ*, 520, 24
- Dekel A., Rees M. J., 1987, *Nature*, 326, 455
- Driver S. P., et al., 2011, *MNRAS*, 413, 971

- Driver S. P., Davies L. J., Meyer M., Power C., Robotham A. S., Baldry I. K., Liske J., Norberg P., 2016, *Astrophys. Sp. Sci. Proc.*, 42, 205
- Drlica-Wagner A., et al., 2018, *ApJS*, 235, 33
- Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, *MNRAS*, 390, L64
- Dvornik A., et al., 2017, *MNRAS*, 468, 3251
- Dvornik A., et al., 2018, *MNRAS*, 479, 1240
- Dvornik A., Zoutendijk S. L., Hoekstra H., Kuijken K., 2019, *A&A*, 627, A74
- Eardley E., et al., 2015, *MNRAS*, 448, 3665
- Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J., 2013, *The Messenger*, 154, 32
- Einstein A., 1916, *Ann. Phys.*, 354, 769
- Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
- Erben T., et al., 2005, *Astron. Nachrichten*, 326, 432
- Farrow D. J., et al., 2015, *MNRAS*, 454, 2120
- Fenech Conti I., Herbonnet R., Hoekstra H., Merten J., Miller L., Viola M., 2017, *MNRAS*, 467, 1627
- Foreman-Mackey D., 2016, *J. Open Source Softw.*, 1
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Friedrich O., et al., 2018, *Phys. Rev. D*, 98, 023508
- Gao L., White S. D. M., 2007, *MNRAS*, 377, L5
- Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66
- Geiger B., Schneider P., 1999, *MNRAS*, 302, 118
- Georgiou C., et al., 2019, *A&A*, 622, A90
- Giblin B., et al., 2018, *MNRAS*, 480, 5529
- Grootes M. W., et al., 2018, *MNRAS*, 477, 1015
- Gruen D., et al., 2018, *Phys. Rev. D*, 98, 023507
- Han J., et al., 2015, *MNRAS*, 446, 1356
- Harnois-Déraps J., et al., 2017, *MNRAS*, 471, 1619
- Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399

- Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2016, *MNRAS*, 460, 2552
- Heymans C., et al., 2012, *MNRAS*, 427, 146
- Hikage C., et al., 2019, *Publ. Astron. Soc. Japan*, 71, 1
- Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, *A&A*, 499, 31
- Hildebrandt H., et al., 2012, *MNRAS*, 421, 2355
- Hildebrandt H., et al., 2017, *MNRAS*, 465, 1454
- Hildebrandt H., et al., 2018, preprint (arXiv:1812.06076)
- Hinshaw G., et al., 2013, *ApJS*, 208, 19
- Hoekstra H., 2003, *MNRAS*, 339, 1155
- Hoekstra H., 2014, *Proc. Int. Sch. Phys. Enrico Fermi*, 186, 59
- Hoekstra H., van Waerbeke L., Gladders M. D., Mellier Y., Yee H. K. C., 2002, *ApJ*, 577, 604
- Hoekstra H., Franx M., Kuijken K., Carlberg R. G., Yee H. K. C., 2003, *MNRAS*, 340, 609
- Hoekstra H., Yee H. K. C., Gladders M. D., 2004, *ApJ*, 606, 67
- Hoekstra H., Hsieh B. C., Yee H. K. C., Lin H., Gladders M. D., 2005, *ApJ*, 635, 73
- Hubble E. P., 1926, *ApJ*, 64, 321
- Hubble E., 1929, *Proc. Natl. Acad. Sci.*, 15, 168
- Hudson M. J., Gwyn S. D. J., Dahle H., Kaiser N., 1998, *ApJ*, 503, 531
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Ivezić Ž., et al., 2004, *Astron. Nachrichten*, 325, 583
- Jakobs A., et al., 2018, *MNRAS*, 480, 3338
- Joachimi B., Schneider P., Eifler T., 2008, *A&A*, 477, 43
- Joudaki S., et al., 2017, *MNRAS*, 471, 1259
- Jullo E., et al., 2012, *ApJ*, 750, 37
- Kaiser N., 1984, *ApJ*, 284, L9
- Kaiser N., Squires G., 1993, *ApJ*, 404, 441
- Kannawadi A., et al., 2019, *A&A*, 624, A92

- Kelvin L. S., et al., 2018, *MNRAS*, 477, 4116
- Köhlinger F., et al., 2017, *MNRAS*, 471, 4412
- Krause E., Eifler T., 2017, *MNRAS*, 470, 2100
- Krause E., Hirata C. M., 2010, *A&A*, 523, A28
- Kuijken K., 2011, *The Messenger*, 146, 8
- Kuijken K., et al., 2015, *MNRAS*, 454, 3500
- Kuijken K., et al., 2019, *A&A*, 625, A2
- Lacerna I., Padilla N., Stasyszyn F., 2014, *MNRAS*, 443, 3107
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Laureijs R., et al., 2011, preprint (arXiv:1110.3193)
- Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, *ApJ*, 738, 45
- Leauthaud A., et al., 2012, *ApJ*, 744, 159
- Lemaître G., 1927, *Ann. la Société Sci. Bruxelles*, A47, 49
- Li Y., Hu W., Takada M., 2014a, *Phys. Rev. D*, 89, 1
- Li R., et al., 2014b, *MNRAS*, 438, 2864
- Li R., et al., 2016, *MNRAS*, 458, 2573
- Limousin M., Kneib J. P., Bardeau S., Natarajan P., Czoske O., Smail I., Ebeling H., Smith G. P., 2007, *A&A*, 461, 881
- Lin Y.-T., Mandelbaum R., Huang Y.-H., Huang H.-J., Dalal N., Diemer B., Jian H.-Y., Kravtsov A., 2016, *ApJ*, 819, 119
- Liske J., et al., 2015, *MNRAS*, 452, 2087
- Mahlke M., et al., 2018, *A&A*, 610, A21
- Mandelbaum R., 2017, preprint (arXiv:1710.03235)
- Mandelbaum R., Seljak U., Cool R. J., Blanton M., Hirata C. M., Brinkmann J., 2006, *MNRAS*, 372, 758
- Mandelbaum R., Seljak U., Baldauf T., Smith R. E., 2010, *MNRAS*, 405, 2078
- Mandelbaum R., et al., 2018, *Publ. Astron. Soc. Japan*, 70, 2
- Marian L., Smith R. E., Angulo R. E., 2015, *MNRAS*, 451, 1418
- Martinet N., et al., 2018, *MNRAS*, 474, 712

- Matthee J., Schaye J., Crain R. A., Schaller M., Bower R., Theuns T., 2017, *MNRAS*, 465, 2381
- Maturi M., Bellagamba F., Radovich M., Roncarelli M., Sereno M., Moscardini L., Bardelli S., Puddu E., 2019, *MNRAS*, 485, 498
- McAlpine S., et al., 2016, *A&C*, 15, 72
- McFarland J. P., Verdoes-Kleijn G., Sikkema G., Helmich E. M., Boxhoorn D. R., Valentijn E. A., 2013, *Exp. Astron.*, 35, 45
- McKay M., Beckman R., Canover W., 1979, *Technometrics*, 21, 239
- Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, *MNRAS*, 454, 1958
- Miller L., et al., 2013, *MNRAS*, 429, 2858
- Miyatake H., et al., 2015, *ApJ*, 806, 1
- Miyatake H., More S., Takada M., Spergel D. N., Mandelbaum R., Rykoff E. S., Rozo E., 2016, *Phys. Rev. Lett.*, 116, 041301
- Mo H., van den Bosch F., White S., 2010, *Galaxy Formation and Evolution*. Cambridge University Press, Cambridge, doi:10.1017/CBO9780511807244, <http://ebooks.cambridge.org/ref/id/CBO9780511807244>
- More S., van den Bosch F. C., Cacciato M., Skibba R., Mo H. J., Yang X., 2011, *MNRAS*, 410, 210
- Murray S., Power C., Robotham A., 2013, *A&C*, 3-4, 23
- Nakoneczny S., Bilicki M., Solarz A., Pollo A., Maddox N., Spiniello C., Brescia M., Napolitano N. R., 2019, *A&A*, 624, A13
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Perlmutter S., et al., 1999, *ApJ*, 517, 565
- Petrillo C. E., et al., 2018, *MNRAS*, 482, 807
- Pielorz J., Rödiger J., Tereno I., Schneider P., 2010, *A&A*, 514, A79
- Planck Collaboration et al., 2013, *A&A*, 571, A16
- Planck Collaboration et al., 2016, *A&A*, 594, A1
- Planck Collaboration et al., 2018, preprint (arXiv:1807.06209)

- Press W. H., Schechter P., 1974, *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, doi:10.1086/152650
- Raichoor A., et al., 2014, *ApJ*, 797, 102
- Riess A. G., et al., 1998, *AJ*, 116, 1009
- Robotham A. S. G., et al., 2011, *MNRAS*, 416, 2640
- Roy N., et al., 2018, *MNRAS*, 480, 1057
- Rozo E., et al., 2015, *MNRAS*, 461, 1431
- Rubin V. C., Thonnard N., Ford, W. K. J., 1980, *ApJ*, 238, 471
- Rykoff E. S., et al., 2014, *ApJ*, 785, 104
- Schaller M., et al., 2015, *MNRAS*, 451, 1247
- Schaye J., et al., 2015, *MNRAS*, 446, 521
- Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
- Schneider P., 2003, preprint (arXiv:0306465)
- Schneider P., 2006, in Meylan G., Jetzer P., North P., Schneider P., Kochanek C., Wambsganss J., eds, *Saas-Fee Adv. Course 33 Gravitational Lensing Strong, Weak Micro.* pp 269–451
- Schneider P., Rix H.-W., 1997, *ApJ*, 474, 25
- Seitz S., Schneider P., Ehlers J., 1994, *Class. Quantum Gravity*, 11, 2345
- Seljak U., 2000, *MNRAS*, 318, 203
- Sergeyev A., Spiniello C., Khramtsov V., Napolitano N. R., Bannikova E., Tortora C., Getman F. I., Agnello A., 2018, *Res. Notes AAS*, 2, 189
- Shan H., et al., 2018, *MNRAS*, 474, 1116
- Sheldon E. S., et al., 2004, *AJ*, 127, 2544
- Sifón C., et al., 2015, *MNRAS*, 454, 3938
- Sifón C., Herbonnet R., Hoekstra H., van der Burg R. F. J., Viola M., 2018, *MNRAS*, 478, 1244
- Simon P., Hilbert S., 2017, *A&A*
- Simon P., Hettterscheidt M., Schirmer M., Erben T., Schneider P., Wolf C., Meisenheimer K., 2007, *A&A*, 461, 861

- Singh S., Mandelbaum R., Seljak U., Slosar A., Vazquez Gonzalez J., 2017, *MNRAS*, 471, 3827
- Skrutskie M. F., et al., 2006, *AJ*, 131, 1163
- Sonnenfeld A., Leauthaud A., 2018, *MNRAS*, 477, 5460
- Spergel D., et al., 2015, preprint (arXiv:1503.03757)
- Spiniello C., et al., 2018, *MNRAS*, 480, 1163
- Springel V., et al., 2005, *Nature*, 435, 629
- Takada M., Hu W., 2013, *Phys. Rev. D*, 87, 1
- Takada M., Jain B., 2009, *MNRAS*, 395, 2065
- Taylor E. N., et al., 2011, *MNRAS*, 418, 1587
- The Dark Energy Survey Collaboration 2005, preprint (arXiv:0510346)
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, *ApJ*, 724, 878
- Tinker J. L., George M. R., Leauthaud A., Bundy K., Finoguenov A., Massey R., Rhodes J., Wechsler R. H., 2012, *ApJ*, 755, L5
- Tortora C., Napolitano N. R., Roy N., Radovich M., Getman F., Koopmans L. V. E., Verdoes Kleijn G. A., Kuijken K. H., 2018a, *MNRAS*, 473, 969
- Tortora C., et al., 2018b, *MNRAS*, 481, 4728
- Troxel M. A., et al., 2018, *Phys. Rev. D*, 98, 043528
- Vakili M., et al., 2019, *MNRAS*
- Velander M., et al., 2014, *MNRAS*, 437, 2111
- Velliscig M., et al., 2017, *MNRAS*, 471, 2856
- Viola M., et al., 2015, *MNRAS*, 452, 3529
- Wang Y., Yang X., Mo H. J., van den Bosch F. C., Weinmann S. M., Chu Y., 2008, *ApJ*, 687, 919
- Wang H., Mo H. J., Jing Y. P., 2009, *MNRAS*, 396, 2249
- Wang L., et al., 2013, *MNRAS*, 431, 648
- Wechsler R. H., Tinker J. L., 2018, *ARA&A*, 56, 435
- Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, *ApJ*, 652, 71



- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Wibking B. D., et al., 2019, *MNRAS*, 484, 989
- Wright C. O., Brainerd T. G., 2000, *ApJ*, 534, 34
- Wright A. H., et al., 2017, *MNRAS*, 470, 283
- Wright A. H., et al., 2018, preprint (arXiv:1812.06077)
- Yang X., Mo H. J., van den Bosch F. C., 2006, *ApJ*, 638, L55
- Yang X., Mo H. J., van den Bosch F. C., 2008, *ApJ*, 676, 248
- Zehavi I., et al., 2011, *ApJ*, 736, 59
- Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, *MNRAS*, 443, 3044
- Zu Y., Mandelbaum R., Simet M., Rozo E., Rykoff E. S., 2017, *MNRAS*, 470, 551
- Zuntz J., et al., 2018, *MNRAS*, 481, 1149
- Zwicky F., 1933, *Helv. Phys. Acta*, 6, 110
- de Jong J., et al., 2013, *The Messenger*, 154, 44
- de Jong J. T. A., et al., 2015, *A&A*, 582, A62
- de Jong J. T. A., et al., 2017, *A&A*, 604, A134
- de la Torre S., 2018, Personal communication
- de la Torre S., et al., 2017, *A&A*, 608, A44
- van Uitert E., Hoekstra H., Velandier M., Gilbank D. G., Gladders M. D., Yee H. K. C., 2011, *A&A*, 534, A14
- van Uitert E., et al., 2016, *MNRAS*, 459, 3251
- van Uitert E., et al., 2017, *MNRAS*, 467, 4131
- van Uitert E., et al., 2018a, *MNRAS*, 476, 4662
- van Uitert E., et al., 2018b, *MNRAS*, 476, 4662
- van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, *MNRAS*, 430, 725



---

# Publications

---

## First author:

1. **Dvornik, A.**, Zoutendijk S. L., Hoekstra H., Kuijken K.,  
*The case for two-dimensional galaxy-galaxy lensing*,  
2019, A&A, 627, A74 (**Chapter 4**)
2. **Dvornik, A.**, Hoekstra H., Kuijken K., Schneider P., et al.,  
*Unveiling Galaxy Bias via the Halo Model, KiDS and GAMA*,  
2018, MNRAS, 479, 1240 (**Chapter 3**)
3. **Dvornik, A.**, Cacciato M., Kuijken K., Viola M., et al.,  
*A KiDS weak lensing analysis of assembly bias in GAMA galaxy groups*,  
2017, MNRAS, 468, 3251 (**Chapter 2**)

## Contributed:

1. Georgiou, C., Chisari, N. E., Fortuna, M. C., **et al.**,  
*GAMA+KiDS: Alignment of galaxies in galaxy groups and its dependence on galaxy scale*,  
2019, submitted to A&A
2. Kuijken, K., Heymans, C., **Dvornik, A.**, Hildebrandt, H., de Jong, J.T.A., Wright A.H., et al.,  
*The fourth data release of the Kilo-Degree Survey: ugri imaging and nine-band optical-IR photometry over 1000 square degrees*,  
2019, A&A, 625, A2
3. Petrillo, C. E., Tortora, C., Vernardos, G., Koopmans, L. V. E., Verdoes Kleijn, G., **et al.**,  
*LinKS: Discovering galaxy-scale strong lenses in the Kilo-Degree Survey using Convolutional Neural Networks*,  
2019, MNRAS, 484, 3879
4. Brouwer, M. M., Demchenko V., Harnois-Deraps J., **et al.**,  
*Studying galaxy troughs and ridges using Weak Gravitational Lensing with the Kilo-Degree Survey*, 2018, MNRAS, 481, 5189

5. Grootes, M. W., **Dvornik, A.**, et al, *Galaxy And Mass Assembly (GAMA): Gas Fuelling of Spiral Galaxies in the Local Universe II. – Direct Measurement of the Dependencies on Redshift and Host Halo Mass of Stellar Mass Growth in Central Disk Galaxies*,  
2018, MNRAS, 477, 1015
6. Brouwer, M. M., Visser, M. R., **Dvornik, A.**, et al.,  
*First test of Verlinde's theory of emergent gravity using weak gravitational lensing measurements*,  
2017, MNRAS, 466, 2547
7. Hildebrandt, H., Viola, M., Heymans, C., **et al.**,  
*KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing*,  
2017, MNRAS, 465, 1454
8. Brouwer, M. M., Cacciato, M., **Dvornik, A.**, et al.,  
*Dependence of GAMA galaxy halo masses on the cosmic web environment from 100 deg<sup>2</sup> of KiDS weak lensing data*,  
2016, MNRAS, 462, 4451

---

# Curriculum Vitae

---

I was born on the 13th of January 1990 in Maribor, Slovenia. I've spent my early childhood in the middle of the hilly parts of Slovenian Styria in Lovrenc na Pohorju and Ruše, not far from Maribor. There I started exploring nature and my surroundings with great interest. The interest in natural phenomena was further fuelled by my grandpa's workshop and my father's interests in everything connected with science and engineering. Including astronomy.

I started high school in Maribor in 2005 and there my interests in physics grew from day to day. I was always more interested in astronomy (as a more interesting part of physics) and this passion grew even more when I attended a couple of astronomy youth summer camps during those years. This culminated by enrolling to the Faculty of Mathematics and Physics at University of Ljubljana in 2009, starting my bachelor studies in physics with a focus in astronomy. The bachelor studies at the Faculty of Mathematics and Physics were quite rigorous and thorough, but because of that I know they gave me all the knowledge, experience and necessities to successfully tackle any physics and astronomy problem. I obtained a bachelor's degree in May 2013.

Due to the fact that there was no astronomy master study offered in Ljubljana, I have started looking for a programme abroad. More or less randomly I have picked Leiden as a place to apply to and got accepted. I started my masters in September of 2013. During the two years, I have first worked with Prof. Dr. Ignas Snellen and Dr. Jayne Birkby on secondary transits of exoplanets and then on modelling the stellar and gas contributions to the halo model and small scale galaxy-galaxy lensing signal with Prof. Dr. Henk Hoekstra and Dr. Marcello Cacciato, which was also my master's thesis. I have successfully defended it in August of 2015.

In 2015 I started as a PhD student under the supervision of Prof. Dr. Konrad Kuijken and Prof. Dr. Henk Hoekstra to work on Kilo Degree Survey and testing the gravity with it. As a member of the KiDS collaboration, I have contributed a halo model, helped with the massive undertaking of reducing the optical data from the VLT Survey Telescope hosting the KiDS, and actively helped with my modelling expertise to the other members of the collaboration. During my studies, I have participated in a winter school in Tonale, Italy and attended conferences at several international venues: Cairns, Australia; Santa Barbara, USA, Kingston, Canada; Leiden, The Netherlands. I have presented my work at different colloquia in Ljubljana, Toronto, Waterloo, Davies, Berkeley, Santa Cruz, Stanford and at regular KiDS collaboration meetings in Leiden and Edinburgh.

This autumn I will continue with my scientific work in the newly established German Centre for Cosmological Lensing headed by Prof. Dr. Catherine Heymans and Prof. Dr. Hendrik Hildebrandt in Bochum, Germany.



---

# Acknowledgements

---

After spending 6 years at the Leiden Observatory, it is almost impossible to account for all the people that contributed to the completion of this thesis in a page or two. First of all my thanks goes to the citizens of the Netherlands for funding the scientific research on such a high level and for starting the project I deeply respect – the EU. Without that, getting to this point in my life would be a significantly harder process, if not almost impossible. My thanks also go to the Leiden Observatory and all the people that made my last 4 (6, counting my master studies as well) years a smooth experience, be it for the administrative tasks, formal requirements, and technical support. All this and the excellent working conditions helped me to focus entirely on scientific research.

I thank all the members of the ever-growing ‘lensing group’, past and present, for all the constructive feedback during our progress meetings and all the insights regarding the recent literature we were debating together: Alessandro, Arun, Christos, Cristobal, Dong-Gang, Fabian, Jeroen, Jelte, Lammim, Maciej, Marcus, Marco, Margot, Maria Cristina, Martin, Merijn, Mohammadjavad, Omar, Patricia, Ricardo, Stijn, Valery, Yannick and Yvette. Marcello and Massimo, thank you for your invaluable insights, wisdom, and support during my first years of my PhD. Together with my actual supervisors, Koen and Henk, you have all provided tremendous knowledge from the interactions we had.

Furthermore, I would like to thank all the collaborators I got to work with during my PhD. Without you, the whole journey would be impossible. Thank you for the amazing data products, independent insights, ideas, and suggestions. Thank you, Catherine, Hendrik, Edo, Reiko, Alexandra, Angus, and all other KiDS and Astro-WISE people. Our meetings were all much fun, and I look forward working closely with you in the coming years.

I am grateful to have met David and Clément during my PhD for being there to complain about things and enjoy the Friday borrels together. For all the dinners and drinks either at Lemmy’s or at Bonte Koe (together with Fabian and Emanuele) and for the search of the proper ‘Fée Verte’ in Leiden. I hope we will be able to join in the life path in the near future and continue our discussions around a nice drink.

This brings me to Emanuele. Thank you for being there in all ups and downs I had during those 6 years in Leiden. Thank you for sticking around also after you left the observatory and for all the lunches at Mauro’s. Thank you, Tiago, for the amazing friendship that grew out of unsuspected office share that started your path as a postdoc in Leiden. We will all continue our ramblings about life, Universe and everything else.

I also say thank to all the people I have met at Leiden and shared good memories with: Aayush, Alex, Alberto, Andrej, Anna, Ann-Sophie, Arthur, Bartosz, Bas, Chris, Corentin, Dilovan, Eleonora, Eva, Felix, Fraser, Frits, Gabriela, Gleb, Hidido, Jit, Josha, Julia, Keira, Ken, Kim, Kirstin, Kirsty, Laura, Luke, Lýdia, Mantas, Merel, Michał, Michael, Mihael, Mieke, Nikki, Pablo, Patrick, Paweł, Pim, Ronniy, Sanjana, Sašo, Shourya, Stella, Tommaso, Valeriya, Vincenzo and all the people I might have missed (you all know who you are).

I am lucky that during my stay in Leiden I have managed to share an apartment with really awesome people. Thank you Iva and Łukasz for being awesome room-mates and even more awesome friends! Let that last!

I would like to also thank all the Lindy Hop dancers (and Shaggies) I have met in Leiden and the Netherlands. Without your shared passion and energy I would have long burned out. You all kept my sanity in the worst moments and forced me to balance my work and life. We also had a lot of fun! Thank you Andries, Annette, Astrid, Audrey, Burcu, Charlotte, Christian, Eveline, Fréd, Floor, Hieke, Iris, Karin, Karsten, Katarina, Kristoffer, Krisna, Laylan, Lysanne, Marcel, Martine, Pia, Rabia, Riemer, Tjitske, Vera, Wouter, Zhang, and Ron and Lotte for bringing Lindy to Leiden! Thank you to all the dancers I have ever shared a dance with! Last but not least, thank you, Monia, for your connection, energy, enthusiasm, dances, kindness, world views, keeping me awake, tea, and being around at the hardest moments! See you at the dance floor!

Posebna zahvala pa gre mojim najboljšim prijateljem. Hvala ker ste! Hvala Ana in Blaž, Filip in Filip, Mitja, Marko in Tjaša, za vse pogovore, piva, filme, namizne igre, piknike, sladice, kosila, večerje in druženja. Brez vas itak ne morem! Hvala pa seveda tudi bratu in sestri, Gregorju in Danaji! Hvala za ves smeh in obiske!

Na koncu pa bi se rad zahvalil staršem, Igorju in Evi. Brez vaju nič od zgoraj napisanega ne bi bilo mogoče. Hvala za vajino brezpogojno ljubezen in potrpežljivost. Tole je za vaju.