A&A 650, A113 (2021)
https://doi.org/10.1051/0004-6361/202040108

Astronomy & Astrophysics

The weak lensing radial acceleration relation: Constraining modified gravity and cold dark matter theories with KiDS-1000

Margot M. Brouwer[1,2], Kyle A. Oman[1,3], Edwin A. Valentijn[1], Maciej Bilicki[4], Catherine Heymans[5,6], Henk Hoekstra[7], Nicola R. Napolitano[8], Nivya Roy[8], Crescenzo Tortora[9], Angus H. Wright[6], Marika Asgari[5], Jan Luca van den Busch[6], Andrej Dvornik[6], Thomas Erben[10], Benjamin Giblin[5], Alister W. Graham[11], Hendrik Hildebrandt[6], Andrew M. Hopkins[12], Arun Kannawadi[13], Konrad Kuijken[7], Jochen Liske[14], HuanYuan Shan[15,16], Tilman Tröster[5], Erik Verlinde[2], and Manus Visser[17]

[1] Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
e-mail: margot.brouwer@gmail.com
[2] Institute for Theoretical Physics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[3] Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK
[4] Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668 Warsaw, Poland
[5] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[6] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany
[7] Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands
[8] School of Physics and Astronomy, Sun Yat-sen University, Guangzhou 519082, Zhuhai Campus, PR China
[9] INAF – Osservatorio Astronomico di Capodimonte, Salita Moiariello 16, Napoli 80131, Italy
[10] Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
[11] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia
[12] Australian Astronomical Optics, Macquarie University, 105 Delhi Road, North Ryde, NSW 2113, Australia
[13] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA
[14] Hamburger Sternwarte, University of Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany
[15] Shanghai Astronomical Observatory (SHAO), Nandan Road 80, Shanghai 200030, PR China
[16] University of the Chinese Academy of Sciences, Yuquanlu 19A, Beijing 100049, PR China
[17] Department of Theoretical Physics, University of Geneva, 24 quai Ernest-Ansermet, 1211 Genève 4, Switzerland

Received 10 December 2020 / Accepted 12 April 2021

ABSTRACT

We present measurements of the radial gravitational acceleration around isolated galaxies, comparing the expected gravitational acceleration given the baryonic matter ($g_{\rm bar}$) with the observed gravitational acceleration ($g_{\rm obs}$), using weak lensing measurements from the fourth data release of the Kilo-Degree Survey (KiDS-1000). These measurements extend the radial acceleration relation (RAR), traditionally measured using galaxy rotation curves, by 2 decades in $g_{\rm obs}$ into the low-acceleration regime beyond the outskirts of the observable galaxy. We compare our RAR measurements to the predictions of two modified gravity (MG) theories: modified Newtonian dynamics and Verlinde's emergent gravity (EG). We find that the measured relation between $g_{\rm obs}$ and $g_{\rm bar}$ agrees well with the MG predictions. In addition, we find a difference of at least 6σ between the RARs of early- and late-type galaxies (split by Sérsic index and $u-r$ colour) with the same stellar mass. Current MG theories involve a gravity modification that is independent of other galaxy properties, which would be unable to explain this behaviour, although the EG theory is still limited to spherically symmetric static mass models. The difference might be explained if only the early-type galaxies have significant ($M_{\rm gas} \approx M_\star$) circumgalactic gaseous haloes. The observed behaviour is also expected in Λ-cold dark matter (ΛCDM) models where the galaxy-to-halo mass relation depends on the galaxy formation history. We find that MICE, a ΛCDM simulation with hybrid halo occupation distribution modelling and abundance matching, reproduces the observed RAR but significantly differs from BAHAMAS, a hydrodynamical cosmological galaxy formation simulation. Our results are sensitive to the amount of circumgalactic gas; current observational constraints indicate that the resulting corrections are likely moderate. Measurements of the lensing RAR with future cosmological surveys (such as Euclid) will be able to further distinguish between MG and ΛCDM models if systematic uncertainties in the baryonic mass distribution around galaxies are reduced.

Key words. gravitational lensing: weak – methods: statistical – surveys – galaxies: halos – dark matter – cosmology: theory

1. Introduction

It has been known for almost a century that the outer regions of galaxies rotate faster than would be expected from Newtonian dynamics based on their luminous, or 'baryonic', mass (Kapteyn 1922; Oort 1932, 1940; Babcock 1939). This was also demonstrated by Gottesman et al. (1966) and Bosma (1981) through measurements of hydrogen profiles at radii beyond the optical discs of galaxies, and by Rubin (1983) through measurements of galactic rotation curves within the optical discs. The excess gravity implied by these measurements has generally been attributed to an unknown and invisible substance named dark matter (DM),

a term coined more than 40 years prior by Zwicky (1933) when he discovered the so-called missing mass problem through the dynamics of galaxies in clusters. More recently, new methods such as weak gravitational lensing (Hoekstra et al. 2004; Mandelbaum et al. 2006; Clowe et al. 2006; Heymans et al. 2013; von der Linden et al. 2014), baryon acoustic oscillations (Eisenstein et al. 2005; Blake et al. 2011), and the cosmic microwave background (CMB; de Bernardis et al. 2000; Spergel et al. 2003; Planck Collabration XVI 2014) have contributed unique evidence to the missing mass problem.

Among many others, these observations have contributed to the fact that cold dark matter[1] (CDM) has become a key ingredient of the current standard model of cosmology: the ΛCDM model. In this paradigm, CDM accounts for a fraction $\Omega_{\rm CDM} = 0.266$ of the critical density $\rho_{\rm crit} = 3H_0^2/8\pi G$ in the Universe, while baryonic matter only accounts for $\Omega_{\rm bar} = 0.049$ (Planck Collabration VI 2020). The cosmological constant Λ, which is necessary to explain the accelerated expansion of the Universe (Riess et al. 1998; Perlmutter et al. 1999) and is a special case of dark energy (DE), accounts for the remaining $\Omega_\Lambda = 0.685$ in our flat space-time (de Bernardis et al. 2000).

Although the ΛCDM model successfully describes the observations on a wide range of scales, no conclusive direct evidence for the existence of DM particles has been found so far (despite years of enormous effort; for an overview, see Bertone et al. 2005; Bertone & Tait 2018). Combined with other current open questions in physics, such as the elusive unification of general relativity (GR) with quantum mechanics and the mysterious nature of DE, this leaves room for alternative theories of gravity. Two modified gravity (MG) theories that do not require the existence of particle DM are modified Newtonian dynamics (MOND; Milgrom 1983) and the more recent theory of emergent gravity (EG; Verlinde 2017). In these theories all gravity is due to the baryonic matter (or, in the case of EG, the interaction between baryons and the entropy associated with DE). Hence, one of the main properties of these theories is that the mass discrepancy in galaxies correlates strongly with their baryonic mass distribution.

Such a correlation has indeed been observed, such as via the Tully–Fisher relation (Tully & Fisher 1977) between the luminosity of a spiral galaxy and its asymptotic rotation velocity (Pierce & Tully 1988; Bernstein et al. 1994). This relation was later generalised as the baryonic Tully–Fisher relation (McGaugh et al. 2000; McGaugh 2012) to include non-stellar forms of baryonic matter. Even earlier, astronomers had found a strong correlation between the observed rotation velocity as a function of galaxy radius $v_{\rm obs}(r)$ and the enclosed luminous mass $M_{\rm bar}(<r)$ (Sanders 1986, 1996; McGaugh 2004; Sanders & Noordermeer 2007; Wu & Kroupa 2015). Since $M_{\rm bar}(<r)$ corresponds to the expected gravitational acceleration $g_{\rm bar}(r)$ from baryonic matter, and the observed gravitational acceleration can be calculated through $g_{\rm obs}(r) = v_{\rm obs}^2(r)/r$, this relation has also been named the radial acceleration relation (RAR)[2].

McGaugh et al. (2016, hereafter M16) in particular measured the RAR with unprecedented accuracy, using the *Spitzer* Photometry and Accurate Rotation Curves (SPARC; Lelli et al. 2016) data of 153 late-type galaxies. Their results again showed a tight correlation between $g_{\rm obs}$ and $g_{\rm bar}$, which they could describe using a simple double power law (Eq. (4) in M16) that depends only on $g_{\rm bar}$ and one free parameter: the acceleration scale $g_\dagger$ where Newtonian gravity appears to break down. This rekindled the interest of scientists working on alternative theories of gravity (Lelli et al. 2017a,b; Burrage et al. 2017; Li et al. 2018; O'Brien et al. 2019), but also of those seeking an explanation of the RAR within the ΛCDM framework, employing correlations between the masses, sizes, and DM content of galaxies (Di Cintio & Lelli 2016; Keller & Wadsley 2017; Desmond 2017; Ludlow et al. 2017; Navarro et al. 2017; Tenneti et al. 2018).

Navarro et al. (2017, hereafter N17) used a range of simplifying assumptions based on galaxy observations and DM simulations in order to create an analytical galaxy model including the baryonic and halo components. With this model they reconstruct the RAR inside galaxy discs, in particular the value of a_0, the acceleration scale where the relation transitions from the baryon-dominated to the DM-dominated regime (which is equivalent to $g_\dagger$), and $a_{\rm min}$, the minimum acceleration probed by galaxy discs. Based on their results, they claim that the RAR can be explained within the ΛCDM framework at the accelerations probed by galaxy rotation curves (within the galaxy disc, i.e., $g_{\rm obs} > a_{\rm min}$). However, since their model relies on the fact that luminous kinematic tracers in galaxies only probe a limited radial range, N17 predicted that extending observations to radii beyond the disc (which correspond to lower gravitational accelerations) would lead to systematic deviations from the simple double power law proposed by M16. Although some progress has been made using globular clusters (Bílek et al. 2019a,b; Müller et al. 2021), using kinematic tracers to measure the RAR beyond the outskirts of visible galaxies remains difficult.

The goal of this work is to extend observations of the RAR to extremely low accelerations that cannot currently be detected through galaxy rotation curves or any other kinematic measurement. To this end, we use gravitational lensing: the perturbation of light inside a gravitational potential as described by relativistic theories such as GR. Both weak and strong gravitational lensing were used by Tian et al. (2020) to measure the RAR from observations of 20 galaxy clusters targeted by the CLASH survey. However, due to the high cluster masses, the accelerations probed by these measurements were of the same order as those measurable with galaxy rotation curves. In this work, we use the method of galaxy–galaxy lensing (GGL): the statistical measurement of the coherent image distortion (shear) of a field of background galaxies (sources) by the gravitational potential of a sample of individual foreground galaxies (lenses; for examples, see e.g., Brainerd et al. 1996; Fischer et al. 2000; Hoekstra et al. 2004; Mandelbaum et al. 2006; van Uitert et al. 2016). Using GGL we can measure the average (apparent) density distribution of isolated galaxies up to a radius of 3 Mpc, roughly 100 times larger than the radius of the luminous disc (~30 kpc). At our stellar mass scale of interest – $\log(M_\star/h_{70}^{-2}\,M_\odot) \approx 10.5$ – this radius corresponds to $g_{\rm bar} \approx 10^{-15}\,{\rm m\,s^{-2}}$, which is three orders of magnitude lower than the baryonic accelerations of the M16 rotation curves[3].

[1] DM particles that moved at non-relativistic speeds at the time of recombination, as favoured by measurements of the CMB (Planck Collabration XVI 2014) and the Lyman-α forest (Viel et al. 2013).

[2] Another closely related (though slightly different) relation is the mass-discrepancy acceleration relation, which shows the expected baryonic acceleration against the discrepancy between the baryonic and the observed mass: $M_{\rm obs}-M_{\rm bar}$ (see McGaugh 2004). Although measuring this relation requires the same data, we prefer the RAR because the two observables ($g_{\rm bar}$ and $g_{\rm obs}$) are uncorrelated.

[3] We note that this value of $g_{\rm bar}$ only takes into account the stellar and cold gas mass of the galaxy. In Sect. 4.3 we show that the contributions of additional hot gas, dust and 'missing baryons' could increase this value to $g_{\rm bar} \approx 10^{-14}\,{\rm m\,s^{-2}}$, which is still two orders of magnitude lower than the accelerations measurable with galaxy rotation curves.

Our main goal is to use the lensing RAR of isolated galaxies at lower accelerations (beyond the observable galaxy disc) to distinguish which of the aforementioned MG and ΛCDM models best describe this result. To achieve this, we first measure the total and baryonic density profiles of our galaxies through their GGL profiles and luminosities. These measurements will be performed using 1006 deg^2 of weak lensing data from the Kilo-Degree Survey (KiDS-1000; de Jong et al. 2013; Kuijken et al. 2019), and nine-band photometric data from KiDS and the VISTA Kilo-Degree Infrared Galaxy Survey (VIKING, Edge et al. 2013). We then translate these measurements into the observed and baryonic radial accelerations, $g_{\rm obs}$ and $g_{\rm bar}$. Finally, we compare the resulting RAR to predictions from different MG theories (MOND and EG) and ΛCDM. To test the MG theories, we need to make the assumption that the deflection of light by gravitational potentials (as described in GR) holds in these modified theories, which we motivate in the relevant sections. This work can be seen as an extension of Brouwer et al. (2017), where we tested the predictions of EG using KiDS GGL on foreground galaxies from 180 deg^2 of the Galaxy and Mass Assembly (GAMA) survey. Instead of GAMA, we now use a selection of ∼1 million foreground galaxies from KiDS-1000 to achieve a fivefold increase in survey area.

The ΛCDM predictions will not only be provided by the N17 analytical model, but also by mock galaxy catalogues based on two different DM simulations. One is the Marenostrum Institut de Ciències de l'Espai (MICE) Galaxy and Halo Light-cone catalogue (Carretero et al. 2015; Hoffmann et al. 2015), which is based on the MICE Grand Challenge lightcone simulation (Fosalba et al. 2015a,b; Crocce et al. 2015). The other mock galaxy catalogue is based on a suite of large-volume cosmological hydrodynamical simulations, called the BAryons and HAloes of MAssive Systems (BAHAMAS) project (McCarthy et al. 2017).

Having ∼1 million foreground galaxies at our disposal allows us to select specific galaxy samples, designed to optimally test the predictions from the aforementioned MG and ΛCDM models. Particularly, we note that the analytical models (MOND, EG and N17) mostly focus on the description of individual, isolated galaxies. In order to test them, we select a sample of galaxies whose GGL profiles are minimally affected by neighbouring galaxies (e.g., satellites) within the radius of our measurement. In contrast, the predictions from simulations can be tested with both isolated and non-isolated galaxy samples.

In addition, our sample of ∼350 000 isolated lens galaxies allows us to analyse the RAR as a function of colour, Sérsic index and stellar mass. Because MG and ΛCDM give different predictions regarding the dependence of the RAR on these observables, this allows us to better distinguish between the different models. Specifically: according to the MOND and EG theories the relation between $g_{\rm bar}$ and $g_{\rm obs}$ should remain fixed in the regime beyond the baryon-dominated galaxy disc, and hence be independent of galaxy observables. Within the ΛCDM paradigm, the relation between $g_{\rm bar}$ and $g_{\rm obs}$ is related to the stellar-to-halo-mass relation (SHMR) that is not necessarily constant as a function of galaxy stellar mass or other observables.

Our paper is structured as follows: in Sect. 2 we describe the methodology behind the GGL measurements and their conversion into the RAR, in addition to the theoretical predictions to which we compare our observations: MOND, EG and the N17 analytical DM model. In Sect. 3 we introduce the KiDS-1000 and GAMA galaxy surveys used to perform both the GGL and stellar mass measurements. Section 4 describes the MICE and BAHAMAS simulations and mock galaxy catalogues to which we compare our results. In Sect. 5 we present our lensing RAR measurements and compare them to the different models, first using all isolated galaxies and then separating the galaxies by different observables. Section 6 contains the discussion and conclusion. In Appendix A we validate our isolated galaxy selection, and Appendix B contains a description of the piecewise-power-law method of translating the lensing measurement into $g_{\rm obs}$. Finally, Appendix C shows the comparison of the N17 analytical DM model with our lensing RAR.

Throughout this work we adopt the WMAP 9-year (Hinshaw et al. 2013) cosmological parameters: $\Omega_{\rm m} = 0.2793$, $\Omega_{\rm b} = 0.0463$, $\Omega_\Lambda = 0.7207$, $\sigma_8 = 0.821$ and $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$, which were used as the basis of the BAHAMAS simulation. When analysing the MICE simulations we use the cosmological parameters used in creating MICE, which are: $\Omega_{\rm m} = 0.25$, $\sigma_8 = 0.8$, $\Omega_\Lambda = 0.75$, and $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$. Throughout the paper we use the reduced Hubble constant $h_{70} = H_0/(70\,{\rm km\,s^{-1}\,Mpc^{-1}})$. Due to the relatively low redshift of our lens galaxies ($z \sim 0.2$) the effect of differences in the cosmological parameters on our results is small.

2. Theory

2.1. Mass measurements with weak gravitational lensing

To estimate the gravitational acceleration around galaxies we used GGL: the measurement of the coherent image distortion of a field of background galaxies (sources) by the gravitational potential of a sample of foreground galaxies (lenses). Because the individual image distortions are very small (only ∼1% compared to the galaxy's unknown original shape), this method can only be performed statistically for a large sample of sources. We averaged their projected ellipticity component tangential to the direction of the lens galaxy, $\epsilon_{\rm t}$, which is the sum of the intrinsic tangential ellipticity component $\epsilon_{\rm t}^{\rm int}$ and the tangential shear $\gamma_{\rm t}$ caused by weak lensing. Assuming no preferential alignment in the intrinsic galaxy shapes ($\langle\epsilon_{\rm t}^{\rm int}\rangle = 0$), the average $\langle\epsilon_{\rm t}\rangle$ is an estimator for $\gamma_{\rm t}$. By measuring this averaged quantity in circular annuli around the lens centre, we obtained the tangential shear profile $\gamma_{\rm t}(R)$ as a function of projected radius R. Because our final goal is to compute the observed gravitational acceleration $g_{\rm obs}$ as a function of that expected from baryonic matter $g_{\rm bar}$, we chose our R-bins such that they corresponded to 15 logarithmic bins between $1\times10^{-15} < g_{\rm bar} < 5\times10^{-12}\,{\rm m\,s^{-2}}$. For each individual lens the calculation of these $g_{\rm bar}$-bins was based on the baryonic mass of the galaxy $M_{\rm gal}$ (see Sect. 3.3). In real space this binning approximately corresponds to the distance range used in Brouwer et al. (2017): $0.03 < R < 3\,h_{70}^{-1}\,{\rm Mpc}$.

The lensing shear profile can be related to the physical excess surface density (ESD, denoted ΔΣ) profile through the critical surface density $\Sigma_{\rm crit}$:

$$\Delta\Sigma(R) = \Sigma_{\rm crit}\gamma_{\rm t}(R) = \langle\Sigma\rangle(<R) - \Sigma(R), \quad (1)$$

which is the surface density $\Sigma(R)$ at projected radius R, subtracted from the average surface density $\langle\Sigma\rangle(<R)$ within R. See Sect. 3.1 for more information on how this is computed.

The error values on the ESD profile were estimated by the square-root of the diagonal of the analytical covariance matrix, which is described in Sect. 3.4 of Viola et al. (2015). The full covariance matrix was calculated based on the contribution of each individual source to the ESD profile, and incorporates the correlation between sources that contribute to the ESD in multiple bins, both in projected distance R and in galaxy observable.

2.2. The radial acceleration relation (RAR)

After measuring the lensing profile around a galaxy sample, the next step is to convert it into the corresponding RAR. We started from the ESD as a function of projected radius $\Delta\Sigma(R)$ and the measured stellar masses of the lens galaxies $M_\star$, aiming to arrive at their observed radial acceleration $g_{\rm obs}$ as a function of their expected baryonic radial acceleration $g_{\rm bar}$. The latter can be calculated using Newton's law of universal gravitation:

$$g(r) = \frac{G\,M(<r)}{r^2}, \tag{2}$$

which defines the radial acceleration g in terms of the gravitational constant G and the enclosed mass $M(<r)$ within spherical radius r. Assuming spherical symmetry here is reasonable, given that for lensing measurements thousands of galaxies are stacked under many different angles to create one average halo profile.

The calculation of $g_{\rm bar}$ requires the enclosed baryonic mass $M_{\rm bar}(<r)$ of all galaxies. We discuss our construction of $M_{\rm bar}(<r)$ in Sect. 3.3. The calculation of $g_{\rm obs}$ requires the enclosed observed mass $M_{\rm obs}(<r)$ of the galaxy sample, which we obtained through the conversion of our observed ESD profile $\Delta\Sigma(R)$.

When calculating $g_{\rm obs}$ we started from our ESD profile measurement, which consists of the value $\Delta\Sigma(R)$ measured in a set of radial bins. At our measurement radii ($R > 30\,h_{70}^{-1}$ kpc) the ESD is dominated by the excess gravity, which means the contribution from baryonic matter can be neglected. We adopted the simple assumption that our observed density profile $\rho_{\rm obs}(r)$ is roughly described by a Singular Isothermal Sphere (SIS) model:

$$\rho_{\rm SIS}(r) = \frac{\sigma^2}{2G\pi r^2}. \tag{3}$$

The SIS is generally considered to be the simplest parametrisation of the spatial distribution of matter in an astronomical system (such as galaxies, clusters, etc.). If interpreted in a ΛCDM context, the SIS implies the assumption that the DM particles have a Gaussian velocity distribution analogous to an ideal gas that is confined by their combined spherically symmetric gravitational potential, where σ is the total velocity dispersion of the particles. In a MG context, however, the SIS profile can be considered to represent a simple r^{-2} density profile as predicted by MOND and EG in the low-acceleration regime outside a baryonic mass distribution, with σ as a normalisation constant. The ESD derived from the SIS profile is:

$$\Delta\Sigma_{\rm SIS}(R) = \frac{\sigma^2}{2GR}. \tag{4}$$

From Brouwer et al. (2017) we know that, despite its simple form, it provides a good approximation of the GGL measurements around isolated galaxies. The SIS profile is therefore well-suited to analytically model the total enclosed mass distribution of our lenses, which can then be derived as follows:

$$M_{\rm SIS}(<r) = 4\pi \int_0^r \rho_{\rm SIS}(r')r'^2{\rm d}r' = \frac{2\sigma^2 r}{G}. \tag{5}$$

Now, for each individual observed ESD value $\Delta\Sigma_{{\rm obs},m}$ at certain projected radius R_m, we assumed that the density distribution within R_m is described by an SIS profile with σ normalised such that $\Delta\Sigma_{\rm SIS}(R_m) = \Delta\Sigma_{{\rm obs},m}$. Under this approximation, we combined Eqs. (4) and (5) to give a relation between the lensing measurement $\Delta\Sigma$ and the deprojected, spherically enclosed mass $M_{\rm obs}$:

$$M_{\rm obs}(<r) = 4\Delta\Sigma_{\rm obs}(r)\,r^2. \tag{6}$$

Through Eq. (2), this results in a very simple expression for the observed gravitational acceleration:

$$g_{\rm obs}(r) = \frac{G\,[4\Delta\Sigma_{\rm obs}(r)\,r^2]}{r^2} = 4G\Delta\Sigma_{\rm obs}(r). \tag{7}$$

Throughout this work, we have used the SIS approximation to convert the ESD into $g_{\rm obs}$. In Sect. 4.4 we validate this approach by comparing it to a more elaborate method and testing both on the BAHAMAS simulation.

2.3. The RAR with modified Newtonian dynamics

With his theory, MOND, Milgrom (1983) postulated that the missing mass problem in galaxies is not caused by an undiscovered fundamental particle, but that instead our current gravitational theory should be revised. Since MOND is a non-relativistic theory, performing GGL measurements to test it requires the assumption that light is curved by a MONDian gravitational potential in the same way as in GR. This assumption is justified since Milgrom (2013, while testing the MOND paradigm using GGL data from the Canada-France-Hawaii Telescope Lensing survey), states that non-relativistic MOND is a limit of relativistic versions that predict that gravitational potentials determine lensing in the same way as Newtonian potentials in GR. For this reason GGL surveys can be used as valuable tools to test MOND and similar MG theories, as was done for instance by Tian et al. (2009) using Sloan Digital Sky Survey (SDSS) and Red-sequence Cluster Survey data.

MOND's basic premise is that one can adjust Newton's second law of motion ($F = ma$) by inserting a general function $\mu(a/a_0)$, which only comes into play when the acceleration a of a test mass m is much smaller than a critical acceleration scale a_0. This function predicts the observed flat rotation curves in the outskirts of galaxies, while still reproducing the Newtonian behaviour of the inner disc. In short, the force F becomes:

$$F(a) = m\,\mu\left(\frac{a}{a_0}\right)a, \quad \mu(x \gg 1) \approx 1,\ \mu(x \ll 1) \approx x. \tag{8}$$

This implies that $a \gg a_0$ represents the Newtonian regime where $F_{\rm N} = m\,a_{\rm N}$ as expected, while $a \ll a_0$ represents the 'deep-MOND' regime where $F_{\rm MOND} = m\,a_{\rm MOND}^2/a_0$. In a circular orbit, this is reflected in the deep-MOND gravitational acceleration $g_{\rm MOND} \equiv a_{\rm MOND}$ as follows:

$$F_{\rm MOND} = m\frac{a_{\rm MOND}^2}{a_0} = \frac{G\,Mm}{r^2} \rightarrow g_{\rm MOND} = \sqrt{a_0\frac{GM}{r^2}}. \tag{9}$$

This can be written in terms of the expected baryonic acceleration $g_{\rm bar} = GM/r^2$ as follows:

$$g_{\rm MOND}(g_{\rm bar}) = \sqrt{a_0\,g_{\rm bar}}. \tag{10}$$

This demonstrates that MOND predicts a very simple relation for the RAR: $g_{\rm obs} = g_{\rm bar}$ in the Newtonian regime ($g_{\rm obs} \gg a_0$) and Eq. (9) in the deep-MOND regime ($g_{\rm obs} \ll a_0$). However, since $\mu(a/a_0)$, also known as the interpolating function, is not specified by Milgrom (1983), there is no specific constraint on the behaviour of this relation in between the two regimes. In the

work of Milgrom & Sanders (2008), several families of interpolation functions are discussed. Selecting the third family (given by their Eq. (13)) with constant parameter $\alpha = 1/2$, provides the function that M16 later used to fit to their measurement of the RAR using rotation curves of 153 galaxies. This relation can be written as:

$$g_{\rm obs}(g_{\rm bar}) = \frac{g_{\rm bar}}{1 - {\rm e}^{-\sqrt{g_{\rm bar}/a_0}}}, \tag{11}$$

where $a_0 \equiv g_\dagger$ corresponds to the fitting parameter constrained by M16 to be $g_\dagger = 1.20{\pm}0.26{\times}10^{-10}\,{\rm m\,s^{-2}}$. Since Eq. (11) (equal to Eq. (4) in M16) is also considered a viable version of the MOND interpolation function by Milgrom & Sanders (2008), we will consider it the baseline prediction of MOND in this work. As the baseline value of a_0, we will likewise use the value of $g_\dagger$ measured by M16 since it exactly corresponds to the value of $a_0 = 1.2 \times 10^{-10}\,{\rm m\,s^{-2}}$ considered canonical in MOND since its first measurement by Begeman et al. (1991), using the rotation curves of 10 galaxies.

One of the main characteristics of the MOND paradigm, is that it gives a direct and fixed prediction for the total acceleration based only on the system's baryonic mass, given by Eq. (11). The main exception to this rule is the possible influence by neighbouring mass distributions through the external field effect (EFE), predicted by Milgrom (1983) and studied analytically, observationally and in simulations by Banik & Zhao (2018), Banik et al. (2020), Chae et al. (2020). Since we explicitly selected isolated galaxies in this work (see Appendix A), this effect is minimised as much as possible. However, since total isolation cannot be guaranteed, a small EFE might remain. In order to describe this effect, we used Eq. (6) from Chae et al. (2020):

$$g_{\rm MOND}(g_{\rm bar}) = \nu_e(z)\, g_{\rm bar}, \tag{12}$$

with:

$$\nu_e(z) = \frac{1}{2} - \frac{A_e}{z} + \sqrt{\left(\frac{1}{2} - \frac{A_e}{z}\right)^2 + \frac{B_e}{z}}. \tag{13}$$

Here $z \equiv g_{\rm bar}/g_\dagger$, $A_e \equiv e(1 + e/2)/(1 + e)$, and $B_e \equiv (1 + e)$. The strength of the EFE is parametrised through: $e = g_{\rm ext}/g_\dagger$, determined by the external gravitational acceleration $g_{\rm ext}$. Although the interpolation functions differ, the result of Eq. (13) corresponds almost exactly to the M16 fitting function given in Eq. (11) in the limit $e = 0$ (no EFE). Positive values of e result in reduced values of the predicted $g_{\rm obs}$ at very low accelerations (see Fig. 4 in Sect. 5.2, and Fig. 1 of Chae et al. 2020). It should be noted that this fitting function represents an idealised model and could be subject to deviations in real, complex, 3D galaxies.

2.4. The RAR with emergent gravity

The work of Verlinde (2017, hereafter V17), which is embedded in the framework of string theory and holography, shares the view that the missing mass problem is to be solved through a revision of our current gravitational theory. Building on the ideas from Jacobson (1995, 2016), Padmanabhan (2010), Verlinde (2011), Faulkner et al. (2014), V17 abandons the notion of gravity as a fundamental force. Instead, it emerges from an underlying microscopic description of space-time, in which the notion of gravity has no a priori meaning.

V17 shows that constructing an EG theory in a universe with a negative cosmological constant ('anti-de Sitter') allows for the re-derivation of Einstein's laws of GR. A distinguishing feature of V17 is that it attempts to describe a universe with a positive cosmological constant ('de Sitter'), that is, one that is filled with a DE component. This results in a new volume law for gravitational entropy caused by DE, in addition to the area law normally used to retrieve Einsteinian gravity. According to V17, energy that is concentrated in the form of a baryonic mass distribution causes an elastic response in the entropy of the surrounding DE. This results in an additional gravitational component at scales set by the Hubble acceleration scale $a_0 = cH_0/6$. Here c is the speed of light, and H_0 is the current Hubble constant that measures the Universe's expansion velocity.

Because this extra gravitational component aims to explain the effects usually attributed to DM, it is conveniently expressed as an apparent dark matter (ADM) distribution:

$$M^2_{\rm ADM}(r) = \frac{cH_0 r^2}{6G} \frac{{\rm d}\,[M_{\rm bar}(r) r]}{{\rm d}r}. \tag{14}$$

Thus the ADM distribution is completely defined by the baryonic mass distribution $M_{\rm bar}(r)$ as a function of the spherical radius r, and a set of known physical constants.

Since we measured the ESD profiles of galaxies at projected radial distances $R > 30\,h_{70}^{-1}$ kpc, we can follow Brouwer et al. (2017) in assuming that their baryonic component is equal to the stars+cold gas mass enclosed within the minimal measurement radius (for further justification of this assumption, see Sect. 4.3). This is equivalent to describing the galaxy as a point mass $M_{\rm bar}$, which allows us to simplify Eq. (14) to:

$$M_{\rm ADM}(r) = \sqrt{\frac{cH_0\, M_{\rm bar}}{6\,G}}\, r. \tag{15}$$

Now the total enclosed mass $M_{\rm EG}(r) = M_{\rm bar} + M_{\rm ADM}(r)$ can be used to calculate the gravitational acceleration $g_{\rm EG}(r)$ predicted by EG, as follows:

$$g_{\rm EG}(r) = \frac{GM_{\rm EG}(r)}{r^2} = \frac{GM_{\rm bar}}{r^2} + \sqrt{\frac{cH_0}{6}}\, \frac{\sqrt{GM_{\rm bar}}}{r}. \tag{16}$$

In terms of the expected baryonic acceleration $g_{\rm bar}(r) = GM_{\rm bar}/r^2$, this simplifies even further to:

$$g_{\rm EG}(g_{\rm bar}) = g_{\rm bar} + \sqrt{\frac{cH_0}{6}}\, \sqrt{g_{\rm bar}}. \tag{17}$$

We emphasise that Eq. (14) is only a macroscopic approximation of the underlying microscopic phenomena described in V17, and is thus only valid for static, spherically symmetric and isolated baryonic mass distributions. For this reason, we selected only the most isolated galaxies from our sample (see Appendix A), such that our GGL measurements are not unduly influenced by neighbouring galaxies. Furthermore, the current EG theory is only valid in the acceleration range $g_{\rm bar} < a_0$, often called the deep-MOND regime. Therefore, the prediction of Eq. (17) should be taken with a grain of salt for accelerations $g_{\rm bar} > 1.2 \times 10^{-10}\,{\rm m\,s^{-2}}$. This will not affect our analysis since weak lensing takes place in the weak gravity regime. In addition, cosmological evolution of the H_0 parameter is not yet implemented in the theory, restricting its validity to galaxies with relatively low redshifts. However, we calculated that at our mean lens redshift, $\langle z \rangle \sim 0.2$, using an evolving $H(z)$ would result in only a ~5% difference in our ESD measurements, based on the background cosmology used in this work.

In order to test EG using the standard GGL methodology, we needed to assume that the deflection of photons by a gravitational potential in this alternative theory corresponds to that

in GR. This assumption is justified because, in EG's original (anti-de Sitter) form, Einstein's laws emerge from its underlying description of space-time. The additional gravitational force described by ADM does not affect this underlying theory, which is an effective description of GR. Therefore, we assumed that the gravitational potential of an ADM distribution produces the same lensing shear as an equivalent distribution of actual matter.

2.5. The RAR in ΛCDM

To help guide an intuitive interpretation of the lensing RAR within the framework of the ΛCDM theory, we made use of the simple model of N17, which combines a basic model of galactic structure and scaling relations to predict the RAR. We refer to N17 for a full description, but give a summary here. A galaxy of a given stellar (or baryonic – there is no distinction in this model) mass occupies a DM halo of a mass fixed by the abundance matching relation of Behroozi et al. (2013). The dark halo concentration is fixed to the cosmological mean for haloes of that mass (Ludlow et al. 2014). The baryonic disc follows an exponential surface density profile with a half-mass size fixed to 0.2× the scale radius of the dark halo. This model is sufficient to specify the cumulative mass profile of both the baryonic and dark components of the model galaxy; calculating $g_{\rm obs}$ and $g_{\rm bar}$ is then straightforward. However, since the N17 model is merely a simple analytical description, our main ΛCDM test utilised more elaborate numerical simulations (see Sect. 4).

3. Data

3.1. The Kilo-Degree Survey (KiDS)

We measured the gravitational potential around a sample of foreground galaxies (lenses), by measuring the image distortion (shear) of a field of background galaxies (sources). These sources were observed using OmegaCAM (Kuijken 2011): a 268-million pixel CCD mosaic camera mounted on the Very Large Telescope (VLT) Survey Telescope (Capaccioli & Schipani 2011). Over the past ten years these instruments have performed KiDS, a photometric survey in the *ugri* bands, which was especially designed to perform weak lensing measurements (de Jong et al. 2013).

GGL studies with KiDS have hitherto been performed in combination with the spectroscopic GAMA survey (see Sect. 3.2), with the KiDS survey covering 180 deg^2 of the GAMA area. Although the final KiDS survey will span 1350 deg^2 on the sky, the current state-of-the-art is the 4th Data Release (KiDS-1000; Kuijken et al. 2019) containing observations from 1006 deg^2 survey tiles. We therefore used a photometrically selected 'KiDS-bright' sample of lens galaxies from the full KiDS-1000 release, as described in Sect. 3.3. The measurement and calibration of the source shapes and photometric redshifts are described in Kuijken et al. (2019), Giblin et al. (2021), and Hildebrandt et al. (2021).

The measurements of the galaxy shapes are based on the *r*-band data since this filter was used during the darkest time (moon distance >90 deg) and with the best atmospheric seeing conditions (<0.8 arcsec). The *r*-band observations were co-added using the THELI pipeline (Erben et al. 2013). From these images the galaxy positions were detected through the SExtractor algorithm (Bertin & Arnouts 1996). After detection, the shapes of the galaxies were measured using the *lens*fit pipeline (Miller et al. 2007, 2013), which includes a self-calibration algorithm based on Fenech Conti et al. (2017) that was validated in Kannawadi et al. (2019). Each shape is accompanied by a *lens*fit weight $w_{\rm s}$, which was used as an estimate of the precision of the ellipticity measurement.

For the purpose of creating the photometric redshift and stellar mass estimates, 9 bands were observed in total. The *ugri* bands were observed by KiDS, while the VIKING survey (Edge et al. 2013) performed on the VISTA telescope adds the $ZYJHK_s$ bands. All KiDS bands were reduced and co-added using the Astro-WISE pipeline (AW; McFarland et al. 2013). The galaxy colours, which form the basis of the photometric redshift measurements, were measured from these images using the Gaussian Aperture and PSF pipeline (GAaP; Kuijken 2008; Kuijken et al. 2015).

The addition of the lower frequency VISTA data allowed us to extend the redshift estimates out to $0.1 < z_{\rm B} < 1.2$, where $z_{\rm B}$ is the best-fit photometric redshift of the sources (Benítez 2000; Hildebrandt et al. 2012). However, when performing our lensing measurements (see Sect. 2.1) we used the total redshift probability distribution function $n(z_{\rm s})$ of the full source population. This $n(z_{\rm s})$ was calculated using a direct calibration method (see Hildebrandt et al. 2017 for details), and circumvents the inherent bias related to photometric redshift estimates of individual sources.

We note that this is a different redshift calibration method than that used by the KiDS-1000 cosmology analyses (Asgari et al. 2021; Heymans et al. 2021; Tröster et al. 2021), who used a self-organising map to remove (primarily high-redshift) sources whose redshifts could not be accurately calibrated due to incompleteness in the spectroscopic sample (Wright et al. 2020; Hildebrandt et al. 2021). Following Robertson et al. (in prep.) we prioritised precision by analysing the full KiDS-1000 source sample (calibrated using the direct calibration method) since percent-level biases in the mean source redshifts do not significantly impact our analysis.

For the lens redshifts $z_{\rm l}$, we used the ANNz2 (Artificial Neural Network) machine-learning redshifts of the KiDS foreground galaxy sample (KiDS-bright; see Sect. 3.3). We implemented the contribution of $z_{\rm l}$ by integrating over the individual redshift probability distributions $p(z_{\rm l})$ of each lens. This $p(z_{\rm l})$ is defined by a normal distribution centred at the lens' $z_{\rm ANN}$ redshift, with a standard deviation: $\sigma_z/(1+z) = 0.02$ (which is equal to the standard deviation of the KiDS-bright redshifts compared to their matched spectroscopic GAMA redshifts). For the source redshifts $z_{\rm s}$ we followed the method used in Dvornik et al. (2018), integrating over the part of the redshift probability distribution $n(z_{\rm s})$ where $z_{\rm s} > z_{\rm l}$. In addition, sources only contribute their shear to the lensing signal when $z_{\rm B} + \Delta z > z_{\rm l}$ – when the sum of their best-fit photometric redshift $z_{\rm B}$ and the redshift buffer $\Delta z = 0.2$ is greater than the lens redshift. Hence, when performing the lensing measurement in Sect. 2.1 the critical surface density[4] (the conversion factor between $\gamma_{\rm t}$ and $\Delta\Sigma$, whose inverse is also called the lensing efficiency) was calculated as follows:

$$\Sigma_{\rm crit}^{-1} = \frac{4\pi G}{c^2} \int_0^\infty D(z_{\rm l}) \left(\int_{z_{\rm l}}^\infty \frac{D(z_{\rm l}, z_{\rm s})}{D(z_{\rm s})} n(z_{\rm s})\, {\rm d}z_{\rm s} \right) p(z_{\rm l})\, {\rm d}z_{\rm l}. \quad (18)$$

Here $D(z_{\rm l})$ and $D(z_{\rm s})$ are the angular diameter distances to the lens and the source respectively, and $D(z_{\rm l}, z_{\rm s})$ the distance

[4] As derived in Appendix C of Dvornik et al. (2018), there are two possible definitions of $\Sigma_{\rm crit}$: proper and comoving. In this work we used the proper $\Sigma_{\rm crit}$, and we compute $\Delta\Sigma(R)$ as a function of proper transverse separation R. This choice is reasonable because, within a $3\,h_{70}^{-1}$ Mpc range, the measured ESD profiles are expected to be approximately stationary in proper coordinates.

between them. The constant multiplication factor is defined by Newton's gravitational constant G and the speed of light c.

The ESD profile was averaged (or 'stacked') for large samples of lenses to increase the signal-to-noise ratio (S/N) of the lensing signal. We defined a lensing weight $W_{\rm ls}$ that depends on both the *lens*fit weight $w_{\rm s}$ and the lensing efficiency $\Sigma_{\rm crit}^{-1}$:

$$W_{\rm ls} = w_{\rm s} \left(\Sigma_{\rm crit,ls}^{-1}\right)^2, \tag{19}$$

and used it to optimally sum the measurements from all lens-source pairs into the average ESD:

$$\Delta\Sigma = \frac{1}{1+\mu} \frac{\sum_{\rm ls} W_{\rm ls}\, \epsilon_{\rm t,ls}\, \Sigma_{\rm crit,ls}}{\sum_{\rm ls} W_{\rm ls}}. \tag{20}$$

Here the factor $(1+\mu)$ calibrates the shear estimates Fenech Conti et al. (2017), Kannawadi et al. (2019). Extending the method of Dvornik et al. (2017) to the higher KiDS-1000 redshifts, μ denotes the mean multiplicative calibration correction calculated in 11 linear redshift bins between $0.1 < z_{\rm B} < 1.2$ from the individual source calibration values m:

$$\mu = \frac{\sum_{\rm s} w_{\rm s} m_{\rm s}}{\sum_{\rm s} w_{\rm s}}. \tag{21}$$

The value of this correction is $\mu \approx 0.014$, independent of the projected distance from the lens.

We also corrected our lensing signal for sample variance on large scales by subtracting the ESD profile measured around ∼5 million uniform random coordinates, 50 times the size of our total KiDS-bright sample. These random coordinates mimic the exact footprint of KiDS, excluding the areas masked by the 'nine-band no AW-r-band' mask that we applied to the KiDS-bright lenses (see Sect. 3.3). In order to create random redshift values that mimic the true distribution, we created a histogram of the KiDS-bright redshifts divided into 80 linear bins between $0.1 < z_{\rm ANN} < 0.5$. In each bin, we created random redshift values equal to the number of real lenses in that bin. Because of the large contiguous area of KiDS-1000, we found that the random ESD profile is very small at all projected radii R, with a mean absolute value of only $1.85 \pm 0.75\%$ of the lensing signal of the full sample of isolated KiDS-bright galaxies.

3.2. The Galaxy and Mass Assembly (GAMA) survey

Although the most contraining RAR measurements below were performed using exclusively KiDS-1000 data, the smaller set of foreground galaxies observed by the spectroscopic GAMA survey (Driver et al. 2011) functions both as a model and validation sample for the KiDS foreground galaxies. The survey was performed by the Anglo-Australian Telescope with the AAOmega spectrograph, and targeted more than 238 000 galaxies selected from the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009). For this study we used GAMA II observations (Liske et al. 2015) from three equatorial regions (G09, G12, and G15) containing more than 180 000 galaxies. These regions span a total area of ∼180 deg^2 on the sky, completely overlapping with KiDS.

GAMA has a redshift range of $0 < z < 0.5$, with a mean redshift of $\langle z \rangle = 0.22$. The survey has a redshift completeness of 98.5% down to Petrosian r-band magnitude $m_{r,\rm Petro} = 19.8$ mag. We limited our GAMA foreground sample to galaxies with the recommended redshift quality: $n_{\rm Q} \geq 3$. Despite being a smaller survey, GAMA's accurate spectroscopic redshifts were highly advantageous when measuring the lensing profiles of galaxies (see Sect. 2.1). The GAMA redshifts were used to train the photometric machine-learning (ML) redshifts of our larger sample of KiDS foreground galaxies (see Sect. 3.3). Also, in combination with its high redshift completeness, GAMA allows for a more accurate selection of isolated galaxies. We therefore checked that the results from the KiDS-only measurements are consistent with those from KiDS-GAMA.

To measure the RAR with KiDS-GAMA, we need individual stellar masses $M_\star$ for each GAMA galaxy. We used the Taylor et al. (2011) stellar masses, which are calculated from *ugrizZY* spectral energy distributions[5] measured by SDSS and VIKING by fitting them with Bruzual & Charlot (2003) Stellar Population Synthesis (SPS) models, using the Initial Mass Function (IMF) of Chabrier (2003). Following the procedure described by Taylor et al. (2011), we accounted for flux falling outside the automatically selected aperture using the 'flux-scale' correction.

3.3. Selecting isolated lens galaxies with accurate redshifts and stellar masses

Because of its accurate spectroscopic redshifts, the GAMA lenses would be an ideal sample for the selection of isolated galaxies and the measurement of accurate stellar masses (as was done in Brouwer et al. 2017). However, since the current KiDS survey area is >5 times larger than that of GAMA, we selected a KiDS-bright sample of foreground galaxies from KiDS-1000 that resembles the GAMA survey. We then used the GAMA redshifts as a training sample to compute neural-net redshifts for the KiDS-bright lenses (see e.g., Bilicki et al. 2018), from which accurate stellar masses could subsequently be derived. The details of the specific sample used in this work are provided in Bilicki et al. (2021). Here we give an overview relevant for this paper.

To mimic the magnitude limit of GAMA ($m_{r,\rm Petro} < 19.8$ mag), we applied a similar cut to the (much deeper) KiDS survey. Because the KiDS catalogue does not contain Petrosian magnitudes we used the Kron-like elliptical aperture r-band magnitudes from SExtractor, calibrated for r-band extinction and zero-point offset[6], which have a very similar magnitude distribution. Through matching the KiDS and GAMA galaxies and seeking the best trade-off between completeness and purity, we decided to limit our KiDS-bright sample to $m_{r,\rm auto} < 20.0$. In addition we removed KiDS galaxies with a photometric redshift $z > 0.5$, where GAMA becomes very incomplete.

To remove stars from our galaxy sample, we applied a cut based on galaxy morphology, nine-band photometry and the SExtractor star-galaxy classifier[7]. Through applying the IMAFLAGS_ISO = 0 flag, we also removed galaxies that are affected by readout and diffraction spikes, saturation cores, bad pixels, or by primary, secondary or tertiary haloes of bright stars[8]. We applied the recommended mask that was also used to create the KiDS-1000 shear catalogues[9]. In addition, objects

[5] The spectral energy distributions were constrained to the rest frame wavelength range 3000−11 000 Å.

[6] MAG_AUTO_CALIB = MAG_AUTO + DMAG − EXTINCTION_R

[7] Our star-galaxy separation corresponds to applying the following flags: SG2DPHOT = 0, SG_FLAG = 1, CLASS_STAR < 0.5.

[8] The IMAFLAGS_ISO cut corresponds to applying all MASK values (1, 2, 4, 8, 16, 32 and 64) described in Appendix A.1.1 of Kuijken et al. (2019).

[9] This mask corresponds to the nine-band KiDS MASK bit values 2–11, 13 and 14, described in Appendix A.2 of Kuijken et al. (2019).

that are not detected in all 9 bands were removed from the sample. Our final sample of KiDS-bright lenses consists of ∼1 million galaxies, more than fivefold the number of GAMA galaxies. This increased lens sample allowed us to verify the results from Brouwer et al. (2017) with increased statistics, and to study possible dependencies of the RAR on galaxy observables.

To use the KiDS-bright sample as lenses to measure $g_{\rm obs}$, we needed accurate individual redshifts for all galaxies in our sample. These photometric redshifts $z_{\rm ANN}$ were derived from the full nine-band KiDS+VIKING photometry by training on the spectroscopic GAMA redshifts (see Sect. 3.2) using the ANNz2 (Artificial Neural Network) machine learning method (Sadeh et al. 2016). When comparing this $z_{\rm ANN}$ to the spectroscopic GAMA redshifts $z_{\rm G}$ measured for the same galaxies, we found that their mean offset $\langle(z_{\rm ANN} - z_{\rm G})/(1 + z_{\rm G})\rangle = 9.3 \times 10^{-4}$. However, this offset is mainly caused by the low-redshift galaxies: $z_{\rm ANN} < 0.1$. Removing these reduces the mean offset to $\langle\delta z/(1 + z_{\rm G})\rangle = -6 \times 10^{-5}$, with a standard deviation $\sigma_z = \sigma(\delta z) = 0.026$. This corresponds to a redshift-dependent deviation of $\sigma_z/(1 + \langle z_{\rm ANN}\rangle) = 0.02$ based on the mean redshift $\langle z_{\rm ANN}\rangle = 0.25$ of KiDS-bright between $0.1 < z < 0.5$, which is the lens redshift range used throughout this work for all lens samples.

In order to measure the expected baryonic acceleration $g_{\rm bar}$, we computed the KiDS-bright stellar masses $M_\star$ based on these ANNz2 redshifts and the nine-band GAAP photometry. Because the GAAP photometry only measures the galaxy magnitude within a specific aperture size, the stellar mass was corrected using the 'fluxscale' parameter[10] The stellar masses were computed using the LePhare algorithm (Arnouts et al. 1999; Ilbert et al. 2006), which performs SPS model fits on the stellar component of the galaxy spectral energy distribution. We used the Bruzual & Charlot (2003) SPS model, with the IMF from Chabrier (2003, equal to those used for the GAMA stellar masses). LePhare provides both the best-fit logarithmic stellar mass value 'MASS_BEST' of the galaxy template's probability distribution function, and the 68% confidence level upper and lower limits. We used the latter to estimate the statistical uncertainty on $M_\star$. For both the upper and lower limit, the mean difference with the best-fit mass is approximately: $|\log_{10}\langle M_{\rm lim}/M_{\rm best}\rangle| \approx 0.06$ dex.

Another way of estimating the statistical uncertainty in the stellar mass is to combine the estimated uncertainties from the input: the redshifts and magnitudes. The redshift uncertainty $\sigma_z/\langle z_{\rm G}\rangle = 0.11$ corresponds to an uncertainty in the luminosity distance of: $\sigma(\delta D_{\rm L})/\langle D_{\rm L}\rangle = 0.12$. We took the flux F to remain constant between measurements, such that: $4\pi D_{\rm L}^2 F \propto D_{\rm L}^2 \propto L$. Assuming that approximately $L \propto M_\star$ leads to an estimate:

$$\frac{M_\star + \delta M_\star}{M_\star} = \frac{D_{\rm L}(\bar{z}) + D_{\rm L}(\bar{z} + \delta z)^2}{D_{\rm L}(\bar{z})^2}, \tag{22}$$

which finally gives our adopted stellar mass uncertainty resulting from the KiDS-bright redshifts: $\log_{10}(1 + \delta M_\star/M_\star) = 0.11$ dex. The uncertainty resulting from the KiDS-bright magnitudes is best estimated by comparing two different KiDS apparent magnitude measurements: the elliptical aperture magnitudes 'MAG_AUTO_CALIB' from SExtractor and the Sérsic magnitudes 'MAG_2dphot' from 2DPHOT (La Barbera et al. 2008). The standard deviation of their difference, $\delta m = m_{\rm 2dphot} - m_{\rm calib}$, is $\sigma(\delta m) = 0.69$, which corresponds to a flux ratio of $F_{\rm 2dphot}/F_{\rm calib} = 1.88$ (or 0.27 dex). Using the same assumption, now taking $D_{\rm L}$ to remain constant, results in: $4\pi D_{\rm L}^2 F \propto F \propto L \propto M_\star$. This means our flux ratio uncertainty directly corresponds to our estimate of the $M_\star$ uncertainty. Quadratically combining the 0.11 dex uncertainty from the redshifts and the 0.27 dex uncertainty from the magnitudes gives an estimate of the total statistical uncertainty on the stellar mass of ∼0.29 dex. This is much larger than that from the LePhare code. Taking a middle ground between these two, we have assumed twice the LePhare estimate: $\sigma_{M_\star} = 0.12$ dex. However, we have confirmed that using the maximal estimate $\sigma_{M_\star} = 0.29$ dex throughout our analysis does not change the conclusions of this work, in particular those of Sect. 5.4.

When comparing $M_{\star,\rm ANN}$ with the GAMA stellar masses $M_{\star,\rm G}$ of matched galaxies, we found that its distribution is very similar, with a standard deviation of 0.21 dex around the mean. Nevertheless there exists a systematic offset of $\log(M_{\star,\rm ANN}) - \log(M_{\star,\rm G}) = -0.056$ dex, which is caused by the differences in the adopted stellar mass estimation methods. In general, it has been found impossible to constrain stellar masses to within better than a systematic uncertainty of $\Delta M_\star \approx 0.2$ dex when applying different methods, even when the same SPS, IMF and data are used (Taylor et al. 2011; Wright et al. 2017). We therefore normalised the $M_{\star,\rm ANN}$ values of our KiDS-bright sample to the mean $M_{\star,\rm G}$ of GAMA, while indicating throughout our results the range of possible bias due to a $\Delta M_\star = 0.2$ dex systematic shift in $M_\star$. We estimated the effect of this bias by computing the RAR with $\log_{10}(M_\star) \pm \Delta M_\star$ as upper and lower limits.

In order to compare our observations to the MG theories, the measured lensing profiles of our galaxies should not be significantly affected by neighbouring galaxies, which we call 'satellites'. We defined our isolated lenses (Appendix A) such that they do not have any satellites with more than a fraction $f_{M_\star} \equiv M_{\star,\rm sat}/M_{\star,\rm lens}$ of their stellar mass within a spherical radius $r_{\rm sat}$ (where $r_{\rm sat}$ was calculated from the projected and redshift distances between the galaxies). We chose $f_{\rm M_\star} = 0.1$, which corresponds to 10% of the lens stellar mass, and $r_{\rm sat} = 3\,h_{70}^{-1}$ Mpc, which is equal to the maximum projected radius of our measurement. In short: $r_{\rm sat}(f_{M_\star} > 0.1) > 3\,h_{70}^{-1}$ Mpc. We also restricted our lens stellar masses to $M_\star < 10^{11}\,h_{70}^{-2}\,M_\odot$ since galaxies with higher masses have significantly more satellites (see Sect. 2.2.3 of Brouwer et al. 2017). This provided us with an isolated lens sample of 259 383 galaxies. We provide full details of our choice of isolation criterion and an extensive validation of the isolated galaxy sample in Appendix A. Based on tests with KiDS, GAMA and MICE data we found that this is the optimal isolation criterion for our data. The ESD profile of our isolated sample is not significantly affected by satellite galaxies and that our sample is accurate to ∼80%, in spite of it being flux-limited. Using the MICE simulation we also estimated that the effect of the photometric redshift error is limited.

4. Simulations

In order to compare our observations to ΛCDM-based predictions, we used two different sets of simulations: MICE and BAHAMAS. Here MICE is an N-body simulation, which means that galaxies are added to the DM haloes afterwards, while BAHAMAS is a hydrodynamical simulation that incorporates both stars and gas through sub-grid physics. MICE, however, has a simulation volume at least two orders of magnitude larger than BAHAMAS. Below we explain the details of each simulation, and how we utilised their unique qualities for our analysis.

[10] This fluxscale correction of the stellar mass $M_\star$ was applied to LePhare's best-fit mass value as follows: $M_\star$ = MASS_BEST + (MAG_GAAP_r – MAG_AUTO_CALIB)/2.5, where the latter are the GAAP and calibrated elliptical r-band magnitudes.

4.1. MICE mock catalogues

The MICE N-body simulation contains $\sim 7 \times 10^{10}$ DM particles in a $(3072\,h_{70}^{-1}\,\mathrm{Mpc})^3$ comoving volume (Fosalba et al. 2015a). From this simulation the MICE collaboration constructed a $\sim$5000 deg^2 lightcone with a maximum redshift of $z = 1.4$. The DM haloes in this lightcone were identified using a Friend-of-Friend algorithm on the particles. These DM haloes were populated with galaxies using a hybrid halo occupation distribution (HOD) and halo abundance matching (HAM) prescription (Carretero et al. 2015; Crocce et al. 2015). The galaxy luminosity function and colour distribution of these galaxies were constructed to reproduce local observational constraints from SDSS (Blanton et al. 2003a,b, 2005).

In the MICECATv2.0 catalogue[11], every galaxy had sky coordinates, redshifts, comoving distances, apparent magnitudes and absolute magnitudes assigned to them. Of the total MICE lightcone we used 1024 deg^2, an area similar to the KiDS-1000 survey. We used the SDSS apparent r-band magnitudes m_r as these most closely match those from KiDS (see Brouwer et al. 2018). We could therefore limit the MICE galaxies to the same apparent magnitude as the KiDS-bright sample: $m_r < 20$ mag, in order to create a MICE foreground galaxy (lens) sample. We used the same redshift limit: $0.1 < z < 0.5$, resulting in a mean MICE lens redshift $\langle z \rangle = 0.23$, almost equal to that of GAMA and KiDS-bright within this range. The absolute magnitudes of the mock galaxies go down to $M_r - 5\log_{10}(h_{100}) < -14$ mag, which corresponds to the faintest GAMA and KiDS-bright galaxies. Each galaxy was also assigned a stellar mass $M_\star$, which is needed to compute the RAR (see Sect. 2.2). These stellar masses were determined from the galaxy luminosities L using Bell & de Jong (2001) $M_\star/L$ ratios.

In addition, each galaxy had a pair of lensing shear values associated with it (γ_1 and γ_2, with respect to the Cartesian coordinate system). These shear values were calculated from HEALPIX weak lensing maps that were constructed using the 'onion shell method' (Fosalba et al. 2008, 2015b). The lensing map of MICECATv2.0 has a pixel size of 0.43 arcmin. We did not use MICE results within a radius $R_{\rm res}$ corresponding to 3 times this resolution. We calculated $R_{\rm res}$ and the corresponding $g_{\rm bar}$ using the mean angular diameter distance and baryonic mass of the MICE lens sample. For the full sample of isolated MICE galaxies these values are: $R_{\rm res} = 0.25\,h_{70}^{-1}$ Mpc and $g_{\rm bar} = 6.60 \times 10^{-14}$ m s^{-2}.

At scales larger than this resolution limit, the MICE shears allowed us to emulate the GGL analysis and conversion to the RAR that we performed on our KiDS-1000 data (as described in Sect. 2) using the MICE simulation. To create a sample of MICE background galaxies (sources) for the lensing analysis, we applied limits on the MICE mock galaxies' redshifts and apparent magnitudes, which are analogous to those applied to the KiDS source sample: $0.1 < z < 1.2$, $m_r > 20$ (see Hildebrandt et al. 2017 and Sect. 3.1; uncertainties in the KiDS $z_{\rm B}$ are not accounted for in this selection). We also applied an absolute magnitude cut of $M_r > -18.5$ mag, in order to reproduce the KiDS source redshift distribution more closely.

The MICE mock catalogue also features very accurate clustering. At lower redshifts ($z < 0.25$) the clustering of the mock galaxies as a function of luminosity was constructed to reproduce the Zehavi et al. (2011) clustering observations, while at higher redshifts ($0.45 < z < 1.1$) the MICE clustering was validated against the Cosmic Evolution Survey (COSMOS; Ilbert et al. 2009). The accurate MICE galaxy clustering allowed us to analyse the RAR at larger scales ($>0.3\,h_{70}^{-1}$ Mpc) where clustered neighbouring galaxies start to affect the lensing signal. MICE also allowed us to test our criteria defining galaxy isolation (see Appendix A).

4.2. BAHAMAS mock catalogue

The second set of simulations that we utilised is BAHAMAS (McCarthy et al. 2017). The BAHAMAS suite are smoothed-particle hydrodynamical realisations of $(400\,h_{100}^{-1}\mathrm{Mpc})^3$ volumes and include prescriptions for radiative cooling and heating, ionising background radiation, star formation, stellar evolution and chemical enrichment, (kinetic wind) supernova feedback, supermassive black hole accretion, and merging and thermal feedback from active galactic nuclei (AGN). The simulations were calibrated to reproduce the stellar and hot gas content of massive haloes, which makes them particularly well suited for our study of the matter content around haloes out to distances of 1–3 h_{70}^{-1} Mpc. The masses of DM and baryonic resolution elements are $3.85\times10^9\,h_{100}^{-1}\,M_\odot$ and $7.66\times10^8\,h_{100}^{-1}\,M_\odot$ respectively, and the gravitational softening is fixed at $\epsilon = 4\,h_{100}^{-1}$ kpc $= 5.71\,h_{70}^{-1}$ kpc.

Haloes and galaxies were identified in the simulations using the friends-of-friends (Davis et al. 1985) and SUBFIND (Springel et al. 2001; Dolag et al. 2009) algorithms. We labeled the most massive sub-halo in each Friend-of-Friend group as the 'central' and other sub-haloes as 'satellites'. We constructed an 'isolated' galaxy sample by restricting the selection to central sub-haloes that have no other sub-haloes (satellites or centrals) more massive than 10% of their mass within $3\,h_{70}^{-1}$ Mpc. We randomly selected 100 galaxies per 0.25 dex bin in M_{200} between 10^{12} and $10^{13.5}\,h_{70}^{-2}\,M_\odot$. In the last two bins there were fewer than 100 candidates, so we selected them all. All galaxies have a redshift $z = 0.25$. For each selected galaxy we constructed an integrated surface density map, integrated along the line-of-sight for ±15 comoving h_{100}^{-1} Mpc around the target halo. We also extracted the cumulative spherically averaged mass profile of each target sub-halo, decomposed into DM, stars, and gas. For both the maps and profiles, we included mass contributions from all surrounding (sub)structures: we did not isolate the haloes from their surrounding environment.

We used the integrated surface density map of each galaxy to calculate its mock ESD profile as a function of the projected distance R from the lens centre, in order to mimic the effect of GGL and the conversion to the RAR on the BAHAMAS results. Each pixel on these maps corresponds to 15 comoving h_{100}^{-1} kpc, which in our physical units is: $15/(1+z)\,0.7^{-1}\,h_{70}^{-1}$ kpc $= 17.14\,h_{70}^{-1}$ kpc. The density maps each have a dimensionality of 400×400 pixels. Hence the total area of each map is $(6.86\,h_{70}^{-1}\,\mathrm{Mpc})^2$. In calculating the lensing profiles and RAR with BAHAMAS we followed, as closely as possible, the GGL procedure and conversion to the RAR as described in Sect. 2. We truncated our lensing profiles at 10 times the gravitational softening length: $10\,\epsilon = 0.057\,h_{70}^{-1}$ Mpc, to avoid the numerically poorly converged central region (Power et al. 2003). For a typical galaxy in our sample of isolated BAHAMAS galaxies, this corresponds to $g_{\rm bar} \sim 2.38 \times 10^{-12}$ m s^{-2}.

4.3. The BAHAMAS RAR: Quantifying the missing baryon effect

The calculation of the expected baryonic radial acceleration $g_{\rm bar}$ requires the enclosed baryonic mass $M_{\rm bar}(<r)$ within a spherical

[11] The MICECATv2.0 catalogue is available through CosmoHub (https://cosmohub.pic.es).

radius r around the galaxy centre. Since we are dealing with measurements around isolated galaxies at $R > 30\,h_{70}^{-1}$ kpc, we can approximate $M_{\rm bar}(<r)$ as a point mass $M_{\rm gal}$ mainly composed of the mass of the lens galaxy itself. $M_{\rm gal}$ can be subdivided into stars and gas, and the latter further decomposed into cold and hot gas.

How we obtained the stellar masses of our GAMA, KiDS-bright, MICE and BAHAMAS galaxies is described in Sects. 3 and 4. From these $M_\star$ values, the fraction of cold gas $f_{\rm cold} = M_{\rm cold}/M_\star$ can be estimated using scaling relations based on H I and CO observations. Following Brouwer et al. (2017) we used the best-fit scaling relation found by Boselli et al. (2014), based on the *Herschel* Reference Survey (Boselli et al. 2010):

$$\log(f_{\rm cold}) = -0.69\,\log(M_\star/h_{70}^{-2}\,M_\odot) + 6.63. \quad (23)$$

We applied this equation to all observed and simulated values of $M_\star$ in order to arrive at the total galaxy mass: $M_{\rm gal} = M_\star + M_{\rm cold} = M_\star(1 + f_{\rm cold})$. The spatial distribution of the stellar and cold gas mass are similar (Pohlen et al. 2010; Crocker et al. 2011; Mentuch Cooper et al. 2012; Davis et al. 2013) and can therefore be considered a single mass distribution, especially for the purposes of GGL, which only measures the ESD profile at scales larger than the galaxy disc ($R > 30\,h_{70}^{-1}$ kpc). We illustrate this in Fig. 1, which shows the enclosed mass profiles (upper panel) and RAR (lower panel) for different baryonic components in the BAHAMAS simulation. For these mock galaxies, the stellar mass within $30\,h_{70}^{-1}$ kpc (red star) gives a good approximation of the $M_\star$ distribution across all radii that we consider. We therefore modeled the baryonic mass of our galaxies as a point mass $M_{\rm gal}$, containing both the stellar and cold gas mass.

We recognise that the total baryonic mass distribution $M_{\rm bar}$ of galaxies may include a significant amount of additional mass at larger distances, notably in the hot gas phase. This is illustrated in Fig. 1. In the upper panel, we show the average baryonic mass profile for BAHAMAS galaxies with $1 < M_{200}/(10^{12}\,h_{70}^{-2}\,M_\odot) < 3$. In addition, we show an estimate of the typical baryonic mass profile for galaxies in the same mass range, based on an extrapolation to larger radii of the compilation of observations in Tumlinson et al. (2017); including stars, cold gas ($<10^4$ K, traced by absorption lines such as H I, Na I and Ca II), cool gas (10^4–10^5 K, traced by many UV absorption lines, e.g., Mg II, C II, C III, Si II, Si III, N II, N III), warm gas (10^5–10^6 K, traced by C IV, N V, O VI and Ne VII absorption lines), hot gas ($>10^6$ K, traced by its X-ray emission) and dust (estimated from the reddening of background QSOs, and Ca II absorption). The light blue shaded region therefore illustrates a component of missing baryons predicted by these simulations but not (yet) observed, possibly related to the cosmological missing baryons (e.g., Fukugita et al. 1998; Fukugita & Peebles 2004; Shull et al. 2012). There are several possibilities: (i) there may be additional gas present in a difficult-to-observe phase (e.g., hot, low-density gas, see for instance Nicastro et al. 2018); (ii) the simulations do not accurately reflect reality, for example: galaxies may eject substantially more gas from their surroundings than is predicted by these simulations; (iii) there may be less baryonic matter in the Universe than expected in the standard cosmology based on big bang nucleosynthesis (BBN; Kirkman et al. 2003) calculations and CMB measurements (Spergel et al. 2003; Planck Collabration XVI 2014).

The lower panel of Fig. 1 illustrates the magnitude of the resulting systematic uncertainties in $g_{\rm bar}$. In the ΛCDM cosmology, the expectation at sufficiently large radii is given by

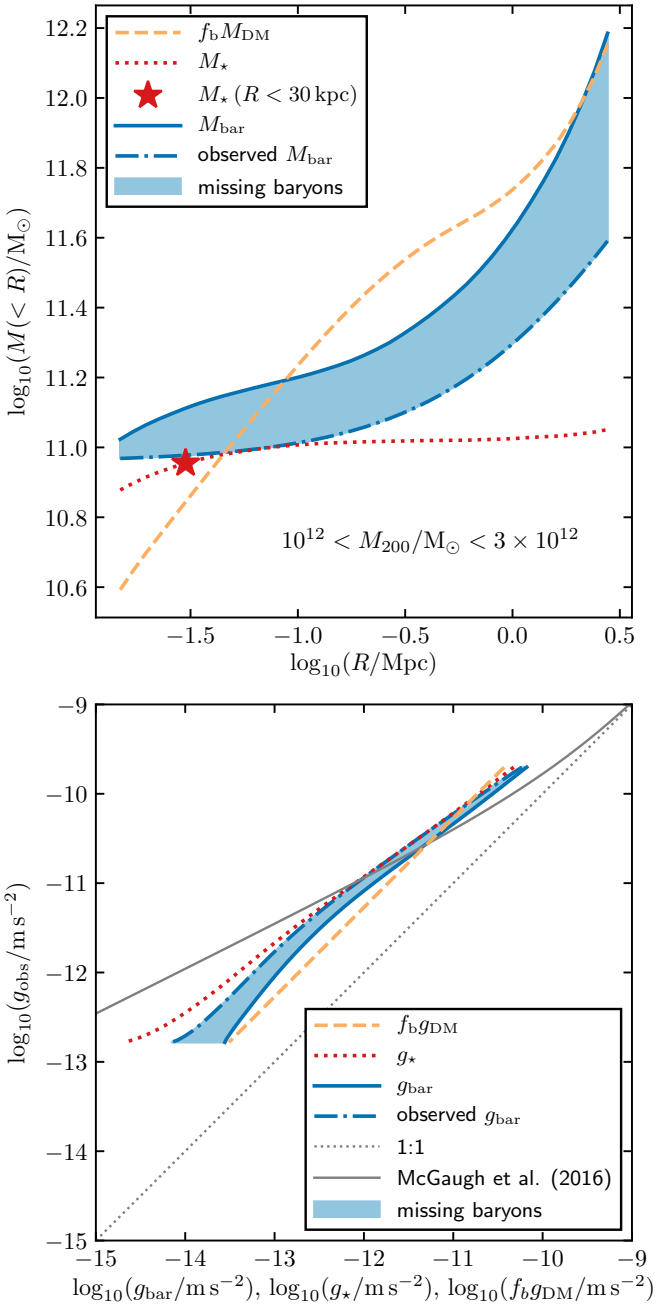

Fig. 1. Mass profiles and RAR of BAHAMAS galaxies. *Upper panel*: cumulative mass profiles of stars (red dotted line) and total baryons (blue solid line) for BAHAMAS galaxies with $1 < M_{200}/(10^{12}\,h_{70}^{-2}\,M_\odot) < 3$. The star marker indicates the stellar mass within a $30\,h_{70}^{-1}$ kpc aperture, indicative of what is typically regarded as the stellar mass of a galaxy. The blue dash-dotted line shows the typical baryonic mass profile of observed galaxies of similar mass, estimated based on an extrapolation of the compilation in Fig. 7 of Tumlinson et al. (2017). In the inner galaxy the discrepancy (light blue shaded region) between the observed and simulated $M_{\rm bar}$ is relatively small, but in the outer galaxy the majority of the baryons predicted to be present in BAHAMAS consist of currently unobserved, missing baryons. The orange dashed line shows the expected baryonic mass profile if the baryon density is everywhere equal to a fixed fraction $f_{\rm b} = \Omega_{\rm b}/\Omega_{\rm m}$ of the local DM density. At large enough radii ($\gtrsim 2\,h_{70}^{-1}$ Mpc), the baryon-to-DM ratio converges to the cosmic average. *Lower panel*: as in upper panel, but in acceleration space. The cosmic baryon fraction provides a strong theoretical upper limit on $g_{\rm bar}$ at low accelerations in the context of the ΛCDM cosmology.

$g_{\rm obs} = f_{\rm b}^{-1} g_{\rm bar}$ where $f_{\rm b}$ is the cosmic baryon fraction $f_{\rm b} = \Omega_{\rm b}/\Omega_{\rm m} = 0.17$ (Hinshaw et al. 2013). BAHAMAS, and generically any ΛCDM galaxy formation simulation, converges to this density at low enough accelerations (large enough radii). The most optimistic extrapolation of currently observed baryons falls a factor of ∼3 short of this expectation, while the stellar mass alone is a further factor of ∼3 lower. The unresolved uncertainty around these missing baryons is the single most severe limitation of our analysis. Given that we are interested in both ΛCDM and alternative cosmologies, we will use the stellar + cold gas mass $M_{\rm gal}$ as our fiducial estimate of the total baryonic mass $M_{\rm bar}$, which is translated into the baryonic acceleration $g_{\rm bar}$, throughout this work. This serves as a secure lower limit on $g_{\rm bar}$. We note that the eventual detection, or robust non-detection, of the missing baryons has direct implications for the interpretation of the results presented in Sect. 5. In Sect. 5.2 we address the possible effect of extended hot gas haloes on $g_{\rm bar}$. We discuss this issue further in Sect. 6.

Concerning $g_{\rm obs}$, omitting the contribution of hot gas will not have a large effect on the prediction within the ΛCDM framework (e.g., from simulations) since the total mass distribution at the considered scales is heavily dominated by DM. Within MG frameworks such as EG and MOND, where the excess gravity is sourced by the baryonic matter, it is slightly more complicated. Brouwer et al. (2017, see Sect. 2.2) carefully modelled the distribution of all baryonic components, based on observations from both GAMA and the literature, including their effect on the excess gravity in the EG framework. They found that, for galaxies with $M_\star < 10^{11}\, h_{70}^{-2}\, M_\odot$, the contribution to the ESD profile (and hence to $g_{\rm obs}$) from hot gas and satellites was small compared to that of the stars and cold gas. Although this analysis was done for the EG theory, the effect of these extended mass distributions within MOND are similar or even less. This allows us to use a point mass $M_{\rm gal}$ as a reasonable approximation for the baryonic mass distribution $M_{\rm bar}(<r)$ within our measurement range when computing $g_{\rm obs}$ as predicted by MOND and EG (see Sects. 2.3 and 2.4).

4.4. The BAHAMAS RAR: Testing the ESD to RAR conversion

We used BAHAMAS to test the accuracy of our SIS method (outlined in Sect. 2.2) in estimating $g_{\rm obs}$ from our GGL measurement of $\Delta\Sigma_{\rm obs}$, by comparing it against the more sophisticated piece-wise power law (PPL) method outlined in Appendix B. As a test system, we used the 28 galaxies from our BAHAMAS sample with $10^{13} < M_{200}/(\,h_{70}^{-2}\, M_\odot) < 10^{13.1}$. We combined these into a stacked object by averaging the individual ESD profiles as derived from their mock lensing maps. The stacked ESD as measured from the lensing mocks is shown in the left panel of Fig. 2. Since the mock ESD profiles are derived from convergence maps (rather than the shapes of background galaxies), they have no associated measurement uncertainty – for simplicity, we assumed a constant 0.1 dex uncertainty, which is similar to that for the KiDS measurements. We also combined the spherically averaged enclosed mass profiles of the galaxies out to $3\, h_{70}^{-1}$ Mpc by averaging them. From this average mass profile we analytically calculated the ESD profile shown in the left panel of Fig. 2. We found that the $\Delta\Sigma$ calculated from the spherically averaged mass profile is ∼0.05 dex higher than the direct measurement of the stacked lensing mocks. This primarily results from the fact that the spherically averaged mass profile does not take into account the additional matter outside the $3\, h_{70}^{-1}$ Mpc spherical aperture, whereas the mock surface density maps are integrated along the line-of-sight for ±15 comoving h_{100}^{-1} Mpc around the lens.

The PPL method described in Appendix B attempts to reproduce the ESD profile by converging to an appropriate volume density profile. The resulting recovered ESD profile and its 68% confidence interval is shown with blue points and error bars in the left panel of Fig. 2 – the fit to the mock data is excellent. In the centre panel we show the enclosed mass profile as recovered by both the PPL and SIS methods, in addition to the true enclosed mass profile. Both estimators recover the profile within their stated errors. The PPL method systematically underestimates it by ∼0.1 dex across most of the radial range. This is directly caused by the difference between the spherically averaged and mock lensing ESD profiles (left panel). The somewhat wider confidence intervals at small radii are caused by the lack of information in the mock data as to the behaviour of the profile at $r < 30\, h_{70}^{-1}$ kpc; the PPL model marginalises over all possibilities. Once the enclosed mass is dominated by the contribution at radii covered by the measurement, the uncertainties shrink. To account for the added uncertainty resulting from the conversion to the RAR, we added 0.1 dex to the error bars of our RAR measurements throughout this work.

The SIS method instead slightly underestimates the enclosed mass at small radii, and overestimates it at large radii. The apparent improved performance relative to the PPL method is actually due to a fortuitous partial cancellation of two errors. First, the SIS calculation suffers from the same underestimation of the spherically averaged enclosed mass profile as the PPL method, due to the difference between the mock lensing and spherically averaged ESD profiles. However, in addition to this, the SIS method assumes a density profile $\rho(r) \propto r^{-2}$ at all radii. At small radii, the power-law slope is in reality about −2.1. This results in a slight overestimate of the enclosed mass, which partially compensates the underestimate described above, resulting in a net underestimate. At larger radii, the slope of the density profile becomes progressively steeper, such that the assumption of an r^{-2} profile increasingly overestimates the enclosed mass, eventually resulting in a net overestimate.

The right panel of Fig. 2 illustrates the resulting uncertainty in the measurement of the RAR. To focus on the influence of the method used to recover $g_{\rm obs}$, we simply used the exact spherically averaged stellar mass profile to calculate $g_\star$, plotted on the x-axis[12]. We found that, for mock lenses within the BAHAMAS simulation, both the SIS and the PPL method yield acceptable and consistent estimates of $g_{\rm obs}$. We note that the BAHAMAS $g_{\rm obs}(g_\star)$ is significantly offset from the RAR as measured by M16; we will return to this point when we compare BAHAMAS to our observations in Sect. 5.3.

5. Results

Tables containing the ESD profile data used to create all results figures (i.e., Figs. 3, 4, 5, 8, 9, 10, A.4 and C.1) can be found online[13].

5.1. Lensing rotation curves

As a final consistency check between the SIS assumption and the PPL method, we applied both methods to the true KiDS-1000 data. Since these methods are only used to convert $\Delta\Sigma(R)$

[12] We do not include the additional gas, which is predominantly in the hot phase, for consistency with the presentation of the results in Sect. 5.
[13] http://kids.strw.leidenuniv.nl/sciencedata.php

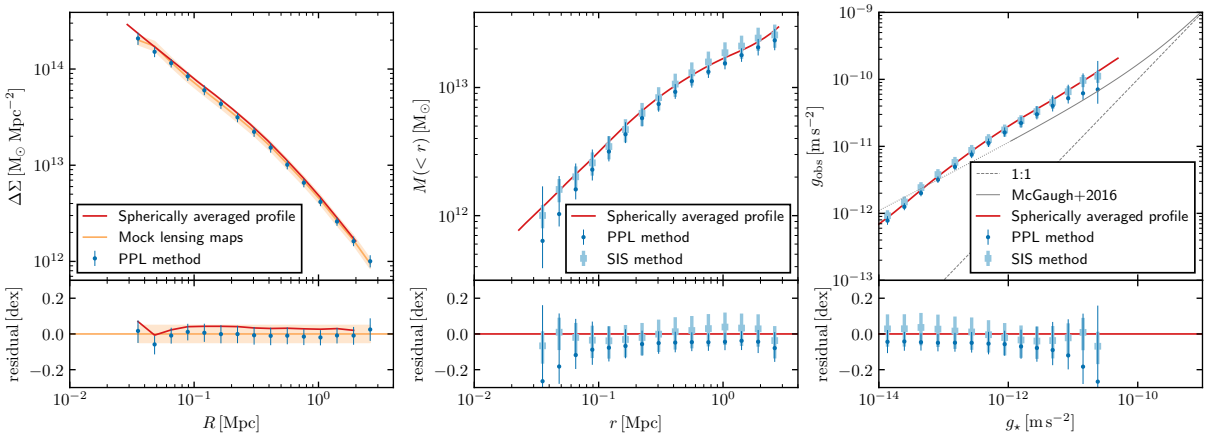

Fig. 2. Illustration of the recovery of the acceleration profile from simulated weak lensing observations. *Left*: average ESD profile of a subset of our sample of BAHAMAS galaxies with $10^{13} < M_{200}/(h_{70}^{-2}\,M_\odot) < 10^{13.1}$, derived from the spherically averaged mass profile (red line) and the mock lensing maps (yellow line, with an assumed 0.1 dex Gaussian uncertainty). The PPL method recovery of the ESD profile is shown with the blue points; error bars represent 68% confidence intervals. *Centre*: SIS (light blue squares) and PPL (dark blue points) method recover the spherically averaged enclosed mass profile. The uncertainties on the SIS points are derived by sampling the uncertainties on the mock lensing ESD profile. *Right*: resulting dynamical acceleration profile $g_{\rm obs}$ and uncertainties, plotted as a function of the acceleration due to stars $g_\star = GM_\star(<r)/r^2$.

into $g_{\rm obs}(r)$, we can leave $g_{\rm bar}$ out of the comparison and plot our results as a function of R. An observable closely related to the RAR that is usually plotted as a function of radius, is the traditional circular velocity curve:

$$v_{\rm circ}(r) = \sqrt{\frac{GM_{\rm obs}(<r)}{r}}, \tag{24}$$

an observable that indeed served as input to the M16 RAR measurement. We applied the SIS method described in Sect. 2.2 to convert our ESD profiles $\Delta\Sigma(R)$ into $v_{\rm circ}(R)$ since substituting Eq. (6) into Eq. (24) gives:

$$v_{\rm circ}(r) = \sqrt{\frac{G\,(4\Delta\Sigma(r)\,r^2)}{r}} = \sqrt{4G\,\Delta\Sigma(r)\,r}. \tag{25}$$

We also applied Eq. (24) to compute $v_{\rm circ}(R)$ from the $M(<R)$ calculated through the PPL method described in Appendix B. We note that both the SIS and PPL method assume spherical symmetry, while in simulations DM haloes are found to deviate from sphericity, which could lead to deviations in the lensing rotation curves (Cuddeford 1993). However, the mean ellipticity of haloes is observed to be small ($\langle|\epsilon|\rangle = 0.174 \pm 0.046$, Schrabback et al. 2021). The stacking of thousands of lenses with approximately random orientations further reduces the impact on the lensing signal, which means the halo ellipticity will not significantly change our results.

Figure 3 shows the lensing rotation curves for isolated KiDS-bright galaxies, divided into four stellar mass bins using the following limits: $\log_{10}(M_\star/h_{70}^{-2}\,M_\odot) = [8.5, 10.3, 10.6, 10.8, 11.0]$. For each bin the mean galaxy mass (stars+cold gas) of the lenses, $\log_{10}\langle M_{\rm gal}/h_{70}^{-2}\,M_\odot\rangle = [10.14, 10.57, 10.78, 10.96]$, is shown at the top of the panel. Showing the data in this way allows us to observe for the first time in this intuitive manner how the circular velocity curves of isolated galaxies continue beyond the observable disc ($r > 30\,h_{70}^{-1}$ kpc). In addition, it provides a consistency check against the SPARC rotation curves (Lelli et al. 2016) that form the basis for the M16 RAR measurement. It is remarkable how well the mean of the SPARC rotation curves and our lensing results correspond at their intersection ($r \sim 30\,h_{70}^{-1}$ kpc). But most importantly, we find that the 'lensing rotation curves' from the SIS assumption are consistent with the ones from the PPL method. Although the SIS assumption results in slightly more scatter, there is very little systematic bias between the results from the two methods, which have a fractional difference of $\langle\log(v_{\rm circ,SIS}/v_{\rm circ,PPL})\rangle = 0.017$ dex. Since this measurement is merely a different way of presenting the observed acceleration, which equals $g_{\rm obs}(r) = v_{\rm circ}^2/r$, we can easily compute that the expected difference in $g_{\rm obs}$ would be $\langle\log(g_{\rm obs,SIS}/g_{\rm obs,PPL})\rangle = 0.038$ dex.

The consistency between the two conversion methods allows us to use the SIS assumption throughout this work. The great advantage of this method is that it allows us to convert GGL profiles binned by baryonic acceleration $\Delta\Sigma(g_{\rm bar})$, into the RAR: $g_{\rm obs}(g_{\rm bar})$. This is not the case for the PPL method, which only works on $\Delta\Sigma(R)$ binned by radius. The former can therefore be applied to any lens sample; the latter only to lenses within a narrow mass range (in order to convert R into $g_{\rm bar}$ using the mean $\langle M_{\rm gal}\rangle$). As explained in Sect. 4.4 we added 0.1 dex to the error bars of all RAR measurements in this work, to account for the added uncertainty from the conversion of the ESD to the RAR. After showing that both methods yield acceptable and consistent estimates of $g_{\rm obs}$, we will show only the SIS measurement when presenting our results in this section to reduce clutter in the figures.

5.2. The RAR of KiDS compared to MG theories

In Fig. 4 we show the RAR, with the observed radial acceleration computed from our lensing measurements through Eq. (7)) on the y-axis. The x-axis shows the expected baryonic (star + cold gas) radial acceleration, where the label serves as a reminder throughout this work that $g_{\rm bar}$ is only computed from the measured stellar masses of the galaxies and an estimate of their cold gas component.

The lensing $g_{\rm obs}$ was measured using the GAMA and KiDS-bright isolated galaxy samples, respectively. Due to its smaller survey area (180 vs. 1006 deg^2), the error bars using GAMA lenses are larger than those using KiDS-bright lenses. However, as explained in Appendix A, the spectroscopic redshifts of the GAMA survey allow for a more reliable selection of the isolated lenses compared to KiDS (which measures photometric redshifts with a $\sigma_z = 0.02$ uncertainty). The effect of this uncertainty on the measured lensing profiles is modelled in Fig. A.3, which shows that the ESD profile of the 'offset' MICE sample diverges

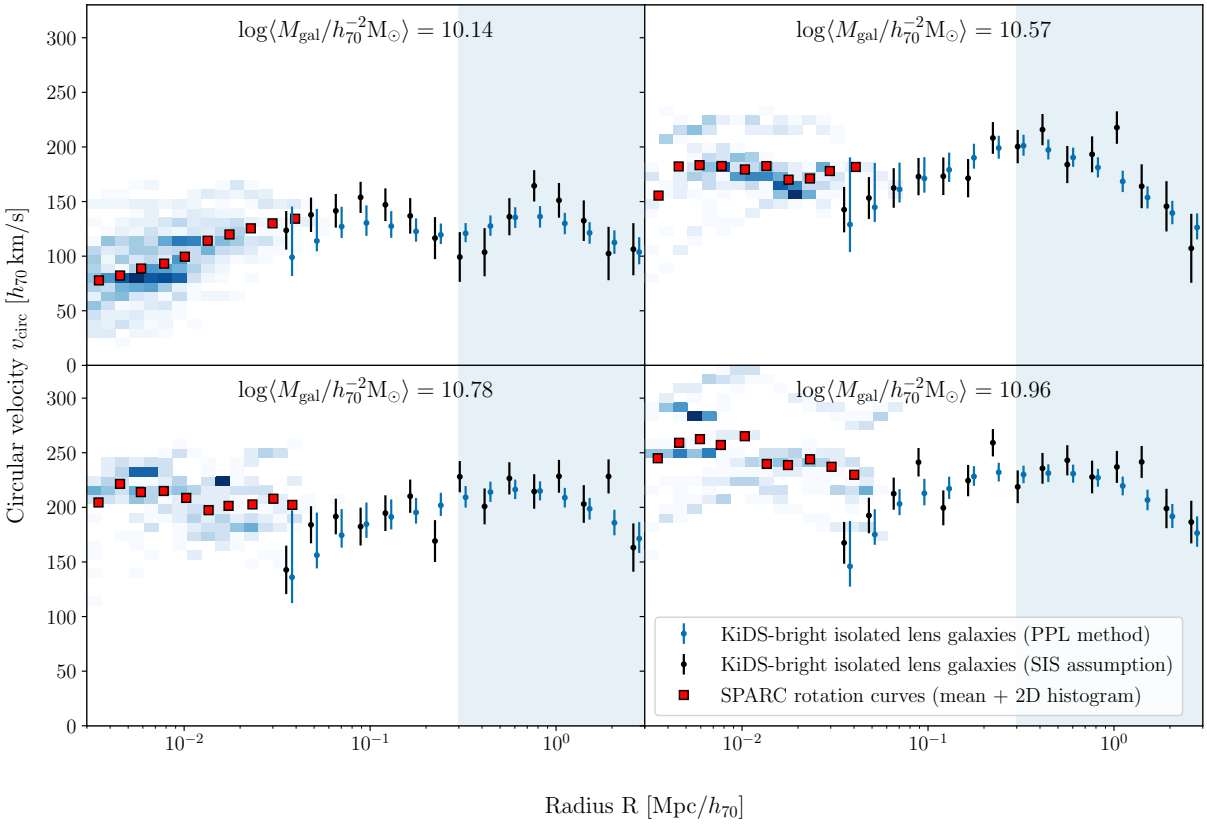

Fig. 3. Measured rotation curves – the circular velocity as a function of radius $v_{\rm circ}(R)$ – of the KiDS-bright isolated lens sample, divided into four stellar mass bins. The mean galaxy mass (stars+cold gas) of the lenses is shown at the top of each panel. The light blue shaded region indicates the radii corresponding to $R > 0.3\,h_{70}^{-1}$ Mpc, where the uncertainty in the photometric KiDS redshifts can affect the isolated lens selection (see Appendix A). The black points (with 1σ error bars) show the result calculated using the SIS assumption, while the blue points (with error bars representing the 16th and 84th percentile of the fits) show the result from the more sophisticated PPL method. Our measurements are consistent between the two methods, and also with the rotation curves from SPARC (all data as the blue 2D histogram, the mean as red squares).

from the truly isolated MICE galaxies at radius $R > 0.3\,h_{70}^{-1}$ Mpc. Age scales, the effect of satellite galaxies on the lensing signal result in a ∼30% increase in $\Delta\Sigma$ due to the contribution of satellite galaxies. We translated this radius into a gravitational acceleration value using Eq. (2), based on the average $M_{\rm gal}$ of the lens sample. In this way we estimate that, for the full sample of isolated KiDS-bright galaxies, the isolation criterion is no longer reliable when $g_{\rm bar} \lessapprox 10^{-13}\,{\rm m\,s^{-2}}$, as indicated by the light blue shaded region in Fig. 4. We note that the GAMA results, which are based on accurate spectroscopic redshift measurements, are still reliable within this region.

The grey band shows the range of possible bias due to a $\Delta M_\star = \pm 0.2$ dex systematic shift in stellar mass. We estimated this range by performing our analysis assuming stellar masses that are 0.2 dex higher than, and then 0.2 dex lower than, their best-fitting $M_\star$ values (see Sect. 3.3). We only show this band once, for the KiDS-bright result, but note that this uncertainty equally affects the GAMA stellar masses (and, indeed, any stellar mass measurement; see Wright et al. 2017).

We compare our results to the M16 RAR measurements (both the full dataset: blue 2D histogram, and the mean: red squares), from SPARC galaxy rotation curves, which cover higher accelerations than our lensing measurements (corresponding to smaller scales: $R < 30\,h_{70}^{-1}$ kpc). At the highest-acceleration end (smallest scales), where $g_{\rm obs}$ is dominated by $g_{\rm bar}$, they follow a one-to-one relation. At lower accelerations (larger scales) their results quickly diverge from unity, signifying the start of the DM dominated regime. We find that these two fully independent RAR observations, respectively from rotation curves and lensing, are in strong agreement[14].

Figure 4 also compares the two MG models, EG and MOND, to our lensing results (for a comparison of these two models with the RAR from SPARC, see Lelli et al. 2017a). As explained in Sects. 2.3 and 2.4, we took the MOND prediction to be equal to the extrapolated M16 fitting function (Eq. (11)), and that of EG as the prediction from Verlinde (2017) for a point mass (Eq. (17)). At high accelerations, the prediction from EG appears to lie above that of MOND and the SPARC data. However, as explained in Sect. 2.4, the prediction of Eq. (17) should be taken with a grain of salt for accelerations $g_{\rm bar} > 1.2 \times 10^{-10}\,{\rm m\,s^{-2}}$. Within our measurement range, the two predictions are almost indistinguishable. Both models are compatible with the GAMA data. The KiDS-bright data points, however, lie systematically above the MG predictions.

To quantify the level of agreement between the acceleration predicted by the different models $g_{\rm mod}$ and the observed $g_{\rm obs}$, we calculated the χ^2 value:

$$\chi^2 = (g_{\rm obs} - g_{\rm mod})^\intercal \cdot C^{-1} (g_{\rm obs} - g_{\rm mod}), \tag{26}$$

where C^{-1} is the inverse of the analytical covariance matrix (see Sect. 2.1). We divided this quantity by the number of degrees of

[14] Because the blinding intended to avoid observer bias in the KiDS-1000 cosmological constraints (Asgari et al. 2021; Heymans et al. 2021; Tröster et al. 2021) only has a small effect on GGL observations, this agreement has been present since the start of our analysis (before the data were un-blinded).

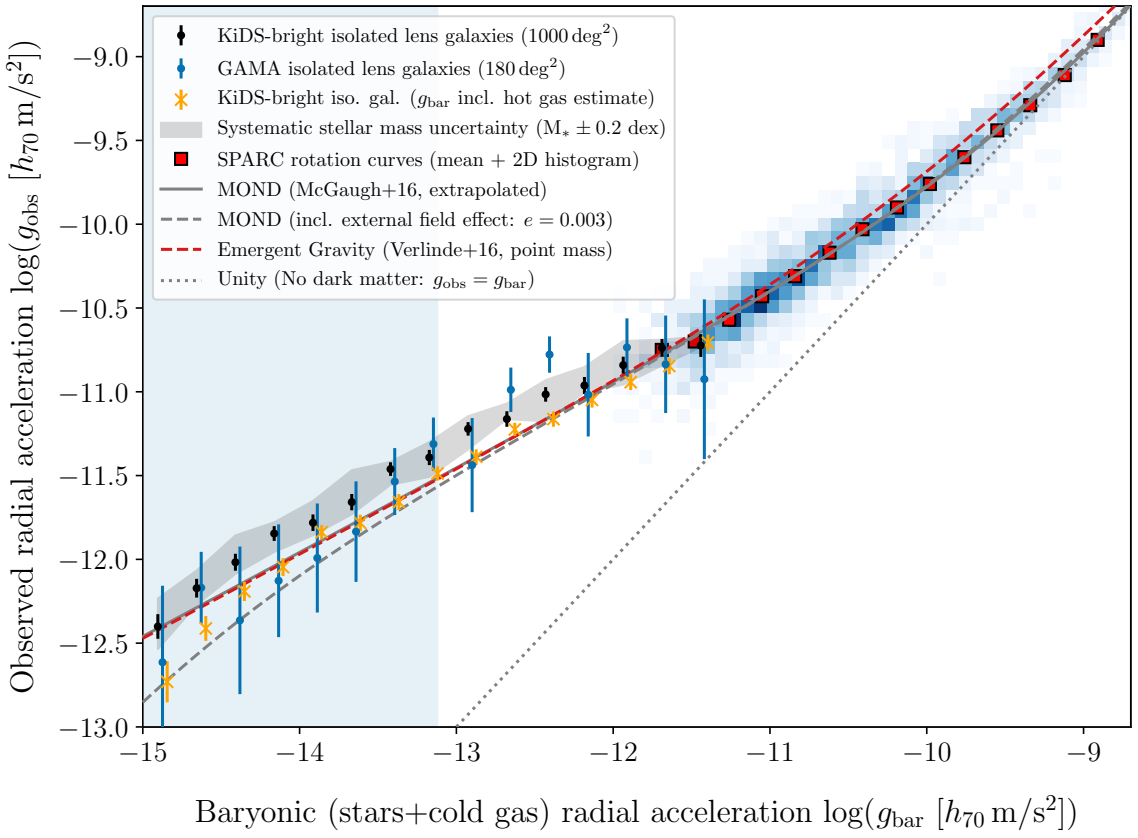

Fig. 4. Measured RAR, which compares the total gravitational acceleration $g_{\rm obs}$ with the expected baryonic acceleration $g_{\rm bar}$ of galaxies. At high accelerations we show the M16 RAR measurements from galaxy rotation curves (all data as the blue 2D histogram, the mean as red squares). Using weak gravitational lensing we were able to extend this measurement to lower accelerations, using both the spectroscopic GAMA and the photometric KiDS-bright isolated lens samples (blue and black points with 1σ error bars). Comparing our lensing observations to two MG models: MOND (the M16 fitting function; grey solid line) and EG (assuming a point mass; red dashed line) we find that GAMA results are in agreement with the two models, while those from KiDS-bright are systematically higher. At very low accelerations (corresponding to $R > 0.3\,h_{70}^{-1}$ Mpc, light blue shaded region) the uncertainty in the photometric KiDS redshifts affects the isolated lens selection, resulting in systematically higher values of $g_{\rm obs}$ due to the possible contribution of satellites. The results from the spectroscopic GAMA survey, however, are still reliable within this region. The impact of stellar mass uncertainty ($\Delta M_\star = 0.2$ dex) on the measurement is shown as the grey band. We show the MOND prediction including the EFE (with $e = 0.003$, see Eq. (13)) as the grey dashed line. In addition, we show the effect on the RAR of KiDS-bright galaxies if $g_{\rm bar}$ contained an additional isothermal hot gas contribution within a $100\,h_{70}^{-1}$ kpc radius, with a nominal gas mass equal to the stellar mass (orange crosses with 1σ error bars). We emphasise that this is only a rough order of magnitude estimate of the possible effect of gaseous haloes, which are extremely difficult to observe.

freedom $N_{\rm d.o.f.}$ of the model, which gives the reduced χ^2 statistic:

$$\chi^2_{\rm red} = \frac{\chi^2}{N_{\rm d.o.f.}} = \frac{\chi^2}{N_{\rm data} - N_{\rm param}}. \quad (27)$$

Here $N_{\rm data}$ is the number of data points in the measurement and $N_{\rm param}$ is the number of free parameters in the model. Since none of the models have free parameters, $N_{\rm d.o.f.}$ is simply the total number of $g_{\rm bar}$-bins (in this case $N_{\rm data} = 15$).

Comparing the GAMA data to the two MG models results in $\chi^2_{\rm red}$-values of 0.8 for both MOND and EG, corresponding to a standard deviation of 0.4σ. This confirms that both models agree well with the GAMA data. When using the KiDS-bright results, neither model provides a good description of the data with: $\chi^2_{\rm red} = 4.6$ and 5.0 for MOND and EG respectively, corresponding to ~ 6 standard deviations ($\sim 6\sigma$). Taking into account the effect of the photometric redshift uncertainty of KiDS-bright by only using the seven data points within the isolation criterion limit ($R < 3\,h_{70}^{-1}$ Mpc) we find: $\chi^2_{\rm red} = 4.0$ for MOND and $\chi^2_{\rm red} = 4.4$ for EG, $\sim 3.8\sigma$ away from a good fit. Considering the $\Delta M_\star = \pm 0.2$ dex uncertainty shown by the grey band (with the data points beyond the isolation criterion limit still removed) leads to $\chi^2_{\rm red} = 1.5$ for $\Delta M_\star = +0.2$ dex and $\chi^2_{\rm red} = 14$ for $\Delta M_\star = -0.2$ dex with respect to MOND, with similar results for EG. Thus, the MOND and EG predictions are able to describe our measurements within the statistical and systematic uncertainties. Whether these models are confirmed or excluded relies heavily on the systematic bias in the stellar mass measurements. This highlights the general point that GGL measurements are now so accurate in determining the total observed mass distribution that improving the RAR measurement primarily depends on obtaining better constraints on the baryonic mass distribution.

This point is highlighted further by the fact that we cannot incorporate measurements of the total baryonic mass distribution into our comparison, in particular those components that have not been detected, such as hot gaseous haloes and missing baryons. This remains a fundamental limitation of all work testing DM or MG theories at large scales (see Sect. 4.3). Although there have been very recent fruitful attempts at a first detection of this barely visible baryonic component (Macquart et al. 2020; Tanimura et al. 2020), there exist no accurate measurements of its distribution around isolated galaxies. However, we can safely continue as long as all estimates of $g_{\rm bar}$ (in the measurements, models and simulations) are based on the same components (in

our case: stars + cold gas). This way our RAR results remain purely observational, based on actual measurements along both axes.

However, a qualitative idea of the possible effect of an additional extended ionised gas component on $g_{\rm bar}$ is depicted in Fig. 4. In addition to our standard stars-and-cold-gas point mass used to calculate $g_{\rm bar}$, we modeled the hot gas as a simple isothermal density profile ($\rho(r) \propto r^{-2}$), truncated at the accretion radius $R_{\rm acc}$. Based on Valentijn (1988), we derived that $R_{\rm acc} \approx 100\,h_{70}^{-1}$ kpc for hot gas haloes around galaxies with $M_\star \approx 10^{11}\,h_{70}^{-2}\,M_\odot$. Finding an accurate estimate of the additional gas mass $M_{\rm gas}$ within this radius is no easy matter. Brouwer et al. (2017) assumed a total hot gas mass $M_{\rm gas} = 3M_\star$, based on results from the OWLS hydrodynamical simulations by Fedeli et al. (2014). They found that, in simulations with AGN feedback, OWLS galaxies with a total mass $M_{200} = 10^{12}\,h_{70}^{-1}\,M_\odot$ (corresponding to $M_\star \approx 10^{10}\,h_{70}^{-2}\,M_\odot$, a lower limit on the typical stellar masses in our sample) have a gas-to-stellar-mass fraction of $M_{\rm gas}/M_\star \approx 3$. One of the few observational scaling relations for hot gas is derived by Babyk et al. (2018), using *Chandra* X-ray observations of 94 early-type galaxies. In their Fig. 7, which shows the X-ray gas mass versus the total galaxy mass, galaxies with $M_{\rm tot} = 10^{12}\,M_\odot$ have gas fractions ranging from 0.1 to 1. However, Babyk et al. (2018) measured both $M_{\rm tot}$ and $M_{\rm gas}$ within 5 effective radii of their galaxies, which means that the hot gas fraction on larger scales could be as high as 3 in extreme cases. These relatively high hot gas masses motivated by the Babyk et al. (2018) observations are possibly biased towards a high X-ray surface brightness and are an order of magnitude higher than the hot gas masses presented in Fig. 7 of Tumlinson et al. (2017). As this gas mass outweighs the possible contribution of various cooler gas and dust components, this case provides a good guide for our evaluation. Based on all these considerations, we assumed a nominal gas-to-stellar-mass fraction of $M_\star/M_{\rm gas} = 1$, emphasising that this is only an order of magnitude estimate due to the challenging nature of observing circumgalactic gas.

In Fig. 4 we include the RAR of KiDS-bright galaxies with our nominal estimate of the hot gas distribution added to $g_{\rm bar}$ on the x-axis. At the highest accelerations measurable by lensing, we find that these results are almost indistinguishable from the original KiDS-bright measurements. As the acceleration decreases, the $g_{\rm bar}$ values including hot gas shift further to the right (higher values) due to the increased enclosed hot gas mass. This causes a steepening downward slope of the RAR, such that it finally diverges from the $g_{\rm obs} \propto \sqrt{g_{\rm bar}}$ relation at very low accelerations ($g_{\rm bar} < 10^{-14}\,{\rm m\,s^{-2}}$). The same effect is as also seen in the BAHAMAS results in Fig. 1. As expected, we find that this steepening of the RAR increases for higher assumed gaseous halo masses $M_{\rm gas}$, and decreases for lower values. This implies that, if gaseous haloes more massive than in our example ($M_{\rm gas} \gtrsim M_\star$) were detected directly and incorporated into the measurement, the observed RAR would diverge from the current MOND and EG predictions at low accelerations.

In the case of MOND a steep downward slope at low accelerations is not expected unless, despite our best efforts, our isolated galaxy sample is not truly isolated. In that case undetected satellites might cause an external field effect (EFE). To evaluate this effect we use the results of Chae et al. (2020) for the isolated SPARC galaxies. Based on their results, we have assumed $e = g_{\rm ext}/g_\dagger = 0.003$ as a reasonable estimate of the external gravitational acceleration $g_{\rm ext}$ compared to the critical acceleration scale $g_\dagger$ (see Sect. 2.3) for our isolated lenses. We use the fitting function in Eq. (13), which represents the EFE for an idealised model of galaxies within their environment, to depict the EFE on the predicted MOND RAR in Fig. 4. The extrapolated M16 fitting function represents the MOND prediction without any EFE ($e = 0$). As expected the MOND prediction including the EFE diverges from the one without, tending towards a steeper downward slope at low accelerations ($g_{\rm bar} < 10^{-12}\,{\rm m\,s^{-2}}$). Hence the EFE moves the MOND prediction away from our main observational result: the lensing RAR from the KiDS-bright sample without an estimate for the additional hot gas, which we explore throughout the rest of this work. We will therefore maintain the use of the M16 fitting function as our main MOND prediction since this represents the optimal case considering our observations. Regarding the KiDS-bright result including an estimate for the hot gas, it turns out that the steeper downward slope resulting from the MOND EFE is not steep enough to be consistent with our measured RAR including an estimate of the additional hot gas. This is illustrated by the fact that, for our chosen value $e = 0.003$, the MOND prediction including EFE and our RAR observation including hot gas reach the same value of $g_{\rm obs}$ at $g_{\rm bar} \approx 10^{-15}\,{\rm m\,s^{-2}}$. However, the observation reaches this depth within a much smaller span in $g_{\rm bar}$ ($-15 < \log_{10}(g_{\rm bar}/{\rm m\,s^{-2}}) < -14$). Choosing a different value for the EFE strength e does not solve this problem, and the effect becomes stronger for higher assumed values of $M_{\rm gas}$. It is therefore unlikely that the MOND EFE can explain the effect of massive ($M_{\rm gas} \gtrsim M_\star$) hot gaseous haloes, if such haloes are detected. In the case of EG it is not yet known whether and, if so, how external gravitational fields affect its prediction (Verlinde, priv. comm.).

5.3. The RAR of KiDS compared to ΛCDM simulations

In this section we compare the KiDS-1000 RAR with numerical ΛCDM simulations[15]. In order to obtain the predictions from these simulations, we applied the same isolation criterion, GGL procedures and RAR conversion to mock galaxy samples from the MICE and BAHAMAS simulations (see Sect. 4). In Fig. 5, BAHAMAS (orange band) is shown as the median result of all lens galaxies, with the upper and lower limit of the band representing the 16th and 84th percentiles. For MICE (red band) we show the result for isolated lenses selected using the true redshifts (lower limit) and using redshifts with a normally distributed random offset of $\sigma_z/(1+z) = 0.02$ (upper limit), in order to emulate the effect of the redshift uncertainty in KiDS on the isolated galaxy selection (see Appendix A). This means that the upper limit of the MICE prediction is considered reliable even at high accelerations (blue shaded region), where uncertainties in the galaxy isolation could affect the RAR measurement. The RAR observations are the same KiDS-bright lensing and M16 rotation curve results as shown in Fig. 4, this time compared to the predictions from the two simulations.

We find a good agreement between the MICE simulation and our measurements. The MICE measurements are limited to the low $g_{\rm bar}$ regime, owing to the resolution of the MICE simulations. The MICE scale limit of $R > 0.25\,h_{70}^{-1}$ Mpc is within the angular scale where satellites missed by the isolation criterion might impact the lensing signal ($R > 0.3\,h_{70}^{-1}$ Mpc, light blue shaded region). The effect of the KiDS-bright redshift uncertainty σ_z on the isolation criterion is however mimicked in the MICE simulation (upper limit of the red band), which means we can safely compare MICE with our low-acceleration measurements. The limited width of red band shows that this effect is relatively small (~30%). The MICE prediction (with the σ_z offset) results in a reduced χ^2 value of $\chi^2_{\rm red} = 2.3$, corresponding to 2.3σ.

[15] The first ΛCDM model we test is that of N17, but find that this simple analytical model is not sufficient to describe our data (see Appendix C).

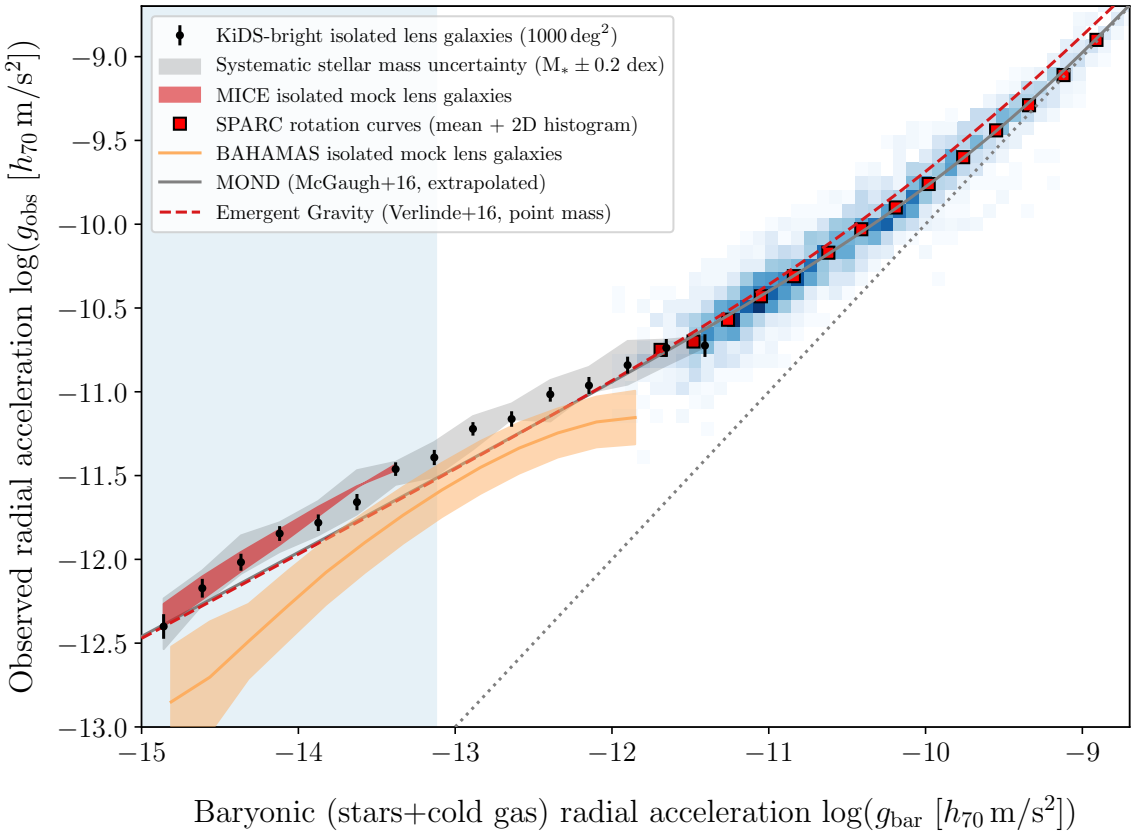

Fig. 5. Measured RAR of the KiDS-bright isolated lens sample (black points with 1σ error bars) compared to two ΛCDM simulations: MICE and BAHAMAS. The accelerations where uncertainty in the photometric KiDS redshifts affects the KiDS-bright isolated lens selection is indicated by the light blue shaded region. The MICE results (red band) emulate the effect of the redshift uncertainty in KiDS, while the BAHAMAS results (orange band) reflect the median and 16th and 84th percentiles of the simulated lens galaxies. The MICE simulation, though limited to low accelerations by its resolution, succeeds in reproducing the lensing data. The result from the BAHAMAS simulation runs approximately parallel to the MICE curve, but underestimates our measurement by 0.5 dex due to the biased SHMR of the BAHAMAS isolated galaxies (see Sect. 5.3).

Figure 5 shows poor agreement between the lensing RAR for isolated BAHAMAS galaxies and the KiDS measurement. The reason for this is straightforward to understand: the BAHAMAS measurement in Fig. 5 runs approximately parallel to both the KiDS and MICE curves, as a result of a constant offset in the stellar-to-halo-mass relation (SHMR) between BAHAMAS and MICE. Both simulations reproduce the observed SHMR in an overall sense, as shown in Fig. 6 of McCarthy et al. (2017) and Jakobs et al. (2018) for BAHAMAS, and guaranteed by construction as described in Carretero et al. (2015) for MICE. However, while in MICE our isolated galaxy sample follows essentially the same SHMR as the parent sample, in BAHAMAS isolated galaxies have, on average, triple the stellar mass at fixed halo mass compared to the global BAHAMAS galaxy population. This difference fully accounts for the 0.5 dex horizontal offset between the MICE and BAHAMAS curves in Fig. 5. The failure of BAHAMAS to reproduce the observed lensing RAR could therefore be regarded as a possible shortcoming of the galaxy formation model used in those simulations, rather than a general failure of their cosmological paradigm. However, we note that the offset in the SHMR as a function of local galaxy density is theoretically expected, and (indirectly) observed (e.g., Dutton et al. 2010; Correa & Schaye 2020). It is therefore curious that MICE, which does not reproduce this observed bias, turns out to be in reasonable agreement with our measurements. The discrepancy between KiDS-bright and BAHAMAS must therefore arise due to some more subtle underlying reason that we have yet to identify; we hope to follow this up in future work. We initially selected BAHAMAS for our analysis due to its large volume – required to produce enough of the rare isolated, relatively massive galaxies of interest – and readily available mock lensing data. It will be interesting to revisit the lensing RAR as cosmological hydrodynamical galaxy formation simulations continue to improve in terms of realism, simulated volume, and resolution.

5.4. The RAR for early- and late-type KiDS galaxies

The large size of the KiDS-bright lens sample gives us the opportunity to divide our lenses into different samples based on observed galaxy parameters. We determined the RAR for isolated galaxies split into two types based on either parameter: bulge-dominated and disc-dominated based on their Sérsic index, and red and blue based on their $u - r$ colour. Although these selections are far from perfect representations of true morphological types, the red and bulge-dominated samples can roughly be identified with canonically early-type (pressure supported) galaxies and the blue and disc-dominated samples with late-type (rotationally supported) galaxies (Driver et al. 2006)[16].

[16] In general, the Sérsic index n does not separate early- and late-type galaxies because dwarf early- and late-type galaxies have similar values of n (Graham 2019). However, dwarf early-type galaxies are not abundant in isolation (Janz et al. 2017), which means that our isolated low-n galaxy sample likely consists of late-type galaxies.

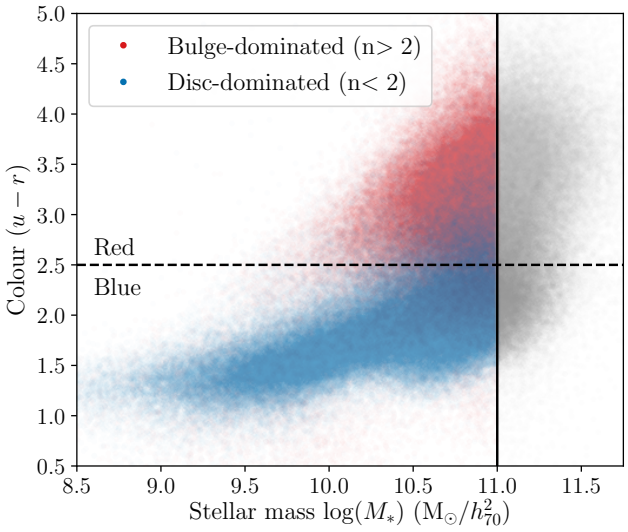

Fig. 6. 2D histogram of the $u-r$ colour and stellar mass of isolated KiDS-bright galaxies. We divide our galaxies into canonically early- and late-type galaxies, based on either Sérsic index n or $u-r$ magnitude. When dividing by Sérsic index, we define bulge-dominated (early-type) galaxies as those with $n > 2$ and disc-dominated (late-type) galaxies as those with $n < 2$ (red and blue points). When dividing colour we define red (early-type) galaxies as those with $m_u - m_r > 2.5$ and blue (late-type) galaxies as those with $m_u - m_r < 2.5$ (above and below the dashed horizontal line).

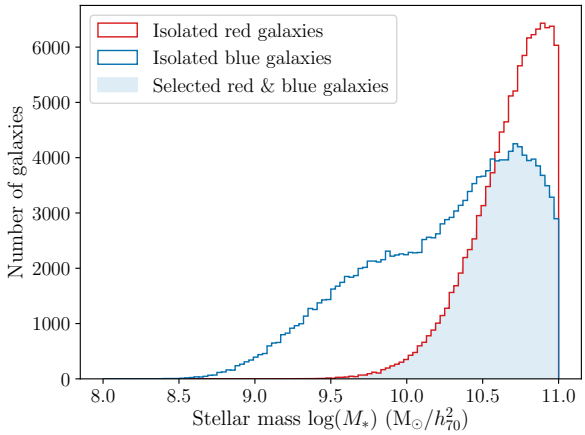

Fig. 7. Stellar mass histogram of the red (early-type) and blue (late-type) isolated KiDS-bright galaxies (red and blue lines), divided by $u-r$ colour ($m_u - m_r \lessgtr 2.5$ mag). To isolate the effect of galaxy type on the RAR from that of $M_\star$, we select two samples with the same stellar mass distribution by randomly removing galaxies from both samples until only the overlapping region (light blue shaded region) remains.

The r-band Sérsic indices n of all KiDS galaxies with S/N > 50 (following Roy et al. 2018) were measured using the 2DPHOT multi-purpose environment for 2D wide-field image analysis (La Barbera et al. 2008). For the colour split, we used the u and r magnitudes measured using the GAAP pipeline (see Sect. 3.1). In Fig. 6 the $u-r$ colour versus stellar mass distribution of isolated galaxies shows the split based on Sérsic index, which defines early-type galaxies as those with $n > 2$ and late-type disc-dominated galaxies as those with $n < 2$. Based on the $u-r$ magnitude distribution of these two populations, we defined our split by galaxy colour as follows: galaxies with $m_u - m_r > 2.5$ mag are defined as red, and those with $m_u - m_r < 2.5$ mag as blue.

In both cases, we aimed to select two samples with the same stellar mass distribution, in order to isolate any possible effect of galaxy type on the RAR from that of $M_\star$. In Fig. 7 we show the $M_\star$ histogram of the two types (in this case based on galaxy colour). From both samples, we removed galaxies until only the overlapping section of both mass distributions remained. Ideally this should give us two samples (red and blue galaxies) with equal stellar mass distributions, shown by the light shaded blue region.

Figure 8 shows the lensing RAR of equal-mass KiDS-bright galaxies split by Sérsic index (left panel) and $u-r$ colour (right panel). For this result, we focus on establishing whether there exists a significant difference between the RAR of the two types. Contrary to previous plots, the effect of a 0.2 dex global systematic bias in $M_\star$ (normally shown by a grey band) is omitted because this affects both measurements in the same way such that their relative difference does not change (the possibility of a colour- or Sérsic index-dependent $M_\star$ bias is discussed below).

We indeed observe a significant difference between the RAR measurements of early and late galaxy types. To quantify this difference, we measured the reduced χ^2 between the RAR measurements by replacing $g_{\rm obs}$ and $g_{\rm mod}$ in Eq. (26) with $g_{\rm obs,E}$ and $g_{\rm obs,L}$ from the early-type (red or bulge-dominated) and late-type (blue or disc-dominated) galaxy samples. The $\chi^2_{\rm red}$ equals $67.8/15 = 4.5$ for the lenses split by Sérsic index, and $134.2/15 = 8.9$ for those split by $u-r$ colour. Taking the full covariance matrix into account we find that even the Sérsic index split, which displays the smallest offset, results in RAR difference with a 5.7σ significance. The mean ratio between the RAR measurements of the two types, $\log_{10}(\delta g_{\rm obs}^{\rm E/L}) = \log_{10}(\langle g_{\rm obs,E}/g_{\rm obs,L}\rangle)$, is 0.17 dex and 0.27 dex for the Sérsic and colour splits respectively.

We address the question whether the observed difference of the RAR between early and late types could be caused by any bias in the stellar mass. To this end, we estimated the systematic stellar mass bias between the two types, defined as $\log_{10}(\delta M_\star^{\rm E/L}) = \log_{10}(\langle M_\star\rangle_{\rm E}/\langle M_\star\rangle_{\rm L})$, that would be required to resolve the difference between their two RAR measurements. When trying to estimate the effect of this bias on the RAR, we had to take into account that $\delta M_\star^{\rm E/L}$ affects both the estimated acceleration from baryonic mass $g_{\rm bar}$ (directly) and the observed acceleration $g_{\rm obs}$ (indirectly, through the equal-mass selection). The bias in baryonic acceleration scales linearly with the bias in $M_\star$, such that: $\log_{10}(\delta g_{\rm bar}^{\rm E/L}) = \log_{10}(\delta M_\star^{\rm E/L})$. Throughout this work, the observed relation between $g_{\rm bar}$ and $g_{\rm obs}$ at the scales measured by lensing has approximately followed $g_{\rm obs} \propto \sqrt{g_{\rm bar}}$. This means that we can roughly estimate the effect on $g_{\rm obs}$ as: $\log_{10}(\delta g_{\rm obs}^{\rm E/L}) \approx \log_{10}(\delta M_\star^{\rm E/L})/2$. Since our measured difference $\delta g_{\rm obs} \gtrsim 0.2$ dex, this means $\log_{10}(\delta M_\star^{\rm E/L})$ should be $\gtrsim 2\log_{10}(\delta g_{\rm obs}^{\rm E/L}) = 0.4$ dex. That is, the observed difference could be resolved by a systematic stellar mass bias between the two types $\gtrsim 0.4$ dex. We will now discuss different sources of a possible systematic bias, and estimate whether they could be the cause of the observed difference.

First, the statistical uncertainty in the $M_\star$ measurements could cause a systematic shift in the two $M_\star$ distributions resulting from Eddington bias (Eddington 1913). We estimated the size of this bias by adding a random offset to the true $\log_{10}(M_\star)$ measurements of KiDS-bright before selecting the two 'equal' stellar mass distributions for red and blue galaxies. Based on

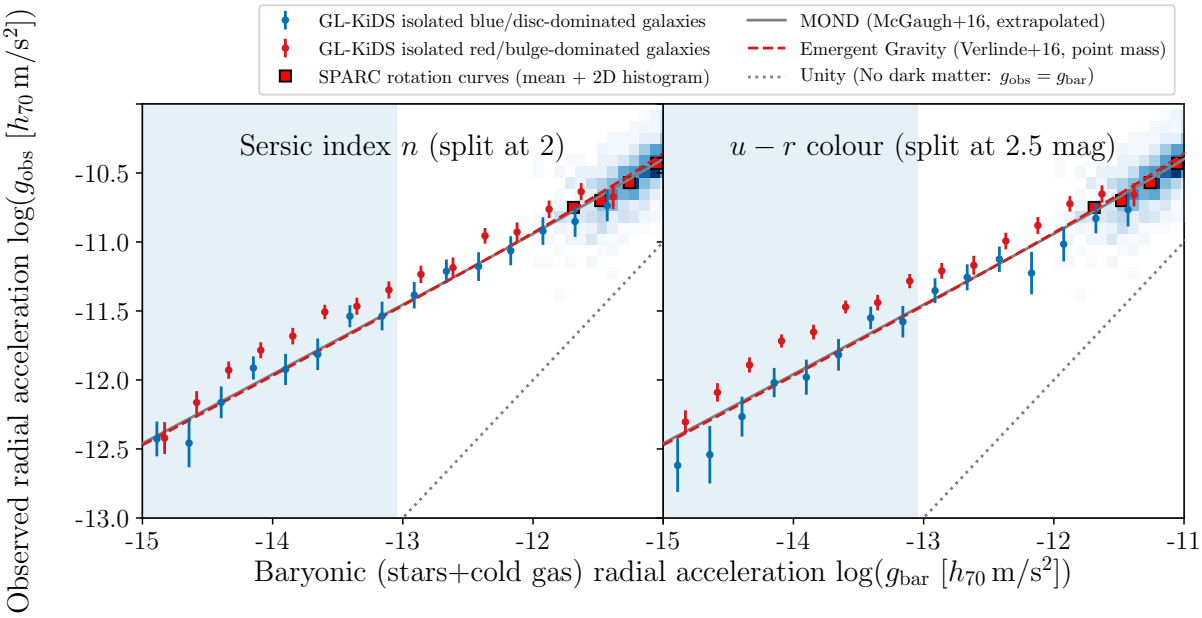

Fig. 8. Measured RAR of the KiDS-bright isolated lenses (points with 1σ error bars) divided into canonically early- and late-type galaxies. In the *left panel*, the lenses are split by Sérsic index ($n \gtrless 2$) into bulge-dominated (red points) and disc-dominated (blue points) galaxies. In the *right panel* they are split by $u - r$ colour ($m_u - m_r \gtrless 2.5$) into red and blue galaxies (with correspondingly coloured points). In both panels we find a significant difference between the RAR measurements of early and late galaxy types. The extrapolated MOND and EG predictions (grey solid and red dashed lines) and the SPARC data (red squares with 2D histogram) are shown as a reference.

our estimate of the statistical uncertainty in the KiDS-bright $M_\star$ (see Sect. 3.3), we drew the random offsets from a log-normal distribution with $\sigma = 0.12$ dex. When looking at the underlying true stellar mass distributions we found that they are indeed not equal, but that the mean stellar masses $\langle M_{\star,\mathrm{E}}\rangle$ and $\langle M_{\star,\mathrm{L}}\rangle$ of the red and blue samples differ by only 0.025 dex. Of course, this method overlooks the fact that the measured $M_\star$ distribution already contains scatter, and is therefore not the true $M_\star$ distribution. Indeed when we apply the random offset multiple times, we see the Eddington bias decrease by ∼5% after every iteration. Therefore, the true Eddington bias is likely to be slightly larger, around 0.027 dex. This is still very small compared to the ≳0.4 dex bias needed, thus it is very unlikely that the difference we observe is caused exclusively by Eddington bias.

Second, there could be systematic errors in the KiDS-bright $M_\star$ measurements that differ between red and blue galaxies (due to e.g., systematic variation of the IMF, SPS model inaccuracies, or systematic errors in the measured redshifts or magnitudes). In order to estimate the size of any systematic biases in the stellar mass, we compared KiDS-bright's $M_{\star,\mathrm{ANN}}$ with GAMA's $M_{\star,\mathrm{G}}$ of exactly the same galaxies. Here $M_{\star,\mathrm{ANN}}$ is based on the nine-band KiDS+VIKING photometry and photometric redshifts z_{ANN} derived by training the ANNz2 (Artificial Neural Network) machine learning method on the spectroscopic GAMA redshifts (see Sect. 3.3), while $M_{\star,\mathrm{G}}$ is based on the *ugrizZY* SDSS+VIKING photometry combined with the spectroscopic GAMA redshifts (see Sect. 3.2). After selecting our samples of blue and red galaxies with the same $M_{\star,\mathrm{ANN}}$ distribution as described above, we indeed found that the $M_{\star,\mathrm{G}}$ distributions are not exactly equal: $\langle M_\star\rangle_\mathrm{E}/\langle M_\star\rangle_\mathrm{L} = 1.4$, corresponding to 0.14 dex. This indicates that using different sets of observations and models to measure $M_\star$ can cause a systematic bias between red and blue galaxies, but that this effect is too small to reach the ≳0.4 dex difference in $M_\star$ needed to explain the ≳0.2 dex difference in the measured RAR.

In conclusion, even when combined the Eddington plus overall systematic measurement bias is at most 0.17 dex, not even half of what is needed. We note that this bias estimation has been carried out using the types split by $u - r$ colour; when split by Sérsic index, the Eddington and other systematic biases between bulge- and disc-dominated galaxies are even smaller (0.021 and 0.12 dex respectively).

Domínguez Sánchez et al. (2019) reported evidence of a varying IMF in massive early-type galaxies. As seen in Fig. 19 of their work, this could cause the global mass-to-light-ratio of these galaxies to increase by as much as 0.09 dex compared to a fixed Chabrier IMF. They find this effect only for their high-mass galaxy sample with a stellar mass of at least $M_\star > 2\times10^{11}\,M_\odot$, and not for their lower-mass sample. Since we limit all our galaxies to $M_\star < 10^{11}\,h_{70}^{-2}\,M_\odot$ (see Sect. 3.3), the varying IMF is not likely to apply to our early-type galaxy sample. However, even if this had been the case, this 0.09 dex difference in $M_\star$ is small compared to the ≳0.4 dex needed to explain the difference in the RAR of early- and late-type galaxies.

The higher values of g_{obs} for red and bulge-dominated galaxies that we find in Fig. 8 are in qualitative agreement with earlier GGL studies. A recent KiDS-1000 lensing study by Taylor et al. (2020) found that, within a narrow stellar mass range near the knee of the SHMR ($M_\star \sim 2-5\times10^{10}\,h_{70}^{-2}\,M_\odot$), galaxy halo mass varied with galaxy colour, specific star formation rate (SSFR), effective radius R_e and Sérsic index n. Although not explicitly mentioned, their Figs. 1 and 6 reveal that their early-type (red, low-SSFR) galaxies have larger halo masses than their late-type (blue, low-n, high-SSFR) galaxies of the same stellar mass. Sérsic parameter coupling between n and R_e, for a fixed galaxy luminosity, may also contribute towards the trends seen among the early-type galaxies in their M_{halo}–n and M_{halo}–R_e diagrams[17]. Much earlier Hoekstra et al. (2005) measured the GGL signal of a sample of 'isolated' Red-sequence Cluster Survey galaxies as

[17] The smaller average size for the early-type galaxies, compared to the late-type galaxies, is because of the different 3D-bulge-to-2D-disc ratios: a fixed stellar mass will fit into a smaller volume if distributed in a bulge rather than a disc.

a function of their rest-frame B-, V-, and R-band luminosity, and found that early-type galaxies have lower stellar mass fractions. In contrast, Mandelbaum et al. (2006) found no dependence of the halo mass on morphology for a given stellar mass below $M_\star < 10^{11}\,M_\odot$, although they did find a factor of two difference in halo mass between ellipticals and spirals at fixed luminosity.

Finding a significantly different RAR at equal $M_\star$ would have interesting implications for galaxy formation models in the ΛCDM framework. In the ΛCDM framework it is expected that the galaxy-to-halo-mass relation, and therefore the RAR, can be different for different galaxy types through their galaxy formation history (Dutton et al. 2010; Matthee et al. 2017; Posti et al. 2019; Marasco et al. 2020). Two parameters that correlate heavily with galaxy formation history are Sérsic index and colour.

Current MG theories do not predict any effect of galaxy morphological type on the RAR, at least on large scales. The MOND paradigm gives a fixed prediction for the relation between $g_{\rm bar}$ and $g_{\rm obs}$ given by Eq. (11). Since the RAR is the observation of exactly this relation, in principle MOND gives a fixed prediction, independent of any galaxy characteristic. As discussed in Sect. 2.3, the main exception is the EFE that could be caused by neighbouring mass distributions. However, Fig. 4 shows that an increase in the EFE only predicts an increase in steepness of the downward RAR slope at low accelerations ($g_{\rm bar} < 10^{-12}\,{\rm m\,s^{-2}}$), while the observed RAR of both early- and late-type galaxies follow approximately the same slope across all measured accelerations. It is therefore unlikely that their amplitude difference can be explained through the EFE.

We will next discuss whether the observed difference in RAR between early and late types is at odds with EG, but first emphasise three caveats of this discussion.

First, the derivation of the EG formalism assumes a spherical mass distribution. Solutions for non-spherical systems do not exist yet. It is not excluded that solutions for large-scale triaxial ellipticals will differ from rotationally supported spiral galaxies. This requires further theoretical study.

Second, the current EG theory predicts ADM fields based exclusively on the static baryonic mass distribution, although very large-scale dynamics can potentially influence the excess gravitational force predicted by EG. It is unknown whether large-scale pressure supported (virialised) systems create an ADM distribution similar to that of rotationally supported galaxies.

Third, we assume here that, to first order, the uncertainty in the KiDS photometric redshifts affects the isolated galaxy selection of both galaxy types in the same way, allowing us to include the full acceleration range into our comparison. However, the well established morphology-density relation predicts a higher density of satellite and dwarf galaxies around early-type galaxies compared to the late types (Dressler 1980; Goto et al. 2003), although we have minimised this effect by selecting isolated galaxies (see Appendix A). It is not yet known whether and, if so, how these external gravitational fields affect the EG prediction.

To address this last caveat, the light blue shaded region in Fig. 8 shows the acceleration scales beyond the KiDS isolation criterion limit ($g_{\rm bar} < 10^{-13}\,{\rm m\,s^{-2}}$), where the presence of satellites might play a role (see Appendix A). But even when we remove all data points inside this region, we obtain a difference $\log_{10}(\delta g_{\rm obs}^{\rm E/L})$ of 0.14 dex and 0.19 dex for the Sérsic and colour split respectively, where the latter has a significance of 3.2σ. Therefore, even at the scales where isolation is certain (corresponding to $R < 0.3\,h_{70}^{-1}$ Mpc), the difference remains significant.

To evaluate the possible effect of circumgalactic hot gas, we computed the RAR of early and late-type isolated galaxies (of the same stellar mass) while including a rough estimate of the hot gas contribution to $g_{\rm bar}$. We used the same model of the nominal hot gas distribution around our galaxies as discussed in Sect. 5.2: an isothermal halo within $100\,h_{70}^{-1}$ kpc, with a mass $M_{\rm gas} = M_\star$. When applying the same hot gas model to both early- and late-type galaxies, we find that there remains a $>6\sigma$ difference between their RARs, both for the split by Sérsic index and $u-r$ colour. However, for this particular gas model, we find that $g_{\rm bar}$ increases in such a way that the RAR of early-type galaxies moves to the right, close to the MG predictions where the RAR of late-type galaxies without circumgalactic gas resides. This means that, in the specific case where early-type galaxies have gaseous haloes with $M_{\rm gas} = M_\star$ while late-type galaxies (of the same stellar mass) have negligible hot circumstellar gas, this would reduce the difference in their RARs to $\sim 4\sigma$. Fine-tuning the $M_{\rm gas}/M_\star$ ratio of early-type galaxies to a slightly higher value, while keeping $M_{\rm gas}/M_\star \approx 0$ for late types, might remove the difference between their RARs. However, as discussed in Sect. 5.2, unbiased X-ray surveys of circumgalactic gas around isolated galaxies are still lacking, which makes it difficult to obtain representative observational data.

In conclusion, unless early-type galaxies have significant circumgalactic gaseous haloes while late types (of the same stellar mass) do not, the difference we find in the RARs of different galaxy types might prove difficult to explain within MG frameworks. In MOND, $g_{\rm bar}$ and $g_{\rm obs}$ should be directly linked through Eq. (11) without any dependence on galaxy type. In EG the effect might be a consequence of yet unexplored aspects of the theory, such as a non-symmetric mass distribution or the effect of large-scale dynamics. To explore whether this is the case, however, more theoretical work is needed. Through the derivative in Eq. (14), EG does include a dependence on the slope of the baryonic density distribution. A shallower slope of $M_{\rm bar}(r)$ increases $M_{\rm ADM}$ and thus $g_{\rm obs}$, which might solve the current tension if early-type galaxies have significantly shallower baryonic mass distributions that extend far beyond $30\,h_{70}^{-1}$ kpc, such as gaseous haloes (although Brouwer et al. 2017 did not find evidence for a significant effect of the baryonic mass distribution on the EG prediction; see their Sect. 4.3). In addition, EG is currently only formulated for spherically symmetric systems. It would be interesting to investigate whether discs and spheroidal galaxies yield different predictions, and whether these differences would extend beyond $30\,h_{70}^{-1}$ kpc.

In a ΛCDM context, our findings would point to a difference in the SHMR for different galaxy types. Recently Correa & Schaye (2020) used SDSS data with morphological classifications from Galaxy Zoo to find that, at fixed halo mass (in the range $10^{11.7}$–$10^{12.9}\,M_\odot$), the median stellar mass of SDSS disc galaxies was a factor of 1.4 higher than that of ellipticals. They found this to be in agreement with the EAGLE simulations, where haloes hosting disc galaxies are assembled earlier than those hosting ellipticals, therefore having more time for gas accretion and star formation.

5.5. The RAR as a function of stellar mass

In addition to splitting by galaxy type, it is interesting to create the RAR for galaxy samples with different stellar mass $M_\star$ (including very low-mass galaxies, 'dwarfs', in Sect. 5.6). In the ΛCDM paradigm, where baryonic and dark matter are described as separate substances, there can in theory be a difference in the SHMR depending on galaxy observables such as stellar mass, which could cause a shift in the measured RAR. This is in contrast with most MG models, which predict a fixed RAR (as is the

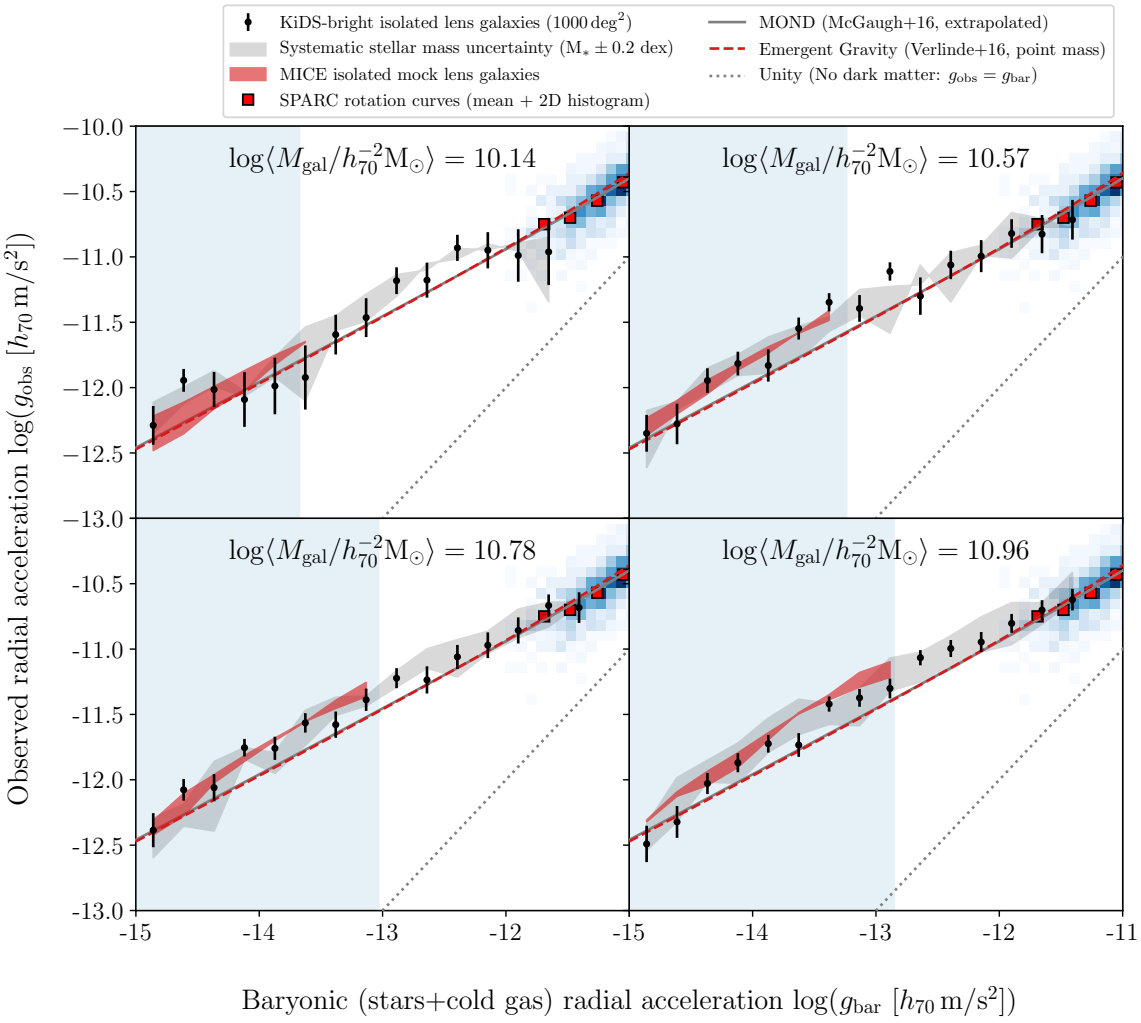

Fig. 9. Measured RAR of isolated KiDS-bright lenses (black points with 1σ error bars) divided into four stellar mass bins. The mean galaxy mass (stars + cold gas) of the lenses is shown at the top of each panel. At increasing stellar mass, the measurements seem to rise above the predictions from MOND (grey solid line) and EG (red dashed line). However, at scales larger than $R > 0.3\,h_{70}^{-1}$ Mpc (light blue shaded region) this could be caused by false positives in the isolated galaxy sample due to the KiDS-bright redshift uncertainty.

case for MOND, and for EG at scales beyond the galaxy disc). In this section, we separated our isolated KiDS-bright lenses into four samples based on $M_\star$. We selected our $M_\star$-bins to obtain a similar S/N of the lensing signal in each bin, resulting in the following limits: $\log_{10}(M_\star/h_{70}^{-2}\,M_\odot) = [8.5, 10.3, 10.6, 10.8, 11.0]$.

Figure 9 shows the lensing measurements and predictions for isolated galaxies split in four stellar mass bins. For each bin the mean galaxy mass (stars + cold gas) of the lenses, $\log_{10}\langle M_{\rm gal}/h_{70}^{-2}\,M_\odot\rangle = [10.14, 10.57, 10.78, 10.96]$, is shown at the top of the panel. Quantifying the difference between MOND (the extended M16 fitting function) and our measurement at all scales results in: $\chi^2_{\rm red} = 117.0/60 = 1.9$, which (noting that the prediction for EG is very similar) excludes both models at the $\sim 4.5\sigma$ level. This result should be taken with caution, however, as at accelerations $g_{\rm bar}$ that correspond to scales larger than $R > 0.3\,h_{70}^{-1}$ Mpc (light blue shaded region) an increasing signal is to be expected since at these distances satellite galaxies missed by our isolation criterion might affect the measurement. Galaxies with higher stellar masses reside in denser neighbourhoods, and therefore tend to have more satellites (see e.g., Baldry et al. 2006; Bolzonella et al. 2010; Brouwer et al. 2016).

The reduced χ^2 values using only the data within $R < 0.3\,h_{70}^{-1}$ Mpc are $\chi^2_{\rm red} = 49.9/31 = 1.6$ for MOND and $51.7/31 = 1.7$ for EG respectively (corresponding to a standard deviation of 2.4 and 2.5σ). Considering the stellar mass uncertainty ($\Delta M_\star = \pm 0.2$ dex), which, if it acts to reduce the observed RAR, results in $\chi^2_{\rm red} = 0.97$ for the extended M16 fitting function (with similar results for EG): a good fit. If the stellar mass uncertainty increases the observed RAR, we find $\chi^2_{\rm red} = 4.6$: a poor fit. This again highlights the grave importance of accurate baryonic mass measurements in determining the RAR, in addition to deep lensing surveys that can detect satellites down to very faint magnitudes. This could be achieved by future cosmology telescopes such as *Euclid* (Laureijs et al. 2011) and *Vera C. Rubin* Observatory, previously called Large Synoptic Survey Telescope (LSST; Dark Energy Science Collaboration 2012). As for the MICE simulation, it matches our measurements reasonably well in every $M_\star$ bin. For the result that includes the photometric redshift uncertainty σ_z in the isolated galaxy selection, we find $\chi^2_{\rm red} = 49.7/30 = 1.7$ (2.5σ).

5.6. The RAR of low-mass (dwarf) late-type galaxies

As a final exploration of different galaxy masses, we attempt to measure the RAR for the lightest lenses in KiDS-bright. Low-mass galaxies are of particular interest to DM and MG

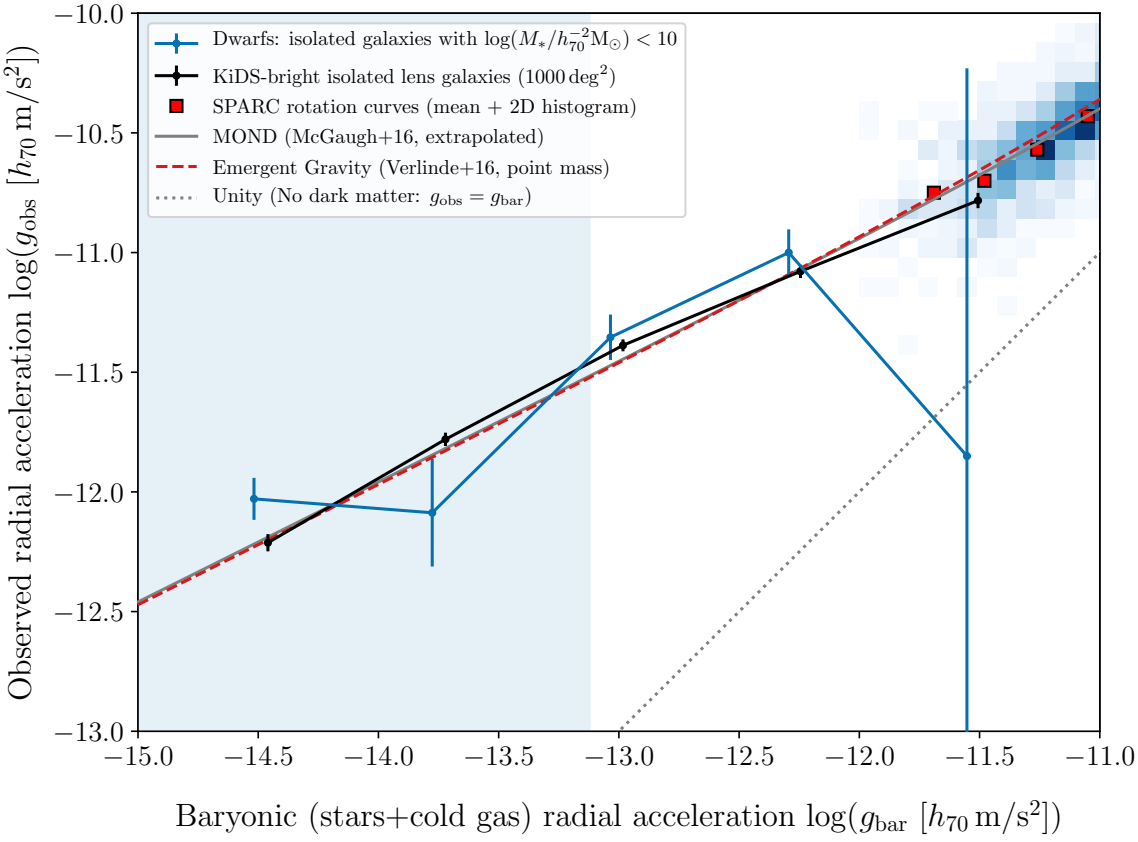

Fig. 10. Measured RAR of KiDS-bright lenses (points with 1σ error bars), respectively for isolated dwarfs ($\log(M_\star/h_{70}^{-2}\,M_\odot) < 10$, blue) and the full isolated galaxy sample ($\log(M_\star/h_{70}^{-2}\,M_\odot) < 11$, black). Due to the low S/N of the dwarf lensing signal, the number of $g_{\rm bar}$-bins is reduced from 15 to 5. We find that the RAR of dwarfs is consistent with that of our regular sample, and with the extrapolated MOND and EG predictions (grey solid and red dashed lines), which are shown as a reference.

researchers as extreme examples that might show eccentric behaviour (e.g., Oman et al. 2016; van Dokkum et al. 2018; Guo et al. 2019), as well as those who attempt to extend the RAR to lower accelerations using galaxy rotation curves (Lelli et al. 2017b; Di Paolo et al. 2019). We therefore select a sample of dwarfs: isolated galaxies with a stellar mass $M_\star < 10^{10}\,h_{70}^{-2}\,M_\odot$ (whereas the full sample of isolated galaxies has $M_\star < 10^{11}\,h_{70}^{-2}\,M_\odot$, see Sect. 3.3). As can be seen in Fig. 6, this sample is dominated by blue, disc-dominated galaxies based on their colours and Sérsic indices ($m_u - m_r > 2.5$ mag and $n < 2$), which means they are likely to be late-type. Since these galaxies are few, and have an even smaller effect on the path of light rays than more massive ones, we needed to reduce the number of bins in $g_{\rm bar}$ from 15 to 5 to obtain sufficient S/N in each bin. Figure 10 shows the resulting RAR measurement of dwarfs compared to the full isolated sample. We do not show the effect of the $\Delta M_\star = \pm 0.2$ dex systematic uncertainty because this would affect both results in the same way. We find that, within its large error bars, the RAR of the dwarfs is consistent with that of the full isolated sample; they both approximately follow the $g_{\rm obs} \propto \sqrt{g_{\rm bar}}$ relation expected by the extended MOND and EG predictions, which are shown as a reference. Hence, we do not find a significant difference in the RAR of dwarf galaxies.

6. Discussion and conclusions

Galaxy-galaxy lensing observations from the fourth data release of the Kilo Degree Survey (KiDS-1000) have extended the RAR of isolated galaxies by nearly 2 orders of magnitude in gravitational acceleration $g_{\rm obs}$, compared to previous measurements based on rotation curves (most notably McGaugh et al. 2016, M16). To compute the lensing RAR, we converted our ESD profiles $\Delta\Sigma(R)$ into the observed gravitational acceleration $g_{\rm obs}$, and our galaxy masses (measured using nine-band KiDS+VIKING photometry) into $g_{\rm bar}$. These measurements allowed us to perform unprecedented tests of two MG models: MOND and EG, as well as tests of DM using the MICE (N-body + semi-analytic) and BAHAMAS (hydrodynamical) simulations. Our conclusions from these observational tests are as follows:

- Figure 3: We find that lensing rotation curves of isolated galaxies, as inferred from GGL measurements, remain approximately flat at scales far beyond the visible disc ($0.03 < R < 3\,h_{70}^{-1}$ Mpc). At the accelerations corresponding to the outskirts of observable galaxies ($R \approx 30\,h_{70}^{-1}$ kpc), our lensing results are in excellent agreement with the SPARC rotation curves (Lelli et al. 2016). These two measurements are obtained by two very different methods, providing independent corroboration of each result.
- Figure 4: At the low accelerations corresponding to GGL scales, the lensing RAR of isolated galaxies approximately follows a $g_{\rm obs} \propto \sqrt{g_{\rm bar}}$ relation. This is in agreement with the expectations from EG (Eq. (17)) and MOND (which we take to be the M16 fitting function, Eq. (11), extrapolated to larger scales). At low accelerations both these models predict a direct relation between observed and baryonic acceleration of this form, with a very similar proportionality

constant[18] of $\sim 1.2 \times 10^{-10}$ m s^{-2}. This reinforces the results of Brouwer et al. (2017), who found that EG provides a good description of ESD profiles measured using 180 deg^2 of KiDS-GAMA data, but with a five times larger survey area. However, this result only remains valid if no massive ($M_{\rm gas} \gtrsim M_\star$) extended baryon distributions, such as as-yet undetected gaseous haloes, are common around our isolated lens galaxies.

– Figure 5: We find that the BAHAMAS simulation underestimates our KiDS-bright lensing RAR. The discrepancy relative to MICE is caused by a bias in the stellar-to-halo-mass-relation (SHMR) of isolated galaxies in BAHAMAS, which is absent in MICE: BAHAMAS galaxies have stellar masses typically three times higher at fixed halo mass than their non-isolated counterparts. Determining which of the two models more accurately captures the true SHMR is clearly crucial to the interpretation of our measurements in the ΛCDM context. Interestingly, the BAHAMAS RAR still has approximately the correct low-acceleration slope, rather than a steeper slope as would naively be predicted based on the $\rho \propto r^{-3}$ outer slopes of the simulated DM haloes. The prediction from MICE (only feasible at low accelerations due to the limited resolution of the simulated lensing measurements) matches our RAR measurements very well.
– The additional lensing power at large radii with respect to the prediction from Navarro et al. (2017, see Appendix C) might be caused by large-scale structure along the line-of-sight to the source, in spite of our efforts to select isolated galaxies. This highlights the crucial importance of simulating the entire measurement process (where possible) when making theoretical predictions, both in ΛCDM and MG, before they can be ruled out. In addition, the need for accurate isolated galaxy selection highlights the importance of large spectroscopic surveys, such as the upcoming 4MOST (de Jong et al. 2019) and Dark Energy Spectroscopic Instrument (DESI; Ruiz-Macias et al. 2021) surveys.
– Figure 8: When we split galaxies into two types based on Sérsic index or $u - r$ colour, we find at least a factor of 1.5 (≃0.2 dex) difference between the respective lensing RAR measurements with a significance of at least 5.7σ. This observed difference could be resolved by a $\gtrsim 0.4$ dex systematic bias between the stellar masses of the two types. However, we calculated that the expected $M_\star$ bias (due to Eddington bias or systematic biases in the $M_\star$ measurement) is at most 0.17 dex. This variation in the RAR based on galaxy type, which is in agreement with Taylor et al. (2020) and Correa & Schaye (2020), could be difficult to explain for MG models that predict a fixed relation between baryonic mass and the total gravitational potential.
– Figure 9: The lensing RAR for galaxy samples split by stellar mass $M_\star$ demonstrated a slight upward trend, away from the fixed predictions of MOND and EG, with increasing $M_\star$. This could be caused by satellite or companion galaxies missed by the isolated galaxy selection due to the KiDS-bright redshift uncertainty, however. With the inclusion of the KiDS isolation criterion limit and accounting for uncertainty in the stellar mass, we find a reasonable agreement between the MG models and observations. This highlights the crucial importance of accurate baryonic mass measurements in determining the RAR, in addition to deep lensing surveys that can detect satellites to down to very faint magnitudes (such as the future *Euclid* space telescope and *Vera C. Rubin* Observatory). The MICE prediction, which is corrected for the KiDS-bright redshift uncertainty, again matches well to our data.
– Figure 10: We find no significantly different RAR, relative to the entire isolated lens sample, for a subsample of the lightest KiDS-bright lenses: isolated dwarf ($M_\star < 10^{10}\, h_{70}^{-2}\, M_\odot$) galaxies.
– Throughout this work, we find that the field of GGL has reached a level of accuracy in the measurement of $g_{\rm obs}$ greater than that of the baryonic acceleration $g_{\rm bar}$. The fact that we have no accurate measurements of the additional hot gas at large radii, and the ambiguity around the cosmological missing baryons, forces us to limit $g_{\rm bar}$ to the contributions of stars and cold gas. In addition, the current 0.2 dex systematic uncertainty in $M_\star$ prevents us from definitively excluding any of the models we test. This shows that, if we want to have any hope of testing DM and MG models using the next generation of cosmological lensing surveys (such as *Euclid* and LSST), we also need to focus on the models and observations needed to accurately measure the baryonic mass distribution in and around galaxies.

We find that galaxy lensing rotation curves continue approximately flat out to $R = 3\, h_{70}^{-1}$ Mpc (where observations are bound to encounter lensing due to surrounding galaxies), which is difficult to explain in a ΛCDM framework that predicts simple NFW-like haloes because of their r^{-3} outer slope (see the N17 model in Appendix C). However, our analysis of the MICE and BAHAMAS simulations shows that the combination of the lenses and the additional structure along the line-of-sight can yield an ESD profile consistent with an $\sim r^{-2}$ density profile for isolated galaxies, even though the lenses have an intrinsic $\sim r^{-3}$ outer profile.

Throughout our analysis we find that the extrapolated M16 fitting function (Eq. (11)), which approximately corresponds to the prediction of both MG models (EG and MOND), holds to scales of $3\, h_{70}^{-1}$ Mpc for isolated galaxies. A fundamental limitation of this measurement is that the additional diffuse gas surrounding galaxies remains difficult to measure, and has therefore not been included in most of this study. By implementing a rough order of magnitude estimate of the hot gas contribution to $g_{\rm bar}$, an isothermal distribution with $M_{\rm gas} = M_\star$ within $100\, h_{70}^{-1}$ kpc, we found that this causes an overall downward shift of the RAR and a steeper downward slope at very low accelerations (see Fig. 4, and also Fig. 1 for a broader discussion of missing baryons). Although the MOND external field effect (EFE) causes a similar steepening of the RAR, we find that the idealised EFE fitting function of Chae et al. (2020) is not steep enough the explain the effect of gaseous haloes. Therefore, a convincing detection of additional gaseous components with a nominal mass of $M_{\rm gas} \gtrsim M_\star$ would move the observed RAR away from the MG predictions ($g_{\rm bar} \propto \sqrt{g_{\rm obs}}$) at very low accelerations ($g_{\rm bar} < 10^{-13}$ m s^{-2}) and towards the DM predictions (where $g_{\rm bar}$ and $g_{\rm obs}$ are independent). A robust non-detection of such massive gaseous haloes in general would likely strengthen the position of MG models. Finding them for early-type galaxies only would reduce the difference between the RAR of early- and late-type galaxies, which otherwise remains unexplained in MG frameworks.

In conclusion, we find that the lensing RAR is a promising method to be used by future cosmological surveys to distinguish between MG and DM models. This can be done by

[18] The proportionality constant $cH_0/6$ in EG is almost equal to the value of $g_\dagger$ found by M16, which is again equal to the $a_0 = 1.2 \times 10^{-10}$ m s^{-2} canonical in MOND.

measuring the RAR including large-scale baryonic mass observations; by simply performing the same comparison with even more accurate lensing and stellar mass measurements; or by further exploring the offset that we have found between the RARs of different galaxy types. All these options require that systematic biases in the stellar and other baryonic mass measurements be reduced.

Acknowledgements. Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018 and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples). GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey 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 www.gama-survey.org. We are indebted to Ian McCarthy, who provided the BAHAMAS data products used in our analysis. Bob Sanders provided us useful comments about the relation between MOND and the M16 fitting function, and about the deflection of photons in MOND. We would also like to thank Federico Lelli, who provided the idea for the lensing rotation curves shown in Fig. 3. Finally, we would like to thank the anonymous referee for insightful questions and useful comments, which helped to increase the value of this work. This work has made use of CosmoHub (Carretero et al. 2017; Tallada et al. 2020). CosmoHub has been developed by the Port d'Información Científica (PIC), maintained through a collaboration of the Institut de Física d'Altes Energies (IFAE) and the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), and was partially funded by the "Plan Estatal de Investigación Científica y Técnica y de Innovación" program of the Spanish government. KAO acknowledges support by the Netherlands Foundation for Scientific Research (NWO) through VICI grant 016.130.338 to M. Verheijen, and support from the European Research Council (ERC) through Advanced Investigator grant to C.S. Frenk, DMIDAS (GA 7786910). MBi is supported by the Polish National Science Center through grants no. 2020/38/E/ST9/00395, 2018/30/E/ST9/00698 and 2018/31/G/ST9/03388, and by the Polish Ministry of Science and Higher Education through grant DIR/WK/2018/12. CH acknowledges support from the European Research Council under grant number 647112, and support from the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck-Humboldt Research Award endowed by the Federal Ministry of Education and Research. HHo acknowledges support from Vici grant 639.043.512, financed by the Netherlands Organisation for Scientific Research (NWO). AHW is supported by an European Research Council Consolidator Grant (No. 770935). MA acknowledges support from the European Research Council under grant number 647112. AD acknowledges the ERC Consolidator Grant (No. 770935). BG acknowledges support from the European Research Council under grant number 647112 and from the Royal Society through an Enhancement Award (RGF/EA/181006). HHi is supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1) as well as an ERC Consolidator Grant (No. 770935). KK acknowledges support from the Royal Society and Imperial College. HYS acknowledges the support from NSFC of China under grant 11973070, the Shanghai Committee of Science and Technology grant (No. 19ZR1466600) and Key Research Program of Frontier Sciences, CAS (No. ZDBS-LY-7013). TT acknowledges support from the European Research Council under grant number 647112, as well as funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement (No. 797794). JLvdB is supported by an ERC Consolidator Grant (No. 770935). The work of MV is funded by the canton of Geneva and the Swiss National Science Foundation, through Project Grants 200020 182513 and NCCR 51NF40-141869 (SwissMAP). This work is partly based on tools and data products produced by GAZPAR operated by CeSAM-LAM and IAP. Computations for the *N*-body simulations were performed in part on the Orcinus supercomputer at the WestGrid HPC consortium (www.westgrid.ca), in part on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund – Research Excellence; and the University of Toronto. We are grateful to https://math.stackexchange.com user Paul Enta for providing an expression for one of the integrals needed in Appendix B. This work has made use of PYTHON (www.python.org), including the packages NUMPY (www.numpy.org) and SCIPY (www.scipy.org). Plots have been produced with MATPLOTLIB (Hunter 2007). Author contributions: All authors contributed to the development and writing of this paper. The authorship list is given in three groups: the lead authors (M. Brouwer, K. Oman, E. Valentijn), followed by two alphabetical groups. The first alphabetical group includes those who are key contributors to both the scientific analysis and the data products. The second group covers those who have either made a significant contribution to the data products, or to the scientific analysis.

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, A&A, 645, A104
Babcock, H. W. 1939, Lick Obs. Bull., 498, 41
Babyk, I. V., McNamara, B. R., Nulsen, P. E. J., et al. 2018, ApJ, 857, 32
Baldry, I. K., Balogh, M. L., Bower, R. G., et al. 2006, MNRAS, 373, 469
Banik, I., & Zhao, H. 2018, J. Astrophys., 1, 1000008
Banik, I., Thies, I., Famaey, B., et al. 2020, ApJ, 905, 135
Begeman, K. G., Broeils, A. H., & Sanders, R. H. 1991, MNRAS, 249, 523
Behroozi, P. S., Marchesini, D., Wechsler, R. H., et al. 2013, ApJ, 777, L10
Bell, E. F., & de Jong, R. S. 2001, ApJ, 550, 212
Benítez, N. 2000, ApJ, 536, 571
Bernstein, G. M., Guhathakurta, P., Raychaudhury, S., et al. 1994, AJ, 107, 1962
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bertone, G., & Tait, T. M. P. 2018, Nature, 562, 51
Bertone, G., Hooper, D., & Silk, J. 2005, Phys. Rep., 405, 279
Bílek, M., Samurović, S., & Renaud, F. 2019a, A&A, 629, L5
Bílek, M., Samurović, S., & Renaud, F. 2019b, A&A, 625, A32
Bilicki, M., Hoekstra, H., Brown, M. J. I., et al. 2018, A&A, 616, A69
Bilicki, M., Dvornik, A., Hoekstra, H., et al. 2021, ArXiv e-prints [arXiv:2101.06010]
Blake, C., Davis, T., Poole, G. B., et al. 2011, MNRAS, 415, 2892
Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003a, ApJ, 594, 186
Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003b, ApJ, 592, 819
Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, AJ, 129, 2562
Bolzonella, M., Kovač, K., Pozzetti, L., et al. 2010, A&A, 524, A76
Boselli, A., Eales, S., Cortese, L., et al. 2010, PASP, 122, 261
Boselli, A., Cortese, L., Boquien, M., et al. 2014, A&A, 564, A66
Bosma, A. 1981, AJ, 86, 1791
Brainerd, T. G., Blandford, R. D., & Smail, I. 1996, ApJ, 466, 623
Brouwer, M. M., Cacciato, M., Dvornik, A., et al. 2016, MNRAS, 462, 4451
Brouwer, M. M., Visser, M. R., Dvornik, A., et al. 2017, MNRAS, 466, 2547
Brouwer, M. M., Demchenko, V., Harnois-Déraps, J., et al. 2018, MNRAS, 481, 5189
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Burrage, C., Copeland, E. J., & Millington, P. 2017, Phys. Rev. D, 95, 064050
Capaccioli, M., & Schipani, P. 2011, The Messenger, 146, 2
Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, MNRAS, 447, 646
Carretero, J., Tallada, P., Casals, J., et al. 2017, in Proceedings of the European Physical Society Conference on High Energy Physics. 5-12 July, 2017 Venice, Italy (EPS-HEP2017), 488
Chabrier, G. 2003, PASP, 115, 763
Chae, K.-H., Lelli, F., Desmond, H., et al. 2020, ApJ, 904, 51
Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, ApJ, 648, L109
Correa, C. A., & Schaye, J. 2020, MNRAS, 499, 3578
Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, MNRAS, 453, 1513
Crocker, A. F., Bureau, M., Young, L. M., & Combes, F. 2011, MNRAS, 410, 1197
Cuddeford, P. 1993, MNRAS, 262, 1076
Dark Energy Science Collaboration 2012, ArXiv e-prints [arXiv:1211.0310]
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
Davis, T. A., Alatalo, K., Bureau, M., et al. 2013, MNRAS, 429, 534
de Bernardis, P., Ade, P. A. R., Bock, J. J., et al. 2000, Nature, 404, 955
de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Exp. Astron., 35, 25
de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, The Messenger, 175, 3
Desmond, H. 2017, MNRAS, 464, 4160
Di Cintio, A., & Lelli, F. 2016, MNRAS, 456, L127
Di Paolo, C., Salucci, P., & Fontaine, J. P. 2019, ApJ, 873, 106
Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, MNRAS, 399, 497
Domínguez Sánchez, H., Bernardi, M., Brownstein, J. R., Drory, N., & Sheth, R. K. 2019, MNRAS, 489, 5612
Dressler, A. 1980, ApJ, 236, 351

Driver, S. P., Allen, P. D., Graham, A. W., et al. 2006, MNRAS, 368, 414
Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, MNRAS, 413, 971
Dutton, A. A., Conroy, C., van den Bosch, F. C., Prada, F., & More, S. 2010, MNRAS, 407, 2
Dvornik, A., Cacciato, M., Kuijken, K., et al. 2017, MNRAS, 468, 3251
Dvornik, A., Hoekstra, H., Kuijken, K., et al. 2018, MNRAS, 479, 1240
Eddington, A. S. 1913, MNRAS, 73, 359
Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560
Erben, T., Hildebrandt, H., Miller, L., et al. 2013, MNRAS, 433, 2545
Faulkner, T., Guica, M., Hartman, T., Myers, R. C., & Van Raamsdonk, M. 2014, J. High Energy Phys., 3, 51
Fedeli, C., Semboloni, E., Velliscig, M., et al. 2014, JCAP, 8, 028
Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, MNRAS, 467, 1627
Fischer, P., McKay, T. A., Sheldon, E., et al. 2000, AJ, 120, 1198
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fosalba, P., Gaztañaga, E., Castander, F. J., & Manera, M. 2008, MNRAS, 391, 435
Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015a, MNRAS, 448, 2987
Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2015b, MNRAS, 447, 1319
Fukugita, M., & Peebles, P. J. E. 2004, ApJ, 616, 643
Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, ApJ, 503, 518
Giblin, B., Heymans, C., Asgari, M., et al. 2021, A&A, 645, A105
Goto, T., Yamauchi, C., Fujita, Y., et al. 2003, MNRAS, 346, 601
Gottesman, S. T., Davies, R. D., & Reddish, V. C. 1966, MNRAS, 133, 359
Graham, A. W. 2019, PASA, 36, e035
Guo, Q., Hu, H., Zheng, Z., et al. 2019, Nat. Astron., 4, 246
Heymans, C., Grocutt, E., Heavens, A., et al. 2013, MNRAS, 432, 2433
Heymans, C., Tröster, T., Asgari, M., et al. 2021, A&A, 646, A140
Hildebrandt, H., Erben, T., Kuijken, K., et al. 2012, MNRAS, 421, 2355
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, A&A, 647, A124
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
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
Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, MNRAS, 447, 1724
Hunter, J. D., et al. 2007, Comput. Sci. Eng., 9, 90
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Jacobson, T. 1995, Phys. Rev. Lett., 75, 1260
Jacobson, T. 2016, Phys. Rev. Lett., 116, 201101
Jakobs, A., Viola, M., McCarthy, I., et al. 2018, MNRAS, 480, 3338
Janz, J., Penny, S. J., Graham, A. W., Forbes, D. A., & Davies, R. L. 2017, MNRAS, 468, 2850
Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, A&A, 624, A92
Kapteyn, J. C. 1922, ApJ, 55, 302
Keller, B. W., & Wadsley, J. W. 2017, ApJ, 835, L17
Kirkman, D., Tytler, D., Suzuki, N., O'Meara, J. M., & Lubin, D. 2003, ApJS, 149, 1
Kuijken, K. 2008, A&A, 482, 1053
Kuijken, K. 2011, The Messenger, 146, 8
Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500
Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2
La Barbera, F., de Carvalho, R. R., Kohl-Moreira, J. L., et al. 2008, PASP, 120, 681
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
Lelli, F., McGaugh, S. S., & Schombert, J. M. 2016, AJ, 152, 157
Lelli, F., McGaugh, S. S., & Schombert, J. M. 2017a, MNRAS, 468, L68
Lelli, F., McGaugh, S. S., Schombert, J. M., & Pawlowski, M. S. 2017b, ApJ, 836, 152
Li, P., Lelli, F., McGaugh, S., & Schombert, J. 2018, A&A, 615, A3
Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, MNRAS, 452, 2087
Ludlow, A. D., Navarro, J. F., Angulo, R. E., et al. 2014, MNRAS, 441, 378
Ludlow, A. D., Benítez-Llambay, A., Schaller, M., et al. 2017, Phys. Rev. Lett., 118, 161103
Macquart, J. P., Prochaska, J. X., McQuinn, M., et al. 2020, Nature, 581, 391
Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, MNRAS, 368, 715
Marasco, A., Posti, L., Oman, K., et al. 2020, A&A, 640, A70
Matthee, J., Schaye, J., Crain, R. A., et al. 2017, MNRAS, 465, 2381
McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, MNRAS, 465, 2936
McFarland, J. P., Verdoes-Kleijn, G., Sikkema, G., et al. 2013, Exp. Astron., 35, 45
McGaugh, S. S. 2004, ApJ, 609, 652
McGaugh, S. S. 2012, AJ, 143, 40
McGaugh, S. S., Schombert, J. M., Bothun, G. D., & de Blok, W. J. G. 2000, ApJ, 533, L99
McGaugh, S. S., Lelli, F., & Schombert, J. M. 2016, Phys. Rev. Lett., 117, 201101
Mentuch Cooper, E., Wilson, C. D., Foyle, K., et al. 2012, ApJ, 755, 165
Milgrom, M. 1983, ApJ, 270, 365
Milgrom, M. 2013, Phys. Rev. Lett., 111, 041105
Milgrom, M., & Sanders, R. H. 2008, ApJ, 678, 131
Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, MNRAS, 382, 315
Miller, L., Heymans, C., Kitching, T. D., et al. 2013, MNRAS, 429, 2858
Müller, O., Fahrion, K., Rejkuba, M., et al. 2021, A&A, 645, A92
Navarro, J. F., Benítez-Llambay, A., Fattahi, A., et al. 2017, MNRAS, 471, 1841
Nicastro, F., Kaastra, J., Krongold, Y., et al. 2018, Nature, 558, 406
O'Brien, J. G., Chiarelli, T. L., Mannheim, P. D., et al. 2019, J. Phys. Conf. Ser., 1239, 012009
Oman, K. A., Navarro, J. F., Sales, L. V., et al. 2016, MNRAS, 460, 3610
Oort, J. H. 1932, Bull. Astron. Inst. Neth., 6, 249
Oort, J. H. 1940, ApJ, 91, 273
Padmanabhan, T. 2010, Rep. Progr. Phys., 73, 046901
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Pierce, M. J., & Tully, R. B. 1988, ApJ, 330, 579
Planck Collabration XVI. 2014, A&A, 571, A16
Planck Collabration VI. 2020, A&A, 641, A6
Pohlen, M., Cortese, L., Smith, M. W. L., et al. 2010, A&A, 518, L72
Posti, L., Fraternali, F., & Marasco, A. 2019, A&A, 626, A56
Power, C., Navarro, J. F., Jenkins, A., et al. 2003, MNRAS, 338, 14
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Robotham, A. S., Norberg, P., Driver, S. P., et al. 2011, MNRAS, 416, 2640
Roy, N., Napolitano, N. R., La Barbera, F., et al. 2018, MNRAS, 480, 1057
Rubin, V. C. 1983, Sci. Am., 248, 96
Ruiz-Macias, O., Zarrouk, P., Cole, S., et al. 2021, MNRAS, 502, 4328
Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, PASP, 128, 104502
Sanders, R. H. 1986, MNRAS, 223, 539
Sanders, R. H. 1996, ApJ, 473, 117
Sanders, R. H., & Noordermeer, E. 2007, MNRAS, 379, 702
Schrabback, T., Hoekstra, H., Van Waerbeke, L., et al. 2021, A&A, 646, A73
Shull, J. M., Smith, B. D., & Danforth, C. W. 2012, ApJ, 759, 23
Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJS, 148, 175
Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726
Tallada, P., Carretero, J., Casals, J., et al. 2020, Astron. Comput., 32, 100391
Tanimura, H., Aghanim, N., Kolodzig, A., Douspis, M., & Malavasi, N. 2020, A&A, 643, L2
Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, MNRAS, 418, 1587
Taylor, E. N., Cluver, M. E., Duffy, A., et al. 2020, MNRAS, 499, 2896
Tenneti, A., Mao, Y.-Y., Croft, R. A. C., et al. 2018, MNRAS, 474, 3125
Tian, L., Hoekstra, H., & Zhao, H. 2009, MNRAS, 393, 885
Tian, Y., Umetsu, K., Ko, C.-M., Donahue, M., & Chiu, I. N. 2020, ApJ, 896, 70
Tröster, T., Asgari, M., Blake, C., et al. 2021, A&A, 649, A88
Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661
Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARA&A, 55, 389
Valentijn, E. A. 1988, A&A, 203, L17
van Dokkum, P., Danieli, S., Cohen, Y., et al. 2018, Nature, 555, 629
van Uitert, E., Cacciato, M., Hoekstra, H., et al. 2016, MNRAS, 459, 3251
Verlinde, E. 2011, J. High Energy Phys., 4, 29
Verlinde, E. 2017, SciPost Phys., 2, 016
Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, Phys. Rev. D, 88, 043502
Viola, M., Cacciato, M., Brouwer, M., et al. 2015, MNRAS, 452, 3529
von der Linden, A., Allen, M. T., Applegate, D. E., et al. 2014, MNRAS, 439, 2
Wright, A. H., Robotham, A. S. G., Driver, S. P., et al. 2017, MNRAS, 470, 283
Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020, A&A, 640, L14
Wu, X., & Kroupa, P. 2015, MNRAS, 446, 330
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zwicky, F. 1933, Helvet. Phys. Acta, 6, 110

Appendix A: Isolated galaxy selection and validation

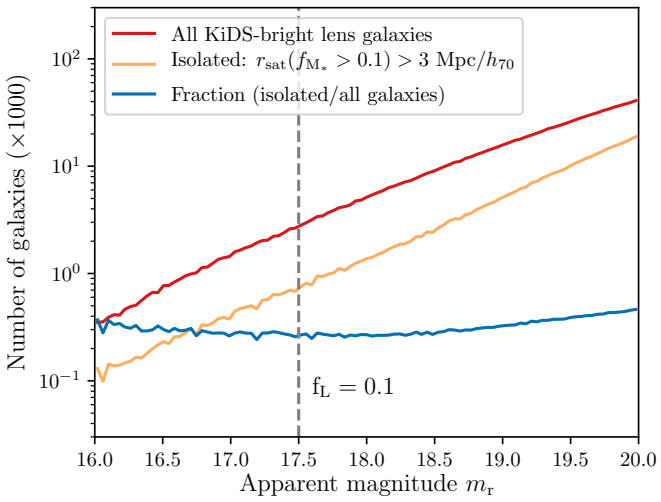

Fig. A.1. Histogram of the number of isolated galaxies (orange line) compared to the total number of galaxies (red line), as a function of apparent r-band magnitude m_r. The dashed vertical line represents the magnitude $m_{\rm bright}$, below which all satellites with a luminosity fraction larger than $f_{\rm L} \equiv L_{\rm sat}/L_{\rm lens} = 0.1$ compared to the lens are still detected. Beyond this limit, the fraction of isolated galaxies (blue line) slightly increases because satellites fainter than the flux limit are not detected, which can cause lenses close to the magnitude limit ($m_{\rm lim} = 20$ mag) to be falsely identified as isolated.

After performing the measurement of the RAR using GGL, our final goal was to compare the results to the different analytical models (Sect. 2) and N-body simulations (Sect. 4) that make specific predictions on the galaxy-halo connection. While the simulations were designed to describe galaxies in their cosmological environment, the analytical models mainly focus on the description of individual galaxies. This means that, in order to test these models, we need to select galaxies that are relatively isolated. We defined our isolated lenses such that they do not have any satellites with more than a fraction $f_{M_\star} \equiv M_{\star,\rm sat}/M_{\star,\rm lens} = 0.1$ of their stellar mass within a spherical radius $r_{\rm sat} = 3\,h_{70}^{-1}$ Mpc (see Sect. 3.3).

Here we validate our isolation criterion using the KiDS-bright and MICE datasets. We find that increasing the value of $r_{\rm sat}$ does not yield any decrease in the 'two-halo term': the GGL signal at larger scales (>0.3 h_{70}^{-1} Mpc) corresponding to the contribution of satellites. This is true both when all lens masses are considered, and when they are restricted to a specific stellar mass: $\log(M_\star/h_{70}^{-2}\,M_\odot) = 10.5 \pm 0.1$. Using both the KiDS-bright and MICE galaxies, we reduce the satellite mass fraction to $f_{M_\star} = 0$ (corresponding to no visible satellites). This also yields no decrease in the two-halo term of the ESD profile since galaxies with $f_{M_\star} \ll 0.1$ are not likely to be observed in a flux-limited survey. When we restrict the total stellar mass $M_{\star,\rm tot}$ of all satellites within $r_{\rm sat}$ to $f_{M_{\star,\rm tot}} < 0.1$ this does not significantly affect the isolated lens sample (i.e., the samples selected with KiDS-bright are >99% overlapping). Using the MICE and KiDS data we also experimented with selecting lenses that are isolated within a conical frustum, defined by a projected radius R and line-of-sight distance range ΔD around the lens. However, significantly increasing ΔD beyond 3 h_{70}^{-1} Mpc has no effect on the ESD profile, until it reduces our number of selected lenses to the point where the S/N does not allow for a significant measurement. Finally, we apply our isolation criterion to the GAMA survey, to compare our current isolated sample with the 'isolated centrals' that we used in Brouwer et al. (2017). These were selected using a more elaborate isolation criterion, which

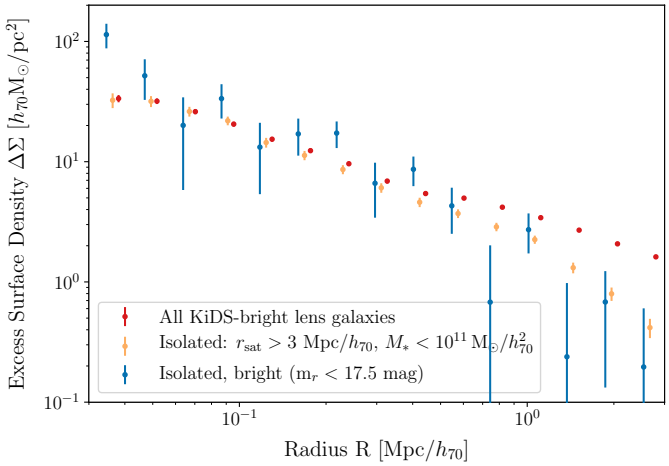

Fig. A.2. Measured ESD profiles of our full sample of isolated KiDS-bright galaxies (red points with 1σ error bars), compared to that of a more reliable 'bright' sample (blue, $m_r < 17.5$ mag), which allows us to see all satellites down to luminosity fraction $f_{\rm L} \equiv L_{\rm sat}/L_{\rm lens} = 0.1$. This is done to assess the effect of the KiDS-bright magnitude limit ($m_r < 20$ mag) on the isolation criterion. Due to the smaller number of lenses, the ESD profile of the bright isolated sample is noisier. Nevertheless, its behaviour on both small and large scales is consistent with the ESD profile of the full isolated sample (orange), indicating that the effect of the magnitude limit is limited.

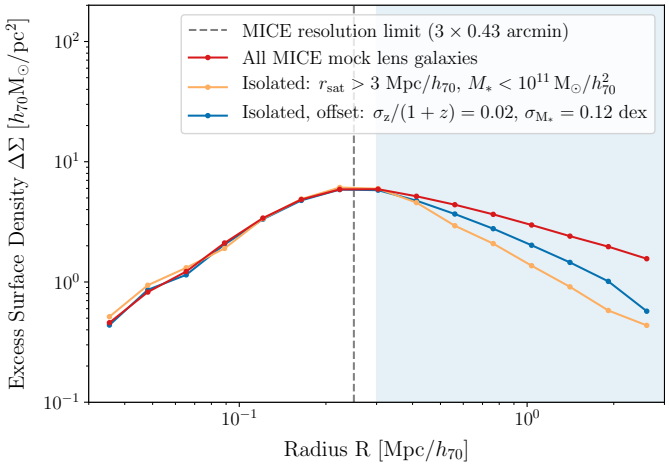

Fig. A.3. Simulated ESD profile of the offset isolated MICE sample (blue line), created by using MICE galaxies with randomly offset redshifts ($\sigma_z/[1+z] = 0.02$) and stellar masses ($\sigma_{M_\star} = 0.12$ dex) when selecting the isolated lenses. This is done in order to mimic the effect of the KiDS-bright measurement uncertainties on the isolation criterion. Compared to the ESD profile of the truly isolated MICE sample (orange line) the offset sample has a ~30% higher signal at large scales due to the contribution of satellites. We therefore take extra care with KiDS-bright results at $R > 0.3\,h_{70}^{-1}$ Mpc (light blue shaded region). Nevertheless, the ESD of the offset isolated MICE sample is significantly lower than that of all MICE galaxies (red line), created without any isolation criterion. In addition, we show the radius corresponding to three times the resolution of the MICE simulation (dashed vertical line), which in the case of the isolated MICE sample is $R < 0.25\,h_{70}^{-1}$ Mpc. Throughout this work, we only use the MICE results beyond this radius.

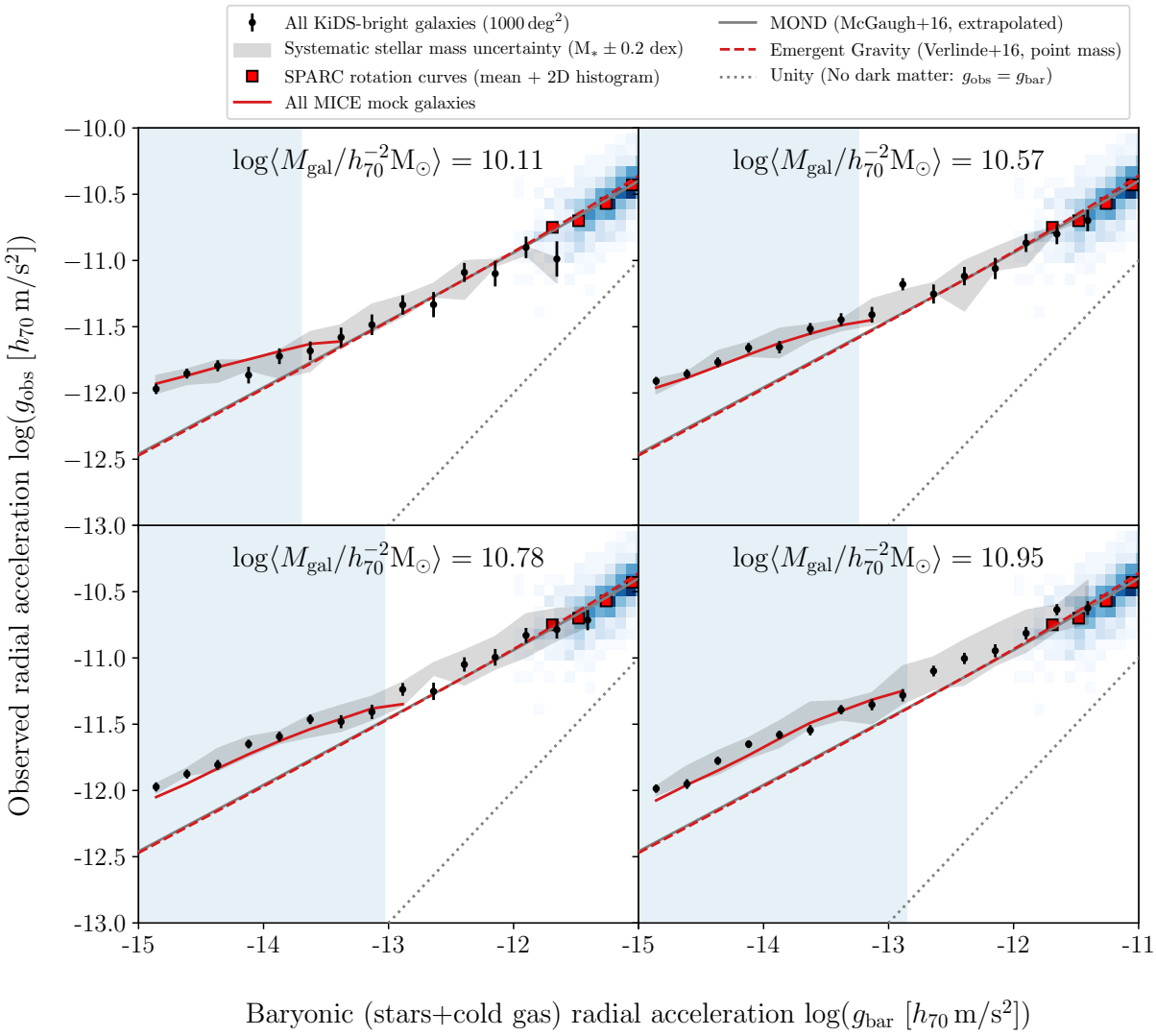

Fig. A.4. Measured RAR of all GL-KiDS lenses (black points with 1σ error bars) divided into four stellar mass bins, with the mean galaxy mass (stars + cold gas) of the lenses shown at the top of each panel. This figure primarily shows the effect on the RAR when the isolation criterion is not applied, which is quite significant and depends on the stellar mass of the galaxies (which correlates with galaxy clustering). The extrapolated M16 and EG predictions (grey solid and red dashed lines) function merely as a reference, showing the approximate location of the isolated galaxy RAR. They do not represent predictions in this case because g_{bar} is calculated using only the baryonic masses of the main lens galaxies (without including the baryonic masses of the satellites). The predictions from the MICE simulation (red line) match with our observations, which shows that the clustering simulated within MICE, driving the low-acceleration upturn due to the two-halo term, is indeed quite accurate.

was driven by the Friends-of-Friends (FoF) group finding algorithm of Robotham et al. (2011). We find that the two isolated galaxy samples are more than 80% overlapping.

However, because both the GAMA survey and the samples designed to mimic it (KiDS-bright and MICE) are flux-limited, satellites that are fainter than the flux limit are not detected. This can cause lenses that are close to the magnitude limit (m_{lim} = 20 mag) to be falsely identified as isolated. This problem is illustrated in Fig. A.1, which shows that the fraction of galaxies assigned to the isolated lens sample increases for higher values of the apparent r-band magnitude m_r. The dashed vertical line represents the magnitude m_{bright}, below which all satellites with a luminosity fraction larger than $f_L \equiv L_{sat}/L_{lens} = 0.1$ compared to the lens are still detected. In the case of KiDS-bright:

$$m_{bright} = m_{lim} - 2.5\log_{10}(f_F = 0.1) = 17.5\,\text{mag}. \quad (A.1)$$

Applying $m_r < m_{bright}$ provided us with an isolated lens sample that should be free of false positives, allowing us to estimate their effect on the ESD profiles. In Fig. A.2 we compare the ESD profiles of all galaxies and isolated galaxies with the more reliable 'bright' sample. The mean stellar masses of the lens samples are very similar for the bright and the full sample: $\log\langle M_\star/h_{70}^{-2}\,M_\odot\rangle$ = 10.78 and 10.77 respectively, and slightly lower for the full isolated sample: $\log\langle M_\star/h_{70}^{-2}\,M_\odot\rangle$ = 10.61. Due to the smaller number of lenses (only 3800), the ESD errors and scatter of the bright isolated sample are much larger than those of the full isolated sample. Nevertheless, it is clear that their ESD profiles show consistent behaviour at both small and large scales. Compared to the total (non-isolated) galaxy sample, both isolated samples show significantly lower lensing signals at large scales (the two-halo term, corresponding to the contribution of satellites). The high level of consistency between the ESD profiles of the full and bright isolated samples indicates that the effect of false positives due to the magnitude limit is limited. In addition, by comparing the expected percentage of true isolated galaxies (32.0%, found in the bright sample) with

the higher percentage found in the 'faint' sample (36.8%, for galaxies with $m_r > 17.5$ mag), we estimated that the expected percentage of false positives is less than 20% of the full sample of isolated galaxies.

Nevertheless, we used the MICE simulations to perform one additional test. We selected the isolated sample of MICE lenses using satellite galaxies that extend to $m_r < 22.5$ mag, such that all satellites with $f_L > 0.1$ can be observed. This paints a similar picture as the bright KiDS sample: although the much smaller sample of isolated galaxies selected using the faint satellites greatly increases the scatter, we find no consistent decrease in the lensing signal at $>0.3\,h_{70}^{-1}$ Mpc scales compared to the original sample of isolated MICE galaxies. All these tests demonstrate the overall robustness of our isolation criterion. In addition, we note that this issue is only relevant when comparing our observations to the theoretical models (EG, MOND and N17). When comparing to the simulations (BAHAMAS and MICE), applying the same isolation criterion to both data and mocks ensured that any issues with the isolated galaxy selection are mimicked.

The major difference between the isolated galaxy selection of the GAMA and mock galaxies compared to KiDS-bright is that for GAMA and the mocks the true redshift values are known, whereas the ANNz2 photometric redshifts of KiDS-bright are only known within a certain standard deviation σ_z (see Sect. 3.3). Since these photometric redshifts were used to calculate the galaxy distances $D(z)$ (using a flat ΛCDM cosmology, ignoring peculiar velocities) they directly affect the observed spherical distances r between the galaxies, a key ingredient of the isolation criterion. The redshift uncertainty also affects the KiDS-bright stellar mass estimates, which influence both the isolation criterion (through $f_{M_\star}$) and the application of the stellar mass limit: $\log(M_\star/h_{70}^{-2}\,M_\odot) < 11$. We assessed the effect of these uncertainties on the isolated galaxy selection by adding a normally distributed random offset with $\sigma_z/(1+z) = 0.02$ to the MICE redshifts, and $\sigma_{M_\star} = 0.12$ dex to its stellar masses. We find that the effect of the mass uncertainty is negligible, but that of redshift uncertainty is significant. Because the random redshift offset decreases the galaxy clustering, it increases the number of galaxies selected by the isolation criterion, adding galaxies that are not truly isolated to the lens sample (as well as excluding some truly isolated galaxies).

The ESD profile of the offset isolated MICE sample is shown in Fig. A.3, compared to the ESD profiles of all MICE galaxies (without any isolation criterion) and the truly isolated MICE sample. At scales $R > 0.3\,h_{70}^{-1}$ Mpc, the ESD of the isolated sample selected using the offset MICE data is ∼30% higher than that of the truly isolated MICE galaxies. When comparing our KiDS-bright lensing measurements to the MICE simulation, we always take this effect into account by mimicking the redshift offset in the simulation. However, for our comparison with the analytical models (MOND, EG and N17) this process is more difficult. When testing these models, we can only use the ESD profile of isolated KiDS-bright lenses within $R < 0.3\,h_{70}^{-1}$ Mpc. For the mean galaxy mass of the KiDS-bright isolated sample ($\log[M_{\rm gal}/h_{70}^{-2}\,M_\odot] = 10.69$) this corresponds to a baryonic acceleration of $g_{\rm gal} > 7.56{\times}10^{-14}$ m s^{-2}. For each RAR measurement resulting from isolated KiDS-bright lenses we will indicate the range in $g_{\rm bar}$ where the measurement is reliable, based on the mean $M_{\rm gal}$ of the appropriate lens sample.

Finally, to indicate the effect of selecting isolated galaxies on our lensing RAR measurements, Fig. A.4 shows the RAR of KiDS-bright and MICE galaxies for all lens galaxies without applying the isolation criterion. Because the clustering of galaxies (and hence the effect of the satellite galaxies) correlates with their stellar mass, we divided the lens galaxies into the same four stellar mass bins as used in Sect. 5.5: $\log_{10}(M_\star/h_{70}^{-2}\,M_\odot) = [8.5, 10.3, 10.6, 10.8, 11.0]$. In that section, Fig. 9 shows the RAR of isolated galaxies in the same stellar mass bins. In both cases, $g_{\rm bar}$ is calculated using only the baryonic masses of the main lens galaxies (i.e., the baryonic masses of the satellites are not included in the x-axis of Fig. A.4). Comparing these two results shows that the effect of our isolated galaxy selection on $g_{\rm obs}$ is very striking: the isolated RAR measurements in Fig. 9 approximately follow the extrapolated M16 and EG predictions, while the non-isolated RAR measurements in Fig. A.4 lie well above these lines at low accelerations ($g_{\rm bar} < 10^{-13}$ m s^{-2}). As expected, the strength of this two-halo term (which shows the amount of clustering) increases with increasing galaxy stellar mass. Again the MICE simulation was able to predict our measurements: $\chi^2_{\rm red} = 51.3/33 = 1.6$ (2.3σ). This shows that the clustering simulated within MICE, which drives the low-acceleration upturn due to the two-halo term, is indeed quite accurate. This was to be expected since the clustering in MICE is constructed to reproduce the SDSS observations at $z < 0.25$ (Zehavi et al. 2011).

Appendix B: Calculating $g_{\rm obs}$ from an ESD profile

To calculate $g_{\rm obs}$ from the ESD profile throughout this work, we have used a simple analytical method that assumes that DM haloes can be roughly approximated with a singular isothermal sphere density model (see Sect. 2.2). To make sure this conversion is robust, we compared it to a more elaborate numerical approach that fits a piece-wise power law (PPL) to the stacked ESD profile, without any assumption on the averaged halo shape except for spherical symmetry. We validate both methods using mock surface density maps from the BAHAMAS simulation in Sect. 4.4.

The PPL method assumes a self-consistent form for the volume density profile $\rho(r)$ and parametrises it as a piece-wise power law constrained to be continuous. This comes at the cost of needing to invert the non-linear function $\Delta\Sigma(\rho)$, which we achieve via an iterative method. We chose to parametrize $\rho(r)$ in terms of N pairs of values (r_n, ρ_n) such that the slope a_n and normalisation b_n of the power law profile segments are:

$$\ln\rho = a_n \ln(r) + b_n \tag{B.1}$$

$$a_n = \frac{\ln(\rho_{n+1}) - \ln(\rho_n)}{\ln(r_{n+1}) - \ln(r_n)} \tag{B.2}$$

$$b_n = \ln(\rho_n) - a_n \ln(r_n) \tag{B.3}$$

$$(a_n, b_n) = \begin{cases} (a_0, b_0) & \text{if } r < r_0 \\ (a_n, b_n) & \text{if } r_n \le r < r_{n+1} \\ (a_{N-1}, b_{N-1}) & \text{if } r \ge r_N\,. \end{cases} \tag{B.4}$$

The ESD profile was measured in a series of discrete radial bins with edges R_m. The representative value at the centre of the bin[19] is $\Delta\Sigma_m = \overline{\Sigma}_m - \Sigma_m$, where $\overline{\Sigma}_m$ is the mean surface density within $\frac{1}{2}(R_m + R_{m+1})$ and Σ_m is the surface density averaged over the interval $[R_m, R_{m+1})$. We give an expression for this discrete ESD profile in terms of the parametric form for $\rho(r)$ given in Eq. (B.4).

The mean enclosed surface density is:

$$\overline{\Sigma}_m = \frac{1}{\pi R_m R_{m+1}} \left[I_1(0, \sqrt{R_0 R_1}, \tilde{a}_0, \tilde{b}_0) + \sum_{k=0}^{m} I_1(\sqrt{R_m R_{m+1}}, \sqrt{R_{m+1} R_{m+2}}, \tilde{a}_m, \tilde{b}_m) \right] \tag{B.5}$$

$$\tilde{a}_m = \frac{\ln(\Sigma_{m+1}) - \ln(\Sigma_m)}{\frac{1}{2}\left(\ln(R_{m+2}) - \ln(R_m)\right)} \tag{B.6}$$

$$\tilde{b}_m = \ln(\Sigma_m) - \frac{1}{2}\tilde{a}_m \ln(R_m R_{m+1}) \tag{B.7}$$

$$I_1(R_i, R_j, \tilde{a}, \tilde{b}) = \frac{2\pi e^{\tilde{b}}}{\tilde{a}+2}\left(R_j^{a+2} - R_i^{a+2}\right), \tag{B.8}$$

and the local surface density is given by:

$$\Sigma_m = \sum_{n=0}^{N-1} \begin{cases} 0 & \text{if } r_{n+1} < R_m \\ \frac{4e^{b_n}}{R_{m+1}^2 - R_m^2}\left(-I_2(r_{n+1}, R_m, a_n)\right) & \text{if } r_n < R_m \text{ and } R_m \le r_{n+1} < R_{m+1} \\ \frac{4e^{b_n}}{R_{m+1}^2 - R_m^2}\left(I_2(r_{n+1}, R_{m+1}, a_n) - I_2(r_{n+1}, R_m, a_n)\right) & \text{if } r_n < R_m \text{ and } r_{n+1} \ge R_{m+1} \\ \frac{4e^{b_n}}{R_{m+1}^2 - R_m^2}\left(I_2(r_{n+1}, r_n, a_n) - I_2(r_{n+1}, R_m, a_n)\right. & \\ \quad \left. + I_2(r_n, R_m, a_n) + I_2(r_{n+1}, R_{m+1}, a_n)\right. & \\ \quad \left. - I_2(r_{n+1}, r_n, a_n)\right) & \text{if } R_m \le r_n < R_{m+1} \text{ and } r_n \ge R_{m+1} \\ \frac{4e^{b_n}}{R_{m+1}^2 - R_m^2}\left(I_2(r_{n+1}, R_{m+1}, a_n) - I_2(r_{n+1}, R_m, a_n)\right. & \\ \quad \left. - I_2(r_n, R_{m+1}, a_n) + I_2(r_n, R_m, a_n)\right) & \text{if } r_n \ge R_{m+1} \\ \frac{4e^{b_n}}{R_{m+1}^2 - R_m^2}\left(I_2(r_{n+1}, r_n, a_n) - I_2(r_{n+1}, R_m, a_n)\right. & \\ \quad \left. + I_2(r_n, R_m, a_n) - I_2(r_{n+1}, r_n, a_n)\right) & \text{if } r_n \ge R_m \text{ and } r_{n+1} < R_m \end{cases} \tag{B.9}$$

$$I_2(r, R, a) = \begin{cases} -\frac{1}{3}R^{a+3}\left(\frac{r^2}{R^2} - 1\right)^{\frac{3}{2}}\, {}_2\mathrm{F}_1\left(\frac{3}{2}, -\frac{a}{2}; \frac{5}{2}; 1 - \frac{r^2}{R^2}\right) & \text{if } r \text{ is finite} \\ \frac{\sqrt{\pi}}{2}\frac{\Gamma\left(-\frac{a+1}{2}\right)}{\Gamma\left(-\frac{a}{2}\right)}\frac{R^{a+3}}{a+3} & \text{if } r = \infty\,, \end{cases} \tag{B.10}$$

where ${}_2\mathrm{F}_1(\cdot,\cdot;\cdot;\cdot)$ is the Gaussian hypergeometric function and $\Gamma(\cdot)$ is the Gamma function. We assumed that the power law slope in the innermost bin continues to $r = 0$, and that the slope in the outermost bin continues to $r \to \infty$. When inverting $\Delta\Sigma(\rho)$, we imposed uninformative priors on the power law slopes except those constraints required to guarantee that the total mass is finite and that the calculation is numerically stable.

[19] Here we define the bin centre as $\frac{1}{2}(R_m + R_{m+1})$, not the logarithmic centre $\sqrt{R_m R_{m+1}}$, which ensures accuracy in the calculation of the mean enclosed surface density.

In order to invert $\Delta\Sigma_m(\rho_n)$, we took as constant the values R_m, $\Delta\Sigma_{\mathrm{obs},m}$ and r_n. We then proposed an initial guess ρ_n, which we perturbed iteratively, calculating the corresponding $\Delta\Sigma_m$ at each iteration and comparing with $\Delta\Sigma_{\mathrm{obs},m}$ via the likelihood function:

$$\ln \mathcal{L} \propto -\frac{1}{2}(\Delta\Sigma_{\mathrm{obs}} - \Delta\Sigma)^{\mathrm{T}} C^{-1} (\Delta\Sigma_{\mathrm{obs}} - \Delta\Sigma), \tag{B.11}$$

where C is the covariance matrix for the $\Delta\Sigma_{\mathrm{obs}}$. We used the package EMCEE (Foreman-Mackey et al. 2013) to estimate the posterior probability distribution of ρ_n, and subsequently of the corresponding $g_{\mathrm{obs},n}$ via integration of the volume density profile.

Appendix C: The RAR of the N17 model

We used the lensing RAR to test the analytical prediction from the ΛCDM-based model created by N17. In Fig. C.1 we show the RAR predicted by this model for a galaxy with a baryonic mass equal to the average stellar + cold gas mass of the lens sample ($\log_{10}\langle M_{\rm gal}\rangle = 10.69$). At higher accelerations there is a good match between the model and the M16 RAR measurements from galaxy rotation curves, which is expected since the N17 model is designed and confirmed to reproduce these results. However, at the lower accelerations unique to our lensing measurements the N17 model underpredicts the $g_{\rm obs}$ amplitude in comparison to our measurements. Due to their large error bars, the GAMA data can still accommodate the analytical prediction: $\chi^2_{\rm red} = 0.90$. The KiDS-bright result, however, excludes the N17 prediction with $\chi^2_{\rm red} = 4.8$, corresponding to 4.3σ. Here we have removed all data points beyond the KiDS isolation limit ($R > 3\,h_{70}^{-1}$ Mpc); therefore, the strong disagreement between N17 and the data is unlikely to be caused by contamination from satellites.

Because of the significant difference in the slope of the model and the data, even taking the $\Delta M_\star = \pm 0.2$ dex uncertainty into account does not result in a better fit. This strong downward slope results from the r^{-3} radial dependence of the Navarro–Frenk–White (NFW) density profile at large scales (where an r^{-2} density profile would instead follow the same slope as the MG predictions in Fig. 4). This effect could be slightly mitigated by taking into account the average DM density of the Universe, which would result in an upward turn towards an r^{-1} slope at $g_{\rm bar} < 10^{14}\,{\rm m\,s^{-2}}$ (as shown by the BAHAMAS prediction of the RAR in the lower panel of Fig. 1). However, components contributing to the large-scale DM profile are not included in the N17 model, which was created to reproduce the RAR at the small scales measured by rotation curves. It is clear from this exercise that, while succeeding to describe the RAR at small scales, this simple model is not sufficient to reproduce the results at the larger scales probed by weak lensing. This requires more elaborate modelling within the ΛCDM paradigm, represented by large cosmological simulations such as BAHAMAS and MICE (see Sect. 4). In Sect. 5.3 we made a fairer comparison using these two simulations, which can mimic the measurement more faithfully.

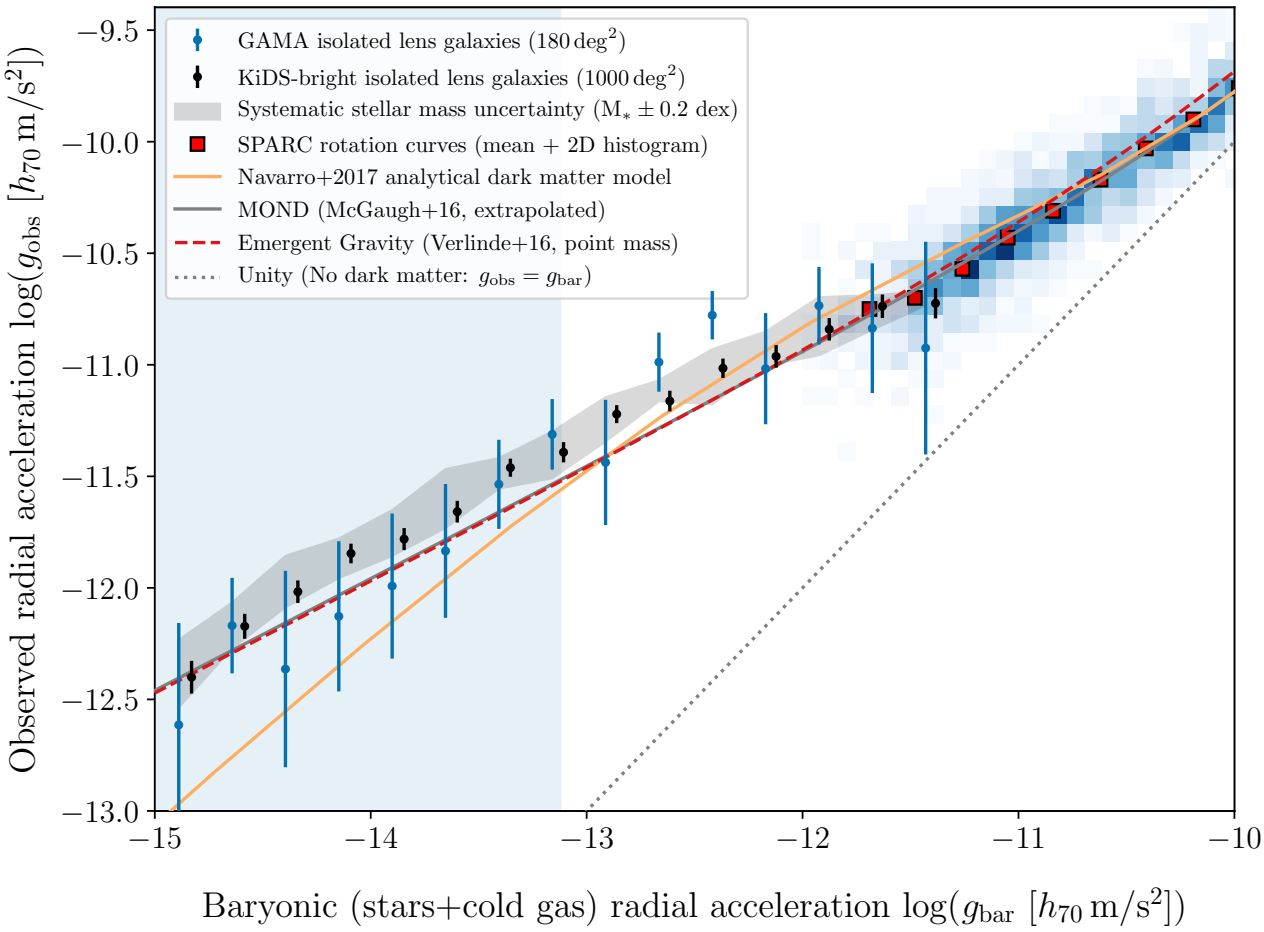

Fig. C.1. Measured RAR of the spectroscopic GAMA and photometric KiDS-bright isolated lens samples (blue and black points with 1σ error bars). We compare our results to the analytical ΛCDM-based model created by N17. At higher accelerations there is a good match between the N17 model and the M16 observations, as expected. However, at lower accelerations the model bends down with respect to the lensing measurements, due to the steep outer slope of the NFW density profile ($\rho \propto r^{-3}$). Large-scale contributions to the total mass distribution, such as the average cosmic DM density, could slightly mitigate this discrepancy. However, these are not implemented into the simple analytical N17 model, which was created to reproduce the RAR at the small scales measured by rotation curves.