KiDS-1000: Combined halo-model cosmology constraints from galaxy abundance, galaxy clustering and galaxy-galaxy lensing

Andrej Dvornik[1], Catherine Heymans[2, 1], Marika Asgari[3, 2], Constance Mahony[1], Benjamin Joachimi[4], Maciej Bilicki[5], Elisa Chisari[6], Hendrik Hildebrandt[1], Henk Hoekstra[7], Harry Johnston[6], Konrad Kuijken[7], Alexander Mead[8, 1], Hironao Miyatake[9, 10, 11], Takahiro Nishimichi[12, 10], Robert Reischke[1], Sandra Unruh[13, 1], and Angus H. Wright[1]

[1] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany
e-mail: `dvornik@astro.rub.de`
[2] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[3] E. A. Milne Centre, University of Hull, Cottingham Road, Hull, HU6 7RX, UK
[4] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
[5] Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668 Warsaw, Poland
[6] Institute for Theoretical Physics, Utrecht University, 3584CC Utrecht, The Netherlands
[7] Leiden Observatory, Leiden University, P.O.Box 9513, 2300RA Leiden, The Netherlands
[8] Department of Computer Science, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
[9] Kobayashi-Maskawa Institute for the Origin of Particles and the Universe (KMI), Nagoya University, Nagoya 464-8602, Japan
[10] Institute for Advanced Research, Nagoya University, Nagoya 464-8601, Japan
[11] Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Chiba 277-8583, Japan
[12] Center for Gravitational Physics and Quantum Information, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan
[13] Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany

ABSTRACT

We present constraints on the flat ΛCDM cosmological model through a joint analysis of galaxy abundance, galaxy clustering and galaxy-galaxy lensing observables with the Kilo-Degree Survey. Our theoretical model combines a flexible conditional stellar mass function, to describe the galaxy-halo connection, with a cosmological N-body simulation-calibrated halo model to describe the non-linear matter field. Our magnitude-limited bright galaxy sample combines 9-band optical-to-near-infrared photometry with an extensive and complete spectroscopic training sample to provide accurate redshift and stellar mass estimates. Our faint galaxy sample provides a background of accurately calibrated lensing measurements. We constrain the structure growth parameter $S_8 = \sigma_8 \sqrt{\Omega_m/0.3} = 0.773^{+0.028}_{-0.030}$, and the matter density parameter $\Omega_{\rm m} = 0.290^{+0.021}_{-0.017}$. The galaxy-halo connection model adopted in the work is shown to be in agreement with previous studies. Our constraints on cosmological parameters are comparable to, and consistent with, joint '3 × 2pt' clustering-lensing analyses that additionally include a cosmic shear observable. This analysis therefore brings attention to the significant constraining power in the often-excluded non-linear scales for galaxy clustering and galaxy-galaxy lensing observables. By adopting a theoretical model that accounts for non-linear halo bias, halo exclusion, scale-dependent galaxy bias and the impact of baryon feedback, this work demonstrates the potential and a way forward to include non-linear scales in cosmological analyses.

Key words. gravitational lensing: weak – methods: statistical – cosmological parameters – galaxies: haloes – dark matter – large-scale structure of Universe.

1. Introduction

In the last quarter of the century we have seen the rise and establishment of the concordance cosmological model which describes the formation and subsequent evolution of the cosmic structure. In this concordance model the Universe at present time is modelled as a flat, cold dark matter (CDM) and cosmological constant (Λ) dominated Universe with a negligible contribution from neutrinos, and gravity described by the law of General Relativity. Furthermore, the initial power spectrum of density fluctuations is assumed to be a single power-law. Such ΛCDM models are described using only six free parameters, which govern the energy densities of baryons, $\Omega_{\rm b}$, and cold dark matter, $\Omega_{\rm dm}$,

the spectral index, $n_{\rm s}$, and normalisation, σ_8, of the density perturbations power spectrum, the Hubble parameter, H_0, and the optical depth of reionisation. The flat geometry, implying that $\Omega_{\Lambda} = 1 - \Omega_{\rm b} - \Omega_{\rm dm}$, is strongly supported by high precision early-Universe measurements of the cosmic microwave background (CMB) temperature fluctuations combined with supernova and/or baryon acoustic oscillation distance measurements (Planck Collaboration et al. 2020; Scolnic et al. 2018; Alam et al. 2021). Formally the models also include the total neutrino mass, but the value of the parameter is too small for the current precision of observations (Gerbino & Lattanzi 2018).

A powerful probe of late-time cosmology is the large-scale distribution of galaxies. Even though the stars contribute neg-

ligibly to the overall energy density of the Universe, the light from stars in galaxies can be used to trace the evolution of the underlying distribution of dark matter in two complementary ways. Firstly, the light-path from distant galaxies is impacted by the distribution of foreground mass. This 'gravitational lensing' effect leads to a correlation between the observed shapes of galaxies, commonly referred to as *cosmic shear*. This observable can be used to probe the statistical properties of the total matter distribution in the Universe, typically quantified through the shape and amplitude of the matter power spectrum (Heymans et al. 2013; Hikage et al. 2019; Hamana et al. 2019; Asgari et al. 2021b; Secco et al. 2022; Amon et al. 2022a; Van Den Busch et al. 2022). Secondly, galaxies are expected to reside within dark matter haloes which form from the highest density peaks in the initial Gaussian random density field (e.g. Mo et al. 2010, and the references therein). Galaxies are therefore tracers of the underlying dark matter distribution, and with an accurate understanding of how biased these tracers are, the measurement of *galaxy clustering* as a function of redshift and scale provides strong constraints on the properties of the ΛCDM model (see for example Alam et al. 2021). It is becoming increasingly common to combine these two different 'two-point' (2pt) statistics, along with a third measurement of the gravitational lensing of background galaxies by foreground galaxies, otherwise known as *galaxy-galaxy lensing*. These joint '3×2pt' large-scale structure cosmological studies already have the precision to directly constrain some cosmological parameters independently of CMB measurements (Heymans et al. 2021; DES Collaboration et al. 2022).

In this analysis we focus on exploiting the significant precision recovered with small-scale measurements of *galaxy clustering* and *galaxy-galaxy lensing*. These non-linear scales are typically excluded from cosmological analyses owing to insufficient or uncertain modelling of the complex relationship between galaxies and the underlying matter distribution on these scales (Davis et al. 1985; Dekel & Rees 1987). *Galaxy bias* is scale-dependent, stochastic and changes as a function of galaxy luminosity, colour and morphological type (Dekel & Lahav 1999; Zehavi et al. 2011; Cacciato et al. 2012; Dvornik et al. 2018). Based on these facts, it is not surprising that galaxy bias is generally considered a nuisance to be marginalised over in the recovery of cosmological constraints. Many studies limit their analyses to scales where the galaxy bias is considered to be linear and scale independent (see for example van Uitert et al. 2018; Yoon et al. 2019; DES Collaboration et al. 2022). An alternative approach uses perturbation theory (Desjacques et al. 2018), to expand galaxy bias modelling into the mildly non-linear regime (e.g. Mandelbaum et al. 2013; Sánchez et al. 2017; Heymans et al. 2021; Pandey et al. 2022). However, as a measurement of small-scale galaxy bias also contains a wealth of information regarding galaxy formation, we argue that it is preferential to utilise all the data, along with an appropriate galaxy bias model, to facilitate joint constraints on both cosmology and galaxy bias.

Given our previous attempts to shine the light on the galaxy bias and its properties (Dvornik et al. 2018), in this analysis we adopt a realistic and physically motivated *halo model* for galaxy bias. Under the assumption that all galaxies reside in dark matter haloes, we adopt a halo occupation distribution (HOD) model, a statistical description for how galaxies are distributed between and within the dark matter haloes (Peacock & Smith 2000; Scoccimarro et al. 2001; Mo et al. 2010; Yang et al. 2009; Cacciato et al. 2013; van den Bosch et al. 2013; Cacciato et al. 2014). When combined with the halo model, which describes the non-linear matter distribution as a sum of spherical dark matter haloes (Seljak 2000; Cooray & Sheth 2002), these models provide a fairly complete, broadly accurate[1], and easy to understand description of galaxy bias, halo masses, and galaxy clustering (Cacciato et al. 2013).

Our approach builds on the cosmological analysis presented in Cacciato et al. (2013); More et al. (2015), where the halo model is used to coherently analyse the clustering of galaxies, the galaxy-galaxy lensing signal (Guzik & Seljak 2002; Yoo et al. 2006; Cacciato et al. 2009) and the galaxy abundances as a function of luminosity or stellar mass (van den Bosch et al. 2013; Cacciato et al. 2013). Furthermore, the same approach was used to study the galaxy-halo connection exclusively, with a fixed cosmology, by Leauthaud et al. (2011); Coupon et al. (2015). This combination of probes is hereafter referred to as '2×2pt+SMF', representing the combination of the two two-point statistics galaxy-galaxy lensing and galaxy clustering with the one-point stellar mass function.

Since early applications, there has been significant interest in using the halo model to interpret large-scale structure probes Seljak et al. (2005); Cacciato et al. (2009); Li et al. (2009). The analysis of 2×2pt statistics, down to non-linear scales, has been shown to lead to tight constraints for both $\Omega_{\rm m}$ and σ_8 (Cacciato et al. 2013; Mandelbaum et al. 2013; More et al. 2015; Wibking et al. 2019). The halo model can also constrain extensions to the standard ΛCDM cosmologies, such as the equation of state of dark energy and neutrino mass (More et al. 2013; Krause & Eifler 2017). The choice of observables is motivated by the focus on a feasibility study on smaller scales which achieves similar precision, thus allowing for a direct and/or independent comparison to cosmic shear studies. In the era of high-precision cosmology, however, Miyatake et al. (2020) showed that the use of only a 'broadly accurate' standard halo model leads to significant offsets in the recovered cosmological parameters from a 2×2pt analysis of HOD-populated numerical simulations. Consistency studies between the observed small-scale clustering and galaxy-galaxy lensing signals cast similar doubt on the accuracy of any standard halo model analysis (Leauthaud et al. 2017; Lange et al. 2021; Amon et al. 2022b).

Arguably the two most flawed approximations in the standard halo model formalism are that (i) haloes, and therefore galaxies, can overlap, and (ii) that haloes trace the matter distribution with a linear and scale independent halo bias. Previous attempts to improve these approximations have used halo exclusion formulations (Giocoli et al. 2010) to solve the first problem, combined with radial bias functions which add scale dependence to the halo bias model (Tinker et al. 2005; van den Bosch et al. 2013; Cacciato et al. 2013).

In this analysis we instead follow the method proposed by Mead & Verde (2021), accounting for both scale-dependent non-linear halo bias and halo exclusion by incorporating the halo bias measured directly from the DARKEMULATOR suite of cosmological simulations (Nishimichi et al. 2019). As shown by Mahony et al. (2022), this necessary upgrade to the standard halo model leads to sufficient accuracy in the recovered cosmological parameters from a 2×2pt+SMF analysis, for the statistical power of current imaging surveys.

Other approximations in a halo model analysis include that the halo mass is the sole variable that determines the properties of the haloes and their occupying galaxies. Galaxy properties and the clustering of haloes are, however, expected to have a

---

[1] The standard halo model of dark matter haloes estimates the true power spectrum to within 20% accuracy on non-linear scales (Mead et al. 2015).

secondary dependence, on their local environment and assembly history (see Wechsler & Tinker 2018, and references therein). Furthermore the adopted halo density profile is modelled from dark matter-only numerical simulations, even though hydrodynamical simulations show that these profiles are modified by the presence of active galactic nuclei (Schaller et al. 2015; Wang et al. 2020). In Debackere et al. (2021), it is shown that to account for baryon physics, it is sufficient to leave the concentration of dark matter haloes free. Amon et al. (2022b) review the literature studies on the impact of these two approximations on 2×2pt halo model studies with luminous red galaxies. Motivated by their conclusions, we choose to adopt nuisance parameters in our halo model to encapsulate the uncertainty of the impact of assembly bias and baryon feedback within our error budget.

In this paper, we analyse the most recent data release from the Kilo-Degree Survey (KiDS-1000, Kuijken et al. 2019; Giblin et al. 2021; Hildebrandt et al. 2021), spanning over 1000 square degrees of imaging in nine bands from the optical through to the near-infrared. Our main 'KiDS-Bright' galaxy sample (Bilicki et al. 2021) benefits from the 180 square degree overlap between KiDS and the spectroscopic Galaxy And Mass Assembly (GAMA) survey (Driver et al. 2011). As an essentially complete spectroscopic survey to $r < 19.8$, GAMA serves as an extensive training set for machine learning and the calibration of different sample selections. The resulting GAMA-trained photometric redshifts and stellar mass estimates for the KiDS-Bright sample have an enhanced accuracy and precision that benefits this galaxy-galaxy lensing and galaxy clustering study. In order to simultaneously constrain cosmology and galaxy bias, we use the 2×2pt+SMF combination of galaxy clustering and galaxy-galaxy lensing, as well as constraints on galaxy abundances in the form of the stellar mass function.

We improve previous related 2×2pt+SMF studies by (i) using a more accurate analytical model with the addition of non-linear halo bias (Mead & Verde 2021), taking into account the halo exclusion and scale dependence, (ii) taking the latest lensing and clustering data from a single survey, and (iii) using the full analytical covariance matrix including the cross-variance between all observables. Our analysis is highly complementary to the emulator based 2×2pt halo model analysis of the Hyper Suprime Camera Survey (HSC, Miyatake et al. 2021).

Throughout this paper, all radii and densities are given in comoving units, 'log' is used to refer to the 10-based logarithm, and 'ln' for natural logarithm. All the quantities that depend on the Hubble parameter adopt units of h, where $h = H_0/100$ km s^{-1}Mpc^{-1}. We also use $\overline{\rho}_m$ as the present day mean matter density of the Universe, $\overline{\rho}_m = \Omega_{m,0} \rho_{crit}$, where $\rho_{crit} = 3H_0^2/(8\pi G)$ and the halo masses are defined as $M = 4\pi r_\Delta^3 \Delta \, \overline{\rho}_m/3$ enclosed by the radius r_Δ within which the mean density of the halo is Δ times $\overline{\rho}_m$, with $\Delta = 200$.

This paper is organised as follows. In Sec. 2, we review our analytical model to compute the galaxy stellar mass function, the galaxy-galaxy correlation function, and the galaxy-galaxy lensing signal using the halo model combined with a model that describes halo occupation statistics as a function of galaxy stellar mass. In Sec. 3, we introduce the 2×2pt+SMF KiDS measurements, specifics of the covariance calculation and our Bayesian analysis methodology. Our main results are presented in Sec. 4, concluding our findings in Sec. 5.

2. The Halo model

The halo model is an analytic framework that can be used to describe the clustering of matter and its evolution in the Universe (Seljak 2000; Peacock & Smith 2000; Cooray & Sheth 2002; van den Bosch et al. 2013; Mead et al. 2015). It is built upon the statistical description of the properties of dark matter haloes (namely the average density profile, large scale bias and abundance) as well as on the statistical description of the galaxies residing in them, using halo occupation distributions (HOD). The model is sufficiently flexible to consistently describe the statistical weak lensing signal around a selection of galaxies, their clustering, abundances and cosmic shear signal.

2.1. Halo model ingredients

We assume that dark matter haloes are spherically symmetric on average, and have density profiles, $\rho(r|M) = M \, u_h(r|M)$, that depend only on their mass M, and $u_h(r|M)$ is the normalised density profile of a dark matter halo. Similarly, we assume that satellite galaxies in haloes of mass M follow a spherical number density distribution $n_s(r|M) = N_s \, u_s(r|M)$, where $u_s(r|M)$ is the normalised density profile of satellite galaxies. All central galaxies are positioned at the centre of their halo: $r = 0$. We assume that the density profile of dark matter haloes follows a Navarro-Frenk-White (NFW) profile (Navarro et al. 1997). Since centrals and satellites are distributed differently, we write the galaxy-galaxy power spectrum, $P_{gg}(k)$, as a combination of the central 'c', satellite 's' and cross power spectrum, with

$$P_{gg}(k) = A_c^2 P_{cc}(k) + 2A_c A_s P_{cs}(k) + A_s^2 P_{ss}(k) \,, \tag{1}$$

and the galaxy-matter power spectrum, $P_{gm}(k)$,

$$P_{gm}(k) = A_c P_{cm}(k) + A_s P_{sm}(k) \,. \tag{2}$$

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

$$\overline{n}_x = \int_0^\infty \langle N_x|M \rangle \, n(M) \, dM \,, \tag{3}$$

where 'x' stands for 'g' (for galaxies), 'c' (for centrals) or 's' (for satellites), $\langle N_x|M \rangle$ is the average number of galaxies given halo mass M, and $n(M)$ is the halo mass function in the following form:

$$n(M) = \frac{\overline{\rho}_m}{M^2} \nu f(\nu) \frac{d\ln\nu}{d\ln M} \,, \tag{4}$$

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

$$P_{cc}^{1h}(k) = \frac{1}{\overline{n}_c} \,, \tag{5}$$

$$P_{ss}^{1h}(k) = \mathcal{P} \int_0^\infty \mathcal{H}_s^2(k, M) \, n(M) \, dM \,, \tag{6}$$

and all other terms are given by:

$$P_{xy}^{1h}(k) = \int_0^\infty \mathcal{H}_x(k, M) \, \mathcal{H}_y(k, M) \, n(M) \, dM \,. \tag{7}$$

Here 'x' and 'y' are either 'c' (central), 's' (satellite), or 'm' (matter), $\mathcal{P}$ is a Poisson parameter that captures the scatter in the number of satellite galaxies at fixed halo mass (in this case a free parameter – we define the $\mathcal{P}$ in detail using equations 24 and 25) and we have defined the mass, central and satellite profiles as

$$\mathcal{H}_\mathrm{m}(k, M) = \frac{M}{\bar{\rho}_\mathrm{m}} \, \tilde{u}_\mathrm{h}(k|M) \,, \tag{8}$$

$$\mathcal{H}_\mathrm{c}(k, M) = \frac{\langle N_\mathrm{c}|M \rangle}{\bar{n}_\mathrm{c}} \,, \tag{9}$$

and

$$\mathcal{H}_\mathrm{s}(k, M) = \frac{\langle N_\mathrm{s}|M \rangle}{\bar{n}_\mathrm{s}} \, \tilde{u}_\mathrm{s}(k|M) \,, \tag{10}$$

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

$$
\begin{aligned}
P_\mathrm{xy}^\mathrm{2h}(k) = P_\mathrm{lin}(k) \int_0^\infty &\mathrm{d}M_1 \, \mathcal{H}_\mathrm{x}(k, M_1) \, b_\mathrm{h}(M_1) \, n(M_1) \\
\times \int_0^\infty &\mathrm{d}M_2 \, \mathcal{H}_\mathrm{y}(k, M_2) \, b_\mathrm{h}(M_2) \, n(M_2) \\
&+ P_\mathrm{lin}(k) \, I_\mathrm{xy}^\mathrm{NL}(k)
\end{aligned}
\tag{11}
$$

where $P_\mathrm{lin}(k)$ is the linear power spectrum, obtained using the Eisenstein & Hu (1998) matter transfer function, and $b_\mathrm{h}(M, z)$ is the halo bias function. We adopt the Tinker et al. (2010) halo bias function which together with their halo mass function provides for consistent normalisation of the halo model integrals. The second term in equation 11 encompasses the beyond-linear halo bias correction β^NL proposed by Mead & Verde (2021) where,

$$
\begin{aligned}
I_\mathrm{xy}^\mathrm{NL}(k) = \int_0^\infty \int_0^\infty &\mathrm{d}M_1 \mathrm{d}M_2 \, \beta^\mathrm{NL}(k, M_1, M_2) \\
&\times \mathcal{H}_\mathrm{x}(k, M_1) \, \mathcal{H}_\mathrm{y}(k, M_2) \\
&\times n(M_1) \, n(M_2) \, b_\mathrm{h}(M_1) \, b_\mathrm{h}(M_2) \,.
\end{aligned}
\tag{12}
$$

β^NL is measured using the DarkQuest emulator (Nishimichi et al. 2019; Miyatake et al. 2020; Mahony et al. 2022), by measuring the non-linear halo-halo power spectrum and then dividing it by the linear matter power spectrum multiplied with the product of linear bias factors (Mead & Verde 2021, Eq. 23). Due to the definition of β^NL, this measurement also holds true for galaxy-galaxy and galaxy-matter correlations. As shown in Mahony et al. (2022), this function is cosmology dependent, but does not account for assembly bias effects. In this paper, owing to the volume-limited mix of all galaxies used in our analysis, we consider any assembly bias to be a subdominant effect. Numerically, the integrals in the halo model are not integrated from zero to infinity, but rather between a wide range of halo masses. Special care has to be taken to account for the masses outside of the integration limits, for which an appropriate correction is applied (as derived in Cacciato et al. (2009), Eqs. 24 and 25, and in Mead et al. (2020); Mead & Verde (2021), Appendices A in both papers). The two-point correlation functions corresponding to these power-spectra are obtained by Fourier transformation:

$$\xi_\mathrm{xy}(r) = \frac{1}{2\pi^2} \int_0^\infty P_\mathrm{xy}(k) \frac{\sin kr}{kr} \, k^2 \, \mathrm{d}k \,. \tag{13}$$

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

$$c(M, z) = 10.14 \left[\frac{M}{(2 \times 10^{12} M_\odot/h)} \right]^{-0.081} (1 + z)^{-1.01} \,. \tag{14}$$

We allow for an additional normalisation $f_\mathrm{h,s}$, such that

$$c_\mathrm{h,s}(M, z) = f_\mathrm{h,s} \, c(M, z) \,, \tag{15}$$

where f_h is the normalisation of the concentration-mass relation for dark matter haloes $\tilde{u}_\mathrm{h}(k|M)$, and f_s is the normalisation of the concentration-mass relation for the distribution of satellite galaxies $\tilde{u}_\mathrm{s}(k|M)$. Our adoption here of separate concentration-mass relations for dark matter haloes and satellite galaxies provides enough flexibility in the model to capture the uncertain impact of baryon feedback (for the scales adopted, Debackere et al. 2020, 2021; Amon et al. 2022b), and it has been used in the literature (Cacciato et al. 2013; Viola et al. 2015; van Uitert et al. 2016; Dvornik et al. 2018) to account for such effects. This additional flexibility is motivated by the fact that in the simulations, the AGN feedback pushes the baryons and dark matter from halo centres towards outskirts, and by that effectively changing the concentration of the matter distribution (Debackere et al. 2020; Mead et al. 2020). This is also supported by observations (Viola et al. 2015), which showed that the preferred value for the concentration normalisation is lower than 1. Using the halo model with these extra parameters is a benefit over the emulators that are based on dark-matter only simulations (as for instance the DarkQuest emulator Nishimichi et al. 2019; Miyatake et al. 2020; Mahony et al. 2022), since they do not offer a simple way to accommodate for such flexibility, nor require simulations (Schneider & Teyssier 2015).

In the halo model we do not consider the mis-centred central term, as for a selection of galaxies the signature is accounted for through the terms for satellite galaxies, which do not reside in the centres of haloes by definition. What is more, the satellite galaxies populate haloes regardless of the existence of central galaxy, which further removes the need for a mis-centred term.

2.2. Conditional stellar mass function

We model the galaxy stellar mass function and halo occupation statistics using the Conditional Stellar Mass Function (CSMF, motivated by Yang et al. 2008; Cacciato et al. 2009, 2013; Wang et al. 2013; van Uitert et al. 2016). The CSMF, $\Phi(M_\star|M)$, specifies the average number of galaxies of stellar mass $M_\star$ that reside in a halo of mass M. In this formalism, the halo occupation statistics of central galaxies are defined via the function:

$$\Phi(M_\star|M) = \Phi_\mathrm{c}(M_\star|M) + \Phi_\mathrm{s}(M_\star|M) \,. \tag{16}$$

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

$$\Phi_\mathrm{c}(M_\star|M) = \frac{1}{\sqrt{2\pi} \ln(10) \, \sigma_\mathrm{c} M_\star} \exp\left[-\frac{\log(M_\star/M_\mathrm{c}^*)^2}{2 \, \sigma_\mathrm{c}^2} \right] \,, \tag{17}$$

and the satellite term as a modified Schechter function,

$$\Phi_\mathrm{s}(M_\star|M) = \frac{\phi_\mathrm{s}^*}{M_\mathrm{s}^*} \left(\frac{M_\star}{M_\mathrm{s}^*} \right)^{\alpha_\mathrm{s}} \exp\left[-\left(\frac{M_\star}{M_\mathrm{s}^*} \right)^2 \right] \,, \tag{18}$$

where σ_c is the scatter between stellar mass and halo mass and α_s governs the power law behaviour of satellite galaxies. Note that M_c^*, σ_c, ϕ_s^*, α_s and M_s^* are, in principle, all functions of halo mass M, but here we assume that σ_c and α_s are independent of the halo mass M. Inspired by Yang et al. (2008), who studied the halo occupation properties of galaxies in the Sloan Digital Sky Survey, we parametrise M_c^*, M_s^* and ϕ_s^* as:

$$M_c^*(M) = M_0 \frac{(M/M_1)^{\gamma_1}}{[1 + (M/M_1)]^{\gamma_1 - \gamma_2}}, \tag{19}$$

$$M_s^*(M) = 0.56 \, M_c^*(M), \tag{20}$$

and

$$\log[\phi_s^*(M)] = b_0 + b_1(\log m_{13}), \tag{21}$$

where $m_{13} = M/(10^{13} M_\odot h^{-1})$. In their analysis of the stellar-to-halo mass relation of GAMA galaxies van Uitert et al. (2016) find that varying the pre-factor of 0.56 in Equation 20 does not significantly affect the results, therefore we retain this normalisation in our analysis. We can see that the stellar to halo mass relation for $M \ll M_1$ behaves as $M_c^* \propto M^{\gamma_1}$ and for $M \gg M_1$, $M_c^* \propto M^{\gamma_2}$, where M_1 is a characteristic mass scale and M_0 is a normalisation. Here γ_1, γ_2, b_0 and b_1 are all free parameters that govern the two slopes of the stellar-to-halo mass relation and the normalisation of the Schechter function. The choice of functional form of the CSMF is motivated by the good performance as seen in previous lensing and combined lensing and clustering studies. In Eq. 19, we adopt an effective stellar-to-halo mass relation for our mixed-population of red and blue galaxies. Bilicki et al. (2021) demonstrate a strong colour-dependence to this relationship, and future studies will investigate including a red/blue galaxy split in our analysis, which can also help to improve the modelling of intrinsic galaxy alignments (e.g. Li et al. 2021).

From the CSMF it is straightforward to compute the galaxy stellar mass function (SMF) and the halo occupation numbers. The galaxy stellar mass function is in this case given by

$$\Phi(M_\star) = \int_0^\infty \Phi(M_\star|M) \, n(M) \, dM, \tag{22}$$

and the average number of galaxies with stellar masses in the range $M_{\star,1} \leq M_\star \leq M_{\star,2}$ is given by:

$$\langle N|M \rangle = \int_{M_{\star,1}}^{M_{\star,2}} \Phi(M_\star|M) \, dM_\star. \tag{23}$$

In order to predict the satellite-satellite term for the galaxy clustering power spectra (Eq. 6), we use that

$$\langle N_s^2|M \rangle = \mathcal{P}(M)\langle N_s|M \rangle^2 + \langle N_s|M \rangle, \tag{24}$$

where $\mathcal{P}(M)$ is the mass-dependent Poisson parameter defined as:

$$\mathcal{P}(M) \equiv \frac{\langle N_s(N_s - 1)|M \rangle}{\langle N_s|M \rangle^2}, \tag{25}$$

which is unity if $\langle N_s|M \rangle$ is given by a Poisson distribution, larger than unity if the distribution is wider than a Poisson distribution (also called super-Poissonian distribution) or smaller than unity if the distribution is narrower than a Poisson distribution (also called sub-Poissonian distribution). We limit ourselves to cases in which $\mathcal{P}(M)$ is independent of halo mass, i.e., $\mathcal{P}(M) = \mathcal{P}$, and we treat $\mathcal{P}$ as a free parameter.

Overall, all the free parameters used to describe the halo occupation distributions and the connection with the dark matter are:

$$\lambda^{\mathrm{HOD}} = [f_h, M_0, M_1, \gamma_1, \gamma_2, \sigma_c, f_s, \alpha_s, b_0, b_1, \mathcal{P}]. \tag{26}$$

Priors on these parameters are broad, assuming wide uniform distributions, similar to the priors used in two studies of the galaxy-halo connection that both used GAMA and KiDS data (van Uitert et al. 2016; Dvornik et al. 2018). Priors and their ranges can be found in Table 2.

2.3. Projected lensing and clustering functions

Once $P_{gg}(k)$ and $P_{gm}(k)$ have been determined, it is fairly straightforward to compute the projected galaxy-galaxy correlation function, $w_p(r_p)$, and the excess surface density (ESD) profile, $\Delta\Sigma(r_p)$. The projected galaxy-galaxy correlation function, $w_p(r_p)$, is related to the real-space galaxy-galaxy correlation function, $\xi_{gg}(r)$, according to

$$\begin{aligned} w_p(r_p) &= 2 \int_0^{r_{\pi,\mathrm{max}}} \xi_{gg}(r_p, r_\pi, z) \, dr_\pi \\ &= 2 \sum_{l=0}^{2} \int_0^{r_{\pi,\mathrm{max}}} \xi_{2l}(s, z) \, \mathcal{L}_{2l}(r_\pi/s) \, dr_\pi. \end{aligned} \tag{27}$$

Here $\xi_{gg}(r_p, r_\pi, z)$ is the redshift-space galaxy-galaxy correlation function, r_π is the redshift-space separation perpendicular to the line-of-sight and $r_{\pi,\mathrm{max}}$ is the maximum integration range used for the data (here we use $r_{\pi,\mathrm{max}} = 233 h^{-1}\mathrm{Mpc}$), $s = \sqrt{r_p^2 + r_\pi^2}$ is the separation between the galaxies, $\mathcal{L}_l(x)$ is the lth Legendre polynomial, and ξ_0, ξ_2, and ξ_4 are given by

$$\xi_0(r, z) = \left(1 + \frac{2}{3}\beta_k + \frac{1}{5}\beta_k^2\right) \xi_{gg}(r, z), \tag{28}$$

$$\xi_2(r, z) = \left(\frac{4}{3}\beta_k + \frac{4}{7}\beta_k^2\right) \left[\xi_{gg}(r, z) - 3J_3(r, z)\right], \tag{29}$$

$$\xi_4(r, z) = \frac{8}{35}\beta_k^2 \left[\xi_{gg}(r, z) + \frac{15}{2}J_3(r, z) - \frac{35}{2}J_5(r, z)\right], \tag{30}$$

where

$$J_n(r, z) = \frac{1}{r^n} \int_0^r \xi_{gg}(y, z) \, y^{n-1} \, dy. \tag{31}$$

and

$$\beta_k = \beta_k(z) = \frac{1}{\bar{b}(z)} \left(\frac{d\ln D(z)}{d\ln a}\right)_z \tag{32}$$

with $a = 1/(1 + z)$ the scale factor, $D(z)$ the linear growth factor, and

$$\bar{b}(z) = \frac{1}{\bar{n}_g(z)} \int_0^\infty \langle N_g|M \rangle \, b_h(M, z) \, n(M, z) \, dM, \tag{33}$$

the mean bias of the galaxies in consideration. Eq. 27 accounts for the large-scale redshift-space distortions due to infall (the 'Kaiser'-effect), which is necessary because the measurements for $w_p(r_p)$ are obtained for a finite r_{max}. We note that whilst this Kaiser (1987) formalism is only strictly valid in the linear

regime, we adopt the non-linear galaxy-galaxy correlation function, $\xi_{gg}(r)$, in Eqs. 28 - 30, with the non-linearities captured through the halo model power spectra in Eq. 13. van den Bosch et al. (2013) show that this modification provides a more accurate correction for the residual redshift space distortions, and that ignoring the presence of residual redshift space distortions leads to systematic errors that can easily exceed 20 percent on scales with $r_p > 10h^{-1}$ Mpc (Cacciato et al. 2013).

The excess surface density profile (ESD), $\Delta\Sigma(r_p)$, is defined as

$$\Delta\Sigma(r_p) = \frac{2}{r_p^2} \int_0^{r_p} \Sigma(R')\, R'\, \mathrm{d}R' - \Sigma(r_p)\,. \tag{34}$$

Here $\Sigma(r_p)$ is the projected surface mass density, which is related to the galaxy-dark matter cross correlation, $\xi_{gm}(r)$, according to

$$\Sigma(r_p) = 2\bar{\rho}_m \int_{r_p}^{\infty} \xi_{gm}(r)\, \frac{r\, \mathrm{d}r}{\sqrt{r^2 - r_p^2}}\,. \tag{35}$$

2.4. Cosmological parameters

The cosmological parameters in our model are described by the vector:

$$\lambda^{cosmo} = [\Omega_m, \sigma_8, h, n_s, \Omega_b]\,. \tag{36}$$

As mentioned in Sect. 1, the goal of this paper is to use the ESD, w_p and SMF data to constrain σ_8 and Ω_m. Because of that, we set the priors for those two parameters to be uninformative and set their ranges following the latest KiDS cosmic shear analysis (Asgari et al. 2021b). The last 3 cosmological parameters are shown to be poorly constrained using the ESD, w_p, and SMF data (Cacciato et al. 2013; Mandelbaum et al. 2013), thus they form a set of secondary cosmological parameters with informative priors. Priors and their ranges can be found in Table 2. In Appendix B we verify that our choice of priors do not inform the main cosmological parameters.

3. Data and sample selection

In this analysis we combined three observables from the Kilo-Degree Survey (KiDS): galaxy abundances in the form of the galaxy stellar mass function, galaxy clustering in the form of the projected galaxy correlation function, and galaxy-galaxy lensing in the form of excess surface density profiles. Our KiDS observations were taken with OmegaCAM (Kuijken 2011), a 268-million pixel CCD mosaic camera mounted on the VLT Survey Telescope. These instruments were designed to perform weak lensing measurements, with the camera and telescope combination providing a fairly uniform point spread function across the field-of-view (de Jong et al. 2013).

We analysed the latest data release of the KiDS survey (KiDS-1000, Kuijken et al. 2019), containing observations from 1006 square-degree survey tiles. Specifics of the survey, the calibration of the source shapes and photometric redshifts are described in Kuijken et al. (2019); Giblin et al. (2021); Hildebrandt et al. (2021), respectively. The companion VISTA-VIKING (Edge et al. 2013) survey has provided complementary imaging in near-infrared bands ($ZYJHK_s$), resulting in a unique deep, wide, nine-band imaging dataset (Wright et al. 2019). The default photo-z estimates provided as part of the KiDS survey were derived with the Bayesian Photometric Redshift approach (BPZ, Benitez 2000).

We used shape measurements based on the r-band images, which have an average seeing of 0.66 arcsec. The galaxy shapes were measured using *lens*fit (Miller et al. 2013), which has been calibrated using image simulations described in Kannawadi et al. (2019). This provides galaxy ellipticities (ϵ_1, ϵ_2) with respect to an equatorial coordinate system, and an optimal weight.

The galaxies used for our lens and clustering sample were taken from the 'KiDS-Bright' sample (Bilicki et al. 2021). This sample mimics the selection of GAMA galaxies (Driver et al. 2011), by applying the condition $m_r < 20.0$. For these galaxies a different method of determining the photometric redshifts was employed using the ANNz2 (Artificial Neural Network) machine learning method (Sadeh et al. 2016), with the spectroscopic GAMA survey, which is 98.5% complete to $r < 19.8$, as a training set (Bilicki et al. 2018, 2021). Comparing the obtained redshifts with the spectroscopic redshifts from the matched galaxies between KiDS-Bright and GAMA, Bilicki et al. (2021) concluded that the ANNz2 photo-z are highly accurate with a mean offset of $\delta_z = 5 \times 10^{-4}$, and a scaled mean absolute deviation scatter of $\sigma_z = 0.018(1 + z)$.

Stellar mass estimates for the KiDS-Bright sample are obtained using the LePHARE template fitting code (Arnouts et al. 1999; Ilbert et al. 2006). In these fits, ANNz2 photo-z estimates are used as input redshifts for each source, treating them as if they were exact, neglecting the percent error associated with the ANNz2 redshift. In practice, this error has little impact on the fidelity of the stellar mass estimates (Taylor et al. 2011). The estimates assume a Chabrier (2003) initial mass function, the Calzetti et al. (1994) dust-extinction law, Bruzual & Charlot (2003) stellar population synthesis models, and exponentially declining star formation histories. The input photometry to LePHARE is extinction corrected using the Schlegel et al. (1998) maps with the Schlafly & Finkbeiner (2011) coefficients, as described in Kuijken et al. (2019).

Bilicki et al. (2021) found that the KiDS-Bright stellar mass estimates are in excellent agreement with independent stellar mass estimates from Wright et al. (2016) that combine GAMA spectroscopic redshifts with multi-wavelength imaging from 21 broadband filters from the far-UV to the far-IR. The median offset is $M_\star^{KiDS}/M_\star^{GAMA} = -0.09 \pm 0.18$ dex. Brouwer et al. (2021) estimated the overall systematic uncertainty on the stellar mass estimates of the KiDS-Bright sample, combining the uncertainty arising from the LePHARE model fit, the photometric redshift scatter, and the difference found when exchanging elliptical aperture magnitudes for Sérsic model magnitudes. They estimated an overall uncertainty of $\sigma_{M_*} = 0.12$ dex for the KiDS-Bright sample. This systematic uncertainty also includes the estimated Eddington systematic bias of ~ 0.027 dex (Brouwer et al. 2021), which is estimated from the population of red and blue galaxies and it is considered a worst-case scenario. We choose to account for both statistical and systematic uncertainty in the stellar mass estimates through the nuisance parameter σ_c, in equation 17, which provides the freedom to model both the intrinsic and measurement noise scatter in the stellar-to-halo mass relation (Leauthaud et al. 2012; Bilicki et al. 2021). Furthermore, as the systematic and statistical uncertainties are comparable in power, the entries in the SMF and cross-covariances are inflated by a factor of 2 to account for the uncertainty arising from Eddington bias and the systematic shift in stellar masses, and not only through the σ_c parameter. Due to the weak cosmology dependence of the SMF, this primarily increases only the uncertainty of our HOD parameters, as the SMF is in the first place used to break degeneracies in our HOD part of the halo model.

3.1. Stellar mass function measurements and sample selection: SMF

Our SMF measurements are performed using the maximum-volume weighting method (Schmidt 1968; Saunders et al. 1990; Cole 2011; Baldry et al. 2012; Wright et al. 2017). We weight each galaxy i by the inverse of the comoving volume over which the galaxy would be visible, given the magnitude limit of the whole sample, $1/V_{\max,i}$. To estimate the number density $\Phi(M_\star)$, we have to derive $M_{\star,\lim}(z)$, the completeness in stellar mass as a function of redshift for our flux-limited sample. For the $1/V_{\max}$ technique, we need to know $z_{\max,i}$, the maximum redshift beyond which galaxy i with stellar mass $M_{\star,i}$ would no longer be part of the subsample (Weigel et al. 2016). This is done by determining the point at which the sample begins to become incomplete. Usually this process contains a potentially biased visual inspection. To avoid any bias, we instead adopt the automated method presented by Wright et al. (2017), using the MassFunc-FrrR package. The algorithm estimates the turn-over point of the number density distribution in bins of comoving distance and stellar mass independently. In each fine bin of comoving distance, we take the mass at the peak density as the mass turn-over point. In each fine bin of stellar mass, we take the largest comoving distance at median stellar mass density as the distance turn-over point. The obtained turn-over points are then fit with a high-degree polynomial resulting in a smooth form for the stellar mass limit as a function of redshift. This limit can be compared to the $M_* - z_{\mathrm{ANNz2}}$ distribution of the full KiDS-Bright galaxies in Fig. 1.

In Fig. 2 we present the stellar mass function of the volume-limited KiDS-Bright sample from the galaxies in the 6 stellar mass bins, $\Phi(M_\star)$, which has a median redshift of $z = 0.25$. This is determined from the galaxy counts within the stellar mass limit, with errors derived analytically in Appendix A. We find good agreement between the KiDS measurement and the stellar mass function from Wright et al. (2018), evaluated at the median redshift of our sample. Wright et al. (2018) is based on from an analysis using spectroscopic data from GAMA, COSMOS and HST. This comparison therefore demonstrates the agreement in the SMF between spectroscopic data and our photometric KiDS-Bright sample of galaxies, demonstrating that our stellar mass estimates are robust to the uncertainty in the photometric redshifts (Taylor et al. 2011; Bilicki et al. 2021; Brouwer et al. 2021).

As galaxy bias is inherently dependent on the stellar mass of the galaxy (Dvornik et al. 2018), we analyse the weak lensing and galaxy clustering of the KiDS-Bright galaxies grouped into 6 stellar mass bins. We choose to limit our analysis to galaxies within the stellar mass range of $9.1 < \log(M_\star/h^{-2}\,M_\odot) \le 11.3$, with the number of bins, and bin limits chosen in such a way to achieve a similar and significant signal-to-noise ratio in all bins. Using the redshift-dependent stellar mass limit, we define upper redshift bounds to ensure each stellar mass bin is volume-limited, as indicated with red boxes in Fig. 1. The lower redshift bound is set to contain 95 percent of the volume-limited sample. The number of galaxies, median stellar mass and redshift of each bin is reported in Table 1.

3.2. Galaxy-galaxy lensing measurement: ESD

As shown in Bilicki et al. (2021), the excellent ANNz2 photometric redshift estimates for the galaxies in the KiDS-Bright sample allow for robust estimates of their physical characteristics, in particular the stellar mass. In this section we combine this
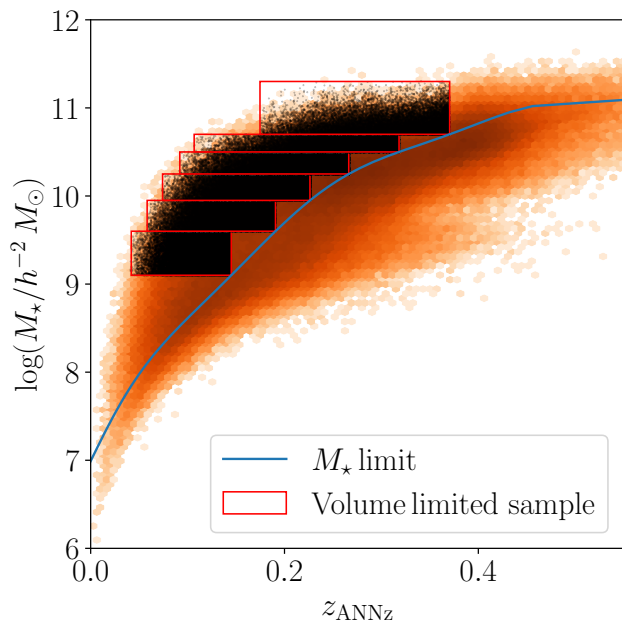

Fig. 1: Galaxy stellar mass as a function of ANNz2 photometric redshift for the KiDS-Bright sample. The full sample is shown with a logarithmic hexagonal density plot. The blue line shows the stellar mass limit determined using the automated method presented by Wright et al. (2017). Red boxes show the six stellar mass bins used in the analysis, with individual galaxies plotted as black dots. The bin ranges were chosen in such a way as to achieve a good signal-to-noise ratio in all bins for our galaxy-galaxy lensing and galaxy clustering measurements.

Table 1: KiDS-Bright stellar mass samples: overview of the number of galaxies/lenses, median stellar masses $M_{\star,\mathrm{med}}$ and median redshifts z_{med}. Stellar masses are given in units of $\log(M_\star/h^{-2}\,M_\odot)$. The final sample of galaxies used is a small subsample of all KiDS-Bright galaxies (~ 1 million).

Bin	Range	N_{gal}	$M_{\star,\mathrm{med}}$	z_{med}
1	(9.1, 9.6]	32 846	9.33	0.12
2	(9.6, 9.95]	35 559	9.77	0.15
3	(9.95, 10.25]	39 487	10.09	0.18
4	(10.25, 10.5]	38 544	10.36	0.22
5	(10.5, 10.7]	28 814	10.58	0.27
6	(10.7, 11.3]	22 560	10.79	0.32

information with accurate shape measurements for more distant KiDS sources from Giblin et al. (2021) to measure the galaxy-galaxy lensing signal. To quantify the weak gravitational lensing signal we use source galaxies from KiDS DR4 with a BPZ photo-z in the range $0.1 < z_{\mathrm{B}} < 1.2$.

The lensing signal of an individual lens is too small to be detected, and hence we compute a weighted average of the tangential ellipticity ϵ_{t} as a function of projected distance r_{p} using a large number of lens-source pairs. In the weak lensing regime this provides an unbiased estimate of the tangential shear, γ_{t}, which in turn can be related to the excess surface density (ESD) $\Delta\Sigma(r_{\mathrm{p}})$, defined as the difference between the mean projected surface mass density inside a projected radius r_{p} and the mean sur-
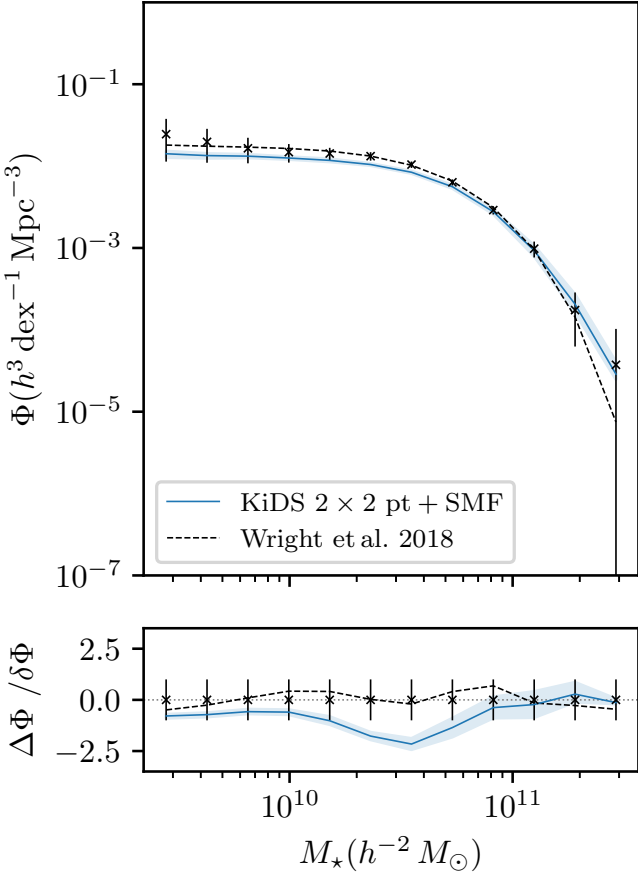
Fig. 2: *Upper panel:* The KiDS-Bright galaxy stellar mass function (crosses) compared to the model from Wright et al. (2018), evaluated at the median redshift of our sample (black dashed). The blue line and shaded region indicates the best fit model and 68% confidence levels of our best fit halo model (Eq. 22). We caution that the quality of the fit cannot be judged by eye, because of the covariance in the data, between the data points and between the other observables. The reduced χ^2 value for this observable is 1.05 (DoF = 14.58, p-value = 0.39), estimated using the method presented in Appendix C. *Lower panel:* The fractional errors on the data and the model, $\Delta\Phi/\delta\Phi$.

face density at r_p (as in Eq. 34, for more details see appendix C from Dvornik et al. 2018).

We compute a weighted average to account for the variation in the precision of the shear estimate, captured by the *lens*fit weight w_s (see Fenech Conti et al. 2017; Kannawadi et al. 2019, for details), and the fact that the amplitude of the lensing signal depends on the source redshift. The weight assigned to each lens-source pair is

$$\widetilde{w}_{ls} = w_s \left(\widetilde{\Sigma}_{cr,ls}^{-1} \right)^2 , \qquad (37)$$

the product of the *lens*fit weight w_s and the square of $\widetilde{\Sigma}_{cr,ls}^{-1}$ – the effective inverse critical surface mass density, which is a geometric term that downweights lens-source pairs that are close in redshift (e.g. Bartelmann & Schneider 2001).

We compute the effective inverse critical surface mass density for each lens using the photo-z of the lens z_l and the full normalised redshift probability density of the sources, $n(z_s)$. The latter is calculated employing the self-organising map calibra-

tion method (Wright et al. 2020) as applied to KiDS DR4 in Hildebrandt et al. (2021). The resulting effective inverse critical surface density can be written as:

$$\widetilde{\Sigma}_{cr,ls}^{-1} = \frac{4\pi G}{c^2} \int_0^\infty (1+z_l)^2 D(z_l) \left(\int_{z_l}^\infty \frac{D(z_l, z_s)}{D(z_s)} n(z_s) \, dz_s \right) p(z_l) \, dz_l , \qquad (38)$$

where $D(z_l)$, $D(z_s)$, $D(z_l, z_s)$ are the angular diameter distances to the lens, source, and between the lens and the source, respectively. For the lens redshifts z_l we use the ANNz2 photo-z of the KiDS-Bright foreground galaxy sample. We implement the contribution of z_l by integrating over the redshift probability distributions $p(z_l)$ of each lens. The lensing kernel is wide and therefore the resulting ESD signals are not sensitive to the small wings of the lens redshift probability distributions. We can thus safely approximate $p(z_l)$ as a normal distribution centred at the lenses photo-z, with a standard deviation $\sigma_z/(1+z_l) = 0.018$ (Bilicki et al. 2021).

For the source redshifts z_s we follow the method used in Dvornik et al. (2018), by integrating over the part of the redshift probability distribution $n(z_s)$ where $z_s > z_l$. Thus, the ESD can be directly computed in bins of projected distance r_p to the lenses as:

$$\Delta\Sigma_{gm}(r_p) = \left[\frac{\sum_{ls} \widetilde{w}_{ls} \epsilon_{t,s} \Sigma'_{cr,ls}}{\sum_{ls} \widetilde{w}_{ls}} \right] \frac{1}{1 + \overline{m}} . \qquad (39)$$

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

$$\overline{m} = \frac{\sum_i w'_i m_i}{\sum_i w'_i} , \qquad (40)$$

is an average correction to the ESD profile that has to be applied to account for the multiplicative bias m in the *lens*fit shear estimates. The sum goes over thin redshift slices for which m_i is obtained using image simulations (Kannawadi et al. 2019), weighted by $w' = w_s D(z_l, z_s)/D(z_s)$ for a given lens-source sample. The value of $\overline{m}$ is -0.003 for the 6 stellar mass bins, independent of the scale at which it is computed.

We note that the measurements presented here are not corrected for the contamination of the source sample by galaxies that are physically associated with the lenses (the so-called 'boost correction'). The impact on $\Delta\Sigma$ is minimal, because of the weighting with the inverse square of the critical surface density in Eq. (38), (see for instance the bottom panel of fig. A4 in Dvornik et al. 2017) and the removal of the sources physically associated with the lens from our signal measurements. The effect of using photometric lenses in the ESD measurements is directly accounted for in our estimator and the covariance matrix. We subtract the signal around random points, which suppresses any large-scale systematics and sample variance (Singh et al. 2017). This empirical 'random' correction for large-scale sample variance has been shown to improve robustness on the measurement scales which are particularly relevant to constrain linear bias (Dvornik et al. 2018). We find the random correction for the KiDS-Bright sample becomes significant at scales $R \gtrsim 3h^{-1}$ Mpc, rising to more than 100% of the ESD signal in the three lowest stellar mass bins. On these large scales the random correction is more than four times larger than the statistical uncertainty (see Appendix D for details). The resulting random-corrected galaxy-galaxy lensing ESD measurements for the six stellar mass bins are shown in Fig. 3
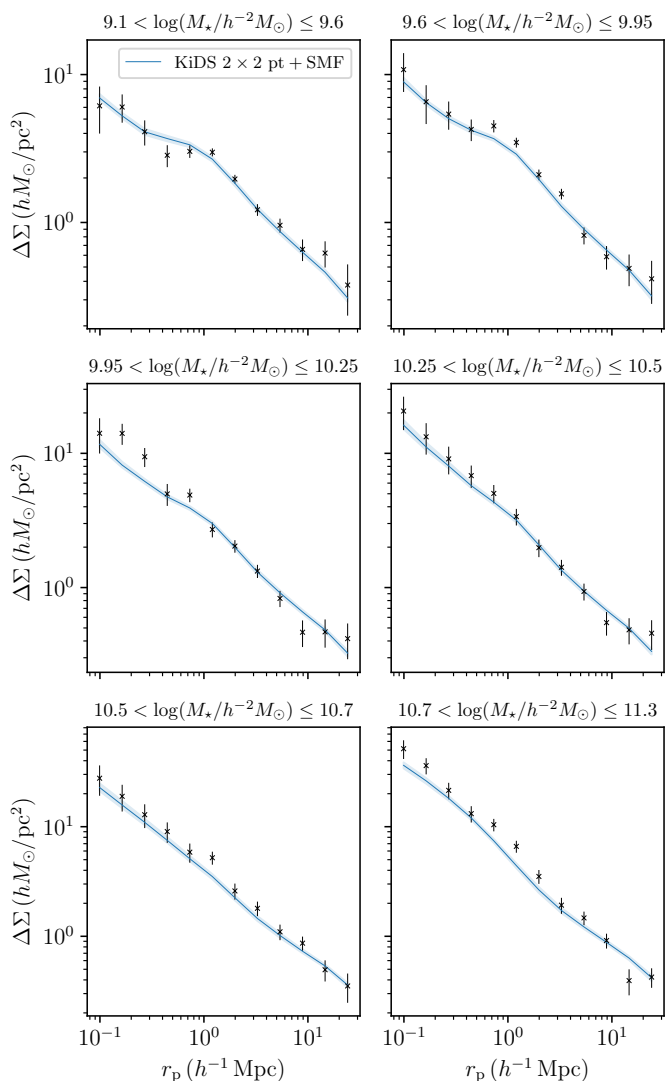
Fig. 3: Galaxy-galaxy lensing: The stacked ESD profiles of the six stellar mass bins in the KiDS-Bright galaxy sample defined in Table 1. The solid lines represent the best-fitting fiducial ESD halo model (Sec. 2.3, Eq. 34) as obtained using an MCMC fit, with the 68 percent confidence interval indicated with a shaded region. We caution that the quality of the fit cannot be judged by eye, because of the covariance in the data between the observed bins and also between the observables. The reduced χ^2 value for this observable is 1.28 (DoF = 73.18, p-value = 0.05), estimated using the method presented in Appendix C.

3.3. Projected galaxy clustering measurements: w_p

We measure the clustering of the KiDS-Bright galaxy sample using the Landy-Szalay (Landy & Szalay 1993) estimator for the galaxy correlation function

$$\hat{\xi}_{gg}(r_p, r_\pi) = \frac{DD - 2DR + RR}{RR}\bigg|_{r_p, r_\pi}. \tag{41}$$

Here we count the number of galaxy-galaxy (DD), random-random (RR) and galaxy-random (DR) pairs, as a function of the pair's transverse r_p and radial r_π comoving separation. The accuracy of galaxy clustering measurements with this estimator depends critically on the quality of the random, R, catalogues. We use the Johnston et al. (2021b) organised random

methodology that has been shown to recover unbiased clustering measurements in a series of mock galaxy catalogue analyses for the KiDS-Bright sample. Using machine learning, we infer the high-dimensional mapping between the observed on-sky galaxy number density and three systematic-tracer variables; atmospheric seeing, point spread function ellipticity and limiting magnitude. Systematically-induced density variations across the survey footprint can then be defined. We randomly distribute clones of the real galaxies across the survey footprint, preserving the on-sky systematic density patterns, and matching the on-sky systematic-tracer properties to that of the clone's parent galaxy. By retaining the photometric properties of the parent for each clone, selection effects are accurately mirrored in the organised randoms for any galaxy sub-sample, for example the 6 different stellar mass bins in our analysis.

The projected clustering correlation function is estimated through an integral over the line-of-sight separation, limited by a maximum defined distance $r_{\pi,\text{max}}$,

$$\hat{w}_p(r_p) = \int_{-r_{\pi,\text{max}}}^{r_{\pi,\text{max}}} \hat{\xi}_{gg}(r_p, r_\pi) \, dr_\pi. \tag{42}$$

When analysing spectroscopic data, this continuous integral is estimated using a discrete sum, typically adopting uniform bins in r_π ranging from $40h^{-1}$Mpc to $100h^{-1}$Mpc (as in for instance Mandelbaum et al. 2010; Farrow et al. 2015). Here the $r_{\pi,\text{max}}$ limits are chosen to maximise the number of correlated galaxy pairs along the line-of-sight in the presence of redshift space distortions, whilst minimising the noise arising from the inclusion of uncorrelated objects. With our KiDS-Bright photometric sample we have an additional uncertainty in the true redshift, $\sigma_z = 0.018(1 + z)$, which translates into an uncertainty on the radial distance of the order $\sim 100h^{-1}$Mpc, This renders the approach taken for spectroscopic samples sub-optimal in terms of signal-to-noise. We therefore choose to follow the approach of Johnston et al. (2021a) who optimised the projected galaxy clustering analysis of the photometric Physics of the Accelerating Universe Survey (PAUS), using dynamic binning in r_π out to a maximum $r_\pi = 233h^{-1}$Mpc. This is motivated by the fact that PAUS photometric redshifts show a similar uncertainty as the KiDS-Bright sample. Using a mock galaxy catalogue, Johnston et al. (2021a) demonstrated that by allowing for an increase in the bin size from small to large values of r_π, their approach maximises the count of physically associated objects, whilst minimising noise at large-r_π with the broader bin size. Given the similar photometric redshift properties of KiDS-Bright and PAUS, we adopt their 12-r_π-bin adapted Fibonacci sequence in our estimator.

Johnston et al. (2021a) analysed mock GAMA galaxy catalogues with PAU-like photometric redshifts to compare the projected clustering correlation function estimator $\hat{w}_p(r_p)$ with the measurements using spectroscopic redshifts. Adopting dynamic binning and random galaxy catalogues that mimic both the position and photometric redshift uncertainty of the real galaxy sample, they found a roughly scale-independent bias with $\hat{w}_p/w_{p^{\text{spec}}} \simeq 0.8$. As such the dynamic binning and organised randoms only partially correct the correlation functions for the dilution introduced by photometric redshift uncertainty. Future work will focus on accounting for this dilution effect accurately in the theoretical prediction. For the purposes of this analysis, however, we choose to include a free dilution parameter $\mathcal{D}$, which is used to correct the galaxy clustering measurements in the following way:

$$\hat{w}_{p,\text{corr}}(r_p) = [1 + \mathcal{D}] \, \hat{w}_p(r_p). \tag{43}$$

We adopt a uniform prior for $\mathcal{D}$ with the range between 0 and 0.3 and use a single parameter to scale all six stellar mass bins. This prior was motivated by a series of mock KiDS-Bright galaxy clustering analysis using MICE2 (Fosalba et al. 2015a,b; Crocce et al. 2015; Carretero et al. 2015; Hoffmann et al. 2015), where we confirmed the findings of Johnston et al. (2021a) and found no strong dependence of the dilution effect on stellar mass. We note that a similar correction was applied to the Dark Energy Survey (DES) photometric clustering measurements (Pandey et al. 2022; DES Collaboration et al. 2022, referred therein as X_{lens}). The prior and motivation behind the introduction of their systematic nuisance parameter differs, however. The resulting projected clustering measurements for the six stellar mass bins are shown in Fig. 4.

3.4. Accounting for the cosmology dependence of distance measures

To obtain estimates of the stellar mass function (SMF, Sect. 3.1, Fig. 2), the galaxy-galaxy lensing (ESD , Sect. 3.2, Fig. 3), and the projected galaxy clustering (w_{p}, Sect. 3.3, Fig. 4), we have adopted a fiducial flat ΛCDM cosmology with $\Omega_{\text{m}} = 0.3$ to compute distances. As such our 2×2pt+SMF data vector is cosmology-dependent, with changes in the fiducial cosmology changing the distance-redshift relation, which in turn shifts galaxies between the stellar mass bins and lens-source pairs between the radial separation bins.

At the mean redshift of the KiDS-Bright sample, the effect of changing Ω_{m} within our prior limits introduces changes in distance estimates at the level of few percent. The approximation that the measurements are effectively independent of cosmological parameters within their observational uncertainties (Mandelbaum et al. 2013; Cacciato et al. 2013) no longer holds for surveys with a statistical power that is similar or better than KiDS.

In this analysis we account for the cosmology dependence of our data vector following the correction procedure presented in More (2013) and More et al. (2015), which modifies the model prediction for each cosmology targeted by the likelihood sampler.

First we define a cosmology-dependent comoving separation $r_{\text{p}}^{\text{model}}$ for our target model, relative to the comoving separation r_{p} that has been used to calculate our data vectors at a fixed fiducial cosmological model,

$$r_{\text{p}}^{\text{model}} = r_{\text{p}}^{\text{fid}} \left[\frac{\chi(z_{\text{med}}, C^{\text{model}})}{\chi(z_{\text{med}}, C^{\text{fid}})} \right]. \qquad (44)$$

Here χ is the comoving distance to the median lens redshift z_{med} in our target cosmological model C^{model}, or in our fiducial cosmological model C^{fid}. The galaxy clustering prediction for our target model is then given by

$$\widetilde{w}_{\text{p}}(r_{\text{p}}) = w_{\text{p}}(r_{\text{p}}^{\text{model}}) \left[\frac{E^{\text{model}}(z_{\text{med}})}{E^{\text{fid}}(z_{\text{med}})} \right], \qquad (45)$$

where $E(z)$ is the Hubble parameter. The galaxy-galaxy lensing prediction for our target model is given by

$$\widetilde{\Delta\Sigma}(r_{\text{p}}) = \Delta\Sigma(r_{\text{p}}^{\text{model}}) \left[\frac{\Sigma_{\text{cr}}^{\text{model}}(z_{\text{med}}, z_{\text{s}})}{\Sigma_{\text{cr}}^{\text{fid}}(z_{\text{med}}, z_{\text{s}})} \right], \qquad (46)$$

where Σ_{cr} is the critical surface density calculated for the median redshift of the lenses z_{med} and a fixed source redshift $z_{\text{s}} = 0.6$. Note that calculating the more precise estimate for Σ_{cr} using Eq. 38 is not necessary in this instance, as Σ_{cr} only has a weak
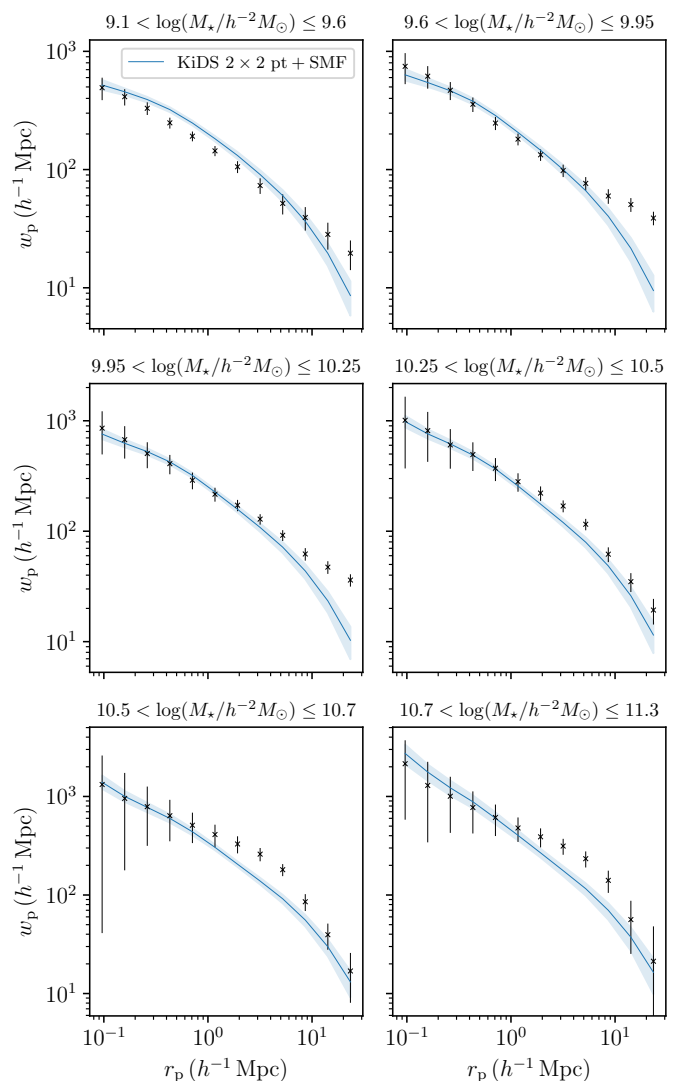
Fig. 4: Galaxy clustering: The projected galaxy clustering signal of the six stellar mass bins in the KiDS-Bright galaxy sample defined in Table 1. The solid lines represent the best-fitting fiducial halo model (Sec. 2.3, Eq. 27) as obtained using an MCMC fit, with the 68 percent confidence interval indicated with a shaded region. We caution that the quality of the fit cannot be judged by eye, because of the covariance in the data, between the observed bins and between the observables. The reduced χ^2 value for this observable is 1.42 (DoF = 71.62, p-value = 0.01), estimated using the method presented in Appendix C.

cosmology dependence. Finally the predictions of abundances of galaxies in the target cosmology is given by

$$\widetilde{n}_{\text{g}} = \overline{n}_{\text{g}}^{\text{model}} \left[\frac{\chi^3(z_{\text{u}}, C^{\text{model}}) - \chi^3(z_{\text{l}}, C^{\text{model}})}{\chi^3(z_{\text{u}}, C^{\text{fid}}) - \chi^3(z_{\text{l}}, C^{\text{fid}})} \right], \qquad (47)$$

which is implicitly correcting the surveyed volume in the stellar mass function calculation. Here the z_{l} and z_{u} are the lower and upper redshift limits in our samples.

3.5. Covariance matrix

The covariance matrix used in this analysis is based on the analytical approach detailed in Dvornik et al. (2018) and Joachimi

et al. (2021), with the addition of the analytical covariance matrix for the SMF and the cross terms between the SMF and 2-point correlation functions. The new terms for the SMF covariance and the cross covariance between the SMF and 2-point functions are presented in Appendix A. Our implementation of the analytical covariance derivation was validated against theory (Pielorz et al. 2010; Takada & Hu 2013; Li et al. 2014; Marian et al. 2015; Krause & Eifler 2017), independent software by Joachimi et al. (2021) and simulations (MICE2 Fosalba et al. 2015a,b; Carretero et al. 2015; Crocce et al. 2015; Hoffmann et al. 2015), following the validation approach of Blake et al. (2020) and Joachimi et al. (2021). Survey area effects on the variance were calculated using the accurate, survey dependent and data based HEALPIX method presented in Joachimi et al. (2021), equation E.10.

3.6. Likelihood and iterative updates

We use Bayesian inference to determine the posterior probability distribution $P(\theta \,|\, \mathbf{d})$ of the model parameters θ, given the data $\mathbf{d}$. According to Bayes' theorem, $P(\theta \,|\, \mathbf{d})$ is:

$$P(\theta \,|\, \mathbf{d}) = \frac{P(\mathbf{d}\,|\,\theta)P(\theta)}{P(\mathbf{d})}, \tag{48}$$

where $P(\mathbf{d}\,|\,\theta)$ is the likelihood of the data given the model parameters, $P(\theta)$ is the prior probability of these parameters, and

$$P(\mathbf{d}) = \int P(\mathbf{d}\,|\,\theta)\,P(\theta)\,\mathrm{d}\theta \tag{49}$$

is the evidence for the model. Since, we do not perform model selection in this analysis, the evidence just acts as a normalisation constant which we do not need to calculate. Given this, the likelihood distribution $P(\mathbf{d}\,|\,\theta)$ is assumed to be Gaussian:

$$P(\mathbf{d}\,|\,\theta) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{C}|}} \exp\left[-\frac{1}{2}\left[(\mathbf{m}(\theta) - \mathbf{d})^T \, \mathbf{C}^{-1} \, (\mathbf{m}(\theta) - \mathbf{d})\right]\right], \tag{50}$$

where $\mathbf{C}$ is the full covariance matrix for all the observables, containing their auto- and cross-correlations, $|\mathbf{C}|$ its determinant, $\mathbf{m}(\theta)$ the model given the parameters θ, and n the number of observable bins. Priors can be found in Table 2. For the Bayesian inference we use the MCMC sampler EMCEE (Foreman-Mackey et al. 2013).

The posterior distribution in such highly multi-dimensional parameter spaces has numerous degeneracies and can be very difficult to sample from. Thus the choice of proposal distributions is very important in order to achieve fast convergence and reasonable acceptance fractions for the proposed walker positions. To do so, we combine the default stretch move in the emcee with the proposal function based on the kernel density estimator of the complementary ensemble of walkers (Foreman-Mackey et al. 2013)[2] in such way that at every step of the sampler run, there is a 50% chance to use one of the proposal methods. This setup has one downside, and that is that it uses many walkers, and thus computing power. On the other hand, the convergence is faster and the resulting auto-correlation times are shorter, giving us shorter MCMC chains overall.

---

[2] The moves are further defined in the documentation of the emcee package at https://emcee.readthedocs.io.

During the MCMC runs we iteratively update the β^{NL} measurement (as it is cosmology dependent), as running the emulator at each step of the chain is computationally not feasible. Thus the β^{NL} measurement is evaluated using the median of the current position of the walkers in the parameter space. This returns an effective value for the non-linear halo bias correction that is, over the run of the MCMC, representative of the median of corrections that would be applied to every single model iteration in the chain. In our pipeline, the number of steps between iterations can be set by the user and we find that updating the β^{NL} values every 20 steps allows for a reasonable run time while providing enough updates to the β^{NL} correction. On the other hand, the covariance matrix is only re-evaluated with the new parameters at the end of the MCMC run and checked. We find that the updated covariance matrix and halo model parameters do not affect the results of our fit as our initial cosmological and HOD parameters were set to the one of Heymans et al. (2021) and van Uitert et al. (2016) which are close to our final results. The covariance matrix is dominated by the shot and/or shape noise on the majority of scales.

To report our result we use two methods to estimate our constraints and parameter values. One method uses the maximum statistics of the marginal posterior distributions for each parameter (MMAX). Here the asymmetric errors are estimated around the maximum point in iso-distribution levels to cover 68% of the marginal distribution. For the second method we use the full posterior distribution to find the best-fitting parameters (maximum a posteriori point, MAP) and use the methodology presented in Joachimi et al. (2021) to associate an error to this measurement with the projected highest posterior density (PJ-HPD) approach. While the former method produces more stable parameter errors, especially when the likelihood surface is sparsely sampled, its point estimates, in general, do not correspond to the best fitting parameter values. In contrast the latter method will in general produce noisier error estimates with unbiased parameter values.

4. Results

Now we turn the focus to our results, presenting cosmological parameter constraints in Sect. 4.1, large-scale analysis in Sect. 4.2, and constraints on the galaxy-halo connection in Sect. 4.3. Further details are presented in Appendix E. To recap, our theoretical 2×2pt+SMF model consists of 17 free parameters, two of which are our main cosmology parameters, with 3 more secondary cosmology parameters that are harder to constrain given the combination of observables, and 11 parameters describing the galaxy-halo connection in the form of the CSMF. With six stellar mass bins, and three observables, our combined data vector consists of 156 data points. In Appendix C we use mock data realisations to estimate the effective number of degrees of freedom for our analysis finding $\nu_{\mathrm{eff}} = 147.55$. Following the likelihood analysis described in Sect. 3.6 we are able to constrain 12 parameters, listed in Table 2 along with their prior ranges. We find the MAP to provide a good fit[3] to the data, with a reduced χ^2 value of 1.07 and $p(\chi^2|\nu_{\mathrm{eff}}) = 0.27$.

We compare the prediction from our fiducial model, and its 68% confidence regions, to the measured galaxy abundance

---

[3] We define an acceptable fit when $p(\chi^2|\nu_{\mathrm{eff}}) \geq 0.003$, corresponding to less than a 3σ event, (see the discussion in Heymans et al. 2021). We note that DES Collaboration et al. (2022) define a more stringent requirement where $p(\chi^2|\nu_{\mathrm{eff}}) \geq 0.01$. We find the goodness of fit for our 2×2pt+SMF analysis, and each individual component of the data vector, to be acceptable given both these definitions.

Table 2: Marginal constraints on all model parameters listed together with their priors.

Parameter	Prior	Fiducial	
		MMAX	MAP+PJ-HPD
$\Omega_{\rm m}$	$[0.1, 0.6]$	$0.290^{+0.021}_{-0.017}$	$0.307^{+0.002}_{-0.031}$
σ_8	$[0.4, 1.2]$	$0.781^{+0.033}_{-0.029}$	$0.801^{+0.013}_{-0.041}$
h	$[0.64, 0.82]$	< 0.726	< 0.711
$\Omega_{\rm b}$	$[0.01, 0.06]$	> 0.01	> 0.01
n_s	$[0.92, 1.1]$	< 1.004	< 0.978
S_8	–	$0.773^{+0.028}_{-0.030}$	$0.809^{+0.001}_{-0.055}$
$f_{\rm h}$	$[0.0, 1.0]$	> 0.645	> 0.939
M_0	$[7.0, 13.0]$	$10.519^{+0.039}_{-0.062}$	$10.521^{+0.062}_{-0.044}$
M_1	$[9.0, 14.0]$	$11.138^{+0.099}_{-0.132}$	$11.145^{+0.136}_{-0.098}$
γ_1	$[2.5, 15.0]$	$7.096^{+2.144}_{-1.406}$	$7.385^{+1.441}_{-1.824}$
γ_2	$[0.0, 10.0]$	0.201 ± 0.010	0.201 ± 0.009
$\sigma_{\rm c}$	$[0.0, 2.0]$	$0.108^{+0.067}_{-0.011}$	$0.159^{+0.007}_{-0.070}$
$f_{\rm s}$	$[0.0, 1.0]$	> 0.377	> 0.84
$\alpha_{\rm s}$	$[-5.0, 5.0]$	$-0.858^{+0.048}_{-0.052}$	$-0.847^{+0.013}_{-0.097}$
b_0	$[-5.0, 5.0]$	$-0.024^{+0.108}_{-0.117}$	$-0.120^{+0.199}_{-0.001}$
b_1	$[-5.0, 5.0]$	$1.149^{+0.091}_{-0.081}$	$1.177^{+0.058}_{-0.096}$
$\mathcal{P}$	$[0.0, 2.0]$	0.403 ± 0.029	$0.417^{+0.024}_{-0.013}$
$\mathcal{D}$	$[0.0, 0.3]$	$0.144^{+0.091}_{-0.085}$	$0.051^{+0.172}_{-0.006}$
$\chi^2_{\rm red}$	–	1.07	
p-value	–	0.27	

Notes: This table lists all the free parameters in our model: the energy density of cold matter $\Omega_{\rm m}$, the normalisation of power spectrum σ_8, the dimensionless Hubble parameter h, the spectral index n_s, the energy density of baryonic matter $\Omega_{\rm b}$, the derived parameter S_8, the normalisation of the concentration-mass relation for dark matter haloes $f_{\rm h}$, the normalisation of stellar-to-halo mass relation M_0, the characteristic scale of the stellar-to-halo mass relation M_1, the slope parameters of the stellar-to-halo mass relation γ_1 and γ_2, the scatter between stellar mass and halo mass $\sigma_{\rm c}$, the normalisation of the concentration-mass relation for distribution of satellite galaxies $f_{\rm s}$, the power law behaviour of satellites $\alpha_{\rm s}$, the normalisation constants of the Schechter function b_0 and b_1, and the Poisson parameter $\mathcal{P}$. Parameters are deemed unconstrained when the marginal probability at 2σ level exceeds 13% of the peak probability (see appendix A of Asgari et al. 2021b). In cases where one side is constrained we report the 1σ lower/upper limit. The MMAX estimate is the marginal maximum statistic, reporting the point of maximum marginal posterior distribution to the iso-posterior levels above and below the maximal point. The MAP+PJ-HPD (maximum posterior with projected joint highest posterior density) estimates are calculated following Joachimi et al. (2021).

SMF in Fig. 2, the galaxy-galaxy lensing ESD in Fig. 3, and the galaxy clustering $w_{\rm p}$ in Fig. 4, for all of the six stellar mass bins. We find that the model reproduces the overall trends in the data, such as the presence of the bump at $\sim 1h^{-1}$ Mpc in ESD due to satellite galaxies, and the fact that the stronger signal is present

where galaxies have higher stellar mass, showing that massive galaxies reside in more massive haloes. We note that some caution is needed when interpreting the results, as the quality of the fit cannot be judged by eye due to highly correlated data points. We find acceptable fits to each component of our 2×2pt+SMF data vector (for details see Appendix C). We note that the poorest fit is found for the $w_{\rm p}$ section of our joint data vector with $p[\chi^2(w_{\rm p})|\nu^{w_{\rm p}}_{\rm eff}] = 0.01$. Whilst a formally acceptable fit, this may indicate that our model is lacking the ability to correctly describe the photometric redshift dilution effect discussed in Sect. 3.3.

4.1. Cosmology constraints

We find the following cosmological parameter constraints from our simultaneous 2×2pt+SMF analysis of the ESD, $w_{\rm p}$ and SMF signals of galaxies in the KiDS-Bright sample,

$$\Omega_{\rm m} = 0.290^{+0.021}_{-0.017}$$
$$\sigma_8 = 0.781^{+0.033}_{-0.029}$$
$$S_8 = 0.773^{+0.028}_{-0.030},$$

where $S_8 = \sigma_8\sqrt{\Omega_{\rm m}/0.3}$, and we quote the maximum statistics of the marginal posterior distributions (MMAX). The remaining cosmological parameters are unconstrained by our analysis, and informed by our choice of prior (see Fig. E.2). In Fig. 5 we present the 68% and 95% confidence levels of the joint two-dimensional, marginalised posterior distribution in the $S_8 - \Omega_{\rm m}$ plane. The 2×2pt+SMF constraints are shown to be in good agreement with constraints from KiDS cosmic shear and KiDS with BOSS 3×2pt constraints (Asgari et al. 2021b; Heymans et al. 2021). They are formally consistent, but in some mild tension with the Planck Collaboration et al. (2020) TT,TE,EE+lowE CMB results. Using the Hellinger distance as a tension measure (see Heymans et al. 2021), the mild tension between our fiducial results and *Planck* is 1.9σ in S_8.

In Fig. 6 we compare our constraints to a broader range of joint-probe large-scale structure analyses, finding consistency with all. Cacciato et al. (2013) and More et al. (2015) adopt a similar methodology to this study, using a halo model formalism to jointly analyse galaxy-galaxy lensing, galaxy clustering and galaxy abundance observables. Cacciato et al. (2013) analysed the Sloan Digital Sky Survey (SDSS-DR7). More et al. (2015) combined data from the Baryon Oscillation Spectroscopic Survey (BOSS) and the Canada France Hawaii Telescope Lensing Survey (CFHTLenS). The improved constraining power in this KiDS analysis reflects the significant increase in depth for the lensing sample relative to SDSS-DR7, and the ten-fold increase in area relative to CFHTLenS. We can also compare to 2×2pt analyses that combine galaxy-galaxy lensing and galaxy clustering that introduce conservative scale cuts to reflect known limitations their adopted galaxy bias models. These include the fiducial analyses from DES, which adopt a linear galaxy bias model[4] (Porredon et al. 2021; Pandey et al. 2022), and from the Hyper Suprime Camera survey (HSC, Miyatake et al. 2021). The HSC analysis is highly complementary to our study as both analyses used a form of halo model with a halo occupation distribution. Miyatake et al. (2021) use the Zheng et al. (2005) HOD built into the dark-matter N-body DarkQuest emulator to predict the 2×2pt observables (Nishimichi et al. 2019; Miyatake et al. 2020). We use the same emulator to calibrate the non-linear

---

[4] Porredon et al. (2021); Pandey et al. (2022) also explore small-scale 2×2pt analyses using non-linear galaxy bias models, finding a 20-30% gain in cosmological constraining power.

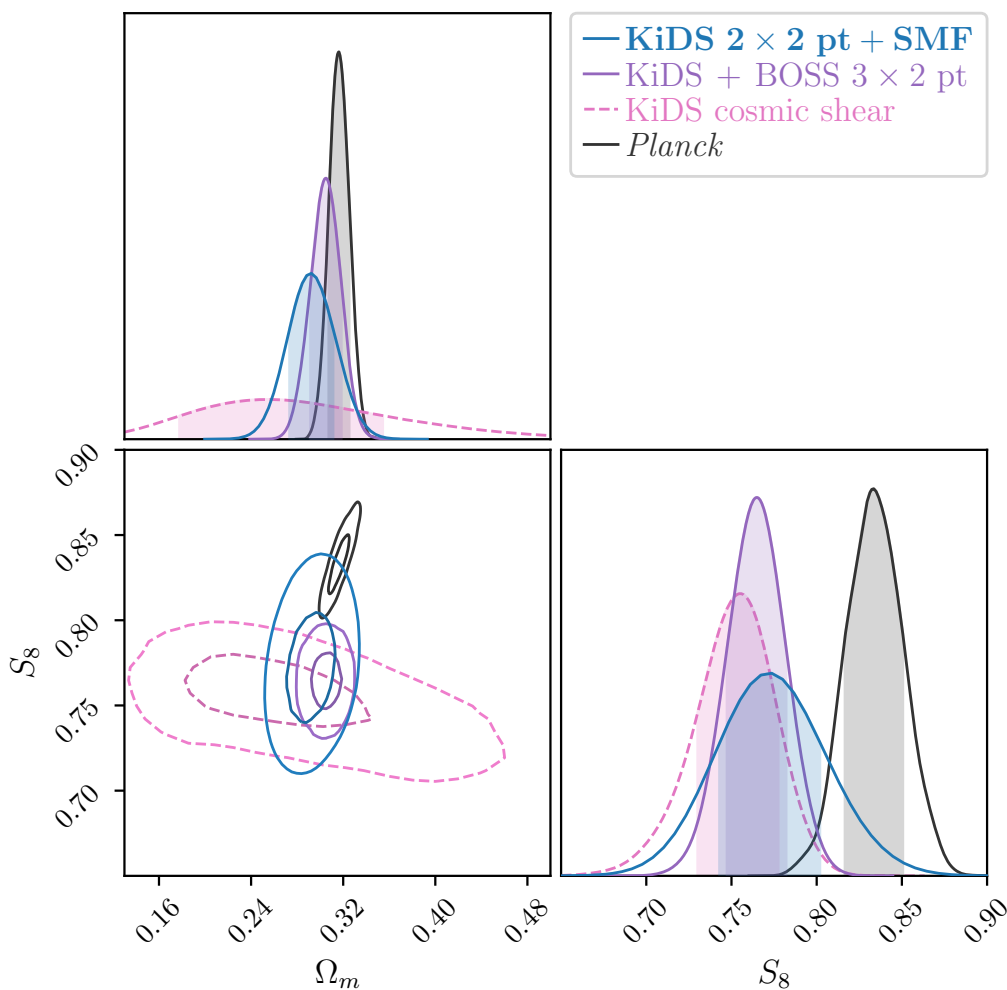Fig. 5: Marginalised constraints for the joint distributions of S_8 and Ω_{m}. The 68% and 95% credible regions for the 2×2pt+SMF fiducial analysis (blue) can be compared with constraints from KiDS cosmic shear (Asgari et al. 2021b, pink), KiDS with BOSS 3×2pt (Heymans et al. 2021, purple), and the CMB Planck Collaboration et al. (2020, black).

halo bias in our model (see Sect. 2.1). Compared to this analysis, HSC adopts more conservative scale cuts arising from concern over unmodeled baryon feedback on small-scales, which our methodology can account for through the free normalisation of the mass-concentration relation(see Sect. 4.3). Miyatake et al. (2021) also choose scale cuts to mitigate small-scale assembly bias for the relatively rare luminous red galaxies in their sample, which cannot be modelled using an HOD. Owing to the volume-limited mix of all galaxies used in our KiDS-Bright analysis, we consider any assembly bias to be a subdominant effect in our theoretical model.

The constraining power of the KiDS 2×2pt+SMF analysis in the $S_8 - \Omega_{\mathrm{m}}$ plane is the same as that of the 3×2pt studies from DES Collaboration et al. (2022), with $\sigma^{S_8}\sigma^{\Omega_{\mathrm{m}}}_{[\mathrm{DES\ 3\times2pt}]}/\sigma^{S_8}\sigma^{\Omega_{\mathrm{m}}}_{[\mathrm{KiDS\ 2\times2pt+SMF}]} = 0.97$. This may be surprising given the five-fold increase in area for DES relative to KiDS, and the addition of the cosmic shear probe in the 3×2pt analysis. This comparison therefore highlights the significant constraining power from the inclusion of non-linear scales in the 2×2pt+SMF analysis that are excluded from the DES 3×2pt analysis. Comparing KiDS 2×2pt+SMF constraints to the KiDS with BOSS 3×2pt analysis (Heymans et al. 2021), we first review the BOSS spectroscopic clustering

constraints where $\sigma^{S_8}\sigma^{\Omega_{\mathrm{m}}}_{[\mathrm{BOSS\ 1\times2pt}]}/\sigma^{S_8}\sigma^{\Omega_{\mathrm{m}}}_{[\mathrm{KiDS\ 2\times2pt+SMF}]} = 1.03$. Finding the same constraining power between these analyses may again be surprising, given the nine-fold increase in area for BOSS relative to KiDS. As such it demonstrates the significant constraining power from non-linear scales when the galaxy bias can be constrained using galaxy-galaxy lensing and galaxy abundance. Comparing to the full 3×2pt analysis we find $\sigma^{S_8}\sigma^{\Omega_{\mathrm{m}}}_{[\mathrm{KiDS+BOSS\ 3\times2pt}]}/\sigma^{S_8}\sigma^{\Omega_{\mathrm{m}}}_{[\mathrm{KiDS\ 2\times2pt+SMF}]} = 0.39$, where the extra constraining power in the 3×2pt analysis is driven by the cosmic shear. Future studies with KiDS will combine 2×2pt+SMF with cosmic shear data, including further development and validation of our adopted halo model methodology.

In Appendix E we explore a number of extensions to our fiducial analysis. In Appendix E.1 we quantify the expected contamination to our observables from intrinsic galaxy alignments and magnification. The contamination levels are found to be negligible relative to our statistical errors, justifying our choice to not account for these astrophysical effects in our model. In Appendix E.2 we demonstrate that our S_8 and Ω_{m} constraints are insensitive to our choice of prior on n_s. In Appendix E.3 we quantify the bias in Ω_{m} without the inclusion of our nuisance parameter $\mathcal{D}$ to model photometric redshift dilution in our galaxy clus-
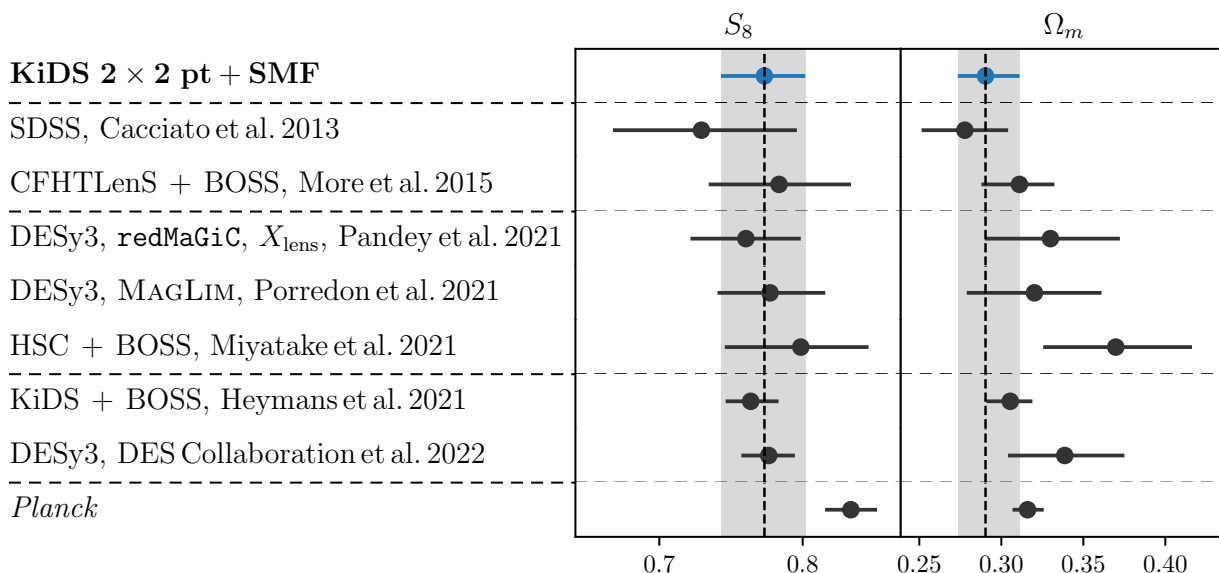
Fig. 6: A joint-probe comparison of S_8 and Ω_m constraints. Our fiducial results (KiDS 2×2pt +SMF) can be compared to a similar 2×2pt +SMF analysis from SDSS (Cacciato et al. 2013); a series of 2×2pt studies including the latest results from DES (Porredon et al. 2021; Pandey et al. 2022) and HSC (Miyatake et al. 2021); 3×2pt analyses from KiDS with BOSS (Heymans et al. 2021) and DES (DES Collaboration et al. 2022). The last entry shows the Planck Collaboration et al. (2020, TT,TE,EE+lowE) constraints. Our results are consistent with all studies, including Planck Collaboration et al. (2020), although we find a mild tension between our S_8 constraints and those from *Planck*, at the level of 1.9σ.

tering measurement. In Appendix E.4 we quantify the impact of assumptions governing the behaviour of satellite galaxies.

4.2. Large-Scale Analysis

Amon et al. (2022b) present a detailed analysis of the uncertainty on the amplitude of the small-scale galaxy-galaxy lensing and galaxy clustering signal for BOSS galaxies that arises from our imperfect knowledge of baryon feedback and assembly bias. They conclude that the introduction of scale cuts with $r_p > 5h^{-1}$ Mpc fully isolates these effects. Following the methodology in Appendix C we determine $\nu_{\text{eff}}^{r_p>5} = 60.31$ for a large-scale only 2×2pt+SMF data vector analysis, calculating the large-scale goodness of fit of our fiducial best-fit model (see Table 2) to be $p[\chi^2(r_p > 5)|\nu_{\text{eff}}^{r_p>5}] = 0.19$. As such we find no sign of tension between our fiducial all-scale analysis and a restricted large-scale analysis.

The majority of the information in our ($r_p > 5$) data vector comes from the SMF as there are only 4 highly correlated data points remaining in the ESD and w_p measurements, per stellar mass bin. We found that a full likelihood analysis of the ($r_p > 5$) data vector was unable to converge in our highly-flexible 17-parameter model space. Where larger scales are used exclusively, either more precise data or the use of a less flexible model is necessary (such as in More 2013; Miyatake et al. 2021; Amon et al. 2022b, amongst others). The extra flexibility afforded in our model is, however, essential when analysing small scales in order to capture baryonic effects (Debackere et al. 2020, 2021)

4.3. Galaxy-halo connection

The powerful aspect of the method used in this work is that it is able to simultaneously constrain both cosmological parameters as well as the halo occupation statistics. The full results are

listed in Table 2 and the marginalised posterior distributions of all the parameters are shown in Fig. E.2. The HOD parameters are tightly constrained, with some strong degeneracies between the parameters M_0, M_1 and γ_1 that govern the characteristic mass of the stellar-to-halo mass relation knee and the high mass slope of centrals, and between the b_0 and b_1 parameters, which govern the normalisation of the satellite CSMF. The first degeneracy is somewhat expected given the data, as the stellar mass function at the high mass end is highly uncertain and dominated by the Eddington bias. The second degeneracy also arises from the fact that both parameters compete for the overall normalisation of the satellite CSMF.

We find the parameters of our HOD model to be in good agreement with the previous studies using GAMA-like galaxies (van Uitert et al. 2016; Dvornik et al. 2018; Bilicki et al. 2021). In order to show the agreement in a more intuitive way, we take the parameters of the stellar-to-halo mass relation and combine them using the same functional form (Eq. 19), which results in the relation shown in Fig. 7. In the same figure we show good agreement with the results from van Uitert et al. (2016), where the HOD parameters were constrained using galaxy-galaxy lensing combined with a SMF for KiDS and GAMA data, adopting a fixed *Planck* cosmology. We also find qualitative agreement with constraints from abundance matching to numerical simulations (Moster et al. 2013) and constraints from a 2×2pt+SMF analysis of COSMOS with a fixed *WMAP5* cosmology (Leauthaud et al. 2012, note here we compare to constraints from the most similar $0.22 < z < 0.48$ COSMOS sample) as well as the CFHTLenS/VIPERS analysis by Coupon et al. (2015). We find our constraints on the scatter of the central CSMF, σ_c, and the low mass end slope of the satellites, α_s, to be in agreement with Yang et al. (2009); Cacciato et al. (2013); van Uitert et al. (2016); Dvornik et al. (2018); Bilicki et al. (2021). The Eddington bias at the high mass end is captured by the σ_c parameter, leaving the other parameters mostly unaffected.
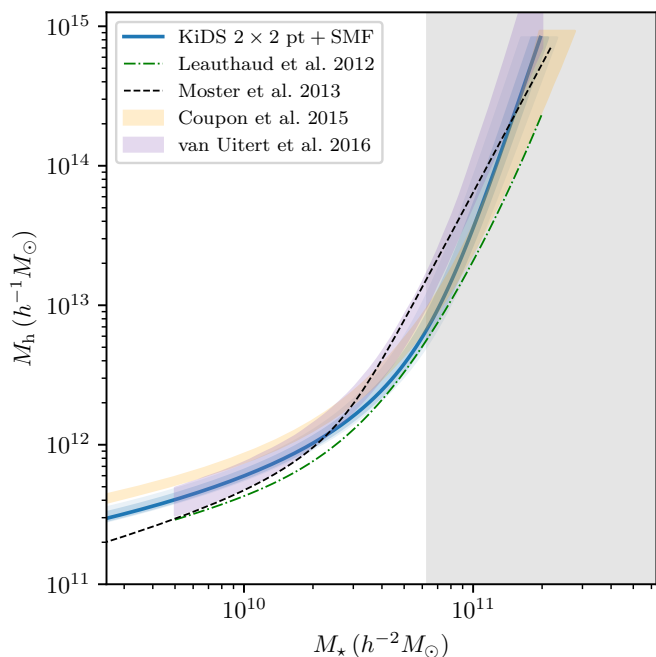
Fig. 7: The stellar-to-halo mass relation, as defined by Eq. 19 using the best-fit HOD parameters from our 2×2pt+SMF analysis (blue). The result can be compared to results from Leauthaud et al. (2012, green), Moster et al. (2013, black), Coupon et al. (2015, orange), and van Uitert et al. (2016, purple). The grey area shows the range in stellar masses where the obtained stellar-to-halo mass relation is extrapolated beyond the range of median stellar masses used in this analysis.

We account for the impact of baryon feedback in our model by allowing for freedom in the normalisation of the mass-concentration relation for both the haloes, f_h, and the satellite galaxy distribution, f_s (Eq. 15). With these independent free parameters we can capture the expected small-scale baryon feedback power suppression shown in hydrodynamical simulations (Debackere et al. 2021; Amon et al. 2022b). We find f_h and f_s to be consistent with 1, with a preference for lower values, with 1σ lower limits $f_h > 0.65$ and $f_s > 0.38$. Our results are consistent with Viola et al. (2015), indicating that the concentrations of real haloes and satellite distributions are smaller than the haloes in dark matter only simulations (see also Debackere et al. 2020, 2021). Future work will compare direct measurements from hydrodynamical simulations (e.g. McCarthy et al. 2017) with our halo model approach to account for the mass dependence of baryonic effects on the radial profiles of dark matter haloes.

5. Discussion and conclusions

In this paper we combined measurements of galaxy clustering, galaxy-galaxy lensing and galaxy abundances in the form of the stellar mass function, in order to simultaneously set constraints on cosmological parameters and galaxy bias. Using a flexible halo model, we analysed the fourth data release of the Kilo-Degree Survey (KiDS) (KiDS-1000, Kuijken et al. 2019), where the source sample, used to measure the galaxy-galaxy lensing signal, has undergone a rigorous study to assess robustness and accuracy (Asgari et al. 2021b; Giblin et al. 2021; Hildebrandt et al. 2021). For our lens sample we used the KiDS-Bright sam-

ple (Bilicki et al. 2021) whose selection was calibrated against a complete and representative spectroscopic sample from the GAMA survey (Driver et al. 2011), with the photometric redshifts calibrated using a neural network (ANNz2, Sadeh et al. 2016). The resulting accuracy of the estimated redshifts for the KiDS-bright sample is sufficient for galaxy-galaxy lensing and galaxy clustering studies.

We used the halo model to analyse our data, building upon the cosmological analyses presented in Cacciato et al. (2013) and More et al. (2015). We used a single halo occupation model to compute the clustering of galaxies and the galaxy-galaxy lensing signal (Guzik & Seljak 2002; Yoo et al. 2006; Cacciato et al. 2009), and the galaxy abundances (van den Bosch et al. 2013; Cacciato et al. 2013). This model was shown to be able to simultaneously constrain cosmology and halo occupation statistics, as well as constrain the extensions to the standard ΛCDM cosmologies, such as the equation of state of dark energy and neutrino mass (More et al. 2013; Krause & Eifler 2017).

We improved upon previous studies by (i) using a more accurate N-body simulation calibrated analytical model, taking into account halo exclusion, scale dependence and the non-linear nature of halo bias (Mead & Verde 2021; Mahony et al. 2022), (ii) using the latest lensing and clustering data from a single survey (KiDS-1000), and (iii) using the full analytical covariance matrix that accounts for cross-covariance between all observables and in particular the cross-covariance between the stellar mass function and two point statistics.

We have adopted a Bayesian approach to constrain our model parameters, using the MCMC to probe the posterior distributions. For a flat ΛCDM cosmology we find $\Omega_m = 0.290^{+0.021}_{-0.017}$ and $S_8 = 0.773^{+0.028}_{-0.030}$, which is consistent and comparable to constraints from 3×2pt studies that also include a cosmic shear observable (Heymans et al. 2021; DES Collaboration et al. 2022). Our results follow the trend seen in other lensing studies for a tension in S_8 when compared to Planck Collaboration et al. (2020). Using the Hellinger distance as a tension metric, this difference is at the 1.9σ level for our 2×2pt+SMF analysis.

Combining galaxy clustering and galaxy-galaxy lensing with cosmic shear measurements has been a standard approach for large-scale structure analyses in recent years (van Uitert et al. 2018; Joudaki et al. 2017; Abbott et al. 2018; DES Collaboration et al. 2022; Heymans et al. 2021). We anticipate that combining our halo model approach with cosmic shear data will allow for additional constraints on astrophysical systematics arising from the intrinsic alignment of galaxies and baryon feedback. So far, intrinsic alignments in cosmic shear analysis have either been included using a non-linear modification of the linear alignment model (Bridle & King 2007, NLA) or a perturbation theory approach (TATT Blazek et al. 2019). For a consistent halo model approach this effect could also be modelled within the framework adopted in this analysis (e.g. Fortuna et al. 2021). In this analysis we have varied the halo concentration parameter to account for baryon feedback. With additional data, a more complex halo model could be adopted allowing for the inclusion of gas observables to constrain baryon feedback through the Sunyaev-Zeldovich effect (Mead et al. 2020; Tröster et al. 2022). The flexibility of the halo model also allows for extensions to the underlying cosmological model without having to employ a myriad of costly simulations to cover a large range of parameters (Cataneo et al. 2019). We therefore see a significant role for our adopted methodology in future cosmological analyses of upcoming large-scale structure surveys.

Author contributions: All authors contributed to writing and development of this paper. The authorship list reflects the lead authors (AD,CH,MA,CM,BJ) followed by an alphabetical group that includes those who are key contributors to the scientific analysis.

References

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, Phys. Rev. D, 98, 043526
Alam, S., Aubert, M., Avila, S., et al. 2021, Phys. Rev. D, 103, 083533
Amon, A., Gruen, D., Troxel, M. A., et al. 2022a, Phys. Rev. D, 105, 023514
Amon, A., Robertson, N. C., Miyatake, H., et al. 2022b, ArXiv e-prints, 1
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, Mon. Not. R. Astron. Soc., 310, 540
Asgari, M., Friswell, I., Yoon, M., et al. 2021a, Mon. Not. R. Astron. Soc., 501, 3003
Asgari, M., Lin, C.-a., Joachimi, B., et al. 2021b, Astron. Astrophys., 645, A104
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, Astron. J., 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, Astron. Astrophys., 558, A33
Baldry, I. K., Driver, S. P., Loveday, J., et al. 2012, Mon. Not. R. Astron. Soc., 421, no
Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291
Benitez, N. 2000, Astrophys. J., 536, 571
Bilicki, M., Dvornik, A., Hoekstra, H., et al. 2021, Astron. Astrophys., 653, A82
Bilicki, M., Hoekstra, H., Brown, M. J. I., et al. 2018, Astron. Astrophys., 616, A69
Blake, C., Amon, A., Asgari, M., et al. 2020, Astron. Astrophys., 642, A158
Blazek, J. A., MacCrann, N., Troxel, M. A., & Fang, X. 2019, Phys. Rev. D, 100, 103506

Bridle, S. & King, L. 2007, New J. Phys., 9, 444
Brouwer, M. M., Oman, K. A., Valentijn, E. A., et al. 2021, Astron. Astrophys., 650, A113
Brown, M. L., Taylor, A. N., Hambly, N. C., & Dye, S. 2002, Mon. Not. R. Astron. Soc., 333, 501
Bruzual, G. & Charlot, S. 2003, Mon. Not. R. Astron. Soc., 344, 1000
Cacciato, M., Lahav, O., van den Bosch, F. C., Hoekstra, H., & Dekel, A. 2012, Mon. Not. R. Astron. Soc., 426, 566
Cacciato, M., van den Bosch, F. C., More, S., et al. 2009, Mon. Not. R. Astron. Soc., 394, 929
Cacciato, M., van den Bosch, F. C., More, S., Mo, H., & Yang, X. 2013, Mon. Not. R. Astron. Soc., 430, 767
Cacciato, M., van Uitert, E., & Hoekstra, H. 2014, Mon. Not. R. Astron. Soc., 437, 377
Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, Astrophys. J., 429, 582
Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, Mon. Not. R. Astron. Soc., 447, 646
Cataneo, M., Lombriser, L., Heymans, C., et al. 2019, MNRAS, 488, 2121
Chabrier, G. 2003, Astrophys. J., 586, L133
Chisari, N. E., Mandelbaum, R., Strauss, M. A., Huff, E. M., & Bahcall, N. A. 2014, Mon. Not. R. Astron. Soc., 445, 726
Cole, S. 2011, Mon. Not. R. Astron. Soc., 416, no
Cooray, A. & Sheth, R. 2002, Phys. Rep., 372, 1
Coupon, J., Arnouts, S., van Waerbeke, L., et al. 2015, Mon. Not. R. Astron. Soc., 449, 1352
Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, Mon. Not. R. Astron. Soc., 453, 1513
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, Astrophys. J., 292, 371
de Jong, J., Kuijken, K., Applegate, D., et al. 2013, The Messenger, 154, 44
Debackere, S. N. B., Schaye, J., & Hoekstra, H. 2020, Mon. Not. R. Astron. Soc., 492, 2285
Debackere, S. N. B., Schaye, J., & Hoekstra, H. 2021, Mon. Not. R. Astron. Soc., 505, 593
Dekel, A. & Lahav, O. 1999, Astrophys. J., 520, 24
Dekel, A. & Rees, M. J. 1987, Nature, 326, 455
DES Collaboration, Abbott, T. M. C., Aguena, M., et al. 2022, Phys. Rev. D, 105, 023520
Desjacques, V., Jeong, D., & Schmidt, F. 2018, Phys. Rep., 733, 1
Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, Mon. Not. R. Astron. Soc., 413, 971
Driver, S. P., Robotham, A. S. G., Obreschkow, D., et al. 2022, Mon. Not. R. Astron. Soc., 515, 2138
Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, Mon. Not. R. Astron. Soc. Lett., 390, L64
Dvornik, A., Cacciato, M., Kuijken, K., et al. 2017, Mon. Not. R. Astron. Soc., 468, 3251
Dvornik, A., Hoekstra, H., Kuijken, K., et al. 2018, Mon. Not. R. Astron. Soc., 479, 1240
Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
Eisenstein, D. J. & Hu, W. 1998, Astrophys. J., 496, 605
Farrow, D. J., Cole, S., Norberg, P., et al. 2015, Mon. Not. R. Astron. Soc., 454, 2120
Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, Mon. Not. R. Astron. Soc., 467, 1627
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publ. Astron. Soc. Pacific, 125, 306
Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2021, Mon. Not. R. Astron. Soc., 501, 2983
Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015a, Mon. Not. R. Astron. Soc., 448, 2987
Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2015b, Mon. Not. R. Astron. Soc., 447, 1319
Georgiou, C., Chisari, N. E., Fortuna, M. C., et al. 2019, Astron. Astrophys., 628, A31
Gerbino, M. & Lattanzi, M. 2018, Front. Phys., 5 [arXiv:1712.07109]
Giblin, B., Heymans, C., Asgari, M., et al. 2021, Astron. Astrophys., 645, A105
Giocoli, C., Bartelmann, M., Sheth, R. K., & Cacciato, M. 2010, Mon. Not. R. Astron. Soc., 408, 300
Guzik, J. & Seljak, U. 2002, Mon. Not. R. Astron. Soc., 335, 311
Hamana, T., Shirasaki, M., Miyazaki, S., et al. 2019, J. Biochem., 166, 1
Heymans, C., Grocutt, E., Heavens, A., et al. 2013, Mon. Not. R. Astron. Soc., 432, 2433
Heymans, C., Tröster, T., Asgari, M., et al. 2021, Astron. Astrophys., 646, A140
Hikage, C., Oguri, M., Hamana, T., et al. 2019, Publ. Astron. Soc. Japan, 71, 1
Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, Astron. Astrophys., 647, A124
Hinton, S. 2016, J. Open Source Softw., 1, 45

Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, Mon. Not. R. Astron. Soc., 447, 1724

Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90

Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, Astron. Astrophys., 457, 841

Joachimi, B. & Bridle, S. L. 2010, Astron. Astrophys., 523, A1

Joachimi, B., Lin, C.-A., Asgari, M., et al. 2021, Astron. Astrophys., 646, A129

Joachimi, B., Mandelbaum, R., Abdalla, F. B., & Bridle, S. L. 2011, Astron. Astrophys., 527, A26

Johnston, H., Joachimi, B., Norberg, P., et al. 2021a, Astron. Astrophys., 646, A147

Johnston, H., Wright, A. H., Joachimi, B., et al. 2021b, Astron. Astrophys., 648, A98

Joudaki, S., Mead, A., Blake, C., et al. 2017, Mon. Not. R. Astron. Soc., 471, 1259

Kaiser, N. 1987, Mon. Not. R. Astron. Soc., 227, 1

Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, Astron. Astrophys., 624, A92

Krause, E. & Eifler, T. 2017, Mon. Not. R. Astron. Soc., 470, 2100

Kuijken, K. 2011, The Messenger, 146, 8

Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, Astron. Astrophys., 625, A2

Landy, S. D. & Szalay, A. S. 1993, Astrophys. J., 412, 64

Lange, J. U., Leauthaud, A., Singh, S., et al. 2021, Mon. Not. R. Astron. Soc., 502, 2074

Leauthaud, A., Saito, S., Hilbert, S., et al. 2017, Mon. Not. R. Astron. Soc., 467, 3024

Leauthaud, A., Tinker, J., Behroozi, P. S., Busha, M. T., & Wechsler, R. H. 2011, Astrophys. J., 738, 45

Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, Astrophys. J., 744, 159

Li, R., Mo, H. J., Fan, Z., et al. 2009, Mon. Not. R. Astron. Soc., 394, 1016

Li, S.-S., Kuijken, K., Hoekstra, H., et al. 2021, Astron. Astrophys., 646, A175

Li, Y., Hu, W., & Takada, M. 2014, Phys. Rev. D - Part. Fields, Gravit. Cosmol., 89, 1

Mahony, C., Dvornik, A., Mead, A., et al. 2022, Mon. Not. R. Astron. Soc., 515, 2612

Mandelbaum, R., Seljak, U., Baldauf, T., & Smith, R. E. 2010, Mon. Not. R. Astron. Soc., 405, 2078

Mandelbaum, R., Slosar, A., Baldauf, T., et al. 2013, Mon. Not. R. Astron. Soc., 432, 1544

Marian, L., Smith, R. E., & Angulo, R. E. 2015, Mon. Not. R. Astron. Soc., 451, 1418

McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, Mon. Not. R. Astron. Soc., 465, 2936

Mead, A. J., Peacock, J. A., Heymans, C., Joudaki, S., & Heavens, A. F. 2015, Mon. Not. R. Astron. Soc., 454, 1958

Mead, A. J., Tröster, T., Heymans, C., Van Waerbeke, L., & McCarthy, I. G. 2020, Astron. Astrophys., 641, A130

Mead, A. J., Tröster, T., Heymans, C., Van Waerbeke, L., & McCarthy, I. G. 2020, A&A, 641, A130

Mead, A. J. & Verde, L. 2021, Mon. Not. R. Astron. Soc., 503, 3095

Miller, L., Heymans, C., Kitching, T. D., et al. 2013, Mon. Not. R. Astron. Soc., 429, 2858

Miyatake, H., Kobayashi, Y., Takada, M., et al. 2020, ArXiv e-prints [arXiv:2101.00113]

Miyatake, H., Sugiyama, S., Takada, M., et al. 2021, ArXiv e-prints [arXiv:2111.02419]

Mo, H., van den Bosch, F., & White, S. 2010, Galaxy Formation and Evolution (Cambridge: Cambridge University Press), 810

More, S. 2013, Astrophys. J., 777, L26

More, S., Miyatake, H., Mandelbaum, R., et al. 2015, Astrophys. J., 806, 2

More, S., van den Bosch, F. C., Cacciato, M., et al. 2013, Mon. Not. R. Astron. Soc., 430, 747

Moster, B. P., Naab, T., & White, S. D. M. 2013, Mon. Not. R. Astron. Soc., 428, 3121

Murray, S., Power, C., & Robotham, A. 2013, Astron. Comput., 3-4, 23

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, Astrophys. J., 490, 493

Nishimichi, T., Takada, M., Takahashi, R., et al. 2019, Astrophys. J., 884, 29

Pandey, S., Krause, E., DeRose, J., et al. 2022, Phys. Rev. D, 106, 043520

Peacock, J. A. & Smith, R. E. 2000, Mon. Not. R. Astron. Soc., 318, 1144

Pielorz, J., Rödiger, J., Tereno, I., & Schneider, P. 2010, Astron. Astrophys., 514, A79

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, Astron. Astrophys., 641, A6

Porredon, A., Crocce, M., Elvin-Poole, J., et al. 2021, ArXiv e-prints [arXiv:2105.13546]

Raveri, M. & Hu, W. 2019, Phys. Rev. D, 99, 043506

Robotham, A. S. G., Norberg, P., Driver, S. P., et al. 2011, Mon. Not. R. Astron. Soc., 416, 2640

Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, Publ. Astron. Soc. Pacific, 128, 104502

Sánchez, A. G., Scoccimarro, R., Crocce, M., et al. 2017, Mon. Not. R. Astron. Soc., 464, 1640

Saunders, W., Rowan-Robinson, M., Lawrence, A., et al. 1990, Mon. Not. R. Astron. Soc., 242, 318

Schaan, E., Takada, M., & Spergel, D. N. 2014, Phys. Rev. D, 90, 123523

Schaller, M., Frenk, C. S., Bower, R. G., et al. 2015, Mon. Not. R. Astron. Soc., 451, 1247

Schlafly, E. F. & Finkbeiner, D. P. 2011, Astrophys. J., 737, 103

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, Astrophys. J., 500, 525

Schmidt, M. 1968, Astrophys. J., 151, 393

Schneider, A. & Teyssier, R. 2015, J. Cosmol. Astropart. Phys., 2015, 049

Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, Astrophys. J., 546, 20

Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101

Secco, L. F., Samuroff, S., Krause, E., et al. 2022, Phys. Rev. D, 105, 023515

Seljak, U. 2000, Mon. Not. R. Astron. Soc., 318, 203

Seljak, U., Makarov, A., Mandelbaum, R., et al. 2005, Phys. Rev. D, 71, 043511

Simon, P. & Hilbert, S. 2018, Astron. Astrophys., 613, A15

Singh, S., Mandelbaum, R., Seljak, U., Slosar, A., & Vazquez Gonzalez, J. 2017, Mon. Not. R. Astron. Soc., 471, 3827

Smith, R. E. 2012, Mon. Not. R. Astron. Soc., 426, 531

Takada, M. & Bridle, S. 2007, New J. Phys., 9, 446

Takada, M. & Hu, W. 2013, Phys. Rev. D - Part. Fields, Gravit. Cosmol., 87, 1

Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, Mon. Not. R. Astron. Soc., 418, 1587

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, Astrophys. J., 724, 878

Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, Astrophys. J., 631, 41

Tröster, T., Asgari, M., Blake, C., et al. 2021, A&A, 649, A88

Tröster, T., Mead, A. J., Heymans, C., et al. 2022, A&A, 660, A27

Unruh, S., Schneider, P., Hilbert, S., et al. 2020, Astron. Astrophys., 638 [arXiv:1910.06400]

van den Bosch, F. C., More, S., Cacciato, M., Mo, H., & Yang, X. 2013, Mon. Not. R. Astron. Soc., 430, 725

Van Den Busch, J. L., Wright, A. H., Hildebrandt, H., et al. 2022, ArXiv e-prints [arXiv:2204.02396]

van Uitert, E., Cacciato, M., Hoekstra, H., et al. 2016, Mon. Not. R. Astron. Soc., 459, 3251

van Uitert, E., Joachimi, B., Joudaki, S., et al. 2018, Mon. Not. R. Astron. Soc., 476, 4662

Viola, M., Cacciato, M., Brouwer, M., et al. 2015, Mon. Not. R. Astron. Soc., 452, 3529

Wang, L., Farrah, D., Oliver, S. J., et al. 2013, Mon. Not. R. Astron. Soc., 431, 648

Wang, Y., Vogelsberger, M., Xu, D., et al. 2020, MNRAS, 491, 5188

Wechsler, R. H. & Tinker, J. L. 2018, Annu. Rev. Astron. Astrophys., 56, 435

Weigel, A. K., Schawinski, K., & Bruderer, C. 2016, Mon. Not. R. Astron. Soc., 459, 2150

Wibking, B. D., Salcedo, A. N., Weinberg, D. H., et al. 2019, MNRAS, 484, 989

Wright, A. H., Driver, S. P., & Robotham, A. S. G. 2018, Mon. Not. R. Astron. Soc., 480, 3491

Wright, A. H., Hildebrandt, H., Kuijken, K., et al. 2019, Astron. Astrophys., 632, A34

Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020, Astron. Astrophys., 640, L14

Wright, A. H., Robotham, A. S. G., Bourne, N., et al. 2016, MNRAS, 460, 765

Wright, A. H., Robotham, A. S. G., Driver, S. P., et al. 2017, Mon. Not. R. Astron. Soc., 470, 283

Yang, X., Mo, H. J., & van den Bosch, F. C. 2008, Astrophys. J., 676, 248

Yang, X., Mo, H. J., & van den Bosch, F. C. 2009, Astrophys. J., 695, 900

Yoo, J., Tinker, J. L., Weinberg, D. H., et al. 2006, Astrophys. J., 652, 26

Yoon, M., James Jee, M., Anthony Tyson, J., et al. 2019, Astrophys. J., 870, 111

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, Astrophys. J., 736, 59

Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, Astrophys. J., 633, 791

Appendix A: Stellar mass function covariance matrix

We derive the covariance of the SMF in direct analogy to the flux-limited case considered in Smith (2012), but neglect the halo occupation variance contribution because it was demonstrated to be always subdominant (Smith 2012). To simplify the expression we neglect the mapping from true stellar mass to observed stellar mass, where the corresponding integrations would appear explicitly as this relation is expected to be fairly tight. The SMF covariance is composed of a shot noise and super-sample covariance (SSC) contribution,

$$\text{Cov}\left[\Phi_\mu^i, \Phi_\nu^j\right] = \text{Cov}^{\text{SN}}\left[\Phi_\mu^i, \Phi_\nu^j\right] + \text{Cov}^{\text{SSC}}\left[\Phi_\mu^i, \Phi_\nu^j\right], \tag{A.1}$$

with

$$\text{Cov}^{\text{SN}}\left[\Phi_\mu^i, \Phi_\nu^j\right] = \delta_{i,j}\,\delta_{\mu,\nu}\,\frac{\Phi_\mu^i}{\Delta M_\star V_{\text{max},\mu}^i}, \tag{A.2}$$

where $\Phi_\mu^i = \Phi^i(M_{\star,\mu})$ and $V_{\text{max},\mu}^i = V_{\text{max}}^i(M_{\star,\mu})$ are the shorthands for the stellar mass function and V_{max} of stellar mass bin μ and generally a tomographic bin i. We define:

$$\widetilde{\Phi}_\mu = \int_0^\infty \Phi(M_{\star,\mu}|M,[z])\,n(M,z)\,b_{\text{h}}(M,z)\,\text{d}M, \tag{A.3}$$

using which the SSC terms are given by

$$\text{Cov}^{\text{SSC}}\left[\Phi_\mu^i, \Phi_\nu^j\right] = \frac{A_{\text{survey}}^2\,f^i f^j}{V_{\text{max},\mu}^i\,V_{\text{max},\nu}^j} \int \text{d}\chi\,\frac{p^i(\chi)}{p_{\text{tot}}(\chi)}\frac{p^j(\chi)}{p_{\text{tot}}(\chi)}\,f_K^2(\chi)\,\sigma_{\text{s}}^2(\chi)\,\widetilde{\Phi}_\mu[z(\chi)]\,\widetilde{\Phi}_\mu[z(\chi)], \tag{A.4}$$

where σ_{s}^2 is the variance of density fluctuations within the angular survey window (see Appendix E of Joachimi et al. (2021) for definition), p^i are the tomographic bin-wise redshift distributions, p_{tot} the overall redshift distribution for all galaxies in the sample/survey, f_K the comoving angular diameter distance, χ the comoving radial distance and f^i the fraction of galaxies in bin i relative to all galaxies. A_{survey} is the survey area.

The cross-variance is derived in close analogy to Takada & Bridle (2007), with the consistency checks by Schaan et al. (2014). The cross-variance receives contributions from two terms,

$$\text{Cov}\left[\Phi_\mu^i, O^{jl}(r_{\text{p}})\right] = \text{Cov}^{\text{CV}}\left[\Phi_\mu^i, O^{jl}(r_{\text{p}})\right] + \text{Cov}^{\text{SSC}}\left[\Phi_\mu^i, O^{jl}(r_{\text{p}})\right], \tag{A.5}$$

a cosmic variance (CV) and a super-sample covariance (SSC) term, respectively. Here we determine the cross-variance with a projected two-point function $O^{jl}(r_{\text{p}})$ either WP, $w_{\text{p}}(r)$, or ESD, $\Delta\Sigma(r)$, in bins j and l, respectively. The cosmic variance contribution is a three-point correlation given by

$$\text{Cov}^{\text{CV}}\left[\Phi_\mu^i, O^{jl}(r_{\text{p}})\right] = \bar{\rho}_{\text{x}} f^i \int \text{d}\chi\,\frac{p^i(\chi)}{p_{\text{tot}}(\chi)} \int \frac{\text{d}k\,k}{2\pi} J_{\text{x}}(kr_{\text{p}})\,(nB_{\text{cmm}})_\mu^{jl}[k,z(\chi)], \tag{A.6}$$

where $\bar{\rho}_{\text{x}} = 1$ and $J_{\text{x}} = J_0$ in the case when observable is w_{p}, and $\bar{\rho}_{\text{x}} = \bar{\rho}_{\text{m}}$ and $J_{\text{x}} = J_2$ in the case when observable is $\Delta\Sigma$. J_n are Bessel functions of n-th kind. Here we have defined the count-matter cross-bispectrum (evaluated for a collapsed triangle) in close analogy to Takada & Bridle (2007). It can be expressed in the halo model formalism as

$$(nB_{\text{cmm}})_\mu(k,z) = \int \text{d}M\,n(M,z)\,\Phi(M_{\star,\mu}|M,[z])\left(\frac{M}{\bar{\rho}_{\text{m}}}\right)^2 \tilde{u}_{\text{h}}^2(k|M) \tag{A.7}$$

$$+ 2P_{\text{lin}}(k,z)\int \text{d}M\,n(M,z)\,\Phi(M_{\star,\mu}|M,[z])\,b_{\text{h}}(M,z)\frac{M}{\bar{\rho}_{\text{m}}}\tilde{u}_{\text{h}}(k|M)\int \text{d}M'\,n(M',z)\,b_{\text{h}}(M',z)\frac{M'}{\bar{\rho}_{\text{m}}}\tilde{u}_{\text{h}}(k|M').$$

Finally, the SSC term reads

$$\text{Cov}^{\text{SSC}}\left[\Phi_\mu^i, O^{jl}(r_{\text{p}})\right] = \bar{\rho}_{\text{x}} A_{\text{survey}} f^i \int \text{d}\chi\,\frac{p^i(\chi)}{p_{\text{tot}}(\chi)} \int \frac{\text{d}k\,k}{2\pi} J_{\text{x}}(kr_{\text{p}})\,\frac{\partial P_{\text{xy}}^{jl}(k,z(\chi))}{\partial \delta_{\text{b}}}\,\sigma_{\text{s}}^2(\chi)\,\widetilde{\Phi}_\mu[z(\chi)], \tag{A.8}$$

where $P_{\text{xy}}(k,z)$ is either $P_{\text{gg}}(k,z)$ for w_{p} or $P_{\text{gm}}(k,z)$ for $\Delta\Sigma$, and the derivative is with respect to a super-survey density fluctuation δ_{b} (Takada & Hu 2013; Dvornik et al. 2018). As the systematic and statistical uncertainties on stellar masses are comparable in power (Brouwer et al. 2021), the entries in the SMF and cross-covariances are inflated by a factor of 2 to account for uncertainty arising from Eddington bias and the systematic shift in stellar masses.
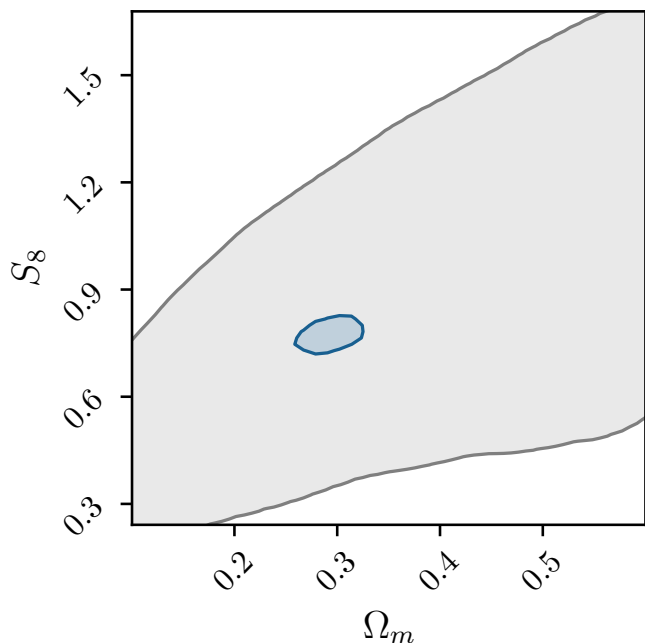
Fig. B.1: Prior range for the $\Omega_{\rm m}$ and S_8 parameters compared to the 2σ contour from our fiducial cosmological analysis.

Appendix B: Prior space

We adopt informative priors for three cosmological parameters; $n_{\rm s}$, $\Omega_{\rm b}$, and h. Our priors are also informative on the parameters which scale the mass-concentration relation for haloes and satellite galaxies; $f_{\rm h}$ and $f_{\rm s}$. In Fig. B.1 we verify that our choice of priors does not inform the $\Omega_{\rm m}$ and S_8 parameters, noting that our prior combination does not result in trivial square prior coverage.

Appendix C: Estimating the goodness of fit

To assess the goodness-of-fit of our model to the data it is necessary to determine the effective number of degrees of freedom, $\nu_{\rm eff}$, (see the discussion in section 6.3 of Joachimi et al. 2021, on why $\nu_{\rm eff} \neq N_{\rm data} - N_\theta$, for a typical cosmology analysis with $N_{\rm data}$ data points and N_θ cosmological parameters). We follow a simulation approach to determine $\nu_{\rm eff}$ following Joachimi et al. (2021); Miyatake et al. (2021).

   We generate 200 noisy mock data vectors, drawing samples from the multivariate distribution defined by the mean, which is our best-fit signal, and the full covariance matrix. We find the maximum posterior point using a Nelder-Mead minimisation and its corresponding χ^2 value for the fit. From the resulting distribution of χ^2 values, shown in Fig. C.1, we fit a χ^2 distribution to find the effective degrees of freedom[5], $\nu_{\rm eff}$ = 147.55, finding that our model provides a good fit to the data with $p(\chi^2|\nu_{\rm eff})$ = 0.27.

   We can use our simulation approach to also determine the goodness of fit for each of the three sections of the data vector: ESD (galaxy-galaxy lensing), $w_{\rm p}$ (galaxy clustering), and SMF (stellar mass function). To estimate the degrees of freedom for each observable we find the χ^2 value for that section of the data vector corresponding to the maximum posterior point that is

---

[5] We note that our simulation approach finds a slightly larger $\nu_{\rm eff}$ compared to the estimation using the Raveri & Hu (2019) approach, for which we find $\nu_{\rm eff}$ = 138.5,
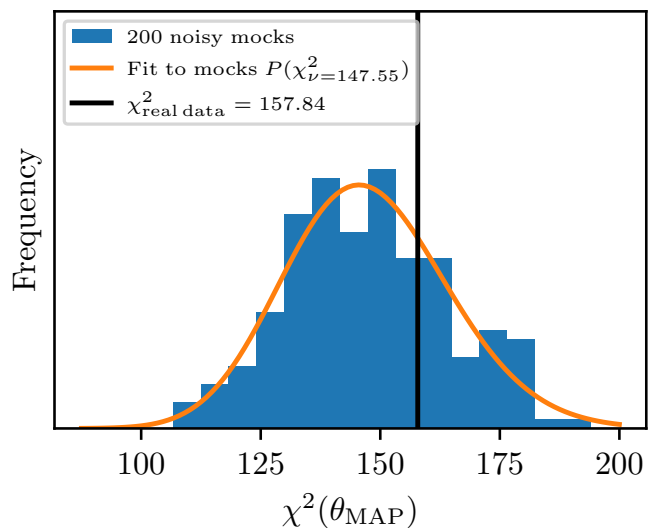

Fig. C.1: An estimation of goodness of fit of the fiducial best fit model at the maximum a posteriori (MAP) values. The histogram shows the distribution of the χ^2 values from 200 noisy mock data vectors (see text for detailed procedure). The orange line shows the fit of the χ^2 distribution to the histogram, obtaining the effective number of degrees of freedom in the data. The black vertical line shows the χ^2 value as obtained from the best fit model to the real data.

found using the full data vector. In other words the partial χ^2 values are not separately minimised. Fig. C.2 presents the resulting χ^2 distributions and fit. We note that some alternative methods to define $\nu_{\rm eff}$, such as Raveri & Hu (2019), would be ill defined for these sub-data vectors owing to the number of free parameters in the model. We find $\nu_{\rm eff}^{\rm ESD}$ = 73.18 with $p[\chi^2({\rm ESD})|\nu_{\rm eff}^{\rm ESD}]$ = 0.05, $\nu_{\rm eff}^{w_{\rm p}}$ = 72.62 with $p[\chi^2(w_{\rm p})|\nu_{\rm eff}^{w_{\rm p}}]$ = 0.01, and $\nu_{\rm eff}^{\rm SMF}$ = 14.58 with $p[\chi^2({\rm SMF})|\nu_{\rm eff}^{\rm SMF}]$ = 0.39. We conclude that the goodness of fit of the full data vector and each of the individual sections is acceptable. This procedure is repeated for the remaining modelling cases considered in this paper.

Appendix D: Galaxy-galaxy lensing random signal

The ratio between the lensing signal measured around random galaxies from the organised randoms catalogue and the fiducial model prediction in our six stellar mass bins is presented in Fig. D.1. A non-zero random signal indicates systematic effects in the ESD signal measured from the KiDS survey. The strong dependence of the random signal on the stellar mass of the lens bin indicates that this systematic is unlikely to be caused by residual uncorrected distortions in the source lensing catalogue, for example those associated with the point spread function of the camera. Such a data-related systematic would impact the ESD signal of each lens bin similarly given the similar source samples (see also Giblin et al. 2021, which presents a series of diagnostic tests). Instead we conclude that this signal arises from large scale structure sample variance that differs as the volume of the sample grows with the increasing stellar mass (Singh et al. 2017). We calculate the error on the mean random signal, measured using 1000 samples of the random catalogue, finding it to be indistinguishable from the thickness of the blue curve, and sufficiently small that we do not include it in our overall error budget. We find that the random correction applied to the data
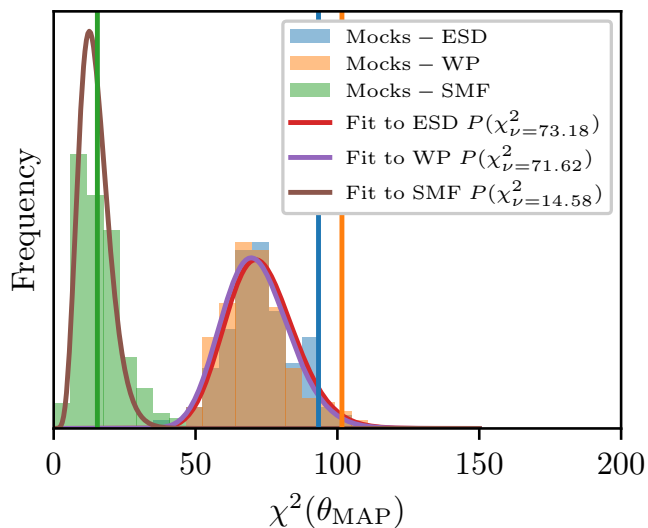
Fig. C.2: Same as Fig. C.1, but for the 3 different observables in our analysis. The fits to the distributions are used to determine the number of degrees of freedom for each observable which are in turn are used to determine the reduced χ^2 values for each of them respectively. Note that the χ^2 distributions are not a result of a maximising the posterior for that section of the mocks. The vertical lines show the χ^2 values from the real data, matching in colours with the histograms.

exceeds 100% in the first three stellar mass bins on large scales, and can be up to 4 times larger than the statistical uncertainty.

Appendix E: Extensions to the fiducial cosmological analysis

In this appendix we review the impact of modelling choices on our fiducial cosmological parameter constraints (Sect. 4) to a series of extensions to our fiducial theoretical model. The results, in terms of S_8 and Ω_m are summarised in Fig. E.1 with the full parameter space quantified in Table E.1 and Fig. E.2.

Appendix E.1: Modelling intrinsic galaxy alignments and magnification

We quantify the contribution of lens galaxy magnification and intrinsic galaxy alignments to our data, which we do not account for in our fiducial model. We estimate the contribution of intrinsic alignments (IA) using the NLA model (Bridle & King 2007) by

$$P_{\rm gI}(k,z) = -A_{\rm IA} C_1 \rho_{\rm crit} \frac{\Omega_{\rm m}}{D(z)} P_{\rm gm}(k,z), \qquad (E.1)$$

where $A_{\rm IA}$ is the amplitude of the intrinsic alignment signal, C_1 is a normalisation constant, $D(z)$ the linear growth factor. We set $C_1 \rho_{\rm crit} = 0.0134$, motivated by Brown et al. (2002). In order to estimate the contribution of intrinsic alignments to the galaxy-galaxy lensing signal, we set $A_{\rm IA} = 1$, which is a good 'worst-case' scenario for our complete magnitude-limited KiDS-Bright sample. We project the $P_{\rm gI}$ power spectrum to the galaxy-galaxy lensing signal. We subtract the additional contribution from the intrinsic alignments from the measured lensing signal and compare the relative difference between the corrected and uncorrected data to quantify the impact of neglecting intrinsic galaxy
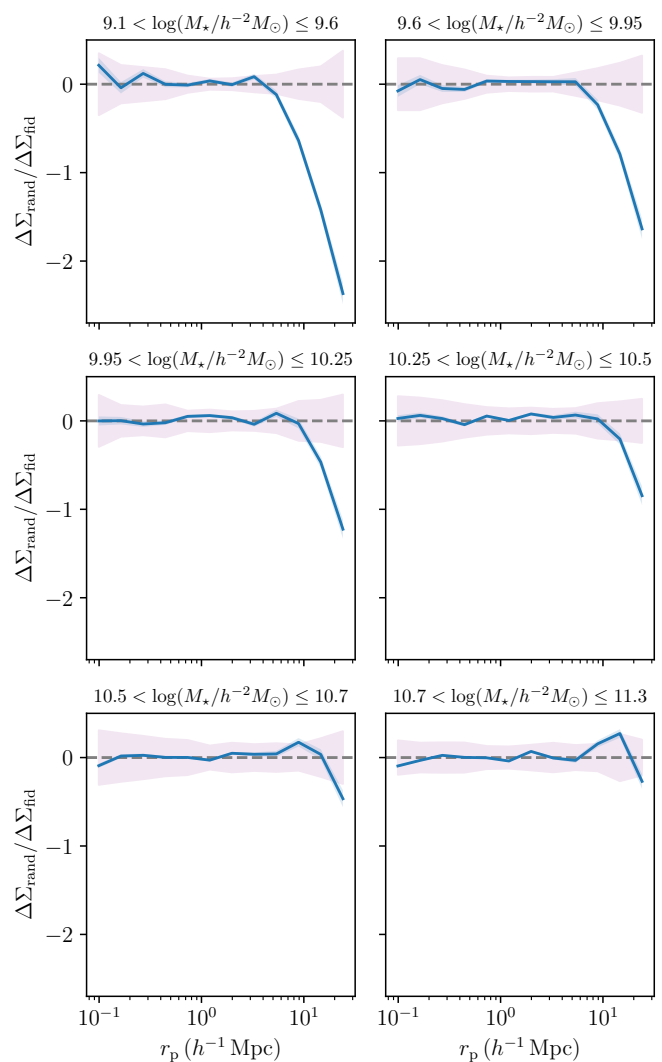
Fig. D.1: The ratio between the lensing signal measured around random galaxies from the organised randoms catalogue and the fiducial model prediction in our six stellar mass bins. For comparison we show the uncertainty of the data as a purple band.
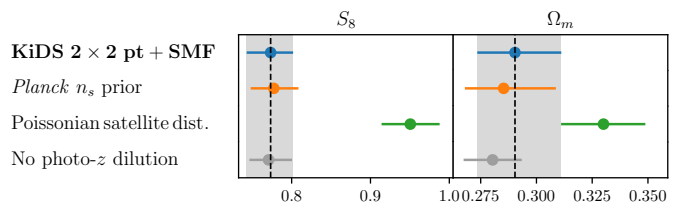

Fig. E.1: Comparison between S_8 and Ω_m values for our fiducial results and the tests for different modelling choices. All results are shown for maximum statistics of the marginal posterior distributions (MMAX) and corresponding credible interval.

alignments in our theoretical model. Such model is only describing the IA well on large scales and becomes increasingly ad hoc on small scales, but it nevertheless provides for a sensible amplitude estimation. Future studies will use the halo model based approach as presented in Fortuna et al. (2021) and Georgiou et al.

Table E.1: Marginal constraints on all model parameters listed together with their priors, for fiducial and extension setups considered.

Parameter	Prior	Fiducial		*Planck* n_s prior		Poissonian satellite dist.	
		MMAX	MAP+PJ-HPD	MMAX	MAP+PJ-HPD	MMAX	MAP+PJ-HPD
$\Omega_{\rm m}$	[0.1, 0.6]	$0.290^{+0.021}_{-0.017}$	$0.307^{+0.002}_{-0.031}$	$0.285^{+0.024}_{-0.017}$	$0.289^{+0.010}_{-0.026}$	0.330 ± 0.019	$0.323^{+0.019}_{-0.011}$
σ_8	[0.4, 1.2]	$0.781^{+0.033}_{-0.029}$	$0.801^{+0.013}_{-0.041}$	$0.792^{+0.035}_{-0.031}$	$0.791^{+0.038}_{-0.020}$	$0.902^{+0.044}_{-0.034}$	$0.953^{+0.009}_{-0.067}$
h	[0.64, 0.82]	< 0.726	< 0.711	< 0.716	< 0.715	< 0.720	< 0.712
$\Omega_{\rm b}$	[0.01, 0.06]	> 0.01	> 0.01	> 0.01	> 0.01	> 0.01	> 0.01
n_s	[0.92, 1.1]	< 1.004	< 0.978	$0.964^{+0.005}_{-0.004}$	$0.965^{+0.003}_{-0.005}$	< 0.988	< 0.979
S_8	–	$0.773^{+0.028}_{-0.030}$	$0.809^{+0.001}_{-0.055}$	$0.777^{+0.031}_{-0.029}$	$0.776^{+0.030}_{-0.023}$	$0.951^{+0.037}_{-0.036}$	$0.989^{+0.020}_{-0.057}$
$f_{\rm h}$	[0.0, 1.0]	> 0.645	> 0.939	> 0.665	> 0.948	> 0.652	> 0.938
M_0	[7.0, 13.0]	$10.519^{+0.039}_{-0.062}$	$10.521^{+0.062}_{-0.044}$	$10.520^{+0.042}_{-0.060}$	$10.537^{+0.051}_{-0.054}$	$10.446^{+0.053}_{-0.044}$	$10.511^{+0.010}_{-0.088}$
M_1	[9.0, 14.0]	$11.138^{+0.099}_{-0.132}$	$11.145^{+0.136}_{-0.098}$	$11.129^{+0.115}_{-0.126}$	$11.168^{+0.080}_{-0.141}$	$11.039^{+0.153}_{-0.123}$	$11.197^{+0.033}_{-0.238}$
γ_1	[2.5, 15.0]	$7.096^{+2.144}_{-1.406}$	$7.385^{+1.441}_{-1.824}$	$7.026^{+2.157}_{-1.454}$	$6.774^{+2.174}_{-1.014}$	$8.364^{+2.831}_{-2.028}$	$6.334^{+4.370}_{-0.343}$
γ_2	[0.0, 10.0]	0.201 ± 0.010	0.201 ± 0.009	0.201 ± 0.011	0.211 ± 0.010	0.201 ± 0.010	0.207 ± 0.007
$\sigma_{\rm c}$	[0.0, 2.0]	$0.108^{+0.067}_{-0.011}$	$0.159^{+0.007}_{-0.070}$	$0.105^{+0.062}_{-0.015}$	$0.098^{+0.063}_{-0.014}$	$0.178^{+0.021}_{-0.023}$	$0.123^{+0.062}_{-0.008}$
$f_{\rm s}$	[0.0, 1.0]	> 0.377	> 0.84	> 0.394	> 0.858	> 0.223	> 0.808
$\alpha_{\rm s}$	[−5.0, 5.0]	$-0.858^{+0.048}_{-0.052}$	$-0.847^{+0.013}_{-0.097}$	$-0.874^{+0.059}_{-0.042}$	$-0.898^{+0.062}_{-0.043}$	$-0.844^{+0.042}_{-0.043}$	$-0.916^{+0.092}_{-0.002}$
b_0	[−5.0, 5.0]	$-0.024^{+0.108}_{-0.117}$	$-0.120^{+0.199}_{-0.001}$	$-0.016^{+0.098}_{-0.028}$	$-0.077^{+0.163}_{-0.118}$	-0.239 ± 0.092	$-0.267^{+0.104}_{-0.058}$
b_1	[−5.0, 5.0]	$1.149^{+0.091}_{-0.081}$	$1.177^{+0.058}_{-0.096}$	$1.139^{+0.091}_{-0.078}$	$1.181^{+0.037}_{-0.111}$	$1.045^{+0.076}_{-0.060}$	$1.019^{+0.092}_{-0.029}$
$\mathcal{P}$	[0.0, 2.0]	0.403 ± 0.029	$0.417^{+0.024}_{-0.013}$	0.403 ± 0.030	$0.411^{+0.014}_{-0.011}$	1.0 fixed	1.0 fixed
$\mathcal{D}$	[0.0, 0.3]	$0.144^{+0.091}_{-0.085}$	$0.051^{+0.172}_{-0.006}$	$0.150^{+0.090}_{-0.082}$	$0.195^{+0.029}_{-0.147}$	$0.067^{+0.082}_{-0.064}$	$0.140^{+0.016}_{-0.132}$
$\chi^2_{\rm red}$	–	1.07		1.07		1.25	
p-value	–	0.27		0.27		0.02	

Notes: This table lists all the free parameters in our model: the energy density of cold matter $\Omega_{\rm m}$, the normalisation of power spectrum σ_8, the dimensionless Hubble parameter h, the spectral index n_s, the energy density of baryonic matter $\Omega_{\rm b}$, the derived parameter S_8, the normalisation of the concentration-mass relation for dark matter haloes $f_{\rm h}$, the normalisation of stellar-to-halo mass relation M_0, the characteristic scale of the stellar-to-halo mass relation M_1, the slope parameters of the stellar-to-halo mass relation γ_1 and γ_2, the scatter between stellar mass and halo mass $\sigma_{\rm c}$, the normalisation of the concentration-mass relation for distribution of satellite galaxies $f_{\rm s}$, the power law behaviour of satellites $\alpha_{\rm s}$, the normalisation constants of the Schechter function b_0 and b_1, and the Poisson parameter $\mathcal{P}$. Parameters are deemed unconstrained when the marginal probability at 2σ level exceeds 13% of the peak probability (see appendix A of Asgari et al. 2021b). In cases where one side is constrained we report the 1σ lower/upper limit. The MMAX estimate is the marginal maximum statistic, reporting the point of maximum marginal posterior distribution to the iso-posterior levels above and below the maximal point. The MAP+PJ-HPD (maximum posterior with projected joint highest posterior density) estimates are calculated following Joachimi et al. (2021).

(2019), which links the IA signal to the properties of the galaxies an their parent dark matter haloes.

The magnification contribution to the total lensing and clustering signal is modelled as

$$\Delta\Sigma_{\rm mag}(r_{\rm p}) = 2(\alpha_{\rm d} - 1)\,\Delta\Sigma_{\rm mm}(r_{\rm p}) \qquad (E.2)$$

and

$$w_{\rm p,mag}(r_{\rm p}) = 4(\alpha_{\rm d} - 1)\,w_{\rm gm}(r_{\rm p}) + 4(\alpha_{\rm d} - 1)^2\,w_{\rm mm}(r_{\rm p}), \qquad (E.3)$$

where $\Delta\Sigma_{\rm mm}(r_{\rm p})$ and $w_{\rm mm}(r_{\rm p})$ are the lensing signal and projected clustering contributions from the convergence power spectrum, respectively (Joachimi & Bridle 2010; Simon & Hilbert 2018; Unruh et al. 2020). The $w_{\rm gm}(r_{\rm p})$ is the same quantity as the $\Sigma(r_{\rm p})$, but without the mean density $\bar{\rho}_{\rm m}$ (see Eq. 35). We estimate the contribution from lens magnification using the values for a KiDS and GAMA like survey, given by Unruh et al. (2020). We adopt $\alpha_{\rm d} = 0.85$ corresponding to a lens redshift of 0.21, and

$\alpha_{\rm d} = 2.11$ corresponding to a lens redshift of 0.36. These values are representative of the magnification effect we would expect for the two highest stellar mass bins in our analysis, with other bins predicted to have a somewhat smaller contribution, due to their redshifts (Unruh et al. 2020). We subtract the magnification contribution from our data to quantify the impact of neglecting magnification in our theoretical model.

The relative contributions of intrinsic alignments and magnification to the ESD and $w_{\rm p}$ observables are presented in Fig. E.3. We present the contribution for the largest stellar mass bin only, as the contribution to magnification and intrinsic alignments is of the same order for all of them. On the same figure we also show the changes to the galaxy-galaxy lensing and galaxy clustering signals if we change the two main cosmological parameters. The contributions of both intrinsic alignments and magnification are well below 1%, as also found in Unruh et al. (2020) and are well within the error budget of the data. Moreover, they are also sub-
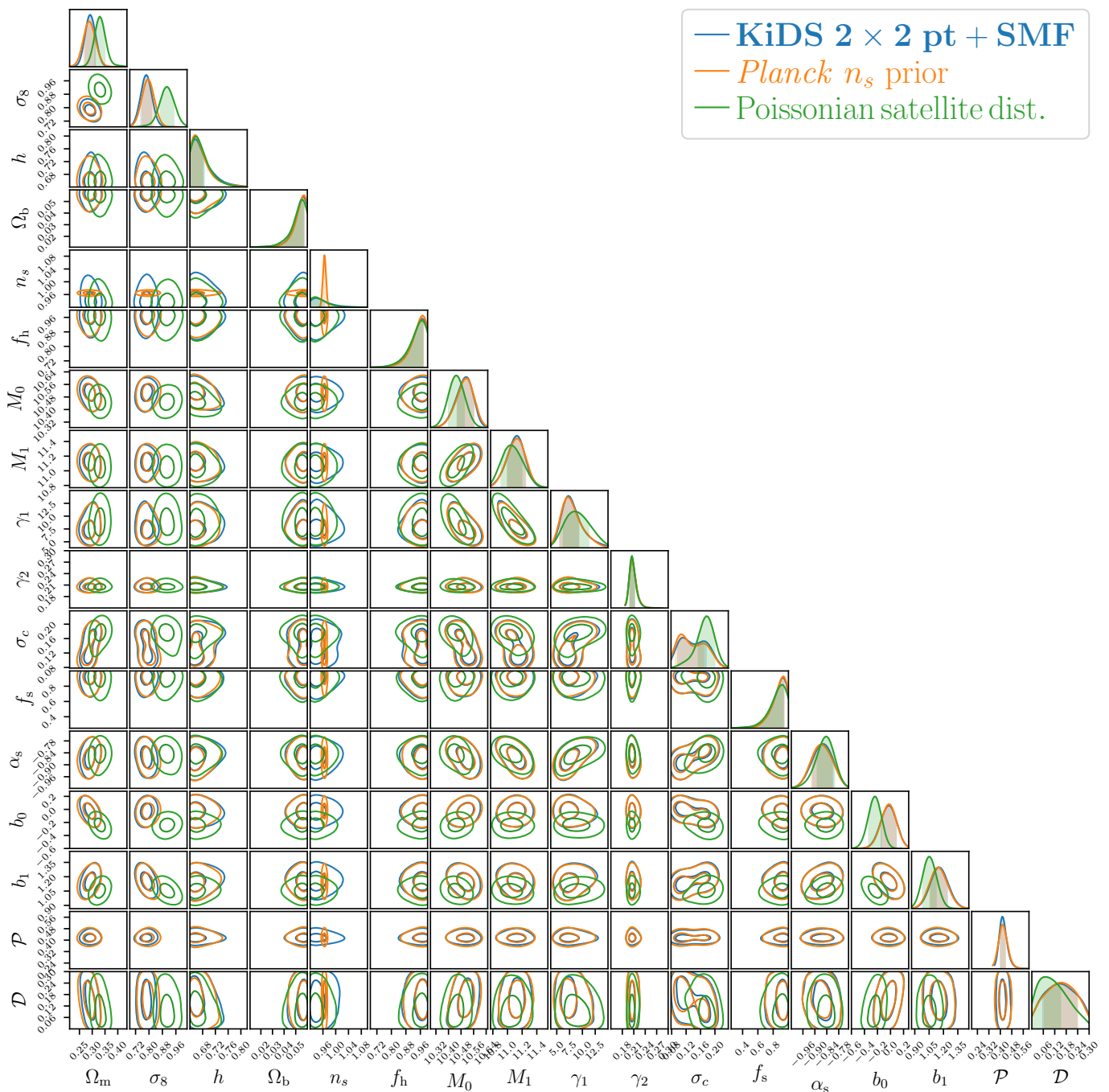
Fig. E.2: The full posterior distributions of the model parameters (where the priors are listed in Table 2). The contours indicate 1σ and 2σ confidence regions.

dominant to the changes due to the cosmological parameters of interest. Such contributions have negligible effects on the overall signal, and are unlikely to be a significant source of bias. To some extent the effects cancel each other out, as for the redshifts of our lens galaxies the magnification effect dilutes the signal, while the intrinsic alignments add a similar contribution.

Appendix E.2: Changing the n_s prior

Inspecting Fig. E.2 we note that the marginalised posteriors of our prior-informed cosmological parameter set, n_s, Ω_b and h, have a tendency to push up against one side of the prior. As the

KiDS 2×2pt+SMF data vector is expected to be insensitive to changes in this parameter set, we conclude that this effect arises from projection effects or the MCMC not fully sampling this part of parameter space. Given that there are no strong degeneracies between this set and the rest of the parameters, and that the set is already well constrained from other studies (see the discussion in appendix B of Heymans et al. 2021) we find no motivation to investigate the impact of widening the priors. Instead we investigated reducing the prior range by adopting a Gaussian prior with the mean and uncertainty fixed to the Planck Collaboration et al. (2020) constraint for n_s. This parameter is the most interesting to investigate of the three, as any tension between small
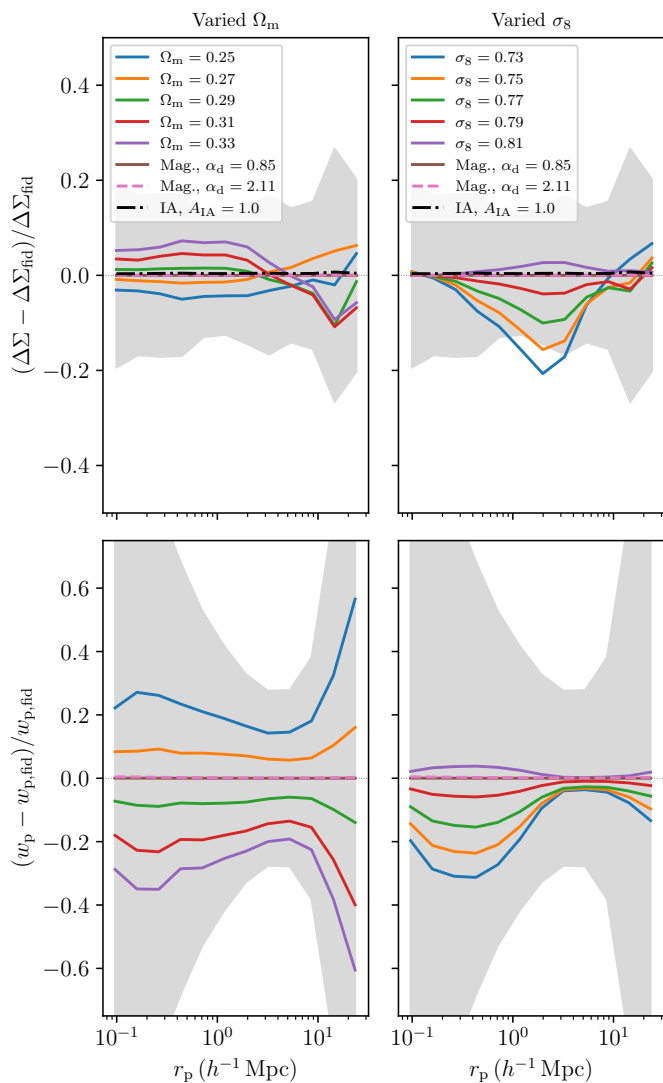
Fig. E.3: Relative difference between the lensing signal with and without the contribution of intrinsic alignments and lens magnification. The intrinsic alignment contribution is modelled with $A_{IA} = 1$, which is a reasonable 'worst case' scenario for the KiDS-Bright sample, and the magnification is modelled with $\alpha_d = 0.85$ and $\alpha_d = 2.11$, typical for GAMA-like lenses at a redshift of 0.21 and 0.36. Grey areas show the error on the data. We also show the changes to the galaxy-galaxy lensing and galaxy clustering signals if we change the two main cosmological parameters. The behaviour is only shown for the largest stellar mass bin, as the contribution to magnification and intrinsic alignments is of the same order for all of them.

and large scales may be expected to manifest in a biased spectral index constraint (Tröster et al. 2021).

The results are presented in Table E.1 and Fig. E.2. We find that the marginal contours for the parameters do not change significantly with the addition of a restrictive prior on n_s. For example, the MMAX estimate of S_8, shifts by 0.13σ, which is consistent with the expected MCMC run-to-run variance (Joachimi et al. 2021). In future studies we will explore the impact of using more restrictive priors on all externally-constrained parameters which we are insensitive to, noting that this may also help to reduce projection bias for posteriors with many parameters.
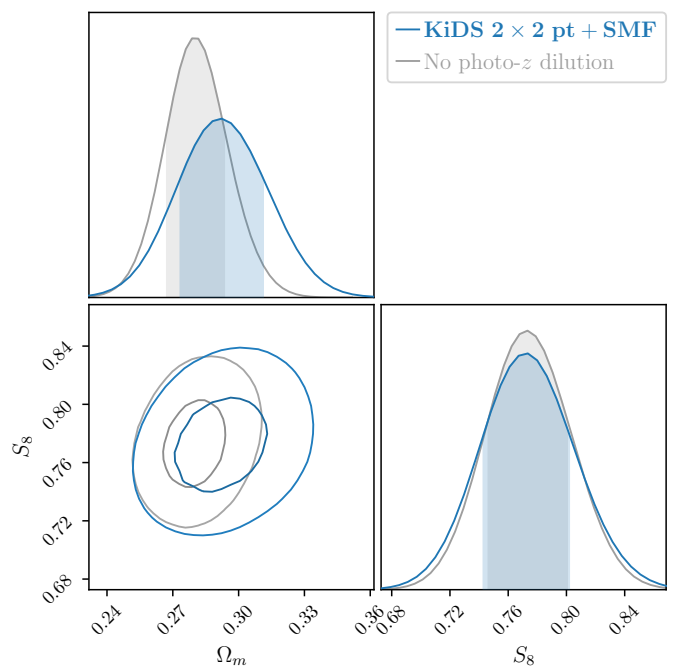

Fig. E.4: Marginalised constraints for the joint distributions of S_8 and Ω_m, showing 68% and 95% credible regions. We compare our fiducial analysis (blue) with an analysis neglecting the impact of photometric redshift uncertainties which dilute the estimated projected galaxy clustering signal (grey).

Appendix E.3: The effect of unmodelled photometric redshift errors on the projected galaxy clustering measurements

In Sect. 3.3 we introduce a free nuisance parameter, $\mathcal{D}$ (Eq. 43), to account for our uncertainty on the amplitude of the true projected galaxy clustering signal. The expected dilution is a result of unaccounted photometric redshift errors in our theoretical model for $w_p(r_p)$. In Figure E.4 we compare our fiducial constraints in S_8 and Ω_m with the constraints from an analysis where the photometric redshift dilution effect is neglected and $\mathcal{D} = 1$. We find that while the omittance of the photometric redshift dilution effect does not impact the S_8 constraints, Ω_m becomes biased and more constrained. This is expected as the galaxy clustering is more sensitive to Ω_m compared to σ_8 (cf. Fig. E.3). This motivates future work to improve the estimator for, or theoretical modelling of, the projected galaxy clustering signal in the presence of photometric redshift errors, following Joachimi et al. (2011); Chisari et al. (2014).

Appendix E.4: Modelling satellite galaxies

In our fiducial model we used the findings of Dvornik et al. (2018) that showed that the occupation distribution of satellite galaxies does not follow a Poisson distribution, and that generally the parameter $\mathcal{P}$ (Eq. 25) is not unity, with our fiducial run preferring a sub-Poissonian behaviour. Following Cacciato et al. (2013), we quantify the impact of removing this flexibility in the model, by fixing the parameter $\mathcal{P}$ to unity,

$$\mathcal{P}(M) \equiv \frac{\langle N_s(N_s - 1)|M\rangle}{\langle N_s|M\rangle^2} \equiv 1 \,, \qquad (E.4)$$

thus assuming that the satellite galaxies obey the Poisson distribution. We run another set of MCMC chains using the same

setup as in the fiducial case, but with one less parameter to constrain. The resulting constraints are shown in Fig. E.2, with marginalised constraints quoted in Table E.1. We find significant shifts for the two main cosmological parameters with $\Omega_{\rm m}$ = 0.330±0.019 and $S_8 = 0.961^{+0.037}_{-0.036}$ and a formally acceptable fit with $p(\chi^2|\nu_{\rm eff})$ = 0.02 for the whole data vector. In this case we find that the fixed Poisson parameter non-trivially affects the other parameters governing the satellites in the halo model. Specifically the normalisation of the satellite conditional stellar mass function, b_0 and b_1, shifts, and these parameters are in turn non-trivially correlated with the main cosmological parameters.

There are several reasons to reject the Poissonian satellite distribution model. Using the GAMA group catalogue (Robotham et al. 2011) we can directly measure $\mathcal{P}(M)$ as a function of the dynamical mass $M_{\rm dyn}$ (Driver et al. 2022). We find the satellite distribution to be sub-Poissonian for $M_{\rm dyn}$ < $10^{13}h^{-1}M_\odot$, ranging from $\mathcal{P}(M_{\rm dyn}) \sim 0.4$ at $M_{\rm dyn} = 10^{11}h^{-1}M_\odot$, to $\mathcal{P}(M_{\rm dyn}) \sim 1$ at $M_{\rm dyn} = 10^{13}h^{-1}M_\odot$. Assuming the GAMA dynamical mass is a reasonable estimate of the halo mass $M_{\rm h}$, using Fig. 7 we expect the satellite distribution to be sub-Poissonian across the full stellar mass range of our 2×2pt+SMF analysis. This finding is supported by hydrodynamical simulations (see figure 7 of Dvornik et al. 2018).

Finally, even though we find an acceptable fit of this model to our full 2×2pt+SMF data vector, there is an unacceptable fit to the $w_{\rm p}$ part of the data vector for this $\mathcal{P}$ = 1 analysis, with $p[\chi^2(w_{\rm p})|\nu_{\rm eff}^{w_{\rm p}}] \sim 10^{-4}$. Our results are therefore consistent with Cacciato et al. (2013), confirming their conclusion that including flexibility in the form of the satellite galaxy model is critically important in order to both constrain the galaxy bias (Cacciato et al. 2012; Dvornik et al. 2018; Asgari et al. 2021a), and to obtain unbiased cosmological parameters.