

Mapping stellar content to dark matter haloes using galaxy clustering and galaxy–galaxy lensing in the SDSS DR7

Ying Zu[★] and Rachel Mandelbaum

McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

Accepted 2015 September 2. Received 2015 September 2; in original form 2015 May 7

ABSTRACT

The mapping between the distributions of the observed galaxy stellar mass and the underlying dark matter haloes provides the crucial link from theories of large-scale structure formation to interpreting the complex phenomena of galaxy formation and evolution. We develop a novel statistical method, based on the halo occupation distribution (HOD) model, to solve for this mapping by jointly fitting the galaxy clustering and the galaxy–galaxy lensing from the Sloan Digital Sky Survey (SDSS). The method, called the *i*HOD model, extracts maximum information from the survey by including ~ 80 per cent more galaxies than the traditional HOD methods, accounting for the incompleteness of the stellar mass samples self-consistently. The derived stellar-to-halo mass relation (SHMR) explains the clustering and lensing of SDSS galaxies over four decades in stellar mass, while successfully predicting the observed stellar mass functions (SMFs). By modelling significantly more galaxies, the *i*HOD breaks the degeneracy between the logarithmic scatter in the stellar mass at fixed halo mass and the slope of the mean SHMR at high masses, without assuming a strong prior on the scatter and/or using the SMF as an input. We detect a decline of the scatter with halo mass, from $0.22^{+0.02}_{-0.01}$ dex below $10^{12} h^{-1} M_{\odot}$ to 0.18 ± 0.01 dex at $10^{14} h^{-1} M_{\odot}$. The model predicts a departure of satellite SMFs from the Schechter form in massive haloes and a linear scaling of satellite number with halo mass. The *i*HOD model can be easily applied to other spectroscopic data sets, greatly improving statistical constraints on the SHMR compared to traditional HOD methods within the same survey.

Key words: gravitational lensing: weak – methods: statistical – galaxies: luminosity function, mass function – cosmology: observations – large-scale structure of Universe.

1 INTRODUCTION

The distribution of galaxy stellar mass in the present-day Universe provides important clues to answering fundamental questions concerning the cosmic assembly of baryons in the Λ cold dark matter (Λ CDM) paradigm (Fukugita, Hogan & Peebles 1998; Kereš et al. 2005; Faucher-Giguère, Kereš & Ma 2011; Davé, Finlator & Oppenheimer 2012). In particular, what fraction of baryons are condensed into stars as opposed to spreading out in the form of gas and dust in and outside of galaxies (Cen & Ostriker 1999; McGaugh et al. 2010; Putman, Peek & Joungh 2012; Shull, Smith & Danforth 2012)? Among those locked in stars, how much stellar mass is stored within the central galaxies of the dark matter haloes (De Lucia & Blaizot 2007; von der Linden et al. 2007), and how is the rest distributed among the satellite galaxies within their larger host haloes (Hansen et al. 2009; Yang, Mo & van den Bosch 2009;

Leauthaud et al. 2012b)? In this paper, we develop a novel statistical approach within the halo occupation distribution (HOD; Jing, Mo & Börner 1998; Ma & Fry 2000; Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Cooray & Sheth 2002; Guzik & Seljak 2002; Yang, Mo & van den Bosch 2003; Zheng et al. 2005; Mandelbaum et al. 2006a; van den Bosch et al. 2007) framework, to solve for the mapping between the stellar mass content and the dark matter haloes using the spatial clustering and the weak gravitational lensing of the Sloan Digital Sky Survey (SDSS; York et al. 2000) spectroscopic sample galaxies. The inferred mapping not only explains the galaxy autocorrelation function (i.e. clustering) and the galaxy–matter cross-correlation (i.e. lensing) successfully, but also correctly predicts the observed stellar mass function (SMF), placing strong constraints on the physics that governs the formation and distribution of galaxies within haloes.

The most comprehensive way to link the galaxy properties to the dark matter haloes is to directly model the expected physical processes involved in the formation and evolution of stars and gas (along with metals and dust), by either running

[★] E-mail: yzu@cmu.edu

hydrodynamic simulations (Hernquist & Katz 1989; Katz, Weinberg & Hernquist 1996; Norman & Bryan 1999; Teyssier 2002; O’Shea et al. 2004; Di Matteo, Springel & Hernquist 2005; Springel et al. 2005; Oppenheimer & Davé 2006; Booth & Schaye 2009; Vogelsberger et al. 2013) or semi-analytical models (SAMs) along the halo merger trees of N -body simulations (Kang et al. 2005; Baugh 2006; Bower et al. 2006; De Lucia et al. 2006; Somerville et al. 2008). The primary advantage of the latter approach is that it is computationally much less expensive compared to the hydrodynamic simulations, albeit with a large number of free parameters. However, even within the hydrodynamic simulations some of the key processes, like the star formation from the collapse of molecular clouds, the stellar feedback from supernovae explosions and galactic winds, and the impact of the active galactic nuclei (i.e. the AGN feedback), are happening on scales well below the resolution limit, so the treatment of the physics is often at the ‘subgrid’ level, i.e. put in by hand using empirical scaling relations (see Benson 2010, for a review). Both the hydrodynamic simulations and the SAMs none the less have enjoyed great success over the past decade, reproducing a wide range of the observed galaxy properties (see Oppenheimer & Davé 2008, Vogelsberger et al. 2014, and Khandai et al. 2015 for the latest hydrodynamic simulations and see Guo et al. 2011 for the recent development in SAM), though still with several possible areas of improvement. Most importantly, because the simulated outputs are directly tied to the prescribed subgrid physics and its parameters, by confronting the predictions of this ‘forward modelling’ approach with new observations, we can constantly furnish our understanding of the galaxy formation physics in a cosmological context, especially the baryonic feedback mechanisms that help sculpt the galaxy SMF (Fontanot et al. 2009). However, the computational complexities of hydrodynamic simulations and the SAM degeneracies make it difficult to come to definite conclusions about galaxy formation theories in certain cases.

Alternatively, one can focus on the *statistical* link between just the stellar content and the dark matter haloes, assuming that the enormous diversity in the individual galaxy assembly histories inside haloes of the same mass would reduce to a stochastic scatter about the mean stellar-to-halo mass relation (SHMR) by virtue of the central limit theorem. The great success of the HOD framework in explaining the clustering and lensing of galaxies, and their dependences on galaxy properties like colour, luminosity, and stellar mass at different redshifts (e.g. Zehavi et al. 2011; Parejko et al. 2013; Guo et al. 2014), further validates this assumption about the statistical simplicity of the SHMR. For a given cosmology, the HOD formalism describes the relationship between the stellar and dark matter in terms of $\langle N_g(M_h) \rangle$, the average number of galaxies of a given type (e.g. central or satellite) within a halo of virial mass M_h , along with the spatial and velocity distributions of galaxies within that halo (see Yang et al. 2003 for a closely related approach, the conditional luminosity function). Furthermore, the HOD is potentially a powerful tool to constrain cosmology (Yoo et al. 2006; Zheng & Weinberg 2007; Cacciato et al. 2013; van den Bosch et al. 2013; More et al. 2015), by exploiting the average bias versus mass relation of the dark matter haloes revealed by the galaxies they contain. The standard HOD modelling is restricted to individual volume-limited samples, each defined by a stringent combination of stellar mass and redshift cuts, which leaves many observed galaxies unused. Moreover, the corresponding HOD parameters are inferred separately for each sample, therefore the SHMR is constrained at a disjunct set of loci where the mean stellar masses of the samples landed.

Recently, Leauthaud et al. (2011, hereafter L11) proposed a new HOD framework by further parametrizing $\langle N_g(M_h) \rangle$ as continuous functions of the threshold stellar mass M_* , i.e. $\langle N_g(> M_* | M_h) \rangle$. In particular, the mean SHMR of central galaxies and its scatter fully specify the expected number of central galaxies of any stellar mass at fixed halo mass, whereas the number of satellite galaxies above a certain stellar mass scales with halo mass in a way that is also halo mass dependent. The advantage of this new framework lies in its unique capability to derive the connection between galaxies and haloes using multiple probes simultaneously within a single global HOD model. Leauthaud et al. (2012a, hereafter L12) demonstrated the efficacy of the L11 model by inferring the SHMR across the entire observed stellar mass range, using the combination of the SMF, angular galaxy clustering, and galaxy–galaxy (g–g) lensing of samples above some critical stellar mass limit (corresponding to high completeness) in the cosmological evolution survey (COSMOS). However, similar to the standard HOD technique, the L11 model requires volume completeness in the stellar mass samples, thereby losing a lot of data in an intrinsically flux-limited survey. Our approach is based on the L11 global HOD framework, and (like in L11) can jointly fit the galaxy clustering and g–g lensing signals of galaxy samples above some critical stellar mass limit, but (unlike in L11) can take into account the incompleteness in a statistically self-consistent way. As a result of the added flexibility of our model, we are able to include significantly more SDSS galaxies at both lower stellar mass and higher redshift, equivalent to almost doubling the survey volume.

Meanwhile, Moster et al. (2010, hereafter M10) and other groups (e.g. Kravtsov et al. 2004; Vale & Ostriker 2004; Conroy, Wechsler & Kravtsov 2006; Shankar et al. 2006; Behroozi, Conroy & Wechsler 2010; Guo et al. 2010) adopted the so-called ‘subhalo abundance matching’ (SHAM) technique to assign stellar masses or luminosities to individual dark matter haloes (including both main and subhaloes) in the N -body simulations. In its simplest form, the SHAM method assumes a monotonic relation between the SMF estimated from galaxy surveys and the halo mass function predicted by the Λ CDM. Instead of using the subhalo mass at the current epoch, most SHAM studies employed the ‘infall mass’, i.e. the mass of the subhalo before its accretion on to the main halo, and instead of a monotonic relation they either assumed some fixed scatter or drew the scatter from external priors. Using the ‘infall mass’ as a proxy for the stellar mass, they found surprisingly good agreement between the predicted and the observed clustering and lensing statistics. The underlying assumption behind SHAM is that the satellite galaxies that live in subhaloes were central galaxies of their own haloes before accretion. After becoming satellites, the dark matter masses of their subhaloes were reduced due to the combined effect of tidal stripping and dynamical friction, whereas their stellar masses in the centre of those subhaloes were much less affected and are thus more closely tied to the infall masses (van den Bosch, Tormen & Giocoli 2005; Reddick et al. 2013). Hearin & Watson (2013) later extended this assumption to further relate the halo formation time to the colour of each galaxy to interpret the colour dependence of the clustering and lensing statistics (also see Kulier & Ostriker 2015, for an interesting proposal of a two-parameter abundance matching scheme). The key distinction between the SHAM and the HOD models is that, by using the infall mass the SHAM method evades the need to parametrize the HOD of the satellite galaxies (but see Guo et al. 2011, on potential numerical issues in tracking subhaloes), which is usually regarded as a nuance in traditional HOD modelling, at the expense of assuming the same SHMRs for the central and the satellite galaxies. However, as we

will demonstrate in this study, the capability of our modified HOD model to maximally extract the information from data enables us to place tight constraint on the satellite HOD as well, thus shedding new insight into the formation and evolution of the satellite galaxies without imposing the central SHMR on the satellites. Moreover, the model breaks the degeneracy between the scatter and the slope of the SHMR, self-consistently deriving the scatter without the need to assume fixed values or external priors.

The paper is organized as follows. Section 2 describes the SDSS data, including the large-scale structure galaxy catalogue and the matched stellar mass catalogue, and the two sets of galaxy samples selected for our analyses. In Section 3, we describe the methods we use to measure galaxy clustering and g - g lensing for these samples. Section 4 introduces our new variant of the HOD model that we use to interpret the clustering and lensing signals, and Section 5 describes the method to theoretically predict these signals using our HOD model. In Section 6, we present the results of our constraint from a Bayesian framework via the new HOD modelling technique. We examine the SMF, the SHMR of the central galaxies, and the conditional SMFs of the satellites predicted by our best-fit model in Section 7. We summarize our main findings and discuss future applications of the new HOD model in Section 8.

Throughout this paper, we assume a Λ CDM cosmology with $(\Omega_m, \Omega_\Lambda, \sigma_8, h) = (0.26, 0.74, 0.77, 0.72)$, same as that used for abundance matching in M10 that we intend to compare with. All the length and mass units in this paper are scaled as if the Hubble constant were $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. In particular, all the separations are comoving distances in units of either $h^{-1} \text{ kpc}$ or $h^{-1} \text{ Mpc}$, and the stellar mass and halo mass are in units of $h^{-2} M_\odot$ and $h^{-1} M_\odot$, respectively. The halo mass is defined by $M_h \equiv M_{200m} = 200 \bar{\rho}_m (4\pi/3) r_{200m}^3$, where r_{200m} is the corresponding halo radius within which the average density of the enclosed mass is 200 times the mean matter density of the Universe, $\bar{\rho}_m$. For the sake of simplicity, $\ln x = \log_e x$ is used for the natural logarithm, and $\lg x = \log_{10} x$ is used for the base-10 logarithm.

2 DESCRIPTION OF THE DATA

Here we describe the spectroscopic data used to define the clustering and the lens samples, and the imaging data used to define the source samples for g - g lensing.

2.1 SDSS main galaxy sample and NYU-VAGC

This study is based on the final data release of the SDSS (DR7; Abazajian et al. 2009), which contains the completed data set of the SDSS-I and the SDSS-II. The survey imaged a quarter of the sky using a drift-scan camera (Gunn et al. 1998) in five photometric bandpasses (u, g, r, i, z ; Fukugita et al. 1996) to a limiting magnitude of $\simeq 22.5$ in the r band. The imaging data were photometrically and astrometrically calibrated (Padmanabhan et al. 2008), and from this imaging data targets were selected for spectroscopic follow-ups with a fibre-fed double spectrograph (Gunn et al. 2006). One of the spectroscopic products is ‘the main galaxy sample’ (MGS; Strauss et al. 2002) that we use in this study as both the tracers of stellar mass clustering and the lenses of background sources. In particular, we obtain the MGS data from the dr72 large-scale structure sample `bright0` of the ‘New York University Value Added Catalogue’ (NYU-VAGC), constructed as described in Blanton et al. (2005). The `bright0` sample includes galaxies with $10 < m_r < 17.6$, where m_r is the r -band Petrosian apparent magnitude, corrected for Galactic extinction. We choose a more relaxed bright limit than the

commonly used `safe0` sample ($14.5 < m_r < 17.6$) to allow higher completeness on the high stellar mass end, where constraint on the mapping between galaxies and haloes is particularly lacking.

Due to the finite size of the fibre plugs, no two targets on the same plate can be closer than 55 arcsec, resulting in ~ 7 per cent of the MGS galaxies with unknown redshifts. A simple remedy for these fibre collisions, as used in the `bright0` sample, is to assign these galaxies the redshifts of their nearest neighbours on the sky, thus exactly preserving the angular clustering signal. For our purpose of measuring the projected correlation w_p at a given redshift z , it would overestimate the clustering signal below the projected physical scale corresponding to the fibre size at that redshift, $r_p^{\text{fb}}(z)$, by including physically distant pairs of galaxies that happen to align within 55 arcsec on the sky. Above $r_p^{\text{fb}}(z)$, however, this ‘nearest-neighbour’ redshift assignment scheme recovers the underlying w_p remarkably well (Zehavi et al. 2005). We thus limit our w_p measurement to those scales above. We do not consider other more sophisticated corrections that would potentially recover the signals below the fibre radius (Guo, Zehavi & Zheng 2012), as r_p^{fb} is always below $0.17 h^{-1} \text{ Mpc}$ across the redshift range of our samples ($z_{\text{max}} = 0.30$). Furthermore, even at the highest redshifts where r_p^{fb} is relatively large, the samples are progressively dominated by more massive galaxies that have a larger correlation length, so the one-halo term is still well resolved in the clustering measurement above r_p^{fb} . For the g - g lensing measurement, this correction using the nearest neighbour slightly blurs the signals around those ‘collided’ lenses in the transverse direction because of the inaccurate conversion from angles to physical separations. This blurring effect can be safely ignored in our analysis as it does not reduce the overall amplitude of the average lensing signal.

In addition to the angular positions and redshifts of the galaxies, the NYU-VAGC also calculated k -corrections for all the MGS galaxies using templates based on stellar population synthesis (SPS) models, while providing approximate stellar mass estimates and star formation histories for all the `bright0` galaxies (Blanton & Berlind 2007). Comparison with the stellar mass estimates from Kauffmann et al. (2003) shows a scatter of only 0.1 dex and mean biases of less than 0.2 dex depending on stellar mass. This tight relationship, as we show later in Section 2.2, provides crucial information for estimating stellar masses for galaxies that do not have valid entries in the MPA/JHU stellar mass catalogue (an updated version of the Kauffmann et al. 2003 catalogue) that we adopted.

To minimize any potential artefacts in the measurements due to irregular survey geometry and areas with low spectroscopic coverage, we use data exclusively within the contiguous area in the North Galactic Cap and from sectors (each defined by a region covered by a unique set of plates) where the angular completeness is greater than 0.8. The final sample used for the galaxy clustering analysis includes 513 150 galaxies over a sky area of 6395.49 deg^2 . A further 5 per cent of the area is eliminated for the lensing analysis, due to the absence of source galaxies in that area (for example, because of poor imaging quality or more conservative masking around bright stars).

2.2 Stellar mass estimate

We employ the stellar mass estimates from the latest MPA/JHU value-added galaxy catalogue.¹ The stellar masses were estimated based on fits to the SDSS photometry following the philosophy

¹ <http://home.strw.leidenuniv.nl/~jarle/SDSS/>

of Kauffmann et al. (2003) and Salim et al. (2007), and assuming the Chabrier (Chabrier 2003) initial mass function (IMF) and the Bruzual & Charlot (2003) SPS model. The estimation is very similar to the spectroscopic fits as described in Kauffmann et al. (2003), but instead of spectral features like the D4000 index and the H δ absorption lines, the broad-band photometry (after correction for emission lines) is used for the fits. The key difference between the photometry-based estimates and the spectroscopic ones is that the stellar mass-to-light ratio (M_*/L) estimated from spectroscopic features is only representative for the region sampled by the fibre, and a constant M_*/L must then be assumed to extrapolate to a total stellar mass. To avoid extrapolation, we employ the total mass estimates obtained through direct fits to the SDSS total photometry, i.e. the `cmodel` magnitudes.

We then match the MPA/JHU stellar mass catalogue to the NYU-VAGC `bright0` sample and identify valid, unambiguous MPA/JHU stellar mass estimates for all but 32 327 (6.3 per cent) of the MGS galaxies. Although it is unclear what causes the failure in estimating stellar mass for these systems, they are statistically indistinguishable from the matched population in redshift, luminosity, flux, and colour. Therefore, given the tight scaling relationship between the NYU-VAGC stellar mass M_*^N and the MPA/JHU mass M_*^M , we can recover the individual photometric stellar mass estimates for these unmatched galaxies in a way that is statistically consistent with the matched ones. We first divide the `bright0` galaxies into red and blue populations based on their $g - r$ colours (> 0.8 or < 0.8 ; k -corrected to $z = 0.1$), because the scaling between the two types of estimates is colour dependent and is much tighter within each single-colour population than the entire sample (Bell & de Jong 2001). Within each colour, we then compute the probability distribution of M_*^M at each fixed M_*^N using the successfully matched galaxies, and assign each unmatched galaxy with M_*^N a random stellar mass M_*^M drawn from that distribution.

Stellar mass estimation is subject to various theoretical uncertainties in predicting the M_*/L , due to the choice of the SPS model, dust extinction law, stellar evolution model, and most importantly, the form of the IMF (see Conroy 2013 for a general review on these topics and L12 for a detailed discussion of the various systematic uncertainties in stellar mass estimation). Assuming no trend in IMF with galaxy type, environment, or redshift,² the range of different IMFs causes uncertainty in the absolute normalization of the M_*/L of factors of two to several depending on the passband (Bell & de Jong 2001), thus shifting the inferred SMF horizontally while maintaining the same shape. The second major source of uncertainty lies in the difficulty in correctly measuring the total fluxes of individual galaxies from aperture photometry. The SDSS pipeline returns a fraction of the total light even when the `cmodel` magnitudes are used (Abazajian et al. 2009), due to a combination of oversubtraction of the sky background and poor model fits to the light profiles (Bernardi et al. 2013). This issue is particularly acute for massive galaxies, where the measured number density of galaxies above $3 \times 10^{11} h^{-2} M_\odot$ may be underestimated by as much as a factor of 5. Unfortunately, there is no well-calibrated correction that can be applied to the MPA/JHU catalogue. Therefore, we use the MPA/JHU catalogue as-is and focus on the particular mapping between the MPA/JHU stellar masses to dark matter haloes. For the purpose of this mapping, both types of systematic errors on the stellar mass are

rather benign, as long as the ranking order of galaxies in their stellar mass estimates is largely unperturbed. Unless otherwise noted, we will refer to the MPA/JHU stellar mass estimate simply as the ‘stellar mass’ and denote it with M_* for the rest of the paper.

2.3 Stellar mass sample selection

The flux-limited nature of surveys like the SDSS makes it very difficult to select volume-limited galaxy samples thresholded or binned in stellar mass, due to the large spread in M_*/L at fixed M_* . However, the M_*/L distribution of the red, quiescent galaxies is much narrower than that of the blue, star-forming galaxies (Gallazzi & Bell 2009), therefore at any given redshift the observed red population has a sharp cutoff at low stellar mass, where the observed galaxies are dominated by the blue population near the flux limit. This phenomenon is best illustrated in Fig. 1(a), where we show the average $g - r$ colour at each (M_*, z) location, colour-coded by the colour bar on top. The map is well separated into two regimes based on the mixing of galaxy colours by the narrow streak of the $\langle g - r \rangle = 0.8$ population, which we hereafter refer to as the ‘mixture’ limit, analogous and related to the ‘flux’ limit on a luminosity versus z diagram.

In Fig. 1(a), above the mixture limit the variation of the average $g - r$ colour with M_* is largely uniform across all redshifts below 0.1, with a gradual and smooth transition at $M_* \sim 10^{10} h^{-2} M_\odot$ as a result of the galaxy evolution physics (a.k.a, downsizing; Cowie et al. 1996). The sharp colour transition across the mixture limit, however, is purely a manifestation of the flux-limited nature of the sample, combined with the fact that quiescent galaxies produce much less light than their star-forming counterparts at the same M_* due to their higher M_*/L ratio. The sharpness of the transition corresponds to the narrow width of the M_*/L distribution of these quiescent galaxies. Therefore, the mixture limit provides a simple empirical guideline for selecting stellar mass samples of relatively high volume-completeness, circumventing the problem of theoretically determining the maximum M_*/L ratio for galaxies at the flux limit of each redshift.

Since our goal is to infer the stellar-to-halo mass mapping for the average galaxy population with a fair mix of both the quiescent and the star-forming galaxies, which occupy dark matter haloes in different ways, we will restrict our analysis to the galaxies above the ‘mixture’ limit (see Section 4.1 for an additional factor for making this choice of restriction). The functional form we adopt to describe the mixture limit $M_*^{\text{mix}}(z)$ is

$$\lg \left(\frac{M_*^{\text{mix}}}{h^{-2} M_\odot} \right) = 5.4 \times (z - 0.025)^{0.33} + 8.0, \quad (1)$$

i.e. the white curves shown in Figs 1(b) and (c), slightly more conservative than the $\langle g - r \rangle = 0.8$ streak in Fig. 1(a). We then define different stellar mass samples above this mixture limit for the clustering and lensing measurements in Figs 1(b) and 1(c), which show the distributions of galaxy number counts and number densities, respectively, on the M_* - z diagram. The thick black boxes in Fig. 1(b) represent a typical sample selection scheme adopted by the traditional HOD analysis. In order to predict the clustering and lensing signals, the HOD model has to assume either volume-completeness of the sample within those boxes, or an ad hoc prescription describing the completeness as function of M_* and/or z (e.g. Miyatake et al. 2015a). Both assumptions are less than ideal due to our ignorance about the M_*/L ratio distribution of galaxies. Furthermore, the rectangular shape of the selections, imposed merely for the sake of modelling convenience, inevitably misses

² For evidence suggesting non-universal IMFs, however, see van Dokkum & Conroy (2010), Conroy & van Dokkum (2012), Cappellari et al. (2012), and Ferreras et al. (2013).

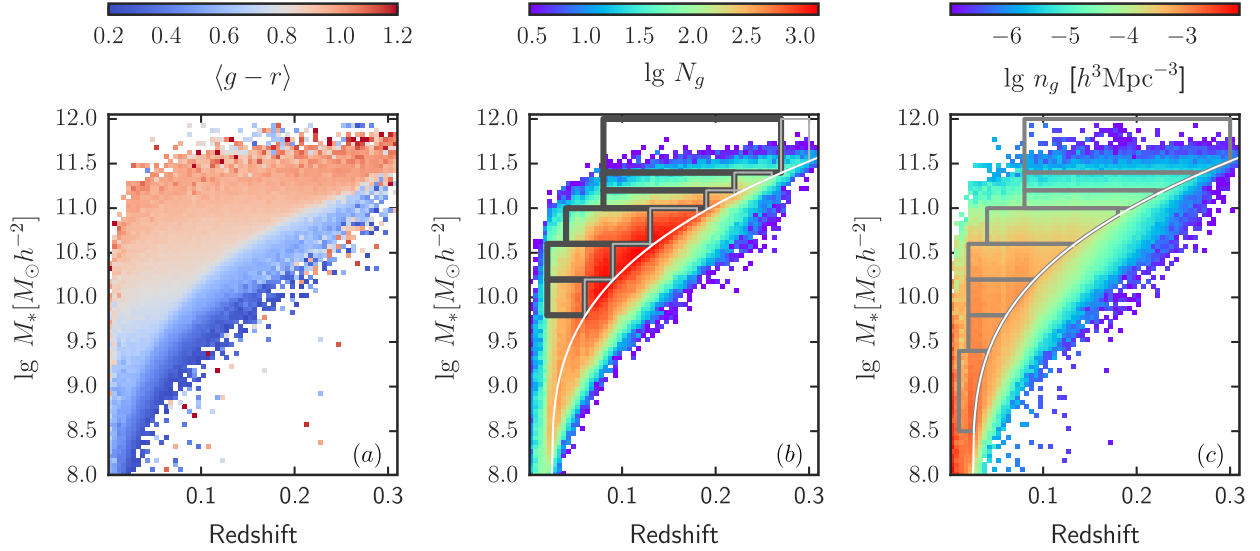


Figure 1. Distributions of the mean $g - r$ rest-frame colour (panel a), the observed number counts (panel b), and the comoving number density (panel c) of SDSS DR7 galaxies on the stellar mass versus redshift plane, with colour bars showing on the top. In the middle panel, the thick ‘boxes’ indicate the galaxy samples selected for the traditional HOD analysis (cHOD), while the thin ‘wedges’ represent the omitted region where galaxies are the most abundant. The grey regions in the right-hand panel are the galaxy samples that were used for our fiducial iHOD analysis. The curved lower boundary of the selection is determined by the mixture limit below which the red galaxy population is severely underrepresented in the spectroscopic survey, and is derived from the distinctive white stripe ($(g - r) = 0.8$) in the left-hand panel (see equation 1).

the regions where the galaxies are the most abundant (thin grey ‘wedges’) on the 2D histogram because they occupy a much larger comoving volume per unit redshift compared to the lower redshift ‘boxes’. In principle, the traditional HOD analysis can include more galaxies inside those ‘wedges’ by adopting finer stellar mass bins with fewer galaxy per sample, but the measurement signal from each individual sample would be very noisy, rendering this scheme highly impractical. In this study we develop two novel improvements over the traditional HOD approach: (1) to statistically account for the sample incompleteness in a self-consistent way, and (2) to be able to predict the clustering and lensing signal for all the galaxies above the mixture limit (thick grey selections in Fig. 1c). We hereafter refer to the traditional HOD approach as the cHOD and our improved version the iHOD, where c and i are loosely tied to ‘completeness’ and ‘incompleteness’, respectively. Table 1 summarizes the basic information of the two sets of sample selections used by the two

modelling methods. In total, we select 314 302 (61 per cent of the bright0 sample) galaxies for the iHOD analysis, and from them 170 483 (54 per cent of the iHOD galaxies) are used for the cHOD analysis.

2.4 Source catalogue

As sources for the g–g lensing measurement, we use a catalogue of background galaxies (Reyes et al. 2012) with a number density of 1.2 arcmin^{-2} with weak lensing shears estimated using the re-Gaussianization method (Hirata & Seljak 2003) and photometric redshifts from Zurich Extragalactic Bayesian Redshift Analyzer (Feldmann et al. 2006). The catalogue was characterized in several papers that describe the data, and use both the data and simulations to estimate systematic errors (see Mandelbaum et al. 2012, 2013; Nakajima et al. 2012; Reyes et al. 2012).

Table 1. The two sets of stellar mass bins used for the iHOD and cHOD analyses, corresponding to the dark and grey thick selections in Figs 1(b) and 1(c), respectively. The iHOD analysis includes two extra stellar mass samples below $\lg M_* = 9.8$, while the higher stellar mass samples have the same binning in $\lg M_*$ and the same minimum redshifts z_{\min} for the two analyses. However, all the iHOD samples have more extended redshift ranges, thresholded by the mixture limit defined in equation (1) on the far side, so the maximum redshift z_{\max} listed for each sample is the maximum redshift of any galaxies in that sample. For the z_{\max} and N_g we list the corresponding numbers for the cHOD analysis in parenthesis. The average large-scale bias for all, central, and satellite galaxies, the average logarithmic halo mass corresponding to the central galaxies, and the satellite fraction of each iHOD sample (with 68 per cent uncertainties) are derived from the fiducial iHOD constraint for the corresponding iHOD samples.

$\lg M_*/h^{-2} M_\odot$	z_{\min}	z_{\max}	N_g	f_{sat}	$\langle b_{\text{all}} \rangle$	$\langle b_{\text{sat}} \rangle$	$\langle b_{\text{cen}} \rangle$	$\langle \lg M_h^{\text{cen}}/h^{-1} M_\odot \rangle$
8.5–9.4	0.01	0.04	13 616	$0.42^{+0.04}_{-0.04}$	$1.12^{+0.02}_{-0.02}$	$1.63^{+0.07}_{-0.07}$	$0.74^{+0.02}_{-0.02}$	$11.16^{+0.16}_{-0.20}$
9.4–9.8	0.02	0.06	16 247	$0.42^{+0.03}_{-0.04}$	$1.15^{+0.01}_{-0.02}$	$1.66^{+0.06}_{-0.06}$	$0.77^{+0.01}_{-0.01}$	$11.44^{+0.09}_{-0.10}$
9.8–10.2	0.02	0.09 (0.06)	46 910 (22 409)	$0.42^{+0.03}_{-0.03}$	$1.19^{+0.01}_{-0.01}$	$1.71^{+0.05}_{-0.05}$	$0.82^{+0.01}_{-0.01}$	$11.74^{+0.05}_{-0.06}$
10.2–10.6	0.02	0.13 (0.09)	96 946 (58 209)	$0.37^{+0.02}_{-0.02}$	$1.26^{+0.01}_{-0.01}$	$1.83^{+0.04}_{-0.04}$	$0.93^{+0.01}_{-0.01}$	$12.15^{+0.03}_{-0.04}$
10.6–11.0	0.04	0.18 (0.13)	102 307 (60 283)	$0.26^{+0.01}_{-0.01}$	$1.40^{+0.01}_{-0.01}$	$2.11^{+0.05}_{-0.04}$	$1.15^{+0.01}_{-0.01}$	$12.68^{+0.02}_{-0.03}$
11.0–11.2	0.08	0.22 (0.19)	24 908 (19 506)	$0.17^{+0.01}_{-0.01}$	$1.73^{+0.01}_{-0.01}$	$2.65^{+0.07}_{-0.06}$	$1.54^{+0.02}_{-0.02}$	$13.21^{+0.02}_{-0.02}$
11.2–11.4	0.08	0.26 (0.22)	10 231 (7427)	$0.11^{+0.02}_{-0.01}$	$2.13^{+0.02}_{-0.02}$	$3.27^{+0.10}_{-0.08}$	$1.99^{+0.02}_{-0.03}$	$13.58^{+0.02}_{-0.02}$
11.4–12.0	0.08	0.30 (0.27)	3137 (2649)	$0.05^{+0.01}_{-0.01}$	$2.84^{+0.06}_{-0.06}$	$4.28^{+0.17}_{-0.13}$	$2.76^{+0.06}_{-0.06}$	$13.96^{+0.03}_{-0.03}$

3 MEASURING GALAXY CLUSTERING AND GALAXY–GALAXY LENSING

In this section, we describe how we measure the projected correlation function from galaxy pair counting and the surface density contrast from g–g lensing. We start by comparing the results of the measurements for the two sets of galaxy samples defined in Table 1, and the reader who is not interested in technical details can skip the remainder of this section.

Fig. 2 compares the clustering and g–g lensing signals measured for the six pairs of iHOD (solid) and cHOD (dashed) samples that share the same stellar mass ranges (marked in the left-hand panel). In each panel, the coloured and the hatched bands illustrate the sizes of the measurement uncertainties (i.e. the diagonal of the error matrices) in the iHOD and the cHOD cases, respectively. At the end of each pair of uncertainty bands, we mark the ratio between the average signal-to-noise ratios (S/N) of the iHOD and the cHOD measurements. The improvement in the S/N of the g–g lensing measurements (right-hand panel) is consistent with the reduction in the Poisson errors by factors of $\sqrt{N_g^i/N_g^c}$, where N_g^i and N_g^c are the numbers of galaxies in the iHOD and the cHOD samples, respectively (the fourth column of Table 1). For the clustering measurements (left-hand panel), the improvement for the three higher mass bins is consistent with the expected decrease in the Poisson error by N_g^i/N_g^c , while for the three lower mass bins the improvement is smaller than expected from assuming the errors are pure Poisson, due to the strong cosmic variance in these samples. These improvements in the S/N of the measurements are crucial to the success of the iHOD model in robustly solving the mapping between galaxies and haloes without resorting to any external information (e.g. the SMF) or strong priors on the model parameters (e.g. the scatter in the SHMR).

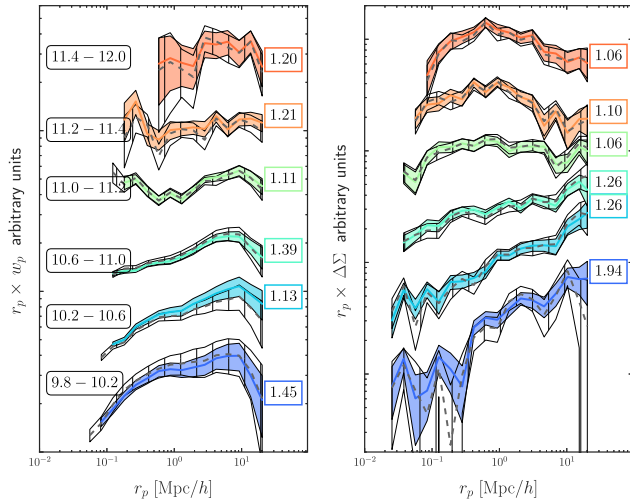


Figure 2. Comparison between the clustering (left) and lensing (right) measurements for the iHOD (coloured solid) and cHOD (grey dashed) samples of the same stellar mass ranges, marked at the beginning of each pair of curves in the left-hand panel. In each panel, the coloured and hatched bands represent the sizes of the uncertainties in the corresponding measurements, and the number marked at the end of each pair of bands is the ratio between the average signal-to-noise ratios of the iHOD and the cHOD measurements, all > 1 because of the extra galaxies included in the iHOD samples. To avoid clutter, the units of the projected correlation function (left) and the g–g lensing signals (right) for each stellar mass bin are scaled arbitrarily.

3.1 Projected galaxy correlation function

We measure the projected correlation function w_p for each galaxy sample by integrating the 2D redshift–space correlation function ξ^s ,

$$w_p(r_p) = \int_{-r_\pi^{\max}}^{+r_\pi^{\max}} \xi^s(r_p, r_\pi) dr_\pi, \quad (2)$$

where r_p and r_π are the projected and the line-of-sight (LOS) co-moving distances between two galaxies. We measure the w_p signal out to a maximum projected distance of $r_p^{\max} = 20 h^{-1}$ Mpc, where the galaxy bias is approximately linear. For the integration limit, we adopt a maximum LOS distance of $r_\pi^{\max} = 60 h^{-1}$ Mpc, where the average pairwise peculiar velocity is very small (Tinker et al. 2008). Based on a mildly modified linear Kaiser formalism, van den Bosch et al. (2013) showed that the w_p measured with $r_\pi^{\max} = 60 h^{-1}$ Mpc should be boosted by a few to 8 per cent depending on r_p , in order to correct for the residual redshift-space distortion (RRSD) effect due to the small but non-zero peculiar velocity beyond $r_\pi > 60 h^{-1}$ Mpc (see their fig. 6). The boost was calibrated for a mock galaxy sample binned in luminosity with different cosmology, therefore it is not suitable to directly apply it to our measurement. However, the impact of the RRSD effect on our result, especially on the SHMR, can be roughly estimated as follows: in the most extreme case that all the constraining power comes from the clustering at large scales, based on the average bias versus halo mass relation for the central galaxies listed in Table 1, we can infer that a 5 per cent boost in w_p translates to 2.5 per cent in galaxy bias, which translates to ~ 0.015 dex in halo mass, shifting the SHMR towards the lower mass end by 0.015 dex. This shift is much smaller than the scatter in the halo mass at fixed stellar mass and our constraint is not solely coming from w_p , therefore the impact of ignoring RRSD should be negligible.

We compute the 2D correlation using the Landy–Szalay estimator (Landy & Szalay 1993),

$$\xi^s(r_p, r_\pi) = \frac{\text{DD} - \text{DR} + \text{RR}}{\text{RR}}, \quad (3)$$

where DD, DR, and RR represent the number counts of pairs of two data galaxies, one data and one random galaxies, and two random galaxies, respectively. The Landy–Szalay estimator has minimal variance and is insensitive to the number of random points (Kerscher, Szapudi & Szalay 2000). For each given galaxy, we find all the neighbouring galaxies within a cylinder of radius r_p^{\max} and height $2r_\pi^{\max}$ centred on that galaxy, using the `LIBKD` package within the `ASTROMETRY.NET` software (Lang et al. 2010). We then count all the pairs in each cell centred on (r_p, r_π) and the three sets of pair counts thus directly give the value of $\xi^s(r_p, r_\pi)$ via equation (3).

The error covariance matrix for each w_p measurement is estimated via the jackknife resampling technique (Norberg et al. 2009). We divide the entire footprint into 200 spatially contiguous, roughly equal-size patches on the sky and compute the w_p for each of the 200 jackknife subsamples by leaving out one patch at a time. For each stellar mass sample, we adopt the sample mean of the 200 subsample measurements as our final estimate of w_p , and the sample covariance matrix as an approximate to the underlying error covariance.

The random galaxy catalogue was constructed following two steps. We first generate a sample of random positions on the sky with 10 times the size of the data catalogue, using the `MANGLE` software (Hamilton & Tegmark 2004; Swanson et al. 2008) and the `bright0` angular selection function provided on the NYU–VAGC website. Secondly, we calculate the 2D joint distribution of

stellar mass and redshift from the data catalogue, and draw 5131 500 random pairs of (M_*, z) values to assign to the stored random positions. In this way, we ensure the random galaxy catalogue has exactly the same angular, radial, and stellar mass joint selection functions as the data galaxies. We apply the same sample selection criterion to the random catalogue and the data sample. As mentioned in Section 2.1, we only use the w_p values down to the physical distance that corresponds to the fibre radius at the maximum redshift of that sample.

The w_p measurements for the eight \mathbf{iHOD} stellar mass samples are shown as the filled circles with error bars in the top sub-panels in Fig. 6. The error bars reflect the diagonal components of the jackknife covariance matrices. Due to the strong cosmic variance effects in the low stellar mass samples ($\lg M_* < 10.6$), overall the off-diagonal components (not shown here) are strong and persist on both small and large scales, while for the high stellar mass ones ($\lg M_* > 11.0$) they are only prominent on scales larger than $5 h^{-1}$ Mpc and between two adjacent distance bins, i.e. along the diagonal blocks. We will refer back to Fig. 6 and discuss in more detail the comparison between the measurements and the predictions from our best-fitting model in Section 6.

3.2 Surface density contrast from galaxy–galaxy lensing

Here we describe how we measure the surface density contrast from g–g lensing. The lensing measurement begins with identification of background source galaxies around each lens (with photometric redshift larger than the lens spectroscopic redshift). Inverse variance weights are assigned to each lens–source pair, including both shape noise and measurement error terms in the variance:

$$w_{ls} = \frac{1}{\Sigma_{\text{crit}}^2(\sigma_e^2 + \sigma_{\text{SN}}^2)}, \quad (4)$$

where σ_e^2 is the shape measurement error due to pixel noise, and σ_{SN}^2 is the rms intrinsic ellipticity (both quantities are per component, rather than total; the latter is fixed to 0.365 following Reyes et al. 2012). Σ_{crit} is the critical surface mass density defined by

$$\Sigma_{\text{crit}}^{-1}(z_l, z_s) \equiv \frac{4\pi G}{c^2} \frac{D_{ls} D_l (1 + z_l)^2}{D_s}, \quad (5)$$

where D_l and D_s are the angular diameter distances to lens and source, and D_{ls} is the distance between them. We use the estimated photometric redshift each source to compute D_s and D_{ls} . The factor of $(1 + z_l)^2$ comes from our use of comoving coordinates.

The projected mass density in each radial bin can be computed via a summation over lens–source pairs ‘ls’ and random lens–source pairs ‘rs’:

$$\Delta \Sigma(r_p) = \frac{\sum_{ls} w_{ls} e_t^{(ls)} \Sigma_{\text{crit}}(z_l, z_s)}{2\mathcal{R} \sum_{rs} w_{rs}}, \quad (6)$$

where e_t is the tangential ellipticity component of the source galaxy with respect to the lens position, the factor of $2\mathcal{R}$ converts our definition of ellipticity to the tangential shear γ_t , and r_p is the comoving projected radius from the lens. The division by $\sum_{rs} w_{rs}$ accounts for the fact that some of our ‘sources’ are physically associated with the lens, and therefore not lensed by it (see; e.g. Sheldon et al. 2004). Finally, we subtract off a similar signal measured around random lenses, to subtract off any coherent systematic shear contributions (Mandelbaum et al. 2005a); this signal is statistically consistent with zero for all scales used in this work.

To calculate the error bars, we also used the jackknife resampling method. As shown in Mandelbaum et al. (2005a), internal estimators

of error bars (in that case, bootstrap rather than jackknife) perform consistently with external estimators of error bars for $\Delta \Sigma$ due to its being dominated by shape noise.

Use of photometric redshifts, which have non-zero bias and significant scatter, gives rise to a bias in the signals that can be easily corrected using the method from Nakajima et al. (2012). This bias is a function of lens redshift, and is properly calculated including all weight factors for each lens sample taking into account its redshift distribution. For typical lens samples in this work, the bias for which we apply a correction is of the order of 1 per cent, far below the statistical errors; the maximum is slightly below 10 per cent.

4 MODEL FOR MAPPING THE STELLAR CONTENT TO HALOES

4.1 A tale of two HODs: \mathbf{cHOD} versus \mathbf{iHOD}

The key to statistically solving the mapping between stellar content and dark matter haloes is $p(M_*, M_h)$, the 2D joint probability density distribution of a galaxy with stellar mass M_* sitting in a halo of mass M_h , normalized so that $\int \int p(M_*, M_h) dM_* dM_h = 1$. For analysing a galaxy sample with stellar mass range $[M_*^0, M_*^1]$ and redshift range $[z_0, z_1]$, it is common practice for HOD models to directly parametrize the occupation number as function of halo mass for the entire sample, $\langle N_g(M_h) \rangle$, which is related to $p(M_*, M_h)$ via

$$\begin{aligned} \langle N_g(M_h) \rangle &= \frac{n_g}{z_1 - z_0} \left(\frac{dn}{dM_h} \right)^{-1} \\ &\times \int_{z_0}^{z_1} \int_{M_*^0}^{M_*^1} p(M_*, M_h) f_{\text{obs}}(M_* | M_h, z) dz dM_*, \quad (7) \end{aligned}$$

where dn/dM_h is the halo mass function, n_g is the total galaxy number density in the Universe (both observed and unobserved), and f_{obs} is the detection rate that varies between 0 and 1. At any given redshift, the detection rate can be explicitly written as

$$f_{\text{obs}}(M_* | M_h, z) = \int_0^{\Gamma_{\text{max}} \equiv M_*/L_{\text{min}}(z)} g(\Gamma | M_*, M_h) d\Gamma, \quad (8)$$

where $g(\Gamma | M_*, M_h)$ is the distribution of the stellar mass-to-light ratio $\Gamma \equiv M_*/L$ of galaxies at fixed M_* within haloes of mass M_h , and $L_{\text{min}}(z)$ is the luminosity threshold corresponding to the flux limit at z . At low redshift, $L_{\text{min}}(z)$ is low (large Γ_{max}), so the integral extends over a wide range of Γ values and gives a result that approaches 1, while at higher redshift, $L_{\text{min}}(z)$ is high (small Γ_{max}), limiting the range of accessible Γ values severely and thus lowering f_{obs} .

In order to describe the sample using a single HOD, f_{obs} must be uniform across the redshift range, i.e. $f_{\text{obs}}(M_* | M_h, z) \equiv f_{\text{obs}}(M_* | M_h)$, which happens if and only if $f_{\text{obs}}(M_* | M_h) \equiv 1$, so that

$$\langle N_g(M_h) \rangle = n_g \left(\frac{dn}{dM_h} \right)^{-1} \int_{M_*^0}^{M_*^1} p(M_*, M_h) dM_*. \quad (9)$$

Following the standard procedure in HOD modelling, we adopt equation (9) for the \mathbf{cHOD} model, writing separate contributions to $p(M_*, M_h)$ for central and satellite galaxies, and derive constraint by applying it to the samples defined in Fig. 1(b). For the sake of comparison to the \mathbf{iHOD} constraint, we do not use the number density of sample galaxies as an input to the \mathbf{cHOD} constraints.

However, due to the increase of the luminosity threshold with redshift for a flux-limited sample, the M_* -selected galaxy populations are strictly stratified in z and the HODs at different redshifts

must be treated separately. At fixed redshift z , the HOD is

$$\langle N_g(M_h|z) \rangle = n_g \left(\frac{dn}{dM_h} \right)^{-1} \int_{M_*^0}^{M_*^1} p(M_*, M_h) f_{\text{obs}}(M_*|M_h, z) dM_*. \quad (10)$$

Therefore, in order to predict $\langle N_g(M_h|z) \rangle$ from $p(M_*, M_h)$ we need to know $f_{\text{obs}}(M_*|M_h, z)$, which is inaccessible to us because of our ignorance of $g(\Gamma|M_*, M_h)$. However, we can rewrite the above equation as

$$\langle N_g(M_h|z) \rangle = \left(\frac{dn}{dM_h} \right)^{-1} \int_{M_*^0}^{M_*^1} p(M_h|M_*) \times [\Phi(M_*) f_{\text{obs}}(M_*|M_h, z)] dM_*, \quad (11)$$

by using Bayes' Theorem $p(M_h|M_*) = p(M_*, M_h)/p(M_*)$ and the definition of the *parent* SMF $\Phi(M_*) \equiv n_g p(M_*)$. To make further progress in our predictions, we adopt the ansatz that above the mixture limit $M_* > M_*^{\text{mix}}(z)$, the dependence of $f_{\text{obs}}(M_*|M_h, z)$ on the halo mass is very weak, i.e. $f_{\text{obs}}(M_*|M_h, z) \sim f_{\text{obs}}(M_*|z)$. Applying this ansatz to equation (11) we arrive at

$$\langle N_g(M_h|z) \rangle = \left(\frac{dn}{dM_h} \right)^{-1} \int_{M_*^0}^{M_*^1} p(M_h|M_*) \Phi_{\text{obs}}(M_*|z) dM_*, \quad (12)$$

where $\Phi_{\text{obs}}(M_*|z) = \Phi(M_*) f_{\text{obs}}(M_*|z)$ is the *observed* SMF at redshift z , directly accessible from the survey. For modelling the samples defined in Fig. 1(c) for the iHOD analysis, we measure the observed galaxy SMF at each redshift, and then obtain the HOD for that redshift slice using equation (12). In this way, we avoid the need to explicitly model the incompleteness as a function of M_* and/or M_h . Although it appears from equation (12) that both the amplitude and the shape of Φ_{obs} are used to derive $\langle N_g(M_h|z) \rangle$, in essence we only use the shape information, because the normalization of $\langle N_g(M_h|z) \rangle$ is irrelevant to the prediction of the clustering and lensing signals. Furthermore, the shape information is used solely for properly weighting individual galaxies during the signal calculation, but not as part of the data that we fit to. For instance, if we randomly throw away half of the galaxies between some arbitrary stellar mass range, so that the observed SMF does not have the right shape, the best-fitting parameters would still stay the same, albeit with larger uncertainties.

Before going any further, here we will lay out the theoretical arguments leading to the ansatz and defer the detailed discussion on its validation using consistency checks later in Sections 6.2 and 7.1. Splitting the galaxies into red and blue populations explicitly in equation (8), we have

$$f_{\text{obs}}(M_*|M_h, z) = \int_0^{\Gamma_{\text{max}}(z)} \{ f_{\text{red}}(M_*|M_h) g_{\text{red}}(\Gamma|M_*, M_h) + [1 - f_{\text{red}}(M_*|M_h)] g_{\text{blue}}(\Gamma|M_*, M_h) \} d\Gamma, \quad (13)$$

where $f_{\text{red}}(M_*|M_h)$ is the intrinsic fraction of red galaxies at given M_* within haloes of mass M_h . To understand the potential dependence of $f_{\text{obs}}(M_*|M_h, z)$ on the halo mass M_h , we first examine the variations of $f_{\text{red}}(M_*|M_h)$ and $g_{\text{red/blue}}(\Gamma|M_*, M_h)$ with M_h separately and then combine them using the above equation. Since the M_*/L ratio Γ is very tightly correlated with galaxy colour c (Bell & de Jong 2001), we can instead look at the colour distributions of the two populations, $g_{\text{red/blue}}(c|M_*, M_h)$, each with a centroid position \bar{c} and a spread Δc . By analysing the variation in the galaxy colour bimodality with stellar mass and projected neighbour density using SDSS, Baldry et al. (2006) found that \bar{c} and Δc of the red and blue populations are stable across different environments of void,

field, and groups and clusters (i.e. halo masses), while the red fraction f_{red} increases continuously with M_* and the local density of the environment, i.e. the so-called mass and environmental quenching, respectively (Peng et al. 2012). This stability in the colour distribution within each coloured population against the environment implies $g_{\text{blue}}(\Gamma|M_*, M_h) \equiv g_{\text{blue}}(\Gamma|M_*)$ and $g_{\text{red}}(\Gamma|M_*, M_h) \equiv g_{\text{red}}(\Gamma|M_*)$, so that equation (13) can be simplified as

$$f_{\text{obs}}(M_*|M_h, z) = f_{\text{red}}(M_*|M_h) G_{\text{red}}(M_*|z) + [1 - f_{\text{red}}(M_*|M_h)] G_{\text{blue}}(M_*|z), \quad (14)$$

where

$$G_{\text{red}}(M_*|z) = \int_0^{\Gamma_{\text{max}}(z)} g_{\text{red}}(\Gamma|M_*) d\Gamma$$

$$G_{\text{blue}}(M_*|z) = \int_0^{\Gamma_{\text{max}}(z)} g_{\text{blue}}(\Gamma|M_*) d\Gamma. \quad (15)$$

are the fractions (from 0 to 1) of the red/blue galaxies with M_* at z that will be observed given the flux limit and the separate red/blue distributions of M_*/L ratios. Above the mixture limit, the completeness of both the red and blue galaxies are relatively high, so that $G_{\text{red}}(M_*|z) \approx G_{\text{blue}}(M_*|z) = f_{\text{obs}}(M_*|z)$ and equation (14) gives $f_{\text{obs}}(M_*|M_h, z) = f_{\text{obs}}(M_*|z)$. Below the mixture limit, however, the observed red galaxies are so scarce that $G_{\text{red}}(M_*|z) \sim 0$, yielding a halo mass-dependent $f_{\text{obs}}(M_*|M_h, z) \simeq [1 - f_{\text{red}}(M_*|M_h)] G_{\text{blue}}(M_*|z)$ because $f_{\text{red}}(M_*|M_h)$ is sensitive to M_h (George et al. 2011).

We expect the ansatz to be largely valid except for the low-mass samples ($M_* < 10^{10} h^{-2} M_\odot$) where the colour bimodality might shift with halo mass because of galaxy evolution (Taylor et al. 2015). We none the less proceed to apply this ansatz to the entire stellar mass range, relying on the fact that the statistical power of the low-mass samples is quite low, and discuss the possible impact on our results later in Section 7.1. Finally, we emphasize that this ansatz assuming $f_{\text{obs}}(M_*|M_h, z) \sim f_{\text{obs}}(M_*|z)$, explicitly made in the iHOD model to account for the fact that for any given M_* we observed fewer intrinsically high- M_*/L systems at higher redshifts, is a much weaker assumption than required by the traditional HOD models, which assume all the high- M_*/L systems were detected in the sample, i.e. $f_{\text{obs}}(M_*|M_h, z) = 1$. In other words, since $f_{\text{obs}}(M_*|M_h, z) = 1$ is a sufficient condition for $f_{\text{obs}}(M_*|M_h, z) = f_{\text{obs}}(M_*|z)$, the iHOD model includes the cHOD model as a subset.

It is important to point out that during the constraint, the iHOD model remains entirely agnostic of the overall amplitude of the SMF, whether it be the parent SMF (normalized by n_g) or the observed ones (normalized by the product of n_g and $f_{\text{obs}}(M_*|z)$), because neither n_g nor $f_{\text{obs}}(M_*|z)$ is known a priori. However, the iHOD model is built on top of the halo mass function, which has a fixed normalization in any given Λ CDM cosmology. Thus, once constrained by the clustering and lensing data, the best-fitting iHOD model would give an explicit prediction for both the shape and the amplitude of the parent SMF. This prediction automatically gives an estimate of $f_{\text{obs}}(M_*|z)$ when compared to the observed SMF at redshift z . The estimated $f_{\text{obs}}(M_*|z)$ thus provides a useful consistency check of the iHOD model. Any significant departure of f_{obs} from unity at stellar mass above the mixture limit would indicate failure of our model assumptions (e.g. the ansatz about the weak dependence of $f_{\text{obs}}(M_*|z)$ on M_h) and/or degeneracies in the model parameters (e.g. residual degeneracy between scatter and amplitude of the SHMR), while a successful model, as we will discover in Section 7.1, would reproduce f_{obs} curves that are slightly below unity. Another consistency check is the comparison between the SHMR constraints from the cHOD and the iHOD models on the

high-mass end – in the regime of $M_* \gg M_*^{\text{mix}}(z)$ where $f_{\text{obs}}(M_*|M_h, z) \rightarrow 1$, the iHOD model becomes identical to the cHOD model and the two set of constraints, as will be shown later in Section 6.2, should be consistent with each other.

In summary, both the cHOD and iHOD models rely on the prediction of $p(M_*, M_h)$. The primary difference is that the cHOD model employs a single HOD by assuming that the galaxy sample is volume-complete in stellar mass, while the iHOD model is able to self-consistently take into account the redshift-dependent selection function of the samples by working in narrow redshift slices and using information from the shape of the observed galaxy SMF at each slice. This improvement in iHOD enables us to include 84 per cent more galaxies than used in cHOD by adding the low-mass galaxies ($M_* < 10^{9.8} h^{-2} M_\odot$) in the local universe as well as more distant galaxies, hence the significant improvement in the S/N of the measurements shown in Fig. 2. This process cannot be extended indefinitely, since at lower stellar mass (below the mixture limit for a given redshift) the basic assumption behind iHOD fails, but the range of stellar masses where it is applicable is still large enough for significant improvements.

4.2 Deriving the two HODs

Here we describe the theoretical framework for predicting the HODs for both the cHOD and iHOD models. In particular, we first model the total number of galaxies (both observed and unobserved) per log(stellar mass) within haloes of some fixed mass, $dN(M_*|M_h)/d \lg M_*$, and then compute the joint probability as

$$p(M_*, M_h) = \frac{\lg e}{M_* n_g} \frac{dN(M_*|M_h)}{d \lg M_*} \frac{dn}{dM_h}. \quad (16)$$

To conform to the traditional HOD notation, we hereafter refer to $dN(M_*|M_h)/d \lg M_*$ simply as $\langle N(M_*|M_h) \rangle$ by implicitly assuming $\Delta \lg M_* \equiv 0.02$ throughout the paper. Given $p(M_*, M_h)$, we can use equation (9) to specify a single HOD for each of the cHOD samples. For the iHOD analysis, however, we also need to compute the *parent* SMF

$$\Phi(M_*) = n_g \int_0^{+\infty} p(M_*, M_h) dM_h, \quad (17)$$

and extract the *observed* SMF $\Phi_{\text{obs}}(M_*)$ directly from the data. The distribution of host halo mass for galaxies at fixed stellar mass is simply

$$p(M_h|M_*) = \frac{p(M_*, M_h)}{p(M_*)} = \frac{n_g p(M_*, M_h)}{\Phi(M_*)}, \quad (18)$$

and we can obtain the HODs for individual redshift slices within each iHOD sample using equation (12). Since the redshift range (0.02–0.30) spans ~ 3 billion years during which only 8 per cent of the total stellar mass observed today formed (cf. equation 15 in Madau & Dickinson 2014), we assume $p(M_*|M_h)$ to be constant with redshift. Furthermore, the constraints on different portions of the SHMR usually come from different redshift ranges, i.e. the low and high M_* portions are from the low and high redshifts, respectively, so it is safe here to ignore any potential redshift evolution. Therefore, all the redshift evolution in the theoretical model comes from the cosmic growth in the halo mass function. To speed up the calculation without loss of accuracy, we adopt the same halo mass function for all the redshift slices within each sample, calculated at the volume-averaged redshift of that sample. We have tested this approximation by comparing the predicted signals with those computed from integrating the halo mass functions over all the redshift slices and the difference is negligible.

4.3 Parametrizing $\langle N(M_*|M_h) \rangle$

Our analytic model for $\langle N(M_*|M_h) \rangle$ has two components: (1) the mean and the scatter of the SHMR for the central galaxies, the combination of which automatically specifies $\langle N_{\text{cen}}(M_*|M_h) \rangle$, and (2) the mean number of satellite galaxies with stellar mass M_* inside haloes of mass M_h , $\langle N_{\text{sat}}(M_*|M_h) \rangle$. We adopt a similar parametrization for the two components as in L11, but allow more of them to vary during the constraint, and predict the signals differently. Here we will briefly describe the functional forms and the model parametrizations.

At fixed halo mass, we assume a lognormal probability distribution for the stellar mass of the central galaxies, with a lognormal scatter. The mean SHMR is then the sliding mean of the lognormal distribution as a function of the halo mass, $f_{\text{SHMR}} \equiv \exp \langle \ln M_*(M_h) \rangle$. The L11 functional form for f_{SHMR} that we adopt for our analysis is defined by Behroozi et al. (2010) via its inverse function,

$$M_h = M_1 m^\beta \exp \left(\frac{m^\delta}{1 + m^{-\gamma}} - \frac{1}{2} \right), \quad (19)$$

where $m \equiv M_*/M_{*,0}$. Among the five parameters that describe f_{SHMR} , M_1 and $M_{*,0}$ are the characteristic halo mass and stellar mass that separate the behaviours in the low- and high-mass ends ($f_{\text{SHMR}}(M_1) = \ln M_{*,0}$). The inverse function starts with a low-mass end slope β , crosses a transitional regime around $(M_{*,0}, M_1)$ dictated by γ , and reaches a high-mass end slope $\beta + \delta$. The fig. 1 in L11 illustrates the different responses of f_{SHMR} to the changes in each of the five parameters.

The lognormal scatter at fixed halo mass is the quadratic sum of the intrinsic scatter and the measurement error. It is important to keep in mind that the intrinsic part of this scatter must be the same for all studies, while studies with different data sets or stellar mass determination methods may have differing measurement error contributions and thus a different total lognormal scatter. L12 considered two models for the scatter, one that is constant and another that includes an empirical stellar mass-dependence in the measurement error across the entire mass range. They found that since the constraint on the overall scatter is primarily driven by the high-mass end where the slope of f_{SHMR} is much shallower, the mass-dependence of the scatter at the low-mass end has little impact on their results. In light of this finding and to focus on the scatter at the high-mass end, we keep the scatter independent of halo mass below M_1 , but allow more freedom in the scatter above the characteristic mass scale, with an extra component that is linear in $\lg M_h$:

$$\sigma_{\ln M_*}(M_h) = \begin{cases} \sigma_{\ln M_*}, & M_h < M_1 \\ \sigma_{\ln M_*} + \eta \lg \frac{M_h}{M_1}, & M_h \geq M_1 \end{cases}. \quad (20)$$

The motivation behind this additional degree of freedom is two-fold: (1) the average measurement uncertainty of the stellar mass estimates decrease with M_* (hence M_h) in the MPA/JHU catalogue, and (2) recently, there is evidence suggesting a smaller scatter at the high-mass end of the SHMR (e.g. Shankar et al. 2014), although in equation (20) the scatter is allowed to either increase or decrease with halo mass. The combination of the mean and the scatter fully specifies the HOD of central galaxies,

$$\langle N_{\text{cen}}(M_*|M_h) \rangle = \frac{1}{\sigma_{\ln M_*}(M_h) \sqrt{2\pi}} \exp \left[-\frac{[\ln M_* - \ln f_{\text{SHMR}}(M_h)]^2}{2\sigma_{\ln M_*}^2(M_h)} \right]. \quad (21)$$

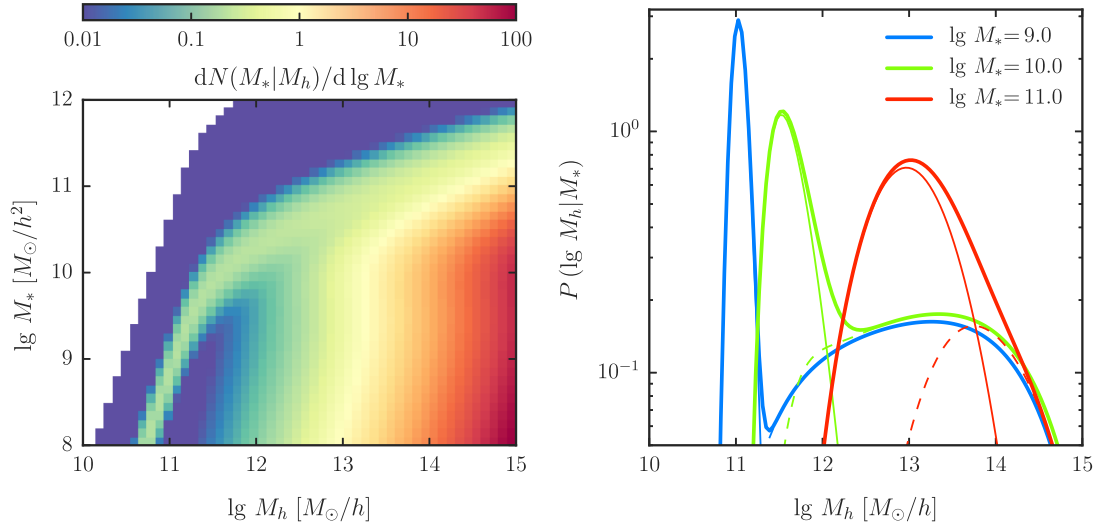


Figure 3. Left-hand panel: the 2D HOD, i.e. the average number of galaxies per dex in stellar mass within haloes at fixed mass, predicted by the best-fitting \mathbf{iHOD} model. Right-hand panel: the probability distributions of the host halo mass for galaxies at three fixed log (stellar mass) listed on the top right. Thick solid, thin solid, and thin dashed lines represent contributions from all, central, and satellite galaxies, respectively. Galaxies of higher stellar mass are much more likely to be central galaxies sitting in high mass haloes, whereas low-mass galaxies are almost as likely to be satellites of high-mass haloes as centrals of their own haloes.

Follow [L11](#), we model $\langle N_{\text{sat}}(M_*|M_h) \rangle$ as the derivative of the satellite occupation number in stellar mass-thresholded samples, $\langle N_{\text{sat}}(> M_*|M_h) \rangle$, which is parametrized as a power of halo mass with an exponential cutoff and scaled to $\langle N_{\text{cen}}(> M_*|M_h) \rangle$ as follows,

$$\langle N_{\text{sat}}(> M_*|M_h) \rangle = \langle N_{\text{cen}}(> M_*|M_h) \rangle \left(\frac{M_h}{M_{\text{sat}}} \right)^{\alpha_{\text{sat}}} \exp \left(-\frac{M_{\text{cut}}}{M_h} \right). \quad (22)$$

Instead of fixing α_{sat} to be 1 as in [L11](#), we allow it to vary during the fit. We parametrize both the characteristic mass of a single satellite-hosting halo, M_{sat} , and the cutoff mass scale, M_{cut} , as simple power-law functions of the threshold stellar mass, so that

$$\frac{M_{\text{sat}}}{10^{12} h^{-2} M_{\odot}} = B_{\text{sat}} \left(\frac{f_{\text{SHMR}}^{-1}(M_*)}{10^{12} h^{-2} M_{\odot}} \right)^{\beta_{\text{sat}}}, \quad (23)$$

and

$$\frac{M_{\text{cut}}}{10^{12} h^{-2} M_{\odot}} = B_{\text{cut}} \left(\frac{f_{\text{SHMR}}^{-1}(M_*)}{10^{12} h^{-2} M_{\odot}} \right)^{\beta_{\text{cut}}}, \quad (24)$$

respectively. In practice, we choose a 0.02 dex bin size in stellar mass for the numerical differentiation of $\langle N_{\text{sat}}(> M_*|M_h) \rangle$.

The left-hand panel of Fig. 3 displays the $\langle N(M_*|M_h) \rangle$ map predicted by the best-fitting \mathbf{iHOD} model in our analysis. The SHMR of the central galaxies is clearly seen as the ‘main sequence’ enveloped by the ‘cloud’ occupied by the satellite galaxies. The right-hand panel shows the probability distribution of host halo masses for galaxies at $M_* = 10^{10} h^{-2} M_{\odot}$ (blue), $10^{11} h^{-2} M_{\odot}$ (green), and $10^{12} h^{-2} M_{\odot}$ (red), each computed from the map in the left-hand panel using the combination of equations (16)–(18). The thin solid and the dashed curves of each colour show the contributions to the total probability distribution from the central and the satellite galaxies, respectively. The central galaxy contribution shows an increasing logarithmic scatter in halo mass with stellar mass, mainly due to the change of slope in the SHMR when going to the high-mass end (where $\sigma_{\ln M_*}$ shrinks slightly as well, i.e. $\eta <$

0). The lower mass satellites, however, populate haloes of much greater diversity than their high-mass counterparts. Meanwhile, the satellite fraction increases towards lower stellar masses, but the total number of galaxies is dominated by central galaxies at all stellar masses.

4.4 Spatial distribution of galaxies within haloes

In addition to the parametrization of $\langle N(M_*|M_h) \rangle$, we also need to model the spatial distribution of galaxies within haloes. We assume the isotropic Navarro–Frenk–White (NFW; Navarro, Frenk & White 1997) density profile for haloes with a concentration–mass relation $c_{\text{dm}}(M_h)$ calibrated from simulations (described further below). Although the NFW density profile is regarded as ‘universal’ only in pure dark matter simulations where the effects of baryons are absent,³ recent studies of nearby rich clusters found that, despite the fact that the baryons tend to flatten the dark matter distribution in the cluster centre, the total matter density (i.e. baryon and dark matter combined) still maintains an NFW shape from the scales of the central brightest cluster galaxy (BCG) out to the virial radius (Newman et al. 2013), probably due to the significant mixing between stars and dark matter as a result of frequent mergers (Laporte & White 2015). For smaller systems like the group and galaxy-scale haloes with $M_h < 10^{14} h^{-1} M_{\odot}$, however, the sum of an NFW and a stellar mass component is required to explain the inner slope of the observed total matter density profiles (Mandelbaum et al. 2006b; Newman, Ellis & Treu 2015). In light of these observational findings, we add a stellar mass component as a point source in the halo centre to the g–g lensing predictions for all the galaxy samples, so that

$$\Delta \Sigma_{\text{stellar}} = \frac{\langle M_* \rangle}{\pi r_p^2}, \quad (25)$$

³ However, see Gao et al. (2008) and Dutton & Macciò (2014) regarding potentially better ‘universality’ when using the Einasto (1965) profile.

where $\langle M_* \rangle$ is the average stellar mass of each stellar mass sample. Although this extra stellar component is not necessary for the central galaxies of very rich clusters, i.e. some galaxies within the highest stellar mass sample, the minimum fitting scale for that sample is above $0.1 h^{-1}$ Mpc (see Appendix A), where the stellar contribution to $\Delta\Sigma$ calculated via equation (25) is negligible.

Central galaxies are placed at the centres of the NFW haloes. For the modelling of w_p , we do not consider the miscentering effect, which is likely to be important only for around 30 per cent of the BCGs (George et al. 2012). In our analysis the galaxy clustering for BCGs is measured down to $\sim 0.2 h^{-1}$ Mpc, therefore largely immune to the miscentering effect which has a kernel $\simeq 75 h^{-1}$ kpc (George et al. 2012). It is worth noting that the miscentering issue in our case is more benign than in other recent papers that have explicitly modelled it (e.g. Miyatake et al. 2015a; More et al. 2015). In those papers, the samples that were being modelled had strict colour and luminosity selection, such that the central galaxy in the halo might not be present in the sample. In our case, with a flux-limited sample that goes to a relatively low flux limit, the central galaxies in the vast majority of group- and cluster-size haloes should be present in the sample, so we only have to contend with small offsets of the central galaxy from the halo centre (rather than a complete misidentification of the central galaxy). For the g-g lensing, the miscentering effect is compounded and somewhat cancelled by the contribution from the subhalo of satellite galaxies (Yoo et al. 2006), which is more important for low stellar mass samples. We will discuss the modelling of the combined effects further in Section 5.3.

We assume an NFW profile for the satellite distribution as well, but with a different amplitude of the concentration–mass relation than the dark matter. In particular, we set $c_{\text{sat}}(M_h) \equiv f_c \times c_{\text{dm}}(M_h)$, where f_c characterizes the spatial distribution of satellite galaxies relative to the dark matter within haloes.

To summarize, we have in total 13 model parameters. Among them $\{\lg M_h^1, \lg M_*^0, \beta, \delta, \gamma\}$ describe the mean SHMR, $\{B_{\text{sat}}, \beta_{\text{sat}}, B_{\text{cut}}, \beta_{\text{cut}}, \alpha_{\text{sat}}\}$ describe the parent HOD of satellite galaxies, $\{\sigma_{\ln M_*}, \eta\}$ describe the logarithmic scatter about the mean SHMR, and f_c is the ratio between the concentrations of the satellite distribution and the dark matter profile.

5 PREDICTING SIGNALS FROM THE HALO MODEL

5.1 Prerequisites and approximations

The goal of our analysis is to infer the SHMR of central galaxies and the HOD of satellite galaxies using the galaxy clustering and the g-g lensing. From the inferred $\langle N(M_*|M_h) \rangle$, we can predict the parent SMF to compare to those empirically reconstructed from the V_{max} method. This route is exactly the reverse of the methodology employed in the SHAM studies (e.g. M10), which infer the SHMR by abundance matching to the V_{max} -estimated SMF and then predict the clustering and/or lensing as a cross-check.

To facilitate a direct comparison with the SHAM results, we adopt the same flat Λ CDM cosmology model as in M10 (listed in Section 1). For the linear matter power spectrum, we use the low baryon transfer function of Eisenstein & Hu (1998), which is a good approximation to the full transfer function on scales well below the BAO scale. To compute the non-linear matter correlation function ξ_{mm} , we use the prescription from Takahashi et al. (2012) (an updated version of the `halofit` prescription from Smith et al. 2003)

to generate the non-linear power spectrum for Fourier transforming to ξ_{mm} . The halo mass function and the halo bias function are from Tinker et al. (2008) and Tinker et al. (2010), respectively. The halo mass–concentration relationship $c_{\text{dm}}(M_h)$ is from the fitting formula of Zhao et al. (2009), which accurately recovers the flattening of halo concentration at high masses. (In DR7 we do not expect an upturn, which only shows up at redshifts beyond 1; see Prada et al. 2012).

For the `cHOD` samples the signal prediction is relatively straightforward, as only one HOD is required to describe each sample. We describe the prediction of w_p and $\Delta\Sigma$ from a single HOD in Section 5.2. Complexity arises when predicting the signals for the `iHOD` samples, each of which contains multiple HODs that describe the stratification of galaxy populations due to the redshift-dependent selection function. We divide each `iHOD` sample into multiple redshift slices of $\Delta z = 0.01$, which corresponds to a comoving width of $\sim 30 h^{-1}$ Mpc across the redshift range of the MGS. This redshift bin size is chosen to make sure the slices are: (1) thin enough for the redshift selection function to be uniform, and (2) thick enough to include both the one and two-halo terms in the LOS direction internally. Within each redshift slice i we predict the clustering and lensing signals, w_p^i and $\Delta\Sigma^i$, in the same way as in the `cHOD` analysis, and combine them to obtain the final predictions for the full sample,

$$w_p = \left(\sum_i N_i^2 w_p^i \right) / \left(\sum_i N_i^2 \right), \quad (26)$$

and

$$\Delta\Sigma = \left(\sum_i N_i w_i \Delta\Sigma^i \right) / \left(\sum_i N_i w_i \right), \quad (27)$$

where N_i is the number of *observed* galaxies in slice i and w_i is the per-galaxy lensing weight at each redshift. w_i is essentially related to the inverse variance weights w_{1s} defined in equation (4), but integrated over the source redshift distribution. It also includes the geometric factor $(1/\Sigma_c^2)$ as well as the fact that an annular bin with a fixed centre value of r_p and width Δr_p contains more source galaxies at lower redshift, due to its larger area on the sky. Since the source catalogue is the same for the g-g lensing in each redshift slice, the pair counting weight for $\Delta\Sigma^i$ is $\propto N_i$ instead of $\propto N_i^2$.

Fig. 4 illustrates the fractional contributions of each individual redshift slice to the predicted total signal (calculated from the two equations above; top sub-panels), and compares the fractional variations among the predicted signals of all redshift slices (bottom sub-panels), using the `iHOD` sample with $\lg M_* = [10.2, 10.6]$ as an example. The dashed curve in each top sub-panel indicates the maximum redshift (0.09) of the corresponding `cHOD` sample with the same stellar mass range. Clearly, the three extra redshift slices above $z = 0.09$ included by the `iHOD` analysis contribute ~ 40 per cent of the total signal from the 11 slices, because of the much higher weights associated with those higher redshift slices.

Since we have used an LOS integration limit of $60 h^{-1}$ Mpc (roughly twice the size of the slice width) to get w_p from ξ , the measured w_p signal will include galaxy pairs that straddle different slices. This effect is not directly reflected in equation (26), so the cross-correlation between different slices might require a separate treatment. However, because the galaxy correlation functions vary very slowly and smoothly between adjacent slices (a few per cent on large scales; see the bottom sub-panels of Fig. 4), the cross terms are most sensitive to the product of the numbers of galaxies in each pair of slices, which is correctly accounted for in equation (26).

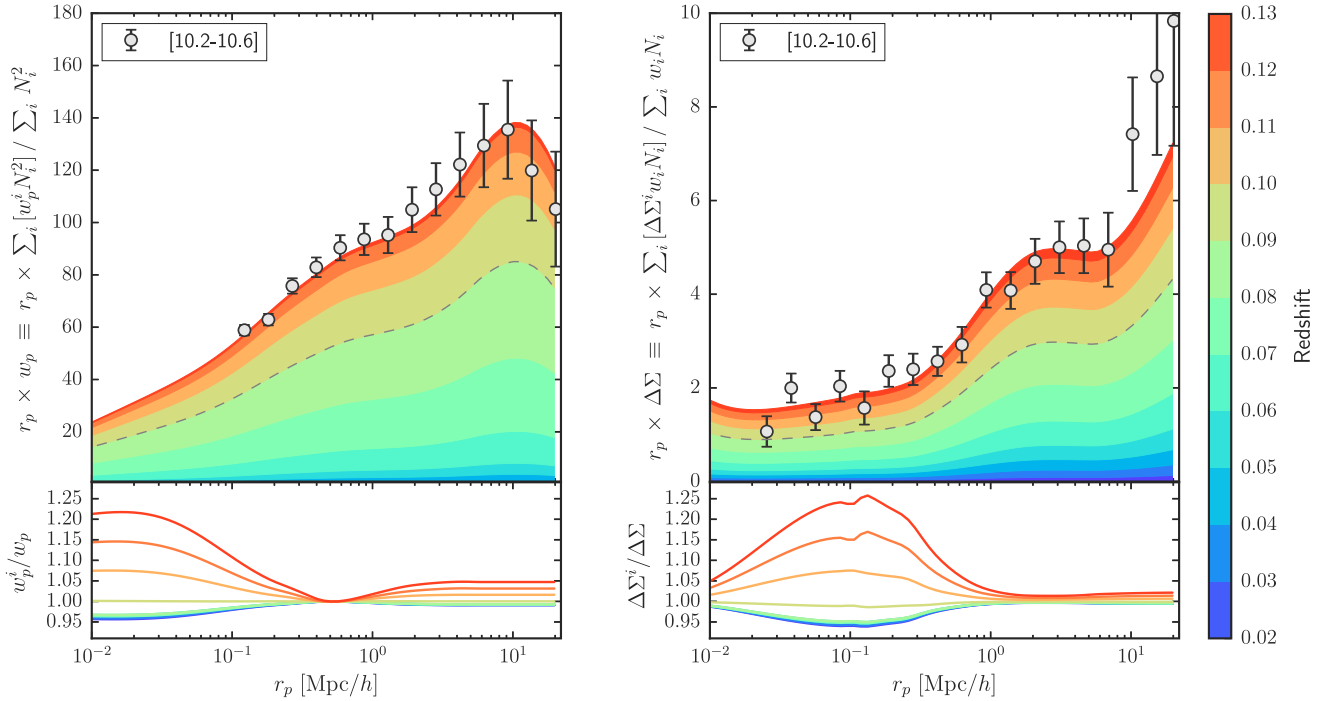


Figure 4. Contributions from individual redshift slices to the predicted clustering (left) and lensing (right) signals of the iHOD sample with $\lg M_* = [10.2, 10.6]$. In each panel, the top sub-panel compares the measured signal (circles with error bars) to the predicted total signal, decomposed into the contributions from each of the 11 redshift slices (colour-coded by the colour bar on the right). The dashed curve in each stack marks the maximum redshift (0.09) of the cHOD sample with the same stellar mass range, indicating a ~ 40 per cent gain in the total signal by switching from cHOD to iHOD. Due to the difference in the weighting schemes, the fractional contributions of the same redshift slice to the total signals are different in the two cases. The bottom sub-panels show the ratio of the signal predicted for each slice over the signal for the whole sample. In both the clustering and lensing cases, the fractional variation is coherent with redshift and below 25 per cent on all scales.

In particular, for two slices i and j , equation (26) makes the assumption that the 3D cross-correlation ξ_{ij} is approximately $(N_i^2 \xi_i + N_j^2 \xi_j) / (2N_i N_j)$, while the more correct form should be $\sqrt{\xi_i \xi_j}$, as the cross-correlation is close to the geometric mean of the two auto-correlation functions ξ_i and ξ_j (Zehavi et al. 2011). When $\Delta \xi / \xi$ is small, the fractional difference between the two forms is of the order $\sim (N_j - N_i)^2 / (N_i N_j) \equiv (\Delta N / N)^2$. Moreover, after integrating ξ via equation (2), the w_p signal is always dominated by the ξ^s values at $r_\pi < 30 h^{-1}$ Mpc at any fixed r_p . We also construct a mock galaxy sample with a similar redshift distribution and bias evolution as the stellar mass samples in the data, and compare the projected correlation functions measured from our approximation in equation (26) to that directly from the pair counting described in Section 3.1. We find that the error induced by adopting equation (26) is no more than 2 per cent on all scales and is thus negligible in our analysis.

It is important to note that the theoretical predictions for w_p and $\Delta \Sigma$ via equation (26) and (27) are independent of the total number of observed galaxies in each sample, i.e. the normalization of the observed SMF – if we analyse only half of the galaxies randomly drawn from the entire bright0 sample, the prediction from the same set of model parameters would not change, and the constraints on the model parameters would stay the same, albeit with larger uncertainties.

5.2 Predicting w_p and $\Delta \Sigma$ from a single HOD

The analytic model for deriving the galaxy clustering and the g–g lensing signals from a single HOD is based on the pre-

scriptions given in a series of papers, including Tinker et al. (2005), Zheng & Weinberg (2007), and Yoo et al. (2006), with improved treatment of the scale-dependent bias, halo exclusion effect, over/underconcentrated satellite distributions from earlier works (e.g. Berlind & Weinberg 2002; Guzik & Seljak 2002; Mandelbaum et al. 2006a). We describe the prescription briefly below and refer the interested readers to the aforementioned three papers for details.

The signals of w_p and $\Delta \Sigma$ are obtained by projecting the 3D real-space galaxy autocorrelation function ξ_{gg} and the galaxy–matter cross-correlation function ξ_{gm} , respectively. The projection of ξ_{gg} to w_p is given by equation (2), while for the g–g lensing it is via

$$\Delta \Sigma(r_p) = \langle \Sigma(< r_p) \rangle - \Sigma(r_p), \quad (28)$$

where

$$\Sigma(r_p) = \bar{\rho}_m \int_{-\infty}^{+\infty} [1 + \xi_{gm}(r_p, r_\pi)] dr_\pi, \quad (29)$$

and

$$\langle \Sigma(< r_p) \rangle = \frac{2}{r_p^2} \int_0^{r_p} r_p' \Sigma(r_p') dr_p'. \quad (30)$$

We are ignoring the effects from the radial window, which is broad enough that it is not relevant at galaxy scales (Baldauf et al. 2010). The formalisms for predicting ξ_{gm} and ξ_{gg} are very similar, as the satellite distribution is merely a discrete realization of NFW tracers with a different concentration than the dark matter. There is, however, one extra contribution to ξ_{gm} and $\Delta \Sigma$ at small scales from the matter retained by the subhaloes that host the satellite galaxies (beside the model-independent stellar mass component,

see equation 25). We will discuss this ‘subhalo’ lensing term in more detail in Section 5.3 and focus on the similar terms shared by ξ_{gg} and ξ_{gm} here.

Let us consider a general scenario in which the correlation is between a primary galaxy catalogue and a secondary catalogue consisting of tracer particles t , whether it be the dark matter (ξ_{gm}) or the same galaxies as the primaries (ξ_{gg}). The correlation signal between the primary and the secondary can be decomposed into two components,

$$\xi_{\text{gt}}(r) + 1 = \left[\xi_{\text{gt}}^{\text{1h}}(r) + 1 \right] + \left[\xi_{\text{gt}}^{\text{2h}}(r) + 1 \right], \quad (31)$$

where ‘1h’ and ‘2h’ refer to the so-called one-halo and two-halo terms, respectively, and the +1 following each ξ_{gt} term is to relate the correlation piece to the corresponding number counts. The ‘1h’ term can be theoretically derived via the simple Peebles & Hauser (1974) estimator,

$$\xi_{\text{gt}}^{\text{1h}}(r) + 1 = \frac{\text{DD}^{\text{1h}}(r)}{\text{RR}(r)}. \quad (32)$$

The $\text{RR}(r)$ term is the expected number of pairs consisting of randomly distributed galaxies and t particles separated by distance between r and $r + dr$,

$$\text{RR}(r) = g(4\pi r^2 dr) \bar{n}_{\text{g}} \bar{\rho}_t V, \quad (33)$$

where \bar{n}_{g} and $\bar{\rho}_t$ are the mean densities of sample galaxies and t particles, respectively, and V is the survey volume. The pre-factor g is 1 if t is a different species than the primary galaxies (ξ_{gm} , with $\bar{\rho}_t \equiv \bar{\rho}_{\text{m}}$), and 1/2 if the two are identical (ξ_{gg} , with $\bar{\rho}_t \equiv \bar{n}_{\text{g}}$). The $\text{DD}(r)$ term requires separate treatment of central and satellite galaxies within each halo first, and then integration over the halo mass function,

$$\begin{aligned} \text{DD}^{\text{1h}}(r) = & V dr \int \frac{dn}{dM_{\text{h}}} \left[\text{DD}_{\text{cen,t}}(M_{\text{h}}) F'_{\text{cen,t}}(r|c_{\text{t}}) \right. \\ & \left. + \text{DD}_{\text{sat,t}}(M_{\text{h}}) F'_{\text{sat,t}}(r|c_{\text{g}}, c_{\text{t}}) \right] dM_{\text{h}}, \end{aligned} \quad (34)$$

where $\text{DD}_{\text{cen,t}}$ and $\text{DD}_{\text{sat,t}}$ are the total numbers of the cen- t and sat- t type of pairs expected within a halo of mass M_{h} , respectively, and $F_{\text{cen,t}}(r)$ or $F_{\text{sat,t}}(r)$ is the cumulative probability distribution of the numbers of each of these two pair types within that halo (although equation (34) requires their respective differential forms, described further below). For ξ_{gg} ,

$$\text{DD}_{\text{cen,g}}(M_{\text{h}}) = \langle N_{\text{cen}}(M_{\text{h}}) N_{\text{sat}}(M_{\text{h}}) \rangle, \quad (35)$$

and

$$\text{DD}_{\text{sat,g}}(M_{\text{h}}) = \langle N_{\text{sat}}(M_{\text{h}}) (N_{\text{sat}}(M_{\text{h}}) - 1) \rangle / 2, \quad (36)$$

while $F'_{\text{cen,g}}(r)$ and $F'_{\text{sat,g}}(r)$ are the galaxy NFW profile of concentration $c_{\text{g}}(M_{\text{h}})$ and the convolution of that galaxy NFW profile with itself, respectively. Similarly for ξ_{gm} ,

$$\text{DD}_{\text{cen,m}}(M_{\text{h}}) = \langle N_{\text{cen}}(M_{\text{h}}) \rangle M_{\text{h}}, \quad (37)$$

and

$$\text{DD}_{\text{sat,m}}(M_{\text{h}}) = \langle N_{\text{sat}}(M_{\text{h}}) \rangle M_{\text{h}}, \quad (38)$$

while $F'_{\text{cen,m}}(r)$ and $F'_{\text{sat,m}}(r)$ are the dark matter NFW profile of concentration $c_{\text{dm}}(M_{\text{h}})$ and the convolution of that dark matter profile with the galaxy NFW profile of concentration $c_{\text{g}}(M_{\text{h}})$, respectively. We use the analytic formula in Zheng & Weinberg (2007) for the convolution of two NFW profiles, either with the same or different concentration parameters. All the $F(r)$ functions are normalized so that $F(r = 2r_{200\text{m}})$ is unity, i.e. we assume both the satellite galaxies

and dark matter are contained within the virial radii of their host haloes. Recently, the study by van Daalen & Schaye (2015) suggested that it is important to account for the matter outside haloes when estimating the small-scale matter power spectrum using the halo model. However, this missing power problem does not exist for predicting galaxy clustering and g-g lensing, as the (sub)haloes occupied by galaxies are generally massive enough to be fully accounted in the calculations.

The ‘two-halo’ terms are relatively straightforward to calculate above the halo exclusion scales ($r_{\text{ex}} \sim 3 h^{-1}$ Mpc, described further below), via

$$\xi_{\text{gg}}^{\text{2h}}(r) = b_{\text{g}}^2 \zeta^2(r) \xi_{\text{mm}}(r), \quad (39)$$

and

$$\xi_{\text{gm}}^{\text{2h}}(r) = b_{\text{g}} \zeta(r) r_{\text{cc}}(r) \xi_{\text{mm}}(r), \quad (40)$$

where b_{g} is the galaxy linear bias, $\zeta(r)$ characterizes the fractional scale-dependence in the large-scale halo bias, and $r_{\text{cc}}(r)$ is the cross-correlation coefficient between the galaxies and the dark matter on relevant scales, defined by

$$r_{\text{cc}}(r) = \frac{\xi_{\text{gm}}(r)}{\sqrt{\xi_{\text{gg}}(r) \xi_{\text{mm}}(r)}}. \quad (41)$$

To a good approximation $r_{\text{cc}}(r)$ is close to unity on scales above r_{ex} (Guzik & Seljak 2001; Weinberg et al. 2004; Baldauf et al. 2010), and so we set $r_{\text{cc}}(r) \equiv 1$ and use the empirical fitting function of $\zeta(r)$ from Tinker et al. (2005). We computed b_{g} as the galaxy occupation number-weighted halo bias

$$b_{\text{g}} = \bar{n}_{\text{g}}^{-1} \int b(M_{\text{h}}) \frac{dn}{dM_{\text{h}}} dM_{\text{h}}, \quad (42)$$

where $b(M_{\text{h}})$ is the halo bias function (see the sixth column of Table 1 for the average linear bias of each iHOD sample).

Within the halo exclusion regime (Cacciato et al. 2009), however, equations (39) and (40) are no longer valid because the centre of one halo cannot sit within the virial radius of another halo, i.e. $r_{\text{ex}} < 2 \times r_{200\text{m}}^{\text{max}}$, where $r_{200\text{m}}^{\text{max}}$ is the maximum halo radius in the calculation. Therefore, the ‘two-halo’ terms at a given distance require explicit integration over pairs of haloes that are too small to run into the centre of each other when separated by that distance. We follow the prescription described in Tinker et al. (2005) for the treatment of halo exclusion and adopt the method in Yoo et al. (2006) to circumvent the issue of unsatisfied integral constraints in the halo mass function and halo bias function (see equation 16 of Yoo et al. 2006, for details).

5.3 Subhalo contribution to $\Delta\Sigma$ in MassiveBlack-II

The extra ‘subhalo’ lensing term, as mentioned in Section 5.2, arises because the satellite galaxies are sitting at the local density peaks (i.e. their subhaloes), rather than random positions within their main haloes (Sheth & Jain 2003; Giocoli et al. 2010). From investigating a suite of N -body cosmological simulations, Mandelbaum et al. (2005b) found that the fractional contribution of the subhalo term on scales below $0.1 h^{-1}$ Mpc is roughly equal to the satellite fraction within the sample, but is cut off on slightly larger scales because of the tidal truncation of those subhaloes inside larger haloes. They discovered that a truncation of the subhaloes at ~ 0.4 times the virial radii gives a good match to the lensing signal measured from the simulations. A subsequent exploration by Yoo et al. (2006), however, estimated that the error induced by ignoring the subhalo contribution is below 10 percent on the relevant scales using a

suite of relatively small smoothed particle hydrodynamic (SPH) simulations. It is still controversial as to whether the modelling of the subhalo lensing is necessary. While some studies (Vander et al. 2014; Hudson et al. 2015) adopted the ‘tidally stripped’ subhalo lensing model proposed in Mandelbaum et al. (2005b), other studies (L12; Coupon et al. e.g. 2015) ignored the lensing from subhaloes, either based on the result of Yoo et al. (2006), and/or the fact that the measured total lensing signal is much less constraining than other probes used in their papers (SMF and galaxy clustering).

We revisit the subhalo lensing issue by employing the MassiveBlack-II⁴ SPH simulation (MB-II; Khandai et al. 2015), a P-GADGET cosmological hydrodynamic simulation evolved with a total of 2×1792^3 dark matter and gas particles (mass resolution \sim several times $10^6 h^{-1} M_\odot$) inside a cubic volume of $100^3 h^3 \text{Mpc}^{-3}$. MB-II is one of the highest resolution simulations of this size which includes a self-consistent model for star formation, black hole accretion, and associated feedback. Thanks to its exquisite capability of resolving subhaloes down to $10^9 h^{-1} M_\odot$ with realistic baryonic effects (for the details on the HOD of simulated galaxies in MB-II, see Tucker et al, in preparation), we can quantify the impact of the often-ignored subhalo lensing term across the entire stellar mass range, by comparing the g–g lensing signals measured from the MB-II galaxies with subhaloes and from a sample of mock galaxies without. Following Yoo et al. (2006), in order to construct a subhalo-less mock galaxy sample, we identify the satellite galaxies in each main halo, and randomize their position angles relative to the halo centre while keeping their halo-centric distances fixed. In this way, we ensure that the only difference between the lensing signals measured from the MB-II galaxies and the mock galaxies on small scales is induced by the presence versus absence of dark matter within the subhaloes.

Fig. 5 shows the results of this experiment in six different stellar mass bins at $z = 0.06$, with M_* increasing from left to right and top to bottom. In each top sub-panel, the satellite fraction and the stellar mass range are marked on the top. The open circles with error bars indicate the measurement for the total g–g lensing signal for that stellar mass bin, decomposed into the central (red solid) and satellite (blue solid) terms. The error bars are derived from jackknife resampling of the simulation volume. The cyan dashed line shows the satellite term measured from the mock galaxies, which decreases rapidly on small scales due to the lack of subhaloes. The magenta dashed line is a crude estimate of the subhalo lensing term via re-scaling the amplitude and truncating the central term (red solid) at $0.4 r_{200m}$. The amplitude is re-scaled by $f_{\text{sat}}/(1 - f_{\text{sat}})$, assuming the subhaloes that host satellites share the same inner matter density profile as those host central galaxies of the same stellar mass, analogous to the ansatz employed in the SHAM technique. The black solid line is the sum of this estimated subhalo term, the directly measured central term, and the satellite term from the mock sample, serving as a rough estimate of the predicted total signal with the subhalo lensing contribution, whereas the black dashed line is the sum of central and mock satellite terms without any subhalo contribution. The difference between the open circles and the black dashed lines represents the magnitude of the error on g–g lensing due to ignoring the subhalo lensing piece, and the difference between the open circles and the black solid lines represents the deficiency of the overly simplified ‘re-scaled central’ subhalo lensing model in overpredicting or underpredicting the signal at various scales.

The effects of ignoring the subhalo lensing term are better illustrated by the bottom sub-panels of Fig. 5 where the black solid and dashed lines shown above are divided by the measured total signals. The grey shaded region indicates the typical uncertainties on the ratios propagated from the measurement uncertainties on $\Delta\Sigma$. The ratio plots clearly demonstrate that, ignoring the subhalo lensing term causes 15–35 per cent systematic underprediction of the total signal on scales below $\sim 0.2 h^{-1} \text{Mpc}$, and the deviation is proportional to the fraction of satellite galaxies in the sample. This contradicts the result from Yoo et al. (2006) who found the effect generally below 10 per cent at all radius. The opposite conclusions drawn from the two sets of simulations could be due to the drastic difference in the resolution and size – MB-II has 2000 times more particles in eight times the volume of the simulation used by Yoo et al. (2006), and/or the average mass of the subhaloes used in the two experiments (i.e. Yoo et al. 2006 could only go to subhaloes above $10^{12} h^{-1} M_\odot$, hence much lower satellite fraction). Once the simple ‘re-scaled central’ subhalo lensing term is added, the predicted signal agrees with the direct measurement within the uncertainties for the three high stellar mass bins in the bottom row, but overpredicts the total signal by 15–35 per cent on the relevant scales for the three low stellar mass bins in the top row. This discrepancy between the simple remedy and the simulation in the low-mass samples is caused by the enhanced tidal attenuation effect on the low-mass subhaloes from their host haloes, besides the usual tidal truncation effect seen for subhaloes of all masses. This attenuation effect is consistent with the findings in Li et al. (2014), who measured the subhalo lensing signal using the satellite lens galaxies selected from the SDSS group catalogue constructed from a redshift-space Friends-of-Friends algorithm (Yang et al. 2007). They found significant $\Delta\Sigma$ signal below projected radius $\sim 0.1 h^{-1} \text{Mpc}$ for satellite lenses located $0.1\text{--}0.5 h^{-1} \text{Mpc}$ away from their respective group centres, and the amplitude of the subhalo lensing signal can be explained by a truncated and attenuated version of those NFW haloes that host central galaxies of the same stellar masses.

Drawing from these findings in the MB-II simulation and the SDSS groups, we model the subhalo contribution to the galaxy-matter cross-correlation function $\xi_{\text{gm}}^{\text{sub}}$ as an attenuated and truncated version of the central term $\xi_{\text{gm}}^{\text{cen}}$,

$$\xi_{\text{gm}}^{\text{sub}}(r|M_*) = \begin{cases} f_t \times \left[\frac{f_{\text{sat}}}{1-f_{\text{sat}}} \xi_{\text{gm}}^{\text{cen}}(r|M_*) \right], & r < r_t \\ 0, & r \geq r_t, \end{cases} \quad (43)$$

where f_t and r_t is the attenuation factor and the truncation radius, respectively. We adopt a truncation radius of $r_t = 0.4 r_{200m}(M_*)$, according to the findings in Mandelbaum et al. (2006a) and Fig. 5. We ignore the dependence of r_t on halocentric distance here, despite in reality the truncation should be a function of subhalo orbital parameters, i.e. the number of pericentre passages. The value of f_t should be stellar mass dependent – we adopt f_t values of 0.5 and 1.0 for stellar mass samples below and above $10^{10} h^{-2} M_\odot$, respectively, informed by the results in Fig. 5.

However, as mentioned in Section 4.4, the small-scale lensing modelling is further complicated by the miscentering effect, which is absent from Fig. 5. As pointed out in Yoo et al. (2006), the miscentering effect smooths the overall $\Delta\Sigma$ signal on small scales with a kernel of length $\sim 0.1 h^{-1} \text{Mpc}$, and is likely to have a large impact for the higher stellar mass bins, i.e. within groups and clusters. The smoothing effect will effectively reduce the predicted signal for the central galaxies at the 10–15 per cent level on scales below $0.1 h^{-1} \text{Mpc}$, where the subhalo lensing effect

⁴ <http://mbii.phys.cmu.edu/>

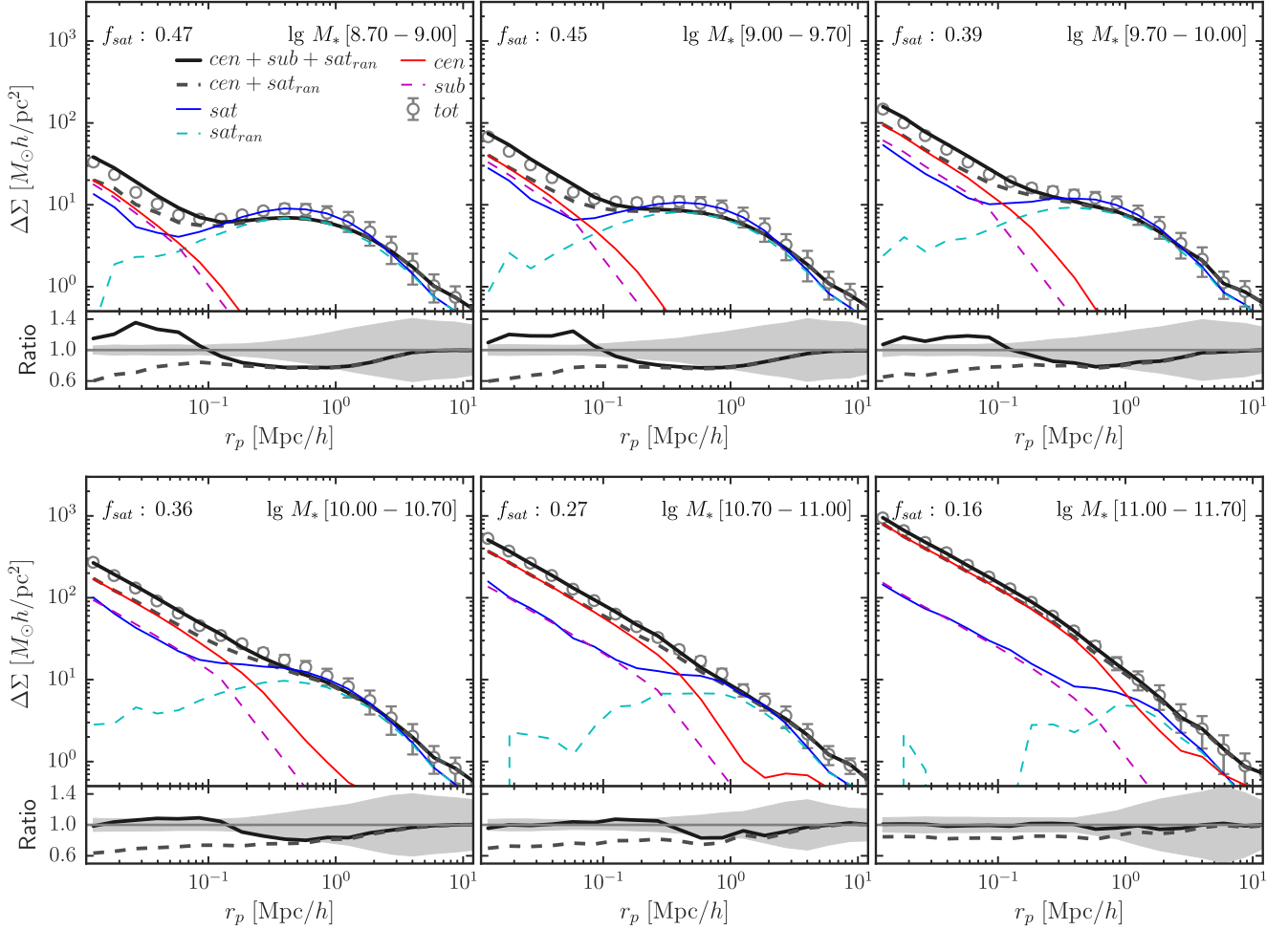


Figure 5. The g–g lensing signal of six different stellar mass samples in the MB-II simulation at $z = 0.06$, highlighting the importance of including the subhalo contribution in the modelling of the surface density contrast $\Delta\Sigma$. In each upper sub-panel, open circles are the total $\Delta\Sigma$ signal measured directly from the simulation, which is the sum of the red and the blue solid curves, indicating the contributions from the centrals and the satellites, respectively. The cyan dashed curve represents the satellite contribution measured after randomizing the position angles of the simulated satellites (distances to the halo centre kept fixed). The magenta dashed curve accounts for the subhalo contribution by re-scaling the central contribution according to the satellite fraction listed on the top left (the values of these f_{sat} are higher than in the data, mostly because the stellar masses in MB-II are systematically higher.) and truncating the 3D density profile at $0.4r_{200\text{m}}$. The black solid curve is the sum of the subhalo (magenta), the central (red), and the randomized satellite (cyan) curves, and the black dashed curve the sum of only the latter two. The ratio of the black solid and the black dashed curves to the open circles is shown on each lower sub-panel. Ignoring the subhalo contribution results in an underestimate of the total signal by 15–35 per cent on small scales, roughly proportional to the satellite fraction in each sample. For the three low-mass samples, using the re-scaled and truncated central curve as a rough estimate for the subhalo contribution overestimates the total signal by 10–35 per cent on small scales, and by ~ 15 per cent in the transitional regime, albeit within the measurement uncertainties indicated by the grey band. For the high-mass samples the black curves agree with the direct measurements within the uncertainties across all scales.

operates in the opposite direction. Since the satellite fraction at the high-mass bins is ~ 20 per cent, this reduction can be effectively absorbed by a low f_t value of ~ 0.5 in equation (43) for the high-mass bins. Therefore, to simultaneously capture both the effects of subhalo lensing at low M_* and miscentering at high M_* on scales below $0.1 h^{-1} \text{ Mpc}$, we adopt a constant attenuation factor of $f_t = 0.5$ in our analysis, regardless of the stellar mass of the sample. We have tried different values of $f_t = 0.5 \pm 0.2$ and found that our results are robust to those changes, due to the relatively large statistical uncertainties of the g–g lensing measurements on small scales (see Fig. 6).

For each redshift slice within the iHOD samples and for each individual cHOD sample, the predictions for the w_p and $\Delta\Sigma$ signals are obtained by projecting ξ_{gg} and ξ_{gm} to 2D, according to equations (2) and (28), respectively. We adopt the same $r_\pi^{\text{max}} = 60 h^{-1} \text{ Mpc}$ for

the theoretical predictions of w_p as for the measurements. Finally, for each iHOD sample we combine the predictions from all its redshift slices to obtain the predictions for that entire sample via equations (26) and (27).

6 PARAMETER CONSTRAINTS

Fig. 6 compares the predictions from our best-fitting model (thick lines) to the measurements from data for the eight iHOD stellar mass samples. The galaxy clustering and g–g lensing results are shown in each pair of the top and bottom sub-panels, respectively. Each best-fitting curve is decomposed into the two-halo (thin dot-dashed), the one-halo central (thin dashed), and the one-halo satellite (thin solid) contributions. For the best-fitting g–g lensing signal of each sample, we also show the contributions from the subhalo lensing term and the

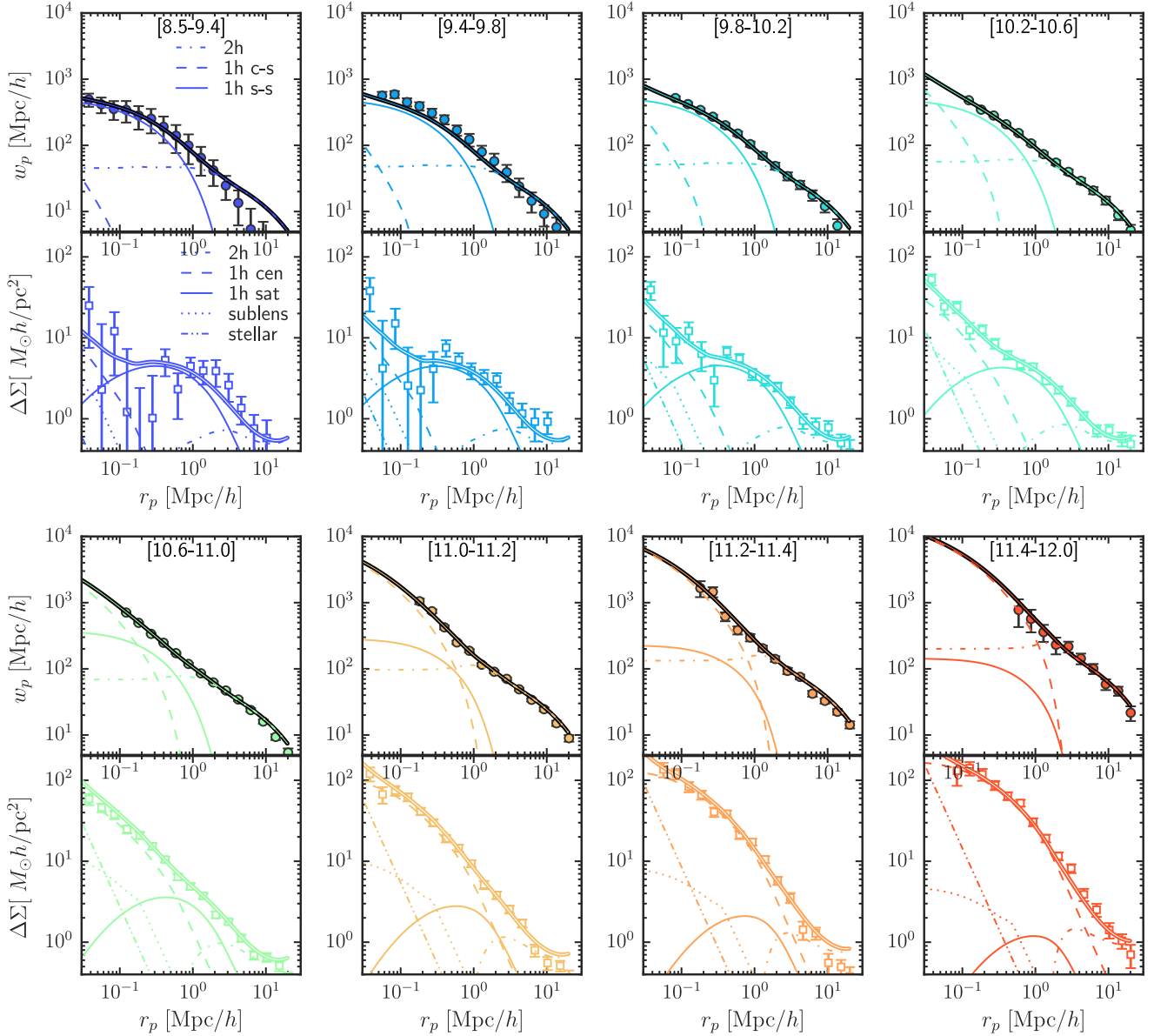


Figure 6. Comparison between the galaxy clustering and g-g lensing measurements from SDSS with the signals predicted by the best-fitting iHOD model, for the eight stellar mass-selected samples. For each sample, the top and bottom panels show the projected correlation function w_p and g-g lensing signal $\Delta\Sigma$, respectively. In each panel, the data points with error bars are the measurements, and the thick line is the best-fitting signal, which is decomposed into the 2-halo (thin solid), the 1-halo central (thin dashed), and the 1-halo satellite (thin dot-dashed) contributions shown underneath. The dotted and the double-dot-dashed lines in the lower sub-panels represent the lensing contributions from the subhalo dark matter and the galaxy stellar mass, respectively. The x-/y-axis ranges are uniform across all panels. The model provides excellent fit to both the clustering and lensing signals of galaxies over four decades in stellar mass.

‘point source’ stellar mass as the thin dotted and double-dot-dashed lines, respectively. The subhalo lensing term begins to dominate the one-halo satellite contribution at around half the virial radius of the average main halo corresponding to that sample, and then the stellar mass term takes over at the galactic scales below tens of h^{-1} kpc. We will come back to Fig. 6 for the detailed comparison between the model fits and the measurements in Section 6.2.

6.1 Likelihood model and bayesian inference

Armed with the capability of predicting the galaxy clustering and g-g lensing signals for any given cHOD or iHOD sample, we can

infer the posterior probability distributions of the model parameters from the measurements within a Bayesian framework, assuming a Gaussian likelihood model and a set of uninformative priors on those parameters.

The observable vector in our likelihood model has two components.

- (1) $w_p^i(r_p^j)$: w_p profile of stellar mass sample i measured at projected radius r_p^j (from r_p^{fb} to $20.0 h^{-1}$ Mpc), for $i \in \{1, \dots, 8\}$ (iHOD) or $\{1, \dots, 6\}$ (cHOD), and $j \in \{1, \dots, n\}$, where n is the number of data points used for each stellar mass samples. Ranked by ascending order in the average stellar mass, the samples have n

Table 2. Description, prior specifications, and posterior constraints of the parameters in the model. All the priors are uniform distributions running across the entire range of possible values for the parameters, and the uncertainties are the 68 per cent confidence regions derived from the 1D posterior probability distributions.

Parameter	Description	Uniform prior range	iHOD	cHOD
$\lg M_h^1$	Characteristic halo mass of the SHMR	[9.5, 14.0]	$12.10^{+0.17}_{-0.14}$	$12.32^{+0.29}_{-0.29}$
$\lg M_*^0$	Characteristic stellar mass of the SHMR	[9.0, 13.0]	$10.31^{+0.10}_{-0.09}$	$10.47^{+0.18}_{-0.21}$
β	Low-mass slope of the SHMR	[0.0, 2.0]	$0.33^{+0.21}_{-0.15}$	$0.54^{+0.29}_{-0.26}$
δ	Controls high-mass slope of the SHMR	[0.0, 1.5]	$0.42^{+0.03}_{-0.04}$	$0.42^{+0.08}_{-0.09}$
γ	Controls intermediate-mass behaviour of the SHMR	[-0.1, 4.9]	$1.21^{+0.18}_{-0.20}$	$1.05^{+0.24}_{-0.26}$
B_{sat}	Normalizes the scaling of M_{sat}	[0.01, 25.0]	$8.98^{+1.18}_{-0.87}$	$11.22^{+2.61}_{-1.99}$
β_{sat}	Slope of the scaling of M_{sat}	[0.1, 1.8]	$0.90^{+0.04}_{-0.05}$	$0.85^{+0.06}_{-0.05}$
B_{cut}	Normalizes the scaling of M_{cut}	[0.0, 6.0]	$0.86^{+0.32}_{-0.37}$	$0.73^{+0.58}_{-0.44}$
β_{cut}	Slope of the scaling of M_{cut}	[-0.05, 1.50]	$0.41^{+0.16}_{-0.15}$	$0.63^{+0.31}_{-0.33}$
α_{sat}	Power-law slope of the satellite HOD	[0.5, 1.5]	$1.00^{+0.03}_{-0.02}$	$1.07^{+0.05}_{-0.05}$
$\sigma_{\ln M_*}$	Low-mass scatter in the SHMR	[0.01, 3.0]	$0.50^{+0.04}_{-0.03}$	$0.42^{+0.07}_{-0.08}$
η	Slope of the scaling of high-mass scatter	[-0.4, 0.4]	$-0.04^{+0.02}_{-0.02}$	$-0.01^{+0.04}_{-0.03}$
f_c	Concentration ratio between satellites and dark matter	[0.1, 3.0]	$0.86^{+0.14}_{-0.11}$	$1.06^{+0.32}_{-0.20}$

$\in \{17, 16, 15, 14, 14, 13, 13, 10\}$ (iHOD) or $n \in \{16, 15, 14, 14, 13, 10\}$ (cHOD), due to the different fibre-collision induced cutoffs on small scales. There are 112 (82) data points in total for the iHOD (cHOD) analysis.

(2) $\Delta \Sigma^i(r_p^k)$: $\Delta \Sigma$ profile of stellar mass sample i measured at radius r_p^k , ranging from r_p^{cut} to $20.0 h^{-1}$ Mpc, where r_p^{cut} is the small-scale lensing cutoff caused by the systematic uncertainties in estimating the ‘boost factor’ for bright samples (see appendix for more details). We use the same set of stellar mass samples as clustering in each analysis. The minimum fitting scale is $25 h^{-1}$ kpc for the samples below $\lg M_* = 10.6$, but increases as the sample gets brighter due to large r_p^{cut} , reaching $0.12 h^{-1}$ Mpc for the highest stellar mass bin. Therefore, $k \in \{1, \dots, n'\}$, where $n' \in \{18, 18, 18, 18, 17, 17, 16, 15\}$ (iHOD) and $\in \{18, 18, 17, 17, 16, 15\}$ (cHOD). There are 137 (101) $\Delta \Sigma$ data points in total for the iHOD (cHOD) analysis.

We model the combinatorial vector \mathbf{x} of the w_p and the $\Delta \Sigma$ components as a multivariate Gaussian ($N = 256$ variables in total for iHOD and $N = 190$ for cHOD), which is fully specified by its mean vector ($\bar{\mathbf{x}}$) and covariance matrix (\mathbf{C}). The Gaussian likelihood is thus

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = |\mathbf{C}|^{-1/2} \exp\left(-\frac{(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{C}^{-1} (\mathbf{x} - \bar{\mathbf{x}})}{2}\right), \quad (44)$$

where

$$\boldsymbol{\theta} \equiv \{\lg M_h^1, \lg M_*^0, \beta, \delta, \gamma, B_{\text{sat}}, \beta_{\text{sat}}, B_{\text{cut}}, \beta_{\text{cut}}, \alpha_{\text{sat}}, \sigma_{\ln M_*}, \eta, f_c\}. \quad (45)$$

We adopt flat priors on the model parameters, with a uniform distribution over a broad interval that covers the entire possible range of each parameter (see the third column of Table 2). The final covariance matrix \mathbf{C} is assembled by aligning the error matrices of w_p and $\Delta \Sigma$ measured for individual samples along the diagonal blocks of the full $N \times N$ matrix. We ignore the weak covariance between w_p and $\Delta \Sigma$ (with the covariance being weak due to the fact that $\Delta \Sigma$ is dominated by shape noise), and between any two measurements of the same type but for different stellar mass samples.

In the parameter inference stage, the posterior distribution is derived using the Markov Chain Monte Carlo (MCMC) algorithm EMCEE (Foreman-Mackey et al. 2013), where an affine-invariant

ensemble sampler is utilized to fully explore the 13D parameter space. For each MCMC chain of the iHOD and cHOD inferences, we perform 90 000 iterations, 30 000 of which belong to the burn-in period for adaptively tuning the steps. To eliminate the tiny amount of residual correlation between adjacent iterations, we further thin the chain by a factor of 10 to obtain our final results.

6.2 Posterior probability distributions

Fig. 7 presents a summary of the inferences for both the iHOD (brown filled) and cHOD (grey open) analyses, showing the 1D posterior distribution for each of the 13 model parameters (diagonal panels), and the 95 and 68 per cent confidence regions for all the parameter pairs (off-diagonal panels). In the panels of the lower triangle, we highlight the results from our fiducial model, i.e. the iHOD model, employing the clustering and g–g lensing from the entire galaxy population above the mixture limit and self-consistently accounting for the survey incompleteness in stellar mass. In those panels, we also show the 95 per cent confidence regions from the iHOD constraints that employ the clustering and the lensing signals separately. The constraint from using the clustering alone (blue solid contours) is significantly tighter than that from the g–g lensing (red dashed contours), due to the much higher S/N in the w_p measurements. However, the g–g lensing is absolutely essential in the joint iHOD analysis, helping break the degeneracy between the scatter parameters ($\sigma_{\ln M_*}$ and η) and the slope of the SHMR on the high-mass end (δ). In each panel of the upper triangle, we compare the constraints from the fiducial model (filled contours) to that of the traditional approach, i.e. the cHOD model, which is limited to 54 per cent of the iHOD galaxies and assumes the samples to be complete (open contours). While the two constraints are largely consistent with each other, the iHOD analysis is able to obtain a much tighter constraint because of the larger span in stellar mass range and the higher number of galaxies (hence the higher S/N of measurements illustrated in Fig. 2). In particular, the iHOD analysis substantially improves the constraints on the pivot of the SHMR ($\lg M_h^1, \lg M_*^0$) and its high-mass end slope δ . The five parameters that describe the satellite HOD benefit the most from the inclusion of low stellar mass samples, with much stronger constraint in

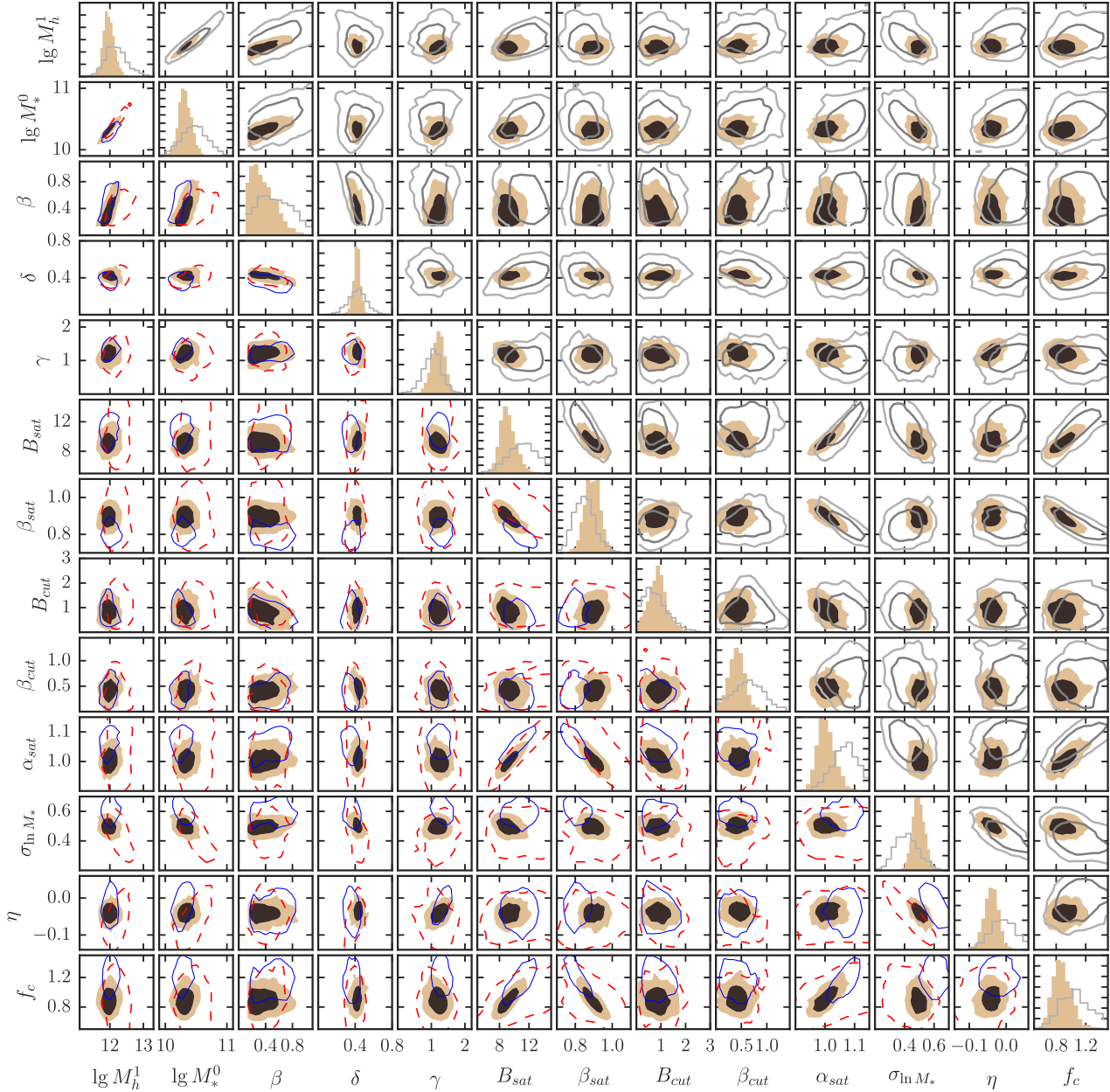


Figure 7. Confidence regions from our iHOD analysis of the galaxy clustering and the g–g lensing data in the 2D planes that comprised of all the pair sets of model parameters. Histograms in the diagonal panels show 1D posterior distributions of individual parameters. For comparison, constraints from the cHOD analysis are only shown in the panels of the upper triangle. Contour levels run through confidence limits of 95 per cent (light brown/grey) and 68 per cent (dark brown/grey) inwards. The tighter iHOD constraint compared to cHOD is due to the combination of improved overall statistics from larger samples and additional information from the low-mass galaxies. In the panels of the lower triangle, blue solid and red dashed contours indicate the 95 per cent confidence regions of the two separate iHOD constraints using the clustering and the lensing data, respectively.

iHOD than in cHOD. The 68 per cent confidence regions of the 1D posterior constraints are listed in Table 2.

The iHOD model favours a scatter of $\sigma_{\ln M_*} = 0.50^{+0.04}_{-0.03}$ about the mean SHMR with a slightly negative slope of $\eta = -0.04 \pm 0.02$, while the cHOD model infers a constant scatter of $\sigma_{\ln M_*} = 0.42^{+0.07}_{-0.08}$. However, the two constraints converge to a similar scatter on the high-mass end, where the iHOD model shrinks its mass-dependent scatter to meet the lower constant value preferred by the cHOD model. While the cHOD does not constrain the concentration ratio f_c well, the iHOD strongly favours the scenario where the

satellite distributions are ~ 15 per cent less concentrated than the dark matter within the same haloes ($f_c = 0.86^{+0.14}_{-0.11}$). This galaxy underconcentration agrees with observational findings in Wang et al. (2014) and Budzynski et al. (2012, but see Watson et al. 2012 for the degeneracy between f_c and the inner slope of the cluster density profiles).

The improvement on the overall constraints from iHOD to cHOD is better illustrated in Fig. 8, where we translate the uncertainties of the individual model parameters to that of the mean SHMR and the satellite HOD separately. In the left-hand panel, the red solid and

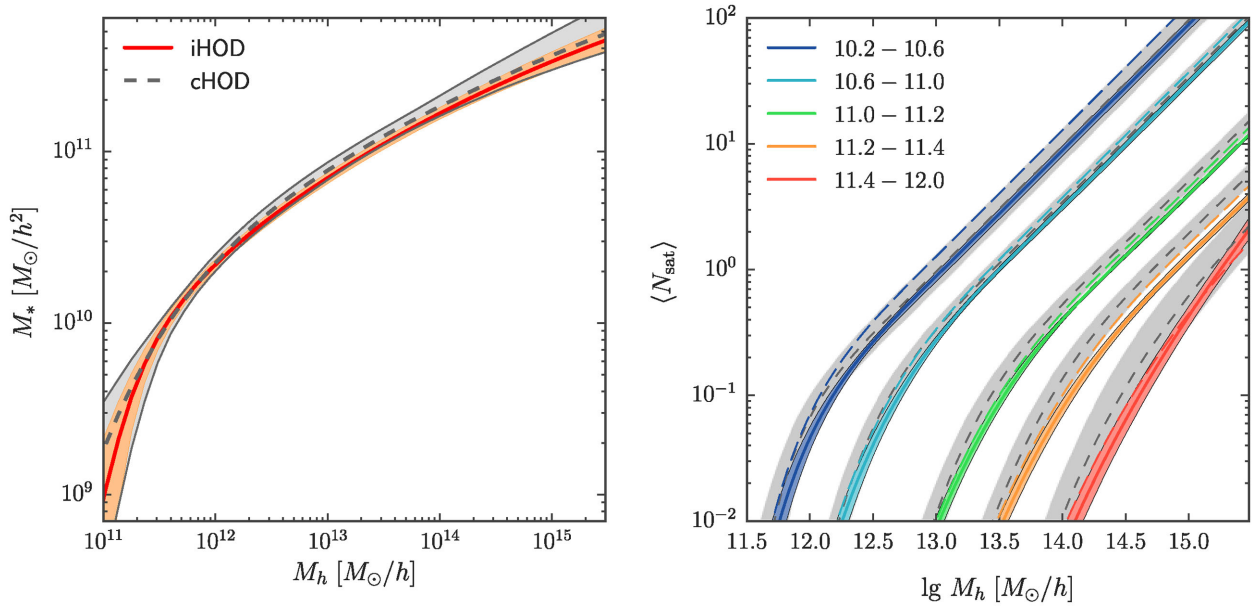


Figure 8. Constraints on the SHMR for the central galaxies (left) and the satellite HODs for four stellar mass bins (right), as a more visually appealing and scientifically informative presentation of the confidence regions shown in Fig. 7. In the left-hand panel, the red solid and the grey dashed lines show the inferred SHMR from the iHOD and cHOD analysis, respectively, and the shaded regions represent the 68 percent confidence range at fixed halo mass. In the right-hand panel, solid coloured and dashed grey lines show the expected satellite HODs for the observed galaxies from the best-fitting iHOD and cHOD model, respectively, with each shaded band indicating the 68 percent confidence range at fixed halo mass; long-dashed coloured lines show the parent (i.e. both observed and unobserved) HODs predicted from the best-fitting iHOD model. The difference between the solid and long-dashed coloured lines implies possible incompleteness in the stellar mass samples.

grey dashed curves are the mean SHMR inferred from the iHOD and cHOD analyses, respectively. The width of each shaded band, i.e. the 68 per cent uncertainty in the mean $\log(\text{stellar mass})$ at fixed halo mass, directly reflects the joint posterior probability distribution of the five parameters ($\lg M_h^1$, $\lg M_*^0$, β , δ , γ) that determine the mean SHMR. The two constraints are consistent with each other, especially on the high-mass end. This is very reassuring – as we mentioned in Section 4.1 that the two models are identical to each other for the high stellar samples, because the completeness approaches unity and is independent of halo mass. As expected from Fig. 7, the uncertainties in the mean SHMR are greatly reduced across all halo masses in the iHOD constraint, with the most significant reduction happening above $M_h = 10^{13} h^{-1} M_\odot$ [by factor of 2 in $\log(\text{stellar mass})$]. Furthermore, both of the mean SHMRs are best constrained at the intermediate mass range around the pivotal point ($\lg M_h^1$, $\lg M_*^0$), due to the highest model sensitivity and the highest S/N at this stellar mass.

Similarly, in the right-hand panel of Fig. 8 we show the constraints on the satellite occupation number predicted for five of the stellar mass samples that were used in the cHOD analysis. The grey dashed lines show the median satellite occupation numbers as functions of halo mass inferred from the cHOD analysis, with their 1σ uncertainties shown as the corresponding grey bands. For comparison, we have predicted two types of satellite HODs for these five cHOD samples from the best-fitting iHOD model, one is for the *parent* satellite galaxies (thin dashed lines, HODs including both observed and unobserved galaxies), and the other for the *observed* galaxies (thick solid lines, HODs including only observed galaxies). To avoid clutter, we only show the uncertainty bands associated with the inferred *observed* satellite HODs, all of which are considerably narrower than the grey bands, especially for the two highest stellar mass samples. This improvement is

very encouraging. As we will show later in Section 7.3, the occupation statistics of the satellite galaxies, often being regarded as a mere nuance in the HOD modelling of galaxy clustering and HOD-based cosmological constraints (Yoo & Seljak 2012), is constrained well enough in the iHOD analysis to provide important insights into the physical formation and evolution of the satellite galaxies.

It is important to point out that, when calculating the HODs for the *observed* galaxies in Fig. 8, we have made use of the actual number of observed galaxies in each sample in order to normalize the amplitude of $\langle N_{\text{sat}}(M_h) \rangle$ using equation (12), while the *parent* HODs are predicted directly from the best-fitting model parameters using equation (9). As mentioned in Section 5.1, the iHOD analysis constrains the host halo distribution of the average galaxy in each sample by matching to the observed clustering and g–g lensing signals, and is thus entirely agnostic of the observed amplitude of the SMFs. Therefore, it is highly non-trivial to discover in the right-hand panel of Fig. 8 that, all the predicted *parent* HODs are lying right above their *observed* counterparts, consistent with our expectation that the galaxy samples above the mixture limit are subject to some mild level of incompleteness. We will inspect closely this consistency later in Section 7.1.

Fig. 6 compares the clustering and g–g lensing signals measured from data to that predicted from the set of best-fitting parameters derived from the median values of their corresponding 1D posterior probability distributions. We refer to this set of ‘median’ parameters as the ‘best-fitting’ parameters for simplicity, despite that it does not produce the maximum posterior probability or likelihood value. The lowest stellar mass sample is subject to severe cosmic variance due to small volume, resulting in a relatively poor w_p fit on large scales. The model also underpredicts the small-scale clustering signal for the $\lg M_* = [9.4, 9.8]$ sample, which is largely dominated

by the CfA2 great wall (Geller & Huchra 1989). Overall, the model provides an excellent fit to the data over almost four decades in stellar mass, across distance scales from the galactic discs to $10\text{--}20 h^{-1} \text{ Mpc}$.

6.3 Systematic uncertainties in the *i*HOD model

6.3.1 Systematic uncertainties from measurements and model assumptions

Observationally, the systematic errors in our fiducial *i*HOD analysis come from the uncertainties in the stellar mass estimates and in the measurements of the projected correlation functions and the g - g lensing signals. As mentioned in Section 2.2, the estimation of M_* is subject to various photometric and model uncertainties that mostly affect the overall normalization of the derived stellar masses, with little perturbation in the ranking order of individual galaxies within the catalogue. Therefore, the inferred mapping between the stellar mass and the dark matter haloes from our analysis can be straightforwardly re-calibrated for other stellar mass estimators.

The uncertainties in the w_p and the $\Delta\Sigma$ measurements are estimated internally from the data via jackknife resampling. Compared to external estimates derived from multiple independent catalogues, the jackknife estimate somewhat overestimates the correlation function errors on scales below $r_p \sim 2 h^{-1} \text{ Mpc}$ (Norberg et al. 2009), while on large scales it likely underestimates the errors because it does not include the cosmic variance above the scales of individual subsamples. Therefore, the jackknife errors for the two lowest stellar mass/redshift samples are underestimated on large scales, because of the small physical size of the jackknife patches. The error budget of these two stellar mass samples, however, is still dominated by the uncertainties on small scales, so the fiducial constraint is insensitive to the underestimation in the error covariance on large scales, despite the inadequate fit to the w_p measurements of the two lowest stellar samples on relevant scales.

The systematic errors in the $\Delta\Sigma$ measurements have additional contributions from the calibration biases mainly related to the shear estimation and the photo- z errors, each at a few per cent level (see Appendix A for the scale-dependent lensing systematics). Mandelbaum et al. (2013) addressed these systematics by conducting a suite of ratio tests (Mandelbaum et al. 2005b), i.e. comparisons of the signal computed using the same lens samples, but with different subsamples of the source catalogue. After applying well-understood corrections, the systematic errors are subdominant compared to the statistical errors and compared to the systematic uncertainties due to modelling assumptions, which we will discuss next.

The theoretical systematic errors in our fiducial *i*HOD analysis have four main sources: (1) simplified model assumptions in the *i*HOD formalism, (2) uncertainties in the theoretical prediction of halo statistics like the halo mass function and the halo bias function, (3) our ansatz of the weak M_h -dependence of the detection rate at fixed M_* , and (4) uncertainties in the combined treatment of the mis-centring and subhalo lensing. We have discussed (3) in detail in Section 4.1 and will comment more on its potential impact in Section 7.1, and the impact from (4) has been discussed and addressed in Section 5.3. Therefore, here we focus on the first two issues and discuss their potential impacts on our model constraints in turn.

As a variant of the HOD formalism, the *i*HOD model inherits the systematic uncertainties associated with the generic HOD models, which assume that the average galaxy content of haloes depends

solely on the halo mass, predicted by the basic excursion set theory of structure formation (Press & Schechter 1974; Bardeen et al. 1986; Bond et al. 1991). However, since the halo assembly histories in cosmological simulations are affected by the large-scale environment (Wang, Mo & Jing 2007; Dalal et al. 2008), some halo properties, including age, concentration, spin, richness, and most importantly clustering, exhibit systematic differences between low- and high-density environments at fixed halo mass (Sheth & Tormen 2004; Gao, Springel & White 2005; Harker et al. 2006; Wechsler et al. 2006; Zhu et al. 2006; Hahn et al. 2007; Jing, Suto & Mo 2007; Li, Mo & Gao 2008; Faltenbacher & White 2010; Croft et al. 2012). This effect, broadly termed ‘halo assembly bias’ (Gao & White 2007), has likely left its imprint on galaxy formation histories as well, resulting in a ‘galaxy assembly bias’. However, hydrodynamic simulations and SAMs predict only a small impact (< 10 per cent) of the halo assembly bias on galaxy clustering statistics (Yoo et al. 2006; Croton, Gao & White 2007; Zu et al. 2008), while observationally a smoking-gun detection of the galaxy assembly bias remains elusive (Berlind et al. 2006; Blanton & Berlind 2007; Wang et al. 2013; Hearin, Watson & van den Bosch 2015; Miyatake et al. 2015b). Additionally, the halo assembly bias becomes only significant for haloes below the characteristic non-linear mass scale, where our error budget is dominated by statistical errors and cosmic variance. However, it is worth noting that halo assembly bias may have a much greater impact on the HOD modelling of colour-selected galaxy samples, and could dominate the error budget on all mass scales for both colours (Zentner, Hearin & van den Bosch 2014).

6.3.2 Systematic uncertainties from cosmology

The cause for the theoretical uncertainties in predictions of the halo mass and bias functions is three-fold: (i) the Universe may have a non- Λ CDM cosmology (for halo statistics beyond Λ CDM; see Bhattacharya et al. 2011; Cui, Baldi & Borgani 2012b; Ichiki & Takada 2012; Murray, Power & Robotham 2013; Zhang et al. 2013; LoVerde 2014) or a different set of cosmological parameters (especially σ_8 ; see Planck Collaboration XIII 2015, for details on the tension between different probes) than the particular Λ CDM cosmology we adopt; (ii) the predictions should be inaccurate due to the unaccounted-for baryonic effects (at 5 to 20 per cent level depending on the mass scale and the feedback models; see Cui et al. 2012a; Cusworth et al. 2014; Velliscig et al. 2014); and (iii) the predictions may be poorly calibrated on the high-mass end where the haloes are rare (at the 5 per cent level for $M_h = 10^{15} h^{-1} M_\odot$ at $z = 0$; see Crocce et al. 2010; Watson et al. 2013; Bocquet et al. 2015). Among the three types of errors, the calibration errors affect few galaxies as the probability of any observed galaxies sitting in a halo with mass above $10^{15} h^{-1} M_\odot$ is extremely low, and are thus negligible in our analysis. Errors due to wrong cosmology, ignored baryonic effects, or the combination of both, change the halo mass and bias functions together in a coherent manner that can be roughly mimicked by changing Ω_m and/or σ_8 within the Λ CDM model. In particular, changing Ω_m at fixed σ_8 shifts the predictions along the halo mass axis almost uniformly (albeit with a minor change in the functional shapes due to the change in the power spectrum), while changing σ_8 at fixed Ω_m modifies the amplitude of the predictions more progressively at higher halo mass (see fig. 7 in Zu et al. 2014, for a pedagogical illustration). We explore the impact of (i) and (ii) by perturbing Ω_m and σ_8 using both theoretical arguments and mock tests below.

In our analysis on large scales, we are effectively using the halo bias versus halo abundance relation compressed from the halo mass and bias functions, i.e. the bias of M_h -thresholded halo samples $b(\bar{n})$ as a function of the comoving number density of that sample, $\bar{n}(>M_h)$. Due to the exponential decline of the abundance of massive haloes, the theoretical uncertainties in this $b(\bar{n})-\bar{n}(>M_h)$ relation are only important at the high-mass end, where the galaxy sample is primarily composed of massive BCGs with small satellite fraction. In the context of the peak background-split theory of biased clustering (Kaiser 1984; Cole & Kaiser 1989; Mo, Jing & White 1996; Sheth & Tormen 1999), for very rare, very highly biased peaks, $b(\bar{n}) \propto \sigma_8^{-1}$, yielding $w_p(\bar{n}) \propto b^2(\bar{n})\xi_{\text{mm}} \propto b^2(\bar{n})\sigma_8^2 \simeq \text{constant}$, and $\Delta\Sigma \propto \Omega_m b(\bar{n})\xi_{\text{mm}} \propto \Omega_m \sigma_8$, respectively. These two approximations hold even in the presence of baryons. Therefore, the clustering is barely affected by the changes in Ω_m or σ_8 , while the lensing is linearly responsive to the change in the product of Ω_m and σ_8 . Since the galaxy clustering measurements have a better overall S/N than the g-g lensing measurements, a 10 per cent change in $\Omega_m \sigma_8$ should produce a shift in the SHMR smaller than 0.06 dex in M_h (if all of our constraints came from large scales). To confirm this, we re-run the analysis using two different combinations of $(\Omega_m, \sigma_8) = (0.26, 0.80)$ and $(0.28, 0.80)$. For the first case where we change σ_8 alone, the shift in the derived SHMR is indiscernible, whereas for the latter case where we increase the product of Ω_m and σ_8 by 12 per cent, the derived SHMR shifts to the lower halo mass by ~ 0.08 dex, corresponding to an increase in the stellar mass at $M_h = 10^{13} h^{-1} M_\odot$ by only ~ 0.03 dex, with little change in the parameters that control the satellite HOD.

The remaining sources of theoretical systematic errors, including the uncertainties in the mean halo mass–concentration relation $c_{\text{dm}}(M_h)$,⁵ calibration in the scale-dependence of halo bias $\zeta(r)$, and the deviation of the cross-correlation coefficient $r_{\text{cc}}(r)$ from unity, are expected to be much smaller than the four main errors discussed above. Summing all the systematic errors in quadrature, we would find the total systematic error comparable to or even dominant over the statistical uncertainties in the measurements. However, as pointed out in Coupon et al. (2015), since each of these systematic errors affects the inference independently with different stellar mass and scale dependences and we fit all of them jointly, the combined impact of these systematic errors should be heavily mitigated by the self-consistency required in the model, therefore becoming subdominant compared to the statistical errors.

7 STELLAR MASS CONTENT WITHIN HALOES

In this section, we explore the implications of our inferred HOD parameters for the connection between stellar and halo mass.

7.1 Stellar mass functions

One of the most important consistency-checks for our fiducial iHOD model is to predict the *parent* SMFs at each redshift, and compare to the *observed* versions as well as the intrinsic SMFs estimated empirically, e.g. from the V_{max} method. As mentioned in Section 5.1, this check echoes the philosophy of the abundance matching

technique, but in practice is carried out in reverse. In particular, the model uses the observed two-point correlation function to infer the mapping from the stellar mass content, including the central and satellite galaxies, to the dark matter haloes, and the mapping in turn gives an exact prediction (via equation 17) for the parent SMF at each redshift – a ‘correlation’ matching as opposed to ‘abundance’ matching.

Fig. 9 presents the results of this consistency check on SMFs at six different redshifts. For each redshift, the upper sub-panel compares the observed (brown histogram), V_{max} -estimated (thin black line), and our inferred parent (thick black line with 1σ uncertainty band) SMFs. The parent SMF, predicted from the best-fitting iHOD model at that redshift, is also decomposed into the central (red dashed) and satellite (blue dot-dashed) contributions, except for the $z = 0.04$ panel where we instead show another measured SMF from Li & White (2009, shifted by 0.1 dex to remove the average bias between the two stellar mass estimators; see figure A1 in their paper). The central galaxies dominate the satellite galaxies in numbers across all stellar masses. The lower sub-panel shows the ratios of the observed (brown line with Poisson errors) and estimated SMFs (thin black line) over the parent one, i.e. the observed fraction f_{obs} . The redshift evolution of the predicted parent SMFs is solely from the cosmic evolution of the halo mass function, and the Panter et al. (2007) SMF is the same in all panels. Although the stellar mass estimates from Panter et al. (2007) were independently derived from the SDSS DR3 spectroscopy, the SMF (also used for the abundance matching in M10) provides a good fit to the observed SMF using the MPA/JHU stellar masses, indicating little systematic offset between the two estimators. Overall, at $M_* > 5 \times 10^{10} h^{-2} M_\odot$ where the completeness of SDSS is expected to be high, the predicted parent SMFs show remarkable consistency with the observed SMFs in both the shape and the amplitude at all redshifts. As mentioned in Section 5.1, the iHOD analysis is completely agnostic to the total number density of the galaxy catalogue (i.e. the normalization of the observed SMFs). Therefore, the agreement seen in Fig. 9 demonstrates that, in order to match the clustering and lensing signal measured in SDSS, the placement of galaxies within the dark matter haloes is uniquely determined so that the expected abundance of the galaxies at each stellar mass, translated from the halo mass function predicted by the ΛCDM Universe, automatically agrees with the observed galaxy SMFs in SDSS.

For the two lowest redshifts, the observed SMFs at the high-mass end are known to be notably incomplete due to two photometric confusions: (1) to avoid saturation and excessive phototalk in the spectrographs, some objects were rejected because they either have saturated centres (mostly bright active nuclei) or are blended with a saturated star; and (2) the image deblending software sometimes overdeblended bright galaxies with large angular extent (Strauss et al. 2002). A somewhat related issue is the oversubtraction of the sky background mentioned in Section 2.2, which results in suppressed flux estimates in bright galaxies. Therefore, limited to the $z < 0.05$ galaxies, the SMF measured by Li & White has a substantially lower amplitude at $M_* > 3 \times 10^{10} h^{-2} M_\odot$ than the Panter et al.’s SMF, the observed SMFs beyond $z > 0.08$, and the parent SMF predicted by our best-fitting model.

We next consider the f_{obs} inferred from the ratio of the V_{max} -estimated SMF to the predicted parent SMF in the lower sub-panels of Fig. 9. Starting with the low-mass end, f_{obs} is always above 60 per cent, and reaches 100 per cent at above $6 \times 10^{10} h^{-2} M_\odot$ until $2 \times 10^{11} h^{-2} M_\odot$, beyond which point the V_{max} method is expected

⁵ We also ignore the dispersion in the halo concentration at fixed M_h , which should have a negligible impact on the average signal calculation as long as it is not strongly correlated with satellite properties.

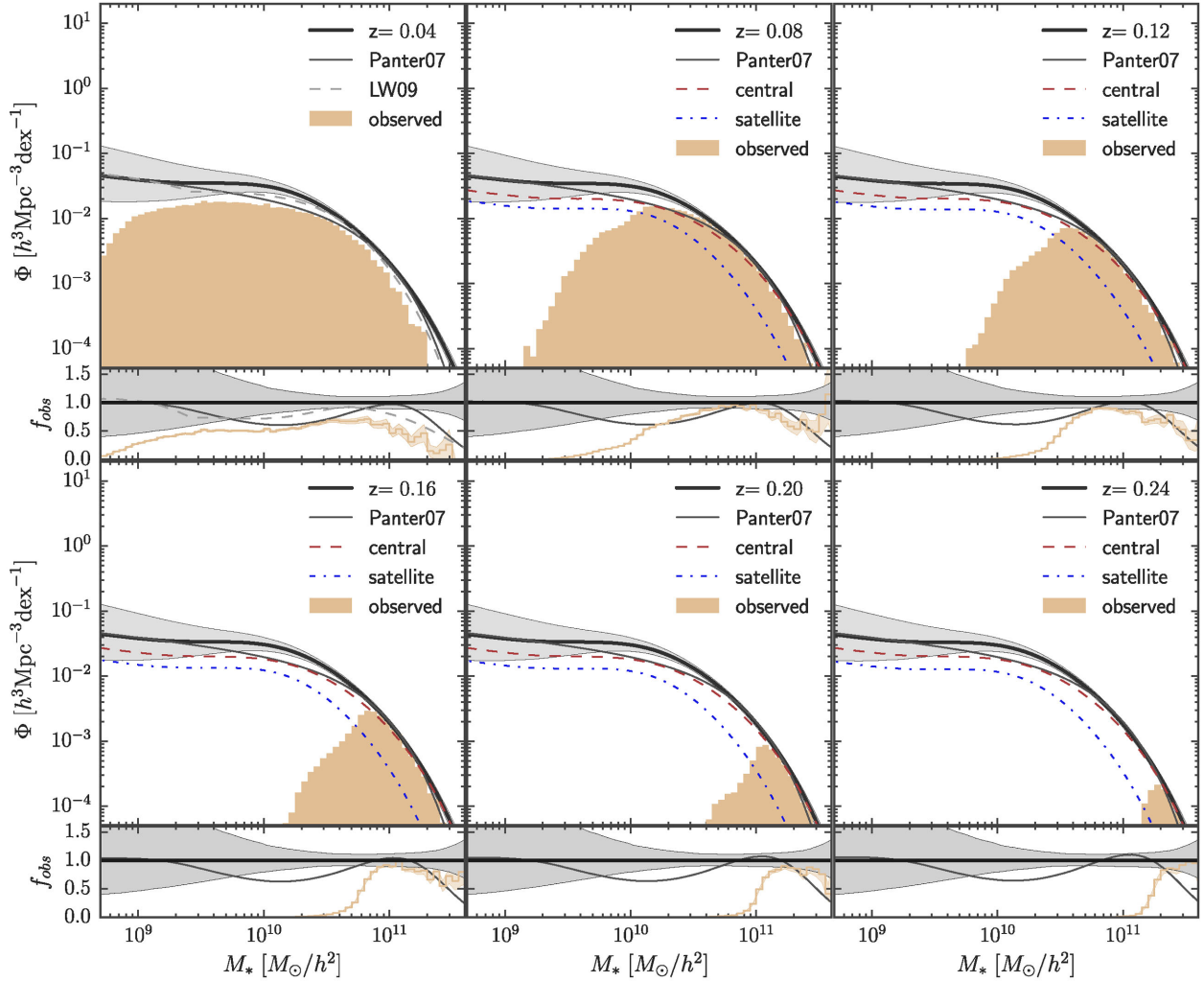


Figure 9. Comparison between the observed (brown histograms) and the parent (thick solid curves with 68 per cent confidence range shown in grey bands) galaxy SMFs predicted from the best-fitting \mathbf{iHOD} model at six different redshifts. In each upper sub-panel, the thin solid curve is the estimated SMF of SDSS DR3 galaxies from Panter et al. (2007). For the $z = 0.04$ panel, we show another SDSS SMF estimated by Li & White (2009) from the $z < 0.05$ galaxies as the thin dashed curve (after correction for the offset between the two stellar mass estimates); for the rest of the panels the predict parent SMF is decomposed into the contributions from the central and satellite galaxies, indicated by the red dashed and blue dot dashed lines, respectively. The lower sub-panels show the ratios of the Panter et al. (2007) (thin black) and the observed galaxy number density (brown) to the predicted SMF from our model. The redshift evolution of the predicted SMF in the model is coming solely from the cosmic evolution of the halo mass function.

to underestimate the galaxy abundance due to the aforementioned photometric issues. Meanwhile at the high-mass end, the \mathbf{iHOD} model predicts a much better match to the observed SMFs than the Panter et al's estimate does, reaching a constant f_{obs} of ~ 90 per cent, especially at the two highest redshifts where the photometric confusions that plague the lower redshifts are absent and the sample of very massive galaxies should be highly complete.

At the mass scale around $10^{10} h^{-2} M_{\odot}$, the f_{obs} inferred from the V_{max} -estimated SMF exhibits a dip at 60–70 per cent level in the lower sub-panels of Fig. 9. There are two main possible sources that could contribute to this dip, one theoretical and one observational. On the one hand, the dip could be a manifestation of the systematic uncertainty in the theoretical model. As seen from Fig. 1(a), this mass range marks the transition of galaxies from a predominantly blue, star-forming population below to a red, quiescent population above on the M_* - z diagram. Given the strong correlation between galaxy colour and M_*/L , we expect that the ansatz we

have made, i.e. assuming a weak dependence of $f_{\text{obs}}(M_*/L)$ on the host halo mass, to be the most fragile at this transitional mass scale. For example, if the M_*/L at fixed stellar mass is skewed higher in more massive haloes (Taylor et al. 2015), the observed $10^{10} h^{-2} M_{\odot}$ galaxies would preferentially reside in host haloes less massive than does a typical $10^{10} h^{-2} M_{\odot}$ galaxy in the Universe, and thus exhibit weaker clustering and g-g lensing signals than their stellar mass suggests. As a result of this masquerade effect, the \mathbf{iHOD} model would place them inside smaller haloes, resulting the overprediction of their abundances. It is worth noting that this theoretical systematic uncertainty is neither absent from the traditional HOD modelling, nor would it be eliminated by adding the SMF data as an extra constraint. On the other hand, the photometric issues that affect the targeting of bright galaxies would also impact this intermediate mass range, rendering the star-forming galaxy population underrepresented in the MGS at the low redshift. The selection effect is easily discernible in Fig. 1(a): at $M_* \sim 10^{10} h^{-2} M_{\odot}$, the

galaxies at redshifts lower than 0.04 have a redder average $g - r$ colour than those at higher redshifts. Since the majority of the detected galaxies with $M_* \sim 10^{10} h^{-2} M_\odot$ comes from $z < 0.08$, they carry heavy weight (large values of $1/V_{\max}$) in the Panter et al.'s and Li & White's SMFs. Therefore, the underrepresentation of these galaxies at low redshift translates to potentially significant underestimation of their abundances. In reality, it is more likely that both the theoretical and observational sources contribute to the dip at some level, and the true parent SMF should lie somewhere between the Panter et al.'s and our predicted curves.

7.2 Centrials: stellar mass to halo mass relation

The most crucial piece for mapping the stellar content back to dark matter haloes is the SHMR of the central galaxies. The top panel of Fig. 10 compares the SHMR and its logarithmic scatter derived from our fiducial iHOD constraints (black solid line with grey band) to that inferred from several methods, including abundance matching. To highlight the shrinkage in scatter at high halo mass predicted by our fiducial constraint, we also show two curves that delineate the boundaries of the constant scatter band with $\eta = 0$. To avoid clutter, here we only show the scatter for the SHMR from our constraint.

Let us first compare to the constraint from M10 (red dashed), which employed the same cosmology as our analysis and matches the halo abundance to the SDSS SMF measured by Panter et al. (2007), albeit with a different functional form for the SHMR and a fixed scatter. The two mean SHMRs show remarkable agreement over $\lg M_h \sim [12-14]$, or $M_* \sim 2 \times 10^{10} - 2 \times 10^{11} h^{-2} M_\odot$, where the data are the most constraining and the completeness is the highest. Above this mass range, the Moster et al.'s relation predicts higher stellar mass for central galaxies than our mean SHMR for a given halo mass, i.e. a steeper SHMR. Considering the lower amplitude of the V_{\max} -estimated SMF employed by the M10 analysis for abundance matching, naively one might expect that the observed massive galaxies have to be pushed to reside in progressively rarer and more massive haloes, making their SHMR shallower instead. However, due to the exponential decline of the halo mass function in this regime, the SHMR is much more sensitive to the change in scatter than the small change in the detection rate. This discrepancy on the high-mass end is mainly caused by the smaller scatter assumed in the M10 analysis ($\sigma_{\ln M_*} = 0.345$, or 0.15 in dex) compared to that inferred by our model ($\sigma_{\ln M_*} \sim 0.40$ at $M_h = 10^{15} h^{-1} M_\odot$). Their smaller scatter allows much fewer low-mass haloes to be considered as hosts for the massive galaxies, thus requiring a steeper slope in order to match the galaxy abundance.

The Behroozi et al. (2010) analysis derives the constraints on the SHMR by abundance matching to the SDSS SMF measured by Li & White (2009) at $z < 0.05$. They carefully accounted for the effect of scatter in the SHMR by varying both the scatters in the true stellar mass at fixed halo mass (intrinsic scatter) and in the observed stellar estimate at fixed true stellar mass (measurement scatter). For the intrinsic scatter, a lognormal prior centred at 0.16 dex with a width of 0.04 dex is placed (i.e. 0.37 ± 0.09 in natural log), and for the measurement scatter a fixed value of 0.07 dex is applied for the SDSS data. After abundance matching, the posterior constraint on the intrinsic scatter is $0.15^{+0.04}_{-0.02}$ dex, consistent with their input prior. However, as clearly shown by the top-left panel of Fig. 9, the Li & White (2009) SMF is underestimated at the high-mass end, thus shifting the mean SHMR horizontally towards higher halo mass. Behroozi et al. (2010) also adopts a slightly higher σ_8 than M10 and our study, which predicts more massive haloes

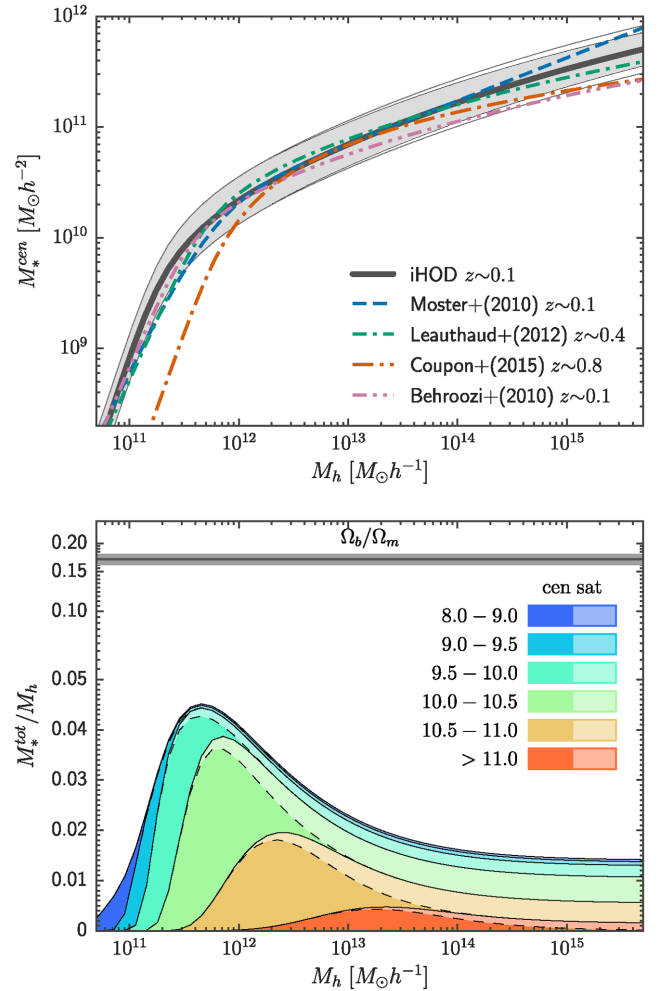


Figure 10. Top panel: comparison between the SHMRs inferred from abundance-matching based methods and our best-fitting iHOD model. The grey band shows the logarithmic scatter predicted by the best-fitting model, while the two enveloping thin curves represent the range of constant scatter with $\eta = 0$. Among the four external constraints used for comparison, the Moster et al. (2010) and the Behroozi et al. (2010) results are based on abundance-matching, while the other two are derived from fitting the clustering, lensing, and SMF jointly. Bottom panel: the total stellar mass to halo mass ratio as a function of halo mass. Coloured layers show the contribution from different stellar mass bins while separate portions of the same coloured layer indicate the relative contribution from centrals and satellites within each stellar mass bin. The horizontal band represents the cosmic baryon to dark matter ratio in the Universe. Note that the y-axis is logarithmic above 0.05.

and further pushes the slope of the mean SHMR shallower than our constraint.

Unlike abundance matching, our fiducial constraint is able to break the degeneracy between the scatter and the slope of the mean SHMR on the high-mass end without assuming any external priors. The leverage comes from the unique constraining power of combining clustering and lensing. On large scales, the galaxy clustering signal is proportional to $b_g^2 \xi_{\text{mm}}$, while the g - g lensing scales with $\Omega_m b_g \xi_{\text{mm}}$. On small scales, w_p is less a probe of the SHMR than $\Delta \Sigma$, which directly measures the average halo mass of individual galaxy samples. For a fixed cosmology, simultaneous fit to w_p and $\Delta \Sigma$ effectively constrains the clustering bias as a function of halo

mass, for which Λ CDM predicts a relatively steep slope on the high-mass end (see the seventh and eighth columns of Table 1). However, as shown by the open contours in the δ - $\sigma_{\ln M_*}$ panel in Fig. 7, the cHOD constraint stills shows substantial degeneracy between the slope parameter δ and the scatter, due to the inadequate S/N in the measurements of w_p and $\Delta\Sigma$. Thanks to the 84 per cent more galaxies used in the iHOD analysis, our fiducial analysis derives much tighter constraint with minimum residual degeneracy (filled contours).

The L12 result shown in Fig. 10 was derived from a slightly higher redshift range ($0.22 < z < 0.48$) than the MGS used by us. Since we have adopted a similar model framework and parametrization proposed in L12, the analysis of L12 is very similar to the cHOD analysis in this paper. However, their analysis adds the measured SMF as part of the constraining data set beside angular clustering and g-g lensing. In this regard, although allowing the scatter to vary freely, the constraint in L12 is somewhat more related to the abundance matching-based methods, because the smaller uncertainties in SMF easily dominate the likelihood of the fit. A straightforward comparison with the mean SHMR of L12 is complicated by the differences in the redshift range of the galaxy samples and the stellar mass estimators (although the COSMOS SMF below $z = 0.48$ agrees very well with that from Panter et al. 2007; see fig. 5 of L12), but the two constraints appear to be qualitatively consistent with each other above the characteristic halo mass. The Coupon et al. (2015) constraint is very similar to L12, derived from a joint lensing, clustering, and abundance analysis, but for galaxies at a much higher redshift ($z \sim 0.8$) and more massive than $10^{10} h^{-2} M_\odot$. Thus, the deviation between the Coupon et al. (2015) SHMR and ours at the low-mass end can be attributed to the evolution of galaxy population from $z = 0.8$ to 0.1. At the high end however, we do not see the steepening of the Coupon et al. (2015) SHMR at $z \sim 0.8$, as hinted by the L12 constraints from $z \sim 0.88$ (see fig. 11 in L12). The L12 and the Coupon et al. (2015) analyses also constrained the scatter in the SHMR, yielding $\sigma_{\ln M_*} \simeq 0.474$ (i.e. 0.206 dex) and $\sigma_{\ln M_*} \simeq 0.506$ (i.e. 0.22 dex; at $M_h \simeq 10^{12} h^{-1} M_\odot$), respectively, in excellent agreement with our constraint of $0.50^{+0.04}_{-0.03}$. This level of agreement is not necessarily expected, given that the scatter includes both intrinsic scatter and measurement uncertainty, the latter of which could differ for the three different samples studied in these papers.

Below the characteristic halo mass, our mean SHMR has a slightly higher amplitude than all the other three SHMRs, implying higher stellar-to-halo mass ratios for the low stellar mass galaxies. This difference echoes the discrepancy seen in the comparison between the parent and measured SMFs at the intermediate and low stellar mass scales. However, all those curves are still consistent with one another within the statistical uncertainties in the inferred mean SHMRs below $M_h \sim 10^{12} h^{-1} M_\odot$ (compare the difference to, e.g. the uncertainty band in the left-hand panel of Fig. 8).

The efficiency of stellar mass assembly at the centre of haloes, characterized by the central stellar-to-halo mass ratio M_*^{cen}/M_h , rises sharply as a function of halo mass at the low-mass scale, probably due to the increasing difficulty for the stellar feedback to drive cold gas out of the steepening gravitational potential wells of haloes (Hopkins, Quataert & Murray 2012). The growth of the central galaxies shifts down into a much lower gear beyond the characteristic mass $\sim 10^{12} h^{-1} M_\odot$, decreasing M_*^{cen}/M_h from its peak ~ 0.04 at $M_h \sim 10^{12} h^{-1} M_\odot$ to below 0.001 at $M_h \sim 10^{15} h^{-1} M_\odot$. AGN feedback is believed to be one of the major culprits that quench the star formation in those central galaxies within high-mass haloes.

However, as shown in Fig. 3, the satellite ‘cloud’ also begins to emerge within haloes above $10^{12} h^{-1} M_\odot$, and a good indicator for the efficiency of stellar mass assembly has to account for the stellar mass of satellites. To this end, the bottom panel of Fig. 10 presents a more complete picture by showing the total stellar mass assembly efficiency, M_*^{tot}/M_h (assuming $h = 0.72$), integrated over all the predicted parent galaxies above $10^8 h^{-2} M_\odot$. We also divide the total efficiency into contributions from six different stellar mass bins marked in the legend. Within the layer of each stellar mass bin, we further split the contribution into two components, the central (shaded colour) and the satellite (tinted colour), separated by a dashed line. For comparison, the horizontal line on top indicates the universal baryon-to-matter mass ratio. As expected, the total efficiency is dominated by the central galaxies until its peak at around $M_h \sim 10^{12} h^{-1} M_\odot$, then it levels off and asymptote to a constant value ~ 0.015 for galaxy groups and clusters, where the satellite contribution dominates. This asymptotic value of the total stellar mass fraction is consistent with observational studies by stacking galaxy catalogues or images inside samples of optical and/or X-ray clusters (Bahcall & Kulier 2014; Budzynski et al. 2014). Galaxies with M_* around $10^{10} h^{-2} M_\odot$ dominate the total stellar mass content within haloes of mass above $10^{12} h^{-1} M_\odot$.

The near-constant total stellar-to-halo mass ratio in group and cluster-size haloes is very intriguing. By comparing the integrated stellar mass to the stacked weak lensing signal at different radii within the MaxBCG clusters, Bahcall & Kulier (2014) found that the total mass of clusters can be largely accounted for by the total dark matter mass associated with all the subhaloes that host the satellite galaxies. Therefore, when the fractional contribution from the central galaxy is negligible,

$$\frac{\sum M_*}{M_h} \bigg|_{M_h} \simeq \left[\int \left(\frac{dN_{\text{sat}}}{d \lg M_*} \bigg|_{M_h} \right) \frac{M_*}{M_h} d \lg M_* \right] / N_{\text{sat}}(M_h), \quad (46)$$

where $(dN_{\text{sat}}/d \lg M_*)|_{M_h}$ is the satellite SMF conditioned at halo mass M_h , and M_h^{sub} is the subhalo mass. If the satellite stellar-to-subhalo mass relation M_*/M_h^{sub} were to roughly follow the SHMR of the centrals, as assumed by our subhalo lensing model and the SHAM methods, the aggregate of the satellite galaxy population inside each halo is likely to be a strong function of the halo mass when $dN_{\text{sat}}/d \lg M_*$ differs from halo to halo. In order for any two haloes with different masses to share the same total stellar-to-halo mass ratio, the SMFs within the two haloes have to be somewhat self-similar – all $dN_{\text{sat}}/d \lg M_*$ have the same shape near the knee of the SMF where most of the stellar mass is stored, with their amplitudes N_{sat} linearly proportional to the halo mass. To test this hypothesis, we explore the conditional SMFs of satellites below in Section 7.3.

The SHMR, specifying the mean logarithmic stellar mass of central galaxies at fixed halo mass, is a convenient choice from a theoretical modelling point of view, but observationally what is often measured is the opposite, i.e. the mean log-halo mass or halo mass at given stellar mass. For instance, as the predecessor of the study in this paper, Mandelbaum et al. (2006a) measured the average log-halo masses for a set of early and late-type galaxy stellar mass samples in the SDSS, via the traditional HOD modelling of their g-g lensing signals. To compare to the result from Mandelbaum et al. (2006a) and other similar studies in the literature, we compute the distribution of the halo mass at fixed stellar mass as

$$p(M_h|M_*^{\text{cen}}) = \frac{p(M_*^{\text{cen}}|M_h)p(M_h)}{p(M_*^{\text{cen}})}, \quad (47)$$

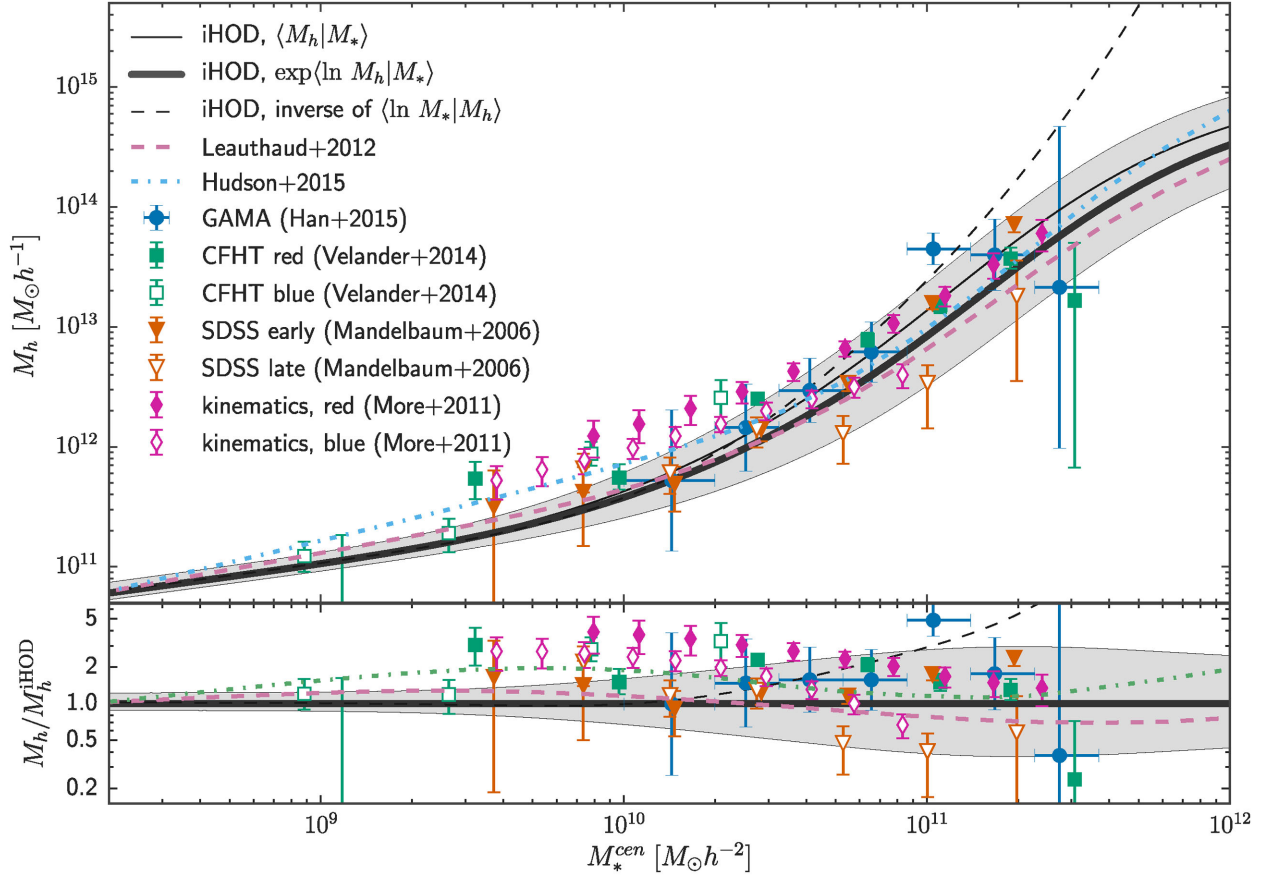


Figure 11. Comparison between the halo-to-stellar mass relations inferred from the best-fitting *iHOD* model and that from various other weak lensing measurements. Top panel: the dashed black curve is just the inverse function of the best-fitting SHMR, i.e. the same black curve in the top panel of Fig. 10 but with x - and y -axis swapped. The thick black curve with a grey band indicate the mean and scatter (not statistical uncertainty on the mean) of the distribution of logarithmic halo mass at fixed stellar mass, while the thin black curve shows the mean halo mass at fixed stellar mass. The magenta dashed and cyan dot-dashed curves are the results from L12 using COSMOS and Hudson et al. (2015) using CFHTLenS, while the blue circles are from the maximum-likelihood weak lensing analysis of GAMA galaxies. The green and the red squares are the average weak lensing masses measured for CFHTLenS and SDSS galaxies, respectively, and within each survey the relations are constrained for red/early-type and blue/late-type galaxies separately using nearly identical HOD prescriptions. The diamonds are measurements from satellite kinematics and are higher than other weak lensing measurements and our predictions on the lower mass end, though consistent on the high-mass end. All the uncertainties shown for the weak lensing mass measurements are 1σ . Bottom panel: similar to the top panel, but showing the ratio of each quantity over the prediction by the best-fitting *iHOD* model.

where $p(M_*^{cen}|M_h)$ is the SHMR of central galaxies (similar to equation 21) and $p(M_h)$ is the halo mass function normalized by the total number density of haloes, while $p(M_*^{cen})$ is proportional to the SMF of central galaxies, calculated from

$$p(M_*^{cen}) = \int_0^{+\infty} p(M_*^{cen}|M_h) p(M_h) dM_h. \quad (48)$$

The halo-to-stellar mass relation, $p(M_h|M_*^{cen})$, computed from the best-fitting SHMR and its scatter shown in the top panel of Fig. 10 via equations (47) and (48), is represented in Fig. 11 in two forms, the mean log-halo mass at fixed stellar mass (thick black solid curve),

$$\langle \ln M_h | M_* \rangle = \int p(M_h | M_*^{cen}) \ln M_h dM_h, \quad (49)$$

and the mean halo mass at fixed stellar mass (thin black solid curve),

$$\langle M_h | M_* \rangle = \int p(M_h | M_*^{cen}) M_h dM_h, \quad (50)$$

respectively. The difference between the two means are small at the low stellar mass end but grows to 0.1 dex at high masses, and

the grey band indicates the 1σ logarithmic scatter about the mean relation. As seen in Fig. 10, the original SHMR flattens on the high-mass end, so the log-scalar in halo mass at fixed stellar mass is much larger than on the low stellar mass end. The 1σ error on the mean relation (not shown here, but see Fig. 8) is comparable to the scatter at the lower stellar mass end, but much smaller than the scatter above $M_* \sim 10^{10} h^{-2} M_\odot$. Also shown in Fig. 11 is the black dashed curve from equation (19) using the best-fitting parameters, i.e. the same thick black curve shown in Fig. 10 except that the x - and y -axis are swapped. The difference between the black dashed and the two black solid curves can easily exceed 0.5 dex for high stellar mass samples, and reaches almost 1 dex for galaxies with $M_* > 5 \times 10^{11} h^{-2} M_\odot$. Therefore, caution should be exercised when trying to compare the theoretical SHMR with the direct observational estimates of halo masses for any galaxy samples selected above $M_* = 10^{10} h^{-2} M_\odot$. The magenta dashed and the cyan dot-dashed curves in Fig. 11 are the halo-to-stellar mass relations inferred from the L12 (COSMOS) and the Hudson et al. (2015) (CFHTLenS) analyses, respectively, showing great consistency between independent constraints from different surveys. The

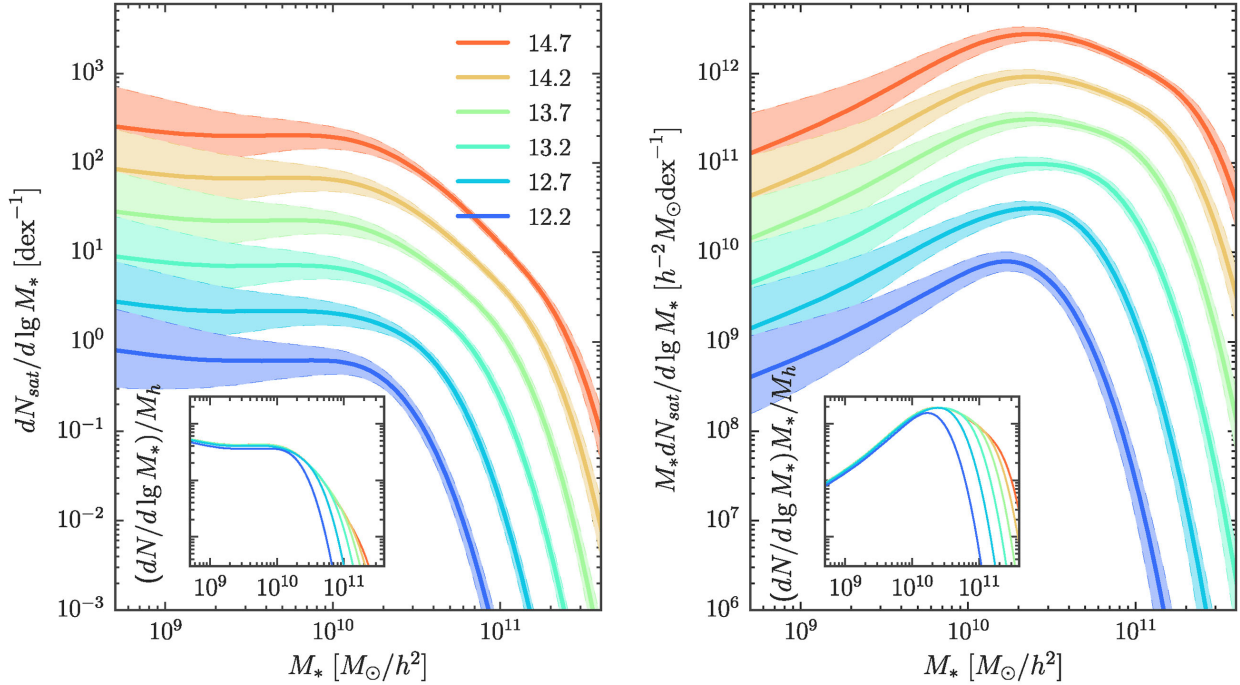


Figure 12. Left-hand panel: satellite SMF dN_{sat}/dM_* conditioned at six different halo masses marked on the top right. The shaded regions indicate the 68 per cent uncertainties derived from the iHOD constraint. The inset panel shows the same set of curves, with each amplitude re-scaled by the corresponding halo mass, highlighting the departure of the conditional SMF from the Schechter functional form in high mass haloes. Right-hand panel: similar to the left-hand panel, but for $M_* dN_{\text{sat}}/d \lg M_*$, the fractional contribution to the total stellar-to-halo mass ratio per dex in M_* . The satellite galaxies with stellar mass around a few times $10^{10} h^{-2} M_\odot$ contribute the most to the total stellar mass content in each halo.

circles and the squares indicate the average halo masses inferred from individual galaxy stellar mass samples using the weak lensing measurements in GAMA (Han et al. 2015) and CFHTLenS (Velander et al. 2014), respectively. The diamonds are the measurement from satellite kinematics (More et al. 2011), which agrees well with weak lensing measurements and our predictions, with slightly higher amplitude on the low-mass end. The lenses used in the Han et al. (2015) analysis were central galaxies of groups and clusters that are almost volume-limited, while for the latter two, the Velander et al. (2014) constraint was separately derived for colour-segregated samples that are flux-limited. Despite the differences in the methods and data sets, our relation is largely consistent with both sets of measurements. The triangles are the measurements from the earlier SDSS DR4 galaxies (Mandelbaum et al. 2006a), also split into subsamples of early- and late-type galaxies. Our analysis is a direct update from the Mandelbaum et al. (2006a) study, using SDSS DR7 (versus DR4) galaxies, updated MPA/JHU stellar mass catalogue, improved photometric redshifts for the source sample, additional measurements (galaxy clustering), and a more sophisticated HOD model, and the two constraints are fully consistent with each other.

Finally, to facilitate the comparison between our constraint and observational studies that measure the mean (log) halo mass of individual galaxy samples, we provide an analytic fitting formula to the best-fitting halo-to-stellar mass relation shown in Fig. 11 (computed from the fiducial iHOD MCMC chain via equation 49),

$$\langle \lg M_h | M_*^{\text{cen}} \rangle = 4.41 \left[1 + \exp(-1.82 (\lg M_*^{\text{cen}} - 11.18)) \right]^{-1} + 11.12 \sin(-0.12 (\lg M_*^{\text{cen}} - 23.37)), \quad (51)$$

where M_h and M_*^{cen} are in units of $h^{-1} M_\odot$ and $h^{-2} M_\odot$, respectively. The above fitting formula is accurate to within 0.15 per cent across the entire stellar mass range above $3 \times 10^8 h^{-2} M_\odot$.

7.3 Satellites: conditional stellar mass function

Instead of imposing some fixed functional form (e.g. the Schechter function) for the conditional stellar mass function (CSMF), we compute the satellite CSMFs as the derivatives of the satellite HODs at fixed halo mass. This model flexibility allows for a more thorough exploration of the shapes of the satellite CSMFs, especially any potential mass-dependent deviations from the Schechter function at both the low and high stellar mass ends. The left-hand panel of Fig. 12 shows the satellite CSMFs inferred for haloes of six different masses marked on the legend, with each shaded band indicating the 68 per cent uncertainty from the iHOD constraint. For low-mass haloes, the form of the CSMFs resemble the Schechter function with a flattening low M_* portion and a sharp exponential cutoff at high M_* , but it starts to exhibit an extra power-law portion at intermediate M_* for haloes more massive than $10^{13} h^{-1} M_\odot$. The inset panel shows the six CSMFs scaled by their corresponding halo mass. All the scaled CSMFs would land on top of one another if the satellite population inside haloes follows an exact homology sequence. Since α_{sat} is strongly constrained to be around unity ($\alpha_{\text{sat}} = 1.000^{+0.03}_{-0.02}$), the number of satellite galaxies above any stellar mass threshold scales linearly with halo mass above M_{cut} (see equation 22), leading to a self-similar behaviour across the low M_* range ($< 10^{10} h^{-2} M_\odot$). At high stellar mass range, however, the homology is broken and an excess population of galaxies of successively higher stellar mass begin to emerge as satellites – they used to be the central galaxies of small haloes that were later accreted by their current-day, more massive host haloes – a physical picture from the hierarchical structure formation of the Λ CDM Universe, recovered by the prediction of our fiducial model, a purely statistical framework that describes the clustering and lensing of galaxies.

To understand the asymptotic behaviour of the total stellar-to-halo mass ratio mentioned in Section 7.2, we show the stellar mass-weighted CSMFs, $M_* dN_{\text{sat}}/d \lg M_*$, in the right-hand panel of Fig. 12. The peaks in those curves confirm the finding from the bottom panel of Fig. 10 that galaxies of $M_* \sim 2 \times 10^{10} h^{-2} M_\odot$ contribute the most amount of stellar mass per dex in M_* to each halo. Again, in the inset panel we scale each quantity by the corresponding halo mass, showing the fractional contribution to the total stellar-to-halo mass ratio per dex in M_* for each halo mass. The asymptotic behaviour of the total stellar mass fraction can be mainly attributed to two factors: (1) the self-similarity of satellite galaxies below $10^{11} h^{-2} M_\odot$ ensures that the total ratio integrated up to $M_* \sim 10^{11} h^{-2} M_\odot$ is the same for all haloes above $5 \times 10^{13} h^{-1} M_\odot$; and (2) the satellite galaxies above $10^{11} h^{-2} M_\odot$ contribute negligibly to the total ratio, so any increase in halo mass brings little change to the total stellar-to-halo mass ratio.

8 CONCLUSIONS

8.1 Methodology

We have developed a novel extension to the HOD framework – namely the *i*HOD model – to solve for the mapping between the observed stellar mass distribution and the underlying dark matter haloes, via modelling the projected galaxy auto-correlation function w_p and the g - g lensing $\Delta\Sigma$ signals. In particular, the *i*HOD model has two main features.

(i) It is able to predict w_p and $\Delta\Sigma$ for all the galaxies above the mixture limit (i.e. the minimum stellar mass for the quiescent galaxies to be detected above the flux limit). This flexibility allows us to include $\simeq 84$ per cent more galaxies than the traditional HOD method, which substantially improves the S/N of the w_p and $\Delta\Sigma$ measurements (by 10 per cent to almost a factor of 2 depending on stellar mass).

(ii) It takes into account the volume incompleteness of galaxy stellar mass samples in a statistically consistent way, eliminating the need to assume completeness or parametrize the redshift-dependent selection functions.

The crucial input to the *i*HOD model is the *shape* of the observed SMF at each redshift, which is employed by the *i*HOD model to construct an HOD for the galaxies at that redshift slice using $p(M_*, M_h)$, the 2D joint probability density distribution of a galaxy with stellar mass M_* sitting in a halo of mass M_h . The clustering and lensing signals of any galaxy sample can then be calculated by combining the signals predicted from the HODs of the individual redshift slices.

The 2D distribution $p(M_*, M_h)$ has two components, the mean and scatter of the SHMR for the central galaxies and the global HOD for the satellite galaxies. We adopt a similar parametrization for the two components as L11, but allow the power-law slope of the satellite HOD (α) and the concentration ratio between the satellite and dark matter profiles (f_c) to vary freely during the constraint. Furthermore, we also re-calibrate the subhalo lensing model in our analysis against the state-of-the-art hydrodynamic simulation MB-II.

8.2 Findings from the SDSS

Thanks to the greatly improved S/N in the w_p and $\Delta\Sigma$ measurements, our fiducial *i*HOD analysis not only breaks the degeneracy between the slope and the scatter of the SHMR, therefore placing

stringent constraint on the link between the dark haloes and their central galaxies, but also derives the conditional SMF of satellite galaxies as a function of halo mass. The inferred SHMR is in good agreement with constraints from the abundance matching methods and the joint clustering, lensing, and galaxy abundance analyses using the L11 framework. We also detect a declining logarithmic scatter about the mean SHMR with halo mass, likely driven by the smaller external scatter in the stellar mass estimate.

For the SMF of satellite galaxies at fixed halo mass, the best-fitting model predicts a departure from the Schechter function in massive groups and clusters, probably a result of mergers and accretion that convert central galaxies of small haloes into satellites of more massive systems. It will be interesting to compare the conditional SMFs of satellites statistically derived from the *i*HOD model, which allows the satellites to have a different SHMR than the central galaxies, to that inferred from abundance matching, which directly assigns satellite stellar masses to subhaloes in the simulations assuming the same SHMR as the central galaxies. The average total number of satellites in each halo, however, is tightly constrained to follow a linear scaling relation with halo mass, while the distribution of those satellites is less concentrated than the underlying smooth distribution of dark matter.

In principle, the *i*HOD model uses only the shape of the observed galaxy SMF for signal calculation and is agnostic of the SMF amplitude – using half of the galaxies randomly drawn from the original catalogue would yield the same best-fitting model parameters, albeit with larger uncertainties. However, the parent SMF predicted from the best-fitting *i*HOD model agrees remarkably well with the observed SMF in both the shape and the amplitude. Since the amplitude of the predicted SMF is translated from the normalization of the halo mass function predicted by the Λ CDM cosmology, this agreement is highly non-trivial, demonstrating the efficacy of the *i*HOD model – in order to match the clustering and lensing signal measured in SDSS, the mapping between the galaxy stellar content and the dark matter haloes is uniquely determined so that the expected galaxy abundance at each M_* , calculated by summing galaxies at that M_* over all the dark matter haloes, automatically gives the observed galaxy SMF in SDSS.

8.3 Future applications

In future work, we will extend the *i*HOD model to provide constraints on the quenching mechanisms that transform galaxies from star-forming to quiescent. The most straightforward approach is to apply the same analysis to just the quenched population (i.e. the galaxies with $g - r > 0.8$ above the mixture limit), and compare the inferred $p_{\text{red}}(M_*, M_h)$ to the best-fitting $p(M_*, M_h)$ in this paper to predict the quenched fraction as a function of M_* and M_h (see e.g. Zehavi et al. 2005; Zehavi et al. 2011, for HOD studies of galaxies segregated by colour). Alternatively, one can do a joint fit for the red and blue galaxies simultaneously, assuming an ad hoc functional form for the quenched fraction (see e.g. Tinker et al. 2013; Rodríguez-Puebla et al. 2015). However, as mentioned in Section 6.3, the galaxy assembly bias needs to be carefully treated or marginalized over to obtain a robust constraint when modelling colour-selected samples.

Going beyond the goal of understanding galaxy formation, HOD modelling of galaxy populations is likely to have a role in cosmological analyses of future large imaging surveys that seek to constrain dark energy using weak lensing, such as the Large Synoptic Survey Telescope (LSST; LSST Science Collaborations & LSST Project 2009), *Euclid* (Laureijs et al. 2011), and *WFIRST* (Spergel

et al. 2015). While cosmological weak lensing was originally identified as a very clean probe of dark energy due to its sensitivity to matter fluctuations, more recent work has identified two major theoretical uncertainties: the effect of baryons on the matter power spectrum, and the intrinsic alignments of galaxy shapes that violate the assumption that any coherent galaxy alignments are due to weak lensing. Leading proposals for the mitigation of these theoretical uncertainties include halo modelling of the galaxy-shear cross-correlation (g - g lensing) in combination with galaxy clustering and cosmic shear, instead of using cosmic shear alone (e.g. Schneider & Bridle 2010; Semboloni et al. 2011; Semboloni, Hoekstra & Schaye 2013; Zentner et al. 2013). Thus, we anticipate that the i HOD model will be an important contribution to such efforts, given that it will enable the use of a larger galaxy sample in understanding these important contaminants to cosmic shear surveys, and thus enable the use of tighter priors on the nuisance parameters for baryonic effects and/or intrinsic alignments when extracting cosmological information from the cosmic shear signal.

As a general yet powerful statistical formalism, the i HOD model can be easily applied to galaxies at higher redshifts, such as the existing data sets from the COSMOS survey (Scoville et al. 2007) used by L12, CFHTLenS (Heymans et al. 2012), the recently finished Baryon Oscillation Spectroscopic Survey (BOSS; Eisenstein et al. 2011; Dawson et al. 2013) and its near-term higher redshift successor eBOSS, and the deeper surveys planned for future facilities such as the Dark Energy Spectroscopic Instrument (Levi et al. 2013), the Subaru Prime Focus Spectrograph (Takada et al. 2014), *Euclid*, and *WFIRST*. Galaxies at higher redshifts, such as luminous red galaxies (LRGs; Eisenstein et al. 2001; Anderson et al. 2012) and emission line galaxies (Comparat et al. 2015), are usually targeted with some complicated colour and flux cuts, so the galaxy samples are always volume-incomplete. For instance, Hoshino et al. (2015) found that the average number of LRG-type central galaxies in the massive haloes does not asymptotically reach unity, contradictory to the assumption in traditional HOD models (Parejko et al. 2013). Therefore, the i HOD model would be an especially valuable tool in constraining the link between galaxies and dark matter haloes at high redshifts.

ACKNOWLEDGEMENTS

We thank Neta Bahcall and Zheng Zheng for helpful discussions. We also thank Jiaxin Han for providing his weak lensing mass measurements and Surhud More for his satellite kinematics measurements. We thank Alexie Leauthaud and David Weinberg for carefully reading an earlier version of the manuscript and for giving detailed comments and suggestions that have greatly improved the manuscript. We also thank the MB-II team for providing the simulated galaxy catalogue and the dark matter particle data. YZ acknowledges the hospitality of the CCAPP at the Ohio State University where he enjoyed a fruitful discussion with participants of the ‘Assembly Bias and the Galaxy-Halo Relation’ Workshop. RM and YZ are supported by the Department of Energy Early Career Program, and the Sloan Fellowship programme.

REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
 Anderson L. et al., 2012, *MNRAS*, 427, 3435
 Bahcall N. A., Kulier A., 2014, *MNRAS*, 439, 2505
 Baldauf T., Smith R. E., Seljak U., Mandelbaum R., 2010, *Phys. Rev. D*, 81, 063531
 Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, *MNRAS*, 373, 469
 Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, 304, 15
 Baugh C. M., 2006, *Rep. Prog. Phys.*, 69, 3101
 Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 379
 Bell E. F., de Jong R. S., 2001, *ApJ*, 550, 212
 Benson A. J., 2010, *Phys. Rep.*, 495, 33
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
 Berlind A. A., Kazin E., Blanton M. R., Pueblas S., Scoccimarro R., Hogg D. W., 2006, preprint (astro-ph/0610524)
 Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, *MNRAS*, 436, 697
 Bhattacharya S., Heitmann K., White M., Lukić Z., Wagner C., Habib S., 2011, *ApJ*, 732, 122
 Blanton M. R., Berlind A. A., 2007, *ApJ*, 664, 791
 Blanton M. R. et al., 2005, *AJ*, 129, 2562
 Blazek J., Mandelbaum R., Seljak U., Nakajima R., 2012, *J. Cosmol. Astropart. Phys.*, 5, 41
 Bocquet S., Saro A., Dolag K., Mohr J. J., 2015, preprint (arXiv:1502.07357)
 Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, *ApJ*, 379, 440
 Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53
 Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
 Budzynski J. M., Koposov S. E., McCarthy I. G., McGee S. L., Belokurov V., 2012, *MNRAS*, 423, 104
 Budzynski J. M., Koposov S. E., McCarthy I. G., Belokurov V., 2014, *MNRAS*, 437, 1362
 Cacciato M., van den Bosch F. C., More S., Li R., Mo H. J., Yang X., 2009, *MNRAS*, 394, 929
 Cacciato M., van den Bosch F. C., More S., Mo H., Yang X., 2013, *MNRAS*, 430, 767
 Cappellari M. et al., 2012, *Nature*, 484, 485
 Cen R., Ostriker J. P., 1999, *ApJ*, 514, 1
 Chabrier G., 2003, *PASP*, 115, 763
 Cole S., Kaiser N., 1989, *MNRAS*, 237, 1127
 Comparat J. et al., 2015, *A&A*, 575, A40
 Conroy C., 2013, *ARA&A*, 51, 393
 Conroy C., van Dokkum P. G., 2012, *ApJ*, 760, 71
 Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
 Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
 Coupon J. et al., 2015, *MNRAS*, 449, 1352
 Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, *AJ*, 112, 839
 Crocce M., Fosalba P., Castander F. J., Gaztañaga E., 2010, *MNRAS*, 403, 1353
 Croft R. A. C., Matteo T. D., Khandai N., Springel V., Jana A., Gardner J. P., 2012, *MNRAS*, 425, 2766
 Croton D. J., Gao L., White S. D. M., 2007, *MNRAS*, 374, 1303
 Cui W., Borgani S., Dolag K., Murante G., Tornatore L., 2012a, *MNRAS*, 423, 2279
 Cui W., Baldi M., Borgani S., 2012b, *MNRAS*, 424, 993
 Cusworth S. J., Kay S. T., Battye R. A., Thomas P. A., 2014, *MNRAS*, 439, 2485
 Dalal N., White M., Bond J. R., Shirokov A., 2008, *ApJ*, 687, 12
 Davé R., Finlator K., Oppenheimer B. D., 2012, *MNRAS*, 421, 98
 Dawson K. S. et al., 2013, *AJ*, 145, 10
 De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2
 De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, *MNRAS*, 366, 499
 Di Matteo T., Springel V., Hernquist L., 2005, *Nature*, 433, 604
 Dutton A. A., Macciò A. V., 2014, *MNRAS*, 441, 3359
 Einasto J., 1965, *Tr. Astrofiz. Inst. Alma-Ata*, 5, 87
 Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
 Eisenstein D. J. et al., 2001, *AJ*, 122, 2267
 Eisenstein D. J. et al., 2011, *AJ*, 142, 72
 Faltenbacher A., White S. D. M., 2010, *ApJ*, 708, 469
 Faucher-Giguère C.-A., Kereš D., Ma C.-P., 2011, *MNRAS*, 417, 2982

- Feldmann R. et al., 2006, *MNRAS*, 372, 565
- Ferreras I., La Barbera F., de la Rosa I. G., Vazdekis A., de Carvalho R. R., Falcón-Barroso J., Ricciardelli E., 2013, *MNRAS*, 429, L15
- Fontanot F., De Lucia G., Monaco P., Somerville R. S., Santini P., 2009, *MNRAS*, 397, 1776
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, 111, 1748
- Fukugita M., Hogan C. J., Peebles P. J. E., 1998, *ApJ*, 503, 518
- Gallazzi A., Bell E. F., 2009, *ApJS*, 185, 253
- Gao L., White S. D. M., 2007, *MNRAS*, 377, L5
- Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66
- Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, *MNRAS*, 387, 536
- Geller M. J., Huchra J. P., 1989, *Science*, 246, 897
- George M. R. et al., 2011, *ApJ*, 742, 125
- George M. R. et al., 2012, *ApJ*, 757, 2
- Giocoli C., Bartelmann M., Sheth R. K., Cacciato M., 2010, *MNRAS*, 408, 300
- Gunn J. E. et al., 1998, *AJ*, 116, 3040
- Gunn J. E. et al., 2006, *AJ*, 131, 2332
- Guo Q., White S., Li C., Boylan-Kolchin M., 2010, *MNRAS*, 404, 1111
- Guo Q. et al., 2011, *MNRAS*, 413, 101
- Guo H., Zehavi I., Zheng Z., 2012, *ApJ*, 756, 127
- Guo H. et al., 2014, *MNRAS*, 441, 2398
- Guzik J., Seljak U., 2001, *MNRAS*, 321, 439
- Guzik J., Seljak U., 2002, *MNRAS*, 335, 311
- Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, *MNRAS*, 375, 489
- Hamilton A. J. S., Tegmark M., 2004, *MNRAS*, 349, 115
- Han J. et al., 2015, *MNRAS*, 446, 1356
- Hansen S. M., Sheldon E. S., Wechsler R. H., Koester B. P., 2009, *ApJ*, 699, 1333
- Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, *MNRAS*, 367, 1039
- Hearin A. P., Watson D. F., 2013, *MNRAS*, 435, 1313
- Hearin A. P., Watson D. F., van den Bosch F. C., 2015, *MNRAS*, 452, 1958
- Hernquist L., Katz N., 1989, *ApJS*, 70, 419
- Heymans C. et al., 2012, *MNRAS*, 427, 146
- Hirata C., Seljak U., 2003, *MNRAS*, 343, 459
- Hopkins P. F., Quataert E., Murray N., 2012, *MNRAS*, 421, 3522
- Hoshino H. et al., 2015, *MNRAS*, 452, 998
- Hudson M. J. et al., 2015, *MNRAS*, 447, 298
- Ichiki K., Takada M., 2012, *Phys. Rev. D*, 85, 063521
- Jing Y. P., Mo H. J., Börner G., 1998, *ApJ*, 494, 1
- Jing Y. P., Suto Y., Mo H. J., 2007, *ApJ*, 657, 664
- Kaiser N., 1984, *ApJ*, 284, L9
- Kang X., Jing Y. P., Mo H. J., Börner G., 2005, *ApJ*, 631, 21
- Katz N., Weinberg D. H., Hernquist L., 1996, *ApJS*, 105, 19
- Kauffmann G. et al., 2003, *MNRAS*, 341, 33
- Kereš D., Katz N., Weinberg D. H., Davé R., 2005, *MNRAS*, 363, 2
- Kerscher M., Szapudi I., Szalay A. S., 2000, *ApJ*, 535, L13
- Khandai N., Di Matteo T., Croft R., Wilkins S. M., Feng Y., Tucker E., DeGraf C., Liu M., 2015, *MNRAS*, 450, 1349
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Kulier A., Ostriker J. P., 2015, *MNRAS*, 452, 4013
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Lang D., Hogg D. W., Mierle K., Blanton M., Roweis S., 2010, *AJ*, 139, 1782
- Laporte C. F. P., White S. D. M., 2015, *MNRAS*, 451, 1177
- Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, *ApJ*, 738, 45 (L11)
- Leauthaud A. et al., 2012a, *ApJ*, 744, 159 (L12)
- Leauthaud A. et al., 2012b, *ApJ*, 746, 95
- Levi M. et al., 2013, preprint ([arXiv:1308.0847](https://arxiv.org/abs/1308.0847))
- Li C., White S. D. M., 2009, *MNRAS*, 398, 2177
- Li Y., Mo H. J., Gao L., 2008, *MNRAS*, 389, 1419
- Li R. et al., 2014, *MNRAS*, 438, 2864
- LoVerde M., 2014, *Phys. Rev. D*, 90, 083530
- LSST Science Collaborations, LSST Project, 2009, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))
- McGaugh S. S., Schombert J. M., de Blok W. J. G., Zagursky M. J., 2010, *ApJ*, 708, L14
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Ma C.-P., Fry J. N., 2000, *ApJ*, 543, 503
- Mandelbaum R. et al., 2005a, *MNRAS*, 361, 1287
- Mandelbaum R., Tasitsiomi A., Seljak U., Kravtsov A. V., Wechsler R. H., 2005b, *MNRAS*, 362, 1451
- Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006a, *MNRAS*, 368, 715
- Mandelbaum R., Seljak U., Cool R. J., Blanton M., Hirata C. M., Brinkmann J., 2006b, *MNRAS*, 372, 758
- Mandelbaum R., Hirata C. M., Leauthaud A., Massey R. J., Rhodes J., 2012, *MNRAS*, 420, 1518
- Mandelbaum R., Slosar A., Baldauf T., Seljak U., Hirata C. M., Nakajima R., Reyes R., Smith R. E., 2013, *MNRAS*, 432, 1544
- Miyatake H. et al., 2015a, *ApJ*, 806, 1
- Miyatake H. et al., 2015b, preprint ([arXiv:1506.06135](https://arxiv.org/abs/1506.06135))
- Mo H. J., Jing Y. P., White S. D. M., 1996, *MNRAS*, 282, 1096
- More S., van den Bosch F. C., Cacciato M., Skibba R., Mo H. J., Yang X., 2011, *MNRAS*, 410, 210
- More S., Miyatake H., Mandelbaum R., Takada M., Spergel D., Brownstein J., Schneider D. P., 2015, *ApJ*, 806, 2
- Moster B. P., Somerville R. S., Maudtsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, *ApJ*, 710, 903 (M10)
- Murray S. G., Power C., Robotham A. S. G., 2013, *MNRAS*, 434, L61
- Nakajima R., Mandelbaum R., Seljak U., Cohn J. D., Reyes R., Cool R., 2012, *MNRAS*, 420, 3240
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Newman A. B., Treu T., Ellis R. S., Sand D. J., Nipoti C., Richard J., Jullo E., 2013, *ApJ*, 765, 24
- Newman A. B., Ellis R. S., Treu T., 2015, preprint ([arXiv:1503.05282](https://arxiv.org/abs/1503.05282))
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Norman M. L., Bryan G. L., 1999, in Miyama S. M., Tomisaka K., Hanawa T., eds, *Astrophysics and Space Science Library*, Vol. 240, Numerical Astrophysics. Kluwer, Dordrecht, p. 19
- O'Shea B. W., Bryan G., Bordner J., Norman M. L., Abel T., Harkness R., Kritsuk A., 2004, preprint ([astro-ph/0403044](https://arxiv.org/abs/astro-ph/0403044))
- Oppenheimer B. D., Davé R., 2006, *MNRAS*, 373, 1265
- Oppenheimer B. D., Davé R., 2008, *MNRAS*, 387, 577
- Padmanabhan N. et al., 2008, *ApJ*, 674, 1217
- Panther B., Jimenez R., Heavens A. F., Charlot S., 2007, *MNRAS*, 378, 1550
- Parejko J. K. et al., 2013, *MNRAS*, 429, 98
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Peebles P. J. E., Hauser M. G., 1974, *ApJS*, 28, 19
- Peng Y.-j., Lilly S. J., Renzini A., Carollo M., 2012, *ApJ*, 757, 4
- Planck Collaboration XIII, 2015, preprint ([arXiv:1502.01589](https://arxiv.org/abs/1502.01589))
- Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, *MNRAS*, 420, 3206
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Putman M. E., Peek J. E. G., Jong M. R., 2012, *ARA&A*, 50, 491
- Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, *ApJ*, 771, 30
- Reyes R., Mandelbaum R., Gunn J. E., Nakajima R., Seljak U., Hirata C. M., 2012, *MNRAS*, 425, 2610
- Rodríguez-Puebla A., Avila-Reese V., Yang X., Foucaud S., Drory N., Jing Y. P., 2015, *ApJ*, 799, 130
- Salim S. et al., 2007, *ApJS*, 173, 267
- Schmidt F., Rozo E., Dodelson S., Hui L., Sheldon E., 2009, *Phys. Rev. Lett.*, 103, 051301
- Schneider M. D., Bridle S., 2010, *MNRAS*, 402, 2127
- Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, *ApJ*, 546, 20
- Scoville N. et al., 2007, *ApJS*, 172, 1

- Seljak U., 2000, MNRAS, 318, 203
- Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, MNRAS, 417, 2020
- Semboloni E., Hoekstra H., Schaye J., 2013, MNRAS, 434, 148
- Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, ApJ, 643, 14
- Shankar F. et al., 2014, ApJ, 797, L27
- Sheldon E. S. et al., 2004, AJ, 127, 2544
- Sheth R. K., Jain B., 2003, MNRAS, 345, 529
- Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
- Sheth R. K., Tormen G., 2004, MNRAS, 350, 1385
- Shull J. M., Smith B. D., Danforth C. W., 2012, ApJ, 759, 23
- Simet M., Mandelbaum R., 2015, MNRAS, 449, 1259
- Singh S., Mandelbaum R., More S., 2015, MNRAS, 450, 2195
- Smith R. E. et al., 2003, MNRAS, 341, 1311
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481
- Spergel D. et al., 2015, preprint ([arXiv:1503.03757](https://arxiv.org/abs/1503.03757))
- Springel V. et al., 2005, Nature, 435, 629
- Strauss M. A. et al., 2002, AJ, 124, 1810
- Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, MNRAS, 387, 1391
- Takada M. et al., 2014, PASJ, 66, 1
- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
- Taylor E. N. et al., 2015, MNRAS, 446, 2144
- Teyssier R., 2002, A&A, 385, 337
- Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, ApJ, 631, 41
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
- Tinker J. L., Leauthaud A., Bundy K., George M. R., Behroozi P., Massey R., Rhodes J., Wechsler R. H., 2013, ApJ, 778, 93
- Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
- van Daalen M. P., Schaye J., 2015, MNRAS, 452, 2247
- van den Bosch F. C., Tormen G., Giocoli C., 2005, MNRAS, 359, 1029
- van den Bosch F. C. et al., 2007, MNRAS, 376, 841
- van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, MNRAS, 430, 725
- van Dokkum P. G., Conroy C., 2010, Nature, 468, 940
- Velander M. et al., 2014, MNRAS, 437, 2111
- Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Dalla Vecchia C., 2014, MNRAS, 442, 2641
- Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031
- Vogelsberger M. et al., 2014, Nature, 509, 177
- von der Linden A., Best P. N., Kauffmann G., White S. D. M., 2007, MNRAS, 379, 867
- Wang H. Y., Mo H. J., Jing Y. P., 2007, MNRAS, 375, 633
- Wang L., Weinmann S. M., De Lucia G., Yang X., 2013, MNRAS, 433, 515
- Wang W., Sales L. V., Henriques B. M. B., White S. D. M., 2014, MNRAS, 442, 1363
- Watson D. F., Berlind A. A., McBride C. K., Hogg D. W., Jiang T., 2012, ApJ, 749, 83
- Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230
- Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71
- Weinberg D. H., Davé R., Katz N., Hernquist L., 2004, ApJ, 601, 1
- Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
- Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153
- Yang X., Mo H. J., van den Bosch F. C., 2009, ApJ, 695, 900
- Yoo J., Seljak U., 2012, Phys. Rev. D, 86, 083504
- Yoo J., Tinker J. L., Weinberg D. H., Zheng Z., Katz N., Davé R., 2006, ApJ, 652, 26
- York D. G. et al., 2000, AJ, 120, 1579
- Zehavi I. et al., 2005, ApJ, 630, 1
- Zehavi I. et al., 2011, ApJ, 736, 59
- Zentner A. R., Semboloni E., Dodelson S., Eifler T., Krause E., Hearin A. P., 2013, Phys. Rev. D, 87, 043509
- Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044
- Zhang Y., Zhang P., Yang X., Cui W., 2013, Phys. Rev. D, 87, 023521
- Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, ApJ, 707, 354
- Zheng Z., Weinberg D. H., 2007, ApJ, 659, 1
- Zheng Z. et al., 2005, ApJ, 633, 791
- Zhu G., Zheng Z., Lin W. P., Jing Y. P., Kang X., Gao L., 2006, ApJ, 639, L5
- Zu Y., Zheng Z., Zhu G., Jing Y. P., 2008, ApJ, 686, 41
- Zu Y., Weinberg D. H., Rozo E., Sheldon E. S., Tinker J. L., Becker M. R., 2014, MNRAS, 439, 1628

APPENDIX A: SMALL-SCALE LENSING CUTOFFS

In this appendix, we describe our choice of minimum scale for modelling of the lensing signals. Our estimator for $\Delta\Sigma$, equation (6), is designed such that our signal includes a correction for the zero shear contributed by ‘source’ galaxies that are at the lens redshift. This correction is often written separately as a ‘boost factor’ (e.g. Sheldon et al. 2004; Mandelbaum et al. 2005a),

$$B(r_p) = \frac{\sum_{ls} w_{ls}(r_p)}{\sum_{rs} w_{rs}(r_p)}, \quad (\text{A1})$$

where the summed weights are over lens–source and random lens–source pairs. $B(r_p) = 1$ if there are no galaxies associated with the lens that are included as sources and if lensing magnification is negligible. Generally, since both clustering and magnification are large at small r_p , $B(r_p)$ is a monotonically decreasing function.

As shown in Mandelbaum et al. (2005a), small-scale systematics due to deblending and sky subtraction can appear as an *increase* in $B(r_p)$ on the smallest scales, followed by the expected decrease on scales where those systematics no longer operate. In the decreasing region, we know that source detection is no longer 100 per cent effective, but what we do not know is the actual level of inefficiency, since $B(r_p)$ gives a single constraint on two completely generate quantities (amount of contamination by physically associated galaxies, and inefficiency in the detection of both real sources and physically associated ones). To properly correct for physically associated sources, we would need to know both of these quantities. We would also have to know how the shear estimates and photo- z estimates of the sources in those regions may have been modified due to the systematic that is causing difficulties with source detection. Due to the lack of sufficient information to model the signal on scales where $B(r_p)$ indicates small-scale systematics, we do not plot or attempt to model $\Delta\Sigma$ there. The quoted minimum scales used for the lensing modelling comes from this consideration.

There are two additional systematics that could, in principle, operate on small scales above our minimum cutoff. The first systematic is the intrinsic alignment of galaxy shapes, pointing coherently in the radial direction with respect to the lens. While intrinsic alignments of bright red galaxies are well-established in the literature (e.g. most recently, Singh, Mandelbaum & More 2015), attempts to measure any radial alignments of faint galaxies in mixed blue and red source populations with respect to the positions of nearby bright galaxies have thus far only resulted in null detections (Blazek et al. 2012) with relatively tight upper limits. Hence we do not consider this to be an important systematic for this work.

Lensing magnification has also been considered as a systematic since, if present, it means that our normalization by $\sum_{rs} w_{rs}$ to correct for physically associated sources is incorrect. As discussed in

Mandelbaum et al. (2005a) and Schmidt et al. (2009), the amount of magnification for a flux- and size-selected sample depends on both the slope of the number counts near the flux limit and the slope of the apparent size distribution near its limit. Moreover, it depends on the slopes of these distributions weighted by whatever per-object weights are used for the lensing analysis. Simet & Mandelbaum (2015) calculated the appropriate slopes for the source catalogue used in this paper, and based on the numbers in their table 1, the ratio of observed number counts with magnification to the unmagnified number counts is very near 1 ($n_{\text{obs}}/n \approx 1-0.03\kappa$). Hence,

even though the best-fitting model for our eight stellar mass samples have convergences that range from 0.0014 to 0.063 at $0.1 h^{-1} \text{Mpc}$, the maximum effect of magnification on the observed counts and therefore the size of the systematic error in the boost factor and lensing calibration at that scale is -0.2 per cent. We therefore neglect this source of error.

This paper has been typeset from a \LaTeX file prepared by the author.