

Galaxy assembly bias: a significant source of systematic error in the galaxy–halo relationship

Andrew R. Zentner,^{1★} Andrew P. Hearin² and Frank C. van den Bosch³

¹Department of Physics and Astronomy & Pittsburgh Particle physics, Astrophysics and Cosmology Center (PITT PACC), University of Pittsburgh, Pittsburgh, PA 15260, USA

²Fermilab Center for Particle Astrophysics, Fermi National Accelerator Laboratory, Batavia, IL 60510-0500, USA

³Department of Astronomy, Yale University, PO Box 208101, New Haven, CT 06520-8101, USA

Accepted 2014 July 8. Received 2014 July 8; in original form 2013 November 18

ABSTRACT

Methods that exploit galaxy clustering to constrain the galaxy–halo relationship, such as the halo occupation distribution (HOD) and conditional luminosity function (CLF), assume halo mass alone suffices to determine a halo’s galaxy content. Yet, halo clustering strength depends upon properties other than mass, such as formation time, an effect known as *assembly bias*. If galaxy characteristics are correlated with these auxiliary halo properties, the basic assumption of standard HOD/CLF methods is violated. We estimate the potential for assembly bias to induce systematic errors in inferred halo occupation statistics. We construct realistic mock galaxy catalogues that exhibit assembly bias as well as companion mock catalogues with identical HODs, but with assembly bias removed. We fit HODs to the galaxy clustering in each catalogue. In the absence of assembly bias, the inferred HODs describe the true HODs well, validating the methodology. However, in all cases *with* assembly bias, the inferred HODs exhibit *significant systematic errors*. We conclude that the galaxy–halo relationship inferred from galaxy clustering is subject to significant systematic errors induced by assembly bias. Efforts to model and/or constrain assembly bias should be priorities as assembly bias is a threatening source of systematic error in galaxy evolution and precision cosmology studies.

Key words: galaxies: evolution – galaxies: haloes – cosmology: theory – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Theoretical models connecting galaxies to dark matter haloes unlock the predictive power of cosmological N -body simulations. The two most widely used models of the galaxy–halo connection are the halo occupation distribution (HOD, e.g. Berlind & Weinberg 2002; Zheng et al. 2005) and the conditional luminosity function (CLF, e.g. Yang, Mo & van den Bosch 2003; van den Bosch et al. 2013). The central quantity in the HOD is $P(N|M)$, the probability that a halo of mass M hosts N galaxies brighter than some luminosity threshold. The CLF instead models $\Phi(L|M)$, the mean abundance of galaxies of luminosity L that reside in a dark matter halo of mass M . These formalisms are closely related: an HOD can be derived by integrating the CLF against luminosity; a CLF can be derived by differentiating the HOD with respect to luminosity. Both formalisms have been studied extensively to constrain the galaxy–halo connection (an incomplete list of recent examples includes Magliocchetti & Porciani 2003; Yang et al. 2005; Zehavi et al. 2005, 2011; van den

Bosch et al. 2007; Zheng, Coil & Zehavi 2007; Simon et al. 2009; Skibba & Sheth 2009; Zheng et al. 2009; Ross, Percival & Brunner 2010; Geach et al. 2012; Parejko et al. 2013) as well as cosmology (Tinker et al. 2005; Leauthaud et al. 2012; Cacciato et al. 2013; Mandelbaum et al. 2013; More et al. 2013).

All conventional formulations of both the HOD and the CLF assume that galaxy occupation statistics are governed exclusively by the masses of the dark matter haloes hosting the galaxies of interest. We will refer to any such formulation of the HOD or CLF formalisms as the ‘standard’ approach henceforth. In this paper, we explore a class of simple, but well-motivated models for the galaxy–halo connection in which the assumption that halo statistics depend upon mass alone is violated, and demonstrate that the degree to which such violations lead to systematic errors in the inferred relationship between galaxies and haloes can be significant.

That halo mass should be the halo property that most strongly influences the properties of the galaxies within them is now widely accepted and has significant and long-standing theoretical support (e.g. White & Rees 1978; Blumenthal et al. 1984). Related to this, it has also long been recognized that halo mass is the halo property that most strongly influences halo abundance and halo clustering. The

★E-mail: zentner@pitt.edu

theoretical underpinnings of this fact lie in the uncorrelated nature of the random walks describing halo assembly and clustering in the simplest implementations of the excursion set formalism (Press & Schechter 1974; Bond et al. 1991; Lacey & Cole 1993; Mo & White 1996; Zentner 2007). Yet, the spatial distribution of dark matter haloes in dissipationless N -body simulations depends not only on halo mass, but also on additional properties such as formation time (Gao, Springel & White 2005; Wechsler et al. 2006; Gao & White 2007; Li, Mo & Gao 2008). Moreover, the lack of any correlation between halo environment and formation time is an assumption of the simplest excursion set models, rather than a derived property of haloes. Relaxing this assumption, excursion set theory also predicts a correlation between halo formation time and halo environment (e.g. Zentner 2007; Dalal et al. 2008). The dependence of the spatial distribution of dark matter haloes upon properties besides mass is generically referred to as ‘halo assembly bias’, and is typically quantified in terms of the halo two-point correlation function.

The character of assembly bias depends on halo mass relative to M_* , the characteristic collapse mass at a given redshift: for haloes at fixed mass $M \ll M_*$, older haloes cluster more strongly than young (Gao & White 2007); for haloes with $M_{\text{vir}} \gg M_*$, the reverse is true (Wechsler et al. 2006). The physical explanation for this trend was laid out in Zentner (2007) and Dalal et al. (2008). At the high-mass end, assembly bias is expected purely from the statistics of peaks in the initial density field. At the low-mass end, halo assembly bias arises due to the cessation of mass accretion on to haloes residing in dense environments (see also Wang, Mo & Jing 2009; Lacerna & Padilla 2011). This correlation between halo formation time z_{form} and environment suggests that other halo properties that are correlated with z_{form} will also exhibit ‘assembly bias’ trends, including concentration, triaxiality, spin, and velocity anisotropy. Indeed, this is the case (Wechsler et al. 2006; Faltenbacher & White 2010; Lacerna & Padilla 2012). If the properties of the galaxies that reside in a halo are correlated with any of these properties that are known to be correlated with halo assembly, then the standard HOD/CLF assumption will be violated and these models will not be able to predict the clustering statistics of galaxies correctly. Of course, this violation could be sufficiently weak as to be of little practical importance, but whether or not this is the case remains an open question.

Croton, Gao & White (2007) studied the effect of assembly bias in semi-analytic models (SAMs) of galaxy formation on three-dimensional galaxy clustering, finding that assembly bias is an important ingredient in determining the clustering strengths of galaxies. In their models, the effect of assembly bias depends non-trivially on galaxy luminosity and colour. Disconcertingly, Croton et al. (2007) find that the assembly bias effects they detect cannot be accounted for with either halo formation time or concentration alone. This suggests that galaxy properties may depend upon the assembly histories of haloes in a sufficiently complicated manner as to make empirical modelling extremely challenging.

Observational investigations of assembly bias in the galaxy distribution have produced mixed results. Yang, Mo & van den Bosch (2006) studied the cross-correlation between galaxies and galaxy groups in the Two Degree Field Galaxy Redshift Survey (Colless et al. 2001), and found that for groups of the same mass, the correlation strength depends on the star formation rate (SFR) of the central galaxy: at fixed mass, the clustering strength of galaxy groups decreases as the SFR of the central galaxy increases. Wang et al. (2008, 2013) confirmed these findings using much larger data sets obtained from the Sloan Digital Sky Survey (SDSS), showing that the colour-dependence is more prominent in less massive groups,

and demonstrating that these results are consistent with predictions from semi-analytical models. These studies suggest that assembly bias may be present in the observed galaxy distribution at a statistically significant level.¹ On the other hand, Blanton & Berlind (2007) use a different technique as well as an alternative SDSS group catalogue, and found little, if any, evidence for assembly bias on large scales, and only a modest signal on small scales ($r \lesssim 300 h^{-1}$ Mpc). Similarly, Tinker et al. (2008a) has shown that HOD models fit to galaxy clustering measurements predict void statistics in good agreement with the data, providing independent support for the standard HOD assumption that halo mass is the only relevant property that determines galaxy occupation statistics.

Assembly bias is a generic prediction of two related classes of models for the galaxy–halo connection that enjoy significant success in reproducing a wide variety of observed galaxy statistics. The first is the widely used *abundance-matching* technique (Kravtsov et al. 2004; Tasitsiomi et al. 2004; Vale & Ostriker 2004, 2006; Conroy & Wechsler 2009; Guo et al. 2010; Neistein et al. 2011; Rodríguez-Puebla, Drory & Avila-Reese 2012; Simha et al. 2012; Watson, Berlind & Zentner 2012; Kravtsov 2013). In this approach, one assumes that every (sub)halo in the universe hosts a single galaxy at its centre, and that there is a monotonic relation between a halo property (usually something that can serve as a proxy for halo size or potential well depth, such as the maximum circular velocity V_{max}) and the luminosity (or stellar mass) of the galaxy it hosts. Abundance matching using V_{max} has been shown to predict accurately a wide range of statistics of the observed galaxy distribution, including the two-point projected correlation function of galaxies at both low- and high redshifts (Conroy, Wechsler & Kravtsov 2006), the conditional stellar mass function (Hearin et al. 2013b; Reddick et al. 2013), magnitude gap statistics (Hearin et al. 2013a), and galaxy–galaxy lensing statistics (Hearin et al. 2013b). As we will demonstrate explicitly in this paper, *assembly bias is a generic prediction of any abundance-matching technique predicated upon halo circular velocity*. Broadly speaking, this is because halo mass alone does not suffice to specify the halo velocity profile, and halo profiles are correlated with assembly (Gao et al. 2005; Wechsler et al. 2006; Dalal, Lithwick & Kuhlen 2010).

The second class of galaxy–halo models we study in this paper is the recently developed *age distribution matching* (Hearin & Watson 2013). In age matching, galaxy colour at fixed luminosity (or stellar mass) is assumed to be in monotonic correspondence with halo age at fixed V_{max} . As was shown by Hearin & Watson (2013), age matching predicts the observed colour-dependence of galaxy clustering as well as the scaling between galaxy colour and host halo mass remarkably well. In a follow-up study, Hearin et al. (2013b) showed that age matching also provides a good description of the excess surface mass density about galaxies as a function of galaxy colour, as measured from galaxy–galaxy lensing in the SDSS. Age matching explicitly correlates galaxy colour with halo age, so it should not be surprising that age matching naturally predicts galaxy assembly bias. A variety of definitions of halo age exist in the literature and in this paper, we use the same definition of halo age as Hearin & Watson (2013), which we reiterate in Section 2.1.3.

In this paper, we address the following question: To what degree might assembly bias threaten the ability of standard HOD models to draw unbiased inferences and/or make unbiased predictions for the galaxy–halo connection? We focus on the statistic that is

¹ See also Cooper et al. (2010) for a reported detection of assembly bias that does not employ a group-finder.

most often modelled using these techniques, namely, the projected two-point correlation function of galaxies. We take the empirical successes of abundance matching and age matching as motivation to use these models as bases for our assessment. To be clear, there are known weaknesses of these models (e.g. Hearin et al. 2013a,b; Hearin & Watson 2013; Reddick et al. 2013) and they certainly do not provide a complete description of the galaxy–halo connection. However, these models are simple to use, contain assembly bias in a transparent manner, and describe observed galaxy clustering reasonably well. Thus, our approach complements that taken in Pujol & Gaztañaga (2014), who instead study a variety of SAMs which violate the simple assumptions of the standard HOD due to the numerous complex baryonic processes that SAMs parametrize.

We proceed by fitting the projected two-point functions of abundance- and age-matching mock galaxy catalogues to a standard HOD model. We compare these to fits of the two-point function in mock galaxy catalogues with *identical* true HODs, but built to have *no assembly bias*. The degree to which the inferred HODs differ from the true HODs can be used to assess the potential importance of assembly bias (as well as provides an important validation exercise for HOD-based inferences).

We find that reasonable levels of assembly bias in the galaxy population can lead to statistically significant systematic errors in the galaxy–halo connection inferred using standard HOD techniques. This is true for luminosity threshold samples, and quite dramatic for colour-selected sub-samples. Moreover, we show that these biases induce systematic errors in predictions for independent quantities made using HOD parameters inferred from galaxy clustering. Finally, we demonstrate that an independent statistic used previously to diagnose assembly bias, namely the void probability function (VPF; Tinker, Weinberg & Warren 2006), is relatively insensitive to the assembly bias present in abundance/age-matching mock galaxy catalogues. These results suggest that

- (i) inferences drawn regarding the galaxy–halo connection from galaxy clustering should include a significant, and previously neglected, systematic error in their error budgets, and
- (ii) incorporating assembly bias effects into HOD/CLF-like models should be a priority henceforth, including for (re)analyses of existing data sets such as SDSS.

Indeed, as we discuss below, the mock galaxy catalogues that we study necessarily contain fewer galaxies within smaller effective volumes than either existing or forthcoming observational samples. Therefore, the galaxy two-point functions from our mock catalogues exhibit larger errors than observational samples, suggesting that observational samples may be subject to a systematic error from assembly bias that is even more statistically significant than those that we present in this paper.

Our paper is organized as follows. In Section 2, we describe our methods. These include the construction of fiducial mock galaxy distributions based on abundance matching and age matching and our methods for fitting HODs to mock galaxy catalogues. In Section 3, we demonstrate the importance of assembly bias in abundance matching and explain our algorithm for erasing the assembly bias from mock galaxy catalogues based on abundance/age-matching models. We present our results in Section 4. In Section 5, we give examples of quantities that can be predicted (perhaps erroneously) using HODs while neglecting assembly bias and we demonstrate that assembly bias large enough to affect HOD inferences is not easy to diagnose using void statistics. We discuss the implications of our findings in Section 6. We conclude in Section 7 with a summary of our primary results.

2 METHODS

2.1 Mock galaxy catalogues

2.1.1 The Bolshoi simulation and haloes

The bedrock of all of the mock galaxy catalogues used in this study is the high-resolution, collisionless N -body Bolshoi simulation (Klypin, Trujillo-Gomez & Primack 2011). The simulation is based on a Λ cold dark matter cosmological model with $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $\Omega_b = 0.042$, $h = 0.7$, $\sigma_8 = 0.82$, and $n_s = 0.95$. Bolshoi tracks 2048^3 particles in a periodic box with side length $250 h^{-1}$ Mpc, has a particle mass of $m_p \simeq 1.9 \times 10^8 M_\odot$, and a force resolution of $\epsilon = 1 h^{-1}$ kpc. The simulation was run with the Adaptive Refinement Tree Code (ART; Kravtsov, Klypin & Khokhlov 1997; Gottloeber & Klypin 2008). Snapshots and halo catalogues are available at <http://www.multidark.org>. We refer the reader to Riebe et al. (2013) for additional information.

Our mock galaxy sample is based on the ROCKSTAR merger trees and halo catalogues. ROCKSTAR is a phase-space, temporal halo finder capable of resolving Bolshoi haloes and sub-haloes down to $V_{\max} \sim 55 \text{ km s}^{-1}$ (Behroozi et al. 2013a,b). These catalogues are publicly available and can be found at <http://hipacc.ucsc.edu/Bolshoi/MergerTrees.html>.

The dark matter haloes in the redshift-zero catalogues are defined to be regions within which the average density is $\Delta_{\text{vir}} \simeq 360$ times the mean matter density of the Universe when centred on the local density peak. This is often called the ‘virial’ criterion and the mass defined in this way is often referred to as the ‘virial mass’. The virial criterion implies that the relationship between the virial mass and the virial radius of a halo is $M = 4\pi\Omega_m\rho_{\text{crit}}\Delta_{\text{vir}}R_{\text{vir}}^3/3$. ‘Sub-haloes’ are distinct, bound structures within the virial radii of still larger ‘parent’ or ‘host’ haloes. We consider a halo to be a sub-halo of a larger host halo if the density peak on which the sub-halo is centred resides within the virial radius of a more massive halo. The structures of sub-haloes are typically strongly affected by the potentials of their host haloes. Consequently, the masses and radii of sub-haloes do not follow the virial definition given above.

2.1.2 Luminosity-only mock galaxies: abundance matching

We begin building mock catalogues by assigning galaxies of particular luminosities to haloes and sub-haloes. In particular, absolute magnitudes in the r band are assigned to Bolshoi (sub)haloes using the prevalent abundance-matching algorithm. We employ the same implementation of abundance matching described in detail in appendix of Hearin et al. (2013a). In this section, we merely sketch the basic features of this method.

The halo property $V_{\max} \equiv \text{Max} \{ \sqrt{GM(<r)/r} \}$, where $M(<r)$ is the mass enclosed within a distance r of the halo centre, defines the maximum circular velocity of a test particle orbiting in the halo’s gravitational potential well. The abundance-matching technique requires that the cumulative abundance of SDSS galaxies brighter than luminosity² L , $n_g(>L)$, is equal to the cumulative abundance of (sub)haloes with circular velocities larger than V_{\max} , $n_h(>V_{\max})$. This assumption specifies a monotonic relationship between luminosity and V_{\max} , enabling us to assign a unique r -band magnitude to every (sub)halo in the simulation.

In our implementation of abundance matching, we use the halo property V_{peak} , the peak value that V_{\max} obtains throughout the entire

² Typically in the r band as in our study.

assembly history of the halo (Reddick et al. 2013). Our model for the stochasticity in the brightnesses of mock galaxies results in uniform scatter in luminosity of ~ 0.15 dex at fixed V_{peak} ; due to scatter between M_{vir} and V_{peak} , our model has ~ 0.18 dex of scatter in luminosity at fixed M_{vir} . This amount of scatter is in accord with results from satellite kinematics (More et al. 2009) and other abundance-matching studies (Hearin et al. 2013a; Reddick et al. 2013).

2.1.3 Age matching: mock galaxies with luminosities and colours

In addition to assigning luminosities to mock galaxies, we also assign galaxies in our mock catalogues $g - r$ colours in order to study colour-dependent clustering in mock galaxy catalogues. For our mock galaxy samples with both r -band luminosity and $g - r$ colour, we use the publicly available mock catalogue based on ‘age distribution matching’, which can be found at <http://logrus.uchicago.edu/~aphearin>. We refer the interested reader to Hearin & Watson (2013) for a detailed presentation of the age distribution matching technique and Hearin et al. (2013b) for additional applications. In what follows, we offer a brief review of the methods used to construct this catalogue.

Luminosities are first assigned to dark matter haloes by abundance matching as described above in Section 2.1.2. After the luminosity assignments have been made for each galaxy, these galaxies are then assigned $g - r$ colours. This is accomplished by enforcing a monotonic relation between a proxy for halo age and galaxy colour at fixed luminosity. To be more specific, we assign to each halo a formation redshift, z_{starve} , which is designed to mimic the redshift at which the gas supply to the galaxy within the halo was cut off and new star formation began to be suppressed. Operationally, z_{starve} is the maximum of (1) the highest redshift at which the halo mass exceeded $10^{12} h^{-1} M_{\odot}$, (2) the redshift at which the halo was accreted on to another, larger halo and thus became a sub-halo, and (3) the halo formation redshift defined according to the fitting function of Wechsler et al. (2002). Subsequently, all mock galaxies are placed in narrow bins of r -band luminosity and rank ordered by z_{starve} , a proxy for halo age. Colour assignments are made by drawing from the observed colour distribution at fixed luminosity $P_{\text{SDSS}}(g - r | L_r)$, in such a way that the redder galaxies at fixed galaxy luminosity are placed within the older haloes (higher z_{starve}) and the observed colour distribution is imposed upon the mock galaxy sample. As was shown in Hearin & Watson (2013), the resulting mock galaxy distribution exhibits good agreement with the luminosity- and colour-dependent two-point clustering measurements made by the SDSS Zehavi et al. (2011). Moreover, the observed, colour-dependent galaxy clustering and galaxy–galaxy lensing statistics are also describe well by mock galaxy catalogues constructed using this algorithm Hearin et al. (2013b). In the interest of brevity, we will refer to this technique as *age matching* for the remainder of this paper.

2.2 The halo model and the HOD

The primary aim of this paper is to investigate the potential threat of assembly bias to standard HOD parameter inference when fitting models to observed clustering statistics. To do so, we treat the mock galaxy samples described in Section 2.1 as if they were the true universe. Mock catalogues are necessary because we must know the correct answers in order to assess our analysis methods. Moreover, we must be able to construct mock catalogues with

varying levels of assembly bias in order to attribute systematic differences to assembly bias. There are a number of reasons that we utilize the abundance-matching and age-matching mock galaxy catalogues. First, the two-point functions of these catalogues faithfully represent the clustering observed in SDSS and we can have some reason to assume that such catalogues may exhibit at least some of the complexity of the true galaxy population. Secondly, abundance-matching and age-matching catalogues are easy to construct and manipulate. Lastly, and most importantly, these models are the simplest algorithms for assigning galaxies to haloes in a way that exhibits assembly bias. The standard HOD formalism is predicated upon the assumption that assembly bias is zero, so by fitting our mock galaxy distributions with HOD models we can study how the inferred parameters must adjust to compensate for the presence of assembly bias in the mock catalogues.

We fit HOD parameters in a manner aimed at emulating the types of analyses that have been applied to data. In particular, we fit HODs to projected two-point correlation functions using the same halo model formulation described in Tinker et al. (2012). We use the halo mass function from Tinker et al. (2008b) and the halo bias of Tinker et al. (2010) to describe the statistics of haloes. This is augmented by the prescription for scale-dependent halo bias in Tinker et al. (2005). This implementation of the HOD-formalism in conjunction with the halo model formalism, or slight modifications thereof, is relatively standard and similar models have been used in numerous studies including the completed SDSS galaxy clustering analysis of Zehavi et al. (2011) as well as Zehavi et al. (2005), Yang et al. (2005), Zheng et al. (2007), van den Bosch et al. (2007), Zheng et al. (2009), Simon et al. (2009), Abbas et al. (2010), Ross et al. (2010), Watson et al. (2010), Matsuoka et al. (2011), Miyaji et al. (2011), Leauthaud et al. (2011), Leauthaud et al. (2012), Tinker et al. (2012), Geach et al. (2012), Kayo & Oguri (2012), van den Bosch et al. (2013), Tinker et al. (2013), Parejko et al. (2013), and Cacciato et al. (2013), to name some of the recent contributions to this extensive literature.

As described in the introductory section, a standard HOD model is defined by $P(N|M)$, the probability for a halo of mass M to contain N galaxies of the type chosen for analysis (e.g. selected by luminosity, colour, etc.). Keeping with the widely used convention, we describe central galaxies separately from satellite galaxies by assuming that

$$P(N|M) = P(N_{\text{cen}}|M) + P(N_{\text{sat}}|M).$$

For the first moment of $P(N_{\text{cen}}|M)$, we take the mean occupation of central galaxies in haloes of mass M to be

$$\langle N_{\text{cen}} \rangle = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log M - \log M_{\text{min}}}{\sigma_{\log M}} \right) \right]. \quad (1)$$

The scale M_{min} describes the halo mass above which you are likely to have a central galaxy ($\langle N_{\text{cen}} \rangle = 1/2$ at $M = M_{\text{min}}$). Below M_{min} the average number of central galaxies approaches zero with decreasing mass while above M_{min} the number of central galaxies trends asymptotically towards unity. The parameter $\sigma_{\log M}$ determines the sharpness of the transition between $\langle N_{\text{cen}} \rangle = 0$ at low mass and $\langle N_{\text{cen}} \rangle = 1$ at high mass. We describe the HOD of satellite galaxies as a Poisson distribution with mean

$$\langle N_{\text{sat}} \rangle = \left(\frac{M}{M_1} \right)^{\alpha} \exp \left(-\frac{M_{\text{cut}}}{M} \right). \quad (2)$$

The mass scale M_1 is the halo mass scale at which haloes have one satellite galaxy on average. At larger masses, the satellite number increases as a power law of halo mass M , with index α , and

the satellite occupation power law is truncated below masses of M_{cut} . In the interest of simplicity, we present results in which the satellites are distributed about central galaxies following a standard Navarro, Frenk & White (1997) profile with concentrations fixed to 60 per cent of the best-fitting average dark matter concentrations in the Bolshoi simulation (Klypin et al. 2011). This value is a good description of the distributions of satellite galaxies in our mock catalogues and mimics the treatment of satellites in the majority of the preceding literature. We show examples of the manner in which our results change when the concentration parameters describing the satellite distributions are allowed to vary in Appendix A.

In order to fit our mock galaxy data, we fix the cosmological parameters to be identical to the parameters of the Bolshoi simulation and vary the HOD parameters $\log(\sigma_{\log M}/h^{-1} M_{\odot})$, $\log(M_1/h^{-1} M_{\odot})$, $\log(M_{\text{cut}}/h^{-1} M_{\odot})$, and α . The parameter M_{min} is a derived parameter that guarantees the galaxy sample has the correct mean number density, given the remaining HOD parameters. We limit $\sigma_{\log M} > 10^{-3}$, and we caution the reader that marginalized posteriors on individual parameters are sensitive to the allowed range of $\sigma_{\log M}$, but none of our qualitative conclusions are sensitive to this choice. As suggested by Tinker et al. (2012), we also include a multiplicative parameter f_b , which is the ratio of the halo bias that we use to predict the galaxy clustering to the bias predicted by the formula of Tinker et al. (2010). The motivation for introducing f_b is that it can partially account for the fact that the halo bias and the scale-dependence of the halo bias are only imperfectly calibrated from simulations. Following, Tinker et al. (2012), we place a Gaussian prior on the halo bias parameter with $f_b = 1.0 \pm 0.15$. In the case of fitting colour-selected samples, we fit both red and blue samples *simultaneously* and require that a halo have no more than one central galaxy. We enforce this constraint by allowing the red samples to have a central galaxy HOD described by equation (1) above and restricting the blue central galaxy HOD to be the minimum of equation (1) and $1 - \langle N_{\text{cen}}^{\text{red}} \rangle$. Thus, the blue and red sub-samples have *distinct* values for each of the halo model parameters, but these parameters are related through the above constraint. We sample the parameter space using a standard Metropolis–Hastings Monte Carlo Markov Chain (MCMC) procedure. We compute the errors from the mock galaxy samples themselves using jackknife resampling with 25 sub-samples and compute χ^2 in the standard manner using the full covariance matrix derived from the jackknife procedure.

We have explored numerous modifications to our baseline fits. In particular, we have tried dropping the assumption of Poisson statistics for the satellite galaxies and varying separately the concentration of the parameter of the spatial distributions of satellites and find that our qualitative conclusions are insensitive to these assumptions. This is in large part because these parameters alter only small-scale clustering (projected separations below $r_p \lesssim 400 h^{-1} \text{ kpc}$), while the differences that we describe are on large scales, and because the projected two-point correlation function receives a significant contribution from galaxies with relatively large three-dimensional separations (see Section 3.4). We have tested to ensure that neither anisotropy in the distribution of satellite galaxies around their hosts nor the peculiar velocities of satellite galaxies alter our conclusions in any significant manner. To make contact with previous literature, we address briefly both the influences of the parameter f_b and of varying the satellite galaxy spatial distribution in Appendix A. Finally, we have verified that our halo model implementation agrees with implementation used to generate halo model predictions of galaxy clustering in the recent publications of van den Bosch et al. (2013) as well as Tinker et al. (2012) and Reddick et al. (2014) to better than ~ 1 per cent (Tinker & Reddick, private communication).

and, furthermore, that we recover consistent HOD parameters (Reddick, private communication).

Generally, we expect some extant and a number of forthcoming data sets to be significantly *more* sensitive to assembly bias than the HOD fits that we present in Section 4 suggest. Consequently, results derived from observational data may exhibit markedly more significant systematic errors than those that we quote later in this paper. The published SDSS data from Zehavi et al. (2011) have errors that are more than a factor of 4 smaller than the errors that we derive from the Bolshoi simulation on scales $\gtrsim 1 h^{-1} \text{ Mpc}$ for the $M_r < -21$ samples (this factor is approximately $\sim 2-3$ for $M_r < -20$ and $\sim 1.15-2$ for $M_r < -19$). Moreover, in most previous analyses, no nuisance parameter analogous to f_b was marginalized over in order to account for the limited calibration of halo clustering formulas. Taken together, our results in Section 4 strongly suggest that extant inferences drawn from observational data are subject to a significant systematic error associated with the unknown true level of assembly bias.

3 ASSEMBLY BIAS

The simple HODs described in Section 2.2 presume that halo mass is the only halo property that influences the number of galaxies residing in a (host) halo. If this is the case, then it is only necessary to enumerate the properties (e.g. abundance, structure, clustering, etc.) of haloes as a function of their masses, averaging over all other halo properties, and compute galaxy clustering using the standard halo model. The term *assembly bias* is often used broadly to refer to the dependence of host halo clustering on a property other than halo mass. It may be useful to consider this the *assembly bias of haloes*. In so far as the HOD is independent of these other properties, galaxy clustering statistics are unaltered by the assembly bias of haloes.³

What is relevant for galaxy clustering studies is the dependence of the HOD on any property $x \neq M$ upon which halo clustering *also* depends. It may be useful to refer to this as the *assembly bias of galaxies*. Mathematically, galaxy assembly bias is non-zero if and only if there exists some halo property x such that $P(N|M, x) \neq P(N|M)$ and the clustering of host haloes depends upon x at fixed halo mass. If there exists such a property, then $P(N|M, x)$ may place galaxies in haloes that are more or less strongly clustered than the average halo of mass M , and the standard halo model will fail to describe the galaxy clustering correctly.⁴ This definition of assembly bias includes the case in which the galaxy population within a halo depends only on a single halo property y other than mass, $P(N|y)$ (y may be a function of basic halo properties such as mass, concentration, formation time, etc.). In this paper, we estimate how poorly the standard halo model will do for a reasonable example of a mock galaxy sample that exhibits galaxy assembly bias explicitly.

3.1 Assembly bias in abundance matching

While this has not been emphasized in any of the previous work on the subject, non-zero assembly bias is a generic prediction of

³ Note that this may not be the case for matter clustering, as may be probed by gravitational lensing or peculiar velocity statistics, for example. Matter clustering depends upon halo properties such as concentration and shape directly.

⁴ Strictly speaking, knowledge of the x -dependence of only the first two moments of $P(N|M, x)$ is sufficient to determine the impact of galaxy assembly bias on two-point galaxy clustering.

all contemporary sub-halo abundance-matching techniques. In fact, there are at least two distinct effects that are important to consider:

- (i) At fixed mass, halo clustering is known to depend upon halo concentration, c (e.g. Gao et al. 2005; Wechsler et al. 2006; Gao & White 2007; Zentner 2007; Li et al. 2008; Wang et al. 2009; Lacerna & Padilla 2011). Haloes with higher c will have larger circular velocities, V_{\max} . Therefore, abundance matching is more likely to place galaxies of a given luminosity in higher concentration, rather than lower concentration, haloes of the same mass.
- (ii) At fixed mass, host haloes with larger c contain *fewer* sub-haloes (e.g. Zentner et al. 2005) and therefore these hosts contain fewer satellites than host haloes of the same mass with lower concentrations.

Other effects may also be important as well, such as the correlation of the spatial distributions of satellite galaxies within host haloes and host halo environment, including the triaxiality of the satellite galaxy distribution and the alignment of this triaxiality among nearby host halo pairs.

Effect (i) is a consequence of the variety of halo profiles at fixed mass and results in haloes of higher concentration hosting brighter galaxies at fixed mass. For any mock galaxy distribution defined by a luminosity threshold determined by a V_{\max} -based abundance-matching model, it follows that the sample will be biased to contain more highly concentrated haloes relative to a mass threshold sample.

Effect (ii) can be understood in terms of merging and tidal disruption of sub-haloes (see, for example, fig. 14 of Zentner et al. 2005). Halo concentration correlates with formation time so the sense of this effect is intuitive: more highly concentrated haloes assembled their masses earlier, thereby leaving more time for processes such as dynamical friction to deplete their sub-halo populations (see also van den Bosch, Tormen & Giocoli 2005). Moreover, the magnitude of this effect is not small. For example, in our fiducial $M_r < -19$ mock galaxy catalogue, if the population of central galaxies in haloes with virial masses $M \sim 10^{12} M_{\odot}$ is split on halo concentration, central galaxies that reside in haloes in the bottom half of the concentration distribution have *over twice as many satellites* as their higher concentration counterparts. We emphasize that this is a genuine feature of sub-structure content, and not simply due to the effect of sub-haloes on NFW fits: trends of similar magnitude obtain when centrals are instead split on other host halo properties such as halo formation time and environment density.

3.2 Assembly bias in age matching

Age matching is the name of the algorithm that we use to assign $g - r$ colours to galaxies as described briefly in Section 2. Age matching is predicated upon a very simple assumption: older haloes host galaxies with older stellar populations. In Hearin & Watson (2013), halo age is quantified by the property z_{starve} , so that in the language of the HOD, $P(N_{\text{red}}|M, z_{\text{starve}}) \neq P(N_{\text{red}}|M)$, and likewise for N_{blue} . Halo clustering is known to depend upon halo age, so again age matching explicitly introduces galaxy assembly bias into mock catalogues.

3.3 Erasing assembly bias

In order to assess the significance of assembly bias, it is necessary to construct mock catalogues that do not exhibit assembly bias effects, but which have the same $P(N|M)$ as our fiducial catalogues. This isolates the effects of assembly bias from effects due to changes in the HOD. For any model of the galaxy–halo connection, it is

possible to construct new mock catalogues without assembly bias while preserving the exact HOD. We describe our algorithm for erasing assembly bias presently.

First, we divide the *central* galaxies in a given sample into bins of halo mass M . We use 50 logarithmically spaced mass bins spanning $11.5 \leq \log M/M_{\odot} \leq 15$, corresponding to a bin width of 0.07 dex, to minimize effects due to finite binning.⁵ We then assign each central galaxy to a new, randomly selected host halo in the same mass bin, including as candidates those haloes that did not originally host a central galaxy. This erases the memory our mock central galaxies have of all halo properties besides mass, while leaving $\langle N_{\text{cen}}(M) \rangle$ fixed by construction. Since this randomization does not alter satellite populations, this step also leaves the satellite occupation $P(N_{\text{sat}}|M)$ unchanged for all luminosity thresholds.

Next, we assign each *system* of satellite galaxies to a new, randomly selected host halo of the same mass, keeping fixed each satellite’s host-centric spatial position. This reassignment preserves $\langle N_{\text{sat}}(M) \rangle$ as well as the higher order moments of $P(N_{\text{sat}}|M)$ because all of satellites in each system are assigned to the same new host halo. Thus, $P(N_{\text{sat}}|M)$ is identical in the original and randomized catalogues, but any correlation between satellite occupation and host halo properties besides M is erased. Likewise, the intrahost spatial distribution of the randomized satellites has no memory of the host halo assembly history or environment.

Note that our assembly bias-erasing algorithm differs from the procedure adopted in Croton et al. (2007) in a subtle but important way. Croton et al. (2007) assign the entire galaxy population of each host halo to a new, randomly selected host. In particular, the central galaxy *and* its satellites are relocated together as an ensemble. Croton et al. (2007) correctly pointed out that this procedure exactly preserves the one-halo term in the original mock catalogue. Thus, if there is assembly bias present in the one-halo term, its algorithm does not erase it. Effect (ii) discussed in Section 3.1 is an example of assembly bias that impacts the one-halo term, and for the purposes of this paper, this effect must be erased because we wish to construct a counterpart mock catalogue that has zero assembly bias. For this reason, we separately assign centrals and satellites to new host haloes, in accord with the standard HOD assumption that central and satellite galaxy occupation is independent, embodied by the equation

$$P(N_{\text{gal}}|M) = P(N_{\text{cen}}|M) + P(N_{\text{sat}}|M).$$

We employ the exact same procedure described above when erasing assembly bias in the age-matching mock catalogues. In particular, we apply this procedure *independently* to the red and blue samples of mock galaxies. This appropriately mimics the assumption of our HOD model that the clustering of red galaxies is independent from the clustering of blue galaxies, and conversely. By performing the above two-step procedure separately on blue and red galaxy populations, we preserve both red and blue HODs, and leave no trace of assembly bias on either population.

Upon erasing assembly bias using these procedures, we are left with galaxy catalogues with *identical* HODs; however, one set of galaxy catalogues, our fiducial catalogues, have explicit galaxy assembly bias, while the other set of galaxy catalogues *cannot* exhibit galaxy assembly bias. This enables us to estimate how large an affect assembly bias may have on galaxy clustering statistics independent of any HOD fitting.

⁵ We have performed a variety of explicit tests to ensure that our results are insensitive to our choice for bin width.

3.4 The importance of assembly bias

Fig. 1 compares the projected two-point clustering of galaxies in our fiducial abundance-matching catalogues, which exhibit assembly bias, and in our catalogues with assembly bias erased for three different magnitude threshold samples. The effects of assembly bias are not insignificant compared to the errors on the simulation measurements (the hatched regions), and are large compared to the precision of the SDSS measurements. The relative effect of assembly bias is largest on large scales and ranges from approximately ~ 15 per cent on large scales for the $M_r < -19$ threshold sample to ~ 6 per cent for the $M_r < -21$ sample. The fact that the effect is most prominent for the lower luminosity thresholds is consistent with the dependence of halo clustering on formation time, which is more prominent for lower mass haloes (Gao et al. 2005; Wechsler et al. 2006). The relative effect of assembly bias in our abundance-matching mock catalogues is grossly similar to that in the SAMs of Croton et al. (2007). However, in detail Croton et al. (2007) find assembly bias to have a more complex dependence upon luminosity. In particular, their red galaxy sub-sample is consistent with no clustering enhancement due to assembly bias at the highest luminosities, so that the clustering of their brightest luminosity threshold samples exhibit diminished, rather than enhanced, clustering as a result of assembly bias.

Neither set of our mock catalogues suffice for a detailed description of the SDSS clustering data; however, these predictions are

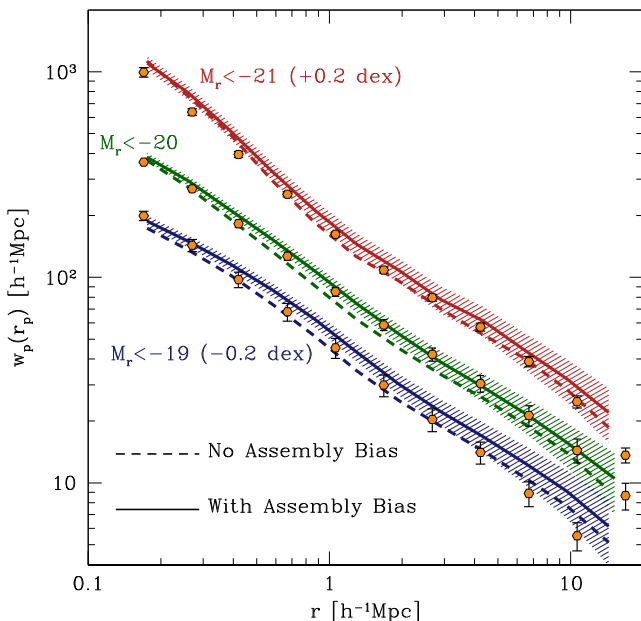


Figure 1. Assembly bias in abundance-matching models. The panel shows the projected two-point correlation functions for several galaxy luminosity threshold samples with $M_r < -21$ (top, offset +0.2 dex for clarity), $M_r < -20$ (middle), and $M_r < -19$ (bottom, offset -0.2 dex). The points with error bars represent measurements of $w_p(r_p)$ from SDSS DR7 (Zehavi et al. 2011). The lines represent the predicted values of $w_p(r_p)$ from abundance-matching mock catalogues based on the Bolshoi simulation (solid) and mock catalogues with precisely the same HODs, but with assembly bias erased (dashed). The hatched regions about the abundance-matching mock catalogue measurements represent the errors on the predicted $w_p(r_p)$ estimated from jackknife resampling of the simulation volume. We show errors only for the assembly bias mock catalogues in the interest of clarity; however, the errors on the $w_p(r_p)$ in models with assembly bias erased are similar.

broadly similar to SDSS clustering, so it is reasonable to suppose that these catalogues exhibit some of the richness of the observed galaxy data and may yield insight into galaxy clustering. For the purposes of this paper, the salient point is that the clustering differences shown in Fig. 1 between the fiducial and assembly bias-erased catalogues will drive our halo model fits to recover (erroneously) *distinct* HODs.

As Fig. 1 shows, the relative size of the effect of galaxy assembly bias on galaxy clustering statistics in these catalogues is large. The clustering is most altered on relatively large scales ($r_p \gtrsim 1 h^{-1}$ Mpc), suggesting that the effect is primarily due to the occupation statistics of central galaxies. This is indeed the case, so it is useful to examine the differences in host halo clustering among our mock catalogues.

In Fig. 2, we compare the host halo populations in our mock catalogues. The top panel of Fig. 2 shows the masses and maximum circular velocities of objects in our catalogues with and without assembly bias. The bottom panel of Fig. 2 compares the clustering of haloes that are selected to have central galaxies in our fiducial catalogues, with assembly bias, to the clustering of host haloes in our catalogues in which assembly bias has been erased. For demonstration purposes, we choose the $M_r < -19$ threshold sample for this example because the galaxy assembly bias is largest for this sample (Fig. 1), and because halo assembly bias is largest in low-mass host haloes (Gao et al. 2005; Wechsler et al. 2006). Recall that the mean occupation statistics of central galaxies in these catalogues are *identical* by construction.

Each point in the *top* panel represents a central galaxy in one or both of our luminosity-only $M_r < -19$ catalogues. Centrals that are common to both catalogues appear as black triangles. The vertical red line illustrates the V_{\max} cut corresponding to $M_r < -19$; the horizontal green line illustrates the cut on halo mass M_h that produces the same corresponding number density of haloes. As discussed above, when randomizing central galaxy occupation we include haloes that did not necessarily host a central galaxy in the fiducial catalogue. Thus, there is no guarantee that a halo hosting a central galaxy in the fiducial catalogue will host a central in the no-assembly-bias counterpart catalogue, and conversely. With red asterisks (green diamonds) we show those host haloes in the $M_r < -19$ fiducial catalogue (erased assembly bias catalogue) that do not appear in the catalogue without (with) assembly bias. The haloes that are common to both catalogues represent approximately ≈ 74 per cent of the host halo population. The remaining ≈ 26 per cent of haloes differ between the two catalogues.

Now we turn attention to the *bottom* panel of Fig. 2. The solid, black line shows the projected correlation function of the haloes that are selected to have central galaxies in both our fiducial mock catalogue (with assembly bias) and in our mock catalogue in which assembly bias has been erased. The dashed and dotted lines show the clustering of the haloes that are unique to the fiducial catalogue and the catalogue with assembly bias erased, respectively. Fig. 2 shows that the haloes that are distinct to the galaxy populations with and without assembly bias are clustered significantly differently. The haloes unique to the fiducial catalogues are a factor of ~ 3 more strongly clustered on small scales ($r_p \lesssim 1 h^{-1}$ Mpc) and a factor of ~ 2 more strongly clustered on large scales ($r_p \gtrsim 10 h^{-1}$ Mpc) than the haloes unique to the galaxy populations with no assembly bias. The difference in host halo clustering shown in Fig. 2 is nearly sufficient to account for the entirety of the differences between the two-point clustering in the $M_r < -19$ samples, even on scales $r_p \lesssim 1 h^{-1}$ Mpc.

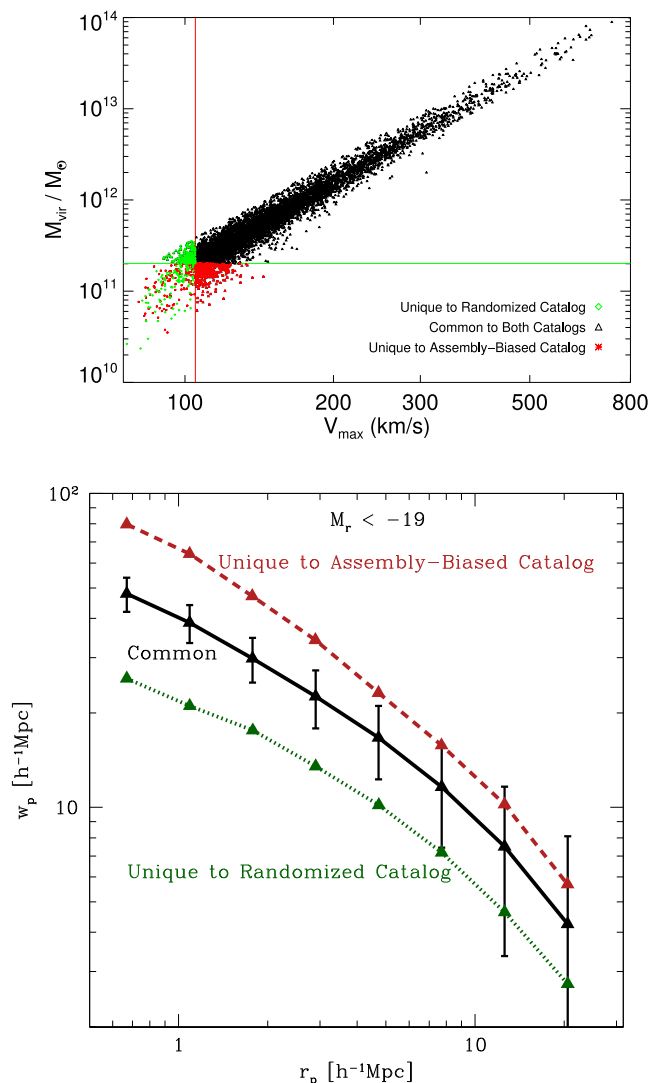


Figure 2. Comparing the haloes that host central galaxies in mock catalogues with and without galaxy assembly bias. Each point in the top panel represents a central galaxy in one or both of our luminosity-only $M_r < -19$ catalogues. In the bottom panel, the black triangles connected by the solid line shows the two-point clustering of haloes selected to have central galaxies in both catalogues. The upper, dashed line shows the two-point clustering of host haloes that are assigned central galaxies only in our fiducial galaxy catalogue that exhibits assembly bias. The lower, dotted line shows the two-point clustering of host haloes that are assigned central galaxies only in the galaxy catalogue in which assembly bias has been erased. All data in this plot refer to the $M_r < -19$ sample with a mean galaxy density of $n_g \simeq 1.57 \times 10^{-2} h^{-1} \text{Mpc}$. Errors are shown only for the two-point function of haloes common to all catalogues in the interest of clarity.

4 RESULTS

Using the methods described in Section 2.2, we fit the clustering of galaxies in luminosity threshold samples as well as red and blue galaxy sub-samples, in both our fiducial mock galaxy catalogues with assembly bias and our galaxy samples in which assembly bias has been removed. In this section, we describe our results with an emphasis on how these fits differ between the mock galaxy samples with and without assembly bias (but with *identical* true HODs).

4.1 Fits to projected clustering

We begin the discussion of our results by stating that, with a single exception, every sample of mock galaxies we study passes a naive ‘goodness-of-fit’ test based on the best-fitting χ^2 , including those galaxy samples exhibiting assembly bias. An immediate and unavoidable conclusion is that the ability to achieve an acceptable fit to projected galaxy clustering data with a standard HOD model has little bearing on the question of whether or not galaxy assembly bias is present in the real Universe.

As we showed in Section 3.4, assembly bias in abundance-matching mock catalogues is strong. The strength of assembly bias predicted by the age-matching assignments of galaxy clustering is, perhaps, extreme because galaxy colour is in monotonic correspondence with halo age at fixed halo V_{max} in this model. Yet nearly all of these mock galaxy samples may be fitted by a galaxy–halo model in which assembly bias is (incorrectly) assumed to be zero.

In Fig. 3, we show our fits to each of the samples we study in this paper. We remind the reader that the correlation function data at different spatial separations are highly correlated, so it is not trivial to estimate χ^2 by inspection of the lines and points in Fig. 3. This is particularly true for the fits to the colour-split samples because both red galaxy and blue galaxy clustering are fit simultaneously, subject to the constraint that there can be only one central galaxy per halo.

With this caveat in mind, Fig. 3 suggests that our fitting procedure is generally quite successful in that it recovers the correct two-point galaxy clustering. In particular, the fits to the samples with no assembly bias are of high quality in all cases. This is, perhaps, not particularly surprising because the models with no assembly bias are consistent with the premises on which the halo model is based and halo clustering statistics have now been calibrated very accurately from cosmological simulations (e.g. Tinker et al. 2008b, 2010). Fig. 3 also demonstrates that our fits to the projected clustering of samples *with* assembly bias generally describe the mock catalogue data quite well. With only a single exception (see below), each of our fits results in a chance probability for attaining larger best-fitting χ^2 values that exceeds $\gtrsim 0.05$.

Let us now discuss the only mock galaxy sample whose best-fitting HOD fails the goodness-of-fit test: the $M_r < -19$, colour-split fiducial catalogue *with* assembly bias. Recall that we fitted the red and blue samples simultaneously subject only to the constraint that there can be no more than one central galaxy per halo. The simultaneous fit to the $M_r < -19$ red and blue galaxy samples results in a best-fitting $\chi^2 \simeq 74.4$; for 16 degrees of freedom, a χ^2 this large or larger would occur by chance with a probability of $\simeq 2 \times 10^{-9}$. This suggests strongly that our halo model description of colour-dependent clustering cannot describe the distribution of galaxies with $M_r < -19$ assigned to Bolshoi haloes through age matching.

Intriguingly, the poor fit to the colour-selected $M_r < -19$ sample occurs in a situation similar to the unacceptable fit to SDSS data for colour-split samples with $-20 \leq M_r \leq -19$ in Zehavi et al. (2011). However, the Zehavi et al. (2011) result is driven by the fact that the blue galaxies are more weakly clustered than their model can accommodate, while in the case of our fits, the model fails because it cannot accommodate the strength of the red galaxy clustering in the age-matching mock galaxy distribution. In particular, we find that the large-scale bias of the red galaxies in this sample is ~ 20 percent higher than in our best-fitting model. In order to accommodate this, it would be necessary to place galaxies in haloes that are ~ 5 times more massive because the halo bias function

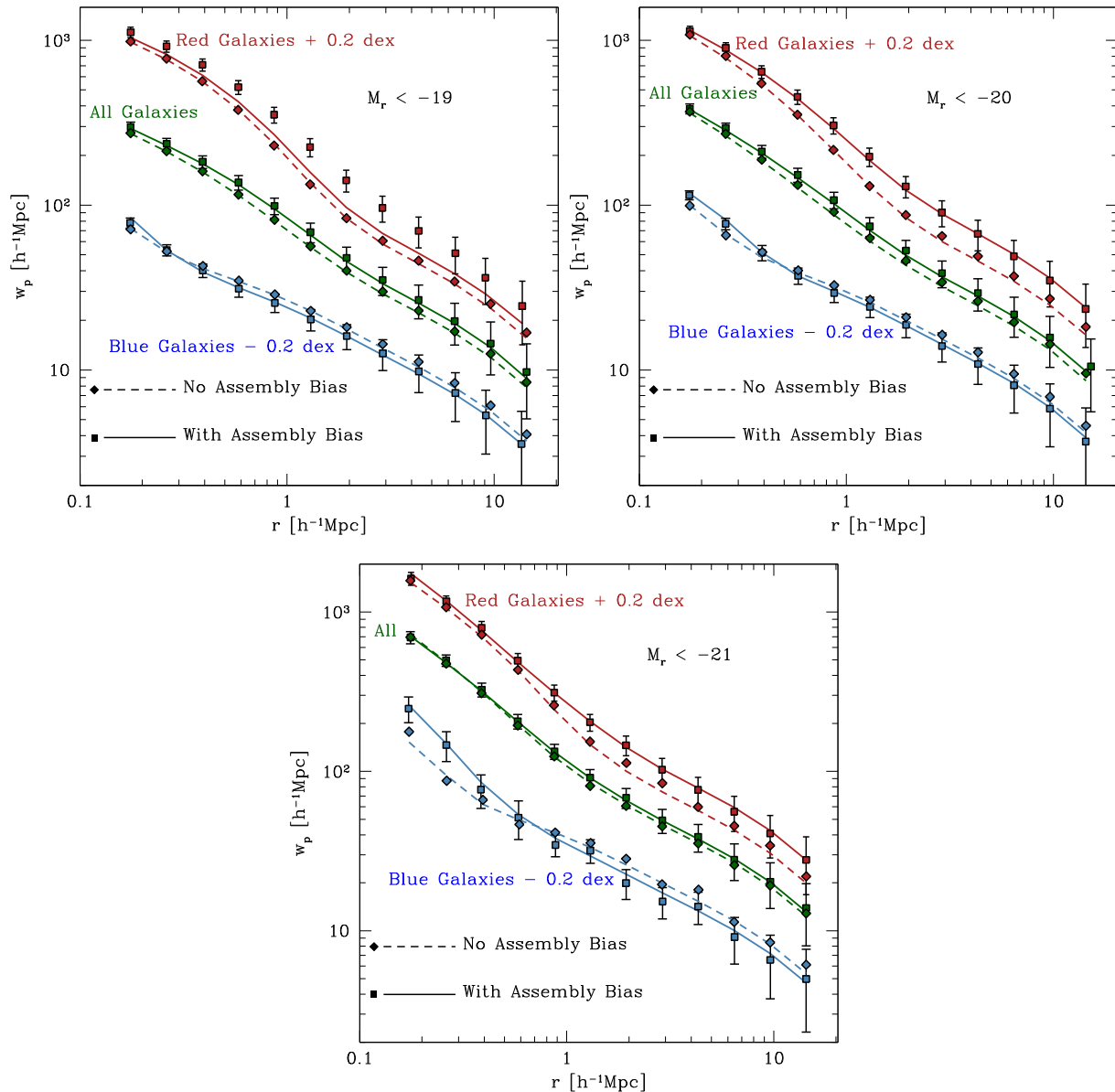


Figure 3. Halo model fits to the $w_p(r_p)$ measured in the Bolshoi simulations. Different panels show fits to samples with different luminosity thresholds, where the brightness cut is indicated in each panel. Solid lines are the best fits to our fiducial models that exhibit assembly bias while dashed lines are the best fits to our models with no assembly bias. The fiducial galaxy samples and their counterparts in which assembly bias has been erased have identical HODs. Mock catalogue data are represented by squares for our fiducial models and by diamonds in the case of erased assembly bias. The topmost pair of mock catalogue data points and curves (in red) correspond to the sub-samples of red galaxies; the bottom pair (in blue) correspond to the sub-samples of blue galaxies; the central pair (in green) correspond to all galaxies in the luminosity threshold sample with no colour selection. For visual clarity, the projected correlations for the red (blue) samples have been offset in the positive (negative) direction by 0.2 dex, and errors are shown only for the fiducial mock catalogues. We emphasize that it is not trivial to determine χ^2 by visual inspection of this plot because the $w_p(r_p)$ data are highly correlated.

is a shallow function of halo mass in the relevant mass range, near $M_{\min} \sim 7 \times 10^{11} h^{-1} M_{\odot}$. However, this increase in mass is impossible because a model with a significantly larger value of M_{\min} does not yield the correct average number density. Moreover, the parameter f_b alone cannot accommodate the strength of the red galaxy clustering for this sample without simultaneously over predicting the clustering of the blue galaxies. For the remainder of this paper, we proceed by giving the results for each of our halo model fits with the caveat that, according to a χ^2 goodness-of-fit test, the fit to the colour-split sample with $M_r < -19$ is unacceptable.

4.2 The character of colour-dependent assembly bias

Before taking a detailed look at our best-fitting HOD parameters in the following section, let us first use Fig. 3 to consider the qualitative imprint that assembly bias leaves on colour-selected galaxy samples. Note that for both the luminosity threshold samples and the red galaxy sub-samples, assembly bias tends to drive galaxy clustering higher on all scales, just as in Figs 1 and 2. This is because the same physical mechanism gives rise to assembly bias in both cases. Abundance-matching-generated luminosity threshold samples are preferentially populated with centrals living in haloes that have

large values of V_{\max} for their masses; such haloes are more strongly clustered, as discussed in Section 3.1. Age matching exaggerates this effect in red galaxy samples, because red centrals are explicitly chosen to reside in the earliest forming haloes, with the highest values of V_{\max} , at a given mass.

On the other hand, the blue galaxy samples in Fig. 3 exhibit distinctly different behaviour in two respects. First, assembly bias drives these galaxies to be more weakly clustered on large scales ($r_p \gtrsim 200\text{--}700 h^{-1} \text{Mpc}$ depending upon the sample under consideration). This is because age matching assigns blue galaxies to haloes that acquired their mass relatively more recently, and it is now well known that relatively-later forming haloes tend to be more weakly clustered than their earlier-forming counterparts at fixed mass.

Secondly, on small scales ($r_p \lesssim 200\text{--}700 h^{-1} \text{Mpc}$), blue galaxy clustering is *strengthened* by assembly bias. This is primarily caused by effect (ii) discussed in Section 3.1: at fixed mass, later-forming host haloes have a larger than average number of satellites. In age matching, haloes with a blue central galaxy are the latest-forming haloes of a given mass. Therefore, at fixed mass, age matching predicts that the presence of a blue central is correlated with having a larger than average number of satellites. Mathematically, this is represented as

$$\langle N_{\text{cen}}^{\text{blue}} N_{\text{sat}} \rangle > \langle N_{\text{cen}}^{\text{blue}} \rangle \langle N_{\text{sat}} \rangle.$$

This small-scale effect is further enhanced by the phenomenon of ‘galactic conformity’ (e.g. Weinmann et al. 2006), a term referring to the observed tendency for a blue (red) central galaxy to host a preferentially blue (red) satellite population. Age matching naturally predicts galactic conformity because the formation time of a host halo correlates with the formation time of its sub-haloes, as both the host and its sub-haloes collapse from the same region of the cosmic density field. Mathematically, this second effect is represented as

$$\langle N_{\text{cen}}^{\text{blue}} N_{\text{sat}}^{\text{blue}} \rangle > \langle N_{\text{cen}}^{\text{blue}} \rangle \langle N_{\text{sat}}^{\text{blue}} \rangle.$$

Note that both of these effects violate the common HOD assumption that the satellite and central HODs are uncorrelated: $P(N_{\text{gal}}|M) = P(N_{\text{cen}}|M) + P(N_{\text{sat}}|M)$.

These two small-scale effects work together to boost the central–satellite pair counts in blue galaxy samples with assembly bias relative to blue samples with no assembly bias. This boost strengthens clustering on small scales where the one-halo term dominates. That the impact on two-point clustering is significant primarily for blue samples derives from two facts. First, it is only for the blue samples that $P(N_{\text{cen}} = 1|M) < 1$ for an appreciable portion of the halo mass range of interest (see Section 4.3 below). Furthermore, early-forming haloes host the reddest galaxies in the age-matching procedure, yet early-forming haloes have fewer satellites because processes have operated for a greater number of dynamical times in such systems (e.g. Zentner et al. 2005; Watson, Berlind & Zentner 2011), with the consequence that the boost in small-scale clustering that may be provided by conformity may be partially cancelled by dynamical evolution.

4.3 Best-fit HOD models

The primary aim of this paper is to emphasize the significance of assembly bias in abundance matching and age matching, and to exploit these differences to estimate the potential for assembly bias to introduce systematic errors in the HOD parameters inferred in galaxy clustering analyses. Therefore, we now turn to examining the HODs inferred from our fits. Before turning to constraints on

individual HOD parameters in the following section, in this section, we begin by examining $\langle N_{\text{gal}} \rangle (M_{\text{vir}})$, the average halo occupation in our acceptable models. We consider this to be the most useful manner in which to represent our results because the posterior distribution of HOD parameters that results from our fits exhibits strong correlations.

4.3.1 HOD fits to $M_r < -19$ samples

In this section, we compare our inferred HODs for the $M_r < -19$ samples to the actual HODs measured in the Bolshoi simulation. We remind the reader that according to a χ^2 goodness-of-fit test, we do not achieve a good fit to the clustering of red galaxies for this luminosity threshold. This is due to the strength of assembly bias in our chosen fiducial model. However, the considerations driving systematic biases in HOD fits to assembly-biased samples are the same for all our luminosity thresholds, and so we consider it instructive to begin discussion of our fits with the $M_r < -19$ samples.

Fig. 4 illustrates $\langle N_{\text{gal}} \rangle (M)$ for both luminosity- and colour-selected $M_r < -19$ galaxy samples. First examine the fits to the luminosity threshold samples in the bottom panels of Fig. 4. In both cases, the inferred HODs are similar to the true HODs. However, in the fits to the fiducial catalogues with assembly bias, the fits are slightly more strongly biased towards a rapid transition from $\langle N_{\text{cen}} \rangle = 0 \rightarrow \langle N_{\text{cen}} \rangle = 1$. The character of this bias will be encountered in many of our fits. It is caused by the stronger clustering of galaxies in the fiducial catalogues, and can be understood as follows. In general, a long tail of non-negligible N_{cen} extending to low mass drives clustering strength down because low-mass haloes are more weakly clustered than high-mass haloes. Our HOD model is unable to exploit halo assembly bias, so our HOD-fit to the fiducial catalogue with assembly bias achieves adequate clustering strength, in part, by making the central galaxy transition sharp, rather than extended; this avoids populating a low-mass tail in N_{cen} , thereby boosting the clustering.

Furthermore, it is also apparent that the average number of satellite galaxies is a steeper function of halo mass in the fits to the fiducial catalogues with assembly bias. In the HOD language, this is reflected in a larger power-law index α , describing the average number of satellites as a function of halo mass. Fits to samples with assembly bias yield significantly larger values of the parameter α because the large-scale clustering of galaxies can be increased by packing relatively more galaxies into more strongly clustered higher mass haloes and fewer galaxies into less strongly clustered, lower-mass haloes. This bias in the number of satellites in large haloes is another feature of our inferred HODs that will be encountered repeatedly in this section.

These trends are more strongly in evidence for the red galaxy sub-samples in the top row of panels in Fig. 4. In particular, in our best-fitting HOD to the red sample with assembly bias, the central galaxy transition is far sharper than the true transition. The effect is of the same character, and more pronounced, in this case because assembly bias in red sub-samples is largely an extreme version of assembly bias in luminosity threshold samples, as discussed in Section 4.2.

An additional distinguishing feature of fits to the colour-selected samples is driven by the need to simultaneously reproduce the blue and red galaxy clustering. A consequence of this is that the bias parameter f_b cannot vary to accommodate the large-scale clustering of the red sample without predicting blue galaxy clustering in excess of the mock catalogue data. Recall that assembly bias in

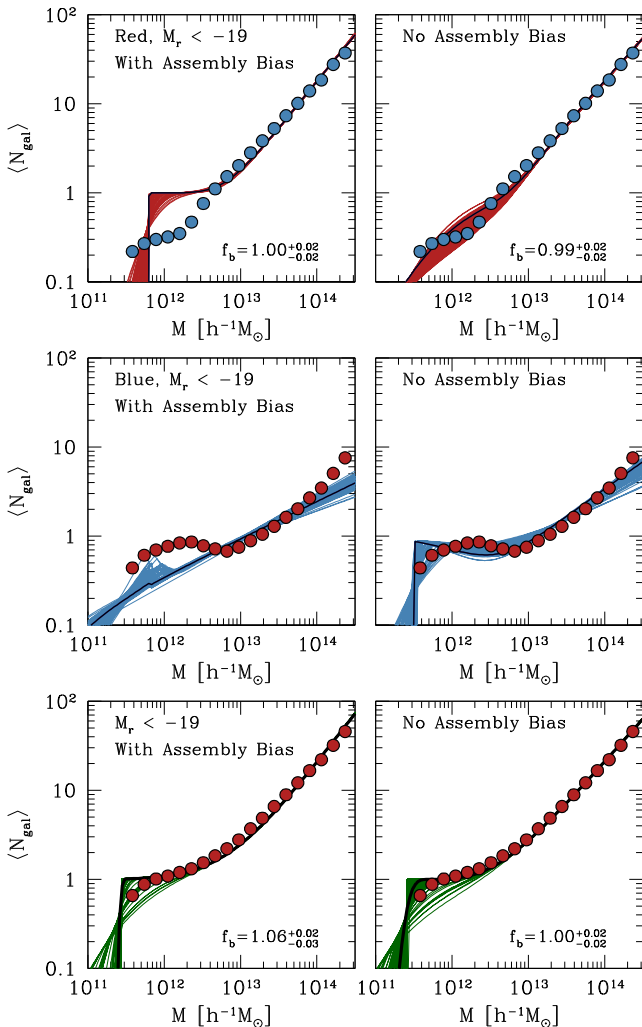


Figure 4. Comparison of the best-fitting HODs for galaxies in the $M_r < -19$ sample with the true HODs in the mock galaxy catalogues (points). In each case, the thick, solid lines represent the best fits and the numerous, thin, coloured lines represent 100 randomly selected HODs from the MCMC chain that are within $\Delta\chi^2 \leq 1$ relative to the best-fitting model. The marginalized constraint on the halo bias nuisance parameter f_b is shown in each of the panels for the red, colour-selected samples and the luminosity threshold samples. Note that the red and blue samples are fit simultaneously, so they are fitted with a common value for f_b .

these models increases red galaxy clustering strength and decreases blue galaxy clustering strength. The marginalized constraints on f_b clearly demonstrate this point. In the threshold samples, the strong clustering of the samples with assembly bias is taken up, in part, by f_b so that the best-fitting value of $f_b = 1.06$ sets halo clustering to be 6 per cent stronger than predicted by the Tinker et al. (2008b) formula. On the contrary, the best-fitting bias parameter from the fits to the colour-split samples is $f_b = 1$. In Appendix A, we show how eliminating the extra freedom afforded by the parameter f_b alters inferred HODs. In short, allowing f_b to vary along with the HOD increases the errors in the inferred HOD parameters and mitigates the biases in the inferred HODs.

Continuing our discussion of the red galaxy samples in the top panels of Fig. 4, it is interesting to note that the true HOD has a feature at $M \approx 2 \times 10^{12} h^{-1} M_\odot$ and $\langle N_{\text{gal}} \rangle \simeq 0.4$ that cannot be accommodated by the standard functional form for the central

galaxy HOD, yet the acceptable HODs go broadly through this feature in a smooth fashion. In contrast, the failure of the fit to recover the correct HOD in the case of the fiducial catalogues with assembly bias is dramatic.

Consider now the blue galaxy samples in Fig. 4. Again, the HOD is recovered comparably well from the sample in which assembly bias has been eradicated, whereas the recovery of the true HOD is significantly poorer in the fiducial sample. However, in this case the sense of the difference is opposite to that in full the luminosity threshold sample. To be specific, the fits are driven to more gradual (rather than more rapid) central galaxy transitions and lower values of α . The reason is because assembly bias *weakens* the large-scale clustering of blue galaxies, as discussed in Section 4.2. Halo model fits attempt to compensate for this by placing as many galaxies as possible in lower mass haloes, thereby broadening the central galaxy transition and reducing α .

To conclude this section, we note that our fits to the clustering of $M_r < -19$ galaxies in the mock catalogues in which assembly bias has been eliminated are significantly less biased than fits to our fiducial samples. In general, the fits to the galaxy catalogues with no assembly bias generally recover the true, input HOD well. This is promising because it suggests that halo model methods have been sufficiently well developed such that when the data are consistent with the premises of the standard HOD implementation (i.e. no assembly bias), they do correctly model the clustering of galaxies in a wide range of reasonable models, and can be used to interpret galaxy clustering data. This is important because there exist few validation exercises that demonstrate this fact in the literature (we are aware of only Reddick et al. (2014), but even in this case the focus is on cosmological parameters, and HOD parameters are not treated in any detail). Moreover, the strength of assembly bias in the real universe remains an open question.

4.3.2 HOD fits to brighter samples

We now consider results pertaining to galaxy samples with brighter luminosity thresholds. We remind the reader that the clustering of *all* samples we consider in this section is adequately fitted by an HOD model, as determined by a χ^2 goodness-of-fit test.

Fig. 5 is analogous to Fig. 4, but here we show the results from fits to $M_r < -20$ mock galaxy samples. In general, the true HOD is recovered relatively well for samples with no assembly bias, particularly for the luminosity threshold sample. On the other hand, the HODs inferred by fitting the clustering to the samples with assembly bias are poorer representations of the true HODs, even for the luminosity-only sample.

The sense of the biases are familiar from our discussion in Section 4.3.1. In particular, in the luminosity threshold sample the central galaxy transition is biased to be significantly sharper than the true transition and the power-law index α is biased higher in order to increase the clustering strength to mimic the effect of assembly bias. More dramatic biases of the same sense are evident for the red galaxy sub-sample. Conversely, the inferred HODs of the blue galaxies are biased towards placing galaxies in low-mass haloes in order to *reduce* the clustering in blue samples (and mimic the effect of assembly bias on blue galaxies). This manifests in two ways. First, the best-fitting HOD model places essentially all blue centrals in low-mass haloes, rather than over a broad range of halo masses. Secondly, blue satellites are inferred to be significantly less abundant within high mass ($M_{\text{halo}} \gtrsim 10^{13} h^{-1} M_\odot$) than they truly are. In this case, the failure is so dramatic that an analyst performing

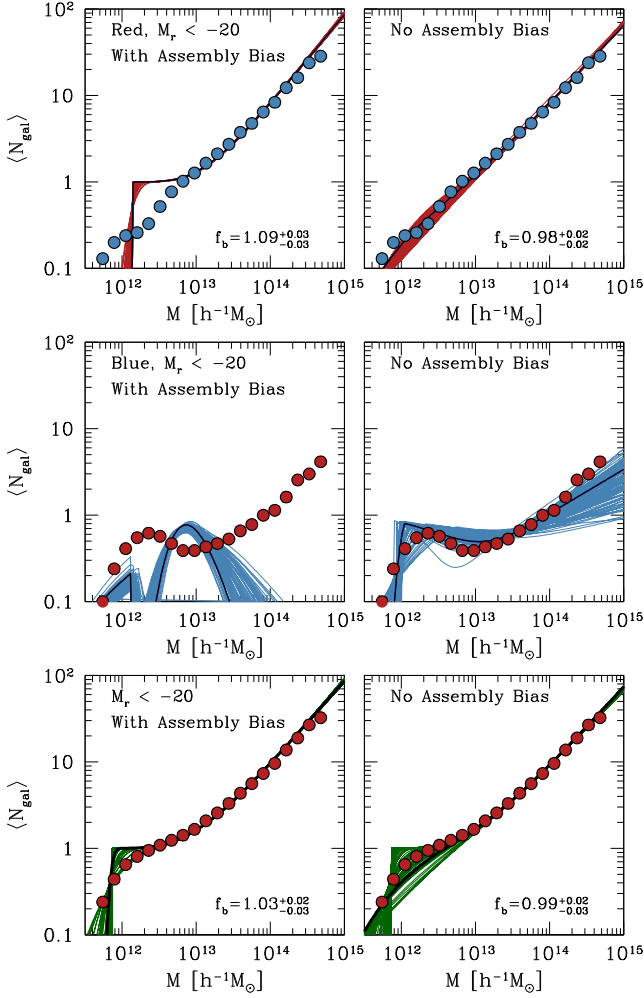


Figure 5. Comparison of the best-fitting HODs for galaxies in the $M_r < -20$ sample with the true HOD in the simulation (points). This figure is the same as Fig. 4, but addresses the higher luminosity samples with $M_r < -20$.

such a fit would recognize it as incorrect. For example, this fit implies no blue satellite galaxies in large clusters. This is manifestly false. In this case, adding additional data, such as galaxy–galaxy lensing (as in Leauthaud et al. 2012; van den Bosch et al. 2013), or galaxy number-to-halo mass ratios (as in Tinker et al. 2012), or additional priors should greatly mitigate against this failure mode.

Fig. 5 makes it apparent that the best-fitting HOD models fail to recover the true halo occupation statistics exhibited by any of our fiducial $M_r < -20$ galaxy samples. In contrast, the HODs of the $M_r < -20$ without assembly bias are recovered comparably well. However, both samples have identical true HODs. We conclude that the failure of the standard HOD model to accurately describe our fiducial galaxy distributions is due in large part to the presence of assembly bias in these samples.

Fig. A1 in the appendix shows inferred HODs for the same samples as in Fig. 5 but in an alternative case in which f_b has been fixed to unity. In this case, the biases are more statistically significant, indicating that including such a nuisance parameter is important for mitigating, in part, biases in inferred HODs due to assembly bias. We have included the f_b bias parameter in order to be conservative, and minimize the systematic errors in HODs induced by assembly bias. However, it is worth noting that our conservative choice to include such a parameter is not standard practice.

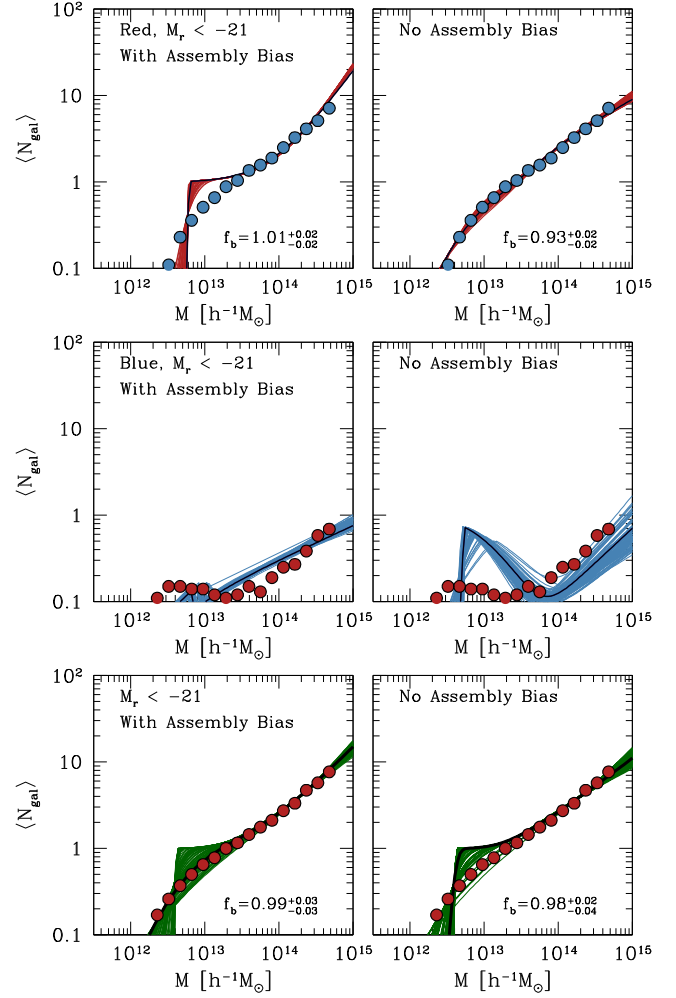


Figure 6. Comparison of the best-fitting HODs for galaxies in the $M_r < -21$ sample with the true HODs in the simulation (points). This is the same as Fig. 4, but addresses this higher luminosity $M_r < -21$ samples.

Fig. 6 depicts the inferred HODs for the $M_r < -21$ mock galaxies. The same broad features that we expounded upon above are evident for the case of red sub-samples. However, the other cases do not qualitatively resemble their counterparts in the lower luminosity samples. In the luminosity threshold samples shown in the bottom row of panels in Fig. 6, the true underlying HOD is recovered more faithfully in the presence of assembly bias, despite the fact that assembly bias is *not* included in the modelling. The bias in the HOD inferred from the mock galaxy sample with no assembly bias is, in part, due to the extra parameter freedom from the halo bias parameter, f_b . The increased clustering due to the erroneously sharp central galaxy transition can be compensated by decreasing f_b ; these parameter shifts result in a slightly better χ^2 despite the offset in the inferred HOD. Additionally, the blue mock galaxies with $M_r < -21$ are an interesting exception to the general trends observed in our other sample analyses. The inferred HODs in these cases certainly differ from each other; however, in neither case are the inferred blue HODs representative of the true, underlying blue galaxy HODs. These results suggest one or more failures of the halo model to describe the clustering of the haloes in these samples.

We summarize these results by reiterating the salient point of this section: fits to the clustering exhibited by mock galaxies with assembly bias typically yield inferred HODs that are biased. Thus in

the absence of independent evidence justifying the assumption that assembly bias is zero, we conclude that conventional implementations of the HOD are ill-equipped to robustly constrain galaxy–halo models with two-point clustering measurements alone. We reach this conclusion even though we have made the conservative choice to marginalize over f_b , whereas what is almost universally done is to assume that halo bias is perfectly calibrated and hold f_b fixed to unity.

4.4 Constraints on individual HOD parameters

The preceding figures make clear the trends in the systematic errors on inferred HOD parameters to be expected when fitting galaxy samples in which galaxy properties are correlated with halo properties other than mass. They also give a reasonable representation of how different the inferred and true HODs can be when analysing real galaxy samples, where assembly bias may or may not be present. Furthermore, the HODs represented in Fig. 4 through Fig. 6 include the covariance among the inferred HOD parameters in each case.

None the less, there is considerable interest in the constraints on individual HOD parameters, and estimates of these parameters can often be of practical use. We turn now to the marginalized constraints on particular HOD parameters in each of our samples. We focus on the three parameters that tend to garner the broadest interest: (1) the mass scale at which the average number of central galaxies per halo is $1/2$, $\log(M_{\min}/h^{-1} M_{\odot})$ (an inferred parameter); (2) the mass scale at which the average number of satellite galaxies is 1, $\log(M_1/h^{-1} M_{\odot})$; and (3) the power-law index of the satellite portion of the HOD, α .

Fig. 7 shows the 1σ marginalized constraints on the derived parameter $\log(M_{\min}/h^{-1} M_{\odot})$ for each of the samples we have fitted with an HOD. While we have derived constraints on the logarithm $\log(M_{\min}/h^{-1} M_{\odot})$, we will refer to these as ‘ M_{\min} constraints’ in the interest of brevity.

For the luminosity threshold samples and the red sub-samples, the constraints on M_{\min} are typically tighter when derived from the fiducial mock galaxy populations with assembly bias. This is largely because these models are driven to have very low $\sigma_{\log M}$ values in order to place galaxies in the most massive haloes possible, thereby boosting clustering. When $\sigma_{\log M}$ is limited to a very narrow range, M_{\min} is also limited to a very narrow range in order to guarantee that the HOD describes a model with the correct average number density of galaxies. Counterintuitively, this also explains why the inferred values of M_{\min} are generally *smaller* in the samples with assembly bias. When $\sigma_{\log M}$ is relatively large, a large fraction of all galaxies in the sample reside in haloes with masses below M_{\min} because the halo mass function increases rapidly as halo mass is decreased. The blue sub-samples run counter to this general trend because their clustering is diminished by assembly bias rather than enhanced (see Section 4.2).

It is clear from Fig. 7 that the differences in the inferred values of M_{\min} between samples with and without assembly bias can be quite significant. Moreover, it is also significant that the precision of the inferred constraints on HOD parameters varies significantly between the models with and without assembly bias. For perspective on this, consider the particular case of the luminosity threshold sample with $M_r < -20$ (the green bands in the middle panel of Fig. 7), in which the systematic difference may not seem egregious. In this case, a fit to the assembly-biased mock galaxy data would rule out the median M_{\min} for the case with no assembly bias by more than $\sim 2\sigma$, despite the fact that the inferred HOD in the case with no assembly bias is an excellent description of the true HOD (Fig. 5).

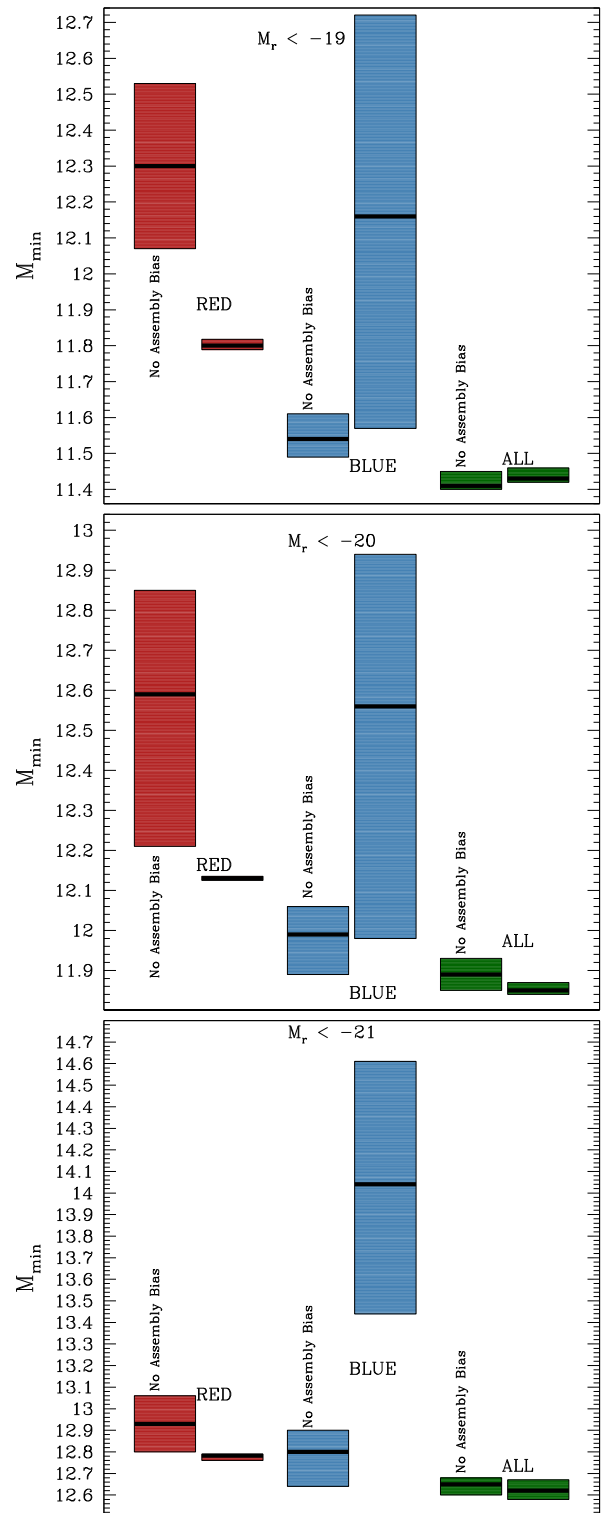


Figure 7. Constraints on $\log M_{\min}/h^{-1} M_{\odot}$ inferred by fitting $w_p(r_p)$ from the mock galaxy catalogues. In each case, the bars span the 1σ marginalized constraint on M_{\min} defined to be symmetric in the sense that the posterior integrates to 68.3 per cent over the domain delineated by the bands while the high- M_{\min} and low- M_{\min} tails outside of the bands each integrate to $(100 - 68.3)\text{per cent}/2 = 15.85\text{ per cent}$ of the posterior. The thick, solid lines show the median values of M_{\min} . From top to bottom, the panels show the results of the $M_r < -19$, $M_r < -20$, and $M_r < -21$ samples. Within each panel, we show the results for the luminosity threshold samples (labelled ‘ALL’ and at the far right), as well as the colour-split samples separately.

Differences of this sort are particularly pronounced for red-selected samples for which the inferred values of M_{\min} differ by significantly more than the statistical errors on M_{\min} . This strongly suggests that the error budgets of HOD analyses of galaxy clustering require a substantial, previously neglected contribution from the unknown strength of galaxy assembly bias in the real Universe. We will return to this point below, and in Section 6.

Fig. 8 depicts the inferred values of the parameter M_1 . As with M_{\min} , the offsets in inferred M_1 values can be significant. In most cases, the systematic offsets in the inferred values of M_1 are comparable to, or larger than, the statistical errors on these inferences. Even for the luminosity threshold samples, the offsets in the constraints on M_1 are significant at all luminosities. Again, there is a relatively general pattern to these systematic offsets. For the luminosity threshold samples and the red galaxy samples, the trend is for the inferred M_1 to be larger in the fiducial models with assembly bias, whereas for the blue mock galaxy samples the inferred M_1 is typically lower in the fiducial mock catalogues. Again, this is driven by the HOD adjusting to increase clustering, by packing galaxies into the most massive haloes possible, in the former cases and to reduce clustering in the latter case.

Following our discussion of M_{\min} and M_1 , Fig. 9 gives the inferred values of the satellite galaxy power-law index α in each of our fits. In the case of α , the systematic offsets induced by assembly bias are comparable to, or larger than, the statistical errors in all cases, including the threshold samples. In the case of the blue galaxies in the fiducial catalogue with $M_r < -20$, the constraint on α is not shown in Fig. 9 because it is significantly negative and depicting this constraint would alter the scale of the figure to the degree that clarity would suffer. The marginalized 1σ constraint on α in this case is $\alpha = -2.04^{+1.04}_{-2.24}$. As we mentioned above, an analyst would likely identify this value to be unphysical, as it predicts essentially zero blue galaxies in large clusters. In our analysis, we did not include any prior on α . For the luminosity threshold samples and for the red galaxies, the trend is for the inferred values of α to be larger in the fiducial samples than in the samples with assembly bias removed. The fits are driven to larger values of α in order to place galaxies preferentially in relatively rare, highly biased haloes, thereby boosting clustering. The values of M_1 inferred from these samples are larger as well in order to keep the total galaxy number density fixed and to mitigate the increase in clustering strength on small scales ($r_p \lesssim 1 h^{-1}$ Mpc), where pairs of galaxies within common host haloes dominate the signal. We refer the reader to Watson et al. (2011) for a more in depth discussion of the factors that determine the relative strength of the small-scale (one-halo) and large-scale (two-halo) clustering of galaxies. As with M_1 , and for the reasons discussed in Section 4.2, the systematic error in α is of the opposite sense for blue galaxies as compared to red galaxies.

The results in this section suggest that galaxy assembly bias at levels that cannot easily be ruled out may have a statistically significant effect on inferences made about the relationship between galaxies and the haloes in which they reside. These biases are best represented in the complete HOD representations shown in Fig. 4 through Fig. 6 because the systematic errors in the conventional HOD parameters (e.g. M_{\min} , M_1 , α) are strongly correlated. Nevertheless, the marginalized constraints on individual HOD parameters exhibit systematic errors that are statistically significant (Fig. 7 through Fig. 9). In the appendix, we demonstrate that removing the freedom provided by the f_b nuisance parameter renders these systematic errors slightly more significant, compared to the statistical errors on the inferred HOD parameters. Consequently, biases will generally be larger if no such nuisance parameter is introduced.

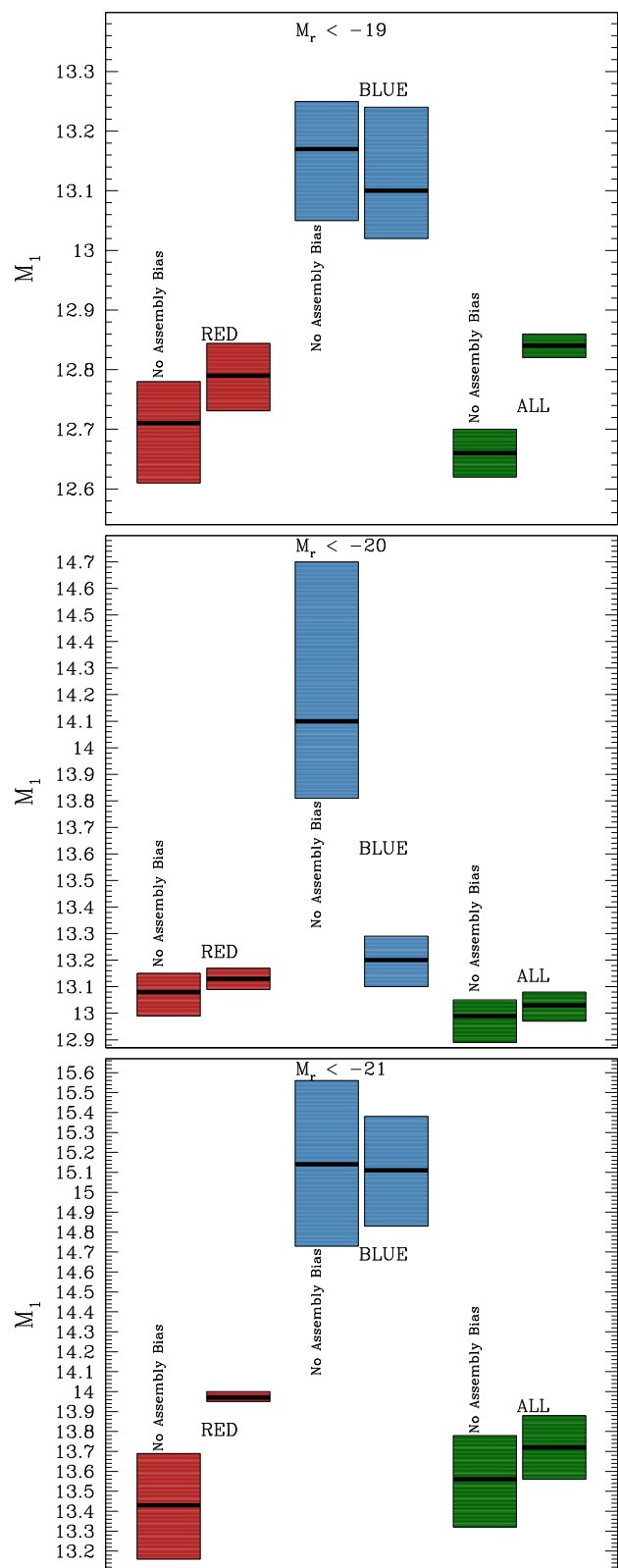


Figure 8. Constraints on $\log M_1/h^{-1} M_\odot$ inferred by fitting $w_p(r_p)$ from the mock galaxy catalogues. In each case, the bars span the 1σ marginalized constraint on M_1 defined as described in Fig. 7. The thick, solid lines show the median values of M_1 . From top to bottom, the panels show the results of the $M_r < -19$, $M_r < -20$, and $M_r < -21$ samples. Within each panel, we show the results for the luminosity threshold samples (labelled ‘ALL’ and at the far right), as well as the colour-split samples separately.

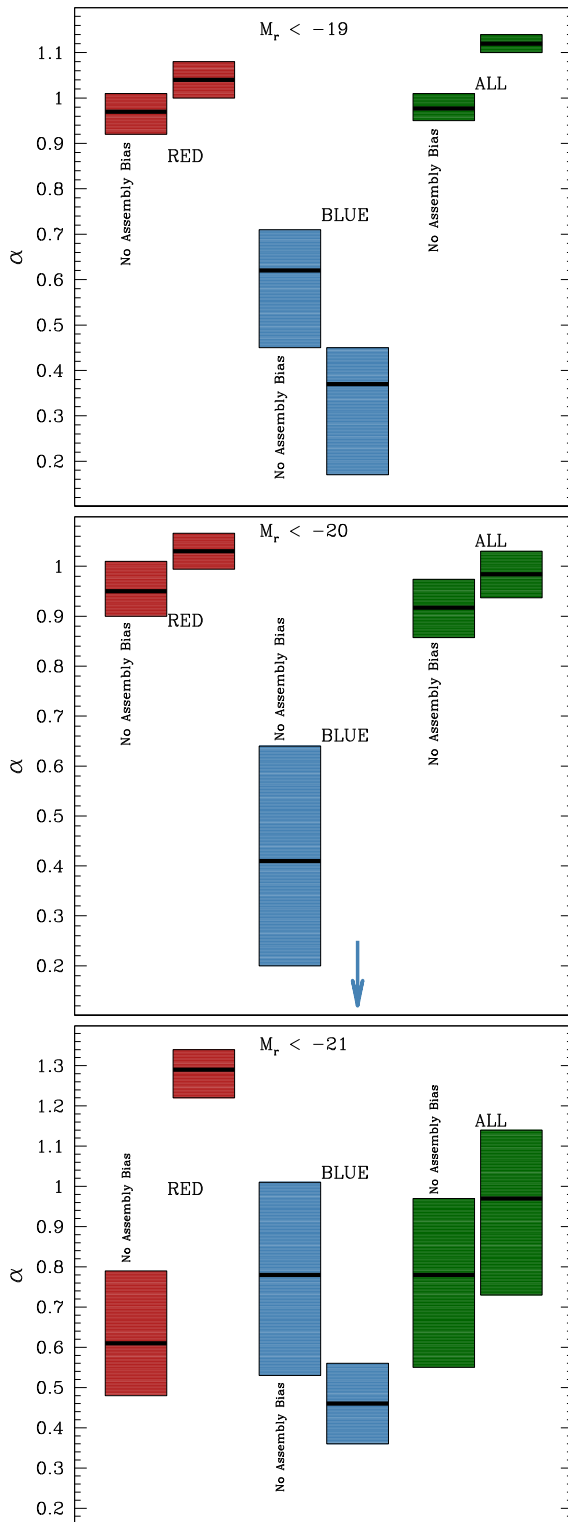


Figure 9. Constraints on α inferred by fitting $w_p(r_p)$ from the mock galaxy catalogues. In each case, the bars span the 1σ marginalized constraint on α defined as described in Fig. 7. The thick, solid lines show the median values of α . From top to bottom, the panels show the results of the $M_r < -19$, $M_r < -20$, and $M_r < -21$ samples. Within each panel, we show the results for the luminosity threshold samples (labelled ‘ALL’ and at the far right), as well as the colour-split samples separately. The constraint for the blue mock galaxies in the fiducial, assembly-biased $M_r < -20$ sample is not shown because it is off the scale of the figure. The marginalized constraint in this case is $\alpha = -2.04^{+1.04}_{-2.24}$.

We also show in the appendix that marginalizing over the internal spatial distributions of satellite galaxies does not mitigate the systematic errors in HODs induced by assembly bias (and can even increase the systematic offset). These demonstrations indicate that nuisance parameters introduced in previous HOD-based studies do not necessarily alleviate the effects of assembly bias on inferred HODs.

At minimum, these results suggest that assembly bias is an observationally relevant effect that is not accounted for in standard HOD/CLF analyses. In fact, it is worth reiterating that we should expect actual data to be significantly *more* sensitive to assembly bias than the mock catalogues we have analysed, with the consequence that inferences based on observational data may contain systematic errors of great statistical significance than those that we quote.

5 ADDITIONAL PREDICTIONS OF THE HALO MODEL AND HOD

The results presented in Section 4 demonstrate that traditional HOD fits to measurements of $w_p(r_p)$ and \bar{n}_g can be significantly altered by the presence of assembly bias. Evidently, when galaxy assembly bias is present in the data, there is sufficient parametric freedom in the HOD to compensate for incorrectly assuming assembly bias to be absent. In the light of these results, we argue that it is necessary to search for statistics other than $w_p(r_p)$ that may be sensitive to assembly bias so that these degeneracies may be broken. In Section 5.1, we investigate the potential for the VPF to detect the presence of galaxy assembly bias.

Once a set of acceptable HOD parameters has been determined by fitting galaxy clustering data, the HOD formalism enables new predictions about galaxy evolution to be made. We present an example of such a prediction in Section 5.2, in which we study the inferred host mass-dependence of satellite quenching, with particular attention to the threat that unknown levels of assembly bias pose for such inferences.

5.1 The void probability function

The amount of galaxy assembly bias in the universe is unknown and cannot be revealed by galaxy clustering statistics alone. Consequently, a natural question to ask is whether some additional statistic describing the galaxy distribution can be used to test for assembly bias. One natural candidate for such a statistic is the VPF, defined as the probability that a spherical region of some radius will be devoid of galaxies. The VPF has been studied previously for precisely this purpose. In Tinker et al. (2006, 2008a), the authors scrutinized particular models with assembly bias. The models they studied had VPFs that are inconsistent with SDSS measurements, even though the clustering in those models *is* consistent with the data, effectively falsifying those models.⁶

In this section, we present the VPF predicted by the mock catalogues explored in Section 4, focusing on the $M_r < -20$ sample for brevity.⁷ We compute the VPF in our mock catalogues by randomly placing 10^6 spheres of a given radius within the simulation box and

⁶ See also Conroy et al. (2005) for an investigation of the information content in the VPF that is independent from the two-point function.

⁷ Recall that the standard HOD interpretation of the clustering in the $M_r < -19$ colour-selected samples with assembly bias is already inconsistent, so the additional value of the VPF is limited in this case.

counting the fraction that are empty. We estimate errors by jackknife resampling over the eight octants of the cubical simulation volume.

Fig. 10 shows the VPFs for four mock galaxy samples with $M_r < -20$. The top panel of Fig. 10 shows results for our luminosity-only mock galaxy catalogues; VPFs in the colour-selected samples appear in the lower panels. Four curves appear in each panel. The *solid black* curve pertains to our fiducial model with assembly bias, the *dashed red* curve to the mock catalogue in which we have erased the assembly bias, but preserved the HOD. To make the remaining two curves, we have populated host haloes in the Bolshoi simulation with an HOD using the parameters of the best-fitting models determined in Section 4. The *blue dot–dashed* curves correspond to HODs fit to the clustering in our fiducial mock galaxy catalogues, the *orange dotted* curves to HOD fits to our mock catalogues without assembly bias. Jackknife-estimated error bars appear for the VPFs in our fiducial models.

In all cases, there is good agreement between the VPFs measured directly from the mock galaxy catalogues without assembly bias (*red dashed* curves) and the VPFs predicted by the HODs inferred from the clustering in these mock catalogues (*orange dotted* curves). This is a reassuring result. A significant discrepancy between these VPFs would only be due to an inadequacy of the halo model that is *not* related to assembly bias because neither of these predictions includes the assembly bias effect. Therefore, this agreement serves as an important validation exercise for halo model predictions of void statistics in the absence of assembly bias.

Next, consider the VPFs for the luminosity threshold samples in the top panel of Fig. 10. The catalogues with assembly bias and with assembly bias erased (the *black solid* and *red dashed* curves, respectively) predict VPFs that are consistent with each other given statistical uncertainties. This comparison isolates the impact of assembly bias in these catalogues on the VPF and the similarity of the two predictions suggests that assembly bias of the strength and character predicted by abundance matching would require volumes significantly larger than that of the Bolshoi simulation to be detected. The impact of assembly bias on the VPFs is relatively minor.

Moving on, consider comparing the VPFs in the luminosity threshold samples to the VPFs predicted from the HODs that best fit the clustering of these samples. The VPF predicted by the mock catalogue with abundance matching (*solid, black* curve) is in excellent agreement with the VPF predicted by the HOD that best fits the two-point clustering in this sample (*blue, dot–dashed* curve). This excellent agreement leads to a somewhat unsettling conclusion, namely, that good agreement between the VPF measured in data and the VPF predicted by an HOD model that has been fitted to the clustering of the same sample *cannot* be used to rule out the presence of significant assembly bias in the underlying galaxy sample.

Consider the middle and bottom panels of Fig. 10, in which we study the VPF in colour-selected sub-samples of our $M_r < -20$ age-matching mock catalogue. The line types are analogous to those in the top panel that we have discussed in detail in the preceding paragraphs. Again, comparing the *red, dashed* and *solid, black* curves yields the true effect of assembly bias on the VPF. For the red galaxy sub-samples, the impact of assembly bias on the VPF is substantial, and the sense of the effect is easy to understand. Age matching preferentially places red galaxies into denser environments, rendering large regions devoid of red galaxies comparably common. Erasing the assembly bias from the catalogues mixes some of these red galaxies into regions of lower density and makes large voids relatively less abundant. The converse is true for blue samples, although in this case we see that the difference this produces in the VPF is negligible.

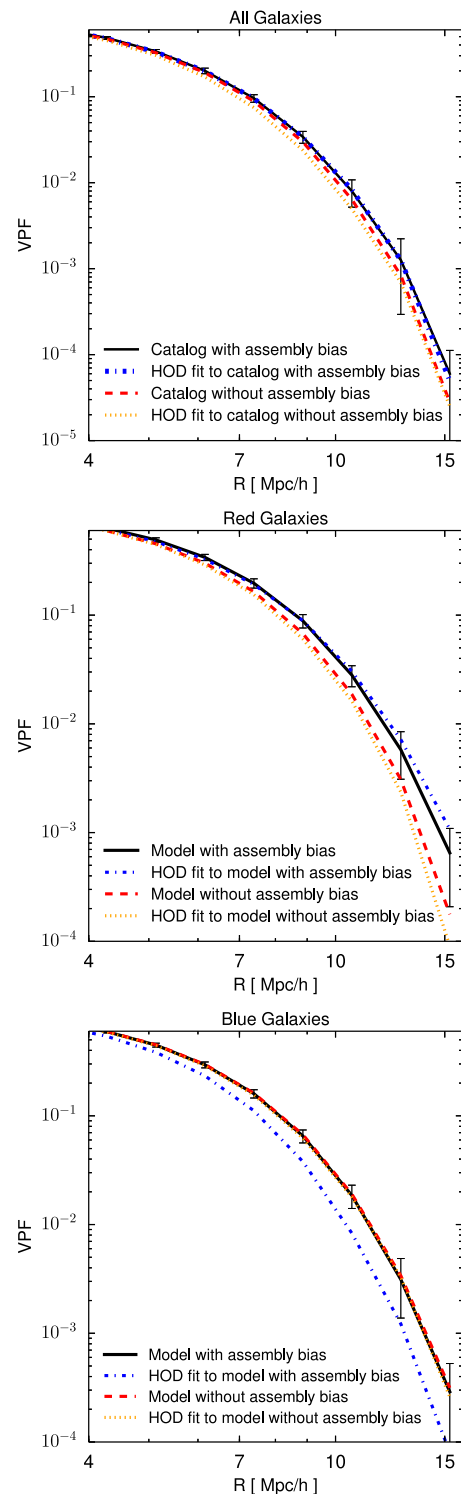


Figure 10. VPF for four of our $M_r < -20$ threshold samples. Results for full luminosity-threshold samples appear in the top panel. The two bottom panels are analogous to the top panel, but show VPFs for our colour-selected sub-samples of the $M_r < -20$ galaxy catalogues. In each panel, the points with error bars represent the VPF in our fiducial mock galaxy catalogue exhibiting galaxy assembly bias. The dashed, red lines give the VPF predicted by our mock galaxy catalogues with assembly bias erased. The dot–dashed, blue lines represent the VPFs predicted by the HOD that best fits the two-point clustering of the fiducial catalogue with assembly bias and, finally, the dotted, orange lines pertain to HOD models fit to the mock galaxy catalogues in which assembly bias has been erased.

Finally, compare the VPFs predicted by the HODs fit to the catalogues with assembly bias (*blue, dot-dashed* curves) and the VPFs measured directly from the catalogues with assembly bias (*black, solid* curves) for our colour-selected samples. For red galaxies, the difference is statistically insignificant. Evidently, the systematic shift in the HOD parameters away from their true values has counterbalanced the effect of assembly bias on the VPF, so that the assembly bias is nearly entirely disguised. This provides another explicit example of a standard HOD model which accurately describes galaxy clustering as well as the VPF, but which has parameters that are significantly biased. However, for the blue sub-samples, this is not true. The VPF predicted by the HOD that best fits the clustering of the catalogue with assembly bias differs significantly from the VPF measured directly from the catalogue with assembly bias. Recall from Section 4.3.2 that the HOD model that best fits the clustering of blue galaxies was grossly incorrect (see the middle left panel of Fig. 5). This biased HOD results in a significantly incorrect VPF, so that in this case, the VPF *does* provide some indication that the true HOD has *not* been correctly recovered.

We conclude that the assembly bias predicted by age matching may be strong enough to be detectable in VPF measurements. However, the VPF signal is neither a smoking gun signature of assembly bias nor is it necessarily an effective cross check for assembly bias. Indeed, with Fig. 10 we have given explicit examples in which assembly bias can be quite strong and yet go entirely undetected in an HOD analysis of galaxy clustering and void statistics. This is unfortunate, as it implies that the VPF alone cannot be relied upon to uncover reasonable levels of galaxy assembly bias in the data.

Our conclusion comes with the caveat that the VPF may be measured to higher precision in both existing and forthcoming dotted survey data and so, may be somewhat more discriminating than our results suggest. The errors are likely only moderately smaller for SDSS DR 7, such as those discussed above, because the effective volumes of the samples are only moderately larger than the volume of the Bolshoi simulation (Zehavi et al. 2011, see the discussion towards the end of Section 4.4). As such, our conclusions are likely to be relevant to the vast majority of extant data analyses. However, spectroscopic surveys of comparably bright galaxies over much larger volumes could result in statistical errors significantly smaller than those in Fig. 10. Conspicuous examples would be the Baryon Oscillation Spectroscopic Survey (BOSS),⁸ the bigBOSS extension,⁹ and the survey to be carried out by the Dark Energy Spectroscopic Instrument (DESI).¹⁰

5.2 The quenching of satellite galaxies

Modelling the galaxy distribution with the HOD is useful beyond the interpretation of observed statistics such as the two-point function. Once a set of parameters has been constrained by fitting to some set of observations, the HOD makes definite predictions for the imprint galaxy evolution physics leaves on halo occupation.

Consider the recent example of Tinker et al. (2013). After fitting Cosmic Evolution Survey (COSMOS) observations with a halo model, the authors demonstrate how their best-fitting HOD parameters can be used to make numerous predictions about galaxy evolution, such as the rate at which central galaxies migrate to the red sequence, and the evolution of the characteristic time-scale

of satellite quenching. In principle, predictions such as these can directly inform our understanding of galaxy evolution, as well as our modelling of the physical processes that govern star formation and quenching. In this section, we study another example of such a prediction: the host halo mass-dependence of satellite galaxy quenching. In particular, we assess the degree to which assembly bias may threaten the program to use HOD fits to projected galaxy clustering to study the impact of the parent halo on star formation in satellite galaxies.

To quantify satellite quenching, we use the *quenched fraction* of satellites, defined as the fraction of satellite galaxies that are red, $F_q^{\text{sat}} = N_{\text{sat}}^{\text{red}} / N_{\text{sat}}$. Although dust obscuration complicates the use of F_q^{sat} to quantify star formation activity (see e.g. Wetzel, Tinker & Conroy 2012), this statistic is used widely throughout the literature for this purpose (van den Bosch et al. 2008; Tinker et al. 2013; Kovac et al. 2014). Moreover, $F_q^{\text{sat}}(M_{\text{host}})$ can be readily computed from our HOD fits to colour-selected samples, permitting a direct comparison between the halo model prediction and the true quenching fraction in the mock catalogues.

Fig. 11 shows the predictions for satellite quenching made by fits to our $M_r < -20$ samples. The *points with error bars* show F_q^{sat} predicted by our mock galaxy catalogues. Note that the catalogues with and without assembly bias have *identical* satellite populations on average (they have exactly the same HODs by construction), so both of these catalogues make *identical* predictions for F_q^{sat} . The *thick, solid, red curve* in Fig. 11 shows the F_q^{sat} predictions from the

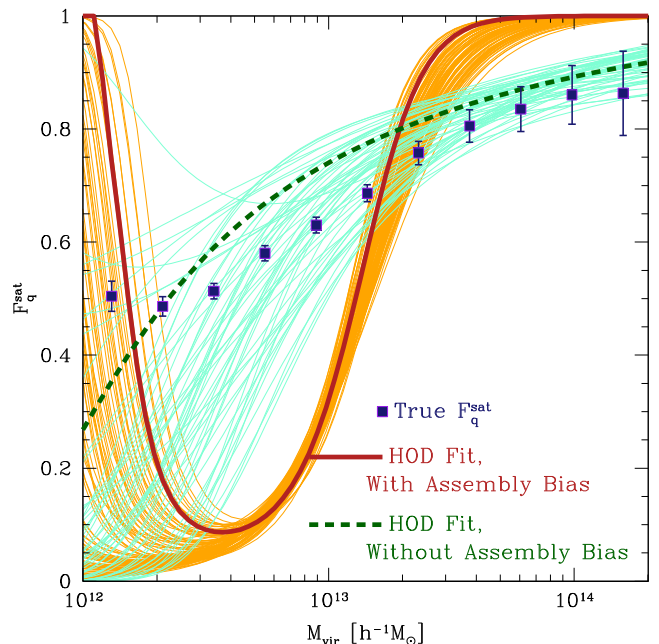


Figure 11. Host halo mass-dependence of satellite galaxy quenching efficiency. The quantity F_q^{sat} is defined as the fraction of satellite galaxies that are red, shown here as a function of the virial mass of the host halo M_{vir} . Points with error bars show satellite quenching in our age-matching catalogues with $M_r < -20$. The solid curve shows the satellite quenching prediction that would be made using the best-fitting HOD to the clustering of red and blue galaxies in our fiducial age-matching mock galaxy populations with assembly bias. The dashed curve shows F_q^{sat} predicted by the best-fitting HOD of the erased-assembly-bias counterpart of the fiducial model. In each case, the thin lines in the background represent the 100 randomly selected samples from HODs with $\Delta\chi^2 < 1$ relative to the best fit, analogous to the samples shown in Figs 4 through 6.

⁸ URL <http://www.sdss3.org/surveys/boos.php>

⁹ URL <https://bigboss.lbl.gov/>

¹⁰ URL <http://desi.lbl.gov/>

HODs fit to the clustering of our fiducial mock galaxy catalogues with assembly bias, while the *thick, dashed, green* curve shows the F_q^{sat} predicted by the best-fitting HODs to our mock galaxy catalogues with assembly bias erased. In each case, the *thinner* lines show F_q^{sat} values predicted by the 100 randomly selected HODs with $\Delta\chi^2 < 1$ in the fits to clustering shown in Fig. 4 through Fig. 6.

Comparing the HOD predictions for the fiducial catalogues with assembly bias to the actual quenching fractions in Fig. 11 illustrates clearly the following point: *if* there is significant galaxy assembly bias in the real universe that is neglected in HOD fits to galaxy clustering, then the conclusions that can be drawn about satellite quenching from such fits may be wildly incorrect. The sense of the discrepancy at high host halo masses follows directly from the HOD fits of Section 4. The catalogues with assembly bias predict stronger red galaxy clustering and weaker blue galaxy clustering on large scales. Consequently, HOD fits to those samples tend to prefer parameters that place a greater number of red satellites in strongly clustered, high-mass haloes and few blue satellite galaxies in massive haloes. The large quenching fractions in the assembly bias fits at low masses ($M_{\text{vir}} \lesssim 3 \times 10^{12} h^{-1} M_{\odot}$) occurs at values of M_{vir} for which $\langle N_{\text{sat}} \rangle \ll 1$, and so is less relevant to measurements of quenching fractions in groups and clusters and its cause is significantly more subtle. In short, the slight increase in small-scale, one-halo clustering in the blue sub-samples with assembly bias (Fig. 3) drives a preference for HODs that place satellites in relatively rarer, higher mass haloes, contrary to the demands of the large-scale clustering (for a more complete discussion of the manner in which HOD parameters affect small- and large-scale clustering, see Watson et al. 2011). This results in a relative deficit of blue satellites in low-mass host haloes and an overestimate of the quenching fraction at low host halo masses.

The quenching fraction discrepancy for the galaxy sample with assembly bias in Fig. 11 is a particularly dramatic consequence of the systematic uncertainty in HOD fits to clustering. Indeed, in the case of the $M_r < -20$, the actual offset in quenching fraction realized in any study would likely not be as dramatic as shown in Fig. 11. Recall that the HOD fit to the assembly biased mock galaxy catalogues in this case give a blue galaxy HOD that is clearly incorrect (Fig. 5). An analyst confronted with these data would likely introduce additional data and/or a prior in order to bring the blue satellite description into closer agreement with expectations. However, in this case, the prediction of the quenching fraction would rely entirely on the reliability of the extra data to constrain the satellite population and/or the prior being truly informative.

In contrast, the *dashed* curves in Fig. 11 give comparably good descriptions of the true, underlying quenching fractions, suggesting that HOD fits to clustering may perform fairly well in predicting similar quantities in the absence of galaxy assembly bias. The exception to this is at low values of host mass $M_{\text{vir}} \lesssim 10^{13} h^{-1} M_{\odot}$, particularly for the brighter sample, where the expected number of satellite galaxies is also small ($\langle N_{\text{sat}} \rangle \ll 1$, see Figs 5 and 6). The residual offsets between the mock galaxy data points and the best-fitting HOD models in this case provide an estimate of the systematic errors on quenching fractions induced by using HOD/CLF-based methods as currently implemented. The influence of halo environment on star formation activity is of central interest in the physical interpretation of astronomical observations of the galaxy distribution. Our results should therefore provide strong motivation for a comprehensive effort to model and constrain the colour-dependence of galaxy assembly bias.

6 DISCUSSION

Assembly bias, the potential for galaxies to preferentially reside in haloes in a manner that is correlated with halo clustering at fixed halo mass, has received significant attention in the recent literature both for its potential to impede our ability to draw reliable inferences from standard HOD/CLF-based statistical studies of galaxy clustering and, quite to the contrary, for its potential to be used as a signal from which new inferences about galaxy formation and evolution may be drawn. None the less, it remains unclear (1) whether or not assembly bias at detectable levels is present in the actual, observed galaxy population, and (2) the degree to which reasonable levels of assembly bias in data may affect the inferences of statistical models of the galaxy–dark matter connection that neglect assembly bias. This is a significant omission because numerous studies infer the statistical relationships between galaxies and the haloes in which they live using standard HOD/CLF-based methods that neglect assembly bias effects (e.g. Zehavi et al. 2005; Yang et al. 2005; van den Bosch et al. 2007, 2013; Zheng et al. 2007, 2009; Simon et al. 2009; Abbas et al. 2010; Watson et al. 2010; Leauthaud et al. 2011, 2012; Matsuoka et al. 2011; Miyaji et al. 2011; Geach et al. 2012; Kayo & Oguri 2012; Tinker et al. 2012, 2013; Cacciato et al. 2013; Parejko et al. 2013). The aim of this paper is to improve upon this situation by studying mock catalogues of galaxies that have been constructed to be broadly representative of the observed galaxy population, yet at the same time exhibit significant levels of assembly bias.

We chose to use halo abundance matching and age matching to build our mock galaxy catalogues to exhibit assembly bias. In Section 3, we discussed why modern abundance-matching techniques all include assembly bias, and we demonstrated that this assembly bias is, indeed, significant by showing that the projected two-point correlation functions predicted using abundance matching differs markedly from the two-point clustering predicted from mock galaxy populations with identical HODs, but no assembly bias. By itself this is an interesting point because abundance matching and age matching yield galaxy populations with clustering statistics that are broadly similar to those observed in large-scale galaxy surveys (e.g. Kravtsov et al. 2004; Tasitsiomi et al. 2004; Vale & Ostriker 2004, 2006; Conroy & Wechsler 2009; Guo et al. 2010; Neistein et al. 2011; Rodríguez-Puebla et al. 2012; Simha et al. 2012; Watson et al. 2012; Kravtsov 2013; Hearin & Watson 2013; Hearin et al. 2013b). This suggests that the assembly bias predicted by abundance matching may not be wildly different from what is observationally permissible.

Assembly bias has been shown to be an important effect in other theoretical contexts as well. Of particular relevance to this paper is the work of Croton et al. (2007). These authors studied the relative effect of assembly bias on the three-dimensional correlation functions of galaxies in SAMs of galaxy formation. They found that halo properties other than mass do influence the properties of galaxies in their models and that these effects lead to significantly altered clustering that is of roughly the same size as the assembly bias effect in abundance matching and age matching. This bolsters our case for using abundance matching to construct simple galaxy catalogues exhibiting assembly bias. Interestingly, Croton et al. (2007) also found that assembly bias could *not* be attributed only to concentration- or formation time-dependent halo clustering, suggesting that the relationship between galaxies and their haloes may be complicated enough to make empirical modelling with sufficient precision to address extant data challenging. In abundance

matching and age matching, the effects of assembly bias arise solely due to the concentrations and formation times of haloes, and this is one of the great benefits of using abundance matching as a sandbox to study assembly bias, but we cannot address these more complex realizations of assembly bias directly.

We built upon our demonstration of assembly bias in abundance matching estimating the degree to which assembly bias can represent a systematic error on the probabilities with which galaxies reside in haloes of particular masses, the HOD. In this first paper on the subject, we have chosen to limit the scope of our study by fixing cosmological parameters to the known, true, underlying cosmology of the simulation that we have used to construct our mock galaxy catalogues and studying systematic errors in HODs only. We will return to cosmological constraints in the presence of assembly bias and study additional observables in a follow-up paper.

Section 4 details our results. In short, we find that neglecting assembly bias at the levels present in our mock catalogues leads to significant systematic errors in inferred HODs (Figs 4–9) in nearly every case we have studied. For mock galaxy samples selected purely on a luminosity threshold, these systematic errors are of a modest absolute size. Systematic errors in M_{\min} and M_1 are often $\lesssim 0.2$ dex, while the offsets in α are $\lesssim 0.2$. Nevertheless, these systematic errors are significant compared to the statistical errors on the inferred HOD parameters in most cases. Systematic errors in inferred HOD parameters are significantly more severe for colour-selected sub-samples of galaxies. These results indicate that traditional HOD analyses of galaxy clustering may be entirely blind to the systematic errors caused by assembly bias.

Motivated by this, in Section 5.1 we investigated the potential to detect assembly bias using the VPF, an example of an auxiliary statistic that has been studied previously for precisely this purpose (Tinker et al. 2006, 2008a). Using our abundance-matching and age-matching mock galaxy catalogues, we constructed explicit examples in which strong levels of assembly bias leave no statistically significant imprint on the VPF, and/or would not be evident in standard HOD analyses of void statistics. This casts doubt that consistency between the observed and HOD-predicted VPF can be interpreted as ruling out assembly bias as a potential systematic (Tinker et al. 2008a). The only case in which this test identifies a problem with the HOD inferred from clustering, is the blue, $M_r < -20$ sub-sample. This case represents a dramatic failure to infer the correct HOD (see Fig. 5) and would also easily be ruled out by a number of other observables, such as group conditional mass functions. Again, this suggests that the VPF is not an especially incisive tool for identifying the effects of reasonable levels of assembly bias in inferred HODs.

Although it is not commonly discussed in the context of assembly bias, we point out that the phenomenon of galactic conformity is squarely at odds with the notion that halo mass alone determines galaxy properties. Galactic conformity refers to the observed tendency for red central galaxies to host a redder satellite population than blue central galaxies residing in haloes of the same mass (Weinmann et al. 2006). This manifestly violates the ‘halo mass only’ assumption of the standard HOD. As discussed in Section 4.2, small-scale clustering may be influenced by this phenomenon in a statistically significant way, though a more focused investigation of this point would be required before more conclusive statements could be made. We leave this as a subject for future work.

Our primary results conclude in Section 5.2 with a case study of the potential threat assembly bias poses to standard HOD studies of galaxy evolution. We presented $F_q^{\text{sat}}(M_{\text{host}})$, the halo mass-dependence of the quenched fraction of satellite galaxies, as an

example of a quantity that could be significantly misestimated from a standard HOD fit that has been compromised by assembly bias (Fig. 11). The drastic consequences that unknown levels of assembly bias may have on the relatively simple statistic F_q^{sat} is particularly interesting in light of recent analyses of COSMOS data (Tinker et al. 2013), in which standard HOD techniques are used to draw conclusions about complex characteristics of the galaxy distribution such as the characteristic quenching time-scale of satellite galaxies, or the migration rate of centrals to the red sequence. Of course, our analysis methods are not directly analogous to COSMOS analysis in Tinker et al. (2013): we have studied a different formulation of the HOD from theirs, and we have focused exclusively on galaxy clustering measurements, whereas they have included galaxy–galaxy lensing data (see below). None the less, Tinker et al. (2013) have demonstrated the potential of the HOD to provide rich information about the history of star formation in galaxies, and so the results in Section 5.2 provide strong motivation to constrain the true level of assembly bias in the data.

It has become increasingly common to fit simultaneously for statistics in addition to two-point galaxy clustering in halo model analyses. For example, many different approaches to galaxy–halo modelling have been brought to bear on galaxy–galaxy lensing measurements (e.g. Leauthaud et al. 2012; Yoo & Seljak 2012; Tinker et al. 2013; Cacciato et al. 2013; Hearin et al. 2013b). The mass-to-number ratio of clusters may also provide additional constraining power on both halo model and cosmological parameters (Tinker et al. 2012; Reddick et al. 2014). One hopes that the independent information provided by additional statistics such as these would break the degeneracies evident in Section 4. However, the results shown in Section 5.1 illustrate that even the relatively strong levels of assembly bias present in our mock catalogues can go entirely undetected in alternative statistics, such as the VPF, that naively seem well suited to this purpose. While we have limited the scope of this paper to projected two-point clustering only, it will be interesting to extend this analysis to itemize the ways in which additional statistics may mitigate systematic errors induced by assembly bias and we will pursue this avenue in a follow-up paper.

We stress that we have attempted to be conservative in our quantification of the potential systematic error induced by unknown levels of assembly bias. In particular, we have marginalized over the parameter f_b , a nuisance parameter introduced in Tinker et al. (2012) and designed to account, in part, for imperfect calibration of halo bias. This enables some of the large-scale clustering offset between mock galaxy samples with and without assembly bias to be absorbed into f_b and, indeed, this is reflected in the inferred values of f_b shown in Fig. 4 through Fig. 6. However, we emphasize that this additional parametric freedom is particularly ineffective at mitigating against assembly bias in the colour-selected samples because assembly bias increases the clustering strength of red galaxies while decreasing the large-scale clustering strength of blue galaxies and f_b cannot accommodate such countervailing demands. Moreover, the fact that assembly bias causes significant systematic errors in the luminosity threshold sample HODs suggests that assembly bias causes a scale-dependent shift in the projected correlation function that cannot be accommodated by a simple shift in large-scale clustering. In Appendix A, we give examples of how our results change when f_b is not marginalized over. In Appendix A, we also show that our results are robust to including additional parametric freedom in the radial distributions of satellites. Further, our jackknife error estimates on the clustering in our mock catalogues are significantly larger at all radii, and for all samples, than the corresponding statistical errors in, for example, the clustering measured in analogous

samples by the SDSS (Zehavi et al. 2011). Thus, if assembly bias is present in the real universe and has comparable strength to that which is present in our mock galaxy catalogues, systematic errors even more severe than what we present here would be present in the HODs inferred from galaxy clustering.

In the absence of definitive studies that constrain assembly bias to negligible levels, it seems prudent to consider inferences drawn about halo occupation statistics from large-scale clustering to be subject to a systematic error that is large compared to its statistical error. In order to mitigate the possibility that assembly bias can induce a systematic error in the inferred statistics of the galaxy distribution, it will be necessary to model assembly bias in parametrized forms and in significantly greater detail than has been attempted before. Explicit, theoretical modelling of *halo* assembly bias has been attempted before (Wechsler et al. 2006). Achieving the necessary precision will require a significant effort involving, in part, precise calibration of halo abundance and clustering as a function of halo properties other than mass.

In addition to a precision calibration of halo assembly bias, a rigorous theoretical formulation of *galaxy* assembly bias will be necessary so that analytical parameters quantifying the character and strength of this effect can be included in likelihood analyses. Indeed, it may be possible to recast the HOD in terms of only a single, distinct halo property (or combination of halo properties) y , $P(N|y)$ as described in Section 3, such that the HOD is a more faithful representation of the relationship between galaxies and their host haloes. In this case, assembly bias is a manifestation of the fact that y , rather than mass, is the halo property that is most directly related to the galaxy population within a halo. Itemizing halo abundance, clustering, and structure as a function of y would enable HOD modelling in terms of this new halo property and mitigate the systematic errors induced by assembly bias. In the mock galaxy samples that we studied in this paper, recasting the HOD and all other ingredients of the halo model in terms of V_{\max} (rather than halo mass) would have been sufficient to describe the galaxy–halo relationship in our mock luminosity threshold samples.

This challenge can be viewed as an opportunity. With the wealth of extant and forthcoming data on galaxy clustering, galaxy–galaxy lensing, and any number of other statistics, it may now be possible to cultivate and constrain a significantly richer empirical relationship between galaxies and their dark matter haloes. This may lead to models that can associate galaxies with haloes based on a number of halo properties, further bridging the gap between the vast amount of existing observational data and direct numerical simulations of galaxy formation in a cosmological context. It is our hope that this study provides motivation to pursue these goals.

Lastly, we note that our study is subject to several noteworthy caveats. First, while we know of no definitive study that rules out significant assembly bias, it is possible that assembly bias is far less prevalent in the true galaxy distribution than it is in models based upon abundance matching. We have constructed explicit examples in which assembly bias is large, induces large systematic errors, and is not easily diagnosed, but we know of no reason that assembly bias must be as large as abundance/age matching predict. Indeed, abundance matching is known to be an inadequate description of observed galaxy clustering statistics in their detail (e.g. Hearin et al. 2013a). Moreover, recent results (Behroozi et al. 2014) indicate that the property we used in our abundance matching, V_{peak} , sometimes occurs during a non-equilibrium phase of halo evolution, and so it may be implausible for V_{peak} to correlate with stellar mass to high precision. In addition, sub-halo incompleteness may also pose

a problem for detailed predictions of galaxy clustering (Wetzel & White 2010; Watson et al. 2012; Guo & White 2014), even in state-of-the-art simulations such as Bolshoi and Millennium (although see also Klypin et al. 2013). We were forced into using such an incomplete model precisely because no model exists that reproduces all of the known properties of the observed galaxy distribution. Related to these points is the fact that our covariance matrices have been estimated from the same mock galaxy catalogues that we have used in our fits. Again, this strategy was necessary because there are few high-resolution simulations available that can be used to construct mock galaxy catalogues over a wide range of luminosities using abundance matching. Lastly, the specific HOD parameterizations and priors used in previous HOD analyses vary greatly from one study to the next. In the results presented in the main body of this paper, as well as in Appendix A, we have not placed any priors on our HOD parameters. We have experimented with a variety of priors and alternative parameterizations, and while the inferred HODs are altered significantly by such choices (emphasizing the fact that priors should be informative), our qualitative conclusions are robust to these choices.

7 SUMMARY

We conclude this paper with the following summary of our primary findings.

- (i) Galaxy assembly bias of considerable strength is a generic prediction of the widely used abundance-matching technique for assigning galaxies to haloes. The same is true of the recently introduced age-matching algorithm for assigning colours to mock galaxies in dark matter simulations. Both of these methods make predictions for the observed galaxy distribution that are in good agreement with a rich variety of SDSS measurements.
- (ii) It is possible to obtain an acceptable fit to galaxy clustering data with a traditional HOD model, even when the strength of assembly bias in the galaxy sample is significant.
- (iii) Assembly bias of the kind predicted by abundance/age matching causes there to be a significant systematic error on the halo–galaxy connection inferred from fits to galaxy clustering. As a similar level of assembly bias has not yet been ruled out, the halo–galaxy connection (whether HOD, CLF, or otherwise) inferred from clustering data should be subject to an additional systematic error that is large compared to statistical errors.
- (iv) The VPF may be useful in constraining the colour-dependence of galaxy assembly bias, but we have constructed explicit examples in which the VPF cannot detect assembly bias even when the systematic errors in HOD parameters derived from galaxy clustering fits are large.
- (v) Uncertainty in the true level of galaxy assembly bias can have a dramatic effect on HOD modelling of the star formation histories of satellite galaxies and may even dominate the error budget in these applications.
- (vi) For future studies of the galaxy–halo connection, including (re)analyses of existing data sets, we recommend a comprehensive effort to model and constrain the true level of galaxy assembly bias, both for colour-selected galaxy samples and samples selected purely on luminosity. To aid this effort, we make publicly available all of the mock catalogues used in this study; these mock catalogues were specifically designed to isolate the effects on the galaxy distribution that are purely due to assembly bias, and can be found at <http://logrus.uchicago.edu/~aphearin>.

ACKNOWLEDGEMENTS

We thank Andreas Berlind, Shaun Cole, Hiram Coombs, Carlos Frenk, Jeff Newman, Risa Wechsler, Idit Zehavi, Zheng Zheng, Ramin Skibba, and particularly Doug Watson for useful discussions throughout various stages of this work. We also thank David Weinberg and Simon White for insightful email exchanges regarding an earlier draft of this manuscript. We thank Kristin Riebe for helping us navigate the MultiDark data base. We are particularly grateful to Jeremy Tinker and Rachel Reddick for comparing the results of their halo model code to our own and for helping us with numerous technical questions related to implementations of the halo model and the HOD. We thank John Fahey for *America*. The work of ARZ is supported by the US National Science Foundation through grant AST 1108802 and by the University of Pittsburgh. Significant portions of this work were completed during visits to The Institute for the Physics and Mathematics of the Universe (IPMU) at the University of Tokyo and we are thankful to IPMU and particularly Alexie Leauthaud, Surhud More, and Rie Ujita for their hospitality. The work of ARZ was also supported by the National Science Foundation under grant PHYS-1066293 and the hospitality of the Aspen Center for Physics. APH is supported by the US Department of Energy under contract no. DE-AC02-07CH11359.

REFERENCES

- Abbas U. et al., 2010, *MNRAS*, 406, 1306
 Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, *ApJ*, 762, 109
 Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, *ApJ*, 763, 18
 Behroozi P. S., Wechsler R. H., Lu Y., Hahn O., Busha M. T., Klypin A., Primack J. R., 2014, *ApJ*, 787, 156
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
 Blanton M. R., Berlind A. A., 2007, *ApJ*, 664, 791
 Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, *Nature*, 311, 517
 Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, *ApJ*, 379, 440
 Cacciato M., van den Bosch F. C., More S., Mo H., Yang X., 2013, *MNRAS*, 430, 767
 Colless M. et al., 2001, *MNRAS*, 328, 1039
 Conroy C., Wechsler R. H., 2009, *ApJ*, 696, 620
 Conroy C. et al., 2005, *ApJ*, 635, 990
 Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
 Cooper M. C., Gallazzi A., Newman J. A., Yan R., 2010, *MNRAS*, 402, 1942
 Croton D. J., Gao L., White S. D. M., 2007, *MNRAS*, 374, 1303
 Dalal N., White M., Bond J. R., Shirokov A., 2008, *ApJ*, 687, 12
 Dalal N., Lithwick Y., Kuhlen M., 2010, preprint ([arXiv:1010.2539](https://arxiv.org/abs/1010.2539))
 Faltenbacher A., White S. D. M., 2010, *ApJ*, 708, 469
 Gao L., White S. D. M., 2007, *MNRAS*, 377, L5
 Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66
 Geach J. E., Sobral D., Hickox R. C., Wake D. A., Smail I., Best P. N., Baugh C. M., Stott J. P., 2012, *MNRAS*, 426, 679
 Gottloeber S., Klypin A., 2008, preprint ([arXiv:0803.4343](https://arxiv.org/abs/0803.4343))
 Guo Q., White S., 2014, *MNRAS*, 437, 3228
 Guo Q., White S., Li C., Boylan-Kolchin M., 2010, *MNRAS*, 404, 1111
 Hearin A. P., Watson D. F., 2013, *MNRAS*, 435, 1313
 Hearin A. P., Zentner A. R., Berlind A. A., Newman J. A., 2013a, *MNRAS*, 433, 659
 Hearin A. P., Watson D. F., Becker M. R., Reyes R., Berlind A. A., Zentner A. R., 2013b, preprint ([arXiv:1310.6747](https://arxiv.org/abs/1310.6747))
 Kayo I., Oguri M., 2012, *MNRAS*, 424, 1363
 Klypin A. A., Trujillo-Gomez S., Primack J., 2011, *ApJ*, 740, 102
 Klypin A., Prada F., Yepes G., Hess S., Gottloeber S., 2013, preprint ([arXiv:1310.3740](https://arxiv.org/abs/1310.3740))
 Kovac K. et al., 2014, *MNRAS*, 438, 717
 Kravtsov A. V., 2013, *ApJ*, 764, L31
 Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, *ApJS*, 111, 73
 Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottloeber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
 Lacerna I., Padilla N., 2011, *MNRAS*, 412, 1283
 Lacerna I., Padilla N., 2012, *MNRAS*, 426, L26
 Lacey C., Cole S., 1993, *MNRAS*, 262, 627
 Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, *ApJ*, 738, 45
 Leauthaud A. et al., 2012, *ApJ*, 744, 159
 Li Y., Mo H. J., Gao L., 2008, *MNRAS*, 389, 1419
 Magliocchetti M., Porciani C., 2003, *MNRAS*, 346, 186
 Mandelbaum R., Slosar A., Baldauf T., Seljak U., Hirata C. M., Nakajima R., Reyes R., Smith R. E., 2013, *MNRAS*, 432, 1544
 Matsuoka Y., Masaki S., Kawara K., Sugiyama N., 2011, *MNRAS*, 410, 548
 Miyaji T., Krumpe M., Coil A. L., Aceves H., 2011, *ApJ*, 726, 83
 Mo H. J., White S. D. M., 1996, *MNRAS*, 282, 347
 More S., van den Bosch F. C., Cacciato M., Mo H. J., Yang X., Li R., 2009, *MNRAS*, 392, 801
 More S., van den Bosch F. C., Cacciato M., More A., Mo H., Yang X., 2013, *MNRAS*, 430, 747
 Nagai D., Kravtsov A. V., 2005, *ApJ*, 618, 557
 Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
 Neistein E., Weinmann S. M., Li C., Boylan-Kolchin M., 2011, *MNRAS*, 414, 1405
 Parejko J. K. et al., 2013, *MNRAS*, 429, 98
 Press W. H., Schechter P., 1974, *ApJ*, 187, 425
 Pujol A., Gaztañaga E., 2014, *MNRAS*, 442, 1930
 Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, *ApJ*, 771, 30
 Reddick R., Tinker J., Wechsler R., Lu Y., 2014, *ApJ*, 783, 118
 Riebe K. et al., 2013, *Astron. Nachr.*, 334, 691
 Rodríguez-Puebla A., Drory N., Avila-Reese V., 2012, *ApJ*, 756, 2
 Ross A. J., Percival W. J., Brunner R. J., 2010, *MNRAS*, 407, 420
 Simha V., Weinberg D., Dave R., Fardal M., Katz N., Oppenheimer B. D., 2012, *MNRAS*, 423, 3458
 Simon P., Hettterscheidt M., Wolf C., Meisenheimer K., Hildebrandt H., Schneider P., Schirmer M., Erben T., 2009, *MNRAS*, 398, 807
 Skibba R. A., Sheth R. K., 2009, *MNRAS*, 392, 1080
 Tasitsiomi A., Kravtsov A. V., Wechsler R. H., Primack J. R., 2004, *ApJ*, 614, 533
 Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, *ApJ*, 631, 41
 Tinker J. L., Weinberg D. H., Warren M. S., 2006, *ApJ*, 647, 737
 Tinker J. L., Conroy C., Norberg P., Patiri S. G., Weinberg D. H., Warren M. S., 2008a, *ApJ*, 686, 53
 Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottloeber S., Holz D. E., 2008b, *ApJ*, 688, 709
 Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottloeber S., 2010, *ApJ*, 724, 878
 Tinker J. L. et al., 2012, *ApJ*, 745, 16
 Tinker J. L., Leauthaud A., Bundy K., George M. R., Behroozi P., Massey R., Rhodes J., Wechsler R., 2013, *ApJ*, 778, 93
 Vale A., Ostriker J. P., 2004, *MNRAS*, 353, 189
 Vale A., Ostriker J. P., 2006, *MNRAS*, 371, 1173
 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., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, *MNRAS*, 387, 79
 van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, *MNRAS*, 430, 725
 Wang Y., Yang X., Mo H. J., van den Bosch F. C., Weinmann S. M., Chu Y., 2008, *ApJ*, 687, 919
 Wang H., Mo H. J., Jing Y. P., 2009, *MNRAS*, 396, 2249
 Wang L., Weinmann S. M., De Lucia G., Yang X., 2013, *MNRAS*, 433, 515
 Watson D. F., Berlind A. A., McBride C. K., Masjedi M., 2010, *ApJ*, 709, 115
 Watson D. F., Berlind A. A., Zentner A. R., 2011, *ApJ*, 738, 22
 Watson D. F., Berlind A. A., Zentner A. R., 2012, *ApJ*, 754, 90

- Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, *ApJ*, 568, 52
- Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, *ApJ*, 652, 71
- Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, *MNRAS*, 366, 2
- Wetzel A. R., White M., 2010, *MNRAS*, 403, 1072
- Wetzel A. R., Tinker J. L., Conroy C., 2012, *MNRAS*, 424, 232
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Yang X., Mo H. J., van den Bosch F. C., 2003, *MNRAS*, 339, 1057
- Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., 2005, *MNRAS*, 358, 217
- Yang X., Mo H. J., van den Bosch F. C., 2006, *ApJ*, 638, L55
- Yoo J., Seljak U., 2012, *Phys. Rev. D*, 86, 083504
- Zehavi I. et al., 2005, *ApJ*, 630, 1
- Zehavi I. et al., 2011, *ApJ*, 736, 59
- Zentner A. R., 2007, *Int. J. Mod. Phys. D*, 16, 763
- Zentner A. R., Bullock J. S., 2003, *ApJ*, 598, 49
- Zentner A. R., Berlind A. A., Bullock J. S., Kravtsov A. V., Wechsler R. H., 2005, *ApJ*, 624, 505
- Zheng Z. et al., 2005, *ApJ*, 633, 791
- Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, 667, 760
- Zheng Z., Zehavi I., Eisenstein D. J., Weinberg D. H., Jing Y. P., 2009, *ApJ*, 707, 554

APPENDIX A: ADDITIONAL NUISANCE PARAMETERS AND HOD RECOVERY

In the main body of the text, we explored the fidelity of the HOD recovered by fitting the projected two-point functions of a variety of mock galaxy samples. In any such fit, a variety of choices must be made regarding the parameters that are allowed to vary in order to describe the clustering. In Section 4, we presented results in which we marginalized over the halo bias, in order to account for imperfect calibration of host halo clustering, and held the spatial distributions of satellite galaxies fixed. In this appendix, we provide examples of how our results change in detail when we alter these assumptions. However, we emphasize that our results do not change qualitatively. In particular, that the inferred HODs are significantly different in fits to samples with and without assembly bias is a robust conclusion.

As a first example, we show the effect of the marginalization over the halo bias nuisance parameter f_b . In particular, Fig. A1 shows the HODs inferred from fits to the projected two-point functions of the $M_r < -20$ samples with the halo bias parameter held fixed at $f_b = 1$. This figure should be compared to Fig. 5 in the main text in order to assess the influence of the bias parameter. Note several things about Fig. A1. First, note that the span of models that provide similarly acceptable fits to the galaxy clustering is narrower in this case, as should be expected because there is less parameter freedom. Secondly, note that the HODs are recovered with similar fidelity in the fits to the samples with no assembly bias. Finally, compared to the fits to the galaxy samples without assembly bias, the fits to the samples with assembly bias exhibit the same systematic differences in the inferred HODs. This suggests that our qualitative conclusions are robust to relatively small uncertainties in the calibration of the halo bias and that the scale-dependence of the clustering is sufficiently different in the models with and without assembly bias as to drive significant differences in the inferred HODs. Fig. A2 depicts the marginalized HOD parameter constraints for the models with $M_r < -20$ and $f_b = 1$. These constraints exhibit the same fundamental trends as described in the main body of the this paper.

We now move on to fits in which we have allowed the spatial distributions of the satellite galaxies to vary during the HOD fitting. This has been done in a number of recent publication in

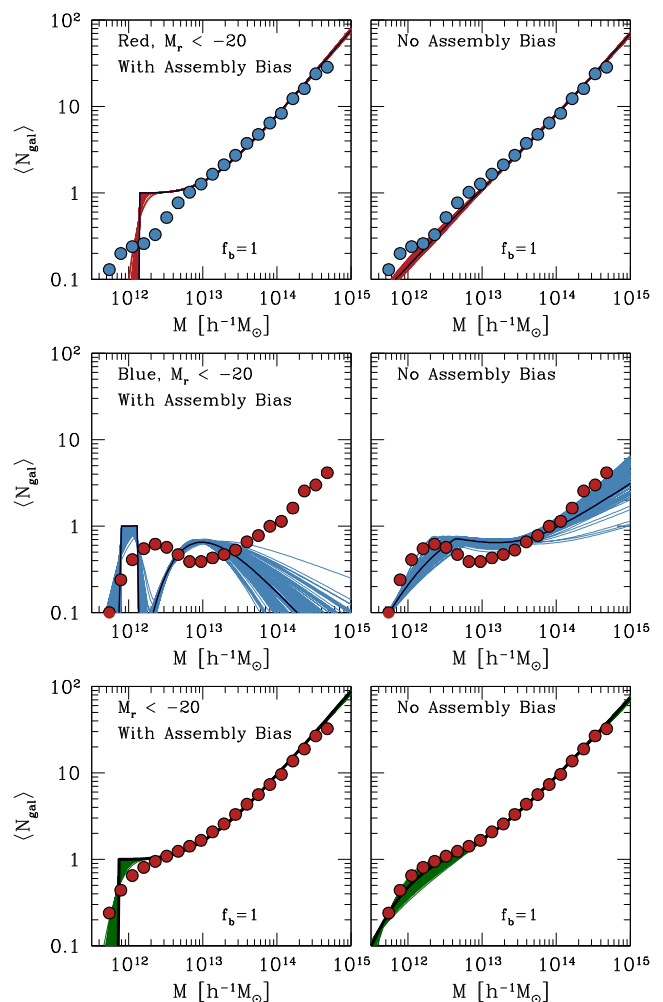


Figure A1. Comparison of the best-fitting HODs for galaxies in the $M_r < -20$ sample with the true HOD in the simulation (points). This figure is the same as Fig. 5 except that these fits were conducted with the halo bias parameter held fixed to $f_b = 1$.

which small-scale galaxy clustering has been fitted with similar models (e.g. Tinker et al. 2012; van den Bosch et al. 2013; Reddick et al. 2014). There are at least two reasons for introducing additional parameter freedom to describe the spatial distributions of the satellite galaxies. Satellite galaxies may not necessarily follow the dark matter distribution and, indeed, neither satellite haloes nor satellite galaxies trace the overall dark matter distributions of their host haloes in simulations (e.g. Zentner & Bullock 2003; Nagai & Kravtsov 2005; Zentner et al. 2005). Moreover, satellite haloes may be distributed in a triaxial configuration about their host haloes and there is some hope that introducing a satellite distribution nuisance parameter can account for this effect without modelling triaxiality directly (though this remains to be demonstrated explicitly). We follow the recent literature and introduce an additional parameter defined to be the ratio of the NFW concentration assumed for the spatial distribution of satellite galaxies to the NFW concentration of the dark matter, $f_{\text{conc}} = c_{\text{sats}}/c_{\text{dm}}$, where c_{dm} is the standard dark matter concentration from the Bolshoi simulation (Klypin et al. 2011b) and c_{sats} is the concentration parameter used to describe the average radial distribution of satellite galaxies. In our standard fits described in Section 4, we held this parameter fixed to $f_{\text{conc}} = 0.6$.

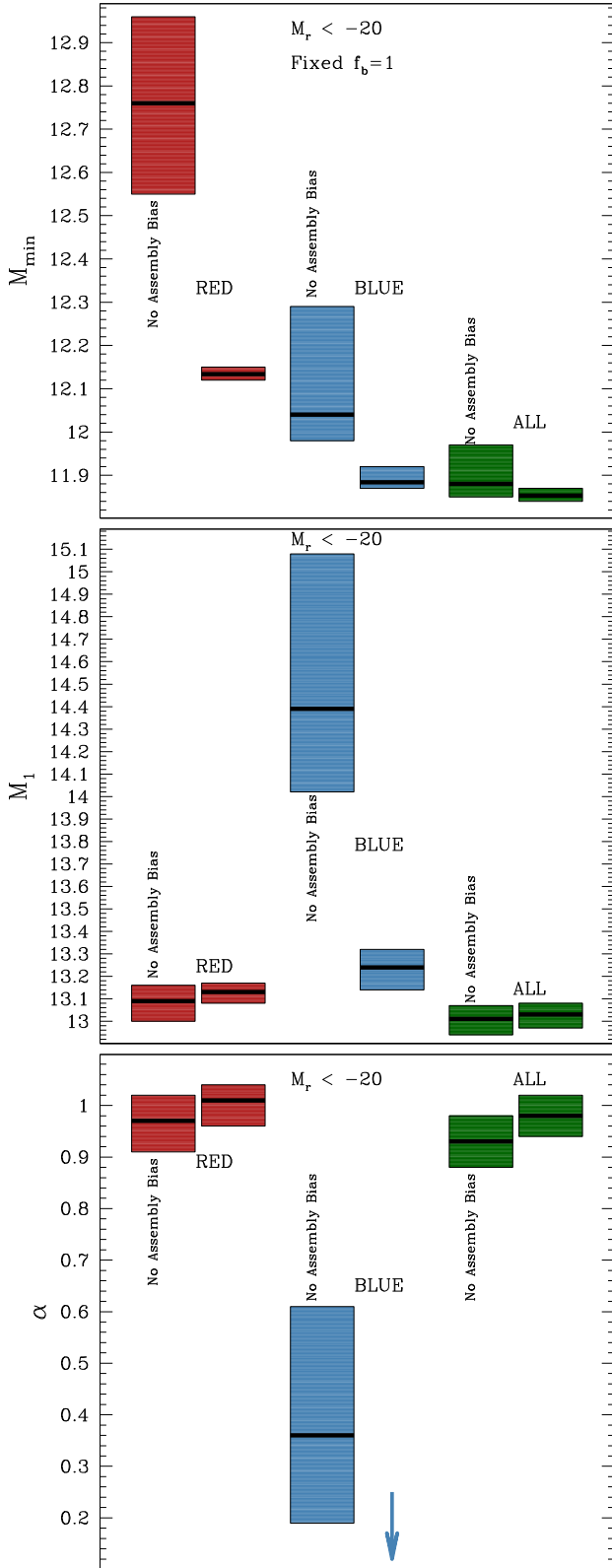


Figure A2. Constraints on HOD parameters inferred from fits with in which the bias nuisance parameter has been held fixed at $f_b = 1$, which corresponds to assuming perfect calibration of halo bias. The top panel shows the inferred constraints on M_{\min} , the middle panel shows inferred constraints on M_1 , and the bottom panel shows inferred constraints on the power-law index of the satellite portion of the HOD, α .

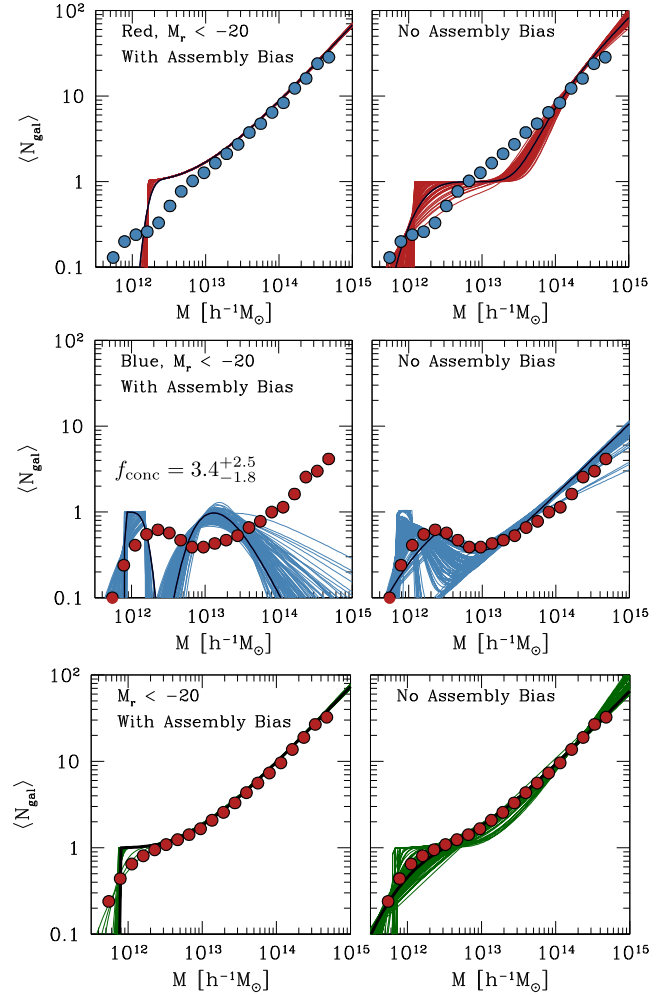


Figure A3. Comparison of the best-fitting HODs for galaxies in the $M_r < -20$ sample with the true HOD in the simulation (points). This figure is the same as Fig. 5 except that these fits were conducted while simultaneously allowing the concentrations of the satellite galaxy distributions to vary. The constraints on f_{conc} are shown in each panel and this parameter is generally poorly constrained.

In this section, we allow f_{conc} to vary while holding $f_b = 1$ for simplicity.

Fig. A3 shows the shift in the inferred HODs when the concentrations of the satellite galaxy distributions were allowed to vary simultaneously with the HOD parameters. In the case of the colour fits, we allowed the blue and red galaxies to have distinct values of the concentration parameter f_{conc} . The results shown in Fig. A3 make clear that including the additional parameter freedom from varying satellite galaxy concentrations does not eliminate the qualitative biases in the inferred HODs that we report in this paper. In fact, new biases are introduced due to the significant degeneracies that exist between the HOD parameters and the galaxy concentration parameters. The sense of the bias can be gleaned by comparing Fig. A3 to either Fig. A1 in this appendix or Fig. 5 in Section 4. Allowing concentrations to vary tends to drive an additional offset in inferred HODs such that satellites become abundant in relatively higher mass haloes in the samples with no assembly bias and vice versa in samples with assembly bias. This is a relatively subtle effect in the luminosity threshold sample (bottom panel of Fig. A3), but it is more evident in the red sub-sample (top panel of Fig. A3). This

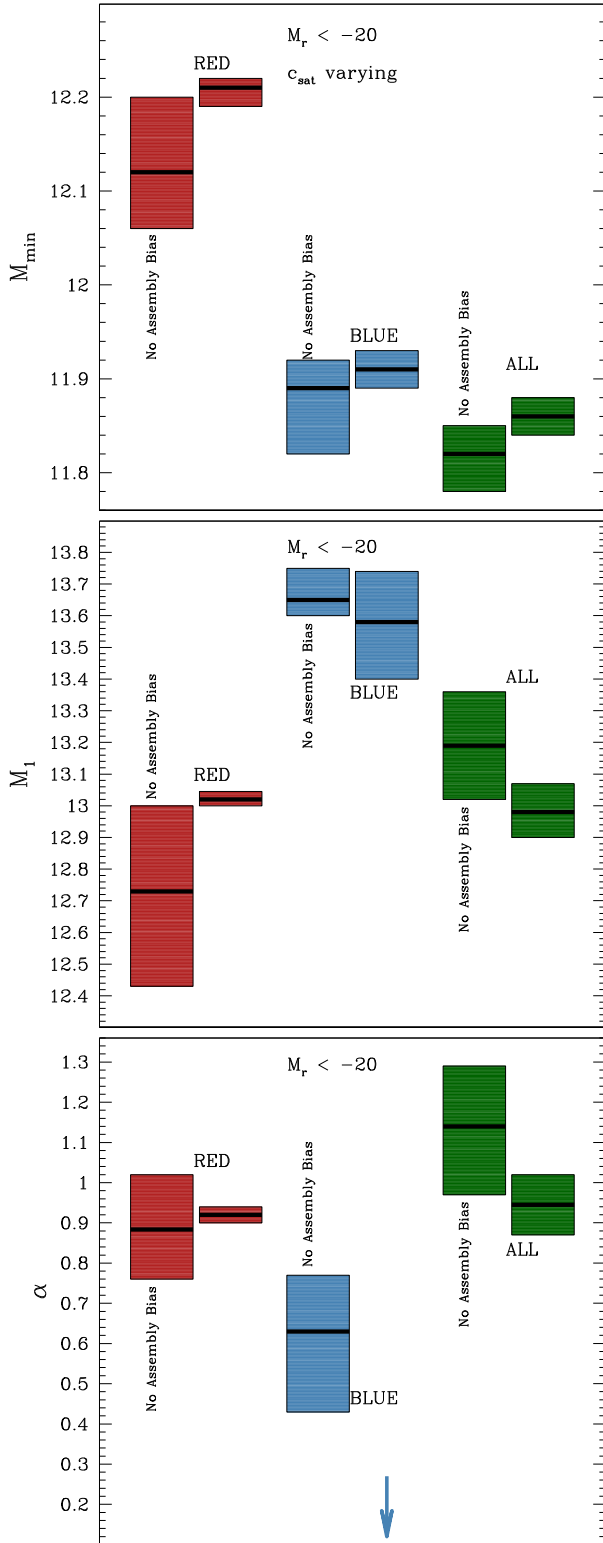


Figure A4. Constraints on HOD parameters inferred from fits with in which the concentration parameter describing the spatial distributions of satellite galaxies are allowed to fit. The top panel shows the inferred constraints on M_{\min} , the middle panel shows inferred constraints on M_1 , and the bottom panel shows inferred constraints on the power-law index of the satellite portion of the HOD, α . Fitting for concentrations in addition to the standard HOD parameters alters the fits notably (compare to Figs 7–9), but does not alter the qualitative point that assembly bias introduces additional errors in HOD parameter inferences.

demonstration suffices to show that varying the satellite galaxy spatial distributions within hosts does not change the qualitative conclusions of our paper that assembly bias can significantly bias inferred HODs from galaxy clustering. While it is possible to explore the degeneracies between satellite galaxy concentration and HOD parameters more thoroughly, such an exploration would be quite complex and we place it beyond the scope of this work.

We depict the marginalized constraints on the HOD parameters in Fig. A4 for the $M_r < -20$ samples. While the constraints shift systematically from our fiducial case, notice that the constraints are offset from each other significantly due only to the effect of assembly bias. Indeed, the extra parameter freedom afforded by fitting concentrations partially overcompensates for the differences between the samples with and without assembly bias and causes the HOD constraints in the threshold samples not split on colour to be *more* significantly offset from each other, rather than less. The conclusion is the same and not surprising. The satellite galaxy concentrations cannot serve as a nuisance parameter to guard against assembly bias effects and, indeed, may exacerbate systematic errors in inferred HODs induced by assembly bias.