Separate Universe calibration of the dependence of halo bias on cosmic web anisotropy

Sujatha Ramakrishnan\textsuperscript{1⋆}, Aseem Paranjape\textsuperscript{1†}

\textsuperscript{1} Inter-University Centre for Astronomy & Astrophysics, Ganeshkhind, Post Bag 4, Pune 411007, India

\begin{abstract}
   We use the Separate Universe technique to calibrate the dependence of linear and quadratic halo bias $b_1$ and $b_2$ on the local cosmic web environment of dark matter haloes. We do this by measuring the response of halo abundances at fixed mass and cosmic web tidal anisotropy $\alpha$ to an infinite wavelength initial perturbation. We augment our measurements with an analytical framework developed in earlier work which exploits the near-Lognormal shape of the distribution of $\alpha$ and results in very high precision calibrations. We present convenient fitting functions for the dependence of $b_1$ and $b_2$ on $\alpha$ over a wide range of halo mass for redshifts $0 \leq z \leq 1$. Our calibration of $b_2(\alpha)$ is the first demonstration to date of the dependence of non-linear bias on the local web environment. Motivated by previous results which showed that $\alpha$ is the primary indicator of halo assembly bias for a number of halo properties beyond halo mass, we then extend our analytical framework to accommodate the dependence of $b_1$ and $b_2$ on any such secondary property which has, or can be monotonically transformed to have, a Gaussian distribution. We demonstrate this technique for the specific case of halo concentration, finding good agreement with previous results. Our calibrations will be useful for a variety of halo model analyses focusing on galaxy assembly bias, as well as analytical forecasts of the potential for using $\alpha$ as a segregating variable in multi-tracer analyses.

\textbf{Key words:} cosmology: theory, dark matter, large-scale structure of the Universe – methods: numerical
\end{abstract}

\section{Introduction}

The large-scale clustering of gravitationally bound haloes of dark matter is a key variable in understanding the formation and evolution of the large-scale structure of the Universe (see Desjacques et al. 2018, for a review). This ‘halo bias’ is known to depend on a number of halo properties such as halo mass (Kaiser 1984; Bardeen et al. 1986; Mo & White 1996; Sheth & Tormen 1999), halo assembly history (Sheth & Tormen 2004; Gao et al. 2005; Wechsler et al. 2006), halo shape, angular momentum and kinematics (Faltenbacher & White 2010) and the local tidal environment (Shen et al. 2006; Hahn et al. 2009; Borzyszkowski et al. 2017; Paranjape et al. 2018; Ramakrishnan et al. 2019). The dependence of halo bias on secondary properties beyond halo mass, generically referred to as ‘halo assembly/secondary bias’, has emerged as a robust prediction of the hierarchical $\Lambda$-cold dark matter (ΛCDM) structure formation paradigm. Typically, halo assembly bias in some halo property $c$ (such as concentration, age, spin, ellipticity, velocity anisotropy, etc.) manifests as a difference in mean bias, at fixed halo mass, between halo populations having large and small values of $c$. Although there has been some analytical progress in describing such trends using simplified models (see, e.g., Zentner 2007; Dalal et al. 2008; Desjacques 2008; Musso & Sheth 2012; Castorina & Sheth 2013), many of these trends show complex behavior, e.g. when multiple secondary variables are studied simultaneously (Lazeyras et al. 2017; Xu & Zheng 2018; Han et al. 2019). A detailed understanding of halo assembly bias from first principles is therefore currently an open problem.

On another front, if the physics of galaxy formation and evolution couples tightly to the mass accretion history of dark matter haloes (White & Rees 1978) – as is routinely assumed in semi-analytic models (SAMs) of galaxy evolution (e.g., Henriques et al. 2015) as well as (sub)-halo abundance matching (SHAM) exercises (Reddick et al. 2013; Hearin & Watson 2013; Zehavi et al. 2019; Contreras et al. 2020) and also confirmed by cosmological hydrodynamical simulations (Chaves-Montero et al. 2016; Bray et al. 2016; Montero-Dorta et al. 2020) – then one expects galaxy assem-
ably bias trends to be apparent in observed galaxy samples. Due to systematic uncertainties in cleanly segregating observed samples, however, such trends have been difficult to establish robustly, with many conflicting results (Lin et al. 2016; Miyatake et al. 2016; Zentner et al. 2016; Montero-Dorta et al. 2017; Tinker et al. 2017; Zu et al. 2017; Tojeiro et al. 2017). A unified framework to understand halo and galaxy assembly bias is therefore currently lacking.

Some recent developments are noteworthy in this context. Studies using dark matter only $N$-body simulations have demonstrated that the local tidal environment of haloes plays a key role in explaining many (if not most) of the halo assembly bias trends studied in the literature. The tidal environment of a halo can be conveniently quantified by the tidal anisotropy $\alpha$ constructed using the tidal tensor of the cosmic web in the vicinity of the halo (Paranjape et al. 2018, see below for details). This variable has been shown to have the strongest correlation with large-scale bias amongst a number of secondary halo properties, and also statistically explains the assembly bias of all these properties (Ramakrishnan et al. 2019). The origins of some of these correlations, such as those between $\alpha$ and the halo age, concentration and velocity anisotropy, can be understood in terms of the dynamics of mass accretion as revealed by using high-resolution zoom simulations of objects accreting in and outside cosmic filaments (Hahn et al. 2009; Borzyszkowski et al. 2011; Li et al. 2016) which provides an exact realization of the peak-background split (Lazeyras et al. 2016). Moreover, when augmented by some basic analytical modeling of the statistical distribution of the underlying variables, the SU technique can provide unprecedented precision in calibrating the relation between large-scale clustering and small-scale halo properties, such as analytical halo models of assembly bias, generating mock halo catalogs with accurate halo assembly bias using low-resolution simulations, forecasting tracer-cosmological constraints, etc., some of which we will discuss below.

The paper is organized as follows. Section 2 describes the SU simulations and halo properties used in this work. In Section 3, we present our calibration of the dependence of $b_1$ and $b_2$ on the tidal anisotropy $\alpha$. In Section 4, we extend the analytical framework mentioned above to include the dependence on both $\alpha$ and a secondary variable $c$ in $b_1$ and $b_2$, focusing on halo concentration as a specific example. We conclude in Section 5. The Appendices present some technical details and calculations relevant to the main text.

2 SIMULATIONS AND HALO PROPERTIES

2.1 Separate Universe simulations

The peak-background split halo bias parameters are defined in terms of the derivative of the mean physical density of haloes with respect to the infinite wavelength density perturbation, i.e., as response coefficients. The response of halo number density to the presence of such a perturbation in a local region of the fiducial FLRW universe is identical to that produced in a universe with a modified cosmology having larger/smaller physical background density depending on the sign of the perturbation. If we denote the infinite wavelength perturbation linearly extrapolated to present day as $\delta_L$, then in practice the SU technique takes a fiducial universe with $\delta_L \neq 0$ and performs an exact mapping to a curved universe with a different spatial curvature, matter density parameter and Hubble constant, all determined by the value of $\delta_L$. We refer the reader to Wagner et al. (2015b) for details of the numerical implementation of $\delta_L \to$ FLRW mapping in $N$-body simulations.

In the following, we give a few details regarding the simulations, halo identification and cleaning procedure, which are identical to Paranjape & Padmanabhan (2017).\(^1\) Hence we refer the reader to the same for a more elaborate discussion. For our fiducial cosmology, we use a flat $\Lambda$CDM model with total matter density parameter $\Omega_m = 0.276$, baryonic matter density parameter $\Omega_b = 0.045$, Hubble constant $H_0 = 100h\text{km/s/Mpc}$ with $h = 0.7$, primordial scalar spectral index $n_s = 0.961$ and $\sigma_8 = 0.811$. Our $N$-body simulations are performed using GADGET-2 (Springel 2005).\(^2\) All the simulations have a box size $L_{\text{box}} = 300/0.7$ Mpc and a particle count of 512\(^3\) each. In addition to the fiducial cosmology, we also use a set of simulations generated with SU technique that correspond to $\delta_L \in \{\pm 0.7, \pm 0.5, \pm 0.4, \pm 0.3, \pm 0.2, \pm 0.1, \pm 0.07, \pm 0.05, \pm 0.02, \pm 0.01, \pm 0.15, \pm 0.25, \pm 0.35\}$. Since physical matter density parameter $\Omega_m h^2$ is constant in all the boxes, the particle mass $m_{\text{part}} = 2.2 \times 10^{10} M_{\odot}$ is the same in all simulations. We have 15 sets of simulations for each $\delta_L$ performed by changing the seed for the random initial conditions, while keeping the seed the same across all $\delta_L$ values in each set. Additionally, 10 realizations of higher resolution (1024\(^3\) particles) $\delta_L = 0$ boxes are also used in order to test for convergence of various quantities computed.

Haloes are identified using ROCKSTAR (Behroozi et al. 2013)\(^3\) which uses a 6-dimensional Friends-of-Friends algorithm to make catalogs of haloes and their properties. From the catalog, only parent haloes are chosen so that the analysis is unaffected by the effects of substructure. Haloes were chosen to have a minimum of 400 particles. Unrelaxed haloes with ‘virial ratio’ $2T/U \geq 2$ are removed from our analysis (see Bett et al. 2007, for a detailed discussion).

In the SU approach, the fiducial universe at redshift $z$ is mapped to a universe with modified cosmology at $\tilde{z}$ and their background densities are related by

\[ \tilde{\rho}(t) = \rho(t) \left( \frac{1 + \tilde{z}}{1 + z} \right)^3. \]

\(^1\) https://bitbucket.org/aparanjape/separateuniversescripts
\(^2\) http://www.mpa-garching.mpg.de/gadget/
\(^3\) https://bitbucket.org/gfcstanford/rockstar
The overdensity of haloes in a $\delta_L \neq 0$ Lagrangian patch is given in terms of differential number density $n(m, \delta_L)$ of haloes between masses $(m, m + dm)$ as follows

$$\delta_L^m(m, \delta_L) \equiv \frac{n(m, \delta_L)}{n(m|\delta_L = 0)} - 1. \quad (2)$$

It can also be related to the underlying dark matter distribution in terms of bias coefficients $b_L^c(m)$ in this manner

$$\delta_L^m(m, \delta_L) = \sum_{n=1}^{\infty} \frac{b_L^c(m)}{n!} \delta_L^n. \quad (3)$$

We will equate the RHS of equations (2) and (3). We have several $\delta_L \neq 0$ simulation boxes each having same number of particles of same particle mass, hence all SU simulations will have identical Lagrangian volume. Thus the numerator and denominator in equation (2) can be replaced by just number count of haloes between mass $(m, m + dm)$ inside our simulation boxes. After computing RHS of equation (2) for all simulations, an average over all 15 realizations is done. The mean and standard deviation of the above average is collected and used to perform a quartic polynomial fit for $\delta_L^m(m, \delta_L)$ as a function of $\delta_L$. Thus, from the best fit values of first and second-order, we obtain the linear and quadratic Lagrangian bias $b_L^c$ and $b_L^2$; the error on these coefficients are obtained from the square root of diagonal elements of the covariance matrix recovered from the fit.

The corresponding Eulerian parameters $b_n$ can be obtained from the relation $(1 + \delta_L^k)(1 + \delta) = 1 + \sum_{n=1}^{\infty} (b_n/n!)|\delta|^n$ (Mo & White 1996) by substituting into it the approximate nonlinear $\delta_L$ derived from spherical evolution: $\delta = \delta_L g(z) + (17/21) \delta_L^2 g(z)^2 + O(\delta_L^3)$ (Bernardeau 1992; Wagner et al. 2015a)

$$b_1 = 1 + b_L^2 g(z)^{-1},$$

$$b_2 = b_L^2 g(z)^{-2} + \frac{8}{21} b_L^4 g(z)^{-1}. \quad (4)$$

Here $g(z)$ $\equiv D(z)/D(0)$ and $D(z)$ is the linear theory growth factor of the fiducial cosmology.

2.3 Local cosmic web environment of haloes

We use the tidal anisotropy variable $\alpha$ introduced by Paranjape et al. (2018) to quantify the halo's nonlinear local environment. We construct this from the eigen values $\lambda_1, \lambda_2, \lambda_3$ of the tidal tensor $\psi_{ij} \equiv \partial^2 \psi/\partial x_i \partial x_j$, where $\psi$ satisfies the normalised Poisson equation $\nabla^2 \psi = \delta$. The halo centric $\alpha$ is then defined as

$$\alpha = \sqrt{q^2/(1 + \delta)}, \quad (5)$$

where $q^2$ and $\delta$ are the halo-centric tidal shear (Heavens & Peacock 1988; Catelan & Theuns 1996) and overdensity respectively,

$$q^2 = \frac{1}{2}[(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2], \quad (6)$$

$$\delta = \lambda_1 + \lambda_2 + \lambda_3. \quad (7)$$

The tidal anisotropy parameter $\alpha$ is in general a proxy for the anisotropy of the environment of halo, those with low $\alpha$ have highly isotropic local environment while those with high $\alpha$ reside in anisotropic filamentary environments. Paranjape (2020) provides theoretically motivated insights into the behavior of $\alpha$.

2.3.1 Measuring $\alpha$ in a fiducial $\delta_L = 0$ simulation

To compute $\alpha$, we start with the density field evaluated on a cubic lattice using Cloud-in-Cell interpolation, which is then Fourier transformed and Gaussian smoothed using a range of smoothing scales $R_s$ to get the Fourier space field $\delta(\mathbf{k}, R_s) = (\mathbf{k}) e^{-k^2 R_s^2 / 2}$. Using this, the tidal tensor $\psi_{ij}(\mathbf{x}, R_s)$ is obtained for various smoothing scales $R_s$ by inverting the normalised Poisson equation and taking derivatives,

$$\psi_{ij}(\mathbf{x}, R_s) = FT \left[ \frac{k_i k_j}{k^2} \delta(\mathbf{k}, R_s) \right]. \quad (8)$$

Then we compute the halo-centric tidal tensor by choosing, for each halo, the tidal tensor centered around the nearest lattice point $x_{halo}$ and then linearly interpolating between the two smoothing scales nearest to the scale $R_s_{\alpha, halo} = 4R_{200c}/\sqrt{5}$ (Paranjape et al. 2018). This scale has been chosen so as to have larger $b_1 \leftrightarrow \alpha$ correlation than $b_1 \leftrightarrow \delta$ correlation while minimising the $\alpha \leftrightarrow \delta$ correlation (see Figure 5 of Paranjape et al. 2018).

2.3.2 Measuring $\alpha$ when $\delta_L \neq 0$

In SU simulations where $\delta_L \neq 0$, we need to be wary of certain subtleties while following our prescription for computing $\alpha$. First is the issue of the units of length. The default unit of measuring length is comoving h$^{-1}$Mpc. We would like to perform all computations in these units in the fiducial cosmology with $h = h_{\text{fid}}$. However, at a cosmic time $t$ (redshift $z$ of the fiducial cosmology) our SU with $\delta_L \neq 0$ corresponds to a snapshot at redshift $\tilde{z}$ in an $N$-body simulation with a different Hubble constant $\tilde{h}$. To ensure that the proper length of the smoothing scale is preserved across SU simulations, the units of length in the SU snapshot are transformed as follows

$$x \rightarrow x \times \frac{h_{\text{fid}}(1 + \tilde{z})}{\tilde{h}(1 + \tilde{z})}. \quad (9)$$
Secondly, we need to modify our CIC algorithm for computing overdensity field $\delta(x)$. Recall that the overdensity can be written as,

$$\delta_{x_n} = \bar{\varrho} - \bar{\varrho} = \frac{\Delta n_{x_n}}{N_{\text{part}}} N_{\text{Grid}} - 1. \quad (10)$$

where $\Delta n_{x_n}$ is the number of dark matter particles contributing to the lattice point $x_n$, $N_{\text{part}}$ is the total number of dark matter particles and $N_{\text{Grid}}$ is the total number of lattice points. Recollect that in order to go from the first equality to the next in equation (10), we make the assumption that the density of the average simulation box $\varrho_{\text{sim}}$ is equal to the average density of the fiducial universe $\bar{\varrho}$. However this is only the case for simulations where $\delta_L = 0$. In the simulations with positive $\delta_L$, $\varrho_{\text{sim}} > \bar{\varrho}$ and when $\delta_L$ is negative, $\varrho_{\text{sim}} < \bar{\varrho}$. The CIC overdensity, after considering this can be computed as

$$\delta_{x_n} = \frac{\Delta n_{x_n} (1 + z)^3}{N_{\text{part}} (1 + z)^3} N_{\text{Grid}} - 1. \quad (11)$$

Lastly, since different SU boxes have different lengths in our default units we alter $N_{\text{Grid}}$ so as to keep the grid size equal. This will keep the CIC density field calculation consistent across different SU simulation boxes. We take $N_{\text{Grid}} = 844^3$ for $\delta_L = 0$ simulations and for other simulations we alter $N_{\text{Grid}}$ to be $844^3(1 + z)^3/(1 + \bar{z})^3$ rounded to the nearest integer.

The first two effects are relatively important and should be mandatorily tracked while the last effect discussed in the section is of lesser importance. This is because the first two modifications lie at the center of SU approach while the last effect plays a significant role only if $\bar{\alpha}$ has not converged.

2.3.3 Distribution of $\alpha$

The tidal anisotropy $\alpha$, for populations in narrow mass ranges, can be Gaussianized by a relatively easy transformation as it has a near-Lognormal distribution. For each mass bin we can standardize the tidal anisotropy $\alpha$ as follows,

$$\bar{\alpha} \equiv \frac{\ln \alpha - \mu_0}{\sigma_0}, \quad (12)$$

where

$$\mu_0 \equiv \langle \ln \alpha | m, \delta_L = 0 \rangle, \quad (13)$$

$$\sigma_0^2 \equiv \text{Var}(\ln \alpha | m, \delta_L = 0). \quad (14)$$

Thus $\bar{\alpha}$ by construction has a standard Gaussian distribution in the $\delta_L = 0$ universe. This can also be seen in Figure 1 where the grey histogram showing the distribution of $\bar{\alpha}$ in $\delta_L = 0$ universe is well approximated by the thick solid black standard Gaussian. However, from the blue and red step histograms of the same figure, we see that this is not the case for $\delta_L \neq 0$ universe. From experimenting with the simulation data for $\delta_L \neq 0$, we see that $\bar{\alpha}$ is still Gaussian but with a systematic shift in mean and variance as $\delta_L$ becomes progressively positive or negative. This encourages us to define the mean and variance for a mass range as a Taylor expansion in powers of $\delta_L$.

$$\mu(m, \delta_L) \equiv \sum_{n=1}^{\infty} \frac{\mu^L_n(m)}{n!} \delta_L^n, \quad (15)$$

$$\sigma^2(m, \delta_L) \equiv 1 + \sum_{n=1}^{\infty} \frac{S^L_n(m)}{n!} \delta_L^n. \quad (16)$$

Figure 2 shows equations (13) and (14) as a function of ‘peak height’ $\nu(m, z)$ for $\delta_L = 0$. In the right panel the data describing redshift 0 and 1 is combined and fit with a universal quadratic polynomial describing the variance of logarithmic tidal anisotropy using $\sigma_0^2 = S_0 + S_1 y + S_2 y^2$. Here $y = \log_{10}(\nu/2.05)$ is the logarithmic peak height. In the left panel a 4 parameter joint fit is performed on the mean value of tidal anisotropy to the polynomial $\mu_0(y, z) = m_{00} (1 - z) + m_{10} z + m_1 y + m_2 y^2$. Thus we have two polynomials corresponding to two data sets at redshift 0 and 1 respectively. The joint fit is produced by minimising the sum of the individual chisquare function. Table 1 provides the best fit values and covariance matrix for the above fits.

| $m_{00}$ | $m_{10}$ | $m_1$ | $m_2$ | $\chi^2$ |
|----------|----------|-------|-------|----------|
| std dev  | 0.001    | 0.012 | 0.058 |          |
| corr $S_0$ | 1.000 | -0.060 | -0.366 | 10.44    |
| corr $S_1$ | -1.000 | 1.000 | 0.874 |          |

Table 1. Best fit coefficients and covariance matrices of quadratic polynomial fits $\mu_0$ and $\sigma_0$ as a function of logarithmic peak height $y = \log_{10}(\nu/2.05)$. Fits were performed with the coefficients defining $\sigma_0^2 = S_0 + S_1 y + S_2 y^2$ and a 4 parameter joint fit for both redshift $z = 0$ and $z = 1$ as follows, $\mu_0(z = 0) = m_{00} + m_{10} y + m_2 y^2$ and $\mu_0(z = 1) = m_{10} + m_1 y + m_2 y^2$. Upper and lower blocks correspond to fits for $m_{00}$ and $S_0$ respectively. In each block the first row gives the best fit values, the second row gives the standard deviation and the last few rows give the correlation coefficients.

The discussions in this section will be useful in the subsequent sections where an analytical framework will be introduced and will rely on a model for the distribution of $\bar{\alpha}$.
Figure 1. Probability distribution of tidal anisotropy \( \tilde{\alpha} \) for different \( \delta_L \), at \( z=0 \) (left panel) and \( z=1 \) (right panel), averaged over 15 realizations. Warmer (cooler) colors are used to denote \( \delta_L < 0 \) (\( \delta_L > 0 \)) respectively and are detailed in the legend. The solid (dotted) linestyles represent data from mass ranges \( m = 6.2-10 \times 10^{12} \) (\( 2-5 \times 10^{13} \)) \( M_\odot \). Solid black curve in each panel shows the standard Gaussian distribution \( p(\tilde{\alpha}) = e^{-\tilde{\alpha}^2/2}/\sqrt{2\pi} \) that we use to approximate the grey \( \delta_L = 0 \) distribution in our analytical framework. The solid blue and red curves are other Gaussians with shifted mean and variance computed from direct measurements (see equations 15 and 16), also used to approximate distribution of \( \tilde{\alpha} \) in \( \delta_L = 0.7 \) and \( \delta_L = -0.7 \) respectively.

Figure 2. Mean and variance of tidal anisotropy \( \alpha \) in the fiducial cosmology as a function of peak height \( \nu \) for redshifts \( z=(0)1 \) as indicated by filled(empty) markers. The dashed line with square markers is computed in the default simulation box having particle count \( 512^3 \) while the dotted line with circular markers is computed in a higher resolution box having particle count \( 1024^3 \). The solid curves are obtained by fitting polynomials as a function of \( y = \log_{10}(\nu/2.05) \) to the mean (variance) in the left (right) panel, as described in section 2.3.3. Best fit values and errors are given in Table 1.
3 FRAMEWORK FOR HIGH-PRECISION BIAS CALIBRATION

3.1 Lognormal Model

This section is a straightforward utilization of the analytic framework developed in (Paranjape & Padmanabhan 2017) which we will refer to as the Lognormal model for halo assembly bias. Here we use the tidal anisotropy $\tilde{\alpha}$ from equation (12) as the assembly bias variable. We can include the dependence of the bias coefficients on $\tilde{\alpha}$ in equations (2) and (3) and write as

$$\delta_L^b (m, \tilde{\alpha}, \delta_L) \equiv \frac{n(m, \tilde{\alpha}, \delta_L)}{n(m, \tilde{\alpha}, \delta_L = 0)} - 1.$$

Combining equations (17), (3) and (2) we can write the dependence of bias coefficients on $\tilde{\alpha}$ in terms of its probability distribution $\mathcal{P}(\tilde{\alpha}|m, \delta_L)$

$$1 + \sum_{n=1}^{\infty} \frac{b_n^L(m, \tilde{\alpha})}{n!} \delta_L^n \equiv \left( 1 + \sum_{n=1}^{\infty} \frac{b_n^L(m)}{n!} \delta_L^n \right) \mathcal{P}(\tilde{\alpha}|m, \delta_L = 0).$$

In the above, we have used Bayes’ theorem to express the number density of haloes in terms of distribution of $\tilde{\alpha}$ like so: $n(m, \tilde{\alpha}, \delta_L) = n(m|\delta_L) \mathcal{P}(\tilde{\alpha}|m, \delta_L)$. As discussed in Section 2.3.3 the probability distribution of $\tilde{\alpha}$ for a fixed mass $m$ and $\delta_L$ is a Gaussian with mean $\mu$ and variance $\sigma$ that can be expressed in powers of $\delta_L$ as shown in equations (15) and (16). Hence it is possible to write out the above expression in powers of $\delta_L$ and equate the coefficients of each power to obtain equations for dependence of each bias coefficient on $\tilde{\alpha}$. In particular, the Lagrangian linear and quadratic bias can be expressed as

$$b_1^L(m, \tilde{\alpha}) = b_1^L(m) + \mu_1^L(m) H_1(\tilde{\alpha}) + \frac{1}{2} \Sigma_1^L(m) H_2(\tilde{\alpha}),$$

$$b_2^L(m, \tilde{\alpha}) = b_2^L(m) + \left\{ \mu_2^L(m) + 2 b_1^L(m) \mu_1^L(m) \right\} H_1(\tilde{\alpha}) + \left\{ \mu_1^L(m)^2 + 2 b_1^L(m) \Sigma_1^L(m) + \frac{1}{2} \Sigma_2^L(m) \right\} H_2(\tilde{\alpha}) + \mu_1^L(m) \Sigma_1^L(m) H_3(\tilde{\alpha}) + \frac{1}{4} \Sigma_2^L(m) H_4(\tilde{\alpha}).$$

For comparison, we also compute the peak-background split bias described in Section 2.2 for the halo populations with $\tilde{\alpha} > 0.675$ and $\tilde{\alpha} < -0.675$ separately. The results, shown as the two sets of points with error bars in Figure 4, agree well with the Lognormal model, but with larger errors. Thus, the Lognormal model is a very convenient noise reduction technique for computing halo assembly bias, as noted previously by Paranjape & Padmanabhan (2017). In Figure 5 we compare the Lognormal model to direct computation of linear halo bias using low-$k$ ($0.02 \leq k/(h \text{Mpc}^{-1}) < 0.1$) measurements of the ratio of halo-matter cross power spectrum to the matter auto-power spectrum. We see that the direct measurements broadly agree with the SU results showing same qualitative trends with an overall reduced strength. The quantitative differences between the two are likely due to the fact that the SU approach probes the infinite wavelength $k \to 0$ modes while any direct measurement will be limited the size of the simulation box considered. The halo bias is also computed in a smaller range of $k$.
formalism. The formalism allows for calculation of bias at
---

Table 2. Tidal anisotropy $\alpha$: Best fit coefficients and covariance matrices of quadratic polynomial fits to $\mu_n^L(y)$ and $\Sigma_n^L(y)$ as a function of logarithmic peak height $y \equiv \log_{10}[\nu(m, z)/1.5]$ for $n=1, 2$ (See Figure 3). The fits were performed in the range $1.1 \leq \nu \leq 2.7$ with the coefficients defining $\mu_n^L/y = \mu_n^L + \mu_{n1y} + \mu_{2ny}^2$ and $\Sigma_n^L/y^2 = \Sigma_n^L + \Sigma_{n1y} + \Sigma_{2ny} y^2$. The upper and lower blocks give these polynomial coefficients for $n=1, 2$ respectively. In each block, the first row gives the least squares best fit values, the second row gives the standard deviation (square root of the diagonal elements of the covariance matrix). The last two rows give the correlation coefficients (elements of the covariance matrix $C_{ij}$ divided by $\sqrt{C_{ii} C_{jj}}$).

---

Figure 3. Lagrangian assembly bias coefficients $\mu_n^L$ and $\Sigma_n^L$ (equations 15 and 16), extrapolated to the measurement redshift by dividing by $g(z)^n$ and shown as functions of $\nu(m_{200}, z)$, for $n = 1$ (left panel) and $n = 2$ (right panel). The points with error bars show measurements from simulations (details in section 2.2). The filled (empty) symbols show measurements at $z = 0$ ($z = 1$). The solid curves are obtained by fitting a quadratic polynomial as a function of $y = \log_{10}(\nu/1.5)$ using the points and errors in the range $1.1 < \nu < 2.8$. The best fit values and errors from this quadratic fit are given in Table 2.

---

modes $(0.02 \leq k/(h\text{Mpc}^{-1}) < 0.03)$ and shown in the same figure with thinner markers. While sample variance makes these measurements noisier, the agreement with SU result improves, an evidence for the susceptibility of direct halo bias measurements to the scale dependence of bias.

We would like to emphasize that the analysis in this section, though interesting for comparing with literature, does not demonstrate the full capability of the Lognormal formalism. The formalism allows for calculation of bias at fixed values of $\alpha$ and $\nu$, which is much more informative than binning in arbitrary percentiles. This has been shown in Figure 6 as the difference between $b_1(\nu, \tilde{\alpha}) - b_1(\nu)$ for a few fixed values of $\alpha$. It is also interesting to note here that the strength of assembly bias in Figure 6 is almost a constant with peak height for lower $[\tilde{\alpha}]$ values, which emphasizes the point made by Paranjape et al. (2018) that tidal anisotropy appears to be more relevant in determining linear halo bias than is halo mass.
3.4 Quadratic halo bias and tidal anisotropy

The quadratic assembly bias i.e, the dependence of quadratic bias $b_2$ on tidal anisotropy, is expected on general grounds but has not, to our knowledge, been demonstrated before. We do so in this Section; both the measurements and the analytical framework above are set up to effortlessly obtain the quadratic bias in addition to the linear bias.

Figure 7 shows the difference in $b_2$ for upper and lower quartiles of the tidal anisotropy $\alpha$. Interestingly, the upper and lower quartiles have opposite signs in all the mass ranges. The upper quartile population having positive values is expected from the extreme non-Gaussianities and non-linearities present in the spatial distribution of haloes in dense filamentary (high $\tilde{\alpha}$) environment. The near-zero, slightly negative $b_2$ of the haloes in isotropic region (low $\tilde{\alpha}$) is more complicated, as it could either have negative skew from being in an underdense void or a positive skew from being in an overdense cluster. There are many examples of tracers that have negative $b_2$ (Fieldman et al. 2001; Guo & Jing 2009; Hoffmann et al. 2019). We can see that the dependence on environment is clearly strong, the relative difference between any quartile $b_1$ and mean $b_1$ is of the order of unity while the relative difference between quartile and mean $b_2$ is of the order of 10.

Unlike $b_1$, the $\alpha$-dependence of $b_2$ is also a strong function of $\nu$, consistent with the expectation that $b_2$ depends

Figure 6. Assembly bias at fixed standardised tidal anisotropy $\tilde{\alpha}$. Each curve is obtained from the Lognormal model by taking the dependence of $b_1(\nu, \tilde{\alpha}) - b_1(\nu)$ (see equation 19) at $\tilde{\alpha} = \pm 2$ (red), $\tilde{\alpha} = \pm 1$ (purple) and $\tilde{\alpha} = 0$ (yellow). The error bands are computed with the same procedure as described in Figure 4.
4 EXTENSION TO OTHER SECONDARY PROPERTIES

Previously, Ramakrishnan et al. (2019) have considered direct halo-by-halo measurements of linear bias \( b_1 \) in standard N-body simulations (Paranjape et al. 2018) and internal property \( c \) as random variables, allowing the correlation between them at fixed halo mass to be defined as assembly bias. Ramakrishnan et al. (2019) showed that the halo bias and internal property are consistent with being conditionally independent given the tidal anisotropy:

\[
p(b_1 | \alpha, c, m) \approx p(b_1 | \tilde{\alpha}, m).
\]  

(21)

Thus, the assembly bias trends \( c \leftrightarrow b_1 \) reflect the two fundamental correlations \( c \leftrightarrow \alpha \) and \( b_1 \leftrightarrow \tilde{\alpha} \). This also implies that, given our formalism for modeling \( b_1(\tilde{\alpha}, m) \), we should also be able to predict \( b_1(\alpha, c, m) \), provided we know the correlation coefficient \( \rho \) between \( \tilde{\alpha} \) and \( c \). We pursue this idea in this section by developing a bivariate model of halo assembly bias.

4.1 Bivariate Lognormal Model

Considering \( b_1 \) as a stochastic property for every halo, we can think of the mean bias at fixed halo mass as the expectation value

\[
\langle b_1|m \rangle = \int db_1 \, p(b_1|m) \, b_1 \equiv b_1(m).
\]

(22)

Similarly, conditional averages of \( b_1 \) can be expressed in terms of appropriate probability distributions as follows.

\[
\langle b_1|c, m \rangle = \int db_1 \, p(b_1|c, m) \, b_1
\]

\[
= \int d\tilde{\alpha} \, p(b_1|\tilde{\alpha}, c, m) \, p(\tilde{\alpha}|c, m) \, b_1
\]

\[
= \int d\tilde{\alpha} \, \langle b_1|\tilde{\alpha}, m \rangle \, p(\tilde{\alpha}|c, m),
\]

(23)

where we marginalized over \( \tilde{\alpha} \) in the second line and assumed the conditional independence of \( b_1 \) on \( c \) at fixed \( \tilde{\alpha} \) in the last line (see equation (21)). This simplifies the expression since we can now replace \( \langle b_1|\tilde{\alpha}, m \rangle \) as

\[
\langle b_1|\tilde{\alpha}, m \rangle = 1 + b_1^c(m, \tilde{\alpha})g(z)^{-1},
\]

(24)

where \( b_1^c(m, \tilde{\alpha}) \) was given in equation (19). We can see that the \( \tilde{\alpha} \) dependence occurs only in the Hermite polynomials, so we need to evaluate the following set of integrals

\[
\langle H_n(\tilde{\alpha})|c, m \rangle = \int H_n(\tilde{\alpha}) p(\tilde{\alpha}|c, m) d\tilde{\alpha},
\]

(25)

So far we have not discussed the distribution of the internal property \( c \). In the case where this is another standard normal Gaussian, then the above integral has an analytic solution,

\[
\langle H_n(\tilde{\alpha})|c, m \rangle = \int H_n(\tilde{\alpha}) p(\tilde{\alpha}|c, m) d\tilde{\alpha} = \rho(m)^n H_n(c),
\]

(26)

where \( \rho(m) \) is the correlation coefficient between \( c \) and \( \tilde{\alpha} \) in the mass bin \( m \) (see Appendix C for details). Putting this back in equation (23) and (24) gives us

\[
b_1(m, z, c) = \langle b_1|c, m, z \rangle,
\]

\[
= b_1(m, z) + \mu_1^c(m, z) p(m, z) H_1(c)
\]

\[
+ \frac{1}{2} \sigma_1^2(m, z) p(m, z) H_2(c).
\]

(27)

Note that by setting \( \rho = 1 \) in the above equation, we can recover equation (19) as it should be in the case of \( c=\tilde{\alpha} \).

4.2 An Example: Halo Concentration

Halo concentration has been extensively used to describe halo assembly bias in the literature although there are several other halo properties in which assembly bias manifests. Despite the large number of studies to describe its assembly bias, there are remarkably few attempts to accurately calibrate the effect (Wechsler et al. 2006; Paranjape & Padmanabhan 2017). Here, we provide an alternate calibration for the dependence of bias on halo concentration within the extended framework described in the previous sections. Halo concentration has an approximate Lognormal distribution which makes it convenient for using its Gaussianized form as the variable \( c \) in the bivariate Lognormal model introduced above.

Denoting halo concentration by \( c_{200b} = R_{200b}/r_s \), where
\[ \chi^2 = e^{\sigma_\sigma^2} - 1 \]

where \( \sigma_\sigma \) and \( \sigma_c \) are the standard deviation of Gaussianized tidal anisotropy \( \tilde{\alpha} \) and concentration \( c \).

Details for obtaining equation (31) are given in Appendix B. However when calculating Pearson’s correlation coefficient for actual data one needs to be wary that it is highly sensitive to outliers. Spearman Correlation is a good alternative which is robust against outliers but its magnitude can differ from the Pearson’s correlation coefficient as required in equation (27).

We have identified three methods which we can use to compute correlation coefficient \( \rho \).

(i) First method: Compute the Pearson’s correlation coefficient \( \rho_{LN} \) between the Lognormal variables from the simulation and analytically obtain \( \rho \) using equation (31).

(ii) Second method: Gaussianize tidal anisotropy and halo concentration and then obtain their correlation coefficient \( \rho \).

(iii) Third method: Compute the Spearman correlation coefficient between the two variables.

Though all the methods should give similar results, they give slightly different values due to non-Gaussianities/outliers in the distribution of \( c \) and \( \tilde{\alpha} \). The distribution of Gaussianized \( c_{200b} \), particularly has a negative skew as well as negative outliers as can be seen in Figure 9. Thus the already weak correlations become increasingly difficult to calculate accurately. We need to identify the method robust to these issues.

After the detailed analysis done in Appendix A, we
choose to work with the first method because we see that the Pearson’s correlation coefficient of Lognormal variables is more robust to negative outliers and downweights them in the calculation of the correlation coefficient. Figure 8 shows \( \rho \) as obtained from the first method as a function of peak height. We choose to fit a third order polynomial to this after analysis with Akaike information criteria with correction (Akaike 1974; Sugnara 1978) for different order polynomials. The best coefficients and covariance matrix are shown in Table 3. We do all the subsequent analysis with this functional form. We have repeated the entire analysis using the other methods and find qualitatively similar results although quantitative details differ.

4.2.2 Comparison with simulations

We separately perform SU calculations as described in Section 2.2 for obtaining peak-background split bias of halo populations for upper and lower quartiles of \( c_{200b} \). In Figure 10, the two sets of points with error bars show the bias for the upper and lower quartiles of \( c_{200b} \). We compare this with the bivariate Lognormal model plotted as solid curves by averaging equation (27) above \( c > 0.675 \) and below \( c < -0.675 \) for the upper and lower quartiles of halo concentration respectively. Error bars are obtained in the same manner as before in the case of assembly bias in \( \alpha \), the covariance matrix from Table 2 is used to construct a trivariate Gaussian distribution and the coefficients \( \mu_L \) and \( \Sigma_L \) are sampled 300 times to obtain convergent error values. \( b^L(m, c) \) is computed each of these times. The standard deviation of the above sample of \( b^L(m, c) \) is plotted as a band around the Lognormal model.

4.3 Can the model predict quadratic assembly bias?

So far in Section 4.2, we have used the conditional independence of linear bias \( b_1 \) and a halo property in fixed tidal environments to predict the linear assembly bias with the property \( c \). This was shown in Ramakrishnan et al. (2019) by treating linear halo bias as a halo-centric property and computing correlation coefficients with other halo-centric quantities. In principle, one could verify the same for quadratic bias by measuring the bispectrum and calculating an analogous ‘halo-by-halo quadratic bias’. Instead of this, here we make an assumption about the conditional independence of \( b_2 \) and an internal property of the halo as follows,

\[
\langle b_2|\alpha, c, m \rangle = \langle b_2|\alpha, m \rangle ,
\]

using which we model the quadratic assembly bias with halo property \( c \). The resulting dependence of \( b_2 \) on halo mass and halo property \( c \) can be written, analogously to equation (27), as

\[
b_2(m, c, z) = b_2 + \{\mu_2^L + 2\mu_2^L(b_1 - 1) + \frac{8}{21}\mu_1^L\} \rho H_1(c) + \{\mu_2^L\}^2 + \Sigma_2^L \rho^2 H_2(c) + \frac{4}{21}\Sigma_1^L \rho^2 H_3(c) + \frac{4}{21}\Sigma_1^L \rho^2 H_4(c) ,
\]

For brevity, we have suppressed the mass and redshift dependence on all terms on the right side of equation above except the Hermite polynomials, which only have \( c \) dependence. We test the accuracy of the above equation in Figure 11. Although the model qualitatively describes the simulation...
points, the overall agreement is poor at low masses. This could just be due to the systematic error in the measurement of second-order terms. It could also be that the assumption of conditional independence in equation (32) breaks down for higher-order non-linear bias coefficients at these mass scales. This is not unexpected since the low mass haloes are a mix of two kinds of populations having contrasting environments, thus making their trends complicated. One sub-population of haloes in isotropic environment behave like ‘standard’ peaks theory/excursion set haloes and their halo concentration is negatively correlated with the tidal environment while the other sub-population lives in a highly anisotropic environment, initially set to become high mass haloes but get tidally truncated by redirected mass flow to filaments and their halo concentration is positively correlated with the environment (Hahn et al. 2009; Paranjape et al. 2018). A fuller exploration of these effects would be possible using direct measurements of the halo bispectrum in different tidal environments, an exercise we leave to future work.

5 SUMMARY

Halo assembly bias is a potential source of systematic uncertainty for cosmological inference from upcoming large-volume galaxy surveys, as well as being a possible channel for enhancing our understanding of galaxy formation and evolution. Our aim in this work has been to develop accurate calibrations of the dependence of halo bias on one of the primary ‘beyond halo mass’ variables responsible for assembly bias, namely, the tidal anisotropy $\alpha$ of the local cosmic web environment of haloes. We used the Separate Universe (SU) technique to calibrate the dependence of linear and quadratic bias $b_1$ and $b_2$, respectively, on halo mass, redshift and $\alpha$. We also showed, using the example of halo concentration, that it is possible to make use of this calibration on web environment to further calibrate the dependence of bias on other secondary properties. Our results can be summarized as follows:

- The tidal anisotropy $\alpha$ has a nearly Lognormal distribution over the entire range of peak height that we studied $1.1 \leq \nu \leq 3.4$, summarized in Table 1.
- We first used the SU approach to numerically calculate $b_1(\alpha, \nu)$ (Figure 4) and $b_2(\alpha, \nu)$ (Figure 7) in quartiles of $\alpha$ and bins of peak height $\nu(m, z)$. This is the first reported detection of quadratic assembly bias with respect to the tidal environment of the halo.
- We also analytically calibrated, with very high precision, the relations $b_1(\alpha, \nu)$ and $b_2(\alpha, \nu)$ as continuous functions of $\alpha$ (i.e., without binning) using the framework developed in Paranjape & Padmanabhan (2017) (see Section 3.1) which exploits the near-Lognormal distribution of $\alpha$, combined with fitting functions $b_1(\nu)$ and $b_2(\nu)$ from the literature for the all-halo results. These results are summarized in equations (19)-(20), Table 2 and Figures 4, 6 and 7, with a comparison to the $\alpha$-dependence of linear bias directly measured in simulations shown in Figure 5.
- Using the conditional independence of large-scale bias on secondary halo properties at fixed $\alpha$ (Ramakrishnan et al. 2019), we then extended this analytic framework to accommodate the dependence of bias on another secondary property, whose distribution has or can be monotonically transformed to have Gaussian distribution (Section 4.1). We demonstrated this technique for the case of halo concentration $c_{200b}$ by calibrating the conditional distribution $p(c_{200b}|\alpha, \nu)$ (Figure 8 and Table 3). We reproduce the known dependence of $b_1(c_{200b}, \nu)$ accurately over our entire dynamic range (Figure 10), while $b_2(c_{200b}, \nu)$ departs from previous results at low $\nu$. We discussed possible reasons for the latter discrepancy in Section 4.3.

Our calibrations of $b_1$ and $b_2$ can potentially be useful in a number of areas:

(i) Self-calibrating cluster surveys which constrain cosmological parameters as well as mass-observable relations (Majumdar & Mohr 2004; Wu et al. 2008; Chiu et al. 2020; Nicola et al. 2020).

(ii) Redshift space distortion (RSD) modeling to constrain cosmic acceleration physics: This can be done by incorporating correlations between large scale bias and velocity dispersion into RSD modeling which can potentially constrain cosmological parameters sensitive to nature of gravity.

(iii) The calibration of $b_2$, $b_1$ on tidal anisotropy and mass provides a possibility to improve the models which use three-point statistics like bispectrum to constrain primordial non-Gaussianities (Jeong & Komatsu 2009; Karagiannis et al. 2018; Gualdi & Verde 2020).

(iv) Analytical forecasts for multitracer analyses that require samples with widely different bias parameters (McDonald & Seljak 2009; Fonseca et al. 2015).

(v) HOD and galaxy modeling to incorporate assembly bias in mock catalogs, potentially for several secondary properties in addition to $\alpha$ and halo concentration discussed here.

We will return to these topics in future work.
ACKNOWLEDGMENTS

We thank Ravi Sheth and Oliver Hahn for useful discussions. We thank the Munich Institute for Astro- and Particle Physics (MIAPP) and the organisers of the programme on Dynamics of Large-Scale Structure (July 2019) for their hospitality while part of this work was completed. The research of AP is supported by the Associateship Scheme of ICTP, Trieste and the Ramanujan Fellowship awarded by the Department of Science and Technology, Government of India. We gratefully acknowledge the use of high performance computing facilities at IUCAA, Pune.

DATA AVAILABILITY

No new data were generated in support of this research. The simulations used in this work are available from the authors upon request.

REFERENCES

Akaike H., 1974, IEEE Transactions on Automatic Control, 19, 716
Baldauf T., Seljak U., Senatore L., Zaldarriaga M., 2011, J. Cosmology Astropart. Phys., 10, 031
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Bernardeau F., 1992, ApJ, 392, 1
Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215
Borzyszkowski M., Porciani C., Romano-Díaz E., Garaldi E., 2013, MNRAS, 432, 436
Bray A. D., et al., 2016, MNRAS, 455, 184
Castorina E., Sheth R. K., 2013, MNRAS, 433, 1529
Catelan P., Theuns T., 1992, MNRAS, 256, 594
Chaves-Montero J., Angulo R., Zennaro M., 2020, arXiv e-prints, p. arXiv:2005.03672
Cole S., 1997, MNRAS, 286, 38
Contreras S., Angulo R., Zennaro M., 2020, arXiv e-prints, p. arXiv:2005.03672
Dalal N., White M., Bond J. R., Shirokov A., 2008, ApJ, 687, 12
Desjacques V., 2008, MNRAS, 388, 638
Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1
Faltenbacher A., White S. D. M., 2010, ApJ, 708, 469
Feldman H. A., Frieman J. A., Fry J. N., Scoccimarro R., 2001, Phys. Rev. Lett., 86, 1434
Fonseca J., Camera S., Santos M. G., Maartens R., 2015, ApJ, 812, L22
Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
Gualdi D., Verde L., 2020, arXiv e-prints, p. arXiv:2003.12075
Guo H., Jing Y. P., 2009, ApJ, 702, 425
Hahn O., Porciani C., Dekel A., Carollo C. M., 2009, MNRAS, 398, 1742
Han J., Li Y., Jing Y., Nishimichi T., Wang W., Jiang C., 2019, MNRAS, 482, 1900
Hearin A. P., Watson D. F., 2013, MNRAS, 435, 1313
Heavens A., Peacock J., 1988, MNRAS, 232, 339
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Hoffmann K., Mao Y., Xu J., Mo H., Wand elt B. D., 2019, MNRAS, 487, 3050
Jeong D., Komatsu E., 2009, ApJ, 703, 1230
Kaiser N., 1984, ApJ, 284, L9
Karagiannis D., Lazanu A., Liguori M., Raccanelli A., Bartolo N., Verde L., 2018, MNRAS, 478, 1341
Lazeyras T., Wagner C., Baldaf u T., Schmidt F., 2016, J. Cosmology Astropart. Phys., 2, 018
Lazeyras T., Musso M., Schmidt F., 2017, J. Cosmology Astropart. Phys., 3, 059
Li Y., Hu W., Takada M., 2016, Phys. Rev. D, 93, 063507
Lin Y.-T., Mandelbaum R., Huang Y.-H., Huang H.-J., Dalal N., Diemer B., Jian H.-Y., Kravtsov A., 2016, ApJ, 819, 119
Majumdar S., Mohr J. J., 2004, ApJ, 613, 41
McDonald P., Seljak U., 2009, J. Cosmology Astropart. Phys., 10, 007
Miyatake H., More S., Takada M., Spergel D. N., Mandelbaum R., Rykoff E. S., Rozo E., 2016, Physical Review Letters, 116, 041301
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Monte ro-Dorta A. D., et al., 2017, ApJ, 848, L2
Monte ro-Dorta A. D., et al., 2020, arXiv e-prints, p. arXiv:2001.01739
Musso M., Sheth R. K., 2012, MNRAS, 423, L102
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neto A. F., et al., 2007, MNRAS, 381, 1450
Nicol a A., Dunkley J., Spergel D. N., 2020, arXiv e-prints, p. arXiv:2006.00008
Paranjape A., 2020, arXiv e-prints, p. arXiv:2006.13954
Paranjape A., Padmanabhan N., 2017, MNRAS, 468, 2984
Paranjape A., Hahn O., Sheth R. K., 2018, MNRAS, 476, 3631
Ramakrishnan S., Paranjape A., Hahn O., Sheth R. K., 2019, MNRAS, 489, 2077
Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30
Shen J., Abel T., Mo H. J., Sheth R. K., 2006, ApJ, 645, 783
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Tormen G., 2004, MNRAS, 350, 1385
Sprin gel V., 2005, MNRAS, 364, 1105
 Sugni ra N., 1978, Communications in Statistics - Theory and Methods, 7, 13
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Tinker J. L., Wetzel A. R., Conroy C., Mao Y.-Y., 2015, MNRAS, 472, 5204
Tojeiro R., et al., 2017, MNRAS, 470, 3720
Torm en G., Bertschinger E., 1996, ApJ, 472, 14
Wagner C., Schmidt F., Chiang C.-T., Komatsu E., 2015a, J. Cosmology Astropart. Phys., 8, 042
Wagner C., Schmidt F., Chiang C.-T., Komatsu E., 2015b, MNRAS, 448, L11
Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Wu H.-Y., Rozo E., Wechsler R. H., 2008, ApJ, 688, 729
Xu X., Zheng Z., 2018, MNRAS, 479, 1579
Zehavi I., Kerby S. E., Contreras S., Jiménez E., Padilla N., Baugh C., 2019, ApJ, 885, 1
Zentner A. R., 2007, International Journal of Modern Physics D, 16, 763
Zentner A. R., Hearin A., van den Bosch F. C., Lange J. U., Villarreal A., 2016, arXiv e-prints, p. arXiv:1606.07817
Zu Y., Mandelbaum R., Simet M., Rozo E., Rykoff E. S., 2017, MNRAS, 470, 551

Separate universe bias & tidal anisotropy 13
APPENDIX A: SENSITIVITY OF CORRELATION COEFFICIENTS TO OUTLIERS

There are three ways to measure the correlation coefficient between two variables $\alpha$ and $c$ as described in the main text in Section 4.2.1. The first method computes Pearson’s correlation coefficient between the halo tidal anisotropy and concentration. These are Lognormal variables and we use equation (31) to obtain the correlation coefficient between their Gaussianized forms. The second method converts them to Gaussianized form first and computes Pearson’s correlation coefficient. The third method simply computes the Spearman rank correlation coefficient between the two variables.

In this section, we want to select the best method of estimating the correlation coefficient out of the three listed above. We do this by considering a toy exercise using a mock sample of 600,000 “haloes”. Each halo is assigned two properties ($\alpha$ and $c$) which are distributed as a bivariate Gaussian. Each of these properties has zero mean and unit variance with the correlation coefficient 0.02. This correlation coefficient is chosen as it is one among the weakest correlations seen in our simulations between $\alpha$ and $c$, hence most sensitive to the presence of negative outliers as this is the most sensitive to the presence of negative outliers as this is especially the case in halo concentration $c$ though not so much in $\alpha$ distribution (see Figure 9). So we construct toy examples where the outlier haloes have negative skewness in $c$, while $\alpha$ is chosen from a standard Gaussian so as to preserve its marginal distribution. The second column shows the percentage of the population that comprises of the outliers. The third column shows the true correlation (excluding the outliers) of the population while the last three columns show how the correlation coefficient deviates from the true correlation for the three methods of computing correlation.

For example, in the first row of the table, ($\alpha \in N(0,1)$, $c \sim N(0,1)$) means that the outliers have $c = -30$ and $\alpha$ is drawn from standard Normal distribution $N(0,1)$ and they comprise 0.16% of the total population. Out of the three methods, the first method is the most robust and closest to the true correlation while the second method is most sensitive to outliers. In fact, in all other examples, the first method is the most robust to the presence of outliers.

The last example, which is an extreme case of large negative outliers in both $\alpha$ and $c$, is used to demonstrate the reason why the first method works better than the rest in the presence of a small population of highly negative outliers. We can see that a true negative correlation of $-0.5$ can turn to an even higher positive correlation of 0.59 when calculated using Pearson’s second method. To understand why the first method works better, let us reconstruct the two Lognormal variables $\alpha = \exp(\tilde{\alpha} + \sigma_\alpha \mu_0)$ and $c = \exp(\tilde{c} + \sigma_c \mu_0)$ where $\mu_0, \sigma_\alpha, \mu_c, \sigma_c$ are as defined in equations (13), (20), (29) and (30). While computing Pearson’s correlation coefficient in a simulation with $N$ haloes, the presence of $\mu_0$ and $\mu_0'$ would cancel to give

$$\rho_{\alpha c} = \frac{N \sum \exp(\tilde{\alpha} + \sigma_\alpha \epsilon_0) \exp(\tilde{c} + \sigma_c \epsilon_0) - \sum \exp(\tilde{\alpha} + \sigma_\alpha \epsilon_0) \sum \exp(\tilde{c} + \sigma_c \epsilon_0)}{\sqrt{N \sum \exp(\tilde{\alpha} + \sigma_\alpha \epsilon_0)^2 (N \sum \exp(\tilde{c} + \sigma_c \epsilon_0)^2 - (\sum \exp(\tilde{\alpha} + \sigma_\alpha \epsilon_0))^2)}^{1/2}}$$

where the summation is over all the haloes. Written in this manner, it becomes easy to see how negative outliers will be exponentiated and thus contribute negligibly to the above summations, thus leaving the correlation coefficient robust to these highly negative outliers. However, this method need not be restricted to be used for suppressing outliers of a negatively skewed Gaussian distribution; the contribution from a positive skew of a near-Gaussian variable $x$ can also be suppressed by this method with additional steps: transform the variable $x \rightarrow -x$ before applying the method and transform the correlation coefficient $\rho \rightarrow -\rho$ after applying the method.

We do not forget that in the attempt to conform to the Gaussian distribution that the model mandates, we have ignored a population of haloes having unusually low concentration, a population that could be physically interesting. One could in principle use Gaussian mixtures to factor in the tail as has been done in Neto et al. (2007), where the distribution is a sum of a larger Gaussian and a smaller one with smaller mean and larger variance. We leave such explorations to future work.

APPENDIX B: CORRELATION COEFFICIENTS (LOG)NORMAL VARIABLES

Let $Z_i$ be a random variable with Lognormal distribution,

$$Z_i = \exp(X_i), \quad X_i \sim N(\mu_1, \sigma_1^2).$$

The mean of $Z_i$ can be written as

$$E(Z_i) = \exp(\mu_1) \exp(\sigma_1^2/2).$$

This can be deduced from the one variable equivalent expression of the Moment generating function $M_X(t)$ for multivariate correlated variables,

$$M_X(t^T X) = \exp(t^T \mu + \frac{1}{2} t^T \Sigma t) \quad (B1)$$

where $X, t, \mu$ are $n$ dimensional vectors and $\Sigma$ is the covariance matrix. Then the expectation value of $Z_i^2$ can also be written as

$$E(Z_i^2) = \langle \exp(2X_i^2) \rangle = \exp(2\mu_1) \exp(2\sigma_1^2)$$

Hence we find the variance of $Z_i$ to be

$$\text{Var}(Z_i) = E(Z_i^2) - E(Z_i)^2 = \exp(2\mu_1 + 2\sigma_1^2) - \exp(2\mu_1 + \sigma_1^2)$$

Now consider two Lognormal variables $Z_i$ and $Z_j$ with correlation coefficient $\rho$. The expectation value of their product is

$$E(Z_i Z_j) = \langle \exp(X_i + X_j) \rangle = M_X(X_i + X_j)(t = (1,1))$$

$$= \exp(\mu_1 + \mu_2) \exp(\sigma_1^2 + \sigma_2^2 + 2\rho \sigma_1 \sigma_2)$$
Now we can find the correlation coefficient $\rho_{LN}$ between two Lognormal variables $Z_i$ and $Z_j$ to be

$$
\rho_{LN} = \rho_{Z_i Z_j} = \frac{E(Z_i Z_j) - E(Z_i)E(Z_j)}{\sqrt{\text{Var}(Z_i)\text{Var}(Z_j)}}
$$

$$
= \frac{e^{\mu_1 + \mu_2} e^{\sigma_1^2 + \sigma_2^2 + 2\rho_{s1}\sigma_1\sigma_2} - e^{\mu_1 + \mu_2} e^{\sigma_1^2 + \sigma_2^2}}{\sqrt{e^{2\mu_i + 2\mu_j} e^{\sigma_1^2 + \sigma_2^2} (e^{\sigma_1^2} - 1)(e^{\sigma_2^2} - 1)}}
$$

$$
\therefore \rho_{LN} = \frac{e^{\rho_{s1} \sigma_1 \sigma_2} - 1}{\sqrt{(e^{\sigma_1^2} - 1)(e^{\sigma_2^2} - 1)}}. \quad (B2)
$$

and we can write

$$
\langle H_n(\tilde{\alpha})|c\rangle = \int d\tilde{\alpha} p(\tilde{\alpha}|c) H_n(\tilde{\alpha})
$$

$$
= \frac{1}{p(c)} \int d\tilde{\alpha} p(\tilde{\alpha}) p(c|\tilde{\alpha}) H_n(\tilde{\alpha})
$$

$$
= \frac{1}{p(c)} \int d\tilde{\alpha} p(c|\tilde{\alpha}) \left( \frac{\partial}{\partial \tilde{\alpha}} \right)^n p(\tilde{\alpha})
$$

$$
= \frac{1}{p(c)} \int d\tilde{\alpha} p(\tilde{\alpha}) \left( \frac{\partial}{\partial \tilde{\alpha}} \right)^n p(c|\tilde{\alpha})
$$

$$
= \frac{\rho^n}{p(c)} \int d\tilde{\alpha} p(\tilde{\alpha}) \int \frac{dk}{2\pi} e^{ik(c-\rho\tilde{\alpha})} (-ik)^n e^{-k^2/(1-\rho^2)/2}
$$

$$
= \frac{\rho^n}{p(c)} \int \frac{dk}{2\pi} e^{ikc} (-ik)^n e^{-k^2/(1-\rho^2)/2}
$$

$$
= \frac{\rho^n}{p(c)} \int \frac{dk}{2\pi} e^{-k^2/2} (-ik)^n e^{-k^2/(1-\rho^2)/2}
$$

$$
= \frac{\rho^n}{p(c)} \int \frac{dk}{2\pi} e^{-k^2/2} H_n(c) p(c)
$$

$$
\therefore \langle H_n(\tilde{\alpha})|c\rangle = \rho^n H_n(c). \quad (C3)
$$

## APPENDIX C: HERMITE POLYNOMIAL INTEGRAL

The probabilist’s Hermite polynomials are defined by

$$
p(s)H_n(s) = (-d/ds)^n p(s) = \int \frac{dk}{2\pi} e^{iks} (-ik)^n e^{-k^2/2}, \quad (C1)
$$

where $p(s) = e^{-s^2/2}/\sqrt{2\pi}$ is the probability density function of a standard normal deviate. All integrals range from $-\infty$ to $\infty$ over the respective variable.

If both $\tilde{\alpha}$ and $c$ are standard normal deviates with correlation coefficient $\rho$, then we have

$$
p(c|\tilde{\alpha}) = \int \frac{dk}{2\pi} e^{ik(c-\rho\tilde{\alpha})} e^{-k^2(1-\rho^2)/2}, \quad (C2)
$$

\[\text{Table A1. Robustness of correlation coefficients to various kinds of outliers. A sample of 600,000 is made by first sampling bivariate normal distribution with mean and variance of both variables 0 and 1 respectively. The true correlation of the sample taken here is -0.16. Outliers are added to this sample as per the table and the correlation coefficients recalculated to check their sensitivity. See Appendix A for description of the method.}\]