Cosmic web anisotropy is the primary indicator of halo assembly bias

Sujatha Ramakrishnan, Aseem Paranjape, Oliver Hahn, Ravi K. Sheth

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **489,** 2977–2996 (2019)
Advance Access publication 2019 August 26

doi:10.1093/mnras/stz2344

Cosmic web anisotropy is the primary indicator of halo assembly bias

Sujatha Ramakrishnan [1]★ Aseem Paranjape [1] Oliver Hahn [2] and Ravi K. Sheth [3,4]

[1] *Inter-University Centre for Astronomy & Astrophysics, Ganeshkhind, Post Bag 4, Pune 411007, India*
[2] *Laboratoire Lagrange, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Blvd de l'Observatoire, CS 34229, F-06304 Nice cedex 4, France*
[3] *Center for Particle Cosmology, University of Pennsylvania, 209 S. 33rd St, Philadelphia, PA 19104, USA*
[4] *The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, I-34151 Trieste, Italy*

Accepted 2019 August 12. Received 2019 August 7; in original form 2019 March 6

ABSTRACT

The internal properties of dark matter haloes correlate with the large-scale halo clustering strength at fixed halo mass – an effect known as assembly bias – and are also strongly affected by the local, non-linear cosmic web. Characterizing a halo's local web environment by its tidal anisotropy α at scales approximately four times the halo radius, we demonstrate that these multiscale correlations represent two distinct statistical links: one between the internal property and α, and the other between α and large-scale ($\gtrsim 30\,h^{-1}$ Mpc) halo bias b_1. We focus on scalar internal properties of haloes related to formation time (concentration $c_{\rm vir}$), shape (mass ellipsoid asphericity c/a), velocity dispersion structure (velocity ellipsoid asphericity c_v/a_v and velocity anisotropy β), and angular momentum (dimensionless spin λ) in the mass range $8 \times 10^{11} \lesssim M_{\rm vir}/(h^{-1}\,{\rm M}_\odot) \lesssim 5 \times 10^{14}$. Using conditional correlation coefficients and other detailed tests, we show that the joint distribution of α, b_1, and *any* of the internal properties $c \in \{\beta, c_v/a_v, c/a, c_{\rm vir}, \lambda\}$ is consistent with $p(\alpha, b_1, c) \simeq p(\alpha)p(b_1|\alpha)p(c|\alpha)$, at all but the largest masses. Thus, the assembly bias trends $c \leftrightarrow b_1$ reflect the two fundamental correlations $c \leftrightarrow \alpha$ and $b_1 \leftrightarrow \alpha$. Our results are unaffected by the exclusion of haloes with recent major merger events or splashback objects, although the latter are distinguished by the fact that α *does not* explain their assembly bias trends. The overarching importance of α provides a new perspective on the nature of assembly bias of distinct haloes, with potential ramifications for incorporating realistic assembly bias effects into mock catalogues of future large-scale structure surveys and for detecting galaxy assembly bias.

Key words: methods: numerical – cosmology: theory – dark matter – large-scale structure of the Universe.

1 INTRODUCTION

The physical connection between the growth and properties of gravitationally collapsed dark matter haloes and the cosmic web environment in which these haloes reside is an interesting and challenging problem in the study of hierarchical structure formation (White & Silk 1979; Eisenstein & Loeb 1995; Bond & Myers 1996; Bond, Kofman & Pogosyan 1996; Monaco 1999; Sheth & Tormen 1999). Although the basic statistical connection between the very large-scale density environment (or halo bias) and halo properties such as mass was already established several decades ago (Kaiser 1984; Bardeen et al. 1986; Bond et al. 1991; Lacey & Cole 1993), subsequent technological improvements in simulating cold, collisionless self-gravitating cosmological systems have revealed several additional features of dark matter haloes.

Primarily, these relate to the striking universality seen in the structure of cold dark matter (CDM) haloes, both in the density (Navarro, Frenk & White 1996, 1997) as well as velocity dispersion profiles (Taylor & Navarro 2001; Ludlow et al. 2010). Later results also indicate a deep connection – which is the focus of this work – between the large-scale halo bias and internal properties of haloes of fixed mass such as formation time, concentration, substructure abundance, shape, velocity dispersion structure, angular momentum, etc. (see e.g. Sheth & Tormen 2004; Gao, Springel & White 2005; Wechsler et al. 2006; Jing, Suto & Mo 2007; Faltenbacher & White 2010). Apart from the intrinsic interest in painting a more complete picture of hierarchical structure formation from first principles, understanding and calibrating these effects also continues to be of interest from the point of view of galaxy formation and evolution (see e.g. Yan, Fan & White 2013; Lin et al. 2016; Tinker et al. 2017; Paranjape, Hahn & Sheth 2018b; Wang et al. 2018; Zehavi et al. 2018; Alam et al. 2019) as well as for precision cosmology (Zentner, Hearin & van den Bosch 2014; McEwen & Weinberg 2018).

★ E-mail: rsujatha@iucaa.in

The dependence of halo bias on halo formation time at fixed mass was termed 'assembly bias' in the early literature on this subject. We will use this term to denote the dependence of bias on any internal property other than mass, although recent results indicate that there could be more than one physical mechanism responsible for establishing these correlations (see e.g. Mao, Zentner & Wechsler 2018; Salcedo et al. 2018; Han et al. 2019).

In general, such correlations between internal halo properties (i.e. quantities defined at length-scales $\lesssim$few $\times\, 100\,h^{-1}$ kpc, say) and large-scale halo bias (measured at scales $\gtrsim$few $\times\, 10\,h^{-1}$ Mpc) can be thought of as remnants of the physics of halo formation in the hierarchical paradigm. For example, excursion set models of halo abundances and clustering generically predict such statistical correlations by connecting the local physics of halo formation to the large-scale halo environment through the long-wavelength correlations present in the initial conditions (see e.g. Zentner 2007; Dalal et al. 2008; Musso & Sheth 2012). These models, however, currently do not correctly reproduce all known assembly bias trends, indicating that they still lack some key physical mechanisms involved in halo formation.

Focusing on the expected local and highly non-linear nature of halo formation, it is then interesting to ask whether one might segregate the correlation between an internal property and large-scale bias into (at least) two distinct contributions: one composed of a connection between the internal property and some feature of the *local* non-linear environment and the other connecting the local environment to the large-scale bias. The latter connection is conceptually exactly the kind of correlation that excursion set models are built to explain, while the former could be a correlation which needs additional physical mechanisms to be included in the dynamical models describing halo formation.

Recent studies indicate that the *cosmic web environment* at relatively small scales (of the order of a few virial radii) plays an important role in the assembly bias due to halo formation epoch (Hahn et al. 2009), mass accretion rate (Fakhouri & Ma 2010; Musso et al. 2018), internal velocity dispersion structure (Borzyszkowski et al. 2017), and halo concentration (Paranjape, Hahn & Sheth 2018a). In particular, these studies have revealed an intimate connection between the nature of assembly bias and the immediate environment of a halo (e.g. whether or not the halo lives in a cosmic filament; see also Wang et al. 2011; Shi, Wang & Mo 2015, who studied the dependence of dynamical variables on the local tidal environment). While this is not unexpected – the protohalo patches from which haloes form are correlated with the linear tidal field (Bond & Myers 1996; Sheth, Mo & Tormen 2001) – and analytical excursion set calculations do predict a statistical correlation between halo bias and formation time or concentration (Musso & Sheth 2012; Castorina & Sheth 2013), the specific role of the non-linear cosmic web in establishing assembly bias effects still lacks a first principles understanding.

In this work, we are interested in clean statistical signatures, using N-body simulations, that the physics of hierarchical halo formation splits into distinct contributions from different length-scales. Previous work by some of us has shown that the *tidal anisotropy* in the immediate vicinity of a halo (see below) plays a key role in determining the assembly bias trends defined by halo concentration, particularly at low masses where a large fraction of haloes reside in filaments (Paranjape et al. 2018a). This appears quite natural in hindsight, since the turn-around radius of material currently infalling on to a halo is a few times the halo radius (around the same scale where Paranjape et al. 2018a, defined the tidal anisotropy), and the only relevant physical mechanism at play for collisionless dark matter is the tidal influence of gravity. Our goal here is to extend these ideas to other halo properties (we will study the halo shape, velocity dispersion tensor, and spin) and statistically assess the importance of the tidal anisotropy as an intermediary in explaining assembly bias in these properties.

The paper is organized as follows: in Section 2, we describe our simulations and the measurements of various internal properties of haloes used in this work. In Section 3, we explore the connection between the halo tidal environment and assembly bias in these properties. We summarize known results before presenting our main findings which indicate that the tidal anisotropy of the cosmic web in the halo vicinity is an important indicator of *all* assembly bias trends. In Section 4, we present tests of potential physical explanations of our results, showing that the connection between tidal anisotropy and assembly bias *cannot be explained* by splashback objects or recent mergers. We conclude with a discussion in Section 5. The Appendix presents convergence studies for our numerical techniques and detailed tests of the robustness of our choice of statistics.

Throughout, we use a spatially flat ΛCDM cosmology with total matter density parameter $\Omega_{\rm m} = 0.276$, baryonic matter density $\Omega_{\rm b} = 0.045$, Hubble constant $H_0 = 100\,h\,{\rm km\,s^{-1}\,Mpc^{-1}}$ with $h = 0.7$, primordial scalar spectral index $n_{\rm s} = 0.961$ and r.m.s. linear fluctuations in spheres of radius $8\,h^{-1}$ Mpc, $\sigma_8 = 0.811$, with a transfer function generated by the code CAMB (Lewis, Challinor & Lasenby 2000).[1]

2 SIMULATIONS AND HALO PROPERTIES

We use N-body simulations of collisionless CDM in cubic, periodic boxes performed using the tree-PM code GADGET-2 (Springel 2005).[2] These simulations, which we briefly describe here, are the same as those used by Paranjape et al. (2018a) in their analysis. We use two configurations: a lower resolution one having 10 independent realizations, and two realizations of a smaller volume, higher resolution box. All boxes were run using $N_{\rm p} = 1024^3$ particles, with the lower (higher) resolution configuration having a box of comoving length $L = 300\,(150)\,h^{-1}$ Mpc, corresponding to a particle mass of $m_{\rm p} = 1.93 \times 10^9\,(2.4 \times 10^8)\,h^{-1}\,{\rm M}_\odot$. The force resolution parameter ϵ in each case was set to 1/30 of the mean comoving interparticle spacing, leading to $\epsilon = 9.8\,(4.9)\,h^{-1}$ kpc for the lower (higher) resolution, while PM forces were computed on a 2048^3 grid in each case.

Initial conditions for the lower (higher) resolution boxes were generated at a starting redshift $z_{\rm in} = 49\,(99)$ using the code MUSIC (Hahn & Abel 2011)[3] with second-order Lagrangian perturbation theory. Haloes were identified using the code ROCKSTAR (Behroozi, Wechsler & Wu 2013a)[4] which performs a Friends-of-Friends (FoF) algorithm in six-dimensional phase space. For the higher resolution boxes, we stored 201 snapshots equally spaced in the scale factor $a = 1/(1+z)$ ($\Delta a = 0.004615$) between $z = 12$ and $z = 0$, which we used to produce merger trees using the code CONSISTENT-TREES (Behroozi et al. 2013b).[5] The simulations and analysis were performed on the Perseus cluster at IUCAA.[6]

[1] http://camb.info
[2] http://www.mpa-garching.mpg.de/gadget/
[3] https://www-n.oca.eu/ohahn/MUSIC/
[4] https://bitbucket.org/gfcstanford/rockstar
[5] https://bitbucket.org/pbehroozi/consistent-trees
[6] http://hpc.iucaa.in

To ensure that our results are not contaminated by substructure and numerical artefacts, we discard all subhaloes identified by ROCKSTAR and further only consider objects whose virial energy ratio $\eta = 2T/|U|$ satisfies $0.5 \leq \eta \leq 1.5$ as suggested by Bett et al. (2007). Our measurements of the tidal environment in the vicinity of the haloes, which we describe below, were performed after Gaussian smoothing on a cubic grid with $N_g = 512^3$ cells. We therefore impose a restriction on the minimum halo mass we study, so as to minimize the contamination to our final results from the resolution imposed by this grid. Based on convergence studies which we discuss below, we choose to analyse haloes with at least 3200 particles for most of the analysis. This gives a mass threshold of 6.2×10^{12} (7.7×10^{11}) $h^{-1}\,\mathrm{M}_\odot$ and leaves us with approximately 19 000 (17 000) objects in each of the lower (higher) resolution boxes.

Throughout, we focus on results at $z = 0$ and study all halo properties as a function of virial mass $M_{\rm vir}$ enclosed in the virial radius $R_{\rm vir}$ as defined using the spherical overdensity prescription of Bryan & Norman (1998). We have checked that qualitatively identical results are obtained when binning haloes according to other mass definitions such as $M_{\rm 200b}$ enclosed inside the radius $R_{\rm 200b}$,[7] or the mass $M_{\rm ell}$ enclosed inside the mass ellipsoid of the halo which is calculated as described in Section 2.3.1 below.

2.1 Measuring halo-by-halo bias

As our indicator of choice for the large-scale density environment of haloes, we use the halo-by-halo bias estimator b_1 described by Paranjape et al. (2018a). A similar variable defined in real space has also been recently used by Han et al. (2019).

This is essentially a halo-centric dark matter overdensity estimate filtered with a window function that is sharp in Fourier space. This sharp k filter is built using k-dependent weights chosen such that the arithmetic mean of b_1 for any population of haloes is identical to the usual Fourier space linear bias of this population, as measured by the ratio of the halo-matter cross-power spectrum $P_{\rm hm}(k)$ and the matter power spectrum $P_{\rm mm}(k)$ at small k.

In detail, denoting the discrete Fourier transform of the dark matter density contrast as $\delta(\mathbf{k})$ evaluated at the grid location $\mathbf{k}$ in Fourier space, the bias for halo h is calculated as

$$b_{1,h} = \sum_{\text{low-}k}\, w_k \left[\left\langle \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}(h)} \delta^*(\mathbf{k}) \right\rangle_k / P_{\rm mm}(k) \right], \tag{1}$$

where $P_{\rm mm}(k) = \langle\, \delta(\mathbf{k})\delta^*(\mathbf{k}) \,\rangle_k$ and $\langle \ldots \rangle_k$ denotes a spherical average over modes contained in a bin of k. The quantity $\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}(h)}$ corresponds to a weighted average of phase factors over the configuration space cell $\mathbf{x}(h)$ containing the halo h, and seven of its neighbouring cells, using weights appropriate for a cloud-in-cell (CIC) interpolation. We sum over low-k modes in the simulation box, using the ranges $0.025(0.05) \lesssim k/(h\,\mathrm{Mpc}^{-1}) \lesssim 0.09$ for the lower (higher) resolution configuration, additionally weighting by the number of modes $w_k \propto k^3$ for logarithmically spaced bins (with $\sum_{\text{low-}k} w_k = 1$).

We emphasize that the resulting bias estimate is an indicator of halo environment at large scales $\gtrsim 30\,h^{-1}\,\mathrm{Mpc}$ where bias is approximately linear and scale independent. This should be contrasted with other estimators employed in the literature, such as marked correlation functions or ratios of correlation functions at scales $\lesssim 10\,h^{-1}\,\mathrm{Mpc}$ (Wechsler et al. 2006; Villarreal et al. 2017; Mansfield & Kravtsov 2019). The interpretation of assembly bias trends of these estimators is likely to be complicated by non-linearity and/or scale dependence of bias (Sunayama et al. 2016; Paranjape & Padmanabhan 2017). See also Han et al. (2019) for tests of linearity at smaller scales.

The primary advantage of using a halo-by-halo estimator of bias is that it allows us to treat halo bias on par with any other halo-centric or internal property. In particular, we are able to directly probe the correlation of the scatter in halo bias with other variables by calculating appropriate correlation coefficients between b_1 and these variables, *without* having to bin haloes. We will build our main analysis below using such correlation coefficients.

2.2 Measuring the halo tidal environment

As our main indicator of a halo's non-linear local environment, we will use the *tidal anisotropy* variable α introduced by Paranjape et al. (2018a). This is constructed using measurements of the tidal tensor at halo locations, as follows.

First, the density field $\delta(\mathbf{x})$ evaluated using CIC interpolation on a cubic lattice is used to evaluate the tidal tensor $\psi_{ij}(\mathbf{x}) \equiv \partial^2\psi/\partial x^i \partial x^j$ by inverting the normalized Poisson equation $\nabla^2\psi = \delta$ in Fourier space. While doing so, we apply a range of Gaussian smoothing filters $\mathrm{e}^{-k^2 R_{\rm G}^2/2}$ to generate multiple smoothed versions $\psi_{ij}(\mathbf{x}; R_{\rm G})$ of the tidal tensor on the lattice. We then interpolate these in configuration space to the location $\mathbf{x}_h$ of halo h and also interpolate in smoothing scales to the size R_h of the halo (see below), thus creating a halo-by-halo catalogue of tidal tensor estimates $\psi_{ij}(\mathbf{x}_h; R_h)$.

Diagonalizing this halo-centric tidal tensor and denoting its eigenvalues by $\lambda_1 \leq \lambda_2 \leq \lambda_3$ (for brevity, we will drop the subscript h in the following), we then construct the halo-centric overdensity δ using

$$\delta = \lambda_1 + \lambda_2 + \lambda_3, \tag{2}$$

and the halo-centric tidal shear q^2 using (Heavens & Peacock 1988; Catelan & Theuns 1996)

$$q^2 = \frac{1}{2}\left[(\lambda_2 - \lambda_1)^2 + (\lambda_3 - \lambda_1)^2 + (\lambda_3 - \lambda_2)^2\right]. \tag{3}$$

The halo-centric tidal anisotropy α is then defined by

$$\alpha = \sqrt{q^2}/\,(1 + \delta)\,. \tag{4}$$

The choice of smoothing scale $R_{\rm G} = R_h$ for each halo is driven by our requirement of a measure of the local halo tidal environment which correlates well with the *large-scale* environment as measured by b_1 above. As shown by Paranjape et al. (2018a), the choice $R_h \sim 4R_{\rm 200b}$ is the largest halo-scaled smoothing radius[8] for which α as defined above correlates more tightly with b_1 than does δ at the same scale (see also Appendix A2).

The measurements of the tidal tensor and associated variables above depend on the choice of grid size used for the original CIC interpolation. For a given grid size, the requirement that the sphere of radius $\sim 4R_{\rm 200b}$ be sufficiently well resolved leads to a lower limit on halo mass. Appendix A1 presents a convergence study using which we conclude that a 512^3 grid is sufficient for our purposes, provided

[7] $R_{\rm 200b}$ is the halo-centric radius which encloses a spherical overdensity of 200 times the background matter density.

[8] In practice, we set $R_h = 4R_{\rm 200b}/\sqrt{5}$, the 'Gaussian equivalent' of the spherical tophat scale $4R_{\rm 200b}$. The factor $\sqrt{5}$ is most easily understood by Taylor expanding the Fourier transforms of the Gaussian and spherical tophat filters and equating the terms proportional to k^2.

we restrict attention to haloes with ≥3200 particles enclosed inside $R_{\rm vir}$. These are the default choices for our analysis.

Fig. A1 also shows that α and δ as defined above are, in fact, positively correlated. This is potentially a cause for concern because any statements regarding the correlation between α and halo properties could simply be reflecting a correlation between δ and those properties (see e.g. Shi & Sheth 2018). To assess the level to which this is true, we perform a detailed comparison of these correlations in Appendices A2 and A3, finding that α is in fact a better indicator of *all* correlations with halo properties than is δ. (We remind the reader that we define both α and δ at scales $\sim 4R_{\rm 200b}$ for the reasons discussed above.)

We also note that other estimators of tidal anisotropy such as $\sqrt{q^2}/(1+\delta)^{\mu}$ with some constant μ can decrease the correlation strength between α and δ. E.g. Alam et al. (2019) find that setting $\mu \simeq 0.55$ works well for $R_{\rm G} = 5\,h^{-1}$ Mpc and haloes selected so as to describe a sample of galaxies in the Sloan Digital Sky Survey. However, the dependence of the value of μ on smoothing scale, halo mass, large-scale environment, or sample selection, and the origin of any specific value, is unclear. We therefore prefer to work with our definition (4), which is a regular function of $1+\delta$, and explicitly check for systematic biases due to correlations with δ.

For example, since α and δ are positively correlated, one might ask whether the variable $\alpha_{(2)} \equiv \sqrt{q^2}/(1+\delta)^2$, which is also a regular function of $1+\delta$, might perform better. Indeed, we find that $\alpha_{(2)}$ correlates very weakly with δ over our entire mass range (see also Haas, Schaye & Jeeson-Daniel 2012, for an alternative tidal variable which also correlates weakly with the isotropic overdensity). However, the $b_1 \leftrightarrow \alpha_{(2)}$ correlation is weaker than the $b_1 \leftrightarrow \alpha$ correlation, and is instead similar to the $b_1 \leftrightarrow \delta$ correlation seen in Fig. A2, thus making $\alpha_{(2)}$ unsuitable for our purposes. Thus, although the tidal anisotropy variable α as defined in equation (4) is strictly a combination of anisotropy and density, its superior correlation with b_1 as compared to pure anisotropy (or pure density) variables makes α our variable of choice for assembly bias studies.

2.3 Measuring internal halo properties

We will study the correlations between the halo environment (as characterized by halo bias b_1 and tidal anisotropy α) and a number of internal halo properties. For the latter, we will focus on *scalar* variables describing the anisotropy of the halo shape and velocity dispersion tensors, halo concentration, and spin. We discuss the measurements of each of these below.

Throughout this work, for any halo we discard particles that are either not contained inside the phase space FoF grouping provided by ROCKSTAR or are gravitationally unbound to the halo. All internal halo properties are therefore calculated using only gravitationally bound particles belonging to the FoF group of each halo.

2.3.1 Mass ellipsoid tensor

As a part of its post-processing analysis, the ROCKSTAR code measures the mass ellipsoid tensor (or shape tensor) M_{ij} of each halo using the iterative procedure prescribed by Allgood et al. (2006). This tensor is evaluated as

$$M_{ij} = \sum_{n\in\text{halo}} x_{n,i}\,x_{n,j}/r_n^2, \tag{5}$$

where $i, j = 1, 2, 3$ refer to the coordinate directions, $\mathbf{x}_n$ is the comoving position of the nth particle in the halo *with respect to the halo centre of mass*, and r_n^2 is the comoving ellipsoidal distance of this particle from the centre of mass given by $r_n^2 = x_n^2 + y_n^2/(b/a)^2 + z_n^2/(c/a)^2$. Here, we defined $a^2 \geq b^2 \geq c^2$ as the ordered eigenvalues of M_{ij}. Since the calculation of the ellipsoidal distance requires knowledge of the eigenvalue ratios, this is done by an iterative procedure with a starting guess of equal eigenvalues and subsequent updates in each iteration after estimating M_{ij} using equation (5) and diagonalizing it. The calculation sets the semimajor axis of the ellipsoid equal to the halo virial radius $R_{\rm vir}$ and sums over all (bound, FoF) particles in the halo. We refer the reader to Allgood et al. (2006) for further details of the procedure.

Denoting the final converged eigenvalues with the same notation $a^2 \geq b^2 \geq c^2$, we use the ratio c/a as a measure of the asphericity of the mass ellipsoid tensor. This variable is convenient since its values are bounded between $0 \leq c/a \leq 1$, with zero corresponding to a highly aspherical halo and unity to a spherical halo. We have checked that using other measures of asphericity which include information on the intermediate axis, such as the triaxiality variable $\mathcal{T} = (a^2 - b^2)/(a^2 - c^2)$ (Franx, Illingworth & de Zeeuw 1991), lead to qualitatively identical results.

2.3.2 Velocity ellipsoid tensor

We have modified ROCKSTAR so as to calculate the velocity ellipsoid tensor which is a measure of the anisotropic velocity dispersion of the dark matter particles constituting a halo. For a halo with N particles, this tensor is given by

$$V_{ij}^2 = \frac{1}{N}\sum_{n\in\text{halo}} \left(v_{n,i} - \langle\, v_i \,\rangle\right)\left(v_{n,j} - \left\langle\, v_j \,\right\rangle\right), \tag{6}$$

where $\mathbf{v}_n$ is the peculiar velocity of nth dark matter particle and $\langle\, \mathbf{v} \,\rangle = \sum_{n\in\text{halo}} \mathbf{v}_n/N$ is the bulk peculiar velocity of the halo.

Similar to the mass ellipsoid tensor, we denote the eigenvalues of V_{ij}^2 by $a_v^2 \geq b_v^2 \geq c_v^2$ and use the ratio c_v/a_v as a measure of the asphericity of the velocity ellipsoid. For consistency with the calculation of the mass ellipsoid tensor, we restrict the sum in equation (6) to be over those (bound, FoF) particles contained inside the mass ellipsoid defined by equation (5).

2.3.3 Velocity anisotropy

A related property of the halo is the velocity anisotropy β defined as (e.g. Binney & Tremaine 1987)

$$\beta = 1 - \sigma_{\rm t}^2/\left(2\sigma_{\rm r}^2\right), \tag{7}$$

where $\sigma_{\rm r}^2$ and $\sigma_{\rm t}^2$ are the radial and tangential velocity dispersion, respectively, of the particles in the halo. These are calculated by first projecting the velocity of each particle in the halo along and perpendicular to the radial direction (defined by the centre of mass) and then computing the variance of each component separately over all particles. As before, we restrict attention to the particles contained inside the mass ellipsoid defined by equation (5). We have modified ROCKSTAR to compute β for each halo alongside the velocity and mass ellipsoid calculations described previously.

Although β is clearly related to the velocity ellipsoid tensor, it is worth keeping in mind that β also crucially depends on the *shape* of the halo when computing the radial and tangential dispersions. Thus, the velocity anisotropy β captures information from the full phase space of the halo, unlike the mass and velocity ellipsoid tensors individually. We return to this point below. For now, we note that this variable takes values in the range $-\infty < \beta \leq 1$, with $\beta = 0$ corresponding to an isotropic velocity ellipsoid and the positive and

negative extremes of the allowed range corresponding, respectively, to radially and tangentially dominated velocity dispersions.

Finally, unlike standard applications which study β as a function of radial distance, here we define the radial and tangential dispersions, and hence β, by averaging over all (bound, FoF) particles in the halo. It would also be interesting to explore the radial dependence of β *vis a vis* the environmental correlations we are focusing on, an exercise we leave for future work.

2.3.4 *Concentration*

By default, ROCKSTAR performs a least-squares fit of the spherically averaged dark matter profile of each halo to the universal NFW form (Navarro et al. 1997)

$$\rho(r) = \frac{\rho_{\rm s}}{(r/r_{\rm s})\,(1 + r/r_{\rm s})^2}, \tag{8}$$

where $\rho_{\rm s}$ is a normalization constant related to the mass of the halo and $r_{\rm s}$ is the scale radius. The halo concentration $c_{\rm vir}$ is then defined as

$$c_{\rm vir} \equiv R_{\rm vir}/r_{\rm s}. \tag{9}$$

Halo concentration correlates well with formation epoch (Navarro et al. 1997; Wechsler et al. 2002; Ludlow et al. 2013), and its dependence on halo mass and environment has been thoroughly studied in the literature (Bullock et al. 2001a; Ludlow et al. 2014; Diemer & Kravtsov 2015, see also below). We include $c_{\rm vir}$ in our analysis as a proxy for formation epoch and to compare with the assembly bias trends of other variables.

2.3.5 *Spin*

The dimensionless spin parameter is given by

$$\lambda \equiv \frac{J\,|E|^{1/2}}{G M_{\rm vir}^{5/2}}, \tag{10}$$

where J is the magnitude of the angular momentum, E the total energy, and $M_{\rm vir}$ the mass of the halo, with G being Newton's constant (Peebles 1969). By default, ROCKSTAR calculates λ for each halo using its bound, FoF particles inside $R_{\rm vir}$; we use this measurement in our analysis below.

We have also checked that using the alternative definition of dimensionless spin λ' proposed by Bullock et al. (2001b, this is also calculated by ROCKSTAR) leads to identical results, where

$$\lambda' \equiv \frac{J_{\rm vir}}{\sqrt{2} M_{\rm vir} R_{\rm vir} V_{\rm vir}} \tag{11}$$

with $J_{\rm vir}$ being the angular momentum inside a sphere of radius $R_{\rm vir}$ containing mass $M_{\rm vir}$, and where $V_{\rm vir} = \sqrt{G M_{\rm vir}/R_{\rm vir}}$ is the halo circular velocity at radius $R_{\rm vir}$.

Similar to halo concentration, the distribution of spin as a function of halo mass and its correlation with other halo properties as well as large-scale environment is also well studied in the literature (e.g. Bullock et al. 2001b; Bett et al. 2007; Rodríguez-Puebla et al. 2016; Johnson et al. 2019, see also below). The measurement of the spin parameter is rather sensitive to the particle resolution, with order unity errors accrued for haloes sampled with a few hundred particles (Oñorbe et al. 2014; Benson 2017), and this can in principle substantially affect any conclusions regarding correlations between spin and other variables. Since we only consider haloes sampled with $\geq$3200 particles, however, we expect these numerical errors in our bins of lowest particle count to be $\lesssim$25 per cent at the object-by-object level (see fig. 3 of Benson 2017). We therefore do not expect any of our conclusions regarding spin assembly bias to be altered as a consequence of particle resolution.

Fig. A6 shows the distributions of each of these variables for a few narrow mass ranges. See Appendix A5 for a discussion of the associated trends.

3 ASSEMBLY BIAS AND TIDAL ENVIRONMENT

In this section, we use measurements of halo bias, tidal anisotropy, and the various internal halo properties discussed in the previous section to assess the nature of the statistical correlations between all these quantities. We start using our simulations to recapitulate some known results on assembly bias, followed by our new statistical analysis.

3.1 Known results

Fig. 1 summarizes previously known assembly bias / secondary bias trends due to halo velocity anisotropy variables β and c_v/a_v (left-hand panel) and halo shape c/a, concentration $c_{\rm vir}$, spin λ, and tidal anisotropy α (right-hand panel). In each panel, upward (downward) triangles indicate the mean halo bias in the upper (lower) quartiles of the respective quantity, at fixed halo mass. Additionally, the circles in the left-hand panel show the mean bias for all haloes at fixed mass.

We see that haloes that are aspherical either in shape (small c/a) or velocity dispersion (small c_v/a_v) are less clustered than more spherical haloes. The split by velocity anisotropy β shows that haloes dominated by more radial orbits ($\beta > 0$) are less clustered than tangentially dominated haloes. Correspondingly, haloes with smaller spin values are less clustered than those with higher spin. The split by halo concentration shows a more complex trend, with highly concentrated haloes being less clustered at high masses but more clustered at low masses, the inversion occurring near $M_{\rm vir} \sim 10^{13}\,h^{-1}\,{\rm M}_\odot$. Finally, haloes in isotropic environments (small α) are substantially less clustered than those in anisotropic environments.

The assembly bias trend with halo concentration (as well as formation time, which we do not show here) has been widely discussed in the literature (see e.g. Wechsler et al. 2006; Jing et al. 2007; Angulo, Baugh & Lacey 2008; Dalal et al. 2008; Desjacques 2008; Faltenbacher & White 2010; Sunayama et al. 2016; Lazeyras, Musso & Schmidt 2017; Paranjape & Padmanabhan 2017). The inversion of the trend is related to the tidal anisotropy of the halo environment; a large fraction of low-mass haloes live in highly anisotropic and biased environments[9] such as cosmic filaments, unlike more isolated haloes which dominate their environment and follow the trends predicted by standard spherical collapse models (Paranjape et al. 2018a). There are also indications that the trend in velocity anisotropy β may be connected to the tidal environment, with low-mass haloes accreting in filaments being dominated by tangential orbits; such haloes should inherit high values of large-scale bias from their parent filaments (Borzyszkowski et al. 2017).

The monotonic dependence of halo bias on halo asphericity c/a and spin λ at fixed mass in the right-hand panel of Fig. 1 is consistent with the trends noted previously in the literature using configuration space definitions of bias (Bett et al. 2007; Gao & White 2007;

[9]We discuss the so-called 'splashback' haloes in Section 4.

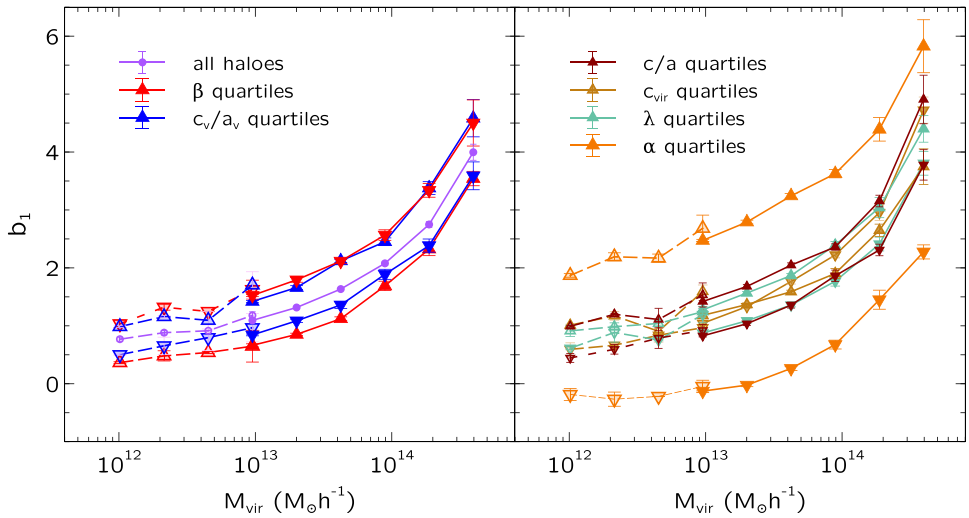

Figure 1. Summary of known assembly (or secondary) bias trends. The symbols joined by lines show measurements of halo bias b_1 (Section 2.1) averaged over haloes in bins of mass $M_{\rm vir}$ for different populations. The circles in the left-hand panel show results for the full halo population in each mass bin. The triangles of different colours in each panel indicate measurements at fixed mass but focusing on haloes in the upper quartile (upward triangles) and lower quartile (downward triangles) of a secondary property. The left-hand panel shows results for the secondary property being velocity anisotropy β (Section 2.3.3) and velocity ellipsoid asphericity c_v/a_v (Section 2.3.2). The right-hand panel shows results for halo shape asphericity c/a (Section 2.3.1), concentration $c_{\rm vir}$ (Section 2.3.4), spin λ (Section 2.3.5), and the tidal anisotropy α (Section 2.2). In each panel, the filled symbols joined with solid lines show the mean over 10 realizations of the lower resolution box, with error bars showing the scatter around the mean, while open symbols joined with dashed lines show measurements using two realizations of the higher resolution box. We see that the tidal anisotropy α has, by far, the strongest trend with halo bias at fixed mass.

Faltenbacher & White 2010; Johnson et al. 2019) (see also van Daalen, Angulo & White 2012, for a study of shape- and spin-dependent clustering at Mpc scales).

As regards the asphericity of the velocity ellipsoid c_v/a_v or related variables, we are unaware of any work other than Faltenbacher & White (2010) that has discussed the corresponding assembly bias trend. It is therefore worth commenting on the nature of this trend before proceeding. We see in the left-hand panel of Fig. 1 that the amplitude of the trend with c_v/a_v is only slightly weaker than that with β. The nature of the trend is quite interesting, however, since it says that haloes with spherical velocity ellipsoids cluster less strongly than aspherical ones. On the one hand, this suggests a potential connection with the trend shown by the asphericity of the shape tensor c/a which is qualitatively identical. On the other, it is also tempting to compare with the trend due to β. Keeping in mind that perfectly spherical velocity ellipsoids would correspond to $\beta = 0$, it is clear that the trend defined by upper and lower quartiles of β is actually sensitive to additional information about haloes with aspherical velocity ellipsoids, by splitting these into radially dominated (upper β quartile) and tangentially dominated (lower β quartile) haloes (cf. the discussion earlier regarding the connection between β and the full phase space of the halo.)

It is clear from Fig. 1 that the trend between halo bias b_1 and the local tidal anisotropy α is the strongest amongst all the secondary bias trends. In fact, defining α at approximately four times the halo radius ensures that this correlation is stronger than that between b_1 and the local overdensity δ of the halo environment measured at the same scale (Paranjape et al. 2018a). Moreover, the definition of α is such that this variable would be statistically independent of the very large-scale overdensity in the (Gaussian random) initial conditions, unlike δ at the same scale (Sheth & Tormen 2002). The fact that α and b_1 correlate so strongly is then highly suggestive of a physical link between these quantities related to the non-linear dynamics of halo formation (see also Castorina et al. 2016). The strength of the $b_1 \leftrightarrow \alpha$ correlation will be important below.

3.2 Disentangling multiscale correlations using conditional correlation coefficients

As discussed in the Introduction, we are interested in identifying a clean statistical signature that contributions from different length-scales might segregate into distinct correlations: one between internal halo properties and the local cosmic web environment and the other between the local web and large-scale halo bias. A convenient approach to addressing this issue is to use the concept of conditional correlation coefficients (Han et al. 2019), as we describe next. This analysis is made possible by our use of a halo-by-halo measurement of bias that does not require haloes to be binned.

Consider three standardized (i.e. zero mean, unit variance) *Gaussian* variables a, b, c with mutual correlation coefficients γ_{ab}, γ_{bc}, and γ_{ca}. The conditional distribution $p(b, c|a)$ is then a bivariate Gaussian with variances ${\rm Var}(b|a) = 1 - \gamma_{ab}^2$, ${\rm Var}(c|a) = 1 - \gamma_{ac}^2$ and the conditional covariance

$$ {\rm Cov}(b, c|a) = \gamma_{bc} - \gamma_{ab}\gamma_{ac} \equiv \gamma_{bc|a}. \tag{12} $$

The key point to note is that, if $\gamma_{\rm bc|a} = 0$, then the conditional distributions of b and c at fixed a are independent: $p(b, c|a) = p(b|a)p(c|a)$. Bayes' theorem then implies that the conditional distribution of c is independent of b: $p(c|a, b) = p(c|a)$. In the present context, to the extent that any statistical correlation between physical variables should ultimately have a physical origin, this would strongly suggest that the statistical connection between c and b is linked by (at least) two physical mechanisms, one connecting c to a and the other connecting a to b.

This discussion shows that the vanishing of $\gamma_{bc|a} = \gamma_{bc} - \gamma_{ab}\gamma_{ac}$ is a useful diagnostic of the conditional independence of c on b. Although we phrased the discussion in terms of a multivariate Gaussian for $p(a, b, c)$, the fact that this distribution is non-Gaussian is not as large a concern as one might have imagined. Rather, the significance of $\gamma_{bc|a} = 0$ is tied to the assumption that c can be well

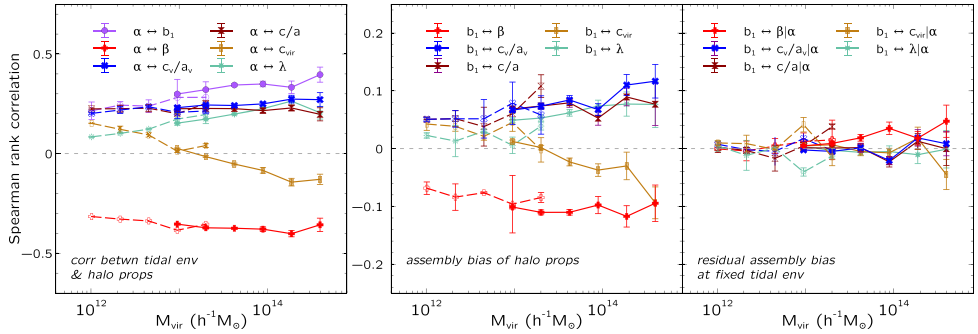

Figure 2. Correlations between internal halo properties, tidal environment, and large-scale bias. (Left-hand panel): Spearman rank correlation coefficients, for haloes in bins of mass $M_{\rm vir}$, between tidal anisotropy α and other halo properties, including $\gamma_{\alpha b_1}$ with large-scale bias b_1 and $\gamma_{\alpha c}$ with internal properties $c \in \{\beta,\, c_v/a_v,\, c/a,\, c_{\rm vir},\, \lambda\}$ (see caption of Fig. 1). In the legend, each coefficient γ_{ab} is represented by the symbol $a{\leftrightarrow}b$. (Middle panel): Assembly bias trends seen using Spearman rank correlation coefficients $\gamma_{b_1 c}$ between halo bias and each internal property c (cf. Fig. 1). (Right-hand panel): Conditional correlation coefficients $\gamma_{b_1 c|\alpha}$ (equation 12) for each internal property c. Note that the vertical axis in the middle and right-hand panels is zoomed in by a factor approximately three as compared to the left-hand panel. The formatting of symbols (filled versus empty) and lines (solid versus dashed) is identical to that in Fig. 1. The right-hand panel shows the main result of this work: each conditional coefficient $\gamma_{b_1 c|\alpha}$ is substantially smaller in magnitude than the corresponding unconditional coefficient $\gamma_{b_1 c}$ in the middle panel. Thus, conditioning on tidal anisotropy α largely accounts for the assembly bias trend of *all* internal halo properties. See the text for a discussion.

approximated by a model which is linear in a and b (see equations 3 and 4 in Bernardi et al. 2003). It is just that, for a multivariate Gaussian, the linear model is exact.

Nevertheless, to minimize systematic errors, we will rely on measurements of Spearman's rank correlation coefficients for each pair of variables, which standardizes all the distributions before computing correlations. Below, we will also discuss tests of the robustness of this choice of statistics.

3.3 Tidal anisotropy as an indicator of assembly bias

Our motivation behind setting up the correlation analysis in the previous section was to explore the possibility that assembly bias correlations between internal halo properties and large-scale bias might be explained using the separate correlations of each of these with some intermediate-scale environmental variable. In this context, it is worth mentioning that previous investigations of assembly bias have failed to identify any single environmental variable that might be responsible for correlations between halo bias and multiple internal halo properties (Villarreal et al. 2017; Xu & Zheng 2018). The fact that tidal anisotropy α shows by far the strongest correlation with halo bias makes α a promising candidate for such a variable.

In the language of the previous section therefore we will now think of a as the tidal anisotropy α, b as halo bias b_1, and c as any one of the internal halo properties $\{\beta,\, c_v/a_v,\, c/a,\, c_{\rm vir},\, \lambda\}$. Below we will also report the results of analysing other permutations and combinations of variables, including using intermediate-scale overdensity δ as the environmental variable.

Fig. 2 shows the main results of this paper. The left-hand panel shows Spearman rank correlation coefficients (for haloes in fixed bins of $M_{\rm vir}$)[10] between the tidal anisotropy α and other halo properties including halo bias b_1 and all internal properties $c \in \{\beta,\, c_v/a_v,\, c/a,\, c_{\rm vir},\, \lambda\}$. This panel summarizes a number of previously known results, including the observations that, at fixed mass, haloes in more anisotropic tidal environments tend to be more strongly clustered ($\alpha{\leftrightarrow}b_1$, Hahn et al. 2009; Paranjape et al. 2018a), more concentrated ($\alpha{\leftrightarrow}c_{\rm vir}$, Paranjape et al. 2018a), more spherical ($\alpha{\leftrightarrow}c/a$, Wang et al. 2011), with higher spin ($\alpha{\leftrightarrow}\lambda$, Hahn et al. 2009; Wang et al. 2011), and have more tangentially dominated velocity distributions ($\alpha{\leftrightarrow}\beta$, Borzyszkowski et al. 2017). Additionally, we see that objects in anisotropic environments also have more spherical velocity ellipsoids ($\alpha{\leftrightarrow}c_v/a_v$), with a correlation very similar at all masses to that between α and the mass ellipsoid asphericity c/a.

The middle panel of Fig. 2 summarizes the known assembly bias trends discussed in Section 3.1. We see that the strength and sign of the correlation coefficients at any halo mass is perfectly consistent with the results of the previous binned analysis (Fig. 1) which focused on the extremes of the distributions of internal halo properties. Note that we have zoomed in on the vertical axis as compared to the left-hand panel; the correlations of halo properties with large-scale bias are weaker (by approximately a factor approximately three in each case) than the respective correlations with the local tidal environment.

The right-hand panel of Fig. 2 shows our main new result: we display the conditional correlation coefficients $\gamma_{b_1 c|\alpha}$ (calculated using equation 12) for each internal property $c \in \{\beta,\, c_v/a_v,\, c/a,\, c_{\rm vir},\, \lambda\}$. The vertical scale is identical to that in the middle panel which showed the corresponding unconditional coefficients using the same scheme for colours and markers. In each case, we see that the conditional coefficients are substantially smaller in magnitude than the corresponding unconditional ones at all masses (by a factor approximately three or so at low masses). In fact, except for β around $\sim 10^{14}\, h^{-1}\, {\rm M}_\odot$ (see below), the conditional coefficients are scattered around zero in all cases over the entire mass range, implying that α is an excellent candidate for the primary environmental variable responsible for halo assembly bias trends. In support of this argument, we find in Appendix A4 (see also below) that conditioning on α performs much better at decreasing assembly bias strength than conditioning on δ at the same scale, despite the fact that α and δ are correlated. Further, in order to quantify exactly how close to zero the conditional coefficients $\gamma_{b_1 c|\alpha}$ are in Fig. 2, we use two methods.

[10]We have checked that all our results are robust to our choice of binning. Namely, we found identical results for all correlation trends when doubling the number of mass bins. Thus our results are unaffected by mass-dependent trends in any correlation.

Table 1. Top 10 conditional correlation coefficients $b_1 \leftrightarrow X \,|\, Y$ rank-ordered by reduced Chi-squared values for comparison to zero. Here X, Y were allowed to be any two of the variables $\{\beta, c_v/a_v, c/a, c_{\rm vir}, \lambda, \alpha, \delta\}$, i.e. treating environmental variables on par with internal halo properties. Chi-squared values were calculated using measurements in nine mass bins, with mean values and errors computed using two realizations of the high-resolution and 10 realizations of the low-resolution simulations. The first column labels the conditional coefficient being tested and the second column reports the value of reduced Chi-squared for nine degrees of freedom. Values below the horizontal line correspond to p-values $<10^{-4}$.

Conditional corr. coeff.	χ^2/dof
$b_1 \leftrightarrow \lambda \,|\, \alpha$	0.90
$b_1 \leftrightarrow c/a \,|\, \alpha$	0.97
$b_1 \leftrightarrow c_v/a_v \,|\, \alpha$	1.76
$b_1 \leftrightarrow c_{vir} \,|\, \alpha$	2.18
$b_1 \leftrightarrow \lambda \,|\, \beta$	3.13
$b_1 \leftrightarrow \beta \,|\, \alpha$	4.34
$b_1 \leftrightarrow c_{\rm vir} \,|\, \delta$	11.89
$b_1 \leftrightarrow c_{\rm vir} \,|\, c_v/a_v$	14.52
$b_1 \leftrightarrow c_{\rm vir} \,|\, c/a$	24.00
$b_1 \leftrightarrow c/a \,|\, c_v/a_v$	29.96

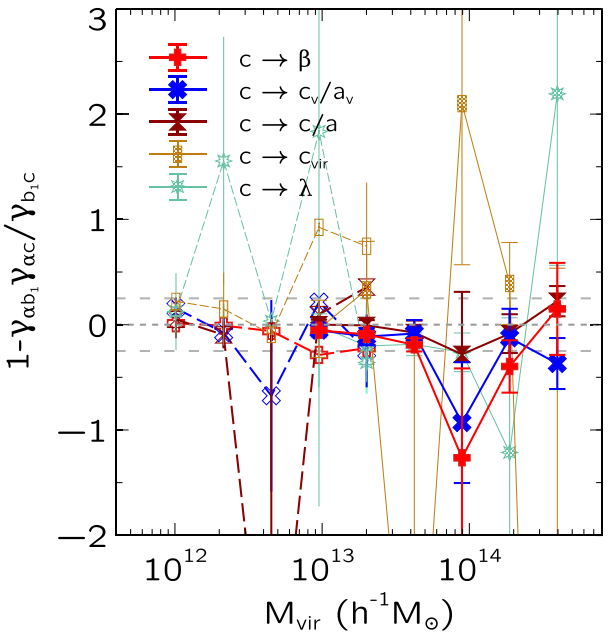

Figure 3. Relative correlation coefficients, calculated as the ratio $\gamma_{b_1 c|\alpha}/\gamma_{b_1 c}$ using measurements from the middle and right-hand panels of Fig. 2 (and formatted identically) for internal halo properties $c \in \{\beta, c_v/a_v, c/a, c_{\rm vir}, \lambda\}$ as indicated. The horizontal dotted line indicates zero and the horizontal dashed lines indicate ± 0.25, i.e. a factor four decrease in the magnitude of $\gamma_{b_1 c|\alpha}$ relative to $\gamma_{b_1 c}$. See the text for a discussion.

The first is a straightforward Chi-squared test which we perform using the results in nine mass bins of the low- and high-resolution boxes where we have reliable error bars on the measurements of $\gamma_{bc|a}$. While performing this test, we also relaxed the assumption of using α as the intermediary between halo properties and bias, exploring multiple other combinations involving either the overdensity δ or one of the internal properties themselves as the intermediary. When the resulting triplet combinations are ordered in increasing order of reduced Chi-squared, we find that triplets involving α as the intermediary produce the best Chi-squared values, while those involving δ perform much worse. Table 1 summarizes these results. We note that the largest discrepancy in our conclusions occurs for the assembly bias variable β at high masses; the residual assembly bias conditioned on α deviates from zero at masses $> 8 \times 10^{13}\, h^{-1}\, {\rm M}_\odot$. This can be seen in the right-hand panel of Fig. 2 and also causes the largest Chi-squared values out of the five assembly bias variables in Table 1.

The second method is to simply construct the ratio $\gamma_{bc|a}/\gamma_{bc}$: if the magnitude of this ratio is small, it means that conditioning on a has indeed substantially decreased the correlation between b and c. This is a particularly useful diagnostic for internal properties such as halo concentration whose correlation with halo bias is the smallest in amplitude of all internal properties. The results are shown in Fig. 3. For the internal properties $\{\beta, c_v/a_v, c/a\}$, we see that the relative correlation coefficient is, in fact, much smaller than unity over nearly the entire halo mass range.

For halo concentration and spin, on the other hand, the relative correlation shows very large fluctuations and noise at higher masses. This is perhaps not surprising considering previous results which suggest that assembly bias signatures at these mass scales are likely caused by other effects (Dalal et al. 2008; Paranjape et al. 2018a). Interestingly, our results from Table 1 and Fig. 3 indicate that α is a particularly good indicator of spin assembly bias in the mass range $\sim 10^{12}$–$10^{14}\, h^{-1}\, {\rm M}_\odot$. This can be compared with the results of Johnson et al. (2019) who found that spin assembly bias can be largely explained using the presence of neighbours of comparable mass. Our results are consistent with theirs, since α represents the anisotropy of the *total* tidal field in the halo vicinity, including the influence of all neighbours.

To summarize, the statistical correlation between large-scale bias b_1 and essentially any internal halo property c that we have studied is consistent with arising from the individual correlations $b_1 \leftrightarrow \alpha$ and $\alpha \leftrightarrow c$, at nearly all halo masses.

3.4 Reliability of chosen statistics

We argued in Section 3.2 that the use of correlation coefficients combined using equation (12) relies essentially on the implicit assumption that the underlying correlations between triplets of variables are linear. Our use of Spearman's rank correlations means that the relevant variables are actually the *ranks* of the physical variables, so that we are dealing with triplets of correlated variables which are individually uniformly distributed. Although the variables are now standardized, their intrinsic correlations are not necessarily linear or even monotonic (see e.g. fig. 12 of Paranjape et al. 2018a, which shows that the median halo concentration is non-monotonic in α at fixed mass), so one might still worry about systematic effects in our analysis. We have therefore performed some explicit tests, which we describe here, to establish the robustness of our conclusions. Our method differs from that of Han et al. (2019) who used Gaussian process regression to explicitly fit for the non-linearity/non-monotonicity of the dependence of halo bias on other variables, thus allowing them to explore a multidimensional bias 'manifold'. Instead, below we demonstrate the robustness of our primary results using direct probes of probability distributions involving b_1, α, and one halo internal property at a time.

We first test the reliability of replacing explicit conditional correlation coefficients (which would require binning of data) with the expression in equation (12) (which uses all available data) in Appendix A4, focusing on the strongest assembly bias signature which is that of the velocity anisotropy β. Fig. A5 shows that explicitly binning in α before computing the correlation coefficient

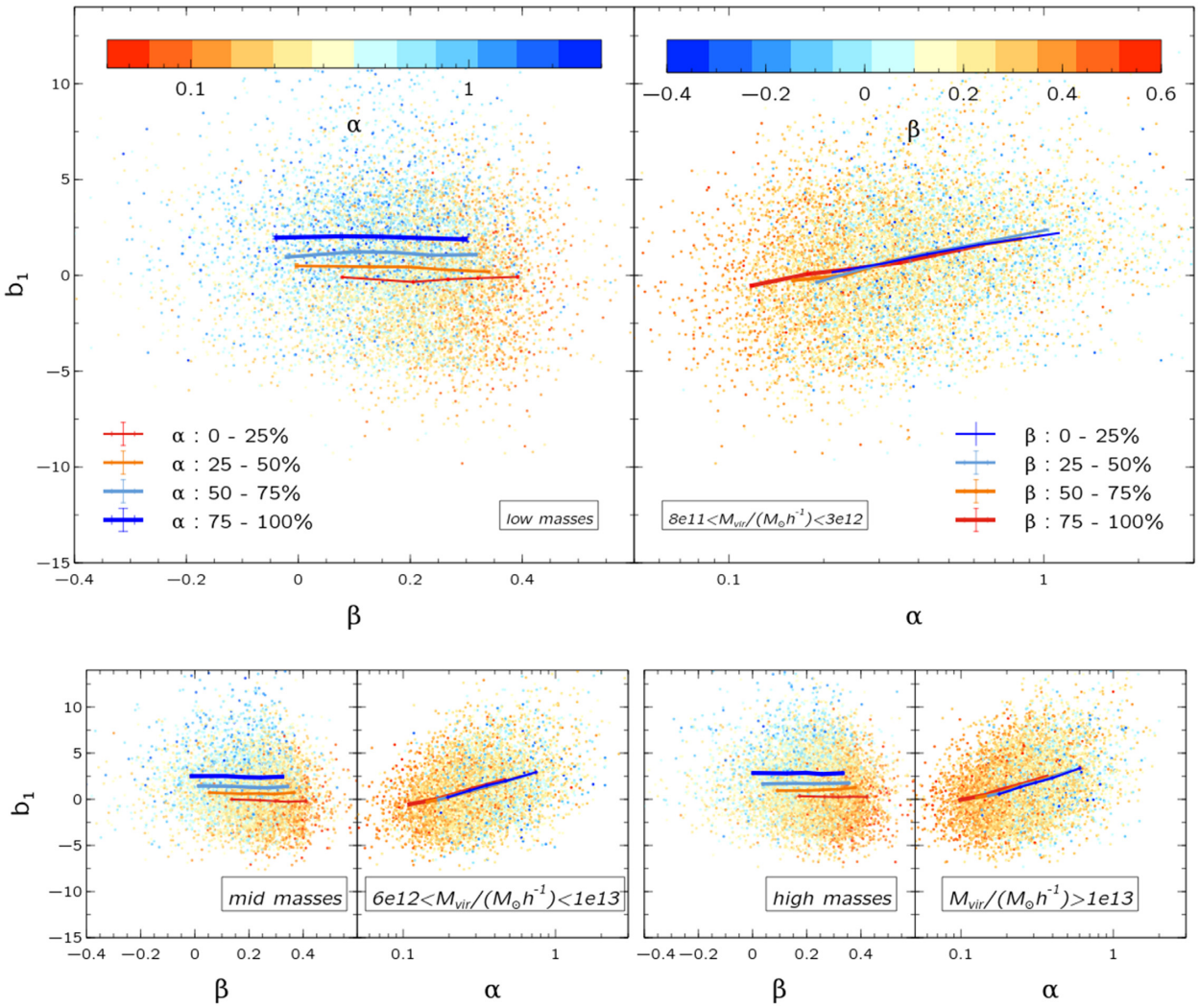

Figure 4. Joint distribution of α, b_1, and β for haloes in three mass ranges: low $[8 \times 10^{11} < M_{\rm vir}/(h^{-1}\,{\rm M}_\odot) < 3 \times 10^{12}]$, mid $[6 \times 10^{12} < M_{\rm vir}/(h^{-1}\,{\rm M}_\odot) < 10^{13}]$, and high $[M_{\rm vir}/(h^{-1}\,{\rm M}_\odot) > 10^{13}]$. (Top left panel): The scatter plot shows β against b_1 with points coloured by α. Each coloured solid line focuses on a quartile of α as indicated in the legend, showing the median b_1 in bins of β (the bins are chosen to be quintiles of β for haloes in each α quartile). (Top right panel): The scatter plot shows α against b_1 with points coloured by β. Similar to the left-hand panel, each coloured solid line now shows the median b_1 in quintiles of α, for haloes selected in a quartile of β as indicated. The results of the two panels are consistent with a correlation structure $p(\alpha, b_1, \beta) \simeq p(\alpha)p(b_1|\alpha)p(\beta|\alpha)$. See the text for a discussion. The bottom panels show the same as the top for mid and high masses.

between b_1 and β does decrease the magnitude of the correlation to nearly zero at all masses and for all α.

To address the concern regarding non-linearity or non-monotonicity of the intrinsic correlations, we focus on three mass ranges: low $[8 \times 10^{11} < M_{\rm vir}/(h^{-1}\,{\rm M}_\odot) < 3 \times 10^{12}]$, mid $[6 \times 10^{12} < M_{\rm vir}/(h^{-1}\,{\rm M}_\odot) < 10^{13}]$, and high $[M_{\rm vir} > 10^{13}\,h^{-1}\,{\rm M}_\odot]$ containing $\sim 10^4$ haloes each and dissect the full distribution of $\{b_1, \alpha, \beta\}$ in Fig. 4. The scatter plots in the top panels of the figure focus on the low-mass range. The top left panel shows the distribution of β and b_1, with the symbols coloured by the value of α. Apart from an overall negative correlation between β and b_1 (cf. middle panel of Fig. 2), we can also see that both these variables are correlated with α by observing that the redder (bluer) points, which correspond to $\alpha \lesssim 0.1$ ($\alpha \gtrsim 1$) are largely confined to the bottom right (top left) of the distribution. Similar conclusions about the correlation between variables can be made from the top right panel which shows the scatter distribution of α and b_1, with the symbols coloured by the value of β.

In order to extract more information on the structure of the joint probability distribution $p(\alpha, b_1, \beta)$, we consider the mean value of bias conditioned on α and β i.e. $\langle\, b_1|\alpha, \beta\,\rangle$. In the three-dimensional space of $\{b_1, \alpha, \beta\}$, this quantity forms a two-dimensional surface whose properties we explore using projections on to the b_1–β and b_1–α planes, as we discuss next.

In the top left panel of Fig. 4, we plot the projection of $\langle\, b_1|\alpha, \beta\,\rangle$ on to the b_1–β plane as solid lines, with each line focusing on haloes in quartiles of α (from red to blue in increasing thickness as α increases.) The overall assembly bias trend between b_1 and β is now visible as the fact that the blue (red) curve having larger (smaller) bias lies towards smaller (larger) β. More interestingly, we see that each of these lines is approximately horizontal; this implies that $\langle\, b_1|\alpha, \beta\,\rangle$ in each quartile of α is independent of β, i.e. $\langle\, b_1|\alpha, \beta\,\rangle \simeq \langle\, b_1|\alpha\,\rangle$. In other words, bias when conditioned on α does not show an assembly bias with β.

Similarly, the projection of $\langle\, b_1|\alpha, \beta\,\rangle$ on to the b_1–α plane is shown in the top right panel as solid lines, with each line focusing

on haloes in quartiles of β (from blue to red in increasing thickness as β increases). All the lines clearly trace out the same locus of positive correlation between b_1 and α, with vertical and horizontal shifts now occurring essentially in perfect tandem as β changes. A simple calculation shows that this is again consistent with the relation $\langle b_1|\alpha, \beta \rangle \simeq \langle b_1|\alpha \rangle$.

These results together give a complete picture of $\langle b_1|\alpha, \beta \rangle$ as being approximately a plane in $\{b_1, \log\alpha, \beta\}$ space and moreover being orthogonal to the b_1–$\log\alpha$ plane. The bottom panels of Fig. 4 show that identical conclusions can be drawn for the mid (bottom left) and low (bottom right) mass bins. We can go even further and ask whether the conditional *variance* $\mathrm{Var}(b_1|\alpha, \beta)$ also displays the same behaviour. Fig. A8 in the Appendix shows that this is indeed the case: the projections of this quantity in the b_1–β and b_1–α planes are consistent with the relation $\mathrm{Var}(b_1|\alpha, \beta) \simeq \mathrm{Var}(b_1|\alpha)$.

These results strongly suggest that the joint distribution $p(\alpha, b_1, \beta)$ itself (as opposed to only its first moment) has a structure consistent with b_1 and β being conditionally independent of each other, when conditioned on α:

$$p(\alpha, b_1, \beta) \simeq p(\alpha)p(b_1|\alpha)p(\beta|\alpha),$$

which then ensures that the trends of $\langle b_1|\alpha, \beta \rangle$ and $\mathrm{Var}(b_1|\alpha, \beta)$ discussed above emerge. Thus, the overall anticorrelation between b_1 and β (assembly bias) is explained by the mutual dependence of these variables on the tidal anisotropy α.

We emphasize that this analysis made *no assumptions* regarding Gaussianity of the variables, monotonicity, or linearity of the trends, etc. We have further verified that essentially identical results are obtained using *all* other internal halo properties considered in this work as well (see Figs A7 and A8 in the Appendix).

The results of this section therefore provide strong support for our claim that the tidal anisotropy α is the primary indicator of assembly bias for a number of internal halo properties. Our tests have further demonstrated that our conclusions are robust to our choice of statistical tools (Spearman rank correlation statistics, with conditional correlation coefficients defined by equation 12). In the next section, we explore other, physical choices related to sample selection which could, in principle, affect our conclusions.

4 THE IMPACT OF SPLASHBACK OBJECTS AND MAJOR MERGERS

The primary analysis of this work presented in Section 3 defined haloes as objects identified as being distinct at the epoch of interest $z = 0$. These haloes therefore also include the small population of so-called 'splashback' haloes (Gill, Knebe & Gibson 2005), which are objects that have passed through one pericentre passage of their eventual host but are currently outside its virial radius. Treating splashback objects equivalently to genuine distinct haloes therefore risks contaminating any signal that involves a correlation with large-scale environment. Indeed, there is considerable evidence that, at low masses, a significant fraction of the assembly bias signal in variables such as halo concentration or age in fact arises from splashback objects (Dalal et al. 2008; Hahn et al. 2009; Sunayama et al. 2016; Villarreal et al. 2017; Mansfield & Kravtsov 2019). It is then important to assess the impact of this small population on our conclusions regarding the influence of the cosmic web environment.

Similarly, the fact that there are strong correlations between tidal environment and internal properties such as halo asphericity in position or velocity space could be connected to the occurrence of recent major merger events. We must therefore also ask whether the cancellations we see in the conditional correlation coefficients in the previous section are related to major mergers.

We address both of these issues in this section, showing that our results are unchanged when excluding splashback haloes or segregating haloes by the epoch of their last major merger.

4.1 Splashback objects

We identify splashback haloes using the output of CONSISTENT-TREES which provides the redshift z_{firstacc} of the 'first accretion' event of each object. This is the epoch at which the main progenitor of the object first passed inside the virial radius of a larger object. Splashback haloes are then objects which are currently not identified as subhaloes (i.e. not inside the virial radius of a larger object; 'PID'$=-1$ according to CONSISTENT-TREES) but have $z_{\mathrm{firstacc}} > 0$. With a fine time resolution in our merger tree which uses 201 snapshots, we expect this criterion to capture most of these objects.

We have repeated the analysis of Section 3 for halo samples excluding splashback haloes and also for the splashback haloes themselves. Since we only have merger histories available for haloes in our high-resolution boxes, we focus on the low-mass range for this analysis. Fig. 5 shows the results. We see in the *top row* that excluding splashback haloes has essentially no impact on our main results, since the correlation coefficients in the left and middle panels, as well as the level of cancellation in the right-hand panel, are nearly identical to the low-mass results of Fig. 2. The black curve in the top right panel shows the fraction of haloes that were excluded as being splashback objects; this is always $\lesssim$2 per cent over this mass range and decreases as expected towards higher masses.

Interestingly, when we repeat the analysis for these splashback objects themselves (*bottom row* of Fig. 5), we see very different behaviour. First, the correlation between α and b_1 at the lowest masses is now weaker in magnitude than other correlations, in stark contrast to the case for distinct haloes. And the right-hand panel shows that, in fact, α has essentially no impact on the assembly bias correlations involving any internal property. (We do not display the results for the two highest mass bins which contain fewer than 20 objects each.) In other words, the cosmic web anisotropy is a very poor indicator of any assembly bias trend for splashback objects. This is physically perhaps not surprising considering the very different accretion and tidal stripping histories of these objects as compared to other genuinely distinct haloes. We discuss this further in Section 5.

4.2 Recent major mergers

The output of CONSISTENT-TREES provides, for each object, the epoch of the last major merger event this object experienced on its main progenitor branch. The definition of a major merger is an event involving the overlap of virial radii of objects with a mass ratio closer to unity than 1:3. We use objects from the higher resolution box as in the earlier analysis and *discard splashback objects* as defined by the criterion of Section 4.1. We segregate the remaining objects by their redshift z_{lmm} of last major merger into two populations: those with recent major mergers which occurred at $z_{\mathrm{lmm}} < 0.4$ (corresponding to <4.3 Gyr of lookback time) and those with major mergers further back in the past. We then compute the same rank correlation coefficients as before and repeat the analysis similar to that shown in Fig. 5.

Fig. 6 shows the results. Our segregation makes the recent major merger population have fewer objects, thus the results have larger

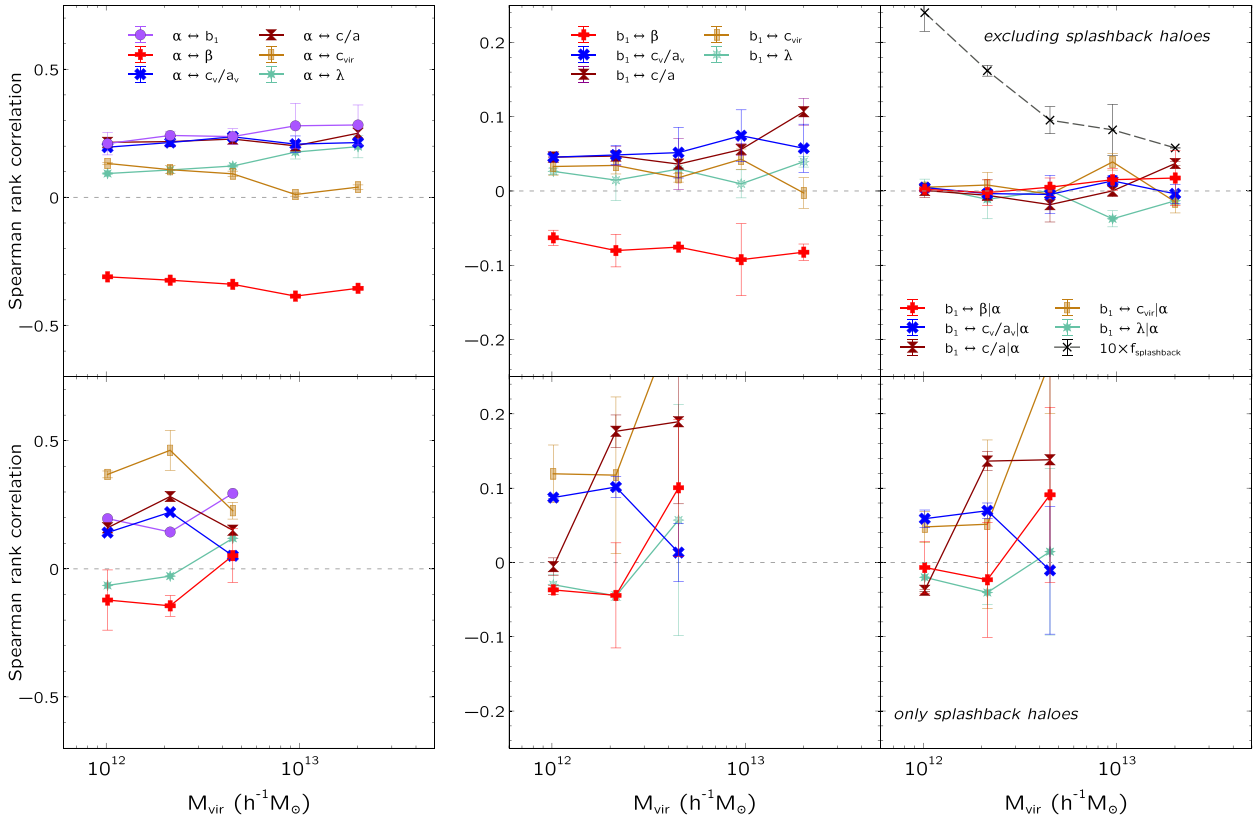

Figure 5. Same as Fig. 2, showing results of the analysis performed separately for splashback and other haloes, restricted to the higher resolution boxes and hence lower masses. Splashback objects were identified as described in Section 4.1. (Top row): Results *excluding splashback objects*; these are essentially identical to those in Fig. 2, with α being a good indicator of all assembly bias trends. The black dashed curve in the top right panel shows $10\times$ the fraction of haloes classified as splashback objects. (Bottom row): Results for *splashback objects only*; in this case, α is a very poor indicator of *any* assembly bias trend. Results are not displayed for the two highest mass bins which contain fewer than 20 objects each. See the text for a discussion.

scatter. Despite this, we see that separating out the population with recent major mergers does not bring out any dramatic difference in our main results, suggesting that both the populations of haloes are influenced similarly by their respective tidal environment as regards their assembly bias trends. We conclude that major merger events are not a likely cause for α being an excellent statistical intermediary in explaining halo assembly bias.

5 DISCUSSION AND CONCLUSIONS

The hierarchical formation of cosmological structure leads to distinct connections between the properties of the cosmic web and its constituent dark matter haloes across a wide range of length-scales. The most striking amongst these are the ones categorized as assembly bias (or secondary bias), in which the large-scale ($\gtrsim$few $\times\, 10\,h^{-1}$ Mpc) clustering strength of haloes shows distinct trends with a number of internal halo properties (defined at scales $\lesssim R_{\rm vir} \sim$ few $\times\, 100\,h^{-1}$ kpc), even at fixed halo mass. Understanding the origin of such correlations across several orders of magnitude in length-scale is of great interest from the point of view of building a complete understanding of structure formation in the ΛCDM framework, and can have consequences for galaxy evolution and precision cosmology.

In this work, we have explored the idea that many (if not all) assembly bias trends in the mass range $8\times10^{11}\,h^{-1}\,{\rm M}_\odot \lesssim M_{\rm vir} \lesssim 5\times10^{14}\,h^{-1}\,{\rm M}_\odot$ could be largely a result of a multiscale connection between internal halo properties and the large-scale environment, with the *local, non-linear cosmic web* environment acting as an intermediary. This is motivated by the expectation that these correlations must be connected to the only physical mechanism at play (gravitational tides) at the most natural intermediate length-scale in the problem (the current turn-around radius for infalling material around a given halo, which is close to approximately four times the halo radius).

We considered *scalar internal properties* related to the shape, velocity dispersion, density profile, and angular momentum of haloes; these include the halo shape asphericity c/a (Section 2.3.1), velocity ellipsoid asphericity c_v/a_v (Section 2.3.2), velocity anisotropy β (Section 2.3.3), concentration $c_{\rm vir}$ (Section 2.3.4), and spin λ (Section 2.3.5). The *large-scale environment* of each halo was characterized using the halo-by-halo bias b_1 of Paranjape et al. (2018a) defined at $\gtrsim 30\,h^{-1}$ Mpc scales (Section 2.1) and, for the *local cosmic web environment*, we considered the halo-centric tidal tensor defined at $\sim 4R_{\rm 200b}$ scales (Section 2.2), focusing on the *tidal anisotropy* variable α (equation 4) introduced by Paranjape et al. (2018a).

Our primary statistical analysis relied on Spearman rank correlation coefficients calculated for pairs of variables. In particular, we argued that the *vanishing of conditional correlation coefficients* defined in equation (12) offers a useful and compact way to assess the strength of multivariate statistical connections (Section 3.2), and we further demonstrated that this technique is robust to all of the assumptions involved in using equation (12) (Section 3.4 and Appendices A4 and A6).

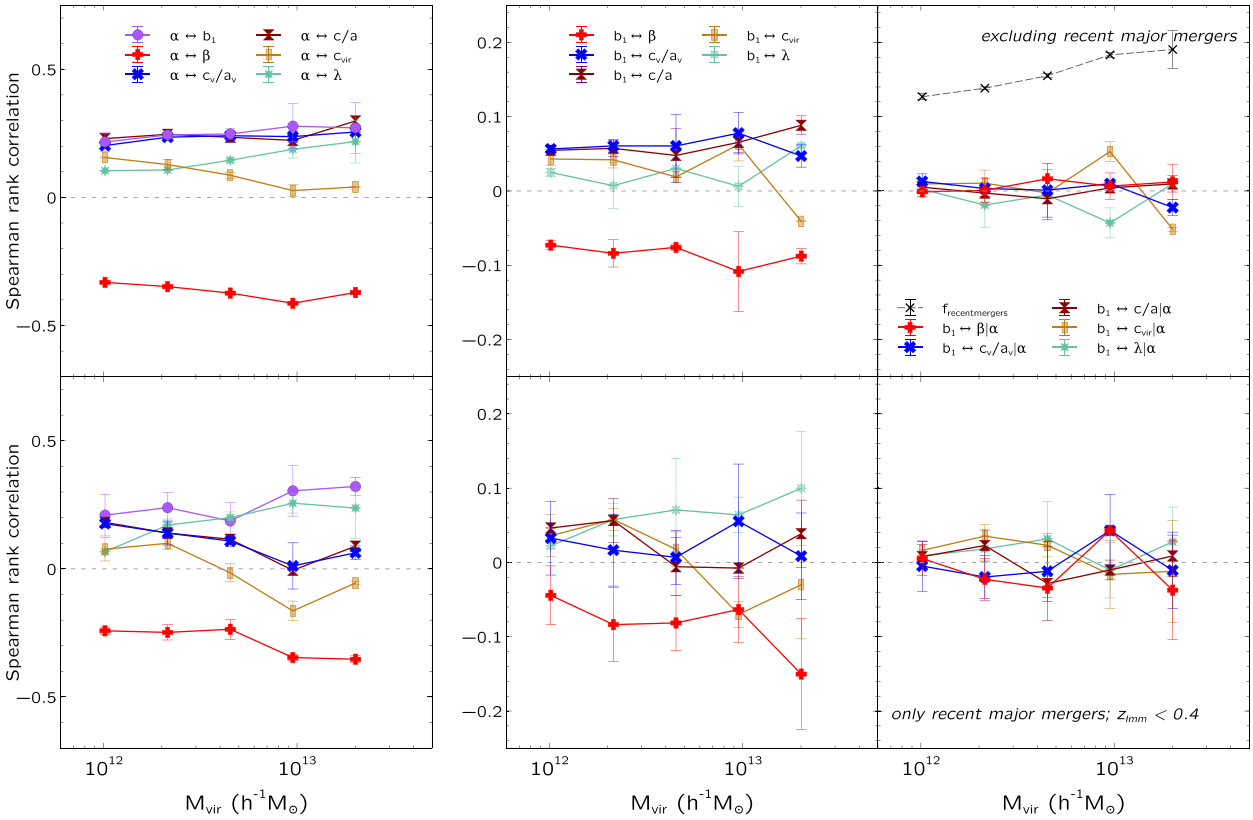

Figure 6. Same as Fig. 5, now splitting haloes into those with recent major merger events and other haloes. The haloes with recent major mergers were required to have their last major merger at a redshift $z_{\rm lmm} < 0.4$ (see Section 4.2 for details). (Top row): Results *excluding haloes with recent major mergers*; these are essentially identical to those in Fig. 2, with α being a good indicator of all assembly bias trends. The black dashed curve in the top right panel shows the fraction of haloes classified as recent major mergers. (Bottom row): Results for *only haloes with recent major mergers*; again, α is still a good indicator of all assembly bias trends. We conclude that α influences assembly bias similarly for both these populations. See the text for a discussion.

Our main results can be summarized as follows:

(i) The tidal anisotropy α shows the strongest correlation by far with b_1 at fixed halo mass amongst *all* halo properties we have considered (Fig. 1 and middle panel of Fig. 2) and correlates strongly with all internal halo properties as well (left-hand panel of Fig. 2). The correlation between α and b_1 in particular is substantially stronger than that between b_1 and pure density or pure anisotropy variables, as discussed in Section 2.2. The variable α is therefore an excellent candidate for an intermediary in explaining assembly bias, more so than the density contrast δ (equation 2) defined at the same scale (Appendices A2 and A3).

(ii) The conditional correlation coefficients $\gamma_{b_1 c|\alpha}$ are substantially smaller in magnitude than the unconditional coefficients $\gamma_{b_1 c}$ for *all* internal halo properties c that we studied, for all but the highest mass scales we consider (right-hand panel of Fig. 2, see also Table 1 and Fig. 3). The joint distribution of α, b_1, and any internal property $c \in \{\beta, c_v/a_v, c/a, c_{\rm vir}, \lambda\}$ is therefore consistent with reflecting only two fundamental correlations $b_1 \leftrightarrow \alpha$ and $c \leftrightarrow \alpha$:

$$p(\alpha, b_1, c) \simeq p(\alpha) p(b_1|\alpha) p(c|\alpha), \tag{13}$$

(Section 3.2, see also Fig. 4 and Appendix A6). Thus, α indeed explains *all* large-scale assembly bias trends, particularly at low halo mass. α defined at approximately four times the halo radius also outperforms the environmental overdensity δ defined at fixed smoothing scales 1–2 h^{-1} Mpc, recently proposed by Han et al. (2019) as an assembly bias indicator (see Appendix A2).

(iii) Our conclusions regarding the role of α are *unchanged upon excluding splashback haloes* from the analysis (Section 4.1, top row of Fig. 5). Interestingly, repeating the analysis for the small population of splashback objects themselves (these are $\lesssim$2 per cent of distinct haloes in our mass range) showed that α is a poor indicator of *any* assembly bias trend for these objects (bottom row of Fig. 5, see also below).

(iv) Our conclusions regarding α are also unchanged when segregating haloes by the presence or absence of a recent major merger event (Section 4.2, Fig. 6).

This wide-ranging effect of α in connecting small and large scales provides a new perspective on the phenomenon of assembly bias of low-mass haloes. There are several indications in the literature that multiple aspects of a halo's tidal environment could play a role in establishing the assembly bias trends of different variables. E.g. being in a non-linear filament affects the mass accretion rate and formation time of an object (due to strong tides, Hahn et al. 2009; Musso et al. 2018) and changes its shape, profile, and velocity dispersion structure (due to strong external flows, Borzyszkowski et al. 2017; Mansfield & Kravtsov 2019). Consistently with this picture, tidal influences on substructure also start well before accretion on to the parent object (Behroozi et al. 2014). Similarly, the presence/absence of neighbours having larger (Hahn et al. 2009; Hearin, Behroozi & van den Bosch 2016; Salcedo et al. 2018) or comparable mass (Johnson et al. 2019), and their corresponding tidal influence, has also been shown to be connected with assembly

bias. (See also Mo et al. 2005; Buehlmann & Hahn 2019, for the related effect of tidal heating due to the formation of cosmic sheets.)

The fact that α simultaneously explains multiple assembly bias trends over a wide range of halo mass suggests that, ultimately, the quantity relevant for assembly bias is the degree of anisotropy of the *current* tidal environment of distinct haloes, evaluated at the current turn-around scale (approximately four times the halo radius). Having fixed this, the specific physical mechanism that affects any particular variable becomes less relevant; we expect it to only play a role in establishing how strongly that variable correlates with the tidal anisotropy.

This has consequences of practical interest, particularly because α is defined at intermediate length-scales. On the one hand, the importance of α as an assembly bias indicator might be exploited to populate low-resolution simulations with otherwise unresolved haloes having the correct assembly bias trends. This would be of immense interest for precision cosmological analyses that would otherwise require high dynamic range as well as tight control on assembly bias related systematics (see e.g. Zentner et al. 2014). On the other hand, α can also be useful in high-resolution, small volume simulations of galaxy formation, where it might be used to predict (albeit with large scatter) the large-scale environment of realistic galaxies. For example, understanding the strength and origin of correlations between α and variables such as stellar mass, star formation rate, metallicity, etc., might help in understanding the expected strength of *galaxy* assembly bias, which has been difficult to detect robustly in observational samples (Lin et al. 2016; Tinker et al. 2017).

To try and understand *why* the variable α, specifically, is such a good assembly bias indicator for distinct haloes, it is worth considering its behaviour for *splashback* haloes. As we showed, *α does not perform well* in explaining the assembly bias of these objects. This is likely a manifestation of the fact that the internal properties of splashback objects, like other substructure, have been dramatically affected by the strong tidal influence of their *host* halo. Since this also includes substantial mass-loss due to tidal stripping and a consequent decrease in radius, it is perhaps not surprising that the tidal environment evaluated at the scale approximately four times the current radius, at the current location, is not a good indicator of the large-scale environment of the splashback object.

It appears, then, that α is a good indicator of assembly bias for objects whose current tidal environment is the most extreme they have ever experienced, and fails for objects whose current environment does not reflect the largest tidal influences that have acted on them. This points towards a novel approach in thinking about substructure in general, in which haloes might be classified by their *tidal history*. Objects that have always been in tidally mild, isotropic environments (small α) would then be distinguished from objects that have spent a considerable fraction of their existence in anisotropic sheets or filaments (large α). Subhaloes and splashback objects would then simply be the extremes of the latter category, objects that have experienced very high tidal forces at some point in their past (not necessarily reflected by their current environment). Of course, for this picture to be consistent, it must also be possible to construct a local tidal indicator of large-scale assembly bias trends for subhaloes and splashback objects, perhaps α defined using the scale of the *host* halo.

We also believe these ideas could be a useful starting point for a dynamical model of the influence of local non-linear tides on internal halo properties, building on, e.g. known results from tidal torque theory for the connection between large-scale tides and halo angular momenta and shapes (see e.g. Catelan & Theuns 1996) and accounting for known correlations between internal halo properties (see e.g. Jeeson-Daniel et al. 2011; Skibba & Macciò 2011). Finally, it would be interesting to extend our analysis to include tensor assembly bias signatures involving alignments between the mass/velocity ellipsoid tensors, angular momenta, and halo-centric tidal tensors. We will return to all these issues in future work.

ACKNOWLEDGEMENTS

We thank R. Srianand, K. Subramanian, T. Padmanabhan, and S. More for useful discussions, and the anonymous referee for a detailed and helpful report. 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. OH acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 679145, project 'COSMO-SIMS'). We gratefully acknowledge the use of high performance computing facilities at IUCAA, Pune.

REFERENCES

Alam S., Zu Y., Peacock J. A., Mandelbaum R., 2019, MNRAS, 483, 4501
Allgood B., Flores R. A., Primack J. R., Kravtsov A. V., Wechsler R. H., Faltenbacher A., Bullock J. S., 2006, MNRAS, 367, 1781
Angulo R. E., Baugh C. M., Lacey C. G., 2008, MNRAS, 387, 921
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
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
Benson A. J., 2017, MNRAS, 471, 2871
Bernardi M. et al., 2003, AJ, 125, 1866
Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215
Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton Univ. Press, Princeton, NJ
Bond J. R., Myers S. T., 1996, ApJS, 103, 1
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
Borzyszkowski M., Porciani C., Romano-Díaz E., Garaldi E., 2017, MNRAS, 469, 594
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Buehlmann M., Hahn O., 2019, MNRAS, 487, 228
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001a, MNRAS, 321, 559
Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001b, ApJ, 555, 240
Castorina E., Sheth R. K., 2013, MNRAS, 433, 1529
Castorina E., Paranjape A., Hahn O., Sheth R. K., 2016, preprint (arXiv: 1611.03619)
Catelan P., Theuns T., 1996, MNRAS, 282, 436
Dalal N., White M., Bond J. R., Shirokov A., 2008, ApJ, 687, 12
Desjacques V., 2008, MNRAS, 388, 638
Diemer B., Kravtsov A. V., 2015, ApJ, 799, 108
Eisenstein D. J., Loeb A., 1995, ApJ, 439, 520
Fakhouri O., Ma C.-P., 2010, MNRAS, 401, 2245
Faltenbacher A., White S. D. M., 2010, ApJ, 708, 469
Franx M., Illingworth G., de Zeeuw T., 1991, ApJ, 383, 112
Gao L., White S. D. M., 2007, MNRAS, 377, L5
Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
Gill S. P. D., Knebe A., Gibson B. K., 2005, MNRAS, 356, 1327
Haas M. R., Schaye J., Jeeson-Daniel A., 2012, MNRAS, 419, 2133
Hahn O., Abel T., 2011, MNRAS, 415, 2101
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., Behroozi P. S., van den Bosch F. C., 2016, MNRAS, 461, 2135
Heavens A., Peacock J., 1988, MNRAS, 232, 339
Jeeson-Daniel A., Dalla Vecchia C., Haas M. R., Schaye J., 2011, MNRAS, 415, L69
Jing Y. P., Suto Y., Mo H. J., 2007, ApJ, 657, 664
Johnson J. W., Maller A. H., Berlind A. A., Sinha M., Holley-Bockelmann J. K., 2019, MNRAS, 486, 1156
Kaiser N., 1984, ApJ, 284, L9
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lazeyras T., Musso M., Schmidt F., 2017, J. Cosmol. Astropart. Phys., 3, 059
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Lin Y.-T., Mandelbaum R., Huang Y.-H., Huang H.-J., Dalal N., Diemer B., Jian H.-Y., Kravtsov A., 2016, ApJ, 819, 119
Ludlow A. D., Navarro J. F., Springel V., Vogelsberger M., Wang J., White S. D. M., Jenkins A., Frenk C. S., 2010, MNRAS, 406, 137
Ludlow A. D. et al., 2013, MNRAS, 432, 1103
Ludlow A. D., Navarro J. F., Angulo R. E., Boylan-Kolchin M., Springel V., Frenk C., White S. D. M., 2014, MNRAS, 441, 378
Mansfield P., Kravtsov A. V., 2019, preprint (arXiv:1902.00030)
Mao Y.-Y., Zentner A. R., Wechsler R. H., 2018, MNRAS, 474, 5143
McEwen J. E., Weinberg D. H., 2018, MNRAS, 477, 4348
Monaco P., 1999, in Giuricin G., Mezzetti M., Salucci P., eds, ASP Conf. Ser., Vol. 176, Observational Cosmology: The Development of Galaxy Systems. Astron. Soc. Pac., San Francisco, p. 186
Mo H. J., Yang X., van den Bosch F. C., Katz N., 2005, MNRAS, 363, 1155
Musso M., Sheth R. K., 2012, MNRAS, 423, L102
Musso M., Cadiou C., Pichon C., Codis S., Kraljic K., Dubois Y., 2018, MNRAS, 476, 4877
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
Oñorbe J., Garrison-Kimmel S., Maller A. H., Bullock J. S., Rocha M., Hahn O., 2014, MNRAS, 437, 1894
Paranjape A., Padmanabhan N., 2017, MNRAS, 468, 2984
Paranjape A., Hahn O., Sheth R. K., 2018a, MNRAS, 476, 3631
Paranjape A., Hahn O., Sheth R. K., 2018b, MNRAS, 476, 5442
Peebles P. J. E., 1969, ApJ, 155, 393
Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS, 462, 893
Salcedo A. N., Maller A. H., Berlind A. A., Sinha M., McBride C. K., Behroozi P. S., Wechsler R. H., Weinberg D. H., 2018, MNRAS, 475, 4411
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Sheth R. K., Tormen G., 2004, MNRAS, 350, 1385
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Shi J., Sheth R. K., 2018, MNRAS, 473, 2486
Shi J., Wang H., Mo H. J., 2015, ApJ, 807, 37
Skibba R. A., Macciò A. V., 2011, MNRAS, 416, 2388
Springel V., 2005, MNRAS, 364, 1105
Sunayama T., Hearin A. P., Padmanabhan N., Leauthaud A., 2016, MNRAS, 458, 1510
Taylor J. E., Navarro J. F., 2001, ApJ, 563, 483
Tinker J. L., Wetzel A. R., Conroy C., Mao Y.-Y., 2017, MNRAS, 472, 2504
van Daalen M. P., Angulo R. E., White S. D. M., 2012, MNRAS, 424, 2954
Villarreal A. S. et al., 2017, MNRAS, 472, 1088
Wang H., Mo H. J., Jing Y. P., Yang X., Wang Y., 2011, MNRAS, 413, 1973
Wang H. et al., 2018, ApJ, 852, 31
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
White S. D. M., Silk J., 1979, ApJ, 231, 1
Xu X., Zheng Z., 2018, MNRAS, 479, 1579
Yan H., Fan Z., White S. D. M., 2013, MNRAS, 430, 3432
Zehavi I., Contreras S., Padilla N., Smith N. J., Baugh C. M., Norberg P., 2018, ApJ, 853, 84
Zentner A. R., 2007, Int. J. Mod. Phys. D, 16, 763
Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044

APPENDIX A: CONVERGENCE AND ADDITIONAL TESTS

In this Appendix, we first present a convergence study for our calculation of tidal variables which justifies our choices for the minimum halo mass threshold in our simulations. We then show that, although the variables α and δ defined in the main text are correlated, the tidal anisotropy α is likely to be a better indicator than the isotropic overdensity δ of all assembly bias, an expectation which is then confirmed in the main text. We also display the one-dimensional probability distributions of all the halo-related variables used in this work, in a few narrow mass ranges. Finally we complete our analysis in Section 3.4 by showing explicitly the structure of distribution of b_1, α, and c for all $c \in \{\beta, c_v/a_v, c/a, c_{\rm vir}, \lambda\}$.

A1 Convergence study

Fig. A1 shows the Spearman rank correlation coefficient $\gamma_{\alpha\delta}$ between α and δ as a function of halo mass. These variables were evaluated as described in Section 2.2 using various grid sizes as indicated in the legend. Results are shown for the low-resolution (markers with solid lines) and high-resolution configuration (markers with dashed lines).

We see that convergence in any given configuration of the simulations is starting to be achieved at grid sizes of $\geq 512^3$ cells.

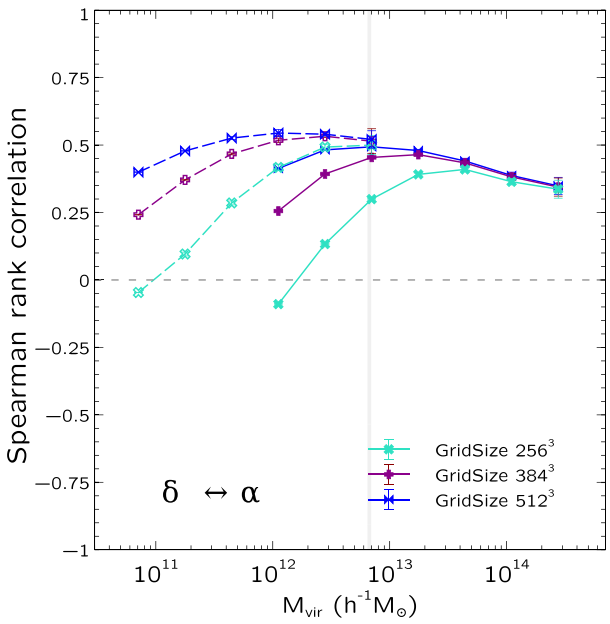

Figure A1. Convergence study of the $\alpha \leftrightarrow \delta$ correlation. The symbols joined by lines of different colours indicate measurements using α and δ (Section 2.2) computed on cubic grids of different sizes as indicated. The formatting of symbols (filled versus empty) and lines (solid versus dashed) is identical to that in Fig. 1. Based on the behaviour of the curves in the overlap region between the higher and lower resolution boxes, we conclude that a 512^3 grid is sufficient for our purposes, provided we restrict attention to haloes with $\geq$3200 particles (shown as the vertical line). See the text for further details and a discussion of the consequences of a positive correlation between α and δ.

Based on the trends seen in Fig. A1, we also choose a minimum halo mass threshold of 3200 particles as a compromise between minimizing the mismatch in $\gamma_{\alpha\delta}$ between the two configurations and retaining enough statistics in the highest mass bin analysed in the high-resolution simulation. A lower mass threshold would increase the mismatch, while a higher threshold such as 4000 particles would minimize the mismatch but make all measurements at the high-mass end of the high-resolution box too noisy to be reliable.

Since the correlation coefficient between α and δ is quite large across all masses, one would worry that any statements about statistical connections between α and other variables such as halo bias or internal halo properties could simply be reflecting a correlation between δ and these properties. Below we demonstrate that this is not the case for any of the correlations we are interested in.

A2 Tidal environment and large-scale bias

Fig. A2 explores the correlations between the environment variables α and δ defined at $\sim 4R_{200b}$ scales and the large-scale environment as measured by halo bias b_1. The dashed curves show the unconditional correlation coefficients $\gamma_{b_1\alpha}$ (red) and $\gamma_{b_1\delta}$ (blue). As already discussed by Paranjape et al. (2018a), these show that $\gamma_{b_1\alpha} > \gamma_{b_1\delta}$, so that α is better correlated with b_1 than is δ at any halo mass. Indeed, Paranjape et al. (2018a) motivated the choice of $4R_{200b}$ as being the largest scale (adapted to the halo size) where this is true across all halo masses (see their fig. 5 and also the discussion below).

The solid curves show the conditional correlation coefficients $\gamma_{b_1\alpha|\delta}$ (red) and $\gamma_{b_1\delta|\alpha}$ (blue). We see that $\gamma_{b_1\delta|\alpha} < \gamma_{b_1\delta}$ by a factor approximately two to three for all halo masses. The conditional coefficient $\gamma_{b_1\alpha|\delta}$, on the other hand, shows a smaller decrement compared to the corresponding unconditional coefficient $\gamma_{b_1\alpha}$. In fact, we curiously also see $\gamma_{b_1\alpha|\delta} \simeq \gamma_{b_1\delta}$ across all masses, so that

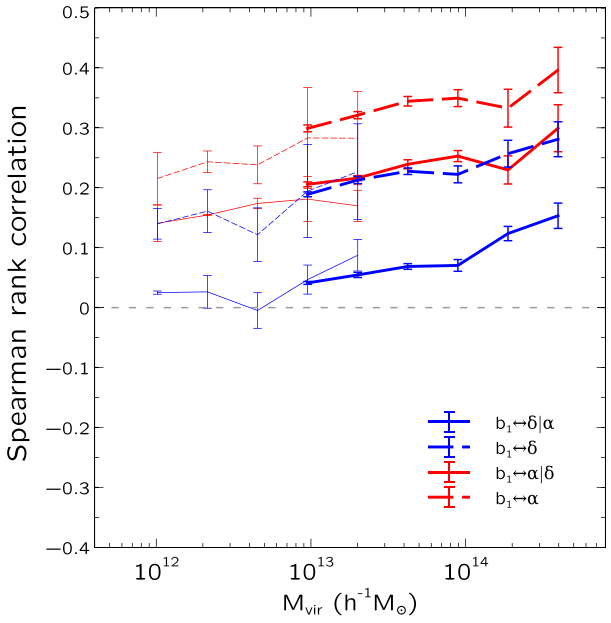

Figure A2. Correlation between tidal environment at $4R_{200b}$ scales (as measured by α and δ) and large-scale environment (measured by halo bias b_1). Curves show the unconditional correlation coefficients $\gamma_{b_1\alpha}$ and $\gamma_{b_1\delta}$ (dashed), as well as the conditional coefficients $\gamma_{b_1\delta|\alpha}$ and $\gamma_{b_1\alpha|\delta}$ (solid). The results indicate that α is a better indicator of large-scale environment than is δ, both in the unconditional and conditional sense.

conditioning on δ does not even decrease the correlation between α and b_1 below the *unconditional* correlation between δ and b_1.

These results indicate that α is a better indicator of large-scale environment than is δ, both in the unconditional and conditional sense.

We have also repeated the analysis of Fig. 2 using α defined at fixed scales of 1 and $2\,h^{-1}$ Mpc, finding that the cancellations leading to small conditional correlation coefficients only occur in the mass range where $4R_{200b} \simeq 1\,h^{-1}$ Mpc, $2\,h^{-1}$ Mpc, respectively. Fig. A3 shows the results. For ease of comparison, we use bins of M_{200b} rather than M_{vir} for this figure. Finally, we have repeated this last analysis using δ instead of α, defined at $4R_{200b}$, $1\,h^{-1}$ Mpc, $2\,h^{-1}$ Mpc. We found that none of these variables perform as well as $\alpha(4R_{200b})$ in producing small conditional correlation coefficients across the entire halo mass range we probe. For brevity, we do not display these results. In a recent study, Han et al. (2019) proposed that δ defined at 1–2 h^{-1} Mpc is a strong candidate for explaining assembly bias trends. Our results indicate that $\alpha(4R_{200b})$ is an even stronger candidate, which can also be understood by the fact that the $\alpha \leftrightarrow b_1$ correlation seen in the left-hand panel Fig. 2 takes values at least comparable to, and usually larger than, the correlation strength between b_1 and the fixed-scale δ's (not shown).

A3 Tidal environment and internal halo properties

Fig. A4 explores the correlations between internal halo properties and the environmental variables α and δ, colour-coded by the internal properties as in previous figures. The solid (dashed) curves show the conditional correlation coefficients $\gamma_{c\delta|\alpha}$ ($\gamma_{c\alpha|\delta}$) for $c \in \{\beta, \lambda\}$. We have chosen these two internal variables as representing the extremes of the trends we discuss here; the other internal variables $\{c_v/a_v, c/a, c_{vir}\}$ show qualitatively identical trends with intermediate strengths. In each case, we find $\gamma_{c\delta|\alpha}$ is substantially smaller in magnitude than $\gamma_{c\alpha|\delta}$ at all but the smallest halo masses we explore, indicating that α accounts for a substantial fraction of the correlation of δ with *all* internal properties. Especially in the case of λ, we see that α accounts for nearly *all* of the correlation between λ and δ.

Taken together, the results shown in Figs A2 and A4 show that α is a much better candidate than δ for an environmental link that could explain assembly bias in any internal halo property. In other words, the anisotropy of the halo tidal environment is expected to be more important than the local density in explaining assembly bias trends.

A4 Explicit conditional correlation

In the main text, we explore the connection between halo tidal environment and assembly bias using the Gaussian-motivated correlation coefficients $\gamma_{bc|a} \equiv \gamma_{bc} - \gamma_{ab}\gamma_{ac}$, where b and c represent halo bias and any internal halo property, respectively, and a represents the environmental variable. Here, we perform an explicit test of this connection by evaluating correlation coefficients in *fixed bins* of the environmental variable. Since binning naturally increases the noise in our measurements, we only display results for the strongest assembly bias trend which is that between b_1 and velocity anisotropy β.

Fig. A5 shows the correlation coefficients $\gamma_{b_1\beta}$ as a function of halo mass, evaluated for haloes in quintiles of δ (left-hand panel) and α (right-hand panel), with the all-halo coefficient repeated in each panel in red. It is visually apparent that fixing α leads to conditional

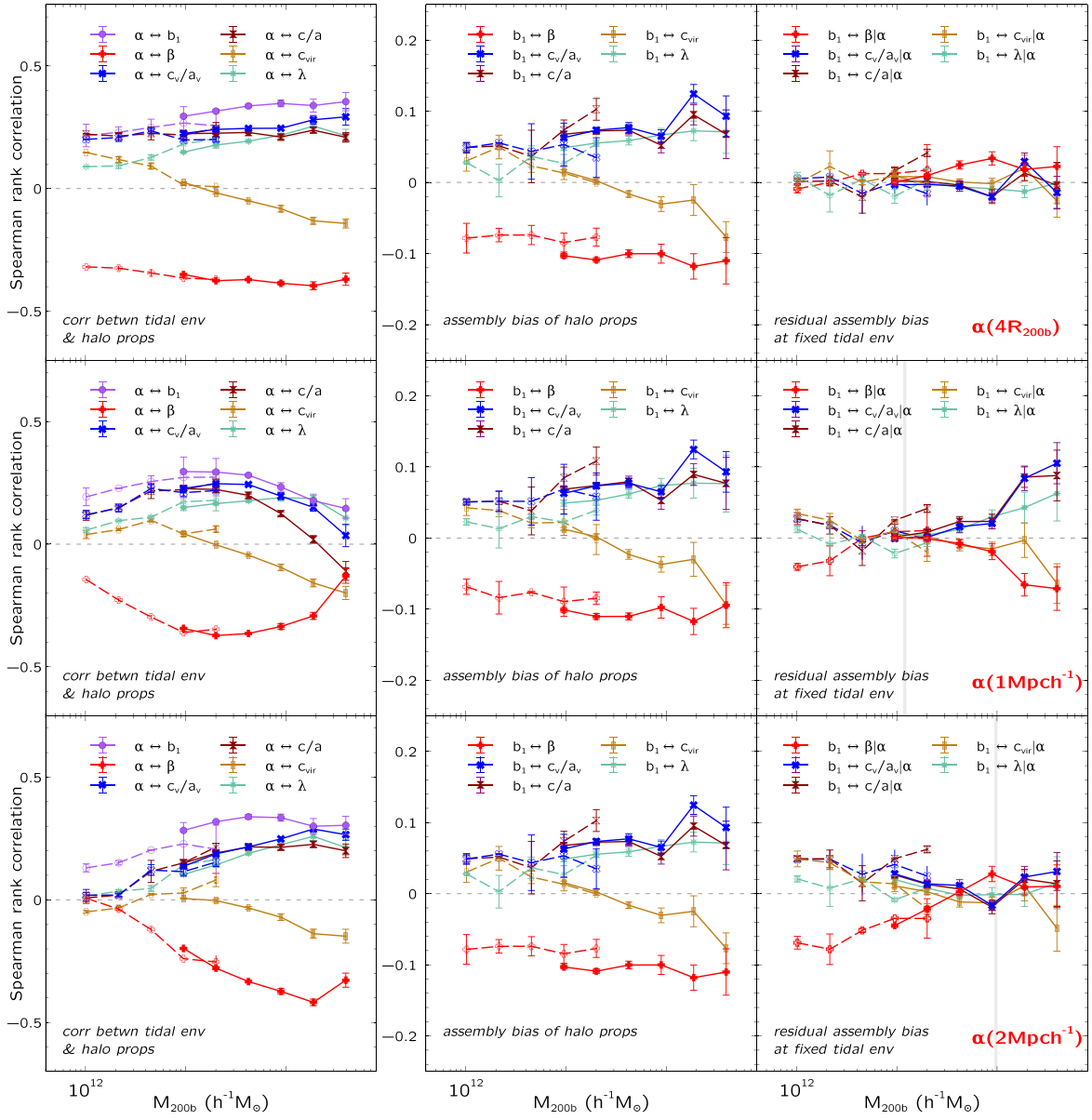

Figure A3. The top panel is identical to Fig. 2 except that we show results as a function of M_{200b} instead of M_{vir}, with tidal anisotropy α still defined at $4R_{200b}$ scales. The middle and lower panels repeat the same analysis for α defined at 1 and 2 h^{-1} Mpc, respectively. We see that the residual assembly bias (right-hand panels) is consistent with being zero only when α is defined at scale $4R_{200b}$. In fact, we see that, when α is defined at 1 or 2 h^{-1} Mpc scales, the residual assembly bias reduces only in the mass range where the haloes have $4R_{200b} \simeq 1$ or 2 h^{-1} Mpc, respectively. The respective mass ranges have been marked with grey vertical lines in the middle and bottom right-hand panels, respectively (this comparison is the reason to use M_{200b}).

correlations that are substantially closer to zero than when fixing δ. We have checked that qualitatively similar results hold for all other internal variables except c_{vir} for which the noise is too large to draw strong conclusions given our simulation set.

A5 Halo properties

Here, we show for reference the distributions of all variables studied in the main text, including halo bias b_1, environmental variables $\{\alpha, \delta\}$, and internal halo properties $\{\beta, c_v/a_v, c/a, c_{vir}, \lambda\}$. See Section 2 for a description of how each of these is measured.

The histograms in Fig. A6 show the individual distributions of all eight variables (different panels, as labelled) for a few narrow mass ranges (different line styles), with several known trends being apparent. We see that haloes are, on average, substantially aspherical in shape (panel c/a) but less so in their velocity ellipsoids (panel c_v/a_v), although there is a clear preference for radially dominated orbits (panel β). The distributions of spin λ and concentration c_{vir} show distinct tails at small values, while those of the environmental variables δ and α are skewed towards large values, and the distributions of b_1 are largely symmetric around the median. All variables except b_1 and λ show noticeable trends with halo mass.

A6 Joint distribution of α, b_1 and internal properties

In Section 3.4 we analysed the full distribution of b_1, α, β and showed that the overall anticorrelation between b_1 and β is consistent with being largely due to α. In this section we complete

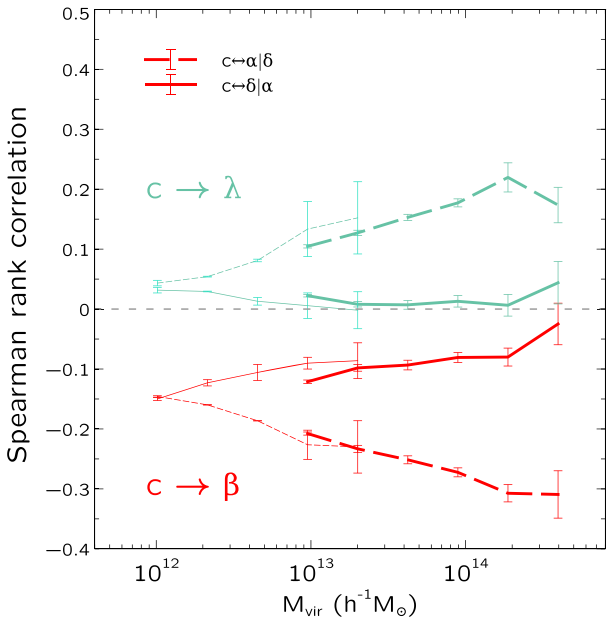

Figure A4. Correlation between tidal environment at $4R_{200b}$ scales (as measured by α and δ) and internal halo properties $c \in \{\beta, \lambda\}$. The curves show the conditional coefficients $\gamma_{c\delta|\alpha}$ (solid) and $\gamma_{c\alpha|\delta}$ (dashed). We see that $|\gamma_{c\delta|\alpha}| < |\gamma_{c\alpha|\delta}|$ at essentially all masses in both cases, indicating that α accounts for a substantial fraction of the correlation of δ with both of these internal properties. We find qualitatively similar results for the other internal properties $\{c_v/a_v, c/a, c_{\rm vir}\}$ (not shown).

this analysis by showing the same for the distribution of b_1, α, and c for all internal properties c in all three mass ranges.

To make the results compact, we will not show scatter plots for the various distributions and instead focus on the conditional mean $\langle\, b_1|\alpha, c\,\rangle$ (as already shown in Fig. 4 for the case $c \to \beta$) and additionally the square root of the conditional variance $\sigma(b_1|\alpha, c) \equiv \left(\left\langle\, b_1^2|\alpha, c\,\right\rangle - \langle\, b_1|\alpha, c\,\rangle^2\right)^{1/2}$, for all $c \in \{\beta, c_v/a_v, c/a, c_{\rm vir}, \lambda\}$. If the general relation (13) is true, then we should expect $\langle\, b_1|\alpha, c\,\rangle \simeq \langle\, b_1|\alpha\,\rangle$ and $\sigma(b_1|\alpha, c) \simeq \sigma(b_1|\alpha)$. Figs A7 and A8 show that this is indeed the case, as we discuss below.

Both figures comprise of subplots focusing on one internal variable c at a time. Fig. A7 (Fig. A8) shows projections of $\langle\, b_1|\alpha, c\,\rangle$ ($\sigma(b_1|\alpha, c)$) in the b_1–c (left subplot panels) and b_1–α planes (right subplot panels) for all three mass ranges (three sets of curves with offsets given for clarity). The bins along each horizontal axis are chosen as quintiles of the respective variable,[11] so that the left subplot panels additionally reveal the b_1–c assembly bias trends for the mean and width of the conditional b_1 distributions as systematic horizontal shifts of the different lines. Fig. A7 for $\langle\, b_1|\alpha, c\,\rangle$ shows results qualitatively identical to those seen in Fig. 4, with the b_1–c projections being approximately horizontal lines in fixed quartiles of α, and the b_1–α projections tracing out common loci in fixed quartiles of c. Fig. A8 extends these results to the (square root of) conditional variance $\sigma(b_1|\alpha, c)$, with the b_1–c projections again being approximately horizontal with α-dependent offsets, and the b_1–α projections tracing out common loci in fixed c-quartiles.

Furthermore, as we see in Fig. A6, the distribution of b_1 is approximately symmetric about its mean, indicating that a Gaussian shape is a reasonable approximation. This would mean that the conditional independence of b_1 and c at fixed α as seen in the conditional mean and variances in fact extends to the entire distribution as discussed above. These results strongly support our main conclusions regarding the role of α in explaining assembly bias.

[11] The marker location on the horizontal axis for each such quintile is chosen as the median of that quintile, averaged over realizations.

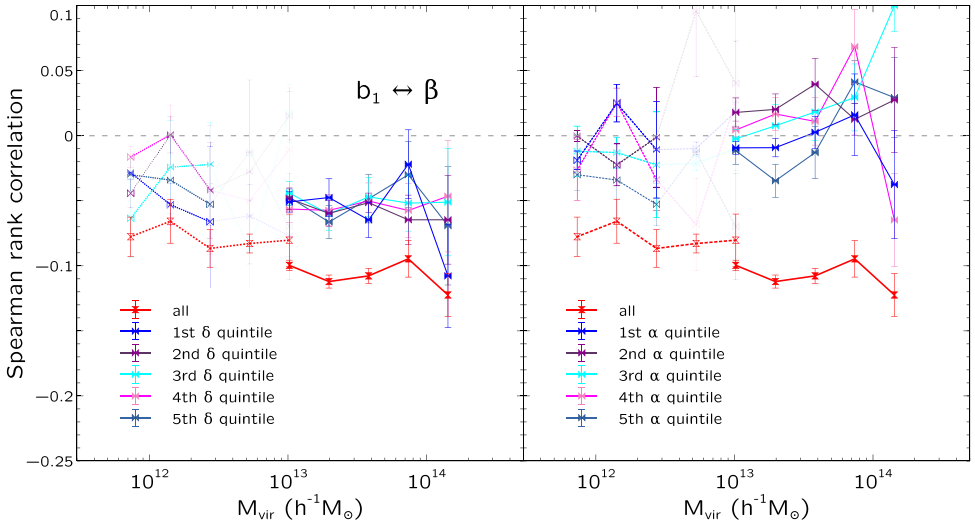

Figure A5. Explicit conditional correlation between halo bias b_1 and velocity anisotropy β as a function of halo mass for haloes in quintiles of δ (left-hand panel) and α (right-hand panel). The all-halo coefficient is shown with red symbols joined by red lines; this is the same in each panel and is repeated from the middle panel of Fig. 2. We see that fixing α leads to conditional correlations that are substantially closer to zero than when fixing δ.

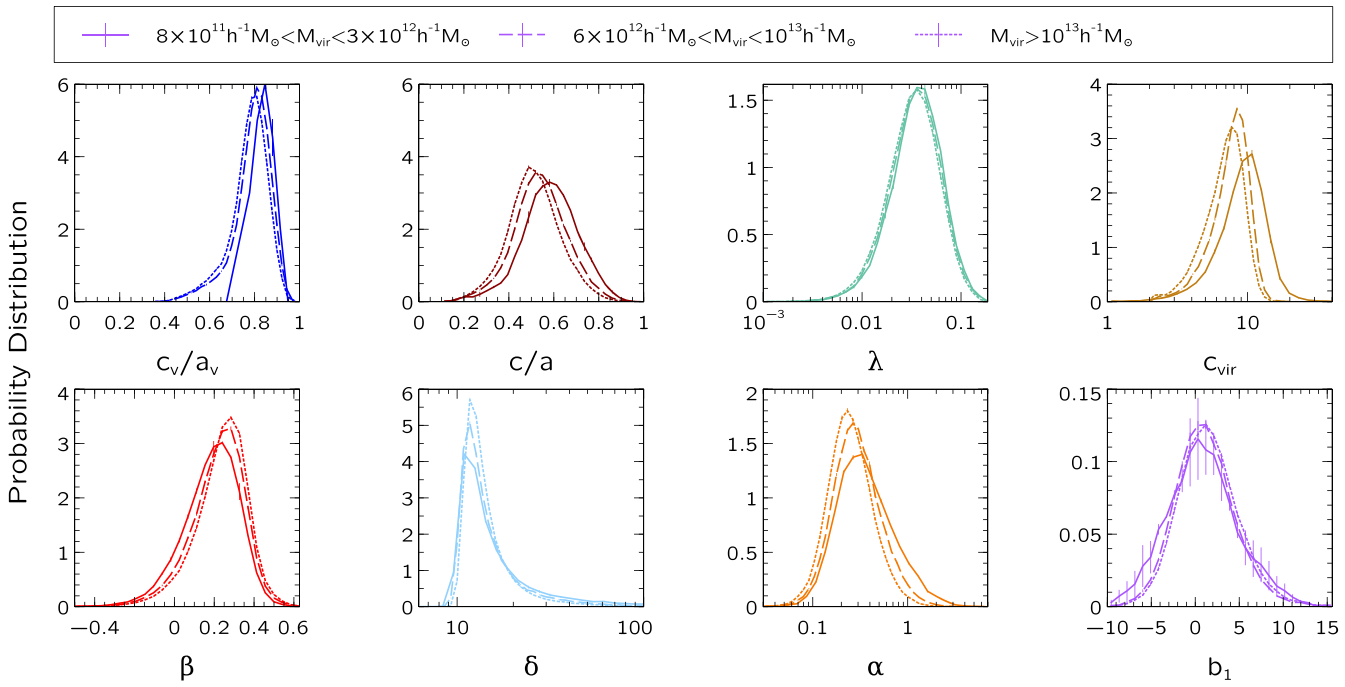

Figure A6. Normalized distributions of all variables used in this analysis (different panels, as indicated). In each panel, the solid, dashed, and dotted curves show measurements for low-, intermediate-, and high-mass haloes, respectively, as indicated in the legend at the top. The lowest mass bin uses measurements from two realizations of the high-resolution box, while the other two mass bins use measurements from 10 realizations of the low-resolution box, with the curves showing the mean and the error bars showing the scatter across realizations.

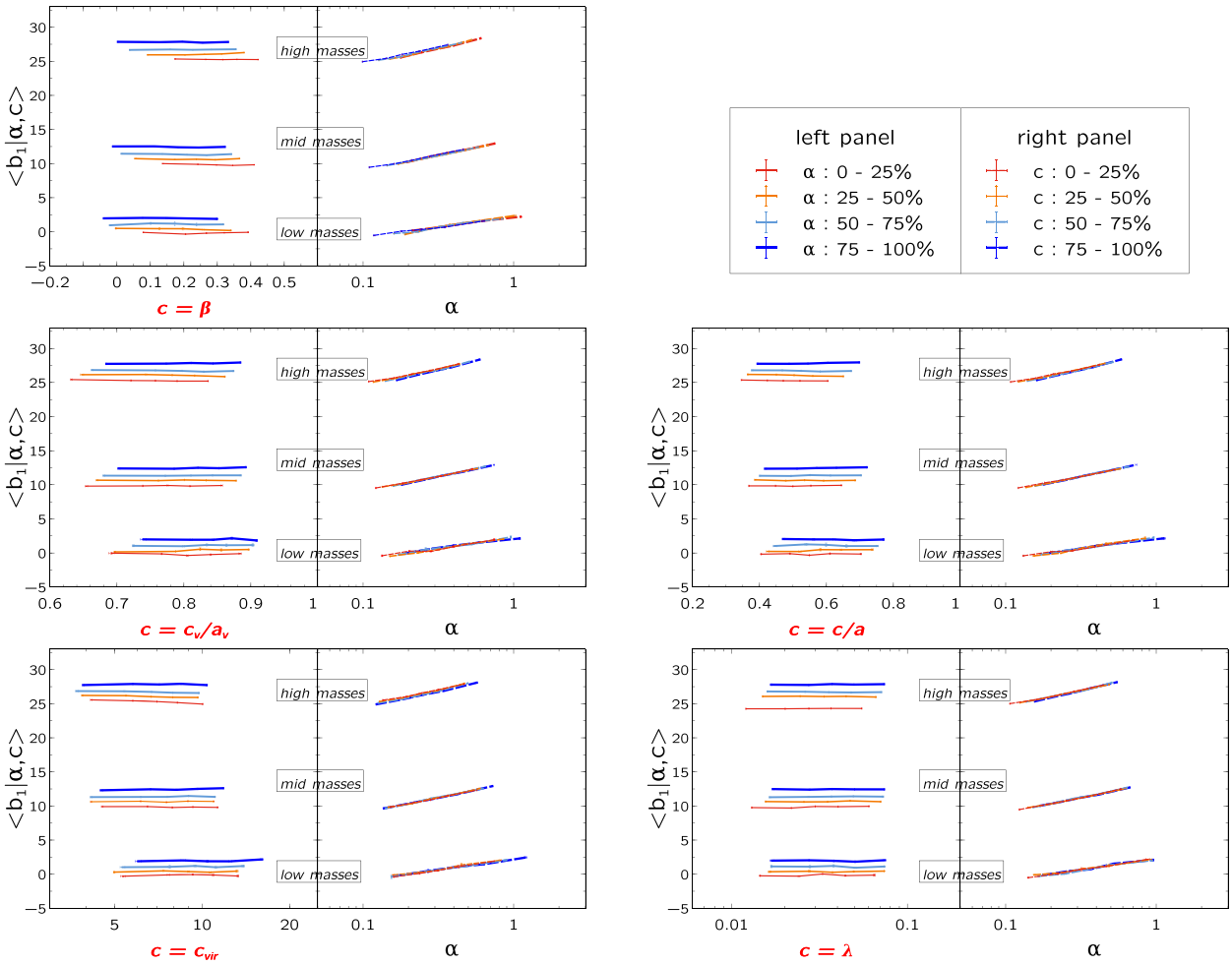

Figure A7. Conditional mean of b_1 at fixed α and $c \in \{\beta, c_v/a_v, c/a, c_{\rm vir}, \lambda\}$: Each subplot shows $\langle\, b_1|\alpha, c\,\rangle$ in two projections: the b_1–c plane for different quartiles of α (coloured lines in the left subplot panels) and the b_1–α plane for different quartiles of c (coloured lines in the right subplot panels). The mapping between line colour and quartiles of α or c is given in the legend at the top right. The results for $c = \beta$ are repeated from Fig. 4. In each panel, results are shown separately for haloes segregated into three mass ranges as in Fig. 4. The low-mass results are averaged over two realizations of the high-resolution box while the mid- and high-mass results are averaged over 10 realizations of the low-resolution box, with error bars in each case showing the error on the respective mean. For clarity, the mid- and high-mass results are also given vertical offsets of +10 and +25, respectively. We see that the results for all internal variables c and for each mass range are consistent with the relation $\langle\, b_1|\alpha, c\,\rangle \simeq \langle\, b_1|\alpha\,\rangle$. See the text for a discussion.

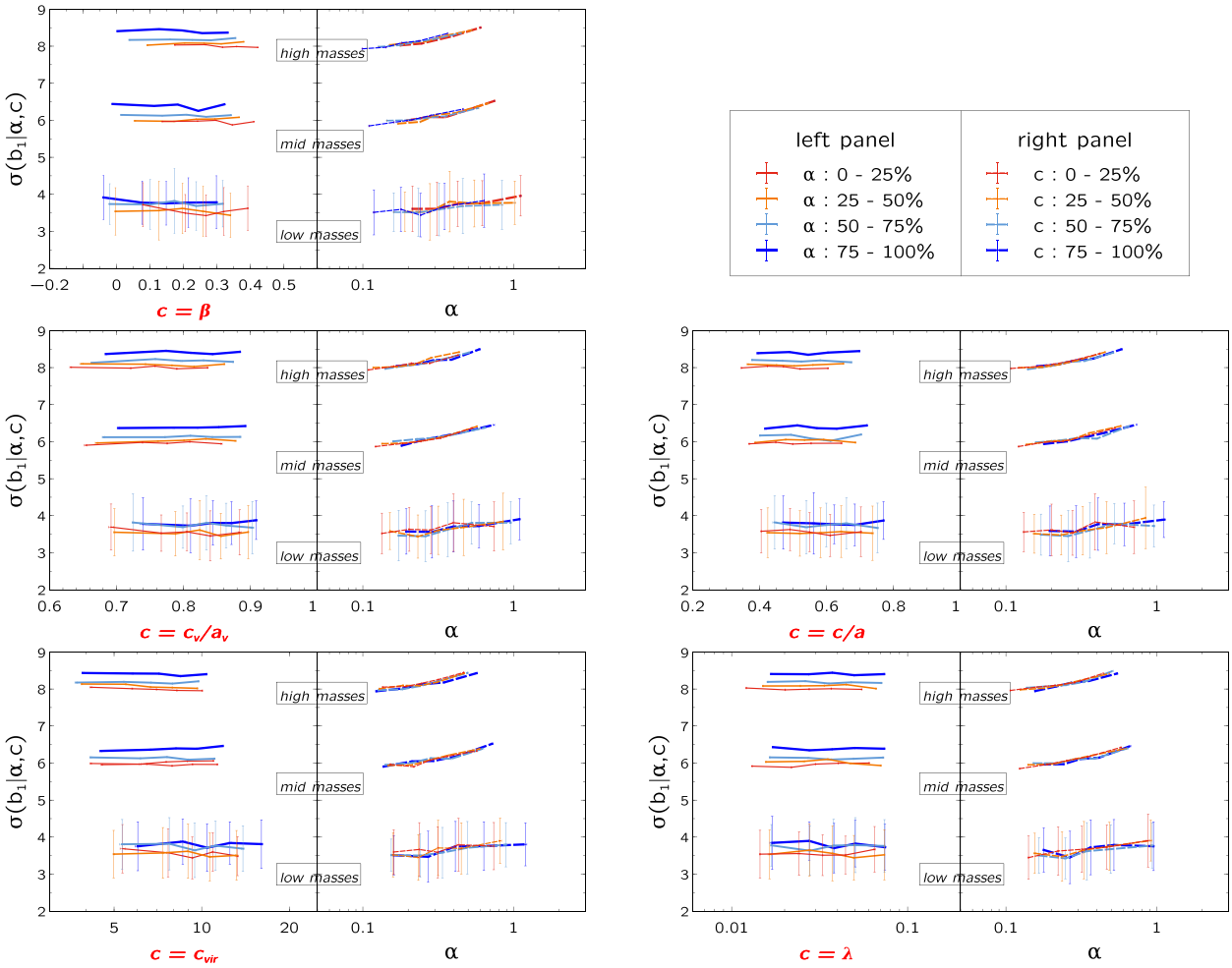

Figure A8. Square root of conditional variance of b_1 at fixed α and $c \in \{\beta,\ c_v/a_v,\ c/a,\ c_{\rm vir},\ \lambda\}$: Same as Fig. A7, showing results for $\sigma(b_1|\alpha, c) \equiv (\langle\, b_1^2|\alpha, c\,\rangle - \langle\, b_1|\alpha, c\,\rangle^2)^{1/2}$. In this case, the mid- and high-mass results in each subplot panel were given vertical offsets of +3 and +5, respectively, for clarity. The large errors in the low-mass results are likely driven by systematic effects in computing object-by-object b_1 values due to the smaller k-space range provided by the smaller volume of the high-resolution box (see also the low-mass histogram of b_1 in Fig. A6). We see that the results for all internal variables c and for each mass range are consistent with the relation $\sigma(b_1|\alpha, c) \simeq \sigma(b_1|\alpha)$. Together with the results of Fig. A7, this shows that the full distribution of b_1, α, and c is consistent with $p(\alpha, b_1, c) \simeq p(\alpha)p(b_1|\alpha)p(c|\alpha)$ for each c. See the text for a discussion.

This paper has been typeset from a TEX/LATEX file prepared by the author.