PAPER

Responses of Halo Occupation Distributions: a new ingredient in the halo model & the impact on galaxy bias

To cite this article: Rodrigo Voivodic and Alexandre Barreira JCAP05(2021)069

View the article online for updates and enhancements.

Responses of Halo Occupation Distributions: a new ingredient in the halo model & the impact on galaxy bias

Rodrigo Voivodic[a,d] **and Alexandre Barreira**[b,c,d]

[a]Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo,
 R. do Matão 1371, 05508-090, São Paulo, SP, Brazil
[b]Excellence Cluster ORIGINS,
 Boltzmannstraße 2, 85748 Garching, Germany
[c]Ludwig-Maximilians-Universität,
 Schellingstraße 4, 80799 München, Germany
[d]Max-Planck-Institut für Astrophysik,
 Karl-Schwarzschild-Straße 1, 85748 Garching, Germany

 E-mail: rodrigo.voivodic@usp.br, alex.barreira@origins-cluster.de

Abstract. Halo occupation distribution (HOD) models describe the number of galaxies that reside in different haloes, and are widely used in galaxy-halo connection studies using the halo model (HM). Here, we introduce and study HOD *response functions* $R_{\mathcal{O}}^g$ that describe the *response* of the HODs to long-wavelength perturbations $\mathcal{O}$. The linear galaxy bias parameters $b_{\mathcal{O}}^g$ are a weighted version of $b_{\mathcal{O}}^h + R_{\mathcal{O}}^g$, where $b_{\mathcal{O}}^h$ is the halo bias, but the contribution from $R_{\mathcal{O}}^g$ is routinely ignored in the literature. We investigate the impact of this by measuring the $R_{\mathcal{O}}^g$ in separate universe simulations of the IllustrisTNG model for three types of perturbations: total matter perturbations, $\mathcal{O} = \delta_m$; baryon-CDM compensated isocurvature perturbations, $\mathcal{O} = \sigma$; and potential perturbations with local primordial non-Gaussianity, $\mathcal{O} \propto f_{\mathrm{NL}}\phi$. Our main takeaway message is that the $R_{\mathcal{O}}^g$ are not negligible in general and their size should be estimated on a case-by-case basis. For stellar-mass selected galaxies, the responses R_{ϕ}^g and R_{σ}^g are sizeable and cannot be neglected in HM calculations of the bias parameters b_{ϕ}^g and b_{σ}^g; this is relevant to constrain inflation using galaxies. On the other hand, we do not detect a strong impact of the HOD response R_1^g on the linear galaxy bias b_1^g. These results can be explained by the impact that the perturbations $\mathcal{O}$ have on stellar-to-total-mass relations. We also look into the impact on the bias of the gas distribution and find similar conclusions.

We show that a single extra parameter describing the overall amplitude of $R_{\mathcal{O}}^g$ recovers the measured $b_{\mathcal{O}}^g$ well, which indicates that $R_{\mathcal{O}}^g$ can be easily added to HM/HOD studies as a new ingredient.

Contents

1 Introduction

The *halo model* (HM) is one of the most popular frameworks to predict the statistics of the large-scale structure (LSS) of the Universe (see ref. [1] for a review and ref. [2] for a recent online calculation tool). The core assumption of the HM is that all mass elements of a given tracer of the LSS, for example galaxies, are found in gravitationally-bound dark matter haloes. The description of the galaxy distribution then becomes effectively a two step process. The first step involves predicting the distribution of haloes in the Universe. Over the years, increasingly accurate N-body methods led to significant advances on this front: gravity-only codes agree to % levels nowadays [3, 4], and a number of accurate fitting formulae for dark matter halo profiles [5, 6], abundance [7, 8] and bias [9, 10] are also available. The second step involves describing the distribution of galaxies inside the haloes, which is an appreciably harder task given the many intertwined astrophysical processes that govern the formation of galaxies and their evolution. One way to tackle this problem is via so-called *Halo Occupation Distribution* (HOD) models [11–21] that describe the galaxy-halo connection via empirical parametrizations of aspects of the galaxy distribution inside halos such as their total number, radial profile and velocity profile.

One common application of the HM and HODs is the direct analytical calculation of N-point correlation functions of the galaxy distribution. In the simplest case of the 2-point correlation function, the calculation is split into the correlation between galaxies that live in the same halo and the correlation between galaxies in different haloes. The HM has also popularly been used to calculate N-point functions of the total matter distribution, although in this case, there is an important diffuse component that is not bound to haloes [22] (as well as other problems [23]) that makes the calculation less accurate; see, however, refs. [24–26] for augmented versions of the HM (some including baryonic physics effects) that have been proving accurate enough for current parameter inference analyses. Another popular application of the HM and HODs is the fast generation of galaxy mock catalogues, in which galaxies are *painted* following some HOD formulae onto halo catalogues generated with gravity-only

simulations. The fact that gravity-only simulations require far less computational resources than direct simulation of galaxy formation makes this a considerable computational advantage. The parameters of the HOD can then be iterated over to reproduce the statistics of some observed galaxy sample and learn about the galaxy-halo connection this way (see e.g. refs. [27–45] for a number of examples of such analyses and ref. [46] for a review).

The mean number of galaxies $\langle N_g(M_h, z) \rangle$ that reside in haloes with mass M_h at redshift z is a central ingredient in HOD modelling. Haloes are normally assumed to contain zero or one central massive galaxy residing at the bottom of their potential wells. The haloes which are massive enough can further contain a number of lower mass satellite galaxies that orbit around the center (figure 3 below shows some the typical shapes of $\langle N_g(M_h, z) \rangle$). For the most part in this paper we will consider galaxies as our specific example of a LSS tracer, but we note that the philosophy behind HOD modelling holds generically to other tracers as well. For example, in 21cm line intensity mapping, the relevant tracer is neutral hydrogen (HI), in which instead of a HOD number, one would use a function describing the total HI mass in haloes [47–49].

In this paper, we focus on the dependence of HOD numbers on the large-scale environment. The environmental dependence of HOD numbers has naturally been the subject of past works in the literature (see e.g. refs. [17, 50–59]), which are often carried out in the context of *assembly bias* (or secondary bias) [60–73], i.e., the dependence of galaxy clustering on properties beyond their host halo mass $M_{\rm h}$. In most of these studies, the environment is normally defined over length scales of a few times the typical virial radii of dark matter haloes ($R_{\rm env} \sim 1 - 10$ Mpc/h) on which the r.m.s. fluctuations of the density field are of order unity. Here, we focus instead on the dependence of $\langle N_g(M_h, z) \rangle$ on the environment defined over much larger linear scales ($R_{\rm env} \gtrsim 50 - 100$ Mpc/h, where the r.m.s. fluctuations are much smaller than unity), which as we will see has important ramifications for how to interpret *galaxy biasing* using the HM. We will go into more details in sections below, but let us expose the context of our work briefly here.

At leading order in perturbation theory and for primordial adiabatic Gaussian scalar fluctuations, the local number density of galaxies can be written as (see ref. [74] for a comprehensive review)

$$n_g(z, \boldsymbol{x}) = n_g(z)\Big[1 + b_1^g(z)\delta_m(z, \boldsymbol{x})\Big], \tag{1.1}$$

where $n_g(z, \boldsymbol{x})$ is the number density of galaxies in a region around position $\boldsymbol{x}$ at redshift z, $n_g(z)$ is the global average over all positions and $\delta_m(z, \boldsymbol{x})$ is a linear total matter density perturbation.[1] In this equation, b_1^g is called the linear local-in-matter-density galaxy bias parameter [75] and it quantifies how *responsive* the number density of galaxies is to the presence of the perturbation δ_m. Using the HM, the global mean number of galaxies is given by integrating the global number density of haloes $n_h(M_h)$ weighted by the global HOD number $\langle N_g(M_h) \rangle$ (we leave the dependence on z implicit to simplify the notation):

$$n_g = \int \mathrm{d}M_h n_h(M_h)\langle N_g(M_h) \rangle. \tag{1.2}$$

From this equation, the local number of galaxies around position $\boldsymbol{x}$ can be worked out by promoting both $n_h(M_h)$ and $\langle N_g(M_h) \rangle$ to be *biased tracers* of δ_m, i.e., to admit a description

---

[1]We distinguish local from global averaged quantities by the presence or absence of the position $\boldsymbol{x}$ in the argument. In this short exposition of the problem we also skip writing the contribution from *stochastic* terms that encapsulate the dependence of galaxy formation on the shorter-wavelength details of the environment.

analogous to that in eq. (1.1):

$$n_h(M_h, \boldsymbol{x}) = n_h(M_h)\big[1 + b_1^h(M_h)\delta_m(\boldsymbol{x})\big]$$
$$\langle N_g(M_h, \boldsymbol{x})\rangle = \langle N_g(M_h)\rangle\big[1 + R_1^g(M_h)\delta_m(\boldsymbol{x})\big], \qquad (1.3)$$

where the parameters b_1^h and R_1^g are the equivalent of b_1^g but for the dark matter halo abundance and the galaxy HOD number, respectively; we will refer to the parameter R_1^g as the *HOD number response function*. It therefore follows that

$$n_g(\boldsymbol{x}) = n_g\left[1 + \frac{1}{n_g}\int \mathrm{d}M_h n_h(M_h)\langle N_g(M_h)\rangle\big(b_1^h(M_h) + R_1^g(M_h)\big)\delta_m(\boldsymbol{x})\right], \qquad (1.4)$$

which can be contrasted with eq. (1.1) to derive the galaxy bias parameter as

$$b_1^g = \frac{1}{n_g}\int \mathrm{d}M_h n_h(M_h)\langle N_g(M_h)\rangle\big(b_1^h(M_h) + R_1^g(M_h)\big). \qquad (1.5)$$

The standard HM lore affirms that the bias of the galaxies is inherited directly from that of the host haloes weighted by the HOD number. This is recovered by the last equation only when $R_1^g = 0$, i.e., assuming that the mean galaxy HOD numbers $\langle N_g(M_h)\rangle$ are the same at all positions in the Universe. There is however no prior reason to expect this to be the case; in fact, given that the dynamics of structure formation are sensitive to whether structures form in overdense, $\delta_m > 0$, or underdense regions, $\delta_m < 0$, it is actually physically plausible to expect the mean number of galaxies inside haloes to display some of this sensitivity.[2] This is an issue that is briefly alluded to in refs. [74, 76], but which to the best of our knowledge has never been specifically addressed and tested in practice before. *This motivates our main goal in this paper*, which is to measure the size and shape of HOD response functions like R_1^g using galaxy formation simulations and discuss the corresponding implications to the modelling of large-scale structure using the HM and HODs.

Concretely, in this paper we measure responses of HOD numbers $\langle N_g(M_h)\rangle$ using separate universe simulations of the IllustrisTNG galaxy formation model [77–79]. Separate universe simulation is a technique that allows to efficiently measure the responses of LSS quantities to long-wavelength perturbations by absorbing the effects of the perturbations into appropriate modifications to the cosmology of the simulation and its box size. These simulations have been run in previous works for the case of total matter density perturbations δ_m [80], primordial baryon-CDM compensated isocurvature perturbations (CIP) σ [81] and potential perturbations with local primordial non-Gaussianity (PNG) $f_{\mathrm{NL}}\phi$ [82] (we describe these perturbations and simulations with more detail in the next section). We measure the responses of the HOD numbers $\langle N_g(M_h)\rangle$ of the simulated galaxies to these three types of perturbations, as well as the response of the gas mass inside haloes for a few example elements tracked in IllustrisTNG. Our main finding is that the HOD response functions are not negligible in general and can lead to sizeable effects on the corresponding galaxy bias parameters

---

[2]This is the general finding of a number of recent studies in the literature [51–57] using definitions of the environment on $1-10$ Mpc/h scales. For example, in very dense environments, mergers take place more frequently, which boosts the number of substructures at fixed halo mass. In contrast, in less dense regions [58], a larger fraction of the build up of halo mass is due to the infall of surrounding diffuse matter. In this paper, we focus on environments defined over large linear scales ($\gtrsim 50-100$ Mpc/h, where $\delta_m \ll 1$) relevant to the calculation of b_1^g using the HM.

calculated using the HM. We also see that modelling the HOD responses as low order polynomials ($n \leq 2$) in $\log M_h$ recovers the correct value of the bias parameters. This all motivates the incorporation of HOD responses as an extra ingredient in traditional HOD studies.

The rest of this paper is organized as follows. In section 2, we introduce the three types of long-wavelength perturbations we consider, as well as the separate universe simulations that we use to measure the corresponding HOD number responses. Section 3 contains our main results on the measurements of $\langle N_g(M_h) \rangle$ and its response functions, and the illustration of the general necessity to take the responses into account within the HM to correctly recover the measured bias parameters. We show results for samples with varying total and stellar mass cuts, as well as for the distribution of hydrogen, carbon and oxygen as a few example gas elements tracked in IllustrisTNG. We summarise and conclude in section 4.

2   Responses and separate universe simulations

Given any quantity Q in LSS, its response functions $R_{\mathcal{O}}^{Q}$ can be defined with all generality via the expansion

$$Q(z, \boldsymbol{x}) = Q(z) \left[1 + \sum_{\mathcal{O}} R_{\mathcal{O}}^{Q}(z) \mathcal{O}(z, \boldsymbol{x}) \right] + \epsilon^{Q}(z, \boldsymbol{x}), \qquad (2.1)$$

where $Q(z, \boldsymbol{x})$ denotes the value of the quantity measured in some local volume around $\boldsymbol{x}$, $Q(z)$ is the corresponding cosmic average and the sum runs over all long-wavelength perturbations $\mathcal{O}$ that can influence the quantity's local value. This equation makes apparent the physical meaning of the $R_{\mathcal{O}}^{Q}$ as the *response* (hence the name) of Q to the presence of the perturbations $\mathcal{O}$. The response functions depend generically on redshift, as well as any other variable Q may also depend on. In this equation, the perturbations $\mathcal{O}$ are assumed to be sufficiently long-wavelength that they can be treated perturbatively [83]. In eq. (2.1), $\epsilon^{Q}(z, \boldsymbol{x})$ is added to absorb all of the dependence of the value of $Q(z, \boldsymbol{x})$ on the shorter-wavelength part of the perturbations $\mathcal{O}$. We note that although this so-called stochastic contribution $\epsilon^{Q}(z, \boldsymbol{x})$ does not correlate with any of the $\mathcal{O}$, it is still important to take it into account in practical descriptions of the statistics of $Q(z, \boldsymbol{x})$; in this paper we focus solely on the response functions inside the squared bracket. We refer the interested reader to ref. [74] for a comprehensive review of the application of the expansion of eq. (2.1) to galaxy number densities as the quantity Q (popularly known as the galaxy bias expansion) and to refs. [84–90] for an application of the same ideas to matter/galaxy power spectra.

In this paper we look into the responses to three types of long-wavelength perturbations $\mathcal{O}$:

1. The first are total matter density perturbations $\mathcal{O} = \delta_m(z, \boldsymbol{x})$, which originate from adiabatic energy density fluctuations in the early Universe;

2. The second are baryon-CDM compensated isocurvature perturbations (CIP) $\mathcal{O} = \sigma(\boldsymbol{x})$ [91–102], which are characterized by perturbations in the baryon density that are exactly compensated by the cold dark matter (CDM) to leave the total matter distribution unchanged. The power spectrum of CIPs, which cannot be generated by single-field inflation models, is remarkably poorly constrained by the CMB data [103]. This has motivated a number of studies of their impact on galaxy statistics, which have shown to be possible to significantly improve on the current observational bounds [99–102];

3. The third type are perturbations of the primordial gravitational potential in local PNG cosmologies, $\mathcal{O} = f_{\mathrm{NL}}\phi(\boldsymbol{x})$; the value of f_{NL} describes the amount of non-Gaussianity via $\phi = \phi_{\mathrm{G}} + f_{\mathrm{NL}}\left[\phi_{\mathrm{G}}^2 - \langle\phi_{\mathrm{G}}^2\rangle\right]$, where ϕ_{G} is a Gaussian random variable and $\langle\rangle$ denotes ensemble average [104]. The tightest observational bounds constrain $f_{\mathrm{NL}} = 0.9 \pm 5.1$ [105] and there is significant interest in improving further on this bound given the power of a detection of $f_{\mathrm{NL}} \neq 0$ to rule out multi-field models of inflation [106–113].

We will focus on the responses of galaxy number densities n_g and galaxy HOD numbers $\langle N_g(M_h)\rangle$. For these two quantities and for the three types of perturbations we consider, the corresponding response expansions are given, respectively, by (recall, we distinguish the local from the global quantities by the presence of the position $\boldsymbol{x}$ in the arguments)

$$n_g(\boldsymbol{x}, z) = n_g(z)\Big[1 + b_1^g(z)\delta_m(z, \boldsymbol{x}) + b_\sigma^g(z)\sigma(\boldsymbol{x}) + b_\phi^g(z)f_{\mathrm{NL}}\phi(\boldsymbol{x})\Big] \tag{2.2}$$

and

$$\langle N_g(M_h, z, \boldsymbol{x})\rangle = \langle N_g(M_h, z)\rangle\Big[1 + R_1^g(M_h, z)\delta_m(z, \boldsymbol{x}) + R_\sigma^g(M_h, z)\sigma(\boldsymbol{x}) + R_\phi^g(M_h, z)f_{\mathrm{NL}}\phi(\boldsymbol{x})\Big],$$

where we skipped writing the stochastic contributions explicitly since we do not study them here.

In the remainder of this section, we outline the main specifications of the separate universe simulations that we use in this paper to measure $b_{\mathcal{O}}^g$ and $R_{\mathcal{O}}^g$. The separate universe simulation technique builds on top of the assumption that the physics that determine the quantity Q act on sufficiently small scales that they regard the large-scale perturbations $\mathcal{O}$ as changes to the background (this is also called the peak-background split approach [114, 115]). Specifically, the separate universe ansatz states that *the formation of structures locally inside long-wavelength perturbations is equivalent to the formation of structures globally in an appropriately modified cosmology.* For the case of $\mathcal{O} = \delta_m$ the modified cosmology has a different mean background total matter density, for $\mathcal{O} = \sigma$ the cosmology has different relative amplitudes of the cosmic fractions of baryons and CDM, and for $\mathcal{O} = f_{\mathrm{NL}}\phi$ the modified cosmology has a different amplitude $\mathcal{A}_s$ of the primordial scalar power spectrum.

The simulations we use here have been described in previous works, and so we shall be brief in descriptions and refer the interested reader to the cited literature for more details and derivations. Next, we describe first the main numerical aspects that are common to all simulations and then comment in turn on the specifics of the three types of separate universe simulations we consider; the main idea and numerical details are summarized in figure 1 and table 1. At the end of this section, we show the measurements of the halo bias $b_{\mathcal{O}}^h$ from the separate universe simulations (the equivalent of the $b_{\mathcal{O}}^g$ in eq. (2.2), but for halo number counts).

2.1 Numerical details

The simulations we use in this work were run with the moving-mesh code AREPO [116, 117] and the IllustrisTNG galaxy formation model [77–79]. The latter is an improved version of the Illustris model [118, 119] and it includes prescriptions for gas cooling and ionization, star formation and feedback, and black hole growth and feedback; refs. [120–124] present and discuss a number of the first key results obtained with IllustrisTNG.

In our main results, we consider simulations run at two mass resolutions. One, which we label as TNG100-1.5, corresponds to a box size $L_{\mathrm{box}} = 75\mathrm{Mpc}/h$ and $N_p = 2 \times 1250^3$ tracer elements; and another, labeled TNG300-2, with $L_{\mathrm{box}} = 205\mathrm{Mpc}/h$ and $N_p = 2 \times 1250^3$. The

Separate Universe δ_{m} Separate Universe σ Separate Universe $\mathcal{A}_{\mathrm{s}}$

$$\delta_m(\mathbf{x}, t)$$

$$\delta_c(\mathbf{x})$$
$$\delta_b(\mathbf{x})$$

$$4 f_{\mathrm{NL}} \phi(\mathbf{x})$$

$$\rho_m(t) \;\rightarrow\; \rho_m(t)\left[1 + \delta_L(t)\right]$$

$$\rho_b(t) \;\rightarrow\; \rho_b(t)\left[1 + \sigma_L\right]$$
$$\rho_c(t) \;\rightarrow\; \rho_c(t)\left[1 - f_b \sigma_L\right]$$

$$\mathcal{A}_s \;\rightarrow\; \mathcal{A}_s\left[1 + \delta\mathcal{A}_s\right]$$
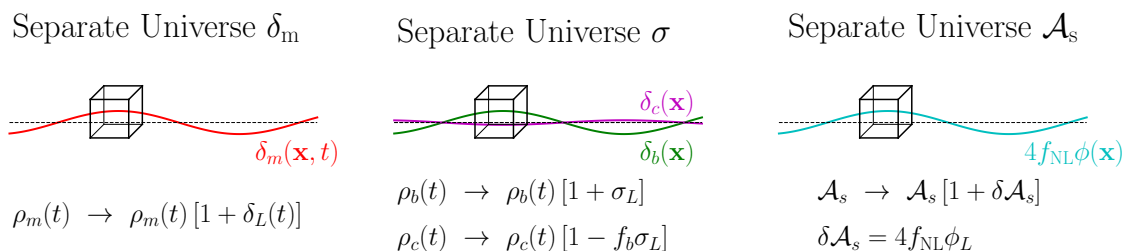$$\delta\mathcal{A}_s = 4 f_{\mathrm{NL}} \phi_L$$

Figure 1. Summary sketch of the separate universe approach to the three types of long-wavelength perturbations considered in this paper. On the left, local structure formation inside long-wavelength matter perturbations δ_m is equivalent to global structure formation in a separate cosmology with modified background matter density ρ_m. In the center, local structure formation inside long-wavelength CIP perturbations σ is equivalent to global structure formation with modified background baryon ρ_b and CDM ρ_c densities, at fixed total matter density. On the right, local structure formation inside long-wavelength primordial gravitational potential perturbations with local PNG is equivalent to global structure formation in a cosmology with modified amplitude of the primordial scalar perturbation power spectrum $\mathcal{A}_s$.

initial conditions were generated at $z = 127$ with the N-GENIC code [125] using the Zel'dovich approximation; the input linear matter power spectrum was calculated with the CAMB code [126]. For the most part in this paper, we show results from the full hydrodynamical simulations with IllustrisTNG (dubbed Hydro), although in section 3.1 we display also results from their gravity-only counterparts (with half the number of mass elements; we dub these as Gravity). There, we consider also an additional set of gravity-only separate universe simulations available for the cases $\mathcal{O} = \delta_m$ and $\mathcal{O} = f_{\mathrm{NL}}\phi$, with $L_{\mathrm{box}} = 560 \mathrm{Mpc}/h$ and $N_p = 1250^3$; we label this resolution simply as $L_{\mathrm{box}} \approx 800 \mathrm{Mpc}$.

Gravitationally bound haloes are identified with a Friends-of-Friends (FOF) code run on the dark matter tracer particles with a linking length $b = 0.2$ times the mean interparticle distance. Inside each halo, subhaloes are found using the SUBFIND algorithm [127]. As is standard, we define the *main central* subhalo to be that which resides at the bottom of the potential of the FOF object, with all of the remaining subhaloes being called *satellites*. When quoting mass values for these structures we always consider the summed mass from all elements that belong to the halo/subhalo. For example, the total stellar mass of a halo is the summed mass of all star particles that are assigned to the halo, which includes those that also belong to its subhalos. Following the standard nomenclature in IllustrisTNG-related literature, we refer to subhalos that contain any mass in stars as *galaxies*.

2.2 The δ_m separate universe

According to the separate universe ansatz, long-wavelength total matter perturbations are regarded by smaller-scale structure formation as a uniform change in the total matter density. Hence, the measurements of the responses to δ_m can be carried out by comparing the results from the Fiducial cosmology with those from cosmologies with modified cosmic total matter density. If $\delta_L(t)$ denotes the amplitude of the total (linear) matter perturbation at physical time t, then the separate universe cosmology is characterized by

$$\tilde{\rho}_m(t) = \rho_m(t)\left[1 + \delta_L(t)\right], \tag{2.3}$$

Name	Ω_{m0}	Ω_{b0}	Ω_{c0}	$\Omega_{\Lambda 0}$	h	n_s	$\mathcal{A}_s$
Fiducial	0.3089	0.0486	0.2603	0.6911	0.6774	0.967	2.068×10^{-9}
High δ_m	0.3194	0.0502	0.2692	0.7146	0.6662	Fiducial	Fiducial
Low δ_m	0.2991	0.0471	0.2520	0.6691	0.6884	Fiducial	Fiducial
High σ	Fiducial	0.0510	0.2579	Fiducial	Fiducial	Fiducial	Fiducial
Low σ	Fiducial	0.0462	0.2627	Fiducial	Fiducial	Fiducial	Fiducial
High $\mathcal{A}_s$	Fiducial	Fiducial	Fiducial	Fiducial	Fiducial	Fiducial	2.171×10^{-9}
Low $\mathcal{A}_s$	Fiducial	Fiducial	Fiducial	Fiducial	Fiducial	Fiducial	1.965×10^{-9}

Table 1. Cosmological parameters of the simulations used in this paper. The High δ_m and Low δ_m simulations are used to measure responses to δ_m; High σ and Low σ are used to measure the responses to CIP perturbations σ; and High $\mathcal{A}_s$ and Low $\mathcal{A}_s$ are used to measure the responses to $\delta\mathcal{A}_s = 4f_{\rm NL}\phi$ (cf. sections 2.2, 2.3 and 2.4). These cosmologies were simulated with AREPO and IllustrisTNG at two numerical resolutions: TNG100-1.5 with $L_{\rm box} = 75$ Mpc/h and TNG300-2 with $L_{\rm box} = 205$ Mpc/h, both with $N_p = 2 \times 1250^3$ mass elements. For all cases except High σ and Low σ, there is an additional numerical resolution for gravity-only dynamics with $L_{\rm box} = 560$ Mpc/h and $N_p = 1250^3$, which we label simply as $L_{\rm box} \approx 800$Mpc. In all cosmologies, we approximate neutrinos as massless (see refs. [128–130] for separate universe simulations where neutrinos and other scale-dependent effects arising in quintessence cosmologies are taken into account).

where a tilde indicates a quantity in the separate universe cosmology. The corresponding cosmological parameters are listed in table 1; the cosmologies High δ_m and Low δ_m mimic the effects of positive and negative amplitude perturbations with present-day values $\delta_L^{\rm High}(z=0) = +0.05$ and $\delta_L^{\rm Low}(z=0) = -0.05$, respectively (note that δ_L evolves with time according to linear theory). The size of the amplitude of these perturbations is chosen as a compromise between having it to be large enough for the measurements to be high signal-to-noise, but small enough to keep higher-order terms (e.g. $\mathcal{O} = \delta_m^2$) in the response expansion negligible.

The different values of h and the different relation between redshift and physical time in the Fiducial, High δ_m and Low δ_m cosmologies imply a number of important adjustments to the box size, FoF linking length, output times and setting up of the initial conditions. These have been discussed at length in previous works [10, 80, 82, 84, 112, 128, 131–135], to which we refer the reader for more details (see e.g. refs. [80, 82] for the descriptions of the actual separate universe simulations we use here). When we quote units with h factors, we always convert to the Fiducial cosmology.

From eqs. (2.2) and (2.3), the bias parameter b_1^g and the $\langle N_g(M_h) \rangle$ response R_1^g can be formally defined as

$$b_1^g(z) = \left. \frac{{\rm d}\ln n_g(z)}{{\rm d}\delta_L(z)} \right|_{\delta_L(z)=0} \quad ; \quad R_1^g(M_h, z) = \left. \frac{{\rm d}\ln \langle N_g(M_h, z) \rangle}{{\rm d}\delta_L(z)} \right|_{\delta_L(z)=0}, \tag{2.4}$$

which we evaluate by finite differencing the results from the Fiducial, High δ_m and Low δ_m simulations. Concretely, we evaluate $R_1^g(M_h, z)$ as

$$R_1^g(M_h, z) = \frac{R_1^{g,{\rm High}\ \delta_m}(M_h, z) + R_1^{g,{\rm Low}\ \delta_m}(M_h, z)}{2}, \tag{2.5}$$

with

$$R_1^{g,\text{High } \delta_m}(M_h, z) = \frac{1}{\delta_L^{\text{High}}(z)} \left[\frac{\langle N_g(M_h, z)\rangle^{\text{High } \delta_m}}{\langle N_g(M_h, z)\rangle^{\text{Fiducial}}} - 1 \right],$$

$$R_1^{g,\text{Low } \delta_m}(M_h, z) = \frac{1}{\delta_L^{\text{Low}}(z)} \left[\frac{\langle N_g(M_h, z)\rangle^{\text{Low } \delta_m}}{\langle N_g(M_h, z)\rangle^{\text{Fiducial}}} - 1 \right], \tag{2.6}$$

where $\langle N_g(M_h)\rangle^{\text{Cosmology}}$ denotes the galaxy HOD number measured in the corresponding cosmology. The calculation of b_1^g is done analogously by finite differencing the total number of galaxies in the simulation box, instead of the number of galaxies inside haloes with mass M_h. In doing so, one actually measures the so-called Lagrangian value $b_1^{g,\text{Lag.}}$, which is related to the (Eulerian) value that we consider in this paper by $b_1^g = b_1^{g,\text{Lag.}} + 1$.[3]

   An important point to note is that we have only one realization of the initial conditions, which prevents us from quoting error bars on our measurements in a statistical ensemble sense. The size of the simulation boxes also makes it hard to estimate errors via resampling of subvolumes. The responses High δ_m and Low δ_m are the forward and backward first-derivatives and are therefore the same in theory. In practice, however, numerical noise and binning effects can drive some differences, and so we take the different between High δ_m and Low δ_m as a rough estimate of the error in our measurements (ref. [82] verified that, at least for b_1^g, these error bars are comparable in size to the estimate based on jackknife resampling presented in ref. [134]).

2.3   The σ separate universe

A long-wavelength CIP is regarded by the structures within it as a uniform change to the cosmic fraction of baryons Ω_b and CDM Ω_c, at fixed total matter Ω_m. With the convention that $\sigma > 0$ corresponds to more baryons (less CDM), then the responses to CIPs can be evaluated using separate universe cosmologies characterized by

$$\tilde{\Omega}_{b0} = \Omega_{b0}\left[1 + \sigma_L\right] \quad ; \quad \tilde{\Omega}_{c0} = \Omega_{c0}\left[1 - f_b\sigma_L\right], \tag{2.7}$$

where σ_L is the amplitude of the long-wavelength CIP and f_b is the ratio of baryon-to-CDM density in the fiducial cosmology; note that σ_L is constant in time because on large-scales the only relevant force is gravity, which acts equally on baryons and CDM. The implementation of these separate universe simulations is straightforward with the only difference relative to the Fiducial being just that the initial linear matter power spectrum should be generated for different values of Ω_{b0} and Ω_{c0} (see also ref. [136] for a numerical study of CIPs in which different transfer functions are used to generate the initial distribution of baryons and CDM). We consider two CIP cosmologies, High σ and Low σ, which correspond to $\sigma_L = +0.05$ and $\sigma_L = -0.05$, respectively (cf. table 1); these are the same simulations used previously in ref. [81].

   The response functions are defined as

$$b_\sigma^g(z) = \left.\frac{\mathrm{d}\ln n_g(z)}{\mathrm{d}\sigma_L}\right|_{\sigma_L=0} \quad ; \quad R_\sigma^g(M_h, z) = \left.\frac{\mathrm{d}\ln \langle N_g(M_h, z)\rangle}{\mathrm{d}\sigma}\right|_{\sigma_L=0}, \tag{2.8}$$

and we evaluate them via finite-differencing analogously to b_1^g and R_1^g above.

---

[3]This follows from the relation between number densities in Eulerian and Lagrangian space.

2.4 The $\mathcal{A}_s$ separate universe

In cosmologies with local PNG ($f_{\rm NL} \neq 0$), long-wavelength perturbations of the primordial gravitational potential are regarded by local structure formation inside them as a rescaling of the amplitude $\mathcal{A}_s$ of the scalar primordial power spectrum [137, 138]. Concretely, using the separate universe picture, it can be shown that the responses to these types of perturbations can be evaluated by comparing the Fiducial cosmology with cosmologies characterized by[4]

$$\tilde{\mathcal{A}}_s = \mathcal{A}_s \left[1 + \delta\mathcal{A}_s\right] \;, \quad \text{where} \;\; \delta\mathcal{A}_s = 4f_{\rm NL}\phi_L, \tag{2.10}$$

with ϕ_L being the amplitude of the primordial potential perturbation. The amplitude $\mathcal{A}_s$ is the only parameter that differs relative to the fiducial cosmology, which makes these separate universe simulations also straightforward to setup: the only change is at the level of the initial conditions, which should be generated with an initial matter power spectrum rescaled by $1 + \delta\mathcal{A}_s$. The two cosmologies we consider, High $\mathcal{A}_s$ and Low $\mathcal{A}_s$, are characterized by $\delta\mathcal{A}_s = +0.05$ and $\delta\mathcal{A}_s = -0.05$, respectively (cf. table 1); these are the same simulations used previously in ref. [82].

Similarly to the other two cases above, the responses to $\mathcal{A}_s$ are defined as

$$b_\phi^g(z) = 4\frac{{\rm dln}\, n_g(z)}{{\rm d}\delta\mathcal{A}_s}\bigg|_{\delta\mathcal{A}_s=0} \qquad ; \qquad R_\phi^g(M_h, z) = 4\frac{{\rm dln}\, \langle N_g(M_h, z)\rangle}{{\rm d}\delta\mathcal{A}_s}\bigg|_{\delta\mathcal{A}_s=0}, \tag{2.11}$$

and we evaluate them via finite-differencing analogously to the responses to δ_m and σ described already above. The factor of 4 in eqs. (2.11) accounts simply for the fact that the responses multiply $f_{\rm NL}\phi$ in eqs. (2.2) and (2.3), but the amplitude rescaling is given by $\delta\mathcal{A}_s = 4f_{\rm NL}\phi$.

2.5 The halo bias parameters

Figure 2 shows the halo mass and redshift dependence of the linear halo bias parameters b_1^h, b_σ^h and b_ϕ^h. These are defined and measured analogously to the galaxy bias parameters, but considering halo instead of galaxy number densities. The result is shown for the three numerical resolutions available for gravity-only dynamics, which all agree to within the precision of our measurements. In the left panels for b_1^h, the black solid line shows the prediction from the fitting formula of ref. [9], whereas in the center and right panels, for b_σ^h and b_ϕ^h, respectively, the solid black line shows the result obtained using the halo mass function fitting formula of ref. [8] with the definition of b_σ^g and b_ϕ^g in eqs. (2.8) and (2.11), respectively. Specifically, we use the formula of ref. [8] to evaluate the mass function at the corresponding cosmologies in table 1, and then compute the derivatives in eqs. (2.8) and (2.11) using finite differences.

These measurements have been presented and discussed with more detail in ref. [82] for b_1^h and b_ϕ^h and ref. [81] for b_σ^h (see also ref. [136]), to which we refer the reader for more details.

---

[4]Local PNG generates a primordial bispectrum that peaks in the squeezed limit, i.e., it generates a coupling between long-wavelength and short-wavelength modes. Concretely, the long-wavelength mode acts as a modified *background* to the local, short-scale power spectrum, which gets modulated as (see refs. [137, 138] or section 7 of ref. [74] for the squeezed bispectrum derivation):

$$P_{\phi\phi}(k_S, \boldsymbol{x}) = \left[1 + 4f_{\rm NL}\phi(k_L)e^{i\boldsymbol{x}.\boldsymbol{k}_L}\right]P_{\phi\phi}(k_S), \tag{2.9}$$

where $P_{\phi\phi}(k_S, \boldsymbol{x})$ is the local power spectrum of the primoridal gravitational potential, $P_{\phi\phi}(k_S)$ is its spatial average, and k_S and k_L denote short- and long-wavelength modes, respectively. Thus, on distance scales sufficiently smaller than $1/k_L$, the effects on structure formation are equivalent to a rescaling of the amplitude of the primordial scalar power spectrum by $1 + \delta\mathcal{A}_s$, where $\delta\mathcal{A}_s = 4f_{\rm NL}\phi_L$.
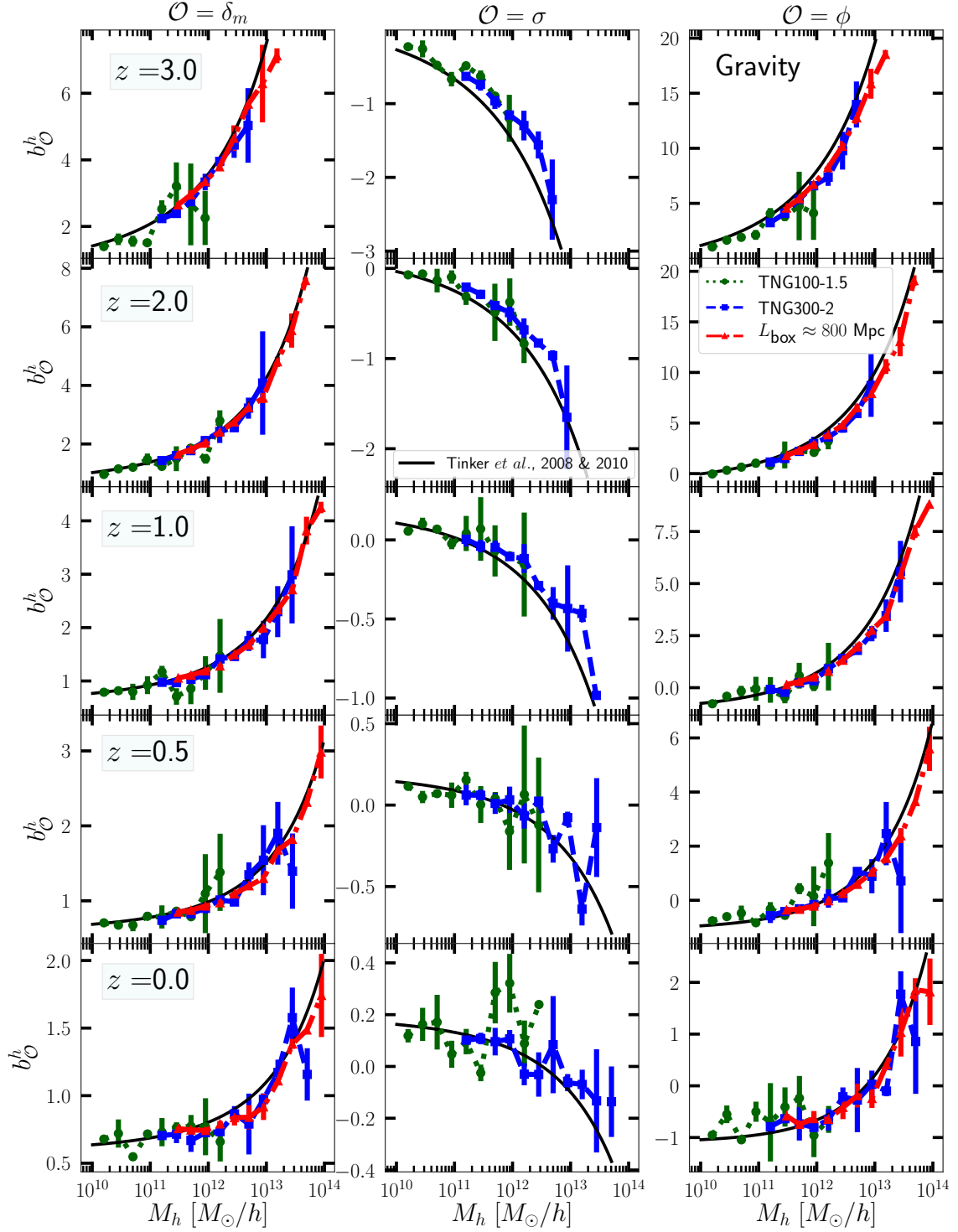
Figure 2. Halo bias parameters b_1^h (left), b_σ^h (center) and b_ϕ^h (right) as a function of total halo mass at different redshifts (different panels) and for the three gravity-only resolutions we consider in this paper, as labeled. The bias parameters are defined by the analog of eq. (2.2) for the halo (not galaxy) number density. In the left panels, the solid black line shows the prediction from the b_1^h fitting formula of ref. [9]. In the center and right panels, the solid line shows the result obtained by applying the separate universe ansatz to the halo mass function fitting formula of ref. [8].

We display them here for completeness and because the halo bias parameters are a crucial ingredient in the HM calculation of the galaxy bias parameters that we show in the next section. Perhaps the only aspect of figure 2 that may be less familiar and whose clarification is worth repeating here has to do with the appreciably different mass dependence of b_σ^h, which is monotonically decreasing, compared to that of both b_1^h and b_ϕ^h, which is monotonically increasing. As explained in ref. [81], this is due to the fact that a boost in Ω_b, at fixed Ω_m, slows down the growth of total matter perturbations on sub-horizon scales between radiation domination and baryon-photon decoupling. This lowers the amplitude of the total matter power spectrum after decoupling for $k > k_{eq}$, which effectively suppresses the subsequent formation of nonlinear structures at later times. On the other hand, boosts in ρ_m and $\mathcal{A}_s$ both enhance hierarchical structure formation overall, hence the monotonic increase of both b_1^h and b_ϕ^h with total halo mass.

We have checked (not shown) that the values of b_1^h and b_ϕ^h measured from the full Hydro versions of the simulations are consistent with those shown in figure 2 for the Gravity case. The same is true for b_σ^h although in this case the Hydro results tend to slightly, but systematically, underpredict the Gravity results at lower masses (cf. figure 1 vs. figure 2 in ref. [81]).

3 HOD responses and their impact on galaxy bias

In this section we present our results on the HOD number responses and their importance in the calculation of galaxy bias using the HM. We have already exposed the problem we wish to address in the Introduction for the case of $\mathcal{O} = \delta_m$ perturbations, but let us re-derive eq. (1.5) again from a slightly different angle and generalize to the case of the $\mathcal{O} = f_{NL}\phi$ and $\mathcal{O} = \sigma$ perturbations as well.

Within the HM framework, eq. (1.2) describes the total number of galaxies n_g in the Universe as the integral over the halo mass function n_h weighted by the HOD number $\langle N_g(M_h) \rangle$:

$$n_g = \int \mathrm{d}M_h\, n_h(M_h) \langle N_g(M_h) \rangle. \tag{3.1}$$

In keeping with the response expansion of eq. (2.1) and the separate universe argument, the galaxy bias parameters are given by the logarithmic derivative of n_g to the perturbations (or changes in cosmology) $\mathcal{O}$, i.e., $b_\mathcal{O}^g = \partial \ln n_g / \partial \mathcal{O}$. Applying this to eq. (3.1) gives

$$\begin{aligned}
b_\mathcal{O}^g &= \frac{1}{n_g} \frac{\partial}{\partial \mathcal{O}} \int \mathrm{d}M_h\, n_h(M_h) \langle N_g(M_h) \rangle \\
&= \frac{1}{n_g} \int \mathrm{d}M_h\, n_h(M_h) \langle N_g(M_h) \rangle \left[\frac{\partial \ln n_h(M_h)}{\partial \mathcal{O}} + \frac{\partial \ln \langle N_g(M_h) \rangle}{\partial \mathcal{O}} \right] \\
&= \frac{1}{n_g} \int \mathrm{d}M_h\, n_h(M_h) \langle N_g(M_h) \rangle \left[b_\mathcal{O}^h(M_h) + R_\mathcal{O}^g(M_h) \right],
\end{aligned} \tag{3.2}$$

where in the last equality we have used the definition of halo bias and the $\langle N_g(M_h) \rangle$ responses as the logarithmic derivative of the halo abundances and HOD numbers, respectively. For the case of $\mathcal{O} = \delta_m$, this equation recovers eq. (1.5).

In our main results below, we compare the values of $b_\mathcal{O}^g$ measured using the separate universe simulations with the result of eq. (3.2), with and without the contribution from

$R^g_{\mathcal{O}}(M_h)$. In our numerical calculations, we evaluate the integral of eq. (3.2) over the range[5] $M_h \in \left[10^{10}; 10^{14}\right] M_\odot/h$ with the quantities n_h, $\langle N_g(M_h)\rangle$, $b^h_{\mathcal{O}}$ and $R^g_{\mathcal{O}}$ obtained by interpolating the simulation measurements. Our HM predictions therefore display some numerical noise inherited from the simulations. This is not ideal from a generic theory prediction perspective (for instance, one could use the solid lines in figure 2 to describe $b^h_{\mathcal{O}}$), but it facilitates our goal to demonstrate that taking into account all of the ingredients of eq. (3.2) (as measured in the simulations) describes well the galaxy bias parameters (measured from the same simulations).

In the remainder of this section, we present our numerical results in turn for (i) subhalos selected by total mass, (ii) galaxies selected by stellar mass, and (iii) for a few example gas elements tracked by the IllustrisTNG model.

3.1 Subhalos selected by total mass

We begin our exploration of the importance of the HOD number responses $R^g_{\mathcal{O}}$ in eq. (3.2) by considering the case of subhalos selected by their total mass $M_{\rm tot}$ in the Gravity simulations. This is not the most realistic scenario from an observational viewpoint, but it is interesting to analyse nonetheless to subsequently compare with the results from other selection criteria below. The higher degree of numerical convergence between different gravity-only resolutions also facilitates the comparison, and in the Gravity case, we can make use also of the larger volume $L_{\rm box} \approx 800{\rm Mpc}$ set of simulations.

Figure 3 shows the mean galaxy HOD numbers $\langle N_g(M_h)\rangle$ measured in the simulations of the Fiducial cosmology for three subhalo total mass samples: $M_{\rm tot} > 10^{11}~M_\odot/h$ (dashed), $M_{\rm tot} > 10^{12}~M_\odot/h$ (solid) and $M_{\rm tot} > 10^{13}~M_\odot/h$ (dotted). The result is shown for the three resolutions TNG100-1.5, TNG300-2 and $L_{\rm box} \approx 800{\rm Mpc}$ (which agree very well with one another), and at different redshifts, as labeled. The redshift and mass dependence of the HOD numbers in figure 3 is in line with the expectation. Namely, $\langle N_g(M_h)\rangle$ transitions from zero to one at around the minimum mass of the corresponding subhalo population (i.e., the main central subhalo cannot be more massive than the host halo), and towards higher halo masses the HOD number increases in power-law fashion reflecting the larger number of satellite subhaloes that reside in massive haloes.

The $\langle N_g(M_h)\rangle$ curves shown in figure 3 and their response functions $R^g_{\mathcal{O}}$ (together with the halo mass function n_h and halo bias $b^h_{\mathcal{O}}$) can be plugged into eq. (3.2) to work out the HM predictions for the galaxy (or subhalo here) bias parameters $b^g_{\mathcal{O}}$. The result is shown in figure 4 for b^g_1 (left), b^g_σ (center) and b^g_ϕ (right), and for different resolutions, redshifts and minimum mass cuts, as labeled. The symbols with error bars show the $b^g_{\mathcal{O}}$ measured directly from the separate universe simulations. The dashed lines show the outcome of eq. (3.2), while the solid lines show the same but with the HOD responses artificially set to zero $R^g_{\mathcal{O}} = 0$; their difference thus measures directly the importance of the $R^g_{\mathcal{O}}$. The shaded bands around the dashed lines indicate the error due to the uncertainty in the HOD response measurements alone. Overall, the figure shows that both HM calculations agree well with the separate universe results, which is indicative of a small size of $R^g_{\mathcal{O}}$ for subhalos selected by their total mass in gravity-only simulations.

---

[5]This is also the range of halo masses we consider in the calculation of bias parameters and responses using the separate universe simulations. We checked that for halo masses above this range our numerical results become strongly affected by the noise originating from having very few such massive objects. On the other hand, the integrand is negligible for mass scales below this range.
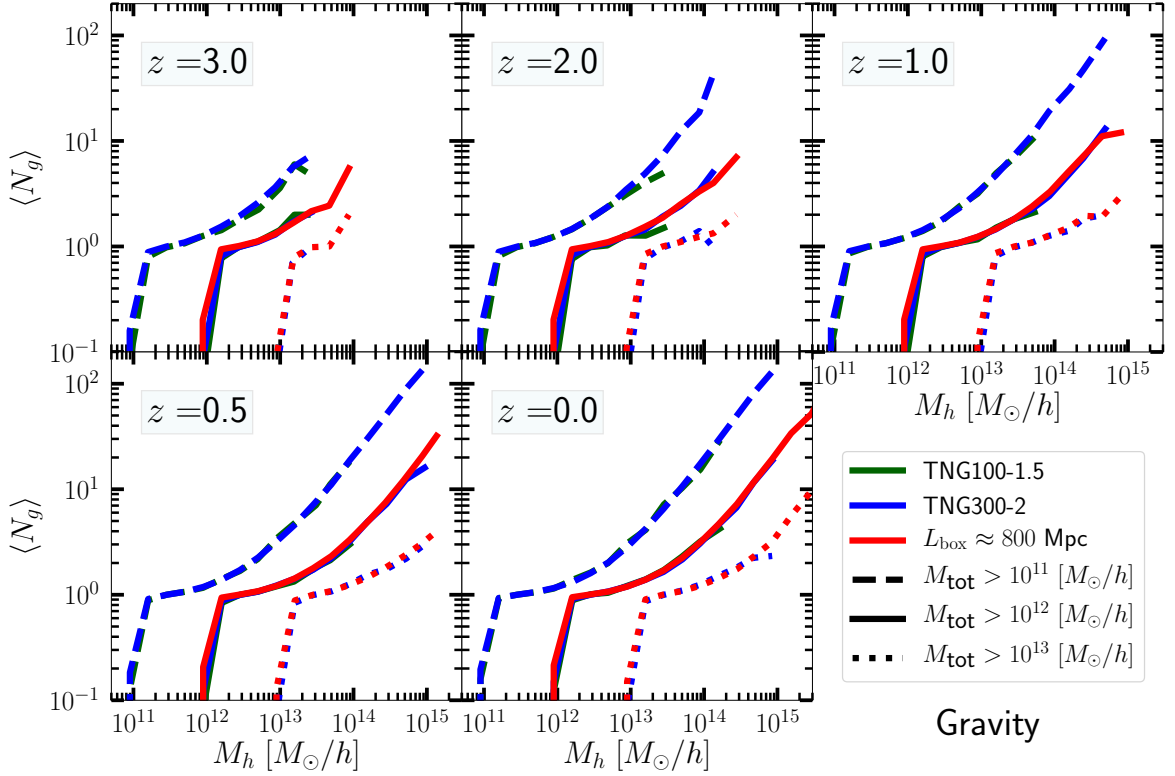
Figure 3. Mean HOD number $\langle N_g(M_h) \rangle$ for three subhalo populations selected by their total halo mass $M_{\text{tot}} > 10^{11}$ $M_\odot/h$ (dashed), $M_{\text{tot}} > 10^{12}$ $M_\odot/h$ (solid) and $M_{\text{tot}} > 10^{13}$ $M_\odot/h$ (dotted). The different panels are for different redshifts, and the different colors are for the three numerical resolutions of the Gravity simulations, as labeled (the different resolutions agree so well that some of the lines almost completely cover the others). The $M_{\text{tot}} > 10^{11}$ $M_\odot/h$ case is not shown for the $L_{\text{box}} \approx 800$Mpc resolution since these objects are poorly resolved. The $M_{\text{tot}} > 10^{13}$ $M_\odot/h$ is not shown for the TNG100-1.5 resolution since there are very few objects with these higher masses.

Figure 5 shows a few illustrative cases of the HOD response functions to better understand the result shown in figure 4 (we always limit ourselves to showing only a few illustrative examples of HOD responses for brevity). In particular, the figure shows the HOD responses measured from the TNG100-1.5 and TNG300-2 resolutions for the subhalo mass sample $M_{\text{tot}} > 10^{12}$ $M_\odot/h$ and for $\mathcal{O} = \delta_m$ at $z = 0$ (left), $\mathcal{O} = \sigma$ at $z = 3$ (center) and $\mathcal{O} = f_{\text{NL}}\phi$ at $z = 2$ (right). The vertical shaded bands mark the range in halo mass that contributes the most to the integral of eq. (3.2), and therefore, where any departures of $R_\mathcal{O}^g$ from zero have the strongest impact. Specifically, the integrand of eq. (3.2) is weighted by

$$W(M_h) = n_h(M_h)\langle N_g(M_h)\rangle, \tag{3.3}$$

which is negligible at both low M_h (as $\langle N_g(M_h) \rangle \to 0$) and high M_h (as n_h is exponentially suppressed). The width of the bands mark the range over which $W(M_h)$ is higher than 0.7 of its maximum value (note that these band widths are merely indicative and the integral is naturally also sensitive to mass scales just outside of it). Indeed, at $z = 0$, the R_1^g measured from both resolutions are compatible with zero near the vertical bands, hence the similarity between the dashed and solid lines at $z = 0$ in the middle left panel of figure 4. On the other hand, at $z = 2$, the R_ϕ^g measured from the TNG100-1.5 resolution appears larger than zero,
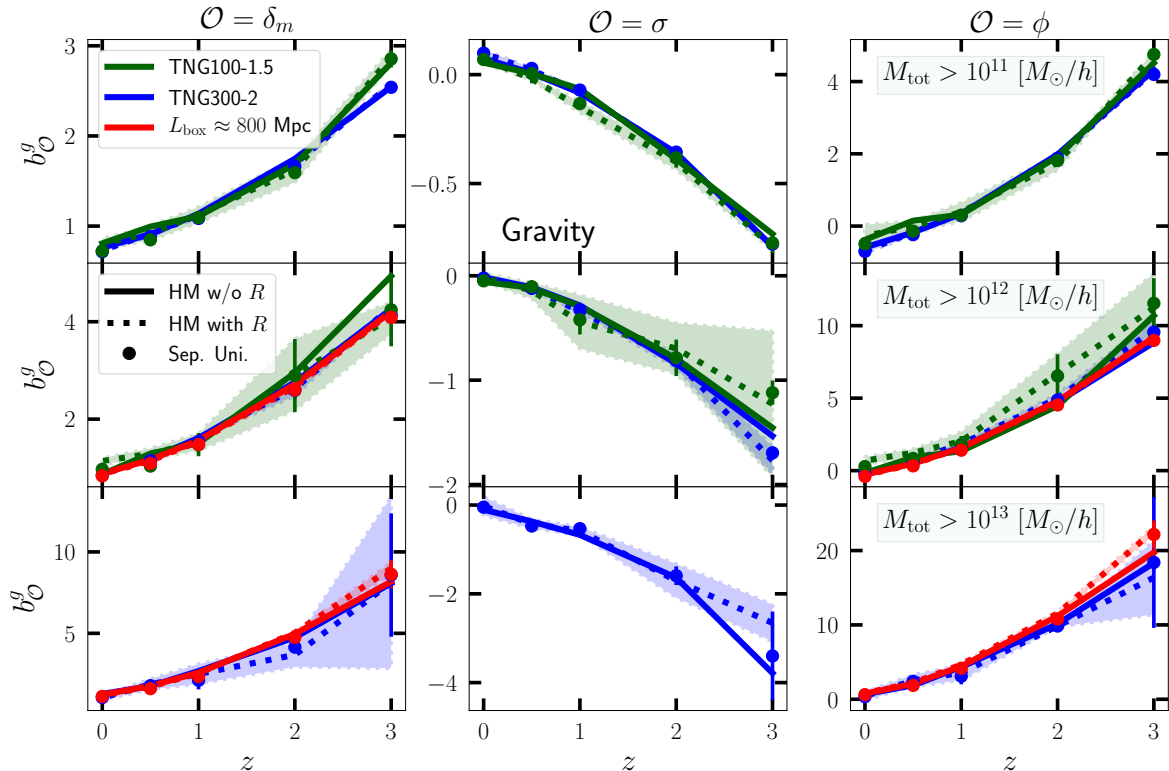
Figure 4. Galaxy bias parameters of the same total-mass-selected subhalo populations shown in figure 3 for the Gravity simulations. The left, center and right panels show the result for b_1^g, b_σ^g and b_ϕ^g, respectively. The three rows are for each of the three subhalo populations, as labeled. In each panel, the symbols with error bars show the $b_\mathcal{O}^g$ measured directly from the separate universe simulations, and the dashed and solid lines show the HM prediction of eq. (3.2) with and without the HOD number responses $R_\mathcal{O}^g$ taken into account.

which explains the upwards shift at $z = 2$ in the middle right panel of figure 4. Similarly, at $z = 3$, the R_σ^g of the TNG100-1.5 resolution also appears larger than zero, hence the upwards shift of the dashed line at $z = 3$ in the middle center panel of figure 4. These small effects are however likely caused by numerical noise in the measurements of $R_\mathcal{O}^g$, as illustrated by the fact that the shaded bands in figure 4 almost always enclose the solid lines.

Overall, the results in this subsection illustrate that the HOD numbers of total-mass selected subhalos do not respond strongly to any of the three types of long-wavelength perturbations we consider in this paper. We have shown this here explicitly for the case of the Gravity simulations, but we have confirmed (not shown) that the same conclusion holds in the case of the Hydro simulations for galaxies selected by their total mass. In the subsections below, the situation becomes more interesting as the responses $R_\mathcal{O}^g$ become larger when the tracers are selected by properties beyond total mass.

3.2 Galaxies selected by stellar mass

We turn our attention now to the case of galaxies selected by their stellar mass. This is a more observationally relevant exercise since the total stellar mass of galaxies can be estimated from observations more robustly than their total mass. Figure 6 shows the HOD numbers $\langle N_g(M_h) \rangle$ for three minimum stellar mass cuts at different redshifts for the TNG100-1.5 and
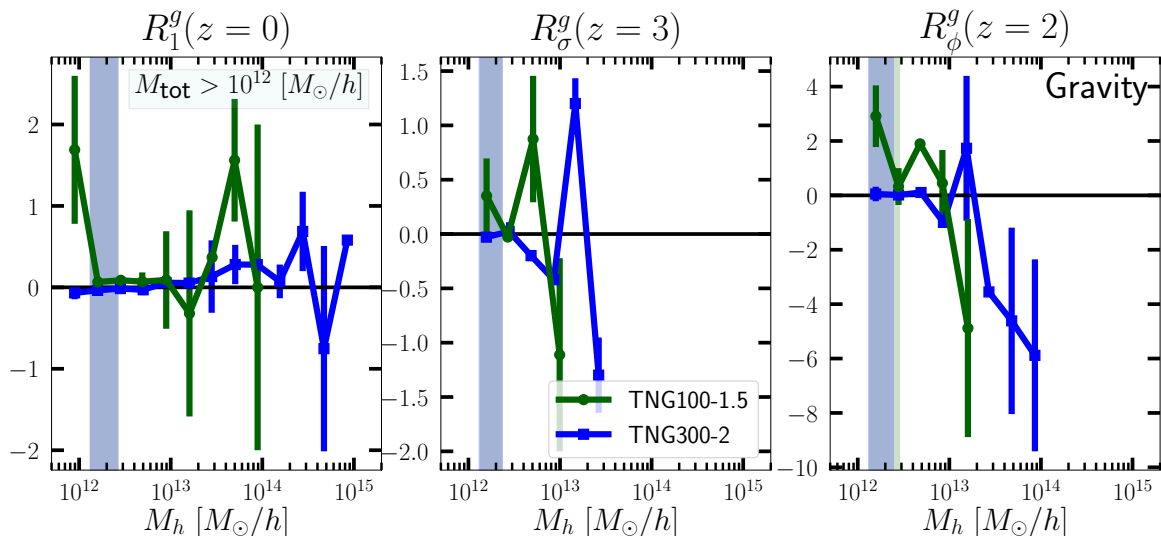
Figure 5. Example HOD number responses measured for the subhalo population with $M_{\rm tot} > 10^{12}~M_\odot/h$ from the Gravity simulations. The different panels are for different perturbations $\mathcal{O}$ and redshifts, and the different colors for the TNG100-1.5 and TNG300-2 resolutions, as labeled. The vertical bands cover the mass range in which the weighting function of eq. (3.3) is larger than 0.7 its maximum value; this roughly marks the range that contributes the most to the integral of eq. (3.2).

TNG300-2 resolutions of the Hydro simulations of the Fiducial cosmology, as labeled. The lower mass sample $M_* > 10^9~M_\odot/h$ is shown only for the TNG100-1.5 resolution since these objects contain $\mathcal{O}(10)$ star particles at TNG300-2 resolution and are not therefore as well resolved. In turn, we do not show the $M_* > 10^{11}~M_\odot/h$ case for the TNG100-1.5 resolution since the measurements are noisier as a result of having much fewer of these objects in this simulation box.

The shape and redshift evolution of the $\langle N_g(M_h)\rangle$ in figure 6 is in line with the expectation discussed already above in figure 3 for total mass selection. A main noteworthy difference here concerns the poorer agreement between the two resolutions: in figure 3, the HOD numbers of the different resolutions are nearly indistinguishable, whereas in figure 6 at fixed halo mass M_h, the TNG100-1.5 case displays a higher number of galaxies with $M_\star > 10^{10}~M_\odot/h$, compared to TNG300-2. This reflects the well known fact that numerical convergence is harder to achieve in hydrodynamical simulations of galaxy formation like IllustrisTNG. In particular, the higher numerical resolution of the TNG100-1.5 simulations makes star formation more efficient, which boosts the amplitude of the stellar mass function (and consequently the corresponding HOD numbers, as shown in figure 6); see also figure A.1 of ref. [121] for a comparison of the stellar mass function at different numerical resolutions in IllustrisTNG. As we will see next, however, the bias and HOD response results of the two resolutions agree very well with each other. This is because the bias and HOD responses are evaluated as ratios of quantities from simulations with the same resolution, which helps to mitigate numerical resolution effects. This agreement is nontrivial and it supports the numerical convergence of our results.

Figure 7 compares the galaxy bias parameters $b_\mathcal{O}^g$ measured for the stellar-mass selected samples with the corresponding HM predictions of eq. (3.2); the latter are again shown for the cases in which the measured $R_\mathcal{O}^g$ are taken into account (dashed) and when they are
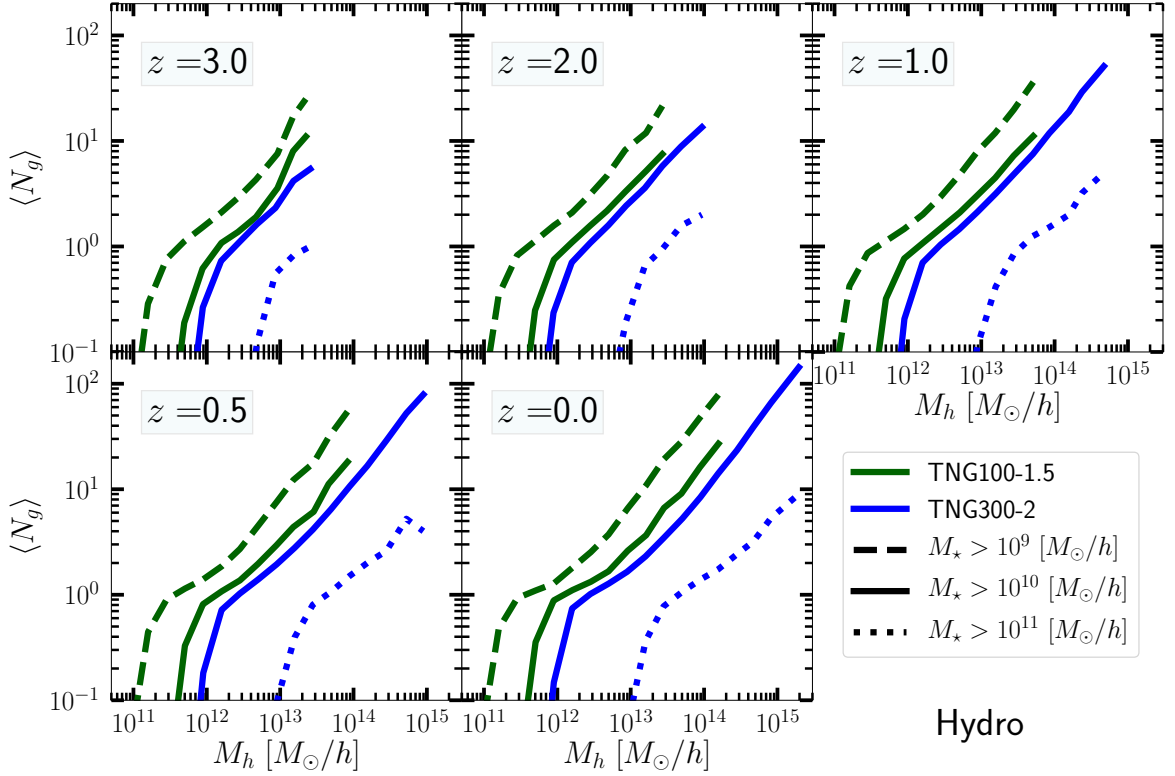
Figure 6. Mean HOD number $\langle N_g(M_h)\rangle$ for three galaxy populations selected by their stellar mass $M_* > 10^9 \ M_\odot/h$ (dashed), $M_* > 10^{10} \ M_\odot/h$ (solid) and $M_* > 10^{11} \ M_\odot/h$ (dotted), at different redshifts and numerical resolutions, as labeled. This is the same as figure 3, but for stellar mass selection in the Hydro simulations, instead of total mass selection in the Gravity simulations. We do not show the $M_* > 10^9 \ M_\odot/h$ case for the TNG300-2 resolution since this includes objects with very few star tracers and may not be well resolved. We also skip showing the $M_* > 10^{11} \ M_\odot/h$ case for the TNG100-1.5, since there are very few objects with these masses in this simulation box.

artificially set to zero (solid). Similarly to the case of total mass selection discussed in the previous subsection, the values of b_1^g under stellar mass selection do not appear dramatically affected by the HOD number response function R_1^g. This is seen by the comparable amplitude between the dashed and solid lines in the left panels of figure 7, with both agreeing well with the separate universe simulation measurements of b_1^g. The left panel in figure 8 shows R_1^g at $z = 0$ for the intermediate mass sample $M_* > 10^{10} \ M_\odot/h$, and it illustrates indeed how $\langle N_g(M_h)\rangle$ does not respond strongly to total matter density perturbations δ_m (the result is compatible with $R_1^g = 0$ within the precision of our measurements).

In contrast, the values of b_σ^g and b_ϕ^g of the stellar-mass selected objects are significantly more affected by the corresponding HOD number responses, as shown by the larger difference between the dashed and solid lines in the center and right panels of figure 7. The center and right panels in figure 8 show, as an example, R_σ^g at $z = 1$ and R_ϕ^g at $z = 3$, which are sizeable and vary between $[0.5 - 2]$ and $[1 - 6]$, respectively, within the halo mass range that contributes the most to eq. (3.2). Note that the impact of the HOD responses on b_σ^g and b_ϕ^g can be quite dramatic in figure 7: the values of b_ϕ^g become non-zero at low redshift, while the boost in b_σ^g can be as large as to result in a change of sign at low redshift.
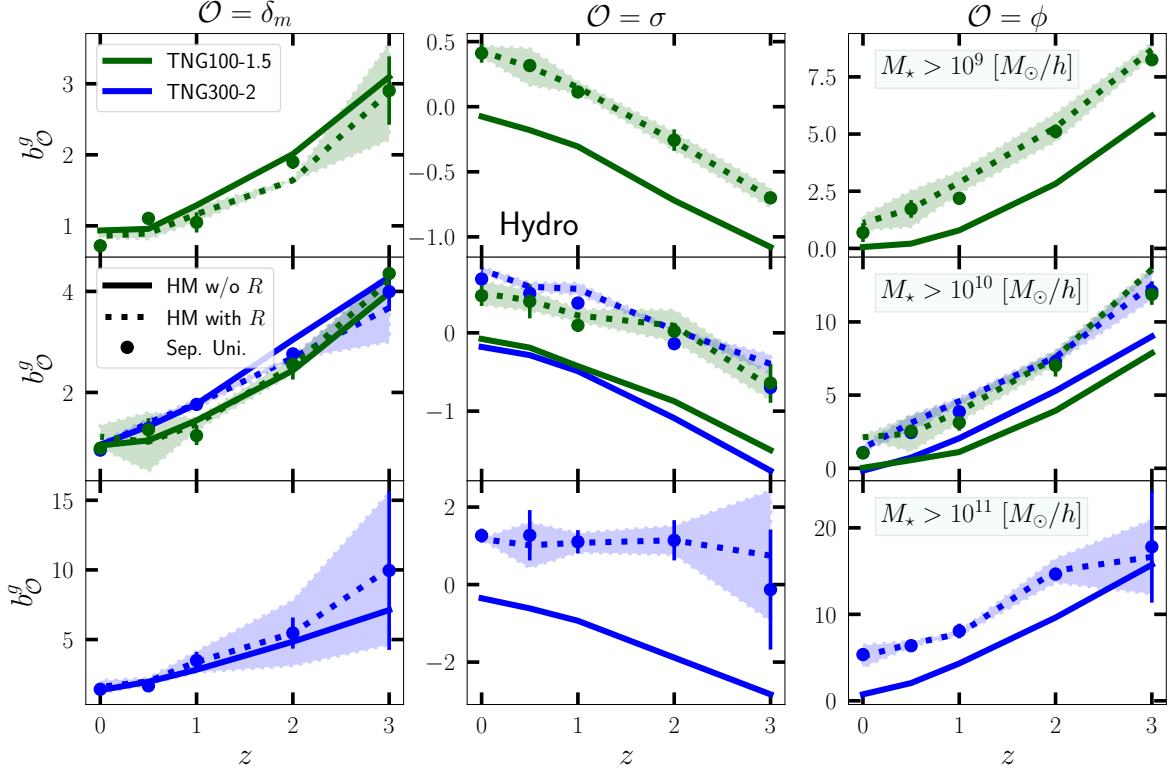
Figure 7. Galaxy bias parameters of the same stellar-mass-selected galaxy populations shown in figure 6. This is the same as figure 3, but for stellar mass selection in the Hydro simulations, instead of total mass selection in the Gravity simulations.

The reason why some $R_{\mathcal{O}}^g$ are sizeable for stellar-mass selected samples (and small for objects selected by their total mass) is interesting and can be traced back to the impact that the $\mathcal{O}$ perturbations have on the relation between the stellar mass and total mass of the galaxies, $M_*(M_{\text{tot}})$. Using the same simulations we use in this paper, ref. [81] showed that the additional baryons supplied by CIPs provide more fuel for star formation, which works to boost the amplitude of the $M_*(M_{\text{tot}})$ relation, i.e. there is more mass in stars at fixed total mass. Further, ref. [82], using also the same simulations as in here, showed that the $\mathcal{O} = f_{\text{NL}}\phi$ perturbations boost the star formation rate more than they boost the total mass accretion rate, and thus, they typically enhance M_* at fixed M_{tot}, as well. On the other hand, ref. [82] found that $\mathcal{O} = \delta_m$ perturbations boost both M_* and M_{tot} by roughly the same amount, thereby preserving the median stellar mass at fixed total mass. If $M_*(M_{\text{tot}})$ is higher in the separate universe cosmology than in the Fiducial, then the same minimum stellar mass cut $M_{*,\text{min}}$ in the two cosmologies corresponds to a lower minimum total mass cut $M_{\text{tot,min}}$ in the separate universe cosmology. Since galaxies with lower total masses are more abundant, that explains the positive response of R_{σ}^g and R_{ϕ}^g that is manifested in figures 7 and 8.

To make this point more specifically, recall from eq. (3.2) that the HOD number responses are defined as the logarithmic derivative of the HOD number w.r.t. the perturbations $\mathcal{O}$

$$R_{\mathcal{O}}^g(M_* > M_{*,\text{min}}|M_h) = \frac{\partial \ln\langle N_g(M_* > M_{*,\text{min}}|M_h)\rangle}{\partial \mathcal{O}}, \tag{3.4}$$
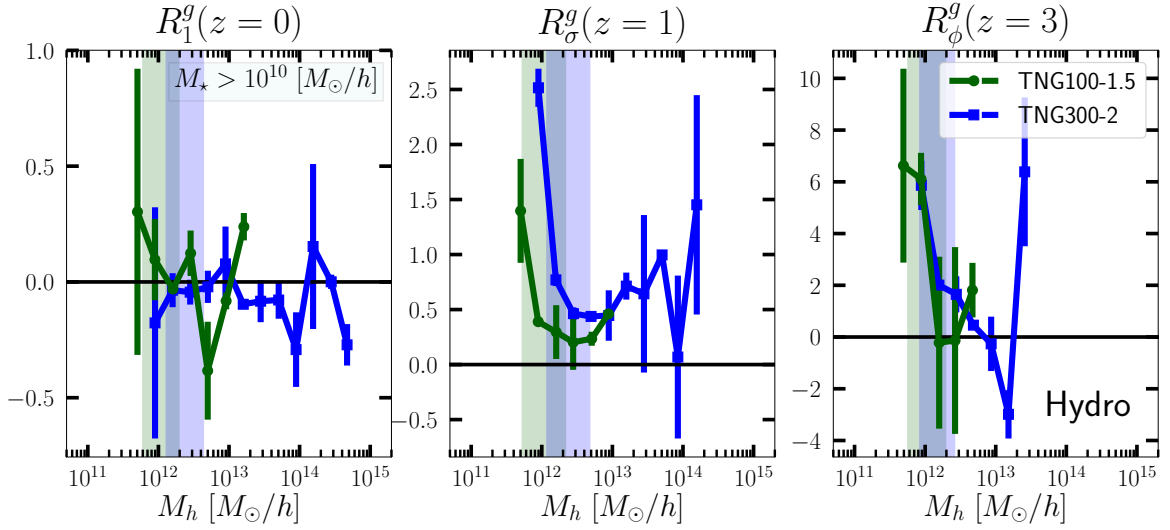
Figure 8. Example HOD number responses measured for the galaxy population with $M_* > 10^{10}\ M_\odot/h$. This is the same as figure 5, but for stellar-mass selected galaxies (and different illustrative $R_\mathcal{O}$ at different z), instead of total mass selected objects.

where we are specializing to the case of galaxy samples selected by a minimum stellar mass cut. The corresponding HOD number can be written as

$$\langle N_g(M_* > M_{*,\min}|M_h)\rangle = \int_{M_{*,\min}}^{\infty} \mathrm{d}M_* \frac{\mathrm{d}\langle N_g(M_*|M_h)\rangle}{\mathrm{d}M_*}$$
$$= \int_{M_{*,\min}}^{\infty} \mathrm{d}M_* \int_{0}^{\infty} \mathrm{d}M_{\mathrm{tot}} \frac{\mathrm{d}\langle N_g(M_{\mathrm{tot}}|M_h)\rangle}{\mathrm{d}M_{\mathrm{tot}}} \mathcal{P}(M_*, M_{\mathrm{tot}}), \quad (3.5)$$

where $\mathrm{d}\langle N_g(M_*|M_h)\rangle/\mathrm{d}M_*$ denotes the number of galaxies in the stellar mass bin $[M_*, M_* + \mathrm{d}M_*]$ that reside in haloes with mass M_h, $\mathrm{d}\langle N_g(M_{\mathrm{tot}}|M_h)\rangle/\mathrm{d}M_{\mathrm{tot}}$ represents the same but for the total galaxy mass bin $[M_{\mathrm{tot}}, M_{\mathrm{tot}} + \mathrm{d}M_{\mathrm{tot}}]$, and $\mathcal{P}(M_*, M_{\mathrm{tot}})$ denotes the probability of a galaxy with total mass M_{tot} to have stellar mass M_*. Using eqs. (3.5) and (3.4), it is straightforward to demonstrate that even if the total-mass selected HOD numbers are non-responsive, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}\mathcal{O}}\left(\frac{\mathrm{d}\langle N_g(M_{\mathrm{tot}}|M_h)\rangle}{\mathrm{d}M_{\mathrm{tot}}}\right) = 0, \quad (3.6)$$

(which the previous subsection suggests is approximately the case) it is still possible to have non-zero responses $R_\mathcal{O}^g$ for the stellar-mass selected samples if the stellar-to-total-mass relation of the galaxies *responds*, i.e., if

$$\frac{\partial \mathcal{P}(M_*, M_{\mathrm{tot}})}{\partial \mathcal{O}} \neq 0. \quad (3.7)$$

Thus, the results from refs. [81, 82] that $\partial \mathcal{P}(M_*, M_{\mathrm{tot}})/\partial \mathcal{O} > 0$ for $\mathcal{O} = f_{\mathrm{NL}}\phi$ and $\mathcal{O} = \sigma$, i.e., galaxies contain more mass in stars at fixed total mass, imply positive HOD number responses, which in turn causes the galaxy bias parameters to become larger, as indeed shown in figure 7. Further, the finding of ref. [82] that $\partial \mathcal{P}(M_*, M_{\mathrm{tot}})/\partial \delta_m \approx 0$ explains why $R_1^g \approx 0$ also for stellar-mass selected galaxies.

In ref. [54], using also stellar-mass selected galaxies in IllustrisTNG, the authors detect a dependence of the HOD numbers on the environment of the haloes; concretely, lower mass haloes are more likely to host a central galaxy in denser environments. Further, the authors find also that the $M_*(M_{\rm tot})$ relation is slightly boosted in denser halo environments. The apparent contradiction of these results with our finding of a negligible impact of $\mathcal{O} = \delta_m$ on $\langle N_g(M_h) \rangle$ and $M_*(M_{\rm tot})$ can be explained by the fact that in ref. [54] the environment is defined over much smaller scales than we do here. Specifically, ref. [54] defines the environment on scales of ≈ 5 Mpc$/h$, where the larger and nonlinear mass fluctuations can affect more strongly the number of mergers and the star formation rate, compared to the lower amplitude, linear fluctuations $\mathcal{O} = \delta_m \ll 1$ we consider here.[6]

The nonzero value of some of the HOD number responses under stellar-mass selection makes it interesting to ask how many more free parameters would be needed to describe the galaxy-halo connection in HOD studies. To roughly address this question, we fitted the $R_{\mathcal{O}}^g$ measured in the separate universe simulations with polynomials in $\log_{10} M_h$ of order n over the range $M_h \in \left[10^{10} ; 10^{14} \right] M_\odot / h$ weighted by $W(M_h)$. These were then plugged into eq. (3.2) to evaluate $b_{\mathcal{O}}^g$. The results are shown in figure 9 for $n = 0, 1, 2$ (colored dashed curves) for the TNG300-2 resolution. Interestingly, the simplest M_h-independent fit ($n = 0$; i.e., one extra parameter describing the amplitude of $R_{\mathcal{O}}^g$) performs remarkably well for all $b_{\mathcal{O}}^g$ measurements shown, significantly improving over the $R_{\mathcal{O}}^g = 0$ case for both $\mathcal{O} = f_{\rm NL}\phi$ and $\mathcal{O} = \sigma$. This is not entirely surprising since only a limited range in M_h contributes strongly to the HM integral over which $R_{\mathcal{O}}^g = $ constant has better chances of being a decent approximation. Naturally, this is just a simple exercise and more work is needed to design appropriate parametrizations of the $R_{\mathcal{O}}^g$. Nonetheless, the result in figure 9 does suggest that parametrizations of the $R_{\mathcal{O}}^g$ can be incorporated in HOD studies without drastically inflating the dimensionality of the parameter space of the galaxy-halo connection. This simple result adds on to efforts in the literature that aim to include environmental dependencies in HOD studies [51, 55, 55, 56, 59, 65, 139–143].

3.3 Different gas elements

In this subsection, we go beyond the case of galaxies as tracers and consider the bias of the gas distribution. This is interesting observationally given the growing interest in using line-intensity mapping observations to constrain cosmology [144], which is a goal that is expected to be realized with the data from upcoming surveys such as HETDEX [145], CHIME [146], HIRAX [147] and SKA [148]. Rather than resolving the light emitted by individual galaxies, line-intensity mapping surveys target the integrated emission from transition lines of atoms and molecules in all galaxies and diffuse gas in the intergalactic medium. The most popular of these lines is perhaps the 21cm spin-flip transition of neutral Hydrogen [149–152], but some attention is being devoted also to other emission lines from carbon monoxide (CO), ionized carbon ([CII]), oxygen (O) and Lyα [153].

The bias parameters of the gas $b_{\mathcal{O}}^{\rm gas}$ are defined analogously to as in eq. (2.2), but with the gas density $\rho_{\rm gas}$ instead of the galaxy number density n_g. Further, for the case of the gas distribution, instead of a mean HOD number, the relevant *halo occupancy quantity* is the mean mass of the gas inside halos with mass M_h, $\langle M_{\rm gas}(M_h) \rangle$. Its response functions are defined similarly to as in eq. (2.3), and we label them as $R_{\mathcal{O}}^{\rm gas}$. The gas bias expression in

---

[6]We stress also that although our results are consistent with $R_1^g \approx 0$, there is still room within the noise for the response to be nonzero. Additional realizations of separate universe simulations would be needed to beat down the statistical error and determine R_1^g more precisely.
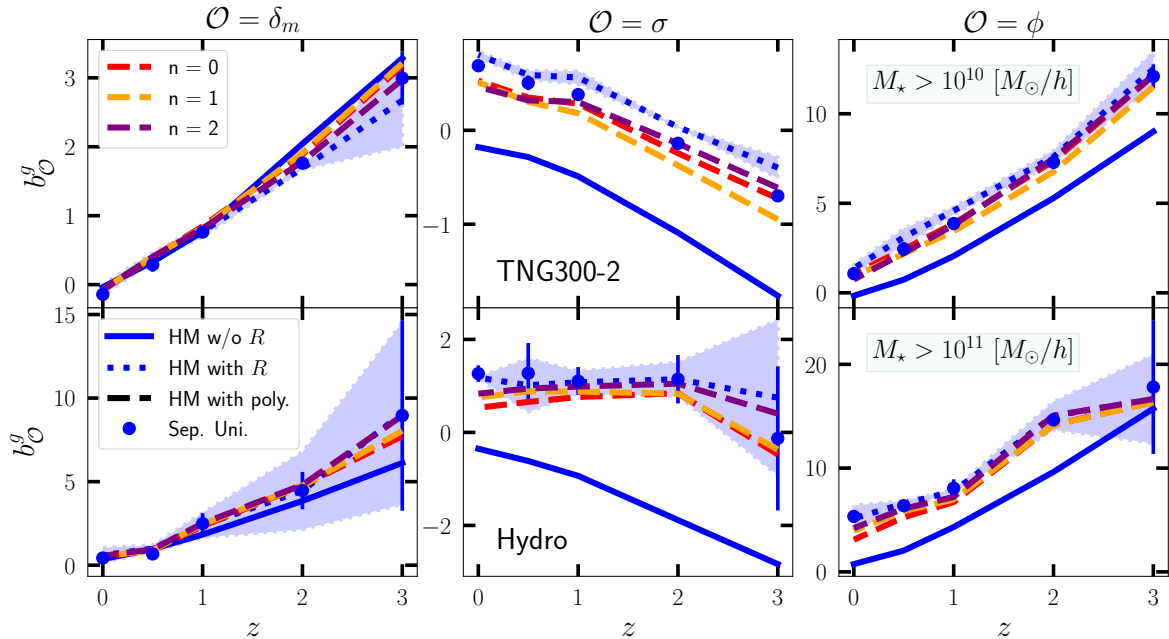
Figure 9. Same as figure 7 for the TNG300-2 resolution (the blue data points and lines are the same as in figure 7), but including also the HM model prediction of eq. (3.2) with the measured $R_{\mathcal{O}}^{\mathrm{g}}$ replaced by polynomial fits of order $n = 0, 1, 2$ (colored dashed lines), as labeled.

the HM is given by a simple generalization of eq. (3.2):

$$b_{\mathcal{O}}^{\mathrm{gas}} = \frac{1}{\rho_{\mathrm{gas}}} \int \mathrm{d}M_h \, n_h(M_h) \langle M_{\mathrm{gas}}(M_h, z) \rangle \Big[b_{\mathcal{O}}^h(M_h) + R_{\mathcal{O}}^{\mathrm{gas}}(M_h) \Big]. \tag{3.8}$$

Again, this equation is routinely used in forecast studies with $R_{\mathcal{O}}^{\mathrm{gas}}(M_h) = 0$. In this subsection we wish to determine the impact of this approximation.

We do not attempt a detailed modelling of the relevant ionized and neutral gas phases in our simulations (see e.g. refs. [48, 154–156]). Rather, we consider simply a few of the gas elements as followed by the IllustrisTNG model to build intuition about $R_{\mathcal{O}}^{\mathrm{gas}}$, and leave a more detailed modelling to subsequent work. Specifically, the example gas elements we look into here are hydrogen (H), carbon (C) and oxygen (O); the corresponding $\langle M_{\mathrm{gas}}(M_h) \rangle$ are shown in figure 10. The result is in line with the expectation that $\langle M_{\mathrm{gas}}(M_h) \rangle$ is a growing function of halo mass. For the case of H (which is present at the initial conditions), this simply reflects the fact that more massive haloes accrete in general more mass (including that in the form of H). The elements C and O are released to the interstellar medium by the stellar evolution and chemical enrichment models, and so the larger mass in C and O in massive haloes follows naturally from these objects containing also more stars. This explains also why the results from the two resolutions agree well for H, but less so for C and O, whose abundance is more directly linked to star formation efficiency.

The gas bias parameters and a few illustrative responses $R_{\mathcal{O}}^{\mathrm{gas}}$ are shown in figures 11 and 12, respectively. The main takeaway points from these figures are similar to those from figures 7 and 8. In particular, the response R_1^{gas} has an appreciably smaller impact on the gas bias parameter b_1^{gas}, compared to the impact of $R_\sigma^{\mathrm{gas}}, R_\phi^{\mathrm{gas}}$ on $b_\sigma^{\mathrm{gas}}, b_\phi^{\mathrm{gas}}$. The origin behind these results is also similar to that discussed in the previous subsection for stellar-
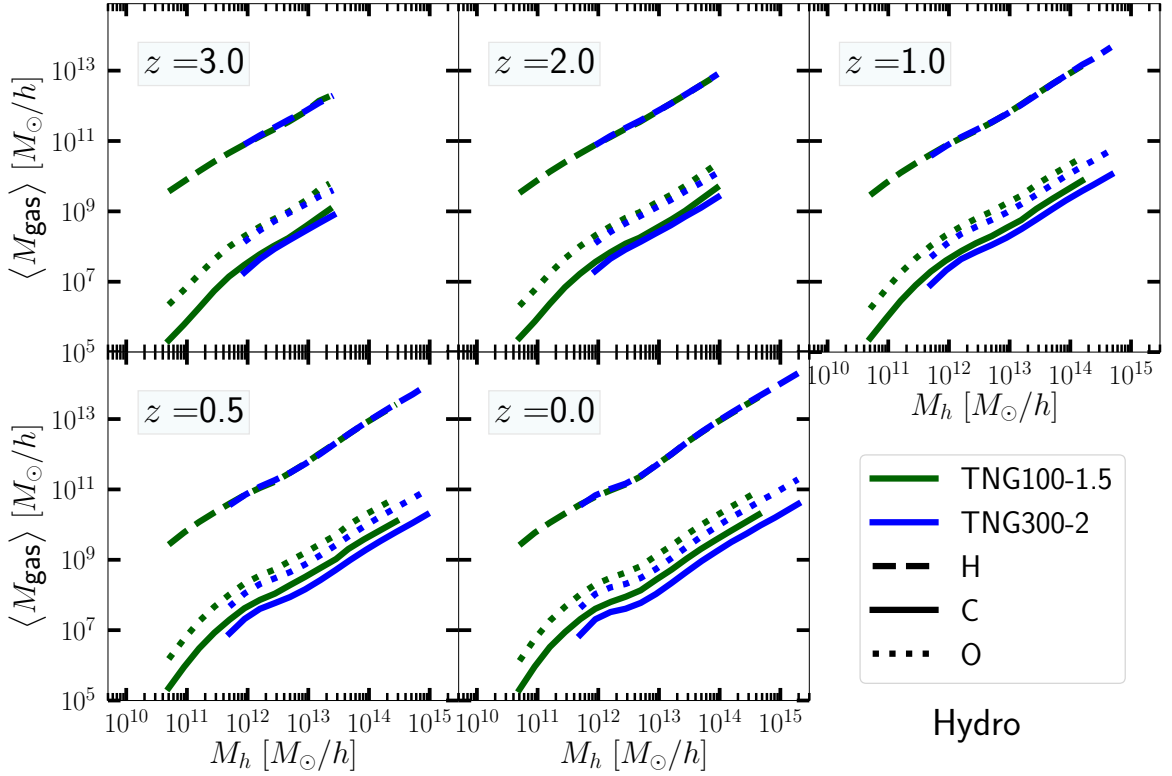
Figure 10. Mean halo gas mass as a function of total halo mass, $\langle M_{\mathrm{gas}}(M_h)\rangle$. The result is shown for hydrogen (H), carbon (C) and oxygen (O), at different redshifts and for the two resolutions TNG100-1.5 and TNG300-2, as labeled.

mass selected galaxies. For example, for the case of the total matter perturbations $\mathcal{O} = \delta_m$, the boost in baryon abundance and star formation inside haloes is roughly matched by the boost in the mass of the haloes, which makes $\langle M_{\mathrm{gas}}(M_h, z)\rangle$ weakly responsive (cf. the R_1^{C} result in the left panel of figure 12, which is compatible with zero). On the other hand, given a positive CIP $\mathcal{O} = \sigma$, H becomes more abundant effectively by definition of a CIP (more baryons at fixed total mass), and the additional fuel supplied for star formation also subsequently increases the abundance of C and O. This makes the halo gas mass responses R_σ^{gas} positive (the center panel of figure 12 shows R_σ^{C} at $z = 1$ as an example), which shifts the gas bias parameters b_σ^{gas} upwards in the center panels of figure 11.

The case of b_ϕ^{gas} displays a couple of interesting features. For C and O, at $z > 1$, the result is in line with the expectation that the increased star formation, increases the abundance of C and O, which shifts b_ϕ^{gas} upwards. At lower redshift, however, the solid and dashed lines on the C and O panels on the right of figure 12 approach one another, which is a consequence of a lowering of the amplitude of R_ϕ^{gas}. This can be contrasted with the b_ϕ^g panels of figure 7 for stellar-mass selected galaxies, in which the impact of the corresponding HOD response is approximately constant with redshift. Perhaps more interesting is the result for H, for which b_ϕ^{H} is shifted downwards at lower redshift, which indicates a negative H mean mass response $R_\phi^{\mathrm{H}} < 0$ (cf. right panel of figure 12). A tentative explanation here could be that the increased activity of baryonic processes such as feedback by active-galactic nuclei (AGN) could at least partially help remove some of the existing H and the excess of C and
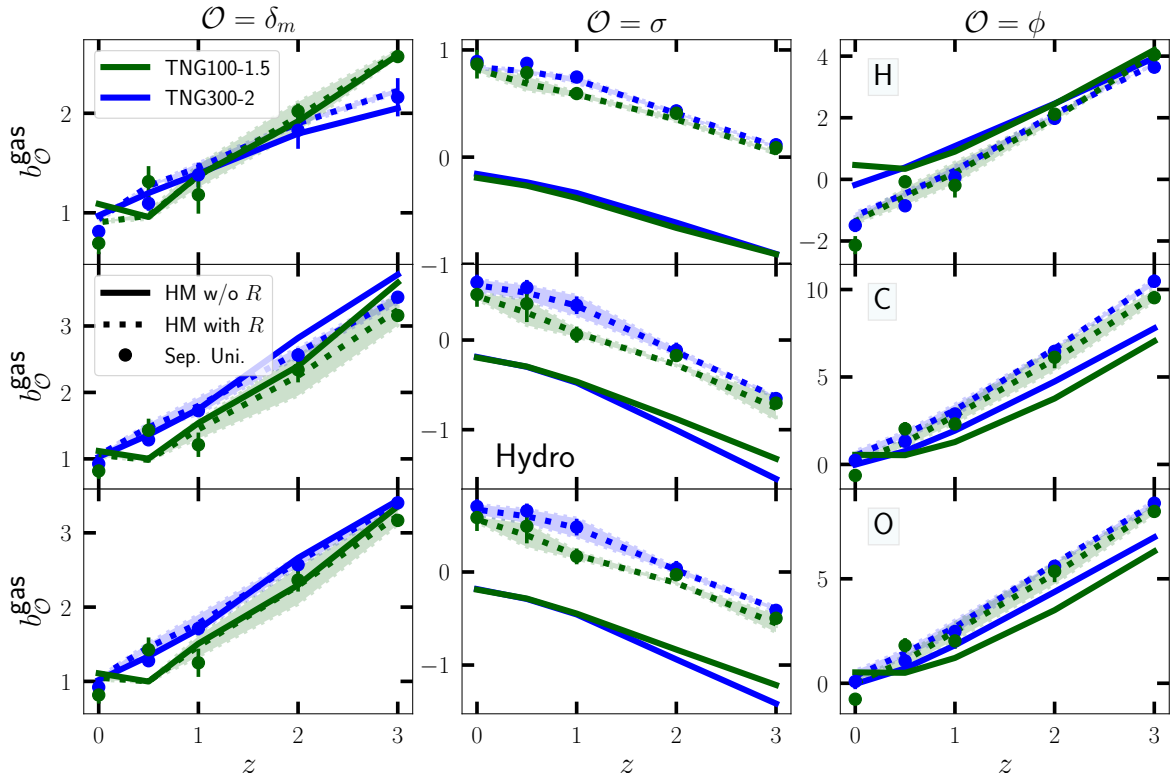
Figure 11. Bias parameters of the gas elements H, C and O shown in figure 10. This is the same as figures 4 and 7, but for gas as a LSS tracer, instead of galaxies selected by their total/stellar mass.

O produced by the enhanced star formation at higher redshift.[7] We leave a more detailed investigation of these features to future work, in which a more robust modelling of different gas phases can also be carried out. We anticipate that the richer the physics of the modelling of the gas phases (including the smaller-scale details of ionization), the richer the set of features that can be imprinted on $R_{\mathcal{O}}^{\mathrm{gas}}$, and consequently, on the corresponding gas bias parameters.

Our results in this subsection demonstrate overall that the response of the gas inside haloes can be sizeable and display a few interesting nontrivial features for $\mathcal{O} = \sigma$ and $\mathcal{O} = f_{\mathrm{NL}}\phi$ perturbations. We would like to emphasize that eq. (3.8) with $R_{\phi}^{\mathrm{gas}} = 0$ is routinely adopted in f_{NL} forecast studies (see e.g. refs. [144, 157–160] for recent examples). Our results here that $R_{\phi}^{\mathrm{gas}} \neq 0$ using the IllustrisTNG model therefore motivates investigations of the impact of $R_{\phi}^{\mathrm{gas}} \neq 0$ on f_{NL} forecasts. Concretely, the signature from f_{NL} on the gas power spectrum is $\propto b_1^{\mathrm{gas}} b_{\phi}^{\mathrm{gas}} f_{\mathrm{NL}}$, and hence, $R_{\phi}^{\mathrm{gas}} > 0$ would drive b_{ϕ}^{gas} upwards, which could potentially increase the signal-to-noise (in robust forecasts, however, uncertainties on the bias parameters should be parametrized and marginalized over [161, 162]).

---

[7]Positive CIPs also increase baryonic effects' activity, but the size of the effect may be weaker compared to the $\mathcal{O} = f_{\mathrm{NL}}\phi$ case and the larger gas fractions that are present since the initial conditions may make it harder to visualize any suppression of its abundance.
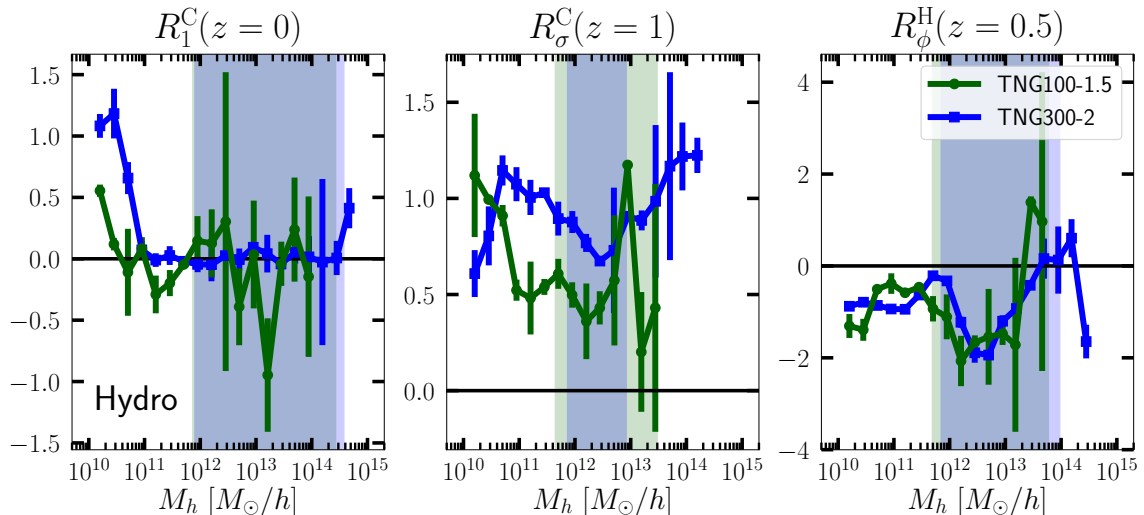
Figure 12. Example halo gas mass responses $R_{\mathcal{O}}^{\text{gas}}$. This is the same as figures 5 and 8, but for the gas in haloes (and different illustrative $R_{\mathcal{O}}$ at different z), instead of galaxies selected by their total/stellar mass.

4 Summary and conclusions

The combination of the HM with HODs is a powerful framework that is widely used to study the clustering of galaxies and construct fast galaxy mock catalogues. The HOD number $\langle N_g(M_h) \rangle$ describes the mean number of galaxies of a given type that reside inside haloes with mass M_h. In this paper, we studied the dependence of $\langle N_g(M_h) \rangle$ on the long-wavelength environment of the haloes as described by HOD response functions $R_{\mathcal{O}}^g$, where $\mathcal{O}$ denotes a type of long-wavelength perturbation. The $R_{\mathcal{O}}^g$ are the analog of the galaxy bias parameters $b_{\mathcal{O}}^g$, but applied to the HODs, instead of galaxy number densities; cf. eqs. (2.2) and (2.3). Despite being a natural ingredient in the HM, the HOD response functions $R_{\mathcal{O}}^g$ have remained largely ignored in the literature. This motivated us to measure them using galaxy formation simulations and study the corresponding impact in HM applications.

We focused on the impact of the HOD responses in the HM prediction of galaxy bias $b_{\mathcal{O}}^g$. The latter is given by a weighted version of the sum of halo bias $b_{\mathcal{O}}^h$ and the HOD responses $R_{\mathcal{O}}^g$, as in eq. (3.2), but in virtually all existing such calculations in the literature, the HOD responses are assumed to be zero, $R_{\mathcal{O}}^g = 0$. To test the validity of this approximation, we used separate universe simulations of the IllustrisTNG galaxy formation model to measure the HOD responses to three types of long-wavelength perturbations (cf. section 2): total matter perturbations, $\mathcal{O} = \delta_m$; primordial potential perturbations with local PNG, $\mathcal{O} = f_{\text{NL}}\phi$; and baryon-CDM CIPs, $\mathcal{O} = \sigma$. Our main results consisted in comparisons between the galaxy bias parameters $b_{\mathcal{O}}^g$ measured directly from the separate universe simulations with the HM galaxy bias prediction when the measured $R_{\mathcal{O}}^g$ are appropriately taken into account or artificially set to zero.

We have shown results from IllustrisTNG simulations run at two numerical resolutions (cf. table 1) and for galaxies/subhalos selected by their total mass (cf. section 3.1) and stellar mass (cf. section 3.2). Beyond galaxies as LSS tracers, we have also studied the impact of the responses of the mean gas mass in haloes $R_{\mathcal{O}}^{\text{gas}}$ on the corresponding gas bias parameters $b_{\mathcal{O}}^{\text{gas}}$ (cf. section 3.3; we considered H, C and O as example case studies).

Our main conclusions can be summarized as follows:

- For objects selected by total mass, the HOD number responses $R_{\mathcal{O}}^g$ are compatible with zero in our simulations and impact $b_{\mathcal{O}}^g$ only weakly (cf. figure 4).

- For stellar mass selected objects, setting $R_1^g = 0$ in the HM remains a good approximation and recovers the measured b_1^g well, but doing the same for R_σ^g and R_ϕ^g drastically underpredicts b_σ^g and b_ϕ^g (cf. figure 7). This can be explained physically by the boost that $\mathcal{O} = \sigma$ and $\mathcal{O} = f_{\mathrm{NL}}\phi$ induce on the stellar-to-total-mass relation of the galaxies (cf. discussion in section 3.2).

- Likewise for the gas distribution, b_1^{gas} remains weakly affected by the mean mass response R_1^{gas}, but b_σ^{gas} and b_ϕ^{gas} are affected strongly by R_σ^{gas} and R_ϕ^{gas} (cf. figure 11). In this case, the $R_{\mathcal{O}}^{\mathrm{gas}}$ can behave differently for different gas elements (e.g., $R_\phi^{\mathrm{C}}, R_\phi^{\mathrm{O}} > 0$, but $R_\phi^{\mathrm{H}} < 0$), which suggests a nontrivial interplay of physical processes that is worth studying further.

- Polynomial fits to $R_{\mathcal{O}}^g$ of order $n \leq 2$ recover well the measured $b_{\mathcal{O}}^g$ for stellar-mass selected samples (cf. figure 9). This suggests that parametrizations of $R_{\mathcal{O}}^g$ can be added to traditional HOD studies of the galaxy-halo connection without introducing too many free parameters.

Our results indicate overall that $R_{\mathcal{O}}^g = 0$ is an approximation that is not always valid and that it should be tested on a case-by-case basis. The sizeable effects we found on b_σ^g and b_ϕ^g find immediate applications in HM-based forecast studies of CIPs and local PNG: the signals in these studies are often $\propto b_\sigma^g, b_\phi^g$, and so misestimates of the bias parameters can lead to *biased* conclusions. Within the precision attained by our separate universe simulation measurements, the impact of the HOD responses to total matter perturbations remained consistent with zero. As argued in section 3.2, this is the expected result if δ_m perturbations do not modify significantly the stellar-to-total-mass relation of the galaxies. This seems to be at least approximately the case in the IllustrisTNG model (given our statistical errors), but may be different in other models of galaxy formation. It would therefore be interesting to measure R_1^g in other state-of-the-art galaxy formation models like EAGLE [163, 164], MAGNETICUM [165], BAHAMAS [166], or HORIZON-AGN [167].

We would like to emphasize the particularly welcoming aspect of the fourth bullet point above that indicates that parametrizations of $R_{\mathcal{O}}^g$ are straightforward to include in HOD studies. In figure 9, we have seen that $R_{\mathcal{O}}^g = $ constant provides a decent first order approximation, in which case only one additional parameter needs to be added to traditional HOD studies. It is important to note also that in this paper we have assumed that the HODs and their responses depend only on halo mass M_h, but it would be interesting to study $R_{\mathcal{O}}^g$ as a function of additional properties such as halo concentration, halo formation time, halo spin, etc.

Our analysis can be extended further to include a selection of tracers that resembles more closely the observations, including for example, an improved modelling of the gas phases that are relevant for future line-intensity mapping studies, or galaxies selected by properties such as star formation rate, color or AGN luminosity. With additional separate universe simulations, it would be also possible to study higher-order response functions (which are defined as higher-order derivatives w.r.t. $\mathcal{O}$) and/or response functions associated with additional perturbations such as large-scale tidal fields $\mathcal{O} = K_{ij}(\boldsymbol{x})K^{ij}(\boldsymbol{x})$ [168–175], higher-derivative operators $\mathcal{O} = \nabla^2 \delta_m(\boldsymbol{x})$ [176] or the ionizing radiation field [177, 178].

Acknowledgments

References

[1] A. Cooray and R.K. Sheth, *Halo Models of Large Scale Structure*, *Phys. Rept.* **372** (2002) 1 [astro-ph/0206508] [INSPIRE].

[2] S.G. Murray, B. Diemer and Z. Chen, *TheHaloMod: An online calculator for the halo model*, arXiv:2009.14066 [INSPIRE].

[3] A. Schneider, R. Teyssier, D. Potter, J. Stadel, J. Onions, D.S. Reed et al., *Matter power spectrum and the challenge of percent accuracy*, *JCAP* **04** (2016) 047 [arXiv:1503.05920] [INSPIRE].

[4] H.A. Winther et al., *Modified Gravity N-body Code Comparison Project*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 4208 [arXiv:1506.06384] [INSPIRE].

[5] J.F. Navarro, C.S. Frenk and S.D.M. White, *A Universal density profile from hierarchical clustering*, *Astrophys. J.* **490** (1997) 493 [astro-ph/9611107] [INSPIRE].

[6] J. Einasto, *On the Construction of a Composite Model for the Galaxy and on the Determination of the System of Galactic Parameters*, *Trudy Astrofizicheskogo Instituta Alma-Ata* **5** (1965) 87.

[7] R.K. Sheth and G. Tormen, *An Excursion Set Model of Hierarchical Clustering: Ellipsoidal Collapse and the Moving Barrier*, *Mon. Not. Roy. Astron. Soc.* **329** (2002) 61 [astro-ph/0105113] [INSPIRE].

[8] J.L. Tinker, A.V. Kravtsov, A. Klypin, K. Abazajian, M.S. Warren, G. Yepes et al., *Toward a halo mass function for precision cosmology: The Limits of universality*, *Astrophys. J.* **688** (2008) 709 [arXiv:0803.2706] [INSPIRE].

[9] J.L. Tinker, B.E. Robertson, A.V. Kravtsov, A. Klypin, M.S. Warren, G. Yepes et al., *The Large Scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests*, *Astrophys. J.* **724** (2010) 878 [arXiv:1001.3162] [INSPIRE].

[10] T. Lazeyras, C. Wagner, T. Baldauf and F. Schmidt, *Precision measurement of the local bias of dark matter halos*, *JCAP* **02** (2016) 018 [arXiv:1511.01096] [INSPIRE].

[11] G. Kauffmann, A. Nusser and M. Steinmetz, *Galaxy formation and large scale bias*, *Mon. Not. Roy. Astron. Soc.* **286** (1997) 795 [astro-ph/9512009] [INSPIRE].

[12] A.J. Benson, S. Cole, C.S. Frenk, C.M. Baugh and C.G. Lacey, *The Nature of galaxy bias and clustering*, *Mon. Not. Roy. Astron. Soc.* **311** (2000) 793 [astro-ph/9903343] [INSPIRE].

[13] U. Seljak, *Analytic model for galaxy and dark matter clustering*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 203 [astro-ph/0001493] [INSPIRE].

[14] J.A. Peacock and R.E. Smith, *Halo occupation numbers and galaxy bias*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 1144 [astro-ph/0005010] [INSPIRE].

[15] R. Scoccimarro, R.K. Sheth, L. Hui and B. Jain, *How many galaxies fit in a halo? Constraints on galaxy formation efficiency from spatial clustering*, *Astrophys. J.* **546** (2001) 20 [astro-ph/0006319] [INSPIRE].

[16] A.A. Berlind and D.H. Weinberg, *The Halo occupation distribution: Towards an empirical determination of the relation between galaxies and mass*, *Astrophys. J.* **575** (2002) 587 [astro-ph/0109001] [INSPIRE].

[17] A.A. Berlind, D.H. Weinberg, A.J. Benson, C.M. Baugh, S. Cole, R. Dave et al., *The Halo occupation distribution and the physics of galaxy formation*, *Astrophys. J.* **593** (2003) 1 [astro-ph/0212357] [INSPIRE].

[18] A.V. Kravtsov, A.A. Berlind, R.H. Wechsler, A.A. Klypin, S. Gottloeber, B. Allgood et al., *The Dark side of the halo occupation distribution*, *Astrophys. J.* **609** (2004) 35 [astro-ph/0308519] [INSPIRE].

[19] Z. Zheng, A.A. Berlind, D.H. Weinberg, A.J. Benson, C.M. Baugh, S. Cole et al., *Theoretical models of the halo occupation distribution: Separating central and satellite galaxies*, *Astrophys. J.* **633** (2005) 791 [astro-ph/0408564] [INSPIRE].

[20] A.R. Zentner, A.A. Berlind, J.S. Bullock, A.V. Kravtsov and R.H. Wechsler, *The Physics of galaxy clustering. 1. A Model for subhalo populations*, *Astrophys. J.* **624** (2005) 505 [astro-ph/0411586] [INSPIRE].

[21] A. Vale and J.P. Ostriker, *Linking halo mass to galaxy luminosity*, *Mon. Not. Roy. Astron. Soc.* **353** (2004) 189 [astro-ph/0402500] [INSPIRE].

[22] R. Voivodic, H. Rubira and M. Lima, *The Halo Void (Dust) Model of Large Scale Structure*, *JCAP* **10** (2020) 033 [arXiv:2003.06411] [INSPIRE].

[23] F. Schmidt, *Towards a self-consistent halo model for the nonlinear large-scale structure*, *Phys. Rev. D* **93** (2016) 063512 [arXiv:1511.02231] [INSPIRE].

[24] A. Mead, J. Peacock, C. Heymans, S. Joudaki and A. Heavens, *An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 1958 [arXiv:1505.07833] [INSPIRE].

[25] M. Cataneo, L. Lombriser, C. Heymans, A. Mead, A. Barreira, S. Bose et al., *On the road to percent accuracy: non-linear reaction of the matter power spectrum to dark energy and modified gravity*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 2121 [arXiv:1812.05594] [INSPIRE].

[26] A. Mead, S. Brieden, T. Tröster and C. Heymans, *HMcode-2020: Improved modelling of non-linear cosmological power spectra with baryonic feedback*, arXiv:2009.01858 [INSPIRE].

[27] L.A. Moustakas and R.S. Somerville, *The masses, ancestors and descendents of extremely red objects: constraints from spatial clustering*, *Astrophys. J.* **577** (2002) 1 [astro-ph/0110584] [INSPIRE].

[28] J.S. Bullock, R.H. Wechsler and R.S. Somerville, *Galaxy halo occupation at high redshift*, *Mon. Not. Roy. Astron. Soc.* **329** (2002) 246 [astro-ph/0106293] [INSPIRE].

[29] Z. Zheng, *Interpreting the observed clustering of red galaxies at Z ~ 3*, *Astrophys. J.* **610** (2004) 61 [astro-ph/0307030] [INSPIRE].

[30] SDSS collaboration, *The Luminosity and color dependence of the galaxy correlation function*, *Astrophys. J.* **630** (2005) 1 [astro-ph/0408569] [INSPIRE].

[31] C. Conroy, R.H. Wechsler and A.V. Kravtsov, *Modeling luminosity-dependent galaxy clustering through cosmic time*, *Astrophys. J.* **647** (2006) 201 [astro-ph/0512234] [INSPIRE].

[32] C. Blake, A. Collister and O. Lahav, *Halo-model signatures from 380,000 SDSS Luminous Red Galaxies with photometric redshifts*, *Mon. Not. Roy. Astron. Soc.* **385** (2008) 1257 [arXiv:0704.3377] [INSPIRE].

[33] D.A. Wake et al., *The 2dF-SDSS LRG and QSO survey: Evolution of the clustering of Luminous Red Galaxies since z = 0.6*, *Mon. Not. Roy. Astron. Soc.* **387** (2008) 1045 [arXiv:0802.4288] [INSPIRE].

[34] Z. Zheng, I. Zehavi, D.J. Eisenstein, D.H. Weinberg and Y. Jing, *Halo Occupation Distribution Modeling of Clustering of Luminous Red Galaxies*, *Astrophys. J.* **707** (2009) 554 [arXiv:0809.1868] [INSPIRE].

[35] SDSS collaboration, *Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity*, *Astrophys. J.* **736** (2011) 59 [arXiv:1005.2413] [INSPIRE].

[36] U. Sawangwit, T. Shanks, F.B. Abdalla, R.D. Cannon, S.M. Croom, A.C. Edge et al., *Angular correlation function of 1.5 million LRGs: clustering evolution and a search for BAO*, *Mon. Not. Roy. Astron. Soc.* **416** (2011) 3033 [arXiv:0912.0511] [INSPIRE].

[37] A. Leauthaud et al., *New constraints on the evolution of the stellar-to-dark matter connection: a combined analysis of galaxy-galaxy lensing, clustering, and stellar mass functions from z = 0.2 to z = 1*, *Astrophys. J.* **744** (2012) 159 [arXiv:1104.0928] [INSPIRE].

[38] F. Beutler, C. Blake, M. Colless, D.H. Jones, L. Staveley-Smith, L. Campbell et al., *The 6dF Galaxy Survey: Dependence of halo occupation on stellar mass*, *Mon. Not. Roy. Astron. Soc.* **429** (2013) 3604 [arXiv:1212.3610] [INSPIRE].

[39] E. Krause, C.M. Hirata, C. Martin, J.D. Neill and T.K. Wyder, *Halo Occupation Distribution Modeling of Green Valley Galaxies*, *Mon. Not. Roy. Astron. Soc.* **428** (2013) 2548 [arXiv:1208.6139] [INSPIRE].

[40] B.A. Reid, H.-J. Seo, A. Leauthaud, J.L. Tinker and M. White, *A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies*, *Mon. Not. Roy. Astron. Soc.* **444** (2014) 476 [arXiv:1404.3742] [INSPIRE].

[41] S. More, H. Miyatake, R. Mandelbaum, M. Takada, D. Spergel, J. Brownstein et al., *The Weak Lensing Signal and the Clustering of BOSS Galaxies II: Astrophysical and Cosmological Constraints*, *Astrophys. J.* **806** (2015) 2 [arXiv:1407.1856] [INSPIRE].

[42] Y. Zu and R. Mandelbaum, *Mapping stellar content to dark matter haloes using galaxy clustering and galaxy–galaxy lensing in the SDSS DR7*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 1161 [arXiv:1505.02781] [INSPIRE].

[43] J. Coupon et al., *The galaxy–halo connection from a joint lensing, clustering and abundance analysis in the CFHTLenS/VIPERS field*, *Mon. Not. Roy. Astron. Soc.* **449** (2015) 1352 [arXiv:1502.02867] [INSPIRE].

[44] R.A. Skibba et al., *Dark Matter Halo Models of Stellar Mass-Dependent Galaxy Clustering in PRIMUS+DEEP2 at 0.2 < z < 1.2*, *Astrophys. J.* **807** (2015) 152 [arXiv:1503.00731] [INSPIRE].

[45] A. Leauthaud et al., *Lensing is Low: Cosmology, Galaxy Formation, or New Physics?*, *Mon. Not. Roy. Astron. Soc.* **467** (2017) 3024 [arXiv:1611.08606] [INSPIRE].

[46] R.H. Wechsler and J.L. Tinker, *The Connection between Galaxies and their Dark Matter Halos*, *Ann. Rev. Astron. Astrophys.* **56** (2018) 435 [arXiv:1804.03097] [INSPIRE].

[47] H. Padmanabhan, A. Refregier and A. Amara, *A halo model for cosmological neutral hydrogen: abundances and clustering*, *Mon. Not. Roy. Astron. Soc.* **469** (2017) 2323 [arXiv:1611.06235] [INSPIRE].

[48] F. Villaescusa-Navarro et al., *Ingredients for 21 cm Intensity Mapping*, *Astrophys. J.* **866** (2018) 135 [arXiv:1804.09180] [INSPIRE].

[49] L. Wolz, S.G. Murray, C. Blake and J.S. Wyithe, *Intensity mapping cross-correlations II: HI halo models including shot noise*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 1007 [arXiv:1803.02477] [ɪɴSPIRE].

[50] K.T. Mehta, *Measuring the universe with high-precision large-scale structure*, Ph.D. thesis, The University of Arizona (2014).

[51] J.E. McEwen and D.H. Weinberg, *The effects of assembly bias on cosmological inference from galaxy-galaxy lensing and galaxy clusters*, arXiv:1601.02693 [ɪɴSPIRE].

[52] I. Zehavi, S. Contreras, N. Padilla, N.J. Smith, C.M. Baugh and P. Norberg, *The Impact of Assembly Bias on the Galaxy Content of Dark Matter Halos*, *Astrophys. J.* **853** (2018) 84 [arXiv:1706.07871] [ɪɴSPIRE].

[53] M.C. Artale, I. Zehavi, S. Contreras and P. Norberg, *The impact of assembly bias on the halo occupation in hydrodynamical simulations*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 3978 [arXiv:1805.06938].

[54] S. Bose, D.J. Eisenstein, L. Hernquist, A. Pillepich, D. Nelson, F. Marinacci et al., *Revealing the galaxy–halo connection in IllustrisTNG*, *Mon. Not. Roy. Astron. Soc.* **490** (2019) 5693 [arXiv:1905.08799] [ɪɴSPIRE].

[55] B. Hadzhiyska, S. Bose, D. Eisenstein, L. Hernquist and D.N. Spergel, *Limitations to the 'basic' HOD model and beyond*, *Mon. Not. Roy. Astron. Soc.* **493** (2020) 5506 [arXiv:1911.02610] [ɪɴSPIRE].

[56] B. Hadzhiyska, S. Bose, D. Eisenstein and L. Hernquist, *Extensions to models of the galaxy–halo connection*, *Mon. Not. Roy. Astron. Soc.* **501** (2021) 1603 [arXiv:2008.04913] [ɪɴSPIRE].

[57] B. Hadzhiyska, S. Tacchella, S. Bose and D.J. Eisenstein, *The galaxy–halo connection of emission-line galaxies in IllustrisTNG*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 3599 [arXiv:2011.05331] [ɪɴSPIRE].

[58] I.G. Alfaro, F. Rodriguez, A.N. Ruiz and D.G. Lambas, *How galaxies populate haloes in very low-density environments - An analysis of the halo occupation distribution in cosmic voids*, *Astron. Astrophys.* **638** (2020) A60 [arXiv:2003.06255] [ɪɴSPIRE].

[59] S. Yuan, B. Hadzhiyska, S. Bose, D.J. Eisenstein and H. Guo, *Evidence for galaxy assembly bias in BOSS CMASS redshift-space galaxy correlation function*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 3582 [arXiv:2010.04182] [ɪɴSPIRE].

[60] L. Gao, V. Springel and S.D.M. White, *The Age dependence of halo clustering*, *Mon. Not. Roy. Astron. Soc.* **363** (2005) L66 [astro-ph/0506510] [ɪɴSPIRE].

[61] L. Gao and S.D.M. White, *Assembly bias in the clustering of dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **377** (2007) L5 [astro-ph/0611921] [ɪɴSPIRE].

[62] R.H. Wechsler, A.R. Zentner, J.S. Bullock and A.V. Kravtsov, *The dependence of halo clustering on halo formation history, concentration, and occupation*, *Astrophys. J.* **652** (2006) 71 [astro-ph/0512416] [ɪɴSPIRE].

[63] N. Dalal, M. White, J.R. Bond and A. Shirokov, *Halo Assembly Bias in Hierarchical Structure Formation*, *Astrophys. J.* **687** (2008) 12 [arXiv:0803.3453] [ɪɴSPIRE].

[64] A.R. Zentner, A.P. Hearin and F.C. van den Bosch, *Galaxy Assembly Bias: A Significant Source of Systematic Error in the Galaxy–Halo Relationship*, *Mon. Not. Roy. Astron. Soc.* **443** (2014) 3044 [arXiv:1311.1818] [ɪɴSPIRE].

[65] A.P. Hearin, A.R. Zentner, F.C. van den Bosch, D. Campbell and E. Tollerud, *Introducing decorated HODs: modelling assembly bias in the galaxy–halo connection*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 2552 [arXiv:1512.03050] [ɪɴSPIRE].

[66] T. Lazeyras, M. Musso and F. Schmidt, *Large-scale assembly bias of dark matter halos*, *JCAP* **03** (2017) 059 [arXiv:1612.04360] [INSPIRE].

[67] J. Shi and R.K. Sheth, *Dependence of halo bias on mass and environment*, *Mon. Not. Roy. Astron. Soc.* **473** (2018) 2486 [arXiv:1707.04096] [INSPIRE].

[68] S. Contreras, I. Zehavi, N. Padilla, C. Baugh, E. Jiménez and I. Lacerna, *The Evolution of Assembly Bias*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 1133 [arXiv:1808.02896] [INSPIRE].

[69] A.D. Montero-Dorta, M.C. Artale, L.R. Abramo, B. Tucci, N. Padilla, G. Sato-Polito et al., *The manifestation of secondary bias on the galaxy population from IllustrisTNG300*, *Mon. Not. Roy. Astron. Soc.* **496** (2020) 1182 [arXiv:2001.01739] [INSPIRE].

[70] B. Tucci, A.D. Montero-Dorta, L.R. Abramo, G. Sato-Polito and M.C. Artale, *The physical origins of low-mass spin bias*, *Mon. Not. Roy. Astron. Soc.* **500** (2020) 2777 [arXiv:2007.10366] [INSPIRE].

[71] A.D. Montero-Dorta, M.C. Artale, L.R. Abramo and B. Tucci, *On the kinetic Sunyaev-Zel'dovich effect as an observational probe for halo spin bias*, arXiv:2008.08607 [INSPIRE].

[72] E. Jimenez, N. Padilla, S. Contreras, I. Zehavi, C. Baugh and A. Orsi, *The assembly bias of emission line galaxies*, arXiv:2010.08500 [INSPIRE].

[73] A.N. Salcedo, Y. Zu, Y. Zhang, H. Wang, X. Yang, Y. Wu et al., *Elucidating Galaxy Assembly Bias in SDSS*, arXiv:2010.04176 [INSPIRE].

[74] V. Desjacques, D. Jeong and F. Schmidt, *Large-Scale Galaxy Bias*, *Phys. Rept.* **733** (2018) 1 [arXiv:1611.09787] [INSPIRE].

[75] J.N. Fry and E. Gaztanaga, *Biasing and hierarchical statistics in large scale structure*, *Astrophys. J.* **413** (1993) 447 [astro-ph/9302009] [INSPIRE].

[76] N. Afshordi, Y.-S. Loh and M.A. Strauss, *Cross-correlation of the Cosmic Microwave Background with the 2MASS galaxy survey: Signatures of dark energy, hot gas, and point sources*, *Phys. Rev. D* **69** (2004) 083524 [astro-ph/0308260] [INSPIRE].

[77] A. Pillepich et al., *Simulating Galaxy Formation with the IllustrisTNG Model*, *Mon. Not. Roy. Astron. Soc.* **473** (2018) 4077 [arXiv:1703.02970] [INSPIRE].

[78] R. Weinberger, V. Springel, L. Hernquist, A. Pillepich, F. Marinacci, R. Pakmor et al., *Simulating galaxy formation with black hole driven thermal and kinetic feedback*, *Mont. Not. Roy. Astron. Soc.* **465** (2016) 3291 [arXiv:1607.03486].

[79] D. Nelson et al., *The IllustrisTNG Simulations: Public Data Release*, arXiv:1812.05609 [INSPIRE].

[80] A. Barreira, D. Nelson, A. Pillepich, V. Springel, F. Schmidt, R. Pakmor et al., *Separate Universe Simulations with IllustrisTNG: baryonic effects on power spectrum responses and higher-order statistics*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 2079 [arXiv:1904.02070] [INSPIRE].

[81] A. Barreira, G. Cabass, D. Nelson and F. Schmidt, *Baryon-CDM isocurvature galaxy bias with IllustrisTNG*, *JCAP* **02** (2020) 005 [arXiv:1907.04317] [INSPIRE].

[82] A. Barreira, G. Cabass, F. Schmidt, A. Pillepich and D. Nelson, *Galaxy bias and primordial non-Gaussianity: insights from galaxy formation simulations with IllustrisTNG*, *JCAP* **12** (2020) 013 [arXiv:2006.09368] [INSPIRE].

[83] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, *Phys. Rept.* **367** (2002) 1 [astro-ph/0112551] [INSPIRE].

[84] C. Wagner, F. Schmidt, C.-T. Chiang and E. Komatsu, *The angle-averaged squeezed limit of nonlinear matter N-point functions*, *JCAP* **08** (2015) 042 [arXiv:1503.03487] [INSPIRE].

[85] C.-T. Chiang, C. Wagner, F. Schmidt and E. Komatsu, *Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum*, *JCAP* **05** (2014) 048 [arXiv:1403.3411] [INSPIRE].

[86] C.-T. Chiang, C. Wagner, A.G. Sánchez, F. Schmidt and E. Komatsu, *Position-dependent correlation function from the SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 10 CMASS Sample*, *JCAP* **09** (2015) 028 [arXiv:1504.03322] [INSPIRE].

[87] K. Akitsu, M. Takada and Y. Li, *Large-scale tidal effect on redshift-space power spectrum in a finite-volume survey*, *Phys. Rev. D* **95** (2017) 083522 [arXiv:1611.04723] [INSPIRE].

[88] A. Barreira and F. Schmidt, *Responses in Large-Scale Structure*, *JCAP* **06** (2017) 053 [arXiv:1703.09212] [INSPIRE].

[89] A. Barreira and F. Schmidt, *Response Approach to the Matter Power Spectrum Covariance*, *JCAP* **11** (2017) 051 [arXiv:1705.01092] [INSPIRE].

[90] Y. Li, M. Schmittfull and U. Seljak, *Galaxy power-spectrum responses and redshift-space super-sample effect*, *JCAP* **02** (2018) 022 [arXiv:1711.00018] [INSPIRE].

[91] C. Gordon and A. Lewis, *Observational constraints on the curvaton model of inflation*, *Phys. Rev. D* **67** (2003) 123513 [astro-ph/0212248] [INSPIRE].

[92] C. Gordon and J.R. Pritchard, *Forecasted 21 cm constraints on compensated isocurvature perturbations*, *Phys. Rev. D* **80** (2009) 063535 [arXiv:0907.5400] [INSPIRE].

[93] G.P. Holder, K.M. Nollett and A. van Engelen, *On Possible Variation in the Cosmological Baryon Fraction*, *Astrophys. J.* **716** (2010) 907 [arXiv:0907.3919] [INSPIRE].

[94] D. Grin, O. Dore and M. Kamionkowski, *Compensated Isocurvature Perturbations and the Cosmic Microwave Background*, *Phys. Rev. D* **84** (2011) 123003 [arXiv:1107.5047] [INSPIRE].

[95] D. Grin, D. Hanson, G.P. Holder, O. Doré and M. Kamionkowski, *Baryons do trace dark matter 380,000 years after the big bang: Search for compensated isocurvature perturbations with WMAP 9-year data*, *Phys. Rev. D* **89** (2014) 023006 [arXiv:1306.4319] [INSPIRE].

[96] C.H. Heinrich, D. Grin and W. Hu, *Lensing Bias to CMB Measurements of Compensated Isocurvature Perturbations*, *Phys. Rev. D* **94** (2016) 043534 [arXiv:1605.08439] [INSPIRE].

[97] J.B. Muñoz, D. Grin, L. Dai, M. Kamionkowski and E.D. Kovetz, *Search for Compensated Isocurvature Perturbations with Planck Power Spectra*, *Phys. Rev. D* **93** (2016) 043008 [arXiv:1511.04441] [INSPIRE].

[98] T.L. Smith, J.B. Muñoz, R. Smith, K. Yee and D. Grin, *Baryons still trace dark matter: probing CMB lensing maps for hidden isocurvature*, *Phys. Rev. D* **96** (2017) 083508 [arXiv:1704.03461] [INSPIRE].

[99] M.T. Soumagnac, C.G. Sabiu, R. Barkana and J. Yoo, *Large scale distribution of mass versus light from Baryon Acoustic Oscillations: Measurement in the final SDSS-III BOSS Data Release 12*, arXiv:1802.10368 [INSPIRE].

[100] S.C. Hotinli, J.B. Mertens, M.C. Johnson and M. Kamionkowski, *Probing correlated compensated isocurvature perturbations using scale-dependent galaxy bias*, *Phys. Rev. D* **100** (2019) 103528 [arXiv:1908.08953] [INSPIRE].

[101] C. Heinrich and M. Schmittfull, *BAO Modulation as a Probe of Compensated Isocurvature Perturbations*, *Phys. Rev. D* **100** (2019) 063503 [arXiv:1904.00024] [INSPIRE].

[102] A. Barreira, G. Cabass, K.D. Lozanov and F. Schmidt, *Compensated Isocurvature Perturbations in the Galaxy Power Spectrum*, *JCAP* **07** (2020) 049 [arXiv:2002.12931] [INSPIRE].

[103] PLANCK collaboration, *Planck 2018 results. X. Constraints on inflation*, *Astron. Astrophys.* **641** (2020) A10 [arXiv:1807.06211] [INSPIRE].

[104] E. Komatsu and D.N. Spergel, *Acoustic signatures in the primary microwave background bispectrum*, *Phys. Rev. D* **63** (2001) 063002 [astro-ph/0005036] [INSPIRE].

[105] PLANCK collaboration, *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **641** (2020) A9 [arXiv:1905.05697] [INSPIRE].

[106] J.M. Maldacena, *Non-Gaussian features of primordial fluctuations in single field inflationary models*, *JHEP* **05** (2003) 013 [astro-ph/0210603] [INSPIRE].

[107] P. Creminelli and M. Zaldarriaga, *Single field consistency relation for the 3-point function*, *JCAP* **10** (2004) 006 [astro-ph/0407059] [INSPIRE].

[108] P. Creminelli, G. D'Amico, M. Musso and J. Norena, *The (not so) squeezed limit of the primordial 3-point function*, *JCAP* **11** (2011) 038 [arXiv:1106.1462] [INSPIRE].

[109] T. Tanaka and Y. Urakawa, *Dominance of gauge artifact in the consistency relation for the primordial bispectrum*, *JCAP* **05** (2011) 014 [arXiv:1103.1251] [INSPIRE].

[110] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, *Galaxy Bias and non-Linear Structure Formation in General Relativity*, *JCAP* **10** (2011) 031 [arXiv:1106.5507] [INSPIRE].

[111] E. Pajer, F. Schmidt and M. Zaldarriaga, *The Observed Squeezed Limit of Cosmological Three-Point Functions*, *Phys. Rev. D* **88** (2013) 083502 [arXiv:1305.0824] [INSPIRE].

[112] L. Dai, E. Pajer and F. Schmidt, *On Separate Universes*, *JCAP* **10** (2015) 059 [arXiv:1504.00351] [INSPIRE].

[113] R. de Putter, O. Doré and D. Green, *Is There Scale-Dependent Bias in Single-Field Inflation?*, *JCAP* **10** (2015) 024 [arXiv:1504.05935] [INSPIRE].

[114] N. Kaiser, *On the Spatial correlations of Abell clusters*, *Astrophys. J. Lett.* **284** (1984) L9 [INSPIRE].

[115] J.M. Bardeen, J.R. Bond, N. Kaiser and A.S. Szalay, *The Statistics of Peaks of Gaussian Random Fields*, *Astrophys. J.* **304** (1986) 15 [INSPIRE].

[116] V. Springel, *E pur si muove: Galiliean-invariant cosmological hydrodynamical simulations on a moving mesh*, *Mon. Not. Roy. Astron. Soc.* **401** (2010) 791 [arXiv:0901.4107] [INSPIRE].

[117] R. Pakmor, V. Springel, A. Bauer, P. Mocz, D.J. Muñoz, S.T. Ohlmann et al., *Improving the convergence properties of the moving-mesh code AREPO*, *Mon. Not. Roy. Astron. Soc.* **455** (2016) 1134 [arXiv:1503.00562] [INSPIRE].

[118] S. Genel, M. Vogelsberger, V. Springel, D. Sijacki, D. Nelson, G. Snyder et al., *Introducing the Illustris Project: the evolution of galaxy populations across cosmic time*, *Mon. Not. Roy. Astron. Soc.* **445** (2014) 175 [arXiv:1405.3749] [INSPIRE].

[119] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu et al., *Introducing the Illustris Project: Simulating the coevolution of dark and visible matter in the Universe*, *Mon. Not. Roy. Astron. Soc.* **444** (2014) 1518 [arXiv:1405.2921] [INSPIRE].

[120] F. Marinacci et al., *First results from the IllustrisTNG simulations: radio haloes and magnetic fields*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 5113 [arXiv:1707.03396] [INSPIRE].

[121] A. Pillepich et al., *First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 648 [arXiv:1707.03406] [INSPIRE].

[122] J.P. Naiman, A. Pillepich, V. Springel, E. Ramirez-Ruiz, P. Torrey, M. Vogelsberger et al., *First results from the IllustrisTNG simulations: a tale of two elements – chemical evolution of magnesium and europium*, *Mon. Not. Roy. Astron. Soc.* **477** (2018) 1206 [arXiv:1707.03401].

[123] V. Springel et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 676 [arXiv:1707.03397] [INSPIRE].

[124] D. Nelson et al., *First results from the IllustrisTNG simulations: the galaxy colour bimodality*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 624 [arXiv:1707.03395] [INSPIRE].

[125] V. Springel, *N-GenIC: Cosmological structure initial conditions*, Astrophysics source code library, 2015.

[126] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473 [astro-ph/9911177] [INSPIRE].

[127] V. Springel, S.D.M. White, G. Tormen and G. Kauffmann, *Populating a cluster of galaxies. 1. Results at $z = 0$*, *Mon. Not. Roy. Astron. Soc.* **328** (2001) 726 [astro-ph/0012055] [INSPIRE].

[128] C.-T. Chiang, W. Hu, Y. Li and M. Loverde, *Scale-dependent bias and bispectrum in neutrino separate universe simulations*, *Phys. Rev. D* **97** (2018) 123526 [arXiv:1710.01310] [INSPIRE].

[129] C.-T. Chiang, Y. Li, W. Hu and M. LoVerde, *Quintessential Scale Dependence from Separate Universe Simulations*, *Phys. Rev. D* **94** (2016) 123502 [arXiv:1609.01701] [INSPIRE].

[130] D. Jamieson and M. LoVerde, *Quintessential Isocurvature in Separate Universe Simulations*, *Phys. Rev. D* **100** (2019) 023516 [arXiv:1812.08765] [INSPIRE].

[131] Y. Li, W. Hu and M. Takada, *Super-Sample Covariance in Simulations*, *Phys. Rev. D* **89** (2014) 083519 [arXiv:1401.0385] [INSPIRE].

[132] Y. Li, W. Hu and M. Takada, *Super-Sample Signal*, *Phys. Rev. D* **90** (2014) 103530 [arXiv:1408.1081] [INSPIRE].

[133] C. Wagner, F. Schmidt, C.-T. Chiang and E. Komatsu, *Separate Universe Simulations*, *Mon. Not. Roy. Astron. Soc.* **448** (2015) L11 [arXiv:1409.6294] [INSPIRE].

[134] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, *Linear response to long wavelength fluctuations using curvature simulations*, *JCAP* **09** (2016) 007 [arXiv:1511.01465] [INSPIRE].

[135] Y. Li, W. Hu and M. Takada, *Separate Universe Consistency Relation and Calibration of Halo Bias*, *Phys. Rev. D* **93** (2016) 063507 [arXiv:1511.01454] [INSPIRE].

[136] H. Khoraminezhad, T. Lazeyras, R.E. Angulo, O. Hahn and M. Viela, *Quantifying the impact of baryon-CDM perturbations on halo clustering and baryon fraction*, *JCAP* **03** (2021) 023 [arXiv:2011.01037] [INSPIRE].

[137] N. Dalal, O. Dore, D. Huterer and A. Shirokov, *The imprints of primordial non-Gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects*, *Phys. Rev. D* **77** (2008) 123514 [arXiv:0710.4560] [INSPIRE].

[138] A. Slosar, C. Hirata, U. Seljak, S. Ho and N. Padmanabhan, *Constraints on local primordial non-Gaussianity from large scale structure*, *JCAP* **08** (2008) 031 [arXiv:0805.3580] [INSPIRE].

[139] H. Gil-Marin, R. Jimenez and L. Verde, *A Halo Model with Environment Dependence: Theoretical Considerations*, *Mon. Not. Roy. Astron. Soc.* **414** (2011) 1207 [arXiv:1008.4583] [INSPIRE].

[140] A. Paranjape, K. Kovac, W.G. Hartley and I. Pahwa, *Correlating galaxy colour and halo concentration: A tunable Halo Model of galactic conformity*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 3030 [arXiv:1503.08212] [INSPIRE].

[141] B.D. Wibking, A.N. Salcedo, D.H. Weinberg, L.H. Garrison, D. Ferrer, J. Tinker et al., *Emulating galaxy clustering and galaxy–galaxy lensing into the deeply non-linear regime: methodology, information, and forecasts*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 989 [arXiv:1709.07099] [INSPIRE].

[142] X. Xu, I. Zehavi and S. Contreras, *Dissecting and Modelling Galaxy Assembly Bias*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 3242 [`arXiv:2007.05545`] [ɪɴSPIRE].

[143] A.N. Salcedo, B.D. Wibking, D.H. Weinberg, H.-Y. Wu, D. Ferrer, D. Eisenstein et al., *Cosmology with stacked cluster weak lensing and cluster–galaxy cross-correlations*, *Mon. Not. Roy. Astron. Soc.* **491** (2020) 3061 [`arXiv:1906.06499`] [ɪɴSPIRE].

[144] J.L. Bernal, P.C. Breysse, H. Gil-Marín and E.D. Kovetz, *User's guide to extracting cosmological information from line-intensity maps*, *Phys. Rev. D* **100** (2019) 123522 [`arXiv:1907.10067`] [ɪɴSPIRE].

[145] G.J. Hill et al., *The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX): Description and Early Pilot Survey Results*, *ASP Conf. Ser.* **399** (2008) 115 [`arXiv:0806.0183`] [ɪɴSPIRE].

[146] K. Bandura et al., *Canadian Hydrogen Intensity Mapping Experiment (CHIME) Pathfinder*, *Proc. SPIE Int. Soc. Opt. Eng.* **9145** (2014) 22 [`arXiv:1406.2288`] [ɪɴSPIRE].

[147] L.B. Newburgh et al., *HIRAX: A Probe of Dark Energy and Radio Transients*, *Proc. SPIE Int. Soc. Opt. Eng.* **9906** (2016) 99065X [`arXiv:1607.02059`] [ɪɴSPIRE].

[148] SKA collaboration, *Cosmology with Phase 1 of the Square Kilometre Array: Red Book 2018: Technical specifications and performance forecasts*, *Publ. Astron. Soc. Austral.* **37** (2020) e007 [`arXiv:1811.02743`] [ɪɴSPIRE].

[149] J.R. Pritchard and A. Loeb, *21-cm cosmology*, *Rept. Prog. Phys.* **75** (2012) 086901 [`arXiv:1109.6012`] [ɪɴSPIRE].

[150] K.W. Masui et al., *Measurement of 21 cm brightness fluctuations at $z \sim 0.8$ in cross-correlation*, *Astrophys. J. Lett.* **763** (2013) L20 [`arXiv:1208.0331`] [ɪɴSPIRE].

[151] E.R. Switzer et al., *Determination of $z \sim 0.8$ neutral hydrogen fluctuations using the 21 cm intensity mapping auto-correlation*, *Mon. Not. Roy. Astron. Soc.* **434** (2013) L46 [`arXiv:1304.3712`] [ɪɴSPIRE].

[152] P. Bull, P.G. Ferreira, P. Patel and M.G. Santos, *Late-time cosmology with 21cm intensity mapping experiments*, *Astrophys. J.* **803** (2015) 21 [`arXiv:1405.1452`] [ɪɴSPIRE].

[153] J. Fonseca, M. Silva, M.G. Santos and A. Cooray, *Cosmology with intensity mapping techniques using atomic and molecular lines*, *Mon. Not. Roy. Astron. Soc.* **464** (2017) 1948 [`arXiv:1607.05288`] [ɪɴSPIRE].

[154] B. Diemer et al., *Modeling the atomic-to-molecular transition in cosmological simulations of galaxy formation*, *Astrophys. J. Suppl.* **238** (2018) 33 [`arXiv:1806.02341`] [ɪɴSPIRE].

[155] G. Popping, D. Narayanan, R.S. Somerville, A.L. Faisst and M.R. Krumholz, *The art of modelling CO, [C I], and [C II] in cosmological galaxy formation models*, *Mon. Not. Roy. Astron. Soc.* **482** (2018) 4906 [`arXiv:1805.11093`].

[156] A.R.H. Stevens, C. del P. Lagos, L. Cortese, B. Catinella, B. Diemer, D. Nelson et al., *Molecular hydrogen in IllustrisTNG galaxies: carefully comparing signatures of environment with local CO and SFR data*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 3158 [`arXiv:2011.03226`] [ɪɴSPIRE].

[157] A. Moradinezhad Dizgah and G.K. Keating, *Line intensity mapping with [CII] and CO(1-0) as probes of primordial non-Gaussianity*, *Astrophys. J.* **872** (2019) 126 [`arXiv:1810.02850`] [ɪɴSPIRE].

[158] D. Karagiannis, A. Slosar and M. Liguori, *Forecasts on Primordial non-Gaussianity from 21 cm Intensity Mapping experiments*, *JCAP* **11** (2020) 052 [`arXiv:1911.03964`] [ɪɴSPIRE].

[159] S. Camera and H. Padmanabhan, *Beyond ΛCDM with H i intensity mapping: robustness of cosmological constraints in the presence of astrophysics*, *Mon. Not. Roy. Astron. Soc.* **496** (2020) 4115 [`arXiv:1910.00022`] [ɪɴSPIRE].

[160] D. Karagiannis, J. Fonseca, R. Maartens and S. Camera, *Probing primordial non-Gaussianity with the power spectrum and bispectrum of future 21 cm intensity maps*, *Phys. Dark Univ.* **32** (2021) 100821 [arXiv:2010.07034] [InSPIRE].

[161] A. Barreira, *On the impact of galaxy bias uncertainties on primordial non-Gaussianity constraints*, *JCAP* **12** (2020) 031 [arXiv:2009.06622] [InSPIRE].

[162] A. Moradinezhad Dizgah, M. Biagetti, E. Sefusatti, V. Desjacques and J. Noreña, *Primordial Non-Gaussianity from Biased Tracers: Likelihood Analysis of Real-Space Power Spectrum and Bispectrum*, *JCAP* **05** (2021) 015 [arXiv:2010.14523] [InSPIRE].

[163] J. Schaye et al., *The EAGLE project: Simulating the evolution and assembly of galaxies and their environments*, *Mon. Not. Roy. Astron. Soc.* **446** (2015) 521 [arXiv:1407.7040] [InSPIRE].

[164] The EAGLE team, *The EAGLE simulations of galaxy formation: Public release of particle data*, arXiv:1706.09899 (2017).

[165] M. Hirschmann, K. Dolag, A. Saro, L. Bachmann, S. Borgani and A. Burkert, *Cosmological simulations of black hole growth: AGN luminosities and downsizing*, *Mon. Not. Roy. Astron. Soc.* **442** (2014) 2304 [arXiv:1308.0333] [InSPIRE].

[166] I.G. McCarthy, J. Schaye, S. Bird and A.M.C. Le Brun, *The BAHAMAS project: Calibrated hydrodynamical simulations for large-scale structure cosmology*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 2936 [arXiv:1603.02702] [InSPIRE].

[167] Y. Dubois et al., *Dancing in the dark: galactic properties trace spin swings along the cosmic web*, *Mon. Not. Roy. Astron. Soc.* **444** (2014) 1453 [arXiv:1402.1165] [InSPIRE].

[168] P. McDonald and A. Roy, *Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS*, *JCAP* **08** (2009) 020 [arXiv:0902.0991] [InSPIRE].

[169] K.C. Chan, R. Scoccimarro and R.K. Sheth, *Gravity and Large-Scale Non-local Bias*, *Phys. Rev. D* **85** (2012) 083509 [arXiv:1201.3614] [InSPIRE].

[170] T. Baldauf, U. Seljak, V. Desjacques and P. McDonald, *Evidence for Quadratic Tidal Tensor Bias from the Halo Bispectrum*, *Phys. Rev. D* **86** (2012) 083540 [arXiv:1201.4827] [InSPIRE].

[171] S. Saito, T. Baldauf, Z. Vlah, U. Seljak, T. Okumura and P. McDonald, *Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum*, *Phys. Rev. D* **90** (2014) 123522 [arXiv:1405.1447] [InSPIRE].

[172] A.S. Schmidt, S.D.M. White, F. Schmidt and J. Stücker, *Cosmological N-Body Simulations with a Large-Scale Tidal Field*, *Mon. Not. Roy. Astron. Soc.* **479** (2018) 162 [arXiv:1803.03274] [InSPIRE].

[173] J. Stücker, A.S. Schmidt, S.D.M. White, F. Schmidt and O. Hahn, *Measuring the Tidal Response of Structure Formation: Anisotropic Separate Universe Simulations using TreePM*, *Mon. Not. Roy. Astron. Soc.* **503** (2021) 1473 [arXiv:2003.06427] [InSPIRE].

[174] S. Masaki, T. Nishimichi and M. Takada, *Anisotropic separate universe simulations*, *Mon. Not. Roy. Astron. Soc.* **496** (2020) 483 [arXiv:2003.10052] [InSPIRE].

[175] K. Akitsu, Y. Li and T. Okumura, *Cosmological simulation in tides: power spectra, halo shape responses, and shape assembly bias*, *JCAP* **04** (2021) 041 [arXiv:2011.06584] [InSPIRE].

[176] T. Lazeyras and F. Schmidt, *A robust measurement of the first higher-derivative bias of dark matter halos*, *JCAP* **11** (2019) 041 [arXiv:1904.11294] [InSPIRE].

[177] F. Schmidt and F. Beutler, *Imprints of Reionization in Galaxy Clustering*, *Phys. Rev. D* **96** (2017) 083533 [arXiv:1705.07843] [InSPIRE].

[178] G. Cabass and F. Schmidt, *A new scale in the bias expansion*, *JCAP* **05** (2019) 031 [arXiv:1812.02731] [InSPIRE].