Bias deconstructed: unravelling the scale dependence of halo bias using real-space measurements

PARANJAPE, Aseem, et al.

Abstract

We explore the scale dependence of halo bias using real-space cross-correlation measurements in N-body simulations and in PINOCCHIO, an algorithm based on Lagrangian Perturbation Theory. Recent work has shown how to interpret such real-space measurements in terms of $k$-dependent bias in Fourier space, and how to remove the $k$-dependence to reconstruct the $k$-independent peak-background split halo bias parameters. We compare our reconstruction of the linear bias, which requires no free parameters, with previous estimates from N-body simulations which were obtained directly in Fourier space at large scales, and find very good agreement. Our reconstruction of the quadratic bias is similarly parameter-free, although in this case there are no previous Fourier space measurements to compare with. Our analysis of N-body simulations explicitly tests the predictions of the excursion set peaks (ESP) formalism of Paranjape et al. for the scale dependence of bias; we find that the ESP predictions accurately describe our measurements. In addition, our measurements in PINOCCHIO serve as a useful, successful consistency check between [...]
Bias deconstructed: unravelling the scale dependence of halo bias using real-space measurements

Aseem Paranjape,1,2 Emiliano Sefusatti,2,3 Kwan Chuen Chan,4 Vincent Desjacques,4 Pierluigi Monaco5,6 and Ravi K. Sheth2,7

1ETH Zürich, Institute for Astronomy, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland
2The Abdus Salam International Center for Theoretical Physics, Strada Costiera, 11, I-34151 Trieste, Italy
3INAF, Osservatorio Astronomico di Brera, Via Bianchi 46, I-23807 Merate (LC), Italy
4Département de Physique Théorique and Center for Astroparticle Physics (CAP), Université de Genève, 24 quai Ernest Ansermet, CH-1211 Genève, Switzerland
5Dipartimento di Fisica – Sezione di Astronomia, Università di Trieste, via Tiepolo 11, I-34131 Trieste, Italy
6INAF, Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34131 Trieste, Italy
7Center for Particle Cosmology, University of Pennsylvania, 209 S. 33rd St, Philadelphia, PA 19104, USA

Accepted 2013 August 20. Received 2013 August 20; in original form 2013 June 4

ABSTRACT
We explore the scale dependence of halo bias using real-space cross-correlation measurements in N-body simulations and in PINOCCHIO, an algorithm based on Lagrangian Perturbation Theory. Recent work has shown how to interpret such real-space measurements in terms of k-dependent bias in Fourier space, and how to remove the k-dependence to reconstruct the k-independent peak-background split halo bias parameters. We compare our reconstruction of the linear bias, which requires no free parameters, with previous estimates from N-body simulations which were obtained directly in Fourier space at large scales, and find very good agreement. Our reconstruction of the quadratic bias is similarly parameter-free, although in this case there are no previous Fourier space measurements to compare with. Our analysis of N-body simulations explicitly tests the predictions of the excursion set peaks (ESP) formalism of Paranjape et al. for the scale dependence of bias; we find that the ESP predictions accurately describe our measurements. In addition, our measurements in PINOCCHIO serve as a useful, successful consistency check between PINOCCHIO and N-body simulations that is not accessible to traditional measurements.

Key words: large-scale structure of Universe.

1 INTRODUCTION
Galaxies, and the dark matter haloes they live in, cluster differently from the underlying dark matter field itself. This halo bias is expected to be non-linear, non-local and stochastic, and understanding its behaviour is a prerequisite to successful program of precision cosmology with large-scale structure. While this non-linearity, non-locality and stochasticity of bias is measured in N-body simulations of cold dark matter (CDM), its precise physical origin remains unclear, and is likely to be influenced by several effects (Desjacques et al. 2010; Baldauf et al. 2012; Chan, Scoccimarro & Sheth 2012; Sheth, Chan & Scoccimarro 2013). In practice, in the absence of accurate analytical predictions of the so-called non-linear bias parameters $b_n$ discussed below, one resorts to fitting these parameters to measurements in N-body simulations (Tinker et al. 2005; Pollack, Smith & Porciani 2012) or marginalizing over them when analysing data from galaxy surveys (e.g. Blake et al. 2011; Sánchez et al. 2012), leading to a potential source of unmodelled systematic effects when attempting to recover information on cosmological parameters.

The language used when discussing halo bias is also not unique. Traditional measurements of bias in simulations are performed in Fourier space. For example, `linear bias’ is typically defined using ratios of power spectra of the halo overdensity $\delta_h(k)$ and matter overdensity $\delta(k)$.

$$b_1^2(k) \equiv \frac{P_{hh}(k)}{P_{mm}(k)} \quad \text{or} \quad b_1(k) \equiv \frac{P_{hm}(k)}{P_{mm}(k)},$$

where $P_{hh}(k) = \langle \delta_h^2 \rangle$, $P_{mm}(k) = \langle \delta^2 \rangle$ and $P_{hm}(k) = \langle \delta \delta_h \rangle$ are the corresponding cross-power spectrum. These ratios are found to be scale independent at large scales (small $k$) as expected from peak-background split
arguments (Kaiser 1984; Mo & White 1996; Sheth & Tormen 1999).

Quadratic bias is typically estimated by measuring (cross-) bispectra of \( b_\| (k) \) and \( \delta (k) \) and modelling them, e.g. by using perturbation theory or halo model arguments combined with a 'local biasing' scheme \( \delta(x) = b_1 \delta(x) + b_2 \delta(x)^2 / 2! + \cdots \) (Fry & Gaztañaga 1993), and in this case the state of the art (Pollack et al. 2012) shows systematic effects associated with, e.g., shot-noise modelling.

The corresponding real-space measurements of bias typically involve gridding the halo and matter density fields on some smoothing scale and then fitting a quadratic relation to the associated scatter plot (e.g. Manera & Gaztañaga 2012). The resulting fits show a dependence on smoothing scale, although it is not easy to interpret this scale dependence in terms of a \( k \)-dependence in Fourier space (Chan & Scoccimarro 2012). (See also Angulo, Baugh & Lacey 2008, for a large-scale real-space treatment that assumes local biasing and recovers scale-independent bias parameters to fourth order.) Recent work using an excursion set approach to the problem has revealed several features of halo bias: (a) Lagrangian Fourier-space bias at any non-linear order is very naturally linked to a particular real-space definition of Lagrangian bias based on cross-correlating the halo density with a suitable transform of the smoothed initial matter density (Musso, Paranjape & Sheth 2012; Paranjape & Sheth 2012a); (b) the excursion set analysis, as well as its extension to peaks theory (Paranjape & Sheth 2012b; Paranjape, Sheth & Desjacques 2013), predicts a specific smoothing-scale dependence of these real-space bias parameters, and hence a specific \( k \)-dependence in Fourier space; and (c) this scale dependence can be unravelled to reconstruct the large-scale, scale-independent bias coefficients using measurements at a finite intermediate smoothing scale. Ultimately, it is the dependence of these scale-independent coefficients on redshift and halo mass that probes the underlying cosmology.

In this paper, we apply these ideas to haloes identified in \( N \)-body simulations of CDM, as well as haloes identified in PINOCCHIO (Monaco, Theuns & Taffoni 2002; Monaco et al. 2013), which is a fast algorithm based on Lagrangian Perturbation Theory (LPT) which provides positions, velocities and merger histories of dark matter haloes. The measurements in the \( N \)-body simulations constitute a direct test of the excursion set peaks (ESP) formalism (Paranjape et al. 2013) which, as we show below, fares very well. The additional measurements in PINOCCHIO then become a very useful (and successful) consistency check between PINOCCHIO and the \( N \)-body simulations on the one hand, and between PINOCCHIO and ESP on the other. Taken together, our results constitute a self-consistent test of real-space measurements of linear and quadratic bias with no free parameters.

The paper is organized as follows. In Section 2, we give details of our simulation data set, including a brief description of the PINOCCHIO algorithm. Section 3 deals with measuring the bias parameters and comparing with theory. We first recapitulate in Section 3.1 the real-space definition of the \( n \)th order bias parameters \( b_n \) and its relation to Fourier-space definitions such as equation (1). Our measurements and the resulting estimates of \( b_1 \) and \( b_2 \) from the data are described in Section 3.2. We find that these estimates, which we make at different smoothing scales, are in good agreement with the corresponding (scale-dependent) predictions of the ESP formalism. This agreement is important because the measurements themselves are completely independent of the ESP formalism.

In Section 3.3, we use the reconstruction algorithm mentioned above, specifically the version described by Paranjape et al. (2013), to obtain estimates of the scale-independent peak-background split parameters \( b_{10} \) and \( b_{20} \) from the estimates of \( b_1 \) and \( b_2 \). The peak-background split bias parameters defined in the excursion set approach directly probe the halo mass function \( f(\delta_c; m) \) through

\[
b_{10} = f^{-1} \left( -\frac{\partial}{\partial \delta_c} \right)^n f, \tag{2}
\]

where \( \delta_c \) is the usual overdensity threshold predicted by spherical collapse (Mo & White 1996; Sheth & Tormen 1999).

If the reconstruction works well, then these estimates of \( b_{10} \) should be independent of the smoothing scale at which the \( b_n \) were measured; we find that this is indeed the case. The linear bias coefficient \( b_{10} \) is also directly probed by the large-scale limit of Fourier-space measurements such as equation (1) (Paranjape & Sheth 2012a). We compare our reconstruction of \( b_{10} \) with the Fourier-space large-scale fit to \( N \)-body simulations provided by Tinker et al. (2010), and find very good agreement. For \( b_{20} \) there are no previous \( N \)-body measurements we can compare with; a comparison with the ESP prediction (Paranjape et al. 2013) shows good agreement. We conclude in Section 4.

We assume a flat \( \Lambda \)-CDM cosmology with Gaussian initial conditions and compute transfer functions using CAMB (Lewis, Challinor & Lasenby 2000)\(^1\) for two different sets of parameter values: \( (\Omega_m, \sigma_8, n_s, h, \Omega_b) = (0.272, 0.81, 0.967, 0.704, 0.0455) \) for the \( N \)-body simulations and \( (0.25, 0.8, 0.95, 0.7, 0.044) \) for PINOCCHIO.

2.1 N-body

Our CDM simulations were run with 1024\(^3\) particles in a cubic box of size 1500\( h^{-1} \)Mpc, with each particle carrying a mass of \( 2.37 \times 10^{11} h^{-1} M_\odot \). Gaussian initial conditions were set at a starting redshift \( z = 99 \), with initial particle displacements implemented using second-order LPT (Crocce, Pueblas & Scoccimarro 2006). The simulations were run using GADGET II (Springel 2005). We use six realizations, with haloes identified using the Spherical Overdensity (SO) halo finder AHF (Gill, Knebe & Gibson 2004; Knollmann & Knebe 2009) which uses a redshift-dependent overdensity criterion motivated by spherical collapse (Eke, Cole & Frenk 1996; Bryan & Norman 1998). We only study haloes having at least 100 particles: this corresponds to halo masses larger than \( \sim 10^{13.5} h^{-1} M_\odot \).

2.2 PINOCCHIO

For a detailed explanation of the PINOCCHIO code we refer the reader to the original paper and to the more recent Monaco et al. (2013) where its parallel implementation and its application to cosmological volumes are presented. Here, we limit ourselves to a quite succinct description.

PINOCCHIO starts from a linear density field generated on a grid in a manner close to the generation of initial conditions in an \( N \)-body simulation. It uses third-order LPT applied to the evolution of a homogeneous ellipsoid to compute the collapse times of 'particles' (grid points); consistently with the excursion set approach, collapse times are computed for many smoothing radii, thus constructing for each particle a 'trajectory' in the plane defined by mass variance and inverse collapse time (the inverse of the growth factor at collapse).

\(^1\) http://lambda.gsfc.nasa.gov/toolbox/tb_camb_form.cfm
At variance with the standard excursion set approach, correlations between trajectories of nearby particles are fully taken into account. An algorithm that mimics the hierarchical assembly of structures is then applied to construct dark matter haloes. These are built in the Lagrangian space, then displaced to their final positions using the Zel’dovich approximation. This algorithm has been tuned to reproduce, to within 5–10 per cent, the ‘universal’ mass function of Warren et al. (2006).

As mentioned above, PINOCCHIO borrows ideas from excursion set theory and, on its own, cannot be considered a completely independent check for the ESP formalism. Nevertheless, these approaches are different enough from each other and from a full-fledged N-body simulation that we believe our analysis provides valuable information on the extent to which each of them reliably captures the underlying physical processes.

We consider here two different set-ups for the PINOCCHIO realizations. The first consists of a box of side $768 h^{-1}$ Mpc sampled by $1024^3$ particles with mass $2.9 \times 10^{10} h^{-1} M_{\odot}$. The second is given instead by a larger box of side $3072 h^{-1}$ Mpc sampled by $2048^3$ particles, each of mass $2.3 \times 10^{11} h^{-1} M_{\odot}$. For the smaller box, we produce nine realizations with different initial seeds, in order to provide a more solid estimate of the uncertainties on our results. In each case, we only study haloes with at least 100 particles.

Monaco et al. (2013) show, using the same realizations, that PINOCCHIO predicts a mass function consistent, within 10 per cent, to the fit of Warren et al. (2006) for the mass range $10^{13}$–$10^{15} h^{-1} M_{\odot}$, while a larger discrepancy is observed, in the large mass tail, with the results of Crocce et al. (2010). In addition, for halo populations characterized by a fixed threshold in mass, a 10 per cent agreement is achieved for the linear bias, determined from measurements of the halo-halo power spectrum. Better results for bias are found at lower masses, where the mass function fit is more accurate.

Fig. 1 shows the halo mass functions output in these PINOCCHIO realizations (left-hand panels) and measured in our N-body simulations (right-hand panels) at two different redshifts, compared with the fit to N-body simulations by Tinker et al. (2008, specifically, using their SO-200mean parameter values), the Sheth & Tormen (1999, hereafter ST) form with their $q = 0.7$ and $p = 0.26$ which we find gives a good fit to the MICE simulations of Crocce et al. (2010), and the analytical ESP calculation from Paranjape et al. (2013). The lower panels in Fig. 1 show the ratio of all quantities with the ESP prediction. (We will use a similar format in all our figures below.) As discussed above, these mass functions agree at the 10–20 per cent level, with larger discrepancies at higher masses. We also note that the halo masses at $z = 0$ in our N-body simulations tend to be ~10 per cent lower than those expected from, e.g. the SO fit by Tinker et al. (2010), and this shows up as a vertical offset in the ratio in the lower panel of the figure. This could be due to the slightly different halo finding criteria in these simulations; the Tinker et al. (2010) fit we use corresponds to haloes identified by them at a density equal to 200 times the mean density at all redshifts, while the halo finder in our simulation finds haloes at $z = 0$ with density ~350 times the mean density and at $z = 1$ with ~200 times the mean density (Eke et al. 1996; Bryan & Norman 1998).

3 HALO BIAS IN REAL SPACE

3.1 Analytical motivation

Before presenting our measurements, we summarize our current understanding of halo bias as motivated by the excursion set approach and recapitulate how the reconstruction argument works. This will also serve to set our notation.

Throughout this paper we will focus on Lagrangian bias defined with respect to the linearly extrapolated initial dark matter density field. We will also denote by $\delta_i$ the traditional spherical collapse barrier for the excursion set random walks, using $\delta_i(z = 0) = 1.686$, and define $v = \delta_i/\sqrt{3}$ where

$$ s \equiv \sigma_v^2(R) \equiv \langle (\delta(R))^2 \rangle = \int d\ln k \Delta^2(k) W(kR)^2 $$

is the linearly extrapolated variance on the Lagrangian scale of the halo, with $\Delta^2(k) \equiv k^3 P(k)/2\pi^2$ the dimensionless linear matter power spectrum at $z = 0$ and $W(kR)$ the smoothing filter which we will take to be a TopHat in real space so that $W(y) = (3/\pi)^{1/2} \sin y - y \cos y$.

Halo bias can be defined in Fourier or real space. For example, linear bias in Fourier space can be defined using ratios of (cross-)power spectra as in equation (1). While these are convenient definitions, e.g. in an N-body simulation, analytical approaches such as the excursion set formalism work naturally in real space, and a priori it is not obvious how the results of the latter should be interpreted in Fourier space. Recent work (Musso & Sheth 2012; Musso et al. 2012; Paranjape & Sheth 2012a) has shown how this connection can be made in practice. In particular, Musso et al. (2012) argued that a useful definition of the nth order Lagrangian halo bias coefficient $b_n$ in real space is as follows. Consider a simulation in which we have identified haloes at some redshift, e.g. $z = 0$. One can now use the particles identified as belonging to a halo in the final non-linear field to define a ‘protohalo’ in the initial conditions, and compute the centre-of-mass of this Lagrangian protohalo using the positions of those same particles in the initial conditions. If there are $N$ haloes (equivalently, protohaloes) in a given mass bin, then the nth order bias coefficient $b_n$ is estimated at some Lagrangian smoothing scale $R_0$ as

$$ b_n = S_n^n/2 \frac{1}{N} \sum_{i=1}^{N} H_n(\delta_0 \sqrt{S_0}) , $$

where $\delta_0$ is the dark matter density contrast in the initial conditions (linearly extrapolated to present epoch) smoothed on the scale $R_0$ and centred on the centre-of-mass in the initial conditions of the ith halo in the bin; $S_0$ is the linearly extrapolated variance at scale $R_0$, $S_0 = \langle \delta_0^2 \rangle = \sigma_v^2(R_0)$, and the $H_n$ are the probabilist’s Hermite polynomials, $H_n(x) = e^{x^2/2} (\partial^n/\partial x^n) e^{-x^2/2}$, with $p_0(x - \mu; \sigma^2)$ a Gaussian in the variable $x$.

The value 1.686 strictly holds only in an $\Omega_m = 1$ cosmology – we use it primarily for ease of comparison with Paranjape et al. (2013) and Tinker et al. (2010) who also used 1.686. The exact value appropriate for the $\Lambda CDM$ cosmologies we study would be different from this number at the ~1 per cent level. This would lead to discrepancies in the mass function, at the highest masses, of the order of 15–20 per cent, whereas in the bias it would cause discrepancies of the order of ~1 per cent which have no impact on our final results.

2 The two PINOCCHIO runs correspond in size and resolution to the MICE768 and MICE3072 runs of the MICE suite Crocce et al. (2010). This choice is motivated by a direct comparison with those N-body simulations (Monaco et al. 2013).
The connection to Fourier-space bias arises as follows. The excursion set analysis (Musso et al. 2012) as well as its ESP extension (Paranjape & Sheth 2012b; Paranjape et al. 2013) predict the following form\(^4\) for the \(b_h\):

\[
b_h = \left( \frac{S_\epsilon}{S_0} \right)^n \sum_{r=0}^{n} \binom{n}{r} b_{\text{esp}} \epsilon^r ,
\]

where

\[
S_\epsilon = \int \ln k \, \Delta^2(k) W(k R) W(k R_0) ,
\]

\[
\epsilon = 2 \ln S_\epsilon / \ln s
\]

are\(^5\) cross-correlations between the mass overdensity field smoothed on the large scale \(R_0\) and on the Lagrangian scale of the halo \(R \propto m^{1/3}\), and the \(b_{\text{esp}}\) are mass-dependent but scale-independent.

\(^4\) The convention for notation in equation (8) differs from the one used in, e.g. Desjacques et al. (2010). The convention here was introduced by Musso et al. (2012) and is adapted to counting powers of \(\epsilon\) in real space which roughly correspond to powers of \(k^2\) in Fourier space.

\(^5\) Strictly speaking, for ESP, \(\epsilon\) should be defined in terms of mixed spectral moments: \(\epsilon = (s/S_\epsilon)(\sigma_{1m}^2/\sigma_{1s}^2)\) where \(\sigma_{1m}^2 = \int \ln k \Delta^2(k) k^2 e^{-2k^2 R_0^2/2} W(k R) ; \sigma_{1s}^2 = \int \ln k \Delta^2(k) k^2 e^{-2k^2 R_0^2/2} W(k R_0)\) and \(R_0\) is matched to \(R\) by demanding \(\langle \delta \delta \delta \rangle = s\) for the reasons discussed by Paranjape et al. (2013). While we implement this in our analysis, we have found that using the second equation in (9) leads to practically identical results.
coefficients whose functional form depends on details of the analysis, such as whether one uses the traditional excursion set calculations of Musso et al. (2012) or the ESP approach of Paranjape et al. (2013).

Among these, the coefficients $b_{n0}$ are somewhat special because they are the logarithmic derivatives of the halo mass function $f(\delta_c, s)$ (Musso et al. 2012) and coincide with the peak-background split bias parameters from equation (2). These are the coefficients one is typically interested in measuring, since they carry information regarding the growth of large-scale structure and are hence sensitive to the underlying cosmology. The other coefficients follow from the cross-correlation calculations advocated by Musso et al. (2012), e.g. for $n = 1, 2$ in the ESP case they can be read off from equations 29 and 30 in Paranjape et al. (2013, also see below).

Musso et al. (2012) showed how the appearance of $\epsilon_s$ in the expression for real-space bias signals a $k$-dependence in Fourier-space bias. Essentially, this is because the real-space cross-correlation that defines $b_n$ can be interpreted in Fourier space by formally introducing $m_n(k)$ and then matching terms in Fourier space with the quantities appearing in equation (8). Roughly, each power of $\epsilon_s$ in real space corresponds to a power of $\ln W(kR)/\ln R$ in Fourier space. So, for example, if $W(kR)$ were a Gaussian filter $W(kR) = e^{-k^2\sigma_s^2}$, then the real-space linear bias $b_1 = (S_x/S_0)(b_{10} + \epsilon_s b_{11})$ would translate in Fourier space as $b_1(k) = b_{10} + (k^2/\sigma_s^2)b_{11}$, where $\sigma_s^2 = \int d\ln k \Delta^2(k)k^2e^{-k^2\sigma_s^2}$. Musso et al. (2012) gave a formal proof that the real-space $b_n$ correspond to integrals over quantities that Matsubara (2012) calls ‘renormalized’ Lagrangian bias coefficients in Fourier space.

A remarkable aspect of the excursion set and ESP frameworks is that there exist linear relations between the $b_n$ which allow all of them to be written in terms of only the coefficients $b_{00}$. For example, for the simplified case of a constant excursion set barrier $B = \delta_c$, one finds (Musso et al. 2012)

$$
\delta_{\epsilon} b_{11} = v^2 - \delta_{\epsilon} b_{10}$$
$$
\delta_{\epsilon}^2 b_{21} = v^2(\delta_{\epsilon} b_{10} - 1) - \delta_{\epsilon}^2 b_{20}$$
$$
\delta_{\epsilon}^2 b_{22} = \delta_{\epsilon}^2 b_{20} + v^2(v^2 - 2\delta_{\epsilon} b_{10} + 1).$$

This means that the scale-dependent coefficients $b_{1}$ and $b_{2}$ can be written as linear combinations of $b_{10}$ and $b_{20}$. And upon measuring $b_{1}$ and $b_{2}$, one can read off the values of $b_{10}$ and $b_{20}$. This is the basis of the reconstruction procedure proposed by Musso et al. (2012), and generalizes to arbitrary order $b_{n0}$.

There is one complication, though. Paranjape et al. (2013) discussed the fact that the barrier appropriate for the excursion set random walks which determine halo masses is not deterministic, but has a mass-dependent scatter, and argued that this is a key ingredient in the analysis if one demands $\sim 10$ percent accuracy when comparing the halo mass function with that measured in simulations. They showed, based on the N-body results of Robertson et al. (2009), that a good model for this barrier is

$$
B = \delta_c + \beta \sqrt{\sigma},$$

where $\beta$ is a stochastic variable drawn from a Lognormal distribution whose mean and variance are fixed by the Robertson et al. (2009) results to be approximately 0.5 and 0.25, respectively. [The results for halo bias are very insensitive to the exact choice for this distribution; this is reassuring since Despali, Tormen & Sheth (2013) suggest slightly different values for the mean and variance of this distribution.]

Define the functions

$$
\mu_n(v, \beta) = v^n H_k(v + \beta),$$

and their averages over $\beta$,

$$
\langle \mu_1 | v \rangle = v (v + \langle \beta | v \rangle)$$
$$
\langle \mu_2 | v \rangle = \langle \mu_1^2 | v \rangle - v^2$$

$$
= v^2 \left( v^2 - 1 + 2v \langle \beta | v \rangle + \langle \beta^2 | v \rangle \right),$$

where

$$
\langle \beta^2 | v \rangle = \frac{\int d\beta \, p(\beta) f_{ESP}(v | \beta) \beta^2}{\int d\beta \, p(\beta) f_{ESP}(v | \beta)}.$$

with $f_{ESP}(v | \beta)$ given in equation 13 of Paranjape et al. (2013). (The second line of equation 13 corrects a typo in equation 37 of Paranjape et al. 2013). Then, the estimate for the reconstructed $b_{10}$ becomes (equation 35 of Paranjape et al. 2013),

$$
\delta_{\epsilon} b_{10} = \frac{1}{(1 - \epsilon_s)} \left[ \frac{\delta_{\epsilon} b_1}{(S_x/S_0)} - \epsilon_s \langle \mu_1 | v \rangle \right],$$

which is straightforward to implement since each term is either easily measurable or calculable.

For $b_{20}$, however, they showed that the exact expression for the reconstruction involves a term $\langle b_1 \mu_1 | v \rangle$ which is cumbersome to keep track of (although it is in principle measurable in the simulation). They therefore proposed a simplification based on assuming $\langle b_1 \mu_1 | v \rangle \rightarrow \hat{b}_1 \langle \mu_1 | v \rangle = \hat{b}_1 \langle \mu_1 | v \rangle$, in which the case estimate for $b_{20}$ is (their equation 36)

$$
\delta_{\epsilon}^2 b_{20} = \frac{1}{(1 - \epsilon_s)^2} \left[ \frac{\delta_{\epsilon}^2 b_2}{(S_x/S_0)^2} - 2\epsilon_s \left( \frac{\delta_{\epsilon} b_1}{(S_x/S_0)} - \langle \mu_1 | v \rangle - \langle \mu_1^2 | v \rangle \right) \right.$$

$$
- \epsilon_s (2 - \epsilon_s) \langle \mu_2 | v \rangle \right].$$

We will make this assumption in what follows, and leave a more detailed analysis of correlations such as $\langle b_1 \mu_1 | v \rangle$ to future work.

3.2 Measurements of $b_1$ and $b_2$

We now present measurements for the linear and quadratic bias coefficients defined by equation (4) and measured in the N-body and PINOCCHIO runs described in Section 2. These measurements, unlike the reconstruction algorithm which we implement in the next subsection, are completely independent of any theoretical input, other than the recognition that weighting by Hermite polynomials (rather than by simple powers of the density field) is the most natural way to define bias factors in a Gaussian random field (Szalay 1988; Musso et al. 2012).

Figs 2 and 3 show the results for $\delta_{\epsilon} b_1$ as measured by summing equation (4) over the haloes identified at $z = 0$ in the N-body simulations and in PINOCCHIO, respectively. The two panels in each figure correspond to different smoothing scales $R_0$. Fig. 3 shows two sets of measurements; the black circles are from the larger simulation box and the grey squares from the smaller box.

For the N-body simulations, the circles in Fig. 2 show the mean over six runs and the error bars are the scatter around this mean for each bin. Similarly, the squares in Fig. 3 show the mean over
Figure 2. Measurements of the linear halo bias $b_1$ in the $N$-body simulations (points with error bars; see the text for details). The two panels show measurements at two different smoothing scales. The smooth curves show the corresponding theory predictions using ESP (solid red) and from Mo & White (1996, dotted black), and the fit to large-scale Fourier-space measurements in $N$-body simulations from Tinker et al. (2010, dashed green). We multiplied the latter two functions with the factor $S_\perp/S_0$ (which slowly varies with mass and is $\sim 1.4$ for $R_0 = 50 \, h^{-1} \text{Mpc}$ and $\sim 1.35$ for $R_0 = 90 \, h^{-1} \text{Mpc}$) to account for the mapping from Fourier to real space. The ESP prediction includes the effect of both $S_\perp/S_0$ as well as $\epsilon_\perp$, although the latter effect is quite small at these scales. Lower panels show the ratios with respect to the ESP prediction, with the grey bands indicating 20 per cent deviations.

Figure 3. Same as Fig. 2, now using measurements in PINOCCHIO. The black circles and grey squares show measurements in the larger and smaller box, respectively. The latter show clear signs of finite volume effects for the larger smoothing scale.

Nine runs, and the error bars the scatter around this mean, for the smaller PINOCCHIO box. For the larger PINOCCHIO box, we had a single run, and the corresponding error bars in Fig. 3 reflect Poisson errors for each mass bin. A comparison of the error bars shows that, at small masses, the Poisson errors are very likely underestimating the scatter in any given bin.

These measurements are susceptible to at least two systematic finite volume effects; the first affects smoothing scales that are a significant fraction of the box size (e.g. $R_0 = 90 \, h^{-1} \text{Mpc}$ in the 768 $h^{-1} \text{Mpc}$ PINOCCHIO box), and the second affects mass bins whose Lagrangian radius is comparable to the smoothing scale [e.g. $\log_{10} \nu^2(m, z = 0) = 1.4$ or $m \sim 4.5 \times 10^{15} \, h^{-1} \text{M}_\odot$ which has $R \sim 15 \, h^{-1} \text{Mpc}$, with $R_0 = 30 \, h^{-1} \text{Mpc}$]. Our choice of $R_0 = 50 \, h^{-1} \text{Mpc}$ tries to minimize these effects, while $R_0 = 90 \, h^{-1} \text{Mpc}$ highlights the first one.

The solid red curves in Figs 2 and 3 are the corresponding predictions of the ESP framework at each smoothing scale (equation 29 of Paranjape et al. 2013), for the respective cosmology. The dashed green lines in each panel are the Lagrangian bias from the fit presented by Tinker et al. (2010). To be consistent in comparing with our real-space measurements, we multiplied this Fourier-space fit with a factor $S_\perp/S_0$ at each smoothing scale. The dotted
black curves are the standard spherical collapse prediction $v^2 - 1$ (Mo & White 1996), which we also multiplied by $S_x/S_0$. This is necessary because, although the Mo–White calculation is based on the excursion set approach, it implicitly assumes that the smoothing window with which $\delta_0$ is defined is a TopHat in $k$-space, for which $S_x/S_0 = 1$, whereas our measurements (along with essentially all other real-space measurements) use a real-space TopHat, a point first made by Paranjape & Sheth (2012a).

The careful reader will have noticed that although the dashed and dotted curves are almost the same in the two panels (they differ only because the multiplicative factor $S_x/S_0$ is about 5 per cent different), the shape of the ESP prediction changes slightly. This is due to the scale dependence introduced by $\epsilon_x$, although this effect is much smaller than the size of the error bars on the measurements. We see that the measurements of $b_1$ agree very well with the ESP prediction. The small systematic differences between the black circles and the red solid lines in each of the figures can be traced back entirely to the respective mass functions; the ESP mass function slightly underpredicts the masses of the PINOCCHIO haloes and slightly overpredicts those in the $N$-body simulation (cf. Fig. 1), which is consistent with the trends seen in Figs 2 and 3.

Figs 4 and 5 show the corresponding measurements with errors for the quadratic bias $\delta^2 b_2$. The solid red curves in these figures are the ESP prediction (equation 30 of Paranjape et al. 2013). The

Figure 4. Measurements of the quadratic halo bias $b_2$ in the $N$-body simulations (points with error bars; see the text for details). The left- and right-hand panels show measurements at two different smoothing scales (the same as in Fig. 2). The smooth curves show the corresponding theory predictions using ESP (solid red) and from Mo, Jing & White (1997, dotted black), the latter being multiplied with the factor $(S_x/S_0)^2$ to account for the mapping from Fourier to real space. In this case, there are no prior $N$-body simulation results to compare with. Insets show zoomed-in views of the lowest masses. Lower panels show the ratios with respect to the ESP prediction, with the grey bands indicating 20 per cent deviations.

Figure 5. Same as Fig. 4, now using measurements in PINOCCHIO. The black circles and grey squares show measurements in the larger and smaller box, respectively. Finite volume effects are apparent at large masses for the measurements in the smaller simulation box.
The appearance of $\langle \beta^j | v \rangle$ which depends on the ESP mass function means that one cannot interpret the reconstructed $\delta_{20}$ as clean tests of the ESP formalism. The original reconstruction prescription of Musso et al. (2012) only involved powers of $v$ (equation 10) rather than integrals over some stochastic variable such as $\beta$, so one might argue that their prescription was in some sense cleaner and unaffected by choices regarding, e.g. the choice of distribution $p(\beta)$. This is misleading, however, since that prescription explicitly assumed a constant deterministic barrier which has been shown by Paranjape et al. (2013) to yield a poor description of the halo mass function, which is better described instead by the ingredients discussed above.

The strength of our approach lies in the following. First, our reconstruction below of $b_{10}$, while model dependent, agrees very well with the fit presented by Tinker et al. (2010) to large-scale Fourier-space measurements in N-body simulations. Moreover, this model-dependence is almost entirely driven by the need to describe the mass function accurately (Paranjape et al. 2013), while the prediction for the bias, in a sense, comes for free. Secondly, our reconstruction of $b_{20}$ then makes no additional assumptions regarding the underlying model, and is in this sense a parameter-free estimate of quadratic bias. The agreement between the scale-dependent measurements of the previous subsection and the corresponding ESP predictions lends support to the expectation that this (albeit model-dependent) reconstruction scheme is correctly capturing the underlying physical processes that lead to halo bias.

The points with error bars in Fig. 6 show our reconstruction of $b_{10}$ (left-hand panel) and $\delta_{20}^2$ (right-hand panel) from measurements of $b_{10}$ and $b_{20}$ in the N-body simulations, for haloes identified at $z = 0$, on four different smoothing scales (two of which are the same as in the previous subsection). The errors were calculated by propagating the errors (i.e. scatter around the mean) on $b_{10}$ and $b_{20}$.

![Figure 6](http://mnras.oxfordjournals.org/)

**Figure 6.** The reconstruction technique described in the text applied to the N-body measurements of linear and quadratic bias at $z = 0$ at four different smoothing scales. Left: reconstructing $b_{10}$. The measurements are in excellent agreement with each other (indicating that the algorithm is working well) and with the fit to Fourier-space measurements from Tinker et al. (2010, dashed green) as well as the ESP prediction (solid red), apart from a minor trend caused by the mass mismatch seen in Fig. 1. For comparison we also show the prediction from Mo & White (1996, dotted black) and the peak-background split prediction using the ST fit to the MICE mass function (the ST mass function with $q = 0.7, p = 0.26$ in equation 2 with $n = 1$; dot–dashed magenta). Right: reconstructing $b_{20}$. The measurements again agree well with each other, although they are noisier than those of $b_{10}$. The smooth curves show the prediction of ESP (solid red), Mo et al. (1997, dotted black) and the peak-background split prediction from the ST fit to the MICE mass function (dot–dashed magenta, the ST mass function with $q = 0.7, p = 0.26$ in equation 2 with $n = 2$). The inset shows a zoom-in of the lowest masses. Note that the predicted values and measurements go through zero close to $\log_{10}v^2 \simeq 0.6$. Lower panels in each case show the ratios with respect to the ESP prediction, with the grey bands indicating 20 per cent deviations.
Fig. 7. Same as Fig. 6, now for the 3072 h^{-1} Mpc PINOCCHIO box at z = 0. In the left-hand panel, the b_{10} measurements are in excellent agreement with each other (with some hints of systematic volume effects at R_0 = 30 h^{-1} Mpc) and with the fit to Fourier-space measurements from Tinker et al. (2010, dashed green) as well as the ESP prediction (solid red). In the right-hand panel, the b_{20} measurements show a systematic decrease as R_0 increases.

Fig. 7 shows the corresponding reconstructions in the 3072 h^{-1} Mpc PINOCCHIO box; in this case, the errors were computed by propagating Poisson errors.

We see in the left-hand panel of Fig. 6 that the reconstruction of b_{10} at all scales gives nearly identical results (indicating that the algorithm is working well in removing the scale dependence). The reconstructed values are also in good agreement with the fit presented by Tinker et al. (2010, dashed green) and the ESP prediction (solid red; equation 29 of Paranjape et al. 2013, with reasonable agreement with the ESP prediction (solid red; equation 30 of Paranjape et al. 2013, with similar behaviour, and the measurements at all scales are in reasonable agreement with each other and with the Fourier-space fit from Tinker et al. (2010) as well as the ESP prediction. The b_{20} measurements again show a systematic drift in the single PINOCCHIO realization and a large realization-to-realization scatter in the N-body results. Additionally, there seems to be a systematic turnover at high masses (log_{10}v^2(m, z = 1) \gtrsim 1.3 or m \gtrsim 6 \times 10^{14} h^{-1} M_{\odot}) in both b_{10} and b_{20} at all smoothing scales, which is more pronounced in the N-body simulation than in PINOCCHIO. Since the PINOCCHIO box has more than eight times the volume of our individual N-body realizations, this is also likely to be a sample variance effect. We discuss other possible sources for these trends in Section 4 below.

4 DISCUSSION

The scale dependence and non-linearity of halo bias is a potential source of systematic effects for upcoming galaxy surveys, which can degrade the measurements of cosmological parameters. Current strategies for dealing with these effects mainly rely on using parametrized functions inspired by fits to N-body simulations (Tinker et al. 2005; Pollack et al. 2012). In principle, though, such effects can be modelled using analytical tools such as the excursion set formalism. While such models predict the linear, scale-independent part of halo bias with reasonable accuracy, they have not fared well until now in predicting these ‘beyond linear’ effects.

We have presented real-space measurements of Lagrangian halo bias at linear and quadratic order using a recently proposed technique (Musso et al. 2012; Paranjape et al. 2013) which exploits the correlation between the locations of halo formation in the initial conditions and the large-scale initial environment. The ‘observables’ in this measurement – namely the scale-dependent bias coefficients...
A. Paranjape et al.

Figure 8. Same as Fig. 6, now using haloes identified at redshift $z = 1$ in the $N$-body simulation. The measurements of $b_{10}$ are in excellent agreement with each other and with the Fourier-space fit from Tinker et al. (2010) as well as the ESP prediction. Both $b_{10}$ and $b_{20}$ now display possible sample variance effects at large masses ($\log_{10} \nu^2(m, z = 1) \gtrsim 1.3$, or $m \gtrsim 6 \times 10^{14} h^{-1} M_\odot$). See the text for a discussion.

Figure 9. Same as Fig. 7, now using haloes identified at redshift $z = 1$ in the $3072 h^{-1} \text{Mpc}$ PINOCCHIO box. See the text for a discussion.

$bn$ (equation 4) – can be estimated using simple measurements in the initial conditions of a simulation, and are direct predictions of the ESP formalism of Paranjape et al. (2013). We find that our measurements in $N$-body simulations (Figs 2 and 4) and in the PINOCCHIO algorithm (Figs 3 and 5) are in very good agreement with expectations using ESP.

Further, the ESP formalism also shows how these scale-dependent measurements of $b_n$ can be converted into scale-independent estimates of the peak-background split parameters $b_{0n}$ defined by equation (2). We find good agreement between our estimates of $b_{10}$ in Figs 6–9 and the fit to large-scale Fourier-space linear bias in $N$-body simulations presented by Tinker et al. (2010), which represents a success of the ESP formalism. Furthermore, the reconstruction of $b_{0n}$ utilizes no additional assumptions other than those required to obtain a good description of the mass function; in this sense, we have presented a parameter-free prescription for estimating non-linear large-scale (i.e. small $k$ or large $R_0$) bias from intermediate-scale measurements (we used $R_0 = 30, 70, 90 h^{-1} \text{Mpc}$).

Our analysis and results raise a number of interesting issues that must be addressed to assess and utilize the full potential of our approach:

(i) We note that our estimates of $b_{10}$ and $b_{20}$ show departures from the theoretical predictions at the largest halo masses; these are especially pronounced in the $N$-body results at $z = 1$ (Fig. 8). As discussed in the text, a likely cause for these is sample variance. Additionally, at small smoothing scales like $R_0 = 30 h^{-1} \text{Mpc}$ the trends could be partly due to the fact that the largest haloes have Lagrangian sizes that approach this size. There is also the possibility
that non-locality induced by effects such as those studied by Sheth et al. (2013) could be contaminating our reconstruction algorithm at high masses. It will be interesting to check whether such non-local, non-spherical effects can be isolated from the recovery of the $b_{nl}$, e.g. using the rotational-invariance motivated orthogonal polynomials associated with peak shapes (Gay, Pichon & Pogosyan 2012; Desjacques 2013) and the tidal field (Sheth et al. 2013). This should require straightforward extensions of our technique to measuring quantities other than the density.

(ii) Another important question is whether our technique, which works at finite scales, can become a realistic, cost-effective method for measuring bias in the late-time, gravitationally evolved Eulerian field. We have presented results using smoothing scales $R_0 \geq 30 \, h^{-1} \, \text{Mpc}$, which are large compared with, e.g., those used by Angulo et al. (2008) in their Eulerian analysis that recovered bias coefficients up to fourth order. The reason for having these large Lagrangian smoothing scales is simply that the most massive protohaloes we work with have Lagrangian sizes that can approach $\sim 15-20 \, h^{-1} \, \text{Mpc}$. Haloes in the Eulerian field are much more compact, so in principle our technique should be applicable at much smaller scales in this case. A bigger issue is the fact that the Eulerian field is weakly non-Gaussian, so using Hermite polynomials as we currently do may not be the best approach. Although one can find natural generalizations of the Hermite polynomials appropriate for non-Gaussian fields (e.g. see appendix B2 of Musso et al. 2012), the question of which basis set to use in extracting non-linear Eulerian bias is still not fully settled.

(iii) Our original goal included a measurement of the $k$-dependence of linear bias as implied by the presence of the quantity $\epsilon_\ast$, in the excursion set prediction in equation (8); however, this effect is too small to be reliably seen given our present error bars. For scale-dependent bias in the Eulerian field, there is also the related question of whether triaxiality of haloes could be a source of significant systematics for our reconstruction technique. The Eulerian field will also be affected by gravitationally induced non-locality (Chan et al. 2012) which would have to be accounted for.

And finally, it will be very interesting to assess to what extent a self-consistent estimate of non-linear and non-local halo bias in galaxy surveys can improve the recovery of cosmological parameters. We leave a detailed exploration of all these issues to future work.

ACKNOWLEDGEMENTS

We thank Roman Scoccimarro for collaboration in the initial stages of this work, and both him and Marcello Musso for useful discussions. KCC and VD acknowledge support by the Swiss National Science Foundation. RKS is supported in part by NSF-AST 0908241.

REFERENCES


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