Cosmological N-body simulations with suppressed variance

Raul E. Angulo1★ and Andrew Pontzen2★

1Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Plaza San Juan 1, Planta-2, E-44001, Spain
2Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

Accepted 2016 May 12. Received 2016 May 12; in original form 2016 March 17

ABSTRACT
We present and test a method that dramatically reduces variance arising from the sparse sampling of wavemodes in cosmological simulations. The method uses two simulations which are fixed (the initial Fourier mode amplitudes are fixed to the ensemble average power spectrum) and paired (with initial modes exactly out of phase). We measure the power spectrum, monopole and quadrupole redshift-space correlation functions, halo mass function and constrain the statistics of fluctuations to greater precision. For instance, reaching 1 per cent accuracy over the whole range of scales would require the simulation of \( \sim 10^5 \) Gpc\(^3\). In this Letter, we propose and test a method to suppress the effect of box variance drastically. We will show that with just two simulations we can achieve the accuracy delivered by tens to hundreds of traditional simulations at the same scale, depending on the particular problem in hand. Briefly, the two simulations:

(i) use a fixed input power spectrum, meaning that we enforce \( \hat{P}^L = P^L \) when generating the initial conditions;
(ii) are paired, so that a hierarchy of effects due to chance phase correlations can be cancelled (Pontzen et al. 2016).

The first condition destroys the statistical Gaussianity of the input field which, at first sight, would seem to limit the usefulness of the approach (see also Neyrinck & Yang 2013). However we will demonstrate empirically and analytically that, by all measures explored here, the non-Gaussian corrections have a negligible effect on ensemble mean clustering statistics.

This Letter is set out as follows. In Section 2, we implement and test our method. We develop an analytic understanding of why the method works in Section 3. Finally, in Section 4 we present our conclusions.

1 INTRODUCTION
Numerical simulations are an essential tool for cosmology, especially for interpreting observational surveys (see Kuhlen, Vogelsberger & Angulo 2012, for a review). They can be deployed to study the impact of a given cosmological ingredient (e.g. Baldi et al. 2014), create virtual galaxy populations (e.g. Overzier et al. 2009), check and develop analytic treatments for structure formation (e.g. Carlson, White & Padmanabhan 2009), and understand systematic and statistical errors in cosmological measurements (e.g. Manera et al. 2015). In the future, simulations could even be used to constrain cosmological parameters (Angulo & Hilbert 2015).

However, a limitation for all the above applications is the sparse sampling of Fourier modes due to the finite extent of the simulation box. A given cosmological simulation is initialized to a particular realization of a Gaussian random field. The power spectrum of the realization, \( \hat{P}^L(k) \), therefore differs from the ensemble mean power spectrum, \( P^L(k) \). Given a box large enough to capture all physical effects (Bagla, Prasad & Khandai 2009), the largest scale modes are still poorly sampled. This, together with the non-linear coupling of small and large scales, implies that several-Gpc size boxes generate statistical errors which limit inferences on 100 or even 10 Mpc scales.

This undersampling effect is closely connected to (though, owing to the non-linear evolution, not precisely the same as) observational cosmic variance. In the observational case, the finite volume that can be achieved by a given survey constitutes an irreducible source of uncertainty. On the other hand the computational variance can be strongly suppressed, at least in principle, until it is smaller than the cosmic variance and other sources of error. This is usually achieved by simulating huge cosmological volumes (e.g. Rasera et al. 2014) or a large number of realizations (e.g. Takahashi et al. 2009). Finite computing resources then generate a tension between the need for large volumes and for high resolution (the latter is required to better resolve the distribution of individual galaxies and their internal structure). Even as supercomputing facilities expand, the tension is becoming more acute as surveys probe larger scales and constrain the statistics of fluctuations to greater precision. For instance, reaching 1 per cent accuracy over the whole range of scales to be probed by Euclid would require the simulation of \( \sim 10^5 \) Gpc\(^3\).

In this Letter, we propose and test a method to suppress the effect of box variance drastically. We will show that with just two simulations we can achieve the accuracy delivered by tens to hundreds of traditional simulations at the same scale, depending on the particular problem in hand. Briefly, the two simulations:

(i) use a fixed input power spectrum, meaning that we enforce \( \hat{P}^L = P^L \) when generating the initial conditions;
(ii) are paired, so that a hierarchy of effects due to chance phase correlations can be cancelled (Pontzen et al. 2016).

The first condition destroys the statistical Gaussianity of the input field which, at first sight, would seem to limit the usefulness of the approach (see also Neyrinck & Yang 2013). However we will demonstrate empirically and analytically that, by all measures explored here, the non-Gaussian corrections have a negligible effect on ensemble mean clustering statistics.

This Letter is set out as follows. In Section 2, we implement and test our method. We develop an analytic understanding of why the method works in Section 3. Finally, in Section 4 we present our conclusions.

★ E-mail: rangulo@cefca.es (REA); a.pontzen@ucl.ac.uk (AP)

© 2016 The Authors
Published by Oxford University Press on behalf of the Royal Astronomical Society
2 COMPARISON WITH AN ENSEMBLE OF SIMULATIONS

2.1 Numerical simulations

All simulations considered in this Letter contain $1024^3$ particles of mass $1.7 \times 10^{12} h^{-1} M_\odot$ inside a box of side $L = 3 h^{-1}$ Gpc. The initial particle positions are computed from an input linear density field $\delta^i$ using 2LPT. The simulation particles are then evolved under self-gravity with a COLA algorithm (Tassev, Zaldarriaga & Eisenstein 2013) using 10 steps from $z = 9$ to 1. The cosmological parameters assumed correspond to those of the Millennium series (Springel 2005): $\Omega_m = 0.25$, $\sigma_8 = 0.9$, and $h = 0.73$.

The COLA algorithm is an approximate $N$-body method, in the sense that the orbits inside high-density regions are not properly integrated. However, the non-linear evolution of intermediate and large scales is accurately captured (Howlett, Manera & Percival 2015; Koda et al. 2016), at a fraction of the computational cost of a traditional $N$-body simulation. This enables the rapid simulation of extremely large volumes, which in turn allows very precise calculations of different statistics that serve as a benchmark for the performance of our method. The total volume of our reference ensemble is $8100 h^{-3}$ Gpc$^3$; more details are given in Chaves-Montero et al. (in preparation).

The only difference between the ensemble of simulations and the pair of fixed simulations is in the input fields $\delta^i(x)$. Because of the finite box size, the Fourier modes for the field are quantized;

$$\delta^i(x) = \sum_i e^{i k_i x} \delta^i_k,$$

where $i$ indexes the possible modes and $\delta^i_k$ is the Fourier amplitude for the mode at wavevector $k_i$. We can choose the indexing such that $k_{i+} = -k_i$; note that, for the field $\delta^i(x)$ to remain real, $\delta^{i+} = \delta^{i-}$.

The reference ensemble of 300 simulations consists of boxes each with $\delta^i_k$ drawn from a Gaussian, zero-mean probability distribution function (PDF). Decomposed into the magnitude $|\delta^i_k|$ and phase $\theta_i = \arg \delta^i_k$, the PDF for each independent mode $i$ is given by

$$P_{\xi}(|\delta^i_k|, \theta_i) \equiv \frac{|\delta^i_k|}{\pi P_i} \exp \left(-\frac{|\delta^i_k|^2}{P_i}\right),$$

where $P_i$ is the discrete version of the power spectrum $P(k)$. In the fixed-power approach, the PDF for mode $i$ is instead given by

$$P_{\xi_F}(|\delta^i_k|, \theta_i) \equiv \frac{1}{2\pi} \delta_\theta \left(|\delta^i_k| - \sqrt{P_i}\right),$$

where $\delta_\theta$ indicates the Dirac delta-function. One can sample from $P_{\xi_F}$ straightforwardly by setting

$$\delta^i_k = \sqrt{P_i} e^{i \theta_i} \frac{\xi}{\sqrt{P_i}},$$

with $\xi$ drawn with uniform probability between 0 and 2$\pi$, and $\theta_i \sim \theta_i$. The second of the pair of simulations is then generated by transforming $\theta_i \rightarrow \pi + \theta_i$ (Pontzen et al. 2016).

Sampling from $P_{\xi_F}$ results in an ensemble that is not equivalent to sampling from $P_{\xi}$. However, $P_{\xi_F}$ can stand in place of $P_{\xi}$ for many practical calculations (the analytic justification is discussed in Section 3). We verified that, despite the fixed amplitudes, the one-point input overdensity PDF in real space, $\delta^i(x)$, is still a Gaussian deviate owing to the central limit theorem. Furthermore, Fig. 1 shows the distribution of overdensities in the initial conditions at $z = 9$. $\delta^{i0}$, averaged over spheres of $8 h^{-1}$ Mpc radius for a subset of traditional simulations (grey lines) and the two paired-and-fixed (orange lines; these overlap almost perfectly). The corresponding PDF for the combined volume of the two paired-and-fixed simulations is shown by the red dots. There is excellent agreement between this characterization of the density fields of traditional and fixed simulations, with both following a near-Gaussian distribution. The mild skewness (which also agrees between the cases) arises from the 2LPT particle displacements.

2.2 Results

Fig. 2 shows the dark matter power spectrum measured from the $z = 1$ outputs. In the top panel, the results of the fixed pair (red circles) are indistinguishable from the traditional ensemble mean (blue line) over all the scales plotted, confirming that the approach correctly predicts the ensemble average power spectrum in linear and in non-linear regimes.

The bottom panel shows deviations with respect to the ensemble mean, in units of the standard deviation of the ensemble, $\sigma(k) = ((P^{NL}(k) - \langle P^{NL}(k)\rangle)^2)^{1/2}$. On scales where evolution is linear (approximately $k < 0.03 h$ Mpc$^{-1}$) the fixed simulations should exactly coincide with linear theory by construction. As expected, the measured power spectrum agrees with the ensemble mean to an accuracy limited only by the statistical errors of the latter, $\sigma(k)/\sqrt{300} \lesssim 2$ per cent of $P(k)$. At larger $k$, non-linear effects – which depend not only on the initial amplitude of Fourier modes but also on phases – become important. Accordingly, the power spectrum of the two individual fixed simulations (orange lines) drift away from the exact mean. However the leading order deviations from the ensemble mean are equal and opposite in sign (Pontzen et al. 2016) between the pair of fixed simulations, so that their average (red dots) has a reduced rms error much below 1 per cent (0.27$\sigma$ over the range $0.03 < k h$ Mpc$^{-1} < 1$). The accuracy of our pair of fixed simulations by this measure is approximately equivalent to averaging over 14 traditional simulations, allowing for a factor 7 reduction in computer time. In particular, the technique suppresses statistical errors on all scales to the point where they are smaller than the impact of numerical parameters (Schneider et al. 2016).

In Fig. 3, we show that the high accuracy of the method also holds in redshift space. In this figure, we plot the monopole (red circles) and quadrupole (green triangles) terms of an expansion of the 2D correlation function in terms of Legendre polynomials. Predictions from the pair of fixed simulations again agree well with the ensemble mean. The same pattern persists where the individual
Simulations with suppressed variance

Figure 2. The power spectrum of the dark matter at $z = 1$. In the top panel, measurements from the ensemble of 300 traditional simulations are shown as grey lines, with the mean shown by a blue line. The solid red circles show the average of the two simulations in the paired-and-fixed set. Finally, the horizontal dotted line marks the shot noise limit. In the bottom panel, we show the differences with respect to the average ensemble measurement, in units of the standard deviation in the ensemble. As in the top panel, red symbols show the final estimate from the pair of fixed simulations. We additionally show residuals in each of the two individual fixed simulations by the orange lines. The envelopes bounded by dashed lines mark a 1 per cent (left) and 0.1 per cent (right) uncertainty in the power spectrum. The fixed pair produces a power spectrum estimate with an rms error of just 0.27σ on non-linear scales $0.03 < k/h \text{Mpc}^{-1} < 1$.

fixed simulations perform best on large scales, while on smaller scales the pairing leads to a substantial cancellation of remaining errors. The overall technique yields a precise prediction for the non-linear correlation function, reaching a 2 per cent accuracy over the whole range of scales investigated (in particular around the baryonic acoustic oscillation peak, whose shape and location is currently driving large simulation campaigns). With traditional ensemble-average techniques, achieving this accuracy would require around 50 simulations of $3 h^{-1}$ Gpc box size.

Having established the accuracy of our simulations for predicting two-point statistics, we now turn to higher-order clustering. The bispectrum is defined (in the limit that the box size is infinite) by

$$B(k_1, k_2, \theta) \delta_{	ext{NL}}(k_1 + k_2 + k_3) = \langle \delta_{	ext{NL}}(k_1) \delta_{	ext{NL}}(k_2) \delta_{	ext{NL}}(k_3) \rangle,$$

where $\delta_{	ext{NL}}(k)$ is the Fourier transform of the non-linear evolved overdensity. We have assumed statistical isotropy in writing $B$ as a function of $\theta$, the angle between the $k_1$ and $k_2$ vectors, and statistical homogeneity imposes the Dirac-delta dependence on the left-hand side. We particularly consider the case where $k_1 = 0.02 h^{-1}$ Mpc and $k_2 = 0.04 h^{-1}$ Mpc to capture the onset of non-linearity, and plot the reduced bispectrum

$$Q(\theta) = \frac{\hat{B}(k_1, k_2, \theta)}{P_{\text{NL}}(k_1)P_{\text{NL}}(k_2) + P_{\text{NL}}(k_1)\hat{P}_{\text{NL}}(k_3) + \hat{P}_{\text{NL}}(k_2)P_{\text{NL}}(k_3) + \hat{P}_{\text{NL}}(k_1)\hat{P}_{\text{NL}}(k_3)},$$

where $\hat{B}$ is the estimated bispectrum from a simulation. The definition of $Q(\theta)$ divides out much of the sensitivity to the power spectrum realization. Accordingly, when we plot this quantity in Fig. 4, each of the two individual fixed simulations exhibit

Figure 3. Same as Fig. 2 but for the monopole (red circles) and the quadrupole (green triangles) of the redshift-space correlation function. The rms error on the paired-and-fixed result is 0.12σ and 0.17σ for the monopole and quadrupole, respectively, meaning that around 50 traditional simulations are required to reach the accuracy of a fixed pair of simulations.

Figure 4. Same as Fig. 2 but for the reduced bispectrum. The configuration plotted corresponds to triangles with two sides fixed at $k_1 = 0.02 h^{-1}$ Mpc$^{-1}$ and $k_2 = 0.04 h^{-1}$ Mpc$^{-1}$, with their angle ranging from 0 to $\pi$. The rms deviation is 0.14σ.
fluctuations of an amplitude comparable to that in traditional realizations. However, the pairing procedure cancels the leading order contribution to these fluctuations because they have odd parity in the input linear density field. Therefore the final estimate from the pair of fixed simulations has an rms deviation from the ensemble average of only 0.14σ over all time. Reaching this accuracy with traditional simulations would again require averaging over 50 (as opposed to two) realizations.

As discussed in Koda et al. (2016), the COLA (N-body) algorithm does not resolve the internal structure of haloes but none the less predicts accurate mass functions for the overall population. Therefore we can meaningfully test the abundance of collapsed objects. In Fig. 5, we show the mass function of dark matter haloes identified using a Friends-of-Friends algorithm (Davis et al. 1985) with a linking length set to l = 0.2. The new method produces results with suppressed fluctuations relative to two Gaussian simulations, with strong cancellations between the pair. The average rms error is 0.47σ, roughly the level expected from four simulations randomly picked from the traditional ensemble.

3 ANALYTIC EXPLORATION

In the previous section, we showed that paired- and fixed simulations are able to predict the average properties of a traditional ensemble. We will now explore the technique from an analytic perspective. The pairing approach has recently been introduced and discussed elsewhere (Pontzen et al. 2016, see particularly section II.C) and so our focus is on the power spectrum fixing. Sampling from the \( P(k) \)-fixed PDF \( P_f \), defined by equation (3), is not equivalent to sampling from the true Gaussian \( P_g \), equation (2). The aim of this section is therefore to motivate more precisely why \( P_f \) reproduces the ensemble average results of \( P_g \).

Expectation values of any \( n \)-point expression with respect to either \( P_f \) or \( P_g \) will be denoted by \( \langle \delta_i, \ldots, \delta_j \rangle_f \) and \( \langle \delta_i, \ldots, \delta_j \rangle_g \), respectively. In the case of the fixed distribution, we can use expression (4) to write that

\[
\langle \delta_i, \ldots, \delta_j \rangle_f = \frac{\sqrt{P_{i_1} \cdots P_{i_n}}}{(2\pi)^n} \times \int_0^{2\pi} d^n\theta \exp \left( i\theta_{i_1} + \cdots + i\theta_{i_n} \right),
\]

where the integral is over the possible \( \theta \) values for all \( N \) modes.

For \( n = 1 \) the single phase factor \( \exp (i \theta) \) averages to zero, and consequently \( \langle \delta_i \rangle_f = \langle \delta_i \rangle_g = 0 \). This result extends to any \( n \)-point correlation for \( n \) odd; we therefore need only consider the even-\( n \) cases further.

For \( n = 2 \), the properties of the two PDFs are indistinguishable:

\[
\langle \delta_i \delta_j \rangle_f = \langle \delta_i \delta_j \rangle_g = \delta_{i,-j} P_i,
\]

where \( \delta_{i,-j} \) is the Kronecker delta equal to 1 when \( i = -j \) and 0 otherwise, and there is no sum implied over repeated indices. The Gaussian result is standard, and the fixed result is obtained by seeing that when \( i \neq -j \), the \( i \) and \( j \) phase integrals in equation (7) evaluate to zero. For \( n = 4 \), the Gaussian result follows by Wick’s theorem:

\[
\langle \delta_i \delta_j \delta_k \delta_l \rangle_f = \delta_{i,-k} \delta_{j,-l} P_i P_k + \delta_{i,-l} \delta_{j,-k} P_i P_j.
\]

The fixed result, again obtained through the use of (7) is similar to the Gaussian case because indices must be ‘paired up’ for their phase integrals to be non-vanishing. The only difference arises in the case where \( \delta_i \delta_j \delta_k \delta_l = |\delta_i|^4 \); here, the Gaussian result is \(|\langle \delta_i \rangle|^4_g = 3P_i^2 \) but in the fixed case we find that \(|\langle \delta_i \rangle|^4_f = P_i^2 \).

The correction (10) is consistent with how the power spectrum of a fixed realization must have zero variance:

\[
\langle (|\delta_i|^2 - P_i)^2 \rangle_f = \langle (|\delta_i|^2)^2 \rangle_f - 2P_i^2 = 0.
\]

Evidently there is a dramatic difference – intentionally so – between fixed and Gaussian statistics: in the linear regime, the fixed \( P(k) \) approach reproduces the ensemble mean with no variance. While this also means that the input trispectrum is unavoidable non-Gaussian by equation (10), the correction only appears when all indices always take the same value (up to sign). We can now explain why most measures of the output non-linear density field are extremely insensitive to this change.

The non-linear density field can be written in standard perturbation theory (SPT; e.g. Bernard et al. 2002) as

\[
\delta_i^{NL} = \delta_i + \sum_{jk} F_{ij}^{(2)} \delta_j \delta_k + \sum_{ijkl} F_{ijkl}^{(3)} \delta_j \delta_k \delta_l + \cdots,
\]

where \( F^{(n)} \) for \( n = 2, 3, \ldots \) are the discretized version of the SPT kernels which in turn are homogeneous, degree-zero, continuous functions of the wavevectors. As a concrete example of an observable correlation in this formalism, we can consider the one-loop SPT non-linear power spectrum with Gaussian statistics:

\[
P_i^{NL-2} \equiv \langle \delta_i^{NL} \delta_i^{NL} \rangle_g \simeq P_i + \sum_{jklm} \left( F_{ij}^{(2)} F_{ik}^{(2)} + 2 \delta_{i,-j} \delta_{i,-k} F_{ijkl}^{(3)} \right) \times \left( \delta_{j,-k} \delta_{l,-m} P_j P_k + \delta_{j,-l} \delta_{k,-m} P_j P_k \right).
\]
Momentum conservation implicit in the $F^{(n)}$s and explicit in the
Kronecker deltas eliminate three of the summations, so that
the overall summation is over just one index. Therefore the magnitude of
the one-loop terms scales proportionally to $N_i P(k)$ where $k$ is
a characteristic scale and $N_i$ is the number of modes around that
scale (as defined by the range of modes over which the relevant $F$

is large). In a continuum limit (i.e. as the box size $L \to \infty$), $N_i$ turns
into the appropriate Fourier-space volume. These simple scaling
behaviours are assured by the degree-zero homogeneity of the $F^{(n)}$

functions.

In the fixed case, expression (13) must be corrected by using
relation (10), giving

$$P^{NL} = P^{NL-g} - 12 F^{(3)}_{ijj} P_i^2 - 2 \sum_j F^{(2)}_{ijj} F_{ijj} P_j^2,$$

(14)

which is valid at one-loop order for the case $k_i \neq 0$. Here most of
the Kronecker delmas have already been summed out; the remaining
summation, by momentum conservation in $F^{(2)}$, only has a contri-
bution at the index $j$ with $k_j = k_i/2$. The overall contribution of
the correction (14) is therefore suppressed relative to the physical
terms in equation (13) by $O(N_k)$.

For the bispectrum with Gaussian statistics, we have

$$B^g_{ijk} = \langle \delta^{NL}_i \delta^{NL}_j \delta^{NL}_k \rangle_g \approx 2 F^{(2)}_{ijj} P_i P_j + \delta_{ij} \delta_{jk} \sum_l F^{(2)}_{lij} P_l + \text{cyc. perms in } ijk$$

(15)

to one-loop order. The second term contributes only for $k_i \Rightarrow 0$. The
correction is now

$$B^f_{ijk} = B^g_{ijk} - 4 F^{(2)}_{ijj} P_i^2 \delta_{ijk}$$

$$- 2 F^{(2)}_{ijj} P_i^2 \delta_{ijk} - \text{cyc. perms in } ijk,$$

(16)

and is non-zero only in the case where $k_j = k_k = -k_i/2$ or $k_i = 0$
(or a cyclic permutation of those configurations). All other bispectra
are unaffected by the changed statistics at this order.

For higher order perturbation theory (or higher $n$ correlations)
the overall pattern established here will remain: the linear $n$-point
correction term (10) will always involve at least one extra Kronecker
delta relative to the physical part (9). For observable correlations,
this implies that either the effect is diluted by a power of a large
factor $N_k$ (as in the case of the one-loop power spectrum) or plays
a role only in a measure-zero part of the continuous function being
studied (as in the case of the one-loop bispectrum).

4 CONCLUSIONS

In this Letter, we have explored a new method to suppress the impact
of undersampling Fourier modes in simulations.

By fixing the initial amplitude of Fourier modes to the ensemble
mean, variance has been eliminated on linear scales. In the
non-linear regime, the suppression is imperfect because phase-
correlation effects begin to impact on the evolved amplitudes.
However by also pairing the simulation with a phase-reversed counterpart
we can average away the leading order imperfections of this type.

We have tested the non-linear dark matter power spectrum, the
deltapoles of the redshift-space correlation function, the reduced
bispectrum and the halo mass function. In all cases, the method is
unbiased (up to the accuracy of our comparison ensemble averages)
and strongly suppresses unwanted variance. These tests were carried
out with a suite of 300 COLA simulations at $z = 1$. The analytic argu-
ments of Section 3 suggest that the accuracy of the results should
be maintained to all redshifts. Similarly, we do not expect results to

change when using more accurate integration methods than COLA,
especially since small-scale gravitational collapse are largely insen-
tive to large-scale correlations. All these points deserve systematic
investigation in future.

Paired simulations can be used with purely Gaussian initial condi-
tions if desired, retaining many of the small-scale benefits we have
discussed. Conversely, single unpaired simulations with fixed
amplitudes can be used, retaining the large-scale benefits. Whenever
fixing is applied, the ensemble statistics are not strictly Gaussian.
The local one-point PDF is, however, unaffected (Fig. 1) and
furthermore our numerical results directly show that a variety of statis-
tics attain the correct, unbiased ensemble mean value. We gave an
analytic discussion of why the non-Gaussianity does not impinge,
arguing that the errors are either strongly suppressed by the large
density of modes or affect only a measure-zero set of correlations.
Fixing the power spectrum does need to be approached with care
but our results underline that it can be a valuable technique.

Straightforward applications are in any comparison to analytic
models, in characterization of the performance of data modelling,
in emulators and in development of fitting functions for non-linear
statistics. It will be particularly valuable to couple the technique to
high-resolution simulations incorporating baryonic effects to mea-
sure galaxy bias, free of the usual difficulties of large-scale vari-
ance. Furthermore the method could be used in combination with
rescaling techniques to quickly predict galaxy clustering statistics as
a function of cosmological parameters (Angulo & White 2010).
All these are crucial steps towards a comprehensive exploitation of
upcoming survey data.

ACKNOWLEDGEMENTS

We thank Jonás Chavez-Montero for providing us with access to the
ensemble of COLA simuations. We thank the Lorentz Center
where this study was initiated, and Oliver Hahn, Carlos Hernandez-
Monteagudo, Aseem Paranjape, Hiranya Peiris, Anže Slosar, and
Matteo Viel for helpful discussions. REA acknowledges support
from AYA2015-66211-C2-2. AP is supported by the Royal Society.

REFERENCES

Angulo R. E., Hilbert S., 2015, MNRAS, 448, 364
Angulo R. E., White S. D. M., 2010, MNRAS, 401, 1796
Bagla J. S., Prasad J., Khandai N., 2009, MNRAS, 395, 918
Baldi M., Villacescusa-Navarro F., Viel M., Puchwein E., Springel V.,
Moscardini L., 2014, MNRAS, 440, 75
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep.,
367, 1
Carlson J., White M., Padmanabhan N., 2009, Phys. Rev. D, 80, 043531
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Howlett C., Manera M., Percival W. J., 2015, Astron. Comput., 12, 109
Koda J., Blake C., Beutler F., Kazin E., Marin F., 2016, MNRAS, 459, 2118
Kuhlen M., Vogelsberger M., Angulo R., 2012, Phys. Dark Universe, 1, 50
Manera M. et al., 2015, MNRAS, 447, 437
Neyrinck M. C., Yang L. F., 2013, MNRAS, 433, 1628
Overzier R. A., Guo Q., Kauffmann G., De Lucia G., Bouwens R., Lemson G.,
2009, MNRAS, 394, 577
Pontzen A., Slosar A., Roth N., Peiris H. V., 2016, Phys. Rev. D, 93, 103519
Rasera Y., Corasaniti P.-S., Alimi J.-M., Bouillot V., Revery V., Balmès L.,
2014, MNRAS, 440, 1420
Schneider A. et al., 2016, J. Cosmol. Astropart. Phys., 4, 047
Springel V., 2005, MNRAS, 364, 1105
Takahashi R. et al., 2009, ApJ, 700, 479
Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmol. Astropart.
Phys., 6, 036

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