

Effects of the size of cosmological N -body simulations on physical quantities – III. Skewness

J. S. Bagla,¹★ Jayanti Prasad²★ and Nishikanta Khandai¹★

¹Harish-Chandra Research Institute, Chhatnag Road, Jhusi, Allahabad 211019, India

²National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Post Bag 3, Ganeshkhind, Pune 411007, INDIA

Accepted 2009 January 30. Received 2009 January 30; in original form 2008 April 21

ABSTRACT

N -body simulations are an important tool in the study of formation of large-scale structures. Much of the progress in understanding the physics of galaxy clustering and comparison with observations would not have been possible without N -body simulations. Given the importance of this tool, it is essential to understand its limitations as ignoring these can easily lead to interesting but unreliable results. In this paper, we study the limitations due to the finite size of the simulation volume. In an earlier work, we proposed a formalism for estimating the effects of a finite box size on physical quantities and applied it to estimate the effect on the amplitude of clustering, mass function. Here, we extend the same analysis and estimate the effect on skewness and kurtosis in the perturbative regime. We also test the analytical predictions from the earlier work as well as those presented in this paper. We find good agreement between the analytical models and simulations for the two-point correlation function and skewness. We also discuss the effect of a finite box size on relative velocity statistics and find the effects for these quantities scale in a manner that retains the dependence on the averaged correlation function $\bar{\xi}$.

Key words: gravitation – methods: N -body simulations – methods: numerical – cosmology: theory – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Large-scale structures like galaxies and clusters of galaxies are believed to have formed by gravitational amplification of small perturbations. For an overview and original references, see e.g. Peebles (1980), Peacock (1999), Padmanabhan (2002) and Bernardeau et al. (2002). Density perturbations are present at all scales that have been observed (Percival et al. 2007; Komatsu et al. 2009). Understanding the evolution of density perturbations for systems that have fluctuations at all scales is essential for the study of galaxy formation and large-scale structures. The equations that describe the evolution of density perturbations in an expanding universe have been known for a long time (Peebles 1974), and these are easy to solve when the amplitude of perturbations is much smaller than unity. These equations describe the evolution of density contrast defined as $\delta(\mathbf{r}, t) = [\rho(\mathbf{r}, t) - \bar{\rho}(t)]/\bar{\rho}(t)$. Here, $\rho(\mathbf{r}, t)$ is the density at \mathbf{r} at time t and $\bar{\rho}$ is the average density in the universe at that time. These are densities of non-relativistic matter, the component that clusters at all scales and is believed to drive the formation of large-scale structures in the universe. Once the density contrast at

relevant scales becomes large, i.e. $|\delta| \geq 1$, the perturbation becomes non-linear and coupling with perturbations at other scales cannot be ignored. The equations that describe the evolution of density perturbations cannot be solved for generic perturbations in this regime. N -body simulations (Bagla & Padmanabhan 1997b; Bertschinger 1998; Bagla 2005; Dolag et al. 2008) are often used to study the evolution in this regime. Alternative approaches can be used if one requires only a limited amount of information and in such a case either quasi-linear approximation schemes (Zel'dovich 1970; Gurbatov, Saichev & Shandarin 1989; Matarrese et al. 1992; Brainerd, Scherrer & Villumsen 1993; Bagla & Padmanabhan 1994; Sahni & Coles 1995; Hui & Bertschinger 1996; Bernardeau et al. 2002) or scaling relations (Davis & Peebles 1977; Hamilton et al. 1991; Jain, Mo & White 1995; Kanekar 2000; Nityananda & Padmanabhan 1994; Peacock & Dodds 1994; Padmanabhan 1996; Padmanabhan et al. 1996; Peacock & Dodds 1996; Ma 1998; Smith et al. 2003) suffice.

In cosmological N -body simulations, we simulate a representative region of the universe. This is a large but finite volume and periodic boundary conditions are often used. Almost always, the simulation volume is taken to be a cube. Effect of perturbations at scales smaller than the mass resolution of the simulation and of perturbations at scales larger than the box is ignored. Indeed, even perturbations at scales comparable to the box are under sampled.

*E-mail: jasjeet@hri.res.in (JSB); jayanti@nra.tifr.res.in (JP); nishi@hri.res.in (NK)

It has been shown that perturbations at small scales do not influence collapse of perturbations at much larger scales in a significant manner (Peebles 1974, 1985; Little, Weinberg & Park 1991; Bagla & Padmanabhan 1997a; Couchman & Peebles 1998; Bagla, Prasad & Ray 2005; Bagla & Prasad 2009). This is certainly true if the scales of interest are in the non-linear regime (Bagla & Padmanabhan 1997a; Bagla & Prasad 2009). Therefore, we may assume that ignoring perturbations at scales much smaller than the scales of interest does not affect results of N -body simulations.

Perturbations at scales larger than the simulation volume can affect the results of N -body simulations. Use of the periodic boundary conditions implies that the average density in the simulation box is same as the average density in the universe, in other words we ignore perturbations at the scale of the simulation volume (and at larger scales). Therefore, the size of the simulation volume should be chosen so that the amplitude of fluctuations at the box scale (and at larger scales) is ignorable. If the amplitude of perturbations at larger scales is not ignorable then clearly the simulation is not a faithful representation of the model being studied. It is not obvious as to when fluctuations at larger scales can be considered ignorable, indeed the answer to this question depends on the physical quantity of interest, the model being studied and the specific length/mass scale of interest as well.

The effect of a finite box size has been studied using N -body simulations and the conclusions in this regard may be summarized as follows.

(i) If the amplitude of density perturbations around the box scale is small ($\delta < 1$) but not much smaller than unity, simulations underestimate the correlation function though the number density of small mass haloes does not change by much (Gelb & Bertschinger 1994a,b). In other words, the formation of small haloes is not disturbed but their distribution is affected by non-inclusion of long wave modes.

(ii) In the same situation, the number density of the most massive haloes drops significantly (Gelb & Bertschinger 1994a,b; Bagla & Ray 2005).

(iii) Effects of a finite box size modify values of physical quantities like the correlation function even at scales much smaller than the simulation volume (Bagla & Ray 2005).

(iv) The void spectrum is also affected by finite size of the simulation volume if perturbations at large scales are not ignorable (Kauffmann & Melott 1992).

(v) It has been shown that properties of a given halo can change significantly as the contribution of perturbations at large scales is removed to the initial conditions but the distribution of most internal properties remains unchanged (Power & Knebe 2006).

(vi) We presented a formalism for estimating the effects of a finite box size in Bagla & Prasad (2006). We used the formalism to estimate the effects on the rms amplitude of fluctuations in density, as well as the two-point correlation function. We used these to further estimate the effects on the mass function and the multiplicity function.

(vii) The formalism mentioned above was used to estimate changes in the formation and destruction rates of haloes (Prasad 2007).

(viii) It was pointed out that the second-order perturbation theory and corrections arising due to this can be used to estimate the effects due to a finite box size (Takahashi et al. 2008). This study focused specifically on the effects on baryon acoustic oscillations.

(ix) If the objects of interest are collapsed haloes that correspond to rare peaks, as in the study of the early phase of reionization, we

require a fairly large simulation volume to construct a representative sample of the universe (Barkana & Loeb 2004; Reed et al. 2008).

In some cases, one may be able to devise a method to ‘correct’ for the effects of a finite box size (Colombi, Bouchet & Schaeffer 1994), but such methods cannot be generalized to all statistical measures or physical quantities.

Effects of a finite box size modify values of physical quantities even at scales much smaller than the simulation volume (Bagla & Ray 2005; Bagla & Prasad 2006). A workaround for this problem was suggested in the form of an ensemble of simulations to take the effect of convergence due to long wave modes into account (Sirk 2005), the effects of shear due to long wave modes are ignored here. However, it is not clear whether the approach where an ensemble of simulations is used has significant advantages over using a sufficiently large simulation volume.

We review the basic formalism we proposed in (Bagla & Prasad 2006) in Section 2. We then extend the original formalism to the cases of non-linear amplitude of clustering and also for estimating changes in skewness and other reduced moments of counts in cells. This is done in Section 3. In Section 4, we confront our analytical models with N -body simulations. We end with a discussion in Section 5.

2 THE FORMALISM

Initial conditions for N -body simulations are often taken to be a realization of a Gaussian random field with a given power spectrum (for details, see e.g. Bagla & Padmanabhan 1997b; Bertschinger 1998; Bagla 2005; Dolag et al. 2008). The power spectrum is sampled at discrete points in the k -space between the scales corresponding to the box size (fundamental mode) and the grid size (Nyquist frequency/mode). Here, k is the wave vector.

We illustrate our approach using rms fluctuations in mass $\sigma(r)$, but as shown below, the basic approach can be generalized to any other quantity in a straightforward manner. In general, $\sigma(r)$ may be defined as follows:

$$\sigma^2(r) = \int_0^\infty \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} W^2(kr). \quad (1)$$

Here, $P(k)$ is the power spectrum of density contrast, r is the comoving length-scale at which rms fluctuations are defined, $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ is the wave number and $W(kr)$ is the Fourier transform of the window function used for sampling the density field. The window function may be a Gaussian or a step function in real or k -space. We choose to work with a step function in real space where $W(kr) = 9(\sin kr - kr \cos kr)^2/(k^6 r^6)$ (see e.g. section 5.4 of Padmanabhan (1993) for further details). In an N -body simulation, the power spectrum is sampled only at specified points in the k -space. In this case, we may write $\sigma^2(r)$ as a sum over these points.

$$\begin{aligned} \sigma^2(r, L_{\text{box}}) &= \frac{9}{V} \sum_k P(k) \left[\frac{\sin kr - kr \cos kr}{k^3 r^3} \right]^2 \\ &\approx \int_{2\pi/L_{\text{box}}}^{2\pi/L_{\text{grid}}} \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} 9 \left[\frac{\sin kr - kr \cos kr}{k^3 r^3} \right]^2 \\ &\approx \int_{2\pi/L_{\text{box}}}^\infty \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} 9 \left[\frac{\sin kr - kr \cos kr}{k^3 r^3} \right]^2 \\ &= \int_0^\infty \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} 9 \left[\frac{\sin kr - kr \cos kr}{k^3 r^3} \right]^2 \end{aligned}$$

$$-\int_0^{2\pi/L_{\text{box}}} \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} 9 \left[\frac{\sin kr - kr \cos kr}{k^3 r^3} \right]^2 = \sigma_0^2(r) - \sigma_1^2(r, L_{\text{box}}). \quad (2)$$

Here, $\sigma_0^2(r)$ is the expected level of fluctuations in mass at scale r for the given power spectrum and $\sigma_1^2(r, L_{\text{box}})$ is what we get in an N -body simulation at early times. We have assumed that we can approximate the sum over the k modes sampled in initial conditions by an integral. Further, we make use of the fact that small scales do not influence large scales to ignore the error contributed by the upper limit of the integral. This approximation is valid as long as the scales of interest are more than a few grid lengths.

In the approach outlined above, the value of σ^2 at a given scale is expressed as a combination of the expected value σ_0^2 and the correction due to the finite box size σ_1^2 . Here, σ_0^2 is independent of the box size and depends only on the power spectrum and the scale of interest. It is clear that $\sigma^2(r, L_{\text{box}}) \leq \sigma_0^2(r)$ and $\sigma_1^2(r, L_{\text{box}}) \geq 0$. It can also be shown that for hierarchical models, $d\sigma_1^2(r, L_{\text{box}})/dr \leq 0$, i.e. $\sigma_1^2(r, L_{\text{box}})$ increases or saturates to a constant value as we approach small r .

At large scales, $\sigma_0^2(r)$ and $\sigma_1^2(r, L_{\text{box}})$ have a similar magnitude, and the rms fluctuations in the simulation become negligible as compared to the expected values in the model. As we approach small r , the correction term $\sigma_1^2(r, L_{\text{box}})$ is constant and for most models it becomes insignificant in comparison with $\sigma_0^2(r)$. In models where $\sigma_0^2(r)$ increases very slowly at small scales or saturates to a constant value, the correction term σ_1^2 can be significant at all scales.

This formalism can be used to estimate corrections for other estimators of clustering, e.g. the two-point correlation (see Bagla & Prasad (2006) for details).

The estimation of the rms amplitude of density perturbations allows us to use the theory of mass function and estimate a number of quantities of interest. For details, we again refer the reader to Bagla & Prasad (2006) but we list important points here.

(i) The fraction of mass in collapsed haloes is underestimated in N -body simulations. This underestimation is most severe near the scale of non-linearity, and falls off on either side. If we consider fractional underestimation in the collapsed fraction then this increases monotonically from small to large scales.

(ii) The number density of collapsed haloes is underestimated at scales larger than the scale of non-linearity. The maximum in collapsed fraction near the scale of non-linearity leads to a change of sign in the effect of a finite box size for the number density of haloes at this scale: at smaller scales, the number density of haloes is overestimated in simulations. This can be understood on the basis of a paucity of mergers that otherwise would have led to formation of high-mass haloes.

(iii) The above conclusions are generic and do not depend on the specific model for mass function. Indeed, expressions for both the Press–Schechter (Press & Schechter 1974) and the Sheth–Tormen (Sheth & Tormen 1999; Sheth, Mo, & Tormen 2001) mass functions are given in Bagla & Prasad (2006), and we have also checked the veracity of our claims for the Jenkins et al. (2001) mass function.

3 REDUCED MOMENTS

In this section, we outline how the formalism and results outlined above may be used to estimate the effect of a finite box size on reduced moments. Reduced moments like the skewness and kurtosis can be computed using the perturbation theory in the weakly non-linear regime (Bernardeau 1994). The expected values of the

reduced moments are related primarily to the slope of the initial or linearly extrapolated $\sigma^2(r)$, as all non-Gaussianities are generated through evolution of the Gaussian initial conditions and the initial $\sigma^2(r)$ characterizes this completely. We can use the expression for $\sigma^2(r)$ as it is realized in simulations with a finite box size to compute the expected values of reduced moments in N -body simulations in the weakly non-linear regime:

$$\begin{aligned} S_3 &= \frac{34}{7} + \frac{\partial \ln \sigma^2}{\partial \ln r} \\ &= \frac{34}{7} + \frac{\partial \ln (\sigma_0^2 - \sigma_1^2)}{\partial \ln r} \\ &= \frac{34}{7} + \frac{\partial \ln \sigma_0^2}{\partial \ln r} + \frac{\partial \ln (1 - \sigma_1^2/\sigma_0^2)}{\partial \ln r} \\ &= S_{30} - S_{31}. \end{aligned} \quad (3)$$

S_{30} is the expected value of S_3 for the given mode, i.e. when there are no box corrections and S_{31} is the correction term in S_3 due to a finite box size. Box-size effects lead to a change in slope of σ^2 , and hence the effective value of n changes. The last term is the offset in skewness in N -body simulations as compared with the expected values in the model being simulated. We would like to emphasize that this expression is valid only in the weakly non-linear regime.

In general, we expect σ_1^2/σ_0^2 to increase as we go to larger scales. Thus the skewness is underestimated in N -body simulations, and the level of under estimation depends on the slope of σ_1^2/σ_0^2 as compared to the slope of σ_0^2 . In the limit of small scales where σ_1^2 is almost independent of scale, we find that the correction is

$$\begin{aligned} S_3 &= \frac{34}{7} + \frac{\partial \ln \sigma_0^2}{\partial \ln r} + \frac{\partial \ln (1 - \sigma_1^2/\sigma_0^2)}{\partial \ln r} \\ &\simeq \frac{34}{7} - (n+3) - \frac{\partial (\sigma_1^2/\sigma_0^2)}{\partial \ln r} + \mathcal{O} \left[\frac{\partial}{\partial \ln r} \left(\frac{\sigma_1^2}{\sigma_0^2} \right)^2 \right] \\ &\simeq \frac{34}{7} - (n+3) \left[1 + \frac{\sigma_1^2}{\sigma_0^2} \right] + \mathcal{O} \left[\frac{\partial}{\partial \ln r} \left(\frac{\sigma_1^2}{\sigma_0^2} \right)^2 \right] \\ &\quad + \mathcal{O} \left(\frac{1}{\sigma_0^2} \frac{\partial \sigma_1^2}{\partial \ln r} \right). \end{aligned} \quad (4)$$

Here, n is the index of the initial spectrum we are simulating. For non-power-law models, this will also be a function of scale. The correction becomes more significant at larger scales and the net effect, as noted above, is to underestimate S_3 .

Similar expressions can be written down for kurtosis and other reduced moments using the approach outlined above. We give the expression for kurtosis below, but do not compute further moments as the same general principle can be used to compute these as well:

$$\begin{aligned} S_4 &= \frac{6071}{1323} + \frac{62}{3} \frac{\partial \ln \sigma^2}{\partial \ln r} \\ &\quad + \frac{7}{3} \left[\frac{\partial \ln \sigma^2}{\partial \ln r} \right]^2 + 2 \frac{\partial^2 \ln \sigma^2}{\partial \ln^2 r} \\ &\simeq \frac{6071}{1323} - \frac{62}{3}(n+3) \left[1 + \frac{\sigma_1^2}{\sigma_0^2} \right] + \frac{7}{3}(n+3)^2 \left[1 - \frac{8}{7} \frac{\sigma_1^2}{\sigma_0^2} \right] \\ &\quad + \mathcal{O} \left[\frac{\partial}{\partial \ln r} \left(\frac{\sigma_1^2}{\sigma_0^2} \right)^2 \right] + \mathcal{O} \left(\frac{1}{\sigma_0^2} \frac{\partial \sigma_1^2}{\partial \ln r} \right). \end{aligned} \quad (5)$$

4 N-BODY SIMULATIONS

In this section, we compare the analytical estimates for finite box-size effects for various quantities with N -body simulations. Such a

Table 1. This table lists characteristics of N -body simulations used in our study.

Model	Description	Cut off (k_c)
A1	Power law, $n = -2.0$	k_f
A2	Power law, $n = -2.0$	$2k_f$
A3	Power law, $n = -2.0$	$4k_f$
B1	Power law, $n = -2.5$	k_f
B2	Power law, $n = -2.5$	$2k_f$
B3	Power law, $n = -2.5$	$4k_f$
C1	Λ CDM, WMAP-5 BF, $L_{\text{box}} = 160 h^{-1}$ Mpc	k_f
C2	Λ CDM, WMAP-5 BF, $L_{\text{box}} = 160 h^{-1}$ Mpc	$2k_f$
C3	Λ CDM, WMAP-5 BF, $L_{\text{box}} = 160 h^{-1}$ Mpc	$4k_f$

Note. Here, the spectral index gives the slope of the initial power spectrum, and the cut-off refers to the wavenumber below which all perturbations are set to zero: $k_f = 2\pi/L_{\text{box}}$ is the fundamental wave mode for the simulation box. All models were simulated using the TREEM code (Bagla 2002; Bagla & Ray 2003; Khandai & Bagla 2008). 256^3 particles were used in each simulation, and the PM calculations were done on a 256^3 grid. Power spectra for both the A and the B series of simulations were normalized to ensure $\sigma = 1$ at the scale of 8 grid lengths at the final epoch if there is no box-size cut-off. A softening length of 0.1 grid lengths was used as the evolution of small-scale features is not of interest in this study. Simulations for both the A and the B series were done with the Einstein-deSitter background and the C series used the WMAP-5 best-fitting (BF) model as the cosmological background, as also for the power spectrum and transfer function.

comparison is relevant in order to test the effectiveness of approximations made in computing the effects of a finite box size. We have made the following approximations.

(i) Effects of mode coupling between the scales that are taken into account in a simulation and the modes corresponding to scales larger than the simulation box are ignored. We believe that this should not be important unless the initial power spectrum has a sharp feature at scales comparable with the simulation size.¹

(ii) Sampling of modes comparable to the box size is sparse, and the approximation of the sum over wave modes as an integral can be poor if the relative contribution of these scales to σ_1 is significant.

Table 1 gives details of the N -body simulations used in this paper. In order to simulate the effects of a finite box size, we used the method employed by Bagla & Ray (2005) where initial perturbations are set to zero for all modes with wavenumber smaller than a given cut-off k_c . The initial conditions are exactly the same as the reference simulation in each series in all other respects. For a finite simulation box, there is a natural cut-off at the fundamental wavenumber $k_f = 2\pi/L_{\text{box}}$ and simulations A1, B1 and C1 impose no other cut-off. These are the reference simulations for the two series of simulations. Simulations A2, B2 and C2 sample perturbations at wavenumbers are larger than $2k_f$ whereas simulations A3, B3 and C3 are more restrictive with non-zero perturbations above $4k_f$. The cut-off of $2k_f$ and $4k_f$ corresponds to scales of 128 and 64 grid lengths, respectively. For the C series of simulations, the cut-off of $2k_f$ and $4k_f$ corresponds to scales of 80 and $40 h^{-1}$ Mpc, respectively.

¹ For example, simulations of baryon acoustic oscillations imprinted in the matter power spectrum may be affected by mode coupling even though the amplitude of fluctuations at the relevant scales is very small (Peebles 1974; Bagla & Padmanabhan 1997a; Takahashi et al. 2008).

The background cosmology was taken to be Einstein-deSitter for the A and B series simulations. The best-fitting Λ cold dark matter (Λ CDM) model from *Wilkinson Microwave Anisotropy Probe 5* (WMAP-5) (Dunkley et al. 2009) was used for the C series of simulations.

In order to ensure that the initial conditions do not get a rare contribution from a large-scale mode, we forced $|\delta_k|^2 = P(k)$ while keeping the phases random for modes $k \geq 6k_f$.

We have chosen to work with models where box-size effects are likely to be significant, particularly with the larger cut-off in wavenumber. This has been done to test our analytical model in a severe situation, and also to further illustrate the difficulties in simulating models with large negative indices.

We present results from N -body simulations in the following section.

4.1 Results

We begin with a visual representation of the simulations. Fig. 1 shows a slice from simulations A1, A2 and A3 at two different epochs. The left-hand panel is for the early epoch when $r_{\text{nl}} = 2$ grid lengths in the model without a cut-off, and the right-hand panel is for $r_{\text{nl}} = 8$ grid lengths. The top row is for the simulation A1, the middle row is for the simulation A2 and the lowest row is for the simulation A3. The relevance of box-size effects is apparent as the large-scale structure in the three simulations is very different even at the early epoch when $r_{\text{nl}} = 2$, much smaller than the effective box size for these simulations. Disagreement between different simulations becomes even more severe as we go to the later epoch with $r_{\text{nl}} = 8$ grid lengths.

Visual appearance for simulations B1, B2 and B3, shown in Fig. 2, follows the same pattern. In this case, the spectral index is closer to -3 than for simulations of the A series shown in Fig. 1, hence the larger scale modes are more important for evolution of perturbations even at small scales. It is interesting to note that the largest underdense region in simulation B1 at early times is already comparable to the box size, and hence we require $L_{\text{box}}/r_{\text{nl}} \gg 128$ for the effects of a finite box size to be small enough to be ignored for simulations of the power-law model with $n = -2.5$. This constraint is even stronger for models with the slope of the power spectrum closer to $n = -3$.

Fig. 3 shows the visual appearance for the C series of simulations. Once again we find a significant change in appearance even with $k_c = 2k_f$ at the earlier epoch, $z = 1$ in this case. This indicates that a box size of $80 h^{-1}$ Mpc is insufficient if we wish to achieve convergence in the large-scale distribution of matter in models of cosmological interest. This reinforces conclusions of Bagla & Ray (2005) where we found that a box size of around $150 h^{-1}$ Mpc is required for convergence in simulations of Λ CDM models.

4.2 Clustering amplitude

The left-hand column in Fig. 4 shows the volume-averaged correlation function ξ for the simulations being studied here. The top-left panel is for simulations A1, A2 and A3 at an early epoch ($r_{\text{nl}} = 2$ grid lengths in the model without a cut-off), and the second panel from top in this column shows ξ for the same simulations at a late epoch ($r_{\text{nl}} = 8$ grid lengths). The corresponding plots in the second column show the same for the simulations in the B series, and the third column is for the C series of simulations. We have shown ξ as a function of scale in these panels. Also shown are the linearly extrapolated values of ξ computed using our formalism for

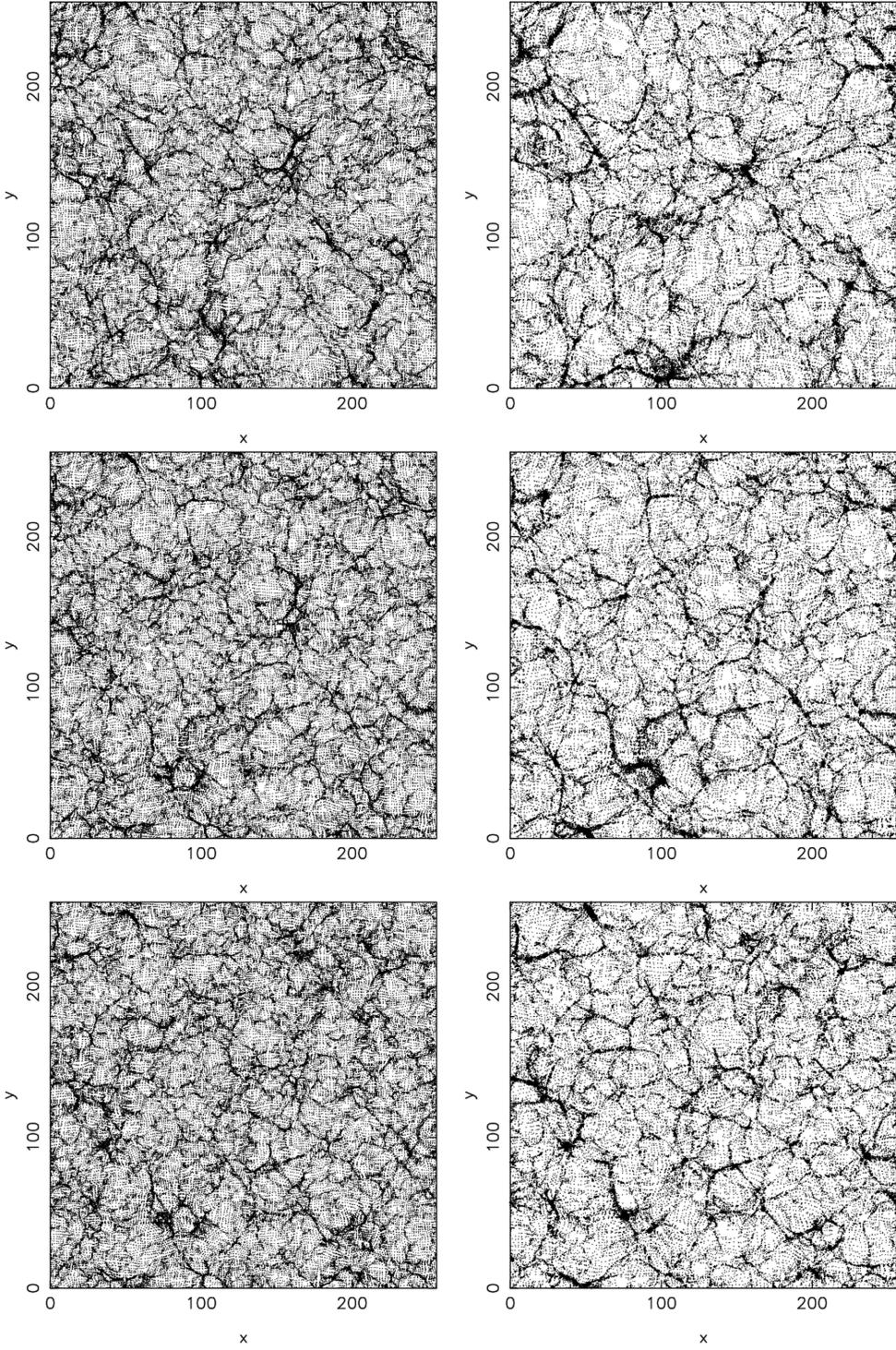


Figure 1. The first, second and third rows in this figure show the slices for models A1, A2 and A3 (see table for details), respectively, at an early epoch when the scale of non-linearity is 2 grid lengths (left-hand column) and a later epoch when the scale of non-linearity is 8 grid lengths (right-hand column).

estimating the effects of a finite box size. Data from N -body simulations are shown as thick curves whereas the theoretical estimate is shown as thin curves with the corresponding line style. It is clear that the analytical estimate for ξ in a finite box captures the qualitative nature of the change from the expected values. The match is better at large scales where ξ is small, and this is expected as the analytical estimate is linearly extrapolated whereas we are comparing it with results from an N -body simulation. Our analysis works

better for the $n = -2$ model used in the A series of simulations and for the Λ CDM model in the C series of simulations as compared to the B series of simulations for the $n = -2.5$ model where it systematically underestimates the suppression of ξ . It is noteworthy that even in this case the differences between the simulation and the analytical model at large scales is of the order of 20–30 per cent whereas the box-size effect changes the clustering amplitude by more than an order of magnitude at some scales. Thus we may state

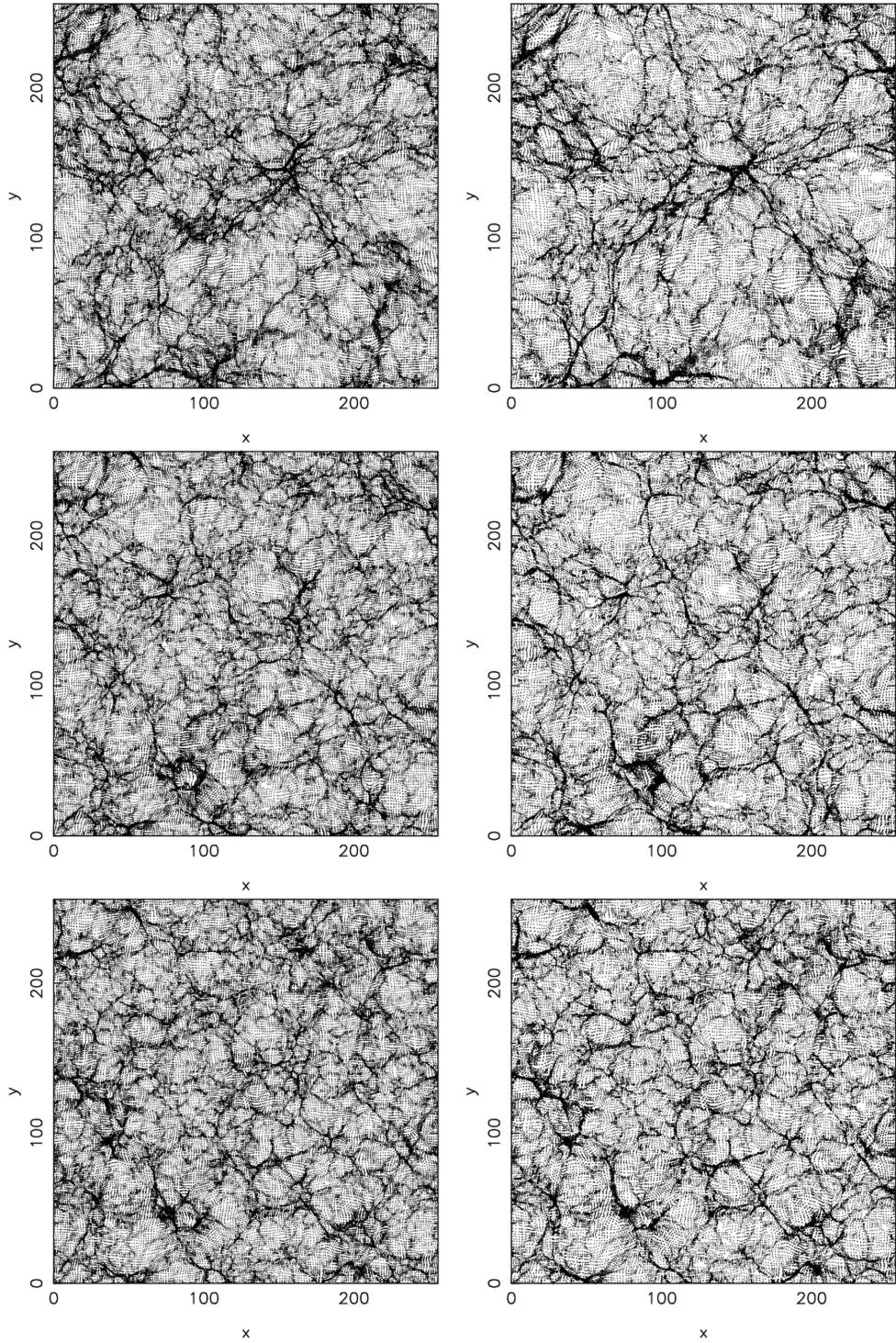


Figure 2. The first, second and third rows in this figure show the slices for the models B1, B2 and B3, respectively, at an early epoch when the scale of non-linearity is 2 grid lengths (left-hand column) and a later epoch (right-hand column) when the scale of non-linearity is 8 grid lengths (right-hand column).

that the model captures the essence of the box-size effects at large scales.

4.3 Skewness

The lower two rows in Fig. 4 show the corresponding plots for S_3 , shown here as a function of scale. Apart from the lines that show S_3 from simulations (thick lines) and our analytical estimate for the

weakly non-linear regime (thin lines), we also show the value of S_3 expected in the weakly non-linear regime in absence of any finite box-size effects for the three series of simulations. The analytical estimate of S_3 , computed using equation (3), matches well with the values in N -body simulation at large scales. It is noteworthy that the match between the two is better for a larger cut-off in wavenumbers. We believe that this is due to sparse sampling of the initial power spectrum at scales comparable to the box size and due to this our

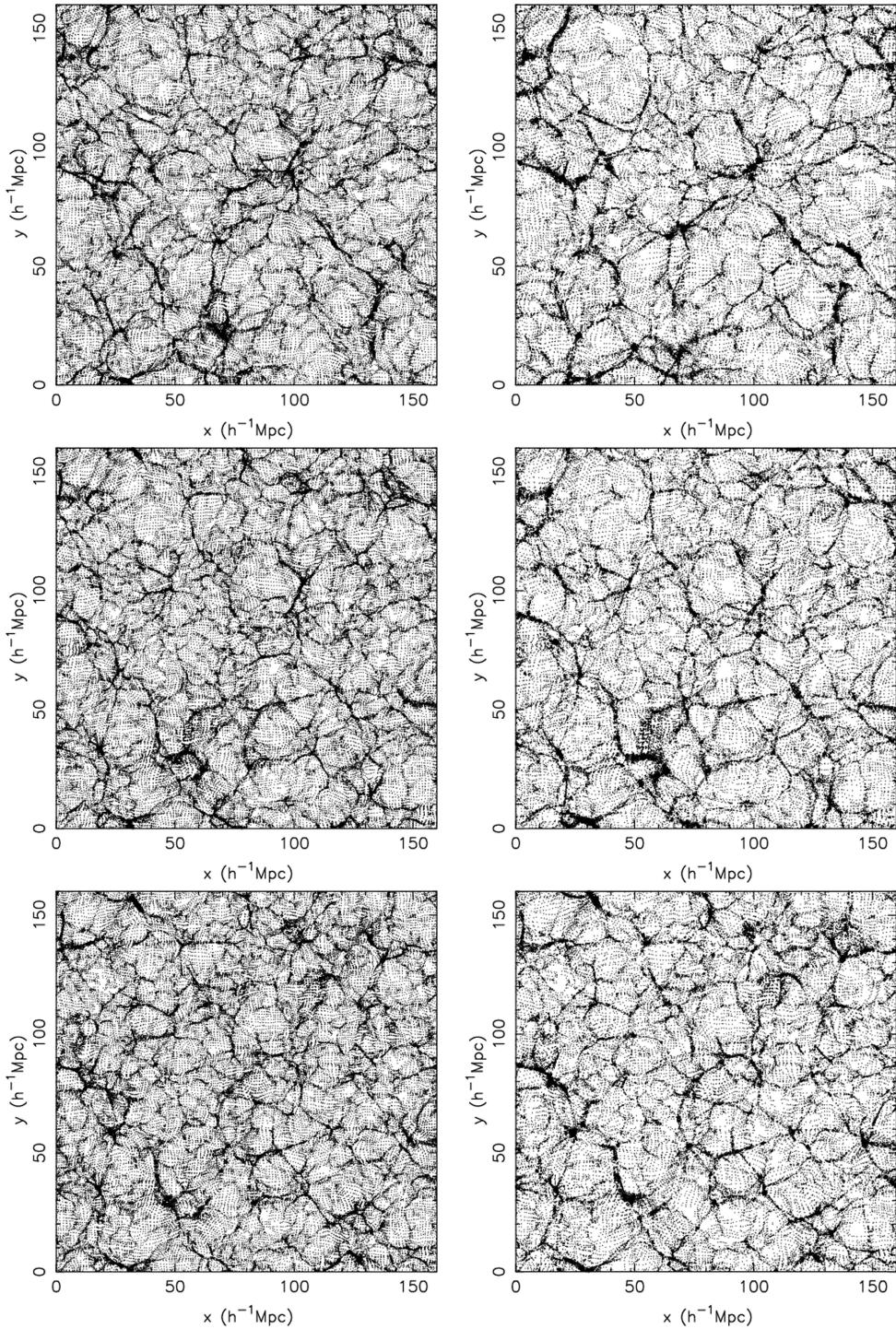


Figure 3. The first, second and third rows in this figure show the slices for the Λ CDM simulations C1, C2 and C3, respectively, at an early epoch ($z = 1$) (left-hand column) and the present epoch ($z = 0$) (right-hand column).

approximation of the sum over wave modes by an integral is not very good.

4.4 Mass function

Fig. 5 shows the number density of haloes for the three series of simulations as a function of mass of haloes. The haloes have been identified using the Friends of Friends method with a linking length

of 0.1 in units of the grid length. Plotted in the same panels are the expected values computed using the Press–Schechter mass function with a correction for the finite box size.

For each series of simulations, and at each epoch, we fitted the value of δ_c to match the simulation with the natural cut-off at the box scale. The same value of δ_c is then used for other simulations of the series.

We find that the features of the mass function are reproduced correctly by the analytical approximation, namely:

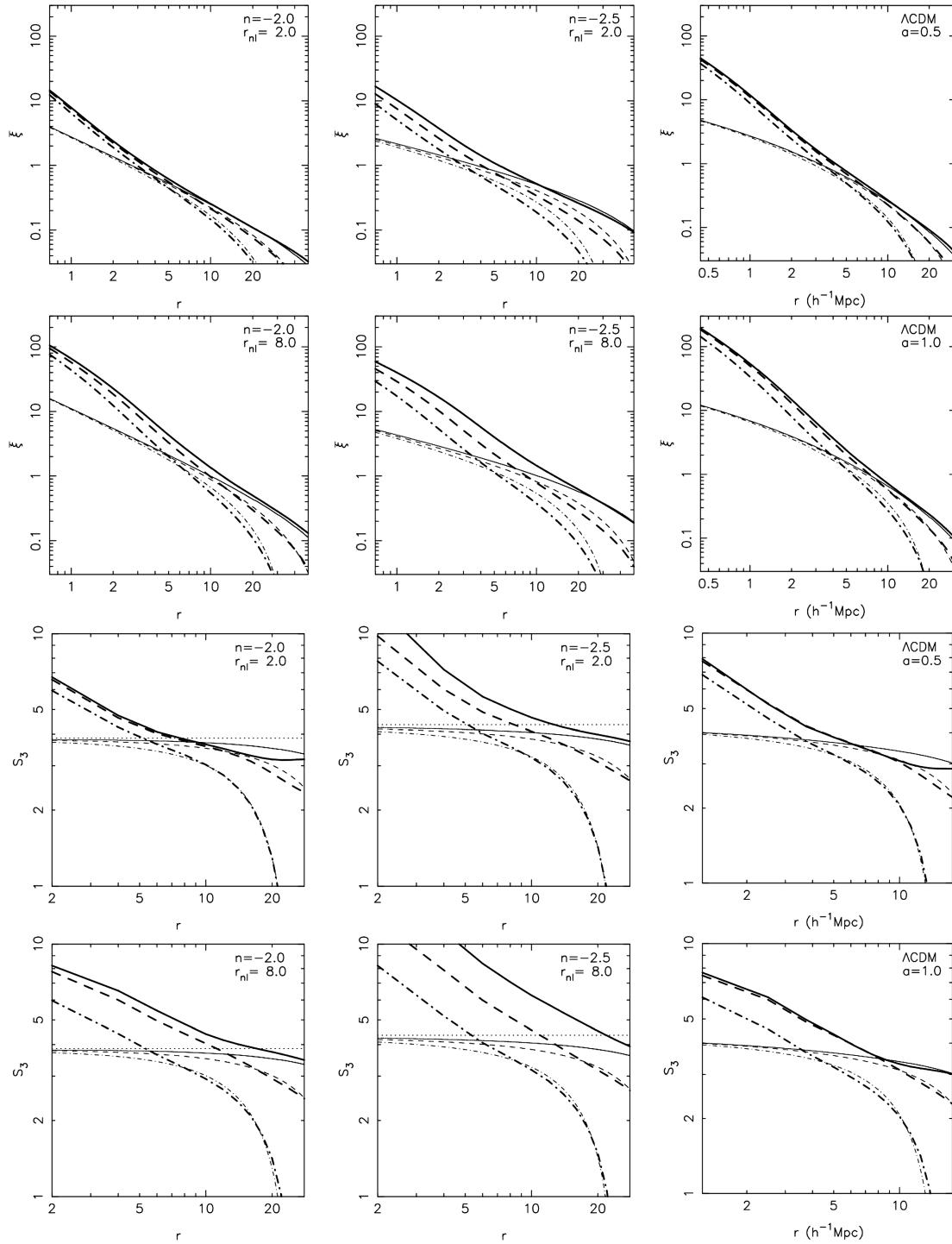


Figure 4. The first two rows in this figure show the average two-point correlation function $\bar{\xi}$, and the next two rows show skewness S_3 . The first and third rows represent the early epoch, and the second and fourth rows represent the later epoch. In all the panels, models with $k_c = k_f$, $k_c = 2k_f$ and $k_c = 4k_f$ are represented by the solid, dashed and dot-dashed lines, respectively. In all the panels, corresponding to every model in simulation (thick lines) theoretical estimates (thin lines) are also shown. Horizontal dashed lines in the lower rows show the expected value of S_3 in the absence of any box-size correction for the power-law models.

- (i) the number density of the most massive haloes declines rapidly as the *effective* box size is reduced and
- (ii) the number density of low-mass haloes increases as the *effective* box size is reduced. This feature is apparent only at the late epoch.

4.5 Velocities

In our discussion of analytical estimates of the effects of a finite box size on observable quantities, we have so far omitted any discussion of velocity statistics. The main reason for this is that the power

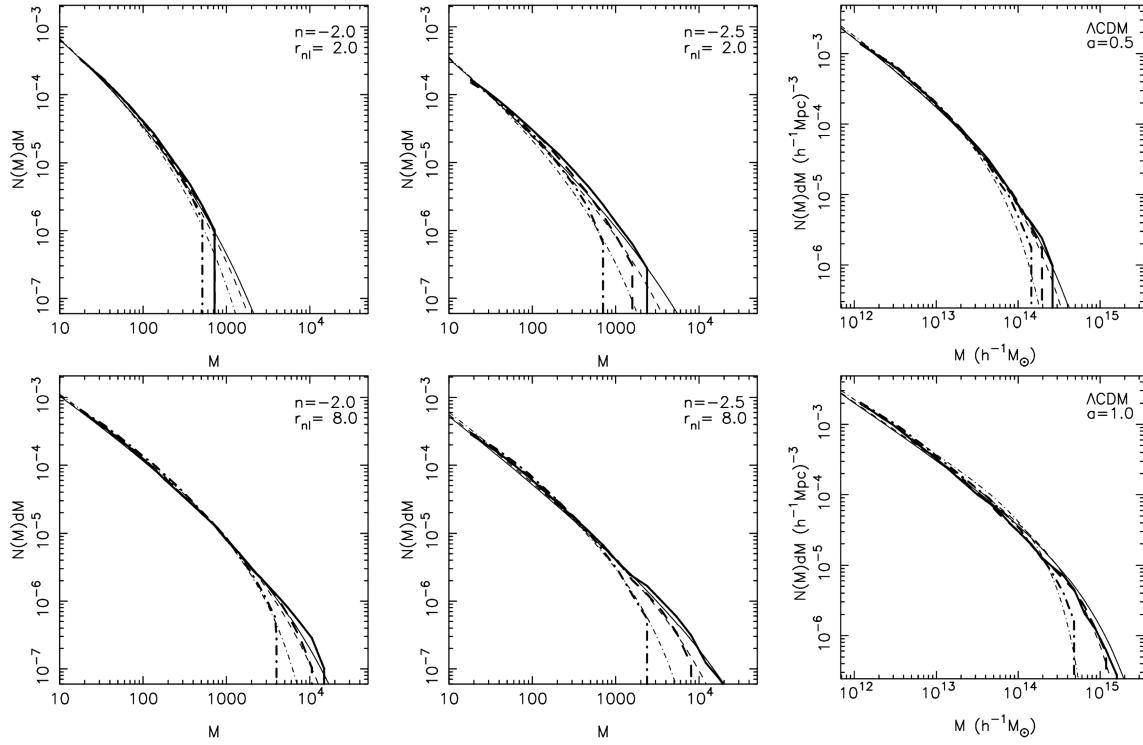


Figure 5. This figure shows the mass function $N(M)$ dM for the three series of simulations. The top row shows this for the early epoch and the lower row corresponds to the late epoch. In all the panels, models with $k_c = k_f$, $k_c = 2k_f$ and $k_c = 4k_f$ are represented by the solid, dashed and dot-dashed lines, respectively. In all the panels, corresponding to every model in simulation (thick lines) theoretical estimates (thin lines) are also shown.

spectrum for velocity is different as compared to the power spectrum for density and one can get divergences for quantities analogous to the second-order estimators analogous to σ^2 for models with $-3 < n \leq -1$. This is due to a more significant contribution of long wave modes to the velocity field than is the case for density. Relative velocity statistics are more relevant on physical grounds, and we use these for an empirical study of the effects of a finite box size on velocities. It is also important to check whether considerations related to velocity statistics put a stronger constraint on the box size required for simulations of a given model.

We measure the radial pair velocity and also the pair velocity dispersion in the simulations used in this work. These quantities are defined as follows:

$$h(r) = -\frac{\langle (\mathbf{v}_j - \mathbf{v}_i) \cdot (\mathbf{r}_j - \mathbf{r}_i) \rangle}{a H r_{ij}^2}, \quad (6)$$

where the averaging is done over all pairs of particles with separation $r_{ij} = |(\mathbf{r}_j - \mathbf{r}_i)| = r$. In practice this is done in a narrow bin in r . Here, a is the scalefactor, H is the Hubble parameter and \mathbf{v}_i is the velocity of the i th particle. Similarly, the relative pair velocity dispersion is defined as

$$\sigma_v^2(r) = \frac{\langle |\mathbf{v}_{ij}|^2 \rangle}{a^2 H^2 r_{ij}^2}, \quad (7)$$

where \mathbf{v}_{ij} is the relative velocity for a pair of particles, and averaging is done over pairs with separation r . Dividing by $a^2 H^2 r_{ij}^2$ gives us a dimensionless quantity and the usefulness of this is apparent from the following discussion.

We have plotted the radial component of pair velocity as a function of scale r in the top two rows of Fig. 6. Panels in these rows show the pair velocity for the different models at early and late epochs. In each panel, we find that the dependence of pair velocity

on r is very sensitive to the small k cut-off used in generating the initial conditions for the simulation. It has been known for some time (Hamilton et al. 1991; Nityananda & Padmanabhan 1994) that h is an almost universal function of ξ . This is certainly true in the linear regime where $h = 2\xi/3$ for clustering in an Einstein-de Sitter universe (Peebles 1980). In order to exploit this aspect, and also to check whether the relation between h and ξ in the weakly non-linear regime is sensitive to the box size, we plot h as a function of ξ at the same scale in the last two rows of Fig. 6. We find that all runs of a series fall along the same line and variations induced by the finite box size are small even in the non-linear regime.

Fig. 7 shows the relative velocity dispersion as a function of scale (top two rows) and also as a function of ξ (lowest two rows). Again, we find that although the relative pair velocity dispersion at a given scale is sensitive to the size of the simulation box, its dependence on ξ is not affected by the large-scale cut-off. Thus we can use estimates of the correction in ξ to get an estimate of corrections in pair velocity statistics.

5 SUMMARY

The conclusions of this paper may be summarized as follows.

(i) We have extended our formalism for estimating the effects of a finite box size beyond the second moment of the density field. We have given explicit expressions for estimating the skewness and kurtosis in the weakly non-linear regime when a model is simulated in a finite box size.

(ii) We have tested the predictions of our formalism by comparing these with the values of physical quantities in N -body simulations where the large-scale modes are set to zero without changing the small-scale modes.

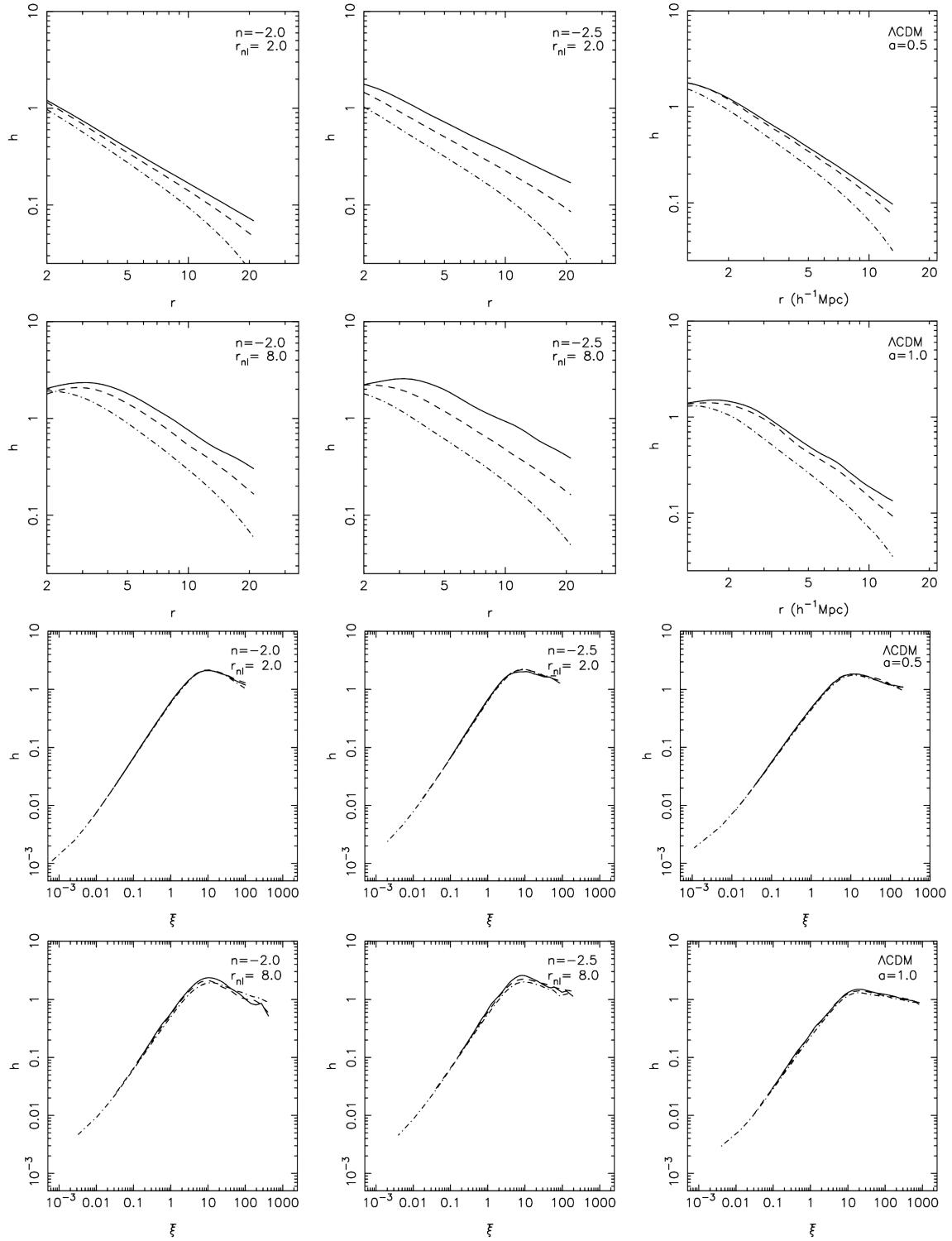


Figure 6. The first two rows show the pair velocity as a function of distance, and the lowest two rows show the pair velocity as a function of average two-point correlation function. The first and third rows represent the early epoch, and second and fourth rows represent the later epoch. In all panels, models with $k_c = k_f$ (A1, B1 and C1), $k_c = 2k_f$ (A2, B2 and C2) and $k_c = 4k_f$ (A3, B3 and C3) are represented by the solid, dashed and dot-dashed lines, respectively.

(iii) We find that the formalism makes accurate predictions for the finite box-size effects on the averaged two-point correlation function ξ and skewness.

(iv) We find that the formalism correctly predicts all the features of the mass function of a model simulated in a finite box size.

(v) We studied the effects of a finite box size on relative velocities. We find that the effects on relative velocities mirror the effects on ξ .

It is desirable that in N -body simulations the intended model is reproduced at all scales between the resolution of the simulation

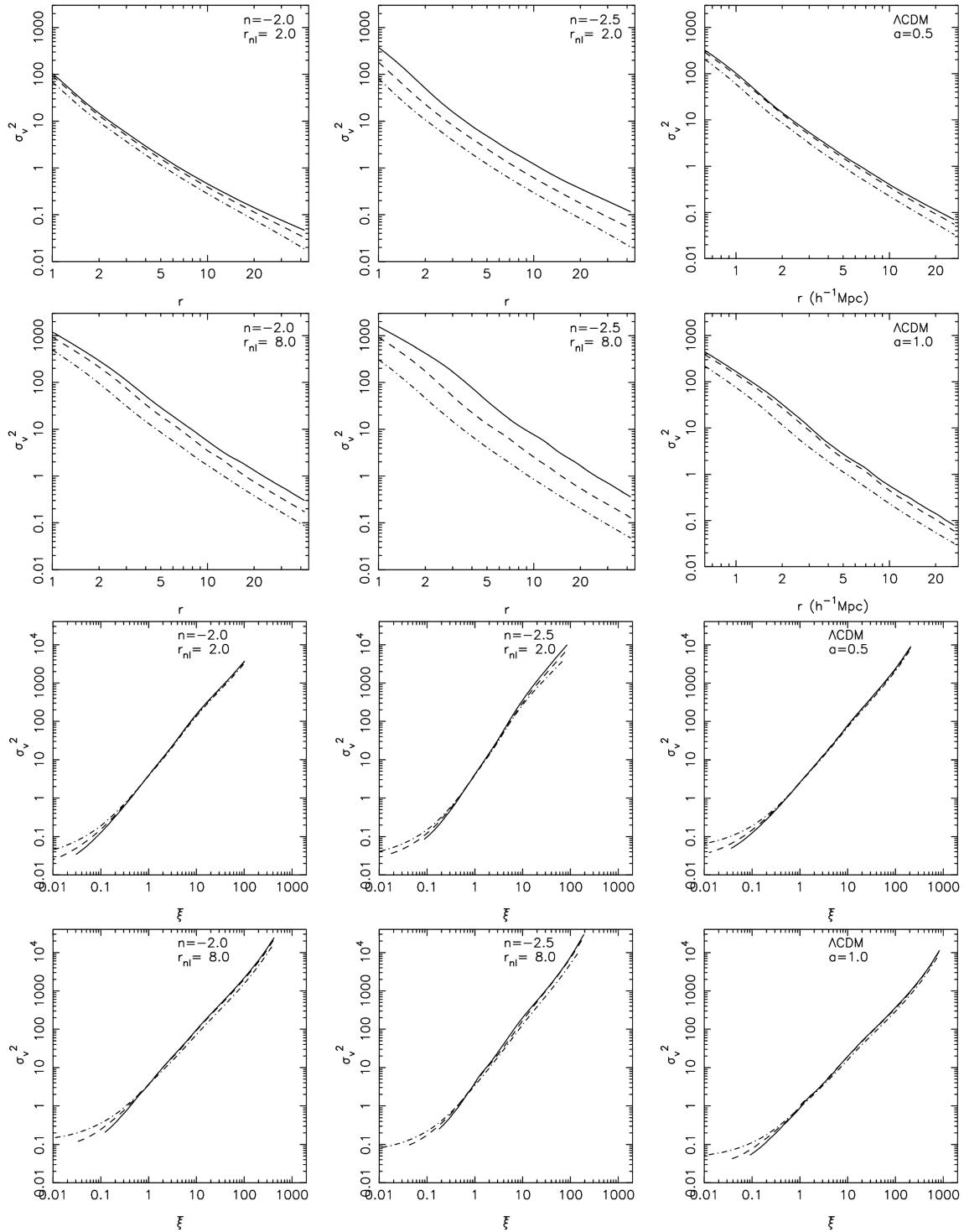


Figure 7. The first two rows show the pair velocity dispersion as a function of distance, and the lowest two rows show the pair velocity dispersion as a function of average two-point correlation function. The first and third rows represent the early epoch, and second and fourth rows represent the later epoch. In all panels, models with $k_c = k_f$ (A1, B1 and C1), $k_c = 2k_f$ (A2, B2 and C2) and $k_c = 4k_f$ (A3, B3 and C3) are represented by the solid, dashed and dot-dashed lines, respectively.

and a fairly large fraction of the simulation box. The outer scale up to which the model can be reproduced fixes the effective dynamical range of simulations. One would like S_3 to be within a stated tolerance of the expected value at this scale. We plot S_{31}/S_{30} for power-law models at the scale $L_{\text{box}}/20$, $L_{\text{box}}/10$ and $L_{\text{box}}/5$ in the left-hand panel of Fig. 8. These are plotted as a function of

$n + 3$. It can be shown that this ratio, as also σ_1/σ_0 are functions of scale only through the ratio r/L_{box} . We find that S_{31}/S_{30} is large for large negative n and decreases monotonically as n increases. This ratio is smaller than 10 per cent only for $n \geq 0.8$ at $r = L_{\text{box}}/5$. The corresponding number for $r = L_{\text{box}}/10$ is $n \geq -1.6$, and for $L_{\text{box}}/20$ is $n \geq -2.8$. Clearly, the effective dynamic range decreases

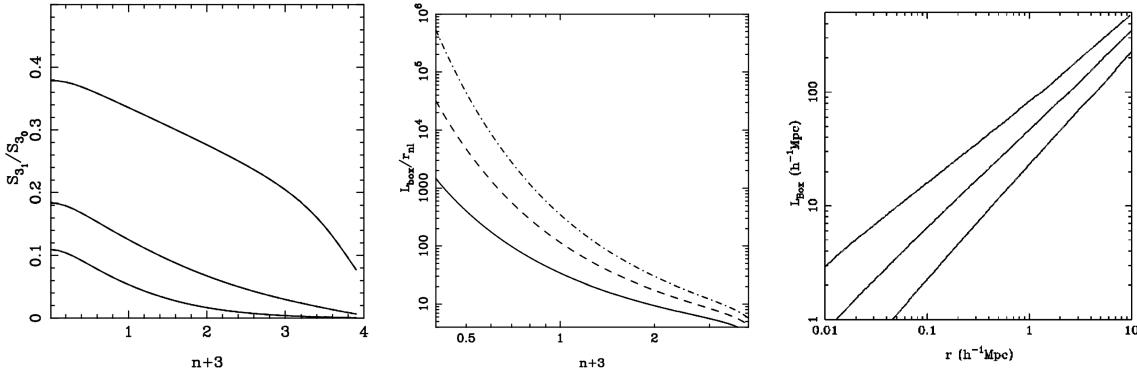


Figure 8. The left-hand panel shows the variation in the fractional correction in S_3 i.e. S_{31}/S_{30} (see equation 4), with the index of power spectrum at the scales $L_{\text{box}}/5$ (top curve), $L_{\text{box}}/10$ (middle curve) and $L_{\text{box}}/20$ (lowest line). For a given tolerance of the error in S_3 due to finite box effects, this gives us the largest scale at which the simulation may be expected to give reliable results. The middle panel shows contours of C_1 at the scale of non-linearity ($\sigma_0 = 1$) for values $C_1 = 0.01$ (top curve), 0.03 (middle curve) and 0.1 (lower curve). The contours are plotted on the $L_{\text{box}}/r_{\text{nl}} - (n + 3)$ plane and indicates the box size required for reliable simulations of a given model. The right-hand panel shows contours of S_{31}/S_{30} for the ΛCDM model that best fits the *WMAP-5* data. Contours shown are for $S_{31}/S_{30} = 0.01, 0.03$ and 0.1 .

rapidly as $n + 3 \rightarrow 0$. This highlights the difficulties associated with simulating such models.

Similarly, one would like σ^2 and σ_0^2 to be comparable at the scale of non-linearity. From requirements of self-similar evolution of power-law models in simulations, we find that at the scale of non-linearity $C_1 \leq 0.03$ is required for the effects of a finite box size to be ignorable. This gives us a lower bound on $L_{\text{box}}/r_{\text{nl}}$ for any given model. The middle panel shows the required $L_{\text{box}}/r_{\text{nl}}$ as a function of $n + 3$ for $C_1 = 0.01, 0.03$ and 0.1 . Here, C_1 is the asymptotic value of σ_1^2 at $r \ll L_{\text{box}}$ and is a fairly good approximation at small scales. We find that the required $L_{\text{box}}/r_{\text{nl}}$ for $n = -2$ is more than 100 for $C_1^2 = 0.03$ at the scale of non-linearity. Thus we need a simulation with $L_{\text{box}} \geq 10^3$ if we are to probe the strongly non-linear regime ($\xi \gg 100$) with some degree of confidence. Requirements for models with $n < -2$ are much more stringent, and for models like $n = -2.5$ even the largest simulations cannot be used to study the asymptotic regime.

To put things in context for the favoured cosmological model, the right-hand panel in Fig. 8 shows contours of S_{31}/S_{30} in the $r - L_{\text{box}}$ plane for the ΛCDM model that best fits the *WMAP-5* data (Dunkley et al. 2009). We find that in order to ensure that the error in skewness is less than 10 per cent at a scale of $10 h^{-1}\text{Mpc}$, we need a simulation box of more than $200 h^{-1}\text{Mpc}$. The required box size is much bigger if the tolerance on error in skewness is smaller. This is a very stringent requirement for simulations of the epoch of reionization where one would like to get the clustering right at the scales of a few Mpc.

Given that the formalism we have proposed works well when compared with simulations, and the fact that calculations in this formalism are fairly straightforward, we would like to urge the cosmological *N*-body simulations community to make use of this formalism. We would like to request simulators to report the fractional corrections to the linearly extrapolated amplitude of clustering and the fractional correction to skewness across the range of scales of interest. This will enable users of simulations to assess potential errors arising due to a finite simulation volume.

ACKNOWLEDGMENTS

Numerical experiments for this study were carried out at cluster computing facility in the Harish-Chandra Research Institute (<http://cluster.mri.ernet.in>). This research has made use of NASA's

Astrophysics Data System. We would like to thank T. Padmanabhan for useful discussions. We thank the anonymous referee for useful comments and suggestions.

REFERENCES

- Bagla J. S., 2002, *JAA&A*, 23, 185
- Bagla J. S., 2005, *Current Science*, 88, 1088
- Bagla J. S., Padmanabhan T., 1994, *MNRAS*, 266, 227
- Bagla J. S., Padmanabhan T., 1997a, *MNRAS*, 286, 1023
- Bagla J. S., Padmanabhan T., 1997b, *Pramana J. Phys.*, 49, 161
- Bagla J. S., Prasad J., 2006, *MNRAS*, 370, 993
- Bagla J. S., Prasad J., 2009, *MNRAS*, 393, 607
- Bagla J. S., Ray S., 2003, *New Astron.*, 8, 665
- Bagla J. S., Ray S., 2005, *MNRAS*, 358, 1076
- Bagla J. S., Prasad J., Ray S., 2005, *MNRAS*, 360, 194
- Barkana R., Loeb A., 2004, *ApJ*, 609, 474
- Bernardeau F., 1994, *ApJ*, 433, 1
- Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, *Phys. Rep.*, 367, 1
- Bertschinger E., 1998, *ARA&A*, 36, 599
- Brainerd T. G., Scherrer R. J., Villumsen J. V., 1993, *ApJ*, 418, 570
- Colombi S., Bouchet F. R., Schaeffer R., 1994, *A&A*, 281, 301
- Couchman H. M. P., Peebles P. J. E., 1998, *ApJ*, 497, 499
- Davis M., Peebles P. J. E., 1977, *ApJS*, 34, 425
- Dolag K., Borgani S., Schindler S., Diaferio A., Bykov A. M., 2008, *Space Science Reviews*, 134, 229
- Dunkley J. et al., 2009, *ApJS*, 180, 306
- Gelb J. M., Bertschinger E., 1994a, *ApJ*, 436, 467
- Gelb J. M., Bertschinger E., 1994b, *ApJ*, 436, 491
- Gurbatov S. N., Saichev A. I., Shandarin S. F., 1989, *MNRAS*, 236, 385
- Hamilton A. J. S., Kumar P., Lu E., Matthews A., 1991, *ApJ*, 374, L1
- Hui L., Bertschinger E., 1996, *ApJ*, 471, 1
- Jain B., Mo H. J., White S. D. M., 1995, *MNRAS*, 276, L25
- Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, *MNRAS*, 321, 372
- Kanekar N., 2000, *ApJ*, 531, 17
- Kauffmann G., Melott A. L., 1992, *ApJ*, 393, 415
- Khandai N., Bagla J. S., 2008, preprint (arXiv:0802.3215)
- Komatsu E. et al., 2009, *ApJS*, 180, 330
- Little B., Weinberg D. H., Park C., 1991, *MNRAS*, 253, 295
- Ma C., 1998, *ApJL*, 508, L5
- Matarrese S., Lucchin F., Moscardini L., Saez D., 1992, *MNRAS*, 259, 437
- Nityananda R., Padmanabhan T., 1994, *MNRAS*, 271, 976

- Padmanabhan T., 1993, Structure Formation in the Universe. Cambridge Univ. Press, Cambridge
- Padmanabhan T., 1996, MNRAS, 278, L29
- Padmanabhan T., 2002, Theoretical Astrophysics, Vol. III: Galaxies and Cosmology. Cambridge Univ. Press, Cambridge
- Padmanabhan T., Cen R., Ostriker J. P., Summers F. J., 1996, ApJ, 466, 604
- Peacock J. A., 1999, Cosmological Physics. Cambridge Univ. Press, Cambridge
- Peacock J. A., Dodds S. J., 1994, MNRAS, 267, 1020
- Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19
- Peebles P. J. E., 1974, A&A, 32, 391
- Peebles P. J. E., 1980, The Large-scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
- Peebles P. J. E., 1985, ApJ, 297, 350
- Percival W. J., 2007, ApJ, 657, 645
- Power C., Knebe A., 2006, MNRAS, 370, 691
- Prasad J., 2007, JA&A, 28, 117
- Press W. H., Schechter P., 1974, ApJ, 187, 425
- Reed D. S., Bower R., Frenk C. S., Jenkins A., Theuns T., 2008, preprint (arXiv:0804.0004)
- Sahni V., Coles P., 1995, Phys. Rep., 262, 1
- Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
- Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
- Sirko E., 2005, ApJ, 634, 728
- Smith R. E. et al., 2003, MNRAS, 341, 1311
- Takahashi R. et al., 2008, MNRAS, 389, 1675
- Zel'dovich Y. B., 1970, A&A, 5, 84

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