THE ASTROPHYSICAL JOURNAL, 513:L99–L102, 1999 March 10

HOW DARK MATTER HALOS CLUSTER IN LAGRANGIAN SPACE

CRISTIANO PORCIANI,[1,2] PAOLO CATELAN,[3] AND CEDRIC LACEY[3]
Received 1998 December 1; accepted 1999 January 15; published 1999 February 3

ABSTRACT

We investigate the clustering of dark matter halos in Lagrangian space in terms of their two-point correlation function. Analyzing a set of collisionless scale-free 128^3 particle *N*-body simulations with spectral indices $n = -2, -1$, we measure the first two Lagrangian bias parameters b_1 and b_2 relating halo and mass correlations. We find that the Mo & White leading-order formula for b_1 describes the clustering of halos with mass $M \gtrsim M_\star$ (where $M_\star$ indicates the characteristic nonlinear mass) quite accurately. Smaller halos turn out to be less clustered in Lagrangian space than predicted by Mo & White. Our findings are consistent with the recent results of Jing for the clustering of halo populations in Eulerian space, demonstrating that the discrepancies between the *N*-body and analytical Mo & White prediction for the bias exist already in Lagrangian space. This shows that a more refined theoretical algorithm for selecting halos in the initial conditions needs to be developed. Finally, we present a very accurate fitting formula for the linear halo bias factor b_1 in Lagrangian space.

Subject headings: cosmology: theory — galaxies: statistics — large-scale structure of universe

1. INTRODUCTION

Virialized dark matter halos in the universe are not distributed in the same way as the underlying dark matter. This is true whether one takes the final positions of the halos in Eulerian space or their initial positions in Lagrangian space (Mo & White 1996, hereafter MW; Catelan et al. 1998, hereafter CLMP; Catelan, Matarrese, & Porciani 1998; Jing 1998; Sheth & Lemson 1998). The numerical tour de force by Jing (1998), who thoroughly investigated with unmatched accuracy the clustering of dark matter halos in Eulerian space, demonstrated explicitly that (1) the halo-to-mass bias is independent of the halo separation (at least in the scale-free case $n = -2$ and in the linear regime) and (2) the MW Eulerian linear bias correctly describes the clustering of halos of masses $M \gtrsim M_\star$, but systematically underpredicts it for any value of the spectral index n if $M \lesssim M_\star$, where $M_\star$ is the typical nonlinear mass. However, it is impossible to understand solely on the basis of the Eulerian investigation whether the discrepancies between the numerical results and the analytical MW predictions are due to (1) a failure of the algorithm for identifying the halo positions in Lagrangian space, (2) the effects of nonlinear shear dynamics (not accounted for in the original MW approach) on the mapping of halo positions from Lagrangian to Eulerian space, or (3) a combination of the two.

In this Letter, we employ scale-free collisionless *N*-body simulations [i.e., with density parameter $\Omega = 1$ and initial power spectra $P(k) \propto k^n$] to investigate the clustering of dark matter halos in Lagrangian space in terms of their two-point correlation function. We compare the halo correlation function to the correlation of the underlying dark matter for $n = -2$, -1 spanning more than 4 (3) orders of magnitudes in halo masses for $n = -2$ ($n = -1$).

We find that the theoretical "underclustering" reported by Jing for masses $M \lesssim M_\star$ is already present in the initial conditions as well (but as "overclustering," since the first-order bias is negative for small masses), and it cannot be due exclusively to the subsequent nonlinear effects of the shear dynamics acting on small scales and unaccounted for in the original spherical collapse model of MW. We then argue that the standard Press-Schechter (1974) approach ("extended" or not), on which MW and CLMP based their speculations, is inadequate for identifying the locations of halos in the initial conditions, and a major effort should be devoted to finding an improved algorithm. Section 2 reports the details of the present investigation. Finally, an accurate fitting formula for the bias b_1 in Lagrangian space is given, which should be considered as the Lagrangian version of Jing's fitting formula for the Eulerian case. Section 3 contains our conclusions.

2. HALO CLUSTERING FROM *N*-BODY SIMULATIONS

2.1. *Simulations and Halos*

The simulations used here are similar to those of Lacey & Cole (1994), who used them to test halo merging histories. They were performed using the AP^{3}M code of Couchman (1991) with 128^3 particles. The force softening used was $L/1280$, with L the size of the periodic box. Initial positions and velocities were generated by displacing particles from a uniform 128^3 grid according to the Zeldovich approximation, assuming an initial scale-free power spectrum and Gaussian statistics. We considered four realizations for two different spectral indices $n = -2, -1$. For each simulation, we recorded, for many epochs, positions and velocities of all particles. The output times (38 for $n = -2$ and 35 for $n = -1$) were chosen so that $M_\star$ increased by a factor $\sqrt{2}$ between subsequent output times. In the last output, $M_\star$ corresponded to 33,748 particles for $n = -2$ and to 28,616 particles for $n = -1$.

For each output time, we selected dark matter halos in the simulations employing the "friends-of-friends" group finder with a linking length equal to 20% of the mean interparticle distance (e.g., Davis et al. 1985). We checked that results obtained using the spherical overdensity group finder (Lacey & Cole 1994) are essentially identical. We excluded halos containing fewer than 20 particles or more than 20,000 particles. We moved all of the particles belonging to a given halo back to their initial (Lagrangian) positions, then computed the position of their center of mass. We used the latter as the "halo position" in Lagrangian space. In such a way, for each output

[1] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218.
[2] Scuola Internazionale Superiore di Studi Avanzati, via Beirut 2-4, 34014 Trieste, Italy.
[3] Theoretical Astrophysics Center, Juliane Maries Vej 30, 2100 Copenhagen Ø, Denmark.

time, we constructed a catalog of halos indicating their mass and position in Lagrangian space.

2.2. *Halo Correlation Function in Lagrangian Space*

We computed the mean correlation function $\bar{\xi}_h$ between halos in a given mass interval, where a bar denotes mass averaged quantities. Self-similar scaling allowed us to combine data from different output times in order to reduce the Poisson fluctuations due to the finite number of halos within the box. We considered every output time containing more than 100 halos in the same mass interval, and we binned the distributions of halo separations in units of $r/R_\star$, where $R_\star^3 \propto M_\star$. Finally, we computed the halo correlation function using the estimator

$$\langle \bar{\xi}_h \rangle = \frac{\sum_i N_i[r/R_\star(z_i)]}{\sum_i N_i^{\rm Poi}[r/R_\star(z_i)]} - 1, \qquad (1)$$

where the index i runs over different output times z_i of the same simulation, $N_i(r/R_\star)$ is the number of halo pairs in the ith output, and $N_i^{\rm Poi}(r/R_\star)$ is the corresponding quantity for a Poisson process with the same number density. The average symbol (angle brackets) is introduced since we considered information coming from the different temporal outputs. In this way, for each mass interval, we collected four realizations of the Lagrangian correlation function of dark matter halos.

This procedure allows us to achieve two goals: (1) to reduce statistical fluctuations by increasing the number of halo pairs and (2) to extend the mass interval and the range of halo separations that may be sampled. In fact, the box size L and the minimum halo mass in the simulation are fixed, while $R_\star$ and $M_\star$ increase with time. This means that the correlation function for halos with $M \gg M_\star$ is measured mainly from the early output times, while that for halos less massive than $M_\star$ comes mostly from later output times. In order to simulate bootstrap resampling, we assigned as the standard error the Poisson error bar multiplied by a factor of $\sqrt{3}$ (Mo, Jing, & Börner 1992).

2.3. *Lagrangian Bias Parameters*

Now we want to test whether the Lagrangian halo correlation ξ_h is related that of the mass (ξ_m) through a relation (see eq. [15] in CLMP and eq. [38] in Porciani et al. 1998)

$$\xi_h\left(\frac{r}{R_\star}\right) = b_1^2 \xi_m\left(\frac{r}{R_\star}\right) + \frac{b_2^2}{2}\,\xi_m^2\left(\frac{r}{R_\star}\right) + \dots\ , \qquad (2)$$

where the symbol b_i denotes the ith Lagrangian bias factor, and ξ_m is calculated according to linear theory. For instance, the extended Press-Schechter approach (Bond et al. 1991) leads to the MW and CLMP expressions for b_1 and b_2, namely $b_1 = \delta_c/\sigma_M^2 - 1/\delta_c$ and $b_2 = (\delta_c^2/\sigma_M^2 - 3)/\sigma_M^2$, where σ_M^2 indicates the mass variance on scale M (Cole & Kaiser 1989; MW; Mo, Jing, & White 1997). Note that, by definition, $\sigma_M = \delta_c$ for $M = M_\star$. To test this biasing model against simulations, we have to consider a finite range of halo masses. Equation (2) then implies $\bar{\xi}_h = \bar{b}_1^2 \xi_m + \frac{1}{2}\bar{b}_2^2 \xi_m^2 + \dots$, where $\bar{b}_k$ is the mean of b_k in the mass interval, weighted by the mass function. In order to obtain statistically reliable values for b_1 and b_2, we compute the averages $\langle \xi_m \rangle$ and $\langle \xi_m^2 \rangle$ *directly* from the initial conditions of the simulation, following the same averaging procedure we used for the halo correlations (in this case, it corresponds to averaging over output times using the volume of the $r/R_\star$ bin as a weight). Note that, after averaging, $\langle \xi_m^2 \rangle$ generally differs from $\langle \xi_m \rangle^2$, so that equation (2) implies $\langle \bar{\xi}_h \rangle = \bar{b}_1^2 \langle \xi_m \rangle + \frac{1}{2}\bar{b}_2^2 \langle \xi_m^2 \rangle + \dots$. This averaging procedure is important to account for the finiteness of the box: in this way, the lack of any Fourier components of the density would be equally experienced by both mass and halo distributions. Finally, to check the reliability of equation (2), we apply the χ^2 test, assuming that every $\langle \bar{\xi}_h \rangle$ is normally distributed around a mean value.

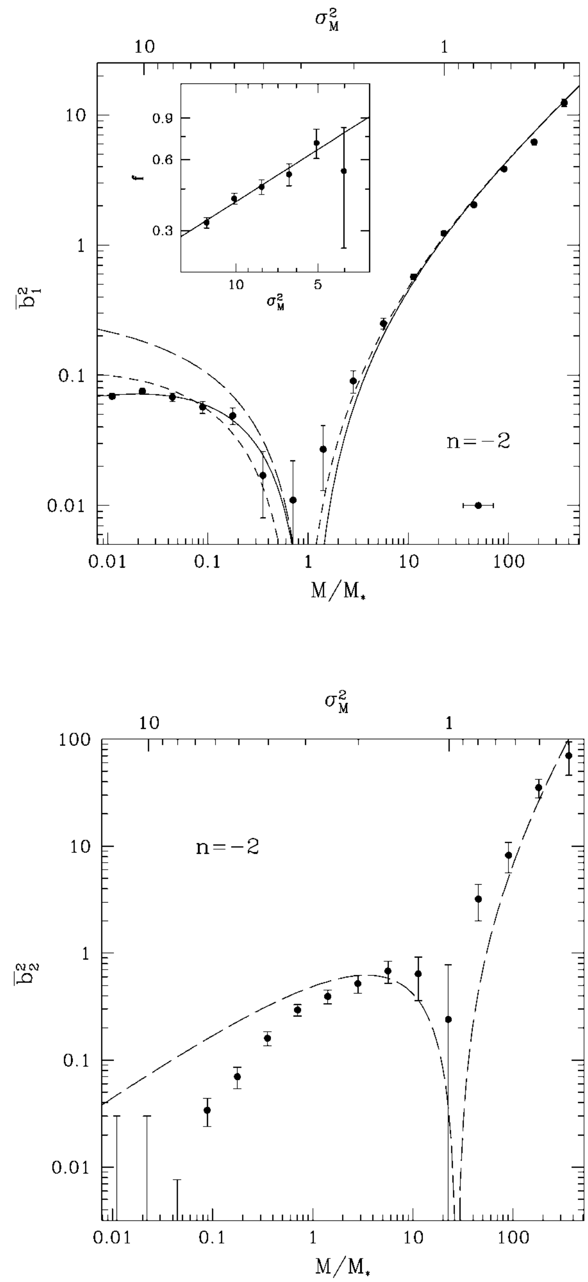

FIG. 1.—*Dots:* Best-fit values of $\bar{b}_1^2$ (*top*) and $\bar{b}_2^2$ (*bottom*) as a function of variance σ_M^2 and halo mass on the same scale for $n = -2$. Error bars denote 1 σ uncertainties. *Long-dashed line:* MW and CLMP Lagrangian bias with $\delta_c = 1.686$. *Short-dashed line:* Jing's fit to the Eulerian bias minus 1. *Continuous line:* Our fitting formula for b_1 given in the main text. Both the theoretical predictions and the fitting formula are averaged over the Press-Schechter mass function. The horizontal error bar in the bottom right-hand corner of the top panel indicates the mass intervals considered. The inset in the top panel shows the best-fit correction $f(\sigma_M)$ to the MW formula for b_1, when $M < M_\star$: $\log f = (0.34 \pm 0.11) - (0.73 \pm 0.12) \log \sigma_M^2$.

Halos are objects of finite size, and we expect exclusion

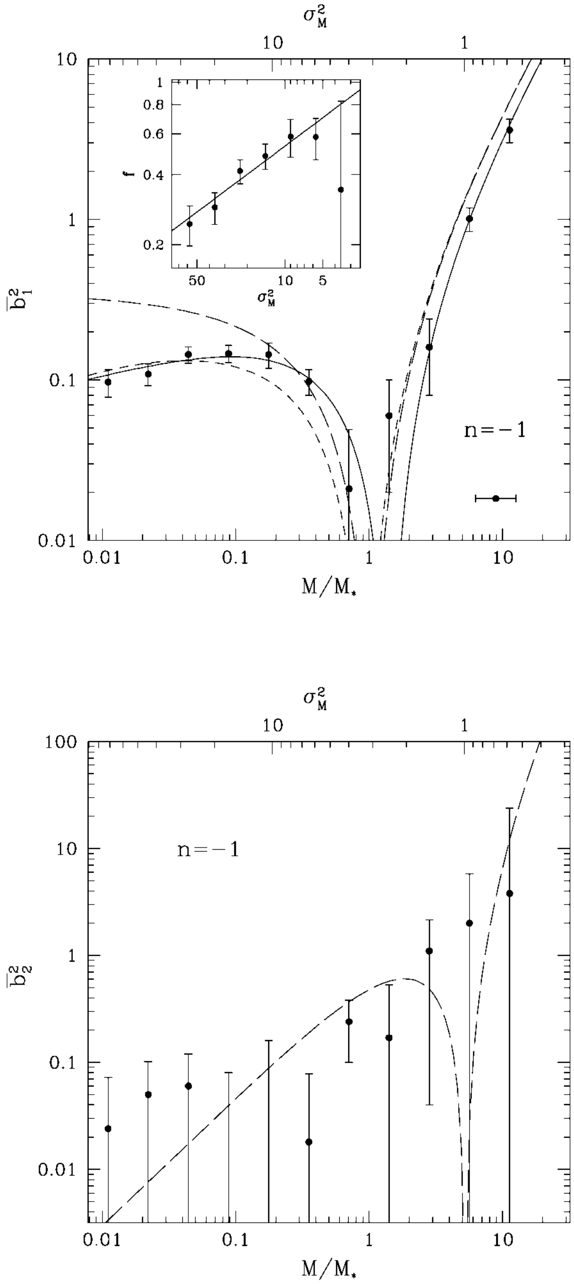

FIG. 2.—As in Fig. 1, but for $n = -1$. In this case, the best fit for f is $\log f = (0.13 \pm 0.13) - (0.41 \pm 0.10) \log \sigma_M^2$.

effects to dominate ξ_h at separations of the order of the halo Lagrangian size R_L. In a sample composed of identical spherical halos, spatial exclusiveness implies $\xi_h = -1$ for $r < 2R_L$ and a strong compensating positive correlation at $r \gtrsim 2R_L$. The magnitude of this positive correlation is related to the volume fraction occupied by the spheres, since $\int_0^\infty dr\, r^2\, \xi_h(r) = 0$. Inspired by this, we decided to exclude from the χ^2 test all data corresponding to separations $r \lesssim 2R_{max}$, where $R_{max} = R_\star(M_{max}/M_\star)^{1/3}$ is the characteristic Lagrangian radius of the largest halo in the mass interval considered. Note that, out of the exclusion region, the additional condition $|\xi_m| < 1$ is not needed to neglect higher order terms in equation (2). This is because, for $i > 1$ and $\sigma_M \to \infty$, $b_i \to 0$ as $\sigma_M^{-\alpha}$ with $\alpha \geq 2$. Anyway, the exact value of the minimum separation considered has been determined on a case-by-case basis, checking for the stability of our results with increasing the number of data points at small separations. We avoided averaging between different realizations of the same power spectrum: the entire data set of up to four values of $\langle \xi_h \rangle$ for each $r/R_\star$ interval is considered in the χ^2 test.

Minimization of χ^2 with nonholonomic constraints $\bar{b}_1^2 \geq 0$ and $\bar{b}_2^2 \geq 0$ is used to obtain estimates of the bias parameters. The resulting χ^2_{min} shows that the a priori bias hypothesis—the bias relation in equation (2)—is acceptable to a 95% confidence level over a wide interval of halo masses. However, for $n = -2$, samples with $M > 16M_*$ are inconsistent with equation (2) at the 99% confidence level. This is caused by the presence of noticeable differences (especially at very large separations) between the halo correlation functions extracted from different realizations. Indeed, very massive halos are exponentially rare and more affected by statistical fluctuations. On the other hand, we could also have underestimated the uncertainty in computing $\langle \xi_h \rangle$.

Figures 1 and 2 show the best-fit results for $\bar{b}_1^2$ and $\bar{b}_2^2$. The error bars are obtained by projecting along the b-axes the contours of constant χ^2 at $\chi^2_{min} + 1$. Concerning $\bar{b}_1^2$, the comparison against theoretical predictions shows a quantitatively good agreement for $M > M_\star$. However, the square bias parameter of lower mass halos is significantly smaller than predicted by Mo & White (meaning that the bias b_1 is *less negative* than the MW value). Findings of this kind have been published by Jing (1998), who investigated the Eulerian halo clustering in the quasi-linear regime, giving an accurate fitting formula for the Eulerian analog of b_1. Our result shows incontrovertably that the effects discovered by Jing are already present in Lagrangian space.

There is no rigorous way to apply Jing's bias-fitting formula in Lagrangian space, since the nonlocal dynamics of the mass density field enters the transformation. However, as a first approach, and only to check the order of magnitude of the effect, we can follow MW in assuming a spherical evolution of the coarse-grained mass density field to map from Lagrangian to Eulerian space. In this case, the first Eulerian bias term is given by $1 + b_1$. Jing's formula, $b_1 \equiv b_{Jing} - 1$, turns out to be a reasonably good description of our data, showing that the underclustering of small halos in Eulerian space detected by Jing is correspondingly already present in their Lagrangian clustering; the intervening dynamics is approximately well accounted for by the standard mapping from Lagrangian to Eulerian space in the laminar regime. We elaborate below our own fitting formula for the linear Lagrangian bias, which turns out to be accurate to 10% over the entire mass range investigated (with the exception of the zero crossing region for b_1).

Our results for b_2 are of course less conclusive than for b_1, mainly because of the larger uncertainties. However, they confirm the information extracted considering b_1. Quantitatively, for $n = -2$, the CLMP expression for b_2 gives a good estimate for halos with $M > M_*$, while it overpredicts the numerical outcome for $\bar{b}_2^2$ when $M < M_*$. For $M \sim M_*$, our results are a factor of ~1.4 smaller than predicted by the CLMP formula. This is the range in which our method is best suited to compute b_2, since b_1 vanishes and the halo correlation function is proportional to ξ_m^2. No firm conclusion can be drawn for $n = -1$, since practically the entire data set only allows one to set an upper limit on $\bar{b}_2^2$. It is not unlikely that, for the less negative spectral indices, bigger simulations are better suited to quantify the halo clustering up to second-order biasing, above all in Lagrangian space. It will be worthwhile to address this point in a future work.

2.4. *Accurate Fitting Formula for* b_1

Our results for b_1 can be accurately parametrized by introducing a mass-dependent, multiplicative correction to the MW formula, namely

$$b_1^2(\sigma_M) = f(\sigma_M)(\delta_c/\sigma_M^2 - 1/\delta_c)^2. \quad (3)$$

For $n = -2$, the original MW formula (i.e., $f \equiv 1$) with $\delta_c = 1.686$, as suggested by the spherical collapse model, describes the data for $M > M_\star$ to within 10% accuracy. Smaller masses instead require $f = 2.19/\sigma_M^{1.46}$. For $n = -1$ and $M > M_*$, our numerical data are very well described by the MW formula for b_1 with a lower collapse threshold $\delta_c \simeq 1.52$. In this case, for $M < M_*$ we obtain $f = 1.35/\sigma_M^{0.82}$. These fitting formulae are extremely accurate in describing our data set. Their simple power-law behavior encourages further theoretical investigation. For instance, with data for more than two spectral indices n, one may attempt to fit the dependence of b_1 on n.

3. DISCUSSION AND CONCLUSIONS

Employing 128^3-body scale-free simulations, we analyzed the clustering of dark matter halos in Lagrangian space. The main results of this investigation can be summarized as follows.

1. Assuming a correlation model as in equation (2), the first two Lagrangian bias factors b_1 and b_2 are strongly mass dependent over the 4 orders of magnitude in mass investigated.
2. The clustering of halos with mass above the nonlinear mass, $M \gtrsim M_\star$, is fairly well described by the MW formula for the linear Lagrangian bias, both for $n = -1$ and $n = -2$.
3. Halos with nonlinear masses $M \lesssim M_\star$ are less clustered (have a smaller correlation amplitude) than what the leading order Lagrangian bias of MW would predict (see also Sheth & Lemson 1998).

When these results are combined with the ones recently obtained by Jing (1998) about the clustering of halos in Eulerian space, we can disentangle the question of whether the discrepancies between Jing's numerical results and the MW theoretical predictions are mainly due to the effects of the nonlinear shear dynamics, effective on smaller scales, or to a possible failure of the halo selection algorithm in the initial conditions—a question actually left unsolved by Jing. Clearly, since, as we showed, the *same* effects discovered by Jing are essentially already present in Lagrangian space, our investigation suggests that it is time to improve on the classical Press-Schechter algorithm for identifying halos in Lagrangian space: since it assumes spherically symmetric collapse, it is not surprising that it fails in correctly counting the small halo masses, where departures from spherical collapse can be cosmologically relevant. The failure of the MW formula for the Lagrangian and Eulerian bias is presumably related to the departure of the halo mass function from the Press-Schechter form at low masses (e.g., Lacey & Cole 1994).

Finally, we derived a fitting formula for the linear Lagrangian bias b_1 that is relevant for accurately predicting the clustering of dark matter halos, above all in the low-mass tail. Our fitting formula can be considered as the Lagrangian equivalent of equation (3) in Jing (1998). These results are highly relevant for predicting the clustering of low-luminosity galaxies, most of which lie in lower-mass halos (e.g., Baugh et al. 1998). Modeling of galaxy clustering found in present and future galaxy redshift surveys will provide an important application of the present results.

C. P. acknowledges the hospitality of the Theoretical Astrophysics Center during the spring of 1997, when this work started. P. C. warmly thanks Sergio Gelato for insightful discussions. We all thank Sabino Matarrese for sharing his ideas on the matter, dark and not. C. P. thanks support from NASA ATP-NAG5-4236 grant and from Italian Ministero dell'Universita è della Ricerca Scientifica e Tecnologica. P. C. and C. L. have been supported by the Danish National Research Foundation at TAC.

REFERENCES

Baugh, C. M., et al. 1998, preprint (astro-ph/9811222)
Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440
Catelan, P., Lucchin, F., Matarrese, S., & Porciani, C. 1998, MNRAS, 297, 692
Catelan, P., Matarrese, S., & Porciani, C. 1998, ApJ, 502, L1
Cole, S., & Kaiser, N. 1989, MNRAS, 237, 1127
Couchman, H. M. P. 1991, ApJ, 368, L23
Davis, M., Efsthatiou, G., Frenk, C., & White, S. D. M. 1985, ApJ, 292, 371
Jing, Y. P. 1998, ApJ, 503, L9
Lacey, C., & Cole, S. 1994, MNRAS, 271, 676
Mo, H. J., Jing, Y. P., & Börner, G. 1992, ApJ, 392, 452
Mo, H. J., Jing, Y. P., & White, S. D. M. 1997, MNRAS, 284, 189
Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347
Porciani, C., Matarrese, S., Lucchin, F., & Catelan, P. 1998, MNRAS, 298, 1097
Press, W. H., & Schechter P. 1974, ApJ, 187, 425
Sheth, R. K., & Lemson, G. 1998, preprint (astro-ph/9808138)