

On the inclusion of environmental effects in the halo model of galaxy clustering

Ravi K. Sheth, Umami Abbas and Ramin Skibba

Department of Physics and Astronomy, University of Pittsburgh, 3941 O'Hara Street,
Pittsburgh, PA 15260, USA email: rks12@pitt.edu

Abstract. Standard models of galaxy formation make specific predictions for how galaxy properties depend on environment. These predictions follow from the assumption that galaxy properties are entirely determined by the formation histories of the halos which host them, and that the distribution of dark halo masses is skewed towards large masses in dense regions. We describe two tests of this assumption which can be made with currently available data.

1. Introduction

We are now in the early years of what is being called the era of precision cosmology; what was a trickle of precise measurements of the large-scale Universe, primarily from exquisitely designed balloon- and space-based measurements of the cosmic background radiation, is growing into a torrent, as data of unprecedented accuracy from large-scale galaxy and cluster surveys floods in. The wealth of data means that, hand-in-hand with precision cosmology, the era of statistical cosmology is upon us.

The start of the year 2003 saw the release of exquisite observations of the microwave sky from the WMAP satellite (Bennett *et al.* 2003). Shortly thereafter, the 2dFGRS (Colless *et al.* 2001, 2003) and the SDSS (Abazajian *et al.* 2003) collaborations released a wealth of data of the optical sky. The auto-correlation functions of the cosmic background radiation temperature, and of the galaxy distribution, and the cross correlation between the two provides strong constraints on the background cosmology. The SDSS dataset has accurate photometry in five bands, so it is well suited to studying a number of galaxy properties, and hence to constraining models of galaxy formation.

Galaxy formation models make specific predictions for how galaxies evolve; in these models, there is a close relation between evolution and environment. The SDSS is not sufficiently deep to allow strong constraints on the models of evolution, but, because it contains a large number of galaxies in a range of environments, it can nevertheless be used to inform galaxy formation models. In what follows, I develop two models of how the environmental dependence of galaxy properties is expected to manifest itself in datasets like the SDSS. Both use the halo model of clustering to illustrate the main features of the argument.

In the halo model (Cooray & Sheth 2002), the nonlinear density field is assumed to be made up of dense objects called halos. At any given time, halos of different masses all have the same density (they are all approximately two hundred times denser than the background). All mass is in such halos, and so all galaxies are also associated with dark halos. Moreover, all physical and statistical quantities of interest are thought of as arising from two types of terms: those which reflect the properties within a single halo, and those which involve more than one halo. For instance, the velocity of a randomly chosen particle is the vector sum of the motion with respect to the center of mass of the halo in which it sits, and the motion of the halo center of mass.

Similarly, the two-point correlation function $\xi(r)$ is determined by the sum of two types of galaxy pairs: pairs in the same halo, and pairs in separate halos. Since the radius of a typical halo at $z = 0$ is less than a Mpc, the one-halo term is negligible on scales larger than a few Mpc. On the small scales where the one-halo term dominates, the shape of $\xi(r)$ is determined by how halo density profiles depend on halo mass, and on how halo abundances depend on mass; on larger scales, $\xi(r)$ is less sensitive to the shapes of halo profiles, and more sensitive to the clustering of the halos themselves.

2. Type-dependent clustering

To illustrate, the power-spectrum, the Fourier transform of $\xi(r)$, is

$$P_{gg}(k) = P_{1h}(k) + P_{2h}(k), \quad (2.1)$$

where

$$P_{1h}(k) = \int dm \frac{dn(m)}{dm} \frac{g_2(m)}{\bar{\rho}^2} |u(k|m)|^2 \quad (2.2)$$

$$P_{2h}(k) \approx \left[\int dm \frac{dn(m)}{dm} \frac{g_1(m)}{\bar{\rho}} b(m) \right]^2 P_{\text{Lin}}(k) \quad (2.3)$$

with

$$\bar{\rho} = \int dm \frac{dn(m)}{dm} g_1(m), \quad (2.4)$$

$$g_1(m) \equiv \langle N_{gal}|m \rangle \quad \text{and} \quad g_2(m) \equiv \langle N_{gal}(N_{gal} - 1)|m \rangle. \quad (2.5)$$

Here $dn(m)/dm$ denotes the number density of halos of mass m , $u(k|m)$ is the Fourier transform of the halo density profile, $g_1(m)$ and $g_2(m)$ denote the first and second factorial moments of the number of galaxies in m -halos, $b(m)$ is the m -halo bias factor, and $P_{\text{Lin}}(k)$ is the power spectrum of the mass in linear theory. The clustering of galaxies is the same as the clustering of the mass only if $g_1(m) \propto m$ and $g_2(m) \propto m^2$; if not, the galaxies are biased tracers of the dark matter distribution. Galaxies of different types cluster differently; for instance, more luminous galaxies cluster more strongly, and the correlation function on Mpc scales is steeper for red galaxies than for blue. In the halo model, this difference arises primarily because the dependence on halo mass of g_1 and g_2 depends on galaxy type. Although, in principle, $u(k|m)$ may also depend on galaxy type (e.g. the red galaxies in a cluster tend to be more concentrated around the cluster center than blue galaxies), accounting for this is only important on scales which are substantially smaller than a Mpc (in most cases of practical interest).

3. Clustering and environment

Most galaxy formation models assume that the formation histories of halos of the same mass are independent of their environments (e.g. the models discussed by Benson in these proceedings). The reason for this assumption is most clearly illustrated by the excursion set approach (Bond *et al.* 1991; Lacey & Cole 1993; Sheth 1998). The idealization is that objects form from the spherical collapse of initially overdense regions; as the collapse proceeds, initially concentric shells remain concentric. Consider such an initially overdense region in a Gaussian random field. The density of this patch is determined by the superposition of waves of many different wavelengths. Small scale waves determine the structure of the interior of the patch, and hence its formation history, whereas large scale modes determine the environment of the patch. In a Gaussian field, these waves

are all independent of one another. So, if the initial fluctuation field was Gaussian (as the CMB suggests), then the assumption that the formation histories patch should be approximately independent of its environment is likely to be accurate (e.g. White 1996). Since these models also assume that the formation history of a halo determines the number of galaxies which form within it, as well as their spatial distribution around the halo center, in these models, the relations $g_1(m)$ and $g_2(m)$ used above, and also the profile shapes $u(k|m)$ are independent of environment. Thus, in these models, all environmental effects arise from the fact that halo abundances depend on environment.

The halo model description of clustering suggests the following test of such models.

- Measure $\xi(r)$ in the full data set.
- Determine the $g_1(m)$ and $g_2(m)$ relations which are required to fit $\xi(r)$.
- Then classify all galaxies by the environment which surrounds them (within 8 Mpc, say). Choose only those galaxies which populate a specified range of environments, and measure their clustering $\xi(r|\delta)$.
- By hypothesis, the relations $g_1(m)$ and $g_2(m)$ do not depend on environment, so they should be independent of δ . Since the spatial distribution of galaxies around halo centers is also assumed to be independent of environment, the only effect of the environment is to modify the halo abundances: $n(m, z|\delta) = [1 + b(m, z)\delta]n(m, z)$. Since this modification is known (e.g. accurate formulae for $b(m, z)$ are available), we can use the relations $g_1(m)$ and $g_2(m)$ measured with relatively high signal-to-noise from the full sample to predict $\xi(r|\delta)$.
- Comparison with the measured shape provides a test of the assumption that the only environmental trends come from the halo abundances.

The poster by Abbas elsewhere in these proceedings illustrates that such a test is certainly feasible with current datasets. The remainder of this section formalizes this argument in the context of the halo model.

We imagine sitting on each m -halo, and counting up the mass M in the volume V which surrounds it. We then select only those halos which are surrounded by volumes V which contain a specified range of M . Let $f(M, V|m)dM$ denote the fraction of m -halos which are surrounded by regions V which contain mass in the range dM about M . In the excursion set model, the number density of such halos is

$$\begin{aligned} n(m)f(M, V|m)dM &= n(m)\frac{f(m|M, V)f(M, V)}{f(m)}dM = \frac{\bar{\rho}}{m}f(m|M, V)f(M, V)dM \\ &= dM n(M, V) N(m|M, V) \end{aligned} \tag{3.1}$$

(Sheth 1998). Thus, the mass density contributed by halos which are embedded in regions of mass $M > M_{min}$ is

$$\bar{\rho}_\delta = \int_{M_{min}}^\infty dM n(M|V) \int_0^M dm N(m|M, V) m = \int_{M_{min}}^\infty dM n(M|V)M. \tag{3.2}$$

In the standard model, the density profile of a halo depends on its mass, but not on the surrounding environment. Therefore, the one-halo term is

$$P_{1h}(k| > M_{min}) = \int_{M_{min}}^\infty dM n(M|V) \int_0^M dm N(m|M, V) \left(\frac{m}{\rho_\delta}\right)^2 |u(k|m)|^2. \tag{3.3}$$

The two-halo term has contributions from pairs which are in the same patch V , and from

pairs in distinct V -patches. The contribution from pairs in different patches is given by

$$P_{2p}(k| > M_{min}) = \left[\int_{M_{min}}^{\infty} dM n(M|V) b(M|V) \times \int_0^M dm \left(\frac{m}{\bar{\rho}_\delta} \right) N(m|M, V) u(k|m) \right]^2 P_{Lin}(k) \quad (3.4)$$

where we have again assumed that $u(k|m, M, V) \rightarrow u(k|m)$. As a check, notice that when $M_{min} \rightarrow 0$ then $\bar{\rho}_\delta \rightarrow \bar{\rho}$. A little algebra shows that $P_{1h}(k| > 0)$ equals the expression for $P_{1h}(k)$ we had previously, and that $P_{2p}(k| > 0)$ reduces to $P_{2h}(k)$.

Galaxies are described simply by replacing the weighting by m in the expression for $\bar{\rho}_\delta$ with a weighting by $\langle N_{gal}|m, M, V \rangle$, which explicitly allows for the fact that the number of galaxies in an m halo may depend on its environment. If it does not, then $\langle N_{gal}|m, M, V \rangle \rightarrow g_1(m)$, and the resulting expression reduces to the standard expression for $\bar{\rho}_g$ in the limit that $M_{min} = 0$. Similarly, in the expression for P_{2p} , one replaces the weighting by $(m/\bar{\rho}_\delta)$ with $g_1(m)/\bar{\rho}_g$, and in the one-halo term, $(m/\bar{\rho}_\delta)^2 \rightarrow g_2(m)/\bar{\rho}_g^2$. These expressions show clearly that the entire environmental dependence of clustering comes from the fact that the halo population depends on environment. Since the most massive halos populate the densest regions which have the largest bias factors, a generic prediction is that galaxies in dense regions are more strongly clustered.

In practice, the analytic calculation is hampered by the fact that while we may observe the overdensity of galaxies δ_g , we do not observe the quantity which affects the halo abundances: the overdensity of the mass δ . We can get a rough idea of the trend if we use the fact that, if V is large, $\delta_g \approx b_{gal}\delta$, where $1 + \delta = M/\bar{\rho}V$. In this case, we can replace δ in the expressions above with δ_g/b_{gal} . In practice, it is more accurate to simply Monte-Carlo the predicted dependence on environment by populating halos in a simulation with the desired g_1 and g_2 relations, and then making the same measurement in the mock galaxy catalog as one does in the data.

4. Galaxies and halo substructure

A simple model for the galaxy distribution associates galaxies with the subclumps of dark matter halos which survive the processes of tidal stripping as they orbit within their parent halos (e.g. Kravtsov *et al.* 2004). If m and M denote the masses of the subclump and the parent respectively, then simulations suggest that, on average, the mass function of halo subclumps is a universal function of m/M . (In fact, Kravtsov *et al.* note that estimating subclump masses is difficult, so some estimate of the circular velocity, which is much less sensitive to the precise definition of where the edge of the subclump is, is almost certainly a better indicator of subclump properties. They find that the subclump distribution is a universal function of v_c/V_c .) In particular, a power-law,

$$\frac{dN_{sat}(m|M)}{dm} = N_0 \left(\frac{M}{m} \right)^\mu \frac{dm}{m} \quad \text{with } \mu \approx 0.9 \quad (4.1)$$

provides a good description (e.g. Ghigna *et al.* 2000). The constant N_0 is set by noting that the mass fraction in subclumps is

$$\int dm \left(\frac{m}{M} \right) \frac{dN_{sat}(m|M)}{dm} = N_0 \frac{1 - (m/M)^{1-\mu}}{1 - \mu} \quad (4.2)$$

where we are imagining that the lower limit of the integral $m/M \ll 1$. Simulations show that this mass fraction is approximately ten percent for all M :

$$N_0 = 0.1 \left(\frac{f}{0.1} \right) \frac{1 - \mu}{1 - (m/M)^{1-\mu}}. \quad (4.3)$$

Moreover, simulations suggest that the cumulative number of massive satellites of M halos follows a Poisson distribution with mean

$$N_{sat}(> m|M) = \int_m^M dm \frac{dN_{sat}(m|M)}{dm} = N_0 \frac{(M/m)^\mu - 1}{\mu} \quad (4.4)$$

(Kravtsov *et al.* 2004).

If there is a monotonic relation between subclump mass and luminosity (or circular velocity), then the number density of subclumps which host galaxies more luminous than L is $N_{sat}(> m(L)|M) \propto [M/m(L)]^\mu$ when $m \ll M$. In addition to populating halo subclumps, it is likely that there is one galaxy at or close to the center of each sufficiently massive halo. If this galaxy has the same $m(L)$ relation as the other satellite galaxies, then the number of galaxies in a halo can be written as

$$\langle N_{gal}|M \rangle = 1 + N_{sat}(> m(L)|M) \quad \text{where } M > m(L) \quad (4.5)$$

and recall that N_{sat} is a Poisson variate with mean $\propto [M/m(L)]^\mu$. The luminosity dependence of clustering seen in the SDSS data is well described by this model provided $m(L)$ is chosen to increase with L so as to match the observed number density.

Recent simulations (e.g. Moore in these proceedings) suggest that there may be a problem with this model, because the spatial distribution of subclumps around the centres of their parent halos appears to be shallower than is observed for the distribution of luminous galaxies around cluster centres in the CNOC survey. However, this is a mass-dependent statement, and leads to the following discussion. Since the baryons, and hence the stellar mass, are almost certainly more centrally concentrated than the dark matter, and the stellar mass is almost certainly more closely related to the mass of the subclump before tidal effects stripped-off the mass in the outer layers, it would be interesting to compare the subclump mass function in the simulations before and after tidal stripping. It is the mass function prior to stripping which is perhaps more relevant for predicting stellar mass and/or luminosity functions, and the luminosity dependence of clustering. Such a measurement is not yet in the literature, although simulations are probably up to the task.

5. Weighted or ‘marked’ correlations

Almost all analyses to date treat galaxies as points without attributes. In fact galaxies have luminosities, sizes, shapes, velocity dispersions, star formation rates, etc. The quality of the data is now sufficiently good that one can imagine measuring the spatial correlations of these attributes. That is to say, rather than measuring clustering as a function of luminosity, one can now measure the clustering of luminosity (or of color, star-formation rate etc.).

The halo-model is the natural language within which to discuss marked correlations. In the standard model, all observables are assumed to be determined by the formation history of the halo, which, in turn, depends on mass. In what follows we will use $W(m|M)$ to denote the observable (W here stands for ‘weight’), and we will assume that the distribution of the weight W at fixed m and M is narrow (i.e., we ignore stochasticity in

the relation between observable W and m and M). Then the two-halo term is particularly straightforward to model. If we define

$$\langle N_{gal}|M \rangle \equiv \int dm N(m|M) \quad \text{and} \quad \langle W|M \rangle \equiv \frac{\int dm N(m|M) W(m|M)}{\int dm N(m|M)}, \quad (5.1)$$

then the average number density of galaxies is

$$\bar{\rho}_g \equiv \int dM \frac{dn(M)}{dM} \int dm N(m|M) \equiv \int dM \frac{dn(M)}{dM} \langle N_{gal}|M \rangle, \quad (5.2)$$

the average weight is

$$\bar{W} \equiv \int dM \frac{dn(M)}{dM} \frac{\langle N_{gal}|M \rangle}{\bar{\rho}_g} \langle W|M \rangle, \quad (5.3)$$

and the two-halo term is

$$\mathcal{W}_{2h}(k) = \left[\int dM \frac{dn(M)}{dM} \frac{\langle N_{gal}|M \rangle}{\bar{\rho}_g} \frac{\langle W|M \rangle}{\bar{W}} b(M) u(k|M) \right]^2 P_{Lin}(k). \quad (5.4)$$

(Notice that when all the weights are equal, $W = \bar{W}$, this expression reduces to the usual expression for the two-halo contribution to the galaxy power spectrum.) If we define $B_W(M) \equiv b(M) \langle W|M \rangle / \bar{W}$, then the expression above shows clearly that the large scale power spectrum is simply a linearly biased version of the dark matter spectrum, where the bias factor depends on the particular observable used as the weight.

The one halo term is slightly more complicated. If we assume that the spatial distribution of subclumps does not depend on subclump mass (in fact, it almost certainly does), and that the luminosity of a galaxy in an m -subclump is independent of whether or not there are other galaxies in the same halo, then

$$\mathcal{W}_{1h}(k) = \int dM \frac{dn(M)}{dM} \frac{\langle N_{gal}(N_{gal} - 1)|M \rangle}{\bar{\rho}_g^2} \frac{\langle W|M \rangle^2}{\bar{W}^2} |u(k|M)|^2. \quad (5.5)$$

To see what this implies, suppose that the weight is proportional to subclump mass. Then

$$\begin{aligned} \langle W|M \rangle &\rightarrow \frac{fM}{\langle N_{gal}|M \rangle}, \quad \text{so} \quad \bar{\rho}_g \bar{W} \rightarrow f\bar{\rho}, \quad \text{and} \quad \frac{\langle W|M \rangle}{\bar{W}} \rightarrow \left(\frac{M}{\bar{\rho}} \right) \frac{\bar{\rho}_g}{\langle N_{gal}|M \rangle}; \\ \mathcal{W}_{1h}(k) &\rightarrow \int dM \frac{dn}{dM} \left(\frac{M}{\bar{\rho}} \right)^2 \frac{\langle N_{gal}(N_{gal} - 1)|M \rangle}{\langle N_{gal}|M \rangle^2} |u(k|M)|^2, \\ \mathcal{W}_{2h}(k) &\rightarrow \left[\int dM \frac{dn}{dM} \left(\frac{M}{\bar{\rho}} \right) b(M) u(k|M) \right]^2 P_{Lin}(k). \end{aligned}$$

If the number of galaxies in M haloes, $p(N_{gal}|M)$, follows a Poisson distribution, then $\langle N_{gal}(N_{gal} - 1)|M \rangle = \langle N_{gal}|M \rangle^2$, so $\mathcal{W}_{1h} \rightarrow \int dM dn/dM (M/\bar{\rho})^2 |u(k|M)|^2$. If the integral is over all M , then this integral equals that for the dark matter. Thus, if the mass-to-light ratio is independent of galaxy mass (it is not) then the galaxy luminosity weighted marked correlation function will have the same shape as that of the dark matter correlation function. A better approximation perhaps, is to use the dynamical mass estimator $R\sigma^2$ as the mark. Such a measurement is within reach in current datasets. More generally, this analysis reveals an important point: if the subclump mass function is a universal function of m/M , then the marked correlation function term $\langle N_{gal}|M \rangle \langle W|M \rangle$ scales with the same power of M as does the mark with m .

The expressions above have no environmental trends other than those which come from the dependence of halo abundances on environment. In essence, they represent the framework within which to describe the predictions of standard galaxy formation models. The extent to which simple halo-model calculations such as this are able to reproduce measurements of marked correlations in real data provides a test of the standard assumption that galaxy properties are more closely related to the formation histories of their parent halos, rather than to their environments.

6. Conclusions

We presented two tests of a fundamental assumption in galaxy formation models: that environmental effects are primarily a consequence of the fact that the halo population depends on environment (e.g. the most massive halos populate the densest regions). One test uses the clustering of galaxies as a function of environment. The other uses the statistics of weighted or marked correlations. In both cases, we used the halo model to illustrate what the standard model predicts for measurements which can easily be made in datasets which are currently available. Application of our results to the SDSS is underway. In addition, it would be interesting to apply these tests to the galaxy formation models of van Kampen (which explicitly include environmental effects) and compare the results with the models of Benson (which explicitly assume that the only environmental effects come from the correlation between halo masses and environment).

Acknowledgements

We are grateful to the organizers, and Antonaldo Diaferio in particular, for scheduling the meeting during Torino's chocolate festival. We would also like to thank some members of Torino's police department for their role in an incident involving a lost wallet. RKS is grateful for the receipt of an IAU grant. This work is supported by NASA and the NSF under grants NAG5-13270 and AST-0307747.

References

- Abazajian, K., *et al.* 2003 *A.J.* **126**, 2081.
Bennett, C. *et al.* 2003 *Ap.J.S.* **148**, 97.
Bond, J. R., Cole, S., Efstathiou, G. & Kaiser, N. 1991 *Ap.J.* **379**, 440.
Colless, M. *et al.* 2001 *Ap.J.* **328**, 1039.
Cooray, A., & Sheth, R. K. 2002 *Phys. Rep.* **372**, 1.
Ghigna, S. *et al.* 2000 *Ap.J.* **544**, 616.
Kravtsov, A. *et al.* 2004 *Ap.J.* submitted, (astro-ph/0308519)
Lacey, C., & Cole, S. 1993 *M.N.R.A.S.* **262**, 627.
Mo, H. J. & White, S. D. M. 1996 *M.N.R.A.S.* **282**, 347.
Sheth, R. K. 1998 *M.N.R.A.S.* **300**, 1075.
White, S. D. M. 1996 In *Cosmology and Large Scale Structure, Les Houches Session LX* (eds. R. Schaefer, J. Silk, M. Spiro & J. Zinn-Justin), pp. 349. Elsevier.