Carnegie Observatories Astrophysics Series, Vol. 3:
Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution, 2004
ed. J. S. Mulchaey, A. Dressler, and A. Oemler (Pasadena; Carnegie Observatories:
http://www.ociw.edu/ociw/symposia/series/symposium3/proceedings.html)

Interpreting SDSS Cluster Masses and Abundances with Mock Catalogs

R. H. WECHSLER
Physics Department, University of Michigan
present address: Center for Cosmological Physics, University of Chicago

Abstract

We describe the creation of large mock galaxy catalogs with properties (including number density, luminosity function, color distribution and evolution, and luminosity and color dependent clustering) that match those of bright galaxies in the Sloan Digital Sky Survey (SDSS). The catalogs associate galaxies with dark matter in the Hubble Volume simulation, which tracks the clustering of dark matter particles on a light cone out to $z \sim 0.6$, and represents a full-sky survey of galaxies brighter than $0.4L_*$ ($M_r - 5\log h < -19.8$) to this redshift. The mock catalogs are used to test color-based cluster finding algorithms that are being applied to galaxies in the SDSS, namely the maxBCG algorithm and the C4 algorithm. For each cluster finding method, we will be able to detail the accuracy with which we can recover the relation between observed cluster properties and dark halo masses, directly predict the cluster abundance function for a given cosmological model, and determine the accuracy with which we can constrain the halo mass function with the full data set.

1.1 Introduction

At completion, the Sloan Digital Sky Survey will detect of order 10^4 massive clusters of galaxies. Since clusters represent the most massive, exponentially decreasing end of the halo mass function, their local abundance has long had promise as a sensitive cosmological probe. However, in order to realize this promise with the large datasets now becoming available, one needs an understanding of how observed clusters are related to the underlying dark matter halos. Pinning down the mass scale of observed clusters is also useful for constraining galaxy formation theories that predict relations between cluster members and their host halos. Here, we describe the creation of mock galaxy catalogs, that populate a simulated dark matter distribution with 'observable' galaxies, with properties similar to those measured in the local universe. Since galaxy formation theory is not yet at the level where it can produce galaxy populations with accurate enough properties, we take an empirical approach in our aim to produce a catalog of galaxies that matches a number of galaxy properties of the actual SDSS data set, including the luminosity and color evolution and the luminosity-color-density relations, in a volume large enough to encompass the entire survey. The mock catalogs are used to test and calibrate cluster finding methods being used on the SDSS data, to calibrate the relation between halo mass and observed cluster properties, and to directly predict the cluster abundance function for a given cosmological model.

1.2 Mock Galaxy Catalogs

1.2.1 Dark Matter

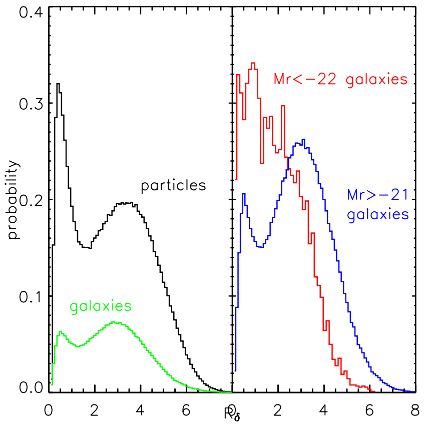
The distribution of dark matter in the Hubble Volume simulation* provides the basis for our mock catalogs. The simulation is of a flat ΛCDM universe, with $\Omega_{\mathrm{m}} = 0.3$, $\sigma_8 = 0.9$, and $h = 0.7$. The evolution of 10^9 particles of mass $2.25 \times 10^{12} h^{-1} \mathrm{M}_\odot$ are traced in a periodic cubical volume with side length $3h^{-1}$Gpc , from z=35 to the present, with a force resolution of $0.1h^{-1}$Mpc . We use the sky survey output described by Evrard et al. (2002), which mimics the collection of data on the past light cone of an observer located at the center of the volume. The large size of the simulation allows the creation of a full-sky survey out to a depth of $z_{\mathrm{max}} = 0.57$. Further simulation details are presented in Evrard et al. (2002).

1.2.2 Galaxy Luminosities and Bias

It has long been recognized (e.g., Kaiser, 1986) that galaxies are biased tracers of the matter distribution. Previous generations of mock catalogs have often used a parameterization of the local bias of the galaxy population as a whole, or related galaxies to dark matter in terms of a "Halo Occupation Distribution" (HOD), which parameterizes the relation between galaxies and their dark matter halos (see e.g., Berlind & Weinberg 2002 and references therein). Since we are particularly interested in testing cluster finding algorithms which depend on galaxy luminosity and color, in this work we would like to go a step beyond previous generations of mock catalogs, and match not only *global* clustering properties, but clustering properties as a function of luminosity and color. This requires a bias scheme which depends on galaxy luminosity and color. A recent advance at reproducing the luminosity dependent clustering has recently been made using the idea of a conditional luminosity function combined with a halo occupation scheme (e.g., Yang, Mo & van den Bosch 2003). We present a complementary method, in which dark matter particles are chosen based on their local density, without respect to the mass of the halos they reside in. This has the advantage that it directly connects the dark matter density field to galaxy properties (luminosities in 5 wavebands), without making assumptions about the form of the HOD or the conditional luminosity function, and that it allows us to populate a simulation with galaxies that reside in halos below the resolved halo mass. This technique is further justified by recent indications that the luminosity of a galaxy does in fact depend more strongly on its local density than on the mass of the group it resides in (Balogh et al., 2004).

We assume that galaxies follow a Schechter luminosity function locally, and use the parameters estimated by Blanton et al. (2003) for SDSS galaxies in the r-band LF at z=0.1 ($\phi_* = 0.0149$, $M_* - 5\log h = -20.44$, and $\alpha = -1.05$). Galaxies are chosen from this luminosity function integrated down to $M_r - 5\log h = -19.8$, giving a galaxy number density of 0.0106 per h^{-3}Mpc^3 (compared with a particle number density of 0.037 per h^{-3}Mpc^3). We treat the dark matter particles as tracers of the underlying density field, and assign a unique dark matter particle to each galaxy, chosen from the distribution $\mathrm{P}(\delta_{\mathrm{dm}} \mid M_r)$, where δ_{dm} is a measure of the local dark matter density around particles The method of setting this distribution as a function of luminosityis described below.

* http://www.physics.lsa.umich.edu/hubble-volume/



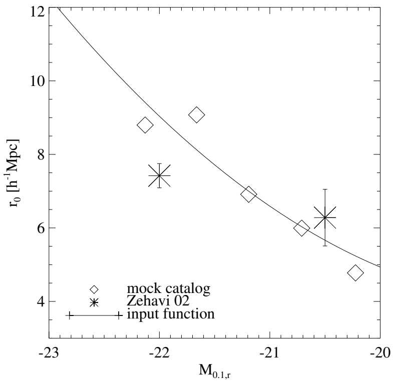


Fig. 1.1. **Left:** Left panel shows the PDF of R_δ for dark matter particles and for galaxies in our fiducial catalog (galaxy probability is normalized to the number of particle from which they are selected). Right panel shows the same PDF for bright and dim galaxies. **Right:** Correlation length as a function of r-band magnitude in our fiducial catalog. This is compared with data points from the SDSS EDR (Zehavi et al 2002), and with the input function we use, interpolated from more recent results (Zehavi et al in prep).

1.2.3 Tuning the luminosity-density relation

We assume that the basic physics setting galaxy luminosities and colors occurs on a mass scale around $M*$, and use the radius which encloses this mass (denoted R_δ) to characterize the local galaxy density. For the Hubble Volume simulation, this corresponds roughly to the distance to the eighth nearest dark matter particle (which encloses $1.8 \times 10^{13} h^{-1} \mathrm{M}_\odot$). Note that we are considering the distribution of a Lagrangian local density (measured on a fixed mass scale), which does not have the familiar lognormal form of Eulerian density measures. We find that the probability distribution of R_δ (directly related to $\delta_{\rm dm}$) around dark matter particles can be represented by a sum of a Gaussian, corresponding roughly to the field population, and a log-normal distribution peaked at higher densities, corresponding to the population of galaxies in clusters (shown in Figure 1.1a (upper left). We make the *assumption* that the probability distribution of R_δ around galaxies of any magnitude has the same functional form as that around particles, for some choice of parameters. This assumption seems quite reasonable; allowing the parameters to be free allows a very wide range of distributions, and we additionally find that both galaxies in a simulation populated with a semi-analytic prescription (Kauffmann et al., 1999) and galactic mass subhalos in a high-resolution simulation have distributions which are well matched by this form.

The distribution has five free parameters, which for a population of galaxies can each be in principle a function of magnitude: the mean and dispersion of the Gaussian and lognormal distributions, and the relative heights of the two distributions (corresponding roughly to the fraction of "field" galaxies). For a given cosmological model, we want to assign galaxies to the simulated dark matter density field by choosing from the parent density PDF in such a way that the galaxy correlation function as a function of magnitude matches the observed data. We assume the correlation function can be fit by a power law, $\xi(r) = (r/r_0)^{-\gamma}$, over the range $1 \leq r \leq 10$. Observed SDSS correlation function measurements are taken from

Zehavi et al (in prep); we interpolate the measured values and their errors to find $r_0(M_r)$ and $\gamma(M_r)$. We then take the following steps to choose 'galaxies' in the simulation that match these clustering properties:

(1) For a gridded range of possible density PDFs, we select simulation particles to satisfy that PDF and directly compute their correlation function, fitting the parameters r_0 and γ.
(2) We limit the range considered slightly, only considering those values which give correlation functions close to those observed in the real data. We place some moderate additional limits that the values cannot be drastically different from values for the particle distribution or from those measured around subhalos in a high resolution simulation.
(3) To this data, we fit a simple model for both r_0 and γ, such that they are each a function of linear-plus-interaction terms in the five parameters specifying the density PDF. This model is able to reproduce the measured values within about 5% for the entire range of parameter space, which is about the same as the cosmic variance errors.
(4) We produce a finely gridded set of points in the PDF parameter space, and use the model in (3) to find the expected r_0 and γ values for each point.
(5) We then step through the galaxies we want to place in the simulation in small luminosity bins and perform the following steps:
   (a) Choose a point from the PDF parameter space, weighting each point by the likelihood that its correlation function parameters are consistent with the observed values for that magnitude (assuming the measured error bars are Gaussian).
   (b) Assign the galaxies to particles chosen from the simulation such that their local density distribution matches the PDF chosen in (5a).
   (c) Add passive luminosity evolution to each galaxy, using the best estimate by Blanton et al. (2003): $M_*(z) = M_*(z=0.1) - Q(z-0.1)$, where $Q = 1.62$.

Following this procedure results in a catalog whose luminosity-dependent correlation function is in agreement with the observed one. We can repeat this process many times to produce multiple catalogs, and find that the uncertainty in assigning the galaxies by this method does not result in significant scatter in the halo occupation (See Wechsler et al in prep for further details.) The density PDFs selected for several magnitude ranges of galaxies are shown in Figure 1.1a. The correlation length of mock galaxies is shown plotted as a function of luminosity in Figure1.1b.

1.2.4 Galaxy Colors

Above, we demonstrated a method of constructing a galaxy catalog that matches the luminosity function and luminosity-dependent clustering in one waveband. However, since the primary goal of this work is to test cluster finding methods which depend sensitively on the evolution of galaxy color with redshift and on clustering in color-space, it's essential to correctly match the distribution and environmental dependence of color and their evolution with redshift. In order to do this, each mock galaxy is assigned photometric properties chosen directly from observed SDSS galaxies. We first construct a volume-limited sample of galaxies with measured redshifts from SDSS DR1 (Abazajian et al., 2003). This sample includes galaxies in the range $0.05 < z < 0.095$. For each galaxy, we measure a local bright galaxy surface density Σ_{10}, equal to the distance to the tenth nearest galaxy, within 1000 km/s, which is brighter than $M_r - 5\log h$=-19.8. In order to properly model the evolution of color with redshift, we use the method of Blanton et al. (2002, `kcorrect v1_11`), which

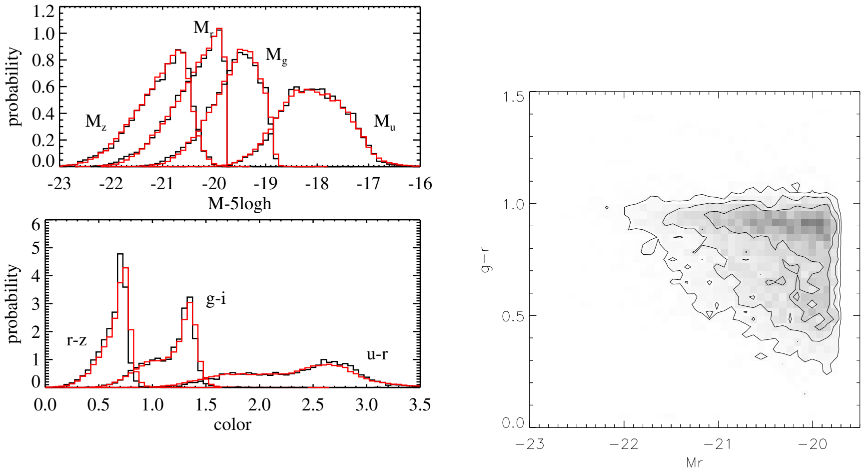


Fig. 1.2. **Top Left:** Galaxy luminosity distribution in the u,g, r, and z for our fiducial mock catalog, compared to a volume limited SDSS DR1 sample. **Bottom Left:** Color distributions in $u-r$, $g-i$, and $r-z$, for a volume limited SDSS sample and our fiducial mock catalog. **Right:** Joint distribution of g-r color and r-band magnitude in our fiducial catalog. This can be compared SDSS data in Blanton et al. (2001) Figure 13.

fits a linear combination of four templates to each set of five observed magnitudes, to measure SED templates for each observed galaxy. For each mock galaxy, we then choose a real galaxy which has a similar r-band luminosity and local galaxy density (as characterized by Σ_{10}), and assign the SED of the real galaxy to the mock one. This SED is then reconstructed at the appropriate redshift, and colors are assigned using these reconstructed magnitudes. Noise is added to each luminosity to mock the effects of seeing.

1.3 Comparison with Galaxy Data

Several comparisons with galaxy data are shown. We create volume limited samples in the redshift range $0.05<z<0.95$ in our fiducial mock catalog and the SDSS DR1 data to facilite the comparison. Figure 1.2a shows the luminosity distribution in SDSS u, g, r, and z, and the color distribution for 3 colors, for both the data and mock catalog. Figure 1.2b) shows the joint distribution of g-r color and r-band luminosity for the same sample. Matching the luminosities, colors, and clustering of the data allow us to find clusters using color based methods begin applied to the SDSS (see Nichol, this proceedings; Annis et al in prep; Miller et al in prep).

1.4 Conclusions

We have constructed mock galaxy catalogs, representing a full-sky survey of galaxies, with magnitudes in five SDSS bands, brighter than $0.4L_*$ out to $z \sim 0.6$, targeted at reproducing the volume and properties of the Sloan Digital Sky Survey. The mock galaxies reproduce a number of features of the local universe including: the redshift distribution, luminosity function, color distribution and evolution, bivariate distribution of luminosity and color, local galaxy density as a function of luminosity and color, and two-point correlation function as a function of luminosity and color. The method descibed here puts constraints on the halo occupation function (see Wechsler et al, in prep); we find similar results to the complimentary methods employed in the SDSS (Zehavi et al and Berlind et al, in prep) though our method makes significantly different assumptions. These catalogs have been used to improve cluster finding algorithms in the SDSS (Miller et al, in prep; Annis et al, in prep). The primary power, however, is that since we have created a catalog with clusters resembling those in the real universe, we can run the same cluster finders on the model as are run on the data — and can thus make direct predictions *in observational space* for the cluster abundance and for a number of cluster observables — implicitly taking systematics in the halo selection with various cluster finders into account. We are working to repeat this processwith a suite of simulations with varied cosmological parameters. In each case, the mocks can directly predict the relation between halo mass and cluster observables (e.g., number of bright galaxies, cluster luminosity and velocity dispersion, cluster size, weak lensing mass); cross checking several of these methods should significantly reduce systematic uncertianties. Catalogs of this type are essential if one hopes to realize the full power of the large upcoming SDSS cluster samples for constraining cosmological models and galaxy formation theories.

Acknowledgments

RHW is grateful to her collaborators, A. Evrard, T. McKay, C. Miller, R. Nicol, J. Annis and D. Bingham, for allowing this work to be presented here. This study used dark matter simulations from the VIRGO collaboration and publicly available data from the SDSS (www.sdss.org).

References

Abazajian, K. et al. 2003, AJ, 126, 2081A
Balogh, M. L. et al. 2004, preprint, astro-ph/0311379
Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M., & Lacey, C. G. 2000, MNRAS, 311, 793
Berlind, A. A. & Weinberg, D. H. 2002, ApJ, 575, 587
Blanton, M. R., Brinkmann, J., Csabai, I., Doi, M., Eisenstein, D., Fukugita, M., Gunn, J. E., Hogg, D. W., & Schlegel, D. J. 2003, AJ, 592, 819
Blanton, M. R. et al. 2001, AJ, 121, 2358
—. 2003, ApJ, 592, 819
Evrard, A. E. et al. 2002, ApJ, 573, 7
Kaiser, N. 1986, MNRAS, 222, 323
Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 303, 188
Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, /mnras, 339, 1057