The Astrophysical Journal, 678:569–577, 2008 May 10

MoLUSC: A MOCK LOCAL UNIVERSE SURVEY CONSTRUCTOR

T. Sousbie,[1,2] H. Courtois,[3,4] G. Bryan,[5] and J. Devriendt[1]

Received 2006 December 6; accepted 2007 November 16

ABSTRACT

This paper presents MoLUSC, a new method for generating mock galaxy catalogs from a large-scale ($\approx 1000^3$ Mpc^3) dark matter simulation, which requires only modest CPU time and memory allocation. The method uses a small-scale ($\approx 150^3$ Mpc^3) dark matter simulation on which the GalICS semianalytic code has been run in order to define the transformation from dark matter density to galaxy density using a probabilistic treatment. MoLUSC is then applied to a large-scale dark matter simulation in order to produce a realistic distribution of galaxies and their associated spectra. This permits the fast generation of large-scale mock surveys using relatively low-resolution simulations. We describe various tests that have been conducted to validate the method and illustrate it by generating a mock Sloan Digital Sky Survey redshift survey.

Subject headings: galaxies: statistics — large-scale structure of universe — methods: *n*-body simulations — methods: numerical — methods: statistical

1. INTRODUCTION

Cosmological simulations provide key contributions to studies of the formation of large-scale structure in the universe. For instance, they allow us to objectively test cosmological models against observations and, in doing so, help to rule out models. However, analysis of real surveys using simulations is difficult due to the biases and limitations inherent to any observational data. In order to quantify those biases and make detailed comparisons to theoretical models, we therefore need to generate realistic mock galaxy catalogs with high-mass resolution that can also span a wide range of length scales. In other words, we need to keep up with observational catalogs, such as the Sloan Digital Sky Survey (SDSS) or the Two Degree Field (2dF) survey, which have been mapping out larger and larger regions of the sky for the past decades and nowadays contain information about the galaxy distribution in a substantial volume of the universe.

A common way to generate mock catalogs is to carry out simulations that model a volume at least as large as the survey in question, while at the same time resolving the smallest dark matter (DM) halos that can host galaxies of interest (e.g., Jing et al. 1998; Yan et al. 2003). Once halos are identified in such a simulation, they can be populated using techniques such as the halo occupation distribution formalism (Peacock & Smith 2000; Zhao et al. 2002). However, for the largest surveys, which can easily span a thousand megaparsecs, it is often prohibitively expensive to run such simulations, both in terms of memory and CPU time.

Instead, we describe a method that does not rely on the identification of individual halos but uses a smaller scale simulation, but one higher mass resolution, to constrain the relation between dark matter density and galaxy density. The higher resolution simulation does resolve all the relevant DM halos, and so it can use a more accurate method to populate the halos. In this paper we use the GalICS semianalytic model on this smaller scale simulation, although in principle any technique to associate halos with galaxies could be used. Our method is similar in spirit to Cole et al. (1998) and Hamana et al. (2002), but our method of computing the effective bias is more sophisticated (as it is based on a full-blown semianalytic method).

The large-scale ($\approx 1000^3$ Mpc^3) cosmological DM simulations presented in this paper were run using the GADGET-2 public code (Springel 2005). GADGET-2 is a massively parallel code for hydrodynamical cosmological simulations, although we use only its dark matter capabilities.

The MoLUSC treatment, i.e., converting DM particles into a distribution with galaxy properties, was made using the GalICS public mock galaxy catalogs. The GalICS code (Hatton et al. 2003) describes hierarchical galaxy formation using the so-called hybrid approach. It processes the outputs of large cosmological *N*-body simulations to get a more realistic description of DM halos and runs a semianalytic model to describe the baryons. Because it keeps a record of the spatial and dynamical information, the hybrid approach opens the way to a detailed treatment of galaxy interaction and merging. The GalICS model explicitly intends to address the issue of the high-redshift star formation rate history in a multiwavelength prospect, from the ultraviolet to the submillimeter range.

2. GENERATING GALAXIES FROM DARK MATTER SIMULATIONS

In the following we make the assumption that galaxy spatial distribution is mainly affected by the underlying dark matter distribution and that all other processes influencing it can be considered to be small stochastic fluctuations with no significant impact on large-scale clustering properties. Let S_l be a large-scale dark matter–only simulation, S_s a small-scale high-resolution dark matter–only simulation, and G_s the resulting galaxy distribution obtained by using GalICS on S_s. The process used by MoLUSC to generate a galaxy distribution $G_l^\star$ out of S_l consists of two main steps:

1. The computation of the bias between dark matter and galaxy distribution in S_s and G_s. This is achieved by
   a) sampling the S_s and G_s density fields over *identical* grids [$\rho_{S_s}(\boldsymbol{r}_i)$ and $\rho_{G_s}(\boldsymbol{r}_i)$ hereafter];

[1] Université de Lyon, CRAL UMR5574, 9 avenue Charles Andre, 69561 Saint Genis Laval Cedex, France.
[2] Institut d'Astrophysique de Paris, UMR7095, Université Pierre and Marie Curie-Paris, 98 bis bd Arago, 75014 Paris, France.
[3] Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, HI 96822.
[4] Université de Lyon, IPNL UMR5822, 4 rue Fermi, 69622 Villeurbanne Cedex, France.
[5] Department of Astronomy, Pupin Physics Laboratories, Columbia University, New York, NY 10027.

b) computing the probability $P[\rho_{G_s}(\boldsymbol{r}_i)|\rho_{S_s}(\boldsymbol{r}_i)]$ that for a given grid node *i*, a dark matter density $\rho_{S_s}(\boldsymbol{r}_i)$ and a galaxy density $\rho_{G_s}(\boldsymbol{r}_i)$ would be measured;
c) computing the probability that for a given value of $\rho_{G_s}(\boldsymbol{r})$ and $\rho_{S_s}(\boldsymbol{r})$ a given spectrum would be assigned to a galaxy located in $\boldsymbol{r}$.

2. The generation of a galaxy distribution $G_l^\star$ that respects the probability distributions computed in the first step, while following the large-scale structure distribution of S_l. This result is obtained by
a) sampling the density field of S_l over a grid [$\rho_{S_l}(\boldsymbol{r}_i)$ hereafter].
b) building a density field $\rho_{G_l^\star}(\boldsymbol{r}_i)$ from $\rho_{S_l}(\boldsymbol{r}_i)$ and the probability distribution $P[\rho_{G_s}(\boldsymbol{r}_i)|\rho_{S_s}(\boldsymbol{r}_i)]$.
c) creating a discrete galaxy distribution whose sampled density field is $\rho_{G_l^\star}(\boldsymbol{r}_i)$ and which has a velocity field based on the underlying dark matter distribution.

2.1. *Bias Analysis*

As explained above, the dark matter to galaxy distribution bias is computed from the sampled density fields of S_s and G_s. There exist many efficient ways of sampling a density field from a discrete point distribution. In this case, we want to keep track of which galaxy contributed to which sampling grid node in order to be able to recover the spectral information. The chosen method uses a truncated Gaussian kernel and consists of considering the *i*th particle as a density cloud $W(\boldsymbol{r}-\boldsymbol{r}_i)$ centered on the particle location $\boldsymbol{r}_i$. For a cubic sampling grid with cell size $\sigma\, h^{-1}$ Mpc, the number density at the *k*th node is then given by

$$n(\boldsymbol{r}_k) = \sum_{i=0}^{N} W(\boldsymbol{r}_k - \boldsymbol{r}_i), \tag{1}$$

and its mass density is

$$\rho(\boldsymbol{r}_k) = \sum_{i=0}^{N} m_i W(\boldsymbol{r}_k - \boldsymbol{r}_i), \tag{2}$$

where m_i is the mass of the *i*th particle and N the total number of particles.

The kernel function is chosen to be a truncated Gaussian of the form

$$W(\boldsymbol{r}) = \frac{A}{(4\pi L^2)^{3/2}} \exp\left(-\frac{\|\boldsymbol{r}\|}{2L^2}\right)\Pi(\Delta\sigma - \|\boldsymbol{r}\|), \tag{3}$$

where A is a normalization constant, L is the smoothing length, Π is the Heaviside function, and Δ sets the truncation length. Given the infinite extent of the Gaussian function, the kernel is truncated in order to reduce the computation time. Experiments show that choosing a sampling length equal to the smoothing length $\sigma = L$ (all information being wiped out at scales smaller than L) and setting the truncation length to at least $\Delta = 3$ provides good results. For this analysis, a value of $\Delta = 5$ was used, the error on the measurement compared to an infinite-extent kernel being of order 10^{-7} in such a case for a homogeneous field.

Once the sampled density fields $\rho_{S_s}(\boldsymbol{r}_i)$ and $\rho_{G_s}(\boldsymbol{r}_i)$ are computed from S_s and G_s, the probability $P(n_G|\rho_S)$ for the galaxy number density at a location $\boldsymbol{r}$ to be $n_G(\boldsymbol{r})$, knowing that the dark matter mass density at this same location is $\rho_S(\boldsymbol{r})$, can be obtained. This is achieved by applying the following equations:

$$\begin{aligned} P(n_G|\rho_S) &\propto \sum_{k=1}^{N_n} V[\rho_S(\boldsymbol{r}_k) - \rho_S]V[n_G(\boldsymbol{r}_k) - n_G], \\ \int_0^\infty P(n_G|\rho_S)\, dn_G &= 1, \end{aligned} \tag{4}$$

where the sum is computed over the N_n nodes of the sampling grids (which are identical for both distributions) and $V(\boldsymbol{r}) = 1$ if $|x| < H/2$ and $V(\boldsymbol{r}) = 0$ otherwise, H being the probability sampling bin size. Figure 1 shows the function $P(n_G|\rho_S)$ for three different redshifts ($z = 3$, 1, and 0) computed from a 512^3 particle simulation enclosed in a 1000 h^{-1} Mpc box dark matter simulation and its GalICS counterpart. As expected, two different regimes exist: when dark matter density is low, no galaxy formation can occur, whereas for high densities, the galaxy number density becomes proportional to the dark matter mass density ($n_{G_s} = b\rho_{S_s}$). Between these two regimes, a large range of galaxy densities can correspond to a given dark matter density depending on the galaxy formation history. At high redshifts (Fig. 1, *top panels*), a change can be observed for $P(n_G|\rho_S)$ at high ρ_S values. This can be explained by the fact that for $z = 3$, the gravitational collapse of the most massive halos has not yet occurred. Hence, the halo mass function is dominated by small-size objects that cannot all be detected within the simulation due to resolution limitations. In fact, as in the hierarchical model framework, large halos are formed by smaller halos fusions, and their galaxy content is weaker than expected at high redshifts.

The function $P(n_G|\rho_S)$ describes how the galaxy distribution maps out the underlying dark matter distribution but does not contain any information about galaxy properties. In order to be able to attribute a spectrum to every galaxy created from S_l, we compute the probability distribution $P[F_i(\lambda)|n_G, \rho_S]$ that a given galaxy located at point $\boldsymbol{r}$ has a given spectrum $F_i(\lambda)$, knowing that the galaxy number density is $n_G(\boldsymbol{r})$ and the dark matter mass density is $\rho_S(\boldsymbol{r})$. This distribution is obtained from G_s and S_s, using the N synthetic spectra $F_i(\lambda)$ generated by GalICS in G_s:

$$\begin{aligned} P[F_i(\lambda)|n_G, \rho_S] &\propto \\ &\sum_{j=1}^{N_n} \delta\big[\rho_S(\boldsymbol{r}_j) - \rho_S\big]\delta\big[n_G(\boldsymbol{r}_j) - n_G\big]W\big(\boldsymbol{r}_i - \boldsymbol{r}_j\big), \\ &\sum_{i=1}^{N} P(F_i(\lambda)|n_G, \rho_S) = 1. \end{aligned} \tag{5}$$

In this equation, $\boldsymbol{r}_i$ is the location of the *i*th galaxy in G_s, $\boldsymbol{r}_j$ the coordinates of the *j*th grid node, and $F_i(\lambda)$ the spectrum associated to the *i*th galaxy. This simply states that the probability for a given type of galaxy to exist at a place where the galaxy and dark matter densities are n_G and ρ_S is proportional to the number of galaxies of that type observed in G_s close to grid nodes, each being weighted by their respective contribution to the value of n_G at that node [hence the factor $W\big(\boldsymbol{r}_i - \boldsymbol{r}_j\big)$]. From a more practical point of view, a list of spectra is attributed to every pixel in Figure 1 with a weight associated to each of them.

2.2. *Generating the Galaxy Distribution*

The first step of the process consists in extracting information from S_s and G_s through the computation of probability distributions $P(n_G|\rho_S)$ and $P(F_i(\lambda)|n_G, \rho_S)$. The second step generates a galaxy distribution $G_l^\star$ out of the large-scale dark matter simulation S_l using $P(n_G|\rho_S)$, and $P[F_i(\lambda)|n_G, \rho_S]$. To do so, we first compute the sampled galaxy number density field $n_{G_l^\star(\boldsymbol{r}_i)}$ corresponding to S_l. Using the same technique and grid parameters as before, the sampled mass density field $\rho_{S_l}(\boldsymbol{r}_i)$ is extracted, and for every grid node *i*, a corresponding value of $n_{G_l^\star(\boldsymbol{r}_i)}$ is randomly selected, following the probability distribution $P(n_G|\rho_S)$. With the large-scale simulation spanning on a larger scale than the reference one, one could expect that densities span on different

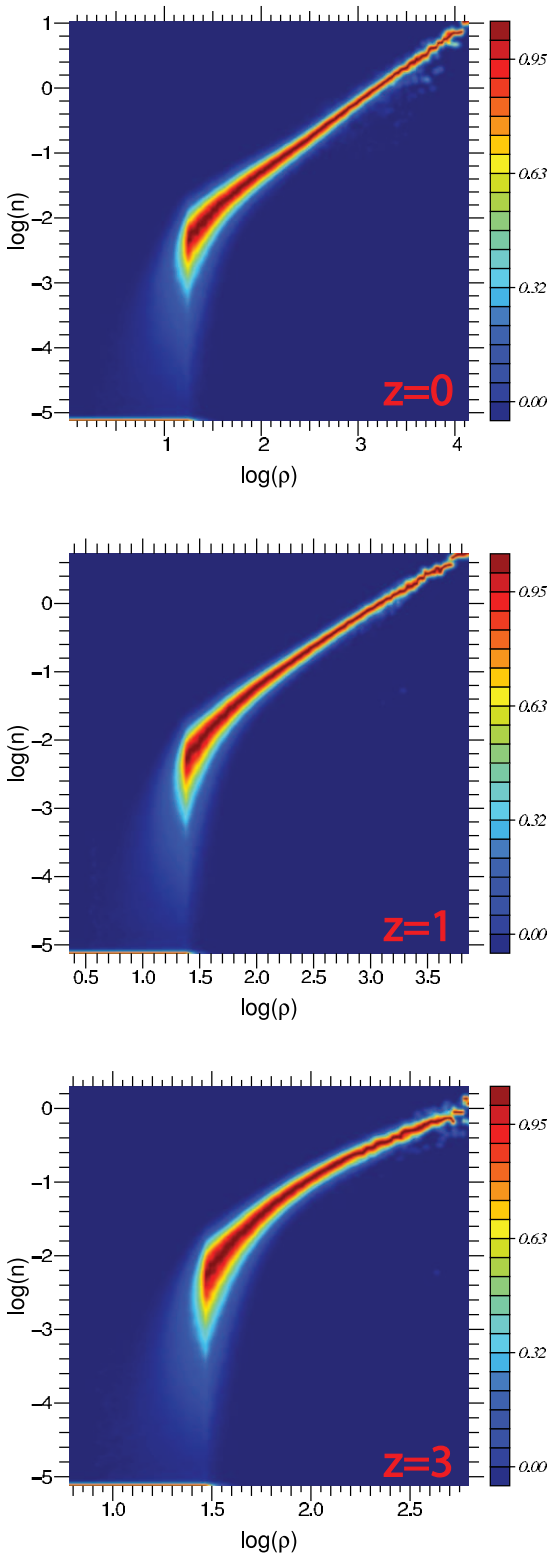

FIG. 1.—Probability $P(n_G|\rho_S)$ that a given galaxy number density n_G corresponds to a DM mass density ρ_S (logarithmic scales) for three different redshifts ($z = 3$, 1, and 0). For low DM densities, no galaxy formation occurs and the galaxy number density is thus 0, whereas it appears to be roughly proportional to the DM density ρ_S for high values of ρ_S. In between these two regimes, a wide range of galaxy densities can correspond to a given DM density. See the main text for an explanation of the change of behavior with redshift.

ranges too. It is thus possible that some densities ρ_S are never measured in the small-scale simulation, and thus the corresponding value of n_G is unknown. This is of no consequence for small values of ρ_S, as one can set $P(n_G|\rho_S) = 0$ in this situation. This is not the case for large ρ_S, but in practise this kind of situation seldom happens: whereas the occurrence of unfrequent events such as large densities should be more probable over large scales, the lack of resolution in the large-scale simulation tends to smooth these peaks. In any case, when this does occur, the value of ρ_S is set to the highest measured value of ρ_s [one could also have linearly interpolated the value of $P(n_G|\rho_S)$, but this appears to have a negligible influence on the resulting distribution].

The galaxy distribution $G_l^\star$ can then be generated by creating a point distribution whose sampled density field is $G_l^\star$. Let

$$\|W(\boldsymbol{r})\| = \sum_{i=1}^{N} W(\boldsymbol{r} - \boldsymbol{r}_i) \tag{6}$$

be the sampled norm of the kernel W, with i the index of a grid node. The fact that we used a *truncated* Gaussian kernel implies that $\|W(\boldsymbol{r})\|$ is not exactly constant, but rather σ-periodic. It is nonetheless possible to show empirically that when $L \leq \sigma$,

$$\frac{\max_{\boldsymbol{r}} \|W(\boldsymbol{r})\| - \min_{\boldsymbol{r}} \|W(\boldsymbol{r})\|}{\min_{\boldsymbol{r}} \|W(\boldsymbol{r})\|} \leq 1\%, \tag{7}$$

where $\min_{\boldsymbol{r}}[f(\boldsymbol{r})]$ and $\max_{\boldsymbol{r}}[f(\boldsymbol{r})]$ are the minimal and maximal values of $f(\boldsymbol{r})$. So as long as $L \leq \sigma$, it is possible to consider that $\|W(\boldsymbol{r})\|$ is a constant and the total number of galaxies $N_{G_l^\star}$ in $G_l^\star$ is

$$N_{G_g^\star} = \frac{1}{\|W(\boldsymbol{r})\|} \sum_{i=1}^{N_g} n_{G_g^\star}(\boldsymbol{r}_i), \tag{8}$$

N_g being the total number of grid nodes.

Following the hypothesis that on the scale at which we are working (i.e., of order 1 h^{-1} Mpc), the galaxy distribution should closely track the dark matter distribution, the galaxy distribution is generated by either turning dark matter particles from S_l into galaxies or removing them from the distribution according to a given criterion. This way, it is assured that the galaxy distribution follows the large-scale distribution of dark matter. Knowing the value of the galaxy number density field $G_l^\star$, the probability Q_i for a dark matter particle to be transformed into a galaxy can be expressed as

$$Q_i \propto \frac{n_G(\boldsymbol{r}_i)}{\rho_S(\boldsymbol{r}_i)}, \tag{9}$$

which can be normalized using the fact that a total of $N_{G_g^\star}$ should be created:

$$Q_i = N_{G_g^\star} \frac{n_G(\boldsymbol{r}_i)}{\rho_S(\boldsymbol{r}_i)} \left[\sum_{j=1} N_S \frac{n_G(\boldsymbol{r}_i)}{\rho_S(\boldsymbol{r}_i)} \right]^{-1}, \tag{10}$$

where i corresponds to the index of one of the N_S dark matter particles in S_l. For every dark matter particle, located at position $\boldsymbol{r}_i$, the local galaxy and dark matter densities $n_G(\boldsymbol{r}_i)$ and $\rho_S(\boldsymbol{r}_i)$ are linearly interpolated from the values at surrounding grid nodes, and every particle is changed into a galaxy or rejected with a probability Q_i. The attribution of a spectrum to every generated galaxy follows the same process: if a galaxy is created at a location where the galaxy and dark matter density are n_G and ρ_S,

then one of the N_s spectra $F_i(\lambda)$ in G_s is attributed to it, following the probability distribution $P[F_i(\lambda)|n_G, \rho_S]$. In the unlikely case in which $Q_i > 1$ (which means that the local galaxy number density is superior to the dark matter number density), a number of galaxies equal to the integer part of Q_i are generated and randomly located on a sphere of radius d centered on $\boldsymbol{r}_i$, d being a random number with a probability distribution $W(d)$ (i.e., identical to the kernel used for the density sampling).

2.3. *The Velocity Field*

Generating a galaxy distribution with the correct velocity field is not an easy task. In order to assign a velocity to every galaxy, MoLUSC relies on the underlying dark matter velocity field, assuming that the velocity distributions for galaxies and dark matter should match. In order to test this hypothesis, we used a 100 h^{-1} Mpc side length, 1024^3 dark matter particles, and a 1024^3 root grid size AMR hydrodynamical simulation with four extra levels of refinement at redshift $z = 2.5$. Galaxy formation is modeled using star particles as well as dark matter particles in such a simulation, so this allows a fair comparison of their respective velocity fields, even though in our case we have to deal with the extra complication that the host halo of a galaxy will not be resolved in general in the low-resolution dark matter simulation.

In order to check the correctness of the method, we identify star particles belonging to the same galaxies and use them to compute the galaxies velocities and center of mass. This allows us to compare the galaxies velocities to the velocity field of the underlying dark matter distribution. For every galaxy, the corresponding dark matter velocity is the mean velocity of the N dark matter particles closest to the galaxy center. In the present case we used $N = 8000$ to mimic the poor resolution of the large-scale simulation used by MoLUSC, but the results appear not to be very sensitive to the exact value of N as long as this value is a fraction of the number of particles in a typical dark matter halo.

Figure 2 (*top*) shows the probability distribution function (PDF) of the angle formed by the galaxies velocities and their underlying dark matter field. As expected, the agreement is excellent. A vast majority of galaxies are in fact moving in a direction within 20° of that of the dark matter. It is nonetheless not obvious that their respective velocities are equal. The PDF of the ratio of the velocity norm for the galaxies V_g and associated dark matter particles $V_{\rm dm}$ is shown on Figure 2 (*bottom*). For most of the galaxies, V_g appears to be about 2 times smaller than $V_{\rm dm}$, the PDF forming quite a marked peak around this value. This phenomenon has already been observed in the literature (e.g., Carlberg et al. 1990) and can be explained by the velocity dispersion of the dark matter within halos, which is greater than that of the galaxies. Since the dark matter velocity is determined in the center of dark matter halos and is subject to dynamical friction, in contrast to collisionless dark matter, it is expected to be smaller, relative to the center of mass of the respective dark matter halo, than that of collisionless dark matter.

The prescription used for generating galaxies velocities in MoLUSC is thus as follows: when a galaxy is created from a dark matter particle, its velocity is chosen to be the same as that of the dark matter particle, with half its norm.

3. CHECKING THE GALAXY DISTRIBUTION

Using the previously described process, one can generate a large-scale galaxy distribution from a large-scale dark matter–only simulation by mimicking the properties of a smaller scale galaxy distribution obtained using sophisticated semianalytical models such as GalICS. To check the quality of the results, we created a galaxy distribution $G^\star$ from a $(100\ h^{-1}\ {\rm Mpc})^3$, 256^3 particle simulation S from the properties of the galaxy distribution G computed by running GalICS on the *same* N-body simulation.[6] Figure 3*a* compares a 40 h^{-1} Mpc thick slice of G with the matching 40 h^{-1} Mpc thick slice of $G^\star$. In the two cases the total number of galaxies is very similar: 30,765 with GalICS as opposed to 30,941 for MoLUSC. Moreover, the scheme makes the distribution look very similar on scales larger than the smoothing length of 1 h^{-1} Mpc (halos, filaments, and voids are located at the same place). Nonetheless, galaxy clusters generated by

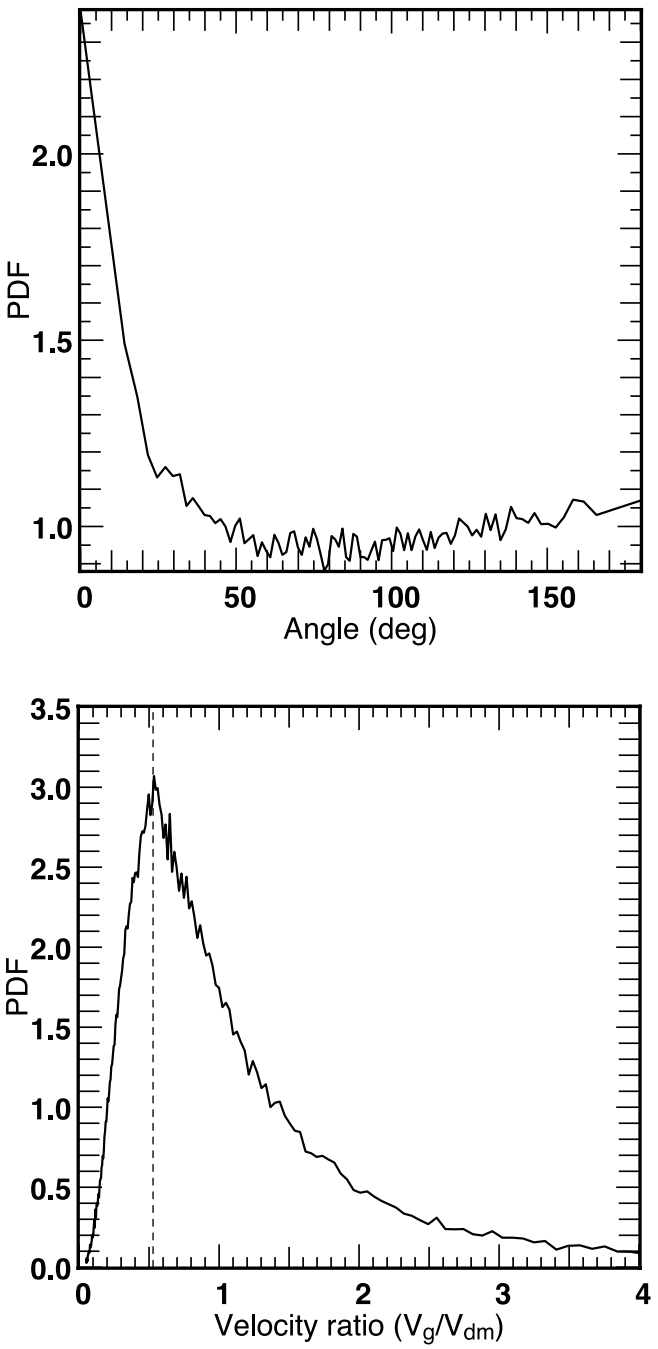

FIG. 2.—PDF of the velocity norm ratio and angle for the galaxies V_g and the corresponding DM field $V_{\rm dm}$ in a $(50\ h^{-1}\ {\rm Mpc})^3$ 1024^3 DM particle and 1024^3 root grid AMR hydrodynamical simulation with four refinement levels at redshift $z = 2.5$. The velocity of the DM is computed using, respectively, the 8000 particles closest to the centers of the 119,214 galaxies identified in the simulation at this redshift.

[6] Available from http://galics.iap.fr.

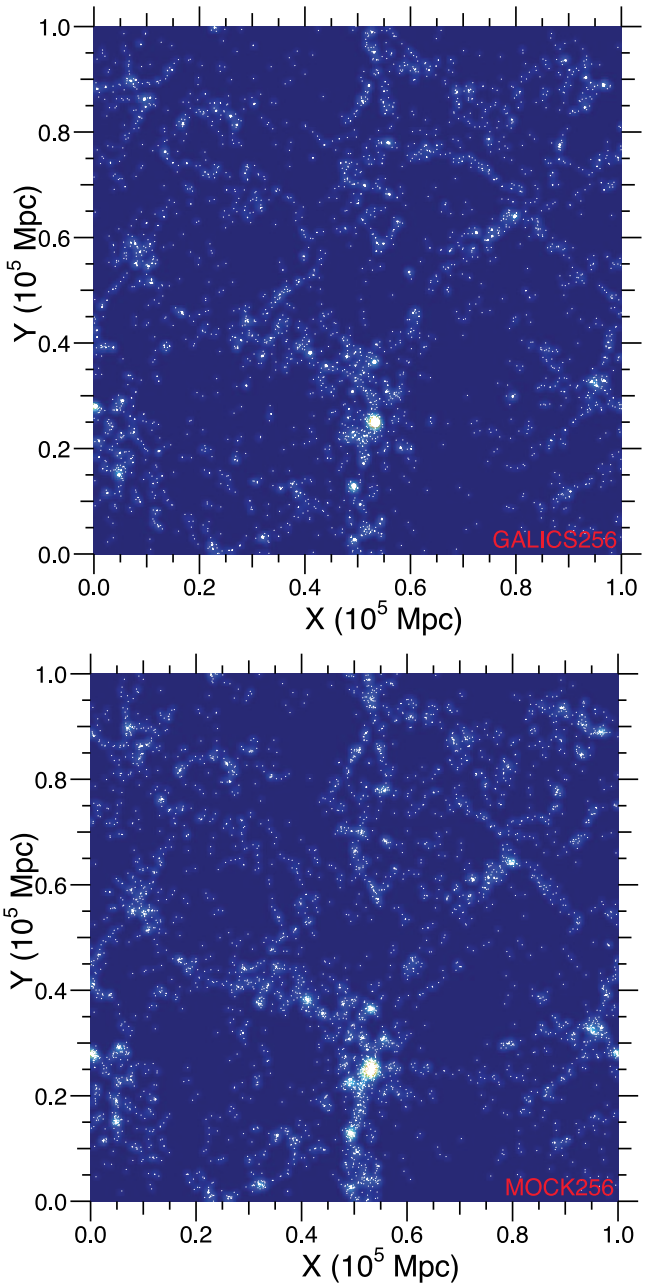

FIG. 3.—Comparison of the two galaxy distributions obtained with GalICS and MoLUSC from the same DM distribution. *Top*: GALICS256, 40 h^{-1} thick slice, 30,765 total galaxies; *bottom*: MOCK256, 40 h^{-1} thick slice, 30,941 total galaxies. The MoLUSC version was computed with a 1 h^{-1} Mpc smoothing length from the DM simulation and its GalICS posttreatment.

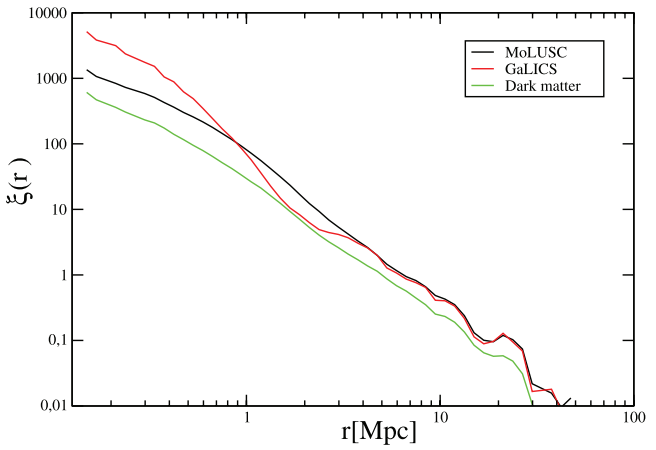

FIG. 4.—Two-point correlation function $\xi(r)$ of the DM distribution (*green curve*) and of the galaxies distributions generated by GalICS (*red curve*) and MoLUSC (*black curve*). The smoothing length used for MoLUSC is $L = 1\ h^{-1}$ Mpc. Part of the particles in the DM simulation have been randomly removed so that the number of points is the same in all cases.

MoLUSC clearly appear more spread out than the ones generated with GalICS. This is due to the Gaussian smoothing that is part of MoLUSC process. The other difference in cluster appearance is in regard to shape, this time constituting a strength of MoLUSC. GalICS uses a spherical collapse approximation for halos such that galaxies are distributed completely spherically within a halo, which is not necessarily the case, however, with MoLUSC, which does not require any halo identification. This results in galaxy clusters that really follow the underlying matter distribution, thus presenting nonsymmetric geometries.

Examination of the two-point correlation functions $\xi(r)$ in Figure 4 confirms the preceding remarks. The main goal of MoLUSC is to reproduce the GalICS galaxy distribution, and it achieves this task remarkably well. A comparison of the correlation functions for G (*red curve*) and the one for $G^\star$ (*black curve*) indicates that they are similar on scales larger than 2 h^{-1} Mpc, while some differences exist on smaller scales. It is clear that below scales of order the smoothing length (1 h^{-1} Mpc here), correlations are weaker with MoLUSC, as they tend to follow the dark matter correlation function (*green curve*) because halos hosting galaxies are no longer resolved in the simulation. Between scales of order the smoothing length and 2 h^{-1} Mpc, the opposite situation appears: whereas the amplitude of the correlation function is too low for G due to the spherical halo approximation made in GalICS, this is not the case in $G^\star$. These results confirm that MoLUSC is a particularly appropriate tool for generating large-scale distributions of galaxies out of a large-scale dark matter–only simulation.

Another interesting feature of the method concerns the excellent preservation of the galaxy distribution topology. This point can be tested using Minkowski functionals, a set of $d + 1$ morphological descriptors that completely describe the morphological properties of a field of d spatial dimensions (e.g., Schmalzing & Buchert, 1997). In dimension 3, the four Minkowski functionals are $\{V_1(\nu),\ V_2(\nu),\ V_3(\nu),\ V_4(\nu)\}$, the volume, area, mean curvature, and Euler characteristic of isodensity contours at normalized density $\nu = (\rho - \bar{\rho})/\sigma$, with $\bar{\rho}$ representing the mean density, and σ its variance. Figure 5 presents the computed Minkowski functionals of the density fields obtained from G and $G^\star$ after smoothing on a scale of 2.3 h^{-1} Mpc. The smoothing scale has been chosen to be of order of the mean intergalactic distance in order to keep the influence of the Poisson noise negligible while sampling small enough scales. The match of the measurements for G and $G^\star$ shows the ability of MoLUSC to produce a galaxy distribution with correct topology. As expected, the agreement slightly worsens when going from V_1 to V_4. As a matter of fact, the Euler characteristic V_4 in particular can be thought of as an alternate count of field extrema (see Colombi et al. 2000), making it very sensitive to small topology changes in the field configuration. Both measurements nonetheless appear to be in very good agreement for G and $G^\star$, which means that on scale over 2.3 h^{-1} Mpc, the topology of the galaxy distribution generated with MoLUSC is as correct as the one generated by semianalytical models such as GalICS. The method thus appears to preserve the filamentary structure very well, which is essential for any

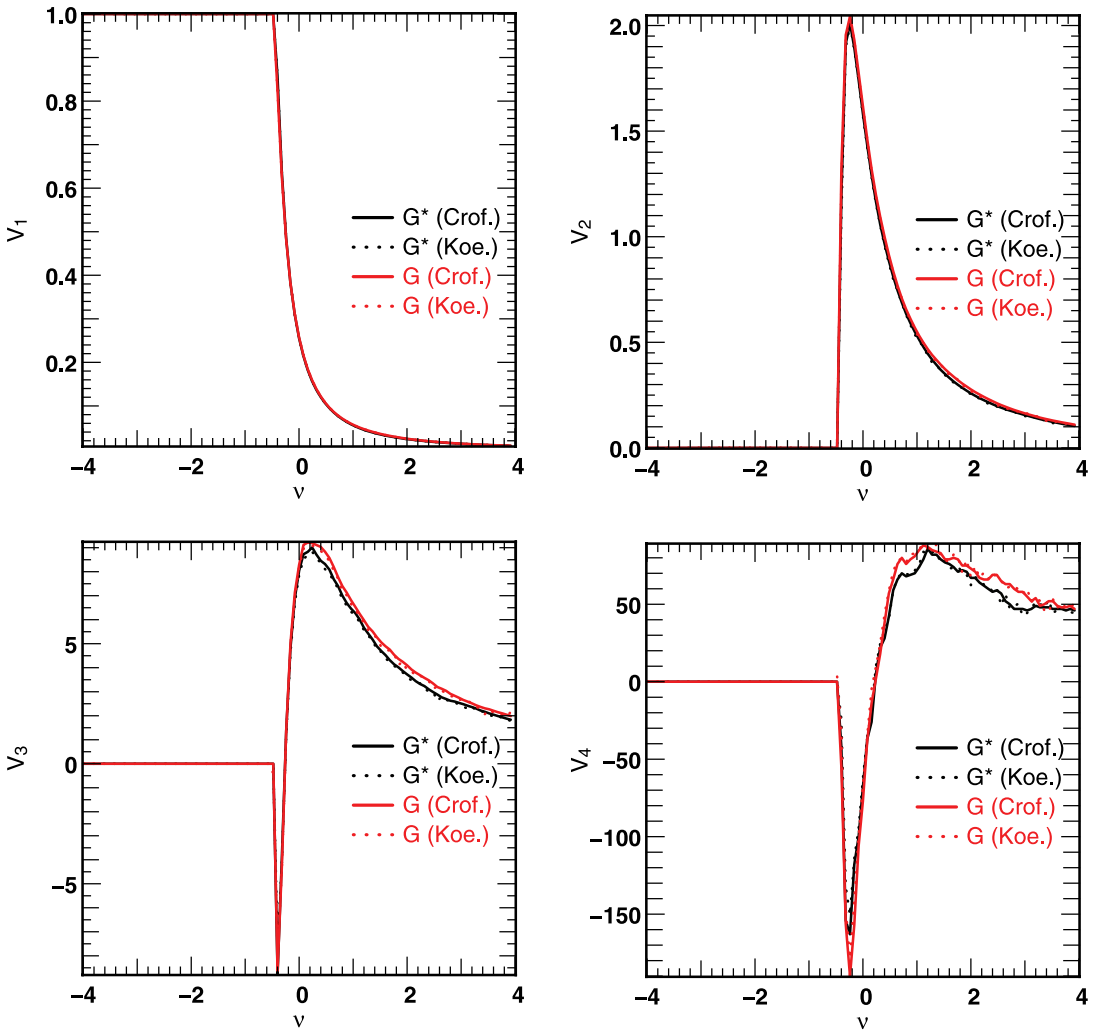

FIG. 5.—Minkowski functionals of the GalICS generated distribution G (*red*) and its MoLUSC equivalent G^* (*black*), computed for a smoothing length of 2.3 h^{-1} Mpc. The dotted lines are computed using the Koenderink method, while the full lines are computed with the Crofton method. The agreement between the two methods shows that the smoothing length is large enough to perform the measurements without being affected by Poisson noise; see, e.g., Schmalzing & Buchert (1997).

topological analysis of large-scale structures such as the ones observed in SDSS.

4. REAL SPACE SNAPSHOT TO REDSHIFT CATALOG CONVERSION

Once the simulation has been populated with galaxies, the primary difference remaining between a simulation and a survey like SDSS is that observing galaxies introduces selection biases that are not present in a simulation. So now that we have obtained a galaxy distribution, we are left with the task of producing a mock catalog emulating these biases. First of all, when galaxy properties are measured with a telescope, often the only measurements taken are of redshift and apparent magnitude in a given filter. All other properties, such as the absolute magnitude or distance, have to be computed from these values. This, of course, introduce errors and limitations, including primarily the following:

1. The spectrum of a galaxy is not constant with wavelength. So as it is redshifted because of the expansion of the universe, the apparent magnitude is not measured in the same wavelength range as the absolute magnitude. This problem can be corrected by using so-called K-corrections, the value of which depends on galaxy morphology, redshift, and filter characteristics.

2. The measured redshift is used to compute a distance on the assumption that it is only due to the expansion of the universe. The peculiar velocities are neglected in the process, giving rise to redshift distortions such as the so-called finger-of-god effect.

3. The geometry of a survey is constrained by practical matters, giving rise to complex geometry, as seen with SDSS on Figure 7.

Most aspects of the procedure used here are quite similar to the one described in detail by Blaizot et al. (2005). We therefore refer the reader to that paper for further details on the aspects common to the two methods.

4.1. *Mock Catalog Construction*

Making mock catalogs consists in extracting from the simulated data cubes a galaxy distribution similar to that which would be observed, taking into account all the observational biases. By randomly placing an observer within the $z = 0$ snapshot and defining a line-of-sight direction, it is possible to compute the *observational* properties of the galaxies belonging to the volume of the galaxy catalog to mimic, paying attention to the fact that the galaxies should be extracted from the snapshot with a redshift corresponding to the distance between the observer and the galaxy. Since the simulated snapshot is potentially smaller than

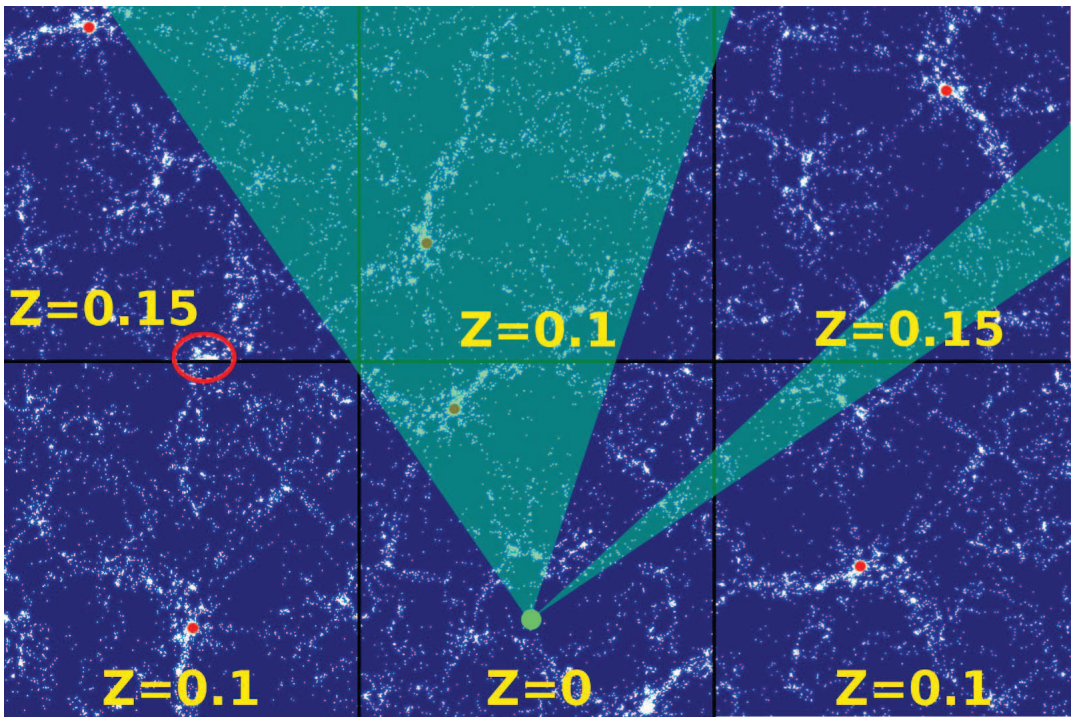

FIG. 6.—Illustration of the random tiling method and the construction of mock catalogs from the simulated cubes. The space is tiled with galaxy cubes at different redshifts that have been randomly transformed (rotation, translation). The red dots show the same cluster of galaxies. The red circle illustrates the problem of clusters lying close to the edge of a box. To create a mock catalog, only the galaxies included in the geometry of the initial catalog (*here in green*) are selected.

the mock survey to be generated, one has to use a method called random tiling, as illustrated by Figure 6. The main drawback comes from the fact that large clusters sitting on the edge of the simulated volume can be cut simply because the snapshot used to build the mock catalog changes due to the randomness of the tiling. This problem is solved using a friends-of-friends–type structure finder: instead of computing observational properties of the galaxies one by one, the computation is performed for every cluster, and depending on the cluster center position, all of its galaxies will either be present or absent in the mock catalogs. The influence of random tiling on the two-point correlation function has been more accurately quantified in Blaizot et al. (2005, § 3).

4.2. *Final Catalogs*

Galaxy catalogs often have complex geometries due to observational constraints. It is important to reproduce these geometries in order to be able to check the influence of edge effects. To do this, we simply build a mask from the real catalog and apply it to the mock catalog. Once the catalog is built, the only remaining step involves reproducing the main observational biases: redshift distortions and spectral redshifting. Redshift distortions are taken into account by using the peculiar velocities of the galaxies obtained from the simulation snapshots. For each galaxy at a given distance we add an additional red/blueshift term derived from its peculiar velocity to the redshift caused by the Hubble flow. The amplitude of this term is simply given by (see Hogg 1999):

$$\delta z = \sqrt{\frac{1+\boldsymbol{v}_p}{1-\boldsymbol{v}_p}} - 1, \tag{11}$$

where $\boldsymbol{v}_p$ is the norm of the peculiar velocity projection along the line of sight. Finally, absolute and apparent magnitudes of galaxies are computed using GalICS synthetic spectra as observed through a given filter $F(\lambda)$. This way, the ith galaxy with redshift $z_i + \delta z_i$ with a rest-frame spectrum $P_i(\lambda)$ has an observer frame spectrum

$$P_i^{\rm obs}(\lambda) = \frac{P_i[\lambda(1+z_i)]}{1+z_i}, \tag{12}$$

and its absolute and apparent magnitudes $M_i(F)$ and $m_i(F)$ are

$$M_i = \int_0^\infty P_i(\lambda) f(\lambda)\, d\lambda \tag{13}$$

and

$$m_i = \int_0^\infty P_i^{\rm obs}(\lambda) f[\lambda(1+z_i)]\, d\lambda. \tag{14}$$

Figure 7 shows the result obtained for a mock SDSS DR4 catalog created with MoLUSC, with different observational biases taken into account.

5. CONCLUSIONS

In this paper we have presented MoLUSC, a tool for extracting mock galaxy surveys out of large-scale pure dark matter numerical simulations. Our method consists in numerically computing how the galaxy distribution maps out the underlying dark matter distribution at a given redshift, using a high-resolution small-scale simulation posttreated with a semianalytic model. We then propose an algorithm that enables us to reproduce the galaxy distribution characteristics from a snapshot of any size and resolution while preserving the statistical properties of the high-resolution galaxy distribution. Finally, galaxy properties are

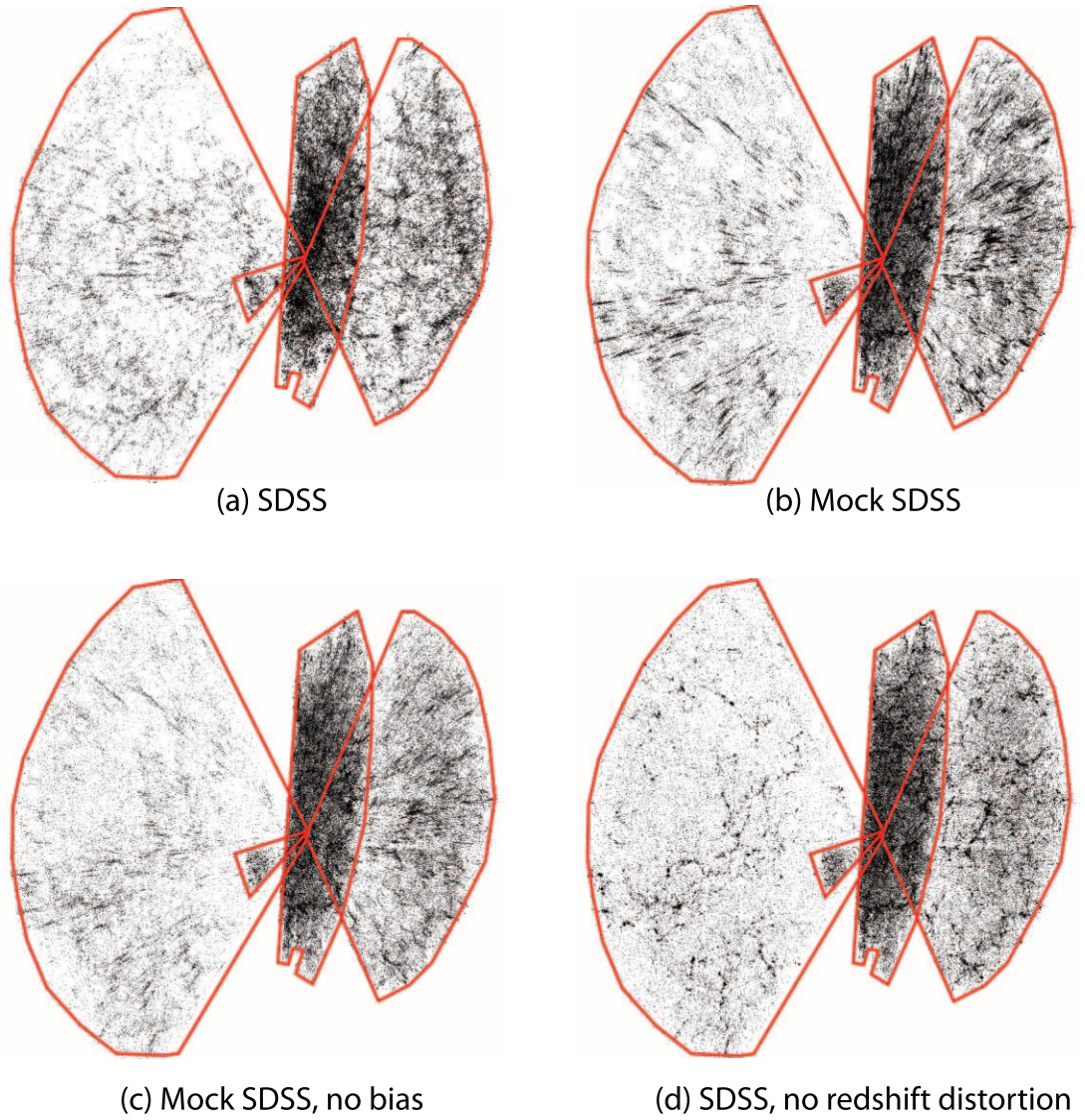

FIG. 7.—Comparison between the observed SDSS DR4 (*a*) survey (limited to 350 Mpc) and three different virtual SDSS surveys (*b*, *c*, and *d*) built from the large-scale low-resolution DM simulation. The virtual catalog MOCK (*b*) is made taking into account all the observed selection effects and observational artifacts, such as the fingers of god. The MOCK-NB (*c*) neglects the bias between galaxies and DM. This results in a density field with less contrast. In MOCK-NBNF (*d*) neither the bias nor the redshift distortions were accounted for. As a result of this absence of observational artifacts, the galaxy clusters appear almost spherical, rather than elongated along the line of sight.

computed as measured by an observer, i.e., bias effects such as fingers of god, *K*-corrections on magnitudes, and catalog geometry are numerically computed to allow a fair comparison to the observed data. We stress that this step is crucial in order to study the formation and evolution of the large-scale structures in the universe.

Although in this paper we have only illustrated how MoLUSC can being be used to generate mock catalogs of the SDSS, work is currently in progress to apply it to the deep fields of the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) and to take into account a variety of cosmologies. Once again, the idea is to provide observers with mock catalogs matching the survey selection effects, such as survey geometry, magnitude depth, and galaxy magnitudes, which are computed in the corresponding survey set of filters. Due to the large number of comparisons between real and mock catalogs that are made possible using MoLUSC, we chose to defer these matters to future papers.

Last but not least, we emphasize that the main advantage of this tool over similar existing methods is that it provides a realistic galaxy distribution out of any large dark matter–only simulation without using powerful computer hardware and thus makes possible faster astrophysical analysis of the present large DM simulations.

We would like to thank Thomas Buchert and Jens Schmalzing for providing us with their Minkowski functionals measurement tool beyond and the anonymous referee for his constructive suggestions. This work was supported by the Space Interferometer Mission Dynamics of Nearby Galaxies (SIMDOG) Key Project, by a European Marie Curie predoctoral grant, and by the French Horizon Project (http://www.projet-horizon.fr). G. B. acknowledges support from NSF grants AST 05-07161, AST 05-47823, and AST 06-06959.

REFERENCES

Blaizot, J., Wadadekar, Y., Guiderdoni, B., Colombi, S. T., Bertin, E., Bouchet, F., Devriendt, J. E. G., & Hatton, S. 2005, MNRAS, 360, 159
Carlberg, R. G., Couchman, H. M. P., & Thomas, P. A. 1990, ApJ, 352, L29
Cole, S., Hatton, S., Weinberg, D. H., & Frenk, C. S. 1998, MNRAS, 300, 945
Colombi, S., Pogosyan, D., & Souradeep, T. 2000, Phys. Rev. Lett., 85, 5515
Hamana, T., Colombi, S. T., Thion, A., Devriendt, J. E. G. T., Mellier, Y., & Bernardeau, F. 2002, MNRAS, 330, 365
Hatton, S., Devriendt, J. E. G., Ninin, S., Bouchet, F. R., Guiderdoni, B., & Vibert, D. 2003, MNRAS, 343, 75
Hogg, D. W. 1999, preprint (astro-ph/9905116)
Jing, Y. P., Mo, H. J., & Boerner, G. 1998, ApJ, 494, 1
Peacock, J. A., & Smith, R. E. 2000, MNRAS, 318, 1144
Schmalzing, J., & Buchert, T. 1997, ApJ, 482, L1
Springel, V. 2005, MNRAS, 364, 1105
Yan, R., Madgwick, D. S., & White, M. 2003, ApJ, 598, 848
Zhao, D., Jing, Y. P., Börner, G. 2002, ApJ, 581, 876