

## Creating synthetic universes in a computer

Carlton M Baugh

*Phil. Trans. R. Soc. A* 2008 **366**, doi: 10.1098/rsta.2008.0192, published 13 December 2008

---

### References

**This article cites 36 articles, 16 of which can be accessed free**  
<http://rsta.royalsocietypublishing.org/content/366/1884/4381.full.html#ref-list-1>

### Subject collections

Articles on similar topics can be found in the following collections

[cosmology](#) (26 articles)

[observational astronomy](#) (21 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

# Creating synthetic universes in a computer

BY CARLTON M. BAUGH\*

*Department of Physics, Institute for Computational Cosmology,  
University of Durham, South Road, Durham DH1 3LE, UK*

Cosmologists regularly generate synthetic universes of galaxies using computer simulations. Such catalogues have an essential role to play in the analysis and exploitation of current and forthcoming galaxy surveys. I review the different ways in which synthetic or ‘mock’ catalogues are produced and discuss the different physical processes that the models attempt to follow, explaining why it is important to be as realistic as possible when trying to forge the Universe.

**Keywords:** cosmology; computer simulation; galaxy formation; dark matter; dark energy

## 1. Introduction

Astronomers are traditionally a patient bunch, passively waiting for photons to arrive from the depths of the Universe. However, a new breed of impatient experimental cosmologists has evolved over the past 20 years. A combination of increasing computing power and a developing theory for the formation of structure in the Universe means that it is now feasible to run computer simulations that can recreate the appearance of the Universe. The picture on large scales is taking shape nicely. Small-amplitude primordial density fluctuations in the early Universe, perhaps seeded during a period of extremely rapid expansion called inflation, grow through gravitational instability. On the scale of galaxies, the physics becomes more complicated and controversial. Many calculations, numerical experiments, need to be carried out to find the right ingredients to make galaxies in a hierarchical universe.

Our view of the Universe is on the verge of being opened up beyond all recognition, due to technological advances that have greatly sped up the rate at which we can map galaxies. Astronomers use the resulting maps or surveys of galaxy positions to constrain the basic parameters that control the evolution and fate of the Universe and to try to extend our knowledge of how galaxies formed and evolve. Synthetic universes have a central role to play in helping to interpret the implications of such maps. In this paper, I explain the rationale behind producing mock galaxy catalogues (§2) before reviewing the main techniques used (§3) and giving an example of the effects that are captured by numerical simulations (§4).

\*c.m.baugh@durham.ac.uk

One contribution of 12 to a Triennial Issue ‘Astronomy’.

## 2. Why build a synthetic universe?

The Universe contains a wealth of structure on different scales, ranging from planets to stars to galaxies and clusters of galaxies. The range of length and time scales and the variety and complexity of processes responsible for shaping these structures make simulating the Universe a daunting task. Before explaining why it is useful for cosmologists to attempt to build a synthetic universe of galaxies, let me first set out the steps that are involved in producing such a catalogue.

In essence, the production of a mock catalogue is a two-stage process. The first stage is to generate a volume filled with galaxies, without applying any pre-selection based on their properties. Various approaches to carrying out this first stage have been developed in the literature and these are reviewed in §3. Broadly speaking, the methods used can be classed as either ‘physical’, in which a calculation is made that attempts to follow the physics of galaxy formation, or ‘statistical’, whereby a scheme is devised to generate a set of points to be treated as galaxies. In general, the physical approach may result only in an approximation to the observed Universe, which, depending on one’s purposes, may require some adjustment to better match observations. The statistical approach, by construction, should yield an exact copy of an observation, but will provide little further insight. The second stage in the construction of a synthetic universe is to apply selection criteria to the idealized galaxy catalogue produced in stage 1 in order to mimic a particular observational survey. These criteria can include (i) applying an angular mask to extract the region of sky observed and (ii) applying a radial selection function, to take into account the flux limit of the survey.

The angular mask records which parts of the sky have been mapped. The overall extent of the angular mask in right ascension and declination is set by the amount of telescope time available and the observing strategy, e.g. the survey footprint may have been chosen to avoid high-extinction regions of the Milky Way. Patches within the survey boundary may also be masked out if they contain bright, overexposed stars or could not be mapped due to the geometry and sampling of the field of view of the telescope. The radial selection function gives the probability of including a galaxy in the survey as a function of distance from the observer. The radial selection applies the survey flux limit to the idealized volume of galaxies produced in stage 1. Only progressively brighter galaxies will be detected at greater distances in the survey. Some surveys may be defined by more than a simple flux limit, and may also require selection in colour or angular diameter. The radial selection itself may vary with the angular position on the sky due to corrections for Galactic dust extinction or due to the survey strategy (i.e. how many times a particular patch on the sky has been visited). Another consideration that has an impact on the radial selection function is the completeness of the survey. This could describe, for example, the fraction of target galaxies for which a spectroscopic redshift was successfully measured. One way to deal with incompleteness is to define a weight for each galaxy. In a (hypothetical) fully complete survey, the weight of each galaxy will be unity. If galaxies are missing, their weight may be redistributed to their neighbours in some way. By applying the selection criteria that describe the survey, we degrade the idealized volume of galaxies to resemble as closely as possible our observed galaxy survey.

There are many levels on which mocks are useful in the analysis and exploitation of galaxy surveys, and in the testing and development of theoretical models. In an idealized volume of galaxies, statistical analyses are straightforward. We ‘see’ all galaxies equally well and know the volume sampled by different galaxies accurately. It is therefore trivial to count galaxies to form luminosity functions (i.e. the number of galaxies per unit volume with a given luminosity) and to measure their clustering (i.e. to quantify their spatial distribution) by comparing the local density of galaxies with a well-defined mean density. In a real survey, the complex angular and radial selection make these tasks non-trivial. Faint galaxies are visible over a smaller volume than bright galaxies, and this has to be taken into account when estimating the luminosity function of galaxies. Often, we do not know *a priori* which is the best or optimal way to analyse a particular survey. This is where the real strength of mock catalogues becomes apparent. In the case of the mock, we know the true underlying result beforehand, e.g. the values of the cosmological parameters adopted, and can use this privileged knowledge to test how well different estimators perform. We can use the mocks to devise new estimators and to tune their performance. If a sufficient number of mocks are generated, which sample the population of possible galaxy surveys, e.g. reproducing the sampling variance due to large-scale density fluctuations, then the scatter between measurements made in each mock can be used to estimate the random and systematic errors on a measurement. Finally, mock catalogues provide an ideal, and in some cases the only, platform on which to confront theoretical predictions with observational data. Quite often, the predictions of hierarchical galaxy formation models are impossible to generate analytically and require the machinery of a simulation code. The comparison with data is usually strongly affected by the selection criteria that define the observational dataset.

### 3. A survey of methods

In this section, we review the various techniques commonly used in the literature to produce mock galaxy catalogues. We have split these into two classes: those used to generate density fields, and those that take these density fields and populate them with galaxies. In most cases, a two-step process is followed, whereby, in the first step, a density field is generated and then, in the second step, this field is used to generate a mock galaxy catalogue. However, in some analyses, the second step is not used, which implies that galaxies are assumed to be a faithful tracer of the underlying density field.

#### (a) *Generating density fields*

The simplest and quickest way to produce a density field is to make a realization of a Gaussian field (or some transformation of a Gaussian, such as a lognormal distribution). Such realizations typically have limited spatial resolution due to the Fourier transform used to set up the field. The typical application is to study the distribution of errors on clustering measurements on large scales. The low computational cost of such mocks means that hundreds or even thousands of examples can be generated, allowing the full covariance matrix to be estimated for a measurement (e.g. Percival *et al.* 2001).

Next in order of increasing computational cost are methods that generate more accurate realizations of density fields using perturbation theory techniques: the publicly available PINpointing Orbit-Crossing Collapsed HIERarchical Objects (PINOCCHIO) code introduced by Monaco *et al.* (2002a) and the PTHALOS code of Scoccimarro & Sheth (2002). These codes have been tested against full numerical simulations (see below; e.g. Monaco *et al.* 2002b). The advantage of these codes over full simulations is their speed, which allows large numbers of mocks to be generated. The memory requirements of PINOCCHIO are similar to those of a full numerical simulation.

The most accurate and computationally expensive means to generate a density field is to model the hierarchical growth of fluctuations due to gravitational instability using an  $N$ -body simulation (e.g. Springel *et al.* 2005). The scale of the calculation depends upon the intended use of the simulation. Springel *et al.* used 10 billion particles to trace the dark matter within a computational volume of side  $500h^{-1}$  Mpc. One of the primary uses of this simulation is to provide accurate merger histories of dark matter haloes for use in galaxy formation models (see §3b). Angulo *et al.* (2008) used a much larger volume (albeit with poorer mass resolution than that achieved by Springel *et al.*) to model the growth of fluctuations accurately on the scale of baryonic acoustic oscillations (BAOs). These authors also ran an ensemble of lower-resolution calculations, with the same box size, to estimate the scatter on the measurement of the power spectrum on large scales. Efstathiou & Moody (2001) proposed generating large numbers of low-resolution simulations to estimate the covariance matrix of large-scale structure measurements.

### (b) Putting in galaxies

Methods for populating density fields with galaxies can be split into empirical and physical techniques.

Early work associated galaxies with peaks in a Gaussian density field (Kaiser 1984; White *et al.* 1987). By selecting peaks in excess of a particular height or rarity, the clustering amplitude of the ‘galaxies’ would be enhanced relative to that of the underlying mass. Cole *et al.* (1998) investigated empirical models based on this idea, using either the initial or final density field as an input. The probability of selecting a dark matter particle to be a galaxy was assumed to be a function of the local smoothed density of matter. The parameters in the functional form were tuned by comparing the resulting galaxy clustering with observational estimates. Cole *et al.* found that biasing schemes based on the final matter density with two parameters were able to reproduce a power-law form for the two-point galaxy correlation function.

The next step up in sophistication is to use information about the distribution of dark matter haloes rather than the smoothed density field to populate a volume with galaxies.

The galaxy halo occupation distribution (HOD) records the mean number of galaxies passing a particular observational selection as a function of the mass of the host dark matter halo (Benson *et al.* 2000; Peacock & Smith 2000; Cooray & Sheth 2002). An extensive analytical framework exists to compute the two-point and higher-order correlation functions from the HOD (Scoccimarro *et al.* 2001). Given a density field in which dark matter haloes can be identified, it is

straightforward to generate a realization of particular HOD, using dark matter particles or substructures from the dark haloes to place galaxies. An extension of this technique to connect the properties of galaxies to their host halo mass was recently introduced by [Conroy & Wechsler \(submitted\)](#).

In the conditional luminosity function approach introduced by [Yang \*et al.\* \(2003\)](#), the luminosity function of galaxies within a halo is allowed to have a dependence on halo mass. The luminosity function is typically assumed to have a Schechter function form, with mass-dependent parameters. This approach has been used to generate mock catalogues with the clustering and group properties observed in local surveys ([Yang \*et al.\* 2004](#)).

Physical approaches involve *ab initio* calculations of the fate of baryons in a cold dark matter universe. Full numerical simulations are extremely expensive to run and can only cover relatively small volumes. Simulations that cover large enough volumes to allow robust predictions for galaxy clustering find it challenging to achieve the resolution required to model the physics of the baryons (see, however, [Pearce \*et al.\* \(1999\)](#) for an example of how this problem can be turned to the simulator's advantage). Much work is still needed to improve the predictions of the galaxy mass function in such approaches ([Berlind \*et al.\* 2003](#)). An alternative approach, called semi-analytical modelling, has its roots in the seminal work of [White & Rees \(1978\)](#), who proposed that galaxies form inside dark matter haloes and that the efficiency of galaxy formation has to be modulated by some form of feedback process. By adopting certain assumptions and approximations, semi-analytical models are a fast, efficient and flexible means with which to generate predictions for the whole galaxy population (for a review of this approach see [Baugh \(2006\)](#)). Semi-analytical models predict the broad-band fluxes, emission-line luminosities and many basic galaxy properties, such as galaxy sizes. Therefore, the same selection criteria as used to define observational samples can be applied to the semi-analytical output, thereby allowing predictions to be made for the abundance and clustering strength of galaxies; these quantities are inputs in the empirical methods outlined above. The current generation of models incorporated into the Millennium Simulation of [Springel \*et al.\* \(2005\)](#) is able to match a wide range of observational data, both locally and at high redshift ([Bower \*et al.\* 2006](#); [Croton \*et al.\* 2006](#); [De Lucia \*et al.\* 2006](#)). By incorporating the semi-analytical model into an  $N$ -body simulation, information about the spatial distribution of galaxies is added. [Cole \*et al.\* \(2005\)](#) used a semi-analytical model incorporated into a gigaparsec box simulation to model scale-dependent effects on the large-scale galaxy power spectrum, due to galaxy bias and nonlinearities (see also [Sánchez & Cole 2008](#)).

#### 4. Do we really need to go this far?

In terms of accuracy, the preferred combination of techniques from §3 would be to use an  $N$ -body simulation to produce a realization of the matter distribution and then to combine this with the most accurate calculation of the galaxy formation process in order to populate the whole volume with galaxies. Currently, this means using a semi-analytical galaxy formation code. This is also one of the more expensive combinations from the point of view of the effort

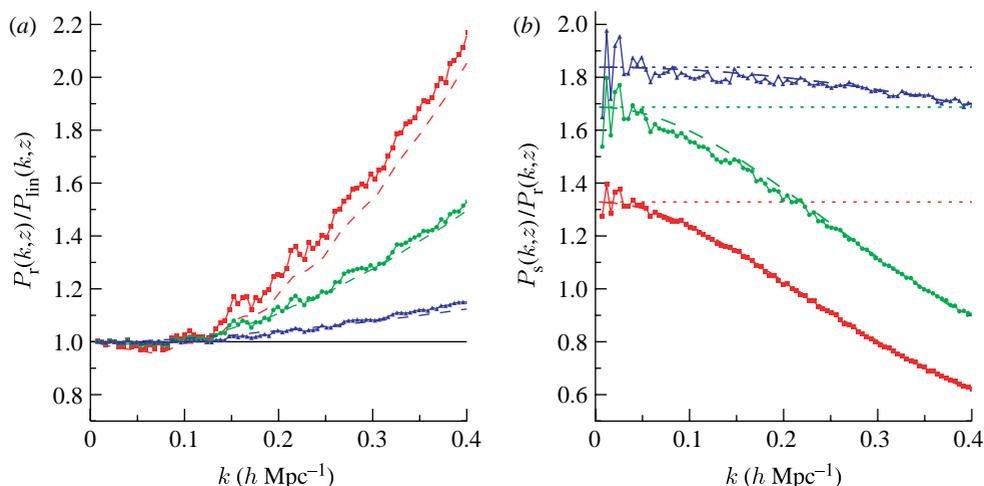


Figure 1. (a) The nonlinear growth of the matter power spectrum in real space. The power spectrum measured at the output redshifts (red,  $z=0$ ; green,  $z=1$ ; blue,  $z=3$ ) has been divided by the power spectrum in the initial conditions, scaled by the linear growth factor. Any deviation from a ratio of unity is a deviation from linear perturbation theory. The dashed lines show empirical predictions of the nonlinear power spectrum taken from [Smith \*et al.\* \(2003\)](#). (b) The ratio of the matter power spectrum measured in redshift space to that measured in real space (red,  $z=0$ ; green,  $z=1$ ; blue,  $z=3$ ). The horizontal lines show a prediction for the ratio from linear perturbation theory. The damping from this value can be described by a simple model (see [Angulo \*et al.\* \(2008\)](#)). Adapted with permission from [Angulo \*et al.\* \(2008\)](#).

required to set up the machinery and the computation time required. In this section we justify this choice by using an illustration that involves measuring the large-scale structure of the Universe.

The application considered is the modelling of the appearance of BAOs in future galaxy surveys. The BAOs are an imprint of sound waves that perturbed the matter distribution in the early Universe, prior to matter–radiation decoupling. The original intention was to use the BAO as a standard ruler, with the argument being that we know the BAO scale precisely from measurements of the power spectrum of temperature anisotropies in the cosmic microwave background. The apparent scale of the BAO standard ruler as measured in galaxy surveys depends upon the geometry and expansion history of the Universe, and is therefore sensitive to the nature of the dark energy, which is thought to be responsible for the accelerating expansion of the Universe. Following proof-of-concept work (e.g. [Blake & Glazebrook 2003](#); [Seo & Eisenstein 2003](#)), it is now apparent that the BAOs are not quite a standard ruler and, furthermore, they can be altered by a range of effects ([Seo & Eisenstein 2005, 2007](#); [Smith \*et al.\* 2007, 2008](#); [Angulo \*et al.\* 2008](#); [Crocce & Scoccimarro 2008](#); [Sánchez \*et al.\* in press](#); [Seo \*et al.\* in press](#)). Careful modelling is required to produce accurate predictions for the form of the BAOs and to help devise estimators to exploit these features in order to extract unbiased constraints on the dark energy.

[Angulo \*et al.\* \(2008\)](#) combined a gigaparsec side  $N$ -body simulation with a semi-analytical model of galaxy formation to show, step by step, how the power spectrum of galaxy clustering differs from the predictions of linear perturbation

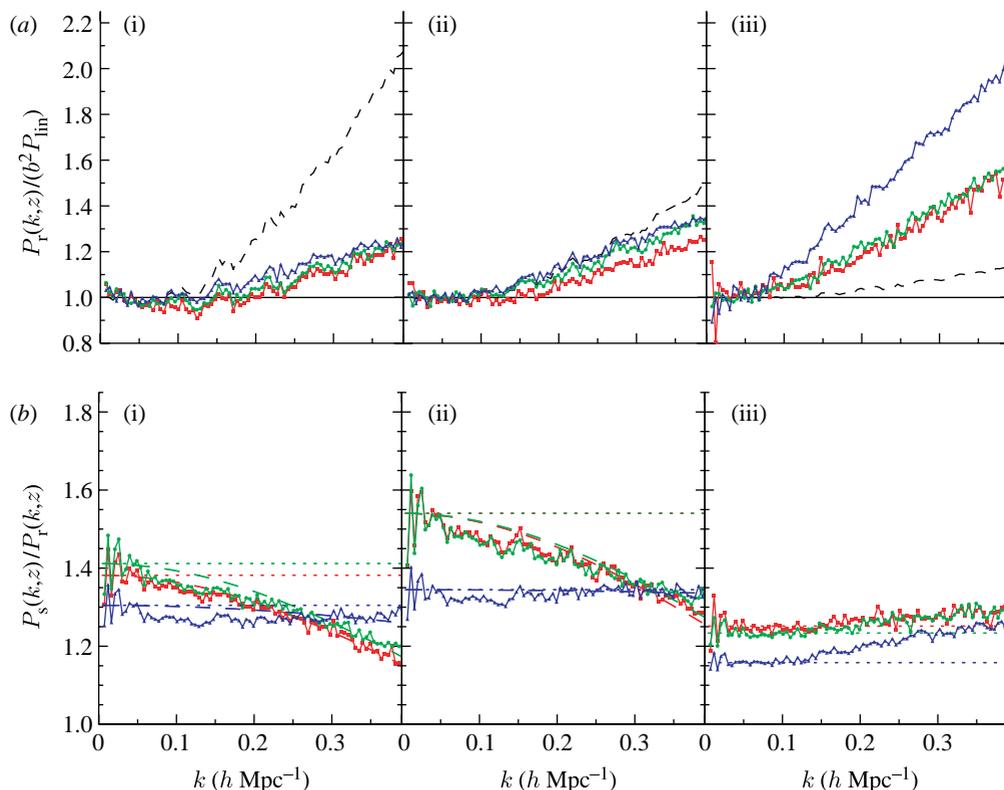


Figure 2. (a) The power spectrum of dark matter haloes in real space, divided by the linear perturbation theory prediction, after taking into account the asymptotic bias factor  $b$ , at redshifts (i)  $z=0$ , (ii)  $z=1$  and (iii)  $z=3$ . The mass range and bias for each halo sample are as follows. (i) blue,  $[4.2, \infty] \times 10^{12} M_{\odot}/h$ ,  $b=1.06$ ; green,  $[0.9, 1.7] \times 10^{12} M_{\odot}/h$ ,  $b=0.81$ ; red,  $[5.4, 7.0] \times 10^{11} M_{\odot}/h$ ,  $b=0.87$ ; (ii) blue,  $[3.6, \infty] \times 10^{12} M_{\odot}/h$ ,  $b=1.81$ ; green,  $[0.9, 1.6] \times 10^{12} M_{\odot}/h$ ,  $b=1.22$ ; red,  $[5.4, 7.0] \times 10^{11} M_{\odot}/h$ ,  $b=1.22$ ; (iii) blue,  $[2.0, \infty] \times 10^{12} M_{\odot}/h$ ,  $b=4.37$ ; green,  $[0.8, 1.1] \times 10^{12} M_{\odot}/h$ ,  $b=3.04$ ; red,  $[5.4, 6.5] \times 10^{11} M_{\odot}/h$ ,  $b=2.84$ . If the bias was *independent* of the scale, the solid lines would lie on top of the dashed line, which shows how the dark matter power spectrum deviates from linear perturbation theory. (b) The ratio of the redshift-space to the real-space power spectrum measured for dark matter haloes. The redshifts and mass ranges are as in panels (a)(i)–(iii). The horizontal dotted lines show the ratio predicted by linear theory and the dashed lines show an empirical fit to the ratio. Adapted with permission from Angulo *et al.* (2008).

theory, the original model used to describe the BAOs, even though the BAO imprint is seen on scales in excess of  $100h^{-1} \text{ Mpc}$ , which are normally assumed to be safely in the linear regime.

The first effect that changes the form of the BAOs is the nonlinear evolution of fluctuations in the dark matter (figure 1). This phenomenon distorts the shape of the power spectrum away from the linear perturbation theory prediction. There is a characteristic dip in power on large scales (wavenumbers  $k=2\pi/\lambda < 0.1h \text{ Mpc}^{-1}$ ) and an enhancement in power on small scales. Intriguingly, the scale at which the largest dip in power occurs is close to the sound horizon scale and to a low-order harmonic of the fundamental mode in the box size ( $L=500h^{-1} \text{ Mpc}$ ) most commonly used in the literature to model the BAO.

Somewhat fortuitously, the enhancement of power due to nonlinear evolution is almost completely cancelled out by the damping of dark matter fluctuations when we mimic how they would be viewed in a survey (figure 1). However, this should be ignored as a fortunate coincidence of little relevance to the problem of modelling the galaxy power spectrum. Galaxies live inside dark matter haloes, so the next step should be to compare the power spectrum of haloes with that of the underlying mass. The power spectrum of haloes shows a different deviation away from the linear theory power spectrum than we find for the mass, which means that the haloes show a scale-dependent bias, even on very large scales (figure 2). The sign of the bias (i.e. whether or not the haloes are more or less clustered than the mass) depends upon the mass of the halo under consideration. The situation is even more complicated when the contribution of peculiar velocities is introduced. In general, for the matter distribution, when clustering is measured in redshift space (i.e. including the contribution of the gravitationally induced peculiar motions to the redshift, on top of the Hubble expansion, which alters the inferred radial position), we expect to see an increase in the amplitude of the power spectrum on large scales, due to coherent bulk flows out of voids and into dense filaments, and a decrease in power on small scales, due to randomized motions inside virialized structures (Kaiser 1987; Peacock & Dodds 1994; figure 1). For haloes, one might expect the enhancement due to bulk flows to be smaller if the haloes are highly biased and for the small-scale damping to be absent, since, by definition, we do not have virialized structures made up of haloes. Figure 2 shows that this guess at the behaviour of the redshift space halo power spectrum is correct in some cases but very wrong in others. Part of the reason for this discrepancy is that haloes are biased, nonlinear tracers of the density and the predictions for redshift space distortions are based on linear theory.

Finally, in figure 3, we show how the power spectra measured for different galaxy samples differ from the linear perturbation theory predictions. Again, a scale-dependent bias is clearly evident, with a magnitude that depends upon the precise way in which the sample has been constructed. All of these distortions from the linear theory power spectrum need to be taken into account in order to extract cosmological information from the BAO.

## 5. The future

Forthcoming galaxy surveys will present a number of new challenges for the builders of mock catalogues. The most significant of these is the need to model the large look-back times covered by new deep photometric (e.g. The Dark Energy Survey, Pan-STARRS) and spectroscopic (e.g. the Baryon Oscillation Spectroscopic Survey (BOSS), the Physics of the Accelerating Universe (PAU) survey and ESA's Euclid mission concept) surveys. Whereas local surveys such as the Two-Degree Field Galaxy Redshift Survey (2dFGRS; with a median redshift of  $z \approx 0.1$ ) could be modelled using a single output from an  $N$ -body simulation, surveys such as Pan-STARRS (an expected median redshift of  $z \approx 0.5$ , with a tail extending beyond  $z \sim 1$ ) will require a full treatment of the evolution of structure along the past light cone of an observer. Approximate calculations using a series of snapshots from  $N$ -body simulations have started

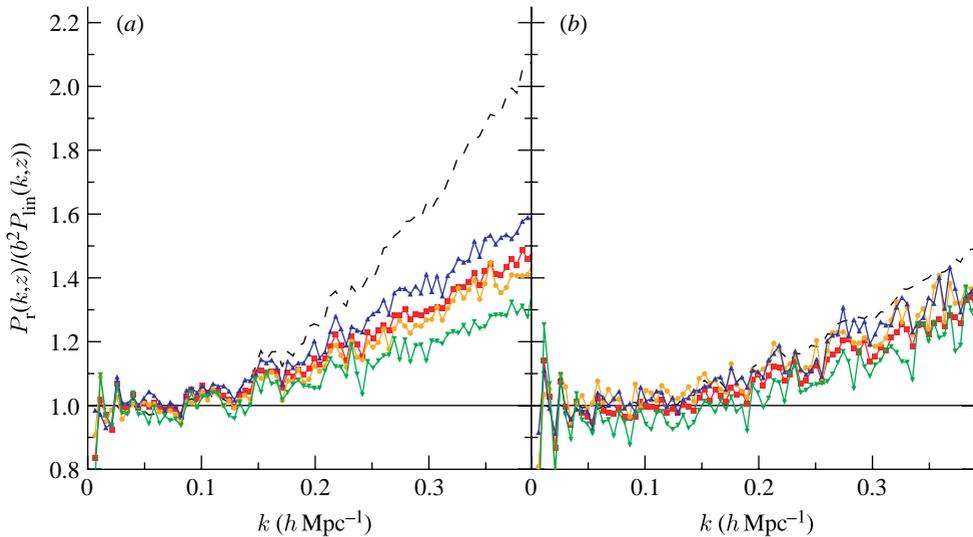


Figure 3. The power spectrum of different samples of galaxies, measured in real space, divided by a scaled version of the linear theory prediction, which takes into account the asymptotic bias  $b$ . (a) Redshift  $z=0$ : red squares, number density of  $5.0 \times 10^{-4} h^3 \text{Mpc}^{-3}$ ,  $b=1.18$ ; orange circles,  $2.5 \times 10^{-4} h^3 \text{Mpc}^{-3}$ ,  $b=1.33$ ; blue triangles, red (R-I),  $b=1.32$ ; green triangles, strong EW(OII),  $b=1.06$ . (b) Redshift  $z=1$ : as in panel (a) except for  $b$  values; red squares,  $b=1.34$ ; orange circles,  $b=1.31$ ; blue triangles,  $b=1.39$ ; green triangles,  $b=1.31$ . The properties and spatial distribution of galaxies are predicted using a semi-analytical model. Adapted with permission from [Angulo \*et al.\* \(2008\)](#).

to take this effect into account ([Blaizot \*et al.\* 2005](#); [Fosalba \*et al.\* submitted](#)). Such calculations are adequate for photometric surveys, in which only photometric redshift estimates are available and there are significant errors in the radial positions of galaxies. However, for a spectroscopic survey such as Euclid ([Cimatti \*et al.\* in press](#)), a more accurate scheme interpolating the positions of dark matter particles or haloes between snapshots will be required. The sheer size of the new surveys is daunting: Pan-STARRS is over 50 times the volume of the 2dFGRS and Euclid is 1000 times bigger. To model the large-scale structure of the Universe on such scales, ultra-large-volume simulations will be required (e.g. [Fosalba \*et al.\* submitted](#)). New techniques will be required to populate such runs with galaxies. A significant improvement will be required in the quality of the model predictions themselves. Many new surveys will have multi-wavelength elements and will push measurements of galaxy properties to new levels. Currently, if a semi-analytical model does not agree exactly with the available observational data, we can resort to applying some sort of scaling to the model predictions to force precise agreement and therefore a more faithful mock. Ultimately, of course, we desire a model that does reproduce all of the observations as closely as possible. The next generation of surveys will depend upon a computer-generated universe for their full exploitation, but, in return, will help to drive further refinements and improvements to the models.

C.M.B. is supported by a Royal Society University Research Fellowship.

## References

- Angulo, R., Baugh, C. M., Frenk, C. S. & Lacey, C. G. 2008 The detectability of baryonic acoustic oscillations in future galaxy surveys. *Mon. Not. R. Astron. Soc.* **383**, 755–776. (doi:10.1111/j.1365-2966.2007.12587.x)
- Baugh, C. M. 2006 A primer on hierarchical galaxy formation: the semi-analytical approach. *Rep. Prog. Phys.* **69**, 3101–3156. (doi:10.1088/0034-4885/69/12/R02)
- Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M. & Lacey, C. G. 2000 The nature of galaxy bias and clustering. *Mon. Not. R. Astron. Soc.* **311**, 793–808. (doi:10.1046/j.1365-8711.2000.03101.x)
- Berlind, A. A. *et al.* 2003 The halo occupation distribution and the physics of galaxy formation. *Astrophys. J.* **593**, 1–25. (doi:10.1086/376517)
- Blaizot, J., Wadadekar, Y., Guiderdoni, B., Colombi, S., Bertin, E., Bouchet, F. R., Devriendt, J. E. G. & Hatton, S. 2005 MoMaF: the mock map facility. *Mon. Not. R. Astron. Soc.* **360**, 159–175. (doi:10.1111/j.1365-2966.2005.09019.x)
- Blake, C. & Glazebrook, K. 2003 Probing dark energy using baryonic oscillations in the galaxy power spectrum as a cosmological ruler. *Astrophys. J.* **594**, 665–673. (doi:10.1086/376983)
- Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S. & Lacey, C. G. 2006 Breaking the hierarchy of galaxy formation. *Mon. Not. R. Astron. Soc.* **370**, 645–655. (doi:10.1111/j.1365-2966.2006.10519.x)
- Cimatti, A. *et al.* In press. SPACE: the Spectroscopic All-sky Cosmic Explorer. *Exp. Astron.* (<http://arXiv.org/abs/0804.4433>)
- Cole, S., Hatton, S., Weinberg, D. H. & Frenk, C. S. 1998 Mock 2dF and SDSS galaxy redshift surveys. *Mon. Not. R. Astron. Soc.* **300**, 945–966. (doi:10.1046/j.1365-8711.1998.01936.x)
- Cole, S. *et al.* 2005 The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *Mon. Not. R. Astron. Soc.* **362**, 505–534. (doi:10.1111/j.1365-2966.2005.09318.x)
- Conroy, C. & Wechsler, R. Submitted. Connecting galaxies, halos, and star formation rates across cosmic time. (<http://arXiv.org/abs/0805.3346>)
- Cooray, A. & Sheth, R. 2002 Halo models of large scale structure. *Phys. Rep.* **372**, 1–129. (doi:10.1016/S0370-1573(02)00276-4)
- Croce, M. & Scoccimarro, R. 2008 Nonlinear evolution of baryon acoustic oscillations. *Phys. Rev. D* **77**, 023 533. (doi:10.1103/PhysRevD.77.023533)
- Croton, D. J. *et al.* 2006 The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Mon. Not. R. Astron. Soc.* **365**, 11–28. (doi:10.1111/j.1365-2966.2005.09675.x)
- De Lucia, G., Springel, V., White, S. D. M., Croton, D. & Kauffmann, G. 2006 The formation history of elliptical galaxies. *Mon. Not. R. Astron. Soc.* **366**, 499–509. (doi:10.1111/j.1365-2966.2005.09879.x)
- Efstathiou, G. & Moody, S. J. 2001 Maximum likelihood estimates of the two- and three-dimensional power spectra of the APM galaxy survey. *Mon. Not. R. Astron. Soc.* **325**, 1603–1615. (doi:10.1046/j.1365-8711.2001.04575.x)
- Fosalba, P., Gaztañaga, E., Castander, F. J., Manera, M. Submitted. The onion universe: all sky lightcone simulations in shells. (<http://arXiv.org/abs/0711.1540>)
- Kaiser, N. 1984 On the spatial correlations of Abell clusters. *Astrophys. J. Lett.* **284**, L9. (doi:10.1086/184341)
- Kaiser, N. 1987 Clustering in real space and in redshift space. *Mon. Not. R. Astron. Soc.* **227**, 1–21.
- Monaco, P., Theuns, T. & Taffoni, G. 2002a The PINOCCHIO algorithm: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field. *Mon. Not. R. Astron. Soc.* **331**, 587. (doi:10.1046/j.1365-8711.2002.05162.x)
- Monaco, P., Theuns, T., Taffoni, G., Governato, F., Quinn, T. & Stadel, J. 2002b Predicting the number, spatial distribution and merging history of dark matter halos. *Astrophys. J.* **564**, 8–14. (doi:10.1086/324182)

- Peacock, J. A. & Dodds, S. J. 1994 Reconstructing the linear power spectrum of cosmological mass fluctuations. *Mon. Not. R. Astron. Soc.* **267**, 1020–1034.
- Peacock, J. A. & Smith, R. E. 2000 Halo occupation numbers and galaxy bias. *Mon. Not. R. Astron. Soc.* **318**, 1144–1156. (doi:10.1046/j.1365-8711.2000.03779.x)
- Pearce, F. R., Jenkins, A., Frenk, C. S., Colberg, J. M., White, S. D. M., Thomas, P. A., Couchman, H. M. P., Peacock, J. A. & Efstathiou, G. 1999 A simulation of galaxy formation and clustering. *Astrophys. J.* **521**, L99–L102. (doi:10.1086/312196)
- Percival, W. J. *et al.* 2001 The 2dF Galaxy Redshift Survey: the power spectrum and the matter content of the Universe. *Mon. Not. R. Astron. Soc.* **327**, 1297–1306. (doi:10.1046/j.1365-8711.2001.04827.x)
- Sánchez, A. G. & Cole, S. 2008 The galaxy power spectrum: precision cosmology from large-scale structure? *Mon. Not. R. Astron. Soc.* **385**, 830–840. (doi:10.1111/j.1365-2966.2007.12787.x)
- Sánchez, A. G., Baugh, C. M. & Angulo, R. In press. What is the best way to measure baryonic acoustic oscillations? *Mon. Not. R. Astron. Soc.* (<http://arXiv.org/abs/0804.0233>)
- Scoccimarro, R. & Sheth, R. K. 2002 P<sub>THALOS</sub>: a fast method for generating mock galaxy distributions. *Mon. Not. R. Astron. Soc.* **329**, 629–640. (doi:10.1046/j.1365-8711.2002.04999.x)
- Scoccimarro, R., Sheth, R., Hui, L. & Jain, B. 2001 How many galaxies fit in a halo? Constraints on galaxy formation efficiency from spatial clustering. *Astrophys. J.* **546**, 20–34. (doi:10.1086/318261)
- Seo, H.-J. & Eisenstein, D. J. 2003 Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys. *Astrophys. J.* **598**, 720–740. (doi:10.1086/379122)
- Seo, H.-J. & Eisenstein, D. J. 2005 Baryonic acoustic oscillations in simulated galaxy redshift surveys. *Astrophys. J.* **633**, 575–588. (doi:10.1086/491599)
- Seo, H.-J. & Eisenstein, D. J. 2007 Improved forecasts for the baryon acoustic oscillations and cosmological distance scale. *Astrophys. J.* **665**, 14–24. (doi:10.1086/519549)
- Seo, H.-J., Siegel, E. R., Eisenstein, D. J. & White, M. In press. Non-linear structure formation and the acoustic scale. *Astrophys. J.* (<http://arXiv.org/abs/0805.0117>)
- Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G. & Couchman, H. M. P. 2003 Stable clustering, the halo model and non-linear cosmological power spectra. *Mon. Not. R. Astron. Soc.* **341**, 1311–1332. (doi:10.1046/j.1365-8711.2003.06503.x)
- Smith, R. E., Scoccimarro, R. & Sheth, R. K. 2007 Scale dependence of halo and galaxy bias: effects in real space. *Phys. Rev. D* **75**, 063 512. (doi:10.1103/PhysRevD.75.063512)
- Smith, R. E., Scoccimarro, R. & Sheth, R. K. 2008 Motion of the acoustic peak in the correlation function. *Phys. Rev. D* **77**, 043 525. (doi:10.1103/PhysRevD.77.043525)
- Springel, V. *et al.* 2005 Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* **435**, 629–636. (doi:10.1038/nature03597)
- White, S. D. M. & Rees, M. J. 1978 Core condensation in heavy halos—a two-stage theory for galaxy formation and clustering. *Mon. Not. R. Astron. Soc.* **183**, 341–358.
- White, S. D. M., Frenk, C. S., Davis, M. & Efstathiou, G. 1987 Clusters, filaments, and voids in a universe dominated by cold dark matter. *Astrophys. J.* **313**, 505. (doi:10.1086/164990)
- Yang, X., Mo, H. J. & van den Bosch, F. C. 2003 Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies. *Mon. Not. R. Astron. Soc.* **339**, 1057–1080. (doi:10.1046/j.1365-8711.2003.06254.x)
- Yang, X., Mo, H. J., Jing, Y. P., van den Bosch, F. C. & Chu, Y. Q. 2004 Populating dark matter haloes with galaxies: comparing the 2dFGRS with mock galaxy redshift surveys. *Mon. Not. R. Astron. Soc.* **350**, 1153–1173. (doi:10.1111/j.1365-2966.2004.07744.x)

## AUTHOR PROFILE

**Carlton M. Baugh**



Dr Carlton Baugh has just completed a Royal Society University Research Fellowship at Durham University, where he is now a reader. His research interests include semi-analytical models of galaxy formation and measuring the large-scale structure of the Universe. Dr Baugh was part of the team of UK and Australian astronomers who carried out and analysed the Two-Degree Field Galaxy Redshift Survey. He is now working to produce mock catalogues for Pan-STARRS and is part of the Euclid consortium to build the ultimate survey of the Universe from space.