

The spatial distribution of cold gas in hierarchical galaxy formation models

Han-Seek Kim,^{1*} C. M. Baugh,¹ A. J. Benson,² S. Cole,¹ C. S. Frenk,¹ C. G. Lacey,¹ C. Power³ and M. Schneider¹

¹*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE*

²*Theoretical Astrophysics, Caltech, MC350-17, 1200 E. California Blvd., Pasadena, CA 91125, USA*

³*Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH*

Accepted 2011 February 17. Received 2011 February 16; in original form 2010 March 1

ABSTRACT

The distribution of cold gas in dark matter haloes is driven by key processes in galaxy formation: gas cooling, galaxy mergers, star formation and reheating of gas by supernovae. We compare the predictions of four different galaxy formation models for the spatial distribution of cold gas. We find that satellite galaxies make little contribution to the abundance or clustering strength of cold gas selected samples, and are far less important than they are in optically selected samples. The halo occupation distribution function of present-day central galaxies with cold gas mass $>10^9 h^{-1} M_{\odot}$ is peaked around a halo mass of $\approx 10^{11} h^{-1} M_{\odot}$, a scale that is set by the AGN suppression of gas cooling. The model predictions for the projected correlation function are in good agreement with measurements from the H I Parkes All-Sky Survey. We compare the effective volume of possible surveys with the Square Kilometre Array with those expected for a redshift survey in the near-infrared. Future redshift surveys using neutral hydrogen emission will make possible measurements of the baryonic acoustic oscillations that are competitive with the most ambitious spectroscopic surveys planned in the near-infrared.

Key words: galaxies: evolution – galaxies: formation – galaxies: luminosity function, mass function – large-scale structure of Universe.

1 INTRODUCTION

Cold gas is central to galaxy formation yet little is known about how much there is in the Universe at different epochs and how this gas is distributed in dark matter haloes of different mass.

The neutral hydrogen content of galaxies has been probed at high redshifts ($z > 2$) using the absorption of the Lyman α line by gas clouds along the line of sight to distant quasars (e.g. Lanzetta et al. 1991; Wolfe et al. 1995; Storrie-Lombardi, Irwin & McMahon 1996; Peroux et al. 2005; Wolfe, Gawiser & Prochaska 2005). A complementary probe of the atomic hydrogen content of galaxies and the physical state of the gas is the 21-cm line. A blind survey of 21-cm line absorption of gas illuminated by background radio sources has been proposed as an unbiased probe of damped Lyman α clouds, which would extend to objects with high dust content, unlike surveys for damped Lyman α absorbers (Kanekar & Briggs 2004; Kanekar et al. 2009). The 21-cm absorption line and absorption lines associated with molecular hydrogen have been detected in the same damped Lyman α system (Srianand et al. 2010).

The 21-cm line is a forbidden transition and the detection of the line in emission for low-gas masses and high redshifts is challenging. It is only in recent years that a robust and comprehensive census of atomic hydrogen (H I) in the local Universe has been made possible through the H I Parkes All Sky Survey (Barnes et al. 2001; Zwaan et al. 2003, 2005). This work is being extended to lower mass systems by the ALFALFA survey (Giovanelli et al. 2005; Martin et al. 2010). Despite this progress, the highest redshift direct detection of H I in emission is very firmly confined to the local Universe at $z = 0.25$ (Catinella et al. 2008). (Lah et al. 2009 have reached $z = 0.34$ by stacking a sample of optically selected galaxies; see also Verheijen et al. 2007.) However, over the coming decade, this situation is expected to change dramatically with the construction of new, more sensitive radio telescopes such as Apertif¹ (Verheijen et al. 2008), the precursors of the Square Kilometre Array, MeerKAT² (Booth et al. 2009) and ASKAP³ (Johnston et al. 2008), and the Square Kilometre Array⁴ itself (Schilizzi, Dewdney & Lazio 2008).

¹ <http://www.astron.nl/general/apertif/apertif>

² <http://www.ska.ac.za/meerkat>

³ <http://www.atnf.csiro.au/SKA>

⁴ <http://www.skatelescope.org>

*E-mail: h.s.kim@durham.ac.uk

Apertif is a prototype focal plane array which is being developed for the Westerbork Synthesis Radio Telescope, and will increase its field of view by a factor of 25 and double the available bandwidth (Verheijen et al. 2008). The Widefield ASKAP *L*-band Legacy All-sky Blind survey (Johnston et al. 2008) will reach a median redshift of $z \sim 0.05$, compared with $z \sim 0.009$ reached by HIPASS, putting our knowledge of the local H I Universe on a par with that in the optical due to the 2dFGRS and SDSS. MeerKAT will be more sensitive than ASKAP but with a smaller field of view, pushing H I detections to higher redshifts. The SKA will revolutionize our understanding of galaxy formation and cosmology, uncovering the H I Universe out to high redshifts. One of the major science goals is to better characterize the evolution of dark energy with redshift. The SKA is expected to provide competitive constraints on the nature of dark energy through high-accuracy measurement of large-scale structure in the galaxy distribution over a look-back time representing a significant fraction of the age of the Universe (Albrecht et al. 2006). This conclusion currently rests on very uncertain calculations which we seek to place on a firmer, more physical footing in this paper.

Modelling the abundance and clustering of H I sources is challenging. A number of possible approaches have been tried: empirical modelling, which relies upon the observations of H I in the Universe; the fully numerical approach, which uses cosmological gas dynamics simulations to model the H I content of galaxies from first principles; and semi-analytical modelling, which we use in this paper. Empirical estimates have been attempted despite the paucity of observational results for guidance (Abdalla & Rawlings 2005; Abdalla, Blake & Rawlings 2010). Such calculations require an assumption about the evolution of the H I mass function over a broad redshift interval. The only constraint on this assumption is the integrated density of H I, which can be compared with the results inferred from quasar absorption features, which themselves require corrections for unseen low column density systems and dust extinction (Storrie-Lombardi et al. 1996). The empirical approach does not predict the clustering of H I sources. Further assumptions and approximations are necessary to extend this class of modelling so that predictions can be made for galaxy clustering. Another layer of approximation in this class of modelling has been motivated by observations that suggest that H I sources tend to avoid the centres of clusters and that clusters do not host an important population of satellites (e.g. Waugh et al. 2002; Verheijen et al. 2007). This led Marin et al. (2010) to make a one-to-one connection between halo mass and H I mass. However, the nature of the relation is uncertain and several possibilities are explored by Marin et al. based on different assumptions about the evolution of the H I mass function.

Ideally, a physically motivated model which follows the sources and sinks of cold gas is needed. Gas dynamic simulations are computationally expensive and are typically restricted to small computational volumes, which makes it impossible to accurately follow the growth of structure to the present day. An example is provided by Popping et al. (2009), who carry out a smoothed particle hydrodynamics simulation in a $32 h^{-1}$ Mpc box. The H I mass function in the simulation is in very poor agreement with the observational estimate of Zwaan et al. (2005), underpredicting the abundance of galaxies of H I mass $10^{10} M_{\odot}$ by a factor of 30, which the authors put down to the small computational volume, and overpredicting low-mass systems by a factor of 2. Clustering predictions are limited to scales smaller than a few Mpc due to the small box size. Furthermore, it is important to be aware that gas dynamic simulations do not have the resolution to follow all of the processes in galaxy formation directly

and in all cases resort to what are essentially semi-analytical rules to treat subresolution physics.

Currently the most promising route to making physical and robust predictions for the H I in the Universe is semi-analytical modelling of galaxy formation (see Baugh 2006; Benson 2010). This type of model includes a simplified but physically motivated treatment of the processes which control the amount of cold gas in a galaxy: gas cooling, galaxy mergers, star formation and reheating of gas by supernovae. These calculations are quick and can rapidly cover the haloes in a cosmological volume. Baugh et al. (2004) presented predictions for the mass function of cold gas galaxies in the GALFORM semi-analytical model of Cole et al. (2000). One issue which must be dealt with is that the models predict only the total mass of cold gas, which includes helium, and both atomic and molecular hydrogen.⁵ Baugh et al. assumed a fixed ratio of molecular to atomic hydrogen. Obreschkow & Rawlings (2009a,b) developed an empirical model based on observations and theoretical arguments by Blitz & Rosolowsky (2006) in which this ratio could vary from galaxy to galaxy. Obreschkow & Rawlings (2009b) applied this ansatz to the semi-analytical model of de Lucia & Blaizot (2007; see also Obreschkow et al. 2009).

In the first paper in this series, we compared the predictions of a range of semi-analytical models for the mass function of H I (Power, Baugh & Lacey 2010). Despite the different implementations of the physical ingredients used in the models and the different emphasis placed on various observations when setting the model parameters, the predictions show generic features. Power et al. found that there is surprisingly little variation in the predicted H I mass function with redshift, and that the models make similar predictions for the rotation speed and size of H I systems. The models predict the mass of cold gas and so a conversion is required to turn this into a H I mass. Currently the most uncertain step is the assumption about what fraction of hydrogen is in atomic form and what fraction is molecular. Power et al. presented predictions for two cases, one in which all the model galaxies are assumed to have a fixed molecular to atomic hydrogen ratio ($H_2/H I$) and the other in which this ratio varies from galaxy to galaxy, depending upon the local conditions in the galactic disc (Blitz & Rosolowsky 2006). The assumption of a variable $H_2/H I$ ratio results in a dramatic reduction in the number of H I sources in the tail of the redshift distribution.

In this paper we look at the distribution of cold gas in galaxies as a function of halo mass. In particular we look at the halo occupation distribution (HOD) for H I galaxies, which gives the mean number of galaxies of a given H I mass as a function of dark matter halo mass, and the clustering of H I galaxies. Using this information, we assess the potential of the SKA to measure the baryonic acoustic oscillation (BAO) signal. We briefly review the GALFORM model in Section 2, explaining the differences between the four models that we consider. We then look at the HOD of cold gas galaxies in Section 3, in which we also present predictions for the clustering of cold gas galaxies at different redshifts and compare it to the measured clustering at the present day. In Section 4 we compare the performance of future redshift surveys in the optical and using H I emission for measuring the properties of the dark energy. We present a summary along with our conclusions in Section 5.

⁵ After this paper was submitted, Lagos et al. (2010) presented a version of the GALFORM model which tracks the atomic and molecular hydrogen components separately.

2 GALAXY FORMATION MODELS AND BASIC PREDICTIONS

Semi-analytical models of galaxy formation invoke simple, physically motivated recipes to follow the fate of the baryons in a universe in which structure in the dark matter grows hierarchically (White & Rees 1978; White & Frenk 1991; Kauffmann, White & Guiderdoni 1993; Cole et al. 1994; for a review of this approach, see Baugh 2006). The current generation of models include a wide range of phenomena, ranging from the heating of the intergalactic medium, which affects the cooling of gas in low-mass haloes, to the suppression of cooling flows in massive haloes due to heating by accretion of matter on to supermassive black holes (e.g. Bower et al. 2006; Croton et al. 2006; Cattaneo et al. 2007; Monaco, Fontanot & Taffoni 2007; Lagos, Cora & Padilla 2008). In this paper, we use the Durham semi-analytical galaxy formation code `GALFORM` to make predictions for the cold gas content of galaxies in dark matter haloes of different masses. This code was introduced by Cole et al. (2000) and has been developed in a series of papers (Benson et al. 2003; Baugh et al. 2005; Bower et al. 2006; Font et al. 2008). The code predicts a wide range of properties for the galaxy population in the context of a spatially flat cold dark matter cosmology with a cosmological constant.

In this paper we consider four different models run using `GALFORM`. Two of these are available from the Millennium Archive;⁶ these are the Bower et al. (2006; hereafter Bow06) and Font et al. (2008) models (hereafter Font08). The third model is a modified version of the Bow06 model (which we label as MHIBow06), which is discussed in more detail below. In this model a small number of parameters have been adjusted from the values used in Bow06 in order to produce a better match to the cold gas mass function estimated by Zwaan et al. (2005). These three models are set in the Millennium Simulation, a cosmological N -body simulation which follows the growth of structure in the dark matter (Springel et al. 2005). The fourth model (denoted by GpcBow06) is set in a different background cosmology which is in better agreement with recent measurements of the cosmic microwave background and the large-scale structure of the Universe (Sanchez et al. 2009).⁷ The Bow06, Font08 and MHIBow06 models use merger histories extracted from the Millennium Simulation. The GpcBow06 model uses Monte Carlo generated merger trees as described below. When we make predictions for the spatial distribution of galaxies in the GpcBow06 model, we use the GigaParsec simulation run at the Institute for Computational Cosmology (GPICC; Baugh et al., in preparation), which uses 10 billion particles to model the hierarchical clustering of mass in a simulation cube $1000 h^{-1}$ Mpc on a side. To keep the number of models manageable, we do not consider the Baugh et al. (2005) model in this paper. This model was included in the study by Power et al. (2010). The star formation recipe used in the MHIBow06 model is based on that used in Baugh et al. (2005).

The MHIBow06 and GpcBow06 models use the same basic physical ingredients as the Bow06 model but with a change in the assumed star formation time-scale. The Font08 model is based on

Bow06, with a modification to the cooling prescription. We discuss these differences in more detail below. We first discuss some of the ingredients which are varied between the models, in order to introduce some of the parameter definitions used in `GALFORM`.

All of the models we consider in this paper include the suppression of cooling flows in massive haloes, as a result of the energy released following accretion of matter on to a central supermassive black hole (Bower et al. 2006; Malbon et al. 2007; Fanidakis et al. 2011). A halo is assumed to be in quasi-hydrostatic equilibrium if the time required for gas to cool at the cooling radius, $t_{\text{cool}}(r_{\text{cool}})$, exceeds a multiple of the free-fall time at this radius, $t_{\text{ff}}(r_{\text{cool}})$:

$$t_{\text{cool}}(r_{\text{cool}}) > \frac{1}{\alpha_{\text{cool}}} t_{\text{ff}}(r_{\text{cool}}), \quad (1)$$

where α_{cool} is an adjustable parameter, whose value controls the sharpness and position of the break in the optical luminosity function. The cooling flow in the halo is then shut down completely if the luminosity released by accretion of matter on to the supermassive black hole (SMBH) exceeds the cooling luminosity. The energy released by accretion depends on the mass of the SMBH (see, for example, Fanidakis et al. 2011).

The models also include the ejection of cooled gas into the hot halo due to heating by supernovae. The strength of supernovae feedback is defined by the factor β :

$$\beta = (V_{\text{hot}}/V_{\text{disc}})^{\alpha_{\text{hot}}}. \quad (2)$$

The rate at which gas is reheated is β times the star formation rate. Here V_{disc} is the circular velocity of the disc at its half-mass radius and V_{hot} and α_{hot} are parameters. A similar equation holds for supernova feedback in the galactic bulge. In `GALFORM` the parameters V_{hot} and α_{hot} are set without reference to the number of supernovae. The primary constraints on these parameters are the shape of the luminosity function, the slope of the disc rotation speed–luminosity relation and the scale size of discs (see Cole et al. 2000).

The Bow06 and Font08 models use a star formation time-scale in discs, τ_* , which is proportional to the galactic dynamical time, τ_{dyn} , and is given by

$$\tau_* = \epsilon_*^{-1} \tau_{\text{dyn}} (V_{\text{disc}}/200 \text{ km s}^{-1})^{\alpha_*}, \quad (3)$$

where α_* and ϵ_* are adjustable parameters ($\alpha_* = -1.5$ in both cases). The dynamical time is defined as $\tau_{\text{dyn}} = r_{\text{disc}}/V_{\text{disc}}$. In contrast, the MHIBow06 and GpcBow06 models adopt a star formation time-scale which does not depend on the galactic dynamical time. Instead, in these cases, the time-scale is given by

$$\tau_* = \tau_*^0 (V_{\text{disc}}/200 \text{ km s}^{-1})^{\alpha_*}, \quad (4)$$

where τ_*^0 and α_* are adjustable parameters (again, in both cases, $\alpha_* = -1.5$); this parametrization was used in Baugh et al. (2005).

The Font08 model includes an improved treatment of the ram-pressure stripping of hot-gas atmospheres of satellite galaxies, motivated by the hydrodynamic simulations of McCarthy et al. (2008). Also in this model, the yield of metals per solar mass of stars formed is increased by a factor of 2 over the default but rather uncertain value expected of a standard solar neighbourhood stellar initial mass function. These changes are motivated in part by the desire to improve the predictions of the Bow06 model for the colour–magnitude relation of central and satellite galaxies in groups. The revision to the stellar yield reddens the colour of all galaxies in the Font08 model compared with Bow06. The change in the cooling model changes the relative abundance of galaxies in the red and blue populations at low luminosities. In the Font08 model, there are more faint blue satellite galaxies than in the Bow06 model. These

⁶ <http://galaxy-catalogue.dur.ac.uk:8080/Millennium/>

⁷ The cosmological parameters used in the Millennium simulation are a matter density $\Omega_0 = 0.25$, a cosmological constant $\Lambda_0 = 0.75$, a Hubble constant $H_0 = 73 \text{ km s}^{-1} \text{ Mpc}^{-1}$, a primordial scalar spectral index $n_s = 1$, baryon density $\Omega_b = 0.045$ and fluctuation amplitude $\sigma_8 = 0.9$. In the Sanchez et al. (2009) best-fitting model these parameters become $\Omega_0 = 0.26$, $\Lambda_0 = 0.74$, $H_0 = 71.5 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $n_s = 0.96$, $\Omega_b = 0.044$ and $\sigma_8 = 0.8$.

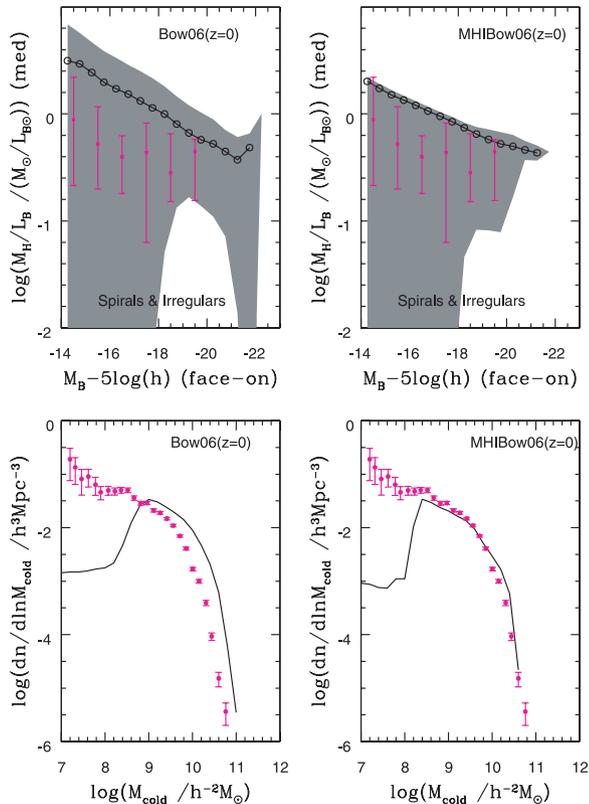


Figure 1. The predicted ratio of neutral hydrogen mass to B -band luminosity (upper panels) and the cold gas mass function (lower panels) in the Bow06 (left-hand panels) and MHI Bow06 models (right-hand panels). In the upper panels, the magenta points show observational estimates of the hydrogen mass to luminosity ratio using data from Huchtmeier & Richter (1988) (H I) and Sage (1993) (H₂). The black points show the median ratio predicted by the models and the grey shading shows the 20–80 percentile range of the predicted distribution. We assume that 76 per cent by mass of the cold gas predicted by the models is neutral hydrogen. In the lower panels, the magenta points show the cold gas mass function derived from the H I mass function estimated by Zwaan et al. (2005). Here, a constant H₂/H I ratio of 0.4 has been assumed to convert the H I measurement into a cold gas mass.

galaxies are starved of freshly cooled gas in Bow06 and so have redder stellar populations. The predicted colours in the Font08 model are in much better agreement with the observed colour–magnitude relation measured by Weinmann et al. (2006).

The motivation for the MHI Bow06 model is clear from Fig. 1. This plot shows the galactic neutral hydrogen mass to optical lumi-

nosity ratio and the cold gas mass function at the present day. Note that when we plot the mass function (lower panels of Fig. 1), cold gas masses are plotted in units of $h^{-2} M_{\odot}$ rather than $h^{-1} M_{\odot}$, which is the unit used in the simulation. This ensures that the observational units (which depend upon the square of the luminosity distance) are matched. The Bow06 model predicts a gas mass to luminosity ratio with the wrong zero-point and slope. Since this model gives an excellent match to the local optical luminosity function, the discrepancy in the gas-to-luminosity ratio results in a poor match to the cold gas mass function. The MHI Bow06 model uses the star formation time-scale given by equation (4) and also adopts a different value for the AGN feedback free parameter, α_{cool} (equation 1; see Table 1). The right-hand panels of Fig. 1 show that the MHI Bow06 model is in much better agreement with the observed gas-to-luminosity ratio and cold gas mass function for cold gas masses in excess of $\sim 3 \times 10^8 h^{-2} M_{\odot}$. The MHI Bow06 model also produces a good match to the observed local galaxy luminosity functions in the optical and near-infrared. Note that the models predict the mass of cold gas, which includes helium, atomic hydrogen and molecular hydrogen. The observed mass function in the lower panels of Fig. 1 is measured in terms of the atomic hydrogen (H I) content of galaxies. To convert this into a cold gas mass, we have assumed a fixed ratio of molecular to atomic hydrogen and corrected for the mass fraction of helium (see Power et al. 2010). We have also tried applying a variable H₂/H I conversion, as advocated by Blitz & Rosolowsky (2006) and Obreschkow & Rawlings (2009a). Whilst this improves the comparison of the Bow06 H I mass function with the observational estimate of Zwaan et al. (2005), to obtain a satisfactory match it is still necessary to change the model parameters, as we do in the MHI Bow06 model. We shall return to this point in Section 5.

The GpcBow06 model starts from the Bow06 model, with small adjustments made to the galaxy formation parameters to obtain a good match to the optical luminosity function (this is required because the cosmological model is different from that used in Bow06) and also to reproduce the observed H I mass function. The GpcBow06 model uses Monte Carlo merger trees generated using the algorithm devised by Parkinson, Cole & Helly (2008).

Fig. 2 shows the cold gas mass function predicted by the four models at $z = 0, 1$ and 2 . The Bow06 and Font08 models overpredict the abundance of galaxies with a given cold gas mass at $z = 0$ compared with the observational estimate by Zwaan et al. (2005). On the other hand, the cold gas mass functions of the MHI Bow06 and GpcBow06 models agree well with the observations for masses in excess of $10^{8.5} h^{-2} M_{\odot}$. The discrepancy between the predictions and observations at lower masses is not due to the finite resolution of the N -body halo merger trees. The turnover can be traced back to the modelling of the photoionization of the intergalactic medium

Table 1. The values of selected parameters which differ between the models. The columns are as follows: (1) The name of the model; (2) The equation used to calculate the star formation time-scale, τ_{\star} ; (3) The value of ϵ_{\star} or τ_{\star}^0 used in the star formation time-scale; (4) The AGN feedback parameter, α_{cool} , (equation 1); (5) The supernova feedback parameter, V_{hot} (equation 2); (6) The source of halo merger histories; (7) Comments giving model source or key differences from published models.

	τ_{\star}	ϵ_{\star} or τ_{\star}^0 (Gyr)	α_{cool}	V_{hot} (km s ⁻¹)	Merger tree	Comments
Bow06	Equation (3)	0.0029	0.58	485	N -body	Bower et al. (2006)
Font08	Equation (3)	0.0029	0.70	485	N -body	Font et al. (2008)
MHI Bow06	Equation (4)	8	0.62	485	N -body	Modified cooling recipe in satellites from Bow06
GpcBow06	Equation (4)	4	0.72	390	Monte Carlo	Different background cosmology and modified star formation recipe from Bow06

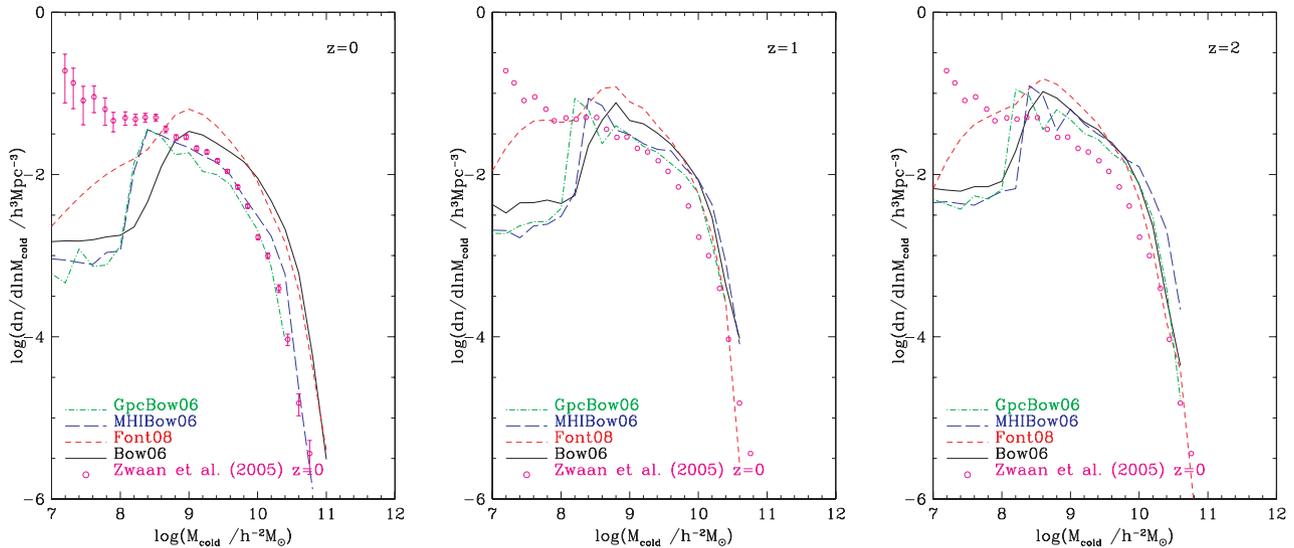


Figure 2. The cold gas mass function predicted in the four models at $z=0$ (left), $z=1$ (middle) and $z=2$ (right). Different colours and line types correspond to different models as indicated by the legend. The points show the local ($z=0$) observational estimate of the cold gas mass function inferred from the H I mass function of Zwaan et al. (2005) (see text in Section 3.2 for details of the conversion). These data are reproduced without error bars in the $z=1$ and $z=2$ panels as a reference from which to illustrate the evolution of the mass function.

and the impact this has on the cooling of gas in low-mass haloes. In all cases a particularly simple approach is taken to model this effect, whereby cooling in low circular velocity haloes ($v_c < v_{\text{cut}}$) is suppressed at redshifts below the redshift at which the Universe is assumed to have been reionized (z_{cut}). The parameters adopted ($v_{\text{cut}} = 50 \text{ km s}^{-1}$ and $z_{\text{cut}} = 6$) may overestimate the impact of this effect according to recent simulations by Okamoto, Gao & Theuns (2008). The form of the observed H I mass function at low masses could give interesting constraints on the modelling of photoionization and supernova feedback (Kim et al., in preparation). Here we focus on the more massive galaxies which dominate the overall H I content of the Universe.

3 THE SPATIAL DISTRIBUTION OF COLD GAS

We now compare the predictions of the four galaxy formation models for the spatial distribution of cold gas with one another and with observations. To understand the spatial distribution of cold gas, we first look at the HOD (Benson et al. 2000; Peacock & Smith 2000; Berlind & Weinberg 2002). This allows us to quantify the number of galaxies above a given cold gas mass, as a function of dark matter halo mass (Section 3.1). We present predictions for the correlation function of galaxies selected by their cold gas mass in Section 3.2.

3.1 The halo occupation distribution

3.1.1 Variation of cold gas mass with halo mass

Before considering the HOD directly, it is instructive to first look at how the cold gas mass of galaxies varies with the mass of their host dark matter halo, which we plot in Fig. 3 for the Bow06 model. The median cold gas mass as a function of host halo mass is plotted separately for central and satellite galaxies. There is a tight correlation between the cold gas mass of a central galaxy and its host halo mass for galaxies in haloes less massive than $\sim 3 \times 10^{11} h^{-1} M_{\odot}$. In haloes more massive than this, AGN feedback suppresses gas cooling and there is a dramatic break in the galaxy cold gas mass–

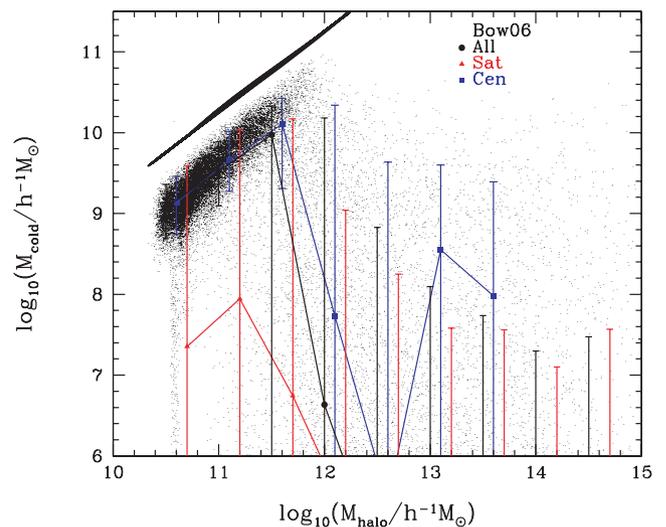


Figure 3. The cold gas mass of galaxies in the Bow06 model as a function of the mass of their host dark matter halo. The black points show individual galaxies. The symbols joined by lines show the median cold gas mass as a function of halo mass, for central galaxies (blue), satellite galaxies (red) and all galaxies (black). The bars show the 10–90 percentile range of the distribution of cold gas masses. All galaxies, including those with zero cold gas mass are included when computing the median and percentile range. The solid black line shows the cold gas mass a galaxy would have if all the available baryons in its halo were in the form of cold gas in one object.

halo mass relation, with an accompanying increase in the scatter. The galaxies with the largest mass of cold gas do not lie in the most massive dark matter haloes, but reside instead in haloes with masses $\sim 10^{12} h^{-1} M_{\odot}$. The predicted cold gas mass–halo mass relation is remarkably similar to that inferred observationally (Wyithe et al. 2009). Another conclusion that is readily apparent from Fig. 3 is that the bulk of the baryons associated with a dark matter halo are not in the form of cold gas. The solid line in this plot shows the mass a galaxy would have if all of the available baryons in the halo

were in the form of cold gas in one object, assuming the universal baryon fraction. The points are some way below this line for three reasons: (1) in most haloes, the bulk of the baryons are in the hot phase; (2) some baryons are in the form of stars; (3) there is more than one galaxy in most haloes.

3.1.2 Cold gas halo occupation distributions

We now examine the predictions for the HOD of galaxy samples constructed according to cold gas mass. The HOD gives the mean number of galaxies which satisfy a given selection criterion as a function of halo mass, and can be broken down into the contribution from the central galaxy in a halo and its satellite galaxies. In the case of optically selected galaxy samples, the HOD is commonly described by a step function for central galaxies and a power law for satellite galaxies (Peacock & Smith 2000; Seljak 2000; Berlind & Weinberg 2002; Zheng 2004). Many attempts have been made to interpret the clustering of optically selected galaxy samples using the HOD formalism (Magliocchetti & Porciani 2003; van den Bosch, Mo & Yang 2003; Zehavi et al. 2005; Yang et al. 2005; Tinker et al. 2007; Wake et al. 2008; Kim et al. 2009). In contrast, there are few studies of the clustering of galaxies selected on the basis of their atomic hydrogen mass using the HOD formalism (Wyithe et al. 2009; Wyithe & Brown 2010; Marin et al. 2010).

Fig. 4 shows the typical form predicted by the models for the HOD of galaxies selected by their cold gas mass. The left-hand panel shows the HOD for galaxies in the Bow06 model which have cold gas masses in excess of $3 \times 10^9 h^{-2} M_\odot$, chosen to have the same H I mass cut as HIPASS. For this mass threshold, the abundance of central galaxies is sharply peaked around a halo mass of $\sim 2 \times 10^{11} h^{-1} M_\odot$. The HOD of satellite galaxies reaches unity in haloes which are a hundred times more massive. In these haloes, the central galaxy has a cold gas mass below the cut-off; there is essentially zero chance of finding a halo which contains a central galaxy and a satellite galaxy above this cold gas mass threshold.

However, this does not imply that it is impossible to find more than one galaxy per halo with cold gas masses above the threshold, simply that when this occurs (i.e. once a sufficiently massive halo is considered), both galaxies will be satellites.

For comparison, we also plot in the left-hand panel of Fig. 4 the traditional form adopted for the HOD of central galaxies (i.e. a step function). The minimum halo mass in this case is set by the requirement that the step function reproduces the number of central galaxies in the Bow06 model. The step function HOD is markedly different to the predicted HOD, which is closer to a Gaussian. A similar conclusion about the peaked form of the central galaxy HOD was postulated by Zehavi et al. (2005) for blue central galaxies. Wyithe et al. (2009) model the clustering of galaxies in the HIPASS survey by adopting a step function for the central galaxy HOD and a truncated power law for satellite galaxies, such that haloes above some mass cut contain no satellites. The truncation point lies in the halo mass range $10^{14} - 10^{15} h^{-1} M_\odot$, depending on the slope of the satellite HOD. As we shall see later on, whilst this truncation is not predicted by any of the models, this has little impact on the abundance or clustering of the galaxies.

In Fig. 4, the HOD of central galaxies in the Bow06 model drops far below unity above a halo mass of $\sim 10^{12} h^{-1} M_\odot$. In this model there is very little cold gas in haloes more massive than this due to the shut down of the cooling flow by AGN heating. To illustrate this, in the middle panel of Fig. 4 we vary the halo mass which marks the onset of AGN heating by changing the value of the α_{cool} parameter (see equation 1). Reducing the value of α_{cool} results in the halo mass in which cooling stops being shifted to higher masses. In the absence of AGN heating (i.e. $\alpha_{\text{cool}} = 0$), the central galaxy HOD still drops below the unity in the most massive haloes ($M_{\text{halo}} > 10^{13} h^{-1} M_\odot$) due to the longer cooling time of the gas in these haloes. These haloes typically have a lower formation redshift and thus a lower gas density and are also hotter; hence they have a longer cooling time. Cold gas is also depleted by star formation in these haloes.

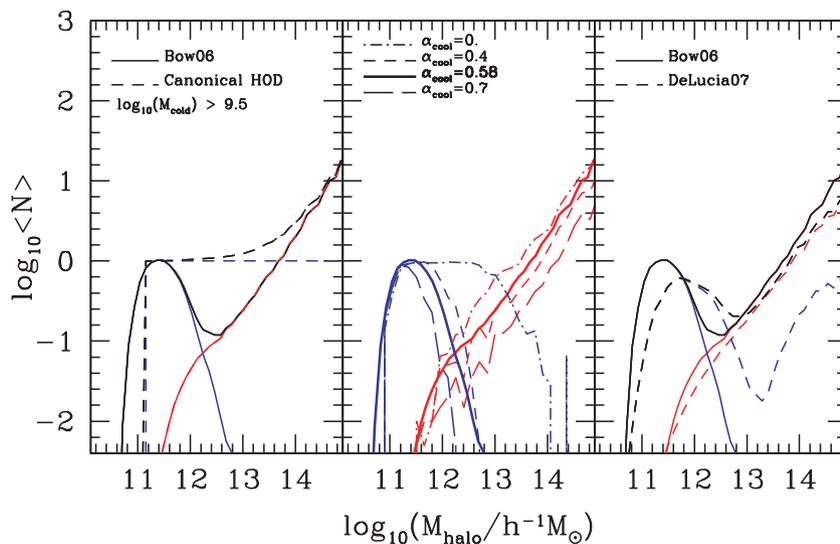


Figure 4. The predicted halo occupation distribution (HOD) of galaxies with cold gas mass in excess of $10^{9.5} h^{-2} M_\odot$, chosen to match the sample of galaxies for which Wyithe et al. (2009) estimated the HOD for in HIPASS. The left-hand panel shows the HOD predicted in the Bow06 model (solid lines: blue shows the central galaxy HOD, red shows satellites and black shows the overall HOD). The dashed blue line shows a step function designed to reproduce the number of central galaxies in Bow06. The dashed black line shows this step function combined with the model HOD for satellites. The central panel shows the impact on the HOD of changing the halo mass above which AGN feedback stops the cooling flow. The fiducial Bow06 model corresponds to $\alpha_{\text{cool}} = 0.58$. In a model with a larger value of α_{cool} , the onset of cooling suppression shifts to lower mass haloes. The right-hand panel compares the HOD predicted by Bow06 (solid lines) with that in the model of De Lucia & Blaizot (2007), for the same cold gas mass threshold (dashed lines). The colour coding is the same in each panel.

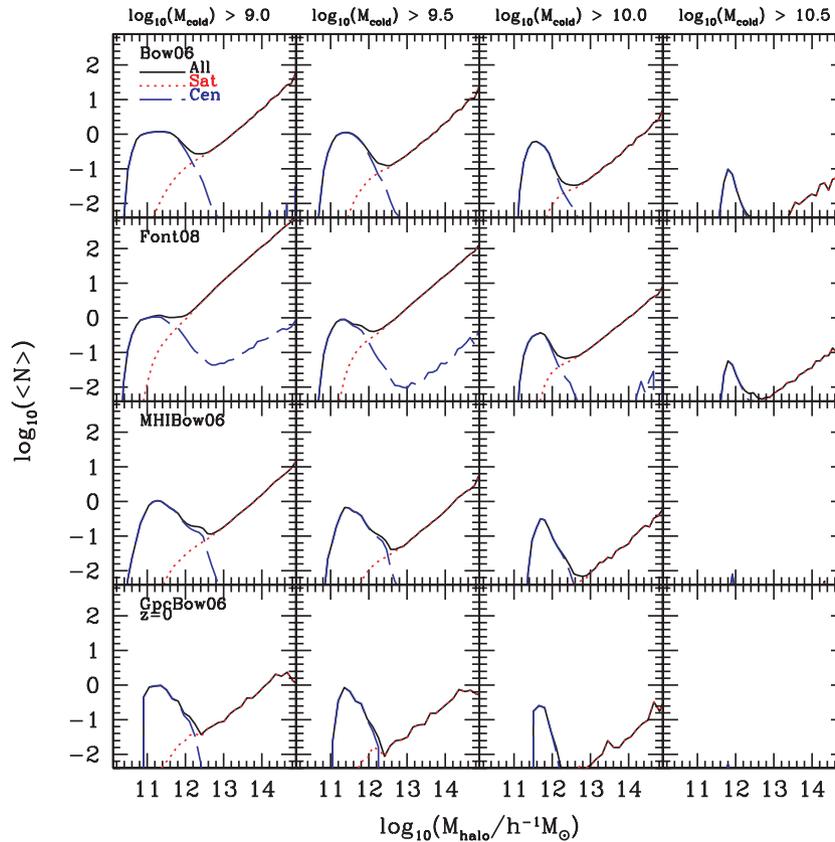


Figure 5. The halo occupation distribution, i.e. the mean number of galaxies per halo which pass the selection labelled at the top of each column, at $z = 0$ for galaxy samples defined by cold gas mass thresholds. The blue dashed curves show the contribution from central galaxies, the red dotted curves show satellite galaxies and the black solid curves show the overall HOD. Each row corresponds to a different model, and each column to a different cold gas mass threshold, as labelled.

We shall see later that the peaked HOD for central galaxies is common to all of the GALFORM models considered, particularly at low redshift. We now examine whether or not this feature is peculiar to the way AGN feedback is implemented in GALFORM by comparing the Bow06 predictions with those of De Lucia & Blaizot (2007; hereafter the DeLucia07 model). The right-hand panel of Fig. 4 shows that the central galaxy HOD in the DeLucia07 model is somewhat broader than that predicted in Bow06, and even increases beyond a halo mass of $\sim 2 \times 10^{13} h^{-1} M_{\odot}$. However, as we shall demonstrate further on in this section, this upturn has little impact on the predicted clustering. The suppression of gas cooling in the DeLucia07 semi-analytical model is smoother than in GALFORM (see Croton et al. 2006 for a description of the implementation of radio mode feedback). Some gas is permitted to cool in haloes with hot gas atmospheres in the DeLucia07 model, with the cooling rate modified by accretion on to the central SMBH. In GALFORM, the cooling flow and heating rate are assumed to balance exactly whenever there is a quasi-hydrostatic hot halo and the Eddington luminosity of the black hole exceeds the cooling luminosity, turning off gas cooling completely.

Figs 5–7 show the HOD in the four Durham models at $z = 0, 1$ and 2. Each column shows the HOD predicted for a different cold gas mass threshold, with the mass cut increasing to the right. The rows show the different models introduced in Section 2. For the most massive cold gas mass threshold plotted in Fig. 5, the mean occupation number in the MHIBow06 and GpcBow06 models is less than 1 galaxy per 100 haloes. In the Bow06 model, the HOD

peaks at a halo mass just under $10^{12} h^{-1} M_{\odot}$, with around 1 in 10 such haloes hosting a central galaxy with cold gas mass above the threshold.

The size of the departure from the traditionally assumed step function HOD for central galaxies at $z = 0$ in Fig. 5 varies in proportion to the ‘strength’ of AGN feedback for the Bow06, Font08 and MHIBow06 models (see Table 1). Although the GpcBow06 model has the weakest AGN feedback, the deviation from a step function is largest in this case since this model adopts weaker supernovae feedback than the other models (as a result of being set in a different cosmology, with a lower density fluctuation amplitude). The departure of the central galaxy from a step function form is less pronounced at $z = 1$ (Fig. 6). This is because fewer haloes have quasi-static hot gas haloes and those which do host lower mass SMBH [see Fanidakis et al. (2011) for plots showing how the mass of SMBH is built up over time in the models]. These trends continue in Fig. 7, which shows the HOD for the GALFORM models at $z = 2$. The HOD of central galaxies is now better approximated by a step function. The HODs become noisy for massive haloes as such objects are extremely rare at this redshift. The central galaxy HOD in the Font08 model has a Gaussian form centred on halo masses of a few times $10^{11} h^{-1} M_{\odot}$. The HOD displays an upturn for more massive haloes which is reminiscent of the HOD in the DeLucia07 model. In Font08, this mass could be brought in by merging satellites, which will have a higher cold gas mass than in the other Durham models. The central galaxy HOD becomes closer to the canonical step function form with increasing redshift.

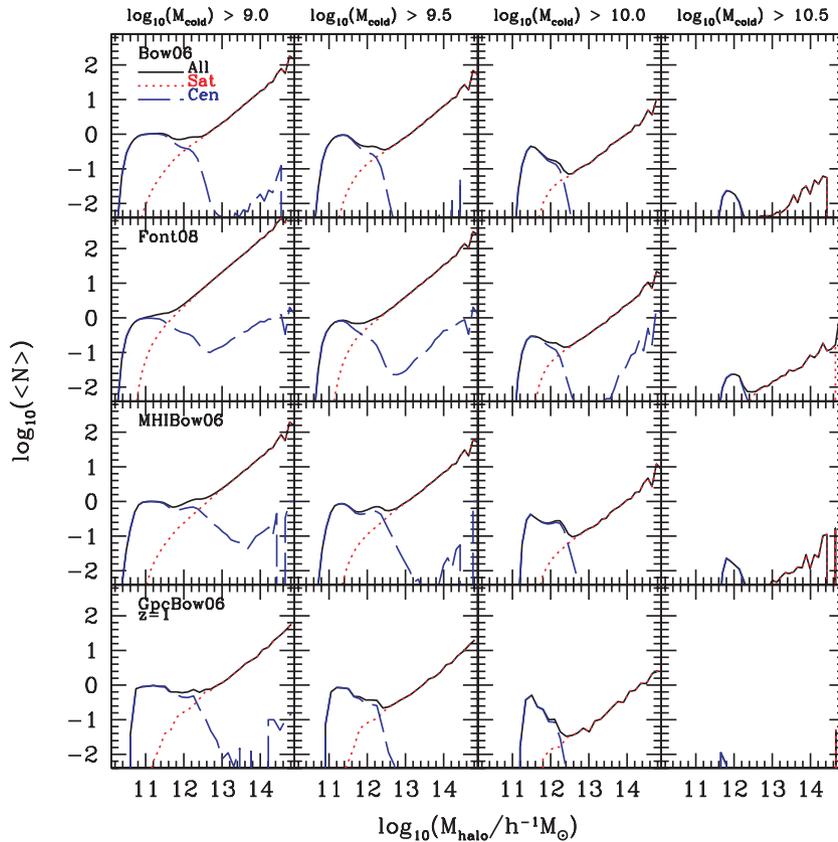


Figure 6. The halo occupation distribution at $z = 1$ for samples defined by a threshold cold gas mass. The blue dashed curves show the contribution from central galaxies, the red dotted curves show satellite galaxies and the black solid curves show all galaxies. Each row shows a different model as labelled, using the notation given in Section 2. Each column corresponds to a different cold gas mass threshold as labelled.

Fig. 5 shows that the amplitude of the HOD for satellite galaxies in the Bow06 model is higher than in the MHIBow06 and GpcBow06 models. This is due in part to the Bow06 model predicting a higher abundance of galaxies by cold gas mass than is observed (see Fig. 2). The Font08 model predicts many more satellite galaxies than the other models (~ 10 times more for the two lowest cold gas mass thresholds). This can be traced back to the modified cooling model in Font08, which means that satellites accrete gas that cools from their incompletely stripped hot haloes. Also, some of the gas which is reheated by supernovae in the satellite is allowed to recool on to the satellite rather than being incorporated into the main hot halo. The amplitude of the HOD for satellite galaxies at $z = 1$ (Fig. 6) in the Bow06, MHIBow06 and GpcBow06 models is higher than predicted at $z = 0$. Star formation depletes the cold gas by $z = 0$. The power-law slope of the satellite HOD is remarkably constant regardless of cold gas mass threshold, redshift or galaxy formation model, with $N_{\text{sat}} \propto M_{\text{halo}}^{0.8}$. The predicted slope is in good agreement with the best fitting value determined from clustering in the HIPASS sample, with Wyithe et al. (2009) reporting a slope of 0.7 ± 0.4 .

3.1.3 Comparing HODs for optical and cold gas mass selection

We next compare the model predictions for the HOD of an optically selected galaxy sample with those of cold gas mass selected samples. Fig. 8 shows the HOD for samples defined by a cold gas mass threshold of $10^{10} h^{-2} M_{\odot}$ in the first four columns, with each column showing the predictions for a different model. In the right-hand

column, we plot the HOD for a sample in which galaxies are selected on the basis of their r -band luminosity in the GpcBow06 model. The optical luminosity cut is chosen such that galaxies brighter than this limit ($M_r - 5 \log h < -21.06$) have the same number density as the sample selected by cold gas mass in the GpcBow06 model. As we have already remarked, the HODs for the cold gas samples have similar properties, with a peaked HOD for central galaxies which declines rapidly with increasing halo mass, and a power-law HOD for satellites. The HOD for central galaxies in the optical sample shows a local bump for halo masses just below $10^{12} h^{-1} M_{\odot}$, but overall rises gradually, reaching unity at a halo mass of $\sim 3 \times 10^{14} h^{-1} M_{\odot}$. The bump is due to the implementation of AGN feedback. The central galaxy HOD drops after the bump as AGN feedback ‘switches on’ in these haloes. Central galaxies hosted by massive haloes are bright in the r -band, whilst possessing too little gas to be included in the cold gas sample.

The remaining rows of Fig. 8 show the steps which connect the HOD predictions to the effective bias of the galaxy samples, which tells us the clustering amplitude. In the lower two rows of this plot we have switched to plotting quantities on a linear scale. In the second row of Fig. 8, the HOD is multiplied by the abundance of the host dark matter haloes, giving the contribution to the number density of galaxies as a function of halo mass. The abundance of the host dark matter haloes is computed using the prescription of Sheth et al. (2001), which gives a good match to simulation results. Beyond the break in the mass function, the number of haloes per unit volume drops exponentially. This means that satellite galaxies, whose HOD is described by a moderate power law, do not

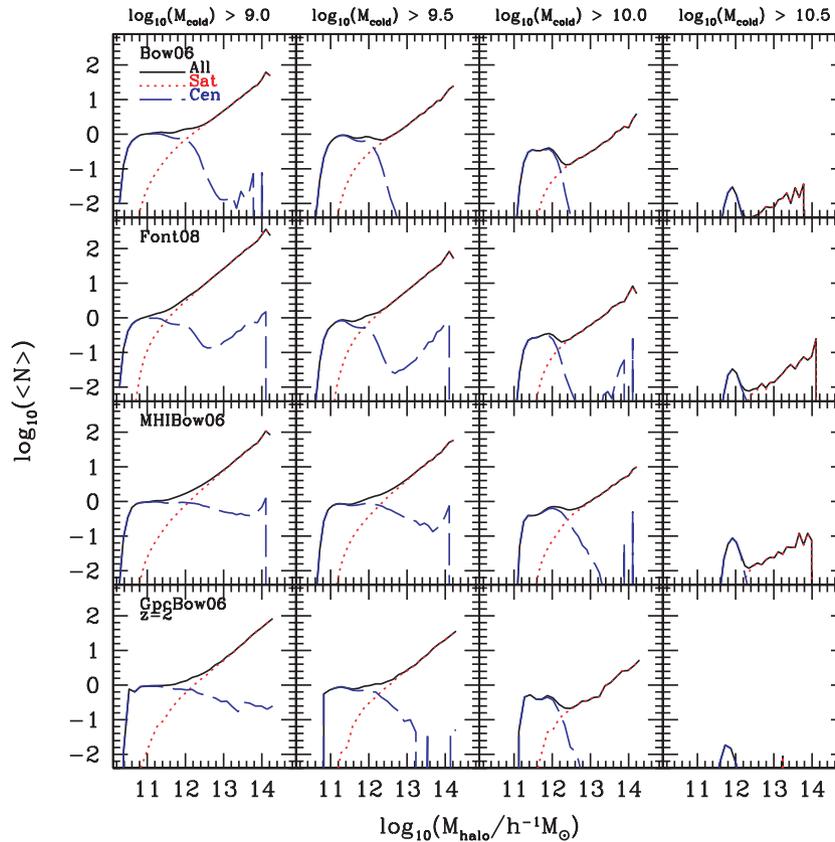


Figure 7. The halo occupation distribution at $z = 2$. As before, the blue dashed curves show the contribution from central galaxies, the red dotted curves show satellite galaxies and the black solid curves show all galaxies. Each row shows a different model as described in Section 2. Each column corresponds to a different cold gas mass threshold as labelled.

contribute significantly to the number of galaxies per unit volume. This is true of samples defined either by cold gas mass or r -band luminosity. The abundance of galaxies is sharply peaked for the cold gas samples. For the optical sample, the galaxy number density has a sharp peak just below a halo mass of $10^{12} h^{-1} M_{\odot}$ and then shows a broad distribution and an appreciable contribution from more massive haloes. In the bottom row of Fig. 8, we plot the number density of galaxies multiplied by the bias factor as a function of halo mass (as computed using the prescription of Sheth et al. 2001). The square of the bias gives the factor by which the autocorrelation function of haloes is boosted on large scales relative to the correlation function of the dark matter. The halo bias increases rapidly beyond the break in the mass function, which increases the influence of satellite galaxies on the effective bias (e.g. Angulo, Baugh & Lacey 2008b). Nevertheless, for the cold gas mass samples satellite galaxies still make a negligible contribution to the clustering amplitude on large scales, as quantified by the effective bias. Satellite galaxies make a modest contribution to the effective bias in the r -band sample. This contribution increases if the luminosity cut is made fainter. In summary, the models predict that galaxies with cold gas mass in excess of $10^{10} h^{-2} M_{\odot}$ are predominately central galaxies hosted by dark matter haloes of mass $10^{12} h^{-1} M_{\odot}$. These haloes are less massive than the characteristic halo mass at $z = 0$ in the cosmologies used and so the bias factor of these samples is below unity; they are subclustered compared to the dark matter. In contrast, the r -band sample has an effective bias with a significant contribution from more massive haloes which have a larger bias factor. The bias factor for the r -band selected samples is therefore greater than unity

and clustering length is larger than it is for cold gas sample (see Fig. 10 later).

Finally, in Fig. 9 we compare the spatial distribution of r -band selected galaxies with that of galaxies chosen on the basis of their cold gas mass ($M_{\text{cold}} > 10^{10} h^{-2} M_{\odot}$ in the GpcBow06 model). Again the r -band magnitude limit ($M_r - 5 \log h < -21.06$) is chosen to match the abundance of galaxies in the cold gas sample. The grey circles represent dark matter haloes. The circle radius and darkness are proportional to halo mass. The cold gas selected galaxies follow the filamentary structure and tend to avoid high-density regions. The difference in the number of satellite galaxies (red circles) is obvious between the cold gas and optical samples. The satellites are found in more massive haloes. This difference in the spatial distributions provides a visual impression of the differences in the HODs plotted in Fig. 8. The stronger clustering of the optical samples in principle means that it should be easier to measure the power spectrum of galaxy clustering using these tracers. However, the key consideration, as we will see in Section 4, is how the product of the number density of galaxies and their power spectrum amplitude changes with redshift. This quantity controls the ‘contrast’ of the power spectrum signal against the noise which arises from having discrete tracers of the density field.

3.2 Predictions for the clustering of cold gas

In this section we present the predictions of the galaxy formation models introduced in Section 2 for the two-point correlation

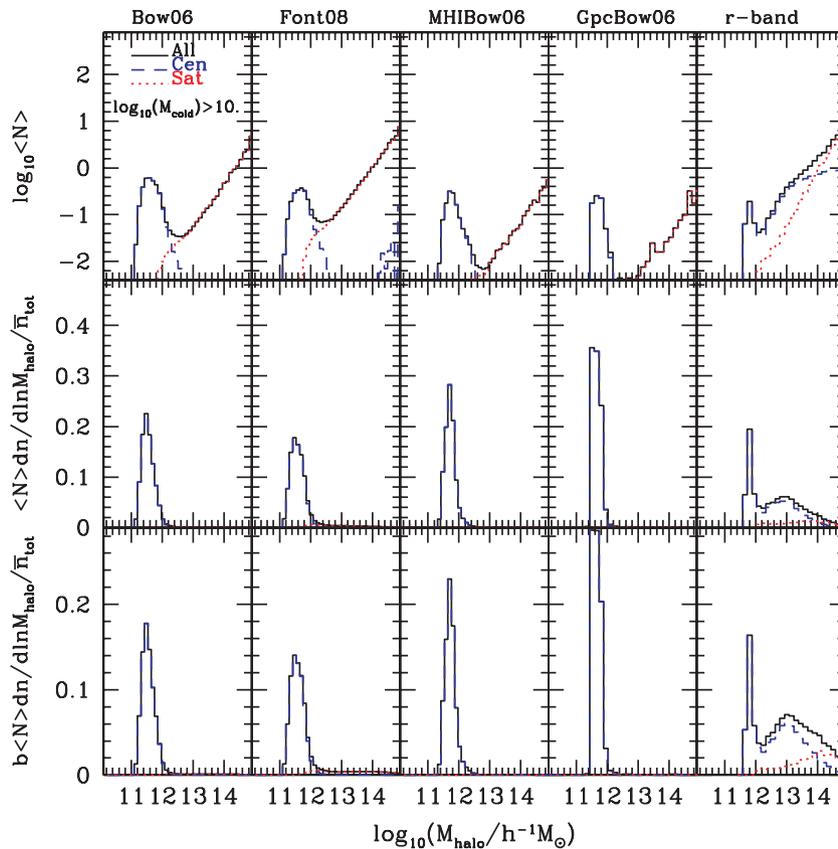


Figure 8. The steps relating the number of galaxies per halo to the strength of galaxy clustering shown using the GALFORM models. The first row shows the HOD as a function of halo mass. The second row shows the HOD multiplied by the abundance of dark matter haloes as a function of halo mass, $dn/d \ln M_{\text{halo}}$, (as computed using the prescription of Sheth, Mo & Tormen 2001) with the y-axis plotted on a linear scale. \bar{n}_{tot} is the number density of galaxies which satisfy the selection criteria (i.e. in cold gas mass for columns 1–4 or r -band luminosity in column 5). The integral of these curves is proportional to the number density of galaxies. The bottom row shows the HOD times the halo mass function times the bias factor as a function of halo mass. The area under the curves in this case gives the effective bias of the galaxy sample. The first four columns show the model predictions for galaxies with cold gas mass in excess of $M_{\text{cold}} > 10^{10} h^{-2} M_{\odot}$. The fifth column shows an r -band selected sample in the GpcBow06 model, with the magnitude limit ($M_r - 5 \log h < -21.06$) chosen such that the number of galaxies matches that in the cold gas sample in this model. As before, the contribution of central galaxies is shown by blue dashed lines, satellite galaxies by red dotted lines and all galaxies by black solid lines.

function. To predict the galaxy distribution of the GpcBow06 model, we generated galaxy samples using the GPICC simulation.

We start in Fig. 10 by comparing the spatial two-point autocorrelation function of a galaxy sample defined by a threshold cold gas mass ($M_{\text{cold}} > 10^{10} h^{-2} M_{\odot}$) in real (solid black line) and redshift space (dashed black line). The correlation function is computed in redshift space using the distant observer approximation. In this case, one of the coordinate axes is chosen as the line of sight and the peculiar velocity of the galaxy in that direction is added to the real-space position along the same axis, after applying a suitable scaling to convert from velocity units to distance units. For the largest pair separations plotted, the correlation function in redshift space has the same shape as the real-space correlation function, but a larger amplitude. The magnitude of the shift in amplitude agrees approximately with the expectation of Kaiser (1987). This effect is caused by coherent bulk flows towards overdense regions. On pair separations between 0.3 and $1 h^{-1}$ Mpc, the real and redshift space correlation functions are very similar. They diverge on smaller scales, where the predictions are noisy simply because there are few galaxy pairs at these separations.

This behaviour can be contrasted with the clustering in the optically selected sample, which is shown by the red lines in Fig. 10. As with the cold gas sample, there is a shift in the clustering amplitude

when measured in redshift space for pair separations $r > 3 h^{-1}$ Mpc. However, the size of the shift is smaller for the optically selected sample, which is consistent with the bias of this sample being greater than unity and larger than the bias of the cold gas selected sample. The real-space correlation function of the optical sample is steep on small scales, reflecting the contribution of satellite galaxies within common dark matter haloes. There is a substantial reduction in the clustering amplitude in redshift space on these scales in the optical sample, again driven by satellite galaxies. This is the so-called ‘fingers of God’ redshift space distortion, whereby randomized peculiar velocities of the satellites within the gravitational potential of the cluster make the cluster appear elongated.

The real-space correlation function cannot be estimated directly from a galaxy redshift survey. A related quantity is the projected correlation function which can be obtained from the two-point correlation function measured in bins of pair separation parallel (π) and perpendicular (σ) to the line of sight, $\xi(\sigma, \pi)$ (e.g. Norberg et al. 2001):

$$\frac{\Xi(\sigma)}{\sigma} = \frac{2}{\sigma} \int_0^{\infty} \xi(\sigma, \pi) d\pi. \quad (5)$$

In the limit that the integral over the radial pair separation is taken to infinity, this quantity is free from redshift space distortions (see

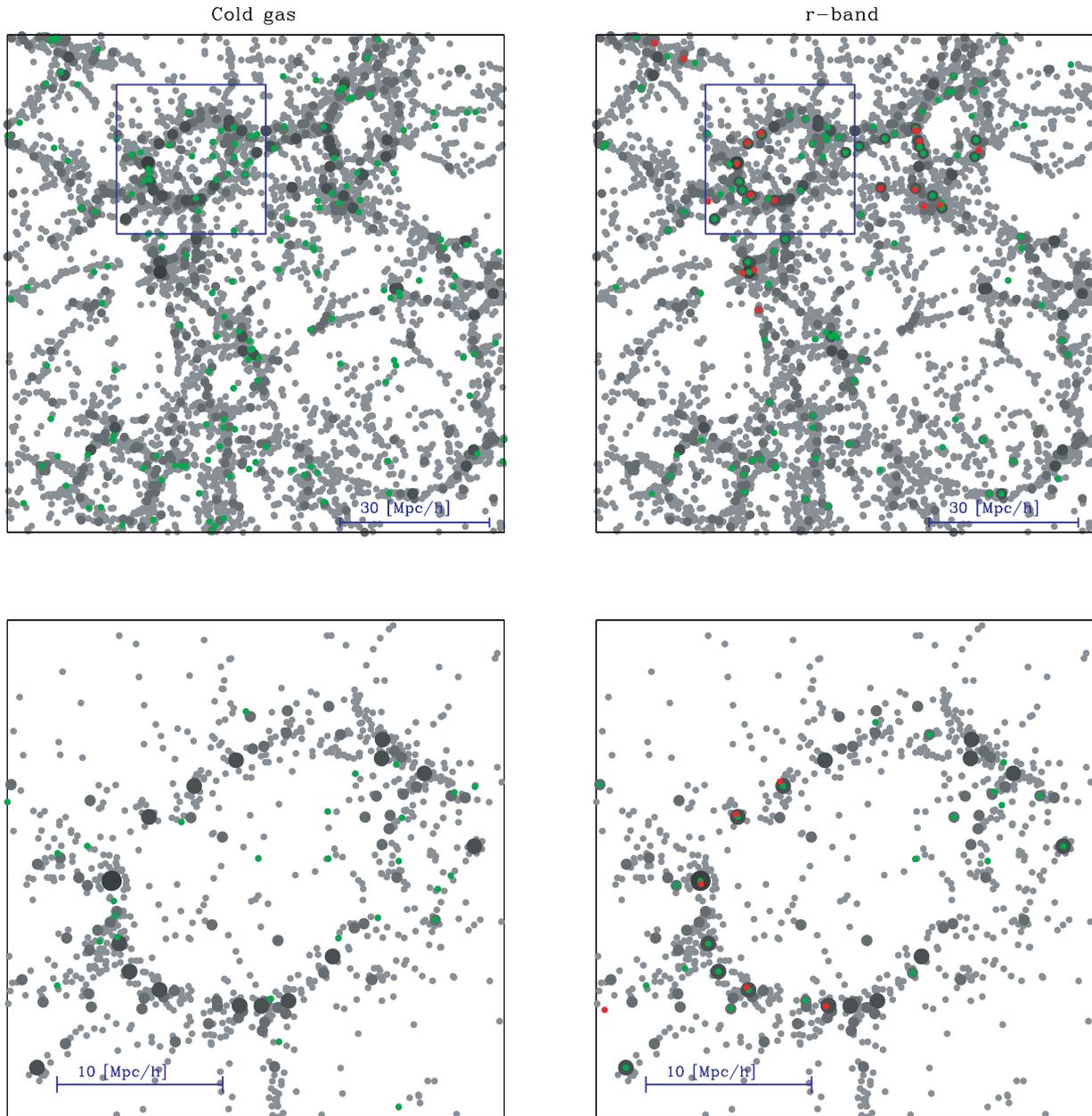


Figure 9. The spatial distribution of galaxies and dark matter haloes in the GpcBow06 model at $z = 0$. Dark matter is shown in grey and the size and darkness of the circle used to plot the dark matter halo increase with mass. Galaxies selected by cold gas mass ($M_{\text{cold}} > 10^{10} h^{-2} M_{\odot}$) and r -band luminosity ($M_r - 5 \log h < -21.06$) are plotted in the left- and right-hand panels, respectively. The top row shows a slice of $100 h^{-1}$ Mpc on a side and $10 h^{-1}$ Mpc thick. The bottom row shows a zoom into a region of $30 h^{-1}$ Mpc on a side and $10 h^{-1}$ Mpc thick, which corresponds to the blue square in the top row. The green circles represent central galaxies and the red circles show satellite galaxies.

Norberg et al. 2009 for an illustration of the impact of imposing a finite upper limit on the integral).

Fig. 11 shows the projected correlation function predicted in the four models for a range of cold gas mass samples at $z = 0, 1$ and 2. The columns show the results for different cold gas mass thresholds, and the rows correspond to different models. The solid black lines in each panel show the projected correlation function measured for the dark matter in the Millennium Simulation (recall that the GpcBow06 model has a different cosmology and so should be compared to a consistent dark matter correlation function which will be slightly different from that in the Millennium simulation on these scales). Overall, the three lowest mass samples at $z = 0$ are less

clustered than the dark matter. The most massive threshold sample we consider at this redshift has a similar clustering amplitude to the dark matter. At $z = 1$, the bias of the three lowest mass samples is close to unity, with the projected clustering of galaxies being very close to that of the dark matter. At $z = 2$, the cold gas samples are more clustered than the dark matter and correspondingly have effective biases greater than unity. This evolution in the bias is due to the adoption of a fixed cold gas mass threshold. At high redshift, galaxies with a large cold gas mass will tend to be found in more massive haloes.

Across the different models there is a small spread in clustering amplitude for a given cold gas mass sample, with remarkably similar

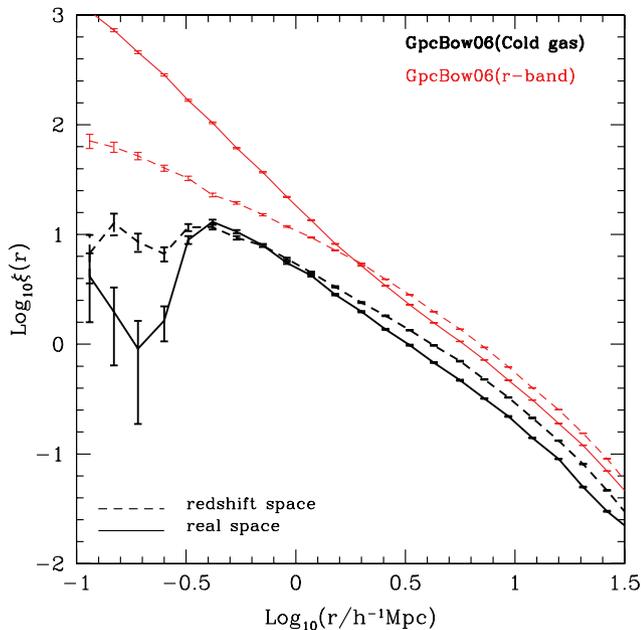


Figure 10. The real space (solid) and redshift space (dashed) correlation function predicted for galaxies in the GpcBow06 model at $z = 0$. The black lines show the correlation function of galaxies with cold gas mass $M_{\text{cold}} > 10^{10} h^{-2} M_{\odot}$ and the red lines show the clustering of galaxies selected to be brighter than a threshold r -band luminosity, chosen to match the abundance of galaxies in the cold gas sample. The error bars show the Poisson error on the pair count in each bin of radial separation.

predictions made for the projected correlation function. Fig. 11 shows that the differences start to appear at $z = 1$ and become larger by $z = 0$. The model which shows the largest difference from the others is Font08. On small scales in the two lowest mass threshold samples, this model has an appreciably higher amplitude projected correlation function than the other models. This feature can be traced back to the HODs plotted in Fig. 5. Due to the revised cooling model used in Font08, this model has more satellite galaxies in the low-mass samples, which boosts the one halo term in the correlation function.

Finally, we compare the predicted correlation functions with an observational estimate from Meyer et al. (2007), which was made using the H I Parkes All-Sky Survey (HIPASS) Catalogue (HICAT; Meyer et al. 2004). In order to make this comparison, we need to convert the cold gas mass output by the models into an atomic hydrogen mass. We assume that 76 per cent by mass of the cold gas is hydrogen. Here we adopt a fixed ratio of molecular (H_2) to atomic (H I) hydrogen of $\text{H}_2/\text{H I} = 0.4$ [see Power et al. (2010) for a discussion]. The H I mass, $M_{\text{H I}}$, is therefore obtained from the cold gas mass M_{cold} by applying the conversion

$$M_{\text{H I}} = 0.76 M_{\text{cold}} / (1 + 0.4). \quad (6)$$

With this relation, the sample analysed by Meyer et al. is equivalent to a cold gas mass threshold of $M_{\text{cold}} > 10^{9.5} h^{-2} M_{\odot}$. The comparison between the model predictions and the observational estimate is presented in Fig. 12. The correlation function predicted by the MHIBow06 model agrees remarkably well with the observational estimate. The GpcBow06 and Bow06 models predict too low a clustering amplitude. The Font08 model gives a reasonable match on intermediate and large scales, but somewhat overpredicts the clustering amplitude on small scales, hinting that this model has too many gas-rich satellites in massive haloes.

4 MEASURING DARK ENERGY WITH FUTURE H I REDSHIFT SURVEYS

In this section we show how redshift surveys of H I-selected galaxies can be used to detect BAOs in the galaxy power spectrum (Section 4.1), and we assess the relative performance of future H I and optical/near-infrared redshift surveys in measuring the large-scale structure of the Universe, first by contrasting the effective volumes sampled by different survey configurations (Section 4.2) and then by applying a Fisher matrix calculation to compare the error forecast on the dark energy equation of state parameter for different surveys (Section 4.3).

4.1 The appearance of baryonic acoustic oscillations

The BAO signal measured from the power spectrum estimated using a sample defined by cold gas mass is shown in Fig. 13. We use the galaxy distribution in the GPICC simulation generated using the GpcBow06 model. The GPICC simulation box is $1 h^{-1} \text{Gpc}$ on a side which allows the evolution of the BAO to be modelled accurately. To show the BAO more clearly, we have divided the measured power spectrum by a smooth reference power spectrum which contains no BAO features. This ‘no wiggle’ power spectrum is defined differently for the linear theory power spectrum and for the spectra measured from the GPICC simulation. For the linear theory prediction, which is shown by the curves in Fig. 13, the reference is based on the no-wiggle parametrization of the power spectrum given by Eisenstein & Hu (1998). The no-wiggle prediction includes the impact of a non-zero baryon component on the width of the turnover in the matter power spectrum, but does not contain BAO. The ratio of the linear theory power spectrum, $P^{\text{L}}(k)$, to the no-wiggle prediction, P_{nw}^{L} , is smoothed by damping the oscillations to represent the impact of non-linear growth and redshift-space distortions (e.g. Eisenstein et al. 2005; Sanchez, Baugh & Angulo 2008):

$$R_{\text{lin}}(k) = \left(\frac{P^{\text{L}}}{P_{\text{nw}}^{\text{L}}} - 1 \right) \times \exp \left(- \frac{k^2}{2k_{\text{nl}}^2} \right) + 1, \quad (7)$$

where k_{nl} is the damping scale and is treated as a free parameter.

The overall shape of the no-wiggle power spectra measured from the simulation is different from the linear theory prediction due to the non-linear growth of fluctuations and redshift-space distortions [see Angulo et al. (2008a) for a step-by-step illustration of these effects]. We model this change in shape by multiplying the no-wiggle version of the linear theory spectrum by a third-order polynomial

$$P_{\text{g}}(k) = (1 + Ak + Bk^2 + Ck^3) P_{\text{nw}}^{\text{L}}(k). \quad (8)$$

The free parameters A , B and C are chosen to give the best match to the overall shape of the measured power spectrum. All points up to $k = 0.4 h \text{Mpc}^{-1}$ were included in the fit and given equal weight. This approach is more straightforward and robust than using a spline fit to a coarsely binned measured spectrum, which is sensitive to the number of k -bins used.

We show in Fig. 13 the BAO signal in the GpcBow06 model at $z = 0, 1$ and 2 for galaxies selected by their cold gas mass ($M_{\text{cold}} > 10^{10} h^{-1} M_{\odot}$; left-hand column), as an illustration of how a cold gas mass selected sample traces this large-scale structure feature. In the right-hand column of Fig. 13, we compare this with the BAO signal expected for an r -band selected sample of galaxies that have same number density at each redshift as the cold gas sample. The reference power spectrum is defined as described above, using the third-order polynomial fit to the measured spectrum in each case. Fig. 13 shows that we should be able to measure the BAO

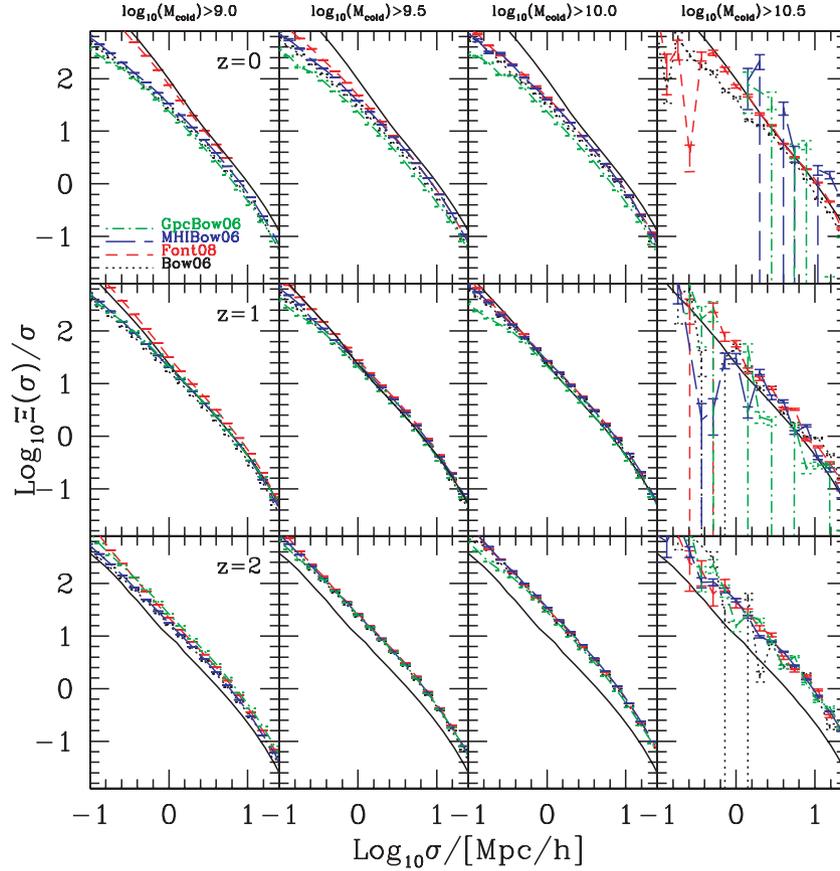


Figure 11. The projected correlation function for cold gas mass selected samples at $z = 0$ (top), 1 (middle) and 2 (bottom), plotted as a function of pair separation transverse to the line of sight, σ . Each column shows the predictions for a different cold gas mass threshold, as indicated by the label. The predictions of the models are distinguished by different line types and colours, as shown by the key in the upper-left panel. The solid black lines in each panel show the projected correlation function of the dark matter measured in the Millennium simulation (note that the GpcBow06 model uses a different cosmology and has different dark matter correlation functions).

feature just as well using a sample selected by cold gas mass as with an optically selected sample with the same number density.

4.2 The effective volumes of different survey configurations

Many ongoing and proposed redshift surveys have the goal of determining the nature of dark energy by measuring the BAO signal in the galaxy power spectrum. A powerful way to compare the expected performance of different surveys for measuring large-scale structure is to estimate their effective volumes (see, for example, Orsi et al. 2009). This is essentially an indicator of the ‘useful’ survey volume which determines the size of the error bar on the measured power spectrum. The effective volume is defined as (Feldman, Kaiser & Peacock 1994)

$$V_{\text{eff}}(k, z) = \int_{z_{\text{min}}}^{z_{\text{max}}} \left[\frac{\bar{n}(z) P_g(k, z)}{1 + \bar{n}(z) P_g(k, z)} \right]^2 \frac{dV}{dz}, \quad (9)$$

where all quantities are expressed in comoving coordinates and dV/dz is the differential comoving volume. To calculate the effective volume, we therefore need to know the number density of galaxies ($\bar{n}(z)$) down to a given survey flux limit and the effective bias [$b(z)$], both as functions of redshift. In this calculation, we obtain the galaxy power spectrum using the linear relation involving galaxy bias and the dark matter power spectrum: $P_g(k, z) = P_{\text{dm}}(k, z=0) b^2(z) D^2(z)$, where P_g is the galaxy power spectrum, $P_{\text{dm}}(k, z=0)$ is the linear

theory dark matter power spectrum at $z = 0$, $b(z)$ is the effective bias, and $D(z)$ is the growth factor of the dark matter.

To make predictions for the effective volume of the SKA, we need to convert the cold gas mass predicted by the models into an H I line flux, which we do following the prescription set out in Power et al. (2010). A key step is the assumption about the fraction of neutral hydrogen which is in molecular form as opposed to atomic hydrogen. Power et al. (2010) adopted two prescriptions: a fixed fraction of 40 per cent as used by Baugh et al. (2005) and a variable fraction, as used by Obreschkow & Rawlings (2009a). In the latter case, the $\text{H}_2/\text{H I}$ ratio varies from galaxy to galaxy as a power law in gas pressure, as found observationally by Blitz & Rosolowsky (2006). We shall refer to these two scenarios as the fixed and variable $\text{H}_2/\text{H I}$ ratio cases. Power et al. (2010) showed that the high-redshift tail of the redshift distribution in the variable $\text{H}_2/\text{H I}$ case is substantially suppressed compared with the fixed $\text{H}_2/\text{H I}$ ratio case.

Following Power et al. (2010) the observed flux from a galaxy at 21 cm is given by

$$S_{\text{obs}} = \frac{3}{16\pi} \frac{hc A_{12}}{m_{\text{H}}} M_{\text{H I}} \frac{1}{D_L^2(z)} \frac{1}{\Delta V_{\text{los}}} (1+z), \quad (10)$$

where A_{12} is the Einstein coefficient which gives the spontaneous rate of transition from the upper to lower hyperfine levels of the hydrogen ground state, m_{H} is the mass of the hydrogen atom, $M_{\text{H I}}$ is the mass of H I in the galaxy, $D_L(z)$ is the luminosity distance to

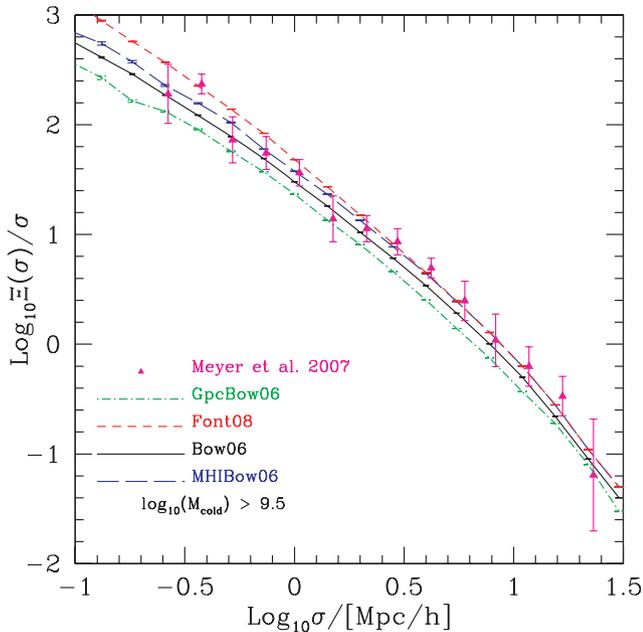


Figure 12. The projected galaxy correlation function at $z = 0$. The points with error bars show an observational estimate made from the HIPASS catalogue by Meyer et al. (2007). The lines show the model predictions for galaxies more massive than $M_{\text{cold}} > 10^{9.5} h^{-2} M_{\odot}$, a threshold chosen to match the selection of galaxies in the HIPASS sample. The results for different models are shown by lines with different colours and line types, as indicated by the key.

redshift z and ΔV_{los} is the rest-frame line-of-sight velocity width of the galaxy. We assume that the linewidth of the 21-cm emission is sampled by many channels and these are combined to yield the flux integrated over the full rotationally broadened width of the line. The model discs are assumed to have random inclinations. The sensitivity of a dual polarization radio receiver can be written in terms of a limiting rms flux as

$$\frac{S_{\text{rms}}}{1.626 \mu\text{Jy}} = \left(\frac{A_{\text{eff}}}{\text{km}^2} \right)^{-1} \left(\frac{T_{\text{sys}}}{50 \text{ K}} \right) \left(\frac{\Delta \nu_{\text{rec}}}{\text{MHz}} \right)^{-1/2} \left(\frac{\tau}{\text{h}} \right)^{-1/2}, \quad (11)$$

where A_{eff} is the effective area of the telescope, T_{sys} is the system temperature which is a measure of the instrument noise, $\Delta \nu_{\text{rec}}$ is the frequency bandwidth and τ is the integration time. By combining equations (10) and (11) we can determine whether or not a given model galaxy would be detectable by a survey.

We evaluate the performance of several survey configurations which are intended to span the range of possibilities currently under discussion for future H I galaxy redshift surveys (see, for example, Abdalla et al. 2010; Faulkner et al. 2010). The variables considered are (i) the sensitivity of the telescope, as quantified by the ratio of the effective area to the system temperature, $A_{\text{eff}}/T_{\text{sys}}$; (ii) the field of view (FoV) of the telescope; and (iii) the scaling, if any, of the integration time with redshift.

We consider two values of $A_{\text{eff}}/T_{\text{sys}}$, one appropriate to ASKAP, $A_{\text{eff}}/T_{\text{sys}} = 82 \text{ m}^2 \text{ K}^{-1}$ (Johnston et al. 2008) and the other for the SKA, $A_{\text{eff}}/T_{\text{sys}} = 20\,000 \text{ m}^2 \text{ K}^{-1}$, which for $A_{\text{eff}} = 1 \text{ km}^2$ corresponds to a system temperature of $T_{\text{sys}} = 50 \text{ K}$. The FoV of a single pointing of a telescope determines the integration time for a given point on the sky, once the total solid angle and the total time available to conduct the survey are specified. We assume that all surveys cover $20\,000 \text{ deg}^2$ in 1 yr. A larger FoV therefore corresponds to a longer net integration time at each location within the survey area.

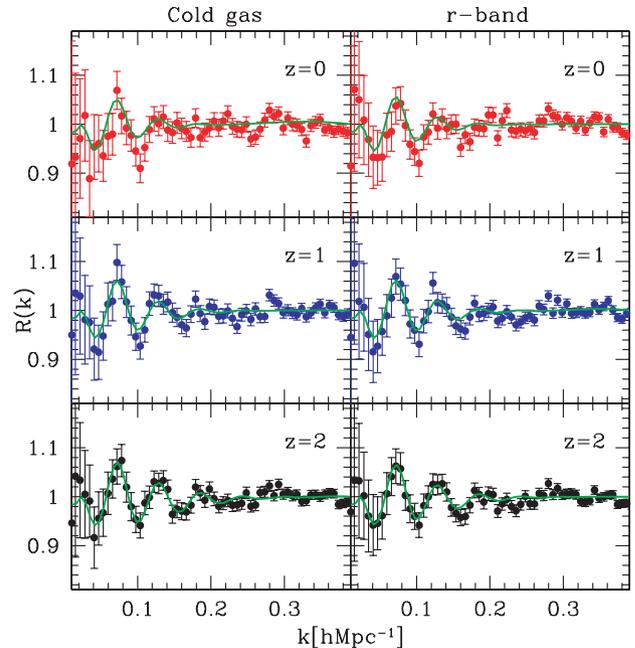


Figure 13. The baryonic acoustic oscillations in the galaxy power spectrum. To display the BAO more clearly, we have divided the predicted spectra by smooth fits, as described in the text. The points show the power spectra predicted by the GpcBow06 model at $z = 2$ (bottom), $z = 1$ (middle) and $z = 0$ (top). The left-hand column shows the power spectra measured for galaxies with cold gas mass ($M_{\text{cold}} > 10^{10} h^{-1} M_{\odot}$). The right-hand column shows the BAO in a sample selected in the r -band with the same number density of galaxies as the cold gas sample at that redshift. The smooth green line shows the linear theory power spectrum, divided by a smooth reference power spectrum, after filtering to damp the higher harmonics (see text). The errors plotted on the power spectrum depend on the number density of galaxies and the simulation volume [see equation (3) in Angulo et al. 2008a].

For ASKAP, we adopt a FoV of 30 deg^2 . For the SKA we consider two cases: FoV = 1 and 100 deg^2 . As remarked upon above, the nominal integration time per observation is controlled by the FoV, the total survey area and the total duration allowed for the survey. Radio telescopes are tuned to operate at different frequencies in order to pick up the rest-frame 21-cm line emission from galaxies at different redshifts. In the case of a single dish with a single detector element, the FoV is set by the diffraction limit c and scales as ν^{-2} . The shift to lower observed frequencies to pick up the redshifted 21-cm emission from $z > 0$ galaxies leads to an increase in the FoV. A survey strategy which tiles the designated solid angle uniformly without overlap between pointings at the rest-frame frequency of the 21-cm line will result in an ‘oversampling’ of the solid angle at $z > 0$. This can be translated into an increase in the effective integration time, which then scales with redshift as $(1+z)^2$ (see Abdalla & Rawlings 2005). The exact scaling of the effective integration time with redshift depends on the design of the telescope (e.g. dishes with focal plane arrays or aperture arrays), the survey strategy and the speed of the data communications and the processing of the signals. Faulkner et al. (2010) argue that aperture arrays can potentially yield survey speed which are limited only by ‘software’ – the signal processing speed – rather than the actual detection hardware. One obstacle which may stand in the way of realizing such survey speeds is the increase in sky noise below a frequency of $\approx 450 \text{ MHz}$, which we have not taken into account. This would affect the detectability of galaxies at $z > 2$; in any case, Figs 14 and 15 show that none

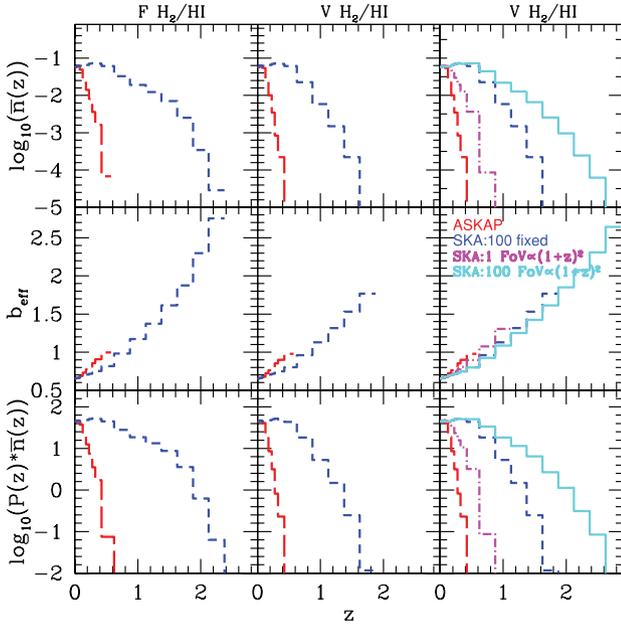


Figure 14. The quantities used to compute the effective volume of a redshift survey, as predicted by the GpcBow06 model. We show predictions for the four survey configurations listed in the text and labelled in the middle-right panel. The top row shows the number density of galaxies as a function of redshift, the middle row the effective bias of the galaxy samples and the bottom row shows the product of the galaxy number density and galaxy power spectrum, $P_{\text{gal}}(k)$, at $k = 0.2 h \text{Mpc}^{-1}$. Note that selected curves are plotted in each column to make different comparisons between the predictions. The first column shows the predictions for the ASKAP and SKA: 100 fixed configurations in the case that the ratio $\text{H}_2/\text{H I}$ is held fixed. The second and third columns show how the predictions change when this ratio is allowed to vary from galaxy to galaxy and with redshift. In the final column, we compare all of the telescope configurations, including two cases in which the effective integration time scales with redshift as $(1+z)^2$ (see text for details).

of the survey configurations considered is effective in this redshift range. We therefore consider SKA scenarios in which the effective integration time scales as $(1+z)^2$, in addition to integration times that are fixed and independent of redshift. In summary, the four cases we examine are

- (i) FoV = 30 deg^2 , $A_{\text{eff}}/T_{\text{sys}} = 82 \text{ m}^2 \text{ K}^{-1}$, fixed, redshift-independent integration time, which we label as ASKAP in Figs 14 and 15 and Table 2.
- (ii) FoV = 100 deg^2 , $A_{\text{eff}}/T_{\text{sys}} = 20\,000 \text{ m}^2 \text{ K}^{-1}$, fixed integration time, labelled SKA:100 FoV fixed.
- (iii) FoV = 1 deg^2 , $A_{\text{eff}}/T_{\text{sys}} = 20\,000 \text{ m}^2 \text{ K}^{-1}$, effective integration time scaling as $(1+z)^2$, labelled SKA:100 FoV $\propto (1+z)^2$.
- (iv) FoV = 100 deg^2 , $A_{\text{eff}}/T_{\text{sys}} = 20\,000 \text{ m}^2 \text{ K}^{-1}$, effective integration time scaling as $(1+z)^2$, labelled SKA:100 FoV $\propto (1+z)^2$.

Fig. 14 shows the model predictions for the quantities needed to calculate the effective volume for various telescope configurations, calculated using the GpcBow06 model. The predictions are divided into columns to show the impact of different assumptions. A comparison of the first and second columns of Fig. 14 shows how the assumption about the $\text{H}_2/\text{H I}$ ratio affects the predictions: the first column shows the predictions for a fixed $\text{H}_2/\text{H I}$ ratio and the second column shows the output for a variable ratio. For $z > 1$ the number density of sources drops more steeply with the variable conversion factor than with the fixed one. The third col-

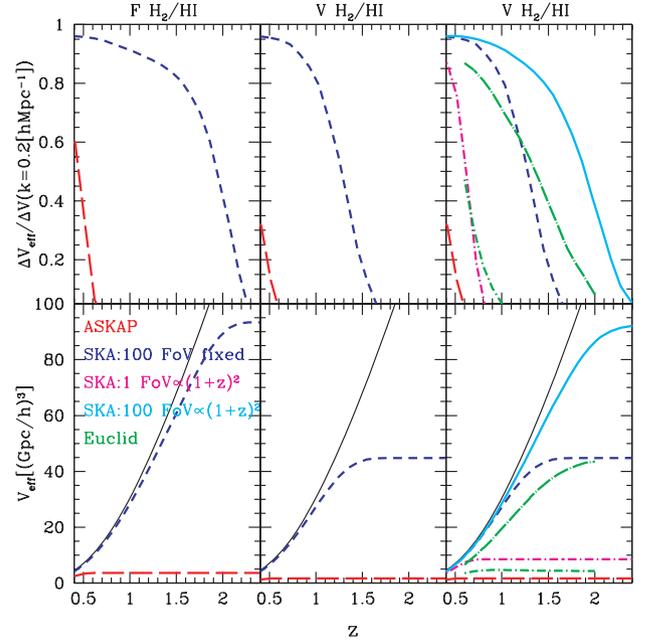


Figure 15. The effective volume per hemisphere of H I-selected samples predicted by the GpcBow06 model. The upper panel shows the differential effective volume divided by the geometrical volume for narrow bins in redshift. The lower panel shows the cumulative volume out to a given redshift. The results for different telescope configurations are shown by different line styles and colours as indicated by the key. The green curves show the predictions for a spectroscopic survey down to $H = 22$ (33 per cent redshift success rate; green dot-long-dashed line) and a slitless survey down to an $\text{H}\alpha$ flux limit of $5 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ (green dot-short-dashed line), as computed by Orsi et al. (2010). The black solid line in the lower panel shows the available geometrical volume per hemisphere. The optical and H I surveys are assumed to cover the same solid angle, one hemisphere.

umn shows the predictions with a variable $\text{H}_2/\text{H I}$ ratio and contrasts predictions with and without a dependence of the effective integration time on redshift. When the integration time scales as $(1+z)^2$ (thick cyan line), the number density of sources declines more slowly with redshift. The number density of galaxies in the ASKAP configuration falls rapidly with redshift (red lines). The decline in the abundance of ASKAP galaxies is slightly more modest in the case of a fixed $\text{H}_2/\text{H I}$ ratio than it is with a variable ratio. In the case of the SKA:100 FoV fixed configuration, the number density of galaxies drops by around a factor of 5 between $z = 0$ and $z = 1$ in the fixed $\text{H}_2/\text{H I}$ ratio case. By $z \sim 2$, the number density of galaxies is over 100 times lower than it is at $z = 0$. The middle panels of Fig. 14 show that the bias changes much more modestly than the number density of galaxies does, increasing by a factor of 4 with redshift over the entire redshift range plotted. The increase in effective bias cannot therefore compensate for the dramatic drop in the abundance of galaxies in the high-redshift tails of the distributions plotted in the upper panels of Fig. 14. The effective volume of a survey configuration no longer increases with redshift once the product of the galaxy number density and the galaxy power spectrum drops below unity. In this regime, the power spectrum signal is swamped by shot noise ($P_{\text{shot}} = 1/\bar{n}$) arising from the use of discrete galaxies to trace the continuous density field, and the volume added in this regime does not contribute significantly to the statistical power of the survey. The product $\bar{n}P$ is plotted in the lower panels of Fig. 14.

In the lower-right panel of Fig. 14, we include two predictions for redshift surveys in the near-infrared taken from Orsi et al. (2009),

Table 2. The forecast constraints on a constant dark energy equation of state, w , when marginalizing over Ω_m for effective survey volumes corresponding to the different assumptions about the H_2/H_1 ratio and the survey FoV. The numbers shown are ratios of $\sigma(w)$, the 1σ error on w , forecast relative to those obtained for the SKA:100 fixed case (FoV 100deg², integration time independent of redshift) with a variable H_2/H_1 ratio. The first column gives the name of the survey configuration, the second column gives the assumption made about the H_2/H_1 ratio, the third column gives the assumption about the scaling of the effective integration with redshift (note that columns 2 and 3 do not apply to the near-infrared redshift surveys listed in the bottom two rows of the table) and columns four and five gives the relative error on w for *WMAP* and *Planck* priors, respectively.

Survey	H_2/H_1	t_{eff}	<i>WMAP</i> prior	<i>Planck</i> prior
ASKAP	fixed	fixed	10.3	3.7
ASKAP	variable	fixed	14.8	3.8
SKA:100	fixed	fixed	0.6	0.6
SKA:100	variable	fixed	1.0	1.0
SKA:100	fixed	$\propto (1+z)^2$	0.4	0.4
SKA:100	variable	$\propto (1+z)^2$	0.6	0.6
H α	-	-	2.6	2.2
$H = 22$	-	-	0.8	0.8

who followed the same procedure we have set out above, but for different galaxy selection criteria. The predictions for a redshift survey to $H = 22$ with a 33 per cent redshift sampling rate (green dot-long-dashed line) and for a slitless survey to an H α flux limit of 5×10^{-16} erg s⁻¹ cm⁻², again with a 33 per cent redshift measurement rate (green dot-short-dashed line), are plotted for comparison. In both cases the surveys cover 20 000 deg².

Our predictions can be compared with those of Abdalla et al. (2010). These authors adopt a slightly different definition of the rms flux limit, based on a study of how the sky background, and hence T_{sys} varies with frequency (Abdalla, private communication). Also, a different dish configuration was adopted, with a correspondingly different constant in equation (11). Taken together, this means that our predictions for a 10σ flux limit should be compared with the 5σ predictions of Abdalla et al. These authors also assume that the velocity width which appears in the definition of the flux limit (see their equation 3) corresponds to a channel width (typically $\Delta V = 30$ km s⁻¹), whereas we assume the full width of the line emission is sampled by many channels. Abdalla et al. make empirical assumptions about the evolution of the H I mass function with redshift. The predictions of our semi-analytical model with the variable H_2/H_1 ratio are similar to their model C. Lastly, Abdalla et al. assume an effective bias of unity for H I emitters, independent of redshift. Our predictions for the effective survey volume are similar given these caveats and conditions.

The effective volumes for the different survey configurations are plotted in Fig. 15. The upper panel shows the fraction of the geometrical volume in differential redshift shells that is sampled by the effective volume probed by the different surveys and the lower panels show the cumulative effective volumes. The surveys track the geometrical volume available until the redshift at which $\bar{n}P < 1$. This is clear from the lower panel of Fig. 15, in which the effective volume curves flatten once this redshift is reached. The effective volume covered by the SKA configurations is larger than that expected to be covered by the Euclid survey for a fixed H_2/H_1 ratio. In the event that the variable H_2/H_1 ratio is adopted, then the effective volumes sampled by Euclid and SKA are comparable.

4.3 The forecast error on the dark energy equation of state

The effective volume gives a broad brush view of the potential performance of a survey. In order to get a more quantitative impression, we need to make a forecast of the error on the parameter of interest, which in our case is the dark energy equation of state parameter, w . This will allow us to assess if the volume sampled by the survey is at a redshift which is useful for constraining the value of w . The conclusions will depend to some extent on the dark energy model adopted. The fiducial model we use is a flat cold dark matter universe with a cosmological constant. The cosmological constant has little influence on the cosmological expansion above $z \approx 1.5$ –2. Hence, a difference in effective volume between survey configurations at these redshifts is likely to have little impact on how well w can be measured. This behaviour could change if we adopted a different dark energy model, such as one with appreciable amounts of dark energy at early epochs (see, for example, the plots of Hubble parameter and luminosity distance in Jennings et al. 2010).

To make the forecast of the error on w for a particular survey configuration, we use a Fisher matrix approach, closely following the calculation in Seo & Eisenstein (2003).

The observable quantity is the ratio of the measured power spectrum, including the BAO signal, to a smooth reference spectrum, which has no BAO features, as outlined in Section 4.1. The Fisher matrix encodes the sensitivity of the measured ratio to different cosmological parameters, assuming Gaussian errors. The derived error on the cosmological parameters is given by the square root of the diagonal of the inverse of the Fisher matrix.

Our goal is to compare the different survey configurations, so we use a number of approximations to simplify the calculation. In particular, we work in the flat sky approximation, ignore the impact of redshift space distortions on the appearance of the BAO and neglect any evolution of the power spectrum over bins of redshift of width 0.1. Under these assumptions, the Fisher matrix (for arbitrary parameters) obtained from the power spectrum is given by (Tegmark, Taylor & Heavens 1997; Seo & Eisenstein 2003),

$$F_{ij} = \sum_{i=1}^{N_z} \int_{k_{\min}}^{k_{\max}} \frac{\partial \ln R(k, z_i)}{\partial p_i} \frac{\partial \ln R(k, z_i)}{\partial p_j} \times V_{\text{eff}}(k, z_i) \frac{4\pi k^2 dk}{2(2\pi)^3}, \quad (12)$$

where R is the measured power spectrum divided by a smooth reference, as given by equation (7) and the effective volume $V_{\text{eff}}(k, z)$ is given by equation (9). The integration is over the wavenumber interval $k_{\min} = 0.02 h \text{ Mpc}^{-1}$ to $k_{\max} = 0.2 h \text{ Mpc}^{-1}$. The summation is over redshift bins of width $\Delta z = 0.1$. To isolate the cosmological constraints which come from the BAO scale, we ignore any information stored in the amplitude of the power spectrum and assume the power spectrum is sensitive to the cosmological parameters only through the observed angular and radial distance scales. The explicit dependence, as given in Seo & Eisenstein (2003), is

$$P_{\text{obs}}(k_{\text{ref}\perp}, k_{\text{ref}\parallel}, z) = \frac{D_A^2(z)_{\text{ref}} H(z)}{D_A^2(z) H(z)_{\text{ref}}} \times P_{\text{true}}\left(\sqrt{k_{\perp}^2 + k_{\parallel}^2}, z\right), \quad (13)$$

where $k_{\text{ref}\perp} \equiv k_{\perp} D_A(z)/D_A(z)_{\text{ref}}$ and $k_{\text{ref}\parallel} \equiv k_{\parallel} H(z)_{\text{ref}}/H(z)$ relate the wavenumbers inferred via an assumed cosmological model and the true physical scales in the power spectrum. The power spectrum defined by equation (13) is then divided by a smooth reference spectrum to form the ratio R .

Table 2 shows the relative constraints forecast on a constant dark energy equation of state, w , when marginalizing over the matter density, Ω_m , for different survey configurations. The numbers shown are the ratios of $\sigma(w)$, the 1σ error on w , forecast for different H I surveys relative to those forecast for the SKA:100 fixed case (SKA with FoV 100deg^2 , integration time independent of redshift) with a variable $\text{H}_2/\text{H I}$ ratio. These calculations include a prior on Ω_m derived from a Fisher matrix for either the *WMAP5* or *Planck* CMB missions. We marginalize the CMB constraints over the spectral index n_s , Hubble parameter h , spectrum normalization A_s , baryon density Ω_b and optical depth to reionization, τ . We consider only constraints from the temperature–temperature CMB power spectra, but include an additional constraint on τ from the *WMAP5* temperature–polarization cross-power spectra (with no such constraint for *Planck*). We model only the cosmic variance and beam smearing (at 90 GHz for *WMAP* and 143 GHz for *Planck*) in the power spectrum noise models. The resulting priors are $\sigma(\Omega_m) = 0.0108$ for *WMAP5* and $\sigma(\Omega_m) = 0.00203$ for *Planck*. Note that it is not necessary to include h as a parameter in the BAO forecasts because when we consider only distance measurements, as the h dependence factors out by choosing to measure distances in units of, e.g. $h^{-1}\text{Mpc}$. This is why we marginalize over h to get the prior on Ω_m from the CMB. Table 2 shows that typical SKA configurations are expected to yield an error on w comparable to that of the $\text{H}\alpha$ survey considered by Orsi et al. (2009).

5 SUMMARY AND CONCLUSIONS

The cold gas content of galaxies and its variation with halo mass lie at the core of the galaxy formation process. The amount of cold gas in a galaxy is set by the balance between a number of competing processes. The cold gas supply comes from the cooling of gas from the hot halo and the accretion of cold gas following mergers with other galaxies. Star formation and supernova feedback act as sinks of cold gas. Semi-analytical simulations model all of these processes in the context of structure formation in the dark matter and so are ideally suited to make predictions for the distribution of cold gas between haloes of different mass. Since the models can make a wide range of predictions, their parameters are set by the requirement that a variety of observed galaxy properties be reproduced, not just the local H I data. The model predictions can be tested by measurements of the clustering of H I-selected galaxy samples, and are invaluable to plan surveys to measure the large-scale structure of the Universe with the next generation of radio telescopes.

In this paper we have compared the predictions for the distribution of cold gas in dark matter haloes in four versions of the Durham semi-analytical galaxy formation model, *GALFORM*. The Bower et al. (2006) and Font et al. (2008) models are publicly available from the Millennium Archive. Currently, most semi-analytical galaxy formation models predict the total cold gas mass and do not make a distinction between atomic or molecular hydrogen [for exceptions to this, see Fu et al. (2010) and Lagos et al. (2010)]. In order to compare with observations of the H I content of galaxies, an assumption is needed for the molecular to atomic hydrogen ratio (Blitz & Rosolowsky 2006; Obreschkow & Rawlings 2009a). Despite this uncertainty, the Bower et al. and Font et al. models still overpredict the local abundance of galaxies as a function of their cold gas mass. This excess is straightforward to fix, with the primary adjustment made to the model star formation time-scale. This modified model, based on Bower et al. (2006) is still able to reproduce the optical luminosity functions at the level enjoyed by Bower et al. We also considered a galaxy formation model set in a different cosmology,

to take advantage of a N -body simulation with a large enough box size to accurately model BAOs. This model also adopted a modified star formation time-scale to better match the local H I mass function.

The predictions from the different models have several features in common. In agreement with observations, satellite galaxies are relatively unimportant in samples selected by cold gas mass. This is true even in the Font et al. (2008) model in which satellites retain some of their hot haloes, depending on their orbit within the main halo, and can hence continue to accrete cooling gas. Samples constructed according to a cold gas mass threshold are dominated by central galaxies in haloes around $10^{11} h^{-1} M_\odot$. The HOD of central galaxies is peaked in halo mass, rather than being a step function as is the case for optical samples. As the cold gas mass cut is increased, the width of the central galaxy HOD increases and the amplitude drops. The peaked nature of the HOD of central galaxies is due to the suppression of gas cooling in massive haloes following heating by AGN. We found the same general form for the HOD in an independent model by de Lucia & Blaizot (2007), in which the implementation of AGN/radio mode feedback is different from that in *GALFORM*.

The relative abundance of central and satellite galaxies has an impact on the form of the predicted correlation function. The correlation function of a galaxy sample selected by cold gas mass is remarkably similar on small scales in real and redshift space. For pair separations in excess of a few Mpc, the redshift space correlation function has a higher amplitude than in real space, as expected given the effective bias of the sample (Kaiser 1987). In contrast, for an optically selected sample with the same number density of galaxies, the correlation is steeper in real space for pair separations below $1 h^{-1}\text{Mpc}$ and is damped in redshift space on these scales, due to the greater influence of satellite galaxies in massive haloes. On larger scales there is a more modest boost in the clustering amplitude in redshift space, due to the larger effective bias of the optical sample. The clustering predictions for H I-selected galaxies are in reasonable agreement with the measurements by Meyer et al. (2007). The clustering in the modified version of the Bower et al. model (MHIBow06) best agrees with the HIPASS results.

One of the primary science goals of the Square Kilometre Array (SKA) is to make a high-precision measurement of large-scale structure in the galaxy distribution. By measuring the apparent size of BAOs at a particular redshift, the cosmological distance to that redshift can be derived, thereby constraining the equation of state of the dark energy. By combining the galaxy formation model with a very large volume N -body simulation ($1 h^{-3} \text{Gpc}^3$), we have been able to demonstrate that galaxy samples constructed on the basis of cold gas mass can trace the BAO with the same fidelity as a near-infrared selected sample with the same number density of galaxies.

The key remaining question is how effectively do H I and optical/near-infrared redshift surveys sample the available geometrical volume and how does this translate into an error on the dark energy equation of state parameter? The effective survey volume varies substantially between H I surveys of different duration and for different assumptions about the split between atomic and molecular hydrogen. However, at least for the case of a cosmological constant, these differences occur in a redshift range which has little impact on the derived error on the equation of state. We find that H I surveys are comparable in accuracy to the most ambitious near-infrared spectroscopic surveys currently under discussion, particularly if an increase in survey speed with redshift can be achieved, and will give a factor of ≈ 3 times smaller error on w than a slitless $\text{H}\alpha$ redshift survey; all are bona fide Stage V experiments in the Dark Energy Task Force nomenclature (Albrecht et al. 2006). The assumption

about the ratio of molecular to atomic hydrogen is one of the major uncertainties at present, and leads to larger differences in the predicted counts of H I emitters than the choice of galaxy formation model. The fraction of molecular hydrogen is thought to depend upon the local conditions in the interstellar medium. This question requires further modelling (e.g. Krumholz, McKee & Tumlinson 2009), augmented by observations of the H I and CO distribution in nearby galaxies, for example by H I surveys on the SKA pathfinder MeerKAT and CO measurements using the Atacama Large Millimeter/submillimeter Array (Wootten 2008).

ACKNOWLEDGMENTS

We acknowledge a very helpful report from the referee and useful comments from Barbara Catinella and private communications from Filipe Abdalla. H-SK acknowledges support from the Korean Government's Overseas Scholarship. CSF acknowledges a Royal Society Wolfson Research Merit Award. This work was supported in part by grants from the Science and Technology Facilities Council at Durham and Leicester (CP).

REFERENCES

- Abdalla F. B., Rawlings S., 2005, *MNRAS*, 360, 27
 Abdalla F. B., Blake C., Rawlings S., 2010, *MNRAS*, 401, 743
 Albrecht A. et al., 2006, preprint (arXiv:0609591)
 Angulo R., Baugh C. M., Frenk C. S., Lacey C. G., 2008a, *MNRAS*, 383, 755
 Angulo R., Baugh C. M., Lacey C. G., 2008b, *MNRAS*, 387, 921
 Barnes D. G. et al., 2001, *MNRAS*, 322, 486
 Baugh C. M., 2006, *Rep. Prog. Phys.*, 69, 3101
 Baugh C. M., Lacey C. G., Frenk C. S., Benson A. J., Cole S., Granato G. L., Silva L., Bressan A., 2004, *New Astron. Rev.*, 48, 1239
 Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Sivia L., Bressan A., Benson A. J., Cole S., 2005, *MNRAS*, 356, 1191
 Benson A. J., 2010, *Phys. Rep.*, 495, 33
 Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, *MNRAS*, 311, 739
 Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, *ApJ*, 599, 38
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
 Blitz L., Rosolowsky E., 2006, *ApJ*, 650, 933
 Booth R. S., de Blok W. J. G., Jonas J. L., Fanaoff B., 2009, preprint (arXiv:0910.2935)
 Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
 Catinella B., Haynes M. P., Giovanelli R., Gardner J. P., Connolly A. J., 2008, *ApJ*, 685, 13
 Cattaneo A. et al., 2007, *MNRAS*, 377, 63
 Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, *MNRAS*, 271, 781
 Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168
 Croton D. J. et al., 2006, *MNRAS*, 365, 11
 De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2
 Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
 Eisenstein D. J. et al., 2005, *ApJ*, 633, 560
 Fanidakis N., Baugh C. M., Benson A. J., Bower R. G., Cole S., Done C., Frenk C. S., 2011, *MNRAS*, 410, 53
 Faulkner A. J. et al., 2010, SKA Design Study White Paper, SKADS Technical paper 44.
 Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23
 Font A. S. et al., 2008, *MNRAS*, 389, 1619
 Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, *MNRAS*, 409, 515
 Giovanelli R. et al., 2005, *AJ*, 130, 2613
 Huchtmeier W. K., Richter O.-G., 1988, *A&A*, 203, 237
 Jennings E., Baugh C. M., Angulo R. E., Pascoli S., 2010, *MNRAS*, 401, 2181
 Johnston S. et al., 2008, *Exp. Astron.*, 22, 151
 Kaiser N., 1987, *MNRAS*, 227, 1
 Kanekar N., Briggs F. H., 2004, *New Astron. Rev.*, 48, 1269
 Kanekar N., Prochaskar J. X., Eillion S. L., Chengalur J. N., 2009, *MNRAS*, 396, 381
 Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201
 Kim H. S., Baugh C. M., Cole S., Frenk C. S., 2009, *MNRAS*, 400, 1527
 Krumholz M. R., McKee C. F., Tumlinson J., 2009, *ApJ*, 693, 216
 Lagos C. D. P., Cora S. A., Padilla N. D., 2008, *MNRAS*, 388, 587
 Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J., 2010, preprint (arXiv:1011.5506).
 Lah P. et al., 2009, *MNRAS*, 399, 1447
 Lanzetta K. M., McMahon R. G., Wolfe A. M., Turnshek D. A., Hazard C., Lu L., 1991, *ApJS*, 77, 1
 Magliocchetti M., Porciani C., 2003, *MNRAS*, 346, 186
 Malbon R. K., Baugh C. M., Frenk C. S., Lacey C. G., 2007, *MNRAS*, 382, 1394
 Marin F. A., Gnedin N. Y., Seo H. J., Vallinotto A., 2010, *ApJ*, 718, 972
 Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, *ApJ*, 723, 1359
 McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, *MNRAS*, 383, 593
 Meyer M. J. et al., 2004, *MNRAS*, 350, 1195
 Meyer M. J., Zwaan M. A., Webster R. L., Brown M. J. I., Staveley-Smith L., 2007, *ApJ*, 654, 702
 Monaco P., Fontanot F., Taffoni G., 2007, *MNRAS*, 375, 1189
 Norberg P. et al., 2001, *MNRAS*, 328, 64
 Norberg P., Baugh C. M., Gaztiaga E., Craton P. J., 2009, *MNRAS*, 369, 19
 Obreschkow D., Rawlings S., 2009a, *MNRAS*, 394, 1857
 Obreschkow D., Rawlings S., 2009b, *ApJ*, 696, 129
 Obreschkow D., Croton D., DeLucia G., Khochfar S., Rawlings S., 2009, *ApJ*, 698, 1467
 Okamoto T., Gao L., Theuns T., 2008, *MNRAS*, 390, 920
 Orsi A., Baugh C. M., Lacey C. G., Cimatti A., Wang Y., Zamorani G., 2010, *MNRAS*, 405, 1006
 Parkinson C. D., Cole S., Helly J., 2008, *MNRAS*, 383, 557
 Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
 Peroux C., Dessauges-Zavadsky M., D'Odorico S., Kim T. S., McMahon R. G., 2005, *MNRAS*, 363, 479
 Popping A., Dave R., Braun R., Oppenheimer B. D., 2009, *A&A*, 504, 15
 Power C., Baugh C. M., Lacey C. G., 2010, *MNRAS*, 406, 43
 Sage L. J., 1993, *A&A*, 100, 537
 Sanchez A., Baugh C. M., Angulo R., 2008, *MNRAS*, 390, 1490
 Sanchez A. G., Crocce M., Cabre A., Baugh C. M., Gaztanaga E., 2009, *MNRAS*, 400, 1643
 Schilizzi R. T., Dewdney P. E. F., Lazio T. J. W., 2008, in Stepp L. M., Gilmozzi R., eds, *Proc. SPIE Vol. 7012, Ground-based and Airborne Telescopes II*. SPIE, Bellingham, p. 70121I
 Seljak U., 2000, *MNRAS*, 318, 2039
 Seo H. J., Eisenstein D. J., 2003, *ApJ*, 598, 720
 Sheth R. K., Mo H. J., Tormen G., 2001, *MNRAS*, 323, 1
 Springel V. et al., 2005, *Nat*, 435, 629
 Srianand R., Gupta N., Petitjean P., Noterdaeme P., Ledoux C., 2010, *MNRAS*, 405, 1888
 Storrie-Lombardi L. J., Irwin M. J., McMahon R. G., 1996, *MNRAS*, 282, 1330
 Tegmark M., Taylor A. N., Heavens A. F., 1997, *ApJ*, 480, 22
 Tinker J. L., Norberg P., Weinberg D. H., Warren M. S., 2007, *ApJ*, 659, 877
 van den Bosch F. C., Mo H. J., Yang X., 2003, *MNRAS*, 345, 923
 Verheijen M., Van Gorkom J. H., Szomoru A., Dwarakanath K. S., Puggianti B. M., Schiminovich D., 2007, *ApJ*, 668, 9
 Verheijen M. A. W., Oosterloo T. A., van Cappellen W. A., Bakker L., Ivashina M. V., van der Hulst J. M., 2008, in Minchin R., Momjian E., eds, *AIP Conf. Proc. Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window*. Am. Inst. Phys., New York, p. 265
 Wake D. et al., 2008, *MNRAS*, 387, 1045

- Waugh M. et al., 2002, MNRAS, 337, 641
Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Wolfe A. M., Lanzetta K. M., Foltz C. B., Chaffee F. H., 1995, ApJ, 454, 698
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861
Wooten A., 2008, Astrophys. Space Sci., 313, 9
Wyithe J. S. B., Brown M. J. I., 2010, MNRAS, 404, 876
Wyithe J. S. B., Brown M. J. I., Zwaan M. A., Meyer M., 2009, MNRAS, submitted (arXiv:0908.2854)
- Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., 2005, MNRAS, 339, 1057
Zehavi I. et al., 2005, ApJ, 630, 1
Zheng Z., 2004, ApJ, 610, 61
Zwaan M. A. et al., 2003, AJ, 125, 2842
Zwaan M. A., Meyer M. J., Satveley-Smith L., Webster R. L., 2005, MNRAS, 359, 30

This paper has been typeset from a \TeX/L\TeX file prepared by the author.