Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **465,** 111–122 (2017) doi:10.1093/mnras/stw2779
Advance Access publication 2016 October 30

The spatial distribution of neutral hydrogen as traced by low H I mass galaxies

Han-Seek Kim,[1⋆] J. Stuart. B. Wyithe,[1,2] C. M. Baugh,[3] C. d. P. Lagos,[2,4] C. Power[2,4] and Jaehong Park[1]

[1] *School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia*
[2] *Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), 44 Rosehill Street Redfern, NSW 2016, Australia*
[3] *Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[4] *International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*

Accepted 2016 October 25. Received 2016 October 23; in original form 2016 March 7

ABSTRACT

The formation and evolution of galaxies with low neutral atomic hydrogen (H I) masses, $M_{\rm H\,I} < 10^8\,h^{-2}$ M$_\odot$, are affected by host dark matter halo mass and photoionization feedback from the UV background after the end of reionization. We study how the physical processes governing the formation of galaxies with low H I mass are imprinted on the distribution of neutral hydrogen in the Universe using the hierarchical galaxy formation model, GALFORM. We calculate the effect on the correlation function of changing the H I mass detection threshold at redshifts $0 \le z \le 0.5$. We parametrize the clustering as $\xi(r) = (r/r_0)^{-\gamma}$ and we find that including galaxies with $M_{\rm H\,I} < 10^8\,h^{-2}$ M$_\odot$ increases the clustering amplitude r_0 and slope γ compared to samples of higher H I masses. This is due to these galaxies with low H I masses typically being hosted by haloes with masses greater than $10^{12}\,h^{-1}$ M$_\odot$, and is in contrast to optically selected surveys for which the inclusion of faint, blue galaxies lowers the clustering amplitude. We show the H I mass function for different host dark matter halo masses and galaxy types (central or satellite) to interpret the values of r_0 and γ of the clustering of H I-selected galaxies. We also predict the contribution of low H I mass galaxies to the 21 cm intensity mapping signal. We calculate that a dark matter halo mass resolution better than $\sim 10^{10}\,h^{-1}$ M$_\odot$ at redshifts higher than 0.5 is required in order to predict converged 21 cm brightness temperature fluctuations.

Key words: galaxies: abundances – galaxies: formation – large-scale structure of Universe – radio lines: galaxies.

1 INTRODUCTION

In hierarchical structure formation theory, the distribution of galaxies is related to the distribution of dark matter in the Universe. Studies of the spatial distribution of galaxies as a function of colour, luminosity, type and stellar mass therefore give important information about where galaxies form and which haloes host them Brown, Webster & Boyle (2000; Norberg et al. 2002; Madgwick et al. 2003; Loh et al. 2010; Zehavi et al. 2011; Campbell et al. 2015).

Neutral hydrogen is a key ingredient of galaxy formation as it traces the processes of gas accretion, outflow from galaxies and star formation. It is therefore important to understand the relation between neutral hydrogen and dark matter haloes. However, understanding galaxy formation and evolution using neutral hydrogen observations is hampered by the weakness of the signal compared to optical observations. The last decade has seen improvements in both the volume and flux limit of surveys which probe the H I content of galaxies from the H I Parkes All Sky Survey (HIPASS; Zwaan et al. 2005), the Arecibo Legacy Fast ALFA (ALFALFA; Martin et al. 2010) and the Arecibo Ultra-Deep Survey (Hoppmann et al. 2015). Studies of the distribution of H I-selected galaxies in the local Universe have shown that these galaxies are an unbiased (weakly clustered) galaxy population Meyer et al. (2007); Martin et al. (2012); Papastergis et al. (2013).

There have been previous studies of the large-scale structure of galaxies based on their neutral hydrogen using the halo occupation distribution formalism Wyithe et al. (2009) and hierarchical galaxy formation models Kim et al. (2011, 2013b). These works have focused on high H I mass galaxies ($M_{\rm H\,I} > 10^{9.25}\,h^{-2}$ M$_\odot$), which are mainly central galaxies located in dark matter haloes of masses $\sim 10^{12}\,h^{-1}$ M$_\odot$. In contrast, red massive galaxies have either consumed most of their gas (neutral hydrogen) in forming stars or have expelled it by feedback processes Power, Baugh &

⋆ E-mail: hansikk@unimelb.edu.au

Lacey (2010); Kim et al. (2011); Lagos et al. (2011a), as shown by their very low neutral gas content Young et al. (2011); Serra et al. (2012); Boselli et al. (2014b). H I-deficient galaxies also result from ram pressure stripping by hot gaseous atmospheres around satellite galaxies in groups and clusters, and tidal interactions between galaxies in high-density environments Font et al. (2008); McCarthy et al. (2008); Cortese et al. (2012); Boselli et al. (2014a); Bahé & McCarthy (2015). These processes, all affect the distribution of satellite galaxies in groups and clusters.

Kim et al. (2015) demonstrated the importance of photoionization feedback for galaxy formation and evolution in low-mass dark matter haloes, and the low-mass end of the H I mass function in the local Universe (see also Lagos et al. 2011a; Kim et al. 2013b). Kim et al. (2015) showed that the contribution of low-mass dark matter haloes to the low-mass end of the H I mass function changes dramatically when different photoionization feedback modelling schemes are applied. Therefore, it is natural to expect these models to affect the predicted clustering of low H I mass galaxies.

The number of individual galaxies detected in H I will greatly increase through ongoing and upcoming large volume H I-selected galaxy surveys using the SKA (Square Kilometre Array; e.g. Baugh et al. 2004; Power et al. 2010; Kim et al. 2011) and its pathfinders, such as ASKAP (Australian Square Kilometre Array Pathfinder; cf. Johnston et al. 2008) and MeerKAT (Meer Karoo Array Telescope; cf. Jonas 2007) in the Southern hemisphere, FAST (Five-hundred-meter Aperture Spherical Telescope)[1] and APERTIF (APERture Tile In Focus)[2] in the Northern hemisphere. In addition, the BINGO (the BAO in Neutral Gas Observations; Battye et al. 2012), CHIME (the Canadian Hydrogen Intensity Mapping Experiment; Bandura et al. 2014) and SKA are expected to detect the distribution of neutral hydrogen at $0.5 < z < 3$ in the Universe using the 21 cm intensity mapping technique which does not require individual H I-selected galaxies to be resolved Wyithe & Loeb (2007); Wolz et al. (2014); Santos et al. (2015). There have been several theoretical models predicting the H I intensity during the post-reionization era. Each model includes a prescription for the relation between H I and dark matter halo masses. This results in the implicit assumption that the H I traces the underlying matter distribution with a linear bias (Bagla, Khandai & Datta 2010; Guha Sarkar et al. 2012; Padmanabhan, Choudhury & Refregier 2015; Villaescusa-Navarro et al. 2016). For example, Bagla et al. (2010) propose phenomenological functions for this relation based on the observation that haloes much more massive than a few times 10^{11} $\mathrm{M}_{\odot}$ do not host significant amounts of H I (Catinella et al. 2013; Dénes, Kilborn & Koribalski 2014). This simple model successfully explains both observations of the abundance of damped Lyman α absorbers at redshifts between 2.4 and 4, and the bias of H I-selected galaxies in the local Universe (Villaescusa-Navarro et al. 2014; Padmanabhan, Choudhury & Refregier 2015; Sarkar, Bharadwaj & Anathpindika 2016). On the other hand, using zoom hydrodynamic simulations, Villaescusa-Navarro et al. (2016) found that the mass in neutral hydrogen of dark matter haloes increases monotonically with halo mass, and can be well described by a power law. Here, we show the importance of the contribution of galaxies with low H I masses to the large-scale distribution of neutral hydrogen in the Universe, and why they cannot be neglected if an accurate measurement the H I intensity mapping power spectrum is desired.

[1] Details at http://fast.bao.ac.cn/en/.

[2] Details at http://www.astron.nl/jozsa/wnshs.

The structure of the paper is as follows. In Section 2, we provide a brief overview of the hierarchical galaxy formation models used. In Section 3, we show the clustering of H I-selected galaxies from the models and the comparison between the models and observations. In Section 4, we extend predictions of H I galaxy clustering to higher redshifts and show predictions for 21 cm brightness fluctuations at different redshifts in Section 5. We finish with a summary of our results in Section 6.

2 THE GALAXY FORMATION MODEL

We use the semi-analytical galaxy formation model GALFORM (cf. Cole et al. 2000; Baugh 2006; Lacey et al. 2016) described in Lagos et al. (2012) and Kim et al. (2015) to predict the properties of galaxies which form and evolve in the ΛCDM cosmology. GALFORM models the key physical processes involved in galaxy formation, including the gravitationally driven assembly of dark matter haloes, radiative cooling of gas and its collapse to form centrifugally supported discs, star formation, feedback from supernovae and active galactic nuclei (AGN), and photoionization. We implement GALFORM in halo merger trees extracted from the Millennium-II cosmological *N*-body simulation[3] which has a 100 h^{-1} Mpc box size (Boylan-Kolchin et al. 2009).

We use the Dhalo merger trees (D-Trees; Jiang et al. 2014). These were designed for galaxy formation modelling and form the basis of the Durham semi-analytic galaxy formation model, GALFORM. The D-Trees algorithm is based on the friends-of-friends algorithm, FOF (Davis et al. 1985). The FOF haloes are decomposed into subhaloes by SUBFIND (Springel 2005). In the D-Trees, subhaloes have consistent membership over time in the sense that once a subhalo is accreted by a Dhalo it never demerges. In contrast to the FOF haloes, in this process we also split some FOF haloes into two or more Dhaloes when SUBFIND substructures are well separated and only linked into a single FOF halo by bridges of low-density material. The central is associated with the main subhalo. Satellite galaxies were originally centrals until their FOF halo merged with a larger one, and are associated with the corresponding subhalo, or the most bound particle from this subhalo if the subhalo itself can no longer be identified. Full details of the Dhaloes can be found in Merson et al. (2013) and Jiang et al. (2014).

The particle mass of the simulation is $6.89 \times 10^{6}\,h^{-1}\,\mathrm{M}_{\odot}$ and we retain haloes down to 20 particles. In the extreme case where all available baryons are in H I, the H I mass of a galaxy in such a halo would be $\sim 1.8 \times 10^{6}\,h^{-2}\,\mathrm{M}_{\odot}$. Thus, the resolution limit of the Millennium-II simulation is sufficient to model the H I mass function over the range of current observations (i.e. down to $M_{\mathrm{H\,I}} \sim 10^{6}\,\mathrm{M}_{\odot}$).

Lagos et al. (2011a) extended GALFORM by modelling the splitting of cold gas in the interstellar medium (ISM) into its H I and H_2 components and by explicitly linking star formation to the amount of H_2 present in a galaxy. Lagos et al. (2011a) compared empirically and theoretically derived star formation laws (cf. Blitz & Rosolowsky 2006; Krumholz, McKee & Tumlinson 2009) with a variety of observations and found that the empirical law of Blitz & Rosolowsky (2006, see also Leroy et al. 2008) is favoured by these data.

[3] The cosmological parameters adopted for the Millennium simulations are total matter density $\Omega_{\mathrm{M}} = 0.25$, baryon density $\Omega_{\mathrm{b}} = 0.045$, dark energy density $\Omega_{\Lambda} = 0.75$, Hubble parameter $H_0 = 73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, the primordial scalar spectral index $n_{\mathrm{s}} = 1$ and the fluctuation amplitude $\sigma_8 = 0.9$.

Another important physical process affecting H I masses in galaxies is reionization. The standard reionization model in GALFORM suppresses cooling in galaxies with host halo circular velocity $V_{\rm circ}$ below a threshold $V_{\rm cut}$ below redshift $z_{\rm cut}$ (see Benson et al. 2002). Lagos et al. (2011b) adopted $V_{\rm cut} = 30$ km s^{-1} and $z_{\rm cut} = 10$. Lagos et al. (2012) changed two parameters corresponding to the star formation time-scale during starbursts, $\tau_{\rm min} = 100$ Myr and $f_{\rm dyn} = 50$ when the star formation time-scale is calculated as $\tau_{\rm SF} = \min(f_{\rm dyn} \times \tau_{\rm dyn,\,bulge}, \tau_{\rm min})$, to better match the observed rest-frame UV luminosity function over the redshift range $z = 3$–6 (hereafter Lagos2012). Here, $\tau_{\rm dyn,\,bulge}$ is the dynamical time-scale of the bulge.

Kim et al. (2015, hereafter Kim15) introduced a new model for photoionization feedback in order to correctly describe the low-mass end of the H I mass function ($10^6\,h^{-2}\,{\rm M}_{\odot} < M_{\rm H\,I} < 10^8\,h^{-2}\,{\rm M}_{\odot}$) estimated from the ALFALFA survey at $z = 0$. In contrast to Lagos2012, Kim et al. (2015) adopted the redshift-dependent photoionization feedback model (Dijkstra et al. 2004) described in Sobacchi & Mesinger (2013). Kim et al. (2015) use a redshift-dependent $V_{\rm cut}$ given by

$$V_{\rm cut}(z)[{\rm km\,s^{-1}}] = V_{\rm cut0}(1+z)^{\alpha_v}\left[1-\left(\frac{1+z}{1+z_{\rm IN}}\right)^2\right]^{2.5/3}, \qquad (1)$$

where $z_{\rm IN}$ is the redshift of UV background exposure, $V_{\rm cut0}$ is the circular velocity of dark matter haloes at $z = 0$ below which photoionization feedback suppresses gas cooling and α_v parametrizes the redshift dependence of the velocity cut. We use the best-fitting set of parameters in Kim et al. (2015), ($V_{\rm cut0}$, α_v) = (50 km s^{-1}, -0.82) at $z_{\rm IN} = 10$, which was found by comparing the predicted H I mass function to the measurements by the ALFALFA survey. The predicted H I mass function using this set of parameters shows an improved match with observations compared to the Lagos2012 model. Note that Sobacchi & Mesinger (2013) suggested $\alpha_v = -0.2$ and a critical halo mass that is 10 times larger than the one we used for a UV background intensity of 1×10^{21} erg s^{-1} Hz^{-1} cm^{-2} sr^{-1} (for more details, see Kim et al. 2015).

3 LARGE-SCALE DISTRIBUTION OF NEUTRAL HYDROGEN IN THE LOCAL UNIVERSE

We show clustering predictions for galaxies selected using different H I mass thresholds in Section 3.1. To compare the spatial distributions of galaxies selected by different H I mass criteria, we investigate the contributions of different host dark matter halo masses and types of galaxies to the H I mass function in Section 3.2.

We quantify the large-scale distribution of galaxies using the two-point correlation function. This is written in terms of the excess probability, compared to a random distribution, of finding another galaxy at a separation r,

$$dP = \bar{n}^2[1+\xi(r)]\delta V_1 \delta V_2, \qquad (2)$$

where $\bar{n}$ is the mean number density of the galaxy sample and δV_1, δV_2 are volume elements. In the case of an H I-selected galaxy catalogue produced from a galaxy formation model combined with a periodic *N*-body simulation, we can measure the correlation function using

$$1+\xi(r) = D_{\rm H\,I}D_{\rm H\,I}/(\bar{n}_{\rm H\,I}^2 V {\rm d}V), \qquad (3)$$

where $D_{\rm H\,I}D_{\rm H\,I}$ is the number of pairs of H I-selected galaxies with separations in the range r to $r+\delta r$, $\bar{n}_{\rm H\,I}$ is mean number density of H I-selected galaxies, V is the volume of the simulation and dV is the differential volume corresponding to r to $r+\delta r$.

The correlation function can be described by a power law using two parameters as

$$\xi(r) = (r/r_0)^{-\gamma}, \qquad (4)$$

where r_0 indicates the correlation length at which $\xi(r_0) = 1$, corresponding to twice the probability relative to a random distribution. The value of r_0 is related to the typical host dark matter halo mass of the selected galaxies, and γ is related to the number of satellite galaxies in a host dark matter halo. Here, the parameters r_0 and γ are derived by minimizing χ^2 over separations between 1 and 10 h^{-1} Mpc, and considering γ values between 0 and 5.

3.1 H I mass-dependent clustering

The ALFALFA survey measures the H I mass function down to H I masses of $\sim 10^6\,h^{-2}\,{\rm M}_{\odot}$. This low-mass limit provides an important test for galaxy formation models. The Kim15 model is able to reproduce the observed H I mass function from ALFALFA by adopting a redshift-dependent photoionization feedback model (see Section 2 and also Fig. 2).

In Fig. 1, we show clustering predictions for galaxies selected using different H I mass thresholds. The clustering predictions for the two highest H I mass thresholds ($M_{\rm H\,I} > 10^8\,h^{-2}\,{\rm M}_{\odot}$ and $M_{\rm H\,I} > 10^9\,h^{-2}\,{\rm M}_{\odot}$) are very similar, both in shape (γ) and the correlation length (r_0). However, the two lowest H I mass thresholds ($M_{\rm H\,I} > 10^6\,h^{-2}\,{\rm M}_{\odot}$ and $M_{\rm H\,I} > 10^7\,h^{-2}\,{\rm M}_{\odot}$) show larger correlation lengths (larger r_0) and also steeper correlation functions (larger γ) than the two high H I mass samples (see values of r_0 and γ in Table 1 and Fig. 7 at $z = 0$). Overall, the trend of clustering as a function of H I mass threshold shows that galaxies become more clustered as the H I mass threshold *decreases*.

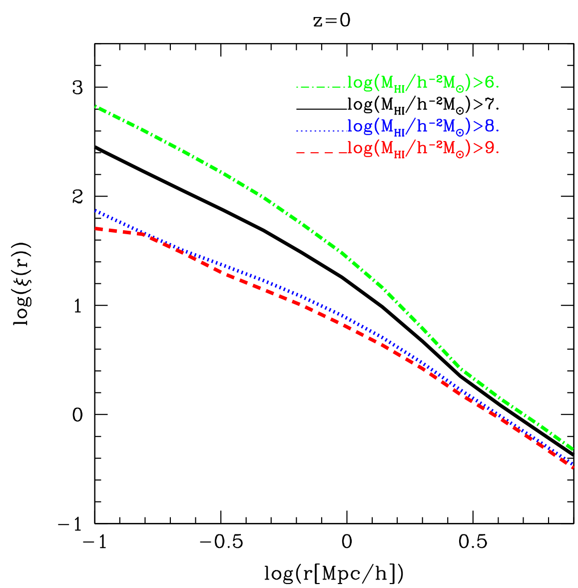

Figure 1. The real-space correlation function at $z = 0$ predicted for different H I mass thresholds using the Kim15 model, as labelled. Reprinted from Kim et al. (2015).

Table 1. The best-fitting parameters ($r_0[h^{-1}\mathrm{Mpc}]$ and γ) in equation (4) for predicted correlation functions of different H I mass thresholds at $z = 0$, 0.2 and 0.5 in the Kim15 model. In the square bracket at $z = 0$, we show the r_0 and γ values following Meyer et al. (2007) using the projected correlation function.

H I mass threshold	$r_0[r_{0,\,\mathrm{pow}}]$ ($z = 0$)	γ [$\gamma_{,\mathrm{pow}}$] ($z = 0$)	r_0 ($z = 0.2$)	γ ($z = 0.2$)	r_0 ($z = 0.5$)	γ ($z = 0.5$)
$>10^6\,h^{-2}\,\mathrm{M}_\odot$	5.12 [4.91]	1.87 [1.99]	4.61	1.81	3.84	1.62
$>10^7\,h^{-2}\,\mathrm{M}_\odot$	4.72 [4.42]	1.74 [1.91]	4.06	1.65	3.19	1.49
$>10^8\,h^{-2}\,\mathrm{M}_\odot$	3.99 [3.71]	1.54 [1.80]	3.34	1.48	2.77	1.39
$>10^9\,h^{-2}\,\mathrm{M}_\odot$	3.77 [3.52]	1.50 [1.76]	3.29	1.45	2.95	1.38

3.2 Understanding the H I mass-dependent clustering in the Kim15 model

The clustering of galaxy samples can be broken down into the contribution of the host dark matter haloes Berlind & Weinberg (2002); Kravtsov et al. (2004); Kim et al. (2011, 2012). We therefore investigate the contributions to the H I mass function from different galaxies and different mass host dark matter haloes.

The top-left panel of Fig. 2 shows the contribution of different types of H I-selected galaxies (black solid line for all galaxies, blue dotted line for central galaxies and red dashed line for satellite galaxies) to the H I mass function. Central galaxies are the dominant (over 90 per cent) type of galaxy when we focus on galaxies with H I masses $M_{\mathrm{H\,I}} > 10^{8.5}\,h^{-2}\,\mathrm{M}_\odot$. However, the contributions of central galaxies and satellite galaxies to the H I mass function are similar for $M_{\mathrm{H\,I}} < 10^8\,h^{-2}\,\mathrm{M}_\odot$. This is consistent with the findings from hydrodynamic simulations (Davé et al. 2013).

The top-right panel of Fig. 2 shows the H I mass functions for galaxies in different bins of host dark matter halo mass. Note that the host dark matter halo mass, M_{halo}, is the mass of the associated host Dhalo. Most galaxies with H I masses $>10^{7.5}\,h^{-2}\,\mathrm{M}_\odot$ reside in host dark matter haloes with masses between 10^{10} and $10^{12}\,h^{-1}\,\mathrm{M}_\odot$. Host dark matter haloes with mass $<10^{10}\,h^{-1}\,\mathrm{M}_\odot$ mainly contribute to the low-mass end of the H I mass function, $M_{\mathrm{H\,I}} < 10^7\,h^{-2}\,\mathrm{M}_\odot$, while making a negligible contribution to the H I mass function at H I masses higher than $10^8\,h^{-2}\,\mathrm{M}_\odot$. The most massive host dark matter haloes ($>10^{14}\,h^{-1}\,\mathrm{M}_\odot$) contribute less than 10 per cent to the abundance of galaxies at all H I masses.

We show host dark matter halo contributions divided into central (bottom-left panel of Fig. 2) and satellite galaxies (bottom-right panel of Fig. 2). For central galaxies, the dominant contribution to the H I mass function at $M_{\mathrm{H\,I}} > 10^{7.5}\,h^{-2}\,\mathrm{M}_\odot$ is from host dark matter haloes with masses $10^{10}\,h^{-1}\,\mathrm{M}_\odot < M_{\mathrm{halo}} < 10^{12}\,h^{-1}\,\mathrm{M}_\odot$. Central galaxies which reside in higher mass host dark matter haloes ($10^{12}\,h^{-1}\,\mathrm{M}_\odot < M_{\mathrm{halo}} < 10^{14}\,h^{-1}\,\mathrm{M}_\odot$ and $>10^{14}\,h^{-1}\,\mathrm{M}_\odot$) represent a very small fraction of H I-selected galaxies. Host dark matter haloes with masses less than $10^{10}\,h^{-1}\,\mathrm{M}_\odot$ mainly host central galaxies with $M_{\mathrm{H\,I}} < 10^7\,h^{-2}\,\mathrm{M}_\odot$.

Photoionization feedback in the model suppresses the cooling of gas in central galaxies which have circular velocities less than $\mathrm{V}_{\mathrm{cut}}$. Therefore, the suppressed cooling of gas in central galaxies introduces the departure from a Schechter-like function. On the other hand, the abundance of satellite galaxies in massive dark matter haloes increases monotonically towards low H I mass. The reason why the H I mass function (HIMF) of satellite galaxies does not show a dip at low H I masses is that the satellites were mainly formed before reionization.

We find that the number of H I-selected galaxies that are satellites increases with decreasing H I mass for all host dark matter halo masses. Furthermore, at $M_{\mathrm{H\,I}} < 10^7\,h^{-2}\,\mathrm{M}_\odot$, H I satellite galaxies dominate the number density. These satellite galaxies tend to lie in host dark matter haloes of mass $M_{\mathrm{halo}} > 10^{12}\,h^{-1}\,\mathrm{M}_\odot$.

In Fig. 1, we found that the correlation length r_0 becomes larger as the H I mass threshold decreases for the selected galaxy samples. This is due to the larger contributions of high-mass host dark matter haloes ($M_{\mathrm{halo}} > 10^{12}\,h^{-1}\,\mathrm{M}_\odot$) to the H I mass function (see top-right panel of Fig. 2) at $M_{\mathrm{H\,I}} < 10^8\,h^{-2}\,\mathrm{M}_\odot$. In addition, the contribution of satellite galaxies residing in the highest mass host dark matter haloes ($M_{\mathrm{halo}} > 10^{14}\,h^{-1}\,\mathrm{M}_\odot$) at $M_{\mathrm{H\,I}} < 10^8\,h^{-2}\,\mathrm{M}_\odot$ leads to a boost in the clustering amplitude at small separations (see Fig. 1). We also find that the slope of the correlation function decreases with increasing H I mass due to the reduced contribution of satellite galaxies.

We show the distribution of host dark matter halo mass as a function of H I mass in galaxies in the top-left panel of Fig. 3. The symbols show the median host dark matter halo mass as a function of H I mass for central galaxies (blue squares), satellite galaxies (red triangles) and all galaxies (black circles). The bars show the 10th–90th percentile range of the distribution of host dark matter halo masses. The host dark matter halo mass of central galaxies increases as the H I mass increases. In contrast, the median host dark matter halo mass for satellite H I-selected galaxies is constant over the entire range of H I masses (median value $\sim 10^{13}\,h^{-1}\,\mathrm{M}_\odot$).

We also show in the top-right panel of Fig. 3 the *total* H I mass contained in all galaxies within a halo. The symbols show the median total H I mass as a function of host dark matter halo mass, for central galaxies (open blue squares), all satellite galaxies (open red triangles) and all galaxies (open black circles) in a halo. The bars show the 10th–90th percentile range of the distribution of total H I masses. The total H I mass in a halo is contributed mainly by the central galaxy for host dark matter haloes less massive than $10^{12}\,h^{-1}\,\mathrm{M}_\odot$ and satellite galaxies for host dark matter haloes more massive than $10^{12}\,h^{-1}\,\mathrm{M}_\odot$. Note that these measurements have not been made observations yet and thus these are model predictions that need to be verified by future observations. In this model, there is very little H I gas in the central galaxy haloes more massive than this due to the shutdown of the cooling flow by AGN heating (Kim et al. 2011). The total H I mass of satellites (adding all H I masses of satellites together in a halo) increases as the halo mass increases. This predicted relation between total H I mass in a host dark matter halo and dark matter halo mass would be the physically motivated input to make H I intensity mapping mock observations based on dark haloes using large volume dark matter only simulations. The bottom panel of Fig. 3 shows the neutral hydrogen gas fraction, $M_{\mathrm{H\,I+H_2}}/M_\star$, in bins of H I mass. $M_{\mathrm{H_2}}$ is the molecular hydrogen mass and $M_\star$ is the stellar mass. The symbols show the median neutral hydrogen gas fraction as a function of the H I mass of galaxies, for central galaxies (blue squares) and satellite galaxies (red triangles). The bars show the 10th–90th percentile ranges of the distributions of neutral hydrogen gas fractions. The overall neutral

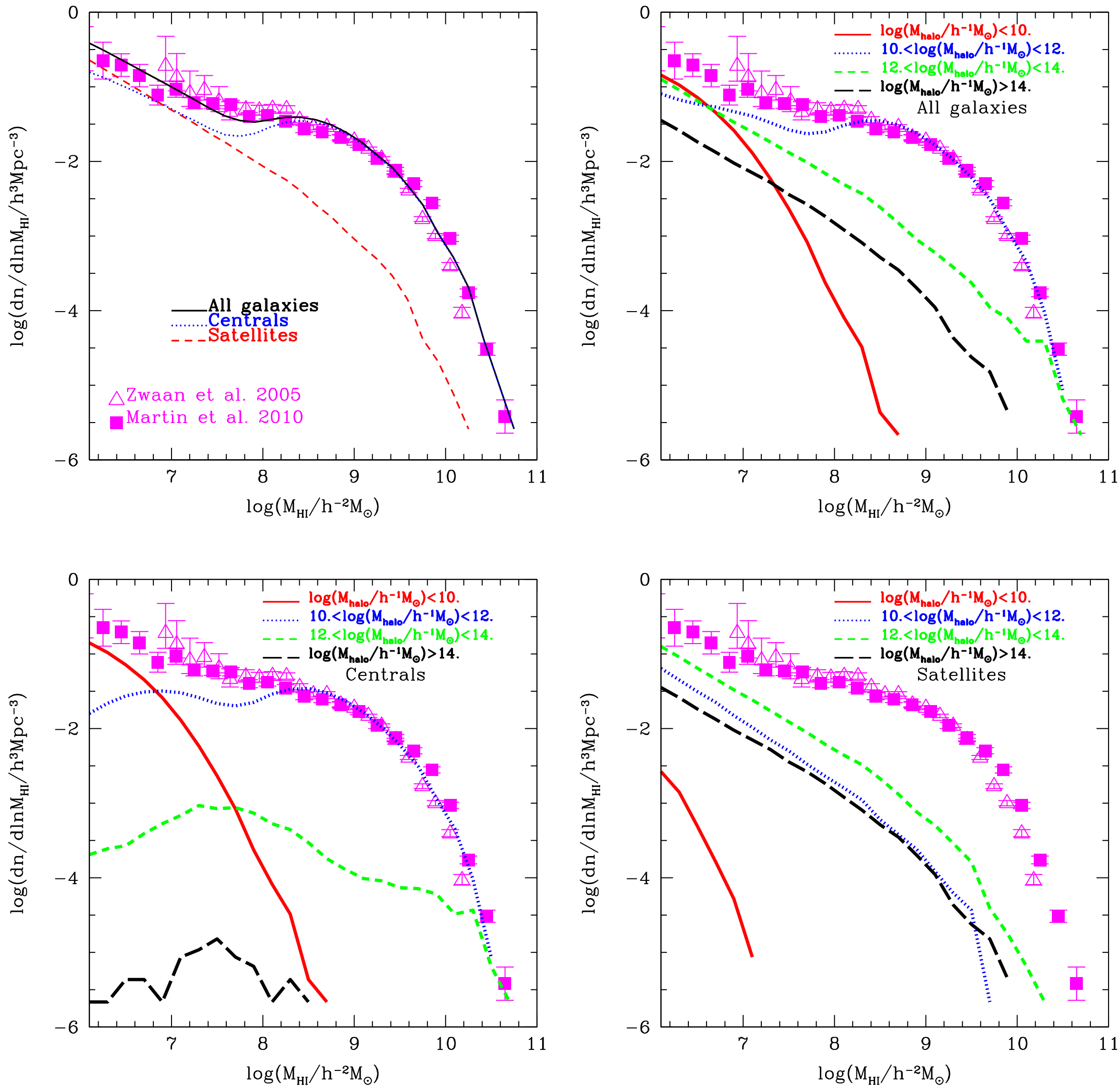

Figure 2. The predicted H I mass function in the Kim15 model. The top-left panel shows the contribution of central and satellite galaxies to the HIMF (originally plotted in Kim et al. 2015). The top-right panel shows the contribution of different masses of host dark matter haloes. The bottom panels show the contribution of different masses of host dark matter haloes for central galaxies (left-hand panel) and satellite galaxies (right-hand panel). The symbols correspond to observational estimations from the local Universe from HIPASS (open triangles; cf. Zwaan et al. 2005) and ALFALFA (filled squares; cf. Martin et al. 2010). $M_{\rm halo}$ is the mass of the associated Dhalo.

hydrogen gas fraction for satellites is lower than centrals. Because the hot gas in the subhaloes around satellites is removed instantaneously and is transferred to the hot gas reservoir of the main halo once a galaxy becomes a satellite in the model. This removal of the hot gas prevents the satellite galaxy to accrete newly cooled gas, condemning it to have ever lower gas fractions as star formation continues (see discussion in Lagos et al. 2011a for the competing environment and star formation processes). The feedback effects are also imprinted on the low H I mass end from the suppression of cooling of gas by the photoionization and the high H I mass end from the shutting down of the cooling flow by the AGN feedback for centrals.

3.3 Imprinted photoionization feedback on H I clustering

In the previous subsection, we presented predictions for H I galaxy clustering and the mass function. By comparing these predictions with the observed H I mass function and clustering of H I-selected galaxies from HIPASS Meyer et al. (2007) and ALFALFA (Martin et al. 2012, see Kim et al. 2011, 2013a, 2015), we can test the galaxy formation model. Martin et al. (2012) showed that H I-selected galaxies display an anisotropic clustering pattern, and are less clustered than dark matter on scales <5 Mpc. Furthermore, Papastergis et al. (2013) found no evidence for H I mass-dependent clustering over the H I mass range between $10^{8.5}$ and $10^{10.5}$ $M_{\odot}$. Meyer et al. (2007) used 4315 sources from the HIPASS survey to measure the clustering of H I-selected galaxies. Martin et al. (2012) used a sample of ∼10 150 galaxies ($M_{\rm H\,I} > 10^{6.2}\,h^{-2}$ $M_{\odot}$) and Papastergis et al. (2013) used a sample of ∼6000 galaxies detected by the ALFALFA 21 cm survey ($M_{\rm H\,I} > 10^{7.5}\,h^{-2}$ $M_{\odot}$), to measure the clustering properties of H I-selected galaxies. Whilst these surveys represent a tremendous step forwards in terms of blind surveys of the H I content of galaxies, it is important to bear in

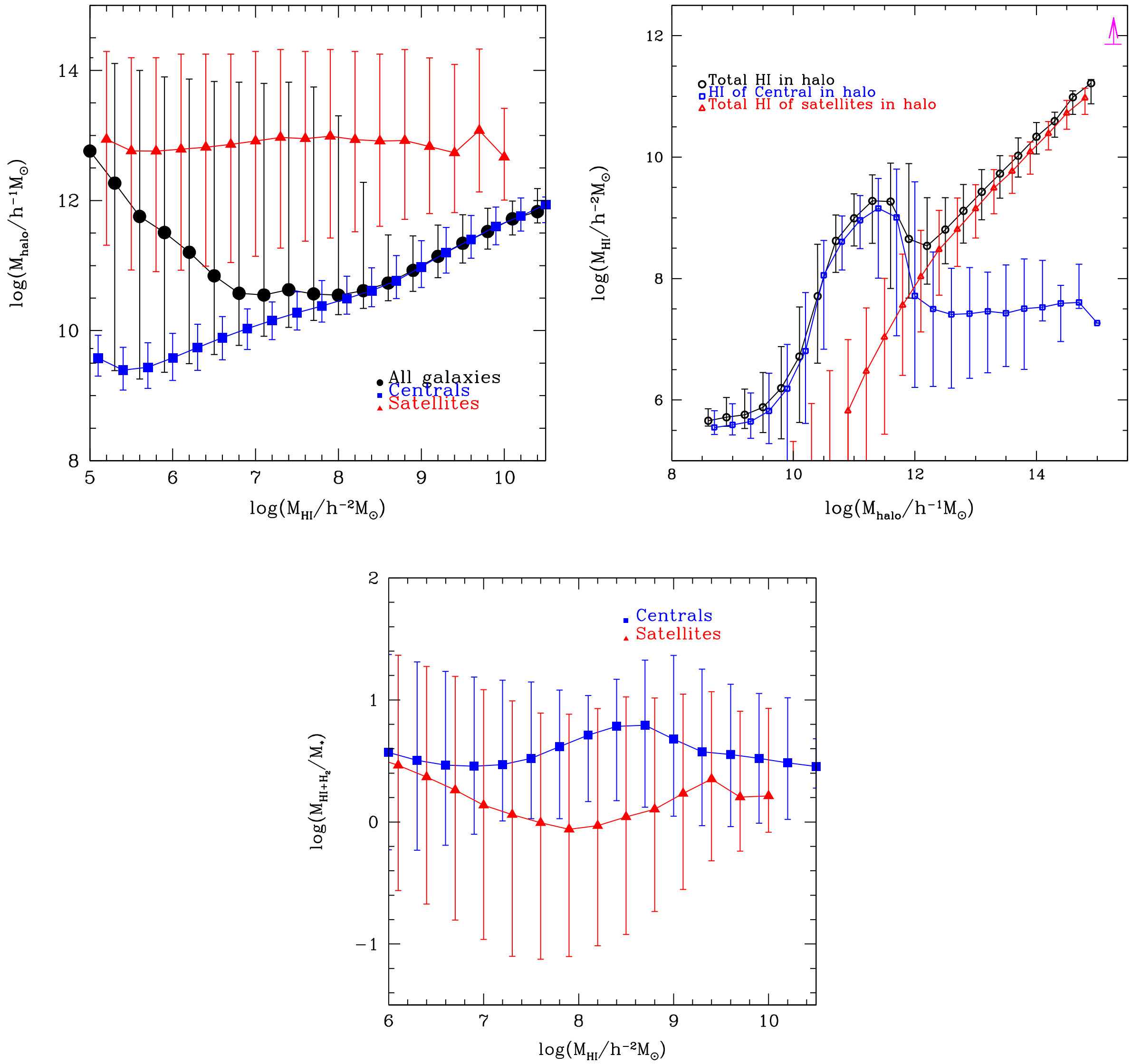

Figure 3. The top-left panel shows the host halo mass plotted in bins of H I mass at $z = 0$. The symbols show the median dark matter mass as a function of the H I mass of galaxies, for central galaxies (blue squares), satellite galaxies (red triangles) and all galaxies (black circles). The bars show the 10th–90th percentile ranges of the distributions of host dark matter halo masses. The top-right panel shows the total H I mass contained in a halo at $z = 0$. The symbols show the median total H I mass as a function of host dark matter halo mass, for central galaxies (open blue squares), all satellite galaxies (open red triangles) and all galaxies (open black circles) in a halo. The bars show the 10th–90th percentile ranges of the distributions of total H I masses. In this figure, we include all galaxies which have H I mass larger than $10^4\,h^{-1}\,\mathrm{M_\odot}$. The magenta arrow represents the lower limit of the overall H I mass of the Coma cluster by summing H I masses of 223 H I-detected late-type galaxies in Gavazzi et al. (2006). The total mass of the Coma cluster from Kubo et al. (2007) is shown for comparison. M_{halo} is the mass of the associated Dhalo. The bottom panel shows the neutral hydrogen gas fraction in bins of H I mass. The symbols show the median neutral hydrogen gas ($M_{\mathrm{H\,I+H_2}}$) fraction as a function of the H I mass of galaxies, for central galaxies (blue squares) and satellite galaxies (red triangles). $M_{\mathrm{H_2}}$ is the molecular hydrogen mass and $M_\star$ is the stellar mass. The bars show the 10th–90th percentile ranges of the distributions of neutral hydrogen gas fractions.

mind when considering clustering measurements made from them that they contain orders of magnitude fewer galaxies and sample smaller volumes than local optical surveys, such as the two-degree-Field Galaxy Redshift Survey (Colless et al. 2001), or Sloan Digital Sky Survey (York et al. 2000).

The left-hand panel of Fig. 4 shows the H I mass functions predicted by the Lagos2012 (red dotted line) and the Kim15 (black solid line) models. The Kim15 model applied a new scheme for photoionization feedback in order to improve the agreement of the predicted low-mass end of the H I mass function with the estimation from the ALFALFA survey at $z = 0$. This work is based on the model in Kim et al. (2015) which improves the modelling of the lower mass end of H I mass function. Here, we also show the ALFALFA observations for comparison to the model, which extend to low masses than the HIPASS data.

Surveys of H I-selected galaxies are not volume-limited and low H I mass galaxies are visible over a smaller volume than high H I mass galaxies. The blue dashed line in the bottom-left sub-panel shows the distribution of H I mass in the sample of Martin et al. (2010) used to estimate the H I mass function.

The sample used in Martin et al. (2010) from the ALFALFA survey has complete source extraction for 40 per cent of its total sky area. Martin et al. (2010) measure the H I mass function from a sample of ∼10 150 galaxies and apply both the $1/V_{\mathrm{max}}$ (Schmidt 1968)

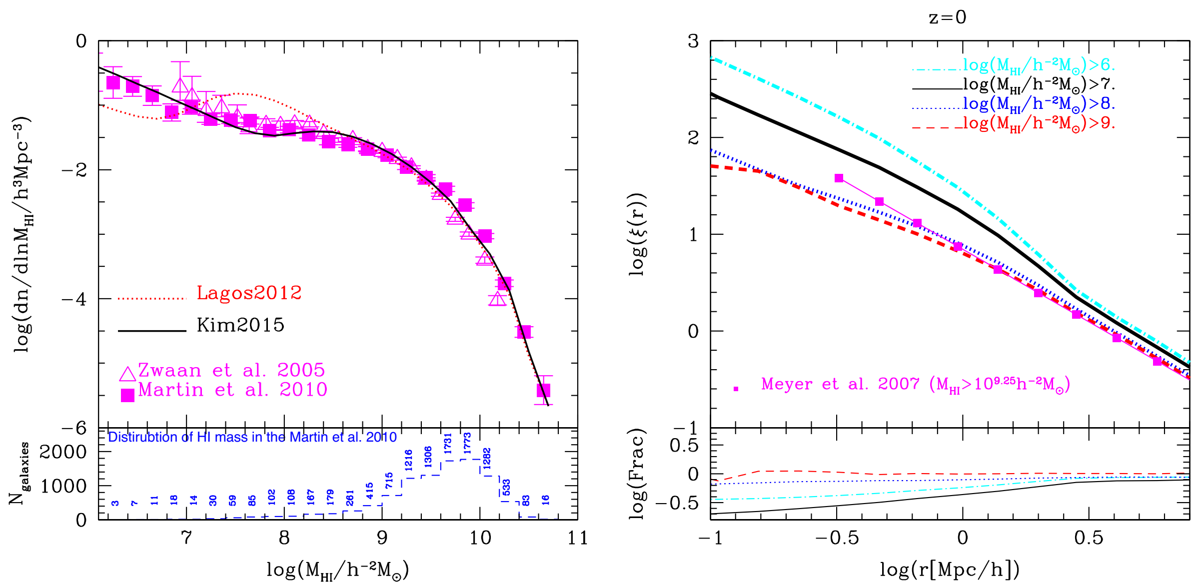

Figure 4. The left-hand panel shows the H I mass function at $z = 0$ predicted by the Lagos2012 (red dotted line) and the Kim15 (black solid line) models compared with local Universe observations from HIPASS (open triangles; cf. Zwaan et al. 2005) and ALFALFA (filled squares; cf. Martin et al. 2010). The bottom-left sub-panel shows the distribution of H I mass in the sample of Martin et al. (2010) to estimate the H I mass function. The right-hand panel shows the correlation function of different H I mass threshold samples at $z = 0$ for the Kim15 model. The bottom-right sub-panel shows the ratio between the predictions of the Lagos2012 and the Kim15 models for the same H I mass threshold samples labelled in the top-right sub-panel. The magenta filled squares connected by the solid line in the top-right sub-panel show the measurement of Meyer et al. (2007) using their best-fitting r_0 and γ values for galaxies with H I masses $>10^{9.25}\,h^{-2}\,\mathrm{M}_\odot$ from the local Universe.

and the Two-Dimensional Step-Wise Maximum Likelihood methods Zwaan et al. (2005). In this work, we use the H I mass function estimated using the $1/V_{\mathrm{max}}$ method when comparing with our model predictions.[4]

The right-hand panel of Fig. 4 shows the correlation functions for different H I mass thresholds in the Kim15 model, compared with observations from Meyer et al. (2007) using the best-fitting r_0 and γ values for galaxies with H I masses $>10^{9.25}\,h^{-2}\,\mathrm{M}_\odot$. The bottom-right sub-panel shows the ratio between the predictions of the Lagos2012 and the Kim15 models, for the same H I mass thresholds used in the second top sub-panel. These models differ in their treatment of photoionization feedback as described in Section 2. The right-hand panel of Fig. 4 shows that H I-selected galaxies down to $M_{\mathrm{H\,I}} > 10^7\,h^{-2}\,\mathrm{M}_\odot$ could constrain the properties of photoionization feedback (which is also imprinted on the H I mass function, Kim et al. 2015), and the contribution of satellite galaxies to the distribution of H I-selected galaxies from H I clustering measurements through the γ value of the correlation function.

Fig. 4 illustrates that observational H I clustering measurements which contain H I galaxies with $M_{\mathrm{H\,I}} < 10^8\,h^{-2}\,\mathrm{M}_\odot$ are necessary to constrain the modelling of galaxy formation, particularly how photoionization affects small galaxies Kim et al. (2015), and the contribution of satellite galaxies to the distribution of H I galaxies.

4 DISTRIBUTION OF NEUTRAL HYDROGEN AT HIGH REDSHIFTS

We now investigate the predicted clustering of H I-selected galaxies at high redshift in ongoing and future H I galaxy surveys.

Power et al. (2010) showed that the H I mass function undergoes little evolution from $z = 1$ to 0 (see also, Kim et al. 2011; Lagos et al. 2011a). This is due to a competition between gas cooling, the transition from H I to H_2 and star formation, together with mass ejection by winds. This is a generic feature predicted by galaxy formation models for high H I mass galaxies (see Power et al. 2010). However, this picture could break down for H I-poor galaxies, which reside in low-mass host dark matter haloes. These haloes are more sensitive to photoionization feedback, which hampers gas cooling. Here, we investigate the evolution of the H I mass function. The left-hand panel of Fig. 5 shows the predicted H I mass functions at $z = 0$, 0.2 and 0.5 (corresponding to maximum redshifts expected to be observed by the WALLABY[5] ($z \sim 0.2$), DINGO[6] ($z \sim 0.5$) and LADUMA[7] ($z \sim 0.5$)), together with the H I mass function from the ALFALFA survey at $z = 0$ for reference. The largest difference between the H I mass functions at $z = 0.2$ (or 0.5) and $z = 0$ is predicted to be for galaxies with masses in the range $10^7\,h^{-2}\,\mathrm{M}_\odot < M_{\mathrm{H\,I}} < 10^{8.5}\,h^{-2}\,\mathrm{M}_\odot$. The circular velocities of the host dark matter haloes of these galaxies are between ~ 30 km s^{-1} and ~ 50 km s^{-1} in the Kim15 model. These velocities correspond to galaxies which become affected by photoionization feedback between $z = 0.5$ and 0.

[4] The $1/V_{\mathrm{max}}$ method treats each individual galaxy by weighting the galaxy counts by the maximum volume $V_{\mathrm{max},\,i}$. This strategy allows the inclusion of low-mass galaxies (only detected in the nearby Universe) in the same sample as rare high-mass galaxies (only found in larger volumes). In addition, the weights may be adjusted in order to correct for a variety of selection effects, large-scale structure effects and missing volume within the data set.

[5] http://askap.org/wallaby; the Widefield ASKAP L−band Legacy All-sky Blind surveY.

[6] http://internal.physics.uwa.edu.au/~mmeyer/dingo/welcome.html; Deep Investigations of Neutral Gas Origins.

[7] http://www.ast.uct.ac.za/laduma/node/6; Looking at the Distant Universe with the MeerKAT Array.

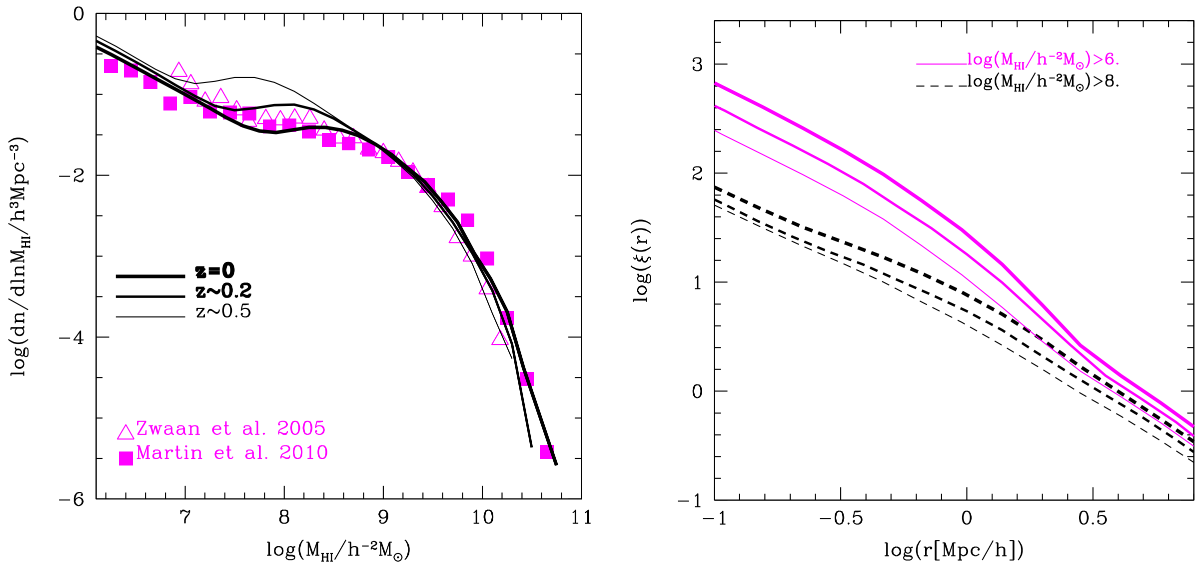

Figure 5. Left-hand panel: the H I mass function predicted by the Kim15 model at $z = 0$, 0.2 and 0.5 as labelled, compared with local Universe observations from HIPASS (open triangles; cf. Zwaan et al. 2005) and ALFALFA (filled squares; cf. Martin et al. 2010). Right-hand panel: the real-space correlation function predicted for two different H I mass thresholds samples at redshifts $z = 0$, 0.2 and 0.5, as labelled in the left-hand panel.

The right-hand panel of Fig. 5 shows the predicted correlation functions at $z = 0$, 0.2 and 0.5 for two different H I mass thresholds ($M_{\rm H\,I} > 10^6\,h^{-2}$ M$_\odot$ and $M_{\rm H\,I} > 10^8\,h^{-2}$ M$_\odot$). The correlation functions at higher redshift have lower correlation length (r_0) and slope (γ) than at lower redshift for the same H I mass thresholds (see Table 1). The parameters r_0 and γ were derived by minimizing χ^2 of predicted correlation functions of the form shown in e quation (4) over separations between 1 and 10 h^{-1} Mpc and considering γ values between 0 and 5. The separation range was chosen based on the approach in Meyer et al. (2007) and Martin et al. (2012), who used projected separations <10 h^{-1} Mpc and Papastergis et al. (2013) used projected separation between 0.5 and 10 h^{-1} Mpc.

Fig. 6 shows the projected correlation functions in the model by projecting galaxy samples on to the 2D x-y plane within the z range ∼40 h^{-1} Mpc by applying the distant-observer approximation. The points with error bars show an observational estimate made from the ALFALFA catalogue by Martin et al. (2010) at local Universe. The projected correlation functions from the model do not show a power-law shape. To see the effect of the power-law assumption, we estimate the r_0 and γ values following Meyer et al. (2007) using the projected correlation function from the model. We have also included the r_0 and γ values from these results in Table. 1 (in the square bracket) and Fig. 7 for $z = 0$ (connected open triangles by the dotted line) for comparison between the values from the real-space correlation function and extracted values from the projected correlation function in the model.

In the left-hand panel of Fig. 7, we show variations on the predicted correlation length (r_0) at different redshifts and H I mass thresholds. In agreement with the $z = 0$ predictions, we find that lower H I mass threshold samples show a larger correlation length than higher H I mass threshold samples at all redshifts in contrast to optically selected galaxy samples (cf. red galaxy samples show an increasing r_0 at faint luminosities). This is because the contribution of galaxies hosted by high mass host dark matter haloes increases with decreasing H I mass threshold (see Fig. 2).

In the right-hand panel of Fig. 7, we show the predicted correlation function slope (γ) for different redshifts and H I mass thresh-

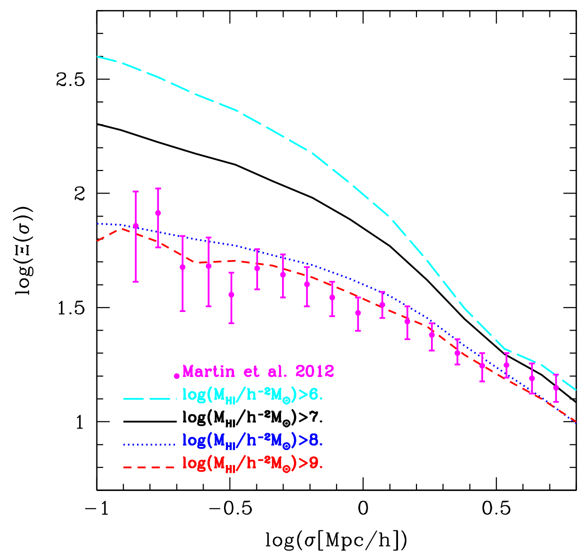

Figure 6. The projected galaxy correlation function at $z = 0$ for different H I mass threshold samples from the model. σ indicates the projected separations. The points with error bars show an observational estimate made from the ALFALFA catalogue by Martin et al. (2010) at local Universe.

olds. γ displays similar trends with redshift and H I mass threshold to r_0 (left-hand panel of Fig. 7). Samples selected using lower H I mass thresholds show steeper correlation functions compared to samples using higher H I mass thresholds due to the contribution of satellite galaxies hosted by host dark matter haloes of $M_{\rm halo} > 10^{14}\,h^{-1}$ M$_\odot$. In addition, the correlation length (r_0) and slope (γ) increase as redshift decreases from $z = 0.5$ to 0. This implies that galaxies at low redshift are more clustered than higher redshift galaxies selected above the same H I mass threshold, which is consistent with the hierarchical growth of structures. For comparison, the symbols with error bars show the best-fitting values of r_0 (the left-hand panel of Fig. 7) and γ (the right-hand panel of Fig. 7) estimated from observations in the local Universe of H I-selected for galaxies with $M_{\rm H\,I} > 10^{9.25}\,h^{-2}$ M$_\odot$ (Meyer et al. 2007), $M_{\rm H\,I}$

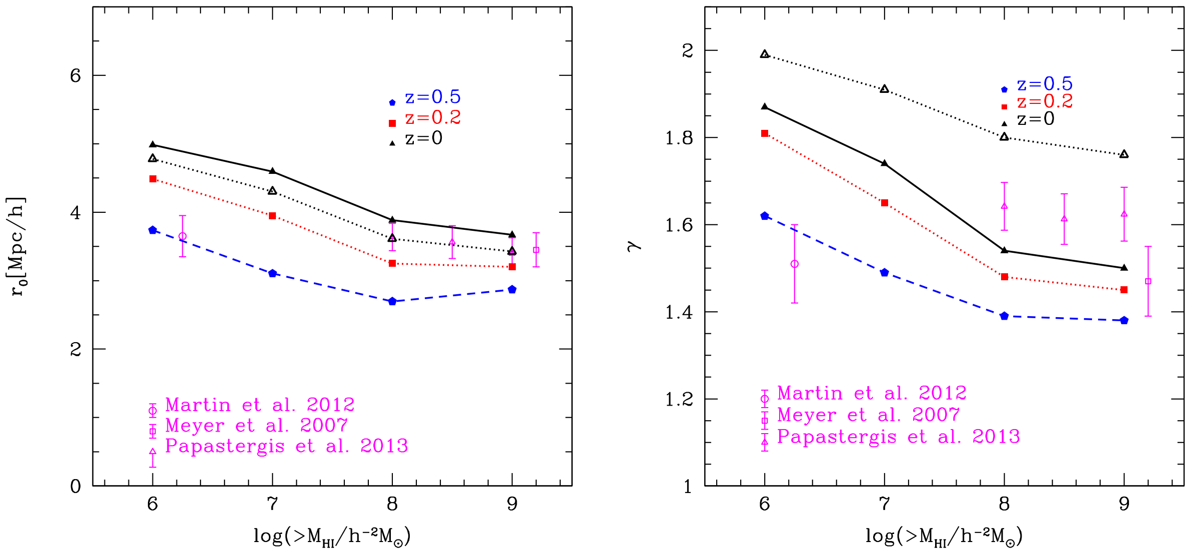

Figure 7. Left-hand panel: the correlation length (r_0) of the correlation function for different H I mass thresholds at $z = 0$, 0.2 and 0.5, as labelled. Right-hand panel: the slope (γ) of the correlation function for different H I mass thresholds at $z = 0$, 0.2 and 0.5, as labelled. For reference, the symbols with error bars (the standard deviation) show the best-fitting r_0 and γ values from observations of H I-selected galaxies measured by Meyer et al. (2007), Martin et al. (2012) and Papastergis et al. (2013) in the local Universe. The open triangles show the $r_{0,\,\mathrm{power}}$ (left-hand panel) and γ_{pow} (right-hand panel) from the projected correlation function of the model following Meyer et al. (2007).

$> 10^{6.2}\,h^{-2}$ M$_\odot$ (Martin et al. 2012) and $M_{\mathrm{H\,I}} > 10^8$, $10^{8.5}$ and 10^9 h^{-2} M$_\odot$ (Papastergis et al. 2013).

Note that the predictions are in reasonable agreement with the observations for $M_{\mathrm{H\,I}} > 10^8\,h^{-2}$ M$_\odot$. However, the model does not reproduce the clustering estimated by Martin et al. (2012) for samples with $M_{\mathrm{H\,I}} > 10^{6.2}\,h^{-2}$ M$_\odot$. The values of r_0 do not change significantly if the value of γ is held fixed at the value found in the observational study. On the other hand, we remind the reader that the observational clustering measurements come from small volumes, which can be affected by cosmic variance. Although Martin et al. (2012) attempted to correct for the selection function of the sample, to account for the fact that low H I mass galaxies are seen over a much smaller volume than high H I mass galaxies, the cleanest approach for measuring galaxy clustering is to use volume-limited samples. However, this is not yet feasible with current H I surveys as this requires many galaxies to be omitted from the analysis (see for example, the volume-limited samples used in an optical galaxy survey by Norberg et al. 2001). The assumption of a power law in the observed estimate of r_0 may also introduce a bias increasing the value of γ and/or decreasing r_0 value.

5 EFFECT OF GALAXIES WITH LOW H I MASSES ON THE SIGNAL FOR 21CM INTENSITY MAPPING

As noted above, clustering studies of low H I mass-selected galaxies will be restricted to $z \sim 0.5$, even within the deepest H I surveys planned (e.g. DINGO). An alternative way to study the clustering of galaxies down to low H I masses and at high redshifts ($0.5 < z < 3$) is through 21 cm intensity mapping, which measures the fluctuation in the 21 cm signal from unresolved H I-selected galaxies (Wyithe & Loeb 2007). This is a new strategy to economically map much larger volumes of the Universe, by measuring the collective 21 cm emission from all galaxies within some volume set by the angular resolution of the telescope and the frequency interval samples.

We predict the 21 cm intensity mapping signal using the Kim15 model. First, we divide the volume of the simulation into 256^3 cells (0.39 Mpc h^{-1} cell size corresponds to ~ 1 arcmin at $z = 1$, but the value is arbitrary). We then sum the mass of all H I in galaxies in each cell, and measure the resulting fluctuations of 21 cm intensity.

The power spectrum of 21 cm brightness fluctuations is defined in Wyithe & Brown (2010) as

$$P_{21\mathrm{cm}}(k, z) \simeq T_{\mathrm{b}}^2 x_{\mathrm{H\,I}}^2(z)\,\langle b(z)\rangle^2\, P_{\mathrm{DM}}(k, z), \qquad (5)$$

where $T_{\mathrm{b}} = 23.8\sqrt{(1+z)/10}$ mK. The term $< b(z) >$ is the H I mass-weighted halo bias, $P_{\mathrm{DM}}(k, z)$ is the power spectrum of the dark matter and $x_{\mathrm{H\,I}}(z)$ is the fraction of neutral atomic hydrogen $x_{\mathrm{H\,I}}(z) = \Omega_{\mathrm{H\,I}}(z)/0.76/\,\Omega_{\mathrm{b}}$, where $\Omega_{\mathrm{H\,I}}(z)$ is the neutral atomic hydrogen density and Ω_{b} the baryon density of the Universe. The predicted $\Omega_{\mathrm{H\,I}}(z)$ from the Kim15 model agrees well with observations up to $z \sim 2$ without any assumption about the relationship between H I mass and dark matter halo mass as was assumed in Bagla et al. (2010, see Kim et al. 2015 and Rhee et al. 2016). The predicted H I mass increases monotonically with the halo mass above $10^{12}h^{-1}$ M$_\odot$ and can be well described by a power law which is consistent with the zoom-in hydrodynamic simulation predictions in Villaescusa-Navarro et al. (2016). Villaescusa-Navarro et al. (2016) only show this relation above $M_{\mathrm{halo}} > 10^{13}$ M$_\odot$ because of the volume limit in their simulation. Equivalently, we can predict the intensity mapping signal using the power spectrum of the H I mass of galaxies which gives us information about the H I mass distribution from the model, with no assumptions about linear bias relative to the underlying dark matter distribution, $P_{\mathrm{H\,I}}(k, z)$,

$$P_{21\mathrm{cm}}(k, z) = T_{\mathrm{b}}^2 \Omega_{\mathrm{H\,I}}(z)/0.76/\,\Omega_{\mathrm{b}} \times P_{\mathrm{H\,I}}(k, z). \qquad (6)$$

Fig. 8 shows the predicted 21 cm brightness temperature power spectrum predicted by the Kim15 model for four different H I mass thresholds. The amplitude of the 21 cm brightness temperature power spectrum decreases as the H I mass threshold increases, while the slope of the 21 cm power spectrum increases with increasing H I mass threshold. We find that the 21 cm intensity mapping power

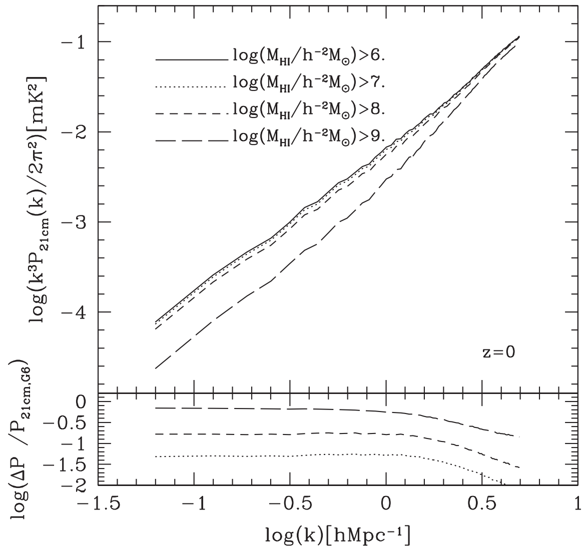

Figure 8. The predicted 21 cm brightness temperature power spectrum as a function of H I mass threshold, as labelled, at $z = 0$. We show the $\Delta P/P_{21\mathrm{cm,\,G6}}$ ratio in the bottom panel. $P_{21\mathrm{cm,\,G6}}$ is the 21 cm power spectrum using all galaxies with $M_{\mathrm{H\,I}} \geq 10^6\,h^{-2}\,\mathrm{M}_\odot$, and ΔP is the difference between power spectra of larger mass threshold and $P_{21\mathrm{cm,\,G6}}$.

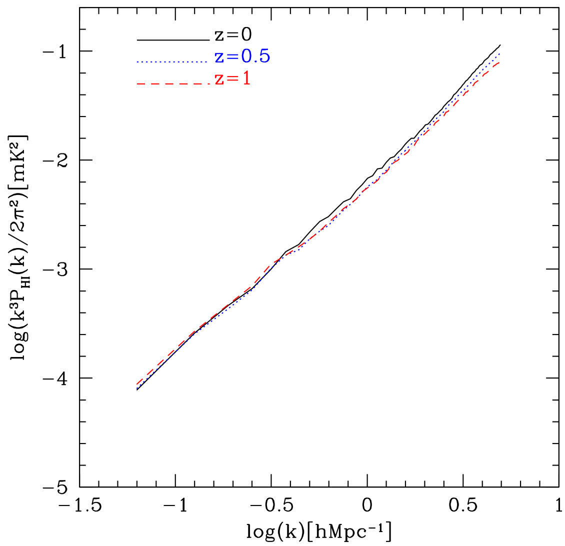

Figure 9. The predicted 21 cm brightness temperature power spectrum ($P_{21\mathrm{cm,\,G6}}$) for the Kim15 model at $z = 0$ (black solid line), 0.5 (blue dotted line) and 1 (red dashed line).

spectrum converges for $M_{\mathrm{H\,I}} > 10^7\,h^{-2}\,\mathrm{M}_\odot$. We also show the ratio $\Delta P\,/\,P_{21\mathrm{cm,\,G6}}$ in the sub-panel, where ΔP is the difference between a 21 cm power spectrum using the galaxies above a given H I mass threshold and the 21 cm power spectrum using all galaxies with $M_{\mathrm{H\,I}} \geq 10^6\,h^{-2}\,\mathrm{M}_\odot$ ($P_{21\mathrm{cm,\,G6}}$). Note that we choose the $P_{21\mathrm{cm,\,G6}}$ to compare with other H I mass thresholds because $M_{\mathrm{H\,I}} \sim 10^6\,h^{-2}\,\mathrm{M}_\odot$ is the minimum H I mass which can be trusted in the Kim15 model based on the Millennium-II simulation (which has a minimum halo mass of few times $10^8\,h^{-1}\,\mathrm{M}_\odot$).

We extend the prediction of the 21 cm brightness temperature fluctuation to higher redshifts in Fig. 9. There are three main factors driving the evolution of the 21 cm brightness temperature power spectrum. The first is $\Omega_{\mathrm{H\,I}}(z)$, which is the neutral hydrogen density. The second is the bias governed by the mass of host dark matter haloes hosting the H I-selected galaxies (r_0 and γ). The third is H I mass contributions from each galaxy (i.e. H I masses) to H I mass intensity fluctuations. The predicted value of $\Omega_{\mathrm{H\,I}}(z)$ at $z = 1$ is larger than at $z = 0.5$ and 0 (see fig. 9 in Kim et al. 2015). However, the clustering of host dark matter haloes hosting the H I-selected galaxies at $z = 0$ is stronger, particularly at small separations (compare Fig. 1 to Fig. 5 and see Fig. 7). Thus, these three main competing drivers lead to different amplitudes and slopes for the predicted 21 cm brightness temperature power spectra at different redshifts. The third driver results in the largest difference between the H I galaxy correlation function results and 21 cm intensity mapping power spectrum results.

We now focus on the contribution of unresolved galaxies. Power et al. (2010) showed that in the Millennium simulation, galaxies with $M_{\mathrm{H\,I}} < 10^8\,h^{-2}\,\mathrm{M}_\odot$ cannot be properly resolved (and so the completeness of the simulation drops below this H I mass). We explore the effect of unresolved galaxies on the 21 cm signals from intensity mapping (Fig. 10; see also Table 2). We first ignore H I masses in galaxies that reside in the dark matter haloes (both host and subhaloes) less massive than $1.72 \times 10^{10}\,h^{-1}\,\mathrm{M}_\odot$ in the Millennium-II simulation (corresponding to the dark matter halo mass resolution of the larger Millennium simulation). The difference in 21 cm brightness temperature fluctuations increases at higher redshifts (see long dashed lines in sub-panels of Fig. 10). This shows that low H I mass galaxies cannot be ignored if we aim to provide accurate predictions of the 21 cm brightness temperature fluctuations at higher redshifts.

The Green Bank Telescope has pioneered the detection of the large-scale structure using the technique of 21 cm intensity mapping Chang et al. (2010); Masui et al. (2013); Switzer et al. (2013). In the near future, more sensitive experiments will improve upon this work, including BINGO, CHIME and the SKA which will carry out 21 cm intensity mapping measurements at different redshifts. We show the predicted sensitivity of SKA in the second and third panels of Fig. 10 at $k \sim 0.1\,h\,\mathrm{Mpc}^{-1}$ based on calculations of Santos et al. (2015). This predicted sensitivity of the SKA is one of main drivers of the effort behind the modelling of the H I content of galaxies down to low H I mass galaxies ($M_{\mathrm{H\,I}} < 10^7\,h^{-2}\,\mathrm{M}_\odot$) using a high enough resolution simulation with a physically motivated galaxy formation model to properly include these low H I mass galaxies.

6 SUMMARY

We have predicted the clustering of H I galaxy samples as a function of H I mass threshold. The clustering of H I-selected galaxies is very similar for H I-selected galaxies with masses greater than $10^8\,h^{-2}\,\mathrm{M}_\odot$. This similarity can be explained in terms of the contribution of host dark matter haloes to the H I mass function. We estimate power-law fits for the predicted correlation functions and derive the correlation length r_0 and slope γ, as a function of redshift. We find that r_0 and γ increase as the H I mass threshold decreases, which is the contrary to the expectations for optically selected galaxy samples. In addition, the predictions of clustering for different redshifts show that the correlation length and slope increase as redshift decreases from $z = 0.5$ to 0. We calculate the contribution of low H I mass galaxies to 21 cm intensity mapping ($z = 0$, $\sim$0.5 and $\sim$1). We find that it is important to model the H I mass function down to low H I masses of $\sim 10^7\,h^{-2}\,\mathrm{M}_\odot$ in order to correctly predict shape and amplitude of forthcoming observations of power spectra of 21 cm intensity fluctuations. We also show that the importance of low H I mass galaxies to the 21 cm brightness fluctuations increases at higher redshifts. Dark matter halo mass resolution in simulations must be sufficient (at least better than

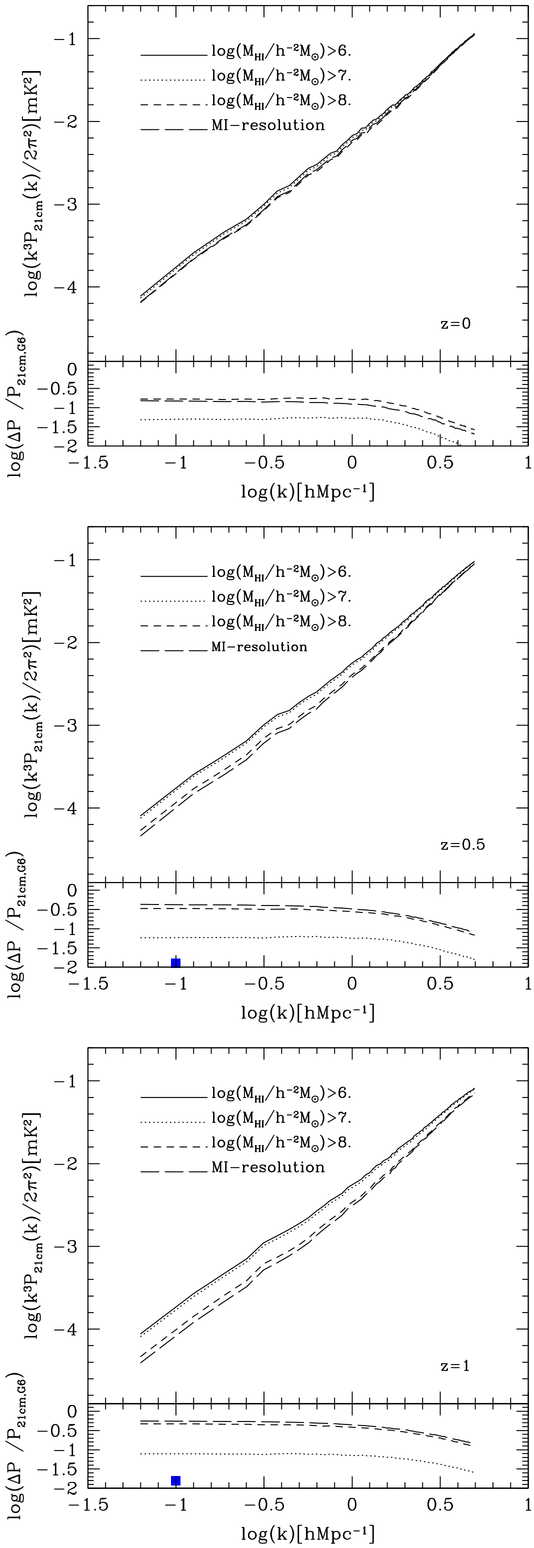

Figure 10. The predicted 21 cm brightness temperature power spectrum for the Kim15 model as a function of H I mass threshold at high redshifts, $z = 0$ (top panel), 0.5 (middle panel) and $z = 1$ (bottom panel). The long dashed line shows the measurement using the H I-selected galaxies in the host dark matter haloes which have the Millennium dark matter halo resolution imposed (labelled as MI-resolution). We show $\Delta P / P_{21\mathrm{cm,\,G6}}$, the fluctuation between thresholds and $P_{21\mathrm{cm,\,G6}}$, in the sub-panels. The filled squares in the middle and bottom panels show the predicted constraint (noise over signal) at $k \sim 0.1\ h\ \mathrm{Mpc}^{-1}$ expected for the SKA1-MID (Santos et al. 2015).

Table 2. The predicted dimensionless 21 cm brightness temperature $\Delta^2(k,z)[\mathrm{mk}^2] = k^3 P_{21\mathrm{cm}}(k,z)/2/\pi^2$ at $k_\mathrm{p} \sim 0.126\ h\,\mathrm{Mpc}^{-1}$ of different H I mass thresholds and the Millennium simulation dark matter resolution case at $z = 0$, 0.5 and 1 in the model.

H I mass threshold (or Millennium resolution)	$\Delta^2(k_\mathrm{p},0)$	$\Delta^2(k_\mathrm{p},0.5)$	$\Delta^2(k_\mathrm{p},1)$
$>10^6\ h^{-2}\ \mathrm{M}_\odot$	2.375e-4	2.5398e-4	2.674e-4
$>10^7\ h^{-2}\ \mathrm{M}_\odot$	2.283e-4	2.389e-4	2.463e-4
$>10^8\ h^{-2}\ \mathrm{M}_\odot$	2.056e-4	1.698e-4	1.422e-4
Millennium resolution	2.049e-4	1.479e-4	1.197e-4

$\sim 10^{10}\ h^{-1}\ \mathrm{M}_\odot$) in order to properly predict the 21 cm brightness temperature signal. With correct modelling of the distribution of H I-selected galaxies, 21 cm intensity mapping surveys will provide a window to understanding of the small mass galaxy formation physics.

ACKNOWLEDGEMENTS

We thank an anonymous referee for helpful comments on the manuscript. H-SK is supported by Discovery Early Career Researcher Awards from the Australian Research Council (DE140100940). CP thanks Simon Driver and Aaron Robotham for helpful discussions. CP is supported by DP130100117, DP140100198, and FT130100041. This work was supported by an STFC rolling grant at Durham. CL is funded by the ARC project DE150100618. CMB acknowledges receipt of a Research Fellowship from the Leverhulme Trust. The calculations for this paper were performed on the ICC Cosmology Machine, which is part of the DiRAC Facility jointly funded by the STFC, the Large Facilities Capital Fund of BIS, and Durham University. Part of the research presented in this paper was undertaken as part of the Survey Simulation Pipeline (SSimPL; http://www.astronomy.swin.edu.au/SSimPL/). The Centre for All-Sky Astrophysics is an Australian Research Council Centre of Excellence, funded by grant CE110001020.

REFERENCES

Bagla J. S., Khandai N., Datta K. K., 2010, MNRAS, 407, 567
Bahé Y. M., McCarthy I. G., 2015, MNRAS, 447, 969
Bandura K., Addison G. E., Amiri M., Bond J. R., Campbell-Wilson D., Connor L., Cliche J.-F., Davis G., 2014, in Stepp L. M., Gilmozzi R., Hall H. J., eds, Proc. SPIE Conf. Ser. Vol. 9145, Ground-based and Airborne Telescopes V. SPIE, Bellingham, p. 914522
Battye R. A. et al., 2012, preprint (arXiv:e-prints)
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
Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2002, MNRAS, 333, 156
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Blitz L., Rosolowsky E., 2006, ApJ, 650, 933
Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014a, A&A, 564, A66
Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Gavazzi G., Lagos C., Saintonge A., 2014b, A&A, 564, A67
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Brown M. J. I., Webster R. L., Boyle B. J., 2000, BAAS, 32, 1562
Campbell D. J. R. et al., 2015, MNRAS, 452, 852
Catinella B. et al., 2013, MNRAS, 436, 34
Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Colless M. et al., 2001, MNRAS, 328, 1039
Cortese L. et al., 2012, A&A, 540, A52
Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dénes H., Kilborn V. A., Koribalski B. S., 2014, MNRAS, 444, 667
Dijkstra M., Haiman Z., Rees M. J., Weinberg D. H., 2004, ApJ, 601, 666
Font A. S. et al., 2008, MNRAS, 389, 1619
Gavazzi G., O'Neil K., Boselli A., van Driel W., 2006, A&A, 449, 929
Guha Sarkar T., Mitra S., Majumdar S., Choudhury T. R., 2012, MNRAS, 421, 3570
Hoppmann L., Staveley-Smith L., Freudling W., Zwaan M. A., Minchin R. F., Calabretta M. R., 2015, MNRAS, 452, 3726
Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, MNRAS, 440, 2115
Johnston S. et al., 2008, Exp. Astron., 22, 151
Jonas J., 2007, Proc. Sci., MeerKAT science and technology. SISSA, Trieste, PoS(MRU)0077
Kim H.-S., Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., Power C., Schneider M., 2011, MNRAS, 414, 2367
Kim H.-S., Lacey C. G., Cole S., Baugh C. M., Frenk C. S., Efstathiou G., 2012, MNRAS, 425, 2674
Kim H.-S., Wyithe J. S. B., Raskutti S., Lacey C. G., Helly J. C., 2013a, MNRAS, 428, 2467
Kim H.-S., Power C., Baugh C. M., Wyithe J. S. B., Lacey C. G., Lagos C. D. P., Frenk C. S., 2013b, MNRAS, 428, 3366
Kim H.-S., Wyithe J. S. B., Power C., Park J., Lagos C. d. P., Baugh C. M., 2015, MNRAS, 453, 2315 (kim15)
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 699, 850
Kubo J. M., Stebbins A., Annis J., Dell'Antonio I. P., Lin H., Khiabanian H., Frieman J. A., 2007, ApJ, 671, 1466
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J., 2011a, MNRAS, 416, 1566
Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011b, MNRAS, 418, 1649
Lagos C. d. P., Bayet E., Baugh C. M., Lacey C. G., Bell T. A., Fanidakis N., Geach J. E., 2012, MNRAS, 426, 2142
Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Loh Y.-S. et al., 2010, MNRAS, 407, 55
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
Madgwick D. S. et al., 2003, MNRAS, 344, 847
Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359
Martin A. M., Giovanelli R., Haynes M. P., Guzzo L., 2012, ApJ, 750, 38
Masui K. W. et al., 2013, ApJ, 763, L20
Merson A. I. et al., 2013, MNRAS, 429, 556
Meyer M. J., Zwaan M. A., Webster R. L., Brown M. J. I., Staveley-Smith L., 2007, ApJ, 654, 702
Norberg P. et al., 2001, MNRAS, 328, 64
Norberg P. et al., 2002, MNRAS, 336, 907
Padmanabhan H., Choudhury T. R., Refregier A., 2015, MNRAS, 447, 3745
Papastergis E., Giovanelli R., Haynes M. P., Rodríguez-Puebla A., Jones M. G., 2013, ApJ, 776, 43
Power C., Baugh C. M., Lacey C. G., 2010, MNRAS, 406, 43
Rhee J., Lah P., Chengalur J. N., Briggs F. H., Colless M., 2016, MNRAS, 460, 2675
Santos M. et al., 2015, Proc. Sci., Cosmology from a SKA HI intensity mapping survey. SISSA, Trieste, PoS(AASKA14)019
Sarkar D., Bharadwaj S., Anathpindika S., 2016, MNRAS, 460, 4310
Schmidt M., 1968, ApJ, 151, 393
Serra P. et al., 2012, MNRAS, 422, 1835
Sobacchi E., Mesinger A., 2013, MNRAS, 432, L51
Springel V., 2005, MNRAS, 364, 1105
Switzer E. R. et al., 2013, MNRAS, 434, L46
Villaescusa-Navarro F., Viel M., Datta K. K., Choudhury T. R., 2014, J. Cosmol. Astropart. Phys., 9, 050
Villaescusa-Navarro F. et al., 2016, MNRAS, 456, 3553
Wolz L., Abdalla F. B., Blake C., Shaw J. R., Chapman E., Rawlings S., 2014, MNRAS, 441, 3271
Wyithe J. S. B., Brown M. J. I., 2010, MNRAS, 404, 876
Wyithe J. S. B., Loeb A., 2007, MNRAS, 375, 1034
Wyithe J. S. B., Warszawski L., Geil P. M., Oh S. P., 2009, MNRAS, 395, 311
York D. G. et al., 2000, AJ, 120, 1579
Young L. M. et al., 2011, MNRAS, 414, 940
Zehavi I. et al., 2011, ApJ, 736, 59
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

This paper has been typeset from a TEX/LATEX file prepared by the author.