Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **486,** 5124–5138 (2019)
Advance Access publication 2019 April 24

doi:10.1093/mnras/stz1118

The H I content of dark matter haloes at $z \approx 0$ from ALFALFA

Andrej Obuljen [1,2]★ David Alonso,[3,4] Francisco Villaescusa-Navarro,[5] Ilsang Yoon [6] and Michael Jones [7]

[1] *SISSA – International School for Advanced Studies, Via Bonomea 265, I-34136 Trieste, Italy*
[2] *INFN – National Institute for Nuclear Physics, Via Valerio 2, I-34127 Trieste, Italy*
[3] *School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF24 3AA, UK*
[4] *Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[5] *Center for Computational Astrophysics, 162 5th Ave, New York, NY 10010, USA*
[6] *National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA*
[7] *Instituto de Astrofísica de Andalucía, Glorieta de la Astronomía, Granada E-18008, Spain*

Accepted 2019 April 16. Received 2019 February 6; in original form 2018 May 10

ABSTRACT

We combine information from the clustering of H I galaxies in the 100 per cent data release of the Arecibo Legacy Fast ALFA survey, and from the H I content of optically selected galaxy groups found in the Sloan Digital Sky Survey to constrain the relation between halo mass M_h and its average total H I mass content $M_{\rm H\,I}$. We model the abundance and clustering of neutral hydrogen through a halo-model-based approach, parametrizing the $M_{\rm H\,I}(M_h)$ relation as a power law with an exponential mass cut-off. To break the degeneracy between the amplitude and low-mass cut-off of the $M_{\rm H\,I}(M_h)$ relation, we also include a recent measurement of the cosmic H I abundance from the α.100 sample. We find that all data sets are consistent with a power-law index $\alpha = 0.48 \pm 0.08$ and a cut-off halo mass $\log_{10} M_{\rm min}/(h^{-1}{\rm M}_\odot) = 11.18^{+0.28}_{-0.35}$. We compare these results with predictions from state-of-the-art magnetohydrodynamical simulations, and find both to be in good qualitative agreement, although the data favours a significantly larger cut-off mass that is consistent with the higher cosmic H I abundance found in simulations. Both data and simulations seem to predict a similar value for the H I bias ($b_{\rm H\,I} = 0.878^{+0.022}_{-0.023}$) and shot-noise power ($P_{\rm SN} = 94^{+20}_{-18}\,[h^{-1}\,{\rm Mpc}]^3$) at redshift $z = 0$.

Key words: galaxies: haloes – large-scale structure of Universe – methods: data analysis – cosmology: observations.

1 INTRODUCTION

The Λ cold dark matter (ΛCDM) model has become the most successful theory framework that is able to explain a wide variety of cosmological observations, from the temperature and polarization anisotropies in the cosmic microwave background (CMB; Planck Collaboration XIII 2016) to the spatial distribution of galaxies at low redshift (Alam et al. 2017). Some of the free parameters of this model are connected with open questions in fundamental physics, such as possible deviations from a pure cosmological constant, the cosmic abundance of dark matter, or the sum of neutrino masses. The main aim of modern cosmological experiments is to determine the value of those parameters with the best possible combination of precision and accuracy.

In this endeavour, the statistics of the matter distribution contains an enormous amount of information to potentially constrain the value of these cosmological parameters. Unfortunately, the matter distribution is not directly observable, but can only be inferred through tracers of it such as galaxies, quasars, and cosmic neutral hydrogen. In particular, 21 cm intensity mapping (Battye, Davies & Weller 2004; McQuinn et al. 2006; Chang et al. 2008; Loeb & Wyithe 2008; Wyithe & Loeb 2008; Peterson et al. 2009; Bagla, Khandai & Datta 2010; Battye et al. 2013; Masui et al. 2013; Switzer et al. 2013; Bull et al. 2015; Anderson et al. 2018) has recently become one of the main contenders in the quest to map out the three-dimensional cosmic density field out to the highest possible redshifts. In spite of the significant observational challenges of this technique, mostly associated with the presence of strong and complex radio foregrounds (Santos, Cooray & Knox 2005; Wolz et al. 2014; Alonso et al. 2015; Shaw et al. 2015; Wolz et al. 2015), intensity mapping (IM) offers a unique way to produce fast and economical, three-dimensional maps of the overdensity of neutral hydrogen (H I) in the Universe. For this reason, intensity mapping has been put forward as an ideal method to probe cosmology on large scales.

However, the properties of H I, especially in terms of clustering, are still not fully understood. This is due to a number of reasons:

★ E-mail: andrej.obuljen@uwaterloo.ca

the early stage of IM as an observational probe, the difficulty of detecting the faint 21cm line for a sufficiently large number of sources at high redshifts, and the possibly conflicting evidence (Castorina & Villaescusa-Navarro 2017) coming from observations of low-redshift H I surveys (Zwaan et al. 2005a; Martin et al. 2012), the Lyman-α forest (Zwaan et al. 2005b; Noterdaeme et al. 2012; Zafar et al. 2013; Crighton et al. 2015) and the clustering of damped Lyman α systems (Rao, Turnshek & Nestor 2006; Pérez-Ràfols et al. 2018). Understanding H I is vital both for cosmology and astrophysics, since it also plays a vital role in understanding the star formation history (Kennicutt 1998).

At linear order, the amplitude of the 21 cm power spectrum at redshift z is proportional to the product of the H I bias $b_{\rm H\,I}(z)$ and its cosmic abundance $\Omega_{\rm H\,I}(z) = \rho_{\rm H\,I}(z)/\rho_c(z=0)$, where $\rho_{\rm H\,I}(z)$ is the mean H I density at redshift z and $\rho_c(z=0)$ is the critical density at $z=0$. While the value of $\Omega_{\rm H\,I}(z)$ is relatively well constrained in the redshift range $z \in [0, 5]$ by several observations (Rao, Turnshek & Nestor 2006; Zwaan & Prochaska 2006; Lah et al. 2007; Martin et al. 2010; Songaila & Cowie 2010; Braun 2012; Noterdaeme et al. 2012; Delhaize et al. 2013; Rhee et al. 2013; Crighton et al. 2015), the value of the H I bias is poorly known (Basilakos et al. 2007; Martin et al. 2012; Guo et al. 2017). Thus, a better model of the H I bias would allow us to (1) improve our understanding of the astrophysical processes governing the abundance and evolution of H I across time, (2) improve the design of future 21 cm experiments and optimize their main science cases, and (3) produce more accurate forecasts for the constraining power of these observations. One of the indirect goals of this paper is to measure the H I bias at $z \approx 0$.

In the absence of better data, the halo model (Smith et al. 2003) offers an alternative method to predict the abundance and clustering of H I after including two extra ingredients: a relation between total halo mass and H I mass $M_{\rm H\,I}(M_h)$, and a model for the distribution of H I within each halo $\rho_{\rm H\,I}(r|M_h)$. However, these extra degrees of freedom must be constrained using available data before this method can be useful to predict the cosmic H I signal. This has been done in the past by combining low-redshift data from H I surveys and column-density information from observations of the Lyman α forest at higher redshifts (Castorina & Villaescusa-Navarro 2017; Padmanabhan, Refregier & Amara 2017), often revealing apparent tensions between data sets. In this paper, we will use a self-consistent framework to constrain the $M_{\rm H\,I}$–M_h relation using the mass-weighed clustering of H I galaxies detected by the Arecibo Legacy Fast ALFA survey (ALFALFA), as well as their abundance in haloes extracted from galaxy groups found in the Sloan Digital Sky Survey (SDSS) galaxy survey. We will also explore the possibility of constraining the shape of the H I profile and the impact of modelling assumptions on our results.

This paper is organized as follows. In Section 2, we describe the theoretical framework we use to characterize the abundance and clustering of H I. We outline the data employed in this work in Section 3. The methods used to analyse the data and compare with the theory predictions are illustrated in Section 4. The main results of this work are shown in Section 5. We discuss the results and summarize the conclusions of this work in Section 6.

2 H I HALO MODEL

Numerical simulations show that almost all the H I in the post-reionization Universe is inside dark matter haloes (Villaescusa-Navarro et al. 2014, 2018). Thus, one can use the halo-model formalism (Cooray & Sheth 2002; Smith et al. 2003) to study the abundance and clustering of cosmic neutral hydrogen. The purpose of this paper is to constrain the H I-mass-to-halo-mass relation $M_{\rm H\,I}(M_h)$ from direct measurements in selected galaxy groups, as well as from the clustering of H I sources. Extending the halo model to predict the properties of H I requires additional assumptions about the relation between the H I mass and the halo mass as well as the distribution of H I itself inside haloes. We follow a prescription similar to that developed recently by Padmanabhan et al. (2017) and Castorina & Villaescusa-Navarro (2017).

We start by assuming that, on average, the H I content of haloes depends solely on their mass, and we parametrize the $M_{\rm H\,I}(M_h)$ relation as (Castorina & Villaescusa-Navarro 2017; Padmanabhan et al. 2017; Villaescusa-Navarro et al. 2018)

$$M_{\rm H\,I}(M_h) = M_0 \left(\frac{M_h}{M_{\rm min}}\right)^{\alpha} \exp\left(-\frac{M_{\rm min}}{M_h}\right). \quad (1)$$

In this model, the overall normalization M_0 can be immediately associated with the cosmic H I fraction $\Omega_{\rm H\,I} \equiv \rho_{\rm H\,I}/\rho_c$ at $z=0$, where ρ_c is the critical density. Both quantities are related through

$$\Omega_{\rm H\,I} \equiv \frac{\bar{\rho}_{\rm H\,I}}{\rho_{\rm c}} = \frac{1}{\rho_{\rm c}} \int_0^{\infty} {\rm d}M_h \, n(M_h) \, M_{\rm H\,I}(M_h), \quad (2)$$

where $n(M_h)$ is the halo mass function. The other two free parameters of the model are α, which describes the scaling of $M_{\rm H\,I}$ with halo mass, and the low-mass cut-off $M_{\rm min}$, which represents the threshold mass needed for a halo to host H I. This mass cut-off is expected, since the gravitational potential of small haloes is not deep enough to trigger the clustering and cooling of the hot gas heated by the UV background (Villaescusa-Navarro et al. 2018).

On small scales, the clustering of H I is dominated by its distribution within the halo (i.e. the so-called one-halo term). Although our constraints will be based solely on the shape of the correlation function on larger scales, we use two different models for the H I density profile, in order to quantify the effect of this assumption on the final results:

(i) *Altered NFW profile*: this is the model introduced and used in Maller & Bullock (2004), Barnes & Haehnelt (2014), Padmanabhan et al. (2017), and Villaescusa-Navarro et al. (2018) and assumes the radial profile of the form:

$$\rho_{\rm H\,I}(r|M_h) \propto (r + 3/4r_s)^{-1}(r + r_s)^{-2}, \quad (3)$$

where $r_{\rm s}$ is the scale radius of the H I cloud, and is related to the halo virial radius $R_{\rm v}(M_h)$ by the concentration parameter – $c_{\rm H\,I}(M_h, z) \equiv R_{\rm v}(M_h)/r_{\rm s}$. We follow Bullock et al. (2001) and Macciò et al. (2007) and use a mass-dependent concentration parameter given by

$$c_{\rm H\,I}(M_h, z=0) = 4\, c_{\rm H\,I,0} \left(\frac{M_h}{10^{11}{\rm M}_{\odot}}\right)^{-0.109}. \quad (4)$$

.

(ii) *Exponential profile*: this is the model implemented in Padmanabhan et al. (2017), and given by

$$\rho_{\rm H\,I}(r|M_h) \propto \exp(-r/r_{\rm s}). \quad (5)$$

In both cases the proportionality factors are automatically fixed by requiring that the H I mass be given by the volume integral of the density profile up to the halo virial radius $R_{\rm v}(M)$.

$$M_{\rm H\,I}(M_h) = 4\pi \int_0^{R_v} {\rm d}r \, r^2 \rho_{\rm H\,I}(r|M_h). \quad (6)$$

Thus, both profiles are described by one additional free parameter, $c_{\rm H\,I,0}$. The normalized H I density profile in Fourier space for the

altered NFW profile is given in Padmanabhan et al. (2017, see their equation A3), while the exponential profile is simply

$$u_{\rm H\,I}(k|M_h) = \frac{1}{(1+k^2 r_s^2)^2}. \quad (7)$$

.

The halo model prediction (Castorina & Villaescusa-Navarro 2017; Padmanabhan et al. 2017; Villaescusa-Navarro et al. 2018) for the H I power spectrum, is given by the sum of a one-halo and a two-halo term:

$$P_{\rm H\,I,1h}(k) = F_2^0(k), \qquad P_{\rm H\,I,2h}(k) = P_{\rm lin}(k)\left[F_1^1(k)\right]^2, \quad (8)$$

$$F_\beta^\alpha(k) \equiv \int n(M_h) b^\alpha(M_h) \left[\frac{M_{\rm H\,I}(M_h)}{\bar{\rho}_{\rm H\,I}} u_{\rm H\,I}(k|M_h)\right]^\beta {\rm d}M_h, \quad (9)$$

where $n(M_h)$ is the halo mass function, $b(M_h)$ is the halo bias and $P_{\rm lin}(k)$ is the linear matter power spectrum. For the halo mass function and bias, we use the parametrizations of Tinker et al. (2010), derived from numerical simulations, and we adhere to halo masses defined by a spherical overdensity parameter $\Delta = 180$

$$M_h = \frac{4\pi}{3}\rho_c \Omega_{\rm m} \Delta R_{\rm v}^3. \quad (10)$$

Finally, our basic clustering data vector is the 2D projected correlation, given by the projection of the 3D correlation function along the line of sight. This can be computed directly from the power spectrum as

$$\Xi(\sigma) = \int_{-\infty}^{\infty} {\rm d}\pi\, \xi(\pi,\sigma) = \int_0^\infty \frac{k\,{\rm d}k}{2\pi}\left[P_{\rm H\,I,1h}(k) + P_{\rm H\,I,2h}(k)\right] J_0(k\sigma), \quad (11)$$

where $J_0(x)$ is the order-0 cylindrical Bessel function. To accelerate the computation of $\Xi(\sigma)$, we made use of FFTLog (Hamilton 2000).

Our theoretical model therefore depends on four free parameters $\theta = \{M_0, M_{\rm min}, \alpha, c_{\rm H\,I,0}\}$. We fix all cosmological parameters to values compatible with the latest ΛCDM constraints measured by Planck (Planck Collaboration XIII 2016) ($H_0 = 70\,{\rm km\,s^{-1}Mpc^{-1}}$, $\Omega_{\rm m} = 0.3075$, $n_s = 0.9667$, $\sigma_8 = 0.8159$).[1]

3 DATA

3.1 The α.100 data set

The Arecibo Fast Legacy ALFA (Arecibo *L*-band Feed Array) survey, or ALFALFA[2] (Giovanelli et al. 2005), is a blind extragalactic H I survey performed using the Arecibo radio telescope. The main goal of ALFALFA is to quantify and study the properties of the H I content of the local Universe ($z \lesssim 0.05$). It represents a significant improvement over previous H I surveys, with a beam FWHM of ∼3.5 arcmin, an rms noise of ∼2.4 mJy and a spectral resolution of $\sim 10\,{\rm km\,s^{-1}}$.

Previous clustering analyses of the ALFALFA samples used the 40 per cent (Haynes et al. 2011; Martin et al. 2012; Papastergis et al. 2013), and 70 per cent (Guo et al. 2017) data releases (labelled α.40 and α.70). Our analysis makes use of the final data release (Haynes et al. 2018), containing ∼31 500 sources up to a redshift of $z = 0.06$ and covering approximately 7000 deg^2 in two continuous regions at either side of the Galactic plane. Sources with good detection significance ($S/N > 6.5$), classified as 'code-1', represent the main sample (∼81 per cent of the total 31 502 sources). Most of the remaining sources, classified as 'code-2', correspond to lower signal-to-noise detections ($S/N > 4.5$) with known optical counterparts. The remaining ∼5 per cent of the catalogue is mostly composed of high-velocity clouds of galactic provenance. We use only code-1 sources in the clustering analysis described in Section 4.1, and both code-1 and code-2 objects in the direct measurement of the H I content of galaxy groups (Section 4.2). For each source, the catalogue provides information about their angular coordinates, heliocentric radial velocity, radial velocity in the CMB frame, 21cm flux, line width and H I mass. H I masses for all objects can also be obtained from their distance and 21 cm flux as

$$m_{\rm H\,I} = (2.356 \times 10^5 M_\odot)\, D^2\, S_{21}, \quad (12)$$

where D is the distance to the source in Mpc, S_{21} is the integrated flux in units of Jy km s^{-1}, and $m_{\rm H\,I}$ is the source's H I mass.[3]

In the clustering analysis, the radial velocities $v_{\rm cmb}$ are used to assign radial distances to sources through their redshift $z_{\rm cmb} = v_{\rm cmb}/c$, using the cosmological parameters listed in Section 2. Due to the radio frequency interference (RFI), we make additional cuts and following Papastergis et al. (2013) we remove the sources outside $700\,{\rm km\,s^{-1}} < cz_{\rm cmb} < 15000\,{\rm km\,s^{-1}}$. After performing these cuts in the raw data, we are left with 24 485 code-1 sources and 5365 code-2 sources. Fig. 1 shows the angular distribution of all sources used in this work. The black lines delineate the survey boundaries used for in the clustering analysis. These cuts further reduce the clustering sample to 23 438 objects.

3.2 The SDSS group catalogue

To assign the H I-detected sources to dark matter haloes, we cross-match the SDSS galaxies and the ALFALFA sources and determine the group membership of the cross-matched galaxies using a galaxy group catalogue, following the procedure described in Yoon & Rosenberg (2015). We use the SDSS DR7 group catalogue[4] updated from the DR4 group catalogue (Yang et al. 2007). The catalogue uses galaxies in the SDSS DR7 spectroscopic sample with $0.01 \leq z \leq 0.2$ and redshift completeness $C > 0.7$. The group finding algorithm has been extensively tested using mock galaxy redshift survey catalogues and has proven to be successful in associating galaxies that reside in a common halo (Yang et al. 2005). In particular, this halo-based group finder works well for poor groups and identifies groups with only one member (i.e. isolated galaxies). The group halo masses are determined down to $M_h = 10^{11.8} h^{-1}{\rm M}_\odot$ using two methods: ranking by luminosity and from the stellar mass of member galaxies. Although we used the luminosity-ranked group halo mass, the results do not change if the stellar-mass-ranked halo mass is used instead. The group finder has been shown to correctly select more than 90 per cent of the true haloes with $M_h \geq 10^{12} h^{-1}{\rm M}_\odot$ (Yang et al. 2007), which allows us to reliably

[1] We fix the expansion rate to $h = 0.7$ instead of its best-fitting measurement $h = 0.6774$ to match the choice made in Jones et al. (2018) to measure $\Omega_{\rm H\,I}$. We will report our final results as a function of $h_{70} \equiv H_0/70\,{\rm km\,s^{-1}Mpc^{-1}}$.

[2] http://egg.astro.cornell.edu/alfalfa/

[3] To distinguish between the H I mass of ALFALFA sources and the total H I mass associated to a given dark matter halo, we label the latter $M_{\rm H\,I}$ and the former $m_{\rm H\,I}$.

[4] http://gax.sjtu.edu.cn/data/Group.html

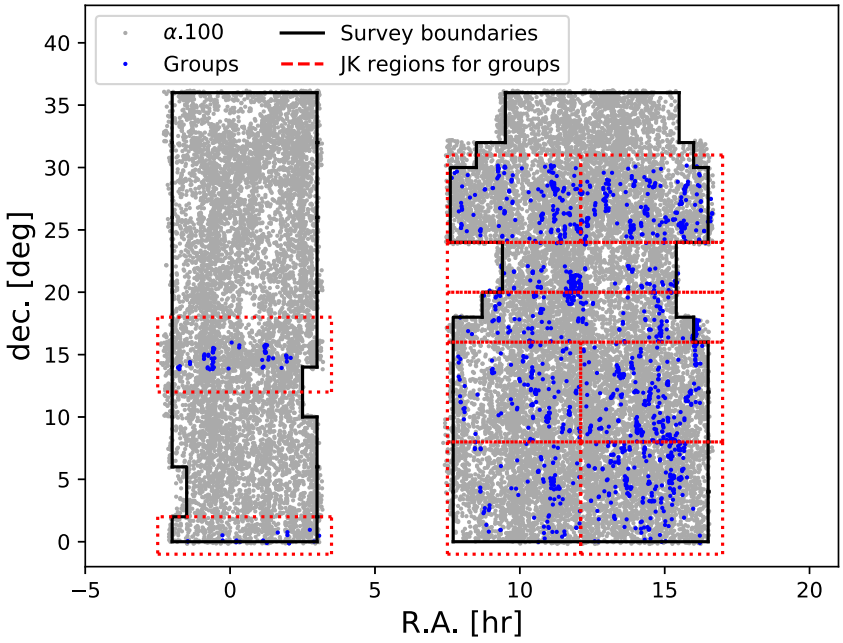

Figure 1. Sky distribution of the H I-selected galaxies from α.100 sample (grey dots). The black lines show the survey boundaries used in our clustering analysis (in which all sources outside the boundaries were omitted). The H I sources associated with groups in the SDSS DR7 group catalogue are highlighted in blue. The dotted red lines show the jackknife regions used to estimate the cosmic variance uncertainties of the H I mass function in groups (see Section 4.2).

study our galaxy samples within groups and clusters with halo mass $10^{12.50}\,h^{-1}\mathrm{M}_\odot \leq M_h \leq 10^{15.04}\,h^{-1}\mathrm{M}_\odot$.

For the virial radius of groups with halo mass M_h, we adopt the radius R_{180} that encloses an overdensity $\Delta = 180$ times larger than mean density (Yang et al. 2007):

$$R_{180} = 1.26h^{-1}\mathrm{Mpc}\left(\frac{M_h}{10^{14}h^{-1}\mathrm{M}_\odot}\right)^{1/3}(1+z_{\mathrm{group}})^{-1}, \quad (13)$$

which is based on the WMAP3 cosmological model parameters Spergel et al. (2007), $\Omega_\mathrm{m} = 0.238$, $\Omega_\Lambda = 0.762$ and $H_0 = 100h$ $\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, where $h = 0.73$. While these parameters differ slightly from those used in this study, this does not significantly impact the results at the low redshifts of our sample ($z < 0.055$). We also note that the DR7 group catalogue has significant overlap only with the 70 per cent ALFALFA data release, and therefore no new information is gained by using the complete ALFALFA sample (α.100 data set).

Fig. 1 shows, in blue, the ALFALFA sources identified as members in the group catalogue, as well as the jackknife regions used to compute the cosmic variance uncertainties for our estimate of the H I mass function in groups (dotted red lines, see Section 4.2).

4 METHOD

We derive constraints on the H I content of dark matter haloes by using the clustering properties of H I galaxies weighted by their H I content, as well as direct measurements of the H I content of galaxy groups. We describe the procedures used to compile these two data vectors and their associated covariances here.

As discussed in Section 1, our main interest is to quantify the properties of the total H I density inhomogeneities, since these are the relevant proxy of the density fluctuations measured by 21 cm intensity mapping. To do so, our main assumption will be that the properties of the full H I density field can be inferred from the properties of H I-selected sources as measured by ALFALFA when weighed by their H I mass. This simplifying assumption should be a good approximation as long as the sources detected by ALFALFA account for a significant portion of the total H I mass. The validity of this assumption can be quantified to some extent by examining the measurements of the H I mass function measured by the ALFALFA collaboration in Jones et al. (2018), extrapolating it below the detection limit. This calculation shows that, for a conservative threshold of $m_{\mathrm{H\,I,lo}} = 10^8\,\mathrm{M}_\odot$, less than 5 per cent of the total H I would lie in sources not observed by ALFALFA. Thus, assuming that the tilt of the H I mass function does not vary sharply on smaller masses, the contribution from diffuse or undetected sources to the observables considered here is negligible given the uncertainties in our measurements. This is even more so for measurements of the H I clustering, given that the clustering bias of H I sources has been shown to be only weakly dependent on H I mass (Papastergis et al. 2013). Even in the case of the measurement of the H I content in galaxy groups (see Section 4.2), where this contribution can rise to $\lesssim$30 per cent, we will explicitly show that the impact of the missing H I mass on our results is minimal.

4.1 The projected two-point correlation function

Previous studies (Martin et al. 2012; Papastergis et al. 2013; Guo et al. 2017) have measured two-point correlation function (2PCF) of H I-selected galaxies to determine their relation with the underlying dark matter density field. These studies have found that this sample has a low value of the clustering amplitude compared to the dark matter field (i.e. H I-selected galaxies have a low bias – $b_{\mathrm{H\,I,g}}$). Under the assumption described above, the same measurement can be performed on the 2PCF of H I-selected galaxies weighed by their H I mass to obtain a measurement of the total H I bias $b_{\mathrm{H\,I}}$, which plays a key role on 21 cm intensity mapping studies. We

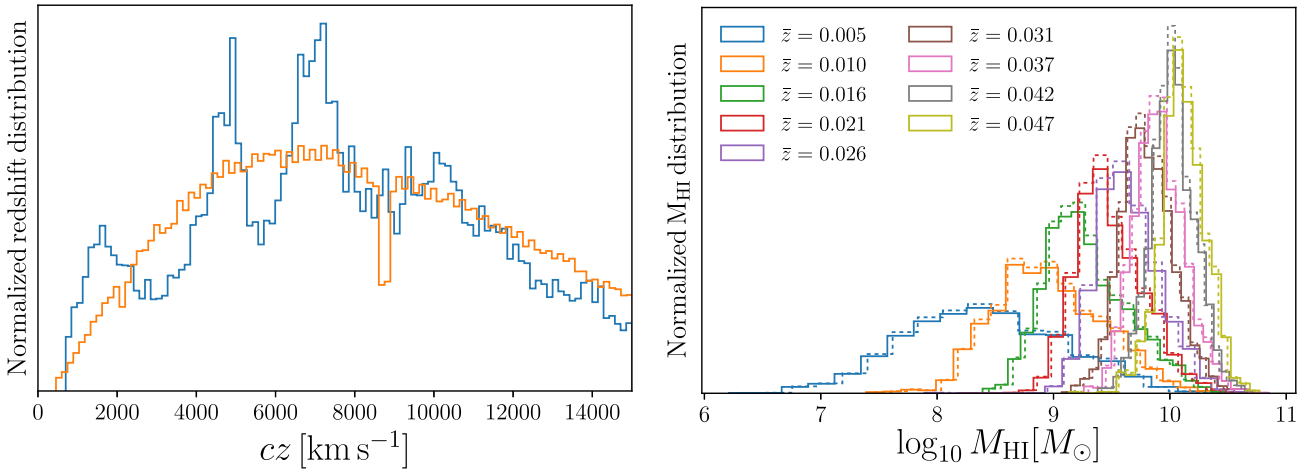

Figure 2. *Left:* normalized redshift distribution in the data (blue) and the constructed random catalogue (orange). *Right:* the H I mass distribution in the data (solid line) and the constructed random catalogue (dashed line) in different redshift bins (see legend).

describe the procedure used to estimate the 2PCF and its uncertainty here.

We begin by estimating the 2D 2PCF $\xi(\pi, \sigma)$ as a function of the distance between pairs of objects along the line of sight (π) and in the transverse direction (σ). For this, we use the Landy and Szalay estimator (Landy & Szalay 1993), given by

$$\xi(\pi, \sigma) = \frac{\mathrm{DD}(\pi, \sigma) - 2\mathrm{DR}(\pi, \sigma) + \mathrm{RR}(\pi, \sigma)}{\mathrm{RR}(\pi, \sigma)}, \quad (14)$$

where DD is the normalized histogram of unique weighted pairs of sources separated by a distance (π, σ) found in the data catalogue:

$$\mathrm{DD}(\pi, \sigma) = \frac{\sum_{i=1}^{N} \sum_{j>i} w_i w_j \, \Theta(\pi_{ij}; \pi, \Delta\pi) \, \Theta(\sigma'_{ij}; \sigma, \Delta\sigma)}{\sum_{i=1}^{N} \sum_{j>i} w_i w_j}.$$

Here, π_{ij} is the distance between the i-th and j-th objects along the line of sight (and similarly for the transverse distance σ_{ij}), and $\Theta(x \in (x_1, x_2)) = 1$ when $x \in (x_1, x_2)$ and 0 otherwise. RR is defined similarly for unique pairs of objects belonging to a random catalogue with statistical properties similar to those of the data (e.g. in terms of spatial and weights distribution) but no intrinsic clustering. Finally, DR is given by all pairs of data-random objects. The weights w_i assigned to each object are described below.

4.1.1 Random catalogue

The random catalogue needed to compute the correlation function should follow the same redshift, angular and weights distribution observed in the data. We use the area cuts reported in Jones et al. (2018) to define the survey footprint These are shown in Fig. 1, as black lines, and we discard all sources outside these boundaries. The angular positions of the random objects are then generated by drawing random coordinates with a constant surface density within this area.

We assign redshifts to the random objects by accounting for both the radial selection function described in Papastergis et al. (2013, see their fig. 4) and for RFI incompleteness, using the completeness function presented in the same paper (see their fig. 6). We achieve this by keeping a point with distance d in the random catalogue with a probability corresponding to the product of the selection and RFI completeness functions at d. The final normalized redshift distribution in both the data and the random catalogue is shown in the left-hand panel of Fig. 2.

The points in the random catalogue must also be assigned mass weights following the same $m_{\mathrm{H\,I}}$ distribution as the data. To achieve this, we split the random and the data set in 10 redshift bins. In each redshift bin, we give each random point an H I mass randomly sampled from the data in the same bin. The resulting H I mass distributions are shown in the right-hand panel of Fig. 2.

4.1.2 Weights

The sample we use is not volume-limited, and the objects near the peak of the selection function will dominate the measured correlation function. In order to avoid this, we apply optimal pairwise weights $w_{i,j} = w_i \times w_j$, where w_i is given by Peebles (1980) and Feldman, Kaiser & Peacock (1994)

$$w_i = \frac{m_{\mathrm{H\,I,i}}}{1 + 4\pi n(d_i) J_3(r_{ij})}, \quad (15)$$

where $n(d_i)$ is the number density of the sample at the distance d_i to the i-th source, r_{ij} is the comoving separation between both objects, and J_3 is an integral over the real-space isotropic correlation function:

$$J_3(r) = \int_0^r r'^2 \, \xi(r') \mathrm{d}r'. \quad (16)$$

Implementing these weights requires an assumption about the shape and amplitude of $\xi(r)$. For these, we follow Martin et al. (2012) and use $\xi(r) = (r/r_*)^{-1.51}$, with $r_* = 3.3\,h^{-1}$ Mpc. In fact, we find that fixing $J_3(r)$ to $J_3(r = 38\,h^{-1}\,\mathrm{Mpc}) = 2962\,\mathrm{Mpc}^3$ is enough to obtain a close-to-optimal correlation function (see Fig. 3). When implementing these weights, we approximated the number density as $n(d) = n_0 \exp(-(d/d_0)^\gamma)$ where $n_0 = 0.23(h^{-1}\,\mathrm{Mpc})^{-3}$, $d_0 = 31.18\,h^{-1}$ Mpc, and $\gamma = 0.99$. These numbers were obtained by fitting the distance distribution of objects in the random catalogue. Note also that equation (15) already includes the $m_{\mathrm{H\,I}}$ weights needed to recover the clustering properties of the total H I density.

Using this formalism, the measurement of the correlation function was carried out using the code `CUTE` (Alonso 2012). We adopted a logarithmic binning in σ in the range $\sigma \in [0.11, 52)\,h^{-1}$ Mpc with $\Delta \log_{10} \sigma/(h^{-1}\,\mathrm{Mpc}) = 0.12$, and we used

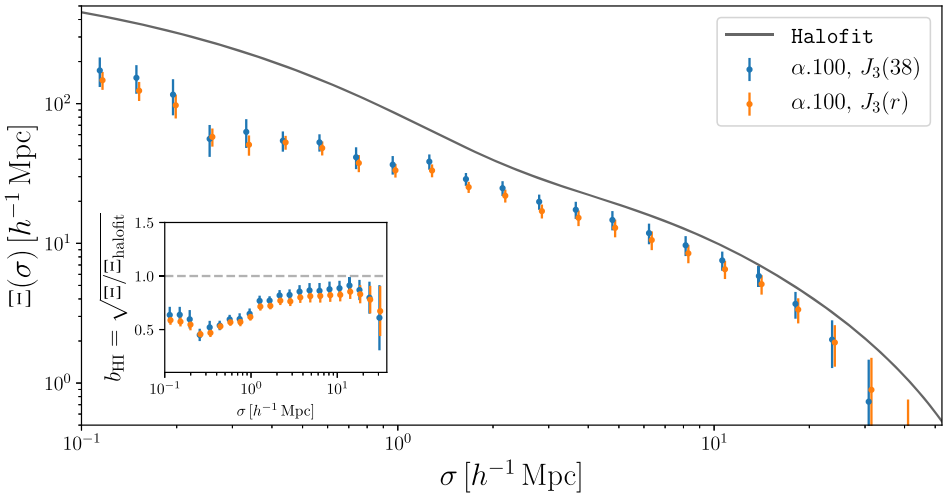

Figure 3. 2D projected correlation function. The points with error bars show $m_{\rm HI}$-weighted correlation function computed for the $\alpha.100$ data set, while the black solid line shows the `HaloFit` prediction for the matter correlation function at $z = 0$. Orange points show the measurements using pairwise weights that depend explicitly on the pair separation (equation 15), while the blue points correspond to the case of fixing $J_3(r)$ to $J_3(38\,h^{-1}\,\mathrm{Mpc})$, independent of separation. The impact of the choice of weighting scheme is found to be negligible. The inset shows the scale-dependent H I bias $b_{\rm HI}$ as a ratio of the measurement with respect to the matter correlation function. Orange points have been slightly shifted to the right.

59 linear bins of π in the range $\pi \in [0.5, 59.5)\ h^{-1}\,\mathrm{Mpc}$. In order to eliminate the effect of redshift-space distortions and be able to compare our measurements with the real-space theoretical prediction, we compute the projected correlation function $\Xi(\sigma)$ by integrating $\xi(\pi, \sigma)$ along the line of sight:

$$\Xi(\sigma) = \int_{-\infty}^{\infty} \mathrm{d}\pi\, \xi(\sigma, \pi) \simeq 2 \sum_{0}^{\pi_{\rm max}} \xi(\sigma, \pi) \Delta\pi, \tag{17}$$

where, as in Martin et al. (2012), we used $\pi_{\rm max} = 30\,h^{-1}\,\mathrm{Mpc}$.

Fig. 3 shows the measured H I-mass-weighted, projected correlation function (points with error bars) together with the prediction for the projected correlation function of the total matter overdensity, obtained from the `HaloFit` model for the matter power spectrum (Takahashi et al. 2012). The scale-dependent H I bias is shown in the inset of the same figure as the square root of the ratio of both quantities. The measured $b_{\rm HI}$ is in good agreement with the measurement of the bias of H I-selected galaxies presented in Martin et al. (2012). This is to be expected, given the observation that the clustering of H I sources shows little or no dependence on H I mass.[5]

4.1.3 *Covariance matrix*

We estimate the uncertainties on the measured projected correlation function using the jackknife resampling method (Lupton 1993; Zehavi et al. 2002). We divide the survey footprint into $N = 156$ contiguous patches covering $\sim$40 deg^2 each. We remove one patch at a time and measure the projected correlation function in the remaining area. The jackknife estimate of the covariance matrix is

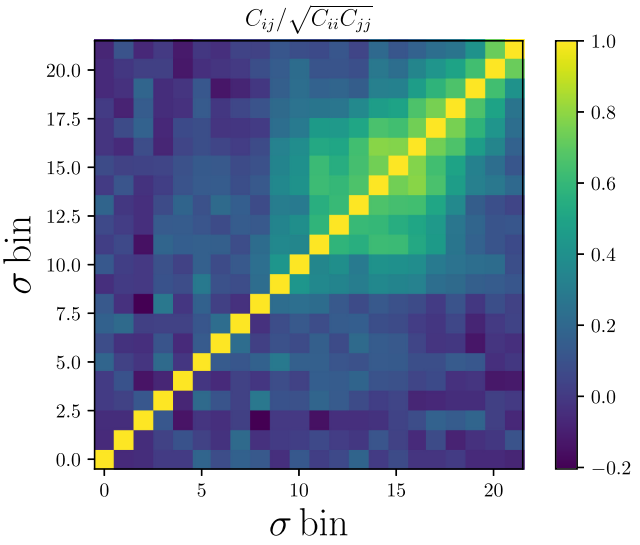

Figure 4. Jackknife correlation matrix for the projected two-point correlation function. We use 22 logarithmic bins in the transverse separation σ in the range $\sigma \in [0.11, 30.8)\,h^{-1}\,\mathrm{Mpc}$.

then given by

$$C_{ij} = \mathrm{Cov}\left(\Xi_i, \Xi_j\right) = \frac{N_s - 1}{N_s} \sum_{p=0}^{N_s} \left(\Xi_i^p - \bar{\Xi}_i\right)\left(\Xi_j^p - \bar{\Xi}_j\right). \tag{18}$$

Here, Ξ_i^p is the correlation function measured in the i-th bin after omitting the p-th patch and $\bar{\Xi}_i$ is the average of Ξ_i^p over all patches. Fig. 4 shows resulting correlation matrix $r_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$.

Ultimately, we are interested in the inverse covariance matrix. The inverse of the jackknife covariance is a biased estimate of the true inverse covariance, and we correct for this bias with an overall normalization factor (Hartlap, Simon & Schneider 2007):

$$C^{-1} \rightarrow \frac{N_{\rm s} - N_{\rm b} - 2}{N_{\rm s} - 1}\, C^{-1}, \tag{19}$$

[5]Note that Guo et al. (2017) observe a significant dependence on H I mass above $10^9\,\mathrm{M}_\odot$. This possible dependence at high masses, however, does not alter our assumption that the ALFALFA sources can be used to study the properties of the overall H I distribution, including all structures below ALFALFA's detection limit.

where $N_s = 156$ is the number of jackknife samples and $N_b = 22$ is the number of σ bins used in the analysis.

4.2 H I content in groups

As described in Section 3.2, we also include direct constraints on the $M_{\rm HI}(M_h)$ relation in our analysis, coming from the matching of ALFALFA sources to optical members of galaxy groups with calibrated halo mass detected in the SDSS group catalogue. To minimize a potential bias due to the incomplete coverage of the sky-projected area for each group, we estimate volume-correction factors for few large groups near the ALFALFA survey boundary. An estimate of the H I mass of each group is made by directly summing the masses of all ALFALFA member sources and applying the corresponding area correction factor, which is almost negligible for most of the groups. In general, this estimate of the group H I mass would be biased low, since the estimator will miss all ALFALFA sources with no optical counterparts lying in the comoving volume of each group, as well as any diffuse or unresolved H I component. The first cause of this bias (the sources with no optical detections) should have a negligible impact on this study, since it affects only ∼6 per cent (Haynes et al. 2011) of all ALFALFA sources, and most of those are expected to be galactic high-velocity clouds, and not extragalactic in nature. To quantify and minimize the impact of contributions from undetected H I components, we estimate the H I mass function (i.e. the $m_{\rm HI}$ distribution of ALFALFA sources) in bins of group halo mass. The exact procedure is as follows:

(i) We separate the SDSS group catalogue into seven logarithmically spaced bins of halo mass in the interval $\log_{10}M_h/(h^{-1}{\rm M}_\odot) \in [12.50, 15.04]$. The top panel of Fig. 5 shows the number of H I sources lying in each of these mass bins.

(ii) In each bin, we estimate the H I mass function $\phi(m_{\rm HI})$ using all the member sources found in the ALFALFA data set. For this, we use the 2D stepwise maximum likelihood (2DSWML) estimator described below.

(iii) In order to extrapolate below the detection limit, we model the measured mass function as a Schechter function with the form

$$\phi(m_{\rm HI}) = \ln(10)\,\phi_*\left(\frac{m_{\rm HI}}{M_*}\right)^{\alpha_s+1}\exp\left(-\frac{m_{\rm HI}}{M_*}\right). \quad (20)$$

(iv) For each halo mass bin, we compute the corresponding H I mass (and its uncertainty) by integrating over the reconstructed H I mass function, propagating all uncertainties as described below. We also compute a second estimate of the H I mass by integrating over the measured, model-independent 2DWSML mass function. This can only be done within the range of H I masses covered by ALFALFA, and the comparison of these two estimates then allows us to quantify the systematic uncertainty associated with undetected H I sources.

(v) We account for the finite width of the halo mass bins and for the uncertainty in the halo mass measurements by including both effects in the theory prediction. To do so, our predicted H I mass for a given halo mass bin is given by

$$M^b_{\rm HI} = \frac{\int {\rm d}\log_{10} M_h\, n(M_h)\, p(\log_{10} M_h|b)\, M_{\rm HI}(M_h)}{\int {\rm d}\log_{10} M_h\, n(M_h)\, p(\log_{10} M_h|b)}, \quad (21)$$

where $p(\log_{10}M_h|b)$ is the probability that a halo with mass M_h be in bin b. We model the errors in M_h to be Gaussianly distributed in

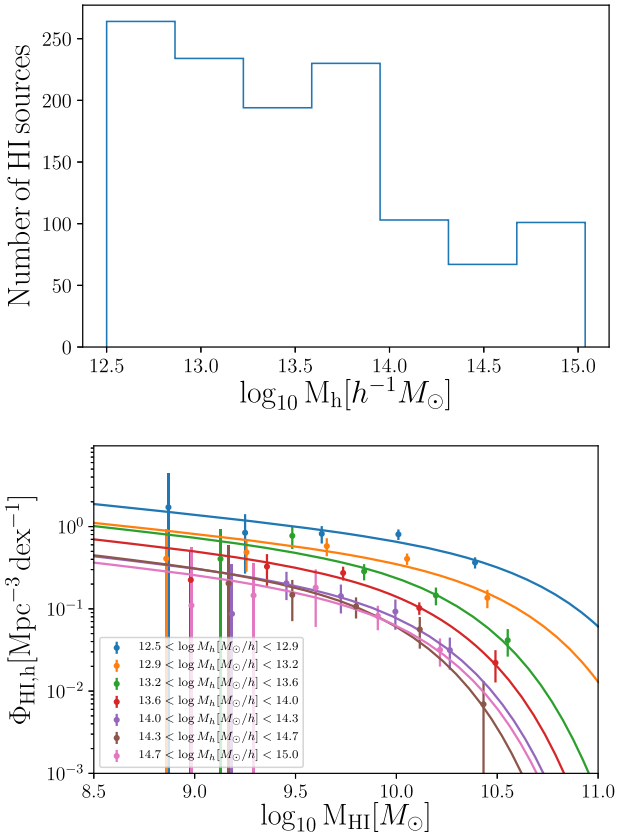

Figure 5. *Top:* the number of H I sources in the SDSS group catalogue lying in each halo mass bin after RFI and 50 per cent completeness cuts. *Bottom:* H I mass functions estimated from the SDSS group catalogue using 2DSWML method in different halo mass bins (see legend).

$\mu \equiv \log_{10}M_h$, in which case

$$p(\mu|b) = \frac{1}{2(\mu^f_b - \mu^i_b)}\left[{\rm erf}\left(\frac{\mu-\mu^i_b}{2\sigma}\right) - {\rm erf}\left(\frac{\mu-\mu^f_b}{2\sigma}\right)\right],$$

where (μ^i_b, μ^f_b) are the edges of the halo mass bin. For the halo mass scatter σ we use $\sigma = 0.3$ dex (Yang et al. 2007). We find both effects to be small, producing shifts in the final parameters below the 1σ level.

The list of reconstructed H I masses as a function of group halo mass is then appended to the correlation function described in the previous section to form the total data vector.

4.2.1 *The 2DSWML mass function estimator*

The idea of stepwise maximum-likelihood estimators has been applied in the past to reconstruct the luminosity function from a magnitude-limited sample (Efstathiou, Ellis & Peterson 1988; Cole et al. 2001; Jones et al. 2016). The method is non-parametric, modelling the luminosity function as sum of top-hat functions, and finding their amplitudes by maximizing the likelihood of the observed sample. The latter is possible interpreting the luminosity function as a probability distribution. The same logic was applied by Martin et al. (2010) and Jones et al. (2018) to estimate the H I mass function of ALFALFA sources, with the added complication that the completeness of the sample depends on both the H I flux S_{21} and

the 21cm line width W_{50}. This gives rise to the two-dimensional stepwise maximum-likelihood estimator (2DSWML), which we describe briefly here. To simplify the notation, we will define here $\mu \equiv \log_{10} m_{\rm HI}/{\rm M}_\odot$ and $w \equiv \log_{10} W_{50}/{\rm km\,s^{-1}}$.

The probability that a source g is detected with mass μ_g and line width w_g at distance d_g (within an interval $\Delta\mu,\ \Delta w$) is given by

$$p_g = \frac{\phi(\mu_g, w_g)\,\Delta\mu\,\Delta w}{\int_{-\infty}^{\infty} {\rm d}w \int_{\mu_{\rm lim}(d_g,w)}^{\infty} {\rm d}\mu\,\phi(\mu, w)}, \tag{22}$$

where $\phi(\mu, w)$ is the joint distribution of H I masses and line widths. Let us now model $\phi(\mu, w)$ as a 2D stepwise function, taking constant values in intervals of μ and w. Then, maximizing the log-likelihood $\mathcal{L} = \prod_g p_g$, we obtain an expression for the best-fitting amplitudes $\phi_{i,j}$ in the i-th interval of μ and the j-th interval of w:

$$\phi_{i,j} = n_{i,j} \left[\sum_g \frac{H_{g,ij}}{\sum_{i',j'} H_{g,i'j'}\phi_{i',j'}}\right]^{-1}, \tag{23}$$

where g runs over all sources in the sample, $n_{i,j}$ is the number of galaxies in bin (i, j), and $H_{g,ij}$ is the mean completeness of the sample in that bin for sources at a distance $d = d_g$. The completeness function was determined as described in Martin et al. (2010). We imposed a hard cut on m and w, using only bins with completeness > 50 per cent. We verified that our results did not vary significantly with more stringent completeness cuts.

Note that equation (23) gives $\phi_{i,j}$ recursively as a function of itself, and in practice $\phi_{i,j}$ is found through an iterative process. Once a converged solution for $\phi_{i,j}$ has been found, the H I mass function is obtained by marginalizing over W_{50}:

$$\phi_i = \sum_j \phi_{i,j}\,\Delta w. \tag{24}$$

Finally, this method is able to determine $\phi_{i,j}$ up to an overall normalization constant. We fix this by matching the integral of $\phi(m_{\rm HI},\ W_{50})$ to the total number of ALFALFA sources in each halo mass bin divided by the comoving volume covered by the corresponding haloes, as described in appendix B of Martin et al. (2010).

The bottom panel of Fig. 5 shows the estimated H I mass functions in each halo mass bin used in this analysis, together with their best-fitting Schechter models. For this figure, the mass functions were normalized dividing by the total volume enclosed within the virial radii of all groups in each halo mass bins. Note that, since we only use $\phi(m_{\rm HI})$ to estimate the $M_{\rm HI}(M_h)$ relation, our results are independent of this volume, and only depend on the total number of H I sources and galaxy groups in each M_h bin.

4.2.2 *Error propagation*

The uncertainties in the $M_{\rm HI}(M_h)$ relation inferred from the H I richness of groups, as described above, are driven by the errors in our estimate of the mass function in each M_h bin. Four main sources of uncertainty contribute to these errors (Jones et al. 2018), and we account for them as follows:

(i) *Poisson*: with each measurement of ϕ_i we associate a Poisson-counting error given by $\sigma(\phi_i) = \phi_i/\sqrt{N_i}$, where N_i is the number of sources contributing to the i-th $m_{\rm HI}$ bin.

(ii) *Sample variance*: the uncertainty associated with the stochastic variation in ϕ_i induced by the particular density fluctuations covered by the survey volume of ALFALFA was quantified through the jackknife resampling method described in Section 4.1. In this case, we use the 10 jackknife regions shown as red dotted lines in Fig. 1.

(iii) *Mass measurement errors*: the H I mass of each source is inferred from its 21 cm flux and its radial comoving distance. Both quantities have associated measurement uncertainties that propagate into $m_{\rm HI}$, shifting sources between different H I mass bins. To account for this, we generated 100 random realizations of the $\alpha.70$ catalogue by adding a random Gaussian error to the distances and fluxes of all sources (with a standard deviation given by their estimated error). We re-computed the H I masses and corresponding $\phi(m_{\rm HI})$ for each realization (see equation 12), and estimate the uncertainty associated to these errors from the scatter of all realizations.

(iv) *Line width measurement errors*: errors in W_{50} also affect our measurement of the 2DSWML mass function, by shifting sources between different W_{50} bins. The associated uncertainties were estimated from 100 random realizations, following the same procedure described above for mass measurement errors.

We added the errors associated with these four sources in quadrature to find the final uncertainties on ϕ_i.

Once ϕ_i and its uncertainties have been measured, we find the best-fitting Schechter models in each M_h bin. To avoid overfitting, given the relatively small number of points in which we estimate the mass function for each bin, we fix the tilt of the Schechter function to its best-fitting value for the overall H I mass function as reported by Jones et al. (2018), $\alpha_s = -1.25$. The best-fitting Schechter functions in each M_h bin are shown as solid lines in Fig. 5.

To estimate the uncertainties in the Schechter parameters (ϕ_*, M_*), we sample their likelihood running a Markov chain Monte Carlo (MCMC). For any point (ϕ_*, M_*) in these chains, the corresponding H I mass for haloes in the b-th M_h bin can be estimated as

$$\begin{aligned} M^b_{\rm HI} &= \frac{V_b}{N^b_{\rm group}} \int_0^\infty \phi(m_{\rm HI})\, m_{\rm HI}\, {\rm d}\log_{10} m_{\rm HI} \\ &= \frac{V_b \phi_*\, M_*}{N^b_{\rm group}}\, \Gamma\,(2+\alpha)\,, \end{aligned} \tag{25}$$

where V_b is the uncorrected volume spanned by all groups in the b-th M_h bin and $N^b_{\rm group}$ is the corresponding number of groups. Our final estimate of the $M_{\rm HI}(M_h)$ relation (and its uncertainty) from the galaxy group data is then given by the mean of $M_{\rm HI}$ (and its scatter) across all points in the MCMC chain. Finally, we correct our results for self-absorption as described in Jones et al. (2018). The results are shown as orange points with error bars in Fig. 6.

Since our measurement of $M^b_{\rm HI}$ involves extrapolating the H I mass function to very small masses, below the ALFALFA detection limit at the group's redshift, it is worth quantifying the impact of this extrapolation on our results. We do so here by comparing the fiducial measurement of $M^b_{\rm HI}$ described above, with two alternative estimates:

(i) The first estimator is given by directly integrating the measured 2DSWML mass function over the available range of H I masses in ALFALFA. Labelling the 2DSWML in the b-th halo mass bin as ϕ^b_i, this alternative estimate is given by

$$\tilde{M}^b_{\rm HI} = \frac{V^b}{N^b_{\rm group}} \sum_i \phi^b_i\, 10^{\mu_i}\, \Delta\mu. \tag{26}$$

The uncertainty on $\tilde{M}^b_{\rm HI}$ can be estimated trivially from the uncertainties on ϕ^b_i. Since $\tilde{M}^b_{\rm HI}$ and $\hat{\phi}^b_i$ are linearly related, the uncertain-

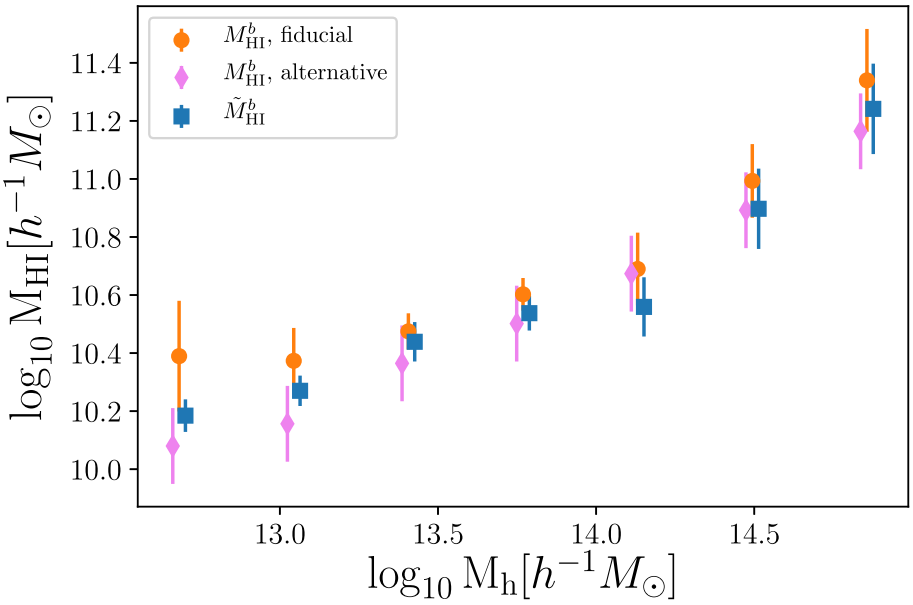

Figure 6. Estimated total $M_{\rm H\,I}$ in each halo mass bin obtained from the H I mass functions using three different methods. The total $M_{\rm H\,I}$ estimated by fitting the H I mass functions using the Schechter parametrization and accounting for the missing H I mass are shown with orange points. The error bars are computed by propagating the Schechter parameter uncertainties. The blue squares show the results of directly integrating the H I mass functions over the available range of H I masses, i.e. without extrapolation. The error bars in this case are computed by propagating the 2DSWML mass function uncertainties in quadrature. The violet diamonds show the second alternative estimate, found by rescaling the best-fitting H I mass function in each halo mass bin (see Section 4.2). The corresponding error bars are computed from the uncertainties in the mass function found by Jones et al. (2018).

ties on ϕ^b_i, quantified as described above, can be propagated into $\tilde{M}^b_{\rm H\,I}$ in quadrature.

(ii) The second estimator is produced by rescaling the best-fitting H I mass function found by Jones et al. (2018) in each halo mass bin. The rescaling factor for each group in the bin is estimated as the ratio of the observed number of sources found in that group to the number expected given the 2DSWML estimate of Jones et al. (2018) accounting for sample completeness at the distance to the group. $M^b_{\rm H\,I}$ is then estimated by applying equation (25) to the Schechter function found in Jones et al. (2018) rescaled by the factor above.

Unlike our fiducial estimator, this alternative method has no free parameters, and can therefore be used to explore the possible consequences of overfitting the per-bin mass functions based on a small number of objects. The main drawback of this estimator is that, by constructions, it assumes that the $m_{\rm H\,I}$ distribution in groups is the same as in the field.

These alternative measurements of the $M_{\rm H\,I}(M_h)$ relation are shown as blue squares and pink diamonds with error bars in Fig. 6.

As could be expected, the measurements corresponding to the first alternative estimator are consistently below our fiducial estimates generated from the integral of the Schechter functions, with the missing mass corresponding to the contribution of sources below the ALFALFA detection limit. However, the associated mass difference is mostly below ∼25 per cent of our fiducial mass measurements throughout the full mass range. Since this offset is always smaller than the 1σ statistical uncertainties, we find the impact of extrapolating the mass function to lower masses to be minimal. Note also that the blue error bars are consistently smaller than the orange ones. This is also to be expected, since the errors on $\tilde{M}^b_{\rm H\,I}$ estimated as described above, do not account for the additional uncertainty associated with mass below the detection limit.

The second estimator, based on extrapolating the overall H I mass function, agrees well with our fiducial measurements in general, although it is noticeably lower in the two lowest M_h bins. This is caused by the larger value of M_* preferred by our Schechter fits in the low-M_h bins. This result is consistent with previous measurements of the H I mass function around the region of the Virgo cluster, which suggest that massive ($\sim 10^{15}$ M$_\odot$) haloes have a smaller M_* than the field. Although this could be caused by ram pressure or tidal stripping, a better understanding of this result will require a more detailed study of the H I content in low-mass haloes in both data and simulations (Villaescusa-Navarro et al. 2018). In any case, both estimates of $M^b_{\rm H\,I}$ are compatible within present uncertainties, and therefore we conclude that our measurements of this quantity are robust with respect to the method used to estimate it.

5 RESULTS

5.1 Fiducial results

We produce constraints on the three parameters of the $M_{\rm H\,I}(M_h)$ relation (equation 1), $\theta \equiv \{\log_{10}M_0, \log_{10}M_{\rm min}, \alpha\}$, from a joint data vector composed of three parts:

(i) Measurements of the projected correlation function $\Xi(\sigma)$ (see Section 4.1) in $N_\Xi = 17$ logarithmic bins of σ between 0.43 and 30.8 h^{-1} Mpc. We use the altered NFW H I density profile described in Section 1 as our fiducial model for the small-scale correlation function. We study the impact of this choice, as well as the choice of scale cuts in Section 5.2.

(ii) Direct measurements of the $M_{\rm H\,I}(M_h)$ relation (see Section 4.2) in the $N_M = 7$ logarithmic bins of halo mass shown in Fig. 6. Our fiducial measurements consist of the $M_{\rm H\,I}$ estimates derived from the integral of the best-fitting Schechter H I mass functions in each M_h bin. We show the impact of extrapolating the H I mass function below ALFALFA's detection limit on our results in Section 5.3.

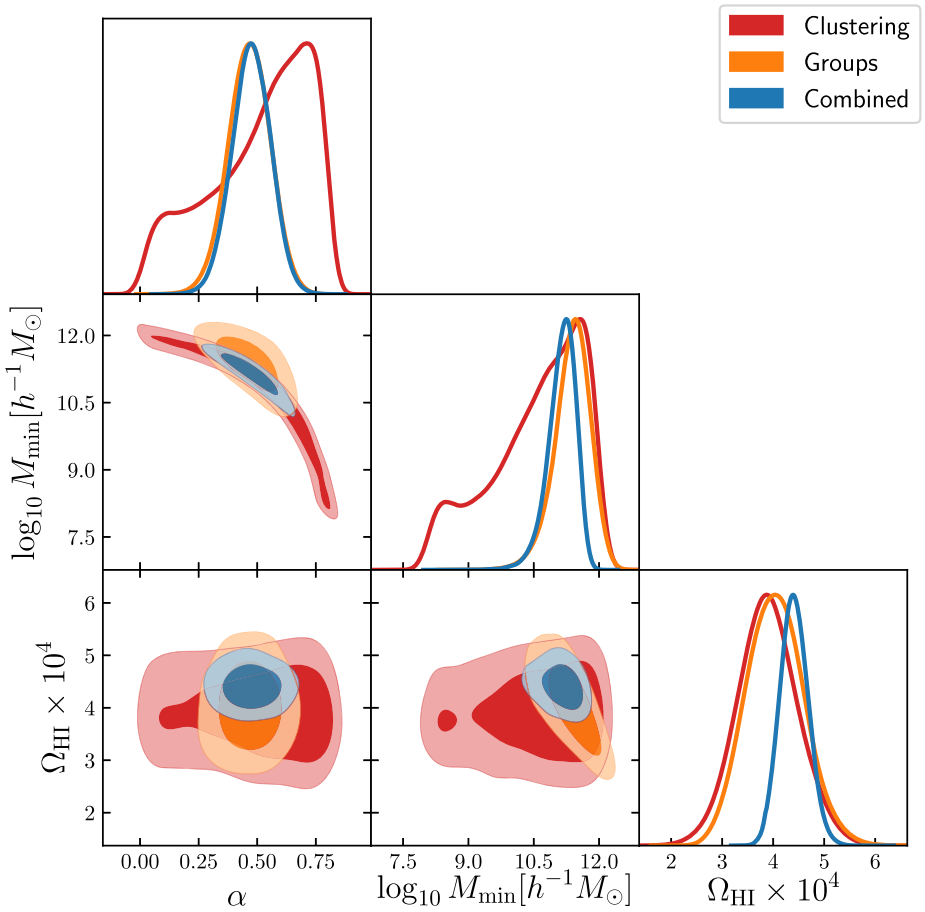

Figure 7. Final constraints on the parameters of the $M_{\rm H\,I}(M_h)$ relation. Results are shown for the combination of clustering data and the $\Omega_{\rm H\,I}$ measurement (red), for the measurements of $M_{\rm H\,I}$ in groups and $\Omega_{\rm H\,I}$ (light orange) and for the combination of the three data sets (blue).

(iii) One measurement of the cosmic H I abundance $\Omega_{\rm H\,I} = (3.9 \pm 0.1\,({\rm stat.}) \pm 0.6\,({\rm syst})) \times 10^{-4}$ from ALFALFA's $\alpha.100$ sample, as reported by Jones et al. (2018). In terms of the halo model, the cosmic abundance receives contributions from the H I content of haloes with arbitrarily small masses. Since our direct measurements of the $M_{\rm H\,I}(M_h)$ relation do not go below $\log_{10} M_h/(h^{-1}\,{\rm Mpc}) \simeq 12.5$, this additional data point allows us to break the degeneracy between the overall amplitude M_0 and the minimum halo mass $M_{\rm min}$ of the $M_{\rm H\,I}(M_h)$ relation.

Our fiducial data vector $\mathbf{d}$ therefore contains $N_\Xi + N_M + 1 = 25$ elements, which we use to constrain the three-parameter model of the $M_{\rm H\,I}(M_h)$ relation (in addition to the profile concentration parameter $c_{\rm H\,I,0}$, which we marginalize over). Assuming Gaussian statistics for $\mathbf{d}$, and in the absence of priors, the posterior distribution of the model parameters θ is given by

$$\chi^2 \equiv -2 \log p(\theta|\boldsymbol{d}) = [\boldsymbol{d} - \boldsymbol{t}(\theta)]^T\, \hat{\mathbf{C}}^{-1} [\boldsymbol{d} - \boldsymbol{t}(\theta)], \qquad (27)$$

where $\boldsymbol{t}(\theta)$ is the theoretical prediction for $\boldsymbol{d}$, described in Section 2, and $\hat{\mathbf{C}}$ is the covariance matrix of our measurements.

We build the covariance matrix $\hat{\mathbf{C}}$ as a block-diagonal matrix, where the first $N_\Xi \times N_\Xi$ block is given by the covariance matrix of the correlation function measurements (see Fig. 4). We assume the remaining $N_M + 1$ elements (corresponding to the H I abundance in groups and the cosmic H I abundance) to be uncorrelated with the correlation function measurements, and that their statistical uncertainties are also uncorrelated among themselves. These measurements are, however, correlated through some of their systematic uncertainties. In particular, the calibration of the absolute flux scale in ALFALFA dominates the systematic error budget in the measurement of $\Omega_{\rm H\,I}$ and $M^b_{\rm H\,I}$, and should affect all of these quantities in the same manner, rescaling them by an overall factor. In order to incorporate this correlation in our analysis, we add, to the statistical covariance matrix described above, a systematic component that is fully correlated across the last $N_M + 1$ measurements and with an amplitude 0.6×10^{-4} in the $\Omega_{\rm H\,I}$–$\Omega_{\rm H\,I}$ component. Note that the measurement of the projected correlation function is immune to the effects of an overall rescaling factor, and therefore the corresponding part of the systematic contribution to the covariance matrix is fixed to 0.

Finally, given the residual degeneracy between $M_{\rm min}$ and M_0 in our parametrization, we choose to show all our results in terms of $(\alpha, \log_{10} M_{\rm min}, \Omega_{\rm H\,I})$ instead, but will also provide the corresponding best-fitting value and uncertainty on $\log_{10} M_0$. We use broad top-hat priors for all parameters, with $c_{\rm H\,I,0} \in [0, 100]$, $\alpha \in [0, 2]$, $\log_{10} M_{\rm min}/h^{-1}{\rm M}_\odot \in [8, 13]$, and $\Omega_{\rm H\,I} \times 10^4 \in [0, 20]$. In all cases, we show constraints on $\log_{10} M_0$, $\log_{10} M_{\rm min}$, and α marginalized over the concentration parameter $c_{\rm H\,I,0}$.

We sample the likelihood in equation (27) using the publicly available implementation of the Markov chain Monte Carlo algorithm `emcee` (Foreman-Mackey et al. 2013). The resulting constraints on the $M_{\rm H\,I}(M_h)$ parameters are shown in Fig. 7 for different data combinations. We find compatible constraints from the clustering and groups data separately. Our marginalized

constraints on the $M_{\rm H\,I}(M_h)$ parameters are $\alpha = 0.48 \pm 0.08$, $\log_{10} M_{\rm min}/(h^{-1}{\rm M}_\odot) = 11.18^{+0.28}_{-0.35}$, $\log_{10} M_0/h^{-1}{\rm M}_\odot = 9.44^{+0.31}_{-0.39}$. The maximum-likelihood values are a good fit to the data in all cases, with a $\chi^2 = 15.4$ for 21 degrees of freedom for the full data vector. Although the clustering data are not able to jointly measure α and $\log_{10}M_{\rm min}$, and the groups data dominate the final uncertainties, clustering is still important in tightening the constraints (see e.g. the α–$\log_{10}M_{\rm min}$ plane). In particular, we find that, within this model, the clustering measurements allow us to reduce the uncertainty on $\Omega_{\rm H\,I}$ with respect to the mass-function measurement of Jones et al. (2018), obtaining $\Omega_{\rm H\,I} = 4.40^{+0.28}_{-0.27} \times 10^{-4}$.

Fig. 8 shows our best-fitting $M_{\rm H\,I}(M_h)$ relation (red solid line), together with its 1σ uncertainty (shaded area) as well as our fiducial measurements of this relation on galaxy groups (blue points with error bars). The measurements from the DR7 group catalogue are shown in blue. In order to jointly reproduce the measured H I content in high-mass haloes as well as the measured total H I abundance, the model predicts a sharp drop in H I content below a halo mass $\log M_h/h^{-1}{\rm M}_\odot \sim 11.5$. The figure also shows, as black points, the $M_{\rm H\,I}(M_h)$ relation measured in the IllustrisTNG-100 magnetohydrodynamic simulation (Villaescusa-Navarro et al. 2018) from a cosmological volume of $(75\,h^{-1}\,{\rm Mpc})^3$. The error bars represent the 1σ halo-to-halo variation on $M_{\rm H\,I}(M_h)$. Although, overall, we find good agreement between our results and the simulation, for very small halo masses, the amplitude of $M_{\rm H\,I}(M_h)$ differs significantly between our results and IllustrisTNG. This is however expected, given that the value of $\Omega_{\rm H\,I}$ in IllustrisTNG is $\simeq 7.5 \times 10^{-4}$, i.e. roughly a factor of 2 larger than the ALFALFA measurement used here. Although our model predicts a larger low-mass cut-off than is found in simulations, existing data on halo masses below the range probed by the SDSS group catalogue are not incompatible with this prediction. To illustrate this, Fig. 8 also shows the H I and halo masses measured for the Milky Way and M31 (Draine 2011; Braun et al. 2009).

Finally, Fig. 9 shows our measurement of the projected correlation function (blue points) together with the best-fitting prediction and associated uncertainties (red line and shaded area) and the dark matter correlation function from `HaloFit` (black solid line) scaled by our best-fitting $b^2_{\rm HI}$ (see Section 6). It is worth noting that our theory prediction does not require for us to invoke assembly bias. This is not surprising, given that our measurements are dominated by the highest $M_{\rm H\,I}$ sources, whereas, as shown by Guo et al. (2017), this effect is most prominent on low masses.

5.2 Impact of small scales

On small scales, the halo-model prediction of the two-point correlation function is dominated by the shape of the H I density profile. It is therefore important to evaluate whether our assumptions regarding the distribution of H I within each halo impacts our results on their overall H I content.

The blue and light-orange contours in the top panel of Fig. 10 show the constraints on the $M_{\rm H\,I}(M_h)$ relation derived from the measurements of the projected correlation function for the exponential and altered NFW profiles described in Section 2, respectively. Constraints are shown for the full range of scales ($\sigma \in (0.11, 30.8)\,h^{-1}$ Mpc) and combined with the ALFALFA measurement of $\Omega_{\rm H\,I}$. The figure shows that the constraints on the $M_{\rm H\,I}(M_h)$ parameters (particularly in terms of uncertainty) depend significantly on the model used to describe the distribution of H I within each halo. This is an undesirable feature, since we aim to constrain the global parameters of the $M_{\rm H\,I}$–M_h relation, given the large uncertainties in the actual shape of the H I density profile. On sufficiently large scales, in the two-halo regime, this dependence should become negligible. We have verified this by removing all data points with $\sigma > 0.43\,h^{-1}$ Mpc. These results are shown in Fig. 10 in green and red for the exponential and altered NFW profiles, respectively. The dependence on the choice of profile, in terms of constraining power, vanishes in this regime. We thus use this restricted range of scales and the altered NFW profile for our fiducial analysis. Although the choice of profile in this regime is not relevant, we note that Villaescusa-Navarro et al. (2018) find that the altered NFW profile with an exponential cut-off on small scales is better able to fit measurements from hydrodynamical simulations.

5.3 Low-mass extrapolation

As described in Section 4.2, our measurement of the H I content of galaxy groups is based on extrapolating the H I mass functions measured in bins of halo mass beyond the detection threshold of ALFALFA. This is a legitimate approach as long as the range of masses covered by our sample constitute the main contribution to the total H I budget, in which case we only incur in a small systematic effect when extrapolating the abundance of low-H I sources. We have shown that the mass deficit is generally below $\sim$20 per cent of each individual H I mass measurement, and is always within the 1σ uncertainties. The right-hand panel of Fig. 10 shows the impact of this systematic on our final constraints on the parameters of the $M_{\rm H\,I}(M_h)$ relation. The figure shows the constraints derived from our fiducial $M_{\rm H\,I}$ measurements in red, as well as the contours corresponding to our two alternative estimates: summing over the 2DSWML mass function (blue) and re-scaling the global H I mass function (light orange). The constraints derived from both estimates are compatible, with a negligible shift in the best-fitting $\log_{10}M_{\rm min}$. We therefore conclude that any residual systematics in the method used to measure the H I content as a function of halo mass in the group catalogue is subdominant.

6 DISCUSSION

We have placed constraints on the distribution of neutral hydrogen in dark matter haloes as a function of halo mass. To do so, we have used the H I-weighed clustering of 21 cm sources detected by ALFALFA, as well as the abundance of those sources in haloes identified in the galaxy group catalogue compiled from the SDSS DR7 data. Our results show a power-law relation between $M_{\rm H\,I}$ and M_h at large halo masses with an exponent $\alpha = 0.48 \pm 0.08$. This relation is exponentially suppressed on masses below $\log_{10} M_{\rm min}/(h^{-1}{\rm M}_\odot) = 11.18^{+0.28}_{-0.35}$. Although this suppression is not directly measurable in the data, given the mass range of the group catalogue, it can be inferred indirectly by combining the group data with the total H I abundance measured by ALFALFA and our measurement of the two-point correlation function.

The constraints derived individually from our two data sets are compatible between themselves and with the combined constraints, and in all cases we find the model in equation (1) to be a good fit to the data. It is worth emphasizing the fact that, although the clustering data are not able to break the degeneracy between α and $M_{\rm min}$, even when combined with the measurement of $\Omega_{\rm H\,I}$, it is vital to improve the constraints derived from the combination of the H I abundance in groups and $\Omega_{\rm H\,I}$. In fact we find that, within our model, clustering information is able to significantly reduce the final uncertainties on $\Omega_{\rm H\,I}$ compared with direct measurements of this quantity from the H I mass function. Furthermore, the clustering properties of the H I

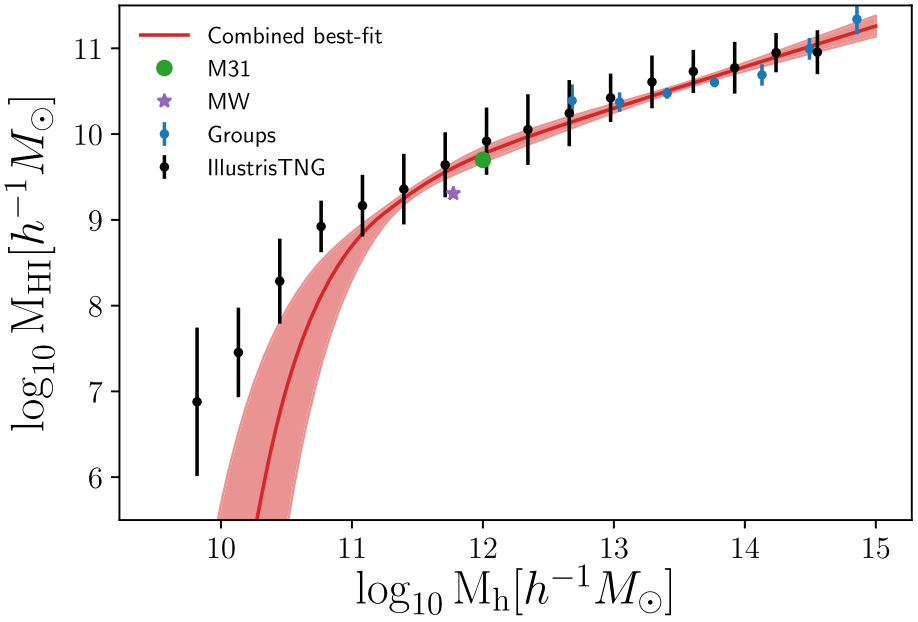

Figure 8. Combined best-fitting $M_{\rm H\,I}(M_h)$ relation (red solid line) together with the 1σ uncertainty (red shaded region) using three data sets: the projected mass-weighted correlation function $\Xi(\sigma)$, the direct estimates of the $M_{\rm H\,I}(M_h)$ relation from the galaxy group catalogue (shown also as blue points with error bars) and the measurement of the cosmic H I abundance $\Omega_{\rm H\,I}$ in Jones et al. (2018). For comparison, we show the results from the IllustrisTNG magnetohydrodynamic simulation (black points, Villaescusa-Navarro et al. 2018) with the error bars corresponding to the typical per-halo scatter. We also show measurements of M_h and $M_{\rm H\,I}$ for individual galaxies: the Milky Way (purple star) and M31 (green circle).

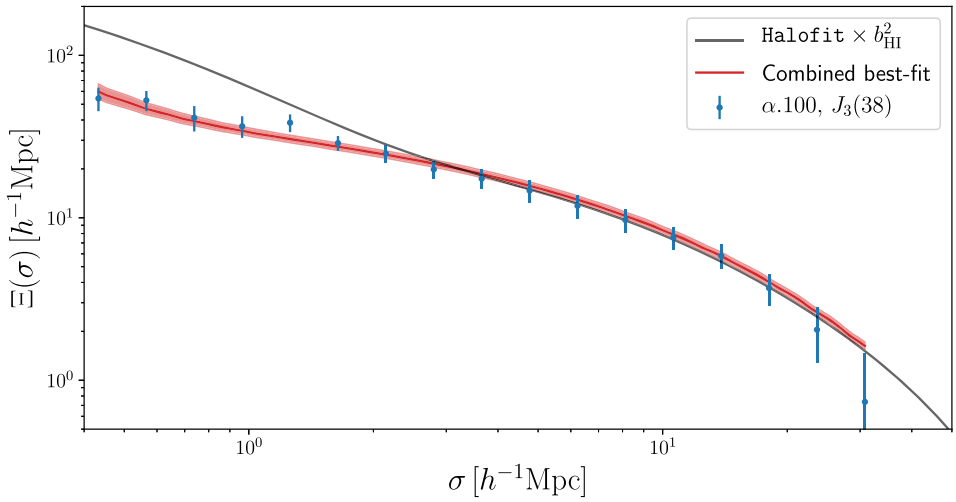

Figure 9. Predicted projected correlation function for our best-fitting $M_{\rm H\,I}(M_h)$ relation (red solid line) and its 1σ uncertainty (red shaded region). The blue points with error bars show the direct measurements from ALFALFA, and the black solid line corresponds to the `HaloFit` prediction for the matter correlation function scaled by our best-fitting $b_{\rm H\,I}^2$ (see Section 6).

are arguably the most relevant piece of information for future 21 cm intensity mapping studies, and this information is potentially better summarized by the projected correlation function data used here.

Recently Villaescusa-Navarro et al. (2018) have aimed at characterizing the $M_{\rm H\,I}(M_h)$ relation from state-of-the-art magnetohydrodynamic simulations, and it is therefore relevant to explore the level of agreement between these simulated results and our data-driven constraints. In terms of the overall $M_{\rm H\,I}(M_h)$ relation, this comparison is best summarized in Fig. 6. We find that our results agree well with those of Villaescusa-Navarro et al. (2018) at $z = 0$ for large halo masses ($M_{\rm h} \gtrsim 10^{12.5}\,{\rm M}_\odot\,h^{-1}$), and that our best-fitting model as well, as the simulated data, are in good agreement with individual H I mass measurements. However, we observe that the $M_{\rm H\,I}(M_h)$ relation derived from simulations departs significantly from our best-fitting model on the low-mass end, predicting significantly higher H I masses. This disagreement is correlated with the higher value of $\Omega_{\rm H\,I} \sim 7 \times 10^{-4}$ measured in IllustrisTNG, which is also the measurement that allows us to place constraints on the cut-off mass scale. The fact that the radiation from the sources is not accounted for in IllustrisTNG may explain the differences in the value of $\Omega_{\rm H\,I}$ and on the average H I mass inside small haloes.

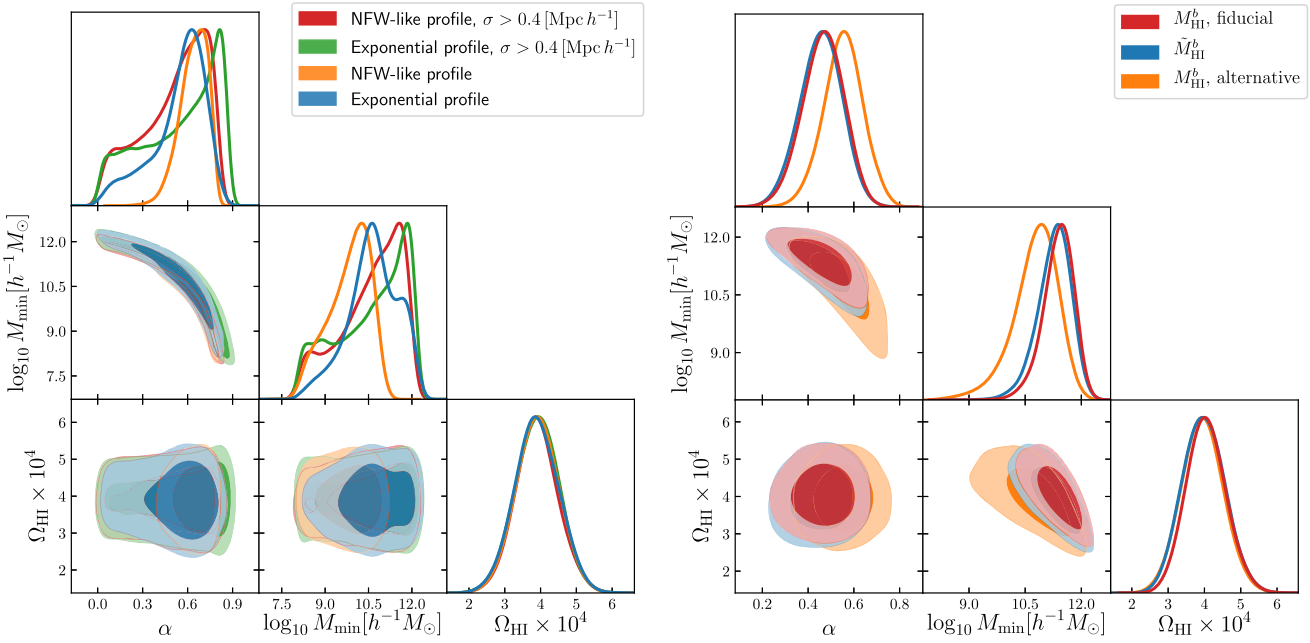

Figure 10. *Left:* constraints on the $M_{\rm H\,I}(M_h)$ relation derived from the clustering analysis under different scale cuts and choices of H I density profile. We derive our final scale cuts by demanding final constraints that do not depend on the choice of profile. *Right:* constraints on the $M_{\rm H\,I}(M_h)$ relation for different estimates of the H I mass in galaxy groups. Our fiducial measurements are shown in blue, while the red and light-orange contours show the results from the two alternative estimates described in Section 4.2 (see also Fig. 6).

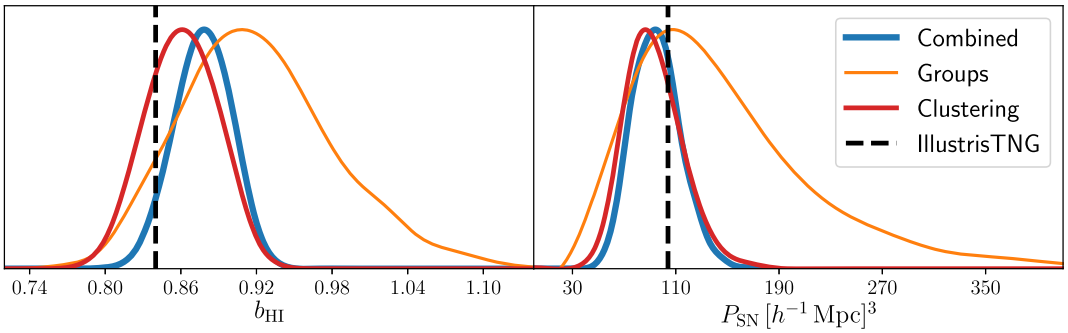

Figure 11. Posterior distributions for the large-scale H I bias (*left*) and shot-noise power spectrum (*right*) predicted from different combinations of our fiducial data vector: clustering+$\Omega_{\rm H\,I}$ (red), groups+$\Omega_{\rm H\,I}$ (light orange), and all data (blue). The vertical dashed line shows the result found in the IllustrisTNG simulation (Villaescusa-Navarro et al. 2018).

For the purposes of predicting the clustering properties of H I in future 21 cm experiments, two quantities are needed beyond $\Omega_{\rm H\,I}$: the large-scale H I bias $b_{\rm H\,I}$ and the shot-noise level $P_{\rm SN}$. Given our model for the $M_{\rm H\,I}(M_h)$ relation, we can make predictions for these two quantities within the halo model ($b_{\rm H\,I} = F^1_1(k=0)$, $P_{\rm SN} = F^0_2(k=0)$, see equation 9), which we can then directly compare with the values found by Villaescusa-Navarro et al. (2018). The results of this comparison are shown in Fig. 11: our constraints on both quantities ($b_{\rm H\,I} = 0.878^{+0.022}_{-0.023}$, $P_{\rm SN} = 94^{+20}_{-18}\,[h^{-1}\,\mathrm{Mpc}]^3$) are in good agreement with the values predicted by IllustrisTNG at $z = 0$. Although this result may seem at odds with the disagreement between data and simulation in terms of the total $\Omega_{\rm H\,I}$, this can be understood as due to the relatively higher contribution from larger mass objects to these two quantities, for which our results agree with those of IllustrisTNG. It is also interesting to note that, even though the clustering data alone are not able to break the degeneracies between the $M_{\rm H\,I}(M_h)$ parameters, they drive the constraints on both $b_{\rm H\,I}$ and $P_{\rm SN}$.

Our measurement of the $M_{\rm H\,I}(M_h)$ relation can be translated into a limiting circular velocity to host H I. Defining this as the circular velocity associated with a minimum halo mass such that 98 per cent of the cosmic H I is contained within heavier objects (see Villaescusa-Navarro et al. 2018), we find $V_{\rm circ} = 51^{+11}_{-10}\,\mathrm{km\,s^{-1}}$. This is in tension with the value found in Villaescusa-Navarro et al. (2018, $V_{\rm circ} = 34\,\mathrm{km\,s^{-1}}$), which is correlated with the higher cut-off halo mass measured in the data and shown in Fig. 6.

The results presented here are also interesting beyond future cosmological 21 cm studies, as they provide insight into the distribution of neutral hydrogen across structures of different masses. Furthermore, our direct measurement of the $M_{\rm H\,I}(M_h)$ relation is based on the characterization of the H I mass function for sources within galaxy groups, and have revealed hints about the relative

dependence of the H I mass distribution on halo mass, with higher H I knee masses found on lower mass haloes. In general, the behaviour of the $M_{\rm H\,I}(M_h)$ relation in the low-mass end ($M_h \lesssim 10^{12}\,{\rm M}_\odot\,h^{-1}$) is still somewhat uncertain, and its study will benefit in the future from higher quality data and improved analysis methods.

We must also emphasize that the $M_{\rm H\,I}(M_h)$ relation contains a huge amount of astrophysical information. In the high-mass end, the strength of processes such as AGN feedback, ram pressure, and tidal stripping will leave its signature on the value of α (Villaescusa-Navarro et al. 2016), while on the low-mass end the presence of the UV background and the minimum mass to trigger self-shielding will determine the shape and amplitude of $M_{\rm H\,I}(M_h)$. Our results can be used in combination with hydrodynamic simulations or semi-analytic models (Lagos et al. 2014; Zoldan et al. 2017) to improve our knowledge on the role of different astrophysical processes.

ACKNOWLEDGEMENTS

We thank David Spergel for useful comments and discussions. AO and DA thank the Center for Computational Astrophysics, part of the Flatiron Institute of the Simons Foundation, for their hospitality. AO is supported by the INFN grant PD 51 INDARK. DA acknowledges support from the Beecroft trust and from the Science and Technology Facilities Council (STFC) through an Ernest Rutherford Fellowship, grant reference ST/P004474/1. The work of FVN is supported by the Simons Foundation. MGJ is supported by a Juan de la Cierva formación fellowship and also acknowledges support from the grant AYA2015-65973-C3-1-R (MINECO/FEDER, UE) and the "Centro de Excelencia Severo Ochoa" program of the Spanish Science Ministry under grant SEV-2017-0709. We acknowledge the work of the entire ALFALFA collaboration team in observing, flagging, and extracting the catalogue of galaxies used in this work. We also acknowledge the use of `CosmoMC/GetDist` (Lewis & Bridle 2002), `CAMB` (Lewis, Challinor & Lasenby 2000), `IPython` (Perez & Granger 2007), `Matplotlib` (Hunter 2007), and `NumPy/SciPy` (van der Walt, Colbert & Varoquaux 2011).

REFERENCES

Alam S. et al., 2017, MNRAS, 470, 2617
Alonso D., 2012, preprint (arXiv:1210.1833)
Alonso D., Bull P., Ferreira P. G., Santos M. G., 2015, MNRAS, 447, 400
Anderson C. J. et al., 2018, MNRAS, 476, 3382
Bagla J., Khandai N., Datta K. K., 2010, MNRAS, 407, 567
Barnes L. A., Haehnelt M. G., 2014, MNRAS, 440, 2313
Basilakos S., Plionis M., Kovač K., Voglis N., 2007, MNRAS, 378, 301
Battye R. A., Davies R. D., Weller J., 2004, MNRAS, 355, 1339
Battye R. A., Browne I. W. A., Dickinson C., Heron G., Maffei B., Pourtsidou A., 2013, MNRAS, 434, 1239
Braun R., 2012, ApJ, 749, 87
Braun R., Thilker D. A., Walterbos R. A. M., Corbelli E., 2009, ApJ, 695, 937
Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, ApJ, 803, 21
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559
Castorina E., Villaescusa-Navarro F., 2017, MNRAS, 471, 1788
Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303
Cole S. et al., 2001, MNRAS, 326, 255
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Crighton N. H. M. et al., 2015, MNRAS, 452, 217
Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398
Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium. Princeton Univ. Press, Princeton, NJ
Efstathiou G., Ellis R. S., Peterson B. A., 1988, MNRAS, 232, 431
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Giovanelli R. et al., 2005, AJ, 130, 2598
Guo H., Li C., Zheng Z., Mo H. J., Jing Y. P., Zu Y., Lim S. H., Xu H., 2017, ApJ, 846, 61
Hamilton A. J. S., 2000, MNRAS, 312, 257
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Haynes M. P. et al., 2011, AJ, 142, 170
Haynes M. et al., 2018, ApJ, 861, 49
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jones M. G., Papastergis E., Haynes M. P., Giovanelli R., 2016, MNRAS, 457, 4393
Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS, 477, 2
Kennicutt R. C.,, Jr., 1998, ApJ, 498, 541
Lagos C. D. P., Baugh C. M., Zwaan M. A., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A. M., van Kampen E., 2014, MNRAS, 440, 920
Lah P. et al., 2007, MNRAS, 376, 1357
Landy S. D., Szalay A. S., 1993, AJ, 412, 64
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Loeb A., Wyithe J. S. B., 2008, Phys. Rev. Lett., 100, 161301
Lupton R., 1993, Statistics in Theory and Practice. Princeton Univ. Press, Princeton, NJ
Macciò A. V., Dutton A. A., van den Bosch F. C., Moore B., Potter D., Stadel J., 2007, MNRAS, 378, 55
McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815
Maller A. H., Bullock J. S., 2004, MNRAS, 355, 694
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
Noterdaeme P. et al., 2012, A&A, 547, L1
Padmanabhan H., Refregier A., Amara A., 2017, MNRAS, 469, 2323
Papastergis E., Giovanelli R., Haynes M. P., Rodríguez-Puebla A., Jones M. G., 2013, ApJ, 776, 43
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ,
Perez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Pérez-Ràfols I. et al., 2018, MNRAS, 473, 3019
Peterson J. B. et al., 2009, in astro2010: The Astronomy and Astrophysics Decadal Survey. p. 234 preprint (arXiv:0902.3091)
Planck Collaboration XIII, 2016, A&A, 594, A13
Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610
Rhee J., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., van der Hulst T., 2013, MNRAS, 435, 2693
Santos M. G., Cooray A., Knox L., 2005, ApJ, 625, 575
Shaw J. R., Sigurdson K., Sitwell M., Stebbins A., Pen U.-L., 2015, Phys. Rev. D, 91, 083514
Smith R. E. et al., 2003, MNRAS, 341, 1311
Songaila A., Cowie L. L., 2010, ApJ, 721, 1448
Spergel D. N. et al., 2007, ApJS, 170, 377
Switzer E. R. et al., 2013, MNRAS, 434, L46
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
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

Villaescusa-Navarro F. et al., 2018, ApJ, 866, 135
Wolz L., Abdalla F. B., Blake C., Shaw J. R., Chapman E., Rawlings S., 2014, MNRAS, 441, 3271
Wolz L., et al., 2015, MNRAS, 464, 4938
Wyithe J. S. B., Loeb A., 2008, MNRAS, 383, 606
Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005, MNRAS, 356, 1293
Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153
Yoon I., Rosenberg J. L., 2015, ApJ, 812, 4
Zafar T., Péroux C., Popping A., Milliard B., Deharveng J.-M., Frank S., 2013, A&A, 556, A141
Zehavi I. et al., 2002, ApJ, 571, 172
Zoldan A., De Lucia G., Xie L., Fontanot F., Hirschmann M., 2017, MNRAS, 465, 2236
Zwaan M. A., Prochaska J. X., 2006, ApJ, 643, 675
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005a, MNRAS, 359, L30
Zwaan M. A., van der Hulst J. M., Briggs F. H., Verheijen M. A. W., Ryan-Weber E. V., 2005b, MNRAS, 364, 1467

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