Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **509,** 3268–3284 (2022) https://doi.org/10.1093/mnras/stab3164
Advance Access publication 2021 November 1

The ALFALFA H I velocity width function

Kyle A. Oman [1,2]★

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

Accepted 2021 October 27. Received 2021 September 29; in original form 2021 August 19

ABSTRACT

We make the most precise determination to date of the number density of extragalactic 21-cm radio sources as a function of their spectral line widths – the H I velocity width function (H I WF) – based on 21 827 sources from the final 7000 deg^2 data release of the Arecibo Legacy Fast ALFA (ALFALFA) survey. The number density of sources as a function of their neutral hydrogen masses – the H I mass function (H I MF) – has previously been reported to have a significantly different low-mass slope and 'knee mass' in the two sky regions surveyed during ALFALFA. In contrast with this, we find that the shape of the H I WF in the same two sky regions is remarkably similar, consistent with being identical within the confidence intervals implied by the data (but the overall normalization differs). The spatial uniformity of the H I WF implies that it is likely a stable tracer of the mass function of dark matter haloes, in spite of the environmental processes to which the measured variation in the H I MF are attributed, at least for galaxies containing enough neutral hydrogen to be detected. This insensitivity of the H I WF to galaxy formation and evolution can be exploited to turn it into a powerful constraint on cosmological models as future surveys yield increasingly precise measurements. We also report on the possible influence of a previously overlooked systematic error affecting the H I WF, which may plausibly see its low-velocity slope steepen by ∼40 per cent in analyses of future, deeper surveys. Finally, we provide an updated estimate of the ALFALFA completeness limit.

Key words: galaxies: abundances – galaxies: luminosity function, mass function – dark matter – radio lines: galaxies.

1 INTRODUCTION

One of the fundamental predictions of the current concordance cold dark matter (ΛCDM) cosmology is the number density as a function of mass of self-bound dark matter haloes, termed the halo mass function (HMF; Frenk et al. 1988; Jenkins et al. 2001; Sheth & Tormen 2002). Once the parameters of the ΛCDM model are specified, the HMF in the absence of galaxy formation, e.g. as realized in a cosmological *N*-body simulation, is a prediction with no free parameters. The CDM HMF is well-approximated as a power law with a slope $\phi(M) \propto M^{-1.9}$ over $\gtrsim 8$ orders of magnitude below the cutoff at the scale of the largest collapsed structures at the present day (Angulo et al. 2012). In a more realistic scenario, low-mass haloes lose their gas, and therefore a fraction of their mass, early in the history of the Universe, stunting their growth and leading to a small (0.1 dex) shift in the low-mass HMF, but no overall change in slope (Sawala et al. 2015, 2016). A precision measurement of the HMF would therefore constitute an excellent test of the predictions of the CDM model, especially at the low-mass end where the predictions of other dark matter models differ (e.g. warm dark matter models; Bode, Ostriker & Turok 2001).

As the HMF is not directly measurable, we must instead rely on the kinematics of visible tracers orbiting in dark matter haloes, or other gravitational effects such as lensing, to provide indirect constraints. The expected structure of CDM haloes implies that their circular velocity[1] profiles $v_{\rm circ}$, related to the enclosed mass within radius r as $v_{\rm circ}^2 = GM(<r)/r$, have a broad plateau whose amplitude $v_{\rm max}$ is tightly correlated with the halo mass (Navarro, Frenk & White 1996, 1997). The number density of sources as a function of the characteristic speed of rotation of those sources (as revealed by kinematic tracers), which we will term the 'velocity function' (VF), therefore gives a means to constrain the HMF observationally. In practice, connecting the actual kinematic tracer observed – such as a spectral line width – and $v_{\rm max}$ requires some additional information and/or modelling.

In the case of the width of the 21-cm emission line of neutral hydrogen (H I), the maximum circular velocity of a halo in which a sufficiently extended disc of H I gas is rotating is approximately half the full width at half-maximum (FWHM) of the line w_{50}, provided the inclination of the disc is accounted for, such that $v_{\rm max} \sim \frac{1}{2} w_{50}/\sin(i)$, where i is the inclination angle. The need to correct for inclination is problematic, as surveys of line widths covering representative volumes currently do not resolve the spatial structure of the gas, necessitating reliance on optical imaging to estimate i (e.g. Zwaan, Meyer & Staveley-Smith 2010). There are numerous additional issues which need to be accounted for in an attempt to obtain $v_{\rm max}$ from a measurement of w_{50} (see e.g. Verheijen 2001; Verbeke

★ E-mail: kyle.a.oman@durham.ac.uk

[1]To avoid ambiguity, we adopt a notation in which v denotes velocity, and V, volume.

et al. 2017, for detailed discussions, in addition to the references in the following list): the gas orbits may not be circular (Macciò et al. 2016); the gas disc may not be sufficiently extended to reach the flat portion of the circular velocity curve (Brook & Di Cintio 2015b; Brook & Shankar 2016; Macciò et al. 2016; Ponomareva, Verheijen & Bosma 2016; Brooks et al. 2017); the gas disc may not lie in a single plane (Ponomareva et al. 2016); the system may not be in dynamical equilibrium; the emission may be confused with neighbouring sources (Jones et al. 2015; Chauhan et al. 2019); the gas disc may be partially supported by turbulent or thermal pressure (Brook & Di Cintio 2015b; Ponomareva et al. 2016); etc. There have been several attempts to work around these issues and recover the HMF of gas-rich galaxies (e.g. Zavala et al. 2009; Zwaan et al. 2010; Papastergis et al. 2015 – hereafter P15; Li et al. 2019; Dutta, Khandai & Rana 2021). Whilst in principle each of the issues may be addressed directly with sufficiently detailed observations and modelling, assembling a large enough sample of objects to pursue this approach has so far proven to be prohibitively expensive.

An observationally more straightforward approach is to simply measure the number density of sources as a function of w_{50}: the H I velocity width function (H I WF). Using the H I WF as a constraint on cosmology then requires a prediction for the H I WF expected in a given model, and some understanding of the possible degeneracies in this prediction across various models, but this is in principle easier to achieve than it is to solve the inverse problem of inferring the HMF from the H I WF (Papastergis et al. 2011). These considerations lead us to focus on the H I WF, and omit further discussion of the H I VF, in this work.

The first direct measurements of the H I WF were by Zavala et al. (2009) and Zwaan et al. (2010), using an early partial (6 per cent) release of the Arecibo Legacy Fast ALFA[2] (ALFALFA) survey, and the H I Parkes All-Sky Survey (H I PASS), respectively. It was immediately recognized (and had been anticipated, from indirect estimates) that the observed low-velocity width slope would be very difficult to reconcile with theoretical expectations – there is an apparently severe overabundance of haloes predicted to exist which, apparently, fail to host observable galaxies. This result has been confirmed repeatedly (Papastergis et al. 2011; P15; Klypin et al. 2015), including by the measurement we present below, and possible means to reconcile theory and measurement further discussed in the literature (e.g. Zavala et al. 2009; Brook & Di Cintio 2015a,b; Macciò et al. 2016; Papastergis & Shankar 2016; Brooks et al. 2017; Dutton, Obreja & Macciò 2019). We defer further discussion of this fascinating topic to future work. In the present study, we present a new measurement of the H I WF based on the catalogue from the now-completed ALFALFA survey. We focus on testing the compatibility of the data with a hypothesis that is well-motivated in the ΛCDM cosmology: that the HMF, as encoded in the H I WF, is spatially invariant. That is, different sub-volumes of the full survey should have similar HMFs, which may be reflected by a similarity between their H I WFs.

Spatial variations in the H I WF have been measured before. Zavala et al. (2009) were able to measure a factor of ∼3 difference in number density in between source counts in a high- and low-density region in a very early ALFALFA survey catalogue, despite having only 15 detected sources in the lower density region. In the context of estimating a systematic uncertainty on their (total) measurement, Zwaan et al. (2010, see their fig. 7) noted that whilst the overall number density in four quadrants of H I PASS differed noticeably, the overall shape of the H I WF appeared similar in each. A similar situation is hinted at in fig. 8 of Papastergis et al. (2011), showing the H I WF for two regions on the sky covered by an early portion of the ALFALFA survey, but these authors comment on this point only very briefly. Moorman et al. (2014) compared the H I WF of ALFALFA galaxies in voids with those in the walls of the cosmic web. Their analysis suggests that the two populations are inconsistent with being drawn from a single underlying distribution, but they conclude that statistical uncertainties prevent them from claiming a significant difference in the H I WF shape. Below, we explore spatial variations in the H I WF leveraging the higher precision afforded by the large number ($\sim 2 \times 10^4$) of extragalactic sources detected in the full ALFALFA survey.

This article is structured as follows. We outline the characteristics of the ALFALFA survey and catalogue, and our selection of galaxies therefrom, in Section 2. Our methodology for the measurement of the H I WF, using the $1/V_{\rm eff}$ estimator, is described in Section 3. We present our measurement, including separately for independent subvolumes of the survey, in Section 4, and discuss its implications in Section 5. We summarize in Section 6.

2 THE ALFALFA SURVEY

The ALFALFA survey (Giovanelli et al. 2005) mapped $\sim 7000\,{\rm deg}^2$ of sky at 21-cm wavelengths out to distances of $\sim 250\,{\rm Mpc}$ ($cz \lesssim 18\,000\,{\rm km\,s^{-1}}$). The survey is composed of two contiguous areas on sky, one in the northern Galactic hemisphere, visible from Arecibo at night during the spring, the other in the southern Galactic hemisphere, visible in the autumn. Following Jones et al. (2018, hereafter J18), we will refer to the two areas as the 'spring' and 'fall' regions, respectively.[3]

Candidate extragalactic H I sources were identified using a matched-filtering approach (Saintonge 2007), supplemented by some sources identified by direct inspection of the raw data cubes. This initial catalogue was curated by hand to confirm or reject each individual detection, and to assign optical counterparts to detections wherever possible, resulting in the $\alpha.100$ extragalactic source catalogue described in Haynes et al. (2018). The catalogue lists the coordinates (for the H I and associated optical source), redshifts, 21-cm line flux density and width, distance, signal-to-noise ratio, and H I mass of sources, and their uncertainties where relevant; we refer to Haynes et al. (2018) for details of the determination of these parameters. The '100' of $\alpha.100$ refers to the 100 per cent completed survey. To compare with earlier work, we also make some use of the earlier $\alpha.40$ catalogue (Haynes et al. 2011, hereafter H11), which is very similar but covers only 40 per cent of the total survey area on sky.

We define a selection of sources from the $\alpha.100$ catalogue for use in measuring the H I WF, closely following the equivalent selection used in J18 for the H I mass function (H I MF). We use all 'Code 1' (i.e. S/N > 6.5) sources whose H I coordinates fall within the survey footprint (J18, tables D1–D4) and have a recessional velocity in the CMB frame of $< 15\,000\,{\rm km\,s^{-1}}$. We adopt the same minimum line width at FWHM[4] $\log_{10}(w_{50}/{\rm km\,s^{-1}}) > 1.2$ as J18, but *do not* impose a minimum H I mass (J18 used $\log_{10}(M_{\rm HI}/{\rm M}_\odot) > 6$). This includes four additional sources. Unlike the low-mass end of the H I MF, the

[2] Arecibo L-band Feed Array.

[3] See their fig. 1 and tables D1–D4 for details of the survey footprint; we use the fiducial, not the 'strict' footprint throughout this work.

[4] We use log and $\log_{10}$ to denote the natural and base-10 logarithms, respectively.

low-velocity end of the H I WF is somewhat sensitive to this choice (though the counting uncertainties are, of course, large). We use only sources above the bivariate 50 per cent completeness of the survey, which depends on flux and line width. Rather than the limit derived from the $\alpha.40$ data reported in H II, we use an updated completeness limit (equation A5) derived from the $\alpha.100$ data, see Appendix A for details. These cuts yield a sample of 21 827 sources, of which 13 620 and 8207 are in the spring and fall regions, respectively – the bulk of the difference with respect to the sample of J18 is due to the updated completeness limit.

Our initial measurement of the $\alpha.100$ H I WF included one obvious outlier point: the estimate in a bin centred at $\sim 90\,\mathrm{km\,s^{-1}}$ was offset upward by a factor of $\sim$2 relative to adjacent bins, and had a statistical uncertainty a factor of $\sim$6 greater than those of its neighbours. We determined that this irregularity was driven primarily by two sources (AGC 749235 and AGC 220210). Both of these are in the vicinity on the sky of the 'Coma I cloud', a group of galaxies at a distance of ~ 16.1 Mpc and with an unusually large peculiar velocity of $-840\,\mathrm{km\,s^{-1}}$ (e.g. Kashibadze, Karachentsev & Karachentseva 2018, see their fig. 10). AGC 749235 has a distance of $D = 6.2 \pm 1.3$ Mpc in the $\alpha.100$ catalogue, but the appearance of its optical counterpart PGC 5059199 suggests a larger distance (Kaisin & Karachentsev 2019). We have been unable to locate a redshift-independent distance for AGC 220210, listed at 2.8 ± 0.6 Mpc in $\alpha.100$, or its optical counterpart SDSS J121323.34+295518.3, in the literature. However, this is clearly the same object as KK 127, at 121322.7 + 295518 (J2000), which has as Tully-Fisher distance of 17.3 Mpc (Kashibadze et al. 2018, table 3). We have adopted distances of 16.1 ± 4.3 and 17.3 ± 3.5 Mpc for AGC 749235 and AGC 220210, respectively, and have re-derived their H I masses accordingly. There are likely other ALFALFA sources with distance errors (beyond those reflected in the uncertainties quoted in the $\alpha.100$ catalogue) – these remain a source of systematic uncertainty for the H I WF measurement which we will not attempt to address further in the present study. We note that, following J18 (but unlike P15), we prefer to retain nearby galaxies in our analysis in order to prevent a possible artificial suppression of the low-velocity end of the H I WF.

3 THE $1/V_{\mathrm{EFF}}$ MAXIMUM LIKELIHOOD ESTIMATOR

The $1/V_{\mathrm{eff}}$ maximum likelihood estimator (as described in Zwaan et al. 2005, Section 2, including discussion of the difference with respect to the 2DSWML estimator) is very similar to, but subtly distinct from, the bivariate step-wise maximum likelihood (2DSWML) estimator (e.g. as described in Martin et al. 2010, appendix B). We refer to the above references for full details, but, briefly, the $1/V_{\mathrm{eff}}$ method determines the effective volume in which each source observed in the course of a survey would have been detectable by that survey, accounting for the large-scale structure in the actual surveyed volume.[5] The estimator is two-dimensional in the sense that the completeness limit of the ALFALFA survey is a function of both the integrated 21-cm line flux S_{21}, and the line width w_{50}. We adopt the 50 per cent completeness limit[6] given in H11 (equations 4 and 5). By summing the 'weights' $1/V_{\mathrm{eff}}$ of sources in 2D bins in H I mass M_{HI}, determined as usual from the flux and distance as

$$M_{\mathrm{HI}}/\mathrm{M}_{\odot} = 2.36 \times 10^5 (D/\mathrm{Mpc})^2 \left(S_{21}/\mathrm{Jy\,km\,s^{-1}}\right), \quad (1)$$

and w_{50}, one obtains the H I mass-velocity width function. Summing this along the velocity-width axis yields the H I MF, whilst summation along the mass axis gives the H I WF. The H I MF and H I WF are therefore 'marginalizations' along the two axes of the same underlying distribution.

We use the same implementation of the $1/V_{\mathrm{eff}}$ estimator as was used by P15. We have verified that we can exactly reproduce their measurement of the H I WF (their fig. 2, reproduced in the upper left panel of Fig. 1 with green cross symbols) by reverting to the $\alpha.40$ catalogue and completeness limit as input, and approximately by trimming the $\alpha.100$ catalogue to the spring $\alpha.40$ footprint, using the $\alpha.100$ completeness limit (shown with grey points in the upper left panel of Fig. 1). As a further check of consistency with previously published results, we have verified that we reproduce the H I MF of J18 (their fig. 2, reproduced with red crosses in the lower left panel of Fig. 1). Our determination based on the $\alpha.100$ catalogue, shown with the black points, lies systematically above this previous measurement. However, if we revert to the $\alpha.40$ completeness limit (H11, equations 4 and 5), our measurement closely follows and is fully statistically consistent with this prior measurement. A detailed explanation of the few remaining small differences with respect to these prior results is given in Appendix B.

4 THE ALFALFA H I WF

Our measurement of the H I WF based on the full $\alpha.100$ catalogue is shown with the black points and error bars in the upper left panel of Fig. 1. In this case, we account only for counting (Poisson) uncertainties; other statistical uncertainties, and systematic uncertainties, are discussed below. This measurement appears to be in reasonable agreement with the determination by P15 (green crosses), but this is due to a chance partial cancellation of two competing systematic effects. First, that sample is drawn only from the spring region of the survey (the $\alpha.40$ subset of the $\alpha.100$ spring region, see H11, fig. 1) – if we restrict our galaxy sample to same region on sky, we obtain a measurement with a higher overall normalization. However, by chance, our updated completeness limit (see Appendix A) also tends to increase the normalization of the H I WF– were the P15 repeated with a matching completeness limit, it would lie significantly above our $\alpha.100$ measurement (see also Appendix B). In summary, our measurement and that of P15 should be taken to differ significantly, but the reason for this is clear: the regions of the sky covered by the two input catalogues have significantly different overall number densities of H I sources.

Although many previous analyses report only the counting uncertainties on the H I WF, the measurement uncertainties in distance, flux, and velocity-width for each galaxy also make a significant contribution to the statistical uncertainty budget. We estimate the total statistical uncertainty by randomly drawing a new value for each of D, S_{21}, and w_{50} from a Gaussian distribution with centre and width given by the values and uncertainties reported in the $\alpha.100$ catalogue, rejecting any unphysical (e.g. negative) values, then re-derive all derived quantities (e.g. H I mass). We then measure the H I WF for the resulting catalogue, again using the $1/V_{\mathrm{eff}}$ estimator. We repeat this process 1000 times and average the probability distributions in each velocity-width bin, then adopt the median and 16th to 84th

[5]The values $1/V_{\mathrm{eff}}$ are loosely analogous to the $1/V_{\mathrm{max}}$ weights in the classical estimator of Schmidt (1968).

[6]This definition approximates the sensitivity function of the survey as a step function, transitioning from a value of 0 to a value of 1 at the 50 per cent completeness limit. The true transition is a smooth gradient, but we adopt this approximation to facilitate numerical calculations, following P15.

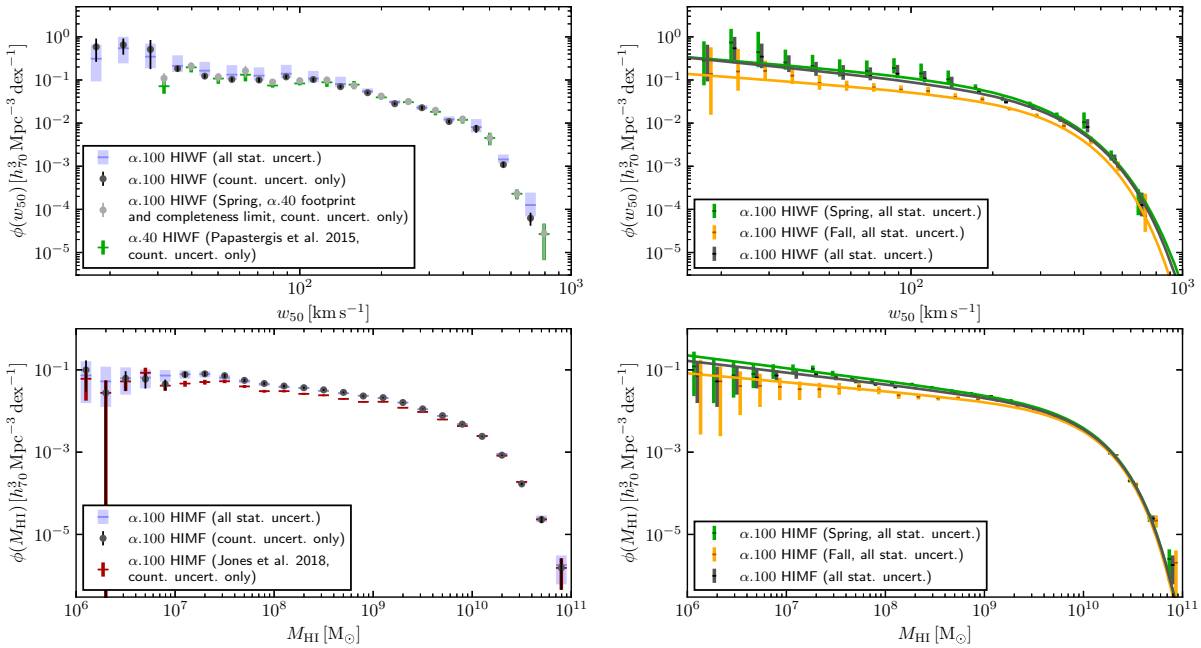

Figure 1. *Upper left:* the H I WF as measured from the complete α.100 release of the ALFALFA survey. The black points with error bars show the measurement accounting only counting (Poisson) uncertainties, whilst the blue bars and boxes also take into account distance, flux, and line-width measurement uncertainties. All error bars represent 68 per cent confidence intervals. We show for comparison the measurement of P15 (green crosses) based on the α.40 catalogue, and our measurement if we use the α.100 catalogue cut back to the same area on sky as used by P15, and reverting to the H11 completeness limit (grey points). *Lower left:* our measurement of the H I MF measured from the α.100 catalogue accounting only for counting uncertainties (black points), and for all statistical uncertainties (blue bars and boxes), which closely reproduces the previously published measurement by J18 (red crosses). *Upper right:* The H I WF measured from the α.100 catalogue, separately for the ALFALFA spring (green) and fall (orange) sky areas; uncertainties are 68 per cent confidence intervals and include all statistical uncertainties. The black bars and boxes are identical to the blue ones from the upper left panel; the green and orange ones have been slightly offset horizontally for clarity. The curves show the modified Schechter functions (equation 3) describing the data (see Section 4). The shapes of the orange and green curves are consistent: within the uncertainties, they differ only in their normalizations. *Lower right:* the H I MF measured from the α.100 catalogue, separately for the spring (green) and fall (orange) sky areas – the black bars and boxes reproduce the blue ones from the lower left panel. The curves show the Schechter functions (equation 2) describing the data (see Section 4). The shapes of the curves differ significantly, as also reported by J18 (see their figs 3 and 7).

percentile interval as an estimate of the total statistical uncertainty. This is shown with the blue bars and boxes in the upper left panel of Fig. 1. (We repeat the same exercise for the H I MF, the result is shown with the blue bars and boxes in the lower left panel of Fig. 1.)

These blue bars and boxes are reproduced in black in the upper right panel of Fig. 1. We have also repeated the same process for two independent subsets of our α.100 sample, corresponding to the spring (green bars and boxes) and fall (orange bars and boxes) areas on sky. The two subsets of the survey are treated fully independently, i.e. we re-determine the $V_{\rm eff}$ values for each source using only the sources in the same subset in the calculation. The spring and fall H I WFs have similar shapes (quantified below), but are offset 'vertically' in overall number density. This further illustrates the main driver of the disagreement with the α.40-based measurement of P15: that sample is drawn only from the spring region, which lies slightly above the full α.100 H I WF.

The H I *mass* function has a significantly different shape in the two sky regions: the spring sky has a steeper low-mass slope, and a slightly higher 'knee' mass, where the approximately exponential decline at high masses begins. The difference in shape was extensively discussed in J18, to which we refer for full details. We illustrate this difference in shape in the lower right panel of Fig. 1, which is analogous to the upper right panel, but for the H I MF, rather than the H I WF. We parametrize the shape of the H I MF with a Schechter function $\phi_{\rm S}$ (e.g. J18):

$$\frac{\phi_{\rm S}(M_{\rm HI})}{\phi_\star} = \log(10)\left(\frac{M_{\rm HI}}{M_\star}\right)^{\alpha+1} {\rm e}^{-M_{\rm HI}/M_\star} \quad (2)$$

where $\phi_\star$ is the overall normalization, $M_\star$ is the 'knee mass', and $\alpha + 1$ is the low-mass slope. We constrain the values of these parameters by sampling the posterior probability distribution for the parameters for each of our 1000 re-sampled catalogues using a Markov chain Monte Carlo sampling.[7] We use the likelihood function $\log \mathcal{L} \propto -\sum_i((\phi_i - \phi_{\rm S}(M_{{\rm HI},i}))/\sigma_i)^2$, where the index i runs over the points, and σ_i is the Poisson error estimate on ϕ_i. We then take the union of the 1000 samplings as an estimate of the posterior probability distribution accounting for all statistical uncertainties.[8] The median parameter values and their 95 per cent

[7]We use the affine-invariant ensemble sampler implementation EMCEE (Foreman-Mackey et al. 2013).

[8]We could also have chosen to use the H I MF with our estimates of the full statistical uncertainties, i.e. the bars and boxes in the right-hand panels of Fig. 1, to constrain the Schechter function parameters directly, but we feel that our adopted approach likely better reflects the true confidence intervals for the parameters. However, we have also checked explicitly that the two approaches give statistically consistent results, for both the H I MF and H I WF.

Table 1. Parameter values and 95 per cent confidence intervals for the (modified) Schechter functions describing the H I MF and H I WF. We give both the counting (Poisson) uncertainties alone, and the total statistical uncertainties. See Section 5 for a discussion of the uncertainty budget of the H I WF, including systematics, and see J18 for a detailed discussion of the uncertainty budget of the α.100 H I MF. The values given here are for the 2.5th, 50th, and 97.5th percentiles of the marginalized distributions; the curves in Fig. 1 correspond to the median values when all statistical uncertainties are accounted for – see Section 4 for further details.

Sample	Schechter function parameters (H I MF) $\log_{10}\phi_\star$ $[h_{70}^3\,\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}]$	$\log_{10}M_\star$ $(\mathrm{M}_\odot)$	α	
	Counting (Poisson) uncertainties only			
α.100	$-2.23^{+0.01}_{-0.01}$	$9.90^{+0.01}_{-0.01}$	$-1.27^{+0.01}_{-0.01}$	
α.100 Spring	$-2.18^{+0.02}_{-0.02}$	$9.89^{+0.01}_{-0.01}$	$-1.27^{+0.01}_{-0.01}$	
α.100 Fall	$-2.29^{+0.02}_{-0.02}$	$9.89^{+0.02}_{-0.02}$	$-1.22^{+0.02}_{-0.02}$	
	All statistical uncertainties			
α.100	$-2.26^{+0.02}_{-0.02}$	$9.92^{+0.01}_{-0.01}$	$-1.29^{+0.02}_{-0.02}$	
α.100 Spring	$-2.23^{+0.03}_{-0.02}$	$9.92^{+0.02}_{-0.02}$	$-1.31^{+0.03}_{-0.02}$	
α.100 Fall	$-2.30^{+0.03}_{-0.03}$	$9.90^{+0.02}_{-0.02}$	$-1.22^{+0.03}_{-0.03}$	
	Modified Schechter function parameters (H I WF)			
Sample	$\log_{10}\phi_\star$ $[h_{70}^3\,\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}]$	$w_\star$ $[\mathrm{km\,s}^{-1}]$	α	β
	Counting (Poisson) uncertainties only			
α.100	$-1.67^{+0.21}_{-0.16}$	300^{+48}_{-58}	$-0.56^{+0.21}_{-0.16}$	$2.1^{+0.4}_{-0.3}$
α.100 Spring	$-1.51^{+0.28}_{-0.22}$	279^{+65}_{-78}	$-0.50^{+0.31}_{-0.22}$	$1.9^{+0.5}_{-0.4}$
α.100 Fall	$-2.14^{+0.15}_{-0.12}$	382^{+37}_{-45}	$-0.82^{+0.16}_{-0.13}$	$2.7^{+0.5}_{-0.4}$
	All statistical uncertainties			
α.100	$-1.67^{+0.76}_{-0.42}$	307^{+127}_{-226}	$-0.63^{+1.12}_{-0.41}$	$2.0^{+1.3}_{-1.0}$
α.100 Spring	$-1.46^{+0.74}_{-0.58}$	268^{+173}_{-243}	$-0.50^{+1.75}_{-0.56}$	$1.8^{+1.6}_{-1.0}$
α.100 Fall	$-1.85^{+0.64}_{-0.36}$	310^{+105}_{-251}	$-0.48^{+1.64}_{-0.40}$	$2.1^{+1.1}_{-1.2}$

confidence intervals are summarized in Table 1, and these median parameter values are used to plot the curves in the right-hand panels of Fig. 1.

The situation is qualitatively different when we compare the H I *width* function for the spring and fall skies. We quantify their shapes, proceeding similarly to above, but using a modified Schechter function $\phi_{\rm MS}$, as is conventional for the H I WF (e.g. P15):

$$\frac{\phi_{\rm MS}(w_{50})}{\phi_\star} = \log(10)\left(\frac{w_{50}}{w_\star}\right)^{\alpha} \mathrm{e}^{-(w_{50}/w_\star)^{\beta}} \quad (3)$$

where $\phi_\star$ is again the normalization, $w_\star$ is the 'knee velocity width', α is the low-mass slope,[9] and the additional parameter β controls how sharply the exponential decay enters at high velocity widths. The parameter values are again summarized in Table 1, and we show the 1D and 2D marginalized posterior probability distributions in Fig. 2 (dark grey solid, and green and orange filled, contours and histograms). While there is a significant difference in the overall normalization $\phi_\star$ (panels g, h, i, and j in Fig. 2), the three parameters (α, β, $w_\star$) describing the shape of the H I WF in the spring and fall skies have statistically consistent values. This implies that, unlike the H I MF, the H I WF of gas-rich galaxies is consistent with having a universal shape, as might arise from an underlying universal HMF. However, we note that the confidence intervals are noticeably wider than for the cognate parameters describing the shape of the H I MF. We next discuss possible caveats to, and implications of, this observation.

(Our measured H I MF and H I WF amplitudes and uncertainties, including all those shown in Fig. 1 (for α.100), are tabulated in Appendix C.)

[9]Note this is not $\alpha + 1$ as in equation (2); we have defined the functions for ease of comparison with earlier work based on ALFALFA.

5 DISCUSSION

We first briefly comment further on the statistical uncertainties in our measurements (Section 5.1) before moving on to possible systematic uncertainties (Section 5.2). Finally, we outline our interpretation of our measurement in Section 5.3.

5.1 Statistical uncertainties

When accounting only for counting uncertainties, the 95 per cent confidence regions for the spring and fall regions involving the normalization $\phi_\star$ (Fig. 2 panels g, h, and i, dotted green and orange contours) are well-separated. When all statistical uncertainties are accounted for, although the 1D marginalized distributions for $\phi_\star$ (panel j, filled histograms) overlap much more than they did when accounting only for counting uncertainties (green and orange dotted histograms), it is clear from some of the 2D projections (especially

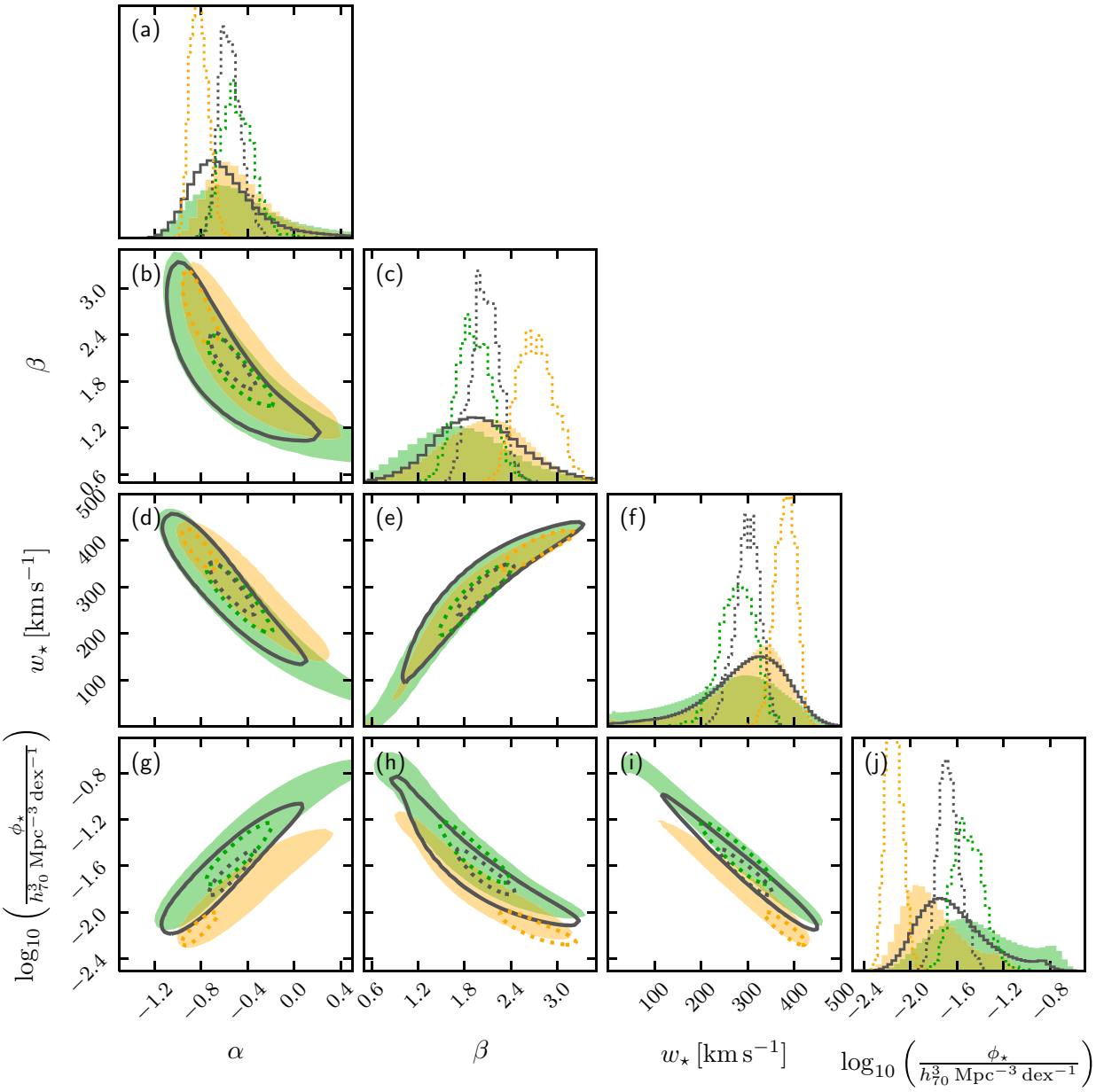

Figure 2. 1D and 2D marginalized posterior probability distributions for the parameters of the modified Schechter function (equation 3). All contours are drawn at the 95 per cent confidence level. Dark grey dotted contours and histograms correspond to the fit to the H I WF of the full α.100 sample when only counting (Poisson) uncertainties are accounted for. Dark grey solid contours and histograms are similar, but include all statistical uncertainties (see Section 4 for details). Dark green and orange contours and dotted histograms show the parameter constraints for the spring and fall sky areas, respectively, accounting only for counting uncertainties, whilst the light green and orange contours and filled histograms show the constraints when all statistical uncertainties are accounted for. The confidence regions for the spring and fall regions overlap significantly, except where the normalization, $\phi_\star$, is involved (panels g, h, i, and j). A similar figure for the H I MF parameters is included in Appendix D.

panels g and i) that there is still an evidence for a difference in the H I WF normalization.

We note that the importance of the measurement uncertainties relative to the counting uncertainties is much greater for the H I WF than it is for the H I MF (e.g. J18, table 1) – this is because the w_{50} uncertainties have a negligible influence on the H I MF, but contribute significantly to the H I WF uncertainty budget.

By happenstance, the confidence region for the fall α.100 sample when not accounting for measurement uncertainties is located in a low probability density region, near the edge[10] of its parent distribution (i.e. including all measurement uncertainties), as can be seen in Fig. 2 (panels b, d, e, g, h, and i). We point out this statistical

[10]Though not shown, in all cases, the peaks of all 2D marginalized distribution in Fig. 2 lie approximately at the centres of the plotted 95 per cent confidence regions.

curiosity only to note that it does not influence our interpretation below.

We include a figure similar to Fig. 2 for the H I MF parameters in Appendix D, Fig. D1.

5.2 Systematic uncertainties

In the context of our main focus in this work, that is, assessing whether the shapes of the H I WFs in the spring and fall sky regions are indeed indistinguishable at a level of precision accessible using $\alpha.100$, we are primarily interested in possible systematic errors which could have independent effects in the two regions. Systematic errors expected to affect parameter estimates in both regions in the same way will widen the confidence intervals somewhat, but not affect our qualitative conclusions. We refer to J18, e.g. table 1 as a guide to some of the sources of systematic uncertainty to be considered.

In addition, we discuss in Section 5.2.6 a systematic effect which has not to our knowledge been previously explored in detail and may see the low-velocity slope of the H I WF increase by ∼40 per cent in future, deeper surveys.

5.2.1 Survey boundary

The measurement of the H I WF is sensitive to the survey boundary chosen. The fiducial boundary which we have used throughout our analysis aims to maximize the area whilst avoiding the irregular coverage near the edges due to slight differences in the beginning and end in right ascension of ALFALFA drift scans. We have explicitly checked the effect of instead adopting the 'strict' boundary defined in J18 tables D1–D4: this slightly widens the confidence intervals for all parameters, but leaves our conclusions unaffected.

5.2.2 Sample variance

We have already established that the H I WF is consistent with having the same shape in the spring and fall areas of the ALFALFA survey, within the statistical uncertainties. We now explore the possibility of spatial variations on smaller scales. We follow J18 and quantify this by jackknifing the calculation of the H I WF: we split the survey area into 42 approximately equal-area sub-regions (27 and 15 in the spring and fall regions, respectively) and determine the H I WF, removing each sub-region in turn. For the spring region, this yields an estimate of the systematic uncertainty due to 'sample variance' of (0.006, 0.02, 2, 0.006) for the parameters (α, β, $w_\star$/km s^{-1}, $\log_{10}(\phi_\star/h_{70}^3$ Mpc^{-3}dex^{-1})), and of (0.02, 0.04, 5, 0.02) for the fall region. Even if this systematic error were to conspire to move the two posterior probability distributions directly apart in parameter space (which we deem unlikely), they would still overlap substantially.

5.2.3 Distance model

In order to estimate the influence of the distance model on our estimate, we replace the Masters (2005) flow model used in the fiducial analysis presented above with a very different model: we simply assume Hubble flow distances $D = v_{\rm CMB}H_0$ for all sources. We then repeat our analysis, imposing a minimum distance of 7 Mpc (as in P15), instead of 0 Mpc, to avoid the region giving rise to the most severe distance errors due to peculiar velocities. Even with these radically different distance estimates, the posterior probability distributions (accounting for all statistical uncertainties) for α, β, and $w_\star$ still overlap substantially.

We note that the H I WF is more sensitive to distance errors than the H I MF. This is because if, for example, the distance of a source is underestimated, its mass will consequently also be underestimated (equation 1). This in turn causes the effective volume ($V_{\rm eff}$) in which the source would be accessible to the survey to be underestimated, causing an overestimate of the true abundance of similar sources. For the H I WF, this leads directly to an overestimate of the number density in the bin containing the w_{50} of the source. Such errors can easily be severe, as seen in Section 2 above. For the H I MF, however, the affected bin is that of the (incorrect) mass estimated for the source, where the number density is higher than that associated with the actual source in question (due to the negative slope), leading to much smaller relative errors. An overestimated distance, on the other hand, causes $V_{\rm eff}$ to be overestimated, effectively down-weighting the contribution of the source to the H I WF and H I MF.

5.2.4 Absolute flux and velocity width calibration

The absolute flux calibration of the ALFALFA survey directly influences the H I MF (Schechter function) parameter $M_\star$, as adjusting the calibration directly causes a 'horizontal' shift of the H I MF. The derivation of the H I WF, however, effectively marginalizes over this effect, making it essentially insensitive to the flux calibration. The equivalent effect for the H I WF would be a systematic error in the determination of w_{50}. The w_{50} measurements in the $\alpha.100$ catalogue are corrected for instrumental broadening following Springob et al. (2005); the correction depends primarily on the signal-to-noise ratio of the spectrum in question. Whilst some residual systematic bias likely remains after this correction, it is likely to be very small, and it seems unlikely that it would affect sources in the two sky areas differently. For instance, if line widths for all low signal-to-noise sources (considering those above the cut for Code 1 sources) in the spring sky are slightly overestimated, then the same is likely to be true for all sources in the fall sky, leading to an equal 'horizontal' shift in the H I WFs in both regions – we note that the distribution of signal-to-noise ratios for sources in our spring and fall H I WF samples are near-identical.

5.2.5 Adopted completeness limit

A mismatch between the true sensitivity of the ALFALFA survey and that implied by the completeness limit assumed (see Appendix A) in the analysis in Section 4 can lead to systematic biases in the recovered H I MF and H I WF. However, for reasonable variations in the assumed limit, the changes to both are small. For example, using the completeness limit of H11 derived from the $\alpha.40$ catalogue, which is offset 'down' by 0.02 dex in S_{21} relative to that we derive in the Appendix, the parameters of the modified Schechter function describing the H I WF of the full survey (including all statistical errors) change by $-0.10\,h_{70}^3$ Mpc^{-3} dex^{-1} ($\phi_\star$), $+4$ km s^{-1} ($w_\star$), $+0.03$ (α) and 0.0 (β), in all cases by much less than our quoted uncertainties.

We have also considered whether the completeness limit of the survey may differ between the two survey regions. We assess this by following the approach of H11 (Section 6), using the sources from the two survey regions as separate inputs. We find tentative evidence that the survey is slightly shallower in the fall region (offset from our fiducial 50 per cent completeness limit by 0.009 dex in S_{21} at all w_{50}), whilst in the spring region the coverage may be

slightly deeper (by 0.011 dex), for a net difference between the two regions of 0.02 dex. We note that this is one of the rare systematic uncertainties which can realistically cause opposite biases in the H I WF measurements in the two regions, however the quantitative differences are small: the 0.02 dex offset is similar in magnitude to the offset between our fiducial 50 per cent completeness limit and that from H11 just discussed, and causes similarly small changes to the parameters describing the H I WF shape. The changes in the H I MF shape following reasonable changes to the assumed completeness limit are likewise small (though in this case sometimes comparable to the statistical uncertainties, which are relatively much smaller for the Schechter function parameters than the modified Schechter function parameters). Our assessment is therefore that the uncertainty in the exact completeness limit of the ALFALFA survey is insufficient to qualitatively affect our conclusions.

5.2.6 *The unseen portion of the galaxy population*

In the $1/V_{\rm eff}$ estimator, any 2D bin in $M_{\rm HI}$ and w_{50} in which the survey has a detection count of zero makes no contribution to the H I MF or H I WF, even if the number density of such sources could be quite high and zero were detected simply because the survey is only capable of detecting them when they are very nearby. Formally, the severity of the systematic error due to this effect is unbounded: there is always a space in the survey volume to hide very high abundances of low-mass and/or high-velocity-width galaxies. Its importance must therefore be evaluated in the context of some prior assumption about the properties of the intrinsic population of galaxies sampled by the survey. We show in this section that for ALFALFA, this source of systematic error could plausibly cause the estimate of the low-velocity slope of the H I WF to be too shallow by $\sim$0.2, whilst the impact on the H I MF is likely much less severe.[11] (The other parameters are also affected, but we focus on the slope to streamline our discussion.) We note that, to our knowledge, this systematic error affects nearly all[12] previously published measurements of both the H I MF and H I WF; the severity of the error will depend on the details of the survey and analysis in question.

To illustrate the source of the error, we construct a simple mock survey, which we label 'mock (a)'. Full details are given in Appendix E; we summarize here. We draw a large set of H I velocity widths from a distribution described by a modified Schechter function with parameters close to those measured for ALFALFA, $(w_\star/{\rm km\,s^{-1}}, \alpha, \beta) = (320, -0.5, 2.2)$. Distances are randomly assigned such that the source population is uniformly distributed in the survey volume. We then assign an H I mass to each source such that (i) the H I MF of the population is close to that measured for ALFALFA, within 10 per cent at all masses of a Schechter function with parameters $(\log_{10}(M_\star/{\rm M}_\odot), \alpha) = (9.94, -1.25)$, and (ii) the mock source count distribution (i.e. after applying the selection function as described below) in the $M_{\rm HI}$–w_{50} plane resembles that observed in ALFALFA as closely as possible. This was achieved brute-force 'by hand', for reasons detailed in the appendix. The flux of each mock source was computed from its H I mass and distance, and randomly selected sources were removed from the sample until the number with fluxes and line widths placing them above the ALFALFA 50 per cent completeness limit (H11, equations 4 and 5) was equal to 21 827, the total count used in our analysis in Section 4. We recorded two catalogues, the 'intrinsic population' including all the mock sources, and the 'mock survey' including only those above the completeness threshold. The $M_{\rm HI}$–w_{50} distributions of both catalogues are shown in the main left panel of Fig. 3 with the contours and heat map, respectively.

The right-hand panels of Fig. 3 clearly illustrate the origin of the systematic error. The contours are repeated from the left-hand panel – this is the intrinsic population which the $1/V_{\rm eff}$ estimator should ideally reconstruct. The heat map in this panel shows the source counts estimated for the entire survey volume using the $1/V_{\rm eff}$ method. Whilst the incompleteness-corrected counts closely approximate the intrinsic population across all areas where the mock survey source count is >0, counts in cells where there are 0 mock surveyed galaxies are left set to 0. This results in an appreciable number of sources in the intrinsic population being 'missed' by the estimator, predominantly towards the upper left of the figure. When the incompleteness-corrected counts are integrated along the two axes to give the H I MF and H I MF, it turns out that the input H I MF is recovered accurately, $(\log_{10}(M_\star/{\rm M}_\odot), \alpha) = (9.94, -1.22)$, but the H I WF has a slope that is noticeably too shallow at the low-velocity end, with $(w_\star/{\rm km\,s^{-1}}, \alpha, \beta) = (321, -0.25, 2.2)$. We have repeated this exercise with many independently realized mock surveys and confirm that the effect is systematic, consistently resulting in a low-velocity slope α of about -0.3.

Given that an intrinsic H I WF slope of $\alpha = -0.50$ results in a slope of about -0.3 recovered by the $1/V_{\rm eff}$ estimator, it is a plausible hypothesis that a steeper intrinsic slope might result in a (biased) recovered slope close to that measured in ALFALFA. We verify this by constructing two more mock samples following the same approach described above, but with intrinsic slopes of $\alpha = -0.7$ – 'mock (b)' – and $\alpha = -1.2$ – 'mock (c)'. The mock samples are illustrated in Fig. 4, similar to the left-hand panel of Fig. 3. These steeper intrinsic slopes are also both underestimated by the $1/V_{\rm eff}$ estimator, giving $\alpha = -0.56$ and -0.78 for mocks (b) and (c), respectively. This confirms (i) that this bias can plausibly lead to a measurement similar to that for ALFALFA ($\alpha = -0.56$) when the true slope is about -0.7 and (ii) that the bias likely does not cause the estimator to 'saturate' (e.g. always return the same slope estimate for ever increasing intrinsic slope) over a reasonably wide range in intrinsic α.

We make one more observation regarding the plausibility of this bias significantly affecting the ALFALFA measurement. We found it challenging to reproduce the shape of the ALFALFA source count distribution in the $M_{\rm HI}$–w_{50} plane, but had the most success when the intrinsic slope was $\alpha \sim -0.7$. This can be seen in Fig. 4: the distribution for mock (b) (lower left) is the closest match to the ALFALFA distribution (upper right). Mock (a; upper right) has a distribution that is rather too narrow at low w_{50}, whilst mock (c; lower right) is somewhat too narrow everywhere. We consider this observation to be suggestive only, because the shape of the observed source count distribution depends on the shape chosen for the intrinsic distribution at fixed w_{50} – there is some freedom in this choice, and we have not exhaustively explored all possibilities.

To our knowledge, the influence of this systematic error on the H I WF has not previously been thoroughly explored, though its influence on the H I MF is acknowledged by Martin et al. (2010).

[11] The two biases – that of the low-velocity slope of the H I WF and that of the low-mass slope of the H I MF – have very different magnitudes because neither the (possible) shape of the intrinsic distribution of galaxies in the $M_{\rm HI}$–w_{50} plane, nor the selection function of the survey are close to symmetric with respect to the two relevant parameters.

[12] The volume-limited, optically selected sample of Klypin et al. (2015) is likely immune. Interestingly, those authors find a somewhat steeper low-width slope for the H I WF than in ALFALFA (e.g. their fig. 10); this discrepancy could plausibly be quantitatively resolved by the bias described in this section.

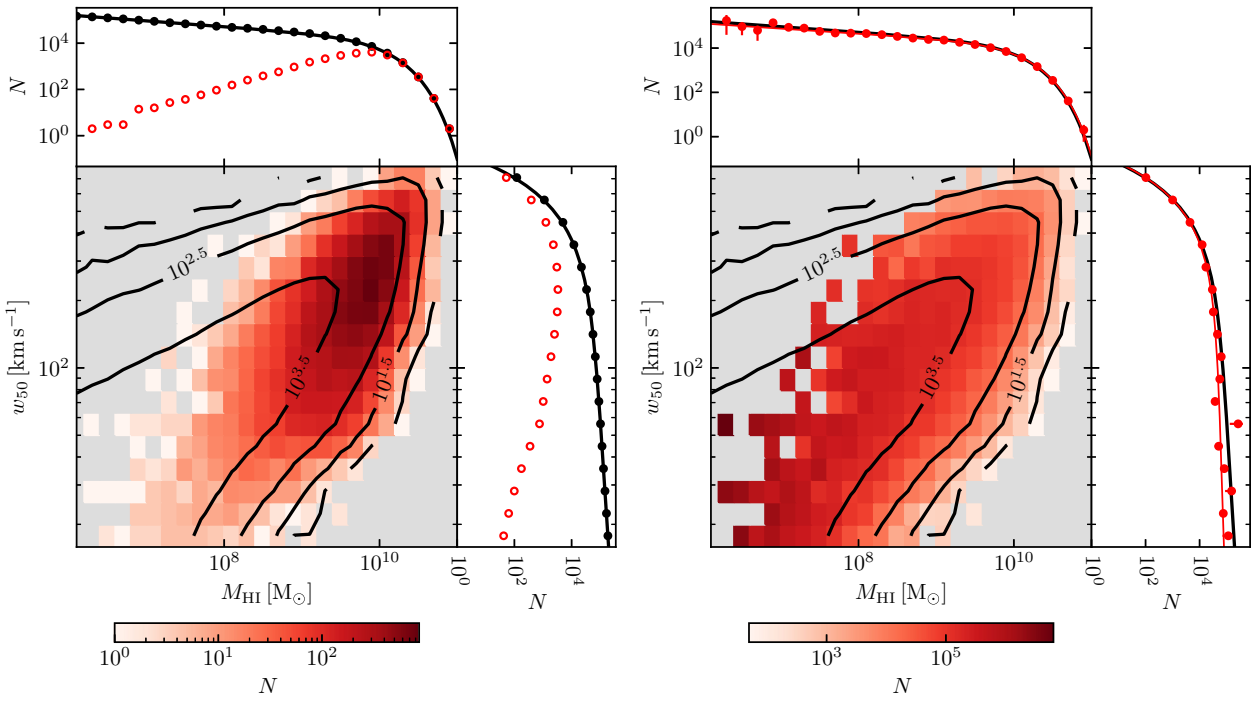

Figure 3. *Left*: a simple mock survey resembling ALFALFA – mock (a) – as described in Section 5.2.6. The contours show the 'intrinsic' source population distribution as a function of $M_{\rm HI}$ and w_{50}. The sources are uniformly distributed throughout the survey volume; the heat map shows those which have a flux and linewidth above the ALFALFA 50 per cent completeness limit, i.e. the 'observed' sources. The upper inset panel shows the H I MF of the intrinsic population (black points) and the observed population (red points), and the Schechter function that the mock sample is constructed to follow (black line). The right inset is similar but shows the H I WF; the sources are drawn from the distribution corresponding to the modified Schechter function shown (black line). *Right*: similar to left, but showing the intrinsic population reconstructed from the observed population using the $1/V_{\rm eff}$ estimator (heat map). The contours in the main panel and black lines in the insets are repeated from left. The insets show the recovered H I MF and H I WF (red points, error bars show counting errors only) and a (modified) Schechter function fit to each (red line). A significant portion of the intrinsic population is missed in the reconstruction in the region where there are no observed sources (low $M_{\rm HI}$, high w_{50}). The H I MF is reasonably well-recovered despite this, but the H I WF is biased towards a shallower slope at the low-velocity end.

They estimate its importance using a mock catalogue constructed by fitting the distribution of observed sources and extrapolating the tail of the distribution. We note that this is qualitatively different from our approach (Appendix E); we use the same form for the distribution as them but assume it describes the intrinsic population, fixing its parameters to match the observed counts once incompleteness is accounted for. Despite this, our conclusions agree: they estimate a bias of about 2 per cent (0.025 out of −1.25) in the low-mass slope of the H I MF; for mocks (a), (b), and (c), the same parameter is underestimated by 2, 0, and 5 per cent, respectively. However, their approach would likely underestimate the bias for the H I WF.

This bias significantly affects the low-velocity slope of the H I WF shown in Fig. 1 is difficult to dispute: the sharp edge of the distribution towards high w_{50} and low $M_{\rm HI}$ reconstructed by the $1/V_{\rm eff}$ estimator (Fig. 3 heat map in right-hand panel[13]) is highly unlikely to be the true edge of the galaxy population in this plane. However, the fraction of galaxies 'missed' due to this error depends entirely on the shape of the distribution precisely where it is not constrained by measurement. Whilst we judge our estimates above to be plausible, definitively settling this issue requires a deeper survey, or perhaps a search optimized for low signal-to-noise, large velocity width sources in existing surveys.

[13]Whilst this is for a mock sample, the equivalent for the actual ALFALFA survey is very similar, e.g. P15 fig. A.1.

5.3 Interpretation

5.3.1 Connection to the HMF

We have shown above that spatial variation in the ALFALFA H I WF is qualitatively different to that in the H I MF. As shown in Fig. 1 and discussed in detail by J18, the latter has a significantly shallower low-mass slope in the lower-density fall survey region than in the higher-density spring survey region. The H I WF, on the other hand, has a low-velocity slope and overall shape that appear to be spatially invariant, at least within the confidence intervals implied by the data. Though there are several effects likely to cause systematic biases in the H I WF measurement, we have not identified any likely to affect the two survey regions significantly differently, suggesting that the spatial invariance may not be a simple coincidence, but instead arise from an underlying spatially invariant HMF. We acknowledge that the confidence intervals for the parameters describing the shape of the ALFALFA H I WF remain relatively wide and will therefore be interested to see whether the insensitivity to density on scales of a few tens of megaparsecs persists in forthcoming, larger surveys, such as WALLABY (Koribalski et al. 2020). In this section, we assume that this spatial invariance is a real feature of the galaxy population and proceed to explore its implications.

Barring some conspiratorial coincidence, the similarity in the shape of the H I WF in the ALFALFA spring and fall skies, despite the difference in the shape of the H I MF, suggests a more fundamental

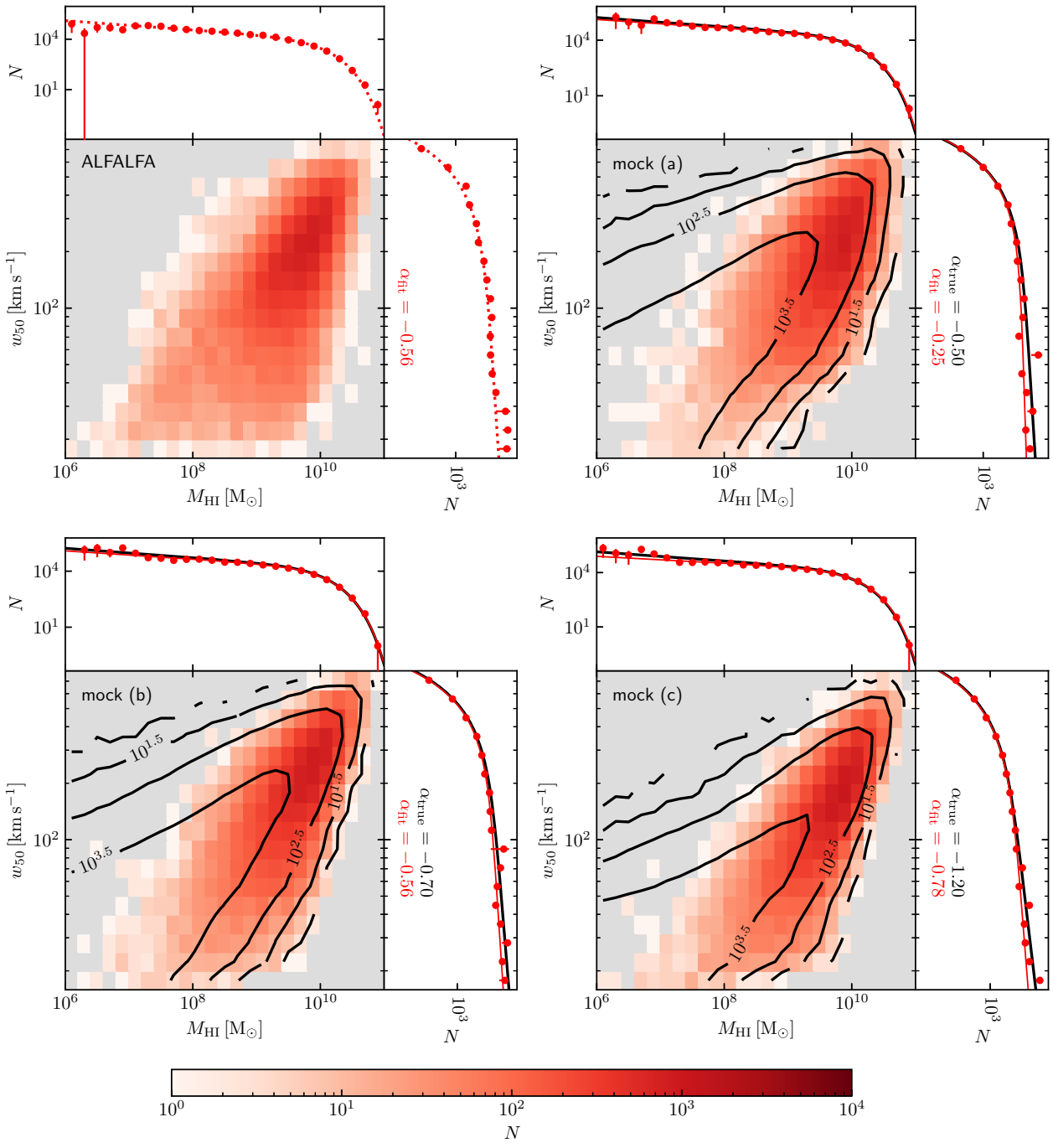

Figure 4. In the upper left group of panels, the main panel shows the observed counts in $\alpha.100$ as a function of $M_{\rm HI}$ and w_{50}. The upper and right-hand panels show the H I MF and H I WF as recovered using the $1/V_{\rm eff}$ estimator (points with error bars), respectively, and (modified) Schechter function fits thereto (dotted lines). The main panel in the other three sets of panels show the intrinsic counts in a mock galaxy population (mocks (a), (b), and (c), as labelled – see Section 5.2.6 and Appendix E) with contour lines, and the mock 'observed' counts for each once the ALFALFA selection function is applied with the colour maps. The upper and right-hand panels in each set show the H I MF and H I WF, respectively: the 'true' function (heavy black line), the $1/V_{\rm eff}$ estimate (points with error bars), and a (modified) Schechter function fit to the estimate (thin red line). The intrinsic and recovered slopes of the H I WF are denoted $\alpha_{\rm true}$ and $\alpha_{\rm fit}$, respectively. The slope is consistently underestimated, with the intrinsic slope of -0.70 for mock (b) resulting in a recovered slope of -0.56, matching the slope of -0.56 measured for $\alpha.100$.

similarity in the underlying galaxy population. We assume here the well-motivated hypothesis that the HMF in the two regions is similar except for its overall normalization (e.g. Crain et al. 2009), and explore where this assumption leads us below.

The 21-cm emission spectrum of a galaxy arises through a combination of the geometry of the system (e.g. radial surface density profile in the disc and vertical density profile falling away from the mid-plane; inclination) and its kinematics (rotation curve,

vertical rotation velocity gradient, and velocity dispersion). That the inclination-corrected line width is often a fair proxy for the maximum circular velocity of the dark matter halo comes about because most of the H I gas rotates at a similar speed: most of the H I is at larger radii in the disc (because the central H I surface density profiles of galaxies are typically fairly flat, e.g. Wang et al. 2016) where the rotation curve is often flat or has a shallow gradient (e.g. Verheijen 2001). This means that substantial changes in the inclination-corrected linewidth of a galaxy are only expected if (i) the gas disc is severely truncated, such that most of the gas is at radii corresponding to the rising part of the rotation curve, or (ii) the total mass distribution changes such that the maximum circular velocity is significantly affected. To preserve the common shape of the H I WF in the spring and fall skies, neither of these effects should be much stronger for galaxies in one sky region compared to the other.

We first consider what happens when a disc of H I gas accretes additional gas, or is stripped or otherwise depleted. It is well established (e.g. Broeils & Rhee 1997; Verheijen & Sancisi 2001; Wang et al. 2016) that the H I mass and diameter of galaxies are very tightly correlated. The authors of the latter reference point out that this implies H I discs grow/shrink in a well-regulated way as gas is accreted or consumed. This is further developed with the help of a series of analytic models by Stevens et al. (2019), who find that galaxies likely evolve almost exactly along the correlation as their gas is consumed or stripped. Naluminsa, Elson & Jarrett (2021) further note that the $M_{\rm HI}$–$D_{\rm HI}$ correlation has no obvious connection to environment (though we feel this point could still be studied in more detail). The persistence of this very tight correlation in spite of even the messier aspects of galaxy evolution leads to a very useful point in the present discussion: if an H I disc is truncated in radial extent, sufficiently that the portion of the rotation curve which it traces does not reach the outer, flat part, it must also necessarily be depleted in H I mass, relative to its halo mass. Therefore, the less a galaxy's inclination-corrected 21-cm spectrum traces the circular velocity of its halo (due to limited extent of the H I disc), the less likely it is to be detected in a flux-limited survey.

We next consider the conditions necessary to cause a substantial change to the maximum circular velocity of a galaxy. The most straightforward way for this to occur is via the stripping of large amounts of material. As the components which usually make the predominant contribution to the maximum circular velocity – that is, dark matter, and for more massive galaxies, stars – are collisionless, tidal stripping is the most important physical process in this context. By their nature, tides act to remove material approximately outside-in. For a typical galaxy, the maximum circular velocity actually does not change very much until a majority fraction of the total mass of the galaxy has been stripped (Peñarrubia, Navarro & McConnachie 2008); for example, a galaxy having lost 50 per cent of its total mass has its maximum circular velocity drop by only $\lesssim$15 per cent. Furthermore, usually no substantial amount of stellar mass is stripped until $\gtrsim$90 per cent of the dark matter has already been lost. The maximum circular velocity is therefore nearly a constant until rather late in the evolution of a satellite galaxy. In the meantime, there has likely been ample opportunity for the neutral gas disc to be stripped by ram pressure, consumed (and not replenished) by star formation, expelled by supernova or AGN winds, or a combination of the above. Thus, those galaxies where the connection between the maximum circular velocity and the (theoretically idealized, e.g. 'pre-infall') halo mass are again the least likely to be detected in a 21-cm survey.

5.3.2 *Utility as a constraint on cosmology*

Putting the above considerations together, we propose the following qualitative interpretation of the shapes of the H I MF and H I WF in the ALFALFA spring and fall skies. We suggest that the spatial variation in the H I MF reflects the relatively fragile nature of H I in galaxies: the atomic gas content responds readily to the local environment (e.g. Jones et al. 2020, and references therein). Thus, the distribution of H I-to-total masses encoded in the H I MF reflects local features, such as the presence of gas-rich filaments in the broad region around the Virgo cluster driving up the low-mass slope of the H I MF in the ALFALFA spring sky, as suggested by J18. We further suggest that this same fragility somewhat counter-intuitively makes the H I WF less sensitive to local variations in the strength of 'environmental processes', because in most cases the H I disc may lose so much mass that it drops out of the survey, or be completely removed/destroyed, before w_{50} begins to be severely affected. To put this concisely, we suggest that environmental processes remove galaxies from the sample contributing to the H I WF more effectively than they re-shape the velocity width distribution of a galaxy population.

This leaves one remaining puzzle: for the shape of the H I WF to remain unaffected, the population of galaxies which are not H I-rich enough to be detected in a survey at any distance should not be differently biased in w_{50} in different survey regions. Such gas-poor galaxies tend to be (i) massive, in terms of stellar or total mass, (ii) satellites or (iii) both. Given this, it seems plausible that, within any reasonably large volume, the fraction of very H I-poor galaxies could be close enough to a constant at any given w_{50} to give rise to the observed spatial invariance of the H I WF. However, further theoretical work to reinforce or refute this hypothesis would be valuable.

To summarize: we argue that the H I WF is a naturally reasonably robust[14] tracer of the HMF of gas-rich galaxies because flux-limited surveys are intrinsically biased against detecting galaxies where w_{50} is a poor tracer of the halo mass. The spatial invariance of the H I WF, therefore, reflects that of the HMF, even when comparing regions of dissimilar average density, and even though the relative gas-richness of the same population (i.e. the H I MF) differs across the same regions.

Provided the interpretation outlined above holds, the position of the H I WF as a cosmologically interesting quantity is reinforced: future measurements (e.g. Koribalski et al. 2020) will deliver what may turn out to be the most stringent constraints on the dark matter HMF available, albeit for a biased selection of galaxies. In our opinion, the most fruitful way to exploit this will be for models to first reproduce the observed H I-bearing galaxy population in sufficient detail within a given cosmological context, then to forward-model (i.e. 'predict') the H I WF for comparison with actual measurements. This is an approach we plan to pursue in future work.

6 CONCLUSIONS

We have presented the first measurement of H I WF using the completed ALFALFA survey (α.100; Section 4). Whilst the shape of the H I MF is significantly different when the spring and fall portions of the survey are compared, the shape of the H I WF in the same two regions is indistinguishable within the uncertainties, although these remain relatively large (Sections 4 and 5.2). We tentatively interpret

[14]Not to be misunderstood as 'direct' or 'ideal'.

this as a signature of the fidelity of the H I line width as a tracer of the dynamical masses of galaxies (Section 5.3). We have also identified a previously oft-overlooked systematic bias particularly affecting the low-velocity slope of the H I WF (Section 5.2.6).

The scenario outlined in Section 5.3 implies a few predictions; it will be interesting to see if these are borne out in future, larger-volume 21-cm surveys, such as WALLABY (Koribalski et al. 2020) or an eventual SKA survey (e.g. Blyth et al. 2015). First and foremost, the spatial invariance of the shape of the H I WF can be confirmed or refuted as the statistical uncertainties shrink. Furthermore, we predict that the H I WF should only maintain its shape when reasonably large (but not necessarily 'cosmologically representative'), contiguous regions are compared. For example, the H I MF measured in a collection of spatially disconnected voids would be expected to drive a strong bias against massive DM haloes, and therefore drive a reduction in $w_\star$ (and likely also some change in α and β). Finally, more sensitive surveys should uncover a previously unseen population of low-mass, high-velocity width galaxies, alleviating the bias discussed in Section 5.2.6. Another interesting possibility expected to be enabled by future surveys is the measurement of an H I VF leveraging large numbers of spatially resolved H I sources to break the conventional reliance on ancillary optical data to estimate inclinations, or even replace line width measures with rotation velocities derived from kinematic models for thousands of sources. We look forward to harnessing the ALFALFA and future H I WF measurements as constraints on galaxy formation and cosmological models.

ACKNOWLEDGEMENTS

We are endebted to E. Papastergis for providing us his implementation of the $1/V_{\rm eff}$ estimator and accompanying notes. We thank M. Jones for providing tabulated results from J18 in electronic form. We thank both of the above, K. Hess and A. Benítez-Llambay for careful readings of draft versions of this article. We thank the anonymous referee for a careful review of the manuscript.

KAO acknowledges support by the European Research Council (ERC) through Advanced Investigator grant to C.S. Frenk, DMIDAS (GA 786910). This research has made use of NASA's Astrophysics Data System.

DATA AVAILABILITY

The $\alpha.100$ data release of the ALFALFA survey, including the optical counterparts from Durbala et al. (2020), is available at egg.astro.cornell.edu/alfalfa/data/.

REFERENCES

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Blyth S. et al., 2015, Proc. Sci., Exploring Neutral Hydrogen and Galaxy Evolution with the SKA. SISSA, Trieste, PoS#128
Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93
Broeils A. H., Rhee M. H., 1997, A&A, 324, 877
Brook C. B., Di Cintio A., 2015a, MNRAS, 450, 3920
Brook C. B., Di Cintio A., 2015b, MNRAS, 453, 2133
Brook C. B., Shankar F., 2016, MNRAS, 455, 3841
Brooks A. M., Papastergis E., Christensen C. R., Governato F., Stilp A., Quinn T. R., Wadsley J., 2017, ApJ, 850, 97
Chauhan G., Lagos C. d. P., Obreschkow D., Power C., Oman K., Elahi P. J., 2019, MNRAS, 488, 5898
Crain R. A. et al., 2009, MNRAS, 399, 1773
Durbala A., Finn R. A., Crone Odekon M., Haynes M. P., Koopmann R. A., O'Donoghue A. A., 2020, AJ, 160, 271
Dutta S., Khandai N., Rana S., 2021, preprint (arXiv:2108.03253)
Dutton A. A., Obreja A., Macciò A. V., 2019, MNRAS, 482, 5606
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Frenk C. S., White S. D. M., Davis M., Efstathiou G., 1988, ApJ, 327, 507
Giovanelli R. et al., 2005, AJ, 130, 2598
Haynes M. P. et al., 2011, AJ, 142, 170 (H11)
Haynes M. P. et al., 2018, ApJ, 861, 49
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Jones M. G., Papastergis E., Haynes M. P., Giovanelli R., 2015, MNRAS, 449, 1856
Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS, 477, 2 (J18)
Jones M. G., Hess K. M., Adams E. A. K., Verdes-Montenegro L., 2020, MNRAS, 494, 2090
Kaisin S. S., Karachentsev I. D., 2019, Astrophys. Bull., 74, 1
Kashibadze O. G., Karachentsev I. D., Karachentseva V. E., 2018, Astrophys. Bull., 73, 124
Klypin A., Karachentsev I., Makarov D., Nasonova O., 2015, MNRAS, 454, 1798
Koribalski B. S. et al., 2020, Ap&SS, 365, 118
Li P., Lelli F., McGaugh S., Pawlowski M. S., Zwaan M. A., Schombert J., 2019, ApJ, 886, L11
Macciò A. V., Udrescu S. M., Dutton A. A., Obreja A., Wang L., Stinson G. R., Kang X., 2016, MNRAS, 463, L69
Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359
Masters K. L., 2005, PhD thesis, Cornell University, New York, USA
Moorman C. M., Vogeley M. S., Hoyle F., Pan D. C., Haynes M. P., Giovanelli R., 2014, MNRAS, 444, 3559
Naluminsa E., Elson E. C., Jarrett T. H., 2021, MNRAS, 502, 5711
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Papastergis E., Shankar F., 2016, A&A, 591, A58
Papastergis E., Martin A. M., Giovanelli R., Haynes M. P., 2011, ApJ, 739, 38
Papastergis E., Giovanelli R., Haynes M. P., Shankar F., 2015, A&A, 574, A113 (P15)
Peñarrubia J., Navarro J. F., McConnachie A. W., 2008, ApJ, 673, 226
Ponomareva A. A., Verheijen M. A. W., Bosma A., 2016, MNRAS, 463, 4052
Saintonge A., 2007, AJ, 133, 2087
Sawala T. et al., 2015, MNRAS, 448, 2941
Sawala T. et al., 2016, MNRAS, 457, 1931
Schmidt M., 1968, ApJ, 151, 393
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Springob C. M., Haynes M. P., Giovanelli R., Kent B. R., 2005, ApJS, 160, 149
Stevens A. R. H., Diemer B., Lagos C. d. P., Nelson D., Obreschkow D., Wang J., Marinacci F., 2019, MNRAS, 490, 96
Verbeke R., Papastergis E., Ponomareva A. A., Rathi S., De Rijcke S., 2017, A&A, 607, A13
Verheijen M. A. W., 2001, ApJ, 563, 694
Verheijen M. A. W., Sancisi R., 2001, A&A, 370, 765
Wang J., Koribalski B. S., Serra P., van der Hulst T., Roychowdhury S., Kamphuis P., Chengalur J. N., 2016, MNRAS, 460, 2143
Zavala J., Jing Y. P., Faltenbacher A., Yepes G., Hoffman Y., Gottlöber S., Catinella B., 2009, ApJ, 700, 1779
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30
Zwaan M. A., Meyer M. J., Staveley-Smith L., 2010, MNRAS, 403, 1969

APPENDIX A: COMPLETENESS OF THE ALFALFA CATALOGUES

The completeness limits for the ALFALFA survey presented in H11 (equations 4 and 5) are derived empirically based on the $\alpha.40$ catalogue. Although, in principle, the fully completed survey should have the same completeness as the 40 per cent completed survey, since the former simply extends the sky coverage of the latter, the $\alpha.100$ catalogue offers the opportunity to both check this, and use the additional data to make a more precise determination of the limit. We have therefore repeated the calculation described in H11 (Section 6) using the $\alpha.100$ 'Code 1' sources. We also repeated the calculation using the $\alpha.40$ 'Code 1' sources to verify that we could reproduce their result. We find that the completeness limit implied by the $\alpha.40$ catalogue is in essentially the same location as reported by H11, though we find a somewhat steeper cutoff as a function of flux (i.e. the spacings between the 90 per cent, 50 per cent, and 25 per cent completeness curves are somewhat narrower) – we tentatively attribute this to small differences in the optimization routines used to fit the number counts and the completeness curves themselves. For completeness, we include our determination of the limits based on the $\alpha.40$ catalogue here, following the notation of H11 for the left-hand sides of the equations:

$$\log_{10} S_{21,90\text{ per cent,Code1}} = \begin{cases} 0.5W - 1.155, & W < 2.5 \\ W - 2.405, & W \geq 2.5 \end{cases} \quad \text{(A1)}$$

$$\log_{10} S_{21,50\text{ per cent,Code1}} = \begin{cases} 0.5W - 1.194, & W < 2.5 \\ W - 2.444, & W \geq 2.5 \end{cases} \quad \text{(A2)}$$

$$\log_{10} S_{21,25\text{ per cent,Code1}} = \begin{cases} 0.5W - 1.214, & W < 2.5 \\ W - 2.464, & W \geq 2.5, \end{cases} \quad \text{(A3)}$$

where $W = \log_{10}(w_{50}/\mathrm{km\,s^{-1}})$. Using the $\alpha.100$ catalogue, we derive the following completeness limits:

$$\log_{10} S_{21,90\text{ per cent,Code1}} = \begin{cases} 0.5W - 1.115, & W < 2.5 \\ W - 2.365, & W \geq 2.5 \end{cases} \quad \text{(A4)}$$

$$\log_{10} S_{21,50\text{ per cent,Code1}} = \begin{cases} 0.5W - 1.170, & W < 2.5 \\ W - 2.420, & W \geq 2.5 \end{cases} \quad \text{(A5)}$$

$$\log_{10} S_{21,25\text{ per cent,Code1}} = \begin{cases} 0.5W - 1.198, & W < 2.5 \\ W - 2.248, & W \geq 2.5. \end{cases} \quad \text{(A6)}$$

Note that these limits imply a completeness slightly (~0.02 dex in S_{21}) worse than what is implied by the $\alpha.40$ catalogue. We use the 50 per cent completeness limit in equation (A5) throughout this work, except where explicitly noted otherwise.

APPENDIX B: REPRODUCTION OF PREVIOUSLY PUBLISHED RESULTS

We can reproduce the H I WF measurement of P15 exactly by using the public $\alpha.40$ galaxy catalogue and the selection cuts listed in that work, but this requires three adjustments. First, P15 used a lower precision table of distances, so one object (AGC 7068) must be removed from the sample as its distance (7.03 Mpc) fails the P15 distance cut ($D > 7.0$ Mpc) at lower precision. Secondly, the original P15 implementation of binning in $M_{\rm HI}$ and w_{50} was subject to floating point roundoff errors, causing some objects on a bin boundary to be counted in the wrong bins, leading to small changes in the H I WF. We have corrected this error, but verified that adopting the original (erroneous) binning scheme results in an exact quantitative reproduction of their result. Thirdly, we must revert to the completeness limit given in H11 (equations 4 and 5). We also recover a near reproduction of the P15 H I WF by beginning with the $\alpha.100$ public catalogue, trimmed back to the $\alpha.40$ spring footprint (the region used by P15), and again applying the same selection cuts and binning as P15, and the H11 completeness limit – this measurement is shown with light grey points in the upper left panel of Fig. 1. The difference between the two measurements is due to the two reasons noted above, as well as two sources (AGC 749309 and AGC 257959) being re-classified as OH megamasers in $\alpha.100$ (Haynes et al. 2018).

The bulk of the difference between our measurement of the H I MF and that reported in J18 is due to our use of an updated completeness limit for $\alpha.100$ (equation A5). Other small quantitative differences are due to: (i) the removal of a few sources re-classified as OH megamasers between the publication of J18 and the $\alpha.100$ catalogue release (Haynes et al. 2018); (ii) our adjusted distances for a few objects (see Section 2); (iii) the fact that we do not impose a minimum $M_{\rm HI}$ (rather than 10^6 $\mathrm{M_\odot}$); (iv) our use of the $1/V_{\rm eff}$ estimator, rather than the nearly equivalent 2DSWML estimator.

APPENDIX C: TABULATED H I WF AND H I MF

In Tables C1 and C2, we tabulate our H I WF and H I MF measurements, i.e. including amongst others all values plotted in the right-hand panels of Fig. 1.

Table C1. Amplitudes and uncertainties of the H I WF for the full $\alpha.100$ survey (columns 2 and 3), and the spring and fall regions separately (columns 4 & 5 and 6 & 7, respectively). For each region, we give both the measurement including only counting uncertainties (left column of each pair) or all statistical uncertainties (right column of each pair), see Section 4 for details of uncertainty estimates. Note that although the logarithms of all amplitudes and uncertainties are given, the counting uncertainties are symmetric on a linear scale and should be interpreted as the 1σ width of a Gaussian distribution (not a lognormal distribution). The intervals including all statistical uncertainties run from the 16th to 84th percentiles of the average of the 1000 probability distributions each corresponding to a Monte Carlo-resampled galaxy catalogue.

	$\log_{10}\phi(w_{50})/h_{70}^3\,\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$					
	$\alpha.100$		$\alpha.100$ Spring		$\alpha.100$ Fall	
$\log_{10} w_{50}/\mathrm{km\,s}^{-1}$	count.	all stat.	count.	all stat.	count.	all stat.
1.2	$-0.23^{+0.19}_{-0.35}$	$-0.51^{+0.33}_{-0.53}$	$-0.30^{+0.18}_{-0.33}$	$-0.55^{+0.44}_{-0.58}$	$-0.44^{+0.28}_{-0.98}$	$-0.86^{+0.62}_{-0.95}$
1.4	$-0.19^{+0.15}_{-0.23}$	$-0.27^{+0.27}_{-0.35}$	$0.03^{+0.17}_{-0.28}$	$-0.13^{+0.32}_{-0.42}$	$-0.72^{+0.15}_{-0.22}$	$-0.80^{+0.52}_{-0.45}$
1.5	$-0.29^{+0.22}_{-0.46}$	$-0.46^{+0.30}_{-0.27}$	$0.04^{+0.27}_{-0.85}$	$-0.36^{+0.47}_{-0.35}$	$-0.72^{+0.14}_{-0.20}$	$-0.79^{+0.24}_{-0.23}$
1.6	$-0.73^{+0.06}_{-0.07}$	$-0.67^{+0.24}_{-0.14}$	$-0.63^{+0.07}_{-0.09}$	$-0.59^{+0.30}_{-0.17}$	$-0.91^{+0.09}_{-0.11}$	$-0.90^{+0.23}_{-0.19}$
1.7	$-0.91^{+0.05}_{-0.05}$	$-0.78^{+0.20}_{-0.12}$	$-0.79^{+0.05}_{-0.06}$	$-0.67^{+0.26}_{-0.14}$	$-1.14^{+0.09}_{-0.11}$	$-1.07^{+0.18}_{-0.16}$
1.8	$-0.99^{+0.04}_{-0.04}$	$-0.88^{+0.22}_{-0.12}$	$-0.91^{+0.05}_{-0.05}$	$-0.78^{+0.26}_{-0.14}$	$-1.11^{+0.07}_{-0.08}$	$-1.13^{+0.12}_{-0.12}$
1.9	$-1.00^{+0.05}_{-0.05}$	$-0.90^{+0.19}_{-0.11}$	$-0.89^{+0.06}_{-0.07}$	$-0.79^{+0.24}_{-0.14}$	$-1.22^{+0.06}_{-0.07}$	$-1.17^{+0.10}_{-0.10}$
2.0	$-0.93^{+0.04}_{-0.05}$	$-0.85^{+0.20}_{-0.11}$	$-0.80^{+0.05}_{-0.06}$	$-0.73^{+0.23}_{-0.12}$	$-1.26^{+0.05}_{-0.05}$	$-1.22^{+0.10}_{-0.08}$
2.1	$-0.99^{+0.04}_{-0.05}$	$-0.97^{+0.17}_{-0.09}$	$-0.88^{+0.05}_{-0.06}$	$-0.85^{+0.23}_{-0.11}$	$-1.25^{+0.06}_{-0.07}$	$-1.26^{+0.10}_{-0.08}$
2.2	$-1.15^{+0.04}_{-0.04}$	$-1.09^{+0.17}_{-0.08}$	$-1.03^{+0.04}_{-0.05}$	$-0.98^{+0.20}_{-0.09}$	$-1.45^{+0.03}_{-0.04}$	$-1.39^{+0.09}_{-0.06}$
2.3	$-1.29^{+0.03}_{-0.03}$	$-1.27^{+0.07}_{-0.04}$	$-1.20^{+0.03}_{-0.04}$	$-1.18^{+0.08}_{-0.05}$	$-1.47^{+0.04}_{-0.04}$	$-1.45^{+0.06}_{-0.05}$
2.4	$-1.55^{+0.03}_{-0.03}$	$-1.51^{+0.05}_{-0.04}$	$-1.47^{+0.03}_{-0.03}$	$-1.44^{+0.06}_{-0.05}$	$-1.71^{+0.03}_{-0.03}$	$-1.67^{+0.05}_{-0.04}$
2.5	$-1.64^{+0.04}_{-0.04}$	$-1.64^{+0.06}_{-0.05}$	$-1.55^{+0.05}_{-0.05}$	$-1.57^{+0.07}_{-0.07}$	$-1.89^{+0.03}_{-0.03}$	$-1.83^{+0.05}_{-0.05}$
2.6	$-1.96^{+0.04}_{-0.04}$	$-1.91^{+0.07}_{-0.06}$	$-1.88^{+0.05}_{-0.05}$	$-1.84^{+0.08}_{-0.07}$	$-2.10^{+0.06}_{-0.07}$	$-2.07^{+0.07}_{-0.07}$
2.7	$-2.12^{+0.08}_{-0.10}$	$-2.09^{+0.18}_{-0.12}$	$-2.00^{+0.09}_{-0.11}$	$-1.98^{+0.23}_{-0.15}$	$-2.50^{+0.06}_{-0.07}$	$-2.41^{+0.10}_{-0.08}$
2.8	$-2.96^{+0.06}_{-0.07}$	$-2.84^{+0.11}_{-0.09}$	$-2.88^{+0.08}_{-0.10}$	$-2.75^{+0.12}_{-0.11}$	$-3.08^{+0.06}_{-0.07}$	$-3.01^{+0.12}_{-0.10}$
2.9	$-4.20^{+0.12}_{-0.18}$	$-3.90^{+0.29}_{-0.26}$	$-4.06^{+0.16}_{-0.25}$	$-3.85^{+0.31}_{-0.29}$	$-4.44^{+0.16}_{-0.25}$	$-4.16^{+0.53}_{-0.36}$

Table C2. As Table C1, but for the H I MF.

	$\log_{10}\phi(M_{\rm HI})/h_{70}^3$ Mpc^{-3} dex^{-1}					
	α.100		α.100 Spring		α.100 Fall	
$\log_{10}M_{\rm HI}/{\rm M}_\odot$	count.	all stat.	count.	all stat.	count.	all stat.
6.1	$-1.00^{+0.23}_{-0.54}$	$-1.14^{+0.35}_{-0.67}$	$-0.74^{+0.23}_{-0.54}$	$-0.92^{+0.36}_{-0.72}$	–	$-1.09^{+0.33}_{-1.48}$
6.3	$-1.56^{+0.30}_{-15.90}$	$-1.28^{+0.36}_{-0.62}$	–	$-1.12^{+0.38}_{-0.69}$	–	$-1.28^{+0.36}_{-1.33}$
6.5	$-1.21^{+0.18}_{-0.30}$	$-1.20^{+0.26}_{-0.40}$	$-1.07^{+0.20}_{-0.38}$	$-1.13^{+0.32}_{-0.54}$	–	$-1.40^{+0.36}_{-0.70}$
6.7	$-1.22^{+0.15}_{-0.23}$	$-1.19^{+0.20}_{-0.26}$	$-1.31^{+0.18}_{-0.32}$	$-1.10^{+0.24}_{-0.34}$	$-1.20^{+0.20}_{-0.40}$	$-1.39^{+0.29}_{-0.52}$
6.9	$-1.35^{+0.11}_{-0.15}$	$-1.14^{+0.13}_{-0.16}$	$-1.50^{+0.14}_{-0.20}$	$-1.02^{+0.16}_{-0.20}$	$-1.25^{+0.15}_{-0.23}$	$-1.41^{+0.22}_{-0.32}$
7.1	$-1.11^{+0.06}_{-0.07}$	$-1.11^{+0.10}_{-0.11}$	$-0.98^{+0.07}_{-0.08}$	$-0.98^{+0.11}_{-0.13}$	$-1.43^{+0.13}_{-0.19}$	$-1.47^{+0.19}_{-0.26}$
7.3	$-1.10^{+0.05}_{-0.06}$	$-1.11^{+0.07}_{-0.08}$	$-0.94^{+0.06}_{-0.07}$	$-0.97^{+0.08}_{-0.09}$	$-1.53^{+0.12}_{-0.16}$	$-1.47^{+0.16}_{-0.21}$
7.5	$-1.14^{+0.04}_{-0.05}$	$-1.19^{+0.06}_{-0.06}$	$-1.01^{+0.05}_{-0.05}$	$-1.09^{+0.07}_{-0.07}$	$-1.44^{+0.10}_{-0.12}$	$-1.44^{+0.13}_{-0.16}$
7.7	$-1.25^{+0.04}_{-0.04}$	$-1.28^{+0.05}_{-0.05}$	$-1.18^{+0.04}_{-0.04}$	$-1.21^{+0.06}_{-0.06}$	$-1.38^{+0.07}_{-0.09}$	$-1.40^{+0.10}_{-0.12}$
7.9	$-1.33^{+0.03}_{-0.03}$	$-1.35^{+0.04}_{-0.04}$	$-1.26^{+0.04}_{-0.04}$	$-1.28^{+0.05}_{-0.05}$	$-1.45^{+0.07}_{-0.08}$	$-1.49^{+0.09}_{-0.10}$
8.1	$-1.39^{+0.03}_{-0.03}$	$-1.41^{+0.04}_{-0.04}$	$-1.31^{+0.03}_{-0.03}$	$-1.33^{+0.04}_{-0.04}$	$-1.56^{+0.06}_{-0.07}$	$-1.62^{+0.08}_{-0.09}$
8.3	$-1.43^{+0.02}_{-0.03}$	$-1.44^{+0.03}_{-0.03}$	$-1.32^{+0.03}_{-0.03}$	$-1.35^{+0.03}_{-0.04}$	$-1.69^{+0.05}_{-0.06}$	$-1.65^{+0.06}_{-0.07}$
8.5	$-1.49^{+0.02}_{-0.02}$	$-1.51^{+0.03}_{-0.03}$	$-1.39^{+0.03}_{-0.03}$	$-1.42^{+0.03}_{-0.03}$	$-1.67^{+0.04}_{-0.04}$	$-1.69^{+0.05}_{-0.06}$
8.7	$-1.55^{+0.02}_{-0.02}$	$-1.57^{+0.02}_{-0.02}$	$-1.47^{+0.02}_{-0.02}$	$-1.50^{+0.03}_{-0.03}$	$-1.68^{+0.03}_{-0.03}$	$-1.68^{+0.04}_{-0.04}$
8.9	$-1.64^{+0.02}_{-0.02}$	$-1.63^{+0.02}_{-0.02}$	$-1.56^{+0.02}_{-0.02}$	$-1.57^{+0.02}_{-0.03}$	$-1.75^{+0.02}_{-0.03}$	$-1.72^{+0.03}_{-0.03}$
9.1	$-1.68^{+0.01}_{-0.01}$	$-1.68^{+0.01}_{-0.01}$	$-1.61^{+0.02}_{-0.02}$	$-1.63^{+0.02}_{-0.02}$	$-1.76^{+0.02}_{-0.02}$	$-1.76^{+0.02}_{-0.02}$
9.3	$-1.79^{+0.01}_{-0.01}$	$-1.79^{+0.01}_{-0.01}$	$-1.73^{+0.01}_{-0.01}$	$-1.74^{+0.02}_{-0.02}$	$-1.88^{+0.01}_{-0.01}$	$-1.88^{+0.02}_{-0.02}$
9.5	$-1.95^{+0.01}_{-0.01}$	$-1.96^{+0.01}_{-0.01}$	$-1.90^{+0.01}_{-0.01}$	$-1.91^{+0.01}_{-0.01}$	$-2.03^{+0.01}_{-0.01}$	$-2.03^{+0.02}_{-0.02}$
9.7	$-2.11^{+0.01}_{-0.01}$	$-2.12^{+0.01}_{-0.01}$	$-2.08^{+0.01}_{-0.01}$	$-2.08^{+0.01}_{-0.01}$	$-2.17^{+0.01}_{-0.01}$	$-2.18^{+0.01}_{-0.02}$
9.9	$-2.32^{+0.01}_{-0.01}$	$-2.32^{+0.01}_{-0.01}$	$-2.28^{+0.01}_{-0.01}$	$-2.28^{+0.01}_{-0.01}$	$-2.39^{+0.01}_{-0.01}$	$-2.39^{+0.01}_{-0.01}$
10.1	$-2.62^{+0.01}_{-0.01}$	$-2.61^{+0.01}_{-0.01}$	$-2.58^{+0.01}_{-0.01}$	$-2.57^{+0.01}_{-0.01}$	$-2.67^{+0.01}_{-0.01}$	$-2.66^{+0.02}_{-0.02}$
10.3	$-3.07^{+0.01}_{-0.01}$	$-3.05^{+0.01}_{-0.01}$	$-3.06^{+0.02}_{-0.02}$	$-3.03^{+0.02}_{-0.02}$	$-3.09^{+0.02}_{-0.02}$	$-3.07^{+0.02}_{-0.02}$
10.5	$-3.77^{+0.03}_{-0.03}$	$-3.73^{+0.03}_{-0.03}$	$-3.74^{+0.03}_{-0.04}$	$-3.70^{+0.04}_{-0.04}$	$-3.83^{+0.05}_{-0.05}$	$-3.78^{+0.05}_{-0.06}$
10.7	$-4.64^{+0.07}_{-0.09}$	$-4.61^{+0.08}_{-0.09}$	$-4.63^{+0.09}_{-0.11}$	$-4.59^{+0.09}_{-0.11}$	$-4.65^{+0.11}_{-0.16}$	$-4.66^{+0.13}_{-0.17}$
10.9	$-5.81^{+0.23}_{-0.53}$	$-5.74^{+0.23}_{-0.47}$	$-5.61^{+0.23}_{-0.53}$	$-5.59^{+0.23}_{-0.52}$	–	$-5.69^{+0.30}_{-1.99}$

APPENDIX D: MARGINALIZED POSTERIOR PROBABILITY DISTRIBUTIONS OF SCHECHTER FUNCTION PARAMETERS

We show the marginalized posterior probability distribution for the Schechter function parameters describing the H I MF in Fig. D1, analogously to those for the H I WF shown in Fig. 2. The same approach as for the H I WF is used to estimate the confidence regions including all statistical uncertainties – see Section 4 for details.

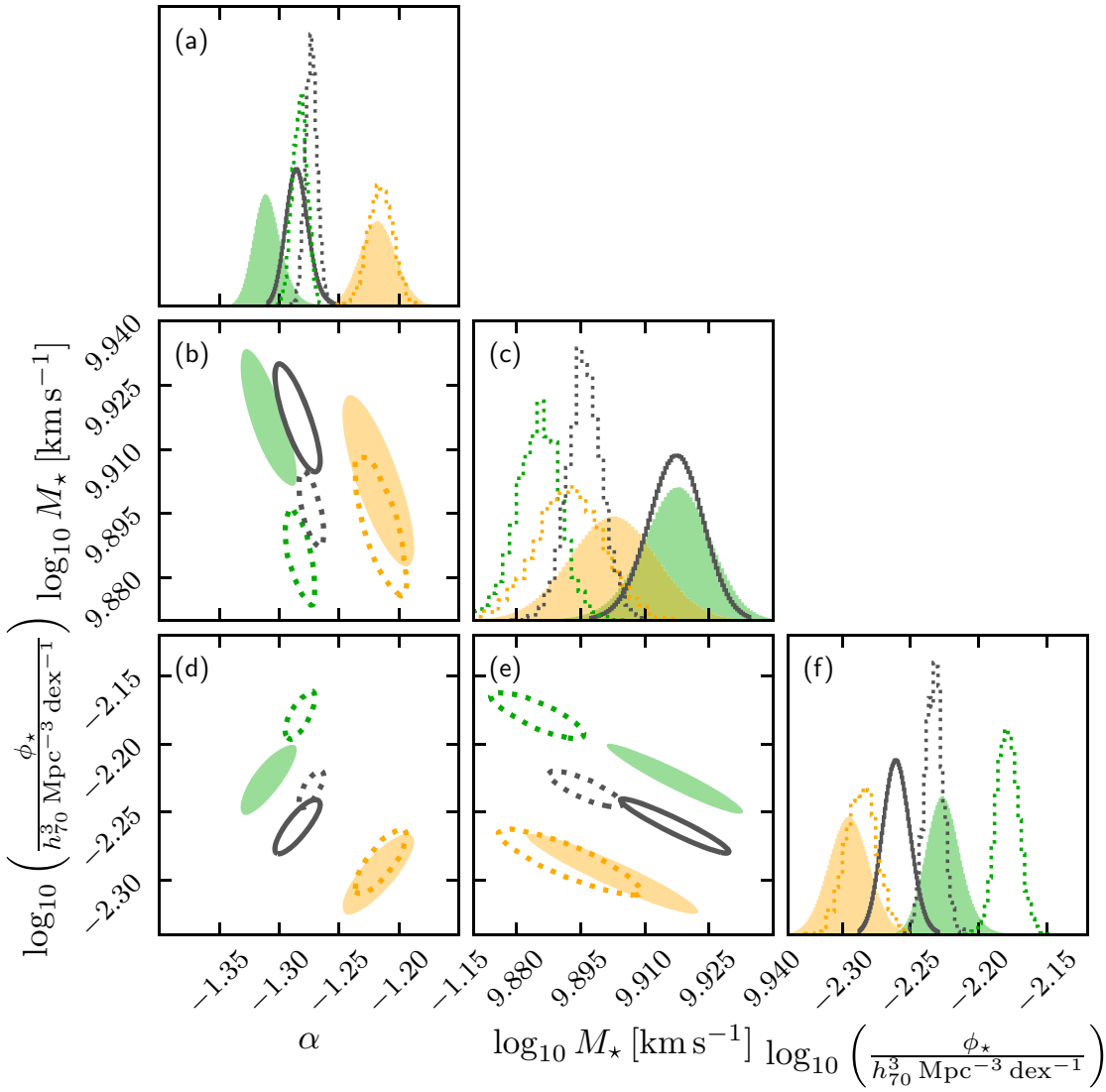

Figure D1. 1D and 2D marginalized posterior probability distributions for the parameters of the Schechter function (equation 2). All contours are drawn at the 95 per cent confidence level. Dark grey dotted contours and histograms correspond to the fit to the H I MF of the full $\alpha.100$ sample when only counting (Poisson) uncertainties are accounted for. Dark grey solid contours and histograms are similar, but include all statistical uncertainties. Dark green and orange contours and dotted histograms show the parameter constraints for the spring and fall sky areas, respectively, accounting only for counting uncertainties, whilst the light green and orange contours and filled histograms show the constraints when all statistical uncertainties are accounted for. The difference in the shape parameters (α, $M_\star$) between the spring and fall regions noted by J18 is clearly visible in panel (b).

APPENDIX E: MOCK SURVEYS

We wish to construct a mock catalogue of 'observed' sources drawn from a known underlying 'intrinsic' source population. The distribution of H I masses of the intrinsic population should be described by a Schechter function, and the velocity width distribution by a modified Schechter function. It is trivial to draw a set of discrete values from the two desired distributions to achieve this, but once this is done satisfying the additional desired constraint that the observed source population resemble that in the ALFALFA survey is not trivial, and we could not arrive at a satisfactory formulation of the problem that is mathematically well-posed and not under-constrained. We therefore opted for a different approach. Since we are most interested in the H I WF, we sampled a set of w_{50} values from the desired distribution. We then assumed a Gumbel distribution (also known as a Type I Extreme Value distribution) form for the distribution of H I masses at fixed w_{50} and that the two parameters describing the distribution, a width β and centroid μ, vary continuously with w_{50}. We iteratively adjusted the scalings of μ and β brute-force 'by hand' until (i) the amplitude of the resulting H I MF matched that desired within 10 per cent in each bin (the bins used are the same as those used in the $1/V_{\rm eff}$ estimator), and (ii) the observed source distribution in the $M_{\rm HI}$–w_{50} plane was an acceptable match to the ALFALFA counts. Criterion (ii) was assessed subjectively, 'by eye' – in combination with constraint (i) requiring a quantitative match proved intractable.

Once an H I mass was assigned to each line width, the sources were distributed random-uniformly in a survey volume and their 21-cm fluxes computed from their distances and H I masses. All sources in this intrinsic source list below the ALFALFA 50 per cent

completeness limit were removed to yield the 'observed' source catalogue.

The Gumbel distribution parameters used to construct the mocks are as follows, where $M = \log_{10}(M_{\rm HI}/{\rm M}_\odot)$ and $W = \log_{10}(w_{50}/{\rm km\,s^{-1}})$. For mock (a):

$$\mu(W) = \begin{cases} 5.47W - 2.22 & \text{if } 1.20 \leq W < 1.76 \\ 3.11W + 1.92 & \text{if } 1.76 \leq W < 2.20 \\ 2.38W + 3.55 & \text{if } 2.20 \leq W < 3.00 \end{cases} \quad \text{(E1)}$$

$$\beta(W) = \begin{cases} -1.00W + 2.90 & \text{if } 1.20 \leq W < 1.80 \\ -0.81W + 2.56 & \text{if } 1.80 \leq W < 2.60 \\ -0.68W + 2.21 & \text{if } 2.60 \leq W < 3.00 \end{cases}, \quad \text{(E2)}$$

for mock (b):

$$\mu(W) = \begin{cases} 6.51W - 3.86 & \text{if } 1.20 \leq W < 1.76 \\ 3.20W + 1.96 & \text{if } 1.76 \leq W < 2.20 \\ 2.21W + 4.14 & \text{if } 2.20 \leq W < 3.00 \end{cases} \quad \text{(E3)}$$

$$\beta(W) = \begin{cases} -0.34W + 2.21 & \text{if } 1.20 \leq W < 1.40 \\ -1.94W + 4.45 & \text{if } 1.40 \leq W < 1.76 \\ -1.00W + 2.79 & \text{if } 1.76 \leq W < 2.20 \\ -0.53W + 1.75 & \text{if } 2.20 \leq W < 3.00 \end{cases}, \quad \text{(E4)}$$

for mock (c):

$$\mu(W) = \begin{cases} 6.05W - 2.44 & \text{if } 1.20 \leq W < 1.76 \\ 2.95W + 3.00 & \text{if } 1.76 \leq W < 2.20 \\ 1.64W + 5.90 & \text{if } 2.20 \leq W < 3.00 \end{cases} \quad \text{(E5)}$$

$$\beta(W) = \begin{cases} 0.4W + 0.94 & \text{if } 1.20 \leq W < 1.40 \\ -2.08W + 4.42 & \text{if } 1.40 \leq W < 1.76 \\ -0.80W + 2.15 & \text{if } 1.76 \leq W < 2.20 \\ -0.31W + 1.09 & \text{if } 2.20 \leq W < 3.00 \end{cases}, \quad \text{(E6)}$$

and the cumulative distribution function of the Gumbel distribution is:

$$F(M|W) = \frac{e^{-e^{-z(M,W)}} - e^{-e^{-z_{\rm min}}}}{e^{-e^{-z_{\rm max}}} - e^{-e^{-z_{\rm min}}}} \quad \text{(E7)}$$

with

$$z(M, W) = \frac{\mu(W) - M}{\beta(W)}, \quad \text{(E8)}$$

$$z_{\rm min} = z(M_{\rm min}, W), \quad \text{(E9)}$$

$$z_{\rm max} = z(M_{\rm max}, W) \quad \text{(E10)}$$

where $M_{\rm min}$ and $M_{\rm max}$ correspond to the ($\log_{10}$ of the) minimum and maximum H I masses to be sampled, in our case 6.0 and 11.0, respectively. To determine a mass for a mock source of velocity width W, one therefore simply draws a random value from the distribution given by equation (E7).

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