The ALFALFA H I mass function: a dichotomy in the low-mass slope and a locally suppressed 'knee' mass

Michael G. Jones,[1,2]★ Martha P. Haynes,[1] Riccardo Giovanelli[1] and Crystal Moorman[3]

[1]*Cornell Center for Astrophysics and Planetary Science, Space Sciences Building, Cornell University, Ithaca, NY 14853, USA*
[2]*Instituto de Astrofísica de Andalucía, Glorieta de la Astronomía s/n, E-18008 Granada, Spain*
[3]*Department of Physics, Lynchburg College, 1501 Lakeside Drive, Lynchburg, VA 24501, USA*

ABSTRACT

We present the most precise measurement of the $z = 0$ H I mass function (HIMF) to date based on the final catalogue of the ALFALFA (Arecibo Legacy Fast ALFA) blind H I survey of the nearby Universe. The Schechter function fit has a 'knee' mass $\log(M_* h_{70}^2/\mathrm{M}_\odot) = 9.94 \pm 0.01 \pm 0.05$, a low-mass slope parameter $\alpha = -1.25 \pm 0.02 \pm 0.1$, and a normalization $\phi_* = (4.5 \pm 0.2 \pm 0.8) \times 10^{-3} \, h_{70}^3 \, \mathrm{Mpc}^{-3} \, \mathrm{dex}^{-1}$, with both random and systematic uncertainties as quoted. Together these give an estimate of the H I content of the $z = 0$ Universe as $\Omega_{\mathrm{H\,I}} = (3.9 \pm 0.1 \pm 0.6) \times 10^{-4} \, h_{70}^{-1}$ (corrected for H I self-absorption). Our analysis of the uncertainties indicates that the 'knee' mass is a cosmologically fair measurement of the $z = 0$ value, with its largest uncertainty originating from the absolute flux calibration, but that the low-mass slope is only representative of the local Universe. We also explore large-scale trends in α and M_* across the ALFALFA volume. Unlike with the 40 per cent sample, there is now sufficient coverage in both of the survey fields to make an independent determination of the HIMF in each. We find a large discrepancy in the low-mass slope ($\Delta\alpha = 0.14 \pm 0.03$) between the two regions, and argue that this is likely caused by the presence of a deep void in one field and the Virgo cluster in the other. Furthermore, we find that the value of the 'knee' mass within the Local Volume appears to be suppressed by 0.18 ± 0.04 dex compared to the global ALFALFA value, which explains the lower value measured by the shallower H I Parkes All Sky Survey (HIPASS). We discuss possible explanations and interpretations of these results and how they can be expanded on with future surveys.

Key words: surveys – galaxies: luminosity function, mass function – radio lines: galaxies.

1 INTRODUCTION

The H I mass function, or HIMF, is the intrinsic distribution of galaxy H I masses in the Universe. Numerous H I surveys (Rosenberg & Schneider 2002; Springob, Haynes & Giovanelli 2005; Zwaan et al. 2005; Martin et al. 2010; Hoppmann et al. 2015) have found this function to be well fit by a Schechter function (Schechter 1976), a power law increasing towards lower masses and an exponential decline at high masses. The two shape parameters of this function are the power-law exponent, usually referred to as the low-mass slope ($\alpha + 1$) on a logarithmic scale, and the 'knee' mass (M_*) where the exponential decline begins.

Current cosmological simulations now include gas and baryonic processes (e.g. Vogelsberger et al. 2014; Crain et al. 2015) that alter the galaxy populations they produce. Many of the baryonic processes included in these models are functions of galaxy environment; for example, they may depend on tidal stripping, the host halo mass, or the background UV radiation field. As H I typically has both a highly extended spatial distribution and is neutral, it is one of the most sensitive baryonic components to galaxy interactions and hard radiation fields. Therefore, the HIMFs, and any variations with environment, are key constraints for the galaxy populations resulting from such simulations.

H I Parkes All Sky Survey (HIPASS; Barnes et al. 2001; Meyer et al. 2004), which covered the entire Southern sky below declination $+2°$ (21 341 deg^2)[1], made one of the first robust measurements

---

[1] HIPASS' Northern extension (Wong et al. 2006) goes up to a Dec. of $+25.5°$; however, this was not included in the HIPASS HIMF calculation and so is not considered in this paper.

★ E-mail: mjones@iaa.es

of the $z = 0$ HIMF (Zwaan et al. 2005), calculating a 'knee' mass of $10^{9.86 \pm 0.03}$ M$_\odot$ and a low-mass slope of -0.37 ± 0.03 (or $\alpha = -1.37 \pm 0.03$). The ALFALFA (Arecibo Legacy Fast ALFA[2]) survey (Giovanelli et al. 2005) observed a smaller portion of the sky ($\sim 6900 \deg^2$) than HIPASS, but with substantially improved sensitivity, angular and velocity resolution, and depth. As Arecibo is located at a latitude of $18°$ N and can operate at a maximum zenith angle of about $20°$, the two surveys have minimal overlap. The ALFALFA 40 per cent ($\alpha.40$) HIMF was calculated by Martin et al. (2010), who found a marginally flatter slope ($\alpha = -1.33 \pm 0.02$) and higher 'knee' mass ($10^{9.96 \pm 0.02}$ M$_\odot$) than found in HIPASS.

Due to its much larger sky coverage the total survey volume of HIPASS is greater than that of ALFALFA; however, ALFALFA's greater sensitivity actually results in a larger volume for the relevant portions of the HIMF. For example, a galaxy of H I mass $\log M_{H I}/M_\odot = 9$ and a velocity width of $100 \, \mathrm{km \, s^{-1}}$ would be detectable by ALFALFA out to a distance of approximately 80 Mpc (based on the completeness limit defined in Haynes et al. 2011), whereas the same galaxy would only be detectable out to about 35 Mpc in HIPASS (at an equivalent completeness level, Zwaan et al. 2004). This means that the maximum volume over which ALFALFA can probe the low-mass slope is about four times larger than what was probed by HIPASS. Similarly, at the HIMF 'knee' mass, ALFALFA can probe out to almost 200 Mpc, whereas the equivalent distance for HIPASS is only 75 Mpc. Again, this means that the volume available in ALFALFA to study the 'knee' mass is over six times greater than for HIPASS.[3]

Zwaan et al. (2005) also looked for environmental dependence of the HIMF, finding that the low-mass slope steepened slightly in high-density environments, and that the 'knee' mass was unaffected. However, as there was no optical galaxy survey covering the HIPASS sky that could be used to define environment, Zwaan et al. (2005) defined environment based on the density of HIPASS sources themselves; therefore, these results are non-trivial to compare with (see Jones et al. 2016, for a more detailed discussion).

Using $\alpha.40$, Moorman et al. (2014) assessed whether the shape of the HIMF was different in large cosmic voids compared to the walls that separate them. They found a small, but significant, decrease in the 'knee' mass of void galaxies and a marginal flattening of the low-mass slope. Jones et al. (2016) used the neighbour density in optical and infrared surveys overlapping with the ALFALFA 70 per cent catalogue ($\alpha.70$) to look for trends based on a galaxy's local environment, finding an increase in the 'knee' mass (similar to Moorman et al. 2014) in higher density environments, but no significant change in the low-mass slope.

With the final ALFALFA 100 per cent catalogue ($\alpha.100$), it is now possible to compare the H I properties of galaxies in two continuous regions spanning thousands of square degrees in the disparate environments covered by the Arecibo Spring and Fall skies. Thus, in this paper we take a different approach to the studies above and focus on the largest scale shifts in environment that we can, namely, the Spring and Fall sky regions of ALFALFA, inside and outside the Local Volume, and finally, proximity to the Virgo region. In addition, we present the most precise blind measurement of the global $z = 0$ HIMF to date along with a thorough analysis of its uncertainties.

The paper is laid out as follows. Section 2 briefly describes the ALFALFA samples used throughout the paper, Section 3 outlines how the HIMF is estimated, and the resulting HIMF and its uncertainty estimates are presented in Section 4. Section 5 covers how the HIMF varies across the different sub-samples which we consider, and possible interpretations of these results are discussed in Section 6. Finally, we draw our conclusions in Section 7. H_0 is assumed to be $70 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$ throughout this paper.

2 THE ALFALFA SAMPLE

The final ALFALFA footprint is approximately $6900 \deg^2$, or about a third the size of the HIPASS footprint. This area is split between two continuous regions in the directions of Virgo ($\sim 7.5 \, \mathrm{h}$ RA to $\sim 16.5 \, \mathrm{h}$ RA) and Pisces ($\sim 22 \, \mathrm{h}$ RA to $\sim 3 \, \mathrm{h}$ RA), with a declination range or approximately $0°$ to $36°$ in both cases. We will refer to these regions as the Arecibo 'Spring sky' and 'Fall sky', respectively, after the seasons in which they were observed. This area was observed blindly using a drift scan strategy that resulted in over 95 per cent time efficiency, including all overheads. Both the data flagging and source extraction were carried out in part by automated processes (Saintonge 2007), but were ultimately completed by a person, who in the case of extragalactic sources would also identify the most likely optical counterpart from available overlapping surveys.

The typical rms noise level in ALFALFA is $2.4 \, \mathrm{mJy}$ per beam (the ALFA beam is 3.3 by 3.8 arcmin at 21 cm) at $10 \, \mathrm{km \, s^{-1}}$ resolution, compared to $13 \, \mathrm{mJy}$ per beam (the Parkes beam diameter is 15.5 arcmin at 21 cm) at $26 \, \mathrm{km \, s^{-1}}$ resolution for HIPASS. The respective redshift ranges of HIPASS and ALFAFLA are $-1280 \, \mathrm{km \, s^{-1}} < cz < 12\,700 \, \mathrm{km \, s^{-1}}$ and $-1600 \, \mathrm{km \, s^{-1}} < cz < 18\,000 \, \mathrm{km \, s^{-1}}$.

The ALFALFA 40 per cent catalogue (Haynes et al. 2011), or $\alpha.40$, contained 40 per cent of the final footprint area, mostly in the Spring sky. The 100 per cent catalogue ($\alpha.100$) has now added significantly more area to the Fall sky as well as expanding the coverage in the Spring sky. These catalogues are available online at http://egg.astro.cornell.edu/alfalfa/data/index.php A summary of the $\alpha.100$ catalogue will be given in an upcoming paper (Haynes et al., in preparation). However, the catalogue description of $\alpha.40$ (Haynes et al. 2011) can be applied to the H I properties listed in $\alpha.100$, as we have done in this paper.

Estimating H I masses for ALFALFA sources requires first an estimate of their line-of-sight distance. To make this estimate ALFALFA uses the combination of a Local Volume flow model (Masters 2005), group assignments based on the 2MASS Redshift Survey (2MRS; Crook et al. 2007; Huchra et al. 2012) and the catalogue of Nearby Optical Galaxies (NOG; Giuricin et al. 2000; Springob 2006), as well as primary and secondary distances available in the literature. In addition, sources in the Virgo region are matched to the Virgo Cluster Catalog (VCC; Binggeli, Sandage & Tammann 1985) and assigned to the relevant clouds in the Virgo cluster. Further details of the distance estimation process are given in Section 4. With these distances the H I mass of the ALFALFA sources are then calculated using the standard equation:

$$\frac{M_{H I}}{M_\odot} = 2.356 \times 10^5 \, D_{Mpc}^2 \, S_{21}, \qquad (1)$$

where D_{Mpc} is the distance to the galaxy in Mpc and S_{21} is its integrated flux in Jy km s^{-1}.

---

[2] ALFA (Arecibo *L*-band Feed Array) is the name of the seven beam feed horn array instrument with which the survey was performed.

[3] For more details on these calculations, please see Appendix A.

3 CALCULATING HIMFs

The HIMF represents the intrinsic number density of galaxies in the Universe as a function of their H I mass. This number density is usually denoted as

$$\phi(M_{\mathrm{H\,I}}) = \frac{\mathrm{d}N_{\mathrm{gal}}}{\mathrm{d}V\,\mathrm{d}\log_{10}(M_{\mathrm{H\,I}})}, \tag{2}$$

where $\mathrm{d}N_{\mathrm{gal}}$ is the average number of galaxies in the volume $\mathrm{d}V$, with H I masses that fall within a small logarithmic bin centred on $M_{\mathrm{H\,I}}$.

Converting the observed number counts (lower panel of Fig. 2) as a function of H I mass to the intrinsic number (Fig. 2, upper panel) is a non-trivial process because ALFALFA is not a volume-limited sample and so the survey sensitivity must be corrected for. Furthermore, the sensitivity limit of an H I survey depends both on the source flux and velocity width. ALFALFA's sensitivity limits are explained in detail in Haynes et al. (2011) and we use the 50 per cent completeness surface for Code 1 sources (signal to noise > 6.5) as our sensitivity threshold throughout this paper.

Large-scale structure (LSS) along the line of sight causes certain regions in redshift to be either sparsely or densely populated (relative to what would be expected in a uniform universe). Unless corrected for along with the survey sensitivity this can impact the form of the calculated HIMF. To account for these effects, we use the two-dimensional stepwise maximum likelihood (2DSWML) estimator (Efstathiou, Ellis & Peterson 1988; Zwaan et al. 2003, 2005; Martin et al. 2010), also known as the $1/V_{\mathrm{eff}}$ method, which has been shown to be robust against the effects of LSS. For a detailed description of the application of this estimator to ALFALFA, refer to Martin et al. (2010) or Papastergis (2013) Appendix A.

By assuming that the form of the HIMF is universal throughout the volume considered the distribution of sources along the line of sight drops out of the maximum likelihood process (and with it the HIMF normalization), preventing adverse effects from LSS, but meaning that the normalization constant must be estimated by other means.

It has been demonstrated numerous times (e.g. Zwaan et al. 2005; Martin et al. 2010) that the HIMF is well fit by a Schechter function

$$\phi(M_{\mathrm{H\,I}}) = \ln(10)\,\phi_* \left(\frac{M_{\mathrm{H\,I}}}{M_*}\right)^{\alpha+1} \mathrm{e}^{-\left(\frac{M_{\mathrm{H\,I}}}{M_*}\right)}, \tag{3}$$

where the fit parameters are ϕ_*, the normalization constant, M_*, the 'knee' mass (we will often use $m_* = \log M_*/\mathrm{M}_\odot$), and the low-mass slope gradient ($\alpha + 1$). These parameters are used throughout this paper to describe the fits to the various HIMFs that we calculate.

An accurate definition is needed of the area on the sky that the survey is complete over. ALFALFA was conducted as a drift scan survey with the ALFA (Arecibo L-band Feed Array) instrument. Each observing night the telescope would be parked at a Dec. between $0°$ and $+36°$ and take data as the sky drifted by between RA 08:00 and 16:30 when observing the Spring part of the sky and between RA 22:00 and 03:00 when observing the Fall part. Each Dec. strip was observed twice in this manner over the course of the survey, with the second observation offset by half a beam width. The vagaries of telescope time allocation are responsible for irregularities near the east and west edges of the sky box, which translated into occasionally poor or incomplete data coverage. To define the edges of the survey area, we inspected the coverage maps by eye, and chose the RA and Dec. extremes such that (as near as possible) the entire area has at least single pass coverage.

The full $\alpha.100$ catalogue contains 25437 high signal-to-noise extragalactic sources. With the sky area trimmed to 6501 deg^2 (Fig. 1) the catalogue contains 24 340 sources. The uncertainty in the determinations of this sky boundary and the impact that this may have on our results is assessed in Section 4. A final boundary is enforced in redshift space at a velocity of 15 000 km s^{-1} (relative to the Cosmic Microwave Background, CMB), beyond which ALFALFA suffers major incompleteness due to radio frequency interference. This further reduces the galaxy count to 23 621. All the sub-samples discussed in the remainder of the paper were drawn from this sample of 23 621 sources.[4]

4 THE ALFALFA HIMF

To calculate the global $\alpha.100$ HIMF, we take the catalogue of sources described in the previous section and impose the following cuts: $6 < \log M_{\mathrm{H\,I}}/\mathrm{M}_\odot$, $1.2 < \log W_{50}/\mathrm{km\,s^{-1}}$, and the 50 per cent completeness limit for high signal-to-noise sources (Haynes et al. 2011, their equations 4 and 5). Note that we have intentionally avoided making a minimum distance cut to remove sources with high fractional distance uncertainties. The lowest mass bins do not appear particularly noisy so such a cut was deemed unnecessary, especially as it can also cause an artificial suppression of the lowest mass bins. As all ALFALFA sources are inspected by eye, the separation between high velocity clouds (HVCs) and low recession velocity dwarf galaxies is expected to be highly reliable. While there does remain some possibility of sources being overlooked due to the bright foreground emission of HVCs, this is expected to be, at most, a handful of sources. Any such sources would also likely fall below the minimum H I mass cut. Furthermore, these 'mistaken identity' candidates are being followed up as part of the Ultra-Compact High Velocity Cloud (UCHVC) project (e.g. Adams, Giovanelli & Haynes 2013; Adams et al. 2016).

These further cuts give a final sample of 22 831 galaxies from which the HIMF is computed using the $1/v_{\mathrm{eff}}$ method. A Schechter function is then fit in linear space to the non-parametric result (shown in Fig. 2) giving the parameters as $\alpha = -1.25 \pm 0.02$, $m_* = 9.94 \pm 0.01$, and $\phi_* = 4.5 \pm 0.2 \times 10^{-3}\,\mathrm{Mpc^{-3}\,dex^{-1}}$, where the fit errors quoted here are due only to the Poisson errors and do not include contributions from distance uncertainty or other random and systematic effects. The following subsections will discuss a number of other sources of error and how we estimate them. Our estimates are collated in Table 1.

4.1 Random flux and distance errors

In order to estimate the error introduced due the uncertainty in source distances, we took a Monte Carlo approach, creating many realizations of the HIMF calculation, altering the sources distances (and therefore masses) each time. There are seven methods that are used to calculate the distance to ALFALFA sources (the assignment occurs in the order listed, top given highest preference):

(i) Literature primary distances are used for all sources that are associated with an optical counterpart for which such a measurement exists.

(ii) Sources that are associated with counterparts in the VCC (Binggeli et al. 1985) are assigned the distance to the relevant Virgo cloud (as described in Hallenbeck et al. 2012).

---

[4] This does not include the samples with alternative distance estimates, for which the maximum velocity cut can change (Appendix B).
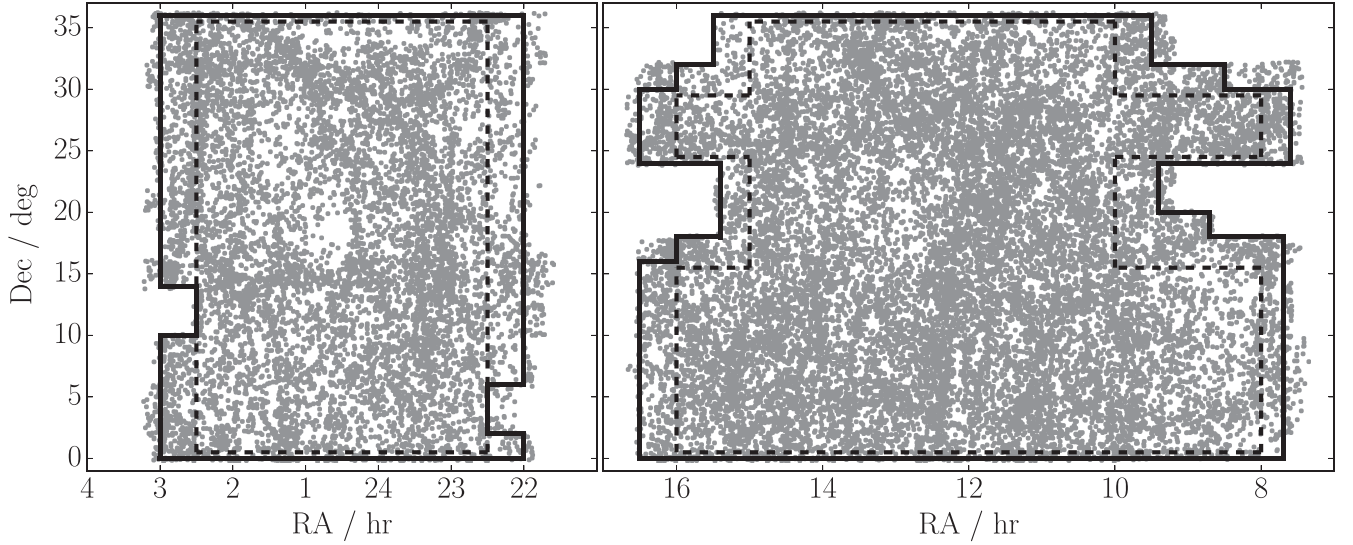
Figure 1. The grey points show the sky positions of the high signal-to-noise extragalactic α.100 sources with $v_{cmb} < 15\,000\,\mathrm{km\,s^{-1}}$. The left side is the ALFALFA Fall sky region, and the right side is the Spring sky region. The solid lines show the boundary that is used to calculate the HIMF, and the dashed lines show the strict boundary that is used for comparison purposes. The strict boundary moves in much further in RA (in order to be very conservative) than in Dec. because most of the variance in the boundary is determined by the time the observations started and stopped drift scanning, not by the top and bottom edges of the drift. The vertices of both boundaries are listed in Appendix D.
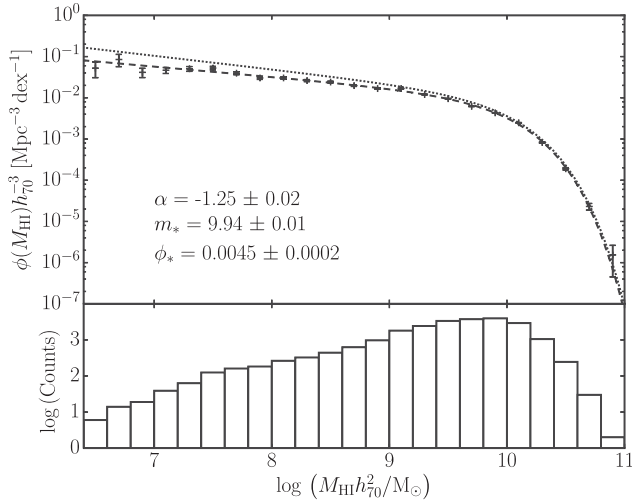

Figure 2. The ALFALFA 100 per cent HIMF. The lower panel shows the number counts in logarithmic H I mass bins, and the points in the upper panel show the calculated intrinsic abundances after correcting for LSS and survey sensitivity using the 2DSWML method. The error bars and the corresponding fit errors displayed here are from Poisson counting errors only. A more detailed error analysis can be found in the text. The dashed line shows the best fit to the data, and the corresponding Schechter function parameters are in the bottom left corner. The dotted line represents the ALFALFA 40 per cent best fit (Martin et al. 2010).

(iii) Literature secondary distances are used for sources with $v_{cmb} < 6000\,\mathrm{km\,s^{-1}}$ that are associated with optical counterparts for which such a measurement exists.

(iv) Galaxies assigned to a group are given the mean recession velocity of the group. For $v_{cmb} < 6000\,\mathrm{km\,s^{-1}}$ a flow model (Masters 2005) is used to estimate the distance.

(v) For group galaxies with $v_{cmb} > 6000\,\mathrm{km\,s^{-1}}$ pure Hubble flow (with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) is assumed.

(vi) The remaining sources with $v_{cmb} < 6000\,\mathrm{km\,s^{-1}}$ are assigned distances from the flow model.

(vii) The remaining sources with $v_{cmb} > 6000\,\mathrm{km\,s^{-1}}$ are assigned distances using Hubble flow.

To account for the uncertainty in the distances to ALFALFA sources, we followed a similar procedure to that of Papastergis, Adams & Romanowsky (2017). We ran 1000 iterations of the HIMF calculation, each time randomly adjusting the distance to each source as follows:

(i) Primary distances and assignments to Virgo are assumed to have a 10 per cent Gaussian uncertainty.

(ii) Secondary distances are drawn from a Gaussian distribution centred on the initial value and with a width of 20 per cent of that value.

(iii) Galaxies assigned to groups have a Gaussian uncertainty introduced due to the uncertainty in H_0, which we take to be $3\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, as well as the inclusion of a group peculiar velocity that is assigned to all members of the group (drawn from a Gaussian distribution centred on 0 and with a width of $300\,\mathrm{km\,s^{-1}}$).

(iv) The remaining sources are given a Gaussian error based on the quadrature sum of the uncertainty in H_0 and the typical value of galaxy peculiar velocities, which we take to be $160\,\mathrm{km\,s^{-1}}$.

In addition to the distance uncertainties, the impact of flux measurement uncertainty was introduced to each source by adding a random Gaussian error with a width of the quoted flux uncertainty in the α.100 catalogue. The output Schechter parameter values of the 1000 resulting iterations were then fit with normal distributions, giving standard deviations of $\sigma_\alpha = 0.007$, $\sigma_{m_*} = 0.004$, and $\sigma_{\phi_*} = 7 \times 10^{-5}\,\mathrm{Mpc^{-3}\,dex^{-1}}$. Combining these in quadrature with the uncertainty in an individual fit (from the Poisson counting errors) gives an estimate of the total random error in the parameter values: α.100 as $\alpha = -1.25 \pm 0.02$, $m_* = 9.94 \pm 0.01$, and $\phi_* = 4.5 \pm 0.2 \times 10^{-3}\,\mathrm{Mpc^{-3}\,dex^{-1}}$.

Table 1. The different contributions to both random and systematic errors in the calculated values of the Schechter function fit to the HIMF. Poisson errors correspond to the Schechter function fit uncertainty due to the finite number of sources contained in each bin of the non-parametric HIMF calculated by the 2DSWML method. The random flux and distance errors are from the uncertainty in measuring a source's flux from its spectrum and from determining its distance from its position and heliocentric velocity. The boundary error is a conservative estimate of the effect of small areas without complete coverage near the edges of the survey. This source of error is impossible to separate from environmental dependence of the HIMF and most of this contribution is likely due to the removal or inclusion of more area covering a given environment, rather than from the effects of non-uniform coverage itself. The cosmic variance term corresponds to our estimate of how much the HIMF may differ in a volume larger than ALFALFA's. Again this is a very conservative estimate taken as the difference between the disparate ALFALFA Spring and Fall skies. Sample variance (as we have called it) is really another form of cosmic variance, but on a scale smaller than the survey volume. This was estimated by jackknifing the sample across 46 approximately equal area, contiguous regions. The systematic distance error due to the choice of flow model was estimated by considering three different possible models. There is also another source of error from the flux estimates that corresponds to the absolute scale of the observations; this is estimated to be accurate to about 10 per cent. Finally, the 2DSWML or $1/v_{\mathrm{eff}}$ method has an intrinsic error associated with it, which we estimated through simulated data sets.

Parameter	Random uncertainties			Systematic uncertainties				
	Poisson	Flux & dist.	Boundary	Cosmic var.	Sample var.	Dist. model	Abs. flux	2DSWML
α	0.01	0.007	~ 0.01	~ 0.1	0.02	~ 0.01	–	0.02
$m_* + 2\log h_{70}$	0.01	0.004	~ 0.01	~ 0.02	0.01	~ 0.01	~ 0.04	0.002
$\phi_*/h_{70}^3 \,\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$	2×10^{-4}	7×10^{-5}	$\sim 3 \times 10^{-4}$	$\sim 6 \times 10^{-4}$	3×10^{-4}	$\sim 1 \times 10^{-4}$	–	–

4.2 Choice of boundary

The ALFALFA survey does not have perfectly binary coverage, and therefore defining the area in which the HIMF will be estimated is somewhat subjective. The survey was performed using a double pass drift-scan strategy and thus the edges are ragged in RA due to differing start time. The nominal boundary was chosen to ensure the full area had at least single coverage (as far as was possible). A very conservative boundary was also created by moving $0.5°$ in Dec. and to the nearest half hour of RA inside the nominal boundary (see Fig. 1). Enforcing this highly conservative boundary cut changes the best-fitting parameters to $\alpha = -1.26 \pm 0.02$, $m_* = 9.94 \pm 0.01$, and $\phi_* = 4.8 \pm 0.2 \times 10^{-3}\,\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$, indicating that for α the uncertainty in the choice of boundary is a similar scale source of error as the Poisson counting uncertainty.

It should be noted, however, that the patchy coverage at the extremes of the survey footprint are unlikely to be the main cause of this shift in the parameter values. Altering the boundary in this way creates a change in two inseparable systematic effects. The first is the intended effect, the elimination of the non-uniform coverage near the edge of the survey. The second is that the footprint of the survey is changed, which means that it covers a slightly different range of local environments. The strict boundary chosen removes area from the edges of the Fall and Spring skies, but not from the vicinity of Virgo, which (as will be discussed below) will act only to steepen the low-mass slope, meaning that the above approach likely gives an overestimate of the impact of the choice of boundary. If no boundary is imposed, α and M_* are unchanged from the values with the nominal boundary (ϕ cannot be calculated without knowing the sky area).

4.3 Cosmic and sample variance

Cosmic variance is a phrase used to mean a number of different things in different contexts; here, we use it to mean that the Universe contains regions of different underlying density, which may potentially impact the form of the mass function (MF). Therefore, unless a survey volume is large enough to contain a fair sample of all of these environments there will be a systematic bias in the resulting MF, regardless of how precisely it can be determined within that volume, and therefore, it will not be representative of the Universe's global MF (at that redshift). To estimate the scale of this effect, we simply take the difference in the HIMFs in the ALFALFA Spring

and Fall regions. These are two regions with completely disparate large-scale environments and so should give a reasonable approximation of how much the parameters can be expected to vary on larger scales (see Section 5.1).

There is another effect of cosmic variance (which here we will label 'sample variance'), which is that it limits how well the HIMF can be determined inside the survey volume because the HIMF may vary across it. To estimate this effect, we split the survey area into 46 approximately equal area, contiguous regions, and jackknifed the HIMF calculation, removing one region each time. The sums of the squared deviations from the mean parameter values (across all jackknife samples), multiplied by $(N - 1)/N$, give estimates of the parameter uncertainties due to sample variance. These are $\sigma_\alpha = 0.02$, $\sigma_{m_*} = 0.01$, and $\sigma_{\phi_*} = 3 \times 10^{-4}\,\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$.

4.4 Flow model uncertainty

ALFALFA uses a local Universe flow model (Masters 2005) and additional corrections (see Section 4.1 and Appendix B) to estimate source distances. In Section 4.1, we test the impact of random errors within this framework, but that does not give an estimate of the systematic uncertainty stemming from the use of flow model itself, as opposed to some other method for estimating distances. As pointed out in Masters, Haynes & Giovanelli (2004), an incorrect flow model can lead to substantial systematic biases in the derived HIMF. To estimate the scale of this uncertainty, we recalculate the HIMF using both the flow model of Mould et al. (2000) and pure Hubble flow only (discussed further in Appendix B). The resulting Schechter function parameters are shown in Table 2. By taking the standard deviation between these three measurements, we obtain a very approximate estimate of the uncertainty due to the distance model we use. These are $\sigma_\alpha \approx 0.01$, $\sigma_{m_*} \approx 0.01$, and $\sigma_{\phi_*} \approx 1 \times 10^{-4}\,\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$.

4.5 Absolute flux scale

Perhaps the largest source of systematic error for M_* is the calibration of the absolute flux scale. The flux scale is calibrated by observing a calibration diode intermittently during drift scans, and an additional correction is applied during the data reduction process after matching the continuum source fluxes in the survey area to known catalogues (Haynes et al. 2011, and references therein). This calibration is estimated to be accurate to about 10 per cent,

Table 2. Schechter function parameter values for each of the ALFALFA samples for which the HIMF was calculated. Note that maximum recession velocities (relative to the CMB) are enforced as derived distance cuts at v_{max}/H_0. The error estimates listed account only for Poisson counting uncertainties, see the text for estimates of the contributions of other error sources. The Spring and Fall samples represent the HIMF in the two separate regions of the ALFALFA survey in the Arecibo Spring and Fall skies. The strict bound samples have highly conservative boundaries that eliminate portions of the survey area that might not have complete coverage, this also changes the area considered that creates a shift in the parameters due to local cosmic variance. The two Virgo samples are for within 3 Mpc of the cluster centre (M87) and a wide slice covering the full ALFALFA Dec. range and 1 h in RA. Finally, the Hubble and Mould et al. (2000) calculations are based on two alternative methods for calculating the source distances.

Sample	$v_{max}/\mathrm{km\,s^{-1}}$	N_{gal}	α	$m_* + 2\log h_{70}$	$\phi_*/h_{70}^3\,\mathrm{Mpc^{-3}\,dex^{-1}}$
$\alpha.100$	15000	22831	-1.25 ± 0.02	9.94 ± 0.01	$4.5 \pm 0.2 \times 10^{-3}$
$\alpha.100$ Spring	15000	14391	-1.29 ± 0.02	9.94 ± 0.02	$4.9 \pm 0.3 \times 10^{-3}$
$\alpha.100$ Fall	15000	8440	-1.15 ± 0.02	9.92 ± 0.02	$4.3 \pm 0.3 \times 10^{-3}$
$\alpha.100$ Strict bound	15000	19268	-1.26 ± 0.02	9.94 ± 0.01	$4.8 \pm 0.2 \times 10^{-3}$
$\alpha.100$ Spring Strict bound	15000	12318	-1.30 ± 0.02	9.94 ± 0.02	$5.0 \pm 0.3 \times 10^{-3}$
$\alpha.100$ Fall Strict bound	15000	6950	-1.15 ± 0.03	9.91 ± 0.02	$4.8 \pm 0.3 \times 10^{-3}$
$\alpha.100a$	4000	3815	-1.22 ± 0.02	9.76 ± 0.04	$6.2 \pm 0.5 \times 10^{-3}$
$\alpha.100a$ Spring	4000	2634	-1.24 ± 0.02	9.75 ± 0.04	$6.2 \pm 0.6 \times 10^{-3}$
$\alpha.100a$ Fall	4000	1181	-1.08 ± 0.03	9.70 ± 0.05	$7.7 \pm 0.8 \times 10^{-3}$
Virgo RA slice	3000	695	-1.23 ± 0.05	9.60 ± 0.10	$2.6 \pm 0.6 \times 10^{-2}$
Extended Virgo cluster	–	272	-1.20 ± 0.12	9.54 ± 0.26	0.30 ± 0.16
$\alpha.100$ (Hubble)	15000	22815	-1.25 ± 0.01	9.94 ± 0.01	$4.6 \pm 0.1 \times 10^{-3}$
$\alpha.100$ (Mould et al.)	15000	22693	-1.26 ± 0.02	9.96 ± 0.01	$4.3 \pm 0.2 \times 10^{-3}$

which corresponds to a systematic uncertainty of 0.04 dex in M_*. However, this uncertainty should not impact the low-mass slope as it acts to shift all sources equally, simply creating a shift in the horizontal direction in the HIMF.

As discussed in detail in van Zee et al. (1997), with carefully constructed observations a flux calibration at the level of a few per cent can be obtained; however, this rises to about 5 per cent when baseline uncertainties are included. The estimate for ALFALFA is larger than this because in a blind survey each observation cannot be tailored with the goal of producing high procession fluxes (as was the case for van Zee et al. 1997), and because it is intended to be conservative, including various sources of uncertainty such as drift in the system gain, standing waves due to interference or continuum emission, and baseline variations. Further details of ALFALFA's calibration can be found in section 5.2 of Haynes et al. (2011).

It should also be noted that ALFALFA's source extraction procedure is non-optimal for sources that are larger than the Arecibo primary beam; therefore, the largest galaxies (in angular extent) may have larger flux uncertainties. However, making the simple approximation that the H I radius of a galaxy is double its optical radius (optical radii taken from the Arecibo General Catalog, a galaxy data base maintained by MPH and RG), we estimate that only 2 per cent of ALFALFA detections have H I extents greater than 4; thus, this is not a major concern.

4.6 The 2DSWML method

The final source of uncertainty we consider is bias in the 2DSWML method itself. To do this, we created 100 mock catalogues of approximately 1 million H I sources each (following the methodology of Jones et al. 2015, to create the mocks), all with the same input HIMF. The large source count was required such that Poisson errors were not a limiting factor. The HIMF for each mock was calculated using the same process as for the real ALFALFA data. As before, the resulting distributions of the Schechter function fit parameters were approximated by Gaussians and gave the uncertainties as: $\sigma_\alpha = 0.02$, $\sigma_{m_*} = 0.002$. σ_{ϕ_*} was not calculated in this analysis, as the 2DSWML method does not return the normalization.

4.7 H I self-absorption

In principle, 21 cm emission can be absorbed by other H I residing in the source galaxy, this is known as H I self-absorption. This effect is often neglected entirely because the cross-section for interaction between an H I atom and a 21 cm photon is very small, and it is a challenging and somewhat controversial topic, with some authors finding that all but the most highly inclined galactic discs are almost transparent to H I (e.g. Giovanelli et al. 1994), while others indicate that typical corrections could be as high as 30 per cent for all galaxies (e.g. Braun et al. 2009; Braun 2012). To assess what impact this has on ALFALFA, and in particular our estimate of $\Omega_{H I}$, we follow a procedure based on Giovanelli et al. (1994) using the H I data from ALFALFA in combination with the exponential radii and axial ratios calculated in Sloan Digital Sky Survey (SDSS) DR14 (Abolfathi et al. 2017).

If a single galaxy could be rotated and viewed from multiple orientations then as it became more edge-on its total integrated H I emission would be expected to decrease somewhat as the line of sight takes a longer path through the disc. By using axial ratios as a proxy for the inclinations of many galaxies (instead of rotating a single galaxy), the rate at which the *average* observed H I mass changes with inclination can be determined. For a sample of galaxies with similar morphologies and H I masses, the average observed H I mass should be independent of inclination, except for self-absorption.

As the vast majority of ALFALFA sources are late-type spiral galaxies we will assume the morphology has a minimal effect. However, the intrinsic axial ratio of galaxies is likely dependent on their mass, so we will exclude dwarf galaxies from this analysis. In addition, it is necessary to remove sources where the minor axis is comparable in angular size to the resolution of the optical image. However, making almost any cut to the sample based on the optical properties is problematic. For example, if this cut is made just at a constant angular resolution then it is creating a distance-dependent effect, but if it is scaled linearly with distance then it creates a very stringent resolution requirement at small distances. We therefore decided (starting from the 22 831 high signal-to-noise sources used to calculate the $\alpha.100$ HIMF) only to consider sources in the distance range of $100 < D/\mathrm{Mpc} < 150$, which effectively sets the minimum

H I mass as log $M_{H I} \sim 9.1$, we also only consider sources that have unambiguously identified counterparts that are 'primary' objects in SDSS, and finally, we require that the major axis exponential radius (in *r*-band) is greater than $10 \times \frac{100 \text{ Mpc}}{D}$ arcsec – with this value even the most inclined galaxies have minor radii of ~ 1 arcsec, thus the measured axial ratio should not be strongly influenced by the image resolution. This leaves a sample of 2022 galaxies. While this selection undoubtedly results in some level of bias with respect to inclination (axial ratio), it is aimed at minimizing those sources of bias of which we are aware.

With this sample, we fit a linear trend, using weighted least squares regression, between $\log (a/b)$ and $\log M_{H I}$. Giovanelli et al. (1994) split their sample into distance bins in order to minimize distance-dependent biases; however, ALFALFA is a considerably better characterized population (from the H I perspective) than their sample was; hence, we instead weight each source by the factor $M_{H I}/V_{\text{eff}}$ (with the v_{eff} values truncated to match the distance boundaries above). This is essentially weighting by mass density, which should produce a correction factor that is appropriate for correcting $\Omega_{H I}$.

We note here that without any weighting the sample is consistent (within 1σ) with galactic discs being entirely transparent to H I emission, except in cases where the disc is almost exactly edge-on. With the mass density weighting, the correction factor is found to be $\Delta \log M_{H I} = (0.13 \pm 0.03) \log(a/b)$, which is consistent with that found by Giovanelli et al. (1994). An increase in the dependence between $\log (a/b)$ and $\log M_{H I}$ is expected when the data are weighted by their v_{eff} values as sources with high-velocity widths (i.e. those that are more inclined) will be underrepresented in the unweighted data due to the selection effects of the survey. Using this relation, the most highly inclined galaxies in the sample require a correction of about 35 per cent, whereas at the log-mean axial ratio ($\log (a/b) = 0.36$) the correction is 11 ± 3 per cent, which we will use as our estimate of the correction required for $\Omega_{H I}$.

We do not use this value to make any correction to the shape or normalization of the HIMF itself. While this correction was designed to be appropriate for the measurement of the total cosmic H I mass density, the correction likely changes across the mass range probed by ALFALFA. Given the difficulties in creating a relatively unbiased sample for the massive galaxies in the survey, we make no attempt to quantitatively estimate the correction to the HIMF's shape due to self-absorption, but we note that it is reasonable to expect the 'knee' mass to also increase by ~ 10 per cent as this region of the HIMF makes the dominant contribution to $\Omega_{H I}$, and that corrections to the low-mass slope are likely quite small, as we expect the self-absorption correction to both decrease with decreasing mass and to do so gradually.

This result indicates that galactic discs are mostly transparent to H I emission, but a small correction is necessary to avoid bias. However, this analysis only considers effects that scale with inclination. Braun (2012) suggests that a large fraction of the H I in galactic discs may reside in high-density clouds (of ~ 100 pc in diameter) that could result in approximately 30 per cent of H I emission being self-absorbed. If these findings are correct then most of this self-absorption would not be apparent in an analysis such as ours because inclination would not be relevant for self-absorption occurring within such compact clouds. Having said this, those results are based on only three galaxies and rely on the accuracy of the modelling of these dense clouds. To really resolve this issue will require extremely deep, ~ 100 pc scale H I imaging of a large sample of galaxies, which is not feasible with existing facilities.

4.8 Final estimates of the HIMF and $\Omega_{H I}$

Adding the random and systematic effects from Table 1 separately in quadrature (with the exception of cosmic variance) gives the final estimate of the HIMF within the ALFALFA volume as $\alpha = -1.25 \pm 0.02 \pm 0.03$, $m_* = 9.94 \pm 0.01 \pm 0.04$, and $\phi_* = (4.5 \pm 0.2 \pm 0.4) \times 10^{-3} \text{ Mpc}^{-3} \text{ dex}^{-1}$, where the first quoted errors correspond to our combined estimates of the 1σ random uncertainties and the second to the systematic uncertainties. This indicates that for sample sizes of $\sim 25\,000$ or fewer, the Poisson uncertainties are reasonable estimates of the random uncertainties, but the systematic uncertainties are likely significantly larger. The inclusion of the uncertainty of cosmic variance gives an estimate of how well ALFALFA can constrain the global $z = 0$ form of the HIMF: $\alpha = -1.25 \pm 0.02 \pm 0.1$, $m_* = 9.94 \pm 0.01 \pm 0.05$, and $\phi_* = (4.7 \pm 0.2 \pm 0.8) \times 10^{-3} \text{ Mpc}^{-3} \text{ dex}^{-1}$.

The value of $\Omega_{H I}$ (the H I content of the Universe) can be estimated from the HIMF by integrating the Schechter function, which gives

$$\Omega_{H I} = \frac{1}{\rho_c} \phi_* M_* \Gamma(\alpha + 2), \tag{4}$$

where ρ_c is the Universe's critical density (calculated assuming $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$). The limits of the integral are taken to be indefinite because both the extreme low- and high-mass regimes make negligible contributions. This gives $\Omega_{H I\text{-corr}} = (3.9 \pm 0.1 \pm 0.6) \times 10^{-4}$, which is consistent at the 2σ level (random) with the value from $\alpha.40$ (Martin et al. 2010), when the scaling for H I self-absorption is removed ($\Omega_{H I\text{-uncorr}} = (3.5 \pm 0.1 \pm 0.5) \times 10^{-4}$).[5] Our value is ~ 15 per cent lower because both the 'knee' mass is lower and the low-mass slope is flatter in $\alpha.100$ than $\alpha.40$, both of which act to decrease $\Omega_{H I}$. This value ($\Omega_{H I\text{-uncorr}}$) is completely consistent with that of HIPASS (Zwaan et al. 2005); however, this agreement hides the substantially different shapes of the HIMF, which do not agree within the random errors. The $\Omega_{H I}$ values are consistent because our 'knee' mass is larger but our low-mass slope is substantially flatter.

Driver & Robotham (2010) estimated that volumes larger than 10^7 Mpc^3 with square survey footprints will have minimal contributions from cosmic variance. The volume over which we calculate the $\alpha.100$ HIMF is approximately half of this value, and that volume is split into two separate pieces (Fig. 1) that acts to reduce the volume needed to overcome cosmic variance. This suggests that for sources that are detectable over almost all this volume (M_* galaxies) our Schechter function parameter values should only be weakly impacted by cosmic variance; however, for low-mass galaxies the accessible volume is only $\sim 10^5 \text{ Mpc}^3$, which may be strongly biased by cosmic variance. This is consistent with our estimates of cosmic variance in Table 1. In other words, while the 'knee' mass of the ALFALFA HIMF should be viewed as cosmologically fair, the low-mass slope should be considered a measurement of the value

---

[5] The systematic uncertainty in $\Omega_{H I}$ is difficult to measure, because it requires estimates of the covariance between the systematic uncertainties in α, m_*, and ϕ_*, which are not available to us. Treating all the systematic uncertainties estimated above as independent would likely lead to an overestimate of systematic uncertainty in $\Omega_{H I}$ (1.0×10^{-4}), while treating them as highly correlated would likely lead to an underestimate (0.1×10^{-4}). We therefore choose only to consider the most dominant source of uncertainty, the impact of the absolute flux calibration of m_*, which produces the quoted level of uncertainty (0.6×10^{-4}). As this was a conservative estimate to begin with and it is the dominant source of uncertainty for $\Omega_{H I}$, it is a reasonable estimate of the total systematic uncertainty.
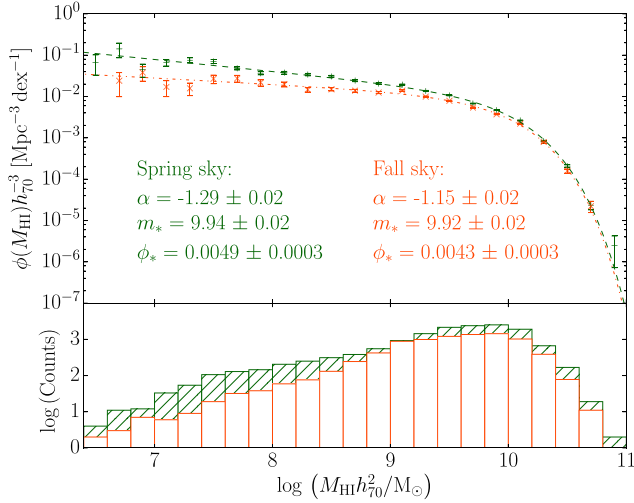
Figure 3. The HIMFs of the $\alpha.100$ Spring and Fall skies. The lower and upper panels are as previously described, with the green plus signs, dashed line, and filled bars corresponding to the Spring sky, and the orange crosses, dotted line and unfilled bars to the Fall sky. The 'knee' mass in the two directions agrees within the uncertainties, but the Fall sky has a significantly flatter low-mass slope.

in the nearby Universe, and it therefore may not be representative of the volume beyond the Local Supercluster.

5 VARIATIONS IN THE HIMF

We have found that the HIMF varies across the ALFALFA volume. In this section, we present these large-scale changes in the shape of the HIMF across the two regions of the survey, the nearby and full samples, and in the direction of the Virgo cluster.

5.1 Comparison of the spring and fall sky

The nearby LSS in the Spring and Fall portions of the ALFALFA footprint is decidedly disparate, with the Virgo cluster and the Local Supercluster dominating the Spring direction, while the Fall sky is sparsely populated due to the deep void in the foreground of the Pisces-Perseus Supercluster (PPS). Contrasting the HIMF in these two directions is the first step to assessing how it may vary between large scale over and under densities. To make the comparison the $\alpha.100$ volume is simply split into the Spring and Fall directions, and the HIMF calculated in each independently. Fig. 3 shows the two HIMF and their Schechter function fits.

Separating the population in this way immediately makes it clear that the Fall region has a substantially flatter low-mass slope than the Spring region, at over 3σ significance (based on Poisson random uncertainties), whereas the 'knee' masses in the two directions agree within their errors. This difference in the slope persists in the nearby Spring and Fall catalogues at similar significance, as the Fall low-mass slope is somewhat lower in the nearby volume (see Table 2).

5.2 Comparison of interior and exterior to 4000 km s^{-1}

To test of large-scale effects with the line-of-sight distance within the ALFALFA volume, we compare the $\alpha.100$ HIMF within ~57 Mpc (4000 km s$^{-1}/H_0$) to that of the full sample. This nearby sample will be referred to as $\alpha.100a$. At this distance, the ALFALFA 50 per cent completeness limit would just include a
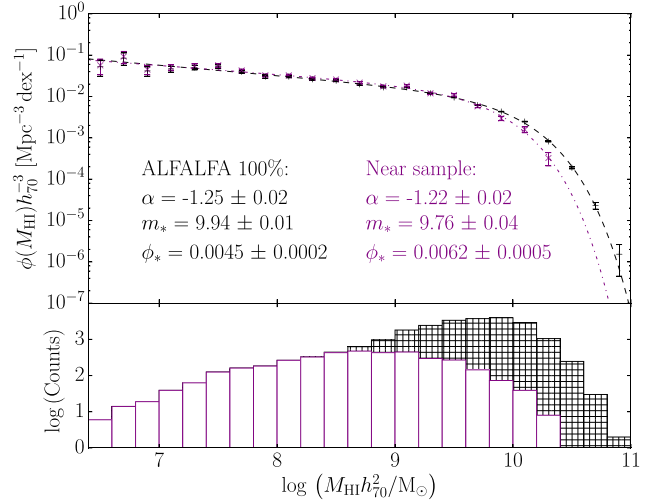

Figure 4. The HIMFs of the $\alpha.100a$ ($v_{cmb} < 4000$ km s^{-1}) and full $\alpha.100$ samples. The lower panel shows the observed counts for the near (unfilled) and full samples (cross hatching), while the upper panel shows the inferred HIMF of those samples with purple crosses and black plus signs, respectively. The Schechter function fits are shown with dash-dot (near) and dashed (full) lines. These two samples appear to have best-fitting 'knee' masses that are different by almost 0.2 dex, but similar low-mass slopes.

source of $\log M_{HI}/M_\odot = 8.5$ and a velocity width of 50 km s^{-1}. Therefore, this represents a reasonable cut-off beyond which few objects on the low-mass slope can be detected. Fig. 4 shows that the low-mass slopes of the nearby sample and the whole $\alpha.100$ are within 1σ (Poisson) agreement, but that the 'knee' mass is locally lower by 0.18 dex. The former is unsurprising because (by construction) the nearby sample contains virtually all the objects detected on the low-mass slope and therefore much below $\log M_{HI} = 10^9$ M$_\odot$ the data points become identical. The latter is less straightforward to explain, with the difference in m_* being at approximately the 4σ (Poisson) level.

5.3 The HIMF of the Virgo cluster

Defining the volume that encompasses the Virgo cluster is nontrivial and there is no single definition, we therefore take two approaches. First, we take simply a box in RA (between 13 and 12 h) and Dec. (between $0°$ and $36°$) that stretches out to a maximum distance of ~43 Mpc (3000 km s$^{-1}/H_0$), as a very crude approximation of the greater Virgo volume, including all the major clouds and filaments that surround the cluster itself. This region contains 695 sources and its HIMF is shown in Fig. 5 by the gold plus signs. Hereinafter, this sample will be referred to as the 'Virgo slice'. We also take the volume within 3 Mpc of the centre of the cluster (defined by the position of M87 and with line-of-sight distances taken as those calculated from the procedure described in the Section 4.1). In the literature, the Virgo cluster is typically defined as a region within $5°$ of M87, whereas our sample extends to approximately double this radius; therefore, we will refer to sample as the 'extended Virgo cluster' sample. This volume was chosen because over 90 per cent of the 272 ALFALFA galaxies it contains are members of the VCC, and for smaller radii the sample size decreases rapidly, while for larger radii the fraction of sources that are in the VCC plummets, likely indicating an increase in contaminants. The HIMF of this volume is also shown in Fig. 5 (red diamonds).
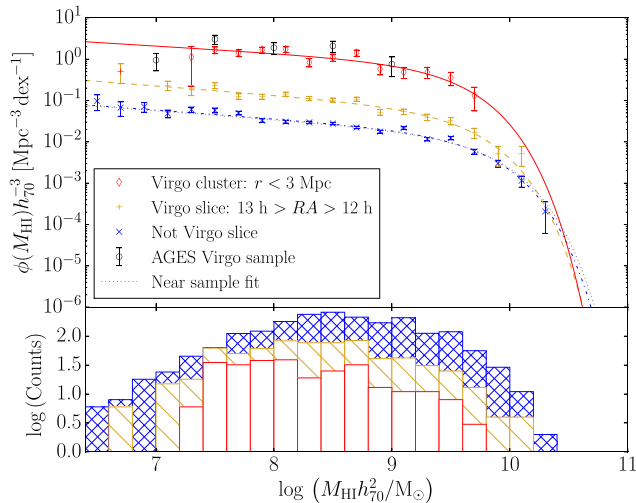
Figure 5. The HIMF of the Virgo region compared to that of the α.100a sample. Again the lower panel shows the raw numbers counts in each mass bin. The red, unfilled bars are the extended Virgo cluster sample, the single hatched gold bars are the Virgo slice sample, the blue cross hatched bars are the α.100a sample minus the Virgo slice, and the black circles are our estimates of the AGES Virgo HIMF (with simplifying assumptions, see Section 6.3). The upper panels show the non-parametric HIMF bin values and the Schechter function fits to each.

The HIMFs of the two Virgo samples are compared to the rest of the α.100a catalogue and the best fit to the entire α.100a catalogue in Fig. 5. It can be seen by eye that all four share approximately the same low-mass slope and 'knee' mass (values displayed in Table 2). Unsurprisingly the extended Virgo cluster is much more densely packed with galaxies than the general ALFALFA sample, and the normalization constant (ϕ_*) increases from the α.100a sample, to the Virgo slice, to the extended Virgo cluster sample. This demonstrates that, other than the normalization, the Virgo HIMF is broadly consistent with the HIMF of the whole ALFALFA sample within 4000 km s^{-1}. It should be noted, however, that the 'knee' mass is quite poorly constrained due to the low number of high H I mass objects in this small volume. Our extended Virgo sample HIMF also appears to be approximately consistent with the observations of Arecibo Galaxy Environment Survey (AGES; Taylor et al. 2012, 2013), although this comparison has a number of caveats that we discuss below.

6 DISCUSSION

The results shown in the previous section demonstrate three distinct trends in the HIMF of ALFALFA:

(i) The low-mass slope is significantly flatter in the Arecibo Fall sky direction than in the Spring sky.

(ii) The 'knee' mass is almost 0.2 dex lower within $\sim$60 Mpc than it is out to $\sim$200 Mpc.

(iii) The shape of the HIMF in the Virgo cluster is consistent with the shape of the HIMF of ALFALFA sources at similar distances.

The first two of these are not directly dependent on each other, as the Fall sample in both the full and nearby catalogues shows the flattening of the low-mass slope (Table 2), but the Fall α.100 sample does not show the suppressed M_* value that is present in all the nearby samples, even though there is a low significance suggestion that it is most suppressed in the Fall α.100 sample (however, note that there is a strong covariance between α and m_*). One natural

conclusion is that these phenomena are the result of environmental dependence, as the Arecibo Spring sky faces towards the centre of the Local Supercluster, whereas the Fall sky faces a foreground void in front of PPS. It is possible that the characteristic H I mass is somewhat reduced within the Local Supercluster and that the low-mass slope is flatter around the PPS foreground void. As the low-mass slope of the entire α.100 sample is necessarily dominated by the galaxies detected nearby (where those on the low-mass slope are detectable), this would explain why the Spring/Fall dichotomy is present regardless of the maximum distance cut-off of a sample. Equally, as the 'knee' mass of the α.100 sample is dominated by the sources detected at $\sim$150 Mpc, where ALFALFA detects most of its M_* galaxies, this would naturally explain why the suppressed 'knee' mass is only apparent in the low-redshift samples. However, before accepting this interpretation we will first demonstrate that we have ruled out a number of potential biases that could have produced similar apparent trends.

6.1 The effect of a small volume

An immediate criticism of these results is that the samples within 4000 km s^{-1} correspond to small volumes that are not cosmologically fair. This is certainly true as the present reality is that current wide-field H I surveys are only capable of measuring the low-mass slope within the Local Volume; however, there are sufficiently many low-mass objects detected in this volume to make a statistically significant detection of a discrepancy between the low-mass slopes of the HIMFs in the Spring and Fall skies. While this may or may not be representative of the behaviour of the H I population as a whole, there is presumably still a physical explanation of this discrepancy between these two starkly different environments. We will return to discuss the low-mass slope later on, and instead focus here on the shift seen in M_* between the local and more distant volumes, and what effect the small (local) volume considered might have on the result.

The simplest bias that could cause an apparent shift in the 'knee' mass would be if the volume considered were sufficiently small that too few M_* galaxies existed in it to accurately constrain the value in that volume. To test this hypothesis, we generated 3000 mock catalogues with an equivalent source density to α.100, but with the small volume of the nearby sample. The input HIMF of the whole sample ($\alpha = -1.25$, $\log M_* = 9.94$) was used, and then cut at the ALFALFA 50 per cent completeness limit (Haynes et al. 2011). The details of the method used to generate the mocks can be found in Jones et al. (2015). The *observed* HIMF was then calculated for each mock, exactly as it would be for ALFALFA. The lowest 'knee' mass found in any of the mocks was $\log M_* = 9.82$, and based on the distribution of all the mock values there is a vanishingly small probability of obtaining a value as extreme as $\log M_* = 9.76$. This demonstrates that the variance in the parameters cannot be attributed to the small number of sources in the Local Volume alone.

In the Local Volume, distances based on redshifts are always a major concern because the Hubble flow velocities are comparable to the magnitude of galaxy peculiar velocities, and without accurate distances, accurate masses cannot be determined. ALFALFA calculates most distances in the Local Volume based on the flow model of Masters (2005), but while this is certainly an improvement over assuming Hubble flow, it does not completely alleviate these concerns. There are however a number of observations that lend support for this method and indicate that it is not the cause of the effects we find. First of all, the Virgo samples show suppressed values of M_*. The extended Virgo sample

contains almost exclusively objects that have been assigned distances from the VCC (see Section 6.3). These are objects that have been identified as part of the Virgo cluster, given the primary distance measured for the relevant cloud. Therefore, the fact that the values of M_* follow the sample pattern, barring the possibility of an exceptionally different 'knee' mass within Virgo, strongly suggests the distances are reliable, at least on average. Furthermore, to reconcile the difference in M_* between the nearby and full samples would require the nearby ($v_{\rm cmb} < 4000 \, {\rm km \, s^{-1}}$) distances to be systematically underestimated by about 25 per cent. The flow model ALFALFA uses gives a velocity correction that is less than 25 per cent of the CMB frame velocity in over 70 per cent of cases for the α.100a sample, with those occurring as positive and negative corrections in approximately equal amounts. This again indicates that a sufficient systematic error could not plausibly be introduced by this method. As a final check, we used two completely different methods to estimate the distances (pure Hubble flow and the flow model of Mould et al. 2000), and in both cases, M_* was still suppressed relative to the full sample (see Appendix B).

6.2 Trends on large angular scales

Fig. 3 shows that the low-mass slope in the Fall sky is flatter at over 3σ (Poisson) significance compared to the Spring sky. The Fall sky contains a deep void (in front of PPS), suggesting that this trend might be associated with the underdensity in the region. However, the low-mass slope of the optical luminosity function (LF) of void galaxies in the SDSS has been found to be consistent with global population (Hoyle et al. 2005; Moorman et al. 2015). Therefore, if the observed trend is associated with the presence of a void then the low-mass H I population would need to respond differently to underdensities than the low-mass optical population. Whether or not this is the case is unclear, but it is possible that the UV background or star formation events could preferentially ionize the H I in low-mass galaxies in voids, leading to different behaviours in the optical and H I low-mass slopes with large-scale environment. To further test this hypothesis requires a population of low-mass galaxies in H I to be detected in other large-scale underdensities. This may be accomplished by the next generation of H I surveys about to be carried out by Square Kilometre Array (SKA) pathfinder facilities, and will certainly be achievable for the SKA itself.

Turning our attention now to the Spring sky, the trend of the steeper low-mass slope appears to be associated with the Virgo direction; however, there are not enough sources in the Virgo cluster alone to dominate the form of the HIMF over the whole Spring sky. The greater Virgo region in the ALFALFA volume is a complex knot of filaments that are in the process of falling on to the cluster, and it appears that we are looking approximately along the length of the main filament (Solanes et al. 2002; Mei et al. 2007). The whole structure is confined to a few hours in RA and a few tens of degrees in Dec., meaning it covers most of ALFALFA's Dec. range, but a much smaller fraction of its RA range (in the Spring sky). To assess to what extent the steeper slope is associated with Virgo rather than the Spring sky as a whole, the HIMF was calculated in 2 h wide (overlapping) bins of RA, spaced every 1 h across the whole α.100 sky. Fig. 6 shows the measured value of α in each RA bin.

A sharp steepening of the low-mass slope is clearly visible centred on 12 h (the approximate direction of the Virgo cluster), where α decreases around to -1.30, before rising again to nearer -1.15. In the Fall sky, there is no visible trend with the same slope being consistent with all bins. Encouragingly, this pattern is preserved when distances are calculated either with the Mould et al. (2000)
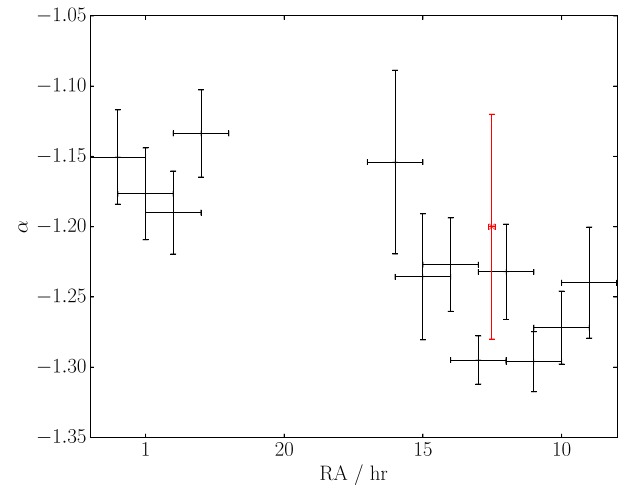

Figure 6. Low-mass slope (α) fit parameter for 2 h wide RA bins spaced across the ALFALFA sky. The values of α are shown for each bin with the black error bars. The single red cross and corresponding error bars indicate the value for the extended Virgo cluster sample. There appears to be a steepening of the low-mass slope that is associated with the direction of the Virgo cluster, but extends much further in RA.

method or with CMB velocities, indicating that it is not an artefact of our distance estimation method.

The location of the extended Virgo cluster sample low-mass slope is also plotted in red on Fig. 6. Although the error bar is large, this indicates that Virgo is not dominating the observed steep slope in this RA slice, and in fact suggests that the slope in Virgo may well be flatter than in its immediate surroundings.

The Virgo cluster itself appears not to be driving the observed steepening of the low-mass slope, yet this phenomenon is strongly associated with the RA of the cluster, which leads us to hypothesise that the filaments connected to Virgo may be driving this shift in α, and that this could be a more general effect of a galaxy's position in LSS (specifically gas-rich filaments).

6.3 The Virgo cluster

The Virgo cluster is known to be H I deficient (e.g. Giovanelli & Haynes 1985; Solanes et al. 2001; Boselli & Gavazzi 2006; Chung et al. 2009; Taylor et al. 2012) and objects falling into Virgo have been seen directly to be losing their gas (Kenney, van Gorkom & Vollmer 2004); therefore, it is perhaps somewhat surprising that its HIMF appears to be generally consistent with the rest of the ALFALFA sample at similar distances (except for the normalization). However, as is shown in Fig. 6 the low-mass slope of Virgo does seem to be considerably flatter than that of its immediate surroundings, even though it is consistent (at 1σ) with the α.100a sample as a whole. Therefore, we make the tentative suggestion that the greater Virgo region might be rich in H I, leading to a steep low-mass slope, but that the cluster instead is H I deficient and, accordingly, has a flatter slope.

Davies et al. (2004) studied the Virgo cluster in H I with the Jodrell Bank Lovell telescope and concluded that there is a dearth of low-mass H I galaxies relative to the HIPASS field population (Zwaan et al. 2003), and this result was confirmed by Gavazzi et al. (2005) with Arecibo observation of the Virgo cluster. However, neither of these works calculated the formal HIMF because they did not weight their detections based on their completeness and sensitivity limits. More recently, the H I population detected by ALFALFA in the

Virgo region was compared with AGES (Taylor et al. 2012, 2013). The two AGES fields within Virgo covered 25 deg^2, sampling both the cluster centre and its outskirts at a sensitivity about a factor of approximately four times greater than ALFALFA. They find a lower fraction of massive galaxies than ALFALFA, but point out that this is not very significant given the small area that they cover and the infrequency of such galaxies. They also find a much higher fraction of low-mass galaxies (below log $M_{HI}/M_\odot < 7.5$) than ALFALFA. However, as no completeness limit is derived for that data set, the formal HIMF is not calculated in those articles. This makes drawing any conclusions as to whether AGES population is or is not consistent with the low-mass slope that ALFALFA finds in Virgo problematic.

AGES detects a considerably higher fraction of low-mass objects in Virgo than ALFALFA does, but this in itself is not surprising as it was more sensitive, and the detected fractions are in agreement in the intermediate mass range where neither sensitivity nor survey area leads to large differences (see Taylor et al. 2013, their fig. 2). The lower fraction of low-mass objects detected by ALFALFA is corrected in the HIMF calculation because these object have much smaller v_{eff} volumes than more massive objects, but without a similar correction for the AGES Virgo sources a direct comparison is not possible. To this end, we make two strong assumptions to permit such a comparison: (1) that AGES is entirely complete within the surveyed regions in the Virgo cluster (A cloud at 17 Mpc) and (2) that the volume surveyed is exactly the area of the rectangular fields times the assumed depth of the cluster, 2.4 Mpc (Mei et al. 2007). With these two assumptions we can make a pseudo-HIMF from the AGES detections, which is shown in Fig. 5. This reveals that the ALFALFA and AGES data sets appear to be largely consistent within Virgo, although there is a slight suggestion that the normalization of the AGES HIMF may be higher, which is not surprising, given that the larger of their two Virgo fields targeted a particularly dense region of the cluster. It can also be seen that although AGES detects lower mass sources due to their greater sensitivity, the corrections applied to ALFALFA imply that there should be considerably more low-mass H I sources than even what AGES detected, i.e. for the lowest mass sources our first assumption appears to be invalid.

The H I deficiency in the cluster would also be expected to suppress the 'knee' mass, at least in the cluster core. Whether or not this has occurred is difficult to determine as M_* is very poorly constrained in the extended Virgo cluster sample. However, the reader may have noted that there are no galaxies at all above an H I mass of $10^{9.8}$ M$_\odot$ (this was also noted by Kent 2008). This may seem like a significant observation, but when the α.100a Spring HIMF is integrated starting from this mass, it shows that (based on a spherical volume of radius 3 Mpc) the extended Virgo cluster sample would only be expected to have fewer than two galaxies with H I masses greater than $10^{9.8}$ M$_\odot$ (assuming the cluster HIMF is 1.5 orders of magnitude more densely populated than the field). Thus, their absence could easily be explained by simple small number statistics. To accurately constrain the 'knee' mass in the Virgo cluster requires a larger sample. Unfortunately, this is not simply a matter of sensitivity, as ALFALFA has more than sufficient sensitivity to detect any H I M_* galaxies at that distance. The statistics are poor solely because a structure like the Virgo cluster does not contain many massive H I galaxies. Therefore, to confidently constrain the 'knee' mass in young clusters like Virgo requires surveying multiple such clusters at a similar H I mass sensitivity to what ALFALFA achieves in Virgo. Such a survey is only feasible with the capabilities of the SKA.
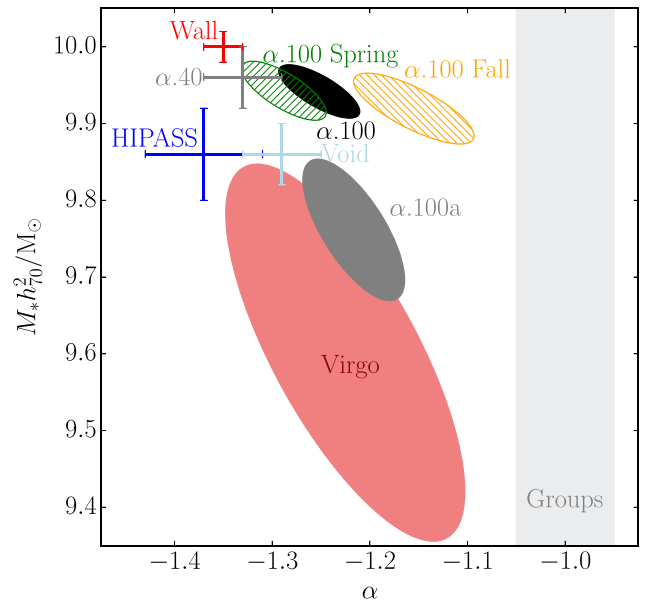

Figure 7. Comparison of various measurements of the HIMF shape parameters with 2σ (Poisson) error bars or ellipses. The solid black ellipse is the HIMF of the full α.100 sample. The green (upward hatching) ellipse is the Spring side of ALFALFA, and the orange (downward hatching) ellipse is the Fall side. Here, the Virgo sample (solid light red ellipse) is the Virgo slice sample covering 1 h of RA centred on the cluster out to a maximum CMB velocity of 3000 km s^{-1}. The α.100a (solid grey) sample covers the full footprint out to 4000 km s^{-1}. The α.40 (grey) and HIPASS (blue) points correspond to the HIMF of the 40 per cent ALFALFA catalogue (Martin et al. 2010) and the HIPASS catalogue (Zwaan et al. 2005). The Void (light blue) and Wall (red) samples are the results of Moorman et al. (2014), based on ALFALFA 40 per cent. The grey band on the right side of the plot shows approximately the typical range of slopes found for individual groups of galaxies. For more details on the α.100 samples, refer to Table 2.

6.4 Comparison and conflicts with previous results

Fig. 7 shows Schechter function shape parameters (the low-mass slope and the 'knee' mass) for the sub-regions discussed in this paper, along with several previous measurements of the HIMF and its variation with environment. In this section, we discuss each of these comparisons and their possible interpretations in turn.

6.4.1 Environmental dependence

The fact that our results suggest a large-scale environmental dependence of the low-mass slope seems to be in conflict with Moorman et al. (2014) and Jones et al. (2016), both of which studied the environmental dependence of the HIMF in ALFALFA (on large and small scales) and concluded that the 'knee' mass does vary on the order of 0.1 dex with environment, but neither found convincing evidence for a change in the low-mass slope. However, there are a number of plausible explanations for why this trend has not been seen before.

Moorman et al. (2014) defined their void and wall environments based on SDSS spectroscopy. These two measurements are shown by the light blue and red error bars in Fig. 7, and labelled as 'Void' and 'Wall'. That work concluded that the 'knee' mass was lower within voids than in walls, and found a tentative flattening of the low-mass slope at low significance. As there is substantial overlap between ALFALFA and SDSS spectroscopy in the Arecibo Spring sky, but not in the Fall sky, this would immediately prevent that

study from finding the large dichotomy between the Spring and Fall skies that we observe.

Jones et al. (2016) focused on a smaller scale definition of environment based on density of neighbours in SDSS DR8 (Aihara et al. 2011) or 2MRS (2MASS Redshift Survey; Huchra et al. 2012) around ALFALFA sources. That work did incorporate the Fall portion of $\alpha.70$ in some of their analysis (when using neighbour densities from 2MRS, which is all sky), but the definition of environment was too small scale to identify the dichotomy found here, and while it may have been large enough angular scale to find the trend associated with the Virgo direction, it was not focused on just the local structures. Therefore, its null result is unsurprising. Jones et al. (2016) also concluded that the shift in the value of M_* was associated with local environment ($\sim 2\,$Mpc scales), with it disappearing when environment was defined on a larger scale ($\sim 10\,$Mpc). Again, this is in apparent conflict with our present results (lower M_* in the Local Supercluster).

Knowing the results of this work, we returned to the environment (neighbour density) quartiles calculated for $\alpha.70$ in Jones et al. (2016) and determined whether $\alpha.100$ sources fell within voids or walls based on the Moorman et al. (2014) analysis. When split into RA bins across the Spring region of ALFALFA the mean environment quartile shows a similar pattern (although inverted) to that of Fig. 6, with the majority of the sources in the immediately vicinity of Virgo being classified as in high-density environments, but the average environment quartile decreases towards the edges of the Spring sky. The fraction of sources classified as in voids also climbs either side of Virgo, with the majority of ALFALFA's low-mass sources to the west of Virgo lying in voids and almost half at the eastern edge of the Spring sky. This pattern suggests that the variation in the low-mass slope is due to environment, with the steepest slopes being in high-density environments around Virgo (but not in the cluster itself), and the flatter slopes in lower density environments. This also raises the possibility that rather than the gas-rich regions connected to the Virgo cluster driving the slope to be steeper in that region, it may be the low-density environments elsewhere that are driving it to be flatter, or perhaps both. A deeper survey than ALFALFA is needed to further explore this dependence in the low-mass slope over a larger volume.

We also found that within the nearby volume the majority of the M_* galaxies were classified as being in the lower half of neighbour densities (compared to the full $\alpha.70$ sample), and that more than twice as many were classified as lying in voids than in walls (compared to a ratio of 2:1 in the opposite sense for the full sample). This would then imply agreement with the findings of Moorman et al. (2014) and Jones et al. (2016) that the 'knee' mass is decreased in low-density environments and that it is causing the observed suppression of the local value.

While all these findings appear to form a coherent picture, caution is advised because this agreement could still be circumstantial and not indicative of the root cause of these variations.

6.4.2 HIPASS

The lower 'knee' mass that we observe in the local Universe conveniently explains the apparent discrepancy in 'knee' masses measured by HIPASS and ALFALFA (Zwaan et al. 2005; Martin et al. 2010). The 'knee' mass that we calculate is 0.08 dex higher than the HIPASS value (0.1 dex compared to Martin et al. 2010). Although HIPASS's nominal volume is larger than ALFALFA's due to its superior sensitivity, the volume in which ALFALFA can detect a

galaxy of $\log M_{H_I} = 10$ is over six times greater than that available to HIPASS (estimated based on 50 per cent completeness limits, Zwaan et al. 2004; Haynes et al. 2011). This, combined with our finding of a lower local value of M_*, suggests that HIPASS's lower value of M_* (see Fig. 7) was caused by their 'knee' mass determination being dominated by the local region where its value appears to be low.

HIPASS also found a steeper low-mass slope than any of the ALFALFA samples. We have seen that the value of α can change on large angular scales, so there remains the definite possibility that the low-mass slope is intrinsically steeper in the Southern sky. However, as HIPASS used pure Hubble flow distances (rather than using a flow model) and implemented their completeness limit slightly differently to ALFALFA, caution is advised when interpreting this difference.

6.4.3 Groups

Studies of the HIMF in individual groups (e.g. Verheijen et al. 2001; Kovac, Oosterloo & van der Hulst 2005; Freeland, Stilp & Wilcots 2009; Kilborn et al. 2009; Pisano et al. 2011; Westmeier et al. 2017) have generally found that they have approximately flat ($\alpha = -1$) low-mass slopes, with the notable exceptions of Stierwalt et al. (2009) and Davies et al. (2011), both of which found very steep ($\alpha < -1.4$) low-mass slopes in the Leo region and across various groups within the AGES footprint, respectively. These results are apparently in tension with those of ALFALFA and HIPASS, which measured a steeper slope ($\alpha \sim -1.3$) and do not find a flattening of the slope in higher density regions (Zwaan et al. 2005; Martin et al. 2010; Moorman et al. 2014; Jones et al. 2016).

The large-scale shifts in the value of α that we have found highlight that it is important for H I studies of individual groups to consider the larger scale environment of groups when comparing their HIMFs to those of wide-field surveys. Although ALFALFA does not cover the sky areas where most of these group studies have been carried out, it is clear from our results that large-scale environmental changes can cause substantial shifts in α, meaning that the position of a group within LSS may be comparably important as the difference in environment between the field and groups. Having said this, none of the sub-regions or previous results plotted in Fig. 7 enter the grey band that approximately represents the findings in galaxy groups, so this does not appear to be a complete explanation of the apparent tension.

There have been several recent indications that group galaxies are 'pre-processed' (e.g. Hess & Wilcots 2013; Odekon et al. 2016; Brown et al. 2017), losing a significant amount of their H I upon joining groups. While this pre-processing could be an explanation for the flatter low-mass slope found in galaxy groups, without reference to the larger scale environment, it does not explain why a transitional HIMF has not been identified with ALFALFA or HIPASS. In addition, low-stellar-mass group centrals may actually have more H I than their counterparts in the field (Janowiecki et al. 2017), which suggests that the apparent amount of processing is probably also a function of group size and potentially position relative to the cosmic web.

An additional complicating factor is that most groups have been studied using interferometric data (again Stierwalt et al. 2009 and Davies et al. 2011 are exceptions, but so are Pisano et al. 2011 and Westmeier et al. 2017), whereas both HIPASS and ALFALFA are single dish surveys. The completeness limit is extremely important when calculating the HIMF, and interferometers suffer from the

additional complication that there is a surface brightness limit as well as an integrated flux limit. This is irrelevant in almost all cases for single dish telescopes, as they seldom resolve any but the most nearby sources. If not adequately accounted for this effect could lead to an under counting of low surface brightness sources and a bias in the low-mass slope. Furthermore, the completeness limit of H I surveys also depends on the source velocity width, because a spectrum can be smoothed to enhance signal to noise. This means that some sources of a given H I integrated flux may be detected, but other sources with the same flux may not be detectable because their velocity width is too wide and their flux is spread across too many channels. In Appendix C, we argue that in the case of groups a correction for this effect should always be applied; however, it typically is not, and this can lead to an underestimation of the slope.

Minchin (2017) argued that by combining flat Schechter function sub-HIMFs (HIMFs of sub-populations or groups) of different 'knee' masses, the global HIMF shape could be recovered, which would resolve the conflict as groups could have flat slopes, provided there was the appropriate distribution of 'knee' masses in the sub-HIMFs across all environments. While this scenario is difficult to disprove, the vast majority of ALFALFA sources are not members of groups, but are centrals in the field (Guo et al. 2017); thus, this approach would amount to summing together a different (flat) sub-HIMF for each galaxy, which prompts the question: what do we mean by an MF and can an individual galaxy have an associated MF? This, however, raises an important point, which is that the ALFALFA population principally probes the H I content of centrals, whereas studies of individual groups probe the H I content of the satellites (as each group can only have one central). Therefore, as they detect distinct populations it is perhaps unsurprising that the two types of H I data sets have a disconnect in the observed HIMF slopes and that a transition has not been identified in ALFALFA.

6.5 Prospects for future Hi surveys

Over the next 5–10 yr, there will be a slew of new blind H I surveys that will be performed by SKA-pathfinder facilities (e.g. Duffy et al. 2012; Giovanelli & Haynes 2016). The direct successors to HIPASS and ALFALFA will be WALLABY (Widefield ASKAP L-band Legacy All-sky Blind surveY) in the Southern hemisphere and the Westerbork–Apertif shallow northern sky survey in the Northern hemisphere. Together these H I surveys will cover the entire sky and detect an order of magnitude more galaxies than ALFALFA. Another class of surveys will also be carried out that perform deep integrations of limited fields. These surveys will come in two varieties: (1) mapping of small fields such as the Westerbork Medium Deep survey and the Deep Investigation of Neutral Gas Origins (DINGO) survey (Meyer 2009), which will detect a comparable number of sources to ALFALFA, but out to $z \sim 0.25$, and (2) extremely deep single pointing surveys that aim to detect H I galaxies out to a redshift of order unity, such as Looking At the Distant Universe with MeerKAT (LADUMA; Holwerda, Blyth & Baker 2012) and the ongoing COSMOS H I Large Extragalactic Survey (CHILES; Fernández et al. 2013).

The results of this paper suggest that these surveys will contribute to the study of the HIMF in two key ways. First, the low-mass slope will be measurable in the field outside the Local Volume for the first time. This will give the first cosmologically fair account of how numerous low-mass, gas-rich galaxies really are at $z = 0$. And second, the 'knee' mass will become measurable beyond $z \approx 0$. Due to the volume of ALFALFA, these next-generation surveys are unlikely to find a significantly different 'knee' mass; however, they

will have the depth and source counts to begin to place constraints on its evolution as a function of z.

While the redshift range covered by SKA-precursor surveys will far exceed what is possible with single dish telescopes (due to the increase in source confusion with redshift, Jones et al. 2015; Elson, Blyth & Baker 2016), interferometric surveys suffer a degradation in their nominal sensitivity for nearby objects that can be resolved over many synthesized beams (suppressing the signal to noise like the square root of the number of beams). This is seldom a problem for single dish telescopes as even the largest dishes have beams that a few arcmin across at 21 cm wavelengths. Furthermore, if outfitted with multibeam receivers capable of forming tens of beams on the sky, the survey mapping speed of the largest existing single dish telescopes could effectively match that of the SKA-pathfinder interferometers (at equivalent nominal sensitivity), making them ideal instruments to detect the very lowest H I mass, and most diffuse, galaxies in the nearby Universe and address the question of what is the threshold mass to form a galaxy.

7 CONCLUSIONS

We have measured the HIMF of the full ALFALFA survey, confirming it is well fit by a Schechter function with parameters $\alpha = -1.25 \pm 0.02 \pm 0.03$, $m_* = 9.94 \pm 0.01 \pm 0.04$, and $\phi_* = (4.5 \pm 0.2 \pm 0.4) \times 10^{-3}\,\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$, where the first error quoted is the estimated random uncertainty due to Poisson counting errors, flux measurement errors, and peculiar velocities, the latter is the estimate of the systematic uncertainty due to boundary effects, sample variance, choice of flow model, the absolute flux scale, and bias inherent to the v_{eff} (or 2DSWML) method. This is the most precise measurement of the HIMF, and the most complete accounting of its uncertainties, to date.

If estimates of the impact of cosmic variance are also included then there is little impact on the systematic uncertainties in the 'knee' mass, which rise to 0.05 dex, but the low-mass slope becomes highly uncertain, with an estimated systematic error of ~ 0.1. This indicates that ALFALFA has made a cosmologically fair measurement of the HIMF 'knee' mass, but although it has a very precise measurement of the low-mass slope, its value only reflects that of the local Universe. The global H I content of the $z = 0$ Universe (corrected for self-absorption) was measured to be $\Omega_{\mathrm{H\,I}} = (3.9 \pm 0.1 \pm 0.6) \times 10^{-4}$, with the large systematic uncertainty arising from the (conservative) estimate of the uncertainty in the absolute flux calibration.

We also investigated differences in the ALFALFA HIMF between the Arecibo Spring and Fall skies, and the local ($v < 4000\,\mathrm{km\,s^{-1}}$) and full volumes ($v < 15000\,\mathrm{km\,s^{-1}}$). We find a clear dichotomy in the low-mass slope of the Spring and Fall skies that was missed by previous studies of the environmental dependence of the ALFALFA population because they focused on the Spring sky only or smaller scale definitions of environment.

The Spring sky is dominated by the Virgo cluster, a very overdense region, whereas, at a similar distance, the Fall sky contains a deep void. This coincident shift in environment and low-mass slope strongly suggests the two are connected. Furthermore, the steepening of the slope appears to be associated with the direction of the Virgo cluster, but the cluster itself does not appear to be the dominant component driving a steeper slope. We therefore hypothesise that the steeper slope may be associated with the gas-rich filaments that are connected to Virgo and feeding the growth of the cluster, or that the flattening of the slope away from the cluster may be due to the impact of low-density environments. To test these hypotheses

will require much deeper surveys than ALFALFA that would be capable of detecting equivalent phenomena, should they exist, in and around other clusters and voids.

The existence of shifts in α on large angular scales also indicates that caution should be exercised when comparing the results of studies of individual galaxy groups to wide-field surveys, as the H I content of their constituent galaxies may be impacted by the group's presence within a larger structure as well as membership of the group itself. We have also discussed several methodological considerations that are relevant to such a comparison, and note that studies of groups detect almost exclusively satellites, while wide-area blind H I surveys consist almost entirely of field centrals.

The other shape parameter of the HIMF, 'knee' mass, appears relatively independent of direction on the sky, indicating that the ALFALFA volume is deep enough to give a cosmologically fair representation of this parameter at $z = 0$. It does, however, appear to have a somewhat lower value in the local Universe (compared to the whole ALFALFA sample), which conveniently explains the lower value found by HIPASS (Zwaan et al. 2005), as HIPASS was a shallower survey than ALFALFA.

Upcoming surveys with SKA pathfinders will be capable of addressing these unresolved issues and will create a much tighter constraint on the H I population as a function of environment, and eventually, redshift. As simulations continue to include more and finer detail gas physics, these surveys will be vital to pointing those efforts in the correct direction. However, the largest existing single dish radio-telescopes may turn out to be the best suited instruments to addressing questions concerning the lowest mass and surface density galaxies in the Universe.

ACKNOWLEDGEMENTS

REFERENCES

Abolfathi B. et al., 2017, ApJS, preprint (arXiv:1707.09322)
Adams E. A. K., Giovanelli R., Haynes M. P., 2013, ApJ, 768, 77
Adams E. A. K., Oosterloo T. A., Cannon J. M., Giovanelli R., Haynes M. P., 2016, A&A, 596, A117
Aihara H. et al., 2011, ApJS, 193, 29
Barnes D. G. et al., 2001, MNRAS, 322, 486
Binggeli B., Sandage A., Tammann G. A., 1985, AJ, 90, 1681
Boselli A., Gavazzi G., 2006, PASP, 118, 517
Braun R., 2012, ApJ, 749, 87
Braun R., Thilker D. A., Walterbos R. A. M., Corbelli E., 2009, ApJ, 695, 937
Brown T. et al., 2017, MNRAS, 466, 1275
Chung A., van Gorkom J. H., Kenney J. D. P., Crowl H., Vollmer B., 2009, AJ, 138, 1741
Crain R. A. et al., 2015, MNRAS, 450, 1937
Crook A. C., Huchra J. P., Martimbeau N., Masters K. L., Jarrett T., Macri L. M., 2007, ApJ, 655, 790
Davies J. et al., 2004, MNRAS, 349, 922
Davies J. I. et al., 2011, MNRAS, 415, 1883
Driver S. P., Robotham A. S. G., 2010, MNRAS, 407, 2131
Duffy A. R., Meyer M. J., Staveley-Smith L., Bernyk M., Croton D. J., Koribalski B. S., Gerstmann D., Westerlund S., 2012, MNRAS, 426, 3385
Efstathiou G., Ellis R. S., Peterson B. A., 1988, MNRAS, 232, 431
Elson E. C., Blyth S. L., Baker A. J., 2016, MNRAS, 460, 4366
Fernández X. et al., 2013, ApJ, 770, L29
Freeland E., Stilp A., Wilcots E., 2009, AJ, 138, 295
Gavazzi G., Boselli A., van Driel W., O'Neil K., 2005, A&A, 429, 439
Giovanelli R., Haynes M. P., 1985, ApJ, 292, 404
Giovanelli R., Haynes M. P., 2016, A&AR, 24, 1
Giovanelli R., Haynes M. P., Salzer J. J., Wegner G., da Costa L. N., Freudling W., 1994, AJ, 107, 2036
Giovanelli R. et al., 2005, AJ, 130, 2598
Giuricin G., Marinoni C., Ceriani L., Pisani A., 2000, ApJ, 543, 178
Guo H., Li C., Zheng Z., Mo H. J., Jing Y. P., Zu Y., Lim S. H., Xu H., 2017, ApJ, 846, 61
Hallenbeck G. et al., 2012, AJ, 144, 87
Haynes M. P. et al., 2011, AJ, 142, 170
Hess K. M., Wilcots E. M., 2013, AJ, 146, 124
Holwerda B. W., Blyth S.-L., Baker A. J., 2012, in Tuffs R. J., Popescu C. C., eds, Proc. IAU Symp. 284, The Spectral Energy Distribution of Galaxies. Cambridge Univ. Press, Cambridge, p. 496
Hoppmann L., Staveley-Smith L., Freudling W., Zwaan M. A., Minchin R. F., Calabretta M. R., 2015, MNRAS, 452, 3726
Hoyle F., Rojas R. R., Vogeley M. S., Brinkmann J., 2005, ApJ, 620, 618
Huchra J. P. et al., 2012, ApJS, 199, 26
Janowiecki S., Catinella B., Cortese L., Saintonge A., Brown T., Wang J., 2017, MNRAS, 466, 4795
Jones M. G., Papastergis E., Haynes M. P., Giovanelli R., 2015, MNRAS, 449, 1856
Jones M. G., Papastergis E., Haynes M. P., Giovanelli R., 2016, MNRAS, 457, 4393
Kenney J. D. P., van Gorkom J. H., Vollmer B., 2004, AJ, 127, 3361
Kent B. R., 2008, PhD thesis, Cornell Univ.
Kilborn V. A., Forbes D. A., Barnes D. G., Koribalski B. S., Brough S., Kern K., 2009, MNRAS, 400, 1962
Kovac K., Oosterloo T. A., van der Hulst J. M., 2005, in Jerjen H., Binggeli B., eds, IAU Colloq. 198: Near-fields Cosmology with Dwarf Elliptical Galaxies. Astron. Soc. Pac., San Francisco, p. 351
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 Univ.
Masters K. L., Haynes M. P., Giovanelli R., 2004, ApJ, 607, L115
Mei S. et al., 2007, ApJ, 655, 144
Meyer M., 2009, in Heald G., Serra P., eds, Proceedings of Panoramic Radio Astronomy: Wide-field 1-2 GHz Research on Galaxy Evolution. Groningen, the Netherlands, p. 15
Meyer M. J. et al., 2004, MNRAS, 350, 1195
Minchin R. F., 2017, ApJ, 847, 49
Moorman C. M., Vogeley M. S., Hoyle F., Pan D. C., Haynes M. P., Giovanelli R., 2014, MNRAS, 444, 3559
Moorman C. M., Vogeley M. S., Hoyle F., Pan D. C., Haynes M. P., Giovanelli R., 2015, ApJ, 810, 108
Mould J. R. et al., 2000, ApJ, 529, 786
Odekon M. C. et al., 2016, ApJ, 824, 110
Papastergis E., 2013, PhD thesis, Cornell Univ.
Papastergis E., Adams E. A. K., Romanowsky A. J., 2017, A&A, 601, L10
Pisano D. J., Barnes D. G., Staveley-Smith L., Gibson B. K., Kilborn V. A., Freeman K. C., 2011, ApJS, 197, 28

Rosenberg J. L., Schneider S. E., 2002, ApJ, 567, 247

Saintonge A., 2007, AJ, 133, 2087

Schechter P., 1976, ApJ, 203, 297

Solanes J. M., Manrique A., García-Gómez C., González-Casado G., Giovanelli R., Haynes M. P., 2001, ApJ, 548, 97

Solanes J. M., Sanchis T., Salvador-Solé E., Giovanelli R., Haynes M. P., 2002, AJ, 124, 2440

Springob C. M., 2006, PhD thesis, Cornell Univ.

Springob C. M., Haynes M. P., Giovanelli R., 2005, ApJ, 621, 215

Stierwalt S., Haynes M. P., Giovanelli R., Kent B. R., Martin A. M., Saintonge A., Karachentsev I. D., Karachentseva V. E., 2009, AJ, 138, 338

Taylor R., Davies J. I., Auld R., Minchin R. F., 2012, MNRAS, 423, 787

Taylor R., Davies J. I., Auld R., Minchin R. F., Smith R., 2013, MNRAS, 428, 459

van Zee L., Maddalena R. J., Haynes M. P., Hogg D. E., Roberts M. S., 1997, AJ, 113, 1638

Verheijen M. A. W., Trentham N., Tully B., Zwaan M., 2001, in Hibbard J. E., Rupen M., van Gorkom J. H., eds, ASP Conf. Ser. Vol. 240, Gas and Galaxy Evolution. Astron. Soc. Pac., San Francisco, p. 507

Vogelsberger M. et al., 2014, MNRAS, 444, 1518

Westmeier T. et al., 2017, MNRAS, 472, 4832

Wong O. I. et al., 2006, MNRAS, 371, 1855

Zwaan M. A. et al., 2003, AJ, 125, 2842

Zwaan M. A. et al., 2004, MNRAS, 350, 1210

Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

APPENDIX A: ALFALFA AND HIPASS VOLUME CALCULATIONS

The calculations in the introduction showing that the accessible volume for ALFALFA is larger for both the low-mass slope and the 'knee' mass are somewhat counter intuitive because the total volume of HIPASS is larger than ALFALFA, and HIPASS continues to detect sources almost to the edge of its bandwidth. When comparing to Zwaan et al. (2005) (their fig. 2), it can be seen that M_* galaxies are detected beyond 75 Mpc, the maximum detection distance calculated for galaxies of $\log M_{\rm HI}/M_{\odot} = 10$ and velocity width of 300 km s^{-1} using equation (4) of Zwaan et al. (2004). However, these detections generally have narrower velocity widths than is typical for such galaxies, which increases their peak flux, making them detectable out to greater distances. The same would be true for ALFALFA, which would cause its volume to grow at a faster rate than HIPASS's (as volume grows like r^3 and the detection distance is greater for ALFALFA), but the outer edge of ALFALFA is determined by a band of heavy Radio Frequency Interference (RFI), which means the volume cannot grow much greater. This creates an odd situation where which survey has a greater volume depends on the mass and the velocity width of the galaxies in question. For galaxies of $\log M_{\rm HI}/M_{\odot} = 10$ with velocity widths greater than ~ 100 km s^{-1}, ALFALFA has the larger volume, but for narrower sources, HIPASS has the larger volume. As most M_* galaxies have broader velocity widths than this, we conclude that ALFALFA has the larger volume for probing the 'knee' mass.

Essentially the same point can also be made in a simpler way, without explicit reference to the survey detection limits, but in doing so hides the details of what is happening. The median distance to an HIPASS M_* galaxy is approximately 50 Mpc, whereas for ALFALFA it is approximately 150 Mpc. Therefore, if the survey areas were the same, the volume accessible in ALFALFA would be 27 times larger for the typical M_* galaxy. As the survey area of HIPASS is less than 27 times that of ALFALFA, the accessible volume for M_* galaxies must be larger for ALFALFA.

For the low-mass slope, the situation is much simpler because these sources in both HIPASS and ALFALFA are detected well away from the outer edge of the bandwidth and the band of RFI that truncates ALFALFA. Thus, lowering the velocity width used to calculate the maximum detection distance just causes the accessible ALFALFA volume to grow faster than the accessible HIPASS volume, as would be naively expected.

APPENDIX B: ALTERNATIVE DISTANCE ESTIMATES

One criticism of this work is that all masses are reliant on distances, which are not measured directly by ALFALFA. Our flow model (Masters 2005) makes corrections to the velocities of nearby objects and then uses Hubble's law to calculate distance estimates. There are also additional corrections for galaxies assigned to groups or the Virgo cluster (described in Section 4.1), or those which have primary or secondary distance measurements in the literature. If the distances that this model produced are systematically bias then some of our finding could be due, in part, to this bias.

To address this issue, we have recalculated the distances to all ALFALFA sources using two alternative methods. The first is our own implementation of the Mould et al. (2000) flow model that included simple spherical infall around Virgo, Shapley, and the Great Attractor. The second is to simply use the CMB velocity of the sources with no corrections at all. The resulting HIMFs for these two distance estimation schemes are shown in Fig. B1. The HIMFs are almost indistinguishable from each other and the fit parameters are consistent within 1σ errors of those of the HIMF based on ALFALFA's flow model.

It is also noticeable that the CMB distances lead to many more nearby sources being excluded (the bottom panel of Fig. B1). Part of the reason for this is because it is not uncommon for nearby sources with large peculiar velocities (often in the direction of Virgo) to have negative CMB velocities. This means they are immediately excluded. The Mould et al. (2000) has the opposite effect and overestimates the number of very low mass sources (relative to ALFALFA's flow model), which indicates it is calculating smaller distances for some of the nearest sources. Despite this difference, the sample with CMB distances retains more objects overall. This
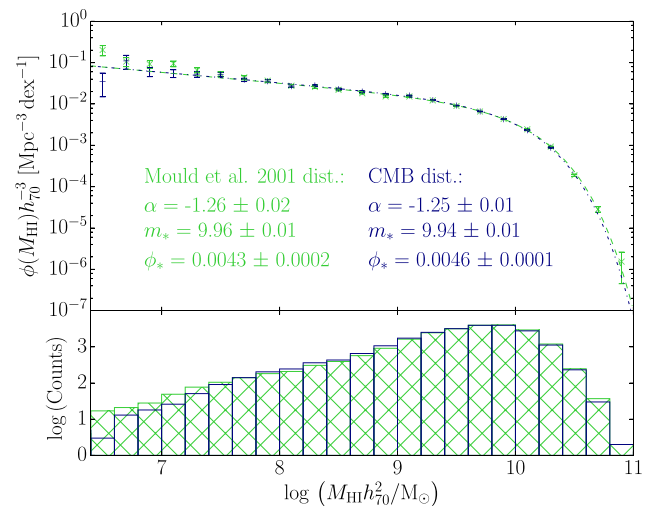

Figure B1. The HIMFs of the $\alpha.100$ sample for Mould et al. (2000) method distances (light green crosses and cross hatched bars) and distances calculated directly from the CMB velocities (dark blue pluses and unfilled bars).

Table B1. As in Table 2, but for samples using the alternative distance estimates.

Sample	$v_{max}/\mathrm{km\,s^{-1}}$	N_{gal}	α	m_*	$\phi_*/\mathrm{Mpc^{-3}\,dex^{-1}}$
$\alpha.100$ (Hubble)	15000	22815	-1.25 ± 0.01	9.94 ± 0.01	$4.6 \pm 0.1 \times 10^{-3}$
$\alpha.100$ Spring (Hubble)	15000	14392	-1.27 ± 0.01	9.94 ± 0.01	$5.0 \pm 0.1 \times 10^{-3}$
$\alpha.100$ Fall (Hubble)	15000	8423	-1.23 ± 0.01	9.94 ± 0.01	$3.9 \pm 0.1 \times 10^{-3}$
$\alpha.100a$ (Hubble)	4000	3804	-1.22 ± 0.02	9.74 ± 0.04	$6.4 \pm 0.6 \times 10^{-3}$
$\alpha.100$ (Mould et al.)	15000	22693	-1.26 ± 0.02	9.96 ± 0.01	$4.3 \pm 0.2 \times 10^{-3}$
$\alpha.100$ Spring (Mould et al.)	15000	14256	-1.30 ± 0.02	9.97 ± 0.01	$4.5 \pm 0.2 \times 10^{-3}$
$\alpha.100$ Fall (Mould et al.)	15000	8437	-1.14 ± 0.02	9.92 ± 0.02	$4.4 \pm 0.2 \times 10^{-3}$
$\alpha.100a$ (Mould et al.)	4000	3472	-1.29 ± 0.02	9.85 ± 0.05	$4.3 \pm 0.5 \times 10^{-3}$

is because at large distances, the Mould et al. (2000) model slightly overestimates the velocities of sources and therefore places them at larger distances than the pure Hubble flow model, which leads to more being cut by the maximum distance limit that we set (v_{max}/H_0). This difference is not immediately apparent from Fig. B1 due to the logarithmic scale, but the counts in the intermediate-mass bins are systematically higher for the Hubble flow sample than the Mould et al. (2000) flow model sample.

The fit parameters for the full list of sub-samples is shown in Table B1.

APPENDIX C: WIDTH CORRECTION FACTOR

In the calculation of the $\alpha.40$ HIMF (Martin et al. 2010), a correction for sources that are undetected due to their velocity widths being too broad was applied; however, we make no correction here because we do not set a minimum distance cut on the sample. The reason why the minimum distance limit is the determining factor is explained below by means of an example.

Imagine that at a given distance D_1 a galaxy of H I mass M_1 is above the completeness limit only if it has a narrow velocity width, but is undetectable if it has a broad velocity width. In the case of a large area survey, there will always be (except for the most extremely nearby sources) nearer sources of the same mass, which, because they are nearer, will be detectable even if they have very broad velocity widths. This means that the 2DSWML algorithm can correctly approximate the form of the mass-width function (the 2D distribution of intrinsic H I masses and velocity widths). Therefore, those sources of mass M_1 that were undetected at D_1 due to their large velocity widths are automatically accounted for by the estimates of v_{eff} for those sources of mass M_1 that were detected.

However, in the case where a minimum distance limit is set, this line of reasoning no longer applies because there are no galaxies in the foreground to act to correct the v_{eff} estimates. Note that this also applies for the v_{max} method, with the additional point that this method does not account for the problem in the first place. As a group is, by definition, confined to a small region, there is always a minimum distance limit applied and thus a width correction factor is necessary. This correction should take the form of an up-weighting of each source based on the fraction of sources of that mass that would be undetectable at each distance. This requires the mass-conditional width function to be measured or assumed for all masses.

When a lower distance limit is set and a correction is not applied, we have invariably found that it acts to flatten the low-mass slope by suppressing the lowest mass bins. This effect is just visible in the extended Virgo cluster HIMF plotted in Fig. 5. This was not corrected for, as it only appears to strongly affect the first bin

(probably due to the proximity of Virgo) and its value is highly uncertain anyway.

APPENDIX D: BOUNDARY VERTICES

The Tables D1, D2, D3, and D4 show the coordinates of the boundary vertices of the Spring sample, the Fall sample, the Spring strict sample, and the Fall strict sample, respectively.

Table D1. Vertices of the boundary of the Spring sample.

Dec. range (deg)	RA min (h])	RA max (h])
0–16	7.7	16.5
16–18	7.7	16.0
18–20	8.7	15.4
20–24	9.4	15.4
24–30	7.6	16.5
30–32	8.5	16.0
32–36	9.5	15.5

Table D2. Vertices of the boundary of the Fall sample.

Dec. range (deg)	RA min (h])	RA max (h)
0–2	22.0	3.0
2–6	22.5	3.0
6–10	22.0	3.0
10–14	22.0	2.5
14–36	22.0	3.0

Table D3. Vertices of the boundary of the Spring strict sample.

Dec. range (deg)	RA min (h)	RA max (h)
0.5–15.5	8.0	16.0
15.5–24.5	10.0	15.0
24.5–29.5	8.0	16.0
29.5–35.5	10.0	15.0

Table D4. Vertices of the boundary of the Fall strict sample.

Dec. range (deg)	RA min (h)	RA max (h)
0.5–35.5	22.5	2.5

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