Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **522,** 4043–4058 (2023) https://doi.org/10.1093/mnras/stad1191
Advance Access publication 2023 April 17

The north–south asymmetry of the ALFALFA H I velocity width function

Richard A. N. Brooks [1,2,3]★ Kyle A. Oman [2,3] and Carlos S. Frenk[2,3]

[1]*Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*
[2]*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[3]*Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

Accepted 2023 April 18. Received 2023 April 13; in original form 2022 November 11

ABSTRACT

The number density of extragalactic 21-cm radio sources as a function of their spectral line widths – the H I width function (H I WF) – is a sensitive tracer of the dark matter halo mass function (HMF). The Lambda cold dark matter model predicts that the HMF should be identical everywhere provided it is sampled in sufficiently large volumes, implying that the same should be true of the H I WF. The Arecibo Legacy Fast ALFA (ALFALFA) 21-cm survey measured the H I WF in northern and southern Galactic fields and found a systematically higher number density in the north. At face value, this is in tension with theoretical predictions. We use the *Sibelius-DARK N*-body simulation and the semi-analytical galaxy formation model *GALFORM* to create a mock ALFALFA survey. We find that the offset in number density has two origins: the sensitivity of the survey is different in the two fields, which has not been correctly accounted for in previous measurements; and the $1/V_{\rm eff}$ algorithm used for completeness corrections does not fully account for biases arising from spatial clustering in the galaxy distribution. The latter is primarily driven by a foreground overdensity in the northern field within 30 Mpc , but more distant structure also plays a role. We provide updated measurements of the ALFALFA H I WF (and H I mass function) correcting for the variations in survey sensitivity. Only when systematic effects such as these are understood and corrected for can cosmological models be tested against the H I WF.

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

1 INTRODUCTION

The standard Lambda cold dark matter (ΛCDM) cosmological model predicts that the number density of self-bound dark matter haloes as a function of mass – the halo mass function (HMF; Frenk et al. 1988) – is well approximated by a power law with slope of $\phi(M) \propto M^{-1.9}$ over almost 20 orders of magnitude below the scale of the largest collapsed structures today (Wang et al. 2020). The low-mass end of the HMF is sensitive to the power spectrum of density fluctuations in the early Universe. For instance, for thermal relic dark matter particles that are lighter than those assumed in ΛCDM models (e.g. warm dark matter (WDM); Avila-Reese et al. 2001; Bode, Ostriker & Turok 2001), there is a cut-off in the power spectrum at smaller scales. The higher velocities of lighter thermal relics naturally suppress the formation of low-mass haloes because free-streaming effects erase the density perturbations that could seed them. Measuring the HMF therefore offers an opportunity to constrain the particle nature of dark matter. However, the HMF is not directly measurable and hence indirect measurements of the HMF must be made instead. One option is to study the abundances of galaxies as a function of the kinematics of visible tracers orbiting within their dark matter haloes.

There is an apparent tension between the number of low-mass dark matter haloes predicted by ΛCDM dark matter-only simulations in which galaxies are expected to form and the number of galaxies with kinematics compatible with inhabiting them measured from observations. This problem was first identified for satellite galaxies of the Milky Way ('too-big-to-fail'; Boylan-Kolchin, Bullock & Kaplinghat 2011) and subsequently resolved by Sawala et al. (2016), who showed that the inclusion of baryons in the simulations changes the theoretical predictions; the inevitable ejection of baryons from haloes at early times (due to supernova 'feedback') reduces the growth rate of haloes such that, at the present time, a halo in the full-physics simulation is about 10 per cent less massive than its counterpart in the dark matter-only simulation. A similar issue applying to field galaxies was identified by Papastergis et al. (2015), and it is so far unclear whether this issue is similarly resolved. In this instance, dwarf galaxies should be hosted by dark matter haloes that are significantly more massive than those implied by measuring the kinematics within the galaxies. If instead lower mass dark matter haloes host dwarf galaxies, observational surveys should measure a much higher number density of galaxies.

The kinematics of observable tracers in a galaxy are linked to the HMF because the maximum circular velocity of a halo, $v_{\rm max}$, is correlated to its mass (Navarro, Frenk & White 1997). Connecting a kinematic tracer to $v_{\rm max}$ often requires additional modelling, e.g. in the case of the 21-cm spectral line width. The 21-cm velocity spectrum is the H I mass-weighted line-of-sight velocity distribution. The 21-cm line width can be parametrized as the full width at half-maximum of the spectrum, w_{50}. If a dark matter halo contains a sufficiently extended H I disc, the maximum circular velocity of the dark matter halo is approximately $v_{\rm max} \sim w_{50}/2\sin(i)$, where

★ E-mail: richard.brooks.22@ucl.ac.uk

i is the inclination angle to the line of sight. However, there are considerations that need to be made when computing $v_{\rm max}$ using this approach. In many cases galaxies will have gas discs that are not sufficiently extended to reach the flat part of the velocity profile (Brook & Di Cintio 2015; Brook & Shankar 2016; Macciò et al. 2016; Ponomareva, Verheijen & Bosma 2016; Brooks et al. 2017). Additionally, the inclination angle is problematic because line-width surveys usually do not spatially resolve the gas structure and therefore rely on optical counterparts of radio sources to get an estimate for i (e.g. Zwaan, Meyer & Staveley-Smith 2010). There are many other difficulties faced when obtaining $v_{\rm max}$ using line-width measurements, such as: the gas orbits may not be circular (Brook & Shankar 2016); the gas disc may not lie in a single plane (Ponomareva et al. 2016); emission may be mistakenly taken to overlap with a neighbouring source (Jones et al. 2015; Chauhan et al. 2019); the gas disc may be partially supported by turbulent or thermal pressure (Brook & Di Cintio 2015; Ponomareva et al. 2016); etc. Instead of attempting to infer $v_{\rm max}$ from line-width measurements, a potentially more straightforward technique is to predict the number density of extragalactic 21-cm sources as a function of w_{50} – the H I width function (H I WF) – and then compare with observational measurements. Comparison of the H I WF from theory to an observational equivalent thus provides an alternative avenue to investigate dark matter than to try and infer an HMF from an observed H I WF.

For the Arecibo Legacy Fast ALFA[1] (ALFALFA; Giovanelli et al. 2005) survey, the H I WF and H I mass function (H I MF) are measured simultaneously as orthogonal integrations of the two-dimensional H I mass-width function. Jones et al. (2018) reported that the H I MF has significantly different global shapes in the 'spring' (northern Galactic hemisphere) and 'fall' (southern Galactic hemisphere) fields of the survey. The low-mass slope is significantly shallower in the fall field. The large-scale environmental dependence of the low-mass slope in the H I MF was tentatively attributed to the presence of the Virgo cluster in the foreground of one half of the survey and a deep void in the other (Jones et al. 2018). The H I WF, on the other hand, has been reported to have a similar shape in both survey fields. Oman (2022) tentatively attributed this similarity in shape to possible environmental effects affecting the shape of the H I MF, but leaving the shape of the H I WF largely unchanged. However, they made no attempt to explain the different overall normalization of the H I MFs and H I WFs in the two regions.

The first measurements of the H I WF came almost simultaneously from two different 21-cm surveys. Zavala et al. (2009) used an early release of the ALFALFA survey with only 6 per cent of the final data available. Meanwhile, Zwaan et al. (2010) used the H I Parkes All-Sky Survey (H I PASS). Both of these measurements revealed that the ΛCDM model apparently overpredicts the abundance of sources at the low-velocity-width end. The same overabundance problem persists in follow-up work after subsequent ALFALFA data releases (Papastergis et al. 2011; Jones et al. 2018), but can be explained (at least to leading order) by the systematic reduction of the total mass and abundance of structures below $v_{\rm max} \sim 100\,{\rm km\,s^{-1}}$ by gas pressure, reionization, supernova feedback, stripping, and truncated accretion (Sawala et al. 2013; see also Macciò et al. 2016; Papastergis & Shankar 2016; Brooks et al. 2017; Dutton, Obreja & Macciò 2019).

Zavala et al. (2009) and Zwaan et al. (2010) both measured differences in the normalization of the H I WF in different subsets of the ALFALFA and H I PASS surveys, respectively, although their shapes are statistically consistent with being identical. It has been variously speculated that what drives the difference in the normalization of the H I WF between the ALFALFA spring and fall fields is a combination of sample variance, distance modelling, the adopted completeness limit (CL), etc. (see Oman 2022, for further discussion).

In this work, we use the *Sibelius-DARK N*-body simulation (McAlpine et al. 2022) that reproduces the local structure of the Universe on scales larger than ≈4 Mpc populated with galaxies using the *GALFORM* semi-analytical model (Lacey et al. 2016) to create mock surveys similar to the ALFALFA survey. The nature of the *Sibelius* simulations allows us to investigate the influence of the spatial clustering of galaxies along the line of sight, as well as possible environmental effects on the global shape of the H I WF. Previous work has only been able to comment speculatively on the origin of the differences in the H I WF between the two fields surveyed by ALFALFA (Jones et al. 2018; Oman 2022). Our mock surveys provide a suitable footing for an investigation into survey systematics that may be responsible for driving the asymmetry between the spring and fall fields; our approach enables us to provide the first quantitative estimates for the magnitude of these effects and assess whether they can explain the observed asymmetry.

The plan of this paper is as follows. In Section 2 we outline the process we use to construct and analyse mock ALFALFA-like 21-cm surveys. In Section 3 we present the ALFALFA and *Sibelius-DARK* plus *GALFORM* H I WFs under various conditions and assumptions. In Section 4 we comment on possible mitigations for variations in sensitivity for future surveys and on the possible origins of qualitative differences between our mock H I MFs and H I WFs and those measured using ALFALFA. We summarize in Section 5.

2 METHODS

In Section 2.1 we summarize the defining properties of the ALFALFA survey, detail how extragalactic sources were identified, and show how the final survey catalogue is obtained. Next, in Section 2.2 we provide an overview of the *N*-body *Sibelius-DARK* simulation and assess the assumption of identical HMFs in both fields. In Section 2.3 we give an overview of the *GALFORM* semi-analytical model. In Section 2.4 we explain how our *GALFORM* mock 21-cm survey is created. Finally, in Section 2.5, we outline the $1/V_{\rm eff}$ statistical estimator used to correct the H I WF for observational incompleteness.

2.1 The ALFALFA survey

The ALFALFA survey (Giovanelli et al. 2005) mapped ∼7000 deg^2 of the sky visible from Arecibo at 21-cm wavelengths out to ∼250 Mpc, or $cz \leq 18\,000\,{\rm km\,s^{-1}}$. ALFALFA was specifically designed to investigate the faint end of the H I MF in the Local Universe. The survey was completed in 2012, and is composed of two separate fields on the sky: one in the northern Galactic hemisphere, visible during the spring, and the other in the southern Galactic hemisphere, visible during the autumn. By convention, these fields are labelled 'spring' and 'fall', respectively (Jones et al. 2018).

Extragalactic sources in the ALFALFA survey were identified using a matched-filtering technique (Saintonge 2007), supplemented with some sources identified from direct inspection of the raw data cubes. These identified sources were subsequently manually checked to confirm or reject each individual detection and to assign optical counterparts to detections where possible. The final product

[1] Arecibo L-band Feed Array.

Table 1. Characteristics of the $\alpha.100$ ALFALFA and mock *Sibelius-DARK* + *GALFORM* survey catalogues. Additionally shown is the information of the *GALFORM* galaxy population before any selection criteria are applied. The assumed distance limit is 200 Mpc. The areas of the spring and fall fields are 1.240 and 0.752 sr, respectively.

	Characteristics of galaxy catalogues – separate CL		
	ALFALFA survey	*GALFORM* survey	*GALFORM* before CL
Number of spring sources, N_s	13 006	12 408	2074 665
Number of fall sources, N_f	7851	8623	1367 320
Number density of spring sources, n_s	0.0040 Mpc^{-3}	0.0038 Mpc^{-3}	0.63 Mpc^{-3}
Number density of fall sources, n_f	0.0039 Mpc^{-3}	0.0044 Mpc^{-3}	0.69 Mpc^{-3}

is the $\alpha.100$ extragalactic source catalogue described in Haynes et al. (2018). This catalogue lists the coordinates (for the H I and associated optical sources), redshifts, 21-cm line flux densities, 21-cm line widths, distances,[2] signal-to-noise ratios and H I masses of sources, and their uncertainties where relevant. The H I mass, $M_{\rm H\,I}$, of a galaxy is determined as usual from the flux and distance as

$$\frac{M_{\rm H\,I}}{\rm M_\odot} = 2.36 \times 10^5 \left(\frac{D}{\rm Mpc}\right)^2 \frac{S_{21}}{\rm Jy\,km\,s^{-1}}. \quad (1)$$

We define a selection of $\alpha.100$ sources from which we measure the H I WF in this work in a similar way to Oman (2022). This includes the choice of only 'Code 1' (i.e. S/N > 6.5) sources whose (RA, Dec.) coordinates fall in the survey footprint[3] (Jones et al. 2018). Instead of the recessional velocity cut $v_{\rm rec} \leq 15\,000$ km s^{-1} used in Oman (2022), we impose a distance cut $d_{\rm mw} \leq 200$ Mpc in order to facilitate comparison with the *Sibelius-DARK* simulation that is contaminated by low-resolution particles from outside of the zoom-in region (McAlpine et al. 2022) beyond this distance. Only sources above the 50 per cent CL of the survey are selected. The determination of the CL for the ALFALFA survey is described in Haynes et al. (2011). For a flux-limited sample drawn from a uniformly distributed population of galaxies, number counts as a function of flux are expected to follow a power law with exponent −3/2. Deviation from this form indicates the onset of incompleteness in the survey. There are 20 857 sources above the 50 per cent CL in the $\alpha.100$ catalogue, of which 13 006 are in the spring field and 7851 are in the fall field (see Table 1). Fig. 1 visualizes the $\alpha.100$ sources within the survey footprint on the sky (upper panels) and in a cone diagram over all declinations (lower panel). Oman (2022) used the global 50 per cent CL that they derived throughout their analysis. We adopt the same CL in some contexts, but also make use of the 50 per cent CL derived separately from each of the two survey fields. The fall CL is slightly shallower than the global CL, by 0.011 dex, while the spring CL is slightly deeper, by 0.009 dex, for a net difference of 0.02 dex. Explicitly,

$$\text{Spring}: \log_{10}\left(\frac{S_{21,50\ \text{per cent}}}{\rm Jy\,km\,s^{-1}}\right) = \begin{cases} 0.5\,W - 1.179 & W < 2.5, \\ W - 2.429 & W \geq 2.5 \end{cases}; \quad (2)$$

$$\text{Fall}: \log_{10}\left(\frac{S_{21,50\ \text{per cent}}}{\rm Jy\,km\,s^{-1}}\right) = \begin{cases} 0.5\,W - 1.159 & W < 2.5, \\ W - 2.409 & W \geq 2.5 \end{cases}, \quad (3)$$

where $W = \log_{10}(w_{50}/{\rm km\,s^{-1}})$. The CLs at other completeness levels (25 per cent, 90 per cent) are given in Appendix A.

[2]Ball et al. (2022) provide independent confirmation that the catalogue distances derived using the Masters (2005) flow model are accurate.
[3]See their tables D1–D4. We adopt the fiducial, not the 'strict' footprint throughout this work.

2.2 *Sibelius-DARK*

The 'Simulations Beyond The Local Universe' (*Sibelius*; Sawala et al. 2022) project aims to connect the Local Group with its cosmological environment. *Sibelius* simulations use ΛCDM initial conditions that are constrained such that the large-scale structure is accurately reproduced; e.g. well-known galaxy clusters such as Virgo, Coma, and Perseus are embedded within the correct large-scale cosmic web, and have appropriate masses. The initial conditions are generated using the *BORG* algorithm ('Bayesian Origin Reconstruction from Galaxies': Jasche & Wandelt 2013) that derives initial conditions using Bayesian inference through forward modelling with Hamiltonian Monte Carlo methods, inferred in this instance from the 2M++ galaxy sample (Lavaux & Hudson 2011). The inference algorithm is fully probabilistic in the sense that it turns the task of reproducing the present non-linear galaxy distribution into a statistical initial conditions problem.

The first simulation from the *Sibelius* project is *Sibelius-DARK* (McAlpine et al. 2022), a realization of a volume constrained within 200 Mpc of the Milky Way. Its volume makes it ideal to compare with the ALFALFA survey, which detects galaxies out to slightly beyond 200 Mpc.

The simulation assumes a flat CDM cosmology with parameters from the Planck Collaboration I (2014): $\Omega_\Lambda = 0.693$, $\Omega_{\rm m} = 0.307$, $\Omega_{\rm b} = 0.048\,25$, $\sigma_8 = 0.8288$, $n_{\rm s} = 0.9611$, and $H_0 = 67.77$ km s^{-1} Mpc^{-3}. We note that Haynes et al. (2018) assume a slightly different cosmology that will cause ∼10 per cent differences in the normalization of the ALFALFA and our mock H I MFs and H I WFs. This is driven by the assumed value of the reduced Hubble constant, h, which when incorporated into the units of the H I MF and H I WF results in a different normalization. Below, we restrict ourselves to qualitative comparisons between ALFALFA and *GALFORM*, so this small quantitative difference does not influence our conclusions.

2.2.1 *The halo mass function of Sibelius-DARK*

In the top panel of Fig. 2, we show the HMF for the spring and fall fields as well as for the median across eight octants of the *Sibelius-DARK* sky (the spring and fall regions have volumes equivalent to 80 and 50 per cent of an octant, respectively; all octants and their properties are detailed in Appendix B). The middle panel of Fig. 2 shows the ratio of the N-th octant's HMF to that of the median. The scatter in the HMF across the octants on the sky is ≲20 per cent around the median.

We quantify how closely we should expect the HMFs in regions of these volumes to agree as follows. Hu & Kravtsov (2003) provide a relation for the fractional variance σ_n in the number density of haloes above a given mass threshold $M_{\rm th}$ in their fig. 2. We can approximate the low-mass end of their sample variance relation (sample variance

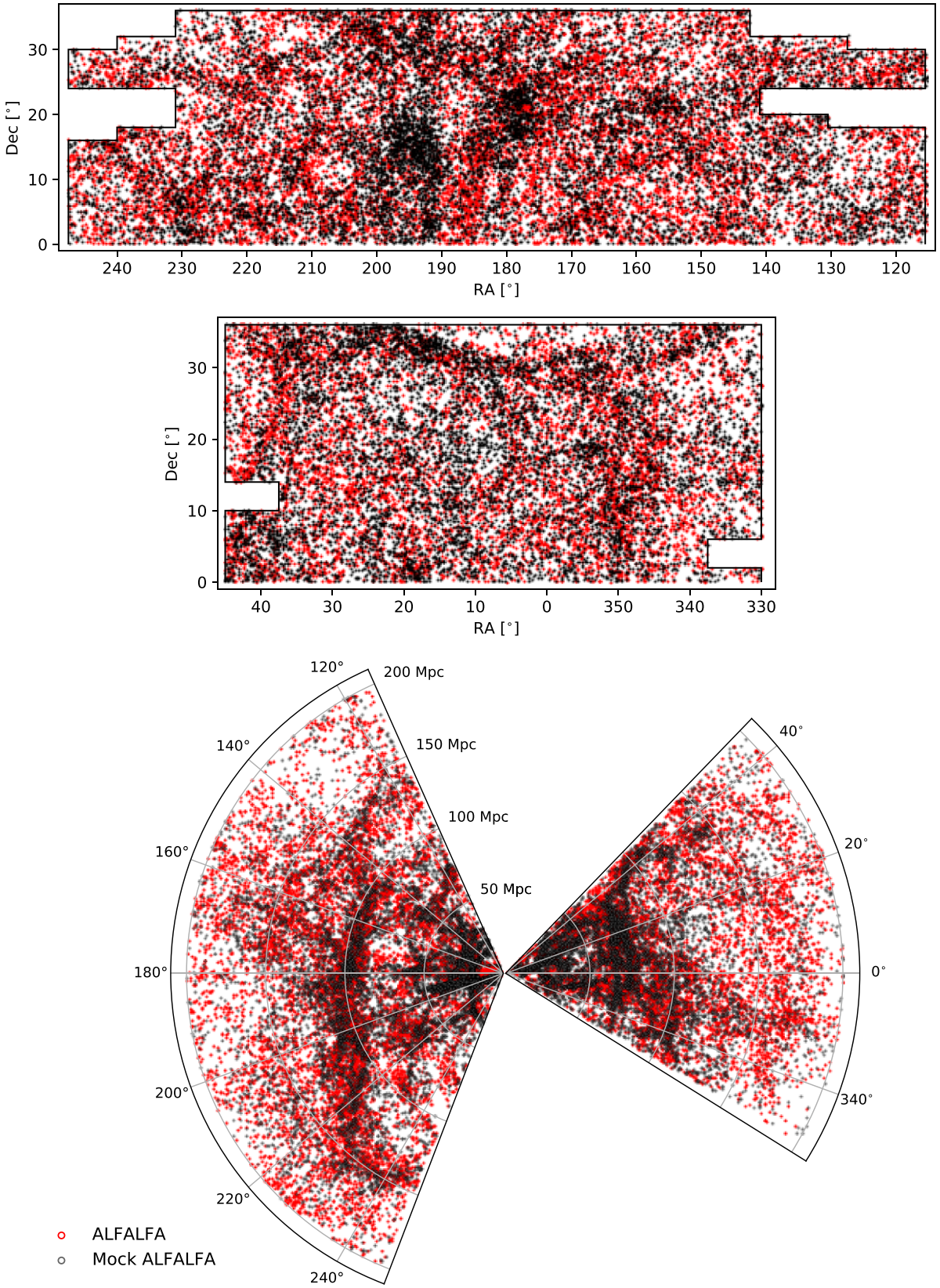

Figure 1. *Top panel:* Distribution of α.100 ALFALFA sources (red points) within the footprint on the sky of the spring field. Black points show sources contained in our ALFALFA-like mock survey catalogue constructed from the *Sibelius-DARK + GALFORM* simulation. *Middle panel:* Same as top panel, but for the fall field. *Bottom panel:* Cone diagram for the survey catalogues over all declinations as a function of distance from the Milky Way and right ascension on the sky.

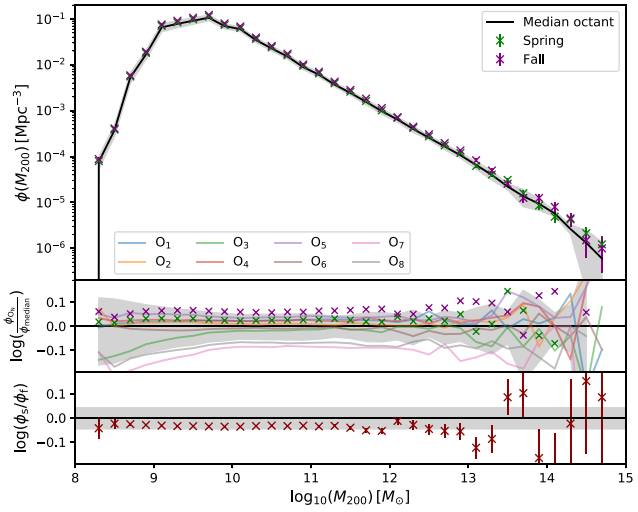

Figure 2. *Top panel*: The HMF for the spring (green crosses with 1σ Poisson uncertainties) and fall (purple crosses with 1σ uncertainties) ALFALFA fields from the $z = 0$ *Sibelius-Dark* plus *GALFORM* output. Additionally, the median HMF for the eight separate octants with its inter-quartile scatter is shown (black crosses). The octants are denoted by O_N ($N = 1$–8); details of the octants are given in Appendix B. *Middle panel*: The ratio of the N-th octant's HMF to the median HMF across octants. Each coloured line represents one of the octants HMF ratios. The shaded region shows the inter-quartile scatter about the median. *Bottom panel*: The ratio of the spring to fall HMF (dark red crosses with 1σ uncertainties). The shaded region indicates a region of ± 10 per cent differences in the HMF.

dominates over shot noise in our mass range of interest) as $\sigma_n \propto k \left(\frac{M_{\rm th}}{10^{13}\,{\rm M}_\odot}\right)^{0.15}$, where k depends on the survey volume. Assuming an 'effective' survey radius of 111 Mpc – the radius of a sphere enclosing the same volume as surveyed in the entire ALFALFA survey – we obtain $k \sim 0.116$. We therefore expect the mass function between 10^9 and $10^{12}\,{\rm M}_\odot$ sampled in a volume equivalent to the ALFALFA survey to scatter by about 5 per cent around the cosmic mean HMF in an equivalent volume [or a factor of $\sqrt{(2)}$ more for a volume equivalent to half of the survey]. Strictly speaking, this calculation applies to cumulative mass functions, but the weak dependence on mass across our range of interest means that it also provides a reasonable order-of-magnitude estimate for the differential mass function.

Given that the fall field has an HMF that is about 8 per cent overdense with respect to the spring field, shown in the lower panel of Fig. 2, and that both fields are overdense with respect to the median (Fig. 2, middle panel), by about 16 and 6 per cent, respectively, are therefore consistent with the expectation for a ΛCDM cosmology.

Further reassurance is provided by the fact that the normalizations of the HMFs in the various octants and in the spring and fall survey regions are broadly consistent with the relative galaxy densities of the northern and southern hemispheres reported for *Sibelius-DARK* by McAlpine et al. (2022, section 3.1.2 and fig. 8) and similarly for the *Simulating the LOcal Web* constrained realization by Dolag et al. (2023), the former highlighting that this difference is entirely consistent (within $1-2\sigma$) with the expectation for a ΛCDM cosmology.

2.3 *GALFORM*

Since the *Sibelius-DARK* simulation is an N-body simulation, we model the evolution of the galaxy population using the *GALFORM* semi-analytical model of galaxy formation.

The *GALFORM* semi-analytical model calculates the evolution of galaxies in hierarchical theories of structure formation. The processes governing galaxy formation and evolution are modelled as sets of coupled non-linear differential equations (White & Frenk 1991). Since the first *GALFORM* models (Cole et al. 1994, 2000), there have been numerous changes and improvements (e.g. Benson et al. 2000; Granato et al. 2000; Baugh et al. 2005). We model the *Sibelius-DARK* simulation galaxy population using the Lacey et al. (2016) variant of *GALFORM*. This is the same variant as used by McAlpine et al. (2022), who present a detailed investigation into the resulting galaxy population (e.g. luminosity function in their fig. 5, stellar mass functions in their fig. 6, etc.). This *GALFORM* variant incorporates different initial mass functions for quiescent star formation versus for starbursts, black hole formation, feedback from supernovae and from active galactic nuclei that suppresses gas cooling in massive haloes, and a new empirical star formation law in galaxy discs based on molecular gas content. A more accurate treatment of dynamical friction acting on satellite galaxies is also introduced, as well as an updated stellar population model.

2.4 Mock 21-cm survey

Four steps are required to construct a mock 21-cm survey from the *Sibelius-DARK* + *GALFORM* simulation: calculation of the galactic circular velocity curve, determination of the amount of H I gas as a function of line-of-sight velocity to produce a 21-cm spectrum, convolution with a kernel to model the thermal broadening of the 21-cm line, and application of the selection criteria consistent with the chosen 21-cm survey.

To define the radial mass profiles of galaxies, we follow Lacey et al. (2016). The relevant mass-bearing components of a galaxy are the dark matter halo, stellar bulge, stellar disc, hot gas, and cold gas. Dark matter haloes are assumed to follow an NFW profile (Navarro, Frenk & White 1996, 1997) described by a virial[4] mass and concentration. Stellar bulges follow a 'de Vaucouleurs' ($r^{1/4}$) law for their surface mass density profile. This profile would require numerical integration to obtain an enclosed mass profile. We follow Lacey et al. (2016) and assume the simpler, though very similar in shape, Hernquist (1990) mass profile for bulges instead. The stellar disc is modelled as an infinitesimally thin disc with an exponentially decaying density profile with a half-mass radius $r_{\rm disc}$. The hot gas is assumed to settle into a spherically symmetric distribution with density profile

$$\rho_{\rm hot}(r) \propto \frac{1}{\left(r^2 + r_{\rm c}^2\right)^2} \tag{4}$$

with gas core radius $r_{\rm c} = 0.1\,r_{\rm vir}$, where $r_{\rm vir}$ is the virial radius of the dark matter halo. The cold gas is modelled as an infinitesimally thin disc with an exponentially decaying surface density profile with the same half-mass radius as the stellar disc, $r_{\rm disc}$. This is partitioned into atomic and molecular gas components, expressed as the fraction of cold gas that is molecular $f_{\rm mol} = \Sigma_{\rm mol}/(\Sigma_{\rm mol} + \Sigma_{\rm atom})$, calculated assuming vertical hydrostatic pressure equilibrium and a gas velocity dispersion of $\sigma_{\rm gas} = 10 \pm 2\,{\rm km\,s^{-1}}$ (for a detailed description, see Lagos et al. 2011). The atomic hydrogen surface density is assumed to be $X_{\rm H} = 0.74$ times the atomic gas surface density. The galaxy components are assumed to be azimuthally symmetric. Circular

[4]We define virial quantities by a sphere enclosing an overdensity that is 200 times the critical density of the Universe, $\rho_{\rm crit} = 3H^2/8\pi G$, and denote them with a 'vir' subscript.

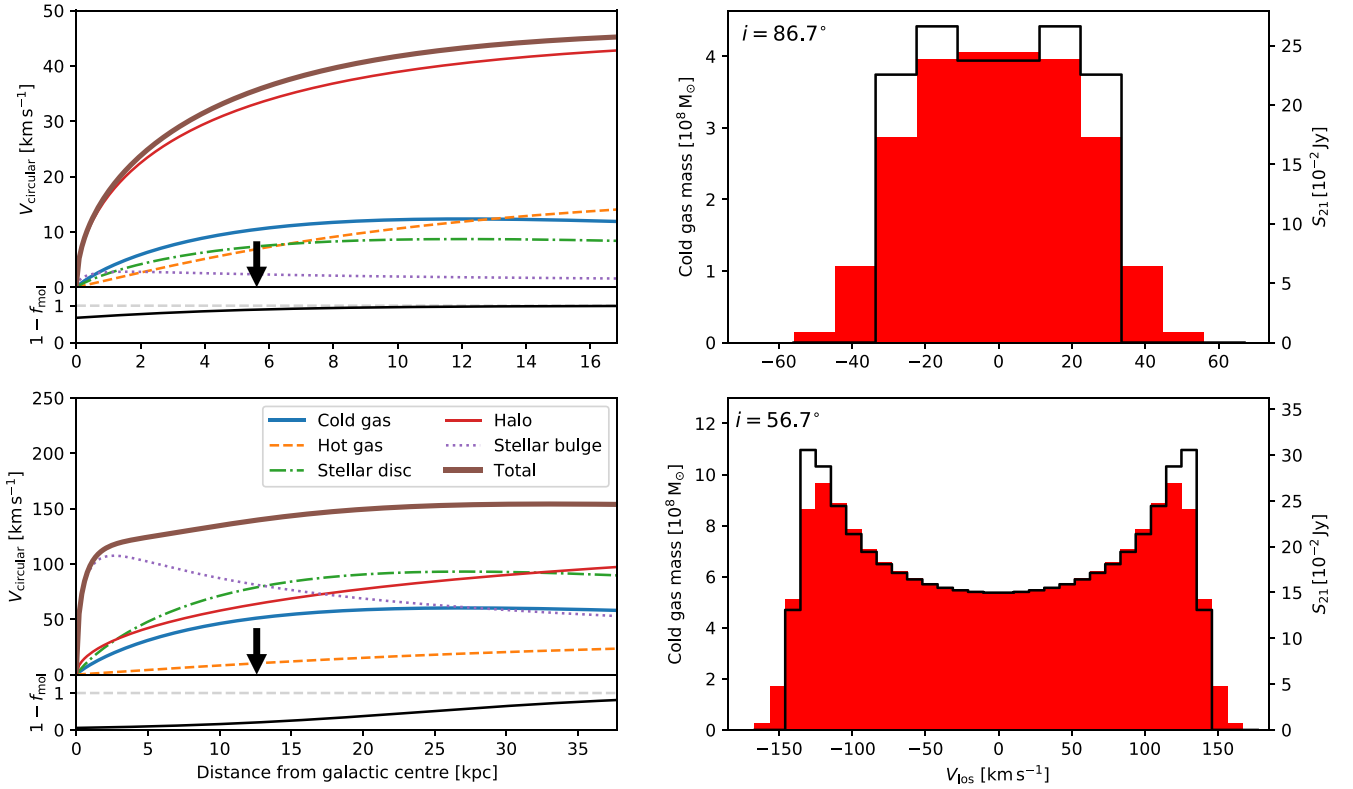

Figure 3. Two examples illustrating our model for generating a mock 21-cm velocity spectrum for *Sibelius-DARK + GALFORM* galaxies. *Left-hand panels:* Examples of the circular velocity curve shown within $3\,r_{\rm disc}$ for a 'low' line width (top) and an 'intermediate' line width galaxy (bottom), including the $(1 - f_{\rm mol})$ atomic fraction profile beneath. The arrow indicates the scale length of the cold gas disc, $r_{\rm disc}$. Contributions from the dark matter halo (red line), stellar bulge (purple dotted line), stellar disc (green dash–dot line), hot gas (orange dashed line), and cold gas (blue solid line) are shown. The sum in quadrature of these components gives the model velocity curve (thick brown line). *Right-hand panels:* The 21-cm velocity spectrum resulting from the circular velocity curves and H I gas surface density profiles for both the case before (black open histogram) and after (red filled histogram) including the gas velocity dispersion, σ_v. The left-hand y-axis shows the distribution by H I mass, while the right-hand y-axis shows the distribution by 21-cm flux, S_{21}. The randomly chosen inclination to the line of sight, i, for each galaxy is given in the top left corners of the panels. 'Low' line width galaxy properties: $M_{\rm vir} = 4.4 \times 10^{10}\,{\rm M}_\odot$, stellar disc mass, $M_{\star,\rm disc} = 1.3 \times 10^{8}\,{\rm M}_\odot$, stellar bulge mass, $M_{\star,\rm bulge} = 1.2 \times 10^{7}\,{\rm M}_\odot$, hot gas mass, $M_{\rm hot} = 1.1 \times 10^{10}\,{\rm M}_\odot$, and cold gas mass, $M_{\rm cold} = 6.9 \times 10^{8}\,{\rm M}_\odot$. 'Intermediate' line width galaxy properties: $M_{\rm vir} = 1.4 \times 10^{12}\,{\rm M}_\odot$, $M_{\star,\rm disc} = 3.3 \times 10^{10}\,{\rm M}_\odot$, $M_{\star,\rm bulge} = 2.8 \times 10^{10}\,{\rm M}_\odot$, $M_{\rm hot} = 1.3 \times 10^{11}\,{\rm M}_\odot$, and $M_{\rm cold} = 3.7 \times 10^{10}\,{\rm M}_\odot$.

velocities for spherically symmetric components are calculated as $v_{\rm c}^2(r) = GM(<r)/r$, while for thin exponential discs we use

$$v_{\rm c}^2 = \frac{2GM_{\rm disc}}{r_{\rm disc}} y^2 \left[I_0(y)K_0(y) - I_1(y)K_1(y)\right], \quad (5)$$

where $M_{\rm disc}$ is the total mass of the disc, $y = r/2r_{\rm disc}$, and I_v and K_v are the modified Bessel functions. The total circular velocity at any radius is then the sum in quadrature over all components. The resulting circular velocity curve, with its decomposition into different components, is shown in the left-hand panels of Fig. 3 for two example galaxies.

We define the 21-cm velocity spectrum as the H I mass-weighted distribution of line-of-sight velocities. To determine the line-of-sight velocities, the inclination i of the galaxy to the observer must be set. This is done by drawing a uniformly distributed random value for $\cos(i)$ between 0 and 1. We have repeated our analysis using 10 unique random seeds to assign galaxy inclinations. These cause only small differences of about 8 per cent or less in the number of galaxies detected in our mock survey at fixed w_{50}.

To calculate how much H I gas contributes at each line-of-sight velocity, the model H I disc is discretized in the radial (50 bins from $r = 0$ to $4\,r_{\rm disc}$) and azimuthal angular (45 bins from $\phi = 0$ to 2π) coordinates. The line-of-sight velocity of each radial-angular element is $v_{\rm circular}(r)\sin i\cos\phi$. The radial extent is sufficient to enclose ≥90 per cent of the cold gas mass, enough to obtain a converged value for w_{50}. We always calculate the total flux of galaxies using the total atomic hydrogen mass, integrated to infinity. We match the effective ALFALFA spectral resolution of $10\,{\rm km\,s^{-1}}$.

We also account for the velocity dispersion of the H I gas. The 21-cm line is thermally broadened, hence influencing the measured value of the 21-cm line width. Following Lagos et al. (2011), we assume an empirically determined amplitude for the H I velocity dispersion of $\sigma_v = 10 \pm 2\,{\rm km\,s^{-1}}$. *GALFORM* also assumes the gas velocity dispersion to be $10 \pm 2\,{\rm km\,s^{-1}}$ (see Lacey et al. 2016, and references therein). We implement this by convolving the spectrum with a Gaussian kernel of standard deviation, σ_v. The right-hand panels in Fig. 3 show 21-cm velocity spectra for the same two example galaxies as shown in the left-hand panels. The example in the upper panels is a galaxy where the velocity dispersion is comparable to the maximum circular velocity. The example in the lower panels is a galaxy where the velocity dispersion is much smaller than the maximum circular velocity. In both cases, the black curves show the spectra before convolution with the thermal broadening kernel, while the red bins show the spectra including thermal broadening. We measure the 21-cm line width, w_{50}, as the full width at half-

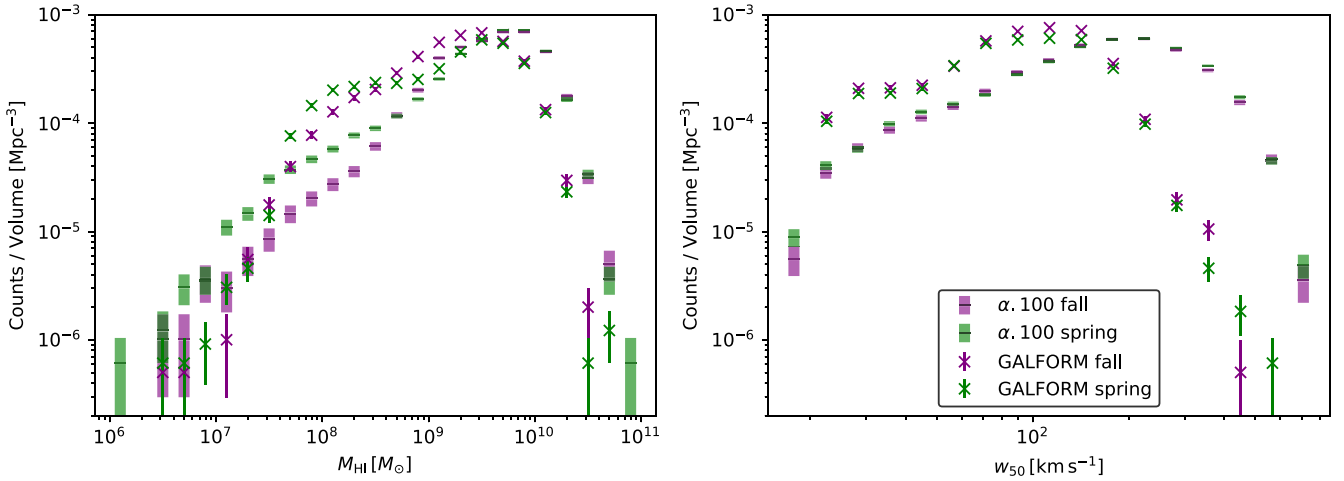

Figure 4. *Left-hand panel:* The observed counts per unit volume as a function of H I mass for the α.100 (dashes) and *GALFORM* (crosses) catalogues. The spring (green) and fall (purple) fields are shown separately. Shaded bands and error bars show Poisson uncertainties. *Right-hand panel:* Similar to the left-hand panel but showing the observed counts per unit volume as a function of 21-cm line width, w_{50}.

maximum of the spectra. We choose w_{50} over alternatives such as the full width at 20 per cent maximum, w_{20}, because w_{50} is less sensitive to noise in the 21-cm spectra, enabling more sources to be used (Zwaan et al. 2010).

To replicate the ALFALFA survey in order to study the observational and statistical effects on the measurements of the H I WF, we first need to create a mock catalogue of *Sibelius-DARK* + *GALFORM* sources that would have been detected by an ALFALFA-like survey. We draw our initial set of galaxies from the $z = 0$ output of the *GALFORM* model.[5]

We apply two criteria to determine which galaxies are included in our mock survey catalogue. First, only sources in the ALFALFA survey footprints should be included (Jones et al. 2018, tables D1–D4). The *Sibelius-DARK* halo catalogues provide (RA, Dec.) coordinates that we use to determine whether a source is included within the footprints. Secondly, the ALFALFA CL must be applied. We implement a continuous CL by linearly interpolating between the 25, 50, and 90 per cent spring and fall CLs given in equations (A1–A6), and linearly extrapolating to 0 and 100 per cent completeness. We draw a uniformly distributed random number between 0 and 1 for each *GALFORM* galaxy. If this number exceeds the survey completeness for the given 21-cm line width and flux of the galaxy, the galaxy is discarded from the catalogue. The result is a mock *GALFORM* survey containing 21 031 sources, of which 12 408 are in the spring field and 8623 are in the fall field (see Table 1, which additionally includes the total number of *GALFORM* sources before the selection criteria are applied). The distribution of the mock survey sources in (RA, Dec.) and distance is overlaid on the ALFALFA survey source distributions in Fig. 1.

For a survey like ALFALFA, there can be instances whereby two or more galaxies fall inside the beam at the same time and overlap in frequency, an effect termed 'beam confusion'. Our method to construct a mock survey does not account for beam confusion. This could plausibly be a limitation when comparing our mock survey to observations. Obreschkow et al. (2013) found that 'confused' galaxies typically have high H I mass and w_{50}, with $M_{\rm HI} > 10^{10}\,{\rm M}_\odot$ and $w_{50} > 300\,{\rm km\,s^{-1}}$, albeit for the H I PASS survey, which has a larger beam than ALFALFA. Subsequently, Chauhan et al. (2019) found lower levels of confusion in these ranges for their mock ALFALFA surveys and that any confusion only reduced the total number of galaxies in their sample by less than 1 per cent. Jones et al. (2015) also found that beam confusion can only slightly change the shape of the H I MF, by no more than would already be allowed by the random errors on the measurements. It therefore seems unlikely that beam confusion is one of the main drivers of the systematic difference between the spring and fall H I WFs, so we omit further discussion of this effect from our analysis below.

Fig. 4 shows the counts in the catalogues per unit volume as a function of H I mass (left-hand panel) and w_{50} (right-hand panel) for the ALFALFA α.100 catalogue (dashes) and *Sibelius-DARK* + *GALFORM* (crosses) for the spring (green) and fall (purple) fields individually. The *GALFORM* H I mass distribution overall appears similar in shape to that of ALFALFA. Noticeable differences include the overdensity of *GALFORM* sources at intermediate masses, $10^8 \lesssim M_{\rm HI}/{\rm M}_\odot \lesssim 10^9$, and the underdensity of sources at the highest H I masses. The *GALFORM* w_{50} distribution, on the other hand, has a starkly different shape to that measured in ALFALFA. There is an overdensity of sources for line widths below $\sim$150 km s^{-1} and an underdensity above. We comment further on these differences in Section 4.

The distributions of sources in the w_{50}–$M_{\rm HI}$ plane for the 'true *GALFORM*' data (detected and undetected galaxies in the simulation), the 'mock *GALFORM*' data (only those simulation sources within the ALFALFA selection criteria), and the ALFALFA data are shown in Fig. 5. The true *GALFORM* data display a bimodal distribution of sources that can be attributed to the relative contribution from satellite and central galaxies, respectively. There are obvious differences in the distribution between the true *GALFORM* and the mock, or real, ALFALFA survey. Given that the simulation will produce many galaxies that are either low in flux and/or possess large values for w_{50}, these sources will not satisfy the ALFALFA selection criteria and hence 'drop out' of the final survey catalogue. Alongside these w_{50}–$M_{\rm HI}$ distributions, we show the relative abundance of sources between the ALFALFA and mock *GALFORM* surveys in

[5]The ALFALFA survey detects galaxies out to $z \sim 0.05$. We have repeated our analysis using the $z = 0.05$ output of the same *GALFORM* model and find only small differences of about 5 per cent or less in the number of galaxies detected in our mock survey at fixed w_{50}.

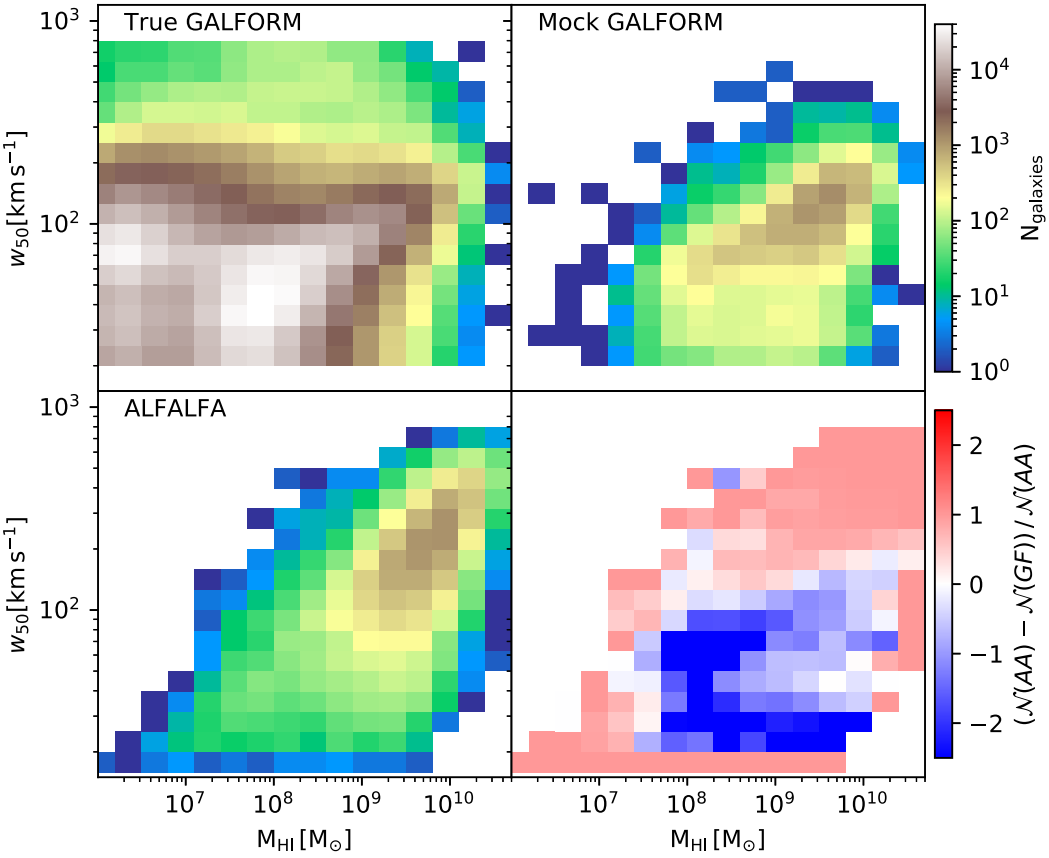

Figure 5. *Upper left:* Number of galaxies across the w_{50}–$M_{\rm H\,I}$ plane in the *Sibelius-DARK + GALFORM* catalogue, including all galaxies (detected and undetected) in the simulation at $z = 0$. The two peaks in the distribution correspond to the contributions from central [peak near $(M_{\rm H\,I}, w_{50}) \sim (10^8\,{\rm M}_\odot,\ 50\,{\rm km\,s^{-1}})$] and satellite galaxies [peak near $(M_{\rm H\,I}, w_{50}) \sim (2 \times 10^6\,{\rm M}_\odot,\ 70\,{\rm km\,s^{-1}})$]. The colour scale represents the number of galaxies per bin in w_{50}–$M_{\rm H\,I}$. *Upper right:* Number of galaxies across the w_{50}–$M_{\rm H\,I}$ plane in the mock *GALFORM* catalogue (sources passing the ALFALFA selection criteria). The colour scale is the same as in the upper left panel. *Lower left:* Number of galaxies across the w_{50}–$M_{\rm H\,I}$ plane in the ALFALFA catalogue. The colour scale is the same as in the upper left and upper right panels. *Lower right*: Relative abundance of sources between the ALFALFA and mock *GALFORM* surveys across the $M_{\rm H\,I}$–w_{50} plane. The two distributions are normalized by their respective integrals across the entire w_{50}–$M_{\rm H\,I}$ plane, subtracted in each bin, and then compared to the normalized abundance of ALFALFA sources in the same bin. Red regions indicate an excess of ALFALFA sources; blue indicates an excess in *GALFORM*.

the bottom right panel of Fig. 5. Generally, the ALFALFA survey contains more sources at the edges of the w_{50}–$M_{\rm H\,I}$ distribution. In particular, the smallest bin in w_{50} ($w_{50} \sim$ 20–30 km s^{-1}) contains no sources for our *GALFORM* mock. This comes from a limitation in our method when accounting for thermal broadening of the 21-cm emission line (see Section 2.4 for details). On the other hand, the mock *GALFORM* survey has an abundance of sources relative to ALFALFA for $(M_{\rm H\,I}, w_{50}) \sim (10^8$–$10^{10}\,{\rm M}_\odot,\ 25$–$150\,{\rm km\,s^{-1}})$. This is consistent with the corresponding 1D $M_{\rm H\,I}$ and w_{50} distributions shown in Fig. 4. We comment further on differences in the w_{50}–$M_{\rm H\,I}$ plane between our mock and the real ALFALFA survey in Section 3.4.

2.5 The $1/V_{\rm eff}$ maximum likelihood estimator

The $1/V_{\rm eff}$ estimator is used to estimate the abundance of undetected galaxies with a given $M_{\rm H\,I}$ and w_{50} from the abundance of those galaxies that were detected. Galaxies can be undetected in the survey due to being low-mass and hence low-flux sources and/or due to having wider line widths that spread their emission over more channels in the detector. The original SWML estimator (Efstathiou, Ellis & Peterson 1988) is applicable to galaxy samples that are integrated-flux limited. The SWML can be extended to become the bivariate stepwise maximum likelihood (2DSWML) estimator for surveys that have a selection function that depends on two observable quantities, such as ALFALFA. The only difference between the $1/V_{\rm eff}$ and 2DSWML estimators is that in the $1/V_{\rm eff}$ case, the effective volumes are iteratively calculated for each individual galaxy, instead of calculated per 2D bin in $M_{\rm H\,I}$–w_{50} space as in the 2DSWML approach (Zwaan et al. 2005; Martin et al. 2010). The effective volumes found for each galaxy are maximum likelihood counterparts of the classical $1/V_{\rm max}$ volumes (Schmidt 1968) with the important difference that the $1/V_{\rm eff}$ method in principle corrects for spatial non-uniformity in the source distribution.

The $1/V_{\rm eff}$ estimator is 2D in the sense that the CL of the ALFALFA survey depends on both the integrated 21-cm flux, S_{21}, and line width, w_{50}, of the source. The implementation of the $1/V_{\rm eff}$ estimator is the same as in Oman (2022) with the only difference being that we adopt separate CLs for the two survey fields. The $1/V_{\rm eff}$ method requires knowledge of the survey CL in order to produce the 2D H I mass-width function. Instead of using the global 50 per cent CL as in Oman (2022, equation A5), we adopt the 50 per cent CL appropriate to each survey field (equations 2 and 3). From the summation of the values of the $1/V_{\rm eff}$ weights in 2D bins in $M_{\rm H\,I}$ and w_{50}, we compute the 2D H I mass-width function. The sum along the w_{50} axis gives the H I MF, while the same along the mass axis gives the H I WF.

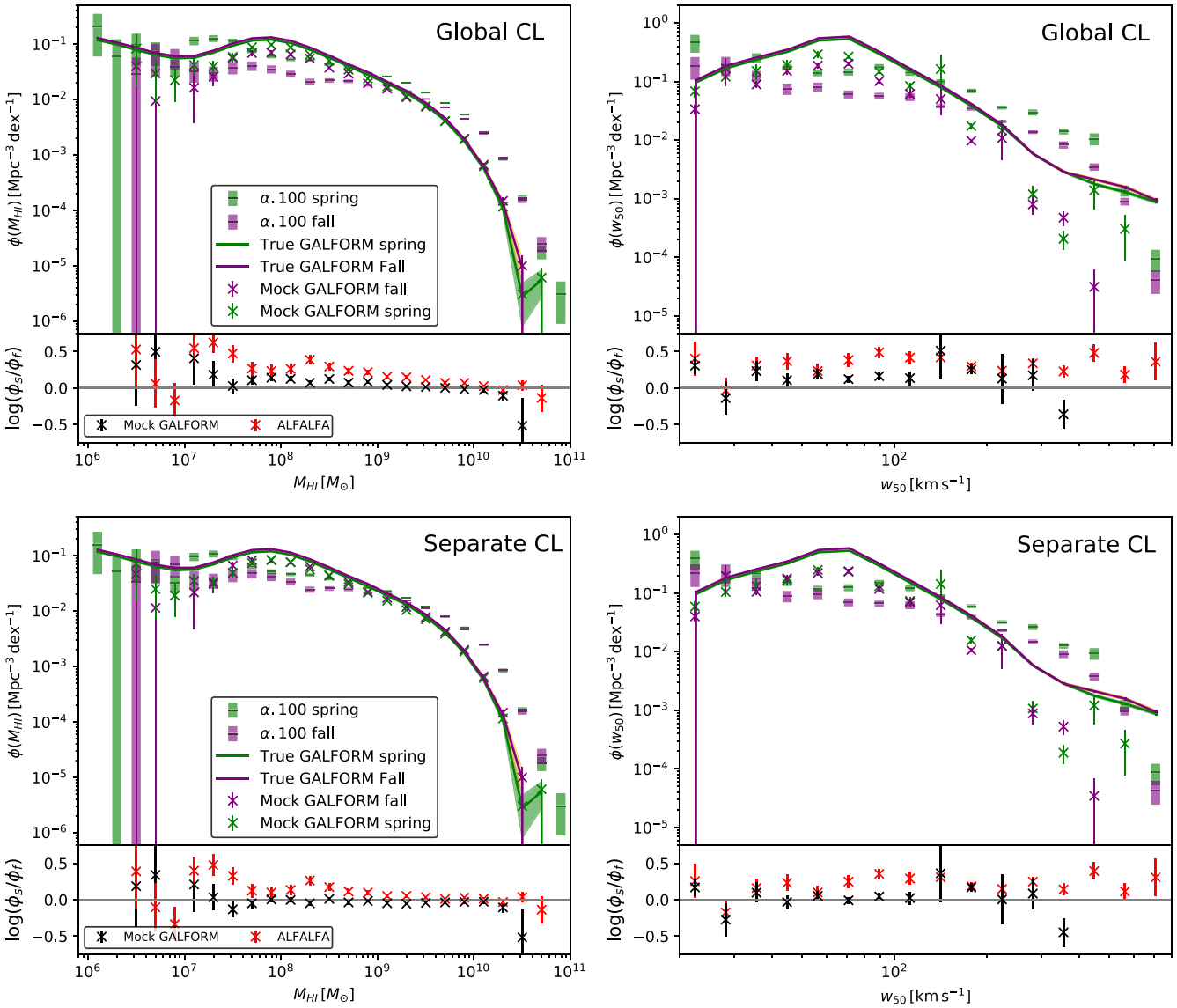

Figure 6. *Upper left panel:* The H I MF calculated by adopting the same CL in the spring and fall fields. The main panel shows the H I MF measured from the α.100 (dashes with shaded 1σ uncertainties) and *GALFORM* catalogues (crosses with 1σ uncertainties shown with error bars), separately for the two ALFALFA fields, spring (green) and fall (purple). Additionally, the 'true' *GALFORM* H I MF (solid lines with 1σ uncertainty shown with shaded band) is shown for the spring and fall fields. The lower sub-panel shows the ratio of the spring and fall *GALFORM* (black crosses) and ALFALFA (red crosses) H I MFs with 1σ uncertainties. *Upper right panel:* The H I WF calculated by adopting the same CL in the spring and fall fields. Lines and symbols are as in upper left panel. *Lower left and right-hand panels:* Similar to upper panels, but showing the H I MF and H I WF calculated by adopting separate CLs in the spring and fall fields.

3 ALFALFA AND *SIBELIUS-DARK* + *GALFORM* H I WIDTH FUNCTIONS

Understanding the origins of spatial variations of the normalization of the ALFALFA H I WF is crucially important in the context of using it as a constraint on cosmology. The ΛCDM cosmological model predicts that the dark matter HMF should be universal (in shape and normalization; e.g. Frenk et al. 1988; Mo & White 1996; Sheth & Tormen 2002; Crain et al. 2009) and therefore similar in the two fields surveyed in ALFALFA, because the volumes sampled are sufficiently large. We have checked this explicitly in the *Sibelius-DARK* simulation: the HMFs in the spring and fall volumes differ by no more than 8 per cent (within their uncertainties) at any halo mass $10^8 < M_{\rm vir}/{\rm M}_\odot < 10^{14}$ (see Section 2.2).

The most straightforward prediction for the H I WF is that it should also have the same shape and normalization (within about 8 per cent) in the two fields (see section 5.3.2, Oman 2022, for a detailed account of the connection between the HMF and the H I WF). Indeed, the *Sibelius-DARK* + *GALFORM* galaxy catalogue confirms this. In the right-hand panels of Fig. 6, we show the H I WF of all galaxies with $M_{\rm H\,I} > 10^6\,{\rm M}_\odot$ in the spring and fall survey fields (regardless of whether they would be detected) with the green and purple solid lines, respectively. Analogously to the HMFs in the two regions, the two H I WF curves differ by no more than 8 per cent at any line width. If the approximately factor of 2 difference in normalization of the H I WF for the spring and fall ALFALFA fields cannot otherwise be explained, then the ΛCDM model could be called into question (Zwaan et al. 2010; Papastergis et al. 2011; Oman 2022, where the first entry highlights the issue in the H I PASS survey and those after

for ALFALFA). We have already outlined above (Section 2.4) that random scatter in the H I WF driven by random inclination angles and the systematic effects due to the redshift evolution of galaxies within the survey volume are much too small to explain the large observed difference in normalization, so we now turn our attention to other potential sources of error.

In Section 3.1 we measure the H I WF of our mock *Sibelius-DARK + GALFORM* survey using exactly the same method as Oman (2022), and compare to their measurement for the ALFALFA survey. Next, in Section 3.2, we repeat the measurement assuming the separately derived CLs for the spring and fall fields in the calculation of the $1/V_{\rm eff}$ weights, and compare to the same approach applied to the $\alpha.100$ catalogue. In Section 3.3 we investigate effect of the clustering of sources along the line of sight on the H I WFs. Finally, in Section 3.4 we investigate the effect of the galaxy distribution in the H I mass – spectral line width plane.

3.1 Fiducial analysis of the mock surveys

We make the measurement of the H I MF and H I WF for the spring and fall mock *Sibelius-DARK + GALFORM* catalogues separately, following the procedure outlined in Oman (2022) where the globally derived CL for the ALFALFA survey (Oman 2022, equation A5) is assumed in the $1/V_{\rm eff}$ estimator. The top left and right-hand panels of Fig. 6, respectively, show the H I MFs and H I WFs measured using this approach. The measurements of Oman (2022), now adapted to retain only sources within 200 Mpc (Section 2.1), are also shown in these panels for comparison.

We restrict ourselves to a qualitative comparison of the $\alpha.100$ ALFALFA (dash marker with shaded box for the spring and fall fields in green and purple, respectively) and 'mock *GALFORM*' (crosses of corresponding colours) H I WFs. We find that the ALFALFA and mock *GALFORM* H I WFs have similar, almost constant low-line-width slopes for $w_{50} \lesssim 100\,{\rm km\,s^{-1}}$. For larger w_{50}, *GALFORM* underpredicts the number density of sources significantly.

Despite the 'true *GALFORM*' curves for the spring and fall fields (green and purple solid lines) differing by less than 8 per cent across the entire range in w_{50}, we find that our mock 21-cm survey has a qualitatively similar offset between the spring and fall H I WFs as is observed in ALFALFA. The number density in the spring field exceeds that in the fall field throughout the line-width range. The median ratio between the spring to fall H I WFs is $\log_{10}(\phi_{\rm s}\,/\,\phi_{\rm f}) = 0.36 \pm 0.09$ (16th–84th percentile scatter about the median) for ALFALFA and 0.17 ± 0.13 for mock *GALFORM*. Our mock *GALFORM* H I WFs capture the shapes of their true counterparts reasonably well, although they do underestimate them by up to a factor of 3 at $w_{50} \lesssim 100\,{\rm km\,s^{-1}}$, and up to a factor of 10 at higher line widths. The reason for this is that there are combinations of H I masses and line widths where galaxies exist, but none are observed (because the survey is not sensitive to detect them); this issue is discussed in detail by Oman (2022, sec. 5.2.6).

We focus our attention on the puzzling result that the mock *GALFORM* H I WFs are systematically offset from each other in the same sense as ALFALFA. We first consider the influence of the choice of CL for each survey field.

3.2 Influence of the survey completeness limit

We repeat the measurement of the H I MF and H I WF for the spring and fall $\alpha.100$ and mock *Sibelius-DARK + GALFORM* catalogues as in Section 3.1, but this time assuming CLs derived separately for the spring and fall ALFALFA fields (equations (2) & 3) in the calculation of the $1/V_{\rm eff}$ weights. The bottom right panel of Fig. 6 shows the measured H I WF using this approach.

The $\alpha.100$ ALFALFA and mock *GALFORM* H I WFs have qualitatively similar shapes to those discussed in Section 3.1. However, the spring and fall mock survey H I WFs are no longer systematically offset from each other. Therefore, when the same CLs used to construct the mock catalogues (Section 2.5) are used to correct them for incompleteness, the $1/V_{\rm eff}$ algorithm correctly recovers the fact that the H I WFs in the two fields have indistinguishable shapes and amplitudes (although the true number densities are still underestimated). Quantifying this, the median ratio between the spring to fall H I WFs is $\log_{10}(\phi_{\rm s}\,/\,\phi_{\rm f}) = 0.05 \pm 0.14$.

For ALFALFA, the systematic offset between the spring and fall H I WFs is reduced when separate spring and fall CLs are used in the calculation of the $1/V_{\rm eff}$ weights, but does not disappear: the median ratio between the spring to fall H I WFs is $\log_{10}(\phi_{\rm s}\,/\,\phi_{\rm f}) = 0.25 \pm 0.09$. Because there is empirical evidence that the ALFALFA survey has different CLs in the spring and fall fields (Haynes et al. 2011; Oman 2022), our view is that these measurements should be taken to supersede those of Oman (2022) when taking the same cut in recessional velocity – we tabulate them in Appendix C for both the H I MFs and H I WFs in the spring and fall regions. We also include a combined measurement for the entire survey in which the $1/V_{\rm eff}$ weights have been derived separately for galaxies in the two regions accounting for the different survey sensitivity in each.

The fact that the spring and fall ALFALFA H I WFs are still offset from each other after correcting for the different CLs in the two fields suggests that this may not be the only systematic effect influencing the measurement. There are a limited number of other factors that can influence the outcome of the $1/V_{\rm eff}$ measurement. We have identified only two: differences in the clustering along the line of sight in the two fields and/or differences in the properties of galaxies in the two fields as reflected by their distribution in the 2D space of $M_{\rm H\,I}$–w_{50}.

In order to focus on these other effects, for our analysis in the remainder of this section we remove the influence of the CL by imposing a very conservative completeness cut on both the ALFALFA survey and on our mock surveys: we impose a CL offset in S_{21} by 0.10 dex above that empirically derived for the fall field (equation 3). We also assume this CL in the calculations of $1/V_{\rm eff}$ weights below.

3.3 Influence of the clustering of sources along the line of sight

Our use of the $1/V_{\rm eff}$ estimator has so far treated the spring and fall fields as fully independent: when calculating the H I WF and H I MF for the spring field, no information is available about the large-scale structure in the fall field, and vice versa. We now wish to investigate whether the different clustering along the line of sight in each field causes a spurious systematic offset between the H I WFs. We proceed as follows. We calculate $1/V_{\rm eff}$ weights for the full $\alpha.100$ catalogue, thereby obtaining effective volumes for each source without any knowledge of the different clustering of sources along the line of sight in each field. The $1/V_{\rm eff}$ algorithm makes no use of the (RA, Dec.) coordinates of sources, so this approach treats the survey as a single contiguous volume. We then separately compute the H I WF for each field using the effective volumes for sources in that field. This exercise should erase the offset between the spring and fall fields if the distribution of sources along the line of sight is its direct cause.

The resulting H I WFs are shown in the top panel of Fig. 7. We label the case for which the survey fields are treated independently as 'independent $V_{\rm eff}$', and the case for which the survey fields are analysed jointly as 'full survey $V_{\rm eff}$'. We find that there is almost no

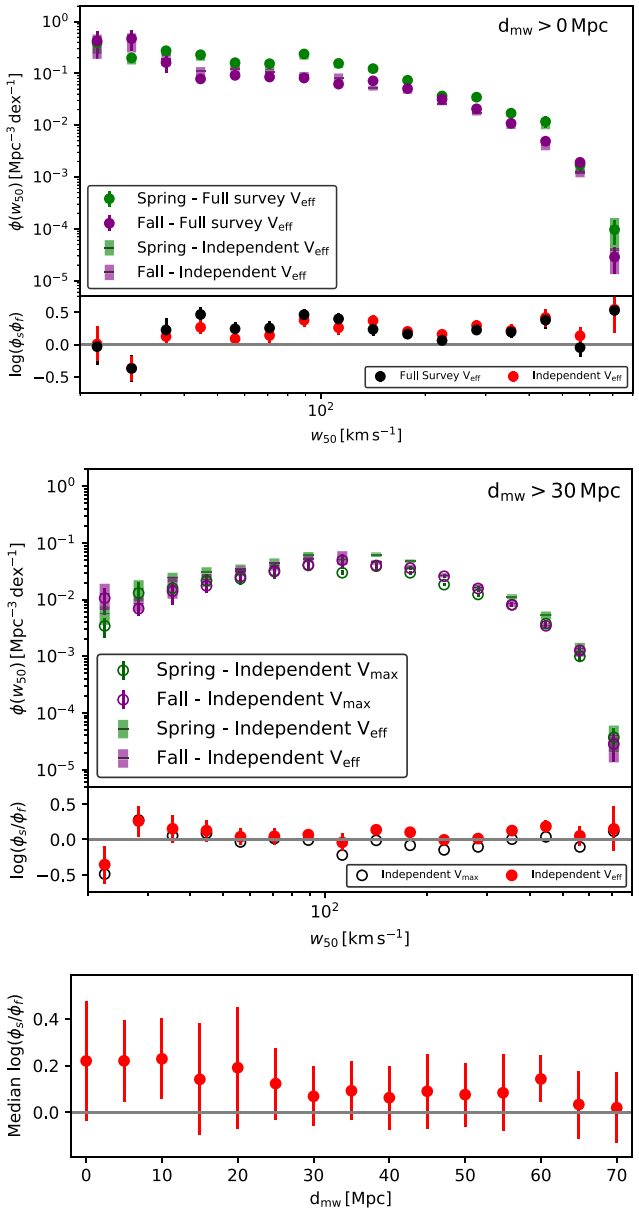

Figure 7. *Upper panel:* The spring (green) and fall (purple) $\alpha.100$ ALFALFA H I WFs as derived using the $1/V_{\rm eff}$ weights assigned independently (described in Section 3.3; dashes with shaded 1σ uncertainties) and from the full survey catalogue (points with error bars showing 1σ uncertainties). The CL adopted has been raised by 0.10 dex above that empirically derived for the fall field. Sources detected at all distances from the Milky Way are included. The lower sub panel shows the spring-to-fall ratio for the independent (red points) and full survey (black points) $1/V_{\rm eff}$ weights. *Middle panel:* The spring (green) and fall (purple) H I WFs derived using the $1/V_{\rm eff}$ (dashes) and $1/V_{\rm max}$ (open circles) weights assigned independently but including only those sources detected at distances $d_{\rm mw} > 30$ Mpc. The lower sub panel again shows the spring-to-fall ratio for the independent $1/V_{\rm eff}$ (red points) and independent $1/V_{\rm max}$ (open black circles) $1/V_{\rm eff}$ weights. *Lower panel:* The median H I WF ratio for spring-to-fall regions (derived using the $1/V_{\rm eff}$ weights) as a function of the minimum distance from the Milky Way at which sources are retained, $d_{\rm mw}$.

change in the systematic offset between the spring and fall H I WFs; the median ratio between the spring to fall H I WFs derived using the independent $V_{\rm eff}$ weights is $\log_{10}(\phi_{\rm s}\,/\,\phi_{\rm f}) = 0.22 \pm 0.11$. Therefore, the different distribution of sources along the line of sight in the two fields is not an important driver of the systematic offset. This leaves the distribution of sources in the 2D $M_{\rm H\,I}$–w_{50} parameter space to be investigated.

3.4 Influence of the galaxy distribution in $M_{\rm H\,I}$–w_{50}

The distribution of detected sources in the $M_{\rm H\,I}$–w_{50} plane is the fundamental input into the $1/V_{\rm eff}$ estimator. Even if the dark matter HMF is identical in the spring and fall fields, should galaxy formation and evolution proceeded differently in the two fields this could lead to differences between their respective H I WFs. We therefore check for measurable differences in the $M_{\rm H\,I}$–w_{50} distribution of sources between the two survey fields, which could be a symptom of such differential galaxy evolution. We show ratios of the $M_{\rm H\,I}$–w_{50} distributions in the two fields in Fig. 8. The intrinsic distribution of *GALFORM* galaxies ('true *GALFORM*', left-hand panel) is slightly overdense in the fall field. The amplitude of the fall overdensity in the $M_{\rm H\,I}$–w_{50} plane is consistent with that in the HMF (see Section 2.2). We interpret this to imply that galaxy formation in this *GALFORM* model proceeds essentially identically in the two fields. There is also some noise due to low number counts at high $M_{\rm H\,I}$ and/or w_{50}.

In the centre panel, we show the ratio of the distributions of *GALFORM* galaxies detected in our mock survey, weighted by their $1/V_{\rm eff}$ weights (calculated separately for the spring and fall fields). Large differences would arise due to the different CLs in the two fields, so we make this comparison after imposing the aforementioned more conservative completeness cut. We repeat the same process for the ALFALFA survey and show the result in the right-hand panel of the figure. In the mock survey, the two distributions agree over most of the parameter space, except around $(M_{\rm H\,I}, w_{50}) \sim (10^{8.5}\,{\rm M}_\odot, 120\,{\rm km\,s^{-1}})$. The ALFALFA survey has a similar feature at the same location, but also has $\log(n_{\rm s}/n_{\rm f}) > 0$ over most of the space. This overall overdensity in the spring field is the same one reflected in the systematically higher number density in the same field in the H I MF and H I WF: these functions are the integrals of the H I mass-width function along the width and mass axes, respectively. We note that this does not necessarily imply that the bias needs be near-uniform across the $M_{\rm H\,I}$–w_{50} plane. Fig. 8 therefore suggests that the difference in number density seems to arise across all line widths and H I masses and is not tied to any particular region in this parameter space. The local overdensity near $(M_{\rm H\,I}, w_{50}) \sim (10^{8.5}\,{\rm M}_\odot, 120\,{\rm km\,s^{-1}})$ bears further investigation.

We have traced the origin of this feature to the difference in the number of detected sources as a function of distance (see Fig. 9) in the foreground of the survey ($d_{\rm mw} \lesssim 30$ Mpc). In Section 3.3 we confirmed that the systematic offset between the H I WFs in the two survey fields is not due to a global difference in the clustering of sources in distance. However, local features could still play a role. Essentially, the $1/V_{\rm eff}$ estimator seems to be incorrectly extrapolating the foreground overdensity in the spring field through the entire survey volume, biasing the spring field to higher number densities. This effect is similar to what would occur if we used the $1/V_{\rm max}$ estimator (Schmidt 1968), which assumes that galaxies are uniformly distributed in space, but less severe – the $1/V_{\rm eff}$ algorithm is intended to compensate for non-uniformity in the galaxy distribution, but does so imperfectly. The spring field in our mock *GALFORM* survey also has slightly more sources than the fall field at small distances, but the effect is much less pronounced, explaining why this particular

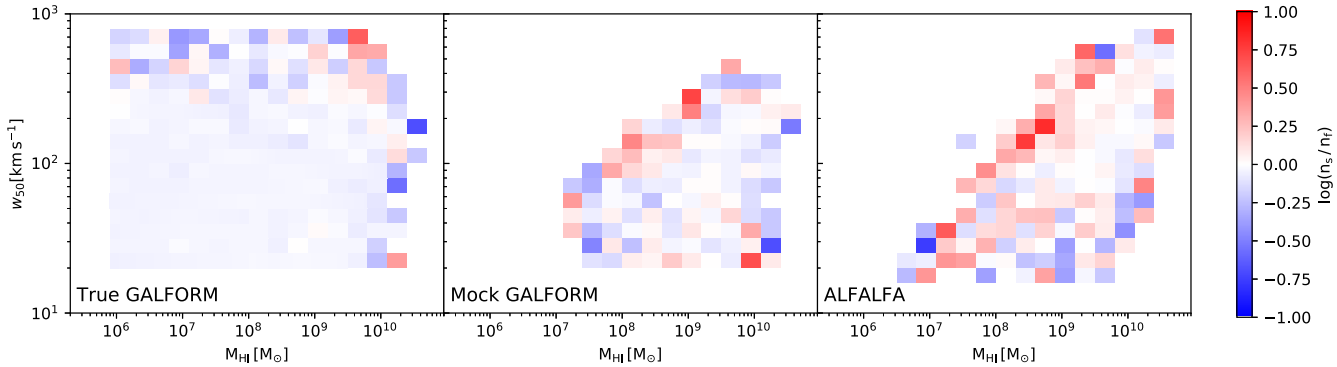

Figure 8. Logarithmic ratio of the abundance per unit volume of sources in the w_{50}–$M_{\rm H\,I}$ plane between the spring and fall survey fields. *Left-hand panel: Sibelius-DARK + GALFORM* catalogue data including all galaxies (detected and undetected). The approximately uniform overdensity in the fall relative to the spring region is due to the different HMFs in the two regions as detailed in Section 2.2. *Middle panel:* Mock *GALFORM* catalogue data (detected sources only) weighted by $1/V_{\rm eff}$ derived independently in each survey field. *Right-hand panel:* ALFALFA catalogue data weighted by $1/V_{\rm eff}$ derived independently in each survey field. In the centre and right-hand panels, only sources above a CL 0.10 dex above that empirically derived for the fall field are included.

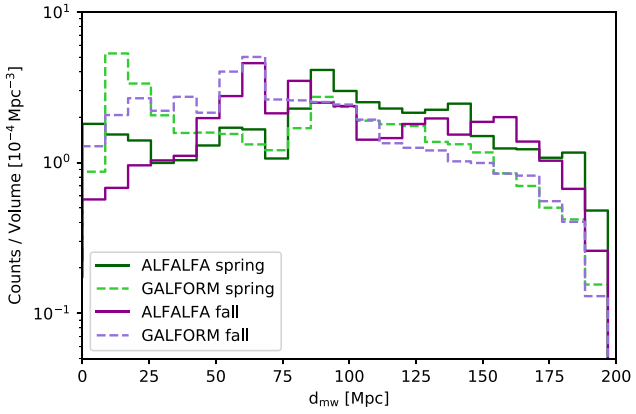

Figure 9. Histogram showing the counts per unit volume as a function of distance from the Milky Way, $d_{\rm mw}$. The α.100 ALFALFA (solid lines) and *GALFORM* (dashed lines) catalogues are shown, split by survey field: spring (ALFALFA: dark green; *GALFORM*: light green) and fall (ALFALFA: magenta; *GALFORM*: light purple).

bias is less pronounced in the mock survey. This can be ascribed to the *Sibelius-DARK* simulation being an imperfect match to the actual local cosmic web. We note an increased density of *GALFORM* sources at ∼15 Mpc in the spring field, associated with the Virgo cluster. We suggest that this is the cause of the localized overdense spring feature at $(M_{\rm H\,I}, w_{50}) \sim (10^{8.5}\,{\rm M}_\odot, 120\,{\rm km\,s^{-1}})$ in Fig. 8.

We confirm this interpretation of the origin of the offset between the ALFALFA spring and fall H I WFs by removing all sources from the ALFALFA catalogue with distances $d_{\rm mw} < 30$ Mpc. This choice of distance cut is further motivated in the lower panel of Fig. 7. Here, we show the median ratio of the spring and fall H I WFs as a function of the minimum distance from the Milky Way at which sources are retained. For all distances beyond about 30 Mpc, the median ratio stays about the same, indicating that the foreground galaxies have the most influence within about this distance. In the middle panel of Fig. 7, we show the H I WFs in the two fields measured with both (i) the conservative CL described above imposed and (ii) sources within 30 Mpc removed. In this case the H I WFs in the two fields are very close to agreement; the median ratio between the spring to fall H I WFs derived using the independent $V_{\rm eff}$ weights is $\log_{10}(\phi_{\rm s}\,/\,\phi_{\rm f}) = 0.07 \pm 0.12$. We attribute the small remaining differences to the $1/V_{\rm eff}$ estimator slightly overcompensating for clustering of sources along the line of sight beyond 30 Mpc. For comparison, we also show the H I WFs in the two fields measured using the $1/V_{\rm max}$ algorithm (i.e. assuming a spatially uniform galaxy distribution) in the middle sub-panel of Fig. 7. In this case, the sign of the offset between spring and fall is reversed (due to an overdensity of sources in the fall field at distances of ∼50–80 Mpc) with a median ratio between the spring to fall H I WFs derived using the independent $V_{\rm max}$ weights being $\log_{10}(\phi_{\rm s}\,/\,\phi_{\rm f}) = -0.01 \pm 0.12$.

In summary, we attribute the differences between the spring and fall ALFALFA H I WFs (e.g. as measured by Oman 2022) to: (i) the adopted CL for the survey; and (ii) the $1/V_{\rm eff}$ estimator incorrectly extrapolating the foreground overdensity in the spring field through the entire survey volume (in that field). Accounting for these systematic effects leads to H I WFs in the spring and fall fields that are consistent with being identical, which is in turn consistent with the two fields having identical HMFs (within a few per cent) as expected in a ΛCDM cosmology.

4 DISCUSSION

In Section 4.1, we comment further on the influence of the choice of CL on the H I WF and the importance for future surveys of mitigating variations in the CL. Then, in Section 4.2 we discuss the possible origins of the qualitative differences between the ALFALFA H I MF and H I WF and those that we obtain from simulations.

4.1 Survey completeness limit

For ALFALFA, the CL is determined empirically. Oman (2022) found that the CLs in the ALFALFA spring and fall fields are different by about 0.02 dex in S_{21} when derived separately. This small quantitative difference was assessed to be insufficient to have a strong influence on the shape (low-velocity slope, and location and sharpness of exponential cut-off at high velocities – i.e. excluding the overall normalization) of the H I WF. We agree with this assessment, but emphasize that the overall normalization of the H I WF is extremely sensitive to the chosen CL. Given this, we have investigated whether differences between the spring and fall CLs could plausibly fully explain the overall difference in number density between the spring and fall H I WFs. Beginning from the

global 50 per cent CL (Oman 2022, eq. A5), we gradually 'raised' the fall and 'lowered' the spring CLs by equal increments until the overall difference in number density in the two fields vanished. We found that the minimum offset between the CLs in order to fully account for the variation would have to be at least 0.1 dex or five times greater than the empirically determined difference. Such a large difference between the CLs is completely incompatible with the $\alpha.100$ catalogue. The difference between the CLs in the two fields can therefore only explain part of the overall difference in amplitude between the spring and fall ALFALFA H I WFs.

Since a difference in sensitivity between the spring and fall fields is apparent, it is worth considering whether there may also be variations in sensitivity internally within each field. Indeed, this has previously been measured: using a catalogue covering 40 per cent of the final surveyed area, Haynes et al. (2011) measured a CL that is 0.02 dex 'deeper' than that measured by Oman (2022) using the entire surveyed area in the spring region.[6]

We suggest that it is worth considering possible measures to mitigate variations in sensitivity across the sky in the design of future 21-cm surveys intending to measure the H I WF (or H I MF). As an illustrative example, a survey built up by covering a wide area to shallow depth and repeating this process to build up sensitivity might be expected to have a more spatially uniform CL in the presence of time-varying radio frequency interference than one where a small field is observed to a final target depth before moving on to a subsequent field.

We also comment briefly on the assumed form of the CL for ALFALFA: a two-segment broken power law. There is what we would characterize as tentative evidence in the ALFALFA catalogue data for departures from this form. For instance, there is a hint of an upturn at low line widths, and the sharp transition between the two power-law segments may not accurately capture what is in reality likely to be a more gradual transition in slope. The relatively small number of sources observed in ALFALFA makes it challenging to make stronger statements, but the order of magnitude increase in source counts expected from the next generation of H I surveys should allow for much more tightly constrained determinations of their CLs.

4.2 Qualitative comparisons of observed and simulated H I MFs and H I WFs

4.2.1 The H I MF

The true *GALFORM* H I MF does not reproduce the almost constant low-mass slope seen for ALFALFA (left-hand panels, Fig. 6). Instead, there is a fall and rise in the number density of sources over the mass range $10^6 < M_{\rm H\,I}/{\rm M}_\odot < 10^8$. This is caused by the relative contributions of central and satellite galaxies to the H I MF, which in *GALFORM* are easily separable. In *GALFORM*, all galaxies are assumed to originate as central galaxies, but when the dark matter halo of a galaxy merges[7] with another more massive halo, that galaxy is assumed to become a satellite in the new larger halo. The evolution of galaxy properties is modelled separately for centrals and satellites. For instance, as soon as a galaxy becomes a satellite, its hot gas halo is instantly stripped away by the ram pressure of the central galaxy and added to the central's hot gas halo. Consequently, no further gas is allowed to cool and accrete on to the satellite. This aggressive stripping influences the cold gas content of a galaxy, causing the dip in the distribution in H I mass, seen in our Fig. 6 to occur for $M_{\rm H\,I} \lesssim 10^8\,{\rm M}_\odot$. The exact location of this transition is affected by the halo mass resolution of the simulation (Lacey et al. 2016, section 4.2.2). However, the dominance of the H I MF by central and satellite galaxies in the high- and low-mass ends, respectively, is independent of the model adopted (Lagos et al. 2011, section 4.1.1). Smoothing these two peaks into a constant low-mass slope for the H I MF may be possible if satellites' hot gas is not instantaneously ram pressure stripped and/or gas is allowed to cool on to satellites.

4.2.2 The H I WF

The true *GALFORM* H I WF when compared to ALFALFA shows a distinct lack of larger line width sources, particularly for $w_{50} > 200\,{\rm km\,s^{-1}}$ (Fig. 6, right-hand panels). The Lacey et al. (2016) *GALFORM* model assumes that atomic gas is distributed in a relatively compact disc (with the same scale length as the stellar disc) in all galaxies. Cases where the emission is dominated instead by a much more diffuse disc, a ring, or even discrete clouds, such as is observed in some massive galaxies (e.g. Oosterloo, Morganti & Sadler 1999; Serra et al. 2012), are not modelled. This leads to severe underestimates of w_{50} for some galaxies where a more realistic, radially extended H I distribution would lead to a wider line. The deficit of galaxies at the highest H I masses (visible in the H I MF), which should be populated by galaxies with large total masses and correspondingly large H I line widths, likely also plays a role.

5 SUMMARY AND OUTLOOK – THE H I WF AS A COSMOLOGICAL CONSTRAINT

The low-line-width end of the H I WF is in principle sensitive to properties of the dark matter through its close relationship with the HMF. Our H I WF derived from the ΛCDM *Sibelius* simulations (Fig. 6) has a much shallower slope than expected from the HMF in the mock-surveyed volume, and is in reasonable qualitative agreement with that measured from ALFALFA observations at low line widths. This reinforces previous work, showing that the ALFALFA and H I PASS H I WFs can be reconciled with a ΛCDM cosmology once relevant galaxy formation processes and observational biases are accounted for (Sawala et al. 2013; Brooks et al. 2017; Dutton et al. 2019). Specifically, this is a consequence of: (i) not every low-mass *Sibelius-DARK* halo hosting a *GALFORM* galaxy; (ii) not every galaxy having substantial amounts of atomic gas ($M_{\rm H\,I} > 10^6\,{\rm M}_\odot$); and (iii) the non-linear mapping between the maximum circular velocity and w_{50}. We have not, however, ruled out interpretations involving a low-mass truncation of the HMF, such as may arise from a WDM particle (e.g. Zavala et al. 2009). We plan to apply our mock survey methodology to a WDM version of the *Sibelius-DARK* volume in future work.

Our approach using a constrained *N*-body realization of the entire ALFALFA survey volume has allowed us to investigate systematic effects inaccessible to previous studies. The *Sibelius-DARK* simulation makes the clear prediction that the spring and fall fields of the survey should have identical HMFs, and our modelling extends this to predicting identical H I MFs and H I WFs. Taken at face value, this is in tension with the observed differences between the spring and fall H I MFs and H I WFs measured from ALFALFA observations (Jones et al. 2018; Oman 2022). We identified two effects that can bias the H I WF and plausibly explain the observed differences between the spring and fall ALFALFA H I WFs:

[6]This could, in principle, be attributed to differences between the algorithms used to derive the CLs, but Oman (2022) rederived a CL for the $\alpha.40$ catalogue (see their equations A2 and A5) and found only much smaller differences with respect to the Haynes et al. (2011) measurement.
[7]As defined by Lacey et al. (2016).

(i) The CLs in the spring and fall survey fields differ, but have previously been assumed to be identical in deriving the $1/V_{\rm eff}$ weights used in deriving the H I WF. We tabulate updated measurements correcting for this in Appendix C.

(ii) The $1/V_{\rm eff}$ algorithm is intended to compensate for galaxy clustering (cf. the classical $1/V_{\rm max}$ method assuming a uniform spatial distribution), but does so imperfectly. A foreground overdensity of H I-rich galaxies in the spring ALFALFA field and an underdensity in the foreground of the fall field drive large systematic errors.

We note that the small remaining systematic offset between the spring and fall H I WFs once the above two effects are accounted for is most likely also attributable to the limited ability of the $1/V_{\rm eff}$ to accurately compensate for differences between the galaxy clustering in distance in the two fields. In this interpretation, the true H I WF is spatially invariant (within about 10 per cent) when sampled in volumes comparable to the volume of the ALFALFA survey; the observed differences between the spring and fall H I WFs in ALFALFA are entirely spurious. Encouragingly, the influence of individual over/underdense regions on the calculation of $1/V_{\rm eff}$ weights can be mitigated by simply surveying a larger area on the sky; the ongoing WALLABY[8] survey (Koribalski et al. 2020) will cover an area about four times wider than that covered by ALFALFA.

The H I WF has the potential to become a stringent test of cosmological models. Realizing this potential will require a deeper understanding of the systematic biases influencing measurements, such as those due to spatial (or temporal) variability in survey sensitivity. Progress on theoretical issues is also needed. Combining semi-analytical models able to simulate directly the large volumes of current and future surveys with hydrodynamical simulations able to resolve in detail the internal structure of atomic gas in galaxies seems a promising path forward. We look forward to the prospect of using the H I WF as a constraint on the nature of dark matter.

ACKNOWLEDGEMENTS

We thank the anonymous referee for a careful review of the manuscript that led to numerous improvements. KAO and CSF acknowledge support by the European Research Council (ERC) through Advanced Investigator grant to C. S. Frenk, DMIDAS (GA 786910). KAO acknowledges support by STFC through grant ST/T000244/1. RANB acknowledges support from both Durham University's Collingwood College Undergraduate Research Internships (UGRI) programme and the Royal Society. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University, and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. This research has made use of NASA's Astrophysics Data System.

DATA AVAILABILITY

The public release of the *Sibelius-DARK* + *GALFORM* simulation data is available at https://virgodb.dur.ac.uk/ as described in (McAlpine et al. 2022, appendix A). Fully reproducing our analysis requires some *GALFORM* galaxy properties not included in the public release (e.g. bulge fractions, stellar half-mass radii, etc.). These can be requested from the Sibelius project team (McAlpine et al. 2022). The $\alpha.100$ catalogue is available at http://egg.astro.cornell.edu/alfalfa/data/index.php.

[8]Widefield ASKAP L-band Legacy All-sky Blind surveY.

REFERENCES

Avila-Reese V., Colín P., Valenzuela O., D'Onghia E., Firmani C., 2001, ApJ, 559, 516
Ball C. J., Haynes M. P., Jones M. G., Peng B., Durbala A., Koopmann R. A., Ribaudo J., O'Donoghue A., 2022, preprint (arXiv:2212.08728)
Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93
Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40
Brook C. B., Di Cintio A., 2015, MNRAS, 453, 2133
Brook C. B., Shankar F., 2016, MNRAS, 455, 3841
Brooks A. M., Papastergis E., Christensen C. R., Governato F., Stilp A., Quinn T. R., Wadsley J., 2017, ApJ, 850, 97
Chauhan G., Lagos C. d. P., Obreschkow D., Power C., Oman K., Elahi P. J., 2019, MNRAS, 488, 5898
Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Crain R. A. et al., 2009, MNRAS, 399, 1773
Dolag K., Sorce J. G., Pilipenko S., Hernández-Martínez E., Valentini M., Gottlöber S., Aghanim N., Khabibullin I., 2023, preprint (arXiv:2302.10960)
Dutton A. A., Obreja A., Macciò A. V., 2019, MNRAS, 482, 5606
Efstathiou G., Ellis R. S., Peterson B. A., 1988, MNRAS, 232, 431
Frenk C. S., White S. D. M., Davis M., Efstathiou G., 1988, ApJ, 327, 507
Giovanelli R. et al., 2005, AJ, 130, 2598
Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, ApJ, 542, 710
Haynes M. P. et al., 2011, AJ, 142, 170
Haynes M. P. et al., 2018, ApJ, 861, 49
Hernquist L., 1990, ApJ, 356, 359
Hu W., Kravtsov A. V., 2003, ApJ, 584, 702
Jasche J., Wandelt B. D., 2013, MNRAS, 432, 894
Jones M. G., Papastergis E., Haynes M. P., Giovanelli R., 2015, MNRAS, 449, 1856
Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS, 477, 2
Koribalski B. S. et al., 2020, Ap&SS, 365, 118
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649
Lavaux G., Hudson M. J., 2011, MNRAS, 416, 2840
McAlpine S. et al., 2022, MNRAS, 512, 5823
Macciò A. V., Udrescu S. M., Dutton A. A., Obreja A., Wang L., Stinson G. R., Kang X., 2016, MNRAS, 463, L69
Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359
Masters K. L., 2005, PhD thesis, Cornell Univ., New York
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Obreschkow D., Ma X., Meyer M., Power C., Zwaan M., Staveley-Smith L., Drinkwater M. J., 2013, ApJ, 766, 137
Oman K. A., 2022, MNRAS, 509, 3268
Oosterloo T., Morganti R., Sadler E., 1999, PASA, 16, 28
Papastergis E., Shankar F., 2016, A&A, 591, A58
Papastergis E., Martin A. M., Giovanelli R., Haynes M. P., 2011, ApJ, 739, 38

Papastergis E., Giovanelli R., Haynes M. P., Shankar F., 2015, A&A, 574, A113
Planck Collaboration I, 2014, A&A, 571, A16
Ponomareva A. A., Verheijen M. A. W., Bosma A., 2016, MNRAS, 463, 4052
Saintonge A., 2007, AJ, 133, 2087
Sawala T., Frenk C. S., Crain R. A., Jenkins A., Schaye J., Theuns T., Zavala J., 2013, MNRAS, 431, 1366
Sawala T. et al., 2016, MNRAS, 457, 1931
Sawala T., McAlpine S., Jasche J., Lavaux G., Jenkins A., Johansson P. H., Frenk C. S., 2022, MNRAS, 509, 1432
Schmidt M., 1968, ApJ, 151, 393
Serra P. et al., 2012, MNRAS, 422, 1835
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Wang J., Bose S., Frenk C. S., Gao L., Jenkins A., Springel V., White S. D. M., 2020, Nature, 585, 39
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
Zavala J., Jing Y. P., Faltenbacher A., Yepes G., Hoffman Y., Gottlöber S., Catinella B., 2009, ApJ, 700, 1779
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30
Zwaan M. A., Meyer M. J., Staveley-Smith L., 2010, MNRAS, 403, 1969

APPENDIX A: COMPLETENESS OF THE ALFALFA CATALOGUES

We present the $\alpha.100$ catalogue CL for the spring and fall fields individually. The global CL was measured and given in Oman (2022, equations A4–A6). The fall CL is slightly shallower than the global CL, by 0.011 dex, while the spring CL is slightly deeper, by 0.009 dex, for a net difference of 0.02 dex. Using the $\alpha.100$ catalogue for spring sources only, we derive the following CLs:

$$\log_{10}\left(\frac{S_{21,90\ \text{per cent}}}{\text{Jy km s}^{-1}}\right) = \begin{cases} 0.5\,W - 1.124 & W < 2.5 \\ W - 2.374 & W \geq 2.5 \end{cases} \quad \text{(A1)}$$

$$\log_{10}\left(\frac{S_{21,50\ \text{per cent}}}{\text{Jy km s}^{-1}}\right) = \begin{cases} 0.5\,W - 1.179 & W < 2.5 \\ W - 2.429 & W \geq 2.5 \end{cases} \quad \text{(A2)}$$

$$\log_{10}\left(\frac{S_{21,25\ \text{per cent}}}{\text{Jy km s}^{-1}}\right) = \begin{cases} 0.5\,W - 1.207 & W < 2.5 \\ W - 2.457 & W \geq 2.5 \end{cases}, \quad \text{(A3)}$$

where $W = \log_{10}(w_{50}/\text{km s}^{-1})$. Similarly, using the $\alpha.100$ catalogue for fall sources only, we derive the following CLs:

$$\log_{10}\left(\frac{S_{21,90\ \text{per cent}}}{\text{Jy km s}^{-1}}\right) = \begin{cases} 0.5\,W - 1.104 & W < 2.5 \\ W - 2.354 & W \geq 2.5 \end{cases} \quad \text{(A4)}$$

$$\log_{10}\left(\frac{S_{21,50\ \text{per cent}}}{\text{Jy km s}^{-1}}\right) = \begin{cases} 0.5\,W - 1.159 & W < 2.5 \\ W - 2.409 & W \geq 2.5 \end{cases} \quad \text{(A5)}$$

$$\log_{10}\left(\frac{S_{21,25\ \text{per cent}}}{\text{Jy km s}^{-1}}\right) = \begin{cases} 0.5\,W - 1.187 & W < 2.5 \\ W - 2.237 & W \geq 2.5 \end{cases}. \quad \text{(A6)}$$

APPENDIX B: *SIBELIUS-DARK* OCTANT DATA

We present in Table B1 the vertices in right ascension and declination for the octants of the *Sibelius-DARK* sky used in Fig. 2.

Table B1. The sky coordinate position definition and number counts of all dark matter haloes contained within bounds of each *Sibelius-DARK* octant. The median octant contains 2.42 million dark matter haloes. A distance cut keeping only sources within $d_{\text{mw}} \leq 200$ Mpc is applied.

Right ascension (°)	Declination (°)	Number of sources ($\times 10^6$)
0–90	0–90	2.61
0–90	−90–0	2.55
90–180	0–90	2.30
90–180	−90–0	2.62
180–270	0–90	2.73
180–270	−90–0	2.40
270–360	0–90	1.93
270–360	−90–0	2.04

APPENDIX C: TABULATED ALFALFA H I WF AND H I MF

In Tables C1 and C2, we tabulate our $\alpha.100$ ALFALFA H I WF and H I MF measurements (see bottom panels of Fig. 6).

Table C1. Amplitudes and uncertainties of the measured $\alpha.100$ H I WF for the spring and fall fields derived assuming the CLs given in equations (A1–A6) and a cut in recessional velocity, $v_{\text{rec}} \leq 15\,000$ km s^{-1}, matching used in Oman (2022). The combined measurement for the entire survey uses the $1/V_{\text{eff}}$ weights calculated separately for the two fields. Note that although the logarithms of all amplitudes and uncertainties are given, the uncertainties are symmetric on a linear scale and should be interpreted as the 1σ width of a Gaussian distribution (not a log-normal distribution).

	$\log_{10}\phi(w_{50})/\text{Mpc}^{-3}\,\text{dex}^{-1}$		
$\log_{10}\ w_{50}/\text{km s}^{-1}$	$\alpha.100$ Spring	$\alpha.100$ Fall	$\alpha.100$
1.25	$-0.35^{+0.18}_{-0.32}$	$-1.36^{+0.24}_{-0.59}$	$-0.53^{+0.18}_{-0.3}$
1.35	$-0.40^{+0.13}_{-0.18}$	$-0.71^{+0.15}_{-0.24}$	$-0.49^{+0.11}_{-0.14}$
1.45	$-0.82^{+0.06}_{-0.08}$	$-0.70^{+0.14}_{-0.20}$	$-0.77^{+0.07}_{-0.09}$
1.55	$-0.66^{+0.07}_{-0.08}$	$-0.88^{+0.09}_{-0.11}$	$-0.74^{+0.06}_{-0.07}$
1.65	$-0.82^{+0.05}_{-0.06}$	$-1.10^{+0.09}_{-0.11}$	$-0.89^{+0.04}_{-0.05}$
1.75	$-0.92^{+0.05}_{-0.05}$	$-1.07^{+0.07}_{-0.08}$	$-0.97^{+0.04}_{-0.04}$
1.85	$-0.90^{+0.06}_{-0.06}$	$-1.20^{+0.06}_{-0.07}$	$-0.10^{+0.05}_{-0.05}$
1.95	$-0.82^{+0.05}_{-0.05}$	$-1.22^{+0.05}_{-0.05}$	$-0.93^{+0.04}_{-0.04}$
2.05	$-0.91^{+0.05}_{-0.06}$	$-1.25^{+0.06}_{-0.07}$	$-1.03^{+0.04}_{-0.05}$
2.15	$-1.06^{+0.04}_{-0.05}$	$-1.42^{+0.03}_{-0.04}$	$-1.14^{+0.04}_{-0.04}$
2.25	$-1.22^{+0.03}_{-0.03}$	$-1.45^{+0.04}_{-0.04}$	$-1.29^{+0.03}_{-0.03}$
2.35	$-1.49^{+0.03}_{-0.03}$	$-1.69^{+0.03}_{-0.03}$	$-1.54^{+0.02}_{-0.02}$
2.45	$-1.57^{+0.05}_{-0.05}$	$-1.87^{+0.03}_{-0.04}$	$-1.66^{+0.04}_{-0.04}$
2.55	$-1.89^{+0.04}_{-0.05}$	$-2.09^{+0.06}_{-0.07}$	$-1.94^{+0.04}_{-0.04}$
2.65	$-2.02^{+0.09}_{-0.11}$	$-2.47^{+0.06}_{-0.07}$	$-2.12^{+0.07}_{-0.08}$
2.75	$-2.90^{+0.08}_{-0.10}$	$-3.07^{+0.06}_{-0.07}$	$-2.87^{+0.05}_{-0.06}$
2.85	$-4.07^{+0.15}_{-0.23}$	$-4.42^{+0.16}_{-0.25}$	$-4.08^{+0.11}_{-0.15}$

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

$\log_{10} M_{\rm H\,I}/{\rm M}_\odot$	$\log_{10}\phi(M_{\rm H\,I})/{\rm Mpc}^{-3}\,{\rm dex}^{-1}$ $\alpha.100$ Spring	$\alpha.100$ Fall	$\alpha.100$
6.1	$-0.80^{+0.23}_{-0.53}$	–	$-1.01^{+0.23}_{-0.53}$
6.3	$-1.28^{+0.30}_{-\infty}$	–	$-1.49^{+0.30}_{-1.57}$
6.5	$-1.09^{+0.20}_{-0.38}$	$-1.52^{+0.30}_{-\infty}$	$-1.20^{+0.18}_{-0.31}$
6.7	$-1.22^{+0.17}_{-0.27}$	$-1.17^{+0.20}_{-0.39}$	$-1.20^{+0.14}_{-0.20}$
6.9	$-1.49^{+0.13}_{-0.18}$	$-1.21^{+0.15}_{-0.23}$	$-1.36^{+0.10}_{-0.14}$
7.1	$-1.01^{+0.07}_{-0.08}$	$-1.47^{+0.14}_{-0.21}$	$-1.14^{+0.06}_{-0.07}$
7.3	$-0.96^{+0.06}_{-0.07}$	$-1.49^{+0.12}_{-0.16}$	$-1.10^{+0.05}_{-0.06}$
7.5	$-1.03^{+0.05}_{-0.05}$	$-1.41^{+0.10}_{-0.12}$	$-1.14^{+0.04}_{-0.05}$
7.7	$-1.19^{+0.04}_{-0.04}$	$-1.38^{+0.08}_{-0.09}$	$-1.25^{+0.04}_{-0.04}$
7.9	$-1.28^{+0.04}_{-0.04}$	$-1.43^{+0.07}_{-0.08}$	$-1.33^{+0.03}_{-0.04}$
8.1	$-1.33^{+0.03}_{-0.03}$	$-1.52^{+0.06}_{-0.07}$	$-1.40^{+0.03}_{-0.03}$
8.3	$-1.34^{+0.03}_{-0.03}$	$-1.66^{+0.05}_{-0.06}$	$-1.44^{+0.02}_{-0.03}$
8.5	$-1.42^{+0.03}_{-0.03}$	$-1.64^{+0.04}_{-0.04}$	$-1.49^{+0.02}_{-0.02}$
8.7	$-1.49^{+0.02}_{-0.02}$	$-1.66^{+0.03}_{-0.04}$	$-1.55^{+0.02}_{-0.02}$
8.9	$-1.58^{+0.02}_{-0.02}$	$-1.72^{+0.02}_{-0.03}$	$-1.63^{+0.01}_{-0.02}$
9.1	$-1.63^{+0.02}_{-0.02}$	$-1.74^{+0.02}_{-0.02}$	$-1.67^{+0.01}_{-0.01}$
9.3	$-1.75^{+0.01}_{-0.01}$	$-1.86^{+0.01}_{-0.01}$	$-1.79^{+0.01}_{-0.01}$
9.5	$-1.92^{+0.01}_{-0.01}$	$-2.01^{+0.01}_{-0.01}$	$-1.95^{+0.01}_{-0.01}$
9.7	$-2.09^{+0.01}_{-0.01}$	$-2.16^{+0.01}_{-0.01}$	$-2.11^{+0.01}_{-0.01}$
9.9	$-2.23^{+0.01}_{-0.01}$	$-2.38^{+0.01}_{-0.01}$	$-2.32^{+0.01}_{-0.01}$
10.1	$-2.59^{+0.01}_{-0.01}$	$-2.67^{+0.01}_{-0.01}$	$-2.62^{+0.01}_{-0.01}$
10.3	$-3.06^{+0.02}_{-0.02}$	$-3.09^{+0.02}_{-0.02}$	$-3.07^{+0.01}_{-0.01}$
10.5	$-3.74^{+0.03}_{-0.04}$	$-3.83^{+0.05}_{-0.05}$	$-3.77^{+0.03}_{-0.03}$
10.7	$-4.62^{+0.09}_{-0.11}$	$-4.65^{+0.11}_{-0.16}$	$-4.64^{+0.07}_{-0.09}$
10.9	$-5.61^{+0.23}_{-0.53}$	–	$-5.81^{+0.23}_{-0.53}$

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