Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **529,** 4192–4209 (2024)
Advance Access publication 2024 March 9

https://doi.org/10.1093/mnras/stae713

MIGHTEE-H I: H I galaxy properties in the large-scale structure environment at $z \sim 0.37$ from a stacking experiment

Francesco Sinigaglia [1,2,3,4★] Giulia Rodighiero [1,2] Ed Elson,[5] Alessandro Bianchetti,[1,2] Mattia Vaccari [6,7,8] Natasha Maddox [9,10] Anastasia A. Ponomareva [11] Bradley S. Frank,[6,7,12,13] Matt J. Jarvis [5,11] Barbara Catinella [14,15] Luca Cortese [14,15] Sambit Roychowdhury,[16] Maarten Baes [17] Jordan D. Collier,[7,18,19] Olivier Ilbert,[20] Ali A. Khostovan [21] Sushma Kurapati,[12] Hengxing Pan [11] Isabella Prandoni [8] Sambatriniaina H. A. Rajohnson [12] Mara Salvato [22] Srikrishna Sekhar [23,7] and Gauri Sharma[5]

Affiliations are listed at the end of the paper

Accepted 2024 March 1. Received 2024 February 7; in original form 2023 October 11

ABSTRACT

We present the first measurement of H I mass of star-forming galaxies in different large scale structure environments from a blind survey at $z \sim 0.37$. In particular, we carry out a spectral line stacking analysis considering 2875 spectra of colour-selected star-forming galaxies undetected in H I at $0.23 < z < 0.49$ in the COSMOS field, extracted from the MIGHTEE-H I Early Science data cubes, acquired with the *MeerKAT* radio telescope. We stack galaxies belonging to different subsamples depending on three different definitions of large-scale structure environment: local galaxy overdensity, position inside the host dark matter halo (central, satellite, or isolated), and cosmic web type (field, filament, or knot). We first stack the full star-forming galaxy sample and find a robust H I detection yielding an average galaxy H I mass of $M_{\rm H\,I} = (8.12 \pm 0.75) \times 10^9\,{\rm M}_\odot$ at $\sim$11.8σ. Next, we investigate the different subsamples finding a negligible difference in $M_{\rm H\,I}$ as a function of the galaxy overdensity. We report an H I excess compared to the full sample in satellite galaxies ($M_{\rm H\,I} = (11.31 \pm 1.22) \times 10^9$, at $\sim$10.2σ) and in filaments ($M_{\rm H\,I} = (11.62 \pm 0.90) \times 10^9$. Conversely, we report non-detections for the central and knot galaxies subsamples, which appear to be H I-deficient. We find the same qualitative results also when stacking in units of H I fraction ($f_{\rm H\,I}$). We conclude that the H I amount in star-forming galaxies at the studied redshifts correlates with the large-scale structure environment.

Key words: galaxies: formation – evolution – emission lines, cosmology: large-scale structure of Universe.

1 INTRODUCTION

The evolution of galaxies, including mass assembly, star formation, and morphological transformations of galaxies, is known to be strongly connected to the availability of fresh molecular Hydrogen (H_2) supporting the star formation process. Star-forming H_2 clumps arise via gravitational instability and collapse out of large diffuse H I clouds, whose existence represents therefore the necessary condition to trigger the formation of new stars. While part of the galactic H I can be produced by either recombination of ionized Hydrogen (H II) or other reprocessing mechanisms of the internal gas, hydrogen accretion on galaxies from the circumgalactic and intergalactic media plays a fundamental role to ensure the availability of H I reservoirs.

In this context, the environment surrounding a galaxy and its physical conditions may assume a primary importance in regulating gas accretion and removal, galaxy interactions, and other relevant evolutionary phenomena. In fact, galaxy surveys have measured the position of millions of galaxies and demonstrated that at cosmological scales smaller than a few hundred Mpc, matter is no longer uniformly distributed, but forms a filamentary pattern called *cosmic web* (e.g. Bond, Kofman & Pogosyan 1996), which is constituted by filaments interconnecting massive knots and surrounding large voids. In the standard inflationary ΛCDM scenario, the cosmic web naturally emerges from gravitational instability and growth of cosmic structures (e.g. Zel'dovich 1970; Bond et al. 1996), happening as a result of the presence of small inhomogeneities in the primordial matter field. While driven by dark matter, the formation of the cosmic web also involves baryons. As a result, gas exhibits different properties (temperature, rotation and dispersion velocities, and ionization state, among others) depending on the environment it lives in (e.g. Martizzi et al. 2019; Galárraga-Espinosa et al. 2020; Sinigaglia et al. 2021, and references therein), and therefore it may impact the evolution of the galaxies it feeds in a different way. The large-scale structure has been shown to correlate with several galaxy properties, for instance star formation history and quenching (e.g. Darvish et al. 2014; Vulcani et al. 2019; Kraljic et al. 2020; Malavasi et al. 2022, and references therein), colour (e.g. Chen et al. 2017; Pandey & Sarkar 2020), stellar mass and its assembly (e.g. Alpaslan et al. 2016; Chen et al. 2017; Malavasi et al. 2017; Kraljic et al. 2020),

★ E-mail: francesco.sinigaglia@phd.unipd.it

angular momentum magnitude and alignment (e.g. Libeskind et al. 2012; Tempel & Libeskind 2013; Krolewski et al. 2019; Barsanti et al. 2022), and the stellar mass–gas metallicity relation (Donnan, Tojeiro & Kraljic 2022), among others.

In this picture, the role of H I has been investigated only in the nearby universe at $z < 0.1$ (e.g. Kleiner et al. 2017; Crone Odekon et al. 2018; Cortese, Catinella & Smith 2021; Tudorache et al. 2022), while it has remained unexplored at higher redshift due to the difficulty in detecting the 21-cm emission line and the general lack of deep complete spectroscopic surveys covering volumes large enough to reconstruct the cosmological large-scale environment.

Statistical approaches such as spectral line stacking (Zwaan, van Dokkum & Verheijen 2001) can be adopted to exploit the property of Gaussianity of the noise and to extract a global mean H I signal out of the investigated population at the expense of the information about the H I content of the individual galaxies constituting the sample. The spectral line stacking technique has been widely used in the last decades to probe the H I content in galaxies at different redshift, and in particular to unveil correlations such as the presence and abundance of H I in galaxy clusters (Zwaan 2000; Chengalur, Braun & Wieringa 2001; Lah et al. 2009; Healy et al. 2021), H I galaxy scaling relations (Fabello et al. 2011a; Geréb et al. 2015; Brown et al. 2017; Bera et al. 2022; Chowdhury, Kanekar & Chengalur 2022c; Sinigaglia et al. 2022c; Bera et al. 2023; Pan et al. 2023), the H I mass function (Pan et al. 2020; Bera et al. 2022), the $M_{\rm H\,I}$ content of AGN host galaxies (Fabello et al. 2011a; Geréb et al. 2013, 2015), the baryonic Tully–Fisher relation (Meyer et al. 2016), the H I cosmic density evolution with redshift (Lah et al. 2007; Delhaize et al. 2013; Kanekar, Sethi & Dwarakanath 2016; Rhee et al. 2018; Bera et al. 2019; Chowdhury et al. 2020; Chen et al. 2021; Chowdhury et al. 2021; Chowdhury, Kanekar & Chengalur 2022a, b, c), the H I content of galaxy groups, and the $M_{\rm H\,I}$–$M_{\rm halo}$ relation (Guo et al. 2020; Chauhan et al. 2021; Roychowdhury et al. 2022; Dev et al. 2023), among others. Spectral line stacking has been successfully applied to other spectral lines as well, such as CO and CII lines (e.g. Decarli et al. 2018; Bischetti et al. 2019; Jolly et al. 2021; Romano et al. 2022).

The problem of the lack of a deep complete spectroscopic coverage on large sky areas can be alleviated in the case where multiwavelength photometric observations can provide an accurate photometric redshift estimation. In this case, photometric redshifts can be used to perform tomographic analysis on the sky in thick redshift slices. In the COSMOS field – studied in this work – the exceptional amount of work over the past decades has made it possible to collect an incredibly large data set, with photometry ranging from the X-ray to the radio domain (see e.g. Laigle et al. 2016; Weaver et al. 2022).

In this work, we perform a systematic investigation of the correlations between the content of H I in galaxies and different definitions of the large-scale structure environment at a median redshift $z \sim 0.37$, then compare to reference results at $z \sim 0$. We perform a H I stacking experiment and measure the amount of H I in different environments, based on subsamples defined from galaxies located in low-/high-density environments, in field/filaments/knots, or being central/satellite/isolated inside their host dark matter halo.

This paper is organized as follows. In Section 2, we present the details of the MIGHTEE survey and of the H I data we make use of. In Section 3, we summarize the large-scale computations and definitions used throughout the work. Section 4 defines the global galaxy sample we analyse and Section 5 introduces our stacking procedure. In Section 6, we present the results we obtain, and in Section 7, we provide a discussion and contextualization of them. We conclude in Section 8.

Throughout the paper, we assume a cosmology as reported from the most recent results from the Planck satellite, with $\Omega_{\rm m} = 0.311$, $h = 0.677$, $\Omega_k = 0$ (TT,TE,EE+lowE+lensing + BAO, Planck Collaboration 2020).

2 HI DATA FROM MIGHTEE

MeerKAT is the SKA precursor located in South Africa and comprises 64 offset Gregorian dishes (13.5 m diameter main reflector and 3.8 m sub-reflector), equipped with receivers in UHF-band ($580 < \nu < 1015$ MHz), L-band ($900 < \nu < 1670$ MHz), and S-band ($1750 < \nu < 3500$ MHz).

The MeerKAT International GigaHertz Tiered Extragalactic Exploration Large Survey Program (MIGHTEE; Jarvis et al. 2016) is a survey, conducted with the *MeerKAT* radio interferometer (e.g. Jonas & MeerKAT Team 2016). MIGHTEE was targeted at L- and S-bands four deep, extragalactic fields (COSMOS, XMM-LSS, CDFS, ELAIS-S1), characterized by a wealth of multi-wavelength data made available by past and ongoing observational efforts.

The MIGHTEE data are acquired in spectral and full Stokes mode, thereby making MIGHTEE a spectral line, continuum, and polarization survey. In this paper, we make use of the Early Science H I spectral line data from MIGHTEE, presented in Maddox et al. (2021). The observations were conducted between mid-2018 and mid-2019 and targeted $\sim$3.5 deg^2 in the XMM-LSS field and $\sim$1.5 deg^2 in the COSMOS field. These observations were performed with the full array (64 dishes) in L-band, using the 4k correlator mode (209 kHz, corresponding to 52 km s^{-1} at $z = 0.23$ and 56 km s^{-1} at $z = 0.49$). Our analysis is limited to the redshift interval $0.23 < z < 0.49$, excluding the spectral bands covering $0.09 < z < 0.23$ and $z > 0.49$ from our analysis because of bad radio frequency interference (RFI) contamination. We use the MIGHTEE-H I data covering only the COSMOS field data because they are the only data currently available in the redshift range of interest ($0.23 < z < 0.49$) as part of the MIGHTEE Early Science data release. At the investigated redshifts, MIGHTEE-H I data have noise with well-behaved Gaussian properties. The median H I noise rms increases with decreasing frequency, ranging from 85 μJy beam^{-1} at $\nu \sim 1050$ MHz to 135 μJy beam^{-1} at $\nu \sim 950$ MHz.

The MIGHTEE–H I Early Science visibilities were processed with the `ProcessMeerKAT` calibration pipeline.[1] The pipeline is CASA-based and performs standard calibration routines and strategies for the spectral line data such as flagging, bandpass, and complex gain calibration. The continuum subtraction was done in both the visibility and image domain using standard CASA routines UVSUB and UVCONTSUB. Residual visibilities after continuum subtraction were imaged using CASA's task TCLEAN with Briggs' weighting (ROBUST = 0.5). Median filtering was applied to the resulting data cubes to reduce the impact of errors due to continuum subtraction. Table 1 summarizes the details presented in this section.

3 LARGE-SCALE STRUCTURE DEFINITION

In this section, we summarize the procedure used to construct the density field and to compute the large-scale scale properties in the COSMOS field, as derived by Darvish et al. (2015). For more details, we refer the reader to Darvish et al. (2015).

[1] https://idia-pipelines.github.io/docs/processMeerKAT

Table 1. Summary of the details of MIGHTEE-H I data presented in Section 2 and used in this paper.

MIGHTEE-H I data Survey parameter	Value
Field	COSMOS
Area	1.5 deg^2
Integration time	16 h
Frequency resolution	209 kHz
Recession velocity resolution	52 km s^{-1} at $z = 0.23$
Frequency range	0.950–1.050 GHz
Recession velocity range	68 952–146 898 km s^{-1}
Beam (FWHM)	14.5 × 11.0 arcsec^2

3.1 Construction of the density field

In absence of a complete spectroscopic coverage over the full ∼2 deg^2 of the COSMOS field, the density field is constructed from the photometric COSMOS2015 catalogue (Laigle et al. 2016). The redshift uncertainty associated with photometric redshifts (photo-zs hereafter) is too large to accurately unveil redshift-space distortion effects. None the less, an accurate photo-z determination ($\Delta z/(1 + z) \lesssim 0.01$) can still enable a robust density field construction, sufficient to study statistical relations between galaxy properties and the surrounding large-scale environment. Following Darvish et al. (2015), a cut in M_* is applied at $\log_{10}(M_*/\mathrm{M}_\odot) \gtrsim 9.6$ to ensure that the photometric sample is 90 per cent complete in M_*. This guarantees that our sample is volume-limited, and hence, we do not expect biases arising from an evolving mass cut across redshift. We will use this fact later in the paper.

The photo-z estimation for the COSMOS2015 catalogue was performed by means of the spectral energy density fitting technique, over >30 photometric bands from the near-ultraviolet (UV) to the near-infrared (IR) domains. When compared to the spectroscopic sample from the zCOSMOS survey (Lilly et al. 2009), the photometric COSMOS sample turns out to have a median accuracy $\Delta z/(1 + z)_{\mathrm{median}} \sim 0.007$, and to have accuracy $\Delta z/(1 + z) \lesssim 0.01$ out to $z \sim 1.2$ [see fig. 1 of Darvish et al. (2015)].

The density field is constructed in adjacent redshift slices of variable width, and each galaxy is associated with a weight corresponding to the fraction of the photo-z probability distribution of the galaxy lying inside each z-slice, i.e. the likelihood of a galaxy having redshift included in that redshift interval. A galaxy is used to construct the density field only in the z-slices where their weight is ≥ 10 per cent.

After having assigned weights to galaxies, Darvish et al. (2015) employed the weighted adaptative kernel smoothing technique (with weights corresponding to the ones discussed above) to estimate the density field, employing a two-dimensional Gaussian kernel with variable variance depending on the local galaxy density. The adopted global smoothing scale is $s = 0.5$ Mpc, roughly the virial radius of galaxy clusters in the COSMOS field (e.g. Finoguenov et al. 2007).

The density field is estimated via interpolation at the galaxy positions, using RA and Dec as angular coordinates and the maximum-likelihood redshift slice as spectral coordinate.

Fig. 1 shows the projected density field, colour-coded from blue to red at increasing density. Even though we perform analysis at $0.23 < z_{\mathrm{spec}} < 0.49$, in this plot, we consider $0.1 < z_{\mathrm{phot}} < 0.6$, i.e. a larger redshift interval, to take into account the uncertainty in photometric redshift estimation of the galaxies with z_{phot} at the edges of the spectroscopic redshift interval ($z_{\mathrm{phot}} \sim 0.23$ and $z_{\mathrm{phot}} \sim 0.49$). One can visually identify overdense regions with spherical or ellipsoidal shapes (knots) and with elongated one-dimensional shapes (filaments), and underdense regions (field).

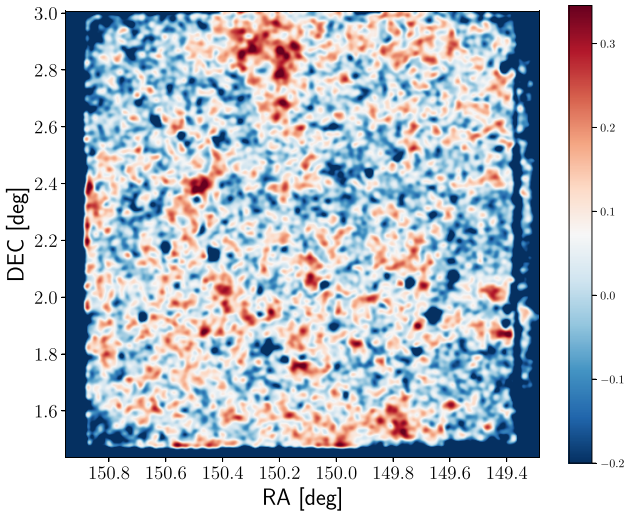

Figure 1. Projected density field, colour-coded from blue to red at increasing density. The plot has been generated by first binning galaxy positions (RA and Dec) of all galaxies in the catalogue with photometric redshift $0.1 < z_{\mathrm{phot}} < 0.6$ in pixels of size $d \sim 10$ arcsec, then applying a two-dimensional Gaussian smoothing with standard deviation $\sigma = 30$ arcsec, and eventually applying a bilinear interpolation. The colourbar is reported in units of $\log_{10}(1 + n_{\mathrm{gal}})$, where n_{gal} is the galaxy number counts per pixel. One can visually identify overdense regions with spherical or ellipsoidal shapes (knots) and with elongated one-dimensional shapes (filaments), as well as underdense regions (field).

3.2 Cosmic web environments definition

The classification of the large-scale cosmic web environment performed by Darvish et al. (2015) consists in the 2D version of the Multi-scale Morphology Filter algorithm (Aragón-Calvo et al. 2007; Darvish et al. 2014). In particular, this algorithm relies on the computation of the local curvature of the density field and the shape of structures, i.e. on geometrical arguments. The local curvature and shape are computed from the eigenvalues λ_i of the rank-2 Hessian matrix (with derivatives computed along the angular coordinates) $\mathcal{H} = \partial_i \partial_j \delta_z(\mathbf{x})$ of the density field δ, at coordinates $\mathbf{x}$ and redshift z. Defining $\lambda_{1,2}$ such that $|\lambda_2| \geq |\lambda_1|$, one defines cosmic web type masks ϵ_k and ϵ_f at coordinates $\mathbf{x}$ as:

(i) $\epsilon_k = 1$ if $\lambda_1 \leq 0$ and $\lambda_2 \leq 0$, 0 otherwise, for knots;
(ii) $\epsilon_f = 1$ if $\lambda_1 \leq 0$, 0 otherwise, for filaments.

This classification implies that the point $\mathbf{x}$ needs to be a local minimum, or a saddle point, in order to sit in a knot or in a filament. If point $\mathbf{x}$ fulfills both the aforementioned $\epsilon = 1$ conditions, one defines the functions:

$$D_k = |\lambda_1|/|\lambda_2|, \; D_f = 1 - D_k \tag{1}$$

to evaluate the local resemblance of an identified structure to a knot and to a filament, respectively. This definition implies that, if $|\lambda_1| \sim |\lambda_2|$, the slope of the density field has a rotationally symmetric structure and the point belongs to a knot ($D_k \sim 1$). Instead, if $|\lambda_1| \ll |\lambda_2|$, a structure is elongated in one dimension and is identified as a filament ($D_f \sim 1$). The Hessian matrix of the density field has been

used as an alternative to the widely used cosmic web classification based on the Hessian matrix of the gravitational tidal tensor and has been shown to be important to parametrize the properties of large-scale structure (see e.g. Heavens & Peacock 1988; Sinigaglia et al. 2021, 2022b).

Subsequently, to control the selection of knots and filaments, the functions $D_{\rm f}$ and $D_{\rm k}$ are non-linearly transformed into $M_{\rm k}$ and $M_{\rm f}$, respectively, as:

$$M_{\rm k} = \exp\left(-\frac{D_{\rm f}}{\beta^2}\right) \tag{2}$$

$$M_{\rm f} = \exp\left(-\frac{D_{\rm k}}{\beta^2}\right) \tag{3}$$

where β is a parameter which regulates the criterion used to select features, set to $\beta = 0.5$ in Darvish et al. (2015). This transform has the effect of enhancing the differences between D_k and D_f, trying to discriminate more clearly between different cosmic web environments. The quantities M_k and M_f should then be seen as boosted knot and filament signals, respectively, and represent auxiliary variables to be used later on in the computations.

Eventually, to maximize the significance of the extracted large-scale structure features and to best separate between such structures and the background, one can not only exploit the signs and the ratios between eigenvalues, but also their magnitude. In fact, the magnitude of eigenvalues will tend to be small for the background due to its fluctuating nature, while it will tend to be large for real structures. This fact can be accounted for by defining the norm of the Hessian as

$$I = 1 - \frac{\sqrt{\lambda_1^2 + \lambda_2^2}}{2\,c^2} \tag{4}$$

where $c = 0.5 \times \max(\sqrt{\lambda_1^2 + \lambda_2^2})$ at each z-slice (Frangi et al. 1998).

The knot and filament probability signals are then evaluated for each pixel on the set of maps obtained using different physical scales L as $S_L = \epsilon \times M \times I$, and eventually the signal S is chosen to be the maximum among the different scales, i.e. $S = \max_L(S_L)$.

At this point, a knot and filament signal (S_k and S_f, respectively) are associated to each galaxy. To conclusively establish what environment a galaxy belongs to, Darvish et al. (2015) found the optimal signal cuts to be parametrized by $t_k = 0.0639 \times z + 0.1142$ and $t_f = 0.0253 \times z + 0.0035$ for knots and filaments, as the best trade-off that guarantees to have a sufficiently large sample and to minimize contamination.

Eventually, a galaxy is classified to belong to a knot, a filament, or to the field as follows:

(i) knot, if $S_k > t_k$ and $S_f < t_f$, or $S_k > t_k$ and $S_f > t_f$ and $S_k > S_f$;
(ii) filament, if $S_k < t_k$ and $S_f > t_f$, or $S_k > t_k$ and $S_f > t_f$ and $S_k < S_f$;
(iii) field, otherwise.

To better understand which density regimes the three classes of cosmic web environment correspond to, Fig. 2 shows the distributions of field (blue), filaments (orange), and knots (green) galaxies as a function of galaxy overdensity[2] (in units $\log_{10}(1+\delta)$, to visually enhance the differences). As expected, the histograms clearly evidence that field galaxies mostly live in low-density regions, filaments cover the intermediate-density regime, while knots galaxies are preferentially found in high-density regions.

[2]We notice that Darvish et al. (2015) denote with'overdensity' $1+\delta$, with $\delta = \rho_{\rm g}/\bar{\rho}_{\rm g} - 1$ and $\rho_{\rm g}$ and $\bar{\rho}_{\rm g}$ the galaxy density and mean galaxy density respectively. For consistency with cosmological studies we refer to only δ as'overdensity'.

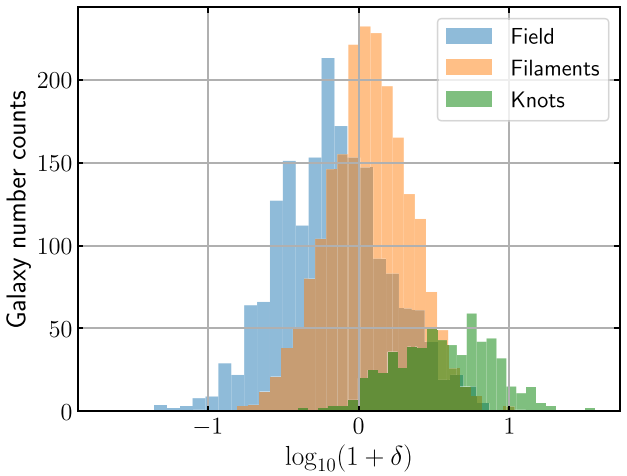

Figure 2. Distributions of field (blue), filaments (orange), and knots (green) galaxies as a function of galaxy overdensity (in units $\log_{10}(1+\delta)$, to visually enhance the differences). As expected, the histograms clearly evidence that field galaxies mostly live in low-density regions, filaments cover the intermediate-density regime, while knots galaxies are preferentially found in high-density regions.

3.3 Central, satellite, and isolated galaxies

To define whether each galaxy in the analysed sample is a central or satellite in a dark matter halo, Darvish et al. (2017) first define a catalogue of galaxy groups and afterwards flag as central the most massive galaxy in a group and as satellite the remaining galaxies of the group. Isolated galaxies, which are found not to be associated with any group, have two possible interpretations. One is that they are central galaxies with too faint satellites to be detected. Alternatively, they can be regarded as satellite galaxies which have been ejected from their group, or as the product of an earlier merger of a galaxy pair.

Galaxy groups are determined via the application of a friends-of-friends algorithm (Huchra & Geller 1982), based on the ansatz that any pair of galaxies closer than a given critical distance (also called *linking length*) belongs to the same group. The definition of the linking length is therefore of primary importance to obtain a reliable galaxy group classification. The angular and redshift separation between two galaxies i and j are defined by Darvish et al. (2017) as

$$\Delta\theta_{ij} \leq b_{\rm ang}\, D_c(z)^{-1}\, n(z)^{-1/2}\,,\ |z_i - z_j| \leq 1/\left(b_z\, \sigma_{\Delta z/(1+z)}\right) \tag{5}$$

where $D_c(z)$ is the comoving distance and $n(z)$ is the median number density of galaxies at redshift z, and $b_{\rm ang}$ and b_z are free parameters set to $b_{\rm ang} = 1.3$ and $b_z = 1.5$. We refer to Darvish et al. (2017) for a thorough discussion of the results.

4 SAMPLE SELECTION

In this section, we present the details about the sample selection.

4.1 Cross-match with spectroscopic catalogues

The galaxy sample we make use of in this work consists of star-forming galaxies at redshift $0.23 < z < 0.49$ in the COSMOS

field (Scoville et al. 2007), with spectroscopic redshift available from public surveys. Compared to the one used in Sinigaglia et al. (2022c), the catalogue has been updated with the addition of new sources. In particular, we cross-match all the galaxies contained in the catalogue compiled by Darvish et al. (2015, 2017; including both star-forming and passive galaxies) with the list of spectroscopic redshifts compiled by querying publicly available catalogues from multiple surveys of the COSMOS field (Khostovan et al., in preparation), and keep only those galaxies for which we were able to identify a spectroscopic counterpart. As shown in Sinigaglia et al. (2022c), the spectroscopic counterpart sample of the COSMOS photometric catalogue is complete (in terms of representativity of the photometric sample) down to $\log_{10}(M_*) \sim 9.5$, roughly the same mass limit used to define the catalogue used in this work. At this point, we select only star-forming galaxies from the resulting sample, based on a rest-frame colour–colour $NUV - r/r - J$ plane selection, extracted directly from the Laigle et al. (2016) catalogue[3] In particular, according to this selection, quiescent galaxies are defined as those featuring $M_{NUV} - M_r > 3(M_r - M_J) + 1$ and $M_{NUV} - M_r > 3.1$, while the remaining galaxies are regarded as star forming.

After applying these cuts, the resulting catalogue is found to contain 1835 more galaxies with respect to the catalogue employed in Sinigaglia et al. (2022c), uniformly spread over the probed redshift range ($0.23 < z < 0.49$).

Through a further cross-match with the radio-selected AGN catalogue built by Smolčić et al. (2017), we estimate the fraction of AGNs in our sample to be ~ 3.5 per cent. We do not exclude AGNs from our sample, and their impact on results from H I spectral stacking will be addressed in future publications of the MIGHTEE collaboration. Eventually, by relying on the spectroscopic redshifts, we also exclude from our sample all galaxies whose H I emission is expected at radio frequencies strongly affected by RFI. While we cross-checked that RFI does not have a major impact on our stacking results, we prefer to limit our sample to RFI-free regions to have a higher degree of control of systematics.

After the application of all the selection criteria described above, our final sample consists of 2875 galaxies, with a median redshift $z \sim 0.37$.

Fig. 3 illustrates the $\log_{10}(M_*)$ (left) and z (right) distributions of galaxies, for different values of the local overdensity field δ ($\delta \leq 0$, $\delta > 0$; first row), position inside the host dark matter halo (central, satellite, or isolated; second row), and the cosmic web environment they belong to (field, filaments, or knots; third row). We also report average and median M_* and z in Table 2. By visually inspecting these distributions, one can easily notice that there are similarities and differences between them. To quantify how similar these distribution are, we have run two-samples one-dimensional Kolmogorov–Smirnov tests comparing galaxy properties distributions in different environments. We report the results in Appendix A, showing the resulting p-values graphically in Figs A1 and A2. In the case of the KS test comparing the M_* distribution for different subsample, $\delta > 0$, centrals and knots galaxies are the subsamples which tend to yield $p < 0.05$. In the case of the KS test comparing the z distributions, we can reject the null hypothesis in most of the cases.

None the less, it may be misleading to conclude that there are selection biases.

In fact, the systematic differences occurring between high-density and low-density environments (both in terms δ and of cosmic web types) are known effects in the literature. The stellar (luminosity) mass functions of star-forming galaxies in high-density environments feature an excess of probability in the high-mass (bright) end and, consequently, a dearth of probability towards low masses (luminosities; see e.g. Bolzonella et al. 2010; Davidzon et al. 2016). Therefore, the differences between the M_* distributions do not arise as a result of selection biases. Rather, they originate as a natural consequence of the diversity of physical processes shaping the mass assembly history in distinct environment. To minimize the impact of such systematic differences, we adopt as target quantity of our stacking not only $M_{\rm H\,I}$ but also the H I fraction $f_{\rm H\,I} \equiv M_{\rm H\,I}/M_*$, as will be explained in more details in Section 5.

As for the z distributions, the Kolmogorov–Smirnov tests tell that we can reject the null hyphotesis in most of the cases. However, we also notice that the deviations between the z distributions of the different subsamples do not show systematic trends, but rather random fluctuations.

Fig. 4 shows the $M_{\rm NUV} - M_{\rm r}$ (left) and $M_{\rm r} - M_{\rm J}$ (right) colour distributions, with the same scheme as Fig. 3. The distributions of different subsamples do not present evident differences, except for a cutoff of the knots galaxies at redder colours towards the bluer end. However, this latter fact is in part due to galaxies in such high overdensities naturally being redder – a well-known fact from the literature – but it is also enhanced by the poor statistics of the knots sample and by the resulting larger statistical fluctuations. In any case, it regards only a tiny fraction of the knots galaxies subsample. This means that the comparison we are performing relates subsamples with consistent underlying colour - and hence star formation rates (SFR) - distributions.

We therefore conclude that there are no obvious selection biases at this point.

4.2 Cosmic variance assessment

For a robust analysis of correlations between the large-scale structure and H I galaxy properties to be robust, we need to address the possible effect of cosmic variance. In fact, the universe starts to be homogeneous on scales $\sim$ hundreds of Mpc. If the volume is not large enough, the results may be biased by the fact that we are looking at a specific cosmic realization.

To address this and to take into account that the surveyed volume extends over a large radial distance, we compute the comoving volume probed by our observations. Given the assumed cosmology, the observed comoving area is $\Delta{\rm RA} \times \Delta{\rm DEC} = 24.2 \times 20.1\,{\rm cMpc}^2$ at $z = 0.23$ and $\Delta{\rm RA} \times \Delta{\rm DEC} = 48.0 \times 40.1\,{\rm cMpc}^2$ at $z = 0.49$, while the total observed comoving radial distance is $d = 949.6\,{\rm cMpc}$. Overall, the total volume is $V = 1.07 \times 10^6\,{\rm cMpc}^3$, equivalent to a cubic volume with side $l \sim 102.3\,{\rm cMpc}$. We stress this volume corresponds to the volume spanned by the spectroscopic sample, not by the entire MIGHTEE data cube, the latter being in principle much larger. In fact, existing spectroscopic surveys in the COSMOS field beyond the nearby universe cover mostly the central part of the field, leaving the angular outskirt of the MIGHTEE data unprobed.

Even though the total volume is not large enough to host outlier rare cosmic structures such as super-clusters or super-voids, it

[3]We notice that, even though a more recent COSMOS photometric catalogue has been made publicly available (Weaver et al. 2022), we prefer to stick to the Laigle et al. (2016) catalogue for consistency with the computations performed by Darvish et al. (2015, 2017). Also, at the probed redshift, the gain in NIR photometry—which is the main progress between the Laigle et al. (2016) and the Weaver et al. (2022) catalogues—is not significant. What drives the accuracy at this redshift are the medium bands, and they did not change from one catalogue to the other.

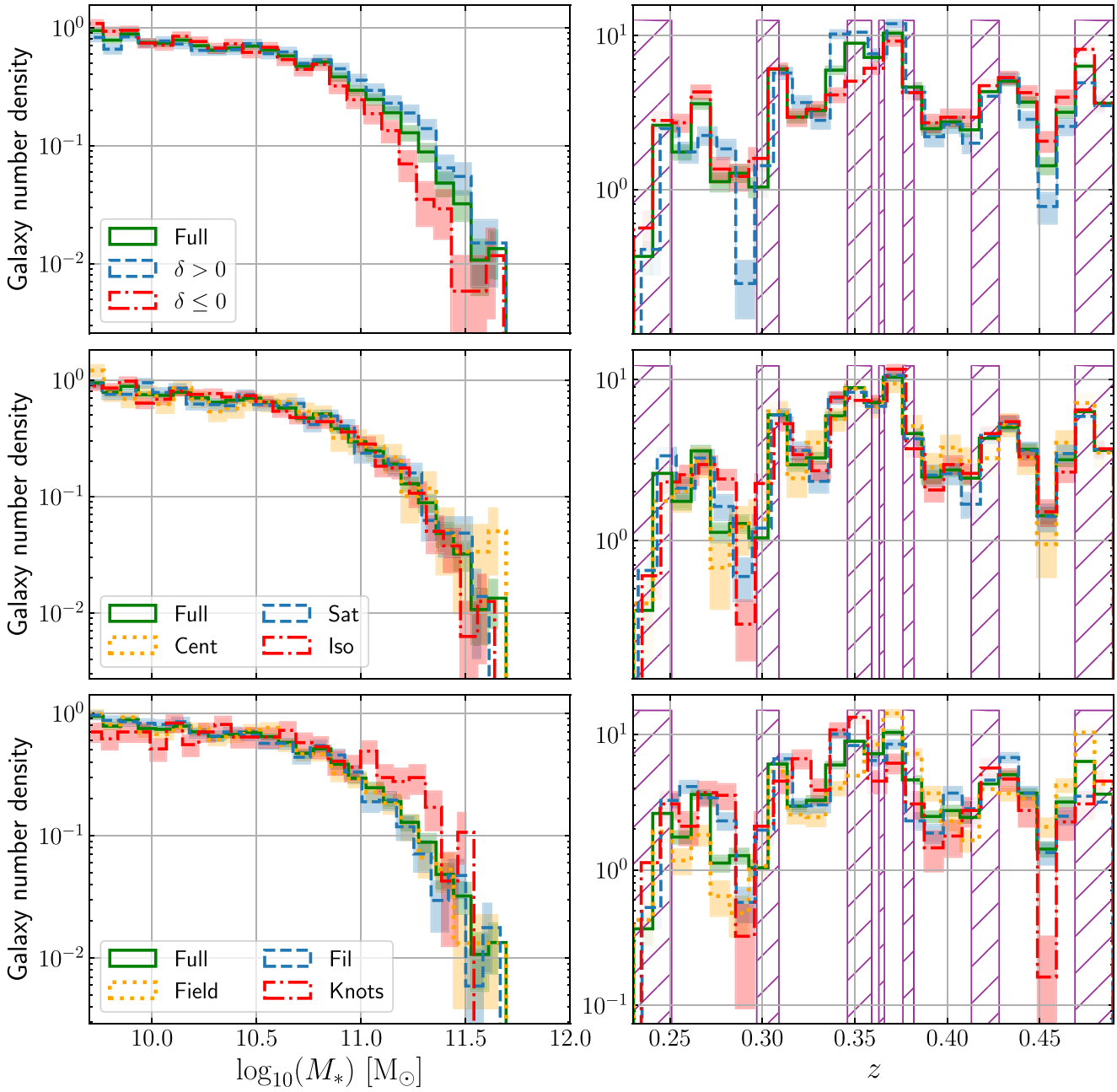

Figure 3. $\log_{10}(M_*)$ (left) and z (right) distributions of the full sample (green solid) and different subsamples investigated in this work. Top row: high-density $\delta > 0$ (blue dashed) and low-density $\delta \leq 0$ (red dashed–dotted) galaxies. Mid row: central (yellow dotted), satellite (blue dashed), and isolated (red dashed–dotted) galaxies. Bottom row: field (yellow dotted), filaments (blue dashed), and knots (red dashed–dotted) galaxies. Poisson uncertainties are shown as shaded areas. In the right column, purple hatched areas indicate frequency intervals strongly affected by RFI, for which we adopt the RFI flagging and masking described in Section 5.

encompasses a volume comparable to state-of-the-art cosmological hydrodynamic simulations. Therefore, we conclude that, while the probed volume is not very large in a cosmological sense and some cosmic variance effects may still be present, this should not constitute a major issue.

5 STACKING PROCEDURE

Throughout the paper, we adopt a standard spectral line stacking procedure (see e.g. Healy et al. 2019; Sinigaglia et al. 2022a). We start by extracting H I cubelets around the galaxies in our sample from the full data cubes, with aperture (3 ×FWHM of the beam major axis, 3 ×FWHM of the beam major axis, ±2000 km s^{-1}) in (RA, Dec., velocity). We choose these apertures to ensure that the whole flux emitted by galaxies is included in the cubelets. The angular aperture corresponds to ∼130 physical kpc at $z = 0.23$ (i.e. minimum aperture), larger than the typical H I disk size (see e.g. Wang et al. 2016; Rajohnson et al. 2022, for the $M_{\rm HI}$ − size relation at $z = 0$). This choice allows us to not underestimate the flux, which may happen if a smaller aperture is chosen instead, and leaves us only

Table 2. Results from the stacking runs performed on different galaxy subsamples. The first column lists the investigated case, the second column reports the number of stacked galaxies for each case, the third column the (uncorrected, i.e. as measured from the data) resulting $M_{\rm H\,I}$ and its associated uncertainties (*stands for 3σ upper limit in the case of non-detection), the fourth column the SNR, the fifth column the percentage confusion correction that we apply to the mass measurement, the sixth column the corrected (by confusion and RFI-masking uncertainty) $M_{\rm H\,I}$ measurements and their associated uncertainties, and the seventh column the figure and panel where the related spectrum is shown.

Sample	N. gal.	$M_{\rm H\,I}$ $[\times 10^9\,{\rm M}_\odot]$	SNR	Conf.	$M_{\rm H\,I,\,corr}$ $[\times 10^9\,{\rm M}_\odot]$	$f_{\rm H\,I}$	$\langle M_* \rangle$ $[\times 10^9\,{\rm M}_\odot]$	$M_{*,\,\rm med}$ $[\times 10^9\,{\rm M}_\odot]$	$\langle z \rangle$	$z_{\rm med}$
Full	2875	9.02 ± 0.76	11.8	10 %	8.12 ± 0.75	0.55 ± 0.05	32.93 ± 10.71	16.41 ± 10.71	0.360	0.370
$\delta \leq 0$	1522	7.86 ± 1.04	7.5	4 %	7.54 ± 1.10	0.53 ± 0.11	26.09 ± 8.55	14.17 ± 8.55	0.360	0.371
$\delta > 0$	1353	10.08 ± 1.04	9.6	18 %	8.27 ± 0.94	0.69 ± 0.08	39.00 ± 13.51	19.82 ± 13.51	0.359	0.368
Centrals	451	5.31*	<3	10 %	5.31*	0.61*	35.95 ± 10.35	15.86 ± 10.35	0.362	0.371
Satellites	1172	13.46 ± 1.32	10.2	16 %	11.31 ± 1.22	0.79 ± 0.08	32.75 ± 10.67	16.19 ± 10.67	0.360	0.369
Isolated	1252	8.37 ± 0.90	9.3	4 %	8.04 ± 0.95	0.69 ± 0.09	32.01 ± 10.87	17.07 ± 10.87	0.359	0.370
Field	1122	6.17 ± 1.21	5.1	5 %	5.86 ± 1.27	0.42 ± 0.10	31.11 ± 10.31	15.75 ± 10.31	0.371	0.373
Filaments	1324	12.77 ± 0.90	14.1	9 %	11.62 ± 0.90	0.93 ± 0.09	31.11 ± 9.79	15.48 ± 9.79	0.354	0.362
Knots	369	4.11*	<3	36 %	4.11*	0.39	45.28 ± 15.90	22.74 ± 15.90	0.348	0.344

with the problem of subtracting the flux contamination by nearby sources. Optical coordinates and spectroscopic redshifts are used to define the center of the cubelets.

Afterwards, we integrate each cubelet over angular coordinates, obtaining a spectrum. Each spectrum at observed frequency $\nu_{\rm obs}$ is then de-redshifted to its rest-frame frequency $\nu_{\rm rf}$ through $\nu_{\rm rf} = \nu_{\rm obs}(1 + z)$ and converted to units of velocity as $v = cz$. As the frequency bin width is fixed throughout the cubes ($\Delta\nu = 209$ kHz), velocity bins have different widths as a function of redshift. Therefore, to ensure that all the spectra are binned the same manner in the spectral direction, spectra are resampled to a reference spectral template, with velocity bin width $\Delta v = 100\,{\rm km\,s^{-1}}$.

We convert spectra from units of flux to units of $M_{\rm H\,I}$ (per velocity channel; e.g. Zwaan et al. 2001):

$$M_{\rm HI}(v) = (2.356 \times 10^5)\, D_{\rm L}^2\, S(v)\, (1+z)^{-1}\,{\rm M_\odot\,km^{-1}\,s} \tag{6}$$

where $D_{\rm L}$ is the luminosity distance of the galaxy in units Mpc, $S(v)$ is the 21-cm spectral flux density in units Jy and $(1 + z)^{-1}$ is a correction factor accounting for the flux reduction due to the expansion of the Universe. Lastly, we co-add all the spectra together. The stacked spectrum can then be expressed as

$$\langle M_{\rm HI}(v) \rangle = \frac{\sum_{i=0}^{n_{\rm gal}} M_{{\rm HI},i}(v) \times w_i \times f_i}{\sum_{i=0}^{n_{\rm gal}} w_i \times f_i^2} \tag{7}$$

where $n_{\rm gal}$ is the number of co-added spectra, and f_i and w_i are the average primary beam transmission and the weight assigned to each source. This equation implements primary beam correction following the procedure detailed in Geréb et al. (2013). Eventually, we integrate the resulting stacked spectrum over $M_{\rm H\,I}$, in order to obtain an average $M_{\rm H\,I}$ estimate. We fix the integration limits to $v_{\rm lim} \pm 500\,{\rm km\,s^{-1}}$.

Moreover, we perform our stacking experiments assuming as target quantity for the spectra not only $M_{\rm H\,I}$, but also $f_{\rm H\,I}$, as anticipated in Section 4. For the latter case, we divide each spectrum – given in units $M_{\rm H\,I}$ – by the M_* of the galaxy. In this way, from the final stacked spectrum we are able compute the average $\langle f_{\rm HI} \rangle = \langle M_{\rm HI}/M_* \rangle$.

A customary choice for the weighting scheme is to weight each spectrum for some power of the inverse of its variance. That is, one can weight the ith spectrum by $w_i = 1/\sigma_i^\gamma$, where σ_i is the rms of the ith spectrum. The cases $\gamma = 1$ (Lah et al. 2007) and $\gamma = 2$ (Fabello et al. 2011b) have been widely employed and tested in the literature. Some authors have proposed to also weight each spectrum by distance (see e.g. Delhaize et al. 2013; Hu et al. 2019). Here we adopt the perspective of Hu et al. (2019) of leaving γ as a free parameter (the exponent was applied to distance in the case of Hu et al. (2019)), and study the evolution of the SNR and of $M_{\rm H\,I}$ as a function of γ. We report the results of this test in Appendix B. It turns out that the choice $\gamma = 1$ maximises the SNR in all the studied subsamples. Therefore, we choose $\gamma = 1$. Since we are working with a volume-limited sample, this does not induce a selection bias due to the evolution of the M_* with redshift. Furthermore, we observe that $M_{\rm H\,I}$ slightly decreases with increasing γ. This latter effect is due to two different facts occurring simultaneously:

(i) The root mean square (rms) per channel slowly increases towards to the lower frequency part of the covered band. As a result, weights display an anticorrelation with redshift. This can be clearly seen in the left-hand panel of Fig. B2, where we plot the weights (assuming $\gamma = 1$) as a function of redshift as blue circles. The anticorrelation can be appreciated by looking at the orange dashed curve, representing the median of the weights in different redshift bins. The existing correlation between weights and redshift indicates that the larger γ the more the weighting scheme downweights galaxies at higher redshift;

(ii) $M_{\rm H\,I}$ is found to undergo a significant evolution with redshift across the probed redshift range. To demonstrate this, we have subdivided the full redshift range into two redshift subsamples – a low-redshift subsample at $0.23 < z < 0.35$ and a high-redshift subsamples at $0.35 < z < 0.49$. In both cases, we find robust detections, the former at $\sim 6.6\sigma$, the latter at $\sim 9.9\sigma$. The resulting average H I masses that we measure from the stacks are $M_{\rm HI} \sim (5.91 \pm 0.90) \times 10^9\,{\rm M_\odot}$ and $M_{\rm HI} \sim (11.64 \pm 1.18) \times 10^9\,{\rm M_\odot}$ for the low-z and high-z subsamples, respectively. I.e., the difference in $M_{\rm H\,I}$ is a factor ~ 2. These findings evidence a strikingly large evolution with redshift of $M_{\rm H\,I}$. It is hard to conclusively tell whether this difference is entirely due to actual redshift evolution, or whether some non-obvious selection or cosmic variance effects contribute to enlarge the difference. We will follow-up this aspect in future work.

The combination of these two effects causes a downweighting of the H I-rich galaxies at higher redshift when we apply a weighting scheme.

We notice that this effect probably has an irrelevant impact on low-z stacking experiments, where data typically are found to have a much more stable and weakly evolving rms per channel as a function of frequency. However, this aspect becomes important in cases where a large bandwidth with and evolving noise level is surveyed, such as the present case.

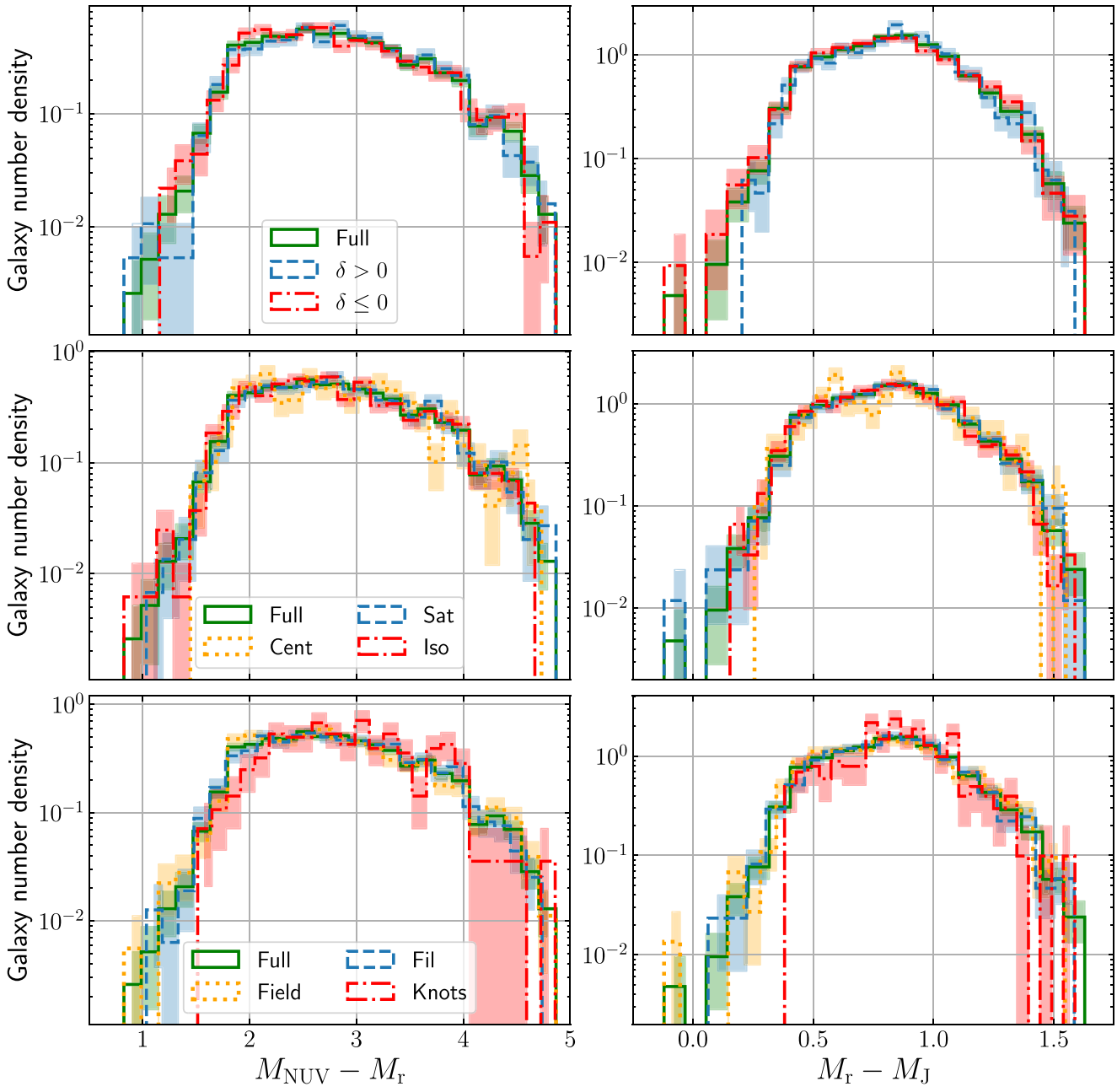

Figure 4. Same as Fig. 3, but for rest-frame colours $M_{\rm NUV} - M_{\rm r}$ (left) and $M_{\rm r} - M_{\rm J}$ (right).

Despite all the arguments we have laid down above, we notice that (i) the change from $\gamma = 0$ (unweighted case) to $\gamma = 1$ is relatively tiny, and (ii) the variation occurs coherently in all the subsamples. Therefore, while the dependence of $M_{\rm H\,I}$ on the weighting scheme may induce a slight variation in the overall normalization, it does not change the relative differences in $M_{\rm H\,I}$ (within reasonable fluctuations). This allows us to draw conclusions on the differences in $M_{\rm H\,I}$ between the different subsamples without worrying about the choice of the weighting scheme.

To evaluate the uncertainty associated with our measurements, we rely on the baseline of the spectra as follows. The 1σ noise uncertainty (in units $M_{\rm H\,I}$) is evaluated by computing the rms of the off-line noisy channels $\sigma_{\rm rms}$ of the stacked spectrum, i.e. those channels outside the spectral interval integrated to compute $M_{\rm H\,I}$.

To further confirm the legitimacy of our detection, we also generate a reference spectrum obtained by stacking noise spectra (one noise spectrum per galaxy) extracted at randomized positions. The positions of the noise spectra are obtained by adding a fixed angular offset to the centre of each galaxy cubelet in a random direction and defined over the same spectral range as the corresponding galaxy cubelet. The angular offset (100 arcsec) is chosen to guarantee that the reference spectrum is extracted close to the galaxy spectrum, although without overlaps. Also, we double-check that the reference spectrum of each galaxy has no overlaps with other

known optical galaxies, reject it and draw a new one if there is any overlap.

We compute the integrated signal-to-noise ratio of the final stacked spectrum as:

$$\mathrm{SNR} = \sum_{i}^{N_{\mathrm{ch}}} \langle S_i \rangle \, / (\sigma_{\mathrm{rms}} \sqrt{N_{\mathrm{ch}}}) \tag{8}$$

where $\langle S_i \rangle$ is the stacked spectrum, and N_{ch} is the number of channels over which the integration is performed (e.g. Healy et al. 2019).

To mitigate the impact of RFI on our stacking results, we adopt the following procedure. We first identify frequency bands severely characterized by strong RFI by studying the variation of the rms per channel as a function of frequency. After identifying bad RFI frequency windows, we first exclude from the sample galaxies with central frequency falling within one of these frequency intervals. Eventually, because a spectrum considered as valid based on the previous criterion can still have some portions overlapping to RFI-affected channels, we mask those portions by setting the flux to zero.

Since this is the first time (to the knowledge of the authors) that this local masking procedure is applied to RFI-affected channels, we address its impact on the results by comparing the values of the average $M_{\mathrm{H\,I}}$ obtained from stacking in the different subsamples before and after masking. We report the results of this test in Appendix C, Table C1. It turns out that such a masking do not induce a systematic effect in the probed subamples, but rather a stochastic deviation from the unmasked case. In addition, the magnitude of the deviation is typically tiny, with a maximum (in absolute value) of -9.2 per cent for the $\delta > 0$ subsample. However, as can be seen in Table C1, the SNR always increases after masking. This demonstrates that the off-line RFI masking is advantageous in terms of gain in SNR, and does not introduce systematics in the stacking procedure. We treat the random deviation introduced by the implementation of such a masking by adding a conservative 10 per cent uncertainty in quadrature to the error coming from the measurement.

We address the problem of flux contamination due to source confusion using detailed *MeerKAT*-like simulated data cubes, built with the same setup as MIGHTEE-H I observations. In particular, we use the Obreschkow & Meyer (2014) flux-limited mock galaxy catalogue, based on the SKA Simulated Skies semi-analytic simulations (S^3-SAX), and therefore on the physical models described in Obreschkow et al. (2009a, b, c), to inject galaxies with realistic H I masses and clustering into a blank synthetic datacube matching the same angular and spectral size as our observations (see also Elson, Blyth & Baker 2016; Elson, Baker & Blyth 2019). Then, using the same methodology presented in Elson et al. (2016), we decomposed the spectrum extracted for each target galaxy into contributions from the actual target, and contributions from nearby contaminating galaxies. Following this procedure, we estimate the average level of contamination for the full sample to be ~ 10 per cent. As we are explicitly investigating environmental trends, which implies considering galaxies lying in regions with different clustering properties, we need to take into account that this fact will be translated into a different confusion contribution, depending on whether a galaxy lives e.g. in a low-density or in a high-density environment. While repeating the same computation performed by Darvish et al. (2015, 2017) to define the large scale structure environments on the simulated data cubes goes beyond the scope of the paper, we handle the confusion correction by rescaling the average detected H I mass of each subsample by the ratio between the mean overdensity in the regions where galaxies in the subsample live and the mean overdensity of the full sample. I.e.:

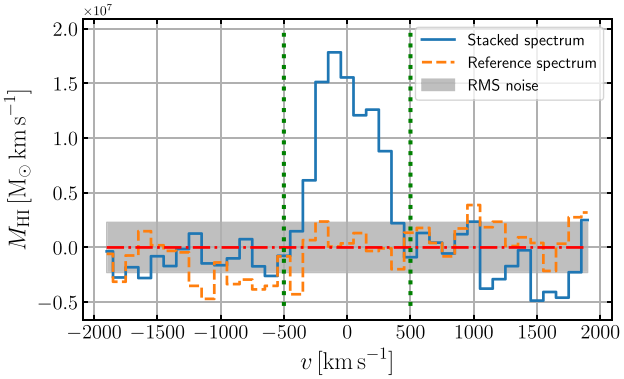

Figure 5. $M_{\mathrm{H\,I}}$ stacked spectrum obtained from the full sample. The blue solid line represents the stacked spectrum, the orange dashed line and the grey shaded area represent the stacked reference spectrum and its rms, respectively, and the green dotted vertical lines represent the mass integration limits.

$$\Delta M_{\mathrm{H\,I,conf,i}} = \frac{\Delta M_{\mathrm{H\,I,conf,tot}} \times \delta_i}{\delta_{\mathrm{tot}}}, \tag{9}$$

where $\Delta M_{\mathrm{H\,I,\,conf,\,i}}$ is the final confusion contribution to be subtracted to each subsample i, δ_i is the average overdensity at the position of each galaxy in the ith subsample, $\Delta M_{\mathrm{H\,I,\,conf,\,tot}}$ is the confusion contribution from the full sample ($\Delta M_{\mathrm{H\,I,conf,tot}} \sim 10$ per cent, obtained from simulations as described above), and δ_{tot} is the average overdensity at the position of each galaxy for the full sample. This computation is applied to all the subsamples investigated in this work (full, $\delta > 0/\delta \leq 0$, central/satellites/isolated, field/filaments/knots), and the quantitative results are expressed in percentage and listed in Table 2.

6 RESULTS

We report on the results we obtain applying the methodology described in the previous sections, and discuss them in the light of previous findings presented in literature.

Figs. 5–8 report our resulting stacked spectra for the full sample, and for subsamples based on galaxy overdensity field, galaxy position inside the host dark matter haloes, and cosmic web environment, respectively. In all cases, we display the stacked spectrum as a blue solid line, the stacked reference spectrum and its rms as an orange dashed line and a grey shaded area, respectively, and the two vertical lines marking the integration limits as green dotted lines. We report inside each panel the case each spectrum corresponds to, while we present in Table 2 the numerical findings, i.e. the number of stacked galaxies, the estimated average $M_{\mathrm{H\,I}}$ and associated uncertainty as measured from the data and after correcting for source confusion and RFI-masking uncertainty, the confusion correction that we apply, the SNR, the average and median M_* of each subsample, and the average and median redshift of each subsample, the median M_*.

Overall, we find robust $>5\sigma$ H I detections in all the studied case, except for the central and knot galaxies subsamples, for which we find a non-detection. The resulting stacked spectra, after applying the careful RFI flagging described in Section 5, appear to be well-behaved, and do not display major anomalous features.

In addition, we quantify the statistical significance of the deviation of the measured average $M_{\mathrm{H\,I,\,X}}$ (average $f_{\mathrm{H\,I,\,X}}$) of a subsample X from the measured average $M_{\mathrm{H\,I,\,Y}}$ (average $f_{\mathrm{H\,I,\,X}}$) of a subsample Y

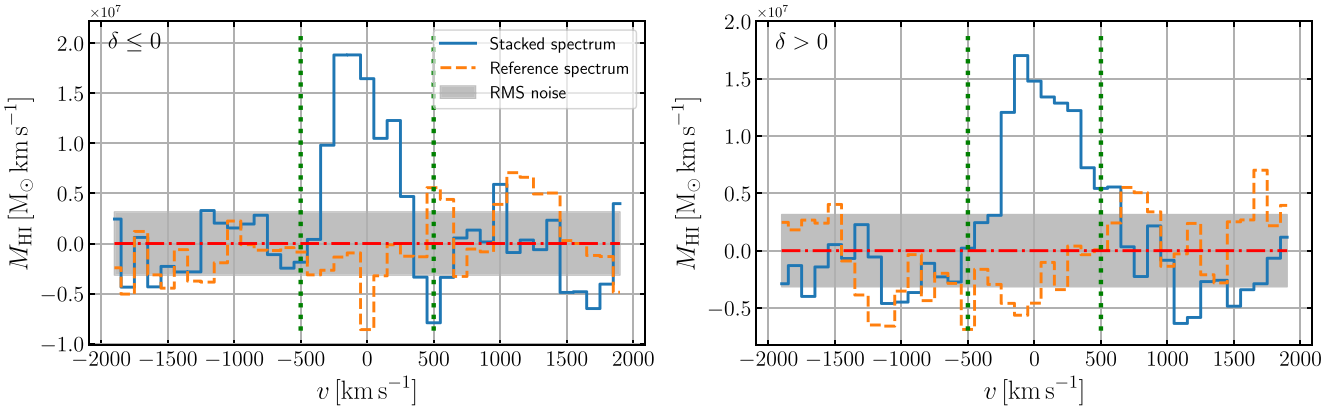

Figure 6. $M_{\rm H\,I}$ stacked spectrum obtained from the subsamples of galaxies in underdense ($\delta \leq 0$, left) and overdense ($\delta > 0$, right) regions, respectively. Symbols as in Fig. 5.

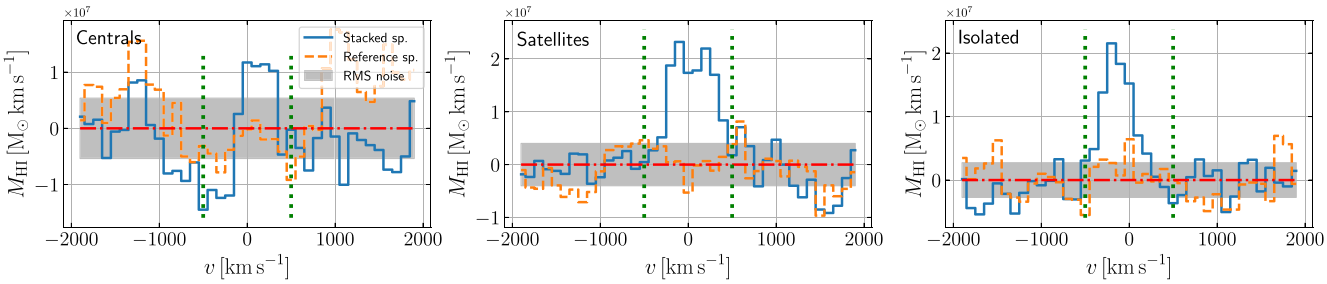

Figure 7. $M_{\rm H\,I}$ stacked spectrum obtained from the subsamples of central (left), satellite (centre), and isolated (right) galaxies, respectively. Symbols as in Fig. 5.

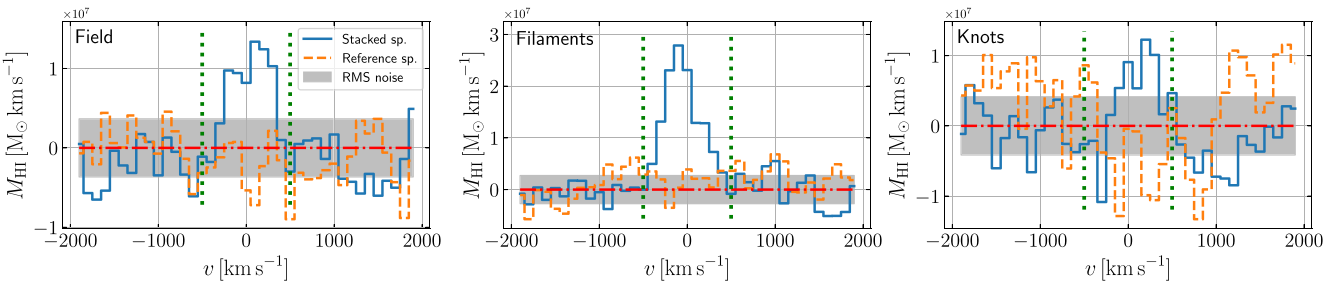

Figure 8. $M_{\rm H\,I}$ stacked spectrum obtained from the subsamples of galaxies sitting in the field (left), in filaments (centre), or in voids (right). Symbols as in Fig. 5.

as:

$$s = \frac{|M_{\rm HI,X} - M_{\rm HI,Y}|}{\sqrt{\sigma_{\rm X}^2 + \sigma_{\rm Y}^2}} \quad , \tag{10}$$

where $\sigma_{\rm X}$ and $\sigma_{\rm Y}$ are the uncertainties on $M_{\rm H\,I}$ ($f_{\rm H\,I}$) of subsamples X and Y, respectively. We graphically report the results in Figs 9 and 10.

We show the results of our stacking experiments in the $M_{\rm H\,I} - M_*$ plane in Fig. 11. We plot the results differentiating between distinct definitions of environment: overdensity field (top), central/satellite/isolated (middle), field/filament/knot galaxies (bottom). We also plot the scaling relations at $z \sim 0.37$ from Sinigaglia et al. (2022c) (hereafter S22) and the one at $z \sim 1$ from Chowdhury et al. (2022c) (hereafter C22). Since we are performing stacking, we adopt as estimate for the M_* of each subsample its average M_* (reported in Table 2).

We start by commenting on the results as a function of the definition of environment based on the galaxy overdensity field. Here, high- (Fig. 6, right-hand panel) and low-density (Fig. 6, left-hand panel) $M_{\rm H\,I}$ measurements tell that galaxies sitting in high-density environments are slightly H I-richer than the ones living in low-density regions. None the less, the measurements in the two subsamples are compatible within uncertainties with the $M_{\rm H\,I}$ estimate obtained from the full sample ($<\sim 0.5\sigma$ in both cases), and are compatible with one another (at $\sim 0.5\sigma$). The same results qualitatively apply also in terms of $f_{\rm H\,I}$, with larger significance ($\sim 1.5\sigma$ significance between the $\delta > 0$ and the full sample, and $\sim 1.2\sigma$ between the $\delta > 0$ and the $\delta \leq 0$ subsamples).

Then, we analyse the results related to the definition of environment based on galaxy position inside the host dark matter halo. It turns out that:

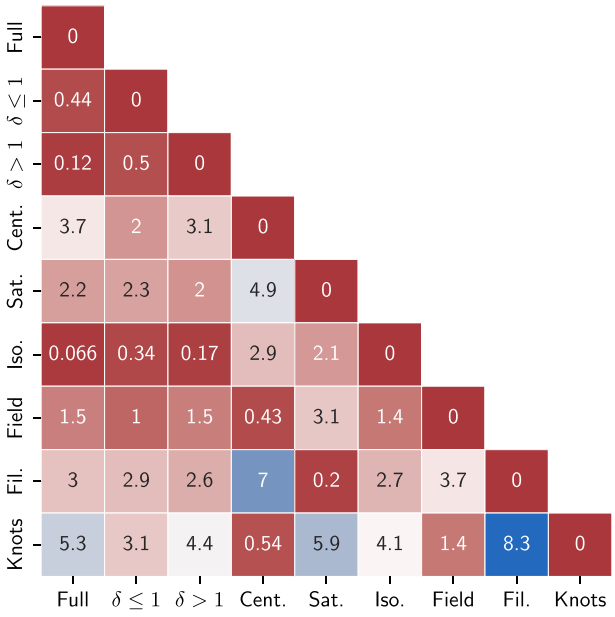

Figure 9. Statistical significance of the deviation in the average $M_{\rm H\,I}$ between different subsamples. The plot is colour-coded from red to blue from the lowest to the highest significance values.

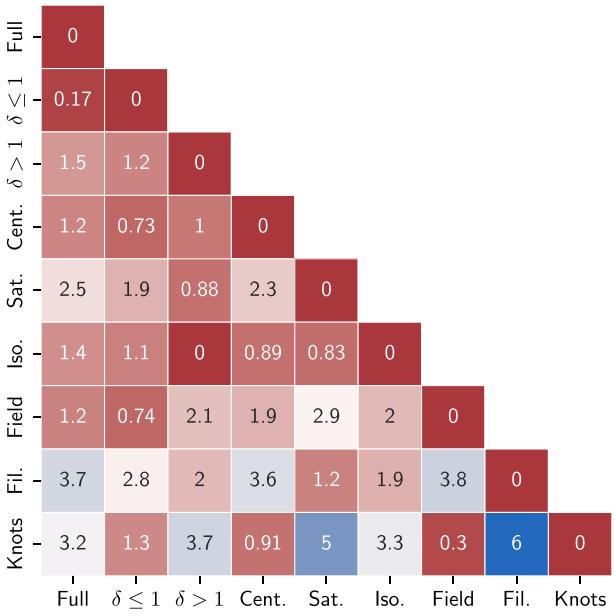

Figure 10. Statistical significance of the deviation in the average HI fraction $f_{\rm H\,I}$ between different subsamples. The plot is colour-coded from red to blue from the lowest to the highest significance values.

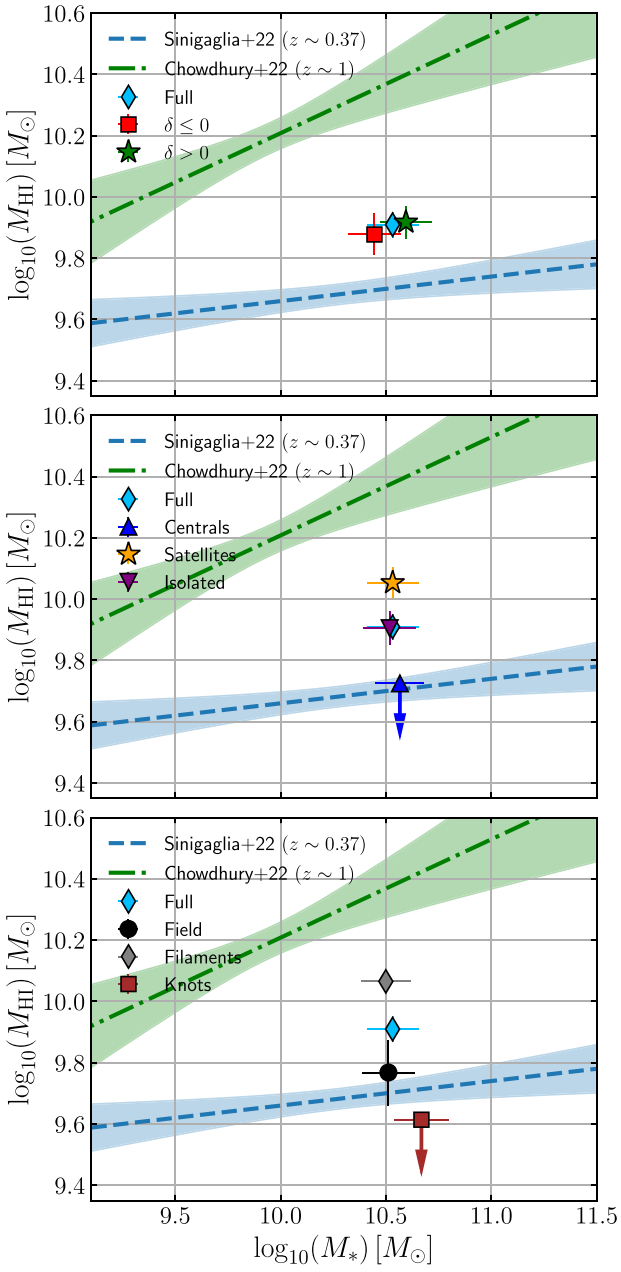

Figure 11. $M_{\rm H\,I}$ results as obtained from our stacking analysis in different definitions of environment. We compare our results to the $M_{\rm H\,I} - M_*$ scaling relations presented in S22 at $z \sim 0.37$ and C22 at $z \sim 1$. We report the findings related to different definitions of environment in different rows: galaxy overdensity (top), central/satellite/isolated (mid), field/filaments/knots (bottom). We also always overplot the data point corresponding to the full sample, for comparison purposes.

(i) Isolated galaxies subsamples feature an average $M_{\rm H\,I}$ extremely similar to the one measured from the full sample and compatible within statistical uncertainties with it. None the less, here again when looking at $f_{\rm H\,I}$ we find the isolated galaxies to be $\sim 1.4\sigma$ H I-richer than the full sample;

(ii) Central galaxies yield a non-detection both for the $M_{\rm H\,I}$ and the $f_{\rm H\,I}$ spectra. In this case, we place a 3σ upper limit on $M_{\rm H\,I}$ and $f_{\rm H\,I}$. The former is found to be a factor ~ 1.5 H I poorer than the full sample in $M_{\rm H\,I}$, corresponding to a $\sim 3.7\sigma$ deviation in $M_{\rm H\,I}$. Conversely, the upper limit for $f_{\rm H\,I}$ actually exceeds the H I fraction of the full sample, which implies that we are not able to draw any conclusion in that sense;

(iii) Satellite galaxies are found to be H I rich, exceeding both the $M_{\rm H\,I}$ and the $f_{\rm H\,I}$ of the full sample by a factor ~ 1.4. This implies a deviation of $\sim 2.2\sigma$ in $M_{\rm H\,I}$ and $\sim 2.5\sigma$ in $f_{\rm H\,I}$. Moreover, satellite galaxies are H I-richer than the upper limit for central galaxies by a

factor $\sim$2.1 in $M_{\rm H\,I}$ and a factor $\sim$1.3 in $f_{\rm H\,I}$, at $\sim 4.9\sigma$ and $\sim 2.3\sigma$, respectively.

Finally, we examine the $M_{\rm H\,I}$ measurements obtained by splitting the full sample into subsamples depending on field, filaments, or knots membership. We find this definition of environment to be the one that correlates more with the $M_{\rm H\,I}$ of galaxies. In this case:

(i) Field galaxies are found to be a factor $\sim$1.4 H I-poorer than the full sample in $M_{\rm H\,I}$ and a factor $\sim$1.3 H I-poorer in $f_{\rm H\,I}$, at $\sim 1.5\sigma$ and $\sim 1.2\sigma$, respectively;
(ii) filament galaxies are found to be a factor $\sim$1.4 H I-richer than the full sample in $M_{\rm H\,I}$ and a factor $\sim$1.7 H I-richer in $f_{\rm H\,I}$, at $\sim 3\sigma$ and $\sim 3.7\sigma$, respectively;
(iii) Knots galaxies yield a non-detection. Also in this case, we place a 3σ upper limit on $M_{\rm H\,I}$ and $f_{\rm H\,I}$. Such an upper limit is a factor $\sim$2 H I-poorer than the full sample in $M_{\rm H\,I}$ and a factor $\sim$1.4 H I-poorer in $f_{\rm H\,I}$, at $\sim 5.3\sigma$ and $\sim 3.2\sigma$, respectively. Moreover, the difference between knot and filament galaxies (taking the upper limit as reference for knot galaxies) is a factor $\sim$2.8 in $M_{\rm H\,I}$ and a factor $\sim$2.4 in $f_{\rm H\,I}$, at $\sim 8.3\sigma$ and $\sim 6\sigma$, respectively

As expected, the source confusion correction that we compute through equation Section 9 is larger the denser the environment, with a minimum of $4-5$ per cent for $\delta \leq 0$, isolated, and field galaxies, whereas the correction is as large as $\sim$ 36 per cent in knots. We also notice that we do not actually apply the confusion correction to the non-detections, as it would be meaningless.

7 DISCUSSION

Overall, we report a tendency of intermediate-density cosmic web environments to host galaxies with larger $M_{\rm H\,I}$ and satellite galaxies to be more H I-rich than central galaxies, the latter at (nearly) fixed stellar mass (but not at fixed halo mass). However, the $M_{\rm H\,I}$ correlates more with the cosmic web environments rather than with local galaxy overdensity, suggesting that environment-specific processes are in act in shaping the H I content of galaxies sitting therein.

7.1 Central, satellite, and isolated galaxies

As anticipated, our findings at $z \sim 0.37$ suggest that satellite galaxies are HI-richer than central galaxies and isolated galaxies at fixed stellar mass $\log_{10}(M_*/{\rm M}_\odot) \sim 10.2$ ($\sim 2.5\sigma$). However, it is not straightforward to interpret these results, since we are here performing a comparison at (nearly) fixed stellar mass, and not at fixed halo mass. This implies that the subsample of satellite galaxies we are investigating here are likely to be, on average, satellite of more massive central galaxies. For instance, following Shuntov et al. (2022), at $0.2 < z < 0.5$ central galaxies of stellar mass $\log_{10}(M_*/{\rm M}_\odot) \sim 10.2$ belong to haloes of mass $\log_{10}(M_{\rm h}/{\rm M}_\odot) \sim 12$, while satellite galaxies belong to haloes of mass $\log_{10}(M_{\rm h}/{\rm M}_\odot) \sim 12.5$–13 with characteristic central galaxy stellar mass $\log_{10}(M_*/{\rm M}_\odot) \sim 10.7$–11. Such a study was performed without splitting the sample into star-forming and passive galaxies, so both centrals and satellites in our sample of star-forming galaxies are likely to belong to haloes of lower masses with respect to the figures quoted above. However, the result remains qualitatively the same.

To put our results into context, we compare them to previous findings presented in the literature. Using xGASS data (Catinella et al. 2018) at $z \sim 0$, Janowiecki et al. (2017) find that for $\log_{10}(M_*/{\rm M}_\odot) < 10.2$, central galaxies in groups feature a H I fraction comparable to the one of isolated galaxies at the same M_* as the one probed in this work and even smaller H I fraction at larger M_*. The same trend with M_* is also mimicked by the SFR. It is important to notice that Janowiecki et al. (2017) analyse the full galaxy population (star-forming + quenched), while here we are focusing only on star-forming galaxies. By taking into account this difference and the fact that at $z \sim 0$ star-forming galaxies are known to be H I-richer than passive galaxies, the results found in this work appear to agree with the ones of Janowiecki et al. (2017).

Our results appear to also be in contrast with the combination of the observational results by Guo et al. (2021) and Brown et al. (2017) and of the simulation-based results by Stevens et al. (2019), showing that at the fixed average M_* studied in this work satellite galaxies at $z \sim 0$ are H I-poorer than central galaxies (see also Cortese et al. 2021, for a review). However, again such studies investigate the full galaxy population, while we restrict our work to star-forming galaxies. Therefore, it is not possible to establish a fair comparison between those references and our results.

7.2 Galaxies in knots, filaments, and in the field

Related to the H I distribution in the cosmic web, our analysis yields a significantly larger $M_{\rm H\,I}$ in filaments than in knots and in the field. In this sense, we observe a significant increase in $M_{\rm HI}$ from the field to filaments, i.e. from low-density to moderate/high-density environments, and then an even more significant decrease in $M_{\rm H\,I}$ from filaments to knots.

While it is not trivial to establish a consistent comparison between our definition of large-scale structure environment, inherited from Darvish et al. (2015, 2017), and the distinct definitions adopted in different works at $z \sim 0$, several authors find large galaxy groups and galaxy clusters – broadly speaking corresponding to our knots H I to be H I-deficient at $z \sim 0$ (e.g. Giovanelli & Haynes 1985; Solanes et al. 2001; Gavazzi et al. 2013; Dénes, Kilborn & Koribalski 2014; Odekon et al. 2016; Zabel et al. 2022). For instance, quantifying the H I deficiency as the logarithmic difference between the expected $M_{\rm H\,I}$ from scaling relations and the observed one (Haynes & Giovanelli 1984):

$$\rm H\,I_{def} = \log_{10}(\mathit{M}_{H\,I,exp}) - \log_{10}(\mathit{M}_{H\,I,obs}) \quad (11)$$

Odekon et al. (2016) study the slope of the $\rm H\,I_{def} - M_*$ relation for blue star-forming galaxies probed by ALFALFA (Haynes et al. 2011, 2018), and find that at fixed M_* the slope of such a relation rapidly increases with M_{200}.[4] Therefore, this means that the bigger the galaxy groups, the less H I is likely to be found in galaxies in such groups at fixed M_*. Furthermore, focusing on satellite galaxies only, Brown et al. (2017) perform a stacking-based analysis on the ALFALFA sample and report that more massive haloes tend to host satellite galaxies with lower H I-to-M_* ratio, at fixed M_*.

The decrease in $M_{\rm H\,I}$ from filaments to knots indicates the presence of some gas removal or consumption mechanism. For instance, ram pressure stripping H I well-known to be a major responsible for gas removal and ionization in clusters and in particular for H I stripping at $z \sim 0$ (e.g. Wang et al. 2020; Kleiner et al. 2023) H I may be removing H I from galaxies and determine the drop in H I amount from filaments to knots. In fact, ram-pressure stripping has been successfully observed to be in act in galaxy clusters up to the redshifts we are investigating ($0.3 < z < 0.5$, Moretti et al. 2022) and beyond ($z \sim 0.7$, Boselli et al. 2019), suggesting that it may have been even

[4] M_{200} is the mass contained within a radius inside of which the mean interior density is 200 times the mean density of the Universe.

more efficient in the past than in the present-day Universe (Moretti et al. 2022).

Despite the broad qualitative agreement with existing results at $z \sim 0$, we highlight three facts to be taken into account when performing such a comparison. First, we reiterate that we are not including passive galaxies in our study. Secondly, knots and clusters are not defined the same way.[5] As a results, knots may contain one or more clusters, but will typically encompass a larger volume, including clusters outskirts and regions beyond clusters. Thirdly, the knot galaxies subsample is the most likely to be subject to cosmic variance among the different samples investigated here. These aspects make it difficult to perform a quantitative fair comparison with $z \sim 0$ results.

Regarding the comparison between the H I content in filaments and in the field, it is worth mentioning that Kleiner et al. (2017) find no significant difference between galaxies in filaments and the control sample constituted by galaxies far away from the filament spines at $z \sim 0$. However, therein filaments are defined in a different way than in this paper, hence it is again not straightfoward to perform an adequate comparison (see e.g. Libeskind et al. 2018, for a review of methods to quantitatively define the cosmic web). On the other hand, low-density environments such as cosmic voids are known to be H I-poorer than filaments in cosmological hydrodynamic simulations (e.g. Martizzi et al. 2019), supporting our findings.

Unfortunately, the limited statistics, as well as the lack of a complete spectroscopic coverage, makes it hard to perform a more detailed comparison between our definition of knots and the over-densities used in other works, as well as further investigating whether the bulk of H I sits in a specific region of knots – e.g. in the outskirt or towards the centre – or whether it is spread across the whole radial profile. We leave this investigation for future works relying on the full area ($\sim$20 deg^2) covered by the MIGHTEE survey.

8 SUMMARY AND CONCLUSIONS

In this work, we have presented the first study of the dependence of H I content in galaxies on the large-scale structure, at a median redshift $z \sim 0.37$.

In particular, we select a M_*-complete star-forming galaxy sample ($\log_{10}(M_*/\mathrm{M}_\odot) > 9.6$) in the COSMOS field (Scoville et al. 2007) based on a colour-colour selection ($(M_{NUV} - M_r > 3(M_r - M_J) + 1$, $M_{NUV} - M_r > 3.1$), Laigle et al. 2016) and with measured spectroscopic redshift (Khostovan et al., in preparation).

We extract individually undetected H I galaxy spectra from H I data cubes delivered by the MIGHTEE survey, covering the full COSMOS field. We rely on a spectral line stacking approach to perform an average $M_{\rm H\,I}$ detection out of the extracted spectra for each of the studied subsample. After a careful flagging and a detailed treatment of RFI, our full sample consists of 2875 galaxies.

We investigate the dependence of the H I content in galaxies on the large scale structure. We perform stacking on subsamples defined by restricting the parent full sample according to different definitions of the large-scale environment. In particular, we focus on the local galaxy overdensity field, the position inside the host dark matter halo (central, satellite, or isolated), and the cosmic web environment that galaxies live in (field, filaments, knots). These information are computed for each of the galaxies constituting our sample through a tomographic analysis based on photo-zs, and made available by Darvish et al. (2015, 2017). We stress that the evaluation of the cosmic web environment as done in Darvish et al. (2015), and hence, in this work, is quantitative and based on the curvature tensor.

Our findings can be summarized as follows:

(i) We find a robust $\sim$11.8σ detection of $M_{\rm H\,I}$ for the full sample. While such a result is not fully compatible with the S22 scaling relation, the galaxy sample used in this work is not the same as in S22, and the true uncertainty associated to the scaling relation is unknown. Furthermore, the evolution of $M_{\rm H\,I}$ with redshift, as well as the impact of weighting scheme, makes the stacking result subject to variation depending on the reference frame one chooses. We will investigate impact of these aspects in a forthcoming paper (Bianchetti et al., in preparation) We point out that, because the sample is 90 per cent complete in M_* down to $\log_{10}(M_*/\mathrm{M}_\odot) \sim 9.6$, this $M_{\rm H\,I}$ measurement can be exploited to compute the H I density parameter $\Omega_{\rm HI}$. We leave the investigation of this aspect for future works.

(ii) Isolated galaxies yield $M_{\rm H\,I}$ and $f_{\rm H\,I}$ measurements very similar to the one from the full sample. Central galaxies yield a non-detection and the 3σ upper limit we place corresponds to a $M_{\rm H\,I}$ value which exceeds by a factor $\sim$1.5 the one from the full sample ($\sim$3.7σ deviation). Satellite galaxies are found to be H I-rich, exceeding both the $M_{\rm H\,I}$ and the $f_{\rm H\,I}$ of the full sample by a factor $\sim$1.4, at $\sim$2.2σ and $\sim$2.5σ, respectively. Moreover, satellite galaxies are HI-richer than the upper limit for central galaxies by a factor $\sim$2.1 in $M_{\rm H\,I}$ and by a factor $\sim$1.3 in $f_{\rm H\,I}$, at $\sim$4.9σ and $\sim$2.3σ respectively.

(iii) Field galaxies are H I-poorer than the full sample by a factor $\sim$1.4 in $M_{\rm H\,I}$ and by a factor $\sim$1.3 in $f_{\rm H\,I}$, at $\sim$1.5σ and $\sim$1.2σ, respectively. Filament galaxies are H I-richer than the full sample by a factor $\sim$1.4 in $M_{\rm H\,I}$ and by a factor $\sim$1.7 in $f_{\rm H\,I}$, at $\sim$3σ and $\sim$3.7σ, respectively. Knot galaxies yield a non-detection and the upper limit we derive is H I-poorer than the the full sample by a factor $\sim$2 in $M_{\rm H\,I}$ and by a factor $\sim$1.4 in $f_{\rm H\,I}$, at $\sim$5.3σ and $\sim$3.2σ, respectively. Moreover, such an upper limit is H I-poorer than the the filament galaxies sample by a factor $\sim$2.8 in $M_{\rm H\,I}$ and by a factor $\sim$2.4 in $f_{\rm H\,I}$, at $\sim$8.3σ and $\sim$6σ, respectively

(iv) The interpretation of our results seem to indicate that there is a general tendency of the studied galaxy population to feature a larger $M_{\rm H\,I}$ towards intermediate-density cosmic web environments and the outskirts of dark matter haloes (although at fixed stellar mass, and not fixed halo mass).

We point out that the studied volume may still feature some cosmic variance effects, especially regarding the results related to knots. We argue that future MIGHTEE-H I data beyond the Early Science dataset will allow us to significantly enlarge the probed volume at $0.23 < z < 0.49$, from the $\sim$1.5 deg^2 of the COSMOS field used in this paper, to $\sim$20 deg^2 of the final data release.

Furthermore, we also notice once again that we are neglecting the contribution to the H I budget from red passive galaxies. We leave this aspect to be investigated in future works (Rodighiero et al., in preparation).

We conclude that this work paves the way to future investigations exploring the connection between the H I properties in galaxies at these redshifts, and the large-scale structure environment.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the anonymous referee for the insightful comments they provided. The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which

[5]We notice that Darvish et al. (2015) uses the nomenclature'clusters' for knots. We prefer here to use'knots', to avoid confusion in the terminology.

is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. F.S. acknowledges the support of the doctoral grant funded by the University of Padova and by the Italian Ministry of Education, University and Research (MIUR) and the financial support of the *Fondazione Ing. Aldo Gini* fellowship. G.R. acknowledges the support from grant PRIN MIUR 2017–20173ML3WW_001. M.V. acknowledges financial support from the Inter-University Institute for Data Intensive Astronomy (IDIA), a partnership of the University of Cape Town, the University of Pretoria and the University of the Western Cape, and from the South African Department of Science and Innovation's National Research Foundation under the ISARP RADIOSKY2020 and RADIOMAP + Joint Research Schemes (DSI-NRF Grant Numbers 113121 and 150551) and the SRUG HIPPO Projects (DSI-NRF Grant Numbers 121291 and SRUG22031677). L.C. acknowledges support from the Australian Research Council Discovery Project and Future Fellowship funding schemes (DP210100337, FT180100066). Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. I.P., F.S. and G.R. acknowledge support from INAF under the Large Grant 2022 funding scheme (project 'MeerKAT and LOFAR Team up: a Unique Radio Window on Galaxy/AGN co-Evolution'). The authors also acknowledge the use of the ilifu cloud computing facility - www.ilifu.ac.za, a partnership between the University of Cape Town, the University of the Western Cape, the University of Stellenbosch, Sol Plaatje University, the Cape Peninsula University of Technology and the South African Radio Astronomy Observatory. The Ilifu facility is supported by contributions from the Inter-University Institute for Data Intensive Astronomy (IDIA - a partnership between the University of Cape Town, the University of Pretoria, the University of the Western Cape and the South African Radio Astronomy Observatory), the Computational Biology division at UCT and the Data Intensive Research Initiative of South Africa (DIRISA).

DATA AVAILABILITY

The MIGHTEE-HI spectral cubes and source catalogue will be released as part of the first data release of the MIGHTEE survey.

REFERENCES

Alpaslan M. et al., 2016, MNRAS, 457, 2287
Aragón-Calvo M. A., Jones B. J. T., van de Weygaert R., van der Hulst J. M., 2007, A&A, 474, 315
Barsanti S. et al., 2022, MNRAS, 516, 3569
Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2019, ApJ, 882, L7
Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2022, ApJ, 940, L10
Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2023, ApJ, 950, L18
Bischetti M., Maiolino R., Carniani S., Fiore F., Piconcelli E., Fluetsch A., 2019, A&A, 630, A59
Bolzonella M. et al., 2010, A&A, 524, A76
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
Boselli A. et al., 2019, A&A, 631, A114
Brown T. et al., 2017, MNRAS, 466, 1275
Catinella B. et al., 2018, MNRAS, 476, 875
Chauhan G., Lagos C. d. P., Stevens A. R. H., Bravo M., Rhee J., Power C., Obreschkow D., Meyer M., 2021, MNRAS, 506, 4893
Chen Y.-C. et al., 2017, MNRAS, 466, 1880
Chen Q. et al., 2021, MNRAS, 508, 2758
Chengalur J. N., Braun R., Wieringa M., 2001, A&A, 372, 768
Chowdhury A., Kanekar N., Chengalur J. N., Sethi S., Dwarakanath K. S., 2020, Nature, 586, 369
Chowdhury A., Kanekar N., Das B., Dwarakanath K. S., Sethi S., 2021, ApJ, 913, L24
Chowdhury A., Kanekar N., Chengalur J. N., 2022a, ApJ, 931, L34
Chowdhury A., Kanekar N., Chengalur J. N., 2022b, ApJ, 937, 103
Chowdhury A., Kanekar N., Chengalur J. N., 2022c, ApJ, 941, L6
Cortese L., Catinella B., Smith R., 2021, PASA, 38, e035
Crone Odekon M., Hallenbeck G., Haynes M. P., Koopmann R. A., Phi A., Wolfe P.-F., 2018, ApJ, 852, 142
Darvish B., Sobral D., Mobasher B., Scoville N. Z., Best P., Sales L. V., Smail I., 2014, ApJ, 796, 51
Darvish B., Mobasher B., Sobral D., Scoville N., Aragon-Calvo M., 2015, ApJ, 805, 121
Darvish B., Mobasher B., Martin D. C., Sobral D., Scoville N., Stroe A., Hemmati S., Kartaltepe J., 2017, ApJ, 837, 16
Davidzon I. et al., 2016, A&A, 586, A23
Decarli R. et al., 2018, ApJ, 854, 97
Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398
Dénes H., Kilborn V. A., Koribalski B. S., 2014, MNRAS, 444, 667
Dev A. et al., 2023, MNRAS, 523, 2693
Donnan C. T., Tojeiro R., Kraljic K., 2022, Nature Astron., 6, 599
Elson E. C., Blyth S. L., Baker A. J., 2016, MNRAS, 460, 4366
Elson E. C., Baker A. J., Blyth S. L., 2019, MNRAS, 486, 4894
Fabello S., Kauffmann G., Catinella B., Giovanelli R., Haynes M. P., Heckman T. M., Schiminovich D., 2011a, MNRAS, 416, 1739
Fabello S., Catinella B., Giovanelli R., Kauffmann G., Haynes M. P., Heckman T. M., Schiminovich D., 2011b, MNRAS, 411, 993
Finoguenov A. et al., 2007, ApJS, 172, 182
Frangi A. F., Niessen W. J., Vincken K. L., Viergever M. A., 1998, in Wells W. M., Colchester A., Delp S.eds, Medical Image Computing and Computer-Assisted Intervention—MICCAI'98. Springer Berlin Heidelberg, Berlin, Heidelberg, p. 130
Galárraga-Espinosa D., Aghanim N., Langer M., Gouin C., Malavasi N., 2020, A&A, 641, A173
Gavazzi G. et al., 2013, A&A, 553, A90
Geréb K., Morganti R., Oosterloo T. A., Guglielmino G., Prandoni I., 2013, A&A, 558, A54
Geréb K., Morganti R., Oosterloo T. A., Hoppmann L., Staveley-Smith L., 2015, A&A, 580, A43
Giovanelli R., Haynes M. P., 1985, ApJ, 292, 404
Guo H., Jones M. G., Haynes M. P., Fu J., 2020, ApJ, 894, 92
Guo H., Jones M. G., Wang J., Lin L., 2021, ApJ, 918, 53
Haynes M. P., Giovanelli R., 1984, AJ, 89, 758
Haynes M. P. et al., 2011, AJ, 142, 170
Haynes M. P. et al., 2018, ApJ, 861, 49
Healy J., Blyth S. L., Elson E., van Driel W., Butcher Z., Schneider S., Lehnert M. D., Minchin R., 2019, MNRAS, 487, 4901
Healy J. et al., 2021, A&A, 650, A76
Heavens A., Peacock J., 1988, MNRAS, 232, 339
Hu W. et al., 2019, MNRAS, 489, 1619
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Janowiecki S., Catinella B., Cortese L., Saintonge A., Brown T., Wang J., 2017, MNRAS, 466, 4795
Jarvis M. et al., 2016, in MeerKAT Science: On the Pathway to the SKA. p. 6, preprint (arXiv:1709.01901)
Jolly J.-B. et al., 2021, A&A, 652, A128
Jonas J., MeerKAT Team, 2016, in MeerKAT Science: On the Pathway to the SKA. p. 1
Kanekar N., Sethi S., Dwarakanath K. S., 2016, ApJ, 818, L28
Kleiner D., Pimbblet K. A., Jones D. H., Koribalski B. S., Serra P., 2017, MNRAS, 466, 4692
Kleiner D. et al., 2023, A&A, 675, A108
Kraljic K. et al., 2020, MNRAS, 491, 4294
Krolewski A., Ho S., Chen Y.-C., Chan P. F., Tenneti A., Bizyaev D., Kraljic K., 2019, ApJ, 876, 52
Lah P. et al., 2007, MNRAS, 376, 1357
Lah P. et al., 2009, MNRAS, 399, 1447
Laigle C. et al., 2016, ApJS, 224, 24

Libeskind N. I., Hoffman Y., Knebe A., Steinmetz M., Gottlöber S., Metuki O., Yepes G., 2012, MNRAS, 421, L137
Libeskind N. I. et al., 2018, MNRAS, 473, 1195
Lilly S. J. et al., 2009, ApJS, 184, 218
Maddox N. et al., 2021, A&A, 646, A35
Malavasi N. et al., 2017, MNRAS, 465, 3817
Malavasi N., Langer M., Aghanim N., Galárraga-Espinosa D., Gouin C., 2022, A&A, 658, A113
Martizzi D. et al., 2019, MNRAS, 486, 3766
Meyer S. A., Meyer M., Obreschkow D., Staveley-Smith L., 2016, MNRAS, 455, 3136
Moretti A. et al., 2022, ApJ, 925, 4
Obreschkow D., Meyer M., 2014, preprint (arXiv:1406.0966)
Obreschkow D., Croton D., De Lucia G., Khochfar S., Rawlings S., 2009a, ApJ, 698, 1467
Obreschkow D., Heywood I., Klöckner H. R., Rawlings S., 2009b, ApJ, 702, 1321
Obreschkow D., Klöckner H. R., Heywood I., Levrier F., Rawlings S., 2009c, ApJ, 703, 1890
Odekon M. C. et al., 2016, ApJ, 824, 110
Pan H., Jarvis M. J., Allison J. R., Heywood I., Santos M. G., Maddox N., Frank B. S., Kang X., 2020, MNRAS, 491, 1227
Pan H. et al., 2023, MNRAS, 525, 256
Pandey B., Sarkar S., 2020, MNRAS, 498, 6069
Planck Collaboration VI, 2020, A&A, 641, A6
Rajohnson S. H. A. et al., 2022, MNRAS, 512, 2697
Rhee J., Lah P., Briggs F. H., Chengalur J. N., Colless M., Willner S. P., Ashby M. L. N., Le Fèvre O., 2018, MNRAS, 473, 1879
Romano M. et al., 2022, A&A, 660, A14
Roychowdhury S. et al., 2022, ApJ, 927, 20
Scoville N. et al., 2007, ApJS, 172, 1
Shuntov M. et al., 2022, A&A, 664, A61
Sinigaglia F., Kitaura F.-S., Balaguera-Antolínez A., Nagamine K., Ata M., Shimizu I., Sánchez-Benavente M., 2021, ApJ, 921, 66
Sinigaglia F., Elson E., Rodighiero G., Vaccari M., 2022a, MNRAS, 514, 4205
Sinigaglia F., Kitaura F.-S., Balaguera-Antolínez A., Shimizu I., Nagamine K., Sánchez-Benavente M., Ata M., 2022b, ApJ, 927, 230
Sinigaglia F. et al., 2022c, ApJ, 935, L13
Smolčić V. et al., 2017, A&A, 602, A2
Solanes J. M., Manrique A., García-Gómez C., González-Casado G., Giovanelli R., Haynes M. P., 2001, ApJ, 548, 97
Stevens A. R. H. et al., 2019, MNRAS, 483, 5334
Tempel E., Libeskind N. I., 2013, ApJ, 775, L42
Tudorache M. N. et al., 2022, MNRAS, 513, 2168
Vulcani B. et al., 2019, MNRAS, 487, 2278
Wang J., Koribalski B. S., Serra P., van der Hulst T., Roychowdhury S., Kamphuis P., Chengalur J. N., 2016, MNRAS, 460, 2143
Wang J., Xu W., Lee B., Du M., Overzier R., Shao L., 2020, ApJ, 903, 103
Weaver J. R. et al., 2022, ApJS, 258, 11
Zabel N. et al., 2022, ApJ, 933, 10
Zel'dovich Y. B., 1970, A&A, 5, 84
Zwaan M. A., 2000, PhD thesis, Rijksuniversiteit Groningen
Zwaan M. A., van Dokkum P. G., Verheijen M. A. W., 2001, Science, 293, 1800

APPENDIX A: KOLMOGOROV-SMIRNOV TESTS

We report here the results of the two-samples Kolmogorov-Smirnov (KS) tests we performed in order to further characterize the distributions of the galaxy properties displayed in Fig. 3.

Figs A1 and A2 show the p-values for the 1D KS tests comparing M_* (Fig. A1), and z (Fig. A2) distributions for different subsamples, obtained by applying distinct cuts depending on the definition of environment. Assuming as confidence threshold $\alpha = 0.05$, we reject the null hypothesis that two samples come from the same distribution if $p > \alpha = 0.05$. In the case of the KS test comparing the M_* distribution for different subsample, $\delta > 0$, centrals and knots galaxies are the subsamples which tend to yield $p < 0.05$. In the case of the KS test comparing the z distributions, we can reject the null hypothesis in most of the cases. This means that the majority of

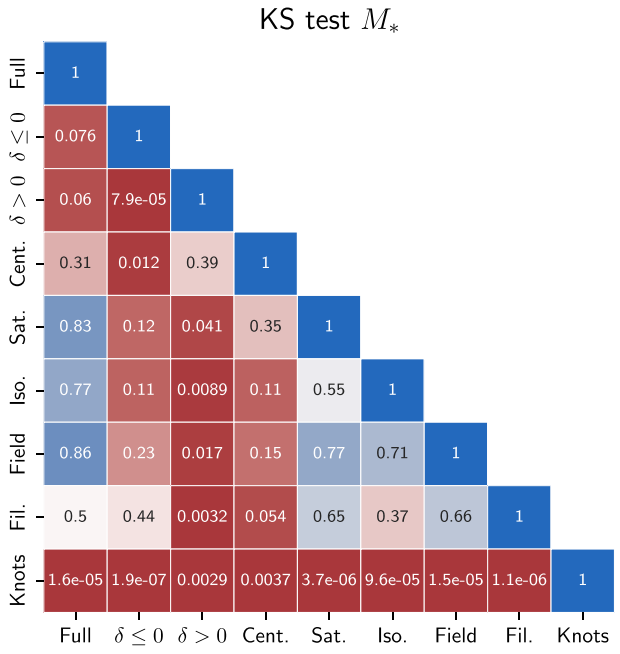

Figure A1. p-values resulting from Kolmogorov–Smirnov tests comparing M_* distributions of different subsamples. The plot is colour-coded from red to blue from the lowest to the highest significance values.

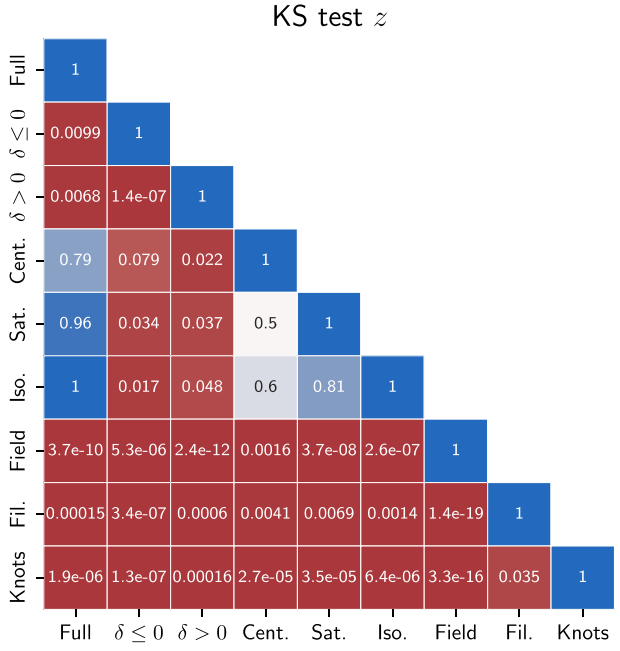

Figure A2. p-values resulting from Kolmogorov–Smirnov tests comparing z distributions of different subsamples. The plot is colour-coded from red to blue from the lowest to the highest significance values.

the subsamples do not come from the same underlying distribution of z.

APPENDIX B: IMPACT OF THE WEIGHTING SCHEME

In this Section we test the impact of the weighting scheme on the average $M_{\rm H\,I}$ that we measure from stacking.

As anticipated in Section 5, we define the weight corresponding to the ith spectrum as $w_i = 1/\sigma_i^{\gamma}$, where σ_i is the rms of the ith spectrum. Subsequently, following a procedure similar in spirit to the one of Hu et al. (2019), we study the evolution of the SNR and of $M_{\rm H\,I}$ as a function of γ. We report the results of the tests in Table B1 and in Fig. B1. We report the evolution of SNR (left panel) and of $M_{\rm H\,I}$ (right panel) as a function of γ. For the sake of clarity of visualization, we report here the results related to the full sample and the filament and satellite galaxies subsamples, and the same we do also for the numerical results in Table B1. However, we stress that we have tested all the subsamples and the same conclusions hold for all of them. It turns out that the choice $\gamma = 1$ maximises the SNR in all the studied subsamples. Therefore, we choose $\gamma = 1$. Since we are working with a volume-limited sample, this does not induce a selection bias due to the evolution of the M_* with redshift. Furthermore, we observe that $M_{\rm H\,I}$ slightly decreases with increasing γ.

To investigate the origin of the aforementioned effects, we study possible correlations between the weights (assuming $\gamma = 1$) associated with the galaxies belonging to the full sample in the left panel and redshift. Also, we split the full sample into two redshift subsamples—a low-redshift subsample at $0.23 < z < 0.35$ and a high-redshift subsample at $0.35 < z < 0.49$ – and compare the resulting average $M_{\rm H\,I}$ to one we compute from the full sample. We plot the weights as a function of redshift in the left panel of Fig. B2, showing our data points as blue circles and the median of the weight distributions in different redshift bins as an orange dashed curve.

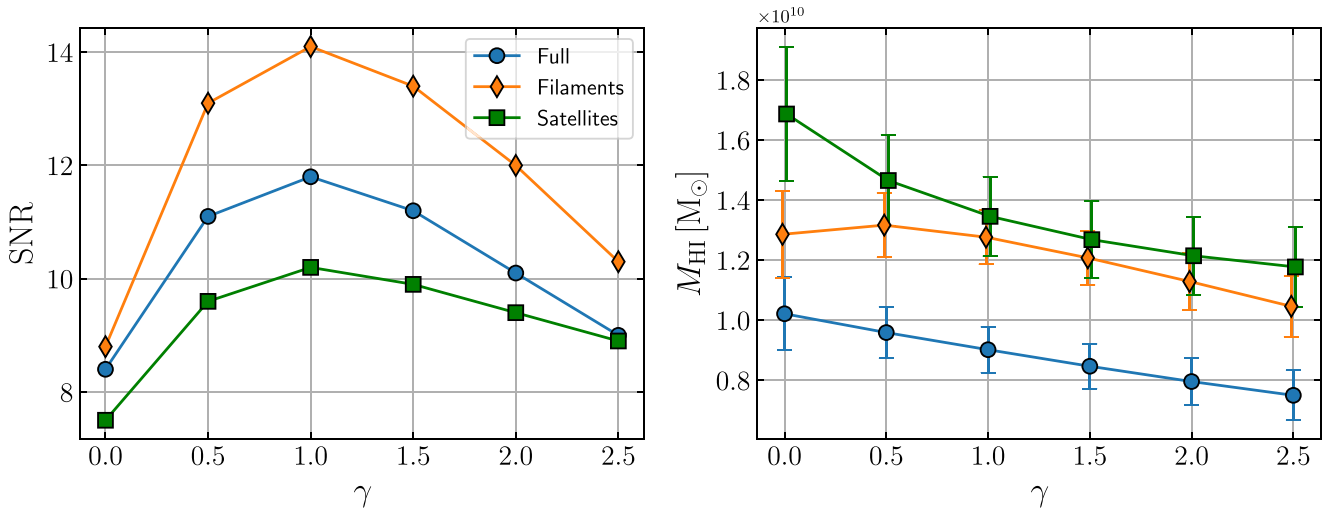

Figure B1. Left: SNR as a function of the exponent γ, for the full sample (blue circles), the filament galaxies subsample (orange diamonds), and the the satellite galaxies subsample (green squares). Right: $M_{\rm H\,I}$ as a function of γ, with the same colours scheme as the left panels. The same data are tabulated in Table B1.

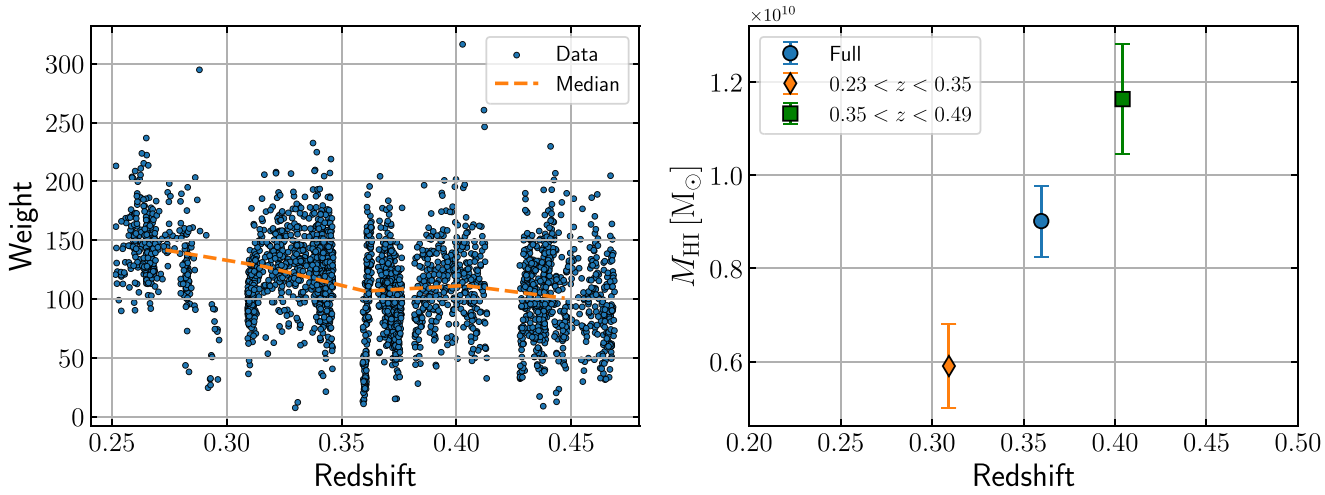

Figure B2. Left: Weights ($\gamma = 1$) as a function of redshift. The median (orange dashed) evidences that there is redshift evolution in the weights distribution. Right: $M_{\rm H\,I}$ as a function of redshift, for the full sample (blue circle), for the subsample at $0.23 < z < 0.35$ (orange diamond), and for the $0.35 < z < 0.49$ subsample (green square).

Table B1. Results of the tests on the impact of the weighting scheme from the stacking runs performed on the full sample (top), and the filament (middle) and satellite (bottom) galaxies subsamples. The first column lists the different tested values of γ, the second column the average $M_{\rm H\,I}$ and its associated uncertainties measured from the data, the third column the SNR, and the fourth column the average weighted redshift. These results are graphically shown in Fig. B1.

	Full sample		
γ	$M_{\rm H\,I}[\times 10^9\,{\rm M}_\odot]$	SNR	$\langle z\rangle$
0.0	10.22 ± 1.21	8.4	0.366
0.5	9.59 ± 0.86	11.1	0.363
1.0	9.02 ± 0.76	11.8	0.360
1.5	8.46 ± 0.75	11.2	0.357
2.0	7.96 ± 0.79	10.1	0.355
2.5	7.50 ± 0.83	9.0	0.353
	Filaments		
γ	$M_{\rm H\,I}[\times 10^9\,{\rm M}_\odot]$	SNR	$\langle z\rangle$
0.0	12.86 ± 1.45	8.8	0.360
0.5	13.17 ± 1.17	13.1	0.375
1.0	12.77 ± 0.90	14.1	0.354
1.5	12.08 ± 0.90	13.4	0.351
2.0	11.29 ± 0.94	12.0	0.348
2.5	10.46 ± 1.01	10.3	0.346
	Satellites		
γ	$M_{\rm H\,I}[\times 10^9\,{\rm M}_\odot]$	SNR	$\langle z\rangle$
0.0	16.87 ± 2.25	7.5	0.366
0.5	14.65 ± 1.53	9.6	0.362
1.0	13.46 ± 1.42	10.2	0.360
1.5	12.68 ± 1.28	9.9	0.357
2.0	12.15 ± 1.29	9.4	0.354
2.5	11.77 ± 1.33	8.9	0.352

Then, we display $M_{\rm H\,I}$ as a function of redshift in the right panel of Fig. B2, where the blue circle stands for the full sample, the orange diamond for the low-redshift ($0.23 < z < 0.35$) subsample, and the green square for the high-redshift ($0.35 < z < 0.49$) subsample. We report an anticorrelation between weights and redshift and a positive correlation between $M_{\rm H\,I}$ and redshift.

We discuss the implications of these findings in Section 5.

APPENDIX C: IMPACT OF RFI MASKING

In this section we address the effect of masking the RFI-affected channels of spectra which have been accepted in the selection procedure. This the case when the spectroscopic redshift of a galaxy—used to define the central frequency channel of the spectrum—does not fall within a frequency region affected by RFI, but some other channel(s) of the spectrum do. In our methodology, the flux in such channels is set to zero. In what follows, we dub this procedure'off-line' RFI masking.

We perform the following experiment. For each subsample, we run the stacking pipeline both with and without off-line RFI masking. We report the results in Table C1. Therein, the first column lists the different investigated subsamples. The second and third columns report the average $M_{\rm HI}$ with its associated uncertainty and the SNR, respectively, before RFI masking. The third and fourth columns report the average $M_{\rm H\,I}$ with its associated uncertainty and the SNR, respectively, after RFI masking. The fifth column reports the (signed) percentage deviation $\Delta M_{\rm H\,I} = 100 \times (M_{\rm H\,I,w/\,mask}/M_{\rm H\,I,wo/\,mask} - 1)$. The sixth, seventh, and eight columns report respectively the mean, average, and maximum percentage of channels which are masked in our procedure.

We notice that (i) the off-line RFI masking always increases the SNR with respect to the unmasked case, and (ii) such a procedure does not seem to introduce a systematic effect in the final results, and (iii) the $M_{\rm H\,I}$ after masking deviates < 10 per cent in al the studied cases, and $\lesssim 5$ per cent in the majority of them.

This demonstrates that the off-line RFI masking is advantageous in terms of gain in SNR, and does not introduce systematics in the stacking procedure. We treat the random deviation introduced by the implementation of such a masking by adding a conservative 10 per cent uncertainty in quadrature to the error coming from the measurement.

Table C1. Results of the tests on the impact of the off-line RFI masking. The first column lists the different investigated subsamples. The second and third columns report the average $M_{\rm H\,I}$ with its associated uncertainty and the SNR, respectively, before RFI masking. The third and fourth columns report the average $M_{\rm H\,I}$ with its associated uncertainty and the SNR, respectively, after RFI masking. The fifth column reports the (signed) percentage deviation $\Delta M_{\rm H\,I} = 100 \times (M_{\rm H\,I,w/\,mask}/M_{\rm H\,I,wo/\,mask} - 1)$. The sixth, seventh, and eight columns report respectively the mean, average, and maximum percentage of channels which are masked in our procedure.

Sample	$M_{\rm H\,I}[\times 10^9\,{\rm M}_\odot]$ (wo/ mask)	SNR (wo/ mask)	$M_{\rm H\,I}[\times 10^9\,{\rm M}_\odot]$ (w/ mask)	SNR (w/mask)	$\Delta M_{\rm H\,I}$ (per cent)	Mean mask. chan. (per cent)	Med. mask. chan. (per cent)	Max. mask. chan. (per cent)
Full	8.63 ± 0.82	10.5	9.02 ± 0.76	11.8	−4.3	22.2	20.2	50.5
$\delta \leq 0$	8.11 ± 1.20	6.7	7.86 ± 1.04	7.5	+3.2	21.1	19.2	50.5
$\delta > 0$	9.15 ± 1.17	8.1	10.08 ± 1.04	9.6	−9.22	23.2	21.2	49.5
Satellites	12.90 ± 1.45	8.9	13.46 ± 1.32	10.2	−4.1	22.5	21.2	49.5
Isolated	7.96 ± 1.30	6.1	8.37 ± 0.90	9.3	−5.0	22.0	20.2	49.5
Field	6.51 ± 1.33	4.9	6.17 ± 1.21	5.1	+5.5	22.6	20.2	50.5
Filaments	12.42 ± 0.97	12.7	12.77 ± 0.90	14.1	−2.3	22.2	21.2	49.5

[1]*Department of Physics and Astronomy, Università degli Studi di Padova, Vicolo dell'Osservatorio 3, I-35122, Padova, Italy*
[2]*INAF - Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122, Padova, Italy*
[3]*Instituto de Astrofísica de Canarias, Calle Via Láctea s/n, E-38205, La Laguna, Tenerife, Spain*
[4]*Departamento de Astrofísica, Universidad de La Laguna, E-38206, La Laguna, Tenerife, Spain*
[5]*Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Rd, 7535 Bellville, Cape Town, South Africa*
[6]*Inter-university Institute for Data Intensive Astronomy, Department of Physics and Astronomy, University of the Western Cape, 7535 Bellville, Cape Town, South Africa*
[7]*Inter-university Institute for Data Intensive Astronomy, Department of Astronomy, University of Cape Town, 7701 Rondebosch, Cape Town, South Africa*
[8]*INAF - Istituto di Radioastronomia, via Gobetti 101, I-40129 Bologna, Italy*
[9]*School of Physics, H.H. Wills Physics Laboratory, Tyndall Avenue, University of Bristol, Bristol, BS8 1TL, UK*
[10]*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*
[11]*Oxford Astrophysics, Denys Wilkinson Building, University of Oxford, Keble Rd, Oxford, OX1 3RH, UK*
[12]*Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa*
[13]*South African Radio Astronomy Observatory, 2 Fir Street, Black River Park, Observatory, 7925, South Africa*
[14]*International Centre for Radio Astronomy Research (ICRAR), University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*
[15]*ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*
[16]*University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 München, Germany*
[17]*Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, 9000 Gent, Belgium*
[18]*School of Science, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia*
[19]*CSIRO, Space and Astronomy, PO Box 1130, Bentley, WA, 6102, Australia*
[20]*Aix Marseille Univ, CNRS, CNES, LAM, Marseille 13013, France*
[21]*Laboratory for Multiwavelength Astrophysics, School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA*
[22]*Max Planck Institute for Extraterrestrial Physics, Giessembachstrasse 1, D-857498, Garching, Germany*
[23]*National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, NM 87801, USA*

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