

Measuring the baryonic Tully–Fisher relation below the detection threshold

Hengxing Pan^{1,2,3*}, Matt J. Jarvis^{1,3}, Anastasia A. Ponomareva^{1,3}, Mario G. Santos^{1,4},
James R. Allison^{1,3}, Natasha Maddox^{1,5} and Bradley S. Frank^{4,6,7}

¹Department of Physics and Astronomy, University of the Western Cape, Cape Town 7535, South Africa

²Purple Mountain Observatory, 2 West Beijing Road, Nanjing 210008, China

³Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

⁴South African Radio Astronomy Observatory (SARAO), 2 Fir Street, Observatory, 7925, South Africa

⁵Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany

⁶The Inter-University Institute for Data Intensive Astronomy (IDIA), and University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa

⁷Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa

Accepted 2021 September 8. Received 2021 September 8; in original form 2021 June 14

ABSTRACT

We present a novel 2D flux density model for observed HI emission lines combined with a Bayesian stacking technique to measure the baryonic Tully–Fisher relation below the nominal detection threshold. We simulate a galaxy catalogue, which includes HI lines described with either Gaussian or busy function profiles, and HI data cubes with a range of noise and survey areas similar to the MeerKAT International Giga-Hertz Tiered Extragalactic Exploration (MIGHTEE) survey. With prior knowledge of redshifts, stellar masses, and inclinations of spiral galaxies, we find that our model can reconstruct the input baryonic Tully–Fisher parameters (slope and zero-point) most accurately in a relatively broad redshift range from the local Universe to $z = 0.3$ for all the considered levels of noise and survey areas and up to $z = 0.55$ for a nominal noise of $90 \mu\text{Jy/channel}$ over 5 deg^2 . Our model can also determine the $M_{\text{HI}} - M_*$ relation for spiral galaxies beyond the local Universe and account for the detailed shape of the HI emission line, which is crucial for understanding the dynamics of spiral galaxies. Thus, we have developed a Bayesian stacking technique for measuring the baryonic Tully–Fisher relation for galaxies at low stellar and/or HI masses and/or those at high redshift, where the direct detection of HI requires prohibitive exposure times.

Key words: methods: statistical – galaxies: fundamental parameters – radio lines: galaxies.

1 INTRODUCTION

The Tully–Fisher relation (TFR, Tully & Fisher 1977) is an empirical relationship that links the luminosity of a disc galaxy to its rotational velocity (Sorce et al. 2014; Ponomareva et al. 2018; Lelli et al. 2019). The TFR remains one of the most studied dynamical scaling relations of disc galaxies, and has been shown to hold over a wide wavelength range (Ponomareva et al. 2017), in different environments (Willick 1999; Abril-Melgarejo et al. 2021) and for galaxies of different morphological types (Chung et al. 2002; Courteau et al. 2003; Bedregal, Aragón-Salamanca & Merrifield 2006; Karachentsev, Kaisina & Kashibadze Nasonova 2017). The TFR was first established as a powerful tool to make redshift-independent distance estimates for spiral galaxies, with the goal of deriving the local peculiar velocity field (Tully et al. 2014).

A more elementary form of the TFR is the so-called baryonic Tully–Fisher relation (bTFR), which relates the baryonic component of a galaxy (characterized by its total baryonic mass) to its total dynamical mass (including the dark matter; DM), characterized by the rotational velocity. Although the exact nature of the bTFR remains unclear, it is one of the most fundamental relations in the framework

of galaxy formation and evolution since it places strong constraints on the properties of galaxies produced by cosmological simulations (e.g. Glowacki, Elson & Davé 2020b; Dubois et al. 2020).

One of the open questions in the comparison of these simulations and observations is the presence or absence of curvature at the low-mass end of the bTFR. While curvature is a generic prediction of the models (Trujillo-Gomez et al. 2011; Desmond 2012), because it arises from the need to achieve low baryon fractions for both high- and low-mass haloes (Papastergis et al. 2012), it has not been observed even in the bTFR studies of heavily gas-dominated galaxies (Begum et al. 2008; Papastergis, Adams & van der Hulst 2016; Iorio et al. 2017). However, since the low-mass ($\lesssim 10^9 M_\odot$) end of the bTFR is always poorly constrained due to the faintness of the atomic hydrogen (HI) signal, which usually lies below the detection threshold (HI being the main mass contributor), there is still no definitive answer regarding the slope of the bTFR at this mass range.

At low redshift ($z < 0.1$), HI is used as a kinematic tracer for the bTFR studies as it extends far beyond the optical radius, better tracing the whole of the DM halo. At higher redshifts, carbon monoxide (CO) and optical lines (e.g. H α , [O III]) have been used as kinematic tracers for studying the evolution of the bTFR (Übler et al. 2017; Topal et al. 2018; Tiley et al. 2019), since the HI emission line has for a long time remained undetectable beyond the local Universe due to its intrinsic faintness (Fernández et al. 2016; Chowdhury et al. 2020).

* E-mail: hengxing.pan@physics.ox.ac.uk

The study of the bTFR evolution is extremely important since it can provide valuable insights on the mechanisms of galaxies' growth and help place further constraints on the cosmological simulations of galaxy formation and evolution. Significant evolution of the bTFR would point to the imbalance in the mass assembly of the baryons and the DM haloes due to the effect of the halo on the formation of the central galaxy. For example, a recent study of the state-of-the-art cosmological simulations SIMBA (Davé et al. 2019) found a clear change in the visual distribution, as well as the linear best-fitting parameters to the bTFR over the redshift range between $z = 0$ and $z = 1$, mostly due to the differences in the merger histories of the DM haloes (Glowacki, Elson & Davé 2020a). In that work, they make predictions for the future H I surveys, which need to be tested.

To date, there has been no definitive observational conclusion regarding the evolution of the bTFR, due to the different kinematic tracers and samples used. Ideally, one would need to compare the statistical properties of the H I-based bTFR (slope, scatter, and zero-point) at different redshifts, as well as use similar methods to measure these properties (Bradford, Geha & van den Bosch 2016). Unfortunately, this remains intractable even with the modern state-of-the-art observational facilities. For example, the H I data from the MIGHTEE survey (Jarvis et al. 2016; Maddox et al. 2021), which is being undertaken with the Meer Karoo Array Telescope (MeerKAT, Jonas & MeerKAT Team 2016), will not detect individual galaxies with a typical H I gas mass ($\sim M_{\text{H I}}^*$) beyond $z \sim 0.3$ (Maddox et al. 2021). Thus, to understand the evolution of the bTFR, and hence the evolution of gas in galaxies and its gravitational interplay with the DM haloes, one way to progress forward is to use stacking techniques in order to extend the accessible redshift and mass range by measuring a statistical signal (Delhaize et al. 2013; Rhee et al. 2018).

Meyer et al. (2016) showed that the TFR can be recovered via the spectral stacking technique by comparing the statistical properties of the stacked K -band TFR with the direct measurements from the HI Parkes All-Sky (HIPASS) Survey (Zwaan et al. 2005). Moreover, they showed that no knowledge of the individual galaxy inclinations, or even the inclination distribution of the data, is needed to accurately recover the properties. However, this technique is based on the classic stacking approach, where the spectral line data at the known locations and redshifts of many galaxies are co-added to improve the signal-to-noise ratio, thus losing information on individual galaxies contributing to the stacked spectra by averaging out their properties.

In this paper, we develop a novel approach and extend the Bayesian stacking technique, which was introduced for measuring the H I Mass Function below the detection threshold (Pan et al. 2020, hereafter P20), to the measurement of the bTFR. This work provides the framework to measure the H I-based bTFR beyond the local Universe, including low H I mass galaxies.

This paper is organized as follows. In Section 2, we describe our simulated galaxy catalogues and H I data cubes. In Section 3, we detail our method for modelling the 2D flux density and present an extended Bayesian stacking technique based on P20. We show the results of implementing this technique to our simulated data sets in Section 4 and present our conclusions in Section 5. Throughout the paper, we adopt a concordance cosmology with a Hubble constant $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$, total matter density $\Omega_m = 0.308$ and dark energy density $\Omega_\Lambda = 0.692$ (Planck Collaboration XIII 2016).

2 SIMULATED DATA

In order to test our method to measure the bTFR below the nominal detection threshold of H I galaxies in a survey, we first require a simulated H I data cube populated with galaxies of known baryonic

mass. To simulate such a cube, we need to construct the H I 21-cm line flux distribution per frequency channel assuming reasonable emission-line profiles, a source-count model, and the expected noise properties for a typical survey.

2.1 Generating the galaxy sample

First, we simulate an optical catalogue of galaxies with known stellar masses using an assumed galaxy stellar mass function (SMF). The SMF represents the intrinsic number density of galaxies in the Universe as a function of their stellar mass at a given redshift. We use a Schechter function model (Schechter 1976) along with a pure density evolution term to characterize the evolution of the SMF with redshift:

$$\phi(M_*, z) = \ln(10)\phi_* \left(\frac{M_*}{M_*^*}\right)^{\alpha_*+1} e^{-\frac{M_*}{M_*^*}(1+z)}, \quad (1)$$

where $\phi_* = 1.5 \times 10^{-3}$, $M_*^* = 1.3 \times 10^{11} M_\odot$, and $\alpha_* = -1.1$ correspond to the normalization, characteristic mass, and faint-end slope, respectively.

We then use the assumed SMF to generate a sample of simulated galaxies over a volume similar to the one expected for the MIGHTEE-H I survey (i.e. $\approx 20 \text{ deg}^2$ over the redshift range of $0 < z < 0.55$). With the big survey volume and exceptional multiwavelength ancillary data, MIGHTEE-H I is an ideal test bed for our novel approach for constraining the bTFR. We note that the exact form of the SMF does not impact on our results, although in reality an accurate account of the stellar mass incompleteness for the parent galaxy sample may be needed to account for any possible biases that arise from sample selection.

2.2 H I and stellar mass relation of spiral galaxies

The relationship between the H I and the stellar mass in galaxies is not straightforward. For example, it was shown by Maddox et al. (2015) that it is not linear and tends to flatten out at higher stellar masses. Parkash et al. (2018) considered various morphological samples of galaxies and found that for a sample of spiral galaxies this relationship can be described as

$$\log_{10}(M_{\text{H I}}) = 0.35[\log_{10}(M_*) - 10] + 9.45, \quad (2)$$

with 68 per cent of the H I masses within 0.4-dex uncertainty ($\sigma_{\text{H I}}$), where $M_{\text{H I}}$ and M_* are the galaxy H I and stellar mass, respectively. Thus, we generate the population of spiral galaxies with H I masses defined as

$$\log_{10}(M_{\text{H I}}) = 0.35[\log_{10}(M_*) - 10] + 9.45 + \mu, \quad (3)$$

where μ is a random variable that follows a normal distribution with mean 0 and standard deviation $\sigma_{\text{H I}} = 0.4 \text{ dex}$.

Fig. 1 shows the simulated H I mass as a function of stellar mass, with lower and upper stellar mass limits of $M_* = 10^7$ and $10^{11} M_\odot$ for our simulated survey. The stellar mass range we use is applicable to most H I-bearing galaxies. Of course, a colour or morphological selection would help to better determine the H I masses; however, this selection mostly affects the highest mass objects, which do not change our results significantly due to their low numbers as we will demonstrate in Section 4. Fig. 2 shows the H I mass as a function of redshift, from which it is clear that the vast majority of H I galaxies lie below the nominal noise threshold of the MIGHTEE survey (indicated by the coloured lines). Our goal here is to take these faint H I galaxies into account.

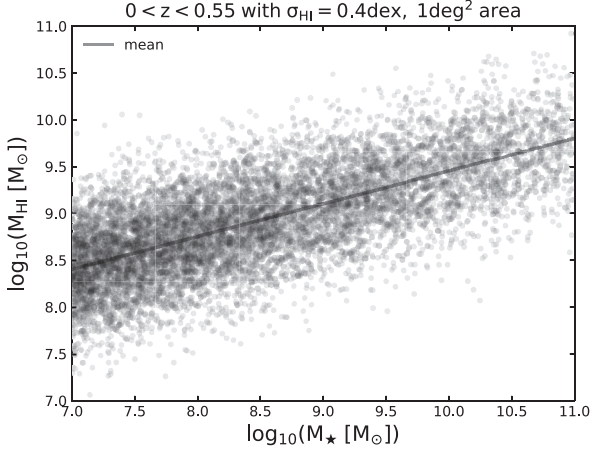


Figure 1. Simulated HI and stellar mass relation with $\sigma_{\text{HI}} = 0.4$ dex on the HI mass.

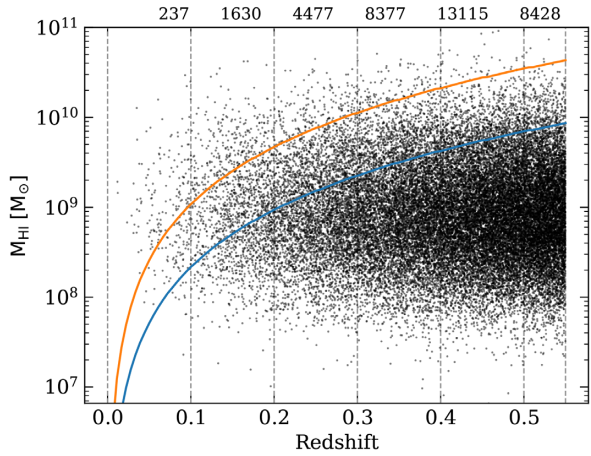


Figure 2. Simulated HI mass versus redshift over a 1 deg^2 area with a scatter of $\sigma_{\text{HI}} = 0.4$ dex on the HI mass, and lower and upper stellar mass limits of $M_\star = 10^7$ and $10^{11} M_\odot$. The colour-coded lines indicate the $\sigma_n = \sqrt{N_{\text{ch}}}\sigma_{\text{ch}}v$ and $5\sigma_n$ detection threshold. The numbers listed at the top of the figure indicate the number of galaxies in each $\Delta z = 0.1$ redshift bin.

Table 1. The parameters of the busy function for simulating the HI lines. We note that only parameters b_1 , b_2 , r_c , and x_p are randomly sampled. The centroid of the error function, x_e , is fixed at the redshift of the source. The parameter r_c is the original parameter c weighted by the half-width w .

Parameter	Meaning	Probability distribution
b_1, b_2	Steepness of line flanks	Uniform $\in [0.3, 0.7]$
r_c	Ratio of peak to trough	Uniform $\in [1.5, 2]$
$ x_p - x_e $	Difference of centroids	Uniform $\in [0, 0.3w]$

We note that the assumed HI and stellar mass relation is rather simplistic than realistic as more low HI mass galaxies are expected from HI surveys and the scatter on this relation may be mass dependent. A more sophisticated model, such as a 2D distribution of galaxies as a function of HI and stellar mass (or even extended to a 3D distribution if the velocity width is also taken as a variable), could be implemented to better quantify the $M_{\text{HI}} - M_\star$ relation. However, this would just add a few more free parameters to our model and will not influence the main results as the majority of spiral galaxies follows equation (2) well. At least, over a range of

masses $9 < \log_{10}(M_{\text{HI}}/M_\odot) < 11$, the $M_{\text{HI}} - M_\star$ relation is fairly linear, and at masses lower than that, the constraining power on the bTFR is less strong as we will see in the results. Although a more complex treatment could be applied in future to real data, this is beyond the scope of this paper where we wish to highlight the key features of the method.

2.3 The baryonic Tully–Fisher relation

To estimate the total baryonic mass of our simulated galaxies, we calculate the total mass of a galaxy disc as $M_{\text{disc}} = M_\star + M_{\text{gas}}$, assuming $M_{\text{gas}} = 1.4 M_{\text{HI}}$ to account for the presence of helium and metals (Arnett 1996). We can then predict the rotational velocity of a galaxy (V_{rot}) using an assumed bTFR:

$$A \left(\frac{W}{200 \text{ km s}^{-1}} \right)^\alpha = M_{\text{disc}}, \quad (4)$$

where $W = 2V_{\text{rot}}$ is the corrected width of the HI line profile, and we take the normalization (i.e. zero-point) $A = 2 \times 10^9$ and the slope $\alpha = 4$ as our canonical model (McGaugh et al. 2000). We simulate the inclination $\sin(i)$ as uniformly distributed from 0 to 1, so the observed velocity width (W_{50}) is defined as:

$$W_{50} = W \times \sin(i). \quad (5)$$

2.4 Profiles of the HI emission lines

To determine whether we can measure the bTFR below the noise threshold, we need to simulate realistic HI line profiles. Therefore, we adopt a Gaussian profile to describe the shape of the HI flux density distribution of dwarf and face-on galaxies:

$$G(v) = \frac{a}{\sigma\sqrt{2\pi}} \times e^{-\frac{(v-v_0)^2}{2\sigma^2}}, \quad (6)$$

where $\sigma = W_{50}/2.3548$, v_0 is the central frequency at a given redshift, and a is the integrated flux determined by the HI mass and redshift. Whereas for inclined spirals, we use the busy function (BF, Westmeier et al. 2014) for incorporating the double-horn profiles:

$$B_1(v) = \frac{a}{4} \times (\text{erf}[b_1\{w + v - x_e\}] + 1) \times (\text{erf}[b_2\{w - v + x_e\}] + 1) \times (c|v - x_p|^n + 1), \quad (7)$$

where w is the half-width of the observed profile; $w \sim W_{50}/2$. We also find that setting the parameter $c = (r_c - 1)w^{-n}$ allows for more realistic profiles to be generated for broad lines. This relation defines a new fixed parameter r_c , replacing the original parameter c and describing the ratio of peak flux to the trough flux, as we find that the peak flux of a double-horn line is approximately proportional to the flux at $|v - x_p| = w$ for broad lines. We note that r_c is only an indicator of that ratio rather than a precise definition. Nevertheless, it provides an intuitive way of controlling the shape of generated lines by the BF. The probability distribution of these input parameters for generating the HI lines with BF profiles is listed in Table 1.

We limit the BF profiles to the more edge-on galaxies with the velocity widths larger than 120 km s^{-1} . Given the velocity width per channel in our cubes $\sim 5.5 \text{ km s}^{-1}$ at $z = 0$, and at least a few channels are needed to resolve the double horns on the line profile, we chose the following specific criteria for simulating our HI lines:

$$S(v) = \begin{cases} B_1(v) & \text{for } \sin(i) \geq 0.5 \text{ and } W_{50} \geq 132 \text{ km s}^{-1}, \\ G(v) & \text{otherwise.} \end{cases} \quad (8)$$

We also need to determine the velocity width W given by the bTFR (equation 4) and the amplitude a fixed by the integrated flux $S =$

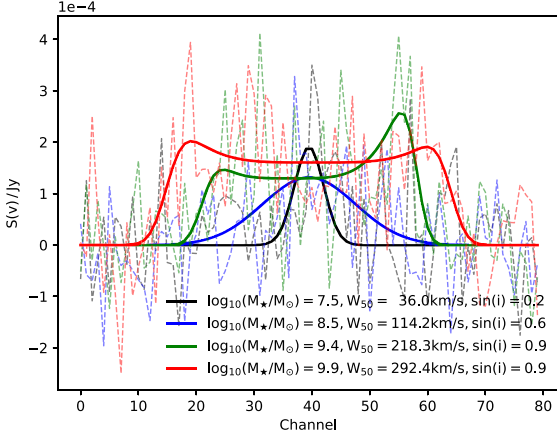


Figure 3. Examples of flux density as a function of channel at $0.1 < z < 0.2$. The solid lines are pure signals while the dashed lines are signals added to the noise. The central channels of the spectra are fixed at channel 40.

$\int S(v)dv$. The integrated flux can be converted from the H I mass under the usual assumptions (i.e. optically thin gas, Meyer et al. 2017) via

$$M_{\text{HI}} = 2.356 \times 10^5 D_L^2 (1+z)^{-1} S, \quad (9)$$

where the H I mass (M_{HI}) is in units of solar masses, the luminosity distance to the galaxy (D_L) is in Mpc, and the integrated flux (S) is in Jy km s^{-1} . The $(1+z)$ factor is needed when S is expressed in units of Jy km s^{-1} rather than Jy Hz .

Finally, we populate the simulated cubes with the emission lines, including Gaussian noise with standard deviation similar to that expected for the MIGHTEE survey. Specifically, we adopt $\sigma_{\text{ch}} = 90 \mu\text{Jy/channel}$, which is the expected noise per 26 kHz channel for MIGHTEE. The simulated cubes have a spatial resolution of 5 arcsec and a total of 19 504 spectral channels from 913.4 MHz to 1420.5 MHz. We note that the real cubes from the new generation of H I surveys are likely to have lower spatial resolution than this, and source confusion may increase the noise slightly; however, here we demonstrate the overall methodology.

The measured flux from the H I emission line $S_m(v)$ is a combination of the intrinsic flux from the source $S(v)$ plus the contribution from the noise $S_n(v)$:

$$S_m(v) = S(v) + S_n(v), \quad (10)$$

where $S_n(v) = \text{Normal}(0, \sigma_{\text{ch}})$ for this simulation. Fig. 3 shows a few examples of flux density that have been redshifted to $0.1 < z < 0.2$ as a function of the frequency channel with mixed Gaussian and BF profiles. As expected from the bTFR, and from the assumed $M_{\text{HI}} - M_*$ relation, the M_* mass correlates with the line width when corrected for the galaxy inclination.

3 THE STATISTICAL MODEL

Being able to simulate the emission lines in essence indicates that we can build a model to determine the bTFR for H I galaxy surveys. Even though we will not be able to measure the H I mass of individual galaxies below the nominal detection threshold, for surveys like MIGHTEE, which have excellent ancillary data, we will have information on the stellar masses, inclinations, and redshifts. With the shape of the $M_* - M_{\text{HI}}$ relation, knowing the M_* of a galaxy gives us a predicted probability density function (PDF) for the M_{HI} (i.e. to generate H I samples shown in Fig. 1), which leads to the PDFs

for the integrated flux (S), gas mass ($M_{\text{gas}} = 1.4 M_{\text{HI}}$), and therefore the disc mass ($M_{\text{disc}} = M_* + M_{\text{gas}}$).

Based on the assumed bTFR, we can further predict the PDF for the velocity width of the H I line profile of a single galaxy and for the integrated flux that determines the amplitude. Along with the assumed H I line profiles, we can then predict the PDF for the H I flux density of the galaxy in each frequency channel. Conversely, if we do not know the parameters A and α for the bTFR, we can then constrain them with the observed H I emission lines in return, which is the aim of this work.

3.1 Modelling the 2D flux density

Given our model, the probability of having the measured fluxes for a single source, $P(S_m(v))$, can be expressed as

$$P(S_m(v)) = \int dM_{\text{HI}} P(M_{\text{HI}}) \prod_v P_n(S_m(v) - S(v, M_{\text{HI}})), \quad (11)$$

where $P(M_{\text{HI}})$ is the PDF for the M_{HI} at a fixed stellar mass, and P_n follows the noise distribution of $\text{Normal}(0, \sigma_{\text{ch}})$ in each channel independently from other channels. $S(v, M_{\text{HI}})$ is the modelled 2D flux density as a function of channel number and H I mass. However, the relationship between the H I mass and flux density $S(v)$ is non-linear; thus, the $S(v, M_{\text{HI}})$ must be solved numerically.

In fact, we model the $S(v, M_{\text{HI}})$ for each channel by generating a distribution of H I samples for a given stellar mass and then propagating them into the final flux density as we generate the signals. We use the quantities with a tilde on the top of their symbols as the samples that meet a given distribution, e.g. $\log_{10}(\tilde{M}_{\text{HI}})$ indicates a group of H I samples with their mass following a distribution of $\text{Normal}(\log_{10}(M_{\text{HI}}), \sigma_{M_{\text{HI}}})$.

If we only take into account the uncertainties of emission lines caused by the scatter on the $M_{\text{HI}} - M_*$ mass relation, then the model for predicting the velocity width W_{50} and the integrated flux S is composed of the following equations:

$$\log_{10}(\tilde{M}_{\text{HI}}) = 0.35[\log_{10}(M_*) - 10] + 9.45, \quad (12)$$

$$\tilde{M}_{\text{disc}} = M_* + 1.4 \tilde{M}_{\text{HI}}, \quad (13)$$

$$A \left(\frac{\tilde{W}}{200 \text{ km s}^{-1}} \right)^\alpha = \tilde{M}_{\text{disc}}, \quad (14)$$

$$\tilde{W}_{50} = \tilde{W} \times \sin(i), \quad (15)$$

$$\tilde{S} = \tilde{M}_{\text{HI}} / (2.356 \times 10^5 D_L^2 (1+z)^{-1}). \quad (16)$$

For modelling the Gaussian lines, we take $S(v) = G(v)$. For the double-horn lines, we adopt a simplified version of the BF with six free parameters as

$$B_2(v) = \frac{a}{4} \times (\text{erf}[b(w + v - x)] + 1) \times (\text{erf}[b(w - v + x)] + 1) \times (c|v - x|^n + 1). \quad (17)$$

Here, we replace b_1, b_2 with b , and x_p, x_e with x compared with equation (7). If we set $n = 4$ and $c = (r_c - 1)w^{-n}$, along with the predicted integrated flux and velocity width, the final free parameters of $B_2(v)$ are reduced to be the steepness of the line flanks b and ratio of peak to trough r_c . We then take these two parameters as nuisance variables in the model.

From equation (17), it can be found that $B_2(v)$ is symmetric once we have $b_1 = b_2$ and $x_p = x_e$. The aim is to make a simplified form (B_2) of the BF that can model the width and amplitude of the BF B_1

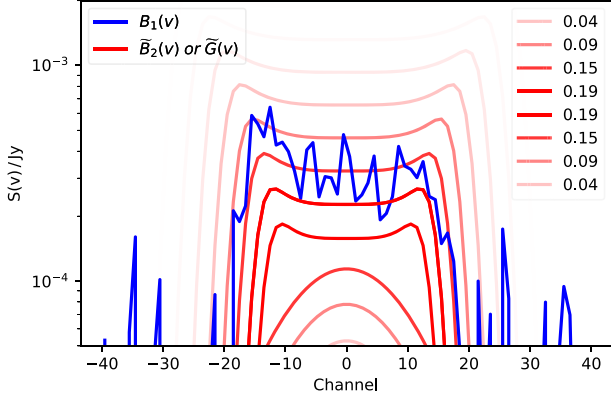


Figure 4. An example of modelling the asymmetric B_1 profile with a 2D flux density model as a function of channel number and H I mass. The blue line is the generated H I emission line added to the noise at $z \sim 0.06$. The red lines are the modelled flux densities with a range of H I masses corresponding to their probabilities indicated in the right legend, including line profiles of a symmetric busy function B_2 and Gaussian function.

for the simulated lines. Based on this equation, we can then model the intrinsic flux densities on each frequency channel for the simulated H I galaxy samples as

$$\tilde{S}(v) = \begin{cases} \tilde{B}_2(v) & \text{for } \sin(i) \geq 0.5 \text{ and } \tilde{W}_{50} \geq 132 \text{ km s}^{-1}, \\ \tilde{G}(v) & \text{otherwise.} \end{cases} \quad (18)$$

In Fig. 4, we show an example of the $B_1(v)$ added to the channel noise in blue for a galaxy with $\log_{10}(M_*/M_\odot) = 8.15$ and $\log_{10}(M_{\text{HI}}/M_\odot) = 9.14$, and the modelled $B_2(v)$ and $\tilde{G}(v)$ profiles in red with a range of H I mass centred at the averaged H I mass of $\log_{10}(M_{\text{HI}}/M_\odot) = 8.8$ from equation (12). The transparency of red lines on the right legend indicates the probability of having the corresponding H I mass. Since the averaged H I mass in our model is slightly lower than the simulated H I mass for the survey, the $B_2(v)$ profile with a maximum probability of 19 per cent is also slightly below the observed (blue) line. With higher or lower H I masses than the $\log_{10}(M_{\text{HI}}/M_\odot) = 8.8$, the probabilities get smaller as expected. In particular, the double-horn profile $B_2(v)$ will change to a Gaussian profile $G(v)$ when the line width W_{50} is less than 132 km s^{-1} . We emphasize that the parameters b and r_c for the B_2 profile are set to be the mean value of the input ranges, while parameters b_1 , b_2 , and r_c for the B_1 profile are randomly assigned.

We note that the $\tilde{S}(v)$ is essentially the 2D flux density $S(v, M_{\text{HI}})$ that we are trying to solve in equation (11), and there are in total five parameters for modelling the flux densities for a given survey, three (A , α , and σ_{HI}) of which are used for characterizing the assumed bTFR and $M_{\text{HI}} - M_*$ relation, with the extra two (b and r_c) to further describe the shape of the measured H I lines. One may wonder why we need these two nuisance variables to model these lines, since we can only predict the line strengths and widths. The point is that modelling the dip of the H I line between the double horns is an important element for our model to be accurate. None of the simpler functions, like Gaussian or top-hat profiles, are able to characterize this dip; therefore, they will inevitably cause a model bias.

In conclusion to this section, we emphasize that the prior knowledge of stellar masses, inclinations, and redshifts of the H I galaxy samples is required for modelling these 2D flux densities effectively. In reality, these data may not be accurate or complete, and eventually more uncertainties may need to be introduced. That is precisely why we are now developing this approach and intend to implement it in

the MIGHTEE fields, where exceptional multiwavelength ancillary data are available. Of course, there may also be non-Gaussian noise or correlated noise in the real H I data, and we can address this issue with a fitted noise distribution to the real data as mentioned in P20. Nevertheless, this method is a useful first step towards the application to real data.

3.2 The likelihood function

Based on Bayes' theorem, the probability of the model given the data is proportional to the likelihood function (i.e. the probability of the data given the model). With the assumption that the behaviour of the noise on intrinsic flux densities of different sources is independent, the likelihood for all the sources having the measured flux densities with our model in an H I survey is given by

$$\mathcal{L} \propto \prod_{\text{source}} P(S_m(v)|A, \alpha, \sigma_{\text{HI}}, b, r_c), \quad (19)$$

where $P(S_m(v)|A, \alpha, \sigma_{\text{HI}}, b, r_c)$ is the probability of having the measured flux density $S_m(v)$ for a single source from equation (11), given the described 5-parameter model above.

As the PDF for the H I mass is characterized in the logarithmic space, we rewrite equation (11) as

$$P(S_m(v)|A, \alpha, \sigma_{\text{HI}}, b, r_c) = \int d \log_{10}(M_{\text{HI}}) P(\log_{10}(M_{\text{HI}})) \prod_v P_n(S_m(v) - S(v, M_{\text{HI}})), \quad (20)$$

where $P(\log_{10}(M_{\text{HI}}))$ follows a normal distribution with mean H I mass determined by equation (2) and standard deviation σ_{HI} . The upper and lower limits of the integral are set to be the mean H I mass ± 2 dex.

At $z < 0.1$, we marginalize the probability over the extra parameters b and r_c to suppress the effect of low-number statistics as

$$P(S_m(v)|A, \alpha, \sigma_{\text{HI}}) = \int d \log_{10}(M_{\text{HI}}) P(\log_{10}(M_{\text{HI}})) \int db P(b) \int dr_c P(r_c) \prod_v P_n(S_m(v) - S(v, M_{\text{HI}})), \quad (21)$$

where $P(b)$ and $P(r_c)$ follow uniform distributions as listed in Table 2.

In practice, we take the logarithm of the likelihood function:

$$\ln \mathcal{L} = \sum_{\text{source}} \log(P(S_m(v))) + \text{constant}. \quad (22)$$

3.3 Parameter estimation

We use MULTINEST (Feroz, Hobson & Bridges 2009; Buchner et al. 2014) to sample the prior parameter space. MULTINEST is an efficient and robust tool for the Bayesian inference, which produces the posterior parameter samples with an associated error estimate. Priors capture our knowledge of a parameter before any new experimental evidence is taken into account. These are listed in Table 2.

3.4 Survey parameters

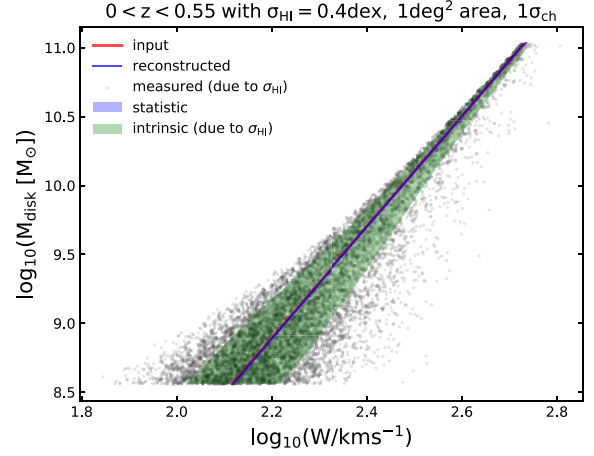
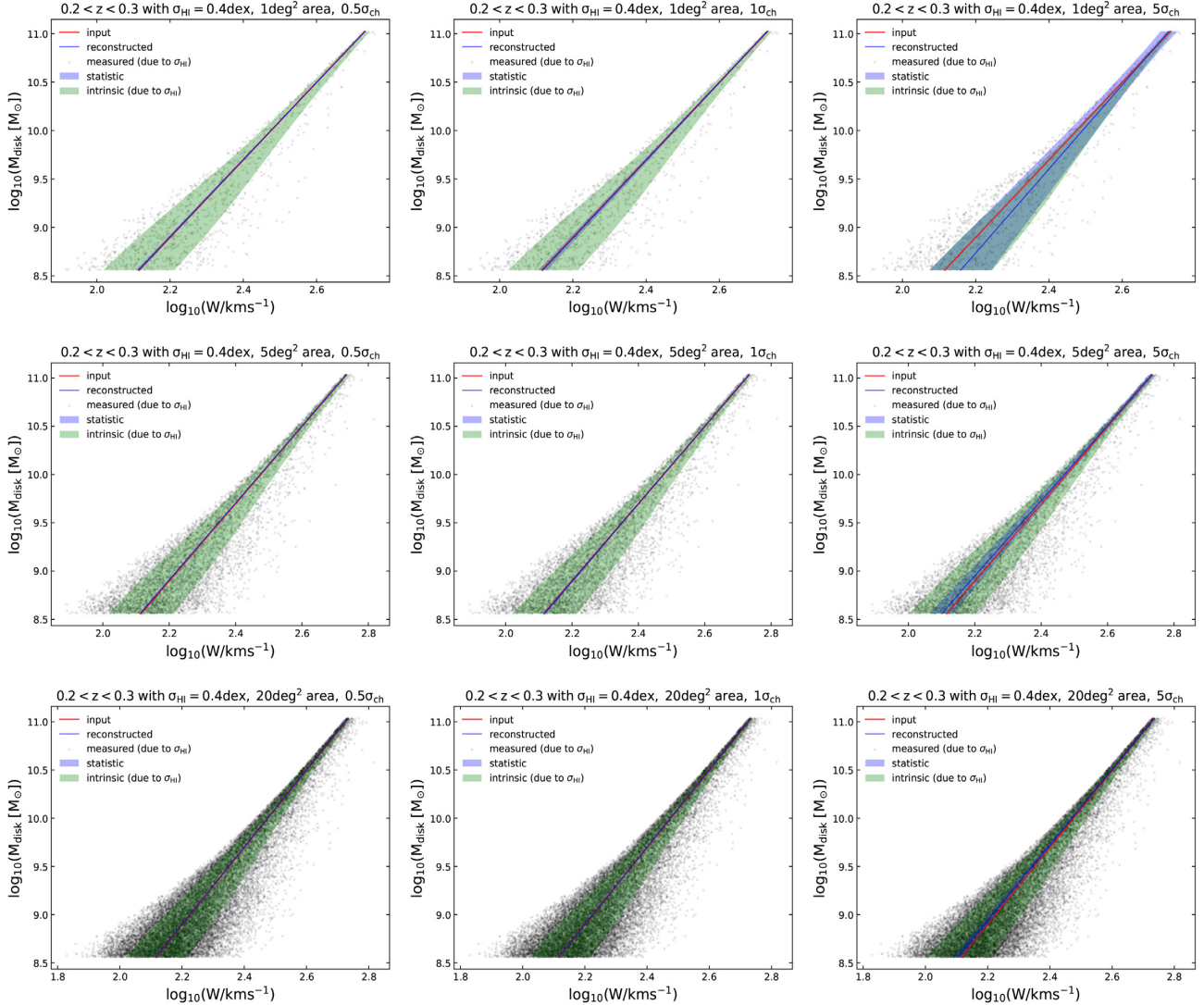
Following P20, we initially assume a 1 deg^2 sky area with noise $\sigma_{\text{ch}} = 90 \mu\text{Jy/channel}$ over the redshift range of $0 < z < 0.55$ as our baseline survey. We then investigate the effect of changing the noise properties of the spectral line cube and the survey area to investigate how they affect our ability to measure the bTFR. Specifically, we change the rms of the spectral line cubes to $0.5\sigma_{\text{ch}}$ and $5\sigma_{\text{ch}}$ and

Table 2. The input parameters of the model for measuring the Tully–Fisher relation, H I and stellar mass relation, and double-horn profiles.

Parameter	Input	Prior probability distribution
$\log_{10}(A)$	9.3	Uniform $\in [7, 12]$
α	4	Uniform $\in [2.5, 7.5]$
σ_{HI}	0.4	Uniform $\in [0.1, 0.9]$
b	0.3, 0.7	Uniform $\in [0, 1]$
r_c	1.5, 2	Uniform $\in [1, 3]$

increase the survey area from 1 deg^2 to 20 deg^2 . We also simulate a second group of galaxies for surveys with the same parameter setup but with only Gaussian emission line profiles to investigate the effect of fitting double-horn profiles. To demonstrate the key elements of our work, we incorporate the scatter in the $M_{\text{HI}} - M_\star$ relationship into our model but fix the slope and normalization, although we note that these could also be included in the modelling.

Here, we set the redshift bin width to be 0.1 for studying the effect of redshift evolution on our approach. This binning scheme is

**Figure 6.** The reconstructed baryonic Tully–Fisher relation from a broad z-bin with a $1\sigma_{\text{ch}}$ noise.**Figure 5.** Reconstructed baryonic Tully–Fisher relation from a single z-bin ($0.2 < z < 0.3$) with the survey area increased from 1 to 20 deg^2 from top to bottom panels. Left: reconstructed from $0.5\sigma_{\text{ch}}$ noise. Middle: $1\sigma_{\text{ch}}$ noise. Right: $5\sigma_{\text{ch}}$ noise.

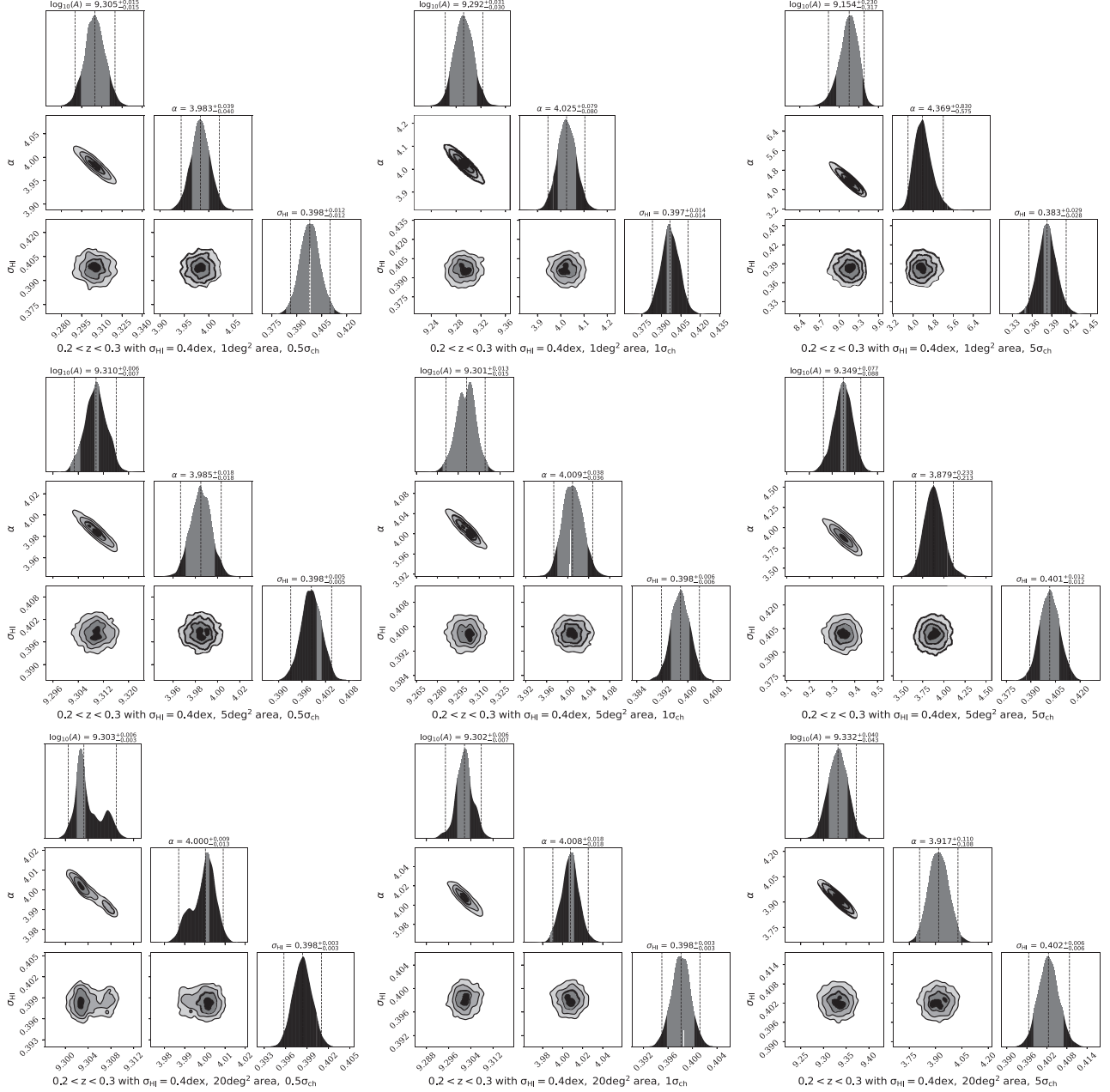


Figure 7. The posterior distributions of reconstructed parameters from a single z -bin ($0.2 < z < 0.3$) with the survey area increased from 1 to 5 deg^2 and up to 20 deg^2 from top to bottom panels. Note that the scales of axes are different across the panels. Left: reconstructed from $0.5\sigma_{\text{ch}}$ noise. Middle: $1\sigma_{\text{ch}}$ noise. Right: $5\sigma_{\text{ch}}$ noise.

relatively coarse rather than optimal to demonstrate that our approach is insensitive to the way of binning the data sets. In principle, a smaller bin width to further extract hidden information could be used.

4 RESULTS

4.1 The reconstructed Tully–Fisher relation from our baseline survey

Fig. 5 shows the reconstructed bTFR with $\sigma_{\text{H1}} = 0.4 \text{ dex}$ on the $M_{\text{H1}} - M_*$ relation at $0.2 < z < 0.3$. The result in the top middle panel is the reconstructed bTFR for our baseline survey. The input (simulated) relation is shown by the red line (in the absence of any

scatter). The simulated disc masses M_{disc} are denoted by the black points. We note that these disc masses are estimated using the prior knowledge of the stellar mass and the $M_{\text{H1}} - M_*$ relation along with the scatter (without knowing the accurate H1 mass). The velocity widths W on the abscissa are the input values based on the input H1 mass. Therefore, the scatter is present only in the measured disc masses (grey points). In reality, the scatter may be higher as there is also scatter in the H1 mass to the velocity width for individual galaxies, which would manifest in the scatter on the bTFR itself.

The best-fitting model of the bTFR using our simulation is shown with the blue line, and the 68 percent credible intervals estimated from the posterior distributions of reconstructed parameters A and α are denoted by the blue region. The H1 signal for those galaxies with

Table 3. Reconstructed parameters of our bTFR model from 1, 5, and 20 deg² surveys with changing noise. Note that we show only the results from combining the H I samples at $0 < z < 0.3$ as we find that this redshift range provides the strongest constraints on our model.

Parameter	$\times \sigma_{\text{ch}}$	$z < 0.1$	$0.1 < z < 0.2$	$0.2 < z < 0.3$	$0.3 < z < 0.4$	$0.4 < z < 0.5$	$0.5 < z < 0.55$	$0 < z < 0.3$
1 deg ²								
$\log_{10}(A)$	0.5	$9.301^{+0.001}_{-0.001}$	$9.302^{+0.003}_{-0.004}$	$9.305^{+0.008}_{-0.007}$	$9.293^{+0.013}_{-0.008}$	$9.323^{+0.015}_{-0.016}$	$9.363^{+0.033}_{-0.031}$	$9.299^{+0.001}_{-0.000}$
$\log_{10}(A)$	1.0	$9.294^{+0.001}_{-0.001}$	$9.303^{+0.007}_{-0.007}$	$9.292^{+0.015}_{-0.016}$	$9.271^{+0.027}_{-0.033}$	$9.338^{+0.041}_{-0.042}$	$9.447^{+0.055}_{-0.052}$	$9.294^{+0.001}_{-0.001}$
$\log_{10}(A)$	5.0	$9.298^{+0.008}_{-0.007}$	$9.393^{+0.042}_{-0.045}$	$9.153^{+0.133}_{-0.153}$	$9.417^{+0.127}_{-0.151}$	$9.337^{+0.185}_{-0.271}$	$9.48^{+0.206}_{-0.443}$	$9.3^{+0.006}_{-0.008}$
α	0.5	$4.003^{+0.003}_{-0.004}$	$4.009^{+0.012}_{-0.009}$	$3.983^{+0.02}_{-0.02}$	$4.028^{+0.03}_{-0.038}$	$3.925^{+0.046}_{-0.04}$	$3.884^{+0.087}_{-0.082}$	$4.024^{+0.001}_{-0.001}$
α	1.0	$4.021^{+0.006}_{-0.005}$	$4.01^{+0.02}_{-0.021}$	$4.025^{+0.04}_{-0.039}$	$4.113^{+0.089}_{-0.076}$	$3.938^{+0.103}_{-0.105}$	$3.702^{+0.148}_{-0.139}$	$4.033^{+0.002}_{-0.003}$
α	5.0	$4.03^{+0.038}_{-0.032}$	$3.791^{+0.133}_{-0.122}$	$4.369^{+0.374}_{-0.328}$	$3.863^{+0.424}_{-0.379}$	$4.476^{+0.77}_{-0.551}$	$3.395^{+1.32}_{-0.6}$	$4.023^{+0.013}_{-0.015}$
$\sigma_{\text{H I}}$	0.5	$0.384^{+0.012}_{-0.013}$	$0.419^{+0.009}_{-0.009}$	$0.398^{+0.006}_{-0.006}$	$0.392^{+0.005}_{-0.005}$	$0.397^{+0.005}_{-0.005}$	$0.407^{+0.007}_{-0.007}$	$0.404^{+0.004}_{-0.003}$
$\sigma_{\text{H I}}$	1.0	$0.399^{+0.016}_{-0.012}$	$0.421^{+0.01}_{-0.01}$	$0.397^{+0.007}_{-0.007}$	$0.396^{+0.007}_{-0.006}$	$0.39^{+0.007}_{-0.007}$	$0.415^{+0.01}_{-0.009}$	$0.405^{+0.006}_{-0.004}$
$\sigma_{\text{H I}}$	5.0	$0.399^{+0.019}_{-0.018}$	$0.431^{+0.015}_{-0.014}$	$0.383^{+0.014}_{-0.014}$	$0.396^{+0.017}_{-0.017}$	$0.366^{+0.02}_{-0.021}$	$0.383^{+0.044}_{-0.055}$	$0.408^{+0.009}_{-0.007}$
5 deg ²								
$\log_{10}(A)$	0.5	$9.299^{+0.000}_{-0.000}$	$9.305^{+0.001}_{-0.001}$	$9.31^{+0.003}_{-0.003}$	$9.293^{+0.004}_{-0.005}$	$9.321^{+0.009}_{-0.007}$	$9.327^{+0.016}_{-0.016}$	$9.3^{+0.000}_{-0.000}$
$\log_{10}(A)$	1.0	$9.299^{+0.001}_{-0.000}$	$9.306^{+0.003}_{-0.003}$	$9.301^{+0.006}_{-0.007}$	$9.292^{+0.01}_{-0.012}$	$9.321^{+0.015}_{-0.024}$	$9.35^{+0.031}_{-0.036}$	$9.299^{+0.000}_{-0.000}$
$\log_{10}(A)$	5.0	$9.298^{+0.002}_{-0.002}$	$9.339^{+0.02}_{-0.019}$	$9.349^{+0.04}_{-0.044}$	$9.442^{+0.059}_{-0.061}$	$9.25^{+0.104}_{-0.122}$	$9.21^{+0.172}_{-0.216}$	$9.3^{+0.002}_{-0.003}$
α	0.5	$4.007^{+0.000}_{-0.000}$	$3.999^{+0.004}_{-0.003}$	$3.985^{+0.009}_{-0.009}$	$4.032^{+0.014}_{-0.014}$	$3.952^{+0.021}_{-0.022}$	$3.947^{+0.044}_{-0.044}$	$4.007^{+0.000}_{-0.000}$
α	1.0	$4.008^{+0.001}_{-0.001}$	$3.993^{+0.009}_{-0.009}$	$4.009^{+0.02}_{-0.019}$	$4.06^{+0.032}_{-0.031}$	$3.955^{+0.051}_{-0.044}$	$3.886^{+0.086}_{-0.082}$	$4.008^{+0.000}_{-0.001}$
α	5.0	$4.009^{+0.003}_{-0.003}$	$3.928^{+0.054}_{-0.054}$	$3.879^{+0.114}_{-0.112}$	$3.632^{+0.162}_{-0.16}$	$4.151^{+0.362}_{-0.303}$	$4.0^{+0.588}_{-0.452}$	$4.007^{+0.004}_{-0.002}$
$\sigma_{\text{H I}}$	0.5	$0.398^{+0.007}_{-0.005}$	$0.398^{+0.004}_{-0.004}$	$0.398^{+0.003}_{-0.003}$	$0.397^{+0.002}_{-0.002}$	$0.399^{+0.002}_{-0.002}$	$0.401^{+0.003}_{-0.003}$	$0.4^{+0.001}_{-0.002}$
$\sigma_{\text{H I}}$	1.0	$0.398^{+0.006}_{-0.007}$	$0.401^{+0.004}_{-0.004}$	$0.398^{+0.003}_{-0.003}$	$0.395^{+0.003}_{-0.003}$	$0.399^{+0.003}_{-0.003}$	$0.401^{+0.005}_{-0.005}$	$0.402^{+0.002}_{-0.002}$
$\sigma_{\text{H I}}$	5.0	$0.405^{+0.008}_{-0.009}$	$0.402^{+0.006}_{-0.006}$	$0.401^{+0.006}_{-0.006}$	$0.407^{+0.007}_{-0.007}$	$0.404^{+0.009}_{-0.01}$	$0.407^{+0.015}_{-0.017}$	$0.404^{+0.004}_{-0.003}$
20 deg ²								
$\log_{10}(A)$	0.5	$9.302^{+0.000}_{-0.000}$	$9.307^{+0.001}_{-0.001}$	$9.303^{+0.004}_{-0.001}$	$9.298^{+0.005}_{-0.003}$	$9.316^{+0.005}_{-0.005}$	$9.308^{+0.01}_{-0.007}$	$9.302^{+0.000}_{-0.000}$
$\log_{10}(A)$	1.0	$9.302^{+0.000}_{-0.000}$	$9.303^{+0.001}_{-0.001}$	$9.302^{+0.003}_{-0.003}$	$9.319^{+0.004}_{-0.007}$	$9.321^{+0.008}_{-0.011}$	$9.32^{+0.014}_{-0.014}$	$9.302^{+0.000}_{-0.000}$
$\log_{10}(A)$	5.0	$9.302^{+0.000}_{-0.000}$	$9.316^{+0.009}_{-0.009}$	$9.332^{+0.021}_{-0.022}$	$9.347^{+0.03}_{-0.032}$	$9.247^{+0.063}_{-0.064}$	$9.477^{+0.068}_{-0.084}$	$9.302^{+0.000}_{-0.000}$
α	0.5	$4.0^{+0.000}_{-0.000}$	$3.992^{+0.002}_{-0.002}$	$4.0^{+0.004}_{-0.008}$	$4.022^{+0.009}_{-0.012}$	$3.97^{+0.012}_{-0.012}$	$3.983^{+0.022}_{-0.022}$	$4.0^{+0.000}_{-0.000}$
α	1.0	$4.0^{+0.000}_{-0.000}$	$4.002^{+0.004}_{-0.004}$	$4.008^{+0.008}_{-0.009}$	$3.96^{+0.017}_{-0.013}$	$3.957^{+0.026}_{-0.024}$	$3.945^{+0.041}_{-0.042}$	$4.0^{+0.000}_{-0.000}$
α	5.0	$3.999^{+0.001}_{-0.001}$	$3.967^{+0.023}_{-0.025}$	$3.917^{+0.058}_{-0.056}$	$3.934^{+0.084}_{-0.083}$	$4.384^{+0.2}_{-0.19}$	$3.42^{+0.214}_{-0.199}$	$3.999^{+0.001}_{-0.001}$
$\sigma_{\text{H I}}$	0.5	$0.398^{+0.004}_{-0.002}$	$0.4^{+0.002}_{-0.002}$	$0.398^{+0.001}_{-0.001}$	$0.399^{+0.001}_{-0.001}$	$0.401^{+0.001}_{-0.001}$	$0.403^{+0.002}_{-0.001}$	$0.402^{+0.001}_{-0.001}$
$\sigma_{\text{H I}}$	1.0	$0.402^{+0.003}_{-0.003}$	$0.4^{+0.002}_{-0.002}$	$0.398^{+0.001}_{-0.002}$	$0.4^{+0.001}_{-0.001}$	$0.402^{+0.001}_{-0.002}$	$0.405^{+0.002}_{-0.002}$	$0.401^{+0.001}_{-0.001}$
$\sigma_{\text{H I}}$	5.0	$0.405^{+0.005}_{-0.005}$	$0.404^{+0.003}_{-0.003}$	$0.402^{+0.003}_{-0.003}$	$0.404^{+0.004}_{-0.004}$	$0.405^{+0.005}_{-0.005}$	$0.412^{+0.008}_{-0.008}$	$0.404^{+0.002}_{-0.002}$

low disc mass is very faint and more susceptible to the influence of the noise. Thus, the statistical uncertainty from the measurements becomes larger at low mass. On the other hand, the noise becomes less influential for the more massive galaxies. However, since the number of most massive galaxies is low, the strongest constraints on the bTFR are provided by the middle upper range of disc mass at $M_{\text{disc}} \sim 10^{10.5} M_{\odot}$. This effect can be seen in the top right-hand panel of Fig. 5, where we show the constraints on the bTFR for a significantly elevated noise. Therefore, a different form of SMF or colour selection that affects only the number of objects at the highest mass end will have little impact on our results.

The intrinsic uncertainty of the bTFR, propagated from the reconstructed $\sigma_{\text{H I}}$ on the $M_{\text{H I}} - M_{\star}$ relation, is shown with the green area. Low-mass galaxies tend to be more H I dominated; therefore, the constant $\sigma_{\text{H I}}$ will have a stronger impact on the lower mass end compared to the higher mass end, which tend to be dominated by stellar mass.

Our ability to constrain the redshift dependence of $\sigma_{\text{H I}}$ is consistent with P20 (i.e. the strongest constraints on the $M_{\text{H I}} - M_{\star}$ relation are provided by the middle redshift range). This is due to the fact that the power of Bayesian stacking technique relies on data in the

sweet-spot of having a sufficiently large number of galaxies that dominate over the noise at a given integrated flux limit. However, the constraints on the bTFR parameters (i.e. A and α) are getting stronger at lower redshifts as the bTFR parameters are very relevant to the H I line profile that are more prominent from nearby sources. We list all the reconstructed parameters across the redshift bins in Table 3. If we assume that the bTFR does not evolve with redshift, then we can reconstruct bTFR from the much broader redshift range defined by the limit of our observational bandwidth (Fig. 6). This is obviously better constrained in comparison to the single redshift bins as expected, and all the posterior distributions are shown in Figs 7 and 8. We also represent the reconstructed parameters in Fig. 9 for an easy visual inspection.

We note again that there are five free parameters to be constrained for the simulated H I samples at $z > 0.1$, as we need the extra two parameters for describing the shape of the double horns, and there is no need to marginalize over them due to the large source density in the high-redshift regime, but we can treat them as nuisance variables since we are interested only in measuring the bTFR here; therefore, we do not show them in Fig. 7. Nevertheless, as we demonstrate in Fig. 8, the reconstructed nuisance parameters b and r_c are also very

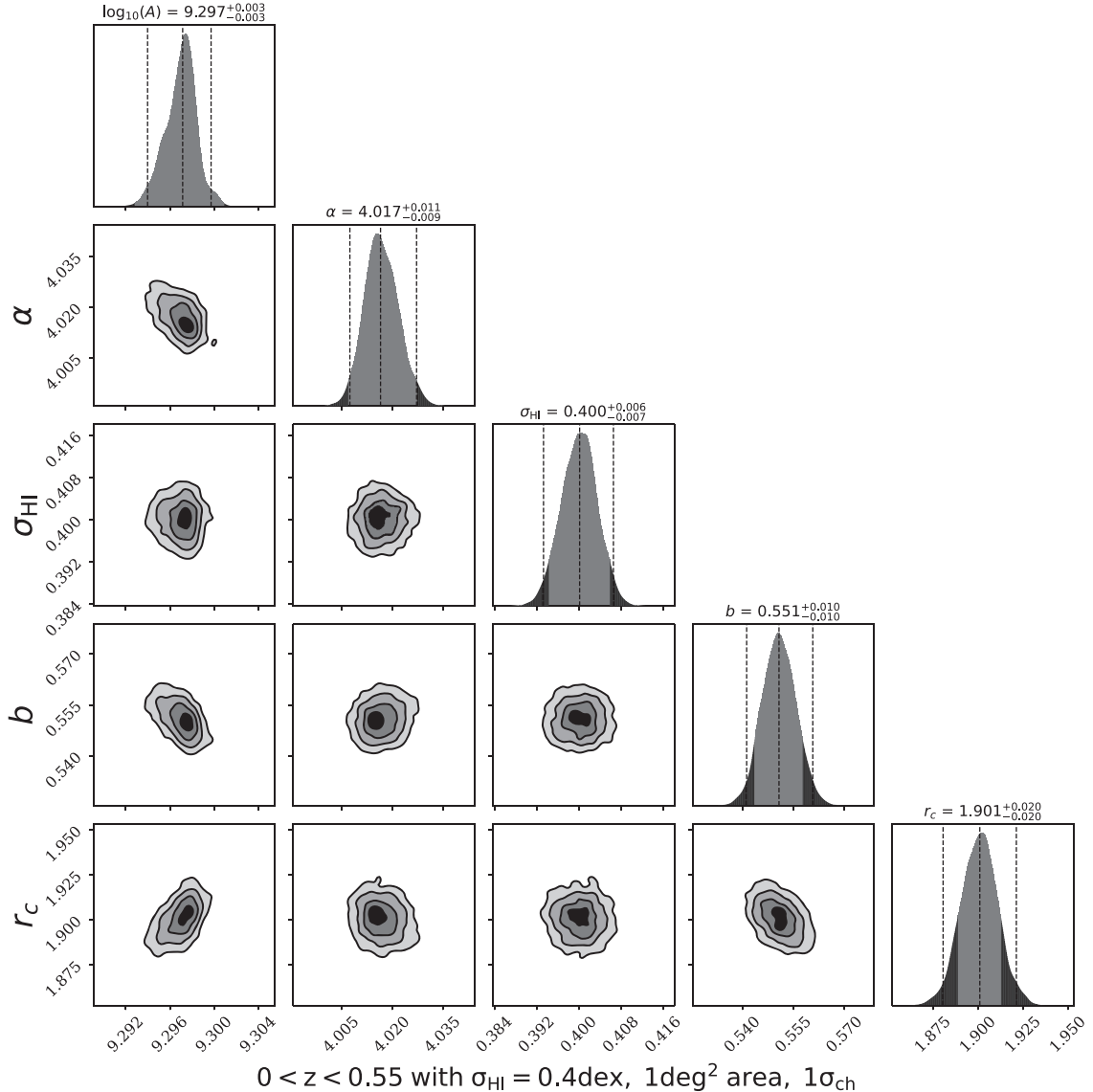


Figure 8. The posterior distributions of reconstructed parameters from a broad z -bin with a $1\sigma_{\text{ch}}$ noise.

close to their input ranges, which indicates that our model can also provide constraints of the line steepness and the depth of the trough.

To evaluate the effect of the double-horn profiles on our results, we have also simulated a group of HI lines with only Gaussian profiles, where the nuisance parameters are not needed. For the same survey setup, we find that the uncertainties for the reconstructed parameters A , α , and σ_{HI} increase slightly. This means that the double-horn profiles are in fact contributing positively to measuring the bTFR, since their profiles with steep line flanks are more easily characterized in the noisy flux densities than the Gaussian profiles for our simulations.

4.2 The effects of changing the noise and survey area

We also show the reconstructed bTFR from the cubes with $0.5\sigma_{\text{ch}}$ and $5\sigma_{\text{ch}}$ Gaussian noise in the left-hand and right-hand panels of Fig. 5 and with a range of survey areas from top to bottom panels. This layout is the same for Fig. 7.

The obvious effect of the increased noise is shown with the blue area of the top right-hand panel of Fig. 5, where the statistical uncertainties increase significantly, following the increase in noise by a factor of 5. This also has an impact on the redshift range that best constrains σ_{HI} for the $M_{\text{HI}} - M_{\star}$ relation, but the effect on the redshift dependence of the bTFR parameters is less significant, since the ratio of signal to noise is greatly reduced at all redshifts.

Increasing the survey area has great benefits for measuring the bTFR accurately, especially in the presence of a strong noise as the increase in number of galaxies compensates for the increase in instrumental noise, i.e. the effect of \sqrt{N} . At the higher redshifts, the $5\sigma_{\text{ch}}$ noise significantly reduces the overall signal-to-noise ratio to the point where even the largest survey area cannot remedy this impact. Nevertheless, the increase of the survey area by a factor of 5 clearly provides better (about twice stronger on average) constraints for all the parameters as indicated most clearly by the last column of Table 3 where we take into account all the single z -bins at $z < 0.3$ together.

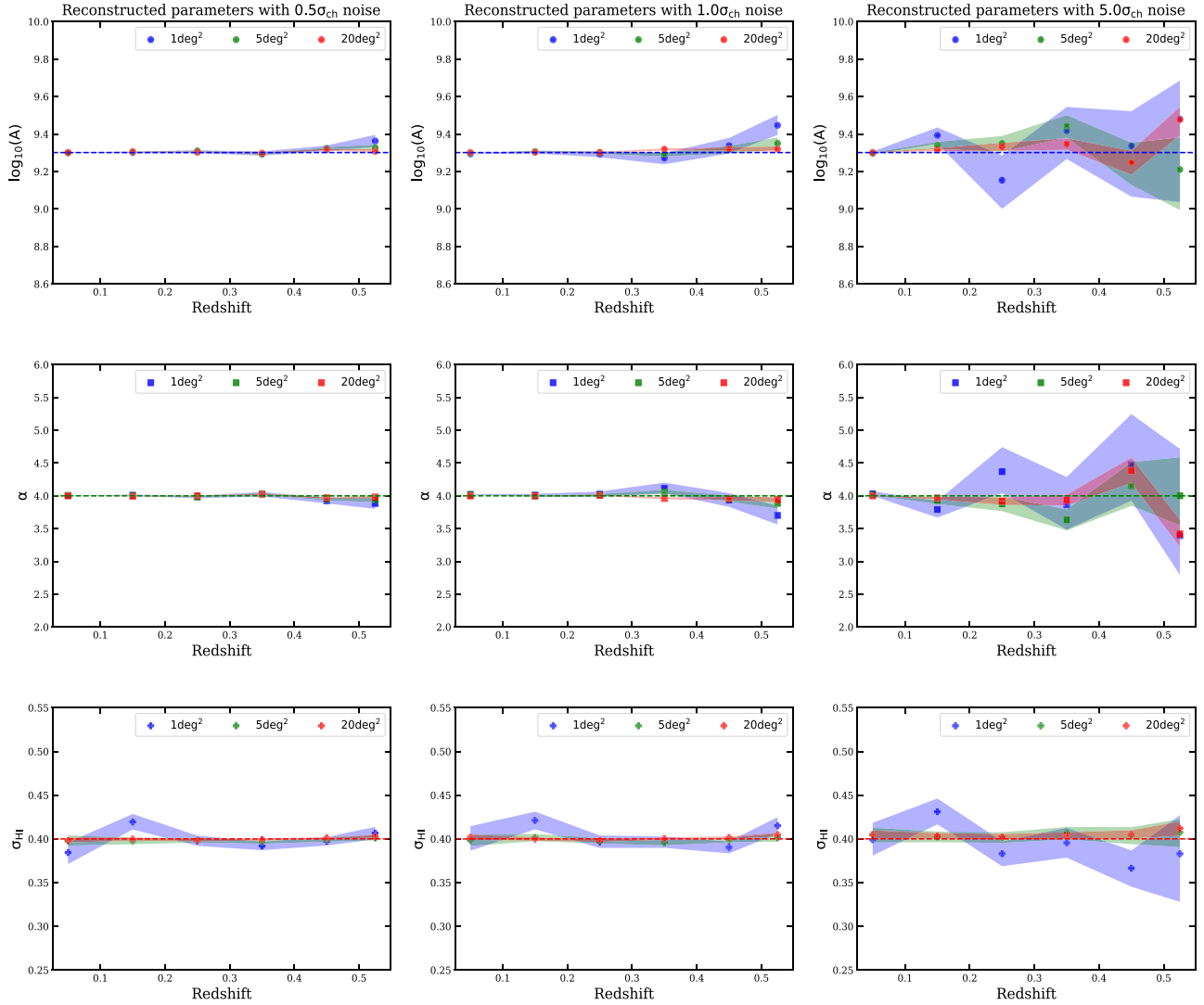


Figure 9. Reconstructed parameters of our bTFR model from the simulated surveys as a function of redshift. Left: reconstructed from $0.5\sigma_{\text{ch}}$ noise. Middle: $1\sigma_{\text{ch}}$ noise. Right: $5\sigma_{\text{ch}}$ noise.

5 CONCLUSIONS

We present a 2D flux density model for the observed H I emission lines from galaxies and a Bayesian stacking technique for measuring the bTFR and the $M_{\text{HI}} - M_{\star}$ relation over the redshift range $0 < z < 0.55$, below the detection threshold and down to $M_{\star} = 10^7 M_{\odot}$. We simulate the galaxy catalogues and H I cubes with a range of noise properties and survey areas, and we find that:

- (i) Our model can reproduce the input bTFR parameters most accurately in a redshift range from the nearby Universe to $z = 0.3$ for all levels of noise and sky areas and up to $z = 0.55$ for the nominal level of noise and the wider areas than that of the baseline survey. This is due to the strong signal-to-noise ratio (SNR) of H I lines, and the number of galaxies covering a wide mass range with a MIGHTEE-like survey, which is important to constrain the normalization and slope of the bTFR.
- (ii) The strongest constraints on the $M_{\text{HI}} - M_{\star}$ relation are provided by the middle redshift range ($0.2 < z < 0.3$) due to a sufficiently large number of galaxies that dominate over the noise at a given integrated flux limit.

(iii) Our model for recovering the bTFR does not perform as well with higher noise at higher redshifts ($z > 0.3$) as expected. This is predominantly due to the reduced number of sources that lie just above the nominal noise threshold.

(iv) With a variety of shapes of the double-horn profiles, our technique shows an excellent potential for measuring the bTFR and the $M_{\text{HI}} - M_{\star}$ relation at high redshifts, where the direct measurements are not possible due to the intrinsic faintness of the H I signal. Moreover, the recovery of the velocity and H I mass of the low-mass galaxies will be a major step forward in better constraining the statistical properties of the bTFR. Our model can also predict the distribution in the shape of H I emission lines below the detection threshold and shed light on the dynamics of galaxies, which would remain undetected otherwise. When we apply our technique to real data, it will be possible to perform direct comparisons between the observations and cosmological simulations of galaxy formation and evolution beyond the local Universe and for the first time in the low-mass regime.

Finally, it is important to note that since our approach for measuring the bTFR is in fact based on the Bayesian stacking technique used

in P20, this approach will inevitably inherit the similar limitations listed at the end of P20 such as dealing with the real noise and instrumental effects in radio observations. We will address these by applying our novel technique to the real MIGHTEE survey data set in our upcoming work.

ACKNOWLEDGEMENTS

We are grateful to the anonymous referee for helpful comments that have improved this paper. HP, MJJ, and MGS acknowledge support from the South African Radio Astronomy Observatory (SARAO) towards this research (<http://www.sarao.ac.za>). MGS also acknowledges support from the National Research Foundation (grant no. 84156). This work was also supported by the China Scholarship Council and the Science and Technology Facilities Council (STFC) of the United Kingdom. AAP acknowledges the support of the STFC consolidated grant ST/S000488/1. NM acknowledges support from the Bundesministerium für Bildung und Forschung (BMBF) award 05A20WM4.

DATA AVAILABILITY

All data used in this paper are available on request to the corresponding authors.

REFERENCES

- Abril-Melgarejo V. et al., 2021, *A&A*, 647, A152
 Arnett D., 1996, *Supernovae and Nucleosynthesis: An Investigation of the History of Matter from the big bang to the Present*. Princeton University Press, Princeton
 Bedregal A. G., Aragón-Salamanca A., Merrifield M. R., 2006, *MNRAS*, 373, 1125
 Begum A., Chengalur J. N., Karachentsev I. D., Sharina M. E., 2008, *MNRAS*, 386, 138
 Bradford J. D., Geha M. C., van den Bosch F. C., 2016, *ApJ*, 832, 11
 Buchner J. et al., 2014, *A&A*, 564, A125
 Chowdhury A., Kanekar N., Chengalur J. N., Sethi S., Dwarakanath K. S., 2020, *Nature*, 586, 369–372
 Chung A., van Gorkom J. H., O’Neil K., Bothun G. D., 2002, *AJ*, 123, 2387
 Courteau S., Andersen D. R., Bershadsky M. A., MacArthur L. A., Rix H.-W., 2003, *ApJ*, 594, 208
 Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
 Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, *MNRAS*, 433, 1398
 Desmond H., 2012, preprint ([arXiv:1204.1497](https://arxiv.org/abs/1204.1497))
 Dubois Y. et al., 2020, *A&A*, 651, A109
 Fernández X. et al., 2016, *ApJ*, 824, L1
 Feroz F., Hobson M. P., Bridges M., 2009, *MNRAS*, 398, 1601
 Glowacki M., Elson E., Davé R., 2020a, *MNRAS*, 507, 3267

- Glowacki M., Elson E., Davé R., 2020b, *MNRAS*, 498, 3687
 Iorio G., Fraternali F., Nipoti C., Di Teodoro E., Read J. I., Battaglia G., 2017, *MNRAS*, 466, 4159
 Jarvis M. et al., 2016, in *Proceedings of MeerKAT Science: On the Pathway to the SKA*. 25–27 May. Stellenbosch, South Africa, p.6
 Jonas J., MeerKAT Team, 2016, in *Proceedings of MeerKAT Science: On the Pathway to the SKA*. 25–27 May. Stellenbosch, South Africa, p. 1
 Karachentsev I. D., Kaisina E. I., Kashibadze Nasonova O. G., 2017, *AJ*, 153, 6
 Lelli F., McGaugh S. S., Schombert J. M., Desmond H., Katz H., 2019, *MNRAS*, 484, 3267
 Maddox N., Hess K. M., Obreschkow D., Jarvis M. J., Blyth S. L., 2015, *MNRAS*, 447, 1610
 Maddox N. et al., 2021, *A&A*, 646, A35
 McGaugh S. S., Schombert J. M., Bothun G. D., de Blok W. J. G., 2000, *ApJ*, 533, L99
 Meyer S. A., Meyer M., Obreschkow D., Staveley-Smith L., 2016, *MNRAS*, 455, 3136
 Meyer M., Robotham A., Obreschkow D., Westmeier T., Duffy A. R., Staveley-Smith L., 2017, *PASA*, 34, 52
 Pan H., Jarvis M. J., Allison J. R., Heywood I., Santos M. G., Maddox N., Frank B. S., Kang X., 2020, *MNRAS*, 491, 1227
 Papastergis E., Cattaneo A., Huang S., Giovanelli R., Haynes M. P., 2012, *ApJ*, 759, 138
 Papastergis E., Adams E. A. K., van der Hulst J. M., 2016, *A&A*, 593, A39
 Parkash V., Brown M. J. I., Jarrett T. H., Bonne N. J., 2018, *ApJ*, 864, 40
 Planck Collaboration XIII, 2016, *A&A*, 594, A13
 Ponomareva A. A., Verheijen M. A. W., Peletier R. F., Bosma A., 2017, *MNRAS*, 469, 2387
 Ponomareva A. A., Verheijen M. A. W., Papastergis E., Bosma A., Peletier R. F., 2018, *MNRAS*, 474, 4366
 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
 Schechter P., 1976, *ApJ*, 203, 297
 Sorce J. G., Tully R. B., Courtois H. M., Jarrett T. H., Neill J. D., Shaya E. J., 2014, *MNRAS*, 444, 527
 Tiley A. L. et al., 2019, *MNRAS*, 482, 2166
 Topal S., Bureau M., Tiley A. L., Davis T. A., Torii K., 2018, *MNRAS*, 479, 3319
 Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, *ApJ*, 742, 16
 Tully R. B., Fisher J. R., 1977, *A&A*, 500, 105
 Tully R. B., Courtois H., Hoffman Y., Pomarède D., 2014, *Nature*, 513, 71
 Übler H. et al., 2017, *ApJ*, 842, 121
 Westmeier T., Jurek R., Obreschkow D., Koribalski B. S., Staveley-Smith L., 2014, *MNRAS*, 438, 1176
 Willick J. A., 1999, *ApJ*, 516, 47
 Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, *MNRAS*, 359, L30

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