Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **484,** 1007–1020 (2019)
Advance Access publication 2018 November 22

doi:10.1093/mnras/sty3142

Intensity mapping cross-correlations II: HI halo models including shot noise

L. Wolz,[1,2★] S. G. Murray,[3,2] C. Blake[2,4] and J. S. Wyithe[1,2]

[1] *School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[2] *ARC Centre of Excellence for All-Sky Astrophysics (CAASTRO)*
[3] *ICRAR, Curtin Institute of Radio Astronomy, GPO Box U1987, Perth, WA 6845, Australia*
[4] *Centre for Astrophysics & Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia*

Accepted 2018 November 17. Received 2018 October 25; in original form 2018 March 6

ABSTRACT

H I intensity mapping data traces the large-scale structure matter distribution using the integrated emission of neutral hydrogen gas (H I). Cross-correlation of the intensity maps with optical galaxy surveys can mitigate foreground and systematic effects, but has been shown to significantly depend on galaxy evolution parameters of the H I and the optical sample. Previously, we have shown that the shot noise of the cross-correlation scales with the H I content of the optical samples, such that the shot noise estimation infers the average H I masses of these samples. In this paper, we present an adaptive framework for the cross-correlation of H I intensity maps with galaxy samples using our implementation of the halo model formalism which utilizes the halo occupation distribution of galaxies to predict their power spectra. We compare two H I population models, tracing the spatial halo and the galaxy distribution, respectively, and present their auto- and cross-power spectra with an associated galaxy sample. We find that the choice of the H I model and the distribution of the H I within the galaxy sample have little impact for the shape of the auto- and cross-correlations, but significantly affects the measured shot noise amplitude of the estimators, a finding we confirm with simulations. We demonstrate parameter estimation of the H I halo occupation models and advocate this framework for the interpretation of future experimental data, with the prospect of determining the H I masses of optical galaxy samples via the cross-correlation shot noise.

Key words: cosmology: theory – cosmology: large-scale structure of Universe – radio lines: galaxies.

1 INTRODUCTION

The cosmological evolution of our Universe can be tested via probes of the statistics of large-scale structure. Common techniques include measuring the Baryon Acoustic Oscillations (BAOs), which act as a standard ruler for distance measures constraining the Cosmic acceleration (see e.g. Reid et al. 2012; Anderson et al. 2014), as well as galaxy clustering which employs the positions of galaxies to measure their cosmological power spectrum (for instance Seljak et al. 2005; Percival et al. 2007). Both BAO measurements and galaxy clustering require the determination of millions of galaxy positions over large volumes in order to minimize statistical uncertainties. Traditionally, optical telescopes have been employed for cosmological measurements as radio telescopes are limited in sensitivity. Beyond the nearby Universe, the determination of redshifts at radio frequencies, at which the spectrum is close to featureless, is extremely challenging. The most notable radio spectral line, at a rest-wavelength of 21 cm, is caused by the spin-flip of the neutral hydrogen (H I) and is comparably weak. It has only been directly detected up to $z = 0.36$ in a single object (Fernandez et al. 2016), and in the statistically averaged spectrum via H I stacking up to $z \approx 0.32$ (Rhee et al. 2018). To circumvent these limitations, H I intensity mapping provides a novel technique to map the large-scale structure distribution as traced by neutral hydrogen gas via low-resolution observations of the integrated and unresolved 21 cm emission of multiple objects.

After H I intensity mapping was proposed as a test of cosmology more than a decade ago (see Battye, Davies & Weller 2004; Wyithe, Loeb & Geil 2008; Chang et al. 2008), Pen et al. (2009) reported the first detection of structure in H I intensity maps of the local Universe. Later, Chang et al. (2010) reported a detection in observations around $z \approx 0.8$. The challenges in detecting the H I intensity mapping power spectrum arise due to the weakness of the redshifted H I signal in comparison to the radio foregrounds in combination with the radio telescope systematics, see e.g. Switzer

★ E-mail: laura.wolz@unimelb.edu.au

Published by Oxford University Press on behalf of the Royal Astronomical Society

et al. (2015), Wolz et al. (2017a), and Harper et al. (2018). The cross-correlation signal of an H I map with an overlapping galaxy survey is insensitive to many of these systematics and increases the significance of detection. The detection of the cosmological distribution via the power spectrum (Masui et al. 2013) was achieved by measurements of the Green Bank telescope at medium redshift $z = 0.8$ in cross-correlation with the WiggleZ Dark Energy survey (Drinkwater et al. 2010), constraining the H I energy density and the H I bias to $\Omega_{\rm H\,I} b_{\rm H\,I} = 0.63^{+0.23}_{-0.15} \times 10^{-3}$ (Switzer et al. 2013). A more recent detection has also been made, using the cross-correlation of the H I intensity maps of the Parkes telescope with the 2dF Galaxy Redshift Survey at $z \approx 0.08$ (Anderson et al. 2018). The analysis presents a 5-sigma detection of the cross-power spectrum with a significant drop of the power on smaller scales, $k \approx 1.5\,h\,{\rm Mpc}^{-1}$, indicating a strong anticorrelation of H I with the red galaxy sample.

The future of H I intensity mapping looks very promising, as a large number of purpose-built instruments are under design and construction. The instruments can be divided into three categories: single-dish telescopes similar to the pioneering Green Bank and Parkes telescopes or those equipped with multibeam receivers (e.g. BINGO Battye et al. 2013), dish interferometers such as HIRAX (Newburgh et al. 2016), and cylindrical dish interferometers such as CHIME (Bandura et al. 2014) or Tianlai (Chen 2012). Additionally, the Square Kilometre Array, an international radio interferometer with unprecedented scale and sensitivity, will conduct H I intensity mapping for wide ranges of redshifts $0 < z < 6$ (Bull et al. 2015; Santos et al. 2015). Two SKA pathfinder projects, MeerKAT and the Australian SKA Pathfinder (ASKAP), are capable of intensity mapping observations, and will be able to explore different observational techniques such as the employment of the array in single-dish mode (Santos et al. 2017) or phased array feeds, in preparation for the SKA observations to commence in the next decade. Forecasts predict that the future SKA H I intensity mapping experiments will be able to measure distances via BAOs to a level that is comparable to Stage IV optical experiments as well as obtaining new constraints on higher, unobserved redshifts (Bull et al. 2015). The forthcoming intensity maps will also set new constraints on non-Gaussianity through measuring the ultra-large scales of the power spectrum (Camera et al. 2014). For all mentioned experiments, the cross-correlation of the H I intensity mapping signal with galaxy surveys will be a crucial test for systematics, and most likely be the first observable to deliver new scientific results.

In addition to cosmological parameters, the amplitude and the clustering power of the H I intensity mapping power spectrum depends on the distribution of the neutral hydrogen gas with respect to the underlying matter field, and additionally for cross-correlations on the observed optical galaxy sample. Recently, H I models based on available data (see e.g. Padmanabhan 2018; Padmanabhan, Refregier & Amara 2017), predictive theories (Chen 2012), as well as hydrodynamical simulations (Villaescusa-Navarro et al. 2018) have been proposed to deliver the theoretical framework for interpretation of the intensity mapping signal.

In this work, we extend existing H I models to predict the cross-correlation of intensity maps with galaxy surveys to enhance the interpretation of existing and forthcoming data, and provide a framework to include halo occupation parameters into the cosmological analysis of future measurements (for forecasts, see Pourtsidou, Bacon & Crittenden 2015; Sarkar et al. 2016; Pourtsidou, Bacon & Crittenden 2017). In Wolz, Blake & Wyithe (2017b), we have shown that the shot noise in the cross-power spectrum, which is caused by the discrete nature of galaxy data, scales with the average H I mass per optical galaxy. Hence, intensity mapping data can be employed to determine an average H I mass for any overlapping galaxy sample. This allows determination of global scaling relations between star formation activity as traced by the optical sample and their gas contents, for redshifts well beyond the current limits for direct gas detection. In this work, we present a theoretical framework which correctly determines the shot noise contribution given the H I parameters of the distribution, and which can be employed to fit the H I parameters and shot noise in future observational data.

In this paper, we first briefly introduce the halo model framework, along with our chosen numerical implementation (HALOMOD) in Section 2. Here, we also introduce the employed halo occupation models for galaxies and H I models and present theoretical equations for the H I autopower spectra and their respective cross-power spectra. In Section 3, we describe our method of producing lognormal realizations of joint optical and H I samples, which we will use to verify our theoretical formalism. In Section 4, we review the current understanding of shot noise on power spectra and discuss its implementation in HALOMOD. We present and examine the comparison of theory with lognormal simulations in Section 5. In the following Section 6, we demonstrate how HALOMOD can be used to constrain H I parameters via MCMC parameter estimation. We discuss our findings and present the conclusions in Section 7.

2 HALO MODEL DESCRIPTION

The halo model (Peacock & Smith 2000; Cooray & Sheth 2002) is a highly successful description of the cosmological density field that uses empirical models of the internal properties of dark matter halos to access nonlinear scales. It has been employed, along with a prescription for the abundance of galaxy tracers within halos termed the *halo occupation distribution* (HOD), to predict the spatial statistics of various galaxy populations, typically in order to constrain various properties of the selected sample (Zheng et al. 2005; Zehavi et al. 2011; Beutler et al. 2013). It has recently been extended to the domain of H I abundance by Padmanabhan & Refregier (2017); Padmanabhan et al. (2017).

The halo model is based on the assumption that all material is sequestered into discrete halos, which are, in turn, self-similar objects that scale exclusively as a function of their mass. Consequently, knowledge of the spatial arrangement of the halo centres combined with a knowledge of their internal profiles, how these scale with the halo's mass, and the abundance of halos at any given mass, yields a full statistical description of the matter field down to arbitrarily small scales in real space. Likewise, assuming that any given tracer inhabits halos with an abundance exclusively as a function of their mass, the statistics of the tracer field may also be determined.

In summary, to describe the two-point statistics of a tracer field (or the cross-correlation of tracers), one requires the following ingredients:

(i) The nonlinear matter power spectrum (Smith et al. 2003).

(ii) The radial profile of the tracer within the halo, $\rho(r)$; we typically employ the standard NFW profile (Navarro, Frenk & White 1997), but also check the modified, or 'cored" NFW employed by Padmanabhan & Refregier (2017).

(iii) The mass function of halos, $n(m)$; we use the fitting formula of Tinker et al. (2008).

(iv) The abundance and distribution of tracers within halos, $N(m)$; we describe our choices for this component further in Section 2.2.

(v) The concentration–mass relation, $c(m)$, which defines how the profile scales with halo mass; we use the fit of Duffy et al. (2008).

(vi) The bias of halos of a given mass, $b(m)$; we use the function determined by Tinker et al. (2010).

Additionally, the effects of halo exclusion can be modelled, such that pairs of the tracer that are very close are probabilistically assigned to the same halo and excluded from the counts between different halos, to avoid double-counting. We omit this modelling for this introductory work, but note that its inclusion is trivial within the HALOMOD package that we use.

2.1 The HALOMOD package

All halo model calculations performed in this work use the HALOMOD PYTHON library[1] (Murray et al., *in prep.*). This library is built on the HMF package[2] (Murray, Power & Robotham 2013), which handles the cosmology, linear power spectra, and mass functions. The HALOMOD code provides many models for halo profiles, halo bias, concentration–mass relations, HODs, and halo exclusion, along with the necessary framework to combine these to produce spatial statistics.

A key feature of the HMF framework which is extended to HALOMOD is the simplicity of defining new component models and ‘plugging” them into the calculations. Thus, for instance, it is simple to define a new HOD model from a standard galaxy HOD, and is instantly usable within the framework without having to modify the source code. We note that versions of both HMF and HALOMOD that calculate the results of this paper can be obtained via the `feature/HIHOD` branch of each.

2.2 Halo occupation distributions

In our study, we require several HOD models: one which describes the full galaxy count population, another which describes a particular optically selected sample count, and a model which describes the H I occupation. We use variants of the simple HOD parametrization of Zehavi et al. (2005) (Z05) in all cases. This model depends on three parameters: the minimum halo mass to be occupied by a galaxy $M_{\rm min}$, the characteristic halo mass M_1 which marks the turnover of the broken power law, and the power-law coefficient α of the satellite HOD. We extend the parametrization by adding the maximum (cut-off) halo mass $M_{\rm max}$ as a parameter.

In general, the HOD can be split into two separate classes of objects; central galaxies $N_{\rm cen}$ located at the centre of the halo, and satellite galaxies $N_{\rm sat}$ that trace the halo's density profile. The Z05 model assigns the following parametrizations to each component:

$$\langle N_{\rm cen}(m)\rangle = \begin{cases} 1, & M_{\rm min} < m < M_{\rm max} \\ 0, & \text{otherwise.} \end{cases} \qquad (1)$$

$$\langle N_{\rm sat}(m)\rangle = \begin{cases} (m/M_1)^\alpha, & M_{\rm min} < m < M_{\rm max} \\ 0, & \text{otherwise.} \end{cases} \qquad (2)$$

When stating that a sample may be described by a separation of central galaxy and satellite galaxies, we furthermore assume (in this paper) that this separation is due to the central having a much higher probability of existence within the sample than its associated satellites. This may be understood easily in terms of optical samples, in which the central galaxy is typically much brighter than the satellites. To approximate the effect of this a priori knowledge, our HALOMOD algorithms assert in such cases that a central galaxy *must* be present before any satellites. In this case, the average total occupation is accurately given by the following definition, which ensures that the total occupation is zero whenever the central occupation is zero, but otherwise yields the expected sum of central and satellite:

Table 1. HOD parameters for all models considered in this work. All masses are given as $\log_{10}$ and in units of $M_\odot\,h^{-1}$.

Model	$M_{\rm min}$	$M_{\rm max}$	α	M_1	$\log A_{\rm H\,I}$
Galaxy *field*	11.0	17.0	0.5	11.0	-
Galaxy *sample*	11.5	17.0	0.45	11.0	-
H I continuous	11.0	17.0	0.7	11.0	11.0
H I discrete	11.0	17.0	0.7	11.0	11.0

$$\langle N(m)\rangle = \langle N_{\rm cen}(m)\rangle(1 + \langle N_{\rm sat}(m)\rangle). \qquad (3)$$

HALOMOD does not limit the form or parametrization of the HODs and more complex models can be assumed.

In this paper, we adopt two fiducial galaxy HODs, in this study, referred to as *sample* and *field*, where we assume that the galaxy *field* model is a description of all optically detectable and H I emitting galaxies and *sample* is an optically detected subsample of the *field*. The HOD parameters of all galaxy and H I models can be found in Table 1, where we choose representative values for all parameters in our toy models.

In the following, we postulate two variations of the H I HOD model in the framework of the halo model. For demonstration purposes, we base the parametrization of the H I HODs on the parametrization established for optical galaxies, the Z05 model. More physically motivated and data-driven H I models have been suggested in the literature, see Padmanabhan & Refregier (2017), Paul, Choudhury & Paranjape (2018), and Villaescusa-Navarro et al. (2018). We refrain from choosing a specific parametrization as H I evolution is to-date poorly constrained by data, and the choice of the H I HOD does not limit the validity of our study. Most suggested models are parametrized by a similar amount of parameters (4)–(6) and can be easily implemented in HALOMOD and their quantitative predictions can be studied with our methods.

Continuous H I distribution. In this scenario, we assume that the H I continuously traces the dark matter halo following an independent density profile, for example a cored NFW profile as in Padmanabhan et al. (2017). This implies that the H I is not associated with galaxies and there are no central or satellite contributions to the density. This model is best suited to describe the cold gas distribution at the early stages of galaxy formation at the end of the epoch of reionization and resembles seminumerical approaches to cosmological simulations of intensity maps (e.g. Alonso, Ferreira & Santos 2014).

We alter the Z05 HOD to describe the H I mass distribution, by adding an extra normalization $A_{\rm H\,I}$ in units of $M_\odot\,h^{-1}$ to scale the distribution to produce typical H I masses, increasing the number of parameters to five.

$$\langle M_{\rm H\,I}(m)\rangle = \begin{cases} A_{\rm H\,I}\left((m/M_1)^\alpha + 1\right), & M_{\rm min} < m < M_{\rm max} \\ 0, & \text{otherwise.} \end{cases} \qquad (4)$$

The additional term +1 in the HOD is introduced to simplify comparison with our second H I model.

Discrete H I distribution. In this model, we assume that the H I is *on average* following the underlying dark matter halo density profile throughout the halo, but specify that the H I mass is co-located

[1] Source code at https://github.com/steven-murray/halomod.
[2] Available at https://github.com/steven-murray/hmf

with the underlying galaxy *field*. Thus, the H I in any given halo is discretely located. This model describes a stage of galaxy evolution in which most H I is confined within galaxies and intergalactic cold gas is negligible in the intensity maps. The approach predicts a similar distribution to semi-analytic simulations which model the cold gas abundances within star-forming regions (e.g. Lagos et al. 2014; Kim et al. 2017).

We model this case in a similar fashion to galaxies, in which we split the HOD contributions into central and satellite components.

$$\langle M_{\rm H\,I}^{\rm cen}(m)\rangle = \begin{cases} A_{\rm H\,I}, & M_{\rm min} < m < M_{\rm max} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

and the satellite part by

$$\langle M_{\rm H\,I}^{\rm sat}(m)\rangle = \begin{cases} A_{\rm H\,I}\ [(m/M_1)^{\alpha}\ ], & M_{\rm min} < m < M_{\rm max} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

We note that this H I model has a dependence on the model defining the underlying galaxy *field* with which it is colocated. While the HOD itself as defined above requires no knowledge of the underlying *field* HOD, and thus the autopower spectrum of the H I is fully self-defined, its cross-correlation with an optical sample relies on the actual distribution of H I within the halo, which we have described as being dependent on the *field*. We will see that this information will be necessary in order to define theoretical cross-correlations, Poisson noise, and also to self-consistently produce joint simulations.

2.3 Galaxy power spectra

The power spectrum is divided into its 2-halo and 1-halo contribution, $P(k) = P_{2h}(k) + P_{1h}(k)$. The 2-halo term closely follows the linear matter power spectrum $P_{\rm lin}(k)$ and, in its most general form applicable to cross-correlation on large scales, the 2-halo term is expressed as

$$P_{2h}^{ij}(k) = b_i(k) b_j(k) * P_{\rm lin}(k), \tag{7}$$

where b_i is the effective bias of the ith probe, given as

$$b_i(k) = \frac{1}{\bar{n}_g}\int {\rm d}m\ n(m)\ b(m) \langle N^i(m)\rangle\ u_i(k|m). \tag{8}$$

Here, $b(m)$ is the halo bias, $u(k|m)$ is the Fourier transform of the halo mass profile, with mass m following the NFW model, and $\bar{n}_g$ is given by the number density of the galaxies, computed as

$$\bar{n}_g = \int {\rm d}m\ n(m) \langle N(m)\rangle, \tag{9}$$

where $n(m)$ is the halo mass function. For more details on the implementation, please refer to Murray et al. (2013).

The 1-halo term is given by the clustering within the halos and depends on the number of central and satellite galaxies. For the autocorrelation of one probe, this results in

$$P_{1h}(k) = \frac{1}{\bar{n}_g^2}\int {\rm d}m\ n(m) \left[\langle N_{\rm cen} N_{\rm sat}\rangle\ u(k|m) + \frac{1}{2}\langle N_{\rm sat}(N_{\rm sat}-1)\rangle\ u^2(k|m) \right]. \tag{10}$$

The first term depends on the expectation value of the number of central–satellite pairs per halo multiplied by the halo mass profile and the second term on the expectation value of the number of satellite–satellite pairs per halo mass multiplied by the self-convolved mass profile. Since in our model there can only ever be either zero or one central galaxy in a halo, and under the assumption that the central galaxy is always the first of the halo to be included in a sample, we have $\langle N_{\rm cen} N_{\rm sat}\rangle = \langle N_{\rm cen}\rangle\langle N_{\rm sat}\rangle$. Furthermore, for Poisson-distributed X, $\langle X(X-1)\rangle \equiv \langle X\rangle^2$, which means (assuming the satellite occupation is Poisson-distributed) that

$$P_{1h}(k) = \frac{1}{\bar{n}_g^2}\int {\rm d}m\ n(m) \Big[\langle N_{\rm cen}\rangle\langle N_{\rm sat}\rangle\ u(k|m) + \frac{1}{2}\langle N_{\rm sat}\rangle^2\ u^2(k|m) \Big]. \tag{11}$$

This form is convenient, as it only depends on the mean occupation functions which we have defined above.

For the cross-correlation of two different galaxy samples which follow different HODs and density profiles, the analogue of equation (10) is

$$P_{1h}^{ij}(k) = \frac{1}{\bar{n}_i \bar{n}_j}\int {\rm d}m\ n(m)\ \Big[\left\langle N_{\rm cen}^i N_{\rm sat}^j \right\rangle\ u_j(k|m) + \left\langle N_{\rm cen}^j N_{\rm sat}^i \right\rangle\ u_i(k|m) + \left\langle N_{\rm sat}^i N_{\rm sat}^j \right\rangle\ u_i(k|m)\, u_j(k|m) \Big]. \tag{12}$$

In general, we cannot further reduce this equation, because it is not guaranteed that the absence of a central galaxy in one sample necessitates the absence of satellites (as well as central) in a different sample. However, if the central HOD happens to be a step-function, so that at any mass either all or none of the haloes have centrals, the central–satellite term decomposes as before. We note that this is an *extra condition*, which was not required for equation (11). This allows us to re-write the equation as follows:

$$P_{1h}^{ij}(k) = \frac{1}{\bar{n}_i \bar{n}_j}\int {\rm d}m\ n(m) \Big[\left\langle N_{\rm cen}^i \right\rangle \left\langle N_{\rm sat}^j \right\rangle\ u_j(k|m) + \left\langle N_{\rm cen}^j \right\rangle \left\langle N_{\rm sat}^i \right\rangle\ u_i(k|m) + \left(\left\langle N_{\rm sat}^i \right\rangle \left\langle N_{\rm sat}^j \right\rangle + \sigma_i \sigma_j R^{ij} - Q \right)\ u_i(k|m)\, u_j(k|m) \Big], \tag{13}$$

where R^{ij} is the correlation of the satellite occupation between the probes, and σ_i the standard deviation of the satellite occupation, which for a Poisson occupation is simply $\sqrt{\langle N_{\rm sat}\rangle}$. Q is equal to the expected number of shared points between the samples unless either tracer is continuously spatially distributed which results in $Q = 0$.

In general, R^{ij} is constrained to be within (− 1, 1) and depends on the complicated physical interactions of the two tracer populations. However, for the toy models that we employ in this paper, it is possible to provide a better description which we present in detail in Appendices A1 and A2.

2.4 H I power spectra

Following the same arguments as in the previous section, we may derive the power spectrum of H I density fluctuations for both cases presented in Section 2.2. The 2-halo term of the H I power spectra for both models is similar to equations (7) and (8) with the galaxy HOD substituted by the H I occupation $\langle M_{\rm H\,I}(m)\rangle$ of the respective model, such that

$$b_{\rm H\,I}(k) = C_{\rm H\,I}\int {\rm d}m\ n(m)\ b(m) \langle M_{\rm H\,I}(m)\rangle\ u_{\rm H\,I}(k|m), \tag{14}$$

where the coefficient $C_{\rm H\,I}$ is described below. The H I halo density profile $u_{\rm H\,I}(k|m)$ is commonly defined as a modified (or cored) NFW profile (Padmanabhan et al. 2017) which in real space reads as

$$\rho_{\mathrm{H\,I}}(r) = \frac{\rho_0 r_s^3}{(r + 0.75 r_s)(r + r_s)^2}, \tag{15}$$

where r_s is the scale radius of the dark matter halo which is defined as $r_s = r_{\mathrm{vir}}/c(m)$ and r_{vir} is the virial radius of the halo. We refrain from the use of a H I-specific parametrization of r_s and adopt the concentration–mass relation fit from Duffy et al. (2008). In HALOMOD, we employ the analytic expression of the Fourier transform $u_{\mathrm{H\,I}}(k)$ of this profile (Padmanabhan et al. 2017).

H I intensity maps are measured in brightness temperature $T_{\mathrm{H\,I}}$. To follow this convention, we convert all power spectra into temperature units, using a conversion $C_{\mathrm{H\,I}}$, given by

$$C_{\mathrm{H\,I}} = \frac{3 A_{12} h_{\mathrm{P}} c^3 (1+z)^2}{32 \pi m_{\mathrm{H}} k_{\mathrm{B}} \nu_{21}^2 H(z)} \tag{16}$$

with h_{P} being the Planck's constant, k_{B} the Boltzmann constant, m_{H} the mass of the hydrogen atom, A_{12} the emission coefficient of the 21 cm line transmission, and ν_{21} the rest frequency of the 21 cm emission. $H(z)$ is the Hubble parameter at redshift z. All presented studies are for redshift $z \approx 0$. The plotted H I power spectra are given in units of $\mathrm{K}^2(\mathrm{Mpc/h})^3$ and cross-power spectra as $\mathrm{K}(\mathrm{Mpc}/h)^3$ if not stated otherwise.

The predicted mean brightness temperature for each H I model can be determined via

$$\overline{T_{\mathrm{H\,I}}} = C_{\mathrm{H\,I}} \int \mathrm{d}m\, n(m) \langle M_{\mathrm{H\,I}}(m) \rangle. \tag{17}$$

The mean H I brightness temperature is directly proportional to the H I energy density $\Omega_{\mathrm{H\,I}}$ which makes it a desired observable when conducting H I intensity mapping experiments.

Continuous H I distribution. The 1-halo term of the autopower spectrum in this case, with lack of satellite components, can be written as

$$P_{1h}^{\mathrm{H\,I,cont}}(k) = C_{\mathrm{H\,I}}^2 \int \mathrm{d}m\, n(m) \langle M_{\mathrm{H\,I}}(m) \rangle^2\, u_{\mathrm{H\,I}}(k|m)^2, \tag{18}$$

while the cross-correlation with a galaxy sample g is

$$\begin{aligned} P_{1h}^{g\mathrm{H\,I,cont}}(k) = \frac{C_{\mathrm{H\,I}}}{\bar{n}_g} \int &\mathrm{d}m\, n(m) u_{\mathrm{H\,I}}(k|m) \\ &\times \left[(u_g(k|m)\, \langle N_{\mathrm{sat}}^g(m) \rangle \langle M_{\mathrm{H\,I}}(m) \rangle + R^{g\mathrm{H\,I}}) \right. \\ &\left. + \langle N_{\mathrm{cen}}^g(m) \rangle \langle M_{\mathrm{H\,I}}(m) \rangle \right], \end{aligned} \tag{19}$$

where $R^{g\mathrm{H\,I}}$ is a galaxy–H I correlation coefficient. As there is no central–satellite split in the H I HOD, the clustering is simplified into two terms – one in which the satellite galaxies pair with the H I profile, and another in which the single (possible) central galaxy pairs with the H I profile. We fiducially consider a value of $R = 0$ for this work, which implies that the H I mass is uncorrelated with the galaxy occupation. The more detailed derivation of the correlation factor R and an example for a correlated toy model can be found in Appendix A1.

Discrete H I distribution. The 1-halo power spectrum of the discrete H I model can be written similarly to equation (11), assuming that the positions of the satellite occupation are Poisson-distributed:

$$\begin{aligned} P_{1h}^{\mathrm{H\,I,dsc}}(k) = C_{\mathrm{H\,I}}^2 \int \mathrm{d}m\, n(m) \Big[&u_{\mathrm{H\,I}}(k|m) \langle M_{\mathrm{H\,I}}^{\mathrm{sat}}(m) \rangle \langle M_{\mathrm{H\,I}}^{\mathrm{cen}}(m) \rangle \\ &+ \frac{1}{2} \langle M_{\mathrm{H\,I}}^{\mathrm{sat}}(m) \rangle^2 u_{\mathrm{H\,I}}(k|m)^2 \Big]. \end{aligned} \tag{20}$$

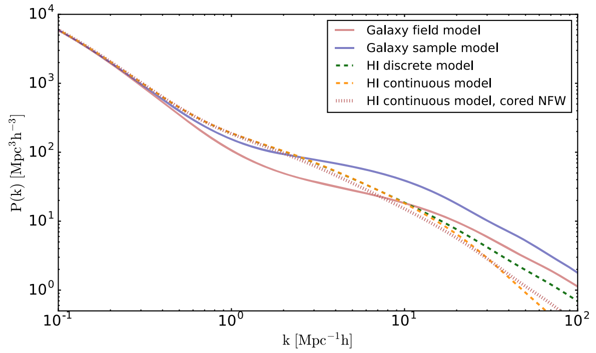

Figure 1. The autopower spectra predicted by our model for the case of galaxy *field* population, galaxy *sample*, H I continuum, and H I discrete model. The H I power spectra are normalized by the square of the mean temperature predicted by each model using equation (17) for presentation purposes. Note that, by construction, both H I models predict the same mean brightness temperature.

The 1-halo term of the H I cross-correlation with a galaxy sample reads as

$$\begin{aligned} P_{1h}^{g\mathrm{H\,I,dsc}}(k) = \frac{C_{\mathrm{H\,I}}}{\bar{n}_g} \int \mathrm{d}m\, n(m) \big[&\left(\langle N_{\mathrm{cen}}^g \rangle \langle M_{\mathrm{H\,I}}^{\mathrm{sat}} \rangle \right) u_{\mathrm{H\,I}}(k|m) + \\ &\left(\langle M_{\mathrm{H\,I}}^{\mathrm{cen}} \rangle \langle N_{\mathrm{sat}}^g \rangle \right) u_g(k|m) + \\ &\left(\langle N_{\mathrm{sat}}^g \rangle \langle M_{\mathrm{H\,I}}^{\mathrm{sat}} \rangle \right) u_g(k|m)\, u_{\mathrm{H\,I}}(k|m) \big]. \end{aligned} \tag{21}$$

We note the absence of the correlation term, R. This is due to exact colocation of the H I with the optical galaxies, as explained in detail in Appendix A2. Briefly, in this model, H I abundance depends only on the properties of the galaxy in which it is situated, and this galaxy, by construction, has no correlation with other galaxies. Therefore, all correlations are expressed at a separation of zero, and do not affect the shape of the 1-halo term. This may alternatively be seen as the exact cancellation of the correlation term with the Q term in equation (13). The autopower spectrum predictions of HALOMOD are shown in Fig. 1, where we show the two models of galaxy power spectra, called *field* and *sample*, with HOD parameters defined in Table 1. In yellow and green, we compare the H I power spectra of the continuous and discrete models, where we renormalize the H I spectra through division by $\overline{T_{\mathrm{H\,I}}}^2$. We can see that for $k < 10\,\mathrm{h\,Mpc^{-1}}$, both models' predictions closely agree. For $k > 10\,\mathrm{h\,Mpc^{-1}}$, the continuous H I model falls off more quickly since the 1-halo term contains no central–satellite contribution in the discrete case. In this figure, we also demonstrate how the cored H I profile alters the power of the 1-halo term in comparison to the standard NFW profile. In the remainder of the paper, we employ the standard NFW profile for our computations such that the comparison of the cases are focused on their clustering terms rather than on the impact of the density profile.

The cross-power spectrum prediction of both H I models with the two galaxy models are shown in Fig. 2. The differences in the two H I models are negligible over all scales k. Furthermore, even the differences in the two different galaxy models are very small compared to the variation in their autopower spectrum. The agreement of the two models is by construction as they follow the same H I HOD parameters and we implement the continuous case to be the sum of the central and satellite terms of the discrete model. Therefore, they correlate in a similar fashion with the galaxy samples which is not to be expected in general.

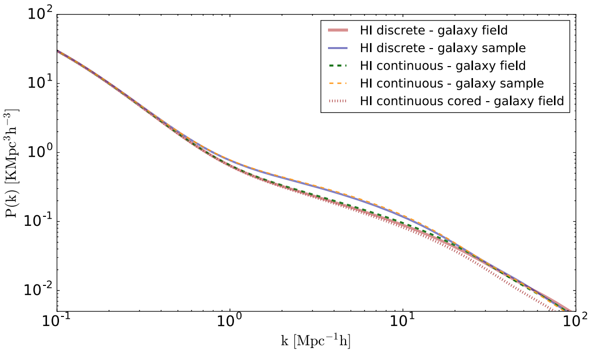

Figure 2. The cross-power spectra predicted by our model for the case of galaxy field population, galaxy sample, H I continuum, and H I discrete model. The H I cross-power spectra are normalized by the mean temperature predicted by each model using equation (17) for presentation purposes. Note that, by construction, both H I models predict the same mean brightness temperature.

In both figures, we neglect the shot noise, also referred to as the Poisson Noise (PN) contribution $P_{\rm PN}$, and we will discuss its contribution in detail in the following sections.

3 LOGNORMAL SIMULATIONS

In order to test the accuracy of the analytic routines within HALOMOD, we create a number of mock realizations of the tracer populations. As this is done explicitly to test the routines, the simulations are prepared to mimic the assumptions of the halo model formalism at the simplest level. In this section, we describe the method used to generate these simulations.

3.1 Galaxy populations

We consider a cube of volume L^3(Mpc/h)^3 with N^3 grid cells, in which we generate a lognormal density field (Coles & Jones 1991) using the POWERBOX package.[3] We choose matching input power spectra and parameters to HALOMOD to ensure comparability of the results using a flat Planck15 cosmological model (Planck Collaboration et al. 2016) with $\Omega_m = 0.307$, $\Omega_b = 0.0486$, and $H_0 = 67.74$km/(Mpc s).

We choose a minimum halo mass $M_{\rm min,\,h}$ such that all halos containing galaxies in our sample lie above the threshold. We then draw a number density of halo masses $n_h = \int_{M_{\rm min,h}} n(m){\rm d}m$ from the halo mass function distribution. These halos are placed probabilistically within the grid volume, with the probability of landing in a certain cell given by its relative density. The final positions of each halo are drawn randomly within each cell, rendering subgrid scales highly inaccurate. We note that the mass of each halo does not affect its placement, which effectively means that the halo bias is unity for all masses. When comparing simulations to theory, we therefore set the theoretical halo bias to unity in HALOMOD.

Finally, we use the resultant halo catalogue, with masses and positions, as the scaffolding on which to assign the tracer population. Here, we will describe the methods used for producing a single tracer population, suitable for comparing with autospectra. We use a routine in which for each halo i we perform the following steps:

(i) Sample a single number (zero or one) C_i from a Bernoulli distribution with mean $\langle N_{\rm cen}(m_i)\rangle$

(ii) If $C_i = 1$, place a galaxy at $\vec{x}_i$ and continue, else proceed to next halo.

(iii) Sample a number $N^i_{\rm sat}$ from a Poisson distribution with mean $\langle N_s(m_i)\rangle$.

(iv) If $N^i_s > 0$, sample $N^i_{\rm sat}$ radii, r^i_j from the halo's profile, $\rho(r, m)$, and sample (θ_j, ϕ_j) isotropically to yield 3D co-ordinates, $\vec{x}^i_j$ centred at the origin.

(v) Assign $N^i_{\rm sat}$ galaxies to positions $\vec{x}_i + \vec{x}^i_j$.

We note that this procedure does not take into account halo exclusion – halos are allowed to overlap arbitrarily – and thus to reproduce the results analytically also requires no halo exclusion model.

In our simulations, we first apply the steps outlined above using the *field* HOD to create a galaxy catalogue which is assumed to contain all available galaxies. We then create a subsample of the galaxy catalogue which follows the HOD of the galaxy *sample*. Similarly for steps (*i*) and (*ii*), for each galaxy in the *field* catalogue we draw a single number $C_i \in (0, 1)$ from the Bernoulli distribution with mean $P_i = \langle N_{\rm sample}(m_i)\rangle/\langle N_{\rm field}(m_i)\rangle$ to determine if the galaxy is part of the *sample*. The positions of the galaxies are kept identical. We note that this procedure does not strictly retain the Poisson-distributed nature of the satellite galaxies in the *sample*. Nevertheless, the mean is retained, and we do not expect the departure from Poisson statistics to be significant.

3.2 H I populations

Continuous H I distribution. For the continuous model, we assign an H I mass to each halo produced by the lognormal realizations, where we draw the H I masses according to the input H I HOD at halo mass m assuming a Gaussian distribution with a standard deviation $\sigma_{\rm H\,I} = 0.25\langle M_{\rm H\,I}(m)\rangle$. In order to mimic the continuous H I distribution throughout the halo, we convolve the resulting H I mass with a density profile using the following method. We note that any arbitrary density profile independent of the underlying halo density profile can be used in this routine.

According to the convolution theorem, the convolution of the H I masses with any given profile is a multiplication in Fourier space, which is more computationally efficient. However, generally, the halo profile is a function of halo mass m_i. In order to reduce computation, we apply a projection algorithm for the convolution. The H I masses are therefore binned according to their halo mass into $N_{\rm bin}$ bins. We create $N_{\rm bin}$ cubes with each H I mass located at their respective halo centre position. Each cube is Fourier-transformed and multiplied by the Fourier-transformed profile of the mean halo mass $\overline{m}_i$ of the respective halo mass bin. We then sum all cubes to create the final intensity mapping cube. For the case of the NFW profile, the algorithm converges for $N_{\rm bin} = 25$.

We note that the continuous H I distribution is based on the same underlying halo distribution of the lognormal realization, but is independent of the *field* or *sample* galaxy densities and satellite positions.

Discrete H I distribution. In the discrete model, the H I HOD is associated with an underlying galaxy *field* HOD $\langle N_{\rm field}\rangle$ which describes the distribution of all H I emitting objects. In our algorithm, the galaxy *field* is drawn from $\langle N_{\rm field}\rangle$ as described in Section 3.1. We then assign the H I mass of each galaxy from a Gaussian distribution with mean $\langle M_{\rm H\,I,field}\rangle = \langle M_{\rm H\,I}(m_i)\rangle/\langle N_{\rm field}(m_i)\rangle$ and standard deviation $\sigma_{\rm H\,I} = 0.25\langle M_{\rm H\,I,field}(m_i)\rangle$ for satellites and centrals, respec-

[3] Available at https://github.com/steven-murray/powerbox.

tively. The assumption that $\langle M_{\rm H\,I,field}\rangle = \langle M_{\rm H\,I}(m_i)\rangle / \langle N_{\rm field}(m_i)\rangle$ is only true if the probability of selecting a galaxy is independent of H I mass, which precludes the use of this algorithm for creating correlated samples. This model allows for the galaxy *field* HOD and the H I HOD to follow independent models and parametrizations within the limitation that $M_{\rm min}$ and $M_{\rm max}$ of the H I sample cannot be outside the defined galaxy mass range. In our study, we choose $M_{\rm min,\,HI} = M_{\rm min,\,field}$ and $M_{\rm max,\,HI} = M_{\rm max,\,field}$ for simplicity. Additionally, the H I mass can be scaled by an independent H I density profile, similarly to the continuous case. In our study, we set the H I profile equal to the NFW profile of the underlying galaxy *field*.

3.3 Correlated populations

The procedures described above produce galaxy catalogues and H I intensity maps useful for determining their 1-halo clustering and Poisson noise. It is nontrivial to populate a physically motivated model for correlated galaxy–H I samples in the framework of HALOMOD. Commonly, the H I mass of galaxies is associated with their star-formation activity and other more complex mechanisms depending on the galaxy's evolutionary state, which is beyond the scope of our work.

As previously stated, in the continuous H I case, the correlation factor R is determined through the dependence of the galaxy numbers on the H I mass per halo, or vice versa (a basic example of this is set out in Appendix A1), and this impacts the 1-halo contribution of the cross-power spectrum.

For the discrete H I case, we demonstrated that a correlation between the H I distribution and the galaxy abundances has no impact on the 1-halo term. However, if H I masses and the galaxy abundances are correlated, the averaged H I mass per galaxy over the *sample* is modulated and hence the amplitude of the cross-shot noise is changed, as we detail in the following section. As one of our primary concerns is investigating the cross-shot noise, we demonstrate this effect with correlated simulations through the following procedure.

In order to create a H I correlation, we either up- or downweight the H I masses of the galaxies in the *sample* by drawing for each galaxy*i* a Gaussian variable $\delta M_{i,\rm H\,I}$ with zero mean and multiplying the absolute value $|\delta M_{i,\rm H\,I}|$ with a weighting factor $w = \{+1, -1\}$. When assigning H I masses, we then add $w \times |\delta M_{i,\rm H\,I}|$ to the mean H I mass at m given by the H I HOD $\langle M_{\rm H\,I}(m)\rangle$. This implies that all galaxies in the *sample* either have higher or lower H I mass than the mean of the Gaussian. Galaxies which are not part of the *sample* are not affected by the weighting and their H I masses fluctuate around the mean. This process slightly alters the measured brightness temperature of the H I intensity maps. However, if the galaxy *sample* is a small enough subsample of the whole galaxy *field*, this effect will be minor.

4 POISSON NOISE

4.1 Autopower spectra

The additive shot noise contribution to the power spectrum, also referred to as Poisson noise in the literature, is due to the finite number of data points used to probe a continuous field. In galaxy surveys, the shot noise is caused by the finite number of galaxies in the sample employed to trace the matter field. The resulting Poisson noise on the power spectrum is scale-independent with the amplitude equal to the inverse of the galaxy density:

$$P_{\rm PN} = \frac{1}{n_g} = \frac{1}{N_g/V}. \tag{22}$$

The total power measured from galaxy survey data is $P(k) = P_{2h}(k) + P_{1h}(k) + P_{\rm PN}$.

The shot noise of a galaxy distribution is not strictly Poissonian. Deviations from the Poisson limit were examined by e.g. (Hamaus, Seljak & Desjacques 2011; Baldauf et al. 2013; Paech et al. 2017). The deviations are caused by halo exclusion, nonlinear clustering on small scales, and satellite galaxy distributions, where the fraction of satellite galaxies can determine if the noise is sub- or super-Poissonian (Baldauf et al. 2013).

In the halo model context, the shot noise of the halo power spectrum may be determined by the $k \to 0$ limit of the 1-halo term of the power spectrum, which results in the Poisson limit. This approach is correct when treating tracers without subsampling the halo with satellite populations. For galaxy populations including a central / satellite split, the $k \to 0$ limit of the 1-halo term does not result in the Poisson limit and overestimates the shot noise. Ginzburg, Desjacques & Chan (2017) investigate the shot noise expression for dark matter, halos, and tracers in the halo model framework, considering galaxy populations with satellites. They derive correction terms to the 1-halo term to accurately determine the deviations from the Poissonian noise on scales $k \ll 1\,h\,{\rm Mpc}^{-1}$. For our scales of interest where the shot noise dominates the overall power for $k \gg 1\,h\,{\rm Mpc}^{-1}$, the shot noise must converge towards the Poisson limit of the 1-halo term neglecting the satellite correlations (Ginzburg et al. 2017). For the remainder of this study, we will only consider the Poisson limit of the shot noise and use the terms Poisson noise and shot noise interchangeably.

We derive Poisson limit of the shot noise as

$$P^{g}_{\rm PN} = \left(\int {\rm d}m\; n(m) \sum_{i={\rm sat,cen}} \langle N^{i}(m)\rangle \right)^{-1}. \tag{23}$$

In intensity mapping, the nature of the shot noise depends on the H I model used. In general, the shot noise is given by the standard deviation (or second moment) of the observed field (see Breysse et al. 2017and Kovetz et al. 2017), in this case the H I mass distribution is such that

$$P^{\rm H\,I}_{\rm PN} = C^{2}_{\rm H\,I} \int {\rm d}m\; n(m) \langle M_{\rm H\,I}(m)\rangle^{2}. \tag{24}$$

In the halo model context, this is equal to taking the $k \to 0$ limit of the 1-halo term neglecting the existence of satellite distributions, similar to equation (18), see also Castorina & Villaescusa-Navarro (2017) for a similar result.

In our specific case of the continuous model, the H I masses are sampled per halo, which means that the number of samples is equivalent to the number of halos. However, H I is not discretely populated, but convolved with the halo profile which results in a continuous map of the H I in voxel space. From a strict definition of Poisson noise originating from discrete sampling and resulting in a scale-independent noise, this means that the H I continuous power spectrum does not contain a Poisson noise contribution. The absence of H I shot noise in the continuous case is due to the strict smoothness of the H I distribution tracing the halo. Alternatively, one could think of the 1-halo term as the Poisson contribution which is convolved by the halo profile.

For our H I discrete model, we assume that the H I masses are sampled per galaxy, rather than halo, so we need to determine the second moment of the H I distribution *per galaxy* where the H I per galaxy is given as $\langle M^i_{\rm H\,I,field}(m)\rangle = \langle M^i_{\rm H\,I}(m)\rangle / \langle N^i_{\rm field}(m)\rangle$ with $i = \{{\rm cen, sat}\}$, (again we note that this is strictly only correct if H I and galaxy abundances are uncorrelated). The resulting Poisson noise of this model is

$$P_{\rm PN}^{\rm H\,I,dsc} = C_{\rm H\,I}^2 \int {\rm d}m\ n(m) \sum_{i={\rm sat,cen}} \langle M^i_{\rm H\,I,field}(m)\rangle^2 \langle N^i_{\rm field}\rangle. \qquad (25)$$

We do usually not know either the HOD of the underlying galaxy *field*, or the H I HOD, in order to determine the H I per galaxy as a function of halo mass. In practice, the Poisson noise can be modelled as a single additive number and fit to observations.

4.2 Cross-power spectra

The shot noise in the cross-power spectrum of two galaxy samples is determined by the galaxy density of the overlap of the two samples (see e.g. Smith 2009). If the two galaxy samples are mutually-exclusive, the amplitude of the Poisson noise in the power spectrum is zero.

As outlined in the previous paragraph, the continuous H I power spectrum does not contain a scale-independent Poisson noise contribution. Similarly, there is no Poisson noise generated in the cross-correlation of a continuum and a discrete galaxy sample, as the H I distribution is assumed to be completely smooth and hence no additional sampling noise can correlate with the sampling noise of the galaxies. Again, alternatively, one could think of the sampling noise being incorporated in the 1-halo term as the Poisson contribution is convolved by the smooth H I profile.

The discrete H I model can be approached similarly to the case of two galaxy samples where shot noise is determined by the cross-section. In intensity mapping, it is assumed that each object emits H I and contributes to the H I maps. The cross-section of the H I maps and the *sample* is hence the number density of the *sample* and the Poisson noise is inversely proportional to the galaxy number density. The H I contribution to the Poisson noise is determined by the average H I emission of the galaxies in the *sample*. This general expression for the cross-shot noise can also be derived considering the $k \rightarrow 0$ limit of the H I-galaxy 1-halo term in absence of satellite populations. We derive the Poisson noise of the cross-correlation of the discrete case as

$$\begin{aligned} P_{\rm PN}^{\rm gH\,I,dsc} =& C_{\rm H\,I}\left(\int {\rm d}m\ n(m) \sum_{i={\rm sat,cen}} \langle M^i_{\rm H\,I,field}(m)\rangle \langle N^i_{\rm sample}\rangle\right) \\ & \times \left(\int {\rm d}m n(m)\langle N(m)\rangle\right)^{-1}. \end{aligned} \qquad (26)$$

This equation agrees with the derivation in Wolz et al. (2017b), where it was shown that the Poisson noise is directly proportional to the averaged H I mass per galaxy in the *sample*. This also implies that the amplitude of the Poisson noise is sensitive to any correlations between H I and the abundance of galaxies in the *sample*. In the following, we verify these expressions by comparing the HALOMOD predictions to simulations and showcase how the Poisson noise can be fit in order to determine the averaged H I masses of galaxy samples.

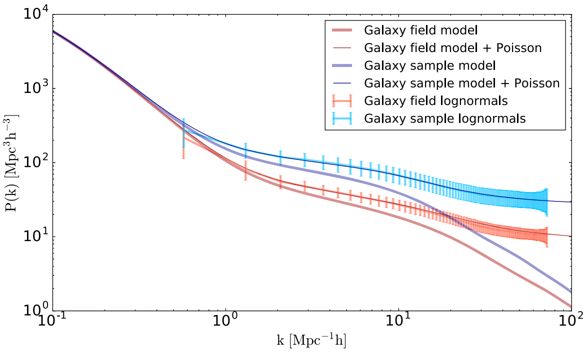

Figure 3. The galaxy power spectra predicted by HALOMOD for the entire galaxy field and the selected galaxy sample in comparison with an average power spectrum of 100 lognormal realizations with a box of length $15{\rm Mpc\,h^{-1}}$ drawn from the respective galaxy HOD. We show HALOMOD predictions including and excluding Poisson noise contribution.

5 COMPARISON OF HALOMOD WITH SIMULATIONS

5.1 Auto- and cross-power spectra

We run a suite of lognormal simulations with different box sizes with length $L \in \{15, 25, 50, 100\}{\rm Mpc\,h^{-1}}$ and $N = 200$ pixels per side to create a valid comparison for all relevant scales k. We find that the HALOMOD prediction agrees well with the lognormal simulations on all scales. In order to resolve the scales dominated by the H I and cross-Poisson noise, we closely inspect simulations with volume $V = (15{\rm Mpc/h})^3$, which are presented in the following figures. We simulate 100 realizations of each lognormal field and the error bars of the following plots are given by the standard deviation of these realizations.

In Fig. 3, we show the comparison of the autopower spectra of the lognormal galaxy population models with the associated HALOMOD prediction. The power spectrum measurements from the lognormal realizations naturally contain Poisson noise contributions which we add to the HALOMOD predictions using equation (23). We see that the HALOMOD prediction including the Poisson noise is in agreement with estimates from the lognormal simulations. In this plot, we show our two galaxy models, *field* and *sample* (cf. Table 1). The galaxy densities of the populations are predicted by HALOMOD as $n_{\rm sample} = 0.036({\rm h/Mpc})^3$ and $n_{\rm field} = 0.110({\rm h/Mpc})^3$ and estimated from the realizations as $n_{\rm sample} = 0.036 \pm 0.0069({\rm h/Mpc})^3$ and $n_{\rm field} = 0.107 \pm 0.017({\rm h/Mpc})^3$.

In Fig. 4, the H I power spectra of the lognormal realizations using the continuous and discrete model are shown in comparison to the analytic HALOMOD predictions in units of ${\rm K}^2\,({\rm Mpc/h})^3$. For the continuous case describing smooth H I distributions within halos independent of galaxy positions, we can see that the average of the simulations and the analytic prediction agree very well within the errors. In the discrete model, the H I distribution is colocated with the galaxy positions and hence this model includes a Poisson noise contribution as described in Section 4. We add the theoretical prediction of the Poisson noise using equation (25) to the predictions of HALOMOD. The combined amplitude is in agreement with the estimates of the lognormal distributions. Both H I models follow the same HOD parameter model, except the discrete model uses two additional parameters to describe the shape of the underlying galaxy field HOD ($\alpha_{\rm field}$ and $M_{1,{\rm field}}$). By construction, both H I models predict the same H I brightness

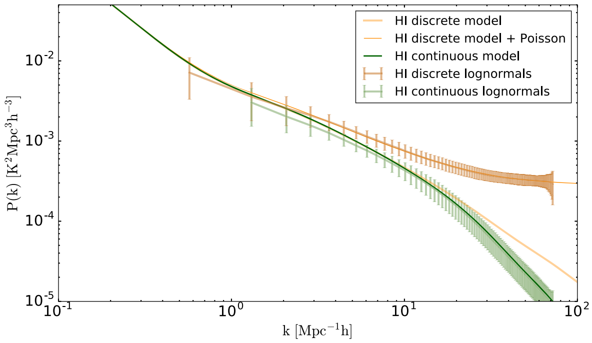

Figure 4. The H I power spectra predicted by HALOMOD for the H I continuous model and the H I discrete model in comparison with an average power spectrum of 100 lognormal realizations with a box of 15Mpc h^{-1} drawn from the respective galaxy HOD. Note that the H I continuous model does not include a scale-independent Poisson noise contribution since it is estimated from a continuum field. We show HALOMOD predictions including and excluding Poisson noise contribution.

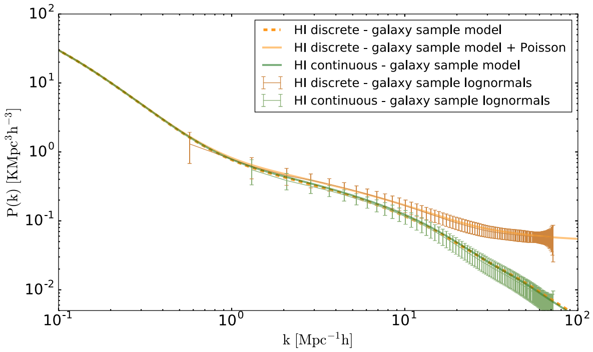

Figure 5. The cross-power spectra predicted by HALOMOD for the H I continuous model with the galaxy *sample*, and the H I discrete model with the galaxy *sample*, in comparison with a average power spectrum of 100 lognormal realizations of a box of length 15Mpc h^{-1} drawn from the respective H I and galaxy HOD. Note that the H I continuous model does not include a scale-independent Poisson noise contribution since it is estimated from a continuum field. We show HALOMOD predictions including and excluding Poisson noise contribution.

temperature $\overline{T_{\rm H\,I}} = 0.0050K$. The lognormal simulations of the continuous case produce $\overline{T_{\rm H\,I}} = 0.0049 \pm 0.0013K$ and in the discrete model produce $\overline{T_{\rm H\,I}} = 0.0048 \pm 0.001K$. The errors in these measurements increase with $\sigma_{\rm H\,I}$, the scatter with which the H I masses per object were drawn from the H I HOD.

The cross-power spectra of the galaxy sample with the two H I models are presented in Fig. 5. Even though the theory calculation of the two models does not predict any visible deviation on all considered scales, we observe that the inclusion of Poisson noise in the discrete model considerably increases the power in the range $k \gtrsim 2\,\mathrm{h\,Mpc}^{-1}$. The theoretical prediction of the cross-Poisson noise is added to HALOMOD using equation (26). As previously discussed, the cross-Poisson noise scales with the H I content of the galaxy population averaged over all halo masses, in this case for the galaxy *sample*. For this galaxy *sample*, we can measure an average H I mass of $\log_{10}(\overline{M}_{\rm H\,I,sample}/\mathrm{M}_{\odot}\mathrm{h}) = 11.202$ from the lognormal realizations, which is very close to the prediction of the theoretical model with $\log_{10}(\overline{M}_{\rm H\,I,sample}/\mathrm{M}_{\odot}\mathrm{h}) = 11.208$. We note that in the considered lognormal realizations with volume (15Mpc/h)^3, the mean number of galaxies in the *sample* is relatively low, with 121.

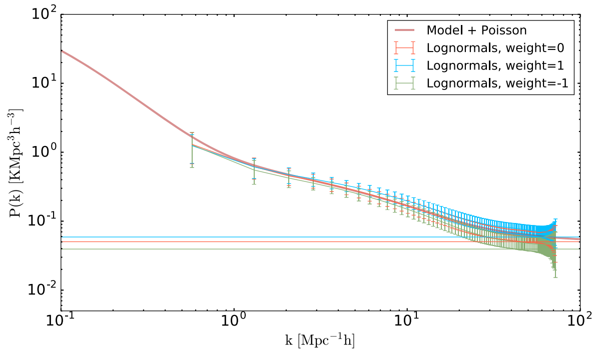

Figure 6. The cross-power spectra predicted by HALOMOD for the H I discrete model with the galaxy *sample* with different H I weighting, in comparison with a average power spectrum of 100 lognormal realizations with a box of length 15Mpc h^{-1} drawn from the respective H I and galaxy HOD. The dependence of the shape of the power spectra on the H I weighting is negligible but the amplitude of the Poisson noise changes significantly. We show HALOMOD predictions including and excluding Poisson noise contribution.

5.2 Correlated populations

In the above example, the distribution of the H I within the galaxy *field* for each halo mass *m* follows a Gaussian distribution with standard deviation $\sigma_{\rm H\,I}$. Thus, there is no dependence of the H I content on the galaxy occupation within the *sample*. In reality, the amount of H I present in the galaxy depends on its evolutionary state. In general terms, blue, star-forming galaxies are expected to be H I-rich whereas red, quiescent galaxies are H I-deficient. In this work, focusing on the concept of Poisson noise in intensity mapping, we do not concern ourselves with details such as luminosity functions which would be required to accurately model these dependencies.

In order to mimic the effect that a correlation between luminosity and H I mass would impose on the Poisson noise, we assume that the galaxy *sample* describes the HOD of a specific type of galaxy which is correlated or anticorrelated with the H I content as described in Section 3.3. This correlation, as predicted, has no effect on the H I autopower or the shape of the cross-power, but it changes the amplitude of the cross-Poisson noise as the averaged H I mass per galaxy in the *sample* is modified. Fig. 6 presents the result of the weighting of the H I for the galaxy *sample*. To demonstrate the change in the Poisson noise, we added the measured Poisson noise from the lognormal simulations with coloured horizontal lines to Fig. 6.

In general, a specific mathematical model of the correlation of two samples is not available, and so we cannot determine the Poisson noise amplitude *a priori*. However, the HALOMOD theory predictions can be used as a tool to fit the pure Poisson noise contribution as well as measure the deviation compared to an uncorrelated sample.

6 PARAMETER ESTIMATION

In this section, we demonstrate the utility of the HALOMOD algorithms to recover the parameters of a specific H I model via a Monte Carlo Markov Chain (MCMC) maximum likelihood fit. We use the Python package EMCEE (Foreman-Mackey et al. 2013) and fit the theory prediction of each H I model to the estimated power spectra of the lognormals with box size (15Mpc/*h*)^3, which optimally resolves the shot noise regime of the power spectra. We fit the averaged power spectra of 100 realizations, as well as 10 individual realizations. The variance of each power spectrum measurement is given by

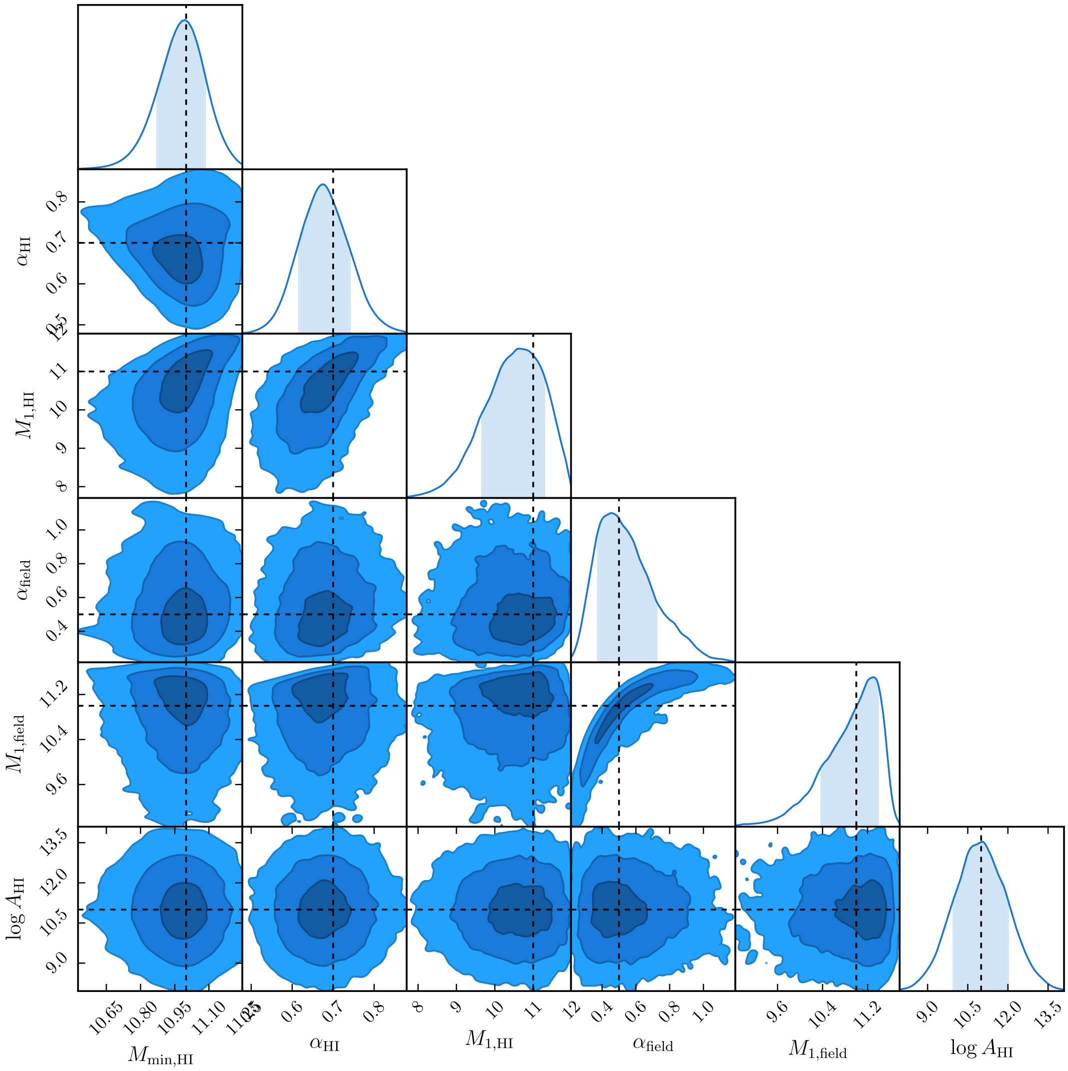

Figure 7. The likelihood contours of the MCMC fit to the realization-averaged power spectrum of the discrete H I model, using the H I autocorrelation and its Poisson noise to constrain the HOD parameters of HALOMOD. The dashed lines indicate the input parameter values. All masses are given as $\log_{10}$ and in units of $\rm M_{\odot}\,h^{-1}$.

the standard deviation of all realizations which includes Cosmic variance, fluctuations in the number densities due to the small box size and variations due to the population of H I masses. We note that due to the limited size of the box, which measures a minimum scale $k \approx 1.0\,\rm h\,Mpc^{-1}$, our parameter fitting is limited to the scales dominated by the shot noise. The overall fitting could be improved using a wider range of wavenumbers, however, some intensity mapping experiments such as the interferometer ASKAP are only sensitive to similar scales $k > 1.0\,\rm h\,Mpc^{-1}$. For all cases, we only fit H I parameters employing very tight priors for parameters of the galaxy *sample* distribution. We do not attempt to fit the maximum halo mass $M_{\rm max,H\,I}$ as for our tested box size, the abundance of these high mass halos is very low and the cut-off cannot be tested.

We ran MCMCs with a total of 10^6 samples and tested convergence of the chains via the Gelman–Rubin criteria where all parameters passed with a threshold of $R_{\rm GR} = 1.1$. We set Gaussian priors with the standard deviations of all HOD parameters α given as $\sigma(\alpha_i) = 0.3$ with $i = \{\rm H\,I, field\}$ and the standard deviation of all other HOD parameters as $\sigma(\theta_i) = 1.0$. We note that our results, in particular for the discrete case in which the model is primarily fit to the constant amplitude of the Poisson noise, are not independent of the chosen priors. In particular, the H I amplitude parameter, $\log A_{\rm H\,I}$, is constrained by the prior and can not be fit efficiently by the MCMC unless a total temperature constraint is imposed. The best-fit values are derived by cumulative statistics as the marginalized parameter likelihoods exhibit non-Gaussian characteristics, as can be seen in Fig. 7.

The resulting posteriors of the MCMCs of the autopower spectra of both H I models are displayed in Figs 7 and 8. In Tables 2, and 3, we present the outcomes of the parameter estimation for the auto- and cross-power spectra of both H I models. We individually fit the power spectra rather than perform a joint analysis as in many upcoming experiments only one or the other will be available due to limitations in the quality of data or lack of an optical galaxy sample. The parameter fits can be extremely biased due to the fluctuations in the lognormal realizations. In order to derive mean parameter fits and the expected variance including Cosmic variance while remaining computationally feasible, we run MCMCs on the mean power spectrum of all 100 realizations presented under names {Auto, Cross} in each table, in addition to running

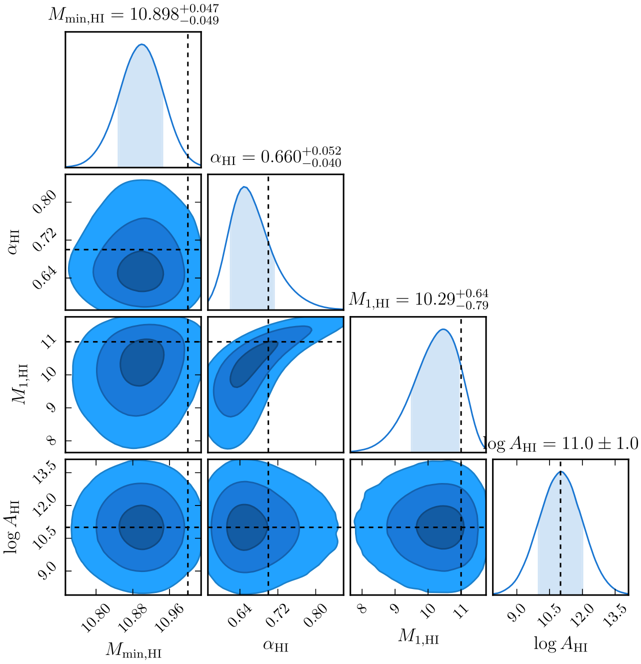

Figure 8. The likelihood contours of the MCMC fit to the realization-averaged power spectrum of the continuous H I model, using the H I-galaxy *sample* autocorrelation and to constrain the HOD parameters of HALOMOD. The dashed lines indicate the input parameter values. All masses are given as $\log_{10}$ and in units of $\mathrm{M}_{\odot}\,\mathrm{h}^{-1}$.

MCMCs on 10 individual realizations, and presenting the mean and standard deviation of their fits under $\{\mathrm{Auto}\sum_{\mathrm{i}=1}^{10},\ \mathrm{Cross}\sum_{\mathrm{i}=1}^{10}\}$ in the tables. Whereas the constraints given by the MCMCs of the mean demonstrate the degeneracy within the halo model parameters, the standard deviation over 10 realizations shows limitations due to Cosmic variance.

Table 2 and Fig. 7 presents the parameter constraints of the discrete H I model, where the theory is primarily fit to the Poisson noise amplitude in the given k range. From cross-correlation, we derive the ensemble-averaged H I mass of the galaxy sample from the estimated parameters which is given by the numerator of the cross Poisson noise. The input parameters correspond to $\langle \log_{10}(M_{\rm H\,I,\mathit{g}}/\mathrm{M}_{\odot}h)\rangle = 9.77$ and the mean of the $N = 10$ realizations gives $\langle \log_{10}(M_{\rm H\,I,\mathit{g}}/\mathrm{M}_{\odot}h)\rangle_N = 9.89 \pm 0.33$.

Table 3 and Fig. 8 display the results of the continuous H I model, where only four parameters need to be estimated in the auto- and cross-correlation. The individual parameter constraints are much tighter due both to the fewer number of parameters, and the fact that the spectrum shape is not dominated by a single Poisson noise term. The uncertainties due to Cosmic variance are comparable to the H I discrete case.

7 SUMMARY AND DISCUSSION

In this study, we present a new, adaptive description of the intensity mapping autopower spectrum and cross-power spectrum with galaxy surveys in the halo model framework (using HALOMOD). We introduce two different implementations for the description of H I populations; the continuous H I model which populates H I within Dark Matter halos following a smooth profile, and the discrete H I model which co-locates H I masses with the positions of an underlying galaxy field, where both H I and field can follow independent HOD descriptions. The models represent the opposite ends of the spectrum of currently used H I simulations. We inspect the impact of the different H I models on the shapes of the auto- and cross-power spectra and find that the H I power spectra of both models only differ on scales $k > 10\,\mathrm{h\,Mpc}^{-1}$, caused by the additional central-satellite contributions in the 1-halo term of the discrete model. The prediction of the 1-halo terms of both H I models in the cross-power spectrum with galaxies are very similar if the same H I halo profiles are used.

We verified our analytic predictions with a set of lognormal realizations, and find that the major difference between the models is the presence or absence of shot noise contributions. We review the current understanding of shot noise in galaxy and H I intensity mapping data and state analytic expressions to determine the amplitude of the shot noise given the underlying HODs. Most notably, the shot noise on the cross-power spectrum directly scales with the averaged H I mass of the optical galaxies, which is well-defined in the halo model framework.

We examine the shot noise properties of both H I models and find that the implementation of the continuous H I models has no Poisson noise contribution to any power spectra due to the continuous, smooth nature of the H I density field. The shot noise of the discrete H I model is correctly predicted by HALOMOD for the autocorrelation and cross-correlation with a galaxy *sample*, given the H I content is independent of the galaxy sample abundances. In our examples, the Poisson noise contributions dominate the amplitude of the overall power spectra from scales .The cross-Poisson noise is proportional to the averaged H I mass per galaxy in the *sample* and, as such, can be used to determine the average H I mass of galaxy samples without directly observing their H I content. Our HALOMOD implementation is the first tool to predict the cross-Poisson noise given H I and galaxy HOD parameters and will be useful for experimental forecasts as well as observational interpretations.

We demonstrate how the H I model parameters of the HALOMOD predictions can be fit to the simulations using MCMC techniques. These fits also estimate derived H I properties such as the average brightness temperature, which is directly proportional to $\Omega_{\rm H\,I}$, and the averaged H I mass per galaxy in the cross-correlation with galaxy *sample*, a quantity of great interest in future cross-correlation experiments on small scales. This way, HALOMOD has the potential to estimate the unknown parameters of the H I distribution traced by the H I intensity maps, as well as determining the averaged H I masses of galaxy samples in intensity mapping cross-correlation experiments.

We note that our study exclusively focuses on the impact of the halo occupation parameters on the power spectra and Poisson noise. We have not considered the degeneracy of cosmological parameters and halo occupation parameters, but on the scales considered in this work the effect of galaxy evolution dominates. We note that nonlinear effects of the power spectrum and peculiar velocities alter the shape and amplitude of the 1-halo term, however, on small enough scales, the contribution of the Poisson noise is considerably more dominant than the 1-halo term. In these cases, the HALOMOD prediction of the cross-Poisson noise could be added to a more sophisticated power spectrum model which includes these effects or the fits could be performed to the projected correlation function to suppress redshift-space distortions.

In this project, we did not employ data-motivated H I models as we aim to demonstrate a maximally flexible framework for H I auto- and cross-power spectrum. The qualitative results on the Poisson noise predictions do not depend on the specific parametrisation of

Table 2. Discrete H I model: Marginalized parameter likelihoods given by the MCMC fit to the averaged auto- and cross-power spectrum, and the mean and the standard deviation of the MCMC fit to 10 realizations marked with $\sum_{i=1}^{10}$, indicating the Cosmic variance. All masses are given as $\log_{10}$ and in units of $\mathrm{M_\odot\,h^{-1}}$.

Model	$M_{\rm min,\,HI}$	$\alpha_{\rm H\,I}$	$M_{1,\rm H\,I}$	$\alpha_{\rm field}$	$M_{1,\rm field}$	$\log A_{\rm H\,I}$
Input	11.0	0.7	11.0	0.5	11.0	11.0
Auto	$10.98^{+0.10}_{-0.11}$	$0.677^{+0.064}_{-0.060}$	$10.54^{+0.74}_{-0.88}$	$0.52^{+0.20}_{-0.14}$	$11.00^{+0.39}_{-0.62}$	11.0 ± 1.0
Auto $\sum_{i=1}^{10}$	10.945 ± 0.2429	0.704 ± 0.1388	10.67 ± 0.522	0.548 ± 0.0214	10.941 ± 0.1453	11.007 ± 0.0296
Cross	$10.97^{+0.18}_{-0.23}$	0.73 ± 0.12	$10.72^{+0.86}_{-0.95}$	$0.55^{+0.22}_{-0.16}$	$10.99^{+0.45}_{-0.66}$	$11.01^{+1.00}_{-1.02}$
Cross $\sum_{i=1}^{10}$	10.883 ± 0.3171	0.753 ± 0.1375	10.79 ± 0.2711	0.544 ± 0.0491	10.883 ± 0.2265	10.998 ± 0.0128

Table 3. Continuous H I model: Marginalized parameter likelihoods given by the MCMC fit to the averaged auto- and cross-power spectrum, demonstrating the degeneracies within the HOD parameters, and the mean and the standard deviation of the MCMC fit to 10 realizations marked with $\sum_{i=1}^{10}$, indicating the Cosmic variance. All masses are given as $\log_{10}$ and in units of $\mathrm{M_\odot\,h^{-1}}$.

Model	$M_{\rm min,\,HI}$	$\alpha_{\rm H\,I}$	$M_{1,\rm H\,I}$	$\log A_{\rm H\,I}$
Input	11.0	0.7	11.0	11.0
Auto	$10.898^{+0.047}_{-0.049}$	$0.660^{+0.052}_{-0.040}$	$10.29^{+0.64}_{-0.79}$	11.0 ± 1.0
Auto $\sum_{i=1}^{10}$	10.852 ± 0.3594	0.611 ± 0.1013	10.347 ± 0.4343	11.004 ± 0.0087
Cross	$10.866^{+0.106}_{-0.095}$	$0.714^{+0.083}_{-0.075}$	$10.61^{+0.84}_{-0.90}$	11.0 ± 1.0
Cross $\sum_{i=1}^{10}$	10.877 ± 0.3458	0.626 ± 0.1561	10.816 ± 0.3106	10.998 ± 0.0093

the H I model. Our adaptable framework allows to easily import any shape and parametrization of the H I-to-halo relation and examine their predictions. In future work, more data-driven models will be implemented in order to compare predictions with observations.

ACKNOWLEDGEMENTS

LW is supported by an ARC Discovery Early Career Researcher Award (DE170100356). This research was conducted by the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020. This work was performed on the gSTAR national facility at Swinburne University of Technology. gSTAR is funded by Swinburne and the Australian Government's Education Investment Fund.

REFERENCES

Alonso D., Ferreira P. G., Santos M. G., 2014, MNRAS, 444, 3183

Anderson L. et al., 2014, MNRAS, 441, 24

Anderson C. J. et al., 2018, MNRAS, 476, 3382

Baldauf T., Seljak U., Smith R. E., Hamaus N., Desjacques V., 2013, Phys. Rev. D, 88, 083507

Bandura K. et al., 2014, in Ground-based and Airborne Telescopes V. p. 914522preprint (arXiv:1406.2288)

Battye R. A., Davies R. D., Weller J., 2004, MNRAS, 355, 1339

Battye R. A., Browne I. W. A., Dickinson C., Heron G., Maffei B., Pourtsidou A., 2013, MNRAS, 434, 1239

Beutler F. et al., 2013, MNRAS, 429, 3604

Breysse P. C., Kovetz E. D., Behroozi P. S., Dai L., Kamionkowski M., 2017, MNRAS, 467, 2996

Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, ApJ, 803, 21

Camera S., Santos M. G., Ferreira P. G., Maartens R., 2014, JPhCS, 12004, JPhCS.566

Castorina E., Villaescusa-Navarro F., 2017, MNRAS, 471, 1788

Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303

Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463

Chen X., 2012, in International Journal of Modern Physics Conference Series. p. 256preprint (arXiv:1212.6278)

Coles P., Jones B., 1991, MNRAS, 248, 1

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1

Drinkwater M. J. et al., 2010, MNRAS, 401, 1429

Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, MNRAS, 390, L64

Fernandez X., Van Gorkom J. H., Gim H., Yun M. S., Momjian E. CHILES Team, 2016, in American Astronomical Society Meeting Abstracts #227. p. 323.04

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Ginzburg D., Desjacques V., Chan K. C., 2017, Phys. Rev. D, 96, 083528

Hamaus N., Seljak U., Desjacques V., 2011, Phys. Rev. D, 84, 083509

Harper S., Dickinson C., Battye R., Roychowdhury S., Browne I., Ma Y.-Z., Olivari L., Chen T., 2018, MNRAS, 478, 2416

Kim H.-S., Wyithe J. S. B., Baugh C. M., Lagos C. d. P., Power C., Park J., 2017, MNRAS, 465, 111

Kovetz E. D. et al., 2017, preprint (arXiv:1709.09066)

Lagos C. D. P., Baugh C. M., Zwaan M. A., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A. M., van Kampen E., 2014, MNRAS, 440, 920

Masui K. W. et al., 2013, ApJ, 763, L20

Murray S. G., Power C., Robotham A. S. G., 2013, Astron. Comput., 3, 23

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Newburgh L. B. et al., 2016, in Ground-based and Airborne Telescopes VI. p. 99065Xpreprint (arXiv:1607.02059)

Padmanabhan H., 2018, Peering towards Cosmic Dawn, 216, IAUS...333

Padmanabhan H., Refregier A., 2017, MNRAS, 464, 4008

Padmanabhan H., Refregier A., Amara A., 2017, MNRAS, 469, 2323

Paech K., Hamaus N., Hoyle B., Costanzi M., Giannantonio T., Hagstotz S., Sauerwein G., Weller J., 2017, MNRAS, 470, 2566

Paul N., Choudhury T. R., Paranjape A., 2018, MNRAS, 479, 1627

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Pen U.-L., Staveley-Smith L., Peterson J. B., Chang T.-C., 2009, MNRAS, 394, L6

Percival W. J. et al., 2007, ApJ, 657, 645

Planck Collaboration et al., 2016, A&A, 594, A13

Pourtsidou A., Bacon D., Crittenden R., 2015, Phys. Rev. D, 92, 103506
Pourtsidou A., Bacon D., Crittenden R., 2017, MNRAS, 470, 4251
Reid B. A. et al., 2012, MNRAS, 426, 2719
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
Santos M. et al., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 19
Santos M. G. et al., 2017, preprint (arXiv:1709.06099)
Sarkar T. G., Datta K. K., Pal A. K., Choudhury T. R., Bharadwaj S., 2016, J. Astrophys. Astron., 37, 26
Seljak U. et al., 2005, Phys. Rev. D, 71, 103515
Smith R. E. et al., 2003, MNRAS, 341, 1311
Smith R. E., 2009, MNRAS, 400, 851
Switzer E. R. et al., 2013, MNRAS, 434, L46
Switzer E. R., Chang T.-C., Masui K. W., Pen U.-L., Voytek T. C., 2015, ApJ, 815, 51
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Villaescusa-Navarro F. et al., 2018, ApJ, 866, 135
Wolz L. et al., 2017a, MNRAS, 464, 4938
Wolz L., Blake C., Wyithe J. S. B., 2017b, MNRAS, 470, 3220
Wyithe J. S. B., Loeb A., Geil P. M., 2008, MNRAS, 383, 1195
Zehavi I. et al., 2005, ApJ, 630, 1
Zehavi I. et al., 2011, ApJ, 736, 59
Zheng Z. et al., 2005, ApJ, 633, 791

APPENDIX A: CORRELATION MODELS FOR H I-GALAXY SAMPLES

A1 Continuous H I model

In this model, the gas is not colocated with observed galaxies, forming a spatially-independent smooth profile within the halo. In reality, we expect that while the *averaged* profile of the gas is smooth, it will be lumpy on scales much smaller than the halo radius. It is then expected that the prospects of observing a galaxy in the sample may be dependent on the H I density around the location of the galaxy. However, dealing with this general situation, in which spatial scales within the halo are correlated according to the typical size of the H I 'lumps" is rather difficult, and we may consider two extreme cases in more detail. The first is that in which the lumps are infinitely broad, or rather that the H I profile is perfectly smooth for every halo. The second is that in which the 'lumps" are Dirac-δ functions, but this is equivalent to the discrete H I model which we consider in the following subsection.

Suppose that the H I profile of every halo is always completely smooth, and is constant with the underlying halo mass. Suppose also that there is a distribution of H I masses for a given halo mass, for which the mean is $\langle M_{\mathrm{H\,I}}(m)\rangle$, and the variance is $\sigma^2_{\mathrm{H\,I}}$ (the distribution remains arbitrary, but one may like to think of it as a Gaussian or Lognormal). If a particular halo has an H I mass from the upper-tail of its distribution, then the H I density of that halo is increased uniformly everywhere in the halo, because it is necessarily completely smooth. Now consider a sample of observed galaxies. The probability of finding a galaxy at any point in a given halo may depend on the density of the H I in that location (in fact, it may depend on much more than that, for example, it may depend on the H I density in nearby locations, or the dynamical state of the H I rather than just its abundance, but these are considered to be minor complications which we will ignore). However, since the density of H I at any given location is determined by the density at all other locations, or rather, the density at any location is fully specified by the total H I mass in the halo – due to its smoothness – the total expected number of observed galaxies in the halo is completely determined by its H I mass. Summarily, we have the following system:

$$\begin{aligned} M^i_{\mathrm{H\,I}} &\sim \phi(m_{\mathrm{H\,I}}, m), \\ N^i &\sim \mathrm{Poisson}(f(M^i_{\mathrm{H\,I}})), \end{aligned} \tag{A1}$$

where f is some function which converts the actual H I mass of a halo into the expected number of observed galaxies. While this function is arbitrary, a simple but flexible toy model is such that $f = n_0(M^i_{\mathrm{H\,I}}/A)^\gamma$. Letting $\tilde{A}(m)$ be the average amount of H I per galaxy in halos of mass m, we obtain that $\langle N\rangle = \langle M_{\mathrm{H\,I}}\rangle/\tilde{A}$. Furthermore, we find that

$$\tilde{A} = \frac{A^\gamma}{n_0}\frac{\langle M_{\mathrm{H\,I}}\rangle}{\langle M^\gamma_{\mathrm{H\,I}}\rangle}. \tag{A2}$$

We focus hereafter on some special cases, $\gamma \in (-1, 0, 1)$, corresponding to anticorrelated, uncorrelated, and correlated cases. For these, we have

$$\tilde{A} = \begin{cases} (An_0)^{-1}\frac{\langle M_{\mathrm{H\,I}}\rangle}{\langle M^{-1}_{\mathrm{H\,I}}\rangle}, & \gamma = -1 \\ \frac{\langle M_{\mathrm{H\,I}}\rangle}{n_0}, & \gamma = 0 \\ \frac{A}{n_0}, & \gamma = 1. \end{cases} \tag{A3}$$

We note that the distribution of N in such a setup is not necessarily Poisson, as we generally assume it to be. Nevertheless, it is not likely to be significantly different to Poisson, and in any case, this fact does not affect the rest of our calculations.

We wish to calculate the value of $\langle N_s M_{\mathrm{H\,I}}\rangle$. This can be achieved by using the law of total expectation,

$$\begin{aligned} \langle N_s M_{\mathrm{H\,I}}\rangle &= \int dm' m'\phi(m', m)\sum_{k=0}^{\infty} k\mathrm{Pois}(f(m')) \\ &= \sum_{k=0}^{\infty}\frac{n_0}{A^{\gamma k}(k-1)!}\int dm'\ m'^{\gamma k+1}\phi(m', m)e^{-n_0(m'/A)^\gamma}. \end{aligned} \tag{A4}$$

If we assume that ϕ is a Gaussian distribution, then for masses m at which the expected number of galaxies is large, since the Poisson distribution tends to a Gaussian, the result tends to

$$\langle N_s M_{\mathrm{H\,I}}\rangle = \frac{n_0 A^{-\gamma}}{\sqrt{2\pi}\sigma}\int dm'\ m'^{1+\gamma}e^{-(m'-\langle M_{\mathrm{H\,I}}\rangle)^2/2\sigma^2_{\mathrm{H\,I}}}. \tag{A5}$$

While this is in general unsolvable, it yields solutions for our three cases of interest:

$$\langle N_s M_{\mathrm{H\,I}}\rangle \approx \begin{cases} n_0 A, & \gamma = -1 \\ n_0\langle M_{\mathrm{H\,I}}\rangle = \langle N\rangle\langle M_{\mathrm{H\,I}}\rangle, & \gamma = 0 \\ n_0\frac{\langle M_{\mathrm{H\,I}}\rangle^2+\sigma^2_{\mathrm{H\,I}}}{A}, & \gamma = 1. \end{cases} \tag{A6}$$

Thus, the correlation function is (in the large-N limit):

$$\begin{aligned} R(m) &= \frac{\langle N_s M_{\mathrm{H\,I}}\rangle - \langle N\rangle\langle M_{\mathrm{H\,I}}\rangle}{\sqrt{\langle N\rangle}\sigma_{\mathrm{H\,I}}} \\ &= \begin{cases} \frac{1}{\sigma_{\mathrm{H\,I}}}\sqrt{\frac{\langle M_{\mathrm{H\,I}}\rangle}{\tilde{A}}}\left(\frac{1}{\langle M^{-1}_{\mathrm{H\,I}}\rangle} - \langle M_{\mathrm{H\,I}}\rangle\right), & \gamma = -1 \\ 0, & \gamma = 0 \\ \frac{\sigma_{\mathrm{H\,I}}}{\sqrt{\tilde{A}\langle M_{\mathrm{H\,I}}\rangle}}, & \gamma = 1. \end{cases} \end{aligned} \tag{A7}$$

The question of how to calculate $\langle M^{-1}_{\mathrm{H\,I}}\rangle$ remains. We find that a useful empirical formula is such that

$$\langle M^{-1}_{\mathrm{H\,I}}\rangle \approx \frac{\langle M_{\mathrm{H\,I}}\rangle^2 + \sigma^2_{\mathrm{H\,I}}}{\langle M_{\mathrm{H\,I}}\rangle^3} \tag{A8}$$

when $M_{\rm H\,I}$ is a Gaussian variable, and $\langle M_{\rm H\,I}\rangle/\sigma_{\rm H\,I} > 3$. This latter condition must be obeyed in any case to ensure the description is physically appropriate, otherwise a significant part of the probability density puts $M_{\rm H\,I} < 0$. Under this approximation, the correlation function becomes

$$R_{\gamma=-1}(m) = -\sigma_{\rm H\,I}\langle M_{\rm H\,I}\rangle \frac{\sqrt{\frac{M_{\rm H\,I}}{\bar{A}}}}{\langle M_{\rm H\,I}\rangle^2 + \sigma_{\rm H\,I}^2} \tag{A9}$$

A2 Discrete H I model

The result in the case of the discrete model, in which the ‘lumps” are Dirac-δ functions, is much simpler. The typical assumption of the galaxies obeying a Poisson distribution carries with it the assumption that they are spatially independent, and we have by construction specified that the H I components are also spatially independent. This implies that while the probability of observation of a galaxy at a certain point may be dependent on the H I in that location, it is entirely uncorrelated with any other point. This means that all correlations exist at a separation of zero, which is not represented in the power spectrum at all. Alternatively one may consider equation (13), in which the total contribution of the satellite–satellite term has a $-Q$ term, which accounts for all self-pairs. After subtracting the self-pairs, no other pairs contain any correlations, and so, in general, we have that

$$\langle N_s M_{\rm H\,I}\rangle - Q = \langle N_s\rangle\langle M_{\rm H\,I}\rangle. \tag{A10}$$

Nevertheless, while these correlations cannot change the *shape* of the power spectrum, they do affect the level of shot-noise present. This is simple to conceptualise; since the shot-noise depends on the average H I mass within galaxies of the *sample*, a correlation which favours observing galaxies which contain a higher H I mass will therefore accordingly increase the shot-noise, and vice versa. The net result is a constant additive factor to the observed power spectrum, and thus detailed modelling will not usually be required – the constant may just as well be fit and then interpreted.

In practice, one may conceive of the correlation occurring in multiple ways. In any of these, if the mean H I mass of the sample can be calculated, it can be used to directly infer the amplitude of the shot noise.

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