



















H_I Simulations for Cosmology with the SKA Observatory

Tommaso Ronconi ^{1,2} Gabriella De Lucia ^{3,7} Marta Spinelli ^{4,19} Pascal Hitz ⁵
Fabio Fontanot ^{3,7} Francesco Sinigaglia ^{6,7} Lizhi Xie ⁸ Robert M. Yates ¹⁸ José
Luis Bernal ²⁰ Anna Bonaldi ⁹ Stefano Camera ^{10,11,12} Isabella P. Carucci ³
Steven Cunnington ^{13,14} Francisco-Shu Kitaura ^{21,22} Joël Mayor ⁵ Rajesh Mondal
¹⁵ Giulia Piccirilli ^{10,11} and Digvijay Wadekar ^{16,17}

¹*INAF – Institute of Radioastronomy (IRA), Via Gobetti 101, 40129 Bologna, Italy*

²*Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, 34136 Trieste, Italy*

³*INAF – Osservatorio Astronomico di Trieste, Via Tiepolo 11, 34131 Trieste, Italy*

⁴*Observatoire de la Côte d’Azur, Laboratoire J-L Lagrange, Boulevard de l’Observatoire, Nice, France*

⁵*ETH Zurich – Institute for Particle Physics and Astrophysics, 8093 Zurich, Switzerland*

⁶*University of Geneva, 24 rue du Général-Dufour, 1211 Genève, Switzerland*

⁷*IFPU: Institute for Fundamental Physics of the Universe, Via Beirut, 2, 34151 Trieste, Italy*

⁸*Tianjin Normal University, Binshuixidao 393, 300387, Tianjin, China*

⁹*SKA Observatory, Jodrell Bank, Lower Whittington, Macclesfield, SK11 9FT, UK*

¹⁰*Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy*

¹¹*INFN, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy*

¹²*INAF, Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese, Italy*

¹³*Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth, PO1 3FX, UK*

¹⁴*Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK*

¹⁵*Department of Physics, National Institute of Technology Calicut, Calicut 673601, Kerala, India*

¹⁶*Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland, 21218, USA*

¹⁷*School of Natural Sciences, Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA*

¹⁸*Centre for Astrophysics Research, University of Hertfordshire, Hatfield, AL10 9AB, UK*

¹⁹*Department of Physics & Astronomy, University of the Western Cape, Cape Town 7535, South Africa*

²⁰*Instituto de Física de Cantabria (IFCA), CSIC-Univ. de Cantabria, Avda. de los Castros s/n, E-39005 Santander, Spain*

²¹*Instituto de Astrofísica de Canarias, Vía Láctea s/n, E-38205, La Laguna, Tenerife, Spain*

²²*Departamento de Astrofísica, Universidad de La Laguna, E-38206, La Laguna, Tenerife, Spain*

E-mail: tommaso.ronconi@inaf.it, gabriella.delucia@inaf.it

We present a comparative overview of state-of-the-art methods for modelling the distribution of neutral hydrogen (H I) in the post-reionization Universe, developed in preparation for upcoming SKAO cosmological surveys. Our aim is to assess how different physical and empirical assumptions reflect into predictions for key observables such as the cosmic H I density, the H I mass function, and the H I-halo mass relation. We consider both: (i) semi-analytical approaches that self-consistently evolve baryonic components within dark matter merger trees through physically motivated prescriptions and (ii) empirical schemes tailored to different observables and based on fast approximations designed for large ensemble studies. By comparing the predictions from the different methods considered, we find overall consistency in integrated quantities such as Ω_{HI} , yet systematic differences in the detailed shape and scatter of the H I-halo mass relation and its redshift evolution. Semi-analytical models offer physically grounded predictions but depend on assumed prescriptions, while empirical methods provide flexibility and computational efficiency at the expense of robustness in extrapolated regions of the parameter space. The increasing number of H I measurements from SKA precursors and pathfinders (including surveys with MeerKAT, ASKAP, and FAST) will provide critical observational constraints to refine and calibrate current simulation methodologies. In turn, increasingly realistic H I simulations play a key role in interpreting these data, guiding survey design and analysis strategies, in preparation for the advent of SKAO data.

1 Introduction

Modelling of neutral hydrogen (H I) across cosmic time is a key requirement for forthcoming large-scale surveys with the SKA Observatory (SKAO; [Braun et al., 2019](#)) and its precursors, that will exploit H I-based observables to address both astrophysical and cosmological questions. On large scales, the distribution and evolution of H I trace the underlying matter density field (e.g. [Chang et al., 2010](#); [Masui, 2013](#); [Padmanabhan et al., 2015](#); [Spinelli et al., 2020](#)), while on galactic scales, the H I reservoir fuels star formation and regulates the interplay between accretion, feedback, and the enrichment of the interstellar and circumgalactic media (e.g. [Furlanetto et al., 2006](#); [Pritchard and Loeb, 2010](#); [Crain et al., 2017](#); [Diemer et al., 2019](#); [Reina-Campos et al., 2022](#)). By connecting the dark and baryonic components of the Universe, H I serves as a bridge between cosmology and galaxy evolution, providing insight into structure formation, gas cycling, and the efficiency of galaxy growth. In the post-reionization era, H I is thus a powerful tracer of both the expansion history of the Universe, the physics of baryons and their interplay with the dark sector (e.g. [Blyth et al., 2015](#); [Santos et al., 2015](#)).

The SKAO observational capabilities will allow a variety of survey strategies. For cosmological studies, these include H I galaxy surveys, exploiting resolved detections of the 21-cm emission line to perform cosmological analyses via, for example, redshift-space distortions and baryon acoustic oscillations ([Mayor et al., 2026](#); [Nasirudin et al., 2026](#)), and H I intensity mapping (IM), where the integrated 21-cm signal from unresolved sources is measured to probe large-scale modes with high efficiency ([Wolz et al., 2026](#); [Mazumder et al., 2026](#)). Each approach requires accurate theoretical predictions and realistic mock catalogues to design optimal survey strategies, validate analysis pipelines, and interpret forthcoming observations.

The generation of H I mock data relies on a broad spectrum of modelling techniques, ranging from fully hydrodynamical simulations (e.g. [Crain et al., 2015, 2017](#); [Villaescusa-Navarro et al., 2018](#); [Stevens et al., 2019](#)) to semi-analytical galaxy formation models and empirical or statistical prescriptions calibrated against observations. While hydrodynamical simulations capture the detailed interplay between baryons and dark matter, they are still computationally expensive and limited in volume ([Lagos et al., 2026](#)). Semi-analytical and empirical methods, by contrast, offer the flexibility and scalability required for cosmological applications, enabling the production of large-volume realisations and light-cones suitable for the design and interpretation of SKAO surveys.

This Chapter provides an overview and comparison of state-of-the-art techniques developed to model the H I content of the post-reionization Universe, from galaxy-level catalogues to large-scale intensity maps. These methods are relevant to many of the scientific topics discussed in other chapters of this volume, including cosmological inference from 21-cm intensity mapping ([Wolz et al., 2026](#); [Majumdar et al., 2026](#)), as well as studies of H I-selected galaxies for cosmology and astrophysics ([Mayor et al., 2026](#); [Nasirudin et al., 2026](#)). By presenting a diverse set of approaches, this Chapter aims to give an overview of assumptions and limits of available methodologies for the design of future H I surveys and the interpretation of their data in preparation for the SKAO era. In fact, H I simulations are essential for forecasting survey performance and for validating analysis pipelines. They provide controlled environments in which to test the impact of instrumental effects, foreground removal, and systematic uncertainties on recovered cosmological signals, as discussed in [Spinelli](#)

et al. (2026). Mock realisations are also crucial for estimating covariance matrices and quantifying cosmic variance in higher-order statistics (Majumdar et al., 2026). As the community advances toward the first SKAO observations, the interplay between realistic H I simulations, mock survey generation, and data-analysis methodologies will be central to transforming 21-cm observations into robust cosmological constraints.

This Chapter is organised into three main Sections. In Sec. 2, we introduce a representative, though not exhaustive, sample of methodologies developed to simulate neutral hydrogen, encompassing both H I galaxy catalogues and approaches tailored for H I intensity mapping (IM). In Sec. 3, we present a comparative analysis of the selected methodologies based on the simulations described above. Finally, in Sect. 5, we conclude by outlining the main results of our comparison, highlighting the general trends, limitations, and strengths of the different methods and providing an outlook on future developments informed by upcoming observations.

2 Methods

In recent years, substantial progress has been made in modelling the interstellar medium (ISM) within cosmological hydrodynamical simulations, with several suites of zoom-in or small-volume simulations now able to follow gas cooling down to temperatures of $\lesssim 10$ K (Agertz et al., 2020; Reina-Campos et al., 2023; Feldmann et al., 2023; Schaye et al., 2025). Such treatments, however, require extremely high mass and spatial resolution, and are therefore unfeasible for large cosmological volumes. To overcome these limitations, a range of complementary methods have been developed, that aim to predict the large-scale distribution of H I while avoiding the high computational cost of full hydrodynamical calculations.

In this Section we focus on five representative methods, selected to illustrate the heterogeneous strategies currently available for simulating the large-scale H I distribution. These include both semi-analytical and empirical frameworks and differ in their complexity, predictive power, and calibration strategies.

The sample of methods presented here is by no way exhaustive. Several additional approaches have been proposed in the last ~ 15 years. For example, the semi-numerical pipeline of Mondal et al. (2015b,a, 2017) combines a parallelized particle-mesh N -body simulation¹ to generate the dark matter density field with a friends-of-friends halo finder² to identify collapsed structures. Neutral hydrogen is then assigned to haloes following the H I–halo mass relation proposed by Bagla et al. (2010). The resulting H I distribution is used to construct redshift-space 21-cm light-cones for observability forecasts. Empirical H I host halo relations can also be applied to lightcones with semi-empirical astrophysical properties obtained from e.g., UniverseMachine (Behroozi et al., 2019), coherent with simulated emission from other tracers of the LSS, as done in Sato-Polito et al. (2023). Other groups have proposed techniques that exploit machine learning to associate an H I distribution to the output of dark matter only simulations. For example, Wadekar et al. (2020) introduced an augmented H I–halo mass relation that extends the traditional $M_{\text{HI}}-M_h$ scaling

¹<https://github.com/rajeshmondal18/N-body>

²<https://github.com/rajeshmondal18/FoF-Halo-finder>

by incorporating secondary halo properties $\{i_h\}$ such as environment, concentration, spin, and formation epoch. The relation $M_{\text{HI}} = f_{\text{ML}}(M_h, \{i_h\})$ is learnt using random forest and symbolic regression techniques trained on hydrodynamical simulations (Villaescusa-Navarro et al., 2018). Therefore, the approach explicitly accounts for assembly-bias that is traditionally neglected in standard empirical approaches. Subsequently, Wadekar et al. (2021) presented a convolutional neural network (CNN) model that maps the dark matter density field, $\rho_m(\mathbf{x})$, to the corresponding HI density, $\rho_{\text{HI}}(\mathbf{x})$, producing high-resolution HI boxes from coarse-grained dark matter inputs and enabling fast generation of realistic large-scale HI fields.

In certain applications, multiple ($O(100)$) realisations of HI simulations are required—such as to estimate covariances in clustering measurements or to reconstruct signal loss due to background cleaning in HI intensity mapping experiments (e.g. Switzer et al., 2015; Cunnington et al., 2023; MeerKLASS Collaboration et al., 2025). A fast mock-generation process is essential in these cases and many of the previously mentioned techniques, while accurate, are still too computationally demanding to this aim. A more direct approach for generating density fields consists in producing realisations from a prescribed power spectrum. Such fields, by construction, reproduce their input power spectrum with variance determined by the simulation volume (i.e. the sample variance). The simplest implementation involves generating Gaussian random realisations of the input $P(k)$; however, a log-normal transformation can be applied, at negligible additional cost, to obtain a field whose probability distribution function more closely resembles that of the cosmic matter density (Shin et al., 2017). Such simulations can be realised, for example, by adapting to the HI case the Generator for Large Scale Structure (GLASS; Tessore et al., 2023), that generates highly detailed and realistic galaxy survey simulations at high angular resolution while efficiently optimising computational resources. Additional realism can be incorporated by, for instance, Poisson-sampling galaxies or adding velocity-field realisations (e.g. Agrawal et al., 2017; Murray, 2018). Then, the line luminosity of each galaxy can be sampled from an input luminosity function to include a realistic distribution of intensities in the resulting map, as done in SIMPLE (Lujan Niemeyer et al., 2023).

In the following sub-sections, we describe the specific methods used in this work, that form the basis for the comparison presented in Sec. 3.

2.1 Semi-Analytical Models

In the semi-analytical approach, the evolution of the baryonic components of dark matter haloes is modelled using simple but physically motivated prescriptions that are coupled to a description of the assembly history of dark matter haloes, which in the jargon is referred to as ‘merger tree’. The latter can be either based on analytical approaches (e.g. extended Press-Schechter formalism, eventually with corrections based on results from high-resolution dark matter simulations - see for example Parkinson et al. 2008) or extracted from numerical simulations (e.g. Tweed et al., 2009). By adopting this approach, it is possible to express the galaxy formation process essentially through a set of (coupled) differential equations that govern the exchanges of mass and energy between the different baryonic components of a galaxy. Given our limited understanding of the physical processes at play, the description involves a number of free parameters (and parametrizations) that are typically tuned to reproduce a given subsample of observational measurements in the local

Universe.

The approach is very flexible and is characterized by limited computational costs, which gives access to a large dynamic range in mass and spatial resolution in large computational volumes allowing an efficient exploration of the parameter space and of the influence of specific physical mechanisms. However, assumptions need to be made to compensate for the lack of an explicit treatment of the gas dynamics, and for the lack of information about the spatial distribution of baryonic matter.

The two models introduced below are state-of-the-art models that find roots in the original model developed around the ‘Munich’ group (Springel et al., 2001; De Lucia et al., 2004; Croton et al., 2006; De Lucia and Blaizot, 2007, and references therein). They are both coupled to substructure based merger trees, i.e. the resolution of the backbone simulation is high enough to resolve the cores of haloes accreted onto larger structures (up to some limit cosmic time, depending on the resolution, in time and mass, of the simulation). Below, we give an overview of the two models, with references to the original work where all details can be found.

2.1.1 GAEA (*sam1*)

The GALaxy Evolution and Assembly (GAEA) model is based on the original model introduced in (De Lucia and Blaizot, 2007) but it has been significantly updated since then. In particular, the most recent rendition of the model De Lucia et al. (2024) includes: (i) a detailed chemical enrichment scheme that accounts for the finite stellar lifetimes and the associated recycling of metals, energy and gas (De Lucia et al., 2014); (ii) an updated stellar feedback scheme that is partially based on results from high-resolution hydrodynamical simulations (Hirschmann et al., 2016); (iii) an updated treatment for gas accretion onto supermassive black holes and the associated feedback (Fontanot et al., 2020) and an updated treatment for the stripping of hot gas associated with infalling satellites and cold gas stripping from satellite galaxies (Xie et al., 2020). The model also includes a treatment for the partition of cold gas in its atomic and molecular gas component, and a star formation law that is based on the molecular gas (Xie et al., 2017). The model version adopted in this Chapter uses empirical relations based on the hydrostatic pressure of the disc (Blitz and Rosolowsky, 2006, previous work shows that alternative prescriptions lead to very similar predictions at the galaxy masses and redshift ranges considered here).

GAEA is tuned to reproduce the observed galaxy stellar mass function and the AGN luminosity function up to $z \sim 3$, and the H I and H₂ galaxy mass functions measured in the local Universe. It has been shown to successfully reproduce a number of observational measurements over a wide range of cosmic epochs. Relevant for this paper: model predictions are in quite good agreement with observational measurements of scaling relations between the amount of gas in different phases and galaxy stellar mass in the local Universe, both for central and satellite galaxies (Xie et al., 2020, see also Zoldan et al. 2017). In addition, the model reproduces the observed clustering strength of galaxies selected on the basis of their H I mass (Fontanot et al., 2025). In Spinelli et al. (2020) the authors presented a comprehensive analysis of model predictions for the H I content of dark matter haloes, based on a previous version of our model and extending up to $z \sim 5$.

In the default run used for this Chapter, GAEA is coupled to high-resolution cosmological N-body simulations (Fontanot et al., 2025). In Cammelli et al. (2025) the model has been coupled to merger

trees extracted from PINOCCHIO (PINpointing Orbit Crossing Collapsed Hierarchical Objects, Monaco, 2016), which allows access to higher resolution cosmological boxes and the construction of several independent mocks with limited computational time. Finally, dedicated software has been developed to associate an H I emission line with each simulated galaxy in post-processing (Mayor et al., 2026). Developed for GAEA but also tested on L-Galaxies, this allows realistic modelling, validated against observational data, of line parametrisation (e.g. width) statistics, scaling relations (e.g. Tully-Fisher) and H I galaxy survey selection systematics. This software, as well as outputs from GAEA, are used in other Chapters in this Book (i.e. Nasirudin et al., 2026; Mayor et al., 2026) to derive predictions for H I galaxy survey number counts, redshift distributions, and H I galaxy bias as functions of instrument sensitivities.

HIP-POP: Building upon the baryonic prescriptions derived from the GAEA semi-analytical model, the H I Probe Populator (HIP-POP³) provides a fast, approximate method to populate large-scale dark matter halo catalogues with neutral hydrogen for intensity mapping applications, where the limited angular resolution does not require modelling of individual galaxies.

The key ingredient is the $M_{\text{HI}} - M_{\text{halo}}$ relation, modelled following Spinelli et al. (2020) as:

$$M_{\text{HI}}(M_h) = M_h \left[a_1 \left(\frac{M_h}{10^{10}} \right)^\beta e^{-\left(\frac{M_h}{M_{\text{break}}} \right)^\alpha} + a_2 \right] e^{-\left(\frac{M_{\text{min}}}{M_h} \right)^{0.5}}, \quad (1)$$

where M_{HI} and M_{halo} are the hydrogen and halo mass, while a_1 , β , α , M_{break} , a_2 , and M_{min} are free parameters. Their redshift evolution is derived from fits to GAEA outputs across multiple snapshots, using a Gaussian likelihood and the MULTINEST sampler (Feroz et al., 2009) through an MPI-enabled Python wrapper (Zwart et al., 2016). Parameter values and scatter are interpolated in redshift using spline fits up to $z = 1$.

This semi-empirical framework allows for the rapid generation of large-volume, full-sky H I mock maps consistent with the adopted semi-analytical galaxy formation model.

2.1.2 L-Galaxies (sam2)

As mentioned above, L-Galaxies (Croton et al., 2006; Guo et al., 2011; Henriques et al., 2015, 2020; Yates et al., 2024) has roots back to similar models on which GAEA is based. However, the models have developed completely independently over the past ~ 15 years and differ significantly in terms of the physical prescriptions adopted. The L-Galaxies version adopted in this Chapter (Yates et al., 2024) includes the following advanced prescriptions: (i) a detailed chemical enrichment scheme including binary stars (Yates et al., 2013, 2024) that tracks a full complement of 118 chemical elements produced from 16 different enrichment sources; (ii) an explicit treatment for dust production and destruction (Vijayan et al., 2019; Yates et al., 2024); (iii) a modelling for radially resolved gas and stellar discs (Fu et al., 2013).

As for GAEA, L-Galaxies includes an explicit treatment for the partition of cold gas in H I and H₂. The model version considered in this Chapter follows the metallicity- and density-dependent formalism from McKee and Krumholz (2010). This formalism returns an H₂ fraction of 0 at

³<https://github.com/spinemart/HIP-POP>

densities below $\sim 30 M_{\odot}\text{pc}^{-2}$ and 1 at densities above $\sim 100 M_{\odot}\text{pc}^{-2}$. The L-Galaxies partitioning prescription returns a good match to the $z = 0$ H I mass function (by construction) and to the M_{HI}/M_* fraction (Henriques et al., 2020). However, the cosmic H I and H₂ mass densities are lower than inferred from observations at $z \sim 3 - 4$, as it is the case for many other semi-analytical models, including GAEA. Likely, this offset is due to the lack of cold gas modelling at the interface between the ISM and CGM (Yates et al., 2021).

The L-Galaxies source code is fully open source and available on [GitHub](#), with detailed installation instructions provided on the [L-Galaxies website](#). Outputs from various L-Galaxies versions are also available for download from the [Millennium database](#) (free registration required), with selected outputs from the latest version of the model also available via [Zenodo](#).

2.2 Empirical Models

An alternative approach to link dark matter haloes with the expected properties of H I galaxies, or more generally with the large-scale distribution of neutral hydrogen, is to employ simulation methodologies that are, to some extent, directly tuned to observational data of the baryonic properties they aim to reproduce. The primary goal of these methods is not to explicitly model the physical processes responsible for the emergence of the observables, but rather to rely on empirical assumptions and approximations in order to generate outputs that reproduce the same statistical properties of the selected galaxy population observed at a given stage of its evolution.

These models are designed to bypass the uncertainties associated with sub-grid physics in simulations of galaxy formation and evolution. Instead, they rely on predefined recipes either to sample populations of mock objects with given observable properties or to assign such properties to a background simulated DM field. While, on the one hand, these models are computationally convenient, on the other, their predictive power is inherently limited. The reason for this is twofold: first, their outputs are only as reliable as the observational datasets used for calibration; and second, when the model predictions extend beyond the range over which they were tuned, they rely on extrapolations that may not accurately represent the real Universe.

In most cases and in the framework relevant for this Chapter, the main goal of these approaches is to deliver a credible statistical description of clustering as a function of redshift. This is achieved by assuming a specific H I mass-halo mass relation ($M_{\text{HI}} - M_{\text{halo}}$ or H I HM). The latter can be conditioned to environmental properties derived from physically motivated arguments or based on results from a parent hydrodynamical simulation. With additional assumptions, the techniques can be extended to provide explicit H I distributions for galaxies populating dark matter haloes. In this Chapter, we consider predictions from three different empirical methods. In the following subsections, we provide a summary of their specific implementations.

2.2.1 T-RECS + SCAM (emp1)

This is a publicly available pipeline that simulates radio sources by overlaying them on a halo/subhalo distribution derived from cosmological N-body simulations. The pipeline enables the generation of full-sky light-cones of H I galaxies along with their continuum counterparts. Full details are given in [Ronconi et al. \(2026\)](#). The main components of the method can be summarized as follows:

- N-body simulations provide a structural framework to identify dark matter haloes and sub-haloes, that host galaxies (both the line-emitting and continuum populations). Specifically, the pipeline employed in this Chapter uses a full-sky light-cone created by tiling snapshots from the DEMNUni simulation set (Carbone et al., 2016; Parimbelli et al., 2022) up to redshift $z \sim 5$, specifically leveraging the high-resolution run (HR-DEMNUni, e.g. Hernández-Molinero et al., 2024).
- Catalogues of continuum and line-emitting galaxies are generated using T-RECS (Tiered Extragalactic Radio Continuum Simulations, Bonaldi et al., 2018, 2023), based on empirical measurements. The mock populations are sampled independently to reproduce observed statistics, such as the H I mass function (up to redshift $z \sim 0.5$) and the variation of H I flux with mass, as well as the source size and morphology distributions.
- To ensure that the synthetic galaxies are accurately coupled with the dark matter distribution, the pipeline applies the SCAM (Sub-halo Clustering and Abundance Matching) method (Guo et al., 2016), implemented in the Python library SCAMPy (Ronconi et al., 2020). SCAM follows a two-step approach: first, two independent Halo Occupation Distribution (HOD) models are used to select the sub-haloes that host the two populations considered. These are tuned to match the clustering data of H I galaxies, star-forming galaxies and active galactic nuclei independently. As a second step, a Sub-halo Abundance Matching algorithm is used to connect synthetic T-RECS sources with the DM haloes and sub-haloes selected in the first step.

The final mock catalogues provide detailed information for H I line emission, continuum sources, and their dark matter host halos. These products include comprehensive source catalogues (up to redshift $z < 0.5$ for the H I galaxies and up to $z = 5$ for continuum sources) and pixelated maps. All codes described above are publicly available. Given that the independent populations are associated to the same light-cone, cross-catalogues can be extracted as well.

In the first release (Ronconi et al., 2026), two catalogues are made available:

1. The first catalogue is tuned to reproduce a realistic spatial distribution of sources by exploiting existing measurements of the two-point correlation function. The parameters of the HOD model are therefore calibrated by maximizing a Gaussian likelihood that simultaneously accounts for both the one-point and two-point statistics, namely the observed number density of sources and their clustering signal:

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2} \left[\chi_{n_g}^2(\boldsymbol{\theta}) + \chi_{\xi_r}^2(\boldsymbol{\theta}) \right], \quad (2)$$

where $\boldsymbol{\theta}$ denotes the free parameters of the HOD model, n_g is the spatial number density of sources, and ξ_r represents the scale-dependent two-point correlation function in real space. The observational datasets employed to constrain the HODs of the two main source populations in T-RECS are: the three-dimensional correlation function of H I-selected galaxies from Martin et al. (2012), and the angular correlation function of continuum-selected sources (SFGs and AGN) from Hale et al. (2017).

2. A second, deeper catalogue is generated by neglecting the clustering term in Eq. (2), thereby omitting an explicit modelling of the two-point statistics. In this case, the two independent HOD models are calibrated solely to reproduce the expected redshift evolution of the source number density, as predicted for the SKA-AA4 survey specifications (Hartley et al., 2023).

2.2.2 *H I Halo Model (emp2)*

Hitz et al. (2025) present a fast method for simulating the cosmological distribution of post-reionization H I based on a halo model approach, that is tailored for forecasting and forward modelling of intensity mapping experiments. The method can be summarised as follows:

- **Past Light Cone of Dark Matter Halos:** the method employs the approximate but fast PINOCCHIO code (Monaco et al., 2002) to simulate the past light cone of dark matter halos. PINOCCHIO is based on Lagrangian Perturbation Theory together with the Extended Press & Schechter formalism for non-spherical collapse (Bond et al., 1991; Monaco, 1995, 1997) to efficiently generate halo catalogues while maintaining clustering properties at the 10% level accuracy compared to N-body simulations. Hitz et al. (2025) describes a simulation suitable for the survey configuration of the HIRAX (Hydrogen Intensity and Real-time Analysis eXperiment, Newburgh et al. 2016; Crichton et al. 2022) radio interferometer array. Specifically, a light cone covering declinations between -15° and -45° in the frequency range 700–800 MHz (corresponding to redshifts 0.77–1.03) is generated. PINOCCHIO is run with parameters allowing to achieve a halo mass resolution of $4.3 \times 10^9 M_\odot$, leading to an estimated missing H I fraction of 2–3% over the considered frequency range.
- **H I-Halo Mass Relation:** the H I mass is subsequently assigned to dark matter halos based on their masses using a H I-halo mass relation constrained by observational data (Padmanabhan et al., 2017):

$$M_{\text{HI}}(M_{\text{halo}}, z) = \alpha f_{\text{H,c}} M_{\text{halo}} \left(\frac{M_{\text{halo}}}{10^{11} h^{-1} M_\odot} \right)^\beta \exp \left[- \left(\frac{v_{\text{c},0}}{v_{\text{c}}(M, z)} \right)^3 \right], \quad (3)$$

where $f_{\text{H,c}} = (1 - Y_{\text{p}})\Omega_{\text{b}}/\Omega_{\text{m}}$ is the cosmic fraction of H I with respect to all matter, M_{HI} and M_{halo} are the H I and halo mass, respectively, v_{c} is the virial velocity of a host halo, while α , β and $v_{\text{c},0}$ are free fitting parameters. In their work, the authors obtain the best-fit values $\alpha = 0.09$, $\beta = -0.58$, and $\log v_{\text{c},0} = 1.56$. The virial velocity as a function of halo mass (Barnes and Haehnelt, 2014) in the exponential introduces a cut-off in H I mass at low halo masses, reflecting that smaller haloes cannot effectively self-shield H I from ionizing radiation.

- **Map Generation:** in the final step, the past light cone is divided into 20 frequency bins, each with a bandwidth of 5 MHz, to generate pixelated HEALPix maps with an *NSIDE* parameter of 512. Maps are generated for dark matter and H I halo masses, halo counts, and H I brightness temperature. The latter is computed using the relation from Bull et al. (2015).

The simulation has been validated by comparing the mass function, mass density, three-dimensional power spectrum, and angular power spectrum of the maps with predictions from an analytical dark matter and H I halo model, implemented in PyCosmo (Refregier et al., 2018; Tarsitano et al., 2021;

Moser et al., 2022; Hitz et al., 2025), a Python package for theoretical cosmological predictions. The simulated data set, including the light-cone halo catalogue, snapshot halo catalogues ($1h^{-3}\text{Gpc}^3$ box at $z = 0.8, 0.9, 1$), as well as the HEALPix maps, is available at <https://cosmology.ethz.ch/research/software-lab/cosmological-neutral-hydrogen-simulation.html>.

2.2.3 Environment-dependent HIHM relation (emp3)

This method produces full-sky mock catalogues of H I sources built directly on a light-cone, with a smooth redshift evolution extending up to $z \approx 3.8$. It combines approximate dark matter simulations with a model that links dark matter haloes to their expected neutral hydrogen content. The process can be summarised in three main steps:

- **Dark matter light-cone:** the dark matter distribution is generated with the WebON code (Kitaura and Sinigaglia, 2026), that implements the *Augmented Lagrangian Perturbation Theory* (ALPT; Kitaura and Hess, 2013). In this framework, the growth of cosmic structures is followed by displacing particles from their initial to final positions using precomputed displacement fields. The light-cone is built by interpolating outputs at different redshifts and applying a correction based on the Zel’dovich approximation (Angulo and White, 2010). This provides a fast, approximated, but accurate way to reproduce the large-scale structure expected from gravitational evolution.
- **Populating the light-cone with haloes:** the dark matter field is then populated with dark matter haloes using the Hicobian bias model (Coloma-Nadal et al., 2024), that connects the matter distribution to the expected halo abundance through a non-linear and stochastic relation (e.g., Kitaura et al., 2014). The model accounts for the diversity of cosmic environments (such as voids, filaments, sheets, and knots) by classifying each region according to its local gravitational field (Hahn et al., 2007; Heavens and Peacock, 1988; McDonald and Roy, 2009; Sinigaglia et al., 2021; Sinigaglia et al., 2022; Kitaura et al., 2022; Coloma-Nadal et al., 2024). The model parameters are calibrated using the ABACUS SUMMIT suite of N -body simulations (Maksimova et al., 2021; Garrison et al., 2019, 2021), exploring a broad range of cosmological scenarios around the Planck18 cosmology (Planck Collaboration et al., 2020).
- **Assigning halo and H I properties:** once the number of haloes is defined in each region, the method assigns them positions, velocities, and masses. Halo positions are initially matched to dark matter particles and then slightly adjusted to reproduce the observed small-scale clustering (Forero Sánchez et al., 2024). Redshift-space distortions are modelled following Kaiser (1987), including both coherent large-scale motions and random velocities within haloes (e.g., Kitaura et al., 2012b,a; Heß et al., 2013; Kitaura et al., 2014, 2016; Bos et al., 2019; Sinigaglia et al., 2022, 2024b,a). Halo masses are drawn from analytic mass functions, while the amount of neutral hydrogen is assigned using the HIHM relation measured in hydrodynamical simulations from Villaescusa-Navarro et al. (2018). This relation is extended to include environmental effects, allowing the H I content to depend not only on halo mass but also on the surrounding large-scale structure (Sinigaglia and Kitaura, 2026). All parameters are redshift-dependent and are interpolated between simulation snapshots to produce a continuous, redshift-evolving H I catalogue.

3 Comparison

Having introduced the main methodologies developed by the different teams, we now turn to a comparison between some of their relevant outputs. The goal of this exercise is not to establish a ranking among models, but rather to illustrate how different methodological assumptions and calibration strategies reflect in the predicted statistical properties of neutral hydrogen across cosmic time. This comparison also highlights the relative strengths and limitations of the available approaches. A summary of the models contributing to this comparison is provided below:

- sam1:** the GAEA semi-analytical model described in Sec. 2.1.1 (results shown as green diamonds);
- sam2:** the L-GALAXIES semi-analytical model described in Sec. 2.1.2 (results shown as light-blue plus markers);
- emp1:** divided into emp1.1 (lavender upward triangles) and emp1.2 (golden squares), corresponding to the two T-RECS+SCAMPY catalogues described in Sec. 2.2.1;
- emp2:** the H I halo-model introduced in Sec. 2.2.2 (blue downward triangles);
- emp3:** the environment-dependent HIHM-relation method from Sec. 2.2.3 (orange crosses).

The comparison presented in this Section is neither homogeneous nor exhaustive, as the models differ in several fundamental aspects:

- **Baryonification strategy:** the sam1 and sam2 models follow a semi-analytical approach, while the remaining models are empirical, each relying on distinct prescriptions for connecting dark matter haloes and/or sub-haloes with baryonic observables.
- **Cosmological parameters:** these depend on the underlying dark-matter simulations used to generate the mock data. Specifically, all methods have been, or can be, run with different cosmologies. In this Section, we use simulation runs consistent with a *flat*- Λ CDM cosmology, which, though, vary minimally in the specific parameters chosen. The complete set of cosmological parameters employed in each case is listed in Tab. 1.
- **Simulated properties:** sam1, sam2 and emp1 provide galaxy-level information, whereas emp2 and emp3 only provide the total neutral hydrogen content associated with host haloes.
- **Geometry:** sam1 and sam2 output data in comoving simulation boxes at discrete redshifts, while the other three models produce full-sky light-cones. For both emp2 and emp3, results can be extracted both from a light-cone and from the underlying simulation boxes from which it was constructed.
- **Available redshifts:** the redshift coverage varies among the models, being related to the depth of the light-cones for the empirical models, and determined by the granularity of the snapshots for the semi-analytic models.

As a consequence of this heterogeneity, a direct one-to-one comparison across all statistics is not always feasible. In the following subsections, we specify on a case-by-case basis which models are compared, under what conditions, and for which physical quantities.

Table 1: Fiducial Flat- Λ CDM cosmological parameters used for the present manuscript. Note that these values are not necessarily the only possible set of parameters available for the listed method, only the parameters used for the measurements presented in Sec. 3. *References:* (1) [Baugh et al. \(2019\)](#); (2) [Springel et al. \(2005\)](#); (3) [Boylan-Kolchin et al. \(2009\)](#); (4) [Carbone et al. \(2016\)](#); (5) [Parimbelli et al. \(2022\)](#); (6) [Hernández-Molinero et al. \(2024\)](#); (7) [Monaco et al. \(2002\)](#); (8) [Kitaura and Sinigaglia \(2026\)](#).

DM Simulation	h	Ω_{CDM}	Ω_{Λ}	Ω_b	σ_8	n_s	Reference
Semi-Analytic Models							
sam1: <i>GAEA</i> (Sec. 2.1.1)							
P-Millennium	0.6777	0.307	0.693	0.04825	0.8288	0.9611	(1)
sam2: <i>L-Galaxies</i> (Sec. 2.1.2)							
Millennium	0.673	0.2663	0.685	0.0487	0.829	0.96	(2)
Millennium-II	0.673	0.2663	0.685	0.0487	0.829	0.96	(3)
Empirical Models							
emp1: <i>TRECS+SCAM</i> (Sec. 2.2.1)							
DEMNUi-HR	0.670	0.320	0.680	0.050	0.834	0.962	e.g. (4, 5, 6)
emp2: <i>HI Halo Model</i> (Sec. 2.2.2)							
PINOCCHIO	0.6736	0.3153	0.6847	0.0493	0.8111	0.9649	(7)
emp3: <i>HIHM relation</i> (Sec. 2.2.3)							
ALPT	0.6777	0.307	0.693	0.04825	0.8288	0.9611	(8)

3.1 HI cosmic density

Here we compare the H I cosmic density derived for the different simulations considered in this Chapter. This has been obtained by summing up all the hydrogen contained either in the box or along the light-cone, and dividing the resulting volume density by the critical density parameter for the corresponding cosmology:

$$\Omega_{\text{HI}}(z) = \frac{1}{V(z)\rho_{\text{crit},0}} \sum_i M_{\text{HI}}^i \quad (4)$$

where $V(z)$ is the volume of each redshift slice considered in the case of light-cones, while it corresponds to the volume of the simulation box for sam1 and sam2 (only snapshots are available for these). For each different method, the sum runs over all simulated systems in the catalogue: galaxies for sam1, sam2 and emp1 and host haloes for emp2. The critical density at redshift zero, $\rho_{\text{crit},0}$ is computed consistently with the cosmological parameters of the background simulation.

In Fig. 1 the results of the sam1, sam2, emp1 and emp2 methods are reported with different coloured markers and error-bars. The redshift dimension error corresponds to the thickness of the redshift

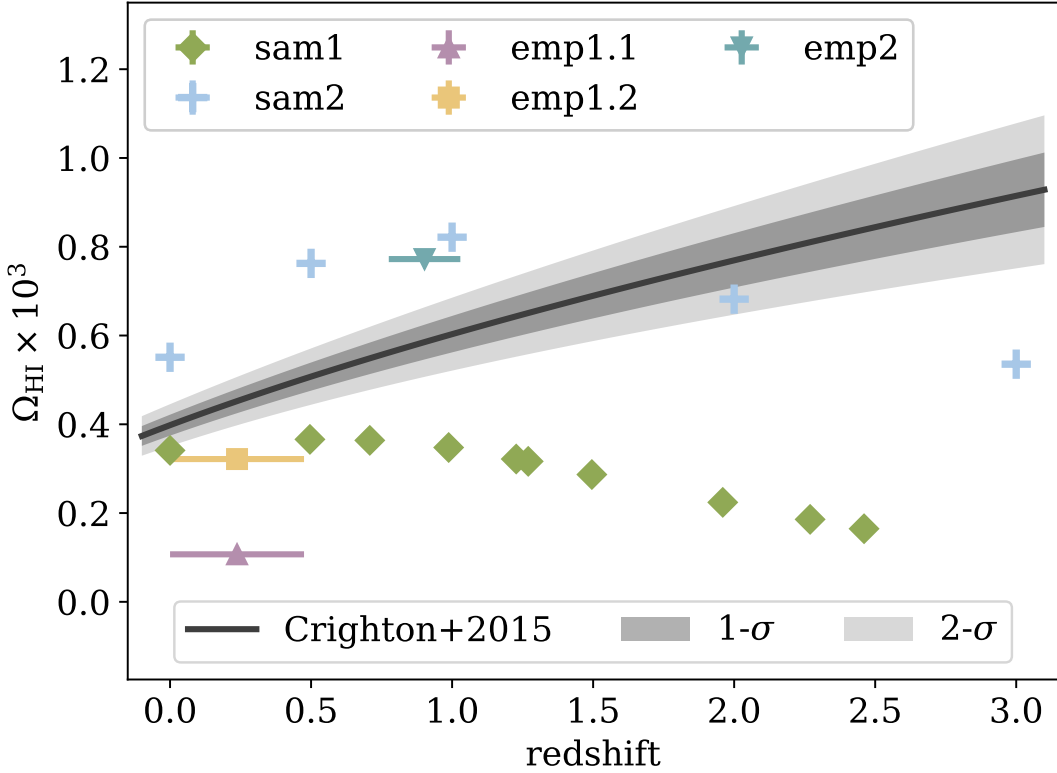


Figure 1: Measurements of Ω_{HI} at different redshifts from 3 methods out of the 4 listed in the Chapter. For comparison, we also mark the empirical relation found in [Crighton et al. \(2015\)](#) that represents a fit to observational data.

slice considered and is thus not present for sam1 and sam2 whose boxes are associated to discrete redshift values. The Ω_{HI} -dimension error was computed by bootstrap resampling both the boxes and the redshift slices. Since it varies between 0.01%–0.1%, for all the models and redshifts considered, it is not visible in the Figure.

As a term of comparison, we also show (solid black line with shaded grey regions marking the 1σ and 2σ confidence intervals) the empirical relation for the evolution of the hydrogen density parameter derived by [Crighton et al. \(2015, hereafter C15\)](#). The authors compiled measurements from independent studies, converted them to a common cosmology, and computed average values of the density parameter from redshift $z = 0$ up to $z = 5$. The solid line with its confidence regions in Fig. 1 represents a power-law of the form $\Omega_{\text{HI}} = A(1+z)^\gamma$, fitted to the dataset considered in C15. The confidence regions were obtained by Gaussian sampling around the mean and standard deviation of the best-fitting parameters: $A = (4.00 \pm 0.24) \times 10^{-4}$ and $\gamma = 0.60 \pm 0.05$. We note that [Péroux and Howk \(2020, hereafter PH20\)](#) later fitted the same functional form to an updated compilation of measurements, finding consistent values, namely $A = (4.6 \pm 0.2) \times 10^{-4}$ and $\gamma = 0.57 \pm 0.04$. In this work, we report only the C15 relation, as most of the models considered here were calibrated on data predating 2020. We also note that both C15 and PH20 rely on H I emission measurements at low redshift, and on absorption-based estimates to constrain the neutral

hydrogen abundance at earlier cosmic times.

Both the sam1 and sam2 models are calibrated to reproduce the H I galaxy mass function at redshift $z \sim 0$, and predict its evolution at higher redshift through physically motivated prescriptions. As shown in the figure, the green diamond symbols corresponding to predictions from sam1 are systematically below the empirical relation with deviations increasing towards higher redshift. As discussed in [Spinelli et al. \(2020\)](#) and [Yates et al. \(2021\)](#), this behaviour is at least in part expected, since the H I budget predicted by SAMs only accounts for the gaseous content of galaxies and does not include gas in filaments and in the circum-galactic medium (CGM). The sam2 model qualitatively reproduces the observed decreasing trend below $z \sim 1$, but also deviates from observational measurements at higher redshifts. [Yates et al. \(2021\)](#) showed that this is the case even when accounting for sub-resolution haloes and for missing H I in the IGM (see also [Di Gioia et al., 2020](#)).

The emp1 model also simulates galaxies, but uses an empirical approach. The emp1.1 run shown as a pink triangle in Fig. 1 matches the clustering properties of the H I galaxy population analysed in [Martin et al. \(2012\)](#) which is calibrated on a sub-sample of the α .40 catalogue lacking a large fraction of low-H I-mass sources. For this reason, the total neutral hydrogen budget available from its mock sources is only $\sim 35\%$ of the observational constraint. On the other hand, the deeper emp1.2 run of the same model reaches the same Ω_{HI} value measured on all the other datasets available in the same redshift bins. Indeed, this run is explicitly tuned to reproduce the expected number counts of objects in the local Universe, instead of their 2-point statistics.

For the emp2 method we have access to a light-cone within a narrow redshift interval, $0.77 < z < 1.03$, as it has been originally developed in the context of preparatory studies for HIRAX. Their adopted configuration and mass resolution allows the recovery of more than 97 per cent of the expected H I density. Fig. 1 shows that the cosmic density from emp2 is very close to that predicted by sam2 at $z = 1$ and actually above the fitting formula by [Crighton et al. \(2015\)](#). This difference arises from the used empirical H I-halo mass relation adopted from [Padmanabhan et al. \(2017\)](#), which is characterised by a cut-off at considerably low masses (i.e. $\log M_{\text{HI}}/M_{\odot} \lesssim 10\text{--}11$), due to the inefficiency of H I self-shielding (see Sec. 2.2.2).

3.2 H I mass function

We measure here the H I mass function of galaxies from sam1, sam2 and emp1. We do not include the other models in this comparison because they do not allow to directly extract information on the H I mass function of galaxies from their outputs: even though the emp2 model is based on methodology ([Padmanabhan et al., 2017](#)) that allows to extract the H I mass function from the H I content of haloes, this information is not available for single galaxies. Similarly, emp3 does not natively provide information on individual galaxies but only on the total host halo hydrogen content.

While sams provide a fixed redshift binning, emp1 is defined continuously in redshift, and is upper limited at $z \leq 0.5$. We limited this comparison to 3 redshift slices: the 2 redshift values for which information are available from all the three simulation methods considered here, and redshift $z = 1$ where only results from sam1 and sam2 are available. For the emp1 case, we consider measurement for thin slices of the light-cone, spanning the intervals $z \in [0.0, 0.015)$ and $z \in [0.475, 0.525)$.

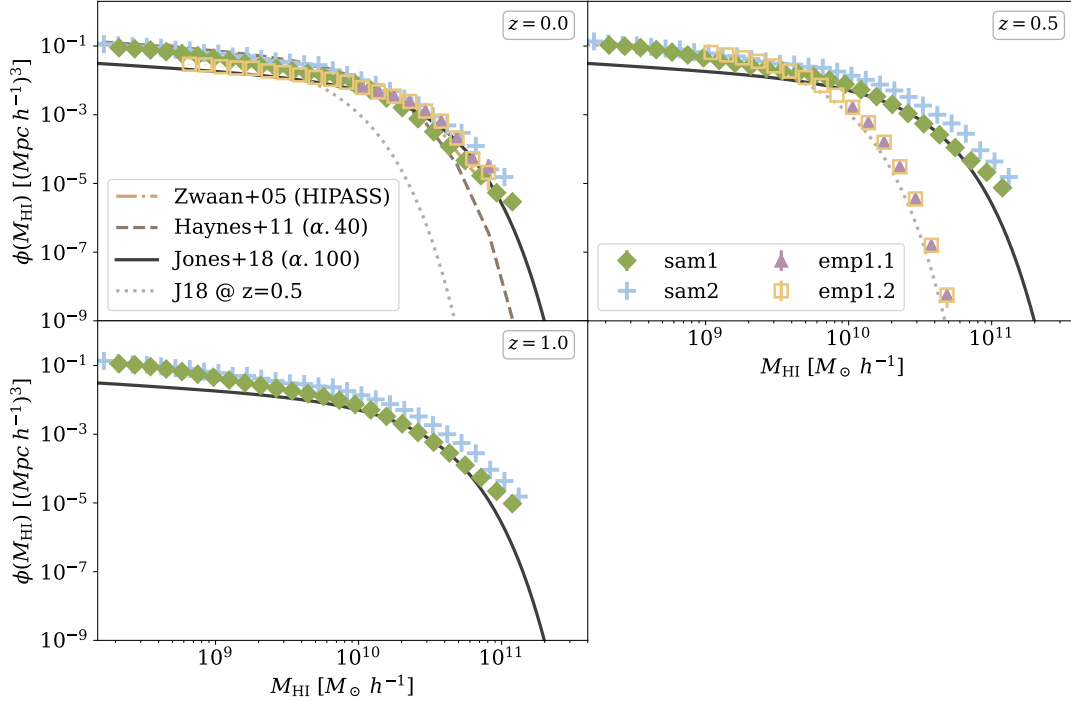


Figure 2: HI mass functions at redshift $z = 0$ (left panel) and $z = 0.5$ (right panel). Measurements from sam1 (green diamonds), sam2 (light-blue plus symbols), emp1.1 (lavender triangles) and emp1.2 (empty golden squares) are shown along with analytical models fitted to data. The models shown in the upper left panel represent fit to different survey data: HIPASS (bronze dashed-dotted line, Zwaan et al., 2005), ALFALFA-40% (gray dashed line, Haynes et al., 2011) and ALFALFA-100% (black solid line Jones et al., 2018). On the 2 upper panels we also show an extrapolation of the redshift-dependent model fitted by Bonaldi et al. (2023) to data from redshift $z \sim 0$ (Jones et al., 2018) to $z \sim 0.3$ (Bera et al., 2022; Paul et al., 2023).

The results of this comparison are shown in the three panels of Fig. 2, for the three redshift values considered. Green diamonds and light-blue plus symbols are for sam1 and sam2 respectively, while the other two markers correspond to the 2 light-cones of the emp1 model: lavender triangles markers are for emp1.1 and empty golden squares for emp1.2.

In the upper left panel, we over-plot different fit to survey data from the literature, as a term of comparison. The bronze dashed-dotted line shows the HIPASS mass function (Zwaan et al., 2005); the gray dashed line corresponds to the results obtained with 40% of the ALFALFA survey (Haynes et al., 2011), while the black solid line shows the most up-to-date estimate from the full ALFALFA survey (Jones et al., 2018, J18 hereafter). Note that all these Schechter-like mass function fits have been obtained mainly with data at redshift $z = 0$. In the upper- and lower-left panels, we only show the most recent J18 fit. Along with these fairly established fitting functions, we also mark the behaviour, extrapolated up to $z = 0.5$, using the redshift-dependent model presented in Bonaldi et al. (2023), tuned to reproduce the limited amount of data available for redshift $z > 0$, namely the Schechter function obtained in Bera et al. (2022) at $z = 0.35$ and the one obtained by indirectly deriving the HIMF from the HI power spectrum at redshifts $z = 0.32$ and $z = 0.44$ (Paul et al., 2023). This HIMF model is the one used in Bonaldi et al. (2023) to sample sources for the T-RECS

mock-catalogues and is thus representative of the simulated dataset of emp1.

At redshift $z = 0$, all methods considered reproduce well the observational data by construction, with an excess of galaxies with H I masses below $\sim 10^{10}$ for the sam1 and sam2 methods. The main difference between the 4 simulations considered is on the H I-mass resolution limit. The sam1 and sam2 models can be considered complete down to $M_{\text{HI}} \gtrsim 10^{8.5} M_{\odot}/h$ and $M_{\text{HI}} > 10^7 M_{\odot}/h$, respectively (see original papers). The emp1 model is limited by construction to higher mass values due to the necessity of reproducing clustering data (i.e., roughly, only sub-haloes with higher masses are selected as galaxy hosts), with emp1.1 complete up to $M_{\text{HI}} \gtrsim 10^{10} M_{\odot}/h$ and emp1.2 $M_{\text{HI}} \gtrsim 10^9 M_{\odot}/h$. In all simulations, the high mass end of the H I mass function is limited by the limited size of the simulated volume.

An important difference between the models considered in this section is visible in the upper right panel of Fig. 2, for the H I mass function at $z = 0.5$. While the sam1 and sam2 models predict very little evolution of the H I mass function up to $z \sim 1$, the emp1 model is tuned to reproduce a strong evolution of the H I mass function between $z = 0$ and $z = 0.5$ (gray dotted line). As mentioned above, little observational information is currently available beyond $z = 0$. However, the situation is rapidly evolving with on-going surveys, such as WALLABY (Koribalski et al., 2020), MIGHTEE-HI (Maddox, N. et al., 2021) and LADUMA (Blyth et al., 2016), and will dramatically change when the SKAO will become operational.

3.3 H I to halo mass relation

The H I to halo mass (HIHM) relation provides insight on the connection between dark matter halos and their baryonic content.

Instead of performing a direct measurement, for sam1 we employed HIP-POP⁴, an approximate method calibrated on sam1 statistics. It therefore encodes the same information as sam1 within the validity range of the fits, namely $\log M_{\text{halo}} \in [9.5, 13.5]$ for halo mass and $z \in [0, 1]$ for redshift. The emp2 method can construct light-cones by adopting the HIHM relation form fitted by Padmanabhan et al. (2017); we therefore use this analytical relation directly for comparison, since the current implementation of emp2 does not include scatter around its functional form and thus coincides exactly with it. For emp1, given the limited depth of the light-cone, we obtained results only at low redshift.⁵

For methods producing galaxy catalogues (i.e. sam1, sam2 and emp1), the HIHM relation is computed by summing the H I masses of all galaxies belonging to a given halo. This approach is prone to underestimates of the total H I content, as neither the faintest galaxies nor the diffuse, unbound neutral hydrogen are included in the total (as already commented in Sec. 3.1).

In Fig. 3 we show, in all panels but the one on the lower right, the distribution of H I mass per halo of given mass as obtained from the catalogues available for sam2 (upper left panel), emp1 (upper right

⁴See the dedicated paragraph in Section 2.1.1. Note that, HIP-HOP has been trained on data from a version of GAEA pre-dating the one used to measure the statistics presented in this Section. Nevertheless, the difference is negligible with respect to the accuracy of our comparative analysis.

⁵We show only the emp1.2 result, as emp1.1 contains the same information at lower resolution.

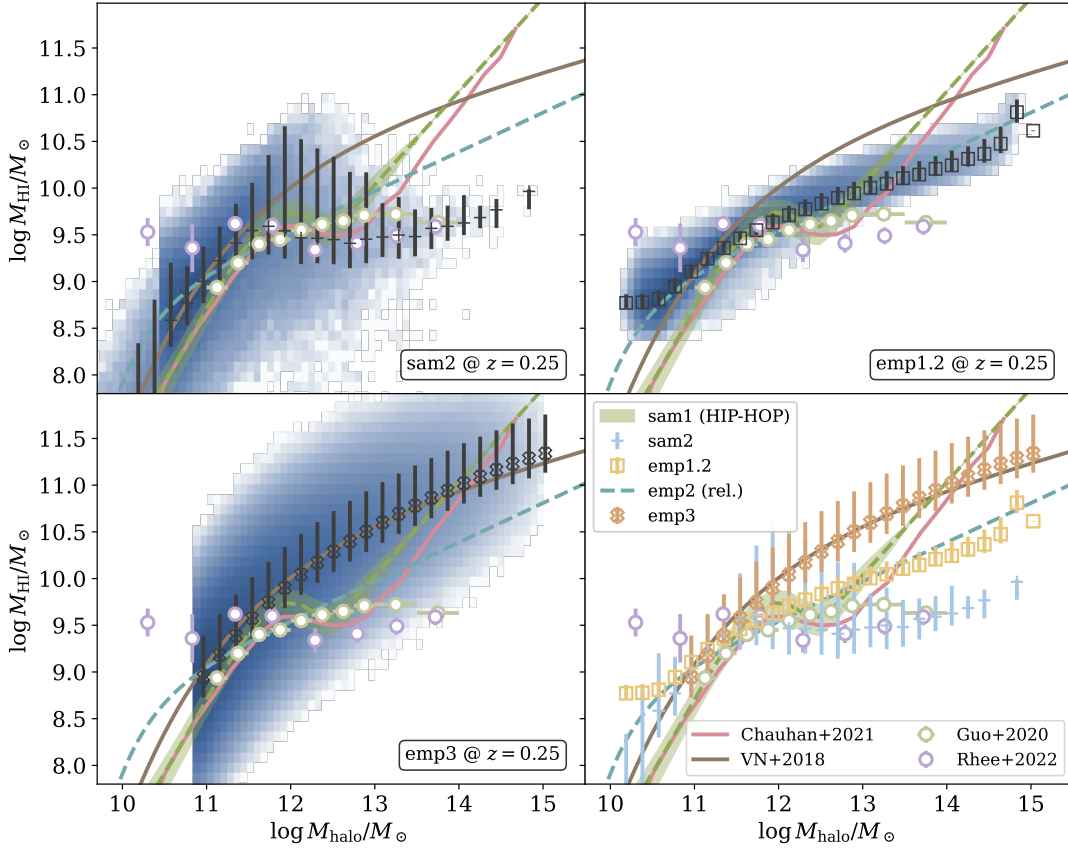


Figure 3: Comparison of the H I –halo mass (HIHM) relation at $z = 0.25$ obtained from all the models considered in this work. The upper panels and lower left panel show the percent abundance of haloes in the $(M_{\text{halo}}, M_{\text{HI}})$ plane for sam2 (upper left), emp1 (upper right) and emp3 (lower left). Blue shaded regions indicate the density of simulated haloes, while empty black squares with error bars mark the mean and standard deviation of $\log M_{\text{HI}}$ in bins of $\log M_{\text{halo}}$. Circles with error bars correspond to observational measurements of central galaxies in clusters (Guo et al., 2020; Rhee et al., 2022), and solid lines represent predictions from hydrodynamical simulations (Villaescusa-Navarro et al., 2018; Chauhan et al., 2021). The lower right panel summarizes all models and reference datasets for direct comparison.

panel) and emp3 (lower left panel). The distributions are shown as blue shaded regions while the empty black squares with vertical error bars indicate the mean and standard deviation of $\log M_{\text{HI}}$ in bins of $\log M_{\text{halo}}$ for each of these distributions. These summary statistics are all shown together in the lower right panel, with the same colouring convention adopted in this Section, for comparison. We also show the results of sam1 and emp2 in all the panels with dashed lines.

As a reference, we also show results from the literature in all the panels: circles with error bars correspond to observational measurements of central galaxies in clusters (Guo et al., 2020; Rhee et al., 2022), while solid lines show predictions from hydrodynamical simulations (Villaescusa-Navarro et al., 2018; Chauhan et al., 2021). The lower right panel provides a direct comparison between all datasets and methods. The semi-analytical models (both sam2 shown in the upper left panel, and sam1 marked by the HIP-HOP fit) show the same dip present in Chauhan et al.

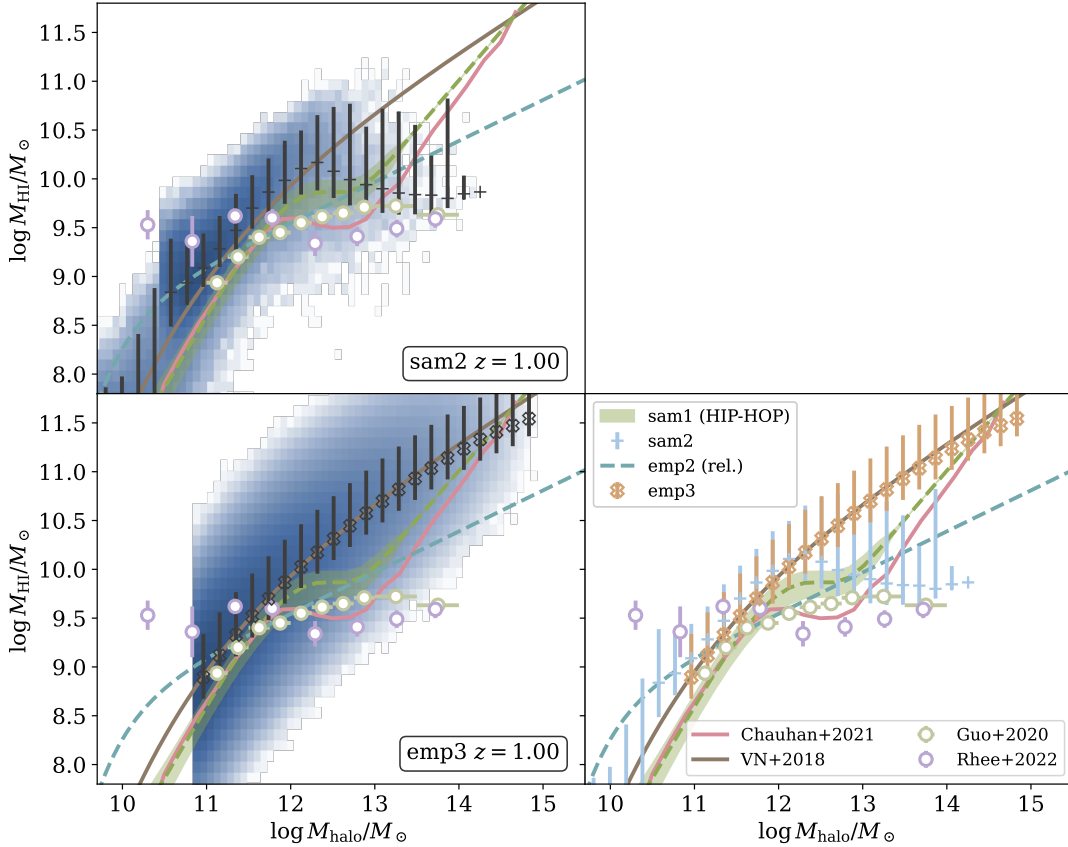


Figure 4: Same as Fig. 3, but for redshift $z = 1$. The upper and lower left panels show results from the sam2 and emp3 models, respectively.

(2021), that is a consequence of AGN feedback. We can also see the transition from the higher-resolution catalogue (run on Millennium-II) to the lower-resolution catalogue (run on Millennium) at $\log(M_{\text{halo}}/M) \sim 10.5$ in the density binning for sam2 in the upper left panel.

The emp1 method (upper right panel) exhibits a normalisation and shape of the relation comparable with observational constraints but no dip. Over the mass range considered, it is also close to the relation based on emp2, marked by the blue dashed line (all panels). However, contrary to emp2, the distribution obtained from emp1 is not the result of an explicit tuning but rather an indirect outcome of the sub-halo abundance-matching scheme. The emp3 model (lower left panel) was tuned to reproduce the relation from Villaescusa-Navarro et al. (2018) and thus, it matches it closely across the full mass range.

In Fig. 4 we present the same comparison as in Fig. 3, at $z \approx 1$. This redshift value is out of the emp1 light-cone and, thus, the corresponding upper right panel is not present. The overall behaviour of the different models considered is the same described for Fig. 3, i.e. the differences between the methods and observational constraints do not appear to change significantly up to $z \sim 1$. We note that there is a fairly good agreement across simulation methodologies and observational constraints at intermediate halo masses, while the results are significantly more scattered at large halo masses

where, due to the limited volumes of current HI surveys, constraints from observational datasets are scarce. Additional observational measurements (both in the radio and in the optical wavelength regime) will enable a better characterization of the HIHM relation, both for the most massive haloes and at halo masses $\sim 10^{12} M_{\odot}$, where observational uncertainties on halo mass affect the strength of the dip that some models predict.

3.4 Marked power spectrum

The marked power spectrum is a weighted generalisation of the standard two-point clustering statistic, designed to probe correlations between large-scale structure and a local property (the *mark*) of the underlying field (Sheth, 2005; White, 2016; Satpathy et al., 2019). Given a tracer density field $n(\mathbf{x})$ and an associated scalar mark $m(\mathbf{x})$, the marked density contrast is defined as

$$\delta_m(\mathbf{x}) = \frac{m(\mathbf{x}) n(\mathbf{x})}{\langle m n \rangle} - 1, \quad (5)$$

where $\langle \cdot \rangle$ denotes the spatial mean. The *marked power spectrum* is then the Fourier-space two-point statistic of this marked field:

$$P_m(k) = \langle \tilde{\delta}_m(\mathbf{k}) \tilde{\delta}_m^*(\mathbf{k}) \rangle, \quad (6)$$

that reduces to the standard matter power spectrum $P(k)$ when the mark is $m(\mathbf{x}) = 1$ everywhere.

The $P_m(k)$ is sensitive to how clustering depends on the chosen mark, which can encode non-linear or environmental effects. Common choices include marks that depend on the local density, halo mass, or other internal properties such as the stellar mass, the gas fraction, or specifically in our case, the HI mass content of the objects studied. In this way, the marked spectrum traces the coupling between large-scale modes and small-scale physics, extending beyond what is captured by $P(k)$ alone.

For our simulated samples, we define the mark as a function of the neutral hydrogen mass M_{HI} (e.g. Massara et al., 2021). We compare results from the $z = 1$ snapshots of sam1, emp2 and emp3 (note that, for sam1, measurements of $P_m(k)$ are also made available in Spinelli et al., 2020). In all cases⁶ model results are shown only for $k_{\text{min}} < k < 12k_{\text{Nyq}}$, where k_{min} is determined by the simulation size, while the k_{Nyq} is the Nyquist frequency of the simulation, i.e. the mean inter-tracer separation in real space. The black line in Fig. 5 is the linear $P_{\text{lin}}(k, z = 0)$ of matter at redshift $z = 0$ evolved up to $z = 1$ with the growth factor (i.e. multiplying by $[D(z = 1)/D(z = 0)]^2$). It has been obtained by running CAMB with approximated Planck2018 parameters (Planck Collaboration et al., 2020). The growth factor is computed with the `scampy.cosmology` module of the SCAMPy (Ronconi et al., 2020) library, using rounded Planck2018 parameters (extracted from `astropy.cosmology.Planck2018` for consistency with the rest of the measurements).

All the methods show a slight loss in power for the largest scales (lowest k values) with respect to the linearly evolved $P(k)$. The different values of k_{min} here reflect the different sizes of the background DM simulations used by the authors. In particular, emp2 and emp3 use approximated methods to build halo catalogues, that allow lower values of k_{min} . On the other hand, sam1 computes the $P_m(k)$ from the galaxy distribution instead of haloes, which results in a higher value of k_{Nyq} . Besides

⁶We have used estimators consistent with those available in the Pylians library (pylians3.readthedocs.io).

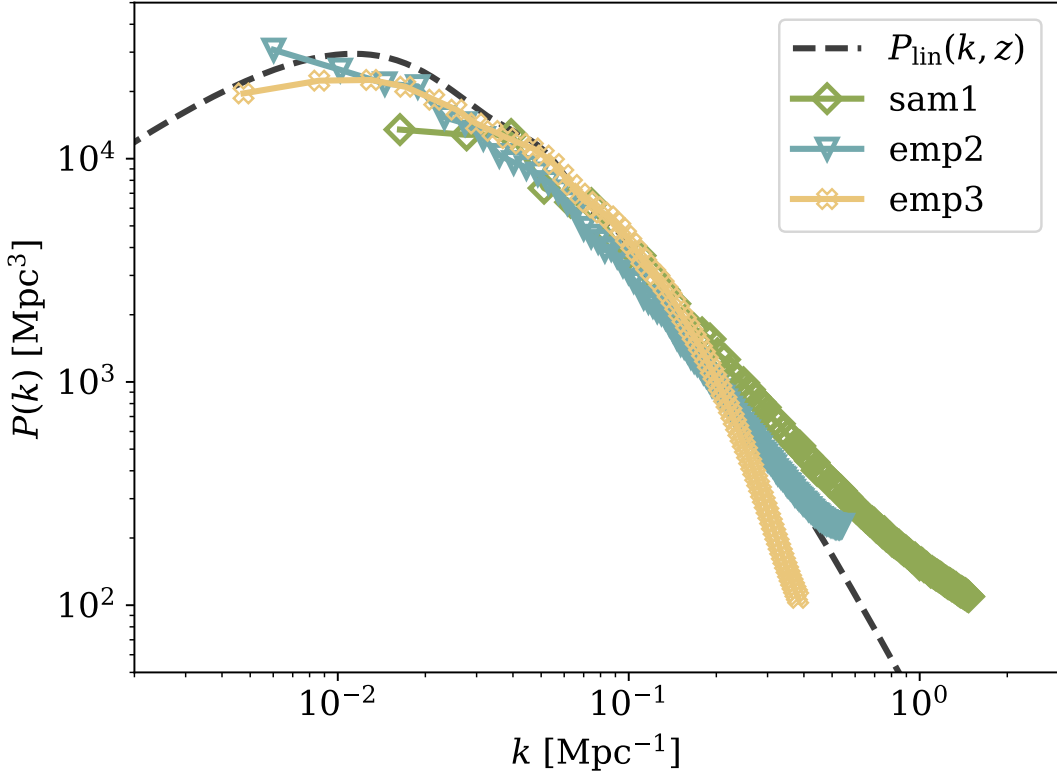


Figure 5: Marked power spectrum measured on the redshift $z = 1$ boxes of sam1 (green diamonds connected by a green solid line), emp2 (blue downward triangles connected by blue solid line) and emp3 (orange crosses connected by the orange solid line). As a term of comparison, we over-plot as a black dashed line, the linear power spectrum of matter evolved linearly up to the same redshift of the simulations.

this, galaxies are expected to be more affected by non-linear effects than dark matter haloes, as reflected by the excess power of the green diamonds at $k > 0.1$ compared to both the empirical methods and the linear case. The lower amplitude of $P_{\text{HI}}(k)$ measured from emp2 with respect to the linear (matter) case, together with the fact that this simulation is nearly complete in terms of the cosmological neutral hydrogen budget, is consistent with a low H I bias, $b_{\text{HI}} \lesssim 1$, i.e. H I behaves as an almost unbiased tracer of the underlying dark-matter density field at these redshifts (see e.g. [Castorina and Villaescusa-Navarro 2017](#); [Villaescusa-Navarro et al. 2018](#); [Obuljen et al. 2023](#)). On the other hand, the abrupt loss of power of emp3, at scales $k > 0.2 \text{ Mpc}^{-1}$, is a numerical effect and is most likely due to aliasing generating from the coarser interpolation mesh used to perform the measurement.

Overall, the three methods analysed here show consistent results on scales $-2 \lesssim \log(k/\text{Mpc}^{-1}) \lesssim -1$. At lower k -values, differences are mostly driven by cosmic variance, while at larger k -values the scatter is less straightforward to attribute to a single cause. Instead, it likely arises from a combination of factors, among which resolution and the nature of the modelled objects play a principal role.

4 Context within the SKAO observational capabilities

The models described in this Chapter are developed in the context of forthcoming observations with the SKAO, whose capabilities will open several complementary routes to probe neutral hydrogen across cosmic time. In the post-reionization regime, and especially for *SKA-Mid*, simulated H I maps and catalogues are not only a forecasting tool, but also the main bridge between physical models of how H I populates haloes and galaxies and the survey observables that will be measured. This connection is particularly important because different observing strategies constrain different parts of the modelling, rather than all aspects equally well.

The telescopes of the SKAO, *SKA-Mid* and *SKA-Low*, probe complementary redshift ranges and physical regimes. *SKA-Mid* is the key facility for the post-EoR applications discussed in this Chapter, since it will probe low to intermediate redshifts, roughly $z \lesssim 3$, through both resolved H I galaxy surveys and H I intensity mapping. By contrast, *SKA-Low* is aimed mainly at higher redshifts and the EoR, where semi-numerical methods (such as 21cmFAST, Mesinger et al., 2011; Murray et al., 2020) remain the natural tool for forecasting the large-scale 21-cm signal. In the *SKA-Mid* context, the simulation methods reviewed here are especially well matched to survey design because they link the H I content of galaxies and haloes to observable quantities such as number counts and clustering (see Sec. 3), and, with further post-processing brightness-temperature fluctuations and linewidths (e.g. Mayor et al., 2026; Nasirudin et al., 2026).

Resolved H I galaxy surveys will be particularly valuable for constraining the ingredients of semi-analytic and empirical models that govern the detectable galaxy population. The most directly testable components are the halo occupation of H I galaxies (Sec. 3.3, available for all the models presented here), the redshift evolution of the H I mass function (Sec. 3.2, available for sam1, sam2 and emp1), and flux selection effects which vary model by model. The comparison among different modelling approaches detailed in this Chapter shows that predicted H I galaxy number counts and large-scale bias can differ by factors of a few (Sec. 3.4), and in some cases by nearly an order of magnitude at higher redshift (as evident from the comparison in Sec. 3.1), especially in the regime relevant to *SKA-Mid* Band 1 surveys (see also Nasirudin et al., 2026). This means that future H I galaxy observations with *SKA-Mid* will be especially effective at testing how different models assign H I to galaxies and haloes, and at identifying which prescriptions should be tuned to best reproduce the abundance and clustering of the detectable populations.

Intensity mapping (see Wolz et al. (2026) for a review), targeted by emp2 (Sec. 2.2.2) and emp3 (Sec. 2.2.3) models directly, constrains a different layer of the same problem. On large scales, single-dish H I intensity mapping is primarily sensitive to the combination of the cosmic H I abundance and its large-scale bias, making it especially powerful for baryon acoustic oscillations, redshift-space distortions, and other large-volume cosmological measurements (Santos et al., 2015; Spinelli et al., 2020). Because it captures the integrated emission from the full galaxy population, including sources far below the detection threshold, it provides a direct handle on the global H I field rather than only on individually resolved systems. However, large-scale intensity mapping alone is less sensitive to the detailed internal distribution of H I within haloes. Those aspects are more effectively constrained by interferometric observing mode, which reaches smaller scales and can probe the one-halo regime, where the signal depends on the H I-halo mass relation, the radial profile of H I

within haloes, and the level of shot noise and halo-to-halo scatter (Chen et al., 2021; Spinelli et al., 2020; Villaescusa-Navarro et al., 2018; Hitz et al., 2025). In this sense, single-dish and interferometric intensity mapping play distinct but complementary roles: the former is best suited to large-scale cosmology, while the latter is more informative about the astrophysical structure of the H_I field itself.

Different telescope configurations and deployment phases (e.g. baseline and upgraded array setups) will impact the achievable sensitivity, angular resolution, and survey speed, and thus determine the range of scales and redshifts that can be effectively probed. From the modelling side, however, the main lesson is that the approaches reviewed in this Chapter should be seen not simply as alternative simulation methods, but as a complementary toolkit whose components are constrained by different classes of SKAO observations. Physically motivated models remain essential for connecting observables to the underlying baryonic processes that regulate the H_I content of galaxies and haloes (see also Lagos et al., 2026), while empirical and fast approximate methods are crucial for producing the large ensembles of mock realizations needed for survey optimization, covariance estimation, and end-to-end pipeline validation (e.g. SKA Cosmology SWG et al., 2020; Hartley et al., 2023; Bonaldi et al., 2023).

A key aspect of the interplay between simulations and observations is the forward modelling of instrumental effects. Realistic H_I simulations, as those presented in this Chapter, can be processed through end-to-end pipelines that include beam convolution, noise realisations, survey masks, and foreground contamination, allowing direct comparison with expected observables such as angular and three-dimensional power spectra (e.g. Harrison et al., 2026a,b). This approach is essential for assessing the impact of systematics on cosmological measurements, including baryon acoustic oscillations, redshift-space distortions, and higher-order statistics.

The methodologies presented in this Chapter are highly needed to exploit the full potential of the SKAO, as demonstrated by the many Chapters in this book that make direct use of simulated data or of closely related techniques (including: Bertacca et al., 2026; Asorey&Hale et al., 2026; Harrison et al., 2026a,b; Wolz et al., 2026; Cunnington&Wang et al., 2026; Majumdar et al., 2026; Mayor et al., 2026; Nasirudin et al., 2026; Baker et al., 2026; Camera et al., 2026; Spinelli et al., 2026). The simulation strategies discussed here are broadly applicable across different array-assembly phases of the Observatory, and can be adapted to both AA* and AA4. In practice, the main telescope specifications that affect the simulation design are the assumed survey area, depth, angular resolution, and limiting flux, many of which remain to be finalized at the time of writing. Some recent works (e.g. Ronconi et al., 2026) adopt survey assumptions motivated by proposed SKAO configurations (SKA Cosmology SWG et al., 2020; Hartley et al., 2023); however, such choices should still be regarded as indicative, since the final observing strategies will depend on the approved key science programmes and the capabilities available in the corresponding deployment phase. By enabling robust predictions for H_I observables across different regimes, simulations play, and will continue to play, a central role in bridging theoretical modelling and observational analysis, both by enabling forecasts for a wide range of science cases and by supporting the interpretation of forthcoming data. More generally, modelling the expected observables is one of the main tools through which the survey strategy itself can be optimized, by identifying which combinations of

area, depth, and observing mode are best suited to maximize the scientific return of the Observatory.

5 Discussion and future prospects

The comparison presented in this Chapter highlights the diversity of existing approaches used to connect dark matter structures with the expected neutral hydrogen content in the Universe. The analysed methods differ significantly in their physical assumptions and calibration strategy. Furthermore, while some methods provide information for individual galaxies, others only give halo-integrated quantities, making a one-to-one comparison more difficult for some observables.

Overall, the various models provide qualitatively consistent correlations between the H_I and halo mass and a quite good agreement for integrated quantities, such as Ω_{HI} (Sec. 3.1) or the mean H_I bias (Sec. 3.4), suggesting that large-scale statistics are relatively robust against the choice of modelling framework. At least part of this agreement is driven by the fact that models are tuned to reproduce the same observational data obtained in the local Universe. As a consequence, larger differences are found when considering model prediction at higher redshifts where observational constraints are more scarce or sometimes based on different tracers (as is the case for the cosmic density of H_I).

Systematic differences emerge when comparing the detailed shape and scatter of the HIHM relation or the evolution of the H_I distribution with redshift. Part of these differences can be traced back to the intrinsic nature of the models. Semi-analytical approaches, such as *sam1* (Sec. 2.1.1) and *sam2* (Sec. 2.1.2), self-consistently follow baryonic processes and remain sensitive to the adopted prescriptions. Empirical methods, on the other hand, are computationally efficient and directly calibrated on observations (as in *emp1* from Sec. 2.2.1 and *emp2* from Sec. 2.2.2) or hydrodynamical predictions (as in *emp3* from Sec. 2.2.3).

In general, any behaviour in regimes (in terms of e.g. halo mass or redshift) beyond the calibration range considered, are extrapolations that should be considered with caution. The different nature of the models, and the weaknesses and advantages briefly discussed, naturally lead to a different tool selection when addressing specific questions/problems. Clearly, when developing specific algorithms for data analysis, empirical methods provide an efficient and natural framework for testing in realistic cases, at least over the redshift range where observational data are available to construct the models. However, it is difficult to translate the information obtained employing statistical methods into constraints on physical processes driving baryonic evolution unless explicitly accounting for them, like done for example in semi-analytic models. As mentioned above, the observational information currently available beyond $z = 0$ is scarce but the situation is rapidly evolving with ongoing survey efforts and will be ultimately revolutionized as the SKAO will move into operation. The tools we have discussed will play a crucial role for the exploitation of the huge amount of information that will be collected in the coming years.

Beyond the analysis presented here, all methods described in this work can be post-processed to derive the corresponding H_I brightness temperature fields enabling additional applications, for example, in the framework of intensity mapping studies (e.g. [Spinelli et al., 2026](#)), by calculating the 21cm emission lines ([Mayor et al., 2026](#)). Furthermore, at least some of the methods discussed

in this Chapter, also predict many other additional properties, beyond masses and positions. For example, both the semi-analytic models (Sec. 2.1) provide information on AGN feedback, gas partitioning and evolutionary properties, useful for understanding, e.g., the formation processes of the sources modelled. The HIHM-relation-based model (Sec. 2.2.3), provides environmental information and, the T-RECS-based method (Sec. 2.2.1), also includes emission properties for other populations of sources, providing a natural framework for, e.g., cross-correlation studies.

The diversity in outcomes, far from being a limitation, illustrates the sensitivity of H_I modelling to both physical assumptions and methodological choices. Together, these results emphasise the importance of maintaining a heterogeneous modelling ecosystem. Each approach captures complementary aspects of the complex connection between dark matter and baryons. Different scientific problems call for different modelling strategies rather than a single "best" prescription: one algorithm may excel for halo-scale H_I assignments, another for large-scale intensity mapping, and another for hybrid clustering or cross-correlation forecasts. This Chapter's objective has been to illustrate this methodological diversity and to inform the reader about the regimes in which each model is likely to perform more reliably.

Looking ahead, the new generation of H_I surveys, including ASKAP's WALLABY (mapping up to $z \sim 0.1$ over $\sim 14\,000 \text{ deg}^2$ with $\sim 30''$ resolution, [Koribalski et al., 2020](#)), and deep surveys with MeerKAT such as MIGHTEE-HI and LADUMA (probing to $z \sim 0.3$ and $z \sim 1.4$, respectively, [Maddox, N. et al., 2021](#); [Blyth et al., 2016](#)), will yield independent observational constraints on the neutral hydrogen distribution. These upcoming data will allow modelling paradigms, such as those presented in the previous Sections, to be recalibrated and validated in preparation of the SKAO era. By benchmarking models against real survey outputs, it will be possible to gradually refine parameter choices, to reduce systematic uncertainties, and to build a more robust interpretative framework for SKAO's unprecedented H_I survey capabilities.

Author List Ordering

The authors of this Chapter are grouped into three tiers reflecting their contribution. The first tier, ordered by contribution, includes TR, GDL, and MS, who led the conceptual development and writing of this work. The second tier (FF, PH, FS, LX, RMY), ordered alphabetically, includes authors who contributed actively through the production and analysis of the simulation results presented. The third tier, also ordered alphabetically, includes authors who participated in the scientific discussion, interpretation of results, and revision of the manuscript.

Acknowledgements

PH is partially supported by grant CRSII5_193826 from the Swiss National Science Foundation, and by SERI as part of the SKACH consortium. JLB acknowledges funding from the grant UC-LIME (PID2022-140670NA-I00), financed by MCIN/AEI/ 10.13039/501100011033/FEDER, UE.

References

- O. Agertz et al. *MNRAS*, 491(2):1656–1672, Jan. 2020. doi: 10.1093/mnras/stz3053.
- A. Agrawal et al. *JCAP*, 10:003, 2017. doi: 10.1088/1475-7516/2017/10/003.
- R. E. Angulo and S. D. M. White. *MNRAS*, 405(1):143–154, June 2010. doi: 10.1111/j.1365-2966.2010.16459.x.
- J. Asorey&Hale et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Asorey01.
- J. S. Bagla, N. Khandai, and K. K. Datta. *Monthly Notices of the Royal Astronomical Society*, 407(1):567–580, 08 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.16933.x. URL <https://doi.org/10.1111/j.1365-2966.2010.16933.x>.
- T. Baker et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Baker01.
- L. A. Barnes and M. G. Haehnelt. *Monthly Notices of the Royal Astronomical Society*, 440(3): 2313–2321, 04 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu445. URL <https://doi.org/10.1093/mnras/stu445>.
- C. M. Baugh et al. *MNRAS*, 483(4):4922–4937, Mar. 2019. doi: 10.1093/mnras/sty3427.
- P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy. *Mon. Not. Roy. Astron. Soc.*, 488(3): 3143–3194, 2019. doi: 10.1093/mnras/stz1182.
- A. Bera, N. Kanekar, J. N. Chengalur, and J. S. Bagla. *The Astrophysical Journal Letters*, 940 (1):L10, nov 2022. doi: 10.3847/2041-8213/ac9d32. URL <https://doi.org/10.3847/2041-8213/ac9d32>.
- D. Bertacca et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Bertacca01.
- L. Blitz and E. Rosolowsky. *ApJ*, 650(2):933–944, Oct. 2006. doi: 10.1086/505417.
- S. Blyth et al. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 128, Apr. 2015. doi: 10.22323/1.215.0128.
- S. Blyth et al. In *MeerKAT Science: On the Pathway to the SKA*, page 4, Jan. 2016. doi: 10.22323/1.277.0004.
- A. Bonaldi et al. *Monthly Notices of the Royal Astronomical Society*, 482(1):2–19, 09 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty2603. URL <https://doi.org/10.1093/mnras/sty2603>.
- A. Bonaldi et al. *Monthly Notices of the Royal Astronomical Society*, 524(1):993–1007, 06 2023. ISSN 0035-8711. doi: 10.1093/mnras/stad1913. URL <https://doi.org/10.1093/mnras/stad1913>.
- J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. *ApJ*, 379:440, Oct. 1991. doi: 10.1086/170520.
- E. G. P. Bos, F.-S. Kitaura, and R. van de Weygaert. *MNRAS*, 488(2):2573–2604, Sept. 2019. doi: 10.1093/mnras/stz1864.
- M. Boylan-Kolchin et al. *MNRAS*, 398(3):1150–1164, Sept. 2009. doi: 10.1111/j.1365-2966.2009.15191.x.
- R. Braun et al. Anticipated performance of the square kilometre array – phase 1 (ska1), 2019. URL <https://arxiv.org/abs/1912.12699>.
- P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos. *The Astrophysical Journal*, 803(1):21, apr 2015.

- doi: 10.1088/0004-637X/803/1/21. URL <https://dx.doi.org/10.1088/0004-637X/803/1/21>.
- S. Camera et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Camera01.
- V. Cammelli et al. *MNRAS*, 536(1):851–870, Jan. 2025. doi: 10.1093/mnras/stae2663.
- C. Carbone, M. Petkova, and K. Dolag. *Journal of Cosmology and Astroparticle Physics*, 2016(07):034, jul 2016. doi: 10.1088/1475-7516/2016/07/034. URL <https://dx.doi.org/10.1088/1475-7516/2016/07/034>.
- E. Castorina and F. Villaescusa-Navarro. *MNRAS*, 471(2):1788–1796, Oct. 2017. doi: 10.1093/mnras/stx1599.
- T.-C. Chang, U.-L. Pen, K. Bandura, and J. B. Peterson. *Nature*, 466(7305):463–465, July 2010. doi: 10.1038/nature09187.
- G. Chauhan et al. *Monthly Notices of the Royal Astronomical Society*, 506(4):4893–4913, 07 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab1925. URL <https://doi.org/10.1093/mnras/stab1925>.
- Z. Chen, L. Wolz, M. Spinelli, and S. G. Murray. *Monthly Notices of the Royal Astronomical Society*, 502(4):5259–5276, 04 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab386. URL <https://doi.org/10.1093/mnras/stab386>.
- J. M. Coloma-Nadal et al. *arXiv e-prints*, art. arXiv:2403.19337, Mar. 2024. doi: 10.48550/arXiv.2403.19337.
- R. A. Crain et al. *MNRAS*, 450(2):1937–1961, June 2015. doi: 10.1093/mnras/stv725.
- R. A. Crain et al. *MNRAS*, 464(4):4204–4226, Feb. 2017. doi: 10.1093/mnras/stw2586.
- D. Crichton, M. Aich, and e. a. Adam Amara. *Journal of Astronomical Telescopes, Instruments, and Systems*, 8(1):011019, 2022. doi: 10.1117/1.JATIS.8.1.011019. URL <https://doi.org/10.1117/1.JATIS.8.1.011019>.
- N. H. M. Crighton et al. *Monthly Notices of the Royal Astronomical Society*, 452(1):217–234, 07 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv1182. URL <https://doi.org/10.1093/mnras/stv1182>.
- D. J. Croton et al. *MNRAS*, 365(1):11–28, Jan. 2006. doi: 10.1111/j.1365-2966.2005.09675.x.
- S. Cunnington et al. *Mon. Not. Roy. Astron. Soc.*, 523(2):2453–2477, 2023. doi: 10.1093/mnras/stad1567.
- S. Cunnington&Wang et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Cunnington01.
- G. De Lucia and J. Blaizot. *MNRAS*, 375(1):2–14, Feb. 2007. doi: 10.1111/j.1365-2966.2006.11287.x.
- G. De Lucia, G. Kauffmann, and S. D. M. White. *MNRAS*, 349(3):1101–1116, Apr. 2004. doi: 10.1111/j.1365-2966.2004.07584.x.
- G. De Lucia et al. *MNRAS*, 445(1):970–987, Nov. 2014. doi: 10.1093/mnras/stu1752.
- G. De Lucia, F. Fontanot, L. Xie, and M. Hirschmann. *A&A*, 687:A68, July 2024. doi: 10.1051/0004-6361/202349045.
- S. Di Gioia, S. Cristiani, G. De Lucia, and L. Xie. *MNRAS*, 497(2):2469–2485, Sept. 2020. doi: 10.1093/mnras/staa2067.
- B. Diemer et al. *MNRAS*, 487(2):1529–1550, Aug. 2019. doi: 10.1093/mnras/stz1323.

- R. Feldmann et al. *MNRAS*, 522(3):3831–3860, July 2023. doi: 10.1093/mnras/stad1205.
- F. Feroz, M. P. Hobson, and M. Bridges. *MNRAS*, 398:1601–1614, Oct. 2009. doi: 10.1111/j.1365-2966.2009.14548.x.
- F. Fontanot et al. *MNRAS*, 496(3):3943–3960, Aug. 2020. doi: 10.1093/mnras/staa1716.
- F. Fontanot et al. *A&A*, 699:A108, July 2025. doi: 10.1051/0004-6361/202452029.
- D. Forero Sánchez et al. *JCAP*, 2024(7):001, July 2024. doi: 10.1088/1475-7516/2024/07/001.
- J. Fu et al. *MNRAS*, 434(2):1531–1548, Sep 2013. doi: 10.1093/mnras/stt1117.
- S. R. Furlanetto, S. Peng Oh, and F. H. Briggs. *Physics Reports*, 433(4):181–301, 2006. ISSN 0370-1573. doi: <https://doi.org/10.1016/j.physrep.2006.08.002>. URL <https://www.sciencedirect.com/science/article/pii/S0370157306002730>.
- L. H. Garrison, D. J. Eisenstein, and P. A. Pinto. *MNRAS*, 485(3):3370–3377, May 2019. doi: 10.1093/mnras/stz634.
- L. H. Garrison et al. *MNRAS*, 508(1):575–596, Nov. 2021. doi: 10.1093/mnras/stab2482.
- H. Guo et al. *Monthly Notices of the Royal Astronomical Society*, 459(3):3040–3058, 04 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw845. URL <https://doi.org/10.1093/mnras/stw845>.
- H. Guo, M. G. Jones, M. P. Haynes, and J. Fu. *The Astrophysical Journal*, 894(2):92, may 2020. doi: 10.3847/1538-4357/ab886f. URL <https://doi.org/10.3847/1538-4357/ab886f>.
- Q. Guo et al. 413:101–131, May 2011. doi: 10.1111/j.1365-2966.2010.18114.x.
- O. Hahn, C. Porciani, C. M. Carollo, and A. Dekel. *Monthly Notices of the Royal Astronomical Society*, 375(2):489–499, feb 2007. doi: 10.1111/j.1365-2966.2006.11318.x. URL <https://doi.org/10.1111%2Fj.1365-2966.2006.11318.x>.
- C. L. Hale et al. *Monthly Notices of the Royal Astronomical Society*, 474(3):4133–4150, 11 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx2954. URL <https://doi.org/10.1093/mnras/stx2954>.
- I. Harrison et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026a. arXiv search: Report number AASKAII/Harrison01.
- I. Harrison et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026b. arXiv search: Report number AASKAII/Harrison02.
- P. Hartley et al. *Monthly Notices of the Royal Astronomical Society*, 523(2):1967–1993, 05 2023. ISSN 0035-8711. doi: 10.1093/mnras/stad1375. URL <https://doi.org/10.1093/mnras/stad1375>.
- M. P. Haynes et al. *AJ*, 142(5):170, Nov. 2011. doi: 10.1088/0004-6256/142/5/170.
- A. Heavens and J. Peacock. *Monthly Notices of the Royal Astronomical Society*, 232(2):339–360, 05 1988. ISSN 0035-8711. doi: 10.1093/mnras/232.2.339. URL <https://doi.org/10.1093/mnras/232.2.339>.
- B. M. B. Henriques et al. 451:2663–2680, Aug. 2015. doi: 10.1093/mnras/stv705.
- B. M. B. Henriques et al. *MNRAS*, 491(4):5795–5814, Feb. 2020. doi: 10.1093/mnras/stz3233.
- B. Hernández-Molinero, C. Carbone, R. Jimenez, and C. P. Garay. *Journal of Cosmology and Astroparticle Physics*, 2024(01):006, jan 2024. doi: 10.1088/1475-7516/2024/01/006. URL <https://dx.doi.org/10.1088/1475-7516/2024/01/006>.
- S. Heß, F.-S. Kitaura, and S. Gottlöber. *MNRAS*, 435(3):2065–2076, Nov. 2013. doi: 10.1093/mnras/stt1428.
- M. Hirschmann, G. De Lucia, and F. Fontanot. *MNRAS*, 461(2):1760–1785, Sept. 2016. doi:

- 10.1093/mnras/stw1318.
- P. Hitz et al. *Journal of Cosmology and Astroparticle Physics*, 2025(04):003, Apr 2025. doi: 10.1088/1475-7516/2025/04/003. URL <https://dx.doi.org/10.1088/1475-7516/2025/04/003>.
- M. G. Jones, M. P. Haynes, R. Giovanelli, and C. Moorman. *MNRAS*, 477(1):2–17, June 2018. doi: 10.1093/mnras/sty521.
- N. Kaiser. *MNRAS*, 227:1–21, July 1987. doi: 10.1093/mnras/227.1.1.
- F. S. Kitaura and S. Hess. *MNRAS*, 435:L78–L82, Aug. 2013. doi: 10.1093/mnras/slt101.
- F.-S. Kitaura and F. Sinigaglia. *arXiv e-prints*, art. arXiv:2603.15834, Mar. 2026. doi: 10.48550/arXiv.2603.15834.
- F.-S. Kitaura et al. *MNRAS*, 427(1):L35–L39, Nov. 2012a. doi: 10.1111/j.1745-3933.2012.01340.x.
- F.-S. Kitaura, S. Gallerani, and A. Ferrara. *MNRAS*, 420(1):61–74, Feb. 2012b. doi: 10.1111/j.1365-2966.2011.19997.x.
- F. S. Kitaura, G. Yepes, and F. Prada. *MNRAS*, 439:L21–L25, Mar. 2014. doi: 10.1093/mnras/slt172.
- F.-S. Kitaura et al. *MNRAS*, 457(1):L113–L117, Mar. 2016. doi: 10.1093/mnras/slw009.
- F. S. Kitaura, A. Balaguera-Antolínez, F. Sinigaglia, and M. Pellejero-Ibáñez. *MNRAS*, 512(2): 2245–2265, May 2022. doi: 10.1093/mnras/stac671.
- B. S. Koribalski et al. *Astro. & Space Sci.*, 365(7):118, July 2020. doi: 10.1007/s10509-020-03831-4.
- C. D. Lagos et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Lagos01.
- M. Lujan Niemeyer, J. L. Bernal, and E. Komatsu. *Astrophys. J.*, 958(1):4, 2023. doi: 10.3847/1538-4357/acfef4.
- Maddox, N. et al. *A&A*, 646:A35, 2021. doi: 10.1051/0004-6361/202039655. URL <https://doi.org/10.1051/0004-6361/202039655>.
- S. Majumdar et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Majumdar01.
- N. A. Maksimova et al. *MNRAS*, 508(3):4017–4037, Dec. 2021. doi: 10.1093/mnras/stab2484.
- A. M. Martin, R. Giovanelli, M. P. Haynes, and L. Guzzo. *The Astrophysical Journal*, 750(1): 38, apr 2012. doi: 10.1088/0004-637X/750/1/38. URL <https://dx.doi.org/10.1088/0004-637X/750/1/38>.
- E. Massara et al. *Phys. Rev. Lett.*, 126(1):011301, Jan. 2021. doi: 10.1103/PhysRevLett.126.011301.
- K. W. Masui. *Advancing precision cosmology with 21 cm intensity mapping*. PhD thesis, University of Toronto, Canada, Jan. 2013.
- J. Mayor et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Mayor01.
- J. Mayor et al. Simulations of the 21cm emission line for upcoming large-scale hi galaxy surveys, 2026. URL <https://arxiv.org/abs/2602.21058>.
- A. Mazumder et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Mazumder01.
- P. McDonald and A. Roy. *Journal of Cosmology and Astroparticle Physics*, 2009(08):020–020, aug 2009. doi: 10.1088/1475-7516/2009/08/020. URL <https://doi.org/10.1088/1475-7516/2009/08/020>.

- 2F1475–7516%2F2009%2F08%2F020.
- C. F. McKee and M. R. Krumholz. *ApJ*, 709(1):308–320, Jan. 2010. doi: 10.1088/0004-637X/709/1/308.
- MeerKLASS Collaboration et al. *Mon. Not. Roy. Astron. Soc.*, 537(4):3632–3661, 2025. doi: 10.1093/mnras/staf195.
- A. Mesinger, S. Furlanetto, and R. Cen. *Monthly Notices of the Royal Astronomical Society*, 411(2):955–972, 02 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.17731.x. URL <https://doi.org/10.1111/j.1365-2966.2010.17731.x>.
- P. Monaco. *ApJ*, 447:23, July 1995. doi: 10.1086/175853.
- P. Monaco. *Monthly Notices of the Royal Astronomical Society*, 290(3):439–455, 09 1997. ISSN 0035-8711. doi: 10.1093/mnras/290.3.439. URL <https://doi.org/10.1093/mnras/290.3.439>.
- P. Monaco. *Galaxies*, 4(4):53, Oct. 2016. doi: 10.3390/galaxies4040053.
- P. Monaco, T. Theuns, and G. Taffoni. *Monthly Notices of the Royal Astronomical Society*, 331(3): 587–608, 04 2002. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2002.05162.x. URL <https://doi.org/10.1046/j.1365-8711.2002.05162.x>.
- R. Mondal, S. Bharadwaj, and S. Majumdar. *Monthly Notices of the Royal Astronomical Society*, 456(2):1936–1947, 12 2015a. ISSN 0035-8711. doi: 10.1093/mnras/stv2772. URL <https://doi.org/10.1093/mnras/stv2772>.
- R. Mondal et al. *Monthly Notices of the Royal Astronomical Society: Letters*, 449(1):L41–L45, 02 2015b. ISSN 1745-3925. doi: 10.1093/mnrasl/slv015. URL <https://doi.org/10.1093/mnrasl/slv015>.
- R. Mondal, S. Bharadwaj, and K. K. Datta. *Monthly Notices of the Royal Astronomical Society*, 474(1):1390–1397, 11 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx2888. URL <https://doi.org/10.1093/mnras/stx2888>.
- B. Moser et al. *Astronomy and Computing*, 40:100603, 2022. ISSN 2213-1337. doi: <https://doi.org/10.1016/j.ascom.2022.100603>. URL <https://www.sciencedirect.com/science/article/pii/S2213133722000373>.
- S. G. Murray. *Journal of Open Source Software*, 3(28):850, 2018. doi: 10.21105/joss.00850. URL <https://doi.org/10.21105/joss.00850>.
- S. G. Murray et al. *Journal of Open Source Software*, 5(54):2582, 2020. doi: 10.21105/joss.02582. URL <https://doi.org/10.21105/joss.02582>.
- A. Nasirudin et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Nasirudin01.
- A. Nasirudin, P. Bull, and I. Ye. *arXiv e-prints*, art. arXiv:2604.26886, Apr. 2026. doi: 10.48550/arXiv.2604.26886.
- L. B. Newburgh, K. Bandura, and e. a. M. A. Bucher. In H. J. Hall, R. Gilmozzi, and H. K. Marshall, editors, *Ground-based and Airborne Telescopes VI*, volume 9906, page 99065X. International Society for Optics and Photonics, SPIE, 2016. doi: 10.1117/12.2234286. URL <https://doi.org/10.1117/12.2234286>.
- A. Obuljen, M. Simonović, A. Schneider, and R. Feldmann. *Phys. Rev. D*, 108(8):083528, Oct. 2023. doi: 10.1103/PhysRevD.108.083528.
- H. Padmanabhan, T. R. Choudhury, and A. Refregier. *MNRAS*, 447(4):3745–3755, Mar. 2015. doi:

- 10.1093/mnras/stu2702.
- H. Padmanabhan, A. Refregier, and A. Amara. *Monthly Notices of the Royal Astronomical Society*, 469(2):2323–2334, 04 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx979. URL <https://doi.org/10.1093/mnras/stx979>.
- G. Parimbelli et al. *Journal of Cosmology and Astroparticle Physics*, 2022(11):041, nov 2022. doi: 10.1088/1475-7516/2022/11/041. URL <https://dx.doi.org/10.1088/1475-7516/2022/11/041>.
- H. Parkinson, S. Cole, and J. Helly. *MNRAS*, 383(2):557–564, Jan. 2008. doi: 10.1111/j.1365-2966.2007.12517.x.
- S. Paul, M. G. Santos, Z. Chen, and L. Wolz. *arXiv e-prints*, art. arXiv:2301.11943, Jan. 2023. doi: 10.48550/arXiv.2301.11943.
- Planck Collaboration et al. *A&A*, 641:A6, Sept. 2020. doi: 10.1051/0004-6361/201833910.
- J. Pritchard and A. Loeb. *Nature*, 468(7325):772–773, Dec. 2010. doi: 10.1038/468772b.
- C. Péroux and J. C. Howk. *Annual Review of Astronomy and Astrophysics*, 58 (Volume 58, 2020):363–406, 2020. ISSN 1545-4282. doi: <https://doi.org/10.1146/annurev-astro-021820-120014>. URL <https://www.annualreviews.org/content/journals/10.1146/annurev-astro-021820-120014>.
- A. Refregier, L. Gamper, A. Amara, and L. Heisenberg. *Astronomy and Computing*, 25:38–43, 2018. ISSN 2213-1337. doi: <https://doi.org/10.1016/j.ascom.2018.08.001>. URL <https://www.sciencedirect.com/science/article/pii/S2213133717300963>.
- M. Reina-Campos et al. *MNRAS*, 517(3):3144–3180, Dec. 2022. doi: 10.1093/mnras/stac1934.
- M. Reina-Campos et al. *MNRAS*, 521(4):6368–6382, June 2023. doi: 10.1093/mnras/stad920.
- J. Rhee et al. *Monthly Notices of the Royal Astronomical Society*, 518(3):4646–4671, 10 2022.
- T. Ronconi, A. Lapi, M. Viel, and A. Sartori. *Monthly Notices of the Royal Astronomical Society*, 498(2):2095–2113, 07 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa2201. URL <https://doi.org/10.1093/mnras/staa2201>.
- T. Ronconi et al. *arXiv e-prints*, art. arXiv:2603.25650, Mar. 2026. doi: 10.48550/arXiv.2603.25650.
- M. Santos et al. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 19, Apr. 2015. doi: 10.22323/1.215.0019.
- G. Sato-Polito, N. Kokron, and J. L. Bernal. *Mon. Not. Roy. Astron. Soc.*, 526(4):5883–5899, 2023. doi: 10.1093/mnras/stad2498.
- S. Satpathy, R. A C Croft, S. Ho, and B. Li. *MNRAS*, 484(2):2148–2165, Apr. 2019. doi: 10.1093/mnras/stz009.
- J. Schaye et al. *arXiv e-prints*, art. arXiv:2508.21126, Aug. 2025. doi: 10.48550/arXiv.2508.21126.
- R. K. Sheth. *MNRAS*, 364(3):796–806, Dec. 2005. doi: 10.1111/j.1365-2966.2005.09609.x.
- J. Shin et al. *Astrophys. J.*, 843(1):73, 2017. doi: 10.3847/1538-4357/aa74b9.
- F. Sinigaglia and F.-S. Kitaura. *arXiv e-prints*, art. arXiv:2606.04830, June 2026. doi: 10.48550/arXiv.2606.04830.
- F. Sinigaglia et al. *The Astrophysical Journal*, 921(1):66, nov 2021. doi: 10.3847/1538-4357/ac158b. URL <https://dx.doi.org/10.3847/1538-4357/ac158b>.
- F. Sinigaglia et al. *ApJ*, 927(2):230, Mar. 2022. doi: 10.3847/1538-4357/ac5112.
- F. Sinigaglia, F.-S. Kitaura, K. Nagamine, and Y. Oku. *arXiv e-prints*, art. arXiv:2407.03918, July

- 2024a. doi: 10.48550/arXiv.2407.03918.
- F. Sinigaglia et al. *A&A*, 682:A21, Feb. 2024b. doi: 10.1051/0004-6361/202346931.
- SKA Cosmology SWG et al. *Publications of the Astronomical Society of Australia*, 37, 2020. ISSN 1448-6083. doi: 10.1017/pasa.2019.51. URL <http://dx.doi.org/10.1017/pasa.2019.51>.
- M. Spinelli et al. *MNRAS*, 493(4):5434–5455, Apr. 2020. doi: 10.1093/mnras/staa604.
- M. Spinelli et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Spinelli01.
- V. Springel, S. D. M. White, G. Tormen, and G. Kauffmann. *MNRAS*, 328(3):726–750, Dec. 2001. doi: 10.1046/j.1365-8711.2001.04912.x.
- V. Springel et al. *Nature*, 435(7042):629–636, June 2005. doi: 10.1038/nature03597.
- A. R. H. Stevens et al. *MNRAS*, 483(4):5334–5354, Mar. 2019. doi: 10.1093/mnras/sty3451.
- E. R. Switzer et al. *Astrophys. J.*, 815(1):51, 2015. doi: 10.1088/0004-637X/815/1/51.
- F. Tarsitano et al. *Astronomy and Computing*, 36:100484, 2021. ISSN 2213-1337. doi: <https://doi.org/10.1016/j.ascom.2021.100484>. URL <https://www.sciencedirect.com/science/article/pii/S221313372100038X>.
- N. Tessore et al. 2 2023. doi: 10.21105/astro.2302.01942.
- D. Tweed et al. *A&A*, 506(2):647–660, Nov. 2009. doi: 10.1051/0004-6361/200911787.
- A. P. Vijayan et al. *MNRAS*, 489(3):4072–4089, Nov. 2019. doi: 10.1093/mnras/stz1948.
- F. Villaescusa-Navarro et al. *ApJ*, 866(2):135, Oct. 2018. doi: 10.3847/1538-4357/aadba0.
- D. Wadekar, F. Villaescusa-Navarro, S. Ho, and L. Perreault-Levasseur. *arXiv e-prints*, art. arXiv:2012.00111, Nov. 2020. doi: 10.48550/arXiv.2012.00111.
- D. Wadekar, F. Villaescusa-Navarro, S. Ho, and L. Perreault-Levasseur. *ApJ*, 916(1):42, July 2021. doi: 10.3847/1538-4357/ac033a.
- M. White. *JCAP*, 2016(11):057, Nov. 2016. doi: 10.1088/1475-7516/2016/11/057.
- L. Wolz et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Wolz01.
- L. Xie et al. *MNRAS*, 469(1):968–993, July 2017. doi: 10.1093/mnras/stx889.
- L. Xie, G. De Lucia, M. Hirschmann, and F. Fontanot. *MNRAS*, 498(3):4327–4344, Nov. 2020. doi: 10.1093/mnras/staa2370.
- R. M. Yates et al. *MNRAS*, 435(4):3500–3520, Nov 2013. doi: 10.1093/mnras/stt1542.
- R. M. Yates, C. Péroux, and D. Nelson. *MNRAS*, 508(3):3535–3550, Dec. 2021. doi: 10.1093/mnras/stab2837.
- R. M. Yates et al. *MNRAS*, 527(3):6292–6311, Jan. 2024. doi: 10.1093/mnras/stad3419.
- A. Zoldan et al. *MNRAS*, 465(2):2236–2253, Feb. 2017. doi: 10.1093/mnras/stw2901.
- M. A. Zwaan, M. J. Meyer, L. Staveley-Smith, and R. L. Webster. *MNRAS*, 359(1):L30–L34, May 2005. doi: 10.1111/j.1745-3933.2005.00029.x.
- J. T. L. Zwart, D. Price, and G. Bernardi. HIBAYES: Global 21-cm Bayesian Monte-Carlo Model Fitting, June 2016.