# 21 cm intensity mapping cross-correlation with galaxy surveys: Current and forecasted cosmological parameters estimation for the SKAO

Maria Berti [1,2,3,4]★ Marta Spinelli [5,6,7,8] and Matteo Viel[2,3,4,6,9]

[1]*Département de Physique Théorique and Center for Astroparticle Physics, Université de Genève, Quai E. Ansermet 24, CH-1211 Genève 4, Switzerland .*
[2]*SISSA- International School for Advanced Studies, Via Bonomea 265, I-34136 Trieste, Italy.*
[3]*INFN – National Institute for Nuclear Physics, Via Valerio 2, I-34127 Trieste, Italy.*
[4]*IFPU, Institute for Fundamental Physics of the Universe,Via Beirut 2, I-34151 Trieste, Italy.*
[5]*Institute for Particle Physics and Astrophysics, ETH Zürich, Wolfgang Pauli Strasse 27, CH-8093 Zürich, Switzerland.*
[6]*INAF, Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, I-34131 Trieste, Italy.*
[7]*Department of Physics and Astronomy, University of the Western Cape, Robert Sobukhwe Road, Bellville, 7535, South Africa.*
[8] *Laboratoire Lagrange, UCA, OCA, CNRS, Blvd de l'Observatoire, CS 34229, F-06304 Nice cedex 4, France.*
[9]*Italian Research Center on High Performance Computing, Big Data and Quantum Computing, Via Magnanelli 2, I-40033, Bologna, Italy .*

## ABSTRACT

We present a comprehensive set of forecasts for the cross-correlation signal between 21 cm intensity mapping and galaxy redshift surveys. We focus on the data sets that will be provided by the SKAO for the 21 cm signal, DESI and Euclid for galaxy clustering. We build a likelihood which takes into account the effect of the beam for the radio observations, the Alcock–Paczynski effect, a simple parametrization of astrophysical nuisances, and fully exploit the tomographic power of such observations in the range $z = 0.7$–1.8 at linear and mildly non-linear scales ($k < 0.25h \, \mathrm{Mpc}^{-1}$). The forecasted constraints, obtained with Monte Carlo Markov Chains techniques in a Bayesian framework, in terms of the six base parameters of the standard $\Lambda$CDM model, are promising. The predicted signal-to-noise ratio for the cross-correlation can reach $\sim$50 for $z \sim 1$ and $k \sim 0.1h \, \mathrm{Mpc}^{-1}$. When the cross-correlation signal is combined with current Cosmic Microwave Background (CMB) data from Planck, the error bar on $\Omega_\mathrm{c} h^2$ and $H_0$ is reduced by factors 3 and 6, respectively, compared to CMB only data, due to the measurement of matter clustering provided by the two observables. The cross-correlation signal has a constraining power that is comparable to the autocorrelation one and combining all the clustering measurements a sub-per cent error bar of 0.33 per cent on $H_0$ can be achieved, which is about a factor 2 better than CMB only measurements. Finally, as a proof of concept, we test the full pipeline on the real data measured by the *MeerKat* collaboration (Cunnington et al. 2022) presenting some (weak) constraints on cosmological parameters.

**Key words:** cosmological parameters – large-scale structure of Universe – radio lines: general.

## 1 INTRODUCTION

Efforts to seek new physics beyond the standard cosmological model, which is grounded in cold dark matter and the cosmological constant ($\Lambda$CDM), are currently centred on addressing the $H_0$ and $S_8$ tensions. These involve modifications of the theoretical framework and the exploration of new observables with the potential to be sensitive to various scales and redshifts while being subject to distinct systematics and statistical errors compared to well-established probes.

Neutral hydrogen (HI) has recently emerged as a new quantitative tracer of the large-scale structure of the Universe (e.g. Ansari et al. 2012; Pritchard & Loeb 2012; Santos et al. 2015) via the intensity mapping (IM) technique. The 21 cm IM line is produced by the hyperfine structure spin-flip transition of the electron in atomic hydrogen (Furlanetto, Oh & Briggs 2006) and, compared to other observables, has the great advantage of probing large volumes in an efficient way at the expense of a relatively poor angular resolution.

Several planned and ongoing experiments, like compact interferometers (e.g. CHIME, Bandura et al. 2014; Amiri et al.2023 ), CHORD or HIRAX (Newburgh et al. 2016) or single-dish telescopes (such as GBT, Masui et al. 2013; Wolz et al. 2022; or FAST, Hu et al. 2020) aim at measuring the IM signal (Bharadwaj et al. 2001; Battye, Davies & Weller 2004; McQuinn et al. 2006; Chang et al. 2008; Seo et al. 2010; Kovetz et al. 2017; Villaescusa-Navarro et al. 2018), and some of them have achieved the detection of the HI signal in cross-correlation with galaxy surveys (Chang et al. 2010; Masui et al. 2013; Anderson et al. 2018; Wolz et al. 2022), since this measurement is likely to be less prone to systematics like foregrounds.

The SKA Observatory (SKAO)[1] consists of SKA-Low and SKA-Mid telescopes, which will be located in Australia and South Africa, respectively. Using the SKA-Mid telescope array as a collection of single-dishes (Battye et al. 2013; Santos et al. 2015), it will be possible to perform 21 cm IM at cosmological scales up to redshift $z \sim 3$ (SKA Cosmology SWG 2020). The SKAO is currently under construction and MeerKAT, the SKA-Mid precursor, has been con-

---

★ E-mail: mberti@sissa.it

[1]https://www.skao.int/

ducting IM survey for cosmology (Santos et al. 2017, MeerKLASS). Preliminary data analyses have provided interesting results on the potential of the *MeerKAT* telescope IM surveys (Wang et al. 2021; Irfan et al. 2022) along with a first MeerKLASS detection of the HI signal in cross-correlation with WiggleZ galaxies (Cunnington et al. 2022). More recently, another breakthrough has been reached with the detection at cosmological scales of HI with IM in the auto-power spectrum (Paul et al. 2023). However, mitigating the foregrounds and their impact on the extracted signal, both in cross and autocorrelation, is challenging and several foreground removal techniques have been proposed (Alonso et al. 2015; Wolz et al. 2016; Carucci, Irfan & Bobin 2020; Matshawule et al. 2021; Cunnington et al. 2021; Irfan & Bull 2021; Soares et al. 2021; Spinelli et al. 2021; Soares et al. 2022).

From the theoretical point of view, it is of great importance to refine the forecasts for the 21 cm IM, both alone and in combination with other probes, to optimize the survey design in order to enhance the signal-to-noise ratio (Villaescusa-Navarro et al. 2015; Bull et al. 2016; Jiang et al. 2023), to address the non-linear scales modelling in the context of the MeerKat detection (Padmanabhan et al. 2023), to investigate the cross-correlation in the connection with galaxy formation models (Spinelli et al. 2020), and to estimate the impact of fundamental new physics on the observables, like non-Gaussianities or dark energy (Jolicoeur, Maartens & Dlamini 2023; Wu et al. 2023).

In Berti, Spinelli & Viel (2023), we built on the formalism of Blake (2019), Cunnington et al. (2020, 2022), and Soares et al. (2021), and study the redshift-space 21 cm power spectrum monopole and quadrupole, forecasting the constraining power of SKAO observations within multiple redshift bins. In this work, we extend this analysis by including in our pipeline the modelling of the 21 cm and galaxy clustering cross-correlation signal. As before, we focus on the parameters of the ΛCDM model and exploit the exquisite tomographic nature of the 21 cm IM signal.

For the 21 cm IM, we mimic SKA-Mid observations following the SKAO Redbook prescriptions (SKA Cosmology SWG 2020). Regarding galaxies, we rely on the mocking of data sets for the galaxy clustering signal which could be provided soon by Dark Energy Spectroscopic Survey (DESI; Aghamousa et al. 2016; Vargas-Magaña et al. 2018) and the Euclid mission (Blanchard et al. 2020). We will build a full likelihood integrated within the CosmoMC code (Lewis & Bridle 2002; Lewis 2013) and compute constraints through Markov Chain Monte Carlo (MCMC) techniques. We assess the constraining power of our mock 21 cm data set in cross-correlation with galaxy clustering alone and combined with CMB data.

We note that forecasts for future IM observations based on the Fisher matrix formalism have also been presented in Obuljen et al. (2018), Viljoen, Fonseca & Maartens (2020), and Karagiannis, Maartens & Randrianjanahary (2022). Our work expands on previous studies in the following aspects: (i) we constrain the complete set of cosmological parameters, (ii) we cross-correlate the signal with the most recent forecasts for state-of-the-art galaxy surveys, (iii) we combine the 21 cm data within multiple redshift bins with the Planck latest available results, and (iv) we conduct complete MCMC analyses, to estimate the full posterior distribution.

The structure of the paper is as follows. The methodology, discussed more extensively in Berti et al. (2023), is briefly reviewed in Section 2. The building of the mock observations is detailed in Section 3. Results are discussed in Section 4, where we present the forecasted constraints obtained from 21 cm and galaxy clustering cross-correlation alone (Section 4.1) and in combination with the latest Planck 2018 CMB data (Section 4.2). Results obtained with data measured in Cunnington et al. (2022) are shown in Section 4.3. A summary of the results and our conclusions is outlined in Section 5.

We also provide two useful appendixes that discuss the Alcock–Paczynski (AP) effect's impact and a further cross-check on the attained signal on present data sets.

## 2 MODELLING THE CROSS-CORRELATION SIGNAL

The analysis presented in this work for the 21 cm× galaxy clustering power spectrum is an extension of the one discussed in Berti et al. (2023). Since we adopt analogous formalism and framework of the previous study, in the following, we review only the essential information. Having defined the cosmological model we consider in Section 2.1, we discuss the 21 cm auto-power spectrum and the galaxy power spectra in Sections 2.2 and 2.3, which enter the error estimation. The model for the cross-correlation power spectrum and the power spectrum multipole expansion are presented in Sections 2.4 and 2.5.

### 2.1 Fiducial cosmological model

We work within the standard cosmological model framework, i.e. the ΛCDM model. We perform our analysis using the following six parameters to define the fiducial cosmology: $\Omega_b h^2$ and $\Omega_c h^2$, which describes the density of the baryonic and cold dark matter, respectively, the scalar spectral index $n_s$, the normalization of the primordial power spectrum $A_s$, the Thomson scattering optical depth due to reionization $\tau$, and $\theta_{MC}$, that is connected to the angular scale of the sound horizon at decoupling. Moreover, we will focus on the derived parameters $H_0$, i.e. the current expansion rate in km s$^{-1}$ Mpc$^{-1}$ and $\sigma_8$, the root mean square matter fluctuations today in linear theory.

Through all this work, we assume a universe described by a Planck 2018 (Planck Collaboration VI 2020) fiducial cosmology, i.e. $\{\Omega_b h^2 = 0.022383, \Omega_c h^2 = 0.12011, n_s = 0.96605, \ln(10^{10} A_s) = 3.0448, \tau = 0.0543, H_0 = 67.32$ km s$^{-1}$ Mpc$^{-1}$, $\Sigma m_\nu = 0.06$eV$\}$, where $\Sigma m_\nu$ is the sum of neutrino masses in eV.

### 2.2 Model for the observed 21 cm signal power spectrum

The 21 cm non-linear power spectrum can be modelled as (Kaiser 1987; Villaescusa-Navarro et al. 2018; SKA Cosmology SWG 2020)

$$P_{21}(z, k, \mu) = \bar{T}_b^2(z) \left[ \left( b_{HI}(z) + f(z)\mu^2 \right)^2 P_m(z, k) + P_{SN}(z) \right], \quad (1)$$

where $\bar{T}_b$ is the HI mean brightness temperature, $b_{HI}$ is the HI bias, $f$ is the growth rate, $\mu = \hat{k} \cdot \hat{z}$ is the cosine of the angle between the wave number and the line of sight, $P_m(z, k)$ is the non-linear matter power spectrum, and $P_{SN}$ is the shot noise term.

For the evolution in redshift of the brightness temperature, we use the parametrization defined in Battye et al. (2013). Given that we lack an analytical model, $b_{HI}(z)$ and $P_{SN}(z)$ at a given redshift are computed by interpolating numerical results from hydrodynamical simulations (Villaescusa-Navarro et al. 2015; Villaescusa-Navarro et al. 2018). The growth rate $f(z)$ and the non-linear matter power spectrum $P_m(z, k)$ are, instead, computed numerically by means of the Einstein–Boltzmann solver CAMB[2] (Lewis, Challinor & Lasenby 2000).

---

[2]See https://CAMB.info/. Note that non-linear corrections to the matter power spectrum are computed with the HALOFIT (Smith et al. 2003) version from Mead et al. (2016).

To mimic a realistic observation, we introduce the effect of a Gaussian telescope beam, as a suppression of the power spectrum on scales smaller than the beam's full width at half maximum (Battye et al. 2013; Villaescusa-Navarro, Alonso & Viel 2017; Cunnington et al. 2020; Soares et al. 2021; Cunnington 2022). The corresponding damping factor $\tilde{B}(z, k, \mu)$ can be written in terms of the beam's physical dimension $R_{\text{beam}}$, as

$$\tilde{B}(z, k, \mu) = \exp\left[\frac{-k^2 R_{\text{beam}}^2(z)(1 - \mu^2)}{2}\right]. \tag{2}$$

In a real-world scenario, one must consider the possibility of having chosen the wrong fiducial cosmology. This can be taken into account with the AP modifications (Alcock & Paczynski 1979). Anisotropies along the radial and transverse direction can be modelled as[3]

$$\alpha_\perp(z) = \frac{D_A(z)}{D_A^{\text{fid}}(z)} \quad \text{and} \quad \alpha_\parallel(z) = \frac{H^{\text{fid}}(z)}{H(z)}. \tag{3}$$

Here, $D_A^{\text{fid}}(z)$ and $H^{\text{fid}}(z)$ are the values of the angular diameter distance and the Hubble parameter at redshift $z$ predicted by the fiducial cosmology. The AP parameters $\alpha_\perp$ and $\alpha_\parallel$ modify the overall amplitude of the power spectrum and the wave vectors. The wave vector components along and transverse to the line of sight are then distorted as

$$q = \frac{k}{\alpha_\perp}\sqrt{1 + \mu^2\left(\frac{\alpha_\perp^2}{\alpha_\parallel^2} - 1\right)} \tag{4}$$

and

$$\nu = \frac{\alpha_\perp \mu}{\alpha_\parallel\sqrt{1 + \mu^2\left(\frac{\alpha_\perp^2}{\alpha_\parallel^2} - 1\right)}}, \tag{5}$$

where $k$ and $\mu$ are the assumed fiducial values of the wave vectors.

The observed 21 cm power spectrum, marked with the symbol $\hat{}$, including the beam smoothing, the AP effects and the instrumental noise, is then

$$\hat{P}_{21}(z, k, \mu) = \frac{1}{\alpha_\perp^2 \alpha_\parallel}\tilde{B}^2(z, q, \nu)P_{21}(z, q, \nu) + P_{\text{N}}(z), \tag{6}$$

where $P_{21}(z, q, \nu)$ is defined in equation (1), but computed on the new variables $q$ and $\nu$. The SKAO-like instrumental noise $P_{\text{N}}$ can be modelled using the instrument specifications as in equation 9 of Berti et al. (2023).

We note that, in this work, we expand the modelling of Berti et al. (2023), where we neglected the AP contribution in the first approximation. A discussion on the effect of the inclusion of the AP distortions on the cosmological parameter constraints is presented in Appendix A.

## 2.3 Model for the galaxy power spectrum

The simplest parametrization of the galaxy power spectrum at a given redshift can be written as

$$P_{\text{g}}(z, k, \mu) = \left(b_{\text{g}}(z) + f(z)\mu^2\right)^2 P_{\text{m}}(z, k) + \frac{1}{\bar{n}_{\text{g}}(z)}, \tag{7}$$

where $b_{\text{g}}$ is the galaxy bias and $\bar{n}_{\text{g}}$ is the galaxy number density. The term $1/\bar{n}_{\text{g}}$ is the shot noise term for the galaxy power spectrum.

In this work, we use values of $b_{\text{g}}$ and $\bar{n}_{\text{g}}$ in agreement with the official expected values for the planned galaxy surveys, as discussed in Section 3.2.

Due to the different observing techniques, the galaxy power spectrum is not affected by the beam correction. The AP distortions, instead, are the ones described in the previous section for $P_{21}$. Therefore, the observed galaxy power spectrum we consider is

$$\hat{P}_{\text{g}}(z, k, \mu) = \frac{1}{\alpha_\perp^2 \alpha_\parallel}P_{\text{g}}(z, q, \nu). \tag{8}$$

## 2.4 The cross-correlation signal power spectrum

To predict the cross-correlation power spectrum between the 21 cm signal and galaxy clustering, we use the following model [see e.g. Cunnington et al. (2022); Casas et al. (2023)]

$$P_{21,\text{g}}(z, k, \mu) = \bar{T}_b\left(b_{\text{HI}} + f\mu^2\right)\left(b_{\text{g}} + f\mu^2\right)P_{\text{m}}(z, k, \mu), \tag{9}$$

where all the quantities appearing here are defined in the previous sections. In the expression above, we do not make explicit the redshift dependence of the brightness temperature, the bias, and the growth rate for the sake of notation easiness. Moreover, it can be shown that the shot noise contribution for the cross-correlation power spectrum is negligible (Castorina & Villaescusa-Navarro 2017; Villaescusa-Navarro et al. 2018)

Taking into account the IM beam effect and the AP distortions, the observed cross-correlation signal becomes

$$\hat{P}_{21,\text{g}}(z, k, \mu) = \frac{1}{\alpha_\perp^2 \alpha_\parallel}r\tilde{B}_\perp(z, q, \nu)P_{21,\text{g}}(z, q, \nu), \tag{10}$$

with $r$ being a cross-correlation coefficient that accounts for unknown effects that may modify the theoretical estimate of the correlation degree.[4]

## 2.5 Multipole expansion

The non-isotropic power spectrum can be decomposed using Legendre polynomials $\mathcal{L}_\ell(\mu)$. The coefficients of the expansion, i.e. the multipoles of the power spectrum, are given by

$$\hat{P}_{\text{X},\ell}(z, k) = \frac{(2\ell + 1)}{2}\int_{-1}^{1} d\mu\, \mathcal{L}_\ell(\mu)\hat{P}_{\text{X}}(z, k, \mu), \tag{11}$$

with X being either the 21 cm IM (X = 21), the galaxy clustering (X = g) or their cross-term (X = 21, g). In this work, we use the auto-power spectrum and cross-correlation monopoles, for which $\ell = 0$ and $\mathcal{L}_0(\mu) = 1$. In particular, we focus on forecasting the cross-correlation power spectrum monopole $\hat{P}_{21,\text{g},0}(z, k)$. In the following, for clarity of notation, we drop the subscript 0 and simply refer to the monopoles as $\hat{P}_{21,\text{g}}(z, k)$, $\hat{P}_{\text{g}}(z, k)$ and $\hat{P}_{21}(z, k)$.

# 3 CONSTRUCTING THE MOCK CROSS-CORRELATION DATA

In this section, we construct mock data sets of cross-correlation measurements from future surveys. The 21 cm and galaxy surveys we take into account are presented in Section 3.1. Details on the construction of the synthetic data set and the analysis framework are given in Sections 3.2 and 3.3.

---

[3]In the literature, several definitions of $\alpha_\perp$ and $\alpha_\parallel$ have been proposed, e.g. Gil-Marín et al. (2017), Hand et al. (2017), and D'Amico et al. (2020). We follow the one of e.g. Blanchard et al. (2020) and Soares et al. (2021).

[4]The definition of $r$ is not unique (see e.g. the discussion in Cunnington et al. (2022)). In this work, we consider it an overall constant for simplicity, given that $r$ is considered as a nuisance parameter (Section 3.3.2).

**Table 1.** Assumed specifications for SKA-Mid Wide band 1 (SKA Cosmology SWG 2020), DESI ELG (Aghamousa et al. 2016; Casas et al. 2023), and Euclid-like surveys (Blanchard et al. 2020). For simplicity, we refer to SKA-Mid as SKAO, to DESI ELG as DESI, and to Euclid-like as Euclid. We collect the used effective redshifts $z$ and bin widths $\Delta z$, the galaxy biases $b_g$ and number densities $\bar{n}_g$, that we express in units of $[10^{-4}h^3 \text{ Mpc}^{-3}]$, the 21 cm IM bias $b_{HI}$, and the 21 cm power spectrum shot noise $P_{SN}$, in units of $[(h^{-1} \text{ Mpc})^3]$.

| SKAO | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Band frequency range | | | 0.35–1.05 GHz | | | | | | |
| Corresponding redshift range | | | 0.35–3 | | | | | | |
| Dish diameter $D_{dish}$ | | | 15 [m] | | | | | | |

| SKAO × Euclid | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Observed redshift range | | | 0.9–1.8 | | | | | | |
| Overlapping survey area | | | 10 000 [deg$^2$] | | | | | | |
| Corresponding $\Omega_{sur}$ | | | 3.0 [sr] | | | | | | |

| $z$ | 1. | | 1.2 | | 1.4 | | | 1.65 | |
|---|---|---|---|---|---|---|---|---|---|
| $\Delta z$ | 0.2 | | 0.2 | | 0.2 | | | 0.3 | |
| $b_g$ | 1.46 | | 1.61 | | 1.75 | | | 1.9 | |
| $\bar{n}_g$ | 6.86 | | 5.58 | | 4.21 | | | 2.61 | |
| $b_{HI}$ | 1.49 | | 1.60 | | 1.71 | | | 1.84 | |
| $P_{SN}$ | 124 | | 114 | | 101 | | | 85.0 | |

| SKAO × DESI | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Observed redshift range | | | 0.7–1.7 | | | | | | |
| Overlapping survey area | | | 5000 [deg$^2$] | | | | | | |
| Corresponding $\Omega_{sur}$ | | | 1.5 [sr] | | | | | | |

| $z$ | 0.75 | 0.85 | 0.95 | 1.05 | 1.15 | 1.25 | 1.35 | 1.45 | 1.55 | 1.65 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\Delta z$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| $b_g$ | 1.05 | 1.08 | 1.11 | 1.14 | 1.18 | 1.21 | 1.25 | 1.28 | 1.32 | 1.36 |
| $\bar{n}_g$ | 11.2 | 8.32 | 8.16 | 5.14 | 4.49 | 4.19 | 1.57 | 1.35 | 0.921 | 0.344 |
| $b_{HI}$ | 1.35 | 1.40 | 1.46 | 1.52 | 1.57 | 1.63 | 1.68 | 1.73 | 1.78 | 1.84 |
| $P_{SN}$ | 132 | 130 | 126 | 122 | 116 | 111 | 105 | 98 | 91 | 85 |

## 3.1 Survey specifications

### 3.1.1 21 cm intensity mapping

The main focus of our analysis is the 21 cm IM signal that can be measured with the SKAO telescope. We consider, in particular, a cosmological survey with the SKA-Mid telescope in single-dish mode, following SKA Cosmology SWG (2020). We assume a Wide Band 1 survey that covers a sky area of 20 000 deg$^2$ in the frequency range 0.35–1.05 GHz (i.e. the redshift range 0.35–3). The used SKAO specifications are summarized in Table 1.

### 3.1.2 Galaxy surveys

We assume a Euclid-like and a DESI-like spectroscopic galaxy survey. For Euclid, following what has been proposed in Blanchard et al. (2020), we consider observations within four different redshift bins in the range of 0.9–1.8. The assumed values of the galaxy bias and number density computed at each effective redshift are presented in Table 1.

To obtain a cross-correlation signal, one must take into account observations of the same portion of the sky. In agreement with other studies in the literature, we assume a 10 000 deg$^2$ overlapping patch of the sky observed by the SKAO and a Euclid-like survey.

Hereafter, we simply refer to the Euclid survey, where is understood that a Euclid-like survey as the one described above is intended.

To construct cross-correlation measurements between the SKAO and DESI, we follow Casas et al. (2023). We focus on the DESI Emission Line Galaxies (ELG), as they probe a redshift range similar to the one covered by Euclid, i.e. 0.7–1.7, making easier a direct comparison between the two experiments. In Table 1, we report the assumed values of the galaxy bias and number density at each effective redshift and we consider an overlapping area between DESI ELG and SKAO of 5000 deg$^2$. The smaller area overlap with respect to a Euclid-like survey is forced by the different hemisphere locations of the two telescopes.

## 3.2 Mock data sets

We construct two different mock data sets for the 21 cm and galaxy clustering cross-correlation power spectrum. One mimics an SKAO × Euclid analysis and the other a SKAO × DESI one, for the redshift bins and survey specifications described in Section 3.1 and Table 1.

The scales that are accessible with the observations are fixed by the volume probed with the surveys in a given redshift bin. In Fourier space, the largest scale available at each effective redshift is $k_{min}(z) = 2\pi/\sqrt[3]{V_{bin}(z)}$, where $V_{bin}(z)$ is the volume of each redshift bin, which we compute as

$$V_{bin}(z) = \Omega_{sur} \int_{z-\Delta z/2}^{z+\Delta z/2} dz' \frac{dV}{dz'd\Omega}$$
$$= \Omega_{sur} \int_{z-\Delta z/2}^{z+\Delta z/2} dz' \frac{cr(z')^2}{H(z')}. \tag{12}$$

with $r(z)$ being the comoving distance and $\Omega_{sur}$ the survey are in steradians (see Table 1). The smallest scale is, instead, imposed by the size of the SKAO telescope beam, due to the damping effect. It can be estimated as $k_{max}(z) = 2\pi/R_{beam}(z)$. At scales smaller than $k_{max}$, the signal is dominated by the beam providing no relevant information on cosmology. Finally, we choose a fixed k-bin width as

a function of redshift $\Delta k(z)$ in order to be large enough for modes to be independent, i.e. $\Delta k(z) \sim 2k_{\min}(z)$.

Assuming a set of $N$ measurements at redshift $z$ of the cross-correlation power spectrum $\hat{P}_{21,g}(k)$ at scales $\{k_1, \ldots, k_N\}$, we estimate the error on at each point as (see e.g. (Smith 2009; Cunnington et al. 2022))

$$\hat{\sigma}_{21,g}(k) = \frac{1}{\sqrt{2N_{\mathrm{modes}}(k)}} \sqrt{\hat{P}_{21,g}^2(k) + \hat{P}_{21}(k)\hat{P}_g(k)}, \tag{13}$$

where $\hat{P}_{21,g}$ is the cross-correlation power spectrum defined in equation (10), $\hat{P}_{21}$ and $\hat{P}_g$ are the 21 cm and the galaxy power spectrum introduced in equations (6) and (8), respectively. Here, $N_{\mathrm{modes}}$ is the number of modes per $k$ and $\mu$ bin, computed as

$$N_{\mathrm{modes}}(z, k) = \frac{k^2 \Delta k(z)}{4\pi^2} V_{\mathrm{bin}}(z). \tag{14}$$

At each central redshift $z$ and data point $k$, we compute the cross-correlation power spectrum for SKAO × Euclid observations, labelled as $\hat{P}_{21,g}^{\mathrm{Euclid}}(z, k)$, the one for SKAO × DESI, $\hat{P}_{21,g}^{\mathrm{DESI}}(z, k)$, and the corresponding errors, as discussed above. In Table 1, we gather some of the used redshift-dependent quantities. The resulting mock data sets are shown in Fig. 1.

We highlight that in equation (13) the 21 cm IM instrumental noise enters the estimate of the errors through $\hat{P}_{21}$. However, we find that the contribution of SKAO-like instrumental noise is minimal. This is not the case, instead, for a *MeerKAT*-like noise, which induces a non-negligible contribution to the error estimate, as in the analysis of Section 4.3.

### 3.3 Numerical analysis

In order to exploit the constraining power of the mock data set presented in Section 3.2, we define a likelihood function and then set up the framework to constrain the cosmological parameters by adopting a Bayesian approach. Given a set of observations and a theory that depends on given parameters, the Bayes theorem links the posterior distribution to the likelihood function. The high-dimensional posterior can then be sampled using MCMC methods (see e.g. Gilks, Richardson & Spiegelhalter 1995). Following Berti et al. (2023), we build a working pipeline to conduct full MCMC analyses on 21 cm and cross-correlation observations. We test this pipeline by forecasting the constraining power of the data sets described above.

#### 3.3.1 Likelihood function for the 21 cm multipoles

Given a set of measurements at scales $\{k_1, \ldots, k_N\}$ and redshift $z$, to compute the likelihood function we define the vector $\Theta(z) = \left( \hat{P}_{21,g}(z, k_1), \ldots, \hat{P}_{21,g}(z, k_N) \right)$. The logarithmic likelihood is computed as

$$-\ln[\mathcal{L}] = \sum_z \frac{1}{2} \Delta\Theta(z)^{\mathrm{T}} \mathsf{C}^{-1}(z) \Delta\Theta(z), \tag{15}$$

where $\Delta\Theta(z) = \Theta^{\mathrm{th}}(z) - \Theta^{\mathrm{obs}}(z)$, the difference between the values of $\Theta(z)$ predicted from theory and observed. Here, $\mathsf{C}(z)$ is the covariance matrix computed as $\mathsf{C}(z) = \mathrm{diag}(\hat{\sigma}_{21,g}^2(z, k_1), \ldots, \hat{\sigma}_{21,g}^2(z, k_N))$. We consider independent redshift bins, i.e. we simply sum over the contribution from each central redshift.

Fig. 2 shows the signal-to-noise ratios as a function of $k$ in each redshift bin for both the constructed mock data sets. We observe that the signal to noise decreases at higher redshifts. The behaviour and
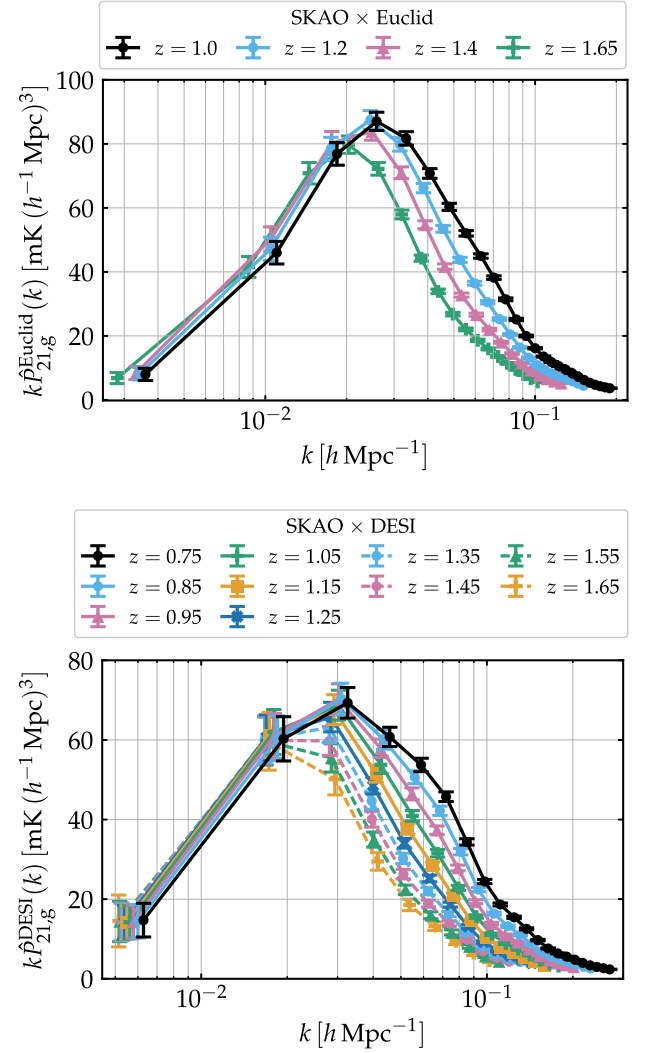


**Figure 1.** Mock data set for SKAO × Euclid (upper panel) and SKAO × DESI (lower panel) observations. The considered redshift bins are different for the two galaxy surveys. We refer to Section 3 for the extended discussion on how the signal and the errors are computed.

orders of magnitude found here are compatible with the results for the 21 cm power spectrum multipoles in Soares et al. (2021); Berti et al. (2023).

We conduct an MCMC analysis varying the six parameters describing the $\Lambda$CDM model, i.e. we vary $\{\Omega_b h^2, \Omega_c h^2, n_s, \ln(10^{10}A_s), \tau, 100\theta_{\mathrm{MC}}, \Sigma m_\nu, P_{\mathrm{SN},i}\}$ assuming wide flat priors on each of the parameters. Results on other parameters, such as $H_0$ and $\sigma_8$, are derived from results on this set. To perform the study, we develop a likelihood code integrated with the MCMC sampler CosmoMC[5] (Lewis & Bridle 2002; Lewis 2013). We further expand on the code we implemented and used in Berti et al. (2022, 2023) including the computation of the theoretical expectations for the 21 cm and galaxy clustering cross-correlation power spectrum and the relative likelihood function at different redshift. We recall that each redshift bin is considered independent,

---

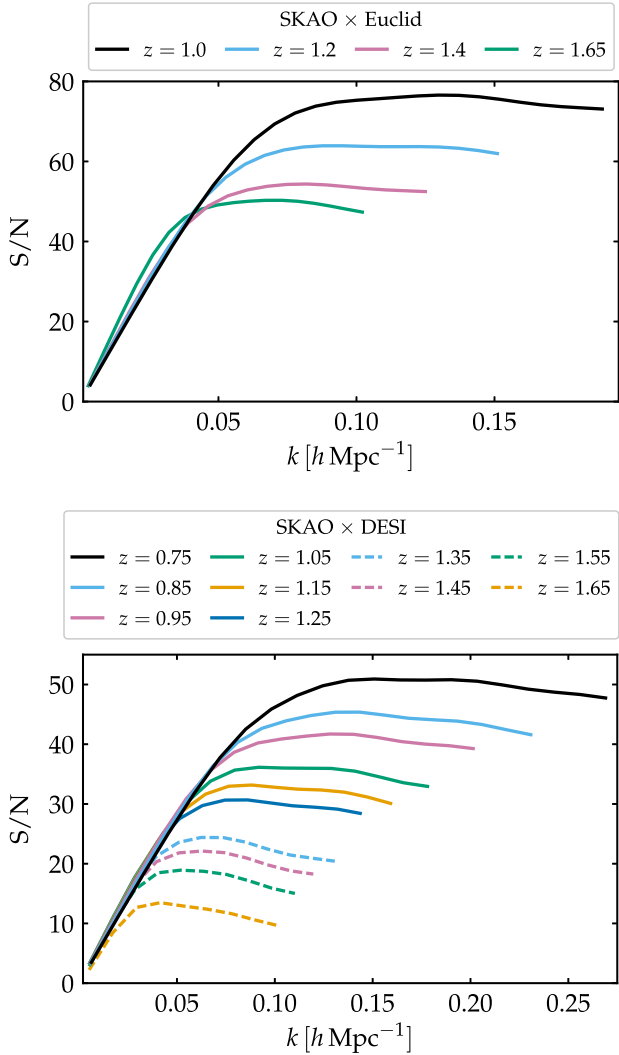[5]See https://cosmologist.info/cosmomc.

**Figure 2.** Predicted signal-to-noise ratio as a function of $k$ for SKAO $\times$ Euclid (upper panel) and SKAO $\times$ DESI (lower panel) mock observations.

thus we consider a diagonal covariance matrix constructed with the forecasted errors.

### 3.3.2 Nuisance parameters

As in Berti et al. (2022, 2023), along with the cosmological parameters we implement different nuisances. Indeed, the access to the matter clustering is not direct as it appears in equation (10) in combination with the brightness temperature and the HI bias and the galaxy bias. These quantities, although the scientific community hopes to obtain external measurements (e.g. the total neutral hydrogen density as a function of redshift, key unknown for the brightness temperature, is one of the scientific goals of the MeerKAT survey Laduma), may need to be treated as unconstrained quantities in a pessimistic scenario. To take into account this lack of knowledge, we allow for combinations of these parameters to vary in the MCMC run, thus leaving free the overall amplitude of the power spectrum. The contribution from the nuisances is then marginalized out in the final analysis.

To be completely agnostic, for the cross-correlation power spectrum we include in the nuisances also the correlation coefficient $r$ and the galaxy bias. Thus, we consider the following three combinations of parameters $\sqrt{r\overline{T}_b}b_{HI}\sigma_8$, $\sqrt{r\overline{T}_b}b_g\sigma_8$, and $\sqrt{r\overline{T}_b}f\sigma_8$, where we renormalized the matter power spectrum as $P_m/\sigma_8^2$.

Given that all the parameters are redshift-dependent quantities, the actual number of nuisances is three times the number of redshift bins. This translates into $4 \times 3$ nuisance parameters for SKAO $\times$ Euclid and $10 \times 3$ for SKAO $\times$ DESI. Especially in the latter case, the high number of parameters to vary can impact the numerical efficiency of the MCMC computations. Following what is already done in Berti et al. (2023), for SKAO $\times$ DESI only we reduce the number of nuisances by constraining their redshift evolution through a polynomial parametrization. Rewriting $N(z) = az^3 + bz^2 + cz + d$ for $N(z) = \sqrt{r\overline{T}_b}b_{HI}\sigma_8$, $\sqrt{r\overline{T}_b}b_g\sigma_8$, $\sqrt{r\overline{T}_b}f\sigma_8$, we implement as nuisances the coefficient of the polynomial $a$, $b$, $c$, and $d$, reducing the number of nuisance parameters from 30 to 12.

In the following, with the label 'nuisances' or 'nuis.' we refer to the parameters described above. For each nuisance, we assume a wide flat prior in the range [ 1, 1].

### 3.3.3 CMB likelihoods and data sets

In this study, we combine our mock 21 cm and galaxy clustering cross-correlation data sets with Planck 2018 (Planck Collaboration VI 2020). The CMB likelihood includes the high-$\ell$ TT, TE, EE lite likelihood in the interval of multipoles $30 \leq \ell \leq 2508$ for TT and $30 \leq \ell \leq 19696$ for TE, EE. Lite likelihoods are calculated with the `Plik lite` likelihood (Planck Collaboration V 2020). Instead for the low-$\ell$ TT power spectrum, we use data from the `Commander` component-separation algorithm in the range $2 \leq \ell \leq 29$. We also adopt the Planck CMB lensing likelihood and the low EE polarization power spectrum, referred to as lowE, in the range $2 \leq \ell \leq 29$, calculated from the likelihood code `SimAll` (Planck Collaboration III 2020). In the rest of the paper with the label 'Planck 2018' we refer to the combination TT, TE, EE + low-$\ell$ + lowE + lensing.

## 4 RESULTS

We present in this section the results of our analysis. We first explore the constraining power of the mock cross-correlation data, with and without nuisances (Section 4.1). We then combine the mock data sets with Planck CMB data (Section 4.2). Finally, in Section 4.3 we present the constraints we obtain on the cosmological parameters for the published measurement of the *MeerKAT* $\times$ WiggleZ cross-correlation power spectrum presented in Cunnington et al. (2022).

As a reference, throughout this analysis, we compare results from the cross-correlation forecast with the best results obtained with the 21 cm multiples in Berti et al. (2023), i.e. the fully non-linear monopole and quadrupole data set that we label as '$\hat{P}_0 + \hat{P}_2$'. Note that we expand on the results of Berti et al. (2023) by introducing the AP, in order to be consistent with the modelling of the cross-correlation used in this work. Thus, '$\hat{P}_0 + \hat{P}_2$' here include AP effects. For a discussion on the impact of AP distortions, we refer to Appendix A.

We stress that when combining auto-power spectrum and cross-correlation forecasts, we neglect the covariance between the two data sets in the first approximation.

We show the marginalized 1D and 2D posteriors for the studied set of parameters. Note that 68 per cent confidence level constraints are presented as percentages with respect to the marginalized mean value.
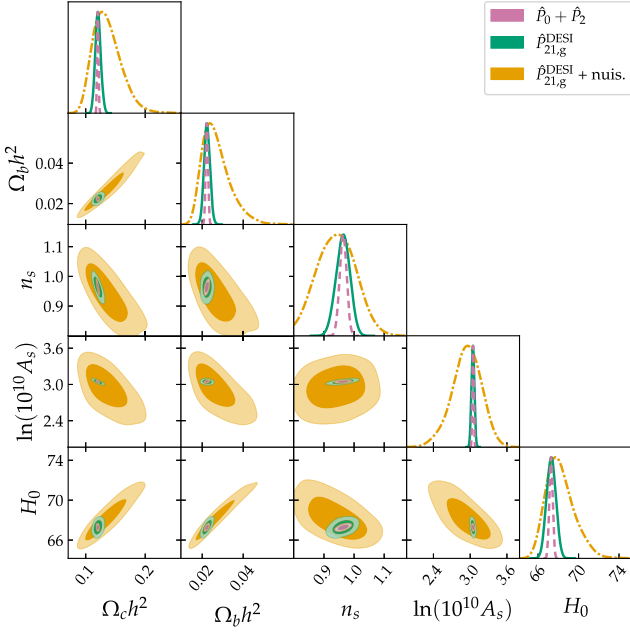
**Figure 3.** Joint constraints (68 per cent and 95 per cent confidence regions) and marginalized posterior distributions on a subset of the cosmological parameters. The label '$\hat{P}_0 + \hat{P}_2$' (dashed lines) stands for the forecasted 21 cm power spectrum monopole and quadrupole observations (see Appendix A). '$\hat{P}_{21,g}^{DESI}$' refers to the mock cross-correlation power spectrum data set constructed above. The label 'nuis.' (dashed-dotted lines) indicates that we vary the nuisance parameters along with the cosmological ones. The relative constraints are listed in Table 2.
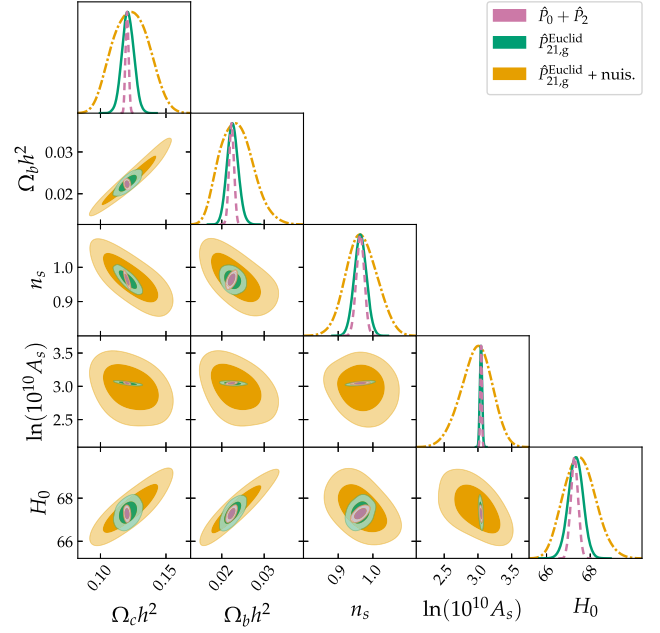
**Figure 4.** Joint constraints (68 per cent and 95 per cent confidence regions) and marginalized posterior distributions on a subset of the cosmological parameters. The label '$\hat{P}_0 + \hat{P}_2$' (dashed lines) stands for the forecasted 21 cm power spectrum monopole and quadrupole observations (see Appendix A). '$\hat{P}_{21,g}^{Euclid}$' refers to the mock cross-correlation power spectrum data set constructed above. The label 'nuis.' (dashed-dotted lines) indicates that we vary the nuisance parameters along with the cosmological ones. The relative constraints are listed in Table 2.

## 4.1 Probing the constraining power of future 21 cm× galaxy clustering data

In Figs 3 and 4, we present the forecasted posterior distributions we obtain for the SKAO × DESI and SKAO × Euclid mock data sets we construct in this work. We show only some of the model parameters for brevity.

We obtain comparable results for both the SKAO × Euclid and the SKAO × DESI analysis. Looking at the 2D contours, we observe that the correlations between the cosmological parameters are similar and in line with the results obtained with the 21 cm multipoles. The marked degeneracy in the $H_0 - \Omega_c h^2$ plane, found in previous works (Berti et al. 2022, 2023), is present also for the cross-correlation power spectrum case. As discussed in Bardeen et al. (1986), measuring cosmological observables that strongly depend on the matter power spectrum, as $\hat{P}_m$ does, fixes the shape of $P_m$. This translates into fixing the quantity $\Omega_m h$, which, in turn, induces the strong correlation $\Omega_c h^2 \propto H_0$. This feature is particularly relevant when combining 21 cm observations with CMB data, as discussed in the next section.

As expected from the signal to noise estimates of Fig. 2, better constraints are obtained for the SKAO × Euclid data set (Table 2). Although DESI probes the same redshift range of Euclid and even with a higher number of redshift bins, Euclid will have a larger sky area of overlap with SKAO, suggesting that a larger sky coverage increases the constraining power more than the number of redshift bins. As expected, the best constraints are the ones obtained with the 21 cm multipoles, in particular for $\Omega_c h^2$ and $H_0$. Indeed, the $\hat{P}_0 + \hat{P}_2$ is constructed to sample a wider redshift ($z = 0$–3) and scales range (up to $k \sim 1\, h/$Mpc at low redshifts). It is interesting, however, to see that, despite these differences, cross-correlation results are still

able to deliver competitive constraints. This makes a strong case for cross-correlation studies, especially in light of the reduced challenges in terms of residual systematics from the 21 cm IM observations.

Adding the nuisance parameters, i.e. assuming no prior knowledge of the astrophysics at play, has the effect of varying the overall amplitude of the cross-correlation power spectrum. This translates into a broadening of the constraints, in particular on $A_s$. Moreover, the 2D contours are generally broader and show less clear correlations, except for the $H_0 - \Omega_c h^2$ and $H_0 - \Omega_b h^2$ planes. Although the shape is stretched, the $H_0 - \Omega_c h^2$ degeneracy is still marked.

In Fig. 5, we show the results on the full set of cosmological parameters for the combination of the two cross-correlation data sets and the 21 cm multipoles one. Note that we do not report the constraints on $\tau$, due to the fact that the considered probes are not sensitive to this parameter. We compare the results with the ones for the 21 cm multipoles and the $\hat{P}_{21,g}^{DESI}$ data set as a reference. In order to explore a more realistic scenario, we include the nuisance parameters. We observe that combining SKAO × DESI and SKAO × Euclid improves the constraints obtained with the two data sets separately. Including also the 21 cm multipoles lead to the best result. With observations from 21 cm probes only in the pessimistic case of including the nuisances, we are able to achieve constraints on the cosmological parameters comparable with Planck CMB observations.

We conclude that 21 cm IM observations in cross-correlation with galaxy clustering seem to have a reduced constraining power with respect to 21 cm auto-power spectrum measurements. However, when combined with the latter, they improve the constraints, showing that the cross-correlation signal carries complementary cosmological information.

**Table 2.** Marginalized percentage constraints on cosmological parameters at the 68% confidence level. We show the results obtained using different combinations of forecasted data sets. The label '$\hat{P}_0 + \hat{P}_2$' stands for the forecasted 21 cm power spectrum monopole and quadrupole observations (see Appendix A). '$\hat{P}_{21,g}^{Euclid}$' and '$\hat{P}_{21,g}^{DESI}$' refer to the mock cross-correlation power spectrum data sets constructed above. The label 'nuis.' indicates that we vary the nuisance parameters along with the cosmological ones.

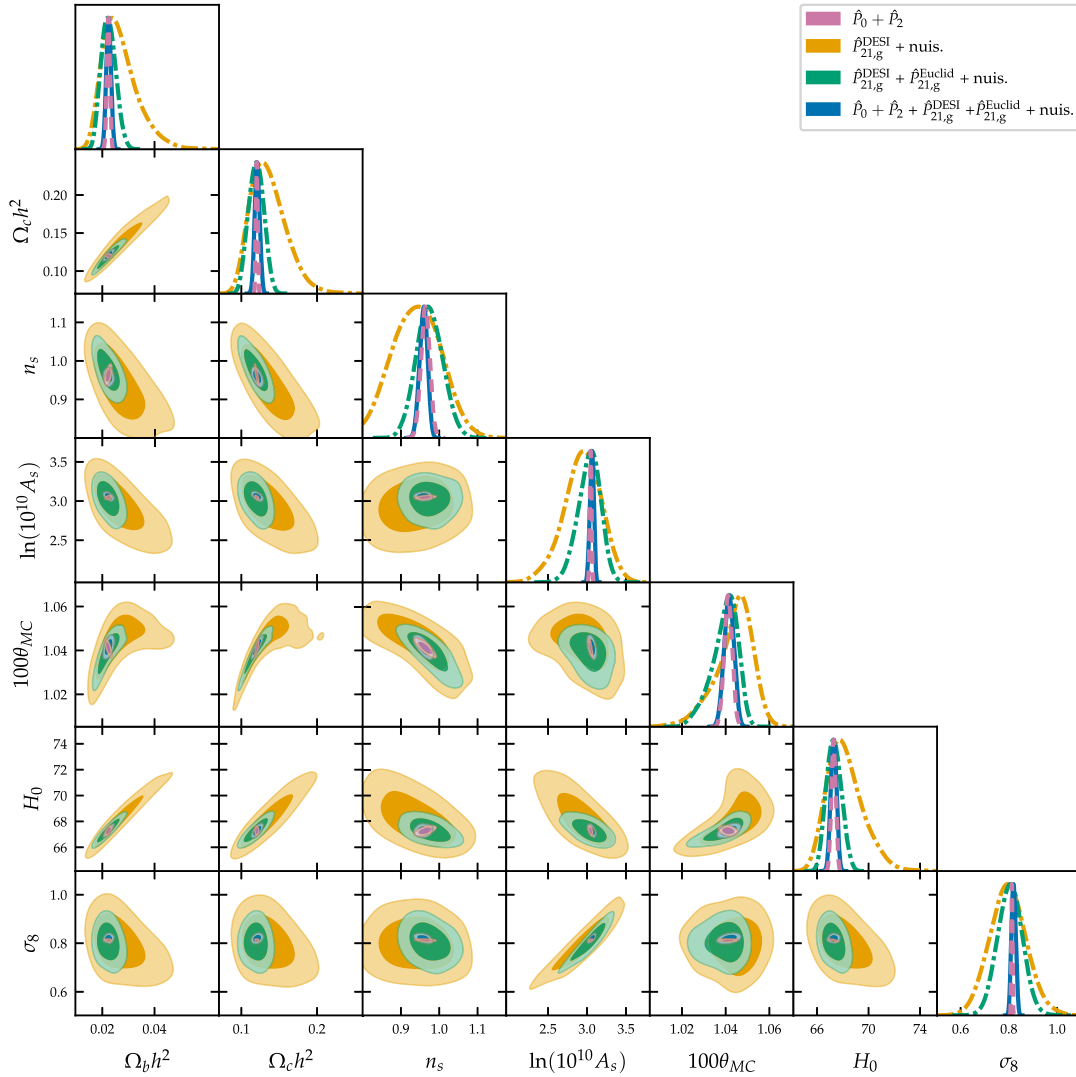| Parameter | $\hat{P}_0 + \hat{P}_2$ | $\hat{P}_{21,g}^{DESI}$ | $\hat{P}_{21,g}^{DESI}$ + nuis. | $\hat{P}_{21,g}^{Euclid}$ | $\hat{P}_{21,g}^{Euclid}$ + nuis. | $\hat{P}_{21,g}^{DESI} + \hat{P}_{21,g}^{Euclid}$ + nuis. | $\hat{P}_0 + \hat{P}_2 + \hat{P}_{21,g}^{DESI} + \hat{P}_{21,g}^{Euclid}$ + nuis. |
|---|---|---|---|---|---|---|---|
| $\Omega_b h^2$ | 2.59% | 6.43% | 23.11% | 5.78% | 16.99% | 12.52% | 3.89% |
| $\Omega_c h^2$ | 0.99% | 3.81% | 16.63% | 3.75% | 11.87% | 8.59% | 2.67% |
| $n_s$ | 1.19% | 2.43% | 6.79% | 1.82% | 4.59% | 3.56% | 1.08% |
| $\ln(10^{10}A_s)$ | 0.37% | 0.78% | 8.08% | 0.54% | 7.62% | 4.73% | 0.81% |
| $100\theta_{MC}$ | 0.17% | 0.39% | 0.75% | 0.30% | 0.62% | 0.54% | 0.21% |
| $H_0$ | 0.25% | 0.69% | 1.96% | 0.49% | 1.07% | 0.87% | 0.33% |
| $\sigma_8$ | 0.29% | 0.40% | 9.41% | 0.58% | 10.03% | 6.37% | 1.11% |



**Figure 5.** Joint constraints (68 per cent and 95 per cent confidence regions) and marginalized posterior distributions on a subset of the cosmological parameters. The label '$\hat{P}_0 + \hat{P}_2$' (dashed lines) stands for the forecasted 21 cm power spectrum monopole and quadrupole observations (see Appendix A). '$\hat{P}_{21,g}^{Euclid}$' and '$\hat{P}_{21,g}^{DESI}$' refer to the mock cross-correlation power spectrum data sets constructed above. The label 'nuis.' (dashed-dotted lines) indicates that we vary the nuisance parameters along with the cosmological ones. The relative constraints are listed in Table 2.

**Table 3.** Marginalized percentage constraints on cosmological parameters at the 68 per cent confidence level. We show the results obtained using different combinations of forecasted data sets in combination with Planck. The label 'Planck 2018' stands for TT, TE, EE + lowE + lensing, while the label '$\hat{P}_0 + \hat{P}_2$' refers to the forecasted 21 cm power spectrum monopole and quadrupole observations (see Appendix A). '$\hat{P}^{\text{Euclid}}_{21,\text{g}}$' and '$\hat{P}^{\text{DESI}}_{21,\text{g}}$' refer to the mock cross-correlation power spectrum data set constructed above. The label 'nuis.' indicates that we vary the nuisance parameters along with the cosmological ones.

| Parameter | Planck 2018 | $+ \hat{P}_0 + \hat{P}_2$ | $+ \hat{P}^{\text{DESI}}_{21,\text{g}}$ | $+ \hat{P}^{\text{DESI}}_{21,\text{g}} + \hat{P}_0 + \hat{P}_2$ + nuis. | $+ \hat{P}^{\text{Euclid}}_{21,\text{g}}$ | $+ \hat{P}^{\text{Euclid}}_{21,\text{g}}$ + nuis. | $+ \hat{P}^{\text{DESI}}_{21,\text{g}} + \hat{P}^{\text{Euclid}}_{21,\text{g}}$ + nuis. | $+ \hat{P}^{\text{DESI}}_{21,\text{g}} + \hat{P}^{\text{Euclid}}_{21,\text{g}} + \hat{P}_0 + \hat{P}_2$ + nuis. |
|---|---|---|---|---|---|---|---|---|
| $\Omega_b h^2$ | 0.64% | 0.47% | 0.48% | 0.52% | 0.48% | 0.50% | 0.49% | 0.46% |
| $\Omega_c h^2$ | 0.99% | 0.26% | 0.39% | 0.55% | 0.36% | 0.48% | 0.42% | 0.31% |
| $n_s$ | 0.42% | 0.28% | 0.33% | 0.34% | 0.32% | 0.32% | 0.32% | 0.30% |
| $\ln(10^{10} A_s)$ | 0.46% | 0.16% | 0.12% | 0.44% | 0.12% | 0.45% | 0.44% | 0.43% |
| $\tau$ | 13.44% | 5.98% | 5.39% | 12.09% | 5.50% | 11.82% | 11.89% | 11.38% |
| $100\theta_{\text{MC}}$ | 0.03% | 0.02% | 0.03% | 0.03% | 0.03% | 0.02% | 0.03% | 0.03% |
| $H_0$ | 0.79% | 0.13% | 0.25% | 0.41% | 0.21% | 0.33% | 0.28% | 0.14% |
| $\sigma_8$ | 0.73% | 0.24% | 0.24% | 0.69% | 0.24% | 0.70% | 0.69% | 0.67% |

## 4.2 Combining 21 cm × galaxy clustering with CMB observations

Most recent forecast analyses find 21 cm IM future observations to be a pivotal cosmological probe, highly complementary to CMB observations (SKA Cosmology SWG 2020). Indeed, in Berti et al. (2023), we found that observations of the 21 cm power spectrum multipoles contribute significantly to improving the constraints and reducing the degeneracies on the cosmological parameters. In this section, we investigate the effects of combining 21 cm and galaxy clustering cross-correlations with CMB measurements.

For consistency, we first run the Planck likelihood in our framework and reproduce constraints in agreement with the Planck 2018 results. We then study the effect of adding the $\hat{P}^{\text{Euclid}}_{21,\text{g}}$ and $\hat{P}^{\text{DESI}}_{21,\text{g}}$ data sets and the two combined. As in Section 4.1, we compare the results we obtain with the constraints from the 21 cm power spectrum multipoles.

Table 3 shows the percentage constraints for this analysis. We observe that adding $\hat{P}^{\text{Euclid}}_{21,\text{g}}$ or $\hat{P}^{\text{DESI}}_{21,\text{g}}$ to Planck 2018 data reduces the estimated constraints with respect to the Planck alone results. The effect is prominent for $\Omega_c h^2$ and $H_0$, for which the error is reduced by a factor of $\sim$3, and $A_s$, with a factor of $\sim$4 decrease. As one can see from Fig. 6, in the $H_0 - \Omega_c h^2$ plane the effect is ascribable to the correlation directions. Indeed, with Planck observations $H_0$ and $\Omega_c h^2$ are anti-correlated, while they are positively correlated with $\hat{P}_{21,\text{g}}$. Combining the two removes the degeneracy and reduces errors. The effect is also particularly evident for $A_s$ since the CMB probes the quantity $A_s \exp(-2\tau)$ and the matter power spectrum, which is constrained by 21 cm data is sensitive to $S_8$, which is in turn degenerate with the optical depth to re-ionization as measured from the CMB. Therefore, adding 21 cm data effectively removes the degeneracies.

When nuisance parameters are taken into account, as expected the improvement on the constraints is softened. In particular for $A_s$, and consequently $\sigma_8$, the constraining power is lost when the parameter space is open to the nuisances. Varying the nuisances corresponds to effectively changing the amplitude of the power spectrum and, thus, it results in worsened constraints on $A_s$.

The effects observed for the cross-correlation data sets combined with CMB are qualitatively comparable to the results obtained for the 21 cm power spectrum multipoles. This confirms that when combining different kinds of 21 cm observations with CMB data the improvement in the constraints is always driven mainly by the breaking of the degeneracy in the $\Omega_c h^2 - H_0$ plane. Indeed, our analysis reveals that even a less constraining measurement, such as the 21 cm and galaxy cross-correlation, is effective in improving the errors on $\Omega_c h^2$ and $H_0$ if it presents a sufficiently marked correlation among these parameters.

To better prove this point, in Fig. 6, we compare the effect of combining Planck data with the 21 cm multipoles, $\hat{P}^{\text{DESI}}_{21,\text{g}}$ and $\hat{P}^{\text{Euclid}}_{21,\text{g}}$, and the three combined. We find that even with the nuisances, results from $\hat{P}^{\text{DESI}}_{21,\text{g}} + \hat{P}^{\text{Euclid}}_{21,\text{g}}$ (orange contours) are similar to the constraints from the multipoles, for which, instead, the nuisances are kept fixed as a best-result reference (pink contours). The main difference resides in the loss of constraining power on $A_s$ and the related parameters, which is however ascribable to the inclusion of the nuisances. Further adding the 21 cm multipoles to $\hat{P}^{\text{DESI}}_{21,\text{g}} + \hat{P}^{\text{Euclid}}_{21,\text{g}}$ (green contours),[6] does not impact the constraining power

---

[6]Note that in this case also the 21 cm power spectrum nuisances are varied as in Berti et al. (2023).
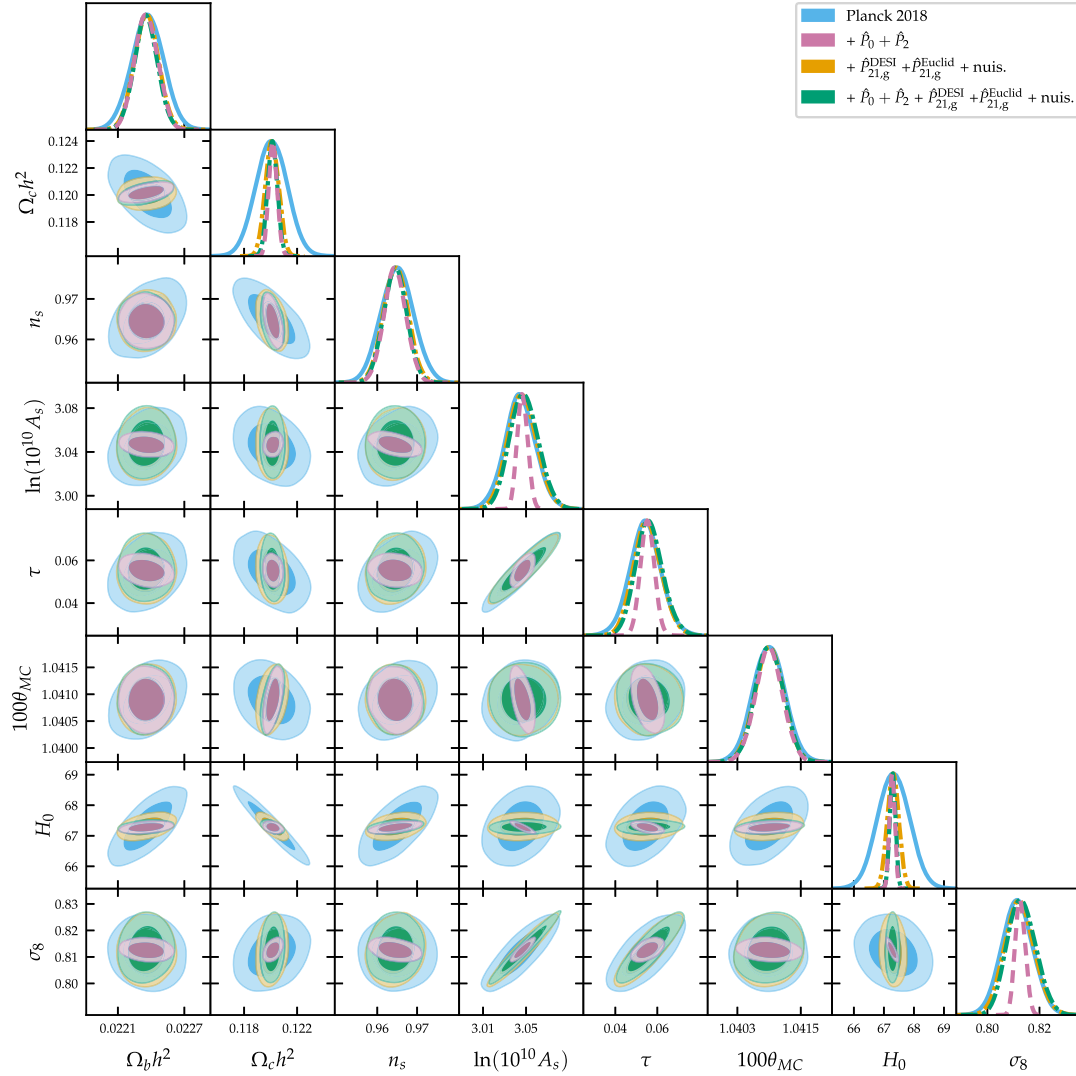
**Figure 6.** Joint constraints (68 per cent and 95 per cent confidence regions) and marginalized posterior distributions on a subset of the cosmological parameters. The label 'Planck 2018' stands for TT, TE, EE + lowE + lensing, while the label '$\hat{P}_0 + \hat{P}_2$' (dashed lines) stands for the forecasted 21 cm power spectrum monopole and quadrupole observations (see Appendix A). '$\hat{P}_{21,g}^{Euclid}$' and '$\hat{P}_{21,g}^{DESI}$' refer to the mock cross-correlation power spectrum data sets constructed above. The label 'nuis.' (dashed-dotted lines) indicates that we vary the nuisance parameters along with the cosmological ones. The relative constraints are listed in Table 3.

or the shape of the correlations. This confirms that the 21 cm probe is pivotal in breaking the CMB degeneracy in the $\Omega_c h^2 - H_0$ plane and the effect is relevant already at the level of cross-correlation or with SKAO precursor power spectrum measurements (see also Berti et al. 2022).

We conclude that cross-correlations measurements of 21 cm IM and galaxy clustering are a key cosmological probe complementary to CMB observations and, in combination with Planck, their forecasted constraining power is compatible with the one from 21 cm power spectrum multipole measurements with the SKAO.

### 4.3 State-of-the-art cosmological parameters constraints from the MeerKAT × WiggleZ detection

Cosmological 21 cm observations with the SKAO will be possible in the upcoming years when the Observatory will be fully operational.

However, the SKA-Mid pathfinder, MeerKAT is already taking data and its first cosmological surveys are promising. Recently, a power spectrum detection with the MeerKLSS survey, the IM survey with MeerKAT, in cross-correlation with galaxy clustering data has been made at the $7.7\sigma$ level (Cunnington et al. 2022). The analysis pipeline we develop in this work is constructed to be ready to use with real cross-correlation power spectrum measurements. Therefore, we decide to test our methodology on the published results available for MeerKLASS. In the following, we present the result we obtain on the cosmological parameters constraints. We refer the interested reader to Appendix B for the technical consistency checks we run on the adopted power spectrum model and the predicted signal-to-noise ratio.

We tune the parameters of the likelihood function to match the settings of the observed data. Instead of the SKAO specifications, we use the MeerKLASS survey parameters, i.e. we consider a 200 dg$^2$ survey area and dishes of a diameter of $D_{dish} = 13.5$ m. The observed
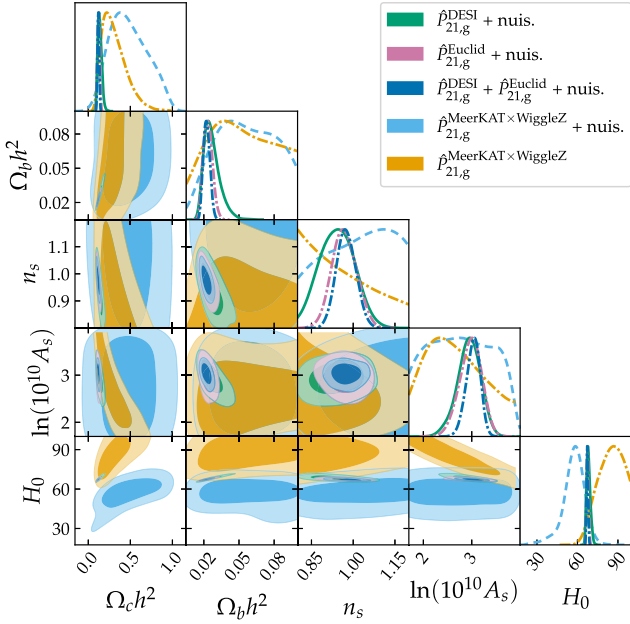
**Figure 7.** Joint constraints (68 per cent and 95 per cent confidence regions) and marginalized posterior distributions on a subset of the cosmological parameters. The labels '$\hat{P}_{21,g}^{\rm Euclid}$' and '$\hat{P}_{21,g}^{\rm DESI}$' refer to the mock cross-correlation power spectrum data sets constructed above. $\hat{P}_{21,g}^{\rm MeerKAT\times WiggleZ}$ refers to the cross-correlation power spectrum detection. The label 'nuis.' (dashed-dotted lines) indicates that we vary the nuisance parameters along with the cosmological ones. The relative constraints are listed in Table 4.

**Table 4.** Marginalized percentage constraints on cosmological parameters at the 68 per cent confidence level. We show the constraints on the cosmological parameters obtained with the MeerKAT × WiggleZ cross-correlation power spectrum detection. The label 'nuis.' indicates that we vary the nuisance parameters along with the cosmological ones. The symbol '—' stands for unconstrained.

| Parameter | $\hat{P}_{21,g}^{\rm MeerKAT\times WiggleZ}$ | $\hat{P}_{21,g}^{\rm MeerKAT\times WiggleZ}$ + nuis. |
|---|---|---|
| $\Omega_b h^2$ | — | — |
| $\Omega_c h^2$ | $0.314^{+0.079}_{-0.18}$ | $0.48^{+0.19}_{-0.27}$ |
| $100\theta_{MC}$ | $1.090^{+0.061}_{-0.083}$ | $1.051^{+0.085}_{-0.070}$ |
| $n_s$ | — | — |
| $\ln(10^{10}A_s)$ | — | — |
| $\tau$ | — | — |
| $H_0$ | $84^{+10}_{-7}$ | $57^{+8}_{-6}$ |
| $\sigma_8$ | $0.974^{+0.068}_{-0.092}$ | $1.04^{+0.30}_{-0.49}$ |

effective redshift is $z = 0.43$ with a bin width of $\Delta z = 0.059$. The signal is observed in cross-correlation with the WiggleZ 11 h galaxy survey (Drinkwater et al. 2010; Drinkwater et al. 2018). When we do not include the nuisance parameters, we use the measured galaxy bias value $b_{\rm g} = 0.911$ (Blake et al. 2011) and cross-correlation factor $r = 0.9$ (Khandai et al. 2011). Other parameters and theoretical predictions are left unchanged.

We present the cosmological parameters constraints resulting from our MCMC analysis in Fig. 7 and Table 4. We observe that the state-of-the-art constraining power is limited with respect to the results forecasted for SKAO × Euclid and SKAO × DESI, as expected due to the wider redshift ranges, probed scales, and survey area. Single bin *MeerKAT* observations are not yet able to constrain the

complete set of cosmological parameters. However, the degeneracies between the parameters match the ones expected from our forecasts. In particular, the $H_0 - \Omega_c h^2$ correlation is clearly visible, although much less prominent. From these real measurements, we can infer new information on the marginalized mean value of the cosmological parameters. We find that all the constraints are compatible with the Planck results.

The most constrained parameters are $\Omega_c h^2$ and $H_0$, proving that 21 cm observations will be most useful to constrain them and their derived parameters. When fixing the nuisances, we find a high central value for $H_0$, although with a large error. We believe that this is not a physical effect, but is rather coming from a mismatch between the assumed brightness temperature value in our analysis and the one that seems to describe the observed data (see Appendix B for a more in-depth discussion). The conservative result is then the one in which nuisances are taken into account. In this case, we find that *MeerKAT* × WiggleZ data prefer a lower value of $H_0$, although the significance is not high enough to draw firm conclusions.

From the constraints on the nuisances, one could estimate the value of $\Omega_{\rm HI}$. With our analysis we find the constraints on the nuisance parameters from real data to be too wide to infer a meaningful result.

Lastly, although we do not show here the results, we test the effect of combining *MeerKAT* × WiggleZ data with Planck 2018 observations. We find that the measured cross-correlation power spectrum does not increase significantly the constraining power of CMB observations, leaving the constraints and the 2D contours mostly unchanged.

Although the constraining power of real detection is not yet competitive with other probes, the quality of the current 21 cm IM observations in cross-correlation with galaxy clustering will improve sharply in the upcoming years and will soon become a useful independent cosmological probe. Moreover, the forecasted results for future surveys are very promising.

## 5 CONCLUSIONS

In this work, we forecast the constraints on the $\Lambda$CDM cosmological parameters for power spectrum measurements of 21 cm IM in cross-correlation with galaxy clustering. Modelling the cross-power spectrum as in Cunnington et al. (2022), we forecast mock observations of the SKAO cross-correlated with DESI and Euclid-like surveys. We test the constraining power of such probes alone and combined with the latest Planck CMB observations. Note that our modelling does not include possible residual foreground and systematics contamination.

We follow the SKAO Red Book (SKA Cosmology SWG 2020) proposal and simulate single-dish observations with the SKA-Mid telescope in Band 1 (frequency range 0.35–1.05 GHz). We cross-correlate this signal with a Euclid-like spectroscopic survey (Blanchard et al. 2020) and the DESI ELGs one (Casas et al. 2023; Raichoor et al. 2023) in the redshift range 0.7–1.7. Assuming a Planck 2018 fiducial cosmology, we construct two data sets of observations within multipole redshift bins. To test the constraining power on the cosmological parameters of our mock observations, we implement a likelihood function for the cross-correlation power spectrum, fully integrated with the MCMC sampler CosmoMC. We include a discussion on the impact of our lack of knowledge on the baryonic physics involved in the computation of the 21 cm power spectrum as nuisance parameters in the analysis.

We first focus on assessing the constraining power of cross-correlation observations alone, compared to the results we obtain for the 21 cm multipoles. We, then, combine the two to investigate

if they carry complementary information. The results of our analysis can be summarized as follows.

We find that SKAO power spectrum measurements in cross-correlation with galaxy clustering have a constraining power comparable to the 21 cm auto-power spectrum. The SKAO × DESI and SKAO × Euclid data sets we construct are able to constrain the cosmological parameters up to the sub- per cent level. They seem to be particularly effective on $H_0$, on which we obtain constraints between the 0.49 per cent and the 1.96 per cent from 21 cm and galaxy clustering cross-correlation alone. The tightest constraints are achieved when we combine 21 cm power spectrum multipoles with the cross-correlation mock observations, for which we obtain a 0.33 per cent constraint on $H_0$, a value that is competitive with Planck.

When combining the cross-correlation mock measurements with CMB data, we find that they are pivotal to reducing the errors on the cosmological parameters. The effect is particularly prominent for $\Omega_c h^2$ and $H_0$, for which the errors are reduced by a factor between 2.5–1.8 and 3.8–2, respectively. Again, the best result is obtained by combining all the 21 cm probes together. In this case, the error with respect to Planck alone results is reduced by a factor of 3.2 for $\Omega_c h^2$ and 5.6 for $H_0$, with the nuisance parameters taken into account.

Lastly, we test our analysis pipeline on the recent data for the cross-power spectrum between *MeerKAT*, the SKA-Mid pathfinder, and WiggleZ galaxy clustering Cunnington et al. (2022). We find that state-of-the-art observations have limited constraining power on the complete set of cosmological parameters. However, the main features of the marginalized constraints are compatible with the forecasted results of this work.

To conclude, our analysis supports the case of 21 cm and galaxy clustering cross-correlation measurements. In combination with CMB observations, cross-correlations will be able to provide competitive constraints on the cosmological parameters comparable to the ones obtained with the auto-power spectrum. The working pipeline presented in this work is found to be compatible and easily employable with real observations. The analysis we carry out can be straightforwardly adapted to forecast constraints on the neutrino mass and beyond $\Lambda$CDM models. These extensions are currently under study.

## ACKNOWLEDGEMENTS

## DATA AVAILABILITY

Access to the original code is available upon reasonable request to the corresponding author.

## REFERENCES

Aghamousa A. et al., 2016, The DESI Experiment Part II: Instrument Design . preprint (arXiv:1611.00037)
Alcock C., Paczynski B., 1979, Nature, 281, 358
Alonso D., Bull P., Ferreira P. G., Maartens R., Santos M., 2015, ApJ, 814, 145
Amiri M. et al., ., 2023, Astrophys. J., 947, 16
Anderson C. J. et al., 2018, MNRAS, 476, 3382
Ansari R. et al., 2012, A&A, 540, A129
Bandura K. et al., 2014, Proc. SPIE Int. Soc. Opt. Eng., 9145, 22
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Battye R. A., Browne I. W. A., Dickinson C., Heron G., Maffei B., Pourtsidou A., 2013, MNRAS, 434, 1239
Battye R. A., Davies R. D., Weller J., 2004, MNRAS, 355, 1339
Bernal J. L., Breysse P. C., Gil-Marín H., Kovetz E. D., 2019, Phys. Rev. D, 100, 123522
Berti M., Spinelli M., Haridasu B. S., Viel M., Silvestri A., 2022, J. Cosmol. Astropart. Phys., 2022, 018
Berti M., Spinelli M., Viel M., 2023, MNRAS, 521, 3221
Bharadwaj S., Nath B. B., Nath B. B., Sethi S. K., 2001, J. Astrophys. Astron., 22, 21
Blake C. et al., 2011, MNRS, 415, 2876
Blake C., 2019, MNRAS, 489, 153
Blanchard A. et al., 2020, A&A, 642, A191
Bull P. et al., 2016, Phys. Dark Univ., 12, 56
Carucci I. P., Irfan M. O., Bobin J., 2020, MNRAS, 499, 304
Casas S., Carucci I. P., Pettorino V., Camera S., Martinelli M., 2023, Phys. Dark Univ., 39, 101151
Castorina E., Villaescusa-Navarro F., 2017, MNRAS, 471, 1788
Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463
Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303
Cunnington S. et al., 2022, MNRAS, 518, 6262
Cunnington S., 2022, MNRAS, 512, 2408
Cunnington S., Irfan M. O., Carucci I. P., Pourtsidou A., Bobin J., 2021, MNRAS, 504, 208
Cunnington S., Pourtsidou A., Soares P. S., Blake C., Bacon D., 2020, MNRAS, 496, 415
D'Amico G., Gleyzes J., Kokron N., Markovic K., Senatore L., Zhang P., Beutler F., Gil-Marín H., 2020, JCAP, 2020, 005
Drinkwater M. J. et al., 2010, MNRAS, 401, 1429
Drinkwater M. J. et al., 2018, MNRAS, 474, 4151
Furlanetto S., Oh S. P., Briggs F., 2006, Phys. Rept., 433, 181
Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2017, MNRAS, 465, 1757
Gilks W., Richardson S., Spiegelhalter D., 1995, in Press C.R.C., Chapman and Hall/CRC Interdisciplinary Statistics, Markov Chain Monte Carlo in Practice. Taylor and Francis, p. 512, Available at: https://books.google.it/books?id=TRXrMWY_i2IC
Hand N., Seljak U., Beutler F., Vlah Z., 2017, J. Cosmol. Astropart. Phys., 2017, 009
Hu W., Wang X., Wu F., Wang Y., Zhang P., Chen X., 2020, MNRAS, 493, 5854
Irfan M. O. et al., 2022, MNRAS, 509, 4923
Irfan M. O., Bull P., 2021, MNRAS, 508, 3551
Jiang Y.-E. et al., 2023, Res. Astron. Astrophys., 23, 075003
Jolicoeur S., Maartens R., Dlamini S., 2023, Eur. Phys. J. C, 83, 320
Kaiser N., 1987, MNRAS, 227, 1
Karagiannis D., Maartens R., Randrianjanahary L., 2022, J. Cosmol. Astropart. Phys., 11, 03
Karagiannis D., Slosar A., Liguori M., 2020, J. Cosmol. Astropart. Phys., 2020, 052
Khandai N., Sethi S. K., Di Matteo T., Croft R. A. C., Springel V., Jana A., Gardner J. P., 2011, MNRAS, 415, 2580
Kovetz E. D. et al., 2017, Line-Intensity Mapping: 2017 Status Report . preprint (arXiv:1709.09066)
Lewis A., 2013, Phys. Rev. D, 87, 103529
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Masui K. W. et al., 2013, ApJ, 763, L20

Matshawule S. D., Spinelli M., Santos M. G., Ngobese S., 2021, MNRAS, 506, 5075

McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815

Mead A., Heymans C., Lombriser L., Peacock J., Steele O., Winther H., 2016, MNRAS, 459, 1468

Newburgh L. B. et al., 2016, Proc. SPIE Int. Soc. Opt. Eng., 9906, 99065X

Obuljen A., Castorina E., Villaescusa-Navarro F., Viel M., 2018, J. Cosmol. Astropart. Phys., 2018, 004

Padmanabhan H., Maartens R., Umeh O., Camera S., 2023, The HI intensity mapping power spectrum: insights from recent measurements . preprint (arXiv:2305.09720)

Paul S., Santos M. G., Chen Z., Wolz L., 2023, preprint (arXiv:2301.11943)

Planck Collaboration III, 2020, A&A, 641, A3

Planck Collaboration V, 2020, A&A, 641, A5

Planck Collaboration VI, 2020, A&A, 641, A6

Pritchard J. R., Loeb A., 2012, Rep. Prog. Phys., 75, 086901

Raichoor A. et al., 2023, AJ, 165, 126

Santos M. G. et al., 2015, PoS, AASKA14, 019

Santos M. G. et al., 2017, in MeerKAT Science: On the Pathway to the SKA. preprint (arXiv:1709.06099)

Seo H.-J., Dodelson S., Marriner J., Mcginnis D., Stebbins A., Stoughton C., Vallinotto A., 2010, ApJ, 721, 164

SKA Cosmology SWG., 2020, Publ. Astron. Soc. Aust., 37, e007

Smith R. E. et al., 2003, MNRAS, 341, 1311

Smith R. E., 2009, MNRAS, 400, 851

Soares P. S., Cunnington S., Pourtsidou A., Blake C., 2021, MNRAS, 502, 2549

Soares P. S., Watkinson C. A., Cunnington S., Pourtsidou A., 2022, MNRAS, 510, 5872

Spinelli M., Carucci I. P., Cunnington S., Harper S. E., Irfan M. O., Fonseca J., Pourtsidou A., Wolz L., 2021, MNRAS, 509, 2048

Spinelli M., Zoldan A., De Lucia G., Xie L., Viel M., 2020, MNRAS, 493, 5434

Vargas-Magaña M., Brooks D. D., Levi M. M., Tarle G. G., 2018, in 53rd Rencontres de Moriond on Cosmology. p. 11, preprint (arXiv:1901.01581)

Viljoen J.-A., Fonseca J., Maartens R., 2020, J. Cosmol. Astropart. Phys., 2020, 054

Villaescusa-Navarro F. et al., 2018, ApJ, 866, 135

Villaescusa-Navarro F., Alonso D., Viel M., 2017, MNRAS, 466, 2736

Villaescusa-Navarro F., Viel M., Alonso D., Datta K. K., Bull P., Santos M. G., 2015, J. Cosmol. Astropart. Phys., 2015, 034

Wang J. et al., 2021, MNRAS, 505, 3698

Wolz L. et al., 2022, MNRAS, 510, 3495

Wolz L., Tonini C., Blake C., Wyithe J. S. B., 2016, MNRAS, 458, 3399

Wu P.-J., Li Y., Zhang J.-F., Zhang X., 2023, Sci. China Phys. Mech. Astron., 66, 270413

# APPENDIX A: ON THE IMPACT OF INCLUDING THE ALCOCK–PACZYNSKI EFFECTS

With respect to the analysis in Berti et al. (2023), in this work, we extended our likelihood code to include the Alcock–Paczynski distortions, which are used in other works (e.g. Bernal et al. 2019; Soares et al. 2021; Casas et al. 2023). We neglected AP effects in the first approximation because we assumed to know the true cosmology, given that it is the one we input when constructing the mock data set. In this section, we give an overview of how the cosmological parameter constraints are affected by the addition of AP effects.

Contrary to what we naively expected, implementing the AP modifications significantly improves the constraints. In the upper panel of Fig. A1, we compare the effects of different mock data sets. Our reference (orange lines and contours) is the 21 cm power spectrum monopole ($\hat{P}_0$) and quadrupole ($\hat{P}_2$) mock data set that we construct in Berti et al. (2023). This data set forecasts SKAO observations in multipole redshift bins in the range 0–3, i.e. for
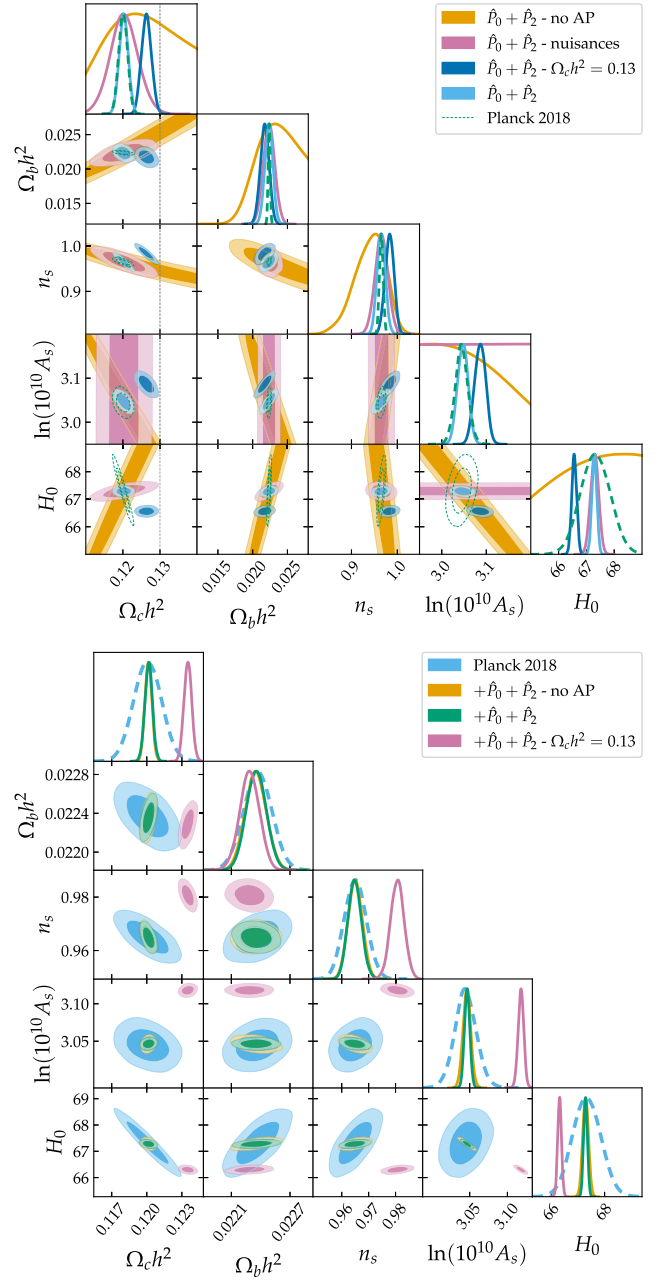


**Figure A1.** Joint constraints (68 per cent and 95 per cent confidence regions) and marginalized posterior distributions on a subset of the cosmological parameters. We show results for the 21 cm multipoles alone (upper panel) and combined with CMB observations (lower panel). The label 'Planck 2018' (dashed lines) stands for TT, TE, EE + lowE + lensing, while the label '$\hat{P}_0 + \hat{P}_2$' stands for the forecasted 21 cm non-linear power spectrum monopole and quadrupole observations, with and without ('no AP') AP effects taken into account. '$\hat{P}_0 + \hat{P}_2 - \Omega_c h^2 = 0.13$' labels the results obtained for a mock data set with a value of $\Omega_c h^2$ different from the assumed fiducial cosmology. The label 'nuis.' indicates that we vary the nuisance parameters along with the cosmological ones.

six bins centred at {0.25, 0.75, 1.25, 1.75, 2.25, 2.75} and with a width of $\Delta z = 0.5$. The nuisances parameters for this data set are $\bar{T}_b b_{\rm HI} \sigma_8$, $\bar{T}_b f \sigma_8$, and the HI shot noise, for the non-linear 21 cm power spectrum.

Using the exact same framework, but implementing also the AP distortions of the amplitude and of the wave vectors as described in Section 2.2 (light blue contours), we find a crucially improved constraining power. E.g., for $\Omega_c h^2$ with only 21 cm observations, we recover the Planck constraint (dashed green lines and contours). On $H_0$, instead, we find even better constraints than Planck. When adding the nuisances (pink contours), the improvement is reduced, although still significantly better than the no AP case. We believe that the extra dependence on $H(z)$ that is introduced in the observable with the AP modifications is the cause of the improved constraining power.

Dealing with mock observations fabricated by ourselves, we have the advantage of knowing the true cosmology. We, thus, further test the impact of AP by creating a data set with a different value of $\Omega_c h^2 = 0.13$. We, however, do not change the fiducial cosmology, for which $\Omega_c h^2 = 0.12011$. Running the MCMC analysis on this data set we find consistent results. The $\Omega_c h^2$ constraint is pushed towards the true value, resulting in a constraint in between the true value and the assumed fiducial one. The errors, instead, are left unchanged, although the 2D contours are rigidly shifted. Thus, it seems that assuming the wrong cosmology impacts only the mean marginalized values of the parameters.

However, the smoking gun of having assumed the wrong cosmology is the probability distribution of the AP parameters themselves. Although we do not show the results here, we implemented the time-dependent $\alpha_\perp$ and $\alpha_\parallel$ as derived parameters and computed the marginalized constraints. We find that when the true cosmology matches the fiducial one, the $\alpha_\perp$, and $\alpha_\parallel$ marginalized posteriors are centred around one. When, instead, the true cosmology is not the assumed one, although still compatible with one the constraints, both 1D and 2D, are clearly shifted. Thus, one can test their assumptions by looking at the AP parameters constraints.

The lower panel of Fig. A1 shows the results for the same exercise, but combining 21 cm observations with Planck data.

We conclude that even when the whole set of cosmological parameters is used, the AP distortions are crucial not only to take into account our lack of knowledge of the true underlying cosmology but also to increase the constraining power on the cosmological parameters in matter power spectrum dependent probes as the 21 cm IM one.

## APPENDIX B: TEST ON THE MOCK DATA CONSTRUCTION PROCEDURE

To test the procedure, we follow to construct the mock data sets, we compare our predictions to the measured cross-correlation data published in Cunnington et al. (2022).

As in Section 4.3, we adjust the parameters of our formalism to mimic MeerKAT observations in the redshift bin centred at $z = 0.43$ and with $\Delta z = 0.059$. We find that with our pipeline we predict fewer $k$-bin in a wider scale range and a slightly different value for the brightness temperature, due to a small difference in the theory-predicted and measured value of $\Omega_{HI}$. Correcting for the brightness temperature results in cross-correlation power spectrum values more in agreement with observations (Fig. B2). However, this is not enough to reproduce the observed signal-to-noise ratio. Indeed, we find that varying the brightness temperature changes the power spectrum and the errors consistently, not impacting the signal-to-noise. Instead, adjusting the $k$-bin width to match the one in Cunnington et al. (2022) is enough to better reproduce the observed signal-to-noise ratio, as shown in Fig. B1.

We conclude that, compared to the state of the art, the pipeline we adopt in this work is consistent with real observational data. We
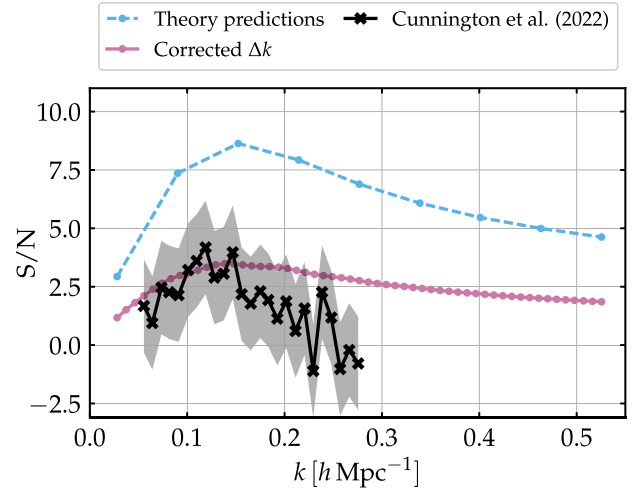


**Figure B1.** Signal-to-noise ratio as a function of $k$. We compare real data observations ('Cunnington et al. (2022)'), with the signal-to-noise predicted by the formalism adopted in this work. The grey shaded area shows the $2\sigma$ region for the observed signal to noise.
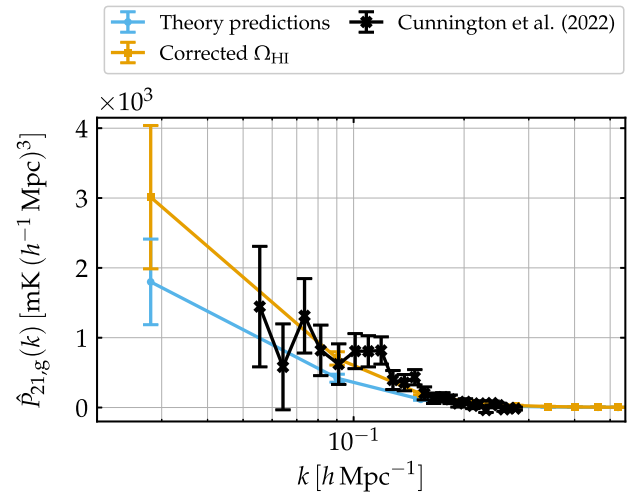


**Figure B2.** Observed and predicted 21 cm *MeerKAT* observations in cross-correlation with WiggleZ galaxy clustering.

are, however, more optimistic about the accessible scales and bin widths. Differences in the predicted power spectrum amplitude, i.e. the brightness temperature, are taken into account when opening the parameter space to the nuisance parameters.

## APPENDIX C: ON THE IMPACT OF FOREGROUND CONTAMINATION

The aim of the work presented in this paper is the development and testing of a pipeline for future 21 cm and galaxy clustering cross-correlation assuming the ideal scenario where all the systematics in the data have been already treated. It is, however, interesting to explore the effect of the key challenge of IM i.e. foreground contamination.

We follow e.g. Karagiannis, Slosar & Liguori (2020) and mimic the effect of the foreground as a loss of observed large scales by
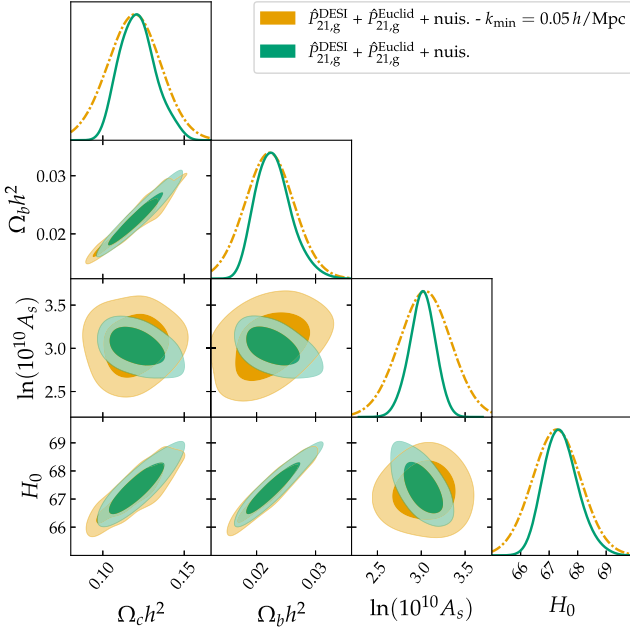
**Figure C1.** Joint constraints (68 per cent and 95 per cent confidence regions) and marginalized posterior distributions on a subset of the cosmological parameters. We compare the results presented in Section 4.1 (green, continuous lines) with what is obtained mimicking the effect of the foregrounds by cutting the data set at $k_{\min} = 0.05 \, h/\text{Mpc}$ (yellow, dashed lines). The label 'nuis.' indicates that we vary the nuisance parameters along with the cosmological ones.

cutting our mock data set at $k_{\min} = 0.05 \, h/\text{Mpc}$. Fig. C1 shows the comparison between the result presented above and the scale-cut data set for a reference case. As expected, foreground removal worsens the constraints on the cosmological parameter. However, it is worth noticing that the correlation between $\Omega_c h^2$ and $H_0$ is left unchanged.

As a consequence, when combining the scale-cut cross-correlation mock data set with measured Planck data, we observe that the same results are achieved with or without accounting for foreground effects. We can then conclude that, although real data could be affected by more complex systematics, the loss of some of the large scale in the cross-power spectrum does not prevent the increase in constraining power obtained combined with Planck results.

## APPENDIX D: COMPARISON WITH GALAXY CLUSTERING

In Section 4.1, we compare results on the cosmological parameter constraints from the cross-correlation of 21 cm and galaxy clustering observations with the ones obtained from 21 cm auto-power spectrum measurements. For completeness, in what follows we also forecast the constraining power of galaxy clustering auto-power spectrum measurements.

The pipeline we built for the 21 cm observables can be extended to include the analysis of galaxy clustering auto-power spectrum monopole $P_g(k, z)$ by implementing an additional likelihood function. For this purpose, we construct a toy data set of future DESI and Euclid-like observations keeping the same framework of observed redshifts, scales and volumes presented in Section 2. We highlight

that a more realistic forecast that includes the full constraining power of $P_g(k, z)$ measurements goes beyond the goal of this exercise.
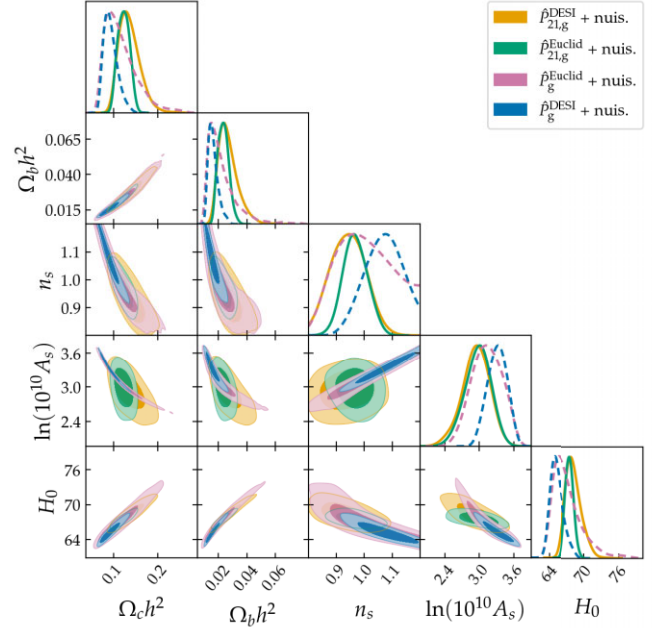


**Figure D1.** Joint constraints (68 per cent and 95 per cent confidence regions) and marginalized posterior distributions on a subset of the cosmological parameters. We compare the results presented in Section 4.1 from measurements of the 21 cm and galaxy clustering cross-correlation power spectrum $\hat{P}_{21,g}$ (continuous lines) with the ones from galaxy clustering auto-power spectrum $\hat{P}_g$ (dashed lines). The label 'nuis.' indicates that we vary the nuisance parameters along with the cosmological ones.

As in Section 2, our galaxy clustering data set is constructed from a theory predictions for the galaxy clustering auto-power spectrum $\hat{P}_g(k, z)$ (equation 8) and using the survey specifications for values of the bias and the shot noise (Section 3.1). Following the formalism used for the 21 cm auto-power spectrum (Berti et al. 2022), errors on each single data point are estimated as

$$\hat{\sigma}_g(k) = \frac{1}{\sqrt{2 N_{\text{modes}}(k)}} \sqrt{\int_{-1}^{1} d\mu \left( \hat{P}_g(k) + \frac{1}{\bar{n}_g} \right)^2}.$$

We stress that to be consistent with our settings, we consider limited sky volumes and do not investigate fully non-linear scales. We thus expect to obtain with our mock data set pessimistic constraints with respect to the full potential of stage IV surveys.

Our results for the mock Euclid and DESI-like observations are compared with results from cross-correlations in Fig. D1. We find that, in our framework, galaxy clustering provides constraints qualitatively comparable with cross-correlation. The main differences are noticeable in the shape of the 2D contours: we find stronger degeneracies in the galaxy clustering case between the cosmological parameters, while with cross-correlation some of these degeneracies are partially removed. As expected, the marked $\Omega_c h^2 - H_0$ degeneracy is present also for galaxy clustering alone, due to the fact that it derives from a measure of the matter power spectrum shape (see Berti et al. 2023).

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