Inferring the astrophysics of reionization and cosmic dawn from galaxy luminosity functions and the 21-cm signal

Jaehong Park [1]★ Andrei Mesinger,[1] Bradley Greig[2,3] and Nicolas Gillet [1]

[1]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[2]*ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D), University of Melbourne, Parkville, Melbourne, VIC 3010, Australia*
[3]*School of Physics, The University of Melbourne, Parkville, Melbourne, VIC 3010, Australia*

ABSTRACT

The properties of the first galaxies, expected to drive the Cosmic Dawn and the Epoch of Reionization, are encoded in the 3D structure of the cosmic 21-cm signal. Parameter inference from upcoming 21-cm observations promises to revolutionize our understanding of these unseen galaxies. However, prior inference was done using models with several simplifying assumptions. Here we introduce a flexible, physically motivated parametrization for high-z galaxy properties, implementing it in the public code 21CMFAST. In particular, we allow their star formation rates and ionizing escape fraction to scale with the masses of their host dark matter haloes, and directly compute inhomogeneous, sub-grid recombinations in the intergalactic medium. Combining current *Hubble* observations of the rest-frame UV luminosity function (UV LF) at high-z with a mock 1000-h 21-cm observation using the Hydrogen Epoch of Reionization Arrays, we constrain the parameters of our model using a Monte Carlo Markov Chain sampler of 3D simulations, 21CMMC. We show that the amplitude and scaling of the stellar mass with halo mass are strongly constrained by LF observations, while the remaining galaxy properties are constrained mainly by 21-cm observations. The two data sets compliment each other quite well, mitigating degeneracies intrinsic to each observation. All eight of our astrophysical parameters are able to be constrained at the level of $\sim$10 per cent or better. The updated versions of 21CMFAST and 21CMMC used in this work are publicly available.

Key words: galaxies: high-redshift – intergalactic medium – dark ages, reionization, first stars – diffuse radiation – early Universe – cosmology: theory.

1 INTRODUCTION

The birth of the first luminous sources in our Universe heralded the end of the cosmic Dark Ages. This so-called Cosmic Dawn (CD) culminated in the final phase transition of hydrogen in the intergalactic medium (IGM): the Epoch of Reionization (EoR). Understanding these cosmic epochs is key to understanding the properties of the first structures of our Universe. Unfortunately, it is likely that the bulk of the first galaxies are too faint to be observed directly, even with upcoming space-based telescopes such as James Webb Space Telescope (JWST; Gardner et al. 2006). Luckily, these unseen objects can be studied indirectly through their imprint in the IGM, using the redshifted 21-cm line. The 21-cm line from neutral hydrogen can map the ionization and thermal state of the IGM well into the infancy of the CD, making it a revolutionary probe of the early Universe (e.g. Hogan & Rees 1979; Scott & Rees 1990; Gnedin & Ostriker 1997; Madau, Meiksin & Rees 1997; Shaver

et al. 1999; Tozzi et al. 2000; Gnedin & Shaver 2004; Furlanetto, Oh & Briggs 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012).

For the past decade, extensive efforts to detect the 21-cm signal have been made. These include global (average) signal experiments such as the Experiment to Detect the Global EoR Signature (EDGES;[1] Bowman & Rogers 2010), the Shaped Antenna measurement of the background RAdio Spectrum (SARAS;[2] Patra et al. 2013), Sonda Cosmológica de las Islas para la Detección de Hidrógeno Neutro (SCI-HI; Voytek et al. 2014), the Large-aperture Experiment to detect the Dark Age (LEDA;[3] Price et al. 2018), and Broadband Instrument for Global HydrOgen ReioNisation Signal (BIGHORNS; Sokolowski et al. 2015). Indeed, Bowman et al. (2018) recently detected a feature in the global signal at $z \approx 17$,

---

[1]http://loco.lab.asu.edu/edges https://www.haystack.mit.edu/ast/arrays/Edges
[2]http://www.rri.res.in/DISTORTION/saras.html
[3]http://www.tauceti.caltech.edu/leda/

★ E-mail: jaehong.park@sns.it

though concerns remain about its interpretation (e.g. Hills et al. 2018). Ongoing interferometer experiments aiming to detect the power spectrum of the signal include the Murchison Wide Field Array (MWA;[4] Bowman et al. 2013; Tingay et al. 2013), Low Frequency Array (LOFAR;[5] van Haarlem et al. 2013), and Precision Array for Probing Epoch of Reionisation (PAPER;[6] Parsons et al. 2010). In the near future, next generation interferometers, such as Hydrogen Epoch of Reionization Array (HERA;[7] DeBoer et al. 2017) and Square Kilometre Array (SKA;[8] Mellema et al. 2013; Koopmans et al. 2015), will allow us to measure the spatial fluctuations of the 21-cm signal over a wider range of redshift with higher signal to noise. With these instruments, we should eventually have 3D maps of the first billion years of our Universe!

This upcoming wealth of 21-cm data will allow us to constrain the bulk properties of the underlying galaxies at a hitherto unseen level of precision. Current EoR observations can provide some insight into the general evolution of reionization (e.g. Choudhury & Ferrara 2006; Mitra, Choudhury & Ferrara 2011; Bouwens et al. 2015b; Price, Trac & Cen 2016; Greig & Mesinger 2017a; Gorce et al. 2018). However, the sheer volume of upcoming 21-cm data, and the fact that the signal is sensitive to both the ionization and thermal state of the IGM, could usher in a new era of precision astrophysical cosmology using standard Bayesian frameworks. Bayesian 21-cm parameter inference has been made using on-the-fly sampling of 3D simulations (Greig & Mesinger 2015, 2017b; Greig & Mesinger 2018), as well as interpolating simulations over a parameter grid (Mesinger, Ewall-Wice & Hewitt 2014; Pober et al. 2014; Liu et al. 2016; Mesinger, Greig & Sobacchi 2016; Fialkov et al. 2017; Hassan et al. 2017). Recently, neural networks have also been used to predict parameters from 21-cm power spectra (PS; Shimabukuro & Semelin 2017), emulate simulations to bypass on-the-fly Monte Carlo Markov Chain (MCMC) sampling (Kern et al. 2017; Schmit & Pritchard 2018), and to directly recover parameters from 21-cm images (Gillet et al. 2018).

However, these early 21-cm inference studies made several simplifying assumptions about the properties of galaxies and IGM structures. For example, most studies assume that the stellar mass of galaxies scales linearly with the mass of the host halo (though see e.g. Hassan et al. 2017), and/or that the ionizing escape fraction is a constant. Analytical approaches (e.g. Mason, Trenti & Treu 2015; Mutch et al. 2016; Sun & Furlanetto 2016) based on observations of luminosity functions (LFs), as well as hydrodynamic simulations (e.g. Gnedin & Kaurov 2014; Paardekooper, Khochfar & Dalla Vecchia 2015; Ocvirk et al. 2016; Xu et al. 2016; Kimm et al. 2017; Katz et al. 2018), suggest that both of these quantities have a more complex dependence on the halo properties. Prior studies also made simplifying assumptions about the role of IGM recombinations and how feedback suppresses star formation in small-mass haloes.

Motivated by observations of high-redshift galaxy LFs, here we generalize the astrophysical parametrization used in the 21-cm modelling code 21CMFAST.[9] We allow both the stellar mass and the escape fraction to have a power-law scaling with the mass of the host dark matter halo. Moreover, we directly compute sub-grid, inhomogeneous recombinations, following the approach

of Sobacchi & Mesinger (2014), removing the often-used yet ad hoc ionizing photon horizon parameter, $R_{\rm mfp}$. The resulting eight-parameter astrophysical model is both physically motivated and flexible enough to accommodate a large variety of galaxy formation scenarios.

We show how current LF observations strongly inform the scaling of star formation rate (SFR) with halo mass; however, they leave most of the remaining galaxy parameters unconstrained even with the addition of reionization observables from the CMB and QSO spectra. Using an MCMC sampler of 3D simulations, 21CMMC,[10] we present parameter forecasts for HERA as an upcoming 21-cm interferometer in combination with current LF observations. We show how the strong synergy between the two observations can result in most of the astrophysical parameters being constrained to the level of ∼10 per cent or better.

The outline of this paper is as follows. We begin in Section 2 by describing the astrophysical model including our new empirical parametrization of the galaxy properties. In Section 3, we present the UV LFs resulting from our model. In Section 4, we compute a mock 21-cm observation for a fiducial parameter choice. We briefly summarize our MCMC sampler of 3D simulations, 21CMMC, in Section 5. In Section 6, we show the resulting constraints on our astrophysical parameters, using the observed UV LFs and the mock 21-cm signal, both individually and combining the data sets. Finally, we summarize our results in Section 7. We assume a standard ΛCDM cosmology based on the *Planck* 2016 result (Planck Collaboration XIII 2016a): $(h, \Omega_{\rm m}, \Omega_{\rm b}, \Omega_{\Lambda}, \sigma_8, n_{\rm s}) = (0.678, 0.308, 0.0484, 0.692, 0.815, 0.968)$. Unless stated otherwise, we quote all quantities in comoving units.

2 ASTROPHYSICAL MODEL

In this section, we introduce a new parametrization for the SFR, ionizing escape fraction, and their scaling with the mass of the host dark matter haloes. We stress that our simple model does not directly follow individual galaxy evolution, making it only applicable for *an ensemble average* of the galaxy population, residing in haloes of a given mass. We note that only ≳10 Mpc 21-cm structures (i.e. ionized and heated regions) are large enough to be detected even with SKA; these structures likely form with the combined effort of ∼100–1000 sources. Therefore, the implicit ensemble averaging below is reasonably well justified.

2.1 Galaxy UV properties

We start with the common assumption that the stellar mass of a galaxy, M_*, can be related to the mass of the host halo, $M_{\rm h}$ (Kuhlen & Faucher-Giguère 2012; Dayal et al. 2014; Behroozi & Silk 2015; Mitra, Roy Choudhury & Ferrara 2015; Mutch et al. 2016; Sun & Furlanetto 2016; Yue, Ferrara & Xu 2016):

$$M_*(M_{\rm h}) = f_* \left(\frac{\Omega_{\rm b}}{\Omega_{\rm m}} \right) M_{\rm h}, \qquad (1)$$

where f_* is the fraction of galactic gas in stars. Consistent with observations of the faint galaxy population (e.g. see Behroozi & Silk 2015), we take f_* to have a power-law dependence on the dark

[4]https://www.haystack.mit.edu/ast/arrays/mwa/

[5]http://www.lofar.org

[6]http://eor.berkeley.edu

[7]http://reionization.org

[8]https://astronomers.skatelescope.org

[9]https://github.com/andreimesinger/21cmFAST

[10]https://github.com/BradGreig/21CMMC

matter halo mass[11]:

$$f_*(M_h) = f_{*,10} \left(\frac{M_h}{10^{10}\,\mathrm{M_\odot}} \right)^{\alpha_*}, \tag{2}$$

where $f_{*,10}$ is the fraction of galactic gas in stars normalized to the value in haloes of mass $10^{10}\,\mathrm{M_\odot}$ and α_* is the power-law index. We impose a physical upper limit of $f_* \leq 1$.

We assume that the SFR can be expressed on average as the total stellar mass divided by a characteristic time-scale[12]:

$$\dot{M}_*(M_h, z) = \frac{M_*}{t_* H(z)^{-1}}, \tag{3}$$

where $H(z)^{-1}$ is the Hubble time and the star formation time-scale, t_*, is a free parameter that we allow to vary between zero and unity.[13]

Similarly, we allow the escape fraction, f_{esc}, to scale with halo mass according to

$$f_{esc}(M_h) = f_{esc,10} \left(\frac{M_h}{10^{10}\,\mathrm{M_\odot}} \right)^{\alpha_{esc}}, \tag{4}$$

where $f_{esc,10}$ is the normalization of the ionizing UV escape fraction and α_* is the power-law scaling of f_{esc} with halo mass. α_{esc} is likely to be negative, as SNe can more easily evacuate low-column density channels from shallower DM potentials; the escape of ionizing photons is thought to be determined by the covering fraction of these low-column density channels (e.g. Paardekooper et al. 2015; Xu et al. 2016; Katz et al. 2018). As for the stellar fraction, we impose a physical upper limit of $f_{esc} \leq 1$.

Star formation in small galaxies is expected to be quenched due to SNe feedback, photoheating feedback, or inefficient gas accretion (e.g. Shapiro, Giroux & Babul 1994; Giroux, Sutherland & Shull 1994; Hui & Gnedin 1997; Barkana & Loeb 2001; Springel & Hernquist 2003; Mesinger & Dijkstra 2008; Okamoto, Gao & Theuns 2008; Sobacchi & Mesinger 2013a,b). We account for this suppression with a redshift-independent duty cycle.

$$f_{duty}(M_h) = \exp\left(-\frac{M_{turn}}{M_h} \right). \tag{5}$$

The duty cycle in our model can be thought of as the fraction of haloes of a given mass, which host stars/galaxies. It is unity for larger halo masses, but as one approaches smaller haloes, not all of them will be hosting galaxies. A fraction $[1 - \exp(-M_{turn}/M_h)]$ of haloes are unable to host a star-forming galaxy, likely because they have not accreted gas efficiently or because a negligible amount of prior star formation sterilized the halo through feedback. The remaining $\exp(-M_{turn}/M_h)$ of haloes manage to host stars with an efficiency

---

[11]For the purposes of modelling reionization, we do not care about the massive haloes that host AGNs bright enough to quench star formation. These haloes are far too rare at high redshifts to contribute to reionization (see e.g. Bouwens et al. 2015b).

[12]Applying the chain rule, $dM_*/dt = (dM_*/dM_h)(dM_h/dt)$, and noting that the Hubble time scales with cosmic time during matter domination, it can be shown that our model implies a halo growth rate of $\dot{M}_h \propto M_h^{\alpha_*} t^{-1}$. In the future, we will consider relaxing this relation, allowing an arbitrary scaling with time, if this is motivated by high-z data.

[13]Note that this is the only redshift dependence of our model. All other parameters are assumed to be functions of only the halo mass. We note that having the star formation scale with the Hubble time is analogous to having it scale with the dynamical time of DM haloes, $t_{dyn} \sim 1/\sqrt{G\bar\rho} \sim 1/\sqrt{180G\bar\rho(z)}$, where ~180 is the mean overdensity of a halo in the spherical collapse model and $\bar\rho(z)$ is the background density. Since at high-z the Universe is matter dominated, we have $\bar\rho = \rho_{crit} = 3H(z)^2/8\pi G$, thus $t_{dyn} \propto H(z)^{-1}$.

f_*. This is something expected from theory, as one approaches cooling/feedback thresholds, and is also seen in hydrodynamic simulations (e.g. fig. 3 in O'Shea et al. 2015; the left-hand panel of fig. 22 in Xu et al. 2016; Gillet et al. in preparation).

2.2 Galaxy X-ray properties

By heating the IGM prior to the bulk of reionization, X-rays are expected to play a dominant role during the CD (e.g. Pritchard & Furlanetto 2007; McQuinn & O'Leary 2012; Mesinger, Ferrara & Spiegel 2013). For each simulation cell at a given spatial position **x**, and redshift z, 21CMFAST computes the angle-averaged specific X-ray intensity (in units of $\mathrm{erg\,s^{-1}\,keV^{-1}\,cm^{-2}\,sr^{-1}}$), by integrating the comoving X-ray emissivity, $\epsilon_X(\mathbf{x}, E, z)$, back along the light cone:

$$J(\mathbf{x}, E, z) = \frac{(1+z)^3}{4\pi} \int_z^\infty dz' \frac{c\,dt}{dz'} \epsilon_X e^{-\tau}, \tag{6}$$

where $e^{-\tau}$ accounts for attenuation by hydrogen and helium in the IGM (see equation 16 of Mesinger, Furlanetto & Cen 2011). The comoving specific emissivity is calculated in the emitted frame, $E_e = E(1 + z')/(1 + z)$, and is given by

$$\epsilon_X(\mathbf{x}, E_e, z') = \frac{L_X}{\mathrm{SFR}} \left[(1 + \bar\delta_{nl}) \int_0^\infty dM_h \frac{dn}{dM_h} f_{duty} \dot{M}_* \right], \tag{7}$$

where $\bar\delta_{nl}$ is the mean, non-linear overdensity of the shell around $(\mathbf{x}, z)$, and the term inside square brackets corresponds to the SFR density along the light cone. The conditional halo mass function (HMF), $\frac{dn}{dM_h}(M_h, z|R, \delta_R)$, is obtained by normalizing the conditional Press–Schechter HMF (Lacey & Cole 1993; Somerville & Kolatt 1999) so as to have the mean of the Sheth–Tormen HMF (Sheth & Tormen 1999, 2002), as discussed in Mesinger et al. (2011) (see also Barkana & Loeb 2004, 2008). Here the galaxy duty cycle, $f_{duty}(M_h)$ (see equation 5), accounts for inefficient star formation inside small-mass haloes, and the SFR, $\dot{M}_*(M_h, z)$, depends on halo mass and redshift as specified in equation (3).

The term, L_X/SFR, is the specific X-ray luminosity per unit star formation escaping the host galaxies in units of $\mathrm{erg\,s^{-1}\,keV^{-1}\,M_\odot^{-1}\,yr}$. We assume that the specific luminosity follows a power law in photon energy, i.e. $L_X \propto E^{-\alpha_X}$, and adopt $\alpha_X = 1$, consistent with models and observations of local high-mass X-ray binaries (HMXBs) over the relevant energy range (e.g. Fragos et al. 2013). We normalize the specific X-ray luminosity using the integrated soft-band ($<2\mathrm{keV}$) luminosity per SFR (in units of $\mathrm{erg\,s^{-1}\,M_\odot^{-1}\,yr}$),

$$L_{X<2\,\mathrm{keV}}/\mathrm{SFR} = \int_{E_0}^{2\,\mathrm{keV}} dE_e\, L_X/\mathrm{SFR}, \tag{8}$$

where E_0 is the X-ray energy threshold below which photons are absorbed inside the host galaxies. This X-ray energy threshold depends on the density of the interstellar medium (ISM) and metallicity (Das et al. 2017). The upper limit of the integral is motivated by the fact that the mean free path of ~ 2 keV photons is roughly the Hubble length at these redshifts; thus, harder photons do not contribute to IGM heating during the CD (e.g. McQuinn 2012).

The soft-band luminosity from equation (8) is kept as a free parameter, while we fix the spectral index, α_X, to unity. Keeping the spectral index constant is motivated by the fact that the 21-cm power spectra are very insensitive to the spectral index when using this parametrization (e.g. see fig. 1 in Greig & Mesinger 2017b).
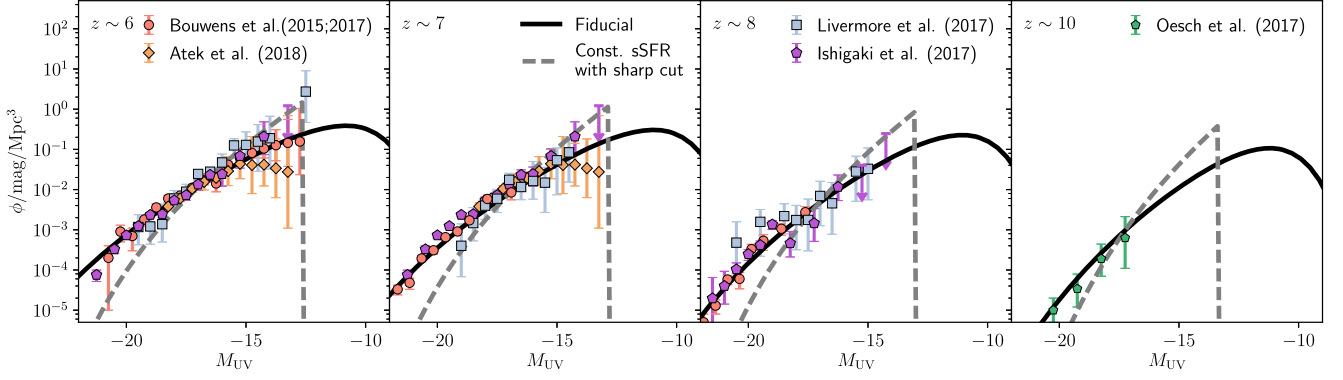
Figure 1. LFs corresponding to our fiducial parameters used to make the mock 21-cm observation below (*black solid lines*), together with observed LFs (*coloured points*). Note that LFs of Ishigaki et al. (2017) and Atek et al. (2018) are duplicated at $z \sim 6$ and 7, because their selection criteria (*i*-dropout) select intermediate-redshift ($z \sim 6$–7) galaxies. The dashed curves illustrate common simplifications made in previous 21-cm studies: (i) fixing $\alpha_* = 0$, thus setting a constant stellar mass to halo mass for all galaxies; and (ii) having a sharp suppression of faint galaxies (i.e. with $f_{\rm duty}$ transitioning from 1 to 0 at $M_{\rm turn}$).

2.3 Inhomogeneous IGM recombinations

Recombinations can impact the progress and topology of reionization via the interplay of ionizing sources and dense IGM structures (so-called Lyman limit systems). If reionization is 'photon-starved' as suggested by emissivity estimates from the Lyman alpha forest, recombinations would 'stall' the growth of large H II regions (e.g. Miralda-Escudé, Haehnelt & Rees 2000; Ciardi et al. 2006; McQuinn et al. 2007; Finlator et al. 2012; Kaurov & Gnedin 2014; Sobacchi & Mesinger 2014). In seminumerical simulations, this effect is usually crudely accounted for with a maximum horizon for ionizing photons (commonly denoted as $R_{\rm mfp}$), which is usually taken to be redshift independent and homogeneous (e.g. Mesinger et al. 2011; Alvarez & Abel 2012; Greig & Mesinger 2017a).

Here we directly compute the local, sub-grid recombinations, according to Sobacchi & Mesinger (2014) (see also Hutter 2018). Specifically, each simulation cell at a spatial location **x** and redshift, z, keeps track of its hydrogen recombination rate according to

$$\frac{dn_{\rm rec}}{dt}(\mathbf{x}, z) = \bar{n}_{\rm H} \alpha_{\rm B} \Delta_{\rm cell}^{-1} \int_0^{180} [1 - x_{\rm HI}]^2 \, P_V \Delta^2 d\Delta \,, \quad (9)$$

where $\Delta_{\rm cell} \equiv 1 + \delta_{\rm nl}$ is the overdensity on the size of the simulation cell, $\Delta \equiv n/\bar{n}$ is the sub-grid overdensity, $P_V(\Delta, \Delta_{\rm cell}, z)$ is the volume-averaged PDF of Δ (with the functional form specified by Miralda-Escudé et al. 2000 and adjusted for the cell's overdensity according to Sobacchi & Mesinger 2014), $\alpha_{\rm B}$ is the case-B recombination coefficient evaluated at a temperature of 10^4 K, and $x_{\rm HI}(\Delta, \Gamma, z)$ is the neutral fraction at the overdensity Δ with the attenuation of the local, inhomogeneous ionizing background Γ accounted for using the analytic expression from Rahmati et al. (2013). The upper limit of the integral is motivated by the mean density of haloes in the spherical collapse model, since by definition recombinations inside galaxies are accounted for in the source terms. However, in practice, this limit is unimportant at the redshifts of interest as gas already starts to self-shield at much lower densities for realistic models of Γ; thus, large densities do not contribute to IGM recombinations.

The reionization field in 21CMFAST is then computed by comparing the cumulative number of ionizing photons in a given region of scale R to the corresponding number of baryons *plus* the average, cumulative number of recombinations inside that region:

$$\bar{n}_{\rm rec}(\mathbf{x}, z, R) = \left\langle \int_{z_{\rm ion}}^z \frac{dn_{\rm rec}}{dt} \frac{dt}{dz} dz \right\rangle_R \,, \quad (10)$$

where $z_{\rm ion}(\mathbf{x})$ is the redshift at which a given cell was first ionized. For more details, see section 3 in Sobacchi & Mesinger (2014).

2.4 Summary of the free parameters in our model

Our new model has eight free parameters. Here we summarize these parameters, also listing them in Table 1, together with the fiducial values and allowed ranges for the MCMC. We stress that the fiducial values are only used when we generate a mock 21-cm observation (see Section 6.2); for the MCMC using UV LFs (Section 6.1), we take currently available observations.

(i) $f_{*, 10}$: the normalization of the fraction of galactic gas in stars at high-z, f_*, evaluated for haloes of mass $10^{10} \, {\rm M}_\odot$. Our fiducial value used to generate a mock 21-cm signal is $f_{*, 10} = 0.05$ and we allow the parameter to vary in the range of $\log_{10}(f_{*, 10}) = [-3, 0]$.

(ii) α_*: the power-law scaling of f_* with halo mass. When making a mock 21-cm observation, we take a fiducial value of $\alpha_* = 0.5$ (e.g. Behroozi & Silk 2015; Ocvirk et al. 2016; Mirocha, Furlanetto & Sun 2017) and we allow the parameter to vary in the range of $\alpha_* = [-0.5, 1]$ in our MCMCs.

(iii) $f_{\rm esc, 10}$: the normalization of the ionizing UV escape fraction of high-z galaxies, $f_{\rm esc}$, evaluated for haloes of mass $10^{10} \, {\rm M}_\odot$. When making a mock 21-cm observation, we take a fiducial value of $f_{\rm esc, 10} = 0.1$ and for our MCMCs we allow the parameter to vary in the range of $\log_{10}(f_{\rm esc, 10}) = [-3, 0]$.

Table 1. The astrophysical parameters of our model, together with the fiducial values used for the mock 21-cm signal, and the allowed range for the MCMC studies. We also note the choice of prior in the final column: flat in either linear or log space within the quoted range. See Section 2.4 for additional details.

Parameter	Fiducial value	Units	Allowed range	Flat prior
$f_{*, 10}$	0.05	–	0.001–1	log
α_*	0.5	–	-0.5–1	linear
$f_{\rm esc, 10}$	0.1	–	0.001–1	log
$\alpha_{\rm esc}$	-0.5	–	-1–0.5	linear
$M_{\rm turn}$	5×10^8	${\rm M}_\odot$	10^8–10^{10}	log
t_*	0.5	–	0–1	linear
$\frac{L_{\rm X<2keV}}{\rm SFR}$	$10^{40.5}$	erg s^{-1} M$_\odot^{-1}$ yr	10^{38}–10^{42}	log
E_0	0.5	keV	0.1–1.5	linear

(iv) α_{esc}: the power-law scaling of f_{esc} with halo mass. We take a fiducial value of $\alpha_{esc} = -0.5$. As mentioned earlier, we expect α_{esc} to be negative as SNe can more easily evacuate low-column density channels from shallower potential wells (e.g. Razoumov & Sommer-Larsen 2010; Yajima, Choi & Nagamine 2011; Ferrara & Loeb 2013; Paardekooper et al. 2015; Xu et al. 2016; Kimm et al. 2017; Katz et al. 2018). We allow the parameter to vary in the range of $\alpha_{esc} = [-1, 0.5]$.

(v) t_*: the star formation time-scale taken as a fraction of the Hubble time, $H^{-1}(z)$. We take a fiducial value of $t_* = 0.5$ and we allow the parameter to vary in the range of $t_* = (0, 1]$.

(vi) M_{turn}: the turnover halo mass below which the abundance of active star-forming galaxies is exponentially suppressed according to a duty cycle of $\exp(-M_{turn}/M_h)$. When making a mock 21-cm observation, we take a fiducial value of $M_{turn} = 5 \times 10^8\,M_\odot$ and in the MCMCs we allow the parameter to vary in the range of $\log_{10}(M_{turn}) = [8, 10]$. Here the lower limit is motivated by the atomic cooling threshold, while the upper limit is motivated by the faint end of current UV LFs (see Fig. 3).

(vii) E_0: the minimum X-ray photon energy capable of escaping the galaxy; softer photons are absorbed by the ISM of high-z galaxies. Motivated by the hydrodynamic simulations used in Das et al. (2017), we take a fiducial value of $E_0 = 0.5$ keV and we allow the parameter to vary in the range of $E_0 = [0.1, 1.5]$. Analogously, this range corresponds to $\log_{10}(N_{H\,I}/cm^2) = [19.3, 23.0]$.[14]

(viii) $L_{X<2\,keV}/\mathrm{SFR}$: the normalization of the soft-band X-ray luminosity per unit star formation, computed over the band E_0 – 2 keV. When making a mock 21-cm observation, we assume the X-ray binary composite SED of Fragos et al. (2013), with the ISM attenuation from Das et al. (2017), resulting in a fiducial value of $L_{X<2\,keV}/\mathrm{SFR} = 10^{40.5}$ erg s^{-1} M$_\odot^{-1}$ yr. In our MCMCs, we allow the parameter to vary in the range $\log_{10}(L_{X<2\,keV}/\mathrm{SFR}) = [38, 42]$.

3 CORRESPONDING UV LUMINOSITY FUNCTIONS

We can write the non-ionizing UV LFs from our model as

$$\phi(M_{UV}) = \left[f_{duty} \frac{dn}{dM_h} \right] \left| \frac{dM_h}{dM_{UV}} \right|, \tag{11}$$

where, as previously noted, the term in brackets is the number density of active star-forming galaxies. The final term on the RHS encodes the conversion of halo mass to UV magnitude. We evaluate this assuming that the SFR is proportional to the rest-frame UV luminosity of a galaxy:

$$\dot{M}_*(M_h, z) = \mathcal{K}_{UV} \times L_{UV}. \tag{12}$$

We assume that the conversion factor, $\mathcal{K}_{UV}$, is constant and adopt $\mathcal{K}_{UV} = 1.15 \times 10^{-28}$ M$_\odot$ yr^{-1}/erg s^{-1} Hz^{-1} following Sun & Furlanetto (2016)[15] (see also e.g. Kennicutt 1998; Madau, Pozzetti & Dickinson 1998; Bouwens et al. 2012). Finally, we relate the UV luminosity to magnitude using the usual AB magnitude
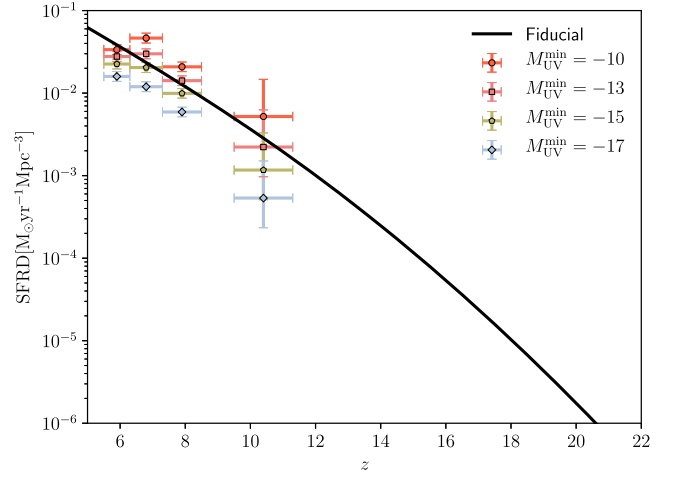

Figure 2. SFR density for our fiducial model (*solid line*). The circles, squares, pentagons, and diamonds represent estimates from Bouwens et al. (2015b), obtained by extrapolating LFs down to minimum magnitudes of $M_{UV}^{min} = -10, -13, -15,$ and -17, respectively.

relation (Oke & Gunn 1983),

$$\log_{10}\left(\frac{L_{UV}}{erg\ s^{-1}\ Hz^{-1}} \right) = 0.4 \times (51.63 - M_{UV}). \tag{13}$$

For illustration purposes, in Fig. 1 we show the rest-frame UV LFs corresponding to the fiducial model parameters we use to make mock 21-cm observations (Table 1), as well current observations of UV LFs. The observations show scatter between various groups, in particular at the faint end that is dominated by lensing uncertainties (Bouwens et al. 2015a, 2016; Livermore, Finkelstein & Lotz 2016; Ishigaki et al. 2017; Atek et al. 2018). Our fiducial parameter choices are consistent with current observations. More important than the fiducial parameter choices is the flexibility of our model to reproduce the main features of the LFs, *within a physical framework based on the HMFs*. It is important to note that empirical models of LF, which are not directly rooted in the assumption that galaxies sit in haloes (such as the Schechter function), require ad hoc tuning to capture the redshift evolution inherent to the HMF. Indeed, Oesch et al. (2017) show that simple estimates based on the HMF can more accurately predict very high-redshift SFRs, compared with empirical ones.

Additionally, in Fig. 2 we show the star formation rate density (SFRD) for our fiducial model parameters (see equation 7). For a comparison, we also show the estimates of Bouwens et al. (2015b), obtained by extrapolating their observed LFs down to minimum magnitudes of $M_{UV}^{min} = -17, -15, -13,$ and -10, truncating them sharply beyond those values. The SFRD corresponding to our fiducial choice of $M_{turn} = 5 \times 10^8$ M$_\odot$ is roughly comparable to the sharp cut-off assumption of Bouwens et al. (2015b) for $M_{UV}^{min} = (-10)-(-13)$.

4 CORRESPONDING 21-CM SIGNAL

4.1 Computing the signal

The 21-cm signal is commonly expressed as the offset of the 21-cm brightness temperature, $\delta T_b(\nu)$, relative to the temperature of the cosmic microwave background (CMB), T_{CMB} (e.g. Furlanetto

---

[14]The conversion to column densities is computed assuming a unity optical depth for a metal-free column of neutral ISM.

[15]This conversion depends on the IMF, as well as the dust content of the galaxy. In the above, we ignore dust extinction for the faint galaxies and high redshifts of interest (e.g. Bouwens et al. 2012; Capak et al. 2015), and assume a Salpeter IMF. However, we note that these uncertainties can roughly be subsumed in our f_* parameter.

2006):

$$\delta T_b \approx 27 x_{\rm HI}(1 + \delta_{\rm nl}) \left(\frac{H}{{\rm d}v_r/{\rm d}r + H} \right) \left(1 - \frac{T_{\rm CMB}}{T_S} \right)$$
$$\times \left(\frac{1+z}{10} \frac{0.15}{\Omega m h^2} \right)^{1/2} \left(\frac{\Omega_b h^2}{0.023} \right), \quad (14)$$

where $x_{\rm HI}$ is the neutral fraction, T_S is the gas spin temperature, $\delta_{\rm nl} \equiv \rho/\bar{\rho} - 1$ is the gas overdensity, $H(z)$ is the Hubble parameter, ${\rm d}v_r/{\rm d}r$ is the gradient of the line-of-sight component of the velocity, and all quantities are evaluated at redshift $z = \nu_0/\nu - 1$, where ν_0 is the 21-cm frequency.

To compute the various fields in the above equation, we use the seminumerical simulation 21CMFAST (Mesinger & Furlanetto 2007; Mesinger et al. 2011). 21CMFAST computes the evolved density and velocity fields from an initial high-resolution Gaussian realization, using second-order LPT (e.g. Scoccimarro 1998). Then, 21CMFAST estimates the ionization field from the evolved IGM density field by comparing the cumulative number of ionizing photons to the number of neutral atoms plus cumulative recombinations within spheres of decreasing radii (e.g. Furlanetto, Zaldarriaga & Hernquist 2004). Specifically, a voxel at coordinates $(\mathbf{x}, z)$ is flagged as fully ionized if it satisfies

$$n_{\rm ion}(\mathbf{x}, z | R, \delta_R) \geq (1 + \bar{n}_{\rm rec})(1 - \bar{x}_e), \quad (15)$$

with $\bar{n}_{\rm rec}$ given by equation (10) and the cumulative number of IGM ionizing photons per baryon produced inside a spherical region of scale R and corresponding overdensity δ_R given by

$$n_{\rm ion} = \bar{\rho}_b^{-1} \int_0^\infty {\rm d}M_h \frac{{\rm d}n(M_h, z | R, \delta_R)}{{\rm d}M_h} f_{\rm duty} M_* f_{\rm esc} N_{\gamma/b}, \quad (16)$$

where $\bar{\rho}_b$ is the mean baryon density in the region, $N_{\gamma/b}$ is the number of ionizing photons per stellar baryon, which we set at 5000, motivated by a Salpeter IMF (in principle, this parameter is largely degenerate with f_*). Note that in our new formulation, the commonly used 'ionizing efficiency' parameter, $\zeta = f_* f_{\rm esc} N_{\gamma/b}$, is broken-up into its constituent parts, with $f_*(M_h)$ and $f_{\rm esc}(M_h)$ now both being functions of halo mass (as discussed in Section 2.1). The final term on the RHS of equation (15) is a small correction factor for pre-ionization by X-rays, discussed below; in practice, this term is negligible for realistic models (see Mesinger et al. 2013).

The above procedure is used to compute the inhomogeneous topology of reionization, consisting of (almost) fully ionized and neutral regions. However, due to their long mean free paths, X-ray and soft UV photons are able to penetrate even the neutral cosmic patches distant from galaxies. These radiation fields help determine the spin temperature. To calculate T_S, 21CMFAST follows the evolution of the ionized fraction inside the neutral IGM, x_e, the kinetic temperature, T_K, and the incident Lyman α background. The ionization fraction and the kinetic temperature in each voxel are solved following

$$\frac{{\rm d}x_e(\mathbf{x}, z')}{{\rm d}z'} = \frac{{\rm d}t}{{\rm d}z'} \left[\Gamma_{\rm ion, X} - \alpha_A C x_e n_b f_H \right], \quad (17)$$

$$\frac{{\rm d}T_K(\mathbf{x}, z')}{{\rm d}z'} = \frac{2}{3k_B(1 + x_e)} \frac{{\rm d}t}{{\rm d}z'} \sum_p Q_p$$
$$+ \frac{2T_K}{3n_b} \frac{{\rm d}n_b}{{\rm d}z'} - \frac{T_K}{1 + x_e} \frac{{\rm d}x_e}{{\rm d}z'}, \quad (18)$$

where n_b is the total (H + He) baryonic number density at $(\mathbf{x}, z')$, ε_p is the heating rate per baryon for process p in erg s^{-1}, α_A is the case-A recombination coefficient, C is the clumping factor on the

scale of the simulation cell, k_B is the Boltzmann constant, f_H is the hydrogen number fraction, $\Gamma_{\rm ion, X}$ is the ionizing background from X-rays, and Q_p is the heating rate per baryon associated with the process 'p'; we include Compton heating and X-ray heating.

The heating and ionization rates per baryon inside the mostly neutral IGM are calculated with (see also e.g. Baek & Ferrara 2013; Madau & Fragos 2017; Eide et al. 2018)

$$Q_X(\mathbf{x}, z) = \int_{{\rm Max}[\nu_0, \nu_{\tau=1}]}^\infty {\rm d}\nu \frac{4\pi J}{h\nu} \sum_i (h\nu - E_i^{\rm th}) f_{\rm heat} f_i x_i \sigma_i \quad (19)$$

and

$$\Gamma_{\rm ion, X}(\mathbf{x}, z) = \int_{{\rm Max}[\nu_0, \nu_{\tau=1}]}^\infty {\rm d}\nu \frac{4\pi J}{h\nu} \sum_i f_i x_i \sigma_i F_i$$
$$F_i = (h\nu - E_i^{\rm th}) \left(\frac{f_{\rm ion, HI}}{E_{\rm HI}^{\rm th}} + \frac{f_{\rm ion, HeI}}{E_{\rm HeI}^{\rm th}} + \frac{f_{\rm ion, HeII}}{E_{\rm HeII}^{\rm th}} \right) + 1, \quad (20)$$

where $i = $ H I, He I, He II denotes the atomic species, f_i is the corresponding number fraction, x_i is the ionization fraction of the cell's species, σ_i is the ionization cross-section, $E_i^{\rm th}$ is the ionization threshold energy of species i, and $f_{\rm heat}$ and $f_{\rm ion, j}$ are the fraction of the primary electron's energy going into heating and secondary ionizations of species j, respectively. The angle-averaged specific X-ray intensity $J(\mathbf{x}, E, z)$ is computed from equations (6) and (7).

With the gas kinetic temperature calculated according to the above equations, the spin temperature can be approximated as a weighted average of the CMB and gas temperatures. Specifically, we have

$$T_S^{-1} = \frac{T_{\rm CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}. \quad (21)$$

Here, T_α is the colour temperature that is closely rated to the gas temperature through multiple Lyman α scatterings (Field 1959). For T_S not to be equal to the CMB temperature (and hence for us to obtain a signal), either the collisional coupling coefficient, x_c, or the Wouthuysen–Field (Wouthuysen 1952; Field 1958; WF) coupling coefficient, x_α, needs to be non-negligible. The former is only efficient in the IGM at $z \gtrsim 30$ while the latter is set by the Lyman α background. 21CMFAST computes the Lyman series background from both X-ray excitation of H I and from direct stellar emission of photons in the Lyman bands, using the composite stellar spectra of Barkana & Loeb (2005). It scales with the SFRs implied by our model, in a manner analogous to equation (20). For more details on the calculations, interested readers are encouraged to consult Mesinger & Furlanetto (2007) and Mesinger et al. (2011).

4.2 Mock 21-cm observation

Using the fiducial parameters listed in Table 1, we generate a mock realization of the 21-cm signal. Our simulation box is 500 Mpc on a side, computed on a 256^3 grid, downsampled from 1024^3 initial conditions. When performing the MCMC, we create 3D simulations on-the-fly, whose dimensions are 250 Mpc on a side on a 128^3 grid. As in previous works, we use a different random seed (and corresponding density realization) for the mock observation than we do for the MCMC inference below. Power spectra are generated from light cones, using the approach from Greig & Mesinger (2018). More details, including the PS of our simulations, can be found in Appendix B.

Fig. 3 shows a 2D light-cone slice of the 21-cm signal, which corresponds to our fiducial model parameters. A 2D slice through the light cone is shown in the top panel, the average brightness temperature is shown in the second panel, and the PS at $k = 0.1\,{\rm Mpc}^{-1}$
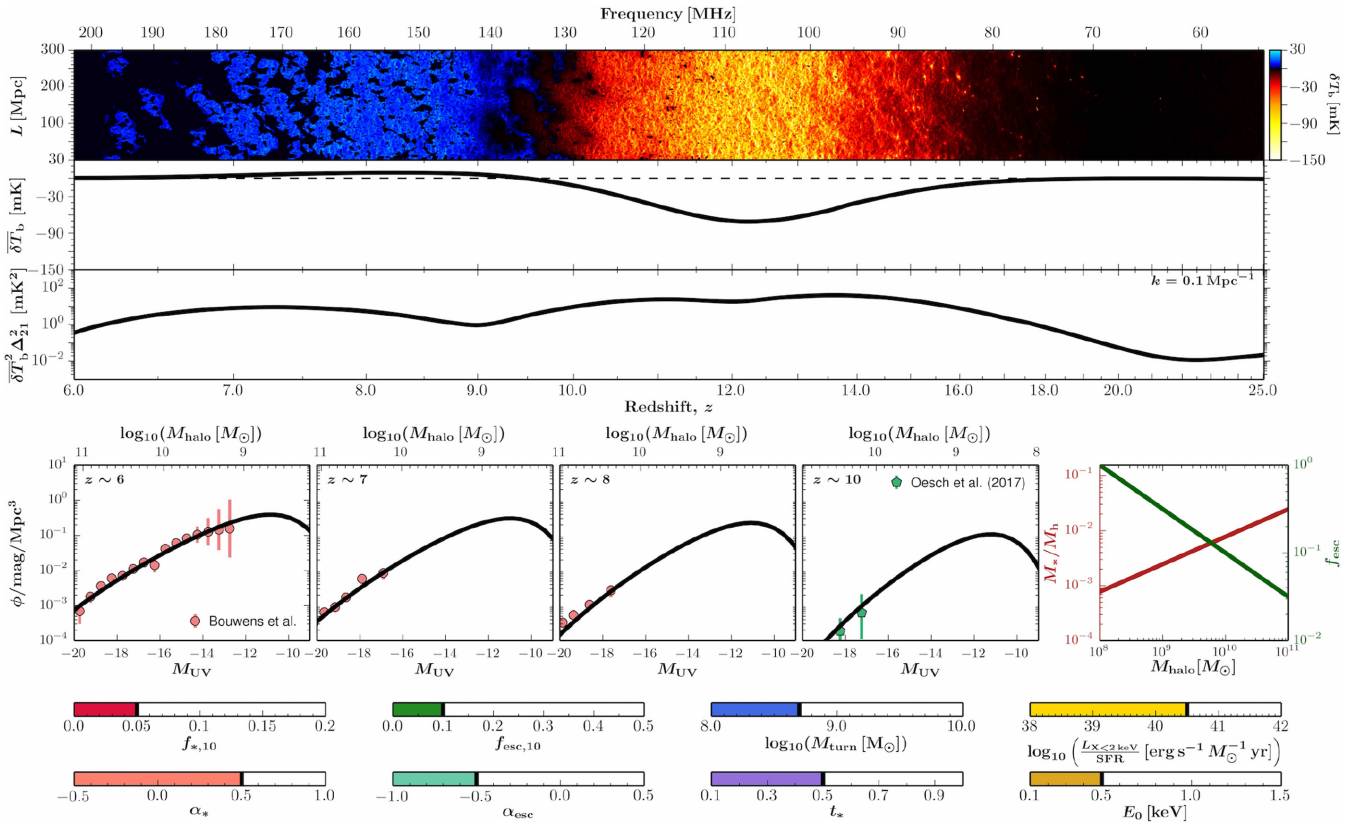
Figure 3. The 21-cm signal together with the UV LFs corresponding to our fiducial model parameters. The top three panels show an ∼1 Mpc slice through the 3D light cone of 21-cm signal, the average brightness temperature offset, and the PS at $k = 0.1\,\mathrm{Mpc}^{-1}$, respectively. The left four panels in the middle show corresponding LFs with observations from Bouwens et al. (2016) for $z \sim 6$, Bouwens et al. (2015a) for $z \sim 7$–8, and Oesch et al. (2017) for $z \sim 10$, respectively. The rightmost panel in the middle shows the stellar mass per halo mass (*left axis*) and the escape fraction (*right axis*) as functions of halo mass. Toggles on the bottom represent the fiducial parameter values. For movies showing how these observables change with changes in the astrophysical parameters, see http://homepage.sns.it/mesinger/Videos/parameter_variation.mp4.

is shown in the third panel. The corresponding LFs and scalings of the stellar mass per halo mass and the escape fraction are shown in the bottom panels.

For these parameter choices, we obtain an end to reionization, which is consistent with current observations (e.g. McGreer, Mesinger & D'Odorico 2015); however, the Epoch of Heating (EoH) and epoch of WF coupling are delayed compared with our previous works (e.g. Mesinger et al. 2016; Greig & Mesinger 2018). Our new fiducial model has the star formation efficiency, f_*, increase with the halo mass, rather than remain constant as we had done previously. The ionizing escape fraction, $f_{\rm esc}$, has the opposite scaling with halo mass for our fiducial choices; as reionization depends on the product of these quantities, the EoR timing is unchanged. However, the EoH is governed by X-rays, which are unaffected by $f_{\rm esc}$.[16] As the importance of small-mass galaxies increases with redshift, the new scaling of $f_*(M_{\rm h})$ results

in a delayed EoH compared to previous $f_* =$ constant. We note that Mirocha et al. (2017) reached a similar conclusion for the global 21-cm signal, using a similar parametrization to the one we use here. Consistent with their follow-up work in Mirocha & Furlanetto (2018), this implies that if the claim of a detection of the EoH at $z \sim 17$ made recently by EDGES (Bowman et al. 2018) is proven genuine, star formation would either need to extend to very small haloes below the atomic cooling threshold, and/or star formation would need to be more efficient in small-mass haloes below current LF limits, implying a break in the power-law scaling of $f_*(M_{\rm h})$. Currently, we are extending our model to capture an additional and separate population of sources residing in smaller, molecularly cooled haloes, in an effort to quantify these claims (Qin et al., in preparation).

The other notable difference between Fig. 3 and previous results is that the EoH and WF coupling epochs are less separated in time. This is due to two factors: (i) the decreasing star formation efficiency with halo mass, discussed above, resulting in a delayed and subsequently more rapid CD (an effect similar to having the dominant population of star-forming galaxies sitting in more massive haloes whose fractional abundance evolves more rapidly); and (ii) our fiducial value of $L_{\rm X, <2keV}$/SFR is larger than most previous studies. Our new value is motivated by the theoretical HMXB models of Fragos et al. (2013), whereas previously we used the empirical scalings of Mineo, Gilfanov & Sunyaev (2012) obtained from local star-forming galaxies. Due to its dependence on metallicity (Basu-Zych

---

[16]Note that in our model, the minimum X-ray energy escaping the ISM is set by the average H I column density of early galaxies, which we take to be *independent* from the UV escape fraction, $f_{\rm esc}$. This is motivated by the emerging physical picture in which the ionizing escape fraction is determined by the covering fraction of low-column density sightlines in early galaxies resulting from feedback events. This low-value tail of the column density distribution (determining $f_{\rm esc}$) is not sensitive to the mean of the distribution (determining E_0) (e.g. Verhamme et al. 2015; Xu et al. 2016; Das et al. 2017; Pallottini et al. 2017).

et al. 2013; Brorby et al. 2016; Lehmer et al. 2016), the X-ray luminosity to SFR for the first galaxies is expected to be roughly an order of magnitude larger than for local ones.

The fact that the epoch of WF coupling and EoH are more coincident in time is evidenced by the smaller separation between the corresponding peaks of the large-scale power, driven by spatial fluctuations in the Ly α coupling coefficient and gas temperature, respectively (e.g. Pritchard & Furlanetto 2007). Moreover, the global absorption signal has a reduced minimum, as the heating commences before all of the IGM has its spin temperature coupled to the gas kinetic temperature. Similarly, the peak in the power spectrum associated with the EoH is reduced, as the cross-terms from the coupling coefficient and gas temperature have a negative contribution to the power amplitude (see the discussion in Pritchard & Furlanetto 2007 and Mesinger et al. 2016).

5 SAMPLING ASTROPHYSICAL PARAMETER SPACE WITH 21CMMC

In this section, we provide a summary of 21CMMC (Greig & Mesinger 2015) used to constrain the astrophysical parameters described in Section 2.4. For further details, interested readers are referred to Greig & Mesinger (2015, 2017b) and Greig & Mesinger (2018).

21CMMC is an MCMC sampler of 3D reionization simulations. To explore the astrophysical parameter space of CD and reionization, 21CMMC adopts a massively parallel MCMC sampler COSMOHAMMER (Akeret et al. 2013) that uses the EMCEE PYTHON module (Foreman-Mackey et al. 2013) based on the affine invariant ensemble sampler (Goodman & Weare 2010). At each proposed MCMC step, 21CMMC calculates an independent 3D light-cone realization of the 21-cm signal, using an optimized version of 21CMFAST. Then, it calculates a likelihood by comparing PS of the sampled 21-cm signal against the mock observation (see Appendix B), defined as

$$\delta \bar{T}_{\rm b}^2 \Delta_{21}^2(k, z) \equiv \frac{k^3}{2\pi^2 V} \delta \bar{T}_{\rm b}^2(z) \left\langle |\delta_{21}(\mathbf{k}, z)|^2 \right\rangle_k, \quad (22)$$

where $\delta_{21}(\mathbf{x}, z) \equiv \delta \bar{T}_{\rm b}(\mathbf{x}, z)/\delta \bar{T}_{\rm b}(z) - 1$. Note that we limit the k space range from 0.1 to 1.0, corresponding roughly to limits on the foreground noise and the shot noise, respectively.

As in previous works, we adopt a modelling uncertainty, accounting for inaccuracies in our seminumerical models. We take a constant uncertainty of 20 per cent on the sampled 21-cm PS, motivated by comparisons to RT simulations (Zahn et al. 2011; Ghara, Choudhury & Datta 2015; Hutter 2018). We note that with further comparisons, these modelling uncertainties can be better characterized and accounted for. Moreover, we include Poisson uncertainties on the sampled 21-cm PS, roughly consistent with cosmic variance for these scales (Mondal et al. 2015). These two uncertainties are added in quadrature with the total noise PS in equation (25).

We account for redshift space distortions along the line of sight using the relation

$$\mathbf{s} = \mathbf{x} + \frac{(1 + z)}{H(z)} v_{\parallel}(\mathbf{x}), \quad (23)$$

where $\mathbf{s}$ and $\mathbf{x}$ are the redshift and real space signal, respectively. For details of this implementation, see Greig & Mesinger (2018) (see also Mao et al. 2012; Jensen et al. 2013).

5.1 Telescope noise

We calculate noise on the mock 21-cm observation using the python module 21CMSENSE (Pober et al. 2013, 2014). First, we generate the thermal noise PS at each uv cell according to (e.g. Morales 2005; McQuinn et al. 2006; Pober et al. 2014):

$$\Delta_{\rm N}^2(k) \approx X^2 Y \frac{k^3}{2\pi^2} \frac{\Omega'}{2t} T_{\rm sys}^2, \quad (24)$$

where XY^2 is a scalar factor converting observed bandwidths and solid angles to comoving distance, Ω' is a beam-dependent factor derived in Parsons et al. (2014), t is the integration time within a particular k mode, and $T_{\rm sys}$ is the system temperature. Then, the total noise power at a given Fourier mode k, with an assumption of Gaussian errors for the cosmic-variance term, is expressed as

$$\delta \Delta_{\rm T+S}^2(k) = \left(\sum_i \frac{1}{[\Delta_{{\rm N},i}^2(k) + \Delta_{21}^2(k)]^2} \right)^{-\frac{1}{2}}, \quad (25)$$

where $\Delta_{21}^2(k)$ is the 21-cm PS from the mock observation.

For our fiducial instrument, we take the HERA design described in Beardsley et al. (2015): a core consisting of 331 dishes. We assume a 1000-h observation, spread across 180 nights at 6 h per night, and an observing bandwidth coverage of 50–250 MHz. We note that previous studies using a reduced parameter set have shown comparable constraints with SKA when using the PS statistic (e.g. Greig & Mesinger 2017b). However, these claims might need to be re-evaluated for our expanded parametrization. We postpone this to future work, as this paper is mainly a proof of concept for the benefit of combining observables.

6 RESULTS: RECOVERY OF ASTROPHYSICAL PARAMETERS

We now quantify how current and upcoming observations are able to constrain our model parameters. We use two main observations: (i) current high-z UV LFs; and (ii) 21-cm PS from a mock 21-cm observation described in Section 4.2. We first quantify the utility of each in turn, before combining them. We also include current EoR constraints, but as we show below, these do not improve parameter constraints beyond what is available with (i) and (ii). Our results are summarized in Table 2, where we list recovered median values for our model parameters together with 68 per cent confidence regions for each data set used in the MCMC.

6.1 Using only galaxy luminosity functions

The LFs from our model depend on four free parameters: $f_{*, 10}$, α_*, $M_{\rm turn}$, and t_*. We begin by quantifying how current observations can constrain these parameters. To compute the likelihood in our MCMC, we use the $z \sim 6$ LFs from Bouwens et al. (2016), $z \sim 7$–8 LFs from Bouwens et al. (2015a), and $z \sim 10$ LFs from Oesch et al. (2017). Using these data points and corresponding error bars, we compute the χ^2 likelihood at each of the four redshifts, and multiply them together when calculating the posterior. We only use data points at $M_{\rm UV} > -20$, as our model does not account for active galactic nucleus (AGN) feedback or dust extinction in bright galaxies since these are far too rare to be relevant for reionization and CD.

We stress that we are *not* trying to rigorously quantify constraints available with current UV LF observations. In order to do this properly, one should account for systematic uncertainties, combining the

Table 2. Summary of the median recovered values and 1σ errors for the eight-parameter astrophysical model, obtained from our MCMC procedure for each combination of data sets listed below. The LF observations are from Bouwens et al. (2015a, 2016) and Oesch et al. (2017), the τ_{e} constraints are from Planck Collaboration XLVII (2016b), the dark fraction constraints are from McGreer et al. (2015), while the 21-cm data correspond to PS extracted from a mock 1000-h observation with HERA331.

	Parameters							
	$\log_{10}(f_{*,\,10})$	α_*	$\log_{10}(f_{\mathrm{esc},\,10})$	α_{esc}	$\log_{10}(M_{\mathrm{turn}})$	t_*	$\log_{10}\left(\frac{L_{\mathrm{X<2keV}}}{\mathrm{SFR}}\right)$	E_0
					$(\mathrm{M_\odot})$		$(\mathrm{erg\,s^{-1}\,M_\odot^{-1}\,yr})$	(keV)
Fiducial values	-1.30	0.50	-1.00	-0.50	8.7	0.5	40.50	0.50
LF only	$-1.25^{+0.20}_{-0.39}$	$0.50^{+0.07}_{-0.06}$	$-$	$-$	$8.68^{+0.40}_{-0.41}$	$0.51^{+0.30}_{-0.30}$	$-$	$-$
LF $+\,\tau_{\mathrm{e}}\,+$ the dark fraction	$-1.21^{+0.18}_{-0.30}$	$0.50^{+0.07}_{-0.07}$	$-0.91^{+0.42}_{-0.35}$	$-0.13^{+0.44}_{-0.53}$	$8.65^{+0.44}_{-0.41}$	$0.55^{+0.28}_{-0.27}$	$-$	$-$
21-cm only	$-1.29^{+0.18}_{-0.21}$	$0.38^{+0.23}_{-0.31}$	$-0.99^{+0.24}_{-0.21}$	$-0.42^{+0.26}_{-0.27}$	$8.80^{+0.27}_{-0.26}$	$0.46^{+0.17}_{-0.14}$	$40.46^{+0.07}_{-0.07}$	$0.50^{+0.04}_{-0.04}$
21-cm $+$ LF	$-1.20^{+0.14}_{-0.14}$	$0.47^{+0.06}_{-0.06}$	$-1.10^{+0.16}_{-0.18}$	$-0.48^{+0.14}_{-0.18}$	$8.76^{+0.19}_{-0.23}$	$0.56^{+0.21}_{-0.16}$	$40.49^{+0.05}_{-0.06}$	$0.50^{+0.03}_{-0.03}$

various estimates from different groups, some of which are shown in Fig. 1. We postpone such an investigation for future work. By using observations from a single group, we are somewhat overestimating the current constraining power of UV LFs, illustrating their future potential when/if systematics can be better understood and we can have 'concordance' LFs.

The constraints on our four model parameters that determine LFs are denoted with the solid green curves in the triangle plot of Fig. 4, and are summarized in the first row of Table 2. Given the allowed range of parameter space, the most robust constraints we obtain are on $\alpha_* = 0.50^{+0.07}_{-0.06}(1\sigma)$. This parameter most strongly affects the *slope* of the LFs, which are very well determined by current observations (see also very similar conclusions reached by Mirocha et al. 2017, who use a similar galaxy formation model).

By contrast, current LFs only set an upper limit for M_{turn}, ruling out $M_{\mathrm{turn}} \lesssim 9.5$ due to the presence of faint galaxies. The marginalized 1D PDF below this value is relatively flat, due to the large uncertainties at the faint end, and also to the lack of an identifiable turnover in the observational data sets we use.

Our remaining two galaxy model parameters are only constrained as a ratio, as is evidenced by the strong degeneracy of the green curve in the $f_{*,\,10} - t_*$ panel of the triangle plot. This is understandable as the LF in our model is determined only by the SFR, which scales as f_*/t_*.

6.1.1 Do current constraints on the reionization history improve parameter inference and allow us to constrain the escape fraction?

The 1500 Å UV LFs do not tell us anything about the ionizing escape fraction of these galaxies. Fortunately, we have additional probes of the first billion years, which directly measure the timing of the EoR. If high-redshift galaxies are responsible for driving the EoR, as seems highly likely, perhaps combining LF observations with EoR observations will allow us to constrain f_{esc}. Indeed, similar approaches combining LFs and EoR measurements have been used by several studies to constrain the escape fraction and its redshift evolution (e.g. Haardt & Madau 2012; Kuhlen & Faucher-Giguère 2012; Mitra, Ferrara & Choudhury 2013; Robertson et al. 2013; Price et al. 2016).

In this section, we expand our model parameter space to include $f_{\mathrm{esc},\,10}$ and α_{esc}, and include two additional observational data sets in our MCMC. We use the EoR constraints that are the least model dependent: (i) the electron scattering optical depth to the CMB

(Planck Collaboration XLVII 2016b); and (ii) the dark fraction of pixels in QSO spectra (McGreer et al. 2015). For (i), we use the latest estimate of the optical depth $\tau_{\mathrm{e}} = 0.058 \pm 0.012(1\sigma)$ from Planck Collaboration XLVII (2016b). For (ii), we use the upper limit from McGreer et al. (2015) $x_{\mathrm{H{\scriptsize I}}} < 0.06 + 0.05(1\sigma)$ at $z = 5.9$. The reionization history from each sample in the chain is compared against these observations, and the corresponding χ^2 likelihoods are multiplied together with the LF likelihoods.

The resulting parameter constraints are shown with the blue dashed curves in the triangle plot of Fig. 4, and summarized in the second row of Table 2. The additional EoR observations do not improve constraints on the four star formation model parameters studied in the previous section; these are determined almost entirely by LF observations. From the marginalized 1D PDFs, the normalization parameter, $\log_{10}(f_{\mathrm{esc},\,10})$ is constrained to a 1σ percentage error of $\sim$40 per cent. Although this is two times larger than we can obtain with 21-cm signal in the next section, it is the best constraint currently available. It shows a mild degeneracy with the star formation parameters, $f_{*,\,10}$ and t_*, which are mostly constrained with LF observations. The scaling of the escape fraction with halo mass, α_{esc}, is entirely unconstrained, as evidenced by the flat PDFs, which are very similar to our flat priors.

In Appendix C, we also consider measurements of the ionizing emissivity at lower redshifts. These can be used as upper limits for the galaxy ionizing emission, since at those redshifts ($z \lesssim 5$), the contribution from AGNs could be non-negligible. We show that current measurements, although not adding constraining power to most of the galaxy formation parameters, do reduce the 1σ error for the escape fraction down to $\sim$25 per cent. This is however a very model-dependent measurement, and so we leave it out of our fiducial data sets.

6.1.2 Corresponding luminosity functions inferred from the data

One powerful benefit of our model is that it allows us to infer LFs down to the faint galaxies and high redshifts, inaccessible by current observations. These LFs, corresponding to the astrophysical parameter constraints, are shown in the top right corner of Fig. 4, with the blue hatched region corresponding to the 95 per cent C.L. From the panels, we see explicitly that the faint end, $M_{\mathrm{UV}} \gtrsim -14$, is poorly constrained, consistent with our broad marginalized limits on M_{turn}. However, due to the tight constraints on α_* and the ratio $f_{*,\,10}/t_*$, current observations allow us to place very
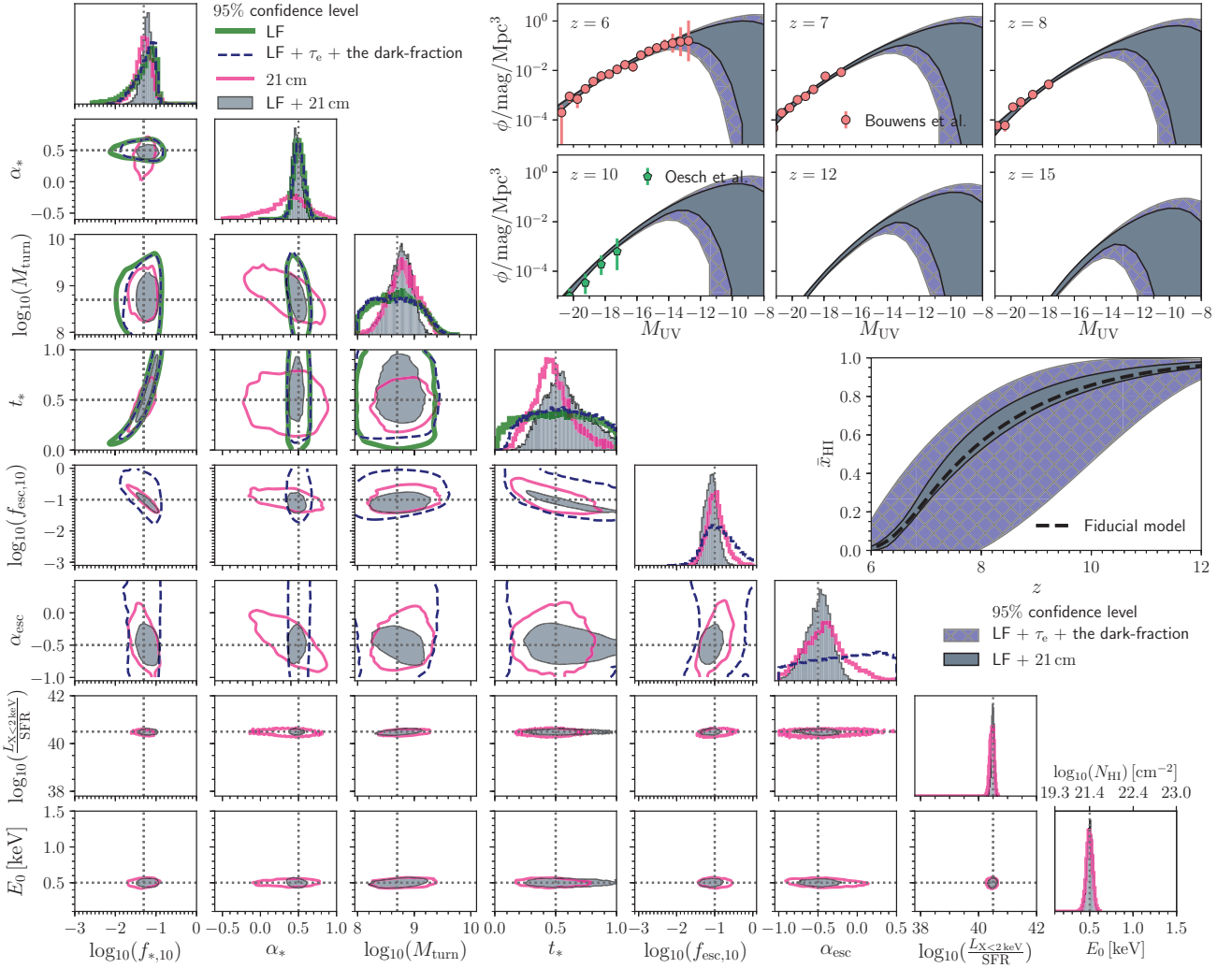
Figure 4. Corner plot showing parameter constraints for the various data sets used (see the legend): 2D marginalized joint posterior distributions are shown in the bottom left corner with 1D marginalized PDFs along the diagonal. The green solid lines, blue dashed lines, pink solid lines, and shaded regions represent 95 per cent confidence levels for constraints using data sets of LF only, LF $+ \tau_e +$ the dark fraction, the mock 21-cm PS, and both LF $+$ the mock 21-cm PS, respectively. Top-right panels: Recovered 95 per cent confidence levels of the LFs corresponding to the posterior of our model. The shaded regions with hatch (blue) and shaded regions (grey) represent constraints using LF $+ \tau_e +$ the dark fraction and using 21-cm with the UV LFs, respectively. Middle-right: Corresponding constraints on the global evolution of the IGM neutral fraction, $x_{\mathrm{H\,I}}(z)$. The dashed line represents the fiducial model used to make the mock 21-cm observation. The shaded regions with hatch (blue) and shaded regions (grey) represent the recovered 95 per cent confidence levels for constraints using LF $+ \tau_e +$ the dark fraction and using both 21-cm with the UV LFs, respectively.

tight constraints on $M_{\mathrm{UV}} \lesssim -14$ galaxies. This is true even at very high redshifts, e.g. $z \sim 15$, where we have *no* observations currently. These Bayesian constraints are a by-product of our model assumption that there is a redshift-independent relation between the stellar mass and halo mass, and that the stellar mass is built-up with a rate that scales inversely with the Hubble time. These assumptions allow us to constrain high-z LFs even with current observations.

6.2 Using only 21-cm signal

We now turn to the constraining power of the 21-cm signal. In Greig & Mesinger (2017b) and Greig & Mesinger (2018), we showed that mock 21-cm PS alone could constrain a simpler

parametrization of galaxy properties.[17] Our model here is more sophisticated/flexible with more free parameters, and so we expect constraints to be weaker.

---

[17]Mirocha, Harker & Burns (2015) presented parameter constraints from measurements of the inflection points of *the global 21-cm signal*, using an idealized instrument. Their results are directly comparable to constraints from the PS in Greig & Mesinger (2017b). Depending on the fiducial model, they recover four astrophysical parameters to roughly $\sim$10 per cent precision, while Greig & Mesinger (2017b) recover five comparable astrophysical parameters to of the order of $\sim$1 per cent precision, using 21-cm PS. This improvement in precision despite a larger model parameter space is due to the power spectrum encoding information about the structure of the signal, which breaks degeneracies in the timing obtained from the global signal (e.g. fig. 2 in Greig & Mesinger 2015), even when realistic noise and foreground avoidance is assumed for the interferometer.

Unlike in the previous section, here we have no current observations to use for our MCMC. We therefore use a mock 1000-h 21-cm observation, generated from different cosmological initial conditions, as described in Section 4.2. This mock observation is created using the fiducial parameters shown in Table 1, and denoted with the vertical and horizontal dotted lines in our corner plot. Although these fiducial choices are consistent at 1σ with those recovered from actual LF observations, they do not correspond to the ML values. As such, the LF data and 21-cm data do not converge to a single set of parameters, and thus we slightly underestimate the potential of combining the two measurements, quantified in the following section. This is a reasonably conservative assumption, as there could be unknown systematics presented in either observation that could pull the posterior towards different values.

In Fig. 4 (pink solid lines), we show the 95 percentiles for each of the eight free parameters along with the 1D marginalized PDFs. In the top right of the figure, we show the recovered median values of the IGM neutral fraction, $\bar{x}_{\rm H_I}$, with 95 per cent confidence levels. It is clear that we recover the parameters used in the mock observation. Specifically, the recovered 68 per cent confidence intervals are $[\log_{10}(f_{*,10}), \alpha_*, \log_{10}(f_{\rm esc,10}), \alpha_{\rm esc}, \log_{10}(M_{\rm turn}), t_*, \log_{10}(L_{\rm X<2keV}/{\rm SFR}), E_0] = (15, 71, 23, 63, 3, 34, 0.2, 8)$ per cent.

From the marginalized 1D PDFs, we see that the 21-cm signal is not very sensitive to α_* alone. The LF observations, discussed in the previous section and shown with the green curve, are much more powerful at constraining this scaling of the SFR with halo mass.

Other parameters are recovered at either comparable confidence as using LF alone, or with improved confidence. For example, the turnover halo mass, $\log_{10}(M_{\rm turn})$, is constrained to a 1σ percentage error of 3, indicating that the 21-cm signal can inform us on the turnover scale that is not captured by LF observations. However, it does show a degeneracy with α_*. Both $M_{\rm turn}$ and α_* help in determining which DM mass scale hosts the dominant population of star-forming galaxies that drive the 21-cm signal: increasing either shifts the population towards higher mass haloes, and vice versa. As the large-scale 21-cm power is sensitive to the bias of the average (unseen) source population (e.g. McQuinn et al. 2007; McQuinn & D'Aloisio 2018), our mock observation constrains a combination of these two parameters.

There is also a degeneracy between the normalization and halo mass scalings of the escape fraction and the star formation efficiency, as evidenced by the $f_{*,10}$ versus $f_{\rm esc,10}$ and α_* versus $\alpha_{\rm esc}$ panels. This is understandable as the EoR, which is at the lowest redshifts for which the telescopes are most sensitive (see Appendix B), only depends on the product of f_* and $f_{\rm esc}$ (cf. equation 16). On the other hand, the EoH and WF coupling epochs only depend on f_*, ameliorating the degeneracy.

We also note a degeneracy in $f_{*,10}$ versus t_*, although it is smaller than when using only LFs. In contrast with LFs that are only sensitive to the instantaneous SFR, the 21-cm EoR signal more strongly depends on the *cumulative* SFR (i.e. the stellar mass), since the average IGM recombination time-scale is longer than the duration of the EoR; thus, once a cosmic IGM patch is ionized, it generally stays ionized. Nevertheless, since the comoving specific emissivity, which is used for the EoH and WF coupling epochs (equation 7), is still proportional to $f_{*,10}/t_*$, the degeneracy is not completely broken.

Finally, we note that the X-ray properties of the first galaxies, inaccessible with UV LFs, are very strongly constrained with the 21-cm signal. In particular, the soft-band X-ray luminosity per unit SFR can be constrained at the level of ~ 0.1 per cent while the minimum X-ray energy escaping the galaxies (which is related to the typical

ISM column density) can be constrained at ~ 1–10 per cent, as seen from the 1D marginalized PDFs. [18]

6.3 Using both LFs and the 21-cm signal

Finally, we show parameter constraints if both the LF observations and the mock 21-cm observations are used when computing the likelihood. The resulting marginalized distributions are shown as shaded regions in the triangle plot of Fig. 4, and the corresponding 2σ constraints on the EoR history are shown with the grey lines in the inset of the figure.

As expected, all of the constraints are either similar to or improved when combining both data sets. As noted earlier, these results are conservative in that the ML values for the LF-only MCMC are not used to create the mock 21-cm signal; thus, the two data sets pull the posterior towards slightly different values (cf. the $f_{*,10}$ 1D PDFs), crudely mimicking the impact on unknown systematics.

In general, the two data sets are fairly complementary, with 21-cm providing the bulk of the constraining power. $M_{\rm turn}, t_*, f_{\rm esc,10}, \alpha_{\rm esc}, L_{\rm X<2keV}/{\rm SFR}$, and E_0 are determined almost entirely by the 21-cm signal. α_* is determined almost entirely by the LFs, while $f_{*,10}$ is constrained by both data sets to a comparable degree. This is also evident in the corresponding LF constraints, in which the bright end is constrained by both data sets, while the faint end is more strongly constrained by 21-cm. Moreover, the $f_{*,10}$ - t_* degeneracy is significantly mitigated by combining the data sets.

From the middle-right panel, we see that the EoR history is entirely constrained by 21-cm. Although knowing the EoR history is less remarkable than knowing the various galaxy properties in the triangle plot, it enables 21-cm observations to tightly constrain τ_e: an important systematic for CMB studies (e.g. Liu et al. 2016).

In summary, the 1σ percentage errors on our parameters from the combined data sets are $[\log_{10}(f_{*,10}), \alpha_*, \log_{10}(f_{\rm esc,10}), \alpha_{\rm esc}, \log_{10}(M_{\rm turn}), t_*, \log_{10}(L_{\rm X<2keV}/{\rm SFR}), E_0] = (12, 13, 15, 33, 2.4, 33, 0.14, 6)$ per cent. [19]

7 CONCLUSION

In the near future, we will detect the 3D structure of the cosmic 21-cm signal. This signal promises to be a treasure trove of physics, informing us on the properties of the otherwise unseen population of galaxies driving the EoR and CD.

Here we develop an expanded, flexible model for galaxy formation, implementing it in the 21-cm modelling code 21CMFAST. In particular: (i) we allow both the stellar mass and the ionizing escape fraction to be a function of the mass of the host halo;

---

[18] This statement is true for our fiducial parameter set used to calculate the mock observation. As quantified by Gillet et al. (2018), if the ISM attenuation of early galaxies is much larger than we expect, such that only hard X-rays escape to heat the IGM (see also Mesinger et al. 2013; Fialkov & Barkana 2014), E_0 will not be recovered. This is due to the strong dependence of the absorption cross-section to photon energy, making the EoH insensitive to the hard X-rays.

[19] We note that the tightest constraints we obtain are on $L_{\rm X<2keV}/{\rm SFR}$, constrained to ~ 1.4 per cent. However, this is only strictly true in the context of our model. For example, if one allows $L_{\rm X<2keV}/{\rm SFR}$ to vary with host halo mass or time, it will be less tightly constrained due to the additional free parameter(s). Nevertheless, the power of our fully Bayesian framework is that when we have an actual observation, we can easily test whether or not the data prefer a more complicated model, using the evidence to perform model selection.

(ii) we implement a duty cycle that suppresses star formation inside low-mass haloes; and (iii) we directly incorporate sub-grid recombinations based on the local density and ionization history.

Using an MCMC sampler of 3D simulations, 21CMMC, we constrain the eight free parameters of our galaxy model using: (i) current observations of high-z LFs; (ii) mock 21-cm PS as measured by a 1000-h integration with HERA; and (iii) and a combination of (i) and (ii).

Using only UV LFs allows us to constrain the scaling of the star formation efficiency with halo mass, and the ratio of $f_{*,10}/t_*$. Folding in EoR observations allows us to additionally weakly constrain the normalization of the ionizing escape fraction, $f_{esc,10}$, but not its dependence on the halo mass.

Including the mock 21-cm PS when performing inference allows us to mitigate these degeneracies, constraining even the ionizing escape fraction and two additional X-ray properties: (i) the soft-band X-ray luminosity per unit star formation and (ii) the minimum X-ray energy escaping the galaxies (analogous to the typical ISM column density). The halo mass scaling, and to a lesser extent the normalization, of the stellar mass is mostly constrained by the LFs. The remaining parameters are mostly constrained by the 21-cm PS. Combining the two parameter sets, we recover all of the parameters at the level of ∼10 per cent or better, with only mild degeneracies remaining.

Our flexible framework makes it easy to tie galaxy observations to the corresponding 21-cm signal. Moreover, 21-cm forecasts can be made from more detailed semi-analytic models of galaxy formation, by casting them into our framework. These improvements to our modelling and inference codes are made publicly available at 21CMFAST (https://github.com/andreimesinger/21cmFAST) and 21CMMC (https://github.com/BradGreig/21CMMC).

REFERENCES

Akeret J., Seehars S., Amara A., Refregier A., Csillaghy A., 2013, Astron. Comput., 2, 27
Alvarez M. A., Abel T., 2012, ApJ, 747, 126
Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, MNRAS, 479, 5184
Aubert D., Deparis N., Ocvirk P., 2015, MNRAS, 454, 1012
Baek S., Ferrara A., 2013, MNRAS, 432, L6
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Barkana R., Loeb A., 2004, ApJ, 609, 474
Barkana R., Loeb A., 2005, ApJ, 626, 1
Barkana R., Loeb A., 2008, MNRAS, 384, 1069
Basu-Zych A. R. et al., 2013, ApJ, 762, 45

Beardsley A. P., Morales M. F., Lidz A., Malloy M., Sutter P. M., 2015, ApJ, 800, 128
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Behroozi P. S., Silk J., 2015, ApJ, 799, 32
Bouwens R. J. et al., 2012, ApJ, 754, 83
Bouwens R. J. et al., 2015a, ApJ, 803, 1
Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015b, ApJ, 811, 140
Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2016, ApJ, 843, 129
Bowman J. D. et al., 2013, PASA, 30, e031
Bowman J. D., Rogers A. E. E., 2010, Nature, 468, 796
Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67
Brorby M., Kaaret P., Prestwich A., Mirabel I. F., 2016, MNRAS, 457, 4081
Capak P. L. et al., 2015, Nature, 522, 455
Chardin J., Haehnelt M. G., Aubert D., Puchwein E., 2015, MNRAS, 453, 2943
Choudhury T. R., Ferrara A., 2006, MNRAS, 371, L55
Ciardi B., Scannapieco E., Stoehr F., Ferrara A., Iliev I. T., Shapiro P. R., 2006, MNRAS, 366, 689
D'Aloisio A., McQuinn M., Maupin O., Davies F. B., Trac H., Fuller S., Upton Sanderbeck P. R., 2018a, preprint (arXiv:1807.09282)
D'Aloisio A., McQuinn M., Davies F. B., Furlanetto S. R., 2018b, MNRAS, 473, 560
Das A., Mesinger A., Pallottini A., Ferrara A., Wise J. H., 2017, MNRAS, 469, 1166
Dayal P., Ferrara A., Dunlop J. S., Pacucci F., 2014, MNRAS, 445, 2545
DeBoer D. R. et al., 2017, PASP, 129, 045001
Eide M. B., Graziani L., Ciardi B., Feng Y., Kakiichi K., Di Matteo T., 2018, MNRAS, 476, 1174
Ferrara A., Loeb A., 2013, MNRAS, 431, 2826
Fialkov A., Barkana R., 2014, MNRAS, 445, 213
Fialkov A., Cohen A., Barkana R., Silk J., 2017, MNRAS, 464, 3498
Field G. B., 1958, Proc. IRE, 46, 240
Field G. B., 1959, ApJ, 129, 536
Finlator K., Oh S. P., Özel F., Davé R., 2012, MNRAS, 427, 2464
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fragos T. et al., 2013, ApJ, 764, 41
Furlanetto S. R., 2006, MNRAS, 371, 867
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181
Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485
Ghara R., Choudhury T. R., Datta K. K., 2015, MNRAS, 447, 1806
Gillet N., Mesinger A., Greig B., Liu A., Ucci G., 2018, preprint (arXiv:1805.02699)
Giroux M. L., Sutherland R. S., Shull J. M., 1994, ApJ, 435, L97
Gnedin N. Y., Kaurov A. A., 2014, ApJ, 793, 30
Gnedin N. Y., Ostriker J. P., 1997, ApJ, 486, 581
Gnedin N. Y., Shaver P. A., 2004, ApJ, 608, 611
Goodman J., Weare J., 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Gorce A., Douspis M., Aghanim N., Langer M., 2018, A&A, 616, A113
Greig B., Mesinger A., 2015, MNRAS, 449, 4246
Greig B., Mesinger A., 2017a, MNRAS, 465, 4838
Greig B., Mesinger A., 2017b, MNRAS, 472, 2651
Greig B., Mesinger A., 2018, MNRAS, 477, 3217
Haardt F., Madau P., 2012, ApJ, 746, 125
Hassan S., Davé R., Finlator K., Santos M. G., 2017, MNRAS, 468, 122
Hills R., Kulkarni G., Meerburg P. D., Puchwein E., 2018, preprint (arXiv:1805.01421)
Hogan C. J., Rees M. J., 1979, MNRAS, 188, 791
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Hutter A., 2018, MNRAS, 477, 1549
Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2017, ApJ, 854, 73
Jensen H. et al., 2013, MNRAS, 435, 460

Katz H., Kimm T., Haehnelt M., Sijacki D., Rosdahl J., Blaizot J., 2018, MNRAS, 478, 4986
Kaurov A. A., Gnedin N. Y., 2014, ApJ, 787, 146
Kennicutt R. C.,, Jr, 1998, ARA&A, 36, 189
Kern N. S., Liu A., Parsons A. R., Mesinger A., Greig B., 2017, ApJ, 848, 23
Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 466, 4826
Koopmans L. et al., 2015, Proc. Sci., The Cosmic Dawn and Epoch of Reionisation with SKA. SISSA, Trieste, PoS#1, available at: http://adsabs.harvard.edu/abs/2015aska.confE...1K
Kuhlen M., Faucher-Giguère C. A., 2012, MNRAS, 423, 862
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lehmer B. D. et al., 2016, ApJ, 825, 7
Liu A., Pritchard J. R., Allison R., Parsons A. R., Seljak U., Sherwin B. D., 2016, Phys. Rev. D, 93, 043013
Livermore R. C., Finkelstein S. L., Lotz J. M., 2016, ApJ, 835, 1
Lusso E., Worseck G., Hennawi J. F., Prochaska J. X., Vignali C., Stern J., O'Meara J. M., 2015, MNRAS, 449, 4204
Madau P., Fragos T., 2017, ApJ, 840, 39
Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429
Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106
Mao Y., Shapiro P. R., Mellema G., Iliev I. T., Koda J., Ahn K., 2012, MNRAS, 422, 926
Mason C. A., Trenti M., Treu T., 2015, Proc. Int. Astron. Union, 11, 33
McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499
McQuinn M., 2012, MNRAS, 426, 1349
McQuinn M., D'Aloisio A., 2018, JCAP, 2018, 16
McQuinn M., O'Leary R. M., 2012, ApJ, 760, 3
McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815
McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043
McQuinn M., Oh S. P., Faucher-Giguère C.-A., 2011, ApJ, 743, 82
Mellema G. et al., 2013, Exp. Astron., 36, 235
Mesinger A., Dijkstra M., 2008, MNRAS, 390, 1071
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Mesinger A., Ferrara A., Spiegel D. S., 2013, MNRAS, 431, 621
Mesinger A., Ewall-Wice A., Hewitt J., 2014, MNRAS, 439, 3262
Mesinger A., Greig B., Sobacchi E., 2016, MNRAS, 459, 2342
Mineo S., Gilfanov M., Sunyaev R., 2012, MNRAS, 419, 2095
Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1
Mirocha J., Furlanetto S. R., 2018, MNRAS, 483, 1980
Mirocha J., Harker G. J. A., Burns J. O., 2015, ApJ, 813, 11
Mirocha J., Furlanetto S. R., Sun G., 2017, MNRAS, 464, 1365
Mitra S., Choudhury T. R., Ferrara A., 2011, MNRAS, 413, 1569
Mitra S., Ferrara A., Choudhury T. R., 2013, MNRAS, 428, L1
Mitra S., Roy Choudhury T., Ferrara A., 2015, MNRAS, 454, L76
Mitra S., Choudhury T. R., Ferrara A., 2018, MNRAS, 473, 1416
Mondal R., Bharadwaj S., Majumdar S., Bera A., Acharyya A., 2015, MNRAS, 449, L41
Morales M. F., 2005, ApJ, 619, 678
Morales M. F., Wyithe J. S. B., 2010, ARA&A, 48, 127
Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 250
O'Shea B. W., Wise J. H., Xu H., Norman M. L., 2015, ApJ, 807, L12
Ocvirk P. et al., 2016, MNRAS, 463, 1462
Oesch P. A., Bouwens R. J., Illingworth G. D., Labbe I., Stefanon M., 2017, ApJ, 855, 105
Okamoto T., Gao L., Theuns T., 2008, MNRAS, 390, 920
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544
Pallottini A., Ferrara A., Bovino S., Vallini L., Gallerani S., Maiolino R., Salvadori S., 2017, MNRAS, 471, 4128
Parsons A. R. et al., 2010, AJ, 139, 1468
Parsons A. R. et al., 2014, ApJ, 788, 106
Patra N., Subrahmanyan R., Raghunathan A., Udaya Shankar N., 2013, Exp. Astron., 36, 319
Planck Collaboration XIII, 2016a, A&A, 594, A13
Planck Collaboration XLVII, 2016b, A&A, 596, A108
Pober J. C. et al., 2013, AJ, 145, 65
Pober J. C. et al., 2014, ApJ, 782, 66
Price D. C. et al., 2018, MNRAS, 478, 4193
Price L. C., Trac H., Cen R., 2016, preprint (arXiv:1605.03970)
Pritchard J. R., Furlanetto S. R., 2007, MNRAS, 376, 1680
Pritchard J. R., Loeb A., 2012, Rep. Prog. Phys., 75, 086901
Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, MNRAS, 430, 2427
Razoumov A. O., Sommer-Larsen J., 2010, ApJ, 710, 1239
Robertson B. E. et al., 2013, ApJ, 768, 71
Schmit C. J., Pritchard J. R., 2018, MNRAS, 475, 1213
Scoccimarro R., 1998, MNRAS, 299, 1097
Scott D., Rees M. J., 1990, MNRAS, 247, 510
Shapiro P. R., Giroux M. L., Babul A., 1994, ApJ, 427, 25
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Shimabukuro H., Semelin B., 2017, MNRAS, 468, 3869
Sobacchi E., Mesinger A., 2013a, MNRAS, 432, L51
Sobacchi E., Mesinger A., 2013b, MNRAS, 432, 3340
Sobacchi E., Mesinger A., 2014, MNRAS, 440, 1662
Sokolowski M. et al., 2015, PASA, 32, e004
Somerville R. S., Kolatt T. S., 1999, MNRAS, 305, 1
Springel V., Hernquist L., 2003, MNRAS, 339, 312
Sun G., Furlanetto S. R., 2016, MNRAS, 460, 417
Tingay S. J. et al., 2013, PASA, 30, e007
Tozzi P., Madau P., Meiksin A., Rees M. J., 2000, ApJ, 528, 597
van Haarlem M. P. et al., 2013, A&A, 556, A2
Verhamme A., Orlitová I., Schaerer D., Hayes M., 2015, A&A, 578, A7
Voytek T. C., Natarajan A., Jáuregui García J. M., Peterson J. B., López-Cruz O., 2014, ApJ, 782, L9
Worseck G. et al., 2014, MNRAS, 445, 1745
Wouthuysen S. A., 1952, AJ, 57, 31
Xu H., Wise J. H., Norman M. L., Ahn K., O'Shea B. W., 2016, ApJ, 833, 84
Yajima H., Choi J.-H., Nagamine K., 2011, MNRAS, 412, 411
Yue B., Ferrara A., Xu Y., 2016, MNRAS, 463, 1968
Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, MNRAS, 414, 727

APPENDIX A: FLEXIBILITY OF THE FUNCTIONAL FORM FOR LFS

Here we quantify further the claim that our analytical model, based on the HMF, is flexible enough to fit 'reasonable' LFs. To do so, we make use of several hydrodynamic cosmological simulations, which form part of the ongoing PRACE tier-0 project GAFFER. The simulations were generated using the cosmological code, EMMA (Aubert, Deparis & Ocvirk 2015), which includes a classical star formation recipe and supernova feedback [Deparis et al. (in preparation)]. As part of GAFFER, we perform many simulations, varying parameters such as star formation efficiency, star formation density threshold, and supernova efficiency. The simulations have a box length of $10\,\mathrm{Mpc}$ and resolve halo masses down to $10^8\,\mathrm{M_\odot}$. They will be presented in an upcoming work [Gillet et al. (in preparation)].

For our purposes here, we take four simulation results that have among the best agreement with existing LF observations, but are different at the faint end and high redshifts at which we have little or no data. We can thus test the ability of our analytical framework to capture diverse, yet physically reasonable LFs.
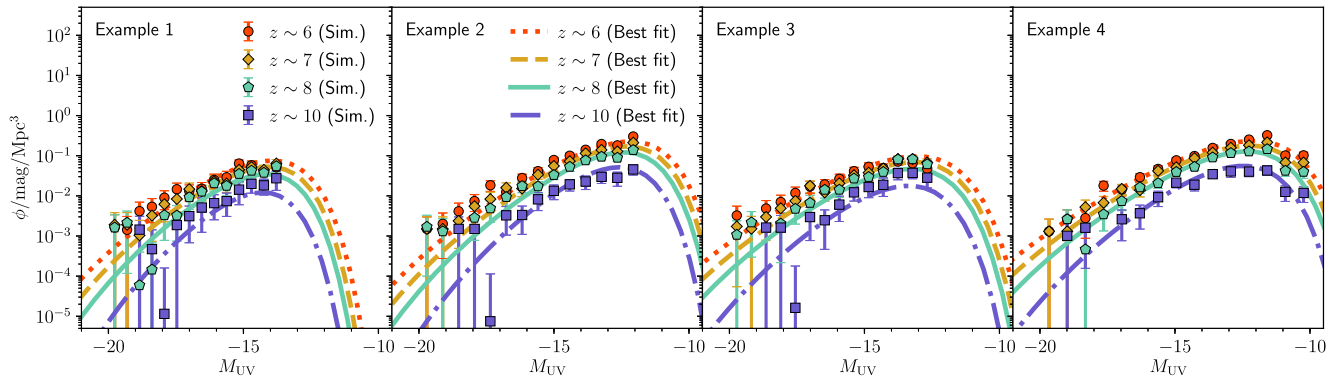
Figure A1. LFs from hydrodynamic simulations are shown with points, together with our best-fitting model. The hydrodynamic simulations were chosen to have good agreement with observed LFs at the bright end, but different trends at the faint end/high-z. We note that our analytic model, based on the HMF, does not have any free parameters that regulate redshift evolution.

We run an MCMC of our model parameters using the LFs from EMMA as a mock observation and find maximum likelihood parameters. We include Poisson errors for the numbers of both dark matter haloes and star particles, adding them in quadrature. Fig. A1 shows LFs generated from the simulation and the corresponding LFs with our maximum likelihood parameters. We find that in all four examples, our model is sufficiently flexible to fit the simulated LFs reasonably well.

APPENDIX B: 21-CM POWER SPECTRA

The light cone of the mock 21-cm observation is generated from 500 Mpc side length coeval cubes with a 256^3 grid, smoothed down from a high-resolution density field of 1024^3. To compute the mock 21-cm PS, we follow the same approach as Greig & Mesinger (2018). We split the light cone into equal comoving distance boxes and calculate the 21-cm PS (equation 22) for each separate box.

For the MCMC samples, we generate a light cone from 250 Mpc side length coeval cubes with a 128^3 grid, smoothed down from a 512^3 density field, but using different initial conditions. The mock observation is split into equal comoving distance boxes equivalent to the box length of the sampled boxes (i.e. 250 Mpc). Then, we compute the 21-cm PS from the same comoving scale for both the mock observation and the MCMC samples. Since the light cones extend from $z = 6$ (~ 200 MHz) to $z = 26.8$ (~ 50 MHz), this generates a total of 12 independent 'chunks'.

Fig. B1 shows the 21-cm PS for the mock observation generated (solid lines) and a sample 21-cm PS generated from the 250 Mpc box, using the same fiducial parameters but using a different initial seed. The shaded regions show the estimated noise corresponding to the mock 21-cm PS. We assume HERA for the noise estimate with a core design consisting of 331 dishes and a 1000-h observation. In each panel, we denote the central redshift for each 'chunk' of the light cone.
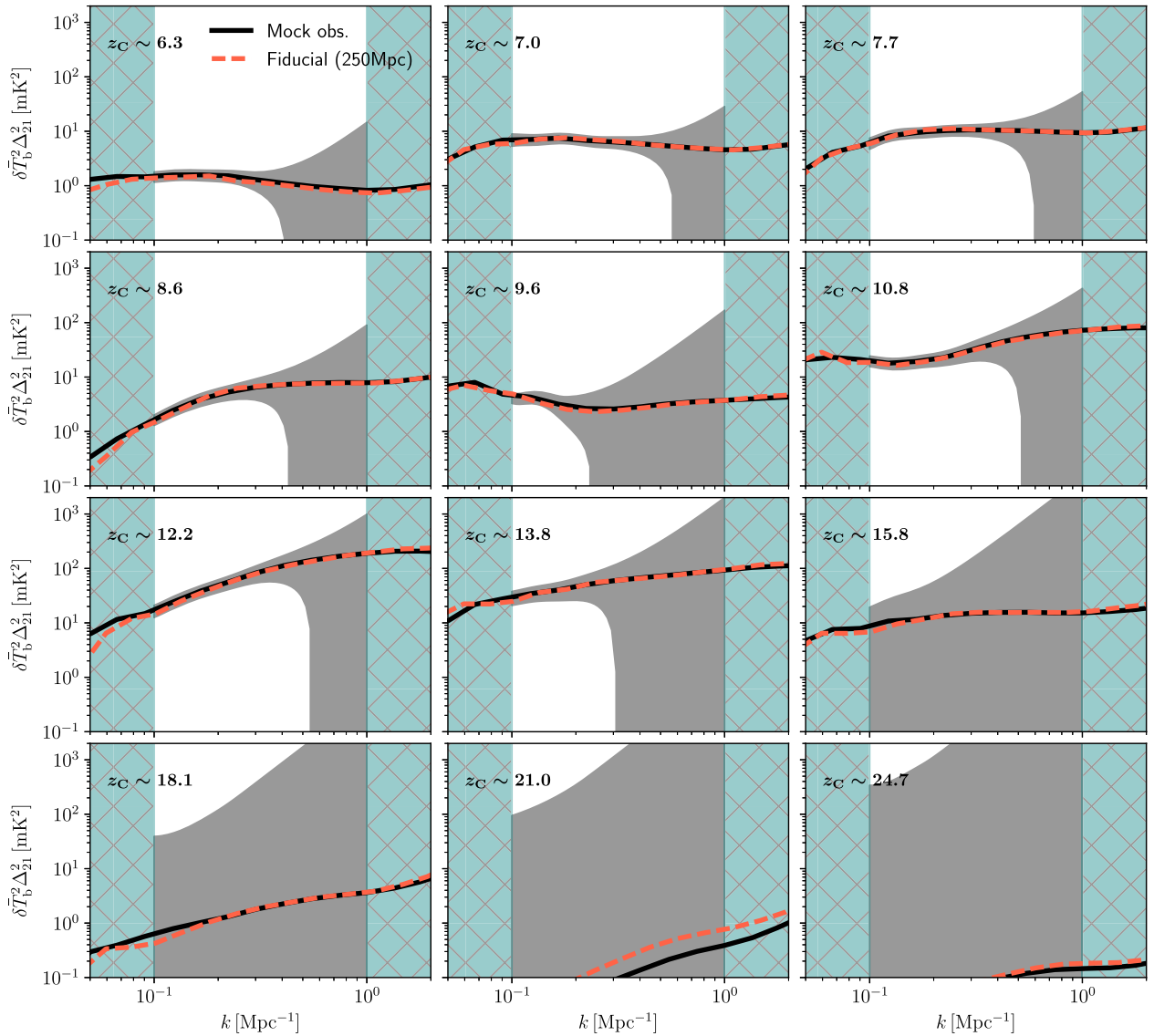
Figure B1. The 21-cm PS from the mock observation (solid lines), and corresponding 1σ errors assuming a 1000-h observation with HERA331. The dashed lines represent the MCMC sample with the fiducial parameters, but from a different random seed. The hatched regions represent k modes outside of our fitting range of $k = 0.1$–$1\ \mathrm{Mpc}^{-1}$. z_C denotes the central redshift of each *'chunk'* of the light cone.

APPENDIX C: IONIZING EMISSIVITY

Another potentially important data set on the high-z source population is the ionizing emissivity as estimated from the Lyman α forest. Here we study how this additional data set can further inform our models (cf. Choudhury & Ferrara 2006; Kuhlen & Faucher-Giguère 2012; Mitra et al. 2013; Bouwens et al. 2015b).

The ionizing emissivity is estimated by using the opacity measured from high-z quasar spectra. Post reionization, the optical depth in the IGM scales roughly as $\tau_{\mathrm{Ly}_\alpha} \propto T^{-0.7} \Delta_\mathrm{b}^2 / \Gamma$, where T is the gas temperature, Δ_b is the gas density in units of the cosmic mean, and Γ is the photoionization rate. The ionizing emissivity, ε, can then be estimated using the post-reionization relation $\Gamma \propto \lambda_{\mathrm{mfp}}^{912} \varepsilon$, where $\lambda_{\mathrm{mfp}}^{912}$ is the mean free path of ionizing photons. This emissivity can then be directly compared to our model prediction from equation 16.

This procedure is non-trivial for several reasons. First, the Lyman α forest is only sensitive enough at $z \lesssim 5$ to provide a reasonable estimate of the emissivity. The galaxies at these post-EoR redshifts could evolve beyond what is expected during the first billion years, due to feedback processes. Thus, they are not the same population that we are modelling. More importantly, although galaxies are expected to dominate the EoR, it is likely that the contribution of AGN ramps up soon afterwards and thus cannot be ignored at these lower redshifts (e.g. Haardt & Madau 2012; Chardin et al. 2015; Mitra, Choudhury & Ferrara 2018). We therefore take the emissivity estimates at $z \sim 5$ as *upper limits* to our galaxy emissivities.

Additionally, as explained above, we require knowledge of the IGM temperature, density, and mean free path in order to estimate the emissivity from the forest. This can be tricky by $z \sim 5$, with the mean free path being especially difficult to constrain to high precision. Moreover, spatial fluctuations in these quantities can bias estimates.

Here we explore the utility of IGM emissivity upper limits for our parameter study, using the estimates from D'Aloisio et al. (2018b). These authors estimated the ionizing emissivity at $4.8 < z < 5.8$ based on the measurement of $\tau_{\mathrm{Ly}_\alpha}$ by Becker et al.
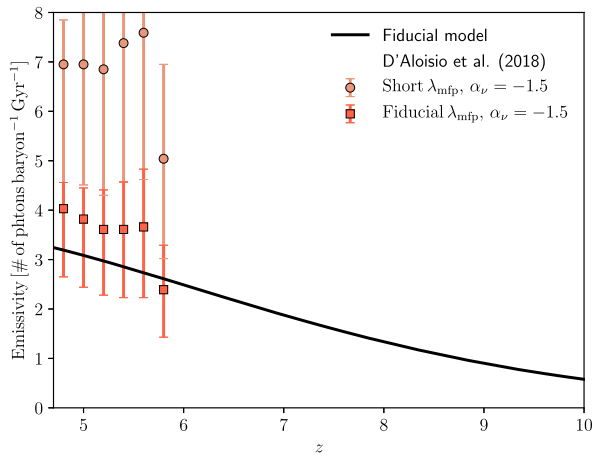
Figure C1. Redshift evolution of the ionizing emissivity. The solid line represents the prediction of our fiducial model. The squares and circles with error bars represent the measured emissivity by D'Aloisio et al. (2018b) with their fiducial mean free path and short mean free path, respectively. We note that to convert units we assume the ionizing specific luminosity follows a power law, $L_\nu \propto \nu^{\alpha_\nu}$, and adopt $\alpha = -1.5$, which is similar quantity in Lusso et al. (2015).

(2015). They post-processed simulations to compute a spatially varying photoionization rate, Γ, and rescaled it to fit the observed τ_{Ly_α} under the assumption $\lambda_{\mathrm{mfp}}^{912}(\mathbf{x}) \propto \Gamma^{2/3}(\mathbf{x})/\Delta(\mathbf{x})$, where $\Delta(\mathbf{x})$ is the local matter density and $\lambda_{\mathrm{mfp}}^{912}$ is the mean free path of ionizing photons. This rescaling provides the ionizing emissivity, ε_{912}, with the relation $\Gamma \propto \lambda_{\mathrm{mfp}}^{912}\epsilon_{912}$. They use three models for the mean free path, which they refer to as fiducial, intermediate, and short. The fiducial one is consistent with the mean free path measurements of

Worseck et al. (2014), which are at $z \leq 5.2$, though D'Aloisio et al. (2018b) argue that this might be an overestimate due to a bias from including the proximity zone in the mean free path calculation.

The resulting estimates of the ionizing background in the fiducial and short mean free path models are shown as points with error bars in Fig. C1. To convert the emissivity to number of photons per baryon per Gyr, we assume that the specific emissivity provided by D'Aloisio et al. (2018b) follows a power law, $L_\nu \propto \nu^{\alpha_\nu}$, and adopt $\alpha = -1.5$, consistent with Lusso et al. (2015) (see also e.g. McQuinn, Oh & Faucher-Giguère 2011; D'Aloisio et al. 2018a). With the solid curve, we also show the emissivity from our fiducial parameter set, used to generate the mock 21-cm signal. This emissivity is roughly consistent with the fiducial mean free path model of D'Aloisio et al. (2018b).

Given the uncertainties in these estimates, how constraining is the emissivity for our parameter space? To quantify this, we use the fiducial mean free path estimates of D'Aloisio et al. (2018b) (which are lower and thus more constraining) at $z \sim 5.4$ and 5.8 as upper limits (allowing for an additional AGN contribution as discussed above). Specifically, we take a flat prior for values lower than the points, and then a one-sided Gaussian decreasing for higher emissivities with the sigma reported by these authors. We then re-run our MCMC for the UV astrophysical parameters, with and without this additional data set.

The resulting constraints are shown in Fig. C2. Even for the conservatively strong prior of using the fiducial emissivity estimates (as opposed to the higher ones provided by D'Aloisio et al. 2018b), the constraints are quite comparable to those already presented in Fig. 4. However, we see that a 1σ percentage error for the escape fraction is reduced from ~ 40 to ~ 25 per cent, while the power law of the escape fraction is still not constrained.
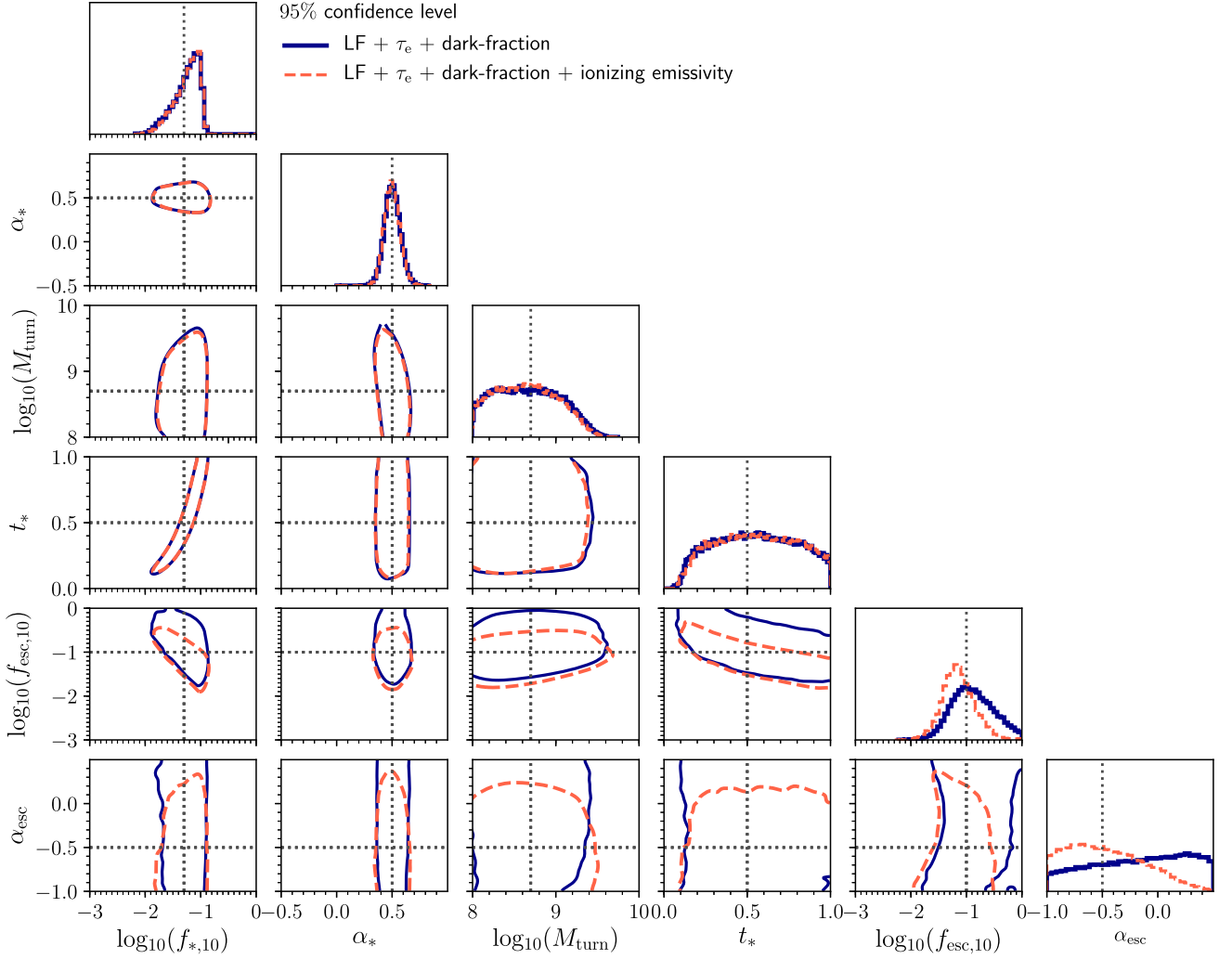
Figure C2. Marginalized joint posterior distributions for UV galaxy properties with and without a prior on the emissivity. The solid (blue) and dashed (red) lines represent 95 per cent confidence levels for constraints using LF $+ \tau_e +$ the dark fraction (same as in Fig. 4), and when additionally using the ionizing emissivity. The minor relative differences between these curves demonstrate that the ionizing emissivity currently has little additional constraining power for our model.

This paper has been typeset from a T_EX/L^AT_EX file prepared by the author.