

Research Article

Percent-level timing of reionisation: Self-consistent, implicit-likelihood inference from XQR-30+ Ly α forest data

Yuxiang Qin^{1,2} , Andrei Mesinger³ , David Prelogović^{3,4} , George Becker⁵ , Manuela Bischetti^{6,7} , Sarah Bosman^{8,9} , Frederick Davies⁹ , Valentina D’Odorico^{3,7} , Prakash Gaikwad⁹ , Martin Haehnelt¹⁰ , Laura Keating¹¹ , Samuel Lai¹² , Emma Ryan-Weber^{2,13} , Sindhu Satyavolu^{14,15} , Fabian Walter^{9,16} , and Yongda Zhu¹⁷ 

¹Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT, Australia, ²ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia, ³Scuola Normale Superiore, Pisa, Italy, ⁴Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy, ⁵Department of Physics & Astronomy, University of California, Riverside, CA, USA, ⁶Dipartimento di Fisica, Università di Trieste, Sezione di Astronomia, Trieste, Italy, ⁷INAF – Osservatorio Astronomico di Trieste, Trieste, Italy, ⁸Institute for Theoretical Physics, Heidelberg University, Heidelberg, Germany, ⁹Max-Planck-Institut für Astronomie, Heidelberg, Germany, ¹⁰Kavli Institute for Cosmology and Institute of Astronomy, Cambridge, UK, ¹¹Institute for Astronomy, University of Edinburgh, Edinburgh, UK, ¹²Commonwealth Scientific and Industrial Research Organisation (CSIRO), Space & Astronomy, Bentley, WA, Australia, ¹³Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, VIC, Australia, ¹⁴Tata Institute of Fundamental Research, Mumbai, India, ¹⁵Institut de Física d’Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Bellaterra, Barcelona, Spain, ¹⁶National Radio Astronomy Observatory, Pete V. Domenici Array Science Center, Socorro, NM, USA and ¹⁷Steward Observatory, University of Arizona, Tucson, AZ, USA

Abstract

The Lyman alpha (Ly α) forest in the spectra of $z > 5$ quasars provides a powerful probe of the late stages of the epoch of reionisation (EoR). With the recent advent of exquisite datasets such as XQR-30, many models have struggled to reproduce the observed large-scale fluctuations in the Ly α opacity. Here we introduce a Bayesian analysis framework that forward-models large-scale lightcones of intergalactic medium (IGM) properties and accounts for unresolved sub-structure in the Ly α opacity by calibrating to higher-resolution hydrodynamic simulations. Our models directly connect physically intuitive galaxy properties with the corresponding IGM evolution, without having to tune ‘effective’ parameters or calibrate out the mean transmission. The forest data, in combination with UV luminosity functions and the CMB optical depth, are able to constrain global IGM properties at percent level precision in our fiducial model. Unlike many other works, we recover the forest observations without invoking a rapid drop in the ionising emissivity from $z \sim 7$ to 5.5, which we attribute to our sub-grid model for recombinations. In this fiducial model, reionisation ends at $z = 5.44 \pm 0.02$ and the EoR mid-point is at $z = 7.7 \pm 0.1$. The ionising escape fraction increases towards faint galaxies, showing a mild redshift evolution at fixed UV magnitude, M_{UV} . Half of the ionising photons are provided by galaxies fainter than $M_{UV} \sim -12$, well below direct detection limits of optical/NIR instruments including *JWST*. We also show results from an alternative galaxy model that does not allow for a redshift evolution in the ionising escape fraction. Despite being decisively disfavoured by the Bayesian evidence, the posterior of this model is in qualitative agreement with that from our fiducial model. We caution, however, that our conclusions regarding the early stages of the EoR and which sources reionised the Universe are more model-dependent.

Keywords: Cosmology; theory; dark ages; reionisation; first stars; early Universe; galaxies: high-redshift; intergalactic medium

(Received 28 November 2024; revised 30 January 2025; accepted 24 February 2025)

1. Introduction

The epoch of reionisation (EoR) is a fundamental milestone in the evolution of our Universe. Its timing and spatial fluctuations encode invaluable information about the intergalactic medium (IGM) and the first galaxies. Recent years have witnessed a dramatic increase in the number and quality of observations probing the EoR, including upper limits on the cosmic

21-cm power spectrum (Mertens et al. 2020; Trott et al. 2020; HERA Collaboration et al. 2023), the polarisation anisotropy of the cosmic microwave background (CMB; Planck Collaboration et al. 2020; Reichardt et al. 2021), and the IGM Lyman- α (Ly α) damping-wing absorption seen in spectra of high-redshift quasars (Eduardo Bañados et al. 2018; Wang et al. 2020) and star-forming galaxies (Pentericci et al. 2018; Umeda et al. 2024; Heintz et al. 2024).

Arguably the most mature of EoR datasets is the Ly α forest. More than two decades of observational efforts have provided over 70 high-quality quasar spectra at $z > 5.5$ (Fan et al. 2002, 2006b,a; Willott et al. 2007; Becker et al. 2015; Wu et al. 2015; Eduardo Bañados et al. 2016; Jiang et al. 2016; Eilers, Davies, & Hennawi 2018; Yang et al. 2020; D’Odorico et al. 2023). These data provide

Corresponding author: Yuxiang Qin; Email: Yuxiang.L.Qin@gmail.com.

Cite this article: Qin Y, Mesinger A, Prelogović D, Becker G, Bischetti M, Bosman S, Davies F, D’Odorico V, Gaikwad P, Haehnelt M, Keating L, Lai S, Ryan-Weber E, Satyavolu S, Walter F and Zhu Y. (2025) Percent-level timing of reionisation: Self-consistent, implicit-likelihood inference from XQR-30+ Ly α forest data. *Publications of the Astronomical Society of Australia* 42, e049, 1–21. <https://doi.org/10.1017/pasa.2025.35>

unparalleled statistics over large volumes of the IGM. As such, the Ly α forest is one of the few EoR probes that is not sensitive to the biased environments proximate to the ionizing sources.

The high quality and quantity of Ly α forest data provide an invaluable stress test on our understanding of the EoR, as they are quite sensitive to missing components in our theoretical and systematic models. For instance, the observed large-scale fluctuations in the Ly α optical depth cannot be reproduced by the simplest, uniform ultraviolet background (UVB) models at $z > 5.2$ (Becker et al. 2015; Bosman et al. 2022). Various theoretical models have attempted to reproduce the observations by increasing fluctuations in the IGM temperature, mean free path (MFP) of ionizing photons, ionizing emissivity, and/or including an ongoing, patchy reionisation (D’Aloisio, McQuinn, & Trac 2015; Davies & Furlanetto 2016; D’Aloisio et al. 2017, 2018; Chardin, Puchwein, & Haehnelt 2017; Kulkarni et al. 2019; Keating et al. 2020; Meiksin 2020; Nasir & D’Aloisio 2020; Asthana et al. 2024). However, moving beyond ‘*this particular model is (in)consistent with the data*’ to ‘*this is the distribution of IGM and galaxy properties inferred from the data*’ is considerably more challenging, and can only be achieved in a physically-motivated, efficient Bayesian inference framework.

Previous work that reproduced the data relied heavily on *effective* (i.e. not physically interpretable) parameters and/or ad-hoc assumptions that ignore or fine-tune the redshift evolution of the mean transmission flux. For example, several studies found that in order to reproduce the forest data, the UV ionizing emissivity in their simulations has to be tuned to drop rapidly towards the end of the EoR, with up to a factor of 2 decrement over just $\Delta z \sim 0.5$ (~ 100 Myr at these redshifts; e.g. Kulkarni et al. 2019; Ocvirk et al. 2021; Fig. 6). Such short time-scales for the UVB evolution are difficult to justify physically (e.g. Sobacchi & Mesinger 2013) or to reconcile with the observed gradual evolution of the cosmic star formation rate (SFR) density from bright galaxies (Bouwens et al. 2015b; Oesch et al. 2018). Indeed subsequent analysis pointed to unresolved substructure in the simulations as a possible explanation (e.g. see section 5.4 in Qin et al. 2021, and the recent analysis in Cain et al. 2024). Alternatively, simulations that tune the ionizing MFP without modelling the time evolution of HII regions and/or adopt effective parameters for inhomogeneous recombinations are also difficult to interpret as they only provide a somewhat opaque proxy for cosmic reionisation (e.g. Choudhury, Paranjape, & Bosman 2021; Gaikwad et al. 2023; Davies et al. 2024).

Ideally, one should use a self-consistent model in which the redshift evolution of the patchy reionisation is simulated directly from the galaxies that drive it. This would allow us to set well-motivated priors on *physical* parameters that can be constrained by complementary galaxy observations (e.g. Park et al. 2019; Mutch et al. 2024). Anchoring the EoR models on galaxies also allows us to constrain earlier epochs where we have no forest measurements, since structure evolution (i.e., the halo mass function) is comparably well understood (e.g. Sheth et al. 2001) and we have complementary observations of UV luminosity functions (LFs) that constrain how halos are populated with galaxies at these high redshifts.

However, such self-consistent modelling of the EoR is inherently extremely challenging, due to the enormous dynamic range of relevant scales. Fluctuations in the Ly α forest are correlated on scales larger than ~ 100 cMpc (e.g. Becker et al. 2021; Zhu et al. 2021), while galaxies and IGM clumps are on sub-kpc scales (e.g. Schaye 2001; Emberson, Thomas, & Alvarez 2013; Park et al.

2016; D’Aloisio 2020). As a result, current simulations must rely on sub-grid prescriptions that have to be calibrated against observations or other more detailed, higher resolution simulations.

Here, we present an updated Bayesian inference framework for the high-redshift Ly α forest that is arguably free from ‘effective’ parameters. We sample physically-intuitive galaxy scaling relations to compute large-scale lightcones of the Ly α opacity using 21cmFAST (Mesinger, Furlanetto, & Cen 2011; Murray et al. 2020). This self-consistently connects galaxy properties to the state of the IGM that is shaped by their radiation fields. We account for missing small-scale structure by calibrating to the Sherwood suite of high-resolution hydrodynamic simulations (Bolton et al. 2017). This calibration allows us to eliminate the poorly-motivated hyperparameters we previously used to account for missing systematics and/or physics (Qin et al. 2021, hereafter Q21). For each astrophysical parameter combination, we forward model the forest transmission, comparing against the observations (Bosman et al. 2022) using an *implicit* likelihood. We present the resulting joint constraints on reionisation and galaxy properties, implied by the combined data from the Ly α forest, UV LFs, and CMB optical depth.

This paper is organised as follows. We summarise the extended XQR-30 Ly α forest data in Section 2, and introduce our Bayesian framework for forward-modelling Ly α forests in Section 3. After summarizing the complementary observations and free parameters used in this work in Sections 3.5 and 3.6, we present results in Section 4 including the recovered properties of the IGM and those of the underlying galaxies. We then discuss the implication to our understanding of reionisation in Section 5, before concluding in Section 6. In this work, we adopt cosmological parameters from Planck ($\Omega_m, \Omega_b, \Omega_\Lambda, h, \sigma_8, n_s = 0.312, 0.0490, 0.688, 0.675, 0.815, 0.968$; Planck Collaboration et al. 2016). Distance units are comoving unless otherwise specified.

2. The Ly α opacity distributions from XQR-30+

The ultimate XSHOOTER legacy survey of quasars at $z \sim 5.8$ –6.6 (XQR-30) is a ~ 250 -h programme using the Very Large Telescope (VLT) at the European Southern Observatory (ESO; D’Odorico et al. 2023). While XQR-30 contains 30 high-quality quasar spectra, Bosman et al. (2022) assembled 67 sightlines at these redshifts by combining XQR-30 with archival spectra. We refer to this extended dataset as XQR-30+.

The Ly α transmission in these spectra was quantified by the commonly-used ‘effective optical depth’, $\tau_{\text{eff}} \equiv -\ln\langle\mathcal{F}_\alpha\rangle_{\Delta z=0.1}$. Here $\mathcal{F}_\alpha(\lambda)$ is the continuum-normalised flux in the Ly α forest, which is averaged over segments of width $\Delta z = 0.1$ (roughly corresponding to ~ 40 cMpc at these redshifts). Non-detections (2σ) were assigned lower limits on τ_{eff} corresponding to twice the mean flux noise in the corresponding segment. The full XQR-30+ sample has at least ~ 10 estimates of τ_{eff} in each redshift bin spanning $z = 5.1, 5.2, \dots, 6.1$. We show the cumulative distribution functions (CDFs) of these τ_{eff} estimates in Fig. 3, where we also compare them to our fiducial posterior. For more details on how the observations were processed, see Bosman et al. (2022).

3. Forward modelling

We use the public simulation code, 21cmFAST^a (Mesinger & Furlanetto 2007; Mesinger, Furlanetto, & Cen 2011; Murray et al.

^a<https://github.com/21cmfast/21cmFAST>.

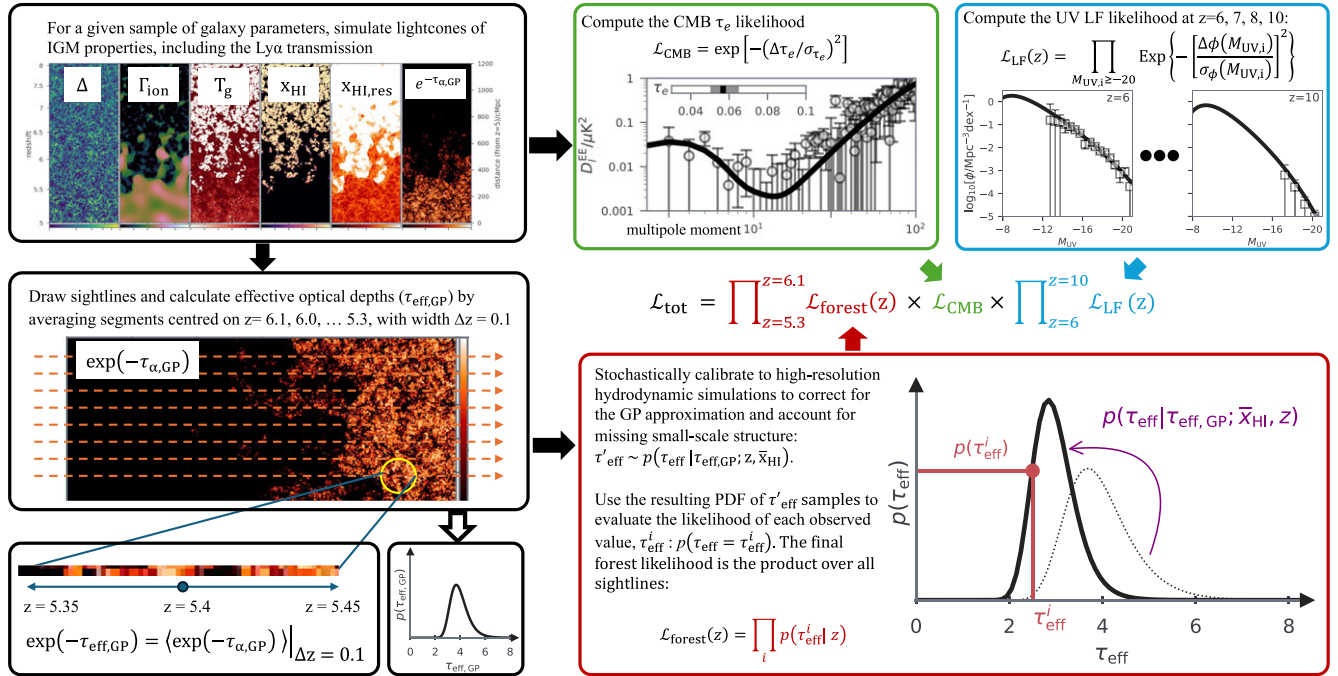


Figure 1. A flow chart showing the steps involved in computing the likelihood for a single sample of astrophysical parameters. See text for more details.

2020), to compute 3D lightcones of the Ly α IGM opacity. A single forward model and the corresponding likelihood evaluation are summarised in the flow chart of Fig. 1 and consist of the following steps:

1. Simulate large-scale 3D lightcones of the IGM density ($\Delta \equiv \rho/\bar{\rho}$), neutral fraction due to inhomogeneous reionisation (x_{HI}), photo-ionisation rate (Γ_{ion}), IGM temperature (T_g), residual neutral fraction inside the ionised IGM ($x_{\text{HI,res}}$) and corresponding Ly α opacity (top left panel of Fig. 1);
2. Construct mock quasar sightlines and compute the effective optical depth by binning the sightlines over the same redshift intervals as the XQR-30+ observation (lower left panels of Fig. 1);
3. Account for missing small scales by calibrating these effective optical depths against high-resolution hydrodynamic simulations. Use the resulting probability density function (PDF) of calibrated τ_{eff} to evaluate the likelihood of the observed values (lower right panel of Fig. 1);
4. Multiply this forest likelihood with the corresponding UV LF and CMB likelihoods in order to obtain the total likelihood of this parameter sample (upper right panels in Fig. 1; c.f. Section 3.5).

We discuss this procedure in detail below, emphasizing the improvements over our previous analysis in Q21.

3.1 Galaxy models

Our galaxy models are based on the semi-empirical parametrisation in Park et al. (2019). We assume power laws relating the fraction of galactic baryons in stars (f_*) and the UV ionizing escape fraction (f_{esc}) to the host halo mass (M_{vir}):

$$f_* = \min \left[1, f_{*,10} \left(\frac{M_{\text{vir}}}{10^{10} M_{\odot}} \right)^{\alpha_*} \right] \quad (1)$$

and

$$f_{\text{esc}} = \min \left[1, f_{\text{esc},10} \left(\frac{M_{\text{vir}}}{10^{10} M_{\odot}} \right)^{\alpha_{\text{esc}}} \left(\frac{1+z}{8} \right)^{\beta_{\text{esc}}} \right], \quad (2)$$

where $f_{*,10}$, α_* , $f_{\text{esc},10}$, α_{esc} and β_{esc} are free parameters. Compared to Park et al. (2019) and our previous analysis in Q21, here we allow for an additional redshift dependence of f_{esc} at a given halo mass through the parameter β_{esc} (e.g. Haardt & Madau 2012; Kuhlen & Faucher-Giguère 2012; Mutch et al. 2016). Note that f_* and f_{esc} have to be in the range from zero to unity as they are fractions.

The average SFRs of galaxies over the past 100 Myr are computed as $\text{SFR} = M_* / [\tau_* H^{-1}(z)]$, where $M_* \equiv f_* M_{\text{vir}} \Omega_b / \Omega_m$ is the stellar mass, and τ_* is an additional free parameter corresponding to the characteristic star formation time-scale in units of the Hubble time, $H^{-1}(z)$, which scales as the halo dynamical time during matter domination. Note that the 1 500 Å rest-frame luminosity used in photometric UV LF observations is sensitive to star formation over the previous 100 Myr (e.g., Flores Velázquez et al. 2021). When forward-modeling UV LFs, we adopt the conversion factor, $L_{\text{UV}}/\text{SFR} = 8.7 \times 10^{27} \text{ erg s}^{-1} \text{ Hz}^{-1} M_{\odot}^{-1} \text{ yr}$ (e.g. Madau & Dickinson 2014).

We also assume only a fraction $f_{\text{duty}} \equiv \exp[-M_{\text{turn}}/M_{\text{vir}}]$ of halos host star-forming galaxies. Here, M_{turn} characterises the halo mass below which star formation becomes inefficient due to feedback and/or atomic cooling limits and is left as a free parameter.

Below we explore two galaxy models, differing in their treatment of the ionizing escape fraction:

Table 1. Posterior distribution ([16, 84]th percentiles) and Bayesian evidence of the galaxy models used in this work. The Bayes ratio indicates a very strong preference for the *Evolving*_{f_{esc}} model, according to Jeffrey’s scale (e.g. Jeffreys 1939).

	$\log_{10} f_{*,10}$	α_*	$\log_{10} f_{\text{esc},10}$	α_{esc}	β_{esc}	τ_*	$\log_{10} (M_{\text{turn}}/M_{\odot})$	$\ln \mathcal{B}^*$
Prior range	$[-2, -0.5]$	$[0, 1]$	$[-3, 0]$	$[-1, 0.5]$	$[-3, 3]$	$(0, 1]$	$[8, 10]$	–
<i>Evolving</i> _{f_{esc}} †	-1.51 ± 0.03	0.48 ± 0.05	$-1.52^{+0.12}_{-0.10}$	$-0.94^{+0.09}_{-0.04}$	$-1.61^{+0.27}_{-0.21}$	$0.27^{+0.02}_{-0.01}$	$8.10^{+0.16}_{-0.07}$	1
<i>Constant</i> _{f_{esc}} ‡	-1.42 ± 0.04	0.51 ± 0.07	$-1.08^{+0.09}_{-0.06}$	$-0.46^{+0.09}_{-0.13}$	Fixed at 0	0.34 ± 0.02	$8.46^{+0.51}_{-0.35}$	–17.5

* Bayes ratio w.r.t. *Evolving*_{f_{esc}} in natural logarithmic scale.

† Galaxies have a mass-dependent and time-evolving escape fraction.

‡ Galaxies have a mass-dependent and time-independent escape fraction.

1. *Constant*_{f_{esc}} – the ionizing escape fraction is a function of halo mass only and is constant with redshift (fixing β_{esc} to zero in equation 2). Note that this model does effectively allow for the population-averaged escape fraction to evolve with redshift, since f_{esc} depends on halo mass and the halo mass function evolves with redshift. This sets a ‘characteristic’ halo mass that drives both the timing and morphology of reionisation.
2. *Evolving*_{f_{esc}} – the ionizing escape fraction is a function of halo mass and evolves with redshift (treating both α_{esc} and β_{esc} as free parameters in equation 2). Note that adding an explicit redshift dependence to the escape fraction at a fixed halo mass gives the *Evolving*_{f_{esc}} model the flexibility to decouple the EoR/UVB morphology from the mean EoR history.

In this work we perform inference with both models, comparing their Bayesian evidences. We find that the data strongly prefer *Evolving*_{f_{esc}}, and we therefore refer to this model as ‘fiducial’. We list the posterior distribution and Bayesian evidence of these two models in Table 1.

3.2 Large-scale IGM simulations

Our simulation boxes are 250 cMpc on a side. Realisations of Gaussian initial conditions are computed at $z = 300$ on a 640^3 grid, with the density fields evolved down to $z = 5$ using second order Lagrangian perturbation theory (2LPT; Scoccimarro 1998) and smoothed down to a final resolution of 128^3 . Galaxy abundances are identified from the evolved density fields using excursion-set theory (Mesinger, Furlanetto, & Cen 2011), and assigned properties including the stellar mass, SFR, ionizing escape fraction and duty cycle according to the galaxy models discussed in the previous section.

Reionisation is modeled with the excursion-set approach (Furlanetto, Zaldarriaga, & Hernquist 2004), accounting for inhomogeneous recombinations (Sobacchi & Mesinger 2014). Unlike Q21, here we include a correction for photon-conservation (Park, Greig, & Mesinger 2022), which further decreases the need for the nuisance hyperparameters used in our previous work. Specifically, a cell is flagged as ionised when the cumulative number of ionizing photons per baryon reaching it:

$$\bar{n}_{\text{ion}} = \int dM_{\text{vir}} \phi f_{\text{duty}} M_* f_{\text{esc}} n_{\gamma} \rho_b^{-1}, \quad (3)$$

exceeds the cumulative number of recombinations per baryon ($\bar{n}_{\text{rec}}$; accounting for unresolved substructure with the analytic framework of Sobacchi & Mesinger 2014):

$$\bar{n}_{\text{ion}} \geq (1 + \bar{n}_{\text{rec}}). \quad (4)$$

In the above equations, ϕ , n_{γ} and ρ_b represent the conditional halo mass function, number of ionizing photons per stellar baryon which we fix at 5 000, and baryon density. The averaging is performed over spherical regions around each cell for radii $R \leq R_{\text{MFP,LLS}}$. Here $R_{\text{MFP,LLS}}$ corresponds to the MFP through the ionised IGM and is governed by damped Ly α systems (DLAs), Lyman limit systems (LLSs) and other unresolved systems with lower column densities (Nasir et al. 2021; Feron et al. 2024; Georgiev, Mellema, & Giri 2024).

Before the end of the EoR, the total MFP determining the local ionizing background is set by a combination of $R_{\text{MFP,LLS}}$ and the distance to the surrounding neutral IGM, $R_{\text{MFP,EoR}}$, i.e. $R_{\text{MFP}}^{-1} = R_{\text{MFP,LLS}}^{-1} + R_{\text{MFP,EoR}}^{-1}$ (e.g. Alvarez & Abel 2012). The reionisation topology computed with our excursion-set algorithm determines the local (inhomogeneous) $R_{\text{MFP,EoR}}$ around each cell. However, since we do not directly resolve the spatial distribution of LSSs and DLAs when these become rare/biased, we assume a homogeneous value for $R_{\text{MFP,LLS}} = 66 [(1+z)/6.3]^{-4.3}$ cMpc at $z \leq 6$ motivated by post-EoR measurements (Worseck et al. 2014; see also Songaila & Cowie 2010 and Becker et al. 2021).^b In future work we will expand our model to additionally sample the mean and variance of $R_{\text{MFP,LLS}}$, allowing us to extend our analysis to even lower redshifts.

With the above, we compute the local photoionisation rate as

$$\Gamma_{\text{ion}} = (1+z)^2 R_{\text{MFP}} \sigma_{\text{H}} \frac{\alpha_{\text{UVB}}}{\alpha_{\text{UVB}} + \beta_{\text{H}}} \times \int dM_{\text{vir}} \phi f_{\text{duty}} \frac{M_*}{\tau_* H^{-1}} f_{\text{esc}} n_{\gamma} m_p^{-1} \quad (5)$$

where m_p is the proton mass, $\alpha_{\text{UVB}} = 2$ corresponds to the effective spectral index of the UVB (see Becker & Bolton 2013; D’Aloisio et al. 2019), $\beta_{\text{H}} = 2.75$ and $\sigma_{\text{H}} = 6.3 \times 10^{-18} \text{cm}^2$ characterise the photo-ionisation cross-section $\sigma(\nu) = \sigma_{\text{H}} \left(\frac{\nu}{\nu_{\text{H}}} \right)^{\beta_{\text{H}}}$ with ν_{H} corresponding to the Lyman limit. After a cell is ionised, its residual neutral fraction is determined assuming photo-ionisation equilibrium:

$$x_{\text{HI, res}} f_{\text{ion,ss}} \Gamma_{\text{ion}} = \chi_{\text{HeII}} \Delta \bar{n}_{\text{H}} (1 - x_{\text{HI, res}})^2 \alpha_{\text{B}} \quad (6)$$

where $\bar{n}_{\text{H}}$ is the mean hydrogen number density while Δ is the cell’s overdensity, $\chi_{\text{HeII}} \sim 1.08$ accounts for singly ionised helium, α_{B} is the case-B recombination coefficient, and $f_{\text{ion,ss}}$ accounts for gas self-shielding (Rahmati et al. 2013; see also Chardin,

^b At $z > 6$ where we do not have direct measurements, we set $R_{\text{MFP,LLS}} = 42$ cMpc. The value of $R_{\text{MFP,LLS}}$ at these high redshifts is highly uncertain, depending on the heating history of the IGM (Emberson, Thomas, & Alvarez 2013; Park et al. 2016; D’Aloisio 2020). However, during reionisation the MFP is dominated by the reionisation topology (i.e. $R_{\text{MFP,LLS}} > R_{\text{MFP,EoR}}$; Fig. 5 and Sobacchi & Mesinger 2014). Thus the exact value of $R_{\text{MFP,LLS}}$ at $z > 6$ should have a negligible impact on the EoR and the corresponding Ly α opacity distributions for realistic scenarios (see also Cain et al. 2023).

Kulkarni, & Haehnelt 2018). Note that sub-grid physics are implemented (Sobacchi & Mesinger 2014) when calculating recombinations with the sub-grid density unresolved by our simulation cells assumed to follow a volume-weighted distribution of $P_V(\Delta_{\text{sub}}, z)$ from Miralda-Escudé, Haehnelt, & Rees (2000). However, we use the cell's mean overdensity when computing the Ly α optical depth, which neglects unresolved opacity fluctuations when calibrating to the hydrodynamic simulations below. This will be improved in future work.

The IGM temperature (T_g) is tracked following McQuinn & Upton Sanderbeck (2016):

$$T_g^\gamma = T_{\text{ion},I}^\gamma \left[\left(\frac{\mathcal{Z}}{\mathcal{Z}_{\text{ion}}} \right)^3 \frac{\rho_b}{\rho_{b,\text{ion}}} \right]^{\frac{2\gamma}{3}} \frac{\exp(\mathcal{Z}^{2.5})}{\exp(\mathcal{Z}_{\text{ion}}^{2.5})} + T_{\text{lim}}^\gamma \frac{\rho_b}{\rho_b}, \quad (7)$$

where we denote $\mathcal{Z} = (1+z)/7.1$ and use the subscript 'ion' to indicate values at the time the cell was first ionised for convenience. $\gamma = 1.7$ is the equation of state index while $T_{\text{lim}} = 1.8\mathcal{Z} \times 10^4$ K (Hui & Gnedin 1997; Theuns et al. 1998; Puchwein et al. 2015) and $T_{\text{ion},I} = 2 \times 10^4$ K (D'Aloisio et al. 2019) are the relaxation and post I-front temperatures, respectively. Note that the scatter in $T_{\text{ion},I}$ has a negligible impact on the Ly α forest (e.g. Davies et al. 2019).

Finally, we compute the associated Ly α optical depth of each 1.95 cMpc simulation cell using a form of the Fluctuating Gunn-Peterson Approximation (FGPA; Gunn & Peterson 1965; Weinberg et al. 1999) for ionised cells:

$$\tau_{\alpha,\text{GP}} = \sqrt{\frac{3\pi\sigma_T}{8}} f_\alpha \lambda_\alpha c H^{-1} n_H x_{\text{HI}}. \quad (8)$$

Here σ_T , $f_\alpha = 0.416$, $\lambda_\alpha = 1216$ Å and n_H are the Thomson cross-section, Ly α oscillator strength, Ly α rest-frame wavelength, and hydrogen number density, respectively. Finally, we compute the effective optical depth, $\tau_{\text{eff},\text{GP}}$, following the same definition as the observation (see Section 2).

The FGPA approximates the cross-section of Ly α absorption as a Dirac delta function at resonance and ignores peculiar velocities of the gas. In the following section we discuss how we use high-resolution hydrodynamic simulations to correct for the FGPA, accounting for missing small-scale structure. This represents the main improvement of this work over our previous analysis in Q21.

3.3 Accounting for missing small-scale structure

As mentioned above, our large-scale IGM simulations have a cell size of 1.95 cMpc. This is a factor of few larger than the typical Jeans length in the ionised IGM and the width of the Ly α cross-section at resonance. As a result, we use the FGPA in equation (8) instead of directly integrating over the full Voigt profile for the Ly α cross-section, σ_α , and accounting for gas peculiar velocities:

$$\tau_\alpha = \int \frac{dz}{1+z} c H^{-1} n_H x_{\text{HI}} \sigma_\alpha. \quad (9)$$

Does this approximation impact our modelled τ_{eff} distributions?

The fact that τ_{eff} is defined over $\Delta z = 0.1$ (corresponding to roughly 20 of our IGM simulation cells) would suggest that this summary statistic is mostly sensitive to (resolved) large-scale fluctuations in flux. However, not resolving small-scale structures can effectively alias power towards large scales (e.g. Viel et al. 2005; Kooistra, Lee, & Horowitz 2022). Here we use a high-resolution hydrodynamic simulation from the Sherwood suite

(Bolton et al. 2017) to compare $\tau_{\text{eff},\text{GP}}$ obtained from the low-resolution FGPA (equation 8) against the correct calculation (equation 9).

We use a simulation with a cubic volume of $80h^{-1}$ cMpc on a side and 2×512^3 particles. It was performed using an updated version of *Gadget-2* (Springel et al. 2005) and with a slightly different Λ CDM cosmology ($\Omega_m, \Omega_b, \Omega_\Lambda, h, \sigma_8, n_s = 0.31, 0.048, 0.69, 0.68, 0.83, 0.96$; Planck Collaboration et al. 2016). The modelled universe is exposed to a Haardt & Madau (2012) UVB switched on at $z = 15$. Although this homogeneous UVB does not account for the effects of patchy reionisation, the simulated Ly α forests from Sherwood, which has a spatial resolution of < 60 kpc, agree very well with observational data at $z \leq 5$ (Viel et al. 2013; Bolton et al. 2017) and therefore can be used to calibrate our forward models in the post reionisation regime (i.e. $\bar{x}_{\text{HI}} = 0$).

We project sightlines along each axis of the $z = 5$ snapshot.^c This results in 5 000 segments of length $80 h^{-1}$ cMpc, which we bin to $\Delta z = 0.1$. As the physical scale corresponding to $\Delta z = 0.1$ changes with redshift, we repeat the binning for all redshifts spanned by the data, $z = 5.1, 5.2, 5.3, \dots, 6.1$. For each bin, we compute the 'true' effective optical depth (τ_{eff} ; i.e. averaging the flux obtained using equation 9). We then recompute the effective optical depth of each segment assuming the same approximations we make in our large-scale IGM simulations (down-sampling the resolution and applying equation 8) to obtain the corresponding $\tau_{\text{eff},\text{GP}}$. We are then left with pairs of $\tau_{\text{eff}} - \tau_{\text{eff},\text{GP}}$, which act as samples of the conditional probability of having a *true* τ_{eff} given the corresponding FGPA value $\tau_{\text{eff},\text{GP}} \sim p(\tau_{\text{eff}} | \tau_{\text{eff},\text{GP}}; z, \bar{x}_{\text{HI}} = 0)$.

We then generalise this conditional probability to higher neutral fractions. Specifically, we randomly place spherical neutral IGM patches in the Sherwood box until we obtain an HI filling factor of $\bar{x}_{\text{HI}}$, repeating the above procedure to obtain $p(\tau_{\text{eff}} | \tau_{\text{eff},\text{GP}}; z, \bar{x}_{\text{HI}})$. We assume a log normal distribution peaked at a constant value of 4 cMpc for the radii of these HI patches. This is motivated by the results of Xu, Yue, & Chen (2017), who find a very modest evolution in the neutral patch size distribution during the final stages of the EoR.

The resulting $\tau_{\text{eff}} - \tau_{\text{eff},\text{GP}}$ samples at $z = 5$ are shown in Fig. 2 where $\tau_{\text{eff}} = \tau_{\text{eff},\text{GP}}$ is marked by a diagonal line in each panel. At low values of the neutral fraction, the FGPA tends to overestimate the true value of the effective optical depth. This is especially evident at large overdensities with high values of τ_{eff} (Kooistra, Lee, & Horowitz 2022). As the neutral fraction increases, this bias decreases. However, the scatter in $p(\tau_{\text{eff}} | \tau_{\text{eff},\text{GP}}; z, \bar{x}_{\text{HI}})$ increases significantly. At $x_{\text{HI}} \gtrsim 0.5$ when damping-wing absorption becomes significant, the FGPA starts to instead underestimate the effective optical depth.

We fit these samples with kernel density estimators (KDEs) in order to obtain an analytic form for $p(\tau_{\text{eff}} | \tau_{\text{eff},\text{GP}}; z, \bar{x}_{\text{HI}})$ that can be evaluated when forward modelling (c.f. Fig. 1). Specifically, we use 2D conditional Gaussian distributions from the `conditional_kde`^d package to fit the samples of $1/\tau_{\text{eff}} - 1/\tau_{\text{eff},\text{GP}}$ (as the reciprocal of the optical depth more closely follows a Gaussian distribution). The parameters of the Gaussian

^cUnfortunately, we did not have snapshots available at every redshift probed by observations, $z = 5.1, 5.2, 5.3, \dots$. We therefore perform our calibration only using the $z = 5$ snapshot and assume the conditional distribution functions to be self-similar at other redshifts. We crudely tested this assumption by scaling the density field by its mean evolution, finding only a $\tau_{\text{eff}} \lesssim 0.5$ shift in the resulting conditionals. We plan to improve the calibration in the future using more snapshots from higher resolution simulations.

^dhttps://github.com/dprelogo/conditional_kde.

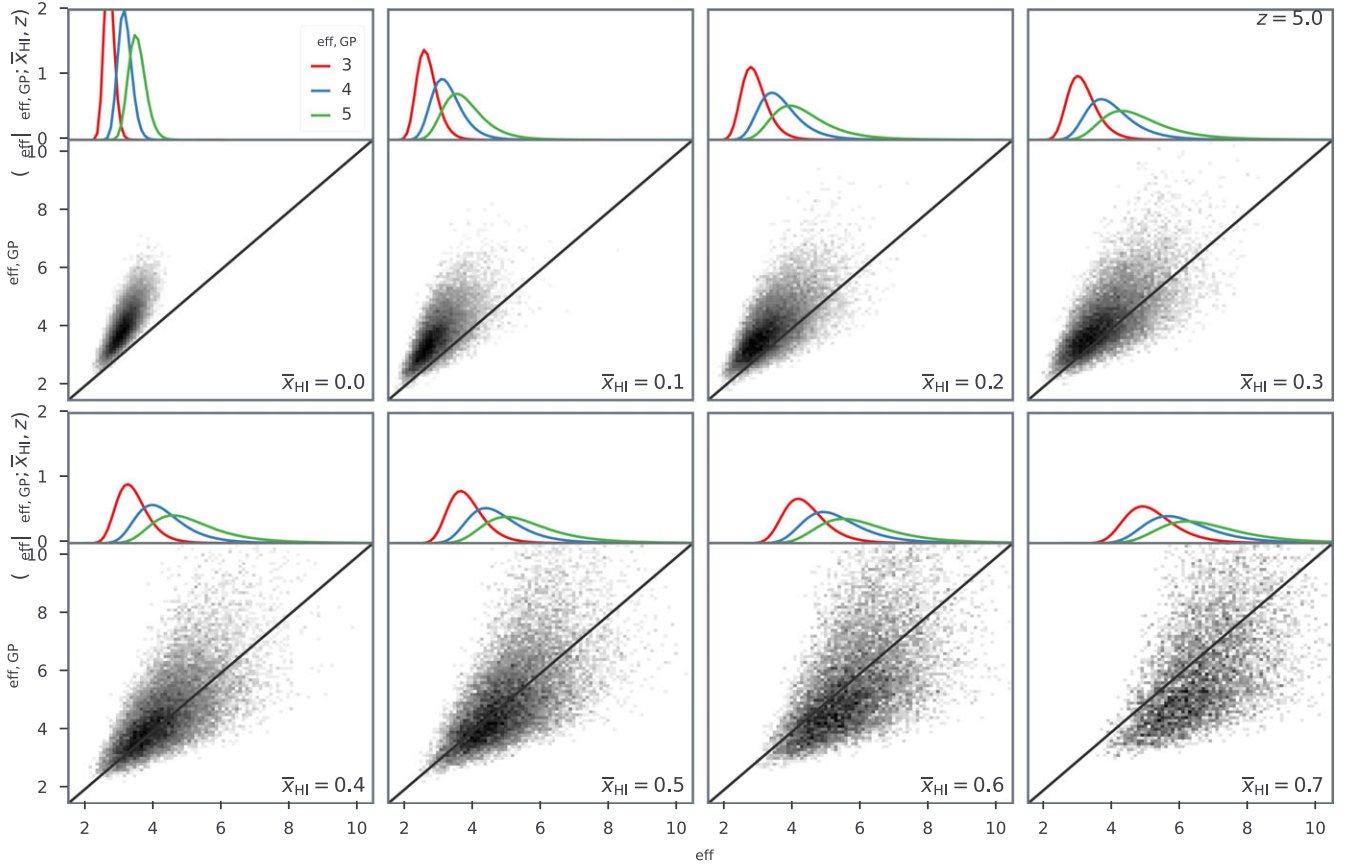


Figure 2. Lower sub-panels: comparisons of the Ly α effective optical depth calculated using the full integral over the Ly α cross-section at the highest available resolution (τ_{eff}), to those calculated assuming the FGPA ($\tau_{\text{eff,GP}}$). Both calculations use the Sherwood hydrodynamic simulation, with the latter obtained by down-sampling to the same low resolution adopted in our IGM forward-models and ignoring peculiar velocities. These sub-panels show pairs of $\tau_{\text{eff}} - \tau_{\text{eff,GP}}$ at different values of the mean neutral fraction, $\bar{x}_{\text{HI}}$ – an incomplete EoR is approximated by randomly placing spherical neutral patches in the simulation box until the desired filling factor of $\bar{x}_{\text{HI}}$ is reached. These distributions of ($\tau_{\text{eff}}, \tau_{\text{eff,GP}}$) pairs are fit with KDE, resulting in a conditional probability distribution function $p(\tau_{\text{eff}} | \tau_{\text{eff,GP}}, \bar{x}_{\text{HI}}, z)$, which is employed to correct our forward-modelled IGM lightcones for missing small scales. Upper sub-panels: example τ_{eff} distributions conditioned at $\tau_{\text{eff,GP}} = 3, 4$ and 5 .

kernels (means and standard deviations) are explicit functions of redshift and the neutral fraction,^c allowing us to easily evaluate $p(\tau_{\text{eff}} | \tau_{\text{eff,GP}}, z, \bar{x}_{\text{HI}})$ at any neutral fraction and redshift. We show some examples of the fitted conditional distributions in the upper sub-panels of Fig. 2.

3.4 Computing the forest likelihood

For each $\tau_{\text{eff,GP}}(\bar{x}_{\text{HI}}, z)$ calculated using equation (8) on our IGM lightcones, we obtain a random sample from the conditional distributions discussed in the previous subsection: $\tau_{\text{eff}} \sim p(\tau_{\text{eff}} | \tau_{\text{eff,GP}}, z, \bar{x}_{\text{HI}})$. Therefore, this leads to a set of effective optical depths that are stochastically corrected for missing substructure in the FGPA method. We additionally account for uncertainty in the continuum reconstruction by adding $\ln(\mathcal{R})$ to every τ_{eff} sample. Here, $\mathcal{R}$ is a random number following a normal distribution centred at unity with a standard deviation of 10%, typical of the continuum reconstruction relative errors (Bosman et al. 2022). Note that we do not account for wavelength correlations in

the reconstruction errors or the actual ‘usable’ range of observed wavelengths in each quasar spectrum (see more in Bosman et al. 2022); we plan on including these in future work. We fit the resulting histograms of τ_{eff} in each of the redshift bins defined by the data to obtain the PDFs, $p(\tau_{\text{eff}}; z)$.

These PDFs are our theoretical expectation of the real Universe, for a given model and choice of astrophysical parameters. Therefore, each *observed* value of τ_{eff}^i at z^i corresponds to a sample from the theoretical PDF, with a corresponding likelihood $p(\tau_{\text{eff}} = \tau_{\text{eff}}^i; z = z^i)$. For non-detections, we take τ_{eff}^i to be the 2σ lower limit implied by the noise (Bosman et al. 2022).

It is worth noting that the likelihood distribution, $p(\tau_{\text{eff}}, z)$, is forward-modeled, meaning it is sampled by running the simulator many times, varying the most relevant sources of stochasticity together with the astrophysical parameters. This is known as an implicit likelihood, as it avoids the need to specify an explicit likelihood function such as a Gaussian, when comparing data to model. Instead, the simulator itself generates data realisations allowing the observed data to be treated as a sample of this data distribution, for the correct astrophysical parameters. This makes it particularly powerful for cases where the likelihood is too complex or intractable to express analytically (Cranmer, Brehmer, & Louppe 2020; Davies et al. 2024). Indeed, inferences using an implicit

^cThroughout this paper, we list the fitted functional dependencies of probability distributions to the right of a semi-colon. Thus $p(\tau_{\text{eff}} | \tau_{\text{eff,GP}}, z, \bar{x}_{\text{HI}})$ is a conditional probability of τ_{eff} given $\tau_{\text{eff,GP}}$, whose parameters (mean and sigma) are functions of z and $\bar{x}_{\text{HI}}$.

likelihood (also called simulation-based inference) are becoming increasingly popular in this field (e.g. Zhao, Mao, & Wandelt 2022; Prelogović & Mesinger 2023; Davies et al. 2024; Greig et al. 2024,?) as most EoR datasets do not have an analytically-tractable likelihood; and common assumptions of Gaussian pseudo-likelihoods can result in biased posteriors (see Prelogović & Mesinger 2023).

We obtain the final forest likelihood by multiplying the implicit likelihoods over all XQR-30+ quasars, i , and over all redshift bins used in the analysis. Specifically, we take $\mathcal{L}_{\text{forest}} = \prod_{z=5.3}^{z=6.1} \prod_i p(\tau_{\text{eff}} = \tau_{\text{eff}}^i; z)$. We do not include data at $z \leq 5.2$ in order to make our likelihood more sensitive to the EoR (see e.g. Bosman et al. 2022 who showed that the EoR ends sometime before $z \sim 5.2$).

Note that this procedure does not account for higher order correlations in the mapping from $\tau_{\text{eff, GP}}$ to τ_{eff} . Moreover, it assumes that each $\Delta z = 0.1$ (~ 40 cMpc) segment is an independent sample of $p(\tau_{\text{eff}}; z)$; i.e. we ignore the covariance between the $\Delta z = 0.1$ segments extracted from a single quasar spectrum. We expect the covariance on such large scales to have only a minor impact on the total likelihood. Nevertheless, we plan on relaxing this approximation in future work in which we will use simulation-based inference for the total likelihood, accounting for large scale correlations in both the effective optical depths and reconstruction errors.

3.5 Combining with complementary observations

We also account for complementary, independent data when performing inference. Specifically, we compute additional likelihood terms for: (i) the galaxy non-ionising UV LFs well-established by *Hubble* at $6 \leq z \leq 10$ (Bouwens et al. 2015b, 2016; Oesch et al. 2018); and (ii) the CMB polarisation power spectra observed by *Planck* (Planck Collaboration et al. 2020). These two datasets are independent and mature, and can therefore be interpreted robustly. Unlike Q21, we do not include a likelihood term for the pixel Dark Fraction (Mesinger 2010; McGreer, Mesinger, & D’Odorico 2015) as this statistic is also based on Lyman forests and therefore is technically not fully independent from the optical depth distributions discussed above. Thus our total likelihood consists of the product of three terms: $\mathcal{L}_{\text{tot}} = \mathcal{L}_{\text{forest}} \times \mathcal{L}_{\text{LF}} \times \mathcal{L}_{\text{CMB}}$, where the final two correspond to the LF and CMB likelihoods, discussed further below.

We construct the UV LF likelihood following Park et al. (2019). Specifically, we assume a Gaussian likelihood in each magnitude bin, $M_{\text{UV},i}$, with a negligible covariance between bins (see e.g. Leethochawalit et al. 2023 for an alternative approach). The UV LF likelihood is thus $\mathcal{L}_{\text{LF,tot}} = \prod_{z=6}^{z=10} \prod_i \exp \left\{ - \left[\Delta\phi(M_{\text{UV}}, i) / \sigma_{\phi}(M_{\text{UV}}, i) \right]^2 \right\}$, where $\Delta\phi$ is the difference between forward-modelled and observed galaxy number densities in a given magnitude and redshift bin, and the corresponding observational uncertainties are σ_{ϕ} . As dust is thought to significantly suppress the UV LFs in the bright end (Mason, Trenti, & Treu 2023; Qin, Balu, & Wyithe 2023), we only consider magnitudes fainter than $M_{\text{UV}} = -20$ to avoid modelling dust attenuation, and use the redshift range between $z = 6$ and 10 spanning the EoR.

We construct the *Planck* CMB likelihood as a two-sided Gaussian on the Thomson scattering optical depth summary statistic inferred by Qin et al. (2020): $\tau_e = 0.0569^{+0.0073}_{-0.0066}$. Specifically, we take the form $\mathcal{L}_{\text{CMB}} = \exp \left[- (\Delta\tau_e / \sigma_{\tau_e})^2 \right]$, where

$\Delta\tau_e$ represents the difference between the forward-modelled and measured optical depths while σ_{τ_e} is the observational uncertainty. Note that Qin et al. (2020) found very little difference in the inferred posteriors when using a likelihood defined directly on the E-mode polarisation power spectra compared to using a Gaussian likelihood on the τ_e summary derived from the power spectra. We thus use the latter as it is much more computationally efficient.

3.6 Summary of model parameters and associated priors

Before showing our inference results, we summarise the free parameters used in our galaxy models and their associated prior ranges (see also Table 1).

1. $\log_{10} f_{*,10} \in [-2, -0.5]$: the fraction of galactic baryons in stars, normalised at $M_{\text{vir}} = 10^{10} M_{\odot}$. This parameter sets the normalisation of the stellar-to-halo mass relation, and its prior range is motivated by observations and simulations of high-redshift galaxies (Dayal et al. 2014; Mutch et al. 2016; Behroozi et al. 2019; Bird et al. 2022; Stefanon et al. 2021).
2. $\alpha_* \in [0, 1.0]$: the power law index relating the stellar fraction to the halo mass. This parameter determines the slope of the stellar-to-halo mass relation. Observations of the faint end of the UV LFs suggest more efficient star formation in more massive galaxies (Bouwens et al. 2015b; Oesch et al. 2018), motivating our prior range.
3. $\log_{10} f_{\text{esc},10} \in [-3, 0]$: the amplitude of the power-law relating the UV ionising escape fraction to halo mass, normalised at $M_{\text{vir}} = 10^{10} M_{\odot}$. The wide prior reflects the large uncertainties in both low-redshift observations (Vanzella et al. 2010, 2016; Boutsia et al. 2011; Nestor et al. 2013; Guaita et al. 2016; Grazian et al. 2016; Shapley et al. 2016; Bian et al. 2017; Steidel et al. 2018; Naidu et al. 2018; Fletcher et al. 2019; Izotov et al. 2021; Pahl et al. 2021) and reionisation simulations (Kostyuk et al. 2023; Choustikov et al. 2024; Mutch et al. 2024).
4. $\alpha_{\text{esc}} \in [-1, 0.5]$: the power law slope of the UV ionising escape fraction to halo mass relation. Galaxy simulations seem to suggest boosted Lyman continuum leakage in less massive galaxies as supernovae evacuate low column density channels from shallow gravitational potentials (Paardekooper, Khochfar, & Dalla Vecchia 2015; Xu et al. 2016; Kostyuk et al. 2023; Mutch et al. 2024). This motivates a wider negative range in the prior, although we caution that this is highly uncertain and therefore still allow positive values in our prior (e.g. Ma et al. 2015; Naidu et al. 2020; Rosdahl et al. 2022; Bhagwat et al. 2024).
5. $\beta_{\text{esc}} \in [-3, 3]$: the power law scaling index of the UV ionising escape fraction as a function of redshift, used only in $\text{Evolving_}f_{\text{esc}}$. The prior is somewhat arbitrary with the upper and lower limits allowing f_{esc} to scale similarly.^f

^fBecause of the wide prior, we save computational time by initially performing a fast, approximate likelihood estimate, which shows that the posterior peaks at negative values of β_{esc} . We then perform our fiducial inference on a narrower prior range $\beta_{\text{esc}} \in [-3.0, 0]$ to save computational overheads, but scale the Bayesian evidence to account for the missing prior volume (see e.g. Murray et al. 2022).

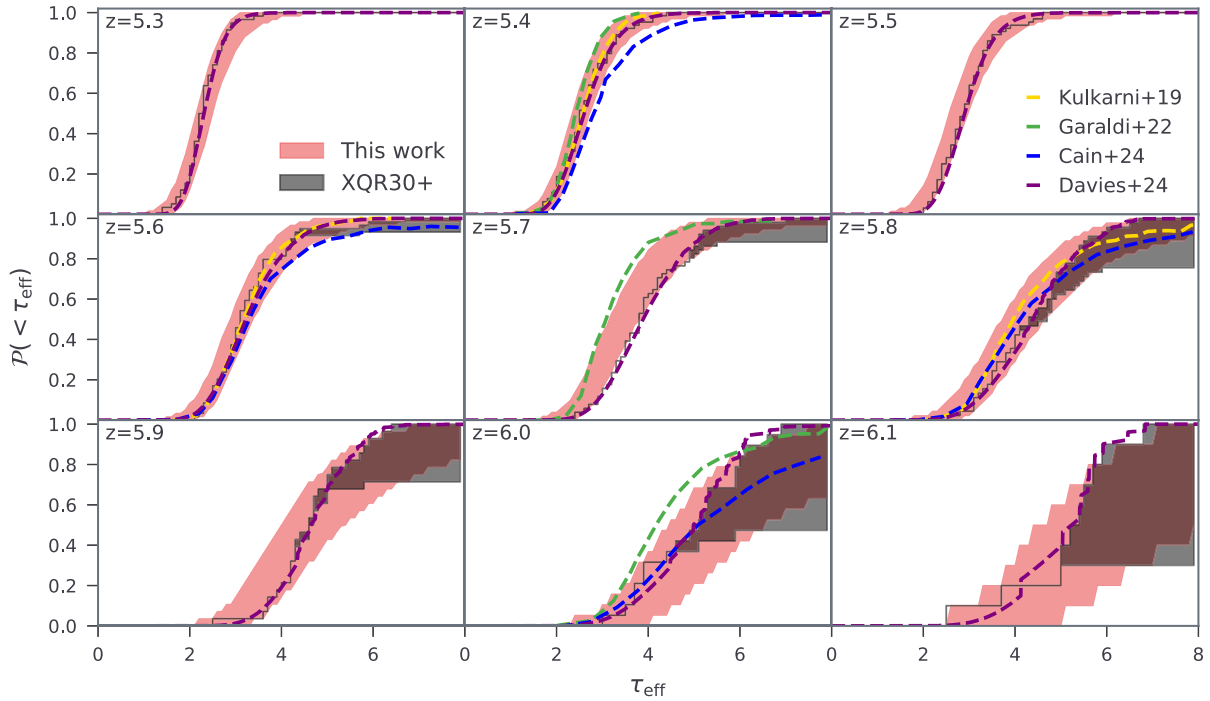


Figure 3. Inferred τ_{eff} CDFs from our fiducial model from $z=5.3$ to 6.1 . To account for cosmic variance, we randomly select from each model in the posterior the same number of sightlines as in the XQR-30+ observational dataset. The red regions indicate the 95% C.I. For comparison, the XQR-30+ observations are shown in grey with non-detections denoted with the shaded regions spanning the flux range between zero and double the noise (Bosman et al. 2022). A number of theoretical results are shown for comparison (Kulkarni et al. 2019; Garaldi et al. 2022; Cain et al. 2024; Davies et al. 2024, optimistic; with earlier works using slightly different binning for τ_{eff}).

6. $\tau_* \in (0, 1]$: the star formation timescale in units of the Hubble time. The flat prior encompasses extreme cases where the entire stellar mass is formed in an instantaneous burst event or gradually built over the age of the universe.
7. $\log_{10}(M_{\text{turn}}/M_{\odot}) \in [8, 10]$: the characteristic halo mass below which star formation becomes exponentially suppressed. The lower and upper limits of the flat prior are motivated by the atomic cooling threshold and the faintest, currently observed high-redshift galaxies (e.g. Bouwens et al. 2016, 2015b; Oesch et al. 2018), respectively.

4. Fiducial inference results

As can be seen from Table 1, the Bayesian evidence ratios suggest that the data have a *very strong* preference (Jeffreys 1939) for the Evolving- f_{esc} model. We therefore treat this model as ‘fiducial’, presenting its posterior in this section, before comparing it to the Constant- f_{esc} model in the following section.[§] Alternatively, one could do Bayesian model averaging to combine the derived IGM and galaxy properties from different models; however the evidence ratio in this case is so strongly skewed towards Evolving- f_{esc} , that the model-averaged posteriors would just follow the Evolving- f_{esc} ones.

[§]Note that having a much higher Bayesian evidence does not necessarily mean that the model is ‘the correct’ one. Compared to the alternate model, the fiducial one is more flexible and predictive given the observed data, without wasted prior volume.

We show the posteriors in the space of galaxy parameters in Appendix A. Here we focus on the inferred IGM properties and galaxy scaling relations.

4.1 Effective optical depth distributions

In Fig. 3, we plot the recovered optical depth CDFs in red enclosing the 95% confidence interval (C.I.). Observational data from Bosman et al. (2022) are shown in grey. The red-shaded regions are constructed from the posterior samples, each having the same number of randomly selected sightlines per redshift bin as in the data to account for cosmic variance. We note that the cosmic variance dominates the widths of the CDFs, especially at the highest redshifts.

We see from the figure that our fiducial model excels at recovering the observed τ_{eff} CDFs throughout this redshift range – despite individual τ_{eff} data being used in the likelihood, our model can recover both the mean *and* the shape of the observed optical depth distribution. We stress that most previous work either used hyperparameters to account for the mean opacity evolution (e.g., Q21), calibrated the models to have the same mean opacity as the data and/or treated each redshift bin independently (e.g. Kulkarni et al. 2019; Meiksin 2020; Cain et al. 2024; Gaikwad et al. 2023;^h Davies et al. 2024). Some more expensive coupled hydrodynamic and radiative-transfer simulations such as CODA and THESAN (Ocvirk et al. 2021; Garaldi et al. 2022) do not directly tune their

^hIn order to constrain the MFP and UVB using Kolmogorov–Smirnov test statistics, Gaikwad et al. (2023) treated non-detections in slightly different ways when calculating the CDFs from data and their model.

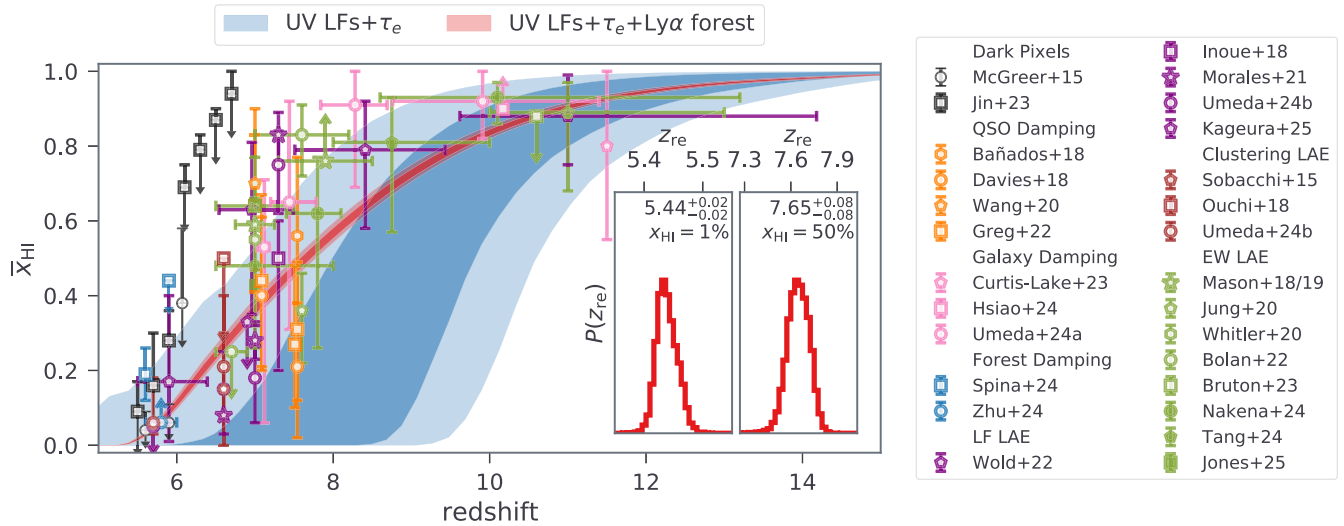


Figure 4. The inferred EoR history using our fiducial model. The blue-shaded region uses only UV LFs and CMB τ_e data (a likelihood of $\mathcal{L}_{\text{LF}} \times \mathcal{L}_{\text{CMB}}$), while the red additionally includes the Ly α forest τ_{eff} distributions (likelihood of $\mathcal{L}_{\text{forest}} \times \mathcal{L}_{\text{LF}} \times \mathcal{L}_{\text{CMB}}$). In both cases, the dark (light) regions indicate the 68% and 95% C.I. The XQR-30+ forest data are very constraining; including them makes the posterior transition from being prior-dominated to being likelihood-dominated. PDFs of the redshifts corresponding to $\bar{x}_{\text{HI}} = 0.01$ and 0.5 are presented in the inset panels, showing that in our fiducial model reionisation ends at $z = 5.44 \pm 0.02$ and the EoR mid-point is at $z = 7.7 \pm 0.1$. Estimates of the ionisation state of the universe coming from other probes are also shown for illustrative purposes including the dark pixel upper limits (McGreer, Mesinger, & D’Oroico 2015; Jin et al. 2023), Lyman- α damping-wing absorption in QSOs (Eduardo Bañados et al. 2018; Davies et al. 2018; Wang et al. 2020; Greig et al. 2022; see also Mesinger & Haiman 2004; Greig et al. 2017; Greig, Mesinger, & Bañados 2019), in galaxies (Curtis-Lake et al. 2023; Hsiao et al. 2024; Umeda et al. 2024), or in forests (Spina et al. 2024; Zhu et al. 2024), Lyman- α equivalent widths (Mason et al. 2019; Jung et al. 2020; Whitler et al. 2020; Bolan et al. 2022; Bruton et al. 2023; Nakane et al. 2024; Tang et al. 2024; Jones et al. 2025; see also Mesinger et al. 2015), and the LF (Inoue et al. 2018; Morales et al. 2021; Wold et al. 2022; Umeda et al. 2024; Kageura et al. 2025) or clustering of Ly α emitters (Sobacchi & Mesinger 2015; Ouchi et al. 2018; Umeda et al. 2024), most of which are consistent with our results despite not being included in the inference.

simulations to reproduce the forest data; however their predicted CDFs do not agree with the data as well as most of the other previously mentioned works. For illustration, we show some of these results in Fig. 3.

4.2 EoR history

In Fig. 4, we show the main result of this work – the inferred reionisation history in our fiducial model. In blue we show the posterior resulting from using only the $\mathcal{L}_{\text{LF}}$ and $\mathcal{L}_{\text{CMB}}$ likelihood terms. This roughly corresponds to our previous state of knowledge, without using the forest data.¹ From the blue region we see that the CMB optical depth and the UV LFs do not result in tight constraints on the EoR history. The UV LFs loosely constrain the evolution of the star formation rate density (SFRD), while the CMB optical depth additionally constrains the corresponding ionising escape fraction (see the parameter posterior shown in blue in Fig. A1). Given that these constraints are not tight, the posterior is prior dominated (as opposed to being likelihood dominated). Since we chose broad priors, allowing the ionising escape fraction to extend to unity, most of the posterior volume traces a relatively early reionisation, with midpoints around $z = 8$ –9.

The red shaded region in the figure shows what happens to the posterior when we further include the Ly α forest data, i.e. with the

total likelihood of $\mathcal{L}_{\text{forest}} \times \mathcal{L}_{\text{LF}} \times \mathcal{L}_{\text{CMB}}$. The EoR history, $\bar{x}_{\text{HI}}(z)$, of the maximum-a-posteriori (MAP) model is listed in Table A1, and is well fit by a rational function

$$f(z) = \frac{m_0 + m_1 z + m_2 z^2 + m_3 z^3}{n_0 + n_1 z + n_2 z^2 + n_3 z^3}, \quad (10)$$

with parameters $\{m_0, m_1, m_2, m_3, n_0, n_1, n_2, n_3\} = \{292.6, -105.47, 7.824, 0.312, -24.3, 22.9, -4.96, 0.694\}$. It is obvious that the τ_{eff} data are extremely constraining, resulting in a very narrow posterior. The uncertainties are over an order of magnitude smaller than without the forest data, with most of the history constrained to better than $\Delta z \sim 0.1$ at the 68% C.I. The forest data *require* the EoR to be ongoing below $z \leq 6$ (see also the previous results in Choudhury, Paranjape, & Bosman 2021 and Q21). From the inset panels in the figure, we see that in this fiducial model reionisation ends at $z = 5.44 \pm 0.02$ and the EoR mid-point is at $z = 7.7 \pm 0.1$. Consequently, the inferred CMB optical depth is also tightly constrained with $\tau_e = 0.0589 \pm 0.001$ (1σ) compared to $\tau_e = 0.0571 \pm 0.006$ when the forest is not included.

Perhaps surprisingly, the forest data tightly constrain the EoR history at redshifts beyond where we have forest data, $z > 6.3$. These constraints are *indirect*, coming from the combination of HMF evolution, the SFR to halo mass implied by UV LF observations, and the ionising escape fraction scalings required to match the forest. The forest in particular provides a firm anchor for our models. The forest data requires a photon-starved end to reionisation, with recombinations starting to balance ionisations, in order to smoothly transition into the post EoR regime (Bolton & Haehnelt 2007; Sobacchi & Mesinger 2014). Such a ‘soft-landing’ is difficult to achieve with small-box EoR simulations (e.g., Barkana & Loeb 2004) and/or with those that

¹The literature has many additional estimates of the EoR history that we do not include in our inference (see for instance data points shown in Fig. 4). As mentioned above, interpreting these observations is very challenging and prone to observational and modelling systematics. Robust interpretation would require dedicated forward-models of each observation and associated systematics. In any case, these alternate probes only weakly constrain the EoR history using current data (Mesinger & Haiman 2004; Greig et al. 2022; Bruton et al. 2023; Ouchi et al. 2018; Reichardt et al. 2021). We therefore expect our results to not be impacted by the inclusion of additional datasets.

cannot resolve recombinations in the late EoR stages (c.f. Fig. 6 in Sobacchi & Mesinger 2014, and Qin et al. 2021; Cain et al. 2024). Such limitations tend to result in an overly rapid evolution of the late EoR stages, which in turn requires ad-hoc corrections/tuning (e.g., a very rapid drop in emissivity) in order to match forest data (see Fig. 7 and associated discussion). The fact that our box sizes are 250 Mpc and that sub-grid recombinations are computed analytically (and thus not limited by resolution), likely allows us to capture this ‘soft-landing’ preferred by the forest data. We caution however that these constraints on the EoR history at $z > 6.3$ are indirect, and as such become increasingly model-dependent at increasingly higher redshifts. We will revisit this in the future using alternate galaxy models that evaluate star-forming duty cycles based on cooling efficiencies and feedback, and include an additional population of early, molecular-cooling galaxies, which might dominate the ionising background at $z > 10$ –15 (e.g., Qin et al. 2020; Muñoz et al. 2022; Ventura et al. 2024).

For illustration, we also plot various independent estimates of the IGM neutral fraction using other probes:

1. Our results are consistent with upper limits from the dark pixel fraction, applied to smaller forest samples (McGreer, Mesinger, & D’Odorico 2015; Jin et al. 2023). These constraints are the most model-independent and probe large volumes of the high-redshift IGM.
2. A significant portion of these probes relies on IGM damping-wing absorption observed in high-redshift spectra. These measurements depend heavily on the accurate modeling of the intrinsic $\text{Ly}\alpha$ profile, particularly its red side:
 - (a) Observations of high-redshift QSOs, including DESJ0252-0503 at $z = 7$ (Wang et al. 2020), ULASJ1120+0641 at $z = 7.09$ (Mortlock et al. 2011), J1007+2115 at $z = 7.51$ (Yang et al. 2020), and ULASJ1342+0928 at $z = 7.54$ (Eduardo Bañados et al. 2018) are generally consistent with our posterior distribution (e.g., Eduardo Bañados et al. 2018; Davies et al. 2018; Greig et al. 2022), lending confidence that these analyses can be reasonably trusted despite the associated systematics and modelling challenges (see also Mesinger & Haiman 2004; Hennawi et al. 2024; Kist, Hennawi, & Davies 2024).
 - (b) Deep Spectroscopic observations from JWST have extended damping-wing analysis to higher redshifts using bright galaxies such as JADES-GS-z11 at $z = 11.5$ (Curtis-Lake et al. 2023) and MACS0647-JD at $z = 10.2$ (Hsiao et al. 2024) as well as stacked spectra spanning $z = 7$ –12 (Umeda et al. 2024). These results are broadly consistent with our inferred EoR history, although high-redshift measurements (e.g., MACS0647-JD) suggest a more neutral early universe. However, we caution the observed sample at these high redshifts are extremely limited and acknowledge our conclusions regarding the early stages of the EoR are also more model-dependent.
 - (c) Damping-wing absorption adjacent to Gunn-Peterson troughs can also be measured using stacked QSO spectra at lower redshifts. Recent results from Spina et al. (2024) and Zhu et al. (2024) reveal

the presence of neutral regions at the end of EoR. Their inferred neutral fractions align with our posterior distribution, except at $z = 5.6$, where discrepancies may arise from the simplistic smoothed step function used to model local neutral fractions near Gunn-Peterson troughs (see more in Spina et al. 2024).

3. Probes involving $\text{Ly}\alpha$ emission from galaxies come with significant uncertainties due to poor constraints on the $\text{Ly}\alpha$ emerging into the IGM.
 - (a) Most inferred neutral fractions from $\text{Ly}\alpha$ LF studies are consistent with our results (e.g., Inoue et al. 2018; Wold et al. 2022; Kageura et al. 2025). However, Morales et al. (2021) report a rapid evolution in the reionisation history, finding a lower x_{HI} at $z = 6.6$ and a higher x_{HI} at $z = 7.3$ compared to our results. Their assumption of a fully ionised universe at $z = 6$ may lead to an underestimation of the neutral fraction at all redshifts, aligning their $z = 6.6$ estimate more closely to ours while elevating the $z = 7.3$ neutral fraction to > 0.83 . Similarly, recent work by Umeda et al. (2024) also reports a high neutral fraction at $z = 7.3$ with $x_{\text{HI}} \sim 0.75$. Given that the survey areas ($\sim 1 \text{ deg}^2$) in these studies suggest minimal cosmic variances, the discrepancy may point to an evolution in the intrinsic $\text{Ly}\alpha$ profile between these redshifts.
 - (b) The redshift evolution of LAE clustering offers smaller uncertainties in intrinsic profile modeling compared to the LF. Current clustering data at $z = 6.6$, enabled by the Subaru narrow-band filter at 921 nm, are consistent with our inferred $x_{\text{HI}} \sim 0.25$ (Sobacchi & Mesinger 2015; Ouchi et al. 2018; Umeda et al. 2024).
 - (c) JWST has significantly expanded direct measurements of $\text{Ly}\alpha$ equivalent width of EoR galaxies, with the inferred neutral fractions largely aligning with our results (Bruton et al. 2023; Nakane et al. 2024; Tang et al. 2024; Jones et al. 2025). Similarly, earlier studies using ground-based telescopes are also mostly consistent with our inferred EoR history. Notable exceptions include Mason et al. (2019, $x_{\text{HI}} > 0.76$ at $z \sim 8$) and Bolan et al. (2022, $x_{\text{HI}} \sim 0.83$ at $z \sim 8$), which suggest a more neutral universe and a rapid reionisation timeline compared to our results. These discrepancies likely arise from a combination of small sample sizes and significant uncertainties in modeling the intrinsic $\text{Ly}\alpha$ profiles, highlighting the need for larger, more robust datasets to constrain the high-redshift neutral fraction more accurately.

In summary, our EoR posterior is qualitatively consistent with most of these estimates, despite not including them in our likelihood.

4.3 UVB and MFP evolution

Fig. 5 shows the inferred redshift evolution of the photo-ionisation rate and MFP in our fiducial model. The forest data are able to constrain these global IGM quantities at percent level precision. The total MFP converges to our assumed uniform value for the ionised IGM post EoR at $z \lesssim 5.2$ (i.e. $R_{\text{MFP,LLS}}$). Neutral patches during the

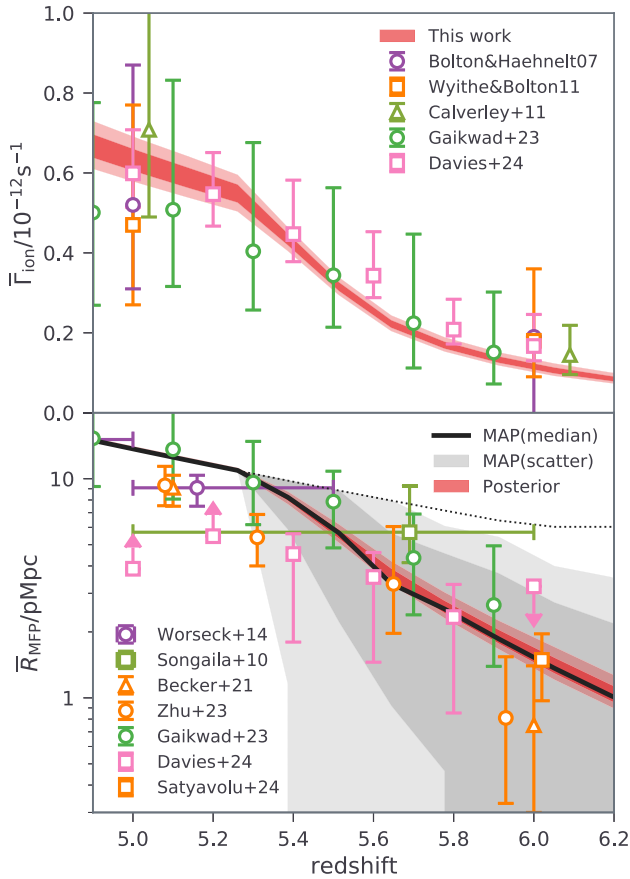


Figure 5. The posterior of our fiducial model in the space of the mean photo-ionisation rate (top panel) and proper mean free path (bottom panel). As in the previous figure, the dark (light) shaded region corresponds to 68% (95%) C.I. In the lower panel, we additionally show the volume distribution of the MFPs from the MAP model (median and scatters). The dotted line indicates the assumed $R_{\text{MFP,LLS}}$. Various previous estimates from the forests (Bolton & Haehnelt 2007; Wyithe & Bolton 2011; Calverley et al. 2011; Worseck et al. 2014; Songaila & Cowie 2010; Becker et al. 2021; Gaikwad et al. 2023; Zhu et al. 2023; Davies et al. 2024; Satyavolu et al. 2024) are also shown with their 68% error bars. Our results are in general agreement with these independent estimates, despite not having used them in the inference.

EoR contribute increasingly to the MFP at earlier stages, as discussed in Section 3.2 (see also Roth et al. 2024). This results in a more rapid drop in the MFP from $z \sim 5$ to 6 than would be expected in simple, uniform-UVB, post-EoR models (e.g., Becker et al. 2021).

In the figure we also show several independent estimates from the literature. These come from: (i) adjusting simulated Ly α optical depths to match the observed flux evolution (e.g., Bolton & Haehnelt 2007); (ii) estimating the column-density evolution of HI absorbers (Songaila & Cowie 2010); (iii) modelling the size evolution of quasar near zones (Wyithe & Bolton 2011); (iv) modelling flux profiles around near zones (Calverley et al. 2011; Worseck et al. 2014; Becker et al. 2021; Zhu et al. 2023; Satyavolu et al. 2024); and (v) co-varying the MFP and UVB to match forest fluctuations independently at each redshift (Davies et al. 2024; Gaikwad et al. 2023). Our results are generally in good agreement with these independent estimates, despite the fact that we do not use them in our analysis. Our recovered MFP at $z \sim 6$ is on the upper end of the 68% error bars from Becker et al. (2021), Zhu et al. (2023). This mild tension could point to additional systematics in these

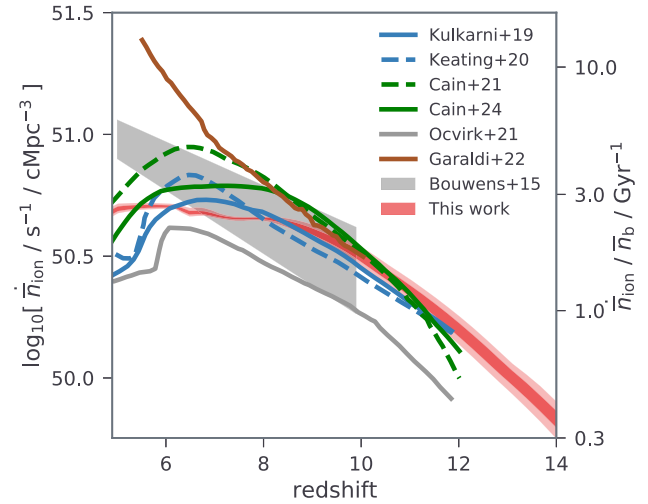


Figure 6. The inferred UV ionising emissivity, $\bar{\eta}_{\text{ion}}$. On the left axis we denote the number of ionizing photons per time per comoving volume, while on the right axis we show the number of ionizing photons per time per baryon. As in the previous figure, the dark (light) shaded red region corresponds to 68% (95%) C.I. For comparison, we include other estimates from: (i) simulations tuned to match the forest opacity distributions (Kulkarni et al. 2019; Keating et al. 2020; Cain et al. 2021); (ii) coupled hydrodynamic and radiative-transfer simulations (Garaldi et al. 2022; Ocvirk et al. 2021); and (iii) a simple empirical relation based on assuming a constant escape fraction and SFRD extrapolated down to a fixed limiting magnitude of $M_{\text{UV}} = -13$ (Bouwens et al. 2015a).

observational interpretations (see more in Satyavolu et al. 2024) and/or missing physics in our models, such as gas relaxation (e.g., Park et al. 2016; D’Aloisio 2020) or the inhomogeneous post I-front temperature (e.g., D’Aloisio et al. 2019; Davies et al. 2019). We plan on investigating these effects in future work.

In the bottom panel of Fig. 5, we additionally present the volume distribution of the MFPs derived from our MAP model. The black curve represents the median while the grey shaded regions indicate 68% and 95% of the volume distribution. We see that 68% of the volume has an MFP determined by $R_{\text{MFP,EoR}}$ at $z \sim 5.5$, even though reionisation completes at $z = 5.44$. This finding underscores that the assumed functional form of $R_{\text{MFP,LLS}}$ likely has a minor impact on the MFP at these EoR redshifts. Nevertheless in future work we will additionally sample the uncertainties in the mean and scatter of $R_{\text{MFP,LLS}}$.

4.4 UV ionising emissivity

In Fig. 6, we present the inferred ionising emissivity evolution, $\bar{\eta}_{\text{ion}}(z)$ (c.f. equation 3). Unlike many previous studies (e.g., Kulkarni et al. 2019; Keating et al. 2020; Cain et al. 2021), we reproduce the Ly α opacity distribution without requiring a sharp drop in the emissivity at $z \lesssim 7$. Such a rapid drop in the emissivity would be difficult to reconcile with the more gradual evolution implied by observations of galaxy UV LFs (e.g., Bouwens et al. 2015a), as it requires either fast evolving feedback in faint galaxies (e.g., Ocvirk et al. 2021; though see e.g., Sobacchi & Mesinger 2014; Katz et al. 2020) or in their ionising escape fractions. Even under both such putative scenarios, it is difficult to physically justify cosmological evolution that is more rapid than characteristic time-scales during this epoch which are generally $\gtrsim 200$ Myrs (e.g., the duration of the EoR, halo dynamical and/or sound crossing times; c.f. Sobacchi & Mesinger 2013). As discussed in Section 4.2, one possible explanation is that simulating the end stages of the

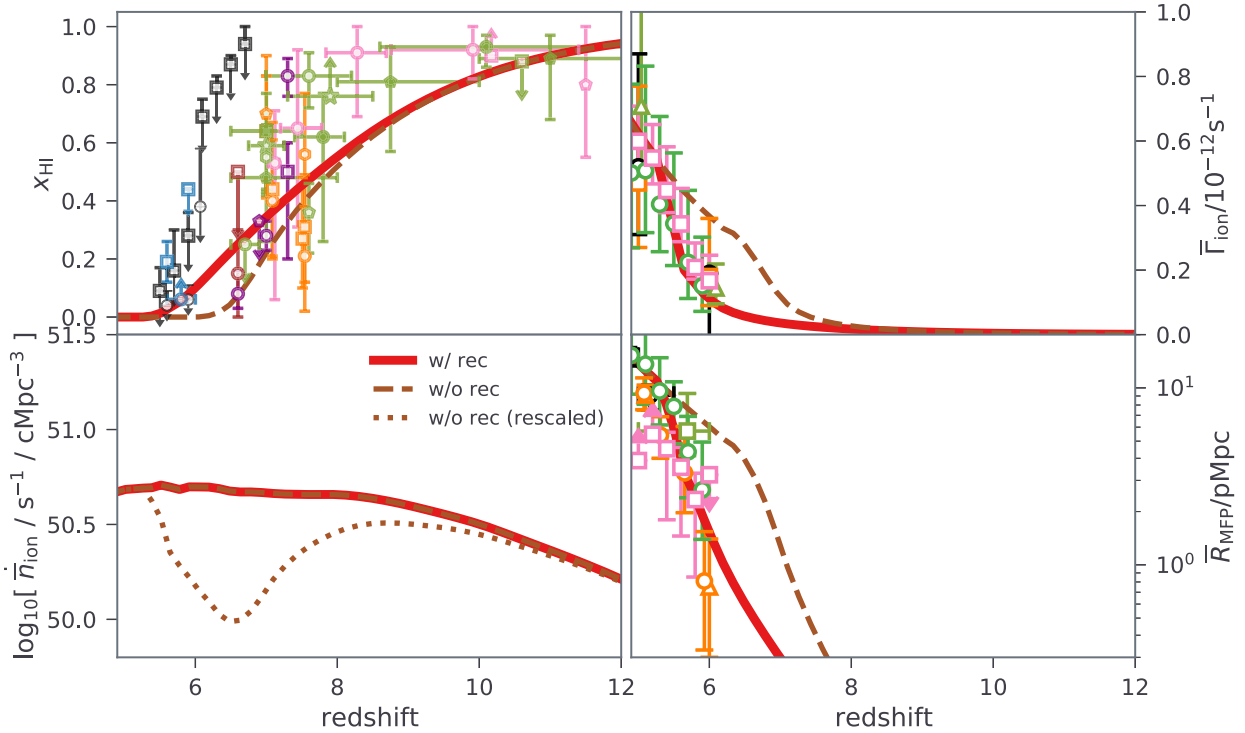


Figure 7. Comparison of the MAP model with and without recombinations. Clockwise from the upper left panel, we show the mean EoR history, photoionisation rate, proper MFP and ionizing emissivity. In the bottom left panel we also show the emissivity rescaled by the ratio of $\bar{\Gamma}_{\text{ion}}$ from *w/ rec* to that from *w/o rec*, roughly mimicking what would be required for the emissivity to compensate for the missing recombinations.

EoR requires very large-scale and very high-resolution hydrodynamic simulations to track the rapid evolution of self-shielding in the IGM and the strong spatial correlation between ionizing sinks and sources. Our calibrated sub-grid approach could allow us to capture the relevant recombination physics without requiring very high resolution (Sobacchi & Mesinger 2014).

To better quantify this claim, we rerun the MAP model, turning off inhomogeneous recombinations. Fig. 7 shows the predicted mean EoR history, photoionisation rate, MFP and ionizing emissivity. In the absence of sub-grid recombinations, the end of reionisation progresses significantly more rapidly, leading to a correspondingly sharp rise in both the MFP and photo-ionisation rate (see also Sobacchi & Mesinger 2014). Since our models fix the post-reionisation MFP to $R_{\text{MFP, LLS}}$, these quantities eventually asymptote to the same values as in the fiducial run. In the emissivity sub-panel, we also adjust the emissivity by rescaling it with the ratio of $\bar{\Gamma}_{\text{ion}}$ from *w/ rec* to that from *w/o rec*. We see that by matching the UVB (which roughly corresponds to what forest observations would require for the *w/o rec* case), the emissivity would need to decrease rapidly during the second half of the EoR, countering the premature rise in the MFP caused by the lack of recombinations. This lends further credibility to our claim that unresolved, inhomogeneous recombinations are responsible (at least in part) for the rapid drop in the emissivity required by some large-scale hydro simulations in order to match the forest data.

We note from Fig. 6 that our inferred emissivity is consistent with simple estimates assuming a constant escape fraction and integrating the observed UV LFs down to $M_{\text{UV}} = -13$. This is somewhat coincidental, since in our fiducial model more than

half of the ionizing photons are provided by galaxies fainter than $M_{\text{UV}} > -13$ due to a strong M_{UV} -dependence of the ionizing escape fraction (see later Fig. 10). As discussed further in Section 5, the forest data combined with UV LFs strongly constrain the redshift evolution of the EoR and the ionizing emissivity. However, determining which galaxies produce the ionizing photons responsible is more model dependent.

4.5 Effective clumping factor in HII regions

Modelling the complex interplay between ionizing sinks and sources during the EoR is best achieved with large-scale numerical simulations. However, simple analytic estimates of the EoR history can be very convenient and help build physical intuition. A common choice is the following (e.g., Bouwens et al. 2015a):

$$\dot{Q}_{\text{HII}} = \dot{n}_{\text{ion, H}} - Q_{\text{HII}}/t_{\text{rec, H}} \quad (11)$$

where

$$\bar{t}_{\text{rec, H}} \equiv \frac{1}{C_{\text{eff}} \bar{n}_{\text{H}} \alpha_{\text{B}}} = \left(\frac{1+z}{6} \right)^{-3} \left(\frac{C_{\text{eff}}}{3} \right)^{-1} \left(\frac{T}{10^4 \text{K}} \right)^{0.75} \text{ Gyr}. \quad (12)$$

Here, $Q_{\text{HII}} \sim 1 - \bar{x}_{\text{HI}}$ and $\dot{Q}_{\text{HII}}$ are the volume filling factors of HII regions and its growth rate while $t_{\text{rec, H}}$ is a characteristic recombination time-scale parameterised by an ‘effective’ clumping factor, C_{eff} . The first term on the right-hand side of equation (11) is the ionizing emissivity per hydrogen atom while the second term approximates the global recombination rate per hydrogen atom. This equation is especially useful in high-redshift galaxy studies, as it allows us to connect the ionizing emissivity from galaxies to the EoR history simply by assuming some value of C_{eff} to capture the

impact of recombinations. Common choices for C_{eff} range from ~ 1 –10 (e.g., Bouwens et al. 2015a; Mason et al. 2018; Davies et al. 2019; Bruton et al. 2023).

One can compute the recombination rate in a given patch of the IGM by defining $C_{\text{eff}} = \langle n_{\text{HII}}^2 \rangle / \langle n_{\text{HII}} \rangle^2$, where the averaging is performed over the ionised hydrogen (accounting for self-shielding and using *local* values of temperature and ionisation rates; e.g. Finlator et al. 2012) with n_{HII} being the local HII density. However, when estimating the *global* recombination rate to be used in equation (11) there is not an obvious way of defining C_{eff} in terms of other global IGM quantities. In particular, ionizing sources and sinks are strongly correlated on large scales. Recombinations preferentially occur in regions proximate to galaxies that were the first to reionise, which have biased, time-evolving, and spatially fluctuating properties.

Here, we investigate what choice of C_{eff} can give the same EoR history as the MAP parameter set in our fiducial model. Specifically, we assume the EoR history and emissivity of our MAP model (c.f. Figs. 4 and 6), and solve for $C_{\text{eff}}(z)$ using equation (11). The resulting effective clumping factor is plotted as a red curve in the bottom panel of Fig. 8. We see C_{eff} starts around unity¹ and then rises rapidly towards the late stages of reionisation when ionisation fronts penetrate deeper into overdensities (e.g., Furlanetto, Oh, & Briggs 2006; Finlator et al. 2012; Sobacchi & Mesinger 2014; Cain et al. 2024; Davies, Bosman, & Furlanetto 2024).

Fundamentally, C_{eff} cannot be a constant during the EoR. We illustrate EoR histories resulting from common assumptions of a constant $C_{\text{eff}} = 1, 3, 10$ in the top panel of Fig. 8. All curves assume the same emissivity as the MAP. However, no constant choice of C_{eff} can reproduce the EoR history of the MAP (red curve).

We offer a fit for $C_{\text{eff}}(z)$ using a rational function (equation 10) with coefficients $\{m_0, m_1, m_2, m_3, n_0, n_1, n_2, n_3\} = \{238.9, -94.35, 11.76, -0.404, 22.6, -3.97, -0.877, 0.1636\}$. This can be used in analytic models to approximate the EoR history resulting from a given emissivity. In future work we will quantify how sensitive this effective clumping factor is to different reionisation or emissivity models.

4.6 Galaxy UV LFs and scaling relations

In Fig. 9 we show the inferred UV LFs for our fiducial model. As in Fig. 4, the blue shaded regions correspond to our posterior without including forest data (i.e. only including UV LF data and τ_e), while the red regions additionally include the τ_{eff} distributions from XQR-30+. In the figure we also show observational estimates from both *Hubble* and *JWST*, with blue points highlighting those *Hubble* datasets that are used in the likelihood (see Section 3.5). The MAP model and [16, 84]th percentiles are also listed in Table A2.

From the figure we see that the *Hubble* estimates we use in the likelihood already constrain the inferred UV LFs at magnitudes brighter than -17 , where we have observational estimates. Overall, the predictions also remain consistent with recent *JWST* measurements (e.g., Donnan et al. 2023; Harikane et al. 2023; Willott et al. 2024), though observations at $10 < z \lesssim 13$ appear slightly higher. At $z \sim 16$, however, UV variability

¹The clumping factor can also decrease rapidly during earlier stages of reionisation as the gas relaxes from an increase in the Jeans mass after it is photo-heated (e.g., Emberson, Thomas, & Alvarez 2013; Park et al. 2016; D’Aloisio 2020). However, it is likely that X-ray preheating (e.g., HERA Collaboration et al. 2022, 2023) diminishes this evolution in practice.

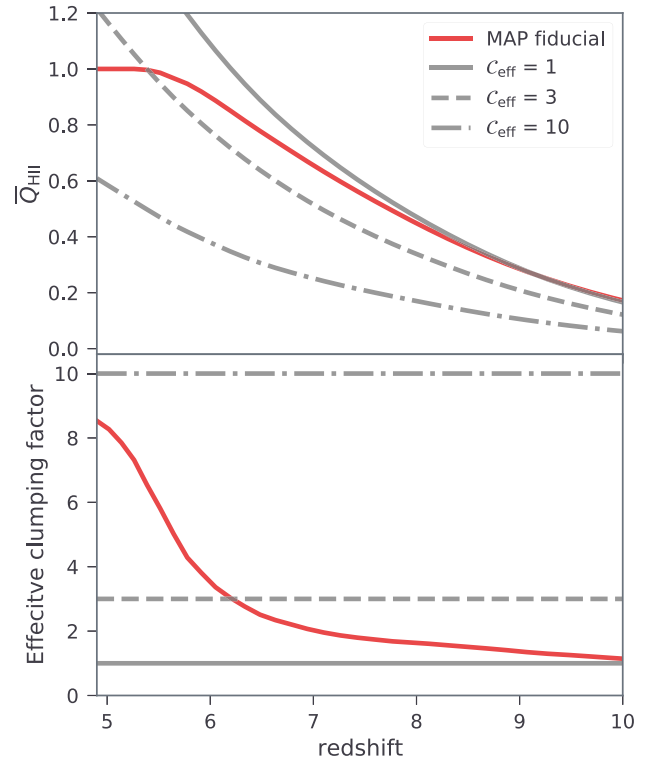


Figure 8. Top panel: Evolution of the HII filling factor from the MAP model (red curve), together with analytic estimates using equation (11) assuming the same mean ionizing emissivity as the MAP but taking a constant ‘clumping factor’. Curves corresponding to $C_{\text{eff}} = 1, 3, 10$ are shown in grey. Bottom panel: the effective clumping factor obtained by solving equation (11) for $C_{\text{eff}}(z)$ when assuming the EoR history and emissivity from the MAP model.

(Shen et al. 2023; Nikolić et al. 2024) sourced by enhanced star formation (e.g., Qin, Balu, & Wyithe 2023; Wang et al. 2023; Chakraborty & Choudhury 2024) or differences in stellar populations (e.g., Ventura et al. 2024; Yung et al. 2024) may be indeed necessary to explain the observed trends. This offset suggests that a redshift evolution in f_* (or τ_*) might also be needed in our model, similar to the adjustments made for f_{esc} (see equation 2). We will explore this further as JWST data continue to mature. On the other hand, the posteriors in blue widen greatly at fainter magnitudes, since there is no observational consensus regarding a faint-end turn-over (see also Gillet, Mesinger, & Park 2020 and Atek et al. 2024). Once we include the forest data however, the posteriors shrink significantly at the faint end of the UV LF. The forest data imply significant star formation in galaxies down to the atomic cooling limit ($M_{\text{UV}} \sim -10$).

Note that the Ly α forest is not sensitive enough to directly distinguish between different reionisation morphologies (see section 5.3 in Q21). Therefore the preference for star formation in UV faint galaxies indirectly comes from the mean EoR history shown in Fig. 4. Abundant, faint galaxies appear earlier and evolve more slowly, compared to rare, bright galaxies (e.g., Behroozi et al. 2019). Therefore they more naturally drive the kind of slower reionisation histories with a ‘soft-landing’, which is preferred by the Ly α forest. Yet, given sufficient flexibility in assigning ionizing escape fractions, one could in principle force a bright-galaxy-dominated EoR to have the same history as the one shown in Fig. 4, which is driven by faint galaxies. This, in practice, is constrained

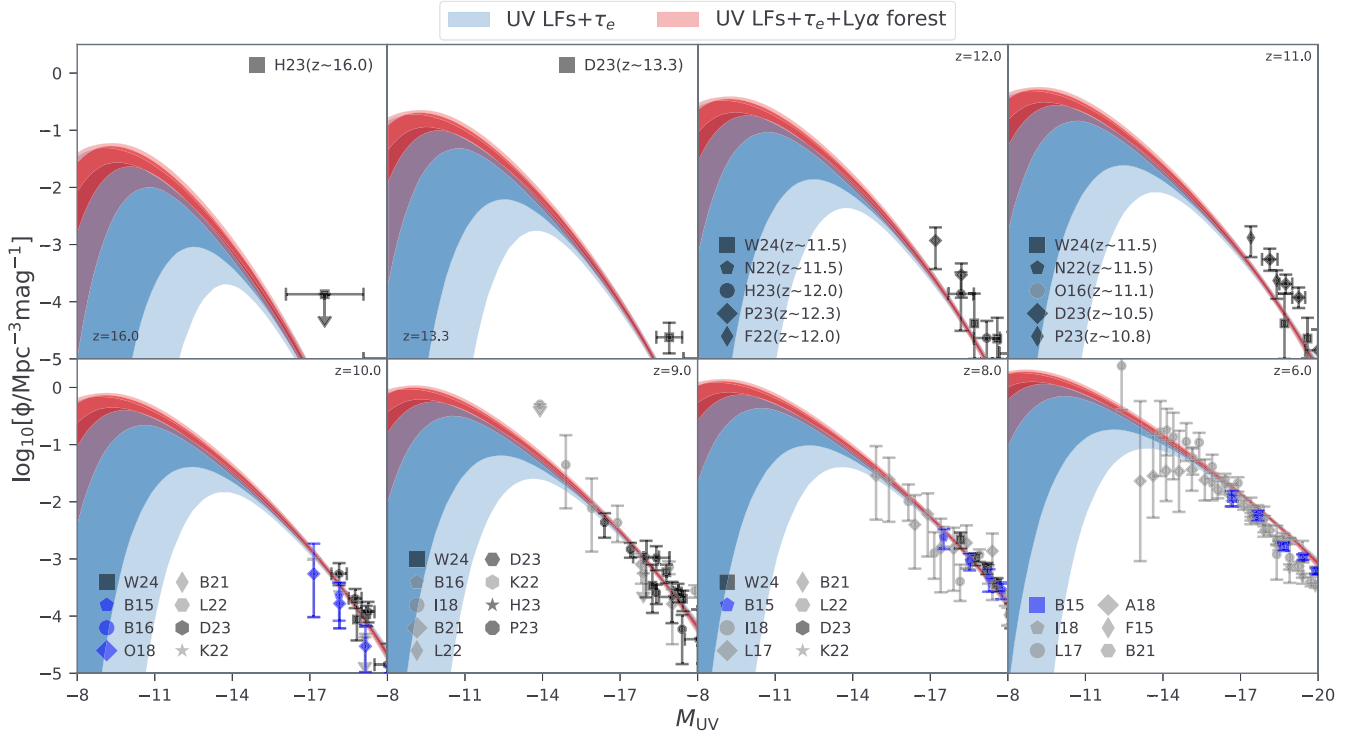


Figure 9. The inferred galaxy UV luminosity function. As in the previous figure, the dark (light) shaded region corresponds to 68% (95%) C.I. Observed luminosity functions are grouped into pre-JWST (light grey; Bouwens et al. 2016, 2015b; Finkelstein et al. 2015; Oesch et al. 2016, 2018; Livermore, Finkelstein, & Lotz 2023; Atek et al. 2018; Ishigaki et al. 2018; Bhatawdekar et al. 2019; Bouwens et al. 2016; Kauttman et al. 2022; Leethochawalit et al. 2023) with those used in the likelihood (see Section 3.5) highlighted in dark blue, and results using recent JWST observations (dark grey; Donnan et al. 2023; Finkelstein et al. 2022; Harikane et al. 2023; Naidu et al. 2022; Pérez-González et al. 2023; Willott et al. 2024).

by the fact that the escape fraction cannot exceed unity, and that bright galaxies reside in the exponential tail of the mass function. Thus very extreme evolutions in the ionizing escape fractions, exceeding unity, would be required for our model to have a ‘bright galaxy dominated EoR’ that is consistent with Ly α forest data. Such models are not in our prior volume.

We further quantify the contribution of faint galaxies to the EoR in Fig. 10. In the top panel we plot the CDF of the galaxies contributing to the ionizing background at $z = 6, 8, 11$ as functions of M_{UV} . There is a mild evolution with redshift, but in general we find that galaxies fainter than $M_{UV} \gtrsim -12$ contribute more than half of the ionizing photons that have reionised the universe. Galaxies above the current direct detection limits of $M_{UV} \lesssim -15$ only contribute a few percent to the ionizing photon budget with gravitational lensing enabling the detection of even fainter sources and extending constraints on their contribution (e.g., Vanzella et al. 2023, 2024). This highlights the power of the IGM as a democratic probe of the emissivity of all galaxies.

In the bottom panel of Fig. 10 we show the *mean* ionizing escape fraction as a function of UV magnitude, at the same three redshifts, $z = 6, 8, 11$. We see that the data prefer a population-averaged f_{esc} that increases towards faint galaxies. This is consistent with simulations including RAMSES by Kimm & Cen (2014), First Billion Years (Paardekooper, Khochfar, & DallaVecchia 2015), Renaissance (Xu et al. 2016) and Cosmic Dawn II (Lewis et al. 2020) while others such as FIRE-II (Ma et al. 2020), SPHINX (Rosdahl et al. 2022), IllustrisTNG

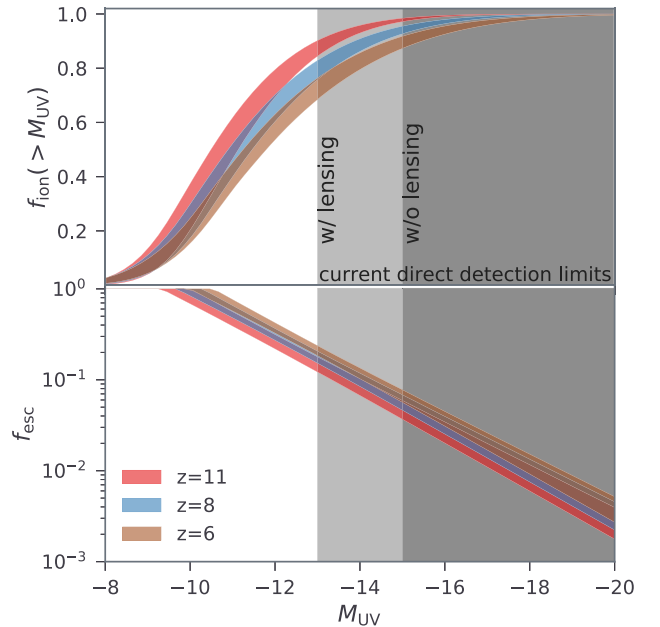


Figure 10. The inferred (68% C.I.) ionizing contribution of galaxies as a function of their UV magnitudes at $z = 6, 8$ and 11 . The top panel shows the normalised cumulative number of ionizing photons while the bottom panel shows the escape fraction. Our results imply reionisation is driven by faint galaxies, far below current direct detection limits (roughly corresponding to the grey shaded region).

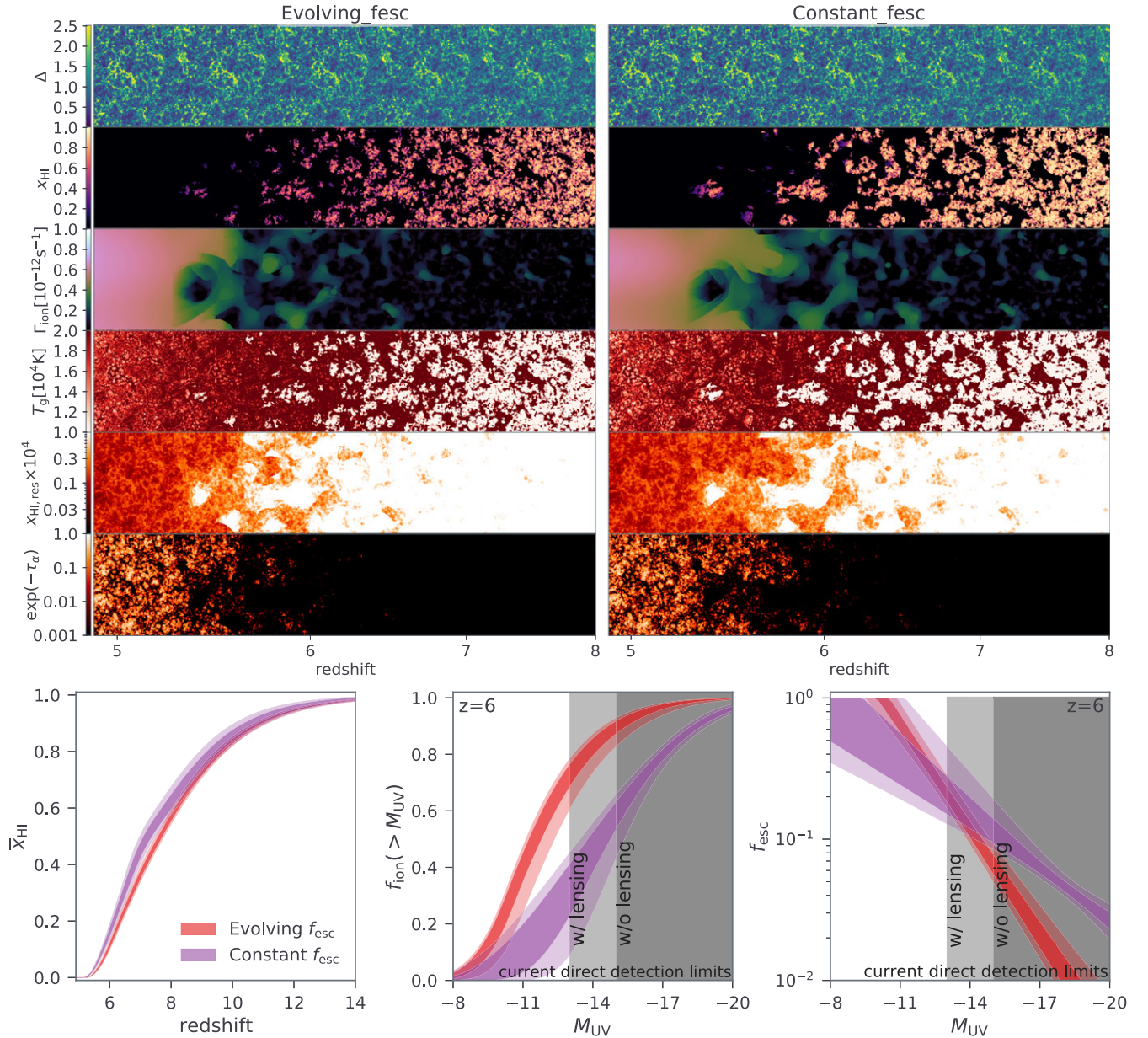


Figure 11. Upper panels: lightcones of MAP models from `Evolving_f_esc` and `Constant_f_esc`. From top to bottom, the panels correspond to the overdensity (Δ), neutral hydrogen fraction (x_{HI}), locally-averaged UVB (Γ_{ion}), temperature (T_g), residual neutral fraction within the ionised regions ($x_{\text{HI, res}}$) and Ly α transmission. Bottom panels: Similar to Figs. 4 and 10 ($z=6$ only) but for comparisons between `Evolving_f_esc` and `Constant_f_esc` and showing the 68% and 95% C.I.s of their posterior distributions. Although the two models reach qualitatively the same conclusions about the EoR, the fiducial `Evolving_f_esc` model favors an EoR that is driven by ultra-faint galaxies close to the atomic cooling threshold, resulting in a slightly more extended and patchy EoR.

(Kostyuk et al. 2023), THESAN (Yeh et al. 2023), SPICE (Bhagwat et al. 2024) show different trends in their results. As the posterior distribution of β_{esc} peaks at ~ -1.6 , the ionizing escape fraction at a given halo mass decreases at earlier times. However, as shown in this panel, such a redshift dependence becomes very mild when f_{esc} is plotted against UV magnitude. The fact that such a mild redshift evolution in the ionizing escape fraction is so strongly preferred by the Bayesian evidence (`Evolving_f_esc` vs `Constant_f_esc`) highlights again the incredible constraining power of the XQR-30+ forest data.

5. How do the results depend on our model?

In the previous section we presented constraints on IGM and galaxy properties using the `Evolving_f_esc` model, which was strongly favored by the data. Here, we explore how our main conclusions are affected by the choice of galaxy model. Specifically, we compare our fiducial model to `Constant_f_esc` where the UV ionizing escape fraction is solely dependent on the host halo mass. If our results remain largely unaffected by the choice of galaxy model, this would increase confidence in their robustness, regardless of their relative Bayesian evidences.

Fig. 11 shows lightcones corresponding to the MAP in both models (*upper panels*) and their posteriors for the EoR history (*bottom left panel*), the cumulative contribution to the $z = 6$ ionizing background of galaxies below a given UV magnitude (*bottom middle panel*), and the ionizing escape fraction as a function of UV magnitude (*bottom right panel*). Note that the inferred τ_{eff} CDFs look indistinguishable to those shown in Fig. 3. Both models suggest a qualitatively similar conclusion – reionisation finishes at $z < 5.5$ with the process primarily driven by ionizing photons emitted by faint galaxies. The end (corresponding to $\bar{x}_{\text{HI}} = 0.01$) and midpoint of reionisation are at $z = 5.33 \pm 0.03$ and $z = 7.2 \pm 0.1$ in the `Constant_f_esc` model, respectively, compared to $z = 5.44 \pm 0.02$ and $z = 7.7 \pm 0.1$ in our fiducial `Evolving_f_esc` model.

From the bottom panels of Fig. 11 we see that the models differ quantitatively in which galaxies drive reionisation. The `Constant_f_esc` model prefers the EoR to be driven by slightly brighter galaxies, with half of the ionizing photons being contributed by $M_{\text{UV}} \gtrsim -14$ galaxies (compared to $M_{\text{UV}} \gtrsim -12$ in the fiducial model). This is due to the fact that without the additional flexibility of a time-evolving f_{esc} , the `Constant_f_esc` model results in an EoR history that is too rapid compared with what the data prefer. It is a testament to the constraining power of the XQR-30+ data that only a small redshift evolution in the ionizing escape fraction (c.f. bottom panel of Fig. 10) results in a much higher Bayesian evidence for the `Evolving_f_esc` model.

6. Conclusions

The Ly α forests observed in the spectra of high-redshift quasars provide critical insight into the final stages of reionisation. In this work, we introduced a novel framework of 21cmFAST that integrates large-scale lightcones of IGM properties and incorporates unresolved sub-grid physics in the Ly α opacity, calibrated against high-resolution hydrodynamic simulations for missing physics on small scales. By sampling only 7 free parameters that are capable of characterizing the average stellar-to-halo mass relation, UV ionizing escape fraction, duty cycle and timescale of stellar buildup for high-redshift galaxies, we performed Bayesian inference against the latest Ly α forest measurement from XQR-30+ complemented by the observed high-redshift galaxy UV LFs and the CMB optical depth. We demonstrated that current data can constrain global IGM properties with percent-level precision.

One of the key outcomes of our model is the ability to reproduce the large-scale fluctuations in Ly α opacity without requiring a sharp decline in the ionizing emissivity from $z \sim 7$ to 5.5, a feature that has been invoked by several other models. In particular, our fiducial model finds reionisation occurs at $z = 5.44 \pm 0.02$ with a midpoint at $z = 7.7 \pm 0.1$. The ionizing escape fraction in this model increases towards fainter galaxies, exhibiting only a mild redshift evolution at a fixed UV magnitude. This suggests that half of the ionizing photons responsible for reionisation are sourced by galaxies fainter than $M_{\text{UV}} \sim -12$, which lie below the detection threshold of current optical and near-infrared instruments including JWST.

Additionally, we explored an alternative galaxy model that limits the redshift evolution in the ionizing escape fraction, allowing it to only vary with the host halo mass and reducing the number of free parameters to 6. Although this model demonstrates lower Bayesian evidence relative to our fiducial case, the posteriors for the evolution of IGM properties are in qualitative agreement. This lends confidence that our conclusions on the progress of the

EoR are robust. The models do differ somewhat on which galaxies were driving reionisation, with the lower evidence model suggesting galaxies fainter than $M_{\text{UV}} \sim -14$ provided half the ionizing photon budget (compared to $M_{\text{UV}} \sim -12$ for the fiducial model).

Future observations both in the Ly α forest and in direct galaxy surveys, will be crucial to further refining these models and improving our understanding of which galaxies drive reionisation as well as the early stages of the EoR (where we currently only have indirect constraints). Our Bayesian framework, allowing us to connect galaxy properties to IGM evolution in a physically-intuitive manner, represents a significant step forward, offering a versatile and efficient tool for interpreting upcoming observational data.

Acknowledgement. The authors gratefully acknowledge the HPC RIVR consortium and EuroHPC JU for funding this research by providing computing resources of the HPC system Vega at the Institute of Information Science (project EHPC-REG-2022R02-213). This work also made use of OzSTAR and Gadi in Australia. YQ acknowledges HPC resources from the ASTAC Large Programs, the RCS NCI Access scheme. YQ is supported by the ARC Discovery Early Career Researcher Award (DECRA) through fellowship #DE240101129. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project no. CE170100013. A.M. acknowledges support from the Ministry of Universities and Research (MUR) through the PRIN project ‘Optimal inference from radio images of the epoch of reionisation’, the PNRR project ‘Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing’, and the PRO3 Scuole Programme ‘DS4ASTRO’. VD acknowledges financial support from the Bando Ricerca Fondamentale INAF 2022 Large Grant ‘XQR-30’. MGH has been supported by STFC consolidated grants ST/N000927/1 and ST/S000623/1.

Data availability statement. The data underlying this article will be shared on reasonable request to the corresponding author.

References

- Alvarez, M. A., & Abel, T. 2012, *ApJ*, **747**, 126
- Asthana, S., Haehnelt, M. G., Kulkarni, G., Aubert, D., Bolton, J. S., & Keating, L. C. 2024, arXiv e-prints, arXiv:2404.06548.
- Atek, H., et al. 2024, *Natur*, **626**, 975
- Atek, H., Richard, J., Kneib, J.-P., & Schaerer, D. 2018, *MNRAS*, **479**, 5184
- Bañados, E., et al. 2016, *ApJS*, **227**, 11
- Bañados, E., et al. 2018, *Natur*, **553**, 473
- Barkana, R., & Loeb, A. 2004, *ApJ*, **609**, 474
- Becker, G. D., & Bolton, J. S. 2013, *MNRAS*, **436**, 1023
- Becker, G. D., Bolton, J. S., Madau, P., Pettini, M., Ryan-Weber, E. V., & Venemans, B. P. 2015, *MNRAS*, **447**, 3402
- Becker, G. D., D’Aloisio, A., Christenson, H. M., Zhu, Y., Worseck, G., & Bolton, J. S. 2021, *MNRAS*, **508**, 1853
- Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, *MNRAS*, **488**, 3143
- Bhagwat, A., Costa, T., Ciardi, B., Pakmor, R., & Garaldi, E. 2024, *MNRAS*, **531**, 3406
- Bhatawdekar, R., Conselice, C. J., Margalef-Bentabol, B., & Duncan, K. 2019, *MNRAS*, **486**, 3805
- Bian, F., Fan, X., McGreer, I., Cai, Z., & Jiang, L. 2017, *ApJ*, **837**, L12
- Bird, S., Ni, Y., Di Matteo, T., Croft, R., Feng, Y., & Chen, N. 2022, *MNRAS*, **512**, 3703
- Bolan, P., et al. 2022, *MNRAS*, **517**, 3263
- Bolton, J. S., & Haehnelt, M. G. 2007, *MNRAS*, **382**, 325
- Bolton, J. S., Puchwein, E., Sijacki, D., Haehnelt, M. G., Kim, T.-S., Meiksin, A., Regan, J. A., & Viel, M. 2017, *MNRAS*, **464**, 897
- Bosman, S. E. I., et al. 2022, *MNRAS*, **514**, 55
- Boutsia, K., et al. 2011, *ApJ*, **736**, 41

- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Caruana, J., Holwerda, B., Smit, R., & Wilkins, S. 2015a, *ApJ*, **811**, 140
- Bouwens, R. J., et al. 2015b, *ApJ*, **803**, 34
- Bouwens, R. J., et al. 2016, *ApJ*, **830**, 67
- Bouwens, R. J., P. A. Oesch, M. Stefanon, G. Illingworth, I. Labbé, N. Reddy, H. Atek, et al. 2021, *AJ*, **162**, 47
- Bruton, S., Lin, Y.-H., Scarlata, C., & Hayes, M. J. 2023, *ApJ*, **949**, L40
- Cain, C., D'Aloisio, A., Gangolli, N., & Becker, G. D. 2021, *ApJ*, **917**, L37
- Cain, C., D'Aloisio, A., Gangolli, N., & McQuinn, M. 2023, *MNRAS*, **522**, 2047
- Cain, C., D'Aloisio, A., Lopez, G., Gangolli, N., & Roth, J. T. 2024, *MNRAS*, **531**, 1951
- Cain, C., Lopez, G., D'Aloisio, A., Munoz, J. B., Jansen, R. A., Windhorst, R. A., & Gangolli, N. 2024, arXiv e-prints, arXiv:2409.02989.
- Calverley, A. P., Becker, G. D., Haehnelt, M. G., & Bolton, J. S. 2011, *MNRAS*, **412**, 2543
- Chakraborty, A., & Choudhury, T. R. 2024, *JCAP*, **2024**, 078
- Chardin, J., Kulkarni, G., & Haehnelt, M. G. 2018, *MNRAS*, **478**, 1065
- Chardin, J., Puchwein, E., & Haehnelt, M. G. 2017, *MNRAS*, **465**, 3429
- Choudhury, T. R., Paranjape, A., & Bosman, S. E. I. 2021, *MNRAS*
- Choustikov, N., et al. 2024, *MNRAS*, **529**, 3751
- Cranmer, K., Brehmer, J., & Louppe, G. 2020, *PNAS*, **117**, 30055
- Curtis-Lake, E., et al. 2023, *NatAs*, **7**, 622
- D'Aloisio, A., McQuinn, M., Davies, F. B., & Furlanetto, S. R. 2018, *MNRAS*, **473**, 560
- D'Aloisio, A., McQuinn, M., Maupin, O., Davies, F. B., Trac, H., Fuller, S., & Upton Sanderbeck, P. R. 2019, *ApJ*, **874**, 154
- D'Aloisio, A., McQuinn, M., & Trac, H. 2015, *ApJ*, **813**, L38
- D'Aloisio, A., McQuinn, M., Trac, H., Cain, C., & Mesinger, A. 2020, *ApJ*, **898**, 149
- D'Aloisio, A., Upton Sanderbeck, P. R., McQuinn, M., Trac, H., & Shapiro, P. R. 2017, *MNRAS*, **468**, 4691
- D'Odorico, V., et al. 2023, *MNRAS*, **523**, 1399
- Davies, F. B., Bosman, S. E. I., & Furlanetto, S. R. 2024, arXiv e-prints, arXiv:2406.18186.
- Davies, F. B., et al. 2024, *ApJ*, **965**, 134
- Davies, F. B., & Furlanetto, S. R. 2016, *MNRAS*, **460**, 1328
- Davies, F. B., et al. 2018, *ApJ*, **864**, 142
- Davies, J. E., Mutch, S. J., Qin, Y., Mesinger, A., Poole, G. B., & Wyithe, J. S. B. 2019, *MNRAS*, **489**, 977
- Dayal, P., Ferrara, A., Dunlop, J. S., & Pacucci, F. 2014, *MNRAS*, **445**, 2545
- Donnan, C. T., et al. 2023, *MNRAS*, **518**, 6011
- Eilers, A.-C., Davies, F. B., & Hennawi, J. F. 2018, *ApJ*, **864**, 53
- Emberson, J. D., Thomas, R. M., & Alvarez, M. A. 2013, *ApJ*, **763**, 146
- Fan, X., Narayanan, V. K., Strauss, M. A., White, R. L., Becker, R. H., Pentericci, L., & Rix, H.-W. 2002, *AJ*, **123**, 1247
- Fan, X., et al. 2006a, *AJ*, **132**, 117
- Fan, X., et al. 2006b, *AJ*, **131**, 1203
- Feron, J., Conaboy, L., Bolton, J. S., Chapman, E., Haehnelt, M. G., Keating, L. C., Kulkarni, G., & Puchwein, E. 2024, *MNRAS*, **532**, 2401
- Finkelstein, S. L., et al. 2022, *ApJ*, **940**, L55
- Finkelstein, S. L., Jr. et al. 2015, *ApJ*, **810**, 71
- Finlator, K., Peng Oh, S., Özel, F., & Davé, R. 2012, *MNRAS*, **427**, 2464
- Fletcher, T. J., Tang, M., Robertson, B. E., Nakajima, K., Ellis, R. S., Stark, D. P., & Inoue, A. 2019, *ApJ*, **878**, 87
- Flores Velázquez, J. A., et al. 2021, *MNRAS*, **501**, 4812
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, *PhR*, **433**, 181
- Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004, *ApJ*, **613**, 1–15
- Gaikwad, P., et al. 2023, *MNRAS*, **525**, 4093
- Garaldi, E., Kannan, R., Smith, A., Springel, V., Pakmor, R., Vogelsberger, M., & Hernquist, L. 2022, *MNRAS*, **512**, 4909
- Georgiev, I., Mellema, G., & Giri, S. K. 2024, *MNRAS*, **536**, 3689
- Gillet, N. J. F., Mesinger, A., & Park, J. 2020, *MNRAS*, **491**
- Grazian, A., et al. 2016, *A&A*, **585**, A48
- Greig, B., Mesinger, A., & Bañados, E. 2019, *MNRAS*, **484**, 5094
- Greig, B., Mesinger, A., Davies, F. B., Wang, F., Yang, J., & Hennawi, J. F. 2022, *MNRAS*, **512**
- Greig, B., Mesinger, A., Haiman, Z., & Simcoe, R. A. 2017, *MNRAS*, **466**, 4239
- Greig, B., Prelogović, D., Mirocha, J., Qin, Y., Ting, Y.-S., & Mesinger, A. 2024, *MNRAS*, **533**, 2502
- Greig, B., Prelogović, D., Qin, Y., Ting, Y.-S., & Mesinger, A. 2024, *MNRAS*, **533**, 2530
- Guaity, L., et al. 2016, *A&A*, **587**, A133
- Gunn, J. E., & Peterson, B. A. 1965, *ApJ*, **142**, 1633
- Haardt, F., & Madau, P. 2012, *ApJ*, **746**, 125
- Harikane, Y., et al. 2023, *ApJS*, **265**, 5
- Heintz, K. E., et al. 2024, arXiv e-prints, arXiv:2404.02211.
- Hennawi, J. F., Kist, T., Davies, F. B., & Tamanas, J. 2024, arXiv e-prints, arXiv:2406.12070.
- HERA Collaboration, et al. 2023, *ApJ*, **945**
- HERA Collaboration, et al. 2022, *ApJ*, **924**
- Hsiao, T. Y.-Y., et al. 2024, *ApJ*, **973**, 8
- Hui, L., & Gnedin, N. Y. 1997, *MNRAS*, **292**, 27
- Inoue, A. K., et al. 2018, *PASJ*, **70**
- Ishigaki, M., Kawamata, R., Ouchi, M., Oguri, M., Shimasaku, K., & Ono, Y. 2018, *ApJ*, **854**, 73
- Izotov, Y. I., Worseck, G., Schaerer, D., Guseva, N. G., Chisholm, J., Thuan, T. X., Fricke, K. J., & Verhamme, A. 2021, *MNRAS*, **503**, 1734
- Jeffreys, H. 1939, *The Theory of Probability* (Clarendon Press).
- Jiang, L., et al. 2016, *ApJ*, **833**, 222
- Jin, X., et al. 2023, *ApJ*, **942**, 59
- Jones, G. C., et al. 2025, *MNRAS*, **536**, 2355
- Jung, I., et al. 2020, *ApJ*, **904**, 144
- Kageura, Y., et al. 2025, arXiv e-prints, arXiv:2501.05834.
- Katz, H., et al. 2020, *MNRAS*, **494**, 2200
- Kauffmann, O. B., et al. 2022, *A&A*, **667**, A65
- Keating, L. C., Weinberger, L. H., Kulkarni, G., Haehnelt, M. G., Chardin, J., & Aubert, D. 2020, *MNRAS*, **491**, 1736
- Kimm, T., & Cen, R. 2014, *ApJ*, **788**, 121
- Kist, T., Hennawi, J. F., & Davies, F. B. 2024, arXiv e-prints, arXiv:2406.12071.
- Kooistra, R., Lee, K.-G., & Horowitz, B. 2022, *ApJ*, **938**, 123
- Kostyuk, I., Nelson, D., Ciardi, B., Glatzle, M., & Pillepich, A. 2023, *MNRAS*, **521**, 3077
- Kuhlen, M., & Faucher-Giguère, C.-A. 2012, *MNRAS*, **423**, 862
- Kulkarni, G., Keating, L. C., Haehnelt, M. G., Bosman, S. E. I., Puchwein, E., Chardin, J., & Aubert, D. 2019, *MNRAS*, **485**, L24
- Leethochawalit, N., Roberts-Borsani, G., Morishita, T., Trenti, M., & Treu, T. 2023, *MNRAS*, **524**, 5454
- Lewis, J. S. W., et al. 2020, *MNRAS*, **496**, 4342
- Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, *ApJ*, **835**, 113
- Ma, X., Kasen, D., Hopkins, P. F., Faucher-Giguère, C.-A., Quataert, E., Kereš, D., & Murray, N. 2015, *MNRAS*, **453**, 960
- Ma, X., Quataert, E., Wetzel, A., Hopkins, P. F., Faucher-Giguère, C.-A., & Kereš, D. 2020, *MNRAS*, **498**, 2001
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Mason, C. A., et al. 2019, *MNRAS*, **485**, 3947
- Mason, C. A., Trenti, M., & Treu, T. 2023, *MNRAS*, **521**, 497
- Mason, C. A., Treu, T., Dijkstra, M., Mesinger, A., Trenti, M., Pentericci, L., de Barros, S., & Vanzella, E. 2018, *ApJ*, **856**, 2
- McGreer, I. D., Mesinger, A., & D'Odorico, V. 2015, *MNRAS*, **447**, 499
- McQuinn, M., & Upton Sanderbeck, P. R. 2016, *MNRAS*, **456**, 47
- Meiksin, A. 2020, *MNRAS*, **491**, 4884
- Mertens, F. G., et al. 2020, *MNRAS*, **493**, 1662
- Mesinger, A. 2010, *MNRAS*, **407**, 1328
- Mesinger, A., Aykatalp, A., Vanzella, E., Pentericci, L., Ferrara, A., & Dijkstra, M. 2015, *MNRAS*, **446**, 566
- Mesinger, A., & Furlanetto, S. 2007, *ApJ*, **669**, 663
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, **411**, 955
- Mesinger, A., & Haiman, Z. 2004, *ApJ*, **611**, L69
- Miralda-Escudé, J., Haehnelt, M., & Rees, M. J. 2000, *ApJ*, **530**, 1–16
- Morales, A. M., Mason, C. A., Bruton, S., Gronke, M., Haardt, F., & Scarlata, C. 2021, *ApJ*, **919**, 120
- Mortlock, D. J., et al. 2011, *Natur*, **474**, 616
- Muñoz, J. B., Qin, Y., Mesinger, A., Murray, S. G., Greig, B., & Mason, C. 2022, *MNRAS*, **511**, 3657

- Murray, S., Greig, B., Mesinger, A., Muñoz, J., Qin, Y., Park, J., & Watkinson, C. 2020, *JOSS*, **5**, 2582
- Murray, S. G., Bowman, J. D., Sims, P. H., Mahesh, N., Rogers, A. E. E., Monsalve, R. A., Samson, T., & Konakon-dula Vydula, A. 2022, *MNRAS*, **517**, 2264
- 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
- Mutch, S. J., Greig, B., Qin, Y., Poole, G. B., & Wyithe, J. S. B. 2024, *MNRAS*, **527**, 7924
- Naidu, R. P., Forrest, B., Oesch, P. A., Tran, K.-V. H., & Holden, B. P. 2018, *MNRAS*, **478**, 791
- Naidu, R. P., et al. 2022, *ApJ*, **940**, L14
- Naidu, R. P., Tacchella, S., Mason, C. A., Bose, S., Oesch, P. A., & Conroy, C. 2020, *ApJ*, **892**, 109
- Nakane, M., et al. 2024, *ApJ*, **967**, 28
- Nasir, F., Cain, C., D'Aloisio, A., Gangolli, N., & McQuinn, M. 2021, *ApJ*, **923**, 161
- Nasir, F., & D'Aloisio, A. 2020, *MNRAS*, **494**, 3080
- Nestor, D. B., Shapley, A. E., Kornei, K. A., Steidel, C. C., & Siana, B. 2013, *ApJ*, **765**, 47
- Nikolić, I., Mesinger, A., Davies, J. E., & Prelogović, D. 2024, arXiv e-prints, arXiv:2406.15237.
- Ocvirk, P., Lewis, J. S. W., Gillet, N., Chardin, J., Aubert, D., Deparis, N., & Thélie, É. 2021, *MNRAS*, **507**, 6108
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., & Stefanon, M. 2018, *ApJ*, **855**, 105
- Oesch, P. A., et al. 2016, *ApJ*, **819**, 129
- Ouchi, M., et al. 2018, *PASJ*, **70**, S13
- Paardekooper, J.-P., Khochfar, S., & Dalla Vecchia, C. 2015, *MNRAS*, **451**, 2544
- Pahl, A. J., Shapley, A., Steidel, C. C., Chen, Y., & Reddy, N. A. 2021, *MNRAS*, **505**, 2447
- Park, H., Shapiro, P. R., Choi, J.-h., Yoshida, N., Hirano, S., & Ahn, K. 2016, *ApJ*, **831**, 86
- Park, J., Greig, B., & Mesinger, A. 2022, *MNRAS*, **517**, 192
- Park, J., Mesinger, A., Greig, B., & Gillet, N. 2019, *MNRAS*, **484**, 933
- Pentericci, L., et al. 2018, *A&A*, **619**, A147
- Pérez-González, P. G., et al. 2023, *ApJ*, **951**, L1
- PLANCK Collaboration, et al. 2016, *A&A*, **594**, A13
- PLANCK Collaboration, et al. 2020, *A&A*, **641**, A6
- Prelogović, D., & Mesinger, A. 2023, *MNRAS*, **524**, 4239
- Puchwein, E., Bolton, J. S., Haehnelt, M. G., Madau, P., Becker, G. D., & Haardt, F. 2015, *MNRAS*, **450**, 4081
- Qin, Y., Balu, S., & Wyithe, J. S. B. 2023, *MNRAS*, **526**, 1324
- Qin, Y., Mesinger, A., Bosman, S. E. I., & Viel, M. 2021, *MNRAS*, **506**, 2390
- Qin, Y., Mesinger, A., Park, J., Greig, B., & Muñoz, J. B. 2020, *MNRAS*, **495**, 123
- Qin, Y., Poulin, V., Mesinger, A., Greig, B., Murray, S., & Park, J. 2020, *MNRAS*, **499**, 550
- Rahmati, A., Pawlik, A. H., Raičević, M., & Schaye, J. 2013, *MNRAS*, **430**, 2427
- Reichardt, C. L., et al. 2021, *ApJ*, **908**, 199
- Rosdahl, J., et al. 2022, *MNRAS*, **515**, 2386
- Roth, J. T., D'Aloisio, A., Cain, C., Wilson, B., Zhu, Y., & Becker, G. D. 2024, *MNRAS*, **530**, 5209
- Satyavolu, S., Kulkarni, G., Keating, L. C., & Haehnelt, M. G. 2024, *MNRAS*, **533**, 676
- Schaye, J. 2001, *ApJ*, **559**, 507
- Scoccimarro, R. 1998, *MNRAS*, **299**, 1097
- Shapley, A. E., Steidel, C. C., Strom, A. L., Bogosavljević, M., Reddy, N. A., Siana, B., Mostardi, R. E., & Rudie, G. C. 2016, *ApJ*, **826**, L24
- Shen, X., Vogelsberger, M., Boylan-Kolchin, M., Tacchella, S., & Kannan, R. 2023, *MNRAS*, **525**, 3254
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*, **323**, 1–12
- Sobacchi, E., & Mesinger, A. 2013, *MNRAS*, **432**, 3340
- Sobacchi, E., & Mesinger, A. 2014, *MNRAS*, **440**, 1662
- Sobacchi, E., & Mesinger, A. 2015, *MNRAS*, **453**, 1843
- Songaila, A., & Cowie, L. L. 2010, *ApJ*, **721**, 1448
- Spina, B., Bosman, S. E. I., Davies, F. B., Gaikwad, P., & Zhu, Y. 2024, *A&A*, **688**, L26
- Springel, V., et al. 2005, *Natur*, **435**, 629
- Stefanon, M., Bouwens, R. J., Labbé, I., Illingworth, G. D., Gonzalez, V., & Oesch, P. A. 2021, *ApJ*, **922**, 29
- Steidel, C. C., Bogosavljević, M., Shapley, A. E., Reddy, N. A., Rudie, G. C., Pettini, M., Trainor, R. F., & Strom, A. L. 2018, *ApJ*, **869**, 123
- Tang, M., Stark, D. P., Topping, M. W., Mason, C., & Ellis, R. S. 2024, *ApJ*, **975**, 208
- Theuns, T., Leonard, A., Efstathiou, G., Pearce, F. R., & Thomas, P. A. 1998, *MNRAS*, **301**, 478
- Trott, C. M., et al. 2020, *MNRAS*, **493**, 4711
- Umeda, H., et al. 2024, arXiv e-prints, arXiv:2411.15495.
- Umeda, H., Ouchi, M., Nakajima, K., Harikane, Y., Ono, Y., Xu, Y., Isobe, Y., & Zhang, Y. 2024, *ApJ*, **971**, 124
- Vanzella, E., et al. 2016, *ApJ*, **825**, 41
- Vanzella, E., et al. 2010, *ApJ*, **725**, 1011
- Vanzella, E., et al. 2023, *A&A*, **678**, A173
- Vanzella, E., et al. 2024, *A&A*, **691**, A251
- Ventura, E. M., Qin, Y., Balu, S., & Wyithe, J. S. B. 2024, *MNRAS*, **529**, 628
- Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, *PhRvD*, **88**, 043502
- Viel, M., Lesgourgues, J., Haehnelt, M. G., Matarrese, S., & Riotto, A. 2005, *PhRvD*, **71**, 063534
- Wang, F., et al. 2020, *ApJ*, **896**, 23
- Wang, Y.-Y., Lei, L., Yuan, G.-W., & Fan, Y.-Z. 2023, *ApJ*, **954**, L48
- Weinberg, D., et al. 1999, in *Evolution of Large Scale Structure: From Recombination to Garching*, ed. A. J. Banday, R. K. Sheth, & L. N. da Costa, 346
- Whitler, L. R., Mason, C. A., Ren, K., Dijkstra, M., Mesinger, A., Pentericci, L., Trenti, M., & Treu, T. 2020, *MNRAS*, **495**, 3602
- Willott, C. J., et al. 2007, *AJ*, **134**, 2435
- Willott, C. J., et al. 2024, *ApJ*, **966**, 74
- Wold, I. G. B., et al. 2022, *ApJ*, **927**, 36
- Worseck, G., et al. 2014, *MNRAS*, **445**, 1745
- Wu, X.-B., et al. 2015, *Natur*, **518**, 512
- Wyithe, J. S. B., & Bolton, J. S. 2011, *MNRAS*, **412**, 1926
- Xu, H., Wise, J. H., Norman, M. L., Ahn, K., & O'Shea, B. W. 2016, *ApJ*, **833**, 84
- Xu, Y., Yue, B., & Chen, X. 2017, *ApJ*, **844**, 117
- Yang, J., et al. 2020, *ApJ*, **897**, L14
- Yeh, J. Y.-C., et al. 2023, *MNRAS*, **520**, 2757
- Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., Wilkins, S. M., & Gardner, J. P. 2024, *MNRAS*, **527**, 5929
- Zhao, X., Mao, Y., & Wandelt, B. D. 2022, *ApJ*, **933**, 236
- Zhu, Y., et al. 2024, *MNRAS*, **533**, L49
- Zhu, Y., et al. 2021, *ApJ*, **923**, 223
- Zhu, Y., et al. 2023, *ApJ*, **955**, 115

Appendix A. Detailed posteriors

Fig. A1 presents the marginalised 1D and 2D posterior distributions of the model parameters of various models discussed in the work, including the fiducial model $\text{Evolving_}f_{\text{esc}}$, and this model without XQR-30+ data, as well as $\text{Constant_}f_{\text{esc}}$. Tables A1 and A2 list the inferred neutral fraction, UVB, MFP and galaxy UV LFs for the MAP model and [16, 84]th percentiles of $\text{Evolving_}f_{\text{esc}}$.

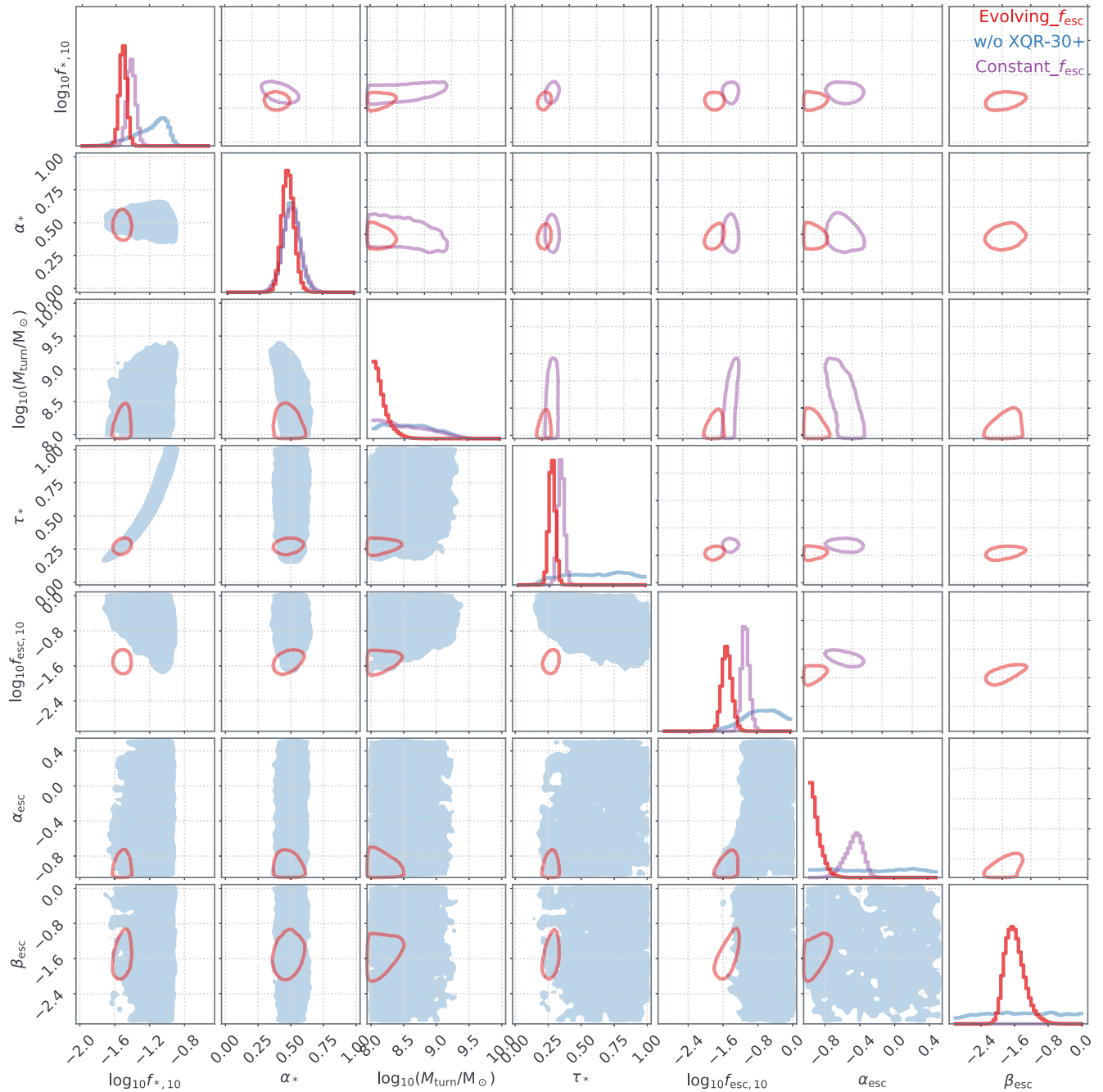


Figure A1. Marginalised 1D and 2D posterior distributions of model parameters from the fiducial model $\text{Evolving_}f_{\text{esc}}$ (red), and this model without XQR-30+ (blue) as well as $\text{Constant_}f_{\text{esc}}$ (purple). Regions inside the curves or indicated in shades represent the 95th percentiles.

Table A1. The inferred neutral fraction, photoionizing rate and MFP for the MAP model and the [16, 84]th percentiles (see also Figs. 4 and 5).

z	$\bar{x}_{\text{HI}}$			z	$\bar{x}_{\text{HI}}$			z	$\bar{\Gamma}_{\text{ion}}/10^{-12}\text{s}^{-1}$			$\bar{r}_{\text{MFP}}/\text{pMpc}$		
	MAP [†]	16th	84th		MAP [†]	16th	84th		MAP	16th	84th	MAP	16th	84th
15.20	0.99	0.99	0.99	14.88	0.99	0.99	0.99	8.68	0.01	0.01	0.01	0.04	0.04	0.05
14.57	0.99	0.99	0.99	14.26	0.98	0.98	0.99	8.49	0.01	0.01	0.01	0.05	0.05	0.06
13.96	0.98	0.98	0.98	13.67	0.98	0.98	0.98	8.30	0.01	0.01	0.02	0.06	0.06	0.08
13.38	0.97	0.97	0.98	13.10	0.97	0.97	0.97	8.12	0.02	0.02	0.02	0.08	0.07	0.09
12.82	0.96	0.96	0.97	12.55	0.96	0.96	0.96	7.94	0.02	0.02	0.02	0.09	0.09	0.11
12.29	0.95	0.95	0.96	12.03	0.94	0.94	0.95	7.77	0.02	0.02	0.02	0.12	0.11	0.14
11.77	0.93	0.93	0.94	11.52	0.92	0.92	0.93	7.60	0.02	0.02	0.03	0.14	0.13	0.17
11.28	0.91	0.91	0.92	11.04	0.90	0.90	0.91	7.43	0.03	0.02	0.03	0.17	0.16	0.20
10.80	0.89	0.89	0.90	10.57	0.87	0.87	0.89	7.26	0.03	0.03	0.03	0.21	0.20	0.25
10.34	0.86	0.86	0.87	10.12	0.84	0.84	0.85	7.10	0.03	0.03	0.04	0.26	0.25	0.30
9.90	0.82	0.82	0.84	9.69	0.80	0.80	0.82	6.94	0.04	0.04	0.04	0.32	0.30	0.37
9.48	0.77	0.77	0.79	9.27	0.75	0.75	0.77	6.78	0.04	0.04	0.05	0.40	0.38	0.45
9.07	0.72	0.72	0.75	8.87	0.70	0.70	0.72	6.63	0.05	0.05	0.06	0.50	0.47	0.57
8.68	0.67	0.67	0.69	8.49	0.64	0.64	0.66	6.48	0.06	0.06	0.07	0.62	0.59	0.71
8.30	0.61	0.61	0.63	8.12	0.57	0.58	0.60	6.34	0.07	0.07	0.08	0.79	0.74	0.89
7.94	0.54	0.54	0.57	7.77	0.51	0.51	0.54	6.19	0.08	0.08	0.09	1.02	0.97	1.15
7.60	0.47	0.48	0.50	7.43	0.44	0.44	0.47	6.05	0.10	0.10	0.11	1.37	1.29	1.53
7.26	0.40	0.41	0.44	7.10	0.37	0.37	0.40	5.91	0.13	0.12	0.14	1.86	1.77	2.04
6.94	0.33	0.34	0.36	6.78	0.29	0.30	0.33	5.78	0.16	0.16	0.18	2.55	2.44	2.78
6.63	0.26	0.26	0.29	6.48	0.22	0.23	0.25	5.64	0.20	0.21	0.23	3.32	3.46	3.89
6.34	0.19	0.19	0.21	6.19	0.15	0.15	0.17	5.51	0.32	0.30	0.33	5.65	5.32	5.86
6.05	0.11	0.12	0.14	5.91	0.08	0.08	0.10	5.39	0.43	0.42	0.45	8.23	7.94	8.47
5.78	0.05	0.05	0.07	5.64	0.03	0.03	0.04	5.26	0.54	0.53	0.57	10.91	10.88	11.06
5.51	0.01	0.01	0.02	5.39	0.00	0.00	0.01	5.14	0.58	0.56	0.61	12.12	12.12	12.32
5.26	0.00	0.00	0.00	5.14	0.00	0.00	0.00	5.02	0.62	0.60	0.65	13.46	13.46	13.68
5.02	0.00	0.00	0.00	4.90	0.00	0.00	0.00	4.90	0.66	0.64	0.70	14.96	14.95	15.19

[†]The inferred MAP $\bar{x}_{\text{HI}}(z)$ can be represented by a simple ratio of two polynomials with high accuracy ($|\Delta\bar{x}_{\text{HI}}| < 0.01$) using the following functional form: $x_{\text{HI}}(z) = (292.6 - 105.47z + 7.824z^2 + 0.312z^3)/(-24.3 + 22.9z - 4.96z^2 + 0.694z^3)$.

Table A2. The inferred galaxy UV luminosity functions for the MAP model and the [16, 84]th percentiles (see also Fig. 9).

M_{UV}	z	$\log_{10} [\phi / \text{Mpc}^{-3} \text{mag}^{-1}]$			z	$\log_{10} [\phi / \text{Mpc}^{-3} \text{mag}^{-1}]$			z	$\log_{10} [\phi / \text{Mpc}^{-3} \text{mag}^{-1}]$			z	$\log_{10} [\phi / \text{Mpc}^{-3} \text{mag}^{-1}]$		
		MAP	16th	84th		MAP	16th	84th		MAP	16th	84th		MAP	16th	84th
−20.0	16	−8.40	−8.51	−8.33	13.3	−6.54	−6.63	−6.48	12	−5.75	−5.83	−5.70	11	−5.19	−5.26	−5.14
−19.0		−7.18	−7.27	−7.15		−5.57	−5.65	−5.54		−4.89	−4.96	−4.87		−4.41	−4.47	−4.39
−18.0		−6.12	−6.22	−6.10		−4.73	−4.80	−4.71		−4.14	−4.20	−4.12		−3.73	−3.78	−3.71
−17.0		−5.20	−5.30	−5.17		−3.98	−4.06	−3.96		−3.47	−3.54	−3.46		−3.11	−3.18	−3.10
−16.0		−4.38	−4.50	−4.35		−3.31	−3.41	−3.29		−2.87	−2.96	−2.85		−2.56	−2.64	−2.55
−15.0		−3.65	−3.79	−3.62		−2.72	−2.83	−2.69		−2.33	−2.43	−2.31		−2.06	−2.15	−2.04
−14.0		−3.00	−3.16	−2.97		−2.18	−2.30	−2.15		−1.84	−1.95	−1.82		−1.61	−1.71	−1.59
−13.0		−2.43	−2.60	−2.41		−1.70	−1.84	−1.68		−1.40	−1.53	−1.38		−1.20	−1.31	−1.18
−12.0		−1.95	−2.13	−1.93		−1.29	−1.44	−1.27		−1.02	−1.16	−1.00		−0.84	−0.97	−0.82
−11.0		−1.57	−1.78	−1.56		−0.96	−1.14	−0.95		−0.71	−0.88	−0.70		−0.54	−0.70	−0.54
−10.0		−1.34	−1.63	−1.33		−0.75	−1.00	−0.75		−0.52	−0.74	−0.51		−0.35	−0.56	−0.35
−9.0	10	−1.37	−1.88	−1.29	9	−0.76	−1.19	−0.70	8	−0.50	−0.90	−0.46	6	−0.33	−0.69	−0.29
−8.0		−1.83	−2.79	−1.57		−1.13	−1.93	−0.92		−0.83	−1.56	−0.65		−0.62	−1.28	−0.45
−20.0		−4.67	−4.74	−4.63		−4.20	−4.26	−4.16		−3.78	−3.83	−3.74		−3.09	−3.13	−3.05
−19.0		−3.97	−4.02	−3.95		−3.57	−3.62	−3.55		−3.21	−3.26	−3.19		−2.64	−2.67	−2.62
−18.0		−3.35	−3.40	−3.34		−3.01	−3.05	−3.00		−2.71	−2.75	−2.69		−2.23	−2.26	−2.22
−17.0		−2.79	−2.85	−2.78		−2.50	−2.55	−2.49		−2.24	−2.29	−2.23		−1.85	−1.88	−1.84
−16.0		−2.28	−2.35	−2.27		−2.03	−2.10	−2.02		−1.82	−1.87	−1.81		−1.49	−1.53	−1.48
−15.0		−1.82	−1.90	−1.80		−1.61	−1.68	−1.59		−1.42	−1.49	−1.41		−1.15	−1.21	−1.14
−14.0		−1.40	−1.49	−1.38		−1.21	−1.30	−1.20		−1.06	−1.14	−1.04		−0.83	−0.90	−0.82
−13.0		−1.01	−1.12	−0.99		−0.85	−0.95	−0.84		−0.72	−0.81	−0.70		−0.54	−0.61	−0.52
−12.0		−0.67	−0.80	−0.66		−0.53	−0.65	−0.52		−0.42	−0.52	−0.40		−0.26	−0.35	−0.25
−11.0	10	−0.39	−0.54	−0.39		−0.26	−0.40	−0.26		−0.16	−0.28	−0.15		−0.02	−0.13	−0.01
−10.0		−0.21	−0.40	−0.20		−0.08	−0.26	−0.07		0.03	−0.14	0.03		0.17	0.03	0.17
−9.0		−0.17	−0.50	−0.14		−0.03	−0.33	−0.00		0.09	−0.19	0.11		0.26	0.04	0.27
−8.0		−0.42	−1.03	−0.28		−0.24	−0.80	−0.12		−0.08	−0.58	0.03		0.18	−0.21	0.24