Cross-correlation of the Lyα and C IV forests as a probe of the intergalactic medium enrichment

Sheng, X.; Tie, S.S.; Hennawi, J.F.

Citation

Sheng, X., Tie, S. S., & Hennawi, J. F. (2025). Cross-correlation of the Lyα and C IV forests as a probe of the intergalactic medium enrichment. *Monthly Notices Of The Royal Astronomical Society*, *542*(4), 3144-3153. doi:10.1093/mnras/staf1378

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **542,** 3144–3153 (2025) https://doi.org/10.1093/mnras/staf1378
Advance Access publication 2025 August 20

Cross-correlation of the Ly α and C IV forests as a probe of the intergalactic medium enrichment

Xin Sheng [1,2]★ Suk Sien Tie [1] and Joseph F. Hennawi [1,3]

[1] *Department of Physics, University of California, Santa Barbara, CA 93106, USA*
[2] *Department of Physics, Columbia University, New York, NY 10027, USA*
[3] *Leiden Observatory, Leiden University, Niels Bohrweg 2, NL-2333 CA Leiden, the Netherlands*

Accepted 2025 August 8. Received 2025 August 3; in original form 2024 October 2

ABSTRACT

The enrichment of the intergalactic medium (IGM) provides important information about the history of galaxy formation and evolution, because the amount, extent, and distribution of metals in the IGM are related to the star formation history and outflow mechanisms of galaxies. Traditional approaches that rely on the detection of discrete absorption lines are challenging to implement at high redshifts ($z \gtrsim 5$) due to the worsening detector sensitivity combined with increasing sky background in the near-infrared (NIR), resulting in low signal-to-noise spectra. Using mock skewers drawn from the `Nyx` simulations, we present the cross-correlation between the Ly α and C IV forests as a probe to constrain the enrichment topology and metallicity of the IGM. We simulate the IGM at $z = 4.5$ using the `Nyx` hydrodynamical simulation and model IGM enrichment by assuming that certain massive haloes enrich the gas to some radius. We compute the two-point cross-correlation between the Ly α and C IV forests and investigate its dependence on model parameters. Using mock data sets consisting of 25 quasars with signal-to-noise ratio/pixel of 50 at $z = 4.5$, we show that cross-correlation can effectively constrain the minimum halo mass $\log(M)$, the enrichment radius R, and the metallicity [C/H]. Moreover, the combination of the autocorrelation of the C IV forests and the cross-correlation between the Ly α and C IV forests substantially improves the precision of these constraints, with $\gtrsim 2\times$ improvements in measurement precision.

Key words: methods: numerical – intergalactic medium – quasars: absorption lines.

1 INTRODUCTION

Observations of a variety of heavy elements in the intergalactic medium (IGM) through quasar absorption lines raise important questions regarding the origin of these metals. C IV is detected at different redshifts, from $z < 1$ (Cooksey et al. 2010), through $z \sim 2-4$ (Schaye et al. 2003; Simcoe et al. 2011), and up to $z \sim 6$ (Bosman et al. 2017; Codoreanu et al. 2018). Meanwhile, O IV and Mg II absorption lines have also been detected and studied at various redshifts (Aguirre et al. 2008; Pieri et al. 2010; Chen et al. 2017). To account for the widespread distribution of these metals, galactic outflows have been proposed as a primary mechanism for transporting heavy elements into the IGM. Cosmological simulations have been performed to investigate the properties and influences of galactic outflows (Aguirre et al. 2001; Oppenheimer & Davé 2006). However, the efficacy of these outflows and the characteristic mass of the sourcing galaxies remain uncertain (Scannapieco et al. 2006; Booth et al. 2012).

Two methods have been commonly used to measure the metallicity of the IGM. The first method relies on the detection of discrete absorption lines and then performing the Voigt profile fitting to obtain their column densities, e.g. Simcoe, Sargent & Rauch (2004) measured a median intergalactic abundance of [O/H] = 2.7 and [C/H] = 3.1 at $z \sim 2.5$ based on a sample of seven discrete absorbers detected from quasar spectra observed with Keck's High Resolution Echelle Spectrometer (Keck/HIRES). Using a similar line detection method performed at $z \sim 4.3$ with the Folded port InfraRed Echellette (FIRE) spectrograph (Simcoe et al. 2010), Simcoe et al. (2011) measured a $2-3\times$ lower median abundance of [C/H], hinting at the possibility of an evolutionary trend in the metallicity of the IGM. However, at higher redshifts, especially at $z > 5$ where C IV redshifts into the saturated hydrogen wavelength range, such measurements of discrete absorbers are sparse due to the worsening sensitivity of detectors combined with the higher sky background noise in the near-infrared (NIR). As such, measurements of discrete absorption lines at high redshifts probe the circumgalactic medium (CGM) instead of the IGM (Ryan-Weber et al. 2009; Bosman et al. 2017). The second method aims to measure the IGM metallicity by analysing the statistical properties of the absorbers rather than individual absorption lines. One such statistical measurement is the so-called pixel optical method (Aguirre, Schaye & Theuns 2002; Schaye et al. 2003; Aguirre et al. 2004, 2008), which measures the optical depth of a metal absorption line as a function of the H I optical depth. This measurement is then converted into a relation of metal abundance as a function of gas density using simulations

---

★ E-mail: xs2525@columbia.edu

and assuming the UV background (UVB) model. Another statistical approach is to measure the flux correlation of metal-line forests, first proposed by Hennawi et al. (2021), in which they investigated the autocorrelation of the Mg II forest at $z = 7.5$ as a probe of metallicity and neutral fraction by treating the flux field as a continuous random field and assuming a simple toy model in which the IGM is uniformly enriched with metals. Tie et al. (2022) simulated a non-uniform metal enrichment and investigated this method on the C IV forest at $z = 4.5$ as a probe of the metallicity and enrichment topology parameters.

Correlation techniques applied to observational data provide quantitative constraints on metal enrichment in the IGM. For example, Tie et al. (2024) employed autocorrelation of Mg II absorption lines to constrain the [Mg/H] < -3.45 around $z \sim 6.715$ and [Mg/H] < -3.75 around $z \sim 6.235$ based on a data set comprising 10 quasars observed with Keck/MOSFIRE (Multi-Object Spectrograph for Infrared Exploration), Keck/NIRES (Near-Infrared Echellette Spectrometer), and VLT/X-shooter.Inspired by the doublet structure, Karaçaylı et al. (2023) calculated the power spectrum of C IV, Si IV, and Mg II using data from the Dark Energy Spectroscopic Instrument (DESI), revealing evidence that the abundance of C IV is higher than previous estimates.

Following the work of Hennawi et al. (2021) and Tie et al. (2022), we study the cross-correlation function between C IV and Ly α forests as a complementary probe of the IGM enrichment at $z = 4.5$. We choose to perform our analyses with carbon, as it is one of the most abundant elements in the IGM after hydrogen and oxygen. In addition, the double feature of the C IV transition line at rest-frame wavelengths 1548.20 and 1550.78 Å, resulting in an autocorrelation peak at $\Delta v = 498$ km s^{-1}, lends itself naturally to correlation function analyses.

Autocorrelation methods have yielded valuable insights, but also present inherent limitations. In particular, their quadratic dependence on the metal absorption line depth can lead to degeneracies in the parameter space, where different parameter combinations produce similar autocorrelation profiles. In contrast, cross-correlation methods depend linearly on the metal absorption line depth, resulting in a distinct degeneracy structure that can help mitigate these issues. Similar inspirations have been applied in calculating the cross-power spectrum of Ly α and Ly β forests, where the constraints provided by the Ly α forest alone show significant degeneracies (Iršič & Viel 2014; Wilson, Iršič & McQuinn 2021). Furthermore, since the Ly α forest traces gas with $N_{\rm H\,I} = 10^{14} - 10^{17}$ cm^{-2}, the cross-correlation between the C IV and Ly α forests is a more accurate probe of the IGM metallicity compared to the autocorrelation of the forest as done in Tie et al. (2022), who do not model the IGM metallicity as a function of density. Overall, integrating both auto and cross-correlation techniques captures complementary properties of the gas and enhances the precision of metal abundance measurements.

The study of the Ly α forest and metal absorption lines has a long history, with early investigations dating back to decades ago (Sargent et al. 1980; Steidel 1990; Ellison et al. 2000). Over the decades, significant advancements in observational instrumentation have led to dramatic improvements in both signal-to-noise ratios and spectral resolution. In particular, the advent of the *James Webb Space Telescope* (*JWST*) has allowed higher spectral resolution and more precise measurements of the Ly α and metal forests, showing signal-to-noise ratio (SNR) $\approx$ 50–200 (Appleton et al. 2023; Christensen et al. 2023; Prieto-Lyon et al. 2023; Heintz et al. 2025). The integration of these high-quality *JWST* observations with our refined correlation methodologies holds considerable promise for investigating the processes that govern IGM enrichment at high redshifts.

The paper is organized as follows: in Section 2, we present the simulations we use to model the IGM and models for the distribution of metals, and we describe how we calculate the correlation function, and how we constrain the model parameters using mock observations. In Section 3, we present our results and constraints. We show that by combining the autocorrelation and cross-correlation functions, the uncertainty of the constraint can be significantly reduced by a factor of 2 to 5. In Section 4, we discuss our results and conclude.

2 METHODS

In this section, we introduce the methods that we use to simulate the IGM, generate the spectrum, and calculate the cross-correlation function.

2.1 Simulation and model

We use the same simulation as Tie et al. (2022), which we briefly summarize here. We simulate the IGM using the simulation code `Nyx`, which is an *N*-body hydrodynamical code designed to simulate the IGM (Almgren et al. 2013; Lukić et al. 2014). Our simulation assumes Lambda cold dark matter (ΛCDM) cosmology with $\Omega_m = 0.3192$, $\Omega_\Lambda = 0.6808$, $\Omega_b = 0.04964$, $h = 0.6704$, $\sigma_8 = 0.826$ and $n_s = 0.9655$, which agrees with the cosmological constraints from the cosmic microwave background (CMB; Planck Collaboration VI 2020). Initial conditions were generated using the MUSIC code (Hahn & Abel 2011) with a transfer function generated by CAMB (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012). The simulation starts from $z = 159$ and assumes that reionization occurs instantaneously when the ultraviolet (UV) background is turned on at $z = 6$. We analyse the simulation snapshot at $z = 4.5$, and our simulation box has 4096^3 grid cells and a length of 100 h^{-1} cMpc on each side, resulting in a pixel scale of 24 h^{-1} ckpc. The simulation outputs the overdensity, temperature, and peculiar velocity of the IGM. We adopt the same metal distribution model as Tie et al. (2022), who generated a non-uniform distribution of metals assuming massive haloes can enrich their surroundings out to some maximum radius with some constant metallicity. The enrichment of the IGM is determined by three parameters: the enrichment radius R describes the radial extent of a halo's enrichment; the minimum halo mass M sets a minimum mass on the enriching haloes that can enrich the surroundings, and the metallicity [C/H] describes the metal abundance of the IGM. We define metallicity [C/H] $\equiv \log_{10}(Z/Z_\odot)$, where Z is the ratio of the number of carbon atoms to the number of hydrogen atoms and $Z_\odot$ is the corresponding solar ratio with $\log_{10}(Z_\odot) = -3.57$ (Asplund et al. 2009).

2.2 Calculating the Ly α and C IV forests

To calculate the cross-correlation function, we must first generate the Ly α and C IV forests. The optical depth for the Ly α forests is calculated as

$$\tau_{\rm H} = \tau_{\rm H,0} \int \frac{x_{\rm H\,I}\Delta}{\sqrt{\pi}} \exp\left[-\left(\frac{v' - v}{b}\right)^2\right] \frac{{\rm d}v'}{b}, \quad (1)$$

where $\Delta = \rho/\langle\rho\rangle$ is the overdensity, v' is the velocity after Doppler transformation, b is the Doppler broadening parameter, and $\tau_{\rm H,0}$ is the Gunn–Peterson optical depth

$$\tau_{\mathrm{H},0} = \frac{\pi e^2 f_{lu,\mathrm{H}} \lambda_{lu,\mathrm{H}} n_{\mathrm{H}}}{m_e c \mathrm{H}(z)}, \tag{2}$$

where $f_{lu,\mathrm{H}}$ is the oscillator strength, and $\lambda_{lu,\mathrm{H}}$ is the wavelength of the transition $l \rightarrow u$. Similarly, the optical depth of the C IV forest is

$$\tau_{\mathrm{C}} = \tau_{\mathrm{C},0} \int \frac{x_{\mathrm{C\,IV}} \Delta}{\sqrt{\pi}} \exp\left[-\left(\frac{v' - v}{b}\right)^2\right] \frac{\mathrm{d}v'}{b}, \tag{3}$$

where $x_{\mathrm{C\,IV}}$ is the fraction of C IV, and

$$\tau_{\mathrm{C},0} = \frac{\pi e^2 f_{lu,\mathrm{C}} \lambda_{lu,\mathrm{C}} n_{\mathrm{C}}}{m_e c \mathrm{H}(z)}, \tag{4}$$

where n_{C} is the number density of carbon. Tie et al. (2022) computed the fraction of C IV using CLOUDY on a grid of hydrogen densities, gas temperatures, and metal abundance. The transmitted flux F is

$$F_{\mathrm{H}} = e^{-\tau_{\mathrm{H}}} \tag{5}$$

$$F_{\mathrm{C\,IV}} = e^{-\tau_{\mathrm{C}}} \tag{6}$$

For given enrichment topology and metallicity, 10 000 skewers were drawn randomly from one face of the selected `Nyx` simulation box. Fig. 1 shows skewers of overdensity, optical depth of H I, transmitted flux of H I, and transmitted flux of C IV for one model. The overdensity is obtained from the `Nyx` simulation. The optical depths of hydrogen and C IV are given by equations (1) and (3), respectively. The total transmitted flux is the addition of the red and blue transmitted flux. The overdensity plot correlates with the optical depth of the hydrogen. The strong absorption line in H I is caused by the large scattering cross-section. The C IV absorption lines are much weaker than those of hydrogen.

2.3 Generating the mock observations

In our simulation, one random skewer corresponds to one spectral line, which is generated by a quasar behind a series of individual absorbers and the IGM. The statistical method is valid only if we can apply it to one of such skewers. Therefore, we generate mock observations to test the validity of our constraint. The mock observations represent the spectral lines we could observe in reality. We generate mock data sets for all models ranging from $\log(M) = 8.5 - 11.0\,\mathrm{M}_{\odot}$ with $\mathrm{d}[\log(M)] = 0.1$, $R = 0.1 - 3.0$ cMpc with $\mathrm{d}R = 0.1$, $[\mathrm{C/H}] = -2.0 - -4.5$ with $\mathrm{d}[\mathrm{C/H}] = 0.1$, which gives a total of 20 280 models. We generate 10^6 realizations of each data set for each model. One of the goals is to make a direct comparison between the autocorrelation in Tie et al. (2022) and the cross-correlation in this paper. Therefore, we keep the set-up of the mock data set the same in both papers, where full width at half-maximum (FWHM) $= 10$ km s^{-1}. We introduce Gaussian random noises into our mock data set and the signal-to-noise ratio (SNR) is 50 per pixel. We assume the Ly α and C IV forests have a path-length of $\mathrm{d}z = 0.8$, which corresponds to the range from the Ly β to the Ly α line. To ensure the same total path-length between the autocorrelation and cross-correlation study, we assume the number of the QSO sightlines to be 25 here, which gives a desired total path-length of $\Delta z = 20$. The spectra are convolved with a Gaussian line spread function with FWHM $= 10\,\mathrm{km\,s}^{-1}$. Our spectral sampling is 3 pixels per resolution element.

3 RESULTS

In this section, we compute the cross-correlation function of the Ly α and C IV forests and use it to constrain the physical parameters of the IGM.

3.1 Cross-correlation function

To compute the cross-correlation function of the C IV and Ly α forests extracted from the simulation, we first calculate the flux fluctuation of the C IV and Ly α forests as follows,

$$\delta_{\mathrm{C}} = \frac{F_{\mathrm{C}} - \langle F_{\mathrm{C}} \rangle}{\langle F_{\mathrm{C}} \rangle}, \quad \delta_{\mathrm{H}} = \frac{F_{\mathrm{H}} - \langle F_{\mathrm{H}} \rangle}{\langle F_{\mathrm{H}} \rangle}, \tag{7}$$

where δ_{H} and δ_{C} represent the flux fluctuation of the Ly α and C IV forests and $\langle F \rangle$ is the mean flux computed from all 10 000 skewers, respectively. The cross-correlation function $\xi_f(\mathrm{d}v)$ computed at a velocity $\mathrm{d}v$ is defined as the expectation value of all pixel pairs separated by $\mathrm{d}v$,

$$\xi_f(\mathrm{d}v) = \langle \delta_{\mathrm{H}}(v) \delta_{\mathrm{C}}(v + \mathrm{d}v) \rangle. \tag{8}$$

Similarly, the autocorrelation functions of the H I and C IV forests are

$$\xi_{\mathrm{H}}(\mathrm{d}v) = \langle \delta_{\mathrm{H}}(v) \delta_{\mathrm{H}}(v + \mathrm{d}v) \rangle \tag{9}$$

and

$$\xi_{\mathrm{C}}(\mathrm{d}v) = \langle \delta_{\mathrm{C}}(v) \delta_{\mathrm{C}}(v + \mathrm{d}v) \rangle. \tag{10}$$

We create Figs 2, 3, and 4 by averaging 10 000 noiseless skewers for each given model, where each skewer is convolved with a Gaussian line spread function with FWHM $= 10$ km s^{-1}. In Fig. 2, ξ_{C} has a peak at $\Delta v = 498$ km s^{-1} due to the doublet feature of the C IV transition line at rest-frame wavelengths 1548.20 and 1550.78 Å. Fig. 3 shows ξ_{H}, which rises at small scales and averages to zero at large scales. Compared to ξ_{C}, it is featureless because the hydrogen transition does not have a significant doublet. The amplitude of ξ_{H} is also significantly higher than that of ξ_{C} because the Ly α forest has larger flux fluctuations in general (see Fig. 1). For ξ_f in Fig. 4, since the abundance of C IV correlates with the abundance of hydrogen, we also have a similar but wider peak around that velocity separation. The ξ_f is flatter and has an overall larger amplitude, due to the larger flux fluctuations coming from δ_{H}, compared to ξ_{C}. The enrichment topology and metallicity influence both the peak and the shape of the cross-correlation function. For instance, the peak drops when we increase the minimum halo mass (less galaxies contribute towards enrichment), decrease the enrichment radius (the metals have smaller volume filling fraction), and decrease the metallicity.

3.2 Constraint on enrichment parameters

We quantify the precision of the enrichment parameters inferred from the mock observations detailed in Section 2.3, which consists of 25 quasars with FWHM $= 10$ km s^{-1} and SNR $= 50$. We assume a multivariate Gaussian likelihood

$$L(\hat{\xi}(\Delta v)) = \frac{1}{\sqrt{(2\pi)^k \det(\boldsymbol{C})}} \exp(-\frac{1}{2} \boldsymbol{d}^T \boldsymbol{C}^{-1} \boldsymbol{d}), \tag{11}$$

where $\boldsymbol{d} = \hat{\xi}(\Delta v) - \xi(\Delta v | \log(M), R, [\mathrm{C/H}])$ is the difference between the correlation function measured from the mock data set $\hat{\xi}(\Delta v)$ and the model correlation function $\xi(\Delta v | \log(M), R, [\mathrm{C/H}])$,

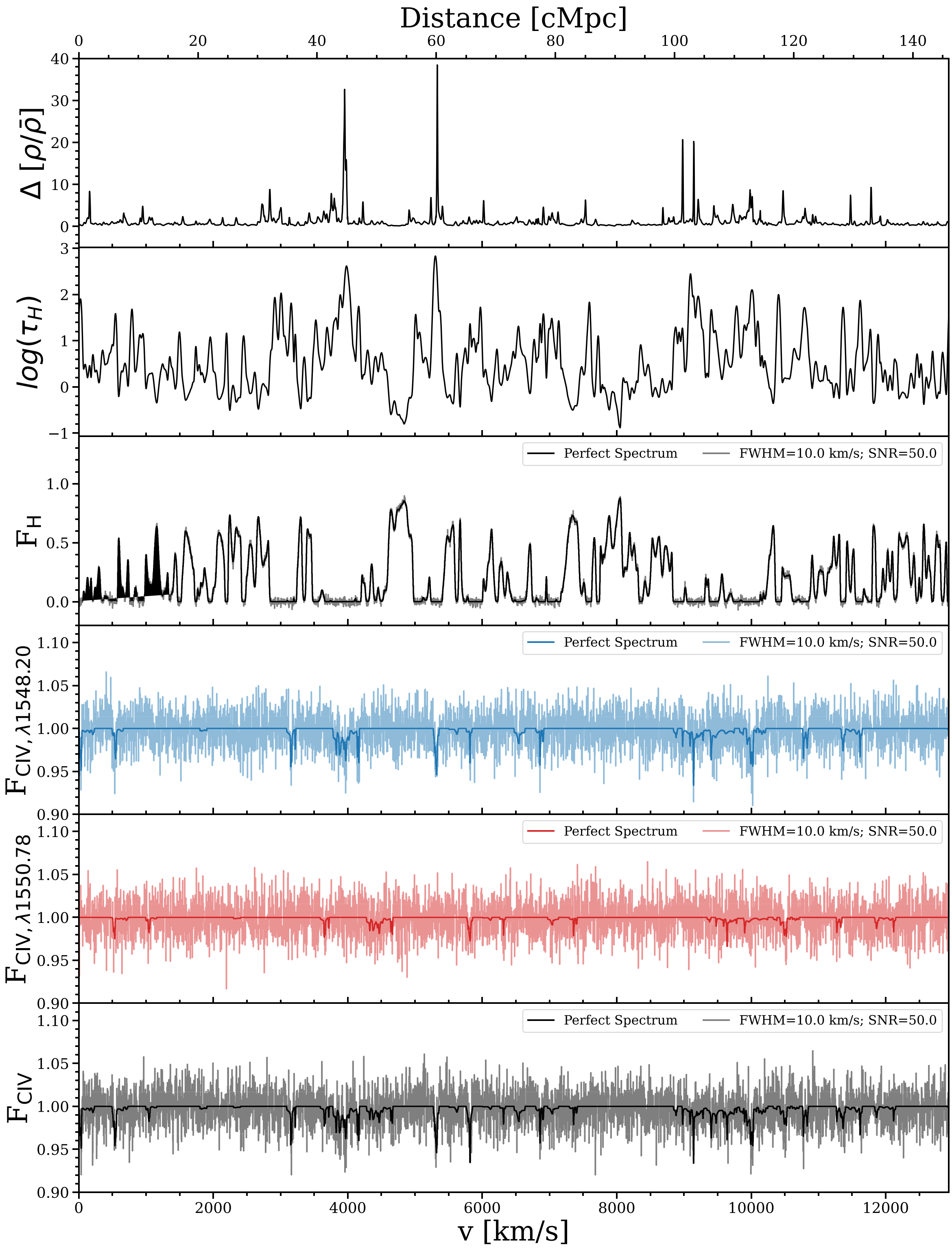

Figure 1. Skewers of various properties for hydrogen and C IV at the same line of sight, with FWHM = 10 km s^{-1} and SNR = 50. From top to bottom: overdensity, hydrogen optical depth, hydrogen transmitted flux, C IV transmitted flux at 1548.20 Å, C IV transmitted flux at 1550.78 Å, and the total C IV transmitted flux. The noisy transmitted flux is calculated by adding Gaussian random noise on the noiseless flux.

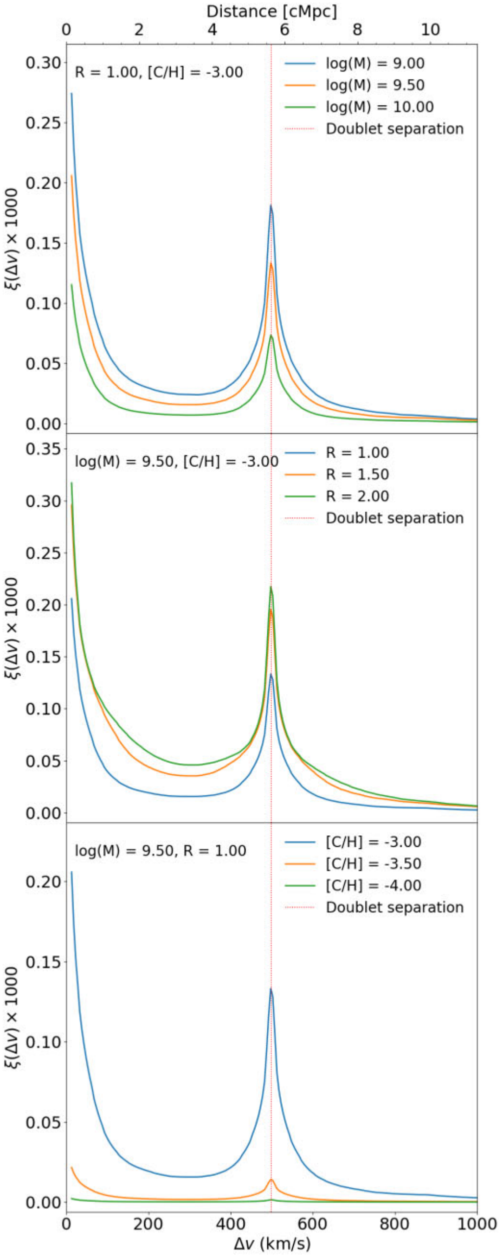

Figure 2. Autocorrelation function of the C IV forest for different minimum halo masses (top), enrichment radii (middle), and metallicities (bottom). The minimum halo mass is in $M_\odot$ and the enrichment radius is given in cMpc. The vertical dashed line indicates the doublet separation of C IV at 498 km s^{-1}.

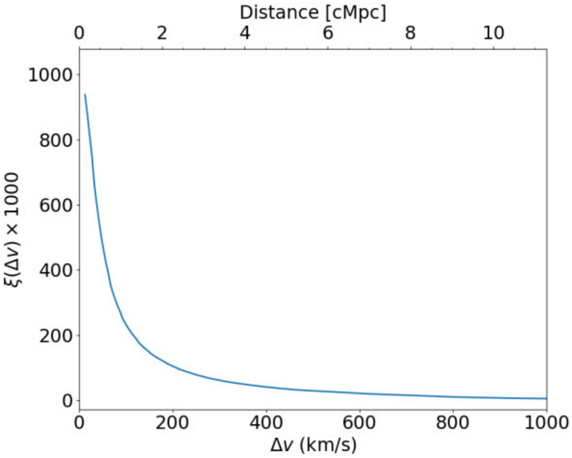

Figure 3. Autocorrelation function of the Ly α forest. The function shows no correlation peak except at 0 km s^{-1}, but the overall correlation is stronger due to the higher density of H I.

and $\boldsymbol{C}$ is the covariance matrix, which has elements

$$C_{ij} = \langle [\hat{\xi}(\Delta v) - \xi(\Delta v)]_i [\hat{\xi}(\Delta v) - \xi(\Delta v)]_j \rangle, \tag{12}$$

where i and j indicate different bins of Δv and the angle brackets denote averaging over 10^6 realizations of each mock data set. The correlation matrix is defined as

$$\mathrm{Corr}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}. \tag{13}$$

Fig. 5 shows an example of a joint correlation matrix for the model $\log(M) = 9.5$, $R = 1.1$ cMpc and $[\mathrm{C/H}] = -3.5$, including the autocorrelation matrices $\mathrm{Corr}_{ij}[\xi_f, \xi_f]$ and $\mathrm{Corr}_{ij}[\xi_\mathrm{C}, \xi_\mathrm{C}]$, and the cross-correlation matrix $\mathrm{Corr}_{ij}[\xi_\mathrm{C}, \xi_f]$. To clarify, the 'autocorrelation matrix' here refers to the correlation matrix between the same ξ, while the 'cross-correlation matrix' refers to the correlation matrix between different ξ. The ξ_C are calculated based on Tie et al. (2022), with a bin size of 10 km s^{-1} from $\Delta v = 10 - 2000$ km s^{-1}. We compute the ξ_f with a bin size of 10 km s^{-1} from $\Delta v = 10 - 1000$ km s^{-1} and 50 km s^{-1} from $\Delta v = 1000 - 2000$ km s^{-1}. This is done in order to minimize the computational cost while keeping high accuracy, as the main feature of the cross-correlation is most apparent around $10 - 1000$ km s^{-1}, while the function approaches zero in greater than 1000 km s^{-1} (as shown in Fig. 4). The bright diagonal structure in $\mathrm{Corr}_{ij}[\xi_f, \xi_f]$ and $\mathrm{Corr}_{ij}[\xi_\mathrm{C}, \xi_\mathrm{C}]$ is expected, given that the same bins should be the most highly correlated. The two darker diagonal structures parallel to the centre one indicate correlation at ± doublet separation of the C IV forest around each bin. $\mathrm{Corr}_{ij}[\xi_f, \xi_\mathrm{C}]$ reveals how ξ_f and ξ_C co-vary. It has similar but much darker diagonal structures at the centre and ± doublet separation. Such correlation is expected since both ξ_f and ξ_C trace the same large-scale structures. Nevertheless, the correlation is weak since they trace different species and are subject to independent noise realizations. In this example, the diagonal structure of the $\mathrm{Corr}_{ij}[\xi_f, \xi_\mathrm{C}]$ has a mean value of ~ 0.3, and the median value of the whole $\mathrm{Corr}_{ij}[\xi_f, \xi_\mathrm{C}]$ matrix is only ~ 0.1. Therefore, ξ_f carries information that is only slightly degenerate with ξ_C, and combining them should yield a significant improvement in the constraining power.

We begin by evaluating the likelihood in equation (11) on our original parameter grid of size 26 × 29 × 26 and creating a coarser likelihood grid. Then, we obtain samples of the parameter vector from the posterior distribution using the Markov Chain Monte Carlo (MCMC) sampler, EMCEE (Foreman-Mackey et al. (2013)). In order to increase the speed and precision of the MCMC sampling, we run the sampling on a finer likelihood grid of size 261 × 291 × 261 with bin size of 0.01 in $\log(M)$, R, and [C/H]. Directly computing the likelihood at every fine-grid point would be prohibitively expensive, so we adopt a hybrid interpolation scheme: we first compare the likelihood of each model grid p with the maximum likelihood p_max to determine the interpolation method. For the low-likelihood region ($\ln p/p_\mathrm{max} < -10$), we linearly interpolate the likelihood. For the high-likelihood region ($\ln p/p_\mathrm{max} > -10$), we linearly interpolate the correlation function of the model parameters

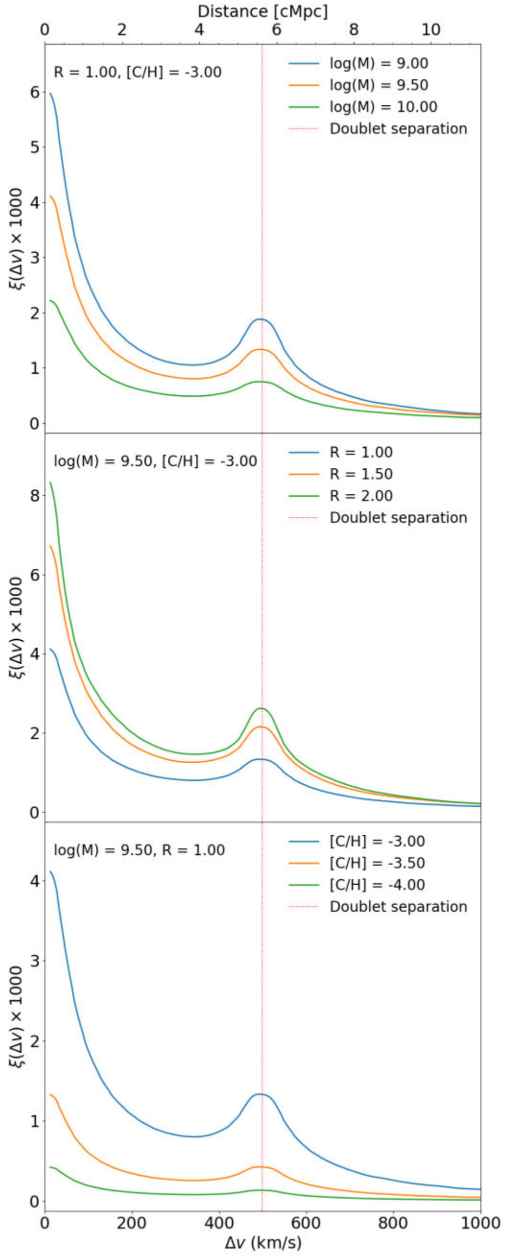

Figure 4. Cross-correlation function of the C IV forest for different minimum halo masses (top), enrichment radii (middle), and metallicities (bottom).

(ξ) on to the finer parameter grid and evaluate equation (11). The mock data $\hat{\xi}$ and the binning size Δv are fixed throughout the process. Therefore, we save computational resources from the low likelihood region, while keeping high accuracy in the high likelihood region. We run 30 walkers with 100 000 steps for three parameters, burning the first 1000 steps for each walker to reach equilibrium. The inferred parameters are the mean of the samples, while the 1σ error is obtained from the posterior distribution.

As previously discussed, $\xi_{\rm C}$ and ξ_f probe complementary aspects of the enrichment. As shown in Fig. 5, the two are not fully correlated, so combining them is expected to yield tighter constraints on the model. We construct the combined model by multiplying the likelihoods on each grid point of $\xi_{\rm C}$ and ξ_f, and then performing MCMC sampling on the resulting joint probability distribution. This method effectively assumes that the two data sets are independent, which is not strictly true in realistic observations – $\xi_{\rm C}$ and ξ_f would typically be computed from the same set of sightlines, and their joint covariance matrix, including cross-covariances between bins, should be used. Nevertheless, we adopt this simplified combination as a proof of concept. The resulting constraints are likely optimistic, with realistic uncertainties expected to be larger by a factor of up to $\lesssim \sqrt{2}$. This approximation is justified in our case, given that the measured correlation matrix $\mathrm{Corr}_{ij}[\xi_{\rm C}, \xi_f]$ is not significant. Figs 6 and 7 show mock data sets and the inference result in for two different models. From the plots, the peak of the correlation is captured at the doublet separation. The inferred model, which is based on the mock data, is close to the true model. In Fig. 6, although the cross-correlation data show some variance around the inferred model, the peak remains robust across different skewers, consistently marking the doublet structure. Variations in the topology of the absorbers affect the smooth component, leading to differences across individual skewers; however, the peak structure remains similar in all cases. Consequently, even though the smooth region may be skewed when averaging over different skewers, the distinct peak is preserved. The corner plots show that the degeneracies in auto and cross-correlation are different. The combined probability solves the degeneracies and improves the prediction of the model.

The χ^2-test is performed to check the goodness of the fit,

$$\chi^2_\nu = \frac{\boldsymbol{d_m}^T \boldsymbol{C}^{-1} \boldsymbol{d_m}}{\nu} \tag{14}$$

where $\boldsymbol{d_m}$ is the difference between the measured and predicted correlation function, and ν is the number of degrees of freedom. In our current proof-of-concept we have treated each velocity bin as independent when computing ν. However, since our spectral bins have finite width and smoothing, both Ly α and C IV measurements are intrinsically correlated, especially between neighboring bins (and at C IV doublet separations). Therefore, the effective degrees of freedom are lower than ν for both $\xi_{\rm C}$ and ξ_f. As a result, one often finds $\chi^2_\nu \lesssim 1$ even when the model describes the data well. Due to the same reason, the correlated fluctuations clusters and the fitting might look noisy by eyes. Despite that, for all of our models, χ^2_ν is close to unity, confirming acceptable fits under our independence approximation.

We observe that the MCMC samples reach the grid boundaries in some cases, as seen in the corner plots. This is a natural limitation of using a finite grid. Although such edge-hitting behaviour may potentially bias the results by limiting the exploration of the full parameter space, in our analysis the 1σ confidence regions remain well within the grid boundaries. Moreover, this issue can be mitigated in practice. If the observational data lie near or beyond the grid edges, the model grid can be extended to ensure the parameter space is not artificially truncated.

The complete inference results are shown in Tables 1 and 2. The cross-correlation has similar performance in constraining the enrichment parameters compared to autocorrelation, and it is less sensitive to the spectrum fluctuation and metal contamination. Moreover, better constraint is achieved by simply combining these two correlations together. From the table, the combined model shows significant improvements in precision. In Table 1, the precision is increased by ∼ ±0.3 dex for log M, ∼ ±0.1 dex for R, and ∼ ±0.2 dex for [C/H]. In Table 2, the precision is increased by ∼ ±0.2 dex for log M, ∼ ±0.1 dex for R, and ∼ ±0.2 dex for [C/H].

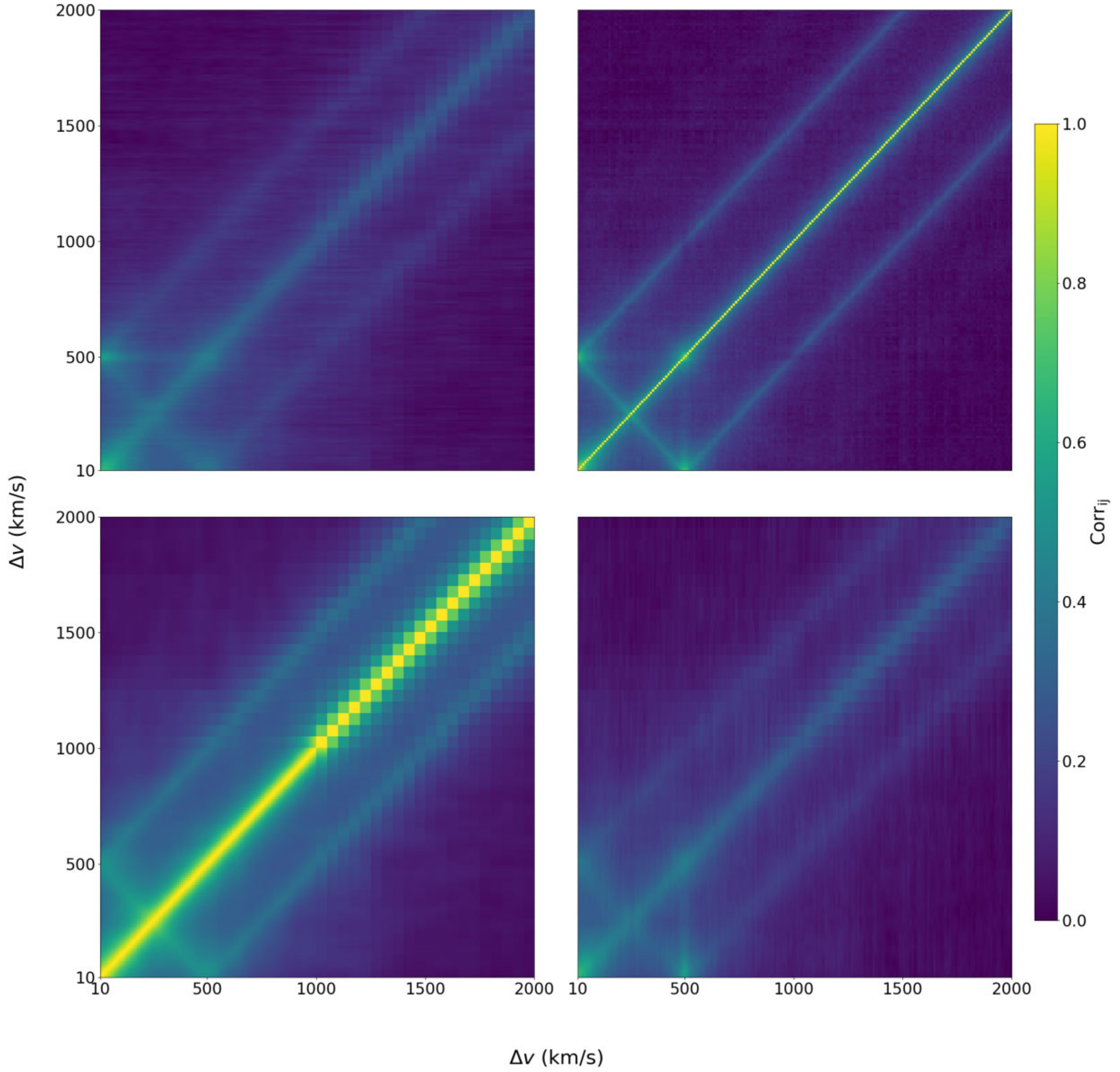

Figure 5. The joint correlation matrix for parameters $\log(M) = 9.5$, R = 1.1 cMpc and [C/H] = 3.5, including $\mathrm{Corr}_{ij}[\xi_f, \xi_{\mathrm{C}}]$ (top left), $\mathrm{Corr}_{ij}[\xi_{\mathrm{C}}, \xi_{\mathrm{C}}]$ (top right), $\mathrm{Corr}_{ij}[\xi_f, \xi_f]$ (bottom left), and $\mathrm{Corr}_{ij}[\xi_{\mathrm{C}}, \xi_f]$ (bottom right). For ξ_{C}, the resolution is 10 km s^{-1} for $\Delta v = 10 - 2000$ km s^{-1}. For ξ_f, the resolution is 10 km s^{-1} for $\Delta v = 10 - 1000$ km s^{-1}, and 50 km s^{-1} for $\Delta v = 1000 - 2000$ km s^{-1}. The symmetrical diagonal structure in $\mathrm{Corr}_{ij}[\xi_f, \xi_f]$ represents the correlation between the two forests. Two diagonal lines have $\pm\Delta v = 498$ km s^{-1}, which corresponds to the doublet separation of the C IV forest. The darker diagonal structure in $\mathrm{Corr}_{ij}[\xi_{\mathrm{C}}, \xi_f]$ shows the correlation between two different ξ functions. Since the correlation is not significant, combining the ξ_f and ξ_{C} together will improve the constraining power.

4 DISCUSSION AND CONCLUSION

In this work, we investigate the cross-correlation function of the C IV and Ly α forests and use it to constrain the metallicity and enrichment topology of the IGM around the redshift $z = 4.5$. We use Nyx to simulate the $z = 4.5$ IGM and model the enrichment distribution as non-homogeneous. We consider the enrichment parameters as the minimum halo mass, the enrichment radius, and the metallicity. We generate skewers of the C IV and Ly α forests for various models, compute the cross-correlation, and generate mock observations to investigate the parameter constraints. The peak of the cross-correlation function drops when we increase the minimum halo mass, decrease the enrichment radius, and decrease the metallicity. Cross-correlation effectively constrains the parameters $\log(M)$, R, and [C/H]. Furthermore, combining auto and cross-correlation enhances our ability to constrain $\log(M)$, R, and [C/H] with greater precision.

The presence of the CGM can bias the measurements of the underlying IGM signal Tie et al. (2022). Both the C IV and Ly α forests may be contaminated by lower redshift metal systems. However, because we use the cross-correlation between the C IV and Ly α forests, the influence of CGM absorbers is minimized, as the Ly α forest primarily traces the diffuse IGM gas. Consequently, contamination from lower redshift systems is expected to have little impact on the cross-correlation function. The only potential source of contamination arises from the lower redshift C IV absorption lines present in the Ly α forest that could correlate

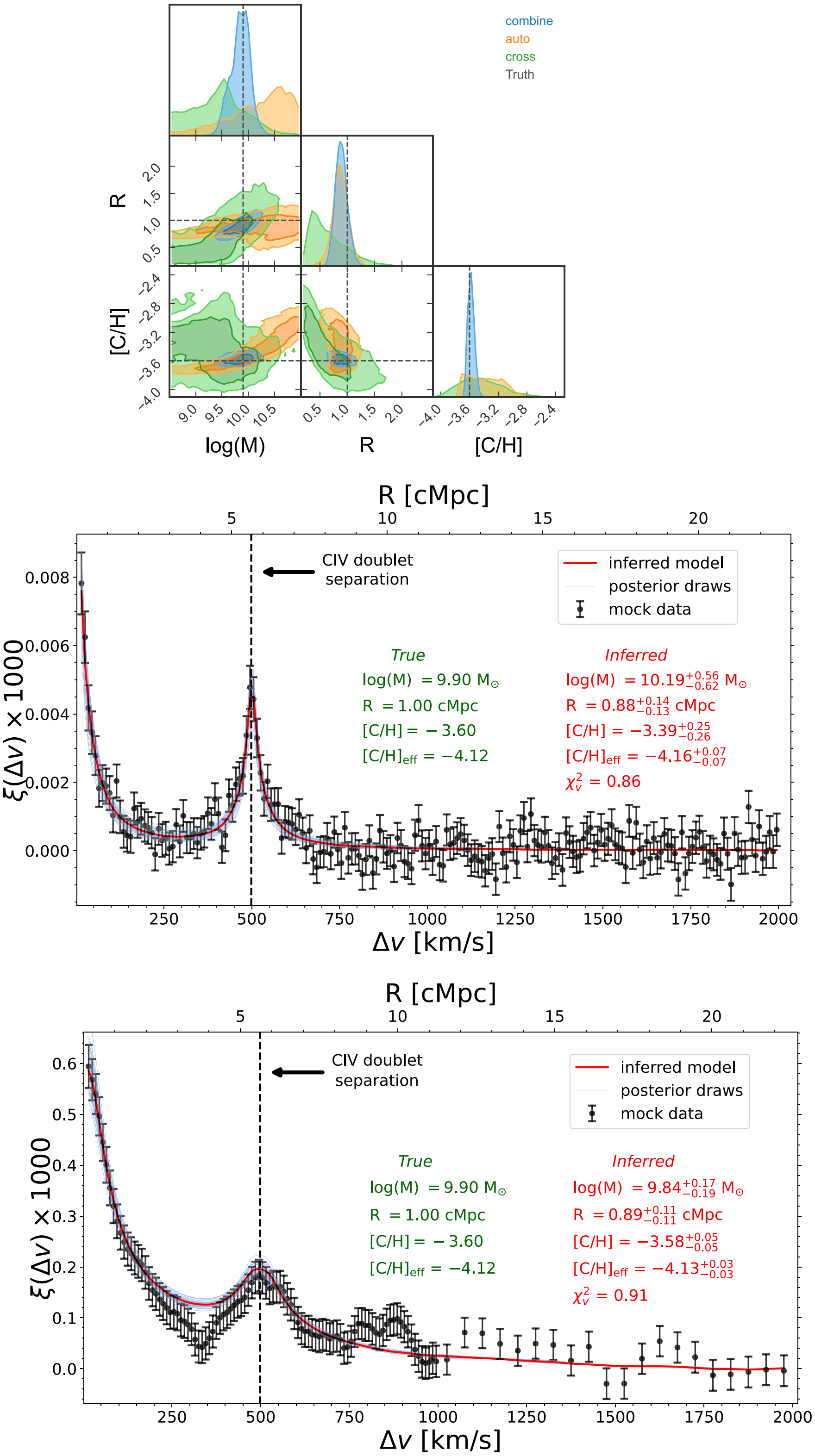

Figure 6. Results for the model $\log(M) = 9.9\,\mathrm{M}_{\odot}$, $R = 1.0$ cMpc, and [C/H] = −3.60. The mock data set has 25 QSOs, $\mathrm{d}z = 0.8$, and SNR = 50. **Top**: Inference result of the mock data set showing the posterior probability contour, using autocorrelation ξ_{C}, cross-correlation ξ_{f}, and the combined likelihood. The truth is indicated with a dashed line. **Middle**: Autocorrelation function and the inferred result. **Bottom**: Cross-correlation function and the inferred result for the combined model.

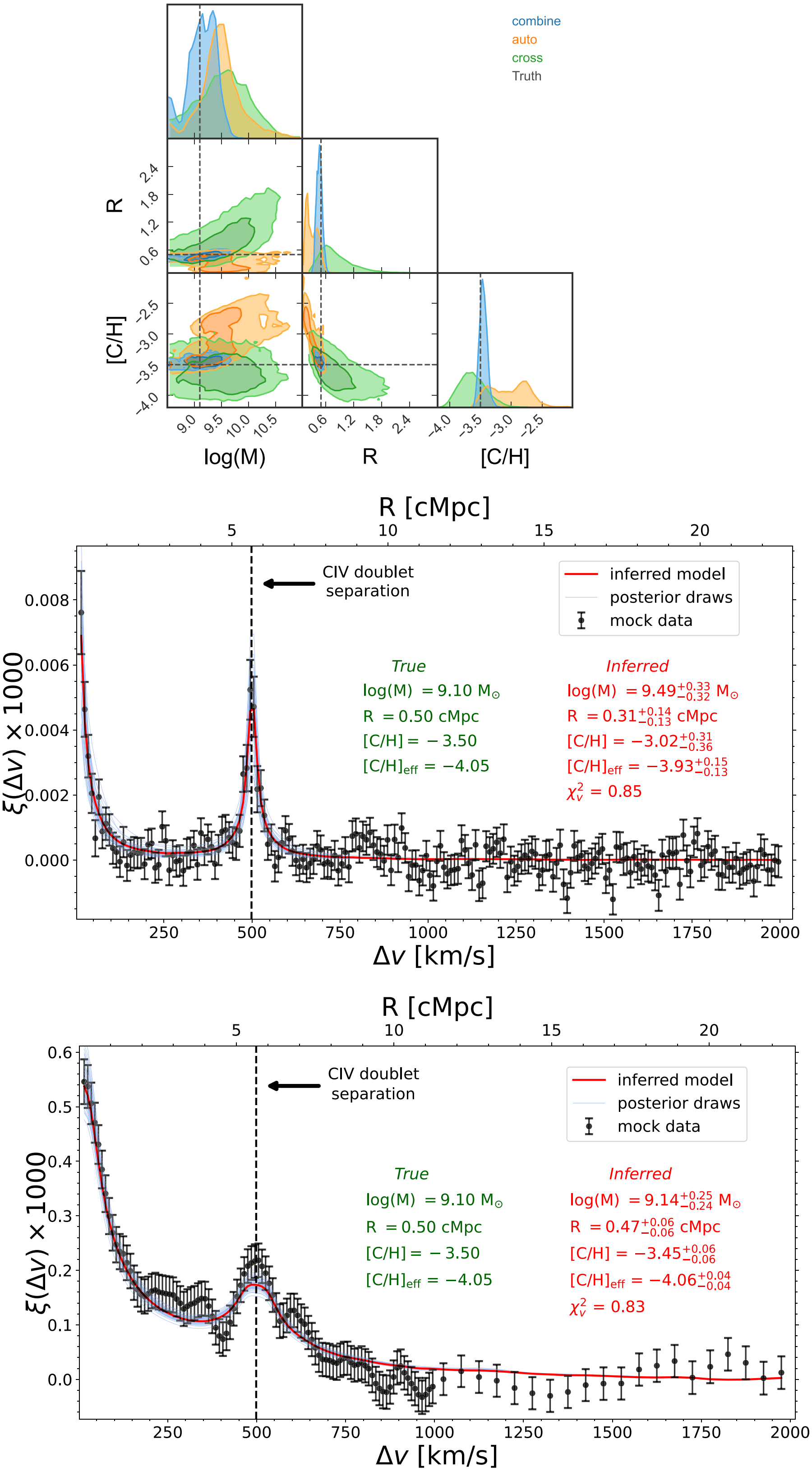

Figure 7. Like Fig. 6 with $\log(M) = 9.1\mathrm{M}_\odot$, $R = 0.5$ cMpc, and [C/H] $= -3.50$.

Table 1. True parameters and inferred parameters in Fig. 6.

Parameters	True model	Auto	Cross	Combined
$\log(M)$ ($\mathrm{M}_{\odot}$)	9.90	$10.19^{+0.56}_{-0.62}$	$9.44^{+0.52}_{-0.58}$	$9.84^{+0.17}_{-0.19}$
R (cMpc)	1.00	$0.88^{+0.14}_{-0.13}$	$0.69^{+0.37}_{-0.34}$	$0.89^{+0.11}_{-0.11}$
[C/H]	−3.60	$-3.39^{+0.25}_{-0.26}$	$-3.45^{+0.30}_{-0.30}$	$-3.58^{+0.05}_{-0.05}$
$[\mathrm{C/H}]_{\mathrm{eff}}$	−4.12	$-4.16^{+0.07}_{-0.07}$	$-4.04^{+0.16}_{-0.15}$	$-4.13^{+0.03}_{-0.03}$

Table 2. True parameters and inferred parameters in Fig. 7.

Parameters	True model	Auto	Cross	Combined
$\log(M)$ ($\mathrm{M}_{\odot}$)	9.10	$9.49^{+0.33}_{-0.32}$	$9.53^{+0.47}_{-0.52}$	$9.14^{+0.25}_{-0.24}$
R (cMpc)	0.50	$0.31^{+0.14}_{-0.13}$	$0.83^{+0.34}_{-0.33}$	$0.47^{+0.06}_{-0.06}$
[C/H]	−3.50	$-3.02^{+0.31}_{-0.36}$	$-3.62^{+0.21}_{-0.21}$	$-3.45^{+0.06}_{-0.06}$
$[\mathrm{C/H}]_{\mathrm{eff}}$	−4.05	$-3.93^{+0.15}_{-0.13}$	$-4.15^{+0.10}_{-0.09}$	$-4.06^{+0.04}_{-0.04}$

with the higher redshift C IV forest, due to their identical velocity separation. Nevertheless, this effect is negligible since the autocorrelation peak is much smaller than the cross-correlation peak (i.e. see Fig. 6). Thus, in this work, we did not account for metal-line contamination.

Traditional methods that rely on measuring discrete absorption lines face significant challenges at high redshift due to limited telescope sensitivity. In contrast, correlation techniques show great promise for resolving metallicity in these distant regions. In particular, combining the cross-correlation between the C IV and Ly α forests with the autocorrelation of the Ly α forest significantly improves precision in constraining IGM enrichment parameters. In future work, we plan to apply our approach to observed quasar spectra, and with more sensitive observations – such as those from the *JWST* – we expect to further refine these constraints.

ACKNOWLEDGEMENTS

We acknowledge helpful conversations with the ENIGMA group at UC Santa Barbara. We specifically thank Molly Wolfson and Teng Hu for useful advice. JFH acknowledges support from the National Science Foundation under Grant No. 2307180.

DATA AVAILABILITY

Data are available upon request to the authors.

REFERENCES

Aguirre A., Hernquist L., Schaye J., Katz N., Weinberg D. H., Gardner J., 2001, ApJ, 561, 521

Aguirre A., Schaye J., Theuns T., 2002, ApJ, 576, 1

Aguirre A., Schaye J., Kim T.-S., Theuns T., Rauch M., Sargent W. L. W., 2004, ApJ, 602, 38

Aguirre A., Dow-Hygelund C., Schaye J., Theuns T., 2008, ApJ, 689, 851

Almgren A. S., Bell J. B., Lijewski M. J., Lukić Z., Andel E. V., 2013, ApJ, 765, 39

Appleton P. N. et al., 2023, ApJ, 951, 104

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Booth C. M., Schaye J., Delgado J. D., Dalla Vecchia C., 2012, MNRAS, 420, 1053

Bosman S. E. I., Becker G. D., Haehnelt M. G., Hewett P. C., McMahon R. G., Mortlock D. J., Simpson C., Venemans B. P., 2017, MNRAS, 470, 1919

Chen S.-F. S. et al., 2017, ApJ, 850, 188

Christensen L. et al., 2023, A&A, 680, A82

Codoreanu A., Ryan-Weber E. V., García L. Á., Crighton N. H. M., Becker G., Pettini M., Madau P., Venemans B., 2018, MNRAS, 481, 4940

Cooksey K. L., Thom C., Prochaska J. X., Chen H.-W., 2010, ApJ, 708, 868

Ellison S. L., Songaila A., Schaye J., Pettini M., 2000, AJ, 120, 1175

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

Hahn O., Abel T., 2011, MNRAS, 415, 2101

Heintz K. E. et al., 2025, A&A, 693, A60

Hennawi J. F., Davies F. B., Wang F., Oñorbe J., 2021, MNRAS, 506, 2963

Howlett C., Lewis A., Hall A., Challinor A., 2012, J. Cosmol. Astropart. Phys., 2012, 027

Iršič V., Viel M., 2014, J. Cosmol. Astropart. Phys., 2014, 024

Karaçaylı N. G. et al., 2023, MNRAS, 522, 5980

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

Lukić Z., Stark C. W., Nugent P., White M., Meiksin A. A., Almgren A., 2014, MNRAS, 446, 3697

Oppenheimer B. D., Davé R., 2006, MNRAS, 373, 1265

Pieri M. M., Frank S., Mathur S., Weinberg D. H., York D. G., Oppenheimer B. D., 2010, ApJ, 716, 1084

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

Prieto-Lyon G. et al., 2023, ApJ, 956, 136

Ryan-Weber E. V., Pettini M., Madau P., Zych B. J., 2009, MNRAS, 395, 1476

Sargent W. L. W., Young P. J., Boksenberg A., Tytler D., 1980, ApJS, 42, 41

Scannapieco E., Pichon C., Aracil B., Petitjean P., Thacker R. J., Pogosyan D., Bergeron J., Couchman H. M. P., 2006, MNRAS, 365, 615

Schaye J., Aguirre A., Kim T.-S., Theuns T., Rauch M., Sargent W. L. W., 2003, ApJ, 596, 768

Simcoe R. A., Sargent W. L. W., Rauch M., 2004, ApJ, 606, 92

Simcoe R. A. et al., 2010, Ground-based and Airborne Instrumentation for Astronomy III, 7735, 469

Simcoe R. A. et al., 2011, ApJ, 743, 21

Steidel C. C., 1990, ApJS, 72, 1

Tie S. S., Hennawi J. F., Kakiichi K., Bosman S. E. I., 2022, MNRAS, 515, 3656

Tie S. S., Hennawi J. F., Wang F., Onorato S., Yang J., Bañados E., Davies F. B., Oñorbe J., 2024, MNRAS, 535, 223

Wilson B., Iršič V., McQuinn M., 2021, MNRAS, 509, 2423