Simulating the effect of high column density absorbers on the one-dimensional Lyman α forest flux power spectrum

Keir K. Rogers,[1]★ Simeon Bird,[2]† Hiranya V. Peiris,[1,3] Andrew Pontzen,[1] Andreu Font-Ribera[1] and Boris Leistedt[4]†

[1]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[2]*Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA*
[3]*Oskar Klein Centre for Cosmoparticle Physics, Stockholm University, AlbaNova, Stockholm SE-106 91, Sweden*
[4]*Department of Physics, New York University, 726 Broadway, New York NY 10003, USA*

ABSTRACT

We measure the effect of high column density absorbing systems of neutral hydrogen (H I) on the one-dimensional (1D) Lyman α forest flux power spectrum using cosmological hydrodynamical simulations from the Illustris project. High column density absorbers (which we define to be those with H I column densities $N(\mathrm{H\,I}) > 1.6 \times 10^{17}$ atoms cm^{-2}) cause broadened absorption lines with characteristic damping wings. These damping wings bias the 1D Lyman α forest flux power spectrum by causing absorption in quasar spectra away from the location of the absorber itself. We investigate the effect of high column density absorbers on the Lyman α forest using hydrodynamical simulations for the first time. We provide templates as a function of column density and redshift, allowing the flexibility to accurately model residual contamination, i.e. if an analysis selectively clips out the largest damping wings. This flexibility will improve cosmological parameter estimation, for example, allowing more accurate measurement of the shape of the power spectrum, with implications for cosmological models containing massive neutrinos or a running of the spectral index. We provide fitting functions to reproduce these results so that they can be incorporated straightforwardly into a data analysis pipeline.

Key words: large-scale structure of universe – cosmology: theory.

1 INTRODUCTION

The Lyman α forest (a series of neutral hydrogen absorption lines in the spectra of quasars) is a uniquely powerful probe of the clustering of matter at redshifts from about $z = 2$ to 6 (Croft et al. 1998; McDonald et al. 2000, 2005b; Palanque-Delabrouille et al. 2013; Viel et al. 2013; Iršič et al. 2017a) and from sub-Mpc to hundreds of Mpc scales. The one-dimensional (1D) Lyman α forest flux power spectrum (along the line of sight) is particularly sensitive to small-scale clustering in the quasi-linear regime and provides important constraints on extended cosmological models that suppress small-scale power (Seljak et al. 2005; Palanque-Delabrouille et al. 2015; Armengaud et al. 2017; Iršič et al. 2017b,c; Yeche et al. 2017), notably those containing massive neutrinos and warm dark matter. This small-scale information complements the larger scales probed by the angular power spectrum of the cosmic microwave background (CMB). For example, the best upper limit on the sum of neutrino

masses (Palanque-Delabrouille et al. 2015) comes from combining CMB data from the Planck Collaboration (Planck Collaboration XI 2016) with the 1D Lyman α forest power spectrum as measured from Sloan Digital Sky Survey (SDSS)-III/Baryon Oscillation Spectroscopic Survey (BOSS) Data Release 9 (DR9) quasar spectra (Eisenstein et al. 2011; Dawson et al. 2013; Palanque-Delabrouille et al. 2013).

Future surveys like the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016a,b) will further improve constraints on extended cosmological models. Font-Ribera et al. (2014) forecast 1σ errors on a DESI measurement of the sum of neutrino masses to be 0.017 eV.[1] Considering that the lower limit on the sum of neutrino masses from neutrino oscillation experiments is 0.06 eV (Forero, Tórtola & Valle 2014; Gonzalez-Garcia, Maltoni & Schwetz 2014; Esteban et al. 2017), this would constitute at least

---

[1] This is the full forecasted constraint considering a combination of *Planck* CMB data, DESI broad-band galaxy power spectrum, DESI broad-band Lyman α forest flux power spectrum and ∼100 high-resolution Lyman α forest quasar spectra.

★ E-mail: keir.rogers@fysik.su.se
† NASA Einstein Fellow.

a 3σ detection. Furthermore, the 1D Lyman α forest flux power spectrum probes the primordial power spectrum on the smallest currently accessible scales, $k \sim 4\,\mathrm{Mpc}^{-1}$. Including Lyman α forest data will improve constraints on the running of the spectral index (which quantifies deviations from a pure power-law spectrum) by a factor of 2, reaching 1σ errors of ± 0.002 (Font-Ribera et al. 2014). This would provide new insights into early universe physics, potentially ruling out classes of models of inflation. Importantly, it will also provide a unique independent cross-check at small scales of the primordial power spectrum shape inferred from CMB measurements at large scales.

Achieving these limits requires marginalization over the uncertain impact of a number of astrophysical effects on the 1D Lyman α forest power spectrum. In particular, this includes broadened absorption features from high column density absorbers. High column density absorbers are usually classified as either damped Lyman α absorbers (DLAs), with column densities $N(\mathrm{H\,\textsc{i}})$ exceeding $2 \times 10^{20}\,\mathrm{atoms\,cm}^{-2}$ (Wolfe et al. 1986), or Lyman-limit systems (LLSs), which correspond to $2 \times 10^{20}\,\mathrm{atoms\,cm}^{-2} > N(\mathrm{H\,\textsc{i}}) > 1.6 \times 10^{17}\,\mathrm{atoms\,cm}^{-2}$. Both types of system produce broad damping wings, which extend large distances in redshift space. If not accounted for, they will bias cosmological parameter estimation from the Lyman α forest. The systems are formed at peaks of the underlying density distribution; consequently, they cluster more strongly than the forest itself (Font-Ribera et al. 2012).

To remove the bias induced by damped absorbers, one can fit a model for their effect on power spectra. The most widely used approach (McDonald et al. 2005a) is now more than a decade old. Although this model was adequate for the data available at the time, future surveys will be substantially more constraining and therefore demand tighter control over systematics. Furthermore, there have been significant improvements in theoretical modelling of these systems (e.g. Pontzen et al. 2008; Bird et al. 2015). An updated model for the effects of high column density absorbers is therefore both timely and essential in order to achieve the forecasted cosmological constraints from future surveys.

Different column densities correspond to gas at different physical densities, so that simulations suitable for modelling the forest are often not suited to reproducing high column density systems. The Lyman α forest is largely insensitive to the physics of galaxy formation since it is sourced by gas at below mean density; the primary uncertainties arise from cosmological parameters and the thermal history of the intergalactic medium. Conversely, high column density absorbers arise largely from regions within or around galaxies and are thus very sensitive to the physics of galaxy formation and less sensitive to large-scale cosmology. It is consequently essential to model the effect of high column density absorbers using simulations that include detailed galaxy formation physics and can thus reproduce their characteristics and statistics.

In Lyman α forest studies, damping wings are sometimes 'clipped' (i.e. removed or masked) from quasar spectra (e.g. see Lee et al. 2013, for details of the process for BOSS DR9 spectra). However, not all damping wings are identified and many will remain in the spectra, especially in noisier spectra where they are harder to spot and for lower density absorbers (i.e. LLSs) that have narrower wings. Therefore, in the final cosmological parameter estimation from the 1D Lyman α forest power spectrum, the effect of residual high column density absorbers is modelled as a multiplicative scale-dependent bias of the power spectrum with an amplitude (reflecting the level of residual contamination) that is fitted and marginalized (Palanque-Delabrouille et al. 2015).

The functional form of this model (i.e. its scale and redshift dependence) is based on the measurements made in McDonald et al. (2005a).

McDonald et al. (2005a) investigated the effect with lognormal model mock quasar spectra (i.e. generated without hydrodynamical simulations; details of their generation are given in McDonald et al. 2006), since the numerical simulations available at the time were not large enough to generate spectra encompassing the full width of damping wings. They then probe the effect of high column density absorbers on the Lyman α forest by inserting damping wings in mock spectra at the peaks of the lognormal field, based on the observationally determined column density distribution function (CDDF). They find a systematic effect on the observed 1D Lyman α forest power spectrum that is maximized on scales corresponding to the width of a damped system and which has negligible redshift evolution (considering three redshift slices at $z = [2.2, 3.2, 4.2]$). They provide a single template to fit their bias measurement, including the effect of all LLSs and DLAs together. However, as discussed above, in current data analysis pipelines, damping wings are removed from quasar spectra in a way that preferentially removes higher density systems. Therefore, when the template is used in parameter inference, it may not correctly model the bias of the *residual* contamination, which will have a different CDDF to the total – the clipping of the survey spectra changes the survey CDDF. The bias will have a different scale-dependence (not just amplitude), since this is driven by the distribution of the widths of damping wings remaining in quasar spectra.

In this work, we investigate the effect of high column density absorbers on the 1D Lyman α forest power spectrum as a function of their column density and redshift using hydrodynamical simulations of galaxy formation from the Illustris project (Nelson et al. 2015; Vogelsberger et al. 2014a). Comparison to relevant observations has shown that Illustris reproduces the observed CDDF and spatial clustering of high-density systems (Vogelsberger et al. 2014a; Bird et al. 2014, see Section 3.1 for more details) at the 95 per cent confidence level. Spectra are generated from this simulation, then separated into categories according to the maximum column density within each spectrum (see Section 2 for more details). We measure the 1D flux power spectrum of each of these types of spectrum and measure the (multiplicative) bias of each type compared to the power spectrum of the Lyman α forest alone. We make this measurement at multiple redshifts and so probe the redshift evolution of this effect.

We discuss high column density absorbers in more detail in Section 2. In Section 3, our methodology in going from hydrodynamical simulations to measurements of the 1D flux power spectrum is explained. We present our main results in Section 4. These results are discussed in Section 5, and in Section 6, we present the templates that we have fitted to our measurements. Finally, conclusions are drawn in Section 7.

2 DAMPED LYMAN α ABSORBERS AND LYMAN-LIMIT SYSTEMS

High column density absorbers are regions of neutral hydrogen (H I) gas that are above a column density threshold of $N(\mathrm{H\,\textsc{i}}) > 1.6 \times 10^{17}\,\mathrm{atoms\,cm}^{-2}$. By contrast, lower column density absorbers form the Lyman α forest. The absorption lines formed by high column density absorbers are broadened, forming damping wings and hence absorption in the spectrum away from the location of the absorbing gas. The damping wings have a characteristic Voigt profile, which is a convolution of a Gaussian profile (caused by Doppler broadening)

Table 1. The neutral hydrogen (H I) column density limits [$N(\mathrm{H\,I})_{min}$, $N(\mathrm{H\,I})_{max}$] that define the categories of absorbing systems used in this work. The columns on the right show the percentage of spectra (at each redshift z that is considered) in our $(106.5\,\mathrm{Mpc})^3$ simulation box (Vogelsberger et al. 2014a; Nelson et al. 2015, Illustris-1) where the highest-density system belongs to a given category.

Absorber category	$N(\mathrm{H\,I})_{min}$	$N(\mathrm{H\,I})_{max}$	per cent of spectra in $(106.5\,\mathrm{Mpc})^3$ simulation at				
	(atoms cm^{-2})		$z = 2.00$	$z = 2.44$	$z = 3.01$	$z = 3.49$	$z = 4.43$
Lyman α forest	0	1.6×10^{17}	77.7	69.6	57.4	45.7	22.0
LLSs	1.6×10^{17}	1×10^{19}	10.6	14.9	21.8	27.0	36.6
Sub-DLA	1×10^{19}	2×10^{20}	5.9	8.1	11.4	14.3	20.1
Small DLA	2×10^{20}	1×10^{21}	3.1	4.1	5.5	7.8	12.8
Large DLA	1×10^{21}	∞	2.7	3.3	3.9	5.2	8.5

and a Lorentzian profile (caused by natural or collision broadening). The width of these wings in velocity space increases with the column density of the absorbing system. High column density absorbers are then usually classified as either DLAs, whose damping wings are considered significantly broadened and which correspond to $N(\mathrm{H\,I}) > 2 \times 10^{20}$ atoms cm^{-2} (Wolfe et al. 1986), or LLSs, which correspond to column densities in the range $2 \times 10^{20} > N(\mathrm{H\,I}) > 1.6 \times 10^{17}$ atoms cm^{-2}.

In this work, we aim to investigate the effect of high column density absorbers (and especially their damping wings) on the 1D Lyman α forest flux power spectrum, as a function of their column density (and redshift). We therefore use a more refined classification of high column density absorbers based on their column densities, in particular accounting for the fact that higher density LLSs do have wide damping wings. Table 1 shows the column density limits that define our categories, as well as the percentage of simulated spectra (see Section 3.1) where the highest density system is a given type and hence is the main contaminant. The overall percentage of spectra contaminated by high column density absorbers (LLSs, sub-DLAs, small and large DLAs) increases with redshift because the H I CDDF increases at higher densities at higher redshifts, but there are always more LLSs than DLAs.

3 METHOD

We first outline the method we have used and then explain the steps in more detail in the following subsections (Section 3.1 to 3.3).

(1) We use a cosmological hydrodynamical simulation from the Illustris project (Vogelsberger et al. 2014a; Nelson et al. 2015) and generate mock spectra on a grid (562 500 in total, each at a velocity resolution of $25\,\mathrm{km\,s^{-1}}$ and with a typical length of $\simeq 8\,000\,\mathrm{km\,s^{-1}}$). We repeat this for a number of redshift slices at which the Lyman α forest is observed ($z = [2.00, 2.44, 3.01, 3.49, 4.43]$). (See Section 3.1.)

(2) For each redshift slice, we separate the spectra according to the highest column density system within that spectrum using the absorber categories defined in Table 1. For each absorber category (and the total set of spectra), we measure the 1D flux power spectrum (i.e. along the line of sight, integrating over transverse directions) using a fast Fourier transform (FFT). (See Section 3.2.)

(3) We then measure the (multiplicative) bias of the flux power spectra from each category relative to the 1D flux power spectrum of the Lyman α forest, as a function of absorber type (i.e. maximum column density) and redshift (see Section 3.3). We fit parametric models to these bias measurements and provide these templates in Section 6.

3.1 Hydrodynamical simulations and mock spectra

Our main results make use of snapshots from the highest resolution (in terms of both dark matter particles and hydrodynamical cells) cosmological hydrodynamical simulation from the Illustris project (Vogelsberger et al. 2014a; Nelson et al. 2015, Illustris-1[2]). The simulation adopts the following cosmological parameters: $\Omega_m = 0.2726$, $\Omega_\Lambda = 0.7274$, $\Omega_b = 0.0456$, $\sigma_8 = 0.809$, $n_s = 0.963$ and $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, where $h = 0.704$ (Vogelsberger et al. 2014b). The box has a volume in comoving units of $(106.5\,\mathrm{Mpc})^3$, and we consider snapshots at redshifts $z = [2.00, 2.44, 3.01, 3.49, 4.43]$.

The Illustris simulations use the moving mesh code AREPO (Springel 2010). The galaxy formation physics implemented is of relevance to dense regions of neutral hydrogen gas, and therefore we describe it briefly here. The subgrid models include prescriptions for supernova (Springel & Hernquist 2003; Vogelsberger et al. 2013) and active galactic nuclei (AGNs; Springel, Di Matteo & Hernquist 2005; Sijacki et al. 2007) feedback (Bird et al. 2014 showed that the properties of DLAs are quite insensitive to the details of AGN feedback), radiative cooling, star formation, and metal enrichment of gas. Self-shielding is implemented as a correction to the photoionization rate, which is a function of hydrogen density and gas temperature. The potential ionizing effect of local stellar radiation within the most dense absorbers (i.e. large DLAs; e.g. Fumagalli et al. 2011) is neglected. Pontzen et al. (2010) found this effect to be negligible and accurate calculations in any case require physics on parsec scales, well below the resolution of the simulation (it can then be viewed as part of the unresolved physics included in the above feedback prescriptions). More details of these models are given in Vogelsberger et al. (2013) and Bird et al. (2014). Gravitational interactions are computed using the TreePM approach (Springel 2005).

We require that these simulations accurately reproduce the necessary statistics of high column density absorbers that are observed in surveys. As a means of quantifying this, we can first consider the CDDF of neutral hydrogen over relevant column densities ($N(\mathrm{H\,I}) > 1.6 \times 10^{17}$ atoms cm^{-2}). Vogelsberger et al. (2014a) make a comparison of the CDDF as produced by Illustris centred at $z = 3$ to the distribution observed in a number of surveys (Prochaska, O'Meara & Worseck 2010, for LLSs; Zafar et al. 2013, for sub-DLAs; Noterdaeme et al. 2009, for DLAs]. In particular, the distributions are consistent with the feature in the CDDF around the DLA threshold, where the distribution rises, being reproduced well (the results of Bird, Garnett & Ho 2017 from SDSS-III DR12

---

[2] The simulation we use is publically available at http://www.illustris-project.org/data.

spectra are also consistent for DLAs). Bird et al. (2014) showed that the AREPO code with the above hydrodynamical models can produce values of the DLA halo bias (at $z = 2.3$), which are in agreement with measured values from real surveys (Font-Ribera et al. 2012), indicating that the clustering of high column density absorbers is well reproduced. Bird et al. (2015) compared the distribution function of velocity widths of low ionization metal absorbers associated with DLAs as produced by the simulations at $z = 3$ to the distribution observed in Neeleman et al. (2013). The data points are within the 68 per cent confidence interval of the simulated distribution. This suggests that the simulations are reproducing the kinematics, and thus, the host halo distribution, of high column density absorbers. One potential caveat is that these simulations are found to produce too high a total incidence rate of DLAs when compared to observations (Noterdaeme et al. 2012) at $z = 2$ (Bird et al. 2014). However, the overall incidence rate is absorbed into a normalization that must in any case be allowed to float during analysis of clipped spectra (as discussed in Section 5).

For each snapshot, we generate mock spectra containing only the Lyman α absorption line (i.e. with a rest wavelength of 1215.67 Å) from neutral hydrogen. We do this on a square grid of 562 500 spectra, in the plane perpendicular to a direction that we define as the line of sight. Each spectrum extends the full length of the simulation box with periodic boundary conditions, giving a size in velocity (or 'redshift') space of {7111, 7501, 8000, 8420, 9199} km s^{-1}, respectively, at $z = [2.00, 2.44, 3.01, 3.49, 4.43]$.[3] We first measure the optical depth τ in velocity bins of size 25 km s^{-1} along the spectrum.[4] We further convolve our spectra with a Gaussian kernel of FWHM = 8 km s^{-1}, setting the simulated spectrographic resolution. We then calculate the transmitted flux $\mathcal{F} = e^{-\tau}$. In this way, the spectra we have constructed are insensitive to contamination from other absorption (or emission) lines, estimation of the emitted quasar continuum (which here is effectively set to unity) or instrumental noise. In each spectrum pixel, we are also able to measure the column density (integrated along the line of sight in each bin) of neutral hydrogen, which we use in measuring the maximum density systems in each spectrum (Section 3.2).

3.2 1D flux power spectrum

We separate our spectra into the absorber categories (Lyman α forest, LLSs, sub-DLAs, small and large DLAs) defined in Table 1 according to the maximum column density system within each spectrum. We search for the highest column density integrated over any four neighbouring velocity bins; this amounts to a comoving length along the line of sight of {1.50, 1.42, 1.33, 1.27, 1.16} Mpc, respectively, at $z = [2.00, 2.44, 3.01, 3.49, 4.43]$. The categorization is insensitive to the number of neighbouring velocity bins that we use, as the boundaries between categories differ by orders of magnitude in column density. Moreover, the method is efficient in identifying high column density absorbers since they are vastly more dense than the surrounding gas forming the Lyman α forest.[5] We have chosen a length that is much larger than the most extensive DLAs as found by

recent studies (Krogager et al. 2012), and so we are sure to integrate over the full length of any high column density absorbers. Our definition of high column density absorbers includes blends, where a number of smaller, lower column density systems have been added together. In this way, we have associated with each spectrum the most dominant absorbing system and in the case where high column density absorbers are identified, these are the main contamination to the spectrum through their associated damping wings. The percentage of spectra in each absorber category at each redshift slice is given in Table 1.

We measure the 1D flux power spectrum of all the spectra and each absorber category at each redshift slice. The 1D power spectrum $P^{1D}(k_{||}, z)$ is defined as the integral of the 3D power spectrum $P^{3D}(k_{||}, \boldsymbol{k}_\perp, z)$ over directions perpendicular to the line of sight:

$$P^{1D}(k_{||}, z) = \int P^{3D}(k_{||}, \boldsymbol{k}_\perp, z) \frac{\mathrm{d}\boldsymbol{k}_\perp}{(2\pi)^2}, \tag{1}$$

where the wavevector $\boldsymbol{k} = [k_{||}, \boldsymbol{k}_\perp]$ is conjugate to velocities in real space and so is measured in units of inverse velocity (e.g. s km^{-1}). We also use the convention of absorbing the 2π into the conjugate variable.[6]

To measure P^{1D} for an individual line of sight, we first calculate the fluctuation in each velocity $v_{||}$ bin $\delta_{\mathcal{F}}(v_{||}) = \frac{\mathcal{F}(v_{||})}{\langle\mathcal{F}\rangle} - 1$, where $\langle\mathcal{F}\rangle$ is the average flux over all spectra at each redshift (Croft et al. 1998). We calculate the 1D Fourier transform along the line of sight $\hat{\delta}_{\mathcal{F}}(k_{||})$ using a FFT-based method since we have evenly spaced velocity bins. We then estimate the 1D flux power spectrum for each sightline $P^{1D}_{Raw}(k_{||}) = |\hat{\delta}_{\mathcal{F}}(k_{||})|^2$. Finally, we estimate the 1D flux power spectrum in equation (1) for each absorber category i by (e.g. Palanque-Delabrouille et al. 2013)

$$P^{1D}_i(k_{||}, z) = \left\langle \frac{P^{1D}_{Raw}(k_{||}, z)}{W^2(k_{||}, \Delta v, R)} \right\rangle_i, \tag{2}$$

where we explicitly indicate that the raw 1D power spectra depend on redshift z. The average is taken over spectra of a given category (or all spectra for the total power spectrum) at each redshift slice. The window function $W(k_{||}, \Delta v, R)$ that is divided out arises from the binning in velocity space (Δv) and the simulated spectrographic resolution R:

$$W(k_{||}, \Delta v, R) = \exp\left(-\frac{1}{2}(k_{||}R)^2\right) \times \frac{\sin(k_{||}\Delta v/2)}{k_{||}\Delta v/2}, \tag{3}$$

where $\Delta v = 25$ km s^{-1} and $R = 3.40$ km s^{-1} (not to be confused with the spectrographic resolving power; see Section 3.1). We then have an estimate of the 1D flux power spectrum for each absorber category of spectra at each redshift slice.

3.3 Modelling the effect of high column density absorbers

The total 1D flux power spectrum of a set of spectra $P^{1D}_{Total}(k_{||}, z)$ can be expressed as a weighted sum of the 1D flux power spectra calculated in equation (2) for each absorber category i:

$$P^{1D}_{Total}(k_{||}, z) = \sum_i \alpha_i(z) P^{1D}_i(k_{||}, z), \tag{4}$$

where $\alpha_i(z)$ are the fraction of spectra in each absorber category at each redshift (as given in Table 1 for our simulated ensemble of spectra). In a real survey, $\alpha_i(z)$ may change from their raw values due to the attempt to clip (i.e. remove) high column density absorbers

---

[3] We convert the comoving length of the box to a proper velocity by the Hubble law.

[4] For comparison, BOSS DR9 spectra are binned at a velocity resolution of 69.02 km s^{-1} (Lee et al. 2013).

[5] We have explicitly tested the impact of doubling or halving the number of neighbouring velocity bins we use on the 1D flux power spectra we measure in each absorber category. We find that the maximum absolute difference in any power spectrum bin is a negligible 0.2 per cent.

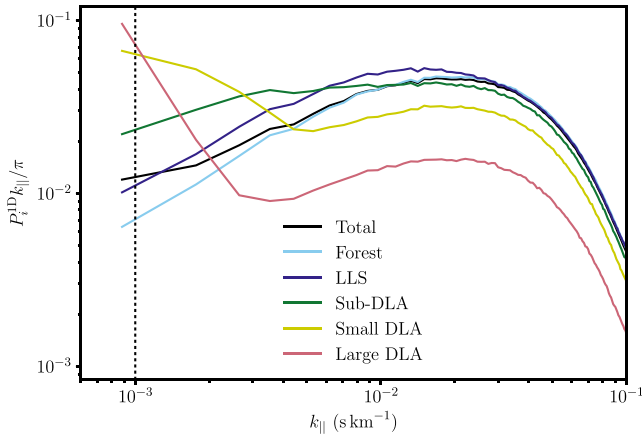[6] I.e. we define the Fourier transform as $\delta(k) = \int \delta(x) e^{-ikx} \mathrm{d}x$.

Figure 1. The 1D flux power spectra of different categories of spectra, as a function of line-of-sight scale $k_{||}$ at redshift $z = 2.00$. The different categories are as follows: the total from our full simulated sample of spectra, spectra containing only Lyman α forest, and spectra contaminated by different types of high column density absorber (LLSs, sub-DLAs, small and large DLAs). The vertical dashed line shows the largest scale probed by the BOSS DR9 1D Lyman α forest flux power spectrum; by comparison, the largest scale probed by our analysis at this redshift is larger at 9×10^{-4} s km^{-1}. The definitions of the different categories of absorber are given in Table 1. (See Section 6 for the full intermediate redshift evolution.)

discussed in Section 1. We can rearrange equation (4) to isolate the 1D flux power spectrum of the Lyman α forest alone:

$$P_{\mathrm{Total}}^{\mathrm{1D}}(k_{||}, z) = P_{\mathrm{Forest}}^{\mathrm{1D}}(k_{||}, z)$$
$$\times \left[\alpha_{\mathrm{Forest}}(z) + \sum_{i \neq \mathrm{Forest}} \alpha_i(z) \frac{P_i^{\mathrm{1D}}(k_{||}, z)}{P_{\mathrm{Forest}}^{\mathrm{1D}}(k_{||}, z)} \right]. \quad (5)$$

In this way, we have isolated the effect of spectra containing high column density absorbers on the 1D flux power spectrum of the Lyman α forest as a multiplicative bias (i.e. the terms in square brackets).[7] This matches the general form of modelling this effect in previous studies, as explained in Palanque-Delabrouille et al. (2015, based on the results in McDonald et al. 2005a), but now additionally probing the bias as a function of column density (i.e. by using the different absorber categories). We discuss in more detail in Section 5 our motivations for using this particular form of the bias (as opposed to, for example, an additive bias). Using the 1D flux power spectra we have calculated in Section 3.2, we are able to measure the fractions in equation (5) ($P_i^{\mathrm{1D}}(k_{||}, z) / P_{\mathrm{Forest}}^{\mathrm{1D}}(k_{||}, z)$), and we present the results in Section 4.

4 RESULTS

Fig. 1 shows the 1D flux power spectra of different subsets of sight-lines that we have measured from our simulations (see Section 3.2 and in particular equation 2) at redshift $z = 2.00$. The different subsets shown are as follows: the total as would be measured if no distinction between different types of spectra was made, spectra containing only Lyman α forest (i.e. the ensemble that is uncontaminated by high column density absorbers), and spectra contaminated

---

[7] We could simplify this form further by asserting the fact that $\sum_i \alpha_i(z) = 1$ to remove the parameter $\alpha_{\mathrm{Forest}}(z)$, but it is useful to keep this form as we explain in Section 6.

by different categories of high column density absorber, as defined in Table 1. We first note that the total 1D flux power spectrum at any redshift slice can be reconstructed as a weighted sum of the other 1D flux power spectra for each absorber category at that redshift (see Section 3.3 and in particular equation 4). The weights are the fraction of spectra in each category (the values we measure for our simulated ensemble are given in Table 1). We can estimate the fractional (1σ) statistical error on each power spectrum data point as $1/\sqrt{N_i}$, where N_i is the number of input modes (i.e. simulated spectra) per data point i. This assumes that data points and input modes are independent. This is largest for the large DLA power spectrum at $z = 2.00$, which has 15 188 input simulated spectra giving an error of 0.81 per cent, and smallest for the forest power spectrum at $z = 2.00$, which has 437 063 input simulated spectra, giving an error of 0.15 per cent. All the other uncertainties for each measured power spectrum range in-between these values and can be computed from Table 1.

The total power spectrum deviates from the Lyman α forest power spectrum at all redshifts, showing there is a bias from contamination of spectra by high column density absorbers. This bias can be deconstructed as a function of column density by looking at the power spectra of different absorber categories. The power spectra of high column density absorbers have a distinctive increase on large scales (small $k_{||}$). This is caused by self-correlations across the width of damping wings, which (as discussed in Section 2) can be modelled by a Voigt profile (a convolution of a Gaussian and a Lorentzian). Therefore, the power spectrum of high column density absorbers (on large scales) is connected to the Fourier transform of a Voigt profile. This increases for higher column density systems since there is more line broadening, and starts on larger scales for higher column density systems since the damping wings are wider. (See Appendix A for more analysis and discussion of the Voigt profile model.) On small scales, all the power spectra converge to a scaled version of the Lyman α forest flux power spectrum. This reflects the fact that contaminated spectra do contain some uncontaminated spectral pixels. The amplitude of the small-scale power spectrum reflects the fraction of spectra that is uncontaminated, increasing for lower column density systems since their damping wings are narrower. There is some sensitivity to the length of our simulated spectra, which primarily manifests in our results as the amplitude of the small-scale residual Lyman α forest power in the contaminated power spectra. This is because longer simulated contaminated spectra would have a larger fraction with residual Lyman α forest. This is discussed further and explicitly modelled such that this effect is removed in Section 6.

Fig. 2 shows 1D flux power spectra as in Fig. 1, but for more of the redshift slices that we consider ($z = [2.00, 2.44, 3.49, 4.43]$), for spectra containing only Lyman α forest and spectra contaminated by large DLAs. The Lyman α forest flux power spectrum has the expected shape, amplitude and redshift evolution, matching observations (e.g. Palanque-Delabrouille et al. 2013), and reflecting the fact that it is an integral of a (biased) matter power spectrum. A peculiarity of the Lyman α forest flux power spectrum is that its amplitude *increases* with redshift (unlike the linear matter power spectrum); this is because neutral hydrogen is more abundant at higher redshift and so there is more absorption in quasar spectra (i.e. the Lyman α forest becomes a more negatively biased tracer of the matter distribution). By contrast, it can be seen that the large-scale correlations associated with the large DLAs are converging to a single point as redshift changes. This reflects the fact that these correlations arise from individual damping wings, which do not evolve with redshift.
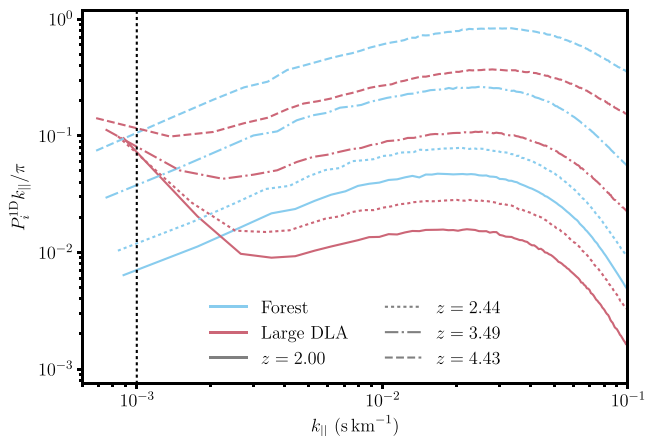
Figure 2. As Fig. 1, but showing more of the redshift slices that we consider (for $z = [2.00, 2.44, 3.49, 4.43]$), for spectra containing only Lyman α forest and spectra contaminated by large DLAs.
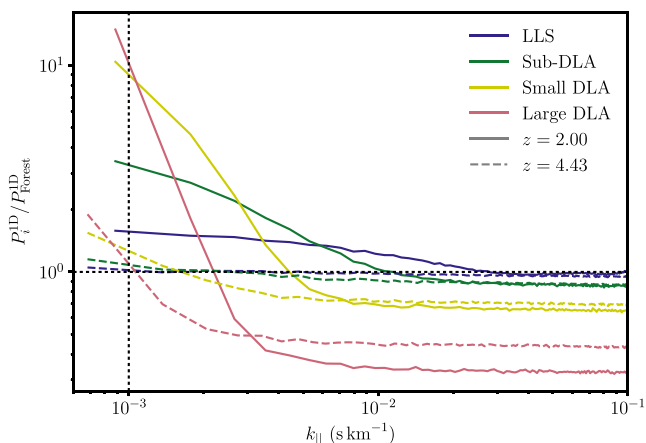

Figure 3. The multiplicative bias of high column density absorbers to the 1D Lyman α forest flux power spectrum, as a function of line-of-sight scale $k_\parallel$ and redshift z, i.e. the ratio of the 1D flux power spectrum of spectra contaminated by high column density absorbers (LLS, sub-DLAs, small and large DLAs) over spectra containing only Lyman α forest. The vertical dashed line shows the largest scale probed by the BOSS DR9 1D Lyman α forest flux power spectrum. The definitions of the different categories of high column density absorber are given in Table 1. The different line styles correspond to different redshift slices, showing the maximum and minimum redshifts that we consider. (See Section 6 for the full intermediate redshift evolution.)

Fig. 3 shows the same 1D flux power spectra as in Figs 1 and 2, but now as ratios between the flux power of spectra contaminated by high column density absorbers and the flux power of spectra containing only Lyman α forest, for $z = 2.00$ and 4.43. These ratios are the quantities to which we fit our templates (see Section 6) as part of our bias model (see Section 3.3). Plotted in this form, it is clear that the large-scale corrections associated with damping wings increase with column density of the damped system. The corrections also decrease with increasing redshift because the Lyman α forest flux power spectrum (on the denominator of the ratio) increases with redshift. On small scales, the ratios converge to a constant value, which reflects the fraction of a line of sight that is uncontaminated (see above). The redshift evolution of this constant value is driven by the transformation from comoving

to velocity space: Spectra are longer in velocity space at higher redshift (despite being drawn from the same comoving length of the simulation). Conversely, the width of damping wings does not change with redshift (for a given column density) because this width just arises from the physical processes within the hydrogen gas (rather than cosmological evolution). Therefore, the fraction of spectra uncontaminated by the damping wings increases with redshift.

5 DISCUSSION

We first discuss and summarize the results we have presented in Section 4. Using our measurements from cosmological hydrodynamical simulations, we have been able to confirm and characterize the effect of high column density absorbers on the 1D Lyman α forest flux power spectrum as a function of column density, scale and redshift. There are distinctive large-scale correlations across the widths of individual damping wings (a 'one-halo' term) arising from high column density absorbers that are seen to bias the 1D flux power spectrum of a set of spectra, relative to the power spectrum of the Lyman α forest alone (Fig. 1). These correlations persist for all the high column density absorbing systems that we identify (i.e. for all column densities $N(\text{H\,I}) > 1.6 \times 10^{17}$ atoms cm^{-2}). Our results can be further understood by relating the shape and amplitude of the large-scale power spectrum of spectra contaminated by high column density absorbers to the Fourier transform of the Voigt profile that is normally used to model damping wings (due to the combination of physical effects that broaden absorption lines; see Appendix A). We find evidence in our simulation results that the 1D flux power spectrum of high column density absorbers does not evolve with redshift (Fig. 2). This reflects the fact that the Voigt profiles of damping wings depend only on column density (i.e. the physical processes within high column density absorbing regions) and not redshift (i.e. cosmological evolution; see equation A1).

The most recent previous investigation into the effect of high column density absorbers on the Lyman α forest was performed by McDonald et al. (2005a, see also Croft et al. 1999; Viel et al. 2004). These authors measured a single bias function for the 1D Lyman α forest flux power spectrum (at each redshift they consider) that includes the combined effect of all high column density absorbers (i.e. all LLSs and DLAs). Our results are qualitatively similar to those of the previous study; however, by investigating different absorber categories based on column density ranges (Table 1), we have shown that the form of the bias as a function of wavenumber depends strongly on column density.

This will have implications for any parameter inference from the 1D flux power spectrum. For instance, the analysis by Palanque-Delabrouille et al. (2015) uses a single multiplicative bias model for the Lyman α forest flux power spectrum based on the results in McDonald et al. (2005a).[8] The model has a free amplitude that is allowed to vary (reflecting the level of contamination in a given survey) and is then marginalized. The shape of this model is therefore based on the observed CDDF of high column density absorbers. However, as discussed in Section 1, in the measurement of the 1D flux power spectrum, high column density absorbers in the quasar spectra are clipped out in the hope of removing noise

---

[8] The model is reported in Palanque-Delabrouille et al. (2015) as $1 - 0.2\,\alpha_{\text{DLA}}\,[1/(15000.0\,k_\parallel - 8.9) + 0.018]$, where α_{DLA} is the free amplitude.

(Lee et al. 2013; Palanque-Delabrouille et al. 2013). This process changes the CDDF of high column density absorbers by preferentially removing higher column density systems that are easier to spot in the noisy spectra. Hence, the shape of the bias from residual high column density absorbers is different (as we have shown in Section 4) and the model used by Palanque-Delabrouille et al. (2013) may not be flexible enough to account for this, especially at the level of accuracy required by future surveys. Our measurements provide a set of templates for the effect of different absorber categories as a function of column density. By using our templates as part of the model in equation (5), it is now possible to more accurately characterize the bias of the residual contamination. We also find evidence for redshift dependence of the fractional effect of high column density absorbers on the forest power spectrum (driven by the changing amplitude of the forest power spectrum), which is also not included in the current model, but is incorporated into our templates. Fits allowing incorporation of our new results into future pipelines are described in Section 6.

We now discuss our motivations for some of the choices made in our analysis. We have chosen to present our main results as the 1D flux power spectra of different sets of simulated spectra, where we have categorized spectra according to the maximum column density system within each spectrum. This means that we are measuring the power spectra of ensembles of spectra that are contaminated to similar extents, rather than the flux power spectra of high column density absorbers alone. Furthermore, a consequence of this categorization is that within the spectra of a given category, there may be high column density absorbers of a lower column density (e.g. there may be LLSs in the large-DLA category of spectra). In the first instance, this does not affect our results because the power spectrum measurements we have made (Section 4) and the templates that we construct (Section 6) include the effect of this possible additional lower column density contamination. A subtlety arises because the amount of additional lower column density contamination will be partly sensitive to the length of simulated spectra, since longer spectra have a greater chance of being contaminated. However, the damping wings of the highest column density systems already produce zero transmitted flux over a significant fraction of the length of our simulation box, so that the presence of possible additional high column density absorbers will make very little difference in any case. We tested this by inserting an LLS into a spectrum contaminated by a large DLA, which reduced the total transmitted flux by 0.07 per cent. By carrying out this insertion test with a 'control' scenario without the additional contamination, we are able to show that this subtlety will have negligible impact on our conclusions and the validity of our templates.

Finally, we comment on the particular form of our preferred model for the bias of high column density absorbers to the 1D Lyman α forest flux power spectrum (as shown in equation 5). We model the bias as a multiplicative correction, rather than, for example, an additive form. First, this matches the form of the currently used model (as shown in Palanque-Delabrouille et al. 2015). Moreover, an additive form would require either the separation of high column density absorbers and the Lyman α forest in the simulated spectra, or a complete physical understanding of how the two components interact at the ensemble level. The former is not trivial for our analysis since we are not inserting high column density absorbers (as previous studies have done), but are simultaneously simulating the low and high column density regions of gas. We avoid the latter due to any remaining physical uncertainties and instead form a parametric multiplicative model based on our simulated results (see Section 6).

6 TEMPLATES FOR THE EFFECT OF HIGH COLUMN DENSITY ABSORBERS

To aid incorporation in future pipelines, we have produced fits to the biases induced by contaminants in our different column density bins. The parametric form of our templates is

$$\frac{P_i^{1D}(k_{||}, z)}{P_{Forest}^{1D}(k_{||}, z)} = \left(\frac{1+z}{1+z_0}\right)^{-3.55} \times \frac{1}{(a(z)e^{b(z)k_{||}} - 1)^2} + c(z), \quad (6)$$

where

$$a(z) = a_0 \left(\frac{1+z}{1+z_0}\right)^{a_1}, \quad b(z) = b_0 \left(\frac{1+z}{1+z_0}\right)^{b_1},$$

$$c(z) = c_0 \left(\frac{1+z}{1+z_0}\right)^{c_1}, \quad (7)$$

and the pivot redshift $z_0 = 2.00$. $[a_0, a_1, b_0, b_1, c_0, c_1]$ are free parameters that we fit simultaneously in $k_{||}$ and z space for each absorber category $i \in \{\text{LLSs, sub-DLA, small DLA, large DLA}\}$. We fit using the Levenberg–Marquardt algorithm (Levenberg 1944; Marquardt 1963).[9]

Fig. 4 shows the result of these fits (dashed lines) with the raw ratios measured from the simulation (solid lines); the corresponding parameter values are given in Table 2. These can be used to reconstruct a final model for the bias of spectra containing high column density absorbers by using equation (5). The model described by equation (6) characterizes the results we have measured in our simulations, and the one described through equation (7) allows interpolation of our results to intermediate redshifts that we have not explicitly probed. (Use of the model outside the limits of redshift and scale we have considered would constitute an extrapolation, but this should not be necessary since our measurements bracket the main redshifts and scales of interest to Lyman α forest studies.) No strong physical meaning should be attached to its terms, although we can motivate the first term on the right-hand side of equation (6) as being the (reciprocal of the) main term of the redshift evolution of $P_{Forest}^{1D}(k_{||}, z)$ as found by Palanque-Delabrouille et al. (2013, using a maximum likelihood estimator). In this way, the parametric form isolates the redshift evolution from $P_{Forest}^{1D}(k_{||}, z)$ and then fits the residual redshift evolution using the terms in equation (7). The best-fitting values of the exponents in equation (7) (as given in Table 2) are small, indicating that most of the redshift evolution can indeed be ascribed to the expected cosmological evolution of $P_{Forest}^{1D}(k_{||}, z)$.

Our results are dependent on the length of our simulated spectra. This manifests in the value of the constant that the ratios $P_i^{1D}(k_{||}, z)/P_{Forest}^{1D}(k_{||}, z)$ have at high $k_{||}$, which is set by the fraction of the length of contaminated spectra that are unaffected by damping wings and contain only Lyman α forest. Since the incidence rates of high column density absorbers are such that one per contaminated spectrum is most likely, a longer spectrum will have a larger fraction that is uncontaminated, causing the constant value at high $k_{||}$ to rise with spectrum length. However, in an analysis of observational data, this will be absorbed into a free parameter. We have used a parametric form for our templates such that all this

---

[9] We were able to further validate our modelling by initially fitting using a subset of the available redshift slices and using this preliminary fit to predict the results at $z = 3.01$. We found the model to accurately predict the results at this intermediate redshift, acting as a form of successful blind test for our model. Our final best-fitting parameters use all available data.
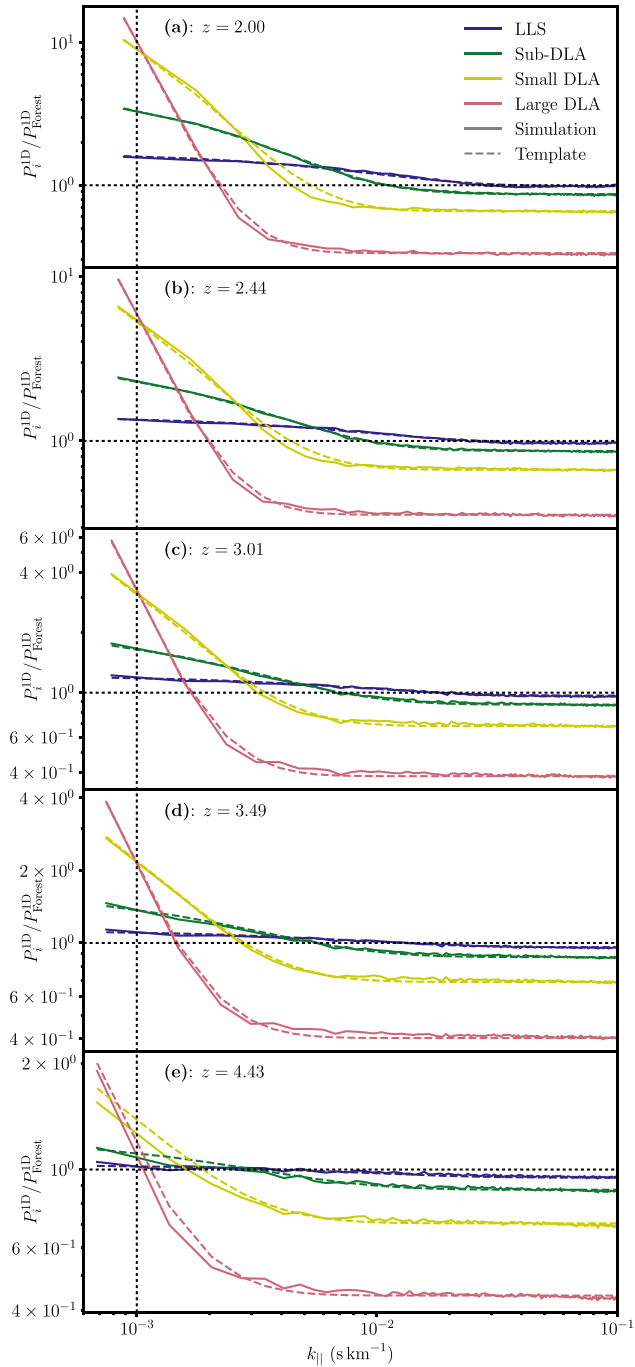
Figure 4. The multiplicative bias of high column density absorbers to the 1D Lyman α forest flux power spectrum, as a function of line-of-sight scale $k_\parallel$ and redshift z, i.e. the ratio of the 1D flux power spectrum of spectra contaminated by high column density absorbers (LLS, sub-DLAs, small and large DLAs) over spectra containing only Lyman α forest. The solid lines are these ratios as measured in the hydrodynamical simulations; the dashed lines are our best-fitting templates to these measurements. The functional form of our templates is given in equation (6) and the best-fit values of the model parameters are given in Table 2. The vertical dashed lines show the largest scale probed by the BOSS DR9 1D Lyman α forest flux power spectrum. The definitions of the different categories of high column density absorber are given in Table 1. *From top to bottom*, we show the templates for simulated results at increasing redshift [(a): $z = 2.00$; (b): $z = 2.44$; (c): $z = 3.01$; (d): $z = 3.49$; (e): $z = 4.43$].

Table 2. Best-fitting values of the parameters in our templates for the bias of spectra contaminated by high column density absorbers on the 1D Lyman α forest flux power spectrum. The template parameters are defined in equations (6) and (7). Values are shown for each high column density absorber category. The definitions of the different categories of high column density absorber are given in Table 1.

Absorber category	Template parameter values					
	a_0	a_1	b_0	b_1	c_0	c_1
LLSs	2.2001	0.0134	36.449	− 0.0674	0.9849	− 0.0631
Sub-DLA	1.5083	0.0994	81.388	− 0.2287	0.8667	0.0196
Small DLA	1.1415	0.0937	162.95	0.0126	0.6572	0.1169
Large DLA	0.8633	0.2943	429.58	− 0.4964	0.3339	0.4653

dependence is measured by the term $c(z)$.[10] By inserting equation (6) into equation (5), it can be seen that the term $c(z)$ is degenerate with $\alpha_{\mathrm{Forest}}(z)$, and hence, these terms can be combined and allowed to vary. It follows that the full parametric form of our model for the effect of high column density absorbers on the 1D Lyman α forest flux power spectrum is

$$
P_{\mathrm{Total}}^{\mathrm{1D}}(k_\parallel, z) = P_{\mathrm{Forest}}^{\mathrm{1D}}(k_\parallel, z) \left[\alpha_0(z) + \sum_{i \neq \mathrm{Forest}} \alpha_i(z) \left(\frac{1+z}{1+z_0} \right)^{-3.55} \right.
$$
$$
\left. \times \frac{1}{(a(z)e^{b(z)k_\parallel} - 1)^2} \right]. \tag{8}
$$

When using this model in inference from the 1D Lyman α forest power spectrum $P_{\mathrm{Forest}}^{\mathrm{1D}}(k_\parallel, z)$, it will be necessary to vary five free parameters α_0 and α_i, where i indexes each high column density absorber category. In this way, the column density, scale and redshift dependence of the effect of high column density absorbers is fully determined by our templates, whereas the relative impact of each absorber category is fitted since this is specific to the survey at hand, as well as the details of any clipping of damping wings that changes the survey CDDF. (See Section 5 for more discussion of these details.) Note that the parameter α_0 is degenerate with factors that rescale the mean flux and could be omitted in an end-to-end analysis.

Fig. 5 compares the model we have constructed to the existing model presented in Palanque-Delabrouille et al. (2015) and based on the results in McDonald et al. (2005a). There is a broad agreement between the existing model and our model for the total contamination of high column density absorbers, although our model is less steep in its scale dependence. We also show our model applied to a possible 'residual' contamination, i.e. under the assumption that all DLAs are identified and clipped out in an analysis, leaving only contamination from LLSs and sub-DLAs (e.g. as assumed by Bautista et al. 2017). The model for this lower column density residual contamination has a shallow scale dependence that the model of McDonald et al. (2005a) is unable to characterize. The use of our more flexible model will avoid potential biases due to mischaracterization of the scale dependence of the residual contamination, thus improving estimation of cosmological effects such as massive neutrinos or the tilt of the primordial power spectrum.

We now discuss the prior probability distributions that can be adopted for $\alpha_i(z)$ in any inference using the model we have presented. The $\alpha_i(z)$ are technically not independent parameters, but

---

[10] It can then be understood why we do not factor out the redshift evolution of $P_{\mathrm{Forest}}^{\mathrm{1D}}(k_\parallel, z)$, as we do for the first term on the right-hand side of equation (6).
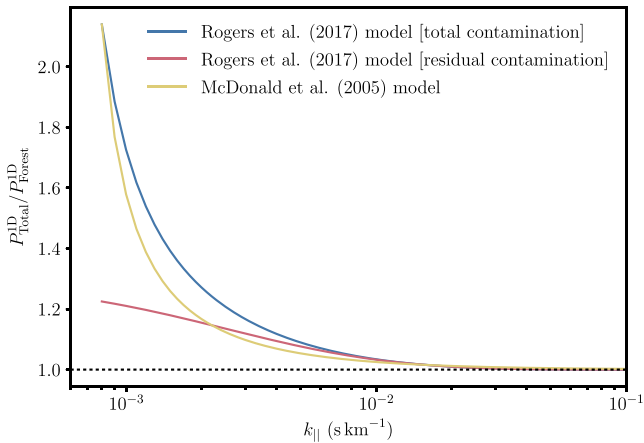
Figure 5. A comparison of the existing multiplicative bias model (McDonald et al. 2005a; Palanque-Delabrouille et al. 2015) for the effect of high column density absorbers on the 1D Lyman α forest power spectrum and the model constructed in this paper using our results from hydrodynamical simulations. For our model, we show an example weighting of the different absorber categories for the full contamination from high column density absorbers on our simulated ensemble of spectra; and an example based on a possible "residual" contamination after the clipping of DLAs (i.e. only LLSs and sub-DLAs remaining). For comparison, the model of McDonald et al. (2005a) is rescaled to have the same amplitude on the largest and smallest scales considered.

are each related to integrals of the H I CDDF for a particular survey over the appropriate column density ranges (and absorption distance per sightline). The effect of spectrum clipping, which changes the survey CDDF, can be modelled by applying a weighting function to the CDDF, which down-weights higher column densities, which are easier to spot and remove. If one wanted to reduce the dimensionality of these nuisance parameters, in particular in redshift space, they could be replaced by a parametrization that quantifies deviations from the expected redshift evolution of the CDDF with only one or two parameters (rather than a parameter for each redshift bin considered). We leave the details of the construction of prior distributions to individual analyses, since the precise considerations will be survey-specific.

To conclude this section, we present a summary of the steps required to incorporate our final model for the effect of high column density absorbers into future 1D Lyman α forest analyses:

(i) Our model describes the effect of quasar spectra contaminated by high column density absorbers as a multiplicative bias to the 1D Lyman α forest flux power spectrum, as given by equation (8). It can therefore be incorporated into a pipeline at the stage of flux power spectrum interpretation to marginalize over effects of these absorbers.

(ii) The free parameters are $\alpha_i(z)$, where i indexes different categories of high column density absorber (as given in Table 1). Our model is of use to any Lyman α forest survey that contains spectra that may be contaminated by high column density absorbers (both LLSs and DLAs). The relative impacts of different categories of high column density absorbers will be determined in the estimation of posterior distributions of these nuisance parameters. Although normalization is necessarily floating, the model fully specifies the scale, column density and redshift dependence of the effect of high column density absorbers, using the results we have measured from hydrodynamical simulations.

(iii) In a survey that does not clip its quasar spectra, strong priors can be given for the free parameters of our model, based on the expected or measured H I CDDF.

(iv) In a survey that does clip its quasar spectra in an attempt to remove high column density absorbers (and therefore changes the survey CDDF), strong priors can still be given for our model parameters, assuming a model can be constructed for the effect of the clipping process on the CDDF. This will constitute some re-weighting of the CDDF.

(v) In order to reduce the dimensionality of our nuisance parameters, rather than having a separate parameter for each redshift bin in a given analysis, one could parametrize the redshift evolution by a simple deviation from the CDDF with only one or two numbers.

7 CONCLUSIONS

We have used a cosmological hydrodynamical simulation (Illustris; Vogelsberger et al. 2014a; Nelson et al. 2015) to investigate the effect of high column density absorbing systems of neutral hydrogen and their associated damping wings on the 1D Lyman α forest flux power spectrum. We find that the effect of high column density absorbers on the Lyman α forest flux power spectrum is a strong function of column density. Accounting for this change in scale-dependence with column density will remove a source of bias in cosmological inference from the Lyman α forest. Previous models (Palanque-Delabrouille et al. 2015) combine the effect of all high column density absorbers together (i.e. all neutral hydrogen column densities $N(\mathrm{H\,I}) > 1.6 \times 10^{17}$ atoms cm^{-2}) based on the CDDF in the raw spectra (McDonald et al. 2006). However, the damping wings of some high column density absorbers are clipped out in the final analysis (Lee et al. 2013), which preferentially removes higher density systems (because they are easier to spot) and changes the column density distribution in the residual contamination. Our results apply for both clipped and unclipped survey spectra, since we separately model the effect for different column densities of the dominant absorber, allowing us to accurately account for the contamination in the 1D flux power spectrum. We discuss in Section 6 the practicalities of employing our model in future analyses.

The shape and amplitude of the distortions in the power spectrum due to a damped absorber depend on its column density because they are driven by the width of the damping wings; i.e. the dominant effect is a 'one-halo' term. We defer investigation of potential 'two-halo' terms to future work, where we measure the effect of high column density absorbers on the 3D Lyman α forest flux power spectrum.

We anticipate that our model will help realize forecasted cosmological constraints from upcoming surveys like DESI. For example, Font-Ribera et al. (2014) forecast that DESI will have the constraining power to make a $\sim$3σ detection of the sum of neutrino masses (in combination with *Planck* CMB data); and they show the power of the 1D Lyman α forest power spectrum in probing the primordial power spectrum, e.g. halving the 1σ error on the running of the spectral index, with implications for inflationary models. It will be necessary to use the models we have presented here, alongside carefully constructed priors on the residual CDDF, to remove degeneracies between the effect of high column density absorbers and cosmological effects.

REFERENCES

Armengaud E., Palanque-Delabrouille N., Yèche C., Marsh D. J. E., Baur J., 2017, MNRAS, 471, 4606

Bautista J. E. et al., 2017, A&A, 603, A12

Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313

Bird S., Haehnelt M., Neeleman M., Genel S., Vogelsberger M., Hernquist L., 2015, MNRAS, 447, 1834

Bird S., Garnett R., Ho S., 2017, MNRAS, 466, 2111

Croft R. A. C., Weinberg D. H., Katz N., Hernquist L., 1998, ApJ, 495, 44

Croft R. A. C., Weinberg D. H., Pettini M., Hernquist L., Katz N., 1999, ApJ, 520, 1

Dawson K. S. et al., 2013, AJ, 145, 10

DESI Collaboration et al., 2016a, preprint (arXiv:e-prints)

DESI Collaboration et al., 2016b, preprint (arXiv:e-prints)

Eisenstein D. J. et al., 2011, AJ, 142, 72

Esteban I., Gonzalez-Garcia M. C., Maltoni M., Martinez-Soler I., Schwetz T., 2017, J. High Energy Phys., 1, 87

Font-Ribera A. et al., 2012, J. Cosmol. Astropart. Phys., 11, 059

Font-Ribera A., McDonald P., Mostek N., Reid B. A., Seo H.-J., Slosar A., 2014, J. Cosmol. Astropart. Phys., 5, 023

Forero D. V., Tórtola M., Valle J. W. F., 2014, Phys. Rev. D, 90, 093006

Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, MNRAS, 418, 1796

Gonzalez-Garcia M. C., Maltoni M., Schwetz T., 2014, J. High Energy Phys., 11, 52

Humlicek J., 1979, J. Quant. Spectrosc. Radiat. Transfer, 21, 309

Iršič V. et al., 2017a, MNRAS, 466, 4332

Iršič V., Viel M., Haehnelt M. G., Bolton J. S., Becker G. D., 2017b, Phys. Rev. Lett., 119, 1302

Iršič V. et al., 2017c, Phys. Rev. Lett., 96, 3522

Krogager J.-K., Fynbo J. P. U., Møller P., Ledoux C., Noterdaeme P., Christensen L., Milvang-Jensen B., Sparre M., 2012, MNRAS, 424, L1

Lee K.-G. et al., 2013, AJ, 145, 69

Levenberg K., 1944, Q. J. Appl. Math., II, 164

Marquardt D. W., 1963, J. Soc. Ind. Appl. Math., 11, 431

McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., Ostriker J. P., 2000, ApJ, 543, 1

McDonald P., Seljak U., Cen R., Bode P., Ostriker J. P., 2005a, MNRAS, 360, 1471

McDonald P. et al., 2005b, ApJ, 635, 761

McDonald P. et al., 2006, ApJS, 163, 80

Morton D. C., 2004, Astrophys. J. Suppl. Ser., 151, 403

Neeleman M., Wolfe A. M., Prochaska J. X., Rafelski M., 2013, ApJ, 769, 54

Nelson D. et al., 2015, Astron. Comput., 13, 12

Noterdaeme P., Petitjean P., Ledoux C., Srianand R., 2009, A&A, 505, 1087

Noterdaeme P. et al., 2012, A&A, 547, L1

Palanque-Delabrouille N. et al., 2013, A&A, 559, A85

Palanque-Delabrouille N. et al., 2015, J. Cosmol. Astropart. Phys., 11, 011

Planck Collaboration XI, 2016, A&A, 594, A11

Pontzen A. et al., 2008, MNRAS, 390, 1349

Pontzen A. et al., 2010, MNRAS, 402, 1523

Prochaska J. X., O'Meara J. M., Worseck G., 2010, ApJ, 718, 392

Seljak U. et al., 2005, Phys. Rev. D, 71, 103515

Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010, MNRAS, 401, 791

Springel V., Hernquist L., 2003, MNRAS, 339, 289

Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

Viel M., Haehnelt M. G., Carswell R. F., Kim T.-S., 2004, MNRAS, 349, L33

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502

Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031

Vogelsberger M. et al., 2014a, Nature, 509, 177

Vogelsberger M. et al., 2014b, MNRAS, 444, 1518

Wolfe A. M., Turnshek D. A., Smith H. E., Cohen R. D., 1986, ApJS, 61, 249

Yeche C., Palanque-Delabrouille N., Baur J., du Mas des BourBoux H., 2017, JCAP, 06, 047

Zafar T., Péroux C., Popping A., Milliard B., Deharveng J.-M., Frank S., 2013, A&A, 556, A141

APPENDIX A: ONE-DIMENSIONAL FLUX POWER SPECTRUM OF A VOIGT PROFILE

As discussed in Section 2, the broadened absorption lines of high column density absorbers are usually modelled by a Voigt profile. A Voigt profile is a convolution of a Lorentzian profile and a Gaussian profile. It therefore appropriately models the combination of the main physical processes that broaden atomic transition lines: the Lorentzian profile from , for example, natural or collisional broadening and the Gaussian profile from ,for example, Doppler broadening. The optical depth as a function of wavelength $\tau(\lambda)$ is the product of the line-of-sight column density N and the atomic absorption coefficient $\alpha(\lambda)$ (e.g. Humlicek 1979):[11]

$$\tau(\lambda) = N\alpha(\lambda) = N \frac{\sqrt{\pi}e^2}{4\pi\epsilon_0 m_{\mathrm{e}}c^2} \frac{f\lambda_{\mathrm{t}}^2}{\Delta\lambda_{\mathrm{D}}} u(x, y), \tag{A1}$$

where the fundamental physical constants have their usual meaning, f is the oscillator strength of the atomic transition, λ_{t} is the transition wavelength and the Doppler wavelength 'shift' associated with a gas of temperature T for an ion of mass m_{ion},

$$\Delta\lambda_{\mathrm{D}} = \frac{\lambda_{\mathrm{t}}}{c} \left(\frac{2k_{\mathrm{B}}T}{m_{\mathrm{ion}}} \right)^{\frac{1}{2}}. \tag{A2}$$

$u(x, y)$ is an unnormalized form of the Voigt function (the normalization is already expressed in the pre-factors of equation A1), specifically the real part of the Faddeeva function:

$$w(z) = \mathrm{e}^{-z^2}\mathrm{erfc}(-iz) = u(x, y) + iv(x, y), \tag{A3}$$

where $\mathrm{erfc}(x)$ is the complementary error function and $z = x + iy$. x and y are, respectively, the wavelength difference from the line centre λ_{c} and the natural width of the transition, in units of the Doppler shift:

$$x(\lambda) = \frac{\lambda - \lambda_{\mathrm{c}}}{\Delta\lambda_{\mathrm{D}}}; \quad y = \frac{\Gamma\lambda_{\mathrm{t}}^2}{4\pi c} \frac{1}{\Delta\lambda_{\mathrm{D}}}, \tag{A4}$$

where Γ is the damping constant of the transition, i.e. the inverse of the time-scale for the electron to remain in the upper level of the transition in the vacuum. For the Lyman α transition, $f = 0.4164$,
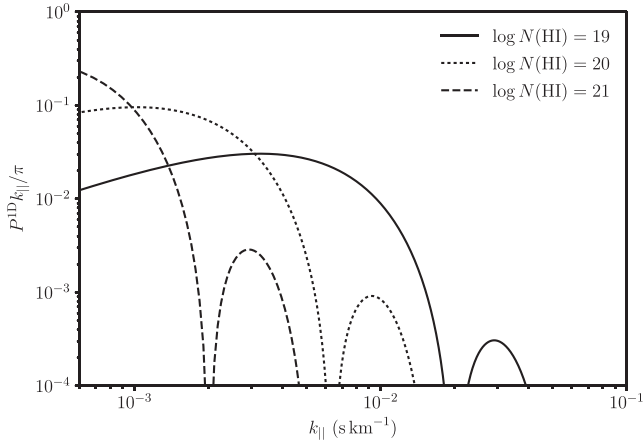
---

[11] Equation (A1) is valid in SI units.

Figure A1. The 1D flux power spectra of Voigt profiles of broadened Lyman α absorption lines as generated by different column densities of neutral hydrogen $N(\mathrm{H\,I})$, as a function of line-of-sight scale $k_{||}$ (the units of $N(\mathrm{H\,I})$ are atoms cm^{-2}).

$\lambda_t = 1215.67$Å, $m_{\mathrm{ion}} = m_{\mathrm{proton}}$ and $\Gamma = 6.265 \times 10^8$ Hz (Morton 2004). For the column densities that we consider, we assume a gas temperature $T \approx 10^4$ K. In order to calculate the 1D flux power spectrum arising from these Voigt profiles, the same procedure is followed as in Section 3.2, i.e. we form flux spectra and carry out a Fourier transform. We transform from wavelengths to velocities by $\Delta v/c = \Delta \lambda / \lambda$.

Fig. A1 shows the 1D flux power spectra of Voigt profiles as given by equation (A1) for the Lyman α absorption line for three different column densities of neutral hydrogen $N(\mathrm{H\,I}) = [10^{19}, 10^{20}, 10^{21}]$ atoms cm^{-2}, spanning the column densities for LLSs and DLAs. This figure should be compared with Fig. 1 in Section 4, which shows the 1D flux power spectra we have measured in the hydrodynamical simulations. The trends in Fig. A1 broadly support the arguments made in Section 5, relating the large-scale power spectrum of simulated spectra contaminated by high column density absorbers to the power spectrum of relevant Voigt profiles. The shape of the large-scale power spectrum of the Voigt profiles is similar in amplitude and scale-dependence as the excesses on large scales for the 1D flux power spectra of simulated spectra in high column density absorber categories. Moreover, these excesses get steeper, increase in amplitude and become prominent on larger scales for higher column densities, both in the simulated and analytic spectra. This reflects the fact that a higher column density means wider damping wings and so correlations on larger scales. In the analytic power spectra in Fig. A1, we observe oscillations in the power spectrum on smaller scales that rapidly decrease in amplitude. These are not observed in the fully simulated power spectra since the oscillations are orders of magnitude lower in amplitude than the flux power spectrum of residual Lyman α forest (see Fig. 1). Furthermore, in our results, we are effectively averaging over a number of column densities in each column density bin (or absorber category) that we consider; this will have the additional effect of averaging out these smaller scale oscillations in the power spectrum to form a smoother scaling.

This paper has been typeset from a TeX/LaTeX file prepared by the author.