Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

LyaCoLoRe: synthetic datasets for current and future Lyman-α forest BAO surveys

James Farr,[a,1,2] **Andreu Font-Ribera,**[a]
Hélion du Mas des Bourboux,[b] **Andrea Muñoz-Gutiérrez,**[c]
F. Javier Sánchez,[d,e,3] **Andrew Pontzen,**[a]
Alma Xochitl González-Morales,[f,g] **David Alonso,**[h] **David Brooks,**[a]
Peter Doel,[a] **Thomas Etourneau,**[i] **Julien Guy,**[j] **Jean-Marc Le Goff,**[i]
Axel de la Macorra,[c] **Nathalie Palanque-Delabrouille,**[i]
Ignasi Pérez-Ràfols,[k] **James Rich,**[i] **Anže Slosar,**[l] **Gregory Tarle,**[m]
Duan Yutong[n] **and Kai Zhang**[j]

[a]University College London, Gower Street, London, WC1E 6BT, U.K.
[b]Department of Physics and Astronomy, University of Utah,
115 S. 1400 E., Salt Lake City, UT 84112, U.S.A.
[c]Instituto de Física, Universidad Nacional Autónoma de México,
A.P. 70-264, 04510, México D.F., México
[d]Department of Physics and Astronomy, University of California,
Irvine, CA 92697, U.S.A.
[e]Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL, U.S.A.
[f]Departamento de Física, DCI, Campus Léon, Universidad de Guanajuato,
37150, Léon, Guanajuato, México
[g]Consejo Nacional de Ciencia y Tecnología,
Av. Insurgentes Sur 1582. Colonia Crédito Constructor,
Del. Benito Juárez, C.P. 03940, México D.F., México
[h]Department of Physics, University of Oxford, Keble Road, Oxford, OX1 3RH, U.K.
[i]IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
[j]Lawrence Berkeley National Laboratory,
1 Cyclotron Road, Berkeley, CA 94720, U.S.A.
[k]Sorbonne Université, CNRS/IN2P3,
Laboratoire de Physique Nucléaire et de Hautes Energies, LPNHE,
4 Place Jussieu, F-75252 Paris, France
[l]Brookhaven National Laboratory, 2 Center Road, Upton, NY 11973, U.S.A.

---

[1] Corresponding author.
[2] https://orcid.org/0000-0002-9817-533X.
[3] https://orcid.org/0000-0003-3136-9532.

https://doi.org/10.1088/1475-7516/2020/03/068

[m]Physics Department, University of Michigan Ann Arbor, MI 48109, U.S.A.
[n]Physics Department, Boston University,
590 Commonwealth Avenue, Boston, MA 02215, U.S.A.

E-mail: james.farr.17@ucl.ac.uk

Received December 18, 2019
Revised March 2, 2020
Accepted March 2, 2020
Published March 31, 2020

Abstract. The statistical power of Lyman-α forest Baryon Acoustic Oscillation (BAO) measurements is set to increase significantly in the coming years as new instruments such as the Dark Energy Spectroscopic Instrument deliver progressively more constraining data. Generating mock datasets for such measurements will be important for validating analysis pipelines and evaluating the effects of systematics. With such studies in mind, we present `LyaCoLoRe`: a package for producing synthetic Lyman-α forest survey datasets for BAO analyses. `LyaCoLoRe` transforms initial Gaussian random field skewers into skewers of transmitted flux fraction via a number of fast approximations. In this work we explain the methods of producing mock datasets used in `LyaCoLoRe`, and then measure correlation functions on a suite of realisations of such data. We demonstrate that we are able to recover the correct BAO signal, as well as large-scale bias parameters similar to literature values. Finally, we briefly describe methods to add further astrophysical effects to our skewers — high column density systems and metal absorbers — which act as potential complications for BAO analyses.

Keywords: baryon acoustic oscillations, dark energy experiments, Lyman alpha forest, redshift surveys

ArXiv ePrint: 1912.02763

Contents

1 Introduction

Our understanding of the expansion history of the Universe has progressed enormously over the last quarter of a century. The discovery of accelerating expansion from the "standardisable candles" of supernovae [1, 2] brought the idea of dark energy to the fore, and it is now considered a vital component of the cosmic inventory. Indeed, efforts to improve our measurements of its properties are at the forefront of current cosmological research, and it is a primary motivation behind a number of surveys past, present and future.

Several of these surveys have focussed on using the "standard ruler" of Baryon Acoustic Oscillations (BAO) [3] in their efforts to understand the Universe's expansion. This fixed-scale imprint on structure formation was first measured from the correlation function [4] and power spectrum [5] of galaxy samples from the Sloan Digital Sky Survey (SDSS) and 2dF Galaxy Redshift Survey respectively. A number of similar measurements have been made in subsequent years, focussing on using galaxies [e.g. 6–9] and quasars (QSOs) [e.g. 10] as tracers of the matter density. These tracers cover redshift ranges $z \sim 0.1 - 1.0$ and $z \sim 1.2 - 1.7$ respectively.

An alternative tracer exists in the form of the Lyman-α (Lyα) forest: a sequence of absorption features that appears in the spectra of high-z QSOs as a result of Lyα absorption of light in the neutral hydrogen gas between QSO and observer. These spectral features thus trace the density of neutral hydrogen gas in the inter-galactic medium (IGM) along the line of sight [11]. Indeed, analytical models developed during the 1990s showed that the Lyα forest absorption closely traces the distribution of dark matter on scales larger than the Jeans length [e.g. 12–14]. The Lyα forest should, then, provide a suitable means to extend measurements of cosmic expansion via BAO to earlier in the Universe's history. Measuring such a signal was first discussed in [15], while the 3D correlation of flux transmission was first studied in [16]. The BAO signal was first detected from measurements of the Lyα auto-correlation using data from data release 9 (DR9) of the Baryon Oscillation Spectroscopic Survey (BOSS) of SDSS-III [17–19], with subsequent improvements in DR11 [20] and DR12 [21], as well as DR14 of the extended Baryon Oscillation Spectroscopic Survey (eBOSS) [22]. The cross-correlation between the Lyα forest and QSOs was first measured in BOSS DR9 [23], with the first detection of BAO coming in DR11 [24], and improvements made in DR12 [25] and eBOSS DR14 [26].

The upcoming Dark Energy Spectroscopic Instrument (DESI) [27] will be able to advance these measurements greatly. Over the 5 years of its operation, it will measure approximately 800,000 QSO spectra with $z > 2.0$, 3 times as many as in the final eBOSS dataset (approximately 270,000). Ahead of such an increase in statistical power, it is vital to be able to sufficiently test analysis pipelines to ensure that they do not introduce any biases. Equally, it is important to be able to quantify exactly how secondary astrophysical effects will impact upon BAO measurements. The best way to carry out both of these tests is through the development of mock datasets [e.g. 28–30] — synthetic realisations of a survey for which cosmological and astrophysical parameters can be easily controlled. Producing such datasets must be computationally inexpensive in order to allow for generation of a large number of realisations, but the data must also provide realistic representations of the survey itself.

In this work, we introduce a package designed to produce mock datasets for current and future Lyα forest BAO analyses, `LyaCoLoRe`. In section 2, we describe the methods used to generate such datasets, including the use of a Gaussian random field to generate the 3D correlations and the subsequent post-processing to yield realistic skewers of transmitted flux fraction. The methods to determine the optimal values of parameters used in these transformations are detailed in section 3. We then verify that the datasets are able to fulfil their purpose for BAO analyses in section 4, measuring correlation functions in the same way as recent analyses from BOSS and eBOSS. In section 5, we introduce and briefly test additional astrophysical effects that `LyaCoLoRe` is able to include, before summarising and concluding in section 6.

2 Making the mocks

The requirement of mocks to be computationally inexpensive but also large in volume prohibits the use of hydrodynamical or full N-body simulations in their construction. Instead, Gaussian random field methods can be used to generate a linear density field in a large box. This method does not capture non-linear evolution, generating data based solely on an initial power spectrum, but is orders of magnitude faster than state of the art simulations. From an initial Gaussian field, a number of options are available to model the physical density. Most straightforwardly, using a lognormal approximation provides a semi-analytic and phys-

JCAP03(2020)068

ically plausible ($\rho > 0$) model of the density field, but breaks down beyond weakly non-linear scales [31, 32]. Formalisms such as Lagrangian perturbation theory [see 33, for a review] can help extend to mildly non-linear scales, while COLA [34] methods subsequently use a small number of N-body code timesteps to further improve modelling of non-linearities. The choice of density approximation method depends on the scales of interest and the computational budget for the task at hand. The presence of non-linear structure is not of vital importance to BAO measurements, particularly at $z \geqslant 2$ where the Lyα forest is observed [19]. As such, Gaussian random field methods are well suited to the production of Lyα BAO mock datasets, and using a lognormal approximation for the physical density is sufficient for current analyses [28]. Studies that are more heavily dependent on accurate reproduction of small-scale effects may require alternative techniques to be used. Having generated a physical density field, tracers such as QSOs can be placed at its peaks via Poisson sampling according to an input bias and number density, and line-of-sight skewers can be drawn by interpolating within the box.

Converting density skewers to mimic the transmitted flux fraction of the Lyα forest then requires a significant degree of post-processing. Despite the speed of Gaussian random field methods, resolution higher than $O(1)$ Mpc/h is not possible within the computational bounds of mock production due to memory limitations. As a result, the 1D power spectrum of the skewers $P_{\rm 1D}(k_\parallel)$ — the power spectrum measured only from modes lying along the line of sight of each skewer — is greatly suppressed. This subsequently affects the errors on our BAO measurements, as the 3D flux power spectrum of the Lyα forest has a significant contribution to its error that is proportional to the 1D power spectrum, known as aliasing noise [15]. As such, we must boost the 1D power spectrum by the addition of small-scale fluctuations in order to ensure that our BAO errors behave correctly. Further, we must convert from density to optical depth at each point of each skewer. The details of this relationship are complex, but in the context of Gaussian random field mocks we are constrained to using a simple approximation such as the fluctuating Gunn-Peterson approximation (FGPA) [35]. Methods such as those discussed in [36] offer improved physical intuition behind the structure of the Lyα forest, but developing techniques to apply them in the context of full-survey mock datasets is beyond the scope of this work. Finally, we must add redshift-space distortions to our skewers. These distortions occur as a result of peculiar velocities in the IGM, and we observe them as an anisotropy in measurements of power spectra and correlation functions.

In this work, we use `CoLoRe` [37] to generate our initial Gaussian skewers, as described in section 2.1. We then present the package `LyaCoLoRe`, which is able to convert `CoLoRe`'s output into realistic skewers of transmitted flux fraction. The methods used in this transformation are described in section 2.2. Finally, in section 2.3, we discuss the computational requirements of running both of these packages. The output skewers from `LyaCoLoRe` then require the addition of instrumental noise and combination with a QSO continuum before they can be considered realistic spectra. This can be carried out in the context of DESI by a package called `desisim`,[1] which is not discussed in this work.

2.1 `CoLoRe`: cosmological lognormal realisations

The `LyaCoLoRe` mocks originate from a program called `CoLoRe`,[2] a highly parallelised code initially designed to produce large catalogues of multiple tracers with the same underlying density field [37]. In this work, we use `CoLoRe`'s lognormal density model for speed, though

---

[1] Publicly available at https://github.com/desihub/desisim.

[2] Publicly available at https://github.com/damonge/CoLoRe.

first and second order Lagrangian perturbation theory methods are also available. Making use of these functionalities would constitute a natural extension of this work, and efforts are ongoing to do so.

From this density field, `CoLoRe` can produce a number of observables such as cosmic shear, intensity maps, CMB lensing and integrated Sachs-Wolfe maps. Most importantly in the context of this work, it is also able to draw line-of-sight skewers from each object to a central observer, interpolating the Gaussian field at intermediate points. This final functionality makes `CoLoRe` well suited for Lyα forest mocks. The basic steps that `CoLoRe` takes in computing such skewers are outlined in the 5-stage process below:

1. Generate a Gaussian random field δ_C at $z = 0$ in a Cartesian box according to an input power spectrum.

2. Compute a corresponding radial velocity in each cell using the gradient of the Newtonian gravitational potential ϕ:

$$v_r(z=0) = -\frac{2f_0}{3H_0^2\Omega_M}(\mathbf{e}_r \cdot \nabla)\phi(z=0), \tag{2.1}$$

   where f_0 is the logarithmic growth rate at $z = 0$, H_0 is the Hubble constant, Ω_M is the matter density parameter, and $\mathbf{e}_r$ is the radial unit vector.

3. Calculate the redshift of each cell (taking the centre of the box as the observer) using a given input cosmology, and de-evolve the fields to that redshift using the corresponding linear growth factor.

4. Carry out a lognormal transformation of the Gaussian field, and Poisson sample it using an input number density $n(z)$ and bias $b(z)$ to obtain a set of sources (QSOs in our case).

5. Compute line-of-sight skewers from each source to the centre of the box by interpolating the initial Gaussian field and the radial velocity field.

The final output from `CoLoRe` is a set of QSOs and corresponding Gaussian field skewers, as well as values of cosmological variables along the skewers. The QSOs have the correct 3D clustering properties on large scales, as demonstrated briefly in appendix A and in more detail in [37]. The skewers also have the correct 3D correlations, as demonstrated in section 4.

2.2 `LyaCoLoRe`

While `CoLoRe` is able to produce skewers with 3D, large-scale correlations matching a given input in a short timeframe, its “raw” output requires significant post-processing before it can be considered a realistic representation of the Lyα forest. To implement these stages of processing, we have developed a `Python` module under the name `LyaCoLoRe`.[3] This code transforms `CoLoRe`’s output into realistic skewers of transmitted flux fraction. The following sections describe the key methods that `LyaCoLoRe` uses to do so, with each step represented visually in figure 1.

---

[3] Publicly available at https://github.com/igmhub/LyaCoLoRe.

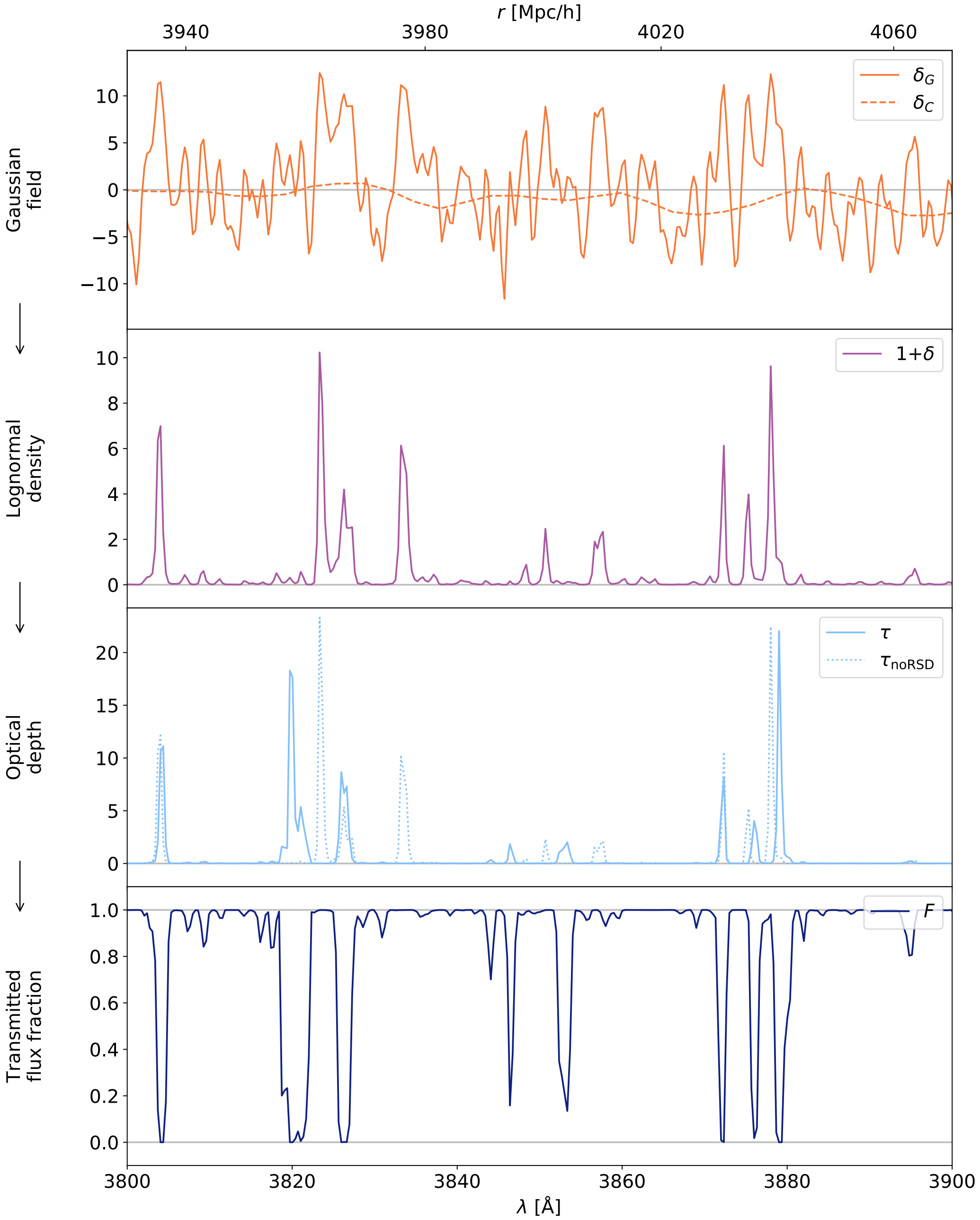

Figure 1. A sample skewer shown at the different stages of transformation from "raw" Gaussian `CoLoRe` output to a final `LyaCoLoRe` flux skewer. The top panel shows the addition of small-scale power to the skewer as described in section 2.2.1, converting δ_C to δ_G. The transition to the second panel shows the lognormal transformation from section 2.2.2, and moving to the dotted line of the third panel shows the fluctuating Gunn-Peterson approximation (FGPA) transformation from the same section. The application of redshift-space distortions (RSDs), as described in section 2.2.3, shifts the dotted line to the solid line in this third panel. The final transformation from optical depth to flux, as described in section 2.2.4, maps the third to the bottom panel. Here, the Hubble flow is used to map distances (top horizontal axis) to observed wavelengths (bottom horizontal axis).

2.2.1 Adding small-scale power

In order that the memory requirements of running `CoLoRe` do not become overwhelmingly large, we are limited to using a grid of 4096^3 cells. Requiring that this encloses the volume of a full Lyα survey limits us to using a low-resolution grid, with cells in `CoLoRe`'s raw output of $O(1)$ Mpc$/h$. In the context of the Lyα forest, we observe clustering on scales down to the Jeans Length, approximately $100\,\mathrm{kpc}/h$ [38] and an order of magnitude lower than the resolution we can feasibly achieve. While BAO is a large-scale phenomenon, imposing that the synthetic data has approximately the right small-scale properties ensures that the covariance matrices in our final analyses are realistic. We address this by first interpolating `CoLoRe`'s Gaussian skewers — labelled as δ_C — to a smaller cell size, using nearest grid point (NGP) interpolation in order to avoid introducing additional smoothing.

We then generate a set of new, independent Gaussian skewers δ_ϵ on the grid of smaller cells according to an input 1D power spectrum. We take the k-dependence of this 1D power spectrum to follow that used in [39]:

$$P_{\mathrm{1D}}(k) \propto [1 + (k/k_1)^n]^{-1}, \tag{2.2}$$

where the normalisation is chosen to ensure unit variance. The additional skewers are then scaled by a common factor in order to control the variance in the extra power added. This factor is allowed to vary along the length of the skewers, effectively adding a redshift-dependency to the extra power. Hence, we write this factor as $\sigma_\epsilon(z)$. The parameters n and k_1, as well as the function $\sigma_\epsilon(z)$ are free, and we choose them according to the process described in section 3, aiming to achieve the correct 1D power spectrum across a range of redshifts. The new skewers are then simply added to each of the existing ones to form our final Gaussian skewers δ_G:

$$\delta_G(z, \mathbf{x}) = \delta_C(\mathbf{x}) + \sigma_\epsilon(z)\delta_\epsilon(\mathbf{x}). \tag{2.3}$$

The top panel of figure 1 shows a sample skewer before and after the extra small-scale power is added. As the additional skewers are independent from one another, there are no correlations between the structures added to each of the skewers. When we measure the 3D correlation function, we ignore contributions from pixel-pairs in the same skewer and so this process of adding small-scale power will not affect the 3D correlations of the Gaussian field beyond simply adding noise.

It is worth noting that we could have chosen to add extra small-scale fluctuations to the velocity field and achieved the same correct 1D power spectrum. However, allowing parameters describing extra small-scale velocities to vary freely would require the re-computation of the redshift-space distortions weights matrix (see section 2.2.3) at each step of the tuning process (see section 3). This is a considerably more time-consuming procedure than simply carrying out the inverse Fourier transform of equation (2.2). As such, we choose to only add small-scale fluctuations to the Gaussian field and assign to each of the small cells the velocity of the nearest large `CoLoRe` cell.

2.2.2 Transformation to optical depth

In `LyaCoLoRe`, the transformation from skewers of the Gaussian field to ones of optical depth is governed by two equations. The first of these is known as a lognormal transformation. This approximates the density of the baryonic matter field closely by using a lognormally-distributed variable [32], introducing a degree of non-linearity. This is normalised so that we may define a deviation δ from the mean density as:

$$1 + \delta(z, \mathbf{x}) = \frac{\rho(z, \mathbf{x})}{\bar{\rho}(z)} = \exp\left[D(z)\delta_G(z, \mathbf{x}) - D^2(z)\frac{\sigma_G^2(z)}{2}\right], \tag{2.4}$$

where $D(z)$ is the linear growth factor at redshift z; $\delta_G(z, \mathbf{x})$ is the Gaussian field value from equation (2.3); $\sigma_G(z)$ is the standard deviation of this Gaussian field and $\rho(z, \mathbf{x})$ is the lognormal density at redshift z and position $\mathbf{x}$. This transformation is shown by the transition from the top to the second panel in figure 1.

The second equation allows us to transform these deviations in density into an approximation of the optical depth at each point. Assuming adiabatic expansion implies a tight relationship between temperature and density of the form $\mathrm{d}\ln T/\mathrm{d}\ln\rho = \gamma - 1$ [40]. If we further assume photoionization equilibrium, the temperature of the gas approximately determines the number of neutral hydrogen atoms $n_{\mathrm{HI}} \propto \rho^2 T^{-0.7}$ for a given baryonic matter density ρ [41]. As the optical depth τ is proportional to n_{HI} [42], these two assumptions allow us to provide an approximation for τ given ρ known as the fluctuating Gunn-Peterson approximation (FGPA) [32, 35]:

$$\tau(z, \mathbf{x}) = \tau_0(z)[1 + \delta(z, \mathbf{x})]^{\alpha(z)}, \tag{2.5}$$

where $\tau_0(z)$ is a normalisation determined by the gas temperature and the photoionisation rate, and $\alpha(z) = 2 - 0.7(\gamma(z) - 1)$ is determined by the temperature-density relation. These parameter functions $\tau_0(z)$ and $\alpha(z)$ are free, and the method for choosing them is described in section 3. The transformation to optical depth is shown by the transition from the second panel to the dotted line of the third panel in figure 1.

2.2.3 Adding redshift-space distortions

The Lyα forest exists as a sequence of absorption features due to the gradient in the recessional velocity of the IGM caused by the Universe's expansion. Features are redshifted according to their distance from the observer, appearing in a spectrum at an observed wavelength $\lambda_{\mathrm{obs}} = \lambda_\alpha(1 + z)$ for λ_α the Lyα wavelength, and z the absorption redshift. However, peculiar velocities in a region of gas cause its redshift to differ from that due to expansion alone. These effects are known as redshift-space distortions (RSDs), and can be induced by a number of different effects. In particular, RSDs due to gravitationally-induced linear velocities in the IGM are calculated by `CoLoRe`: as mentioned in section 2.1, it produces velocity skewers quantifying this effect by calculating the gradient of the Newtonian gravitational potential.

The transition from real- to redshift-space in each skewer can be thought of as an integral over velocity space of the real-space optical depth field multiplied by a kernel K:

$$\tau(s) = \int \tau(x) K\Big(s - x - v_r\big(x|T(x)\big)\Big)\mathrm{d}x, \tag{2.6}$$

where x and s are velocity coordinates along the skewer in real- and redshift-space respectively, v_r is the radial peculiar velocity, and T is the temperature. The choice of K depends on the complexity of the physical effects that you wish to capture. Choosing a suitable Gaussian kernel allows the inclusion of thermal broadening effects: the apparent spreading of the gas's optical depth contribution in redshift-space due to random thermal velocities of the gas atoms. This is implemented as an option within `LyaCoLoRe`, the details of which are described in appendix B. However, we find that the width σ_v of this Gaussian kernel is often smaller than the typical cell size used in `LyaCoLoRe` when adding small-scale fluctuations. Thus, the net effect of accounting for this physical process is small, and so for the purposes of this work we choose the most straightforward option, setting $K(x) = \delta^D(x)$ for δ^D the Dirac delta function. This shifts the optical depth along each skewer according to the peculiar velocity, and does not attempt to include any further physical effects.

In order to implement equation (2.6), we determine a matrix of weights W_{ij} for each skewer to map its real-space cells τ^x_j to redshift-space cells τ^s_i via the matrix equation $\tau^s_i = W_{ij}\tau^x_j$. The matrix W_{ij} depends on the velocities in the skewer as well as the choice of kernel K, and the details of its calculation can be found in appendix B. Our implementation conserves the integrated optical depth along each line of sight (ignoring pixels which are shifted to un-observed wavelengths). The matrix W_{ij} is near-diagonal and filled mostly by zeros. It can thus be stored in the form of a sparse matrix, and applied to any additional absorption transitions (see section 5.2), reducing both the computation time and memory requirements of adding RSDs to the skewers.

The addition of RSDs (without thermal broadening) to a sample optical depth skewer is shown by the transition from the dotted to the solid line in the third panel of figure 1.

2.2.4 Final transmission skewers

In one final stage, we convert from skewers of optical depth τ to transmitted flux fraction F via the equation:

$$F(s) = \exp\big[-\tau(s)\big], \tag{2.7}$$

and interpolate onto a wavelength grid of the user's choice to obtain $F(\lambda)$, where $\lambda = \lambda_\alpha(1+z)$. These skewers are then written to disc.

This final transformation can be seen in the transition between the solid lines in the third and fourth panels of figure 1. It is worth noting that, while the signal in the lognormal density deviation $1+\delta$ and optical depth τ skewers is dominated by over-dense regions, the signal in flux F becomes saturated (equal to 0) at these points and does not carry a great deal of information. Rather, the intermediate density regions — where the density is high enough to cause some absorption but not so high that saturation occurs — are those from which the most information can be gleaned.

2.3 Computational requirements

In the realisations presented in this work, we specify that `CoLoRe` generates a 4096^3 cell box as a compromise between resolution and memory usage, given the large volume that we must cover in order to realistically represent a Lyα forest survey. Generating approximately 7.5M QSOs (across the whole sky) and drawing subsequent skewers produces a dataset sufficient for a DESI-like survey, allowing for a significant degree of flexibility in the final survey strategy and number densities. The computational cost of producing one such dataset is relatively low, provided suitable multi-node, multi-core computational facilities are available. Running `CoLoRe` using the input data and options specified in section 4.1 in parallel across 32 Haswell compute nodes (each with 32 cores and 128 GB of memory) on the National Energy Research Scientific Computing Centre's *Cori* machine requires approximately 18 minutes to run, equivalent to approximately 300 CPU hours. The large number of nodes is necessary to improve the speed of the code and to satisfy its memory requirements — a total of approximately 920 GB is needed for each run of this size. If such facilities are not available, then the box size must be reduced or the resolution lowered.

The precise requirements for running `LyaCoLoRe` depend strongly on the exact choices of input options. As an example, converting 800k skewers — similar to the number that will be observed by DESI — from `CoLoRe`'s Gaussian output to realistic transmission skewers including RSDs (though not thermal broadening effects) requires only 4 minutes when spread across the same 32 nodes mentioned previously. If such computational facilities are not available, then running `LyaCoLoRe` is still possible as its memory requirements are much lower than `CoLoRe`.

A very small test dataset of 1000 skewers is available within the `LyaCoLoRe` repository. It is straightforward to run `LyaCoLoRe` on this data on any standard laptop to generate sample skewers or to explore the functionality of the code.

3 Parameter tuning

A number of parameters are defined in the various transformations described in section 2.2, namely n, k_1, $\sigma_\epsilon(z)$, $\tau_0(z)$ and $\alpha(z)$ (see equations (2.2), (2.3) and (2.5) for definitions). These are all free parameters, and we would like to be able to choose their values so that our final skewers have particular properties. Specifically, we aim to match the 1D power spectrum $P_{\rm 1D}(k,z)$, mean transmitted flux fraction $\bar{F}(z)$ and large-scale bias $b_{\delta,F}(z)$ (as defined in equation (4.2)) to literature values. Ignoring RSDs and the shape of the 1D power spectrum would allow the problem to be treated analytically, but unfortunately such simplifications are unrealistic. As such, it is not obvious how to choose our parameters correctly, and a more complex process is necessary.

We aim to solve this problem via a minimisation process. We first define a function that we will aim to minimise, and which takes the following steps:

1. Generate sample skewers in F corresponding to a given set of parameter values using the methods described in section 2.2.
2. Measure the 1D power spectrum, mean flux and large-scale bias of these skewers at a selection of redshift values.
3. Evaluate the deviation of each measurement at each redshift from literature results.
4. Quantify this deviation with a single number.

In step 2, we measure $P_{\rm 1D}$ and $\bar{F}$ straightforwardly, excluding cells that sit at a rest-frame wavelength above 1200 Å. We measure $b_{\delta,F}$ by calculating the response of $\bar{F}$ to a small deviation in the average density field: $b_{\delta,F} = (1/\bar{F})\,{\rm d}\bar{F}/{\rm d}\delta$ [43]. The literature values referred to in step 3 are the fitting function from the BOSS DR9 $P_{\rm 1D}$ measurement from [44], the fitting function of the mean flux measurement from [45] and the bias value and redshift evolution determined by the BOSS DR12 combined Lyα auto- and cross-correlation analysis in [25]. Using these literature results as targets, we compute a weighted error for each measurement at each redshift value. When computing the error on the $P_{\rm 1D}$, we prioritise the low-k modes by using a k-dependent error weighting. For $k < 0.02$ s km^{-1}, this is proportional to $1/(1+(k/k_0)^2)$ where $k_0 = 0.01$ s km^{-1}. This ensures the modes most relevant for a BAO analysis — those with $k \lessapprox 0.005$ s km^{-1} [15, 46] — are prioritised over less important, high-k modes. Beyond $k = 0.02$ s km^{-1}, we ignore any errors as our finite cell size makes it unreasonable to expect realistic power at these scales, and these modes were not measured by BOSS. We sum the errors in quadrature over all k-modes using this weighting to produce an overall error on P_{1D}. In step 4, the errors on each measurement at each redshift value are summed in quadrature, and a single number produced. This number quantifies how well a given parameter set is able to produce realistic data, as measured by our specified properties. A standard minimisation routine can then be used to minimise it over the space of input parameters. We use `Minuit` [47], as implemented by the python module `iminuit`[4] to do so.

We introduce a number of simplifications to improve the speed of the minimisation. We assume that $\log \tau_0$ and $\log \sigma_\epsilon$ follow the functional form:

$$\log(X) = A_0 + A_1 \log[(1+z)/(1+z_0)], \tag{3.1}$$

where $z_0 = 3.0$ and the A_i are scalar parameters. In the case of $X = \tau_0$, we fix $A_1 = 4.5$ [48]. Further, we assume that $\alpha(z)$ takes a constant value of 1.65 across redshifts [49] (equivalent to a value of $\gamma = (2-\alpha)/0.7+1$ of 1.5, in reasonable agreement with the literature [e.g. 50, 51]). With these simplifications, we end up with a 5-parameter minimisation problem: 1 parameter describing the normalisation of $\tau_0(z)$; 2 describing the normalisation and z-dependence of $\sigma_\epsilon(z)$; and 2 describing the shape of the 1D power of the small scale fluctuations (n and k_1). At each call of the routine, we produce sample skewers at a point in parameter space, and compute their 1D power spectra, mean flux and bias parameter values in 7 redshift bins of width $\Delta z = 0.2$ centred at points evenly spaced between $z = 2.0$ and 3.2. We run this procedure using $\sim 55,000$ skewers to obtain an initial estimate, and increase this to $\sim 220,000$ skewers in order to fine tune the optimisation.

We also introduce a parameter a_v by which we multiply the velocities in our skewers in order to match the amount of anisotropy in the clustering of the Lyα forest to literature values. This is not because the velocities from `CoLoRe` are incorrect — when using `CoLoRe`'s unmodified velocities, we obtain the correct level of anisotropy in the QSO auto-correlation (see appendix A) — but is a result of the approximations in our recipe to estimate F. These approximations define the relationship between `CoLoRe`'s initial Gaussian field and our final flux skewers, and thus their inherent assumptions will affect the large-scale flux biases. While the density bias b_δ is matched to literature values in our tuning process, this is not the case for the RSD parameter β (defined in section 4.3). As such, the somewhat crude nature of our approximations results in an unrealistic value of β, and we must introduce our parameter a_v in order to correct for this shortcoming. Improving on our approximations by using more detailed modelling [e.g. 36] might allow us to avoid introducing a_v, but that is beyond the scope of this work. We fix $a_v = 1.3$ when tuning; it is computationally costly to leave it free as a change in a_v requires re-computation of the RSD weights matrix W_{ij} (see section 2.2.3). The value is chosen on an ad hoc basis to match approximately the RSD parameter β measured from BOSS DR12 data [25].

The final values of the transformation parameters are $\log[\tau_0(z)] = 1.48 + 4.5 \log x$, $\alpha(z) = 1.65$, $\log[\sigma_\epsilon(z)] = 6.02 + 0.276 \log x$, $n = 0.732$, $k_1 = 0.0341$ and $a_v = 1.3$, where $x = [(1+z)/(1+z_0)]$ and numerical values are rounded to three significant figures where appropriate. These are the default values used by `LyaCoLoRe`. The tuning process is effective, matching literature values of P_{1D}, $\bar{F}$ and $b_{\delta,F}$ to within 10% at almost all relevant k-modes and z values. As an example, the P_{1D} measured across ~ 7.5M skewers is shown in figure 2. We only plot 4 redshift bins and a limited number of k-modes here for clearer visualisation.

---

[4] Publicly available at https://github.com/iminuit/iminuit.

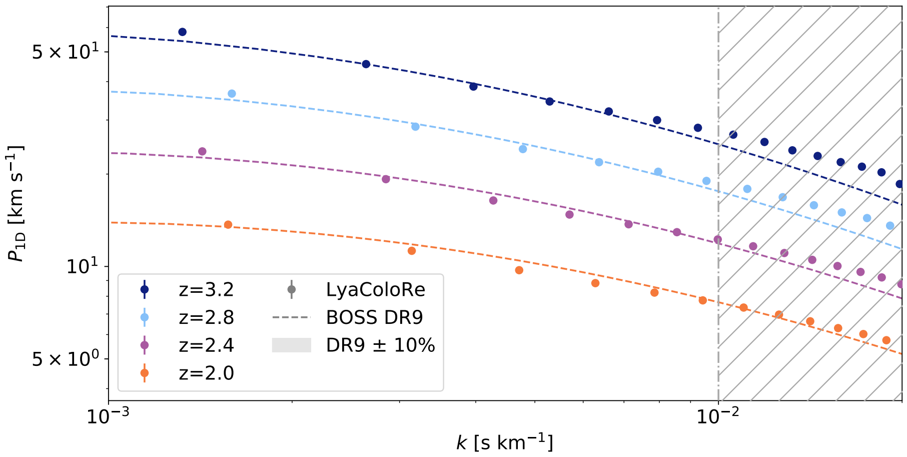

Figure 2. The 1D power spectrum as measured from one realisation of `LyaCoLoRe` mocks. The tuning process aims to match the measured P_{1D} to that from BOSS DR9 data [44] for k-modes that affect BAO analysis, as described in detail in section 3. Modes to the left of the dot-dash line at $k = 0.01$ s km^{-1} are the most important in this respect [15, 46], and these all lie within 10% of our target P_{1D}, as indicated by the shaded areas. Modes to the right of the dot-dash line are not important in the context of BAO, and are not prioritised in our tuning procedure.

4 Verifying the mocks

The primary motivation for creating the `LyaCoLoRe` mocks is to provide realistic sets of test skewers for BAO analyses from Lyα forest surveys. Evidently then, it is important to verify that the fundamental physical quantities studied by such analyses are correctly reproduced in the mock datasets. We thus seek to test that the BAO signal is present and unbiased in our mock datasets. Section 4.1 describes the inputs we use in generating a collection of mock datasets; section 4.2 explains how we measure the correlation functions from each realisation, taking the skewers in F directly from `LyaCoLoRe`'s output; and finally section 4.3 shows how we fit to a model. We do not visually compare the correlation functions measured from mocks to those from data, since our mock measurements are not affected by distortions from continuum fitting. Instead we compare fitted parameter values in order to assess the performance of our mock datasets.

4.1 Generating realisations

The input power spectrum that we use in step 1 of section 2.1 is generated by the Boltzmann solver `CAMB` [52] using the *Planck* Collaboration's 2015 parameters for a flat, ΛCDM cosmology [see column 1 of table 3 in 53]. We generate the field in a box of 4096^3 cells, stipulating that this covers a redshift range $0.0 \leqslant z \leqslant 3.79$: a volume large enough to contain a DESI-like survey. This results in a grid of total size $\sim (9.8\ \mathrm{Gpc}/h)^3$, with each cell $\sim (2.4\ \mathrm{Mpc}/h)^3$ in dimensions. The QSO number density function is based on estimates from SDSS-III data in Stripe 82 [54]. This is considered to represent an optimistic estimate of the photometric capability of targeting for DESI, and results in ~ 3.7M QSOs[5] above $z = 1.8$ across the whole sky. We use as an input QSO bias the fitting function defined in equation 19 of [55], which

---

[5] This is lower than the 7.5M quoted in section 2.3 as we no longer require the previously mentioned flexibility to adapt to different observing strategies in our realisations, and thus can reduce the QSO number density to more realistic values (approximately 59 QSOs per square degree).

is based on clustering measurements from the BOSS DR12 QSO sample [56]. When running `LyaCoLoRe`, we use a cell size of $0.25\,\mathrm{Mpc}/h$, and tune the parameters of our transformations according to the methods described in section 3.

For the purposes of this work we generate 10 such realisations, each with unique random seeds, and stack our results in order to test `LyaCoLoRe` as stringently as possible. This is approximately equivalent to 30 times the final number of Lyα QSOs with $z \geq 2.1$ that will be observed by DESI. It is worth noting that the signal to noise ratio will be significantly greater than 30 times that of DESI, as our skewers of $F(\lambda)$ do not include any instrumental noise, nor do they require any continuum fitting (as mentioned in section 2).

4.2 Measuring correlation functions

We test the BAO signal in our mock realisations in the standard way, by measuring correlation functions using the contrast in flux transmission:

$$\delta_F(\lambda) = \frac{F(\lambda)}{\bar{F}(\lambda)} - 1, \tag{4.1}$$

where $\bar{F}(\lambda)$ is the mean value of $F(\lambda)$ in each pixel over all skewers for which that cell corresponds to rest-frame wavelength $\lambda_r \in [1040, 1200]$ Å. The skewers of $F(\lambda)$ are taken straight from the processes described in section 2, with no further steps such as addition of continua or instrumental noise. This allows us to test the methods of section 2 to as high a degree of precision as possible, but consequently our covariance matrices may not necessarily be representative of true measurements.

We would like to measure the 3D Lyα auto-correlation and the 3D Lyα-QSO cross-correlation, the standard measurements made by recent Lyα BAO analyses from BOSS and eBOSS. Both are estimated using the `Package for IGM Cosmological-Correlations Analyses` (`picca`).[6] We measure these correlations separately in 3,072 `HEALPix` [57] pixels on the sky for each of the 10 realisations, and treat the resultant measurements as a set of 30,720 independent subsamples. In order to compute the correlation functions more quickly, we rebin pixels in our final transmission skewers into larger pixels of width 3×10^{-4} log(Å) in log-wavelength. This enables us to use a larger number of skewers and thus reduce our errors, without compromising the large-scale properties of the correlations or incurring large computational costs.

Our computation of the 3D Lyα auto-correlation follows that of recent Lyα forest BAO analyses [21, 22]. We first define a grid of bins in parallel and perpendicular separation between pairs of pixels — $r_\parallel$ and $r_\perp$ respectively — where each bin is $4\,\mathrm{Mpc}/h \times 4\,\mathrm{Mpc}/h$ in size, and the maximum separation is $200\,\mathrm{Mpc}/h$ in each direction. Pixel pairs are assigned to one of these bins by using a fiducial cosmology to convert from wavelength and angular separations to comoving distances parallel and perpendicular to the line-of-sight. The correlation is then computed as a weighted sum of products of pixel pairs of δ_F within each bin. We restrict ourselves to include only contributions from the Lyα absorption in the Lyα region, ignoring delta pixels outside the rest-frame wavelength range [1040,1200] Å. The covariance matrix is estimated straightforwardly by calculating the scatter between our set of 30,720 subsamples.

The 3D Lyα-QSO cross-correlation is also computed in line with recent analyses of BOSS and eBOSS data [25, 26], as a weighted sum of pixels of δ_F within bins of parallel and

---

[6] Publicly available at https://github.com/igmhub/picca.

perpendicular separation. We use the same bin size as in the auto-correlation, but are able to extend our minimum value of $r_{\parallel}$ to $-200\,\mathrm{Mpc}/h$ as the pixel-pixel pair symmetry of the auto-correlation is not present in the pixel-QSO pairs of the cross-correlation. As for the Lyα auto-correlation, we restrict the rest-frame wavelength range of our δ_F pixels to [1040,1200] Å, and we estimate our covariance matrix from the scatter between our 30,720 subsamples.

4.3 Fitting the correlation functions

Having measured the 3D Lyα auto- and Lyα-QSO cross- correlations, we fit a model to our measurements to obtain the location of the BAO peak and check that no significant shift has been introduced. We also seek to measure the bias parameters of our tracers: Lyα flux F and QSOs. These are defined by the relationship between the power spectra of the tracers, $P_F(\mathbf{k})$ and $P_{\mathrm{QSO}}(\mathbf{k})$, and the power spectrum of dark matter $P(\mathbf{k})$ [58]:

$$P_F(\mathbf{k}) = [b_{\delta,F} + b_{\eta,F} f \mu^2]^2 P(\mathbf{k}), \tag{4.2}$$

$$P_{\mathrm{QSO}}(\mathbf{k}) = [b_{\delta,\mathrm{QSO}} + f \mu^2]^2 P(\mathbf{k}). \tag{4.3}$$

Here, the large-scale biases of flux and QSOs are $b_{\delta,F}$ and $b_{\delta,\mathrm{QSO}}$. The parameter $b_{\eta,F}$ is the velocity gradient bias of flux, which serves to quantify the effect of RSDs. This is often expressed alternatively using $\beta = f b_{\eta,F}/b_{\delta,F}$. The value of $b_{\eta,QSO}$ is 1 by default as QSOs are conserved under RSDs (unlike F) and so it is held fixed [58]. The Lyα-QSO cross- power spectrum follows naturally from (4.2) and (4.3) as:

$$P_{F\times\mathrm{QSO}}(\mathbf{k}) = [b_{\delta,F} + b_{\eta,F} f \mu^2][b_{\delta,\mathrm{QSO}} + f \mu^2] P(\mathbf{k}). \tag{4.4}$$

We fit a model of the correlation functions to each of the measurements individually, and then to both correlations jointly. We use the same models as recent eBOSS analyses [22, 26] but ignore terms relating to systematics not present in our realisations, such as metal absorbers and high column density systems (HCDs). As we do not add continua to our skewers, we need not worry about the distortion of the correlations by the removal of long wavelength modes in the continuum fitting process, as occurs in real analyses. Thus, we do not need to consider distortion matrices, the standard method for taking these effects into account [introduced for the auto- and cross- correlations respectively in 21, 25]. The relevant terms are described using Kaiser models [58], as described in section 4.1 of [22] for the Lyα auto-correlation, and section 5.1 of [26] for the Lyα-QSO cross-correlation. We use the same cosmology as used to generate the input power spectrum of `CoLoRe` to produce the smooth and peak components of the fiducial model power spectrum.

The fit is carried out leaving free the parameters describing the position of the BAO peak in the perpendicular and parallel directions:

$$\alpha_{\parallel} = \frac{D_H(z)/r_d}{[D_H(z)/r_d]_{\mathrm{fid}}}, \quad \alpha_{\perp} = \frac{D_A(z)/r_d}{[D_A(z)/r_d]_{\mathrm{fid}}}, \tag{4.5}$$

where $D_H(z) = c/H(z)$, as well as parameters describing the bias and RSDs of the Lyα-forest, $b_{\eta,F}$ and $\beta_F = f b_{\eta,F}/b_{\delta,F}$. We also leave free 2 parameters that describe the smoothing of the model power spectrum in the parallel and perpendicular directions, which help to account for the effects of the low-resolution of our `CoLoRe` grid. When fitting the Lyα-QSO cross-correlation individually, we fix the value of the QSO bias $b_{\delta,\mathrm{QSO}}$ to the input value in order to avoid degeneracies, though when we fit jointly with the Lyα auto-correlation we are able to leave it free.

Having defined our models, the fits are then carried out using `picca`. We fit only on separations $40 < r$ [Mpc/h] < 160 as the lognormal density approximation used in both `CoLoRe` and `LyaCoLoRe` begins to break down on scales smaller than this, and we are not able to fit the shape of the correlation function well at these separations. Further, the QSOs cannot be expected to be correctly clustered on the smallest scales due to the low resolution of the `CoLoRe` box. To determine an effective redshift of our measurements, we consider pixel-pixel/pixel-QSO pairs which fall in bins A which satisfy $80 < r_A$ [Mpc/h] < 120, i.e. the bins that cover the BAO peak. The value of $z_{\rm eff}$ is then given by a weighted average of the redshifts of pairs in these bins.

The measured Lyα auto- and Lyα-QSO cross-correlations are shown in the left and right panels of figure 3 respectively, along with the model from the combined fit.[7] We plot the correlations as $\xi(r)$ in bins of $|\mu| = |r_{||}|/r$, where $|\mu|$ close to 0 indicates correlations close to perpendicular to the line of sight, and $|\mu|$ close to 1 indicates correlations close to parallel to the line of sight. The model appears to be a good fit to the measurement on the scales that we fit over, and the BAO peak is correctly placed. The two measurements deviate slightly from the model either side of the BAO bump in the highest $|\mu|$ bin, but this deviation is very small and is noticeable due to the extremely small error bars on our measurements.

The parameters from the individual and combined fits are shown in table 1. The α parameters for each fit are all consistent with 1 to within 1σ. Any deviation in the α parameters from 1 is certainly less than 0.2%, and so can be considered insignificant in the context of a DESI-like survey. Thus, the mock production pipeline up to this stage can be said to introduce no clear systematic bias within the capabilities of a current or near-future instrument.

In order to compare the values of biases and βs to BOSS DR12 values [table 4 of 25], we first use the published functional forms of each parameter's redshift evolution to match the effective redshift of the BOSS DR12 measurements to that of our measurements. Having done so, we find that the two sets of values are very similar, with our measurements all lying within the 1σ errors on the BOSS DR12 values. In particular, the values of $b_{\delta,F}$ in each of our fits are almost identical to the BOSS DR12 value, demonstrating the effectiveness of our tuning of this parameter (see section 3). We do not compare the value of $\beta_{\rm QSO}$ to BOSS DR12 measurements as our input QSO bias takes a different value at this redshift. However, the value of $b_{\delta,{\rm QSO}}$ deduced from the joint fit is consistent with the input value (as shown at the bottom of the column showing the Lyα-QSO only fit). As such, we can consider the mocks to fulfil the basic criteria required of them, and thus they appear sufficient for a DESI-like survey. We do not assess the χ^2 of the fits as we do not expect our covariance matrices to be representative of those one would expect from a real survey given the lack of noise in our skewers.

5 Adding secondary astrophysical effects

A key purpose of creating mock datasets is to quantify the impact of secondary astrophysical effects on our measurements so that we may assess any biases that they could induce in our cosmological inference. When generating realisations of the synthetic data, we may choose to add or not to add different effects to each realisation, or to vary the strength of a given effect across a range of values. The resultant impact on BAO measurements can then be quantified. In Lyα forest analyses, two of the most pertinent effects are the presence of high

---

[7] Note that error bars are present for all points, but are often exceedingly small and thus obscured by the points themselves.

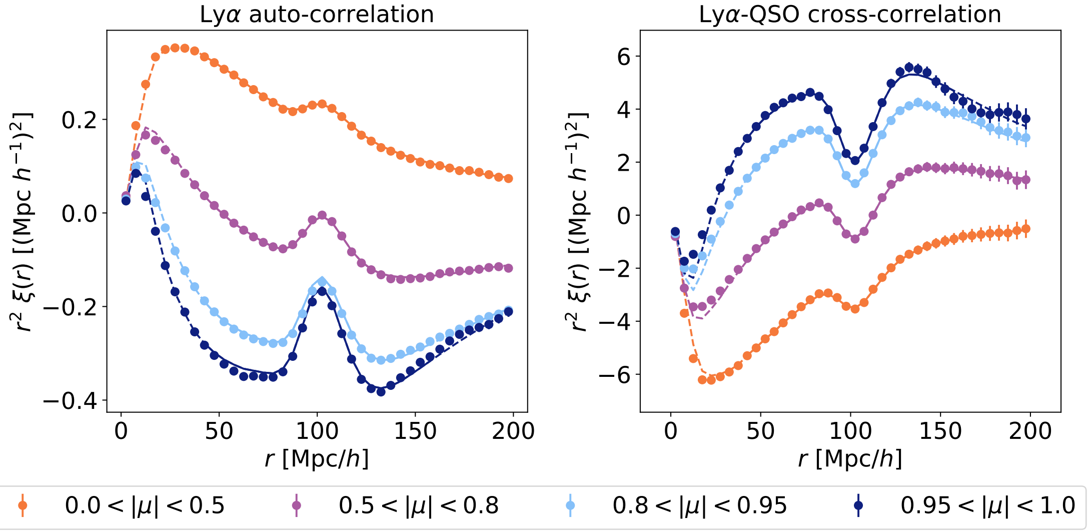

Figure 3. The correlation functions measured from 10 realisations of `LyaCoLoRe` datasets combined, and the best fit lines with parameters as described in the third column of table 1. The left panel shows the Lyα auto-correlation, while the right panel shows the Lyα-QSO cross-correlation. Each plot panel shows the same 4 bins in $|\mu| = |r_{\parallel}|/r$. Note that the correlations presented here do not have any distortion from continuum-fitting and so should not be visually compared with the equivalent plots from recent BOSS/eBOSS data.

	`LyaCoLoRe`			BOSS DR12
Parameter name	Lyα	Lyα-QSO	Lyα + Lyα-QSO	Lyα + Lyα-QSO
$\alpha_{\parallel}$	1.000 ± 0.002	1.001 ± 0.002	1.000 ± 0.001	
$\alpha_{\perp}$	0.998 ± 0.002	1.000 ± 0.002	0.999 ± 0.001	
$b_{\eta,F}$	-0.204 ± 0.0004	-0.201 ± 0.0009	-0.203 ± 0.0004	-0.206 ± 0.012
β_F	1.627 ± 0.008	1.624 ± 0.012	1.624 ± 0.007	1.650 ± 0.081
β_{QSO}		0.261	0.261 ± 0.0007	
$b_{\delta,F}$	-0.121 ± 0.0006	-0.120 ± 0.0009	-0.121 ± 0.0005	-0.121 ± 0.004
b_{QSO}		3.701	3.700 ± 0.009	

Table 1. Parameters from model fits of the Lyα auto-correlation and Lyα-QSO cross-correlation functions measured from 10 realisations of `LyaCoLoRe` mocks combined. The relevant results from the BOSS DR12 combined fit [25] — those to which our values of $b_{\delta,F}$ and $b_{\eta,F}$ are tuned — are presented in the rightmost column at the same effective redshift as our measurements. The parameters in the first segment of the table are those used in the minimisation process which determines the best fit to our correlations, while those in the second segment are calculated subsequently.

column density systems (HCDs) and additional absorption transitions. `LyaCoLoRe` is able to compute both of these effects, and the methods it uses to do so are described in section 5.1 and section 5.2 respectively (alternative implementations of these effects are also possible). Once computed, `LyaCoLoRe` stores skewers of metal absorption and a table of HCDs in its output. These can then be added to the Lyα skewers during subsequent stages of the pipeline by packages such as `desisim`.

We do not present here a full study of the effects of HCDs and additional transitions on a BAO analysis. Rather, we simply illustrate in section 5.3 that their implementations within `LyaCoLoRe` are broadly correct and achieve the correct levels of clustering. We leave the study of these effects as systematics in a BAO analysis for future work.

5.1 Adding HCDs

HCDs occur in particularly dense regions of gas, and contain a number of subclasses determined by HI column density. Typically, we define regions with column density $N_{\rm HI} > 2 \times 10^{20}$ cm^{-2} as Damped Lyα absorbers (DLAs), and regions with column density $10^{17.2} < N_{\rm HI} < 2 \times 10^{20}$ cm^{-2} as Lyman Limit Systems (LLSs) [59]. In detailed Lyα forest analyses, it is important to be able to identify HCDs as their high densities broaden their absorption profiles, impacting on inferred values of F over a significant wavelength range [60, 61]. Further, HCDs are of scientific interest in and of themselves [e.g. 62–66]. As such, being able to add HCDs to our mocks is important in maximising their realism.

We first determine potential HCD locations by computing a threshold value of the Gaussian field, set by an input bias $b_{\rm HCD}(z)$. In our realisations, we choose $b_{\rm HCD}(z) = 2.0$ to be constant with redshift, and in line with [66]. This picks out peaks in the field that are sufficiently dense to host an HCD. We then Poisson sample the potential locations according to an input number density $n_{\rm HCD}(z)$. This number density is imported from the default model of the IGM physics package `pyigm`[8] [67], which is fitted to a selection of literature results [summarised in table 1 of 68]. The sampling is carried out before adding small-scale power (section 2.2.1), as we would like the HCDs to correlate with the 3D fluctuations rather than the 1D extra power. A column density is then allocated to each HCD using a given redshift distribution — again from the default model of `pyigm` — and a radial velocity is determined using `CoLoRe`'s output. The resulting catalogue of HCDs can then be interpreted by a package such as `desisim`, which is able to calculate the absorption profile of the HCD using a Voigt template, and insert it into the final spectrum.

5.2 Including additional absorption transitions

As with HCDs, absorption from additional transitions are an important level of detail to add to our mocks and are of significant scientific interest [e.g. 55, 69–71]. Additional transitions have a rest-frame absorption wavelength different to that of Lyα, and so absorption from gas at the same redshift appears at different observed wavelengths in spectra. Conversely, absorption from two different transitions can appear at the same observed wavelength even though the regions of gas hosting the absorbers are far apart physically. As a result, the presence of such absorption transitions acts to contaminate our measurements of Lyα flux, and thus our correlation functions and resultant BAO measurements. Such transitions include Lyman-β (Lyβ), as well as from silicon, oxygen and carbon gas, for example.

Similar to the method to add HCDs described in section 5.1, it would also be reasonable to place additional absorption transitions using a Poisson-sampled "density-peak" approach, as metals are typically produced in high density regions of the universe. However, we choose to follow the methods of previous works [16, 30], assuming that the optical depth of each additional transition is proportional to that of the Lyα absorber. In the context of these mocks, the most important feature of these additional transitions that we seek to replicate is the strength of their 3D clustering, as it is this that will quantify any impact upon BAO

---

[8] Publicly available at https://github.com/pyigm/pyigm.

Name	Rest-frame wavelength [Å]	Relative absorption strength
Lyα	1215.67	1.0
Lyβ	1025.72	0.1901
SiII (1260)	1260.42	3.542×10^{-4}
SiIII (1207)	1206.50	1.8919×10^{-3}
SiII (1193)	1193.29	9.0776×10^{-4}
SiII (1190)	1190.42	1.28478×10^{-4}

Table 2. Details of a small selection of additional absorption transition that can be used in `LyaCoLoRe`. The absorption strength for each absorber X has been tuned to match approximately the bias value $b_{\delta,X}$ found in literature [21, 22, 25]. It is possible to add more absorbers straightforwardly, but the absorption strengths have not been calibrated beyond those listed above. These absorbers are those included in the skewers from which the correlation function in figure 4 is measured.

measurements. In order to do so, we simply require an absorption strength (relative to Lyα) and a rest-frame wavelength for each additional transition that we wish to include. Having calculated the skewers of optical depth in real space, we scale them differently for each absorption transition according to the transition's relative strength. For an additional transition X, we obtain $\tau_X = A_X \tau_\alpha$, where A_X is the relative strength and τ_α is the Lyα optical depth as defined in equation (2.5). We then apply RSDs (using the same weights matrix W_{ij} as for Lyα), and convert to $F_X(\lambda)$ separately for each X according to its rest-frame wavelength. For each line of sight, the separate $F_X(\lambda)$ skewers are then interpolated onto a common wavelength grid and are combined multiplicatively.

This method ensures that RSDs are correctly applied to each additional absorption transition, and we may tune the absorption strength in order to achieve the correct large-scale bias — and thus the correct 3D clustering — for each transition. A small selection of additional transitions and their relative strengths are shown in table 2. These are the transitions most important to a Lyα BAO analysis, though further transitions can be added straightforwardly if needed. These strengths have been tuned to approximately match bias values presented in the literature [21, 22, 25].

5.3 Testing astrophysical effects

We assess the methods of section 5.1 and section 5.2 by first computing the 3D Lyα-HCD cross-correlation. The methods used to do so are largely the same as used to compute the 3D Lyα-QSO cross-correlation, as described in section 4.2. One significant difference is that we restrict the HCDs in our calculation of the Lyα-HCD cross-correlation to lie in the rest-frame wavelength range $[1040, 1100]$ Å, far from the background QSO. This restriction is necessary to prevent the correlation between Lyα flux and QSOs from significantly affecting our measurements close to the line of sight, as is discussed further in appendix C. An effect can still be seen in the two μ-bins closest to the line of sight at large values of r, though this is mostly beyond the fitted range and so we are still able to measure the degree of clustering in the HCDs well. Future studies may prefer to model this effect in order to avoid reducing the HCD catalogue in this way, but such work is beyond the scope of this analysis. As in section 4, we measure correlations on each of our 10 realisations, and combine the measurements.

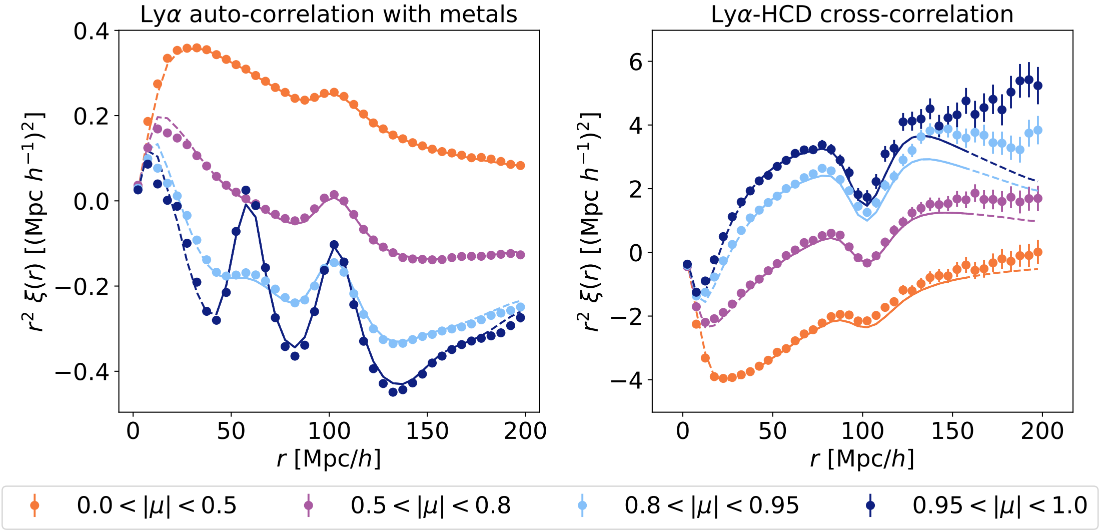

Figure 4. The correlation functions measured from 10 realisations of `LyaCoLoRe` combined, demonstrating the additional astrophysical effects that can be included in its skewers. The left panel shows the flux auto-correlation measured from skewers including metal absorption, from which we measure the metal absorber biases presented in table 3. The right panel shows the Lyα-HCD cross-correlation. The subtleties of this measurement — including the discrepancy at large-r — are discussed in appendix C. Note that the correlations presented here do not have any distortion from continuum-fitting and so should not be visually compared with the equivalent plots from recent BOSS/eBOSS data.

	bias $\times 10^3$		
Absorber	`LyaCoLoRe`	BOSS DR12	eBOSS DR14
SiII (1260 Å)	-1.70 ± 0.04	-1.5 ± 1.2	-2.5 ± 1.3
SiIII (1207 Å)	-3.3	-3.3 ± 1.3	-8.2 ± 1.0
SiII (1193 Å)	-3.28 ± 0.03	-3.5 ± 0.9	-4.6 ± 1.0
SiII (1190 Å)	-4.55 ± 0.03	-4.4 ± 0.9	-5.1 ± 1.0

Table 3. The biases of the metal absorbers used in our realisations of `LyaColoRe`, along with the values from BOSS DR12 [21] and eBOSS DR14 [22] for comparison. The bias of SiIII (1207) is held fixed to the DR12 value as the “bump” that it creates in the correlation function is at $r = 21\,\mathrm{Mpc}/h$, below the minimum separation of $40\,\mathrm{Mpc}/h$ used in our fits. The values from `LyaColoRe` are all within 1σ of those from BOSS DR12, indicating that the absorption strengths used in our realisations (see table 2) result in the correct levels of large-scale clustering.

The measurement of the Lyα-HCD cross-correlation is shown in the right panel of figure 4. Here, we fit for a model in the same way as for the Lyα-QSO cross-correlation. Carrying out a combined fit with the Lyα auto-correlation from section 4 allows us to measure the HCD bias $b_{\delta,\mathrm{HCD}}(z_{\mathrm{eff}}) = 2.26 \pm 0.02$. Strictly, this is not consistent with the redshift-constant input value of $b_{\delta,\mathrm{HCD}} = 2.0$ (as motivated by [66]). There are a number of potential reasons for such a shift, but given the errors on current measurements of $b_{\delta,\mathrm{HCD}}$ from data (approximately 10% in [66]), we do not investigate the agreement further at this stage.

We then compute the 3D auto-correlation from skewers of F that include contributions from the additional absorption transitions in table 2 (on top of Lyα absorption). The method used to do so is identical to that described for the 3D Lyα auto-correlation in section 4.2. We only include contributions from pixels that lie in the rest-frame wavelength range $[1040, 1200]$ Å, and so we do not include any absorption from the Lyβ absorber as its rest-frame wavelength is below this range. As such, from here on we refer to the additional absorption transitions as "metals". As in section 4, we measure correlations on each of our 10 realisations, and combine the measurements.

The measurement of the auto-correlation with metal absorbers is shown in the left panel of figure 4. By comparison with figure 3, the effect of including these metals in the skewers is clearly significant, particularly in the near-line of sight $0.95 < |\mu| < 1.0$ bin. Notably, we can clearly see a at approximately 55–60 Mpc/h as a result of SiII (1190 Å) and SiII (1193 Å) absorption, as well as a peak at approximately 21 Mpc/h from SiIII (1207 Å). This final peak is not included in our fit as it is below the minimum separation. Less visually obvious, but more important to the BAO analysis, is the effect of absorption from SiII (1260 Å), which causes a bump at 105 Mpc/h, very close to the BAO peak.

In our fit of this correlation, we model the effect of metal absorbers in the same way as in [22], summing contributions to the model power spectrum from each combination of pairs of absorbers. In table 3, we show the biases for each of our metal absorbers, as well as the values from BOSS DR12 [21] and eBOSS DR14 [22] for comparison. The bias of SiIII (1207 Å) is held fixed to the DR12 value as the peak that it creates in the correlation function is at $r = 21$ Mpc/h, below the minimum value used in our fits. The `LyaColoRe` values sit within 1σ of those from BOSS DR12, demonstrating that the levels of clustering given by the absorption strengths in table 2 are similar to those found in data. Of course, each absorption strength can be tuned further so that the bias of the relevant absorber more closely matches any given value.

6 Summary & conclusions

In this work we have presented `LyaCoLoRe`, a tool for creating mock Lyα forest datasets when used in conjunction with a Gaussian random field code such as `CoLoRe`. We first use `CoLoRe` to generate skewers from a Gaussian random field, avoiding the use of N-body or hydrodynamical simulations due to the limited volume and high computational cost of such methods. `LyaCoLoRe` is then able to transform the output into realistic skewers of transmitted flux fraction, with a number of properties defined by an automatic tuning process. The process is computationally efficient, making it suitable for generating large numbers of realisations of mocks with different input data and parameters.

We then demonstrate the effectiveness of `LyaCoLoRe`'s output, generating a number of skewers equivalent to approximately 30 realisations of DESI and measuring the Lyα auto- and Lyα-QSO cross-correlations. Fitting these measurements with an appropriate model gives BAO peak positions that are consistent with the input cosmologies to within 0.2%, and certainly within the capabilities of an instrument such as DESI. In addition, the biases of the Lyα forest and of QSOs are shown to be very similar to those derived from BOSS DR12 data. As such, we conclude that the mock datasets generated by `LyaCoLoRe` are suitable for the BAO analyses of current and upcoming surveys such as eBOSS and DESI.

Finally, we demonstrate two additional capabilities of the `LyaCoLoRe` package in adding correlated high column density systems (HCDs) and additional absorption transitions to the skewers. We leave a full analysis of the impact of such features on a BAO analysis to a future work, but demonstrate that the HCDs are clustered approximately correctly on large scales, and that the additional transitions affect the Lyα auto-correlation in the expected manner.

Mock datasets such as those generated by `LyaCoLoRe` are of use to the BAO analyses of Lyα forest surveys in a number of ways. They are able to provide robust tests of analysis pipelines, while they can also help in assessing the impact of astrophysical effects — such as HCDs and additional absorption transitions — on BAO measurements. Finally, they can be used to provide evidence when making decisions regarding the planning of large surveys, such as in targeting and survey strategy. As such, we hope that `LyaCoLoRe` will be of use for Lyα BAO surveys both present and future.

Acknowledgments

We thank Nicolas Busca for valuable discussion, and aiding with the analysis of `LyaCoLoRe` datasets via `picca`. We also thank Matthew Pieri and Satya Gontcho A Gontcho for their comments and feedback through the DESI review process. JF was supported by a Science and Technology Facilities Council (STFC) studentship. AFR acknowledges support by an STFC Ernest Rutherford Fellowship, grant reference ST/N003853/1. AP was supported by the Royal Society. AP and AFR were further supported by STFC Consolidated Grant number ST/R000476/1. This work was partially supported by the UCL Cosmoparticle Initiative. This research is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH1123, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; additional support for DESI is provided by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to the National Optical Astronomy Observatory; the Science and Technologies Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the National Council of Science and Technology of Mexico, and by the DESI Member Institutions. The authors are honored to be permitted to conduct astronomical research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

JCAP03(2020)068

A The quasar auto-correlation

We measure the quasar (QSO) auto-correlation on ten QSO catalogues from ten realisations of `CoLoRe` and combine our results. Correlations are computed as the weighted sum of pairs of QSOs in a grid of parallel and perpendicular separation bins. We divide the sky into `HEALPix` pixels, computing "data-data", "data-random" and "random-random" correlations in each one using a random catalogue of QSOs. This random catalogue has the same number density distribution of QSOs as that in the mock data, and is generated by `LyaColoRe`. The different correlation types are then combined using the Landy-Szalay estimator [72], and the covariance is estimated via sub-sampling across `HEALPix` pixelisations of all 10 realisations (as described in section 4.2). As in section 4.2, all correlations are computed using `picca`.[9] A Kaiser model [58] is then fitted to the measurement, leaving free parameters describing the location of the BAO peak and the QSO bias $b_{\delta,\mathrm{QSO}}$. As in section 4.3, we also leave free parameters describing the smoothing of the input power spectrum in the parallel and perpendicular directions. As in section 4.3, we fit only in the range $40 < r$ [Mpc/h] < 160

---

[9] Publicly available at https://github.com/igmhub/picca.

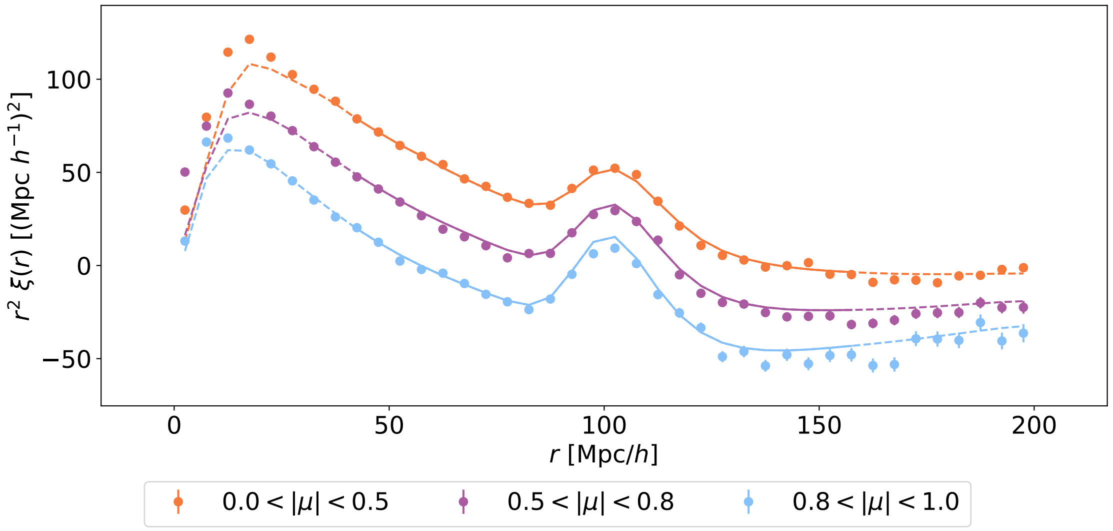

Figure 5. The auto-correlation of QSOs, as measured from ten realisations of `CoLoRe`. The fit is generally good in the fitted region, though the correlation on smaller scales than this is evidently too high.

as the lognormal approximation begins to break down below this range. The resultant fit is very good in the fitted region, as shown in figure 5. We measure a QSO bias of 3.57 ± 0.01 at an effective redshift of $z = 2.20$, consistent with the input value of 3.56 to within 1σ.

B Redshift-space distortions: implementation details

As described in section 2.2.3, adding RSDs to our skewers requires the calculation of a matrix of weights W_{ij} to map each skewer's real-space cells τ_j^x to redshift-space cells τ_i^s via the matrix equation $\tau_i^s = W_{ij}\tau_j^x$. W_{ij} is determined by representing each cell as a top-hat function in real space, mapping this profile into redshift space according to the choice of kernel K, and calculating the overlap with each redshift-space cell:

$$W_{ij} = \int_{s_j^l}^{s_j^u} P(s - x_i - v_{r,i}|T_i, d_i)\mathrm{d}s, \tag{B.1}$$

where s_j^l and s_j^u are the lower and upper boundaries of cell j in redshift-space, and $P(x|T,d)$ describes the profile of the real-space cell when mapped into redshift space. $P(x|T,d)$ is dependent on the distance from the centre of the cell x, the temperature of the gas T and the half-width of the cell d. The form of P is determined by the choice of K (defined in equation (2.6)):

$$P(x|T,d) = \frac{1}{2d}\int_{-d}^{d} K(x-y|T)\mathrm{d}y. \tag{B.2}$$

As such, in the case of K chosen to be a Dirac delta function, the redshift-space cell is represented by a top-hat function (as it was in real space).

In order to account for thermal broadening when adding RSDs to our skewers, we must instead choose our kernel K to be defined by

$$K(x|T) = \frac{1}{\sqrt{2\pi}\sigma_v(T)} \exp\left(-\frac{x^2}{2\sigma_v^2(T)}\right), \tag{B.3}$$

where $\sigma_v(T)$ is the thermal velocity dispersion, which we approximate as in [49] by

$$\sigma_v(T) = 9.1\left(\frac{T}{10,000\text{K}}\right)^{1/2} \text{ km s}^{-1}, \tag{B.4}$$

for temperature $T(z, \mathbf{x}) = T_0(z)\rho(z, \mathbf{x})^{\gamma(z)-1}$. As described in section 3, for the purposes of this work we fix $\gamma = 1.5$. We also fix $T_0 = 10,000\,\text{K}$ in line with [16] and consistent with literature values [49–51]. Of course, these values can easily be updated to follow a more complex redshift dependence for any uses of `LyaCoLoRe` where thermal broadening effects become significant. Evaluating equation (B.2) for this choice of K yields a cell profile in redshift space defined by

$$P(x|T, d) = \frac{1}{4d}\left[\text{erf}\left(\frac{x+d}{\sqrt{2}\sigma_v(T)}\right) - \text{erf}\left(\frac{x-d}{\sqrt{2}\sigma_v(T)}\right)\right], \tag{B.5}$$

and the matrix of weights can then be computed as per equation (B.1).

C The Lyα-HCD cross-correlation

In figure 4, we showed the cross-correlation between Lyα absorption and high column density systems (HCDs) from ten realisations of `LyaCoLoRe`, comparing it to a linear theory model similar to that used to describe the cross-correlation with quasars (QSOs). This model assumes that HCDs have the same clustering as dark matter halos, with a large-scale bias of approximately 2.0. However, in a QSO survey, HCDs are only detected when they are absorbing light from a background QSO, and this observational bias is not taken into account in our modelling. In this appendix, we propose that this bias results in an asymmetry in the measured correlation function. We present a qualitative description of this effect and explain our choice to use only HCDs detected far away from the QSO in figure 4 in this context.

In the left panel of figure 6 we show the same measurement of the Lyα-HCD cross-correlation as in the right panel of figure 4, this time plotting the correlation against $r_\parallel$ in 3 narrow bins of $r_\perp$. The solid lines show the model obtained by fitting this measurement jointly with the Lyα auto-correlation. The model is generally able to fit the measurement well, though some small residuals remain at large $r_\parallel$. These are visible at large separations in the right panel of figure 4, accentuated due to the factor of r^2 in that plot.

In the right panel of figure 6 we plot the Lyα-HCD cross-correlation measured on one realisation of `LyaCoLoRe`, this time using a full HCD catalogue (with no maximum rest-frame wavelength). The solid lines are the exact same lines as in the left panel. It is clear from this plot that there is a strong asymmetry in the data, and the model used to fit the data in the left panel does not fit this measurement well.

We propose that this asymmetry is a consequence of the observational bias that is inherently present in our HCD sample, and the dependence on the Lyα-QSO cross-correlation that this induces. According to the density-QSO cross-correlation, a QSO q will tend to have dense regions of gas around it. In relation to an HCD X in q's spectrum, these dense regions will be located at small $r_\perp$ and $r_\parallel \simeq r_{Xq}$, the distance between X and q (as X is constrained to lie directly along the line of sight between q and the observer). This preferential location of dense regions of gas will imprint a feature in the correlation between X and neighbouring skewers of δ_F at these specific separations. Referring to the diagram in figure 7, we can see that the cells of δ_F in Region 1 will tend to be significantly biased according to the Lyα-QSO

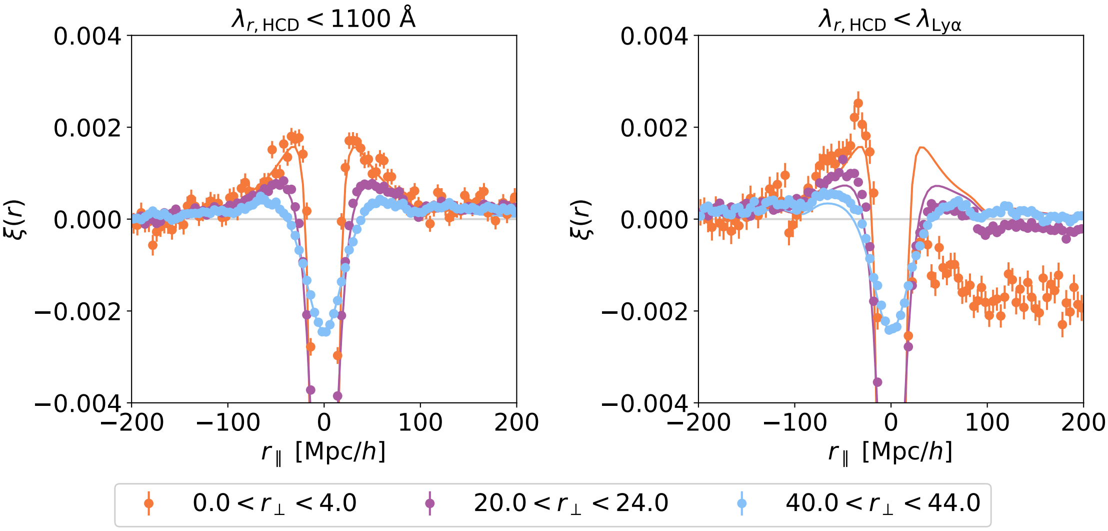

Figure 6. The Lyα-HCD cross-correlation, plotted against $r_\parallel$ for different bins of $r_\perp$. The left panel shows the combined measurement from ten realisations using an HCD catalogue that only includes HCDS with rest-frame wavelength less than 1100 Å (as in the right panel of figure 4). The right panel shows the correlation measured from one realisation when using an HCD catalogue that includes HCDs in the full rest-frame wavelength range, up to $\lambda_{\mathrm{Ly}\alpha} = 1215.67$ Å. The solid lines in both panels show the same fitted correlation as in figure 4: the joint fit of the Lyα auto-correlation and Lyα-HCD cross-correlation from ten realisations of `LyaCoLoRe`.

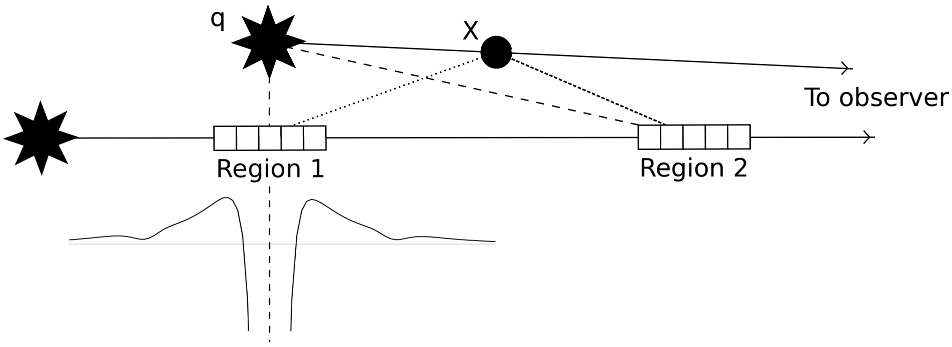

Figure 7. A diagram showing the geometry of the setup involved when measuring the Lyα-HCD cross-correlation between two near-parallel skewers. Given the proximity of the QSO q to “Region 1” of the lower skewer, we expect to measure biased values of δ_F for cells corresponding to that region. The cells' values will tend to be reduced or boosted according to the Lyα-QSO cross correlation, as indicated beneath “Region 1”. This biasing is then imprinted on the correlation between an HCD X and the skewer.

cross-correlation. Thus, we will see a feature in the correlation between HCD X and its neighbouring skewer corresponding to this region. The shape of this feature is determined by the shape of the Lyα-QSO cross-correlation at small r_t, as shown beneath Region 1. The cells in Region 2 will not be significantly affected by the presence of q, and so we would not expect to see a feature here.

Summing over HCD-pixel pairs in order to compute the full Lyα-HCD cross-correlation will average out most of the signal, but a small, asymmetric residual will remain, as seen in the right panel of figure 6. The contribution from each HCD will carry a similar signature but the signature will be centred at different values of $r_{\parallel}$ due to the different values of r_{Xq} for each X-q pair. Certainly though, the sign of r_{Xq} will always be the same as an HCD is always less distant than its host QSO. Using `picca`'s definition of the sign of $r_{\parallel}$, this means that $r_{Xq} > 0$ for all X and q. As a result, we will see a reduction of the Lyα-HCD cross-correlation for all $r_{\parallel} > 0$, due to the strong reduction in δ_F at the centre of regions such as Region 1 in figure 7. This is only apparent for $r_{\perp}$ small as the reduced area shown in figure 7 is narrow. We also see a secondary effect: a boost in the Lyα-HCD cross-correlation for small, negative $r_{\parallel}$. This is a result of the small boost in δ_F on the right-hand side of Region 1 in figure 7, which appears at $r_{\parallel} < 0$ for HCDs that are very close to their host QSOs. This effect extends to larger values of $r_{\perp}$ due to the greater width of the boosted area (relative to that which is reduced).

Whilst interesting, these effects are very small. In order to assess their visibility in current/future studies, we would need to carry out tests using a more realistic mock dataset. This would involve using the entire data reduction pipeline — including continuum fitting and the use of a distortion matrix — and is beyond the scope of this work. As an approximate comparison, we observe that the size of the deviation of points in the $0.0 < r_{\perp} < 4.0$ bin in the right panel of figure 6 is approximately an order of magnitude smaller than the size of the error bars in the uppermost two panels of figure 2 of [66].[10]

Making such an extreme cut in rest-frame wavelength greatly reduces the number of HCDs in our catalogue. In this work we use approximately 30 times the number of skewers as DESI will have, and so this reduction does not cause us any concern. For studies from real surveys, however, maximising the scientific value of their data will be of much greater importance. As such, we would recommend the development of a new model to account for the effects described above using the measured Lyα-QSO cross-correlation. Alternatively, a catalogue of random HCDs, uncorrelated with the Lyα forest, could be generated and used to quantify these effects before accounting for them appropriately. Either way, further tests are needed in order to understand more fully the effect described in this appendix, particularly if new modelling is required for future Lyα-HCD cross-correlation measurements.

References

[1] Supernova Search Team collaboration, *Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant*, *Astron. J.* **116** (1998) 1009 [`astro-ph/9805201`] [INSPIRE].

[2] Supernova Cosmology Project collaboration, *Measurements of Ω and Λ from 42 High-Redshift Supernovae*, *Astrophys. J.* **517** (1999) 565 [`astro-ph/9812133`] [INSPIRE].

[3] P.J.E. Peebles and J.T. Yu, *Primeval Adiabatic Perturbation in an Expanding Universe*, *Astrophys. J.* **162** (1970) 815 [INSPIRE].

[4] SDSS collaboration, *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies*, *Astrophys. J.* **633** (2005) 560 [`astro-ph/0501171`] [INSPIRE].

---

[10] It should be noted that [66] includes only HCDs at least 5000 km/s away from their host quasar, equivalent to a rest-frame wavelength cut of approximately 1195 Å. We choose to use $\lambda_{r,\mathrm{HCD}} < \lambda_{\mathrm{Ly}\alpha}$ in the right panel of figure 6 in order to explain the relationship between the geometry of the problem and the observed effect more clearly.

[5] 2dFGRS collaboration, *The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications*, Mon. Not. Roy. Astron. Soc. **362** (2005) 505 [`astro-ph/0501174`] [INSPIRE].

[6] SDSS collaboration, *Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample*, Mon. Not. Roy. Astron. Soc. **401** (2010) 2148 [`arXiv:0907.1660`] [INSPIRE].

[7] F. Beutler et al., *The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant*, Mon. Not. Roy. Astron. Soc. **416** (2011) 3017 [`arXiv:1106.3366`] [INSPIRE].

[8] C. Blake et al., *The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations*, Mon. Not. Roy. Astron. Soc. **418** (2011) 1707 [`arXiv:1108.2635`] [INSPIRE].

[9] BOSS collaboration, *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, Mon. Not. Roy. Astron. Soc. **470** (2017) 2617 [`arXiv:1607.03155`] [INSPIRE].

[10] M. Ata et al., *The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2*, Mon. Not. Roy. Astron. Soc. **473** (2018) 4773 [`arXiv:1705.06373`] [INSPIRE].

[11] H.G. Bi, G. Boerner and Y. Chu, *An alternative model for the Ly-alpha absorption forest*, Astron. Astrophys. **266** (1992) 1.

[12] R.-y. Cen, J. Miralda-Escudé, J.P. Ostriker and M. Rauch, *Gravitational Collapse of Small-Scale Structure as the Origin of the Lyman-Alpha Forest*, Astrophys. J. **437** (1994) L9 [`astro-ph/9409017`] [INSPIRE].

[13] P. Petitjean, J.P. Mcket and R. Kates, *The Lyα forest at low redshift: tracing the dark matter filaments*, Astron. Astrophys. **295** (1995) L9 [`astro-ph/9502100`] [INSPIRE].

[14] J. Miralda-Escudé, R.-y. Cen, J.P. Ostriker and M. Rauch, *The Lyα Forest from Gravitational Collapse in the Cold Dark Matter + Λ Model*, Astrophys. J. **471** (1996) 582 [`astro-ph/9511013`] [INSPIRE].

[15] P. McDonald and D. Eisenstein, *Dark energy and curvature from a future baryonic acoustic oscillation survey using the Lyman-α forest*, Phys. Rev. **D 76** (2007) 063009 [`astro-ph/0607122`] [INSPIRE].

[16] A. Slosar et al., *The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data*, JCAP **09** (2011) 001 [`arXiv:1104.5244`] [INSPIRE].

[17] N.G. Busca et al., *Baryon acoustic oscillations in the Lyα forest of BOSS quasars*, Astron. Astrophys. **552** (2013) A96 [`arXiv:1211.2616`] [INSPIRE].

[18] A. Slosar et al., *Measurement of baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9*, JCAP **04** (2013) 026 [`arXiv:1301.3459`] [INSPIRE].

[19] D. Kirkby et al., *Fitting methods for baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9*, JCAP **03** (2013) 024 [`arXiv:1301.3456`] [INSPIRE].

[20] BOSS collaboration, *Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars*, Astron. Astrophys. **574** (2015) A59 [`arXiv:1404.1801`] [INSPIRE].

[21] J.E. Bautista et al., *Measurement of baryon acoustic oscillation correlations at $z = 2.3$ with SDSS DR12 Lyα-Forests*, Astron. Astrophys. **603** (2017) A12 [`arXiv:1702.00176`] [INSPIRE].

[22] V. de Sainte Agathe et al., *Baryon acoustic oscillations at $z = 2.34$ from the correlations of Lyα absorption in eBOSS DR14*, Astron. Astrophys. **629** (2019) A85 [`arXiv:1904.03400`] [INSPIRE].

[23] A. Font-Ribera et al., *The large-scale quasar-Lyman α forest cross-correlation from BOSS*, JCAP **05** (2013) 018 [`arXiv:1303.1937`] [INSPIRE].

[24] BOSS collaboration, *Quasar-Lyman α forest cross-correlation from BOSS DR11: Baryon Acoustic Oscillations*, JCAP **05** (2014) 027 [arXiv:1311.1767] [INSPIRE].

[25] H. du Mas des Bourboux et al., *Baryon acoustic oscillations from the complete SDSS-III Lyα-quasar cross-correlation function at $z = 2.4$*, Astron. Astrophys. **608** (2017) A130 [arXiv:1708.02225] [INSPIRE].

[26] M. Blomqvist et al., *Baryon acoustic oscillations from the cross-correlation of Lyα absorption and quasars in eBOSS DR14*, Astron. Astrophys. **629** (2019) A86 [arXiv:1904.03430] [INSPIRE].

[27] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, arXiv:1611.00036 [INSPIRE].

[28] J.M.L. Goff et al., *Simulations of BAO reconstruction with a quasar Ly-α survey*, Astron. Astrophys. **534** (2011) A135 [arXiv:1107.4233] [INSPIRE].

[29] A. Font-Ribera, P. McDonald and J. Miralda-Escudé, *Generating mock data sets for large-scale Lyman-α forest correlation measurements*, JCAP **01** (2012) 001 [arXiv:1108.5606] [INSPIRE].

[30] J.E. Bautista et al., *Mock Quasar-Lyman-α forest data-sets for the SDSS-III Baryon Oscillation Spectroscopic Survey*, JCAP **05** (2015) 060 [arXiv:1412.0658] [INSPIRE].

[31] P. Coles and B. Jones, *A lognormal model for the cosmological mass distribution*, *Mon. Not. Roy. Astron. Soc.* **248** (1991) 1 [INSPIRE].

[32] H. Bi and A.F. Davidsen, *Evolution of Structure in the Intergalactic Medium and the Nature of the Lyα Forest*, Astrophys. J. **479** (1997) 523 [astro-ph/9611062] [INSPIRE].

[33] F. Bernardeau, S. Colombi, E. Gaztañaga and R. Scoccimarro, *Large-scale structure of the Universe and cosmological perturbation theory*, Phys. Rept. **367** (2002) 1 [astro-ph/0112551] [INSPIRE].

[34] S. Tassev, M. Zaldarriaga and D. Eisenstein, *Solving large scale structure in ten easy steps with COLA*, JCAP **06** (2013) 036 [arXiv:1301.0322] [INSPIRE].

[35] R.A.C. Croft, D.H. Weinberg, N. Katz and L. Hernquist, *Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyα Forest*, Astrophys. J. **495** (1998) 44 [astro-ph/9708018] [INSPIRE].

[36] V. Iršič and M. McQuinn, *Absorber Model: the Halo-like model for the Lyman-α forest*, JCAP **04** (2018) 026 [arXiv:1801.02671] [INSPIRE].

[37] D. Alonso and collaborators, *`CoLoRe`: Cosmological lognormal realisations*, in preparation, (2020).

[38] M. Walther et al., *A New Precision Measurement of the Small-scale Line-of-sight Power Spectrum of the Lyα Forest*, Astrophys. J. **852** (2018) 22 [arXiv:1709.07354] [INSPIRE].

[39] P. McDonald et al., *The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey*, Astrophys. J. Suppl. **163** (2006) 80 [astro-ph/0405013].

[40] L. Hui and N.Y. Gnedin, *Equation of state of the photoionized intergalactic medium*, Mon. Not. Roy. Astron. Soc. **292** (1997) 27 [astro-ph/9612232] [INSPIRE].

[41] L. Hui, N.Y. Gnedin and Y. Zhang, *The Statistics of Density Peaks and the Column Density Distribution of the Lyα Forest*, Astrophys. J. **486** (1997) 599 [astro-ph/9608157] [INSPIRE].

[42] J.E. Gunn and B.A. Peterson, *On the Density of Neutral Hydrogen in Intergalactic Space*, Astrophys. J. **142** (1965) 1633 [INSPIRE].

[43] P. McDonald, *Toward a Measurement of the Cosmological Geometry at $z \sim 2$: Predicting Lyα Forest Correlation in Three Dimensions and the Potential of Future Data Sets*, Astrophys. J. **585** (2003) 34 [astro-ph/0108064] [INSPIRE].

[44] N. Palanque-Delabrouille et al., *The one-dimensional Lyα forest power spectrum from BOSS*, *Astron. Astrophys.* **559** (2013) A85 [arXiv:1306.5896] [INSPIRE].

[45] G.D. Becker, P.C. Hewett, G. Worseck and J.X. Prochaska, *A refined measurement of the mean transmitted flux in the Lyα forest over $2 < z < 5$ using composite quasar spectra*, *Mon. Not. Roy. Astron. Soc.* **430** (2013) 2067 [arXiv:1208.2584] [INSPIRE].

[46] M. McQuinn and M. White, *On estimating Lyα forest correlations between multiple sightlines*, *Mon. Not. Roy. Astron. Soc.* **415** (2011) 2257 [arXiv:1102.1752] [INSPIRE].

[47] F. James and M. Roos, *Minuit — a system for function minimization and analysis of the parameter errors and correlations*, *Comput. Phys. Commun.* **10** (1975) 343 [INSPIRE].

[48] U. Seljak, *Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Ly-α forest*, *JCAP* **03** (2012) 004 [arXiv:1201.0594] [INSPIRE].

[49] P. McDonald, J. Miralda-Escudé, M. Rauch, W.L.W. Sargent, T.A. Barlow and R. Cen, *A Measurement of the Temperature-Density Relation in the Intergalactic Medium Using a New Lyα Absorption-Line Fitting Method*, *Astrophys. J.* **562** (2001) 52 [*Erratum ibid.* **598** (2003) 712] [astro-ph/0005553] [INSPIRE].

[50] M. Ricotti, N.Y. Gnedin and J.M. Shull, *The Evolution of the Effective Equation of State of the Intergalactic Medium*, *Astrophys. J.* **534** (2000) 41 [astro-ph/9906413] [INSPIRE].

[51] H. Hiss et al., *A New Measurement of the Temperature-density Relation of the IGM from Voigt Profile Fitting*, *Astrophys. J.* **865** (2018) 42 [arXiv:1710.00700] [INSPIRE].

[52] A. Lewis, A. Challinor and A. Lasenby, *Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models*, *Astrophys. J.* **538** (2000) 473 [astro-ph/9911177] [INSPIRE].

[53] PLANCK collaboration, *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [arXiv:1502.01589] [INSPIRE].

[54] N. Palanque-Delabrouille et al., *The extended Baryon Oscillation Spectroscopic Survey: Variability selection and quasar luminosity function*, *Astron. Astrophys.* **587** (2016) A41 [arXiv:1509.05607] [INSPIRE].

[55] S. Gontcho A Gontcho, J. Miralda-Escudé, A. Font-Ribera, M. Blomqvist, N.G. Busca and J. Rich, *Quasar — CIV forest cross-correlation with SDSS DR12*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 610 [arXiv:1712.09886] [INSPIRE].

[56] P. Laurent et al., *A 14 h^{-3} Gpc^3 study of cosmic homogeneity using BOSS DR12 quasar sample*, *JCAP* **11** (2016) 060 [arXiv:1602.09010] [INSPIRE].

[57] K.M. Górski et al., *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, *Astrophys. J.* **622** (2005) 759 [astro-ph/0409513] [INSPIRE].

[58] N. Kaiser, *Clustering in real space and in redshift space*, *Mon. Not. Roy. Astron. Soc.* **227** (1987) 1 [INSPIRE].

[59] A.M. Wolfe, D.A. Turnshek, H.E. Smith and R.D. Cohen, *Damped Lyman-Alpha Absorption by Disk Galaxies with Large Redshifts. I. The Lick Survey*, *Astrophys. J. Suppl.* **61** (1986) 249.

[60] A. Font-Ribera and J. Miralda-Escudé, *The effect of high column density systems on the measurement of the Lyman-α forest correlation function*, *JCAP* **07** (2012) 028 [arXiv:1205.2018] [INSPIRE].

[61] K.K. Rogers, S. Bird, H.V. Peiris, A. Pontzen, A. Font-Ribera and B. Leistedt, *Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers*, *Mon. Not. Roy. Astron. Soc.* **476** (2018) 3716 [arXiv:1711.06275] [INSPIRE].

[62] M. Pettini, L.J. Smith, D.L. King and R.W. Hunstead, *The Metallicity of High-Redshift Galaxies: The Abundance of Zinc in 34 Damped Lyα Systems from* $z = 0.7$ *to* 3.4, *Astrophys. J.* **486** (1997) 665 [astro-ph/9704102] [INSPIRE].

[63] J.X. Prochaska et al., *The UCSD HIRES/Keck I Damped Lyα Abundance Database. IV. Probing Galactic Enrichment Histories with Nitrogen*, *Publ. Astron. Soc. Pac.* **114** (2002) 933 [astro-ph/0206296] [INSPIRE].

[64] H. Padmanabhan, T.R. Choudhury and A. Refregier, *Modelling the cosmic neutral hydrogen from DLAs and 21-cm observations*, *Mon. Not. Roy. Astron. Soc.* **458** (2016) 781 [arXiv:1505.00008] [INSPIRE].

[65] I. Pérez-Ràfols, J. Miralda-Escudé, A. Arinyo-i Prats, A. Font-Ribera and L. Mas-Ribas, *The cosmological bias factor of damped Lyman alpha systems: dependence on metal line strength*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 4702 [arXiv:1805.00943] [INSPIRE].

[66] I. Pérez-Ràfols et al., *The SDSS-DR12 large-scale cross-correlation of damped Lyman alpha systems with the Lyman alpha forest*, *Mon. Not. Roy. Astron. Soc.* **473** (2018) 3019 [arXiv:1709.00889] [INSPIRE].

[67] J.X. Prochaska et al., *pyigm/pyigm: Initial release for publications*, (November 2017), DOI.

[68] J.X. Prochaska, P. Madau, J.M. O'Meara and M. Fumagalli, *Towards a unified description of the intergalactic medium at redshift* $z \approx 2.5$, *Mon. Not. Roy. Astron. Soc.* **438** (2014) 476 [arXiv:1310.0052] [INSPIRE].

[69] M.M. Pieri et al., *Probing the circumgalactic medium at high-redshift using composite BOSS spectra of strong Lyman* α *forest absorbers*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) 1718 [arXiv:1309.6768] [INSPIRE].

[70] M. Blomqvist et al., *The triply-ionized carbon forest from eBOSS: cosmological correlations with quasars in SDSS-IV DR14*, *JCAP* **05** (2018) 029 [arXiv:1801.01852] [INSPIRE].

[71] H. du Mas des Bourboux et al., *The Extended Baryon Oscillation Spectroscopic Survey: Measuring the Cross-correlation between the MgII Flux Transmission Field and Quasars and Galaxies at* $z = 0.59$, *Astrophys. J.* **878** (2019) 47 [arXiv:1901.01950] [INSPIRE].

[72] S.D. Landy and A.S. Szalay, *Bias and Variance of Angular Correlation Functions*, *Astrophys. J.* **412** (1993) 64 [INSPIRE].