Title
The Lyman-α forest from LBGs: First 3D correlation measurement with DESI and prospects for cosmology

Journal
Journal of Cosmology and Astroparticle Physics, 2025(12)

ISSN
1475-7516

Authors
Herrera-Alcantar, Hiram K
Armengaud, Eric
Yèche, Christophe
et al.

Publication Date
2025-12-01

DOI
10.1088/1475-7516/2025/12/053

Journal of Cosmology
and Astroparticle
Physics

PAPER • OPEN ACCESS

The Lyman-α forest from LBGs: First 3D correlation measurement with DESI and prospects for cosmology

To cite this article: Hiram K. Herrera-Alcantar *et al* JCAP12(2025)053

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Received: *July 30, 2025*
Revised: *October 15, 2025*
Accepted: *November 5, 2025*
Published: *December 22, 2025*

The Lyman-α forest from LBGs: First 3D correlation measurement with DESI and prospects for cosmology

Hiram K. Herrera-Alcantar et al.

Full author list at the end of the paper

E-mail: herreraa@iap.fr

Abstract: The Lyman-α (Lyα) forest is a key tracer of large-scale structure at redshifts $z > 2$, traditionally studied using the spectra of luminous but relatively rare quasars. In this work, we explore the viability of using the fainter yet significantly more abundant Lyman Break Galaxies (LBGs) as alternative background sources for Lyα forest studies. We analyze 4,151 Lyα forest skewers extracted from LBG spectra obtained in the DESI pilot surveys conducted in the COSMOS and XMM-LSS fields. From this dataset, we present the first measurement of the Lyα forest auto-correlation function derived exclusively from LBG spectra, probing comoving separations up to 48 h^{-1} Mpc at an effective redshift of $z_{\rm eff} = 2.70$. The measured LBG Lyα forest auto-correlation is consistent with that derived from DESI DR2 quasar Lyα forest spectra at a comparable redshift, validating the use of LBGs as reliable background sources for Lyα forest analyses. In addition, we measure the cross-correlation between the LBG Lyα forest and the positions of 13,362 galaxies, demonstrating that this observable serves as a sensitive diagnostic for assessing the precision and accuracy of galaxy redshift estimates, and for identifying and correcting systematic offsets. Finally, using both synthetic LBG spectra and Fisher matrix forecasts, we show that a future wide-area survey covering ~5,000 deg^2, targeting 1,000 LBGs per square degree at signal-to-noise levels comparable to our sample, could enable LBG-based Lyα forest baryon acoustic oscillation (BAO) measurements with expected uncertainties of $\sigma_{\alpha_{\rm ISO}} = 0.4\%$ (isotropic) and $\sigma_{\alpha_{\rm AP}} = 1.3\%$ (Alcock-Paczynski). This performance is further enhanced when combining the BAO analysis with a Lyα forest Full Shape (FS) approach, yielding a predicted uncertainty of $\sigma^{\rm FS}_{\alpha_{\rm AP}} = 0.6\%$. These results open a new avenue for precision cosmology at high redshift using the Lyα forest in dense LBG samples.

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

ArXiv ePrint: 2507.21852

https://doi.org/10.1088/1475-7516/2025/12/053

JCAP12(2025)053

Contents

JCAP12(2025)053

1 Introduction

The large-scale statistical properties of the matter distribution encode a wealth of information about our Universe, tracing its expansion history, primordial fluctuations and non-Gaussianities, and the growth of cosmic structure and the formation of galaxies. For that reason, a number of surveys such as DESI [2, 3], EUCLID [4], PFS [5], and WEAVE [6] are ongoing or planned to measure these cosmological structures over a wide range of scales and redshifts. At redshifts $2 < z \lesssim 4$, a remarkably powerful probe is the Lyman-α (Lyα) forest. The Lyα forest is a set of absorption features imprinted onto the optical spectra of background sources. These arise from resonant Lyman-α absorption and scattering due to the presence of neutral hydrogen along the line-of-sight. Given that most of this Lyα forest signal comes from regions located in the diffuse intergalactic medium (IGM) [7], these spectroscopic features encode information on the physical properties of this medium [8].

At the same time, the Lyα forest can be considered as a biased tracer of the large-scale matter distribution. The 3D correlations between absorption features from different lines-of-sight, first measured in [9], can be interpreted using a linear clustering model, valid on large, linear scales, and incorporating relevant astrophysical and instrumental contaminants. As a consequence, the baryon acoustic oscillation (BAO) feature at a typical redshift of $z \sim 2.3$ could be measured from the large Lyα forest samples collected by the BOSS and eBOSS surveys [10–12] and more recently by the DESI survey [13, 14]. With the highest precision achieved on the isotropic BAO distance scale delivered by DESI data release 2 (DR2) results ([14], hereafter `DESIDR2-Ly`α) at the 0.7 % level. It should be noted that these BAO measurements combine two correlation signals: the auto-correlation of the Lyα forest, and the cross-correlation between Lyα absorption and the spatial distribution of galaxy tracers, as pioneered in [15].

Until now, most of the background sources used to carry out Lyα forest observations were quasars (also denoted as QSOs). Indeed quasars are among the brightest optical sources in the Universe, with a strong continuum flux in the rest-frame UV wavelength range. Therefore spectra with a good signal-to-noise ratio can be measured from those sources up to high redshift, in relatively short exposure times with existing telescopes. However quasars are rare objects in our Universe, especially when restricting to those with redshift $z > 2.1$, which have their Lyα forest in the optical observed wavelength range. The quasar luminosity function was measured over a wide range of redshift and magnitude from SDSS spectroscopic observations in [16, 17]: predictions from models fitted in [17] indicate that, in the redshift range $2 < z < 3$, there are only ~ 140 quasars per deg^2 with magnitude $r < 24$, and down to the deeper limit of $g < 25$ the expected number count is ~ 230 per deg^2. In practice, the DESI survey obtained spectra for 60 high-redshift ($z > 2.1$) quasars per deg^2, using a photometric selection [18] limited to $r < 23$. Such a sky density, while unprecedented over a wide footprint, remains a limiting factor for several scientific applications. For instance, it restricts measurements of line-of-sight absorption cross-correlations between close quasar pairs [19, 20], and also affects the detection of the BAO feature, which remains in the shot-noise-limited regime for DESI [21].

In order to significantly increase the sky density of Lyα forest samples, background sources other than quasars are needed. Lyman Break Galaxies (LBGs) are prime candidates, as these star-forming galaxies are among the most readily detected at high redshifts. Their distinctive break at the Lyman limit (912 Å) enables efficient selection from broadband photometry [22]. However, LBGs are typically fainter than quasars, so extracting the Lyα forest signal from their spectra requires longer exposure times, larger telescopes, or both. Despite these challenges, the Lyα forest in high signal-to-noise ratio LBG spectra has already been exploited in relatively small-area surveys such as CLAMATO [23, 24] and LATIS [25]. While the primary goal of these programs was to map the three-dimensional distribution of the IGM at high redshift and to identify structures in the cosmic web (e.g., protoclusters, voids, filaments), they also enabled measurements of the Lyα-galaxy cross-correlation at $z > 2$ on scales $r < 30\ h^{-1}$ Mpc by combining forests extracted from both quasars and LBGs [e.g., 26–29]. These efforts complement works probing the Lyα-galaxy connection based exclusively on quasar forests [e.g., 30–33].

JCAP12(2025)053

In this article, we extend these efforts by demonstrating the potential of using the Lyα forest in LBG spectra alone as an independent tracer of large-scale structure in the context of wide-area spectroscopic surveys. We use optical spectra of high-redshift galaxies collected during DESI pilot observations in the COSMOS and XMM-LSS fields, as described in Section 2. In Sections 3 and 4, we present the first 3D Lyα forest auto-correlation measurement using only LBG forests, covering scales from 1 to 48 h^{-1} Mpc, along with the Lyα cross-correlation with both LBGs and Lyman-α emitters (LAEs). We show that these measurements can be described with the same modeling framework applied to DESI Lyα forest correlations from quasars. Finally, in Section 5, we provide forecasts for high-redshift BAO measurements, highlighting that future wide-area surveys with DESI-like capabilities can exploit the Lyα forest in LBG spectra to deliver competitive cosmological constraints at redshifts beyond those accessible with quasars alone.

2 Dataset

DESI is a fully robotic, highly multiplexed spectroscopic instrument installed on the 4-meter Mayall Telescope at Kitt Peak National Observatory [34]. It delivers simultaneous spectra for up to 5,000 targets over a 3.2° diameter field of view [3, 35–37]. The majority of DESI's dark-time observing program [38] is dedicated to its main survey, which aims to acquire spectra for roughly 50 million galaxies, including several million quasars at $z \gtrsim 1.6$. In addition, DESI has conducted dedicated observing campaigns in the COSMOS and XMM-LSS (hereafter XMM) fields during both the survey validation and main-survey phases [39], including targeted pilot programs to test and refine observation strategies for high-redshift galaxies such as Lyman-break galaxies (LBGs) and Lyman-α emitters (LAEs).

In the following section, we describe the target selection strategies and the automated redshift estimation pipeline used to construct the LBG and LAE samples for this work. Preliminary analyses of the Ly$\alpha\times$LBG and Ly$\alpha\times$LAE cross-correlations revealed systematic redshift offsets of $\Delta v_{\rm LBG} = -169 \pm 22$ km s^{-1} for LBGs and $\Delta v_{\rm LAE} = -241 \pm 20$ km s^{-1} for LAEs. To mitigate potential biases in the clustering signal, such as a shift in the cross-correlation along the line-of-sight due to these offsets, we apply a correction to the redshift estimates in all subsequent measurements. Details of these corrections are provided in appendix A.

2.1 Target selection

LBG spectra exhibit a pronounced flux "break" just blueward of the 912 Å Lyman limit in the rest frame, primarily caused by absorption from neutral hydrogen in stellar atmospheres and the interstellar medium (ISM). At high redshifts ($z > 2$), additional Lyα forest absorption caused by the intergalactic medium, further suppresses the flux blueward of 1216 Å. These characteristic features enable efficient identification of Lyman-break galaxies (LBGs) at $z \sim 2.3$ via broad-band photometry, particularly through dropout selection in the u-band.

Our target selection employs the u-dropout technique, based on deep imaging from the Hyper Suprime-Cam survey [HSC; 40], complemented by u-band imaging from the Canada-France-Hawaii Telescope's CFHT Large Area U-band Deep Survey [CLAUDS; 41], obtained with MegaCam.

We adopt two complementary u-dropout selection strategies. The first, described in [22] (hereafter R24 [22]), applies color cuts in the $u-g$ vs $g-r$ plane, alongside a magnitude threshold in the r band. The second approach, presented in [42], utilizes a random forest algorithm trained on HSC+CLAUDS *ugriz* photometry artificially degraded to the depth of the Ultraviolet Near-Infrared Optical Northern Survey [UNIONS; 43], forecasting selection performance for upcoming surveys such as DESI-II.

A smaller subset of our sample is selected using medium-band photometry. In this case, Lyman-alpha emitters (LAEs) are identified by flux excesses in narrow-band filters, while LBGs are selected based on flux decrements between adjacent bands. We emphasize that throughout this work, the term "LAE" *refers exclusively to the photometric selection technique*, rather than to the strength of the Lyα emission.

All LBG and LAE targets presented here were observed in DESI pilot programs conducted between 2021 and 2024 in the COSMOS and XMM fields, with effective exposure times ranging from 2 to 5 hours. In our Lyα forest analysis (see Section 3), only the LBGs serve as background sources. The median r-band magnitude of the LBG sample is $r = 23.94$.

2.2 Automated redshift measurement

Spectra of the LBG and LAE targets were obtained with the DESI 3-arm spectrographs, which cover the wavelength range 3,600-9,800 Å with a spectral resolution between 2,000 and 5,000 [34]. Given the redshift range of these sources, most of the key spectral features fall within the blue arm (3,600-5,930 Å), where the resolution varies from 2,000 to 3,000. Raw spectra were processed using the DESI spectroscopic pipeline [44], and multiple exposures of the same object were coadded with weights proportional to their signal-to-noise ratio (SNR). The vast majority of the spectra were obtained using a dithering strategy, in which each target was typically observed multiple times with different fibers (and, in some cases, different spectrographs) across overlapping tiles following a rosette pattern. This configuration averages out potential systematics from sky-line subtraction residuals and minimizes fiber and spectrograph-dependent effects, which reduces possible biases in the clustering measurements.

Redshift estimation proceeds in two stages. First, a convolutional neural network (CNN) identifies whether the spectrum corresponds to an LBG or LAE and assigns a preliminary redshift. Second, a template-fitting algorithm refines the redshift determination.

The CNN is a modified version of `QuasarNET` [45, 46], which DESI uses for quasar classification. For LBG classification, we retrained the CNN using 14 interstellar absorption lines and 2 emission lines (see table 3 in R24 [22]). Each of the 16 lines are assigned a CNN confidence level (CL) and are sorted in descending order. A spectrum is classified as an LBG or LAE if its fifth-highest CL exceeds a chosen threshold. Additional details on the CNN algorithm and performance can be found in section 4.1 of R24 [22].

To refine spectral classification and the redshift estimate, the redshift of the line with the highest CNN confidence is used as a prior for the `redrock` software,[1] DESI's standard classifier and redshift-fitting tool. We employ a customized version of `redrock` that includes four LBG templates derived from visually confirmed spectra in R24 [22]: one template without Lyα emission and three with increasing Lyα emission strength.

[1] https://github.com/desihub/redrock [47]

To quantify the quality of the redshift measurements of our COSMOS LBG sample, we cross-matched our dataset with the high-confidence (CL > 0.95) spectroscopic redshifts from the COSMOS spectroscopic redshift compilation [48]. We define catastrophic failures as sources with a velocity offset $|\Delta v| > 3,000$ km s^{-1} relative to the reference spectroscopic redshift. The resulting outlier rates are 16.5%, 13.3%, 11.1%, and 8.3% for thresholds of CL $> 0.97, 0.99, 0.995$, and 0.999, respectively. These values are consistent with those reported in table 4 of `R24` [22] and confirm that higher CL cuts substantially reduce the catastrophic outlier rate, although a non-negligible fraction remains even at our strictest cut. As will be noted in Section 4.2, sample purity and catastrophic redshift outliers directly impact the inferred clustering amplitude on the correlation functions. Therefore, it will be an important consideration for future large-scale clustering analyses with LBG forests. We return to this point in Section 6.

For the non-catastrophic objects, we find excellent agreement with the reference catalog, with a median velocity offset of only ~ 8 km s^{-1} and a scatter of $\sigma_{\rm MAD} = 1.4826 \times {\rm MAD}(|\Delta v|) \sim 180$ km s^{-1}. This demonstrates that, aside from the catastrophic failures, the redshifts of our CNN+`redrock` pipeline are robust and suitable for clustering analyses.

For our baseline analysis, we adopt a confidence threshold of CL > 0.995 to enhance sample purity while preserving statistical power. We also consider thresholds in the range $0.97 <$ CL < 0.999 for robustness checks. Table 1 summarizes the number of LBGs, LAEs, and accepted Lyα forests (see Section 3) in COSMOS and XMM as a function of CL threshold. At CL > 0.995, the sample comprises 4,931 LBGs, 8,431 LAEs, and 4,151 forests; the angular distribution of the LBG targets is illustrated in Figure 1. The survey footprints correspond to approximately 6 and 5 deg^2 in COSMOS and XMM, respectively.

We note that approximately 70% of LBGs in our sample are in COSMOS, while 70% of LAEs are in XMM. This imbalance in the field distributions is a direct result of the strategy employed by the DESI pilot survey campaigns. Different targeting strategies were tested in each field; specifically, the COSMOS field observations were optimized for LBG selection to enable future Lyα tomography studies, while LAEs were targeted mainly in XMM where medium-band imaging was available first. Future DESI special programs are planned to observe both fields with more uniform strategies, which will provide larger and more balanced samples for future analyses.

3 Data analysis

One of the main goals of this work is to present the first measurement of the Lyα forest auto-correlation (Ly$\alpha\times$Lyα) extracted solely from LBG spectra, together with the corresponding cross-correlations between Lyα absorption and LBG (Ly$\alpha\times$LBG) and LAE (Ly$\alpha\times$LAE) positions. Our analysis builds on the `DESIDR2-Ly`α [14] framework for quasar-forest studies, implemented via the public `picca` package.[2]

Below, we outline the steps of our correlation-function measurement pipeline, highlighting the minor differences with respect to the quasar-based pipeline. For further methodological details, we refer the reader to the `DESIDR2-Ly`α [14] reference and to section 3 of [50].

[2] `https://github.com/igmhub/picca` [49]

JCAP12(2025)053

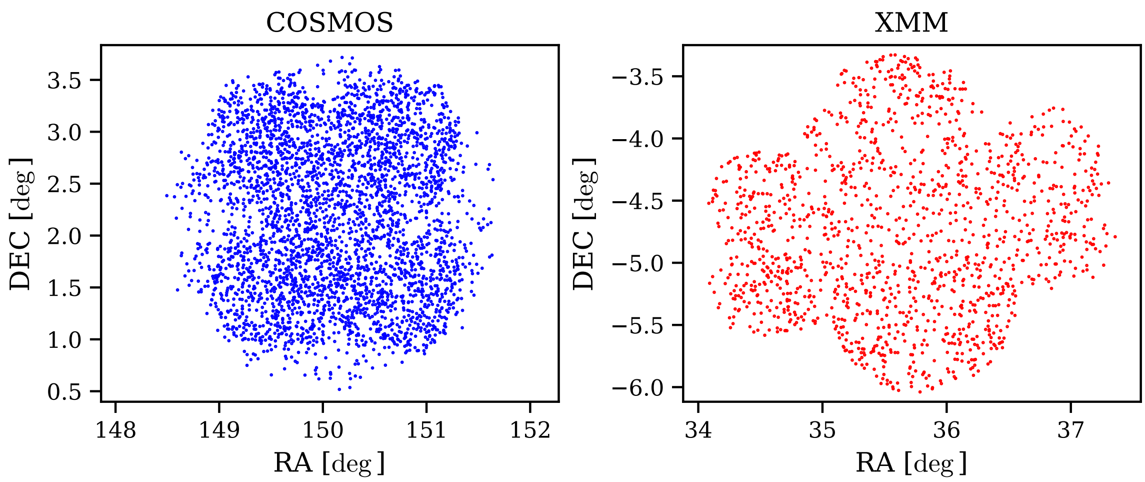

Figure 1. Angular positions of the CL > 0.995 LBG sample in the COSMOS (left) and XMM (right) fields used for our analysis. The COSMOS footprint covers approximately 6 deg^2, while the XMM footprint spans about 5 deg^2.

CNN Confidence Level	Survey	LBG tracers	LAE tracers	Lyα Forests (Section 3)
CL > 0.97	COSMOS	4,847	3,495	4,117
	XMM	1,988	7,100	1,555
CL > 0.99	COSMOS	3,951	2,733	3,395
	XMM	1,598	6,430	1,258
CL > 0.995	COSMOS	3,517	2,416	3,027
	XMM	1,414	6,015	1,124
CL > 0.999	COSMOS	2,714	1,848	2,336
	XMM	1,080	4,969	859

Table 1. Statistics of LAEs, LBGs, and selected Lyα forest skewers for different CL cuts.

3.1 Continuum fitting and extraction of transmitted flux field

The first step in our analysis is to preprocess the spectra. Following `DESIDR2-Ly`α [14], this normally comprises a series of operations but we do not apply all of them here. We mask bad pixels identically to `DESIDR2-Ly`α [14], removing regions affected by galactic absorption, residual sky-line emission, and CCD defects flagged by the DESI reduction pipeline [44] (e.g., cosmic ray contamination or hardware-related issues). On the other hand, we do not recalibrate the fluxes as the redshift range ($2.0 < z < 3.5$) of our LBG sample is not wide enough to completely cover a calibration region (typically C III; see table 3 of [50]) in the 3,600-5,423 Å observed frame wavelength range that we have adopted for our analysis.

Unlike quasar forest analyses, we do not mask Damped Lyα Absorbers (DLAs) or Broad Absorption Line (BAL) features. BALs are quasar-specific phenomena and are not expected to appear in LBG spectra. DLAs, however, are present in LBG spectra and are expected to impact our measurements. Unfortunately, no dedicated DLA catalog exists for our LBG sample. Existing DLA detection algorithms, developed for quasar spectra, cannot be straightforwardly applied here due to the differing SNR characteristics and different spectral shapes LBGs with respect to quasars. Furthermore, DLAs in low-SNR quasar spectra (SNR < 3) are typically not masked in quasar forest analyses (e.g., DESIDR2-Lyα [14]), and most of our LBG spectra fall below this SNR threshold. Adapting any of the existing quasar-based DLA finders [e.g., 51–53] is beyond the scope of this work.

After spectral preprocessing, we extract the transmitted-flux field,

$$\delta_g(\lambda) = \frac{f_g(\lambda)}{\overline{F}(\lambda) C_g(\lambda)} - 1, \tag{3.1}$$

where g indexes each LBG, $f_g(\lambda)$ is the observed flux, $\overline{F}(\lambda)$ is the mean transmission, and $C_g(\lambda)$ is the unabsorbed LBG rest-frame spectrum, hereafter referred to as the continuum. We limit our analysis to the $1040 - 1205$Å rest-frame wavelength interval and the $3600 < \lambda_{\rm obs} < 5423$ Å observed wavelength range, which corresponds to a redshift range of $1.98 < z < 3.5$.

As in quasar analyses, we model $\overline{F}(\lambda) C_g(\lambda)$ by a universal rest-frame continuum, $\overline{C}_{\rm LBG}(\lambda_{\rm RF})$, modified by a first-degree polynomial in $\Lambda = \log\lambda$ with two free parameters, a_g and b_g,

$$\overline{F}(\lambda)\ C_g(\lambda) = \overline{C}_{\rm LBG}(\lambda_{\rm RF}) \left(a_g + b_g \frac{\Lambda - \Lambda_{\rm min}}{\Lambda_{\rm max} - \Lambda_{\rm min}} \right). \tag{3.2}$$

We determine $\overline{C}_{\rm LBG}$, a_g, and b_g through an iterative likelihood maximization process, simultaneously fitting the pixel-variance,

$$\sigma_g^2 = \eta(\lambda)\ \sigma^2_{{\rm pipe},g}(\lambda) + \sigma^2_{\rm LSS}(\lambda) \left[\overline{F}(\lambda) C_g(\lambda)\right]^2, \tag{3.3}$$

where $\sigma^2_{{\rm pipe},g}(\lambda)$ is the reported variance from the data processing pipeline [44], $\eta(\lambda)$ corrects for small inaccuracies in the noise estimate,[3] and $\sigma^2_{\rm LSS}(\lambda)$ is a free parameter that captures the intrinsic Lyα flux variance. Because our LBG spectra is dominated by noise, we were unable to robustly constrain $\sigma_{\rm LSS}$ in our analysis. We nevertheless include this term to maintain a consistent methodology with the DESI Lyα quasar analyses and verified that omitting it leads to negligible differences in the measured correlation functions. Future, larger LBG spectroscopic samples should enable a more precise determination of $\sigma_{\rm LSS}$.

Additionally, we apply quality cuts to remove forest skewers with (i) negative fitted continuum, (ii) negative mean SNR in the forest region, or (iii) fewer than 150 valid pixels. This yields 3,027 accepted forests in COSMOS and 1,124 in XMM. The median SNR of spectra in the forest region is 0.35 Å^{-1}, which is about 4.5 times lower than the median SNR of 1.57 Å^{-1} found in the quasar Lyα forests sample used in DESIDR2-Lyα [14]. The distribution of SNR as a function of r-band magnitude is shown in Figure 2.

Figure 3 shows the mean rest-frame continuum $\overline{C}_{\rm LBG}$ in both fields, revealing excellent agreement between COSMOS and XMM. Moreover, both samples exhibit the characteriztic absorption features from stellar and interstellar lines, consistent with those identified in

[3]The DESIDR2-Lyα [14] pipeline treats $\eta(\lambda)$ as a free parameter and found values close to unity with small fluctuations. Given the limited size of our dataset, we decided to fix $\eta(\lambda) = 1$.

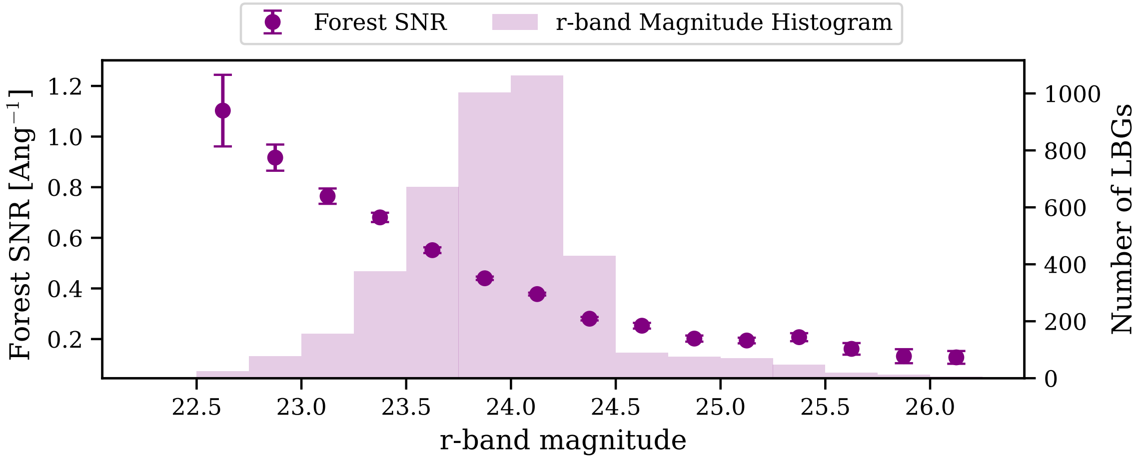

Figure 2. Filled histogram: r-band magnitude distribution of LBGs with CL > 0.995 in the COSMOS and XMM fields, whose Lyα forests pass our quality cuts. The mean forest SNR as a function of r-band magnitude is shown with purple points; error bars represent the standard error of the mean. The effective exposure times for these observations range from 2 to 5 hours, with a median value of 4.5 hours. A small number of bright LBGs have longer exposures due to repeated observations.

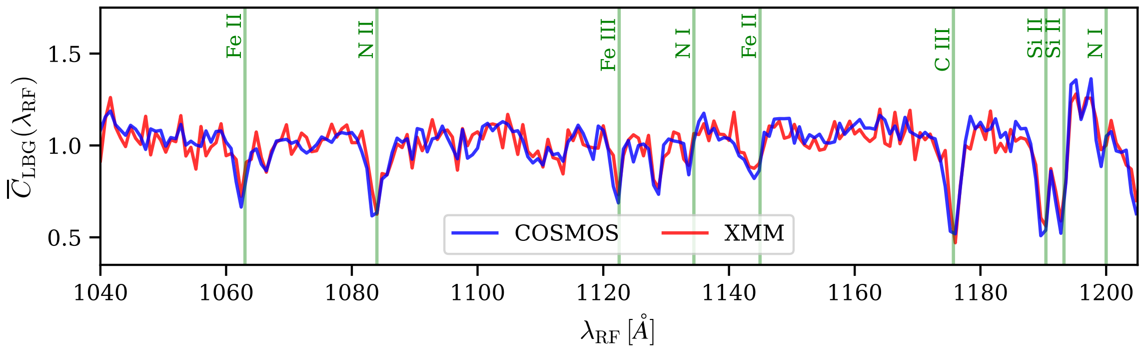

Figure 3. Mean rest-frame continuum in the Lyα region as measured from LBG spectra on the COSMOS (blue line) and XMM (red line) fields. The green shaded vertical lines indicate known spectral line features caused by stellar and interstellar absorption. By construction from the continuum fitting procedure, $\overline{C}_{\rm LBG}$ is arbitrarily normalized.

composite LBG spectra from previous studies [e.g., 22, 25, 54, 55]. Notably, the continuum-fitting procedure as implemented in `picca`, originally developed for quasar Lyα forest analyses, successfully recovers these features in LBG spectra. This demonstrates its robustness and adaptability to a broader class of high-redshift background sources.

3.2 Correlation function measurement

We estimate both the Lyα auto-correlation of the transmitted-flux field $\delta_g(\lambda)$ (described in Section 3.1) and its cross-correlation with the LBG/LAE tracer catalogs introduced in Section 2. Pair separations, either between Lyα pixels in the auto-correlation or between Lyα pixels and galaxies in the cross-correlation, are binned in the line-of-sight ($r_\parallel$) and transverse

($r_\perp$) directions. To compute these separations, angular and redshift differences are converted into comoving distances using the fiducial cosmology of [56].

For both correlation functions, we adopt $\Delta R = 2\ h^{-1}\,$Mpc bins spanning the range $0 < r_\parallel, r_\perp < 48\ h^{-1}\,$Mpc for the auto-correlation, and $0 < |r_\parallel|, r_\perp < 48\ h^{-1}\,$Mpc for the cross-correlations. In the latter case, negative $r_\parallel$ values correspond to configurations where the tracer galaxies are located behind the Lyα pixels along the line of sight.

The auto-correlation in a $M \equiv (r_\parallel, r_\perp)$ bin, is defined as

$$\xi_M^{\alpha\times\alpha} = \frac{\sum_{i,j\in M} w_i^\alpha\ w_j^\alpha\ \delta_i\ \delta_j}{\sum_{i,j\in M} w_i^\alpha\ w_j^\alpha}, \tag{3.4}$$

where $\delta_i \equiv \delta_g(\lambda_i)$, and each Ly$\alpha$ pixel weight, w_i^α, is

$$w_i^\alpha \equiv w(\lambda_i) = \left(\frac{1+z_i}{1+2.25}\right)^{\gamma_\alpha - 1} \left[\eta(\lambda_i)\ \tilde{\sigma}^2_{\mathrm{pipe},g}(\lambda_i) + \eta_{\mathrm{LSS}}\ \sigma^2_{\mathrm{LSS}}(\lambda_i)\right]^{-1}, \tag{3.5}$$

where $\tilde{\sigma}^2_{\mathrm{pipe},g}(\lambda) \equiv \sigma^2_{\mathrm{pipe},g}(\lambda)/[\overline{F}(\lambda)C_g(\lambda)]^2$, $\gamma_\alpha = 2.9$ [57], $\eta(\lambda) = 1$, and η_{LSS} modulates the contribution of the intrinsic Lyα forest flux variance.[4]

Similarly, the Lyα cross-correlation with the galaxy tracer X is

$$\xi_M^{\alpha\times\mathrm{X}} = \frac{\sum_{i,j\in M} w_i^\alpha\ w_j^\mathrm{X}\ \delta_i}{\sum_{i,j\in M} w_i^\alpha\ w_j^\mathrm{X}}, \tag{3.6}$$

with galaxy tracer weights, w_i^X, given by

$$w_i^\mathrm{X} = \left(\frac{1+z_i}{1+2.25}\right)^{\gamma_\mathrm{X} - 1}. \tag{3.7}$$

In this analysis, we adopt $\gamma\mathrm{X} = 1.44$, following the convention commonly used for quasars [59]. While more detailed models exist for the redshift and magnitude dependence evolution of dropout-selected galaxies (e.g., Equation 2.7 in [60]) we opt here for this simplified prescription. Incorporating a more refined evolution model is left for future work.

We estimate the covariance matrix of our correlation measurements by dividing the survey footprint into HEALPix pixels with $N_{\mathrm{side}} = 64$ [61], each covering approximately $0.84\ \mathrm{deg}^2$, which corresponds to a comoving scale of $\sim 70\ h^{-1}\,$Mpc at $z = 2.7$. We compute the correlation function independently within each HEALPix region and derive the covariance matrix using a bootstrap resampling approach, drawing 10,000 realizations from the set of HEALPix pixel subsamples. The estimated covariance is then smoothed following the same recipe as described in section 3.3 of [13].

Due to the relatively small footprint of our analysis (and hence the limited number of independent samples) the estimated covariance may be underestimated. This limitation could be mitigated in future work by analyzing larger sky areas or by extracting the covariance from several realizations of synthetic LBG spectra datasets.

[4] In `DESIDR2-Ly`α [14], η_{LSS} was set to 7.5 based on results from the DESI Early Data Release (EDR) Lyα analysis to minimize the quasar Lyα forest auto-correlation covariance [50, 58]. Here, we fix $\eta_{\mathrm{LSS}} = 1$ as determining the optimal value for LBG forests requires further study.

3.3 Model of the correlation functions

Deriving tight model constraints is not the primary objective of this study. Nevertheless, we include a model of the correlation functions, analogous to that used by `DESIDR2-Ly`α [14], for completeness. Our analysis employs an anisotropic cross-power spectrum model,

$$P^{\alpha\times\mathrm{Y}}(\mathbf{k},z) = b_\alpha b_\mathrm{Y}\left(1+\beta_\alpha\mu_k^2\right)\left(1+\beta_\mathrm{Y}\mu_k^2\right)P_L(k,z)G(\mathbf{k})F_\mathrm{NL}(\mathbf{k}), \tag{3.8}$$

where Y denotes either Lyα or a galaxy tracer, and $\mathbf{k}\equiv(k_\parallel,k_\perp)$ is the wave-number vector with magnitude $k=\sqrt{k_\parallel^2+k_\perp^2}$ and $\mu_k = k_\parallel/k$ representing the cosine of the angle between $\mathbf{k}$ and the line-of-sight. In this model, $P_L(k,z)$ is the linear matter power spectrum computed using `CAMB`[62] under the fiducial cosmology of [56], $G(\mathbf{k}) = \mathrm{sinc}\left(k_\parallel\Delta R/2\right)\mathrm{sinc}\left(k_\perp\Delta R/2\right)$ accounts for the finite binning $\Delta R = 2\ h^{-1}\,\mathrm{Mpc}$ of the correlation functions, and $F_\mathrm{NL}(\mathbf{k})$ captures non-linear corrections at small scales.

In the auto-correlation, F_NL accounts for non-linear structure growth, peculiar velocities, and IGM thermal effects (e.g., broadening and pressure) which are incorporated as specified in Equation 3.6 of [63], with parameters $q_1 = 0.8424$, $q_2 = 0$, $k_v = 0.69516\,h\,\mathrm{Mpc}^{-1}$, $a_v = 0.4015$, $b_v = 1.711$, and $k_p = 21.13\,h\,\mathrm{Mpc}^{-1}$ obtained by interpolating the best-fit model from the L160R25 ACCEL2 hydrodynamical simulation at $z\sim 2.7$ (see table A3 of [64]). For the cross-correlation, the combined effects of galaxy velocity dispersion and statistical redshift estimation errors are modeled by a Lorentzian profile:

$$F_\mathrm{NL}^\mathrm{cross}(\mathbf{k}) = \left[1+k_\parallel^2\sigma_{z,\mathrm{X}}^2\right]^{-\frac{1}{2}}, \tag{3.9}$$

where $\sigma_{z,\mathrm{X}}$ is a free parameter representing the velocity dispersion of the tracer due to both peculiar velocities and redshift errors. In addition, to account for systematic redshift offsets that shift the cross-correlation along the line-of-sight, we introduce a free parameter, $\Delta r_{\parallel,\mathrm{X}}$, which quantifies this shift.

We model the redshift evolution of the linear biases for Lyα and galaxies using a power law, $b_\mathrm{Y}(z) = b_\mathrm{Y}(z_\mathrm{eff})\left[(1+z)/(1+z_\mathrm{eff})\right]^{\gamma_\mathrm{Y}}$, with $\gamma_\alpha = 2.9$ for Lyα and $\gamma_\mathrm{X} = 1.44$ for galaxy tracers, as in section 3.2. For galaxy tracers, the RSD parameter is derived from the relation $\beta_\mathrm{X} = f(z)/b_\mathrm{X}(z)$, where $f(z)\sim\Omega_m^{0.55}(z)$ is the linear growth rate. In our analysis, $f(z)$ is fixed to its value at $z\sim 2.7$ (i.e, $f\sim 0.98$). Note that this relation is not applied to the Lyα forest, for which β_α is treated as independent, though we approximate it as constant in redshift for simplicity, despite evidence of redshift dependence [63, 64].

Additionally, we account for contributions to the correlation functions from contaminants to the Lyα forest such as unmasked High Column Density systems (HCDs), absorption from other atomic transition lines blended within the Lyα forest (hereafter referred to as metals), and correlated noise. These contributions, if not properly modeled, could bias the Lyα forest parameter measurements. The procedures used to incorporate these effects are described in Section 3.3.1.

The final power spectrum model is decomposed into multipoles up to $\ell = 6$ and transformed into correlation function multipoles via a Hankel transform. The resulting model is then interpolated onto an $(r_\parallel, r_\perp)$ grid to yield the following forms for the correlation functions:

$$\xi^{\alpha\times\alpha}_{\rm Total} = \xi^{\alpha\times\alpha} + \sum_m \xi^{\alpha\times m} + \xi^{\rm noise}, \tag{3.10}$$

$$\xi^{\alpha\times{\rm X}}_{\rm Total} = \xi^{\alpha\times{\rm X}} + \sum_m \xi^{{\rm X}\times m}, \tag{3.11}$$

where each m labels a distinct metal transition. Finally, distortions in the correlation functions arising from the continuum fitting procedure are incorporated via a distortion matrix, following the approach described in section III.A of `DESIDR2-Ly`α [14] and in [65].

3.3.1 Contaminants Model

High Column Density systems. HCDs, including Lyman-limit systems (LLS) and DLAs, are biased tracers of the underlying matter density field. When blended into the Lyα forest, their presence introduces a scale-dependent excess of large-scale power into the 3D Lyα forest correlations [66–68], boosting the inferred Lyα linear bias and suppressing the RSD parameter [69]. To account for these effects, we model the Lyα bias and RSD as “effective” parameters that include a scale-dependent HCD contribution:

$$b'_\alpha = b_\alpha + b_{\rm HCD} F_{\rm HCD}(k_\parallel), \tag{3.12}$$

$$b'_\alpha \beta'_\alpha = b_\alpha \beta_\alpha + b_{\rm HCD} \beta_{\rm HCD} F_{\rm HCD}(k_\parallel), \tag{3.13}$$

where $b_{\rm HCD}$ and $\beta_{\rm HCD}$ are the linear bias and RSD parameter associated with the HCD contribution to the transmitted-flux fraction. Note that $b_{\rm HCD}$ is negative and small in this context, as it describes absorption in the transmitted-flux field, rather than the clustering bias of the HCD systems themselves, which is typically ~ 2 [70, 71]. The term $F_{\rm HCD}(k_\parallel) = \exp\left(-L_{\rm HCD} k_\parallel\right)$ describes the scale-dependent contribution of HCDs to the correlation function, with $L_{\rm HCD}$ representing their typical length scale.

Other atomic transitions (metals). Our pipeline assumes that every absorption in the analysis region arises from Lyα but in reality metal absorption blended into the forest introduce spurious correlations most noticeable along the line-of-sight. We include this contamination via metal cross-terms, $\xi^{\alpha\times m}$ in the auto-correlation and $\xi^{{\rm X}\times m}$ in the cross-correlation. Given the scales relevant to our analysis, we model only the contribution from the Si III(1207) transition, treating its linear bias, $b_{\rm Si\,III(1207)}$, as a free parameter and fixing its RSD parameter to $\beta_{\rm Si\,III(1207)} = 0.5$. We neglect metal auto-correlations and cross-correlation between distinct metals as they are subdominant. Strong transitions such as C IV, Si IV, and Mg II can in principle significantly contribute through their auto-correlation [72]; however, preliminary analyses did not show a robust signal detection.

Correlated noise. The DESI data reduction pipeline introduces correlated noise between fibers on the same focal-plane petal [44, 72], which contaminates the Lyα forest auto-correlation. We model this contribution as an additive term,

$$\xi^{\rm noise} = a_{\rm noise} \delta^K(r_\parallel) f(r_\perp), \tag{3.14}$$

where $a_{\rm noise}$ is a free parameter that sets the amplitude of the noise contribution, δ^K is the Kronecker delta in $r_\parallel$, and $f(r_\perp)$ is a decreasing function of $r_\perp$, normalized such that $f(0) = 1$, proportional to the probability that two pixels originate from the same petal.

JCAP12(2025)053

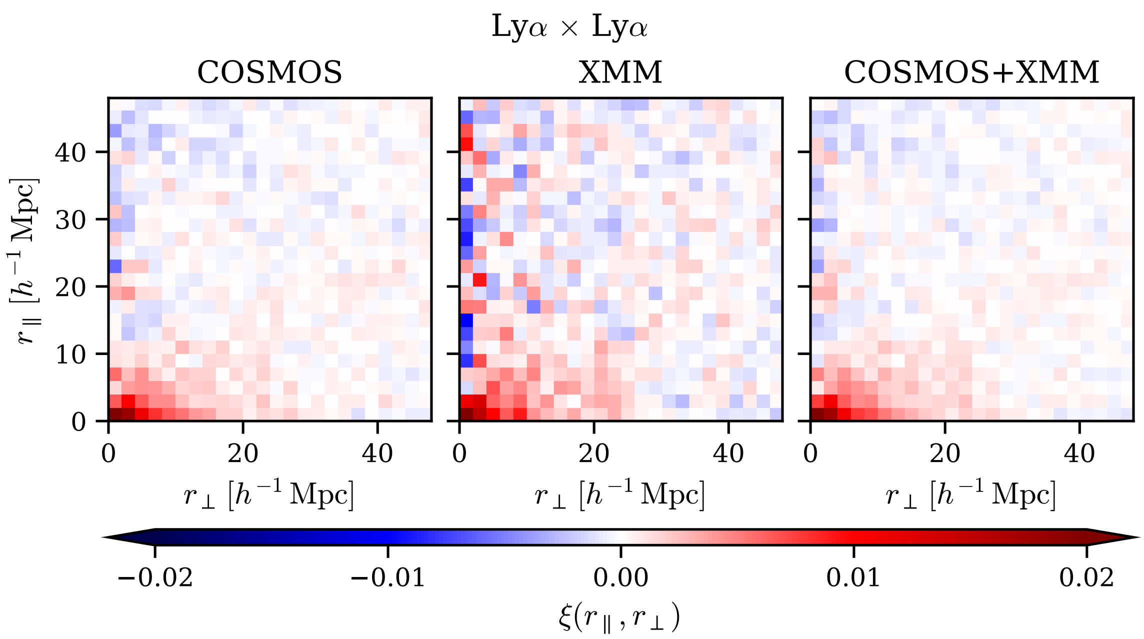

Figure 4. Lyα×Lyα auto-correlation measured from LBG spectra with CL $>$ 0.995 in COSMOS (left), XMM (center), and the combined fields (right), shown over a $(r_\parallel, r_\perp)$ grid. While some visual differences are apparent between COSMOS and XMM, we have verified that the measurements are statistically consistent; these differences are likely due to noise and statistical fluctuations.

4 Results and discussion

4.1 Correlation functions

Now we focus on the Lyα forest auto and cross-correlations with LBG and LAE positions results. Figure 4 shows the LBG forest Lyα×Lyα auto-correlation as a function of $r_\parallel$ and $r_\perp$ in the COSMOS (left) and XMM (center) fields, using LBG spectra with CNN confidence level CL $>$ 0.995. The effective redshift for this measurement is $z_{\rm eff} = 2.70$. Some visual discrepancies appear between the COSMOS and XMM measurements. However, these are likely driven by statistical fluctuations and noise. We have verified that the measurements are consistent within uncertainties. Because the mean rest-frame continuum (see Section 3.1), signal-to-noise ratios, magnitude ranges, and redshift distributions are consistent across both fields, we combine them to improve statistical power. The right panel shows results for the combined sample.

For illustrative purposes, Figure 5 displays the same Lyα auto-correlation signal from COSMOS+XMM, averaged over the first five bins in the transverse (left) and parallel (right) directions. For comparison, we include the Lyα auto-correlation measurement from ∼766,000 quasar forests in the `DESIDR2-Ly`α [14] sample, where we restricted forest pairs to match the effective redshift of our LBG forest signal. Throughout this section, we adjust our notation to Lyα(LBG) and Lyα(QSO) to differentiate the origin of the Lyα forest signal.

Our Lyα(LBG)×Lyα(LBG) auto-correlation measurement represents the first detection of the Lyα forest auto-correlation signal using exclusively LBG spectra, at an effective redshift

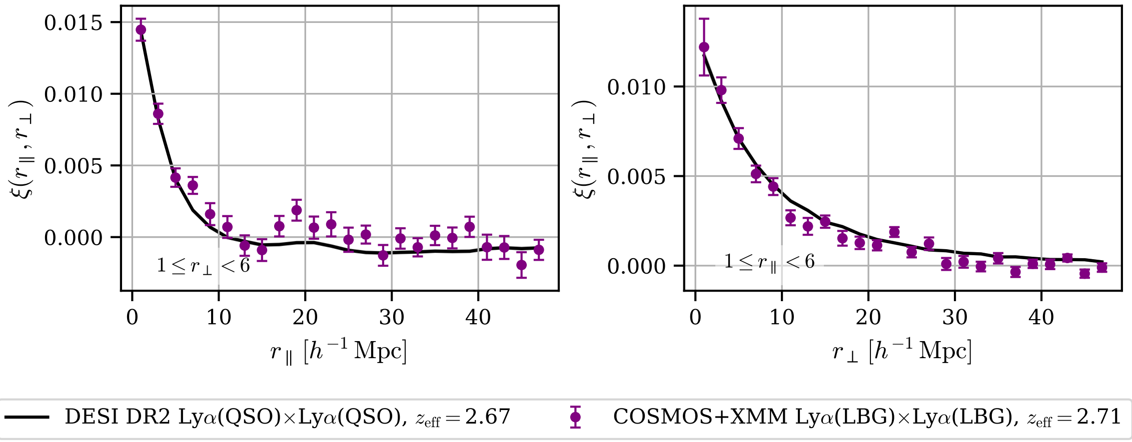

Figure 5. Lyα×Lyα auto-correlation from the COSMOS+XMM LBG forest combined sample as a function of $r_\parallel$ (left) and $r_\perp$ (right), averaged over the $1-6$ h^{-1} Mpc range of the complementary coordinate. The black line shows the results from the `DESIDR2-Ly`α [14] quasar forest sample where forest pairs have been restricted to match the effective redshift of our LBG forest signal. The (almost indistinguishable) shaded black region corresponds to the 1σ uncertainty.

of $z_{\rm eff} \sim 2.7$. While previous works have included LBG forests [e.g., 26–29], they primarily combined them with quasar forests to study Lyα-galaxy cross-correlations.

Our measurement is consistent within uncertainties with the Lyα(QSO)×Lyα(QSO) correlation obtained from the DESI DR2 quasar sample when forest pairs are restricted to match the same effective redshift ($z_{\rm eff}$). We do not observe any additional systematic effects beyond those already present in the quasar analysis. Furthermore, we verified that excluding DLA masking in the DESI DR2 quasar forests introduces only minor differences, well within the uncertainties of our LBG forest measurements.

We have also confirmed that the distortion introduced by the continuum-fitting procedure is comparable between the quasar and LBG-based measurements at the scales relevant to our study. This consistency supports the expectation that both signals should closely match. A notable exception is observed in the Si III(1207) feature at $r_\parallel \sim 20$ Mpc at transverse distances close to the line-of-sight (see Section 3.3.1), which appears more prominent in the LBG sample. Interestingly, the strength of this feature diminishes as we increase the CNN confidence threshold (see appendix B), suggesting that it likely arises from contamination by objects misclassified as LBGs, redshift errors outliers, or from spectral noise.

Similarly, Figure 6 presents the Lyα(LBG)×LBG (top) and Lyα(LBG)×LAE (bottom) cross-correlations on the $(r_\parallel, r_\perp)$ grid for COSMOS, XMM, and the combined dataset. Figure 7 shows the same cross-correlation signals averaged over four $r_\perp$ slices. We include the Lyα(QSO)×QSO cross-correlation results from the aforementioned DESI DR2 quasar forests and the position of quasar tracers, where forest-quasar pairs were restricted to match the effective redshift of our LBG and LAE cross-correlation signals.

We observe that the Lyα(LBG)×LBG and Lyα(LBG)×LAE cross-correlation functions differ from the Lyα(QSO)×QSO signal, showing lower amplitudes at small-scales. This reflects the fact that LBGs [73, 74] and LAEs [75–77] reside in lower-mass halos, resulting in reduced

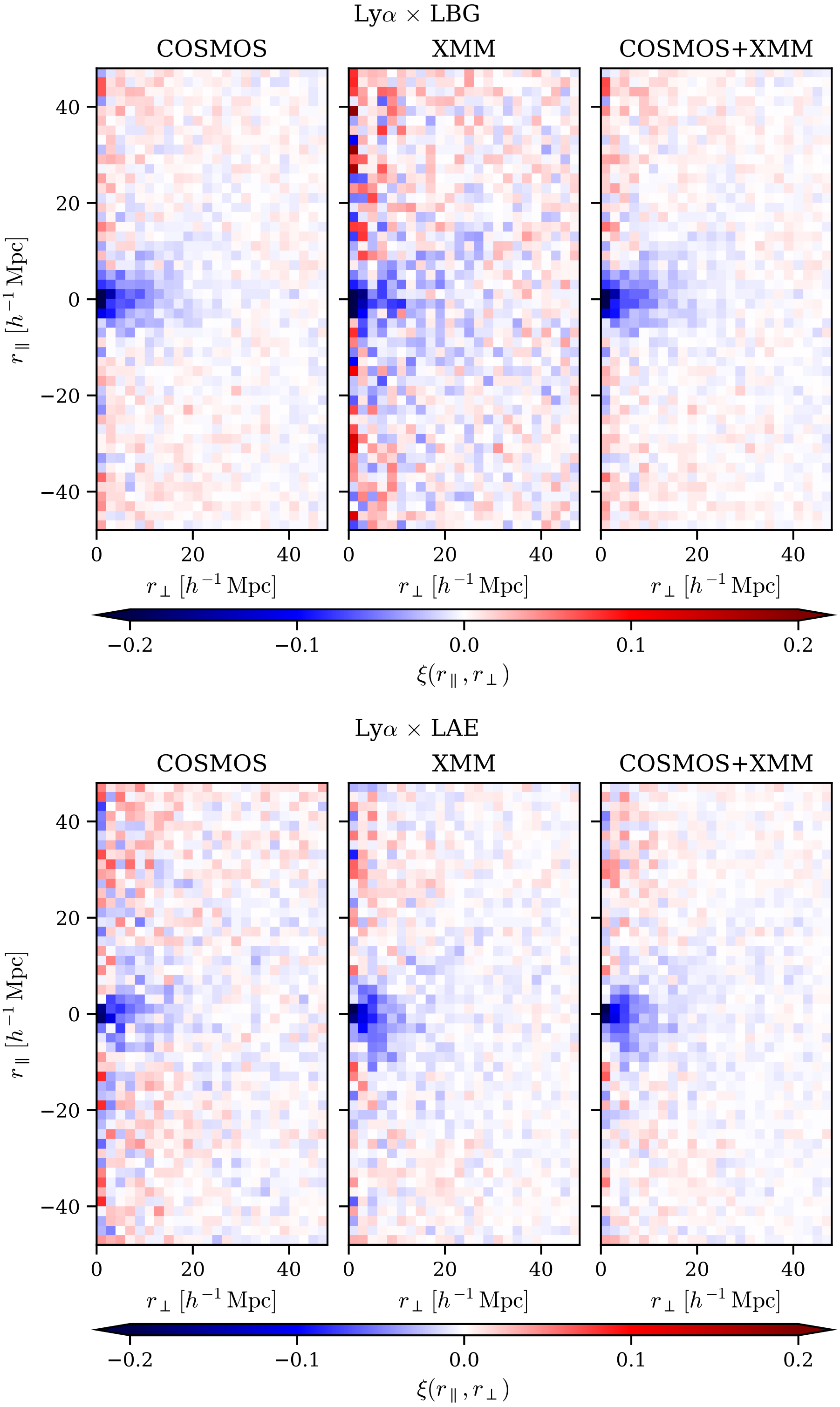

Figure 6. Lyα×LBG (top) and Lyα×LAE (bottom) cross-correlations in COSMOS (left), XMM (center), and the combined (right) samples, shown over a $(r_\parallel, r_\perp)$ grid.

JCAP12(2025)053

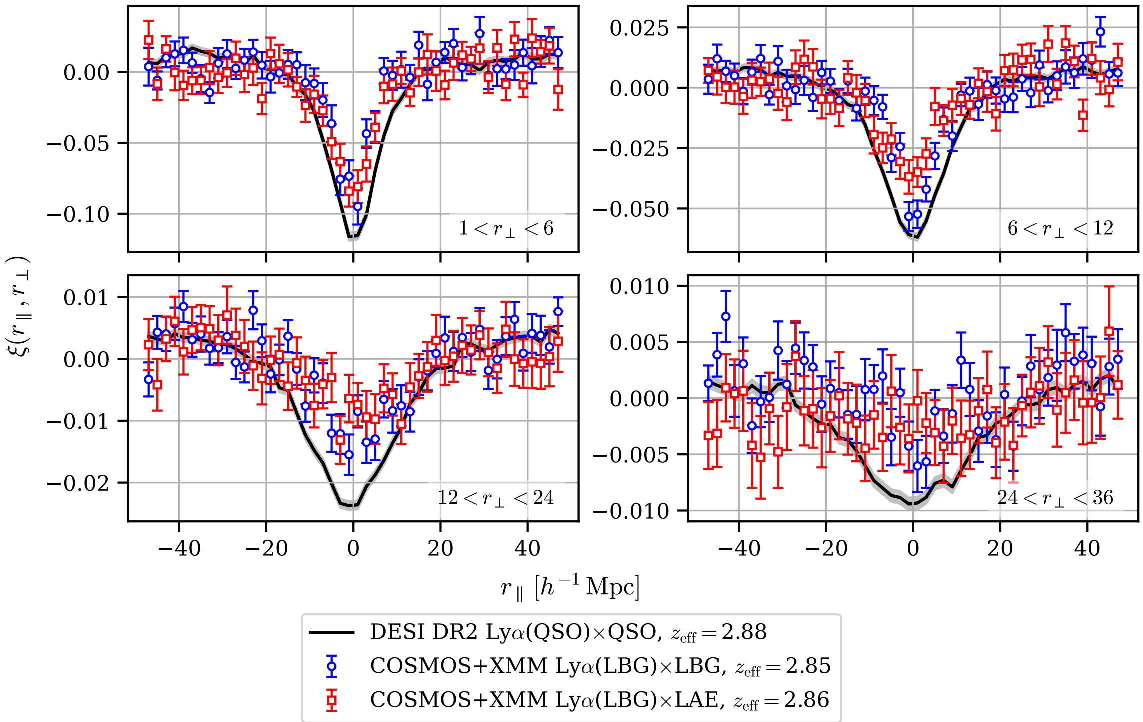

Figure 7. Measured cross-correlation between the Lyα forest in LBG spectra and the positions of LBGs (blue unfilled dots) and LAEs (red unfilled squares) from the COSMOS+XMM combined sample as a function of $r_\parallel$, averaged over four distinct $r_\perp$ bin slices. The black line shows the cross-correlation between the Lyα forest from quasars and quasar tracer positions from the `DESIDR2-Lyα` [14] sample with the shaded region corresponding to 1σ uncertainties. The DESI DR2 forest-quasar pairs have been restricted to ensure that the Lyα(QSO)$\times$QSO signal has an effective redshift comparable to that of our Lyα(LBG)$\times$LBG and Lyα(LBG)$\times$LAE signals.

JCAP12(2025)053

clustering amplitude (lower linear bias) and smaller intrinsic velocity dispersions compared to quasars [78–81]. Moreover, redshift estimate errors introduce additional broadening of the profile. Both the Lyα(LBG)$\times$LBG and Lyα(LBG)$\times$LAE cross-correlations peak at $r_\parallel \sim 0\ h^{-1}$ Mpc, especially in the $1 < r_\perp < 6\ h^{-1}$ Mpc and $6 < r_\perp < 12\ h^{-1}$ Mpc slices, confirming that our systematic redshift offset corrections (see appendix A) are appropriate.

Our detection of the LBG forest auto-correlation, together with its cross-correlations with LBG and LAE positions, shows that correlations based on LBG forests alone can be robustly measured. In contrast to the relatively high SNR LBG spectra used in LATIS and CLAMATO, our measurement relies on lower SNR (~ 0.35 Å^{-1}) data, providing the first direct demonstration of the viability of LBG forests in the context of wide-area surveys such as DESI. These results indicate that LBG forests can complement quasar-based studies by extending clustering measurements to higher source densities and redshifts. Looking ahead, larger datasets such as those anticipated from DESI-II [82] will be required to test this approach at higher precision and to explore whether the BAO feature can be detected from LBG forests. We discuss these prospects further in Section 5.

Fitting range	b_α	β_α	$b_{\rm LBG}$	$\sigma_{z,{\rm LBG}}$ $[h^{-1}\,{\rm Mpc}]$	$\Delta r_{\parallel,{\rm LBG}}$ $[h^{-1}\,{\rm Mpc}]$	$\chi^2_{\rm min}$/D.o.F
Lyα×Lyα+Lyα×LBG						
$1 < r < 28$	$-0.152^{+0.017}_{-0.023}$	$1.46^{+0.23}_{-0.36}$	1.90 ± 0.13	< 0.780	0.04 ± 0.18	$461.29/(462 - 10) = 1.02$
$6 < r < 28$	$-0.187^{+0.028}_{-0.042}$	$1.10^{+0.16}_{-0.45}$	1.85 ± 0.16	< 0.790	-0.37 ± 0.36	$425.71/(438 - 10) = 0.99$
$10 < r < 28$	-0.137 ± 0.043	$1.99^{+0.42}_{-1.1}$	1.93 ± 0.24	< 1.350	-1.09 ± 0.54	$382.31/(402 - 10) = 0.98$
$1 < r < 34$	$-0.167^{+0.015}_{-0.020}$	$1.24^{+0.18}_{-0.26}$	1.88 ± 0.13	< 0.760	0.04 ± 0.19	$711.19/(684 - 10) = 1.06$
$6 < r < 34$	$-0.209^{+0.019}_{-0.031}$	$0.85^{+0.13}_{-0.26}$	1.80 ± 0.16	< 0.800	-0.29 ± 0.36	$672.01/(660 - 10) = 1.03$
$10 < r < 34$	$-0.164^{+0.034}_{-0.046}$	$1.39^{+0.21}_{-0.68}$	1.90 ± 0.25	< 1.220	-0.93 ± 0.54	$629.60/(624 - 10) = 1.03$

Table 2. Parameter constraints at 68% confidence from joint auto+cross-correlation fits, applied to the combined COSMOS+XMM LBG sample. Our full model is described in Section 3.3. We report the best-fit model $\chi^2_{\rm min}$ value divided by the degrees of freedom (D.o.F). Results are shown for several fitting ranges in comoving separation. Entries given as $\mu \pm \sigma$ denote Gaussian posterior distributions, while those of the form $\mu^{+\sigma_1}_{-\sigma_2}$ indicate non-Gaussian posteriors.

4.2 Model parameter constraints

As noted in Section 3.3, deriving tight model constraints is beyond the primary scope of this work. This is primarily because the statistical power of our measurements is insufficient, and also because we adopt a simplified model that does not fully capture the underlying physics on the small scales considered here, potentially biasing the results. Nevertheless, we perform model fitting for completeness and emphasize that *the following results should be interpreted with caution.*

For the remainder of this section, we focus only on the Lyα and LBG-related parameters. A more complete discussion of the other model parameters and their posterior distributions can be found in appendix B.

We adopt a Gaussian likelihood and use the `Vega` package[5] to model the correlation functions. Best-fit parameter values and statistics are estimated using the `iminuit` minimizer [83, 84], while posterior distributions are sampled via the `Polychord` nested sampler [85, 86], both integrated within the `Vega` interface. We adopt wide flat priors for all free parameters, except for the HCD redshift-space distortion parameter, for which we impose a Gaussian prior of $\beta_{\rm HCD} = 0.5 \pm 0.2$. This is broader than the prior used in `DESIDR2-Ly`α [14] to allow for possible redshift evolution of $\beta_{\rm HCD}$, given that our analysis targets higher redshifts.

We carry out auto+cross-correlation joint fits, also accounting for their cross-covariance. We do not present auto or cross-only fits, as their statistical power is significantly lower than that of the combined fit. We also attempted additional fits based on the Lyα×LAE cross-correlation; however, these proved unstable because most of our LAE sample lies in the XMM field, whereas the bulk of our Lyα forest statistics come from COSMOS (see Table 1); therefore, we omit these results from this work.

Table 2 presents the parameter constraints derived from our combined COSMOS+XMM LBG sample auto+cross-correlation fits when varying the comoving separation fitting range. From these results we draw the following conclusions:

- Lyα bias parameters (b_α, β_α): These parameters exhibit significant variation depending on the chosen fitting range, with b_α spanning from -0.209 to -0.137 and β_α ranging from

[5]https://github.com/andreicuceu/vega

JCAP12(2025)053

Variation	b_α	β_α	$b_{\rm LBG}$	$\sigma_{z,\rm LBG}$ $[h^{-1}\,\rm Mpc]$	$\Delta r_{\parallel,\rm LBG}$ $[h^{-1}\,\rm Mpc]$	$\chi^2_{\rm min}/\rm D.o.F$
Reference	$-0.152^{+0.017}_{-0.023}$	$1.46^{+0.23}_{-0.36}$	1.90 ± 0.13	< 0.780	0.04 ± 0.18	$461.29/(462-10) = 1.02$
CNN Confidence level						
CL > 0.97	$-0.156^{+0.018}_{-0.022}$	$1.39^{+0.21}_{-0.36}$	$1.322^{+0.092}_{-0.10}$	< 0.760	-0.09 ± 0.17	$504.20/(462-10) = 1.12$
CL > 0.99	$-0.186^{+0.016}_{-0.021}$	$0.97^{+0.16}_{-0.23}$	1.71 ± 0.12	< 0.760	-0.07 ± 0.18	$492.76/(462-10) = 1.09$
CL > 0.999	$-0.168^{+0.019}_{-0.032}$	$1.36^{+0.20}_{-0.41}$	2.18 ± 0.15	< 0.750	-0.29 ± 0.19	$514.74/(462-10) = 1.14$
Bin width ΔR						
$\Delta R = 1\ h^{-1}\,\rm Mpc$	$-0.163^{+0.017}_{-0.022}$	$1.29^{+0.20}_{-0.31}$	1.79 ± 0.12	< 0.960	-0.16 ± 0.16	$2095.91/(1854-10) = 1.14$
$\Delta R = 3\ h^{-1}\,\rm Mpc$	$-0.174^{+0.020}_{-0.026}$	$1.25^{+0.20}_{-0.35}$	1.86 ± 0.13	< 0.650	-0.02 ± 0.19	$243.75/(207-10) = 1.24$
$\Delta R = 4\ h^{-1}\,\rm Mpc$	$-0.186^{+0.019}_{-0.024}$	$1.09^{+0.20}_{-0.29}$	1.74 ± 0.15	< 1.060	-0.03 ± 0.26	$133.96/(117-10) = 1.25$
Alternative model or priors						
Gaussian $F_{\rm NL}^{\rm cross}$	$-0.154^{+0.017}_{-0.021}$	$1.46^{+0.22}_{-0.37}$	1.89 ± 0.13	< 1.130	0.04 ± 0.20	$461.29/(462-10) = 1.02$
Gaussian prior β_α	$-0.148^{+0.007}_{-0.01}$	1.540 ± 0.095	1.89 ± 0.13	< 0.790	0.04 ± 0.19	$461.80/(462-10) = 1.02$
$F_{\rm NL}^{\rm auto}$ parameters	$-0.154^{+0.017}_{-0.021}$	$1.41^{+0.22}_{-0.34}$	$1.90^{+0.12}_{-0.14}$	< 0.800	0.04 ± 0.19	$461.99/(462-10) = 1.02$
Fixed β_α and HCDs	-0.130 ± 0.004	1.71	1.89 ± 0.13	< 0.740	0.04 ± 0.20	$463.92/(462-6) = 1.02$

Table 3. Parameter constraints at 68% confidence from Ly$\alpha\times$Lyα+Ly$\alpha\times$LBG joint fits from alternative analysis and modeling strategies when fitting over the $1 < r < 28\ h^{-1}\,\rm Mpc$ range. We adopt as a reference the results from the corresponding fitting range from Table 2. All our variations have an effective redshift $z_{\rm eff} = 2.71$, except for the CL > 0.999 variation which has an effective redshift $z_{\rm eff} = 2.72$.

0.85 to 1.99. While including smaller-scale separations tightens statistical uncertainties, it also increases susceptibility to non-linear effects. On the other hand, within the limits of our dataset, including larger-scale separations does not improve parameter stability. The degeneracy of the Lyα parameters with the HCD linear bias parameter, in addition to the limited statistical power of our current dataset further complicates the robust disentanglement of Lyα and HCD contributions. Taken together, these factors prevent a reliable measurement of b_α and β_α within the scope of this analysis.

- LBG bias ($b_{\rm LBG}$): This parameter remains stable across all fitting ranges explored ranging from 1.80 to 1.93. We note that the inferred values are systematically lower than those reported in previous LBG clustering analyses at comparable redshifts [e.g., 22, 73], which typically find $b_{\rm LBG} \sim 2.6$–3.5. We attribute this discrepancy primarily to sample purity. As will be discussed later in this section, the inferred $b_{\rm LBG}$ increases as a function of the CNN confidence level threshold used for selecting LBGs, indicating a strong dependence on contamination, classification reliability, and redshift errors outliers.
- Velocity dispersion ($\sigma_{z,\rm LBG}$): We are unable to robustly determine a central value for this parameter, obtaining only upper limits that vary with the fitting range but consistently remain below $\sim 2.7\ h^{-1}\,\rm Mpc$ at the 95% confidence level. We also note that this parameter is stable against CNN confidence level. This upper bound is relatively smaller than the $\sim 3.7\ h^{-1}\,\rm Mpc$ velocity dispersion measured for `DESIDR2-Ly`α quasars [14].
- Redshift offset ($\Delta r_{\parallel,\rm LBG}$): Across all fitting ranges, the inferred redshift offset is consistent with zero, with the largest deviation reaching only $\sim 2\sigma$ when small scales ($r < 10\ h^{-1}\,\rm Mpc$) are excluded. This supports the effectiveness of the redshift estimate correction described in appendix A.

We further assess the robustness of our measurements by exploring a range of alternative analysis and modeling choices. In Table 3, we summarize the resulting parameter constraints, taking as a baseline the Lyα×Lyα+Lyα×LBG joint fit over the range $1 < r < 28\ h^{-1}$ Mpc. The variations explored include:

- CNN confidence level: We explore alternative LBG selection thresholds of CL > 0.97, 0.99, and 0.999, relative to our baseline of CL > 0.995. Increasing the threshold systematically raises the inferred LBG bias, $b_{\rm LBG}$, consistent with higher sample purity, p, and a reduced fraction of catastrophic redshift errors, f. `R24` [22] report $p \sim 0.94$ and $f \sim 0.10$ at CL > 0.995. Independently, our comparison with the COSMOS spectroscopic compilation in Section 2.2 indicates a similar redshift outlier fraction. Since our sample extends to fainter magnitudes than either the `R24` [22] or the cross-matched COSMOS samples, we expect slightly lower purity and a correspondingly higher fraction of catastrophic redshift errors, which would further suppress $b_{\rm LBG}$. In contrast, we observe no clear trend in the Lyα linear bias b_α with CL, partly due to its degeneracy with β_α and the HCD parameters. This degeneracy is discussed in appendix B, where we also present a complementary test that confirms a similar dependence of b_α on the CL threshold. Together, these results highlight the need for a more refined treatment of interlopers and catastrophic redshift errors in future large-scale clustering analyses with LBG forest.
- Bin width ΔR: We vary the bin width ΔR between 1 and 4 h^{-1} Mpc, compared to our baseline of 2 h^{-1} Mpc. Focusing on the redshift estimate offset parameter $\Delta r_{\parallel,\rm LBG}$, we find values consistent with zero across all choices of ΔR, confirming the robustness of our previous results. In contrast, the parameters b_α, β_α, $b_{\rm LBG}$, and $\sigma_{z,\rm LBG}$ show noticeable sensitivity to the bin width.
- Alternative parametrization for the LBG redshift dispersion: For the effect of velocity dispersion and redshift errors on the cross-correlation, we adopt a Gaussian model, $F_{\rm NL}^{\rm cross}(\mathbf{k}) = \exp\left(-\frac{1}{2}k_\parallel^2\sigma_{z,\rm X}^2\right)^{-\frac{1}{2}}$, instead of the Lorentzian profile introduced in Eq. (3.9). The inferred value in this case is $\sigma_{z,\rm LBG}^2 < 1.71\ h^{-1}$ Mpc at 95% the confidence level, consistent with our previous findings.
- Alternative constraints on β_α and $F_{\rm NL}^{\rm auto}$ parameters: In a first fitting variation, we impose a Gaussian prior $\beta_\alpha = 1.56 \pm 0.1$ based on results from the ACCEL2 hydrodynamic simulations [64]. In a second variation, we use an alternative set of (fixed) parameters for the $F_{\rm NL}^{\rm auto}$ auto-correlation small-scales correction.[6] Both variations again highlight the important sensitivity of our best-fit values for b_α as a function of the fitting strategy details, while the inferred values for $b_{\rm LBG}$ and redshift error parameters are stable.
- Fixed β_α and HCD parameters: We fix the Lyα RSD parameter to $\beta_\alpha = 1.71$ and the HCD parameters to $b_{\rm HCD} = -0.092$, $\beta_{\rm HCD} = 0.56$, and $L_{\rm HCD} = 5.39\ h^{-1}$ Mpc. These values were derived from fits to the Lyα correlation functions of `DESIDR2-Ly`α [14] quasars, where forest-forest and forest-quasar pairs were restricted to match the effective redshift of our LBG sample, and DLAs were not masked, in order to replicate the

[6]Specifically: $q_1 = 0.7768$, $k_v = 1.26735\,h\,{\rm Mpc}^{-1}$, $a_v = 0.6094$, $b_v = 1.645$, and $k_p = 20.09\,h\,{\rm Mpc}^{-1}$ obtained by interpolating the results from table 7 of [63] at $z_{\rm eff} \sim 2.7$

JCAP12(2025)053

conditions of the LBG forest sample. We allow b_α to vary, as it is expected to be affected by sample purity and the redshift outlier fraction, and therefore may deviate from the quasar-based result. Taking the quasar-based measurement of $b_\alpha = -0.144 \pm 0.016$ as a reference, the value inferred from the LBG Lyα forest in this variation is ∼10% smaller in amplitude. This is consistent with the expected redshift outlier fraction in our dataset and provides an independent cross-check of the trends identified in the CNN confidence-level variations. While the effect of catastrophic redshift errors appear to cause the majority of this difference, contributions from statistical fluctuations and the higher noise of LBG spectra compared to quasars cannot be ruled out.

5 Prospects for cosmology

Our detection of the Lyα forest auto and cross-correlations using a relatively low-SNR and small sample of LBG spectra from DESI observations demonstrates the feasibility of performing large-scale clustering analyses with the Lyα forest from dense LBG samples in future wide-area spectroscopic surveys. To quantify the cosmological potential of such measurements, we present forecasts for the expected precision on Baryon Acoustic Oscillation (BAO) and Full-Shape (FS) parameters of the correlation functions. We consider a future survey covering 5,000 deg^2 with a surface density of 1,000 spectroscopically confirmed LBGs per deg^2, assuming target selection similar to that of our COSMOS sample. The expected performance is evaluated using two complementary approaches: a forecast based on synthetic LBG spectra (Section 5.1) and a Fisher-matrix analysis (Section 5.2). For context, we also include analogous quasar-based forecasts over a 5,000 deg^2 footprint of the completed DESI Lyα quasar survey, allowing a direct comparison with the LBG-based results.

We define the parallel and transverse BAO dilation parameters as

$$\alpha_\parallel = \frac{D_H(z_{\rm eff})/r_d}{[D_H(z_{\rm eff})/r_d]_{\rm fid}}, \tag{5.1}$$

$$\alpha_\perp = \frac{D_M(z_{\rm eff})/r_d}{[D_M(z_{\rm eff})/r_d]_{\rm fid}}, \tag{5.2}$$

where D_H is the Hubble distance, D_M the comoving angular diameter distance, and r_d the sound horizon at drag epoch. The subscript "fid" denotes the values given by the Planck 2018 fiducial cosmology [56]. Alternatively, we define the isotropic dilation parameter, $\alpha_{\rm ISO}$ [13] and the Alcock-Paczynski distortion parameter, $\alpha_{\rm AP}$, as

$$\alpha_{\rm ISO} = \frac{D_H^{0.55}(z_{\rm eff})D_M^{0.45}(z_{\rm eff})/r_d^2}{[D_H^{0.55}(z_{\rm eff})D_M^{0.45}(z_{\rm eff})/r_d^2]_{\rm fid}} = \alpha_\parallel^{0.55}\alpha_\perp^{0.45} \tag{5.3}$$

$$\alpha_{\rm AP} = \frac{D_M(z_{\rm eff})/D_H(z_{\rm eff})}{[D_M(z_{\rm eff})/D_H(z_{\rm eff})]_{\rm fid}} = \frac{\alpha_\perp}{\alpha_\parallel}, \tag{5.4}$$

where the exponent weights in $\alpha_{\rm ISO}$ correspond to the optimal combination found in the `DESIDR2-Ly`α [14] analysis. While these coefficients are appropriate for quasar Lyα forest analyses, the optimal values for LBG Lyα forest may differ. This will be clarified once BAO measurements from the LBG forest become available.

As in Section 4, we adopt the notation Lyα(LBG) and Lyα(QSO) throughout this section to explicitly differentiate between the Lyα forest signal extracted from LBG spectra and that from quasars.

5.1 Forecast with LBG synthetic spectra

For the LBG forecasts, we construct a single synthetic dataset of 1,000 LBG/deg^2 over a 5,000 deg^2 footprint, employing the following methodology.

1. Density field simulations: We use the Quasi-Linear implementation of `CoLoRe`[7] to generate a Gaussian random field matching an input matter power spectrum, and draw LBG positions via Poisson sampling assuming a density distribution, $n(z)$, similar to our COSMOS sample, rescaled to 1000 LBGs/deg^2 (see Figure 8). For simplicity, we assume a quasar-like linear bias evolution model.
At this same stage, we generate density and velocity fields along the line-of-sight (also referred to as skewers) associated with each mock LBG. At last, we use `LyaCoLoRe`[8] to post-process these skewers into a transmitted flux fraction field via a Fluctuating Gunn-Petterson Approximation (FGPA), also incorporating small-scale corrections and redshift-space distortions.
2. Synthetic Spectra generation: Each mock LBG is assigned a random r-band magnitude drawn from the observed COSMOS distribution (in redshift and magnitude). Then, we use the `quickquasars` script[9] to generate mock spectra, using a common LBG continuum template (redshifted and magnitude-scaled, accordingly). We use `specsim` package[10] to introduce DESI-like instrumental noise assuming two-hour nominal dark-time exposures (no moon, 1.1" seeing, airmass 1.0) for each LBG. This configuration results in an average in forest SNR of 0.33 Å^{-1} for our mock spectra. We do not include contamination due to metals, HCDs, or redshift errors.

For the quasar forecasts, we generate a corresponding synthetic spectra dataset (also without contaminants) that emulates the characteristics expected for the completed DESI Lyα survey. The procedure follows the same methodology used for the mocks in previous DESI Lyα forest analyses [e.g., 88, 90, 92].

Then, we measure the auto and cross-correlations and associated covariance matrices, following the same recipe described in Section 3, adopting a $\Delta R = 4\ h^{-1}$ Mpc grid up to $0 < r_{\parallel}, r_{\perp} < 200\ h^{-1}$ Mpc separations for the auto-correlations and $0 < \left|r_{\parallel}\right|, r_{\perp} < 200\ h^{-1}$ Mpc for the cross-correlations. We compute the covariance by sub-sampling into `nside=16` HEALpix pixels and account for cross-covariance between the auto and cross-correlations. Due to computational limitations arising from the high LBG target density, we restrict the LBG forest analysis to a 1,000 deg^2 section and rescale the covariance by a factor of 5 to approximate that of the full 5,000 deg^2 footprint. For the quasar analysis, we rescale the covariance matrix by a factor of $\sqrt{14,000/5,000}$ to account for the difference in footprint area between the completed DESI survey and the considered 5,000 deg^2 portion.

[7] https://github.com/damonge/CoLoRe [87, 88]
[8] https://github.com/igmhub/LyaCoLoRe [89]
[9] https://github.com/desihub/desisim/blob/main/py/desisim/scripts/quickquasars.py [90]
[10] https://github.com/desihub/specsim [91]

At last, we use the *forecast mode* of the `Vega` code to extract the expected BAO uncertainties. This mode generates a noiseless synthetic data vector representing the correlation function predicted by a given fiducial model. Forecasted errors are then obtained from the covariance matrix via the Hessian of the likelihood function, ensuring that all survey configurations are evaluated under consistent assumptions. We adopt the best-fit model from the DESIDR2-Lyα [14] analysis as our fiducial template; for the LBG forecasts, we exclude quasar-specific terms and include redshift evolution of the Lyα parameters around $z \sim 2.7$. The BAO scale parameters are fixed to $\alpha_{\parallel} = \alpha_{\perp} = 1$. The mock-based forecast results are reported in Table 4: the first three rows correspond to LBG-based correlations, and the next three (below the dashed line) to their quasar-based analogues.

As a complementary forecast, we use the same mock covariance matrices to estimate Alcock-Paczynski constraints from a Lyα Full Shape (FS) analysis, under the same survey setup and fiducial model. We follow the analysis methodology presented in [93], previously applied to both mock [94] and observational data [95, 96]. This approach extends the information extracted beyond the BAO peak by including a wider range of scales, leading to significantly improved constraints on the AP parameter.

For the LBG case, we perform FS forecasts using three different minimum separation thresholds for the fit, ranging from an optimistic case ($r > 10\ h^{-1}\,$Mpc) to a conservative case ($r > 50\ h^{-1}\,$Mpc), while for quasars we only consider the $r > 30\ h^{-1}\,$Mpc case. The results for the intermediate, and most realistic, choice of $r > 30\ h^{-1}\,$Mpc are shown in the final column of Table 4. For the optimistic and conservative configurations on LBGs, we obtain combined auto+cross-correlation constraints of $\sigma^{\rm FS}_{\alpha_{\rm AP}} = 0.46\%$ and 0.84%, respectively.

5.2 Fisher forecast

To complement our mock-based BAO forecast (Section 5.1), we carry out a Fisher matrix forecast under identical survey conditions. This analysis follows the formalism developed by [97] and subsequently applied in [21, 98]. The forecast is implemented using the publicly available `lyaforecast` code.[11] This approach allows for fast and consistent evaluation of the expected BAO precision across different tracer combinations and survey configurations. Unlike the mock-based forecasts, we do not include Full-Shape constraints here, as this functionality is not currently available in `lyaforecast`.

The Fisher matrix F_{ij} is related to the covariance of our dataset by the relation [99]:

$$F_{ij} = \frac{1}{2}\mathrm{Tr}\left(C^{-1}\frac{\delta C}{\delta\theta_i}C^{-1}\frac{\delta C}{\delta\theta_j}\right), \tag{5.5}$$

where θ are a set of parameters, and $C(\theta) = \langle\delta_i\delta_j\rangle$, for a Gaussian distributed data vector δ with $\langle\delta\rangle = 0$. Then, the best possible unbiased estimates on θ, given by the Cramer-Rao lower bound [100], are related to the Fisher matrix via:

$$\sigma_i \geq (F^{-1})_{ii}, \tag{5.6}$$

where σ_i is the variance on the estimate of parameter θ_i.

[11] https://github.com/igmhub/lyaforecast

Following [97], we construct our covariance matrix in Fourier space as:

$$\left\langle \tilde{\delta}(\mathbf{k})\tilde{\delta}^{*}(\mathbf{k}')\right\rangle = (2\pi)^3\delta^D(\mathbf{k}-\mathbf{k}')\left(P_F(\mathbf{k}) + P_w^{\perp}P_F^{\mathrm{1D}}(k_{\parallel}) + P_N^{\mathrm{eff}}\right), \tag{5.7}$$

where $P_F(\mathbf{k})$ is the 3-dimensional Lyα power spectrum (P3D), $P_F^{\mathrm{1D}}(k_{\parallel})$ is the 1-dimensional Lyα power spectrum (P1D). $P_w^{\perp}$ is the power spectrum of the weight overdensity ($\delta_w(\mathbf{x}_{\perp}) = w(\mathbf{x}_{\perp})/\overline{w} - 1$), where $w(\mathbf{x}_{\perp})$ is the weight assigned to each line of sight. The weight field is treated as white noise in the transverse direction, reflecting the assumption that weights are uncorrelated between different forests. In contrast, within each line of sight, the weights are assumed to be perfectly correlated across pixels. Finally, P_N^{eff} is the weighted noise power, dependent on the target density and spectral noise. A detailed description on the estimation of each of the terms in Eq. (5.7) can be found in [97], here we will just give a brief summary.

In general, $P_w^{\perp}$ depends on the distribution of galaxies with magnitude and redshift, the survey area, target density and forest length. The weighting also depends on pixel variance σ_N, which is also a function of magnitude, redshift and observed wavelength (position within the forest).[12] The noise term P_N^{eff} also depends on all of the parameters mentioned above, but features σ_N explicitly rather than just having dependence in the weights ($P_N^{\mathrm{eff}} \propto \langle\sigma_N^2\rangle$).

We adopt a SNR of 0.33 Å^{-1} for the Lyα(LBG)×Lyα(LBG), Lyα(LBG)×LBG, and Lyα(LBG)×Lyα(LBG)+Lyα(LBG)×LBG correlation functions estimates. This corresponds to the median SNR value found in the mock LBG spectra used in our mock-based forecast. We note that this approximation is used because we lack noise estimates as a function of redshift, magnitude, and observed wavelength.

Taking advantage of the flexibility of the Fisher matrix formalism, we also forecast the uncertainties for the LBG×LBG and QSO×QSO auto-correlation as well as for the Lyα(QSO)×LBG cross-correlation that combines the DESI quasar forest with the LBG distribution from our hypothetical survey. For all forecasts involving quasars, we adopt the SNR distribution expected for the final DESI survey [21]. The luminosity functions used in these forecasts are shown in Figure 8, with solid lines denoting quasars and dashed lines representing LBGs, derived from DESI targeting projections and from our COSMOS sample, respectively.

The Fisher forecast results are presented in the second part of Table 4. The two forecasting methods yield consistent results, with the LBG forest auto and cross-correlations combination agreeing within $\lesssim 5\%$. The Fisher analysis also shows that the Lyα(QSO)×LBG cross-correlation is less constraining than the corresponding Lyα(LBG)×LBG measurement. In agreement with the mock spectra forecasts, quasar-based analyses exhibit larger forecasted uncertainties due to their much lower tracer density, despite the higher SNR of quasars. The dense sampling of LBGs compensates for their lower SNR, yielding overall tighter BAO constraints. Since quasar forest analysis typically probe lower redshifts (e.g., $z_{\mathrm{eff}} = 2.33$ for `DESIDR2-Ly`α [14]) than our LBG forest analysis ($z_{\mathrm{eff}} = 2.7$), the two tracers probe complementary epochs of cosmic evolution. Their joint use therefore enables a continuous mapping of large-scale structure across the range $2 \lesssim z \lesssim 3.5$.

[12]Since we are assuming spectral noise is the same everywhere within a forest, we evaluate relevant terms at a mean forest wavelength, rather than iterating across all pixels.

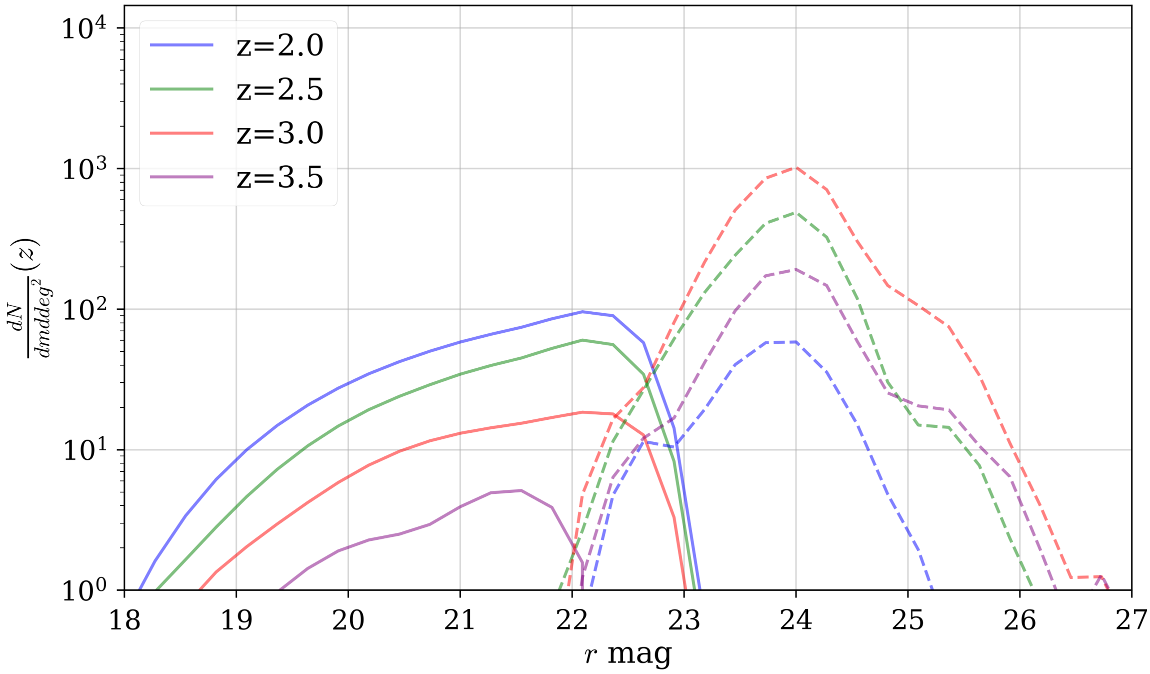

Figure 8. Density distributions, as a function of r-band magnitude and redshift, used for both our mock-based and Fisher forecasts. Solid lines: Quasar distribution for the completed DESI survey, with a mean target density of $60\,\mathrm{deg}^{-2}$. Dashed lines: LBG distribution for a COSMOS-like survey, scaled to a target density $1000\,\mathrm{deg}^{-2}$.

In the transverse direction, the LBG auto-correlation provides the most constraining measurement, owing to the very high LBG target density. We find an average SNR per mode for the LBG auto-correlation of $nP \gtrsim 1$, approaching the regime where cosmic variance becomes significant. In the very low shot-noise limit ($nP \gg 1$), the benefit of combining tracers for BAO constraints is minimal, and only increasing the survey volume improves precision (see [98] for discussion). When combining measurements for the results in Table 4, we therefore adopt the high-noise approximation in which parameter errors are combined through a simple inverse-sum. For the Lyα(LBG)$\times$LBG and Lyα(LBG)$\times$Lyα(LBG) correlations, $nP \sim 0.3$, meaning we may slightly underestimate the total errors. A more accurate treatment would use the full Fisher matrix including cross-covariances between measurements, which we leave for future work.

6 Summary and conclusions

Measurements of the Lyα forest from quasar spectra have delivered groundbreaking cosmological constraints at $z > 2$, including a sub-percent precision BAO measurement at $z_{\mathrm{eff}} = 2.33$ from `DESIDR2-Ly`α [14]. However, such studies remain limited by the relatively low number density of quasars. In this work, we present the first 3D Lyα forest correlation analysis based exclusively on Lyman-break galaxy (LBG) spectra, using a pioneering DESI pilot sample in the COSMOS and XMM fields selected via broadband imaging [22, 42]. By adapting

Correlation	$\sigma_{\alpha_\parallel}$ (%)	$\sigma_{\alpha_\perp}$ (%)	$\sigma_{\alpha_{\rm ISO}}$ (%)	$\sigma_{\alpha_{\rm AP}}$ (%)	$\sigma^{\rm FS}_{\alpha_{\rm AP}}$ (%)
Mock-based Forecast					
Lyα(LBG)$\times$Lyα(LBG)	1.08	1.13	0.57	1.90	0.67
Lyα(LBG)$\times$LBG	1.10	0.95	0.55	1.77	1.00
Lyα(LBG)$\times$Lyα(LBG)+Lyα(LBG)$\times$LBG	0.83	0.72	0.42	1.29	0.57
Lyα(QSO)$\times$Lyα(QSO)	1.82	2.03	0.98	3.31	1.07
Lyα(QSO)$\times$QSO	2.50	2.04	1.23	3.91	2.07
Lyα(QSO)$\times$Lyα(QSO)+Lyα(QSO)$\times$QSO	1.63	1.54	0.83	2.73	0.94
Fisher Forecast					
Lyα(LBG)$\times$Lyα(LBG)	1.11	1.00	0.57	1.80	—
Lyα(LBG)$\times$LBG	1.26	0.86	0.62	1.82	—
Lyα(LBG)$\times$Lyα(LBG)+Lyα(LBG)$\times$LBG	0.87	0.68	0.44	1.34	—
LBG$\times$LBG	1.04	0.59	0.51	1.41	—
Lyα(QSO)$\times$LBG	1.78	1.38	0.89	2.72	—
Lyα(QSO)$\times$Lyα(QSO)	1.89	1.99	0.99	3.34	—
Lyα(QSO)$\times$QSO	2.91	2.35	1.43	4.53	—
Lyα(QSO)$\times$Lyα(QSO)+Lyα(QSO)$\times$QSO	1.15	1.08	0.58	1.92	—
QSO$\times$QSO	4.00	2.58	1.94	5.70	—

Table 4. Forecasted BAO scale uncertainties for a 5,000 deg^2 survey with 1,000 LBGs deg^{-2} and a mean SNR of 0.33 Å^{-1}, using two approaches: synthetic mock-based fits and Fisher matrix formalism. For comparison, we include analogous forecasts for a quasar forest analysis on a 5,000 deg^2 portion of the completed DESI Lyα survey. The Fisher forecasts additionally include the LBG and QSO auto-correlations as well as the Lyα(QSO)$\times$LBG cross-correlation between the forest from DESI quasars in a completely overlapping area with the LBGs in the hypothetical survey. The final column ($\sigma^{\rm FS}_{\alpha_{\rm AP}}$) lists Full Shape (FS) constraints on the Alcock-Paczynski parameter from mock-based forecasts at scales $r > 30\ h^{-1}$ Mpc.

JCAP12(2025)053

the DESI Lyα forest analysis pipeline, we demonstrate that it performs robustly even on lower signal-to-noise ratio (SNR) LBG spectra.

We measure the Lyα forest auto-correlation function on a $0 < r_\parallel, r_\perp < 48\ h^{-1}$ Mpc grid at an effective redshift $z_{\rm eff} = 2.70$, based on 4,151 forests extracted exclusively from LBG spectra. Remarkably, the results show excellent agreement with those derived from the significantly higher-SNR quasar sample in `DESIDR2-Ly`α [14], demonstrating that LBG forests, even at the relatively low SNR of our sample, can reliably trace the intergalactic medium across large volumes. In addition to the auto-correlation, we analyze the Lyα-galaxy cross-correlation using the positions of 4,931 LBGs and 8,431 Lyman-α emitters (LAEs) in the same fields. The Lyα-galaxy cross-correlation provides a powerful tool for calibrating galaxy redshifts, allowing us to correct systematic redshift offsets of $\Delta v_{\rm LBG} = -169 \pm 22$ km s^{-1} for LBGs and $\Delta v_{\rm LAE} = -241 \pm 20$ km s^{-1} for LAEs.

An important distinction between LBG and quasar-based analyses is the significantly higher rate of catastrophic redshift outliers and the lower sample purity in LBG samples. For our DESI LBG dataset, we estimate an outlier fraction of $\sim$10% for galaxies with CNN confidence level CL > 0.995, consistent with the findings of `R24` [22], who additionally report a ∼94% purity at the same threshold. Such contamination directly suppresses the observed clustering amplitude and increases the uncertainty of correlation measurements. In contrast, DESI Lyα quasars typically exhibit catastrophic redshift fractions of $\lesssim 2\%$ and purities exceeding 98% [101, 102]. The relatively larger uncertainty in LBG samples highlights the need for refined mitigation and modeling strategies to fully exploit the statistical power of future dense LBG surveys, alongside a systematic characterization of effects impacting the Lyα forest specifically. Future improvements may build upon the experience from DESI galaxy clustering analyses, which employ redshift-failure weighting schemes and mock-based evaluations of catastrophic redshift impacts, along with additional systematic mitigation strategies [103, 104], as well as from Lyα correlation studies that address correlated noise [72] and other potential systematics that could specifically affect LBG-based forest measurements.

Despite these challenges, our results demonstrate that the Lyα forest extracted from LBG spectra in DESI-like surveys can serve as a complementary cosmological probe to quasar-based analyses and establish a promising path toward mapping large-scale structure at redshifts $z \gtrsim 2$. Looking ahead, we forecast the BAO measurement capabilities of a future spectroscopic survey targeting 1,000 LBGs per square degree over a 5,000 deg^2 footprint, assuming a median SNR of ~ 0.33 Å^{-1}. Using both Fisher-matrix and mock-spectra forecasts independently, we find that combining the Lyα forest auto-correlation and the Ly$\alpha\times$LBG cross-correlation can deliver competitive constraints at $z > 2.6$, with expected uncertainties of $\sigma_{\alpha_\parallel} = 0.8\%$ and $\sigma_{\alpha_\perp} = 0.7\%$, corresponding to $\sigma_{\alpha_{\rm ISO}} = 0.4\%$ in the isotropic and $\sigma_{\alpha_{\rm AP}} = 1.3\%$ in the Alcock-Paczynski (AP) directions. This constraining power is further enhanced when the BAO analysis is complemented by a full-shape analysis of the Lyα forest, yielding a predicted uncertainty of $\sigma_{\alpha_{\rm AP}} = 0.6\%$.

This precision is crucial in light of recent DESI results hinting at possible deviations from the standard dark energy model [105, 106]. Achieving such accuracy at high redshift will require improved sample purity, optimized target selection, deeper imaging, and a careful characterization of systematics, as highlighted by our analysis. This level of performance is within reach of next-generation spectroscopic programs such as DESI-II [82], MUST [107], SPEC-S5 [108], and WST [109], especially when combined with wide-area imaging from the DESI Legacy Surveys [110], UNIONS [43], Euclid [4], and LSST [111]. As such, the Lyα forest from LBGs represents a key component of next-generation high-redshift cosmological probes.

A Correction for the bias in redshift estimate

The automatic redshift estimates for our LBG and LAE samples were calibrated against visually inspected spectra, which primarily rely on strong absorption features from low-ionization lines (Si II(1260), O I(1302) + Si II(1304), C II(1335)) and high-ionization lines (the Si IV(1394), Si IV(1403) doublet). As shown by [54], these interstellar absorption lines are systematically blueshifted relative to the systemic redshift of the galaxy, leading to an underestimation of the true redshift in visual inspections.

In our correlation-function framework (see Section 3), a systematic error in the galaxy redshift translates into a shift of the cross-correlation along the line-of-sight. By measuring the Ly$\alpha\times$LBG cross-correlation using automatic redshift estimates we find a $\Delta r_{\parallel,\rm LBG} = -1.54 \pm 0.20\ h^{-1}$ Mpc offset. Similarly, for Ly$\alpha\times$LAE we measure an offset of $\Delta r_{\parallel,\rm LAE} =$

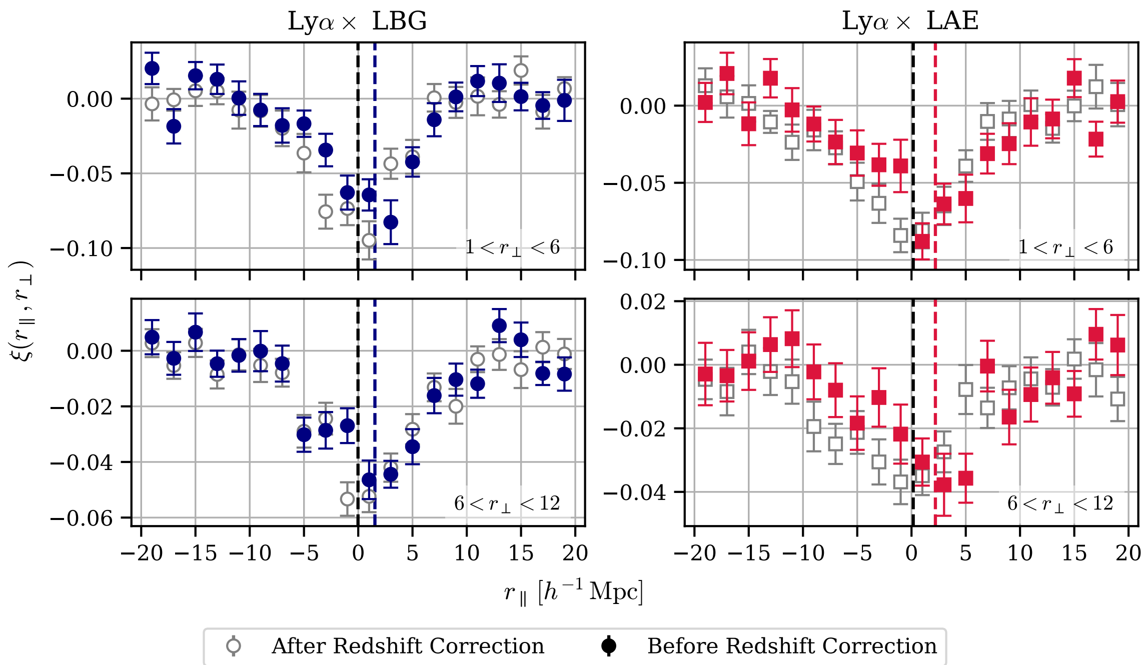

Figure 9. Comparison of the Lyα×LBG (left) and Lyα×LAE (right) cross-correlation signals before (gray open symbols) and after (colored filled symbols) applying the redshift-offset correction. The measurements are shown as a function of transverse separation $r_\parallel$, averaged over two line-of-sight separation bins: $1 < r_\perp < 6$ h^{-1} Mpc (top panels) and $6 < r_\perp < 12$ h^{-1} Mpc (bottom panels). For improved visibility, we restrict the $r_\parallel$ range to $-20 < r_\parallel < 20$ h^{-1} Mpc. Vertical dashed lines indicate the best-fit $\Delta r_{\parallel,\mathrm{X}}$ offsets for each tracer, with colored and black lines corresponding to the results before and after correcting redshift estimates, respectively.

-2.20 ± 0.18 h^{-1} Mpc. These displacements correspond to a velocity offsets of $\Delta v_{\mathrm{LBG}} = -169 \pm 22$ km s^{-1} and $\Delta v_{\mathrm{LAE}} = -241 \pm 20$ km s^{-1} at our effective redshift, respectively.

The LBG velocity offset we measure agrees with the typical -100 to -200 km s^{-1} blueshifts of interstellar absorption lines observed in $z \sim 2-3$ star-forming galaxies [e.g., 54, 112–114]. For LAEs, the inferred offset is in line with values reported for interstellar absorption features in LAE spectra [e.g., 115, 116], which are expected to be comparable to those of LBGs.

We correct for these redshift offsets by applying a shift to each galaxy redshift, using the transformation $z' \to z - \frac{\Delta v_{\mathrm{X}}}{c}(1+z)$, where c is the speed of light. The inferred residual offsets after these corrections are $\Delta r_{\parallel,\mathrm{LBG}} = 0.04 \pm 0.18$ h^{-1} Mpc and $\Delta r_{\parallel,\mathrm{LAE}} = -0.13 \pm 0.17$ h^{-1} Mpc, both consistent with zero which indicates that the procedure effectively removes the systematic bias.

To illustrate the impact of this correction, Figure 9 compares the Lyα×LBG (left panels) and Lyα×LAE (right panels) cross-correlation functions before and after applying the redshift adjustments. The correlations are shown as a function of transverse separation $r_\parallel$, averaged over line-of-sight bins of $1 < r_\perp < 6$ h^{-1} Mpc (top panels) and $6 < r_\perp < 12$ h^{-1} Mpc (bottom panels). Vertical dashed lines mark the measured offsets $\Delta r_{\parallel,\mathrm{X}}$ for each tracer, with colored

JCAP12(2025)053

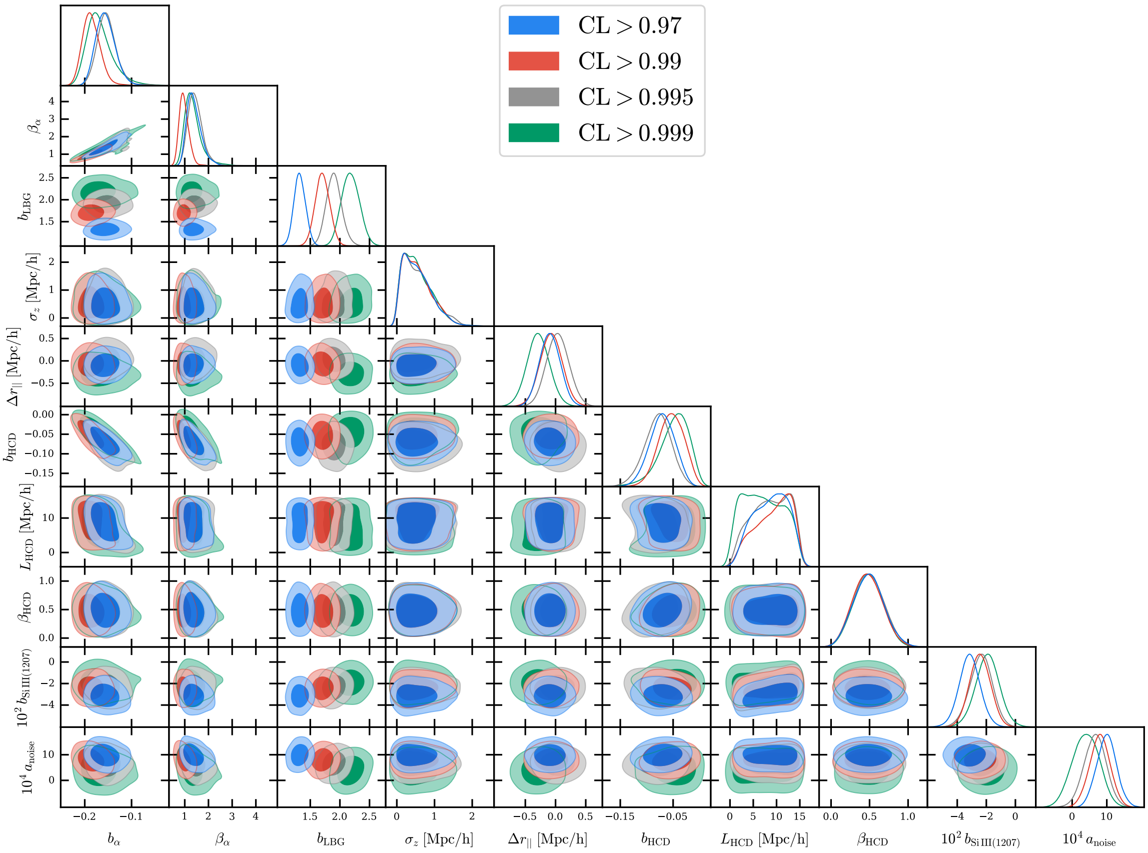

Figure 10. Posterior distributions for every model parameter (see Section 3.3), from the Ly$\alpha\times$Lyα+Ly$\alpha\times$LBG joint-fit of the combined COSMOS+XMM LBG sample fit over $1 < r < 28\ h^{-1}$ Mpc at $z_{\rm eff} \sim 2.70$, for different CNN confidence cuts.

lines indicating the best-fit offset values prior redshift estimate corrections and black lines the results after corrections are applied.

B Impact of sample purity on posterior distributions

In Section 4.2, we presented key constraints on the model parameters derived from the combined fits of the Ly$\alpha\times$Lyα and Ly$\alpha\times$LBG correlation functions. Here, we expand on that summary by showing the full posterior distributions for all free parameters involved in the joint auto and cross-correlation fit over the range $1 < r < 28\ h^{-1}$ Mpc. We also examine the impact of the CNN confidence level (CL) threshold on the resulting parameter estimates. These distributions provide a more complete view of parameter degeneracies and the sensitivity of the results to sample purity. The full triangle plot for Ly$\alpha\times$Lyα+Ly$\alpha\times$LBG is shown in Figure 10.

We begin by examining metal and noise contamination. In our baseline fit, which adopts CL > 0.995, we measure a Si III(1207) linear bias of $b_{\rm Si\,III(1207)} = (-2.36 \pm 0.77) \times 10^{-2}$ and detect correlated noise with an amplitude of $a_{\rm noise} = (6.0 \pm 3.2) \times 10^{-4}$. These values are significantly larger than those obtained from the joint auto and cross-correlation fit

JCAP12(2025)053

of DESIDR2-Lyα [14] quasars, where we measured $b_{\rm Si\,III(1207)} = (-0.97 \pm 0.11) \times 10^{-2}$ and $a_{\rm noise} = (3.39 \pm 0.31) \times 10^{-4}$. Note that in the quasar-based results, forest-forest and forest-quasar pairs were restricted in the correlation-function estimation, and DLAs were not masked, to match the effective redshift and observational conditions of our LBG sample. One might therefore have expected closer agreement with the LBG-based measurements.

A plausible explanation for this discrepancy lies in differences in sample purity: as shown in Figure 10, both $b_{\rm Si\,III(1207)}$ and $a_{\rm noise}$ systematically decrease with increasing CL. This trend suggests that the fitted parameters are sensitive to contamination from misclassified objects or interlopers. Higher CL thresholds (i.e., higher purity) appear to suppress these effects and may help bring the LBG-derived values into closer agreement with those from the quasar-based analysis. For correlated noise, an additional contributing factor may be the lower SNR of the LBG spectra at lower CL thresholds, which can amplify the apparent noise signature in the auto-correlation.

Next, we examine the LBG, Lyα, and HCD parameters. As discussed in section 4.2, we find that the inferred LBG bias increases with CL threshold, while the LBG redshift error parameters remain relatively stable. This behavior suggests a significant fraction of interlopers in the LBG sample at lower CLs, which biases the inferred clustering amplitude. A similar effect is expected for the Lyα linear bias. However, as shown in Figure 10, b_α is degenerate with both the Lyα RSD parameter and the HCD linear bias, complicating the interpretation and limiting our ability to isolate the purity dependence.

In our baseline analysis, we obtain an HCD linear bias of $b_{\rm HCD} = -0.079^{+0.027}_{-0.024}$. For $\beta_{\rm HCD}$, we recover values consistent with the Gaussian prior set to 0.5 ± 0.20, while we are not able to robustly constrain $L_{\rm HCD}$. We expect this limitation to be alleviated as the LBG sample size increases and additional constraints (e.g., masking of DLAs) are applied.

As a complementary test to those presented in Table 3, we also examined the trend in b_α as a function of CL threshold while fixing β_α and the HCD parameters to the values derived from DESIDR2-Lyα [14] quasars at the effective redshift of our sample, with DLAs unmasked to match the LBG conditions. We find a consistent decrease in b_α with increasing sample purity: $b_\alpha = -0.1276 \pm 0.0033$ for CL > 0.97, -0.1286 ± 0.0035 for CL > 0.99, -0.1304 ± 0.0035 for CL > 0.995, and -0.1339 ± 0.0039 for CL > 0.999. This trend is similar as the observed in the LBG linear bias and reinforces the interpretation that sample purity is having a considerable impact on the inferred clustering amplitudes of both the Lyα forest and LBGs.

Acknowledgments

We are grateful to Pat McDonald for developing the original `C++` version of the Lyα fisher forecast code, which served as the basis for the `lyaforecast` code.

HKHA, EA, and CY acknowledge support from the French National Research Agency (ANR) under grant ANR-22-CE31-0009 (HZ-3D-MAP project) and grant ANR-22-CE92-0037 (DESI-Lya project).

CG, LC and AFR acknowledge support from the European Union's Horizon Europe research and innovation programme (COSMO-LYA, grant agreement 101044612). CG is also partially supported by the Spanish Ministry of Science and Innovation (MICINN) under grants PGC-2018-094773-B-C31 and SEV-2016-0588. AFR acknowledges financial support from the Spanish Ministry of Science and Innovation under the Ramon y Cajal program (RYC-2018-025210) and the PGC2021-123012NB-C41 project. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya.

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of High-Energy Physics, under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Humanities, Science and Technology of Mexico (CONAHCYT); the Ministry of Science, Innovation and Universities of Spain (MICIU/AEI/10.13039/501100011033), and by the DESI Member Institutions: https://www.desi.lbl.gov/collaborating-institutions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U. S. National Science Foundation, the U. S. Department of Energy, or any of the listed funding agencies.

The authors are honored to be permitted to conduct scientific research on I'oligam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

Data Availability Statement. This article has associated data in a data repository. All data points presented in the figures of this article will be made publicly available upon publication at https://doi.org/10.5281/zenodo.17242512, in accordance with the DESI data management policy. The DESI Early Data Release and Data Release 1 are publicly accessible at: https://data.desi.lbl.gov/doc/releases/.

References

[1] Hiram K. Herrera-Alcantar et al., https://doi.org/10.5281/zenodo.17242512

[2] DESI collaboration, *The Dark Energy Spectroscopic Instrument (DESI)*, arXiv:1907.10688 [INSPIRE].

[3] DESI collaboration, *The DESI Experiment Part II: Instrument Design*, arXiv:1611.00037 [INSPIRE].

[4] Euclid collaboration, *Euclid preparation. I. The Euclid Wide Survey*, Astron. Astrophys. **662** (2022) A112 [arXiv:2108.01201] [INSPIRE].

[5] PFS Team collaboration, *Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph*, Publ. Astron. Soc. Jap. **66** (2014) R1 [arXiv:1206.0737] [INSPIRE].

[6] G. Dalton et al., *WEAVE: the next generation wide-field spectroscopy facility for the William Herschel Telescope*, in the proceedings of the *Ground-based and Airborne Instrumentation for Astronomy IV*, Amsterdam, Netherlands, 1-6 July, (2012) [DOI:10.1117/12.925950].

[7] A.A. Meiksin, *The Physics of the Intergalactic Medium*, Rev. Mod. Phys. **81** (2009) 1405 [arXiv:0711.3358] [INSPIRE].

[8] M. Walther, J. Oñorbe, J.F. Hennawi and Z. Lukić, *New Constraints on IGM Thermal Evolution from the Lyα Forest Power Spectrum*, *Astrophys. J.* **872** (2019) 13 [arXiv:1808.04367] [INSPIRE].

[9] BOSS collaboration, *The Lyman-alpha forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data*, *JCAP* **09** (2011) 001 [arXiv:1104.5244] [INSPIRE].

[10] BOSS collaboration, *Baryon Acoustic Oscillations in the Ly-α forest of BOSS quasars*, *Astron. Astrophys.* **552** (2013) A96 [arXiv:1211.2616] [INSPIRE].

[11] BOSS collaboration, *Measurement of Baryon Acoustic Oscillations in the Lyman-alpha Forest Fluctuations in BOSS Data Release 9*, *JCAP* **04** (2013) 026 [arXiv:1301.3459] [INSPIRE].

[12] eBOSS collaboration, *The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Lyα Forests*, *Astrophys. J.* **901** (2020) 153 [arXiv:2007.08995] [INSPIRE].

[13] DESI collaboration, *DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman alpha forest*, *JCAP* **01** (2025) 124 [arXiv:2404.03001] [INSPIRE].

[14] DESI collaboration, *DESI DR2 results. I. Baryon acoustic oscillations from the Lyman alpha forest*, *Phys. Rev. D* **112** (2025) 083514 [arXiv:2503.14739] [INSPIRE].

[15] A. Font-Ribera et al., *The large-scale Quasar-Lyman α Forest Cross-Correlation from BOSS*, *JCAP* **05** (2013) 018 [arXiv:1303.1937] [INSPIRE].

[16] SDSS collaboration, *The Sloan Digital Sky Survey Quasar Survey: Quasar Luminosity Function from Data Release 3*, *Astron. J.* **131** (2006) 2766 [astro-ph/0601434] [INSPIRE].

[17] eBOSS collaboration, *The Extended Baryon Oscillation Spectroscopic Survey: Variability Selection and Quasar Luminosity Function*, *Astron. Astrophys.* **587** (2016) A41 [arXiv:1509.05607] [INSPIRE].

[18] E. Chaussidon et al., *Target Selection and Validation of DESI Quasars*, *Astrophys. J.* **944** (2023) 107 [arXiv:2208.08511] [INSPIRE].

[19] E. Rollinde et al., *The Correlation of the Lyman - alpha forest in close pairs and groups of high-redshift quasars: Clustering of matter on scales 1-5 Mpc*, *Mon. Not. Roy. Astron. Soc.* **341** (2003) 1279 [astro-ph/0302145] [INSPIRE].

[20] M.L. Abdul-Karim et al., *Measurement of the small-scale 3D Lyman-α forest power spectrum*, *JCAP* **05** (2024) 088 [arXiv:2310.09116] [INSPIRE].

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

[22] V. Ruhlmann-Kleider et al., *High redshift LBGs from deep broadband imaging for future spectroscopic surveys*, *JCAP* **08** (2024) 059 [arXiv:2404.03569] [INSPIRE].

[23] COSMOS collaboration, *First Data Release of the COSMOS Lyα Mapping and Tomography Observations: 3D Lyα Forest Tomography at $2.05 < z < 2.55$*, *Astrophys. J. Suppl.* **237** (2018) 31 [arXiv:1710.02894] [INSPIRE].

[24] B. Horowitz et al., *Second Data Release of the COSMOS Lyα Mapping and Tomography Observations: The First 3D Maps of the Detailed Cosmic Web at $2.05 < z < 2.55$*, *Astrophys. J. Suppl.* **263** (2022) 27 [arXiv:2109.09660] [INSPIRE].

[25] A.B. Newman et al., *LATIS: The Lyα Tomography IMACS Survey*, *Astrophys. J.* **891** (2020) 147 [arXiv:2002.10676] [INSPIRE].

[26] R. Momose et al., *Environmental Dependence of Galactic Properties Traced by Lyα Forest Absorption: Diversity among Galaxy Populations*, *Astrophys. J.* **909** (2021) 117 [arXiv:2002.07335].

[27] R. Momose et al., *Catch Me if You Can: Biased Distribution of Lyα-emitting Galaxies according to the Viewing Direction*, *Astrophys. J. Lett.* **912** (2021) L24 [arXiv:2104.10580].

[28] A.B. Newman et al., *LATIS: Constraints on the Galaxy–Halo Connection at $z \sim 2.5$ from Galaxy–Galaxy and Galaxy–Lyα Clustering*, *Astrophys. J.* **961** (2024) 27 [arXiv:2311.00184] [INSPIRE].

[29] B. Zhang et al., *Cross-correlations between the CLAMATO Lyman-alpha forest and galaxies within the COSMOS field*, arXiv:2509.07453 [INSPIRE].

[30] K.L. Adelberger, C.C. Steidel, A.E. Shapley and M. Pettini, *Galaxies and intergalactic matter at redshift $Z \sim 3$: Overview*, *Astrophys. J.* **584** (2003) 45 [astro-ph/0210314] [INSPIRE].

[31] K.L. Adelberger et al., *The Connection between galaxies and intergalactic absorption lines at redshift 2<= z <=3*, *Astrophys. J.* **629** (2005) 636 [astro-ph/0505122] [INSPIRE].

[32] R.M. Bielby et al., *The VLT LBG redshift survey – VI. Mapping H i in the proximity of $z \sim 3$ LBGs with X-Shooter*, *Mon. Not. Roy. Astron. Soc.* **471** (2017) 2174 [arXiv:1610.09144].

[33] Y. Liang et al., *Statistical Correlation between the Distribution of Lyα Emitters and Intergalactic Medium H i at $z \sim 2.2$ Mapped by the Subaru/Hyper Suprime-Cam*, *Astrophys. J.* **907** (2021) 3 [arXiv:2008.01733].

[34] DESI collaboration, *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **164** (2022) 207 [arXiv:2205.10939] [INSPIRE].

[35] DESI collaboration, *The Robotic Multiobject Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI)*, *Astron. J.* **165** (2023) 9 [arXiv:2205.09014] [INSPIRE].

[36] DESI collaboration, *The Optical Corrector for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **168** (2024) 95 [arXiv:2306.06310] [INSPIRE].

[37] C. Poppett et al., *Overview of the Fiber System for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **168** (2024) 245.

[38] DESI collaboration, *Survey Operations for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **166** (2023) 259 [arXiv:2306.06309] [INSPIRE].

[39] J. Ratajczak et al., *The Compilation and Validation of the Spectroscopic Redshift Catalogs for the DESI-COSMOS and DESI-XMMLSS Fields*, arXiv:2508.09286 [INSPIRE].

[40] H. Aihara et al., *Second Data Release of the Hyper Suprime-Cam Subaru Strategic Program*, *Publ. Astron. Soc. Jap.* **71** (2019) 114 [arXiv:1905.12221] [INSPIRE].

[41] M. Sawicki et al., *The CFHT Large Area U-band Deep Survey (CLAUDS)*, *Mon. Not. Roy. Astron. Soc.* **489** (2019) 5202 [arXiv:1909.05898] [INSPIRE].

[42] C. Payerne et al., *Selection of high-redshift Lyman-Break Galaxies from broadband and wide photometric surveys*, *JCAP* **05** (2025) 031 [arXiv:2410.08062] [INSPIRE].

[43] S. Gwyn et al., *UNIONS: The Ultraviolet Near-Infrared Optical Northern Survey*, arXiv:2503.13783.

[44] DESI collaboration, *The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **165** (2023) 144 [arXiv:2209.14482] [INSPIRE].

JCAP12(2025)053

[45] N. Busca and C. Balland, *QuasarNET: Human-level spectral classification and redshifting with Deep Neural Networks*, arXiv:1808.09955 [INSPIRE].

[46] D. Green et al., *Using active learning to improve quasar identification for the DESI spectra processing pipeline*, JCAP **10** (2025) 087 [arXiv:2505.01596] [INSPIRE].

[47] S. Bailey et al., *Redrock: Spectroscopic Classification and Redshift Fitting for the Dark Energy Spectroscopic Instrument*, https://github.com/desihub/redrock, *in preparation*.

[48] A.A. Khostovan et al., *COSMOS Spectroscopic Redshift Compilation (First Data Release): 488k Redshifts Encompassing Two Decades of Spectroscopy*, arXiv:2503.00120.

[49] H. du Mas des Bourboux, J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, T. Etourneau et al., *picca: Package for Igm Cosmological-Correlations Analyses. Astrophysics Source Code Library*, (2021).

[50] DESI collaboration, *The Lyman-α forest catalog from the Dark Energy Spectroscopic Instrument Early Data Release*, Mon. Not. Roy. Astron. Soc. **528** (2024) 6666 [arXiv:2306.06312] [INSPIRE].

[51] B. Wang et al., *Deep Learning of Dark Energy Spectroscopic Instrument Mock Spectra to Find Damped Lyα Systems*, Astrophys. J. Supp. **259** (2022) 28 [arXiv:2201.00827] [INSPIRE].

[52] M.-F. Ho, S. Bird and R. Garnett, *Detecting Multiple DLAs per Spectrum in SDSS DR12 with Gaussian Processes*, Mon. Not. Roy. Astron. Soc. **496** (2020) 5436 [arXiv:2003.11036] [INSPIRE].

[53] DESI collaboration, *Construction of the damped Lyα absorber catalog for DESI DR2 Lyα BAO*, Phys. Rev. D **112** (2025) 083510 [arXiv:2503.14740] [INSPIRE].

[54] A.E. Shapley, C.C. Steidel, M. Pettini and K.L. Adelberger, *Rest - frame ultraviolet spectra of Z∼3 Lyman Break Galaxies*, Astrophys. J. **588** (2003) 65 [astro-ph/0301230] [INSPIRE].

[55] J.S. Monzon, J.X. Prochaska, K.-G. Lee and J. Chisholm, *Effective Opacity of the Intergalactic Medium from Galaxy Spectra Analysis*, Astron. J. **160** (2020) 37 [arXiv:2005.09667] [INSPIRE].

[56] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, Astron. Astrophys. **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [arXiv:1807.06209] [INSPIRE].

[57] SDSS collaboration, *The Lyman-alpha forest power spectrum from the Sloan Digital Sky Survey*, Astrophys. J. Suppl. **163** (2006) 80 [astro-ph/0405013] [INSPIRE].

[58] C. Gordon et al., *3D Correlations in the Lyman-α Forest from Early DESI Data*, JCAP **11** (2023) 045 [arXiv:2308.10950] [INSPIRE].

[59] eBOSS collaboration, *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].

[60] M.J. Wilson and M. White, *Cosmology with dropout selection: straw-man surveys & CMB lensing*, JCAP **10** (2019) 015 [arXiv:1904.13378] [INSPIRE].

[61] 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].

[62] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, Astrophys. J. **538** (2000) 473 [astro-ph/9911177] [INSPIRE].

[63] A. Arinyo-i-Prats, J. Miralda-Escudé, M. Viel and R. Cen, *The Non-Linear Power Spectrum of the Lyman Alpha Forest*, JCAP **12** (2015) 017 [arXiv:1506.04519] [INSPIRE].

[64] S. Chabanier et al., *The ACCEL2 project: simulating Lyman-α forest in large-volume hydrodynamical simulations*, Mon. Not. Roy. Astron. Soc. **534** (2024) 2674 [arXiv:2407.04473] [INSPIRE].

[65] N. Busca et al., *The effects of continuum fitting on Lyman-α forest correlations*, JCAP **09** (2025) 020 [arXiv:2506.15262] [INSPIRE].

[66] M. McQuinn and M. White, *On Estimating Lyman-alpha Forest Correlations between Multiple Sightlines*, Mon. Not. Roy. Astron. Soc. **415** (2011) 2257 [arXiv:1102.1752] [INSPIRE].

[67] K.K. Rogers et al., *Simulating the effect of high column density absorbers on the one-dimensional Lyman α forest flux power spectrum*, Mon. Not. Roy. Astron. Soc. **474** (2018) 3032 [arXiv:1706.08532] [INSPIRE].

[68] T. Tan et al., *Modeling of the high column density systems in the Lyman-Alpha forest*, JCAP **11** (2025) 074 [arXiv:2506.13005] [INSPIRE].

[69] A. Font-Ribera and J. Miralda-Escude, *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].

[70] BOSS collaboration, *The large-scale cross-correlation of Damped Lyman Alpha Systems with the Lyman Alpha Forest: First Measurements from BOSS*, JCAP **11** (2012) 059 [arXiv:1209.4596] [INSPIRE].

[71] 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].

[72] DESI collaboration, *Characterization of contaminants in the Lyman-alpha forest auto-correlation with DESI*, JCAP **01** (2025) 140 [arXiv:2404.03003] [INSPIRE].

[73] R. Bielby et al., *The VLT LBG Redshift Survey - III. The clustering and dynamics of Lyman-break galaxies at $z \sim 3$*, Mon. Not. Roy. Astron. Soc. **430** (2013) 425 [arXiv:1204.3635] [INSPIRE].

[74] C. Jose, K. Subramanian, R. Srianand and S. Samui, *Spatial Clustering of High Redshift Lyman Break Galaxies*, Mon. Not. Roy. Astron. Soc. **429** (2013) 2333 [arXiv:1208.2097] [INSPIRE].

[75] R.M. Bielby et al., *The VLT LBG redshift survey – V. Characterizing the z = 3.1 Lyman α emitter population*, Mon. Not. Roy. Astron. Soc. **456** (2016) 4061 [arXiv:1501.01215] [INSPIRE].

[76] M. White et al., *The clustering of Lyman Alpha Emitting galaxies at $z = 2 - 3$*, JCAP **08** (2024) 020 [arXiv:2406.01803] [INSPIRE].

[77] H. Ebina et al., *Clustering analysis of medium-band selected high-redshift galaxies*, arXiv:2509.26467 [INSPIRE].

[78] S. Eftekharzadeh et al., *Clustering of intermediate redshift quasars using the final SDSS III-BOSS sample*, Mon. Not. Roy. Astron. Soc. **453** (2015) 2779 [arXiv:1507.08380] [INSPIRE].

[79] B. Chehade et al., *The 2QDES Pilot : The luminosity and redshift dependence of quasar clustering*, Mon. Not. Roy. Astron. Soc. **459** (2016) 1179 [arXiv:1603.04849] [INSPIRE].

[80] S. de Beer et al., *Resolving the physics of quasar Lyα nebulae (RePhyNe): I. Constraining quasar host halo masses through circumgalactic medium kinematics*, Mon. Not. Roy. Astron. Soc. **526** (2023) 1850 [arXiv:2309.01506] [INSPIRE].

[81] A.M. Eltvedt et al., *The VST ATLAS Quasar Survey – III. Halo mass function via quasar clustering and quasar-CMB lensing cross-clustering*, Mon. Not. Roy. Astron. Soc. **535** (2024) 2105 [arXiv:2410.23372] [INSPIRE].

[82] DESI collaboration, *A Spectroscopic Road Map for Cosmic Frontier: DESI, DESI-II, Stage-5*, arXiv:2209.03585 [INSPIRE].

[83] 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].

[84] H. Dembinski and P.O. et al., *scikit-hep/iminuit*, https://github.com/scikit-hep/iminuit.

[85] W.J. Handley, M.P. Hobson and A.N. Lasenby, *PolyChord: nested sampling for cosmology*, Mon. Not. Roy. Astron. Soc. **450** (2015) L61 [arXiv:1502.01856] [INSPIRE].

[86] W.J. Handley, M.P. Hobson and A.N. Lasenby, *polychord: next-generation nested sampling*, Mon. Not. Roy. Astron. Soc. **453** (2015) 4385 [arXiv:1506.00171] [INSPIRE].

[87] C. Ramírez-Pérez, J. Sanchez, D. Alonso and A. Font-Ribera, *CoLoRe: fast cosmological realisations over large volumes with multiple tracers*, JCAP **05** (2022) 002 [arXiv:2111.05069] [INSPIRE].

[88] DESI collaboration, *Validation of the DESI DR2 Lyα BAO analysis using synthetic datasets*, arXiv:2503.14741 [INSPIRE].

[89] J. Farr et al., *LyaCoLoRe: synthetic datasets for current and future Lyman-α forest BAO surveys*, JCAP **03** (2020) 068 [arXiv:1912.02763] [INSPIRE].

[90] H.K. Herrera-Alcantar et al., *Synthetic spectra for Lyman-α forest analysis in the Dark Energy Spectroscopic Instrument*, JCAP **01** (2025) 141 [arXiv:2401.00303] [INSPIRE].

[91] D. Kirkby, S. Bailey, J. Guy and B.A. Weaver, *Quick simulations of fiber spectrograph response v0.5*, (2016) https://zenodo.org/records/154130.

[92] DESI collaboration, *Validation of the DESI 2024 Lyα forest BAO analysis using synthetic datasets*, JCAP **01** (2025) 148 [arXiv:2404.03004] [INSPIRE].

[93] A. Cuceu, A. Font-Ribera, B. Joachimi and S. Nadathur, *Cosmology beyond BAO from the 3D distribution of the Lyman-α forest*, Mon. Not. Roy. Astron. Soc. **506** (2021) 5439 [arXiv:2103.14075] [INSPIRE].

[94] A. Cuceu et al., *The Alcock–Paczyński effect from Lyman-α forest correlations: analysis validation with synthetic data*, Mon. Not. Roy. Astron. Soc. **523** (2023) 3773 [arXiv:2209.12931] [INSPIRE].

[95] A. Cuceu et al., *Constraints on the Cosmic Expansion Rate at Redshift 2.3 from the Lyman-α Forest*, Phys. Rev. Lett. **130** (2023) 191003 [arXiv:2209.13942] [INSPIRE].

[96] A. Cuceu et al., *DESI DR1 Lyα forest: 3D full-shape analysis and cosmological constraints*, arXiv:2509.15308 [INSPIRE].

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

[98] A. Font-Ribera et al., *DESI and other dark energy experiments in the era of neutrino mass measurements*, JCAP **05** (2014) 023 [arXiv:1308.4164] [INSPIRE].

[99] M. Tegmark, *Measuring cosmological parameters with galaxy surveys*, Phys. Rev. Lett. **79** (1997) 3806 [astro-ph/9706198] [INSPIRE].

[100] C.R. Rao, *Information and the Accuracy Attainable in the Estimation of Statistical Parameters*, in *Breakthroughs in Statistics*, Springer New York (1992), p. 235–247 [DOI:10.1007/978-1-4612-0919-5_16].

[101] D.M. Alexander et al., *The DESI Survey Validation: Results from Visual Inspection of the Quasar Survey Spectra*, *Astron. J.* **165** (2023) 124 [arXiv:2208.08517] [INSPIRE].

[102] DESI collaboration, *Performance of the Quasar Spectral Templates for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **166** (2023) 66 [arXiv:2305.10426] [INSPIRE].

[103] A. Krolewski et al., *Impact and mitigation of spectroscopic systematics on DESI DR1 clustering measurements*, *JCAP* **01** (2025) 147 [arXiv:2405.17208] [INSPIRE].

[104] J. Yu et al., *ELG spectroscopic systematics analysis of the DESI Data Release 1*, *JCAP* **01** (2025) 126 [arXiv:2405.16657] [INSPIRE].

[105] DESI collaboration, *DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations*, *JCAP* **02** (2025) 021 [arXiv:2404.03002] [INSPIRE].

[106] DESI collaboration, *DESI DR2 results. II. Measurements of baryon acoustic oscillations and cosmological constraints*, *Phys. Rev. D* **112** (2025) 083515 [arXiv:2503.14738] [INSPIRE].

[107] C. Zhao et al., *MUltiplexed Survey Telescope: Perspectives for Large-Scale Structure Cosmology in the Era of Stage-V Spectroscopic Survey*, arXiv:2411.07970 [INSPIRE].

[108] Spec-S5 collaboration, *The Spectroscopic Stage-5 Experiment*, arXiv:2503.07923 [INSPIRE].

[109] WST collaboration, *WST - Widefield Spectroscopic Telescope: motivation, science drivers and top level requirements for a new dedicated facility*, *Proc. SPIE Int. Soc. Opt. Eng.* **13094** (2024) 130941O [arXiv:2405.12518] [INSPIRE].

[110] DESI collaboration, *Overview of the DESI Legacy Imaging Surveys*, *Astron. J.* **157** (2019) 168 [arXiv:1804.08657] [INSPIRE].

[111] LSST Dark Energy Science collaboration, *Large Synoptic Survey Telescope: Dark Energy Science Collaboration*, arXiv:1211.0310 [DOI:10.2172/1156445] [INSPIRE].

[112] M. Pettini et al., *The Rest frame optical spectra of Lyman break galaxies: Star formation, extinction, abundances, and kinematics*, *Astrophys. J.* **554** (2001) 981 [astro-ph/0102456] [INSPIRE].

[113] C.C. Steidel et al., *The Structure and Kinematics of the Circum-Galactic Medium from Far-UV Spectra of $z \sim 2$-3 Galaxies*, *Astrophys. J.* **717** (2010) 289 [arXiv:1003.0679] [INSPIRE].

[114] O. Rakic, J. Schaye, C.C. Steidel and G.C. Rudie, *Calibrating Galaxy Redshifts Using Absorption by the Surrounding Intergalactic Medium*, *Mon. Not. Roy. Astron. Soc.* **414** (2011) 3265 [arXiv:1011.1282] [INSPIRE].

[115] T. Hashimoto et al., *Gas Motion Study of Lya Emitters at $z \sim 2$ Using UV and Optical Spectral Lines*, *Astrophys. J.* **765** (2013) 70 [arXiv:1206.2316] [INSPIRE].

[116] T. Shibuya et al., *What is the Physical Origin of Strong Lyα Emission? I. Demographics of Lyα Emitter Structures*, *Astrophys. J.* **785** (2014) 64 [arXiv:1401.1209] [INSPIRE].

JCAP12(2025)053

Author List

Hiram K. Herrera-Alcantar[1,2], Eric Armengaud[2], Christophe Yèche[2], Calum Gordon[3], Laura Casas[3], Andreu Font-Ribera[3], Christophe Magneville[2], Corentin Ravoux[4], J. Aguilar[5], S. Ahlen[6], A. Anand[5], D. Brooks[7], E. Chaussidon[5], T. Claybaugh[5], A. Cuceu[5], K. S. Dawson[8], A. de la Macorra[9], Arjun Dey[26], P. Doel[7], S. Ferraro[5,10], J. E. Forero-Romero[11,12], E. Gaztañaga[13,14,15], S. Gontcho A Gontcho[5,16], A. X. Gonzalez-Morales[17], G. Gutierrez[18], J. Guy[5], C. Hahn[19], D. Kirkby[20], A. Kremin[5], O. Lahav[7], A. Lambert[5], M. Landriau[5], L. Le Guillou[21], M. Manera[22,3], P. Martini[23,24,25], A. Meisner[26], R. Miquel[27,3], A. Muñoz-Gutiérrez[9], S. Nadathur[14], N. Palanque-Delabrouille[2,5], W. J. Percival[28,29,30], F. Prada[31], I. Pérez-Ràfols[32], G. Rossi[33], E. Sanchez[34], D. Schlegel[5], M. Schubnell[35,36], J. Silber[5], D. Sprayberry[26], G. Tarlé[36], B. A. Weaver[26], R. Zhou[5], H. Zou[37]

[1] *Institut d'Astrophysique de Paris. 98 bis boulevard Arago. 75014 Paris, France*
[2] *IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[3] *Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Edifici Cn, Campus UAB, 08193, Bellaterra (Barcelona), Spain*
[4] *Université Clermont-Auvergne, CNRS, LPCA, 63000 Clermont-Ferrand, France*
[5] *Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, U.S.A.*
[6] *Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, MA 02215 U.S.A.*
[7] *Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, U.K.*
[8] *Department of Physics and Astronomy, The University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, U.S.A.*
[9] *Instituto de Física, Universidad Nacional Autónoma de México, Circuito de la Investigación Científica, Ciudad Universitaria, Cd. de México C. P. 04510, México*
[10] *University of California, Berkeley, 110 Sproul Hall #5800 Berkeley, CA 94720, U.S.A.*
[11] *Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, CP 111711, Bogotá, Colombia*
[12] *Observatorio Astronómico, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio H, CP 111711 Bogotá, Colombia*
[13] *Institut d'Estudis Espacials de Catalunya (IEEC), c/ Esteve Terradas 1, Edifici RDIT, Campus PMT-UPC, 08860 Castelldefels, Spain*
[14] *Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth, PO1 3FX, U.K.*
[15] *Institute of Space Sciences, ICE-CSIC, Campus UAB, Carrer de Can Magrans s/n, 08913 Bellaterra, Barcelona, Spain*
[16] *University of Virginia, Department of Astronomy, Charlottesville, VA 22904, U.S.A.*
[17] *Departamento de Física, DCI-Campus León, Universidad de Guanajuato, Loma del Bosque 103, León, Guanajuato C. P. 37150, México*
[18] *Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, U.S.A.*
[19] *Steward Observatory, University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, U.S.A.*
[20] *Department of Physics and Astronomy, University of California, Irvine, 92697, U.S.A.*
[21] *Sorbonne Université, CNRS/IN2P3, Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), FR-75005 Paris, France*
[22] *Departament de Física, Serra Húnter, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*
[23] *Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, U.S.A.*

JCAP12(2025)053

[24] *Department of Astronomy, The Ohio State University, 4055 McPherson Laboratory, 140 W 18th Avenue, Columbus, OH 43210, U.S.A.*
[25] *The Ohio State University, Columbus, 43210 OH, U.S.A.*
[26] *NSF NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, U.S.A.*
[27] *Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys, 23, 08010 Barcelona, Spain*
[28] *Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[29] *Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada*
[30] *Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[31] *Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, s/n, E-18008 Granada, Spain*
[32] *Departament de Física, EEBE, Universitat Politècnica de Catalunya, c/Eduard Maristany 10, 08930 Barcelona, Spain*
[33] *Department of Physics and Astronomy, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea*
[34] *CIEMAT, Avenida Complutense 40, E-28040 Madrid, Spain*
[35] *Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, MI 48109, U.S.A.*
[36] *University of Michigan, 500 S. State Street, Ann Arbor, MI 48109, U.S.A.*
[37] *National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Road, Chaoyang District, Beijing, 100101, P. R. China*

JCAP12(2025)053