

First Ly- α 1D bispectrum measurement in eBOSS

Rodrigo de la Cruz ^a, Gustavo Niz ^a, Vid Iršič ^{b,c}, Corentin Ravoux ^{d,e},
César Ramírez^f and Hiram K. Herrera-Alcantar ^{a,g,h}

^aDepartamento de Física, División de Ciencias e Ingenierías, Universidad de Guanajuato, Campus León, C.P. 37150, León, Mexico

^bKavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, U.K.

^cCavendish Laboratory, University of Cambridge, 19 J.J. Thomson Ave., Cambridge CB3 0HE, U.K.

^dAix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France

^eUniversité Clermont-Auvergne, CNRS, LPCA, 63000 Clermont-Ferrand, France

^fInstitut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, 08193 Bellaterra (Barcelona), Spain

^gUniversité Paris-Saclay, CEA, IRFU, 91191, Gif-sur-Yvette, France

^hSorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France

E-mail: rodrigo.cruz.noriega93@gmail.com, g.niz@ugto.mx,
v.irsic@herts.ac.uk, corentin.ravoux@clermont.in2p3.fr,
cramirez@ifae.es, hiram.herreraalcantar@cea.fr

ABSTRACT: We present the first robust measurement of the one-dimensional Lyman alpha (Ly- α) forest bispectrum using the complete extended Baryon Oscillation Spectroscopic Survey (eBOSS) quasar sample, corresponding to the sixteenth data release (DR16) of the Sloan Digital Sky Survey (SDSS). The measurement employs an FFT estimator over 12 redshift bins, ranging from $z = 2.2$ to $z = 4.4$, and extends to scales of $0.017 \text{ (km/s)}^{-1}$. The used sample consists of 122,066 quasar spectra; however, only the first six redshift bins contain sufficient data to extract a physical bispectrum. To validate and correct the bispectrum measurement, we use synthetic datasets generated from lognormal and 2LPT mocks. Additionally, we detect clear evidence of correlations between Si_{III} absorption lines and the Ly- α forest within the bispectrum signal, which we model using an extension of the approach applied to the analogous one-dimensional power spectrum signal. In this context, the pipeline developed for this study addresses the impact of instrumental and methodological systematics and is

ready for application to larger spectroscopic datasets, such as those from the Dark Energy Spectroscopic Instrument (DESI). Finally, we compare the signal to a simple perturbation theory model, where a χ^2 analysis shows a reasonable fit for specific bispectrum configurations with $z \leq 3.2$; suggesting that higher-order one-dimensional statistics in the Ly- α forest can complement cosmological inference based on the power spectrum in future analyses.

KEYWORDS: galaxy surveys, Lyman alpha forest, non-gaussianity

ARXIV EPRINT: [2410.09150](https://arxiv.org/abs/2410.09150)

Contents

1	Introduction and summary of main results	1
2	Overview of the data and the bispectrum formalism	4
2.1	Observational dataset	4
2.2	Methodology	5
2.3	Ly- α bispectrum formalism and estimator	8
2.4	Theoretical description of B^{1D}	11
3	Data analysis and P^{1D} assessment	12
3.1	Summary of the Ly- α sample	12
3.2	Spectrograph resolution	13
3.3	Noise bispectrum measurement	13
3.4	Side-band bispectrum	16
3.5	1D power spectrum assessment	18
4	Synthetic data correction	20
4.1	Mocks	20
4.2	Continuum-fitting correction	21
4.3	Spectrum pixel masking correction	22
5	Systematic uncertainties estimation	24
6	Results: first measurement of the Ly-α B^{1D} in eBOSS	28
6.1	B^{1D} measurement	28
6.2	Si _{III} cross-correlation	32
7	Conclusions	34
A	Maximal q scale and signal for different triangle configurations	36

1 Introduction and summary of main results

The intergalactic medium (IGM) resides in the low-density regions between galaxies and is important to various astrophysical disciplines. Understanding IGM has been essential since the pioneering work of Gunn and Peterson [1] in the 1960s, in which it was first used to test models of structure formation on the smallest comoving scales (see also [2–5]). Variations in the IGM’s density are observed in quasar (QSO) spectra as a continuous series of absorption features. Although various atomic transitions serve as biased tracers of this density field, the strongest signal arises from the Lyman-alpha (Ly- α) transition of neutral hydrogen atoms within the IGM. These absorption features in QSO spectra are referred to as the Ly- α forest, commonly studied at redshifts between 2 and 5 (see [6] for a review).

The small-scale IGM distribution is captured in the fluctuations of the Ly- α forest superimposed on the quasar continuum, defined as the QSO's unabsorbed spectrum. These fluctuations can be analyzed through different summary statistics, either in Fourier or real space. One approach in Fourier space is through the three-dimensional power spectrum (P^{3D}), a correlation between QSO absorption across multiple lines of sight separated by small angles on the sky (see, for example, [7–10]). Due to the observational nature of the Ly- α forest, there is a high-resolution dataset along lines of sight with a limited sample in the transverse directions, which poses a technical challenge for measuring the P^{3D} . The advent of new Ly- α forest surveys with a greater number of lines of sight, such as the Dark Energy Spectroscopic Instrument (DESI) [11, 12], offers a realistic prospect of measuring the P^{3D} .

A more straightforward approach is to study the Ly- α forest power spectrum averaged over the line of sight, known as the one-dimensional power spectrum (P^{1D}). Early measurements of P^{1D} can be found in [13–16]. In more recent years, medium-resolution surveys have measured P^{1D} using data from the Sloan Digital Sky Survey (SDSS) ([17]), the Baryon Oscillation Spectroscopic Survey (BOSS) and the extended BOSS (eBOSS). These analyses started with catalogues of $\mathcal{O}(10^3)$ spectra [18], to conclude with the larger SDSS DR9 (BOSS) sample used in [19] or the SDSS DR16 (eBOSS) data in [20]. More recently, P^{1D} was measured using the first DESI data using a Fast Fourier Transform (FFT) approach ([21]) and compared with the quadratic maximum likelihood estimator (QMLE) approach from [22] (or [19] for BOSS); both methods yield consistent results. Moreover, high-resolution P^{1D} measurements have also been made using SQUAD data ([23], $\mathcal{O}(4 \times 10^2)$ spectra), KODIAK ([24], $\mathcal{O}(3 \times 10^2)$ spectra), and XQ-100 data ([25], $\mathcal{O}(1 \times 10^2)$ spectra). Further examples of P^{1D} analysis are found in [26–33].

The P^{1D} analysis provides valuable information on structure formation, as seen in [34–37]. It is also a tool for studying neutrino masses, as demonstrated by [38–40]. To extract more information from Ly- α forest data, one can either use higher-order Lyman series (e.g., [41, 42]) or higher-order statistics such as the one-dimensional bispectrum (B^{1D}).

In this work, we focus on higher-order statistics, specifically the bispectrum, which provides a complementary description of non-Gaussian features in the Ly- α forest. The bispectrum quantifies correlations between three Fourier modes, offering insight into the non-linear processes shaping the large-scale structure of the universe ([43–46]). Zaldarriaga et al. [47] mention that gravitational growth induces correlations between large-scale modes and small-scale power, which can be demonstrated through the bispectrum. These three-point correlations can also distinguish between fluctuations caused by large-scale structure in the matter distribution and those arising from non-gravitational processes, such as variations in the continuum emission of quasars. Other physically interesting phenomena can be studied using higher-order statistics, including the recent parity study by [48] employing the four-point correlation function.

Several studies in the literature have explored the use of the Ly- α bispectrum. Notably, Viel et al. [49] measured B^{1D} using 27 high-resolution quasar spectra in the redshift range of 2.0 to 2.4, and concluded that the signal was not strong enough to use it for a cosmological analysis. In contrast, the works of [47] and [50] measured a three-point correlation, which is not exactly what we will define as the B^{1D} . All these contemporary studies [49–51] discuss the advantages of a bispectrum analysis, either to break degeneracies between the mean flux transmission, \bar{F} , and the amplitude of the power spectrum, or to improve constraints

on the model parameters via the joint analysis with the power spectrum [50]; as well as to constrain primordial non-Gaussianities [51]. The accuracy achieved in [49] is reported to be within the 50% relative to the preferred Λ CDM model at the time, while the authors of [50] emphasized the need for a more detailed study of systematics affecting the bispectrum, before it can be considered for cosmological analysis. These two aspects, improved accuracy and understanding of systematic errors, are two of the main goals of this work. In [52], the three-point correlation function was measured using hydrodynamical simulations, decomposing the Ly- α forest into multiple Voigt profile components (clouds). Another interesting work by [53] measured the squeezed-limit cross-bispectrum between the Ly- α power spectrum and the long-wavelength quasar overdensity using simulations. This approach is similar to the position-dependent power spectrum approach ([54]). Additionally, [55] attempted to place stringent constraints on primordial non-Gaussianity by analyzing the measured 3D bispectrum of the Ly- α forest. In [56], they further pursued non-Gaussianity testing, this time by incorporating the three-dimensional 21 cm bispectrum and examining the cross-bispectrum between the 21 cm brightness temperature and Ly- α transmitted flux fields. With this background on the Ly- α bispectrum (B^{1D}) established, we can now describe the main results of this work.

Our primary result is the first measurement of the one-dimensional bispectrum in the Ly- α forest using data from a large galaxy spectroscopic survey, specifically the eBOSS DR16 dataset. We estimate the B^{1D} signal in 12 redshift bins, although only six of these bins yield sufficient statistical significance. After carefully analyzing the noise and other systematics, we show that the signal is compatible with simple models based on either second-order perturbation theory (2OPT) or synthetic mocks generated using LogNormal and second-order Lagrangian Perturbation Theory (2LPT) simulation methods. Figure 1 summarizes our main results on the B^{1D} measurement from eBOSS.

Three-point correlations are defined over triangular configurations in real or Fourier space. In the case of B^{1D} , the function is supported over colinear triangles, defined by two momentum scales (q_0, q_1). The left panel of figure 1 shows the full B^{1D} for all triangular configurations, while the right panel focuses on the isosceles triangles. These isosceles configurations, corresponding to the diagonal in the q_0 - q_1 plane, exhibit the strongest bispectrum signal. Although a more detailed discussion will follow, there is clear redshift evolution in the signal, consistent with the 2OPT model. Moreover, as seen in the density plot in the left panel, periodic darker regions are observed, which result from the cross-correlation between the absorption of the Ly- α and the Si_{III} quasar emission lines. Another notable result is the first measurement of the one-dimensional bispectrum in the sidebands, caused by metals on the red side of the Ly- α peak. Finally, the pipeline presented in this work can be directly applied to the next generation of large-scale spectroscopic Ly- α surveys, such as DESI [12], or the high-resolution quasar data from SQUAD [23].

The rest of the paper is structured as follows: in section 2, we present the three-point statistic formalism and the pipeline to perform the B^{1D} measurement, as well as the theoretical model of the bispectrum based on second-order perturbation theory. Section 3 focuses on data analysis. In the following section, we describe the generation of synthetic spectra using Gaussian random field methods and Lagrangian perturbation theory, extending to mildly nonlinear scales. The treatment of statistical and systematic uncertainties for the B^{1D} measurement is detailed in section 5. In section 6, we present our measurements using

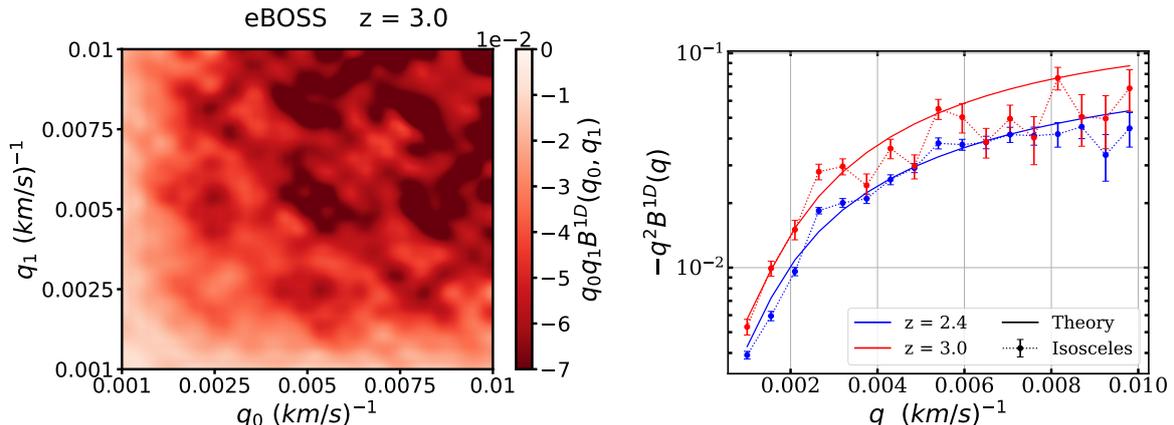


Figure 1. One-dimensional bispectrum in the Ly- α forest using eBOSS DR16 data. *Left:* 2D B^{1D} measurement for all triangle configurations along the line of sight at a redshift of $z = 3.0$. The dark areas are caused by the cross-correlations between Ly- α and Si_{III}. *Right:* the 1D signal is obtained by extracting data from the diagonal of the 2D signal. The data points with error bars depict the B^{1D} measurements for isosceles triangles ($q_0 = q_1$). Results are presented for two specific redshift bins: one at $z = 2.4$ (blue) and the other at $z = 3.0$ (red). The variation in amplitude between these signals is attributed to the evolution of the B^{1D} with redshift. The solid lines, corresponding to the same redshifts as the data points, represent the theoretical modeling of the signal based on second-order perturbation theory. These measurements were made by applying eq. (2.9). The error bars are derived from the square root of the diagonal of the bootstrap covariance matrix and systematic uncertainties added in quadrature.

eBOSS DR16 data, along with comparisons to the second-order perturbation theory mode. Finally, in the last section we provide conclusions.

2 Overview of the data and the bispectrum formalism

2.1 Observational dataset

The DR16 quasar catalogue includes 750,132 confirmed quasars [57], covering wavelengths from 3,600 Å to 10,400 Å, with a spectral resolution of $\Delta\lambda = 0.11$ Å. The dataset spans a redshift range of $0.0 < z < 6.0$. For this paper, we focus only on quasars with a Ly- α forest visible in their spectra, applying a redshift cut of $z_{\text{Ly-}\alpha} \geq 2.1$. This subset consists of 193,346 quasar spectra with Ly- α forests, visually confirmed redshifts, and covering the range $2.1 < z < 4.8$. Furthermore, we limit the wavelength range to $3,650 \text{ \AA} < \lambda < 7,235 \text{ \AA}$ to avoid the CCD camera edges. For quasars with multiple observations, we use the combined spectra from all exposures on a given plate. To study redshift dependence, we consider a sub-sample with $2.2 < z < 4.4$, divided into 12 bins of size $\Delta z = 0.2$, with magnitudes in the r-band between 16.25 and 22.75.

Furthermore, 117,458 spectra containing Damped Lyman Alpha systems (DLA) were identified by [58] within the same redshift range. To detect more DLAs, we apply a simple DLA finder, which first calculates the observed flux power spectrum (fPk) and, using the continuum flux, computes the transmission power spectrum (TPk). If $\langle \text{TPk} \rangle / \langle \text{fPk} \rangle \geq 2.5$, the spectrum is flagged as containing a DLA, followed by manual visual inspections. In some

cases, these visually inspected spectra exhibit significant portions of the Ly- α forest being cut, resulting in the exclusion of those quasars from the sample.

Finally, 27,117 quasar spectra with Broad Absorption Lines (BALs) have been identified [57]. The final catalogue for our analysis excludes these BAL quasars, identified by a non-zero BI_CIV flag, reducing the final sample to 166,229 Ly- α quasars.

2.2 Methodology

The starting point consists of cleaning the initial sample. To that end, we first remove spectra with fewer than 50 pixels in the forest or where the mean spectral resolution is larger than 85 km/s. Moreover, since sky lines affect the P^{1D} and B^{1D} by increasing the pixel noise, we mask several sky lines in each forest ¹ and instead of removing the entire pixel, we interpolate between neighboring pixels. This interpolation is performed before computing the continuum fit to avoid affecting it. However, once we calculate the delta field from the fitted continuum, we set the corresponding sky-line pixels to zero before performing the Fourier transform. We assess the impact of this masking scheme when discussing systematics. Additional steps for cleaning the initial dataset include accounting for galactic extinction and discarding bad spectra, which we define as those with a negative mean flux or no pixels in the Ly- α forest region. Finally, for the initial dataset, we mask DLAs using the catalogue from [58], at the redshift z_{abs} , in the range $[\lambda_C - W/2, \lambda_C + W/2]$, where $\lambda_C = (1 + z_{\text{abs}})\lambda_{\text{Ly-}\alpha}$ and W is the equivalent width given by equation (20) of [27], $W = 7.3(1 + z_{\text{abs}})\sqrt{N_{H_I}/10^{20} \text{ cm}^{-2}} \text{ \AA}$. Additional DLA systems identified by our simple DLA finder are removed from the sample.

For the B^{1D} analysis, we restrict the Ly- α forest region to the range $1050 \text{ \AA} < \lambda_{\text{R.F.}} < 1180 \text{ \AA}$ to avoid contamination of the correlation functions by astrophysical effects near the quasars [20]. We use a pixel size in velocity units of $\Delta v(\lambda) = 69 \text{ km/s}$. This forest region is shown for an average quasar in figure 2. We assume an average signal-to-noise ratio \overline{SNR} ,² greater than 1.0.

The Ly- α forest region spans a redshift range $\Delta z \sim 0.4$ for quasars at $z_{\text{Ly-}\alpha} = 2.4$, $\Delta z \sim 0.5$ at $z_{\text{Ly-}\alpha} = 3.2$, and $\Delta z \sim 0.6$ at $z_{\text{Ly-}\alpha} = 4.4$. To reduce redshift correlations, it is common to split the forest into three subregions (each containing about 170 pixels per QSO spectrum for the eBOSS sample) and assign them to different redshift bins when calculating P^{1D} ([19–21]). However, in this work, we divide the forest into chunks for each line of sight and assign each chunk to its corresponding redshift bin. These chunks correspond to consecutive, non-overlapping sub-forests of different lengths. Each chunk spans at most $\Delta z = 0.2$. In addition to the previous cuts, each chunk must pass two further cuts. To improve the quality of the P^{1D} and B^{1D} estimations, we select chunks with \overline{SNR} greater than 2 and exclude chunks with fewer than 50 pixels, a different approach from [20]. We are now ready to calculate the quasar continuum for each individual chunk.

¹Formally, sky removal implies a convolution of the masks with the field in Fourier space. However, we assume it acts locally, so we only remove the sky lines before extracting the delta field and applying the Fourier transform. Sky lines are emission lines produced by various processes in Earth’s atmosphere, which emit light at specific wavelengths. We use the list of sky lines from <https://github.com/igmhub/picca/tree/master/etc>.

²We define \overline{SNR} as $\langle f(\lambda)/\sigma_p(\lambda) \rangle$, where $f(\lambda)$ is the observed flux and $\sigma_p(\lambda)$ is the estimate of the standard deviation of the flux.

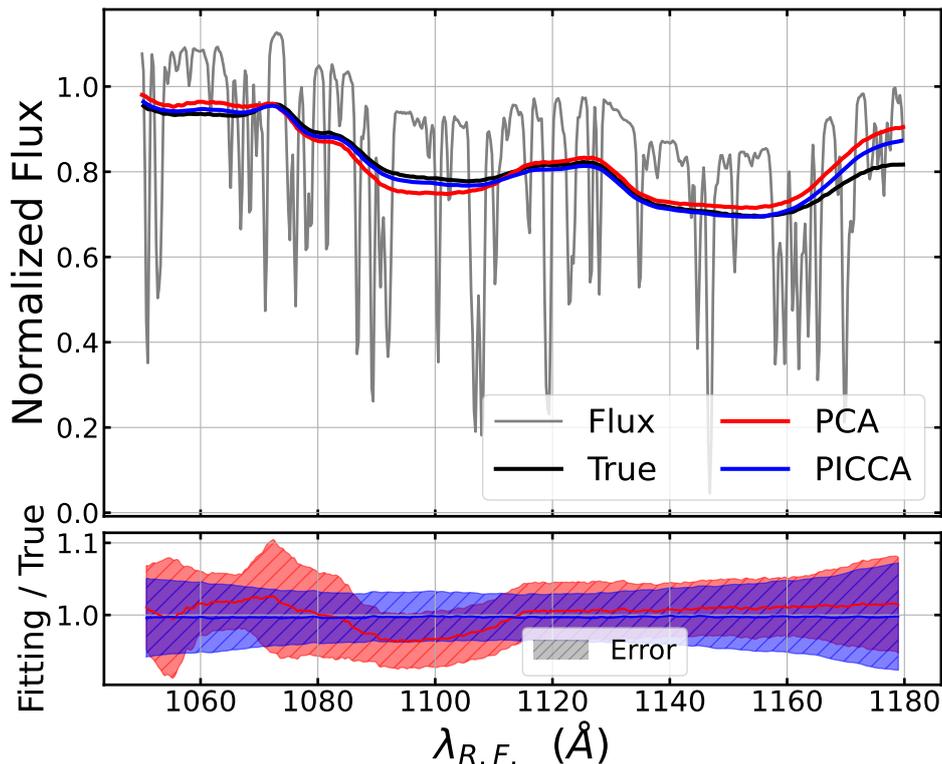


Figure 2. Example of a synthetic quasar spectrum at $z = 2.64$ (gray line) in the rest-frame. The color lines cover our defined forest region, $\lambda_{R.F.} \in [1050, 1180] \text{ \AA}$, which lies between the quasar’s Ly- β (1025.72 \AA) and Ly- α (1215.67 \AA) emission lines. On the top panel, the blue, red and black lines represent, respectively, the quantity $C(\lambda)\overline{F}(\lambda)$ measured using PICCA, PCA and the true continuum in the mocks. The bottom panel shows the average ratio of the continuum fitting and the true continuum (solid lines), as well as the dispersion from the average (shaded bands).

2.2.1 Continuum fitting

The observable quantity in the Ly- α forest is the transmission, $F = e^{-\tau}$, where τ is the optical depth. By definition, it is also expressed as the transmitted flux fraction $F(\lambda) = f(\lambda)/C(\lambda)$, where $f(\lambda)$ is the observed flux and $C(\lambda)$ is the unabsorbed quasar flux. One can define the transmission fluctuation, or “delta” field, $\delta_F(\lambda)$, by

$$\delta_F(\lambda) = \frac{F(\lambda) - \overline{F}(\lambda)}{\overline{F}(\lambda)} = \frac{f(\lambda)}{C(\lambda)\overline{F}(\lambda)} - 1, \quad (2.1)$$

where $\overline{F}(\lambda) = \langle f(\lambda)/C(\lambda) \rangle$ is the transmission ensemble average and $\langle \cdot \rangle$ denotes the ensemble average.

Different strategies can be used to compute the δ_F field. For example, one may choose to estimate the product $C(\lambda)\overline{F}(\lambda)$, as done with the public code PICCA [🔗](https://github.com/gmh/picca)³, or independently estimate the $C(\lambda)$ and $\overline{F}(\lambda)$ functions, as is done when fitting the continuum using Principal Component Analysis (PCA) [59]. In this work, we test both methods for measuring the

³The Package for IGM Cosmological-Correlations Analyses (PICCA) is available at <https://github.com/gmh/picca>.

$\delta_F(\lambda)$ field using mock spectra and compare them with the true continuum. For the PCA method, we employ the *empca* code [60]. Interestingly, we find that with this PCA code, we are not directly measuring $C(\lambda)$, but rather the product $C(\lambda)\overline{F}(\lambda)$, similar to the PICCA method. We observe that calculating $\langle f(\lambda_{\text{obs}})/PCA(\lambda_{\text{obs}}) \rangle_{\lambda_{\text{obs}}}$, with $PCA(\lambda_{\text{obs}})$ the PCA model, closely approximates unity for each wavelength in the forest. This indicates that $\langle f(\lambda_{\text{obs}})/PCA(\lambda_{\text{obs}}) - 1 \rangle_{\lambda_{\text{obs}}} \sim \langle \delta_F(\lambda_{\text{obs}}) \rangle$, confirming that the PCA method, as we use it in this analysis, truly measures $C(\lambda)\overline{F}(\lambda)$. To capture as much variance as possible with the PCA, we normalize each quasar spectrum so that its integral over the forest is equal to a constant, where the value of the constant is chosen to have an integrate flux of one. This step is repeated independently for each line of sight.

Figure 2 compares the $C(\lambda)\overline{F}(\lambda)$ quantity obtained using both the PICCA and PCA methods for a simulated quasar, compared to its true continuum. In both cases, the errors are below 10%, with smaller discrepancies in the central regions of the spectra and increasing towards the edges, particularly near the Ly- α and Ly- β peaks. Generally, the method employing PICCA, as described in detail in [21, 61, 62], appears to provide greater homogeneity over the forest region compared to PCA. This stability is evident as PICCA does not exhibit the long-mode features seen in the PCA results, especially below $\sim 1,120 \text{ \AA}$. These features in the PCA method arise from the selection of eigenvectors used for spectrum reconstruction, where higher principal components correct the mean spectrum of each quasar. As a result, using more principal components introduces greater features in the continuum relative to the true continuum. The emergence of features before $\lambda = 1,120 \text{ \AA}$ suggests that, on average, absorption lines around $\lambda = 1,070 \text{ \AA}$ are shallower, while those around $\lambda = 1,100 \text{ \AA}$ are deeper compared to absorption lines at $\lambda \geq 1,120 \text{ \AA}$, where the depth remains generally constant. Regardless of the continuum fitting choice, it introduces a bias to both the P^{1D} and B^{1D} signals. Continuum fitting generally suppresses the P^{1D} signal relative to the true continuum, particularly on large scales. With PICCA, this suppression is approximately 3%, while with PCA, it is about 5%. We further discuss these biases in the continuum fitting methods when addressing systematics in the bispectrum. The suppression of the B^{1D} signal varies depending on the triangle configuration used, but generally exceeds that of the P^{1D} signal, as discussed in section 4. Despite PICCA demonstrating a smaller deviation from the true continuum, we opt to use the PCA method for continuum extraction, as it provides a clearer redshift evolution of the B^{1D} signal compared to PICCA. In fact, we observe little redshift evolution when using PICCA. With the continuum fitting method established, we can proceed to calculate the delta field, accounting for the contributions of various absorbers involved in the Ly- α forest.

2.2.2 FFT 1D power spectrum

As will become evident in the next section, where we introduce an estimator for the 1D bispectrum of the “delta field”, $\delta_F(\lambda)$ from eq. (2.1), it is also necessary to estimate the 1D power spectrum. Therefore, we begin by deriving the P^{1D} . As discussed in [18, 20, 21], the δ_F field in the Ly- α region can be decomposed into several contributions: one component from the Ly- α peak absorption and another originating from metal absorption. Metal contributions in the Ly- α forest analysis can be further classified into two types: those redward (i.e., at larger

wavelengths) of the Ly- α peak, known as side-bands, and those within the Ly- α region itself. The absorptions redward of the Ly- α peak are independent of the Ly- α peak absorption and are known as side-bands. We define side band one (SB1) in the range $1270 \text{ \AA} < \lambda_{\text{R.F.}} < 1380 \text{ \AA}$ and side band two (SB2) for $1410 \text{ \AA} < \lambda_{\text{R.F.}} < 1520 \text{ \AA}$. Metal absorptions within the Ly- α region are primarily due to the Si_{II} ($\lambda_{\text{SiII}} = 1,190 \text{ \AA}$) and Si_{III} ($\lambda_{\text{SiIII}} = 1,206.50 \text{ \AA}$) emission lines. These metal features are correlated with Ly- α absorption lines and generate oscillations in both the power spectrum and the bispectrum. In summary, the delta field in the Ly- α region can be expressed as a sum of the following contributions:

$$\delta_{\text{F}}(\lambda) = \left(\delta^{\text{Ly-}\alpha}(\lambda) + \delta^{\text{SiII/SiIII}}(\lambda) + \delta^{\text{SB1/SB2}}(\lambda) \right) W(\lambda, R, \Delta\lambda) + \delta^n(\lambda), \quad (2.2)$$

where δ^n represents noise fluctuations, and the function W accounts for the spectrograph resolution, which we will characterize later. The 1D Ly- α power spectrum can be estimated from this decomposition by applying a fast Fourier transform (FFT) to δ_{F} for each chunk separately. The delta pixels are assumed to be equally spaced to allow the use of a simple FFT algorithm. The eBOSS quasar catalogue provides a constant pixel width of $\Delta[\log(\lambda)] = 10^{-4}$, corresponding to $\Delta v(\lambda) = c \ln(10) \Delta[\log(\lambda)] = 69 \text{ km/s}$ in velocity units. As a result, the Fourier space wave-vector, $k = 2\pi/\Delta v$, is measured in $(\text{km/s})^{-1}$.

We now define an estimator for the $P^{1\text{D}}$. In the absence of instrumental effects, the $P^{1\text{D}}$ can be written as the product of two Fourier-transformed $\delta_{\text{F}}(\lambda)$, namely $(2\pi)\delta_{\text{D}}(k - k')P^{\text{raw}} = \langle \delta_{\text{F}}(k)\delta_{\text{F}}(k') \rangle$, where $\delta_{\text{D}}(k)$ is the 1D Dirac delta function. However, when all contributions to the delta field from eq. (2.2) are included, the raw power spectrum becomes

$$P^{\text{raw}}(k) = \left(P^{\text{Ly-}\alpha}(k) + P^{\text{Ly-}\alpha\text{-SiII/SiIII}}(k) + P^{\text{SB1/SB2}}(k) \right) W^2(k, R, \Delta v) + P^n(k). \quad (2.3)$$

In this decomposition, $P^{\text{Ly-}\alpha}$, $P^{\text{Ly-}\alpha\text{-SiII/SiIII}}$, $P^{\text{SB1/SB2}}$, and P^n represent the power spectra corresponding to the components $\delta^{\text{Ly-}\alpha}$, $\delta^{\text{SiII/SiIII}}$, $\delta^{\text{SB1/SB2}}$, and δ^n , respectively, as defined in eq. (2.2). The estimation of the power spectrum, accounting for the different absorbers in the Ly- α forest, was first studied in [18]. The window function W , which characterizes the spectrograph's spectral response and pixel width, is defined in [19] as

$$W(k, R, \Delta v) = \exp\left(-\frac{(kR)^2}{2}\right) \times \frac{\sin(k\Delta v/2)}{k\Delta v/2}, \quad (2.4)$$

where R is the spectrograph resolution measure in $(\text{km/s})^{-1}$. This estimator for the $P^{1\text{D}}$ is our starting point to construct the 1D bispectrum.

2.3 Ly- α bispectrum formalism and estimator

Before constructing an estimator for the Ly- α $B^{1\text{D}}$, let us first review some general properties of three-point statistics in cosmology. The 3D matter three-point correlation function (3PCF),

$$\xi^{(3)}(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) \equiv \langle \delta(\mathbf{x}_0)\delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle,$$

measures the correlation between overdensity triplets in a sample. It depends on 9 degrees of freedom (dofs), which naturally define the vertex coordinates of a triangle in real space. Statistical homogeneity (or equivalently, translation invariance, which allows us to choose one

vertex as the origin) reduces this dependence to 6, which can be interpreted as the two vector sides of the triangle ($\mathbf{r}_1 = \mathbf{x}_1 - \mathbf{x}_0$, $\mathbf{r}_2 = \mathbf{x}_2 - \mathbf{x}_0$). Statistical isotropy further reduces the number of variables to 3, which we choose to be the lengths of the two sides of the triangle ($|\mathbf{r}_1|$ and $|\mathbf{r}_2|$) and their corresponding opening angle ($\mu = \mathbf{r}_1 \cdot \mathbf{r}_2$).

The bispectrum, $B(\mathbf{k}_0, \mathbf{k}_1, \mathbf{k}_2)$, is simply the Fourier transform of the 3PCF. In terms of an ensemble average over the δ fields, it reads

$$\begin{aligned} \langle \delta(\mathbf{k}_0)\delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle &= \int d^3\mathbf{x}_0 d^3\mathbf{x}_1 d^3\mathbf{x}_2 e^{i(\mathbf{x}_0 \cdot \mathbf{k}_0 + \mathbf{x}_1 \cdot \mathbf{k}_1 + \mathbf{x}_2 \cdot \mathbf{k}_2)} \langle \delta(\mathbf{x}_0)\delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle \\ &= \int d^3\mathbf{x}_0 e^{i\mathbf{x}_0 \cdot (\mathbf{k}_0 + \mathbf{k}_1 + \mathbf{k}_2)} \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 e^{i(\mathbf{r}_1 \cdot \mathbf{k}_1 + \mathbf{r}_2 \cdot \mathbf{k}_2)} \xi^{(3)}(\mathbf{r}_1, \mathbf{r}_2) \\ &= (2\pi)^3 \delta^D(\mathbf{k}_0 + \mathbf{k}_1 + \mathbf{k}_2) B(\mathbf{k}_0, \mathbf{k}_1, \mathbf{k}_2), \end{aligned} \quad (2.5)$$

where we have used the homogeneity of the 3PCF. In the last equality, we applied the definition of the Dirac delta function, δ^D , whose consequence is that the \mathbf{k} -vectors must form a triangle. Therefore, the bispectrum also has 6 dof when statistical homogeneity is assumed, and it is further reduced to 3 variables when isotropy is taken into account. As in real space, the resulting isotropic bispectrum can be expressed in terms of the two side lengths of triangles in \mathbf{k} -space ($k_1 = |\mathbf{k}_1|$, $k_2 = |\mathbf{k}_2|$) and their corresponding opening angle ($\alpha = \mathbf{k}_1 \cdot \mathbf{k}_2$). From symmetry arguments, one may expect the signal in Fourier or configuration space to peak for isosceles triangles in random data, with three notable configurations: equilateral, squeezed (when the opening angle approaches zero), and spread (when one side equals the sum of the other two). These configurations are also prominent in Ly- α analysis, as we will demonstrate later.

When restricted to a one-dimensional distribution (e.g., the Ly- α forest along the line of sight), the triangle configurations in both configuration and Fourier space collapse into a single line. In other words, the cosine of the opening angle (α) reduces to ± 1 , leaving only two independent variables. Formally, we split the \mathbf{k} space into parallel ($k_{\parallel} \equiv q$) and perpendicular (\mathbf{k}_{\perp}) directions. Using the 3D-to-1D Fourier space relation (see, e.g., [63, 64])

$$f_q^{(1d)} = \int \frac{d\mathbf{k}_{\perp}^2}{(2\pi)^2} f_{k_{\parallel}=q, \mathbf{k}_{\perp}}^{(3d)},$$

the 1D Bispectrum from equation (2.5) reduces to

$$\langle \delta(q_0)\delta(q_1)\delta(q_2) \rangle = (2\pi) \delta^D(q_0 + q_1 + q_2) B(q_0, q_1, q_2), \quad (2.6)$$

where Dirac's delta function imposes the linear restriction $q_2 = -q_0 - q_1$.

2.3.1 FFT bispectrum estimator for Ly- α

As we proceeded with the P^{1D} , the 1D bispectrum can be estimated by applying the FFT algorithm on δ_F from eq. (2.2) for each chunk, and by estimating the ensemble average of the third moment of $\delta(\Delta v)$ values. In the absence of instrumental effects (such as noise and spectrograph resolution), the B^{1D} can be simply written as the real part of eq. (2.6), namely:

$$B^{\text{raw}}(q_0, q_1, q_2) = \langle \text{Re}[\mathcal{F}(\delta(\Delta v_0)) \cdot \mathcal{F}(\delta(\Delta v_1)) \cdot \mathcal{F}(\delta(\Delta v_2))] \rangle,$$

where $\mathcal{F}(\delta(\Delta v))$ represents the Fourier transform of $\delta_{\text{F}}(\Delta v)$, binned into pixels of width Δv over the Ly- α forest region. As a reminder, momentum conservation, or equivalently the presence of Dirac's delta function, enforces the closure of the momentum variables over the line of sight, restricting the 1D bispectrum to depend on two variables. We select q_0 and q_1 as these variables. As a result, the noiseless raw bispectrum is given by:

$$B^{\text{raw}}(q_0, q_1) = \langle \text{Re}[\delta_{\text{F}}(q_0)\delta_{\text{F}}(q_1)\delta_{\text{F}}(-q_0 - q_1)] \rangle.$$

Notice that it is not necessary to compute the bispectrum over the entire two-dimensional plane because there are several symmetries in the signal. Since $\delta(\Delta v)$ is real, it follows that $\delta(-q) = \delta^*(q)$. Consequently, the bispectrum exhibits the following symmetries

$$\begin{aligned} B(q_0, q_1) &= B(q_1, q_0) = B^*(-q_0, -q_1), \\ B(q_0, q_1) &= B^*(-q_1, q_0 + q_1) = B^*(-q_0, q_0 + q_1), \\ B(q_0, q_1) &= B(-q_0 - q_1, q_1) = B(q_0, -q_0 - q_1). \end{aligned} \quad (2.7)$$

Thus, we only need to compute $B^{\text{raw}}(q_0, q_1)$ for $q_0, q_1 \geq 0$ and can use the first of the above symmetries to construct the full q -dependence shown in figure 1.

When considering the other effects in the delta field that lead to the decomposition in eq. (2.2), the associated third moment in Fourier space simplifies to

$$\begin{aligned} B^{\text{raw}}(q_0, q_1) &= \left(B^{\text{Ly-}\alpha}(q_0, q_1) + B^{\text{Ly-}\alpha\text{-Si}_{\text{II}}/\text{Si}_{\text{III}}}(q_0, q_1) + B^{\text{SB1/SB2}}(q_0, q_1) \right) \\ &\quad \cdot W(q_0, R, \Delta v)W(q_1, R, \Delta v)W(q_0 + q_1, R, \Delta v) \\ &\quad + P^n(q) [P^{\text{raw}}(q_0) + P^{\text{raw}}(q_1) + P^{\text{raw}}(q_0 + q_1)] - 2P_n^2(q). \end{aligned} \quad (2.8)$$

In this decomposition, $B^{\text{Ly-}\alpha}$ and $B^{\text{SB1/SB2}}$ are the auto-bispectra associated with the transmission fluctuation fields $\delta^{\text{Ly-}\alpha}$ and $\delta^{\text{SB1/SB2}}$, respectively, with the same definition as the real part of eq. (2.6) for the corresponding delta field. The transmission fluctuation field attributed to noise fluctuations, δ^n , introduces two contributions to the bispectrum signal, as previously discussed in the literature (see, for example, [65, 66]). The first contribution arises from the cross-term between the noise power spectrum and the raw power spectrum, while the second term comes from the quadratic power of the noise power spectrum in eq. (2.8). Moreover, the auto and cross-correlations between Si_{II} and Si_{III} are subdominant and therefore neglected. All cross terms between two delta fields of $\text{Si}_{\text{II}}/\text{Si}_{\text{III}}$ and one delta field of Ly- α , as well as cross terms with noise fluctuations and side-band metals, are also neglected. The only non-negligible cross terms are

$$\begin{aligned} B^{\text{Ly-}\alpha\text{-Si}_{\text{II}}/\text{Si}_{\text{III}}}(q_0, q_1) &\approx 2\text{Re}[\delta^{\text{Ly-}\alpha}(q_0)\delta^{\text{Si}_{\text{II}}/\text{Si}_{\text{III}}}(q_1)\delta^{\text{Ly-}\alpha}(-q_0 - q_1)] \\ &\quad + \text{Re}[\delta^{\text{Ly-}\alpha}(q_0)\delta^{\text{Ly-}\alpha}(q_1)\delta^{\text{Si}_{\text{II}}/\text{Si}_{\text{III}}}(-q_0 - q_1)], \end{aligned}$$

which corresponds to the correlated absorptions of the Ly- α with either Si_{II} or Si_{III} .

The final FFT estimator for the 1D bispectrum is computed as an average over all available Ly- α forests within a given chunk. Since, at present, we are unable to independently extract the $B^{\text{Ly-}\alpha}(q_0, q_1)$ and $B^{\text{Ly-}\alpha\text{-Si}_{\text{II}}/\text{Si}_{\text{III}}}(q_0, q_1)$ components of eq. (2.8), we combine them

by defining $B^{1D}(q_0, q_1) = \langle B^{\text{Ly-}\alpha}(q_0, q_1) + B^{\text{Ly-}\alpha-\text{SiII/SiIII}}(q_0, q_1) \rangle$. Thus, from eq. (2.8), we conclude that the estimator for B^{1D} is

$$B^{1D}(q_0, q_1) = \left\langle \frac{B_{q_0, q_1}^{\text{raw}} - P^n(q) [P^r(q_0) + P^r(q_1) + P^r(q_0 + q_1)] + 2P_n^2(q)}{W(q_0, R, \Delta v)W(q_1, R, \Delta v)W(q_0 + q_1, R, \Delta v)} \right\rangle - B^{\text{SB1/SB2}}(q_0, q_1), \quad (2.9)$$

where we use the shorter notation $P^r \equiv P^{\text{raw}}$. The eq. (2.9) is our final estimator to measure the 1D bispectrum.

The bispectrum, B^{1D} , is divided into different redshift bins to account for its evolution, as discussed in section 6. The negative sign in the bispectrum is consistent with the findings of [49], and we demonstrate that this negative sign persists for other combinations of q_0 and q_1 . Furthermore, Zaldarriaga et al. [47] suggested that the negative sign might result from higher-order correlations arising due to gravitational growth.

In this work, we focus on two specific triangle configurations: the isosceles and the special case of the squeezed limit. As mentioned by [65] and [67], the triangle configurations with the most prominent signals are those that share a common q -vector ($q_1 = n \cdot q_0$), where n is an integer. In appendix A, we will verify that these two configurations yield the strongest signals. Isosceles triangles are characterized by $q_1 = q_0$, and we have adopted the definition of the squeezed limit from [51].

- *Configuration 1 (isosceles)*: $q_1 = q_0$

$$B(q) = \text{Re}[\langle \delta(q)\delta(q)\delta^*(2q) \rangle]$$

- *Configuration 2 (squeezed)*: $q_0 = q - q_{\text{min}}$ and $q_1 = -q - q_{\text{min}}$

$$B(q) = \text{Re}[\langle \delta(q - q_{\text{min}})\delta^*(q + q_{\text{min}})\delta(2q_{\text{min}}) \rangle]. \quad (2.10)$$

Formally the squeezed limit is when $q_{\text{min}} \rightarrow 0$, but in practice one cannot take that limit given the forest finite range. Therefore, one may use the first bin of the delta field, in order to avoid nonphysical signals, we take q_{min} as the second bin. In the following, we present the flux bispectrum as a function of the wavenumber $q = q_0$. Hereafter, we will refer to the bispectrum of isosceles triangles as B^{1D} , while for the squeezed limit bispectrum, we will use B_{SQ}^{1D} . In the case of the Ly- α forest, configurations close to the squeezed limit may offer an interesting approach for studying primordial non-Gaussianities (e.g., [51, 53, 55]) or consistency conditions when compared with the power spectrum (see, for example, [68–70]).

2.4 Theoretical description of B^{1D}

For the modelling, we use the theoretical expressions of [49], which relate the flux power spectrum and bispectrum using second-order perturbation theory (2OPT). In the Fluctuating Gunn-Peterson Approximation (FGPA), the flux can be related to the density field as

$$F = \exp\{-A(1 + \delta_{\text{IGM}})^\beta\}, \quad (2.11)$$

where A and β depend on redshift. $\beta = 2 - 0.7(\gamma - 1)$, and γ is the power-law index of the gas temperature-density relation, whose values are reported in [71, 72]. This model

assumes that redshift space distortions are not important and neglects thermal broadening and instrumental noise. As a future line of work, we can overcome this assumption using the ideas developed in [73–75]. Furthermore, the model assumes that at the scale of interest $\delta_{\text{IGM}} = \delta_{\text{DM}} = \delta$, together with a Taylor expansion of the density field to second order, namely $\delta(x) = \delta^{(1)}(x) + \delta^{(2)}(x)$. Using FGPA such that $(1 + \delta)^\beta = 1 + \beta\delta + \beta(\beta - 1)/2 \delta^2$, we finally get the following expression for the delta field

$$\delta_{\text{F}} \approx b_1[\delta^{(1)}(x) + \delta^{(2)}(x)] + \frac{b_2}{2} \left(\delta^{(1)}(x)\right)^2, \quad (2.12)$$

with $b_1 = -A\beta$ and $b_2 = -A\beta(\beta - 1 - A\beta)$. The expression for the Ly- α bispectrum follows then by projecting the 3D bispectrum ([65]) along the line-of-sight, namely

$$\begin{aligned} B(q_0, q_1, q_2) = & \left(\frac{12}{7}c_1 + c_2\right) p(q_0)p(q_1) + c_1 \left[\left(q_0q_1 - \frac{2}{7}q_0^2\right) p^{(-1)}(q_0)p(q_1) \right. \\ & \left. + \left(q_1q_0 - \frac{2}{7}q_1^2\right) p^{(-1)}(q_1)p(q_0) + \frac{6}{7}q_0^2q_1^2 p^{(-1)}(q_0)p^{(-1)}(q_1) \right] + \text{cyc.}(0, 1, 2), \end{aligned} \quad (2.13)$$

where $c_1 = 1/b_1$, $c_2 = b_2/b_1^2$ and $p(q)$ is the 1D-Power Spectrum. The spectral moments $p^{(-1)}(q)$ are given by

$$p^{(\ell)}(q) = |q|^{2\ell} p(q) + 2\ell \int_{|q|}^{\infty} k^{-2\ell-1} p(k) dk, \quad (2.14)$$

in this case $\ell = -1$.

We develop eq. (2.13) for both *Configuration 1* and *Configuration 2*. It is essential to consider the sign in the above equations for all q values and account for the symmetry between $p(q_i)$ and $p_i^{(-1)}(q)$. [49] demonstrated that the optimal values that match with synthetic spectra are $A = 0.48$ and $\gamma = 1.5$. In our results, we fit the values of c_1 and c_2 for the full 2D B^{1D} signal, considering all triangle configurations.

3 Data analysis and P^{1D} assessment

The instrumental effects in eBOSS have been extensively investigated in previous studies related to the measurement of P^{1D} (see, for example, [19–21]). However, the impact of these effects on B^{1D} estimation remains unexplored. For this reason, we begin by analyzing these effects in the context of our study. Specifically, we focus on the impact of spectral resolution, instrumental noise, and the metal (side-bands) bispectrum. In addition, we provide a brief overview of our final sample selection and present the P^{1D} estimate within the Ly- α forest, comparing it with previous measurements to assess the effectiveness of our pipeline.

3.1 Summary of the Ly- α sample

As mentioned in section 2.1, our Ly- α forest sample consists of 166,229 quasars. After removing BALs, missing DLAs, or bad spectra using our DLA finder, Ly- α forests with less than 50 pixels, and applying the \overline{SNR} cut, the sample is reduced to 122,066 quasars. Since we work with chunks (correspond to consecutive, non-overlapping sub-forests of different lengths) to measure the B^{1D} , we improve the quality of the low/high-redshift chunks by considering

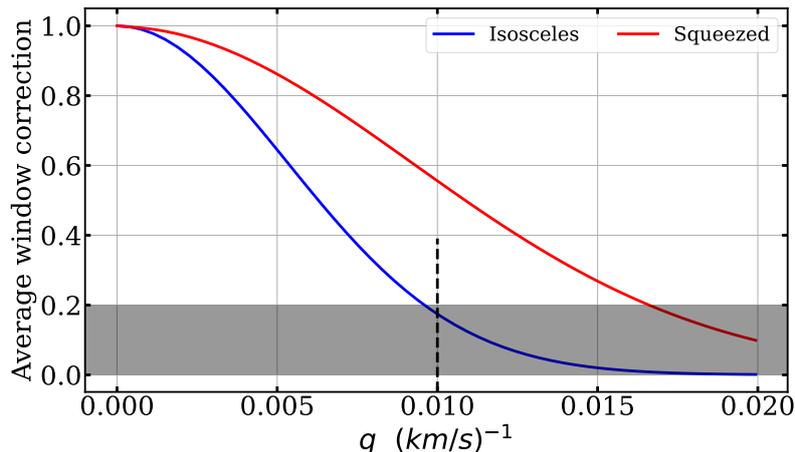


Figure 3. Window function for B^{1D} measurements in the isosceles or squeezed limit configurations. We restrict to a maximal suppression of 80%, which results in a maximal scale of $q \approx 0.01 \text{ (km/s)}^{-1}$ ($q \approx 0.017 \text{ (km/s)}^{-1}$) for the isosceles (squeezed) triangles.

CCD-observed wavelengths in the range $3770 \text{ \AA} \leq \lambda_{\text{obs}} \leq 6687 \text{ \AA}$. After removing chunks with $\overline{SNR} < 2$ and discarding those where the mean spectral resolution exceeds 85 km/s, our final Ly- α forest chunk sample contains 151,621 chunks. For SB1, we have 256,751 chunks.

3.2 Spectrograph resolution

The window function correction, as expressed in eq. (2.4), encompasses two effects on the measurement of P^{1D} and B^{1D} : one due to pixelization (spectrum binning or pixel width) and the other due to spectrograph resolution. While this correction has been thoroughly investigated for SDSS P^{1D} measurements [18–20], its application to B^{1D} measurements remains unexplored. Eq. (2.4) shows that the influence of the spectrograph resolution R is modeled by a Gaussian function. We determine R following the same analysis performed by Chabanier et al. [20] (see also [19] for a detailed description), which accounts for corrections to R as a function of CCD position, fiber number, and wavelength. We then apply this correction to each pixel in every spectrum. Figure 3 shows the average window correction for the two configurations defined in (2.10). These corrections reveal that both resolution and pixelization suppress the isosceles bispectrum signal by just over 80% at $q = 0.01 \text{ (km/s)}^{-1}$. For the squeezed limit bispectrum, the suppression exceeds 90%, similar to that of P^{1D} . The maximum value of q varies depending on the triangle configurations and the Nyquist-Shannon limit: $k_{\text{Nyquist}} = \pi/\Delta v = 0.045 \text{ (km/s)}^{-1}$. However, we choose $k_{\text{max}} = 0.02 \text{ (km/s)}^{-1}$ for P^{1D} and B_{SQ}^{1D} ($q_{\text{max}} = 0.017 \text{ (km/s)}^{-1}$), as the impact of the spectrograph resolution becomes very strong beyond this value. For *Configuration 1*, we set $q_{\text{max}} = 0.01 \text{ (km/s)}^{-1}$ due to the spectrograph’s resolution (see A).

3.3 Noise bispectrum measurement

3.3.1 Estimator of noise power spectrum

One of the systematics that significantly impacts the P^{1D} measurement signal at small scales is the noise power spectrum P^n . Therefore, it is essential to accurately estimate it. Over

the years, two methods have been employed to calculate the P^n : the first involves directly estimating it from the pipeline’s noise, while the second utilizes an exposure difference method. For our bispectrum estimation, we opt for the exposure difference method, which is extensively applied in previous work [19–21]. For each redshift bin, we first fit P^{raw} on q -modes above $q = 0.02 \text{ (km/s)}^{-1}$ with an exponential decay plus a constant $P_{\text{lim}}^{\text{raw}}$. We then compute the ratio β between P_n^{pipeline} and $P_{\text{lim}}^{\text{raw}}$. Our pipeline noise is then defined by $P^n = P_n^{\text{pipeline}}$ when $\beta < 1$, and $P^n = P_n^{\text{pipeline}}/\beta$ otherwise.

3.3.2 Characterization of noise in the bispectrum signal

To demonstrate that the raw bispectrum signal originates from a non-trivial source, we compare it with two distinct, trivial signals. The first comparison is with noise fluctuations as defined by eq. (2.2), with its power spectrum computed using the exposure difference method. This time, our aim is to determine the bispectrum associated with the noise fluctuations, again using the exposure difference method. Specifically, we seek to measure $B^{\text{diff}}(q_0, q_1, q_2) = \text{Re}[\mathcal{F}(n(\Delta v_0)) \cdot \mathcal{F}(n(\Delta v_1)) \cdot \mathcal{F}(n(\Delta v_2))] + \dots$, assuming a Gaussian random field with zero mean, though its dispersion may explain some of the power in the raw data signal. However, as seen in figure 4, this is not the case. Here, $\mathcal{F}(n(\Delta v_i))$ represents the Fourier transform of the normalized exposure difference spectrum, whose power spectrum is expected to be an accurate estimator of P^n in the coadded spectrum. Furthermore, P^n is found to be scale-independent, as expected for white noise. In the case of white noise, its bispectrum is expected to be zero, as shown by the red line in figure 4. The second comparison is with a bispectrum signal obtained by shuffling the delta field at the pixel level ($B_{\text{forest}}^{\text{shuffled}}$). We shuffle each pixel of the delta field randomly, similar to what was done in [49], with the main difference being that they shuffled absorption features instead of pixels, thus preserving the smallest structures. Since we aim to distinguish the raw signal from this other delta field configuration on large scales, we perform the shuffling at the pixel level. Again, the raw signal is not consistent with this shuffled configuration on scales below $q \lesssim 0.01$, as shown in figure 4. To identify the source of bias in the $B_{\text{forest}}^{\text{shuffled}}$ signal, we computed the same for the side-bands. In this case, we found no bias, as the $B_{\text{forest,SB1}}^{\text{shuffled}}$ signal was consistent with zero. Thus, we conclude that the bias arises from the background power associated with the side-bands.

One might wonder which term in eq. (2.9) dominates the signal (excluding B^{metals}). Figure 5 illustrates the contributions from the terms in eq. (2.9). The largest contribution to $B^{1\text{D}}$ comes from the term B^{raw} , while the second contribution comes from the term $P^n \cdot \sum_{i=0}^2 P_{\text{raw}}(q_i)$, as the value of P^n is high, amplifying the sum of the $P^{\text{raw}}(q_i)$ signal. This second term enhances the B^{raw} signal, while the third term represents an asymptote to the combined signal of the first two terms. In the case of mocks, the P^n is even smaller than in eBOSS, so the $B^{1\text{D}}$ is dominated entirely by the B^{raw} signal. We plot the average of B^{raw} , $P^n \cdot \sum_{i=0}^2 P_{\text{raw}}(q_i)$, and $2P_n^2$, computed over the Ly- α forest. As expected, the noise is white, and the term $2P_n^2$ is scale-independent. The term $P^n \cdot \sum_{i=0}^2 P_{\text{raw}}(q_i)$ is important, as it amplifies the $B^{1\text{D}}$ signal, while in the mocks, it is irrelevant given the low noise level. For comparison, figure 5 also shows the raw bispectrum. We select $q_{\text{max}} = 0.01 \text{ (km/s)}^{-1}$ because at higher modes, the effect of the spectrograph resolution becomes significant, suppressing almost 80% of the signal, although this percentage depends on redshift, with a larger effect at lower redshifts.

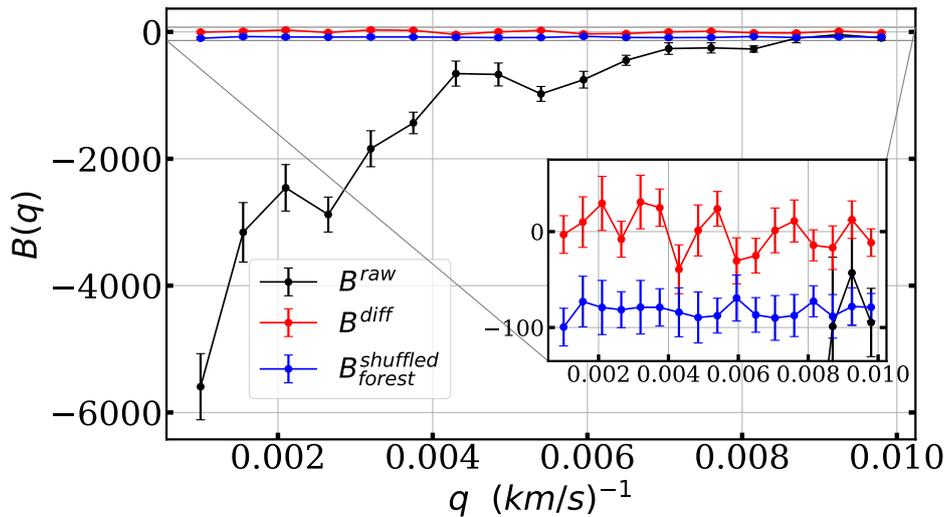


Figure 4. Comparison between the flux bispectrum of observed fluxes (circles blacks), “randomized” absorption lines in the forest spectra (circles blues) and the bispectrum from the exposure difference noise (circles reds). The plot represent *Configuration 1*. The bias present in the randomized bispectrum is given by the background power of the metals in the side-bands. The error bars come from the square root of the diagonal of the bootstrap covariance matrix.

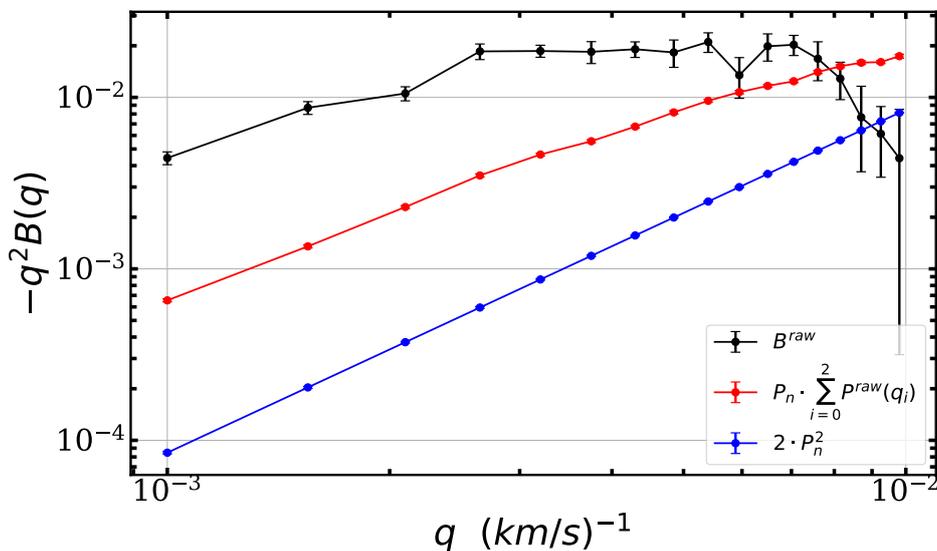


Figure 5. Average bispectrum of the raw signal (black dots), the noise power spectrum times the power spectrum for different q -modes (red dots) and twice the square of the noise power spectrum signal (blue dots). The blue line is a asymptote of the raw bispectrum signal. The red line contribute significantly of the raw bispectrum signal at small scales. The error bars come from the square root of the diagonal of the bootstrap covariance matrix.

3.4 Side-band bispectrum

In section 2.2.2, we discussed the absorption lines redward of the Ly- α emission line caused by metals. These absorptions produce a background power in the 1D Ly- α power spectrum. A similar effect occurs in the case of the Ly- α bispectrum. We assume that the cross-correlation between Ly- α absorption and metals in the side-bands is negligible. In both side-bands that we consider, the delta field ($\delta_F(\lambda)|_{\text{SB}}$) can be expressed as a combination of metal absorption lines and noise fluctuations. Consequently, the calculation of the metal side-bands bispectrum can be written as:

$$B_{\text{SB}}^{1\text{D}}(q_0, q_1) = \left\langle \frac{B_{q_0, q_1}^{\text{raw}}|_{\text{SB}} - P^n(q)|_{\text{SB}} \cdot \left[\sum_{i=0}^2 P^{\text{raw}}(q_i)|_{\text{SB}} \right] + 2P_n^2(q)|_{\text{SB}}}{W(q_0, R, \Delta v)W(q_1, R, \Delta v)W(q_0 + q_1, R, \Delta v)} \right\rangle. \quad (3.1)$$

This uncorrelated background cannot be directly estimated through bispectrum measurements in the Ly- α wavelength region. To address this issue, we measure the bispectra in the side-bands and subtract them from the Ly- α bispectrum, measured over the same redshift range. This approach has been extensively examined for the power spectrum analysis [19–21]. Notice that the method is purely statistical, using a distinct sample of quasars to compute both the Ly- α and metal bispectra within a specific redshift range. In the case of what we call side band one (SB1), the bispectrum includes contributions from all metals with wavelengths greater than 1380 Å, with significant contributions from Si IV and C IV absorption lines. Conversely, our side band two (SB2) contains only C IV absorptions. Therefore, we subtract only the SB1 signal from the Ly- α bispectrum, while using SB2 primarily for consistency checks. We estimate the metal bispectrum over the same observed wavelength range as the Ly- α bispectrum, meaning we use a quasar sample with a lower redshift range than the sample used to estimate the Ly- α bispectrum. For example, in the first redshift bin, $2.1 < z_{\text{Ly-}\alpha} < 2.3$, we measure the bispectrum in SB1 (corresponding to $3770 \text{ \AA} < \lambda_{\text{obs}} < 4012 \text{ \AA}$) using quasars with $z_{\text{qso}} \sim 2.0$, while for the Ly- α bispectrum we use quasars with $z_{\text{qso}} \sim 2.2$. The measurement for isosceles triangles is shown in figure 6. A total of 256,751 chunks were analyzed, undergoing the same cuts as applied to the quasar spectra in the Ly- α forest analysis. The top panel shows the $B_{\text{SB1}}^{1\text{D}}$ stacking across all redshift bins (black dots with error bars), while the red dots represent the bispectrum signal in SB2, from a sample of 237,985 chunks. As expected, the amplitude of the $B_{\text{SB2}}^{1\text{D}}$ signal is smaller than that of $B_{\text{SB1}}^{1\text{D}}$, since SB2 excludes the Si IV absorptions. The bottom panel shows $B_{\text{SB1}}^{1\text{D}}$ for different redshift bins. Figure 7 illustrates the results for the squeezed limit bispectrum. Similar to the isosceles triangles, the signal amplitude of $B_{\text{SQ}}^{1\text{D}}|_{\text{SB2}}$ is smaller than that of $B_{\text{SB1}}^{1\text{D}}$. This configuration clearly displays the oscillations caused by the C IV doublet absorptions on the Si IV forest. Once again, the black dots with error bars in the upper panel represent the stack of all redshift bins of $B_{\text{SB1}}^{1\text{D}}$, while the red dots show the $B_{\text{SQ}}^{1\text{D}}|_{\text{SB2}}$ signal. The bottom panel illustrates $B_{\text{SB1}}^{1\text{D}}$ for different redshift bins.

The solid lines represent the fit to the bispectrum shape of the side-bands. The fitting procedure we employed is inspired by the recent work on DESI [21], and carefully explained in [76]. However, it was originally intended for measuring the side-bands power spectrum, and we have adapted this approach to analyze the side-bands bispectrum. Similar to this previous work on DESI, we account for the absorptions of the Si IV and C IV doublets. According

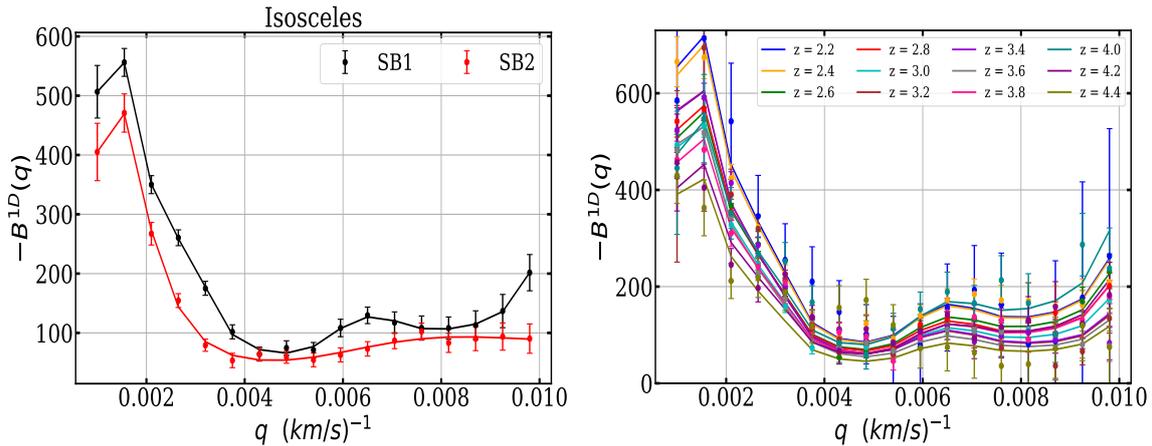


Figure 6. 1D bispectrum measured in the side-bands regions SB1 and SB2 for isosceles triangles. *Top:* average of the B_{SB}^{1D} over all redshift bins. The solid lines represent the fitted model given by eq. (3.2). *Bottom:* B_{SB1}^{1D} as a function of redshift. Each redshift bin is fitted using the product of the fitting in the top panel and a first-degree polynomial in which the two parameters are free. The error bars come from the square root of the diagonal of the bootstrap covariance matrix.

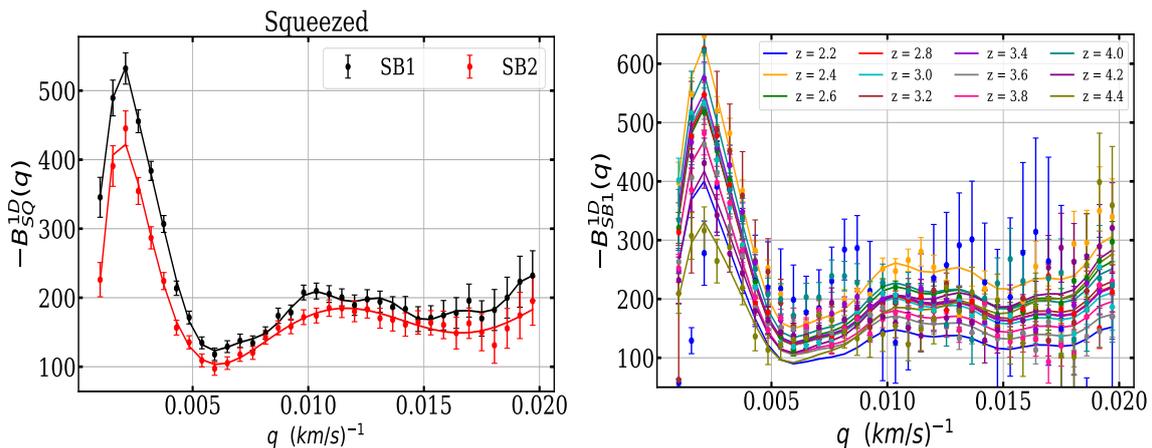


Figure 7. 1D bispectrum measured in the side-bands regions SB1 and SB2 for squeezed limit triangles. *Top:* average of the B_{SQ}^{1D} over all redshift bins. The solid lines represent the fitted model given by eq. (3.2). *Bottom:* B_{SB1}^{1D} as a function of redshift. Each redshift bin is fitted using the product of the fitting in the top panel and a first-degree polynomial in which the two parameters are free. The error bars come from the square root of the diagonal of the bootstrap covariance matrix.

to NIST data ([77]), the Si IV doublets are located at $\lambda_{R.F.}^{Si\ IV_a} = 1,393.76 \text{ \AA}$ and $\lambda_{R.F.}^{Si\ IV_b} = 1,402.77 \text{ \AA}$, while the C IV doublets are at $\lambda_{R.F.}^{C\ IV_a} = 1,548.202 \text{ \AA}$ and $\lambda_{R.F.}^{C\ IV_b} = 1,550.774 \text{ \AA}$. The C IV doublet has a separation of $r = \Delta v_{C\ IV_{a,b}} = 499 \text{ km/s}$, while the Si IV doublet has a separation of $s = \Delta v_{Si\ IV_{a,b}} = 1933 \text{ km/s}$. The presence of an absorption doublet in the side-bands induces oscillations in the bispectrum signal. The periodicity of these oscillations depends on the separation of the doublets. When two absorption doublets are present in the same side-band, the periodicity of the oscillations also depends on both the sum and difference of the separations of the doublets. In the top panels of figures 6 and 7, both

side-bands show significant oscillations caused by the C IV doublet, along with large-scale oscillations due to the difference in the doublet spacing. As expected, the SB1 signal shows additional oscillations due to the Si IV absorptions. The model for $B_{\text{SQ}}^{\text{1D}}$ consists of a sum of a power law and oscillating functions:

$$B_{\text{SB1,model}}^{\text{1D}} = A \cdot \left(\frac{q}{q_2}\right)^{-\epsilon} + \sum_i B_{\text{doublet},i}(q, q_i, A_i, a_i, \psi_i). \quad (3.2)$$

In the isosceles configuration, where there is a $2q$ dependence, the oscillations induced by a doublet have a frequency characterized by the wavenumber $2q = 2\pi/\Delta v_i$. For example, in the top panel of figure 6, the first bump of $B_{\text{SB1}}^{\text{1D}}$ corresponds to $2q = 2\pi/s = 0.0016 \text{ (km/s)}^{-1}$, and the second bump to $2q = 2\pi/r = 0.0062 \text{ (km/s)}^{-1}$. In the case of the squeezed limit configuration, while there is also a $2q$ dependence, it does not affect the signal because q is a minimum value, behaving like a constant. In this case, the frequency of the oscillations is given by $q = 2\pi/\Delta v_i$. In this configuration, the oscillation due to the difference in the velocities of the doublets becomes more evident, as seen in the first bump of the $B_{\text{SQ}}^{\text{1D}}|_{\text{SB2}}$ signal at $2q = \frac{2\pi}{s-r} = 0.0021 \text{ (km/s)}^{-1}$. We model the doublet oscillations using damped sinusoidal functions as follows:

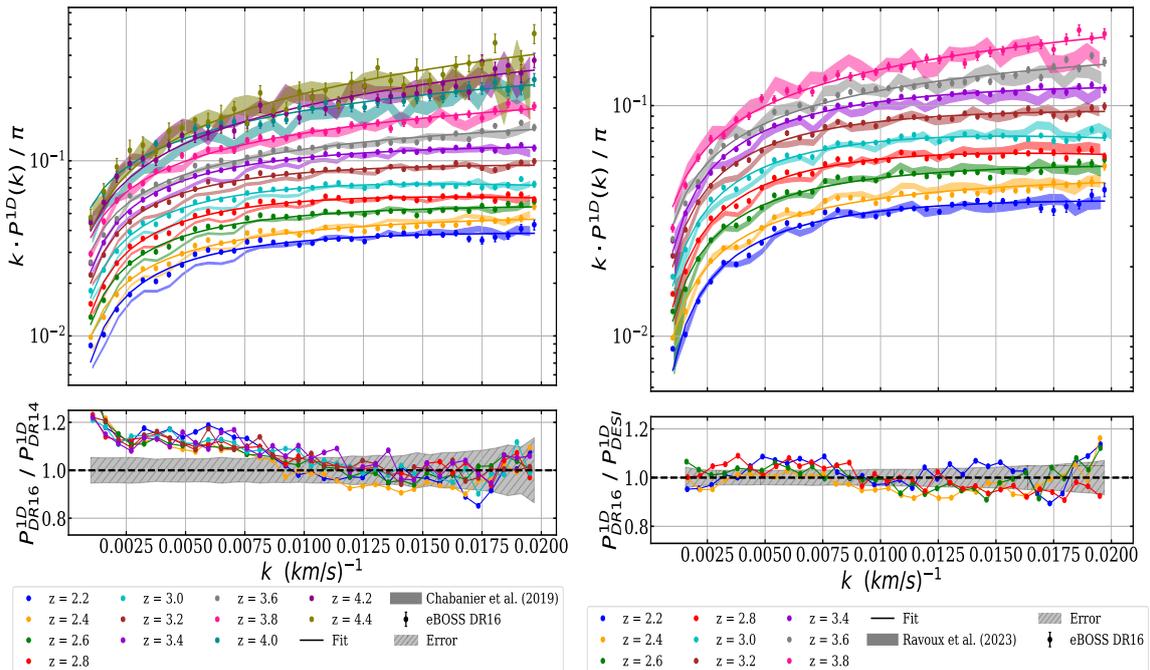
$$B_{\text{doublet},i}(q, q_i, A_i, a_i, \psi_i) = \sum_i A_i e^{-a_i \left(\frac{q}{q_i}\right)} \sin\left(2\pi \left(\frac{q}{q_i}\right) + \psi_i\right). \quad (3.3)$$

In the case of isosceles triangles, $q_i = \left(\frac{\pi}{s}, \frac{\pi}{r}, \frac{\pi}{s+r}, \frac{\pi}{s-r}\right)$. For squeezed limit triangles, $q_i = \left(\frac{2\pi}{s}, \frac{2\pi}{r}, \frac{\pi}{s+r}, \frac{\pi}{s-r}\right)$. The terms $(s+r)$ and $(s-r)$ are derived from a simple model that approximates the transmitted flux contrast (δ_{F}) in the presence of two doublets with velocity separations s and r , i.e., $\delta_{\text{F}} \sim \delta(v) + a\delta(v+s) + b\delta(v+r)$, where the factors a and b indicate the relative strengths of the absorptions of the doublets compared to the background absorption. When calculating the bispectrum of this transmitted flux contrast, oscillating functions arise that vary as $(s+r)$ and $(s-r)$.

The SB1 fitted function is then used to derive the redshift dependence of the side-band bispectrum, shown in the bottom panels of figures 6 and 7. For each redshift bin, we fit the product of the global SB1 fitted function and a first-order polynomial. This fit should not be used for rigorous scientific conclusions, as the χ^2_{ν} values indicate that the model based on the SB1 fitting function does not fully represent the observed data. However, it is sufficient for providing a baseline estimate. We performed the fit independently for each configuration. The fitted parameters are generally consistent between the two configurations, except for the parameters A and A_i , which primarily affect the amplitude of the doublet oscillations.

3.5 1D power spectrum assessment

As shown in eq. (2.9), calculating the bispectrum requires estimating the power spectrum, which includes both the noise power spectrum and the side-band power spectrum (SB1 only), as indicated in eq. (2.3). The power spectrum allows us to assess the quality of our data analysis and the efficiency of our pipeline for calculating the bispectrum (B^{1D}). This is done by comparing our P^{1D} estimation with previous measurements. While numerous P^{1D} measurements have been made in the Ly- α forest, some of which were mentioned in



1D flux power spectrum in eBOSS DR16 data for twelve redshift bins.

Figure 8. 1D flux power spectrum in eBOSS DR16 data for twelve redshift bins. *Left:* comparison between the measurement performed in this work and that of the eBOSS DR14 data [20]. Our measurement with points and error bars, the DR14 measurement with shaded colored areas. The ratio between DR16 and DR14 measurements is showed in the bottom panel. The striped grey area in the bottom panel represent the centered error bar (added in quadrature) of the ratio averaged over all shown redshift bins. We found the same discrepancies reported in [21] at large scales with respect to DR14 data, but we are unable to determinate the source of this difference. *Right:* same comparison, but this time between the measurement performed in this work and that of the DESI data [21]. Both signals show a good agreement for all redshift bins considered. Solid lines in both panels represent the fitted model given by eq. (3.5). The error bars present in our measurements come from the square root of the diagonal of the bootstrap covariance matrix.

section 1, we opted to compare our results with previous measurements from the eBOSS DR14 data ([20]) and the most recent measurements from the DESI early data release ([21]), which draws from large quasar surveys. The FFT estimator for the 1D power spectrum is calculated as an average over all available Ly- α forests in the measurement sample. From eq. (2.3), the P^{1D} estimator is defined as:

$$P^{1D}(k) = \left\langle \frac{P^{\text{raw}}(k) - P^n(k)}{W^2(k, R, \Delta v)} \right\rangle - P^{\text{metals}}(k). \quad (3.4)$$

Figure 8 (left) shows the comparison between our P^{1D} estimate (represented by dots with error bars) and the eBOSS DR14 data (shaded lines) for different redshift bins. The percentage difference between them is approximately 15% at large scales ($k < 0.01$ km/s). A similar discrepancy was reported in [21], though their comparison was between DESI and eBOSS DR14 data. Like them, we are unable to determine the source of this discrepancy. In our case, the continuum fitting provided by PCA also affects the P^{1D} signal at large scales,

but its effect is insufficient to account for the magnitude of this difference. However, at small scales, both measurements are consistent within the error bars, suggesting that we have accurately estimated the noise power spectrum and corrected for the spectrograph resolution present in the window function, as these are the main factors influencing small scales.

We also compare our measurement with the DESI early data release in figure 8 (right). Overall, both measurements are in good agreement across the k range when considering the error bars, with a percentage difference of around 5%. The largest difference is observed around $k \sim 0.013 \text{ km/s}$, likely due to the oscillation of $P_{\text{SB1}}^{1\text{D}}$ in that region compared to the eBOSS data.

We made a fit to the $P^{1\text{D}}$ using the fiducial power spectrum function proposed in [78] given by:

$$\frac{kP(k, z)}{\pi} = A \frac{(k/k_0)^{3+n+\alpha \ln k/k_0}}{1 + (k/k_1)^2} \left(\frac{1+z}{1+z_0} \right)^{B+\beta \ln k/k_0}, \quad (3.5)$$

where $k_0 = 0.009 \text{ (km/s)}^{-1}$ and $z_0 = 3.0$. The continuous thin colored lines in both plots of figure 8 show the fitting.

4 Synthetic data correction

In this section, we examine the biases introduced at each step of the data analysis and assess their impact using simulated spectra. Our goal is to characterize their effect on the $B^{1\text{D}}$, considering continuum fitting, masking of pixels affected by sky lines, and absorption from DLAs as well as noise and spectrograph resolution.

4.1 Mocks

CoLoRe [79], which stands for Cosmological Lofty Realisation, is a package designed to produce various cosmological tracers of the same underlying matter fluctuations, such as spectroscopic galaxies, weak lensing, and CMB lensing. Additionally, it can generate Ly- α absorption features in the spectra of high-redshift quasars, which is of particular interest for this work. Although CoLoRe can produce skewers in transmitted flux fraction, its “raw” output requires significant post-processing before it can be considered a realistic representation of the Ly- α forest. For this purpose, we use LyaCoLoRe  [80], which is capable of generating realistic skewers of transmitted flux fraction. CoLoRe generates the density field from an initial power spectrum using two approaches. The first is based on an initial Gaussian field, where a lognormal approximation is employed to model the physical density, which can help understand the linear evolution of the field. The second approach uses a formalism called Lagrangian Perturbation Theory (LPT) [81], which extends to mildly non-linear scales. In CoLoRe, both first- and second-order LPT (2LPT) methods are available. We estimated the bispectrum using both approaches, and our results indicate that the signals agree within the error bars, with deviations of less than 10% at large scales in both configurations. One would naively expect the bispectrum to be sensitive to the different prescriptions for the non-linear evolution of matter density perturbations in the mocks. However, the fact that

⁴Public software available at <https://github.com/igmhub/LyaCoLoRe>.

both approaches result in a similar bispectrum suggests that signal is dominated by the mapping from density field to the skewers' flux. In conclusion, we choose the simplest and least computationally expensive method, the lognormal approach, to discuss results based on these mocks.

The output skewers from LyaCoLoRe require the addition of instrumental noise and a combination with a QSO continuum before the spectra can be considered realistic. This can be carried out using the *desisim* package [🔗](#).⁵ *Desisim* produces QSO continuum using the *SIMQSO* package [🔗](#),⁶ where the continuum is generated by applying a broken power law and adding emission lines, modeled by Gaussian distributions. In addition, small-scale power can be added to account for clustering on scales up to 100 kpc/h [80], which is smaller than the minimal scale in the bispectrum analysis presented here. Absorption lines in the Ly- α quasar spectra are added by *quickquasars* [🔗](#)⁷ (a DESI code within *desisim*), by multiplying the continuum templates with the transmitted flux of the raw mocks. Furthermore, *quickquasars* can introduce noise to the spectra, metal absorption lines, DLAs, BALs, and tune the pixel width of the wavelength in the spectra. For more details, see [82].

To understand the impact of systematic errors, such as sky lines, DLA contamination, and the continuum fitting procedure, we generated mock noiseless spectra that replicate the numerical density of Ly- α forest chunks per redshift bin. DLA contamination is introduced using a random catalogue. We created five independent realizations with different initial conditions. For each realization, we generated two sets of spectra: one that includes DLA information and the other that does not.

4.2 Continuum-fitting correction

As mentioned in 2.2, we decided to estimate the continuum fitting using PCA. Since we only perform PCA on the blue side of the Ly- α forest, what we are actually estimating is $C(\lambda)\bar{F}(\lambda)$. PCA produces a series of eigenvectors ordered by the fraction of the sample variance they explain. In principle, each eigenvector is a linear combination of the input fluxes. The quasar continuum is produced through the following equation:

$$C\bar{F}_{i,m}(\lambda_{\text{R.F.}}) = \bar{f}(\lambda_{\text{R.F.}}) + \sum_{j=1}^m \alpha_{ij} \xi_j(\lambda_{\text{R.F.}}). \quad (4.1)$$

where $\bar{f}(\lambda_{\text{R.F.}})$ is the mean observed flux, and $\xi_i(\lambda_{\text{R.F.}})$ represent the eigenvectors. We construct the continuum using three principal components.⁸ The continuum-fitting procedure systematically distorts the measured $C(\lambda)\bar{F}(\lambda)$ by suppressing large-scale modes in the B^{1D} measurement. We observed that suppression of the B^{1D} signal at small wavenumbers increases as the number of eigenvectors increases. We tested up to 5 eigenvectors and found that the signal suppression exceeded 25%. Thus, we decided to use only 3 eigenvectors. This choice

⁵Public software available at <https://github.com/desihub/desisim>.

⁶Public software available at <https://github.com/desihub/simqso>.

⁷Public software available at <https://github.com/desihub/desisim/blob/main/py/desisim/scripts/quickquasars.py>.

⁸We initially decompose into 10 eigenvectors but only use the first 3 modes, which in total capture around 30% of the variance. Notice that this methodology is different from constructing the PCA with 3 eigenvectors from the begging, which did not capture well the Ly- α peak.

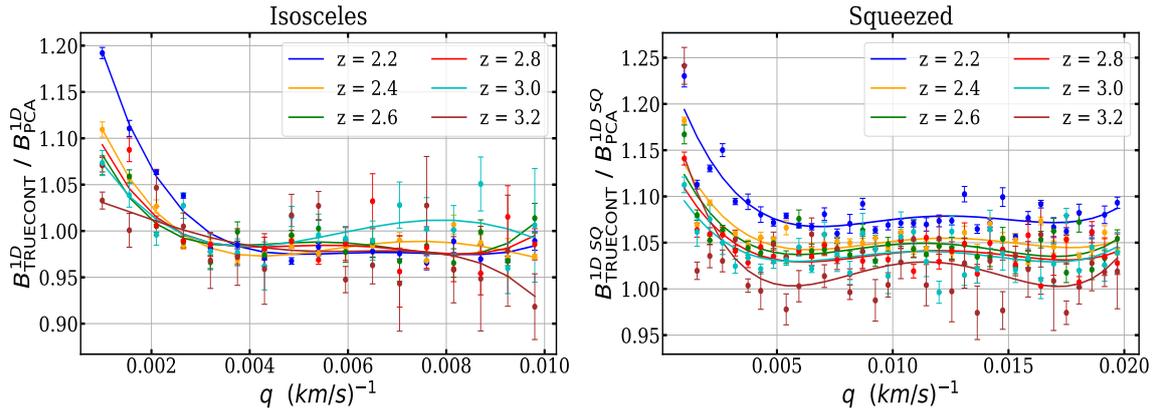


Figure 9. Ratios between the bispectrum obtained using true continuum (TRUECONT) and the one derived with our PCA continuum fitting (PCA) on the combination of 5 mocks, for both triangular configurations considered. Fitting functions are represented by continuous lines and used to correct the bispectrum measurement. The correction is fitted with a four-degree polynomial. The error bars come from the square root of the diagonal of the bootstrap covariance matrix between the five realizations.

reduces the impact of the continuum fitting and ensures that we can adequately capture the diversity of the quasar sample (de la Cruz et al., in preparation).

We define the bias induced by the continuum-fitting as the ratio of the bispectrum computed with the true continuum (TRUECONT) to the bispectrum computed using our standard PCA-based continuum-fitting procedure:

$$b_{\text{cont}}(q, z) = \frac{B_{\text{TRUECONT}}^{1\text{D}}(q, z)}{B_{\text{PCA}}^{1\text{D}}(q, z)}. \quad (4.2)$$

Figure 9 illustrates the measured bias. The top panel shows the isosceles configurations, where suppression of the $B^{1\text{D}}$ signal is evident at small wavenumbers, with slightly excessive power across the remaining wavenumbers. The bottom panel represents the squeezed-limit configurations, where signal suppression is observed across all wavenumbers, with additional suppression particularly noticeable at small wavenumbers. The correction is fitted with a fourth-order polynomial dependence for $2.2 < z < 3.2$. Overall, the impact of the continuum-fitting become more significant in the estimation of the bispectrum compared to its effect on the power spectrum, where the error is approximately 6%.

4.3 Spectrum pixel masking correction

In our pipeline for estimating the $B^{1\text{D}}$ signal, we masked pixels affected by sky lines or DLAs present in the Ly- α forest. We set the affected pixels to a flux value of zero and maintained this value throughout the transmission delta field. This pixel masking introduces a q -dependent bias, which must be quantified.

We begin by estimating the correction due to the masking of the sky lines. We compare the $B^{1\text{D}}$ measured from mocks with sky-line masking (SKYm) to mocks without masking (PCA). The bias parameter used for the masking correction is defined as the ratio between

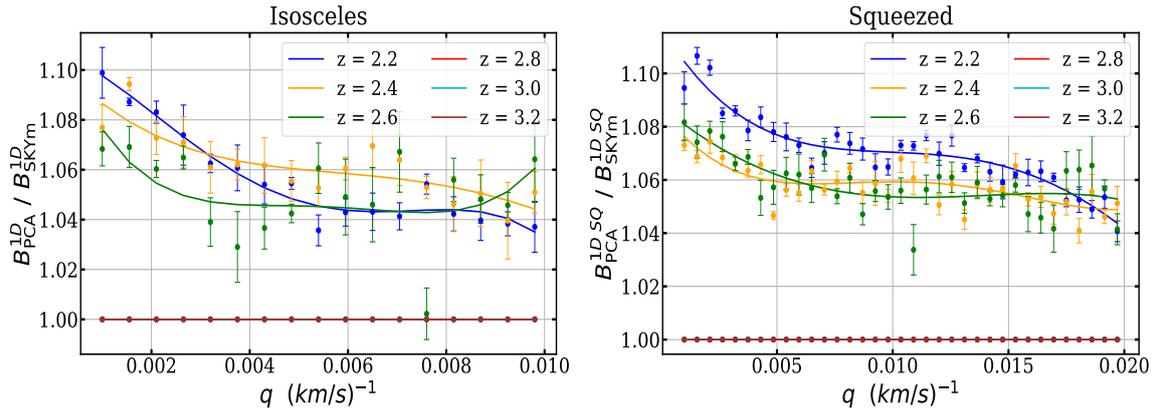


Figure 10. Ratio between the bispectrum obtained using unmasked sky lines pixels (PCA) and the one with masked pixels (SKYm) on the combination of the 5 mocks, for both triangular configurations considered. Four-order polynomial function are employed to fit the corrections in each redshift bin. The error bars come from the square root of the diagonal of the bootstrap covariance matrix between the five realizations.

the unmasked and masked bispectrum:

$$b_{\text{SKYm}}(q, z) = \frac{B_{\text{PCA}}^{1\text{D}}(q, z)}{B_{\text{SKYm}}^{1\text{D}}(q, z)}. \quad (4.3)$$

For the two chosen triangle configurations, the masking effect is most pronounced at large scales and decreases as it approaches smaller scales. Only the redshift bins at $z = 2.2$, 2.4, and 2.6 are affected by sky lines, while the redshift bins at $z = 2.8$, 3.0, and 3.2 remain unbiased. These unbiased redshifts are consistent with the findings of [20], although their analysis focused on the $P^{1\text{D}}$. To model this bias, we employ a fourth-order polynomial, which is used for the final $B^{1\text{D}}$ correction. Figure 10 shows the $b_{\text{SKYm}}(k, z)$ effect for both configurations. The left panel corresponds to isosceles triangles, while the right panel shows the squeezed-limit triangle configuration.

DLAs are added at random locations in the Ly- α forest during the creation of the mocks. For these initial results on the $B^{1\text{D}}$, we are not focused on characterizing the completeness of the DLA finder applied to the data. Instead, we use a ‘truth’ DLA catalogue for masking, making it sufficient to use a random distribution to understand this effect. We mask the ‘truth’ catalogue using the same parameters as those applied to the observed DLA data catalogue. We then compare the $B^{1\text{D}}$ measured from mocks with DLA masking (DLAm) to mocks without masking (PCA). The bias parameter used for the masking correction is defined as the ratio between the unmasked and masked bispectrum:

$$b_{\text{DLAm}}(q, z) = \frac{B_{\text{PCA}}^{1\text{D}}(q, z)}{B_{\text{DLAm}}^{1\text{D}}(q, z)}. \quad (4.4)$$

Figure 11 shows the $b_{\text{DLAm}}(k, z)$ effect for both configurations. The top panel corresponds to isosceles triangles, while the bottom panel represents the squeezed-limit triangle configuration. In both configurations, there is a noticeable effect at small wavenumbers, which increases with redshift. This is consistent with the expectation that DLA contamination in the Ly- α

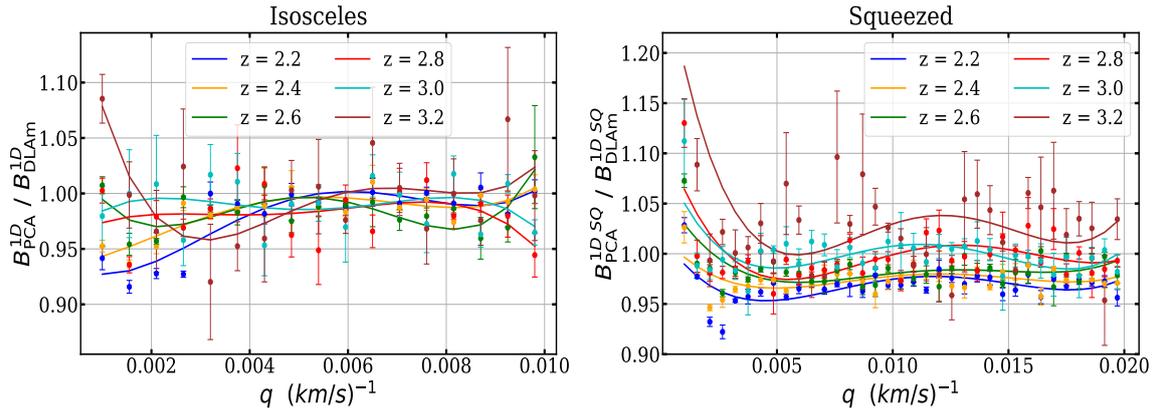


Figure 11. Ratio between the bispectrum obtained using unmasked DLAs pixels (PCA) and the one with masked pixels (DLAm) on the combination of the 5 mocks, for both triangular configurations considered. Four-order polynomial function are employed to fit the corrections in each redshift bin. The error bars come from the square root of the diagonal of the bootstrap covariance matrix between the five realizations.

forest increases with redshift. The impact of DLA masking is smaller compared to the effect of sky-line masking. These results are in good agreement with the P^{1D} measurements from eBOSS DR14 [20]. To model this bias, we employ a fourth-order polynomial, which is used for the final B^{1D} correction.

5 Systematic uncertainties estimation

We estimate the statistical uncertainty (σ_{stat}) of our averaged B^{1D} measurement using the diagonal of the bootstrap covariance matrix.⁹ For our dataset of N chunks, we form a bootstrap dataset by randomly selecting N chunks with replacement. The covariance matrix is then computed as follows:

$$C_{ij}^{B^{1D}} = \frac{1}{N_{\text{chunks}}} \sum_{k=1}^{N_{\text{chunks}}} \left[B_k^{1D}(q_i) - \langle B^{1D}(q_i) \rangle \right] \cdot \left[B_k^{1D}(q_j) - \langle B^{1D}(q_j) \rangle \right], \quad (5.1)$$

$\langle \cdot \rangle$ means average over bootstrap realizations. In eq. (5.1) the B^{1D} term can to be replaced by B_{SQ}^{1D} in the case of the squeezed limit. We generate 2000 bootstrap samples of the input dataset and calculate the corresponding bootstrap covariance matrix. Figure 12 shows the normalized bootstrap correlation matrix $C_{ij} / \sqrt{C_{ii}C_{jj}}$, where we observe zero correlation between adjacent wavenumbers. A slight correlation appears to grow from large to small scales, but this is only seen at small wavenumbers, likely due to the systematics affecting the initial q -modes. Convergence is achieved rapidly, and it may even be sufficient to use 200 bootstrap samples. Figure 13 illustrates the dependence of the error bars on the bispectrum power. The low redshift bins are not clearly separated, and at redshifts $z = 2.2$ and $z = 2.4$, statistical uncertainties intersect at small scales. This intersection is due to a significant

⁹We compute the bootstrap covariance matrix for both configurations independently, rather than across the full 2D signal of the bispectrum.

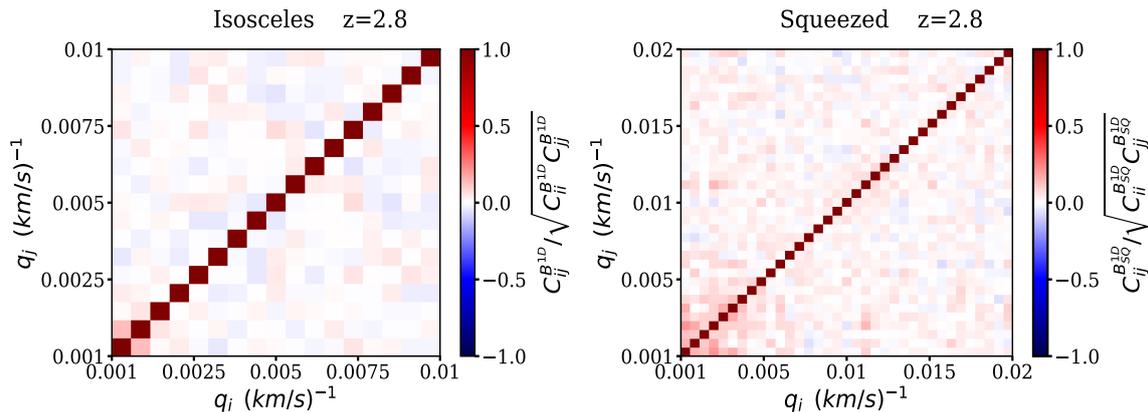


Figure 12. Bootstrap covariance matrices for both triangular configurations considered over eBOSS DR16 data set. We generate 2000 bootstrap samples. There is not correlation between adjacent q -modes, only slightly for large scales.

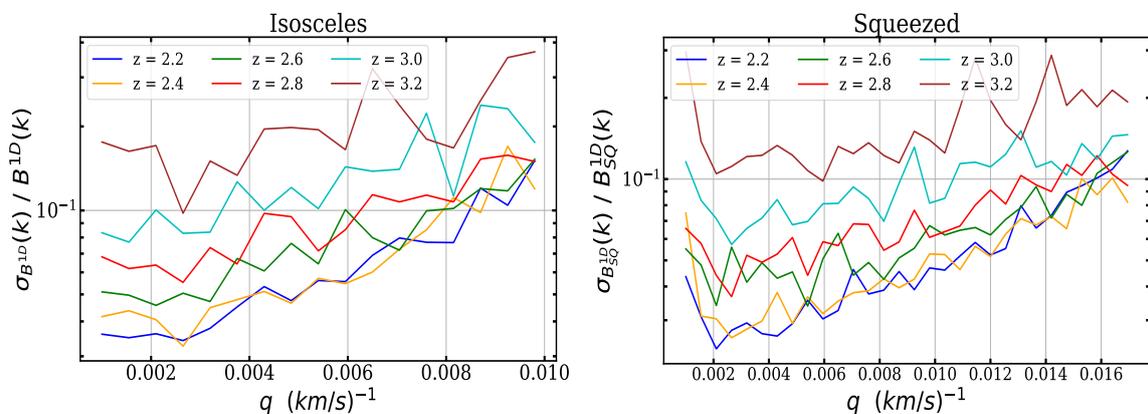


Figure 13. The estimation of the relative statistical uncertainties on the flux bispectrum measurement, for both triangular configurations considered.

increase in noise in the blue spectral band. The variation in the number of chunks present in each redshift bin leads $\sigma(q)/B^{1D}$ to increase as a function of redshift. Furthermore, the substantial change in amplitude at small scales compared to large scales is attributed to the effect of the window function, where the spectrograph’s resolution directly impacts the results.

As explained in the previous sections, starting from eq. (2.8), the measurement of B^{1D} requires characterizing the impact of various instrumental and astrophysical contaminants. Therefore, we need to associate systematic errors (σ_{sys}) with our B^{1D} estimate. The left panel of figure 14 shows the systematic uncertainties for isosceles triangles across different redshift bins and their relative values compared to statistical errors. The right panel of figure 14 presents the systematic uncertainties for squeezed limit triangles in different redshift bins and their relative values with respect to statistical errors. We identify seven systematic uncertainties and made a conservative choice in defining these uncertainties:

- *Continuum fitting*: we assign a systematic error of 30 per cent times the $b_{\text{cont}}(k, z)$ correction computed in 4.2.

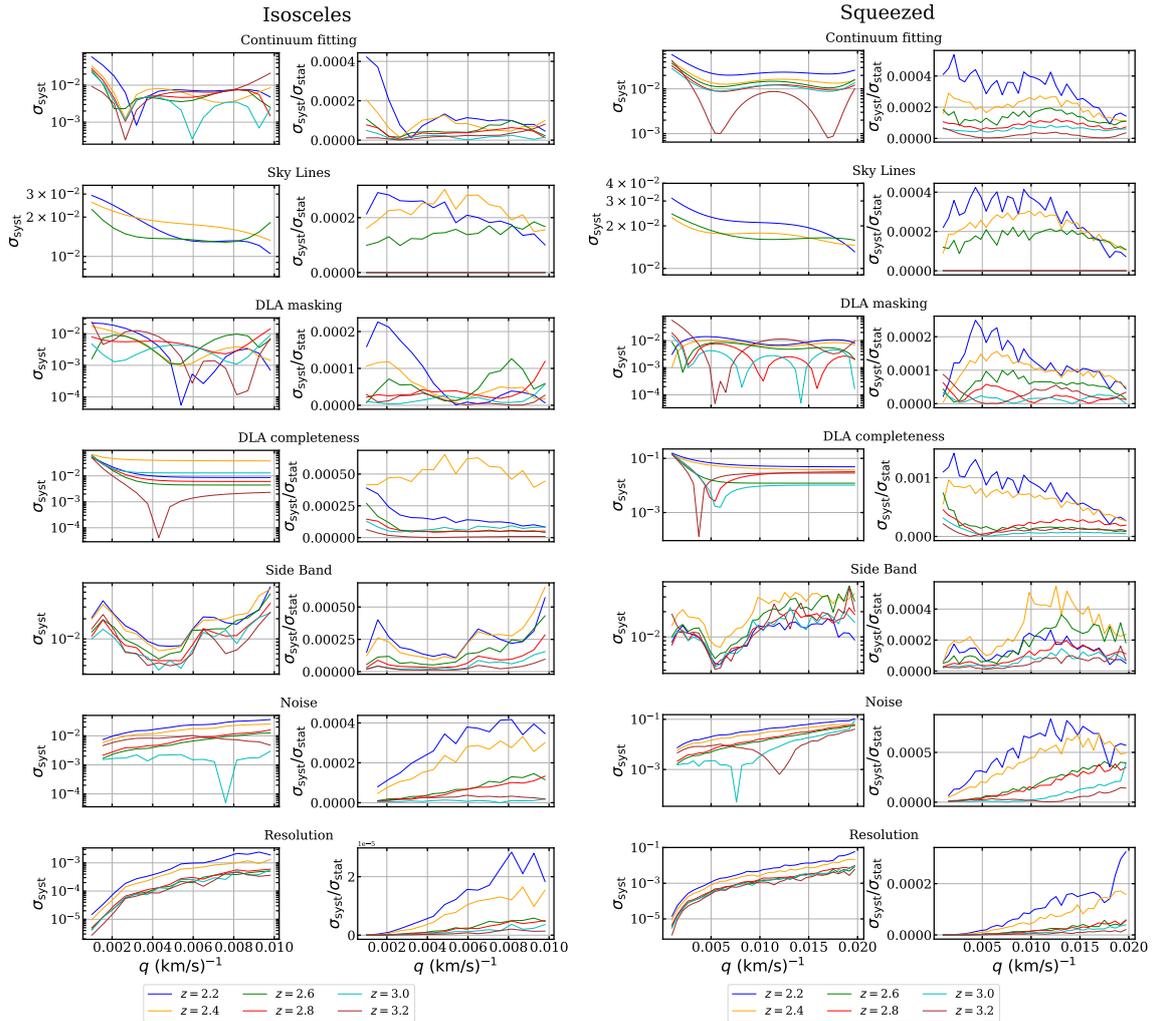


Figure 14. Estimation of the systematic uncertainties σ_{syst} for different redshift bins in the two triangular configurations considered. Each line sub-panel is associated to a systematic considered in this work. The left panels of each triangular configuration show the absolute uncertainties, while the right panels show the relative value between systematic uncertainties and statistical uncertainties σ_{stat} .

- *Sky lines masking*: the effect of the sky emission lines masking on the B^{1D} measurement was determined with synthetic data in 4.3. We define the systematic error associated with each masking as 30 per cent of this correction.
- *DLA masking*: the effect of the DLAs masking on the B^{1D} measurement was also determined using synthetic data in 4.3. We define the systematic error associated with each masking as 30 per cent of this correction.
- *DLA completeness*: we utilize the synthetic mock data described in section 4 to investigate the impact of DLAs on the 1D bispectrum. We calculate the ratio between mocks with DLAs and those without DLAs. This ratio shows a similar effect to the one observed in the 1D power spectrum signal, with small wavenumbers being the most affected. To address this, we employ a model presented by [83] to fit this ratio. Based

on the DLA catalogue provided by [58], the authors report over 90% efficiency using their CNN finder. However, as a conservative measure, we associate a 10% uncertainty with the total impact of DLAs on our B^{1D} signal.

- *Side-band*: we subtract the fitted side-bands bispectrum $B_{SB1,model}^{1D}$, modeled and computed in section 3.4, from the measured B^{1D} in the Ly- α forest to account for metal absorption. While one could use statistical uncertainty as the systematic uncertainty, the bootstrap error estimates might not be accurate, potentially biasing the result. Moreover, because the $B_{SB1,model}^{1D}$ signal depends on the number of side-band quasars in the relevant wavelength range, the statistical estimate of metal power might be incorrect. Projects like DESI will significantly improve the statistics in this area. Therefore, we calculate the ratio between the bispectrum before and after removing the side-band contribution. Allowing for the blending of metal lines as discussed in [30] and [27], we assign a 10 percent error.
- *Noise estimation*: as discussed throughout this text, the noise level in the bispectrum signal depends on both the raw power spectrum and the noise power spectrum, which amplifies the $P^{raw}(k)$ signal. A straightforward method to characterize the noise effect on the B^{1D} signal involves multiplying the noise power spectrum by the raw power spectrum. The procedure for determining the noise level in the quasar spectrum is described in section 3.3. The noise level is corrected using the β term, which varies for each redshift bin and dataset. We assign a systematic uncertainty to the resulting noise power spectrum, equal to 30% of the $(1-\beta)$ term times $q \cdot P(q)$. The maximum value occurs at $z = 2.2$, as this is closer to the edge of the CCD.
- *Resolution*: to investigate the impact of the resolution on the B^{1D} measurement, we conducted a study similar to the one proposed in [20], focusing on the bispectrum rather than the power spectrum. The window function W requires knowledge of the spectral resolution R of the spectrograph, see eq. (2.4). To incorporate its systematic uncertainty, we calculated the average resolution $\langle R \rangle$ across all the skewers contributing to each redshift bin. The systematic uncertainty on the B^{1D} signal is then given by $(2q^3 R \Delta R) \cdot B^{1D}$. The cubic q -term makes the large q -modes more affected by this uncertainty. The average resolution $\langle R \rangle$ also varies with respect to the redshift, from 79 km/s to 68 km/s with larger value for lower redshifts. The low- z bins are more affected by this systematic.

The choice of a 30 percent impact for systematic uncertainty comes from considering a random shift ranging between ‘100% of the correction’ and ‘no correction’, which we describe using a uniform distribution between 0 and 1. The standard deviation of this distribution, equal to $1/\sqrt{12} \sim 0.30$, quantifies the spread among the possible values, resulting in a systematic uncertainty equal to 30% of the correction.

We acknowledge that we have incorporated several ideas previously used to estimate systematic uncertainty in the 1D power spectrum signal. While we have adapted these ideas for the 1D bispectrum signal, we recognize there is still room for improvement in our systematic uncertainty estimation. Continuum fitting could be enhanced either by using the

method employed by `picca` or by developing a new approach that introduces less bias into the bispectrum signal. New galaxy surveys, such as DESI, will further reduce the effect of sky line masking by minimizing the number of masked pixels (see [21]). Similarly, improvements in the number of masked pixels for DLAs are expected. Significant progress is also anticipated in the study of quasar identification with DLAs, leading to improved DLA completeness. Additionally, the influence of other high-column density absorber systems, such as sub-DLAs, small-DLAs, and Lyman Limit systems, should be considered. These systems have been studied by [83] in the context of the power spectrum. We have shown that the modeling of high column density absorbers can also be applied to the bispectrum, suggesting that it should be effective for these other systems as well. As the statistics of quasars with side-bands increase, the associated systematic error can be better characterized. We expect significant improvements in noise estimation for the bispectrum signal, as new methods are developed to quantify its effect. Regarding resolution, DESI is expected to yield considerable improvements. Since σ_{stat} is quite large, the ratio between σ_{stat} and σ_{syst} is small for all systematics.

In future work, we plan to explore new ideas and methods to better quantify systematic effects. Specifically, we aim to investigate the impact of systematics using large samples of more realistic synthetic data, which will provide us with a deeper understanding of their effects on the bispectrum measurement.

6 Results: first measurement of the Ly- α B^{1D} in eBOSS

The B^{1D} measurement was performed using the pipeline described in section 2. Considering all the corrections outlined in previous sections, the final estimation of the 1D bispectrum incorporate the 2D signal from eqs. (4.2), (4.3) and (4.4) into the full bispectrum signal from eq. (2.9). Our results will be presented for the two triangular configurations defined in section 2. While this is primarily for visualization purposes, it is important to emphasize that the analysis was conducted using the 2D bispectrum signal, encompassing all triangular configurations.

6.1 B^{1D} measurement

Using the procedure described in the previous sections, we estimate the 1D bispectrum over 12 redshift bins, from $z_{\text{Ly-}\alpha} = 2.2$ to 4.4, and over 17 Fourier modes, ranging from $q = 1 \times 10^{-3}$ to 0.01 (km/s)^{-1} , for the isosceles triangle configuration. For the squeezed-limit triangle configuration, we calculate 30 modes, ranging from $q = 1 \times 10^{-3}$ to $0.017 \text{ (km/s)}^{-1}$, within the same redshift range. We chose these q -limits because the effect of the window function suppresses the bispectrum signal by 80%. We focus on the low redshift bins from $z_{\text{Ly-}\alpha} = 2.2$ to 3.2 for both configurations, as the higher redshift bins exhibit a noisier bispectrum signal. Figure 15 displays the 1D bispectrum measurement for the isosceles triangles at low redshift. The error bars shown correspond to the square root of the diagonal elements of the bootstrap covariance matrix, with systematic uncertainties added in quadrature. Our measurements appear to be consistent with the previous results of [49], which used a redshift bin from $z_{\text{Ly-}\alpha} = 2.0$ to 2.4, with a signal characterized by large error bars, as shown by the gray shaded line (see figure 15). The colored solid lines, corresponding to the different redshift bins, represent the fit of the theoretical 2OPT model to the observed data. The theory was fitted to the 2D bispectrum signal (see the left panel of figure 1), and the quality of the

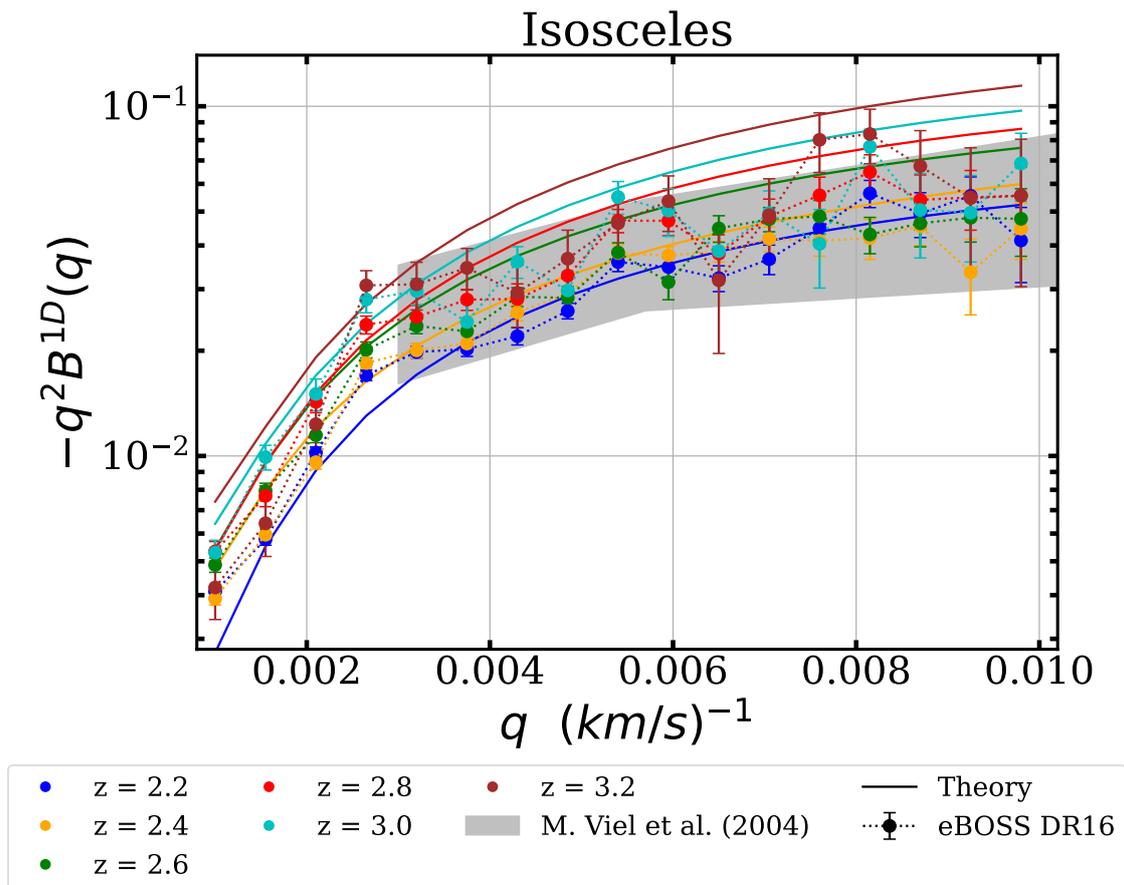


Figure 15. One-dimensional Ly- α bispectrum for isosceles triangles using eBOSS DR16 data, for redshift bins from $z = 2.2$ to $z = 3.2$ (point with error bars). Solid lines represent the fit modeled theoretically by eq. (2.14). The shaded gray area is the previous measurement by [49] in a redshift range of $2.0 < z < 2.4$. The error bars present in our measurements come from the square root of the diagonal of the bootstrap covariance matrix and systematic uncertainties added in quadrature.

z	2.2	2.4	2.6	2.8	3.0	3.2
Isosceles	8.58	13.28	20.33	10.56	10.16	9.19
Squeezed-Limit	165.68	1199.94	1133.73	851.91	843.19	782.06

Table 1. The partial contribution to χ^2 over the number of degrees of freedom for each redshift bin using the best fit model in 2D. As visually appreciated in figures 15 and 16, the isosceles configurations are closer to the simple analytical model than the squeezed ones.

fit was assessed by measuring the χ^2 value for both the isosceles triangle configuration and the squeezed-limit triangle configuration. The reduced χ^2 value for both configurations is presented in table 1. Given the number of degrees of freedom, the reduced χ^2 value indicates a poor fit between the data and the theory in both cases, with a stronger failing evidence in the squeezed-limit configuration (see figure 16). In the case of isosceles triangles, we observe a larger discrepancy between the data and the theory at a redshift of 2.6, which could be due to an underestimation of a systematic effect impacting small scales. Figure 16 displays

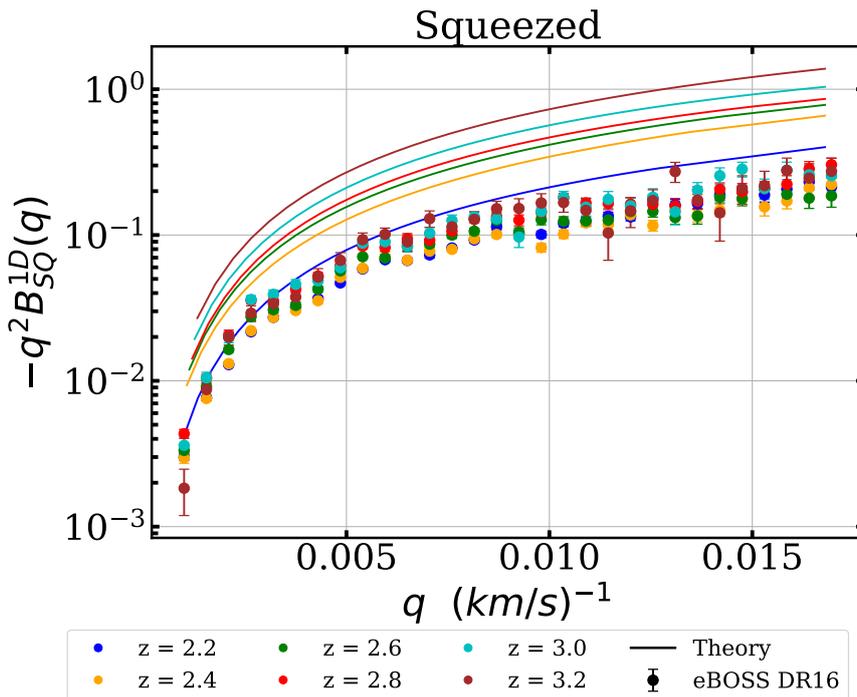


Figure 16. One-dimensional Ly- α bispectrum for squeezed limit triangles using eBOSS DR16 data (points with error bars). Solid lines represent the fit modeled theoretically by eq. (2.14). The error bars present in our measurements come from the square root of the diagonal of the bootstrap covariance matrix and systematic uncertainties added in quadrature.

the B_{SQ}^{1D} signal for the different redshift bins. The poor fit is likely due to the fact that the squeezed-limit configuration depends on the delta field in the minimum Fourier mode (the first Fourier mode, excluding the $q = 0$ mode). This mode is significantly influenced by several systematics, as discussed in section 5, such as continuum fitting and pixel masking due to DLAs or sky lines. The value of q_{\min} in the B_{SQ}^{1D} signal varies for each redshift bin, increasing with redshift and ranging from $q_{\min} = 3.3 \times 10^{-4}$ to 5.6×10^{-4} (km/s) $^{-1}$.

We also theoretically modeled the bispectrum signal using mock data, but the results were unsatisfactory. Given the error bars associated with the bispectrum signal, the χ^2 value indicated a poor fit for both configurations, despite the A and γ values being similar to those of the observed data. Notably, the theoretical modeling of the squeezed-limit signal produced results similar to those of the observed signal; this suggests that the poor fit in this configuration may also be due to the limitations of the theoretical model. As noted in [53], the squeezed-limit bispectrum encodes the impact of large-scale fluctuations on the small-scale power spectrum, representing how the small-scale power spectrum “responds” to large-scale fluctuations. We find that large-scale fluctuations are strongly affected by systematics, while the 2OPT-based theoretical model may inadequately describe small scales. In future work, we will explore the theoretical modeling provided by 2OPT in more detail. It may be necessary to employ a different model or extend the model proposed by [49] by accounting for factors such as peculiar velocities and thermal broadening.

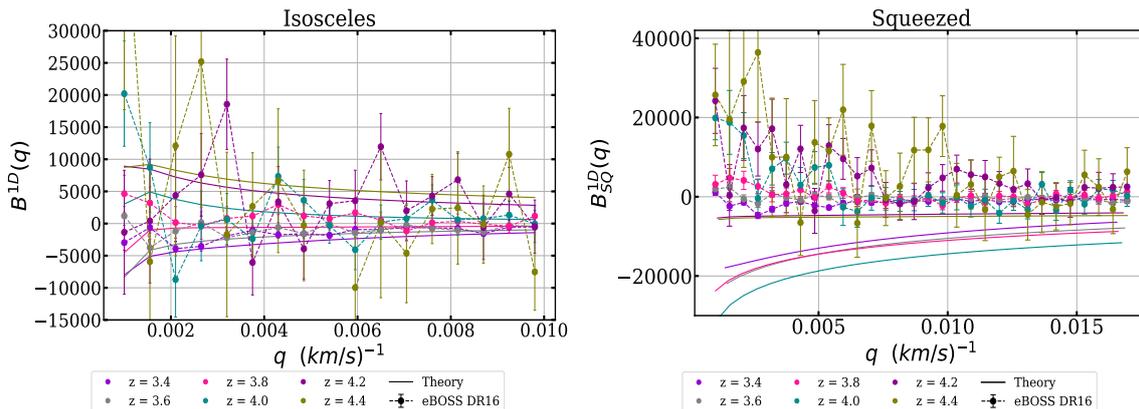


Figure 17. One-dimensional Ly- α bispectrum for redshift bins from $z = 3.4$ to $z = 4.4$ in the two triangular configurations considered. Solid lines represent the fit theoretically modeled by eq. (2.14). The error bars present in our measurements come from the square root of the diagonal of the bootstrap covariance matrix and systematic uncertainties added in quadrature.

Data	z-bin	A	γ	N_{chunks}
eBOSS	2.2	0.39 ± 0.06	1.49 ± 0.50	39972
	2.4	0.47 ± 0.08	1.48 ± 0.49	34511
	2.6	0.53 ± 0.07	1.47 ± 0.43	25882
	2.8	0.62 ± 0.07	1.50 ± 0.32	19537
	3.0	0.69 ± 0.07	1.31 ± 0.32	13402
	3.2	0.76 ± 0.12	1.38 ± 0.48	8335

Table 2. Parameter’s fitting to observed 1D bispectrum for all triple q-configurations for different redshift bins. The γ value (power-law index of the gas temperature-density relation) is estimated from the β parameter in eq. (2.11).

We also measured the bispectrum for the high redshift bins ($3.4 < z_{\text{Ly-}\alpha} < 4.4$); however, the signal was not significantly detected. The bispectrum signal for these redshift bins exhibits large fluctuations around zero, making it challenging to obtain a reliable estimate (see the upper panel of figure 17). In the case of the squeezed triangles (lower panel of figure 17), the signal appears to shift toward positive values. This shift is more noticeable in the last two redshift bins ($z_{\text{Ly-}\alpha} = 4.2$ and 4.4). Nonetheless, due to the large error bars resulting from the low number of chunks, it is difficult to confidently assert this observation.

The fit to the 2D bispectrum signal depends on the parameters c_1 and c_2 , which can be expressed as functions of A and β in eq. (2.11). We present the results of the fit in terms of the A values. Instead of using β , we opted to use γ , the power-law index of the gas temperature-density relation. The results are shown in table 2. Our results for $z = 2.4$ are in agreement with previous measurements of the γ parameter reported by [71, 72], and [84], with differences of less than 3%. [84] used high-resolution quasar spectra from the KODIAQ dataset to estimate the γ parameter for redshift bins similar to those we have considered. Overall, our results are comparable, with differences of less than 5%, except for the redshift bins at $z = 2.6$ and $z = 3.2$, where the discrepancy is as high as 7%.

6.2 Si_{III} cross-correlation

The correlated background due to absorption by Ly- α and Si_{III} from the same gas cloud along the quasar line of sight has been studied in [18] and [19]. They explain that this correlation can be estimated directly from the power spectrum signal. This suggests that we can extend this idea to the bispectrum case as well. Consider two absorption lines, one due to Ly- α and one due to Si_{III}. The Si_{III} absorption line is weaker than the Ly- α absorption by a factor of a and has a separation of $\mu = 2271$ km/s. We can write the total transmission fluctuation field as:

$$\delta_{\text{F}}(v) = \delta(v) + a \cdot \delta(v + \mu), \quad (6.1)$$

where $\delta(v)$ represent the Ly- α component. The Fourier transform of this total transmission field is given by

$$\tilde{\delta}_{\text{F}}(k) = \tilde{\delta}(k) \cdot \left(1 + a \cdot e^{ik\mu}\right). \quad (6.2)$$

The power spectrum is then $P \propto |\tilde{\delta}_{\text{F}}|^2$,

$$P_{\text{F}}(k) = P(k) \left(1 + a^2 + 2a \cos(k\mu)\right), \quad (6.3)$$

while the bispectrum is $B \propto \text{Re}[\tilde{\delta}_{\text{F}}(q_0)\tilde{\delta}_{\text{F}}(q_1)\tilde{\delta}_{\text{F}}(-q_0 - q_1)]$,

$$B_{\text{F}}(q_0, q_1) = B(q_0, q_1) \left[1 + a^3 + a(1 + a) \left(\cos(q_0\mu) + \cos(q_1\mu) + \cos((q_0 + q_1)\mu)\right)\right]. \quad (6.4)$$

Eq. (6.4) describes the oscillations due to the Ly- α /Si_{III} cross-correlation for all triangle configurations along the line of sight. This expression helps us describe the dark areas in the left panel of figure 1. However, we need to model the signal originating solely from the Ly- α forest. Although this modeling could theoretically be based on the approach explained in section 2.4, the results may be insufficient, as a good fit is not achieved for all triangle configurations, as previously shown. For this reason, we decided to model the $B^{1\text{D}}$ signal independently for the two chosen triangle configurations. As a first step, let us derive the appropriate expressions for these configurations from eq. (6.4).

Implementing the isosceles triangles configuration defined as $q = q_1 = q_0$,

$$B^{1\text{D}}(q) = B_{\text{Ly-}\alpha}^{1\text{D}}(q) \left[1 + a^3 + a(1 + a) \left(2 \cos(q\mu) + \cos(2q\mu)\right)\right], \quad (6.5)$$

now over the squeezed limit triangle configuration given by $q_0 = q - q_{\text{min}}$ and $q_1 = -q - q_{\text{min}}$,

$$B_{\text{SQ}}^{1\text{D}}(q) = B_{\text{SQ Ly-}\alpha}^{1\text{D}}(q) \left[1 + a^3 + a(1 + a) \left(2 \cos(q\mu) \cos(q_{\text{min}}\mu) + \cos(2q_{\text{min}}\mu)\right)\right]. \quad (6.6)$$

The bispectrum generated solely by Ly- α absorption lines is modeled using eq. (3.5). Although this equation was originally designed to model $kP^{1\text{D}}(k)/\pi$, it provides a good fit to the quantity $-q^2 B^{1\text{D}}(q)$ for either configuration. We independently fit $B_{\text{Ly-}\alpha}^{1\text{D}}$ and $B_{\text{SQ Ly-}\alpha}^{1\text{D}}$, and then fit the oscillating components of eqs. (6.5) and (6.6), respectively. Table 3 presents the parameter results from eq. (3.5) for the $P^{1\text{D}}$ and the two bispectrum configurations at $z = 2.4$.

We clearly detect, the oscillation pattern in the bispectrum due the Ly- α /Si_{III} cross-correlation, see figure 18. For our first fit to $B^{1\text{D}}$ accounting for Si_{III}, we use a as a extra

Signal	k_1	A	n	α	B	β
B^{1D}	0.009	0.57	-3.52	-0.25	11.99	-9.69
B_{SQ}^{1D}	0.89	0.46	-3.60	-0.10	9.79	-10.28
P^{1D}	0.9	0.31	-4.22	-0.11	13.24	-9.73

Table 3. Parameter’s fitting to observed 1D bispectrum oscillations for both q-configurations at $z = 2.4$, as well as to power spectrum.

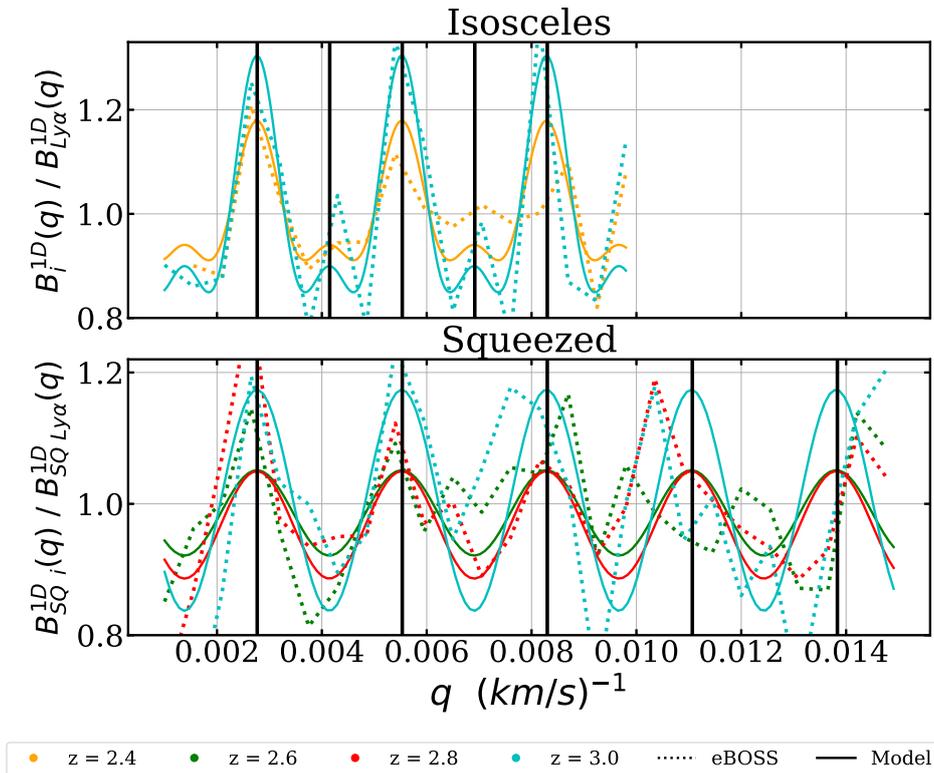


Figure 18. Oscillations present in the 1D bispectrum measurement for the two triangle configuration considered. The oscillations are induced by Ly- α /SiIII cross-correlation. The dotted lines correspond to eBOSS DR16 data, while the solid lines show our best fit modelled by eqs. (6.5) and (6.6). The bispectrum signal without oscillation is computed using (3.5), but fitting to the quantity $-q^2 B^{1D}(q)$.

free parameter of the fit. We found a remarkable improvement in χ^2 , from 38.35 to 8.26 at $z = 3.0$, with similar results for other redshift bins. As in [18], after a very simple fit we find that $a \sim 0.04$, of course this value depends on the redshift.

The top panel of figure 18 shows the oscillations of isosceles triangles configuration as described by eq. (6.5). The largest oscillations are produced by the term ‘ $2 \cos(q\mu)$ ’, with a frequency of $q = 2\pi/\mu$. The term ‘ $\cos(2q\mu)$ ’ induces smaller oscillations in the B^{1D} signal, with a frequency of $q = \pi/\mu$. The vertical lines indicate the peaks of both induced oscillations. The figure clearly shows the smaller amplitude oscillation at $q = 4.1 \times 10^{-3}$ and $6.9 \times 10^{-3} \text{ (km/s)}^{-1}$ in the estimated B^{1D} signal. The lower panel corresponds to the oscillations present in the B_{SQ}^{1D} signal, modeled by eq. (6.6), where the oscillation frequency is $q = 2\pi/\mu$. Similarly, the vertical lines indicate the oscillation frequency.

7 Conclusions

The Lyman alpha forest (Ly- α) is a unique high-redshift, multi-scale probe of the matter distribution in the universe. It can be used to characterize the expansion history of the universe on the largest scales, while also providing insights into the growth of structure, the nature of dark matter, neutrino masses, and primordial features when focusing on the smallest scales. A common approach to extract small-scale information is to use the two-point correlation function in Fourier space along the line of sight, known as the 1D power spectrum (P^{1D}). However, additional information can be obtained by using higher-order correlation functions, whose signal becomes subdominant as the order increases. The next estimator to the P^{1D} in this signal-to-noise hierarchy is the 1D bispectrum (B^{1D}), which correlates three pixels along the line of sight. Thus, an important question is whether we can measure the B^{1D} for the Lyman alpha forest using the current high-precision data from spectroscopic surveys.

In this context, we perform the first measurement of the Lyman alpha forest 1D bispectrum using eBOSS DR16 data in the redshift range $2.2 < z < 4.4$. A statistically significant bispectrum signal is observed in the range $2.2 < z < 3.2$, where the signal is negative, smooth, and exhibits a clear dependence on the redshift. Moreover, the signal clearly reveals oscillations due to the metal absorption lines (particularly from Si_{III}) in the Ly- α forest region, which we effectively account for with a simple model. Our measurement encompasses all possible triangle configurations up to a given maximal frequency, as shown in figure 1. However, we focus on two specific triangle configurations: isosceles and squeezed limit. Figures 15 and 16 display the bispectrum signal for these two configurations, respectively.

The computation cost of the bispectrum scales as $\mathcal{O}(n^2)$, where n is the number of pixels in each Ly- α sub-forest (chunk). For the redshift bin containing the largest number of spectra ($z = 2.2$), the bispectrum calculation takes roughly 4 min on 128 CPU cores of NERSC.¹⁰ The bootstrap computation for the covariance matrices, using 2000 realisations for both bispectrum configurations, takes roughly 3 to 4 times longer than the bispectrum calculation for all redshifts $z \leq 3.2$ in all configurations. These two components represent the most computationally intensive stages of the entire pipeline. From the computational cost to run the analysis presented here, it is plausible to measure the bispectrum using our pipeline on the final 5-year DESI data.

To assess the robustness of the B^{1D} signal, we conduct a thorough investigation of the systematic uncertainties that affect the measurement. We use adapted synthetic data to correct for the impact of masking pixels affected by DLAs and sky lines, as well as to account for the continuum fitting procedure that we choose. These mocks were processed using the same pipeline as the observed data. The major source of uncertainty at small scales arises from the spectrograph's resolution and noise power estimation, with the latter being dominant at low redshifts. In contrast, on the largest scales, the incompleteness of the DLA catalogue is the primary source of uncertainty. There may also be uncertainty due to the incompleteness of the BAL catalogue, but we leave a detailed study to a future publication. Another systematic effect concerns metal absorption on the redder side of the Ly- α peak (known as side bands). We perform the first measurement of the 1D bispectrum

¹⁰<https://www.nersc.gov/>

for these side bands, with that defined closer to the Ly- α peak showing the highest power, as in the P^{1D} case. We characterize this side-band bispectrum using oscillatory functions of the Si IV and C IV doublets, and use it to remove the background power present in the bispectrum signal of the Lyman-alpha forest.

To understand the shape and redshift scaling of the full 1D bispectrum signal, we use a simple model based on second-order perturbation theory, which relates the power spectrum to the bispectrum, together with a simple bias model. We find that this analytic description can reproduce the slope of the estimated bispectrum for both the isosceles triangle configuration and the squeezed limit. However, it cannot accurately reproduce the amplitude of the signal, especially in the squeezed limit configuration where we expect non-linear physics to play a major role. The quality of the fit was assessed by calculating χ^2 for both isosceles and squeezed configurations. In summary, the value of the reduced χ^2 indicates a poor fit between the data and the theory in both configurations, with stronger failing evidence in the squeezed-limit configuration (see table 1). This is to be expected since the squeezed configuration needs better modeling of the smallest scales as well as a better understanding of the involved systematics in the measurement. However, with this simple modeling we are able to estimate the power-law index γ of the gas temperature-density relation for different redshift bins, and the values are in agreement with previous measurements reported in [71, 72], and [84].

We foresee two main avenues to improve the modeling of the bispectrum signal. One approach is to use emulators from computationally expensive hydrodynamical simulations, as recently done with a neural network emulator for the 1D power spectrum [85]. An interesting second approach, which is widely used in the cosmological analysis of spectroscopic discrete tracers (see [86] and references therein), would be to include EFT corrections in the perturbation model and a more complete bias prescription. The EFT terms allow for the effect of short-scale physics, as well as correcting on the largest modes via the infrared resummations, showing encouraging results in recent studies of the 1D lya power spectrum [40, 87]. The EFT approach adds several nuisance parameters that may not be well constrained by the data. Therefore, informative priors from simulations may prove useful in order to perform well with this analytical modeling. One advantage of this analytic modeling is the possibility of including physics beyond the standard cosmological scenario, which will benefit from the analysis using higher-order statistics (see, for example, [88]) and which may be unrealistic to pursue with expensive hydrodynamical simulations.

Although we measure the bispectrum for the high redshift bins $3.4 < z < 4.4$, the signal is quite noisy and exhibits a tendency to change sign. Upcoming spectroscopic surveys, such as DESI, will help clarify whether these high-redshift signals are statistical anomalies or related to the evolution of mean transmission with redshift. We plan to refine our analysis using DESI data, as improvements in noise calibration and spectral resolution, as shown by [21], should enhance signal quality. Furthermore, extending the analysis to more sophisticated analytic models, such as those ideas expressed above (see also [89]), in combination with a clean bispectrum measurement using DESI data and the pipeline presented here, could allow the use of the B^{1D} to place additional constraints on cosmological and intergalactic medium parameters, beyond those provided by the power spectrum alone. As happens with galaxies, using higher-order correlations in addition to the power-spectrum analysis should reduce the uncertainty on the cosmological parameters, through breaking degeneracies and extracting

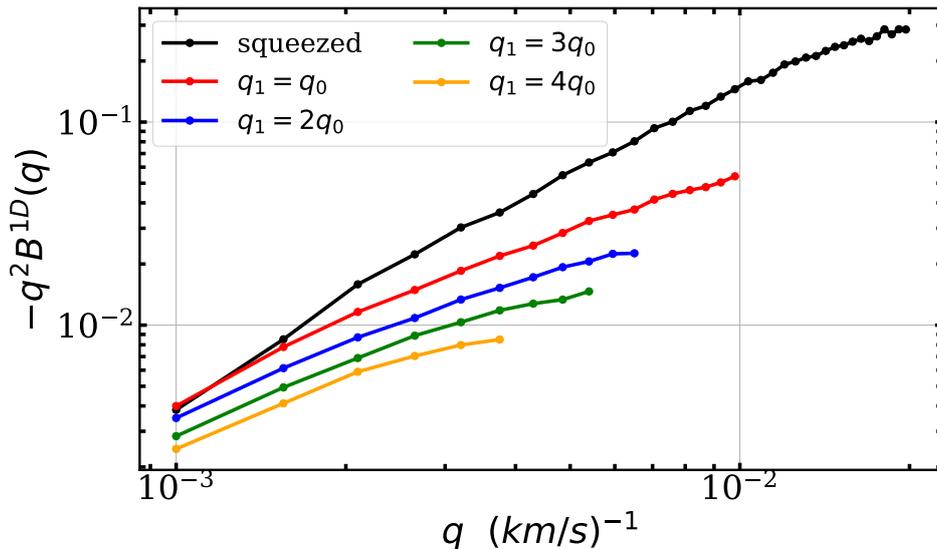


Figure 19. One-dimensional bispectrum for different triangles configurations. The q_{\max} value for the different configurations depends on the size of the forest.

more of the non-gaussian information in the data. It can also open the door to explore new physics which cannot be seen by two-point statistics, such as parity [48].

Acknowledgments

We thank Andreu Font-Ribera and Alma X. Gonzalez Morales for their advise, early discussions on the project, and for providing some of the mock catalogues. We also thank the DESI Ly- α WG members who have contributed indirectly to the infrastructure of the pipeline analysis use here. R.D.L.C. and G.N. acknowledge the computational resources of the DCI-UG DataLab, and the financial support of CONAHCYT, through the graduate fellowship programs and grants “Ciencia de Frontera 2019” No. 102958 and “Ciencia Básica y de Frontera” No. CBF2023-2024-162; the DAIP-UG and the Instituto Avanzado de Cosmología. V.I. is supported by the Kavli Foundation. C.R. Acknowledge the funding from Excellence Initiative of Aix-Marseille University - A*MIDEX, a French “Investissements d’Avenir” program (AMX-20-CE-02 - DARKUNI).

A Maximal q scale and signal for different triangle configurations

The bispectrum is constructed using triangles whose vertices are at modes q_0 , q_1 and $-(q_0 + q_1)$. Among all possible triangle configurations, some exhibit higher power, particularly those that share a common q -mode. This phenomenon is explained in [65] and [67]. We verify this by extracting the signal for different triangle configurations from the complete bispectrum signal (see figure 1). Figure 19 shows the B^{1D} for different common q -vectors ($q_1/q_0 = 1, 2, 3, 4$) and the special case of the squeezed limit (eq. (2.10)). Since the shot-noise in the bispectrum signal is derived from the power spectrum signal, the value of q_{\max} for different triangle configurations corresponds to the k_{\max} value in the P^{1D} signal. This k_{\max} value is determined by the size of

the forest, the spectrograph resolution and the Nyquist-Shannon limit ($k_{\text{Nyquist}} = \pi/\Delta v$). In our P^{1D} estimate, $k_{\text{max}} = 0.02 \text{ (km/s)}^{-1}$. According to eq. (2.9), the power spectrum includes a term that depends on $(q_0 + q_1)$. Since we have chosen triangles of the form $q_1/q_0 = n$, where n is an integer, it follows that $q_0 + q_1 = nq = k_{\text{max}}$. Therefore, $q = q_{\text{max}} = k_{\text{max}}/n$.

References

- [1] J.E. Gunn and B.A. Peterson, *On the Density of Neutral Hydrogen in Intergalactic Space*, *Astrophys. J.* **142** (1965) 1633 [INSPIRE].
- [2] U. Seljak et al., *Cosmological parameter analysis including SDSS Ly α forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy*, *Phys. Rev. D* **71** (2005) 103515 [astro-ph/0407372] [INSPIRE].
- [3] M. Viel, J. Lesgourgues, M.G. Haehnelt, S. Matarrese and A. Riotto, *Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman- α forest*, *Phys. Rev. D* **71** (2005) 063534 [astro-ph/0501562] [INSPIRE].
- [4] N. Palanque-Delabrouille et al., *Hints, neutrino bounds and WDM constraints from SDSS DR14 Lyman- α and Planck full-survey data*, *JCAP* **04** (2020) 038 [arXiv:1911.09073] [INSPIRE].
- [5] S. Goldstein, J.C. Hill, V. Iršič and B.D. Sherwin, *Canonical Hubble-Tension-Resolving Early Dark Energy Cosmologies Are Inconsistent with the Lyman- α Forest*, *Phys. Rev. Lett.* **131** (2023) 201001 [arXiv:2303.00746] [INSPIRE].
- [6] M. McQuinn, *The Evolution of the Intergalactic Medium*, *Ann. Rev. Astron. Astrophys.* **54** (2016) 313 [arXiv:1512.00086] [INSPIRE].
- [7] P. McDonald, *Toward a measurement of the cosmological geometry at $z \sim 2$: predicting Ly α forest correlation in three dimensions, and the potential of future data sets*, *Astrophys. J.* **585** (2003) 34 [astro-ph/0108064] [INSPIRE].
- [8] A. Font-Ribera, P. McDonald and A. Slosar, *How to estimate the 3D power spectrum of the Lyman- α forest*, *JCAP* **01** (2018) 003 [arXiv:1710.11036] [INSPIRE].
- [9] R. de Belsunce, O.H.E. Philcox, V. Irsic, P. McDonald, J. Guy and N. Palanque-Delabrouille, *The 3D Lyman- α forest power spectrum from eBOSS DR16*, *Mon. Not. Roy. Astron. Soc.* **533** (2024) 3756 [arXiv:2403.08241] [INSPIRE].
- [10] M.L. Abdul-Karim, E. Armengaud, G. Mention, S. Chabanier, C. Ravoux and Z. Lukić, *Measurement of the small-scale 3D Lyman- α forest power spectrum*, *JCAP* **05** (2024) 088 [arXiv:2310.09116] [INSPIRE].
- [11] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, arXiv:1611.00036 [INSPIRE].
- [12] DESI collaboration, *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **164** (2022) 207 [arXiv:2205.10939] [INSPIRE].
- [13] T.-S. Kim, M. Viel, M.G. Haehnelt, R.F. Carswell and S. Cristiani, *The power spectrum of the flux distribution in the Lyman α forest of a large sample of UVES QSO absorption spectra (LUQAS)*, *Mon. Not. Roy. Astron. Soc.* **347** (2004) 355 [astro-ph/0308103] [INSPIRE].
- [14] R.A.C. Croft, D.H. Weinberg, N. Katz and L. Hernquist, *Recovery of the power spectrum of mass fluctuations from observations of the Ly α forest*, *Astrophys. J.* **495** (1998) 44 [astro-ph/9708018] [INSPIRE].

- [15] R.A.C. Croft et al., *Towards a precise measurement of matter clustering: Ly α forest data at redshifts 2–4*, *Astrophys. J.* **581** (2002) 20 [[astro-ph/0012324](#)] [[INSPIRE](#)].
- [16] P. McDonald et al., *The Observed probability distribution function, power spectrum, and correlation function of the transmitted flux in the Lyman- α forest*, *Astrophys. J.* **543** (2000) 1 [[astro-ph/9911196](#)] [[INSPIRE](#)].
- [17] SDSS collaboration, *The Sloan Digital Sky Survey: Technical Summary*, *Astron. J.* **120** (2000) 1579 [[astro-ph/0006396](#)] [[INSPIRE](#)].
- [18] P. McDonald et al., *The Ly α forest power spectrum from the Sloan Digital Sky Survey*, *Astrophys. J. Suppl.* **163** (2006) 80 [[astro-ph/0405013](#)] [[INSPIRE](#)].
- [19] N. Palanque-Delabrouille et al., *The one-dimensional Ly- α forest power spectrum from BOSS*, *Astron. Astrophys.* **559** (2013) A85 [[arXiv:1306.5896](#)] [[INSPIRE](#)].
- [20] S. Chabanier et al., *The one-dimensional power spectrum from the SDSS DR14 Ly α forests*, *JCAP* **07** (2019) 017 [[arXiv:1812.03554](#)] [[INSPIRE](#)].
- [21] C. Ravoux et al., *The Dark Energy Spectroscopic Instrument: one-dimensional power spectrum from first Ly α forest samples with Fast Fourier Transform*, *Mon. Not. Roy. Astron. Soc.* **526** (2023) 5118 [[arXiv:2306.06311](#)] [[INSPIRE](#)].
- [22] N.G. Karaçaylı et al., *Optimal 1D Ly α Forest Power Spectrum Estimation — III. DESI early data*, *Mon. Not. Roy. Astron. Soc.* **528** (2024) 3941 [[arXiv:2306.06316](#)] [[INSPIRE](#)].
- [23] M.T. Murphy, G.G. Kacprzak, G.A.D. Savorgnan and R.F. Carswell, *The UVES Spectral Quasar Absorption Database (SQUAD) data release 1: the first 10 million seconds*, *Mon. Not. Roy. Astron. Soc.* **482** (2018) 3458.
- [24] J.M. O’Meara et al., *The First Data Release of the KODIAQ Survey*, *Astron. J.* **150** (2015) 111 [[arXiv:1505.03529](#)] [[INSPIRE](#)].
- [25] S. López et al., *XQ-100: A legacy survey of one hundred $3.5 \lesssim z \lesssim 4.5$ quasars observed with VLT/X-shooter*, *Astron. Astrophys.* **594** (2016) A91.
- [26] M. Walther et al., *A New Precision Measurement of the Small-scale Line-of-sight Power Spectrum of the Ly α Forest*, *Astrophys. J.* **852** (2018) 22 [[arXiv:1709.07354](#)] [[INSPIRE](#)].
- [27] N.G. Karaçaylı et al., *Optimal 1D Ly α forest power spectrum estimation — II. KODIAQ, SQUAD, and XQ-100*, *Mon. Not. Roy. Astron. Soc.* **509** (2022) 2842 [[arXiv:2108.10870](#)] [[INSPIRE](#)].
- [28] V. Iršič et al., *The Lyman α forest power spectrum from the XQ-100 Legacy Survey*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 4332 [[arXiv:1702.01761](#)] [[INSPIRE](#)].
- [29] M. Viel, G.D. Becker, J.S. Bolton and M.G. Haehnelt, *Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman- α forest data*, *Phys. Rev. D* **88** (2013) 043502 [[arXiv:1306.2314](#)] [[INSPIRE](#)].
- [30] A. Day, D. Tytler and B. Kambalur, *Power spectrum of the flux in the Lyman- α forest from high-resolution spectra of 87 QSOs*, *Mon. Not. Roy. Astron. Soc.* **489** (2019) 2536 [[INSPIRE](#)].
- [31] V. Khaire et al., *The power spectrum of the Lyman- α Forest at $z < 0.5$* , *Mon. Not. Roy. Astron. Soc.* **486** (2019) 769 [[arXiv:1808.05605](#)] [[INSPIRE](#)].
- [32] E. Boera, G.D. Becker, J.S. Bolton and F. Nasir, *Revealing Reionization with the Thermal History of the Intergalactic Medium: New Constraints from the Ly α Flux Power Spectrum*, *Astrophys. J.* **872** (2019) 101 [[arXiv:1809.06980](#)] [[INSPIRE](#)].

- [33] P. Gaikwad, R. Srianand, M.G. Haehnelt and T.R. Choudhury, *A consistent and robust measurement of the thermal state of the IGM at $2 \leq z \leq 4$ from a large sample of Ly α forest spectra: evidence for late and rapid He II reionization*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 4389 [[arXiv:2009.00016](#)] [[INSPIRE](#)].
- [34] F.X. Linares Cedeño, A.X. González-Morales and L.A. Ureña-López, *Ultralight DM bosons with an axion-like potential: scale-dependent constraints revisited*, *JCAP* **01** (2021) 051 [[arXiv:2006.05037](#)] [[INSPIRE](#)].
- [35] N. Porqueres, J. Jasche, G. Lavaux and T. Enßlin, *Inferring high-redshift large-scale structure dynamics from the Lyman- α forest*, *Astron. Astrophys.* **630** (2019) A151 [[arXiv:1907.02973](#)] [[INSPIRE](#)].
- [36] S. Bose et al., *ETHOS — an Effective Theory of Structure Formation: detecting dark matter interactions through the Lyman- α forest*, *Mon. Not. Roy. Astron. Soc.* **487** (2019) 522 [[arXiv:1811.10630](#)] [[INSPIRE](#)].
- [37] A. Borde et al., *New approach for precise computation of Lyman- α forest power spectrum with hydrodynamical simulations*, *JCAP* **07** (2014) 005 [[arXiv:1401.6472](#)] [[INSPIRE](#)].
- [38] N. Palanque-Delabrouille et al., *Neutrino masses and cosmology with Lyman- α forest power spectrum*, *JCAP* **11** (2015) 011 [[arXiv:1506.05976](#)] [[INSPIRE](#)].
- [39] C. Yèche, N. Palanque-Delabrouille, J. Baur and H. du Mas des Bourboux, *Constraints on neutrino masses from Lyman- α forest power spectrum with BOSS and XQ-100*, *JCAP* **06** (2017) 047 [[arXiv:1702.03314](#)] [[INSPIRE](#)].
- [40] M.M. Ivanov, M.W. Toomey and N.G. Karaçaylı, *Fundamental Physics with the Lyman- α Forest: Constraints on the Growth of Structure and Neutrino Masses from SDSS with Effective Field Theory*, *Phys. Rev. Lett.* **134** (2025) 091001 [[arXiv:2405.13208](#)] [[INSPIRE](#)].
- [41] B. Wilson, V. Iršič and M. McQuinn, *A measurement of the Ly β forest power spectrum and its cross with the Ly α forest in X-Shooter XQ-100*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 2423 [[arXiv:2106.04837](#)] [[INSPIRE](#)].
- [42] M. Dijkstra, A. Lidz and L. Hui, *Beyond Lyman α : Constraints and Consistency Tests from the Lyman β Forest*, *Astrophys. J.* **605** (2004) 7 [[astro-ph/0305498](#)] [[INSPIRE](#)].
- [43] S. Chongchitnan, *The Lyman- α Forest as a tool for disentangling non-Gaussianities*, *JCAP* **10** (2014) 034 [[arXiv:1408.4340](#)] [[INSPIRE](#)].
- [44] M. Liguori, E. Sefusatti, J.R. Fergusson and E.P.S. Shellard, *Primordial non-Gaussianity and Bispectrum Measurements in the Cosmic Microwave Background and Large-Scale Structure*, *Adv. Astron.* **2010** (2010) 980523 [[arXiv:1001.4707](#)] [[INSPIRE](#)].
- [45] R. Scoccimarro, *The bispectrum: from theory to observations*, *Astrophys. J.* **544** (2000) 597 [[astro-ph/0004086](#)] [[INSPIRE](#)].
- [46] M.M. Ivanov, O.H.E. Philcox, G. Cabass, T. Nishimichi, M. Simonović and M. Zaldarriaga, *Cosmology with the galaxy bispectrum multipoles: Optimal estimation and application to BOSS data*, *Phys. Rev. D* **107** (2023) 083515 [[arXiv:2302.04414](#)] [[INSPIRE](#)].
- [47] M. Zaldarriaga, U. Seljak and L. Hui, *Correlations across scales in the Lyman α forest: Testing the gravitational instability paradigm*, *Astrophys. J.* **551** (2001) 48 [[astro-ph/0007101](#)] [[INSPIRE](#)].
- [48] P. Adari and A. Slosar, *Searching for parity violation in SDSS DR16 Lyman- α forest data*, *Phys. Rev. D* **110** (2024) 103534 [[arXiv:2405.04660](#)] [[INSPIRE](#)].

- [49] M. Viel et al., *The bispectrum of the Lyman α forest at $z < 2-2.4$ from a large sample of UVES QSO absorption spectra (LUQAS)*, *Mon. Not. Roy. Astron. Soc.* **347** (2004) L26 [[astro-ph/0308151](#)] [[INSPIRE](#)].
- [50] R. Mandelbaum, P. McDonald, U. Seljak and R. Cen, *Precision cosmology from the Lyman- α forest: Power spectrum and bispectrum*, *Mon. Not. Roy. Astron. Soc.* **344** (2003) 776 [[astro-ph/0302112](#)] [[INSPIRE](#)].
- [51] M. Viel, E. Branchini, K. Dolag, M. Grossi, S. Matarrese and L. Moscardini, *Primordial non-Gaussianities in the Intergalactic Medium*, *Mon. Not. Roy. Astron. Soc.* **393** (2009) 774 [[arXiv:0811.2223](#)] [[INSPIRE](#)].
- [52] S. Maitra, R. Srianand, P. Gaikwad, T.R. Choudhury, A. Paranjape and P. Petitjean, *Three- and two-point spatial correlations of IGM at $z \sim 2$: cloud-based analysis using simulations*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 6100 [[arXiv:2005.05346](#)] [[INSPIRE](#)].
- [53] C.-T. Chiang, A.M. Cieplak, F. Schmidt and A. Slosar, *Response approach to the squeezed-limit bispectrum: application to the correlation of quasar and Lyman- α forest power spectrum*, *JCAP* **06** (2017) 022 [[arXiv:1701.03375](#)] [[INSPIRE](#)].
- [54] C.-T. Chiang, C. Wagner, F. Schmidt and E. Komatsu, *Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum*, *JCAP* **05** (2014) 048 [[arXiv:1403.3411](#)] [[INSPIRE](#)].
- [55] D.K. Hazra and T.G. Sarkars, *Primordial Non-Gaussianity in the Forest: 3D Bispectrum of Ly- α Flux Spectra Along Multiple Lines of Sight*, *Phys. Rev. Lett.* **109** (2012) 121301 [[arXiv:1205.2790](#)] [[INSPIRE](#)].
- [56] T.G. Sarkar and D.K. Hazra, *Probing primordial non-Gaussianity: The 3D Bispectrum of Ly- α forest and the redshifted 21 cm signal from the post reionization epoch*, *JCAP* **04** (2013) 002 [[arXiv:1211.4756](#)] [[INSPIRE](#)].
- [57] B.W. Lyke et al., *The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release*, *Astrophys. J. Suppl.* **250** (2020) 8 [[arXiv:2007.09001](#)] [[INSPIRE](#)].
- [58] S. Chabanier et al., *The Completed Sloan Digital Sky Survey IV Extended Baryon Oscillation Spectroscopic Survey: The Damped Ly α Systems Catalog*, *Astrophys. J. Suppl.* **258** (2022) 18 [[arXiv:2107.09612](#)] [[INSPIRE](#)].
- [59] I. Pâris et al., *A principal component analysis of quasar UV spectra at $z \sim 3$* , *Astron. Astrophys.* **530** (2011) A50 [[arXiv:1104.2024](#)] [[INSPIRE](#)].
- [60] S. Bailey, *Principal Component Analysis with Noisy and/or Missing Data*, *Publ. Astron. Soc. Pac.* **124** (2012) 1015 [[arXiv:1208.4122](#)] [[INSPIRE](#)].
- [61] H. du Mas des Bourboux et al., *The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Ly α Forests*, *Astrophys. J.* **901** (2020) 153 [[arXiv:2007.08995](#)] [[INSPIRE](#)].
- [62] C. Ramírez-Pérez et al., *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](#)].
- [63] N. Kaiser and J.A. Peacock, *Power spectrum analysis of one-dimensional redshift surveys*, *Astrophys. J.* **379** (1991) 482 [[INSPIRE](#)].
- [64] V. Desjacques and A. Nusser, *Redshift distortions in one-dimensional power spectra*, *Mon. Not. Roy. Astron. Soc.* **351** (2004) 1395 [[astro-ph/0401544](#)] [[INSPIRE](#)].

- [65] S. Matarrese, L. Verde and A.F. Heavens, *Large scale bias in the universe: Bispectrum method*, *Mon. Not. Roy. Astron. Soc.* **290** (1997) 651 [[astro-ph/9706059](#)] [[INSPIRE](#)].
- [66] K.C. Chan and L. Blot, *Assessment of the Information Content of the Power Spectrum and Bispectrum*, *Phys. Rev. D* **96** (2017) 023528 [[arXiv:1610.06585](#)] [[INSPIRE](#)].
- [67] H. Gil-Marín et al., *The power spectrum and bispectrum of SDSS DR11 BOSS galaxies — I. Bias and gravity*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 539 [[arXiv:1407.5668](#)] [[INSPIRE](#)].
- [68] M. Peloso and M. Pietroni, *Galilean invariance and the consistency relation for the nonlinear squeezed bispectrum of large scale structure*, *JCAP* **05** (2013) 031 [[arXiv:1302.0223](#)] [[INSPIRE](#)].
- [69] S. Mooij and G.A. Palma, *Consistently violating the non-Gaussian consistency relation*, *JCAP* **11** (2015) 025 [[arXiv:1502.03458](#)] [[INSPIRE](#)].
- [70] S. Goldstein et al., *Squeezing fNL out of the matter bispectrum with consistency relations*, *Phys. Rev. D* **106** (2022) 123525 [[arXiv:2209.06228](#)] [[INSPIRE](#)].
- [71] P. McDonald, J. Miralda-Escude, M. Rauch, W.L.W. Sargent, T.A. Barlow and R. Cen, *A measurement of the temperature-density relation in the intergalactic medium using a new Lyman- α absorption line fitting method*, *Astrophys. J.* **562** (2001) 52 [Erratum *ibid.* **598** (2003) 712] [[astro-ph/0005553](#)] [[INSPIRE](#)].
- [72] G.C. Rudie, C.C. Steidel and M. Pettini, *The Temperature-Density Relation in the Intergalactic Medium at Redshift $\langle z \rangle = 2.4$* , *Astrophys. J. Lett.* **757** (2012) L30 [[arXiv:1209.0005](#)] [[INSPIRE](#)].
- [73] A. Arinyo-i-Prats, J. Miralda-Escudé, M. Viel and R. Cen, *The Non-Linear Power Spectrum of the Lyman α Forest*, *JCAP* **12** (2015) 017 [[arXiv:1506.04519](#)] [[INSPIRE](#)].
- [74] J.J. Givans and C.M. Hirata, *Redshift-space streaming velocity effects on the Lyman- α forest baryon acoustic oscillation scale*, *Phys. Rev. D* **102** (2020) 023515 [[arXiv:2002.12296](#)] [[INSPIRE](#)].
- [75] L. Verde, A.F. Heavens, S. Matarrese and L. Moscardini, *Large scale bias in the universe. 2. Redshift space bispectrum*, *Mon. Not. Roy. Astron. Soc.* **300** (1998) 747 [[astro-ph/9806028](#)] [[INSPIRE](#)].
- [76] N.G. Karaçaylı et al., *A framework to measure the properties of intergalactic metal systems with two-point flux statistics*, *Mon. Not. Roy. Astron. Soc.* **522** (2023) 5980 [[arXiv:2302.06936](#)] [[INSPIRE](#)].
- [77] A. Kramida, J. Reader et al., *NIST Atomic Spectra Database*, version 5.11, National Institute of Standards and Technology, Gaithersburg, MD, U.S.A. (2009), <https://physics.nist.gov/asd>.
- [78] N.G. Karaçaylı, A. Font-Ribera and N. Padmanabhan, *Optimal 1D Ly α forest power spectrum estimation — I. DESI-lite spectra*, *Mon. Not. Roy. Astron. Soc.* **497** (2020) 4742 [[arXiv:2008.06421](#)] [[INSPIRE](#)].
- [79] 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](#)].
- [80] J. Farr et al., *LyaCoLoRe: synthetic datasets for current and future Lyman- α forest BAO surveys*, *JCAP* **03** (2020) 068 [[arXiv:1912.02763](#)] [[INSPIRE](#)].
- [81] F. Bernardeau, S. Colombi, E. Gaztañaga and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, *Phys. Rep.* **367** (2002) 1 [[astro-ph/0112551](#)] [[INSPIRE](#)].

- [82] 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](#)].
- [83] K.K. Rogers, S. Bird, H.V. Peiris, A. Pontzen, A. Font-Ribera and B. Leistedt, *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](#)].
- [84] K.N. Telikova, P.S. Shternin and S.A. Balashev, *Thermal state of the intergalactic medium at $z \sim 2-4$* , *Astrophys. J.* **887** (2019) 205 [[arXiv:1910.13184](#)] [[INSPIRE](#)].
- [85] L. Cabayol-Garcia, J. Chaves-Montero, A. Font-Ribera and C. Pedersen, *A neural network emulator for the Lyman- α forest 1D flux power spectrum*, *Mon. Not. Roy. Astron. Soc.* **525** (2023) 3499 [[arXiv:2305.19064](#)] [[INSPIRE](#)].
- [86] DESI collaboration, *DESI 2024 V: Full-Shape Galaxy Clustering from Galaxies and Quasars*, [arXiv:2411.12021](#) [[INSPIRE](#)].
- [87] M.M. Ivanov, *Lyman α forest power spectrum in effective field theory*, *Phys. Rev. D* **109** (2024) 023507 [[arXiv:2309.10133](#)] [[INSPIRE](#)].
- [88] A. Aviles and G. Niz, *Galaxy three-point correlation function in modified gravity*, *Phys. Rev. D* **107** (2023) 063525 [[arXiv:2301.07240](#)] [[INSPIRE](#)].
- [89] J. Chaves-Montero et al., *ForestFlow: predicting the Lyman- α forest clustering from linear to nonlinear scales*, *Astron. Astrophys.* **694** (2025) A187 [[arXiv:2409.05682](#)] [[INSPIRE](#)].