

OPEN ACCESS

## DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman alpha forest

To cite this article: A.G. Adame *et al* JCAP01(2025)124

View the [article online](#) for updates and enhancements.

### You may also like

- [LOCALIZATION AND BROADBAND FOLLOW-UP OF THE GRAVITATIONAL-WAVE TRANSIENT GW150914](#)

B. P. Abbott, R. Abbott, T. D. Abbott et al.

- [Multi-messenger Observations of a Binary Neutron Star Merger](#)

B. P. Abbott, R. Abbott, T. D. Abbott et al.

- [DESI 2024 III: baryon acoustic oscillations from galaxies and quasars](#)

A.G. Adame, J. Aguilar, S. Ahlen et al.

RECEIVED: April 15, 2024

REVISED: July 3, 2024

ACCEPTED: July 29, 2024

PUBLISHED: January 30, 2025

## DARK ENERGY SPECTROSCOPIC INSTRUMENT (DESI) SURVEY YEAR 1 RESULTS

# DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman alpha forest



DARK ENERGY  
SPECTROSCOPIC  
INSTRUMENT

U.S. Department of Energy Office of Science

## The DESI collaboration

E-mail: [spokespersons@desi.lbl.gov](mailto:spokespersons@desi.lbl.gov)

**ABSTRACT:** We present the measurement of Baryon Acoustic Oscillations (BAO) from the Lyman- $\alpha$  (Ly $\alpha$ ) forest of high-redshift quasars with the first-year dataset of the Dark Energy Spectroscopic Instrument (DESI). Our analysis uses over 420 000 Ly $\alpha$  forest spectra and their correlation with the spatial distribution of more than 700 000 quasars. An essential facet of this work is the development of a new analysis methodology on a blinded dataset. We conducted rigorous tests using synthetic data to ensure the reliability of our methodology and findings before unblinding. Additionally, we conducted multiple data splits to assess the consistency of the results and scrutinized various analysis approaches to confirm their robustness. For a given value of the sound horizon ( $r_d$ ), we measure the expansion at  $z_{\text{eff}} = 2.33$  with 2% precision,  $H(z_{\text{eff}}) = (239.2 \pm 4.8) (147.09 \text{ Mpc}/r_d) \text{ km/s/Mpc}$ . Similarly, we present a 2.4% measurement of the transverse comoving distance to the same redshift,  $D_M(z_{\text{eff}}) = (5.84 \pm 0.14) (r_d/147.09 \text{ Mpc}) \text{ Gpc}$ . Together with other DESI BAO measurements at lower redshifts, these results are used in a companion paper to constrain cosmological parameters.

**KEYWORDS:** baryon acoustic oscillations, cosmological parameters from LSS, Lyman alpha forest, redshift surveys

ARXIV EPRINT: [2404.03001](https://arxiv.org/abs/2404.03001)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data sample</b>	<b>3</b>
2.1	The quasar catalog	5
2.2	Damped Ly $\alpha$ systems	6
2.3	Broad Absorption Line quasars	7
2.4	Catalog of the Ly $\alpha$ forest	8
2.5	Continuum fitting	9
<b>3</b>	<b>Measurement of correlations</b>	<b>10</b>
3.1	Measurement of the auto-correlation	11
3.2	Measurement of the cross-correlation	15
3.3	Covariance matrix	15
<b>4</b>	<b>Modelling of correlations</b>	<b>18</b>
4.1	Isolating the BAO information	18
4.2	Redshift evolution	18
4.3	Distortion matrix	19
4.4	Metal contamination	19
4.5	Correlated noise from the data processing	20
4.6	HCD contamination	21
4.7	Quasar redshift errors	22
4.8	Quasar radiation (proximity effect)	22
4.9	Small-scales corrections	22
<b>5</b>	<b>Measurement of Baryon Acoustic Oscillations</b>	<b>23</b>
<b>6</b>	<b>Analysis validation</b>	<b>25</b>
6.1	Blinding strategy	26
6.2	Validation using synthetic data	26
6.3	Data splits	30
6.4	Alternative analyses	32
6.5	Conclusion on the validation tests	37
<b>7</b>	<b>Discussion</b>	<b>38</b>
7.1	Cosmological distances	38
7.2	Comparison with previous measurements from SDSS	39
7.3	DESI-SDSS cross-survey covariance	40
7.4	Future work	41
<b>8</b>	<b>Conclusions</b>	<b>42</b>

<b>A Alternative modelling of HCD contaminations</b>	<b>43</b>
<b>B Nuisance parameters</b>	<b>44</b>
<b>C Blinding</b>	<b>46</b>
<b>D Comparison of sampling methods</b>	<b>48</b>
<b>E Significance of BAO shifts</b>	<b>49</b>
<b>F Estimation of the DESI-SDSS covariance</b>	<b>50</b>
<b>The DESI collaboration</b>	<b>60</b>

---

## 1 Introduction

Baryon Acoustic Oscillations (BAO) in the distribution of matter are a unique probe of the cosmic expansion history and the geometry of the Universe [1]. By themselves, BAO measurements at different redshifts enable precise measurements of the energy density parameters. In combination with studies of the Cosmic Microwave Background (CMB), they allow us to better constrain extensions to the  $\Lambda$ CDM model [2, 3].

The Dark Energy Spectroscopic Instrument (DESI, [4]) project aims to measure BAO with unprecedented precision over a wide range of redshifts. DESI is in the midst of a five-year campaign to obtain accurate redshifts for 40 million galaxies and quasars over 14 000 square degrees. The survey started in May 2021 and the upcoming DESI Data Release 1 (DESI DR1, [5]), covering data collected until June 14, 2022, contains about 13 million galaxies, 1.5 million quasars, and 4 million stars over an area of more than 9 500 square degrees.<sup>1</sup> DESI has obtained seven BAO measurements at different redshifts with this DR1 dataset.

The BAO measurements from the clustering of DESI galaxies and quasars at  $z < 2$  are presented in [6]. In this publication we present a BAO measurement at  $z = 2.33$  using the auto-correlation of the Lyman- $\alpha$  (Ly $\alpha$ ) forest dataset from DESI DR1 and its cross-correlation with quasar positions. Neutral hydrogen along the line-of-sight towards high-redshift quasars cause absorption features in their spectra. While gas pressure dominates the distribution of gas on scales of tens of kiloparsecs [7], on larger scales the Ly $\alpha$  forest is a powerful tracer of the density fluctuations [8]. The cosmological interpretation of all DESI DR1 BAO measurements is presented in [9].

While the first BAO measurements [10, 11] used the distribution of galaxies to trace the density fluctuations, the Ly $\alpha$  forest can also be used to extend the BAO measurements to higher redshifts than those available with current galaxy surveys. The Baryon Oscillation Spectroscopic Survey (BOSS, [12]) presented the first BAO measurement using the auto-correlation of the Ly $\alpha$  forest [13–15] with measurements of 50 000 quasar spectra from the 9th data release of the Sloan Digital Sky Survey (SDSS DR9, [16, 17]). With the 11th

<sup>1</sup>9 500 square degrees is the area covered by the dark time survey where quasars are observed, the total area surveyed by DESI with other programs is larger.

data release (DR11, [18]), BOSS presented as well the first BAO measurement from the cross-correlation of quasars and the Ly $\alpha$  forest [19], which doubled the amount of information available from the same dataset.

Updated Ly $\alpha$  BAO measurements were published following subsequent SDSS data releases [20–24], including data from the Extended Baryon Oscillation Spectroscopic Survey (eBOSS, [25]). The final Ly $\alpha$  BAO results used 210 000 Ly $\alpha$  forests from BOSS and eBOSS observations in the 16th data release (SDSS DR16, [26]). These results were published in du Mas des Bourboux et al. (2020) [27] and are referred to as [dMb20](#) in the rest of this article. [dMb20](#) has been the state-of-the art in Ly $\alpha$  BAO measurements until this publication, and we compare the measurements and discuss the methodological differences in the next sections (notably the use of a cross-covariance between the auto-correlation of the Ly $\alpha$  forest and its cross-correlation with quasars). In this work we use over 420 000 Ly $\alpha$  forests, doubling the number of lines of sight used in [dMb20](#).

The structure of this paper is designed to guide readers through the main aspects of the BAO measurements with the Ly $\alpha$  forest using DESI’s first year of data. We start in section 2 with a description of the DESI survey and the datasets used in our analysis, i.e. the Ly $\alpha$  forest absorption and quasar catalogues [28, 29], along with the masking of Broad Absorption Lines and Damped Lyman- $\alpha$  systems found in the spectra. The methodology to extract the Ly $\alpha$  forest fluctuations from the spectra of DESI quasars is presented in [30], and it is summarised in section 2.5. We present the measurement of correlations in section 3 and how we model them in section 4. The methodology employed in both of these sections is described in more detail in [31]. In section 4 we also summarize the (minor) contamination from sky residuals and other pipeline-induced systematics, which is described in detail in a companion paper [32]. In section 5 we present the main result from this work: the most precise BAO measurement at  $z > 2$  to date.

In order to minimize confirmation (and other) biases, the DESI Collaboration used blinding strategies in the BAO measurements. The analysis methodology was entirely developed on blinded measurements, and we were only allowed to *unblind* the measurement once we had fulfilled a long list of validation tests. The blinding strategy used in the Ly $\alpha$  BAO measurement is described in appendix C, and the analysis validation is presented in section 6. In [33] we present the validation tests in more detail. These tests are based on 150 synthetic realisations of our dataset (or mocks). These mocks were generated following the methods described in [34]. In section 7 we contextualize our findings, compare them to the Ly $\alpha$  BAO measurement from [dMb20](#), and discuss a few options to improve future Ly $\alpha$  BAO analyses. We conclude in section 8.

## 2 Data sample

We use quasar spectra collected during the first year of the DESI main survey. The observations were conducted with the Mayall 4-m telescope at Kitt Peak National Observatory, in Arizona, with a new prime focus, multi-fiber spectrograph. It consists of a new corrector equipped with an atmospheric dispersion compensator [35] providing a 3.2 degree diameter field of view, and a focal plane composed of 5000 robotically actuated fibers [36] that distribute the light to 10 spectrographs situated below the telescope, in a temperature-controlled room. Each spectrograph is composed of three arms, blue (3600–5930 Å), red (5600–7720 Å) and near

infrared (7470–9800 Å). In this analysis we use only data from the blue arms. In each blue camera, the light from 500 fibers is dispersed and refocused, forming 500 spectral traces on a 4096x4096 pixel STA4150 CCD. Despite this dense fiber packing, the cross-contamination of the spectra from adjacent fibers was measured to be only of 0.2% or smaller after spectral extraction (for wavelengths shorter than 8900 Å, see [37] section 4.1). This effect is corrected in post-processing and we expect the residual cross-talk to be negligible. The CCD pixel size is  $15\text{ }\mu\text{m}$ , corresponding to  $0.6\text{ \AA}$  in wavelength. The spectrograph line spread function is  $1.8\text{ \AA}$  FWHM in the blue channel, which corresponds to a resolution from 2000 to 3400 depending on the wavelength. Many more details on the instrument can be found in [38].

The main survey started on May 14, 2021 after a survey validation period (see [39] and references therein) used to tune the target selection, the fiber assignment, the exposure times, and exercise the data processing pipeline that is run every day to validate the observations. DESI is running two programs for bright and dark time, switching dynamically between them with the observation conditions (moonlight, but also seeing, sky brightness and sky transparency). The choice of pointing (called a *tile* when combined with a specific selection of targets to observe) and the allocation of fibers to targets is performed automatically during the night [40, 41]. The exposure times are adjusted in real time in order to obtain the same signal to noise for a fiducial target in each pointing. The quasar sample is observed during dark time. The quasar targets that are spectroscopically confirmed and have a redshift  $> 2.1$  are scheduled for re-observation to build up the signal to noise necessary for the Ly $\alpha$  forest measurements. The objective is to acquire four exposures for each quasar, each with an effective exposure time<sup>2</sup> of 1000 sec. We use data collected until June 14, 2022 for this analysis which will be made public in the DESI Data Release 1 (DESI DR1) [5].

Data collected on the mountain are transferred to NERSC (National Energy Research Scientific Computing Center) within minutes and processed by the offline pipeline [37]. CCD images are preprocessed, the spectra extracted and calibrated for each exposure. It is worth noting that the extraction algorithm returns spectra with uncorrelated noise on the same wavelength grid of bin  $0.8\text{ \AA}$  for all fibers and exposures, which simplifies the co-addition (or averaging) of spectra from the same target obtained in different exposures. The data processing pipeline provides 2 sets of co-added spectra, one per spectrograph and tile, combining data of several exposures and nights for the same pointing and fiber allocation, and one per HEALPix pixel [42] on the sky where the full co-added spectra (across exposures and tiles) of all the targets in a HEALPix pixel are saved. We use this later data set for this analysis. These spectra include fluxes, a wavelength array (in vacuum, and in the solar system barycenter frame), an estimate of the flux variance, a mask to flag bad pixels, and a resolution matrix that encodes the line spread function of the instrument [37].

In the subsequent sections, we delineate the processing applied to our quasar sample, both at the catalog and spectral level, including the identification of Damped Lyman- $\alpha$  (DLAs) systems and Broad Absorption Line (BAL) troughs, to yield the ultimate datasets for our analysis. While a comprehensive overview is presented in [30], here we emphasize on the small but yet significant adjustments made to enhance our methodology.

---

<sup>2</sup>The effective exposure time corresponds to the exposure time in ideal observing conditions: at zenith, with a nominal seeing, in photometric conditions [41].

## 2.1 The quasar catalog

Studies of the Ly $\alpha$  forest require a pure sample of quasars with accurately determined redshifts. To meet these stringent criteria, DESI employs a multi-step classification and redshift fitting procedure for the identification of quasar spectra. This process is extensively outlined in [28] (see their figure 9 summarizing the workflow) and complementary information is also provided in [5, 29, 43]. The cornerstone of DESI’s classification and redshift fitting is `Redrock`, a template fitting code [44]. It utilizes Principal Component Analysis (PCA) templates representing three broad object classes (stars, galaxies, and quasars) while scanning a range of redshifts. The determination of the most probable spectral class and redshift for a given spectrum relies on the lowest  $\chi^2$  fit. Additionally, DESI employs two quasar-specific classifiers, referred to as “afterburners”, to enhance the completeness of the quasar sample by  $\sim 10\%$  [28, 29]. One afterburner scrutinizes spectra not classified as quasars by `Redrock` for broad MgII emission; upon detection, it reclassifies the spectrum as a quasar without altering the assigned redshift. The second afterburner, known as `QuasarNET` [45, 46], employs a deep convolutional neural network to identify potential quasar emission features and estimate the redshift. If `QuasarNET` identifies a target as a quasar, there is another `Redrock` fitting iteration, although this time with only quasar templates and a redshift prior based on the `QuasarNET` result.

The resulting DESI DR1 quasar catalog was produced by the three classifiers, and is slated for publication as part of first DESI Data Release (DR1) [5]. This catalog boasts an estimated completeness and redshift purity that surpasses 95% and 98%, respectively [47].

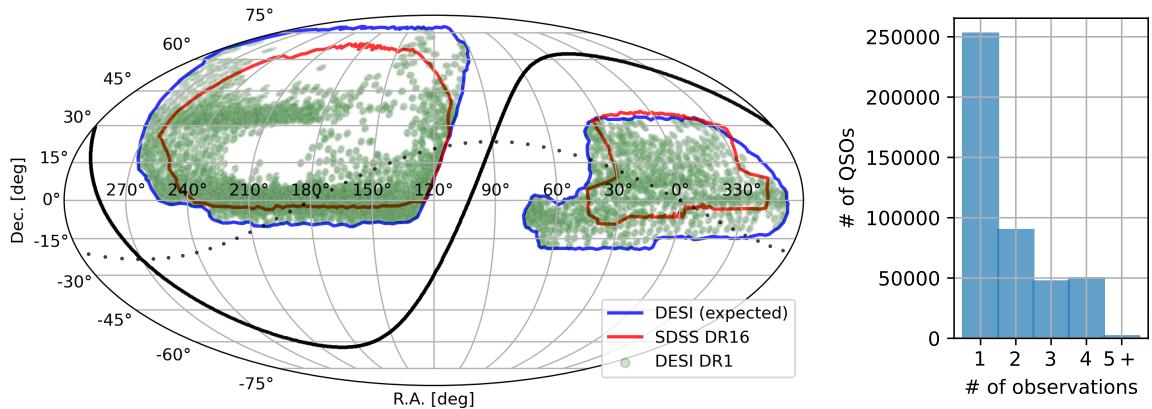
From this catalog we keep only quasars with `ZWARN=0` or `ZWARN=4`, rejecting quasars with any issue reported by the spectroscopic pipeline except for the low  $\Delta\chi^2$  flag of `Redrock` that is irrelevant for QSOs. We also use updated redshifts for this Ly $\alpha$  analysis. These are based on new quasar templates tailored for high redshifts, specifically  $1.4 < z < 7.0$ . Additionally, a slightly modified version of the `Redrock` code was used that integrates a more recent optical depth model from [48]. These subtle adjustments were prompted by the identification of a bias for redshifts  $z > 2$ , as identified by their impact on the Ly $\alpha$  quasar cross-correlation (see [49] for more details). The statistical precision of the updated redshifts is estimated to be better than  $150 \text{ km s}^{-1}$ , with a catastrophic redshift failure rate of  $\sim 2.5\%$  ( $\sim 4\%$ ) for redshifts in error by more than  $3000 \text{ km s}^{-1}$  ( $1000 \text{ km s}^{-1}$ ).

In the left panel of figure 1 we show the spatial distribution of quasars in the DESI DR1 sample (green points), compared to the SDSS DR16 footprint `dMdB20` (red curve), and the expected final footprint DESI (blue curve), while the right panel shows the distribution of number of observations for Ly $\alpha$  quasars in DESI DR1. The number of quasars in the DESI DR1 sample is given in table 1, together with the number of quasar spectra covering each of the rest-frame wavelength “regions” described in section 2.4.

The redshift distribution of DESI DR1 and SDSS DR16 quasars are compared in the left panel of figure 2, together with the redshift distribution of Ly $\alpha$  pixels. In the right panel of the same figure we show the contribution of different redshifts to the measurement of the four correlation functions discussed in section 3. Integrating the curve, one finds that 95% of the measurement of the Ly $\alpha$ ×Ly $\alpha$  auto-correlation comes from  $1.96 < z < 2.8$ , while 95% of the Ly $\alpha$ ×QSO cross-correlation comes from  $1.96 < z < 2.95$ . As discussed in section 3,

Catalogue	Num quasars	Too short	Negative continuum	Valid forests
Total	1529530	—	—	—
Tracers ( $z > 1.77$ )	709565	—	—	—
Ly $\alpha$ , region A	531000	83666 (15.8%)	18931 (3.6%)	428403 (80.7%)
Ly $\alpha$ , region B	199449	53162 (26.7%)	8855 (4.4%)	137432 (68.9%)
C III region	1183522	66506 (5.6%)	5753 (0.5%)	1111263 (93.9%)

**Table 1.** Total number of quasars in DESI DR1 that pass our selection criteria. Those with  $z > 1.77$ , referred to as “tracers”, contribute to the measurement of the Ly $\alpha$ ×QSO cross-correlation. The bottom part of the table shows the statistics of number of lines of sight available in each of the three rest-frame wavelength regions (or forests) discussed in section 2.4. Some of the forests are discarded because they do not have 150 valid pixels (too short), or because they do not have a valid continuum fit (negative continuum). These cuts are described in section 2.5.

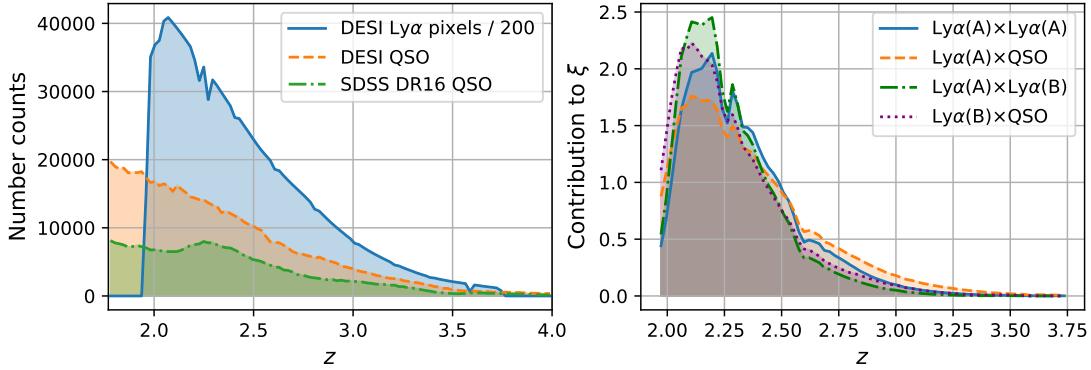


**Figure 1.** Left: expected final DESI (blue) and SDSS-DR16 footprint (red) together with the spatial distribution of DESI DR1 observed quasars (green). For reference we also show the Galactic plane (solid black) and the Ecliptic plane (dotted black). Right: number of observations for Ly $\alpha$  quasars in the DESI DR1 sample.

however, quasars at redshifts as low as  $z = 1.77$  and as high as  $z = 4.16$  can also contribute to the measurement of the cross-correlation.

## 2.2 Damped Ly $\alpha$ systems

The Ly $\alpha$  forest BAO measurement exploits the fact that typical fluctuations in the forest trace matter fluctuations of moderate over density, or under density, along the line of sight. DLA systems are produced by neutral hydrogen concentrations with column density,  $N_{\text{HI}} > 2 \times 10^{20} \text{ cm}^{-2}$ , typically arising in the interstellar medium of high redshift galaxies. Because of the damping wings, each DLA can affect a noticeable fraction of a Ly $\alpha$  forest spectrum (corresponding to thousands of km/s), so even though these systems are rare, it is important to mask the contaminated regions of the spectra, in order to make a robust BAO measurement with uncertainties that can be easily modeled. In DESI we have used two DLA finders, one based on a convolutional neural network (CNN), an algorithm first proposed in [50], and another one based on a Gaussian Process (GP) model [51]. Their adaptation to



**Figure 2.** Left: redshift distribution of quasars in DESI DR1 (orange), compared to the distribution in SDSS DR16 (green) and to the distribution of Ly $\alpha$  pixels (blue, divided by 200). Right: contribution of different redshifts to the four measured correlation functions. In particular, we show the sum of weights used in eqs. (3.5) and (3.6) as a function of redshift, for large transverse separations (to reduce the biasing effect from the clustering of background quasars) and  $r_{\parallel} = 0$  (the contribution varies as a function of  $r_{\parallel}$ , especially for quasar cross-correlations). As described in section 3, we measure Ly $\alpha$  correlations in two different rest-frame wavelength regions of the quasar spectra (A and B).

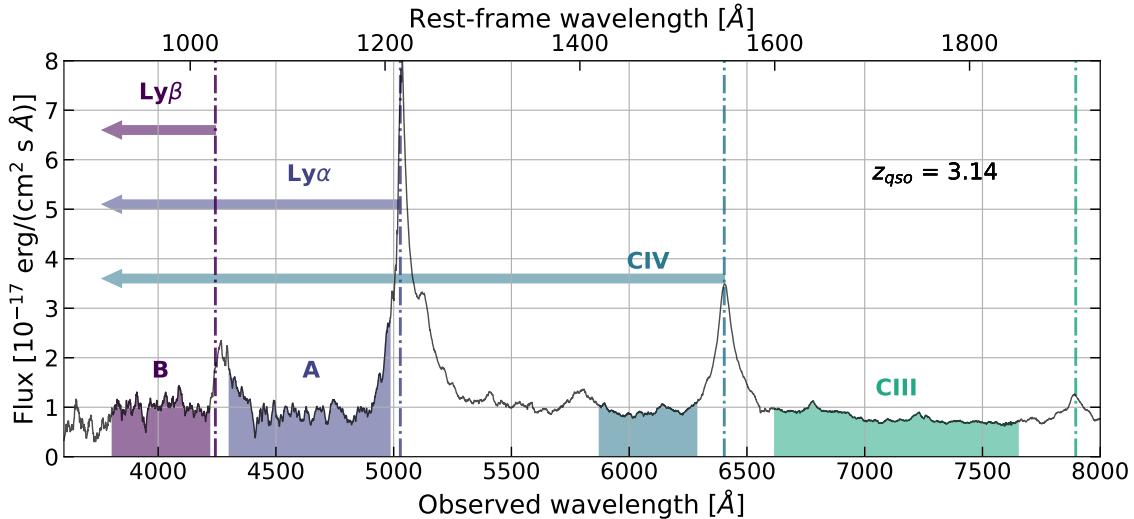
DESI and performance in DESI’s simulated spectra is presented in [52]. The performance of both algorithms in the DESI DR1 dataset will be presented in an upcoming paper.

For the purpose of this work, we construct a DLA catalog based on the combined result of the two DLA finders. We select DLAs found by the two algorithms with a threshold probability of 50%, as done in [30, 31]. However, in order to ensure the purity of the DLA catalog we also limit it to DLAs detected in spectra with signal-to-noise (SNR) larger than 3 (mean SNR within the Ly $\alpha$  forest region). The DLAs included in the final catalog are masked in the spectra before computing the forest fluctuations (see section 2.4) and the contamination from undetected ones is taken into account in the modelling of the correlation functions (see section 4.6).

### 2.3 Broad Absorption Line quasars

Numerous studies of large quasar samples have identified broad absorption line (BAL) troughs in 10–30% of quasars [e.g., 53, 54]. While these features are most commonly seen as blue shifted absorption relative to the CIV emission feature, the absorption is also associated with many other features, including several that contribute absorption in the Ly $\alpha$  forest region [55]. Significant BAL features associated with CIV and other strong emission lines are also expected to introduce some systematic redshift biases and increase redshift errors for a subset of BAL quasars [56]. We consequently need to identify and characterize BALs in the DESI quasar sample to mitigate their impact on the cosmological analysis.

We catalog BALs in the DESI quasar sample following the same approach employed by [57] for the DESI Early Data Release (EDR). This approach fits a series of continuum components to each quasar, and iteratively identifies and masks regions that meet the Absorptivity Index [AI, 58] and Balnicity Index [BI, 59] criteria. For the DR1 BAL catalog, we use the same components developed by [60]. The catalog includes measurements of the



**Figure 3.** QSO spectrum from the first year of DESI data at redshift  $z = 3.14$  (TargetID = 39627581225438176). The region B is highlighted in purple. The region A is highlighted in indigo. The CIV and CIII regions are highlighted in various shades of green. While there is almost no CIII absorption, the CIV absorption spans leftward of the CIV doublet, contaminating the Ly $\alpha$  regions A and B. The Ly $\alpha$  absorption extends into region B. For better visualization, we have chosen a relatively high signal-to-noise spectrum and we have smoothed it by averaging the pixels in groups of 25.

velocity range of each trough, AI and BI values, and other quantities measured relative to the CIV and SiIV features, similar to the DESI EDR catalog. This catalog is used in our analysis to mask the corresponding locations of BAL features before computing the Ly $\alpha$  forest fluctuations (see section 2.4).

## 2.4 Catalog of the Ly $\alpha$ forest

We discuss here the subset of quasar spectra that we use to study the fluctuations in the Ly $\alpha$  forest. We consider two different catalogues that correspond to Ly $\alpha$  absorption in two different regions of the quasar spectra. Region “A” is defined as the rest-frame wavelength range from 1040 to 1205 $\text{\AA}$ ,<sup>3</sup> while region “B” covers the range from 920 to 1020 $\text{\AA}$ . Pixels in the B region are also affected by absorption from other Lyman lines, since it extends beyond the Ly $\beta$ , Ly $\gamma$  and Ly $\delta$  emission lines, and therefore it require special treatment.<sup>4</sup> See figure 3 for an illustration of these regions in a quasar observed with DESI.

In addition to the rest-frame wavelength cut, we also restrict our observed wavelength range from 3600 $\text{\AA}$  to 5772 $\text{\AA}$ . The lower bound is the minimum wavelength provided by the pipeline; the upper bound corresponds to the middle of the overlap region between the blue and the red arms of the spectrographs.<sup>5</sup> These cuts in observer-frame and rest-frame wavelength limit the redshift range of the background quasars from  $z = 2.1$  to  $z = 4.4$  for

<sup>3</sup>Following [30], we extend the A region to 1205 $\text{\AA}$ , going a bit further than the 1200 $\text{\AA}$  limit used in [dMdB20](#).

<sup>4</sup>Note that these regions were referred to as Ly $\alpha$  and Ly $\beta$  regions in [dMdB20](#).

<sup>5</sup>We could extend the analysis to pixels in the red spectrograph, but the gain would be marginal given the limited number of high redshift quasars.

the region A and  $z = 2.6$  to  $z = 5.1$  for the region B. The number of quasars available for these catalogues is given in table 1.

Once we have the initial set of Ly $\alpha$  forests, we apply four different masks to remove bad pixels and astrophysical contaminants. The first mask comes from the data reduction pipeline, which flags bad pixels in the spectra [37]. These are usually caused by cosmic rays or defects in the CCD cameras. As shown in [30], the fraction of pixels masked as a function of observed wavelength is generally constant and below 1%. The second mask consists of a set of narrow windows in observed wavelength: 3933.0 to 3935.8Å, 3967.3 to 3971.0Å, and 5570.5 to 5586.5Å (see [30]). This mask aims to remove galactic absorption not observed in the calibration stars (first two lines) and residual sky lines from the sky subtraction models [37].

The other two masks aim at removing the effects of astrophysical contaminants. In particular, we mask DLAs (see section 2.2) and BALs (see section 2.3). For DLAs, we mask the regions of the Ly $\alpha$  forest where the presence of a DLA decreases the transmitted flux by 20% or more. We then correct the remaining DLA wings using a Voigt profile. For BALs, we follow the procedure described in [61], which we summarize here. We mask the expected locations of all potential BAL features associated with, irrespective of whether or not the absorption is apparent. These include Ly $\alpha$ , NIV, CIII, SiIV, and PV in region A, and OVI, OI, Ly $\beta$ , Ly $\gamma$ , NIII and Ly $\delta$  in region B. Because of this, the mask may remove some path length that is unaffected by BALs. However, we note that there is a net increase in path length compared to completely discarding these spectra as done in previous analyses [e.g. 27]. The paper by [62] investigates the impact of other masking strategies on the BAO analysis.

At this stage, we discard forests that are too short (defined as having less than 150 valid pixels corresponding to a length of 120Å). This is necessary as we later fit the unabsorbed quasar continua to extract the transmitted flux field (see section 2.5), and having forests that are too short interferes with the continuum fitting procedure. The number of forests lost because of this in each of the samples is given in table 1.

Finally, to check for unaccounted-for calibration residuals, we compute the mean transmitted flux fraction in a quasar rest-frame region where there is no Ly $\alpha$  absorption. As in [30], we use the CIII region ( $\lambda_{\text{rf}} \in [1600, 1850]$  Å). We note that previous analysis from eBOSS used longer rest-frame wavelengths, but [30] showed that the measured calibration residuals are the same irrespective of the metal region used to measure them, and there are more spectra available in the C III region. We use this small correction (less than 5% in the relevant wavelengths range) to re-calibrate our fluxes and instrumental noise estimates.

## 2.5 Continuum fitting

The initial step in our data analysis is to compute the transmitted flux field  $\delta_q(\lambda)$ . In general terms, it is defined as the ratio of the observed flux to the expected flux:

$$\delta_q(\lambda) = \frac{f_q(\lambda)}{\bar{F}(\lambda) C_q(\lambda)} - 1, \quad (2.1)$$

where  $\bar{F}(\lambda)$  is the mean transmission and  $C_q$  is the unabsorbed quasar continuum. The sub-index  $q$  indicates that these are line-of-sight (quasar) dependent. We label the process of estimating their product as *continuum fitting*.

The continuum fitting procedure we use is described in detail in [30] and we summarize it here. The product  $\bar{F}C_q(\lambda)$  is taken to be a universal function of the quasar  $\bar{C}(\lambda_{\text{rf}})$ , corrected by a first degree polynomial in  $\Lambda \equiv \log \lambda$  to account for quasar diversity and the redshift evolution of  $\bar{F}(z)$ :

$$\bar{F}(\lambda)C_q(\lambda) = \bar{C}(\lambda_{\text{rf}}) \left( a_q + b_q \frac{\Lambda - \Lambda_{\text{min}}}{\Lambda_{\text{max}} - \Lambda_{\text{min}}} \right). \quad (2.2)$$

Here, the universal function  $\bar{C}(\lambda_{\text{rf}})$  is computed to be the weighted mean (inverse variance weighting) of all the lines of sight entering the analysis. The parameters  $a_q$  and  $b_q$  are estimated by maximizing the likelihood function

$$2 \ln \mathcal{L} = - \sum_i \frac{(f_i - \bar{F}C_q(\lambda_i, a_q, b_q))^2}{\sigma_q^2(\lambda_i, a_q, b_q)} - \sum_i \ln \sigma_q^2(\lambda_i, a_q, b_q), \quad (2.3)$$

where  $\sigma_q^2$  is the total variance of the data, including the contribution from the noise estimated from the pipeline ( $\sigma_{\text{pip},q}^2$ ), modulated by a function  $\eta_{\text{pip}}(\lambda)$  to account for small inaccuracies in the noise estimation, and the intrinsic variance of the Ly $\alpha$  forest,  $\sigma_{\text{LSS}}^2$ , multiplied by  $(\bar{F}C_q)^2$ :

$$\sigma_q^2(\lambda) = \eta_{\text{pip}}(\lambda) \sigma_{\text{pip},q}^2(\lambda) + \sigma_{\text{LSS}}^2(\lambda) (\bar{F}C_q)^2(\lambda). \quad (2.4)$$

We keep the original wavelength bin size of 0.8Å when computing the  $\delta_q(\lambda)$  following the study of [30]. We note that the continuum fitting procedure distorts the  $\delta_q$  field, since line-of-sight fluctuations on scales comparable to the length of a given forest will be suppressed when fitting the  $(a_q, b_q)$  parameters. This is taken into account with the definition of a projected flux transmission field in section 3.1 and in the correlation function model in section 4.3. We perform the continuum fit in each of the analysed regions independently. This means that we have up to 3 independent sets of  $(a_q, b_q)$  per quasar as each quasar can be used in the Ly $\alpha$  regions A and B, and in the calibration region.

Occasionally, the best-fit values of  $a_q$  and  $b_q$ , combined with the shape of  $\bar{C}(\lambda_{\text{rf}})$  result in a negative estimated product  $\bar{F}C_q(\lambda)$ . Whenever this happens for at least one pixel, we deem the continuum fit as problematic and discard the entire region. The number of forests lost because of this in each of the samples is given in table 1. Note that, because the  $a_q$  and  $b_q$  are fitted independently in each of the regions of interest, a particular spectrum might be removed from one of the regions but kept in another.

### 3 Measurement of correlations

In section 2 we have described the three cosmological datasets that we use in our BAO measurement: the catalog of quasar positions (angles and redshifts) and the catalogs of Ly $\alpha$  fluctuations in the rest-frame regions A and B. There are six (3x2) possible correlations that we could use, but following [dMb20](#) we focus on the four correlations that have higher signal to noise: we ignore the auto-correlation of quasars, and the auto-correlation of Ly $\alpha$  fluctuations in the B region. For convenience, in some sections we group the correlations in two subsets: we use the term *auto-correlation* to refer to the combination of the auto-correlation of the Ly $\alpha$  fluctuations in region A, Ly $\alpha$ (A)  $\times$  Ly $\alpha$ (A), and the correlation of Ly $\alpha$

fluctuations in region A with the Ly $\alpha$  fluctuations in region B, Ly $\alpha$ (A)  $\times$  Ly $\alpha$ (B);<sup>6</sup> we use the term *cross-correlation* to refer to the correlation of quasar positions with the Ly $\alpha$  fluctuations in region A, Ly $\alpha$ (A)  $\times$  QSO, and in region B, Ly $\alpha$ (B)  $\times$  QSO.

In this section we summarise the methodology used to measure the different correlations and their covariance matrix, and we refer the reader to [31] for more details. We measure the correlations with the software `picca`<sup>7</sup> [73] that was developed originally for the Ly $\alpha$  analysis of BOSS and eBOSS data, and that was also used to estimate the Ly $\alpha$  fluctuations. The only relevant difference with respect to [31] is that we now include the cross-covariance of the different correlations, which were considered independent in previous analyses. We discuss this in section 3.3.

We measure the correlations in bins of comoving separation along ( $r_{\parallel}$ ) and across ( $r_{\perp}$ ) the line of sight, computed from the angular and redshift separations using a fiducial cosmology based on the best-fit flat  $\Lambda$ CDM model from Planck 2018 [2] (see table 2). We present a single measurement of the correlations averaged over a wide range of redshifts (see figure 2). This is motivated by the fact that the BAO scale in comoving coordinates does not vary much with redshift if the true cosmic expansion is not very far from the fiducial one for those redshifts,<sup>8</sup> and that considering several redshift bins would make the BAO peak less significant and possibly more difficult to fit [74]. As discussed in section 4, we model the correlations at an effective redshift ( $z_{\text{eff}} = 2.33$ ) and report the BAO measurement at that redshift. In our fiducial model and at our effective redshift, the BAO scale corresponds to an angular separation of 1.46 degrees and an observed wavelength separation of 141Å.

### 3.1 Measurement of the auto-correlation

The correlation function measurement relies on a fiducial cosmology to convert the angular separations and redshifts into comoving separations. We use the same approach as in [31] and earlier studies. We first define the longitudinal and transverse separations ( $r_{\parallel}, r_{\perp}$ ) for a pair of measurements ( $i, j$ ) at redshifts ( $z_i, z_j$ ) and angular separation  $\theta_{ij}$  as

$$\begin{aligned} r_{\parallel} &= [D_C(z_i) - D_C(z_j)] \cos(\theta_{ij}/2) , \\ r_{\perp} &= [D_M(z_i) + D_M(z_j)] \sin(\theta_{ij}/2) , \end{aligned} \quad (3.1)$$

where  $D_C(z)$  is the comoving distance and  $D_M(z)$  the angular (or transverse) comoving distance. For our fiducial cosmology with  $\Omega_k = 0$ , they are identical.

The estimator for the correlation function of the transmitted flux field is a simple weighted average. For this purpose, we define the following weight for the flux decrement  $\delta_q(\lambda)$ ,

$$w_q(\lambda) = \left( \frac{1+z_{\lambda}}{1+z_0} \right)^{\gamma_{\alpha}-1} \left[ \eta_{\text{pip}}(\lambda) \left( \frac{\sigma_{\text{pip},q}(\lambda)}{FC_q(\lambda)} \right)^2 + \eta_{\text{LSS}} \sigma_{\text{LSS}}^2(\lambda) \right]^{-1} . \quad (3.2)$$

<sup>6</sup>While this is technically the cross-correlation of two disjoint datasets, we include it in the *auto-correlation* since both datasets contain pixels with Ly $\alpha$  fluctuations, and we use the same model as in the Ly $\alpha$ (A)  $\times$  Ly $\alpha$ (A) correlation.

<sup>7</sup><https://github.com/igmhub/picca>. We used version v9.0.0. We also acknowledge the use of the following packages: `numpy` [63], `scipy` [64], `astropy` [65–67], `mpi4py` [68], `healpy` [69], `matplotlib` [70], `GetDist` [71], `numba` [72], and `fitsio`, <https://github.com/esheldon/fitsio>.

<sup>8</sup>If the fiducial cosmology was significantly different than the truth, the BAO peak would be smeared when averaging over a wide redshift range. However, as shown in the appendix B of [33], varying the fiducial cosmology by  $1\sigma$  (in terms of  $\alpha_{\perp}, \alpha_{\parallel}$ ) causes a bias smaller than  $0.1\sigma$ .

Parameter	Planck (2018) cosmology (TT,TE,EE+lowE+lensing)
$\Omega_m h^2 =$	0.14297
$+ \Omega_c h^2$	0.12
$+ \Omega_b h^2$	0.02237
$+ \Omega_\nu h^2$	0.0006
$h$	0.6736
$n_s$	0.9649
$10^9 A_s$	2.100
$\Omega_m$	0.31509
$\Omega_r$	7.9638e-05
$\sigma_8(z=0)$	0.8119
$r_d$ [Mpc]	147.09
$r_d$ [ $h^{-1}$ Mpc]	99.08
$D_H(z_{\text{eff}} = 2.33)/r_d$	8.6172
$D_M(z_{\text{eff}} = 2.33)/r_d$	39.1879
$f(z_{\text{eff}} = 2.33)$	0.9703

**Table 2.** Parameters of the flat- $\Lambda$ CDM cosmological model used in the analysis, both to compute comoving separations and in the modelling (described in section 4). The first part of the table gives the cosmological parameters, the second part gives derived quantities used in this paper.

The right hand term is the inverse of the variance of the flux decrement (eq. (2.4)), but with an additional scaling term,  $\eta_{\text{LSS}}$ , that increases the contribution of the large scale structure variance compared to the pipeline noise term. This correction was introduced by [30] to improve the precision of the correlation function measurement.<sup>9</sup> Our estimator is sub-optimal because it does not take into account the correlations between neighboring flux decrement values caused by the large scale structure. The adjustment of weights with  $\eta_{\text{LSS}}$  is a simple way to avoid over-weighting pixels that have a high signal to noise ratio but correlated information content. We use a value of  $\eta_{\text{LSS}} = 7.5$ , which was found to minimise the covariance matrix of the Ly $\alpha$  auto-correlation given our wavelength bin size of 0.8 Å [30]. The left hand term is a scaling factor to account for the variation of the amplitude of the correlation function with redshift. It is the optimal term for the measurement of the shape of the correlation function. We set the bias evolution index  $\gamma_\alpha = 2.9$  as in previous studies, following [75] (the value of  $z_0$  does not affect the results as it cancels in the estimator of the correlation function).

In order to estimate the Ly $\alpha$  fluctuations, we have divided the observed flux by a model of the continuum of each quasar (see eq. (2.1)). As noted in section 2.5, the parameters  $(a_q, b_q)$  of this model have been fitted using all the Ly $\alpha$  pixels in the spectrum and this makes our estimated fluctuation at a given pixel ( $\delta_i$ ) dependent on the flux of the other pixels in the spectrum, and therefore on their true Ly $\alpha$  fluctuation ( $\delta_j^t$ ) [8, 76]. Following earlier

<sup>9</sup> $\eta_{\text{LSS}}$  was referred to as  $\sigma_{\text{mod}}^2$  in [30].

work [21, 22], we linearise this relation and approximate the distortion<sup>10</sup> as  $\delta_i = \eta_{ij}^{\text{cont}} \delta_j^t$ , with  $\eta^{\text{cont}}$  defined as

$$\eta_{ij}^{\text{cont}} = \delta_{ij}^K - \frac{w_j^{\text{cont}}}{\sum_k w_k^{\text{cont}}} - \frac{w_j^{\text{cont}} \Lambda_i \Lambda_j}{\sum_k w_k^{\text{cont}} \Lambda_k^2}, \quad (3.3)$$

where  $\delta_{ij}^K$  is the Kronecker delta function,  $w_q^{\text{cont}} = (\bar{F}C_q)^2 / \sigma_q^2$  are the weights used in continuum fitting (eq. (2.3)) and  $\Lambda_i \equiv \log(\lambda_i) - \sum_k w_k^{\text{cont}} \log(\lambda_k) / \sum_k w_k^{\text{cont}}$ .

It is clear from eq. (3.3) that during continuum fitting the weighted mean and the linear trend over each Ly $\alpha$  forest are set to zero. In order to make sure that these are also zero when using the weights  $w$  defined in eq. (3.2), we apply a similar linear projection to the measured fluctuations  $\delta_j$  and define a projected field  $\tilde{\delta}_i = \eta_{ij} \delta_j$ , where  $\eta$  is equivalent to  $\eta^{\text{cont}}$  but using the weights  $w$  instead of  $w^{\text{cont}}$  in eq. (3.3). One can show that, to first order in  $\delta^t$ ,  $\eta$  is also the relation between the projected field and the true fluctuation:<sup>11</sup>

$$\tilde{\delta}_i = \eta_{ij} \delta_j = \eta_{ij} \eta_{jk}^{\text{cont}} \delta_k^t = \eta_{ij} \delta_j^t. \quad (3.4)$$

As a result we do not need to know the exact values of  $\eta_{ij}^{\text{cont}}$ , but only those of  $\eta_{ij}$ . In section 4.3 we will use this linear projection operator  $\eta$  to model the correlations of the projected field  $\tilde{\delta}_q(\lambda)$ .

The correlation function is measured in bins of  $4 h^{-1}$  Mpc ranging from 0 to  $200 h^{-1}$  Mpc along both  $r_{\parallel}$  and  $r_{\perp}$  for a total of 2500 bins. Calling  $M$  a two-dimensional bin of comoving separation, the estimator of the correlation function averaged in the separation bin  $M$  is the weighted average:

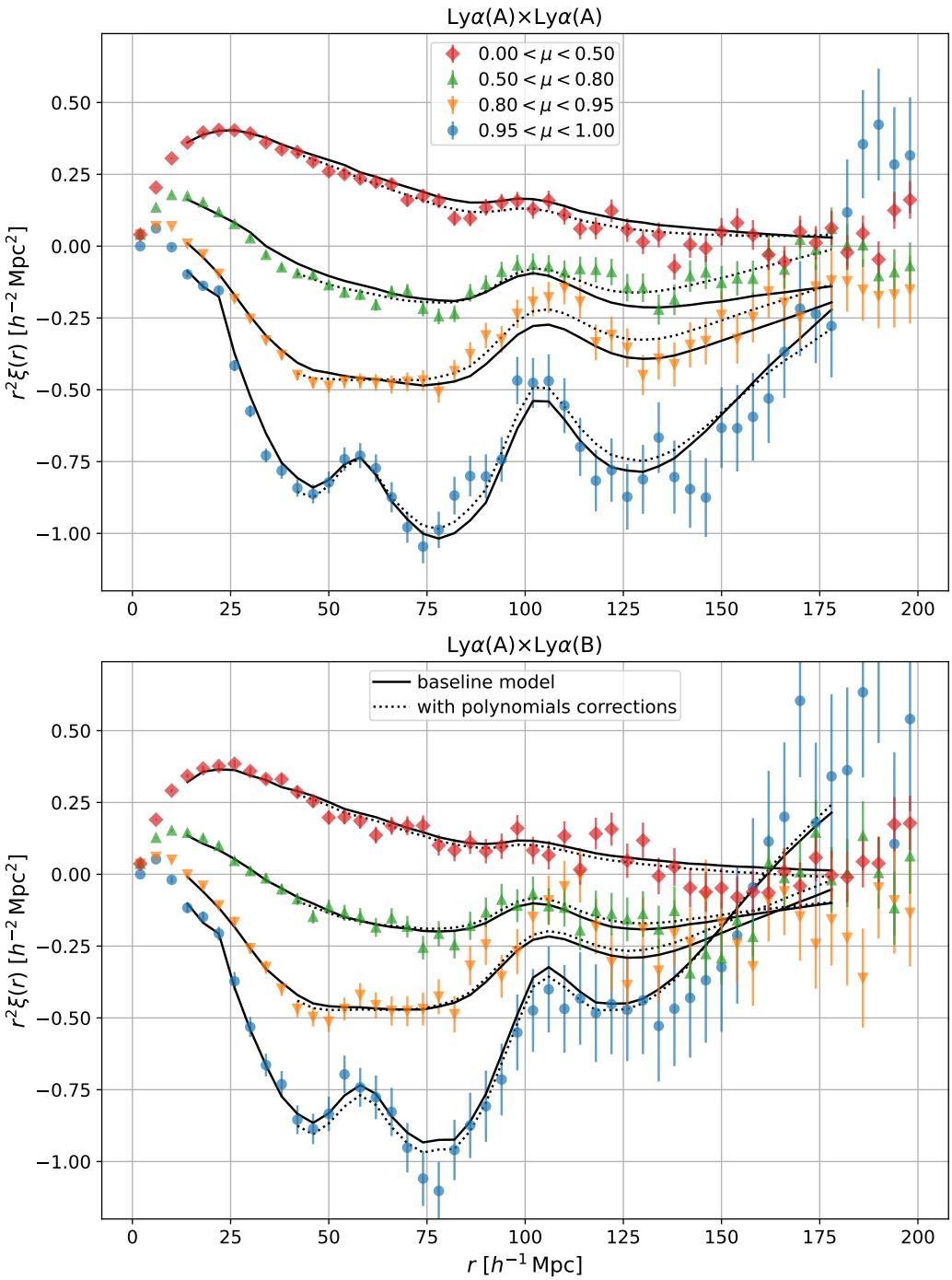
$$\xi_M = \sum_{(i,j) \in M} w_i w_j \tilde{\delta}_i \tilde{\delta}_j / \sum_{(i,j) \in M} w_i w_j \quad (3.5)$$

where  $(i,j) \in M$  are all the pairs of projected transmitted flux measurements  $\tilde{\delta}_i$  and  $\tilde{\delta}_j$  from quasar lines of sight separated by an angle  $\theta_{ij}$  and at redshift (or wavelength)  $z_i$  and  $z_j$ , such that their comoving separation is in  $M$ . We only consider pairs from different quasar spectra to avoid the contribution of correlated residuals within a spectrum caused by continuum fitting errors.

In figure 4 we show the measurement of the Ly $\alpha$  forest auto-correlation when using pixels from region A (Ly $\alpha$ (A)  $\times$  Ly $\alpha$ (A), top panel) and when correlating pixels from region A with pixels in region B (Ly $\alpha$ (A)  $\times$  Ly $\alpha$ (B), bottom panel). We show the measurement as a function of the total separation  $r = (r_{\parallel}^2 + r_{\perp}^2)^{1/2}$  in wedges defined by a range of the cosine between the orientation of the pair and the line-of-sight,  $\mu = r_{\parallel}/r$ . This is obtained by resampling the original data, which has the form of a rectangular 2D grid of  $(r_{\parallel}, r_{\perp})$  bins, into  $(r, \mu)$  bins. We evaluate the uncertainties in those bins using the covariance matrix, the computation of which is described in section 3.3. The resampling results in additional correlation between neighboring  $(r, \mu)$  bins that share fractions of the original bins. We note

<sup>10</sup>Here we ignore uncorrelated sources of noise, like random continuum errors or instrumental noise, since these would not bias the measurement of the 3D correlations.

<sup>11</sup>Picture a scatter plot of  $(x, y)$  points from which one subtracts to  $y$  the mean and slope as a function of  $x$ . The result is obviously independent from any prior subtraction of any mean and slope.



**Figure 4.** Measured Ly $\alpha$  auto-correlation when using pixels from region A (top, colored markers) and when correlating pixels from region A with pixels from region B (bottom), along with the best fit model (solid black curves), described in section 4. The different colors and markers correspond to different orientations with respect to the line-of-sight, with blue correlations being close to the line-of-sight  $0.95 < \mu < 1$ . The dotted curves show the best fit model with additive polynomial corrections (see section 6.4.4).

that this resampling into wedges is performed for display purpose only and is not used for the fit, which is realized on the original rectangular grid.

### 3.2 Measurement of the cross-correlation

With the same notations as in the previous section, we use the following estimator [76] for the quasar Lyman- $\alpha$  cross-correlation in a separation bin  $M$ :

$$\xi_M = \sum_{(i,j) \in M} w_i w_j^Q \tilde{\delta}_i / \sum_{(i,j) \in M} w_i w_j^Q. \quad (3.6)$$

Here  $(i,j) \in M$  stands for pairs made of a projected transmitted flux measurement  $\tilde{\delta}_i$  at a redshift  $z_i$  along a quasar line of sight, and another quasar  $j$  at another redshift  $z_j$  separated by an angle  $\theta_{ij}$  from the first one, such that their comoving separation is in the bin  $M$  (once the redshifts and  $\theta_{ij}$  are converted to comoving separation using equation (3.1)). We use for quasars the weights

$$w_j^Q = [(1+z_Q)/(1+z_0)]^{\gamma_Q-1} \quad (3.7)$$

where  $z_Q$  is the quasar redshift, and we choose  $\gamma_Q = 1.44$  which follows closely the measured bias evolution of quasars [77].

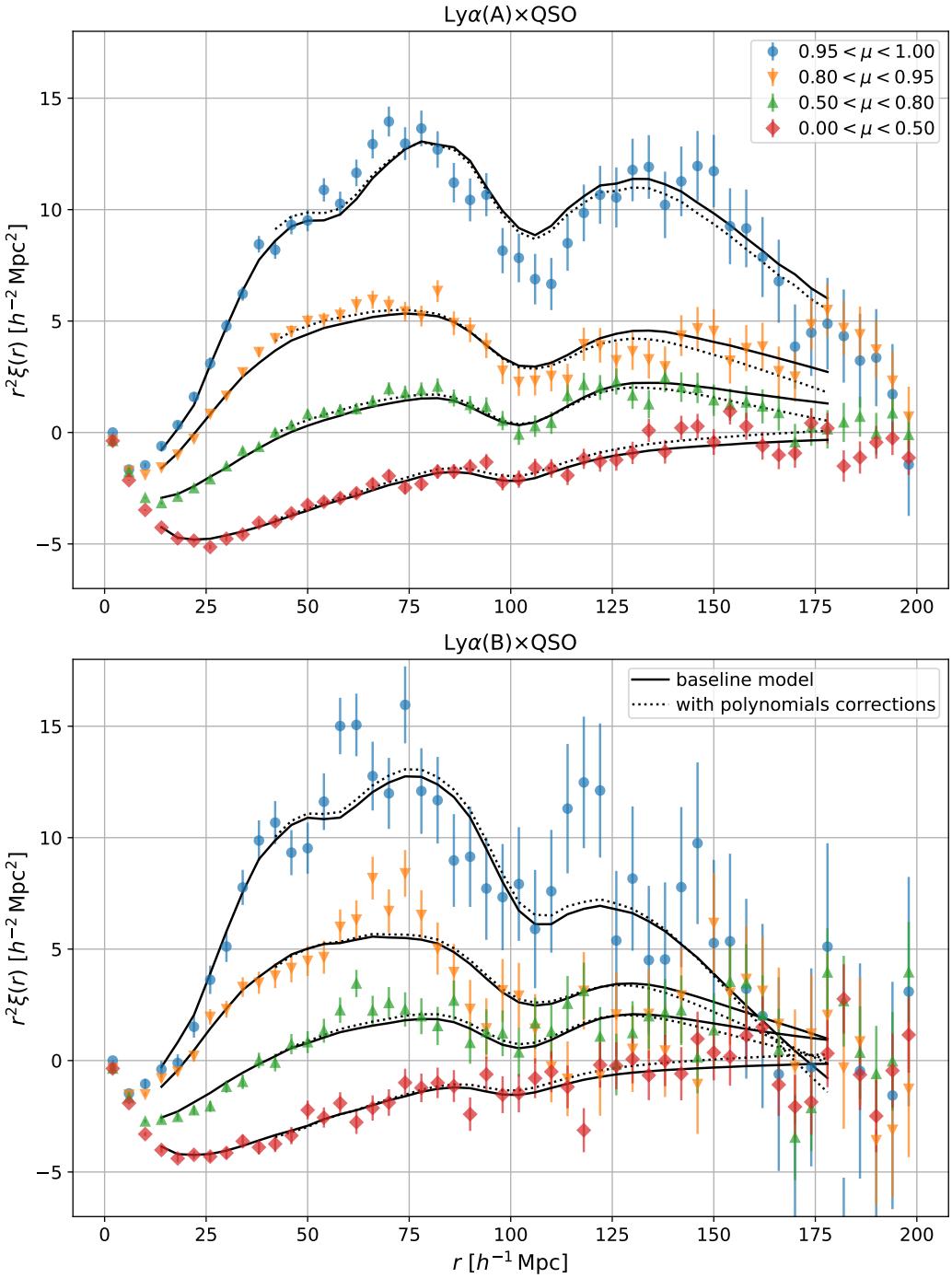
We use the same bin size of  $4 h^{-1}$  Mpc for the cross-correlation but we differentiate positive and negative longitudinal separation, giving us a range from  $-200$  to  $+200 h^{-1}$  Mpc for  $r_{\parallel}$  because the cross-correlation is asymmetric. We define the sign of  $r_{\parallel}$  such that  $r_{\parallel} < 0$  for pairs where the quasar is behind the Ly $\alpha$  transmitted flux measurement. We have 5 000 cross-correlation bins.

In figure 5 we show the measurement of the cross-correlation of quasars with Ly $\alpha$  pixels in region A (Ly $\alpha$ (A)  $\times$  QSO, top panel) and with Ly $\alpha$  pixels in region B (Ly $\alpha$ (B)  $\times$  QSO, bottom panel).

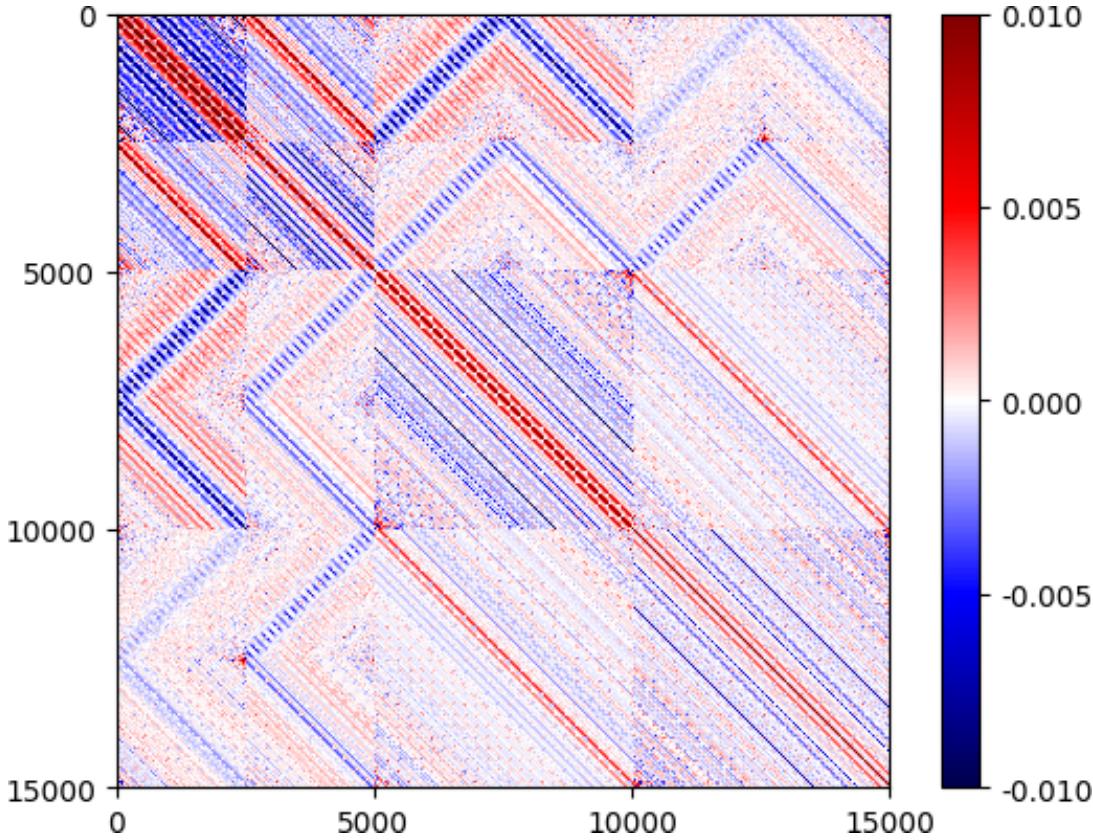
### 3.3 Covariance matrix

We use the method described in more detail in [dMdB20](#) to compute the covariances. In brief, the correlation function is first measured independently in sub-samples defined by `HEALpix` pixels on the sky. Each sub-sample correlation is saved with its weights  $W_M$  in each bin  $M$ , which are denominators in eqs. (3.5) and (3.6). We do not lose or double-count pairs because each possible pair is assigned a unique sub-sample. The combined correlation function is simply the weighted mean of the sub-sample correlations, and its covariance is determined by replacing the unknown covariance of each sub-sample by the square of its difference with the mean. We ignore the cross-covariance between sub-samples, which is negligible for our scales of interest given the size of a `HEALpix` pixel of about  $(250 h^{-1}$  Mpc) $^2$  at  $z \sim 2.3$  for our choice of `NSIDE=16`. The correlation is measured in 1028 pixels, with more than 400 valid forests per pixel on average. In figure 11 we show that using `NSIDE=32` instead has a negligible impact on the BAO parameters.

This noisy estimate of the covariance  $C$  is then smoothed. We replace all the non-diagonal elements of the correlation matrix  $\text{Corr}_{MN} \equiv C_{MN}/(C_{MM}C_{NN})^{1/2}$  in which indices correspond to the same differences  $|r_{\parallel}(M) - r_{\parallel}(N)|$  and  $|r_{\perp}(M) - r_{\perp}(N)|$  by their average.



**Figure 5.** Measured Ly $\alpha$ ×QSO cross-correlation functions in region A (top, colored markers) and region B (bottom) along with the best fit model (solid black curves), described in section 4. The different colors and markers correspond to different orientations with respect to the line-of-sight, with blue correlations being close to the line-of-sight  $0.95 < \mu < 1$ . The dotted curves show the best fit model with additive polynomial corrections (see section 6.4.4).



**Figure 6.** Global correlation matrix for the four 2pt functions included in this analysis, as measured from the scatter between correlations measured in more than 1000 HEALPix pixels, after applying the smoothing discussed in the main text. The first block of  $2500 \times 2500$  in the top left corresponds to the correlation matrix of the  $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{A})$  measurement, while the second block of the same size corresponds to the  $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{B})$  one. The third, larger block of  $5000 \times 5000$  corresponds to the  $\text{Ly}\alpha(\text{A}) \times \text{QSO}$  cross-correlation, with positive and negative values of  $r_{\parallel}$ , and the last block (bottom right) is the correlation matrix of the  $\text{Ly}\alpha(\text{B}) \times \text{QSO}$  measurement. It is clear in that cosmic variance (in the form of off-diagonal stripes in red and blue) is relevant, even thought it is only detected at the sub-percent level. Note that we only have strong correlations ( $> 10\%$ ) for bins with the same value of  $r_{\perp}$  and neighbouring value of  $r_{\parallel}$ , and that we limit the range of the color scale to 1% for visualization purposes.

This method has proven to be a good approximation of the covariance when compared with other methods, like a Gaussian covariance computed with the Wick expansion (see [20] and appendix C of [dMdB20](#)), and it is discussed in detail in the companion paper [33].

We measure with this sub-sampling technique the covariance of the full data set composed of the four correlation functions,  $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{A})$ ,  $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{B})$ ,  $\text{Ly}\alpha(\text{A}) \times \text{QSO}$ , and  $\text{Ly}\alpha(\text{B}) \times \text{QSO}$ . Previous works (including [dMdB20](#)) measured the covariances of each correlation function in isolation, ignoring the cross-covariances between them. Indeed, we show in section 6.2.2 that ignoring the cross-covariance between the correlation functions results in a 10% underestimate of uncertainties on the BAO parameters. The new correlation matrix is shown in figure 6. This full covariance estimation represents the most important change in the methodology from previous analyses of the 3D correlations in the Ly $\alpha$  forest.

## 4 Modelling of correlations

In this section we describe how the correlations are modeled. The approach is similar to the one used in previous BOSS and eBOSS analyses (see [dMdB20](#) for the latest results from eBOSS) except for few methodological changes that are summarised below, and that we describe in more detail in companion papers [31–33]. We now use the software [Vega](#),<sup>12</sup> which is an improved version of the [picca-fitter2](#) software used in these previous  $\text{Ly}\alpha$  BAO analyses.

From the point of view of this analysis, our model has 2 important (BAO) parameters and 15 nuisance parameters that we marginalize over. We start in sections 4.1–4.3 with a simplified version of the model that will allow us to introduce some of the key aspects of a  $\text{Ly}\alpha$  BAO analysis, before describing the modelling of contaminants in sections 4.4–4.8. The priors, best-fit values and uncertainties for the nuisance parameters are presented in appendix B.

### 4.1 Isolating the BAO information

We start by building a model for the (anisotropic) large-scale power spectrum of fluctuations in the  $\text{Ly}\alpha$  forest, based on linear perturbation theory. We use the linear power spectrum of matter fluctuations for our fiducial cosmology (see table 2) evaluated at our effective redshift (see below), and linear bias ( $b_F$ ) and redshift-space distortion ( $\beta_F$ ) parameters. When modelling the cross-correlation with quasars, we introduce another linear bias parameter ( $b_Q$ ) and we follow [78] to model the linear redshift-space distortions of quasars with  $\beta_Q = f/b_Q$ , where  $f$  is the logarithmic growth rate.<sup>13</sup>

We then compute its inverse Fourier transform to obtain a model for the correlation function  $\xi(r_\perp, r_\parallel)$  that we can compare to our measurement. In our fiducial cosmology, there is an excess correlation (the BAO *peak*)<sup>14</sup> at around  $100 h^{-1} \text{Mpc}$ . We then introduce two scaling parameters,  $\alpha_\perp$  and  $\alpha_\parallel$ , which multiply  $r_\perp$  and  $r_\parallel$  respectively in our model, and we vary them in order to better match the BAO peak seen in the measured correlations.

In order to make sure that we only extract BAO information from the fits, we decompose the model of the correlations into a *peak* and a *smooth* component following [15], and the scaling parameters ( $\alpha_\perp, \alpha_\parallel$ ) are only applied to the peak component.<sup>15</sup> Following [15] again, we apply a Gaussian smoothing to the peak component in order to model the non-linear broadening of the BAO feature caused by non-linear growth of structure.

### 4.2 Redshift evolution

When computing the binned correlations in eqs. (3.5) and (3.6), we use the same weights to compute the mean separations ( $r_\perp, r_\parallel$ ) and to compute the mean redshift ( $z$ ) of each bin. We evaluate the model at these coordinates. The redshift evolution of the model is captured by the linear growth of the matter power spectrum ( $\sigma_8(z)$ ), the logarithmic growth rate ( $f(z)$ ), and the redshift evolution of the bias parameters. We define our parameters at an effective redshift ( $z_{\text{eff}} = 2.33$ , see below), and model the redshift evolution of the bias parameters

<sup>12</sup><https://github.com/andreicuceu/vega>. We used version v1.0.0.

<sup>13</sup>Note that the same relation does not apply to the  $\text{Ly}\alpha$  forest, and  $\beta_F$  is an independent parameter [79, 80].

<sup>14</sup>Note that in the  $\text{Ly}\alpha \times \text{QSO}$  cross-correlation we expect a *BAO trough*, instead of a *BAO peak*.

<sup>15</sup>We validate this in section 6.4.4, where we present an alternative analysis where we add up to 48 extra free parameters describing a flexible *broadband* component.

with a power law,  $b(z) = b(z_{\text{eff}}) [(1+z)/(1+z_{\text{eff}})]^{\gamma}$ . Following [27, 31], we use  $\gamma_Q = 1.44$  to describe the redshift evolution of the quasar bias,  $\gamma_\alpha = 2.9$  for bias of the Ly $\alpha$  forest, and we assume that the RSD parameter of the Ly $\alpha$  forest ( $\beta_\alpha$ ) does not vary with redshift.

In order to estimate the effective redshift of our BAO measurement, we compute the mean of the redshifts of each correlation bin in the range  $80 h^{-1} \text{ Mpc} < r < 120 h^{-1} \text{ Mpc}$ , weighted with their inverse variance. We do this separately for the auto-correlation ( $z_{\text{eff}} = 2.339$ ) and for the cross-correlation ( $z_{\text{eff}} = 2.325$ ), and we then compute a simple mean of those two values for the combined BAO measurement (rounded to two decimal values) to obtain the effective redshift of our BAO measurement ( $z_{\text{eff}} = 2.33$ ).

### 4.3 Distortion matrix

As discussed in section 3, the distortions introduced with the continuum fitting have led us to the use of the projected field  $\tilde{\delta}$  defined by eq. (3.3). We must therefore fit the correlations of the projected field  $\tilde{\delta}$  with a model that has suffered the same projection. This is performed with the ‘‘distortion matrix’’ formalism introduced in [21],  $\xi_M = \sum_N D_{MN} \xi'_N$ , where  $\xi_M$  refers to a  $(r_\perp, r_\parallel)$  bin of the distorted model and  $\xi'_N$  to a bin of the undistorted model. The matrices  $D_{MN}$  are constructed using the same linear operators  $\eta$  used to compute the projected field (eq. (3.3)). The elements of the matrices are given by equations (21) and (22) of [dMdB20](#) for, respectively, the auto- and cross-correlations.

Two improvements have been made to the distortion matrix treatment over that of [dMdB20](#). First, the undistorted model bins are now calculated on a grid of  $2 h^{-1} \text{ Mpc}$  (instead of the  $4 h^{-1} \text{ Mpc}$  bins used to measure the correlations). Second, in the  $r_\parallel$  directions, we extend the modelling of the undistorted correlations to  $300 h^{-1} \text{ Mpc}$  rather than to  $200 h^{-1} \text{ Mpc}$ . This improves the accuracy of the distortion calculation at high  $r_\parallel$ , but has a negligible impact on the BAO results (see section 6.4).

The computation of the distortion matrix  $D_{MN}$  is computationally intensive, and following previous work we only use a small fraction of the dataset to approximate it [21, 31]. By default we use 1% of the Ly $\alpha$  pixels, but in section 6.4 we show that doubling that number does not affect the BAO results.

### 4.4 Metal contamination

The Ly $\alpha$  forest is contaminated by absorption from atomic transitions of elements other than neutral hydrogen. The auto-correlation of those absorbers and their cross-correlation with Ly $\alpha$ , QSOs, or other transitions can contribute to the measured Ly $\alpha$  auto-correlation and its cross-correlation with QSOs, and has to be taken into account. We call them *metal* absorbers. Labeling  $\delta_m$  their contribution to the total flux decrement  $\delta$ , the Ly $\alpha$  forest auto-correlation will have contributions of the form

$$\langle \delta \delta \rangle = \langle \delta_\alpha \delta_\alpha \rangle + \sum_m \langle \delta_m \delta_m \rangle + \sum_m \langle \delta_\alpha \delta_m \rangle + \sum_m \sum_{m' \neq m} \langle \delta_m \delta_{m'} \rangle. \quad (4.1)$$

The second term is the contribution of the metal auto-correlations that are present for all foreground absorbers with  $\lambda_m > \lambda_{\min}$  where  $\lambda_{\min}$  is the minimum wavelength of the forest in the quasar rest-frame. The third term is the contribution of the cross-correlation

of metal absorbers with  $\text{Ly}\alpha$ . Only transition wavelengths close to the  $\text{Ly}\alpha$  line will have a significant contribution for the range of longitudinal separation we are studying. The fourth term is the cross-correlation of metals, which can introduce a lot of complexity in the interpretation of the measured correlation function. Fortunately, the metal absorbers are much less abundant than neutral hydrogen and hence have much smaller absorption. We include in the fit the cross-correlation of Si II and Si III absorbers, whose cross-correlation with  $\text{Ly}\alpha$  is also observed, but we neglect the cross-correlation of other foreground absorbers. For the  $\text{Ly}\alpha$  quasar cross-correlation, the situation is simpler, as we have only contributions from absorbers with transition wavelengths close to the Lyman- $\alpha$  line.

Previous analyses based on measurements of the  $\text{Ly}\alpha$  1D correlation [21], or cross-correlation with strong absorbers [81–83], have shown that only a few transitions had to be considered for the cross-correlation terms cited above. Those are Si II lines at 1190Å, 1193Å, and 1260Å, and one Si III line at 1207Å. Other lines are present but can be neglected. While all of the above transitions will also have an auto-correlation term that we account for in the modeling, it is their cross-correlation with  $\text{Ly}\alpha$  that will allow us to differentiate their signals and measure their biases. Other metal transitions at longer wavelengths and lower redshifts only contribute significantly with their auto-correlation. Those cannot be easily separated from the  $\text{Ly}\alpha$  auto-correlation signal. In a companion paper [32], we show that they are dominated by the C IV absorption, and that one can measure their contribution from the auto-correlation in the side bands (at wavelengths larger than the  $\text{Ly}\alpha$  line in the quasar rest-frame). Following a first estimate from their analysis,<sup>16</sup> we use a prior on the effective C IV bias of  $b_{\text{CIV}}^{\text{eff}} = -0.0243 \pm 0.0015$  with  $\beta_{\text{CIV}} = 0.5$  which combines the signal from C IV and other transitions (notably Mg II and Si IV).

We use the same set of metal absorbers and priors when modelling the  $\text{Ly}\alpha(A) \times \text{Ly}\alpha(B)$  correlations. Absorbers with wavelength close to the  $\text{Ly}\beta$  line like the O VI lines at 1032Å and 1038Å do not contribute in cross-correlation with  $\text{Ly}\alpha$  to our measurements because their wavelength is much smaller than the  $\text{Ly}\alpha$  line. Their auto-correlation, peaking at zero separation, does not contribute either because we do not measure the auto-correlation of pixels in the B region,  $\text{Ly}\alpha(B) \times \text{Ly}\alpha(B)$ .

As explained in [31], for each pair of absorbers  $(m, n)$  we compute a metal matrix that provides the mapping between the true co-moving separation  $(r_{\parallel}, r_{\perp})$  of the two absorbers and their apparent separation when assuming both are caused by the Lyman- $\alpha$  transition. In previous works this mapping was computed numerically for a small fraction of the total number of pairs in the sample. This estimation was not precise enough at small separation and expensive in computing time. We now compute uniquely the shift along  $r_{\parallel}$  and ignore the few percent change in  $r_{\perp}$ . This simplification allows us to measure more precisely the effect with one dimensional integrals using the sum of the weights as a function of wavelength.

#### 4.5 Correlated noise from the data processing

The data processing [37] introduces correlated noise among the spectra from fibers of the same spectrograph, which correspond to fibers from a unique petal in the focal plane.<sup>17</sup>

<sup>16</sup>A slightly lower bias value was found from a revised analysis. We show in section 6.4 that this has a negligible impact on the best fit BAO parameters and their uncertainties.

<sup>17</sup>The focal plane is segmented into 10 petals or wedges, see [38].

This contamination is studied in a companion paper [32], where we show that the dominant contribution is the sky background model noise (as in BOSS/eBOSS, see [21] and [dMdB20](#)). We find that the following expression is sufficient to describe this contamination to the DESI Ly $\alpha$  auto-correlation function

$$\xi_{\text{cont}}(r_{\parallel}, r_{\perp}) = a_{\text{noise}} \delta^K(r_{\parallel}) f(r_{\perp}) \equiv a_{\text{noise}} \xi_{\text{noise}}(r_{\parallel}, r_{\perp}) \quad (4.2)$$

with  $a_{\text{noise}}$  an amplitude and  $f(r_{\perp})$  a decreasing function of  $r_{\perp}$  proportional to the fraction of pairs at  $r_{\perp}$  that belong to the same petal. This function is evaluated numerically in [32] assuming pairs at  $z = 2.4$  for the fiducial cosmology to convert angles to co-moving separations. We have  $f(r_{\perp} > 110 h^{-1} \text{ Mpc}) = 0$  when the separation exceeds the size of a petal. We chose an arbitrary normalization  $f(0) = 1$  such that  $a_{\text{noise}}$  is close to the value of this contamination in the first  $(r_{\parallel}, r_{\perp})$  bin.

#### 4.6 HCD contamination

The presence of High Column Density systems (HCDs) in the quasar spectra, including Lyman Limit Systems (LLS,  $\log N_{\text{HI}} > 17.2$ ) and Damped Lyman  $\alpha$  systems (DLA,  $\log N_{\text{HI}} > 20.3$ ), complicates the modeling of 3D correlations in the Ly $\alpha$  forest [84–86]. Like the Ly $\alpha$  forest itself, HCDs are tracers of the underlying matter density and on very large scales (larger than the width of their absorption profiles) their contamination is limited to a change in the linear bias parameters of the Ly $\alpha$  forest (see section 4.2 in [85]). However, the damping wings of DLAs can extend to fairly large line-of-sight separations, and adds an extra scale dependence to the correlation function on scales of tens of megaparsecs.

To diminish the effect of the HCDs on the correlation function, we mask the highly absorbed wavelength range of identified Damped Lyman alpha systems (DLAs). As described in section 2, we have good efficiency for identifying DLAs in the spectra with high signal-to-noise. However, the smoothing effect of unidentified HCDs remains, and this must be modeled.

We use a scale-dependent Ly $\alpha$  bias of the form  $b'_{\alpha} = b_{\alpha} + b_{\text{HCD}} F_{\text{HCD}}(k_{\parallel})$  and a similar form for the RSD parameter  $\beta_{\alpha}$ . As discussed in appendix A (see also [87]), the form of  $F_{\text{HCD}}(k_{\parallel})$  and the magnitude of  $b_{\text{HCD}}$  can be related to the column density distribution of unmasked HCDs and the bias of their host halos. The  $k_{\parallel}$  dependence is given by the Fourier transform of the HCD Voigt profiles (an absorption profile with a Gaussian core and Lorentzian tails). If the column-density distribution of the unmasked HCDs was known, it would then be possible to calculate  $b_{\text{HCD}} F_{\text{HCD}}(k_{\parallel})$ . Unfortunately, at present we do not know precisely the efficiency of the DLA detection in noisy spectra, and we only know approximately the bias of halos hosting DLAs [76, 88, 89].

Because of this, and motivated by the fact that the Fourier transform of a Lorentzian is an exponential function, we model the contamination with  $F(k_{\parallel}) = \exp(-L_{\text{HCD}} k_{\parallel})$  and treat  $b_{\text{HCD}}$  and  $L_{\text{HCD}}$  as free parameters of the model. The parameter  $\beta_{\text{HCD}}$  is poorly determined in the fits and we choose to use a prior  $\beta_{\text{HCD}} = 0.5 \pm 0.1$ , a value motivated by the measured bias of  $b_{\text{DLA}} \sim 2$  of the DLA hosts. In the [dMdB20](#) analysis,  $L_{\text{HCD}} = 10 h^{-1} \text{ Mpc}$  was fixed, but in the present analysis we vary this parameter and find  $L_{\text{HCD}} = (6.51 \pm 0.9) h^{-1} \text{ Mpc}$ . As discussed in section 6.4, variations in the modelling of the HCD contamination have a negligible impact on the BAO results.

## 4.7 Quasar redshift errors

Similar to the impact of random peculiar velocities (or ‘‘Fingers of God’’), random errors in the estimation of quasar redshifts dilute the clustering of quasars along the line of sight. This has an impact in the cross-correlation of quasars and the Ly $\alpha$  forest, as first discussed in [90]. Following [dMdB20](#), in our main analysis we model this smoothing with a Lorentzian with free parameter  $\sigma_z$ , but in section 6.4 we show that using a Gaussian instead has a negligible impact on the BAO results.

A small systematic error in the quasar redshifts would be very difficult to detect in the auto-correlation of quasars. However, such an offset would shift the Ly $\alpha$ ×QSO cross-correlation such that it would no longer peak at  $r_{\parallel} = 0$ .<sup>18</sup> We parameterize this shift with a free parameter  $\Delta r_{\parallel}$ .

The impact of redshift errors in the Ly $\alpha$ ×QSO cross-correlation can be seen as a nuisance in Ly $\alpha$  BAO studies, but as discussed in [49] it also provides a great diagnosis tool to better calibrate the redshifts of quasars.

We find that the quasar redshifts are unbiased, with  $\Delta r_{\parallel} = (0.066 \pm 0.058)h^{-1}$  Mpc. We also find that the combination of random peculiar velocities and redshift errors are described by a Lorentzian with  $\sigma_z = (3.67 \pm 0.14)h^{-1}$  Mpc, significantly smaller than the value reported in the eBOSS analysis of [dMdB20](#),  $\sigma_z = (6.86 \pm 0.27)h^{-1}$  Mpc. The reduced redshift errors can be explained by the updated quasar templates used in `Redrock` (see table 6 of [47]).

## 4.8 Quasar radiation (proximity effect)

Quasars are some of the brightest objects in the Universe. Therefore, we expect them to significantly ionize their surroundings, an effect sometimes referred to as the *proximity effect*. The Ly $\alpha$  forest of neighbouring quasars is therefore affected by two competing effects: the gas density is higher than average (quasars live in high-density regions), but the neutral fraction is lower than average (due to the quasar radiation).

Following [90], we use a simple model to account for the proximity effect in the Ly $\alpha$ ×QSO cross-correlation. In particular, we use the implementation of [dMdB20](#) that assumes isotropic radiation from the quasar, a mean-free path of UV photons of  $\lambda_{\text{UV}} = 300 h^{-1}$  Mpc, and has a single free parameter  $\xi_0^{\text{TP}}$  that sets the amplitude of the contamination.

## 4.9 Small-scales corrections

The BAO parameters only shift the peak component of the model, and as discussed in [15] this is by construction zero on scales smaller than  $80 h^{-1}$  Mpc. Therefore, one could decide to limit the BAO analysis to these very large separations. However, some of the nuisance parameters described in this section can only be constrained when extending the analysis to smaller separations. This is the case for the parameters describing quasar redshift errors, or the parameter describing the Si III line at 1207Å that causes a sharp feature at ( $r_{\perp} \sim 0$ ,  $r_{\parallel} \sim 20 h^{-1}$  Mpc). Moreover, the distortion matrix discussed in section 4.3 spreads the impact of some of these small-scale effects to larger separations (in particular along the line of sight). For this reason, in our main analysis we include the measurement of correlations

---

<sup>18</sup>Remember that the cross-correlation is measured for positive and negative values of  $r_{\parallel}$ .

down to  $r > 10 h^{-1}$  Mpc, and in section 6.4 we show that the BAO results do not depend on the minimum separation.<sup>19</sup>

In order to fit the Ly $\alpha$  auto-correlation to  $10 h^{-1}$  Mpc, we follow previous Ly $\alpha$  BAO analyses and use a multiplicative correction to the model of the Ly $\alpha$  power spectrum. In particular we use the correction from [91], calibrated with hydrodynamical simulations, that models both the effect of non-linearities in the densities and velocities, but also thermal effects in the IGM (thermal broadening, pressure).<sup>20</sup> In section 6.4 we show that this correction has a negligible impact on our BAO parameters.

An equivalent model was proposed in [92] for the cross-correlation of the Ly $\alpha$  forest with halos of intermediate masses. However, the simulations used were too small to contain enough massive halos to accurately study the clustering of quasars. We decided to not include this correction in our model. However, as discussed in section 4.7 we do take into account the impact of non-linear peculiar velocities on the Ly $\alpha$  × QSO cross-correlation.

Finally, following [31] and previous Ly $\alpha$  BAO analyses, we take into account the finite size of our correlation function bins.

## 5 Measurement of Baryon Acoustic Oscillations

After presenting the measured correlations in section 3 and discussing the model we used to describe them in section 4, in this section we summarise the statistical method used to fit the measurement and present the main results of this paper: the measurement of the BAO scale along and perpendicular to the line of sight.

We use the `Vega` package both for the modelling of the correlations and for the parameter inference. We use a Gaussian likelihood, and the main results presented in this section were obtained using the Nested Sampler `Polychord` [93, 94]. The best-fit values are the mean of the posteriors, and the reported uncertainties are the 68% credible intervals. However, this analysis is computationally intensive, and in most of the tests in section 6 we use instead a simpler method: we use the `iminuit` software [95, 96] to find the maximum of the likelihood, and use the derivatives of the likelihood around the best-fit point to estimate a Gaussian posterior. As discussed in appendix D, both BAO estimates are very similar.

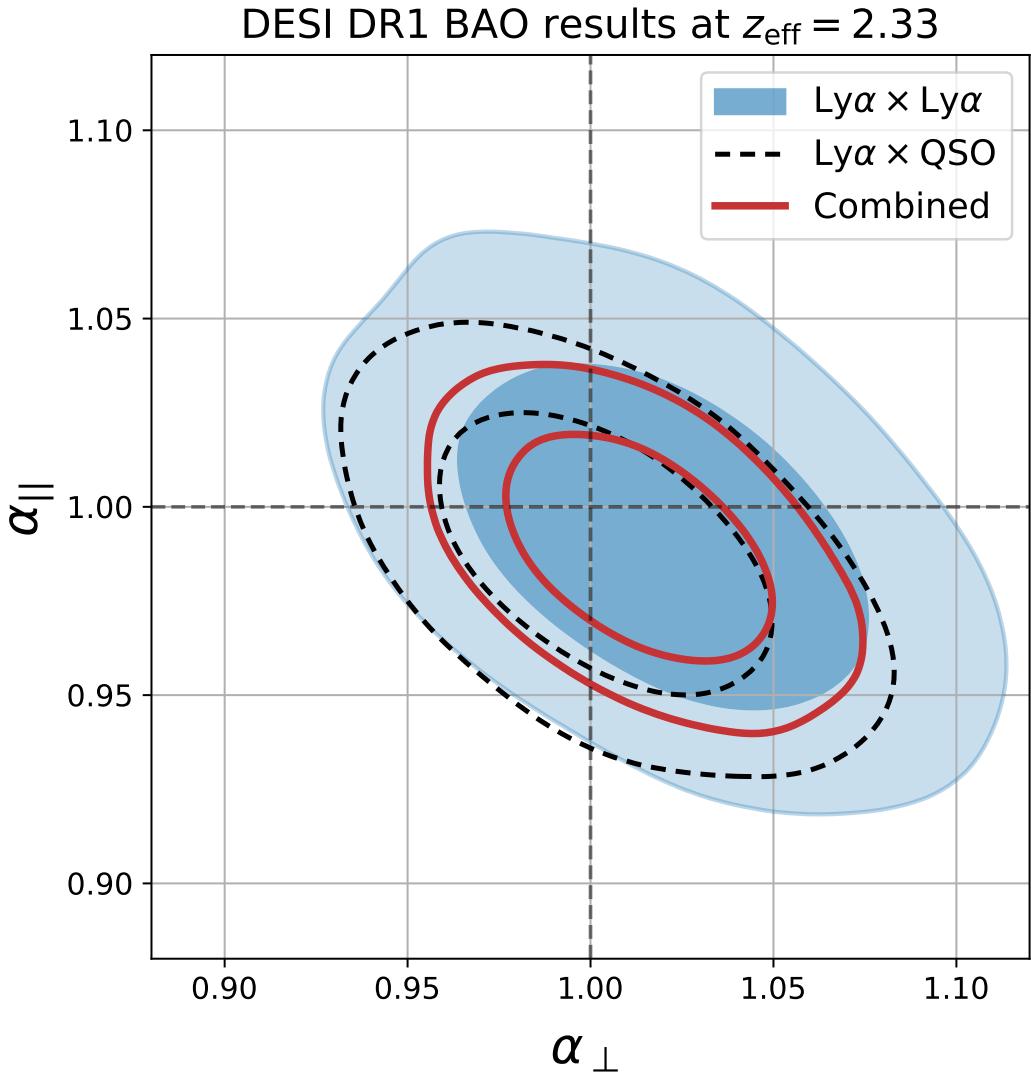
We start with a data vector composed of 4 different 2-point functions, two of them with 2 500 data points (Ly $\alpha$ (A) × Ly $\alpha$ (A) and Ly $\alpha$ (A) × Ly $\alpha$ (B)) and two of them with 5 000 data points (Ly $\alpha$ (A) × QSO and Ly $\alpha$ (B) × QSO), for a total of 15 000 data points. While these are mostly independent, as discussed in section 3 we include their small cross-covariance, and therefore we use a  $15\,000 \times 15\,000$  covariance matrix. However, following `DMdB20` we limit the range of separations used in the fits to  $10 < r < 180 h^{-1}$  Mpc, reducing the number of data points used in the combined fit to 9540 (see table 3).<sup>21</sup>

The constraints on the BAO parameters are listed in table 3 and shown in figure 7, where we show constraints from the Ly $\alpha$  auto-correlation (in filled blue contours, including both Ly $\alpha$ (A) × Ly $\alpha$ (A) and Ly $\alpha$ (A) × Ly $\alpha$ (B)) as well as constraints from the cross-correlation with quasars (dashed black, including both Ly $\alpha$ (A) × QSO and Ly $\alpha$ (B) × QSO). Both measurements

<sup>19</sup>The nuisance parameters do change with the minimum separation included in the fits.

<sup>20</sup>We use the values from the *Planck* simulation in table 7 of [91], interpolated to our effective redshift.

<sup>21</sup>As discussed in section 6.4, the results are not sensitive to the exact range of separations used.



**Figure 7.** Measurements of the BAO parameters along the line of sight ( $\alpha_{\parallel}$ ) and across the line of sight ( $\alpha_{\perp}$ ), with contours corresponding to the 68% and 95% confidence regions. The auto-correlation results (filled blue contours) are the combined measurement of the Ly $\alpha$  forest auto-correlations in the regions A and B. The cross-correlation results (dashed black) are the correlations of the forest in these two regions with quasars. The combined results (solid red) simultaneously fit all four correlations taking into account their cross-covariance, and are the main result of this publication.

are consistent, and their combined constraints (shown in solid red lines) are  $\alpha_{\perp} = 1.013 \pm 0.024$  and  $\alpha_{\parallel} = 0.989 \pm 0.020$ , with a correlation coefficient of  $\rho = -0.48$ .

As discussed in section 4, in addition to the two BAO parameters our model has 15 nuisance parameters that we marginalise over. The number of degrees of freedom in the combined fit is 9523 (9540-17), and the  $\chi^2$  of the best-fit model is 9624.36, with a probability of having a value larger than this of 23%. The best-fit values of the nuisance parameters are discussed in appendix B. Some of these nuisance parameters only affect the auto-correlation or the cross-correlation, and are therefore ignored when fitting these correlations individually (see table 3 and table 5). Moreover, when analysing these correlations separately we are

Parameter	Best fit		
	Combined	$\text{Ly}\alpha \times \text{Ly}\alpha$	$\text{Ly}\alpha \times \text{QSO}$
$\alpha_{\parallel}$	$0.989 \pm 0.020$	$0.993^{+0.029}_{-0.032}$	$0.988^{+0.024}_{-0.025}$
$\alpha_{\perp}$	$1.013 \pm 0.024$	$1.020^{+0.036}_{-0.037}$	$1.005 \pm 0.030$
$\rho_{\alpha_{\parallel}, \alpha_{\perp}}$	-0.48	-0.46	-0.50
$N_{\text{bin}}$	9540	3180	6360
$N_{\text{param}}$	17	12	14
$\chi^2_{\text{min}}$	9624.36	3183.79	6427.41
p-value	0.23	0.42	0.23

**Table 3.** Best fit BAO parameters (mean of the posterior), uncertainties (68% credible intervals) and correlation coefficient  $\rho$  from the three main analyses: auto-correlations ( $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{A})$  and  $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{B})$ ), cross-correlations ( $\text{Ly}\alpha(\text{A}) \times \text{QSO}$  and  $\text{Ly}\alpha(\text{B}) \times \text{QSO}$ ) and their combination. All parameters are given at  $z_{\text{eff}} = 2.33$ . The *p-value* is only accurate for the combined analysis, because in the other analyses we fix the value of one of the nuisance parameters to the best-fit value in the combined analysis (see discussion in appendix B).

not able to break some of the degeneracies between nuisance parameters, and we use extra priors as described in table 5.

In the latest  $\text{Ly}\alpha$  BAO analyses from eBOSS, the auto-correlation provided  $\sim 20\%$  better constraints on  $\alpha_{\parallel}$  than the cross-correlation, while providing  $\sim 10\%$  weaker constraints on  $\alpha_{\perp}$  (see figure 12 in [dMb20](#)). Redshift space distortions in the quasar dataset ( $\beta_Q \sim 0.3$ ) are milder than in the  $\text{Ly}\alpha$  forest ( $\beta_{\alpha} \sim 1.7$ ), reducing the constraining power along the line-of-sight direction. On the other hand, the  $\text{Ly}\alpha$  BAO measurement from DESI DR1 seems to be dominated by the cross-correlation for both  $\alpha_{\perp}$  and  $\alpha_{\parallel}$  (see figure 7). Most  $\text{Ly}\alpha$  quasars in DESI DR1 have only one observation (see right panel of figure 1). In future data releases their signal-to-noise will increase as we collect more observations, and the constraining power of the auto-correlation (doubly affected by noise in the spectra) will increase more than that of the cross-correlation.

## 6 Analysis validation

Baryon Acoustic Oscillations (BAO) imprint a characteristic three-dimensional feature in the measured correlations. On the other hand, most spurious correlations from instrumental systematics and astrophysical contaminants are smooth and featureless.<sup>22</sup> This makes BAO measurements particularly robust.

However, some of the analysis choices presented in section 2 cause small changes in the dataset that introduce statistical fluctuations in the BAO measurement. Examples of these are the observed wavelength range, the rest-frame limits of the  $\text{Ly}\alpha$  regions A and B, or the masking of pixels (due to sky lines, DLAs or BAL features). Moreover, differences in the

<sup>22</sup>An important exception is the contamination from metal lines (mostly Silicon) that cause characteristic *bumps* in line-of-sight correlations, but with a very different angular ( $\mu$ ) dependence that allows us to distinguish between them and the BAO parameters.

quasar redshift estimators cause pixels to fall in and out of the A and B regions, adding an extra source of statistical fluctuations that is also seen in mocks (see appendix B of [dMb20](#)).

## 6.1 Blinding strategy

In order to avoid unconscious or confirmation biases, the development and testing of the analysis framework defined in sections [2–5](#) was done using synthetic datasets and blinded measurements. We considered several blinding strategies for the DESI Ly $\alpha$  BAO analysis, including the possibility of blinding the data at the catalog level as done in the galaxy BAO analysis of DESI [\[6\]](#). However, the presence of known sky lines in the spectra, as well as the presence of absorption lines with small restframe wavelength separations from Ly $\alpha$ , made it challenging to apply a robust blinding to the data at the catalog level.

For this reason, we opted instead for blinding the measured correlation functions following a simple method described in appendix [C](#). In short, we applied an additive correction to the measured correlation functions to mimic a blind shift in the BAO parameters.

We defined a list of tests that we needed to pass in order to validate the analysis before *unblinding* the measurement. First, we validated the analysis using synthetic datasets (or mocks), as explained in detail in a companion paper [\[33\]](#), and as summarised in section [6.2](#). Second, we studied the consistency of the results under various data splits, as discussed in section [6.3](#). Finally, we tested the robustness of the results under variations in the analysis setup, as described in section [6.4](#). We report here test results applied to the unblinded data, but the same tests were first performed on the blinded data set.

## 6.2 Validation using synthetic data

A detailed description of the procedure to generate synthetic DESI spectra (or mocks) for Ly $\alpha$  studies can be found in [\[34\]](#). The analysis validation of the Ly $\alpha$  BAO measurement using mocks is presented in a companion paper [\[33\]](#). Here we give a brief summary of the mocks, and we show some of the main tests validating the analysis.

### 6.2.1 DESI Ly $\alpha$ mocks

We generated DESI mocks from two different sets of fast simulations: 100 realisations of `LyCoLoRe` mocks [\[97, 98\]](#) and 50 realisations of `Saclay` mocks [\[99\]](#). Both sets of mocks use a log-normal description of the density field, and use simplified recipes to distribute quasars and simulate the optical depth of Ly $\alpha$  absorption in redshift space. These recipes were calibrated in order to approximately reproduce the mean flux, the 1D power spectrum, and the large-scale biases of the Ly $\alpha$  forest as measured by the eBOSS Collaboration. The `LyCoLoRe` simulations were also used in the final Ly $\alpha$  BAO analysis of eBOSS presented in [dMb20](#).

As described in [\[34\]](#), these simulations are post-processed with the script `quickquasars` of the `desisim` package,<sup>23</sup> where the DESI specificities are introduced, namely the footprint, signal to noise, spectrograph resolution, quasar redshift errors, etc. At the same time, astrophysical contaminants are introduced such as Damped Lyman alpha systems (DLAs), Broad Absorption Line features (BALs), and absorption from metal lines.

<sup>23</sup><https://github.com/desihub/desisim>. We used version **v0.38.0**.

We have made two small changes in the procedure with respect to the description in [34]. First, we have improved the way in which we imprint the footprint inhomogeneities caused by the survey strategy of DESI. Second, we have slightly modified the recipes to add metal absorption to the mocks in order to better match the amount of metal contamination seen in the data. Both of these changes are discussed in more detail in [33].

We analyse these mocks using the same analysis applied to the data, with the following minor exceptions in the modelling: (i) we ignore the contamination from CIV, the transverse proximity effect and correlated sky residuals, since these are not included in the mocks; (ii) we include an extra smoothing to the model to account for pixelisation effects coming from the finite cell size of the log-normal simulations; (iii) we do not smooth the peak component of our model since we use lognormal simulations that do not capture the non-linear broadening of the BAO peak. These differences are also discussed in more detail in [33].

The contaminants that we cannot study with synthetic data are studied with variations in their modelling in section 6.4 or discussed in a companion paper. For instance, the CIV contamination and the correlated signal from the sky subtraction are studied in [32] as discussed in section 4.5. Given that the fitted amplitudes of these contaminants are uncorrelated with  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  (see figure 14 of appendix B), we are confident that they do not bias our estimate of the BAO parameters. The lack of non-linear broadening of the BAO peak in the mocks results in an underestimation of the BAO uncertainties measured from mocks (see figure 9). As discussed in section 4, when fitting the DESI DR1 measurement we model the broadening using calculations based on Lagrangian Perturbation Theory [15]. This model reproduces the shape of the BAO peak in the correlations measured from Ly $\alpha$  mocks constructed from N-body simulations [100].

Finally, the algorithm used to identify DLAs in the data requires a significant amount of computing time, and therefore we decided to not run it on the many realisations of mocks. Instead, we assume that we find all HCDs with  $\log N_{\mathrm{HI}} > 20.3$ , and none below this column density, and we mask them in the analysis. Similarly, we assume that we can find all BAL features in the data, and mask the corresponding region of spectra accordingly. The impact of BAL completeness is studied in detail in [62].

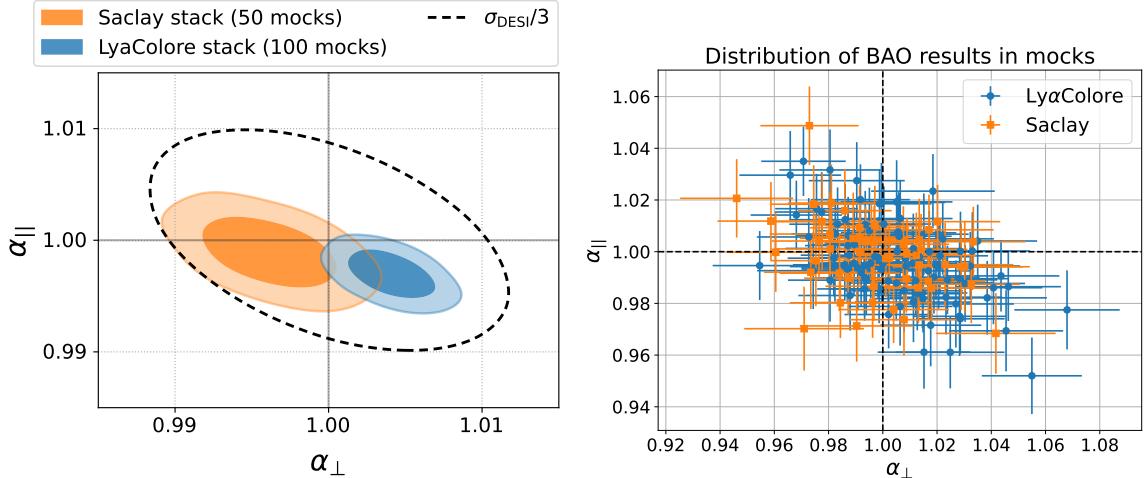
### 6.2.2 Validation of the covariance matrix

In order to validate the covariance matrices of the correlation functions for the purpose of measuring the BAO scale, we measured the quantity

$$\delta_{\alpha} \equiv \left( \partial_{\alpha} M^T C^{-1} \partial_{\alpha} M \right)^{-1} \partial_{\alpha} M^T C^{-1} R \quad (6.1)$$

for each mock, where  $\partial_{\alpha} M$  is the derivative of the best fit model from a stack of mocks with respect to a BAO scale parameter  $\alpha$ ,  $C$  the covariance matrix of the correlation functions and  $R$  the data minus the best fit model from the stack. The covariance  $C$  is determined for each mock independently with the sub-sampling and smoothing method described in section 3.3, so the noise of the covariance itself is directly comparable to that of the true data.

This is not a true fit (which involves a non-linear model plus many nuisance parameters, see section 4) but a straightforward compression of the data and its covariance that is



**Figure 8.** Left: BAO constraints from the “stack” of 100 **Lyacolore** (blue) and 50 **Saclay** (orange) DESI DR1 synthetic datasets, compared to the constraints from data (scaled down by a factor of 3, dashed black ellipse). Right: scatter plot of the best fit  $\alpha_{\perp}$  and  $\alpha_{\parallel}$  from each of the **Lyacolore** (blue) and **Saclay** (orange) mocks. Note the difference of scale between the two plots.

best suited to describe the fluctuations that matter for measuring  $\alpha$ . We also measure the uncertainty on this parameter from the covariance,

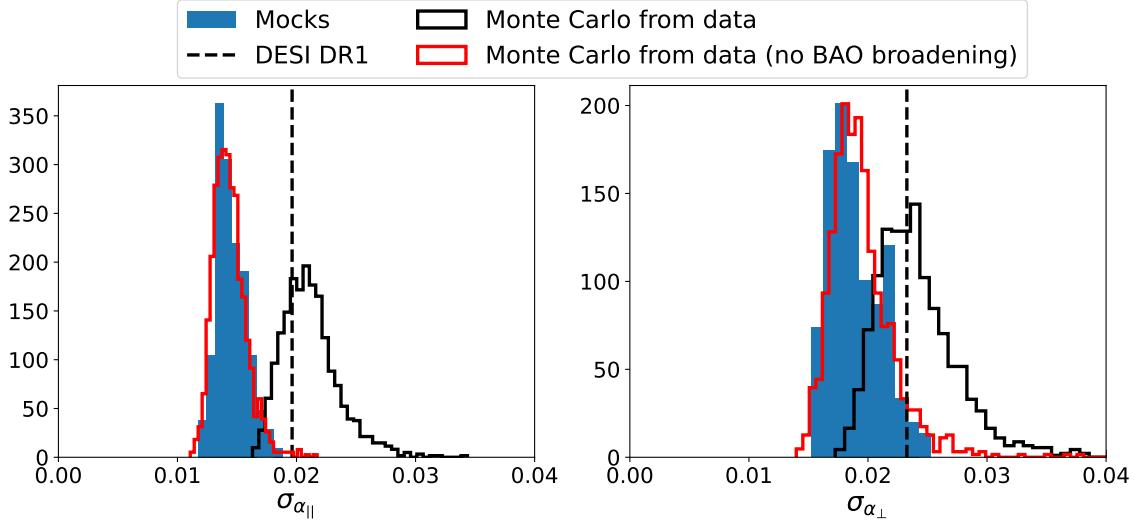
$$\sigma_{\alpha} \equiv \left( \partial_{\alpha} M^T C^{-1} \partial_{\alpha} M \right)^{-1/2}. \quad (6.2)$$

We then measure the rms of the distribution of  $(\delta_{\alpha}/\sigma_{\alpha})$ . If the covariance matrix is correct we expect this distribution to be Gaussian with a rms of 1. We looked at  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  and linear combinations of both (major and minor axes of 2D uncertainty contour). When using the full covariance matrix (including the cross-covariance between the correlation functions), we find a scatter of  $0.96 \pm 0.05$  ( $0.99 \pm 0.06$ ) for  $\alpha_{\parallel}$  ( $\alpha_{\perp}$ ) for the **Lyacolore** mocks, and  $1.07 \pm 0.07$  ( $1.03 \pm 0.07$ ) for the **Saclay** mocks. This shows that the statistical uncertainties derived from the covariance matrix are a good estimate of the dispersion among random mock realizations. This gives us confidence in the covariance matrix at the scales of interest for the measurement of the BAO. As discussed in more detail in [33], the scatter was 10% larger for both sets of mocks when ignoring the cross-covariance between the correlation functions.

### 6.2.3 Validation of BAO estimates

We measured the correlations in 100 **Lyacolore** mocks and 50 **Saclay** mocks mimicking the DESI DR1 dataset, and combine their measurements of the correlations into “stacks of correlations”, with a statistical power much larger than that of the final DESI dataset. The BAO constraints from these stacks are shown in the left panel of figure 8. We used the same cosmology to generate the mocks and to analyse them, so we should expect to recover values of  $(\alpha_{\perp}, \alpha_{\parallel})$  consistent with unity.

In order to proceed with the unblinding, we had set a requirement that the measurement on the stack of many mocks could not present a bias larger than a third of the statistical



**Figure 9.** Distributions of BAO uncertainties along ( $\sigma_{\alpha_{\parallel}}$ , left) and across ( $\sigma_{\alpha_{\perp}}$ , right) the line of sight. The blue histogram shows the distribution from the analysis of 150 DESI DR1 mocks, while the vertical dashed lines are the uncertainties measured in the data. The solid black line shows the distribution of BAO uncertainties from Monte Carlo realisations of the data covariance matrix, when using the best-fit model. The solid red line, on the other hand, shows an equivalent distribution from Monte Carlo realisations generated around a linear model that does not include the non-linear broadening of the BAO peak. These Monte Carlo realisations are discussed in detail in [33].

uncertainty obtained when fitting the blinded data.<sup>24</sup> This corresponded to a tolerance of  $\sim 0.005$  in  $\alpha_{\parallel}$  and  $\sim 0.007$  in  $\alpha_{\perp}$ .

There is a small bias in the measurement of BAO from the stack of 100 LyaCoLoRe mocks (blue contours), but it is smaller than the requirement accuracy (black dashed contours). The difference between the results from the two sets of mocks also suggests that any offsets seen due to analysis problems are at the same level as systematics in the creation of the mocks. Moreover, combining these with the results from the stack of 50 Saclay mocks (orange contours) would further reduce the bias.

#### 6.2.4 Validation of BAO uncertainties

We discuss here the distribution of BAO results when fitting each of the DESI DR1 mocks individually, and use them to validate several aspects of our analysis. In the right panel of figure 8 we show the best-fit values for the BAO parameters ( $\alpha_{\parallel}$  and  $\alpha_{\perp}$ ) for the 50 Saclay and the 100 LyaCoLoRe mocks, with errorbars representing the  $1-\sigma$  uncertainties. In the following we combine the results from these 150 mocks.

Figure 9 shows the distribution of uncertainties on  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  in the mocks (blue histogram) and compares it to the uncertainty measured in DESI DR1 (vertical dashed line). One can see that the BAO uncertainties vary significantly from mock to mock, as expected, but that the uncertainties from DESI DR1 are larger than those from mocks.

<sup>24</sup>Note that the dashed black contour in figure 8 shows the uncertainties on data *after unblinding*, and these are slightly larger than the blinded ones.

The BAO uncertainties from analyses of the DESI DR1 mocks are expected to be a bit smaller than in the data, since the mocks used in this analysis do not have non-linear broadening of the BAO peak. In order to study this, we generated 1 000 Monte Carlo (MC) realisations of the correlation functions using the covariance matrix from the data, adding the random fluctuations to the best-fit model from our main analysis.<sup>25</sup> The distribution of BAO uncertainties from these MC realisations is shown in black in figure 9. It clearly shows that the DESI DR1 Ly $\alpha$  BAO result is not an outlier, and that the constraining power of the mocks is larger than that of the data. In the same figure we show (in red) the distribution of uncertainties from another set of 1 000 MC realisations, where the fluctuations have now been added to a model that ignores the non-linear broadening of the BAO peak. This distribution is in very good agreement with the distribution of uncertainties from the fits on individual mocks, and confirms the hypothesis that the non-linear broadening of the BAO peak degrades the Ly $\alpha$  BAO result significantly.

As discussed in [33], the  $\chi^2$  value from the data is consistent with the distribution of  $\chi^2$  in the mocks. In the same publication, we also look at the distribution of BAO residuals ( $\Delta\alpha_{\parallel}/\sigma_{\alpha_{\parallel}}$  and  $\Delta\alpha_{\perp}/\sigma_{\alpha_{\perp}}$ ). Their rms is found to be of  $1.01 \pm 0.07$  for  $\alpha_{\parallel}$  and  $1.11 \pm 0.06$  for  $\alpha_{\perp}$ , with uncertainties obtained through bootstrap. Those values which are close to one validate our error propagation. We note that this is a more comprehensive test than the one presented in section 6.2.2 because it is based on the results from the full non-linear fit that includes numerous nuisance parameters.

From the tests discussed above we conclude that our BAO estimates on mocks are unbiased at the level of precision required by the current dataset, and that the scatter of best-fit values is consistent with the reported uncertainties.

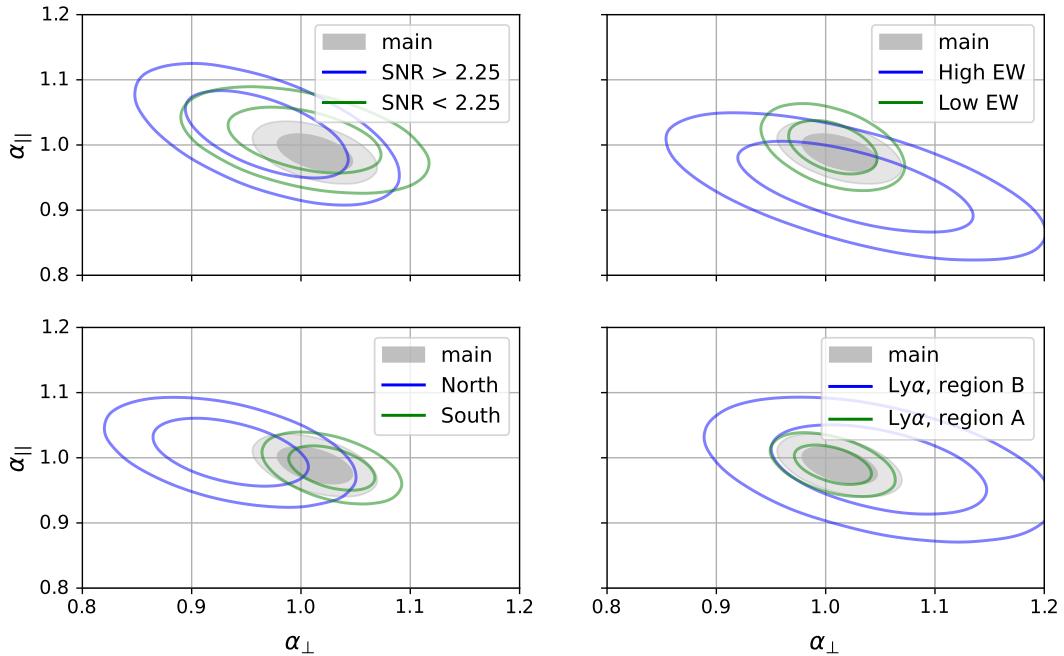
### 6.3 Data splits

The second set of tests that we use to validate the analysis are data splits, where we measure BAO using different subsets of the data. A first data split was already shown in figure 7, where we presented the consistency of BAO measurements from the auto-correlations (including both the Ly $\alpha$ (A)  $\times$  Ly $\alpha$ (A) and the Ly $\alpha$ (A)  $\times$  Ly $\alpha$ (B) correlations) and from the cross-correlations (including Ly $\alpha$ (A)  $\times$  QSO and Ly $\alpha$ (B)  $\times$  QSO). In the bottom right panel of figure 10 we group them instead in correlations that only use pixels in the A region (Ly $\alpha$ (A)  $\times$  Ly $\alpha$ (A) and Ly $\alpha$ (A)  $\times$  QSO, in green) and correlations that use pixels in the B region (Ly $\alpha$ (A)  $\times$  Ly $\alpha$ (B) and Ly $\alpha$ (B)  $\times$  QSO, in blue). The B region does not contain as much information as the A region for several reasons: only quasars at higher redshift have this region in the DESI spectrograph, so the number of Ly $\alpha$  lines-of-sight in the B region is smaller; the B region, as defined in section 2.4, is significantly shorter than the A region, so there are fewer pixels per line-of-sight; finally, the B region is affected by other Lyman lines that add extra variance to the fluctuations.

We will now discuss the consistency of the BAO constraints when splitting the quasar catalog in three ways: by imaging survey used in the target selection; by CIV emission line equivalent width (EW); and by signal-to-noise in the spectrum. In these cases we

---

<sup>25</sup>These Monte Carlo realisations are discussed in more detail in [33].



**Figure 10.** BAO constraints from the main analysis (grey) and from data splits. Top left: low (green) vs high (blue) SNR in the quasar spectrum. Top right: low (green) vs high (blue) CIV equivalent width (EW) in the quasar spectrum. Bottom left: South (green) vs North (blue) imaging used in the quasar target selection. Bottom right: correlations from region A (green) and region B (blue); the A region shows the combined measurement from the auto-correlation of the forest measured in the A region ( $\text{Ly}\alpha \times \text{Ly}\alpha$ ) and the cross-correlation of this region with quasars ( $\text{Ly}\alpha(\text{A}) \times \text{QSO}$ ). The contours labeled region B show the combined measurement of the forest auto-correlation measured in the B region ( $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{B})$ ) and the cross-correlation of this region with quasars ( $\text{Ly}\alpha(\text{B}) \times \text{QSO}$ ).

run alternative end-to-end analyses starting from new sub-catalogs, i.e., fitting new quasar continua, measuring correlations and fitting them separately for each subset of the data.<sup>26</sup>

We start in the bottom left panel of figure 10 by splitting the catalog based on the imaging survey that was used for target selection. Most of the DESI footprint was observed with the DECam camera on the Blanco telescope in Chile, including the entire South Galactic Cap and the southern fraction of the North Galactic Cap [101–103]. We refer to this subset of the data as “South”, while we designate the area that was imaged in the BASS and MzLS surveys at  $\delta > 32.375^\circ$  as “North”. The South sample is significantly larger, as it contains 82% of the quasars (579 166 quasars with  $z > 1.77$  in the South and 130,399 in the North).

In the top right panel of figure 10 we look at a second catalog split based on the CIV EW. We do this split because we expect the shape of the quasar spectral energy distribution to depend on EW, due to the well known anti-correlation between the EW of quasar emission lines and the continuum luminosity known as the Baldwin Effect [104]. We measure the CIV EW of the quasars with `fastspecfit`,<sup>27</sup> finding a median of 37.3 Å for all quasars, and a median of 41.6 Å for  $3\sigma$  measurements of the CIV emission line. We split the sample at

<sup>26</sup>Note that when we split the data set in two by CIV EW or SNR, the density of quasars in each subset is a factor of two lower, so the number of pixel pairs or pixel quasar pairs is about four times smaller.

<sup>27</sup><https://github.com/desihub/fastspecfit>. We use version **v2.4.2**.

39 Å, which produce a low (high) EW sample of 371 751 (337,814) quasars at  $z > 1.77$ . As expected from the Baldwin Effect, the low EW sample is somewhat higher luminosity and has a somewhat higher effective redshift of 2.36 compared to 2.29 for the high EW sample.

Finally, in the top left panel of figure 10 we present the third quasar catalog split, based on mean signal-to-noise ratio (SNR) in the spectra. Instead of splitting the quasar catalog into two subsets of equal size, we chose a SNR threshold of 2.25 such that both subsets have the same weight in the measurement of the auto-correlation function. There are different ways of estimating the SNR of quasar spectra, and we decided to use the mean value of SNR per pixel averaged over the Ly $\alpha$  region, as reported by the `picca` code at the end of the continuum fitting process of the main analysis. This results in a lower SNR catalogue with 321 767 quasars and a higher SNR catalogue with 106 636 quasars. The sum of these catalogues does not match the total size of the catalogue used in the main analysis (709,565 quasars at  $z > 1.77$ ), since we do not detect the forest continuum at  $z < 2$  and therefore do not have a SNR estimate.

With the exception of the North vs South data split, the subsets discussed here share the same footprint and redshift range. However, cosmic variance is a very small component of the covariance matrix and to a first approximation the data splits can be considered as independent. Taking this into account, the BAO constraints on the various data splits are consistent with statistical fluctuations.

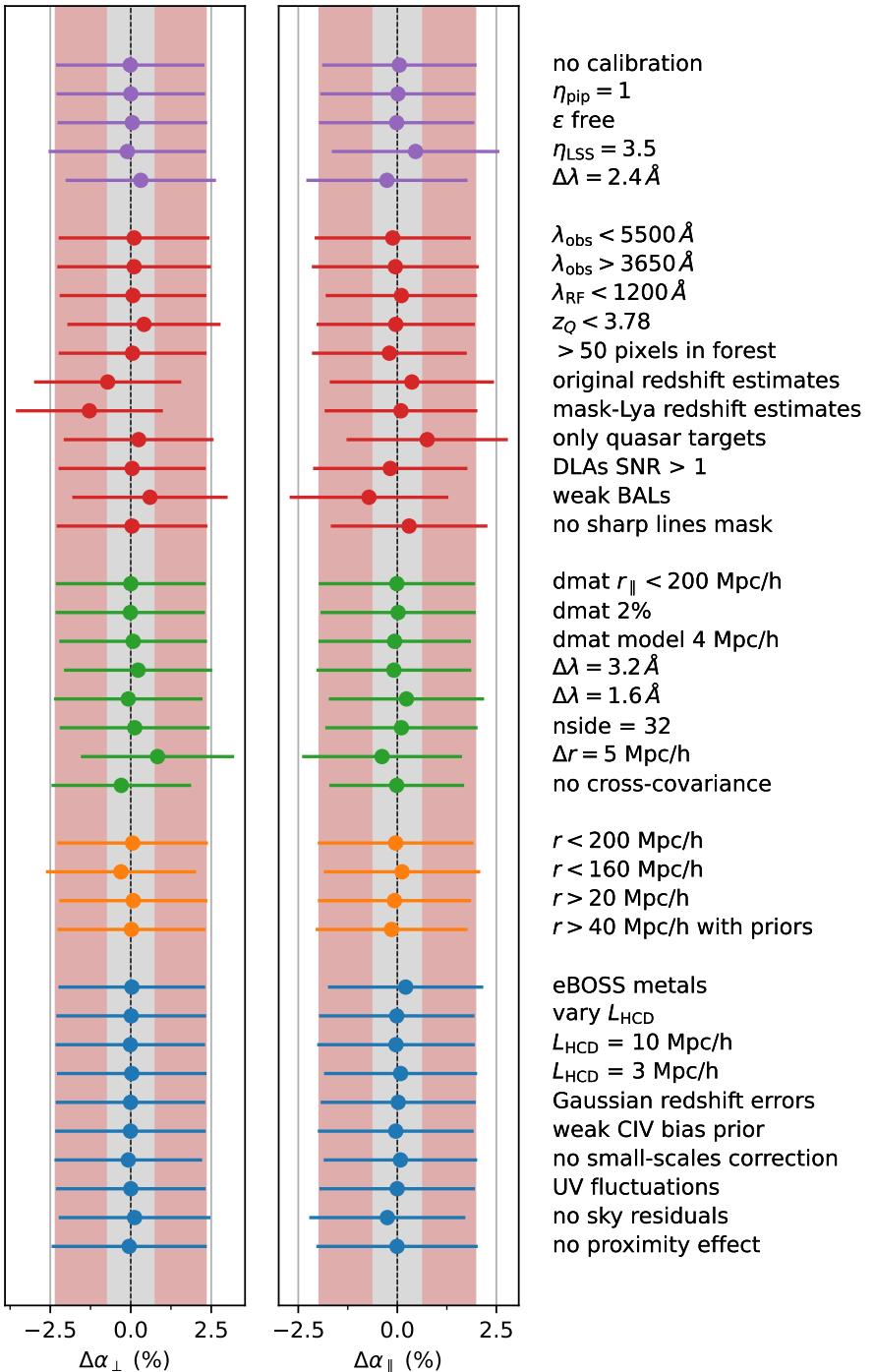
## 6.4 Alternative analyses

In this section we show a final set of validation tests: robustness of the BAO measurement under variations in the analysis configuration. We set a requirement for unblinding that none of the variations would cause a shift on the BAO parameters larger than a third of the statistical uncertainty from the main (blind) analysis. This corresponded to a tolerance of  $\sim 0.005$  in  $\alpha_{\parallel}$  and  $\sim 0.007$  in  $\alpha_{\perp}$ . However, some of the analysis variations result in a small change in the size of the dataset. In these cases we relax the requirement and take into account the probability of the measured shift being caused by statistical fluctuations. These are discussed in appendix E.

In order to reduce the amount of computing time needed in these alternative analyses, we do not run the nested sampler algorithm and only report the maximum likelihood values and Gaussian errors computed by `iminuit`. As discussed in appendix D, the BAO results do not vary significantly with the sampling method used. In some of the alternative analyses `iminuit` has difficulty breaking internal degeneracies between nuisance parameters, particularly between  $L_{\text{HCD}}$  and some of the bias parameters. In order to avoid this, we fix  $L_{\text{HCD}}$  in all the alternative analyses to the best-fit value of the main analysis ( $6.51 h^{-1} \text{Mpc}$ ), and we show in the variation “vary  $L_{\text{HCD}}$ ” that this has a negligible impact on the BAO results.

### 6.4.1 Variations in the estimation of the fluctuations

The method to estimate the Ly $\alpha$  fluctuations starting from the observed quasar spectra is described in section 2.5. In the first set of variations shown (in purple) in figure 11 we quantify the impact on the BAO results when we vary different aspects of the method. In particular we look at the following variations:



**Figure 11.** Shifts in the BAO parameters from alternative analyses, including variations in the method to estimate the fluctuations (purple); variations in the dataset (red); variations in the method to compute correlations and covariances (green); variations in the range of separations used (orange); and variations in the modelling (blue). The red shaded contours show the one  $\sigma$  uncertainty from the main analysis, while the smaller gray area shows the threshold set to these tests ( $\sigma/3$ ). Note that the two parameters are anti-correlated ( $\rho = -0.48$ ). Variations of the dataset (in red) are subject to statistical fluctuations as described in appendix E.

- no calibration: we do not re-calibrate the spectra with the C III region mentioned in section 2.4 and described in [30].
- $\eta_{\text{pix}} = 1$ : we do not apply the re-calibration of the instrumental noise  $\eta(\lambda)$  mentioned in section 2.4 and described in [30].
- $\epsilon$  free: we include an extra term ( $\epsilon$  in [dMdB20](#)) in the computation of the Ly $\alpha$  weights to try to capture quasar diversity.
- $\eta_{\text{LSS}} = 3.5$ : we reduce the contribution from the intrinsic Ly $\alpha$  forest variance to the weights by a factor of two.
- $\Delta\lambda = 2.4\text{\AA}$ : we coadd three pixels into one before the continuum fitting step (as done in [dMdB20](#)). In this variation we also use  $\sigma_{\text{mod}}^2 = 3.1$  as suggested by [30] when coadding DESI data by three pixels.

We continue with a second set of variations (in red) in figure 11 where we look at variations that cause changes in the dataset by removing (or adding) pixels or entire quasars. As discussed in appendix E, these can cause statistical fluctuations in the BAO results. In particular we look at the following variations:

- $\lambda_{\text{obs}} < 5500\text{\AA}$ : we use only Ly $\alpha$  pixels below this observed wavelength ( $\lambda_{\text{obs}} < 5577\text{\AA}$  in the main analysis).
- $\lambda_{\text{obs}} > 3650\text{\AA}$ : we use only Ly $\alpha$  pixels above this observed wavelength ( $\lambda_{\text{obs}} > 3600\text{\AA}$  in the main analysis).
- $\lambda_{\text{RF}} < 1200\text{\AA}$ : we use only Ly $\alpha$  pixels below this rest-frame wavelength ( $\lambda_{\text{RF}} < 1205\text{\AA}$  in the main analysis).
- $z_Q < 3.78$ : we use only quasars with  $z_Q < 3.78$  (highest redshift included in the mocks discussed in section 6.2).
- $> 50$  pixels in forest: we include lines-of-sight with more than 50 valid Ly $\alpha$  pixels (150 in the main analysis).
- original redshift estimates: we use the quasar redshifts in the original run of `Redrock` (slightly biased as discussed in [47]).
- mask-Lya redshift estimates: we use a different estimator for the quasar redshifts, re-running `Redrock` with only wavelengths longer than the Ly $\alpha$  emission line (as done in [dMdB20](#)).
- only quasar targets: we use only quasars that were considered quasar targets.
- DLAs SNR  $> 1$ : we mask DLAs in spectra with SNR  $> 1$  (SNR  $> 3$  in the main analysis).
- weak BALs: we remove spectra with strong BAL features ( $A_I > 840$ , 50% percentile of strongest BALs in eBOSS [61]).
- no sharp lines mask: we do not mask the 4 sharp lines discussed in section 2.4.

Some of the variations (like “mask-Lya redshift estimates”) are slightly outside the threshold of  $1/3\sigma$ . As we discuss in appendix E, these shifts can be explained by statistical fluctuations caused by the addition or subtraction of Ly $\alpha$  pixels from the dataset.

#### 6.4.2 Variations in the measurement of correlations

We now move to the third set of variations (in green) shown in figure 11, where we look at the impact of varying the setup in the measurement of correlations, their covariances and the distortion matrices described in section 3. In particular we look at the following variations:

- dmat  $r_{\parallel} < 200 \text{ Mpc/h}$ : we model the distortion matrix up to  $r_{\parallel} = 200 h^{-1} \text{ Mpc}$  as done in [dMdB20](#) ( $r_{\parallel} = 300 h^{-1} \text{ Mpc}$  in the main analysis).
- dmat 2%: we use 2% of pixels to compute the distortion matrix (1% in the main analysis).
- dmat model 4 Mpc/h: we model the distortion matrix using the same  $4 h^{-1} \text{ Mpc}$  binning as the correlation function ( $2 h^{-1} \text{ Mpc}$  in the main analysis).
- $\Delta\lambda = 3.2 \text{ \AA}$ : we rebin the Ly $\alpha$  fluctuations in groups of 4 pixels (3 in the main analysis).
- $\Delta\lambda = 1.6 \text{ \AA}$ : we rebin the Ly $\alpha$  fluctuations in groups of 2 pixels (3 in the main analysis).
- nside = 32: we measure the correlations in HEALPix pixels defined by nside=32 (16 in the main analysis).
- $\Delta r = 5 \text{ Mpc/h}$ : we use  $5 h^{-1} \text{ Mpc}$  bins in the correlation functions ( $4 h^{-1} \text{ Mpc}$  in the main analysis).
- no cross-covariance: we ignore the cross-covariance of the different correlation functions (as done in eBOSS, [dMdB20](#)).

We do not see any problematic variation related to the measurement of correlations and their covariances.

#### 6.4.3 Variations in the parameter estimation

We move now to the impact of variations in the parameter estimation. We start with the four set of variations (orange) shown in figure 11 by looking at the impact of the range of separations used:

- $r < 200 \text{ Mpc/h}$ : we fit separations with  $r < 200 h^{-1} \text{ Mpc}$  ( $180 h^{-1} \text{ Mpc}$  in the main analysis).
- $r < 160 \text{ Mpc/h}$ : we fit separations with  $r < 160 h^{-1} \text{ Mpc}$  ( $180 h^{-1} \text{ Mpc}$  in the main analysis).
- $r > 20 \text{ Mpc/h}$ : we fit separations with  $r > 20 h^{-1} \text{ Mpc}$  ( $10 h^{-1} \text{ Mpc}$  in the main analysis).

Parameter	Prior
$b_{\text{HCD}}$	$\mathcal{N}(-0.0556, 0.0034)$
$10^3 b_{\text{SiIII}(1207)}$	$\mathcal{N}(-9.78, 0.56)$
$\sigma_v(h^{-1}\text{Mpc})$	$\mathcal{N}(3.66, 0.14)$
$\Delta r_{\parallel}(h^{-1}\text{Mpc})$	$\mathcal{N}(0.067, 0.058)$
$\xi_0^{\text{TP}}$	$\mathcal{N}(0.399, 0.051)$

**Table 4.** Extra Gaussian priors added to some of the variations discussed in section 6.4. They correspond to the best-fit values and uncertainties from the main analysis as reported in table 5.

- $r > 40 \text{ Mpc}/h$  with priors: we fit separations with  $r > 40 h^{-1} \text{ Mpc}$  ( $10 h^{-1} \text{ Mpc}$  in the main analysis). Without the smaller scales we are not able to constraint several nuisance parameters, and therefore we add the informative priors described in table 4.

Finally, in the last set of variations (in blue) shown in figure 11 we look at the impact of different modelling choices. In particular we look at the following variations:<sup>28</sup>

- eBOSS metals: we model the contamination by Silicon lines (metals) following the method used in eBOSS [27] instead of the new method described in section 4.4.
- vary  $L_{\text{HCD}}$ : we vary the parameter  $L_{\text{HCD}}$  in the model of the contamination by HCDs as done in the main analysis. This parameters was fixed to  $L_{\text{HCD}} = 6.51 h^{-1} \text{ Mpc}$  in the other variations to minimise the degeneracies between this and other nuisance parameters. See the related discussion in appendix A.
- $L_{\text{HCD}} = 10 \text{ Mpc}/h$ : we use a fixed value of  $L_{\text{HCD}} = 10 h^{-1} \text{ Mpc}$  to model HCD contamination (free parameter in the main analysis), as was done in [dMb20](#).
- $L_{\text{HCD}} = 3 \text{ Mpc}/h$ : we use a fixed value of  $L_{\text{HCD}} = 3 h^{-1} \text{ Mpc}$  to model HCD contamination (free parameter in the main analysis).
- Gaussian redshift errors: we use a Gaussian distribution to model quasar redshift errors and quasar peculiar velocities (a Lorentzian distribution is used in the main analysis).
- weak CIV bias prior: we use a significantly weaker flat prior on the CIV bias parameter of  $-0.03 < b_{\text{CIV}} < 0$  (instead of  $-0.0243 \pm 0.0015$  in the main analysis).
- no small-scales correction: we ignore the small-scales multiplicative correction from [91] in the modelling of the Ly $\alpha$  auto-correlation.
- UV fluctuations: following the prescription of [21], we model the impact of fluctuations in the UV background [105, 106] on the Ly $\alpha$  forest auto-correlation. We do not detect these fluctuations in our analysis.

<sup>28</sup>The last two variations in this list were added during the refereeing process and were not part of the original analysis validation.

- no sky residuals: we ignore the contamination from correlated sky residuals in the Ly $\alpha$  auto-correlation, discussed in section 4.5.
- no proximity effect: we ignore the *proximity effect*, the impact of quasar radiation in the cross-correlation discussed in section 4.8.

None of the variations in the parameter estimation cause a significant shift in the inference of the BAO parameters. Some of the nuisance parameters, on the other hand, are more sensitive to changes in the analysis setup.

#### 6.4.4 Broadband polynomial corrections

As a final test to demonstrate the robustness of the BAO measurement, we run an alternative analysis where we introduce a flexible but smooth additive component to each of the four modelled correlations. In particular, we follow the procedure of [21, 27] and use (for each correlation) Legendre polynomials  $L_j(\mu)$  of order  $j = 0, 2, 4$  and 6 to describe the angular dependence of the additive terms, divided by powers of  $r^i$  with  $i = 0, 1, 2$  (corresponding to a parabola in  $r^2\xi(r)$ ). The total number of broad-band parameters is therefore 48 (12 for each of the four correlations).

We then fit the baseline model with those additional parameters in the limited separation range  $40\,\text{Mpc}/h < r < 180\,\text{Mpc}/h$  while adding extra priors on nuisance parameters as described in table 4 to break degeneracies between the polynomial coefficients and other parameters. The best fit model with and without those broadband corrections is compared to the data in figures 4 and 5. The shifts in the best fit BAO parameters when adding broadband terms are  $\Delta\alpha_{\parallel} = +0.001$  and  $\Delta\alpha_{\perp} = -0.001$ , and the change in the uncertainties is negligible ( $|\Delta\sigma_{\alpha}| < 0.001$ , both for  $\alpha_{\parallel}$  and  $\alpha_{\perp}$ ). We also note that the value of  $\chi^2$  is reduced by only 47.5 points when adding 48 new parameters.

### 6.5 Conclusion on the validation tests

We have presented numerous validation tests of the analysis using mocks, data splits, variations in choices made for the analysis, and adding ad-hoc broadband terms to improve the fit of the correlations around the BAO peak. All tests where the data set is left unchanged result in variations much smaller than 1/3 of the final statistical error. All data splits and other tests where the data set was altered are consistent with statistical fluctuations.

Adding broadband terms moved the best fit by less than a tenth of the statistical error. The largest offsets were found with mocks with average shifts of  $\Delta\alpha_{\parallel} = -0.003 \pm 0.0014$  and  $\Delta\alpha_{\perp} = +0.004 \pm 0.0018$  for the Ly $\alpha$ Colore mocks (see figure 8). Because we find similar discrepancies among the two sets of mocks (Ly $\alpha$ Colore and Saclay), we can not use those offsets to correct the measurements and we hence have to treat them as systematic uncertainties. Adopting a conservative systematic uncertainty of 0.5%, this results in a 3% (2%) increase of the total uncertainties on the longitudinal (transverse) BAO measurement when combined quadratically with the statistical errors. We consider that this increase in uncertainties is small enough to be ignored. As a result we only report statistical uncertainties in the following sections.

## 7 Discussion

In section 5 we presented a measurement of the BAO parameters ( $\alpha_{\perp}$ ,  $\alpha_{\parallel}$ ) at  $z_{\text{eff}} = 2.33$  from DESI DR1. Thanks to the peak-smooth decomposition introduced in section 4.1, we interpret these parameters as:

$$\begin{aligned}\alpha_{\parallel} &= \frac{D_H(z_{\text{eff}})/r_d}{[D_H(z_{\text{eff}})/r_d]_{\text{fid}}}, \\ \alpha_{\perp} &= \frac{D_M(z_{\text{eff}})/r_d}{[(D_M(z_{\text{eff}})/r_d]_{\text{fid}}},\end{aligned}\quad (7.1)$$

where  $D_M(z)$  is the transverse comoving distance,  $D_H(z) = c/H(z)$  and  $r_d$  is the sound horizon at the drag epoch. Quantities with  $[\cdot]_{\text{fid}}$  are computed with the fiducial cosmology (see table 2). In combination with the  $z < 2$  BAO measurements from [6], our BAO measurement enables the state-of-the-art cosmological constraints presented in [9].

### 7.1 Cosmological distances

Substituting values from our fiducial cosmology, we can rewrite the ( $\alpha_{\parallel}$ ,  $\alpha_{\perp}$ ) constraints as the following constraints on ratios of distances:

$$\begin{aligned}D_H(z_{\text{eff}})/r_d &= 8.52 \pm 0.17, \\ D_M(z_{\text{eff}})/r_d &= 39.71 \pm 0.95,\end{aligned}\quad (7.2)$$

with a correlation coefficient of  $\rho = -0.48$ . For a given value of the sound horizon  $r_d$ , these translate into a measurement of the expansion rate at  $z_{\text{eff}} = 2.33$ :

$$H(z_{\text{eff}}) = (239.2 \pm 4.8) \frac{147.09 \text{ Mpc}}{r_d} \text{ km/s/Mpc} \quad (7.3)$$

and a measurement of the comoving transverse distance to  $z_{\text{eff}}$ :

$$D_M(z_{\text{eff}}) = (5.84 \pm 0.14) \frac{r_d}{147.09 \text{ Mpc}} \text{ Gpc}. \quad (7.4)$$

It is also convenient to report the BAO information as an isotropic dilation parameter

$$D_V(z_{\text{eff}})/r_d = \left( z_{\text{eff}} D_M^2 D_H \right)^{1/3} / r_d = 31.51 \pm 0.44, \quad (7.5)$$

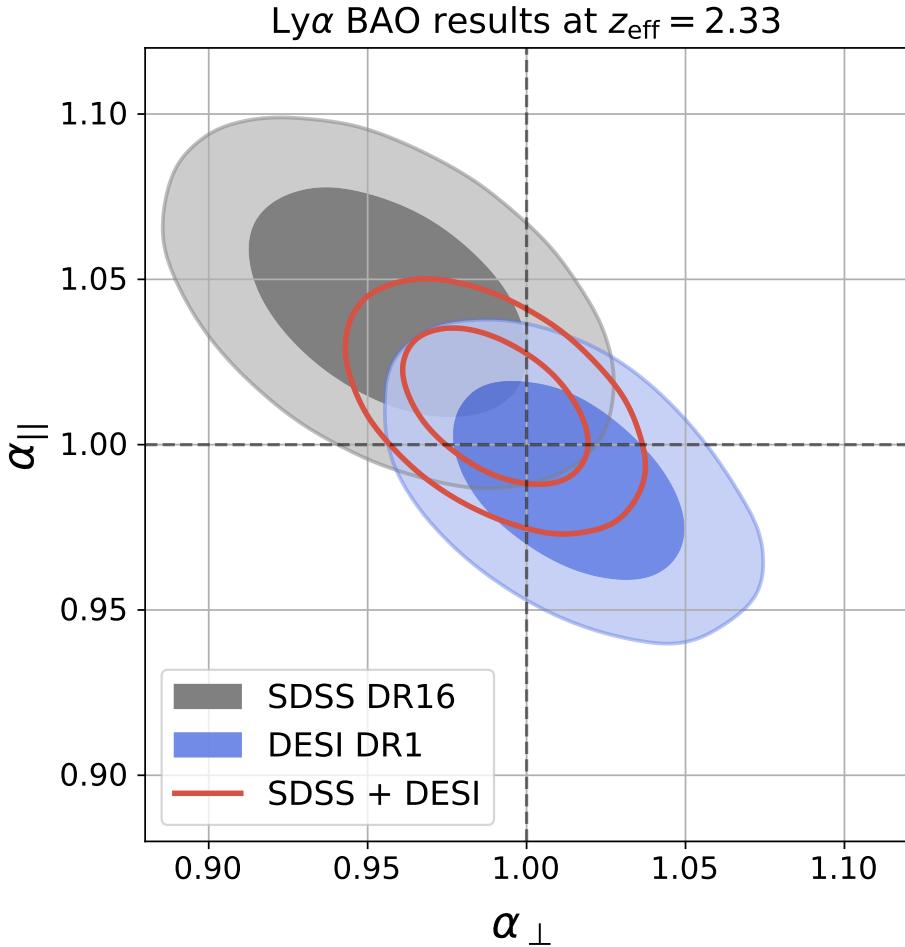
and an anisotropic (or Alcock-Paczynski [107]) parameter  $f_{\text{AP}}$ :

$$f_{\text{AP}}(z_{\text{eff}}) = \frac{D_M(z_{\text{eff}})}{D_H(z_{\text{eff}})} = 4.66 \pm 0.18. \quad (7.6)$$

However, the ratio  $D_V/r_d$  is only the optimal definition of the isotropic BAO parameter in the absence of redshift space distortions. Every BAO measurement will have a different combination of  $(D_H, D_M)$  that will minimise the correlation with  $f_{\text{AP}}$  and will therefore have a smaller relative uncertainty. From the posterior of the Ly $\alpha$  BAO measurement of DESI DR1, we find that this combination is approximately:

$$D_M(z_{\text{eff}})^{9/20} D_H(z_{\text{eff}})^{11/20} / r_d = 17.03 \pm 0.19, \quad (7.7)$$

which corresponds to a 1.1% measurement of the (isotropic) BAO scale at  $z_{\text{eff}} = 2.33$ .



**Figure 12.** Ly $\alpha$  BAO measurement from DESI DR1 (blue) compared to the equivalent measurement from eBOSS using data from SDSS DR16 (dMdB20, gray). The red contours show the combined measurement taking into account the cross-survey covariance as discussed in appendix F.

## 7.2 Comparison with previous measurements from SDSS

Prior to this work, the BAO scale at  $z \sim 2.3$  had been measured by the BOSS and eBOSS collaborations with successively larger Ly $\alpha$  forest datasets [13–15, 19–24, 27]. The BAO measurements from SDSS DR11 [19, 20] showed a mild  $\sim 2.3\sigma$  tension with the best-fit  $\Lambda$ CDM model of Planck. As discussed in dMdB20, this tension gradually disappeared with the addition of more data, and in the last measurement from eBOSS with SDSS DR16 the disagreement was only at the  $1.5\sigma$  level. In figure 12 we compare the BAO measurements from this work (solid blue) with those from dMdB20 using the quasar catalog of SDSS DR16 (solid gray).

The differences between the analyses have been discussed in the previous sections, but we summarise them here. The main difference is the input quasar catalog. There are  $\sim 150\,000$  quasars with  $z > 1.77$  in common in both datasets, representing  $\sim 45\%$  and  $\sim 25\%$  of the quasars in SDSS DR16 and DESI DR1 respectively. DESI targeted quasars down to  $r < 23$ , compared to  $g < 22$  in BOSS and  $g < 22.5$  in eBOSS. We expect most Ly $\alpha$  quasars in DESI to receive four observations by the end of the survey, at which point the distribution

of signal-to-noise per Angstrom in DESI and SDSS quasars will be similar, but in DESI DR1 the majority of quasars have only been observed once and are therefore on average noisier than in SDSS DR16.

The spectrographs are also different: in the blue arm (relevant for Ly $\alpha$  science) DESI has a resolving power  $R$  in the range 2000–3000, compared to 1500–2000 in BOSS/eBOSS. While DESI pixels have a constant width of 0.8Å, SDSS pixels were constant in  $\log \lambda$  with widths ranging from 0.83 to 1.27Å in the relevant range of wavelengths. The spectro-photometry is also greatly improved thanks to an atmospheric dispersion corrector, and the spectrograph optics are much more stable, allowing a finer 2D spectral extraction and improved sky subtraction [37]. The higher quality of the DESI spectra allows us to better determine quasar redshifts, as captured by the  $\sigma_v$  parameter in our fits that is almost a factor of two smaller than the one reported in [dMdB20](#) (see section 4.7 and appendix B).

The *continuum fitting* of quasars for both this work and [dMdB20](#) was performed using the `picca` software, but the code has been re-written and re-structured in a more modular way since the analysis of [dMdB20](#). There have been several minor changes in the methodology as well: we now use the CIII region to re-calibrate the spectra (instead of a region redder than the MgII emission line); we extended the Ly $\alpha$  region to 1205Å (instead of 1200Å); we include quasars with BAL features and mask them following the prescription by [61] (instead of rejecting the entire line of sight); we use modified Ly $\alpha$  weights, presented in [30]. These changes were motivated by the studies of [30] using early DESI data, and the (minor) impact of these changes are discussed in the variations of section 6.4.

The correlations in both analyses were measured with `picca` as well, and the only methodological change is the inclusion of the cross-covariance between correlations discussed in [33]. There are also some minor changes in the modelling, discussed in [31, 33]: the distortion matrix is now modelled at higher resolution and up to  $300 h^{-1}$  Mpc (instead of  $200 h^{-1}$  Mpc); the contamination of metals is modelled in a slightly different way; the model describing the contamination of HCDs now has a free parameter  $L_{\text{HCD}}$  (fixed at  $10 h^{-1}$  Mpc in [dMdB20](#)); the correlated sky residuals are now modelled with an improved method described in [32] that only needs a single free parameter (versus four in [dMdB20](#)). While these changes are a clear improvement in the modelling of the correlations, in figure 11 of section 6.4 we show that their impact on the BAO results is negligible.

Finally, an important difference between the analyses is the effort that we did to validate the analysis pipeline using only blinded measurements and synthetic datasets. As described in section 6, we only *unblinded* the measurements once we had passed a long list of robustness and consistency tests, including variations of the analysis and data splits.

### 7.3 DESI-SDSS cross-survey covariance

Even though there is a large overlap in redshift range and footprint between the Ly $\alpha$  samples of SDSS DR16 and DESI DR1 (see figure 2), the contribution from cosmic variance to these measurements is small, and their cross-survey covariance is smaller than one might naively think. We quantify the cross-covariance in appendix F, where we also use it to combine both

results into a “DESI + SDSS” Ly $\alpha$  BAO measurement:

$$\begin{aligned}\alpha_{\parallel}^{\text{DESI+SDSS}} &= 1.012 \pm 0.016, \\ \alpha_{\perp}^{\text{DESI+SDSS}} &= 0.990 \pm 0.019,\end{aligned}\quad (7.8)$$

with a correlation coefficient of  $\rho = -0.47$ . These contours are also plotted as red contours in figure 12, and can be used in cosmological inference assuming a Gaussian posterior. Given our fiducial cosmology, this gives the following constraints on ratios of distances:

$$\begin{aligned}D_H(z_{\text{eff}})/r_d &= 8.72 \pm 0.14, \\ D_M(z_{\text{eff}})/r_d &= 38.80 \pm 0.76.\end{aligned}\quad (7.9)$$

It is important to note that we have not redone the SDSS analysis with our analysis pipeline. We have taken the  $(2 \times 2)$  Ly $\alpha$  BAO posterior from [dMb20](#), and combined it with our own posterior after taking into account an approximate cross-survey covariance that was computed as described in appendix F. Ignoring the cross-survey covariance would lead to an underestimate of the covariance of the combined result by 10%.

One could do a joint analyses of both surveys, starting by co-adding the roughly 100 000 spectra of quasars at  $z > 2.1$  observed with both telescopes. However, the modelling would be complicated, since there are several differences in the contaminants of both surveys. These differences include: correlated sky residuals extend to different transverse separations; DESI observes fainter quasars, and the effective quasar bias could be different; thanks to the higher spectral resolution of DESI, quasar redshift errors are smaller than in SDSS; the efficiency of the DLA finder might also be quite different, impacting the level of HCD contamination.

The combined “DESI + SDSS” BAO measurement has an uncertainty 20% smaller than the one from DESI DR1. As the DESI survey observes more and more data, the SDSS dataset will gradually add less and less to the joint analysis.

#### 7.4 Future work

There are several known issues that we have not addressed in this publication, and that we leave for future work.

- Non-linear evolution causes a small systematic shift on the BAO peak [\[108–110\]](#). At low redshift, this is expected to be a sub-percent bias, and at high redshift it is expected to be even lower, below our current systematic uncertainties of 0.5% coming from the analysis of mocks (it is a conservative estimate, see section 6.5).
- Relative velocities between dark matter and baryons can bias the position of the BAO peak [\[111\]](#). Studies using hydrodynamical simulations have shown that the bias should be small in the Ly $\alpha$  forest auto-correlation [\[112, 113\]](#), but we do not have equivalent studies for the cross-correlation with quasars.
- The combination of quasar redshift errors and quasar continuum errors can lead to spurious correlations in 3D analyses of the Ly $\alpha$  forest [\[114\]](#). In a companion paper [\[33\]](#), we have analysed synthetic data to show that the impact of this systematic on BAO

measurements is limited to  $0.1\sigma$  of the current statistical uncertainties. However, this could become a problem for the final analysis of DESI, or for other Ly $\alpha$  studies that obtain cosmological information from the full shape of the correlations [115–119].

- In this work we used mocks (synthetic datasets) derived from log-normal density fields. We are working on the development of more complex mocks, including mocks based on N-body simulations [100] and mocks based on Lagrangian Perturbation Theory; these mocks will include the non-linear broadening of the BAO peak discussed in figure 9.
- BAO analyses using galaxy samples at low redshift often use a *reconstruction* algorithm to undo some of the non-linear effects and improve the accuracy and precision of the BAO measurements [120, 121]. However, the reconstruction method is not recommended for surveys limited by shot noise (as opposed to cosmic variance), since the estimates of the displacement field would be noisy and could degrade the BAO precision. While this will be an interesting possibility for future Ly $\alpha$  surveys with higher quasar densities, we do not expect this method to improve the precision of the BAO measurement with the current DESI data set.

## 8 Conclusions

We have measured the three-dimensional correlations in the Ly $\alpha$  forest dataset from the first year of DESI data, as well as its cross-correlation with the position of DESI quasars. Using these correlations we have measured the BAO scale parallel ( $\alpha_{\parallel}$ ) and perpendicular ( $\alpha_{\perp}$ ) to the line of sight with a precision of 2.0% and 2.4% respectively. The statistical uncertainties on the BAO parameters from just one year of DESI observations are already smaller than the ones from [dMb20](#), who used 10 years of BOSS and eBOSS observations.

This analysis is the first Ly $\alpha$  BAO measurement that was fully blinded.<sup>29</sup> It is also the first time that the (small) cross-covariance between the auto-correlation of Ly $\alpha$  fluctuations and its cross-correlation with quasars is taken into account (see section 3.3). In a companion paper [33], we validated the analysis with 150 mocks mimicking the DESI DR1 Ly $\alpha$  sample. We also characterized the correlated noise introduced by the data processing pipeline and the imprint of foreground absorbers on the Ly $\alpha$  auto-correlation in [32]. In parallel, we set a long list of robustness tests that we needed to pass before we could *unblind* the measurements. These are discussed in sections 6.3 and 6.4.

The BAO measurement presented here can be translated into constraints on the following ratio of distances:  $D_H(z_{\text{eff}} = 2.33)/r_d = 8.52 \pm 0.17$  and  $D_M(z_{\text{eff}} = 2.33)/r_d = 39.71 \pm 0.95$ , where  $D_H = c/H(z)$ ,  $D_M(z)$  is the transverse comoving distance and  $r_d$  is the sound horizon at the drag epoch.

This publication is part of a series of publications presenting clustering measurements of the different tracers in DESI DR1 [5]. Besides this Ly $\alpha$  BAO measurement at  $z = 2.33$ , the clustering of galaxies and quasars at  $z < 2$  is presented in [122], the BAO measurements derived from this data set are described in [6], and the cosmological constraints from the

---

<sup>29</sup>The BOSS DR9 analysis of [13–15] was partially blinded, in the sense that during several months the BAO peak was masked when plotting the correlation function.

combined galaxy, quasar and Ly $\alpha$  forests BAO can be found in [9]. A complementary analysis of the two points clustering statistics of galaxies and quasars using a larger range of scales beyond the BAO scale will be presented in [123] with their corresponding cosmological constraints in [124].

In the near future, we will also present other cosmological analyses using the Ly $\alpha$  forest dataset from DESI DR1. We will extract non-BAO information from the full shape of 3D correlations on large, linear scales (see [116–119]), which should further improve the precision of these results. We will also constrain the linear matter power spectrum on small scales using the 1D power spectrum of fluctuations in the Ly $\alpha$  forest (see [125, 126]). DESI has completed approximately three years of observations and has collected more than half of its planned dataset. Upon completion of the five year survey, we expect the precision of the BAO results to improve by a factor of two.

**Data availability.** The data points corresponding to the most relevant figures in this paper are available at Zenodo.<sup>30</sup>

## Acknowledgments

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

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

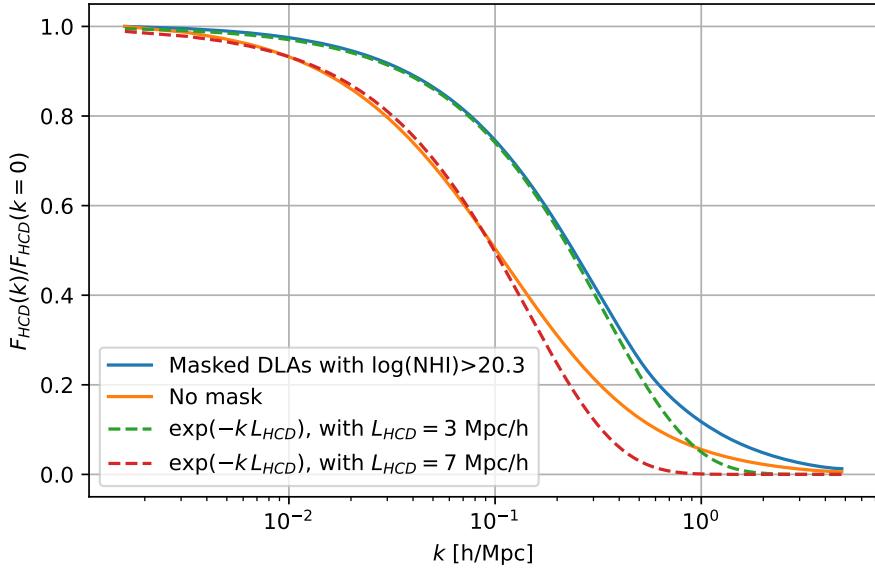
## A Alternative modelling of HCD contaminations

As discussed in section 4.6, the presence of high column density systems (HCDs) changes the flux correlation function by smearing the  $\delta$  field in the radial direction [84–86]. The contamination can be described by introducing a scale-dependent component to the bias that depends on  $k_{\parallel}$  that is added to the normal IGM-induced Ly $\alpha$  bias. Following [21] we write this term as  $b_{\text{HCD}}F_{\text{HCD}}(k_{\parallel})$  where

$$F_{\text{HCD}}(k_{\parallel}, z) = A(z) \int dN_{\text{HI}} f(N_{\text{HI}}, z) V(N_{\text{HI}}, k_{\parallel}, z) \quad (\text{A.1})$$

---

<sup>30</sup><https://doi.org/10.5281/zenodo.10799350>.



**Figure 13.** Comparison of the Voigt model for HCD contamination (solid curves) and the exponential model used in this work (dashed curves), for the HCD column-density distribution predicted by [127]. The orange solid line includes HCDs of all column densities and the blue solid line includes only those with  $\log(N_{\mathrm{HI}}) < 20.3$ , which is appropriate for perfect masking of DLAs. The dashed lines show that exponential models can reproduce the Voigt model, except at large  $k_{\parallel}$  where the Voigt model has a longer tail.

where  $V(N_{\mathrm{HI}}, k_{\parallel})$  is the Fourier transform of the Voigt profile for an HCD of column density  $N_{\mathrm{HI}}$ , and  $f(N_{\mathrm{HI}})$  is the column-density distribution of HCDs. The normalization factor  $A(z)$  can be chosen so that  $F_{\mathrm{HCD}}(k_{\parallel} = 0) = 1$ , in which case  $b_{\mathrm{HCD}}$  is proportional to the product of the HCD halo bias and the mean absorption caused by HCDs (see appendix B in [84] and eq. 4.19 in [85]).

Given our lack of precise knowledge of the HCD distribution  $f(N_{\mathrm{HI}})$ , following [dMdB20](#) we model  $F_{\mathrm{HCD}}(k_{\parallel}) = \exp(-L_{\mathrm{HCD}}k_{\parallel})$  as an exponential with unknown scale parameter  $L_{\mathrm{HCD}}$  that characterizes the typical size of unmasked HCDs.

We compare the functional forms of  $F_{\mathrm{HCD}}(k_{\parallel})$  in figure 13. The solid orange line shows the computation from eq. (A.1) when we use the column density distribution  $f(N_{\mathrm{HI}})$  from [127]. The solid blue line uses the same model, but it only integrates up to  $\log(N_{\mathrm{HI}}) = 20.3$  to mimic the effect of perfectly masking all DLAs. The dashed red and green lines show the exponential model for  $L_{\mathrm{HCD}} = 7$  and  $L_{\mathrm{HCD}} = 3 h^{-1} \mathrm{Mpc}$ , respectively, and they capture fairly well the suppression of power from eq. (A.1).

## B Nuisance parameters

As discussed in section 4, our model to describe the measured correlations has 17 free parameters, including the two BAO parameters  $(\alpha_{\perp}, \alpha_{\parallel})$  and 15 nuisance parameters that we marginalize over. The best-fit values and uncertainties for all 17 parameters are shown in table 5, together with the priors used in the analysis.

While most of the parameters are well constrained by the data and the priors are uninformative, we use informative Gaussian priors for two parameters. These are the linear bias

Parameter	Priors	Best fit		
		Combined	Ly $\alpha$ × Ly $\alpha$	Ly $\alpha$ × QSO
$\alpha_{\parallel}$	$\mathcal{U}[0.01, 2.00]$	$0.989 \pm 0.020$	$0.993^{+0.029}_{-0.032}$	$0.988^{+0.024}_{-0.025}$
$\alpha_{\perp}$	$\mathcal{U}[0.01, 2.00]$	$1.013 \pm 0.024$	$1.020^{+0.036}_{-0.037}$	$1.005 \pm 0.030$
$b_{\alpha}$	$\mathcal{U}[-2.00, 0.00]$	$-0.1078^{+0.0045}_{-0.0054}$	$-0.1078 \pm 0.0036$	$-0.099^{+0.015}_{-0.013}$
$\beta_{\alpha}$	$\mathcal{U}[0.00, 5.00]$	$1.743^{+0.074}_{-0.100}$	$1.745^{+0.076}_{-0.088}$	$2.07^{+0.23}_{-0.35}$
$10^3 b_{\text{SiII}(1190)}$	$\mathcal{U}[-500.00, 0.00]$	$-4.50 \pm 0.64$	$-5.53^{+0.85}_{-0.82}$	$-3.5 \pm 1.1$
$10^3 b_{\text{SiII}(1193)}$	$\mathcal{U}[-500.00, 0.00]$	$-3.05^{+0.63}_{-0.62}$	$-4.16^{+0.81}_{-0.79}$	$-1.76^{+1.13}_{-0.81}$
$10^3 b_{\text{SiII}(1260)}$	$\mathcal{U}[-500.00, 0.00]$	$-4.02 \pm 0.62$	$-4.63 \pm 0.91$	$-4.00 \pm 0.82$
$10^3 b_{\text{SiIII}(1207)}$	$\mathcal{U}[-500.00, 0.00]$	$-9.79 \pm 0.68$	$-10.80^{+0.74}_{-0.75}$	$-8.7 \pm 1.1$
$10^3 b_{\text{CIV(eff)}}$	$\mathcal{N}(-24.3, 1.5)^*$	$-24.3 \pm 1.5$	$-24.5 \pm 1.6$	
$b_{\text{HCD}}$	$\mathcal{U}[-0.20, 0.00]$	$-0.0563^{+0.0045}_{-0.0036}$	$-0.0582 \pm 0.0037$	$-0.053^{+0.013}_{-0.014}$
$\beta_{\text{HCD}}$	$\mathcal{N}(0.500, 0.090)$	$0.625 \pm 0.080$	$0.588^{+0.082}_{-0.081}$	$0.528^{+0.088}_{-0.094}$
$L_{\text{HCD}}(h^{-1}\text{Mpc})$	$\mathcal{U}[0.00, 40.00]$	$6.51^{+0.82}_{-0.96}$		
$b_Q$	$\mathcal{U}[0.00, 10.00]^*$	$3.408 \pm 0.048$		$3.49 \pm 0.10$
$\Delta r_{\parallel}(h^{-1}\text{Mpc})$	$\mathcal{U}[-3.00, 3.00]$	$0.066 \pm 0.058$		$0.077^{+0.061}_{-0.062}$
$\sigma_z(h^{-1}\text{Mpc})$	$\mathcal{U}[0.00, 15.00]$	$3.67 \pm 0.14$		$4.12^{+0.42}_{-0.43}$
$\xi_0^{\text{TP}}$	$\mathcal{U}[0.00, 2.00]$	$0.395 \pm 0.051$		$0.320^{+0.082}_{-0.083}$
$10^4 a_{\text{noise}}$	$\mathcal{U}[0.00, 100.00]$	$3.54 \pm 0.16$	$3.57 \pm 0.17$	
$N_{\text{bin}}$	—	9540	3180	6360
$N_{\text{param}}$	—	17	12	14
$\chi^2_{\text{min}}$	—	9624.36	3183.79	6427.41
p-value	—	0.23	0.42	0.23

**Table 5.** Priors, best-fit values (mean of the posterior) and uncertainties (68% credible intervals) for the 17 free parameters in the fits. When analysing the auto-correlation or the cross-correlation alone, we fix  $L_{\text{HCD}}$  to the best-fit value of the combination ( $L_{\text{HCD}} = 6.51 h^{-1} \text{ Mpc}$ ). This is necessary to break internal degeneracies, but it makes the p-value of these analyses difficult to interpret. When analysing the cross-correlation alone we also use an extra prior on the quasar bias parameter ( $b_Q = 3.5 \pm 0.1$ ) to break the degeneracy with the Ly $\alpha$  biases. Some parameters are not needed when fitting the auto-correlation or the cross-correlation alone.

of CIV absorption,  $b_{\text{CIV(eff)}}$ , and the RSD parameter of absorption caused by HCDs,  $\beta_{\text{HCD}}$ . When fitting the auto- and cross-correlation independently, we are not able to break the degeneracy between  $L_{\text{HCD}}$  and other nuisance parameters, and we fix this value to the best-fit value of the combined analysis ( $L_{\text{HCD}} = 6.51 h^{-1} \text{ Mpc}$ ). Moreover, when fitting the cross-correlation alone we also need to add an informative Gaussian prior on the quasar bias<sup>31</sup> of  $b_Q = 3.5 \pm 0.1$  to break the degeneracy between this parameter and the bias of the Ly $\alpha$  forest ( $b_{\alpha}$ ).

<sup>31</sup>Motivated by preliminary studies of quasar clustering in DESI DR1 [128].

Comparison of the different columns in table 5 shows that the best-fit values from the auto-correlations alone (including regions A and B) are in agreement with those from the fit of the cross-correlations with quasars (also including both regions). However, we would like to add a word of caution when interpreting these nuisance parameters. While we have extensively tested the robustness of the BAO results under different data splits and analysis settings, some of the nuisance parameters do vary significantly with reasonable changes in the analysis choices. For instance, depending on how aggressively we mask DLAs, we obtain different values for the parameters that model the contamination by HCDs ( $b_{\text{HCD}}$ ,  $\beta_{\text{HCD}}$  and  $L_{\text{HCD}}$ ), but the differences also propagate to other parameters that are degenerate with these, including  $b_\alpha$  and  $\beta_\alpha$ .

Finally, in figure 14 we show that none of the 15 nuisance parameters is correlated with any of the BAO parameters ( $\alpha_\perp$ ,  $\alpha_\parallel$ ).

## C Blinding

Our analysis validation was performed entirely on blinded data, with clearly defined requirements that needed to be achieved before unblinding (section 6). The main goal of our blinding strategy was to blind the BAO measurement in a way that does not impede the analysis process. We considered multiple blinding methods, including blinding at the catalog level (as done for the DESI galaxy BAO analyses, see [6]). However, the disadvantage of catalog-level blinding is that metal contamination results in very well-measured peaks along the line-of-sight due to  $\text{Ly}\alpha \times \text{Metal}$  correlations (see section 4.4). Any catalogue-level blinding would have also shifted the positions of these peaks, giving away the direction and magnitude of the blinding. We therefore instead developed a blinding method that only shifts the position of the BAO peak at the level of the measured correlation functions.

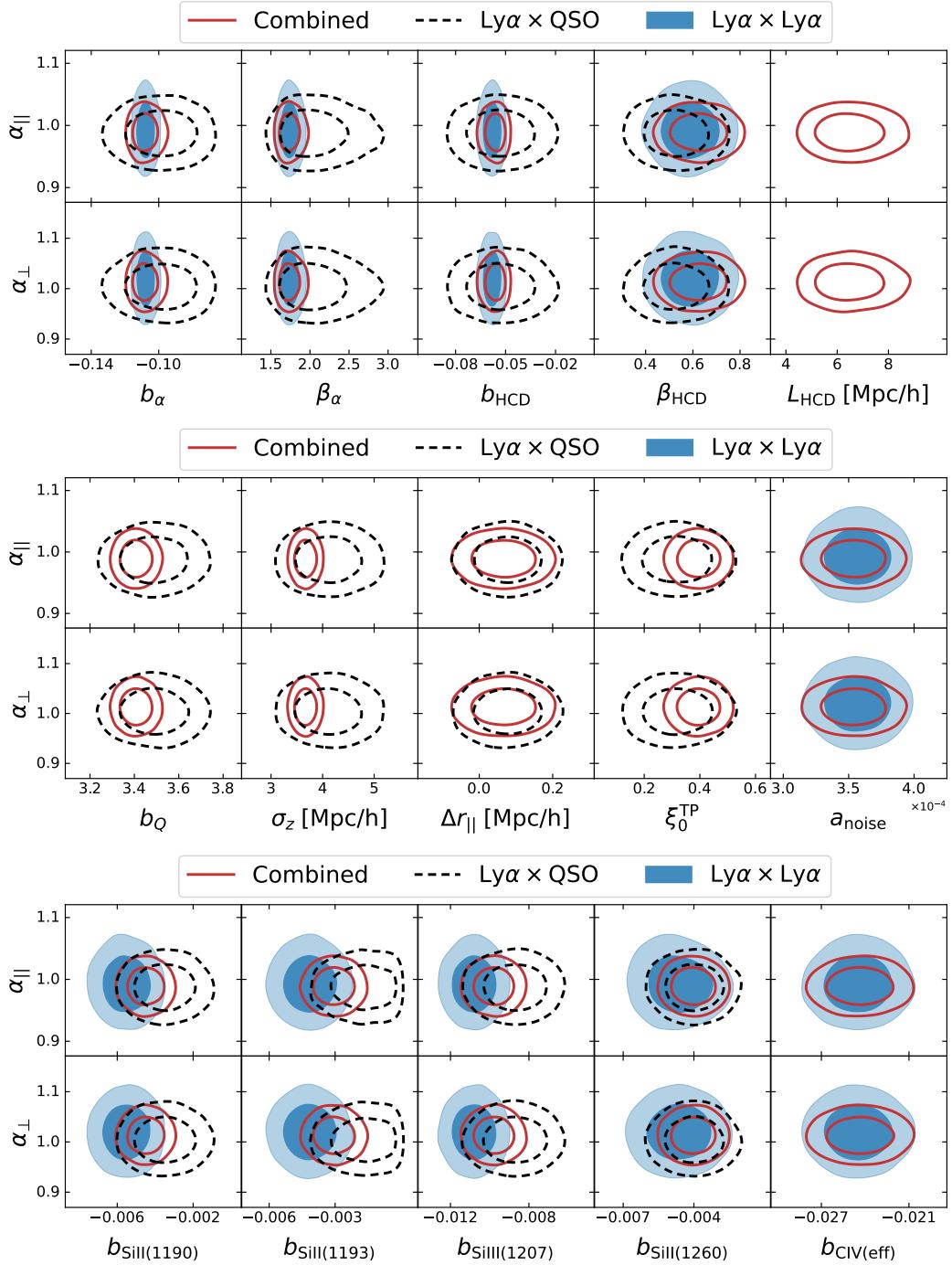
Our blinding strategy starts with a correlation function model  $\xi_t$  with all nuisance parameters set to their best-fit values from [27]. We then used **Vega** to compute the blinding template:

$$\xi_b = \xi_t(\alpha_\parallel = 1 + \Delta\alpha_\parallel, \alpha_\perp = 1 + \Delta\alpha_\perp) - \xi_t(\alpha_\parallel = 1, \alpha_\perp = 1), \quad (\text{C.1})$$

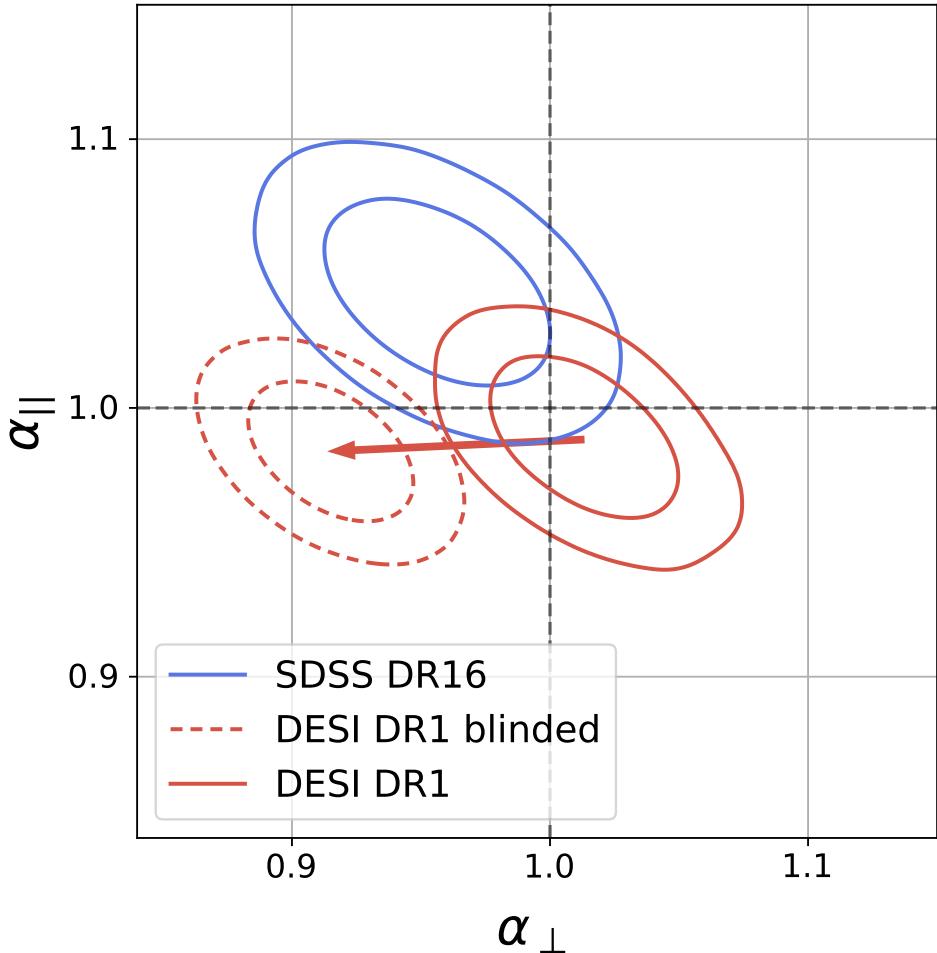
where  $\Delta\alpha_\parallel$  and  $\Delta\alpha_\perp$  were randomly drawn from Gaussian distributions with mean zero and variance equal to two times the uncertainties on  $\alpha_\parallel$  and  $\alpha_\perp$  from [27]. The random values used in the DR1 BAO analysis were  $\Delta\alpha_\parallel = -0.011$  and  $\Delta\alpha_\perp = -0.084$ , i.e., a fairly small shift along the line of sight and a very large blinding shift in the transverse BAO measurement, close to the maximum allowed by the blinding strategy. These values of  $\Delta\alpha_\parallel$  and  $\Delta\alpha_\perp$  were unknown to us and never stored anywhere. We did save the template<sup>32</sup> and our pipeline (**picca**) automatically added this  $\xi_b$  template to the correlation function at the moment of writing it to file, effectively blinding the BAO scale.

The random shift applied to the blinded measurements can be seen in figure 15, where we show the DESI  $\text{Ly}\alpha$  BAO results before (dotted red) and after (solid red) unblinding. The difference in the best fit BAO parameters before and after unblinding ( $\Delta\alpha_\parallel = -0.005$ ,  $\Delta\alpha_\perp = -0.098$ , red arrow in the figure) is consistent with the random shifts applied in the blinding ( $\Delta\alpha_\parallel = -0.011$ ,  $\Delta\alpha_\perp = -0.084$ ).

<sup>32</sup>The format used was carefully chosen to eliminate the possibility of the template being viewed accidentally.



**Figure 14.** Correlation between the BAO parameters ( $\alpha_{\perp}$ ,  $\alpha_{\parallel}$ ) and the 15 nuisance parameters, for the combined analysis (solid red), the cross-correlation alone (dashed black) and the auto-correlation alone (filled blue). Not all nuisance parameters are varied when fitting the auto- and cross-correlations separately. The BAO parameters are not strongly correlated with any of the nuisance parameters.

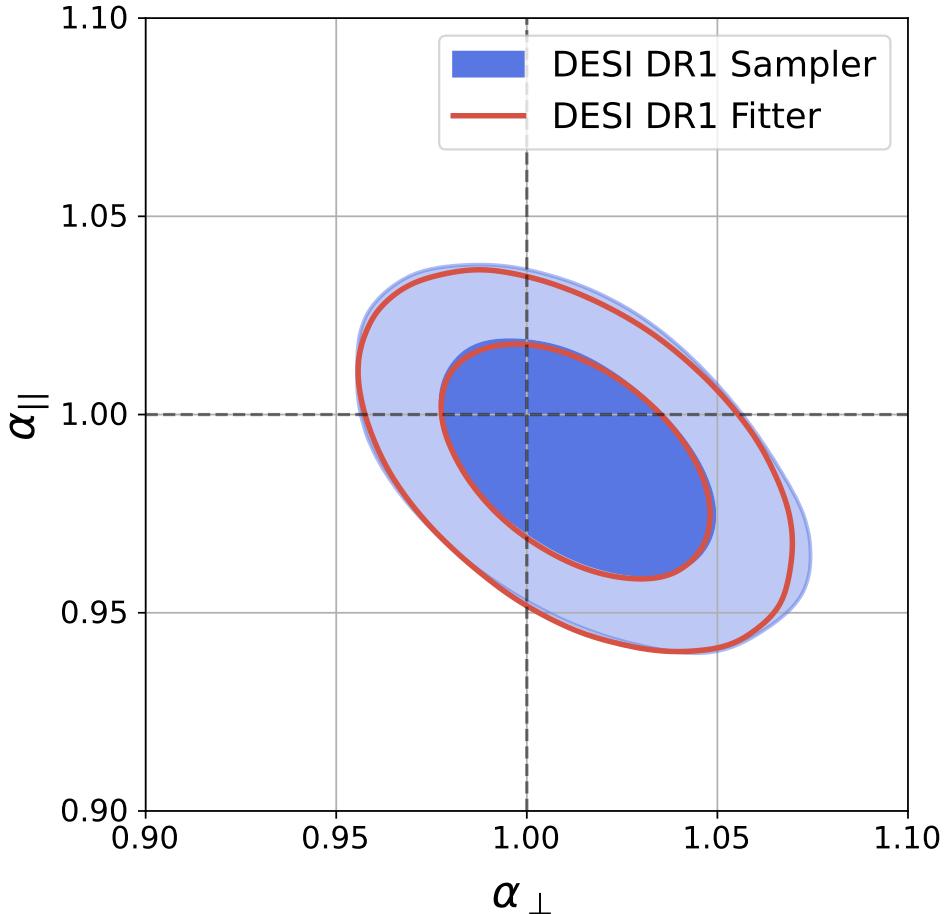


**Figure 15.** BAO parameters along ( $\alpha_{\parallel}$ ) and across the line of sight ( $\alpha_{\perp}$ ) from DESI DR1 (solid red) and from its blinded measurement (dashed red). The blinded measurement was in tension with Planck (dashed gray) and with the results from [dMdB20](#) using SDSS DR16 (solid blue).

We unblinded our analysis on December 8th, 2023, after passing the extensive validation tests described in section 6. We only made two minor changes in the methodology after unblinding. First, we fixed a small bug in `picca` related to the masking of BAL features in the Ly $\alpha$  region B, with an impact on the best-fit BAO parameters smaller than 0.1%. Second, we obtained a more accurate measurement of the CIV bias parameter from [32], with no impact on the BAO results.

#### D Comparison of sampling methods

The main results in this publication (including figure 7) were obtained using the Nested Sampler `Polychord` [93, 94]. The quoted parameter values are given by the mean of the posterior distribution, and the reported uncertainties are the 68% credible regions. However, computing the full posterior distribution is computationally intensive, and in most of the tests in section 6 we instead use a faster approximate method. This involves the use of the `iminuit` package [95] to find the maximum likelihood (minimum  $\chi^2$ ) point in parameter space.



**Figure 16.** 68% and 95% credible regions of BAO parameters along the line-of-sight ( $\alpha_{\parallel}$ ) and across the line-of-sight ( $\alpha_{\perp}$ ). The filled blue contours are based on the full-posterior distribution computed by `Polychord` and the solid red contours are based on the approximate Gaussian fit computed by `iminuit`.

Approximate Gaussian uncertainties are then computed by taking the second derivative with respect to parameters around the best-fit point [96]. In figure 16 we show that both methods lead to very similar BAO contours. Therefore, the faster approximate method is good enough to check for shifts in the BAO position as done in section 6.

In contrast, previous Ly $\alpha$  forest BAO analyses used a frequentist approach to obtain their main results [see e.g. 21, 22, 27]. This involved using the Profile Likelihood method to create a two-dimensional  $\chi^2$  grid of  $\alpha_{\parallel}$  and  $\alpha_{\perp}$ , and then calibrating the size of contours based on  $\Delta\chi^2$  values obtained from large sets of Monte Carlo simulations of the correlation functions. [74] found that BAO measurements obtained with this method agree well with Bayesian results based on the full posterior distribution. Therefore, in this work, we rely on the simpler and often faster method of sampling the full posterior for our main results.

## E Significance of BAO shifts

In section 6.4 we present an exhaustive list of robustness tests, where we look at the impact on the BAO parameters when changing different settings in the analysis. In most

of these variations, the dataset is exactly the same, and therefore we expect shifts in the BAO parameters to be caused exclusively by the analysis settings. Before *unblinding* the measurements we required that none of these variations caused a shift in the BAO parameters larger than a third of the statistical uncertainty from the results on (blinded) data.

In the second set of alternative analyses (in red) in figure 11, however, we also showed variations that caused minor changes in the dataset, by adding / removing quasars from the sample or by adding / removing Ly $\alpha$  pixels from the data vector. In these variations we relaxed the requirement, since changes in the dataset will cause statistical fluctuations in the measurement of the BAO parameters. This explains why the shifts shown in red in figure 11 are somewhat larger than those in the other variations.

Two of the larger shifts correspond to the variations “only quasar targets” and “weak BALs” (see section 6.4 for details on the variations). These variations cause the datasets to be 11% and 8% less constraining than the main analysis (based on the increase in the errorbars on the measured correlations). Following [129], we estimate the statistical fluctuations for these variations to be on average 0.28 and 0.33  $\sigma$  respectively. We conclude that the shifts on BAO parameters in these variations are therefore consistent with statistical fluctuations.

There is another variation that has caused a shift larger than the requirement, and this is the “mask-Lya redshift estimates” ( $\Delta\alpha_{\parallel} = 0.001$ ,  $\Delta\alpha_{\perp} = -0.013$ ). In this variation, we have used an alternative redshift estimator that masks wavelengths bluer than the Ly $\alpha$  emission line of the quasar. The rms of the differences in redshifts above  $z > 2$  is 443 km/s,<sup>33</sup> causing differences in pixel-quasar pairs of order  $4.2 h^{-1}$  Mpc (using the fiducial cosmology to compute  $H(z_{\text{eff}} = 2.33)$ ) and causing statistical fluctuations in the measurement of the cross-correlation (see appendix B of dMb20). Besides the expected effect in the cross-correlation, differences in the redshift estimates have a more subtle effect in the auto-correlation as well. The rest-frame wavelength of a given pixel changes with the quasar redshift, and we find that on average 2% of the pixels of a given Ly $\alpha$  forest are moved in or out of the restframe wavelength range used in the analysis, causing further statistical fluctuations.

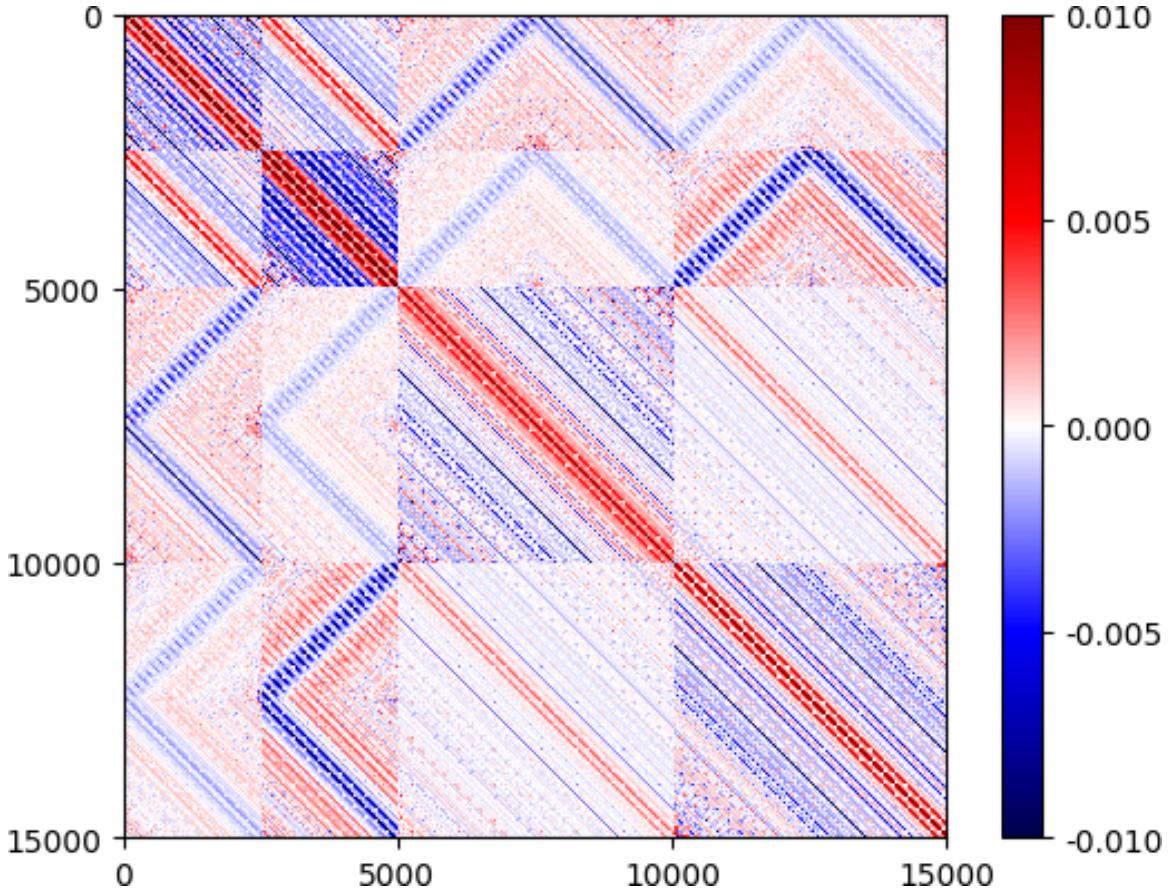
Finally, we want to assess whether the shift caused by the use of a different redshift estimates is a systematic shift or if it can be explained by statistical fluctuations. For this purpose, we perform a bootstrap analysis based on a random sampling with replacement of the HEALpix pixels used to compute the average correlation functions and we generate 1000 bootstrap samples for each of the two analyses. We then fit the BAO parameters for each of these samples and look at the distribution of shifts between the two analyses. We estimate with this technique that the statistical uncertainties on the shifts of  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  are of 0.007 and 0.009, respectively. We conclude from this study that the observed shifts of  $\Delta\alpha_{\parallel} = 0.001$  and  $\Delta\alpha_{\perp} = -0.013$  are consistent with statistical fluctuations.

## F Estimation of the DESI-SDSS covariance

The cross-covariance of the four correlation functions measured with the DESI DR1 Ly $\alpha$  forest dataset is shown in figure 6. As described in section 3.3, we compute a noisy estimate of this  $15\,000 \times 15\,000$  covariance from the scatter of the measurements obtained in different

---

<sup>33</sup>The distribution is not quite Gaussian, and 5.5% of the redshifts have changed by more than 1 000 km/s.



**Figure 17.** Correlation matrix corresponding to the cross-covariance of SDSS DR16 and DESI DR1 correlations. The first block of  $2500 \times 2500$  in the top left corresponds to the correlation matrix of the  $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{A})$  measurement of SDSS, while the second block of the same size corresponds to the same measurement in DESI. The third, larger block of  $5000 \times 5000$  corresponds to the  $\text{Ly}\alpha(\text{A}) \times \text{QSO}$  cross-correlation in SDSS, and the last block (bottom right) has the same measurement in DESI.

HEALPix pixels. Following [27, 31], we smooth the correlation matrix so that it is invertible and we can use it to define a likelihood function.

Here we use the same method to compute the cross-covariance between the DESI DR1 correlations functions and those measured in [dMdB20](#) using the SDSS DR16 Ly $\alpha$  dataset. A matrix describing the covariance of all eight correlations would be  $30\,000 \times 30\,000$ . However, given that the measurements using the B region carry a small amount of information (see the bottom right panel in figure 10) we ignore them in this appendix. In figure 17 we show the correlation matrix of SDSS DR16 and DESI DR1 measurements of  $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{A})$  and  $\text{Ly}\alpha(\text{A}) \times \text{QSO}$ .

Using this cross-covariance and the best-fit model from the combined analysis from table 5, we generated 1 000 Monte Carlo realisations of the two main correlation functions ( $\text{Ly}\alpha(\text{A}) \times \text{Ly}\alpha(\text{A})$  and  $\text{Ly}\alpha(\text{A}) \times \text{QSO}$ ) of SDSS and of DESI, along with the correct cross-covariance between the surveys. We then minimised the likelihood for the SDSS and DESI correlation functions separately, and looked at the correlation between the best-fit BAO

parameters  $(\alpha_{\perp}, \alpha_{\parallel})$  to obtain the correlation matrix of the four BAO parameters:

$$\begin{pmatrix} 1 & \rho(\alpha_{\perp}^S, \alpha_{\parallel}^S) & \rho(\alpha_{\perp}^S, \alpha_{\perp}^D) & \rho(\alpha_{\perp}^S, \alpha_{\parallel}^D) \\ \rho(\alpha_{\parallel}^S, \alpha_{\perp}^S) & 1 & \rho(\alpha_{\parallel}^S, \alpha_{\perp}^D) & \rho(\alpha_{\parallel}^S, \alpha_{\parallel}^D) \\ \rho(\alpha_{\perp}^D, \alpha_{\perp}^S) & \rho(\alpha_{\perp}^D, \alpha_{\parallel}^S) & 1 & \rho(\alpha_{\perp}^D, \alpha_{\parallel}^D) \\ \rho(\alpha_{\parallel}^D, \alpha_{\perp}^S) & \rho(\alpha_{\parallel}^D, \alpha_{\parallel}^S) & \rho(\alpha_{\parallel}^D, \alpha_{\perp}^D) & 1 \end{pmatrix} = \begin{pmatrix} 1 & -0.53 & 0.09 & -0.05 \\ -0.53 & 1 & -0.05 & 0.10 \\ 0.09 & -0.05 & 1 & -0.50 \\ -0.05 & 0.10 & -0.50 & 1 \end{pmatrix}, \quad (\text{F.1})$$

where the superscript  $S$  ( $D$ ) refers to SDSS (DESI) measurements of BAO.

The correlation of the BAO parameters from eBOSS and DESI can be computed separately from their posteriors, as discussed in the main text. For this reason, we modify the matrix above and instead use  $\rho(\alpha_{\perp}^S, \alpha_{\parallel}^S) = -0.45$  and  $\rho(\alpha_{\perp}^D, \alpha_{\parallel}^D) = -0.48$ . We use this modified correlation matrix, and the variances of each measurement reported by each survey, to build a  $4 \times 4$  multi-survey covariance  $C$  for the multi-survey data vector  $d = (\alpha_{\perp}^S, \alpha_{\parallel}^S, \alpha_{\perp}^D, \alpha_{\parallel}^D)$ .

We use these results to compute a combined BAO measurement  $d^{\text{DS}} = (\alpha_{\perp}^{\text{DS}}, \alpha_{\parallel}^{\text{DS}})$  with covariance  $C_{\text{DS}}$  using linear algebra:

$$C_{\text{DS}}^{-1} = S^T C^{-1} S \quad \text{and} \quad C_{\text{DS}}^{-1} d^{\text{DS}} = S^T C^{-1} d, \quad (\text{F.2})$$

where we have defined the matrix  $S$  as:

$$S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (\text{F.3})$$

Using these equations we obtain the following combined BAO measurements:

$$\begin{aligned} \alpha_{\perp}^{\text{DESI+SDSS}} &= 0.990 \pm 0.019, \\ \alpha_{\parallel}^{\text{DESI+SDSS}} &= 1.012 \pm 0.016, \end{aligned} \quad (\text{F.4})$$

with correlation coefficient of  $\rho = -0.47$ .

## References

- [1] D.H. Weinberg et al., *Observational Probes of Cosmic Acceleration*, *Phys. Rept.* **530** (2013) 87 [[arXiv:1201.2434](https://arxiv.org/abs/1201.2434)] [[INSPIRE](https://inspirehep.net/search?p=find+EPRINT+arXiv:1201.2434)].
- [2] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [[arXiv:1807.06209](https://arxiv.org/abs/1807.06209)] [[INSPIRE](https://inspirehep.net/search?p=find+EPRINT+arXiv:1807.06209)].
- [3] EBOSS collaboration, *Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory*, *Phys. Rev. D* **103** (2021) 083533 [[arXiv:2007.08991](https://arxiv.org/abs/2007.08991)] [[INSPIRE](https://inspirehep.net/search?p=find+EPRINT+arXiv:2007.08991)].
- [4] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, [arXiv:1611.00036](https://arxiv.org/abs/1611.00036) [[INSPIRE](https://inspirehep.net/search?p=find+EPRINT+arXiv:1611.00036)].
- [5] DESI collaboration, *DESI 2024 I: Data Release 1 of the Dark Energy Spectroscopic Instrument*, in preparation (2025).

- [6] DESI collaboration, *DESI 2024 III: Baryon Acoustic Oscillations from Galaxies and Quasars*, [arXiv:2404.03000](https://arxiv.org/abs/2404.03000) [INSPIRE].
- [7] M. McQuinn, *The Evolution of the Intergalactic Medium*, *Ann. Rev. Astron. Astrophys.* **54** (2016) 313 [[arXiv:1512.00086](https://arxiv.org/abs/1512.00086)] [INSPIRE].
- [8] BOSS collaboration, *The Lyman-alpha forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data*, *JCAP* **09** (2011) 001 [[arXiv:1104.5244](https://arxiv.org/abs/1104.5244)] [INSPIRE].
- [9] DESI collaboration, *DESI 2024 VI: Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations*, [arXiv:2404.03002](https://arxiv.org/abs/2404.03002) [INSPIRE].
- [10] SDSS collaboration, *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies*, *Astrophys. J.* **633** (2005) 560 [[astro-ph/0501171](https://arxiv.org/abs/astro-ph/0501171)] [INSPIRE].
- [11] 2DFGRS collaboration, *The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications*, *Mon. Not. Roy. Astron. Soc.* **362** (2005) 505 [[astro-ph/0501174](https://arxiv.org/abs/astro-ph/0501174)] [INSPIRE].
- [12] BOSS collaboration, *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *Astron. J.* **145** (2013) 10 [[arXiv:1208.0022](https://arxiv.org/abs/1208.0022)] [INSPIRE].
- [13] BOSS collaboration, *Baryon Acoustic Oscillations in the Ly- $\alpha$  forest of BOSS quasars*, *Astron. Astrophys.* **552** (2013) A96 [[arXiv:1211.2616](https://arxiv.org/abs/1211.2616)] [INSPIRE].
- [14] BOSS collaboration, *Measurement of Baryon Acoustic Oscillations in the Lyman-alpha Forest Fluctuations in BOSS Data Release 9*, *JCAP* **04** (2013) 026 [[arXiv:1301.3459](https://arxiv.org/abs/1301.3459)] [INSPIRE].
- [15] BOSS collaboration, *Fitting Methods for Baryon Acoustic Oscillations in the Lyman- $\alpha$  Forest Fluctuations in BOSS Data Release 9*, *JCAP* **03** (2013) 024 [[arXiv:1301.3456](https://arxiv.org/abs/1301.3456)] [INSPIRE].
- [16] BOSS collaboration, *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems*, *Astron. J.* **142** (2011) 72 [[arXiv:1101.1529](https://arxiv.org/abs/1101.1529)] [INSPIRE].
- [17] BOSS collaboration, *The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey*, *Astrophys. J. Suppl.* **203** (2012) 21 [[arXiv:1207.7137](https://arxiv.org/abs/1207.7137)] [INSPIRE].
- [18] BOSS collaboration, *The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III*, *Astrophys. J. Suppl.* **219** (2015) 12 [[arXiv:1501.00963](https://arxiv.org/abs/1501.00963)] [INSPIRE].
- [19] BOSS collaboration, *Quasar-Lyman  $\alpha$  Forest Cross-Correlation from BOSS DR11: Baryon Acoustic Oscillations*, *JCAP* **05** (2014) 027 [[arXiv:1311.1767](https://arxiv.org/abs/1311.1767)] [INSPIRE].
- [20] BOSS collaboration, *Baryon acoustic oscillations in the Ly $\alpha$  forest of BOSS DR11 quasars*, *Astron. Astrophys.* **574** (2015) A59 [[arXiv:1404.1801](https://arxiv.org/abs/1404.1801)] [INSPIRE].
- [21] BOSS collaboration, *Measurement of baryon acoustic oscillation correlations at  $z = 2.3$  with SDSS DR12 Ly $\alpha$ -Forests*, *Astron. Astrophys.* **603** (2017) A12 [[arXiv:1702.00176](https://arxiv.org/abs/1702.00176)] [INSPIRE].
- [22] BOSS collaboration, *Baryon acoustic oscillations from the complete SDSS-III Ly $\alpha$ -quasar cross-correlation function at  $z = 2.4$* , *Astron. Astrophys.* **608** (2017) A130 [[arXiv:1708.02225](https://arxiv.org/abs/1708.02225)] [INSPIRE].
- [23] EBOSS collaboration, *Baryon acoustic oscillations at  $z = 2.34$  from the correlations of Ly $\alpha$  absorption in eBOSS DR14*, *Astron. Astrophys.* **629** (2019) A85 [[arXiv:1904.03400](https://arxiv.org/abs/1904.03400)] [INSPIRE].

[24] eBOSS collaboration, *Baryon acoustic oscillations from the cross-correlation of Ly $\alpha$  absorption and quasars in eBOSS DR14*, *Astron. Astrophys.* **629** (2019) A86 [[arXiv:1904.03430](https://arxiv.org/abs/1904.03430)] [[INSPIRE](#)].

[25] eBOSS collaboration, *The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *Astron. J.* **151** (2016) 44 [[arXiv:1508.04473](https://arxiv.org/abs/1508.04473)] [[INSPIRE](#)].

[26] eBOSS collaboration, *The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra*, *Astrophys. J. Suppl.* **249** (2020) 3 [[arXiv:1912.02905](https://arxiv.org/abs/1912.02905)] [[INSPIRE](#)].

[27] eBOSS collaboration, *The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Ly $\alpha$  Forests*, *Astrophys. J.* **901** (2020) 153 [[arXiv:2007.08995](https://arxiv.org/abs/2007.08995)] [[INSPIRE](#)].

[28] E. Chaussidon et al., *Target Selection and Validation of DESI Quasars*, *Astrophys. J.* **944** (2023) 107 [[arXiv:2208.08511](https://arxiv.org/abs/2208.08511)] [[INSPIRE](#)].

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

[30] DESI collaboration, *The Lyman- $\alpha$  forest catalogue from the Dark Energy Spectroscopic Instrument Early Data Release*, *Mon. Not. Roy. Astron. Soc.* **528** (2024) 6666 [[arXiv:2306.06312](https://arxiv.org/abs/2306.06312)] [[INSPIRE](#)].

[31] C. Gordon et al., *3D Correlations in the Lyman- $\alpha$  Forest from Early DESI Data*, *JCAP* **11** (2023) 045 [[arXiv:2308.10950](https://arxiv.org/abs/2308.10950)] [[INSPIRE](#)].

[32] DESI collaboration, *Characterization of contaminants in the Lyman-alpha forest auto-correlation with DESI*, [arXiv:2404.03003](https://arxiv.org/abs/2404.03003) [[INSPIRE](#)].

[33] DESI collaboration, *Validation of the DESI 2024 Ly $\alpha$  forest BAO analysis using synthetic datasets*, [arXiv:2404.03004](https://arxiv.org/abs/2404.03004) [[INSPIRE](#)].

[34] H.K. Herrera-Alcantar et al., *Synthetic spectra for Lyman- $\alpha$  forest analysis in the Dark Energy Spectroscopic Instrument*, [arXiv:2401.00303](https://arxiv.org/abs/2401.00303) [[INSPIRE](#)].

[35] DESI collaboration, *The Optical Corrector for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **168** (2024) 95 [[arXiv:2306.06310](https://arxiv.org/abs/2306.06310)] [[INSPIRE](#)].

[36] DESI collaboration, *The Robotic Multiobject Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI)*, *Astron. J.* **165** (2023) 9 [[arXiv:2205.09014](https://arxiv.org/abs/2205.09014)] [[INSPIRE](#)].

[37] DESI collaboration, *The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **165** (2023) 144 [[arXiv:2209.14482](https://arxiv.org/abs/2209.14482)] [[INSPIRE](#)].

[38] DESI collaboration, *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **164** (2022) 207 [[arXiv:2205.10939](https://arxiv.org/abs/2205.10939)] [[INSPIRE](#)].

[39] DESI collaboration, *Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **167** (2024) 62 [[arXiv:2306.06307](https://arxiv.org/abs/2306.06307)] [[INSPIRE](#)].

[40] A.D. Myers et al., *The Target-selection Pipeline for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **165** (2023) 50 [[arXiv:2208.08518](https://arxiv.org/abs/2208.08518)] [[INSPIRE](#)].

[41] DESI collaboration, *Survey Operations for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **166** (2023) 259 [[arXiv:2306.06309](https://arxiv.org/abs/2306.06309)] [[INSPIRE](#)].

[42] K.M. Górski et al., *HEALPix — A Framework for high resolution discretization, and fast analysis of data distributed on the sphere*, *Astrophys. J.* **622** (2005) 759 [[astro-ph/0409513](https://arxiv.org/abs/astro-ph/0409513)] [[INSPIRE](#)].

[43] C. Yèche et al., *Preliminary Target Selection for the DESI Quasar (QSO) Sample*, *Res. Notes AAS* **4** (2020) 179 [[arXiv:2010.11280](https://arxiv.org/abs/2010.11280)] [[INSPIRE](#)].

[44] S. Bailey et al., *Redrock: Spectroscopic Classification and Redshift Fitting for the Dark Energy Spectroscopic Instrument*, in preparation (2024).

[45] N. Busca and C. Balland, *QuasarNET: Human-level spectral classification and redshifting with Deep Neural Networks*, [arXiv:1808.09955](https://arxiv.org/abs/1808.09955) [[INSPIRE](#)].

[46] J. Farr, A. Font-Ribera and A. Pontzen, *Optimal strategies for identifying quasars in DESI*, *JCAP* **11** (2020) 015 [[arXiv:2007.10348](https://arxiv.org/abs/2007.10348)] [[INSPIRE](#)].

[47] DESI collaboration, *Performance of the Quasar Spectral Templates for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **166** (2023) 66 [[arXiv:2305.10426](https://arxiv.org/abs/2305.10426)] [[INSPIRE](#)].

[48] EBOSS collaboration, *Measurements of effective optical depth in the Ly $\alpha$  forest from the BOSS DR12 Quasar sample*, [arXiv:1904.01110](https://arxiv.org/abs/1904.01110) [[DOI:10.3847/1538-4357/ab76bd](https://doi.org/10.3847/1538-4357/ab76bd)] [[INSPIRE](#)].

[49] A. Bault et al., *Impact of Systematic Redshift Errors on the Cross-correlation of the Lyman- $\alpha$  Forest with Quasars at Small Scales Using DESI Early Data*, [arXiv:2402.18009](https://arxiv.org/abs/2402.18009) [[INSPIRE](#)].

[50] D. Parks, J.X. Prochaska, S. Dong and Z. Cai, *Deep learning of quasar spectra to discover and characterize damped Ly $\alpha$  systems*, *Mon. Not. Roy. Astron. Soc.* **476** (2018) 1151 [[arXiv:1709.04962](https://arxiv.org/abs/1709.04962)].

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

[52] B. Wang et al., *Deep Learning of Dark Energy Spectroscopic Instrument Mock Spectra to Find Damped Ly $\alpha$  Systems*, *Astrophys. J. Supp.* **259** (2022) 28 [[arXiv:2201.00827](https://arxiv.org/abs/2201.00827)] [[INSPIRE](#)].

[53] C.B. Foltz et al., *On the Fraction of Optically-Selected QSOs with Broad Absorption Lines in Their Spectra*, *Bull. Am. Astron. Soc.* **22** (1990) 806.

[54] J.R. Trump et al., *A catalog of broad absorption line quasars from the Sloan Digital Sky Survey third data release*, *Astrophys. J. Suppl.* **165** (2006) 1 [[astro-ph/0603070](https://arxiv.org/abs/astro-ph/0603070)] [[INSPIRE](#)].

[55] L. Mas-Ribas and R. Mauland, *The Ubiquitous Imprint of Radiative Acceleration in the Mean Absorption Spectrum of Quasar Outflows*, *Astrophys. J.* **886** (2019) 151.

[56] L.Á. García et al., *Analysis of the impact of broad absorption lines on quasar redshift measurements with synthetic observations*, *Mon. Not. Roy. Astron. Soc.* **526** (2023) 4848 [[arXiv:2304.05855](https://arxiv.org/abs/2304.05855)] [[INSPIRE](#)].

[57] S. Filbert et al., *Broad absorption line quasars in the Dark Energy Spectroscopic Instrument Early Data Release*, *Mon. Not. Roy. Astron. Soc.* **532** (2024) 3669 [[arXiv:2309.03434](https://arxiv.org/abs/2309.03434)] [[INSPIRE](#)].

[58] SDSS collaboration, *Unusual Broad Absorption Line Quasars from the Sloan Digital Sky Survey*, *Astrophys. J. Suppl.* **141** (2002) 267 [[astro-ph/0203252](https://arxiv.org/abs/astro-ph/0203252)] [[INSPIRE](#)].

[59] R.J. Weymann, S.L. Morris, C.B. Foltz and P.C. Hewett, *Comparisons of the emission-line and continuum properties of broad absorption line and normal quasi-stellar objects*, *Astrophys. J.* **373** (1991) 23 [[INSPIRE](#)].

[60] EBOSS collaboration, *Modeling the Spectral Diversity of Quasars in the Sixteenth Data Release from the Sloan Digital Sky Survey*, *Astron. J.* **163** (2022) 110 [[arXiv:2110.07748](https://arxiv.org/abs/2110.07748)] [[INSPIRE](#)].

[61] L. Ennesser, P. Martini, A. Font-Ribera and I. Pérez-Ràfols, *The impact and mitigation of broad-absorption-line quasars in Lyman $\alpha$  forest correlations*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 3514 [[arXiv:2111.09439](https://arxiv.org/abs/2111.09439)] [[INSPIRE](#)].

[62] P. Martini et al., *Validation of the DESI 2024 Lyman Alpha Forest BAL Masking Strategy*, [arXiv:2405.09737](https://arxiv.org/abs/2405.09737) [[INSPIRE](#)].

[63] C.R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357 [[arXiv:2006.10256](https://arxiv.org/abs/2006.10256)] [[INSPIRE](#)].

[64] P. Virtanen et al., *SciPy 1.0 — Fundamental Algorithms for Scientific Computing in Python*, *Nature Meth.* **17** (2020) 261 [[arXiv:1907.10121](https://arxiv.org/abs/1907.10121)] [[INSPIRE](#)].

[65] ASTROPY collaboration, *Astropy: A Community Python Package for Astronomy*, *Astron. Astrophys.* **558** (2013) A33 [[arXiv:1307.6212](https://arxiv.org/abs/1307.6212)] [[INSPIRE](#)].

[66] ASTROPY collaboration, *The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package*, *Astron. J.* **156** (2018) 123 [[arXiv:1801.02634](https://arxiv.org/abs/1801.02634)] [[INSPIRE](#)].

[67] ASTROPY collaboration, *The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package*, *Astrophys. J.* **935** (2022) 167 [[arXiv:2206.14220](https://arxiv.org/abs/2206.14220)] [[INSPIRE](#)].

[68] L. Dalcin and Y.-L.L. Fang, *mpi4py: Status Update After 12 Years of Development*, *Comput. Sci. Eng.* **23** (2021) 47.

[69] A. Zonca et al., *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python*, *J. Open Source Softw.* **4** (2019) 1298 [[INSPIRE](#)].

[70] J.D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Comput. Sci. Eng.* **9** (2007) 90 [[INSPIRE](#)].

[71] A. Lewis, *GetDist: a Python package for analysing Monte Carlo samples*, [arXiv:1910.13970](https://arxiv.org/abs/1910.13970) [[INSPIRE](#)].

[72] S.K. Lam, A. Pitrou and S. Seibert, *Numba: a LLVM-based Python JIT compiler*, in the proceedings of the *International Conference for High Performance Computing, Networking, Storage and Analysis: HPC Transforms*, Austin, Texas, U.S.A., 15 November 2015 [[DOI:10.1145/2833157.2833162](https://doi.org/10.1145/2833157.2833162)] [[INSPIRE](#)].

[73] H. du Mas des Bourboux et al., *picca: Package for Igm Cosmological-Correlations Analyses*. Astrophysics Source Code Library, [ascl:2106.018](https://ascl.net/2106.018) (2021).

[74] A. Cuceu, A. Font-Ribera and B. Joachimi, *Bayesian methods for fitting Baryon Acoustic Oscillations in the Lyman- $\alpha$  forest*, *JCAP* **07** (2020) 035 [[arXiv:2004.02761](https://arxiv.org/abs/2004.02761)] [[INSPIRE](#)].

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

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

[77] EBOSS collaboration, *The Extended Baryon Oscillation Spectroscopic Survey: Measuring the Cross-correlation between the MgII Flux Transmission Field and Quasars and Galaxies at  $z = 0.59$* , *Astrophys. J.* **878** (2019) 47 [[arXiv:1901.01950](https://arxiv.org/abs/1901.01950)] [[INSPIRE](#)].

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

[79] P. McDonald et al., *The observed probability distribution function, power spectrum, and correlation function of the transmitted flux in the Lyman-alpha forest*, *Astrophys. J.* **543** (2000) 1 [[astro-ph/9911196](#)] [[INSPIRE](#)].

[80] P. McDonald, *Toward a measurement of the cosmological geometry at  $z \sim 2$ : predicting Ly $\alpha$  forest correlation in three dimensions, and the potential of future data sets*, *Astrophys. J.* **585** (2003) 34 [[astro-ph/0108064](#)] [[INSPIRE](#)].

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

[82] EBOSS collaboration, *Metal Lines Associated with the Ly $\alpha$  Forest from eBOSS Data*, *Astrophys. J.* **935** (2022) 121 [[arXiv:2206.11385](#)] [[INSPIRE](#)].

[83] S. Morrison et al., *A strong blend in the morning: studying the circumgalactic medium before cosmic noon with strong, blended Ly  $\alpha$  forest systems*, *Mon. Not. Roy. Astron. Soc.* **532** (2024) 32 [[arXiv:2309.06813](#)] [[INSPIRE](#)].

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

[85] A. Font-Ribera and J. Miralda-Escudé, *The Effect of High Column Density Systems on the Measurement of the Lyman  $\alpha$  Forest Correlation Function*, *JCAP* **07** (2012) 028 [[arXiv:1205.2018](#)] [[INSPIRE](#)].

[86] K.K. Rogers et al., *Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers*, *Mon. Not. Roy. Astron. Soc.* **476** (2018) 3716 [[arXiv:1711.06275](#)] [[INSPIRE](#)].

[87] T. Tan et al., *Modeling of the High Column Density systems in The Lyman-Alpha Forest*, in preparation (2024).

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

[89] I. Pérez-Ràfols et al., *The cross-correlation of galaxies in absorption with the Lyman  $\alpha$  forest*, *Mon. Not. Roy. Astron. Soc.* **524** (2023) 1464 [[arXiv:2210.02973](#)] [[INSPIRE](#)].

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

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

[92] J.J. Givans et al., *Non-linearities in the Lyman- $\alpha$  forest and in its cross-correlation with dark matter halos*, *JCAP* **09** (2022) 070 [[arXiv:2205.00962](#)] [[INSPIRE](#)].

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

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

[95] H. Dembinski et al., *scikit-hep/iminuit*, <https://github.com/scikit-hep/iminuit>.

[96] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, *Comput. Phys. Commun.* **10** (1975) 343 [[INSPIRE](#)].

[97] J. Farr et al., *LyaCoLoRe: synthetic datasets for current and future Lyman- $\alpha$  forest BAO surveys*, *JCAP* **03** (2020) 068 [[arXiv:1912.02763](https://arxiv.org/abs/1912.02763)] [[INSPIRE](#)].

[98] 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](https://arxiv.org/abs/2111.05069)] [[INSPIRE](#)].

[99] T. Etourneau et al., *Mock data sets for the Eboss and DESI Lyman- $\alpha$  forest surveys*, *JCAP* **05** (2024) 077 [[arXiv:2310.18996](https://arxiv.org/abs/2310.18996)] [[INSPIRE](#)].

[100] B. Hadzhiyska et al., *Planting a Lyman alpha forest on AbacusSummit*, *Mon. Not. Roy. Astron. Soc.* **524** (2023) 1008 [[arXiv:2305.08899](https://arxiv.org/abs/2305.08899)] [[INSPIRE](#)].

[101] DESI collaboration, *Overview of the DESI Legacy Imaging Surveys*, *Astron. J.* **157** (2019) 168 [[arXiv:1804.08657](https://arxiv.org/abs/1804.08657)] [[INSPIRE](#)].

[102] H. Zou et al., *Project Overview of the Beijing-Arizona Sky Survey*, *Publ. Astron. Soc. Pac.* **129** (2017) 064101 [[arXiv:1702.03653](https://arxiv.org/abs/1702.03653)].

[103] D. Schlegel et al., *DESI Legacy Imaging Surveys Data Release 9*, in preparation (2024).

[104] J.A. Baldwin, *Luminosity Indicators in the Spectra of Quasi-Stellar Objects*, *Astrophys. J.* **214** (1977) 679.

[105] A. Pontzen, *Scale-dependent bias in the baryonic-acoustic-oscillation-scale intergalactic neutral hydrogen*, *Phys. Rev. D* **89** (2014) 083010 [[arXiv:1402.0506](https://arxiv.org/abs/1402.0506)] [[INSPIRE](#)].

[106] S. Gontcho A Gontcho, J. Miralda-Escudé and N.G. Busca, *On the effect of the ionizing background on the Ly $\alpha$  forest autocorrelation function*, *Mon. Not. Roy. Astron. Soc.* **442** (2014) 187 [[arXiv:1404.7425](https://arxiv.org/abs/1404.7425)] [[INSPIRE](#)].

[107] C. Alcock and B. Paczynski, *An evolution free test for non-zero cosmological constant*, *Nature* **281** (1979) 358 [[INSPIRE](#)].

[108] M. Crocce and R. Scoccimarro, *Nonlinear Evolution of Baryon Acoustic Oscillations*, *Phys. Rev. D* **77** (2008) 023533 [[arXiv:0704.2783](https://arxiv.org/abs/0704.2783)] [[INSPIRE](#)].

[109] R.E. Smith, R. Scoccimarro and R.K. Sheth, *Eppur Si Muove: On The Motion of the Acoustic Peak in the Correlation Function*, *Phys. Rev. D* **77** (2008) 043525 [[astro-ph/0703620](https://arxiv.org/abs/astro-ph/0703620)] [[INSPIRE](#)].

[110] H.-J. Seo, E.R. Siegel, D.J. Eisenstein and M. White, *Non-linear structure formation and the acoustic scale*, *Astrophys. J.* **686** (2008) 13 [[arXiv:0805.0117](https://arxiv.org/abs/0805.0117)] [[INSPIRE](#)].

[111] D. Tseliakhovich and C. Hirata, *Relative velocity of dark matter and baryonic fluids and the formation of the first structures*, *Phys. Rev. D* **82** (2010) 083520 [[arXiv:1005.2416](https://arxiv.org/abs/1005.2416)] [[INSPIRE](#)].

[112] C.M. Hirata, *Small-scale structure and the Lyman- $\alpha$  forest baryon acoustic oscillation feature*, *Mon. Not. Roy. Astron. Soc.* **474** (2018) 2173 [[arXiv:1707.03358](https://arxiv.org/abs/1707.03358)] [[INSPIRE](#)].

[113] J.J. Givans and C.M. Hirata, *Redshift-space streaming velocity effects on the Lyman- $\alpha$  forest baryon acoustic oscillation scale*, *Phys. Rev. D* **102** (2020) 023515 [[arXiv:2002.12296](https://arxiv.org/abs/2002.12296)] [[INSPIRE](#)].

[114] S. Youles et al., *The effect of quasar redshift errors on Lyman- $\alpha$  forest correlation functions*, *Mon. Not. Roy. Astron. Soc.* **516** (2022) 421 [[arXiv:2205.06648](https://arxiv.org/abs/2205.06648)] [[INSPIRE](#)].

[115] A. Font-Ribera, P. McDonald and A. Slosar, *How to estimate the 3D power spectrum of the Lyman- $\alpha$  forest*, *JCAP* **01** (2018) 003 [[arXiv:1710.11036](https://arxiv.org/abs/1710.11036)] [[INSPIRE](#)].

- [116] A. Cuceu, A. Font-Ribera, B. Joachimi and S. Nadathur, *Cosmology beyond BAO from the 3D distribution of the Lyman- $\alpha$  forest*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 5439 [[arXiv:2103.14075](https://arxiv.org/abs/2103.14075)] [[INSPIRE](#)].
- [117] A. Cuceu et al., *Constraints on the Cosmic Expansion Rate at Redshift 2.3 from the Lyman- $\alpha$  Forest*, *Phys. Rev. Lett.* **130** (2023) 191003 [[arXiv:2209.13942](https://arxiv.org/abs/2209.13942)] [[INSPIRE](#)].
- [118] A. Cuceu et al., *The Alcock-Paczynski effect from Lyman- $\alpha$  forest correlations: analysis validation with synthetic data*, *Mon. Not. Roy. Astron. Soc.* **523** (2023) 3773 [[arXiv:2209.12931](https://arxiv.org/abs/2209.12931)] [[INSPIRE](#)].
- [119] F. Gerardi et al., *Direct cosmological inference from three-dimensional correlations of the Lyman  $\alpha$  forest*, *Mon. Not. Roy. Astron. Soc.* **518** (2022) 2567 [[arXiv:2209.11263](https://arxiv.org/abs/2209.11263)] [[INSPIRE](#)].
- [120] D.J. Eisenstein, H.-J. Seo, E. Sirk and D. Spergel, *Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak*, *Astrophys. J.* **664** (2007) 675 [[astro-ph/0604362](https://arxiv.org/abs/astro-ph/0604362)] [[INSPIRE](#)].
- [121] DESI collaboration, *Optimal Reconstruction of Baryon Acoustic Oscillations for DESI 2024*, [arXiv:2404.03005](https://arxiv.org/abs/2404.03005) [[INSPIRE](#)].
- [122] DESI collaboration, *DESI 2024 II: Sample definitions, characteristics and two-point clustering statistics*, in preparation (2024).
- [123] DESI collaboration, *DESI 2024 V: Analysis of the full shape of two-point clustering statistics from galaxies and quasars*, in preparation (2024).
- [124] DESI collaboration, *DESI 2024 VII: Cosmological constraints from full-shape analyses of the two-point clustering statistics measurements*, in preparation (2024).
- [125] DESI collaboration, *The Dark Energy Spectroscopic Instrument: one-dimensional power spectrum from first Ly  $\alpha$  forest samples with Fast Fourier Transform*, *Mon. Not. Roy. Astron. Soc.* **526** (2023) 5118 [[arXiv:2306.06311](https://arxiv.org/abs/2306.06311)] [[INSPIRE](#)].
- [126] N.G. Karaçaylı et al., *Optimal 1D Ly $\alpha$  Forest Power Spectrum Estimation — III. DESI early data*, *Mon. Not. Roy. Astron. Soc.* **528** (2024) 3941 [[arXiv:2306.06316](https://arxiv.org/abs/2306.06316)] [[INSPIRE](#)].
- [127] J.X. Prochaska, P. Madau, J.M. O'Meara and M. Fumagalli, *Towards a unified description of the intergalactic medium at redshift  $z \approx 2.5$* , *Mon. Not. Roy. Astron. Soc.* **438** (2014) 476 [[arXiv:1310.0052](https://arxiv.org/abs/1310.0052)] [[INSPIRE](#)].
- [128] E. Chaussidon et al., *Constraints on Primordial Non-Gaussianities for DESI 2024*, in preparation (2024).
- [129] S. Gratton and A. Challinor, *Understanding parameter differences between analyses employing nested data subsets*, *Mon. Not. Roy. Astron. Soc.* **499** (2020) 3410 [[arXiv:1911.07754](https://arxiv.org/abs/1911.07754)] [[INSPIRE](#)].

## The DESI collaboration

A.G. Adame<sup>a</sup>, J. Aguilar<sup>b</sup>, S. Ahlen<sup>Dc</sup>, S. Alam<sup>Dd</sup>, D.M. Alexander<sup>De,f</sup>, M. Alvarez<sup>b</sup>, O. Alves<sup>g</sup>, A. Anand<sup>Db</sup>, U. Andrade<sup>Dh,g</sup>, E. Armengaud<sup>Di</sup>, S. Avila<sup>Dj</sup>, A. Aviles<sup>Dk,l</sup>, H. Awan<sup>Dg</sup>, S. Bailey<sup>Db</sup>, C. Baltay<sup>m</sup>, A. Bault<sup>Dn</sup>, J. Bautista<sup>o</sup>, J. Behera<sup>p</sup>, S. BenZvi<sup>Dq</sup>, F. Beutler<sup>Dr</sup>, D. Bianchi<sup>Ds</sup>, C. Blake<sup>Dt</sup>, R. Blum<sup>Du</sup>, S. Brieden<sup>Dr</sup>, A. Brodzeller<sup>Dv</sup>, D. Brooks<sup>w</sup>, E. Buckley-Geer<sup>x,y</sup>, E. Burtin<sup>i</sup>, R. Calderon<sup>Dz</sup>, R. Canning<sup>aa</sup>, A. Carnero Rosell<sup>Dab,ac</sup>, R. Cereskaite<sup>ad</sup>, J.L. Cervantes-Cota<sup>Dae</sup>, S. Chabanier<sup>Db</sup>, E. Chaussidon<sup>Db</sup>, J. Chaves-Montero<sup>Dj</sup>, S. Chen<sup>Daf</sup>, X. Chen<sup>Dm</sup>, T. Claybaugh<sup>b</sup>, S. Cole<sup>Df</sup>, A. Cuceu<sup>Dag,1</sup>, T.M. Davis<sup>Dah</sup>, K. Dawson<sup>v</sup>, R. de la Cruz<sup>Dai</sup>, A. de la Macorra<sup>Daj</sup>, A. de Mattia<sup>Di</sup>, N. Deiosso<sup>Dak</sup>, A. Dey<sup>Du</sup>, B. Dey<sup>Dal</sup>, J. Ding<sup>am</sup>, Z. Ding<sup>Dan</sup>, P. Doel<sup>w</sup>, J. Edelstein<sup>ao,ap</sup>, S. Eftekharzadeh<sup>aq</sup>, D.J. Eisenstein<sup>ar</sup>, A. Elliott<sup>Das,at</sup>, P. Fagrelius<sup>u</sup>, K. Fanning<sup>Dau,av</sup>, S. Ferraro<sup>Db,ap</sup>, J. Ereza<sup>Daw</sup>, N. Findlay<sup>Daa</sup>, B. Flaugher<sup>y</sup>, A. Font-Ribera<sup>Dj</sup>, D. Forero-Sánchez<sup>Dax</sup>, J.E. Forero-Romero<sup>Day,az</sup>, C. Garcia-Quintero<sup>Dba</sup>, E. Gaztañaga<sup>bb,aa,bc</sup>, H. Gil-Marín<sup>Dbd,bb,be</sup>, S. Gontcho A Gontcho<sup>Db</sup>, A.X. Gonzalez-Morales<sup>Dbf,ai</sup>, V. Gonzalez-Perez<sup>Dbg,a</sup>, C. Gordon<sup>Dj</sup>, D. Green<sup>Dn</sup>, D. Gruen<sup>bh,bi</sup>, R. Gsponer<sup>Daa</sup>, G. Gutierrez<sup>y</sup>, J. Guy<sup>Db</sup>, B. Hadzhiyska<sup>Db,ap</sup>, C. Hahn<sup>Dbj</sup>, M.M. S Hanif<sup>Dg</sup>, H.K. Herrera-Alcantar<sup>Dai</sup>, K. Honscheid<sup>ag,as,at</sup>, C. Howlett<sup>Dah</sup>, D. Huterer<sup>Dg</sup>, V. Iršič<sup>Dbk</sup>, M. Ishak<sup>Dba</sup>, S. Juneau<sup>u</sup>, N.G. Karaçaylı<sup>Dag,bl,as,at</sup>, R. Kehoe<sup>bm</sup>, S. Kent<sup>Dx,y</sup>, D. Kirkby<sup>Dn</sup>, A. Kremin<sup>Db</sup>, A. Krolewski<sup>bn,bo,bp</sup>, Y. Lai<sup>ah</sup>, T.-W. Lan<sup>Dbq</sup>, M. Landriau<sup>Db</sup>, D. Lang<sup>bo</sup>, J. Lasker<sup>Dbm</sup>, J.M. Le Goff<sup>i</sup>, L. Le Guillou<sup>Dbr</sup>, A. Leauthaud<sup>Dam,bs</sup>, M.E. Levi<sup>Db</sup>, T.S. Li<sup>Dbt</sup>, E. Linder<sup>Db,ao,ap</sup>, K. Lodha<sup>Dz,bu</sup>, C. Magneville<sup>i</sup>, M. Manera<sup>Dbv,j</sup>, D. Margala<sup>Db</sup>, P. Martini<sup>Dag,bl,at</sup>, M. Maus<sup>ap</sup>, P. McDonald<sup>Db</sup>, L. Medina-Varela<sup>ba</sup>, A. Meisner<sup>Du</sup>, J. Mena-Fernández<sup>Dbw</sup>, R. Miquel<sup>bz,j</sup>, J. Moon<sup>by</sup>, S. Moore<sup>f</sup>, J. Moustakas<sup>Dbz</sup>, E. Mueller<sup>ad</sup>, A. Muñoz-Gutiérrez<sup>aj</sup>, A.D. Myers<sup>ca</sup>, S. Nadathur<sup>Daa</sup>, L. Napolitano<sup>Dca</sup>, R. Neveux<sup>r</sup>, J. A. Newman<sup>Dal</sup>, N.M. Nguyen<sup>Dg</sup>, J. Nie<sup>Dcb</sup>, G. Niz<sup>Dai,k</sup>, H.E. Noriega<sup>Dl,aj</sup>, N. Padmanabhan<sup>m</sup>, E. Paillas<sup>Dbn,bp</sup>, N. Palanque-Delabrouille<sup>Di,b</sup>, J. Pan<sup>Dg</sup>, S. Penmetsa<sup>bn</sup>, W.J. Percival<sup>Dbn,bo,bp</sup>, M.M. Pieri<sup>cc</sup>, M. Pinon<sup>Di</sup>, C. Poppett<sup>b,ao,ap</sup>, A. Porredon<sup>Dr,cd,at</sup>, F. Prada<sup>Daw</sup>, A. Pérez-Fernández<sup>Daj,by</sup>, I. Pérez-Ràfols<sup>Dce</sup>, D. Rabinowitz<sup>m</sup>, A. Raichoor<sup>Db</sup>, C. Ramírez-Pérez<sup>j</sup>, S. Ramirez-Solano<sup>aj</sup>, M. Rashkovetskyi<sup>Dar</sup>, C. Ravoux<sup>Doi,cf</sup>, M. Rezaie<sup>Dp</sup>, J. Rich<sup>i</sup>, A. Rocher<sup>Dax,i</sup>, C. Rockosi<sup>Dam,bs,cg</sup>, N.A. Roe<sup>b</sup>, A. Rosado-Marin<sup>ch</sup>, A.J. Ross<sup>Dag,bl,at</sup>, G. Rossi<sup>ci</sup>, R. Ruggeri<sup>Dt,ah</sup>, V. Ruhlmann-Kleider<sup>Di</sup>, L. Samushia<sup>Dcj,p,ck</sup>, E. Sanchez<sup>Dak</sup>, C. Saulder<sup>Dby</sup>, E.F. Schlafly<sup>Dcl</sup>, D. Schlegel<sup>b</sup>, M. Schubnell<sup>g</sup>, H. Seo<sup>Dch</sup>, R. Sharples<sup>Dcm,f</sup>, J. Silber<sup>Db</sup>, F. Sinigaglia<sup>Dab,ac</sup>, A. Slosar<sup>cn</sup>, A. Smith<sup>Df</sup>, D. Sprayberry<sup>u</sup>, T. Tan<sup>Di</sup>, G. Tarlé<sup>Dg</sup>, S. Trusov<sup>br</sup>, R. Vaisakh<sup>Dbm</sup>, D. Valcín<sup>Dch</sup>, F. Valdes<sup>Du</sup>, M. Vargas-Magaña<sup>Daj</sup>, L. Verde<sup>Dbx,be</sup>, M. Walther<sup>Dbh,bi</sup>, B. Wang<sup>Dco,cp</sup>, M.S. Wang<sup>Dr</sup>, B.A. Weaver<sup>u</sup>, N. Weaverdyck<sup>Db</sup>, R.H. Wechsler<sup>Dau,cq,av</sup>, D.H. Weinberg<sup>Db,at</sup>, M. White<sup>Dcr,ap</sup>, J. Yu<sup>Dax</sup>, Y. Yu<sup>Dan</sup>, S. Yuan<sup>Dav</sup>, C. Yèche<sup>Di</sup>, E.A. Zaborowski<sup>Dag,as,at</sup>, P. Zarrouk<sup>Dbr</sup>, H. Zhang<sup>Dbn,bp</sup>, C. Zhao<sup>Dcp</sup>, R. Zhao<sup>aa,cb</sup>, R. Zhou<sup>Db</sup>, H. Zou<sup>Dcb</sup>

<sup>a</sup> Instituto de Física Teórica (IFT) UAM/CSIC, Universidad Autónoma de Madrid, Cantoblanco, E-28049, Madrid, Spain

<sup>b</sup> Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, U.S.A.

<sup>c</sup> Physics Department, Boston University, 590 Commonwealth Avenue, Boston, MA 02215, U.S.A.

<sup>d</sup> Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India

<sup>e</sup> Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham, DH1 3LE, U.K.

<sup>f</sup> Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, U.K.

<sup>g</sup> Department of Physics, University of Michigan, Ann Arbor, MI 48109, U.S.A.

<sup>h</sup> Leinweber Center for Theoretical Physics, University of Michigan, 450 Church Street, Ann Arbor, Michigan 48109-1040, U.S.A.

<sup>i</sup> IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

<sup>j</sup> Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra Barcelona, Spain

<sup>k</sup> Instituto Avanzado de Cosmología A. C., San Marcos 11 – Atenas 202. Magdalena Contreras, 10720. Ciudad de México, México

<sup>l</sup> Instituto de Ciencias Físicas, Universidad Autónoma de México, Cuernavaca, Morelos, 62210, México

<sup>m</sup> Physics Department, Yale University, P.O. Box 208120, New Haven, CT 06511, U.S.A.

<sup>n</sup> Department of Physics and Astronomy, University of California, Irvine, 92697, U.S.A.

<sup>o</sup> Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France

<sup>p</sup> Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, U.S.A.

<sup>q</sup> Department of Physics & Astronomy, University of Rochester, 206 Bausch and Lomb Hall, P.O. Box 270171, Rochester, NY 14627-0171, U.S.A.

<sup>r</sup> Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, U.K.

<sup>s</sup> Dipartimento di Fisica “Aldo Pontremoli”, Università degli Studi di Milano, Via Celoria 16, I-20133 Milano, Italy

<sup>t</sup> Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia

<sup>u</sup> NSF NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, U.S.A.

<sup>v</sup> Department of Physics and Astronomy, The University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, U.S.A.

<sup>w</sup> Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, U.K.

<sup>x</sup> Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, U.S.A.

<sup>y</sup> Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, U.S.A.

<sup>z</sup> Korea Astronomy and Space Science Institute, 776, Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Republic of Korea

<sup>aa</sup> Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth, PO1 3FX, U.K.

<sup>ab</sup> Departamento de Astrofísica, Universidad de La Laguna (ULL), E-38206, La Laguna, Tenerife, Spain

<sup>ac</sup> Instituto de Astrofísica de Canarias, C/ Vía Láctea, s/n, E-38205 La Laguna, Tenerife, Spain

<sup>ad</sup> Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, U.K.

<sup>ae</sup> Departamento de Física, Instituto Nacional de Investigaciones Nucleares, Carretera México-Toluca S/N, La Marquesa, Ocoyoacac, Edo. de México C.P. 52750, México

<sup>af</sup> Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, U.S.A.

<sup>ag</sup> Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, U.S.A.

<sup>ah</sup> School of Mathematics and Physics, University of Queensland, 4072, Australia

<sup>ai</sup> Departamento de Física, Universidad de Guanajuato – DCI, C.P. 37150, Leon, Guanajuato, México

<sup>aj</sup> Instituto de Física, Universidad Nacional Autónoma de México, Cd. de México C.P. 04510, México

<sup>ak</sup> CIEMAT, Avenida Complutense 40, E-28040 Madrid, Spain

<sup>al</sup> Department of Physics & Astronomy and Pittsburgh Particle Physics, Astrophysics, and Cosmology Center (PITT PACC), University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, PA 15260, U.S.A.

<sup>am</sup> Department of Astronomy and Astrophysics, UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064, U.S.A.

<sup>a<sub>n</sub></sup> Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>a<sub>o</sub></sup> Space Sciences Laboratory, University of California, Berkeley, 7 Gauss Way, Berkeley, CA 94720, U.S.A.

<sup>a<sub>p</sub></sup> University of California, Berkeley, 110 Sproul Hall #5800 Berkeley, CA 94720, U.S.A.

<sup>a<sub>q</sub></sup> Universities Space Research Association, NASA Ames Research Centre, Moffett Blvd, Mountain View, CA 94035, U.S.A.

<sup>a<sub>r</sub></sup> Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, U.S.A.

<sup>a<sub>s</sub></sup> Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, U.S.A.

<sup>a<sub>t</sub></sup> The Ohio State University, Columbus, 43210 OH, U.S.A.

<sup>a<sub>u</sub></sup> Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Menlo Park, CA 94305, U.S.A.

<sup>a<sub>v</sub></sup> SLAC National Accelerator Laboratory, Menlo Park, CA 94305, U.S.A.

<sup>a<sub>w</sub></sup> Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, s/n, E-18008 Granada, Spain

<sup>a<sub>x</sub></sup> Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

<sup>a<sub>y</sub></sup> Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, CP 111711, Bogotá, Colombia

<sup>a<sub>z</sub></sup> Observatorio Astronómico, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio H, CP 111711 Bogotá, Colombia

<sup>b<sub>a</sub></sup> Department of Physics, The University of Texas at Dallas, Richardson, TX 75080, U.S.A.

<sup>b<sub>b</sub></sup> Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain

<sup>b<sub>c</sub></sup> Institute of Space Sciences, ICE-CSIC, Campus UAB, Carrer de Can Magrans s/n, 08913 Bellaterra, Barcelona, Spain

<sup>b<sub>d</sub></sup> Departament de Física Quàntica i Astrofísica, Universitat de Barcelona, Martí i Franquès 1, E08028 Barcelona, Spain

<sup>b<sub>e</sub></sup> Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (UB), c. Martí i Franquès, 1, 08028 Barcelona, Spain

<sup>b<sub>f</sub></sup> Consejo Nacional de Ciencia y Tecnología, Av. Insurgentes Sur 1582, Colonia Crédito Constructor, Del. Benito Juárez C.P. 03940, México D.F., México

<sup>b<sub>g</sub></sup> Centro de Investigación Avanzada en Física Fundamental (CIAFF), Facultad de Ciencias, Universidad Autónoma de Madrid, ES-28049 Madrid, Spain

<sup>b<sub>h</sub></sup> Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany

<sup>b<sub>i</sub></sup> University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81677 München, Germany

<sup>b<sub>j</sub></sup> Department of Astrophysical Sciences, Princeton University, Princeton NJ 08544, U.S.A.

<sup>b<sub>k</sub></sup> Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, U.K.

<sup>b<sub>l</sub></sup> Department of Astronomy, The Ohio State University, 4055 McPherson Laboratory, 140 W 18th Avenue, Columbus, OH 43210, U.S.A.

<sup>b<sub>m</sub></sup> Department of Physics, Southern Methodist University, 3215 Daniel Avenue, Dallas, TX 75275, U.S.A.

<sup>b<sub>n</sub></sup> Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

<sup>b<sub>o</sub></sup> Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada

<sup>b<sub>p</sub></sup> Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

<sup>b<sub>q</sub></sup> Graduate Institute of Astrophysics and Department of Physics, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan

<sup>b<sub>r</sub></sup> Sorbonne Université, CNRS/IN2P3, Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), FR-75005 Paris, France

<sup>bs</sup> Department of Astronomy and Astrophysics, University of California, Santa Cruz,  
1156 High Street, Santa Cruz, CA 95065, U.S.A.

<sup>bt</sup> Department of Astronomy & Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada

<sup>bu</sup> University of Science and Technology,  
217 Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea

<sup>bv</sup> Departament de Física, Serra Húnter, Universitat Autònoma de Barcelona,  
08193 Bellaterra (Barcelona), Spain

<sup>bw</sup> Laboratoire de Physique Subatomique et de Cosmologie,  
53 Avenue des Martyrs, 38000 Grenoble, France

<sup>bx</sup> Institució Catalana de Recerca i Estudis Avançats,  
Passeig de Lluís Companys, 23, 08010 Barcelona, Spain

<sup>by</sup> Max Planck Institute for Extraterrestrial Physics, Gießenbachstraße 1, 85748 Garching, Germany

<sup>bz</sup> Department of Physics and Astronomy, Siena College,  
515 Loudon Road, Loudonville, NY 12211, U.S.A.

<sup>ca</sup> Department of Physics & Astronomy, University of Wyoming,  
1000 E. University, Dept. 3905, Laramie, WY 82071, U.S.A.

<sup>cb</sup> National Astronomical Observatories, Chinese Academy of Sciences,  
A20 Datun Rd., Chaoyang District, Beijing, 100012, P.R. China

<sup>cc</sup> Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France

<sup>cd</sup> Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB),  
German Centre for Cosmological Lensing, 44780 Bochum, Germany

<sup>ce</sup> Departament de Física, EEBE, Universitat Politècnica de Catalunya,  
c/Eduard Maristany 10, 08930 Barcelona, Spain

<sup>cf</sup> Université Clermont-Auvergne, CNRS, LPCA, 63000 Clermont-Ferrand, France

<sup>cg</sup> University of California Observatories, 1156 High Street, Santa Cruz, CA 95065, U.S.A.

<sup>ch</sup> Department of Physics & Astronomy, Ohio University, Athens, OH 45701, U.S.A.

<sup>ci</sup> Department of Physics and Astronomy, Sejong University, Seoul, 143-747, Korea

<sup>cj</sup> Abastumani Astrophysical Observatory, Tbilisi, GE-0179, Georgia

<sup>ck</sup> Faculty of Natural Sciences and Medicine, Ilia State University, 0194 Tbilisi, Georgia

<sup>cl</sup> Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, U.S.A.

<sup>cm</sup> Centre for Advanced Instrumentation, Department of Physics, Durham University,  
South Road, Durham DH1 3LE, U.K.

<sup>cn</sup> Physics Department, Brookhaven National Laboratory, Upton, NY 11973, U.S.A.

<sup>co</sup> Beihang University, Beijing 100191, China

<sup>cp</sup> Department of Astronomy, Tsinghua University,  
30 Shuangqing Road, Haidian District, Beijing 100190, China

<sup>cq</sup> Physics Department, Stanford University, Stanford, CA 93405, U.S.A.

<sup>cr</sup> Department of Physics, University of California, Berkeley,  
366 LeConte Hall MC 7300, Berkeley, CA 94720-7300, U.S.A.

<sup>1</sup> NASA Einstein Fellow.