

UC Irvine

UC Irvine Previously Published Works

Title

Probing the limits of cosmological information from the Lyman- α forest 2-point correlation functions

Permalink

<https://escholarship.org/uc/item/5tp9s0q5>

Journal

Journal of Cosmology and Astroparticle Physics, 2026(05)

ISSN

1475-7516

Authors

Turner, Wynne

Cuceu, Andrei

Martini, Paul

et al.

Publication Date

2026-05-01

DOI

10.1088/1475-7516/2026/05/010

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

PAPER • OPEN ACCESS

Probing the limits of cosmological information from the Lyman- α forest 2-point correlation functions

To cite this article: Wynne Turner *et al* JCAP05(2026)010

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

You may also like

- [Performance of the CMS high-level trigger during LHC Run 2](#)
A. Hayrapetyan, A. Tumasyan, W. Adam et al.
- [Identification of low-momentum muons in the CMS detector using multivariate techniques in proton-proton collisions at \$\sqrt{s} = 13.6\$ TeV](#)
V. Chekhovsky, A. Hayrapetyan, V. Makarenko et al.
- [Performance of heavy-flavour jet identification in Lorentz-boosted topologies in proton-proton collisions at \$\sqrt{s} = 13\$ TeV](#)
A. Hayrapetyan, A. Tumasyan, W. Adam et al.

RECEIVED: September 19, 2025

REVISED: February 4, 2026

ACCEPTED: March 6, 2026

PUBLISHED: May 5, 2026

Probing the limits of cosmological information from the Lyman- α forest 2-point correlation functions

Wynne Turner et al.

Full author list at the end of the paper

E-mail: turner.1839@osu.edu

ABSTRACT: The standard cosmological analysis with the Ly α forest relies on a continuum fitting procedure that suppresses information on large scales and distorts the three-dimensional correlation function on all scales. In this work, we present the first cosmological forecasts without continuum fitting distortion in the Ly α forest, focusing on the recovery of large-scale information. Using idealized synthetic data, we compare the constraining power of the full shape of the Ly α forest auto-correlation and its cross-correlation with quasars using the baseline continuum fitting analysis versus the true continuum. We find that knowledge of the true continuum enables a $\sim 10\%$ reduction in uncertainties on the Alcock-Paczyński (AP) parameter and the matter density, Ω_m . We also explore the impact of large-scale information by extending the analysis up to separations of $240 h^{-1}\text{Mpc}$ along and across the line of sight. The combination of these analysis choices can recover significant large-scale information, yielding up to a $\sim 15\%$ improvement in AP constraints. This improvement is analogous to extending the Ly α forest survey area by $\sim 40\%$.

KEYWORDS: cosmological parameters from LSS, Lyman alpha forest

ARXIV EPRINT: [2509.14322](https://arxiv.org/abs/2509.14322)



Contents

1	Introduction	1
2	Quasar spectra	4
2.1	Synthetic data	4
2.2	Forest continuum	5
3	Method	6
3.1	Correlations and covariance	6
3.2	Theoretical framework	8
3.3	Implementation	10
4	Compressed analysis results	12
4.1	True continuum	12
4.2	Changes in scale range	13
4.3	Covariance modifications	16
5	Direct cosmology results	17
6	Discussion	18
7	Summary	20
A	Potential feasibility with LyCAN	22
B	Validation of the Hartlap correction	23
	Author List	31

1 Introduction

Ever since the discovery of the universe’s accelerating expansion [1, 2], determining the nature of dark energy has become a central goal in modern cosmology. Multiple independent probes — Type 1a supernovae, the cosmic microwave background, and the clustering of tracers of the matter-density field — consistently confirm this late-time accelerated expansion. Among the most powerful tools for constraining the expansion history are baryon acoustic oscillations (BAO), which provide a standard ruler originating from oscillations in the photon-baryon plasma prior to recombination. While galaxies and quasars are often used as tracers of the matter-density field [3, 4] in which we can measure the BAO scale, their number densities decline with increasing redshift, necessitating alternative high-redshift probes.

The Lyman-alpha ($\text{Ly}\alpha$) forest is the collection of absorption lines seen in the spectra of distant quasars resulting from intervening neutral hydrogen along the line of sight. At redshifts $z > 2$, the $\text{Ly}\alpha$ forest is observable from the ground and becomes a valuable probe for cosmological measurements at high redshifts, as the BAO scale can be measured from the

distribution of matter in the intergalactic medium. The Baryon Oscillation Spectroscopic Survey [BOSS; 5], as part of the third stage of the Sloan Digital Sky Survey [SDSS-III; 6] was the first to obtain cosmological measurements from the BAO scale in the Ly α forest [e.g. 7–9]. This was repeated by its successor, extended-BOSS [eBOSS; 10] within SDSS-IV [11] with a sample of $\sim 250,000$ Ly α forest quasars [e.g. 12, 13].

The Dark Energy Spectroscopic Instrument [DESI; 14–17] is a Stage-IV spectroscopic survey operating on the Mayall 4-meter telescope at Kitt Peak National Observatory [18]. DESI is on schedule to measure the spectra of approximately 50 million galaxies and quasars within the first five years of operation [19]. Approximately one million of these targets are high-redshift ($z \geq 2.1$) quasars whose spectra include the Ly α forest. DESI selects targets with a sophisticated pipeline for survey tiling and optimization [20]. The observations are then processed by the spectroscopic pipeline [21] for further analysis. The main DESI survey began in May 2021 following successful commissioning and survey validation periods [22]. Data Release 1 [DR1; 19] is publicly available, containing data from the first year of the survey, alongside their cosmological interpretation [e.g. 23]. Recently, cosmological results from the first three years of data (DR2) were also released, based on nearly one million Ly α forest quasar spectra [24] and over 14 million lower redshift galaxies and quasars [25]. The results have revealed tension between the best-fit model describing the universe’s expansion history and the Λ CDM concordance model, instead favoring an evolving $w_0 w_a$ CDM model. In addition to the collection of more data, new analysis methods will be crucial in elucidating this tension. In particular, [26] highlights the potential of the Ly α forest to constrain dynamical dark energy at high redshift.

Most cosmological studies with the Ly α forest have only used measurements of the BAO peak position. While robust, this method discards additional information contained in the full shape of the correlation functions. When the fiducial cosmology used to convert observed angular and redshift separations into comoving distances differs from the true cosmology, this introduces an anisotropy in the measured correlation functions. This anisotropy, known as the Alcock-Paczyński [AP; 27] effect, can be extracted to provide even more precise cosmological constraints. It impacts both the BAO scale and the broadband of the correlation functions. A full-shape analysis with the Ly α forest which extracts information from the broadband in addition to the BAO scale can provide up to twice the constraining power as a BAO-only analysis [28–30]. In a full-shape analysis, [31] showed that most of the large-scale cosmological information is captured in compressed parameters that describe the BAO scale alongside anisotropies introduced by the AP effect and redshift space distortions (RSD). The Ly α RSD signal is sensitive to the product of the linear growth rate and the amplitude of matter fluctuations on $8 h^{-1}$ Mpc scales, $f\sigma_8$, but is degenerate with an unknown velocity divergence bias [e.g. 32–36]. However, $f\sigma_8$ can be measured from the quasar RSD signal with the Ly α -QSO cross-correlation. While the joint full-shape Ly α forest analysis allows for a measurement of $f\sigma_8$, its most powerful cosmological constraint comes from the AP signal [e.g. 28]. A full-shape analysis with the Ly α forest has recently been performed for the first time with DESI DR1 data, yielding a 1.6% constraint on the AP effect [26].

Multiple systematics complicate the Ly α forest full-shape analysis. One such systematic is the quasar continuum fitting method. The standard approach estimates a mean continuum

from the dataset and performs a power-law fit to each forest. This process biases the mean and slope of each forest to zero, projecting out large-scale modes and correlating all pixels in the forest. Variations in quasar diversity become coupled with large-scale over- or under-densities during continuum fitting, leading to a distortion of the correlation function on all scales. The current solution is to model this effect with a distortion matrix prior to fitting the model to the measured correlation functions [37, 38]. The distortion matrix relies on several assumptions that must be validated using mock datasets [38]. So far, validation has focused on the BAO and AP parameters in the regimes typically studied [e.g. 26, 30, 39], but it has not yet been extended to other parameters. We do not study potential biases induced by the distortion matrix and instead focus on the idealized case assuming perfect knowledge of this distortion. Of particular importance to this work is that the continuum fitting process removes information about large-scale over- or under-densities that could in principle be recovered if the true unabsorbed continuum were known. Extending the analysis to larger scales may further enhance the recovery of this information.

Recently, alternative continuum fitting approaches have shown promise in accessing this previously lost information. These include principal component analysis [PCA; e.g. 40–42], nonnegative matrix factorization [NMF; e.g. 43], and more sophisticated machine learning methods [44–48]. Approaches such as Quasar Factor Analysis [45] and SPENDERQ [47] use both the red side and the blue side of the Ly α emission line to estimate the unabsorbed forest continuum. More recently, neural network approaches [e.g. 44, 46, 48] predict the unabsorbed continuum independently of the forest region, thereby avoiding the coupling between large-scale structure and quasar diversity that drives distortions in methods that use data in the forest region. However, such predictions are not necessarily free of spurious correlations, for example from correlated continuum errors, and it remains important to test for and characterize these effects. Predicting the continuum independently of the forest also offers other advantages; for instance, LyCAN has been applied to DESI DR1 data to measure the evolution of the effective optical depth [46]. PCA-based approaches, along with the neural networks iQNet [44] and LyCAN [46], have published predictions of the forest continuum based solely on the red side. Of these, the latter was developed to achieve accurate performance on DESI-like spectra. Tools such as LyCAN offer a path forward toward unlocking the full constraining power of the Ly α forest.

The current state-of-the-art analysis [e.g. 24, 49] fits a model to the two-point correlation functions over separations from $r = 30\text{--}180\ h^{-1}\text{Mpc}$, and estimates the covariance matrix from the data by subsampling correlation function measurements from ~ 1000 individual sky regions [24]. While modeling small-scale correlations ($r_{\min} < 30\ h^{-1}\text{Mpc}$) is challenging due to nonlinearities, extending the analysis to larger scales may be more feasible. Full-shape analyses can access additional cosmological information at these larger scales, particularly from the AP effect. However, a significant challenge arises from the fact that the correlation functions are measured in two dimensions, resulting in a large $p \times p$ covariance matrix, where p is the number of correlation function bins. As p increases, it becomes increasingly difficult to accurately estimate the covariance matrix from the data alone.

The goal of this work is to investigate potential gains in cosmological constraining power through two key avenues: (1) assuming knowledge of the true continuum, which eliminates

the distortion introduced by continuum fitting and is expected to better preserve large-scale information, and (2) extracting additional information by extending the analysis to larger scales. Our forecasts focus on the idealized case of perfect continuum knowledge to probe the limit of how much information could be gained with access to the unabsorbed continuum. Throughout this work, we refer to analyses that use the true continuum as *undistorted* methods, as they do not require the distortion matrix. We assess improvements in both compressed and uncompressed (direct) cosmological parameters. Our aim is to quantify the additional constraining power achievable with the idealized undistorted analysis relative to the continuum fitting (*distorted*) method, as well as the benefit of recovering large-scale information in both analyses.

This paper is organized as follows: we describe the synthetic data we use and different ways of estimating the quasar continuum in section 2. In section 3 we describe how we estimate covariance matrices and our forecasting methodology. We present our cosmological forecasts on compressed parameters in section 4, and on direct cosmological parameters in section 5. Throughout this paper we assume the flat Λ CDM cosmology from Planck 2018 [50]. In section 6 we provide a discussion including future work, and we summarize our results in section 7.

2 Quasar spectra

In this section we describe the synthetic quasar spectra that are the foundation of this analysis, as well as how the quasar continuum is modeled. We first describe the synthetic data used for our forecasts in section 2.1. In section 2.2 we summarize the continuum fitting procedure employed in the DESI Ly α forest analysis and compare it to the true unabsorbed continuum.

2.1 Synthetic data

We use 50 independent synthetic realizations (mocks) of the first three years of DESI to compute covariance matrices. The mocks were generated by `quickquasars`,¹ a DESI software suite that simulates realistic quasar spectra and is described in [51]. These mocks include input quasar continua produced by `simqso`,² a package that generates mock quasar spectra and photometry [52]. This package models the continuum as a broken power-law and is optimized to better match the mean continuum observed in eBOSS DR16 [13] using principal component analysis (PCA). These unabsorbed continua are then multiplied by transmission files generated with `LyaCoLoRe`, a package that produces synthetic Ly α forest datasets [53].

The `LyaCoLoRe` mocks were recently updated to include more realistic small-scale clustering and broadening of the BAO peak due to the nonlinear growth of structure [39]. We use these improved mocks for this work. Each mock is representative of the DESI DR2 Ly α forest dataset with realistic noise and a comparable magnitude and redshift distribution. For more details on the DESI mock creation process, see [51] and [39]. For this work, we use mocks without contamination from metals, broad absorption lines (BALs), or optically thick absorbers to isolate the effects of the distortion introduced by continuum fitting. We

¹<https://github.com/desihub/desisim/blob/main/py/desisim/scripts/quickquasars.py>.

²<https://github.com/imcgreer/simqso>.

use the true continua of the mocks for the undistorted analysis, and compare this to the continuum fitting (distorted) analysis to quantify the improvement in constraining power enabled by perfect knowledge of the true continuum. The computation of the covariance matrix is described in section 3.1.

2.2 Forest continuum

The current state-of-the-art in Ly α forest cosmology involves fitting the expected flux in each forest in order to measure the amount of absorption present in each pixel. While this approach effectively captures the BAO signal, it removes information on scales comparable to the forest length, limiting the ability to fully exploit the full shape of the 3D correlation functions.

The DESI Ly α forest analysis pipeline, `picca`, is a package that performs continuum fitting, measures correlation functions and power spectra, and computes the covariance matrices [13]. During continuum fitting, the expected flux $\bar{F}(\lambda)C_q(\lambda)$ is estimated by fitting a two-parameter model to the forest region of each quasar [54]:

$$\bar{F}(\lambda)C_q(\lambda) = \bar{C}(\lambda_{\text{RF}}) \left(a_q + b_q \frac{\Lambda - \Lambda_{\text{min}}}{\Lambda_{\text{max}} - \Lambda_{\text{min}}} \right), \quad (2.1)$$

where $\Lambda \equiv \log \lambda$, $\bar{C}(\lambda_{\text{RF}})$ is the mean continuum measured from the data, $C_q(\lambda)$ is the unknown unabsorbed continuum, and $\bar{F}(\lambda)$ is the mean flux. The parameters a_q and b_q account for variations in quasar diversity and the redshift evolution of the mean flux. This procedure removes information on large scales, since fluctuations in the density field on scales comparable to the forest are absorbed into the fit parameters. It also introduces spurious correlations that affect all scales, which must be taken into account using a distortion matrix. The distortion matrices for the auto- and cross-correlation are defined as [37, 38, 55]:

$$D_{AB}^{\text{auto}} = \frac{1}{W_A} \sum_{i,j \in A} w_i w_j \sum_{i',j' \in B} \eta_{ii'} \eta_{jj'}, \quad (2.2)$$

$$D_{AB}^{\text{cross}} = \frac{1}{W_A} \sum_{i,j \in A} w_i w_j \sum_{i',j' \in B} \eta_{ii'}, \quad (2.3)$$

where $w_{i,j}$ are the weights, A and B refer to the data and model bins, respectively, and the sums are over pixel-pixel pairs for the auto-correlation and pixel-quasar pairs for the cross-correlation. The projection matrices, η_{ij} , are given by:

$$\eta_{ij} = \delta_{ij}^K - \frac{w_j}{\sum_k w_k} - \frac{w_j \kappa_i \kappa_j}{\sum_k w_k \kappa_k^2}, \quad (2.4)$$

where δ_{ij}^K is the Kronecker delta. The quantity $\kappa_k = \log \lambda_k - \overline{\log \lambda_q}$, where λ_i is the observed wavelength in pixel i and $\overline{\log \lambda_q}$ is the mean of $\log \lambda$ in the forest of quasar q . The measured distortion matrix is multiplied by the undistorted correlation function model to achieve a distorted correlation function model, which is subsequently fit to the data. This method has been shown to yield robust, unbiased constraints for BAO and AP parameters [see e.g. 26, 30, 39], but its performance for other parameters has not yet been established [38]. Figure 1 shows an example DESI mock spectrum with the fitted continuum (expected flux) and its true continuum. Continuum prediction models such as LyCAN [46] aim to reconstruct

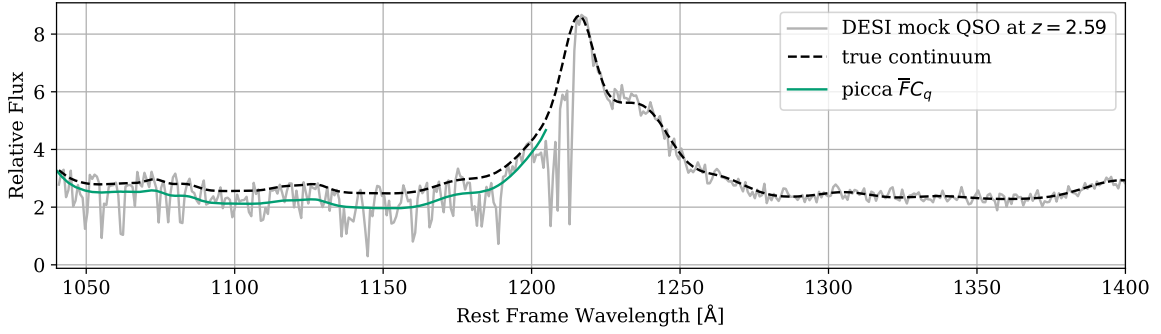


Figure 1. An example DESI DR2 mock quasar spectrum with its true continuum (*black dashed line*) and the product $\overline{F}C_q$ fit by the *picca* pipeline (*teal line*).

this true continuum independently of the forest region to achieve undistorted correlation function measurements. Recent results from LyCAN, presented in appendix A, indicate this may be feasible and therefore motivate this study.

In this work, we present the first analysis of the cosmological impact of bypassing the distortion matrix and preserving all of the large-scale information. To forecast the limiting case of perfect knowledge of the true continuum, we use the true continua from our mock dataset.

3 Method

We perform cosmological forecasts using a noiseless synthetic data vector built from the fiducial cosmology [50] and DESI DR2 mock covariance matrices for the true continuum and continuum fitting analyses. Section 3.1 describes how we compute mock covariances for the undistorted and distorted analyses. We provide an overview of the analysis framework in section 3.2, and in section 3.3 we describe how we perform our forecasts in two different parameter spaces to constrain compressed (section 3.3.1) or direct (section 3.3.2) cosmological parameters.

3.1 Correlations and covariance

We use two versions of synthetic 3D Ly α correlation functions without contaminants. The distorted version includes a model for the distortion effect due to continuum fitting and a distortion matrix computed from one mock, while the undistorted version is based on the true continuum and does not require the distortion matrix. In the distorted analysis, the model for the correlation function is multiplied by the distortion matrix. We generate our synthetic data vector following the framework of [28], built from a model based on cosmological parameters from the Planck 2018 results [50] and nuisance parameters from DESI DR2 [24]. The data vector is noiseless and we calculate covariance matrices from uncontaminated synthetic data based on the first three years of DESI.

We use *picca* to compute correlation functions and synthetic covariance matrices from the mock data described in section 2.1. We only use the covariance matrices since we are only concerned with forecasting, while the noiseless synthetic data used in our forecasts is generated from the fiducial model described in section 3.2. In the baseline continuum fitting analysis,

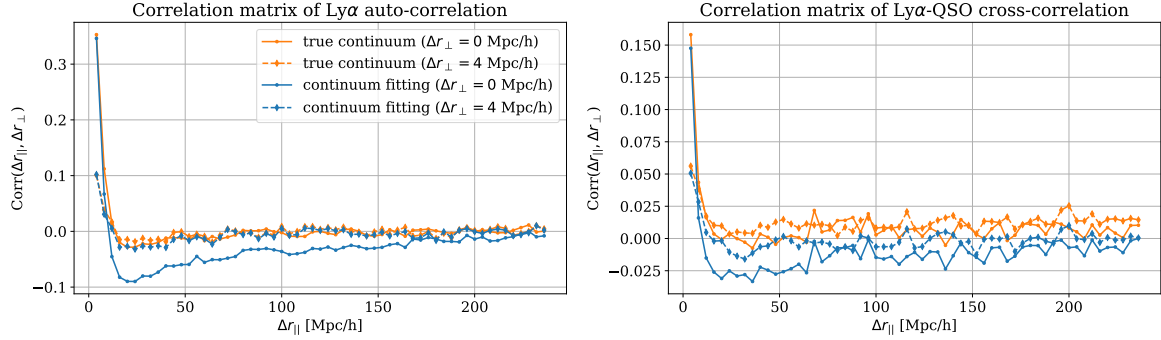


Figure 2. The unsmoothed correlation matrix as a function of line-of-sight separation for the true continuum (undistorted) and continuum fitting (distorted) analyses. We show correlation matrix elements for the Ly α auto-correlation (*left*) and the Ly α -QSO cross-correlation (*right*) in two bins of transverse separation. The correlation quickly approaches zero with increasing separation, especially for the true continuum case which exhibits smaller correlations.

the covariance matrix is normally computed through subsampling of HEALPix pixels [56] with $\text{NSIDE} = 16$. This assumes that the correlation function in each $\sim 250 \times 250 (h^{-1}\text{Mpc})^2$ patch covered by one HEALPix pixel is independent. While this is a good assumption for the continuum fitting analysis where large-scale modes are projected out and the fits only extend to $180 \text{ Mpc}/h$, it may no longer be accurate for an undistorted continuum analysis where we expect large-scale correlations. There are therefore two main considerations when deciding how to compute the correlations and covariance: (1) the size of the HEALPix pixels relative the largest transverse separations, and (2) the number of independent correlation function measurements that we use to derive the covariance. To ensure independence at larger scales, we use larger HEALPix pixels ($\text{NSIDE} = 14$, corresponding to patches of linear size $\sim 280 h^{-1}\text{Mpc}$) to compute the subsamples for the true continuum analysis. Since the larger HEALPix pixels reduce the number of available subsamples for estimating the covariance matrix, we mitigate this by using 50 independent mock realizations, providing a sufficient number of independent correlation function measurements ($N = N_{\text{HEALPix}} \times N_{\text{mocks}} \approx 43000$) relative to the size of the data vector. We compute the covariance up to separations of $240 h^{-1}\text{Mpc}$ in both the transverse and line-of-sight directions to investigate the additional constraining power that may come from larger scales, particularly for the undistorted analysis. For consistency, we compute the covariance matrix for the distorted analysis in the same way. We rescale each covariance matrix to one mock by multiplying them by a factor of 50. We also calculate the cross-covariance between the Ly α auto- and cross-correlation functions as in [49] and [24]. Figure 2 shows representations of the correlation matrix for each analysis for the Ly α auto- and cross-correlation functions. The variance of the undistorted correlation function is about 5–10% larger than that of the distorted correlation function in the line-of-sight bin ($\Delta r_{\perp} = 0 \text{ Mpc}/h$).

In the baseline continuum fitting analysis, the noisy measurement of the covariance matrix from the data is smoothed. This process replaces all off-diagonal elements of the correlation matrix with equal line-of-sight and transverse separations with their average [e.g. 13, 24, 37, 49, 57, 58]. While the smoothing procedure has been validated in the context of

continuum fitting for BAO and AP parameters [26, 30, 58], it has not been validated in the true continuum analysis or for other parameters of interest. We choose not to adopt the assumptions implicit in the smoothing procedure, as they may not hold in the undistorted analysis. They also may not hold at larger separations as the smoothing erases some large-scale structure in the correlation matrix (e.g. due to cosmic variance). For these reasons, we do not smooth our covariance matrices in either analysis, except to compare our results to the current baseline.

Because we use a limited number of independent samples to measure covariance matrices, these estimates will be noisy, particularly at larger separations. Noise in the covariance matrix tends to result in underestimated uncertainties in inferred parameters due to the bias in the inverse covariance [59]. Since we do not smooth our covariance matrices, we apply the Hartlap correction [59, 60] to the inverse covariance during the likelihood analysis. We validate that the Hartlap approximation is accurate and sufficient in the regimes we explore in appendix B.

This work uses the standard estimators of the correlation function and its associated covariance matrix currently employed in the DESI Ly α forest analysis [24, 49]. However, these current estimators are sub-optimal. During correlation function estimation, pixel weights are assigned independently without accounting for covariance between pixels within the same forest (cross-pixel) or between different forests (cross-forest) [e.g. 8, 54], whereas an optimal estimator would take into account the full covariance matrix of all pixels in the survey. The size of the Ly α forest dataset precludes the use of an optimal estimator at present. Instead, [54] introduced an empirical correction term tuned on mock datasets that improves the precision of the correlation function measurements. In the context of the covariance matrix, the subsampling estimator is also sub-optimal. While these estimators have been extensively validated for the traditional continuum fitting analysis, their performance in true continuum analyses remains relatively untested. Developing a more optimal covariance estimator tailored for undistorted continuum analyses will be an important goal for future work.

3.2 Theoretical framework

The correlation functions are computed on grids of comoving coordinates along and across the line of sight. For two tracers at redshifts z_i and z_j separated by an angle $\Delta\theta$, their separations in comoving coordinates are [13, 28, 61]:

$$r_{\parallel} = [D_{C,\text{fid}}(z_i) - D_{C,\text{fid}}(z_j)] \cos(\Delta\theta/2), \quad (3.1)$$

$$r_{\perp} = [D_{M,\text{fid}}(z_i) + D_{M,\text{fid}}(z_j)] \sin(\Delta\theta/2), \quad (3.2)$$

where $D_C = c \int_0^z dz'/H(z')$ is the radial comoving distance, D_M is the transverse comoving distance, and the fid subscript denotes the fiducial cosmology. Throughout this work, we adopt the Planck 2018 Λ CDM model as our fiducial cosmology [50]. The mismatch between the fiducial and true cosmology induces an extra anisotropy in the correlation functions due to the AP effect, which we extract for additional constraining power.

The tracer separations $r_{\parallel}$ and $r_{\perp}$ correspond to the separations computed in the fiducial cosmology. If this differs from the true cosmology, the true separations are instead described by transformations of the fiducial separations, $r'_{\parallel}(r_{\parallel}, r_{\perp}, z)$ and $r'_{\perp}(r_{\parallel}, r_{\perp}, z)$ [28, 62]. Since

the AP effect changes the ratio $r_{\perp}/r_{\parallel}$, these true separations can be related to the fiducial ones through anisotropic rescalings,

$$r'_{\parallel} = q_{\parallel} r_{\parallel}, \quad (3.3)$$

$$r'_{\perp} = q_{\perp} r_{\perp}. \quad (3.4)$$

The AP parameter at effective redshift z is defined as the ratio of the transverse comoving distance $D_M(z)$ to the Hubble distance $D_H(z)$:

$$F_{\text{AP}} = \frac{D_M(z)}{D_H(z)}. \quad (3.5)$$

To isolate the anisotropy introduced by the AP effect, we measure the parameter

$$\phi(z) = \frac{F_{\text{AP}}(z)}{F_{\text{AP},\text{fid}}(z)} = \frac{q_{\perp}(z)}{q_{\parallel}(z)} = \frac{D_M(z)H(z)}{[D_M(z)H(z)]_{\text{fid}}}. \quad (3.6)$$

A measurement of $\phi(z)$ therefore corresponds to a measurement of Ω_m in flat Λ CDM [28]. This information is commonly summarized as ϕ_f , which captures the full-shape AP signal. Alternatively, the AP information can be separated into two components: ϕ_p , which isolates the contribution from the BAO peak, and ϕ_s , which captures the broadband (smooth) contribution. Similarly, we can define an isotropic rescaling of coordinates as

$$\alpha(z) = \sqrt{q_{\perp}(z)q_{\parallel}(z)}. \quad (3.7)$$

The compressed parameters $\alpha(z)$ and $\phi(z)$ will both be unity if the true underlying cosmology is equal to the fiducial cosmology. The parameter $\alpha(z)$ has different interpretations for the peak and broadband (smooth) components. For the peak component, this corresponds to

$$\alpha_p(z) = \sqrt{\frac{D_M(z)D_H(z)/r_d^2}{[D_M(z)D_H(z)/r_d^2]_{\text{fid}}}}, \quad (3.8)$$

where r_d is the sound horizon scale at the end of the drag epoch. The broadband (or smooth) isotropic component α_s is less straightforward to model [28], and is typically marginalized over as most of the cosmological information is contained within α_p and ϕ_f , which contains the AP information from the full shape [30]. A measurement of the BAO scale corresponds to a constraint in the $\{H_0 r_d, \Omega_m\}$ parameter space. When including the full-shape information on Ω_m , this tightens constraints by a factor of ~ 2 [28]. Since this additional constraining power comes from the broadband, the parameter ϕ_s is of particular interest. The key cosmological parameters in our compressed analysis are therefore ϕ_s , ϕ_p , and α_p . We return to this in section 3.3.1.

In order to proceed with the analysis, we need to model the power spectrum of the biased tracers. Following the modeling of [13] and [28], also detailed in [31], the model for the power spectrum is

$$P_{\text{Ly}\alpha}(k, \mu_k, z) = b_{\alpha}^2 (1 + \beta_{\alpha} \mu_k^2)^2 F_{\text{nl},\text{Ly}\alpha}^2(k, \mu_k) P(k, z), \quad (3.9)$$

for the Ly α forest and

$$P_{\times}(k, \mu_k, z) = b_{\alpha} (1 + \beta_{\alpha} \mu_k^2) (b_Q + f(z) \mu_k^2) F_{\text{nl},\text{QSO}}(k_{\parallel}) P(k, z) \quad (3.10)$$

for its cross-power spectrum with quasars. In each of the above equations, $P(k, z)$ is the linear isotropic matter power spectrum, and $\mu_k = k_{\parallel}/k$ where k is the modulus of the wave vector and $k_{\parallel}$ is its projection along the line of sight. For Ly α forest tracers, b_{α} is the linear bias of the Ly α forest, and the Ly α forest RSD parameter $\beta_{\alpha} = b_{\eta, \text{Ly}\alpha} f(z)/b_{\alpha}$ is degenerate with an unknown velocity divergence bias $b_{\eta, \text{Ly}\alpha}$ and the logarithmic growth rate $f(z)$ [32–36]. We treat both b_{α} and β_{α} as nuisance parameters to marginalize over. The $F_{\text{nl}, \text{Ly}\alpha}$ term encodes nonlinear corrections modeled by [63], whose parameters we keep constant. We restrict our main analysis to linear scales ($r \geq 30 h^{-1} \text{Mpc}$), so this choice should not significantly affect our results.

In the cross-power spectrum, the linear bias of quasars b_Q is also relevant, and the RSD term for quasars is $\beta_Q = f(z)/b_Q$. We treat the QSO linear bias as a nuisance parameter and marginalize over it. Lastly, we model the effects of redshift errors and nonlinear peculiar velocities of quasars with the parameter $F_{\text{nl}, \text{QSO}}(k_{\parallel})$, which we define by the Lorentzian function used in [13],

$$F_{\text{nl}, \text{QSO}}(k_{\parallel}) = \sqrt{[1 + (k_{\parallel} \sigma_z)^2]^{-1}}, \quad (3.11)$$

where the free parameter σ_z captures the effects of redshift errors and the velocity dispersion. This is also a nuisance parameter that we marginalize over.

3.3 Implementation

We perform our analysis by fitting the models for the two-point correlation functions to the noiseless synthetic data defined by the fiducial cosmology with the appropriate covariance matrix. During the sampling process, **Vega**³ [64] computes the theoretical correlation functions from the fiducial cosmology by decomposing the biased power spectrum into multipoles and transforming them into the correlation function with the FFTLog algorithm [65]. **Vega** then jointly fits the model correlation functions to the noiseless synthetic data using the user-provided covariance matrix to estimate uncertainties on the nuisance parameters and either compressed (section 3.3.1) or direct cosmological parameters (section 3.3.2).

3.3.1 Compressed analysis

In the compressed analysis, the correlation function is modeled by first decomposing the power spectrum into a peak (wiggles) and broadband (smooth) component [62]. This is done because the BAO peak represents a well-defined feature in the power spectrum, and the decomposition allows us to determine how much additional information is coming from the broadband. Since the BAO peak is localized to $\sim 100 h^{-1} \text{Mpc}$, the ability to constrain its position is robust against a variety of analysis choices and scale ranges. During the decomposition, nonlinear broadening of the BAO peak along and across the line of sight are also added as fixed parameters. The broadening is due to the nonlinear growth of structure, which is represented in the quasi-linear mock dataset [39].

We use **Vega** to fit the model correlation functions to the noiseless synthetic data. We fit the compressed parameters $\{\alpha_p, \alpha_s, \phi_p, \phi_s\}$ where the $\alpha_{p,s}$ parameters denote the isotropic

³<https://github.com/andreicuceu/vega>.

Parameter	Fiducial	68% limits			
		Distorted (180 h^{-1} Mpc)	Undistorted (180 h^{-1} Mpc)	Distorted (240 h^{-1} Mpc)	Undistorted (240 h^{-1} Mpc)
α_p	1.0	1.0 ± 0.0061	1.0 ± 0.0059	1.0 ± 0.0060	1.0 ± 0.0057
α_s	1.0	1.0 ± 0.012	1.0 ± 0.016	1.0 ± 0.012	1.0 ± 0.016
ϕ_p	1.0	1.0 ± 0.021	1.0 ± 0.018	1.0 ± 0.020	1.0 ± 0.018
ϕ_s	1.0	1.0 ± 0.0081	1.0 ± 0.0070	1.0 ± 0.0079	1.0 ± 0.0069
$f\sigma_8$	0.298	0.298 ± 0.025	0.298 ± 0.029	0.298 ± 0.024	0.298 ± 0.028
b_α	-0.1352	-0.1352 ± 0.0014	-0.1352 ± 0.0041	-0.1352 ± 0.0014	-0.1352 ± 0.0040
β_α	1.445	1.445 ± 0.027	1.445 ± 0.041	1.445 ± 0.025	1.445 ± 0.040
b_Q	3.545	3.545 ± 0.041	3.545 ± 0.10	3.545 ± 0.038	3.545 ± 0.10
σ_z	3.18	3.18 ± 0.82	3.18 ± 0.92	3.18 ± 0.79	3.18 ± 0.90
$\Delta r_\parallel$	0.0	0.0 ± 0.13	0.0 ± 0.14	0.0 ± 0.13	0.0 ± 0.13

Table 1. The full set of parameters for the compressed analysis. The second column lists the fiducial values used to generate the synthetic correlations. The last four columns show the one-dimensional marginal constraints in different cases, where the scale indicated on the second line denotes either $r_{\max} = 180$ or $240 h^{-1}$ Mpc. The units for σ_z and $\Delta r_\parallel$ are h^{-1} Mpc.

BAO scale (peak) or broadband (smooth) information, respectively, and $\phi_{p,s}$ denotes the same for the anisotropic (AP) information. We also fit and marginalize over the nuisance parameters $\{b_\alpha, \beta_\alpha, b_Q, \sigma_z, \Delta r_\parallel\}$. Table 1 shows the fit parameters and their fiducial values, including the derived quantity $f\sigma_8$.

3.3.2 Direct fit

In the direct fit framework, we sample cosmological parameters through a direct fit of the full shape of the 3D correlation functions. We use a custom likelihood for the **Cobaya** framework [66, 67] which uses **CAMB** [68] to compute the linear matter power spectrum and interfaces with **Vega** to model the correlation functions. We perform a joint fit to the full shape of the $\text{Ly}\alpha$ autocorrelation and $\text{Ly}\alpha \times \text{QSO}$ cross-correlation by sampling the parameter space with Markov chain Monte Carlo [MCMC; 69–71].

At each sampling step, **Cobaya** generates a template power spectrum over the range $10^{-3} < k < 10 h \text{ Mpc}^{-1}$. **Vega** transforms this full power spectrum into correlation function models for the $\text{Ly}\alpha$ auto- and cross-correlation without any separation into peak and smooth components. Because the decomposition is skipped in the direct fit analysis, this means that we do not add any nonlinear broadening to the BAO peak. The cosmological parameters we sample are $\{H_0, \Omega_m, \Omega_b h^2, A_s, n_s\}$, where H_0 is the Hubble constant, Ω_m is the present-day matter density parameter, Ω_b is the present-day baryon density parameter, and A_s and n_s are the amplitude and slope of the primordial power spectrum, respectively. We also sample the nuisance parameters $\{b_\alpha, \beta_\alpha, b_Q, \sigma_v\}$ and marginalize over them. Table 2 lists the parameters we sample with their fiducial values and priors. Any parameters that we do not explicitly sample are fixed to their Planck 2018 values [50].

Parameter	Fiducial	Prior	68% limits			
			Distorted (180 h^{-1} Mpc)	Undistorted (180 h^{-1} Mpc)	Distorted (240 h^{-1} Mpc)	Undistorted (240 h^{-1} Mpc)
H_0	67.36	$\mathcal{U}(40, 100)$	$69.7^{+2.1}_{-5.4}$	$69.8^{+2.3}_{-5.6}$	$69.5^{+2.0}_{-5.0}$	$69.7^{+2.3}_{-5.2}$
Ω_m	0.31509	$\mathcal{U}(0.01, 0.99)$	0.318 ± 0.011	0.318 ± 0.010	0.318 ± 0.011	0.318 ± 0.010
$\Omega_b h^2$	0.02237	$\mathcal{U}(0.01, 0.05)$	$0.0253^{+0.0022}_{-0.0064}$	$0.0255^{+0.0024}_{-0.0066}$	$0.0251^{+0.0020}_{-0.0060}$	$0.0253^{+0.0024}_{-0.0062}$
$10^9 A_s$	2.1	$\mathcal{U}_{\log(10^{10} A_s)}(0.5, 6)$	$1.98^{+0.29}_{-0.37}$	$1.97^{+0.32}_{-0.45}$	$1.99^{+0.29}_{-0.35}$	$1.97^{+0.32}_{-0.44}$
n_s	0.9649	$\mathcal{U}(0.8, 1.2)$	$0.950^{+0.035}_{-0.020}$	$0.946^{+0.042}_{-0.030}$	$0.951^{+0.032}_{-0.019}$	$0.947^{+0.040}_{-0.029}$
b_α	-0.1352	$\mathcal{U}_{\log(-b_\alpha)}(-2, 0)$	$-0.138^{+0.012}_{-0.0099}$	$-0.138^{+0.014}_{-0.011}$	$-0.137^{+0.012}_{-0.0094}$	$-0.138^{+0.014}_{-0.011}$
β_α	1.445	$\mathcal{U}(0, 5)$	1.448 ± 0.024	1.458 ± 0.046	1.448 ± 0.022	1.458 ± 0.045
b_Q	3.545	$\mathcal{U}_{\log(b_Q)}(-2, 1.3)$	$3.62^{+0.28}_{-0.35}$	$3.64^{+0.35}_{-0.46}$	$3.61^{+0.26}_{-0.33}$	$3.64^{+0.35}_{-0.45}$
σ_z	3.18	$\mathcal{U}(0, 15)$	$2.84^{+1.1}_{-0.67}$	$2.77^{+1.2}_{-0.78}$	$2.87^{+1.1}_{-0.65}$	$2.79^{+1.2}_{-0.75}$

Table 2. The full set of sampled parameters for direct cosmological inference. The second column lists the fiducial values used to generate the synthetic correlations. The priors for sampling are shown in the third column, where $\mathcal{U}$ denotes a uniform prior and the ‘log’ subscript indicates sampling in logarithmic space. We provide the one-dimensional marginal constraints in the last four columns for different cases, where the value corresponds to the mean of the posterior. The scale indicated on the second line denotes either $r_{\max} = 180$ or $240 h^{-1}$ Mpc. The units for H_0 are km/s/Mpc and the units for σ_z are h^{-1} Mpc.

4 Compressed analysis results

In this section, we present the results of our cosmological forecasts on compressed parameters using **Vega**. The following subsections explore three main categories of changes to the standard analysis: the impact of using the true continuum (section 4.1), variations in the scale range (section 4.2), and modifications to the covariance estimation (section 4.3).

Throughout this section, we fit the isotropic scale parameters $\{\alpha_p, \alpha_s\}$ and the AP scale parameters $\{\phi_p, \phi_s\}$. We marginalize over the nuisance parameters.

4.1 True continuum

The true continuum should provide better cosmological measurements, at least on scales that are not measured well in the continuum fitting approach. The exact scales impacted by the continuum fitting are difficult to quantify, as the effect is not localized in real space. If all forests were the same length, the continuum fitting procedure would erase all power on the scale of the forest length and larger, resulting in a sharp cutoff at this low- k mode in the power spectrum. For example, in the mock datasets employed here, the maximum forest length is $\sim 450 h^{-1}$ Mpc, corresponding to a complete loss of power at $k \lesssim 0.014 h \text{ Mpc}^{-1}$ in the distorted analysis. In practice, all forests are different lengths, resulting in a more gradual effect over a range of scales [72].

Figure 3 shows the forecast results from the distorted and undistorted analyses with unsmoothed covariances computed in the same way. Since these forecasts also employ the same scale cuts, this isolates the effect of the true continuum. We show constraints on three cosmological parameters ($\phi_s, \alpha_s, f\sigma_8$) and three nuisance parameters ($b_\alpha, \beta_\alpha, b_Q$).

The greatest improvement is seen in the broadband AP scale parameter ϕ_s , where the true continuum enables a nearly 15% improvement in constraining power over continuum fitting. The constraint on the BAO AP parameter ϕ_p also improves by $\sim 10\%$ with the true continuum, though this is not shown here.

Figure 3 also shows constraints on $f\sigma_8$, the linear biases, and the RSD parameter, all of which degrade when using the true continuum. In several cases, the degeneracy directions shift between the distorted and undistorted analyses, altering how information is recovered. The distortion matrix and projection formalism rely on a number of assumptions that have been validated for the BAO and AP parameters, but not yet for α_s or the bias terms [26, 30, 58]. Additionally, the distorted analysis allows small-scale information below $r_{\min}$ to leak into the fit [38]. This could cause a significant improvement in constraining power in the context of a forecast which fits an idealized synthetic data vector built from the model. In practice, we cannot yet model these small scales sufficiently well. As a result, constraints on these parameters may be biased in the distorted analysis, which could translate to a bias in the inferred growth rate, as seen in [26]. This also implies that our forecast improvement on AP parameters with the true continuum could be underestimated. Importantly, the Ly α forest is not a competitive probe of $f\sigma_8$ compared to galaxy surveys, and its most impactful cosmological constraint beyond BAO comes from the AP effect. Table 1 summarizes the one-dimensional marginal constraints on all parameters for two scale ranges in each analysis type.

The improvement on ϕ_s is also shown in figure 4, where the undistorted analysis yields the tightest constraints across all tested scale cuts. As previously mentioned, the true continuum analysis achieves $\sim 15\%$ tighter constraints than the unsmoothed distorted analysis, and this relative improvement remains roughly constant even as $r_{\max}$ increases. This consistency may be in part due to the fact that most Ly α forests in our mock dataset have lengths exceeding the largest scales probed here. Relative to the current baseline with a smoothed covariance, the improvement from the true continuum is $\sim 10\%$. This is in addition to the improvement with extensions to larger scales (see section 4.2).

The $\sim 15\%$ improvement on ϕ_s corresponds to a covariance rescaling factor of $f_{\text{cov}} \approx 0.85^2$ in the below relation (adapted from eq. 15 in [28]):

$$f_{\text{cov}} \propto \frac{A_{\text{distorted}}}{A_{\text{undistorted}}}, \quad (4.1)$$

where $A_{\text{undistorted}}$ and $A_{\text{distorted}}$ refer to the survey area in the undistorted and distorted analysis, respectively. Therefore, a 15% reduction in uncertainty corresponds to the constraining power achievable with an additional $\sim 40\%$ of Ly α forest survey area.

4.2 Changes in scale range

In addition to changes in the analysis method, additional cosmological information should be accessible when extending the analysis to a broader range of scales. To quantify the information gain from different physical scales, we examine how forecast constraints on ϕ_s change when varying the minimum and maximum separations used in the analysis. The current baseline continuum fitting analysis typically adopts a scale range of $r = 30\text{--}180 h^{-1} \text{ Mpc}$ [24]. Figure 4 shows the improvements in constraints on ϕ_s as a function of these isotropic scale

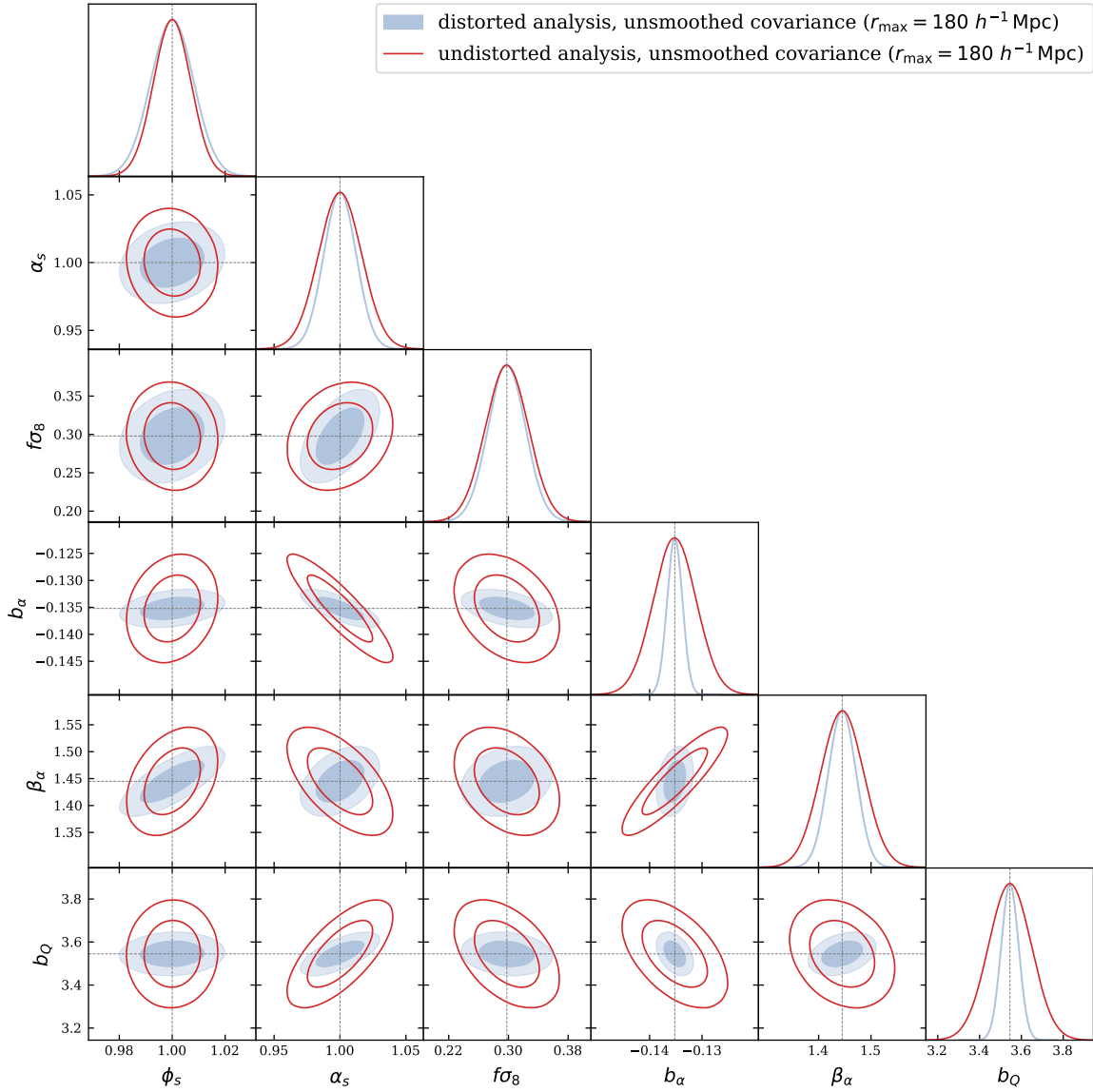


Figure 3. Forecast constraints on compressed cosmological parameters using two analyses at the standard scale range ($r_{\max} = 180 h^{-1} \text{ Mpc}$) with unsmoothed covariances: the continuum fitting (distorted) analysis (*solid blue contours*) and the true continuum (undistorted) analysis (*red contours*). This comparison isolates the impact of continuum fitting versus knowledge of the true continuum, as both analyses use identically computed covariance matrices with the same corrections applied. The advantage of the true continuum is evident in the broadband AP scale parameter, ϕ_s . The BAO AP scale parameter, ϕ_p , also improves with the true continuum but is not shown here (see table 1). See sections 2.2 and 4.1 for further discussion.

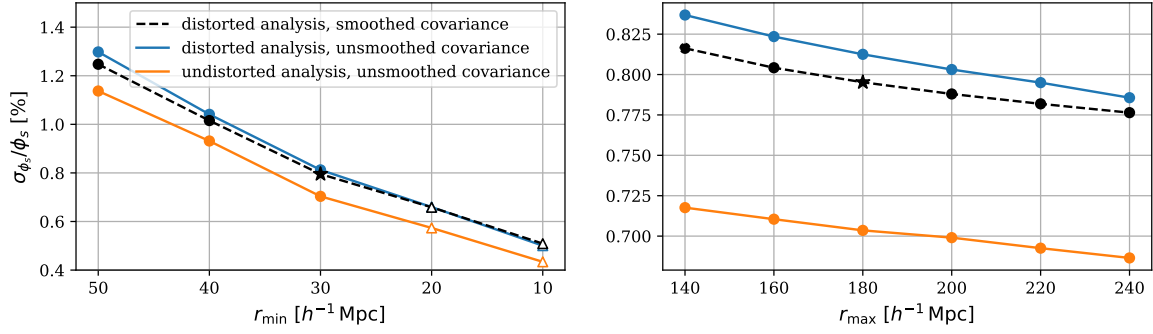


Figure 4. Forecast 1σ precision (%) on the AP broadband scale parameter ϕ_s for changes in the minimum and maximum scale range of the correlation function. The left-hand panel shows these as a function of the minimum isotropic scale cut $r_{\min}$: filled circles indicate a common cut for both the auto- and cross-correlation, while open triangles denote the specified cut in the auto-correlation and a fixed $r_{\min} = 30 h^{-1}$ Mpc in the cross-correlation. The right-hand panel shows the precision as a function of the maximum isotropic scale cut $r_{\max}$. Results from three analysis types are shown: the distorted analysis with a smoothed covariance matrix (*black dashed line*), the distorted continuum fitting analysis with an unsmoothed covariance matrix (*blue line*), and the undistorted true continuum analysis with an unsmoothed covariance matrix (*orange line*). The latter two analyses use covariances computed in the same way and include the Hartlap correction in the likelihood analysis. The current baseline analysis is indicated by the black stars and corresponds to the distorted analysis with a smoothed covariance and scale cuts of $r_{\min} = 30 h^{-1}$ Mpc and $r_{\max} = 180 h^{-1}$ Mpc.

cuts, comparing the current baseline analysis, the distorted analysis using an unsmoothed covariance, and the true continuum (undistorted) analysis with an unsmoothed covariance.

In the left-hand panel of figure 4, we vary $r_{\min}$ while fixing the maximum isotropic separation to $r_{\max} = 180 h^{-1}$ Mpc. The open triangles for $r_{\min} = 10$ and $20 h^{-1}$ Mpc indicate that the $r_{\min}$ choice was adopted for the auto-correlation but fixed at $r_{\min} = 30 h^{-1}$ Mpc for the cross-correlation. This is motivated by the fact that nonlinearities are more difficult to accurately model in the cross-correlation due to the stronger bias of quasars relative to the forest, whereas modeling these smaller scales may be more feasible for the auto-correlation. As shown in the figure, decreasing $r_{\min}$ results in similar improvements for all three analyses. Specifically, reducing $r_{\min}$ from 50 to $10 h^{-1}$ Mpc (in the auto-correlation) leads to a $\sim 60\%$ improvement on ϕ_s . Relative to the typical baseline analysis choice of $r_{\min} = 30 h^{-1}$ Mpc, the improvement enabled by decreasing the minimum scale cut to $r_{\min} = 10 h^{-1}$ Mpc in the auto-correlation is $\sim 40\%$, or $\sim 50\%$ when also exploiting the true continuum. However, for consistency, we adopt the standard minimum scale cut of $r_{\min} = 30 h^{-1}$ Mpc for the main results in this paper.

The right-hand panel of figure 4 shows the effect of increasing the maximum isotropic separation $r_{\max}$ while fixing the minimum scale cut to $r_{\min} = 30 h^{-1}$ Mpc. Increasing $r_{\max}$ from 140 to $240 h^{-1}$ Mpc results in a $\sim 5\%$ gain in information on ϕ_s , including for the distorted analysis. The advantage of the undistorted analysis over the distorted case remains roughly constant as $r_{\max}$ increases, as described in section 4.1. For the main results presented in this paper, we adopt maximum scale cuts of $r_{\max} = 180$ and $240 h^{-1}$ Mpc.

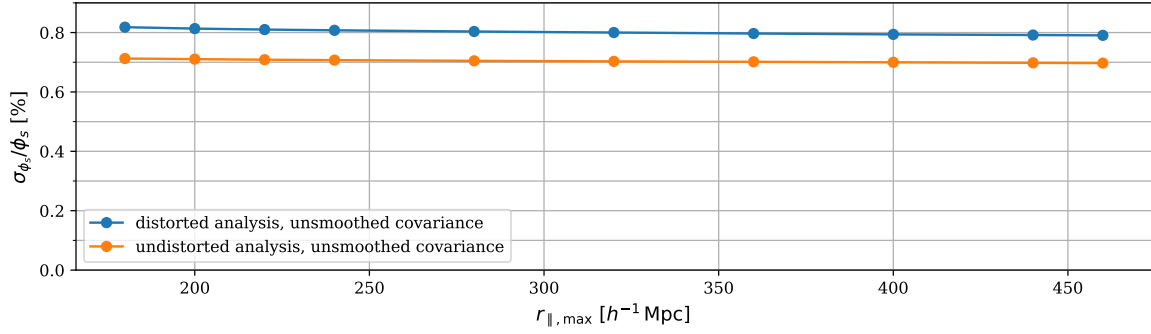


Figure 5. Forecast 1σ precision (%) on the AP broadband parameter ϕ_s with different line-of-sight scale cuts $r_{\parallel, \text{max}}$. The point at $r_{\text{max}} = 180 h^{-1} \text{Mpc}$ uses this cut both along and across the line of sight, while the remaining points fix the maximum transverse separation to $r_{\perp, \text{max}} = 200 h^{-1} \text{Mpc}$. The distorted (*blue*) and undistorted (*orange*) analyses both use unsmoothed covariances computed up to separations of $r_{\parallel, \text{max}} = 460 h^{-1} \text{Mpc}$ along and $r_{\perp, \text{max}} = 200 h^{-1} \text{Mpc}$ across the line of sight. This is in contrast to figure 4, where larger separations are probed in both directions. The undistorted analysis consistently outperforms the distorted case at all scale cuts tested, but information gain at very large scales is minimal.

We also investigate whether additional information can be extracted at even larger scales. To do this, we compute covariance matrices from correlation functions measured up to $r_{\parallel, \text{max}} = 460 h^{-1} \text{Mpc}$ along the line of sight and $r_{\perp, \text{max}} = 200 h^{-1} \text{Mpc}$ in the transverse direction. These anisotropic limits are chosen to ensure a sufficient number of independent sky regions when using $\text{NSIDE} = 16$ HEALPix pixels. In this case, the unsmoothed covariance is estimated from a stack of 150 mocks to compensate for the larger size of the data vector. The results of this test, for both the distorted and undistorted cases, are shown in figure 5. We find that the undistorted analysis consistently outperforms the distorted case by $\gtrsim 10\%$ across all tested scale cuts. The shallow slope of each curve indicates that only minimal additional information is gained at the largest scales; however, separations beyond $200 h^{-1} \text{Mpc}$ are only probed in the line-of-sight direction in this test.

4.3 Covariance modifications

We also explore the effects of choices we make to estimate the covariance. One key choice for estimating the covariance is the size of the HEALPix pixels used to compute individual correlation functions, since this affects the number of samples N , where $N = N_{\text{mocks}} \times N_{\text{HEALPix}}$. In our main analysis, we fix the number of mocks and the HEALPix size as described in section 3.1. The other choice directly affecting the covariance matrix is whether or not to apply the smoothing algorithm that is currently used in the baseline analysis.

In figure 4, the smoothed (current baseline) and unsmoothed distorted analyses show similar constraints ($\lesssim 3\%$) on the AP scale parameter. Therefore, the impact of covariance smoothing in the baseline distorted analysis is insignificant. Although not shown, we did not find a significant difference in the performance on the isotropic BAO parameter α_p between the current baseline and the unsmoothed distorted analysis.

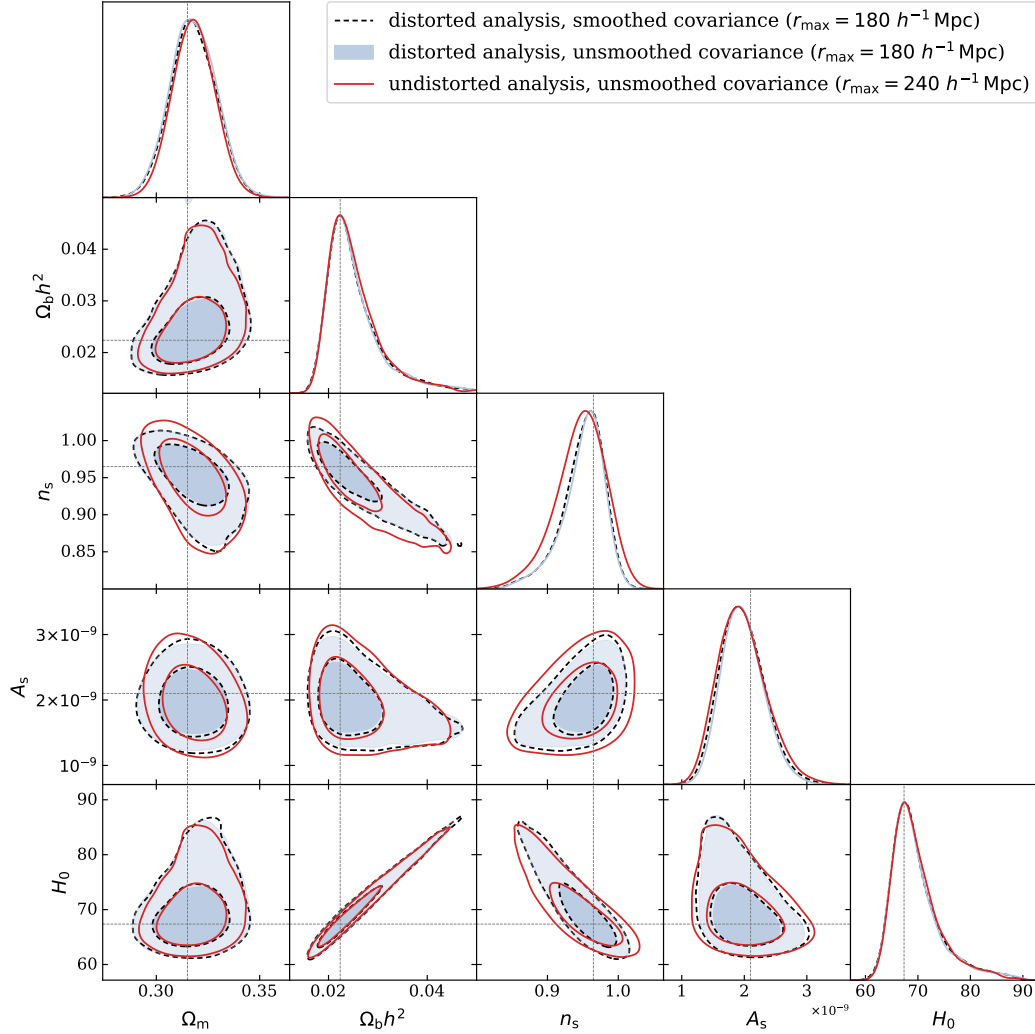


Figure 6. Forecast constraints on direct cosmological parameters obtained using MCMC sampling with *Cobaya*. We compare results from the continuum fitting (distorted) analysis with an unsmoothed (*solid blue contours*) or smoothed (*dashed black contours*) covariance and the standard maximum isotropic scale cut of $r_{\text{max}} = 180 h^{-1} \text{ Mpc}$ to those from the undistorted, unsmoothed covariance analysis with an extended scale range of $r_{\text{max}} = 240 h^{-1} \text{ Mpc}$ (*red contours*). The latter represents the best possible improvement on recovered cosmological information studied in this paper. We find the largest improvement on Ω_m (approximately 10%), which is consistent with our findings for compressed parameters in section 4. The degradation of constraints on other parameters with the true continuum is likely a consequence of the sub-optimal correlation function estimator (see sections 3.1 and 6).

5 Direct cosmology results

While the compressed parameters provide a valuable comparison point to models, they may not capture all of the cosmological information [31]. We therefore also perform direct cosmological inference with *Cobaya*. Figure 6 shows our forecasts for cosmological parameters from a direct fit to the full shape of the correlation functions. This figure shows results from the standard scale range of $r = 30\text{--}180 h^{-1} \text{ Mpc}$ for the distorted analysis with smoothed and

unsmoothed covariances, and a larger scale range of $r = 30\text{--}240\ h^{-1}\text{ Mpc}$ for the undistorted analysis. Table 2 lists the one-dimensional marginal constraints for the two different scale ranges for each of the unsmoothed analyses.

We find a $\sim 10\%$ improvement in Ω_m with the undistorted analysis, consistent with our results from the compressed forecasts. Interestingly, the constraints on the other flat Λ CDM cosmological parameters worsen in the undistorted case. This may be related to the degradation of bias parameter constraints when using the true continuum, as described in section 4.1. However, the $\text{Ly}\alpha$ forest alone provides only modest constraining power on these remaining parameters, which are much more tightly constrained in combination with other probes such as the cosmic microwave background [e.g. 31].

6 Discussion

We showed that knowledge of the true continuum leads to approximately 10% tighter constraints on the AP scale parameter and Ω_m compared to the current continuum fitting approach in the idealized case without contaminants. This improvement holds across all scale cuts tested. Moreover, additional constraining power is accessible at larger scales in both the continuum fitting (distorted) and true continuum (undistorted) analyses, leading to an overall $\sim 15\%$ improvement relative to the current baseline when extending the true continuum analysis to $r_{\text{max}} = 240\ h^{-1}\text{ Mpc}$. Interestingly, forecasts on other parameters such as $f\sigma_8$, n_s , and A_s revealed a degradation in constraining power with the true continuum; however, these parameters are not well constrained by the $\text{Ly}\alpha$ forest and may be prone to bias in the distorted analysis.

The results presented in this work rely on accurate estimates of the covariance matrix. As discussed in section 3.1, the current estimators for the $\text{Ly}\alpha$ two-point correlation function and its covariance are sub-optimal. While they have been extensively validated in the baseline distorted analysis, their impact in an undistorted analysis must be studied further. It is possible that the projection inherent in the distorted analysis results in artificially tight constraints on nuisance parameters and $f\sigma_8$, for example because the distorted analysis allows small-scale information to leak into the fit [38]. As a result, our measured improvements in the AP parameters with the true continuum may be conservative. Directly testing the impact of the sub-optimal estimator would require the development and validation of an optimal estimator. If one wishes to constrain RSD from the 3D $\text{Ly}\alpha$ forest, our results suggest that tighter constraints than the current baseline will first require the development of an optimal estimator for these parameters. In contrast, we find that access to the true continuum could already enable a $\sim 10\text{--}15\%$ improvement on the AP effect with the current estimators.

As shown in figure 3, the undistorted analysis yields different degeneracy directions between several parameters. For instance, the correlation coefficient between the linear bias b_α and the RSD parameter β_α shifts from roughly 0.29 in the distorted analysis to 0.88 in the undistorted case. Perhaps more importantly, the correlation coefficient between ϕ_s and α_s reduces from approximately 0.22 to -0.08 . This may be particularly beneficial for a full-shape cosmological analysis on data, as α_s may be biased by large-scale distortions introduced in the continuum fitting analysis. These changes in degeneracy structure may also be related to the correlation function estimator and the projection in the distorted analysis. Different degeneracy directions between nuisance parameters can also influence the

amount of cosmological information that may be present on different scales. Additionally, the cosmological information appears to be highly sensitive to the off-diagonal structure of the covariance matrix. For example, ignoring the cross-covariance between the auto- and cross-correlations in both the distorted baseline and undistorted analyses reduces the relative improvement on ϕ_s enabled by the true continuum from approximately 14% to 8%. This impact is lessened when comparing instead to the unsmoothed distorted case, where the relative improvement drops more modestly from roughly 16% to 12%. These shifts highlight the importance of including the cross-covariance to obtain more accurate uncertainty estimates and fully realize the benefits of the true continuum.

There may be even more cosmological information accessible at larger scales. We tested this with forecasts using covariance matrices computed up to line-of-sight separations of $r_{\parallel,\text{max}} = 460 h^{-1} \text{ Mpc}$, while we fixed the maximum transverse separation at $r_{\perp,\text{max}} = 200 h^{-1} \text{ Mpc}$ to avoid even larger HEALPix pixels and thus fewer subsamples for the covariance matrix measurement. While we found hints of extra information up to $r_{\text{max}} \approx 450 h^{-1} \text{ Mpc}$, the relative gain is modest, likely due to the weak clustering signal and the fact that we are only probing these large scales along the line-of-sight.

There are several ongoing and upcoming spectroscopic surveys such as DESI (including DESI-II [73]), Euclid [74], and Roman [75] that will extend our reach into the $z > 2$ universe. However, the next major advance in Ly α forest datasets is expected to come from a Stage-V spectroscopic survey with Spec-S5 [76], anticipated in about a decade. In this context, developing new analysis methods that extract additional information from existing and forthcoming data can significantly enhance the scientific return of these surveys. This is particularly timely given the emerging tensions with ΛCDM [23, 25, 77].

Full-shape measurements of the Ly α correlation functions using the new avenues explored in this work offer a promising path to extract more information from current datasets. To realize the gains suggested by this work, several future directions are essential. These include validating that predictions of the true continuum yield accurate two-point correlation measurements, and improving the estimation of the covariance matrix especially with extensions to larger scales. This may entail alternative binning strategies and validation of mock-based covariance estimates. Since we only observe one realization of the universe, it may be necessary to compute the covariance matrix with mocks or analytically, rather than relying on the assumptions currently employed to smooth the data covariance matrix (see section 3.1).

It may also be worthwhile to explore full-shape analyses in multipoles, as this would greatly simplify the covariance estimation and make the gains enabled by the true continuum more tractable. In addition, the true continuum could potentially enhance measurements of the three-dimensional power spectrum, especially in light of recent developments in optimal estimators [e.g. 78, 79]. Furthermore, extending the analysis to larger scales ($r_{\text{max}} > 180 h^{-1} \text{ Mpc}$) will require careful assessment and accurate modeling of potential systematics at these scales, such as UV background fluctuations [e.g. 80–83] and He II reionization [e.g. 84–86]. It will also be important to develop a more optimal estimator of the two-point correlation function, and ideally include the effects of continuum errors in this estimator. This would enable a more robust study of how continuum prediction errors impact cosmological constraints. In the context of the standard continuum fitting approach, the distortion matrix

relies on several assumptions that have so far only been tested within the regime typically studied [38]. Extending the analysis to new regimes will require validating the distortion matrix in those contexts, and may necessitate modifications or more complex algorithms if the underlying assumptions no longer hold. For example, it is necessary to compute the distortion matrix up to sufficiently large line-of-sight separations ($r_{\parallel} = 300 h^{-1} \text{ Mpc}$) to accurately model the auto-correlation up to $r_{\parallel} = 200 h^{-1} \text{ Mpc}$, but this is still insufficient for modeling the cross-correlation to $r_{\parallel} = 200 h^{-1} \text{ Mpc}$ [38]. The reliability of the distortion matrix at large scales will be an important consideration for extending the analysis to $r > 200 h^{-1} \text{ Mpc}$. However, our results indicate that most of the improvement in AP arises from the use of the true continuum, making this approach particularly promising.

Our work focused exclusively on the idealized case of $\text{Ly}\alpha$ forest measurements without contamination from metal absorption lines, optically thick absorbers, or broad absorption lines in the forest, and assuming perfect knowledge of the true continuum. Future studies will need to investigate the impact of these contaminants on the potential gains from the true continuum. Additionally, in practice, any continuum prediction method will introduce errors that will degrade our results. For example, current LyCAN predictions exhibit continuum errors at the $\sim 2\%$ level [46]. Quantifying the impact of these errors on cosmological constraints will be an important direction for future work, as the exact level of degradation will depend on whether the errors are correlated and how well we can model them. We plan to use improved versions of LyCAN to measure the two-point correlation functions with DESI data in a future paper, enabling the next step toward full-shape $\text{Ly}\alpha$ forest cosmology with the true continuum.

7 Summary

We presented the first forecasts with the $\text{Ly}\alpha$ forest auto-correlation and its cross-correlation with quasars that are unaffected by continuum fitting distortion. We performed forecasts on compressed cosmological parameters through maximum likelihood estimation with **Vega** and sampled posteriors of direct cosmological parameters using a custom likelihood for **Cobaya** that interfaces with **Vega**. Given a model power spectrum, **Vega** creates a model for the two-point correlation functions and fits the model to the noiseless synthetic data defined by our fiducial cosmology.

To forecast constraints on cosmological parameters, we required estimates of the covariance matrix for the distorted and undistorted analyses. To study the additional constraining power from larger scales, we computed covariance matrices over separations from $r = 30\text{--}240 h^{-1} \text{ Mpc}$ using 50 mocks with HEALPix pixels covering areas of linear size $\sim 280 h^{-1} \text{ Mpc}$. We applied the Hartlap [59] correction to the inverse covariance matrix in our likelihood analysis and validated its robustness for our specific regime of sample size and number of data bins in appendix B.

We found that using the true continuum and extending the analysis to $r_{\text{max}} = 240 h^{-1} \text{ Mpc}$ can enable up to $\sim 15\%$ tighter constraints on the AP scale parameter and Ω_{m} when compared to the baseline continuum fitting analysis at $r_{\text{max}} = 180 h^{-1} \text{ Mpc}$. This improvement is analogous to an extra $\sim 40\%$ of $\text{Ly}\alpha$ forest survey area. When not exploiting larger scales and instead comparing constraints using the same standard scale range of $r = 30\text{--}180 h^{-1} \text{ Mpc}$, knowledge of the true continuum still results in $\sim 10\%$ tighter constraints on these parameters, analogous to an additional $\sim 25\%$ of $\text{Ly}\alpha$ forest survey area.

To explore even larger scales, we extended our forecasts on compressed parameters to include separations up to $r_{\text{max}} = 460 h^{-1}$ Mpc. This required computing covariance matrices from a larger set of mocks (150), using HEALPix pixels with areas of linear size $\sim 250 h^{-1}$ Mpc. This setup allowed us to probe scales up to $r_{\parallel, \text{max}} = 460 h^{-1}$ Mpc along the line of sight, while limiting the transverse separations to $r_{\perp, \text{max}} = 200 h^{-1}$ Mpc. We found evidence for additional constraining power at these larger scales; however, extending only the line-of-sight range provided minimal improvement.

This work has significant implications for cosmology. Recent DESI results have found evidence for significant tension with the Planck Λ CDM model [23, 25]. While results from the full five-year DESI survey (DR3) may increase the statistical significance of this tension, future major advancements in cosmology will arise from next-generation surveys including Roman, Rubin, and Spec-S5. In the meantime, meaningful progress will rely on improved analysis techniques that can extract more information from existing data. The forecast improvement in the Alcock-Paczyński parameter from this work suggests that the $\text{Ly}\alpha$ forest may soon have the ability to help distinguish between the best-fit Λ CDM and $w_0 w_a$ CDM models. Further, the improved precision on Ω_m offers a promising avenue for shedding light on the emerging tension between DESI and Planck measurements within the Λ CDM model.

Acknowledgments

We thank Naim Karaçaylı and David Weinberg for helpful discussions. WT and PM acknowledge support from the United States Department of Energy, Office of High Energy Physics under Award Number DE-SC-0011726. AC acknowledges support provided by NASA through the NASA Hubble Fellowship grant HST-HF2-51526.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

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

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

Layer type	Size	Activation function
Conv1D (input)	64	ReLU
MaxPooling1D	2	
Conv1D	16	ReLU
MaxPooling1D	2	
Conv1D	48	ReLU
MaxPooling1D	2	
Conv1D	32	ReLU
MaxPooling1D	2	
Flatten		
Dense	96	ReLU
Dense	160	ReLU
Dense	224	ReLU
Dense	256	ReLU
Dense (output)	699	Linear

Table 3. Updated LyCAN architecture. The first column specifies the layer type. The *size* column specifies the number of filters for convolutional layers or neurons for dense layers. In the case of the max pooling layers, pixels are pooled in groups of two. We use the linear activation function for the output layer, and the ReLU activation function for all other layers where applicable. The number of nodes in the output layer is equal to the number of pixels in the predicted continuum. The kernel size is six for the convolutional layers.

A Potential feasibility with LyCAN

The Ly α Continuum Analysis Network [LyCAN; 46] is a convolutional neural network that predicts the unabsorbed quasar continuum between 1040–1600 Å strictly based on the longer-wavelength region (1216–1600 Å). Since the network predicts the unabsorbed continuum independently of the forest region, it retains large-scale information and may eliminate the need for the distortion matrix. LyCAN was trained on a combination of DESI mock spectra representative of the five-year survey and synthetic spectra generated from low- z Hubble Space Telescope (HST)/Cosmic Origins Spectrograph (COS) observations. It achieves excellent performance on spectra from these datasets reserved for testing purposes ($\lesssim 2\%$ median error in the forest region).

We have made several updates to LyCAN since its introduction in [46]. While the original network demonstrated strong overall performance in forest continuum prediction, it introduced some spurious correlations in the 3D Ly α correlation functions relative to the true continuum case. We found that this was due to the addition of systematic perturbations in the synthetic spectra generated from high-resolution HST/COS data to dampen emission line strengths (see section 3.2 of [46]). After modifying this perturbation scheme to be centered around zero, we re-optimized and re-trained LyCAN using the procedures described in section 4.1 of [46]. This updated version of LyCAN yields an undistorted measurement of the correlation function when applied to a DESI mock reserved for testing that is representative of the five-year survey, as illustrated in figure 7. This result motivates this work, as a continuum prediction method that produces undistorted correlation function measurements may soon be within reach. The revised neural network architecture is presented in table 3.

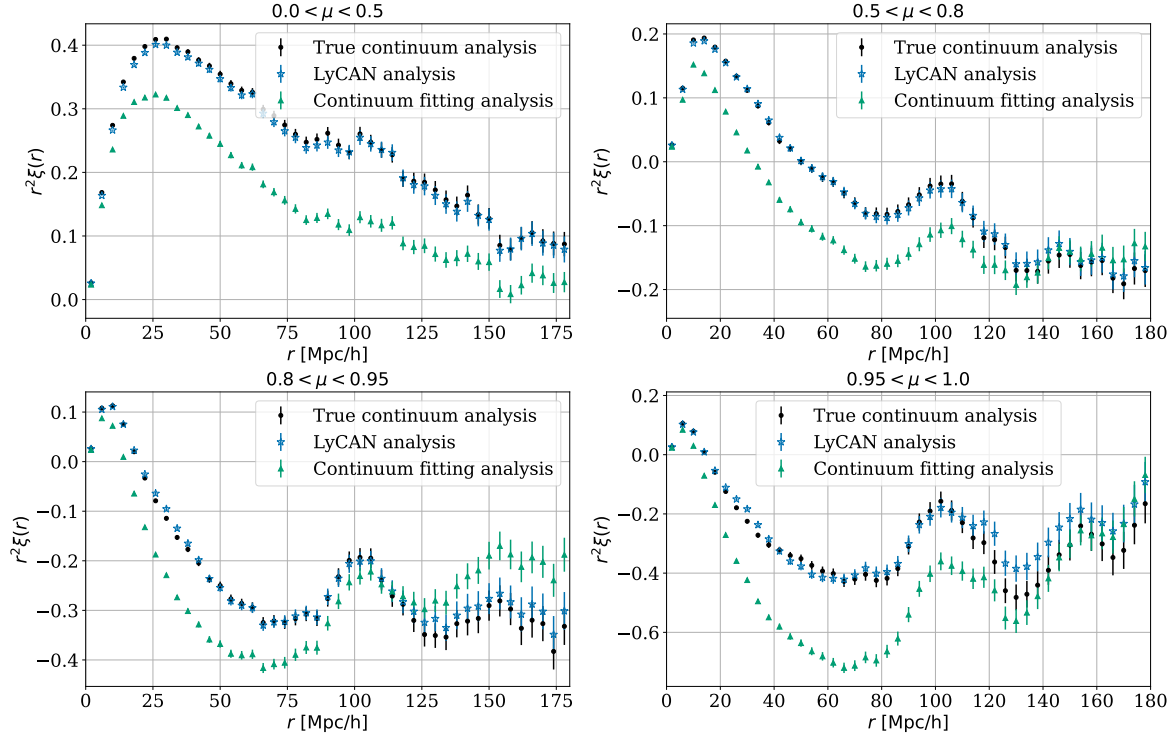


Figure 7. Three measurements of the Ly α autocorrelation function $\xi(r)$ in four wedges of $\mu = r_{\parallel}/r$ on an uncontaminated DESI mock representative of the full five-year survey: the measurement from the standard continuum fitting approach (*teal*), the measurement from the true continuum analysis (*black*), and the measurement from LyCAN-predicted continua (*blue*). These measurements are shown for illustrative purposes, demonstrating the future feasibility of a LyCAN-based full-shape analysis.

LyCAN continuum predictions also show slight biases with respect to signal-to-noise ratio (SNR) and redshift, under-predicting the true continuum in low-SNR or low-redshift spectra and over-predicting it at high redshift [46]. While this is a small effect ($\sim 2\%$, as seen in figure 12 of [46]), it may still warrant correction or further network optimization before LyCAN can be reliably applied to measure undistorted correlation functions. Ongoing efforts aim to improve LyCAN and mitigate these biases in order to achieve accurate correlation function measurements across a wide range of mocks. We therefore do not use LyCAN continuum predictions in our forecasts and instead focus on the idealized case of the true continuum.

B Validation of the Hartlap correction

When a covariance matrix is estimated from a limited number of samples relative to the size of the data vector, its inverse is a biased estimate of the true inverse covariance [e.g. 59, 60, 87]. If left uncorrected, this bias leads to underestimated uncertainties on inferred parameters. A common mitigation is the Hartlap correction [59], which rescales the inverse covariance matrix during likelihood evaluation:

$$h = \frac{N - p - 1}{N}, \quad (\text{B.1})$$

where N is the number of independent samples used to estimate the covariance matrix and p is the size of the data vector.

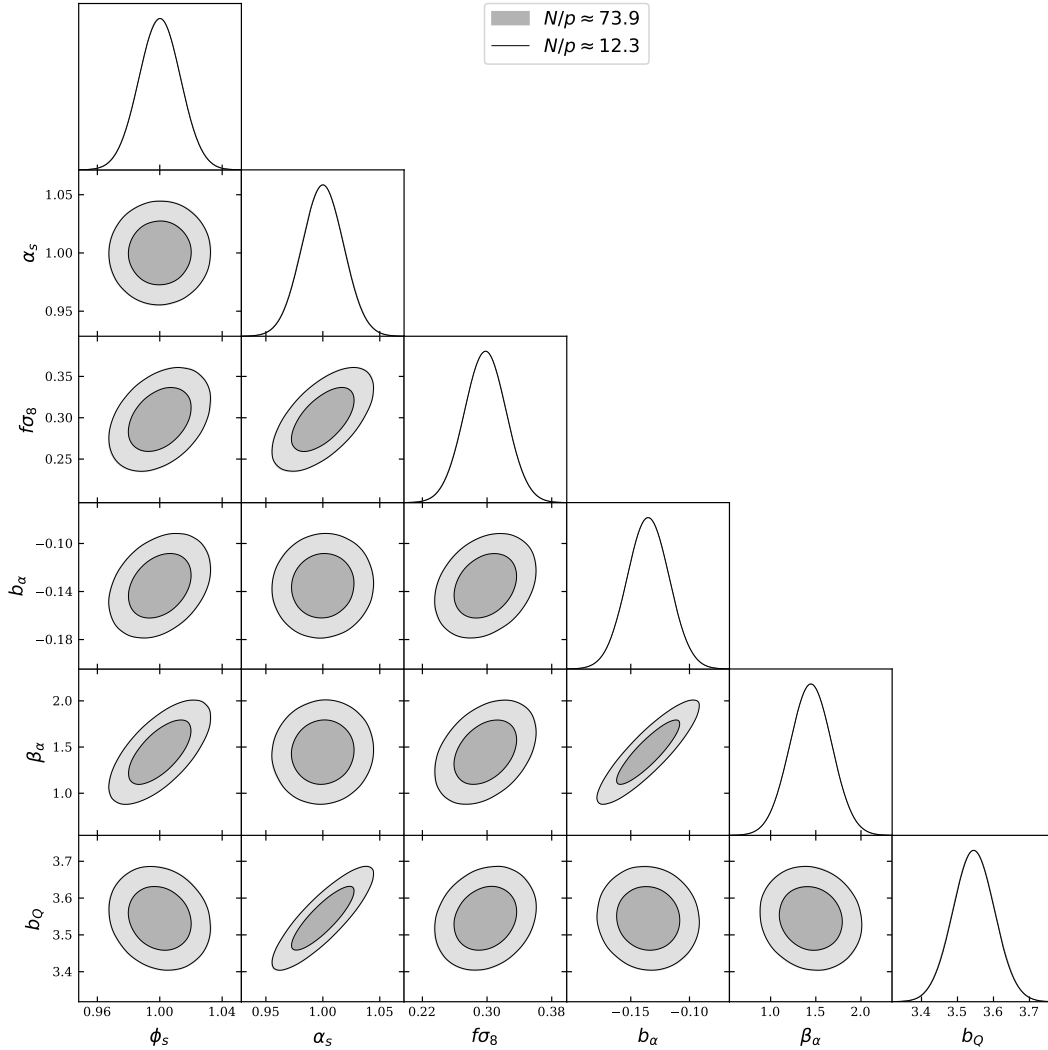


Figure 8. Forecasts using two covariance matrices from fully contaminated NSIDE=16 mocks based on stacks of 300 ($N/p \approx 73.9$, *solid gray contours*) or 50 ($N/p \approx 12.3$, *black contours*). Appropriate Hartlap corrections are applied in each case. The results are indistinguishable, validating the correction in the lower N/p regime.

However, the Hartlap correction is only an approximate solution, and its accuracy degrades as N approaches p . To validate its reliability in our analysis, we test its performance across a range of N/p values. We first compute a reference covariance matrix from 300 fully contaminated mocks in the baseline continuum fitting analysis, which includes separations from $30\text{--}180\ h^{-1}\text{ Mpc}$ in NSIDE=16 HEALPix pixels. This setup corresponds to $N = 344,040$ measurements and a data vector of size $p = 4653$, yielding $N/p \approx 74$. In this regime, the Hartlap correction is negligible and the covariance is expected to be robust. We then compute the same covariance using only 50 fully contaminated mocks, matching the number of realizations available for our uncontaminated analysis. We apply the corresponding Hartlap correction in both cases and perform maximum likelihood analysis. As shown in figure 8, the resulting constraints are indistinguishable, confirming that the Hartlap-corrected inverse covariance is reliable at $N/p \approx 12$.

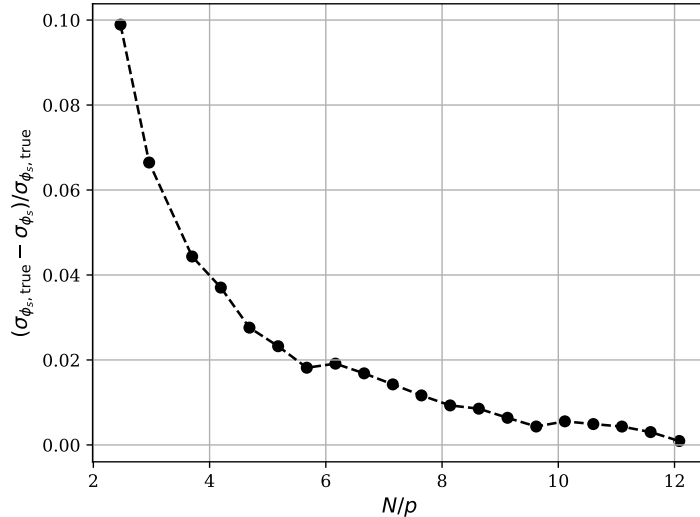


Figure 9. Fractional error in the inferred uncertainty on ϕ_s in different N/p regimes using Hartlap-corrected inverse covariances. We require $N/p \gtrsim 5$ to limit our error on inferred parameter uncertainties to within $\sim 2\%$.

To test the limits of this approximation, we compute additional covariance matrices from the uncontaminated mocks by varying the number of realizations and examining the resulting forecasts for ϕ_s . For each value of N , we apply the Hartlap correction and evaluate the fractional error in the inferred uncertainty on ϕ_s relative to the result from the full 50-mock covariance (our reference “truth”). The results are shown in figure 9. We find that achieving $N/p \gtrsim 5$ is sufficient to limit errors in parameter uncertainties to within $\sim 2\%$. This criterion motivates our use of covariances computed from $\text{NSIDE}=14$ HEALPix pixels up to $240 h^{-1} \text{Mpc}$ in the main analysis.

We note that the Hartlap factor does not account for the additional uncertainty arising from the sample covariance matrix being a single realization of an underlying distribution, which propagates to increased parameter uncertainties [88]. This results in a non-Gaussian likelihood [89], but can be approximated using a Gaussian likelihood and additional correction factors [e.g. 60, 87, 88]. We do not apply these additional corrections since our validation tests demonstrate that the Hartlap correction alone can sufficiently recover the truth in our analysis.

Data Availability Statement. Data points from each figure are available at <https://doi.org/10.5281/zenodo.19927452>.

References

- [1] SUPERNOVA SEARCH TEAM collaboration, *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, *Astron. J.* **116** (1998) 1009 [[astro-ph/9805201](#)] [[INSPIRE](#)].
- [2] SUPERNOVA COSMOLOGY PROJECT collaboration, *Measurements of Ω and Λ from 42 High Redshift Supernovae*, *Astrophys. J.* **517** (1999) 565 [[astro-ph/9812133](#)] [[INSPIRE](#)].

- [3] 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](#)] [[INSPIRE](#)].
- [4] SDSS collaboration, *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies*, *Astrophys. J.* **633** (2005) 560 [[astro-ph/0501171](#)] [[INSPIRE](#)].
- [5] BOSS collaboration, *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *Astron. J.* **145** (2013) 10 [[arXiv:1208.0022](#)] [[INSPIRE](#)].
- [6] 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](#)] [[INSPIRE](#)].
- [7] BOSS collaboration, *Baryon Acoustic Oscillations in the Ly- α forest of BOSS quasars*, *Astron. Astrophys.* **552** (2013) A96 [[arXiv:1211.2616](#)] [[INSPIRE](#)].
- [8] BOSS collaboration, *Measurement of Baryon Acoustic Oscillations in the Lyman- α Forest Fluctuations in BOSS Data Release 9*, *JCAP* **04** (2013) 026 [[arXiv:1301.3459](#)] [[INSPIRE](#)].
- [9] BOSS collaboration, *Quasar-Lyman α Forest Cross-Correlation from BOSS DR11: Baryon Acoustic Oscillations*, *JCAP* **05** (2014) 027 [[arXiv:1311.1767](#)] [[INSPIRE](#)].
- [10] EBOSS collaboration, *The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *Astron. J.* **151** (2016) 44 [[arXiv:1508.04473](#)] [[INSPIRE](#)].
- [11] EBOSS collaboration, *Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies and the Distant Universe*, *Astron. J.* **154** (2017) 28 [[arXiv:1703.00052](#)] [[INSPIRE](#)].
- [12] 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](#)] [[INSPIRE](#)].
- [13] EBOSS collaboration, *The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Ly α Forests*, *Astrophys. J.* **901** (2020) 153 [[arXiv:2007.08995](#)] [[INSPIRE](#)].
- [14] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, [arXiv:1611.00036](#) [[INSPIRE](#)].
- [15] DESI collaboration, *The DESI Experiment Part II: Instrument Design*, [arXiv:1611.00037](#) [[INSPIRE](#)].
- [16] DESI collaboration, *The Optical Corrector for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **168** (2024) 95 [[arXiv:2306.06310](#)] [[INSPIRE](#)].
- [17] C. Poppett et al., *Overview of the Fiber System for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **168** (2024) 245.
- [18] DESI collaboration, *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **164** (2022) 207 [[arXiv:2205.10939](#)] [[INSPIRE](#)].
- [19] DESI collaboration, *Data Release 1 of the Dark Energy Spectroscopic Instrument*, *Astron. J.* **171** (2026) 285 [[arXiv:2503.14745](#)] [[INSPIRE](#)].
- [20] DESI collaboration, *Survey Operations for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **166** (2023) 259 [[arXiv:2306.06309](#)] [[INSPIRE](#)].
- [21] DESI collaboration, *The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **165** (2023) 144 [[arXiv:2209.14482](#)] [[INSPIRE](#)].

- [22] DESI collaboration, *Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **167** (2024) 62 [[arXiv:2306.06307](#)] [[INSPIRE](#)].
- [23] DESI collaboration, *DESI 2024 VII: cosmological constraints from the full-shape modeling of clustering measurements*, *JCAP* **07** (2025) 028 [[arXiv:2411.12022](#)] [[INSPIRE](#)].
- [24] DESI collaboration, *DESI DR2 results. I. Baryon acoustic oscillations from the Lyman alpha forest*, *Phys. Rev. D* **112** (2025) 083514 [[arXiv:2503.14739](#)] [[INSPIRE](#)].
- [25] DESI collaboration, *DESI DR2 results. II. Measurements of baryon acoustic oscillations and cosmological constraints*, *Phys. Rev. D* **112** (2025) 083515 [[arXiv:2503.14738](#)] [[INSPIRE](#)].
- [26] A. Cuceu et al., *DESI DR1 Ly α forest: 3D full-shape analysis and cosmological constraints*, [arXiv:2509.15308](#) [[INSPIRE](#)].
- [27] C. Alcock and B. Paczyński, *An evolution free test for non-zero cosmological constant*, *Nature* **281** (1979) 358 [[INSPIRE](#)].
- [28] A. Cuceu, A. Font-Ribera, B. Joachimi and S. Nadathur, *Cosmology beyond BAO from the 3D distribution of the Lyman- α forest*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 5439 [[arXiv:2103.14075](#)] [[INSPIRE](#)].
- [29] A. Cuceu et al., *Constraints on the Cosmic Expansion Rate at Redshift 2.3 from the Lyman- α Forest*, *Phys. Rev. Lett.* **130** (2023) 191003 [[arXiv:2209.13942](#)] [[INSPIRE](#)].
- [30] A. Cuceu et al., *The Alcock-Paczyński effect from Lyman- α forest correlations: analysis validation with synthetic data*, *Mon. Not. Roy. Astron. Soc.* **523** (2023) 3773 [[arXiv:2209.12931](#)] [[INSPIRE](#)].
- [31] F. Gerardi et al., *Direct cosmological inference from three-dimensional correlations of the Lyman α forest*, *Mon. Not. Roy. Astron. Soc.* **518** (2022) 2567 [[arXiv:2209.11263](#)] [[INSPIRE](#)].
- [32] P. McDonald, *Toward a Measurement of the Cosmological Geometry at $z \sim 2$: Predicting Ly α Forest Correlation in Three Dimensions and the Potential of Future Data Sets*, *Astrophys. J.* **585** (2003) 34 [[astro-ph/0108064](#)] [[INSPIRE](#)].
- [33] BOSS collaboration, *The Lyman- α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data*, *JCAP* **09** (2011) 001 [[arXiv:1104.5244](#)] [[INSPIRE](#)].
- [34] U. Seljak, *Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Lyman α forest*, *JCAP* **03** (2012) 004 [[arXiv:1201.0594](#)] [[INSPIRE](#)].
- [35] J.J. Givans and C.M. Hirata, *Redshift-space streaming velocity effects on the Lyman- α forest baryon acoustic oscillation scale*, *Phys. Rev. D* **102** (2020) 023515 [[arXiv:2002.12296](#)] [[INSPIRE](#)].
- [36] S.-F. Chen, Z. Vlah and M. White, *The Ly α forest flux correlation function: a perturbation theory perspective*, *JCAP* **05** (2021) 053 [[arXiv:2103.13498](#)] [[INSPIRE](#)].
- [37] BOSS collaboration, *Measurement of baryon acoustic oscillation correlations at $z = 2.3$ with SDSS DR12 Ly α -Forests*, *Astron. Astrophys.* **603** (2017) A12 [[arXiv:1702.00176](#)] [[INSPIRE](#)].
- [38] N. Busca et al., *The effects of continuum fitting on Lyman- α forest correlations*, *JCAP* **09** (2025) 020 [[arXiv:2506.15262](#)] [[INSPIRE](#)].
- [39] DESI collaboration, *Validation of the DESI DR2 Ly α BAO analysis using synthetic datasets*, *Phys. Rev. D* **113** (2026) 023520 [[arXiv:2503.14741](#)] [[INSPIRE](#)].
- [40] N. Suzuki et al., *Predicting QSO continua in the Ly- α forest*, *Astrophys. J.* **618** (2005) 592 [[astro-ph/0306577](#)] [[INSPIRE](#)].

- [41] I. Pâris et al., *A Principal Component Analysis of quasar UV spectra at $z \sim 3$* , *Astron. Astrophys.* **530** (2011) A50 [[arXiv:1104.2024](#)] [[INSPIRE](#)].
- [42] K.-G. Lee, N. Suzuki and D.N. Spergel, *Mean-flux Regulated PCA Continuum Fitting of SDSS Lyman-alpha Forest Spectra*, *Astron. J.* **143** (2012) 51 [[arXiv:1108.6080](#)] [[INSPIRE](#)].
- [43] G. Zhu, *Nonnegative Matrix Factorization (NMF) with Heteroscedastic Uncertainties and Missing data*, [arXiv:1612.06037](#).
- [44] B. Liu and R. Bordoloi, *A deep learning approach to quasar continuum prediction*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 3510.
- [45] Z. Sun, Y.-S. Ting and Z. Cai, *Quasar Factor Analysis — An Unsupervised and Probabilistic Quasar Continuum Prediction Algorithm with Latent Factor Analysis*, *Astrophys. J. Suppl.* **269** (2023) 4 [[arXiv:2211.11784](#)] [[INSPIRE](#)].
- [46] W. Turner et al., *New measurements of the Lyman- α forest continuum and effective optical depth with LyCAN and DESI Y1 data*, *Astrophys. J.* **976** (2024) 143 [[arXiv:2405.06743](#)] [[INSPIRE](#)].
- [47] C.H. Hahn et al., *Reconstructing Quasar Spectra and Measuring the Ly α Forest with SpenderQ*, [arXiv:2506.18986](#) [[INSPIRE](#)].
- [48] F. Pistis et al., *Automated quasar continuum estimation using neural networks: A comparative study of deep-learning architectures*, *Astron. Astrophys.* **698** (2025) A292 [[arXiv:2505.10976](#)].
- [49] DESI collaboration, *DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman alpha forest*, *JCAP* **01** (2025) 124 [[arXiv:2404.03001](#)] [[INSPIRE](#)].
- [50] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [[arXiv:1807.06209](#)] [[INSPIRE](#)].
- [51] H.K. Herrera-Alcantar et al., *Synthetic spectra for Lyman- α forest analysis in the Dark Energy Spectroscopic Instrument*, *JCAP* **01** (2025) 141 [[arXiv:2401.00303](#)] [[INSPIRE](#)].
- [52] I. McGreer, J. Moustakas and J. Schindler, *simqso: Simulated quasar spectra generator*, Astrophysics Source Code Library, record [ascl:2106.008](#) (2021).
- [53] J. Farr et al., *LyaCoLoRe: synthetic datasets for current and future Lyman- α forest BAO surveys*, *JCAP* **03** (2020) 068 [[arXiv:1912.02763](#)] [[INSPIRE](#)].
- [54] DESI collaboration, *The Lyman- α forest catalog from the Dark Energy Spectroscopic Instrument Early Data Release*, *Mon. Not. Roy. Astron. Soc.* **528** (2024) 6666 [[arXiv:2306.06312](#)] [[INSPIRE](#)].
- [55] BOSS collaboration, *Baryon acoustic oscillations from the complete SDSS-III Ly α -quasar cross-correlation function at $z = 2.4$* , *Astron. Astrophys.* **608** (2017) A130 [[arXiv:1708.02225](#)] [[INSPIRE](#)].
- [56] K.M. Górski et al., *HEALPix — A Framework for high resolution discretization, and fast analysis of data distributed on the sphere*, *Astrophys. J.* **622** (2005) 759 [[astro-ph/0409513](#)] [[INSPIRE](#)].
- [57] BOSS collaboration, *Baryon acoustic oscillations in the Ly α forest of BOSS DR11 quasars*, *Astron. Astrophys.* **574** (2015) A59 [[arXiv:1404.1801](#)] [[INSPIRE](#)].
- [58] DESI collaboration, *Validation of the DESI 2024 Ly α forest BAO analysis using synthetic datasets*, *JCAP* **01** (2025) 148 [[arXiv:2404.03004](#)] [[INSPIRE](#)].
- [59] J. Hartlap, P. Simon and P. Schneider, *Why your model parameter confidences might be too optimistic: Unbiased estimation of the inverse covariance matrix*, *Astron. Astrophys.* **464** (2007) 399 [[astro-ph/0608064](#)] [[INSPIRE](#)].

- [60] W.J. Percival et al., *The Clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Including covariance matrix errors*, *Mon. Not. Roy. Astron. Soc.* **439** (2014) 2531 [[arXiv:1312.4841](#)] [[INSPIRE](#)].
- [61] EBOSS collaboration, *Baryon acoustic oscillations at $z = 2.34$ from the correlations of Ly α absorption in eBOSS DR14*, *Astron. Astrophys.* **629** (2019) A85 [[arXiv:1904.03400](#)] [[INSPIRE](#)].
- [62] BOSS collaboration, *Fitting Methods for Baryon Acoustic Oscillations in the Lyman- α Forest Fluctuations in BOSS Data Release 9*, *JCAP* **03** (2013) 024 [[arXiv:1301.3456](#)] [[INSPIRE](#)].
- [63] A. Arinyo-i-Prats, J. Miralda-Escudé, M. Viel and R. Cen, *The Non-Linear Power Spectrum of the Lyman Alpha Forest*, *JCAP* **12** (2015) 017 [[arXiv:1506.04519](#)] [[INSPIRE](#)].
- [64] A. Cuceu, A. Font-Ribera and B. Joachimi, *Bayesian methods for fitting Baryon Acoustic Oscillations in the Lyman- α forest*, *JCAP* **07** (2020) 035 [[arXiv:2004.02761](#)] [[INSPIRE](#)].
- [65] A.J.S. Hamilton, *Uncorrelated modes of the nonlinear power spectrum*, *Mon. Not. Roy. Astron. Soc.* **312** (2000) 257 [[astro-ph/9905191](#)] [[INSPIRE](#)].
- [66] J. Torrado and A. Lewis, *Cobaya: Bayesian analysis in cosmology*, Astrophysics Source Code Library, record [ascl:1910.019](#) (2019).
- [67] J. Torrado and A. Lewis, *Cobaya: Code for Bayesian Analysis of hierarchical physical models*, *JCAP* **05** (2021) 057 [[arXiv:2005.05290](#)] [[INSPIRE](#)].
- [68] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473 [[astro-ph/9911177](#)] [[INSPIRE](#)].
- [69] A. Lewis and S. Bridle, *Cosmological parameters from CMB and other data: A Monte Carlo approach*, *Phys. Rev. D* **66** (2002) 103511 [[astro-ph/0205436](#)] [[INSPIRE](#)].
- [70] R.M. Neal, *Taking Bigger Metropolis Steps by Dragging Fast Variables*, [math/0502099](#) [[INSPIRE](#)].
- [71] A. Lewis, *Efficient sampling of fast and slow cosmological parameters*, *Phys. Rev. D* **87** (2013) 103529 [[arXiv:1304.4473](#)] [[INSPIRE](#)].
- [72] M. Blomqvist et al., *Broadband distortion modeling in Lyman- α forest BAO fitting*, *JCAP* **11** (2015) 034 [[arXiv:1504.06656](#)] [[INSPIRE](#)].
- [73] DESI collaboration, *A Spectroscopic Road Map for Cosmic Frontier: DESI, DESI-II, Stage-5*, [arXiv:2209.03585](#) [[INSPIRE](#)].
- [74] EUCLID collaboration, *Euclid. I. Overview of the Euclid mission*, *Astron. Astrophys.* **697** (2025) A1 [[arXiv:2405.13491](#)] [[INSPIRE](#)].
- [75] Y. Wang et al., *The High Latitude Spectroscopic Survey on the Nancy Grace Roman Space Telescope*, *Astrophys. J.* **928** (2022) 1 [[arXiv:2110.01829](#)] [[INSPIRE](#)].
- [76] SPEC-S5 collaboration, *The Spectroscopic Stage-5 Experiment*, [arXiv:2503.07923](#) [[INSPIRE](#)].
- [77] DES collaboration, *Dark Energy Survey: Implications for cosmological expansion models from the final DES baryon acoustic oscillation and supernova data*, *Phys. Rev. D* **113** (2026) 063530 [[arXiv:2503.06712](#)] [[INSPIRE](#)].
- [78] R. de Belsunce et al., *The 3D Lyman- α forest power spectrum from eBOSS DR16*, *Mon. Not. Roy. Astron. Soc.* **533** (2024) 3756 [[arXiv:2403.08241](#)] [[INSPIRE](#)].
- [79] N.G. Karaçaylı and C.M. Hirata, *Light in the dark forest. Part I. An efficient optimal estimator for 3D Lyman-alpha forest power spectrum*, *JCAP* **07** (2025) 085 [[arXiv:2503.15619](#)] [[INSPIRE](#)].
- [80] R.A.C. Croft, *Ionizing Radiation Fluctuations and Large-Scale Structure in the Ly α Forest*, *Astrophys. J.* **610** (2004) 642 [[astro-ph/0310890](#)] [[INSPIRE](#)].

- [81] A. Pontzen, *Scale-dependent bias in the baryonic-acoustic-oscillation-scale intergalactic neutral hydrogen*, *Phys. Rev. D* **89** (2014) 083010 [[arXiv:1402.0506](#)] [[INSPIRE](#)].
- [82] S. Gontcho A Gontcho, J. Miralda-Escudé and N.G. Busca, *On the effect of the ionizing background on the Ly α forest autocorrelation function*, *Mon. Not. Roy. Astron. Soc.* **442** (2014) 187 [[arXiv:1404.7425](#)] [[INSPIRE](#)].
- [83] A. Pontzen, S. Bird, H. Peiris and L. Verde, *Constraints on Ionizing Photon Production from the Large-scale Ly α Forest*, *Astrophys. J. Lett.* **792** (2014) L34 [[arXiv:1407.6367](#)] [[INSPIRE](#)].
- [84] S. Furlanetto and S.P. Oh, *The History and Morphology of Helium Reionization*, *Astrophys. J.* **681** (2008) 1 [[arXiv:0711.1542](#)] [[INSPIRE](#)].
- [85] G.D. Becker, J.S. Bolton, M.G. Haehnelt and W.L.W. Sargent, *Detection of Extended He II Reionization in the Temperature Evolution of the Intergalactic Medium*, *Mon. Not. Roy. Astron. Soc.* **410** (2011) 1096 [[arXiv:1008.2622](#)] [[INSPIRE](#)].
- [86] G. Worseck, J.X. Prochaska, J.F. Hennawi and M. McQuinn, *Early and Extended Helium Reionization Over More Than 600 Million Years of Cosmic Time*, *Astrophys. J.* **825** (2016) 144 [[arXiv:1405.7405](#)] [[INSPIRE](#)].
- [87] W.J. Percival, O. Friedrich, E. Sellentin and A. Heavens, *Matching Bayesian and frequentist coverage probabilities when using an approximate data covariance matrix*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 3207 [[arXiv:2108.10402](#)] [[INSPIRE](#)].
- [88] S. Dodelson and M.D. Schneider, *The Effect of Covariance Estimator Error on Cosmological Parameter Constraints*, *Phys. Rev. D* **88** (2013) 063537 [[arXiv:1304.2593](#)] [[INSPIRE](#)].
- [89] E. Sellentin and A.F. Heavens, *Parameter inference with estimated covariance matrices*, *Mon. Not. Roy. Astron. Soc.* **456** (2016) L132 [[arXiv:1511.05969](#)] [[INSPIRE](#)].

Author List

Wynne Turner^{^{a,b,*}}, Andrei Cuceu^{^{c,1}}, Paul Martini^{^{a,b,d}}, J. Aguilar^c, S. Ahlen^{^e}, A. Anand^{^c},
D. Bianchi^{^{f,g}}, D. Brooks^h, L. Casasⁱ, T. Claybaugh^c, A. de la Macorra^{^j}, B. Dey^{^{k,l}}, P. Doel^h,
S. Ferraro^{^{c,m}}, A. Font-Ribera^{ⁱ}, J.E. Forero-Romero^{^{n,o}}, E. Gaztañaga^{^{p,q,r}},
S. Gontcho A Gontcho^{^{c,s}}, G. Gutierrez^t, H.K. Herrera-Alcantar^{^{u,v}}, K. Honscheid^{^{a,d}},
M. Ishak^{^x}, R. Joyce^{^y}, R. Kehoe^z, D. Kirkby^{^{aa}}, A. Kremin^{^c}, O. Lahav^h, M. Landriau^{^c},
L. Le Guillou^{^{ab}}, M. Manera^{^{ac,i}}, R. Miquel^{^{ad,i}}, A. Muñoz-Gutiérrez^j, S. Nadathur^{^q},
G. Niz^{^{ae,af}}, N. Palanque-Delabrouille^{^{v,c}}, W.J. Percival^{^{ag,ah,ai}}, C. Poppett^{^{c,aj,m}}, F. Prada^{^{ak}},
A.J. Ross^{^{a,b}}, G. Rossi^{al}, E. Sanchez^{^{am}}, D. Schlegel^c, M. Schubnell^{^{an,ao}}, H. Seo^{^{ap}}, J. Silber^{^c},
D. Sprayberry^y, G. Tarlé^{^{ao}}, M. Walther^{^{aq,ar}}, B.A. Weaver^y, R. Zhou^{^c}, H. Zou^{^{as}}

^a Center for Cosmology and AstroParticle Physics, The Ohio State University,
191 West Woodruff Avenue, Columbus, OH 43210, U.S.A.

^b Department of Astronomy, The Ohio State University,
4055 McPherson Laboratory, 140 W 18th Avenue, Columbus, OH 43210, U.S.A.

^c Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, U.S.A.

^d Department of Physics, The Ohio State University,
191 West Woodruff Avenue, Columbus, OH 43210, U.S.A.

^e Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, MA 02215, U.S.A.

^f Dipartimento di Fisica “Aldo Pontremoli”, Università degli Studi di Milano,
Via Celoria 16, I-20133 Milano, Italy

^g INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy

^h Department of Physics & Astronomy, University College London,
Gower Street, London, WC1E 6BT, U.K.

ⁱ Institut de Física d’Altes Energies (IFAE), The Barcelona Institute of Science and Technology,
Edifici Cn, Campus UAB, 08193, Bellaterra (Barcelona), Spain

^j Instituto de Física, Universidad Nacional Autónoma de México,
Circuito de la Investigación Científica, Ciudad Universitaria, Cd. de México C. P. 04510, México

^k Department of Astronomy & Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada

^l 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.

^m University of California, Berkeley, 110 Sproul Hall #5800 Berkeley, CA 94720, U.S.A.

ⁿ Departamento de Física, Universidad de los Andes,
Cra. 1 No. 18A-10, Edificio Ip, CP 111711, Bogotá, Colombia

^o Observatorio Astronómico, Universidad de los Andes,
Cra. 1 No. 18A-10, Edificio H, CP 111711 Bogotá, Colombia

^p Institut d’Estudis Espacials de Catalunya (IEEC),
c/ Esteve Terradas 1, Edifici RDIT, Campus PMT-UPC, 08860 Castelldefels, Spain

^q Institute of Cosmology and Gravitation, University of Portsmouth,
Dennis Sciama Building, Portsmouth, PO1 3FX, U.K.

^r Institute of Space Sciences, ICE-CSIC, Campus UAB,
Carrer de Can Magrans s/n, 08913 Bellaterra, Barcelona, Spain

^s University of Virginia, Department of Astronomy, Charlottesville, VA 22904, U.S.A.

^t Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, U.S.A.

^u Institut d’Astrophysique de Paris. 98 bis boulevard Arago. 75014 Paris, France

^v IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

^x Department of Physics, The University of Texas at Dallas,
800 W. Campbell Rd., Richardson, TX 75080, U.S.A.

^y NSF NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, U.S.A.

^z Department of Physics, Southern Methodist University, 3215 Daniel Avenue, Dallas, TX 75275, U.S.A.

^{aa} Department of Physics and Astronomy, University of California, Irvine, 92697, U.S.A.

- ^{ab} *Sorbonne Université, CNRS/IN2P3, Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), FR-75005 Paris, France*
- ^{ac} *Departament de Física, Serra Húnter, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*
- ^{ad} *Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys, 23, 08010 Barcelona, Spain*
- ^{ae} *Departamento de Física, DCI-Campus León, Universidad de Guanajuato, Loma del Bosque 103, León, Guanajuato C. P. 37150, México*
- ^{af} *Instituto Avanzado de Cosmología A. C., San Marcos 11 - Atenas 202, Magdalena Contreras, Ciudad de México C. P. 10720, México*
- ^{ag} *Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
- ^{ah} *Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada*
- ^{ai} *Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
- ^{aj} *Space Sciences Laboratory, University of California, Berkeley, 7 Gauss Way, Berkeley, CA 94720, U.S.A.*
- ^{ak} *Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, s/n, E-18008 Granada, Spain*
- ^{al} *Department of Physics and Astronomy, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea*
- ^{am} *CIEMAT, Avenida Complutense 40, E-28040 Madrid, Spain*
- ^{an} *Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, MI 48109, U.S.A.*
- ^{ao} *University of Michigan, 500 S. State Street, Ann Arbor, MI 48109, U.S.A.*
- ^{ap} *Department of Physics & Astronomy, Ohio University, 139 University Terrace, Athens, OH 45701, U.S.A.*
- ^{aq} *Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany*
- ^{ar} *University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81677 München, Germany*
- ^{as} *National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Road, Chaoyang District, Beijing, 100101, P.R. China*

* Corresponding author

¹ NASA Einstein Fellow