Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **532,** 3669–3681 (2024)
Advance Access publication 2024 July 4

https://doi.org/10.1093/mnras/stae1610

Broad absorption line quasars in the Dark Energy Spectroscopic Instrument Early Data Release

S. Filbert [1,2,3,4★] P. Martini,[1,2,3] K. Seebaluck,[1,2,3] L. Ennesser [2,3] D. M. Alexander,[5,6] A. Bault,[7] A. Brodzeller [4] H. K. Herrera-Alcantar,[8] P. Montero-Camacho,[9] I. Pérez-Ràfols [10] C. Ramírez-Pérez [11] C. Ravoux [12,13] T. Tan,[14] J. Aguilar,[15] S. Ahlen,[16] S. Bailey,[15] D. Brooks,[17] T. Claybaugh,[15] K. Dawson,[4] A. de la Macorra,[18] P. Doel,[17] K. Fanning,[2,3] A. Font-Ribera [11] J. E. Forero-Romero [19,20] S. Gontcho A Gontcho [15] J. Guy,[15] D. Kirkby,[7] A. Kremin,[15] C. Magneville,[13] M. Manera,[11,21] A. Meisner [22] R. Miquel,[11,23] J. Moustakas,[24] J. Nie,[25] W. J. Percival [26,27,28] F. Prada,[29] M. Rezaie [30] G. Rossi,[31] E. Sanchez [32] M. Schubnell,[33,34] H. Seo [35] G. Tarlé,[34] B. A. Weaver[22] and Z. Zhou[25]

Affiliations are listed at the end of the paper

Accepted 2024 June 3. Received 2024 May 8; in original form 2023 September 7

ABSTRACT

Broad absorption line (BAL) quasars are characterized by gas clouds that absorb flux at the wavelength of common quasar spectral features, although blueshifted by velocities that can exceed $0.1c$. BAL features are interesting as signatures of significant feedback, yet they can also compromise cosmological studies with quasars by distorting the shape of the most prominent quasar emission lines, impacting redshift accuracy and measurements of the matter density distribution traced by the Lyman α forest. We present a catalogue of BAL quasars discovered in the Dark Energy Spectroscopic Instrument (DESI) survey Early Data Release, which were observed as part of DESI Survey Validation, as well as the first two months of the main survey. We describe our method to automatically identify BAL quasars in DESI data, the quantities we measure for each BAL, and investigate the completeness and purity of this method with mock DESI observations. We mask the wavelengths of the BAL features and re-evaluate each BAL quasar redshift, finding new redshifts which are 243 km s^{-1} smaller on average for the BAL quasar sample. These new, more accurate redshifts are important to obtain the best measurements of quasar clustering, especially at small scales. Finally, we present some spectra of rarer classes of BALs that illustrate the potential of DESI data to identify such populations for further study.

Key words: galaxies: active – galaxies: nuclei – galaxies: quasars: absorption lines – galaxies: quasars: emission lines – catalogues.

1 INTRODUCTION

Quasars have been a valuable cosmological tool since the realization by Schmidt (1963) that they were outside the local Universe, and therefore had the potential to probe early into cosmic history. In the intervening decades, quasars have become a standard tool for cosmology at redshifts above $z \sim 1.5$. At these redshifts, the space density of luminous quasars is much higher than in the local Universe (e.g. Hopkins, Richards & Hernquist 2007). Their higher space density combined with their high luminosity make quasars a valuable probe of large-scale structure.

The Sloan Digital Sky Survey (SDSS; York et al. 2000) and especially its extensions to SDSS-III (Eisenstein et al. 2011) and SDSS-IV (Blanton et al. 2017) have targeted large numbers of quasars to study their physical properties, their evolution, and use them as probes of large-scale structure. The SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) and its SDSS-IV successor eBOSS (Dawson et al. 2016) measured the baryon acoustic oscillation feature with quasars as point tracers of the matter-density field (Neveux et al. 2020; Hou et al. 2021), through studies of the matter distribution in the Lyman α (Ly α) forest (Busca et al. 2013; Bautista et al. 2017; de Sainte Agathe et al. 2019; du Mas des Bourboux et al. 2020), and by cross-correlating quasars with Ly α forest absorption (Font-Ribera et al. 2014; Blomqvist et al. 2019).

A complication to these studies is that 10–30 per cent of quasars exhibit BAL troughs (Foltz et al. 1990; Trump et al. 2006). The BAL class of quasars was first defined by Weymann et al. (1991) as quasars with troughs blueshifted by at least 3000 km s^{-1} and absorption velocity widths of at least 2000 km s^{-1}. BAL troughs are most commonly observed associated with the prominent C IV emission line, although are also observed associated with many

★ E-mail: u1477296@umail.utah.edu

other high-ionization features, such as Si IV and N V. Quasars with absorption associated with high-ionization features are dubbed HiBALs. Quasars more rarely exhibit absorption associated with lower ionization features such as Mg II and Fe II, in which case they are called LoBALs, and FeLoBALs when absorption from excited states of Fe II and Fe III are present (Hall et al. 2002; Trump et al. 2006).

BAL features are a complication for cosmological studies for two reasons. The first is that BAL features can significantly distort the shape of the main emission features that are used to determine the redshift of the quasar, such as the C IV emission line. This can confuse and bias algorithms that automatically determine quasar redshifts for large surveys. Specifically, BAL troughs can to lead to an overestimate of the quasar redshift because the absorption can depress the blue wing of one or more spectral features. Secondly, BAL features may also be associated with many of the emission features in the Ly α forest region used for cosmological studies. Highly blueshifted BAL features associated with prominent lines like Ly α and N V, as well as weaker lines such as P V and S IV (Mas-Ribas & Mauland 2019), can produce absorption that is often indistinguishable from the neutral hydrogen absorption that comprises the Ly α forest.

These complications have made it critically important to identify BALs for large-scale structure studies using quasars. With large surveys such as SDSS and now DESI, the use of automated methods to identify and classify BAL quasars is more important than ever. Reichard et al. (2003) were the first to use an automated method to identify BAL quasars in the SDSS early data, presenting 224 visually confirmed BAL quasars. The SDSS DR12 quasar catalogue contains 297 301 quasars, 29 580 of which exhibited BAL features (Pâris et al. 2017). This is 13 per cent of all quasars above $z = 1.57$, as it is only above this redshift that BAL features could be unambiguously identified on the blue side of the C IV emission line. Given the size of these quasar samples, subsequent quasar samples from SDSS DR14 (Pâris et al. 2018) and DR16 (Lyke et al. 2020) used a variety of automated algorithms, including machine-learning methods (Guo & Martini 2019).

The Dark Energy Spectroscopic Instrument (DESI) survey plans to measure about 2.8 million quasars, including over 800 000 at $z > 2.1$ that will probe the Ly α forest (Chaussidon et al. 2023). This further increases the importance of developing automated methods to identify BALs, quantify their impact on quasar redshift errors, and further study how to optimally include BAL quasars in cosmological analyses. This is important because BAL quasars had historically been excluded from Ly α studies (e.g. Bautista et al. 2017; Bourboux et al. 2020), which removes 10–15 per cent of the sample. Ennesser et al. (2022) recently studied the impact of BALs on Ly α forest studies in detail and developed criteria to maintain the vast majority of BAL quasars in Ly α cosmology measurements. They showed that after masking the expected locations of BAL features in the Ly α forest continuum region, they could obtain cosmological results that were consistent with non-BAL quasars. Including BALs after masking absorption features produced a pronounced improvement in the fractional errors in the correlation function both by increasing the total number of sightlines with Ly α measurements and because the BAL quasar spectra had on average higher signal-to-noise ratio (SNR).

In this paper, we describe the algorithm that we plan to use to identify BAL quasars in the DESI survey, as well as present a catalogue of the BAL quasars from Early Data Release (EDR) observations and the first two months of main survey operations (M2). Section 2 has a description of the DESI observations, how we identify BALs, and how we add BALs to mock DESI data sets. In Section 3, we investigate the completeness and purity of the BAL identification algorithm with DESI mock data. We then present the DESI EDR + M2 BAL catalogue in Section 4. This section includes a description of the catalogue data model as well as statistical properties of the BAL quasars. The catalogue contains information about individual BALs, including the locations of all of the absorption troughs associated with C IV, as the velocity offsets of these troughs are critical to masking the features that impact both redshifts and the forest. We then study the impact of BAL features on quasar redshift measurements in Section 5. We quantify the difference in redshifts with and without masking the BAL features, and present new redshifts for the BAL quasars based on our masking approach. In Section 6, we mention some unusual BALs and we conclude in Section 7 with a summary of our main results.

2 DATA

The BALs in this paper were discovered in observations during DESI Survey Validation (SV), which are described in DESI Collaboration (2023a), and the first two months of the main survey (M2). SV extended from 2020 December through 2021 mid-May and M2 from 2021 mid-May to mid-July. The purposes of SV were to test the data quality, optimize target selection, and exercise the operational and analysis programs. The two specific periods of SV included here are Target Selection Validation (also referred to as SV1) and the One-Percent Survey (also referred to as SV3). The SV observations are publicly available as part of the DESI EDR, which is described in DESI Collaboration (2023b). In this section, we briefly present some background information about DESI observations, describe our procedure to identify BALs in these data, and finally describe the mock data sets we use to verify our approach. We constructed separate BAL catalogues for SV1, SV3, and M2. Unless otherwise specified, we perform our analysis with the combination of all three of these catalogues.

2.1 DESI observations

The goal of DESI is to study cosmic acceleration with a spectroscopic survey of 40 million galaxies and quasars in just 5 yr (DESI Collaboration 2016a). The DESI Collaboration plans to use these data to measure distances from the nearby Universe to beyond $z > 3.5$ with the baryon acoustic oscillation method, as well as employ redshift-space distortions to measure the growth of cosmic structures and test potential modifications to general relativity. The DESI Collaboration is conducting this survey with a highly multiplexed fibre-fed spectrograph at the Mayall 4 m telescope of the Kitt Peak National Observatory (DESI Collaboration 2022). This instrument has a 3 deg diameter field of view that was designed to observe 5000 targets in a single observation. The light from each target is fed into one of 10 bench-mounted spectrographs that are located in a climate-controlled room. Each of these spectrographs records light from 360 to 980 nm split among three distinct wavelength channels. The blue channel is especially important for the quasars at $z > 2.1$ that are used to study the Ly α forest. This channel extends from 360 to 593 nm and ranges in resolution from about $2000 - 3500$. The neighbouring red channel covers $560 - 772$ nm with a resolution of about $3500 - 4500$ and the near-infrared channel covers $747 - 980$ nm with a resolution of about $4000 - 5000$. More details of the instrumentation are described in DESI Collaboration (2016b) for the technical design, Silber et al. (2023) for the focal

plane system, and Miller et al. (2023) for the optical corrector and support system.

DESI target selection (Myers et al. 2023) includes both quasars at $z > 0.9$ that are used as direct tracers of the dark matter distribution and $z > 2.1$ quasars that are used to trace the matter distribution in the Ly α forest (Chaussidon et al. 2023). This selection is based on significant imaging from the DESI Legacy Imaging Surveys (Zou et al. 2017; Dey et al. 2019). DESI observes 310 quasar targets per deg^2 and successfully identifies more than 200 quasars per deg^2, including over 60 per deg^2 at $z > 2.1$ (Chaussidon et al. 2023). As they are lower surface density than other targets, the quasars are assigned higher priority for fiber assignment during dark time than the other dark time targets (Luminous Red Galaxies and Emission Line Galaxies). Some fibres are devoted to standard stars for flux calibration and to empty regions to measure and subtract the spectrum of the night sky. DESI is able to achieve superb sky subtraction and flux calibration due to a combination of excellent stability and sophisticated data processing algorithms. One-dimensional spectral extractions, redshifts, and spectral classifications are typically available the morning after the observations. Guy et al. (2023) describe the DESI data processing pipeline in detail and Schlafly et al. (2023) describe survey operations. The software for target selection, the spectroscopic pipeline, and survey operations is publicly available.[1]

The DESI collaboration constructs quasar catalogues from a combination of three classification routines: Redrock[2] (Bailey et al., in preparation), QuasarNET (Busca & Balland 2018; Farr, Font-Ribera & Pontzen 2020b), and the Mg II afterburner. Redrock performs a χ^2 analysis as a function of redshift for a range of spectral templates developed by Brodzeller et al. (2023) to identify the best redshift and spectral classification and correctly classifies the vast majority of quasars (Alexander et al. 2023). Both QuasarNET and the Mg II afterburner are used to check quasar targets that were not classified as quasars by Redrock. QuasarNET is a machine learning algorithm that employs convolutional neural networks for both classification and redshift estimation. The Mg II afterburner searches for broad Mg II emission at the best-fitting Redrock redshift and reclassifies the object as a quasar if Mg II emission is significant, regardless of the spectral type initially provided by Redrock (Chaussidon et al. 2022). Chaussidon et al. (2023) further describe the quasar target selection procedure and Alexander et al. (2023) describe the visual inspection process that we employed to quantify the performance of the redshift and classification algorithms. The quasar catalogues for SV1, SV3, and M2 include all objects classified as quasars by these algorithms.

2.2 BAL identification

There are several methods used to quantify the strength of absorption troughs in the spectra of BAL quasars. The first, and in some sense the strictest definition, is the Balnicity Index (BI) proposed by Weymann et al. (1991). This definition is especially useful for characterizing quasars that have strong, broad absorption from one or more continuous high column density clouds with a large velocity ($> 3000\,\mathrm{km\,s^{-1}}$) relative to the quasar. The equation for BI is

$$BI = -\int_{25000}^{3000}\left[1-\frac{f(v)}{0.9}\right]C(v)dv\,. \tag{1}$$

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

The term $f(v)$ is the normalized flux density of the quasar, which is the observed quasar spectral energy distribution divided by a model fit to the quasar that does not have BAL features. The integration variable v is the velocity displacement blueward of the line centroid, which is usually C IV. In some cases, BI is also calculated relative to Si IV, Mg II, and other lines. C is a constant that is zero unless the term $[1-\frac{f(v)}{0.9}]$ is greater than zero for more than $2000\,\mathrm{km\,s^{-1}}$, in which case it is set to one. The result is that a trough will have a BI value greater than zero only after the quasar flux is more than 10 per cent below the estimated continuum flux for a contiguous span of at least $2000\,\mathrm{km\,s^{-1}}$.

We estimate the error in BI for each BAL quasar classified using the method introduced by Trump et al. (2006), although modified by Guo & Martini (2019) to include the uncertainty in the continuum fitting from principal component analysis (PCA). The equation used for the error in BI is

$$\sigma_{BI}^2 = -\int_{25000}^{3000}\left(\frac{\sigma_{f(v)}^2+\sigma_{PCA}^2}{(0.9)^2}\right)C(v)dv\,, \tag{2}$$

where $\sigma_{f(v)}^2$ is the variance in the normalized flux density and σ_{PCA}^2 is the variance of the PCA fit.

Although BI is a robust estimator of absorption, it misses many BAL features because it starts $3000\,\mathrm{km\,s^{-1}}$ blueward of the line centre and is insensitive to features that are narrower than $2000\,\mathrm{km\,s^{-1}}$. The absorption index (AI), introduced by Hall et al. (2002), is an alternate estimator of the absorption strength. Unlike BI, the presence of absorption is evaluated up to the line centroid. Additionally, $C(v)$ is set to one after the trough extends more than $450\,\mathrm{km\,s^{-1}}$, rather than $2000\,\mathrm{km\,s^{-1}}$ for BI. The equations for AI and the corresponding uncertainty are

$$AI = -\int_{25000}^{0}\left[1-\frac{f(v)}{0.9}\right]C(v)dv \tag{3}$$

and

$$\sigma_{AI}^2 = -\int_{25000}^{0}\left(\frac{\sigma_{f(v)}^2+\sigma_{PCA}^2}{(0.9)^2}\right)C(v)dv\,. \tag{4}$$

The AI criterion captures about $4-5$ times more BALs than the BI criterion, and most importantly it captures BAL features that could compromise precise cosmological measurements. We therefore consider any object with $AI > 0$ to be a BAL quasar, although we also calculate BI for comparison with previous work. Furthermore, we only consider absorption from C IV to classify a quasar as a BAL quasar. We consequently only search for BALs in quasars where the observed spectra extend to $25\,000\,\mathrm{km\,s^{-1}}$ blueward of the C IV line. We also require that the spectra extend redward of C IV to at least 1633 Å in order to accurately fit the quasar spectra with PCA components, which effectively sets an upper limit in redshift. Because of these two constraints, we only look for BALs in quasar spectra between $1.57 < z < 5.0$. This differs slightly from the implementation used by Guo & Martini (2019) which searched for absorption in the redshift range between 1.57 and 5.56. As we only conduct our search around the high ionization C IV line, all of the BALs appear to be high-ionization BALs or HiBALs; however, as we only search in this wavelength range we do not classify the BALs into categories such as HiBALs, LoBALs (low ionization BALs), and FeLoBALs (LoBALs with Fe emission; Hall et al. 2002; Trump et al. 2006; Gibson et al. 2009). We briefly discuss the prospects for identifying other types of BALs in Section 6.

We calculate AI and BI, and other BAL-specific quantities (as in Table 1) with software in the `baltools`[3] repository. We refer

Table 1. Information about all of the BAL measurements included in the BAL catalogue. More information about the columns is in the DESI EDR documentation for this Value Added Catalog and Section 4.1.

DESI BAL catalogue columns			
Column	Name	Data type	Description
65	PCA_COEFFS	FLOAT[5]	Coefficients on PCA eigenbasis for spectrum fit
66	PCA_CHI2	FLOAT	Goodness of fit from PCA components
67	BAL_PROB	FLOAT	Likelihood of BAL (not populated in this catalogue)
68	BI_CIV	FLOAT	Equivalent width of C IV absorption by Balnicity index definition [km s^{-1}]
69	ERR_BI_CIV	FLOAT	Uncertainty in BI from C IV absorption [km s^{-1}]
70	NCIV_2000	INT32	Number of absorption troughs from C IV extending $>$ 2000 km s^{-1}
71	VMIN_CIV_2000	FLOAT[5]	Minimum velocity of BI absorption troughs blueward of C IV [km s^{-1}]
72	VMAX_CIV_2000	FLOAT[5]	Maximum velocity of BI absorption troughs blueward of C IV [km s^{-1}]
73	POSMIN_CIV_2000	FLOAT[5]	Velocity displacement from C IV of normalized flux minimum for each AI absorption trough [km s^{-1}]
74	FMIN_CIV_2000	FLOAT[5]	Value of normalized flux minimum for each BI absorption trough from C IV
75	AI_CIV	FLOAT	Equivalent width of C IV absorption by absorption index definition [km s^{-1}]
76	ERR_AI_CIV	FLOAT	Uncertainty in AI from C IV absorption [km s^{-1}]
77	NCIV_450	INT32	Number of absorption troughs from C IV extending $>$ 450 km s^{-1}
78	VMIN_CIV_450	FLOAT[17]	Minimum velocity of AI absorption troughs blueward of C IV [km s^{-1}]
79	VMAX_CIV_450	FLOAT[17]	Maximum velocity of AI absorption troughs blueward of C IV [km s^{-1}]
80	POSMIN_CIV_450	FLOAT[17]	Velocity displacement from C IV of normalized flux minimum for each AI absorption trough [km s^{-1}]
81	FMIN_CIV_450	FLOAT[17]	Value of normalized flux minimum for each AI absorption trough from C IV
82	BI_SIIV	FLOAT	Equivalent width of SI IV absorption by Balnicity index definition [km s^{-1}]
83	ERR_BI_SIIV	FLOAT	Uncertainty in BI from Si IV absorption [km s^{-1}]
84	NSIIV_2000	INT32	Number of absorption troughs from Si IV extending $>$ 2000 km s^{-1}
85	VMIN_SIIV_2000	FLOAT[5]	Minimum velocity of BI absorption troughs blueward of Si IV [km s^{-1}]
86	VMAX_SIIV_2000	FLOAT[5]	Maximum velocity of BI absorption troughs blueward of Si IV [km s^{-1}]
87	POSMIN_SIIV_2000	FLOAT[5]	Velocity displacement from Si IV of normalized flux minimum for each BI absorption trough [km s^{-1}]
88	FMIN_SIIV_2000	FLOAT[5]	Value of normalized flux minimum for each BI absorption trough from Si IV
89	AI_SIIV	FLOAT	Equivalent width of Si IV absorption by absorption index definition [km s^{-1}]
90	ERR_AI_SIIV	FLOAT	Uncertainty in AI from Si IV absorption [km s^{-1}]
91	NSIIV_450	INT32	Number of absorption troughs from Si IV extending $>$ 450 km s^{-1}
92	VMIN_SIIV_450	FLOAT[17]	Minimum velocity of AI absorption troughs blueward of Si IV [km s^{-1}]
93	VMAX_SIIV_450	FLOAT[17]	Maximum velocity of AI absorption troughs blueward of Si IV [km s^{-1}]
94	POSMIN_SIIV_450	FLOAT[17]	Velocity displacement from Si IV of normalized flux minimum for each AI absorption trough [km s^{-1}]
95	FMIN_SIIV_450	FLOAT[17]	Value of normalized flux minimum for each AI absorption trough from Si IV
96	BALMASK	BYTE	Bitmask specifying whether BAL information was evaluated (see Section 4.1)
103	SNR_CIV	DOUBLE	SNR in region extending from the C IV emission feature to 25 000 km s^{-1} bluewards

to the code that identifies BALs and measures their properties as the BALFINDER. The BALFINDER code was originally developed for SDSS and used a convolution neural network (CNN), as described in Guo & Martini (2019). The implementation of the BALFINDER for DESI does not use a CNN, and instead relies on PCA to fit spectra for a given quasar from the five eigenspectra in Guo & Martini (2019). This PCA fit to each spectrum is iterative, in that any regions of the spectrum with BAL features according to equation (3) are masked before the next iteration. This continues for 10 iterations of fitting and masking. Fig. 1 shows an example PCA fit for a DESI quasar.

The best-fitting PCA representation is used to estimate the quasar continuum if no absorption were present, and is used to compute the normalized flux density $f(v)$ in the equations for AI and BI. Each quasar may have multiple troughs that satisfy the AI and/or BI criteria, and we refer to the number of such troughs as N_{450} and N_{2000}, respectively. The code also records the minimum and maximum velocity for each trough. The sum of the absorption troughs that meet the AI and BI criteria are the AI and BI values for the quasar. All of these data are stored for each BAL quasar, as well as other properties such as the coefficients for each of the eigenspectra used in the PCA fits, the errors σ_{AI} and σ_{BI}, and the χ^2 of the PCA fit. The BAL properties included in the catalogue are summarized in Table 1.

In some rare cases the PCA fitting leads to misidentifications. Based on our visual inspection of hundreds of BAL quasars, this occurs because the PCA components are unable to fit the full range of quasar diversity. Approximately 15 per cent of objects identified as BALs by the BALFINDER which have bad χ^2_{PCA} fits were not initially classified as quasars by Redrock, suggesting that these objects already have strange spectra, and were re-classified as quasars by the afterburners discussed at the end of Section 2.1. In any case, should the initial PCA fit be relatively poor, the BALFINDER may incorrectly identify and mask non-BAL regions, and this masking may compound the misclassification. Based on visual inspection, fits with $\chi^2_{PCA} > 10$ are suspect. These high χ^2_{PCA} values are uncommon, comprising only about 2 per cent of the entire BAL catalogue. The χ^2_{PCA} quantity is in the BAL catalogue as 'PCA_CHI2'.

The presence of high χ^2_{PCA} fits hint at a failure of the templates discussed in Section 2.3 to reproduce quasar spectral diversity. This is one limitation of using mock spectra created using these PCA templates to estimate completeness and purity of the BAL quasar sample as it is unclear whether these templates span the diversity of BAL quasars observed by DESI. Section 2.3 further details the creation of the mock spectra used to estimate the purity and completeness of the BAL quasar sample.

[3]https://github.com/paulmartini/baltools

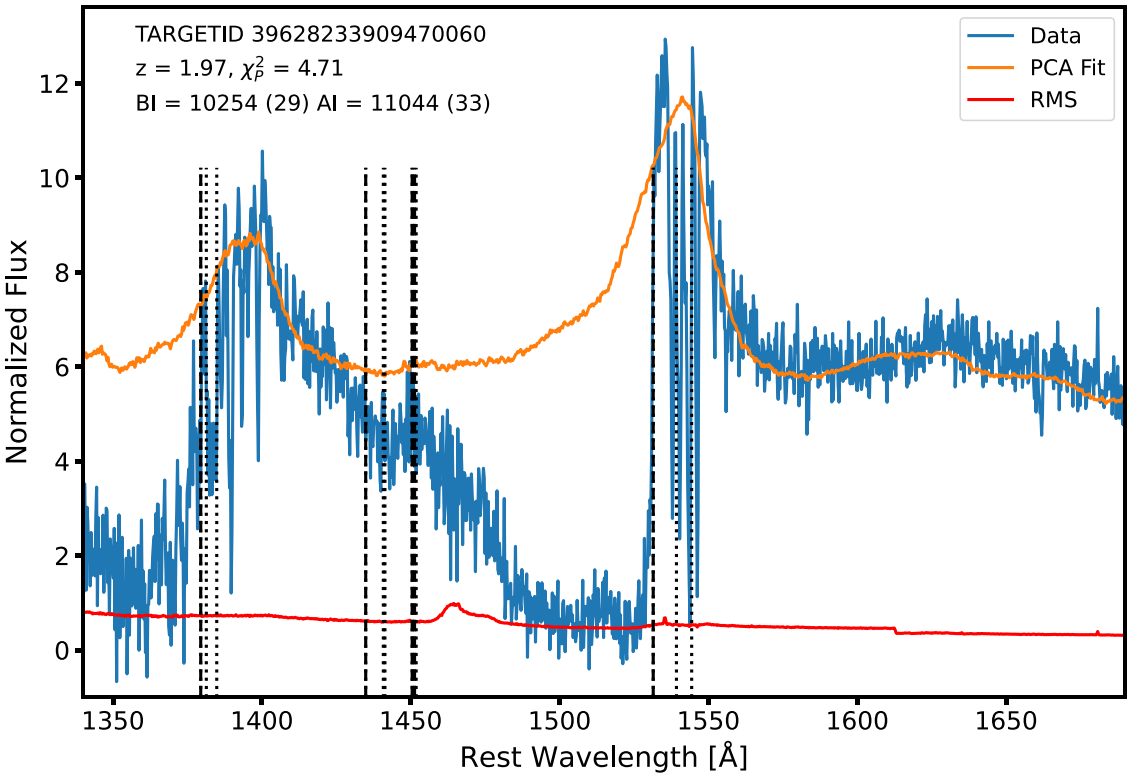

Figure 1. Spectrum of a BAL quasar from early DESI data. The χ^2_P represents the goodness of fit of the PCA fit to the data with rms flux uncertainty. The dotted lines delimit multiple absorption troughs meeting the AI criterion, whereas the thicker dashed lines correspond to BI absorption. The values of AI and BI absorption are shown in the legend, along with their corresponding errors. These absorption definitions and their errors are in equations (1)–(4).

2.3 Mock spectra

We test our BAL identification method using a set of simulated or mock quasar spectra. These mock spectra are generated with QUICKQUASARS, which is part of the DESI simulation package DESISIM[4] (see also Herrera-Alcantar et al. 2023). The QUICKQUASARS code was designed to produce realistic DESI spectra in the DESI footprint with the same magnitude and redshift distribution as DESI quasars, as well as add realistic instrument noise and astrophysical systematics, such as BAL features. For quasars at sufficiently high redshift, the mock spectra also include Ly α forest absorption, damped Ly α absorbers, other high column density systems, and metal lines. The Ly α forest absorption is added with model transmission data generated with the LYACOLORE code described by Farr et al. (2020a).[5]

BALs are added into QUICKQUASARS with a series of empirical BAL templates developed by Niu (2020). The templates were developed from a subset of about 1500 BAL quasars from SDSS DR14 catalogues during tests of a convolutional neural network BAL classifier by Guo & Martini (2019). Each template consists of the normalized absorption features as a function of velocity for a C IV BAL detected in high SNR data after removal of the quasar continuum variations. The complete set of templates were selected to have the same AI and BI distributions as the complete DR14 BAL catalog. Each mock quasar has some probability of being a BAL (the default is 16 per cent) and if a quasar is a BAL, one of the templates is randomly assigned to that quasar. The BAL absorption features are added before the addition of noise or any Ly α and other absorption. The BAL absorption is applied to other lines with the same absorption profile as a function of velocity as the C IV emission line. For each quasar that is assigned BAL absorption, QUICKQUASARS outputs a table of the BAL properties including the BAL template ID, the BAL redshift, and AI and BI values. We describe how we used these data to test the performance of the BALFINDER in Section 3.

[4] https://github.com/desihub/desisim
[5] https://github.com/igmhub/LyaCoLoRe

3 COMPLETENESS AND PURITY

The completeness and purity of BAL catalogues are important to evaluate the impact of BALs on cosmological analysis. It is difficult to reliably quantify the completeness and purity of BAL identification algorithms on real BALs due to the challenge of constructing a useful truth catalogue. Two previous studies by Busca & Balland (2018) and Guo & Martini (2019) relied on human-classified BALs for truth in order to test the performance of their classifiers and estimate what they referred to as the pseudo-completeness and pseudo-purity. The challenge of this approach is the reliability of human classification, which may have biases towards specific classes of BALs, redshifts, and/or SNR. In this paper, we instead quantify the performance with the DESI mocks described in Section 2.3 that include simulated BAL spectra in 16 per cent of the quasars. We compare the BALFINDER output with the truth catalogues to determine the completeness and purity based on the $AI > 0$ criterion.

Table 2 contains estimates of the average completeness and purity of the $AI > 0$ BAL quasar sample. The average completeness and purity for SV1, meant to be representative of the full DESI data, is 56 per cent and 96 per cent respectively, with completeness being slightly lower for the lower average SNR catalogues (SV3 and M2). Although relatively low completeness is not detrimental to cosmo-

Table 2. Statistical properties and numerical characteristics of each of the BAL catalogues. AI > 0 and BI > 0 refer to the number of objects identified with at least one absorption trough for each definition.

	DESI BAL catalogue properties								
Catalog	Total QSOs	QSOs in z range	AI > 0	AI %	BI > 0	BI %	avg(SNR_CIV)	Completeness	Purity
DESI SV1	37 903	21 486	4282	19.93	1001	4.66	4.41	0.56	0.96
DESI SV3	50 993	28 641	5439	18.99	1338	4.67	3.99	0.51	0.97
DESI-M2	280 485	159 772	20 264	12.68	5058	3.17	2.44	0.42	0.99
All	369 381	209 899	29 985	14.29	7397	3.52	2.85	0.45	0.99

logical studies of large-scale structure with quasars, we recognize that applications such as quasar population studies, stacking, and variability require high completeness and purity. Those who wish to use this catalogue for such applications may find the $BI > 0$ cut to be a more pure and complete sample, due to the more extreme absorption regions. This work did not estimate the $BI > 0$ completeness and purity using our mock spectra. Furthermore, masking objects with bad χ^2_{PCA} fits (i.e. >10) could also serve to further improve the purity of the sample.

3.1 Purity

The purity is defined as the number of true BAL quasars recovered by the BALFINDER divided by the total number of BAL quasars identified by the BALFINDER. We use the AI criterion throughout, so that if the BALFINDER identifies any $AI > 0$ troughs, we consider this a BAL quasar. On average, we find that the purity of the BALFINDER is a little under 50 per cent for mock quasars. For comparison, Busca et al. (2013) measured a pseudo-purity of 77 per cent and Guo & Martini (2019) measured a pseudo-purity of 40 per cent. The significant difference relative to Busca et al. (2013) is because we use $AI > 0$ to define BALs, while they use a sample based on visual inspection that is biased towards BALs with stronger features. Our results are more similar to Guo & Martini (2019) because they similarly used the $AI > 0$ threshold to identify BALs.

The 50 per cent pseudo-purity based on the mock spectra is not indicative of the purity we estimate for the BAL catalogue presented in this work. This is because purity (and completeness) depend heavily on SNR. By binning the SNR of simulated BAL quasar spectra and evaluating the purity of each bin, we fit a simple model, similar to that shown in Fig. 2. We find that the purity is very nearly 100 per cent for SNR < 11, but decreases rapidly to 50 per cent for SNR > 14. Using an estimation for the SNR blueward of the C IV feature ('SNR_CIV' column in the catalogue, see Section 4.1), we can then apply the purity function for each quasar to estimate the expected purity of the sample. From this, we estimate that the BAL quasar sample as a whole has a purity of ∼ 99 per cent. This seems reasonable since the average 'SNR_CIV' of the entire BAL sample is ∼ 3.

The decrease in purity for SNR > 11 is caused by the BALFINDER algorithm's failure to fit PCA components to high SNR spectra. This introduces a larger false-positive rate and drives down the purity. However, false positives are not a significant issue for cosmological analysis, as their main impact is that a small fraction of good data are thrown out, rather than the introduction of a source of contamination. This is also why we do not include a criterion based on the significance of the AI measurement, as the number of additional, masked pixels is very small.

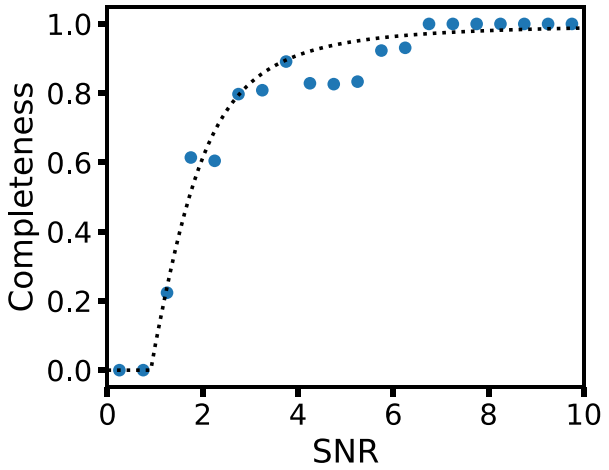

Figure 2. Completeness of the BALFINDER from LYACOLORE mocks as a function of SNR. The completeness is shown in bins of 0.5 up to SNR= 10. Relatively few spectra have higher SNR, and above this threshold the completeness asymptotes to 100 per cent. We use a simple, analytic approximation to the completeness versus SNR (*dotted line*) to determine the average completeness of quasar catalogues with different SNR distributions.

3.2 Completeness

The completeness is defined as the number of BAL quasars successfully recovered by the BALFINDER divided by the total number of true BAL quasars from the mocks. We again use the $AI > 0$ criterion and find the average completeness is about 68 per cent for the mock catalogue. This is less than the 98 per cent pseudo-completeness reported by Busca et al. (2013) and the 97.4 per cent measured by Guo & Martini (2019). We attribute our lower completeness to both our $AI > 0$ criterion and the lower average SNR of our spectra.

As with purity, we investigate the completeness as a function of SNR. The accompanying plot can be seen in Fig. 2. We use the same method to estimate the completeness of the BAL quasar as purity, finding an average completeness of ∼ 45 per cent. Unlike the purity, the completeness is a very strong function of SNR in the low-SNR regime (SNR < 1), while at higher SNR the completeness rapidly asymptotes to ∼ 95 per cent and is consequently comparable to the performance of previous algorithms. Unsurprisingly, it is harder for the algorithm to identify BAL quasars in noisy data, as noise spikes may interrupt a systematic depression of the quasar flux for the 450 km s^{-1} necessary to identify an $AI > 0$ BAL.

Unlike previous work, our truth sample is not biased to only consider BALs in high SNR data, as it does not rely on visual inspection and the BAL features were added independent of the noise. As DESI quasar spectra will generally have lower SNR than SDSS spectra, especially for the tracer quasars at $z < 2.1$, we expect lower completeness. On the other hand, such lower completeness are not a significant cause for concern for cosmological studies, as the

BAL features are less important relative to the noise properties of the data for those spectra.

4 DESI BAL CATALOGUE

We construct the DESI EDR + M2 BAL catalogue from quasar catalogues for SV1, SV3, and M2 as described in Section 2. While we conduct most of our analysis with the combined catalogue, we also analyse them individually as they have somewhat different noise properties. The key difference is that the SV1 data are deeper and therefore higher SNR than what DESI plans to achieve at the end of the main survey, SV3 is representative of the four observations of $z > 2.1$ quasars that DESI aims to achieve at the end of the main survey, and DESI-M2 mostly contains a single observation of each $z > 2.1$ quasar and therefore is lowest in SNR. In the first subsection below we present the key properties of the BAL data model that is provided in the BAL catalogue. The following subsection presents a summary of the BAL properties, and lastly we derive the BAL fractions for the SV1, SV3, and M2 data sets.

4.1 Data model

All of the DESI BAL catalogues have the same information as in a DESI quasar catalogue plus additional columns with the BAL properties. There are a total of 39 additional columns. The 33 determined by the BALFINDER are summarized in Table 1. We add the remaining six after we mask the BAL features and refit the redshifts. Those additional columns are described below in Section 5. Each of these properties is measured by the code provided in the public baltools repository. The remainder of this subsection explains each of the columns of the BAL catalogue in more detail than is presented in Table 1. The online catalogue also includes documentation.[6]

Columns 65 and 66 of the BAL catalogue have information pertaining to the PCA fit used by the BALFINDER and discussed in Section 2.2. The 'PCA_COEFFS' column specifies the coefficients for each of the five eigenspectra used to construct the PCA fit to each quasar (Guo & Martini 2019), and the linear sum of these eigenspectra multiplied by the coefficients is the PCA fit to the spectrum. The column 'PCA_CHI2' is the reduced χ^2 value of this fit. Any object in the catalogue in the redshift range $1.57 < z < 5.0$ will have values in both of these columns, regardless of whether or not there are BAL features in the spectrum.

Column 67 'BAL_PROB' is not calculated in this implementation of the BALFINDER. This column is an artefact of the CNN approach used by Guo & Martini (2019). The value for each object for 'BAL_PROB' will be either zero or −99. A value of zero indicates that the object has been run through the BALFINDER code whereas a value of −99 means that BAL information was not evaluated for the object. The latter will indicate the quasar is outside of the redshift range in which the BALFINDER searches for BALs. We maintain this column for backwards compatibility and in the event we apply a CNN code or other classifier to future DESI quasar catalogues.

Columns 68 through 95 record the strength and location of absorption BAL troughs in the quasar spectra. These values are calculated for the AI and BI criteria relative to both the C IV and Si IV lines. This includes a value for the absorption strength, errors for that value, the number of troughs, velocity ranges for each trough, and the location and depth of the minimum for each absorption feature. For each, a qualifier is specified for clarity in what the value corresponds to. For example, '_2000' corresponds to values where the BI criterion was used and '_450' corresponds to values where the AI criterion was used. Similarly, '_SIIV' corresponds to features blueward of the Si IVλ1398 line and the '_CIV' indicator corresponds to features blueward of the C IVλ1549 line. Since the Si IV line is blueward of C IV, BAL data for the Si IV feature are not available for the lowest-redshift BALs.

The column 'BALMASK' is a bitmask that encodes information about any issues with running the BALFINDER on a given quasar. The definitions for 'BALMASK' are that 0 means the BALFINDER ran successfully, 1 that the BALFINDER was not run, 2 that an object is outside of the BAL z range, and 4 that the redshift difference between the input quasar catalogue and the redshift after BAL masking are greater than 0.001. The bitmasks sum to specify any possible combinations of these cases.

The 'SNR_CIV' column has the average SNR in the wavelength region where we search for BAL features associated with C IV. We calculate the SNR of each pixel as the flux in a given pixel divided by the flux error provided by the spectroscopic pipeline (Guy et al. 2023). The average SNR is the average of all pixels that range from $v = 25\,000\,\mathrm{km\,s^{-1}}$ blueward of the C IVλ1549 emission line to the line centre at $v = 0$ that have $f(v) > 0$.

4.2 Summary properties

The BALs identified in this work are broadly similar to the BALs identified in earlier quasar catalogues from SDSS. For example, the AI and BI distributions of quasars in the DESI sample are qualitatively similar to the distributions in Pâris et al. (2017) and Trump et al. (2006) which examined SDSS DR12 and SDSS DR3 quasars, respectively (see Fig. 3, *left*). The AI distribution shows an excess of BALs at $\log(\mathrm{AI}_{C\,IV}) < 3.3$, which is not as noticeable in the SDSS DR12 or DR14 quasars studied by Pâris et al. (2017) and Guo & Martini (2019). This difference in the AI distributions is likely because the DESI catalogues use a simple $AI > 0$ criterion without consideration of the uncertainty in AI, while the earlier Guo & Martini (2019) and SDSS catalogues used classifiers that rejected measurements of marginal significance. The BI distribution shown in the right panel of Fig. 3 is in better agreement with previous work. This is because BAL features need to be much more prominent to pass the BI criterion, and therefore this quantity is less impacted by a range of SNR values.

We summarize a number of other important properties of the BAL catalogues in Table 2. These include the total number of quasars in each catalogue, the number in the redshift range where we can identify BALs, the number of quasars with one or more $AI > 0$ and $BI > 0$ troughs, and the per cent of quasars in the redshift range that were identified as BAL quasars by each criterion.

We also provide the average SNR of the quasars in the redshift range where we can identify BALs and an estimate of the completeness and purity of the BAL catalogues. The average SNR corresponds to all quasars in the range $1.57 < z < 5$, and not just those identified as BALs. We then use the SNR distribution for each catalogue, combined with a simple function that matches the completeness and purity as a function of SNR presented in Section 3, to calculate the average completeness and purity of each catalogue. Fig. 4 shows the SNR distribution of the BAL and non-BAL quasars. This is the SNR in the region blueward of C IV. Note the mean SNR is higher for the BAL sample, which is expected based on our study of completeness.

[6] https://data.desi.lbl.gov/doc/releases/edr/vac/balqso/

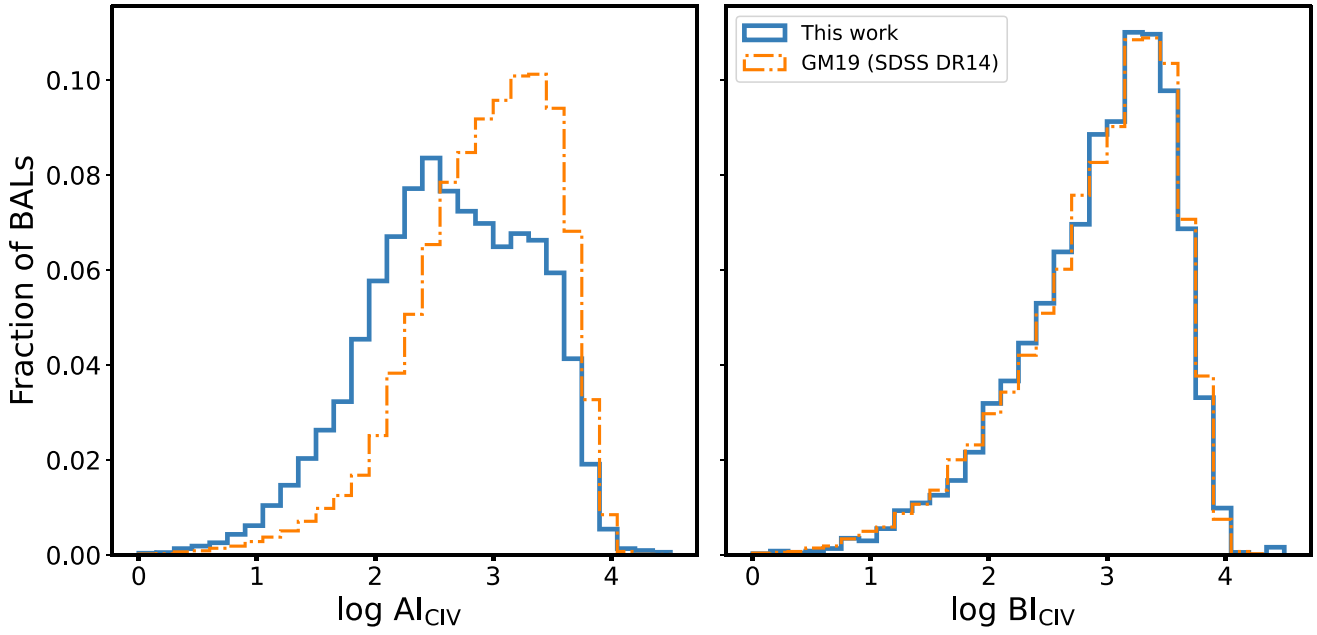

Figure 3. Logarithm of AI (*left panel*) and BI (*right panel*) values for BAL quasars. The distributions are normalized by the total number of quasars identified in each of the samples. The distributions for SDSS DR14 quasars from Guo & Martini (2019) are shown for comparison. The normalizations are different for the AI and BI histograms (see Table 2).

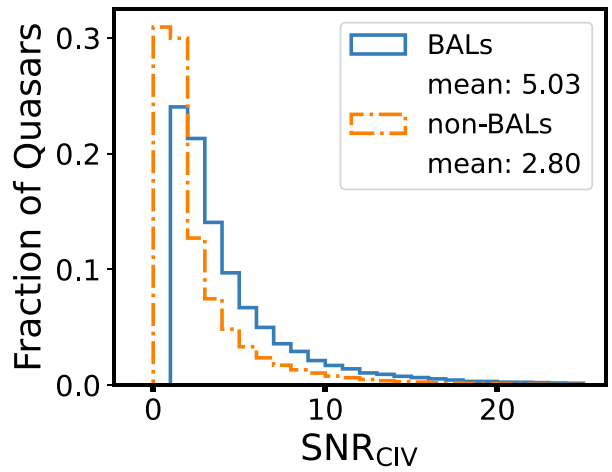

Figure 4. Distribution of SNR in the C IV region for the BAL quasars (*solid blue line*) and non-BAL quasars (*orange dotted line*). The SNR is calculated in the region where we search for absorption, which extends from the systemic wavelength of the C IV emission to 25 000 km s^{-1} blueward of the C IV line. The SNR is the average of the product of the flux and the square root of the inverse variance for each pixel in this range. See Section 4.1 for more details.

4.3 BAL fraction differences

As shown in Table 2, the predicted completeness is highest for SV1, as it has the highest average SNR. This is because SV1 used longer exposure times than planned for DESI in order to obtain greater redshift completeness and thereby better quantify the performance of the target selection algorithms. SV1 also has the highest BAL fractions. The SV3 catalogue is somewhat lower SNR than SV1, as the aim of SV3 was to obtain the same depth expected upon the completion of DESI. Lastly, the DESI-M2 main survey catalogue has the lowest SNR as those spectra typically only have one observation of each Ly α quasar to date, while the goal is to obtain four or five observations of these quasars by the end of the survey. The completeness is consequently lowest for the main survey, and the BAL fractions are correspondingly lowest. We expect the main survey spectra will achieve the SV3 values by the end of the survey.

5 IMPACT ON REDSHIFTS

The presence of BAL absorption features may impact the shape of the C IV and other emission lines, sometimes spectacularly (see e.g. Fig. 1), and produce redshift errors. In this section, we demonstrate that if we mask the locations of BAL features, we obtain better redshift measurements for the BAL quasars. The first subsection describes the masking procedure, and the second subsection describes the new redshift measurements and compares them to the values prior to masking the BAL features. The improvement in redshift has also been demonstrated with mock BAL spectra by García et al. (2023) and with observational data by Brodzeller et al. (2023).

5.1 Masking procedure

Redshift estimation algorithms typically rely on some form of template fitting, and are especially susceptible to redshift errors if the observed quasar is not a great match to the templates. Due to the rich diversity of BALs in quasar spectra, it is not practical to account for all of this diversity with templates, and redshift errors are consequently more common in this subset of the quasar population. To reduce the bias and inherent redshift errors for the BAL sample, we adopted a procedure for masking out the wavelengths associated with the BAL features.

Our approach starts with the BAL features cataloged on the blue side of the C IV line and then assumes that absorption will also be associated with Si IV, N V, and Ly α at the same relative velocities. Absorption is sometimes present in these shorter wavelength features, such as shown in the stacked BAL spectra studied by Mas-Ribas & Mauland (2019), but this is not always

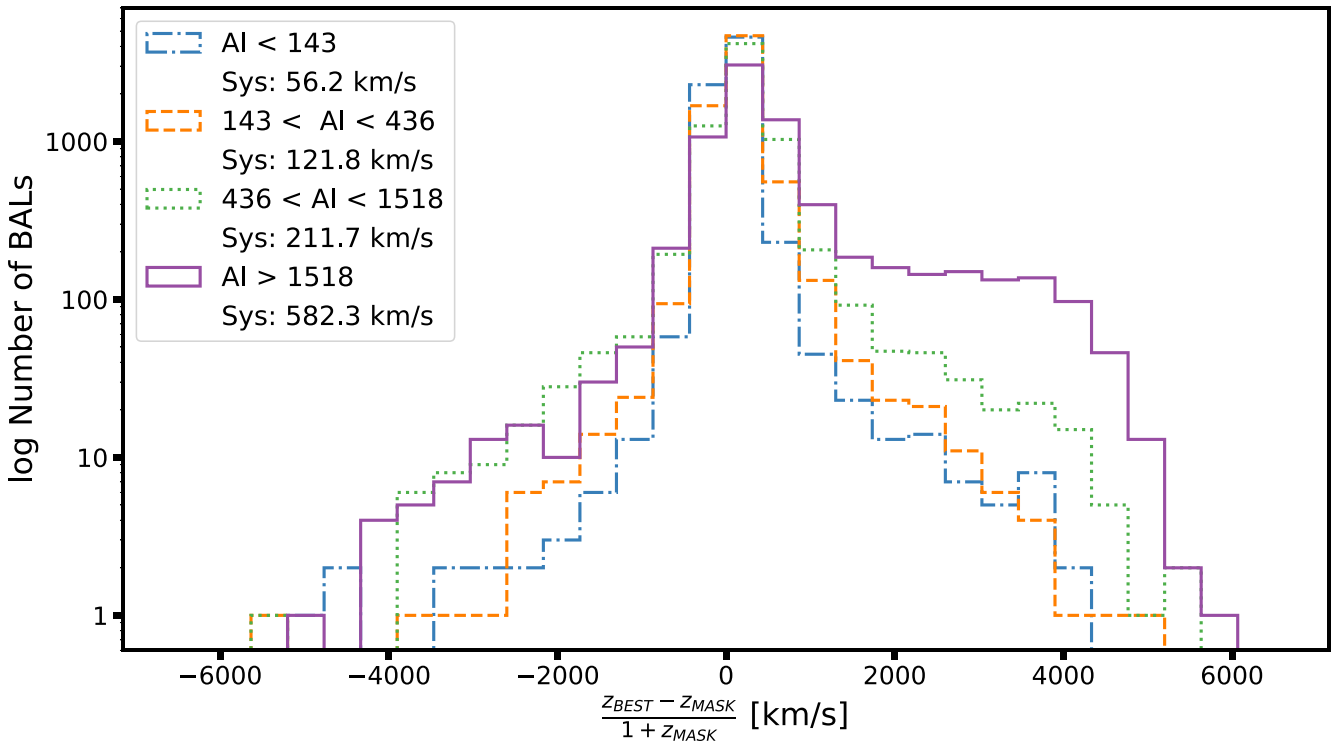

Figure 5. Distribution of velocity differences between the original pipeline redshift and the new redshift after masking BAL features. The distributions are binned into quartiles by AI value. The mean difference is a strong function of AI value in the sense that the velocity different is relatively modest for weak BALs (smallest AI quartile) and increases to 582.3 $\mathrm{km\,s^{-1}}$ for the greatest quartile, which shows redshift errors are more significant for larger AI values. In addition, the velocity differences are strongly skewed towards positive velocity differences, which correspond to a systematic underestimate of the true redshift before masking. This is due to the absorption on the blue side of the line.

the case. Consequently, the masking procedure implemented in this work is a conservative one, tending to overmask the spectra before recomputing the redshift. These emission features are important for redshift measurements, as they are among the strongest emission lines in the observed wavelength range at high redshifts ($z > 1.5$).

The procedure used to mask the absorption features is relatively simple. Each object identified as a BAL quasar by the balfinder is assigned a 'VMIN_CIV_450' and 'VMAX_CIV_450' for each absorption feature. These numbers represent the velocity corresponding to an absorption feature up to 25 000 $\mathrm{km\,s^{-1}}$ blueward of the C IV emission line. For most quasars in the sample, only one absorption trough is present, although in extreme cases there may be many features. We mask these wavelengths by setting the inverse variance to zero for all of the wavelength values between each 'VMIN_CIV_450' and 'VMAX_CIV_450' pair from the BAL catalogue, and repeat this for each of the emission lines considered. We make these changes in a copy of the original quasar spectral data, and the resulting file is identical to the original data except for the changes to the inverse variance values.

5.2 Masked BAL redshifts

Our next step is to rerun Redrock and measure new redshifts for the BAL quasars. The data used to create the quasar catalogues boast a high completeness and purity for quasars, especially in the redshift regime for which BALs are characterized (Chaussidon et al. 2023). Because of this, we only fit the masked BAL quasar spectra with Redrock's quasar spectral templates, rather than all templates, as we are confident that the objects are quasars. We also assume that masked quasar redshifts will not significantly differ from the non-masked redshifts already available in the catalogue. We therefore use the best (pre-masking) redshift to set a Gaussian prior for Redrock, with a width of 0.1.

The outputs from Redrock[7] include the redshift 'Z', redshift error 'ZERR', a warning flag indicating the reliability of the redshift from Redrock 'ZWARN', the χ^2 of the best-fitting template 'CHI2', the difference between the best and second-best fit template 'DELTACHI2', and the spectral classification 'SPECTYPE' (Bailey et al., in preperation). These data are appended to the BAL catalogue with the extra qualifier '_MASK' to make clear that these data come from the run of Redrock after BAL features are masked. Objects which are not BAL quasars will have non-sensical values for all '_MASK' column values, for example, 'Z_MASK' for a non-BAL quasar is set to −99. 'SPECTYPE_MASK' is always 'QSO' for BAL quasars since only quasar templates are used to rerun Redrock after BAL features are masked.

The velocity differences $c(z_{nomask} - z_{mask})/(1 + z_{mask})$ before and after masking range from on order −1000 to +1000 $\mathrm{km\,s^{-1}}$, with the average skewed towards positive values. This offset means that the redshifts after masking are lower than the redshifts before masking, which is consistent with our expectation that the BAL absorption has removed flux from the blue side of the emission lines and biases the line centres redward of their true location. Fig. 5 shows histograms of the velocity differences. The velocity difference averaged over all of the BALs is 243 $\mathrm{km\,s^{-1}}$.

[7]See https://redrock.readthedocs.io/en/latest/api.html for more information.

We separated the BAL sample into four quartiles of AI to determine if there is a trend in the size of the velocity shift with the strength of the absorption. This trend is significant and is shown in Fig. 5. The shift ranges from an average velocity shift of 56.2 km s^{-1} for the first quartile of AI to 582.3 km s^{-1} for the fourth quartile. The tail of the distribution also becomes longer when more highly absorbed objects are in the sample. This result is expected, as more absorption in the quasar spectra corresponds to a larger portion of the spectra with a low flux contribution, and this causes the fits to misinterpret the peak of emission line features.

In addition to this systematic shift in the velocity, we found that 6.7 per cent of the BAL quasars in the sample have changes in redshift greater than the catastrophic error regime required for tracer quasars in DESI ($\Delta v > 1000$km s^{-1}) whereas 9.3 per cent have redshift changes greater than the cut-off required for random errors ($\Delta v > 750$ km s^{-1}; DESI Collaboration 2023a).

6 PROSPECTS FOR RARER BAL CLASSES IN DESI

The large samples of BAL quasars from spectroscopic surveys like SDSS and DESI offer a great opportunity to identify larger samples of rare BALs (e.g. Hall et al. 2002), and the frequency and properties of such systems may lead to new insights into quasar physics. In a study of quasars in the SDSS Third Data Release (DR3), Trump et al. (2006) measured the incidence of HiBALs, LoBALs, and FeLoBALs. They found 26 per cent of quasars are HiBALs with the $AI > 0$ criterion in the redshift range $1.7 \leq z \leq 4.38$ where they could detect C IV absorption. The rate for LoBALs was much lower at 1.31 per cent in the redshift range $0.5 \leq z \leq 2.15$, and the rate for FeLoBALs was lower still at about 0.33 per cent. Many of the rarer classes of BALs were identified through visual inspection, especially the FeLoBALs, as the absorption may be so significant that automated techniques fail. A further complication is that BAL quasars, especially LoBALs and FeLoBALs, tend to have greater reddening than non-BAL quasars (e.g. Hall et al. 1997; Najita, Dey & Brotherton 2000; Morabito et al. 2019).

Alexander et al. (2023) performed visual inspection of a large number of quasars from DESI SV and identified many BAL quasars. We have performed a complimentary visual inspection of BALs identified with the `balfinder` that exhibited poor fits. The example spectra shown in Fig. 6 include ones dominated by high-ionization features of Al III, C IV, and Si IV (*top panel*), and below that examples of a LoBAL quasar with Mg II absorption and two FeLoBALs with prominent absorption that is blueshifted relative to the Fe II multiplets near 2400 and 2600 Å. DESI will observe approximately two million quasars in the redshift range where it could detect BALs associated with Mg II and Fe II, and therefore has the potential to identify several tens of thousands of LoBALs and several thousand FeLoBALs. This could include extremely rare objects like the one studied by Choi et al. (2020).

Many studies of BALs have used their variability to study their physical properties. For example, Filiz Ak et al. (2014) measured changes in the C IV, Si IV, Al III features and compared correlations to disc wind models. One such correlation is that lines of sight with more low ionization material also exhibit broader and deeper C IV troughs. Grier et al. (2016) used multi-epoch observations from several SDSS programs and found that the vast majority show no evidence of acceleration, which implies most of the absorbing material is at large distances from the black hole and/or is not decelerating due to collisions with ambient material along the line of sight. McGraw et al. (2015) studied the absorption line variability of the subclass of FeLoBAL quasars and concluded that the absorbing material can be up to tens of parsecs from the central source. DESI will obtain multiple observations of quasars at $z > 2$ to improve the SNR of the Ly α sample. This will enable variability studies of a significant subset of the many thousands of LoBAL and FeLoBAL quasars that DESI will discover.

Even though the vast majority of DESI spectra will be lower SNR than targeted follow-up studies of BAL quasars, these data will be valuable for stacking analyses. Hamann et al. (2019) created composite spectra based on C IV strength and Al III to investigate weaker features, including ones normally compromised by Ly α forest absorption. They found interesting constraints on the typical densities and distances of the absorbing material, and that the LoBALs tend to have the largest outflow column densities and highest velocities. In another application of stacked BAL spectra, Mas-Ribas & Mauland (2019) found evidence of radiative acceleration through line locking in the C IV line. High SNR stacks of DESI BALs or of individual, rare objects could lead to further insights with the continued growth in sophistication of analysis tools, such as the SIMBAL analysis code (Choi et al. 2022a, b; Leighly et al. 2022).

7 SUMMARY

The DESI BAL catalogues together comprise one of the largest samples of BAL quasars to date, with just under 30 000 BAL quasars identified by our automated searching algorithm. In this paper, we describe the algorithm employed by the DESI Collaboration to identify and measure the properties of BAL quasars. This algorithm identifies BALs via absorption on the blue side of the C IV emission feature, and therefore we only identify BALs in quasars between $1.57 < z < 5$ when this spectral region is clearly visible by DESI. This restriction means that all of the BAL quasars appear to be HiBALs, although we have not conducted a similar, systematic search in the vicinity of low-ionization features such as Mg II. We record the number of absorption troughs that satisfy both the AI and BI criteria and find the distribution of both quantities is similar to previous results from SDSS, except that DESI is somewhat more sensitive to lower AI values. We also record the velocity limits of each absorption trough. The limits of each absorption feature are important to mask spectral regions that may be contaminated by BAL absorption.

We used mock catalogues to analyse the completeness and purity of our algorithm. These mock catalogues use a range of empirical templates derived from high SNR BAL spectra from SDSS that are randomly assigned to the mock quasars. We find that the completeness of our algorithm depends on SNR and decreases significantly for SNR < 2 per pixel. This dependence implies that the completeness will vary between surveys, and we see a difference in the BAL fraction for different DESI catalogues that is consistent with the average completeness of the spectra in each catalogue. The One-Percent Survey was designed to be representative of the final DESI data set. Our results from the SV3 catalogue suggest that the expected fraction of BAL quasars identified by the BALFINDER with at least one AI absorption trough should be $\sim$ 19 per cent once DESI reaches full survey depth.

Historically, one major motivation to identify BALs in cosmological surveys was to eliminate them as a potential source of contamination. One rationale was that BAL features may produce greater redshift errors, which negatively impacts quasar clustering. Another was that BAL absorption features in the Ly α forest region may contaminate measurements of the neutral hydrogen distribution

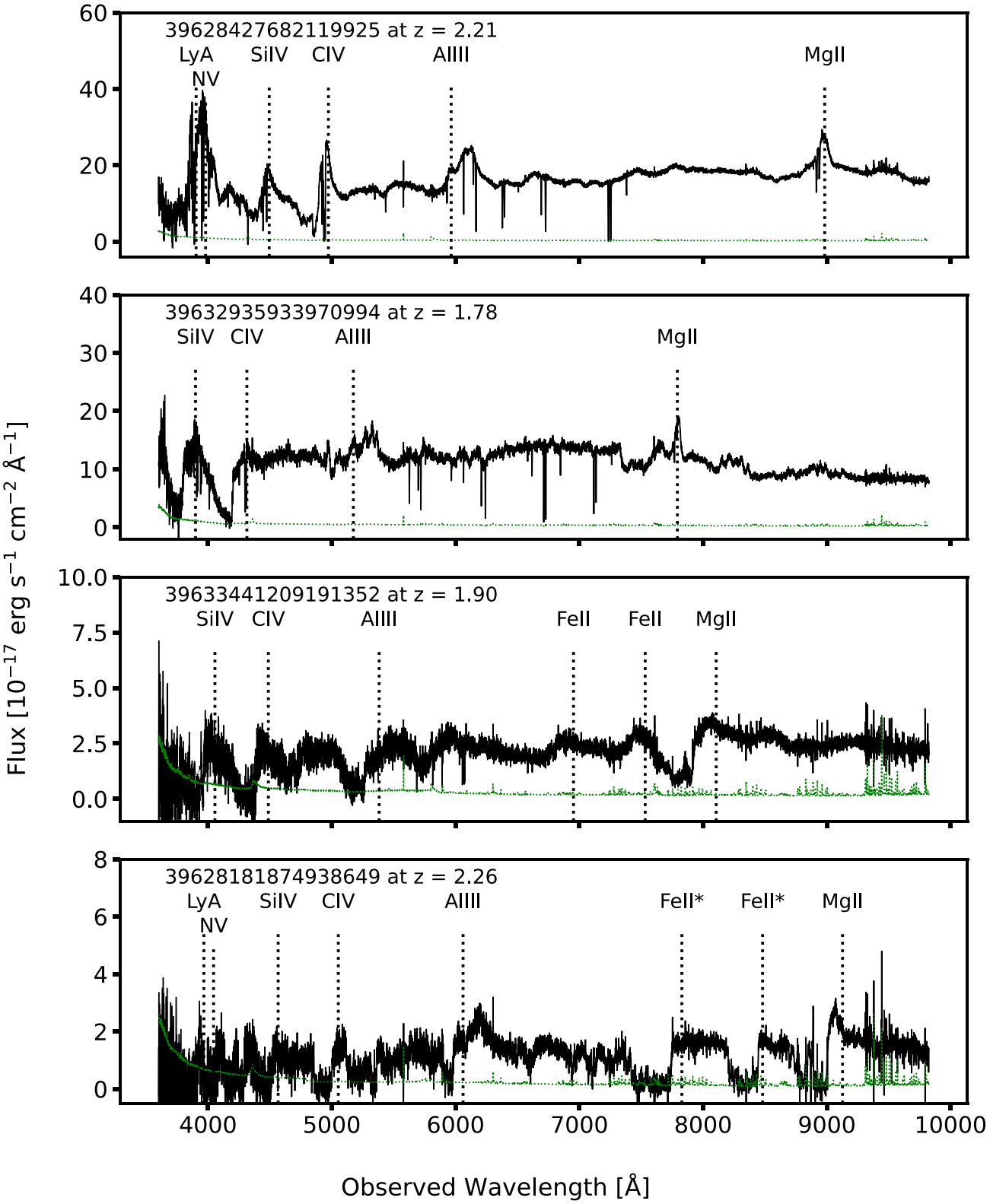

Figure 6. Examples of BALs from the DESI EDR. The top panel shows a BAL at $z = 2.21$ dominated by high-ionization troughs associated with Al III, C IV, and Si IV, although there is weak evidence for some broad, highly blueshifted absorption associated with Mg II. The second panel from the top shows a BAL at $z = 1.78$ with the broad Mg II absorption characteristic of LoBAL quasars, in addition to the common high-ionization features. The third panel shows a BAL at $z = 1.90$ with weak Fe absorption, most notably Fe II absorption, that is characteristic of FeLoBALs. The bottom panel shows a BAL at $z = 2.26$ with the very strong Fe absorption characteristic of FeLoBALs. The flux uncertainty is also shown in each panel (*green, dotted line*).

in quasars above $z > 2$. In order to mitigate the impact of BALs on quasar redshift errors, we mask the BAL features associated with the C IV, Si IV, N V, and Ly α lines and report new redshifts for these quasars. Mock quasar studies by García et al. (2023) show that masking the BAL features decreases the redshift errors to be comparable to those for non-BAL quasars. We show that the velocity shifts of quasars before and after masking are correlated with the strength of absorption and that masking the BAL troughs shifts the redshifts somewhat to the blue. In some cases, the velocity shifts after masking are as large as thousands of km s^{-1}, which are greater than the 1000 km s^{-1} criterion that DESI has adopted to define catastrophic redshift errors, although the average velocity shift is $\sim$ 243 km s^{-1}.

Keeping BAL quasars in DESI increases the statistical power for quasar clustering measurements and increases the number of sightlines that probe the Ly α forest. The projected BAL quasar fraction of $\sim$ 19 per cent identified by the BALFINDER will significantly impact the number of sightlines ultimately kept for the cosmological analysis. Specifically, Ennesser et al. (2022) showed with eBOSS observations that adding back in the 12 per cent of BAL quasars in eBOSS decreased the uncertainties on the autocorrelation function by 12 per cent. While some of the spectral range is masked in BALs, this is offset by the higher average SNR of each BAL quasar relative to non-BAL quasars, or at least relative to quasars where we do not detect BAL features. In the DESI sample, the average SNR per pixel of BALs is about five, as opposed to just under three for non-BAL quasars. This and future BAL catalogues will be an invaluable part of maximizing the scientific return of the DESI survey.

ACKNOWLEDGEMENTS

SMF is grateful for support from Ohio State that includes two College of Arts and Sciences Undergraduate Research Scholarships, summer research support from the Departments of Astronomy and Physics, and scholarships including the Smith Student Support scholarship and the Ann Slusher Tuttle undergraduate scholarship. PM and LE acknowledge support from the United States Department of Energy, Office of High Energy Physics under Award Number DE-SC-0011726. 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 Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: https://www.desi.lbl.gov/collaborating-institutions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U. S. National Science Foundation, the U. S. Department of Energy, or any of the listed funding agencies. The authors are honoured to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

DATA AVAILABILITY

The BAL catalogues for data collected during DESI EDR observations can be downloaded from https://data.desi.lbl.gov/doc/releases/edr/vac/balqso/. Data points for the Figures in this publication can be accessed in the following zenodo repository https://zenodo.org/record/8267633.

REFERENCES

Alexander D. M. et al., 2023, AJ, 165, 124
Bautista J. E. et al., 2017, A&A, 603, A12
Blanton M. R. et al., 2017, AJ, 154, 28
Blomqvist M. et al., 2019, A&A, 629, A86
Brodzeller A. et al., 2023, AJ, 166, 66
Busca N., Balland C., 2018, preprint (arXiv:1808.09955)
Busca N. G. et al., 2013, A&A, 552, A96
Chaussidon E. et al., 2022, MNRAS, 509, 3904
Chaussidon E. et al., 2023, ApJ, 944, 107
Choi H., Leighly K. M., Terndrup D. M., Gallagher S. C., Richards G. T., 2020, ApJ, 891, 53
Choi H., Leighly K. M., Dabbieri C., Terndrup D. M., Gallagher S. C., Richards G. T., 2022a, ApJ, 936, 110
Choi H., Leighly K. M., Terndrup D. M., Dabbieri C., Gallagher S. C., Richards G. T., 2022b, ApJ, 937, 74
Dawson K. S. et al., 2013, AJ, 145, 10
Dawson K. S. et al., 2016, AJ, 151, 44
DESI Collaboration, 2016a, preprint (arXiv:1611.00036)
DESI Collaboration, 2016b, preprint (arXiv:1611.00037)
DESI Collaboration, 2022, AJ, 164, 207
DESI Collaboration, 2024a, AJ, 167, 62
DESI Collaboration, 2024b, AJ, 168, 58
du Mas des Bourboux H. et al., 2020, ApJ, 901, 153
de Sainte Agathe V. et al., 2019, A&A, 629, A85
Dey A. et al., 2019, AJ, 157, 168
Eisenstein D. J. et al., 2011, AJ, 142, 72
Ennesser L., Martini P., Font-Ribera A., Pérez-Ràfols I., 2022, MNRAS, 511, 3514
Farr J. et al., 2020a, J. Cosmol. Astropart. Phys., 2020, 068
Farr J., Font-Ribera A., Pontzen A., 2020b, J. Cosmol. Astropart. Phys., 2020, 015
Filiz Ak N. et al., 2014, ApJ, 791, 88
Foltz C. B., Chaffee F. H., Hewett P. C., Weymann R. J., Morris S. L., 1990, BAAS, 22, 806
Font-Ribera A. et al., 2014, J. Cosmol. Astropart. Phys., 2014, 027
García L. A. et al., 2023, MNRAS, 526, 4848
Gibson R. R. et al., 2009, ApJ, 692, 758
Grier C. J. et al., 2016, ApJ, 824, 130
Guo Z., Martini P., 2019, ApJ, 879, 72
Guy J. et al., 2023, AJ, 165, 144
Hall P. B., Martini P., DePoy D. L., Gatley I., 1997, ApJ, 484, L17
Hall P. B. et al., 2002, ApJS, 141, 267
Hamann F., Herbst H., Paris I., Capellupo D., 2019, MNRAS, 483, 1808
Herrera-Alcantar H. K. et al., 2023, preprint (arXiv:2401.00303)
Hopkins P. F., Richards G. T., Hernquist L., 2007, ApJ, 654, 731
Hou J. et al., 2021, MNRAS, 500, 1201
Leighly K. M., Choi H., DeFrancesco C., Voelker J., Terndrup D. M., Gallagher S. C., Richards G. T., 2022, ApJ, 935, 92
Lyke B. W. et al., 2020, ApJS, 250, 8
Mas-Ribas L., Mauland R., 2019, ApJ, 886, 151
McGraw S. M., Shields J. C., Hamann F. W., Capellupo D. M., Gallagher S. C., Brandt W. N., 2015, MNRAS, 453, 1379
Miller T. N. et al., 2023, preprint (arXiv:2306.06310)
Morabito L. K. et al., 2019, A&A, 622, A15
Myers A. D. et al., 2023, AJ, 165, 50
Najita J., Dey A., Brotherton M., 2000, AJ, 120, 2859
Neveux R. et al., 2020, MNRAS, 499, 210

Niu W., 2020, American Astronomical Society Meeting Abstracts #235, p. 108.06
Pâris I. et al., 2017, A&A, 597, A79
Pâris I. et al., 2018, A&A, 613, A51
Reichard T. A. et al., 2003, AJ, 125, 1711
Schlafly E. F. et al., 2023, AJ, 166, 259
Schmidt M., 1963, Nature, 197, 1040
Silber J. H. et al., 2023, AJ, 165, 9
Trump J. R. et al., 2006, ApJS, 165, 1
Weymann R. J., Morris S. L., Foltz C. B., Hewett P. C., 1991, ApJ, 373, 23
York D. G. et al., 2000, AJ, 120, 1579
Zou H. et al., 2017, PASP, 129, 064101

[1]*Department of Astronomy, The Ohio State University, 140 W 18th Avenue, Columbus, OH 43210, USA*
[2]*Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue 191 West Woodruff Avenue, Columbus, OH 43210, USA*
[3]*Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA*
[4]*Department of Physics and Astronomy, The University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, USA*
[5]*Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[6]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[7]*Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA*
[8]*Departamento de Física, Universidad de Guanajuato - DCI, C.P. 37150, Leon, Guanajuato, México*
[9]*Department of Astronomy, Tsinghua University, 30 Shuangqing Road, Haidian District, Beijing 100190, P.R. China*
[10]*Departament de Física Quàntica i Astrofísica, Universitat de Barcelona, Martí i Franquès 1, E-08028 Barcelona, Spain*
[11]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra Barcelona, Spain*
[12]*Le Département de Physique, Universitat de Barcelona, Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille 13284, France*
[13]*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[14]*Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), Sorbonne Université, CNRS/IN2P3, F-75005 Paris, France*
[15]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[16]*Physics Department, Boston University, 590 Commonwealth Avenue, Boston, MA 02215, USA*
[17]*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[18]*Instituto de Física, Universidad Nacional Autónoma de México, Cd. de México C.P. 04510, México*
[19]*Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, CP 111711, Bogotá, Colombia*
[20]*Observatorio Astronómico, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio H, CP 111711 Bogotá, Colombia*
[21]*Departament de Física, Serra Húnter, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain*
[22]*NSF's NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, USA*
[23]*Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys, 23, E-08010 Barcelona, Spain*
[24]*Department of Physics and Astronomy, Siena College, 515 Loudon Road, Loudonville, NY 12211, USA*
[25]*National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Rd., Chaoyang District, Beijing 100012, P.R. China*
[26]*Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[27]*Perimeter Institute for Theoretical Physics, 31 Caroline St North, Waterloo, ON N2L 2Y5, Canada*
[28]*Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[29]*Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, s/n, E-18008 Granada, Spain*
[30]*Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA*
[31]*Department of Physics and Astronomy, Sejong University, Seoul 143-747, Korea*
[32]*CIEMAT, Avenida Complutense 40, E-28040 Madrid, Spain*
[33]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[34]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
[35]*Department of Physics & Astronomy, Ohio University, Athens, OH 45701, USA*

This paper has been typeset from a TEX/LATEX file prepared by the author.