

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Maximizing the Output of Large Sky Surveys: From the IGM to Milky-Way Dwarf Spheroidal Galaxies

Permalink

<https://escholarship.org/uc/item/3q89j4tz>

Author

Ding, Jiani

Publication Date

2024

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|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**MAXIMIZING THE OUTPUT OF LARGE SKY SURVEYS: FROM
THE IGM TO MILKY-WAY DWARF SPHEROIDAL GALAXIES**

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ASTRONOMY AND ASTROPHYSICS

by

Jiani Ding

September 2024

The Dissertation of Jiani Ding
is approved:

Professor Constance Rockosi, Chair

Professor Puragra GuhaThakurta

Professor Piero Madau

Professor Zheng Cai

Peter Biehl, Vice Provost and Dean of Graduate Studies

Copyright © by

Jiani Ding

2024

Table of Contents

List of Figures	v
List of Tables	viii
Abstract	ix
Dedication	x
Acknowledgments	xi
1 Introduction	1
2 A New Measurement of the Mean Transmitted Flux in the Lyman-α and Lyman-β Forest	7
2.1 Introduction	7
2.2 DATA SET AND METHOD	11
2.2.1 Systematic Errors	14
2.2.2 Final Composite Spectra	16
2.3 Mean Transmitted Flux	18
2.3.1 Ly α Opacity	18
2.3.2 Metal Line Contamination	23
2.3.3 Correction for Optically Thick Absorbers	23
2.3.4 Ly β Relative Opacity	24
2.4 DISCUSSION	27
2.5 Conclusion	31
2.6 Appendix	32
2.6.1 Bayesian Inference Parameters and Priors	32
2.6.2 Covariance Array	32
2.6.3 Convergence of MCMC Algorithm	36
2.6.4 Systematic Errors Estimation from quasar SEDs	41
3 Building DLA catalog for the DESI QSO Spectra	44

4	The Draco dSph Galaxy in DESI	52
4.1	Introduction	52
4.2	Data	54
4.2.1	Photometry Selection	56
4.3	Methods	56
4.4	Results	63
4.4.1	Spatial Distribution of the Sample	63
4.4.2	Velocity Dispersion and Metallicity gradient	66
4.4.3	Surface Density Profile	67
4.5	Discussion	69
4.5.1	Surface Number Density at Large Radii	70
4.5.2	High Probability Member Stars in the Outskirts of Draco	73
4.5.3	Comparison with Other Recent Draco Results	73
4.6	Conclusion	76
5	Unveiling the Outskirt of Sextans: Searching for Stars along the Orbit of Sextans Dwarf through 6D Orbit Integration	87
5.1	Introduction	87
5.2	Data	91
5.2.1	Sextans Sample Selection: Photometry	93
5.3	Gaussian Mixture Model to select candidate Sextans members	95
5.3.1	Spatial Distribution of the Sample	103
5.3.2	High Probability Member Stars in the Outskirts of Sextans	103
5.3.3	Gaussian Mixture Model Sample Validation	109
5.4	Orbits of Candidate Sextans Members	110
5.5	Orbit Integration	111
5.5.1	Computing the Integrals of Motion	112
5.6	K-Nearest Neighbour Classification of Sextans Candidates	113
5.6.1	KNN Classification Results	114
5.7	Discussion	115
5.7.1	KNN Sextans members from the GMM orbit sample	115
5.7.2	Extra-tidal Stars from the GMM orbit sample	116
5.7.3	Sextans member stars from the BHB candidates	117
5.7.4	Difference in extra-tidal structure around Draco and Sextans	118
5.8	Conclusion and Future work	119
6	Conclusion	138

List of Figures

2.1	The composite quasar spectra	13
2.2	Flux ratio of composite quasar spectra from SDSS DR14 and DR7	15
2.3	Same as Fig. 2.2 but after a $S/N > 2.2$ pixel ⁻¹ cut.	16
2.4	Final composite spectra from eBOSS DR14 data in different redshift bins.	17
2.5	MCMC method in our study	20
2.6	MCMC best-fit model to DR14 composite spectra	21
2.7	Same as the previous figure in the redshift interval $z_c = 3.65 - 4.69$. . .	22
2.8	Observational determinations from DR14 composite spectra of the Ly α opacity in the redshift range 2.4 – 4.4	25
2.9	Relative Ly β effective opacity of the IGM from eBOSS DR14 data. The ten redshift bins cover the range $2.8 \leq z \leq 4.6$	26
2.10	The comparison between $\tau_{\text{Ly}\alpha}$ measurement from the best fitted spline evolution for $\tau_{\text{Ly}\alpha}$	28
2.11	Comparison between $\tau_{\text{Ly}\alpha}$ measurement from the best fitted spline evolution for $\tau_{\text{Ly}\alpha}$	29
2.12	The Covariance matrices for DR14 composite spectra in the redshift interval $z_c = 2.85 - 3.55$	37
2.13	Covariance matrices for DR14 composite spectra at $z \geq 3.65$	38
2.14	MCMC final converged traces and kernel densities for the spline points used to represent the quasar continuum SED	39
2.15	MCMC final converged traces and kernel densities for the spline points used to represent the Ly α opacity vs. redshift	40
2.16	Final corner plot for all the parameters we used to represent the continuum and $\tau_{\text{Ly}\alpha}(z)$ in the MCMC fitting.	42
2.17	$\tau_{\text{Ly}\alpha}$ measurement for quasar spectra with three different spectral index	43
3.1	The three-layers CNN architecture used in our project	49
3.2	The data structure of the DLA Finder	50
3.3	My preliminary work on the purity and completeness of the SDSS trained CNN model run on the DESI London mock v09	51

4.1	The region inside the dashed lines is the final selection for the Draco candidates sample from the DESI SV data	57
4.2	The probability distribution for the Draco members predicted from the Gaussian mixture model. From the probability distribution, we define the high probability members to be prob > 0.9	64
4.3	The color-magnitude diagram with the membership probabilities from the mixture model	79
4.4	Spatial distribution for the probability greater than 0.9 star sample with the tidal radius in blue ellipse	80
4.5	The velocity dispersion and the metallicity gradient of our high probability Draco members	81
4.6	The metallicity gradient measured from the mean metallicity of our high probability Draco members	82
4.7	The surface number density for stars with membership probability greater than 0.9, corrected for sample incompleteness	83
4.8	The number of stars as a function of radius	84
4.9	Comparison of proper motion (PM) measurement between this work and previous literature	85
4.10	[Fe/H] versus heliocentric radial velocity distribution and the proper motion distribution of high probability members in our work	86
5.1	Final selection for the Sextans candidate sample from the DESI SV data	94
5.2	Probability distribution for Sextans members identified by the Gaussian mixture model	101
5.3	Color magnitude diagram with membership probabilities from the mixture model	120
5.4	Spatial distribution of the high probability Sextans stars	121
5.5	Comparison of proper motion (PM) measurement between this work and previous literature	122
5.6	The metallicity distribution before and after the distance cut and distance percentage error cut for the GMM high probability sample	123
5.7	Distance modulus distribution for the GMM orbit sample	124
5.8	Rotation curves for the scaled potential from the MWpotential2014 used in our study	125
5.9	$E-L_z$ plot for all stars in the BHB candidate sample	126
5.10	$E-L_z$ plot for all the BHB candidates (black) and the GMM orbit sample (blue)	127
5.11	J_r-J_z plot for all the BHB candidates	128
5.12	$E-L_z$ plot for the contours (16th, 50th 84th) of the training sample for the KNN	129
5.13	$E-L_z$ plot for the performance of the KNN	130
5.14	J_r-L_z plot for the performance of the KNN	131
5.15	J_z-L_z plot for the performance of the KNN	132

5.16	Orbits for the GMM orbit sample	133
5.17	Orbits for the selected BHB members by the KNN	134
5.18	Pericenter versus apocenter for the BHB stars selected by the KNN . . .	135
5.19	[Fe/H] distributions for the selected Sextans members	136
5.20	Distance modulus distribution for the extra-tidal stars	137

List of Tables

2.1	Composite Spectra. Columns give the number of quasars included in each composite, the minimum (z_{\min}) and maximum (z_{\max}) quasar redshifts in each bin, and the mean redshift (z_c) of the composite.	12
2.2	Continuum spline points in rest-wavelength space.	33
2.3	$\text{Ly}\alpha$ opacity (corrected for metal lines but not for LLS) spline points in redshift space (see final_results for details).	34
2.4	Ranges of the prior uniform distributions for the 22 spline points used in our MCMC inference.	35
4.1	Priors for the Gaussian Mixture Model	62
4.2	Values for the Best Fitted Parameters	65
4.3	Values for the Best Fitted Parameters for Walker et al. (2023) sample	74
4.4	Properties for the eight Star Members Outside the Tidal Radius	78
5.1	Priors for the Gaussian Mixture Model	100
5.2	Values for the Best Fitted Parameters	102
5.3	Properties for the 52 Star Members Outside the Tidal Radius	105
5.3	Continued	106
5.3	Continued	107
5.3	Continued	108

Abstract

MAXIMIZING THE OUTPUT OF LARGE SKY SURVEYS: FROM the IGM TO MILKY-WAY DWARF SPHEROIDAL GALAXIES

by

Jiani Ding

The availability of large sky surveys, including the SDSS and DESI, have revolutionized research across various fields from the large-scale structure of the universe to the evolutionary histories of the Milky Way and its satellite galaxies. In my thesis, I utilize machine learning methods to analyze the large datasets from SDSS and DESI. I first show new measurements of the mean transmitted flux in the hydrogen Ly α and a measurement of the relative transmitted flux in Ly β . This measurement uses 27,008 quasar spectra from the Fourteenth Data Release (DR14) of the Extended Baryon Oscillation Spectroscopic Survey (eBOSS). The $\tau_{\text{Ly}\alpha}$ values show a smooth increase by a factor of 5 over the redshift range $z = 2.4 - 4.4$. I then investigate the spatial distribution, kinematics and metallicity of the Draco and Sextans dwarf spheroidal galaxies using DESI data. For Draco, I find minimal evidence for an extended distribution of stars and conclude it has not been significantly tidally disturbed or undergone extensive tidal stripping. For Sextans, through 6D orbit integration, I identify 16 candidate extra-tidal stars and argue that Sextans may have had significant dynamical interactions, possibly with other dwarf galaxies before accretion by the Milky Way.

Acknowledgments

I would like to express my deepest gratitude to Prof. Constance Rockosi for her guidance and insightful ideas over the past year as my PhD advisor. Her support is important for my research and academic growth. Since I began working with Prof. Rockosi in 2020, she has been considerate of both my research and personal life. Her support was especially crucial during the pandemic, a time when I faced significant challenges. Her encouragement and advice have greatly contributed to the quality of my research and thesis. I am profoundly grateful for her support and dedication.

I would also like to sincerely thank Prof. Puragra GuhaThakurta, Prof. Piero Madau and Prof. Zheng Cai for being my thesis review committee, and for providing valuable review and revision advice for my thesis. I want to express my gratitude to Prof. Puragra GuhaThakurta for his valuable general advice. I would like to extend my thanks to Prof. Zheng Cai for his invaluable guidance during the initial two research projects and for his continued support throughout my PhD career. I would also like to thank Prof. Piero Madau for all the contributions to the $\tau_{Ly\alpha}$ project, and all the useful guidance and helpful advice in completing the paper.

Thanks should also go to Prof. J. Xavier Prochaska. I have learned a lot about machine learning techniques from him, and he has provided insightful advice for the $\tau_{Ly\alpha}$ project.

I am also very grateful to all my friends and colleagues for supporting me through my PhD. I would like to thank Qiong Li, Siwei Zou, Xiangyu Jin and Jessie

Peng for their companionship and support during the tough times in my life.

I would like to thank all my family members, especially my parents for providing me with constant support and love.

I would like to sincerely thank my partner, Minghao Yue, for his countless support, meticulous care, helpful life advice, and a forever safe harbor.

Chapter 1

Introduction

The astrophysics community has witnessed the substantial scientific impact and contributions made by large-scale sky surveys over the past decades, such as the Sloan Digital Sky Survey (SDSS; York et al., 2000), Dark Energy Survey (DES; Dark Energy Survey Collaboration et al., 2016), and the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al., 2022a). By providing large and homogeneous datasets throughout our universe, these surveys have revolutionized research across various fields, from the large-scale structures of the universe to the evolutionary histories of the Milky Way and its satellite galaxies.

Similar to how a wide-angle lens can capture the essence of a grand vista, the extensive survey area (covering 8000 square degree of the sky at Galactic latitudes above 30 degree) and depth (g band magnitude up to 23 mag) from the SDSS survey has contributed to comprehensive investigations mapping the distributions of the luminous and non luminous structure in the universe (York et al., 2000).

As a component of the SDSS, the Extended Baryon Oscillation Spectroscopic Survey (eBOSS), was designed to map the large-scale structure of the universe with unprecedented detail. As part of SDSS-IV, eBOSS extend the SDSS surveys by targeting a diverse range of cosmic tracers including galaxies, quasars, and Lyman-alpha forest absorbers Dawson et al. (2016). The quasar sample from eBoss contains over 500,000 quasars with over 140,000 quasars newly discovered since the beginning of SDSS-IV. This sample is detected over 9376 deg^2 (Pâris et al., 2018). One major reasons that this quasar sample is crucial for astrophysical studies is that quasars are luminous spotlights in the distant universe. The absorption lines imprinted by the quasar spectra on the neutral gas is a powerful and unique probe of the large scale structure including the IGM, and the underlying dark matter. The IGM is gas outside the virial radius of galaxies and clusters, spanning the vast spaces between galaxies. Greater than 90% of baryonic mass is located in diffuse matter, and IGM thermal history is connected to the thermal temperature and ionization state histories resulting from structure formation. Therefore, the IGM holds key insights into the reservoir of gas for galaxy formation, the thermal history of the universe, measuring the total ionizing photon production of galaxies, and the underlying matter power spectrum (e.g., Becker et al., 2011; Becker & Bolton, 2013; Palanque-Delabrouille et al., 2013, 2015; Villasenor et al., 2022).

Building on the legacy of SDSS, the Dark Energy Spectroscopic Instrument (DESI) has emerged as the largest current ground-based spectroscopic survey ground-based large spectroscopic survey. With advanced instrumentation (DESI Collaboration et al., 2022b) built on the 4 meter Mayall telescope, DESI is able to observe 5,000 spectra

simultaneously with a resolution 2000-5500 depending on wavelength for a wavelength window from 360 nm to 980 nm. DESI will enhance our ability to probe the fundamental properties of the cosmos and continue the groundbreaking work initiated by the SDSS (DESI Collaboration et al., 2016a). DESI will advance the study of baryon acoustic oscillations (BAO) and the growth of structure with a much larger sample of over 35 million galaxies and quasars. The larger sample, survey volume and survey depth of DESI allows a much refined measurement of cosmological parameters and galaxy formation and evolution (DESI Collaboration et al., 2016b,c)

While DESI's primary mission focuses on the precise measurement of dark energy and the expansion history of the universe, it also includes a significant program dedicated to studying the Milky Way along with its primary 5-year cosmological program (Cooper et al., 2023). The DESI Milky Way Survey (MWS) aims to map the structure, kinematics, and chemical composition of stars in our galaxy, providing new insights into its formation and evolution. The MWS targets stars with r band magnitude up to 21 mag. In the total five year program, the MWS will observe about seven million stellar spectra at Galactic latitudes $|b| > 20$ degrees with a rapid survey rate of ~ 660 targets per square degree.

Among these targets, the dwarf galaxies play a crucial role in understanding the formation and evolution of the Milky Way, as well as constraining the properties of dark matter on small scales. Dwarf galaxies are metal poor, dark matter dominated and represent the faint end of the luminosity function. Therefore, they open up a unique window on the study of the hierarchical formation of galaxies, the gravitational

potential of the MW and dark matter properties on small scales. They are also key to understanding the dark matter distribution, star formation history, chemical and dynamical evolution of the Milky Way (MW) (Bullock et al., 2000; Bullock & Johnston, 2005; Bovill & Ricotti, 2009; Weinberg et al., 2015; Simon, 2019).

The first dSph discovered was Sculptor by Shapley (1938) dating back to 1938. The recent discovery of new dwarf spheroidal galaxies (dSphs) in the Dark Energy Survey (Bechtol et al., 2015a; Koposov et al., 2015a), Sloan Digital Sky Survey (Belokurov et al., 2006a, 2007, 2008; Koposov et al., 2008; Drlica-Wagner et al., 2015), Magellanic Satellites Survey (Torrealba et al., 2018) and similar databases has revolutionized our knowledge of the Milky Way's satellite systems.

In the past few decades, numerous studies have focused on the velocity dispersion, chemical abundance, metallicity gradient as well as the surface density of dwarf galaxies (e.g., Walker et al., 2007, 2008; Kirby et al., 2011). Astrometric and spectroscopic surveys including Gaia (Gaia Collaboration et al., 2016, 2018a) and DESI provide chemical and kinematic information for a tremendous number of stars. There are now many studies focusing on faint structures, like tidal tails and stellar streams, around the dwarf galaxies. In the picture of hierarchical galaxy formation, the differential gravitational forces from the Milky Way gradually strip the stars from the dwarf galaxies through tidal disruption. These tidal forces stretch the stars into tidal tails and elongated streams that trace the orbit of their progenitors Johnston (1998). These structures are sensitive tracers for the underlying Milky Way gravitational potential, the behavior of the dark matter (DM) properties and the DM particle effects on the dSph

mass profiles (e.g., Erkal et al., 2017; Ibata et al., 2020). The addition of radial velocities and distance estimates from spectroscopic surveys and proper motions from the Gaia data release give us unprecedented, precise 6D kinematic information about these structures. It opens up a wide and unique window on the kinematic and spatial distribution of dwarf galaxies, streams and the outer MW halo. These can give us new constraints on the MW accretion history, potential and dark matter density distributions.

Motivated by the previous influential scientific output from all the previous large sky surveys, this dissertation aims at utilizing the unprecedented data sample from SDSS to DESI to delineate a more comprehensive picture of galaxy and large scale structure evolution. With the implementation of machine learning technique to unveil the information available in large sky survey data, this dissertation focus on research from large scale structure traced by the IGM to small scale substructure revealed by Milky Way dwarf galaxies.

In Chapter 2, I present my work using the SDSS data to study the IGM properties. I motivate and discuss the details of a refined procedure for determination of the effective opacity, $\tau_{\text{Ly}\alpha}$, of the Ly α forest (excluding high column density absorbers), and a new measurement of the relative effective opacity, $\Delta\tau_{\text{Ly}\beta}$, of the Ly β forest, based on 27,008 quasar spectra at redshift $2.8 < z < 4.8$ from SDSS DR14.

In Chapter 3, I summarize my contribution to developing a Damped Ly α (DLA) finder based on a CNN algorithm for the DESI survey.

In Chapter 4, I discuss my study using a Gaussian Mixture Model to investigate the spatial distribution, kinematics and metallicity of the Draco dwarf spheroidal galaxy

using data from the Dark Energy Spectroscopic Instrument (DESI) (DESI Collaboration et al., 2023a)

In Chapter 5, I search for tidally stripped stars in the Sextans dSph in order to investigate the history of its dynamical interactions with the MW. I use 6D orbital integration to select candidate Sextans member stars from the Sextans dwarf galaxy and BHB stars from the DESI data. I examine the potential Sextans members giants and halo BHB stars selected from the 6D kinematic information as well as the implication of this technique to search for halo stars associated with other dwarf galaxies.

Chapter 2

A New Measurement of the Mean Transmitted Flux in the Lyman- α and Lyman- β Forest

2.1 Introduction

Intergalactic hydrogen scatters Lyman series radiation and produces a “forest” of absorption lines in the spectra of distant quasars. This forest is a powerful probe of the evolution of cosmic baryons and of the formation of large-scale structures in the Universe, as it traces gas density fluctuations, the underlying dark matter distribution, and the ionization state and temperature of the diffuse intergalactic medium (IGM) following reionization (see, e.g., Meiksin, 2009; McQuinn, 2016; Gnedin & Madau, 2022). The primary second-order statistic derived from spectroscopic data – the 1D power spectrum of the flux distribution in the forest – provides one of the best tools for

measuring the smoothness of the density field and constrain the nature of the dark matter (e.g., “cold” versus “warm”), cosmological parameters, and the thermal and reionization history of the IGM (e.g., Palanque-Delabrouille et al., 2013; Viel et al., 2013; Palanque-Delabrouille et al., 2015; Nasir et al., 2016; Iršič et al., 2017; Yèche et al., 2017; Chabanier et al., 2018; Garzilli et al., 2019; Villasenor et al., 2022).

While the flux power spectrum contains information encoded across different spatial scales, a more basic quantity, the effective optical depth of the forest $-\ln F(z)$, where $F(z) \equiv \langle F^i(z) \rangle$ is the mean transmitted flux at a given redshift, gives insight into the state of diffuse baryons, yields a global measurement of the overall H I content of the highly ionized IGM, and allows for estimates of the intensity of the ionizing background radiation produced by star-forming galaxies and active galactic nuclei (AGN; Bolton et al., 2005; Becker & Bolton, 2013).

Specifically, gas in the IGM comprises most of the baryons in the Universe and approximately follows a density–temperature power-law relation of the form (e.g. Hui & Gnedin, 1997; Puchwein et al., 2015; McQuinn & Upton Sanderbeck, 2016; Villasenor et al., 2021),

$$T = T_0 \Delta^{\gamma-1} \tag{2.1}$$

where $\Delta = \rho_b/\bar{\rho}_b$ is the gas over-density, T_0 is the temperature at the mean cosmic density $\bar{\rho}_b$, and $\gamma - 1$ is the power-law index of the relation. The time evolution of the parameters T_0 and γ is determined by photoheating from hydrogen and helium ionization, recombination, Compton, and expansion cooling, and collisional processes. Assuming the IGM is in photoionization equilibrium, the optical depth of Lyman series

lines can be linked to these properties and thus can in principle constrain T_0 , γ , and the photoionization rate per hydrogen atom Γ_{HI} .

There are different ways to perform a measurement of the mean Ly α transmission through the forest. One is to estimate the unabsorbed continuum in either high or medium resolution spectra and measure the transmitted flux as $F^i = f_{\text{obs}}^i / f_{\text{cont}}^i$, where f_{obs}^i and f_{cont}^i are the observed flux and the unabsorbed continuum at each pixel, respectively (e.g. Songaila, 2004; Kirkman et al., 2005; Dall’Aglio et al., 2008). Perhaps the most accurate determinations with this technique are those of Faucher-Giguère et al. (2008a), who measured the Ly α transmission in 86 high resolution quasar spectra. The common challenge faced by these studies is the difficulty of identifying the peaks of transmission near unity, as the finite resonant scattering opacity in even the most underdense regions of the IGM causes a systematic continuum bias that becomes more severe with increasing redshifts (Faucher-Giguère et al., 2008a). This issue does not affect determinations based on extrapolating the continuum from redward of the Ly α line, as usually done in the case of lower resolution data for which a direct local continuum estimation is not feasible (e.g. McDonald et al., 2005; Pâris et al., 2011).

Another approach is to perform relative measurements using medium resolution quasar spectra in large datasets such as the Sloan Digital Sky Survey (SDSS). Here one exploits the general lack of evolution in the mean unabsorbed quasar continuum to compute, without continuum fitting, the mean transmitted flux $F(z)$, as a fraction of its value at $z \leq 2.5$, then converts to absolute values by scaling to the measurements made by Faucher-Giguère et al. (2008a) from high-resolution data, where continuum

errors are minimal. Using this technique, Becker et al. (2013) presented a new and highly precise measurement of the mean transmitted Ly α flux over $2 < z < 5$ using a sample of 6065 moderate-resolution quasar spectra drawn from SDSS DR7. More recently, Kamble et al. (2020) have used a similar modeling framework, except for allowing spectral diversity in the sample, to measure the effective optical depth in the Ly α forest using 40,035 quasar spectra from SDSS DR12. The higher opacity estimated at $z \lesssim 3$ by these authors compared to Becker et al. (2013) appears to require a 25% weaker ionizing background.

The release of DR14 (Abolfathi et al., 2018) of the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) provides us with a large statistical sample with which to improve over and extend previous results by incorporating the Ly β forest at rest-frame wavelengths between 978 and 1014 Å (Iršič et al., 2013). Because of the smaller cross-section, the study of Ly β absorption allows a better measurement of the equation of state of the IGM at higher overdensities, can break degeneracies when used in conjunction with Ly α absorption, and may yield stronger constraints on feedback processes and other nuances of IGM physics that affect higher density regions but leave low-density structures intact (Dijkstra et al., 2004).

In this Paper, we present a refined determination of the effective opacity, $\tau_{\text{Ly}\alpha}$, of the Ly α forest (excluding high column density absorbers), and a new measurement of the relative effective opacity, $\Delta\tau_{\text{Ly}\beta}$, of the Ly β forest, based on 27,008 quasar spectra at redshift $2.8 < z < 4.8$ from SDSS DR14. The plan is as follows. In §2.2 we describe the dataset and our method of measuring the Ly α and relative Ly β opacities in the

IGM. The main results of this project are presented in §2.3. We discuss our findings and compare them to previous measurements in §2.4.

2.2 DATA SET AND METHOD

In this section we describe how we construct composite spectra from the eBOSS DR14 dataset, and measure the normalized transmitted flux in the Lyman series forest at redshift $z = \lambda_{\text{obs}}/\lambda_n - 1$, where λ_{obs} is the observed wavelength at each pixel and $\lambda_n = 1215.67 \text{ \AA}$ for Ly α and 1025.72 \AA for Ly β . The final dataset consists of 27,008 quasar spectra at $2.8 < z < 4.8$ covering the wavelength interval $3800 - 9200 \text{ \AA}$ with a resolution ranging from $R = 1560$ at 3700 \AA to $R = 2270$ at 6000 \AA (blue spectrograph), and from $R = 1850$ at 6000 \AA to $R = 2650$ at 9000 \AA (red spectrograph). The average emission redshift of quasars in each of our 16 composite spectra and the number of spectra in each composite are listed in Table 2.1.

Rather than fit continua to individual spectra, we shall assume below that the mean unabsorbed spectral energy distribution (SED) of quasars is very similar regardless of redshifts, and that the difference between composites at different epochs essentially reflects the evolution of the mean transmitted flux in the forest. The validity of this premise is illustrated in Figure 2.1, where we plot the ratio between our composites and the reference composite spectrum at $z_c = 2.85$. Longward of Ly α , all ratios approach unity within the errors, as expected in the case of intrinsic SEDs that vary very little with redshift. In contrast, shortward of rest wavelengths 1215 \AA , one observes

Table 2.1: Composite Spectra. Columns give the number of quasars included in each composite, the minimum (z_{\min}) and maximum (z_{\max}) quasar redshifts in each bin, and the mean redshift (z_c) of the composite.

Number of QSOs	z_{\min}	z_{\max}	z_c
4640	2.8	2.9	2.85
4330	2.9	3.0	2.95
3995	3.0	3.1	3.05
3556	3.1	3.2	3.15
2830	3.2	3.3	3.25
2102	3.3	3.4	3.35
1189	3.4	3.5	3.45
991	3.5	3.6	3.55
1100	3.6	3.7	3.65
855	3.7	3.8	3.75
616	3.8	3.9	3.85
360	3.9	4.0	3.95
259	4.0	4.2	4.08
101	4.2	4.4	4.29
52	4.4	4.6	4.49
32	4.6	4.8	4.69

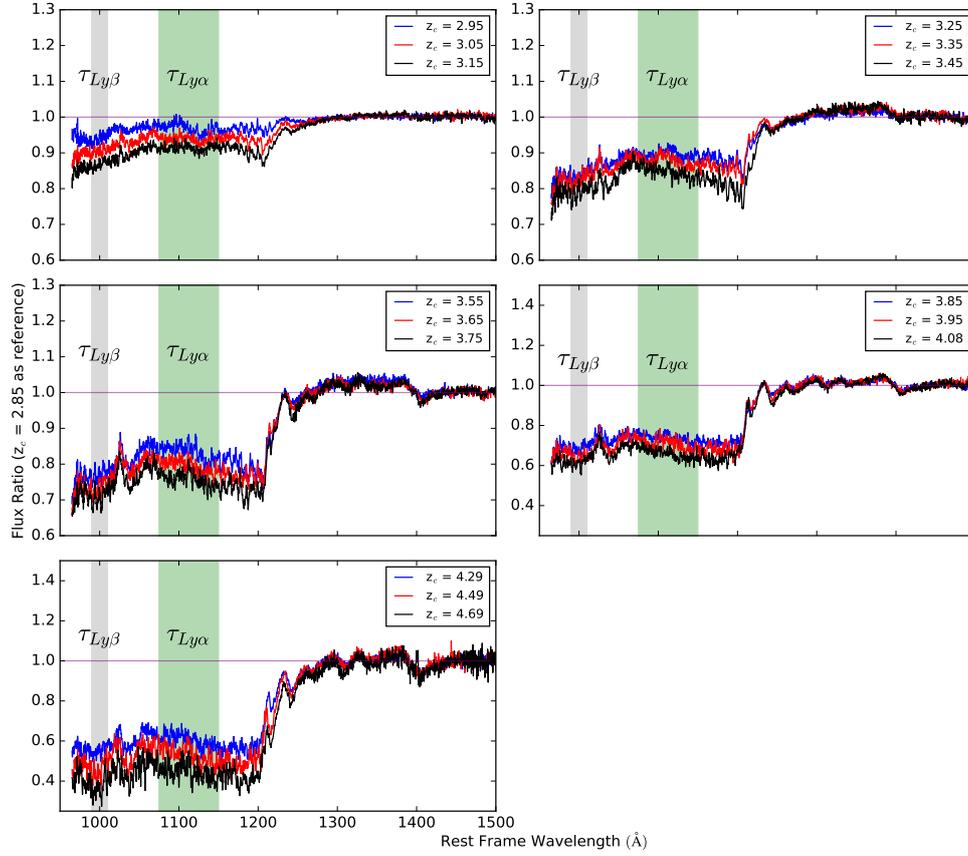


Figure 2.1: The composite quasar spectra used in this work, plotted versus rest-frame wavelength. The grey region denotes the wavelength range used for Ly β measurement and the green region denotes the wavelength ranges used for Ly α measurement. All spectra have been divided by our reference composite at $z_c = 2.85$ and the horizontal line marks unity flux ratio. Redward of Ly α , all ratios approach one, as expected in the case of almost identical intrinsic quasar SEDs.

strong and systematic evolution in the flux ratio indicative of an increasing IGM opacity with increasing redshift. When generating composite spectra in the quasar rest-frame, nearest-pixel values were adopted without interpolation (in order to reduce the covariance in the final composite), and combined by using an unweighted mean method (all sightlines contributing equally). Individual spectra were normalized at the rest-frame wavelength $1440 - 1460 \text{ \AA}$ where the quasar SED is relatively flat (without prominent emission and absorption lines).

2.2.1 Systematic Errors

The initial eSDSS DR14 spectra have systematic errors in the blue flux calibration and sky subtraction (Lan et al., 2018). Though corrected by the pipeline following Margala et al. (2016), DR14 spectra at wavelength $\lambda_{\text{obs}} < 4000 \text{ \AA}$ appear still slightly tilted compared to SDSS DR7 composites of the same quasar sample. This can be seen in Figure 2.2, where the ratio between eBOSS DR14 and SDSS DR7 composites are plotted at four representative redshifts ($z_c = 2.85, 3.25, 3.65, 4.29$). The DR14 spectra have systematically higher fluxes (median ratios always greater than one) across the relevant redshift range. The comparison between DR14 and DR7 composites also suggests a systematic trend of higher flux ratios towards increasing redshifts. We believe this bias is mainly due to differences in the sky subtraction between the two data releases.

In an attempt to reduce systematics, we have made different signal-to-noise (S/N) cuts to our sample spectra and compared the resulting DR14/DR7 composites in different redshift bins. We find that a $S/N > 2.2 \text{ pixel}^{-1}$ cut at rest frame wavelengths

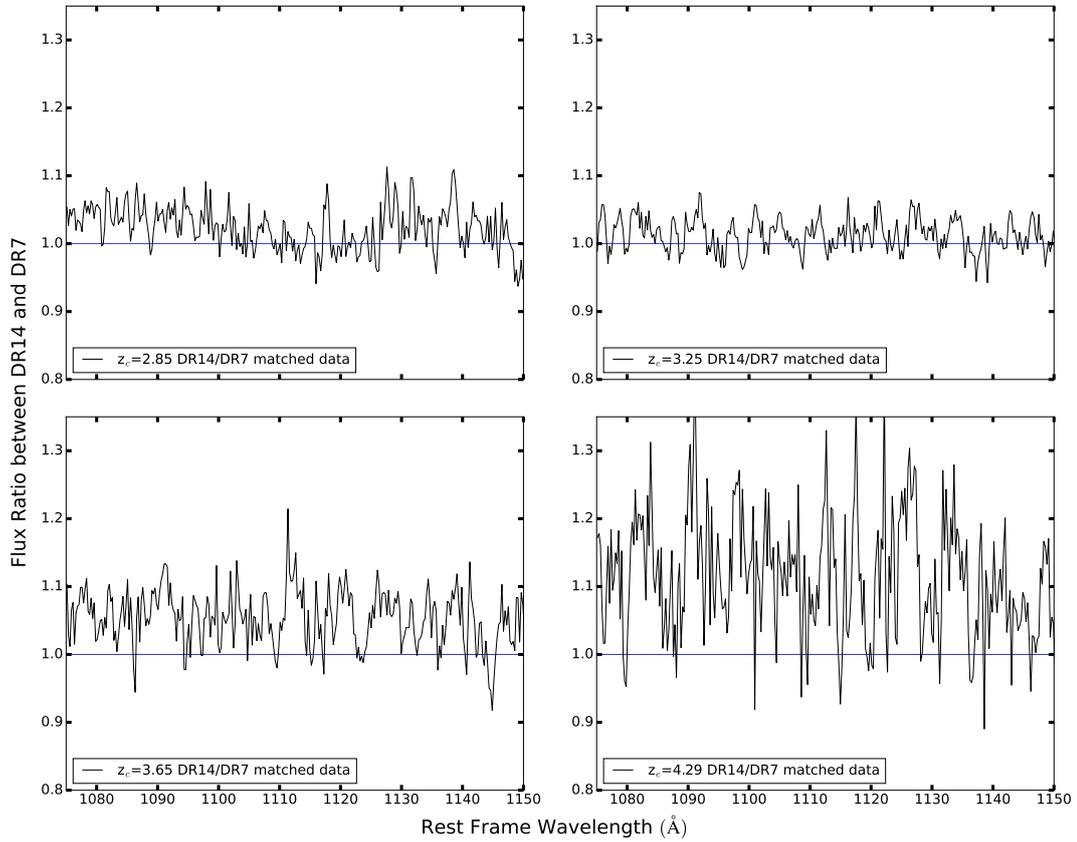


Figure 2.2: Flux ratio of composite spectra (same quasar sample) from SDSS DR14 and DR7 at four different redshifts. Note how DR14 composites have a systematically higher flux level compared to DR7, a bias that increases with increasing redshift.

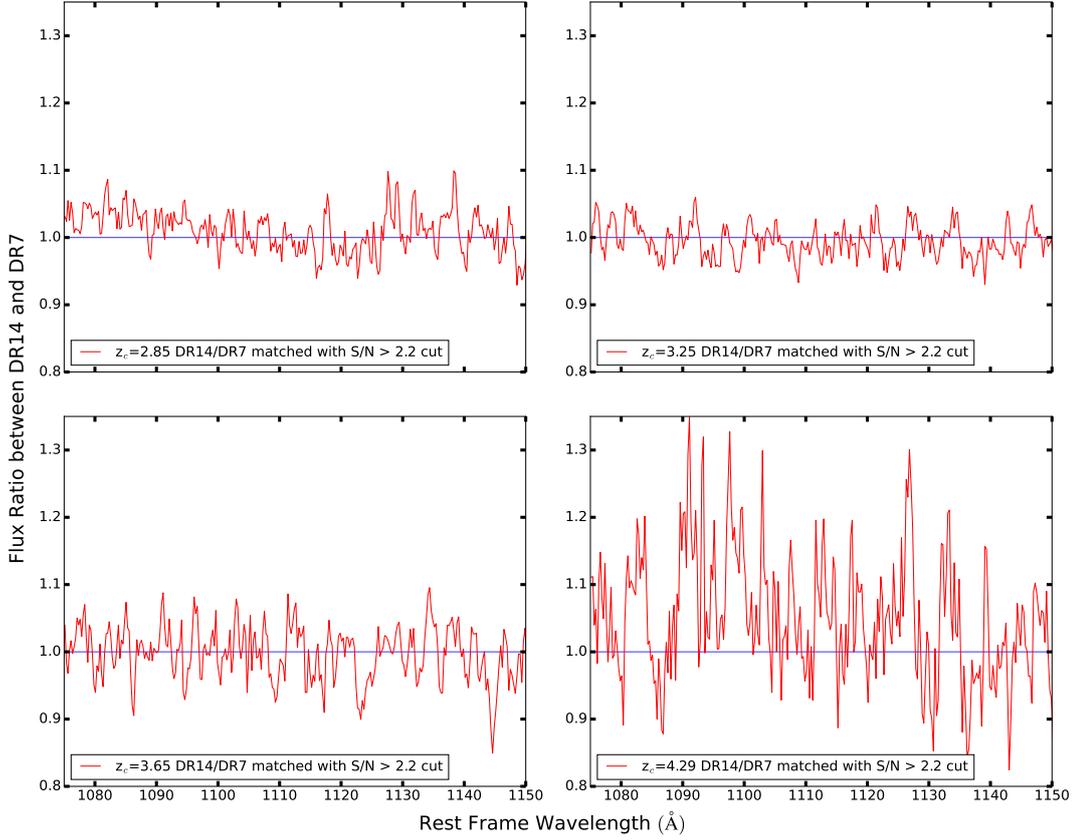


Figure 2.3: Same as Fig. 2.2 but after a $S/N > 2.2 \text{ pixel}^{-1}$ cut.

of $1440 - 1460 \text{ \AA}$ is optimal, as it greatly reduces systematics while only reducing the sample size by half. As shown in Figure 2.3, with this S/N cut the DR14 and DR7 composites now match to within 3% in most redshift bins.

2.2.2 Final Composite Spectra

The final sample consists of 27,008 individual quasar spectra, which are combined in 16 redshift bins of width equal to 0.1 for quasars with redshifts < 4.0 , and 0.2 at higher redshifts because of the decreasing number of targets. We exclude spectra with broad absorption line (BAL) features using the relevant flag in the quasar catalog (Pâris

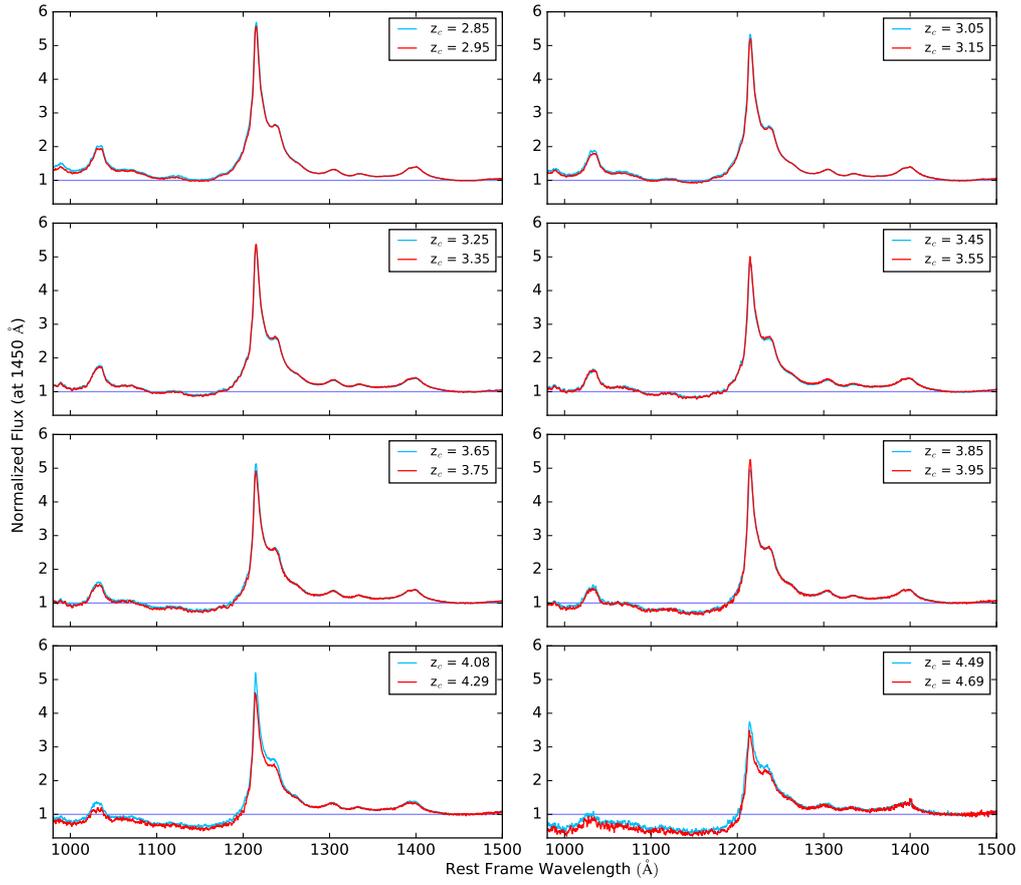


Figure 2.4: Final composite spectra from eBOSS DR14 data in different redshift bins.

et al., 2018), and mask the skylines region from 5570 to 5590 Å. The final composite spectra for eBOSS DR14 are shown in Figure 2.4.

2.3 Mean Transmitted Flux

2.3.1 Ly α Opacity

The mean transmitted flux at the rest-frame wavelength λ_c of a quasar composite with mean emission redshift z_c can be written as

$$F(z) = \frac{f_{\text{obs}}(\lambda_c, z_c)}{f_{\text{cont}}(\lambda_c)}, \quad (2.2)$$

where, in the Ly α forest, $1 + z_{\text{abs}} = \lambda_c(1 + z_c)/\lambda_\alpha$. Becker et al. (2013) measured the ratio of $F(z)$ at two different redshifts z_1 and z_2 by evaluating the ratio of fluxes between composites “c, 1” and “c, 2” at the same rest-frame wavelength,

$$\frac{F(z_1)}{F(z_2)} = \frac{f_{\text{obs}}(\lambda_c, z_{c,1})}{f_{\text{obs}}(\lambda_c, z_{c,2})}, \quad (2.3)$$

a relation that is satisfied when the mean continuum in the composites either does not evolve or has been appropriately corrected vs. redshift. They compute $F(z)$ up to a normalization factor, achieved by fitting a function that matches the observed flux ratios. The flux $F(z)$ is parameterized discretely in bins of $z = 0.1$, with the initial bin at $z = 2.15$, and computed as a fraction of the transmitted flux in that bin, i.e., $F(z)/F(z_c = 2.15)$.

Here, we adopt a modified approach. We assume, as in Becker et al. (2013), that the quasar unabsorbed SED does not change with redshift (see Fig. 2.1), but use Monte Carlo Markov Chain (MCMC) inference to perform a simultaneous fit to all of our DR14 composite spectra using a two component model: 1) the continuum represented by a cubic spline that varies with rest-frame wavelength; and 2) $\tau_{\text{tot}}(z) \equiv -\ln F(z)$, also

modeled as a cubic spline as a function of redshift. We emphasize that τ_{tot} includes all sources of opacity in the Ly α forest region including metals and optically-thick absorbers (discussed below).

Our empirical model for the shape of the quasar continuum uses 12 spline points spaced in wavelength as detailed in the Appendix (2.2). The cubic-spline form for τ_{tot} differs from the power-law function commonly adopted in the literature (but shown to be insufficient by Becker et al. 2013). After experimentation, we chose ten spline points to represent the total Ly α effective opacity as a function of redshift. The prior uniform distribution of the first spline point ranges from 0.2 to 0.8, and the range limits for the next points increase by 0.045 (i.e. 0.245 – 0.845 for the second point, 0.29 – 0.89 for the third, etc (see Table 2.3). Lastly, to yield an absolute estimate of τ_{tot} , the model is constrained by the high-resolution measurements of Faucher-Giguère et al. (2008a) at redshifts 2.0 – 2.8, including our estimate for optically-thick absorbers (§ 2.3.3) which were masked by Faucher-Giguère et al. (2008a). We also note that this is a small correction (≈ 0.015) at these redshifts.

Our Bayesian MCMC analysis centers on the rest-frame wavelength region 1075 – 1150 Å in the composite spectra for two reasons: 1) the ionization state of intergalactic gas is known to be enhanced near quasars (Bajtlik et al., 1988; Lidz et al., 2007), and restricting to wavelengths ≤ 1150 Å mitigates the impact of this “quasar proximity effect”; and 2) the lower bound avoids variation in the quasar continuum from OVI and FeII emission. The prior distributions of each of the 22 MCMC parameters are assumed to be uniform, and their ranges are listed in Table 2.4 of the Appendix.

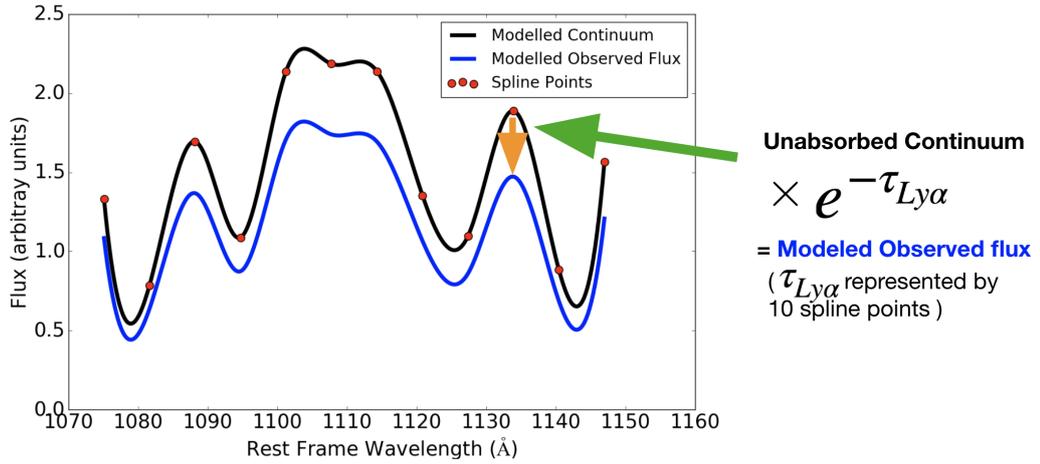


Figure 2.5: Our MCMC method. Red dots: spline points used to represent the unabsorbed quasar continuum. The mock observed flux (solid blue curve) is given by the unabsorbed continuum $\times \exp(-\tau_{Ly\alpha})$, where the function $\tau_{Ly\alpha}(z)$ is represented with 10 spline points. The mock flux is then fitted to the DR14 composite spectra.

The fitting algorithm adopts a multivariate normal log-likelihood distribution, with a covariance matrix generated as described in the Appendix. Our MCMC method is illustrated in Figure 2.5.

In the Appendix we show trace and kernel density plots for all the 22 spline parameters describing the quasar continuum (Fig. 2.14) and $\tau_{Ly\alpha}$ (Fig. 2.15). All parameters appear to have converged at the end of the MCMC sampling. Also in the Appendix, Figure 2.16 depicts the one- and two-dimensional projections of the posterior distributions for the quasar continuum. The best-fit MCMC model is compared to SDSS DR14 composite spectra as a function of redshift in Figures 2.6 and 2.7.

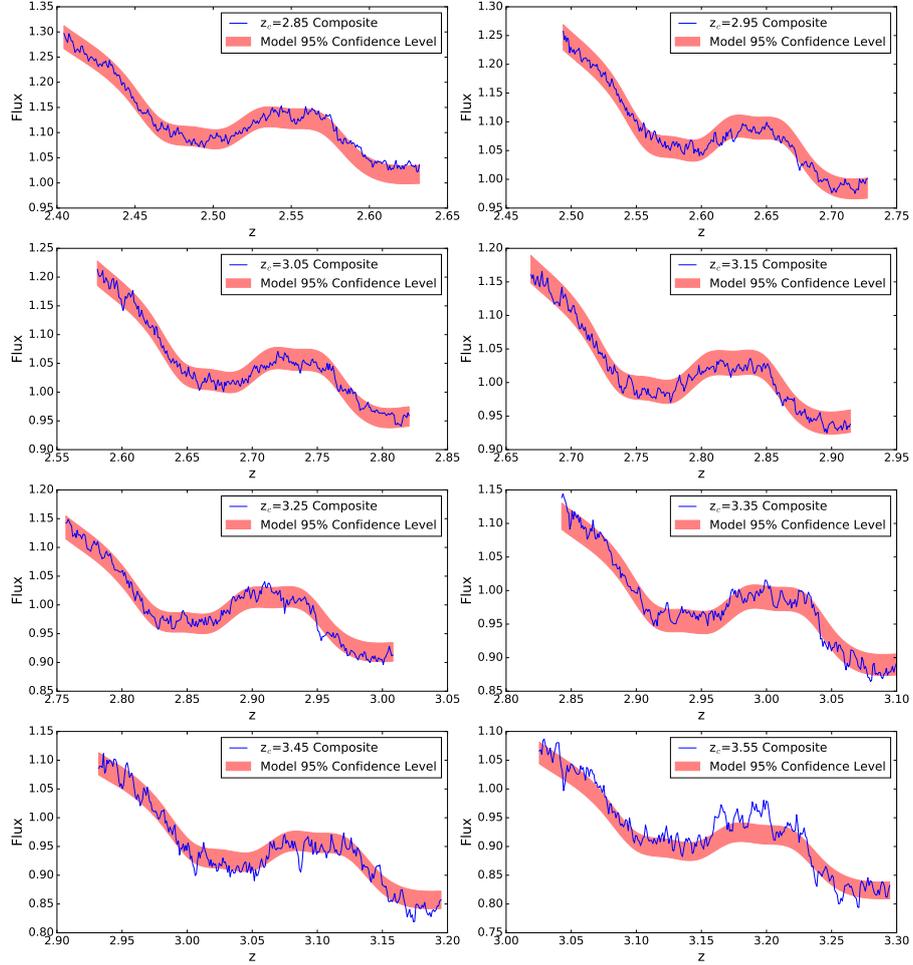


Figure 2.6: MCMC best-fit model to DR14 composite spectra (blue curves) in the redshift interval $z_c = 2.85 - 3.55$. The shaded region corresponds to the 95% confidence range of the transmitted flux marginalized over the posterior distribution.

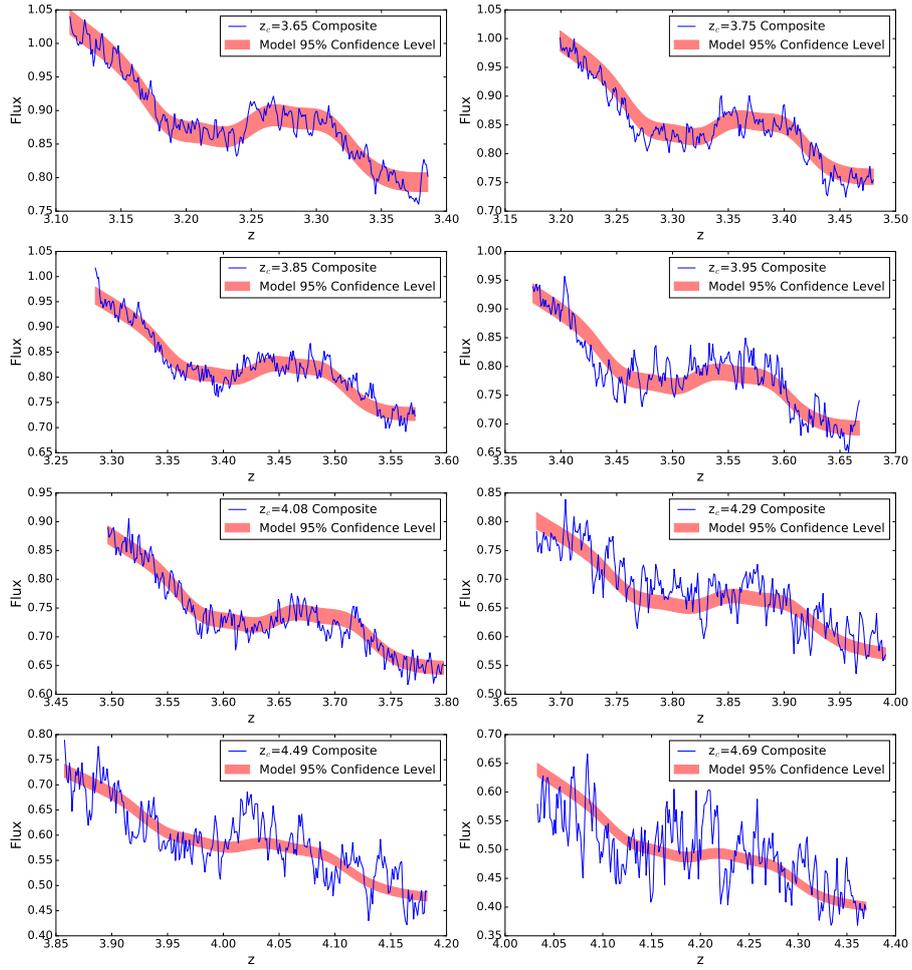


Figure 2.7: Same as the previous figure in the redshift interval $z_c = 3.65 - 4.69$.

2.3.2 Metal Line Contamination

Metal absorption in the Ly α forest is an important systematic component to be modeled for a precise measurement of $\tau_{\text{Ly}\alpha}$. Previous studies have shown that the flux decrement associated with the opacity of metal lines is nearly independent of redshift (Schaye et al., 2003; Kirkman et al., 2005; Becker et al., 2011). Therefore, it presents only an arbitrary constant that we choose not to model. By using the metal-corrected $\tau_{\text{Ly}\alpha}$ measurements of Faucher-Giguère et al. (2008a) at $z = 2.0 - 2.8$ to constrain our MCMC model combined with optically-thick absorbers, we effectively recover an estimate without metals absorption.

2.3.3 Correction for Optically Thick Absorbers

Our Ly α opacity measurements (corrected for metal absorption) include all diffuse intergalactic gas along the line of sight including optically-thick absorbers with $N_{\text{HI}} \geq 10^{17.2} \text{ cm}^{-2}$. These “Lyman-limit systems” (LLS) arise from highly overdense regions in the circumgalactic environment and are difficult to model analytically or in numerical simulations. It is standard practice, therefore, to subtract the contribution of LLS to $\tau_{\text{Ly}\alpha}$. Here, we follow Becker et al. (2013), and compute the integrated flux decrement from LLS as

$$D_{\text{LLS}} = \frac{1+z}{\lambda_\alpha} \int dN_{\text{HI}} \int db f(N_{\text{HI}}, z) W_0(N_{\text{HI}}, b), \quad (2.4)$$

where $f(N_{\text{HI}}, z)$ is the neutral hydrogen column density distribution, and W_0 is the rest-frame equivalent width. We use the $f(N_{\text{HI}}, z)$ distribution for $3.0 < z < 4.5$ simulated

by the package *pyigm* (Prochaska et al., 2014), and integrate Equation (2.4) over the column density range $N_{\text{HI}} = 10^{17.2} - 10^{22} \text{ cm}^{-2}$. We assume that gas column density and Doppler parameters b are statistically independent, and adopt a fixed value of $b = 24 \text{ km s}^{-1}$ (varying b in the range $20 - 30 \text{ km s}^{-1}$ has little impact on our results). The opacity from these optically-thick absorbers is therefore $\tau_{\text{LLS}} \equiv -\ln(1 - D_{\text{LLS}})$.

In our analysis, we first derive a raw measurement of τ'_{tot} from MCMC inference, and then subtract the contribution of optically thick absorbers,

$$\tau_{\text{Ly}\alpha} = \tau'_{\text{tot}} - \tau_{\text{LLS}} \quad (2.5)$$

where τ'_{tot} has been corrected for metal-absorption. The magnitude of the LLS correction is approximately 4-6%. The fully corrected Ly α opacity of the IGM is shown in Figure 2.8. Our flexible spline point model for τ_{tot} reproduces the redshift evolution of DR14 composite spectra more precisely than the standard power-law approximation.

2.3.4 Ly β Relative Opacity

Leveraging the large dataset of eBOSS DR14 and its blue wavelength coverage, we extend our analysis of the opacity of intergalactic hydrogen to include $\tau_{\text{Ly}\beta}$. As there are no existing measurements of quasar UV continuum levels in the Ly β forest, we can only provide here a relative determination, $\Delta\tau_{\text{Ly}\beta}$. We focus on the rest-wavelength range $990 - 1010 \text{ \AA}$ in the composite spectra, the region blueward of the Ly β emission line where the signal-to-noise is higher. As before, we measure the ratio of transmitted fluxes between all other composites and that at the reference redshift $z_c = 2.95$. In

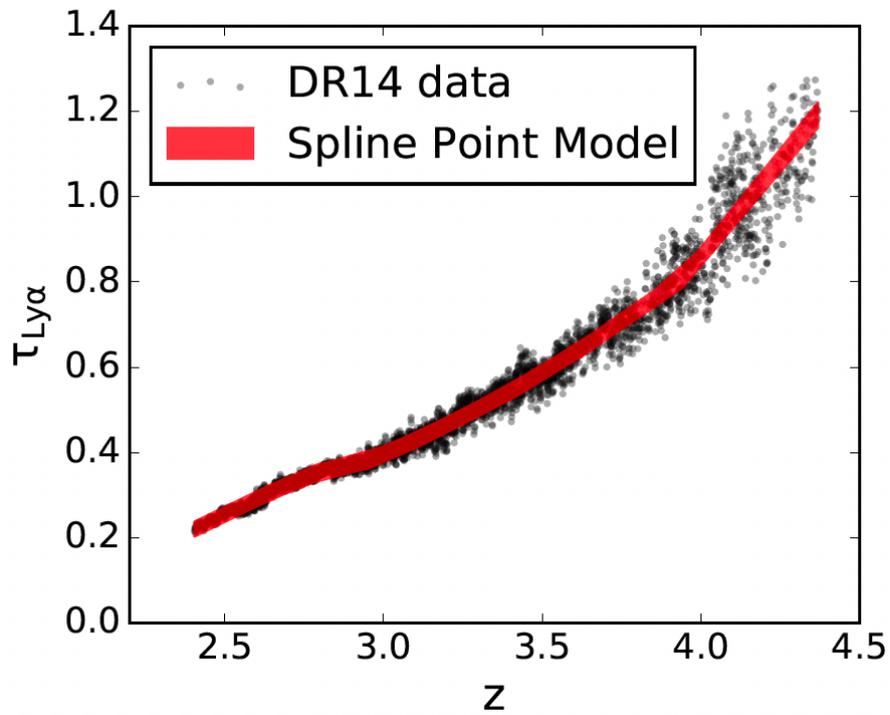


Figure 2.8: Observational determinations from DR14 composite spectra of the Ly α opacity in the redshift range 2.4 – 4.4. The best-fit from MCMC inference (corrected for metal line and LLS absorption) is shown as the shaded region corresponding to the 95% confidence interval of $\tau_{\text{Ly}\alpha}$ marginalized over the posterior distribution.

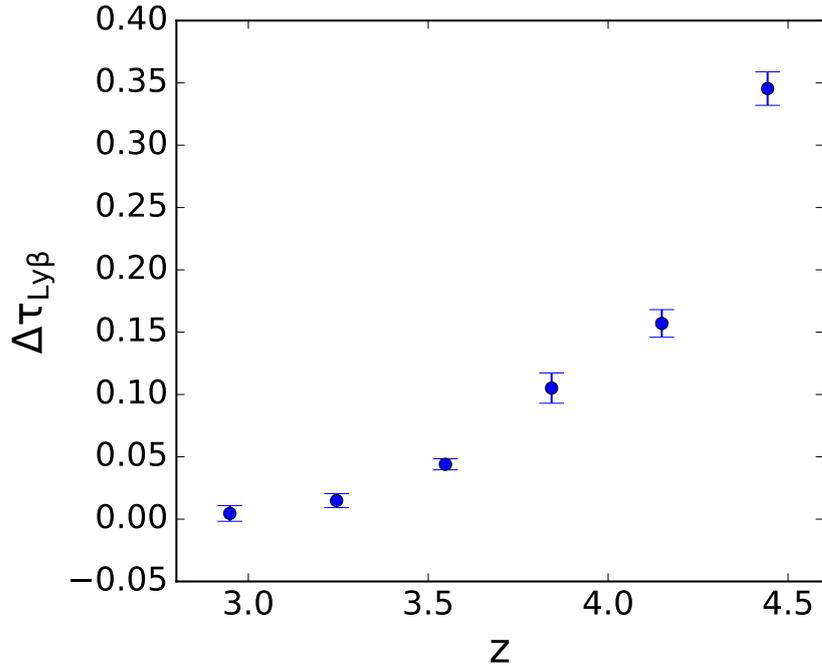


Figure 2.9: Relative Ly β effective opacity of the IGM from eBOSS DR14 data. The ten redshift bins cover the range $2.8 \leq z \leq 4.6$.

each composite, the transmission through the Ly β forest is determined by the combined effective optical depth,

$$\tau_{\alpha+\beta}(\lambda_{\text{obs}}) = \tau_{\text{Ly}\alpha}(z_{\alpha}) + \tau_{\text{Ly}\beta}(z_{\beta}), \quad (2.6)$$

for the overlapping spectral region, with $z_{\alpha} = \lambda_{\text{obs}}/\lambda_{\alpha} - 1$ and $z_{\beta} = \lambda_{\text{obs}}/\lambda_{\beta} - 1$. Therefore, the flux ratio between a pair of composite spectra is sensitive to the difference $\Delta\tau_{\text{Ly}\alpha} + \Delta\tau_{\text{Ly}\beta}$ evaluated at λ_{obs} . We first calculate $\tau_{\alpha+\beta}$, then attain a value for $\Delta\tau_{\text{Ly}\alpha} + \Delta\tau_{\text{Ly}\beta}$ evaluated at λ_{obs} relative to $\tau_{\text{Ly}\beta}(z_c = 2.95)$ from equation (3). We subtract from it the foreground $\Delta\tau_{\text{Ly}\alpha}$ and finally determine $\Delta\tau_{\text{Ly}\beta}(z)$. The evolving relative Ly β opacity in ten redshift bins is shown in Figure 2.9.

2.4 DISCUSSION

We have presented new measurements of the mean transmitted UV flux through cosmic hydrogen using 27,008 quasar spectra from eBOSS DR14, and applied MCMC inference to produce a spline-piecewise fit of the effective Ly α (corrected for metal lines and hydrogen optically thick absorption) and relative Ly β optical depths. In Figure 2.10 we plot our derived intergalactic Ly α opacity along with previous SDSS-based results by Becker et al. (2013), Dall’Aglio et al. (2009), and Pâris et al. (2011), as well as with the high-resolution determinations made by Faucher-Giguère et al. (2008a) – and used to constrain our MCMC model at $z < 2.8$. All of the works yield similar $\tau_{\text{Ly}\alpha}$ estimations at $z < 3$ and our new measurements lie between the locus of data at higher redshifts.

Compared to Becker et al. (2013), the $\approx 10\%$ differences between the two determinations at $z = 3.4 - 4$ may be associated with three different factors. Firstly, the difference in the sky subtraction for the DR7 and DR14 data sets. However, because Becker et al. (2013) estimate that sky subtraction affects only $\approx 5\%$ of the results, we expect this is a minor effect. Secondly, the DR14 sample analyzed here is considerably larger than the DR7 sample. Also, the systematic on the $\tau_{\text{Ly}\alpha}$ measurement caused by the Quasar SED shape discussed in §APPENDIX D may contribute to $\approx 6 - 8\%$ of the measurement difference. Taking the three factors into account, our results are consistent with Becker et al. (2013). Dall’Aglio et al. (2009) combined a redshift-dependent global correction from mock simulated spectra with a local spline interpolation to fit

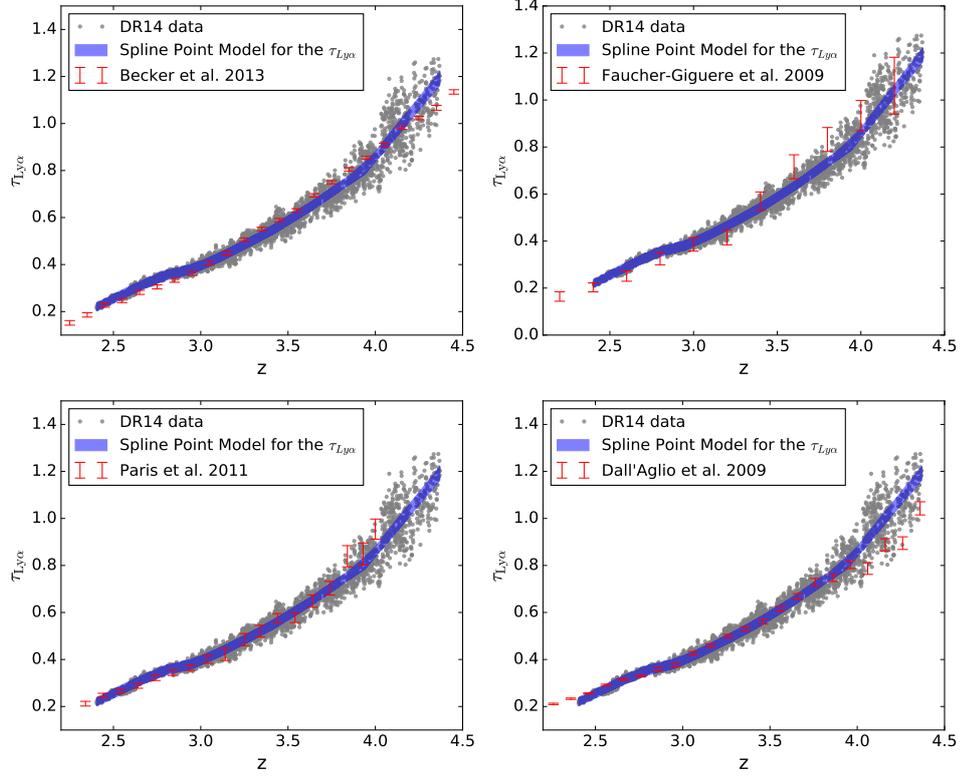


Figure 2.10: The comparison between $\tau_{\text{Ly}\alpha}$ measurement from the best fitted spline evolution for $\tau_{\text{Ly}\alpha}$ (blue region marks 3-sigma confidence level for the fitted $\tau_{\text{Ly}\alpha}$ model) from the MCMC modeling plotted with the DR14 data after correcting metal line absorption and optically thick absorbers and the measurement in Becker et al. (2013); Dall’Aglio et al. (2009); Faucher-Giguère et al. (2008a); Pâris et al. (2011). Upper left: The comparison between our results and measurement in Becker et al. (2013). Upper right: The comparison between our results and measurement in Faucher-Giguère et al. (2008a). Lower left: The comparison between our results and measurement in Pâris et al. (2011). Lower right: The comparison between our results and measurement in Dall’Aglio et al. (2009).

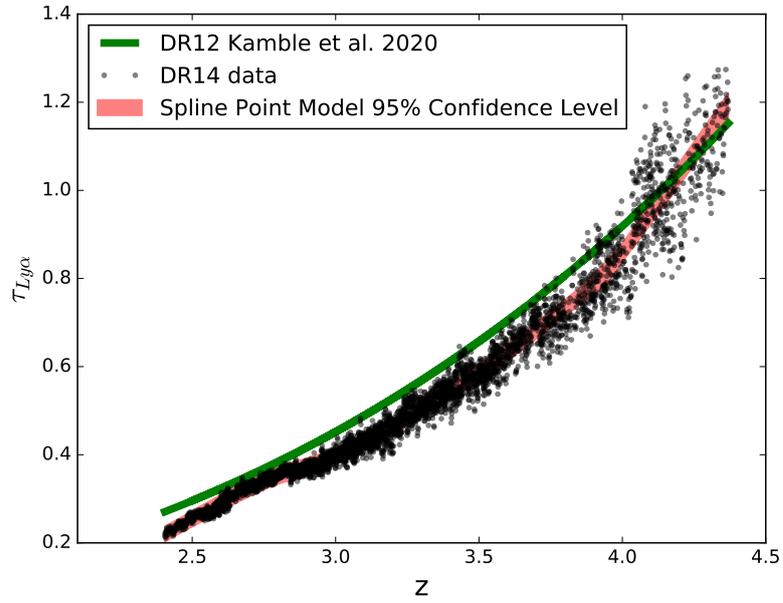


Figure 2.11: Comparison between $\tau_{\text{Ly}\alpha}$ measurement from the best fitted spline evolution for $\tau_{\text{Ly}\alpha}$ (red region marks 3-sigma confidence level for the fitted $\tau_{\text{Ly}\alpha}$ model) from the MCMC modeling plotted with the DR14 data after correcting metal line absorption and optically thick absorbers and the measurement in Kamble et al. (2020) (green curve).

the continuum of 1733 individual quasar spectra. At the medium resolution of SDSS redshift-dependent corrections must be larger, and will be more significant at lower redshifts. There is a small discrepancy between Dall’Aglia et al. (2009) measurement and our results at $z > 4.0$. The reasons for this slight disagreement may be as follows. While fitting the continuum in the medium resolution spectra, the fitting parameters will also have larger error and systematics. Also, in Dall’Aglia et al. (2009), the mock spectra used for fitting the continuum are based on a fixed number density, column density and Doppler parameter distributions of absorbers. This also leads to the systematic effects causing some discrepancy in the evolution of the Ly α opacity.

Finally, a more recent paper by Kamble et al. (2020) utilizes BOSS DR12 data to measure the Ly α opacity in a quasar sample that is similar to ours. These authors model $\tau_{\text{Ly}\alpha}$ as a power-law and fit for the continuum normalization and then marginalize over the latter in seven bins of quasar samples. A comparison between their results and ours shows (Fig. 2.11) a systematically higher optical depth. The majority of the difference originates from their decision not to constrain their model by high-resolution estimates of $\tau_{\text{Ly}\alpha}$ (e.g. Faucher-Giguère et al., 2008a). We also note that while they ignored quasars with foreground damped Ly α systems, their $\tau_{\text{Ly}\alpha}$ do include opacity from LLS at lower column densities. We estimate, however, that this is a relatively small correction ($< 10\%$). Also, our measurement shows a small dispersion with Kamble et al. (2020) at higher redshift. This may be due to the difference in our methods to attain the $\tau_{\text{Ly}\alpha}$ measurement. Our measurement uses a spline point model for the $\tau_{\text{Ly}\alpha}$ measurement while Kamble et al. (2020) assume a power law distribution for the $\tau_{\text{Ly}\alpha}$,

leading to the fact that the discrepancy tends to get small in the high redshift regime. In §APPENDIX D, we also investigate the effect caused by different spectral indexes (similar to the spectral index groups in Kamble et al. (2020)). Taking into account the systematic discussed in §APPENDIX D, our discrepancies with Kamble et al. (2020) are largely relieved.

A recent publication from the large spectroscopic datasets (the Dark Energy Spectroscopic Instrument (Turner et al., 2024) also perform a measurement on $\tau_{Ly\alpha}$ by using a Convolutional Neural Network to reconstruct the Lyman alpha continuum. Our results are consistent with the DESI results within 95% confidence level taking into account the systematic mentioned in §APPENDIX D.

2.5 Conclusion

We measure the $\tau_{Ly\alpha}$ and relative $\tau_{Ly\beta}$ using a novel method from the eboss DR 14 spectroscopic data. Our results are consistent with the majority of $\tau_{Ly\alpha}$ measurements in previous literature. In the future, as we turn to the next generation of large spectroscopic datasets (the Dark Energy Spectroscopic Instrument; DESI/DESI-II DESI Collaboration et al., 2022a), it is evident that quasar sample size is no longer the limiting factor in estimating $\tau_{Ly\alpha}$. Instead, the accuracy and precision is limited by systematic error in data collection (e.g. fluxing) and the assumption that quasar continua are invariant with redshift. Nevertheless, opportunities to extend to higher order Lyman series lines will materialize, both in surveys like DESI but also in higher-resolution

samples.

2.6 Appendix

2.6.1 Bayesian Inference Parameters and Priors

We use a MCMC process to sample the best fitted parameters. All the priors for the 22 parameters are uniform. The location for the first 12 spline points (continuum model) is listed in 2.2. The location for the 10 spline points for the τ_α is listed in 2.3. The range of the value for the uniform prior is listed in 2.4.

2.6.2 Covariance Array

The uncertainty in the composite spectra does not follow standard error propagation because of correlations in the underlying continuum of each quasar. To assess the uncertainty and these correlations, we generated a covariance matrix with standard bootstrap resampling techniques. In each iteration, a new composite was generated from a number \mathcal{N} of randomly drawn individual spectra equal to the total number \mathcal{N} of spectra in that redshift bin, allowing for duplication. This process was repeated 10,000 times to generate the covariance matrix of that composite. The covariance matrices showing correlations in the error budget of the transmitted flux are plotted in Figures 2.12. The structure of the covariance matrix is symmetric, and the variance of the transmitted flux (diagonal elements) dominates over the covariance between all possible flux pairs (off-diagonal terms). The covariance of the off-diagonal terms also increase

Table 2.2: Continuum spline points in rest-wavelength space.

Spline Point	Wavelength λ_c (\AA)
0	1075
1	1082
2	1088
3	1095
4	1101
5	1108
6	1114
7	1121
8	1127
9	1134
10	1140
11	1147

Table 2.3: Ly α opacity (corrected for metal lines but not for LLS) spline points in redshift space (see final_results for details).

	Redshift	Mean Value	hpd 2.5	hpd 97.5
Spline Point 12	2.50	0.27	0.25	0.29
Spline Point 13	2.64	0.32	0.31	0.34
Spline Point 14	2.80	0.38	0.36	0.39
Spline Point 15	2.95	0.40	0.38	0.42
Spline Point 16	3.11	0.46	0.44	0.47
Spline Point 17	3.27	0.52	0.5	0.54
Spline Point 18	3.42	0.58	0.56	0.60
Spline Point 19	3.63	0.68	0.67	0.70
Spline Point 20	3.94	0.85	0.83	0.87
Spline Point 21	4.20	1.09	1.07	1.12

Table 2.4: Ranges of the prior uniform distributions for the 22 spline points used in our MCMC inference.

Parameter	Range
Continuum spline points 0-11	0.6–2.8
Spline Point 12	0.2–0.8
Spline Point 13	0.245–0.845
Spline Point 14	0.29–0.89
Spline Point 15	0.335–0.935
Spline Point 16	0.38–0.98
Spline Point 17	0.47–1.07
Spline Point 18	0.515–1.115
Spline Point 19	0.56–1.16
Spline Point 20	0.605–1.205
Spline Point 21	0.65–1.25

with respect to the redshift.

2.6.3 Convergence of MCMC Algorithm

All parameters appear to have converged at the end of the MCMC sampling. Figures 2.14 and 2.15 show kernel density estimation of the posteriors and the samples of the Markov chain for all 22 spline parameters describing the quasar continuum and $\tau_{\text{Ly}\alpha}$.

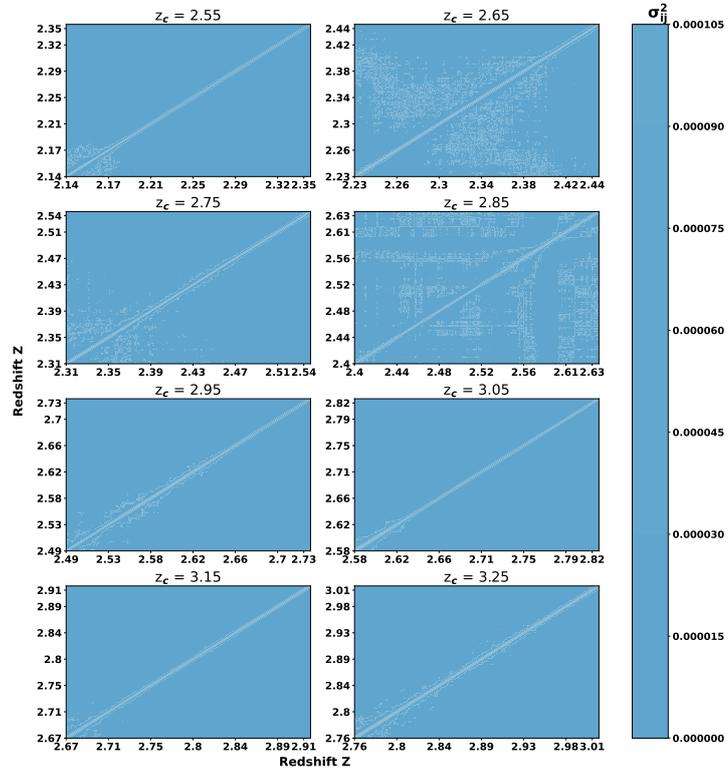


Figure 2.12: The Covariance matrices for DR14 composite spectra in the redshift interval $z_c = 2.85 - 3.55$. All the covariance plot show similar features (symmetric along the diagonal axis). The covariance in the off-diagonal terms are more significant in the higher redshift bins.

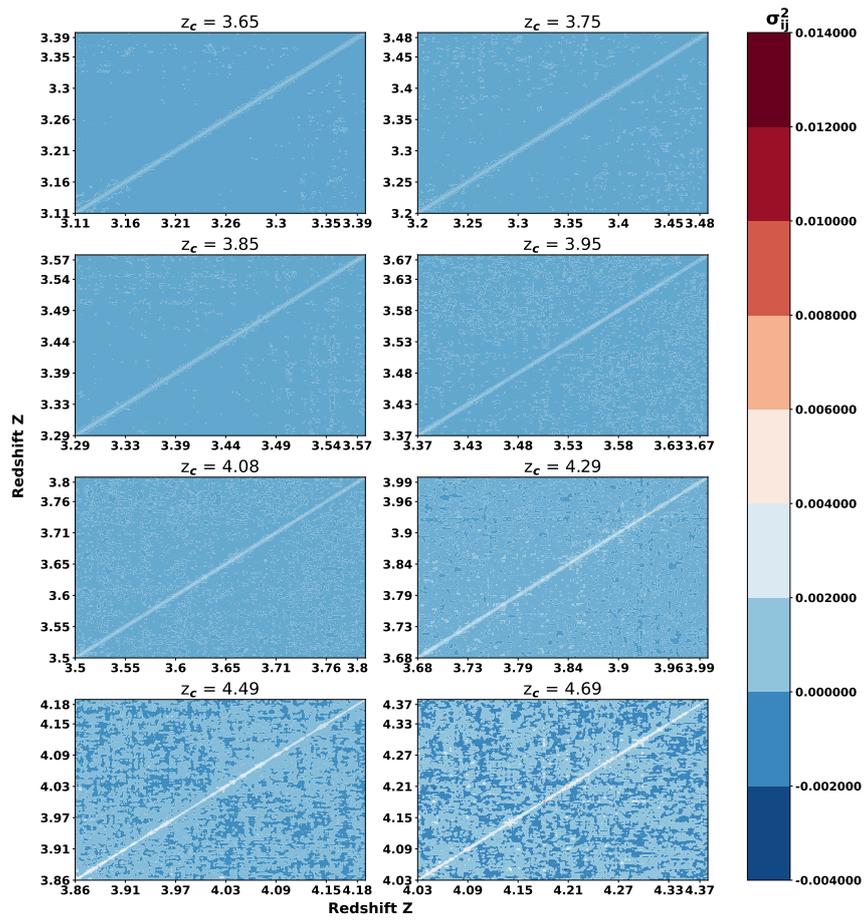


Figure 2.13: Covariance matrices for DR14 composite spectra at $z \geq 3.65$.

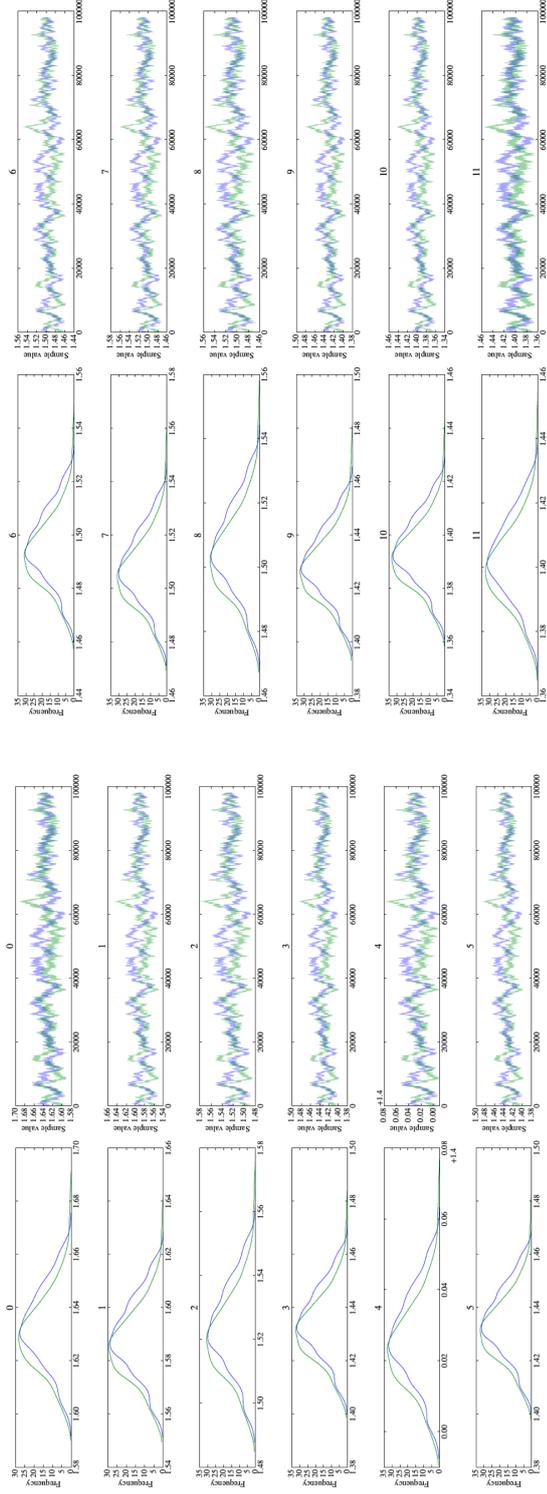


Figure 2.14: MCMC final converged traces and kernel densities for the spline points used to represent the quasar continuum SED. The x -axis shows normalized flux value (left columns in each panel) and the number of the MCMC samples (right columns in each panel). The left panel depicts the smoothed histogram (using kernel density estimation) of the posteriors (left columns) and the samples of the Markov chain (right columns) of the 0-5 spline points and the right panel is same plot for the 6-11 spline points.

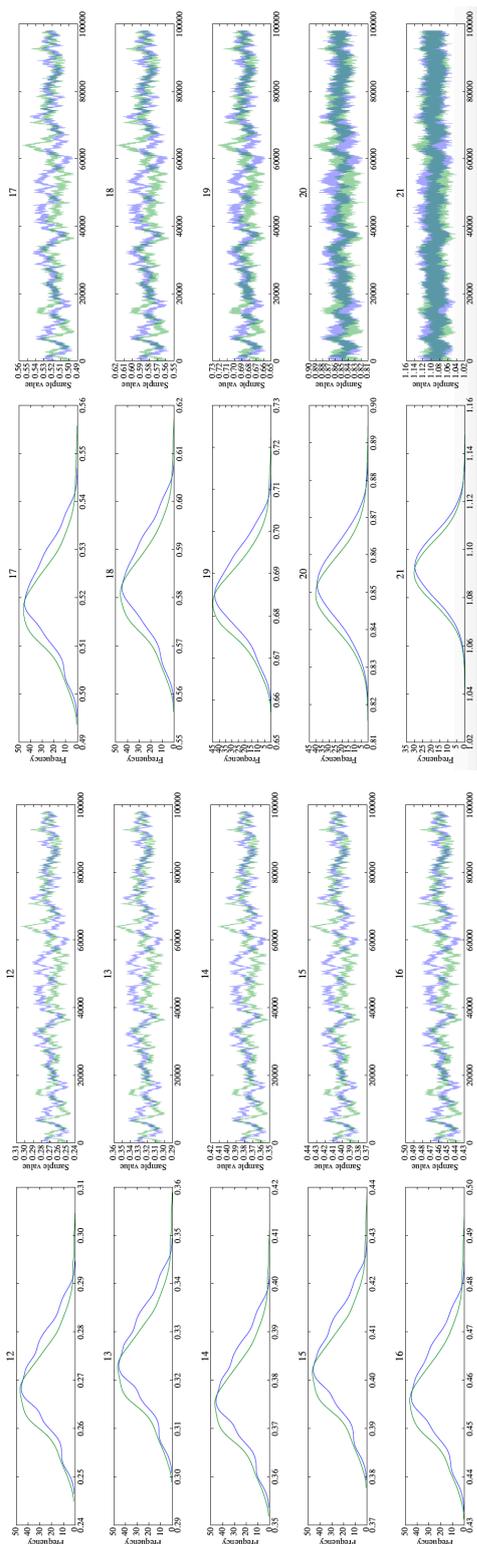


Figure 2.15: MCMC final converged traces and kernel densities for the spline points used to represent the $\text{Ly}\alpha$ opacity vs. redshift. The x -axis shows the value of $\tau_{\text{Ly}\alpha}$ (left columns in each panel) and the number of the MCMC samples (right columns in each panel). The left panel depicts the smoothed histogram (using kernel density estimation) of the posteriors (left columns) and the samples of the Markov chain (right columns) for the first five spline points describing the function $\tau_{\text{Ly}\alpha}(z)$ (12-16 traces), and the right panel shows the next five spline points (17-21 traces).

2.6.4 Systematic Errors Estimation from quasar SEDs

We also perform an estimation of systematic errors in the $\tau_{Ly\alpha}$ measurement caused by different quasar SED shapes. We first fit a power law $f = \alpha\lambda^\beta$ to the shape of the quasar SEDs, where f is the flux of the quasar spectra and β is the spectral index. We then group those quasar spectra by spectral index as in Kamble et al. (2020). We adopt three spectral index bins: $-2.78 < \beta < -2.12$, $-2.12 < \beta < -1.46$, $-1.46 < \beta < -0.8$. We then recompute the composite spectra at all redshifts for the three different spectral index bins and rerun the MCMC process with our spline point method. The measurements of $\tau_{Ly\alpha}$ for the 3 bins are presented in Figure 2.17. From the differences between these $\tau_{Ly\alpha}$ measurements and those from the full sample, we estimate that the uncertainty from the systematics caused by the difference in the spectral index of the quasar SEDs is 6 – 8% at $z < 4$ and 10 – 12% at $z > 4.0$.

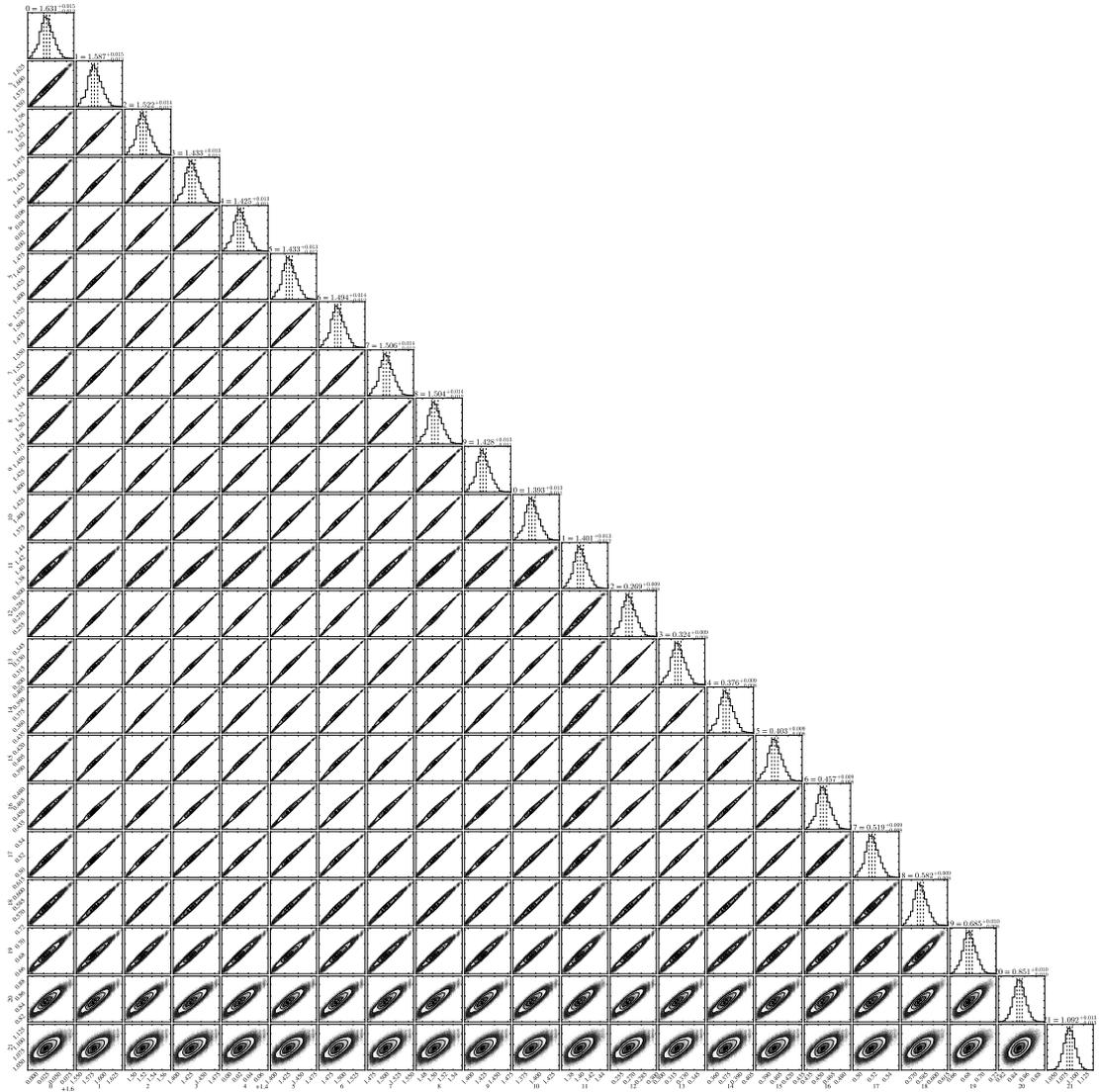


Figure 2.16: Final corner plot for all the parameters we used to represent the continuum and $\tau_{\text{Ly}\alpha}(z)$ in the MCMC fitting.

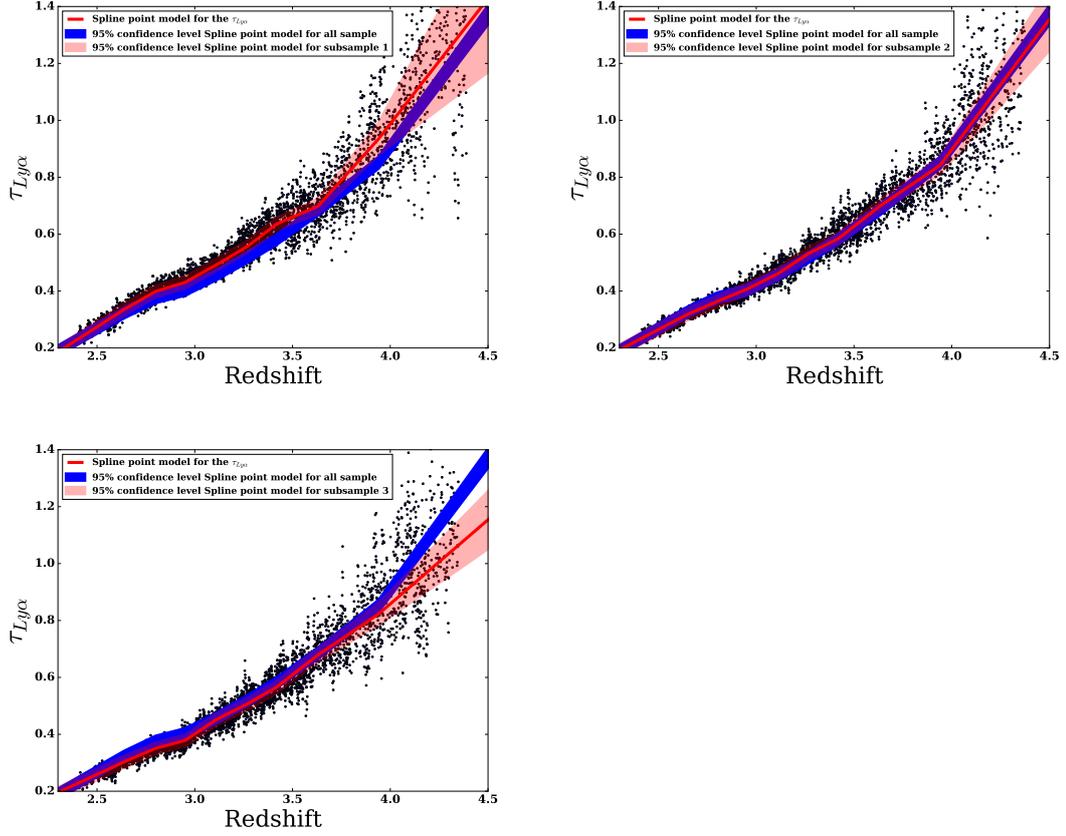


Figure 2.17: $\tau_{Ly\alpha}$ measurement for quasar spectra with three different spectral index ($-2.78 < \beta < -2.12$, $-2.12 < \beta < -1.46$, $-1.46 < \beta < -0.8$). The upper left panel shows the comparison between the $\tau_{Ly\alpha}$ measurement for quasar spectra with $-2.78 < \beta < -2.12$ and the $\tau_{Ly\alpha}$ measured with all the spectra in the sample. The 95% confidence level of the $\tau_{Ly\alpha}$ measurement for quasar spectra with $-2.78 < \beta < -2.12$ is labeled in red and the 95% confidence level of the $\tau_{Ly\alpha}$ measured from all sample is labeled in blue region. The DR14 quasar data with $-2.78 < \beta < -2.12$ is labeled in dark data points. The upper right and lower panels are similar figures for quasar spectra with $-2.12 < \beta < -1.46$ and $-1.46 < \beta < -0.8$.

Chapter 3

Building DLA catalog for the DESI

QSO Spectra

Quasar absorption systems are crucial for studying galaxy evolution and large scale structure. The absorption of Ly α opens a wide window to study gas properties inside and around galaxies as well as neutral gas on large scales in the IGM (Wolfe et al., 1986; Rauch, 1998; DESI Collaboration et al., 2016b; Cai et al., 2017; Pérez-Ràfols et al., 2018). Among these absorption systems, damped Ly α systems (DLAs) (usually this refers to objects with neutral hydrogen column densities $N_{\text{HI}} > 2 \times 10^{20} \text{cm}^{-2}$), serve as significant reservoirs of atomic hydrogen and provide valuable insights into galaxy evolution and large scale structure (Wolfe et al., 2005; Prochaska & Wolfe, 1997; Zafar et al., 2013). For example, DLAs are crucial for examining the circumgalactic medium (CGM) around high-redshift galaxies ($z > 2$).

Historically, DLA detection involved manual inspection of spectral data Prochaska

& Herbert-Fort (2004). In recent decades, automated detection techniques, including the Gaussian Process Algorithm and Convolutional Neural Networks (CNNs), have been employed Garnett et al. (2017); Ho et al. (2020). These novel methods significantly increase the efficiency of DLA detection while maintaining accuracy comparable to visual inspection.

Building a DLA catalog is essential for studying the properties of galaxies and their evolution. Investigating the rate at which galaxies convert their gas into stars is one important approach to study the assembly of galaxies. Therefore, DLAs are key to understanding the gas reservoir for galaxy formation and evolution, and thus shed light on the role of neutral gas in galaxy formation and evolution Prochaska & Wolfe (2009); Noterdaeme et al. (2012).

Moreover, such a catalog is crucial for addressing a significant challenge in Baryon Acoustic Oscillation (BAO) measurements. BAO studies are fundamental for understanding the expansion history of the universe. The BAO signal, imprinted in the distribution of galaxies and intergalactic gas, serves as a "standard ruler" for cosmology. By measuring the large-scale structure of the universe, researchers can infer the rate of expansion at different epochs, providing crucial insights into the nature of dark energy. BAO measurements using the Ly α forest have proven to be particularly effective for high-redshift studies ($z > 2$). These measurements rely on detecting the subtle imprints of BAO in the distribution of Ly α absorbers in quasar spectra (e.g., Bautista et al., 2017). Optically thick absorbers like DLAs introduce bias into the analysis, making an accurate DLA catalog essential for mitigating these contaminants.

As discussed in the thesis introduction, the Dark Energy Spectroscopic Instrument (DESI) survey, covering 14,000 deg², will substantially expand high-redshift QSO observations, providing extensive data for Ly α forest studies and BAO measurements. The 2.4 million QSO spectra observed by DESI will enhance our ability to trace intergalactic gas distribution at $z > 2$. To minimize the bias and systematics introduced by DLAs, it is essential to build a DLA catalog in DESI. This catalog will improve the accuracy of BAO studies using optically thin absorbers, enabling a more precise understanding of the large-scale structure of the universe.

By integrating state-of-the-art detection techniques and extensive survey data, the creation of a detailed DLA catalog will not only advance our knowledge of galaxy evolution but also refine the methods used to map the cosmos through BAO measurements.

This chapter outlines my contribution to the development of a Convolution Neutral Network based DLA finder and the DLA catalog for the DESI survey in Wang et al. (2022). In this paper, we optimized and applied the three-layer CNN model (see 3.1 for the network structure) from Parks et al. (2018) to identify and characterize the damped Ly α systems. The two physical properties (N_{HI} , redshift z) of the damped Ly α systems (DLAs) for mock spectra from the Dark Energy Spectroscopic Instrument (DESI) are the major physical properties we constrained in the algorithm. Our CNN model achieves over 99% classification accuracy for spectra with signal-to-noise ratios (S/N) above 5/pixel and over 97% for spectra with S/N \sim 1/pixel. We also examined the effect of different DLA catalogs (the DLAs from mock catalogs, DLAs detected

by Gaussian process) on the measurement of baryon acoustic oscillations (BAO). The difference in the cosmological fitting parameter for BAO is less than 0.61% compared to an analysis using perfect DLA masking. This discrepancy is smaller than the 1.7% statistical error for the first year of DESI.

I completed the preliminary work for this project by developing a Python package that organizes DESI mock spectra into a format suitable for convolutional neural networks (CNNs) and ran the network on the arranged data. The structure of the DLA finder from DESI original mock spectra to the final output of the ID, N_{HI} and redshift of the DLAs is shown in 3.2. First the flux and error with the meta data of the input DESI mock spectra are extracted. The CNN network slides through each individual spectra and identifies existing DLAs. The final output of each run is a DLA catalog including the ID, redshift and column density N_{HI} of each detected DLAs.

I also ran some preliminary plural in order to improve the performance of the CNN model on the mock spectra. The original CNN model was trained on the SDSS/BOSS spectra. I created the confusion matrix to estimate the completeness and purity of our DLA catalog based on the SDSS training set. The purity is defined as the number of true positive divided by the total number of true positive and false positive. The completeness is defined by the total number of the true positives divided by the total number of true positives and true negatives. The original performance of the CNN model is illustrated in 3.3. This indicates that the original CNN Network trained on the SDSS/BOSS spectra achieved a accuracy around 70 % with a classification confidence level around 0.6, indicating that refined training is in order to further improve the CNN

performance. After the preliminary testing, I helped with tutoring students who are the primary authors on the paper and also ran some tests on retraining the CNN Network on the DESI Mock spectra with uniform N_{HI} and different S/N. Other details for the CNN DLA finder and the BAO-fitting result by masking the DLAs using the DLA catalog can be found in Wang et al. (2022).

The CNN finder algorithm in our study is designed and used to mask the DLAs in the spectra with signal-to-noise (SNR) larger than 3 for the BAO measurement (DESI Collaboration et al., 2024). It is successfully applied in one of the most influential scientific output of the DESI year one data, the BAO measurement from the Ly α forest (DESI Collaboration et al., 2024).

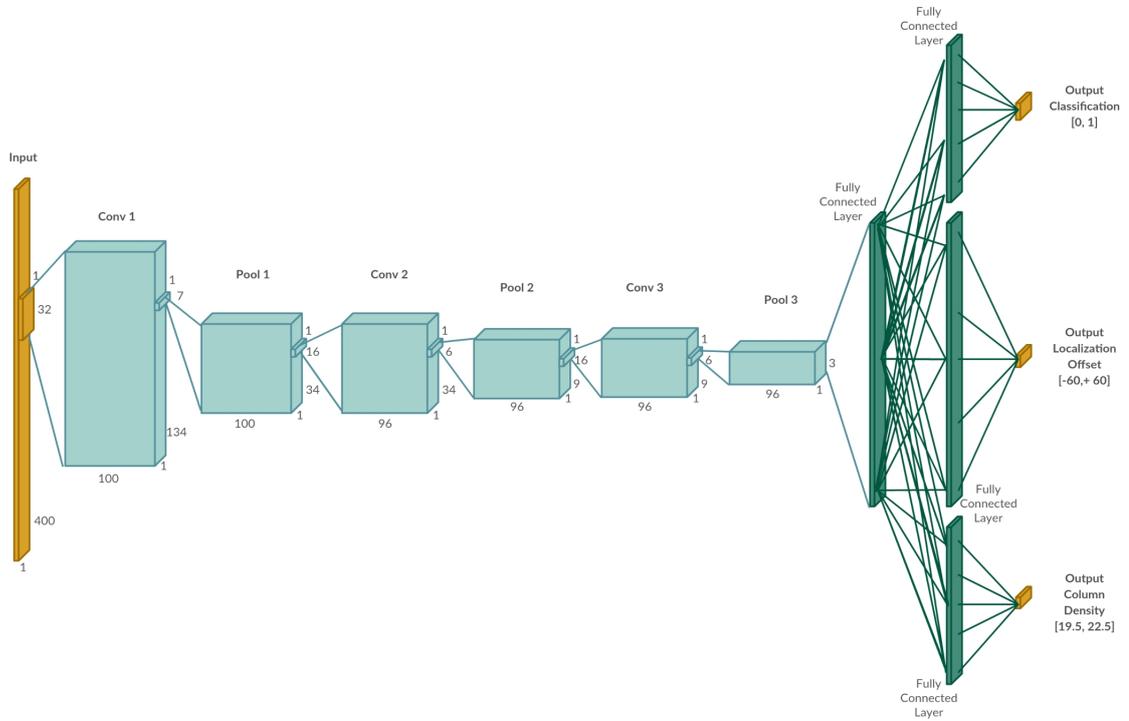


Figure 3.1: The three-layers CNN architecture used in our project (Fig. 6 in Wang et al. (2022)) Parks et al. (2018); Wang et al. (2022), casting DLAs as a 1D image problem. There are three convolutional layers, three pooling layers, one fully connected layer, and three sub-fully connected layers. The final output from the CNN network are: classification, offset (redshift), and column density.

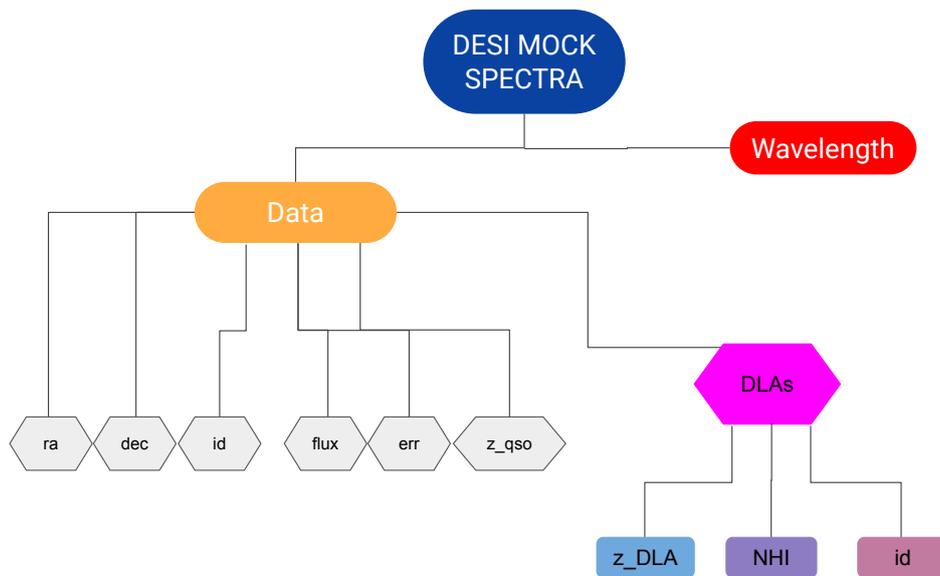


Figure 3.2: The data structure of the DLA Finder. The flux and error with other data are extracted from the input DESI mock spectra. After the CNN algorithm is run on the data, the code outputs a catalog including the redshift, column density and IDs of the detected DLAs.

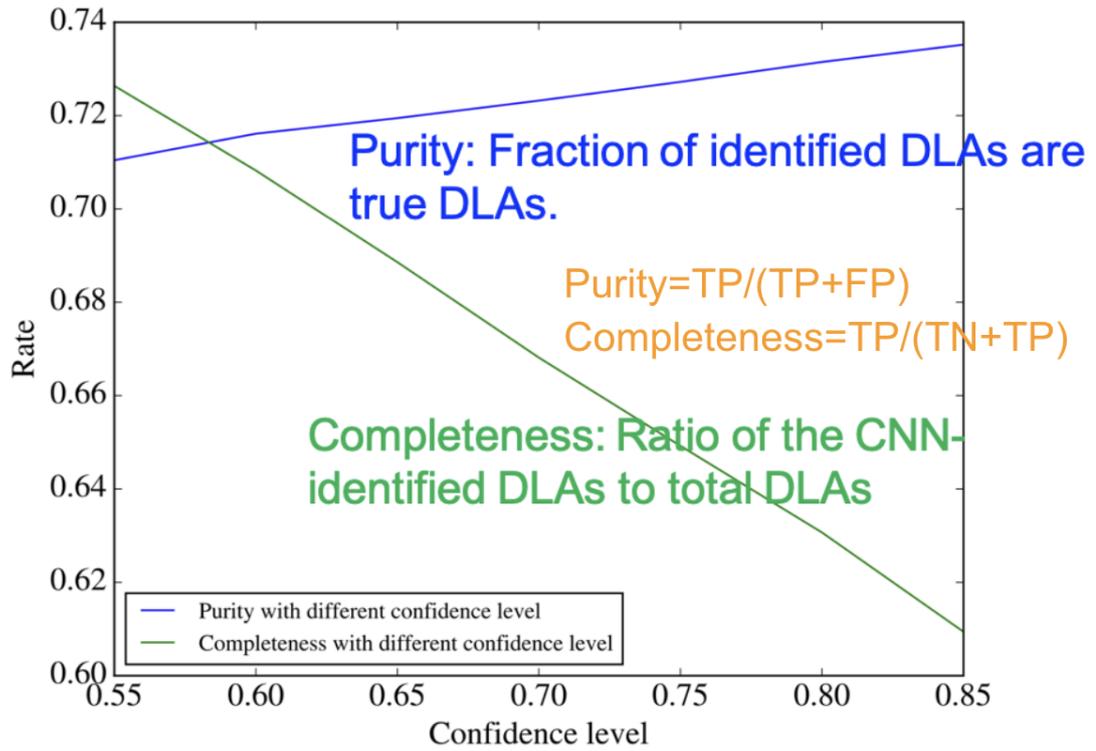


Figure 3.3: My preliminary work on the purity and completeness of the SDSS trained CNN model run on the DESI London mock v09. A comparable figure for the final algorithm after retraining the DESI mock data can be found in Fig. 20 in Wang et al. (2022).

Chapter 4

The Draco dSph Galaxy in DESI

4.1 Introduction

Dwarf galaxies play a crucial role in our understanding of galaxy formation and evolution. They are key to understanding the dark matter distribution, star formation history, chemical and dynamical evolution of the Milky Way (MW) (e.g., Bullock et al., 2000; Bullock & Johnston, 2005; Bovill & Ricotti, 2009; Weinberg et al., 2015). Starting with the classical dwarf spheroidal galaxies of the Milky Way, the discovery of the Sagittarius dSph (Ibata et al., 1994) was followed by many new, faint dwarf galaxies discovered at a rapid pace in the SDSS (e.g., Willman et al., 2005; Belokurov et al., 2006a, 2007, 2008; Koposov et al., 2008). The recent discovery of new dwarf spheroidal galaxies (dSphs) in the Dark Energy Survey (Bechtol et al., 2015a; Koposov et al., 2015a; Drlica-Wagner et al., 2015), Magellanic Satellites Survey (Torrealba et al., 2018) and other data has revolutionized our knowledge of the Milky Way's satellite systems.

These findings reveal a large population of dSphs, with some being potential satellites of the Large Magellanic Cloud (LMC) and the Small Magellanic Cloud (SMC).

Dwarf spheroidal galaxies live in low-mass dark matter halos. The subhalo mass function of a galaxy like the Milky Way and the central density profiles of dwarf galaxies and their subhalos are crucial for measurements of the distribution of dark matter on small scales. Thus dwarf galaxies play an important role in investigating the nature of dark matter itself (Zavala & Frenk, 2019). Data from the Gaia satellite (Gaia Collaboration et al., 2018a) and new photometric surveys have enabled new investigations of dwarf galaxy kinematics and their interactions with the Milky Way (Bechtol et al., 2015a; Wang et al., 2017; Muñoz et al., 2018; Li et al., 2018; Shipp et al., 2018). These and other recent investigations suggest that some of these dwarfs may be surrounded by significant low-surface-brightness stellar tidal features, a phenomenon predicted by simulations even for galaxies not currently experiencing significant mass loss through tidal stripping (Wang et al., 2017). Many studies have shown evidence of stars being stripped off the other dSph satellite galaxies and potential associations with extended stellar streams (e.g., Newberg et al., 2002; Majewski et al., 2003; Law et al., 2005; Koposov et al., 2012; Ibata et al., 2020; Vasiliev et al., 2021; Sestito et al., 2023).

Among the dwarf galaxies, Draco was first discovered by Wilson (1955) using the Palomar 48 inch telescope. Draco is a metal poor dwarf ($[\text{Fe}/\text{H}] = -1.93 \pm 0.01$ from Kirby et al. (2011)), with a 9.67 arcmin half-light radius at a heliocentric distance of 76 kpc (Muñoz et al., 2018). Draco is highly dark matter dominated (Klessen et al., 2003). Its large velocity dispersion made it the subject of interest for many previous

studies, none of which found clear evidence for tidal disruption in Draco (Ségall et al., 2007; Muñoz et al., 2018; Jensen et al., 2024).

The DESI Early Data Release DESI Collaboration et al. (2023a) includes data from the Milky Way Survey Validation (SV) campaign. The SV campaign consists DESI Collaboration et al. (2023b) $> 200,000$ unique stars. The Draco dwarf was observed on two of 8 square degree tiles in the SV campaign. With the addition of heliocentric velocity and metallicity measurements (Cooper et al., 2023; DESI Collaboration et al., 2023a) from DESI in a large area around Draco, we revisit the properties of the Draco dSph especially at and beyond the tidal radius.

This paper is organized as follows. In §4.2, we explain our sample selection. In §4.3, we discuss the Gaussian mixture model used to identify member stars in Draco. In §4.4, we illustrate the results (membership probability, mean line of sight velocity, mean metallicity, mean proper motion) we attained from the Gaussian mixture model. In §4.5 and §4.6, we compare our results to previous literature and present the conclusions.

4.2 Data

The Draco data sample we used in this study was taken as part of the DESI Science Verification program DESI Collaboration et al. (2023b). The spectra are included in the DESI Early Data Release (EDR) DESI Collaboration et al. (2023a). However, we use radial velocities and $[\text{Fe}/\text{H}]$ from a more recent version of the `rvspecfit` pipeline discussed in (Cooper et al., 2023; DESI Collaboration et al., 2023a) that will be included

in a future DESI data release. This version of `rvspecfit` provides measurements of the heliocentric velocity v_{hel} to about 1 km/s accuracy and $[Fe/H]$ accuracy to about 0.2 dex (Cooper et al., 2023).

The sample contains data observed on two different DESI tiles. Each tile is 8 square degrees and targets 5000 objects, including calibration and sky fibers. See section 2.2 in DESI Collaboration et al. (2023a) for more details about DESI observing. The two Draco tiles were observed at a range of moon illumination and observing conditions. The median exposure time on the stars used for this analysis was 6100.26 s. The details for RA/DEC selection and quality cuts for our preliminary selection sample are as follows.

- RA from 253 to 267
- Dec from 55.9 to 59.9
- $16 < r \text{ mag} < 21$
- $\log g < 4$
- `RVS_WARN = 0`
- `RR_SPECTYPE` not QSO
- `PHOT_VARIABLE_FLAG` not VARIABLE

The total number of stars after this preliminary selection is ~ 700 .

4.2.1 Photometry Selection

We use photometric catalogs from the Legacy Survey DR9 (Dey et al., 2019) to improve our selection of Draco members using color-magnitude cuts. We use an old, metal-poor ($[\text{Fe}/\text{H}] = -1.5$) Dartmouth Isochrone with age = 10 Gyr Dotter et al. (2008) and select stars in the range $|g - r - \text{Iso}(r)| < 0.39$, where Iso is the Isochrone value. The horizontal branch is not included in the Dartmouth Isochrone model. We use photometry from the Legacy Survey (Dey et al., 2019) of the globular cluster M92 to define a region for the horizontal branch of an old stellar population like Draco. We select a region 0.6 mag wide in g band centered on the cluster horizontal branch ridgeline and add that to the isochrone selection. The final selection yields 357 stars and is shown in Fig. 4.1.

4.3 Methods

We apply a Gaussian mixture model to the combined dataset of v_{hel} and $[\text{Fe}/\text{H}]$ from DESI (DESI Collaboration et al., 2023a) and proper motions from Gaia DR3 (Gaia Collaboration et al., 2023). We correct the radial velocity for perspective rotation using the equations in Kaplinghat & Strigari (2008). We implement the correction as part of fitting the mixture model parameters, so the correction is made using our best-fit proper motion. All the radial velocity and proper motion data are in the heliocentric frame. We model the radial velocity, $[\text{Fe}/\text{H}]$ and the proper motion distributions as a two-component mixture of the satellite and MW foreground/background components.

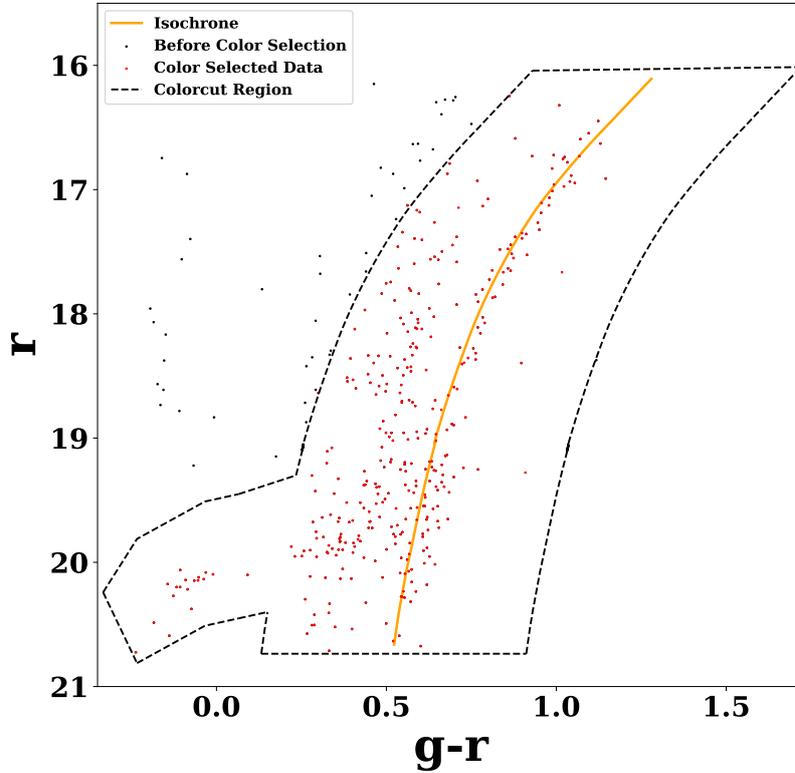


Figure 4.1: The region inside the dashed lines is the final selection for the Draco candidates sample from the DESI SV data. For the color-magnitude cuts used in the selection, we use an old, metal-poor ($[\text{Fe}/\text{H}] = -1.5$) Dartmouth isochrone with a age = 10 Gyr Dotter et al. (2008) and select stars within the Isochrone $g-r \pm 0.39$. We use photometry from the Legacy survey (Dey et al., 2019) of the M92 globular cluster to define a region for the horizontal branch that is not included in the isochrone.

We do not apply a radial density profile in our mixture model in order to avoid biasing our membership probabilities against stars outside the tidal radius. Our model has two components, Draco and the Milky Way (MW) interlopers, and three parameters: v_{hel} , $[\text{Fe}/\text{H}]$, proper motion, fit for each component. The functional form for the likelihood for each component is:

$$L(\bar{x}, \sigma^2; x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right) \quad (4.1)$$

where \bar{x} is the mean of the physical property and σ the combination of the intrinsic dispersion of the physical property and the internal errors of the data. The two component mixture model is a sum of likelihoods:

$$L = (f_{\text{sat}})L_{\text{satellite}} + (1 - f_{\text{sat}})L_{\text{MW}} \quad (4.2)$$

$$L_{\text{satellite}} = L_{v_{\text{hel}},\text{satellite}}L_{[\text{Fe}/\text{H}],\text{satellite}}L_{\text{PM},\text{satellite}} \quad (4.3)$$

$$L_{\text{MW}} = L_{v_{\text{hel}},\text{MW}}L_{[\text{Fe}/\text{H}],\text{MW}}L_{\text{PM},\text{MW}} \quad (4.4)$$

where f_{sat} is the fraction of stars belong to the dSph. Variables $L_{v_{\text{hel}},\text{satellite}}$, $L_{[\text{Fe}/\text{H}],\text{satellite}}$, $L_{\text{PM},\text{satellite}}$ refer to the heliocentric velocity, metallicity and proper motion likelihood components for the dSph. $L_{v_{\text{hel}},\text{MW}}$, $L_{[\text{Fe}/\text{H}],\text{MW}}$, $L_{\text{PM},\text{MW}}$ denote the heliocentric velocity, metallicity and proper motion likelihood components for the MW foreground/background stars.

We list all the parameters used in the model in table 4.1. The four parameters ($v_{\text{hel},\text{satellite}}$, $[\text{Fe}/\text{H}],\text{satellite}$, $\mu_{\alpha \cos \delta,\text{satellite}}$, $\mu_{\delta,\text{satellite}}$) represent the mean of the

physical properties for the galaxy component and $v_{\text{hel}, \text{MW}}$, $[\text{Fe}/\text{H}]_{\text{MW}}$, $\mu_{\alpha \cos \delta, \text{MW}}$, $\mu_{\delta, \text{MW}}$ represent the mean for the Milkyway background. We constrain the dispersion component of the proper motion ($\mu_{\alpha \cos \delta, \text{satellite}}$ and $\mu_{\delta, \text{satellite}}$) for the Draco dSph using previous measurements. We use the value of 9.1 km/s, the value of the Draco velocity dispersion measured by Muñoz et al. (2018) converted to the proper motion (0.025 mas/yr) using a distance of 75.8 kpc for Draco Muñoz et al. (2018). Therefore, we have two parameters $\sigma_{v_{\text{hel}, \text{satellite}}}$, $\sigma_{[\text{Fe}/\text{H}]_{\text{satellite}}}$ for the dSph dispersion and four parameters, $\sigma_{v_{\text{hel}, \text{MW}}}$, $\sigma_{[\text{Fe}/\text{H}]_{\text{MW}}}$, $\sigma_{\mu_{\alpha \cos \delta, \text{MW}}}$ and $\sigma_{\mu_{\delta, \text{MW}}}$ for the MW dispersion. Also, as discussed in section 7.4.2 in Cooper et al. (2023), there is an additional systematic component of ~ 0.9 km/s in the radial velocity errors; thus, we apply a 0.9 km/s in quadrature to the radial velocity errors in our Gaussian Mixture Model.

We model the satellite proper motion component as a multivariate distribution (e.g., Pace et al. (2022)). We include the covariance in the proper motion errors (the cross term in the proper motion error) and intrinsic proper motion dispersion terms.

The log likelihood for this multivariate Gaussian distribution is:

$$L(\bar{\mathbf{x}}, \mathbf{Covm} \mid \mathbf{X}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{Covm}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{Covm}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \quad (4.5)$$

where $\bar{\mathbf{x}}$ represents the parameters for the mean proper motion (μ_{δ} , $\mu_{\alpha \cos \delta}$) of Draco, \mathbf{Covm} represents the covariance matrix, \mathbf{X} represents the observed proper motion data, n represents the number of stars in the sample. For the MW proper motion distribution, we define the intrinsic dispersion by adding the MW dispersion component

to the diagonal term of the PM covariance. The covariance matrix \mathbf{Covm} is now defined as:

$$\text{Cov}(\mu_\delta, \mu_{\alpha \cos \delta}) + \text{Diag}(\sigma_{(\mu_{\alpha \cos \delta})}, \sigma_{\mu_\delta}) \quad (4.6)$$

where $\text{Cov}(\mu_\delta, \mu_{\alpha \cos \delta})$, is the Gaia covariance matrix for proper motions. For the Draco component, we add $\sigma_{\mu_{\alpha \cos \delta, \text{satellite}}}$ and $\sigma_{\mu_{\delta, \text{satellite}}}$, the Draco proper motion dispersion we fixed as explained above in the diagonal. For the background component, we add $\sigma_{\mu_{\alpha \cos \delta, \text{MW}}}$ and $\sigma_{\mu_{\delta, \text{MW}}}$, the parameters fit by the mixture model in the diagonal.

From the above discussion, there are 15 parameters in total for the model (see Table 4.1). In order to estimate the best values for the parameters, we run an MCMC process based on the affine invariant ensemble sampler for Markov Chain Monte Carlo Goodman & Weare (2010) in the *emcee* library Foreman-Mackey et al. (2013). The prior distributions for all the parameters are as follows (see Table 4.1):

- (i) One uniform prior between 0 and 1 for the fraction f_{sat} defined above
- (ii) Four parameters describing the mean and the dispersion of the heliocentric velocity, two for the velocity distribution of the Draco component and two for the Milky way component. For Draco, we use a uniform distribution prior for the mean heliocentric velocity between -400 and 0 and log dispersion between -1 and 3. For the MW foreground/background we use a uniform distribution prior for the mean heliocentric velocity between -400 and 0 and log dispersion between -1 and 3.
- (iii) Four parameters describing the mean and the dispersion of [Fe/H]. For the Draco component we use a uniform distribution prior for the mean [Fe/H] between -4.0 and -0.2 and log dispersion between -1 and 3 For the MW foreground/background compo-

ment we use a uniform distribution prior for the mean $[\text{Fe}/\text{H}]$ between -4 and -0.2 and log dispersion between -1 and 3.

(iv) Two parameters describing the Draco Proper motion. We use a uniform distribution prior for the mean proper motion for both $\mu_{\alpha \cos \delta, \text{satellite}}$ and $\mu_{\delta, \text{satellite}}$ between -2 and 2. Note that the dispersion of $\mu_{\alpha \cos \delta, \text{satellite}}$ and $\mu_{\delta, \text{satellite}}$ for the Draco component are fixed to 0.025.

(v) Four parameters describing the MW Proper motion. We use a uniform distribution prior for the mean proper motion $\mu_{\alpha \cos \delta, \text{MW}}$ and $\mu_{\delta, \text{MW}}$ between -10 and 10 and log dispersion for $\mu_{\alpha \cos \delta, \text{MW}}$ and $\mu_{\delta, \text{MW}}$ between -1 and 1.3. These priors are summarized in Table 4.1.

We checked that we attained an acceptance rate of 0.34 from the MCMC chain, which is between 0.3-0.5.

With the v_{hel} , $[\text{Fe}/\text{H}]$ and proper motion likelihood we described above, we can compute the relative likelihood for each star for membership in Draco and Milky Way. This gives us membership probabilities for the stars in Draco. The probability that a star belongs to the satellite is:

$$p = \frac{(f_{\text{sat}})L_{\text{satellite}}}{(f_{\text{sat}})L_{\text{satellite}} + (1 - f_{\text{sat}})L_{\text{MW}}} \quad (4.7)$$

where the component $L_{\text{satellite}}$ refers to the dSph and the component L_{MW} describes the foreground/background stars. The variable f_{sat} is the fraction of stars that belong to the dSph, and thus $1-f_{\text{sat}}$ is the fraction of stars that belong to the MW foreground/background.

Table 4.1. Priors for the Gaussian Mixture Model

Parameters	Prior Range
f_{sat}	0-1
$v_{\text{hel,satellite}}$	-450 - -150 km/s
$v_{\text{hel,MW}}$	-450 - -150 km/s
$\text{Log } \sigma_{v_{\text{hel,satellite}}}$	-1 - 3
$\text{Log } \sigma_{v_{\text{hel,MW}}}$	-1 - 3
$[\text{Fe}/\text{H}], \text{satellite}$	-3.5 - -0.5
$[\text{Fe}/\text{H}], \text{MW}$	-3.5 - -0.5
$\text{Log } \sigma_{[\text{Fe}/\text{H}], \text{satellite}}$	-1 - 3
$\text{Log } \sigma_{[\text{Fe}/\text{H}], \text{MW}}$	-1 - 3
$\mu_{\alpha \cos \delta, \text{satellite}} / \mu_{\delta, \text{satellite}}$	-2 - 2
$\mu_{\alpha \cos \delta, \text{MW}} / \mu_{\delta, \text{MW}}$	-2 - 2
$\sigma_{\mu_{\alpha \cos \delta, \text{satellite}}} , \sigma_{\mu_{\delta, \text{satellite}}}$	fixed 0.025
$\text{Log } \sigma_{\mu_{\alpha \cos \delta, \text{MW}}} , \text{Log } \sigma_{\mu_{\delta, \text{MW}}}$	-1 - 1.3

4.4 Results

We summarize the best fitted parameters for the Draco heliocentric velocity, $[\text{Fe}/\text{H}]$ and proper motion from the mixture model in Table 4.2. The uncertainty on each fit parameter is calculated from the 16th/84th percentile of the posterior distribution. We compare our parameters to previous results in the next section.

With the best fit mean and dispersion for v_{hel} , $[\text{Fe}/\text{H}]$, $\mu_{\alpha} \cos \delta$ and μ_{δ} we use equation (7) to calculate the membership probability P_i for the i th star of the Draco candidates shown in Fig. 4.1. The distribution of membership probabilities for the Draco candidates is shown in Fig. 4.2. We define high probability member stars to be those with probability > 0.9 .

We plot the candidates, color-coded by their membership probabilities, in the color-magnitude diagram in Fig. 4.3. The ones outside the tidal radius are plotted as the large triangles using the same color scale. High membership probability stars are distributed along the isochrone, showing that stars identified as Draco members by the mixture model are also likely members based on their location in the CMD.

4.4.1 Spatial Distribution of the Sample

Our Gaussian mixture model does not use spatial information or a density profile constraint, minimizing bias against identifying members near and beyond the tidal radius of the galaxy. We plot the spatial distribution of our sample with membership probability > 0.9 in Fig. 4.4 along with the orbit (Qi et al., 2022) of the Draco dSph.

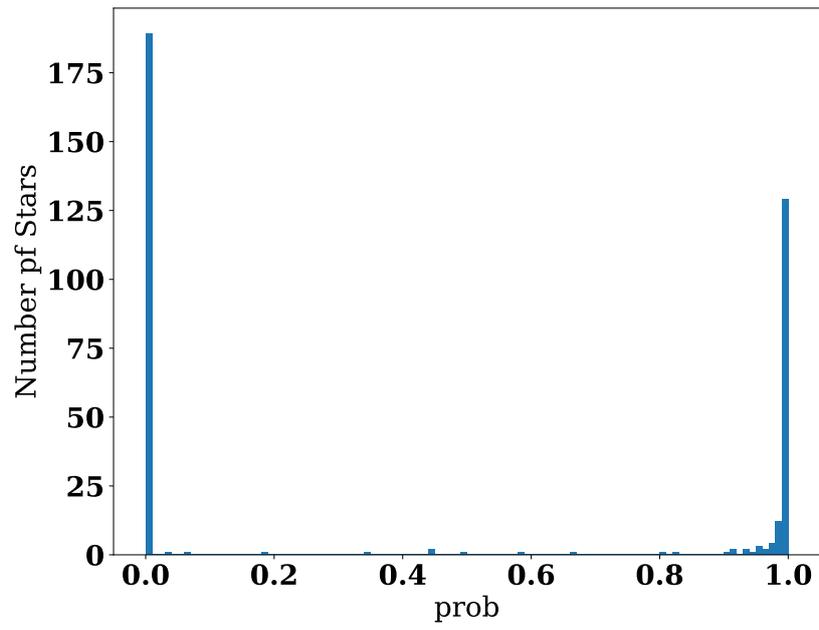


Figure 4.2: The probability distribution for the Draco members predicted from the Gaussian mixture model. From the probability distribution, we define the high probability members to be $\text{prob} > 0.9$

Table 4.2. Values for the Best Fitted Parameters

Draco Parameters	Mean value with error
f_{sat}	0.45 ± 0.026
$V_{\text{hel,satellite}}$	-290.66 ± 0.82
Velocity Dispersion ($\sigma_{V_{\text{hel,satellite}}}$)	$9.68^{+0.70}_{-0.64}$
[Fe/H]	-2.10 ± 0.04
[Fe/H] _{satellite} Dispersion ($\sigma_{[\text{Fe}/\text{H}]_{\text{satellite}}}$)	0.48 ± 0.03
$\mu_{\alpha \cos \delta, \text{satellite}}$	0.03 ± 0.01
$\mu_{\delta, \text{satellite}}$	-0.19 ± 0.01

Note. — Note that the dispersion of $\mu_{\alpha \cos \delta}$ and μ_{δ} for Draco are fixed to 0.025 and we do not fit those two parameters.

We use the orthographic projection of RA and Dec and to the Cartesian frame:

$$X = \cos \delta \sin(\alpha - \alpha_c) \quad (4.8)$$

$$Y = \sin \delta \cos \delta_c - \cos \delta \sin \delta_c \cos(\alpha - \alpha_c) \quad (4.9)$$

where α and δ are the location of a star in the RA and Declination directions, α_c and δ_c are the center coordinates of the dSph. The blue ellipse is the Draco tidal radius of 48.1 arcmin from Muñoz et al. (2018). The ellipse at the tidal radius is projected in the same way as discussed above, using the ellipticity value of 0.29 from Qi et al. (2022). We detect eight stars outside this 48.1 arcmin tidal radius.

4.4.2 Velocity Dispersion and Metallicity gradient

We bin the sample relative to the center of Draco, using an adaptive bin size to have at least 40 stars in each bin from the center of the galaxy to ~ 70 arcmin from the center. We then run a Gaussian mixture model using the same likelihood in equation (2) in each bin to measure the velocity dispersion and metallicity profile of Draco as a function of the distance from the center. The final velocity dispersion (top panel) and metallicity (bottom panel) as a function of distance from the center are shown in Fig. 4.5. Our radial velocity profile is relatively flat, as expected if Draco is embedded in a more massive dark matter halo, and is consistent with the measurement in Walker et al. (2007). See §4.5 for discussion of the metallicity gradient.

4.4.3 Surface Density Profile

We calculate the projected surface density per square arcmin of our Draco member stars using elliptical radial bins, with ellipticity ϵ equal to the measured value for Draco of 0.29. The half-light radius of Draco r_0 is 9.67 arcmin (Muñoz et al., 2018). Between $1.5r_0$ and $8r_0$ we use bin intervals of r_0 . For $r > 8r_0$, we use $2r_0$ bins. We use these elliptical bins to calculate the surface density of high-probability member stars as a function of projected radius r .

We compare with the number density predicted from a Plummer model, a King model (King, 1962) and an exponential model in the top panel of Fig. 4.7. The Plummer model (Plummer, 1911) is:

$$\Sigma(R_e) = \frac{1}{\pi a_h^2 (1 - \epsilon)} (1 + R_e^2/a_h^2)^{-2} \quad (4.10)$$

R_e is the elliptical radius defined as:

$$(R_e^2) = R_x^2 + \frac{R_y^2}{(1 - \epsilon)^2} \quad (4.11)$$

where a_h is the semimajor half-light radius and ϵ is the projected ellipticity. We rotate the frame by the position angle θ (angle measured from north to east),

$$R_x = x \cos(\theta) - y \sin(\theta) \quad (4.12)$$

$$R_y = x \sin(\theta) + y \cos(\theta) \quad (4.13)$$

We normalize the Plummer model to our measurement at 10.95 arcmin, where our sample has low incompleteness and a large number of member stars. Our uncorrected

surface density measurements, plotted in red, are in approximate agreement with the Plummer model for large radii but very discrepant at smaller radii. This is due to the incompleteness of our spectroscopic sample in the central region of Draco. The DESI instrument has a fixed fiber density of 667/square degree. Since Draco was observed with a single assignment of fibers to targets shared by the two DESI tiles (only the center was shifted), this results in higher incompleteness in the central region of the dSph where the stellar density is highest. To correct our density estimates, for each elliptical bin we calculate the ratio between the total number of targets available from the color-magnitude selection (§2.2) and the number that were assigned to fibers and observed. We use that ratio to correct the surface density of our high-probability Draco members. The corrected surface density is shown in blue in the top panel of Fig. 4.7 and is consistent with the Plummer model in the inner part of the dSph. We also compare our model with an exponential profile and a King model profile. A King profile (King, 1962) describes the density of a system in equilibrium in a tidal field. An exponential profile is usually a better fit to the Milky Way satellite dSph surface density profiles (McConnachie & Venn, 2020a). We fit these to the corrected surface density (blue points in the top panel of Fig. 4.7). The best fit exponential curve is labelled in yellow and the King model is labelled in green in Fig. 4.7). Our data agree with all three profiles in the central region, including the King profile. The exponential model is a better fit to the Draco surface density than either the Plummer model or the King model near and outside the tidal radius. In §4.5.1, we discuss the surface density and best fit model in detail.

4.5 Discussion

We find a mean line of sight heliocentric velocity $v_{\text{hel}} = -290.66 \pm 0.82$ km/s with a dispersion $= 9.68^{+0.70}_{-0.64}$ km/s. Our value is consistent with the heliocentric velocity measured in Pace et al. (2022) (-290.7 ± 0.75 km/s) and Muñoz et al. (2018) ($v_{\text{hel}} = -291$ km/s). Our radial velocity dispersion measurement ($\sigma_{v_{\text{hel}}} = 9.68 \pm 0.67$) is also consistent with values quoted in previous work: 9.0 ± 1.1 km/s in Massari et al. (2019) and 9.1 ± 1.2 km/s in Muñoz et al. (2018). The velocity dispersion as a function of radius is shown in Fig. 4.5. This is also consistent with the measurement of the radial velocity as a function of radius from Walker et al. (2008) and of our measurement of the radial velocity dispersion profile of the data from Walker et al. (2023) within the measurement errors (see §4.5.3.1).

Our measurements for the Draco dSph proper motion components are $\mu_{\alpha} \cos \delta = 0.03 \pm 0.01$ and $\mu_{\delta} = -0.19 \pm 0.01$. Figure 4.9 shows our measurements are consistent with previous studies of Draco that used only Gaia EDR3 proper motions and photometry. Qi et al. (2022) measured $\mu_{\alpha} \cos \delta = 0.045 \pm 0.006$, $\mu_{\delta} = -0.188 \pm 0.006$ and Battaglia et al. (2022) found $\mu_{\alpha} \cos \delta = 0.04 \pm 0.01$, $\mu_{\delta} = -0.19 \pm 0.01$.

Our best-fit value for the mean metallicity $[\text{Fe}/\text{H}]$ (-2.10 ± 0.04) is slightly lower than the mean value found by Kirby et al. (2011) (-1.93 ± 0.01) or Kirby et al. (2013) ($[\text{Fe}/\text{H}] = -1.98 \pm 0.01$). Cooper et al. (2023) compare the DESI $[\text{Fe}/\text{H}]$ measurements in star clusters and dwarf galaxies to previous literature measurements from Carretta et al. (2009), Kirby et al. (2011) and D’Orazi et al. (2020). They find that the DESI

rvs $[\text{Fe}/\text{H}]$ measurements are on average 0.13 dex low compared to other studies. This known systematic offset accounts for the difference between our best-fit $[\text{Fe}/\text{H}]$ and previous work.

Our measurement of the width of the metallicity $[\text{Fe}/\text{H}]$ distribution (0.48 ± 0.03) agrees with previous measurements: $\sigma_{[\text{Fe}/\text{H}]} = 0.47$ in Kirby et al. (2011) and $\sigma_{[\text{Fe}/\text{H}]} = 0.42$ in Kirby et al. (2013). In the right panel of the Fig. 4.5, we measure the slope of the metallicity gradient to be 0.01 ± 0.01 dex per arcmin (see §4.4.2 for detail description), smaller than the slope (0.026 ± 0.004 dex per arcmin) measured by Kirby et al. (2011). The metallicity gradient in Kirby et al. (2011) is measured out to $r = 9$ arcmin, near the half-light radius of Draco. To better compare our results to Kirby et al. (2011) over the same radial range, we divide our high probability members into 7 bins and calculate the mean metallicity in each bin in Figure 4.6. There is a clear trend for the metallicity gradient to have steeper slope, more comparable to the value measured by Kirby et al. (2011), within the half-light radius, where the two samples overlap, and then become flatter at large radii.

4.5.1 Surface Number Density at Large Radii

The Draco surface number density profile (Top panel in Fig. 4.7) shows a excess over a King profile and an exponential profile outside the tidal radius. This is more clear in the spatial logarithmic derivative of the surface number density shown in the bottom panel in Fig. 4.7. Our high probability members extend to as far as eight half light radii from the center of Draco. Beyond 5 half light radii, our Draco surface number

density profile shows a clear excess over the King profile and is potentially above the single exponential profile. However, each of the last two bins is populated by one high probability member star. The top panel Fig. 4.8 (green curve) shows that the number of Draco candidates selected for spectroscopy remains roughly constant from the Draco tidal radius out to a projected radial distance of 126 arcmin. We therefore have about the same probability of finding Draco members in our sample over that range of radial distance. However, we find the fraction of stars with membership probability > 0.9 decreases to $< 10\%$ beyond the tidal radius and we identify only eight high probability members there, indicating that there is not a large fraction of Draco members outside the tidal radius. We do not find well-defined tidal tails or other extended structure in the outskirts of Draco, so if tidal interactions have extended the Draco stellar distribution it was long enough ago for the tidal tails to have dispersed. Utilizing the Gaia EDR3 data, Jensen et al. (2024) do not find an extended profile for Draco outside the tidal radius. Previous studies have the same conclusion. Jardel et al. (2013) present a dynamical estimate of the Draco tidal radius of 3 kpc and point out that is likely too large for equilibrium tidal effects to influence the stellar component. Adding the radial velocity and metallicity from DESI data, we find only a small excess surface density above the single exponential profile, in agreement with previous work. An exponential spatial density profile has no direct physical motivation, so there is no particular reason we should expect dSph surface density profiles to follow an exponential. Given the small number of Draco member stars outside the tidal radius we can at most conclude that a single exponential profile is not the best model for the outer spatial density profile of

Draco.

Sestito et al. (2023) discuss the surface density of stars outside the tidal radius for Sculptor and Fornax Dwarf galaxies. They find that Sculptor has a obvious excess above an exponential density profile that can be explained by a model of its tidal interaction with the Milky Way. Fornax is well fit by a single exponential, suggesting no recent tidal disturbance. We find that the surface density of Draco beyond its tidal radius has a small potential excess above an exponential profile, between the cases of Sculptor and Fornax. The orbital parameters of the three dSph galaxies also suggest the impact of tidal interaction with the Milky Way on Draco is between that of Sculptor and Fornax. From the orbital motion information listed in Pace et al. (2022), Draco has pericenter distance $r_{peri} = 58$ kpc and apocenter distance $r_{apo} = 106.3$ kpc. Draco has $f_{peri} = 0.37$, where f_{peri} is defined by $\frac{r_{GC} - r_{peri}}{r_{peri} - r_{apo}}$ and $f_{peri} = 0$ at pericenter and 1 at apocenter. By comparison, Fornax has $f_{peri} = 0.96$, $r_{peri} = 76.7$ kpc and $r_{apo} = 152.7$. Fornax is very close to apocenter and has a larger pericenter distance, which may explain why it shows even less evidence of recent tidal disturbance than Draco. Sculptor has $f_{peri} = 0.39$, $r_{peri} = 44.9$ kpc and $r_{apo} = 145.7$. Sculptor passes closer to the center of the Milky Way than Draco. Sculptor also has a more radial orbit ($e = 0.54$) compared to Draco ($e = 0.3$). This and its smaller pericenter distance imply that Sculptor may have experienced a larger impact from tidal interactions with the Milky Way potential than Draco (Peñarrubia et al., 2008a; Gnedin et al., 1999).

4.5.2 High Probability Member Stars in the Outskirts of Draco

We identify several high probability members near and outside the Draco tidal radius of 48.1 arcmin. There are eight high probability members outside the tidal radius ellipse, as far as eight half light radii. Information for those members is listed in Table 4.4. The spatial distribution for the high probability star members outside the tidal radius is shown in Fig. 4.4. These extra-tidal stars do not appear to be preferentially distributed along Draco’s orbit. Fig. 4.3 demonstrates that these members are distributed along the old population isochrone that describes Draco, further demonstrating their high probability of being members.

4.5.3 Comparison with Other Recent Draco Results

We compare our results with two other recent surveys of the Draco dSph by Walker et al. (2023) and Qi et al. (2022).

4.5.3.1 Draco data in the Walker et al. (2023) spectroscopic catalog

Walker et al. (2023) present their latest reductions of the Magellan/M2FS and MMT/Hectochelle spectroscopic data on multiple dwarf galaxies, including Draco. In order to compare the our outskirts members in Draco with those in the Walker et al. (2023) catalog, we apply our Gaussian Mixture model analysis to their measured parameters v_{rad} (line of sight velocity) and Gaia DR3 proper motion. We use the same prior range as for the DESI data and we adopt the same cuts for radial velocity and proper motion as discussed in Section 4.3. In the two right panels of Fig. 4.10 we com-

Table 4.3. Values for the Best Fitted Parameters for Walker et al. (2023) sample

Parameters	Mean value with error
V_{hel}	-291.622 ± 0.43
$\sigma_{V_{\text{hel}}}$	9.27 ± 0.32
$\mu_{\alpha \cos \delta, \text{walker}}$	0.04 ± 0.01
$\mu_{\delta, \text{walker}}$	-0.18 ± 0.01

pare $[\text{Fe}/\text{H}]$, radial velocity and proper motion for the high probability members we find in our DESI plus Gaia sample with those we find in the Walker et al. (2023) catalog. Using the data from Walker et al. (2023, 2008), our best fit parameters are listed in table 4.3. Our results from the DESI sample are in good agreement with the results from the sample of Walker et al. (2023, 2008), allowing us to compare high probability members outside the tidal radius in the two samples.

4.5.3.2 Comparison of Draco Outskirt Members with Previous Studies

The tidal radius we use in this work and that we use to compare with previous literature is 48.1 arcmin, from Muñoz et al. (2018). Qi et al. (2022) determine Draco membership probabilities using Gaia data. In the lower right panel in Fig. 4.10, we plot the proper motion of the high probability members outside the tidal radius in our work, Qi et al. (2022) and Walker et al. (2023). This panel shows the proper motion

distribution (labelled in yellow) of the two high probability members outside the tidal radius that overlap in the three studies: Walker et al. (2023), Qi et al. (2022) and our sample. The high-probability members outside the tidal radius from the DESI data and the Walker et al. (2023) data have a larger spread in proper motion compared to the members in Qi et al. (2022). This may be because for both samples we have additional data, radial velocity and/or $[\text{Fe}/\text{H}]$, to help determine the membership probabilities. The lower left panel of Fig. 4.10 shows the proper motions and the proper motion errors of our DESI Draco members outside the tidal radius. They are still consistent with membership in Draco after taking into account the errors.

Qi et al. (2022) and our study have two members in common, shown in yellow points in the lower right panel of Fig. 4.10. In our results, these stars have metallicities, velocities and proper motions consistent with Draco and are distributed along an isochrone matching Draco’s properties. The Gaia IDs for these two stars are 1434492516786370176 and 1433949770359264768. Of the other seven members identified outside the tidal radius in Qi et al. (2022), one has a $v_{\text{hel}} < 5$ km/s measured in DESI, the others do not pass the original target selection criteria or were not assigned to a fiber in the DESI SV data.

From our mixture model analysis of the Walker et al. (2023) data, we identify four high probability Draco outskirts members in common with our DESI sample (see §4.5.3.2). Two of them are also identified by Qi et al. (2022) as discussed above. The other two have Gaia ID 1433252164591733504 and 1433193345014967424. The r band magnitude limit of our Draco spectroscopic data is about 0.5 mag fainter than Walker

et al. (2023). We cross match the Walker et al. (2023) high probability outskirts members with our entire DESI Draco sample, regardless of membership probability, and find that DESI did not target the other outskirts members we identify in the Walker et al. (2023) data.

4.6 Conclusion

By combining line of sight velocity and $[\text{Fe}/\text{H}]$ from DESI spectroscopy with Gaia DR3, we measure the radial velocity, mean proper motion and mean metallicity $[\text{Fe}/\text{H}]$ of the Draco dSph and find good agreement with previous work. The DESI SV spectroscopic data on Draco extends to 21 in r -band magnitude and to a radial distance of 140 arcmin from the center of Draco. We identify eight high-probability stars outside the King tidal radius for Draco. We do not find a significant number of extra-tidal stars. This is in agreement with previous studies (Jensen et al., 2024; Jardel et al., 2013; Ségall et al., 2007; Muñoz et al., 2018), that has found very little evidence of an extended stellar distribution or tidal tails. Comparing the distribution of these eight stars in Fig. 4.4 with the orbital direction of Draco, there is no obvious alignment. We identify three high probability member stars outside the tidal radius in common with previous studies, further increasing our confidence that we are identifying stars associated with Draco. The orbital parameters of Draco suggest that it may not have experienced a large impact from tidal interactions with the Milky Way, explaining our and previous results that do not find many members outside the tidal radius or other

evidence of significant tidal interactions with the Milky Way.

Table 4.4. Properties for the eight Star Members Outside the Tidal Radius

Gaia ID	RA	DEC	v_{hel}	[Fe/H]	$\mu_{\alpha} \cos \delta$	μ_{δ}	Prob ¹
1432965054618434560	261.56	57.73	-295.23 ± 1.80	-2.2 ± 0.10	-0.20 ± 0.38	-0.04 ± 0.48	0.998
1434407579513539072	262.57	57.72	-308.46 ± 3.48	-2.2 ± 0.24	-0.25 ± 0.48	-0.09 ± 0.48	0.987
1433888300787161728	258.65	58.12	-285.31 ± 1.37	-2.4 ± 0.08	-0.14 ± 0.39	0.18 ± 0.37	0.999
1433193345014967424	261.22	58.29	-293.87 ± 0.84	-2.1 ± 0.05	0.14 ± 0.26	0.08 ± 0.20	1.0
1434492516786370176	261.91	58.26	-283.98 ± 0.39	-1.9 ± 0.02	0.05 ± 0.13	-0.22 ± 0.14	1.0
1433949770359264768	259.14	58.50	-289.45 ± 0.76	-2.1 ± 0.05	0.09 ± 0.19	-0.21 ± 0.17	1.0
1433996705761927168	259.53	58.51	-295.62 ± 1.29	-1.6 ± 0.06	-0.58 ± 0.42	-0.01 ± 0.60	0.905
1433252164591733504	261.27	58.63	-284.28 ± 1.99	-2.1 ± 0.11	-0.55 ± 0.42	0.49 ± 0.39	0.951

¹Membership Probability

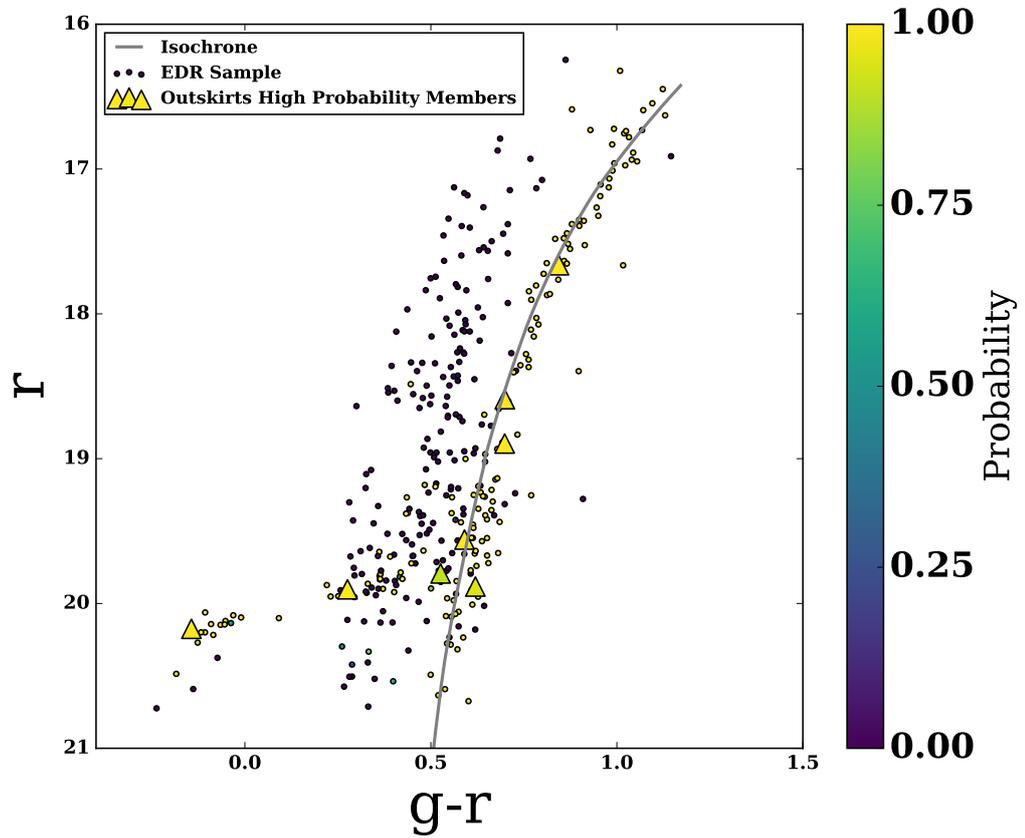


Figure 4.3: The color-magnitude diagram with the membership probabilities from the mixture model. Points are colored according to their probability. The high probability members outside the tidal radius are labelled with triangles. The grey isochrone is defined in Section 4.2.1

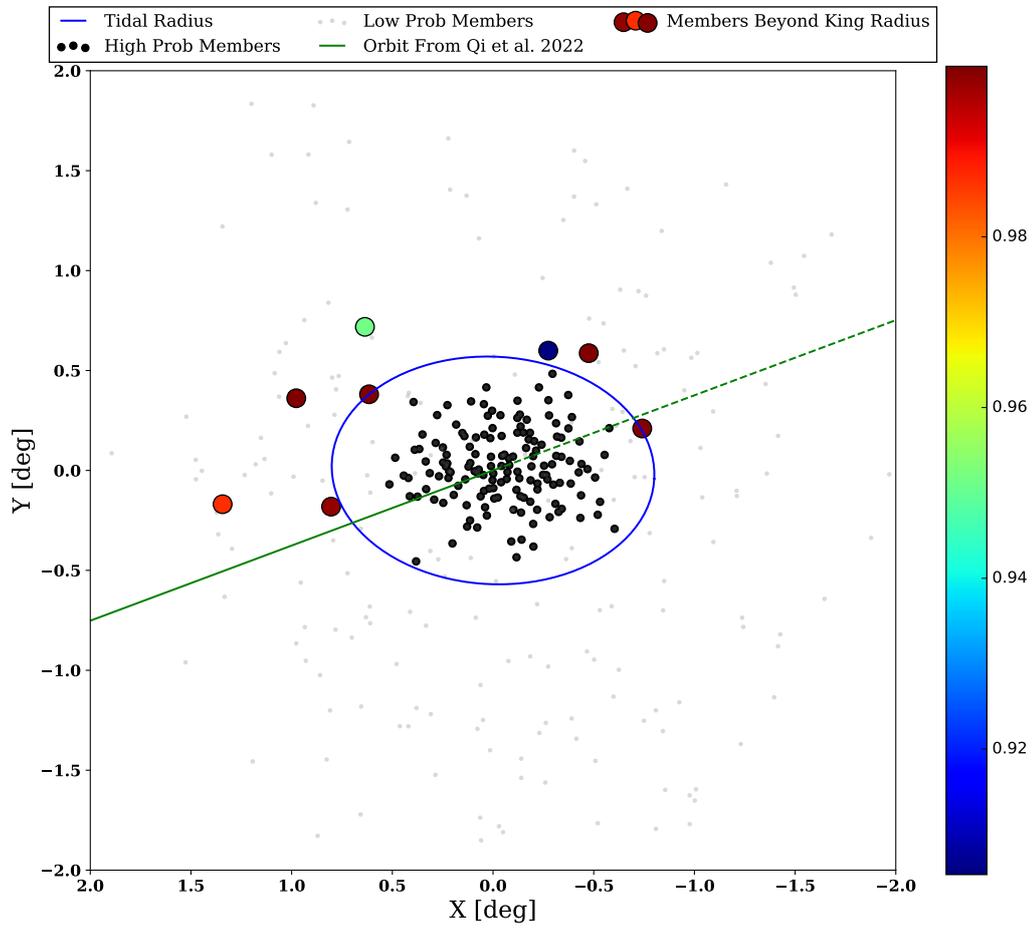


Figure 4.4: Spatial distribution for the probability greater than 0.9 star sample with the tidal radius in blue ellipse. The orbit of the satellite forwards (backwards) in time labelled in the solid (dotted) green line. The eight high probability stars outside the tidal radius are presented in larger circles with the colorbar showing their individual probability.

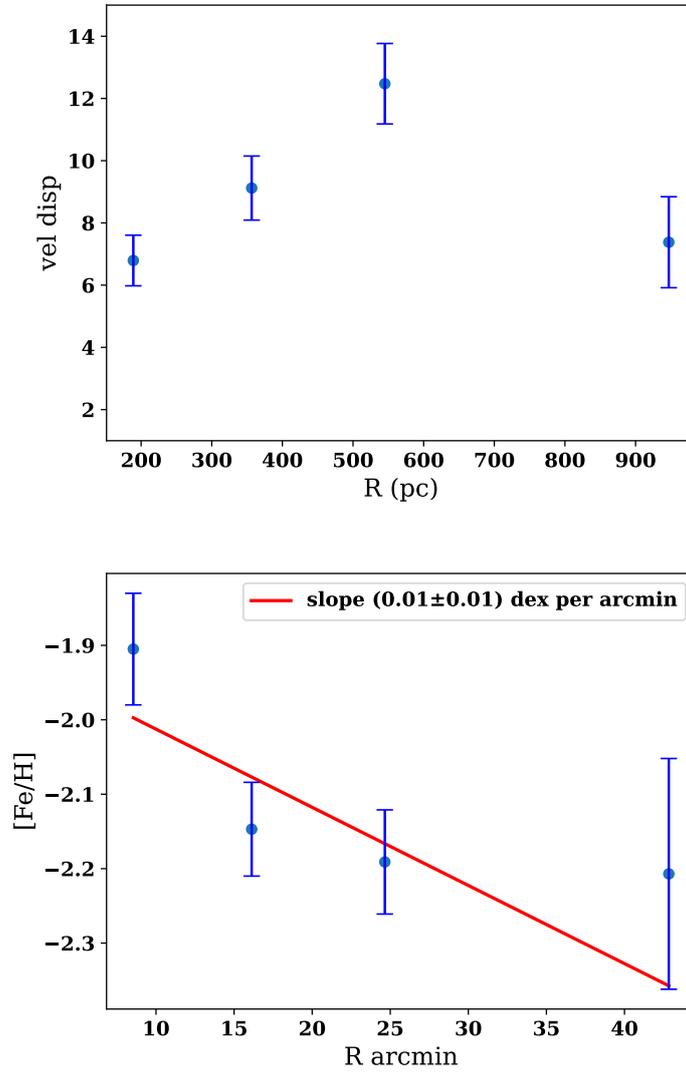


Figure 4.5: The velocity dispersion (Upper panel) and the metallicity gradient (Lower panel) of our high probability Draco members corrected for the perspective rotation as a function of distance from the center in parsec for the velocity dispersion and in arcmin for the metallicity.

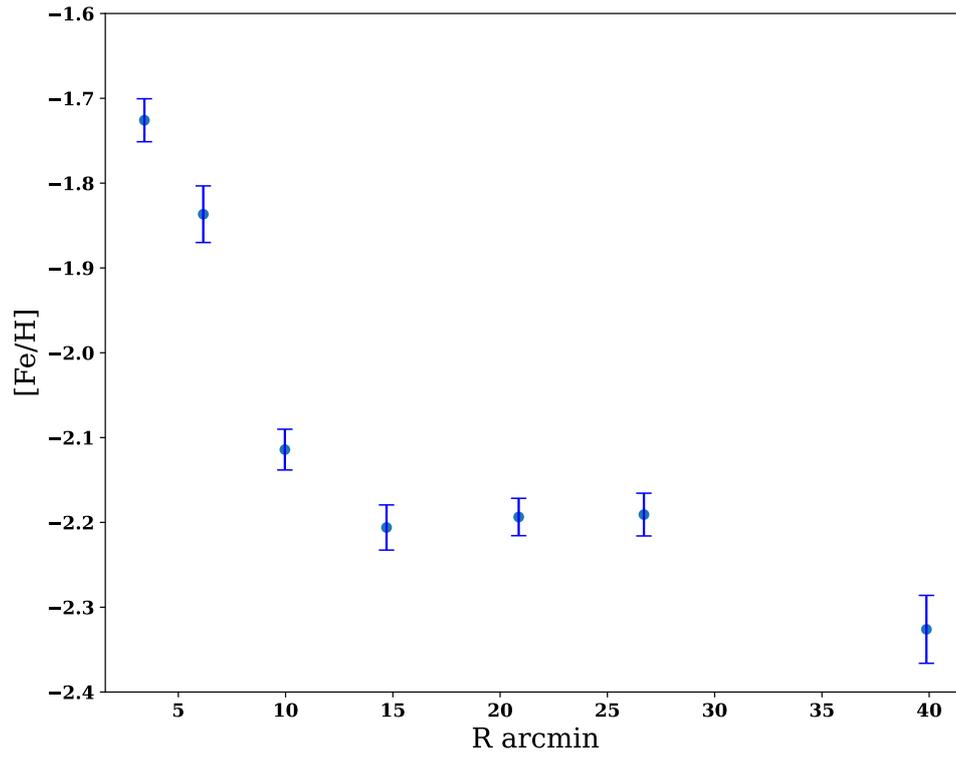


Figure 4.6: The metallicity gradient is measured from the mean metallicity of our high probability Draco members in different bins as a function of the distance from the center.

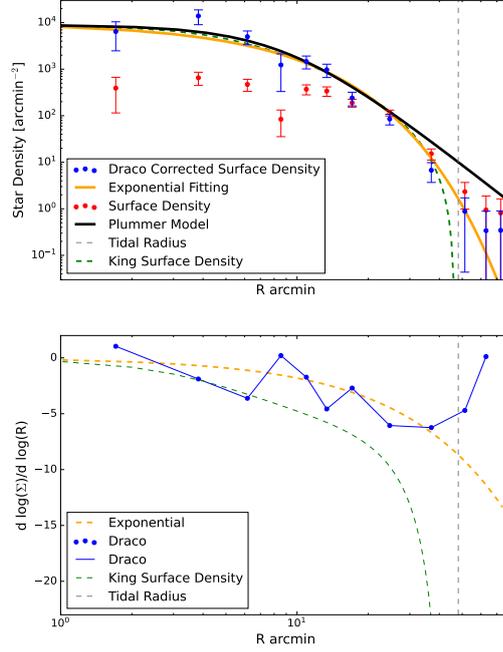


Figure 4.7: Top panel: The surface number density for stars with membership probability greater than 0.9, corrected for sample incompleteness (blue data points with error bars). A Plummer model is shown in black, a King surface density profile (King, 1962) in green and a exponential model in yellow. The red data points are the original measured surface density. The blue data points are the final corrected surface density after taking the incompleteness in the inner radius into account. The grey dash line is the tidal radius measured in Qi et al. (2022). Bottom panel: The spatial logarithmic derivative of Draco stars (blue data) plotted with the exponential model as a dashed yellow line. The green dashed line shows the spatial logarithmic derivative for a King surface number density profile model (King, 1962) without tidal disruption from Peñarrubia et al. (2008b).

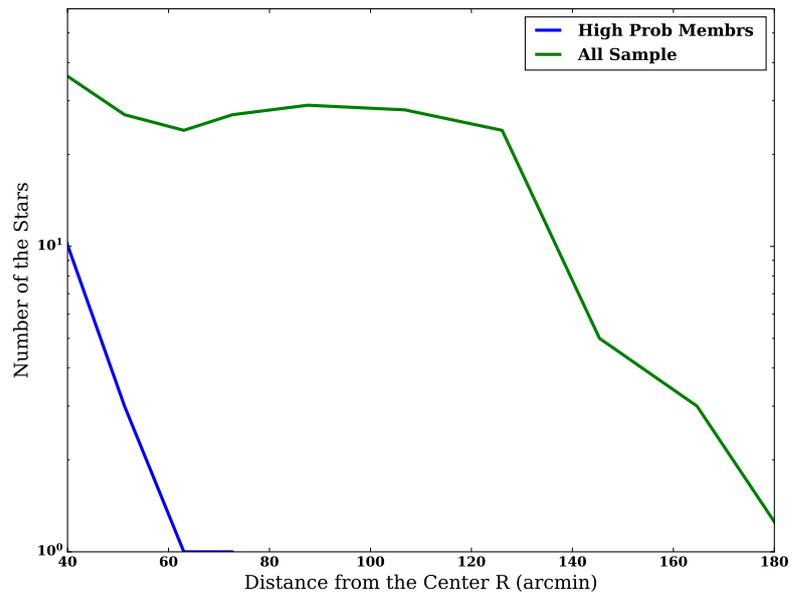


Figure 4.8: The number of stars as a function of radius. Stars with membership probability greater than 0.9 is labelled in blue while the number of all the stars from our sample after sample selection is labelled in green.

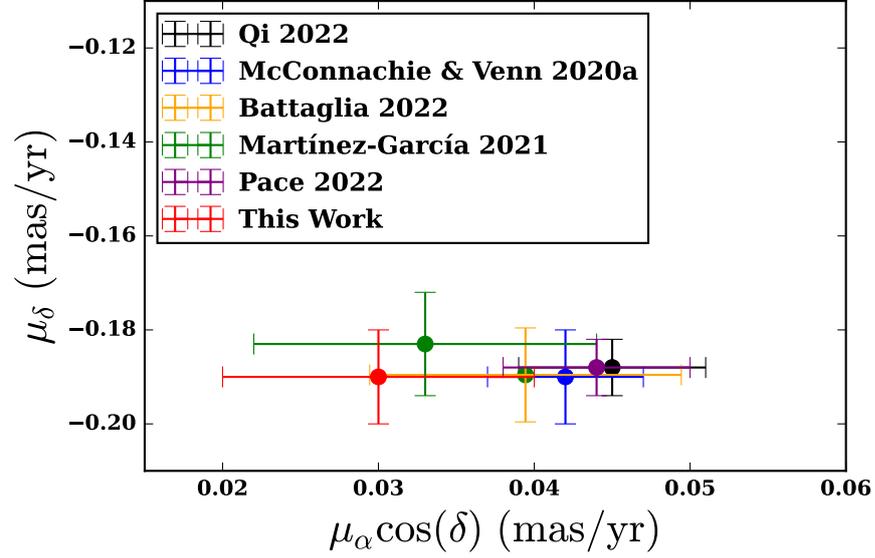


Figure 4.9: Comparison of proper motion (PM) measurement between this work and previous literature. The red error bars show the Draco PM with 1-sigma uncertainties measured in this work. The black error bars are the PM value and uncertainty measured by Qi et al. (2022) using Gaia DR3 data and photometry. The blue error bars show the PM measured by McConnachie & Venn (2020b). The orange data represents PMs from Battaglia et al. (2022), and the green crosses show PMs from Martínez-García et al. (2021). The purple measurement is from Pace et al. (2022). Our PM distributions are consistent with other PM values.

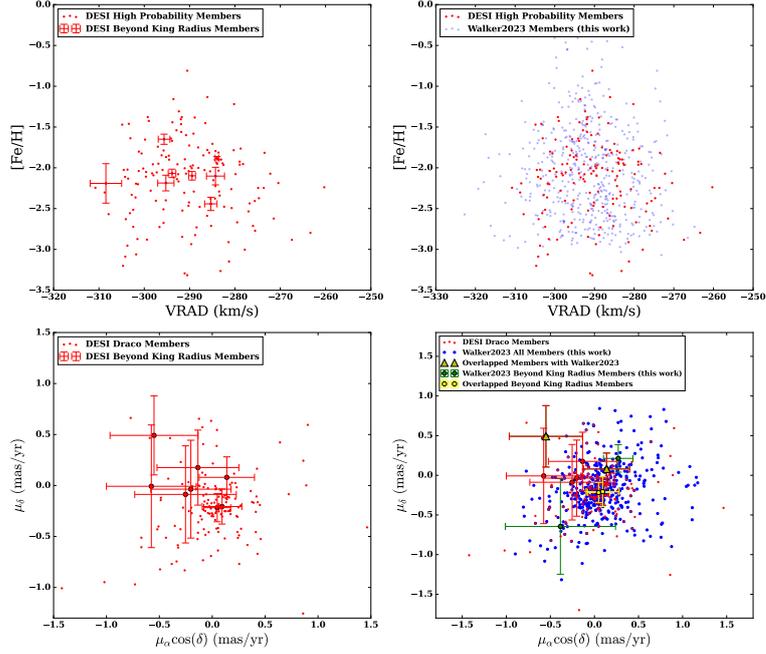


Figure 4.10: $[\text{Fe}/\text{H}]$ versus heliocentric radial velocity distribution and the proper motion distribution of high probability members in our work plotted with the high probability members from Qi et al. (2022) and high probability members we found using data from Walker et al. (2023). The left two panels are the $[\text{Fe}/\text{H}]$ versus radial velocity distribution (upper panel) and the proper motion distribution (lower panel) from this work. The right two panels compare our sample and previous literature. The top right panel shows the $[\text{Fe}/\text{H}]$ versus heliocentric velocity for the high probable members from our sample and Walker et al. (2023). The lower right panel compares the proper motion distribution of our sample, high probability members in Walker et al. (2023) (labelled in green) and two high probability members (labeled in yellow circle) outside the tidal radius matched in all three studies (this work, Walker et al. (2023) and Qi et al. (2022)). The yellow triangle shows the two more matched extra-tidal stars in our sample and in Walker et al. (2023) with the red errors from the DESI data.

Chapter 5

Unveiling the Outskirt of Sextans: Searching for Stars along the Orbit of Sextans Dwarf through 6D Orbit Integration

5.1 Introduction

Large sky surveys including Gaia DR3 (Gaia Collaboration et al., 2023) and DESI (DESI Collaboration et al., 2023a) open up a new window for investigations of the formation and evolution of the Milky-way (MW) galaxy. The combination of precise tangential motion measurements from Gaia (Gaia Collaboration et al., 2023) with radial velocity and distance measurements using spectroscopy from DESI (DESI Collaboration et al., 2023a) allow us to build up a 6-D integral-of-motion space for studying the

dynamical interactions between the Milky Way and its dwarf satellite galaxies (Thomas & Battaglia, 2023; Savino & Posti, 2019).

Features like tidal tails and stellar streams from disrupted dwarf galaxies and star clusters, as well as the velocity dispersion profiles of the dSph, can be used to understand the accretion history of the Milky Way. Studying these features are necessary for diagnosing the underlying gravitational potential as well as testing the Λ CDM model on small scales (Eyre & Binney, 2009; Law & Majewski, 2010; Pawlowski et al., 2015). The velocity dispersion measurement and other kinematic studies of the dwarf galaxies indicate they are highly dark matter dominated (Mateo, 1998; Walker et al., 2007, 2008; Simon et al., 2011; Koposov et al., 2015b). Most of the kinematic models for these systems in previous work assume they are in dynamical equilibrium. However, being accreted into the MW potential, these systems interact and evolve with their host. A large fraction of them may undergo significant tidal stripping, leading to mass loss and unbound stars contaminating and inflating their kinematic profile (Johnston, 1998; Łokas et al., 2012; Klimontowski et al., 2007). Also, tidal stirring may increase the velocity dispersion of dwarfs, making it difficult to measure the mass profile of the dark matter halo and use the kinematic profile of dwarfs to investigate dark matter particle properties (Kazantzidis et al., 2011). Observational evidence for tidal interactions has already been found in some MW dwarf galaxies including Sagittarius (Newberg et al., 2002; Koposov et al., 2012; Ibata et al., 2020; Vasiliev et al., 2021), Sculptor (Sestito et al., 2023), Ursa Minor (Palma et al., 2003) and Carina (McMonigal et al., 2014). In order to attain a precise constraint on the dark matter from the kinematic structure

of these dwarf systems, investigation on the impact of tidal interactions in more MW dwarfs is urgently needed.

Among the known dwarf galaxies of the MW, the Sextans dwarf spheroidal galaxy (hereafter Sextans) presents a potential case for existing tidal interactions. Discovered by Irwin et al. (1990), Sextans is an old metal-poor system with a high mass-to-light ratio and low surface brightness with a half-light radius at 16.5 arcmin (Kirby et al., 2011; Lokas, 2009; Muñoz et al., 2018). Sextans has a large velocity dispersion, which has been attributed to potential tidal stirring, and a radial gradient in stellar population that suggests an extended star formation history Lee et al. (2009). All of these aspects of Sextans remain under investigation.

The DESI Early Data Release (DESI Collaboration et al., 2023a) includes data from the Milky Way Survey Validation (SV) campaign. The SV campaign (DESI Collaboration et al., 2023b) observed more than 200,000 unique stars. The Sextans dSph was observed on two 8 square degree tiles in the SV campaign. With the addition of heliocentric velocity, metallicity and distance measurements (Cooper et al., 2023; DESI Collaboration et al., 2023a) from DESI in a large area around Sextans, we revisit the properties of the Sextans dSph especially at and beyond the tidal radius.

In addition to measuring the properties of Sextans itself, we search for evidence of Sextans' past dynamical interactions with the MW by searching for stars that are on the same orbit, candidate members of Sextans stripped by previous tidal interactions. For this search we utilize the latest DESI Blue Horizontal Branch (BHB) catalog from Byström et al. (in prep). There are 5461 stars in the BHB catalog from the iron

data release, and their distance ranges from local to 120 kpc. All known Milky Way substructures like the Sagittarius tidal stream are removed from the catalog. BHBs are known as an ideal tracer of the stellar halo Kinman1994,Brown2014,Deason2014. Specifically, the well-constrained color-magnitude relation of BHBs allows us to measure their distance using photometry alone (Preston et al., 1991; Belokurov & Koposov, 2016). This gives us accurate 6D kinematic information to identify potential BHB stars along the orbit of Sextans.

There are recent studies of the Sextans dSph using photometric data and Gaia. Roderick et al. (2016) measure the surface density profile and compare it to an exponential profile. The exponential is typically a good fit to dSph density profiles (McConnachie & Venn, 2020a) but they find that Sextans is significantly more extended, providing further evidence for tidal stripping. Qi et al. (2022) use photometry and Gaia proper motions to identify 34 extra-tidal stars in Sextans, again evidence that Sextans has undergone tidal stripping. Our study investigates Sextans using 3D kinematics by adding radial velocities and distance estimates derived from the DESI spectroscopy (Koposov et al., 2024).

In this work, we use data from Gaia DR3 (Gaia Collaboration et al., 2023) and DESI (DESI Collaboration et al., 2023a) to investigate the outer structure of Sextans and search for potential structures like tidal tails and streams produced by the interaction between Sextans and the MW. This paper is organized as follows. In §5.2, we explain our sample selection. In §5.3, we discuss the Gaussian mixture model (GMM) used to identify candidate stars in Sextans. In §5.4, we illustrate the results from the

Gaussian mixture model, including membership probability, mean line of sight velocity, mean metallicity and mean proper motion. In §5.5, we discuss how we model the orbit of Sextans candidates selected from the Gaussian Mixture Model. In §5.6, we illustrate how we use the Sextans high probability members as training set to select potential Sextans members in the stellar halo from the DESI BHB stars. We use a k-nearest neighbour technique (KNN) to search for potential BHB along the Sextans orbit and compute a t-test to evaluate the KNN performance. In §5.7, we discuss the results of the 6D orbit integration of Sextans candidates and physical implications for the tidal interaction between Sextans and the MW. §5.8 discusses future work.

5.2 Data

The Sextans data used in this study were taken as part of the DESI Science Verification program (DESI Collaboration et al., 2023b). The spectra are included in the DESI Early Data Release (EDR) (DESI Collaboration et al., 2023a). However, we use radial velocities and $[\text{Fe}/\text{H}]$ from a more recent version of the `rvspecfit` pipeline discussed in Cooper et al. (2023); DESI Collaboration et al. (2023a) that will be included in a future DESI data release. This version of `rvspecfit` provides measurements of heliocentric velocity v_{hel} to about 1 km/s accuracy and $[\text{Fe}/\text{H}]$ accuracy to about 0.2 dex (Cooper et al., 2023).

The sample contains data observed on two different DESI tiles. We select targets from Gaia DR3 (Gaia Collaboration et al., 2023) and DECaLS DR9 (Dey et al.,

2019) 1.7 deg from the field center with r magnitude > 16 and $g - r < 1.2$. We then select as follows:

1. $|\text{pm} - \text{pm0}| < 2$, where pm0 is the proper motion ($\mu_\alpha \cos \delta = -0.409_{0.008}^{0.009}$ and $\mu_\delta = 0.037_{0.009}^{0.009}$ from Pace et al. (2022)) of Sextans.
2. $w - 3 \times \sigma_w < 1/\text{distance}$, where w is the parallax and distance is the distance to Sextans.
3. Star/galaxy separation cut:
 - $\text{gaia_astrometric_excess_noise} < 1$
 - $\text{gaia_phot_bp_rp_excess_factor} < 1.3 + 0.06 \times \text{gaia_phot_bp_mean_mag} - \text{gaia_phot_rp_mean_mag}^2$

Each tile is 8 square degrees and targets 5000 objects, including calibration and sky fibers. See section 2.2 in DESI Collaboration et al. (2023a) for more details about DESI observing. The two Sextans tiles were observed at a range of moon illumination and observing conditions. The median exposure time on the stars used for this analysis was 1255.1852 s.

The RA/DEC selection and data quality selection suggested in Kuposov et al. (2024) for our preliminary Sextans sample are as follows.

- RA from 150 to 157
- Dec from -4.5 to 1.5
- $16 < r \text{ mag} < 21$

- $\log g < 4$
- $RVS_WARN = 0$
- $RR_SPECTYPE$ not QSO
- $PHOT_VARIABLE_FLAG$ not VARIABLE
- $RUWE < 1.3$

The total number of stars after this preliminary selection is ~ 600 .

5.2.1 Sextans Sample Selection: Photometry

We use photometric catalogs from the Legacy Survey DR9 (Dey et al., 2019) to improve our selection of Sextans members using color-magnitude cuts. We use an old, metal-poor ($[Fe/H] = -1.5$) Dartmouth Isochrone with a age = 10 Gyr (Dotter et al., 2008) and select stars within $|g - r - Iso(r)| < 0.38$, where Iso is the Isochrone value. The horizontal branch is not included in the Dartmouth Isochrone model. We use photometry from the Legacy Survey (Dey et al., 2019) of the globular cluster M92 to define a region for the horizontal branch of an old stellar population like Sextans. We select a region 0.6 mag wide in g band centered on the cluster horizontal branch ridgeline and add that to the isochrone selection. The final selection yields 577 stars and is shown in Fig. 5.1.

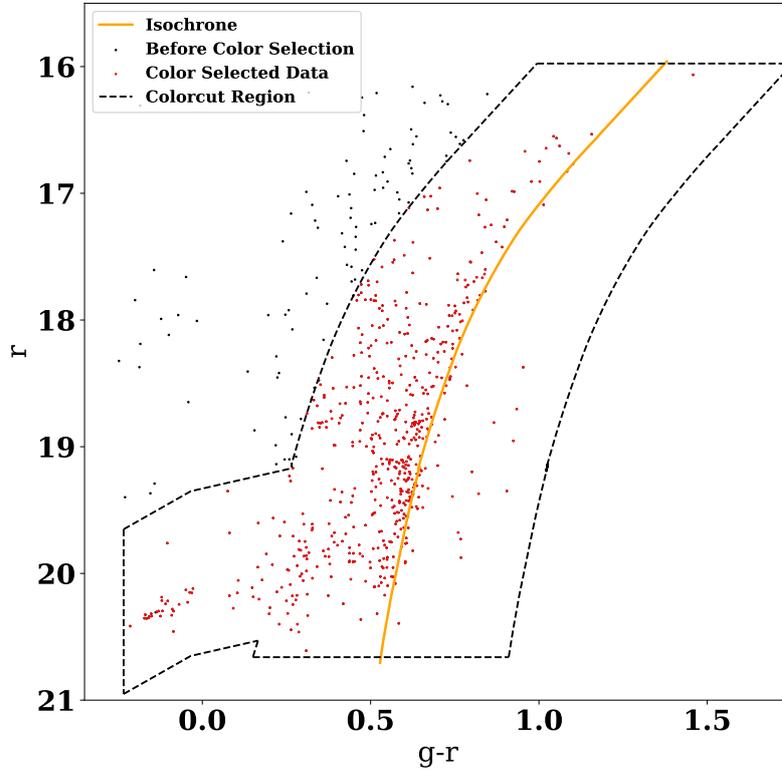


Figure 5.1: The region inside the dashed lines is the final selection for the Sextans candidate sample from the DESI SV data. For the color-magnitude cuts used in the selection, we use an old, metal-poor ($[\text{Fe}/\text{H}] = -1.5$) Dartmouth isochrone with age = 10 Gyr (Dotter et al., 2008) and select stars close to the Isochrone: $g - r \pm 0.38$. We use photometry from the Legacy survey (Dey et al., 2019) of the M92 globular cluster to define a region for the horizontal branch, since that is not included in the isochrone.

5.3 Gaussian Mixture Model to select candidate Sextans members

We use a Gaussian mixture model to identify stars in the Sextans dSph using the combined dataset of v_{hel} and $[\text{Fe}/\text{H}]$ from DESI (DESI Collaboration et al., 2023a) and proper motions from Gaia DR3 (Gaia Collaboration et al., 2023). We use the same method as described in Chapter 4 of this thesis, and we include the description here updated for Sextans.

All the radial velocity and proper motion data are in the heliocentric frame. We model the radial velocity, $[\text{Fe}/\text{H}]$ and the proper motion distributions as a two-component mixture of the satellite and MW foreground/background components. We do not apply a radial density profile in our mixture model in order to avoid biasing our membership probabilities against stars outside the tidal radius. Our model has two components, Sextans and the Milky Way (MW) interlopers, and three parameters: v_{hel} , $[\text{Fe}/\text{H}]$, proper motion, fit for each component. The functional form for the likelihood for each component is:

$$L(\bar{x}, \sigma^2; x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right) \quad (5.1)$$

where \bar{x} is the mean of the physical property and σ the combination of the intrinsic dispersion of the physical property and the internal errors of the data. The two component mixture model is a sum of likelihoods:

$$L = (f_{\text{sat}})L_{\text{satellite}} + (1 - f_{\text{sat}})L_{\text{MW}} \quad (5.2)$$

$$L_{\text{satellite}} = L_{v_{\text{hel}},\text{satellite}}L_{[\text{Fe}/\text{H}],\text{satellite}}L_{\text{PM},\text{satellite}} \quad (5.3)$$

$$L_{\text{MW}} = L_{v_{\text{hel}},\text{MW}}L_{[\text{Fe}/\text{H}],\text{MW}}L_{\text{PM},\text{MW}} \quad (5.4)$$

where f_{sat} is the fraction of stars belong to the dSph. Variables $L_{v_{\text{hel}},\text{satellite}}$, $L_{[\text{Fe}/\text{H}],\text{satellite}}$, $L_{\text{PM},\text{satellite}}$ refer to the heliocentric velocity, metallicity and proper motion likelihood components for the dSph. $L_{v_{\text{hel}},\text{MW}}$, $L_{[\text{Fe}/\text{H}],\text{MW}}$, $L_{\text{PM},\text{MW}}$ denote the heliocentric velocity, metallicity and proper motion likelihood components for the MW foreground/background stars.

We list all the parameters used in the model in table 5.1. The four parameters ($v_{\text{hel},\text{satellite}}$, $[\text{Fe}/\text{H}]_{\text{satellite}}$, $\mu_{\alpha \cos \delta,\text{satellite}}$, $\mu_{\delta,\text{satellite}}$) represent the mean of the physical properties for the galaxy component and $v_{\text{hel},\text{MW}}$, $[\text{Fe}/\text{H}]_{\text{MW}}$, $\mu_{\alpha \cos \delta,\text{MW}}$, $\mu_{\delta,\text{MW}}$ represent the mean for the background. We constrain the dispersion component of the proper motion ($\mu_{\alpha \cos \delta,\text{satellite}}$ and $\mu_{\delta,\text{satellite}}$) for the Sextans dSph using previous measurements. We use the value of 7.9 km/s, the value of the Sextans velocity dispersion measured by Muñoz et al. (2018) converted to the proper motion (0.02 mas/year) using a distance of 86 kpc for Sextans. Therefore, we have two parameters $\sigma_{v_{\text{hel},\text{satellite}}}$, $\sigma_{[\text{Fe}/\text{H}],\text{satellite}}$ for the dSph dispersion and four parameters, $\sigma_{v_{\text{hel},\text{MW}}}$, $\sigma_{[\text{Fe}/\text{H}],\text{MW}}$, $\sigma_{\mu_{\alpha \cos \delta,\text{MW}}}$ and $\sigma_{\mu_{\delta,\text{MW}}}$ for the MW dispersion. Also, as discussed in section 7.4.2 in Cooper et al. (2023), there is an additional systematic component of ~ 0.9 km/s in the radial velocity errors; thus, we apply a 0.9 km/s in quadrature to the radial

velocity errors in our Gaussian Mixture Model.

We model the satellite proper motion component as a multivariate distribution (e.g., Pace et al. (2022)). We include the covariance in the proper motion errors (the cross term in the proper motion error) and intrinsic proper motion dispersion terms.

The log likelihood for this multivariate Gaussian distribution is:

$$L(\bar{\mathbf{x}}, \mathbf{Covm} | \mathbf{X}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{Covm}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{Covm}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \quad (5.5)$$

where $\bar{\mathbf{x}}$ represents the parameters for the mean proper motion ($\mu_\delta, \mu_\alpha \cos \delta$) of Sextans, \mathbf{Covm} represents the covariance matrix, \mathbf{X} represents the observed proper motion data, n represents the number of stars in the sample. For the MW proper motion distribution, we define the intrinsic dispersion by adding the MW dispersion component to the diagonal term of the PM covariance. The covariance matrix \mathbf{Covm} is now defined as:

$$\text{Cov}(\mu_\delta, \mu_\alpha \cos \delta) + \text{Diag}(\sigma_{(\mu_\alpha \cos \delta)}, \sigma_{\mu_\delta}) \quad (5.6)$$

where $\text{Cov}(\mu_\delta, \mu_\alpha \cos \delta)$, is the Gaia covariance matrix for proper motions. For the Sextans component, we add $\sigma_{\mu_\alpha \cos \delta, \text{satellite}}$ and $\sigma_{\mu_\delta, \text{satellite}}$, the mean Sextans proper motion dispersion we fixed as explained above in the diagonal. For the background component, we add $\sigma_{\mu_\alpha \cos \delta, \text{MW}}$ and $\sigma_{\mu_\delta, \text{MW}}$, the parameters fit by the mixture model in the diagonal.

From the above discussion, there are 15 parameters in total for the model (see Table 5.1). In order to estimate the best values for the parameters, we run an MCMC

process based on the affine invariant ensemble sampler for Markov Chain Monte Carlo Goodman & Weare (2010) in the *emcee* library Foreman-Mackey et al. (2013). The prior distributions for all the parameters are as follows (see Table 5.1):

- (i) One uniform prior between 0 and 1 for the fraction f_{sat} defined above
- (ii) Four parameters describing the mean and the dispersion of the heliocentric velocity, two for the velocity distribution of the Sextans component and two for the Milky way component. For Sextans, we use a uniform distribution prior for the mean heliocentric velocity between 50 and 350 and log dispersion between -1 and 3. For the MW foreground/background we use a uniform distribution prior for the mean heliocentric velocity between 50 and 350 and log dispersion between -1 and 3.
- (iii) Four parameters describing the mean and the dispersion of [Fe/H]. For the Sextans component we use a uniform distribution prior for the mean [Fe/H] between -4.0 and -0.2 and log dispersion between -1 and 3 For the MW foreground/background component we use a uniform distribution prior for the mean [Fe/H] between -4 and -0.2 and log dispersion between -1 and 3.
- (iv) Two parameters describing the Sextans Proper motion. We use a uniform distribution prior for the mean proper motion for both $\mu_{\alpha \cos \delta, \text{satellite}}$ and $\mu_{\delta, \text{satellite}}$ between -2 and 2. Note that the dispersion of $\mu_{\alpha \cos \delta, \text{satellite}}$ and $\mu_{\delta, \text{satellite}}$ for the Sextans component are fixed to 0.025.
- (v) Four parameters describing the MW Proper motion. We use a uniform distribution prior for the mean proper motion $\mu_{\alpha \cos \delta, \text{MW}}$ and $\mu_{\delta, \text{MW}}$ between -10 and 10 and log dispersion for $\mu_{\alpha \cos \delta, \text{MW}}$ and $\mu_{\delta, \text{MW}}$ between -1 and 1.3. These priors are summarized

in Table 5.1.

We checked that we attained an acceptance rate of 0.34 from the MCMC chain, which is between 0.3-0.5.

With the v_{hel} , $[\text{Fe}/\text{H}]$ and proper motion likelihood we described above, we can compute the relative likelihood for each star for membership in Sextans and Milky Way. This gives us membership probabilities for the stars in Sextans. The probability that a star belongs to the satellite is:

$$p = \frac{(f_{\text{sat}})L_{\text{satellite}}}{(f_{\text{sat}})L_{\text{satellite}} + (1 - f_{\text{sat}})L_{\text{MW}}} \quad (5.7)$$

where the component $L_{\text{satellite}}$ refers to the dSph and the component L_{MW} describes the foreground/background stars. The variable f_{sat} is the fraction of stars that belong to the dSph, and thus $1-f_{\text{sat}}$ is the fraction of stars that belong to the MW foreground/background.

We summarize the best fitted parameters for the Sextans heliocentric velocity, $[\text{Fe}/\text{H}]$ and proper motion from the mixture model in Table 5.2. The uncertainty on each fit parameter is calculated from the 16th/84th percentile of the posterior distribution. We compare our parameters to previous results in the next section.

With the best fit mean and dispersion for v_{hel} , $[\text{Fe}/\text{H}]$, $\mu_{\alpha} \cos \delta$ and μ_{δ} we use equation (7) to calculate the membership probability P_i for the i th star of the Sextans candidates shown in Fig. 5.1. The distribution of membership probabilities for the Sextans candidates is shown in Fig. 5.2. We define high probability member stars to be those with probability > 0.9 .

Table 5.1. Priors for the Gaussian Mixture Model

Parameters	Prior Range
f_{sat}	0-1
$v_{\text{hel,satellite}}$	50 - 350 km/s
$v_{\text{hel,MW}}$	50 - 350 km/s
$\text{Log } \sigma_{v_{\text{hel,satellite}}}$	-1 - 3
$\text{Log } \sigma_{v_{\text{hel,MW}}}$	-1 - 3
$[\text{Fe}/\text{H}], \text{satellite}$	-3.8 - -0.2
$[\text{Fe}/\text{H}], \text{MW}$	-3.8 - -0.2
$\text{Log } \sigma_{[\text{Fe}/\text{H}],\text{satellite}}$	-1 - 3
$\text{Log } \sigma_{[\text{Fe}/\text{H}],\text{MW}}$	-1 - 3
$\mu_{\alpha \cos \delta,\text{satellite}} / \mu_{\delta,\text{satellite}}$	-2.5 - 2.5
$\mu_{\alpha \cos \delta,\text{MW}} / \mu_{\delta,\text{MW}}$	-2.5 - 2.5
$\sigma_{\mu_{\alpha \cos \delta,\text{satellite}}} , \sigma_{\mu_{\delta,\text{satellite}}}$	fixed 0.02
$\text{Log } \sigma_{\mu_{\alpha \cos \delta,\text{MW}}} , \text{Log } \sigma_{\mu_{\delta,\text{MW}}}$	-1 - 1.3

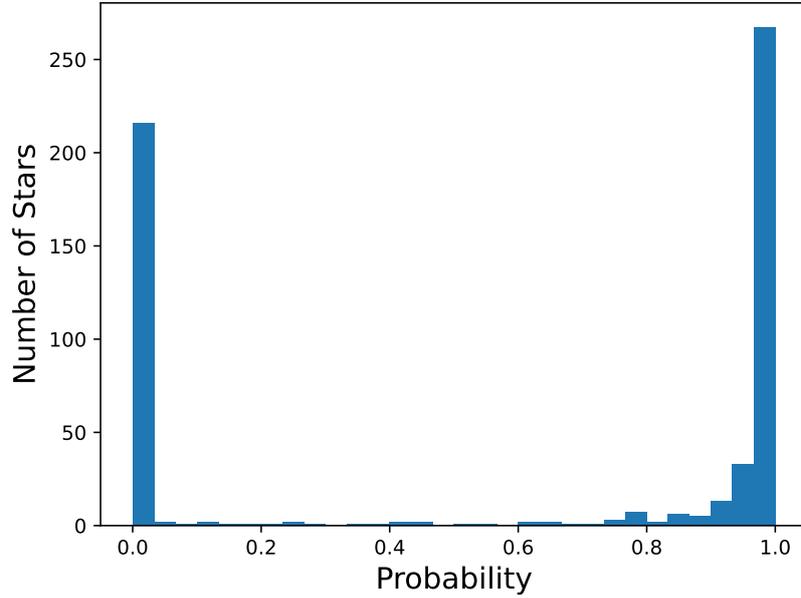


Figure 5.2: Probability distribution for Sextans members identified by the Gaussian mixture model. From the probability distribution, we define high probability members to be $\text{prob} > 0.9$

We plot the candidates, color-coded by their membership probabilities, in the color-magnitude diagram in Fig. 5.3. The ones outside the tidal radius are plotted as large triangles using the same color scale. Most high probability member stars are distributed along the isochrone, showing that stars identified as Sextans members by the mixture model are also likely members based on their location in the CMD.

Table 5.2. Values for the Best Fitted Parameters

Sextans Parameters	Mean value with error
f_{sat}	0.58 ± 0.021
$V_{\text{hel,satellite}}$	225.26 ± 0.62
Velocity Dispersion ($\sigma_{V_{\text{hel,satellite}}}$)	$9.33^{+0.53}_{-0.50}$
[Fe/H]	-2.27 ± 0.04
[Fe/H] _{satellite} Dispersion ($\sigma_{[\text{Fe}/\text{H}]_{\text{satellite}}}$)	0.57 ± 0.03
$\mu_{\alpha \cos \delta, \text{satellite}}$	-0.41 ± 0.01
$\mu_{\delta, \text{satellite}}$	0.04 ± 0.01

Note. — Note that the dispersion of $\mu_{\alpha \cos \delta}$ and μ_{δ} for Sextans are fixed to 0.025 and we do not fit those two parameters.

5.3.1 Spatial Distribution of the Sample

Our Gaussian mixture model does not use spatial information or a density profile constraint, minimizing bias against identifying members near and beyond the tidal radius of the galaxy. We plot the spatial distribution of our sample with membership probability > 0.9 in Fig. 5.4 of the Sextans. We use the orthographic projection of RA and Dec and to the Cartesian frame:

$$X = \cos \delta \sin(\alpha - \alpha_c) \quad (5.8)$$

$$Y = \sin \delta \cos \delta_c - \cos \delta \sin \delta_c \cos(\alpha - \alpha_c) \quad (5.9)$$

where α and δ are the location of a star in the RA and Declination directions, α_c and δ_c are the center coordinates of the dSph. The blue ellipse is the Sextans tidal radius of 60.5 arcmin from Muñoz et al. (2018). The ellipse at the tidal radius is projected in the same way as discussed above, using the ellipticity value of 0.29 from Muñoz et al. (2018).

5.3.2 High Probability Member Stars in the Outskirts of Sextans

We identify several high probability members near and outside the Sextans King tidal radius of 60.5 arcmin based on a King model (Muñoz et al., 2018). There are 52 high probability members outside the tidal radius ellipse, as far as ten half-light radii (16.5 arcmin = 0.413 kpc). Information for those members is listed in Table 5.3. The spatial distribution for the high probability members outside the tidal radius is shown in Fig. 5.4. These extra-tidal stars appear to be preferentially distributed along

the direction perpendicular to the direction of Sextans's orbit. Fig. 5.3 demonstrates that these members (the large triangles) are mostly distributed along the old population isochrone that describes Sextans, further demonstrating their high probability of being members. We have 12 extra-tidal stars in common with those identified by Qi et al. (2022), who use only photometry and Gaia proper motions.

Table 5.3. Properties for the 52 Star Members Outside the Tidal Radius

DESI TARGET ID	RA	DEC	v_{hel}	[Fe/H]	$\mu_{\alpha} \cos \delta$	μ_{δ}	Membership Probability
39627708786806375	153.39	-3.23	225.01 ± 5.55	-3.0 ± 0.34	0.12 ± 0.29	0.08 ± 0.31	0.999
39627708795195171	153.90	-3.24	240.25 ± 3.74	-2.0 ± 0.21	-0.18 ± 0.33	0.35 ± 0.51	0.99
39627714818213887	153.11	-2.89	205.14 ± 3.14	-2.7 ± 0.22	-0.24 ± 0.24	0.13 ± 0.21	0.999
39627714826605804	153.75	-3.09	221.59 ± 6.37	-3.7 ± 0.21	-0.29 ± 0.34	-0.59 ± 0.41	1.0
39627720845428732	152.35	-2.87	217.78 ± 2.77	-2.8 ± 0.13	-0.24 ± 0.35	-0.23 ± 0.44	1.0
39627720866399264	153.56	-2.65	226.75 ± 7.91	-2.8 ± 0.34	0.40 ± 0.44	-0.49 ± 0.48	0.996
39627720870595511	153.91	-2.86	239.12 ± 11.69	-3.4 ± 0.37	-0.83 ± 0.66	1.12 ± 0.77	1.0
39627720870596488	153.96	-2.76	226.91 ± 3.38	-3.2 ± 0.09	-0.11 ± 0.15	-0.14 ± 0.15	1.0
39627726851673736	150.43	-2.55	67.33 ± 58.33	-3.4 ± 0.08	-0.14 ± 0.99	1.16 ± 1.45	0.996
39627726893613809	152.79	-2.45	230.05 ± 2.01	-1.3 ± 0.14	-0.26 ± 0.26	-0.37 ± 0.24	0.918
39627726906197156	153.56	-2.43	234.79 ± 4.95	-2.4 ± 0.39	-0.05 ± 0.36	-0.42 ± 0.46	0.989
39627732941804672	153.49	-2.30	226.06 ± 6.90	-3.1 ± 0.39	-0.61 ± 0.33	0.59 ± 0.46	1.0
39627732945997309	153.66	-2.34	216.39 ± 3.94	-2.8 ± 0.18	-0.24 ± 0.36	0.10 ± 0.46	1.0
39627732945998526	153.71	-2.34	232.77 ± 5.82	-2.3 ± 0.14	-0.07 ± 0.77	0.42 ± 0.80	0.996
39627732945998635	153.72	-2.31	224.29 ± 9.76	-3.2 ± 0.44	0.31 ± 0.94	-1.77 ± 1.15	0.988

Table 5.3. Continued

DESI TARGET ID	RA	DEC	v_{hel}	[Fe/H]	$\mu_{\alpha} \cos \delta$	μ_{δ}	Membership Probability
39627732945999171	153.74	-2.22	221.31 ± 4.91	-2.7 ± 0.30	-0.12 ± 0.52	-0.69 ± 0.55	0.994
39627732954384245	154.09	-2.27	211.30 ± 7.10	-2.6 ± 0.33	0.89 ± 0.79	-0.48 ± 0.69	0.979
39627732962774195	154.65	-2.16	227.15 ± 0.59	-2.4 ± 0.05	-0.41 ± 0.09	0.11 ± 0.09	1.0
39627738994184231	154.21	-1.96	215.23 ± 4.04	-2.1 ± 0.17	-0.72 ± 0.36	0.41 ± 0.39	0.997
39627738998376221	154.34	-1.89	222.79 ± 3.00	-3.1 ± 0.18	-0.39 ± 0.21	-0.38 ± 0.31	1.0
39627738998378719	154.47	-1.99	206.92 ± 1.93	-1.6 ± 0.13	-0.16 ± 0.26	-0.15 ± 0.35	0.944
39627739019346351	155.52	-1.99	223.61 ± 5.35	-2.8 ± 0.27	0.01 ± 0.75	0.24 ± 1.06	0.998
39627744996233649	151.99	-1.71	232.60 ± 5.73	-2.8 ± 0.38	-0.14 ± 0.45	-0.15 ± 0.54	0.999
39627745033981576	154.17	-1.80	234.21 ± 5.80	-2.2 ± 0.31	-0.26 ± 0.27	0.01 ± 0.25	0.999
39627751044420319	152.45	-1.44	222.93 ± 2.46	-2.6 ± 0.14	-0.36 ± 0.17	-0.12 ± 0.15	1.0
39627751073780223	154.23	-1.52	216.66 ± 2.14	-2.4 ± 0.17	-0.21 ± 0.15	-0.16 ± 0.21	0.999
39627751077971286	154.32	-1.46	224.31 ± 1.61	-2.0 ± 0.09	-0.51 ± 0.16	0.18 ± 0.22	1.0
39627751082166143	154.60	-1.50	206.92 ± 6.03	-2.9 ± 0.46	0.49 ± 0.77	1.43 ± 0.82	0.993
39627757088409181	152.58	-1.27	239.13 ± 6.26	-2.9 ± 4.32	-0.67 ± 0.85	0.60 ± 0.51	0.989
39627757088409186	152.58	-1.19	222.98 ± 10.51	-2.1 ± 1.46	-1.01 ± 0.76	0.55 ± 0.64	0.987

Table 5.3. Continued

DESI TARGET ID	RA	DEC	v_{hel}	[Fe/H]	$\mu_{\alpha} \cos \delta$	μ_{δ}	Membership Probability
39627757117770739	154.43	-1.20	230.91 ± 13.97	-3.1 ± 0.96	-0.68 ± 1.23	0.86 ± 1.22	0.983
39627757126157630	154.82	-1.19	220.62 ± 0.95	-2.5 ± 0.06	-0.41 ± 0.11	0.11 ± 0.11	1.0
39627763115626579	151.92	-0.92	227.17 ± 3.91	-2.0 ± 0.17	-0.36 ± 0.52	-0.87 ± 0.77	0.938
39627763128208532	152.66	-0.96	221.23 ± 7.08	-2.9 ± 0.45	-0.43 ± 0.34	-0.46 ± 0.31	0.994
39627763128208822	152.67	-1.07	227.49 ± 7.32	-1.8 ± 0.34	-0.58 ± 0.52	0.22 ± 0.48	0.991
39627763128209967	152.75	-1.07	236.37 ± 2.71	-2.8 ± 0.16	-0.63 ± 0.21	-0.11 ± 0.21	1.0
39627763149180928	153.94	-0.88	204.36 ± 15.52	-2.3 ± 0.44	-0.07 ± 0.80	-0.56 ± 0.84	0.922
39627763157567569	154.36	-1.06	209.71 ± 5.01	-2.4 ± 0.41	0.70 ± 0.41	-0.07 ± 0.37	0.931
39627769163811832	152.39	-0.79	219.82 ± 5.92	-3.4 ± 0.36	-0.85 ± 0.67	0.69 ± 0.53	1.0
39627769168003521	152.52	-0.87	221.30 ± 5.53	-3.0 ± 0.34	-0.84 ± 0.33	0.19 ± 0.29	1.0
39627769172198868	152.81	-0.74	231.20 ± 5.14	-3.3 ± 0.12	-0.18 ± 0.35	-0.07 ± 0.43	1.0
39627769172201274	152.92	-0.80	222.51 ± 9.02	-2.4 ± 0.15	-1.11 ± 0.70	0.62 ± 0.84	0.996
39627769188974921	153.76	-0.73	246.47 ± 8.05	-2.3 ± 0.26	-1.25 ± 0.78	1.05 ± 0.76	0.985
39627769188979685	153.96	-0.81	236.39 ± 6.32	-2.9 ± 0.62	0.42 ± 0.78	-0.71 ± 1.03	0.989
39627769193169433	154.03	-0.69	236.46 ± 7.00	-3.5 ± 0.28	-0.56 ± 0.42	0.17 ± 0.47	1.0

Table 5.3. Continued

DESI TARGET ID	RA	DEC	v_{hel}	[Fe/H]	$\mu_{\alpha} \cos \delta$	μ_{δ}	Membership Probability
39627769193171580	154.14	-0.83	229.39 ± 5.54	-2.0 ± 0.28	-0.23 ± 0.32	-0.09 ± 0.41	0.996
39627769197363872	154.28	-0.82	218.82 ± 11.90	-2.8 ± 0.91	-1.52 ± 1.05	0.36 ± 1.07	0.965
39627775211997526	152.87	-0.52	229.10 ± 2.37	-2.8 ± 0.16	-0.21 ± 0.20	-0.13 ± 0.20	1.0
39627775241357741	154.62	-0.55	229.17 ± 1.79	-2.3 ± 0.13	-0.44 ± 0.18	0.06 ± 0.15	1.0
39627781289546092	155.21	-0.28	234.35 ± 4.73	-1.3 ± 0.20	-0.46 ± 0.45	-0.29 ± 0.49	0.91
39627787304175686	153.59	0.07	218.38 ± 6.82	-2.7 ± 0.41	-0.21 ± 0.87	-0.74 ± 1.05	0.988
39627793331390597	152.87	0.13	232.39 ± 8.29	-1.9 ± 0.12	0.07 ± 0.73	1.49 ± 0.72	0.99

5.3.3 Gaussian Mixture Model Sample Validation

We have 313 stars in total identified by the Gaussian Mixture Model (GMM) as high probability Sextans members. To evaluate these high probability stars from the GMM selection, we compare our measurement of velocity, velocity dispersion, metallicity and mean proper motion (summarized in Table 5.2) to previous values from the literature. Our measurements for the Sextans proper motion components are $\mu_\alpha \cos \delta = -0.41 \pm 0.01$ and $\mu_\delta = 0.04 \pm 0.01$. Figure 5.5 shows our measurements are consistent with previous studies of Sextans that used proper motions and photometry from Gaia DR2 (McConnachie & Venn, 2020b), and Gaia EDR3 (McConnachie & Venn, 2020c; Qi et al., 2022; Pace et al., 2022).

We find a mean line of sight heliocentric velocity $v_{\text{hel}} = 225.26 \pm 0.62$ km/s with a dispersion $= 9.33^{+0.53}_{-0.53}$ km/s. Our value is consistent with the heliocentric velocity measured in Pace et al. (2022) (224.3 ± 0.1 km/s) and Muñoz et al. (2018) ($v_{\text{hel}} = 224.3$ km/s). Our radial velocity dispersion measurement ($\sigma_{v_{\text{hel}}} = 9.33 \pm 0.53$) is higher than the values quoted in previous work: 7.9 ± 1.3 km/s in Massari et al. (2019), 8.4 ± 0.4 km/s from Battaglia et al. (2011) and 7.9 ± 1.2 km/s in Muñoz et al. (2018). This is also higher than the measurement of the radial velocity as a function of radius from Walker et al. (2008). The high velocity dispersion of our GMM sample may indicate our sample is contaminated by foreground MW halo stars. The MW background components fitted by the GMM model has a mean velocity $= 182.36 \pm 4.91$ and $[\text{Fe}/\text{H}] = -1.4 \pm 0.04$, indicating that the Milky Way foreground excluded by the GMM are mainly from thick

disk stars, which have a mean velocity around 180-192 km/s (Anguiano et al., 2020). This may suggest that we may have some other contamination from halo stars.

Our best-fit value for the mean metallicity $[\text{Fe}/\text{H}]$ (-2.27 ± 0.04) is lower than the mean value found by Kirby et al. (2011) (-1.94 ± 0.01), again indicating the existence of halo star contaminants in our sample of high probability Sextans members.

5.4 Orbits of Candidate Sextans Members

As seen in Figure 5.4, there are 52 extra-tidal stars selected as high probability members in Sextans by the GMM analysis. To further evaluate all the candidate Sextans members, especially the extra-tidal members, we integrate the orbits of our Sextans candidates using 6D kinematic information from Gaia and DESI. To further investigate the previous history of tidal interactions and stripping, we also integrate orbits of the DESI BHB star sample from the whole halo to search for stars that match the orbit of Sextans and might have been tidally stripped from it.

However, from the CMD plot of our high probability Sextans members in Fig. 5.3, it is clear that a small fraction of the stars far from the isochrone are selected as Sextans members from the Gaussian Mixture Model. The velocity dispersion we measure is higher than previous measurements (see discussion in §5.3.3), and we measure a lower metallicity for Sextans than previous work. Sextans is a faint, distant, low surface brightness dwarf galaxy, and our Gaussian mixture model does not include distance estimates. Therefore, we also use orbit information to investigate whether our Sextans

sample is significantly contaminated by foreground Milky Way stars.

5.5 Orbit Integration

The 6D orbital information is computed using the spatial information from the photometry, the radial velocity and distance measured from the DESI spectra and the proper motion measurement from Gaia (Gaia Collaboration et al., 2023). The distance measurement used to compute the 6D kinematic information is from Koposov et al. (private communication). Since the pericenter of Sextans is estimated to be around 82.2 kpc in Pace et al. (2022), we discard from our list of candidate Sextans members used for further analysis all those with distance < 35 kpc. We also impose a limit on the fractional distance error ($< 15\%$) to rule out those stars with poorly constrained distance. We call this sample the "GMM orbit sample" hereafter. It contains 43 stars out of the 313 high probability members from our GMM analysis and 6 of the extra-tidal stars out of 52. See Fig. 5.7 for the distance distribution of the Sextans candidates that pass these criteria. In Fig. 5.6, it is clear that the foreground contaminants in our high probability sample from the GMM analysis are mainly metal poor stars, and they likely bias the metallicity measurement from the GMM to a much lower value. To be consistent, we apply the same distance criteria to the BHB sample. The BHB sample used in orbit integration is named our "BHB candidates" hereafter, and has 1258 stars.

We use `Galpy` (Bovy, 2015) to integrate the orbits. For the Milky Way potential, we start with the default potential `MWPotential2014` (Bovy, 2015), where the MW

consists of an NFW halo, a bulge, and a disk. We use a left-handed coordinate system where the disk has positive angular momentum. This potential is light and the orbits integrated from this potential do not agree with the latest measurements (Pace et al., 2022; McMillan, 2017). To relieve this tension, we scale this potential to have a similar halo mass as McMillan (2017). The rotation curves from the two model potentials are shown in Fig. 5.8. At distances greater than 20 kpc, far away from the location of the Sextans pericenter (~ 82 kpc (Pace et al., 2022)), they have the same rotation curve.

5.5.1 Computing the Integrals of Motion

We integrate the orbits of the high probability Sextans members back in time for 7 Gyrs, using 7 Myr time steps, enough time for the Sextans center of mass to go through almost two complete orbits. We use 6D kinematic information (RA, DEC, V_{hel} , d_{hel} , $\mu_{\alpha \cos \delta}$, μ_{δ}), where d_{hel} is the heliocentric distance. We then create 1000 random realizations of the data drawn from distributions defined by the measurement errors on each of these quantities to derive errors on the properties of the orbits. We repeat the same orbit integration process for the BHB stars using 100 realizations to derive the errors.

The integral-of-motion quantities orbital energy E , angular momentum along the z-axis L_z , vertical action (oscillatory motion of a star perpendicular to the Galactic plane) J_z and radial action (oscillatory motion in the radial direction within the Galactic plane) J_r are calculated using the `Galpy` package (Bovy, 2015). The resultant energy E versus L_z plot for all the BHB stars in the sample is shown in Fig. 5.9. The Gaia-

Sausage-Enceladus structure is revealed in the low E and low L_z space in the plot. E - L_z for the GMM orbit sample of Sextans candidates and the BHB candidate sample are shown in Fig. 5.10. The vertical and radial action (J_z and J_r) plot for the two samples is shown in Fig. 5.11. The Sextans center of mass is shown in orange in each plot, and is computed using the median distance estimated from the GMM orbit sample to be consistent.

5.6 K-Nearest Neighbour Classification of Sextans Candidates

There are two main goals to achieve with the GMM orbit sample and BHB candidate sample:

1. Further confirming the Sextans members from the GMM orbit sample.
2. With the population of Sextans member orbits described in the 4D integral of motion space, we aim to identify a similar population of BHB stars in the halo that are potentially tidally stripped stars from Sextans.

Introduced by Fix & Hodges (1951), KNN is a classical and powerful tool used in classification. It classifies unknown data points by finding the most common class among the k closest examples, where k is a hyper-parameter. This technique has been widely used in star and galaxy classification (de Beurs et al., 2022; Mukundan et al., 2024; Zeraatgari et al., 2024). The integral-of-motion quantities provide a four dimensional description of the orbital motion of the Sextans candidates from our GMM

analysis and of the BHB stars. Starting with the four quantities (E, L_z, J_z and J_r), we can use the KNN to create a more pure sample of Sextans candidates from the GMM orbit sample and search for potential Sextans members in the BHB sample.

We first generate a training set for the KNN using mock observations in the 4D space of the orbit integrals of motion. We draw from the distribution of errors of the four quantities (E, L_z, J_z). We create the BHB sample mock observation using 100 random draws per star and the Sextans mock observation sample using 1000 random draws per star. These mock observations are used as training set. The contour plot describing the the stellar density of our training set on the $E-L_z$ plane is shown in Fig. 5.12. It is clear that the training sets well represent the two samples. For the GMM orbit sample training set, we exclude those stars with $L_z < 0$ since these are stars with large errors creating dissociated small contours far from the center of mass of Sextans, and since Sextans members should have a similar rotation direction reflected in $L_z > 0$.

5.6.1 KNN Classification Results

The KNN identifies 25 out of 43 stars in the GMM orbit sample and 25 stars out of 1258 in the BHB candidates as members of Sextans. The $E-L_z$ plot for the KNN selected Sextans members from the BHB candidate sample and KNN selected Sextans members from the GMM orbit sample is shown in Fig. 5.13. The 25 candidates Sextans members from our GMM orbit sample that are classified as Sextans members by the KNN are grouped around the Sextans center of mass, indicating the KNN classification is accurate. From Fig. 5.14 and Fig. 5.15, the stars seem to be more spread in the J_r

direction, but in general most of the data are clustered around the Sextans center of mass in the 4D integral of motion space.

We also perform a t-test evaluation on the E , L_z , J_z and J_r distribution for the KNN-selected Sextans members from the two samples. The p-value from the t-test is 0.23, significantly greater than 0.05, suggesting that these two samples occupy similar spaces in the 4D integral of motion space.

5.7 Discussion

Through implementing the KNN selection using the E , L_z , J_r , J_z of the orbit sample and the BHB candidates sample, we identify a promising sample for the Sextans members from the GMM orbit sample and the BHB candidate sample. Most of the KNN selected GMM orbit sample have similar orbits as the Sextans center of mass, as do 60% of the KNN selected BHBs. More detailed discussion of these results is below.

5.7.1 KNN Sextans members from the GMM orbit sample

A useful way to validate the selected Sextan members from the GMM orbit sample is to trace their orbit in the R-Z plane. In top panel of the Fig. 5.16, we show the R-Z orbits for the members selected from KNN with the orbit of the Sextans center of mass. The star symbols note the current position of the stars. Only five out of the 25 KNN detected Sextans candidates have orbits that seem to be different than the orbit of the center of mass. Those stars are located at a larger radii with much larger errors in distance and velocity compared to the rest of the 20 members identified by the KNN,

providing possible reasons for why they were selected. The orbits from the rest of the 20 stars agree with the orbit of the center of mass taking into account the errors. The bottom panel in Fig. 5.16 shows the orbits of the stars from the GMM orbit sample there were excluded by the KNN. They do not resemble the apocenter, pericenter or shape of the Sextans center of mass orbit, indicating the KNN is able to exclude stars with orbits significantly discrepant from that of the Sextans center of mass. The two panels of Fig. 5.16 suggest that the KNN is capable of distinguishing all the true negative classifications in the GMM orbit sample, but may have some false positive detections, especially for those are likely to be stars with poor distance or velocity measurements.

5.7.2 Extra-tidal Stars from the GMM orbit sample

There are 52 extra-tidal stars initially selected by the GMM. Six of these extra-tidal stars in the GMM orbit sample pass those distance and distance error cuts to be included in our GMM orbit sample. Fig. 5.20 shows the distance modulus distribution for the original 52 extra-tidal stars. There are about 16 foreground stars with distance less than 35 kpc that fail our first distance cut and over 30 stars with a large uncertainty in their distance measurements that fail our second cut. Of the six extra-tidal stars in the GMM orbit sample, three of them are excluded by the KNN. The top panel in Fig. 5.16 illustrates that of the three extra-tidal stars that are selected by the KNN, one star has very similar orbit as the Sextans center of mass. The other two extra-tidal stars selected by the KNN have unbounded orbits. One of the two false positive stars has a current position near the Sextans center of mass but has a three times larger error in

velocity, which may explain why the KNN selects this target. The other unbound star is at a large distance from the current position of the Sextans CM and has a distance error twice as large as the other KNN selected stars. Improving the KNN performance in the case of larger errors is needed.

5.7.3 Sextans member stars from the BHB candidates

We repeat the same R-Z plane orbit plot for the members selected by KNN from the BHB candidate sample. The orbits from BHBs are more diverse, as shown in the top panel of Fig.5.17), though some are close to that of Sextans. Fig. 5.19 shows the metallicity $[\text{Fe}/\text{H}]$ distribution for these KNN-selected BHBs and for the GMM orbit sample selected by the KNN. Note that although we did not consider $[\text{Fe}/\text{H}]$ in the KNN classification, the Sextans and BHB stars classified as members of Sextans by the KNN share a close mean value in $[\text{Fe}/\text{H}]$ (-1.95), indicating that they may have the same origin.

To investigate the KNN classifications further, we look at the apocenter and pericenter of the BHBs selected by the KNN. In Fig. 5.18, the apocenter-pericenter relation shows a distinct correlation, especially for stars with good distances estimates. This indicates that the KNN is not giving random false positives, but is classifying stars as Sextans members with a pattern that we can investigate in order to improve the purity of the sample. The blue box in this figure includes stars with apocenter from 50 kpc to 150 kpc and pericenter from 25 kpc to 100 kpc. We use this box to select BHB star orbits in the KNN selected sample that have apocenter and pericenter close to the

Sextans center of mass orbit. This selects 15 out of 25 of the KNN-selected BHBs.

The BHBs selected from the apocenter-pericenter relation (blue box in Fig. 5.18) have comparable orbits to the Sextans center of mass. The bottom panel of Fig. 5.17 presents the orbits for these 15 promising BHB Sextans members candidates (we refer to this as our gold sample) with orbits resembling that of Sextans. The median proper motion errors for BHBs in the gold sample are 38% smaller than the 10 BHBs with orbits qualitatively different from Sextans, which may explain the false positive detection of 10 BHBs from KNN.

5.7.4 Difference in extra-tidal structure around Draco and Sextans

We identify 16 extra-tidal BHBs and other giants from Sextans through orbit integration, indicating that Sextans may have undergone some tidal stripping or other kinematic process. Comparing to the case study of Draco (See Chapter 4), which has a comparable orbit to Sextans with similar ellipticity and f_{peri} (see definition in Chapter 4.5.1), Draco does not exhibit obvious evidence for the tidal effect. Sextans is located at a comparable distance from the Galactic center as Draco, but with a larger pericenter compared with Draco, suggesting that it should experience few tidal effects from the MW. This is contradict to our observations in the study. Therefore, it is possible that the extra-tidal features we identify in Sextans are from different mechanism than tidal disturbance. Simulations from Deason et al. (2014a) suggest that distant dwarfs may experience mergers – the merger fractions doubles for dwarfs outside the virial radius of their host galaxy. Also, recent observations (Deason et al., 2014c; Yang et al., 2022;

Gao et al., 2023) indicate satellites may fall into their host as a group, and thus dwarf-dwarf merging may happen before or during their in-fall into the MW. This raises the possibility that Sextans may have been involved in a merger before its infall. The dynamical heating from the merger activities contribute to the loss of those extra-tidal stars. This scenario may be further supported by the extended star formation history between the center and outer regions of Sextans (Lee et al., 2009).

5.8 Conclusion and Future work

We identify 50 (25 from GMM orbit sample and 25 from BHB candidate sample) potential Sextans members from the 6D kinematic information. After investigating the orbits, 20 out of 25 KNN selected stars from the GMM orbit sample and 15 out of 25 selected from the BHB sample have comparable orbits to the Sextans center of mass. Sextans is a diffuse, low surface brightness dwarf at large distance. From the orbits and metallicity of the potential Sextans member BHB stars, we may conclude that Sextans seems to have undergone some dynamical heating, possibly in a merger before its infall. The success of our technique in identifying extra-tidal members in Sextans shows promise for using orbit integration information with a KNN like machine learning technique to search for tidally stripped or heated members of Milky Way dwarfs.

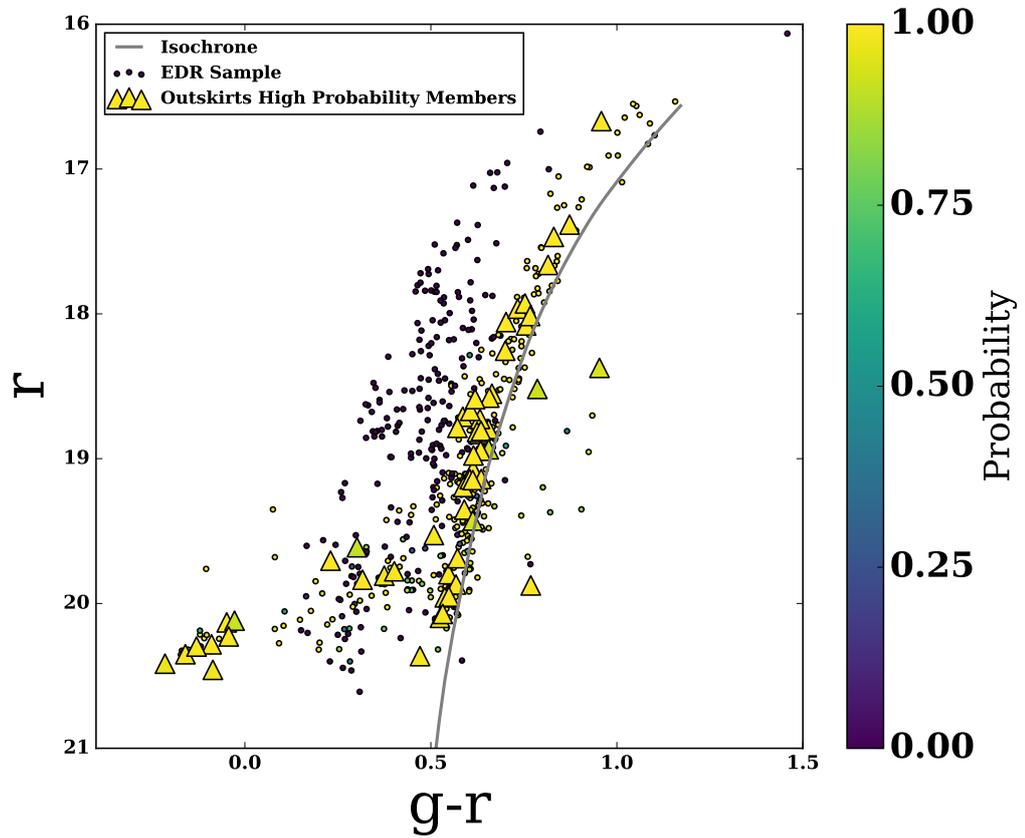


Figure 5.3: Color magnitude diagram with membership probabilities from the mixture model. Points are colored according to their probability. The high probability members outside the tidal radius are labelled with triangles. The grey isochrone is defined in section 5.2.1.

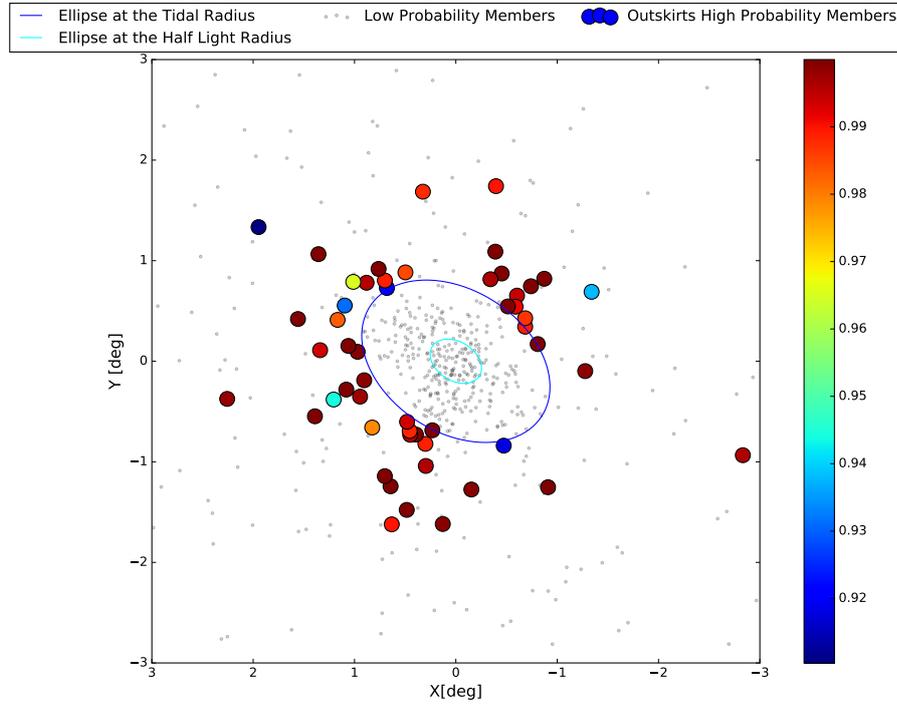


Figure 5.4: Spatial distribution of the high probability Sextans stars. The Sextans Kind tidal radius is the outer blue ellipse, and the inner cyan ellipse is the half-light radius. The 52 high probability stars outside the tidal radius are shown in larger circles. The colorbar shows their membership probability.

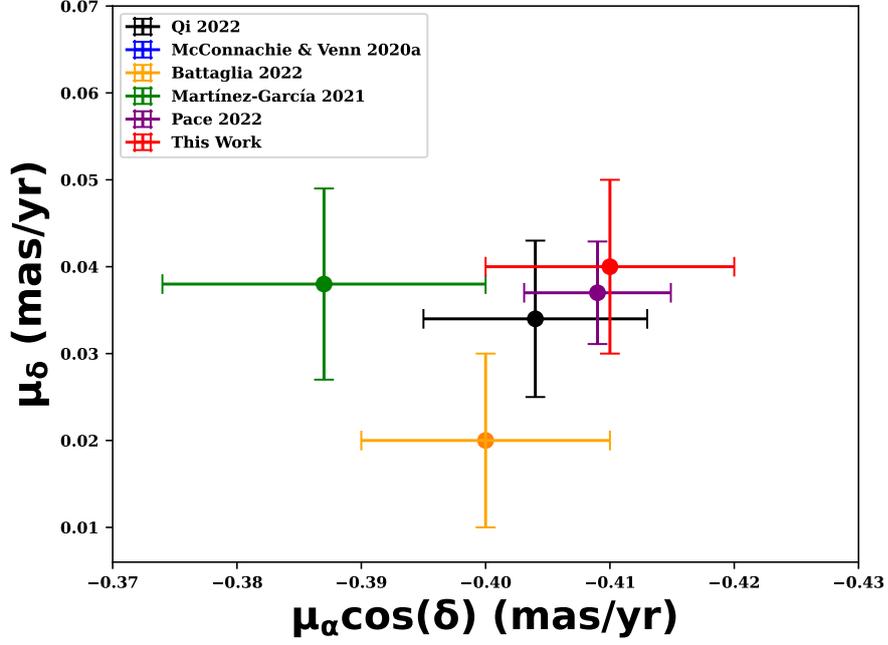


Figure 5.5: Comparison of proper motion (PM) measurement between this work and previous literature. The red error bars show the Sextans PM with 1-sigma uncertainties measured in this work. The black error bars are the PM value and uncertainty measured by Qi et al. (2022) using Gaia DR3 data and photometry. The blue error bars show the PM measured by McConnellie & Venn (2020b). The orange data represents PMs from Battaglia et al. (2022), and the green crosses show PMs from Martínez-García et al. (2021). Our PM distributions are in consistent with other PM values.

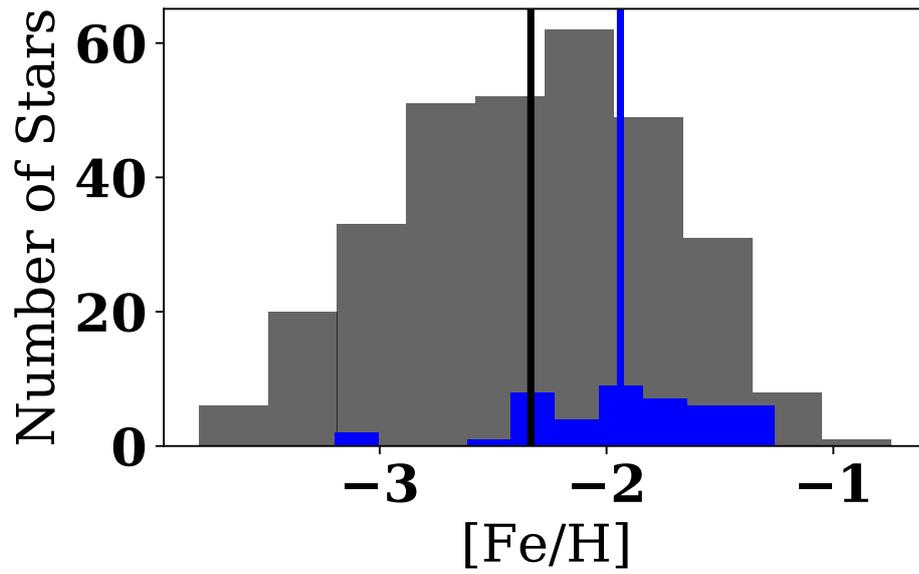


Figure 5.6: The metallicity distribution before and after the distance cut and distance percentage error cut for the GMM high probability sample. The mean metallicity before the distance cut and distance percentage error cut is labelled in black, and the mean metallicity after the cut is labelled in blue.

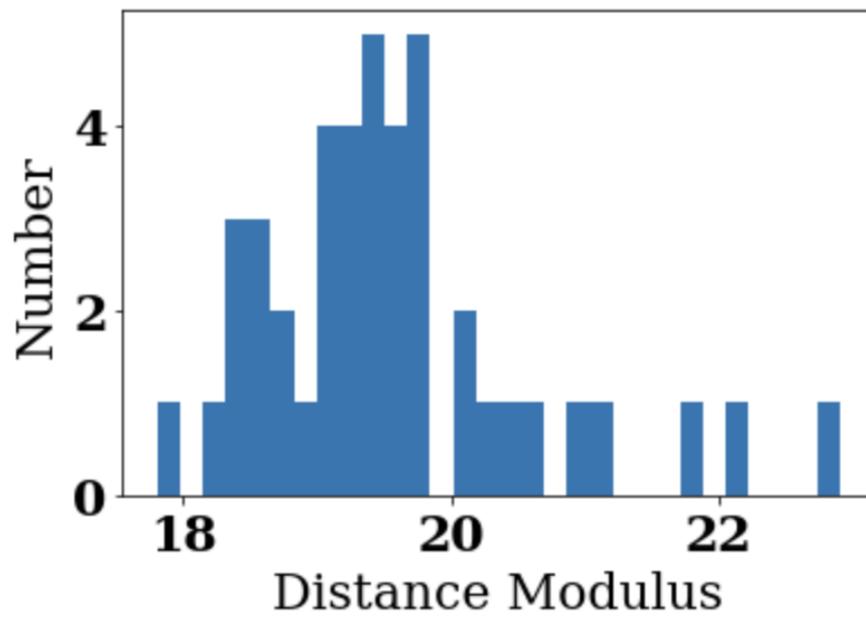


Figure 5.7: Distance modulus distribution for the GMM orbit sample. The distance modulus for the center of mass is at 19.404.

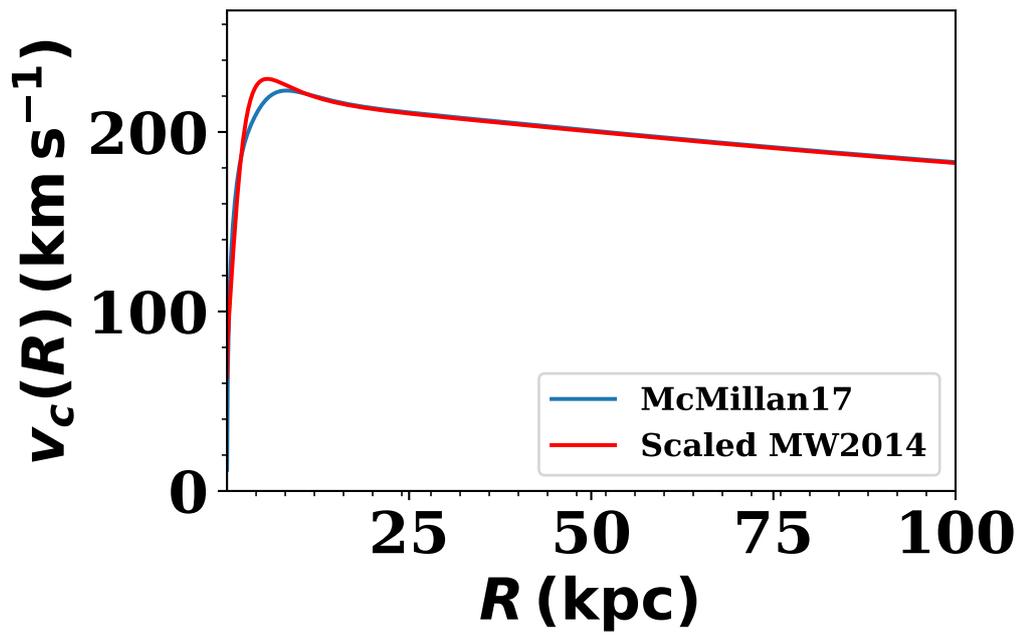


Figure 5.8: Rotation curves for the scaled potential from the MWpotential2014 used in our study (red) compared to the potential in McMillan (2017) (blue).

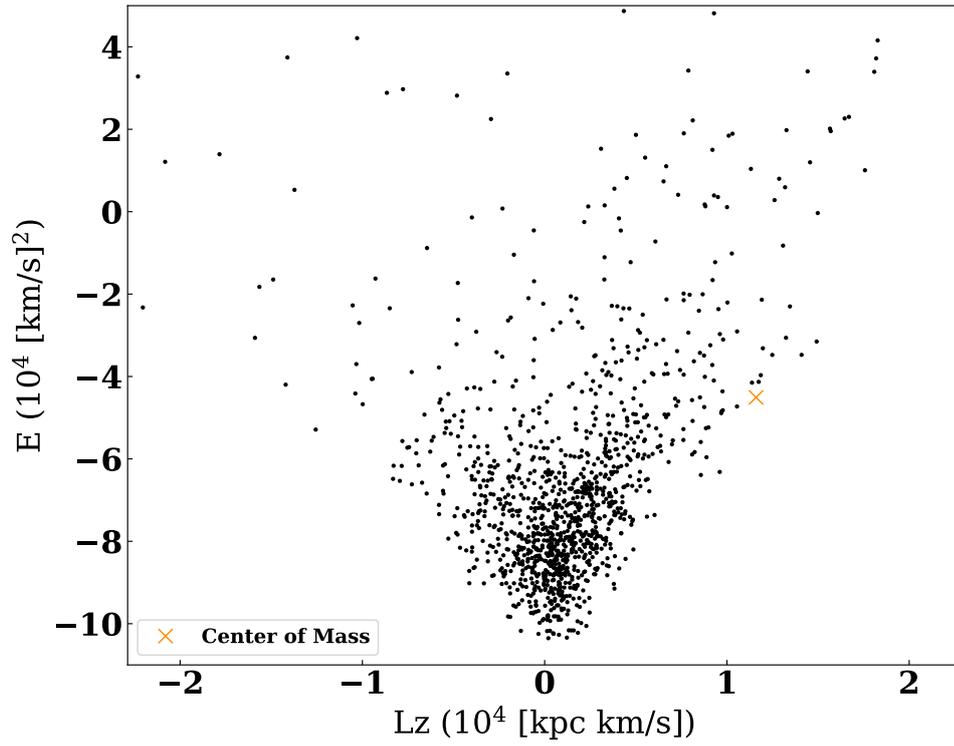


Figure 5.9: E - L_z plot for all stars in the BHB candidate sample. GSE is at low (very negative) energy and near zero L_z , and the sextans CM is the orange X label.

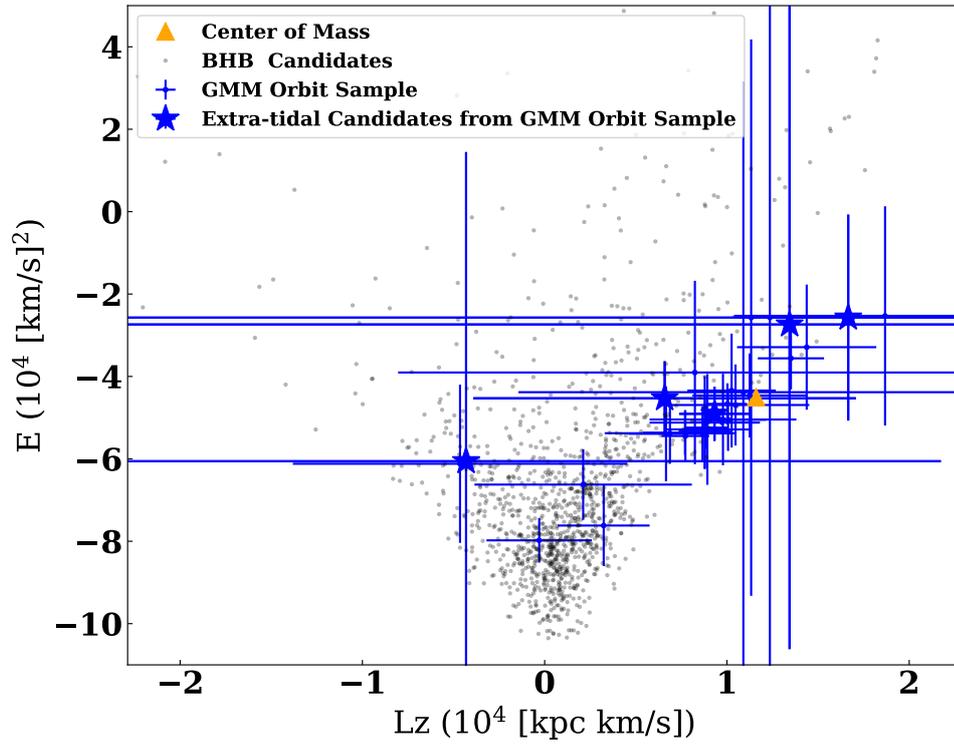


Figure 5.10: E - L_z plot for all the BHB candidates (black) and the GMM orbit sample (blue). Most of the GMM orbit sample have a positive L_z , the Sextans CM is the big orange triangle and most of the GMM sample are near the Sextans CM.

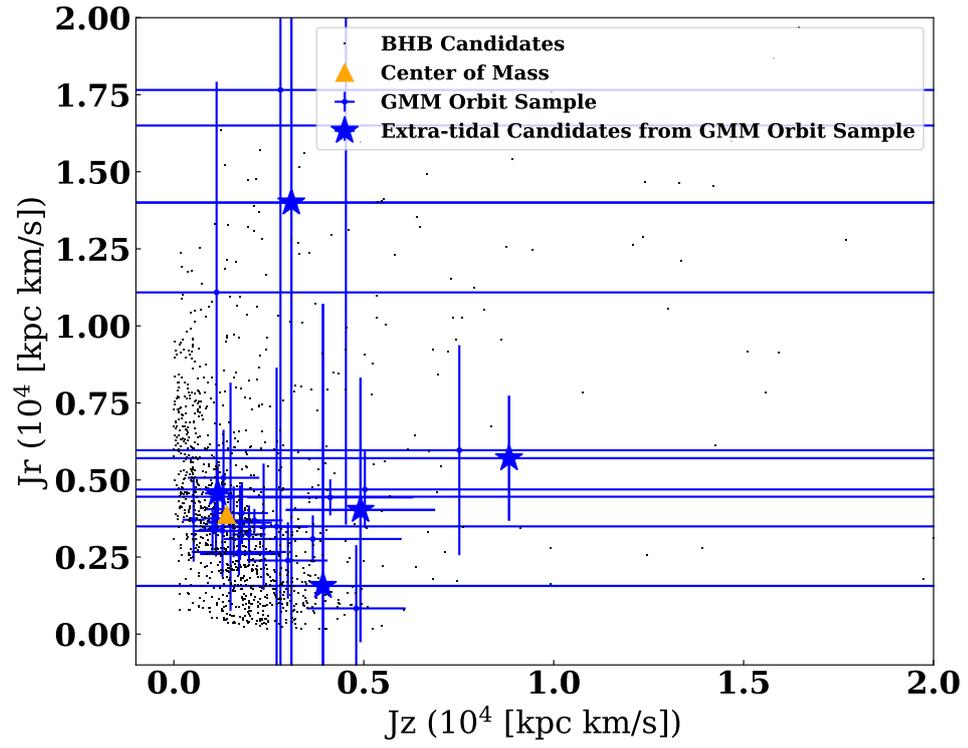


Figure 5.11: J_r - J_z plot for all the BHB candidates labelled in black, GMM orbit sample labeled in blue. There is a cluster around the center of mass.

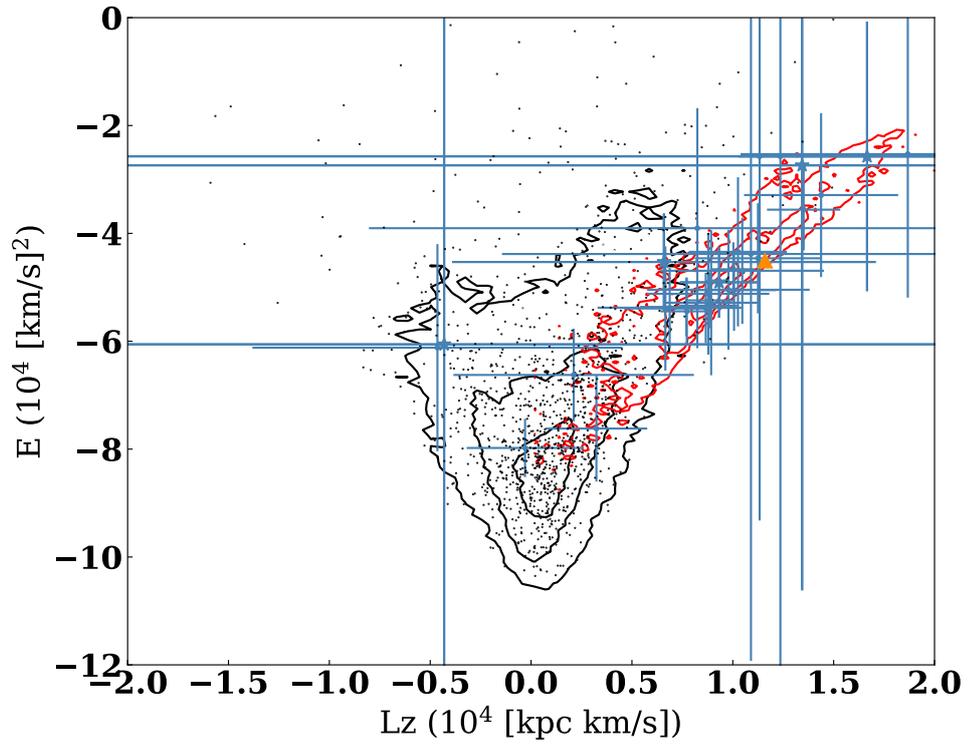


Figure 5.12: E - L_z plot for the contours (16th, 50th 84th) of the training sample for the KNN. The contours for the BHB candidates are labeled in black and the GMM orbit sample are labeled in red. The blue data is the original GMM orbit sample.

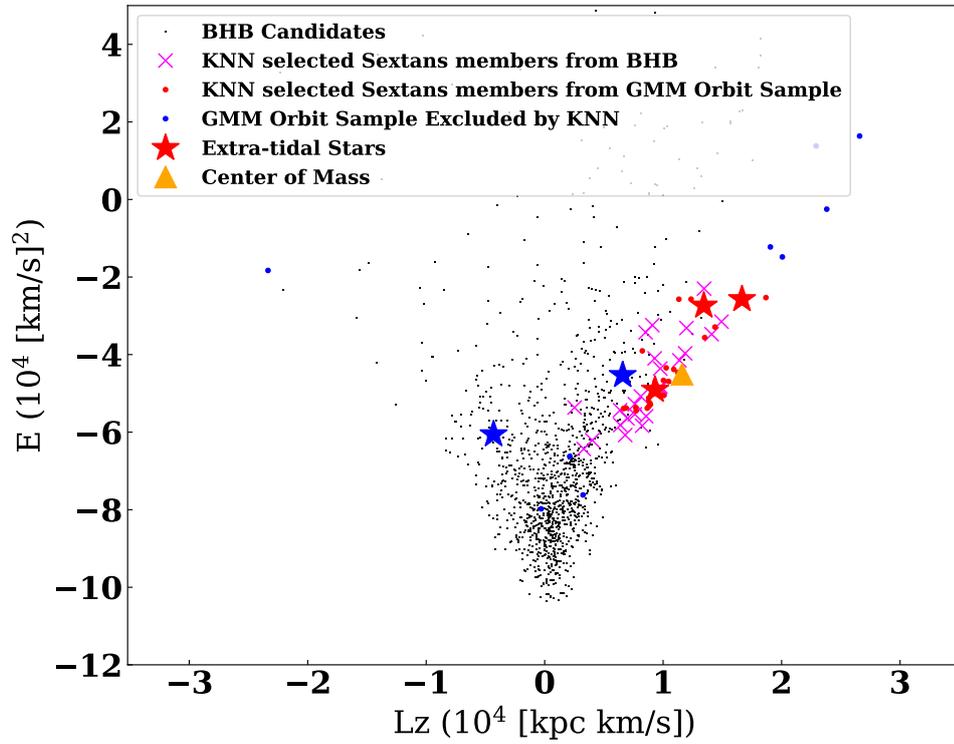


Figure 5.13: $E-L_z$ plot for the performance of the KNN. The GMM orbit sample selected by KNN is labelled in red while the excluded ones are labelled in blue. The BHBs selected by KNN are labelled in purple cross. It is clear that the selected stars are clustered around the center of mass.

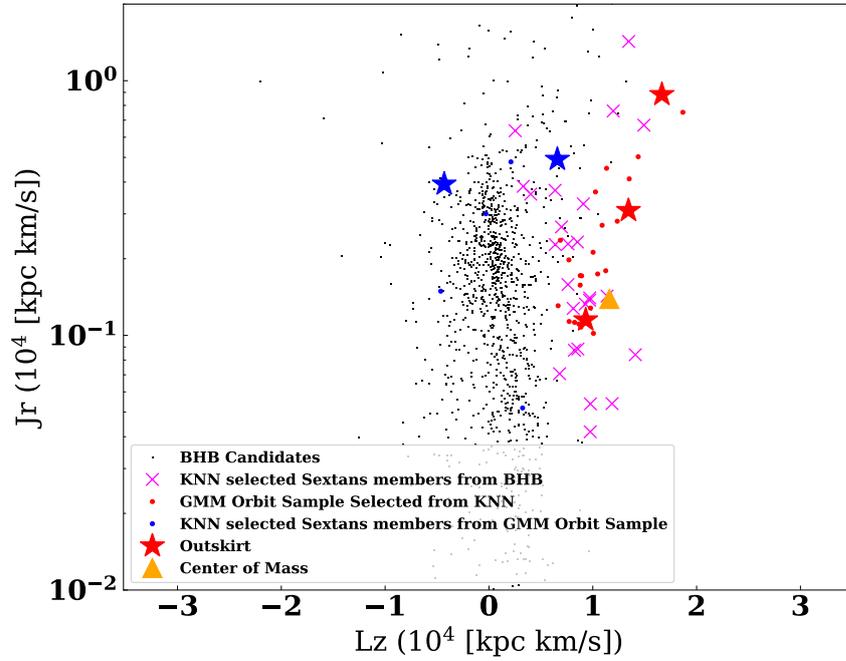


Figure 5.14: J_r - L_z plot for the performance of the KNN. The GMM orbit sample selected by KNN are labelled in red while the excluded ones are labelled in blue. The BHBs selected by KNN are labelled with purple cross. It is clear that the selected stars are clustered around the Sextans center of mass, in L_z from 0-20000 and in J_r from 300-10000.

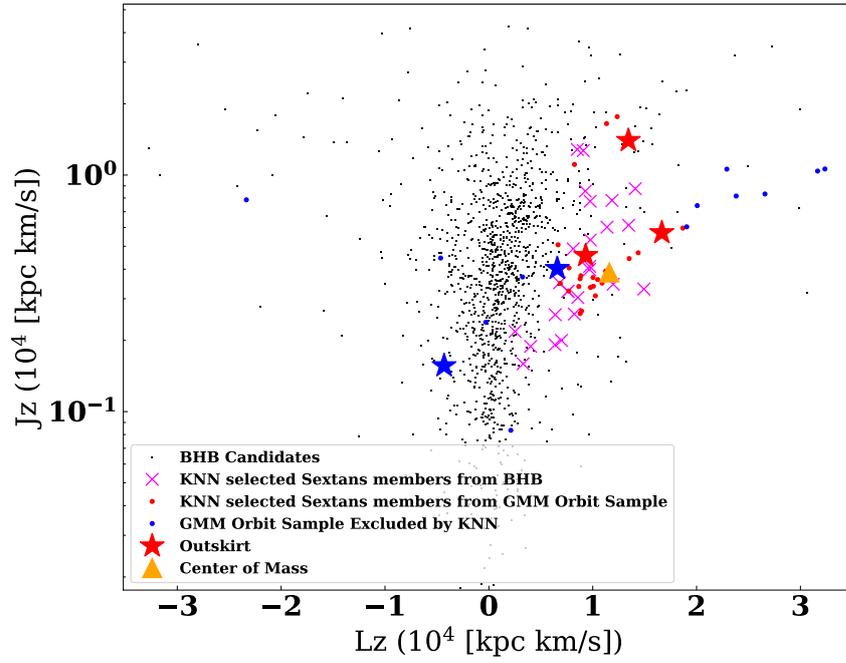


Figure 5.15: J_z - L_z plot for the performance of the KNN. The GMM high probability members selected by KNN are labelled in red while the excluded ones are labelled in blue. The BHBs selected by KNN are labelled in purple cross. We can see a narrower distribution in J_z for the selected stars from 1000 to 10000.

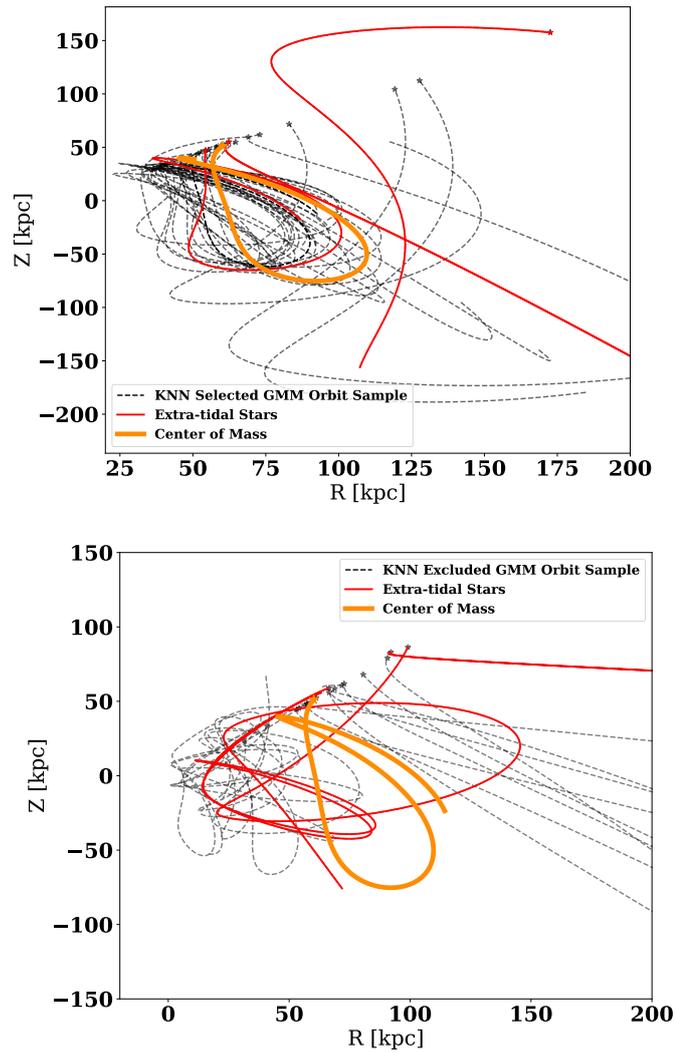


Figure 5.16: Orbits for the GMM orbit sample. Top panel: The GMM orbit sample members selected by the KNN. The star denotes the current position of the stars. The orange curve denotes the center of mass orbit of Sextans. The orbits of all the stars selected by KNN are labelled by the dark dash line. Most of the orbits are similar to the orbit of the center of mass. Bottom panel: The members excluded by KNN. The KNN is capable of excluding all the stars with different orbits than the Sextans center of mass.

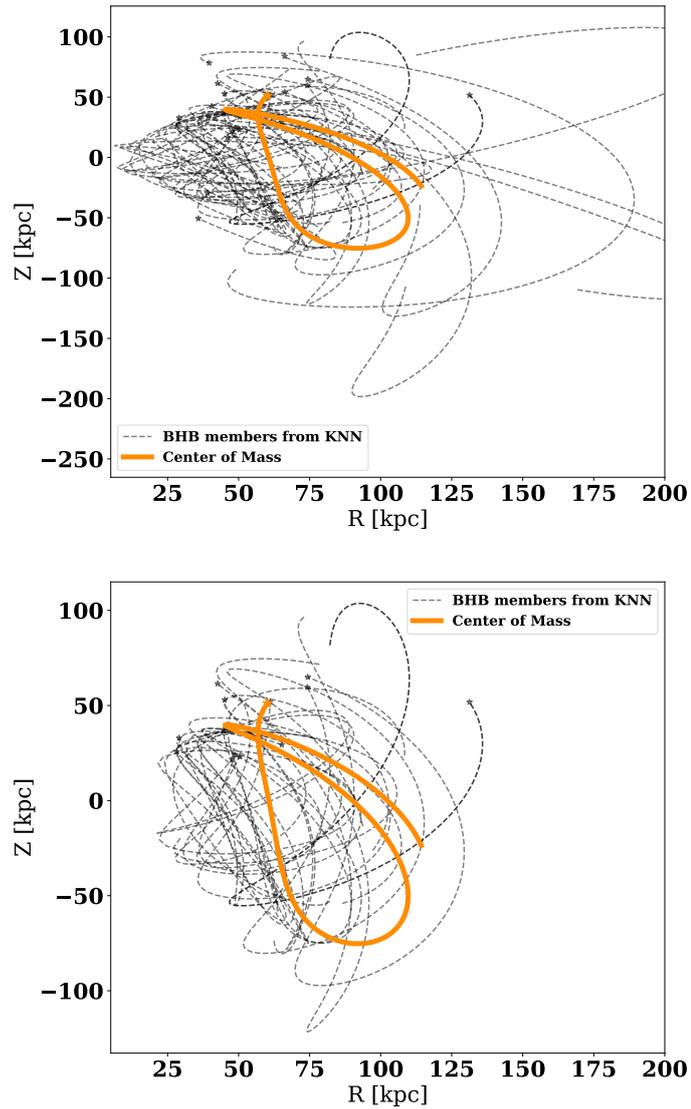


Figure 5.17: Orbits for the selected BHB members by the KNN. Top panel: All the BHB stars selected by KNN. The orange curve denotes the center of mass orbit of Sextans. The orbits of all the BHB selected by KNN are labelled by the dark dash line. Bottom panel: BHB stars selected by KNN with orbits with comparable orbits as the Sextans center of mass (blue box selection in 5.18). The star denotes the current positions of the stars.

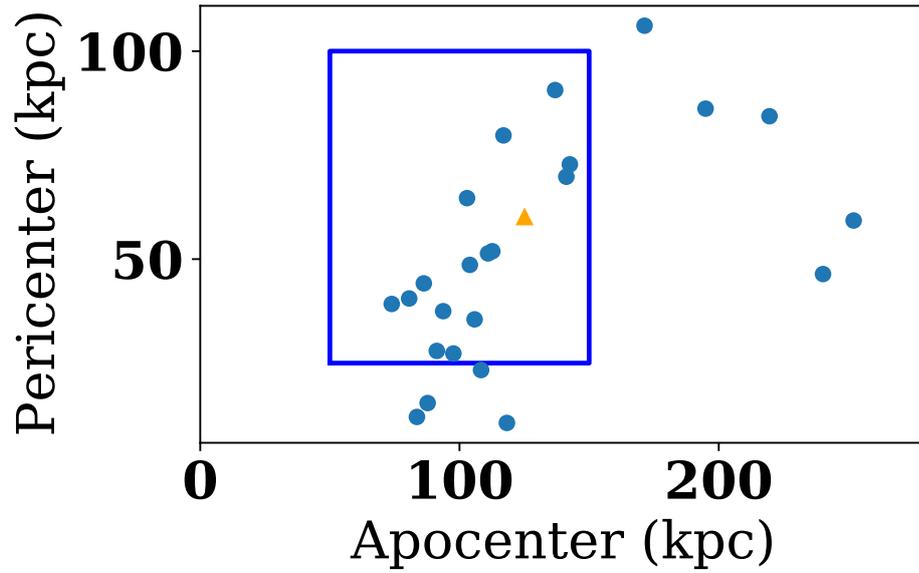


Figure 5.18: Pericenter versus apocenter for the BHB stars selected by the KNN. The orange triangle denotes the Sextans center of mass. The blue points are the apocenter and pericenter calculated from the orbits of the BHB stars. The blue box is our selection of stars have comparable pericenter and apocenter to the center of mass.

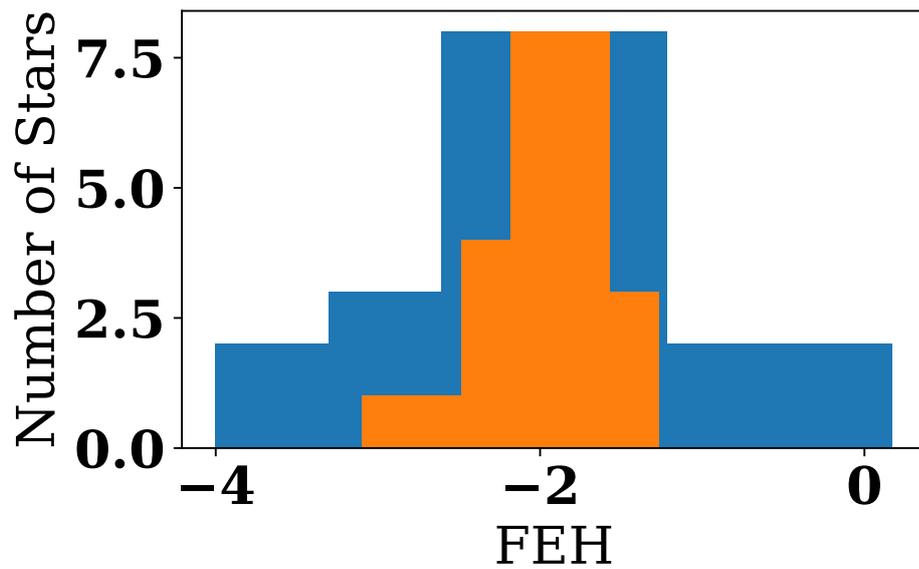


Figure 5.19: [Fe/H] distributions for the selected Sextans members from the BHB sample (labelled in blue) and the GMM orbit sample (labelled in orange). The two samples share the same mean value in [Fe/H].

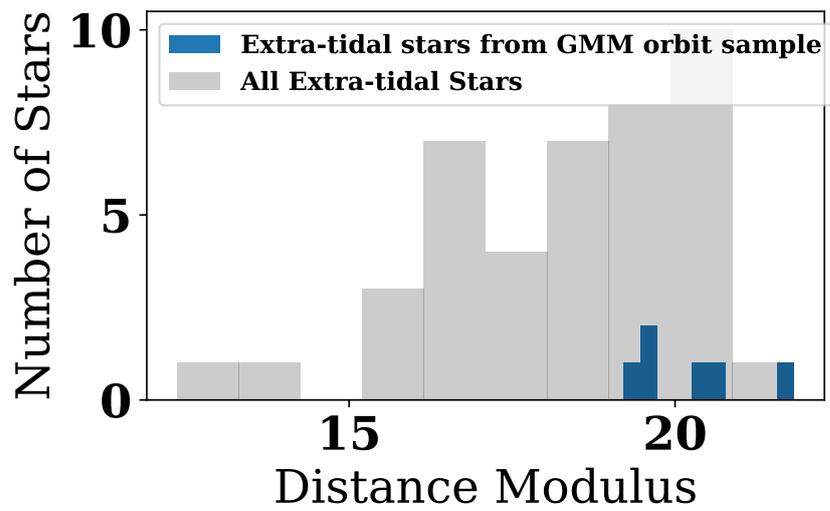


Figure 5.20: Distance modulus distribution for the extra-tidal stars in the original GMM sample (grey) and pre-selected GMM orbit sample (blue). The median distance modulus value of Sextans center of mass is around 19.404.

Chapter 6

Conclusion

In my dissertation, I utilized the datasets from large sky surveys, SDSS and DESI, to unveil the structure of our universe from large-scale structure traced by IGM to small-scale structures traced by nearby dwarf galaxies. The main results of this dissertation are:

1. I present new measurements of the mean transmitted flux in the hydrogen Ly α and a relative transmitted flux measurement in Ly β using 27,008 quasar spectra from the Fourteenth Data Release (DR14) of the Extended Baryon Oscillation Spectroscopic Survey (eBOSS). The $\tau_{\text{Ly}\alpha}$ values in this study show a smooth increase by a factor of 5 over the redshift range $z = 2.4 - 4.4$.

2. I discuss my contribution to the development of a CNN-based DLA finder used in the DESI year one cosmology analysis.

3. I investigate the spatial distribution, kinematics and metallicity of the Draco dwarf spheroidal galaxy. Using a sample of stars selected from the DESI early

data release, I employ a Gaussian mixture model to identify high probability members in Draco using Gaia DR3 proper motions and line of sight velocity and metallicity information derived from DESI spectroscopy. I identify eight high probability members outside the tidal radius. Two of those extra-tidal members also identified in Walker et al. (2015) and Qi et al. (2022), providing further confirmation for previous extra-tidal candidates as well as new candidates for the study of the extended stellar distributions of Draco. I also conclude that there is minimal evidence for the tidal disturbance of Draco, in agreement with previous work and analytic predictions.

4. I study the spatial distribution, kinematics and metallicity of the Sextans dwarf spheroidal galaxy, and focus on searching for extra-tidal candidates using 6D orbit integration and KNN machine learning. I identify 35 potential Sextans members that have comparable orbits to the Sextans center of mass. From the orbits and metallicity of the potential Sextans member BHB stars, we conclude that Sextans has undergone some dynamical heating, possibly in a merger before its infall to the Milky Way. This technique is also promising for the analysis of other dwarf galaxies observed in the DESI and DESI-II surveys.

Bibliography

Abbott, R., Abbott, T. D., Abraham, S., et al. 2021, *ApJL*, 915, L5, doi: 10.3847/2041-8213/ac082e

Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, *ApJS*, 235, 42, doi: 10.3847/1538-4365/aa9e8a

Anguiano, B., Majewski, S. R., Hayes, C. R., et al. 2020, *AJ*, 160, 43, doi: 10.3847/1538-3881/ab9813

Bajtlik, S., Duncan, R. C., & Ostriker, J. P. 1988, *ApJ*, 327, 570, doi: 10.1086/166217

Battaglia, G., Helmi, A., & Breddels, M. 2013, *NewAR*, 57, 52, doi: 10.1016/j.newar.2013.05.003

Battaglia, G., Taibi, S., Thomas, G. F., & Fritz, T. K. 2022, *A&A*, 657, A54, doi: 10.1051/0004-6361/202141528

Battaglia, G., Tolstoy, E., Helmi, A., et al. 2011, *MNRAS*, 411, 1013, doi: 10.1111/j.1365-2966.2010.17745.x

- Bautista, J. E., Busca, N. G., Guy, J., et al. 2017, *A&A*, 603, A12, doi: 10.1051/0004-6361/201730533
- Bechtol, K., Drlica-Wagner, A., Balbinot, E., et al. 2015a, *ApJ*, 807, 50, doi: 10.1088/0004-637X/807/1/50
- . 2015b, *ApJ*, 807, 50, doi: 10.1088/0004-637X/807/1/50
- Becker, G. D., & Bolton, J. S. 2013, *MNRAS*, 436, 1023, doi: 10.1093/mnras/stt1610
- Becker, G. D., Bolton, J. S., Haehnelt, M. G., & Sargent, W. L. W. 2011, *MNRAS*, 410, 1096, doi: 10.1111/j.1365-2966.2010.17507.x
- Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. 2013, *MNRAS*, 430, 2067, doi: 10.1093/mnras/stt031
- Becker, R. H., Fan, X., White, R. L., et al. 2001, *AJ*, 122, 2850, doi: 10.1086/324231
- Belokurov, V., & Koposov, S. E. 2016, *MNRAS*, 456, 602, doi: 10.1093/mnras/stv2688
- Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2006a, *ApJL*, 647, L111, doi: 10.1086/507324
- . 2006b, *ApJL*, 642, L137, doi: 10.1086/504797
- . 2007, *ApJ*, 654, 897, doi: 10.1086/509718
- Belokurov, V., Walker, M. G., Evans, N. W., et al. 2008, *ApJL*, 686, L83, doi: 10.1086/592962

- Bernardi, M., Sheth, R. K., Annis, J., et al. 2003, *AJ*, 125, 1866, doi: 10.1086/367794
- Bolton, J. S., & Becker, G. D. 2009, *MNRAS*, 398, L26, doi: 10.1111/j.1745-3933.2009.00700.x
- Bolton, J. S., Haehnelt, M. G., Viel, M., & Springel, V. 2005, *MNRAS*, 357, 1178, doi: 10.1111/j.1365-2966.2005.08704.x
- Bovill, M. S., & Ricotti, M. 2009, *ApJ*, 693, 1859, doi: 10.1088/0004-637X/693/2/1859
- Bovy, J. 2015, *ApJS*, 216, 29, doi: 10.1088/0067-0049/216/2/29
- Brown, T. M., Tumlinson, J., Geha, M., et al. 2014, *ApJ*, 796, 91, doi: 10.1088/0004-637X/796/2/91
- Bullock, J. S., & Johnston, K. V. 2005, *ApJ*, 635, 931, doi: 10.1086/497422
- Bullock, J. S., Kravtsov, A. V., & Weinberg, D. H. 2000, *ApJ*, 539, 517, doi: 10.1086/309279
- Cai, Z., Fan, X., Bian, F., et al. 2017, *ApJ*, 839, 131, doi: 10.3847/1538-4357/aa6a1a
- Carretta, E., Bragaglia, A., Gratton, R., D'Orazi, V., & Lucatello, S. 2009, *A&A*, 508, 695, doi: 10.1051/0004-6361/200913003
- Casetti-Dinescu, D. I., & Girard, T. M. 2016, *MNRAS*, 461, 271, doi: 10.1093/mnras/stw1337
- Chabanier, S., Palanque-Delabrouille, N., Yèche, C., et al. 2018, arXiv e-prints, arXiv:1812.03554. <https://arxiv.org/abs/1812.03554>

- Cooper, A. P., Kuposov, S. E., Allende Prieto, C., et al. 2023, *ApJ*, 947, 37, doi: 10.3847/1538-4357/acb3c0
- Dall’Aglio, A., Wisotzki, L., & Worseck, G. 2008, *A&A*, 491, 465, doi: 10.1051/0004-6361:200810724
- . 2009, arXiv e-prints. <https://arxiv.org/abs/0906.1484>
- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, *MNRAS*, 460, 1270, doi: 10.1093/mnras/stw641
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, 151, 44, doi: 10.3847/0004-6256/151/2/44
- de Beurs, Z. L., Islam, N., Gopalan, G., & Vrtilik, S. D. 2022, *ApJ*, 933, 116, doi: 10.3847/1538-4357/ac6184
- Deason, A., Wetzel, A., & Garrison-Kimmel, S. 2014a, *ApJ*, 794, 115, doi: 10.1088/0004-637X/794/2/115
- . 2014b, *ApJ*, 794, 115, doi: 10.1088/0004-637X/794/2/115
- Deason, A. J., Bose, S., Fattahi, A., et al. 2022, *MNRAS*, 511, 4044, doi: 10.1093/mnras/stab3524
- Deason, A. J., Belokurov, V., Hamren, K. M., et al. 2014c, *MNRAS*, 444, 3975, doi: 10.1093/mnras/stu1764

- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016a, arXiv e-prints, arXiv:1611.00036, doi: 10.48550/arXiv.1611.00036
- . 2016b, arXiv e-prints, arXiv:1611.00036, doi: 10.48550/arXiv.1611.00036
- . 2016c, arXiv e-prints, arXiv:1611.00037, doi: 10.48550/arXiv.1611.00037
- DESI Collaboration, Abareshi, B., Aguilar, J., et al. 2022a, AJ, 164, 207, doi: 10.3847/1538-3881/ac882b
- . 2022b, AJ, 164, 207, doi: 10.3847/1538-3881/ac882b
- DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2023a, arXiv e-prints, arXiv:2306.06308, doi: 10.48550/arXiv.2306.06308
- . 2023b, arXiv e-prints, arXiv:2306.06307, doi: 10.48550/arXiv.2306.06307
- . 2024, arXiv e-prints, arXiv:2404.03001, doi: 10.48550/arXiv.2404.03001
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168, doi: 10.3847/1538-3881/ab089d
- Dijkstra, M., Lidz, A., & Hui, L. 2004, ApJ, 605, 7, doi: 10.1086/382199
- D’Orazi, V., Oliva, E., Bragaglia, A., et al. 2020, A&A, 633, A38, doi: 10.1051/0004-6361/201936651
- Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, ApJS, 178, 89, doi: 10.1086/589654

- Drlica-Wagner, A., Bechtol, K., Rykoff, E. S., et al. 2015, *ApJ*, 813, 109, doi: 10.1088/0004-637X/813/2/109
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, *ApJ*, 633, 560, doi: 10.1086/466512
- El-Badry, K., Bradford, J., Quataert, E., et al. 2018, *MNRAS*, 477, 1536, doi: 10.1093/mnras/sty730
- Erkal, D., Koposov, S. E., & Belokurov, V. 2017, *MNRAS*, 470, 60, doi: 10.1093/mnras/stx1208
- Eyre, A., & Binney, J. 2009, *MNRAS*, 400, 548, doi: 10.1111/j.1365-2966.2009.15494.x
- Fattahi, A., Navarro, J. F., Frenk, C. S., et al. 2018, *MNRAS*, 476, 3816, doi: 10.1093/mnras/sty408
- Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008a, *ApJ*, 688, 85, doi: 10.1086/592289
- Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008b, *ApJ*, 681, 831, doi: 10.1086/588648
- Fix, E., & Hodges, J. L. 1951
- Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *AJ*, 150, 150, doi: 10.1088/0004-6256/150/5/150

- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306, doi: 10.1086/670067
- Frieman, J. A., Turner, M. S., & Huterer, D. 2008, *ARA&A*, 46, 385, doi: 10.1146/annurev.astro.46.060407.145243
- Fumagalli, M., O’Meara, J. M., & Prochaska, J. X. 2016, *MNRAS*, 455, 4100, doi: 10.1093/mnras/stv2616
- Fumagalli, M., O’Meara, J. M., Prochaska, J. X., Rafelski, M., & Kanekar, N. 2015, *MNRAS*, 446, 3178, doi: 10.1093/mnras/stu2325
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, 595, A1, doi: 10.1051/0004-6361/201629272
- Gaia Collaboration, Helmi, A., van Leeuwen, F., et al. 2018a, *A&A*, 616, A12, doi: 10.1051/0004-6361/201832698
- . 2018b, *A&A*, 616, A12, doi: 10.1051/0004-6361/201832698
- Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018c, *A&A*, 616, A10, doi: 10.1051/0004-6361/201832843
- Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, *A&A*, 674, A1, doi: 10.1051/0004-6361/202243940
- Gao, Y., Gu, Q., Liu, G., et al. 2023, *A&A*, 677, A179, doi: 10.1051/0004-6361/202346753

- Garnett, R., Ho, S., Bird, S., & Schneider, J. 2017, MNRAS, 472, 1850, doi: 10.1093/mnras/stx1958
- Garzilli, A., Magalich, A., Theuns, T., et al. 2019, MNRAS, 489, 3456, doi: 10.1093/mnras/stz2188
- Geha, M., Willman, B., Simon, J. D., et al. 2009, ApJ, 692, 1464, doi: 10.1088/0004-637X/692/2/1464
- Gnedin, N. Y., & Madau, P. 2022, Living Reviews in Computational Astrophysics, 8, 3, doi: 10.1007/s41115-022-00015-5
- Gnedin, O. Y., Hernquist, L., & Ostriker, J. P. 1999, ApJ, 514, 109, doi: 10.1086/306910
- Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65, doi: 10.2140/camcos.2010.5.65
- Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633, doi: 10.1086/148444
- Ho, M.-F., Bird, S., & Garnett, R. 2020, MNRAS, 496, 5436, doi: 10.1093/mnras/staa1806
- Hui, L., & Gnedin, N. Y. 1997, MNRAS, 292, 27, doi: 10.1093/mnras/292.1.27
- Ibata, R., Bellazzini, M., Thomas, G., et al. 2020, ApJL, 891, L19, doi: 10.3847/2041-8213/ab77c7
- Ibata, R. A., Gilmore, G., & Irwin, M. J. 1994, Nature, 370, 194, doi: 10.1038/370194a0

- Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, MNRAS, 442, 1805, doi: 10.1093/mnras/stu936
- Iršič, V., & Viel, M. 2014, JCAP, 12, 024, doi: 10.1088/1475-7516/2014/12/024
- Iršič, V., Viel, M., Berg, T. A. M., et al. 2017, MNRAS, 466, 4332, doi: 10.1093/mnras/stw3372
- Iršič, V., Slosar, A., Bailey, S., et al. 2013, JCAP, 2013, 016, doi: 10.1088/1475-7516/2013/09/016
- Irwin, M. J., Bunclark, P. S., Bridgeland, M. T., & McMahon, R. G. 1990, MNRAS, 244, 16P
- Ivezić, Ž., Sesar, B., Jurić, M., et al. 2008, ApJ, 684, 287, doi: 10.1086/589678
- Jardel, J. R., Gebhardt, K., Fabricius, M. H., Drory, N., & Williams, M. J. 2013, ApJ, 763, 91, doi: 10.1088/0004-637X/763/2/91
- Jena, T., Norman, M. L., Tytler, D., et al. 2005, MNRAS, 361, 70, doi: 10.1111/j.1365-2966.2005.09095.x
- Jensen, J., Hayes, C. R., Sestito, F., et al. 2024, MNRAS, 527, 4209, doi: 10.1093/mnras/stad3322
- Johnston, K. V. 1998, ApJ, 495, 297, doi: 10.1086/305273
- Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864, doi: 10.1086/523619

- Kamble, V., Dawson, K., du Mas des Bourboux, H., Bautista, J., & Scheinder, D. P. 2020, *ApJ*, 892, 70, doi: 10.3847/1538-4357/ab76bd
- Kaplinghat, M., & Strigari, L. E. 2008, *ApJL*, 682, L93, doi: 10.1086/591052
- Kazantzidis, S., Lokas, E. L., Callegari, S., Mayer, L., & Moustakas, L. A. 2011, *ApJ*, 726, 98, doi: 10.1088/0004-637X/726/2/98
- Kepner, J., Fan, X., Bahcall, N., et al. 1999, *ApJ*, 517, 78, doi: 10.1086/307160
- King, I. 1962, *AJ*, 67, 471, doi: 10.1086/108756
- Kinman, T. D., Suntzeff, N. B., & Kraft, R. P. 1994, *AJ*, 108, 1722, doi: 10.1086/117191
- Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, *ApJ*, 779, 102, doi: 10.1088/0004-637X/779/2/102
- Kirby, E. N., Lanfranchi, G. A., Simon, J. D., Cohen, J. G., & Guhathakurta, P. 2011, *ApJ*, 727, 78, doi: 10.1088/0004-637X/727/2/78
- Kirkman, D., Tytler, D., Suzuki, N., et al. 2005, *MNRAS*, 360, 1373, doi: 10.1111/j.1365-2966.2005.09126.x
- Klessen, R. S., Grebel, E. K., & Harbeck, D. 2003, *ApJ*, 589, 798, doi: 10.1086/374816
- Klimentowski, J., Lokas, E. L., Kazantzidis, S., et al. 2007, *MNRAS*, 378, 353, doi: 10.1111/j.1365-2966.2007.11799.x
- Koposov, S., Belokurov, V., Evans, N. W., et al. 2008, *ApJ*, 686, 279, doi: 10.1086/589911

- Koposov, S. E., Belokurov, V., Torrealba, G., & Evans, N. W. 2015a, *ApJ*, 805, 130,
doi: 10.1088/0004-637X/805/2/130
- Koposov, S. E., Belokurov, V., Evans, N. W., et al. 2012, *ApJ*, 750, 80, doi: 10.1088/
0004-637X/750/1/80
- Koposov, S. E., Casey, A. R., Belokurov, V., et al. 2015b, *ApJ*, 811, 62, doi: 10.1088/
0004-637X/811/1/62
- Koposov, S. E., Allende-Prieto, C., Cooper, A. P., et al. 2024, arXiv e-prints,
arXiv:2407.06280, doi: 10.48550/arXiv.2407.06280
- Lan, T.-W., Ménard, B., Baron, D., et al. 2018, *MNRAS*, 477, 3520, doi: 10.1093/
mnras/sty864
- Law, D. R., Johnston, K. V., & Majewski, S. R. 2005, *ApJ*, 619, 807, doi: 10.1086/
426779
- Law, D. R., & Majewski, S. R. 2010, *ApJ*, 714, 229, doi: 10.1088/0004-637X/714/1/
229
- Lee, M. G., Yuk, I.-S., Park, H. S., Harris, J., & Zaritsky, D. 2009, *ApJ*, 703, 692,
doi: 10.1088/0004-637X/703/1/692
- Li, T. S., Simon, J. D., Kuehn, K., et al. 2018, *ApJ*, 866, 22, doi: 10.3847/1538-4357/
aadf91

- Lidz, A., Faucher-Giguère, C.-A., Dall’Aglio, A., et al. 2010, *ApJ*, 718, 199, doi: 10.1088/0004-637X/718/1/199
- Lidz, A., McQuinn, M., Zaldarriaga, M., Hernquist, L., & Dutta, S. 2007, *ApJ*, 670, 39, doi: 10.1086/521974
- Lokas, E. L. 2009, *MNRAS*, 394, L102, doi: 10.1111/j.1745-3933.2009.00620.x
- Lokas, E. L., Majewski, S. R., Kazantzidis, S., et al. 2012, *ApJ*, 751, 61, doi: 10.1088/0004-637X/751/1/61
- Lusso, E., Worseck, G., Hennawi, J. F., et al. 2015, *MNRAS*, 449, 4204, doi: 10.1093/mnras/stv516
- Madau, P. 1995, *ApJ*, 441, 18, doi: 10.1086/175332
- Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003, *ApJ*, 599, 1082, doi: 10.1086/379504
- Margala, D., Kirkby, D., Dawson, K., et al. 2016, *ApJ*, 831, 157, doi: 10.3847/0004-637X/831/2/157
- Martínez-García, A. M., del Pino, A., Aparicio, A., van der Marel, R. P., & Watkins, L. L. 2021, *MNRAS*, 505, 5884, doi: 10.1093/mnras/stab1568
- Massari, D., Helmi, A., Mucciarelli, A., et al. 2019, arXiv e-prints, arXiv:1904.04037, doi: 10.48550/arXiv.1904.04037
- Mateo, M. L. 1998, *ARA&A*, 36, 435, doi: 10.1146/annurev.astro.36.1.435

- McConnachie, A. W. 2012, *AJ*, 144, 4, doi: 10.1088/0004-6256/144/1/4
- McConnachie, A. W., & Venn, K. A. 2020a, *AJ*, 160, 124, doi: 10.3847/1538-3881/aba4ab
- . 2020b, *Research Notes of the American Astronomical Society*, 4, 229, doi: 10.3847/2515-5172/abd18b
- . 2020c, *Research Notes of the American Astronomical Society*, 4, 229, doi: 10.3847/2515-5172/abd18b
- McDonald, P., Seljak, U., Cen, R., et al. 2005, *ApJ*, 635, 761, doi: 10.1086/497563
- McMillan, P. J. 2017, *MNRAS*, 465, 76, doi: 10.1093/mnras/stw2759
- McMonigal, B., Bate, N. F., Lewis, G. F., et al. 2014, *MNRAS*, 444, 3139, doi: 10.1093/mnras/stu1659
- McQuinn, M. 2016, *ARA&A*, 54, 313, doi: 10.1146/annurev-astro-082214-122355
- McQuinn, M., & Upton Sanderbeck, P. R. 2016, *MNRAS*, 456, 47, doi: 10.1093/mnras/stv2675
- Meiksin, A. A. 2009, *Reviews of Modern Physics*, 81, 1405, doi: 10.1103/RevModPhys.81.1405
- Muñoz, R. R., Côté, P., Santana, F. A., et al. 2018, *ApJ*, 860, 66, doi: 10.3847/1538-4357/aac16b

- Muñoz, R. R., Frinchaboy, P. M., Majewski, S. R., et al. 2005, *ApJL*, 631, L137, doi: 10.1086/497396
- Mukundan, K., Nair, P., Bailin, J., & Li, W. 2024, *MNRAS*, doi: 10.1093/mnras/stae1684
- Nasir, F., Bolton, J. S., & Becker, G. D. 2016, *MNRAS*, 463, 2335, doi: 10.1093/mnras/stw2147
- Newberg, H. J., Yanny, B., Rockosi, C., et al. 2002, *ApJ*, 569, 245, doi: 10.1086/338983
- Noterdaeme, P., Balashev, S., Krogager, J. K., et al. 2019, *A&A*, 627, A32, doi: 10.1051/0004-6361/201935371
- Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012, *A&A*, 547, L1, doi: 10.1051/0004-6361/201220259
- O'Meara, J. M., Lehner, N., Howk, J. C., et al. 2015, *AJ*, 150, 111, doi: 10.1088/0004-6256/150/4/111
- Pace, A. B., Erkal, D., & Li, T. S. 2022, *ApJ*, 940, 136, doi: 10.3847/1538-4357/ac997b
- Pace, A. B., & Li, T. S. 2019, *ApJ*, 875, 77, doi: 10.3847/1538-4357/ab0aee
- Palanque-Delabrouille, N., Yèche, C., Borde, A., et al. 2013, *A&A*, 559, A85, doi: 10.1051/0004-6361/201322130
- Palanque-Delabrouille, N., Yèche, C., Baur, J., et al. 2015, *Journal of Cosmology and Astro-Particle Physics*, 2015, 011, doi: 10.1088/1475-7516/2015/11/011

- Palma, C., Majewski, S. R., Siegel, M. H., et al. 2003, *AJ*, 125, 1352, doi: 10.1086/367594
- Pâris, I., Petitjean, P., Rollinde, E., et al. 2011, *A&A*, 530, A50, doi: 10.1051/0004-6361/201016233
- Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, *A&A*, 613, A51, doi: 10.1051/0004-6361/201732445
- Parks, D., Prochaska, J. X., Dong, S., & Cai, Z. 2018, *MNRAS*, 476, 1151, doi: 10.1093/mnras/sty196
- Patrick, J. M., Kopolov, S. E., & Walker, M. G. 2022, *MNRAS*, 514, 1757, doi: 10.1093/mnras/stac1478
- Pawlowski, M. S., McGaugh, S. S., & Jerjen, H. 2015, *MNRAS*, 453, 1047, doi: 10.1093/mnras/stv1588
- Peñarrubia, J., Navarro, J. F., & McConnachie, A. W. 2008a, *ApJ*, 673, 226, doi: 10.1086/523686
- . 2008b, *ApJ*, 673, 226, doi: 10.1086/523686
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Percival, W. J., Cole, S., Eisenstein, D. J., et al. 2007, *MNRAS*, 381, 1053, doi: 10.1111/j.1365-2966.2007.12268.x

- Pérez-Ràfols, I., Font-Ribera, A., Miralda-Escudé, J., et al. 2018, MNRAS, 473, 3019,
doi: 10.1093/mnras/stx2525
- Plummer, H. C. 1911, MNRAS, 71, 460, doi: 10.1093/mnras/71.5.460
- Preston, G. W., Shectman, S. A., & Beers, T. C. 1991, ApJ, 375, 121, doi: 10.1086/
170175
- Prochaska, J. X., & Herbert-Fort, S. 2004, PASP, 116, 622, doi: 10.1086/421985
- Prochaska, J. X., Madau, P., O’Meara, J. M., & Fumagalli, M. 2014, MNRAS, 438, 476,
doi: 10.1093/mnras/stt2218
- Prochaska, J. X., O’Meara, J. M., & Worseck, G. 2010, ApJ, 718, 392, doi: 10.1088/
0004-637X/718/1/392
- Prochaska, J. X., & Wolfe, A. M. 1997, ApJ, 487, 73, doi: 10.1086/304591
- . 2009, ApJ, 696, 1543, doi: 10.1088/0004-637X/696/2/1543
- Puchwein, E., Bolton, J. S., Haehnelt, M. G., et al. 2015, MNRAS, 450, 4081, doi: 10.
1093/mnras/stv773
- Qi, Y., Zivick, P., Pace, A. B., Riley, A. H., & Strigari, L. E. 2022, MNRAS, 512, 5601,
doi: 10.1093/mnras/stac805
- Rauch, M. 1998, ARA&A, 36, 267, doi: 10.1146/annurev.astro.36.1.267
- Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, AJ, 131, 2766, doi: 10.1086/503559

- Roberts, G. O., Gelman, A., & Gilks, W. R. 1997, *The Annals of Applied Probability*, 7, 110
- Roderick, T. A., Jerjen, H., Da Costa, G. S., & Mackey, A. D. 2016, *MNRAS*, 460, 30, doi: 10.1093/mnras/stw949
- Sakamoto, T., & Hasegawa, T. 2006, *ApJL*, 653, L29, doi: 10.1086/510332
- Savino, A., & Posti, L. 2019, *A&A*, 624, L9, doi: 10.1051/0004-6361/201935417
- Schaye, J., Aguirre, A., Kim, T.-S., et al. 2003, *ApJ*, 596, 768, doi: 10.1086/378044
- Schlaufman, K. C., Rockosi, C. M., Allende Prieto, C., et al. 2009, *ApJ*, 703, 2177, doi: 10.1088/0004-637X/703/2/2177
- Ségall, M., Ibata, R. A., Irwin, M. J., Martin, N. F., & Chapman, S. 2007, *MNRAS*, 375, 831, doi: 10.1111/j.1365-2966.2006.11356.x
- Sestito, F., Roediger, J., Navarro, J. F., et al. 2023, *MNRAS*, 523, 123, doi: 10.1093/mnras/stad1417
- Shapley, H. 1938, *Nature*, 142, 715, doi: 10.1038/142715b0
- Sharma, S., Bland-Hawthorn, J., Johnston, K. V., & Binney, J. 2011, *ApJ*, 730, 3, doi: 10.1088/0004-637X/730/1/3
- Shipp, N., Drlica-Wagner, A., Balbinot, E., et al. 2018, *ApJ*, 862, 114, doi: 10.3847/1538-4357/aacdab
- Simon, J. D. 2019, *ARA&A*, 57, 375, doi: 10.1146/annurev-astro-091918-104453

- Simon, J. D., Geha, M., Minor, Q. E., et al. 2011, *ApJ*, 733, 46, doi: 10.1088/0004-637X/733/1/46
- Songaila, A. 2004, *AJ*, 127, 2598, doi: 10.1086/383561
- Suzuki, N. 2006, *ApJS*, 163, 110, doi: 10.1086/499272
- Tegmark, M., Strauss, M. A., Blanton, M. R., et al. 2004, *PhRvD*, 69, 103501, doi: 10.1103/PhysRevD.69.103501
- Theuns, T., Schaye, J., Zaroubi, S., et al. 2002, *ApJL*, 567, L103, doi: 10.1086/339998
- Thomas, G. F., & Battaglia, G. 2023, *A&A*, 675, C1, doi: 10.1051/0004-6361/202142347e
- Torrealba, G., Belokurov, V., Koposov, S. E., et al. 2018, *MNRAS*, 475, 5085, doi: 10.1093/mnras/sty170
- Turner, W., Martini, P., Göksel Karaçaylı, N., et al. 2024, arXiv e-prints, arXiv:2405.06743, doi: 10.48550/arXiv.2405.06743
- Vasiliev, E., Belokurov, V., & Erkal, D. 2021, *MNRAS*, 501, 2279, doi: 10.1093/mnras/staa3673
- Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, *PhRvD*, 88, 043502, doi: 10.1103/PhysRevD.88.043502
- Villasenor, B., Robertson, B., Madau, P., & Schneider, E. 2021, arXiv e-prints, arXiv:2111.00019. <https://arxiv.org/abs/2111.00019>

- . 2022, arXiv e-prints, arXiv:2209.14220, doi: 10.48550/arXiv.2209.14220
- Walker, M. G., Caldwell, N., Mateo, M., et al. 2023, *ApJS*, 268, 19, doi: 10.3847/1538-4365/acdd79
- Walker, M. G., Mateo, M., & Olszewski, E. W. 2008, *ApJL*, 688, L75, doi: 10.1086/595586
- Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2007, *ApJL*, 667, L53, doi: 10.1086/521998
- Walker, M. G., Olszewski, E. W., & Mateo, M. 2015, *MNRAS*, 448, 2717, doi: 10.1093/mnras/stv099
- Wang, B., Zou, J., Cai, Z., et al. 2022, *ApJS*, 259, 28, doi: 10.3847/1538-4365/ac4504
- Wang, M. Y., Fattahi, A., Cooper, A. P., et al. 2017, *MNRAS*, 468, 4887, doi: 10.1093/mnras/stx742
- Weinberg, D. H., Bullock, J. S., Governato, F., Kuzio de Naray, R., & Peter, A. H. G. 2015, *Proceedings of the National Academy of Science*, 112, 12249, doi: 10.1073/pnas.1308716112
- Willman, B., Dalcanton, J. J., Martinez-Delgado, D., et al. 2005, *ApJL*, 626, L85, doi: 10.1086/431760
- Wilson, A. G. 1955, *PASP*, 67, 27, doi: 10.1086/126754

- Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, *ARA&A*, 43, 861, doi: 10.1146/annurev.astro.42.053102.133950
- Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, *ApJS*, 61, 249, doi: 10.1086/191114
- Worseck, G., Prochaska, J. X., O’Meara, J. M., et al. 2014, *MNRAS*, 445, 1745, doi: 10.1093/mnras/stu1827
- Xue, X.-X., Rix, H.-W., Yanny, B., et al. 2011, *ApJ*, 738, 79, doi: 10.1088/0004-637X/738/1/79
- Yang, Y., Hammer, F., Jiao, Y., & Pawlowski, M. S. 2022, *MNRAS*, 512, 4171, doi: 10.1093/mnras/stac644
- Yèche, C., Palanque-Delabrouille, N., Baur, J., & du Mas des Bourboux, H. 2017, *Journal of Cosmology and Astro-Particle Physics*, 2017, 047, doi: 10.1088/1475-7516/2017/06/047
- York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, *AJ*, 120, 1579, doi: 10.1086/301513
- Zafar, T., Péroux, C., Popping, A., et al. 2013, *A&A*, 556, A141, doi: 10.1051/0004-6361/201321154
- Zaldarriaga, M., Hui, L., & Tegmark, M. 2001, *ApJ*, 557, 519, doi: 10.1086/321652
- Zavala, J., & Frenk, C. S. 2019, *Galaxies*, 7, 81, doi: 10.3390/galaxies7040081

Zeraatgari, F. Z., Hafezianzadeh, F., Zhang, Y., et al. 2024, MNRAS, 527, 4677, doi: 10.

1093/mnras/stad3436