A Study of Galactic and Extragalactic Evolution with the MUSE Integral-Field Spectrograph

Zixian Wang
(Purmortal / 王梓先)

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Supervisor:
Prof. Jonathan (Joss) Bland-Hawthorn

Co-supervisors:
Dr. Michael Hayden, Dr. Sanjib Sharma

The University of Sydney
Faculty of Science
School of Physics

2023

Declaration

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Zixian Wang
(Purmortal / 王梓先)

September 12, 2023

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Zixian Wang
(Purmortal / 王梓先)

September 12, 2023

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Prof. Jonathan (Joss) Bland-Hawthorn

September 12, 2023

Always remember the night when I saw the sky full of stars in my grandma's yard.

多年以后，仍然会记起在乡下姥姥家看到漫天繁星的那个遥远的夜晚。

Abstract

The Milky Way is by far the best-studied galaxy in the Universe, and it has long been regarded as a benchmark and ideal laboratory for understanding galaxy evolution. Recent studies compared the Milky Way with its analogues from IFS surveys and revealed common structures such as the B/P bulge, thin and thick disks, etc. This presents a unique opportunity to link Galactic and extragalactic studies to understand galaxy evolution. However, high stellar density regions like the Galactic bulge and the core of globular clusters have not been fully explored but they are essential in the Galactic evolution. This is due to the fiber collision issues of fiber-fed multi-object spectrographs and extreme extinction. Furthermore, direct comparisons of Galactic and extragalactic observations are marred by differences in using individual stars vs. stellar populations and methods of measuring galaxy properties. It is difficult to identify whether the differences between the Milky Way and other galaxies are due to systematic offsets or different evolution processes. This thesis addresses these difficulties using the MUSE integral-field spectrograph. We develop a novel method to measure stellar chemical abundances of individual stars from MUSE. This method is then used to study the age distribution and chemical evolution of MSTO-SGB stars in low-extinction regions of the Galactic center. Our results agree with previous studies but are obtained more efficiently. We further present a tool to generate mock MUSE observations of the Milky Way using GCE models and spectral templates for unbiased comparisons with external galaxies. We use mock data to investigate the reliability of a broadly used tool pPXF in measuring external galaxy properties and explore the impact of dust on these results. This research facilitates a deeper understanding of critical Galactic structures such as the bulge and bridges the gap between Galactic and extragalactic studies for understanding galaxy evolution.

Acknowledgements

The journey of my PhD started with the exchange program in my third-year undergraduate. Thanks to Nanjing University for giving me the opportunity to study abroad, and also I am grateful to those who helped me during my undergraduate, especially my supervisor Jiwei Xie, who improved my confidence in working on astronomy, and always supported and encouraged me. Also, I would like to thank the Chinese Scholarship Council to support the living stipends during my PhD.

Since my research projects are somewhere in between Galactic and extragalactic studies, a four-year PhD is not enough for me to learn everything. Studying is endless. I also deeply feel that many of my works are indeed by “standing on the shoulders of giants”. Therefore, I am grateful to many researchers who have made important contributions to related fields.

I feel very fortunate to have had three supervisors during my PhD. Joss can always point out important research directions for the future where my work could be used, and gave me invaluable suggestions on how to extend my research. Michael acted like my daily supervisor who helped me deal with many things in life and study. He guided me into Galactic archaeology and taught me a lot in various aspects of scientific research, such as how to better present my own opinions, how to cooperate with others, and how to use the opportunity of each conference to talk with others. Sanjib is very talented in various mathematical, statistical, and physical methods. Every time I encountered a problem, he always provides extremely clear clues to help me think of and solve it. He taught me how to highlight my work and discard those aspects that are not important, to make the paper clear and concise.

In addition, I would like to thank a few collaborators. Yuan-Sen Ting and Maosheng Xiang encouraged me all the time, and even patiently taught me when I struggled with degrading the spectra. Jesse van de Sande is the one who helped me the most in SIfA except for my supervisors. He deserves my fourth supervisor and I learned a lot from him. Also thanks to Nic Scott for always answering my questions.

I also want to thank my friends in SIfA, who either helped me when I came to Australia, comforted me when I was upset, helped me adapt to the local culture, or encouraged me when during my graduation. They helped me overcome many difficulties during COVID-19, including Tanda, Zhen, Tania, Sam, Henry, Rebecca, Hyein, Madusha, Manisha, and Penny. And I wish Penny all the best and happy every day. Also thanks to my current office mates and CSC mates, we had a lot of fun together.

I am also very grateful to Marie. Every time I feel sad, I will talk to her and tell my recent moods. She also helped me a lot with scholarship and travel funding. I am also grateful to Scott for supporting me in attending many conferences, helping me with tutoring, and answering my endless questions. Also thanks to ASTRO-3D for providing various funding.

I would also like to thank Kai Yang for helping me out during those struggling periods and wish him all the best in the future. Thanks to Chengyi Wang, having you in the last few months before graduation was enjoyable, and your cooking is so great.

Thanks to YM, who is always with me, encourages me and makes me more mature during my life and study. We have gone through many difficulties together. I hope I can make you happy all the time.

Thanks to my childhood friend Hang Luo, those long phone calls immediately improved my psychological status. Thanks to my parents and kinsfolk for their continuous support during COVID-19.

Finally, I would like to thank myself for somehow always choosing the most difficult road every time, from studying astronomy to studying abroad. I did learn a lot and become stronger and more confident.

这次出国学习的经历始于大三参加的南京大学和悉尼大学的交换项目。我很感激我的母校能给我出去看看的机会，也很感谢本科期间一直帮助我的老师和同学，尤其是谢基伟老师，他提升了我对做天文研究的信心，也一直鼓励和帮助我。同时，我也感激国家基金委能够资助我完成四年的博士学习。

由于我的研究课题处在银河系和河外星系研究的中间，四年的时间远远不足以让我了解全部知识。学习永无止境。在学习的过程我我也深刻的感受到许多研究成果确实是“站在了巨人的肩膀上”，因此，我很感激在相关领域中做出重要贡献的许多学者。

我非常幸运拥有三位导师在博士期间指导我。Joss 每次都能一针见血地指出未来重要的研究方向，以及我的工作能派上用场的地方，并且为我在现有工作的基础上拓展其应用给出了很多宝贵的建议；Michael 作为我的日常导师，帮我处理了许多在生活和学习中大大小小的问题，他是带我进入银河考古学的引路人，并且在科学研究的各个方面，比如如何更好的陈述自己的观点，如何和他人合作，如何利用每次开会的机会和他人交流等教会我很多；Sanjib 在我心目中是各种数理统计方法的全才，每次我遇到问题他都会提供极其清晰的思路来帮助我思考，他教会了我如何抓中工作中最重要的，舍弃那些不重要的，从而让论文的观点更突出，也更容易让读者理解。

除此之外，我还想感谢几位合作者。丁源森和向茂盛在合作期间一直在鼓励我，甚至在我不懂如何处理光谱的时候耐心的教我。Jesse van de Sande 是除我的导师外在整个 SIfA 帮助我最多的人，可以说是我的第四导师。也非常感谢 Nic Scott 经常为我答疑解惑。

另外，我也很感谢在 SIfA 的好朋友，他们或在我刚到悉尼的时候帮助我，或在我失落难过的时候安慰我，或帮助我适应当地文化，或在我即将毕业的时候一直鼓励我，这让帮助我在疫情期间克服了很多困难，包括李坦达，宛振，Tania，Sam，Henry，Rebecca，Hyein，Madusha，Manisha 和 Penny。另祝 Penny 身体健康，开心每一天。也感谢办公室和公派的小伙伴们，和你们在一起玩耍非常快乐。

我还很感谢 Marie，每次难过的时候都会找她聊天，说出我最近的情绪。她也帮助我解决了很多奖学金和旅行经费的问题。我也很感谢 Scott 支持我出去开会，在助教方面帮助我，以及日常解答我的各种问题。感谢 ASTRO-3D 提供的各种经费。

我还要感谢杨凯帮助我走出了许多困难，希望他未来一切顺利。感谢汪成一，在毕业的前最后几个月合租生活非常快乐，做饭非常好吃。

感谢小明，在博士期间一直陪伴我，鼓励我，让我变得更成熟更懂事。我们一起走过了许多风风雨雨，希望我能让你一直幸福。

感谢我的发小罗航，我们的几次长电话对我的心理状况的改善立竿见影。感谢我的父母和亲人们在疫情期间一直提供的支持。

最后，感谢我自己每次都能稀里糊涂的走了最艰难的路，从选择学天文，到出国读书，不过我确实学到了很多，希望今后继续努力，变得越来越强大，越来越自信。

Contents

List of Figures

List of Tables

Summary of Works

This thesis includes one published work:

- Chapter 2 along with Appendix A have been published as Wang, Z., Hayden, M. R., Sharma, S., et al. 2022, MNRAS, 514, 1034

and one work to be submitted to the journal:

- Chapter 4 and Chapter 5 along with Appendix B and Appendix C are to be submitted to Monthly Notices of the Royal Astronomical Society as Wang et al. 2023

These chapters have been reproduced as published or to be submitted, except for minor typographical, reference updates, and formatting changes to keep a consistent style throughout this thesis.

Works completed as part of the thesis are also included:

- Chapter 3 will be drafted and submitted to Monthly Notices of the Royal Astronomical Society after this thesis submission
- Chapter 6 along with Appendix D will be submitted after more comprehensive explorations

Important contributions from others toward the works in this thesis are as follows. Joss Bland-Hawthorn: Project design, general supervision and guidance, advice on research direction and exploration. Michael Hayden: Project design, specific supervision and guidance, data reduction suggestions, advice on scientific interpretation and analysis. Sanjib Sharma: Project design, specific supervision and guidance, data reduction suggestions, advice on scientific interpretation and analysis. Yuan-Sen Ting & Maosheng Xiang: Assistance with data processing. Jesse van de Sande: Assistance with data processing and reduction, advice on scientific interpretation and analysis. Sam Vaughan: Assistance with data reduction. Specific contributions are listed in detail at the beginning of each chapter.

Chapter 1

Introduction

When looking at the winter night sky in the southern hemisphere, we see a long, bright and beautiful band. This is the Milky Way, where our sun has lived for over 5 billion years. The individual white dots we see are stars in our solar neighbourhood orbiting the Milky Way within a few kpc. Observing further with a telescope, we can find fainter galaxies such as Andromeda (M31) and M33, which are outside the Milky Way (hundreds of kpc away) and have similar structures to our Galaxy. With a more powerful and sensitive telescope, we can observe even much fainter galaxies (several to tens of Mpc away) with disk or elliptical structures.

Over centuries, many efforts have been made to understand the formation and evolution of galaxies. Among these, the studies of the Milky Way are the most comprehensive due to our unique vantage point. *Gaia* has provided accurate positions and velocities for more than a billion individual stars (Gaia Collaboration et al., 2018), and spectroscopic surveys have enabled the measurement of over 20 elemental abundances for millions of these stars (e.g., Zhao et al. 2012; De Silva et al. 2015; Majewski et al. 2017). Extragalactic observations have also unravelled many galaxy evolution processes. Despite the challenges of large distances and two-dimensional projections on the sky coordinates, we can still access the luminosity, averaged velocity, age, and metallicity to address general star formation history, merger events, and co-evolution with the surrounding environments (e.g., Sánchez et al. 2012; Croom et al. 2012; Bundy et al. 2015). Therefore, we have two categories of datasets for understanding galaxy evolution: one is our Milky Way, with the largest parameter dimensionality and the smallest uncertainty in a narrow parameter range; the other is external galaxies with fewer parameter dimensions and larger errors, but in a wide parameter range.

Nonetheless, combining these two datasets to construct general evolution theories that apply to all galaxies is a challenging and complex task. For decades, Galactic and extragalactic studies have proceeded along separate paths, each explored deeply but rarely compared to one another. This has resulted in several effects, such as the lack of comparison between the Galactic evolution models and extragalactic results, and the debates on the uniqueness of the Milky Way compared to external galaxies (e.g., Boardman et al. 2020b). Recent efforts to compare the Galactic properties with its analogues have revealed qualitative similarities on certain properties (e.g., Boardman et al. 2020a; Scott et al. 2021; Zhou et al. 2023). However, quantitative conclusions are still difficult to be drawn because of the differences in observation strategies, assumptions and measurement methods between the Galaxy and external systems. Furthermore, explorations are still needed for several aspects of the Milky Way, including the gas evolution, the formation of spiral arms and bar, the evolution history of stellar dense regions (e,g, Galactic center, globular clusters), and the interaction with dwarf and satellite galaxies, etc.

This thesis aims to employ the MUSE integral-field spectrograph to develop tools and algorithms that can facilitate the study of specific regions of the Milky Way (especially the Galactic center), and bridge the gap in comparing properties of the Milky Way with external systems. By linking the Galactic and extragalactic studies, we can construct theoretical models of galaxy formation and evolution that can fit both the Milky Way and external galaxies. In this chapter, we briefly introduce the history of integral-field spectrographs and the MUSE instrument, summarise the current Galactic and extragalactic studies, and provide our thesis goals.

1.1 Integral-field Spectrograph and MUSE

Ever since Isaac Newton first observed the solar spectrum using a simple prism, spectroscopy has become an important tool for astronomers. It allows us to study properties such as temperature, mass, chemical abundances, gas composition, and kinematics of celestial objects by analyzing light absorption or emission features across various wavelengths. Spectrographs are one of the most complex instruments and can be categorized into long-slit, multi-slits, fiber-fed, and integral-field spectrographs. With technological advancements, spectrograph observations have been transformed from single-source to multi-object simultaneous exposure. For spatially extended sources like galaxies, spectrographs have evolved from single-aperture spectroscopy (e.g., 2dF Galaxy Redshift Survey Colless et al. 2001 and 6dF Galaxy Survey Jones et al. 2004) to multiple spectra acquisition at different regions of the sources.

The integral-field spectrograph (IFS) is designed to obtain spatially resolved spectra in the optical/NIR at different locations of the object. The data format of IFS is referred to as a "datacube", which is a 3D array containing two dimensions in spatial and one dimension in spectral. Therefore, it is a combination of 2D images at all wavelength pixels, as shown in Fig. 1.1. IFS has the unique advantage of facilitating studies of specific regions of celestial objects and enables a broad range of scientific topics. The construction of an IFS instrument includes the use of lenslet array (Courtes, 1982), fiber bundles (with/without lenslets, e.g., Vanderriest 1980; Bland-Hawthorn et al. 2011; Bryant et al. 2011), or image slicer (Content, 1997), as shown in Fig. 1.1.

The Multi Unit Spectroscopic Explorer (MUSE, Bacon et al. 2010, 2014) is an IFS instrument for the Very Large Telescope (VLT, 8 m) of the European Southern Observatory (ESO), as shown in Fig. 1.2. During observation, it divides the adaptive optics corrected field-of-view into 24 sub-fields. Each sub-field is then fed into a spectrograph to take the spectrum (called Integral-field Unit, IFU), with an image slicer at the forefront acting as an entrance slit. Taking advantage of the large telescope aperture and increased spatial resolution provided by adaptive optics (see Table 1.1), MUSE can provide spatially re-

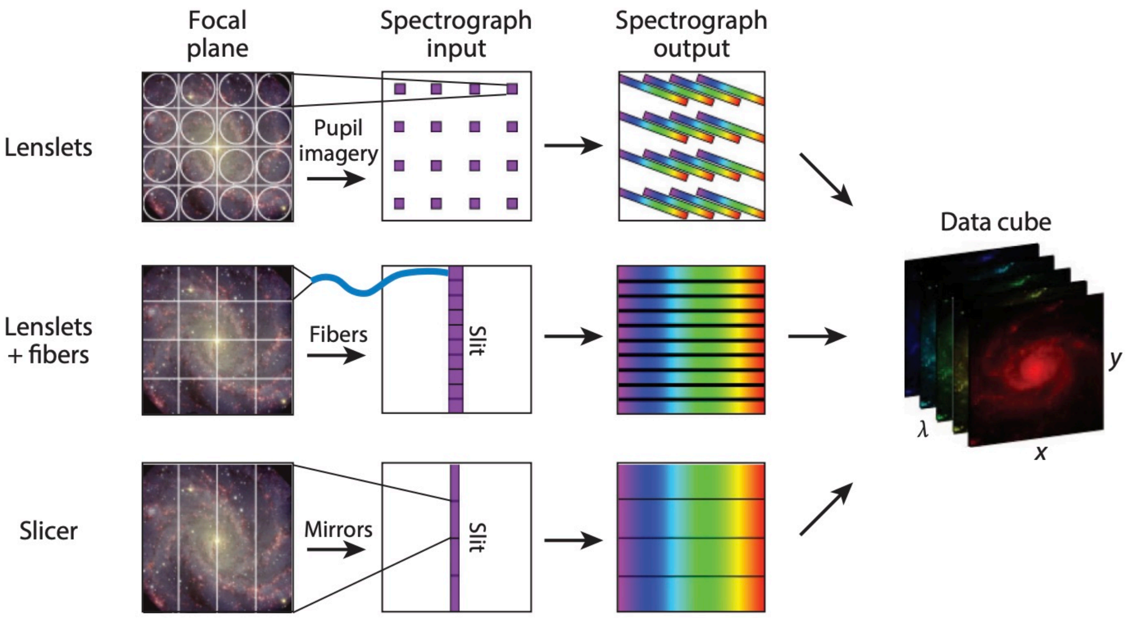

Figure 1.1: Configurations of an integral-field spectrograph datacube. Figure adapted from Eisenhauer & Raab (2015).

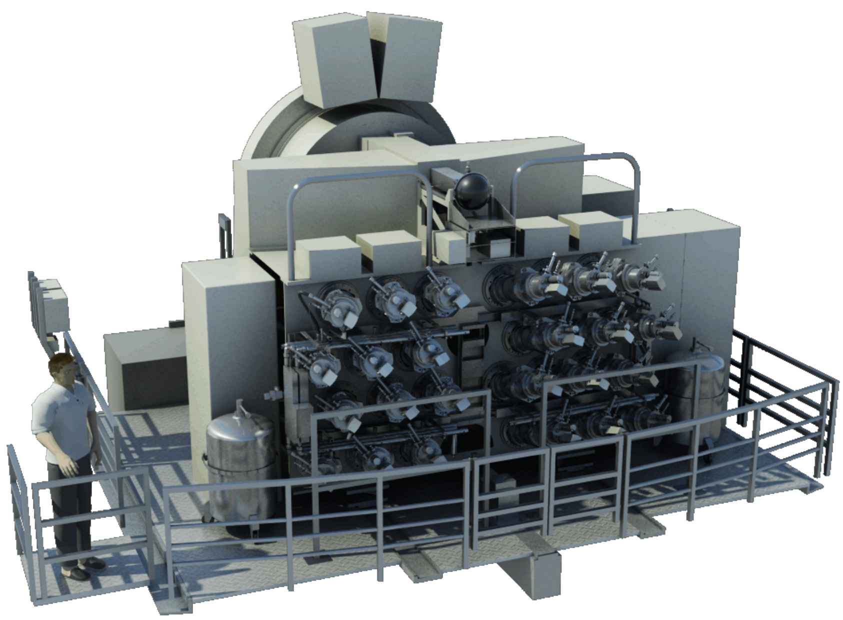

Figure 1.2: 3D modeling of the MUSE instrument. Figure from `https://www.eso.org/sci/facilities/develop/instruments/muse.html`.

Table 1.1: Observational parameters of the MUSE instrument. Table adapted from https://www.eso.org/sci/facilities/develop/instruments/muse.html

Observational Parameters	
Name	Multi Unit Spectroscopic Explorer
Site	Paranal
Telescope	Very Large Telescope (8 m), UT4
Focus	Nasmyth
Type	Integral field spectrograph, imager
Wavelength coverage	465–930 nm
Spectral resolution $\lambda/\Delta\lambda$	2000@460 nm
	4000@930 nm
Observing Mode	Wide Field Mode (WFM)
	Narrow Field Mode (NFM)
First light date	March 2014
Wide Field Mode	
Spatial resolution	0.2×0.2 arcsec^2
Spatial resolution (FWHM)	$0.3 \sim 0.4$ arcsec
Field-of-View	1×1 arcmin^2
Narrow Field Mode	
Spatial resolution	0.025×0.025 arcsec^2
Spatial resolution (FWHM)	$0.03 \sim 0.05$ arcsec
Field-of-View	7.5×7.5 arcsec^2

solved spectroscopic data in greater detail and making it one of the most over-subscribed instruments. To date, MUSE has been used to observe 17,000 fields across the southern sky for a diverse range of studies, including galaxy evolution (e.g., Foster et al. 2021; van de Sande et al. 2023), exoplanets (Irwin et al., 2018; Haffert et al., 2019), Galactic regions (Weilbacher et al., 2015; McLeod et al., 2015), gravitational waves (Tanvir et al., 2017; Levan et al., 2017), gas clouds around quasars (Borisova et al., 2016), dark matter (Massey et al., 2015), stellar-mass black hole (Giesers et al., 2018), and resolved stellar clusters (Husser et al., 2016; Kamann et al., 2018). All the works in this thesis are related to the MUSE instrument.

1.2 The Milky Way and Galactic Archaeology

Our Milky Way is a luminous ($L_\star$) barred spiral (Sb–Sbc) with a boxy/peanut bulge, a dominant disk, and a diffuse stellar halo. It falls in the so-called "green valley" in the colour-magnitude diagram (Mutch et al., 2011; Licquia et al., 2015), which indicates a potential transformation from star-forming (blue cloud) into a quenched galaxy (red sequence) with the lack of gas supplies. But the Galaxy seems to be unusual due to the evidence of ongoing gas accretion in the Magellanic gas stream (Putman et al., 1998; Fox et al., 2014) and the appearance of two luminous star-forming dwarf galaxies (Robotham et al., 2012). The global properties of the Milky Way are summarized in Bland-Hawthorn & Gerhard (2016). Basically, it has a photometric magnitude of $M_V = -21.37$ with color

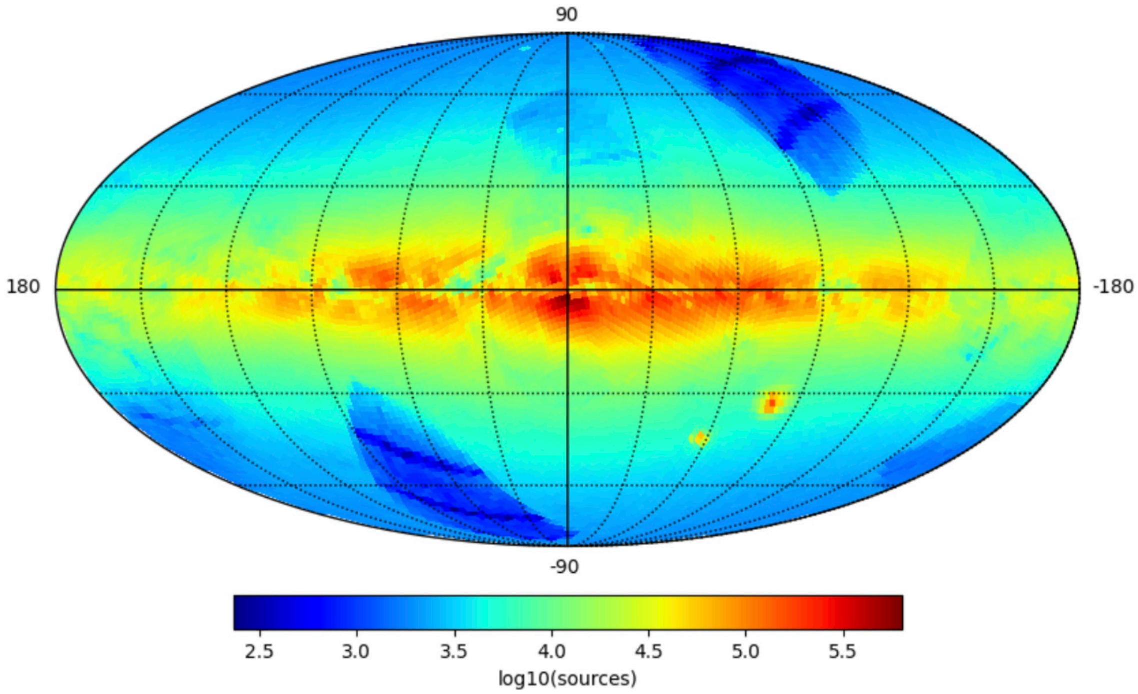

Figure 1.3: Density of *Gaia* DR3 sources in the sky with available in Galactic coordinates. Figure adapted from (Babusiaux et al., 2022).

$B - V = 0.73$. The total stellar mass is $5 \pm 1 \times 10^{10}$ $M_\odot$ with a star formation rate of $\dot{m}_\star = 1.65 \pm 0.19$ $M_\odot$ year^{-1} (Licquia & Newman, 2015). The Sun is within the Galaxy with a Galactocentric distance of $R_{gc} = 8.2 \pm 0.1$ kpc and an offset of $z = 25 \pm 5$ pc (Jurić et al., 2008) from the local disk midplane. Its tangential velocity relative to the Galactic Center Sgr A* is 248 ± 3 km s^{-1}. Due to this unique vantage point, the Milky Way will always be the most important benchmark for galaxy evolution because it provides fully resolved information that few other galaxies can offer (Bland-Hawthorn & Gerhard, 2016). The study of this resolved information gave rise to Galactic archaeology, which aims to unravel the formation and evolution history of the Milky Way using individual stars as fossil records (Freeman & Bland-Hawthorn, 2002; Bland-Hawthorn & Gerhard, 2016). This idea was present in early studies of moving stellar groups (Eggen, 1974). In this section, we will introduce the Galactic stellar components, *Gaia* mission with other spectroscopic surveys, which are the dataset used for Galactic archaeology, chemical cartography, and chemical evolution models. We also point out the unexplored regions that this thesis will focus on.

1.2.1 *Gaia* and Fiber-Fed Spectroscopic Surveys

The *Gaia* satellite (Perryman et al., 2001) is the second space-based astrometric mission conducted by European Space Agency (ESA) and it was launched in December 2013. *Gaia* aims to scan the whole sky with unprecedented sensitivity (~ 10 μas in position) 140 times more accurate than its predecessor *Hipparcos* (Høg et al., 2000) and measure astrometric and photometric parameters of celestial sources including position (α, δ), parallax (ϖ), proper motion (μ_α, μ_δ), radial velocity, magnitude (in G, G_{BP}, and G_{RP} bands), and some stellar atmospheric parameters ($T_{\rm eff}$, $\log g$, and elemental abundances) (Gaia

Collaboration et al., 2016a). As of June 2023, *Gaia* has provided four data releases (Gaia Collaboration et al., 2016b, 2018, 2021, 2022b) including a total number of 1.59 billion sources within $3 < m_G < 21$, in which 1.46 billion sources have full astrometric solutions $(\alpha, \delta, \varpi, \mu_\alpha, \mu_\delta)$ and 33 million stars have radial velocities. The full information and parameter uncertainties can be accessed via the newest *Gaia* DR3 summary website[1]. There have been several studies on the Milky Way structures in greater detail using the six-dimensional phase space information (x, y, z, v_x, v_y, v_z) from *Gaia*, which include the evidence of merger events from substructures in the stellar halo (Helmi et al., 2017), the wraps at the edges of the Galaxy (Poggio et al., 2017), the phase spiral structure (Antoja et al., 2018), and the spiral structure using upper main-sequence stars (Poggio et al., 2021), etc. Studies mentioned above are only a small subset of all the discoveries by the wealthy information of *Gaia*.

The newest *Gaia* DR3 (Gaia Collaboration et al., 2022b) additionally provides stellar atmospheric parameters ($T_{\rm eff}$, $\log g$, [M/H], [X/M] for 12 elements) estimated from RVS spectra ($R \sim 11000$) for 5.5 million stars (see the parameter distribution maps in Gaia Collaboration et al. 2022a), and mean BP/RP spectra ($R \sim 20-115$) for 219 million stars with $m_G < 17.6$ for most of them (see the distribution of BP/RP sources in Fig. 1.3). Importantly, recent studies have developed data-driven models to measure $T_{\rm eff}$, $\log g$, metallicity, revised parallax, and extinction from these 219 million BP/RP spectra with precision better than 0.15 dex in metallicity and $\log g$, and 90 K in $T_{\rm eff}$ (e.g., Andrae et al. 2023; Zhang et al. 2023). This dataset permits comprehensive studies on the three-dimensional dust distribution, metallicity gradient in the LMC, chemodynamics of the Galactic disk, and stellar rotation in open clusters (Andrae et al., 2023), and also provides prospects for larger samples in future *Gaia* DR4.

In addition to *Gaia*, ground-based telescopes with fiber-fed multi-object spectrograph provide essential datasets for the study of Galactic archaeology. Compared to space-based instruments, these surveys either have higher spectral resolution ($R \sim 10^4$) or wider wavelength range from near-ultraviolet (NUV) to the infrared (IR). Although having less sources than space satellites, these ground-based surveys can obtain more stellar parameters with better precision.

Over the past two decades, many Galactic spectroscopic surveys have been conducted to derive millions of individual stellar spectra. With the use of radiation transfer codes (e.g. SME Valenti & Piskunov 1996, Turbospectrum Alvarez & Plez 1998; Plez 2012, MOOG Sneden et al. 2012), many model-driven (e.g. *The Payne* Ting et al. 2019 and *StarNet* Fabbro et al. 2018; Bialek et al. 2020) and data-driven methods (e.g. *The Cannon* Ness et al. 2015 and *Astro-NN* Leung & Bovy 2019) and models in between (e.g. *DD-Payne* Xiang et al. 2019 and *Cycle-StarNet* O'Briain et al. 2021) have been developed to measure accurate stellar atmospheric parameters ($T_{\rm eff}$, $\log g$) and more than 20 elemental abundances, and each of them is important and irreplaceable (Ting & Weinberg, 2022). We summarize detailed information of each survey in Table 1.2 along with *Gaia*-RVS (Recio-Blanco et al., 2022) and *Gaia*-BP/RP (De Angeli et al., 2022). Overall, the combination of astrometric information from *Gaia* and astrophysical parameters from these spectroscopic surveys permits studies on both phase and chemical space, to unravel chemodynamical

[1]https://www.cosmos.esa.int/web/gaia/dr3

Table 1.2: Summary of the major Galactic spectroscopic surveys. Ongoing and scheduled surveys are marked with an asterisk.

Survey	Instrument	Location	Sky coverage	Operation period	Wave range (Å)	Magnitude (mag)	No. of stars (data release)	Resolution ($\lambda/\Delta\lambda$)	σ_{RV} (km s^{-1})	Available parameters	Ref.
RAVE	UKST/6dF	Siding Spring (SSO)	southern	2003 – 2013	8410 – 8795	$9 < I < 12$	451, 783 (DR6)	7500	1.5	$T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], Al, Ni	1,2
APOGEE	SLOAN/APOGEE-N Irénée du Pont/APOGEE-S	Apache Point (APO) Las Campanas (LCO)	both	2014 – 2020	15000 – 17000	$11 < H < 14$	657,000 (DR17)	22500	0.1	$T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], $vsini$, v_{macro}, C, C-I, N, O, Na, Mg, Al, Si, S, K, Ca, Ti, Ti-II, V, Cr, Mn, Fe, Co, Ni, Ce	3
*GALAH	AAT/HERMES	Siding Spring (SSO)	southern	2014 – present	4718 – 7890	$11 < V < 14$	588, 571 (DR3)	28000	0.1	$T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], v_{mic}, v_{broad}, Li, C, O, Na, Al, K, Mg, Si, Ca, Ti, Ti-II, Sc, V, Cr, Mn, Co, Ni, Cu, Zn, Y, Ba, La, Rb, Mo, Ru, Nd, Sm, Eu	4
*LAMOST-LRS	LAMOST/Guoshoujing (郭守敬望远镜)	Xinglong (兴隆)	nothern	2012 – present	3700 – 9000	$r < 19$	10, 336, 752 (DR8)	1800	7	$T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], v_{mic}, C, N, O, Na, Mg, Al, Si, Ca, Ti, Cr, Mn, Fe, Co, Ni, Cu, Ba [a]	5,6
*LAMOST-MRS	LAMOST/Guoshoujing (郭守敬望远镜)	Xinglong (兴隆)	nothern	2018 – present	4950 – 5350 6300 – 6800	$G \sim 15$	4, 684, 915 (DR8)	7500	1	$T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], $vsini$, C, N, O, Mg, Al, Si, S, Ca, Ti, Cr, Ni, Cu	7
Gaia-ESO	VLT/FLAMES	Paranal (ESO)	southern	2011 – 2018	3000 – 11000	$V < 19$	114, 917 (DR5)	5000 – 47000	0.3	$T_{\rm eff}$, $\log g$, v_{mic}, $vsini$, He, Li, C, N, O, Na, Mg, Al, Si, S, Ca, Sc, Ti, V, Cr, Mn, Co, Ni, Cu, Zn, Sr, Y, Zr, Mo, Ba, La, Ce, Pr, Nd, Sm, Eu	8
*WEAVE-GA	WHT/WEAVE	La Palma (ORM)	northern	2022 – present	3660 – 9590 4040 – 6850	$16 < G < 20.7$ $12 < G < 16$	1.8 – 2.6 million 1.1 – 1.6 million	5000 20000	1 ~ 2 0.5	$T_{\rm eff}$, $\log g$, [Fe/H] from both R, light, α-, Fe-peak, and s- and r-process elements from high-R	9
*4MOST	VISTA/4MOST	Paranal (ESO)	southern	2024-Q2 (sched)	3900 – 9500 3926 – 6790	$15 < G < 20$ $G < 15.5$	> 15 million > 3 million	4000 – 7500 18000 – 21000	1 ~ 2 0.3	$T_{\rm eff}$, $\log g$, [Fe/H], and up to 20 elements	10-13
*Milky Way Mapper	SLOAN+Irénée du Pont/BOSS SLOAN+Irénée du Pont/APOGEE	Apache Point (APO) Las Campanas (LCO)	both	2020 – present	3600 – 10400 15000 – 17000	-	5 million	2000 22500	-	$T_{\rm eff}$, $\log g$, [Fe/H], and chemical abundances	14
*Gaia-RVS	Gaia/RVS	L2 (ESA)	whole	2014 – present	8450 – 8720	$G < 17$	5, 591, 594 (DR3)[b]	11500	1.05	$T_{\rm eff}$, $\log g$, [Fe/H], N, Mg, Si, S, Ca, Ti, Cr, FeI, FeII, Ni, Zr, Ce and Nd	15,16
*Gaia BP/RP	Gaia/BP+RP	L2 (ESA)	whole	2014 – present	3300 – 10500	$G < 17.6$	219 million (DR3)	20 – 115	-	$T_{\rm eff}$, $\log g$, [M/H], parallax, $E(B-V)$	17-19

[a] Elemental abundances are measured from LAMOST DR5 by *DD-Payne* (Ting et al., 2017a; Xiang et al., 2019).
[b] Among this number, about 2,513,593 sources have chemical abundances measurements.
References. (1) Steinmetz et al. (2020a); (2) Steinmetz et al. (2020b); (3) Abdurro'uf et al. (2022); (4) Buder et al. (2021); (5) Ting et al. (2017b); (6) Xiang et al. (2019); (7) Liu et al. (2020); (8) Gilmore et al. (2012); (9) Jin et al. (2023); (10) Helmi et al. (2019); (11) Christlieb et al. (2019); (12) Chiappini et al. (2019); (13) Bensby et al. (2019); (14) Almeida et al. (2023); (15) Recio-Blanco et al. (2022); (16) Creevey et al. (2022); (17) De Angeli et al. (2022); (18) Andrae et al. (2023); (19) Zhang et al. (2023)

evolution history of the Milky Way. Some examples will be shown in Section 1.2.2 and Section 1.2.3. In recent years, spectroscopic surveys have entered a new era with the ongoing WEAVE (Jin et al., 2023) and 4MOST (de Jong et al., 2019) surveys in optical, which will obtain stellar spectra down to $G \sim 20$ and many stellar parameters are measurable. The Milky Way Mapper (Almeida et al., 2023) as part of the fifth Sloan Digital Sky Survey (SDSS-V) will focus on the Galactic central and disk regions with unprecedented number of samples for the study of stellar components evolution.

1.2.2 Galaxy Stellar Components

The basic stellar components of the Milky Way include a boxy/peanut bulge in the densest Galactic center with a bar, an extended disk with flaring in the outer region, a diffuse and spherical halo, and several dense stellar regions in the inner halo called star clusters. We show the anatomy of the Milky Way with these components in Fig. 1.4.

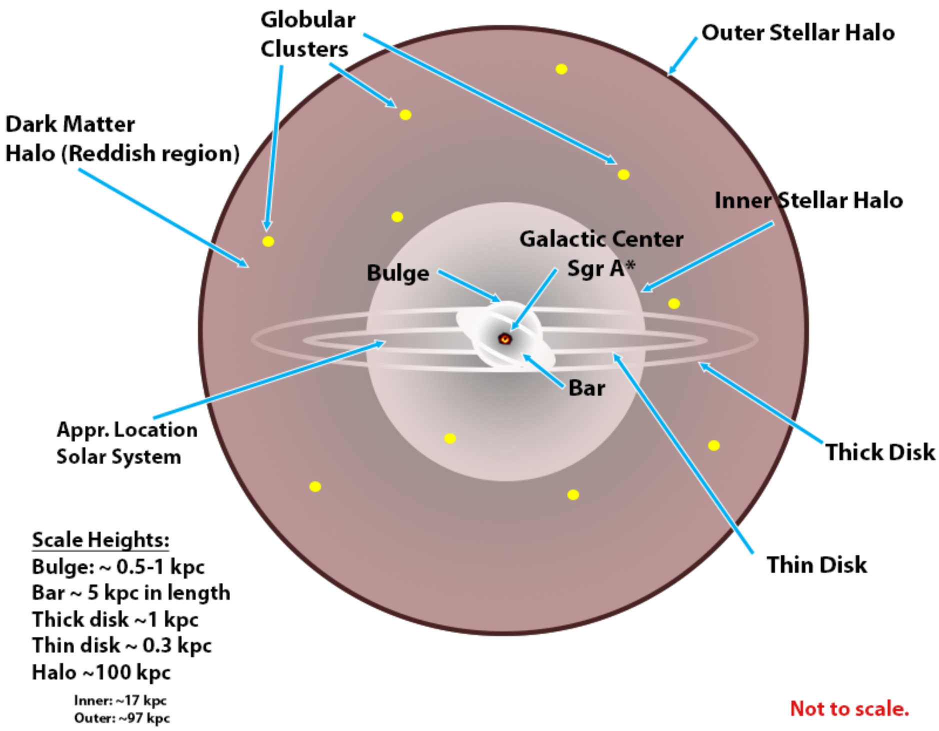

Figure 1.4: The anatomy of the Milky Way with its various components (not to scale). Credit: Matthias Schmitt 2019.

Bulge

The stellar bulge was once considered to be a spherical *classical bulge* with stars formed rapidly at the early stages of the Galaxy through gas-rich mergers (Ortolani et al., 1995; Elmegreen, 1999; Clarkson et al., 2008). But another theory proposed a *pseudo-bulge*, which is a natural consequence of the evolution of stars in the disk (Combes & Sanders, 1981; Raha et al., 1991; Norman et al., 1996; Athanassoula, 2005). In this case, dynamical forces caused stars in the inner disk region to depart from circular orbits and led to the formation of the bulge. Later on, studies using 2MASS star count (Skrutskie et al., 2006) confirmed a boxy/peanut (b/p) or X-shape bulge due to the appearance of the Galactic bar (McWilliam & Zoccali, 2010; Nataf et al., 2010; Wegg & Gerhard, 2013) where stars have a cylindrical rotation (Kunder et al., 2012; Ness et al., 2013b). Therefore, it indicates the bulge has a more complex scenario with multiple formation and evolution processes. Combining the kinematic and chemical observations, Soto et al. (2007); Babusiaux et al. (2010) found two main bulge stellar populations and only the more metal-rich population follows the bar. Rojas-Arriagada et al. (2014) found two equal numbers of metal-rich and metal-poor stars from Gaia-ESO, whereas Ness et al. (2013a,b) find evidence for five populations. They all agree that the metal-rich populations trace the X-shape or the barred bulge. For metal-poor populations, studies using RR Lyrae found that the old population does not participate in the b/p bulge (Dékány et al., 2013; Pietrukowicz et al., 2015), consistent with the results from ARGOS (Ness et al., 2012). It is debated whether these stars are formed from an old bulge in the proto-Galaxy or just a reflection of accreted features with orbits in the pericenter phase due to instabilities with the halo (Babusiaux et al., 2010; Di Matteo et al., 2014), or both. A more recent study using *Gaia*-BP/RP metallicity by Rix et al. (2022) found a centrally concentrated metal-poor and old component with most of the stars at apocenters, but a minority at pericenter that can be taken to $R_{\rm gc} > 10$ kpc. Wylie et al. (2021) combined ARGOS and APOGEE samples (A2A) and conclude that each [Fe/H] maximum have separate disk origins and metal-poor stars originating from a thicker disk than the metal-rich stars.

Disk

The Galactic disk is a flattened and kinematically cold structure than other components. Previous studies have observed that the disk can be broken into two components, a thin and thick disk, based on geometry, chemical abundances, kinematics, and age. However, different definitions and selection criteria for the thin and thick disks do not produce identical sub-samples of stars. Some studies have even argued that the thin and thick disks are not completely distinct components (e.g., Bovy et al. 2012a,b,c; Sharma et al. 2021b). Studies on the shape of the disk began with Gilmore & Reid (1983) that they made the first reliable measurement of the stellar densities vertical to the Galactic plane using photometric parallax and geometrically demonstrated a two-component exponential distribution with scale-heights z around 300 pc and 1450 pc, respectively. Later on, these values are improved to be $350 \sim 750$ and 940 pc (Siegel et al., 2002). The thin disk and thick disk have a stellar mass of $3.5 \pm 1 \times 10^{10}$ $M_\odot$ and $6 \pm 3 \times 10^9$ $M_\odot$ (Bland-Hawthorn & Gerhard, 2016), respectively. In terms of chemistry, high-resolution spectroscopic surveys indicated that most stars in the thin disk have high [Fe/H] and low [α/Fe], whereas stars in the thick disk have low [Fe/H] and high [α/Fe] (Bensby et al., 2003, 2014; Nidever et al., 2014; Hayden et al., 2015). This is the so-called α-bimodality. The thick disk stars also have higher [C/N] from observations of APOGEE DR12 (Masseron & Gilmore, 2015). In the solar neighbourhood, a distinct division appears in the stellar age distribution at around 8 Gyr. Stars younger than this age are primarily located in the thin disk, whereas older stars belong to the thick disk, and a strong correlation is observed between stellar age and $[\alpha/\mathrm{Fe}]$ (Haywood et al., 2013; Bensby et al., 2014), but not in age and [Fe/H] (Edvardsson et al., 1993; Nordström et al., 2004). The $[\alpha/\mathrm{Fe}]$-poor population is also found in the Sagittarius dwarf galaxy but with a different pattern to the thick disk (Hasselquist et al., 2019). In the outer disk, a flaring profile of low $[\alpha/\mathrm{Fe}]$ stars is observed with slowly exponential increase with the Galactocentric radius (Bovy et al., 2016). The flaring is present in almost all mono-age populations with the youngest populations flaring most strongly and is thought to be a result of active accretion history in the early disk life (Mackereth et al., 2017). Combining all the results above, we show in Fig. 1.5 a sketch of the α abundance distribution of the Milky Way, which is adapted from Bland-Hawthorn et al. (2019). We will illustrate the chemical cartography of the disk in greater detail in Section 1.2.3 and connect this to chemical evolution models in Section 1.2.4.

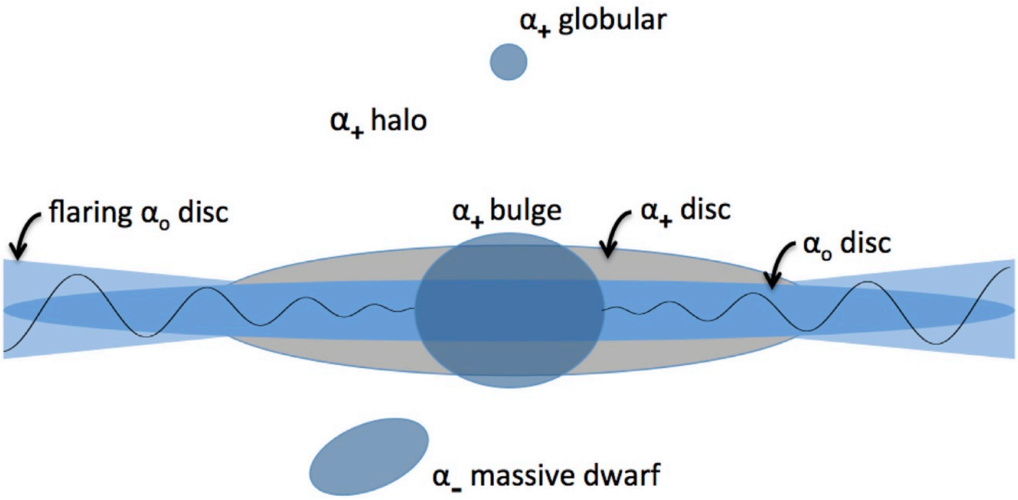

Figure 1.5: Schematic diagram illustrating the structure of the Galactic disc. α_0 and α_+ means $[\alpha/\mathrm{Fe}]$-poor and $[\alpha/\mathrm{Fe}]$-rich component, respectively. Figure adapted from Bland-Hawthorn et al. (2019).

Halo

Stars in the Galactic halo follow a double power-law distribution with a slope of -2.5 ± 0.3 for the inner halo ($5 \sim 25 \pm 10$ kpc) and -3.8 ± 0.1 for the outer halo (Bland-Hawthorn & Gerhard, 2016). Even though the halo contains only 1% of the total stellar mass (Morrison, 1993), it is considered to be an essential component to reveal the merger history of the Galaxy. Most of the mass in the inner halo regions is predicted to originate in a few massive progenitors (Helmi et al., 2002; Wang et al., 2011), where these ancient galaxies were disrupted into debris by the Milky Way (Cooper et al., 2010), and some of them with low-inclination orbits can be deposited in the thick disk (Abadi et al., 2003). Many tracers can be used to study the formation history of the Galactic halo, which includes satellite galaxies (Watkins et al., 2010), globular clusters (e.g., Sohn et al. 2018; Watkins et al. 2019), halo stars (Xue et al., 2008; Kafle et al., 2012), high-velocity stars (Rossi et al., 2017; Monari et al., 2018), stellar streams (Gibbons et al., 2014; Malhan & Ibata, 2019; Erkal et al., 2019), and heated stars from the disk (Zolotov et al., 2009; Tissera et al., 2013). Similar to the bulge's old component, the halo is dominated by old and metal-poor stars formed in the proto-Galaxy. The metallicity of accreted stars is even more metal-poor because the accreted galaxies were less massive than the proto-Galaxy and hence have lower metallicity due to the mass-metallicity relation (Helmi, 2020). In recent years, studies using *Gaia* metallicity and kinematics found the most massive accreted progenitor *Gaia*-Sausage-Enceladus (GSE) (Belokurov et al., 2018; Deason et al., 2018; Helmi et al., 2018), which had a stellar mass of $\sim 6 \times 10^8\ M_\odot$ (Helmi et al., 2018) and merged with the Milky Way around 11 Gyr ago (Xiang & Rix, 2022). In addition, other substructures in the inner halo such as the Helmi streams (Helmi et al., 1999), Sequoia (Myeong et al., 2019), and Thamnos (Koppelman et al., 2019), are also readily apparent in *Gaia* data (Koppelman et al., 2018). Besides, many tiny streams crisscrossing the halo at larger distances have been uncovered as shown in Fig. 1.6. Horta et al. (2023) combined APOGEE with *Gaia* and found the elemental properties of various halo substructures have different nucleosynthetic pathways.

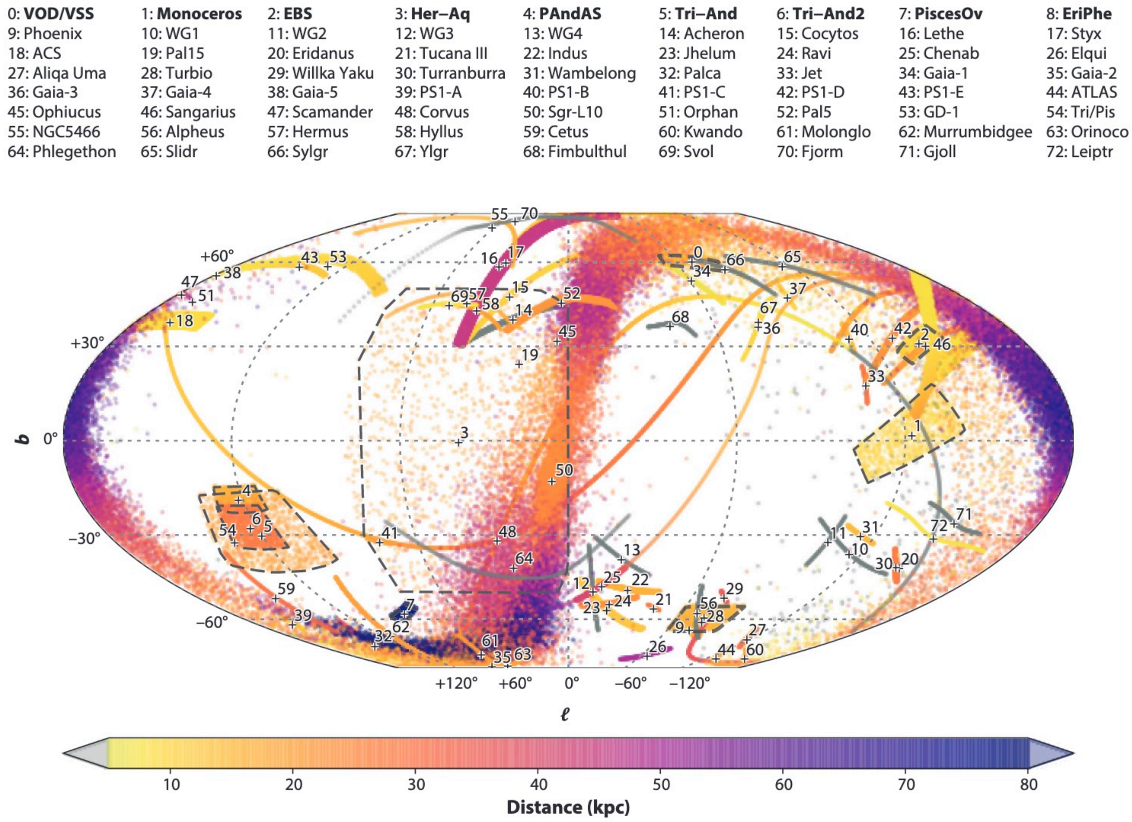

Figure 1.6: Sky distribution of currently known spatially coherent streams and overdensities (indicated as regions delimited by dashed lines, and in boldface in the inset). Figure adapted from Helmi (2020).

Clusters

Stellar clusters can be roughly defined as overdensities of the background stellar distribution of the Milky Way and are gravitationally bound for some length of time (Lada & Lada, 2003), which is very useful to provide a way to study and model stellar evolution and ages (Krumholz et al., 2019). The two traditional categories of star clusters are open clusters (OC) and globular clusters (GC).

Open clusters contain a few hundred loosely bound stars. They reside exclusively in the young and low-[α/Fe] disk and rarely exceed $|z| = 200$ pc (Cantat-Gaudin et al., 2018). They have an average age of $\sim 10^8$ Gyr and are rarely older than 500 Myr (Salaris et al., 2004). Due to their young ages, OCs are identified as tracers of the Galactic spiral arms and star-forming regions. Stars in OCs are formed from the same initial molecular cloud and hence are believed to have the same age and chemical composition (e.g., De Silva et al. 2006, 2007; de Silva et al. 2011; Feng & Krumholz 2014), even though some recent studies on solar twins identified certain inhomogeneous elements with variations around 0.02 dex (Liu et al., 2016a,b, 2019). Under this assumption, it is possible to identify the common birth sites of disrupted OC members only using their chemical signatures, which is the so-called chemical tagging (Freeman & Bland-Hawthorn, 2002).

Globular clusters contain several thousand to one million stars in spherical and gravitationally-bound systems, which were previously difficult to resolve due to the high stellar density. They are mostly in the Galactic halo with higher initial mass and at least 8 Gyr old, many of them are as old as the universe (Chaboyer et al., 1996). GCs are crucial to understanding the origin of the halo. Dynamical analysis of GCs found that about 40% of them contain old stars originating from the disk and bulge, and 35% are from merger events (Massari et al., 2019). GCs are observed to have inhomogeneities and anticorrelations in most of the light elements such as He, C, N, O, Na, and Al (e.g, Carretta et al. 2010; Milone et al. 2018; Mészáros et al. 2020). We refer to the detailed review from Milone & Marino (2022). Heavy-element abundance variations are also observed, but only in a small number of massive GCs (Gratton, 2020). Using integral-field spectrographs such as MUSE, several studies have been conducted to study the kinematics and abundance variations of GCs using resolved individual stars (e.g., Husser et al. 2016; Kamann et al. 2016, 2018; Göttgens et al. 2019; Husser et al. 2020), and a potential central black hole was identified in NGC 6397 (Kamann et al., 2016).

1.2.3 Chemical Cartography of the Disk

In this section, we focus on the detailed chemical cartography of the Galactic disk, including the [α/Fe]-[Fe/H] distribution and the metallicity distribution function (MDF) at different locations with their indications of dynamical processes. Here we refer to results from Hayden et al. (2015) using red giant stars observed by APOGEE DR14 (Abolfathi et al., 2018). We note an earlier version also appeared in Nidever et al. (2014) using red clump stars from APOGEE (see their Fig. 11). For a more general chemical cartography of the entire Milky Way with all the structures, we refer to a more recent study using *Gaia* DR3 (Gaia Collaboration et al., 2022a).

Fig. 1.7 demonstrates APOGEE DR14 stellar density distributions in the [α/Fe]-[Fe/H] plane as a function of Galaxy locations in R and $|z|$ from Hayden et al. (2015). This figure reveals several features: (i): The distribution can be separated into two sequences: one is [Fe/H]-rich and [α/Fe]-poor, and the other is [Fe/H]-poor and [α/Fe]-rich. Whether they are distinct and their relative fractions depend on the Galactic location; (ii): The distribution of the inner Galaxy ($3 < R < 5$ kpc) is close to a single sequence starting from low-[Fe/H], high-[α/Fe] to high-[Fe/H], low-[Fe/H], whereas the two sequences are clearly distinct in the outer Galaxy regions; (iii): The density peaks of the high-[α/Fe] sequences remain the same at different radial zones, whereas for the low-$[\alpha/\mathrm{Fe}]$ sequences the peaks decrease with a larger Galactocentric radius; (iv): The relative fraction of low-$[\alpha/\mathrm{Fe}]$ sequences increases with larger R and smaller $|z|$, indicating a thin disk structure with larger scalelength and larger scaleheight, and flaring in the outer Galaxy; the relative fraction of high-$[\alpha/\mathrm{Fe}]$ sequences increases with smaller R and larger $|z|$, indicating a thick disk structure with shorter scalelength and higher scaleheight. The MDF of stars in each panel is also shown in Fig. 1.8 (adapted from Hayden et al. 2015), where it is clear that the peak of MDF is centred at high metallicity in the inner Galaxy, and low metallicity in the outer Galaxy. The shape of MDF also changes from negative skewness in the inner Galaxy to positive skewness in the outer Galaxy. Moreover, with the increase of $|z|$, MDF has less variation with radius.

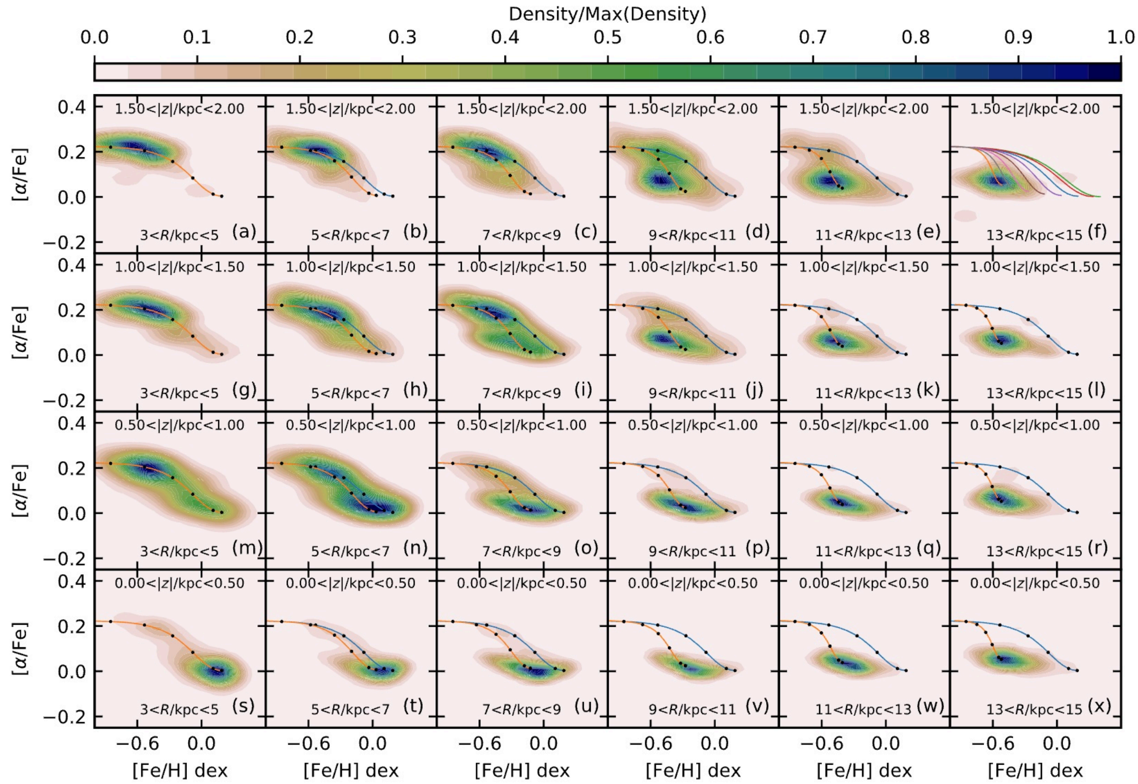

Figure 1.7: Density distribution revealed by APOGEE DR14 stars in the ($[\alpha/\mathrm{Fe}]$, $[\mathrm{Fe/H}]$) panel at different Galaxy locations. In each panel, the sum of density is normalized to 1 and the locations are in cylindrical coordination (R, z) in kpc to the Galactic center. The solid lines show the evolution of abundances at a given stellar birth radius predicted by Sharma et al. (2021b). The blue line is for the birth radius of 4 kpc, while the orange line is for the birth radius corresponding to the central value of R in each bin. The black dots mark the evolution at the ages of 4, 8, 10, 11, 12, and 13 Gyr, with metallicity decreasing with age. Panel (f) shows the model predictions corresponding to birth radii of 1, 2, 4, 6, 8, 10, 12, and 14 kpc, with $[\alpha/\mathrm{Fe}]$ increasing with birth radius. This figure is originated in Hayden et al. (2015) and adapted from Sharma et al. (2021b). We note an earlier version also appeared in Nidever et al. (2014) using red clump stars from APOGEE (see their Fig. 11).

The above chemical cartography from observations of the disk and its variance with Galaxy location is important for an insight into the chemodynamical evolution of the Milky Way

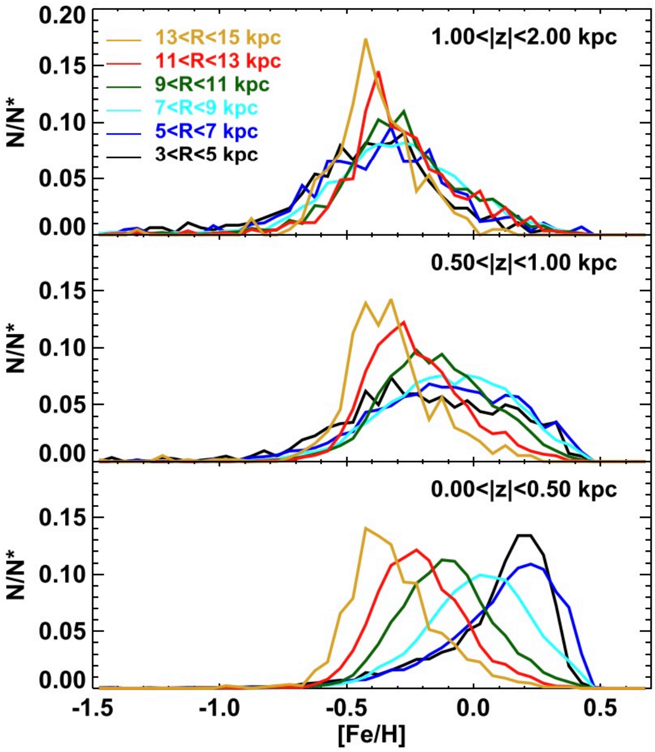

Figure 1.8: Observed metallicity distribution function for the entire sample as a function of Galactocentric radius over a range of distances from the plane. Figure adapted from Hayden et al. (2015).

and provides useful constraints to chemical evolution models and simulations. The skewness of MDF and its variance with the Galactocentric radius is verified to be due to the radial migration or churning by a toy model in Hayden et al. (2015) that stars can migrate away from their birth radius (e.g., Sellwood & Binney 2002). Stellar radial migration can be caused by a variety of reasons including interactions with infalling satellites (Quillen et al., 2009), giant molecular clouds in the disc (Schönrich & Binney, 2009a), or interactions with the bar or spiral arms, or both (e.g., Brunetti et al. 2011; Di Matteo et al. 2013; Daniel & Wyse 2015; Loebman et al. 2016). The vertical structure or flaring in the outer disk is found to be due to the vertical velocity dispersion σ_z decreasing until solar angular momentum L_z and increasing back later on (Sharma et al., 2021b). This is a fundamental relation that exists in all spectroscopic surveys (Sharma et al., 2021a), which may be caused by orbiting satellites, warps, the infall of misaligned gas, and reorientation of the disc (e.g., Minchev et al. 2014, 2015; Grand et al. 2016).

1.2.4 Chemical Evolution Models

Historically, many efforts have been made to build Galactic chemical evolution (GCE) models and compare with observations to understand the evolution processes of the Milky Way. The simplest one is the one-zone model (van den Bergh, 1962; Schmidt, 1963) which assumes stars begin to form from pristine gas that only contains H, He, and a small amount of Li from the Big Bang, and the heavier elements are produced later on by stellar nucleosynthesis processes and supernova. The falling gas from intergalactic medium can also be added into the model over time to moderate the evolution track. One-zone models were later extended into multi-zone by including a series of interconnected galactic rings (Chiappini et al., 1997, 2001) where stars and gas can be exchanged. Some other processes such as inside-out growth indicated by Galactic and extragalactic observations (Bovy et al., 2012b; van der Wel et al., 2014; Rodríguez-Puebla et al., 2017) are also proved to be essential to GCE models to match observations.

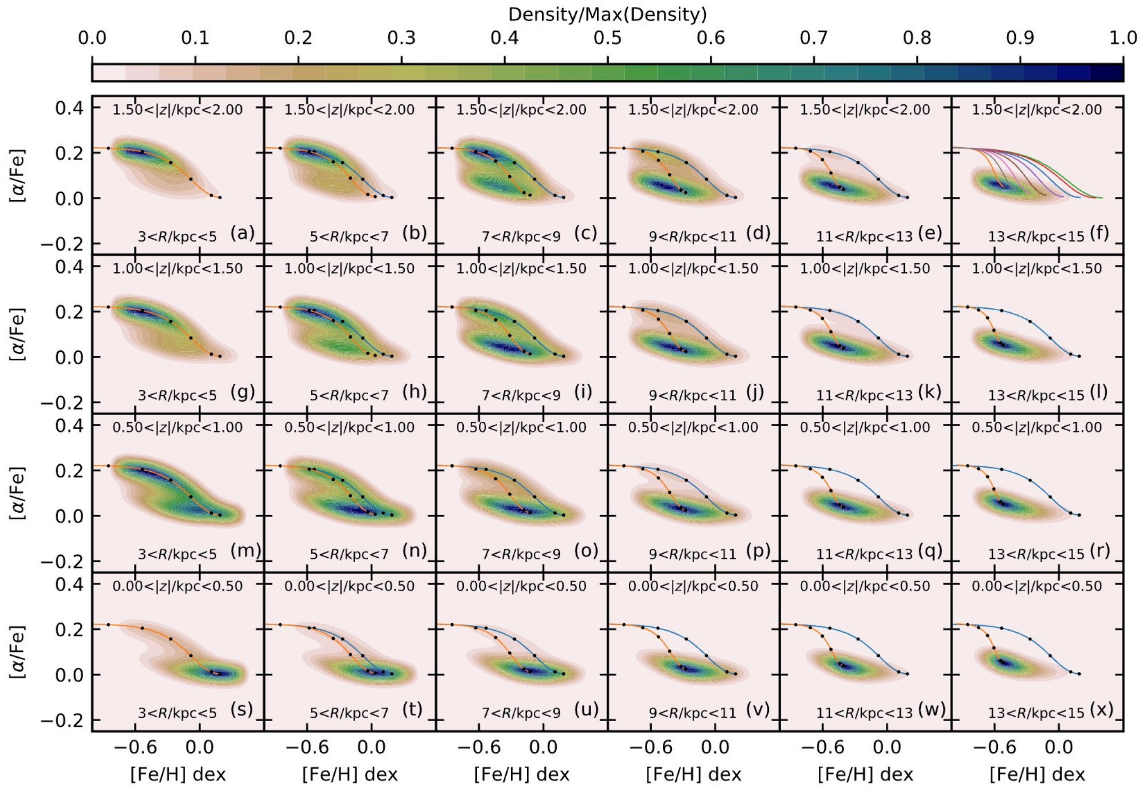

Figure 1.9: Predictions by Sharma et al. (2021b) for stellar density distributions in the ([α/Fe], [Fe/H]) plane at different Galaxy locations. The definitions of solid lines, contours, and black dots are the same as Fig. 1.7. Figure adapted from Sharma et al. (2021b).

One key factor in determining the correction of GCE models is whether it can reproduce the [α/Fe]-[Fe/H] bimodality distribution observed by spectroscopic surveys, as we illustrated in Section 1.2.3. Here we list several GCE models that made this attempt and were mostly cited during the past decades. The two-infall model (originally from Chiappini et al. 1997 and modified later by Haywood et al. 2018; Spitoni et al. 2019; Lian et al. 2020a) assumes two distinct infall episodes with a delay of 4 $\sim$ 6 Gyr. The first episode happens at the beginning of the model and forms metal-poor and [α/Fe]-enhanced stars in the halo and the thick disk. When the metallicity is enriched to $\sim$ 0.4 dex, the second infall brings in much metal-poor gas and resets the gas chemical composition. Then the chemical enrichment repeats and [α/Fe]-poor stars spreads along the wide metallicity regions. On the contrary, another GCE model by Schönrich & Binney (2009a); Sanders & Binney (2015) considers radial migration or radial mixing. They showed that the extended [Fe/H] distribution in low-[α/Fe] regime is due to radial migration where stars change their angular momentum or guiding radius with time, and one smooth episode of gas infall is sufficient to produce [α/Fe]-bimodality. The effect of radial migration is also seen in simulations (e.g., Minchev et al. 2013; Kubryk et al. 2015). Sharma et al. (2021b) extended the GCE model by applying Shu distribution function (Shu, 1969) as the phase-space distribution and a new prescription for velocity dispersion relation from Sharma et al. (2021a). This extended model successfully reproduced the [α/Fe]-[Fe/H] bimodality distributions (see Fig. 1.9) that are consistent with APOGEE observations (see Fig. 1.7) in Hayden et al. (2015).

In addition to [α/Fe] and [Fe/H], spectroscopic surveys also measured other abundances such as $s-$ and $r-$ process elements, which provides a great opportunity to extend the current GCE models by applying stellar nucleosynthetic yields to reproduce more elements. A recent attempt was made by Chen et al. (2022) using the yields from Kobayashi et al. (2020a,b). It will not only help in understanding the comprehensive chemodynamic evolution of the Milky Way but also provides a way to verify the nucleosynthetic processes and better understand stellar evolution.

1.2.5 Unexplored Regions

The above results indicate great progress in understanding how our Milky Way is formed and evolved for different stellar components, which is attributed to astrometric data from *Gaia* and stellar parameters with chemical abundances from many fiber-fed multi-object spectroscopic surveys. However, the majority of these observations only focus on resolvable stellar regions with a limitation of stellar density. In extremely dense regions, once the spatial distance of adjacent stars is smaller than the fiber size (e.g. 30″ for GALAH; 56 – 71.5″ for APOGEE; 268.2″ for LAMOST), it becomes unsolvable. Such regions include the center of the Galactic bulge and globular clusters which are important for understanding the early formation of the Milky Way. But the current fiber-fed spectrograph can only obtain less dense regions in the bulge and outskirts of clusters.

To address individual stars in dense regions, one approach is to use the integral-field spectrograph like MUSE. As is shown in Fig. 1.1, the integral-field spectrograph data contains a spectrum for every pixel. Therefore, it is feasible to resolve individual stars by fitting stacked PFS of stars with observations. Such a method has been implemented by Kamann et al. (2013) to be a tool named PampelMUSE. This tool has been applied to several studies on the center of globular clusters and more than 6000 spectra can be extracted in a MUSE cube of 47 Tuc (Kamann et al., 2016) with $1' \times 1'$ FoV. Fig. 1.10 provides an example of the resolving power of PampelMUSE on adjacent stars, which verifies that there is a great opportunity to resolve individual stellar spectra in the densest Galactic bulge regions, which we will address in Chapter 3.

To study chemodynamic evolution of dense regions, similar to fiber-fed multi-object spectroscopic surveys, there needs a reliable and efficient tool to measure stellar parameters from MUSE stellar spectra. Unfortunately, previous studies on dense regions mainly measured the kinematics of stars or equivalent width of spectral features (e.g., Göttgens et al. 2019; Kamann et al. 2018; Latour et al. 2019; Husser et al. 2020; Latour et al. 2021), and there has not a model-driven or data-driven model being applied on MUSE spectra to measure elemental abundances yet, which we will address in Chapter 2.

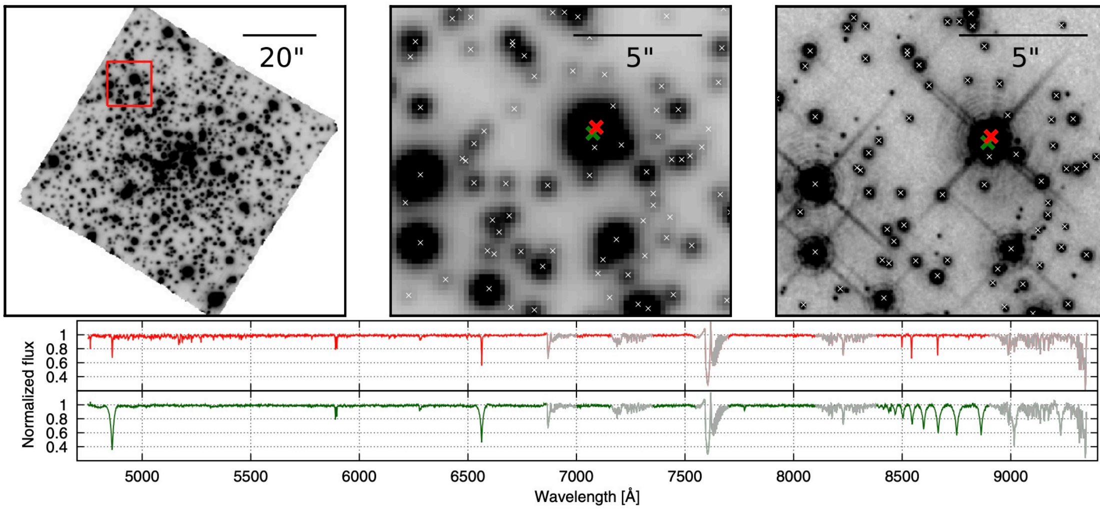

Figure 1.10: Example of the resolving power of MUSE on individual spectra extraction. The top left panel shows a white light image of the central MUSE datacube of NGC 6397. A zoom into the area indicated with a red square is shown in the top middle panel. In the top right panel, the same FoV is shown as seen with HST. All stars for which spectra were extracted are shown with crosses. Spectra extracted from the stars indicated with the red and green crosses are shown in the bottom panel and show clearly distinct features without any contamination from each other. The spectra are continuum subtracted and normalised and the areas dominated by telluric absorption are indicated in grey for clarity. Figure adapted from Husser et al. (2016).

Table 1.3: Summary of IFU surveys on low-redshift galaxies. Ongoing and scheduled surveys are marked with an asterisk.

Survey	Instrument	Location	IFS type	Operation period	Wave range (Å)	Spectral Resolution FWHM(Å)	$\sigma_{\rm inst}$ km s^{-1}	Spatial Sampling (arcsec^2)	No. of galaxies (data release)	redshift	Ref.
SAURON	WHT/SAURON	La Palma	lenslet	1999 – 2003	4800 – 5380	4.8	108	0.8×0.8	72 (E, S0, Sa)	< 0.01	1-3
ATLAS^{3D}	WHT/SAURON	La Palma	lenslet	2007 – 2008	4800 – 5380	4.8	108	0.8×0.8	260 (E, S0)	< 0.01	4
CALIFA	Calar Alto/PMAS/PPAK	CAHA	fiber bundle	2010 – 2016	3745 – 7300 3400 – 4750	6 2.7	150 85	2.7×2.7	667	$0.005 \sim 0.03$	5,6
SAMI	AAT/SAMI	SSO	hexabundles	2013 – 2018	3750 – 5750 6300 – 7400	2.65 1.61	70.4 29.6	1.6×1.6	3068	$0.004 \sim 0.095$	7-10
*HECTOR	AAT/HECTOR	SSO	hexabundles	2023 – present	3727 – 7761	1.3	25	1.6×1.6	15,000	< 0.1	11
MaNGA	SLOAN/BOSS	APO	fiber bundle	2014 – 2020	3630 – 6300 5900 – 10300	$2.86 \sim 3.46$	$50 \sim 80$	2×2	10,010	$0.01 \sim 0.15$	12-15
*MAGPI	VLT/MUSE	Paranal (ESO)	image slicer	2019 – present	4650 – 9300	$2.4 \sim 3$	$35 \sim 80$	0.2×0.2	60	$0.25 \sim 0.35$	16
Fornax	VLT/MUSE	Paranal (ESO)	image slicer	2016 – 2017	4650 – 9300	$2.4 \sim 3$	$35 \sim 80$	0.2×0.2	33	~ 0.0046	17
PHANGS	VLT/MUSE	Paranal (ESO)	image slicer	2017 – 2020	4650 – 9300	$2.4 \sim 3$	$35 \sim 80$	0.2×0.2	19	< 0.005	18
*GECKOS	VLT/MUSE	Paranal (ESO)	image slicer	2022 – present	4650 – 9300	$2.4 \sim 3$	$35 \sim 80$	0.2×0.2	35 (edge-on)	< 0.02	19
TIMER	VLT/MUSE	Paranal (ESO)	image slicer	2016	4650 – 9300	$2.4 \sim 3$	$35 \sim 80$	0.2×0.2	24 (barred)	< 0.01	20
MAD	VLT/MUSE	Paranal (ESO)	image slicer	2015 – 2017	4650 – 9300	$2.4 \sim 3$	$35 \sim 80$	0.2×0.2	38	< 0.013	21

References. (1) Bacon et al. (2001); (2) Emsellem et al. (2004); (3) Kuntschner et al. (2006); (4) Cappellari et al. (2011); (5) Sánchez et al. (2012); (6) Sánchez et al. (2016); (7) Croom et al. (2012); (8) Bryant et al. (2015); (9) van de Sande et al. (2017); (10) Croom et al. (2021); (11) Bryant et al. (2020); (12) Bundy et al. (2015); (13) Yan et al. (2016); (14) Yan et al. (2016); (15) Abdurro'uf et al. (2022);(16) Foster et al. (2021); (17) Sarzi et al. (2022);(18) Emsellem et al. (2022);(19) van de Sande et al. (2023); (20) Gadotti et al. (2019); (21) Erroz-Ferrer et al. (2019)

1.3 Extragalactic Observations

External galaxies are also an essential sample, and being wildly observed to understand galaxy evolution that is different from galaxy to galaxy. Therefore, a large galaxy sample with different types (as shown in the Hubble tuning fork diagram in Fig. 1.11) is ideal for testing galaxy evolution theories. Thanks to the integral-field spectrograph, it is feasible to spatially resolve different components and study specific regions of external galaxies such as the star-forming regions (van Dishoeck & Blake, 1998), galaxy winds (Sharp & Bland-Hawthorn, 2010; Rupke, 2018), circumgalactic medium (CGM, Tumlinson et al. 2017), and velocity distribution (Brough et al., 2011), angular momentum (Emsellem et al., 2011), age, metallicity gradients (Kuntschner et al., 2010), and star formation history (Emsellem et al., 2011). Over the decades, many IFS surveys have been conducted to enlarge the external galaxy sample for a better and more comprehensive study, as is listed in Table 1.3.

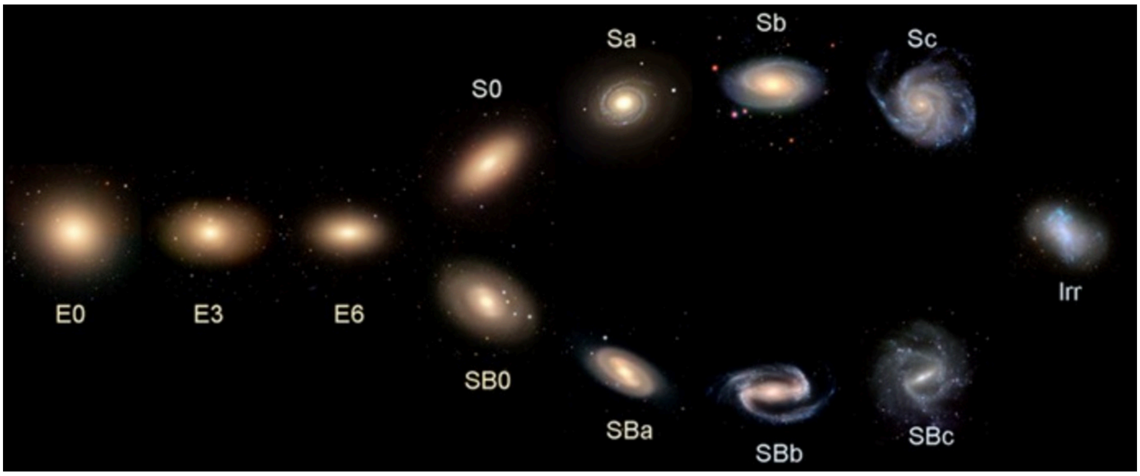

Figure 1.11: Hubble tuning fork used for galaxy classification. Figure adapted from Masters et al. (2019).

1.3.1 Kinematic and Chemical Parameter Measurements

Different from the Milky Way, it is extremely difficult to resolve individual stars in distant external galaxies and measure their kinematic and chemical parameters. The light received from each spaxel is a stack of many stars, which is called integrated spectra. Therefore, an important assumption in extragalactic analysis is that the integrated spectrum is a weighted superposition of spectra of many simple stellar populations (SSPs), which is defined by a coeval stellar population at a single metallicity and abundance pattern (Conroy, 2013). In addition, effects due to stellar velocity (shift, broaden, or skew the spectrum), dust attenuation, and gas emission should also be included. A summary of this whole process is shown in Fig. 1.12.

To measure the kinematics and chemical composition from integrated spectra, an important procedure is to decompose these several complex processes. Many software and algorithms have been developed for this aim since the 2000s such as pPXF (Cappellari & Emsellem, 2004; Cappellari, 2017), STARLIGHT (Cid Fernandes et al., 2005), STECMAP

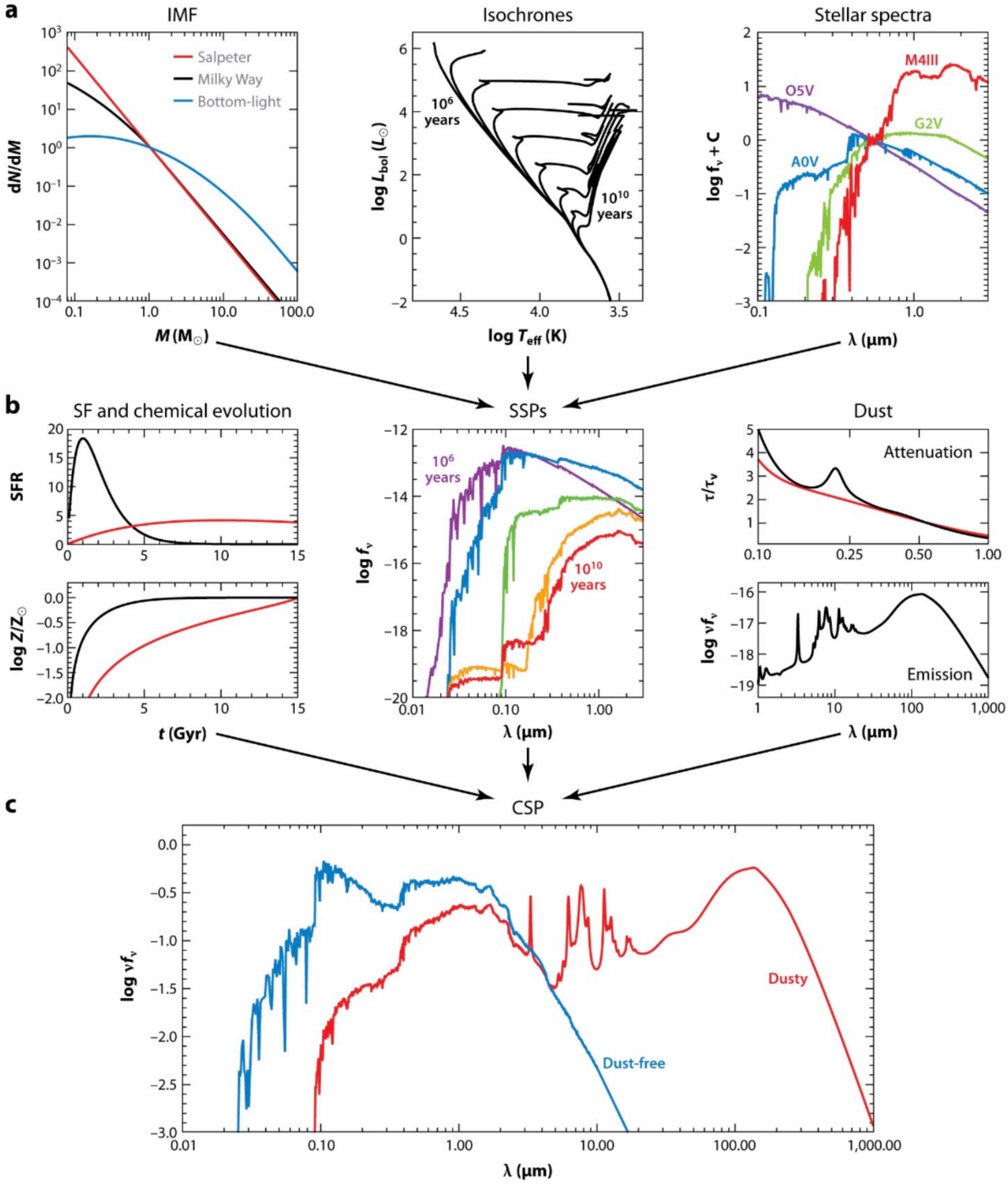

Figure 1.12: Overview of the stellar population synthesis technique. SSPs are generated by using an initial-mass function (IMF), stellar isochrones, and stellar spectra. Some other ingredients for stellar population synthesis include star formation and chemical enrichment, dust attenuation and gas emission. Figure adapted from Conroy (2013).

(Ocvirk et al., 2006), VESPA (Tojeiro et al., 2007), and FIREFLY (Wilkinson et al., 2017). Some semi-analytical models have also been developed in recent years (e.g., Zhou et al. 2022). Here we mainly take pPXF as an example to explain the implementation in detail, but other softwares are more or less based on the same principles. pPXF (Penalized PiXel-Fitting) is used to extract the stellar and gas kinematics, as well as the stellar population of galaxies via full-spectrum fitting. It assumes the observed galaxy spectrum $G_{\rm mod}(x)$ can be synthesized by the equation as follows (adapted from Cappellari 2017):

$$G_{\rm mod}(x) = \sum_{n=1}^{N} w_n \left\{ [T_n(x) * \mathcal{L}_n(cx)] \sum_{k=1}^{K} a_k \mathcal{P}_k(x) \right\} + \sum_{l=0}^{L} b_l \mathcal{P}_l(x) + \sum_{j=1}^{J} c_j S_j(x), \tag{1.1}$$

where x is the wavelength pixel, $\mathcal{L}_n$ is the line-of-sight velocity distribution (LOSVD) for the n-th stellar population and gas spectral template T_n, $\mathcal{P}_l$ and $\mathcal{P}_k$ are multiplicative and additive polynomials with degree k and l to minimize template mismatch, imperfect sky subtraction, or inaccuracies in the spectral calibration. In addition, $\mathcal{P}_k$ can be replaced by a reddening curve (e.g., Calzetti et al. 2000) to measure extinction value $E(B-V)$. S_j are spectra of the sky. During the fitting process, pPXF will find the best solution that matches the observed spectrum by χ^2 minimization and return the LOSVD ($\mathcal{L}_n$) and coefficients of each term (w_n, a_k, b_l, and c_j).

The kinematics of the observed spectrum is represented by $\mathcal{L}_n$, which is normally parameterized using the Gauss-Hermite function introduced by van der Marel & Franx (1993); Gerhard (1993). The Gauss-Hermite function has the form as follows:

$$\begin{aligned} \mathcal{L}(y) &= \frac{\exp\left(-y^2/2\right)}{\sigma\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y)\right], \\ y &= (v - V)/\sigma, \end{aligned} \tag{1.2}$$

where H_m are the Hermite polynomials. The first four moments of this equation are stellar mean line-of-sight velocity $V_{\rm LOS}$, velocity dispersion σ, skewness h_3, and kurtosis h_4. These four moments are broadly used to study the kinematic structures of different galaxy components. The including of h_3 and h_4 is due to the natural build-up of stellar particles $V_{\rm LOS}$ along the edge-on disk (see Fig. 1 of Kregel & van der Kruit 2005). A key feature of pPXF is that it automatically penalizes non-Gaussian solutions to prevent over-fitting and reduce the noise of recovered kinematics. This is achieved by adding a term in the χ^2 equation to amplify the impact of non-Gaussian features with large h_3 or h_4, which are normally found to be nonphysical (Cappellari & Emsellem, 2004). Another useful method to measure kinematics is BAYES-LOSVD (Falcón-Barroso & Martig, 2021), which is a non-parametric framework and can overcome the limitations of Gauss-Hermite function on some peculiar LOSVD shapes.

Among the coefficients obtained by pPXF, w_n represents the weight of each simple stellar population and gas templates. SSP templates spread in a certain metallicity, $[\alpha/\rm Fe]$ and age grid. Therefore, these weights can be used to calculate the mass- or light-weighted average of [M/H], [Fe/H] and age for this observed spectrum, depending on whether the templates are normalized to one $M_\odot$ or $L_\odot$. One can also obtain star formation history (age distribution) and metallicity distribution of this observed spectrum by integrating the weights along the age or metallicity grid.

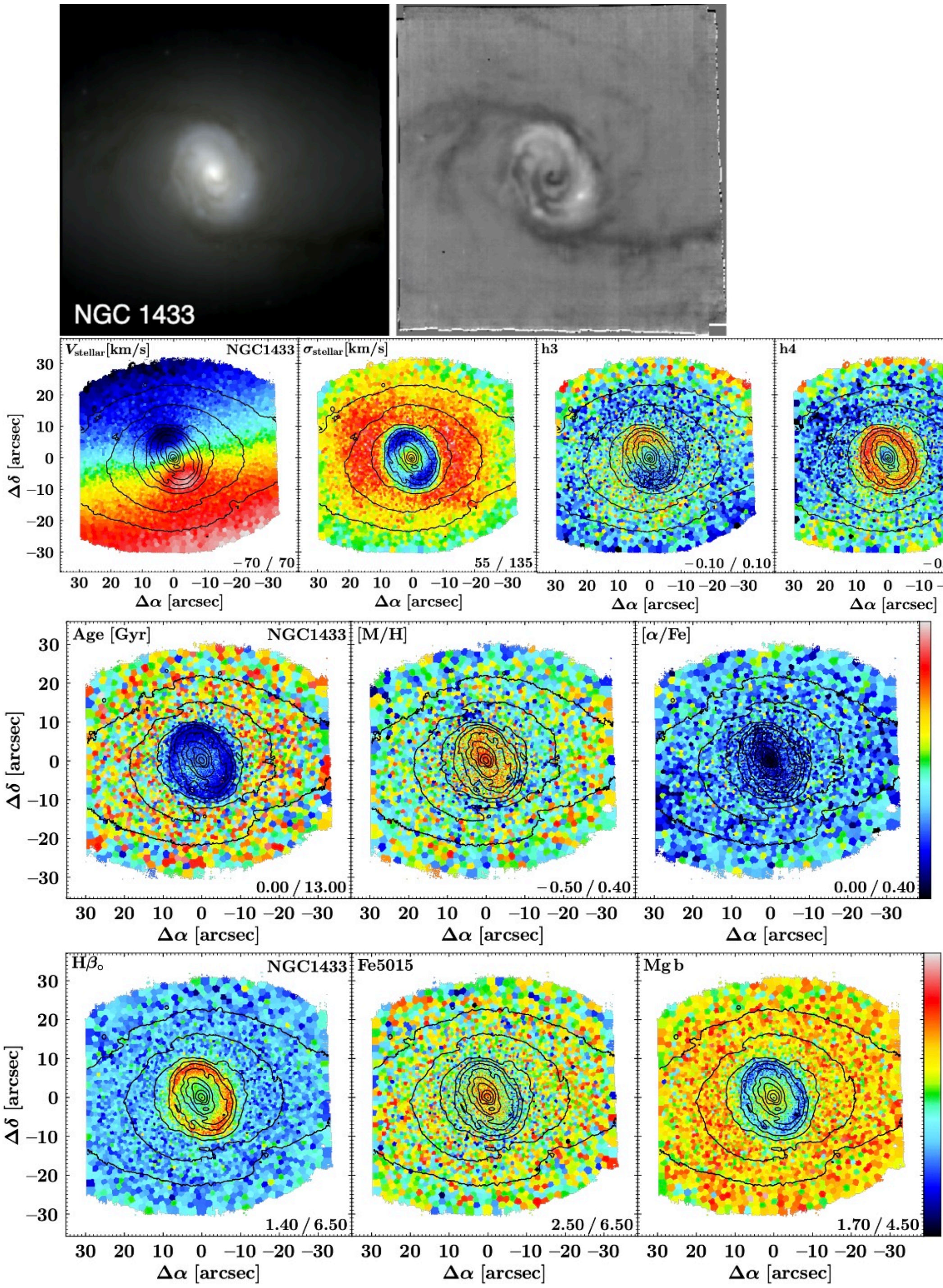

Figure 1.13: Kinematic and chemical composition maps of an example galaxy NGC 1433. Figure adapted from Bittner et al. (2019).

Another complementary method to obtain chemical composition is to measure the line strength of each absorption feature, which is due to the electron transition between different energy levels. The line strength of each absorption feature is calculated by the difference between the total flux of this bandpass and the continuum in the same wavelength region (Faber et al., 1985; Kuntschner et al., 2006). Because of the relation between line strength and [M/H], [α/Fe], and age, it is feasible to measure the SSP-equivalent chemical composition properties by applying Markov chain Monte Carlo (MCMC) algorithm (Foreman-Mackey et al., 2013). This method is also broadly used to infer the stellar information of galaxies in IFS surveys (e.g., Kuntschner et al. 2010; McDermid et al. 2015).

Finally, after applying all the procedures above, one will obtain a comprehensive parameter maps of external galaxies including kinematics, mass- or light-weighted [M/H], [α/Fe], and age, line-strength indices with the inferred SSP-equivalent properties, gas properties, star-forming regions, etc. An example of these property maps is demonstrated in Fig. 1.13.

1.3.2 Explorations on Milky Way Analogues

The progress of our understanding of the Milky Way evolution in Section 1.2 revealed a complex mixture of internal and external processes. In the Universe, there are many disk galaxies that have similar structures to the Milky Way such as the thick disk (Burstein, 1979; Yoachim & Dalcanton, 2006; Comerón et al., 2018; Martínez-Lombilla & Knapen, 2019), b/p bulge (e.g., Georgiev et al. 2019; Kormendy & Bender 2019), stellar radial migration and accretion history (e.g., Zhu et al. 2022). They are thought to have gone through similar formation and evolution history, although some specific processes might be different. Therefore, the GCE models of the Milky Way might also be applied to fit with external galaxies by tuning some parameters. A promising way to better understand galaxy evolution is by combining the Milky Way with other disk galaxies, which is permitted by large samples of extragalactic IFS surveys.

In recent years, several studies attempted to compare some aspects of the Milky Way with the so-called Milky Way analogues (MWAs) both from extragalactic observations and cosmological simulations that have similar integrated properties such as star formation rate, bulge-to-total ratio, etc. Fraser-McKelvie et al. (2019) selected 176 galaxies from the whole SDSS DR7 sample (Abazajian et al., 2009) and compared the mean star formation rate with the Milky Way, they found that even though the Milky Way has a relatively lower SFR, it is not unusual when compared to similar galaxies within 2σ. Kormendy & Bender (2019) found the pseudo-bulge of two galaxies NGC 4565 and NGC 5746 are made of old and [α/Fe]-enhanced stars, similar to the Milky Way. Evans et al. (2020) selected galaxies with the Milky-Way-mass halos from EAGLE simulations and found that $\sim$ 16 percent have an LMC-like accretion, $\sim$ 5 percent have a GES-like event and no further massive merger since $z = 1$, and only $\sim$ 0.65 percent have both LMC and GES. Boardman et al. (2020a) selected 62 Milky Way analogues (MWAs) from MaNGA with the criteria of stellar masses and bulge-to-total ratios and found most of them have flatter stellar and gas-phase metallicity gradients due to different disc scale lengths, and a greater consistency can be found when scaling gradients by these lengths; Mao et al. (2021) spectroscopically observed 127 satellite galaxies around 36 Milky Way analogues and found the number and luminosities of MW satellites are consistent with the distributions from these analogues. Zhou et al. (2023) revisited 138 MaNGA galaxies by fitting the spectra with a semi-analytic chemical evolution model (Zhou et al., 2022) and measured their star formation and chemical enrichment histories. They detected similar [α/Fe] bimodality as the Galactic APOGEE observations (Abdurro'uf et al., 2022) in many of the galaxies.

With the unprecedented observational depth and spatial resolving power of the MUSE instrument. Many MUSE IFS surveys have been proposed to study thin disk structures and interstellar medium on face-on galaxies in the scale of 50 $\sim$ 100 pc, as shown in Table 1.3. Several individual studies on edge-on galaxies have also been carried out and the [α/Fe]-enhanced thick disk and merger signatures were identified (Pinna et al., 2019a,b; Scott et al., 2021; Martig et al., 2021). Particularly, Scott et al. (2021) demonstrated that the MW analogue UGC 10738 has similar metallicity, [α/Fe], and age distributions at different projected R and $|z|$ with to Milky Way, as shown in Fig. 1.14. In addition, [α/Fe]-bimodality distributions at different projected R and $|z|$ are also identified, which supports the commonness of MW's chemical evolution history. In the future, the GECKOS survey (van de Sande et al., 2023) with deep observations on 35 edge-on galaxies provides great opportunities to extend our Galactic chemodynamic evolution theories to more samples and identify more internal and external processes of external galaxies than the current stage.

1.3.3 Challenges of Comparing to the Milky Way

Although all the results above demonstrated that many aspects can be studied by combining the Milky Way with external disk galaxies, there are still challenges to quantitatively making the comparison due to differences in observational strategies, parameter definitions and measuring methods, and the assumptions that are made during the studies. We list all the challenges below:

- The Galactic observations can resolve individual stars, but extragalactic observations provide unresolved integrated spectra.
- The Galactic individual stellar spectra can be used to measure $T_{\rm eff}$, $\log g$, metallicity and more than 20 elemental abundances including [α/Fe], but extragalactic observations perform full-spectrum fitting to obtain averaged age, metallicity, and [α/Fe] which is either mass- or light-weighted.
- Even though the stellar number density distribution of the Milky Way can be transferred to other definitions by correcting the survey or pipeline selections, corrections for the extragalactic projection effect are challenging and rarely considered in previous studies.
- The 6D phase space parameters can be obtained by astrometric satellites like *Gaia*, but extragalactic observations can only provide LOSVD with projected radius and height.
- The Galactic 3D dust distribution can be addressed by measuring $E(B-V)$ values of individual stars with different distances, but it is still unknown how the dust is distributed in external galaxies, and how it affects the kinematic and chemical composition results from full-spectrum fitting.

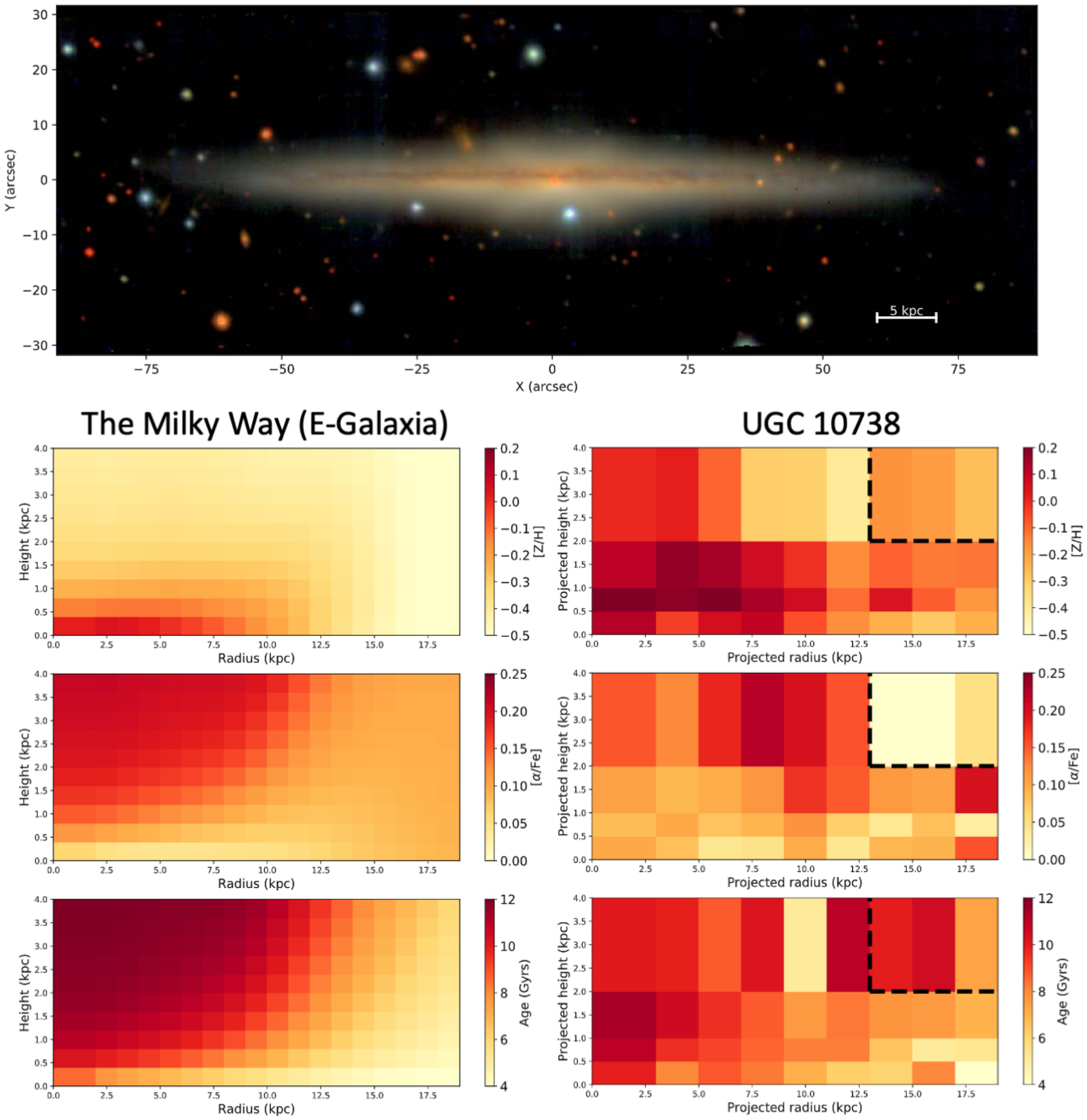

Figure 1.14: Metallicity, $[\alpha/\text{Fe}]$, and age distributions at different radii and heights of the Milky Way and UGC 10738. The distributions of the Milky Way are obtained by using a mock stellar catalog from E-GALAXIA implemented by the chemical evolution model of Sharma et al. (2021b). This mock catalog is chemodynamically consistent with the APOGEE observation on the Milky Way (Hayden et al., 2015), as shown in Fig. 1.7 and Fig. 1.9. Figure adapted from Scott et al. (2021).

- Even without extinction, the limitation of full-spectrum fitting techniques is still unknown for extragalactic studies. Several studies have shown this method can obtain "ill-posed" and nonphysical results, even though a lot of efforts have been made such as penalization on high-order kinematics moments (Cappellari & Emsellem, 2004) and regularization for a smooth star formation history (Emsellem et al., 2011).
- Every galaxy is different. Therefore, different selection criteria can severely affect the number of Milky Way analogues and the conclusions addressed can also be opposite (Boardman et al., 2020b).
- The integrated properties of the Milky Way (e.g., Licquia & Newman 2015; Licquia et al. 2015) are still being updated by the newest and larger stellar samples. And the updated values can be very different from the previous ones (e.g., bulge-to-total ratio is estimated to be $0.150^{+0.028}_{-0.019}$ in Licquia & Newman 2015, but 0.3 ± 0.06 in Bland-Hawthorn & Gerhard 2016), which brings difficulties in selecting Milky Way analogues.

To overcome the above challenges, one potential way is to generate a benchmark of the Milky Way which is easy to be compared to external galaxies directly without any definition and methodology difference. The best way to achieve it is to generate mock integral-field spectroscopic observations of the Milky Way by applying all the knowledge we know on the Galactic evolution. It not only allows us to directly compare the Galactic properties with external galaxies but also permits the feasibility of investigating the limitations of current full-spectrum fitting methods because we know the correct parameter distributions of the mock Milky Way. In addition, we can also address the effect of dust on these measurements for the first time. This comes to the work of this thesis in Chapter 4-6.

1.4 Thesis Goals

The advent of the MUSE instrument and high-spatial-resolution extragalactic observation on nearby galaxies presents a great opportunity to study Galactic archaeology in the MW's dense regions and also helps link the Galactic and extragalactic studies for the understanding of galaxy evolution.

In this thesis, we study on both sides. We create a novel method to measure stellar parameters and chemical abundances for individual stars located in dense regions. Then we apply this method to MUSE observation on the Galactic bulge. We also develop a tool to generate mock Milky Way MUSE observations, which can be used as a bridge to directly compare with external galaxies in kinematics and chemical compositions.

We attempt to further explore the chemical evolution of the Milky Way and extend our theories to other disk galaxies. This is not merely to answer the question if the Milky Way is unique in the Universe but also to investigate if other galaxies have undergone comparable evolutionary processes that could not be addressed in the past.

In Chapter 2, we develop a novel method by combining PampelMUSE with *DD-Payne*, which is a neural network model used to measure stellar parameters from LAMOST spectra, to extract individual stellar spectra in high stellar density regions observed by MUSE and measure $T_{\rm eff}$, $\log g$ with chemical abundances of [Fe/H], [Mg/Fe], [Si/Fe], [Ti/Fe], [C/Fe], [Ni/Fe] and [Cr/Fe].

In Chapter 3, we propose a deep MUSE observation towards the Galactic bulge on low extinction Baade's window with $A_V \sim 1.45$ and obtain stellar spectra of more than 100 MSTO stars with $\mathrm{SNR_{MUSE}} > 80\ \mathrm{pix}^{-1}$. We apply the method in Chapter 2 to measure stellar parameters and combine stellar isochrones to measure the ages of these MSTO stars, which is used to study the age distribution and chemical evolution history of the Galactic center.

In Chapter 4, we introduce GalCraft, a tool to generate mock MUSE observations of the Milky Way using simple stellar population spectral libraries and the mock stellar catalog observationally consistent to the Milky Way. This tool is flexible with a variety of parameters adjustable and the kinematic and chemical composition parameters can be measured from the mock observation in the same way as stated in Section 1.3.1.

In Chapter 5, we test the recovery ability of commonly used software pPXF on kinematic and chemical composition measurements by applying it to GalCraft generated mock MUSE observations and comparing with the true values.

In Chapter 6, we investigate the impact of extinction on the kinematic and chemical composition measurements by adding different dust distributions to GalCraft generated mock MUSE observations.

Finally, we conclude all our results in Chapter 7 and provide prospects that should be done in this topic to achieve our end goal of linking the Galactic and extragalactic studies to understand galaxy evolution.

Chapter 2

Reliable Stellar Abundances of Individual Stars from MUSE Observations

Zixian Wang,[1,2] Michael R. Hayden,[1,2] Sanjib Sharma,[1,2] Maosheng Xiang,[3] Yuan-Sen Ting,[4,5,6,7,8] Joss Bland-Hawthorn,[1,2,9] Boquan Chen[1,2]

[1]Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW, 2006, Australia
[2]ARC Centre of Excellence for All Sky Astrophysics in Three Dimensions (ASTRO-3D)
[3]Max-Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany
[4]Research School of Computer Science, Australian National University, Acton ACT 2601, Australia
[5]Research School of Astronomy and Astrophysics, Australian National University, ACT 2611, Australia
[6]Institute for Advanced Study, Princeton, NJ 08540, USA
[7]Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA
[8]Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA
[9]Miller Professor, Miller Institute, UC Berkeley, Berkeley CA 94720

This chapter is reproduced from Wang et al. (2022), which was published in Monthly Notices of the Royal Astronomical Scocity as "Reliable stellar abundances of individual stars with the MUSE integral-field spectrograph". I wrote all the text, produced all the figures except Fig. 2.11, and wrote the codes for data analysis and interpretation. Sanjib Sharma provided the bottom panel of Figure 2.11 and helped run BSTEP to measure stellar distance and extinction. Joss Bland-Hawthorn, Sanjib Sharma and Michael Hayden supervised the whole work and helped revise and polish the draft. Yuan-Sen Ting and Maosheng Xiang provided comments and the DD-Payne models and codes for estimating the correlation of gradient spectra from DD-Payne and synthetic spectral models.

We present a novel approach to deriving stellar labels for stars observed in MUSE fields making use of data-driven machine learning methods. Taking advantage of the comparable spectral properties (resolution, wavelength coverage) of the LAMOST and MUSE instruments, we adopt the Data-Driven Payne (*DD-Payne*) model used on LAMOST observations and apply it to stars observed in MUSE fields. Remarkably, in spite of instrumental differences, according to the cross-validation of 27 LAMOST-MUSE common stars, we are able to determine stellar labels with precision better than 75K in $T_{\rm eff}$, 0.15 dex in $\log g$, and 0.1 dex in abundances of [Fe/H], [Mg/Fe], [Si/Fe], [Ti/Fe], [C/Fe], [Ni/Fe] and [Cr/Fe] for current MUSE observations over a parameter range of $3800 < T_{\rm eff} < 7000$ K, $-1.5 <$ [Fe/H] < 0.5 dex. To date, MUSE has been used to target 13,000 fields across the southern sky since it was first commissioned six years ago and it is unique in its ability to

study dense star fields such as globular clusters or the Milky Way bulge. Our method will enable the automated determination of stellar parameters for all stars in these fields. Additionally, it opens the door for applications to data collected by other spectrographs having resolution similar to LAMOST. With the upcoming BlueMUSE and MAVIS, we will gain access to a whole new range of chemical abundances with higher precision, especially critical s-process elements such as [Y/Fe] and [Ba/Fe] that provide key age diagnostics for stellar targets.

2.1 Introduction

The formation and evolution of the Milky Way is one of the outstanding questions facing astrophysics today. The study of stars and stellar populations is a pillar of Galactic Archaeology (Freeman & Bland-Hawthorn, 2002), as they contain the chemical imprint of the gas from which they formed and serve as birth tags. This allows stars to be used as a fossil record to unravel the history of the Milky Way. During the past decade, traditional fiber-fed spectrographs employed by most large-scale spectroscopic surveys like RAVE (Steinmetz et al., 2006), LAMOST (Deng et al., 2012; Zhao et al., 2012), *Gaia*-ESO (Gilmore et al., 2012), GALAH (De Silva et al., 2015), and APOGEE (Majewski et al., 2017) combined with astrometric and photometric information from *Gaia* (Gaia Collaboration et al., 2018, 2021) provided detailed chemodynamics for millions of stars from the solar neighborhood to even several kilo-parsecs away (Bland-Hawthorn & Gerhard, 2016).

The large volume of spectroscopic observations is posing great challenges for data analysis, particularly in deriving stellar labels (atmospheric parameters and chemical abundances) precisely and efficiently from spectra. In the past few years, with the use of radiation transfer codes (e.g. SME Valenti & Piskunov 1996, Turbospectrum Alvarez & Plez 1998; Plez 2012, MOOG Sneden et al. 2012), many model-driven (e.g. *The Payne* Ting et al. 2019 and *StarNet* Fabbro et al. 2018; Bialek et al. 2020) and data-driven methods (e.g. *The Cannon* Ness et al. 2015 and *Astro-NN* Leung & Bovy 2019) and models in between (e.g. *DD-Payne* Xiang et al. 2019 and *Cycle-StarNet* O'Briain et al. 2021) have been developed for this purpose. One benefit of the data-driven approach is it can generate a model from a training set sample, rather than traditional physics-based approaches using theoretical stellar atmospheres; this is much easier to do and in general leads to more precise stellar parameters as it can potentially make use of the full wavelength region of the spectra.

However, one limitation of the data-driven methods is that some chemical abundances could be determined by astrophysical correlations in the training set, instead of physically motivated measurements from spectral features. This has a non-negligible impact on the precision of some stellar labels, especially those derived from low-resolution spectra, where many of the absorption features are blended.

To overcome this issue, Ting et al. (2017b, 2018) used the methods outlined in *The Payne* (Ting et al., 2019) to build a neural network to determine stellar labels. They imported a new term in the loss function which takes into account the similarity of gradient spectra from the data-driven model and Kurucz model atmospheres (Kurucz, 1970, 1993, 2005), to regularize the training process and force the model to measure labels from real absorption features. They verified that more than 10 elements can be extracted even from the low-resolution spectra as in LAMOST. Based on this approach (Ting et al., 2017b), Xiang et al. 2019 further developed the Data-Driven Payne (hereafter *DD-Payne*) which used input training labels from the high-resolution spectroscopic surveys APOGEE (Ting et al., 2019) and GALAH (Buder et al., 2018) to determine stellar labels for LAMOST DR5 spectra. These authors found that they could precisely estimate close to 20 stellar labels for more than 6 million spectra. These studies highlight that if one has a reliable training set then one can obtain precise elemental abundances even from low-resolution spectra.

The methods outlined above demonstrate how one can measure stellar labels for millions of spectra using machine learning approaches. However, all of the surveys described previously are conducted using fiber-fed spectrographs, which take spectra by plugging a series of optical fibers onto a plate based on the positions of stars on the sky. Fibers are typically set to have a minimum separation (e.g. $30''$ for GALAH; $56 - 71.5''$ for APOGEE; $268.2''$ for LAMOST) during each observation. And each fiber has a typical diameter of $2 \sim 3''$. This means that these instruments can not observe dense fields such as cores of globular clusters and the very dense regions of Galactic bulge, where the separation of stars is less than $1''$. However, stars in these fields also play an essential role in the Galactic evolution, it is vital to obtain more complete spectroscopic observations in these regions to unravel the history of the Milky Way.

The Multi-Unit Spectroscopic Explorer (MUSE) (Bacon et al., 2010) on the Very Large Telescope is an integral field spectrograph, which is regularly used to observe external galaxies and can be an ideal solution to deal with high stellar density fields. With the advanced development of image slicer and spectrometer, MUSE can perform both imaging and spectroscopy simultaneously. The benefit is that each MUSE spectrum is obtained for a spatial pixel ($0.2'' \times 0.2''$) on the sky rather than the area of the fiber, so it is not restricted by the fiber size or collision effects. The data format of MUSE is a 3-D datacube, with 2D in spatial on the sky and 1D in wavelength. Combining with the PSF-fitting spectral extraction software PampelMUSE (Kamann et al., 2013), MUSE becomes the best instrument to obtain spectra in fields with high stellar density.

There have been several studies on stars in globular clusters using spectra from MUSE observations (Husser et al., 2016; Kamann et al., 2016). Near the core of these clusters, more than 6000 spectra (e.g., 47 Tuc at a distance of 4 kpc analyzed by Kamann et al. 2016) can be extracted on average per MUSE cube (in wide-field mode with a field of view of $1' \times 1'$). Currently, most studies focused exclusively on kinematics (e.g. Kamann et al. 2016, 2018; Valenti et al. 2018). Some studies explored specific spectral region or metallicity (e.g. Göttgens et al. 2019; Husser et al. 2020). However, these spectra contain a wealth of chemical information that has not yet been fully explored. This is because currently there is no robust and easy-to-use method to automatically derive chemical abundances for stars in these fields.

The success of data-driven models in deriving about 20 abundances from LAMOST spectra highlights the potential of other low-resolution surveys, for example MUSE, having wavelength coverage and resolution similar to LAMOST to also derive abundances. Inspired by this, in this paper, we develop an approach based on the *DD-Payne* (Ting et al., 2017b; Xiang et al., 2019), to measure stellar labels from MUSE observations. Using this approach we derive robust stellar labels, including effective temperature ($T_{\rm eff}$), surface gravity ($\log g$), metallicity [Fe/H] and abundances of four α elements, [Mg/Fe], [Si/Fe], [Ca/Fe] and [Ti/Fe].

In Section 2.2, we introduce the data-sets used in our analysis and the spectral extraction method used by us. In Section 2.3, we outline the procedures used to measure stellar labels for MUSE spectra. We do two validation tests in Section 2.4. First, we verify the bias and dispersion of our approach by applying it to common stars between LAMOST and MUSE. Second, for dense fields, we explore the [Fe/H]-[Mg/Fe] distribution of stars towards the bulge region using MUSE observations and compare it with that obtained from the high-resolution APOGEE survey. In Section 2.5, we explore and discuss the dependence of label precision on magnitude and exposure time. In Section 2.6, we discuss the exciting prospects of using MUSE to observe dense stellar fields in the Milky Way and other galaxies. Finally, a summary of our results and conclusions is presented in Section 2.7.

2.2 Data

2.2.1 MUSE

Overview

MUSE (the Multi-Unit Spectroscopic Explorer) (Bacon et al., 2010, 2014) is an optical wide-field integral field spectrograph using the image slicing technique mounted on the UT4 of the Very Large Telescope at the Paranal Observatory in Chile. MUSE operates in two spatial modes, the wide-field mode (WFM) and the narrow-field mode (NFM). The wide-field mode covers a FoV of $1' \times 1'$ with $0.2'' \times 0.2''$ spatial bins. The narrow-field mode covers a FoV of $7.5'' \times 7.5''$ with $0.025'' \times 0.025''$ spatial bins. It operates in two wavelength modes, the nominal mode covering 4800-9300 Å and the extended mode covering 4650-9300 Å. It samples the wavelength in 1.25Å bins at a spectral resolution of R$\sim$3000. To date, more than 13000 datacubes have been observed by MUSE and they are publicly available via the ESO Science Portal Website[1]. MUSE is typically used to conduct studies on external galaxies but it has also been used to study a wide variety of other things, e.g., exoplanets (Irwin et al., 2018; Haffert et al., 2019), Galactic regions (Weilbacher et al., 2015; McLeod et al., 2015) and resolved stellar populations of Globular clusters (Husser et al., 2016; Kamann et al., 2018).

Reductions

The data format of MUSE observations follows the ESO science data product standard for datacubes. Each datacube has three dimensions, two in spatial and one in spectral. The data processing pipeline transforms the raw CCD-based data into fully calibrated and corrected datacubes, which can be directly used for studies. The transformation processes include bias, dark and flat-field reduction, wavelength and flux calibration, line spread function and illumination correction, sky creation, and astrometrically correlation and multiple exposures combination, which are discussed in detail in Weilbacher et al. 2012,

[1] http://archive.eso.org/scienceportal/home

2014, 2020. For this study, all the datacubes we use are published online and have been processed and calibrated using the MUSE Data Processing Pipeline[2].

Stellar Spectral Extraction

To extract stellar spectra, we use the PampelMUSE package (Kamann et al., 2013) which was specifically developed for 3D datacubes in dense fields. It assumes the datacube to be a sum of many overlapping PSFs and sky background. First, a single object is represented by a PSF function such as Moffat and Gaussian, and assuming parameters in the PSF function change smoothly with wavelengths. Next, the PSF photometry is employed to fit each star in each wavelength layer and then all the layers are combined to generate a single stellar spectrum. One great advantage of this method is that it can de-blend the spectra of different stars. Therefore, MUSE observations combined with PampelMUSE become an ideal way to study dense fields.

To get spectra in the field using PampelMUSE, a reference catalog with stellar positions and magnitude is needed as input. Various reference catalogs can be used, such as all-sky surveys (e.g. *Gaia* Gaia Collaboration et al. 2018, 2021, Sky-Mapper Keller et al. 2007 or HST) or smaller surveys for specific purposes. In this study, we mainly use LAMOST DR5 and VVV DR2 (Minniti et al., 2017) as reference catalogs for the two validation test data, respectively. Moreover, a transformation function of the filter band is needed when aligning the MUSE datacube slice with reference catalogs. Here we refer to the transformation function from SVO Filter Profile Service[3] and use the Moffat function as the PSF profile, which has better performance than the Gaussian profile when fitting flux distribution of stars (see Figure 3 in Kamann et al. 2013).

After going through all processes in PampelMUSE, including source position transformation, PSF fitting, sky background estimation and subtraction, and unresolved star reduction, the output is a list of stellar spectra which are identified as resolvable by the algorithm. A few additional cleaning steps (removal of emission lines and non-stellar features) are performed on the spectra and will be discussed in Section 2.4.2.

2.2.2 LAMOST

Overview

The LAMOST Galactic survey (Deng et al., 2012; Zhao et al., 2012; Liu et al., 2014, 2015) is the first dedicated spectroscopic survey to obtain spectra of millions of stars covering most of the sky. LAMOST DR5 has released more than 8 million stellar spectra covering the optical wavelength regime 3700-9000 Å with low-resolution R$\sim$1800. For about 5 million spectra, LAMOST DR5 also provides the basic stellar labels such as $T_{\rm eff}$, $\log g$ and [Fe/H] which were derived with the LAMOST stellar parameter pipeline (LASP; Wu et al. 2011; Luo et al. 2015). There are also several value-added catalogs providing stellar labels with some chemical abundances and extinction via different methods (e.g., Xiang et al. 2015; Li et al. 2016; Xiang et al. 2017). Despite LAMOST's low spectral resolution,

[2] https://www.eso.org/sci/software/pipelines/muse/
[3] https://svo.cab.inta-csic.es/main/index.php

Table 2.1: Comparison of MUSE and LAMOST survey.

Survey	Wavelength range (Å)	Resolution
MUSE	4750-9300	R~3000
LAMOST	3700-9000	R~1800

Ting et al. 2017b, 2018 has verified that ≥10 chemical abundances should be derivable with precision on the level of 0.1 ~ 0.2 dex or better.

Data-Driven Payne

Given the success of machine learning and data-driven models in stellar parameter determination, an effective way to estimate stellar labels from MUSE spectra would be to develop a data-driven model for the MUSE spectra based on stellar labels from other high-resolution spectroscopic surveys. Presently, we do not have enough common stars (ideally $\sim 10^4$) covering a wide range in the parameter space between MUSE and other surveys for a robust training set. However, with LAMOST spectra having similar resolution and wavelength range as MUSE (see Table 2.1), we can apply a model developed for LAMOST to MUSE spectra. Ting et al. 2017b; Xiang et al. 2019 (hereafter T17 & X19) had developed a *DD-Payne* model for LAMOST, which we discuss below. It was shown to have good performance and hence we adopted it for our analysis.

DD-Payne utilizes the neural network interpolator and the fitting technique from *The Payne* (Ting et al., 2019) combined with physical gradient spectra from Kurucz spectral model (Kurucz, 1970, 1993, 2005) to regularize the training process. This model can derive stellar labels of $T_{\rm eff}$, $\log g$, micro-turbulence velocity ($V_{\rm mic}$) and 16 reliable chemical abundances. Compared with other similar models (Ho et al. 2017; Wheeler et al. 2020), the biggest advantage of *DD-Payne* is it can enforce these stellar labels to be measured physically from the spectral features. For $\mathrm{SNR}^{\mathrm{g-band}}_{\mathrm{LAMOST}} > 50\ \mathrm{pix}^{-1}$, the typical theoretical precision of the *DD-Payne* abundances is ~0.05 dex for Fe, Mg, Ca, Ti, Cr and Ni, ~0.1 dex for C, N, O, Na, Al, Si, Mn, and Co. This study demonstrates that while obtaining reliable elemental abundances remains challenging for low-resolution spectra, precise abundances are still derivable for spectra with SNR~50 pix^{-1}.

2.2.3 Validation data and the motivations

We do two different validation tests. Since the model used by us is designed for LAMOST spectra, the question is how precisely does it work when applied to MUSE spectra? Therefore, for the first validation test, we run the model on both the LAMOST spectra and the MUSE spectra of the same star and compare the results. We cross-matched the LAMOST DR5 catalog from X19[4] with all the MUSE observations so far published on the ESO website, and found 79 stars across 162 MUSE datacubes.

The second validation is done by comparing the abundance distribution of stars in the inner Galaxy (bulge region) with those obtained from the high-resolution APOGEE survey. The Galactic bulge region has a very high stellar density and its stars suffer from

[4] http://dr5.lamost.org/doc/vac

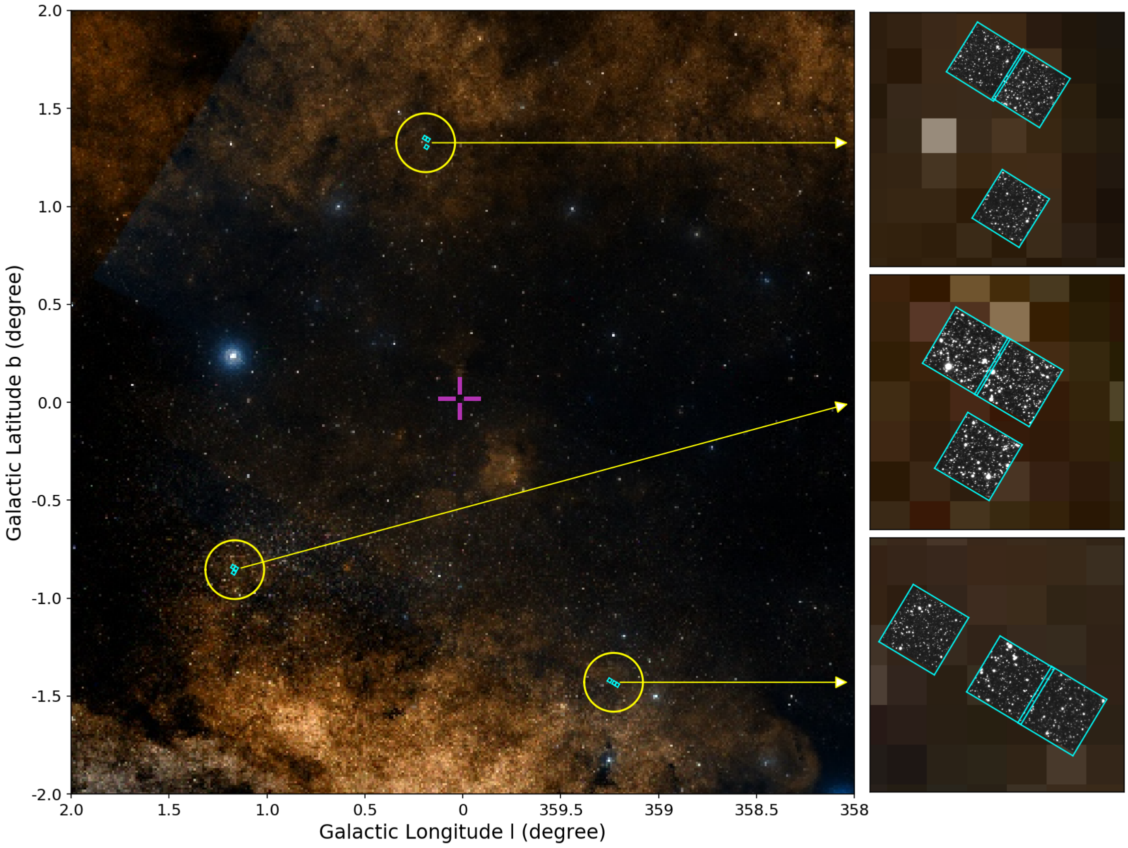

Figure 2.1: Distribution of 9 fields observed by MUSE (0101.B-0381(A), PI: Zoccali) in the Galactic coordinate with the background taken by DSS2. The nine square fields are marked in cyan and are divided into three areas. The purple cross in the middle represents the Galactic center where $(l, b) = (0, 0)$. On the right-hand side are zoomed-in figures of these three areas guided by yellow arrows. The white light picture of the MUSE datacube is shown in each square field.

strong extinction. Hence, validation of bulge stars is important to test the performance of our method in challenging conditions. Therefore, for the second test, we selected 29 datacubes (0101.B-0381(A), PI: Zoccali) covering nine fields that are less than 3 degrees to the Galactic center. They are dense fields and provide a large sample of stars for which to verify our spectroscopic abundances. Figure 2.1 shows the distribution of 9 fields in the Galactic coordinate with the background taken by DSS2. The nine square fields are marked in cyan and are divided into three areas. The purple cross in the middle represents the Galactic center where $(l, b) = (0, 0)$. On the right-hand side are zoom-in figures of these three areas guided by yellow arrows. The white light picture of the MUSE datacube is shown in each square field. For each cube field, there are three or four repetitive exposures. Galactic bulge fields present some unique challenges and need some additional reprocessing, such as low SNR wavelength cut-off, emission line masking, etc, which will be discussed in detail in Section 2.4.2.

2.3 Methods

The feasibility of applying the *DD-Payne* on MUSE observations is due to the similar wavelength range and spectral resolution of the MUSE and LAMOST spectrographs. The processes of deriving stellar labels from MUSE observations include five steps:

- Stellar spectra are extracted from MUSE datacubes by running the automatic spectra extraction algorithm PampelMUSE (Kamann et al., 2013) to obtain spectra at R~3000. This step was talked about in Section 2.2.1.
- The R~3000 MUSE spectra are degraded by using the line spread function (LSF) of these two instruments to R~1800.
- The spectra are normalized by a pseudo-continuum derived in the same way as in T17 & X19, to make it compatible with the LAMOST *DD-Payne* model.
- The degraded and normalized spectra are then fitted by three *DD-Payne* models of X19 to derive stellar labels.
- For each label, the cross-correlation of gradient spectra between *DD-Payne* and Kurucz spectral model (Kurucz, 1970, 1993, 2005) are calculated to check if the label is estimated from real physical features rather than astrophysical correlations.
- All the stellar information is generated and combined to create an output catalog.

The detailed process of each step and the calculating procedure is described in the following sub-sections.

2.3.1 Degrading and normalizing the MUSE spectra

To make *DD-Payne* executable on MUSE spectra, we need to make the MUSE spectra as similar to LAMOST spectra in its wavelength grid and spectral resolution as possible, which leads to the process of degrading and interpolating MUSE spectra. This is the procedure that determines whether we can derive precise stellar labels from MUSE spectra successfully.

To minimize the difference between LAMOST and MUSE, we employ the line spread function (LSF) of both instruments from X19 and Bacon et al. 2017. The line spread function is FWHM in each wavelength grid. The process of degrading is performed by Gaussian kernel smoothing over the whole wavelength with $\sigma(\lambda_{\rm i})$, which is calculated from the equation given by

$$\sigma(\lambda_{\rm i}) = \sqrt{\left(\frac{\mathrm{FWHM_L}(\lambda_{\rm i})}{2.355}\right)^2 - \left(\frac{\mathrm{FWHM_M}(\lambda_{\rm i})}{2.355}\right)^2}, \tag{2.1}$$

where $\mathrm{FWHM_L}(\lambda_{\rm i})$ and $\mathrm{FWHM_M}(\lambda_{\rm i})$ are the LAMOST and MUSE line-spread function, respectively. After degrading, the spectra will be normalized by a pseudo-continuum which is derived by smoothing the spectra with a Gaussian kernel of 50Å in width, which is the same method employed by T17 & X19 for *DD-Payne* training set spectra normalization. This is also an important procedure to ensure the similarity between these two sets of spectra.

Figure 2.2 shows the normalized spectra of one single star observed by both LAMOST (in orange) and MUSE (in blue) and their residuals. The MUSE spectra were degraded using the above method. The gray areas are the masked areas due to telluric lines and other interference in the LAMOST wavelength grid, which will not be processed by *DD-Payne*. This figure demonstrates that for the same star, the difference between the LAMOST and the MUSE spectra is less than 0.03 even in regions with strong absorption lines. This is encouraging and is essential for reliably estimating chemical abundances. For the LAMOST-MUSE cross-validation datacubes, the spectral SNR pix^{-1} is in general higher than 200, which is much more than the average SNR of spectra in LAMOST (~40). After degrading (binning), the SNR is expected to increase even further.

2.3.2 Estimating stellar labels and errors

After degrading and normalizing the MUSE spectra to the LAMOST wavelength grid and spectral resolution, the next step is to estimate the stellar labels using the *DD-Payne* models. There are three models (so-called *DD-Payne-G*, *DD-Payne-A* and *DD-Payne-A-mp* below) trained by X19 with different training sets. The *DD-Payne-G* model used spectra from LAMOST ($\mathrm{SNR}^{\mathrm{g-band}}_{\mathrm{LAMOST}} > 50\ \mathrm{pix}^{-1}$ and $\mathrm{SNR}^{\mathrm{g-band}}_{\mathrm{LAMOST}} > 30\ \mathrm{pix}^{-1}$ for stars with $[\mathrm{Fe/H}] < -0.6$) and stellar labels from GALAH DR2 (Buder et al., 2018) while the *DD-Payne-A* model adopted LAMOST spectra ($\mathrm{SNR}^{\mathrm{g-band}}_{\mathrm{LAMOST}} > 50\ \mathrm{pix}^{-1}$) and stellar labels from APOGEE-*Payne* (Ting et al., 2019) for their respective cross-matched common samples. The *DD-Payne-G* derives stellar labels for $T_{\rm eff}$, $\log g$, [Fe/H] and abundance ratios [X/Fe] for 20 elements: Li, C, O, Na, Mg, Al, Si, Ca, Ti, V, Cr, Mn, Co, Ni, Cu, Zn, Y, Ba, and Eu. *DD-Payne-A* derive stellar labels of $T_{\rm eff}$, $\log g$, micro-turbulence velocity ($V_{\rm mic}$), [Fe/H] and abundance ratios for 15 elements: C, N, O, Na, Mg, Al, Si, Ca, Ti, Cr, Mn, Co, Ni, Cu, and Ba. Since X19 described these models in great detail, we direct the readers to this paper for more information about the selection criteria, validation and comparisons of the performances of different models. However, we plot the parameter ranges of these training sets in $T_{\rm eff}$, $\log g$ and [Fe/H] in Figure 2.4 as square markers for later comparison with the validation data in Section 2.4.1. In this work, we run all these models to test their performances because behind the training sets are GALAH and APOGEE spectra, which have different wavelength regions. The use of all these models will build a complete set of elements and also test the non-negligible systematics among

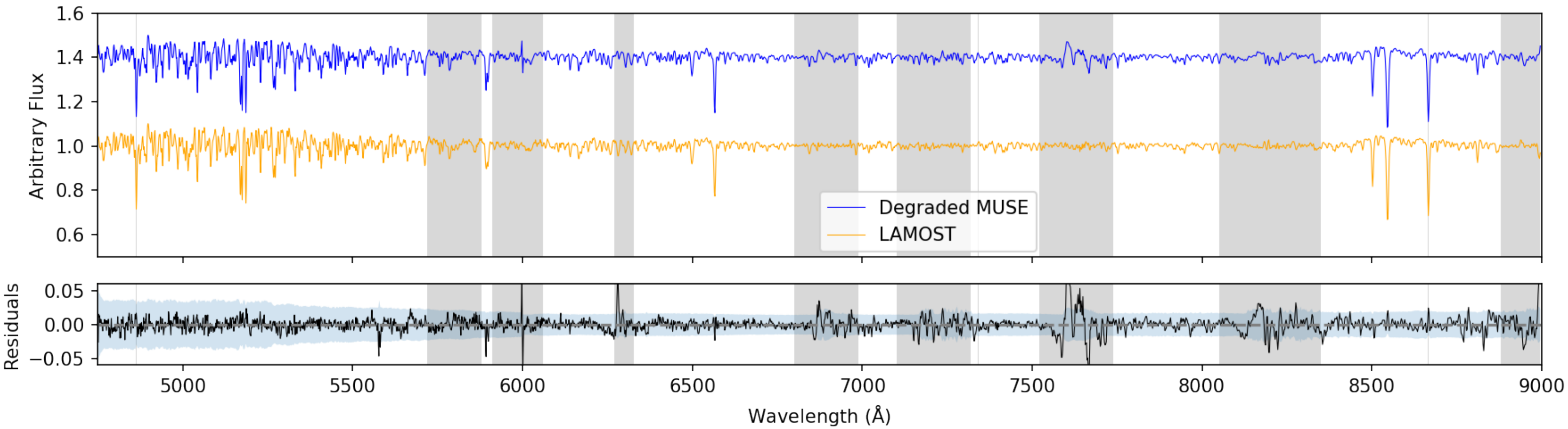

Figure 2.2: Comparison of LAMOST spectra (in yellow) with MUSE spectra (in blue) of the same star. The MUSE spectra have been degraded to the same resolution. Then both spectra were normalized in the same way as X19 with a pseudo-continuum derived by smoothing the spectra with a Gaussian kernel of 50Å in width. Gray areas are the masked pixels (due to telluric lines and other sources of interference) which will not be used when fitting with *DD-Payne*. The bottom plot shows the residuals of these two spectra. From the residuals, we can see that LAMOST spectra and MUSE spectra are similar with residuals less than 0.03. The blue shaded area represents the uncertainty of the residual, which is calculated using the flux error of the LAMOST and MUSE spectra. The flux error of MUSE spectra is re-estimated by multiplying σ in the distribution of flux residuals of the fitting divided by the flux error, see the right panel of Figure 2.3

.

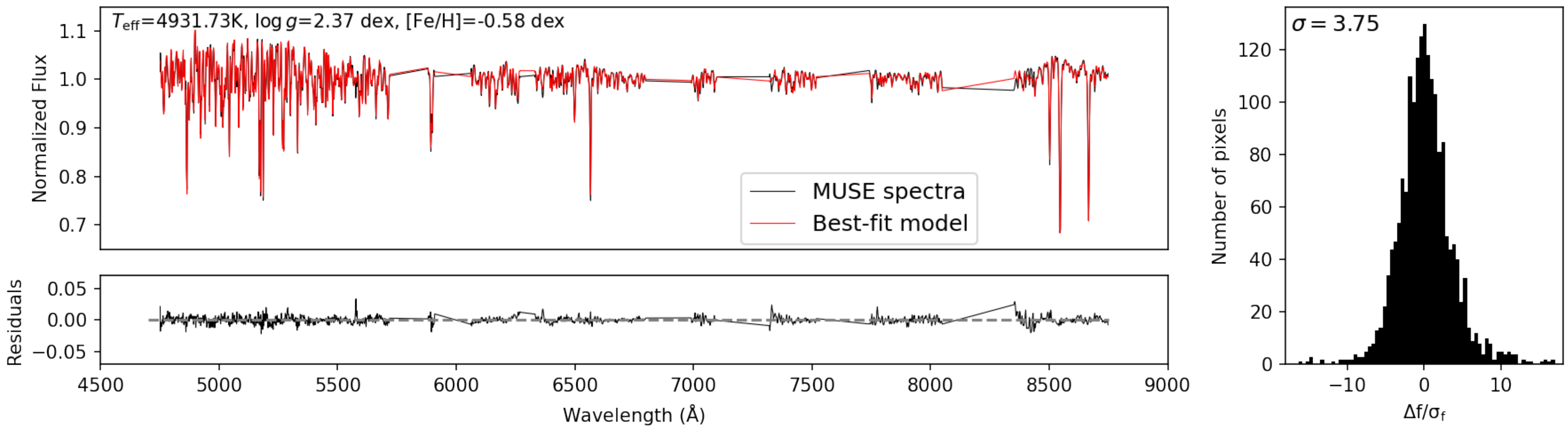

Figure 2.3: Comparison of MUSE spectra (in black) with model predicted spectra (in red) from *DD-Payne-G* on the same star as Figure 2.2. The left bottom plot shows their residuals. The masked regions in grey in Figure 2.2 are not shown in this plot. From the residuals we can see the difference between MUSE spectra and predicted spectra is small, with the residuals less than 0.03 even in absorption regions, demonstrating that *DD-Payne* training models can fit the MUSE spectra very well. The right panel shows the distribution of flux residuals divided by the flux error. A Gaussian fit to the distribution yields a standard deviation of 3.75, which demonstrates the underestimation of flux error of MUSE spectra.

the high-resolution surveys' labels. In addition, for some overlapped elements, we also compare their bias and dispersion, which can provide a flexible choice of models for the users in the future. Based on results from Figure 2 in X19, Li, Sc, V, Zn, Y, Eu were removed due to the weak correlations of the gradients, and will not be considered in this study.

The process of applying the *DD-Payne* to MUSE spectra is as follows: After loading the *DD-Payne* models, same as X19, several wavelength regions are masked including areas with telluric bands, the very red wavelength region (> 8750Å), and a region overlapped by the dichroic (5720−6060Å) where the LAMOST spectral performance is low (Xiang et al., 2015). Each MUSE spectra is then fitted with both the *DD-Payne-G* and *DD-Payne-A* model. Following the procedure outlined in X19, we also run an additional model for stars being identified as metal-poor in *DD-Payne-A* ($[\mathrm{Fe/H}] < -0.6$) to improve the precision for metal-poor stars. This *DD-Payne* APOGEE metal-poor model (*DD-Payne-A-mp*) is trained by X19 based on common metal-poor stars in APOGEE-*Payne* and LAMOST DR5. For stars with $[\mathrm{Fe/H}] < -1.0$ dex, only labels from *DD-Payne-A-mp* model are adopted. For stars with metallicity in $-1.0 < [\mathrm{Fe/H}] < -0.6$ dex, we take the weighted mean value of results from *DD-Payne-A* and *DD-Payne-A-mp* models, with weighting given below by Equation 2.2. For stars with $[\mathrm{Fe/H}] > -0.6$ dex, we adopt the *DD-Payne-A* measurements.

$$\begin{aligned} &\omega_{\mathrm{MP}} = ([\mathrm{Fe/H}]_{\mathrm{MR}} + 0.6)/(-0.4), \\ &\omega_{\mathrm{MR}} = 1 - \omega_{\mathrm{MP}}, \\ &[\mathrm{X/Fe}] = \omega_{\mathrm{MR}} \times [\mathrm{X/Fe}]_{\mathrm{MR}} + \omega_{\mathrm{MP}} \times [\mathrm{X/Fe}]_{\mathrm{MP}}. \end{aligned} \tag{2.2}$$

Here MR is the label from *DD-Payne-A*, and MP is the label from *DD-Payne-A-mp*, ω is the relative weight for each label [X/Fe]. In the following sections, when we refer to the labels derived from *DD-Payne-A* we mean this combined set of results.

In addition, we determine the radial velocity and its error using the Doppler equation at the same time during the fit, which is the same as *The Payne* (Ting et al., 2019).

We also determine two values to represent the quality of fitting. One is the reduced χ^2, calculated given by

$$\chi^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{(P_i - O_i)^2}{e_i^2}, \tag{2.3}$$

where n is the number of wavelength pixels, P_i, O_i and e_i are the best-fit spectra, observed spectra and the flux error in the i-th wavelength pixel. The other is the Pearson correlation coefficient between the pixel values of the observed and the best-fit *DD-Payne* spectra.

Figure 2.3 illustrates the comparison of observed MUSE spectra (in black) with the best fit by *DD-Payne* (in red) for the same star as that in Figure 2.2. The residual spectrum is shown in the bottom panel. The difference is less than 0.03 over most of the spectrum, including areas with strong absorption lines. This demonstrates that *DD-Payne* training models can reproduce the MUSE spectra with high fidelity.

2.3.3 Assessing the precision of labels

Even though our best-fit model spectra are quite similar to observed MUSE spectra, it is not guaranteed that all stellar labels from *DD-Payne* are estimated precisely. Many spectral features suffer from blending due to the low spectral resolution. Since *DD-Payne* is developed to avoid astrophysical correlations, the imprecise stellar labels are due to weak or no absorption features in the fitting wavelength window. When compared to LAMOST spectra, MUSE loses information in the blue regions 3700 – 4750Å, which contains absorption features of many elements. Therefore, we cannot expect the performance of *DD-Payne* on the MUSE spectra to be as good as that on the LAMOST spectra, even if MUSE spectra have higher SNR. To check the precision of estimated stellar labels, we follow T17 & X19's work, and for each stellar label, we compare the gradient spectra of *DD-Payne* with Kurucz spectral model (Kurucz, 1970, 1993, 2005) and calculate their correlation coefficient. If a label is physically measured from the spectral absorption features, the gradient spectra predicted by *DD-Payne* should be very similar to the Kurucz spectral model, with a correlation coefficient close to 1. Otherwise, if a weak or no absorption feature is identified for the label the correlation coefficient will be very low. Due to the computational expense of calculating the theoretical model, we use a template of Kurucz model gradient spectra from X19 (see Table 1) which is calculated based on a list of reference stars covering a wide range in the parameter space, from −2.5 to 0.5 dex in [Fe/H] and 4000 to 7000 K in $T_{\rm eff}$. Therefore, for each star, we compare the gradient spectra from *DD-Payne* with one of the template spectra having the closest parameter distance to the target star. The parameter distance is defined by X19 as

$$D = \sqrt{(\Delta T_{\rm eff}/100{\rm K})^2 + (\Delta \log g/0.2)^2 + (\Delta[{\rm Fe/H}]/0.1)^2}. \tag{2.4}$$

Finally, the correlation coefficient of each label will be provided in the final catalog as well as a flag based on the correlation values to represent whether the label is precise or not. Here we assign the flag of 1 if the correlation coefficient is larger than 0.5, otherwise, the flag is set to 0. In this way, we will know which elements are determined precisely for each star.

2.3.4 Building the catalog

After deriving two series of stellar labels for each star by applying *DD-Payne-G* and *DD-Payne-A* models and assessing the precision of each label value, we compile two catalogs. Table 2.2 presents a description of the columns in the catalogs. Due to different science goals, wavelength coverage, and data quality, the relative performances of the two models are different for different stellar labels. Some elements are estimated better in one model as compared to the other. In normal circumstances, the version of the label with a larger gradient spectra correlation coefficient should be selected. In the following Section 2.4.1, we discuss in detail the bias and precision of stellar labels from these training models.

2.4 Results

In Section 2.3 we gave detailed steps for applying the *DD-Payne* on MUSE spectra, measuring their stellar labels and building the *DD-Payne-A* and *DD-Payne-G* catalogs. To test

Table 2.2: Descriptions of all the column names and information for the DD-Payne stellar label catalog.

Column	Desceiption	Unit
stellar_id	Stellar ID of stars	String
ra	Right ascension from the input catalog (or RAJ2000)	deg
dec	Declination from the input catalog (or DEJ2000)	deg
snr_muse	Spectra signal-to-noise ratio calculated by median of all pixels	pix^{-1}
$T_{\rm eff}$	Effective temperature	K
$T_{\rm eff}$_err	Uncertainty in $T_{\rm eff}$	K
$T_{\rm eff}$_flag	Quality for the value based on the examination of the DD-Payne gradient spectra $\partial f_\lambda/\partial T_{\rm eff}$	
$T_{\rm eff}$_gradcorr	Correlation coefficients of $\partial f_\lambda/\partial T_{\rm eff}$ between the DD–Payne and the Kurucz model	
$\log g$	Surface gravity	in cgs
$\log g$_err	Uncertainty in $\log g$	in cgs
$\log g$_flag	Quality for the value based on the examination of the DD-Payne gradient spectra $\partial f_\lambda/\partial \log g$	
$\log g$_gradcorr	Correlation coefficients of $\partial f_\lambda/\partial \log g$ between the DD–Payne and the Kurucz model	
$\mathrm{V_{mic}}$	Micro-turbulent velocity	km s^{-1}
$\mathrm{V_{mic}}$_err	Uncertainty of micro-turbulent velocity	km s^{-1}
$V_{\rm mic}$_flag	Quality for the value based on the examination of the DD-Payne gradient spectra $\partial f_\lambda/\partial V_{\rm mic}$	
$V_{\rm mic}$_gradcorr	Correlation coefficients of $\partial f_\lambda/\partial V_{\rm mic}$ between the DD–Payne and the Kurucz model	
[Fe/H]	Metallicity	dex
[Fe/H]_err	Uncertainty of metallicity	dex
[Fe/H]_flag	Quality for the value based on the examination of the DD-Payne gradient spectra $\partial f_\lambda/\partial$[Fe/H]	
[Fe/H]_gradcorr	Correlation coefficients of $\partial f_\lambda/\partial$[Fe/H] between the DD–Payne and the Kurucz model	
[X/Fe]	Chemical abundance fraction to iron, sorted by Mg, Si, Ca, Ti	dex
[X/Fe]_err	Uncertainty of chemical abundance	dex
[X/Fe]_flag	Quality for the value based on the examination of the DD-Payne gradient spectra $\partial f_\lambda/\partial$[X/H]	
[X/Fe]_gradcorr	Correlation coefficients of $\partial f_\lambda/\partial$[X/H] between the DD–Payne and the Kurucz model	
red_chi2	Reduced χ^2 of the spectral fit (hereafter χ^2)	
corr_flux	Correlation coefficient of the muse and best-fitting model spectra	

whether our method can measure stellar labels from MUSE observations with high precision, we now do two validations: First, we compare the *DD-Payne*-estimated stellar labels from common stars observed by both MUSE and LAMOST; Second, we use our method to analyze the [Fe/H] and [Mg/Fe] distributions of stars in the bulge and compare our results with that of the previous high-resolution spectroscopic studies. The motivation for each validation has been discussed in detail in Section 2.2.3.

2.4.1 Validation 1: LAMOST-MUSE cross-validation samples

We follow the steps described in Section 2.3 and derive stellar labels of common stars in LAMOST and MUSE using both *DD-Payne-G* and *DD-Payne-A* models. To ensure the precision of labels, we use the selection criteria given by

$$\begin{cases} \mathrm{SNR_{LAMOST}} > 35\mathrm{pix}^{-1} (\sim \mathrm{SNR}^{\mathrm{g-band}}_{\mathrm{LAMOST}} > 23\mathrm{pix}^{-1}) \\ \mathrm{SNR_{MUSE}} > 80\mathrm{pix}^{-1} \end{cases} \tag{2.5}$$

Here $\mathrm{SNR_{LAMOST}}$ and $\mathrm{SNR_{MUSE}}$ is the median signal-to-noise ratio of the LAMOST and MUSE spectra, respectively. We made a cut-off of $\mathrm{SNR_{LAMOST}}$ to ensure the correction of stellar parameters measured from LAMOST spectra, this threshold is looser than the criteria of *DD-Payne* training set ($\mathrm{SNR}^{\mathrm{g-band}}_{\mathrm{LAMOST}} > 30\ \mathrm{pix}^{-1}$) to keep more common stars. The threshold of $\mathrm{SNR_{MUSE}}$ cut-off is more strict than that of $\mathrm{SNR_{LAMOST}}$ because the SNR of MUSE spectra is somehow overestimated after data reduction pipeline, which is confirmed by looking at the spectra of the two validation data and the distribution of flux residuals of the fitting divided by the flux error (the right panel of Figure 2.3). Nevertheless, most MUSE spectra have SNR pix^{-1} greater than 200. This is because these observations typically have long exposure time, as they were targeting external galaxies which are much fainter than the stars we are interested in. Finally, we end up with 27 common stars in *DD-Payne-G* and *DD-Payne-A* results, during which the χ^2 is mostly concentrated in 4~25, meaning the $\mathrm{SNR_{MUSE}}$ is overestimated by a factor of 2~5. Star markers/diamonds in Figure 2.4 represent this cross-validation sample. These stars span a reasonable range of T_{eff}, $\log g$ and [Fe/H] than those from *DD-Payne* training sets. One star is measured by extrapolation, which will be discussed in Section 2.4.1.

Bias and precision check for main labels

In this section, we focus on T_{eff}, $\log g$ and several chemical abundances which we reported with confidence in the abstract. Figure 2.5 shows the differences of *DD-Payne*-estimated stellar labels from MUSE and LAMOST spectra, as a function of [Fe/H]. The median and dispersion of all points are marked in the upper left corner of each panel, where the dispersion is calculated as half of the difference between 15.87 and 84.13 percentile values. The median and dispersion are indicative of the bias μ and precision σ of our label estimates, respectively. The left and right panels show results for the *DD-Payne-G* and *DD-Payne-A* models, respectively. This figure demonstrates that the dispersion of ΔT_{eff} is about 75K, $\Delta \log g$ is around 0.15 dex for both the *DD-Payne-G* and *DD-Payne-A*. As for chemical abundances, the dispersion is less than 0.1 dex in general and we consider this to be precise, as this is roughly the uncertainty in the LAMOST labels themselves. Therefore, most of the stellar labels from MUSE spectra and LAMOST spectra are in

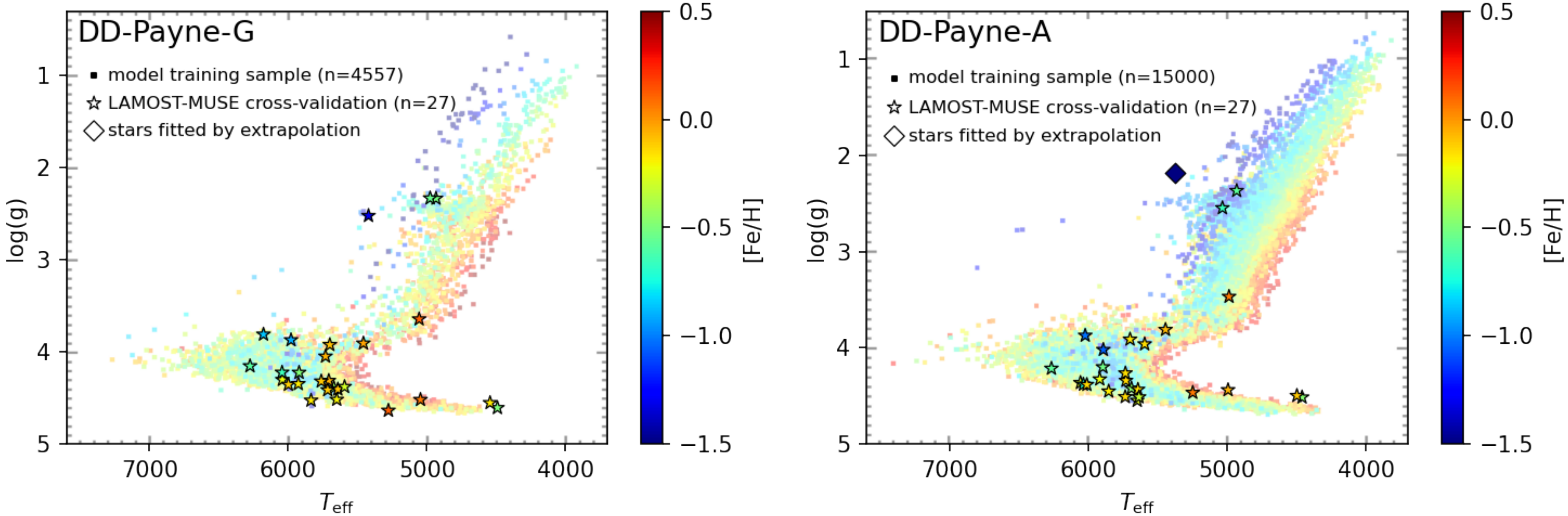

Figure 2.4: Distributions of the *DD-Payne* training set (Figure 5 in X19, in square markers) and LAMOST-MUSE cross-validation samples (in star markers) in the T_{eff}-$\log g$ plane color-coded by [Fe/H]. Left panel: Distribution of the *DD-Payne-G* training stars, the T_{eff} and $\log g$ are from a corrected version GALAH DR2 values (see Section 3 and Appendix A of X19) The number of stars for each sample is noted in each subplot. The LAMOST-MUSE cross-validation samples are selected following the criteria of Section 2.4.1 and span reasonable parameter ranges than those from *DD-Payne* training sets. No stellar parameter is measured by extrapolation.

good agreement in Figure 2.5. Even though the dispersion of some elements (e.g. [Si/Fe] from *DD-Payne-G* and [Cr/Fe] from *DD-Payne-A*) is slightly higher, it is still close to 0.1 dex. The α-elements listed in this figure are some of the most useful ones as they can be used to constrain intrinsic stellar properties such as evolutionary stage, distance, birth radius, extinction, and even age (Ness et al., 2019; Hayden et al., 2022; Sharma et al., 2022). Therefore, this agreement shows the usefulness of *DD-Payne* based estimates from MUSE observations for Galactic studies. The bias in this figure reflects the differences between LAMOST and MUSE instruments and the different performances when losing wavelength in 3700 – 4750Å. However, it can be seen that all of them are smaller than the dispersion. Therefore, in this work we do not perform the bias correction on any label. Whether to do it will be re-investigated when we have a larger sample.

In addition, by comparing the dispersion of the same elements under different model estimates, we also found that except [Fe/H], [Ti/Fe] and [Cr/Fe], all the labels calculated using the *DD-Payne-A* model have lower dispersion than that from the *DD-Payne-G* model. The dispersion of [Mg/Fe] from the *DD-Payne-A* model is even less than 0.05 dex. Since we derived labels from the same spectra, the reason must be the relative performance of the different models. Additionally, when comparing the derived gradient spectra with Kurucz spectral model for these labels, *DD-Payne-G* with lower correlation coefficients, which indicates *DD-Payne-A* model has better performance. This is because the training set of *DD-Payne-A* is larger than that of *DD-Payne-G*, so the measured stellar labels will be more precise.

There is one star in the panel of *DD-Payne-A* measured by extrapolation since the [Fe/H] is less than -1.5 dex, which is shown in the right panel of Figure 4 as a "diamond". By looking at the parameter differences in each panel, we can see from Figure 5 that $T_{\rm eff}$, $\log g$, [Fe/H], [Mg/Fe] and [Si/Fe] measured from MUSE spectra of this star have good agreements with those measured from LAMOST spectra. But for the other parameters, the differences are larger. Note that the LAMOST spectra of this star was also measured by extrapolation, whether this indicates anything about extrapolation will be investigated when we have more stars.

Bias and precision check for all labels

Other than the parameters discussed in the previous section, we also estimated abundances for additional elements, with 23 stellar labels for *DD-Payne-G* and 16 labels for *DD-Payne-A*. However, the dispersion for many of the abundances is not shown as they are much larger than those analyzed in Figure 2.5. This is because, unlike LAMOST, MUSE does not have any pixels in the wavelength range 3700 – 4750Å, which contains a large number of absorption features. In this section, we will further discuss the precision of these labels and compare them with the results of cross-validation samples between LAMOST and GALAH or APOGEE.

We perform the same analysis in Figure 2.5, and derive bias (μ) and the dispersion (σ) for all labels in Figure 2.6 and Figure 2.7. In each plot stellar labels from MUSE are determined in two regions: λ = (4750 − 8750Å) (in blue) and λ = (5991 − 8750Å) (in yellow), corresponding to normal spectra and spectra in high extinction fields such as the Galactic bulge. We also over-plot the dispersion of label differences between those from

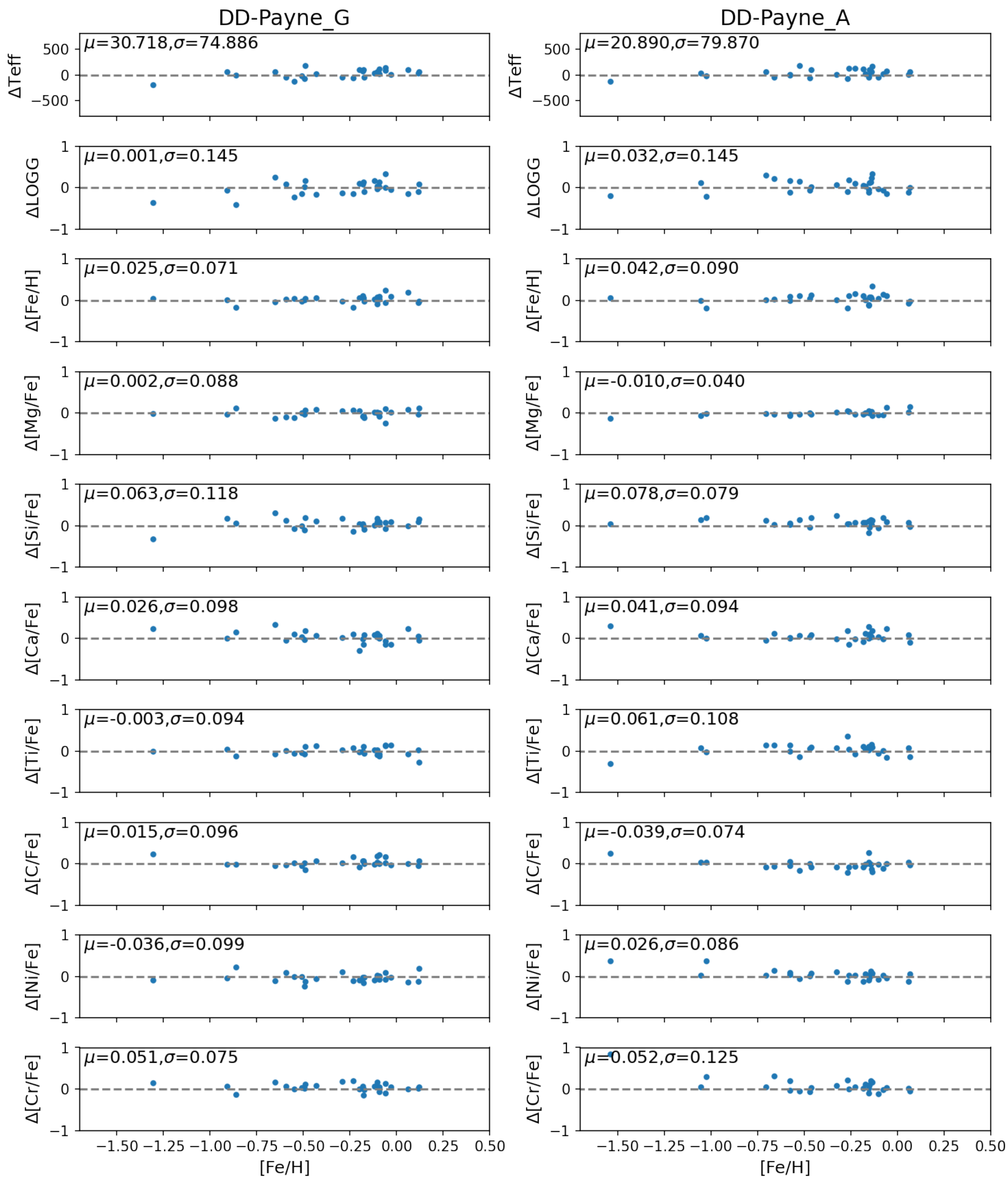

Figure 2.5: Differences of stellar labels derived from MUSE spectra with those from LAMOST spectra using *DD-Payne-G* (left column) and *DD-Payne-A* (right column) models as a function of [Fe/H]from MUSE spectra, respectively. All the stars in these plots have LAMOST spectra with SNR higher than 35 pix^{-1}. The median and dispersion of all points are marked in the upper left corner of each panel, the dispersion is calculated as half of the difference between 15.87 and 84.13 percentile values.

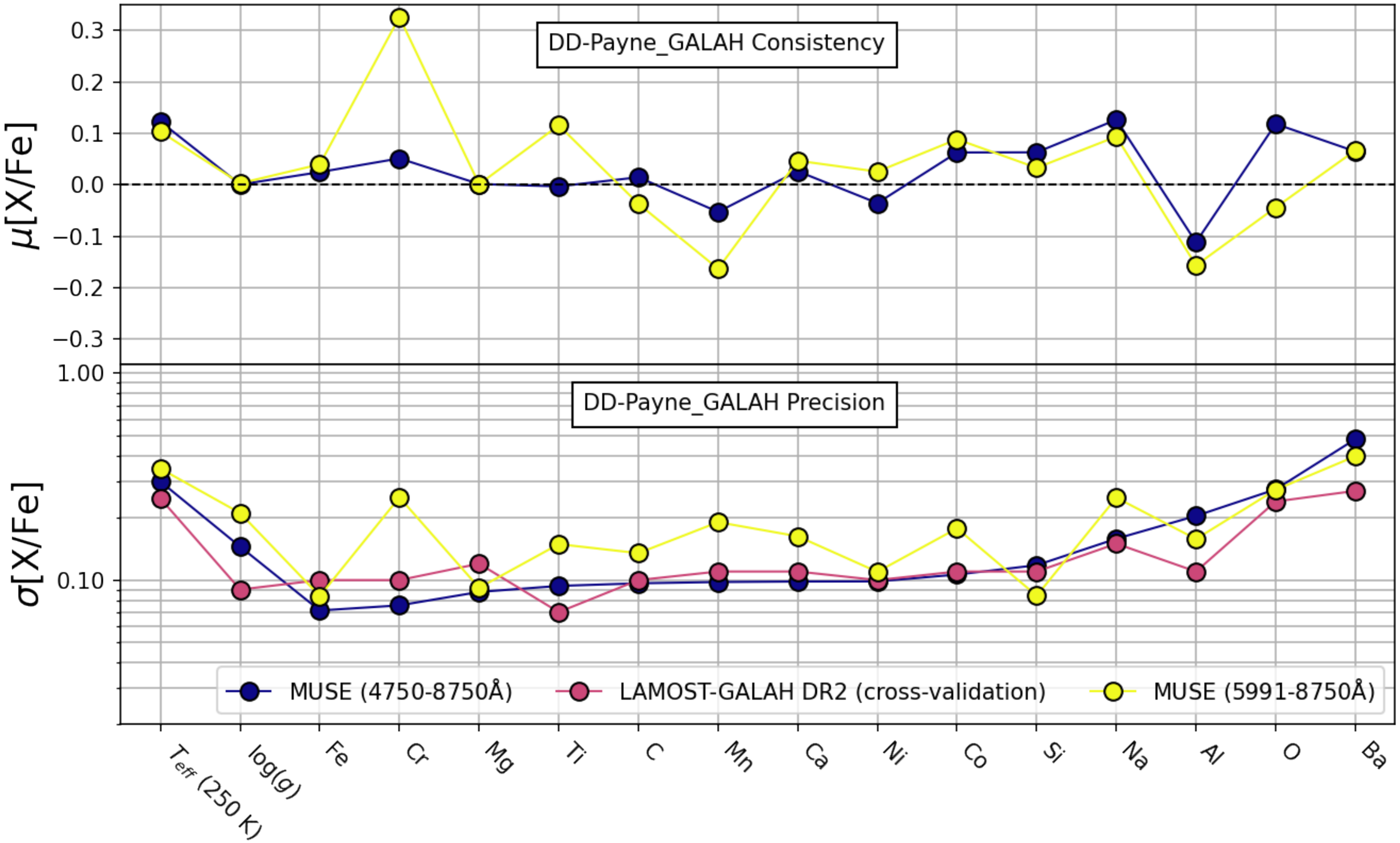

Figure 2.6: The bias (top panel) and precision (bottom panel) of all stellar labels derived from *DD-Payne-G* model, respectively. The bias and precision are calculated in the same way as Figure 2.5 by taking the median and half the difference between 15.87 and 84.13 percentile values. We list all measured labels and arrange them from small to large according to the precision of the elements. We also plotted results in the Bulge case which loses pixels between (4750 – 5991Å). From these two plots, for the *DD-Payne-G* model, Fe, Mg, Ni, Si are precisely estimated with σ close to or better than 0.1 dex when fitting MUSE spectra between both (4750 – 8750Å) and (5991 – 8750Å). As for bias, most of the elements are within 0.1 dex except Co, Mn, Na, and Al.

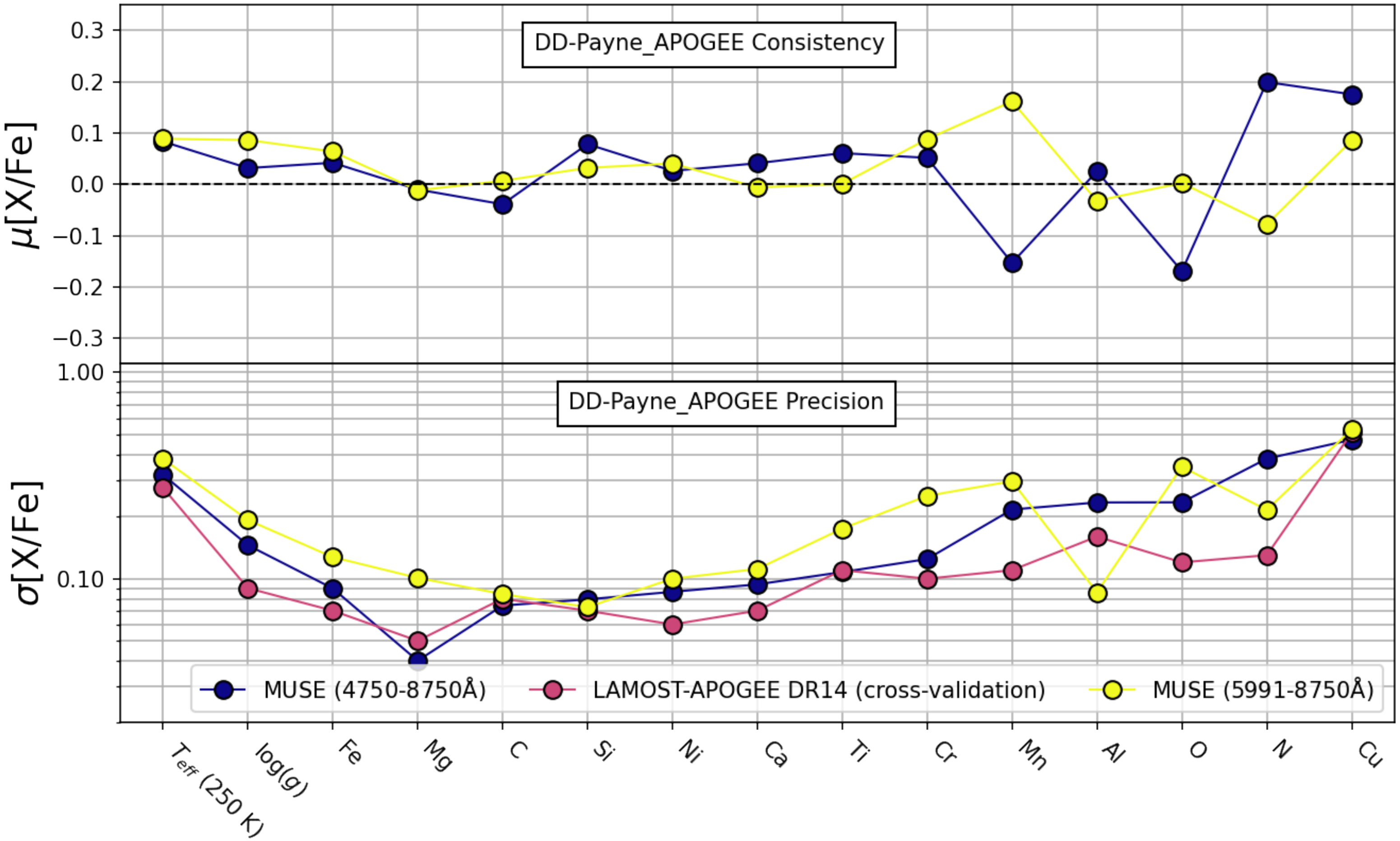

Figure 2.7: The same plot as Figure 2.6 but now using *DD-Payne-A* model. From these two plots, for *DD-Payne-A* model, Fe, C, Si, Mg, Ni, and Ca are well estimated with σ less than 0.1 dex when fitting MUSE spectra between both (4750 – 8750Å) and (5991 – 8750Å). Besides, MUSE (5991 – 8750Å) shows a larger bias and dispersion. As for bias, most of the elements are within 0.1 dex except Mn, O, N, and Cu. Comparing with results from *DD-Payne-G* model, for the same element, *DD-Payne-A* model has better precision.

LAMOST spectra and GALAH DR2 (Buder et al., 2018) or APOGEE-*Payne* (Ting et al., 2019) in red.

For the LAMOST-MUSE cross-validation sample (the blue line), elements such as Na, Al, O, and Ba in *DD-Payne-G* model and Mn, Al, O, N, and Cu in *DD-Payne-A* have high median and dispersion (larger than 0.1 dex). This can be due to two reasons. We find that the bias and dispersion of Na, Al, O, and Ba in cross-validation samples are also high. For these elements, the high dispersion is the inheritance from the *DD-Payne* models. However, for Mn and N, the high dispersion is due to the loss of the blue wavelength region (3700 – 4750Å) from the MUSE spectra compared to the larger wavelength region covered by LAMOST. This can be seen in Figure 2.7, where μ and σ show a large discrepancy between the LAMOST-GALAH/APOGEE cross-validation estimates and the LAMOST-MUSE cross-validation estimates. This is further confirmed by the gradient spectra of these two elements (See Figure 22 in X19), where the dominant absorption features of Mn and N are in 3700 – 4750Å, which the MUSE spectra are lacking.

In general, for most of the labels in these two figures, the dispersion of the cross-validation estimates and MUSE estimates are in agreement. Therefore, it can be verified that the models designed for LAMOST spectra can be applied to MUSE spectra, and measure some of labels ($T_{\rm eff}$, $\log g$, [Fe/H], [Ni/Fe], [Cr/Fe], [C/Fe] and α abundances) with similar precision. It is worth noting that there are several elements in these figures in which MUSE has higher precision than the LAMOST cross-validations, for example, Fe, Cr, Mg in *DD-Payne-G* and Mg in *DD-Payne-A*, this is due to the small number of stars in the common sample of LAMOST and MUSE.

For the LAMOST-MUSE cross-validation estimates in the wavelength region of λ = (5991 – 8750Å) (the yellow line). Both bias and dispersion are larger than when it is fitted in 4750 – 8750Å (the blue line). This is because more absorption features in the region of 4750 – 5991Å are lost. Nevertheless, some elements such as [Fe/H], [Mg/Fe], [Si/Fe] in *DD-Payne-G* and [Fe/H], [Mg/Fe], [Ca/Fe], [C/Fe], [Ni/Fe], [Si/Fe] in *DD-Payne-A* still have dispersion lower than 0.1 dex. This indicates that after further losing pixels in the wavelength range of 4750 – 5991Å due to high extinction from the bulge, we still can recover some stellar labels for these fields.

Theoretical precision prediction

In Section 2.4.1 and Section 2.4.1 we discussed the bias and precision of stellar labels from MUSE spectra in comparison with results from LAMOST spectra and discussed the influence of losing wavelength pixels in 3700 – 4750Å and 4750 – 5991Å. The theoretical precision, which is the standard deviation of each label X (written as X_err in Table 2.2) as a by-product of *DD-Payne*, will also be affected by the changes in wavelength coverage. Since we don't have enough LAMOST and MUSE common stars with a given SNR, we use the method *Chem-I-Calc* (Sandford et al., 2020) to forecast the theoretical precision, which uses *ab initio* spectral model and the Cramér-Rao Lower Bound (CRLB) to quantify the chemical information content of stellar spectra in terms of theoretical precision. Figure 2.8 shows the predicted theoretical precision of stellar labels for a typical solar-metallicity K-Giant with a g-band SNR of 50 pix^{-1} in three different scenarios. By comparing these scenarios, we find that when losing wavelength range in 3700 – 4750Å (cyan versus blue),

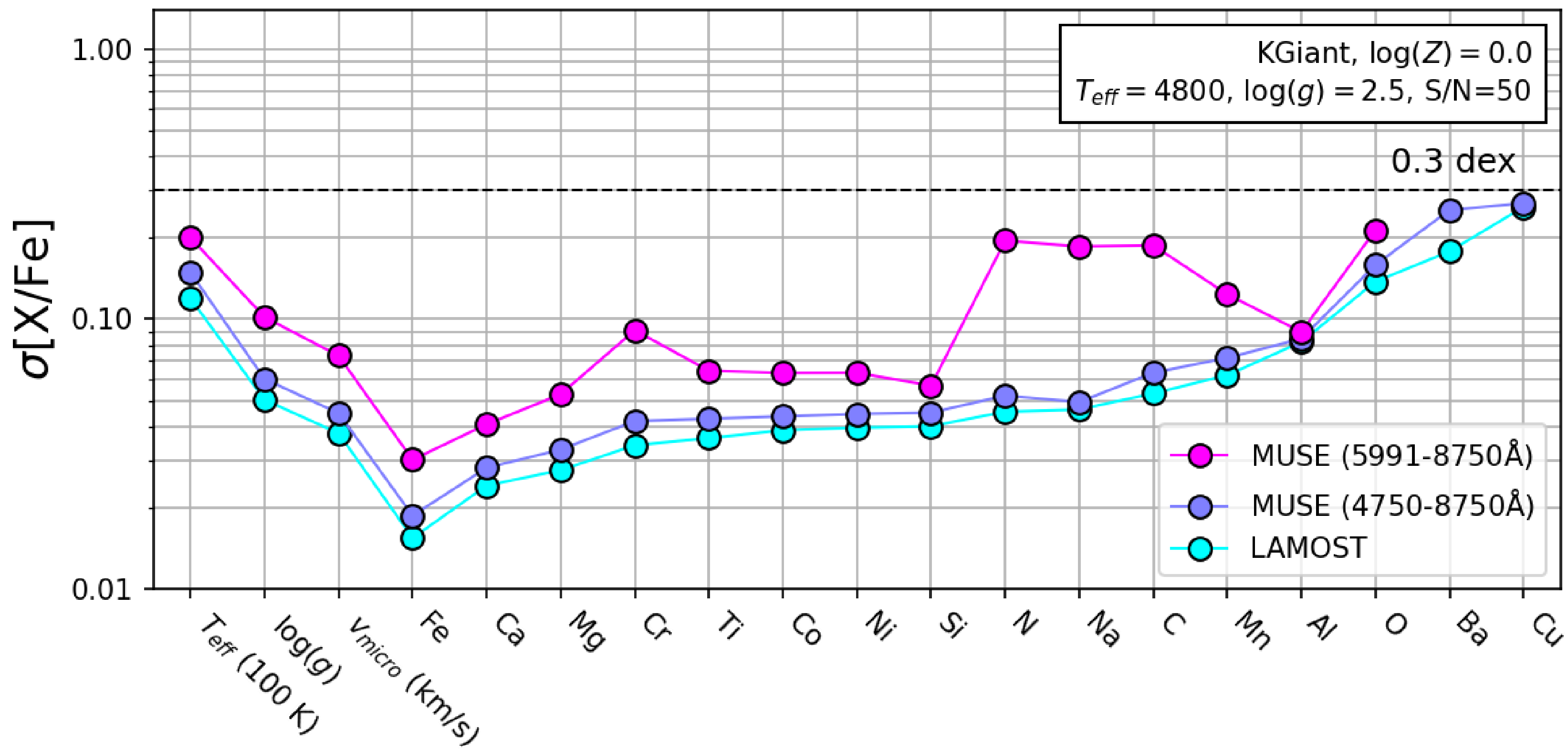

Figure 2.8: Estimation of the theoretical precision for a typical solar-metallicity K-Giant with a g-band SNR of 50 pix^{-1} in three different scenarios. The LAMOST spectra cover the wavelength range between 3700 – 8750Å. Both MUSE (4750 – 8750Å) and MUSE (5991 – 8750Å) have the same resolution as LAMOST. The theoretical precision is predicted by using the *Chem-I-Calc* method which is an efficient method for computing the expected precision of stellar labels determined via full spectral fitting (Sandford et al., 2020). When losing wavelength pixels in wavelength 3700 – 4750Å (LAMOST versus MUSE (4750 – 8750Å), the precision of all the labels is getting slightly worse by an average factor of 1.17, but there is no label having a large change. When further losing wavelength pixels in the range of 4750 – 5991Å (i.e., for MUSE cubes in high extinction fields), the precision of all the labels is worse by an average factor of 2.16. However, there are some labels such as Cr, N, Na, and C which have much larger shifts towards larger uncertainties by a factor of 2.66, 4.30, 4.02, 3.53, respectively due to the loss in wavelength coverage.

the theoretical precision of all the labels is getting slightly worse by an average factor of 1.17, but there is no label having a large change. When further losing wavelength pixels in wavelength 4750 – 5991Å (blue versus pink), the precision of all the labels is getting worse by an average factor of 2.16. However, there are some labels such as Cr, N, Na, and C which have a much greater theoretical precision change with a factor of 2.66, 4.30, 4.02, 3.53, respectively, as we lose critical absorption lines in a more restricted wavelength range. This is shown in the gradient spectra of these labels from Figure 22 of X19, where they show that there are no significant features for these abundances in the remaining wavelength range. Fortunately, precision of labels like $T_{\rm eff}$, $\log g$, [Fe/H], and some α elements are not significantly affected.

2.4.2 Validation 2: observations towards the Galactic bulge

From the above sections we have shown that we can derive stellar labels precisely for individual stars in MUSE observations via the utilization of the *DD-Payne*. Here we want to demonstrate that our method is also useful to determine stellar labels for stars in dense fields. The Milky Way bulge is one of the densest regions of the Galaxy. In the past decades, there have been many fiber-fed spectroscopic surveys focusing on the bulge to study its chemical distribution such as *Gaia*-ESO (Gilmore et al., 2012), ARGOS (Ness et al., 2013a), and GIBS survey (Pasquini et al., 2002), which have revealed multiple stellar components with different mean metallicity. A highlight of this study is from Rojas-Arriagada et al. (2019) who revealed the bimodality in [Mg/Fe]-[Fe/H] distribution using the bulge stars from APOGEE DR14 (Abolfathi et al., 2018). In terms of the integral-field spectroscopic survey, Valenti et al. (2018) observed four fields with MUSE towards the inner bulge and measure radial velocity based on the CaT region of spectra and reached $\sigma V_{\rm gc} \sim 140$ km s^{-1}. They also found that the velocity dispersion peak is symmetric with respect to the distance z from the Galactic plane. To summarize previous bulge studies, chemical element distributions have been studied based on high-resolution fiber-fed spectroscopic surveys, and kinematics has been studied using IFS data. Since our method can precisely determine stellar labels from MUSE spectra, this should enable us to study the chemodynamics of the Milky Way bulge using IFS data.

Data refinement

The bulge data processing begins with the steps outlined in Section 2.3. The spectra are extracted from the 29 bulge datacubes using PampelMUSE with VVV DR2 (Minniti et al., 2017) as the input photometry catalog. Next, the spectra are degraded to LAMOST spectral resolution of R$\sim$1800. Figure 2.9 shows the degraded spectra, both raw and normalized, of one bulge star (VVV-J175141.74-282207.85) extracted from the datacube.

The bulge stars pose several challenges for data reduction. They have large heliocentric distances, which means they are typically faint and have low SNR. These stars also typically lie along lines of sight with high extinction, making SNR quite low in the shorter wavelength regions. Finally, subtracting the night sky emission lines from faint stellar spectra is challenging. This highlights the challenges associated with reducing the spectra in the bulge region and reinforces the need to adopt additional data refinement procedures.

To remove emission lines, we attempted to apply the Zurich Atmosphere Purge (ZAP)

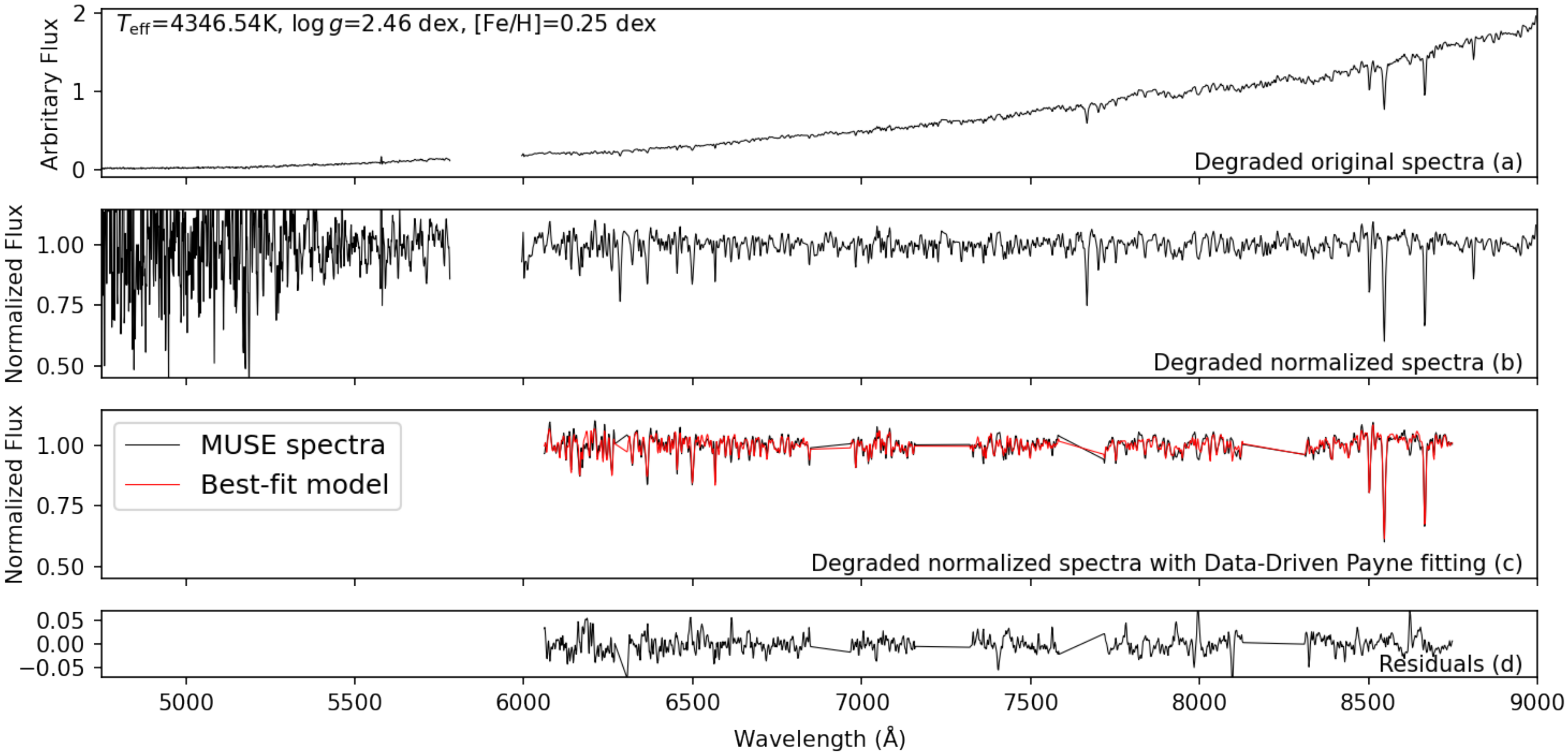

Figure 2.9: The detailed process of spectral processing of one bulge star in the MUSE datacube. (a) Original MUSE spectra after degrading which demonstrates the high extinction in the Galactic center causing almost no flux in short wavelength. (b) Spectra after normalization, which shows the extremely high extinction makes the short-wavelength spectra between (4750 – 5991Å) impossible to fit. (c) The predicted spectra and original MUSE spectra of this star. (d) The residuals of predicted spectra and the original MUSE spectra.

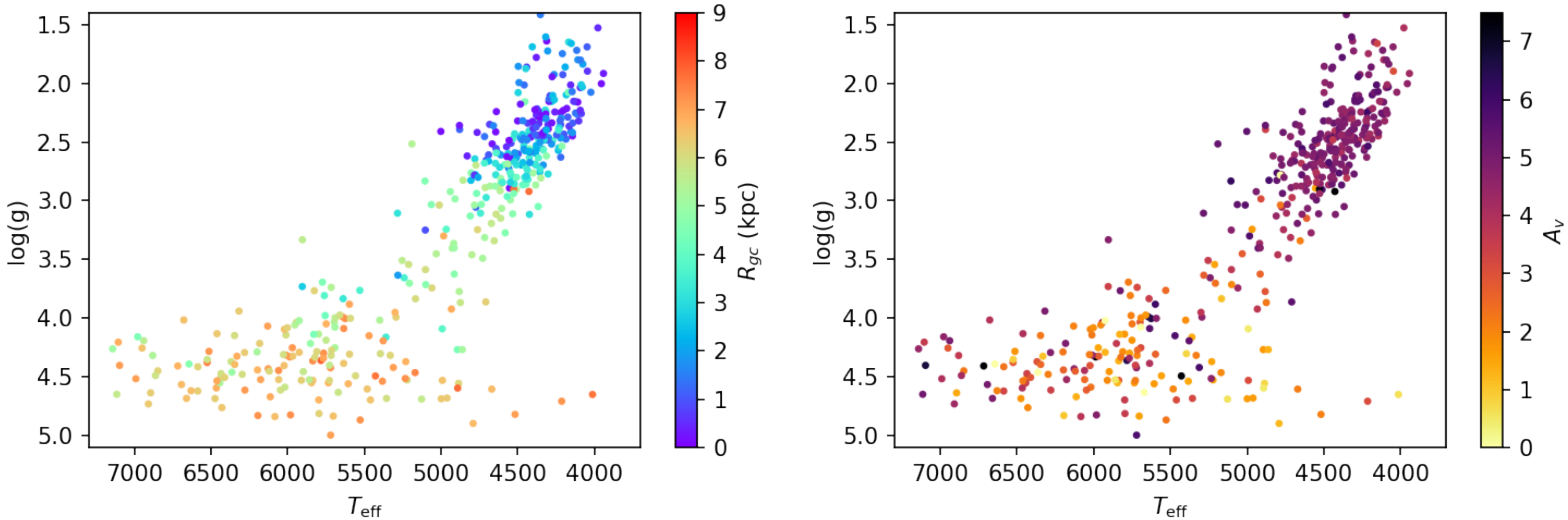

Figure 2.10: H-R diagrams of stars with good fitting to the model defined by Equation 2.6. The left-hand side panel is color-coded with $R_{\rm gc}$ (distance to the Galactic center) in cylindrical Galactocentric coordinates, where distance from the Sun to the Galactic Center is adopted as $R_{\odot}$= 8.2 kpc, $z_{\odot}$ = 25 pc (Bland-Hawthorn & Gerhard, 2016). The right-hand side panel is color-coded with extinction A_V. Both $R_{\rm gc}$ and A_V are predicted by BSTEP (Sharma et al., 2018) taking parameters of $T_{\rm eff}$, $\log g$, [Fe/H], [Mg/Fe] from DD-Payne and m_J, m_{Ks}, m_H magnitude from VVV DR2.

(Soto et al., 2016), which is an approach based on principal component analysis (PCA) to reduce emission lines for faint spectra in datacubes. However, using ZAP has some undesirable side effects. By adjusting some tunable parameters in ZAP, we can increase the efficiency of subtracting the emission lines, but then ZAP has a significant impact on the absorption lines which are desirable for estimating chemical abundances. Strangely, the spectra of bright stars were also found to become noisy. In the end, we find the most effective way to deal with emissions is to replace emission pixels with the average value of adjacent non-emission pixels In this way, we can retain more pixels in the spectra.

In addition, due to extinction in the bulge, the flux and hence the SNR in short-wavelength ranges is not enough to extract useful information (See panel (a) and (b) in Figure 2.9). The NaD doublet line (5803 – 5966Å) is also masked in datacubes that are observed using the adaptive optics module. Therefore, we decided to discard information for wavelengths less than 5991Å. We also mask the following telluric bands; 6270 – 6305Å, 6850 – 6966Å, 7160 – 7320Å, 7585 – 7715Å, 8130 – 8310Å. The widths of the telluric bands are narrower than those used by X19, this was done to increase the number of useful wavelength pixels. This range was determined after many tests to ensure that the unmasked pixels provide useful information about the chemical elements. Figure 2.6 and Figure 2.7 can indicate that when losing pixels in 4750 – 5991Å, the dispersion of some elements is still below 0.1 dex. In panel (c) of Figure 2.9 shows the best-fit model spectra in red, alongside the original spectra in black. It is clear from the comparison between LAMOST and MUSE samples presented in Figure 2.3, that the spectra of bulge stars have more noise, due to a combination of these stars having higher extinction and lower average SNR. In addition, the distribution of residuals between the observed and the best-fit model spectra for a bulge star (Figure 2.9 panel (d)) is larger than that for the star in Figure 2.3, which shows the difficulty of fitting spectra in the bulge fields.

After these refinement and masking steps, stellar labels are derived from two *DD-Payne* models. We adopt the labels from the *DD-Payne-A* model and elect not to use the *DD-Payne-G* model to compare directly with high-resolution APOGEE data which has ample stars in the bulge. Given the loss of precision due to the omission of the blue wavelength region, we only adopt the [Mg/Fe] when studying α abundances for the bulge fields.

Our final MUSE bulge catalog is comprised of 8720 individual spectra containing 2721 unique stars, with the following columns $T_{\rm eff}$, $\log g$, [Fe/H] and [Mg/Fe]. On average, from each cube about 300 spectra were extracted and 40% of them have $\mathrm{SNR_{MUSE}} > 80$ pix^{-1}. Compared with typical MUSE wavelength coverage, the bulge spectra lose the bluer regions of the spectrum, between 4750 – 5991Å, so we adopt more strict selection criteria compared to our LAMOST verification sample with

$$\begin{cases} \mathrm{SNR_{MUSE}} > 80\mathrm{pix}^{-1} \\ \mathrm{corr\_flux} > 0.85 \end{cases} \tag{2.6}$$

Here the $\mathrm{SNR_{MUSE}}$ is the SNR pix^{-1} in wavelength higher than 5991Å. Since there are multiple spectra for each star, we selected the observation having the highest corr_flux between the MUSE and model-predicted spectra, to represent the most physical fitting.

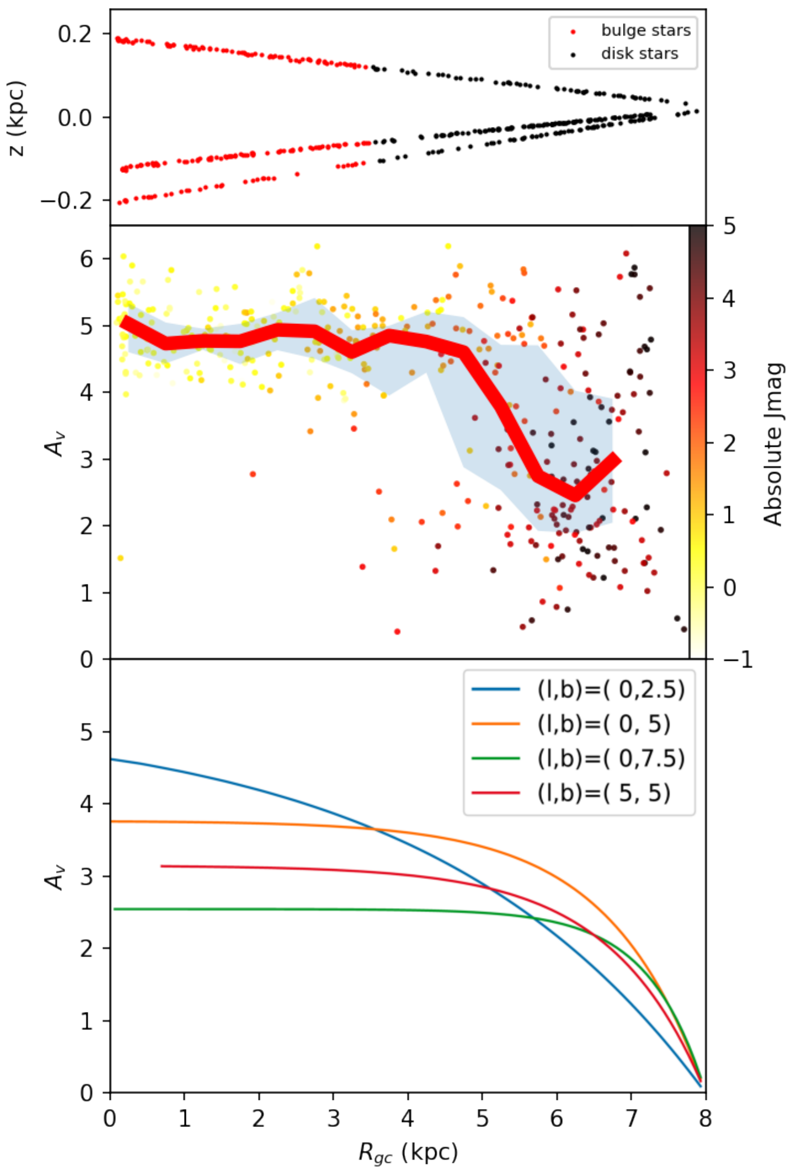

Figure 2.11: Top: Spatial distribution in cylindrical Galactocentric coordinates ($R_{\rm gc}$,z) of stars observed in three nine bulge fields. The distance to the Galactic center is adopted as $R_\odot$= 8.2 kpc, $z_\odot$ = 25 pc, same as Figure 2.10. The red dots are stars with $R_{\rm gc}$< 3.5kpc. **Middle:** Extinction-$R_{\rm gc}$ distribution of all-stars with good fitting to the model defined by Equation 2.6 in dots color-coded by absolute J mag. The red curve shows the median value of each distance bin and the blue fill shows the standard deviation of each bin. **Bottom:** Extinction A_V as a function of Galactocentric radius R for various lines of sight from the Sun. The extinction is based on an analytical model from Sharma et al. 2011 consisting of an exponential disc with warp and flare; scale length $R_{\rm d}$ = 4.2 kpc and scale height 0.88 kpc. From lines towards the center, we could see that with the larger line of sight latitude, the extinction will become a constant more quickly, which is in agreement with the left bottom plot.

Distance estimation and its relationship with extinction

To discriminate bulge stars from other foreground stars, we make use of stellar distance. The distance was estimated by BSTEP (Bayesian stellar labels estimator) (Sharma et al., 2018), a Bayesian scheme that is used for predicting stellar ages, distances, and extinction from a set of observables given by

$$\mathbf{y} = (l, b, m_J, m_{Ks}, T_{\rm eff}, \log g, [\mathrm{Fe/H}]_{\rm eff}) , \tag{2.7}$$

Briefly speaking, for each star, if we know its position, $T_{\rm eff}$, $\log g$, [Fe/H] and magnitude, its intrinsic parameters such as distance, age, and extinction can be derived by making use of isochrones (Sharma, 2017). The isochrones employed in this study are from Bressan et al. 2012. For our catalog, parameters of $T_{\rm eff}$, $\log g$, [Fe/H] are determined by *DD-Payne*. As for m_J, m_{Ks}, they are derived from VVV DR2 catalog.

Figure 2.10 shows the H-R diagram of stars with model spectra with good fitting to the model defined by Equation 2.6 and color-coded by Galactocentric cylindrical radial distance $R_{\rm gc}$ and extinction A_V. We identify bulge stars with $R_{\rm gc}$< 3.5kpc. The left panel shows that bulge stars are predominantly giants. The right panel shows that bulge stars have extinction A_V in the range 4-6 mag. Note, there is a selection bias in the sense that low extinction stars are more likely to be detected. In fact, extinction towards the Galactic center can be as high as $A_V = 50$ mag (Nataf et al., 2016).

In Figure 2.11 we explore the extinction in more detail. The top panel shows the $(R_{\rm gc}, z)$ distribution of stars in nine MUSE cubes. The line of sight of three fields can be seen. Stars with distances beyond the Galactic center have low temperature ($T_{\rm eff}$ < 4300 K) and surface gravity ($\log g$ < 1), which fall outside of the label ranges for which we can reliably determine α abundances, so these stars were removed. The middle panel of Figure 2.11 shows a scatter plot of extinction and $R_{\rm gc}$ color-coded by absolute magnitude in J band. The median value of extinction is shown by the red line, along with 25 and 75 percentile dispersion as a shaded blue region. This figure demonstrates that MUSE can successfully extract bulge giant stars with extinction up to $A_V = 6$ mag. To understand the trend of extinction with $R_{\rm gc}$, in the bottom panel of Figure 2.11 we show extinction A_V as a function of Galactocentric radius $R_{\rm gc}$ for various lines of sight from the Sun as predicted by a theoretical model in *Galaxia* (Sharma et al., 2011) consisting of an exponential disc with warp and flare; scale length $R_{\rm d} = 4.2$ kpc and scale height 0.88 kpc. It can be seen that if the latitude $|b|$ is greater than 5 degrees, the extinction first increases as we go towards smaller R, but then flattens and reaches a plateau at distances 1-2 kpc away from the sun. A similar trend is seen in the observed data in the middle panel. This flattening is due to the finite height of the dusty disc responsible for the extinction. For a given line of sight as long as we are in the dusty disc, the extinction increases with distance, but once we are out of the dusty disc there is no further increase in extinction. The model qualitatively explains the trend seen in the observed data, but there are differences. The extinction for $b = 5$ degrees is not high enough compared to our observations. For lower $|b|$ extinction can be high but the rise with distance is much slower. These differences suggest that the scale height of the disc in the model is too high.

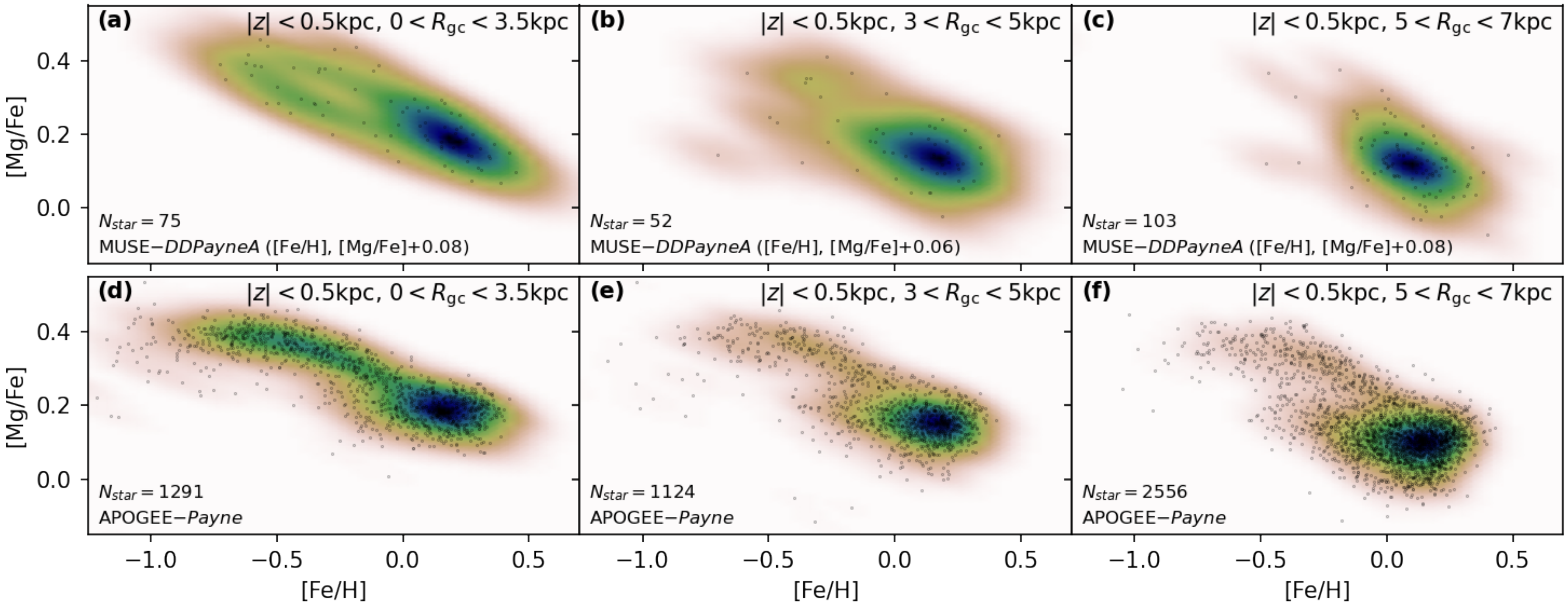

Figure 2.12: Distribution of stars satisfying $|z| < 0.5$ kpc in the ([Mg/Fe]-[Fe/H]) plane as a function of R_{gc}, with R_{gc} increasing from left to right. **Top**: The same distribution of stars in MUSE datacubes towards the Galactic bulge with abundances measured by *DD-Payne-A* model (hereafter MUSE-*DD-Payne-A*). **Bottom**: The [Mg/Fe]-[Fe/H] distribution of stars in APOGEE-*Payne* (Ting et al., 2019). The labels are calibrated by adding an offset for [Mg/Fe], which is due the [Mg/Fe]-T_{eff} trend inherited from APOGEE-*Payne* (see details in Appendix A), respectively. In each panel, the black dots mark stars in this location and the distribution is smoothed by a Gaussian kernel function and then used to plot the color maps. For each R_{gc} bin, the APOGEE-*Payne* and MUSE-*DD-Payne-A* distributions are in agreement with the densest peak at the same location. This validates the precision of the MUSE results in dense regions.

[Mg/Fe]-[Fe/H] **distribution in the Galactic bulge**

We next analyze the abundance distribution of stars towards the bulge. We make some additional quality selections to guarantee all the stars have precisely estimated parameters:

$$\begin{cases} [\mathrm{Fe/H}] > -1.5\mathrm{dex} \\ T_{\mathrm{eff}} > 4300\mathrm{K} \\ [\mathrm{Mg/Fe}]\_\mathrm{gradcorr} > 0.65 \\ \chi^2 < 20 \end{cases} \tag{2.8}$$

Here stars with [Fe/H]< -1.5 dex and $T_{\mathrm{eff}}< 4300$K are removed as it has been demonstrated in X19 that [Mg/Fe] for these giants are erroneous with low gradient spectra correlation coefficients (see their Figure 2). Moreover, we apply a cut in the $\chi^2 < 20$ to remove stars with poor fittings from *DD-Payne* according to the χ^2 distribution of the LAMOST-MUSE cross-validation sample. The removed stars are primarily giants with low T_{eff} ($<$ 4300 K) and $\log g$ ($<$ 1.5 dex), which shows low correlation in gradient spectra of *DD-Payne* (see Figure 2 of X19). We also limit our analysis to stars with correlation coefficient gradient spectra for [Mg/Fe] to larger than 0.65. After applying all the selection criteria, we end up with 75 stars physically located in the Galactic bulge.

Panel (a) and (d) of Figure 2.12 show the [Mg/Fe]-[Fe/H] distribution of stars in MUSE datacubes towards the Galactic bulge with values measured by *DD-Payne-A* model (hereafter MUSE-*DD-Payne-A*) with comparison of the same plot using stars from APOGEE-*Payne* (Ting et al., 2019). We use the Gaussian kernel estimation and plot the density color map. The selection criteria of APOGEE stars is the same as that in Rojas-Arriagada et al. 2019 and the distances were estimated by Santiago et al. 2016. [Mg/Fe] in MUSE-*DD-Payne-A* are calibrated by adding an offset. This is because there is a non-negligible [Mg/Fe]-T_{eff} trend in *DD-Payne-A* model, which is inherited from the APOGEE-*Payne* values in the training set (See Figure 14 in X19). The average temperature differences between stars in APOGEE-*Payne* and MUSE-*DD-Payne-A* in the bulge will lead to roughly 0.08 dex deviation in [Mg/Fe]. So we calibrate this deviation here. This trend is beyond the scope of this study, and more details will be discussed in the Appendix A. In Figure 2.12, both panel (a) and (d) demonstrate two sequences of stars in the bulge: Metal-poor/α-rich sequence with center in ([Mg/Fe], [Fe/H]) space at $(0.36, -0.43)$ and metal-rich/α-poor sequence at $(0.19, 0.18)$. These two centers are in agreement in MUSE-*DD-Payne-A* and APOGEE-*Payne* samples. In addition, the two sequences merge at [Fe/H]~ -0.07 dex in both panels. All these agreements reassure and validate the precision of the MUSE results, and the ability of *DD-Payne* to measure stellar labels in dense fields.

However, The two plots also show differences. The α-rich stars in the MUSE-*DD-Payne-A* are less dense than those in APOGEE-*Payne* and the distribution is also broader. Therefore, the small sample size of MUSE is certainly a limiting factor in comparing the two panels. Another important factor that is different in the two panels is the distribution of stars in the (R_{gc}, z) plane (See the top panel of Figure 2.11), neither is the apparent magnitude selection. MUSE data is restricted to 3 lines of sight, which are all close to the plane.

[Mg/Fe]-[Fe/H] distribution of other regions

In the top panel of Figure 2.11, we can see that in addition to the bulge stars, there are many foreground stars observed in these nine fields. We can also study the [Mg/Fe]-[Fe/H] distribution of these stars and compare them with distributions in APOGEE-*Payne*. Here, we select the stars with R_{gc} between $3 \sim 5$ and $5 \sim 7$ kpc, and plot the [Mg/Fe]-[Fe/H] distributions in the panel (b), (e), (c), (f) of Figure 2.12, respectively. This is similar to Hayden et al. 2015; Sharma et al. 2021b, which used stars in APOGEE DR12 and plotted [α/Fe]-[Fe/H] distributions in different Galactic locations. Here the [Fe/H] and [Mg/Fe] values in MUSE-*DD-Payne-A* are calibrated for the same reason as before. In this figure, for the same R_{gc}, the distributions of stars in MUSE-*DD-Payne-A* and APOGEE-*Payne* samples are in good agreement, with the centers at the same ([Mg/Fe], [Fe/H]) locations. All the panels are dominated by low-α stars, and the overall trends are consistent. Despite the small number of MUSE stars (52 in panel (e) and 103 in panel (f)), the good agreement with APOGEE-*Payne* is once again reassuring and validating the precision of the MUSE results.

2.5 Estimating exposure times to achieve a given uncertainty

The uncertainty σ_X of a stellar label X derived with MUSE using the *DD-Payne* depends on the signal to noise ratio (SNR hereafter) of the spectra, which in turn depends on the exposure time T_{exp} and the magnitude of a star. The brighter the star and longer exposure time, the higher the SNR and thus the more precise the stellar labels will be. Therefore, when doing a survey or planning future observations, one should ensure that the exposure time is long enough to reach the expected uncertainty. In the next sections, we provide formulas to estimate the σ_X as a function of V band magnitude and exposure time, to provide a reference for future MUSE observations. To do this, we first estimate the σ_X as a function of SNR. Next, we estimate the SNR as a function of magnitude V and T_{exp}.

2.5.1 Uncertainty of stellar labels as a function of SNR

Figure 2.13 shows the uncertainty σ_X of a stellar label X as a function of SNR. We firstly derive theoretical uncertainty $\sigma_{X,\mathrm{Theoretical}}$ using *Chem-I-Calc* (Sandford et al., 2020) from a typical K-giant spectra. SNR is estimated from MUSE Exposure Time Calculator[5]. The air-mass and seeing are set as 1.5 and 0.8″. Then σ_X is calculated by

$$\sigma_X = \sigma_{X,\mathrm{Theoretical}}(\mathrm{SNR}) f_{X,\mathrm{corr}}, \tag{2.9}$$

where $f_{X,\mathrm{corr}}$ is the scale factor applied to transform the theoretical uncertainty to the observed uncertainty for each label X. The factor is calculated from MUSE repeat observations by computing the dispersion of pairwise differences as follows.

$$f_{X,\mathrm{corr}} = \text{std-dev}\left(\frac{X_m - X_n}{\sigma^2_{X,\mathrm{Theoretical},m} + \sigma^2_{X,\mathrm{Theoretical},n}}\right), \tag{2.10}$$

[5] https://www.eso.org/observing/etc/bin/gen/form?INS.NAME=MUSE+INS.MODE=swspectr

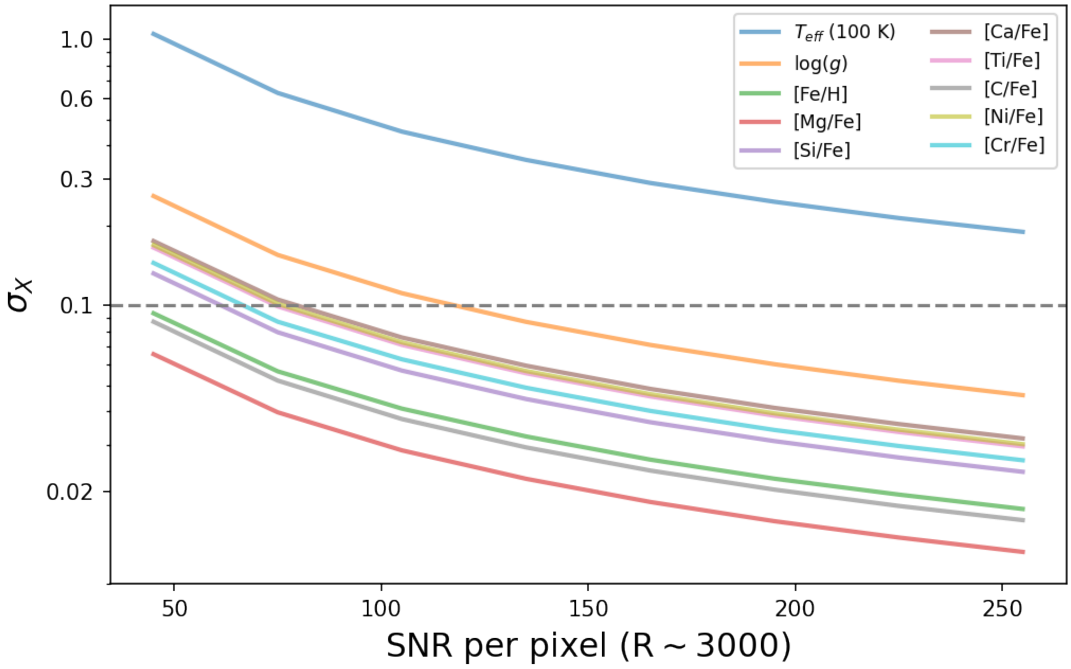

Figure 2.13: Prediction of precision of stellar labels as a function of SNR. It is derived by estimating the theoretical precision using *Chem-I-Calc* from typical K-giant spectra from MUSE Exposure Time Calculator in different SNRs. Then a factor is applied to transfer the theoretical precision to real precision for each label which is calculated by Equation 2.10.

Table 2.3: Parameters for the equation 2.11 of uncertainty as a function of SNR.

stellar label	A_X	B_X	stellar label	A_X	B_X
$T_{\rm eff}$	-0.9875	1.652	[Ca/Fe]	-0.9834	0.868
$\log g$	-0.9922	1.052	[Ti/Fe]	-0.991	0.856
[Fe/H]	-0.9746	0.5823	[C/Fe]	-0.9891	0.5737
[Mg/Fe]	-0.9862	0.4486	[Ni/Fe]	-0.9925	0.8681
[Si/Fe]	-0.9886	0.7546	[Cr/Fe]	-0.9835	0.7855

where m and n represent the repeated observations of a star. $\sigma_{X,\rm Theoretical}$ is the theoretical uncertainty of a spectra as given by *Chem-I-Calc*. Comparing with the error from *DD-Payne*, *Chem-I-Calc* estimates are normally 1.5 times smaller (see Fig. 21 in Sandford et al. 2020), so we multiply this factor. Then $\sigma_{X,\rm Theoretical}$ mentioned in this section means the theoretical uncertainty of multiplying 1.5. Figure 2.13 indicates that the uncertainty will decrease as the SNR increases. Generally, [Mg/Fe] has the lowest uncertainty with $\sigma_{\rm [Fe/H]} < 0.1$ dex for SNR higher than 29.44 pix^{-1}. As for other elements, to make the uncertainty less than 0.1 dex, the SNR needs to be higher than ~39.00, 42.02, 59.54, 79.35, 74.62, 76.24, 65.39 pix^{-1} for [C/Fe], [Fe/H], [Si/Fe], [Ca/Fe], [Ti/Fe], [Ni/Fe] and [Cr/Fe], respectively. To facilitate the calculation of the SNR required to achieve a given uncertainty, we provide analytical fits to the uncertainty curves shown in Figure 2.13. The curves are fitted with the functional form of

$$\log \sigma_X = A_X \log(\mathrm{SNR}) + B_X, \tag{2.11}$$

The coefficients A_X and B_X for each label are listed in Table 2.3.

2.5.2 SNR as a function of magnitude and exposure time

Using equation 2.11, with the known SNR we can obtain the uncertainty of stellar labels from its spectra. We know SNR depends on magnitude V, exposure time $T_{\rm exp}$, and air-mass. Therefore, when making science goals, what we care about is how long the exposure time is needed to achieve the required label uncertainty for stars with different magnitudes. Here we provide an analytical function to calculate SNR with given V magnitude and exposure time $T_{\rm exp}$ as

$$\mathrm{SNR} = \mathrm{SNR}_{\rm 1hr}(V)\sqrt{(T_{\rm exp}/3600\mathrm{s})}. \tag{2.12}$$

Here $\mathrm{SNR}_{\rm 1hr}(V)$ is the SNR estimated from the K-giant spectra by the MUSE Exposure Time Calculator for an exposure of one hour. The analytical function is given by

$$\mathrm{SNR}_{\rm 1hr}(V) = \frac{462.34 \times 10^{0.4(14.83-V)}}{\sqrt{0.00223 + 10^{0.4(14.83-V)}}}, \tag{2.13}$$

This idea is taken from Sharma et al. 2018. For the above equations, we set the air-mass as 1.5 and seeing as 0.8″. Therefore, by combining equations 2.11-2.13, with a given magnitude V and exposure time $T_{\rm exp}$, one can predict the uncertainty of each label.

Figure 2.14 provides the SNR as a function of V magnitude for an one-hour exposure (purple line). We also plot the threshold of maximum magnitude for each label where the

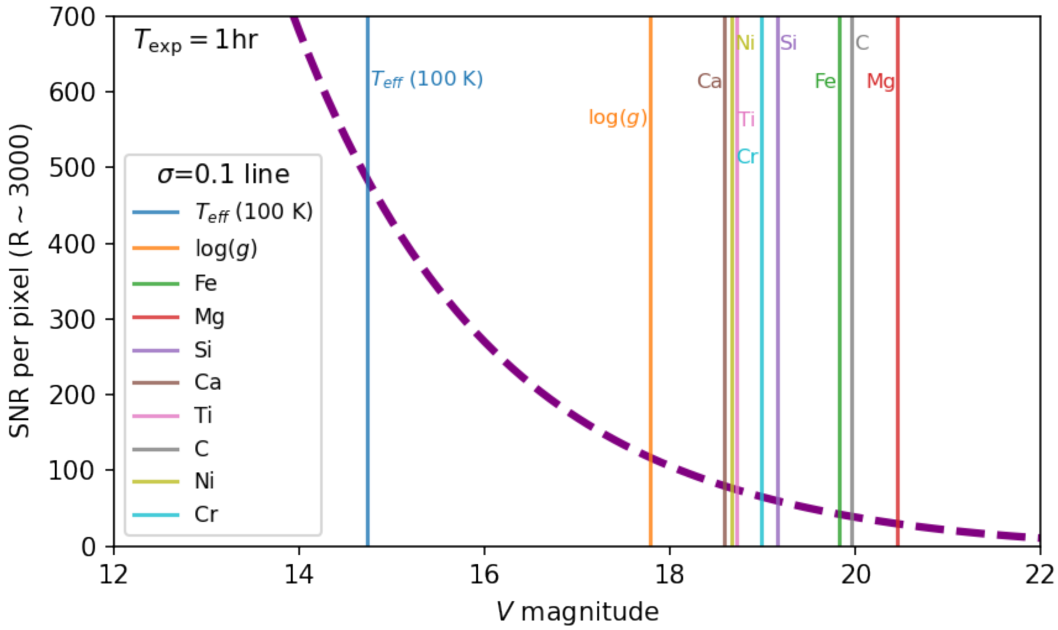

Figure 2.14: SNR of MUSE spectra at $R \sim 3000$ as a function of V magnitude (purple line) for an exposure time of one hour. Each colored line represents the threshold where the uncertainty of a label reaches 0.1 dex (or 10 K), which is derived using equation 2.11. All chemical abundances can be derived with uncertainty less than 0.1 dex till V = 18.5 mag, for [Mg/Fe], [C/Fe] and [Fe/H] even till V = 19.7 mag.

uncertainty reaches 0.1 dex or 10K, which are derived by Figure 2.13. As we can see, all chemical abundances can be derived with uncertainty less than 0.1 dex till V = 18.5 mag, for [Mg/Fe], [C/Fe] and [Fe/H] even till V = 19.7 mag. This figure combining with equations 2.11-2.13 can be used as a guide for future research, providing an indication of the exposure time for different magnitude stars to achieve the expected uncertainty.

2.6 Discussions

2.6.1 Performance of MUSE stellar label extraction under different scenarios

Ideally, the methods outlined in this paper to determine stellar labels would be applicable to all stars in MUSE datacubes. However, the label ranges of the *DD-Payne* training set do not cover the space for all potential observations: for stars outside of this parameter space, for example having very low metallicity ([Fe/H] < -1.5) or very high effective temperatures ($T_{\rm eff} > 7000K$), labels are determined through extrapolation of the model. Despite this, some studies have shown *DD-Payne* can do some moderate extrapolation in the linear region of the gradient spectra varying with labels. This is because *DD-Payne* imposes the gradient spectra to be similar to that of the Kurucz models (Kurucz, 1970, 1993, 2005), which is an advantage of such a hybrid method compared to pure data-driven models. For example, the LAMOST [Fe/H] estimates can be robust down to -2.5 dex (see Fig. 10 of Conroy et al. 2019); *DD-Payne* can also generate reasonable [Ba/Fe] stars with $1 < [\mathrm{Ba/Fe}] < 3$ dex and $T_{\mathrm{eff}} > 7000\mathrm{K}$ (Xiang et al., 2020).

Typically in MUSE data, stellar spectra come from the following types of fields: random individual stars near a galaxy, high extinction fields, and globular clusters. In each of these scenarios the performance of the spectroscopic analysis is different and some of them pose unique challenges for analysis.

For individual stars near a galaxy, they typically have high SNR, as observations in galaxy fields generally have long exposure times. Therefore, as long as the PSF does not overlap with the nearby galaxy, it is relatively simple to extract and estimate the stellar labels. All of the common stars between MUSE and LAMOST belong to this case; one typical spectrum is shown in Figure 2.3. For these stars, stellar labels will have high precision due to their high SNR, as well as the fact that most of them belong to the nearby disk, which are stellar populations that are well covered by the *DD-Payne* training sets.

For stars in high extinction fields, such as the bulge, the spectra are restricted to a narrower range in wavelength and the SNR of the extracted spectra is in general low (due to a combination of large distance and dust). For low SNR, in addition to SNR directly affecting precision, the emission lines are comparatively stronger and more difficult to remove. The above-listed factors lead to lower precision in stellar labels. Hence, in high extinction regions, one can only study giants which are intrinsically bright. Therefore, we can only get reliable label estimates for stars with relatively low extinction ($A_V < 6$ mag) along the bulge lines of sight, which was shown in the middle plot of Figure 2.11. As for stars of $A_V > 6$ mag in the Galactic center, the dust will make the giants difficult to observe with typical MUSE exposure times.

For globular clusters, which have high stellar density, our work will be of particular importance. It will provide an automated way to determine high precision stellar labels for these dense fields down to the center of the cluster; previous studies using fiber-fed spectrographs were often limited to the outskirts of these objects. Additionally, it is possible to increase the precision of stellar parameters and several abundance estimations by making use of full wavelength fitting, which has not been done so far. Previously, the MUSE GC survey (Husser et al., 2016; Kamann et al., 2016) observed 26 globular star clusters, with most of these studies focusing on specific features of spectra or just metallicity, e.g. emission-line studies (Göttgens et al., 2019), the dynamics using CaT absorption lines (Kamann et al., 2018), the relative abundance variations using the equivalent widths (Latour et al., 2019), metallicity distributions (Husser et al., 2020; Latour et al., 2021). Our method has the potential to enable more chemical abundances measured automatically by the full-wavelength fitting frameworks, rather than applying any spectral library to execute the fittings. Even though globular clusters typically have $[\mathrm{Fe/H}] < -1.0$ dex, some of them are still in the parameter range of *DD-Payne*, and the moderate linear extrapolation ability of *DD-Payne* still enables this method to measure robust stellar labels. We will perform the test of it in the future work.

2.6.2 Computational challenges in spectral extraction

In this work, the spectral extraction processes are executed using PampelMUSE (Kamann et al., 2013). The typical time spent on extracting spectra from each datacube with a 24-core CPU and 32GB RAM is ~ 30 minutes and it is independent of the number of stars in the field. Therefore, for the verification data analyzed in this paper (the cross-match of LAMOST and MUSE), we spent more time on spectral extraction for these cubes compared to the datacubes in the bulge, even though there were two orders of magnitude more spectra extracted from the bulge fields. Hence, to perform a survey of all individual stars in publicly available MUSE datacubes, the most time-consuming step will be the extraction of spectra.

For individual stars in sparse fields or more limited studies focused purely on kinematics in dense fields, it is also possible to use other stellar extraction strategies such as aperture photometry in IRAF for each layer subtracted by the sky estimated locally (e.g. Valenti et al. 2018). This approach has the advantage of retaining as many stars as possible and has a more accurate estimate of the sky background, which will lead to reduced artifacts in the final spectra (Valenti et al., 2018). For regions with modest crowding (i.e., < 200 stars per field), aperture photometry and PSF-fitting could yield similar results. However, PSF-fitting is still necessary when dealing with more crowded fields (> 200 stars per field) as there will be significant blending for stars in these dense regions. Kamann et al. 2013 tested the performance of PampelMUSE PSF-fitting using a simulated MUSE datacube towards the center of a globular cluster 47 Tuc. Under the average seeing of 0.8 arcsec, ~5000 useful spectra could be extracted. In good conditions, this value can be 3 or 4 times larger. With severe crowding, aperture photometry will suffer from significant blending between stellar spectra. Therefore, for globular clusters, we still recommend PampelMUSE.

2.6.3 Future research

There are several aspects in the data pipeline that can be improved in the future. In the methods outlined in this paper, stellar labels are estimated from the *DD-Payne*, which depends sensitively on the precision of labels in GALAH DR2 (Buder et al., 2018) and APOGEE-*Payne* (Ting et al., 2019). Given that improved versions of both these catalogs are now available, GALAH DR3 (Buder et al., 2021) and APOGEE DR16 (Ahumada et al., 2020), one needs to update the *DD-Payne* models. In addition, we can train a new *DD-Payne* model if a sufficient number of common stars in MUSE and other high-resolution spectroscopic surveys for a training set is available. The advantage is that there will be no need to degrade the MUSE spectra to a lower resolution, which will help extraction of blended features. In addition, some absorption lines such as [Na/Fe] and [O/Fe] are masked in the LAMOST spectra due to dichroic issues in the wavelength range (5803 – 5966Å). This range can now be used when working in MUSE non-AO mode. It will enable to estimate precise sodium and oxygen abundances, which is essential in identifying different stellar populations in globular clusters (Bastian & Lardo, 2018). The primary difficulty for retraining is having a training set that is both large and covers a wide range of parameter space.

In the future, the method outlined here can be utilized for many purposes. A survey can be executed based on stars from publicly available MUSE datacubes using *Gaia* eDR3 (Gaia

Collaboration et al., 2021) and SkyMapper (Keller et al., 2007) as photometric input catalogs. Our method can be applied to estimate their stellar labels and compile a catalog that will be useful for Galactic archaeology studies. For the Galactic bulge, our method has the opportunity to study the chemistry and dynamics of the Galactic center in greater detail. However, more observational data is needed, such as fields in Baade's window, which are less affected by the dust. In addition, this method can be applied to globular clusters observed by Husser et al. 2016; Kamann et al. 2016 to study their chemistry, and identify multiple stellar populations if any, and their connection to kinematics. Moreover, the interest in using MUSE for spectroscopy of resolved stellar populations has been growing quite strongly, even beyond the limits of the Galaxy (e.g. Roth et al. 2019), our method can also provide parameter estimations for these targets.

With the launch of the BlueMUSE project (Richard et al., 2019), we will obtain spectra in the wavelength range between 3500 − 5800Å, which is also covered by LAMOST. As mentioned previously, the loss of the blue portion of the spectrum in MUSE is the primary reason why we are not able to estimate as many precise chemical elements as those determined from LAMOST spectra. BlueMUSE will provide the opportunity to measure these abundances with higher precision, and for more elements that are currently not available; critically the s-process elements such as [Y/Fe] and [Ba/Fe] which provide key age diagnostics for stellar targets (Hayden et al., 2022). The blue portion will also bring strong absorption features for determining precise abundances of Na, C, N, and O, which are essential for identifying multiple populations in globular clusters. In addition, MAVIS (McDermid et al., 2020), which is the new IFS instrument being built for ESO's VLT AOF (Adaptive Optics Facility, UT4 Yepun) also covers a wavelength range 3700 − 10000Å similar to BlueMUSE and MUSE. With the lower spatial sampling of $0.02'' - 0.05''$, we have the opportunity to study the core of globular clusters with more stars providing an insight into its chemistry and dynamics. Because of the similar wavelength coverage of *DD-Payne* and these instruments, our method can be directly applied to them.

Finally, based on the fact that the study of resolved stellar populations represents one of the major science cases for the European Extremely Large Telescope (ELT) (Gilmozzi & Spyromilio, 2007; Hook, 2009), it is reassuring that our method is paving the way for future studies in multi-dimensional chemical space. For the Galactic bulge, ELT will allow us to measure chemical abundances all the way down to the main sequence turn-off stars (Minniti & Zoccali, 2008). For these targets, we can use BSTEP (Sharma et al., 2018) to map the age distribution and reveal the detailed formation and evolution history of the innermost bulge, the peanut bar, and the long bar.

For Local Group galaxies, the ELT can also derive more stellar spectra of bright main-sequence (potentially even turn-off) stars and giants with necessary SNR than VLT (e.g. McLeod et al. 2019; Carini et al. 2020); this includes the SMC and LMC, as well as more distant galaxies such as Andromeda and M33, extending Galactic Archaeology beyond the Milky Way to the Local Group. With the unprecedented detection sensitivity of ELT, our method can provide a way for future stellar label estimation and science goals for the ELT-related IFS instruments, to measure stellar labels of stars beyond the Local Group (Wyse & Gilmore, 2006; Evans, 2008; Greggio et al., 2012).

2.7 Summary

In this study, we applied *DD-Payne* (Xiang et al., 2019), which was developed for LAMOST spectra, on MUSE datacubes and estimated $T_{\rm eff}$, $\log g$, [Fe/H] and chemical abundances of Mg, Si, Ti, Ca for individual stars in MUSE datacubes.

By comparing our MUSE results with that of LAMOST using common stars, we show that despite instrumental differences, it is possible to extract precise stellar parameters and abundances with typical dispersion of $T_{\rm eff}$, $\log g$ and other chemical abundances less than 75K and 0.15 dex and 0.1 dex, respectively.

In dense fields, which is the unique advantage of the MUSE instrument, we selected 29 datacubes towards the Galactic center. Based on previous studies (Rojas-Arriagada et al., 2019; Hayden et al., 2015; Sharma et al., 2021b), by comparing the [Fe/H]-[Mg/Fe] distribution estimated by our method in the bulge and inner disk region with stars from APOGEE-*Payne*, we found excellent agreements in two sequences with similar center values and overall trends. This indicates the ability of our method to measure chemical abundances in regions with high stellar density and extinction. In addition, we also studies the extinction as a function of $R_{\rm gc}$ in the direction of the Galactic center and compared these results with the prediction from *Galaxia* (Sharma et al., 2011). We found there is a qualitative agreement with the overall trends, but the scale height of the dust predicted by *Galaxia* is higher than observations.

In addition, we provided analytical formulas to predict the precision of each label as a function of V magnitude and exposure time. This can be used for observational proposal designing before a survey to achieve the expected label uncertainty.

In the future, this method can be applied to do a survey for all individual stars in public MUSE cubes and estimate their stellar labels. The result can be a supplementary catalog for all-sky spectroscopic surveys. For the Galactic bulge, more observations are needed to expand the sample size. By combining the chemical and kinematic information, there is a great opportunity to study the evolution and assembly history of the Galactic center. This method can also be applied to MUSE observations on globular clusters, to study the connection between kinematics and multiple populations.

This is the first time to measure stellar labels using full wavelength fitting in dense fields observed by the MUSE instrument. We demonstrate that despite PSF/LSF being different, it is possible to apply a model trained on spectra from one instrument to another. Given a large number of spectrographs having wavelength similar to LAMOST, our results suggest that one can easily do detailed spectroscopic analysis with them. In the future, this method can be used directly for BlueMUSE (Richard et al., 2019) and MAVIS (McDermid et al., 2020). It will also help with the label measurement development for IFS instruments on ELT (Gilmozzi & Spyromilio, 2007), to study Galactic archaeology from the Milky Way to somewhere more than 10 Mpc away.

Acknowledgements

We thank Xu Zhang, David M. Nataf, Thorsten Tepper Garcia, Sven Buder, and Jesse Van de Sande for their very useful comments.

This research has made use of the services of the ESO Science Archive Facility. The results are based on public data released from the MUSE commissioning observations at the VLT Yepun (UT4) telescope. This work has also made use of the data from LAMOST. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences. This work was also based on data products from VVV Survey observations made with the VISTA telescope at the ESO Paranal Observatory under programme ID 179.B-2002.

This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France (DOI: 10.26093/cds/vizier). The original description of the VizieR service was published in Bonnarel et al. 2000; Astropy[6], a community-developed core Python package for Astronomy (Astropy Collaboration et al., 2013, 2018), numpy (Harris et al., 2020), scipy (Virtanen et al., 2020), matplotlib (Hunter, 2007), ZAP (Soto et al., 2016), PampelMUSE (Kamann et al., 2013), DD-Payne (Xiang et al., 2019) and MPDAF (Piqueras et al., 2017).

We acknowledge the University of Sydney HPC service at The University of Sydney for providing HPC resources that have contributed to the research results reported in this paper.

ZW is supported by the China Scholarship Council and Australian Research Council Centre of Excellence for All Sky Astrophysics in Three Dimensions (ASTRO-3D) through project number CE170100013. YST is grateful to be supported by the NASA Hubble Fellowship grant HST-HF2-51425.001 awarded by the Space Telescope Science Institute. YST acknowledges financial support from the Australian Research Council through DECRA Fellowship DE220101520.

Data Availability

The data underlying this article will be shared on reasonable request to the corresponding author.

[6] http://www.astropy.org

Chapter 3

A Deep MUSE View towards Baade's Window

ZIXIAN WANG,[1,2][1] SANJIB SHARMA,[1,2] MICHAEL R. HAYDEN,[1,2] JOSS BLAND-HAWTHORN[1,2,3]

[1]Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW, 2006, Australia
[2]ARC Centre of Excellence for All Sky Astrophysics in Three Dimensions (ASTRO-3D)
[3]Miller Professor, Miller Institute, UC Berkeley, Berkeley CA 94720

This Chapter presents data reduction and early results of MUSE observation towards the Galactic bulge in ESO Period 109 (Proposal PI: Zixian Wang, co-Is: Michael Hayden, Sanjib Sharma, and Joss Bland-Hawthorn). I am the PI of this MUSE proposal and wrote all the text, made all the figures and produced most of the data reduction procedures and figures. Sanjib Sharma helped to plot Fig 3.2 and produced GALAXIA simulated catalog and BSTEP results to measure stellar age, distance, and extinction. Joss Bland-Hawthorn, Sanjib Sharma, and Michael Hayden supervised the whole work.

Current studies of the Galactic bulge are limited to the chemodynamical properties of stars but typically lack precise age information. And previous studies of the age distributions in the bulge have produced contradictory results, primarily due to a lack of reliable age determinations as existing studies can often only target giant stars due to magnitude and extinction limits. We study the age, chemistry, and dynamics of stars from the inner Galactic bulge by executing a 7.2-hour MUSE observation towards one field in the Baade's window. This area has low extinction, with $A_V = 1.45$ mag, allowing MUSE to reach all the way to the Galactic center with main-sequence turn-off and subgiant branch (MSTO-SGB) stars to obtain an accurate age census of the Galactic bulge. After spectra extraction and parameter estimation, we obtained 138 giant and 159 MSTO-SGB stars with $\mathrm{SNR_{MUSE}} > 80\ \mathrm{pix}^{-1}$ from a 3.6-hour exposure. We lost 66 giants and 135 MSTO-SGB stars due to incomplete observation. We find that the V-band magnitude distribution of our sample is consistent with predictions of a simulated catalog by GALAXIA when $m_V < 19.3$, and the $E(B-V)$ - distance relation agrees with the 3D dust map Bayestar19. We apply GMM modelling and separate the [Fe/H] distribution into different components, with the peaks consistent with previous studies within 1σ except for the most metal-poor one. We find the age distribution revealed by MSTO-SGB stars within $R_{\mathrm{gc}} < 3.5$ kpc has three main groups with ages around $10 \sim 12$ Gyr, $7 \sim 9$ Gyr, and $4 \sim 6$ Gyr, respectively. The

$[\alpha/\mathrm{Fe}]$-[Fe/H] distribution of stars in the bulge indicates a "knee" at [Fe/H] = −0.5 dex, which agrees with the chemical evolution model of Sharma et al. (2021b). But the existence of metal-rich stars with [Fe/H] $\sim$ 0.5 dex indicates further improvements on the model. In the future, we propose to observe more low-extinction fields and obtain a larger sample of stars to address the age and chemodynamical distributions of the Galactic bulge in more detail.

3.1 Introduction

The formation and evolution of the Milky Way is one of the outstanding questions facing astrophysics today, and Galactic Archaeology aims to unravel this using stars as our fossil record (Bland-Hawthorn & Gerhard, 2016). Stars contain the chemical imprint of the gas from which they formed, allowing the enrichment history of the Galaxy to be traced through time and space (e.g. Xiang & Rix 2022). The Galactic bulge is one of the fundamental building blocks of the Galaxy but has a complicated structure and formation history. The bulge contains some of the oldest stars born in the Galaxy (e.g., Zoccali et al. 2003; Grieco et al. 2012), but a significant fraction of young stars are also observed (e.g., Buell 2013; Gesicki et al. 2014; Bensby et al. 2013). Moreover, our knowledge of the very precise stellar age distribution and the dynamic properties on the far side of the bulge is limited due to a combination of the large distance of the bulge, high extinction along its line-of-sight, stellar crowding, and smaller telescope diameters (< 4m) of the current fiber-fed multi-object spectroscopic facilities.

During the previous decade(s), many studies using different data have reached a consensus on the morphology, metallicity, and kinematics of the bulge as a function of Galactic coordinates, such as surveys like BRAVA (Kunder et al., 2012), ARGOS (Ness et al., 2013a,b), GIRAFFE (Zoccali et al., 2014, 2017), *Gaia*-ESO (Rojas-Arriagada et al., 2014, 2017), and APOGEE (Rojas-Arriagada et al., 2019; Hasselquist et al., 2020). The bulge hosts a boxy/peanut (B/P) shape structure and is represented by multiple components with different [Fe/H], where the metal-rich stars are much more abundant in the Galactic plane and bar and have a much steeper velocity dispersion gradient as a function of galactic latitude than metal-poor stars, while the metal-poor stars are distributed more spherically near the Galactic center and the velocity varies rather mildly with latitude. A more comprehensive study by combining ARGOS and APOGEE samples (A2A) is in Wylie et al. (2021). However, due to the detection limits of existing spectroscopic surveys, nearly all stars used in these measurements are on the near side of the bulge. The structure and dynamics on the far side of the bulge and inner disk have not been fully observed so far, so the stellar populations of these regions remain unexplored.

While the chemodynamics properties of the nearside of the bulge are coming into focus, accurate age distributions for the stellar populations of the bulge are still lacking, as there have been several contradicting results from different datasets (Nataf, 2016). A purely old bulge with age > 10 Gyr is supported by most of the chemical compositions or photometric age estimates (e.g., Zoccali et al. 2003; Grieco et al. 2012). However, some other studies based on planetary nebulae and microlensed sub-giant and dwarf stars suggest a non-negligible fraction of young populations at around 3 Gyr (Buell, 2013; Gesicki et al., 2014; Bensby et al., 2013). Comparable results have also been found by Schultheis et al. (2017) using the [C/N] ratio in giants from the APOGEE survey. Hasselquist et al. (2020) selected ~6000 high-luminosity ($\log g < 2.0$) and metal-rich ([Fe/H] > 0.5) stars and found at least half of them are old ($>$ 8 Gyr), while the younger stellar populations (2 − 5 Gyr) are strongly constrained to the Galactic plane. Lian et al. (2020b) quantitatively modelled the [Fe/H] and [Mg/Fe] distributions and found the fraction of young stars are subject to the estimated age uncertainty, leading to this number varying between 11% and 29%. Therefore, an accurate age estimation is essential to reveal the star formation history of the bulge.

Main-sequence turn-off (MSTO) and subgiant branch (SGB) stars have long been considered as the most accurate age proxy (e.g. Feltzing et al. 2001; Nissen 2015). Using stellar isochrones, one can measure the age of MSTO-SGB stars from the spectroscopic stellar parameters $T_{\rm eff}$, $\log g$, [Fe/H] and [Mg/Fe] (Sharma et al., 2018). This is because, for a given metallicity, the isochrones are well separated in the colour-magnitude diagram for MSTO-SGB stars (see the orange box in Fig. 3.1). Main sequence dwarfs and stars along the giant branch (area outside the orange box of Fig. 3.1) have a much smaller separation than MSTO-SGBs, making age determination nearly impossible to measure. Age estimates are also available for stars that have asteroseismic observations, or the [C/N] ratio in giants that has been calibrated using asteroseismology (Silva Aguirre et al., 2018), but these age estimates have a factor of 3 $\sim$ 4 times larger uncertainty than the ages from MSTO-SGB stars.

However, due to the large heliocentric distance of the bulge (~8 kpc) and the severe extinction along the line-of-sight, most spectroscopic observations of the bulge are limited to giants. Hence, the only study using these stars in the bulge to date is by using gravitational microlensing. Bensby et al. (2017) used high-resolution spectra of 91 microlensed MSTO stars observed by UVES (Dekker et al., 2000) and found that about 35% of the stars are metal-rich ([Fe/H] > 0) and span ages in between 8 Gyr and 2 Gyr, whereas the metal-poor populations ([Fe/H] < -0.5) are 10 Gyr or older. Note this method relies on rare events – only 91 microlensing events have been observed over more than a decade, making this approach unfeasible for obtaining more than a handful of stars. Therefore, the age distribution of stars outside of the solar neighbourhood is still poorly constrained. Additionally, crowding due to the high stellar density of the bulge makes observing with traditional fiber-fed spectrographs difficult. Typically only the brightest giants can be observed due to large fiber size combined with neighbour magnitude limits and fiber collision requirements.

In this study, we use MUSE (Bacon et al., 2010) to measure the chemical abundances, ages, and kinematics of stars in the bulge, from one field of the low extinction region of Baade's window (e.g., Stanek 1996; Gonzalez et al. 2011). We aim to obtain spectra with SNR around 80 pix^{-1} for stars with $m_V \sim 20.5$. The approach laid out by Wang et al. (2022) demonstrates the ability of MUSE to measure reliable stellar parameters for individual stars in crowded fields, probing areas of the Galaxy unreachable by traditional fiber-fed spectrographs. This, combined with a large field of view, spectral resolution and wavelength coverage, and location on a large telescope, make MUSE the ideal instrument to study the properties of the Milky Way bulge. We will obtain a large sample of MSTO-SGB stars directly in the bulge, from which we will determine the age and metallicity distribution function of this field with high precision, and compare with previous

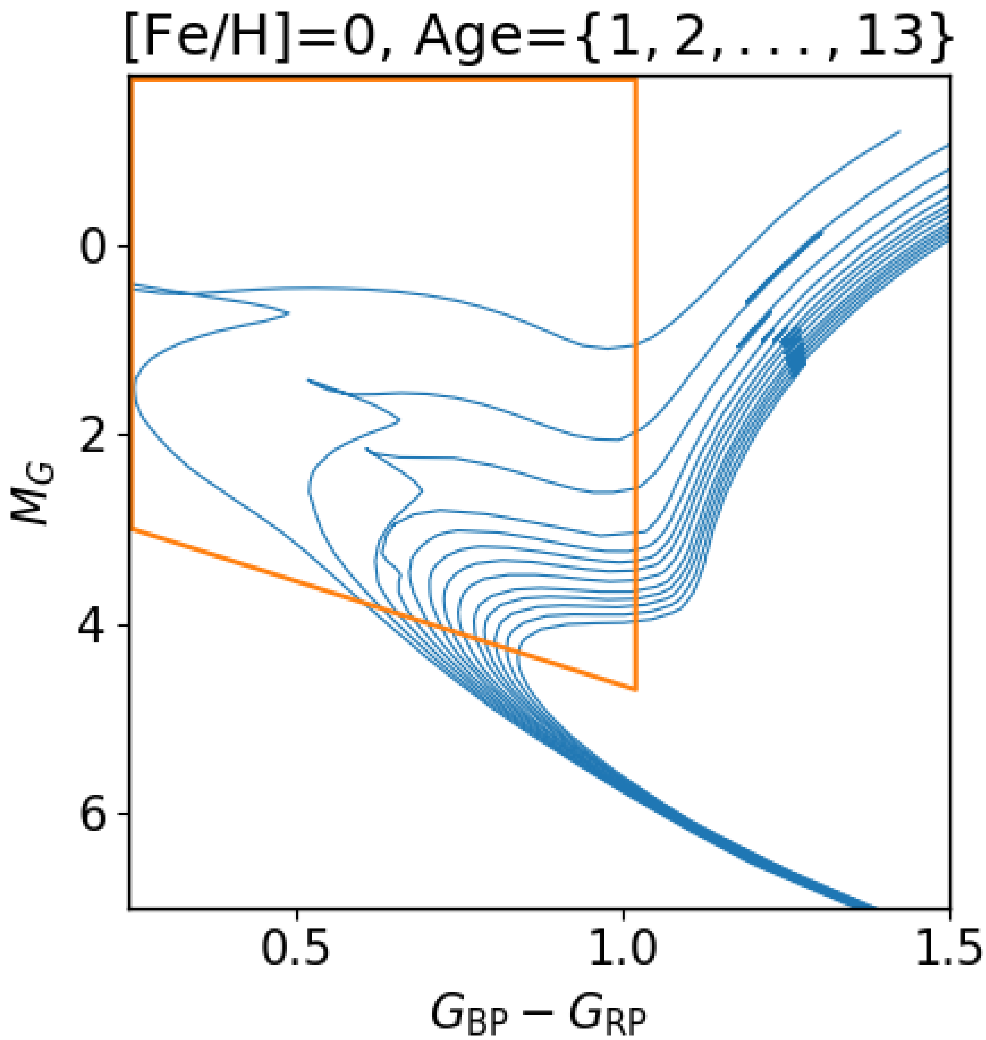

Figure 3.1: A color-magnitude (*Gaia* photometric bands) diagram using PARSEC isochrones (Bressan et al., 2012) with [Fe/H]=0 and age ranging from 1 to 13 Gyr. The region covered by the orange frame represents the MSTO-SGB stars, for which age can be measured from spectroscopic stellar parameters. This is because for the MSTO-SGB stars the isochrones have a wider separation between them than giant and main-sequence stars.

.

results.

3.2 Data and Methods

3.2.1 Field Selection and Photometric Analysis

The observed field in Baade's window is selected with the following criteria: First, we use the 3D Dust Mapping tool Bayestar19 [2] from Green et al. (2019) to generate a high-resolution extinction map of Baade's window and select regions with $A_V < 2$ mag at $R_{gc} = 8$ kpc. Second, we calculate the distance of these fields to the two globular clusters NGC 6522 and NGC 6528 and select fields with distances larger than 20 arcmins to remove GC members. Third, we remove fields having stars with $m_V > 14.5$ mag to prevent CCD saturation due to a long exposure time. The resulting fields are then matched with OGLE-III (Szymański et al., 2011) and the field with the most number of stars is selected. Finally, a field at $(l, b) \sim (1.66, -3.59)$ degrees with 1734 OGLE stars and $A_V = 1.45$ is selected.

We employ a Galaxia (Sharma et al., 2011) simulated catalog to predict the results from observations at the same (l, b) and FoV. First, we select all the Galaxia particles included in the FoV and calculate the apparent magnitude m_V using distances and $E(B-V)$, where $E(B-V)$ is predicted by 3D dust map of Bayestar19 (Green et al., 2019). Then we apply the OGLE-III selection effect by removing stars with $m_V < 21.651$, which is the maximum V-band magnitude from those 1734 OGLE stars on this field. Next, we select MSTO-SGB, giant, and dwarf stars according to the following criteria:

$$\begin{cases} \text{MSTO-SGB: } 3.7 < \log g < 4.3 \text{ and } 5100 < T_{\rm eff}/\text{K} < 6500 \\ \text{Giant: } \log g < 3.7 \\ \text{Dwarf: } \log g > 4.3 \end{cases} \tag{3.1}$$

Fig. 3.2 demonstrates the predicted stellar apparent magnitude distributions of stars in our chosen field. The top panel shows the fraction distribution of MSTO-SGB, giant, and dwarf stars as a function of m_V. This panel demonstrates that giant stars mostly have $m_V < 19.5$, and a majority of MSTO-SGB stars are in $19 < m_V < 20.5$ mag. As for dwarfs, they dominate the region of $m_V > 20.5$ mag. The bottom panel shows the number distribution of heliocentric distances for giant and MSTO-SGB stars. The center of the bulge lies at a distance of 8.2 kpc (Bland-Hawthorn & Gerhard, 2016). The thin and thick line shows the number distribution of stars with $m_V < 19.5$ and $m_V < 20.5$, respectively. This panel indicates the necessity for the observation to reach $m_V = 20.5$ mag, because then we can obtain around 10 times more MSTO-SGB samples concentrated to the Galactic center than those with $m_V < 19.5$ mag. To reach $m_V = 20.5$ mag, we apply the MUSE exposure time calculator [3] and find that under the weather condition of air-mass ~ 1.2 and seeing $\sim 0.7''$ with AO, at least 7.2 hour exposure time is required to satisfy SNR ~ 80 pix^{-1}. In this case, according to the simulated catalog, we are predicted to obtain around 395 MSTO-SGB and 117 giant stars from this field. The MSTO-SGB sample is 3.3 times larger than those from Bensby et al. (2017) over a decade of observation on microlensed events.

[2] http://argonaut.skymaps.info/

[3] https://www.eso.org/observing/etc/bin/gen/form?INS.NAME=MUSE+INS.MODE=swspectr

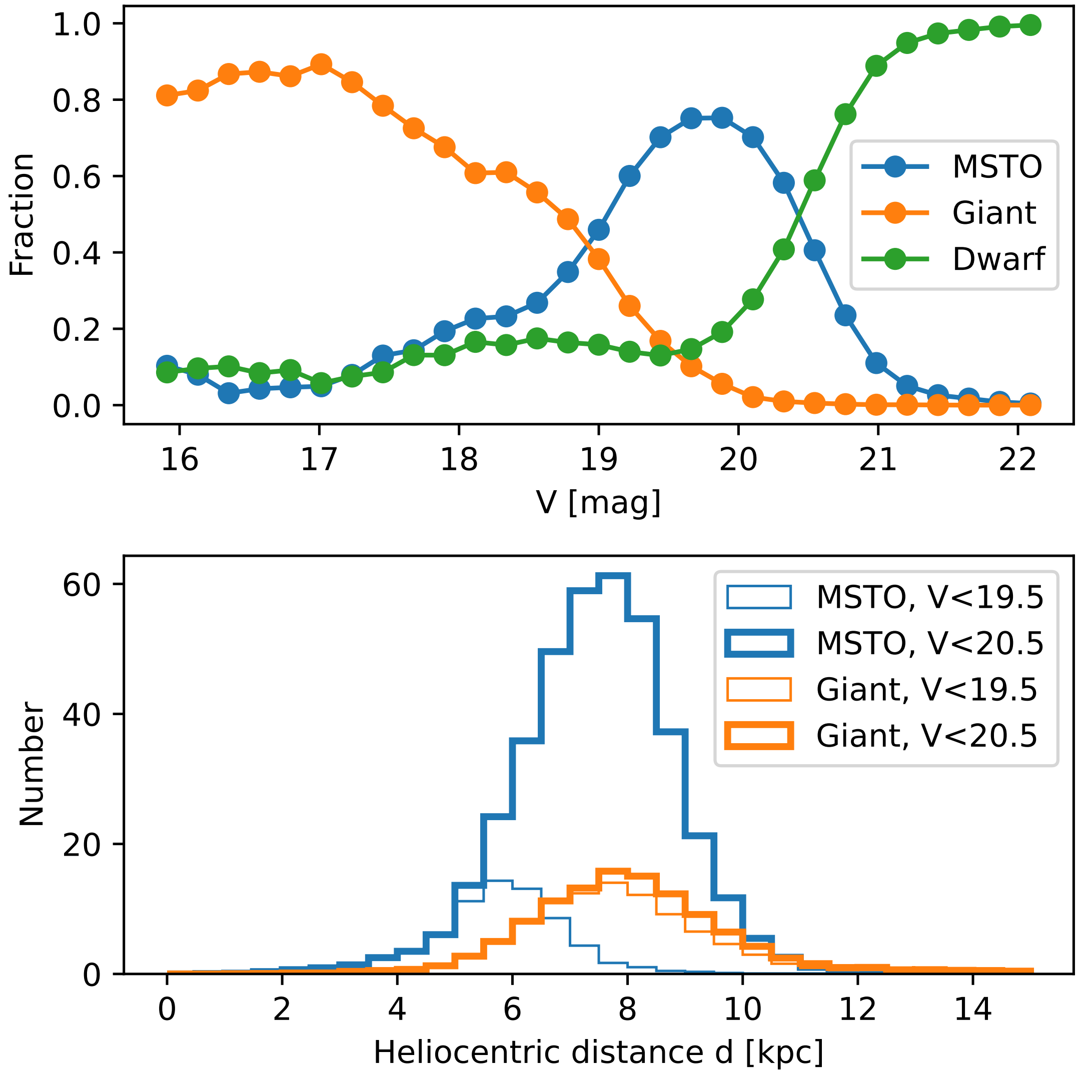

Figure 3.2: Stellar properties of stars in our chosen field based on a simulation by *Galaxia*. **Top:** The fraction MSTO-SGB and giant stars as a function of V-band magnitude. For $m_V > 19.5$ close to 23% of stars are MSTO-SGB stars. **Bottom:** The distribution of heliocentric distances for giant and MSTO stars. The center of the bulge lies at a distance of 8 kpc. With $m_V < 19.5$ (thin blue line) we do not get enough stars in the center of the bulge, but with a limit of $m_V < 20.5$ (thick blue line), we do.

.

Table 3.1: Information of MUSE data (ESO Period 109) towards Baade's window

RA	DEC	Obs Date	Exptime (s)	Airmass	Seeing at Start (″)
18:03:39.37	-29:19:20.6	2022-08-21	1293.00	1.027	0.80
18:03:39.37	-29:19:21.4	2022-08-21	1293.00	1.011	1.09
18:03:39.31	-29:19:20.6	2022-09-02	1293.00	1.049	0.48
18:03:39.31	-29:19:21.4	2022-09-02	1293.00	1.024	0.37
18:03:39.37	-29:19:21.4	2022-09-20	1293.00	1.133	0.81
18:03:39.37	-29:19:21.4	2022-09-20	1293.00	1.193	0.90
18:03:39.31	-29:19:21.4	2022-09-27	1293.00	1.103	0.99
18:03:39.31	-29:19:21.4	2022-09-27	1293.00	1.064	1.32
18:03:39.31	-29:19:20.6	2022-09-28	1293.00	1.250	0.93
18:03:39.31	-29:19:20.6	2022-09-28	1293.00	1.178	1.04

3.2.2 MUSE Observations and data reduction

We proposed a 7.2-hour MUSE observation towards the chosen BW field in Period 109 (Program ID: 109.23HS.001) and were allocated 9.8 hours including overheads. We divided the total allocated time into 10 observing blocks (OB) with each one having 1293×2 s science exposures. We applied a small dithering pattern by a few arcseconds and 90 degrees rotation between on-object exposures to average the spatial signature of the 24 integral-field units (IFU) on the FoV. Finally, 50% percent of the observations were made (equivalent to 3.6 hours) with details listed in Table 3.1.

The data reduction was performed using MUSE pipeline v2.8.7 (Weilbacher et al., 2014, 2016, 2020) under the ESOREFLEX environment (Freudling et al., 2013). The entire reduction cascade includes two main procedures: (1) calibrations of each exposure such as bias, flats, bad pixels map, instrument geometry, illumination, astrometry correction, sky subtraction, line spread function, flux, and wavelength calibration; and (2) a combination of individually processed exposures and construction of the final science data product. Specifically, during sky subtraction, we set 10% of the total spatial pixels to be considered as the sky, which is based on the flux distribution of an OGLE I-band image (Szymański et al., 2011) in the target FoV. Finally, we obtained one combined datacube with a total exposure time of 12,930 s.

3.2.3 Spectra Extraction and Parameter Estimation

We perform the spectral extraction and parameter estimation following procedures in Wang et al. (2022), also see Section 2.3. First, we employ PampelMUSE (Kamann et al., 2013) to extract the spectra by using OGLE-III source list (Szymański et al., 2011) as the input catalog. Next, we degrade the extracted spectra from MUSE spectral resolution ($R \sim 3000$) to LAMOST ($R \sim 1800$) and apply *DD-Payne* (Ting et al., 2017a; Xiang et al., 2019) to measure stellar parameters ($T_{\rm eff}$, $\log g$, [Fe/H], [Mg/Fe], [Si/Fe], [Ti/Fe], [C/Fe], [Ni/Fe] and [Cr/Fe]) by fitting the model with spectra in the wavelength region of [4700, 8750Å]. We obtain a catalog with 1349 stars measured by *DD-Payne-A* and 1498 stars measured by *DD-Payne-G*, respectively. Nearly 45% stars have SNR pix^{-1} larger than 60. Then we re-estimate the uncertainty of each stellar parameter using Equation 2.11 to represent real uncertainties from *DD-Payne* measurements (see details in Section 2.5). Next, we follow the parameter selection recommendation from Xiang et al. (2019) and employ T_{eff}, $\log g$, [Fe/H] from *DD-Payne-A* and take the average of [Mg/Fe], [Si/Fe], [Ti/Fe] from *DD-Payne-G* and [Ca/Fe] from *DD-Payne-A* as a proxy of α elements for later analysis. We also confirmed that using [Fe/H] from *DD-Payne-G*, [Mg/Fe] from *DD-Payne-A*, or weighted average of [Mg/Fe], [Si/Fe], [Ti/Fe], [Ca/Fe] as [α/Fe] do not qualitatively effect our conclusions.

Next, we apply BSTEP (Sharma et al., 2018) to estimate Galactocentric distance (R_{gc}), extinction $E(B-V)$ and stellar age (τ) using PARSEC release v1.2S + COLIBRI stellar isochrones (Marigo et al., 2017) with T_{eff}, $\log g$, [Fe/H], [α/Fe] from *DD-Payne*, and V- and I-band photometry from OGLE-III as input. Even though not all the stars have both photometric magnitudes, we find stars that lack V-band magnitude will be removed anyway by SNR cut-off in the later section criteria. Therefore, we do not lose any stars for this reason. Moreover, after crossmatching our field with other optical or infrared photometric surveys such as VVV (Minniti et al., 2017), *Gaia* (Gaia Collaboration et al., 2022b) and Pan-STARR (Chambers & et al., 2017), we found they all have fewer stars than those in OGLE-III survey. The above trails indicate that our MUSE spectroscopy has surpassed the limitation of the currently deepest photometric survey, and a deeper photometric survey is important for the future study of the bulge.

3.3 Results

In this section, we provide results of stars in the Baade's window using parameters measured by both *DD-Payne* (Xiang et al., 2019; Wang et al., 2022) and BSTEP (Sharma et al., 2018). We talk about the impact of the incomplete observation; parameter distributions (e.g., $E(B-V)$, m_V) of stars in the OGLE field with the comparison to a Galaxia simulated catalog. Additionally, we investigate the multiple components of the metallicity distribution function, as well as the age, [Fe/H], and [α/Fe] distributions of stars in the bulge ($R_{\mathrm{gc}} < 3.5$ kpc). We compare our results mainly with previous studies on 91 microlensed dwarfs (Bensby et al. 2017, hereafter B17) and ARGOS (Ness et al., 2013a) and APOGEE (Rojas-Arriagada et al., 2020) observations at a similar Galactic longitude or latitude. These analyses aim to verify that our stellar parameter measurements are reliable and consistent with previous studies. The comprehensive scientific interpretation will be explored for subsequent publication.

3.3.1 Impact of Incomplete Observation

Since we only got 5 of 10 observation blocks (OBs) being executed, we did not obtain as many MSTO-SGB and giant stars as we expected. Fig 3.3 quantitatively illustrates this observational loss. In this figure, we plot $\mathrm{SNR_{MUSE}} - \log g$ distribution of stars in the selected OGLE field and color-code them by V-band magnitude. We separate the giant and MSTO-SGB stars using the criteria of $\log g$ in Equation 3.1 and fill the regions in orange and blue, respectively. Because we require $\mathrm{SNR_{MUSE}} > 80\ \mathrm{pix}^{-1}$ (the cyan line), having only 50% exposure means stars with $\mathrm{SNR_{MUSE}}$ between $[56.6, 80]\ \mathrm{pix}^{-1}$ are supposed to

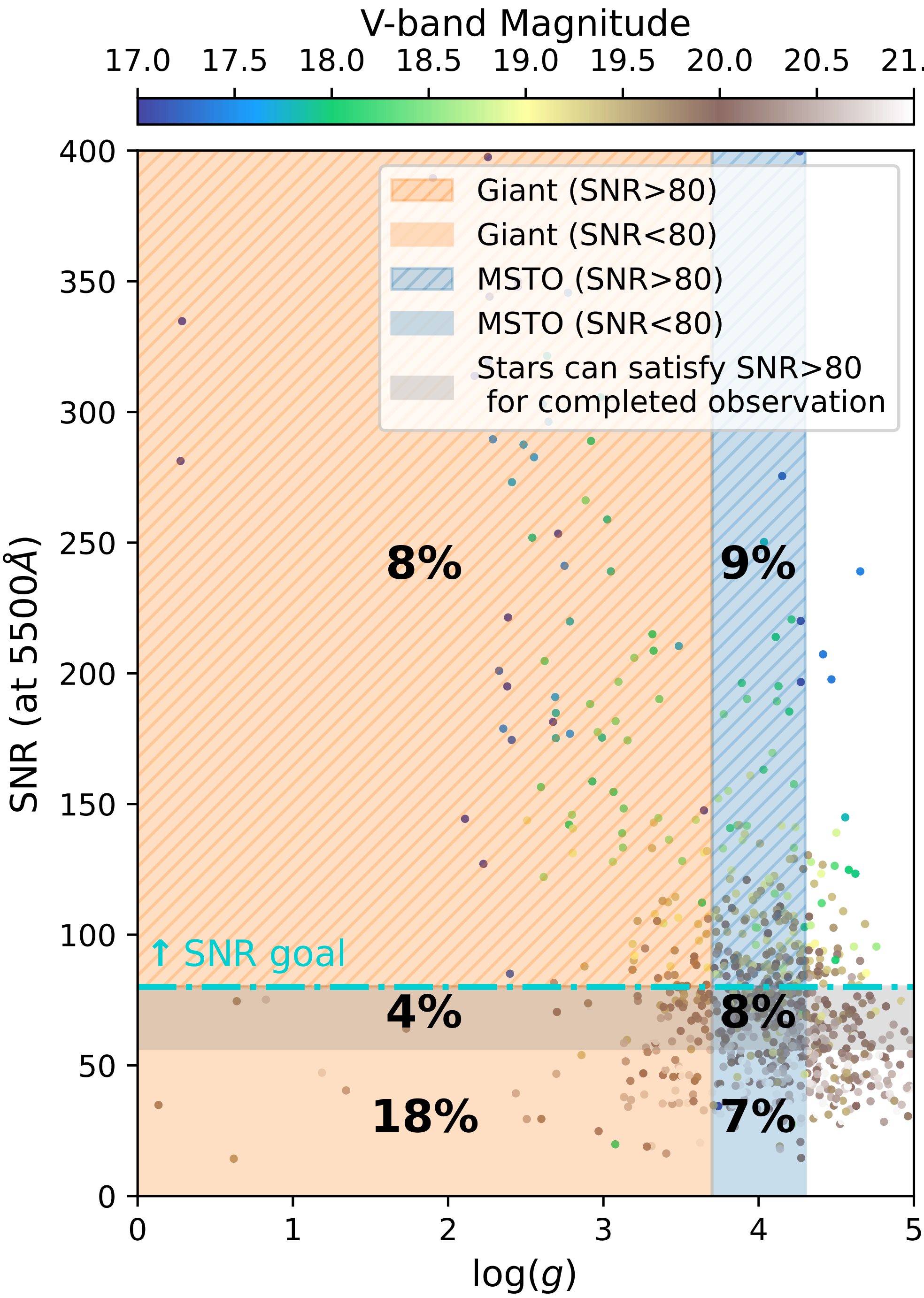

Figure 3.3: $\mathrm{SNR_{MUSE}} - \log g$ distribution of stars towards the Baade's window from MUSE observations color-coded by m_V. The blue and orange regions are MSTO-SGB and giant stars, respectively, which are color-coded by V-band magnitude. The black percentage is the fraction of stars enclosed in each region. Stars above the cyan line have $\mathrm{SNR_{MUSE}} > 80\ \mathrm{pix}^{-1}$, including 8% (138) giants and 9% (159) MSTO-SGB stars from 1734 stars. The grey region with SNR between $[56.6, 80]\ \mathrm{pix}^{-1}$ are stars that were lost due to incomplete observation, which has around 4% (66) giants and 8% (135) MSTO-SGB stars. Note the percentages become uncertain for stars below the cyan line because of the low $\mathrm{SNR_{MUSE}}$ and therefore large $\log g$ uncertainty.

.

satisfy this threshold, but lost due to incomplete observation. Therefore, we mark them in the grey region. The percentages of stars in each enclosed region are texted in the center. This figure indicates that we obtained 8% (138) giants and 9% (159) MSTO-SGB from a total sample of 1734 stars and lost around 4% (66) giants and 8% (135) MSTO-SGB stars. From the current observations we roughly reach $m_V = 20$ at SNR = 80 pix^{-1}. Note we do not make any $\mathrm{SNR_{MUSE}}$ cut-off in this figure, so the percentages become uncertain for stars below the cyan line because of low $\mathrm{SNR_{MUSE}}$ and hence large $\log g$ uncertainties.

3.3.2 Stellar Distributions Compared to Galaxia

Fig. 3.4 demonstrates a comparison of the 1734 OGLE stars with predictions of Galaxia on V-band magnitude and distance. The top two and bottom left panels are V-band magnitude distributions of MSTO-SGB, giant, and dwarf stars classified by Equation 3.1, respectively. The solid lines are predictions from the Galaxia simulated catalog, same as Fig. 3.2. The dashed lines are distributions of OGLE stars toward the Baade's window. Note here we still do not apply any $\mathrm{SNR_{MUSE}}$ cut-off. We find that for stars with $m_V < 19.3$, magnitude distributions from OGLE are consistent with Galaxia prediction. However, the differences appear for stars with $m_V > 19.3$, which could be due to the wrong stellar classification by large $\log g$ uncertainty of faint stars. We will re-investigate it after having more exposure time. The bottom right panel shows the heliocentric distribution of MSTO-SGB and giant stars from Galaxia prediction (solid) and MUSE observation (dashed). This panel indicates that even though we have fewer MSTO-SGB stars than predicted due to incomplete observation, there are still some MSTO-SGB stars close to the Galactic center. The distribution of giant stars is similar to Galaxia prediction, but we have fewer giants with distances less than 7.5 kpc.

The above analysis uses all the 1734 stars in this field. From now on, we demonstrate our results using stars with high-quality measurements by selecting them according to the following criteria:

$$\begin{cases} \mathrm{SNR_{MUSE}} > 80\ \mathrm{pix}^{-1} \\ \text{reduced } \chi^2_{\textit{DD-Payne}} < 20 \\ E(B-V) > 0, \\ [\alpha/\mathrm{Fe}] < 0.4, \end{cases} \tag{3.2}$$

where $E(B-V)$ is estimated by BSTEP and reduced $\chi^2_{\textit{DD-Payne}}$ represents the fitting quality of *DD-Payne*. We made a cut-off of $\mathrm{SNR_{MUSE}}$ because Wang et al. (2022) reported after several tests that stellar parameters can be accurately measured only for stars in the Galactic bulge with $\mathrm{SNR_{MUSE}} > 80$ pix^{-1} (also see Section 2.4.2). The cut-off of $[\alpha/\mathrm{Fe}]$ is to ensure the measurements do not exceed the *DD-Payne* training set range (See Fig. 7 and Fig. 8 of Xiang et al. 2019). In addition, stellar parameters such as [Fe/H], T_{eff}, $\log g$ are input values for BSTEP, so ensuring the accuracy of these parameters is essential for obtaining accurate age, distance, and $E(B-V)$. Note we do not remove stars with $m_V < 19.3$ where a significant proportion of stars might be misclassified as demonstrated in Fig. 3.4. This is because the main factor on $\log g$ uncertainty is $\mathrm{SNR_{MUSE}}$ rather than the magnitude and the dependence of magnitude and $\mathrm{SNR_{MUSE}}$ is not strong in IFS observations, especially on crowded regions. It is possible that a faint star has a high SNR when located at a relatively lower-density area and a bright star has a low SNR when located at the

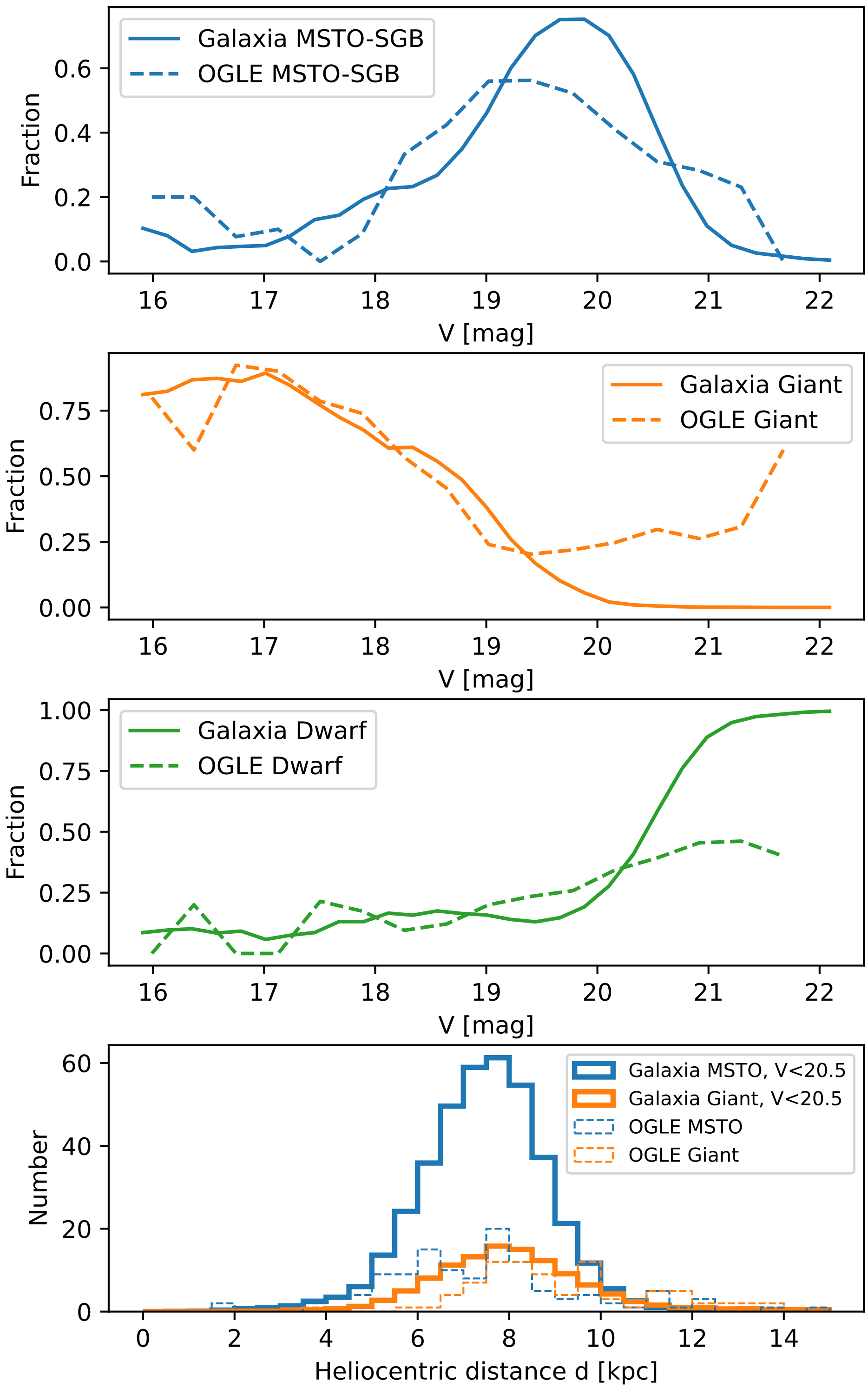

Figure 3.4: Comparison of OGLE stars with Galaxia on V-band magnitude and distance. The top three panels are V-band magnitude distributions of MSTO-SGB, giant, and dwarf stars classified by Equation 3.1, respectively. The solid lines are predictions from Galaxia simulated catalog, same as Fig. 3.2. The dashed lines are distributions of OGLE stars toward the Baade's window. The bottom right panel shows the heliocentric distribution of MSTO-SGB and giant stars from Galaxia prediction (solid) and MUSE observation (dashed).

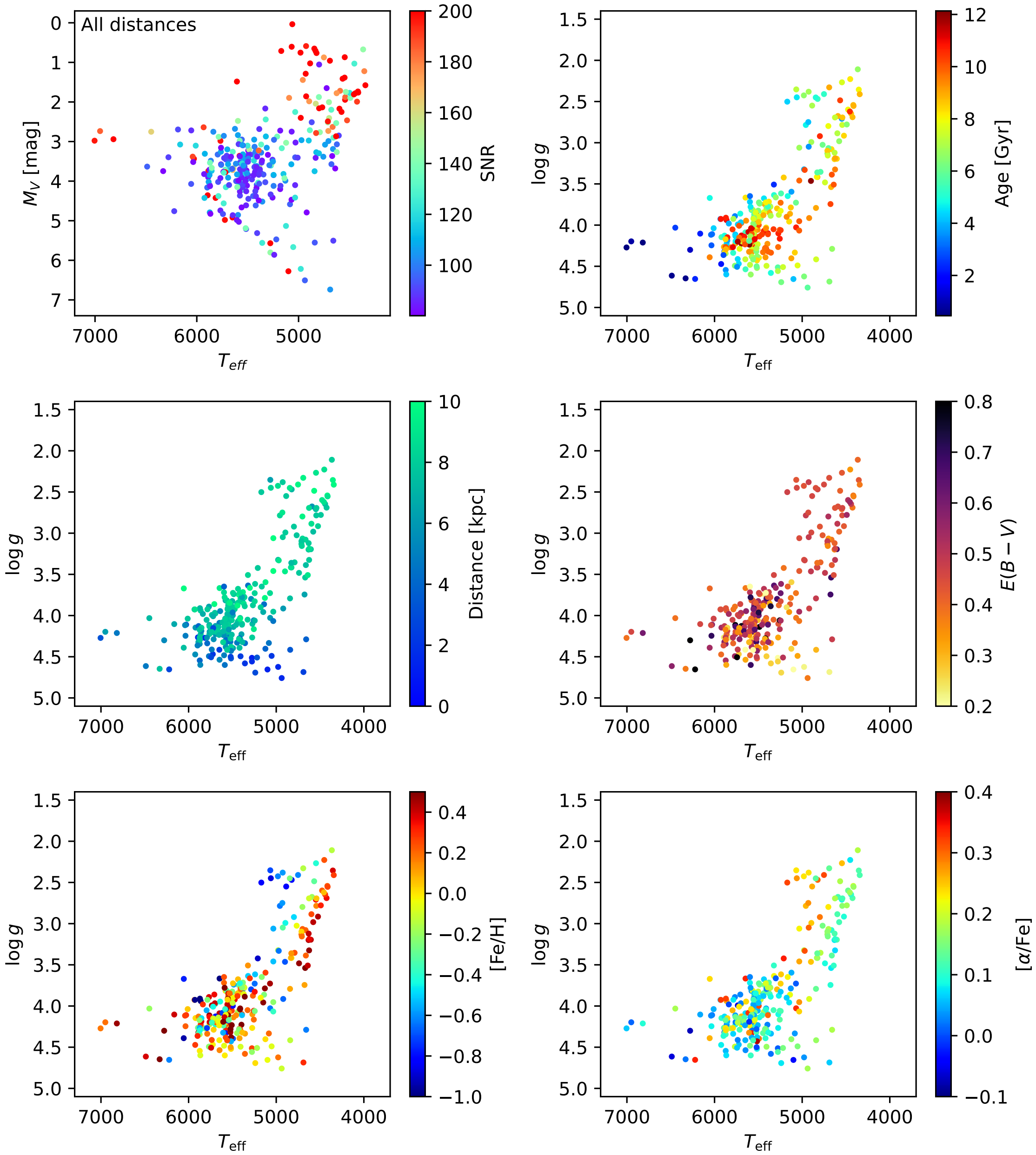

Figure 3.5: Color-magnitude and Kiel diagram of stars toward Baade's window color-coded by SNR, age, distance, $E(B-V)$, [Fe/H], and [α/Fe], respectively. Due to the high luminosity of giant stars, they have larger SNR and can go further than MSTO-SGB stars. The bottom left panel indicates that we have MSTO-SGB stars that can reach the Galactic center. The bottom right panel indicates that all the giant and MSTO-SGB stars have similar $E(B-V)$, which will be investigated later in Fig. 3.6.

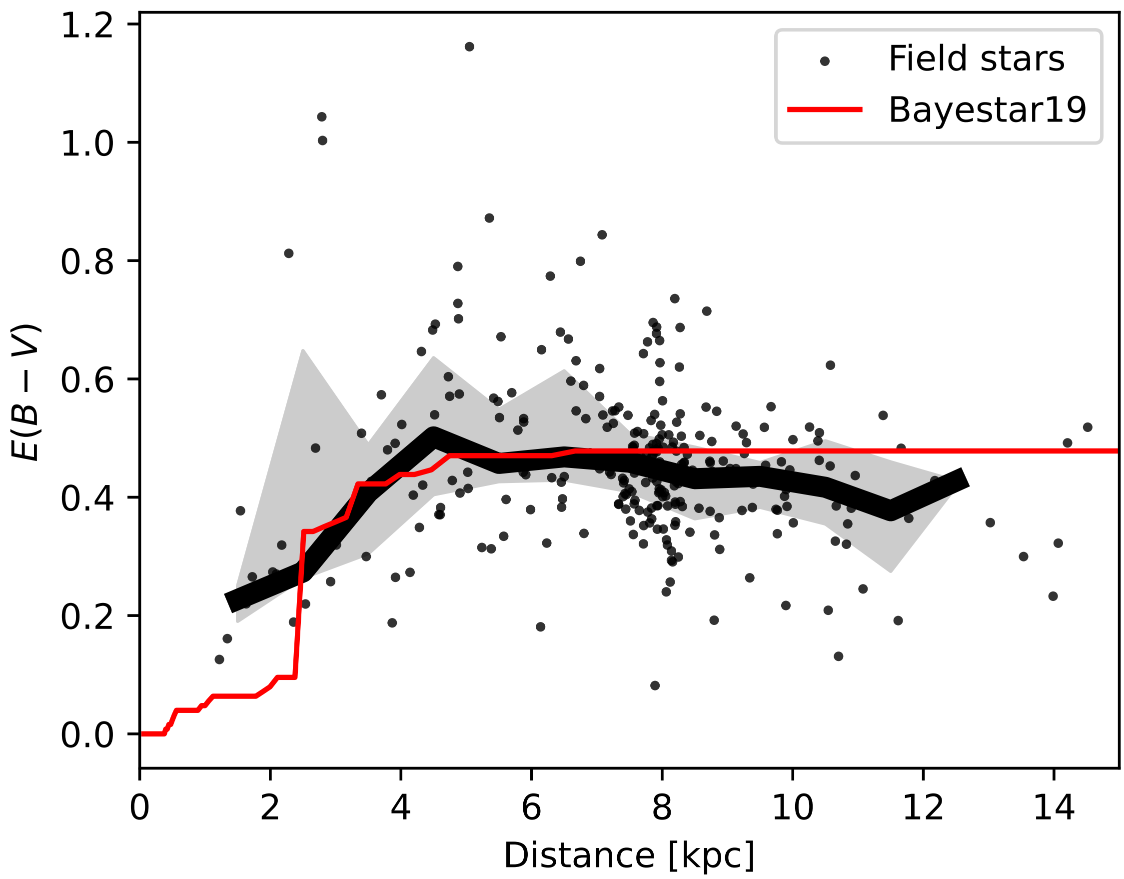

Figure 3.6: Comparison of $E(B-V)$ from field stars in the Baade's window (black dots) and 3D extinction map of Bayestar19 (Green et al. 2019, red line). Most stars in the Baade's window lie on the $E(B-V)$ region of [0.2, 0.6] mag, with the average consistent with the 3D extinction map.

edge of FoV. Therefore, as long as they pass the $\mathrm{SNR_{MUSE}}$ threshold, their $\log g$ should be well estimated. Finally, we have 309 stars left over and identified as high-quality measurements.

Fig. 3.5 demonstrates the colour-magnitude and Kiel diagram of these 309 stars colour-coded by $\mathrm{SNR_{MUSE}}$, age, heliocentric distance, $E(B-V)$, [Fe/H], and [α/Fe], respectively. Due to the high luminosity of giant stars, they have larger $\mathrm{SNR_{MUSE}}$ and can go further than MSTO-SGB stars. The panel color-coded by distances indicates that we have some MSTO-SGB stars that can reach the Galactic center. The panel color-coded by $E(B-V)$ indicates that all the giant and MSTO-SGB stars have similar $E(B-V) \sim 0.5$. We compare stellar $E(B-V)$ measured by BSTEP for these 309 stars with the 3D dust map Bayestar19 from Green et al. (2019) in Fig. 3.6. It can be seen that most stars in the Baade's window lie on the $E(B-V)$ region of [0.2, 0.6] mag, with the average consistent with the 3D extinction map. There is a mismatch between Bayestar19 prediction and observation at distances less than 2 kpc and larger than 8 kpc, respectively. One reason is the lack of stars in these distance ranges. For stars with distances larger than 8 kpc, we think the mismatch is due to the selection effect that only bright stars with low reddening can be observed given our magnitude limit of $m_V = 20.5$. We will re-investigate it in the future when a larger sample is available.

3.3.3 Multiple Metallicity Components

We plot in Fig. 3.7 the [Fe/H] histogram of stars with $R_{\mathrm{gc}} < 3.5$ kpc. For stars with $\mathrm{SNR_{MUSE}} > 80\ \mathrm{pix}^{-1}$, the uncertainty calculated by using Equation 2.11 is σ[Fe/H] $<$ 0.07 dex. On the top panel, the metallicity histogram (grey solid line) is separated into 5 components using GMM modelling, with the Gaussian peaks plotted as vertical grey lines. We compare these peaks with previous studies of Bensby et al. 2017 (blue) and Ness et al. 2013a (red). On the bottom panel, the metallicity histogram is separated into 3 components with the comparison of APOGEE observations (Rojas-Arriagada et al. 2020, green).

By fitting the metallicity distribution function with 5 components (top panel), our Gaussian peaks are located at -0.84 (A), -0.55 (B), -0.19 (C), $+0.17$ (D), and $+0.43$ (E) dex, respectively. According to previous studies, 1845 giant stars at $b = -5°$ from ARGOS survey (Ness et al., 2013a) revealed five [Fe/H] components with the peaks at -1.80 ± 0.10, -1.18 ± 0.10, -0.71 ± 0.02, -0.25 ± 0.02, and $+0.15 \pm 0.02$ dex; B17 also analyzed the metallicity distribution using 91 microlensed dwarf samples at $-4° < b < -2°$ and identified 5 main components with the peaks at -1.09 ± 0.08, -0.63 ± 0.11, -0.20 ± 0.06, $+0.12 \pm 0.04$, and $+0.41 \pm 0.06$ dex. Therefore, except for the most metal-poor peak (A), our results are more consistent with B17 within 1σ. The mismatch of peak (A) is due to the lack of metal-poor stars at [Fe/H]$\sim=$ 1.1 dex in our sample. Comparing the results of this work and B17 to those from the ARGOS survey, both of us misidentify the most metal-poor component with [Fe/H] $= -1.80$ dex, which is because neither of us obtained enough stars at this [Fe/H] range. In this work, stars with [Fe/H] < -1.5 dex are measured by extrapolation of *DD-Payne* during full-spectrum fitting and are removed by selection criteria. In addition, same as B17, we found a very metal-rich component at [Fe/H] = 0.43 dex which ARGOS results did not recognize because it lacks calibration of

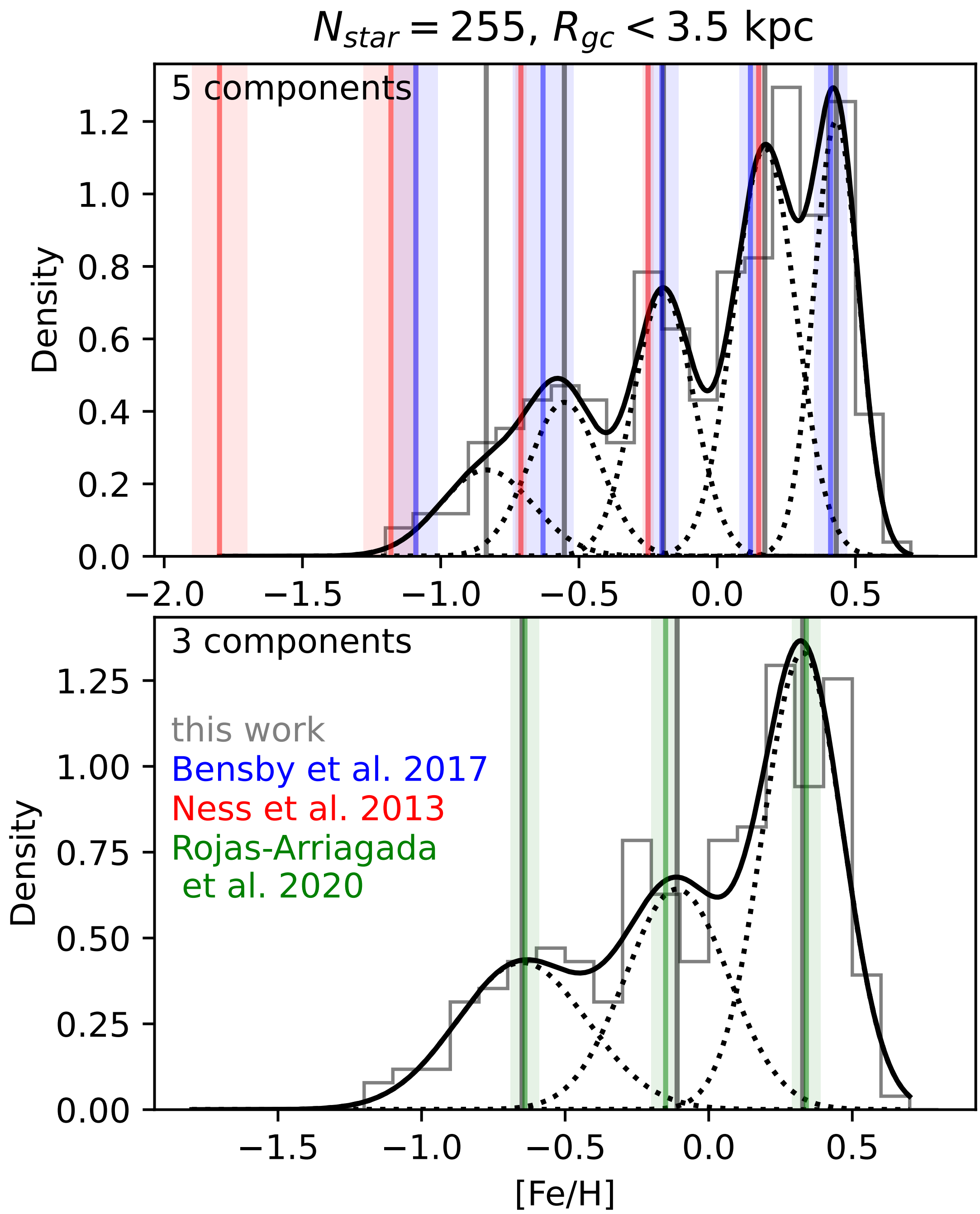

Figure 3.7: [Fe/H] distributions of stars with $R_{\mathrm{gc}} < 3.5$ kpc from our MUSE observations. **Top panel:** metallicity histogram (grey solid line) separated into 5 components using GMM modelling, with the Gaussian peaks plotted as vertical grey lines. We compare these peaks with previous studies using 91 microlensed dwarfs (Bensby et al. 2017, blue) and ARGOS survey (Ness et al. 2013a, red). **Bottom panel:** same metallicity histogram as the top panel but separated into 3 components with the comparison of APOGEE observations (Rojas-Arriagada et al. 2020, green).

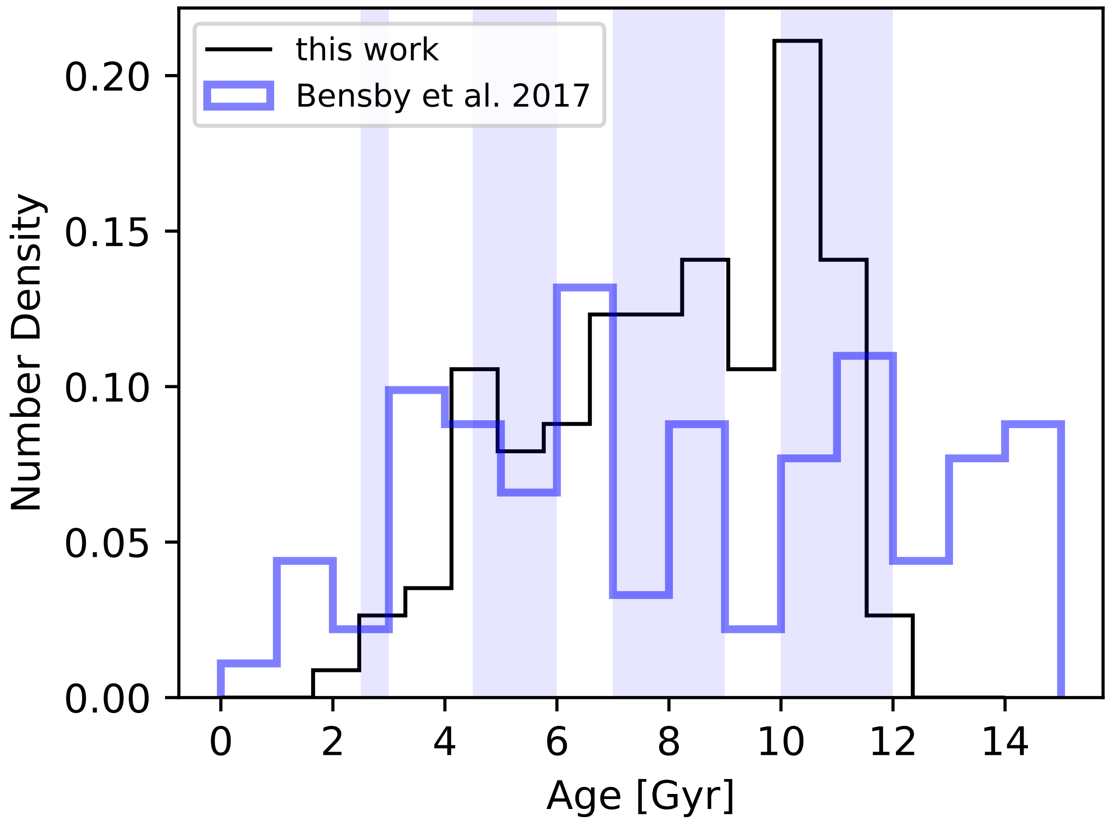

Figure 3.8: Age distribution of MSTO-SGB stars in the bulge (R_{gc} < 3.5 kpc) with a comparison of results using 91 microlensed dwarf samples from Bensby et al. 2017 (blue). According to the histogram, the fractions of age bins in our results are different from those in B17. Nevertheless, after comprehensive statistics using the age probability distribution functions, B17 claimed four major episodes occurring about 11, 8, 6 and 3 Gyr ago (blue regions), and the first three can also be identified in our results.

stars at this high metallicity (priv. comm. from B17). Our results support the component at [Fe/H] = 0.43 dex to be physical.

We also demonstrate the results by fitting with 3 Gaussian components (bottom panel) and the peaks are at −0.65 (A), −0.11 (B), and +0.33 (C) dex. Then we compare them with the previous study of Rojas-Arriagada et al. (2020) using 3292 stars from APOGEE observations at $2.5^\circ < |b| < 4^\circ$, which reports three Gaussian peaks at -0.64 ± 0.05, -0.15 ± 0.05, and $+0.34 \pm 0.05$ dex. Note the uncertainty of Gaussian peaks from Rojas-Arriagada et al. (2020) is according to Monte Carlo analysis (see their Fig. 11). This panel indicates that our results are consistent with APOGEE observations within 1σ. The three peaks can also be identified visually from the histogram.

3.3.4 Age, [Fe/H] and [α/Fe] Distributions

With the age provided by the most precise tracer of MSTO-SGB samples, we can study the stellar age distribution and its relation with [Fe/H] and [α/Fe] in the Galactic central regions. We select MSTO-SGB stars according to Equation 3.1 with R_{gc} < 3.5 kpc and `flag_alpha_fe` = 1 for accurate and physical age and [α/Fe] measurements. Then we remove stars with $\sigma(\text{age}) < 0.01$ Gyr because our sample normally have $0.15 < \sigma(\text{age}) < 3.5$ Gyr. The stars with extremely small σ(age) should normally be very young (< 1 Gyr), but after checking with results using different [Fe/H] and [α/Fe] proxy strategies, these stars are estimated to be $10 \sim 12$ Gyr. Therefore, we think there might be issues with BSTEP when estimating their ages and remove them from the sample. We also double-check that for the other stars, different [Fe/H] and [α/Fe] proxy strategies normally provide similar ages with $\Delta\text{age} < 3$ Gyr.

After cleaning our sample, we plot the age distribution in Fig. 3.8 with a comparison of results using 91 microlensed dwarf samples from B17 (blue). It can be seen that our age distribution spans a wide range from 2 Gyr to 13 Gyr with several peaky features. According to the histogram, the fractions of age bins in our results are different from those in B17. It could be due to the selection effect caused by the different Galactic longitude ranges ($-6° < l < 6°$ for B17 and 1.66° for our sample). After comprehensive statistics using the age probability distribution functions, B17 claimed four major episodes occurring about 11, 8, 6 and 3 Gyr ago (blue regions). Even though we only have around 100 stars and the distribution structures depend on the binning of the data, the first three episodes can also be identified in our results at the regions of $10 \sim 12$ Gyr, $7 \sim 9$ Gyr, and $4 \sim 6$ Gyr. Therefore, we think both our results are physical and indicate the complexity of the star formation history in the Galactic bulge.

We can also study the chemical evolution of the Galactic center by combining age with [Fe/H] and [α/Fe], as shown in Fig. 3.9. For each panel, we also plot in solid lines the evolution of abundances for different birth radius R_b predicted by the chemical evolution model of Sharma et al. (2021b) for comparison. Since the maximum [α/Fe] in our sample is around 0.3 dex, we manually changed α_{max} in Sharma et al. (2021b) model from 0.225 to 0.3 to match the [α/Fe] scale. The top row demonstrates age-metallicity and age-[α/Fe] diagrams from the MSTO-SGB sample color-coded by [α/Fe] and [Fe/H], respectively. From age-metallicity distributions, we can see a metallicity enrichment history from old, metal-poor, and α-rich stars to young, metal-rich, and α-poor stars. The age-[α/Fe] distributions also indicate a decline of [α/Fe] as a function of stellar age. We find that the distributions of both panels are more scattered than the [Fe/H] and [α/Fe] evolution tracks at $R_b < 3$ kpc. One reason is the uncertainty of parameter measurements, the other could be stellar radial migration, which means some of the MSTO-SGB stars in these two panels were born at a larger R_{gc}, and moved inwards the Galactic center later on. Moreover, the metallicity range in the age-[Fe/H] distribution is wider than model predictions and shows a more efficient enrichment history.

The bottom panel of Fig. 3.9 shows the [α/Fe]-[Fe/H] distributions of MSTO-SGB stars color-coded by age. The grey dashed line represents solar-[Fe/H] and solar-[α/Fe]. Firstly, this panel is quantitatively consistent with the results from B17 (see their Fig. 20), where stars are separated into two main sequences, one with α-rich, metal-poor and old stars, the other with α-poor, metal-rich and young stars. The most metal-poor stars in this panel show a relatively flatter [α/Fe] trend with [Fe/H]. At around [Fe/H] = −0.5 dex, [α/Fe] stars decrease with metallicity increasing. At about solar-[Fe/H], [α/Fe] becomes flatter again with the increase of [Fe/H]. There are a non-negligible number of high-[α/Fe] young stars and low-[α/Fe] old stars that are also seen in B17. They challenge

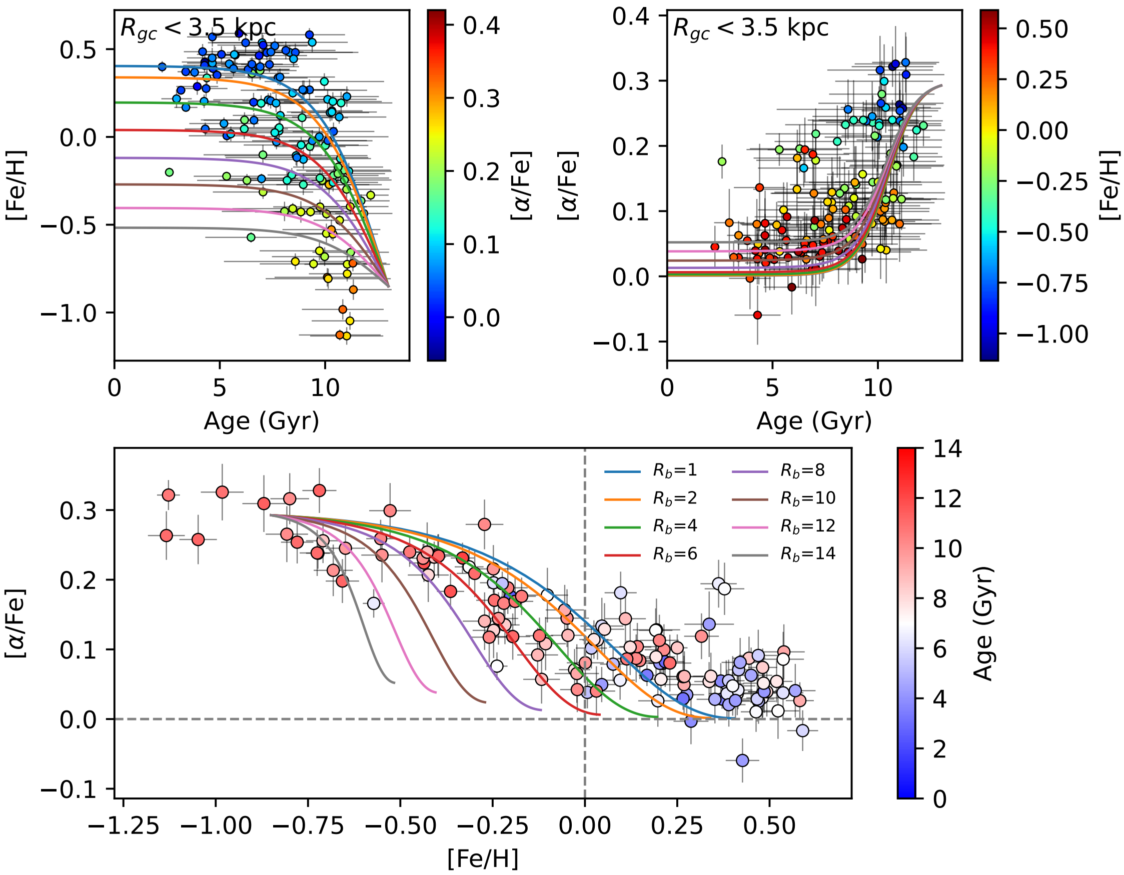

Figure 3.9: Chemical evolution history and [α/Fe]-[Fe/H] distributions of MSTO-SGB stars in the bulge with $R_{\rm gc}$ < 3.5 kpc. The top row demonstrates age-metallicity and age-[α/Fe] diagrams from the MSTO-SGB sample with $R_{\rm gc}$ < 3.5 kpc color-coded by [α/Fe] and [Fe/H], respectively. The bottom panel shows the [α/Fe]-[Fe/H] distributions of MSTO-SGB stars color-coded by age. The grey dashed line represents solar-[Fe/H] and solar-[α/Fe]. We also plot in solid lines the evolution of abundances for different birth radius R_b predicted by the chemical evolution model of Sharma et al. (2021b) for comparison. Since the maximum [α/Fe] in our sample is around 0.3 dex, we manually changed $\alpha_{\rm max}$ from 0.225 to 0.3 to match the [α/Fe] scale.

the tight relation of α-elements with stellar ages. Some studies propose that it is due to star formation events near the edge of the Galactic bar (e.g., Chiappini et al. 2015). Others suggested it is due to the material transfer in binary system (Yong et al., 2016; Hekker & Johnson, 2019; Zhang et al., 2021). Future studies are needed to investigate the origin of these unusual stars. Compared to the abundance evolution tracks predicted by Sharma et al. (2021b), stars in our observation mainly follow the evolution tracks with $R_b = 1 \sim 4$ kpc, and the location of this down-turn or "knee" is consistent with the tracks for the innermost R_b ($1 \sim 2$ kpc). However, we also see some metal-rich stars with [Fe/H]~ 0.5 dex do not follow the model predicted evolution tracks. Note these abundance evolution tracks from Sharma et al. (2021b) are based on stars with $R_{\rm gc}$ between 3 and 15 kpc in APOGEE observations. Therefore, abundance evolution tracks with $R_b < 3$ kpc are calculated by extrapolation, and the mismatch of our results with the model prediction indicates future improvements of chemical evolution models in the Galactic center.

3.4 Summary and Future Studies

In this Chapter, we performed data analysis of a 3.6-hour MUSE observation towards the Baade's window with $A_V = 1.45$ mag and studied the chemical evolution and age distribution using MSTO-SGB stars in the Galactic bulge. We first apply *DD-Payne* (Ting et al., 2017a; Xiang et al., 2019) and BSTEP (Sharma et al., 2018) to measure stellar parameters ($T_{\rm eff}$, $\log g$, and chemical abundances) and distance, extinction $E(B-V)$ and ages in our selected field. The V-band magnitude distributions of giant, MSTO-SGB, and dwarf stars in our sample are consistent with predictions of a simulated catalog by Galaxia with $m_V < 19.3$, and the $E(B-V)$-distance relation agrees with the 3D dust map Bayestar19 (Green et al., 2019). We obtained 138 giant and 159 MSTO-SGB stars with $\mathrm{SNR}_{\rm MUSE} > 80\ \mathrm{pix}^{-1}$ from a 3.6-hour MUSE exposure and lost 66 giants and 135 MSTO-SGB stars due to incomplete observation. Despite this, we still obtain 138 MSTO-SGB stars with $R_{\rm gc} < 3.5$ kpc which can be used to study the [Fe/H], [α/Fe], and age distributions.

The metallicity distribution function is separated into three and five Gaussian components using GMM modelling and respectively compared with previous studies using stars at similar Galactic longitude or latitude. When separated into three components, our results agree with APOGEE observations (Rojas-Arriagada et al., 2020) within 1σ. When separated into five components, our results agree with Bensby et al. (2017) within 1σ except for the most metal-poor components, which is due to the lack of very metal-poor stars in our sample. The age distribution revealed by this dataset shows three groups with ages around $10 \sim 12$ Gyr, $7 \sim 9$ Gyr, and $4 \sim 6$ Gyr, which are consistent with predictions from B17. However, the relative fraction of each age bin is different from B17 which could be due to the difference in the Galactic longitude range. We also demonstrate age-metallicity, age-[α/Fe], [α/Fe]-[Fe/H] distributions using our sample with the comparison of abundance evolution tracks predicted by the chemical evolution model of Sharma et al. (2021b). Most of our stars are located at tracks with $R_b = 1 \sim 4$ kpc. And the [Fe/H] and [α/Fe] evolution history is clearly seen. The [α/Fe]-[Fe/H] distribution of stars in the bulge indicates a "knee" at [Fe/H] $= -0.5$ dex, which agrees with B17 and the chemical evolution model. However, there are several stars with [Fe/H] ~ 0.5 dex shown in our sample but not predicted by the chemical evolution model, which provides clues for future improvements on the models in the Galactic center.

The above results are mainly to verify the reliability of our stellar parameter measurements and the power of MUSE to reveal the complicated star formation history of the Galactic center by comparing them with previous studies. For the subsequent publication, we will also add detailed scientific interpretations of our results. In the future, there are many aspects can be explored using this dataset. Firstly, we can investigate the origin of young [α/Fe]-rich and old [α/Fe]-poor stars in our sample. In addition, we can combine [Fe/H] with velocities of stars to study the chemodynamical distributions similar to Ness et al. (2013b) and compare the results in the near and far sides of the Galactic bulge using giant stars. The abundance evolution tracks at $R_b < 3$ kpc in Sharma et al. (2021b) is predicted by extrapolation. We can use our datasets to verify this relation in the Galactic central regions and help improve the chemical evolution model. Since we only have 50% exposure time being executed in this work, it is necessary to apply more MUSE time to observe more similar low-extinction fields, which can provide a larger sample and help in studying the spatial distribution of chemodynamical properties along the Galactic coordinates. More stars with precise ages are also important to reveal the accurate star formation history of the Galactic bulge. A deeper catalog is also needed to obtain photometric magnitudes for more stars.

Data Availability

The comprehensive catalog containing all the stellar parameters utilized in this work will be available through `https://github.com/purmortal/muse-bulge` following the re-formatting of this chapter for publication.

Chapter 4

A Tool for Creating Mock IFS Datacubes

ZIXIAN WANG,[1,2] MICHAEL R. HAYDEN,[1,2] SANJIB SHARMA,[1,2] JESSE VAN DE SANDE,[1,2] JOSS BLAND-HAWTHORN,[1,2,3] SAM VAUGHAN,[1,2] MARIE MARTIG[4]

[1]Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW, 2006, Australia
[2]ARC Centre of Excellence for All Sky Astrophysics in Three Dimensions (ASTRO-3D)
[3]Miller Professor, Miller Institute, UC Berkeley, Berkeley CA 94720
[4]Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK

This chapter is reproduced from Chapter 1 & 2 of Wang et al. (2023), which is to be submitted to Monthly Notices of the Royal Astronomical Scocity as "The Milky Way in Context: Building an integral-field spectrograph datacube of the Galaxy". I wrote all of the text, produced all the figures, and wrote the codes for data analysis and interpretation. Sanjib Sharma provided E-GALAXIA mock catalog. Joss Bland-Hawthorn, Sanjib Sharma, and Michael Hayden supervised the whole work. Sanjib Sharma and Michael Hayden contributed equally to this work. Jesse van de Sande, Sam Vaughan and Marie Martig helped with using PPXF, scientific interpretations and provided comments.

Our Galaxy is by far the best studied in the Universe, and it has long been regarded as a benchmark and ideal laboratory to test our theories of galaxy evolution. However, due to the different viewpoints, direct comparisons of Galactic and extragalactic observations are marred by many challenges including selection effects, observational and methodological differences. In this chapter, we present a novel code GALCRAFT to address these challenges by generating a mock integral-field spectrograph (IFS) datacube of the Milky Way using simple stellar population (SSP) models and a mock stellar catalog from E-GALAXIA, which is chemodynamically consistent to the Milky Way. The products are in the same data format as external galaxies and enable direct comparisons. With future higher resolution and multi-[α/Fe] SSP templates, GALCRAFT will be useful to identify key signatures such as [α/Fe]-[M/H] distribution at different R and $|z|$ and potentially measure radial migration and kinematic heating efficiency to study detailed chemodynamical evolution of external galaxies. The GALCRAFT code will be publicly available via https://github.com/purmortal/galcraft.

4.1 Introduction

How galaxies formed and evolved remains one of the outstanding questions facing astrophysics today. Due to our unique vantage point, the Milky Way (MW) is by far the best-studied galaxy in the Universe. Precise astrometric data from *Gaia* (Gaia Collaboration et al., 2022b) and accurate chemical abundances of individual stars from large spectroscopic surveys such as GALAH (Buder et al., 2021), APOGEE (Majewski et al., 2017) and LAMOST (Zhao et al., 2012) have been conducted in the last decade for Galactic archaeologists to reveal the detailed chemodynamical picture of the Milky Way (Freeman & Bland-Hawthorn, 2002; Bland-Hawthorn & Gerhard, 2016), including the accretion history and interplay of chemical and dynamical processes (e.g., Xiang & Rix 2022). Particularly, Hayden et al. (2015) demonstrated two distinct sequences (bimodality) in [α/Fe]-[Fe/H] distributions which vary systematically with location R and $|z|$ across the MW, and the skewness of the outer-disk metallicity distribution function (MDF) indicates the important process of radial migration. These two sequences correspond to the thin and thick disk structures, and are found to be common in most disk galaxies (e.g., Yoachim & Dalcanton 2006; Comerón et al. 2018; Martínez-Lombilla & Knapen 2019).

In the last decade, many efforts have been made to develop MW chemical evolution models to reproduce the [α/Fe] bimodality. For example, Schönrich & Binney (2009a) introduced a chemical evolution model that included the radial flow of gas and radial migration of stars along with other related physical processes such as continuous star formation history, stellar yields, gas accretion and outflow. It can reproduce the metallicity distribution of stars and bimodality of the [α/Fe]-[Fe/H] distribution of stars in the solar neighbourhood (Schönrich & Binney, 2009b), where the former is consistent with the Geneva–Copenhagen survey (Nordström et al., 2004); Sanders & Binney (2015) extended the analytical phase-space distribution functions in Binney (2012) to make the models easy to fit data and they successfully tracked the evolution of metallicity; Sharma et al. (2021b) applied Shu distribution function (Shu, 1969) and extended the model in Sanders & Binney (2015) by adding the distribution of [α/Fe] and a new prescription for velocity dispersion relation from Sharma et al. (2021a), and successfully reproduced the [α/Fe]-[Fe/H] bimodality distributions which are quantitatively consistent with APOGEE observations in Hayden et al. (2015). The relation of chemistry and kinematics are also consistent at all Galactic locations.

Even though specific chemical evolution models can reproduce the chemodynamical distributions of our MW successfully, the Milky Way is still only one galaxy. Whether these theories are applicable to other disk galaxies' formation history is still an open question, and it is still under debate whether Milky Way is unique in the Universe. From an extragalactic view, the MW is an Sb-Sbc galaxy positioned at the "green valley" on the color-mass diagram (Mutch et al., 2011; Licquia et al., 2015) and obeys the Tully-Fisher relation (Licquia & Newman, 2016); it shows a low star formation rate (SFR) which is not unusual when compared to similar galaxies (Fraser-McKelvie et al., 2019); The Milky Way appears to be a kinematically typical spiral galaxy for its intrinsic luminosity (Bland-Hawthorn & Gerhard, 2016), but unusually compact in its stellar mass (e.g. Licquia et al. 2016); Its satellite galaxies are fewer in number and more compactly distributed (e.g. Yniguez et al. 2014) compared to other galaxies with similar total stellar mass; In simulations, Mackereth et al. (2018) found chemical bimodality is rare in the EAGLE simulation, while Buck (2020) found it to be present due to gas-rich mergers in all simulated MW-like galaxies from the NIHAO-UHD project.

Over the last decade, integral-field spectroscopy (IFS) instruments have enabled us to obtain integrated-light spectra of galaxies across different regions of the same object. A number of IFS surveys have already been carried out such as SAMI (Croom et al., 2012), MaNGA (Bundy et al., 2015), and CALIFA (Sánchez et al., 2012) to measure kinematics and chemical compositions of hundreds of galaxies. Several studies compared face-on galaxies in these surveys with the MW: Boardman et al. (2020a) selected 62 Milky Way analogues (MWAs) from MaNGA with the criteria of stellar masses and bulge-to-total ratios and found most of them have flatter stellar and gas-phase metallicity gradients due to different disc scale lengths, and a greater consistency can be found when scaling gradients by these lengths; Zhou et al. (2023) revisited 138 MaNGA galaxies by fitting the spectra with a semi-analytic chemical evolution model (Zhou et al., 2022) and measured their star formation and chemical enrichment histories. They detected similar $[\alpha/\mathrm{Fe}]$ bimodality as the Galactic APOGEE observations (Abdurro'uf et al., 2022) in many of the galaxies.

Even though face-on MWAs are ideal for measuring parameter variations crossing different Galactocentric radii R_{gc}, they lack vertical ($|z|$) information, which is essential to distinguish the thin and thick disk. Therefore, edge-on MWAs are more useful in this case to provide parameter distributions at different R and $|z|$. Several studies of nearby edge-on MWAs and lenticular galaxies using MUSE found similar kinematics and bimodal distributions to the MW (e.g., Pinna et al. 2019a,b; Scott et al. 2021; Martig et al. 2021). Particularly, Scott et al. (2021) demonstrated that UGC 10738 has similar $[\alpha/\mathrm{Fe}]$-$[\mathrm{Fe/H}]$ distributions at different projected R and $|z|$ with the MW observation in Hayden et al. (2015) and model predictions in Sharma et al. (2021b), which supports the commonness of MW's chemical distributions.

Despite the efforts above, a direct comparison of MW with its analogues in kinematics and chemistry is still challenging because the observables and methods used for studying the MW are different from those for external galaxies - utilizing properties of individual stars as opposed to integrated quantities with projection effects. Therefore, the comparisons are suffering from systematics and biases (Boardman et al., 2020a); In addition, some key processes such as radial migration and kinematic heating have not been explored on external galaxies, which are also essential to identify the uniqueness of the MW. Therefore, to take the MW as an ideal laboratory and test galaxy evolution theories, one needs to remove these observational and methodological biases, transfer the observables of MW and external systems into the same definition, and apply models that consider both chemical enrichment and kinematic processes. These considerations lead to the development of the tools presented in this work.

In this study, we present a novel code GalCraft to generate mock integral-field spectrograph (IFS) datacubes of the MW with integrated light using simple stellar population (SSP) models and mock stellar catalog from E-Galaxia, which is based on the chemo-dynamical model of Sharma et al. (2021b) (hereafter S21) that is most consistent to the current Galactic observations in both kinematics and chemistry. This mock datacube is in the same format as extragalactic IFS observations. It can be applied to extragalactic data analysis pipelines to measure directly comparable parameters in (age, [M/H], [α/Fe], $V_{\rm LOS}$, σ, h_3, h_4). We describe the ingredients used in this code and detailed procedures of the code in Section 4.2.

4.2 Datacube Generation

The purpose of GalCraft is to take the mock stellar catalog of the Galaxy obtained from E-Galaxia (Sharma et al. 2023, in prep.), and transform it into a mock datacube in three dimensions (two in spatial and one in spectral) as observed by corresponding integral-field spectrographs (IFS) such as MUSE (Bacon et al., 2010), SAMI (Croom et al., 2012) or MaNGA (Bundy et al., 2015). The GalCraft code takes as input the following set of user-specified input parameters: galaxy distance (d), inclination (i), extinction, SSP templates, instrumental properties, as well as a few additional parameters (see the full list in Table 4.1). The produced mock datacube can be processed in the same way as real IFS observations data by codes like Voronoi binning (Cappellari & Copin, 2003), Penalized Pixel-Fitting (pPXF, Cappellari & Emsellem 2004; Cappellari 2017), line-strength indices (e.g., Martín-Navarro et al. 2018), or a combination of them (e.g., The GIST pipeline, Bittner et al. 2019). The results can be compared directly with those from IFS observations of MWAs in terms of mass-weighted parameter maps, line-of-sight velocity distribution (LOSVD), and mass fraction distributions. The ingredients, flexibility, procedures, and computational expenses of this code are described in detail in the following sub-sections.

4.2.1 The Ingredients

Chemical Evolution Model

We apply the analytical chemodynamical model of the Milky Way developed by S21 which can predict the joint distribution of position ($\mathbf{x}$), velocity ($\mathbf{v}$), age (τ), extinction $E(B-V)$, the photometric magnitude in several bands and chemical abundances ([Fe/H], [α/Fe]) of stars in the Milky Way. Compared with previous models (e.g., Chiappini et al. 1997; Haywood et al. 2019), this model included a new prescription for the evolution of [α/Fe] with age and [Fe/H] and a new set of relations describing the velocity dispersion of stars (Sharma et al., 2021a). This model shows for the first time that it can reproduce the [α/Fe]-[Fe/H] distribution of APOGEE observed stars (Hayden et al., 2015) at different radius R and height $|z|$ across the Galaxy. In this model, the origin of two [α/Fe] sequences is due to two key processes: the sharp transition from high-[α/Fe] to low-[α/Fe] at around 10.5 Gyr ago that creates a valley between the two sequences, which is likely due to the delay between star formation and the occurrence of SN Ia; the radial migration or so-called churning is responsible for the large spread of the low-[α/Fe] sequence along the [Fe/H] axis. This model also showed that without churning it is not sufficient to reproduce the double sequences (see their Fig. 6).

Because this chemical evolution model is purely analytical, it is easy to be inserted into the forward-modeling tool E-Galaxia to generate mock stellar catalogs for further analysis.

In addition, several free parameters such as radial migration and heating efficiency are adjustable, which will be useful to implement similar analysis on external galaxies.

E-Galaxia

The ideal way to build the Milky Way datacube is to combine all the stellar and gas-phase observations in the Galaxy with well-measured parameters such as position ($\mathbf{x}$), velocity ($\mathbf{v}$), age (τ) and chemical abundances. However, this is impractical as the Galaxy has hundreds of billions of stars being unexplored and the dust in the disk obscures distant light. An alternative way is to use particle catalogs from N-body/hydrodynamical simulations (e.g., EAGLE Schaye et al. 2015, FIRE-2 Hopkins et al. 2018), but most of the simulations only contain $\sim 10^6$ stellar particles, which is not enough to generate integrated spectra because each spatial bin would only contain less than a hundred particles. This in turn would increase the sampling noise of the integrated spectrum and make observables derived from spectra noisy. Even though oversampling can solve this problem, it is still challenging to find a simulation that is observably identical to the MW in all aspects quantitatively, especially the $[\alpha/\mathrm{Fe}]$ bimodal trends.

Therefore, the approach we take here is to generate a mock stellar catalog to represent the MW. We use E-Galaxia (Sharma et al. 2023, in prep.). This tool is in accordance with the chemical evolution model of S21 and can create a catalog with the user-defined number of particles with parameters including position ($\mathbf{x}, \mathbf{y}, \mathbf{z}$), velocity ($\mathbf{v_x}, \mathbf{v_y}, \mathbf{v_z}$), age ($\tau$), metallicity ($[\mathrm{M/H}]$) and $[\alpha/\mathrm{Fe}]$, where the parameter distributions are chemodynamically consistent to the MW observations. There are other codes like TRILEGAL (Girardi et al., 2005), BESANCON (Girardi et al., 2005), and Galaxia (Sharma et al., 2011) can also create mock catalogs. Compared to E-Galaxia, the underlying models of these codes lack the information of $[\mathrm{Fe/H}]$ and $[\alpha/\mathrm{Fe}]$, and do not have the processes of radial migration and kinematic heating. Furthermore, the ability of E-Galaxia to control the observed properties by regulating the underlying physical process is important for future analysis of external galaxies, whose formation history and dynamical processes are expected to be different from the Milky Way.

Spectral Libraries

To build a mock datacube, one important procedure is to turn particles in E-Galaxia into stellar spectra based on their properties, so a spectra library is needed. Because the integrated light in extragalactic observations is assumed from stellar populations, in GalCraft, we treat each particle as a simple stellar population (SSP). The SSP spectrum describes the spectral energy distribution (SED) of a stellar population with a single age, metallicity, and chemical abundance patterns. An initial mass function (IMF) is assumed, and the stellar population is evolved using a given isochrone (Conroy, 2013). GalCraft currently supports MILES (Vazdekis et al., 2010, 2015) and PEGASE-HR (Le Borgne et al., 2004).

The MILES SSP library (FWHM = 2.51Å, 3500Å $< \lambda <$ 7500Å) is based on the model of Vazdekis (1999). It uses Padova 2000 (Girardi et al., 2000) or BaSTI (Pietrinferni et al., 2004, 2006) isochrones and IMF in Unimodal/Bimodal (Vazdekis et al., 1996), Kroupa Universal/Revised (Kroupa, 2001) and Chabrier (Chabrier, 2003) shapes. For Padova 2000 isochrones, the template grids cover 7 metallicity bins between (-2.32, 0.22) dex, 50 age bins between (0.063, 17.78) Gyr and one scaled-solar [α/Fe] bin (Vazdekis et al., 2010). As for BaSTI isochrones, the template grids cover 12 metallicity bins between (-2.27, 0.40) dex, 53 age bins between (0.03, 14.00) Gyr and two [α/Fe] bins of (0.0, 0.4) dex (Vazdekis et al., 2015). Since [α/Fe] enhancement is essential in this study, for most of the cases we will use the α-variable templates.

The PEGASE-HR library (FWHM = 0.55Å, 3900Å $< \lambda <$ 6800Å) is based on the Elodie 3.1 stellar library (Prugniel & Soubiran, 2001; Prugniel et al., 2007). The SSP models are computed using Padova 1994 (Bertelli et al., 1994) isochrones with a Salpeter (Salpeter, 1955), Kroupa, or top-heavy (Elmegreen & Shadmehri, 2003) IMF. The templates grid consist of 7 metallicity bins between (-2.30, 0.70) dex, 68 age bins between (0.001, 20) Gyr, and one scaled-solar [α/Fe] bin. In this work, the PEGASE-HR templates are mainly used to explore the effect of spectral resolution on pPXF fitting. Therefore, it is still useful even if it lacks variable [α/Fe]. Revised grids interpolated by ULYSS (Koleva et al., 2009) in the same way as Kacharov et al. (2018) for PEGASE-HR are also available. The new grids contain 15 steps in metallicity between -2.3 and 0.7 dex and 50 steps in age between 10 Myr and 14 Gyr.

4.2.2 Configurations of GalCraft

To meet different research purposes, we set a lot of adjustable parameters in GalCraft. Specifically, the adjustable parameters are divided into three groups: `GALPARAMS`, `SSPPARAMS` and `INSPARAMS`, as listed with descriptions in Table 4.1. The user can set up their preferred parameters to obtain the expected datacubes. `GALPARAMS` is for setting the distance, inclination, and position using coordinates transformation; `SSPPARAMS` governs the spectral properties, such as the SSP templates selection, single or variable [α/Fe], spectral resolution, age, [M/H] and [α/Fe] range and the interpolation method to be used to assign each particle a spectrum (details in Section 4.2.3). Based on particles' parameters, there are two options to assign them spectra: one is "nearest", which will assign the spectrum of the nearest template grid. The other is "linear", which will assign the spectrum by piece-wise linear interpolation of templates. The `INSPARAMS` controls the instrument spatial sampling, atmospheric effects (PSF), and the number of spatial bins in each coordinate. We also provide an option to derive the datacube in the format of a specified instrument. Alternatively, the user can also design a hypothetical instrument that does not exist by manually setting up these parameters, which will be useful for future instrument designs.

4.2.3 Procedures of Making a Mock Datacube

The procedures of generating mock MW datacubes are described in detail as the following steps, along with a flowchart in Fig. 4.1:

[i] Use E-Galaxia to generate a mock stellar catalog of the MW, with particles' position ($\mathbf{x}$), velocity ($\mathbf{v}$), age (τ), extinction $E(B-V)$, metallicity ([Fe/H]) and [α/Fe], transfer [Fe/H] to [M/H].

Table 4.1: Configuration parameters adjustable in the GALCRAFT code.

Parameters	Description	Unit
GALPARAMS:	**Observation parameters of the Galaxy**	
distance	Distance of the Galaxy center to the observer	kpc
θ_{zx}	Angles that the Galaxy rotates along the Y axis in the direction from z to x	deg
θ_{yz}	Angles that the Galaxy rotates along the X axis in the direction from y to z	deg
l	Galactic longitude of the center of the Galaxy	deg
b	Galactic latitude of the center of the Galaxy	deg
SSPPARAMS:	**Parameters of the SSP templates**	
model	The SSP model to be used for the spectrum (MILES or PEGASE-HR)	
isochrone	Isochrones used to generate the SSP templates	
IMF	Initial-mass function used to generate the SSP templates	
single_alpha	If use single [α/Fe] or α-variable templates for the spectrum	bool
factor	Oversampling factor of the templates	
FWHM_gal	Spectral resolution in FWHM of the output datacube	Å
dlam	Bin width of the wavelength sampling of the output datacube	Å
age_range	Optional age range (inclusive) in for the SSP models	Gyr
metal_range	Optional metallicity range (inclusive) in for the SSP models	dex
wave_range	Optional wavelength range (inclusive) in for the SSP models	Å
spec_interpolator	Interpolation method to assign the spectrum to the particle given its (age, metallicity, [α/Fe])	
INSPARAMS:	**Instrument properties**	
instrument	Instrument name, when using this parameter, the other parameters will be automatically setup	
spatial_resolution	Spatial resolution of the datacube	arcsec
spatial_nbin	Number of spatial bins in two coordinates, in the format of [n_x, n_y]	
FWHM_spatial	Full-Width Half-Maximum of the PSF function to model the atmosphere	arcsec

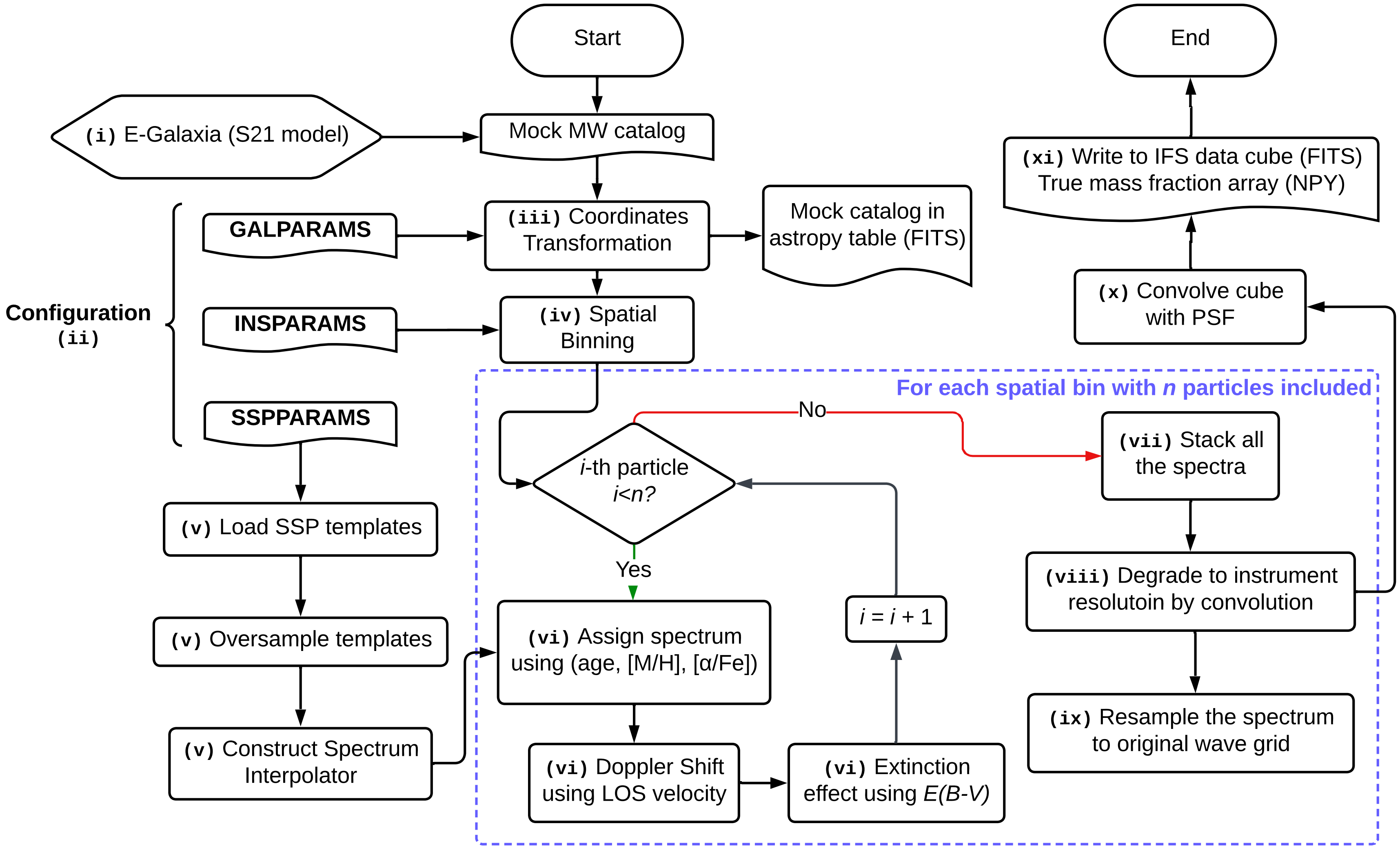

Figure 4.1: A flowchart that demonstrates the workflow of GalCraft code in this work. The numbers on some processes correspond to the steps listed in Section 4.2.3.

[ii] Load the setup parameters of Table 4.1 provided by the user, as well as a list of datacubes to be generated with their center coordinates in (l, b) or (ra, dec).

[iii] Apply coordinates transformation based on `GALPARAMS` parameters to move this Galaxy to a certain distance and rotate it into a defined inclination.

[iv] Apply the spatial binning based on `INSPARAMS`, and find the particles included in each bin, then note the bin number for later use.

[v] Load the SSP templates and make the cutoff based on the (age, [M/H]) ranges in `SSPPARAMS`, then oversample it by a factor of `SSPPARAMS:factor` using `spline` interpolation with the order of three. Next, construct the interpolator which will be used to assign the spectrum in the next step.

[vi] Assign each particle an SSP spectrum based on its (age, [M/H], [α/Fe]) by interpolating the templates with the method defined by `SSPPARAMS:spec_interpolator`. And shift the spectrum due to its line-of-sight velocity ($V_{\rm LOS}$) using the Doppler equation. Then apply the flux calibration due to the particle's distance and extinction.

[vii] Loop over spatial bins to stack all the spectra of particles included and obtain the integrated spectrum for each bin. In the meantime, generate the mass/light-fraction distribution of each spatial bin. The spatial bin numbers are given by `INSPARAMS:spatial_nbin`.

[viii] Degraded the stacked spectrum to the instrument resolution given by `SSPPARAMS:FWHM_gal` using convolution.

[ix] Re-bin the oversampled flux array into the original wave grid or the user-defined interval and wavelength range, according to `SSPPARAMS:dlam` and `SSPPARAMS:wave_range`.

[x] Apply the atmosphere effects by convolving the spaxel with a point spread function (PSF), this can be either a Gaussian or Moffat kernel function with the given `INSPARAMS:FWHM_spatial`.

[xi] Create the fits file header of flux arrays, then combine all the results as datacubes in the format of FITS file.

Other than the above procedures, there are a few notes that need to be clarified to the users as follows:

- The original mock stellar catalog in step (i) should have two coordinate systems, Cartesian coordinates (x, y, z) and Galactic coordinates (l, b), where (l, b) are overlapped with (y, z), respectively. Therefore, by adjusting $(\theta_{zx}, \theta_{yz})$, users can change the inclination of the Galaxy at any angle. This is convenient when compared with real observations.
- The oversampling in step (v) has two purposes: one is to ensure the validity of degrading - when degrading the SSP templates from the original spectral resolution to instrumental resolution (e.g., from MILES to MUSE), the σ of the Gaussian kernel to be convolved with the templates will be smaller than wavelength interval $\Delta\lambda$ without oversampling. In this case, the kernel function will be invalid and the

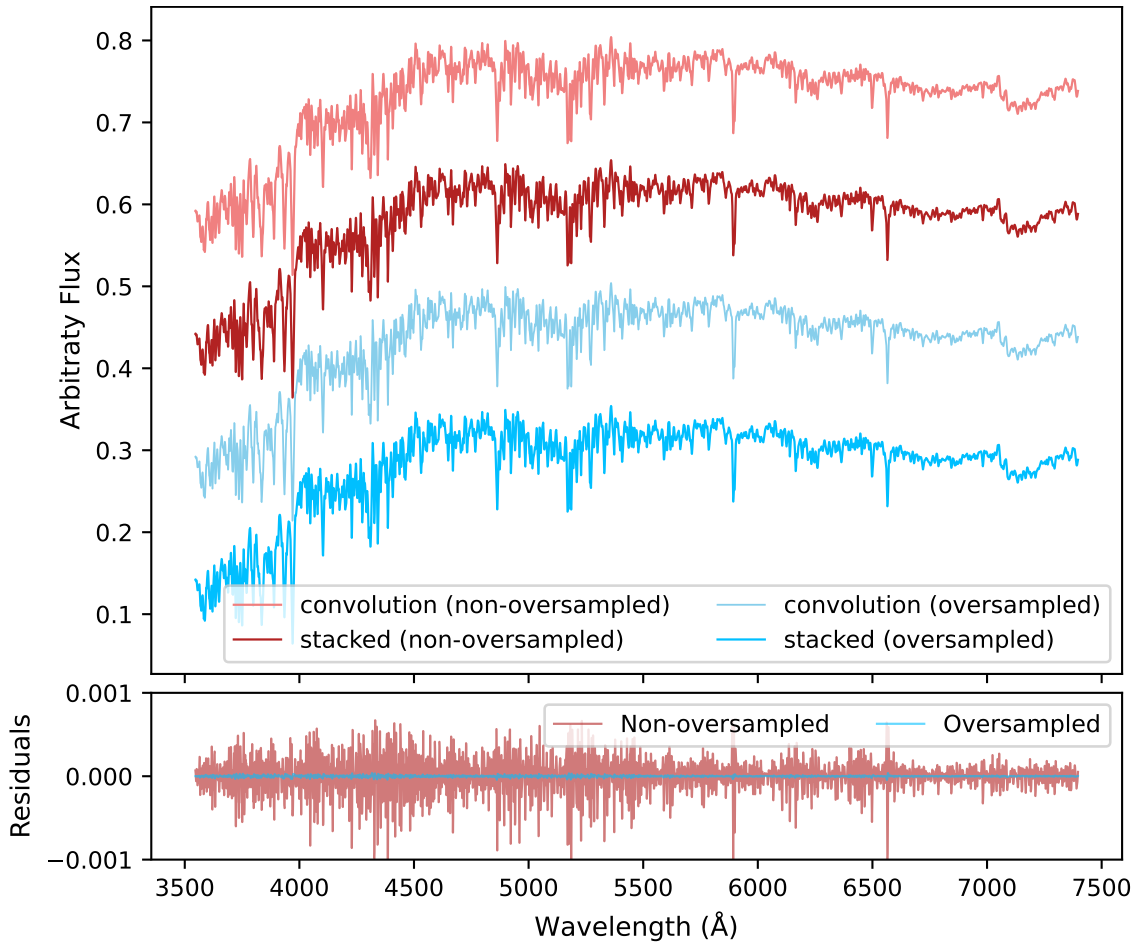

Figure 4.2: Comparison of the mock integrated spectrum generated by using the non-oversampled (in red and light-red) and the oversampled (in blue and light-blue) MILES templates. The light-color lines are spectra convolved by a Gaussian kernel with given (μ, σ), which is the way of pPXF measuring kinematics. The dark-color lines are spectra stacked by 2000 shifted spectrum with different line-of-sight velocities which follow a Gaussian distribution with the same (μ, σ), which is the manner of GalCraft. The bottom panel shows residuals of convolved and stacked spectra. By calculating the residuals of these two lines (shown in the residual panel), we find that oversampling can reduce the deviation between the convolved and stacked spectrum significantly by a factor of ~ 25.

degraded spectrum is still in the original resolution. Another reason is to minimize the sampling error when stacking the spectrum with different line-of-sight velocities. Fig. 4.2 shows an example mock integrated spectrum generated by using the non-oversampled (in red and light-red) and oversampled (in blue and light-blue) MILES SSP templates. The light-color lines are the spectra convolved with a Gaussian kernel by the given mean velocity and dispersion (μ, σ), which is the manner of pPXF during the kinematics measurements. The dark-color lines are stacked from 2000 spectra shifted by Gaussian distributed line-of-sight velocities using the same (μ, σ), which is the manner of GalCraft. By calculating the residuals of these two lines (shown in the residual panel), we find that oversampling can reduce the deviation between the convolved and stacked spectrum significantly by a factor of ~ 25.

- This package can select a portion of particles within the field of view (FoV) of the user-defined instrument to generate the datacube, rather than take a whole catalog into account. This helps to reduce computational expenses. Besides, users can also give a list of cube center coordinates in (l, b) or (ra, dec), this package can automatically generate all these cubes in one execution;
- Other than the datacube FITS file, this package also generates some by-products (optional), such as the array of mass/light-fraction and the number of particles of each spatial bin, the mass/light-weighted [M/H], [α/Fe], age and LOSVD distribution map in the Galactic coordinates, which are the expected answers that pPXF should provide from spectrum fitting, and the mock stellar catalogs of all the particles used to generate each datacube. These by-products are useful for the validation and comparison of measured results with true values in Section 5.1.

4.2.4 Estimate the Sampling Error

Compared to the real galaxy catalogs of the Milky Way (e.g. *Gaia*), one shortcoming of the E-Galaxia mock stellar catalog is the finite number of particles. Although there are $10 \sim 100$ times more particles (10^8) in the mock stellar catalog compared to MWAs in N-body/hydrodynamical simulations, it is inevitable that some spatial bins contain too few particles for robust measurements of their properties. Even for those spatial bins with more than $\sim 10^4$ particles, there is still noise in the parameter distributions compared with the predictions by the chemical evolution model of S21. This under-sampling introduces errors into the integrated spectrum, which we call "sampling error". One way to reduce the sampling error is to generate more particles in the mock stellar catalog. However, this will increase the computational expenses and the use of memory, and exceed most of the high-performance computers' limitations (~ 22 GB for a catalog with 10^8 particles, see details in Section 4.2.5).

Therefore, one has to ensure that the sampling error should be smaller than the Gaussian error which is added to the flux to mimic the real observation. In this way, it is safe to apply the data reduction pipeline on this mock datacube and allow pPXF to derive mathematically reasonable results.

To estimate the sampling error, the GalCraft code provides an option to apply bootstrapping methods. First, it randomly re-selects particles from the original mock stellar

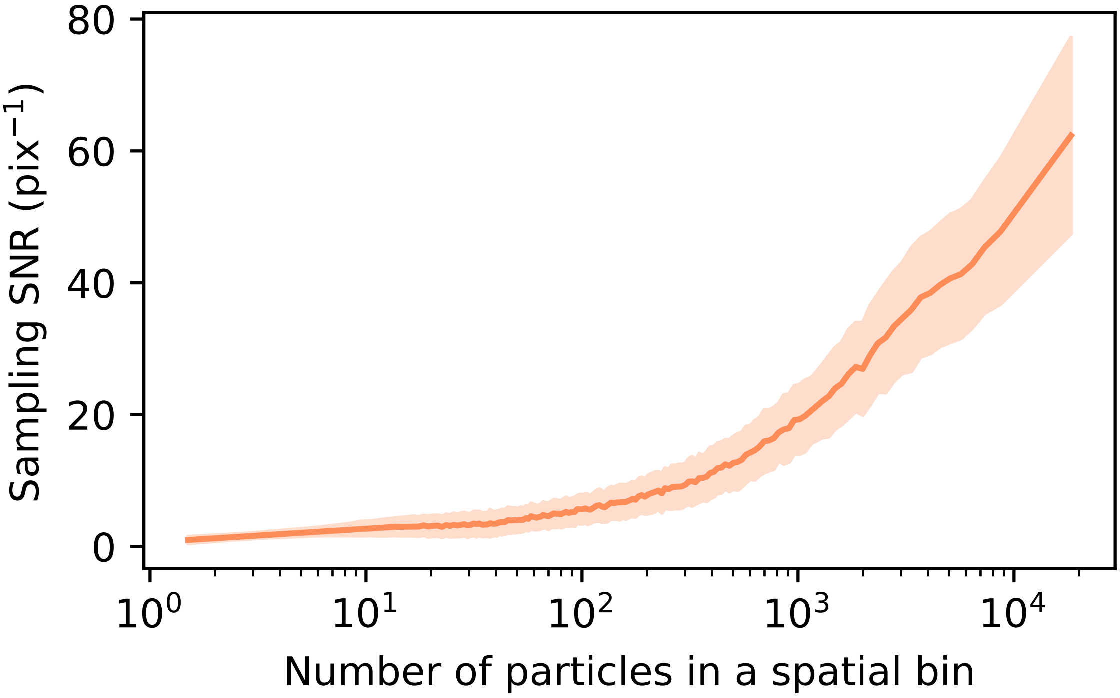

Figure 4.3: The sampling signal-to-noise ratio (SNR) of an integrated spectrum as a function of the number of particles of a spatial bin. The shaded region is the standard deviation of sampling SNRs for spaxels in the same number of particle bin. This is calculated by using original and bootstrapped mock datacubes later in Section 5.1.1. This figure provides an estimation of spectral sampling noise due to the limited number of particles.

catalog with the same particle number 20 times and obtains 20 bootstrapped catalogs. Then, GalCraft uses these catalogs to generate 20 bootstrapped datacubes. Next, the 20 bootstrapped fluxes for each spaxel are used to calculate the standard deviation which represents the sampling flux error of this spaxel. This sampling flux error will be set as the lower limit of the acceptable Gaussian noise when mimicking the real observation. Fig. 4.3 illustrates how the sampling error varies as a function of the number of particles in a given spatial bin. The shaded region is the standard deviation of sampling SNRs for spaxels in the same number of particle bin. We obtain this figure by using the original and bootstrapped mock datacubes later in Section 5.1.1. It can be seen that a spatial bin with $\sim 10^3$ particles can generate a spectrum with sampling SNR ~ 20 pix^{-1}, and a spatial bin with $\sim 10^4$ particles can generate a spectrum with sampling SNR ~ 45 pix^{-1}.

4.2.5 Computational Expenses and Multiprocessing Strategy

The execution time of GalCraft to generate mock datacubes depends mostly on the number of particles included in the instrument FoV and the spectral interpolation method. In general, the execution time is proportional to the number of particles used, and the "nearest" interpolation is three times faster than the "linear" interpolation. In this code, we apply `python-multiprocessing` techniques to speed up the computing. For a typical MUSE FoV containing 6×10^6 particles, the execution time spent with a 24-core CPU (2.50GHz) is ~ 1 hour. Based on this, the users can roughly estimate the execution time they will spend. If taking into account the 20 bootstrapped cubes, we need to run the code 20 more times. Therefore, it is highly recommended to run it on high-performance computers (HPC) or a Cluster where these 21 jobs can be executed simultaneously.

Chapter 5

Testing the Reliability of pPXF using Mock MUSE observations

Zixian Wang,[1,2] Michael R. Hayden,[1,2] Sanjib Sharma,[1,2] Jesse van de Sande,[1,2] Joss Bland-Hawthorn,[1,2,3] Sam Vaughan,[1,2] Marie Martig[4]

[1]Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW, 2006, Australia
[2]ARC Centre of Excellence for All Sky Astrophysics in Three Dimensions (ASTRO-3D)
[3]Miller Professor, Miller Institute, UC Berkeley, Berkeley CA 94720
[4]Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK

This chapter is reproduced from Chapter 3-6 of Wang et al. (2023), which is to be submitted to Monthly Notices of the Royal Astronomical Scocity as "The Milky Way in Context: Building an integral-field spectrograph datacube of the Galaxy". I wrote all of the text, produced all the figures, and wrote the codes for data analysis and interpretation. Sanjib Sharma provided E-Galaxia mock catalog. Joss Bland-Hawthorn, Sanjib Sharma, and Michael Hayden supervised the whole work. Sanjib Sharma and Michael Hayden contributed equally to this work. Jesse van de Sande, Sam Vaughan and Marie Martig helped with using pPXF, scientific interpretations and provided comments.

In this chapter, we test the recovery ability of broadly used spectral fitting code pPXF on kinematics and stellar population properties for an edge-on mock MW cube. In Section 5.1, we test the ability of spectral fitting tools to recover kinematics, stellar population parameters, and mass fraction distributions by applying it to a mock edge-on datacube and compare to the true values from E-Galaxia catalog. In Section 5.2, we discuss the causes of deviations between measured and true values and address potential reasons and future improvements on current data reduction pipelines to obtain better results. We also explore the effect of different fitting strategies and give some caveats of using GalCraft code. In Section 5.3, we talk about future improvements of this code along with SSP models and methods for measuring parameters and other scientific studies that we can do. We confirm that pPXF can identify the thin and thick disk structures using both kinematics and mass-weighted properties. However, PPXF struggles to recover the star formation history overestimating in the time intervals between 2 – 4 Gyr and 12 – 14 Gyr compared to the expected values, which is due to the regularization strategy and the spacing of template grids. Furthermore, we find systematic offsets in kinematics recovery potentially due to deficient spectral resolution and the variation of $V_{\rm LOS}$ with [M/H] and age through a line-of-sight. These systematics can affect the further process of dynamical modeling and interpretations of galaxy assembly histories. A summary is presented in Section 5.4.

5.1 Data Preparation and Parameter Measurements

In this section, we test the property recovery ability of pPXF by applying it on GalCraft generated mock MUSE cubes to measure kinematics ($V_{\rm LOS}$, σ, h_3, h_4), stellar population parameters (age, [M/H], [α/Fe]) and mass fraction distributions of different components in the same way as extragalactic observation data analysis. Then we compare the results with the input true values (calculated using particles' parameters). On the one hand, this test helps to investigate the consistency of parameters measured via broadly applied pipelines in other studies (e.g. Pinna et al. 2019a,b; Scott et al. 2021; Martig et al. 2021), which was not possible previously as the true values of external galaxies are unknown. On the other hand, it also helps to recognize the limitation of studying some of the galaxy evolution processes using current methods and provides direction for potential improvements in the future.

5.1.1 Mock Cube Generation for MUSE Instrument

We generate a mock MUSE observation by GalCraft, using the E-Galaxia catalog that contains 10^8 particles and assumes a distance of 26.5 Mpc and inclination of 86°. The selected distance and inclination are the same as the projection of the galaxy NGC 5746, which was observed by MUSE and has comprehensive analysis in Martig et al. (2021) (hereafter M21). We use MILES α-variable SSP templates (Vazdekis et al., 2015) with the BaSTI stellar isochrones (Pietrinferni et al., 2004) and Kroupa Universal IMF (Kroupa, 2001). The templates have 53 bins in age, 12 bins in [M/H] and 2 bins in [α/Fe] and we apply a "linear" interpolation to assign each particle a spectrum based on its age, [M/H] and [α/Fe]. Following procedures in Section 4.2.3, we generate two mock cubes focusing on the central (N_p = 61575676) and the disk (N_p = 7379847), respectively, as shown in Fig. 5.1. This observation strategy is also the same as M21. No extinction is applied for these cubes. The spectra are degraded to MUSE resolution (FWHM ~ 2.65Å). The total execution time spent by GalCraft on a 24-core CPU for these two cubes is ~ 20 hours. We also generate 2×20 bootstrapped cubes and use 16% and 84% percentiles to calculate the sampling error of each spaxel.

Next, we derive the flux error of the mock MUSE cubes. The flux error depends on many aspects but can be classified into two main categories: the observation conditions (seeing, air-mass, exposure time, etc.) and the instrumental properties (telescope aperture, system efficiency, dark current, read-out noise, etc). For simplicity, we ignore the sky conditions, dark current and read-out noise which only contribute a few percent to total received photons, and assume the MUSE SNR and received photons are defined by

$$\begin{aligned} \mathrm{SNR} &= \frac{f}{e_f} = \sqrt{N}, \\ N &= aft, \end{aligned} \tag{5.1}$$

where f is the flux of the target; e_f is the flux error; N is the received number of photons; t is the exposure time; a is an overall reaction of sky transmission, efficiency and telescope

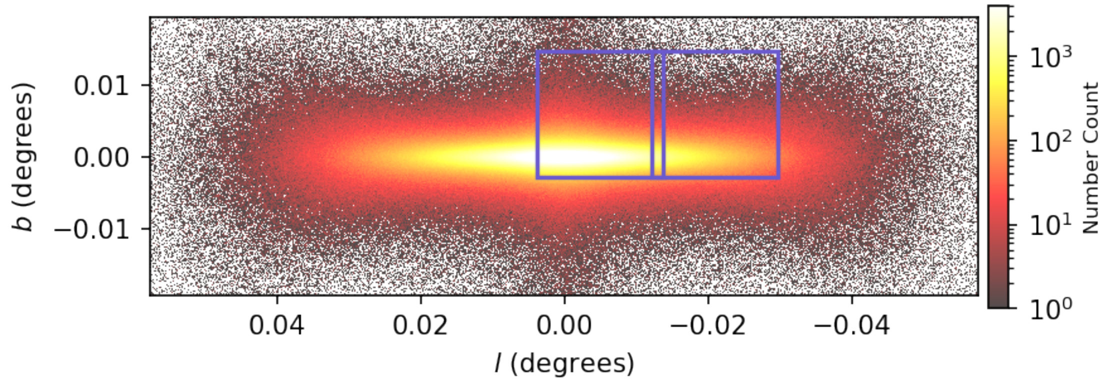

Figure 5.1: Demonstration of two mock datacube positions (in purple) relative to the number distribution of particles in E-GALAXIA mock stellar catalog. The model was set with a distance 26.5 Mpc and inclination 86°. This observation strategy is the same as the MUSE observation on NGC 5746 by Martig et al. 2021.

aperture. Therefore, a should be only dependent on wavelength for the same instrument. By substituting the above equations, the parameter a can be calculated using f, e_f and t from an observation by

$$a = \frac{f}{e_f^2 t}. \tag{5.2}$$

In this work, we take all the bulge and disk observations of NGC 5746 from M21 and fit equation 5.2 as a function of wavelength (λ) using a 4-degree polynomial, which is described by

$$\begin{aligned} a(\lambda) = {} & 4.34274826^{-18}\lambda^4 - 1.43263443^{-13}\lambda^3 + 1.61141240^{-9}\lambda^2 \\ & - 7.29164505^{-6}\lambda + 1.17213217^{-2}, \end{aligned} \tag{5.3}$$

where $a(\lambda)$ is in the unit of $1/10^{-20}$ erg cm^{-2} Å^{-1}. Next, we set the bulge and disk mock cubes to have an exposure time of 2323.06 s and 8357.68 s, respectively, same as MUSE observations of M21, and use the equation $a(\lambda)$ to estimate the flux error e_f of each spaxel. We find that for all spaxels with particle numbers greater than 200, e_f is on average larger than the sampling error by a factor of 1.47, by taking the median of e_f along all wavelengths. Therefore, we multiply e_f by this factor and this is equivalent to the bulge and disk datacube having the exposure time of 1161.53 s and 4178.84 s, respectively. Next, we use this error to add Gaussian noise to all the spaxels. Finally, the two mock cubes are stitched together.

5.1.2 Extracting Galaxy Properties

We apply the GIST pipeline (Bittner et al., 2019) on the stitched mock cube to measure the kinematics and stellar population parameters. The GIST pipeline combines all the tools needed to process the data and the user can obtain final results in a single execution. Here we use a modified version to implement some functionalities that the current public version (v3.0.0) does not have but are needed in this work. A detailed list of modifications is given in Appendix B.

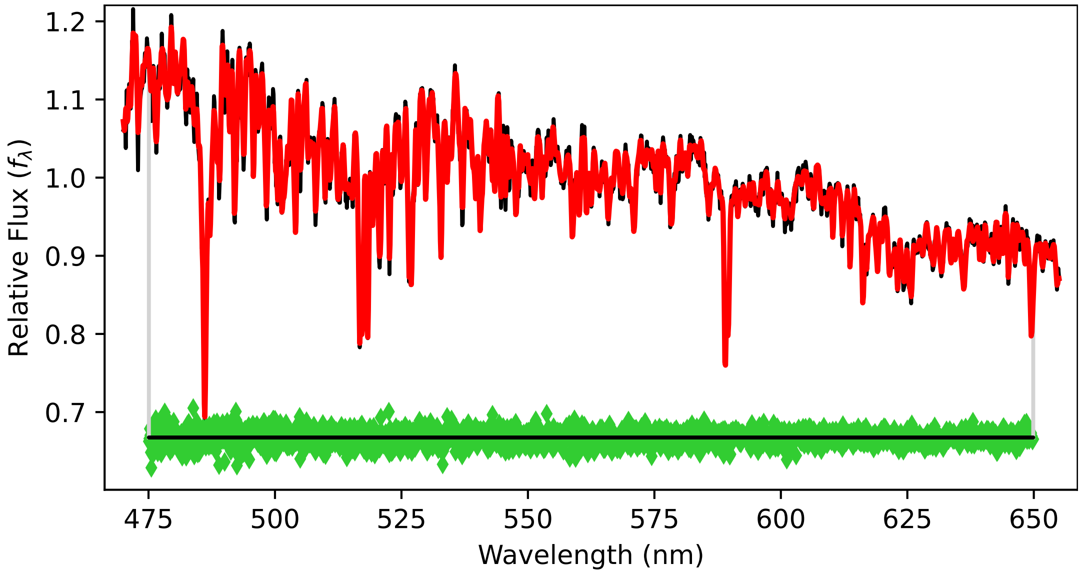

Figure 5.2: A spectral fit example of a Voronoi bin by pPXF in Section 5.1.2. The "observed" spectrum is in black while the best-fit spectrum is in red. Residuals are in green. The fitting area is within the gray lines on both sides.

We run the GIST pipeline in the following steps: First, we apply the Voronoi tessellation software (Cappellari & Copin, 2003) to spatially re-bin the mock cube and increase the SNR to 80 pix^{-1}, which results in 1080 Voronoi bins. Next, we apply the full-spectrum fitting method pPXF (Cappellari, 2017) to measure the stellar kinematics of each Voronoi bin. Here we use the same MILES α-variable templates (Vazdekis et al., 2015) as when we generated the mock cubes and degraded them to the same MUSE spectral resolution. Since there is no emission line or atmosphere effect, we use a wide wavelength range of $(4700, 6800)$Å to fit with the Voronoi binned spectra and remove the first and last 75Å because these two edges are impacted by procedures such as spectral oversampling, rebinning and Doppler shifting, etc. During the fitting, the MILES templates are convolved with a line-of-sight velocity distribution (LOSVD) described by a Gauss-Hermite function to match the Voronoi binned spectrum. We parameterize the LOSVD using four moments, which are mean line-of-sight velocity V_{LOS}, line-of-sight velocity dispersion σ, and the third- and fourth-order moments h_3, h_4. No regularization is applied in this process, and we include a fourth-order additive Legendre polynomial.

After measuring kinematics, we employ pPXF again to obtain the stellar population parameters for each Voronoi bin. We choose the same templates, spectral resolution, and fitting wavelength range as the previous step. Each template is normalized to one solar mass ($\mathrm{M}_\odot$). We use the LOSVDs (V_{LOS}, σ, h_3, h_4) measured in the last step as input, and fix them during the fitting to obtain the weight of each template. The best-fit spectrum is the weighted sum of all the templates. We first set no regularization during the fitting to obtain the initial χ^2, and find the maximum regularization parameter R_{max} allowed by increasing R until the χ^2 is increased by $\Delta\chi^2 = \sqrt{2N}$ to obtain the smoothest solution that is still compatible with the data within 1σ level following McDermid et al. (2015), where N is the number of wavelength pixels considered for the fit. Next, we choose $R = 5$ which is between 0 and R_{max} ($50 \sim 150$) to keep smooth solutions while still allowing for a variation on short timescales of the star formation, which will disappear if R is too large (see similar discussions in Pinna et al. 2019a and M21). Fig. 5.2 is an example of the fitting results for the spectrum of one Voronoi bin. Finally, using the weights, we can calculate the mass-weighted age, [M/H], [α/Fe], and mass fraction distributions.

In addition, we also apply the LS module in the GIST pipeline to compute line-strength indices of each Voronoi bin spectrum in the LIS system (Vazdekis et al., 2010) choosing the definitions provided at a spectral resolution of 8.4 Å. This routine was presented by Kuntschner et al. (2006) and Martín-Navarro et al. (2018). Next, given the relationship between templates' properties (age, [M/H], [α/Fe]) and line strengths, the measured line strengths can be matched to SSP-equivalent parameters by using the MCMC implementation from the package *emcee* (Foreman-Mackey et al., 2013). In this work, we follow M21 and use Hβ_o as an age indicator and Fe5015, Fe5270 and Fe5335, and Mgb to trace metallicity and [α/Fe]. The Monte-Carlo simulation is run for 15 times and each one uses 100 walkers and 1000 iterations to obtain uncertainties.

After measuring all the parameters, in the next sub-sections, we compare the kinematics, mass-weighted stellar population parameters, and mass fraction distributions of the mock cube with the true values to verify the recovery ability of pPXF.

5.1.3 Kinematic Maps

Fig. 5.3 shows the kinematics maps in four moments ($V_{\rm LOS}$, σ, h_3, h_4) of the mock MUSE cubes. The scale of the color bar is given in the second row of the upper left corner of each panel. We calculate the true values in the first row. First, for each Voronoi bin, we calculate the total flux of each particle in the pPXF fitted wavelength region. Then we plot flux-weighted $V_{\rm LOS}$ histogram distribution using all the particles included in this bin. Next, we fit this histogram with a Gauss-Hermite function and obtain four best-fit moments ($V_{\rm LOS}$, σ, h_3, h_4). This method is consistent to the definition of light-weighted kinematics which pPXF is expected to recover during the fitting process. The middle row is the results from pPXF and the bottom rows are residuals of the pPXF results and the true values.

In this figure, the kinematic moments obtained by pPXF have the same trend compared to the true values. Both show two kinematically distinct components: one is aligned to $y \sim 0$ and the thickness increases with x, which has larger absolute $V_{\rm LOS}$ and h_3, and smaller σ; the other has positive and negative y which is in a similar projected radius but vertically higher and has smaller absolute $V_{\rm LOS}$ and h_3, and larger σ. An anti-correlation of h_3 with $V_{\rm LOS}$ which are usually associated with disk-like components (e.g. Krajnović et al. 2008; Guérou et al. 2016; van de Sande et al. 2017) is clearly seen and are similar to MUSE edge-on galaxies studies of Pinna et al. (2019a,b) and M21. In Section 5.1.6, we will define these two components as thin and thick disk.

However, in the residual panels, all these four moments are showing systematic offsets. Compared to the true values, $V_{\rm LOS}$ from pPXF is around 17 km s^{-1} lower above and below the very thin mid-plane ($y \sim 0$) and shows an overestimation around $x \sim [10, 30]$ arcsec. And it becomes higher when going much above the disk ($y >$ 18 arcsec); Around the galaxy center, the residual of $V_{\rm LOS}$ also shows a continuous decrease from negative to positive x; σ is overestimated in the inner galaxy region but underestimated in the outer

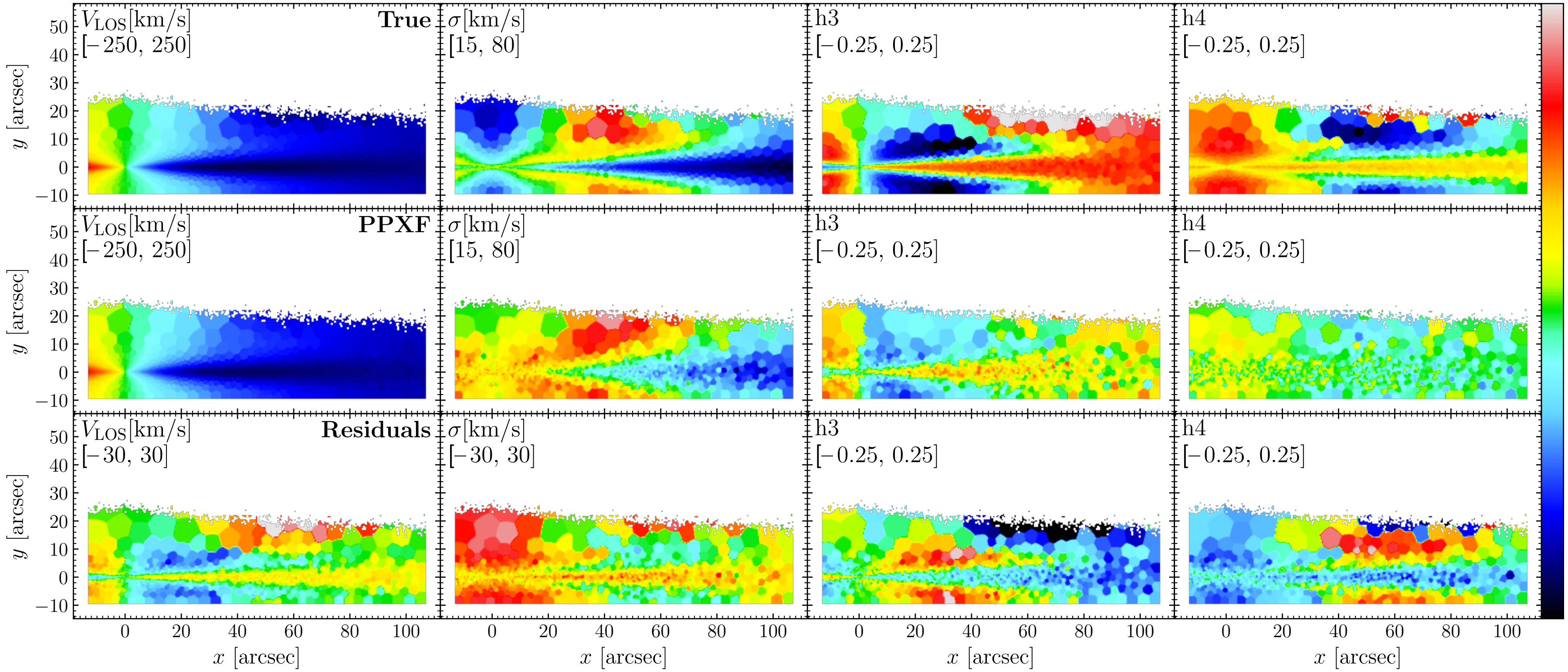

Figure 5.3: Stellar kinematic maps ($V_{\rm LOS}$, σ, h_3, h_4) of the mock MUSE cube generated in Section 5.1.1. The scale of the color bar is given in the second row of the upper left corner for each panel. **Top row**: True values calculated by fitting a Gauss-Hermite function with particles' velocity distribution weighted by their total flux in pPXF fitted wavelength region for each Voronoi bin. **Middle row**: Results from pPXF. **Bottom row**: Residuals of the pPXF results and the true values for each moment. The residual panels indicate results from pPXF have systematic offsets compared to true values.

region; h_3 is overestimated in regions of $x \sim [20, 60]$ arcsec and $y \sim [10, 25]$ arcsec and underestimated in the outer region of $x \sim [60, 110]$ arcsec; h_4 from pPXF has no clear signature of discrete components like σ and h_3 maps, which is also seen in real galaxies results (e.g., Pinna et al. 2019a,b and M21), but the true h_4 map does show kinematically discrete components. The clear structures in these residual panels indicate that it is not because of the fitting uncertainties. We will discuss this issue in detail and provide our solutions in Section 5.2.1.

5.1.4 Mass-weighted Stellar Population Parameter Maps

Fig. 5.4 shows the mass-weighted age, [M/H] and [α/Fe] maps of the mock MUSE cubes. The first row is the true values by calculating the particles' median age, [M/H] and [α/Fe] for each Voronoi bin, which are equivalent to mass-weighted values since each particle is weighted as 1 $M_\odot$. The second row is the results from pPXF with regularization $R = 5$. The third row is the results from pPXF with $R = R_{max}$. The last row shows the residuals of the pPXF results with $R = 5$ and the true values. The overall distributions of these three parameters obtained by pPXF are very close to the true values. This confirms the reliability of using pPXF to measure galaxy chemical information. Especially, the [α/Fe] map from pPXF with $R = 5$ indicates the capability of pPXF to identify [α/Fe]-rich and [α/Fe]-poor populations in the thick and thin disk, respectively, even though only two bins are available. The residuals of [α/Fe] from pPXF with $R = 5$ and the true values are flat and no systematic pattern is found.

However, minor offsets appear in age and [M/H] residual panels: age from pPXF is overestimated in the outer regions while [M/H] is overestimated in the inner center regions. This means the age gradient from pPXF is underestimated but the [M/H] gradient is overestimated. In addition, the [α/Fe] distribution from pPXF results with $R = R_{max}$ is almost uniformly high and much larger than the true values for all the Voronoi bins. This is because when R is very large (~ 100), the pPXF algorithm forces the result to have very smooth template weights in three parameter dimensions (age, [M/H], [α/Fe]). Since there are only two [α/Fe] grids, regularization will force them to have similar weights to achieve smoothness and does not permit large deviations (e.g., more than 0.1 dex). Therefore, it will be challenging to identify [α/Fe] bimodality. Results from pPXF with $R = R_{max}$ also show much underestimation for age gradients than results with $R = 5$. The age and [M/H] gradients are essential properties to help understand the star formation and chemical enrichment processes. A wrong choice of regularization will then easily lead to wrong conclusions. We will explore these offsets in more detail in Section 5.1.6 using mass fraction distributions and the effect of regularization in Section 5.2.3.

5.1.5 SSP-equivalent Maps from Line-Strength Indices

Fig. 5.5 shows the SSP-equivalent age, [M/H] and [α/Fe] maps of the mock MUSE cubes measured by line-strength indices. This figure shows that the main structures we derived from pPXF are also recovered by the line-strength analysis with consistent trends. In the age panel, young populations are in the thin disk, and relatively old populations are in the thick disk and above the bulge. In the [M/H] panel, we see the metallicity gradient from the inner center to the outer Galaxy. In the [α/Fe] panel, we see the α-rich bins in

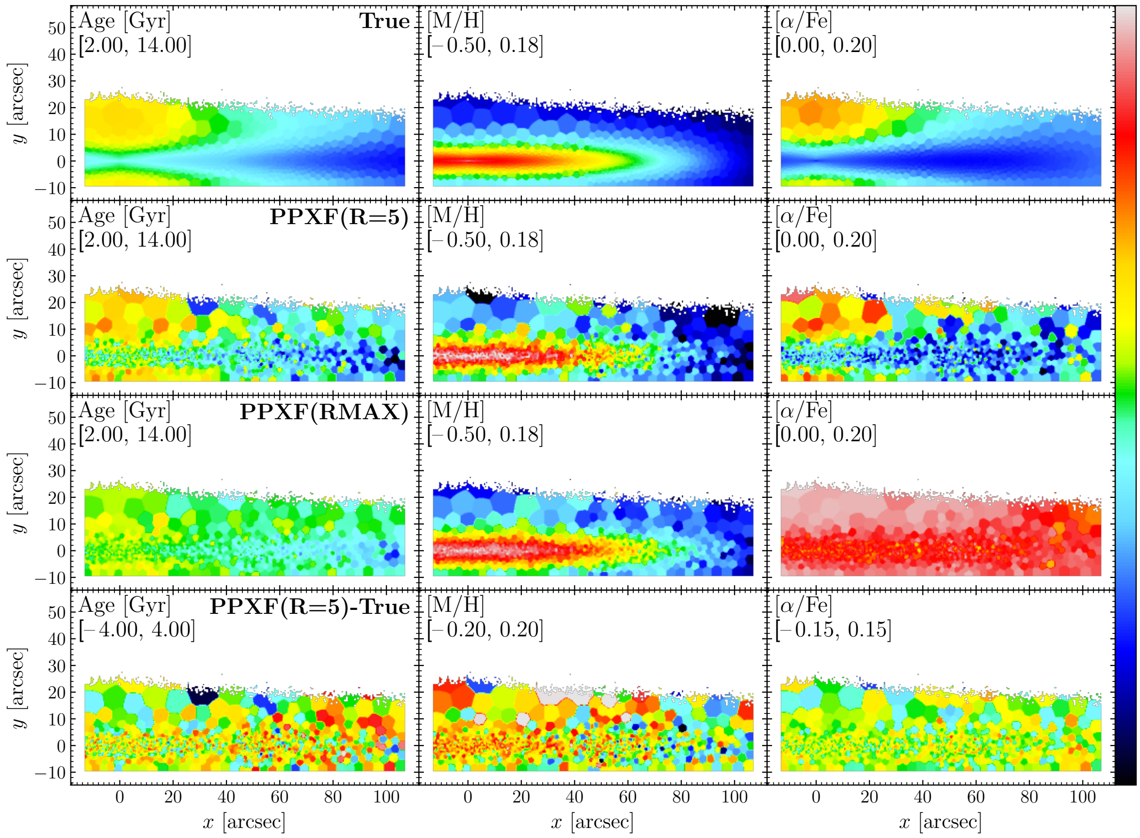

Figure 5.4: Stellar population maps (age, [M/H] and [α/Fe]) of the mock MUSE cube generated in Section 5.1.1. The scale of the color bar is given in the second row of the upper left corner for each panel. **First row**: True values calculated by the mass-weighted average of particles' age, [M/H] and [α/Fe] for each Voronoi bin. **Second row**: Results from pPXF via full-spectrum fitting with regularization $R = 5$. **Third row**: Results from pPXF with $R = R_{max}$, calculated using strategies of McDermid et al. 2015. **Last row**: Residuals of pPXF results with $R = 5$ and the true values. This figure indicates that pPXF results with proper regularization can identify different galaxy components by their stellar population parameters, which are consistent with true values. Applying $R = R_{max}$ during the fitting can smooth the distribution and make different [α/Fe] components indistinguishable. According to residuals, stellar age in the outer region and [M/H] in the center region are overestimated.

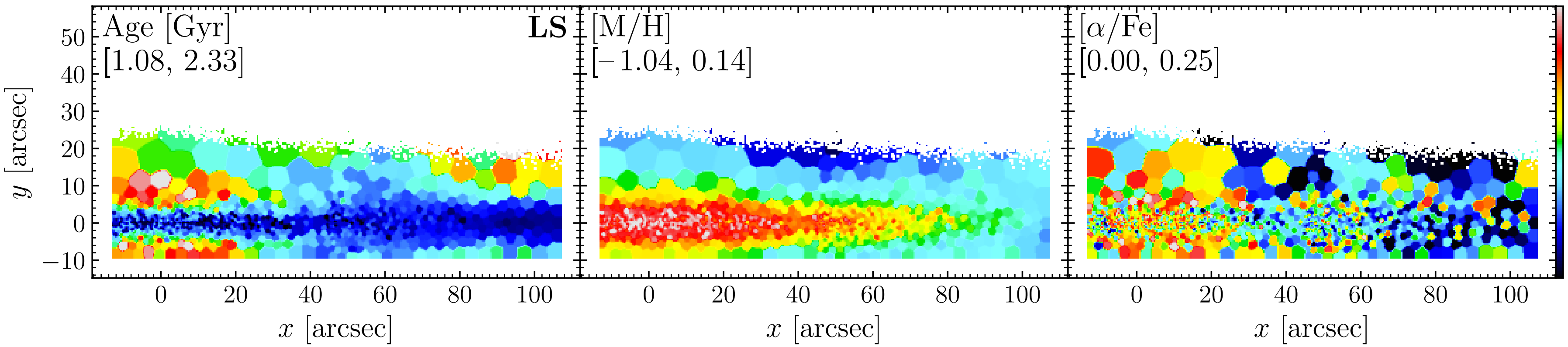

Figure 5.5: SSP-equivalent age, [M/H] and [α/Fe] maps of the mock MUSE cube generated in Section 5.1.1. The parameters are measured by line-strength indices. The scale of the color bar is given in the second row of upper left corner for each panel. These maps are qualitatively similar to the ones obtained from pPXF (Fig. 5.4) and different galaxy components can also be identified. But the parameter ranges differ from pPXF results and the true values.

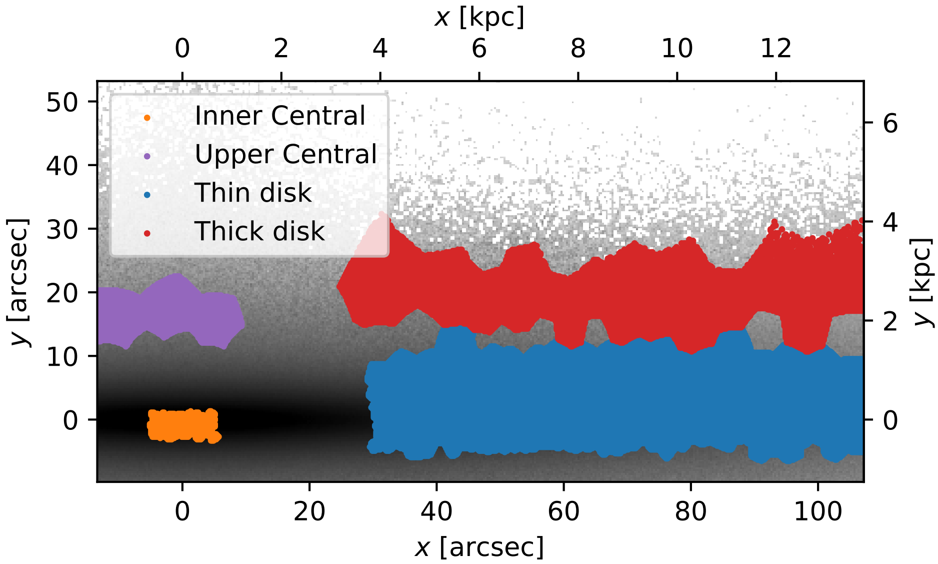

Figure 5.6: Demonstration of the projected thin disk (blue), thick disk (red), upper central (purple), and inner central (orange) regions on top of a grey-scale image of the mock datacube generated in Section 5.1.1. These regions will be used to study the mass fraction distribution of different galaxy components later.

the center and α-poor bins in the outer region. The main difference compared with pPXF results is that the age panel shows a very low range of [1 ~ 2.5] Gyr. This is also seen in M21 (Fig. B1) and because of the Balmer line indices being dominated by young stars. Therefore, the SSP-equivalent ages only reflect the fraction of stars formed within the past Gyr (Serra & Trager, 2007; Trager & Somerville, 2009). The SSP-equivalent [M/H] and [α/Fe] range are much closer compared to pPXF results because young populations do not contribute much to the metal lines, which is also indicated in M21. This figure confirms that both line-strength indices and pPXF analysis can identify α-rich (thick disk) and α-poor (thin disk) populations.

5.1.6 Mass Fraction Distributions of Different Galaxy Components

In addition to calculating the mass-weighted parameters, we can also study the mass fraction distribution of stellar populations along the age and [M/H] space. This is done by using weights of templates from pPXF. Since the templates are weighted by 1 $M_\odot$, the weights represent the stellar population masses. Therefore, we can study the mass distribution of any component of the Galaxy. M21 employed multiple components fitting to a *Spitzer* 3.6 micron image to obtain the boxy/peanut bulge, nuclear disk, and thin/thick disk regions. In Fig. 5.6, we artificially select similar regions based on the location of different components of M21, and name them “upper central”, “inner central”, “thin disk” and “thick disk”, as shown in different colors. We call them “upper central” and “inner central” because we do not have a boxy/peanut bulge and nuclear disk in the chemical evolution model of S21. For each component, the mass weights of all the Voronoi bins are combined together to represent its mass fraction distribution.

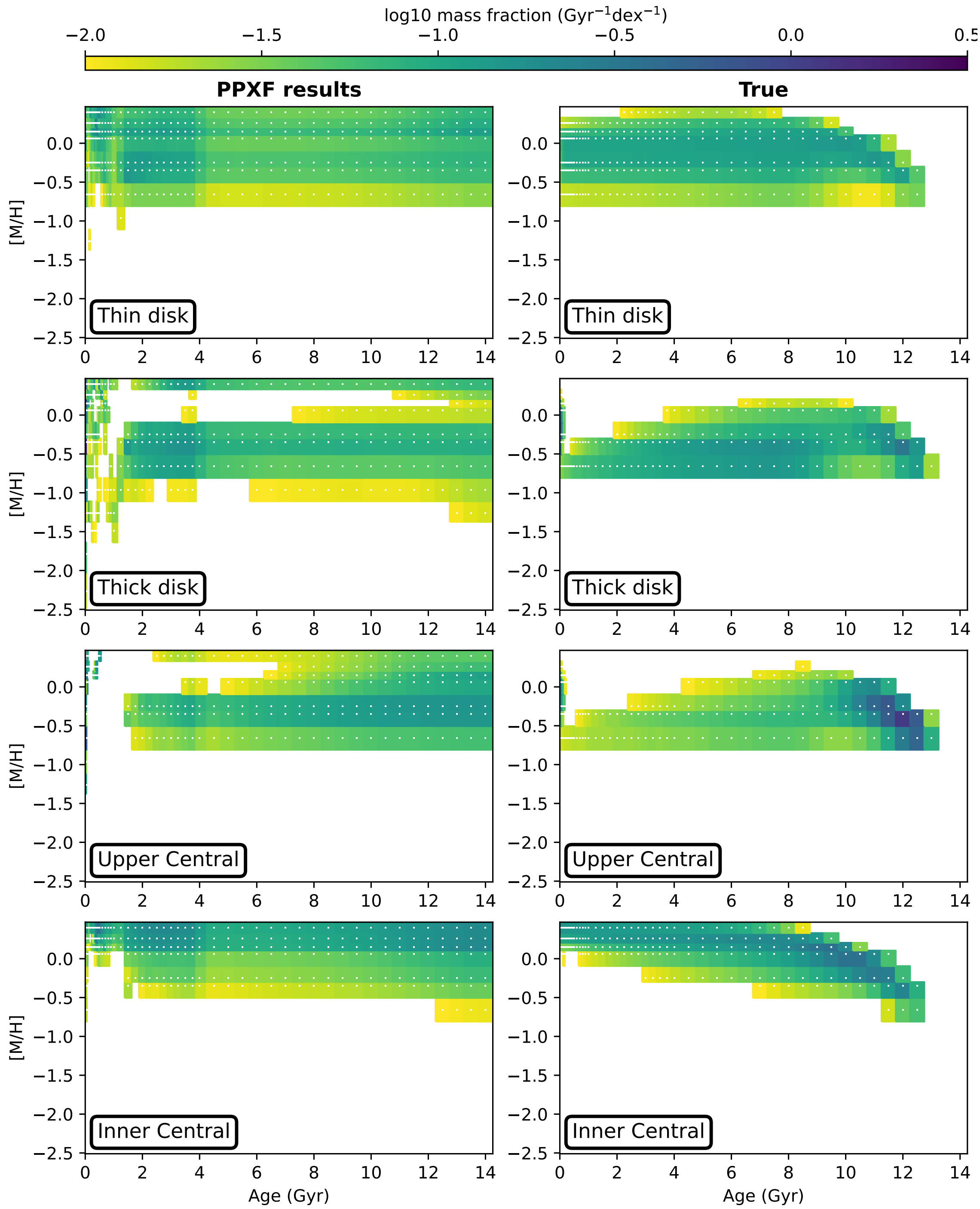

Figure 5.7: Mass fraction distributions of the thin disk, thick disk, inner central, and upper central of the mock datacube generated in Section 5.1.1. Mass fraction weights are normalized to one for each panel. **Left-hand column**: Results from pPXF with R = 5. **Right-hand column**: True values calculated using particles' properties in mock E-Galaxia catalog. The color bar is shown on the top. We only plot the weights above 0.001 Gyr^{-1}dex^{-1}. This figure indicates results from pPXF are qualitatively consistent with the true values. However, there is an overdensity in 2 $\sim$ 4 Gyr and populations around 12 Gyr are smoothed towards older, more metal-rich regions, which are not seen in the true values.

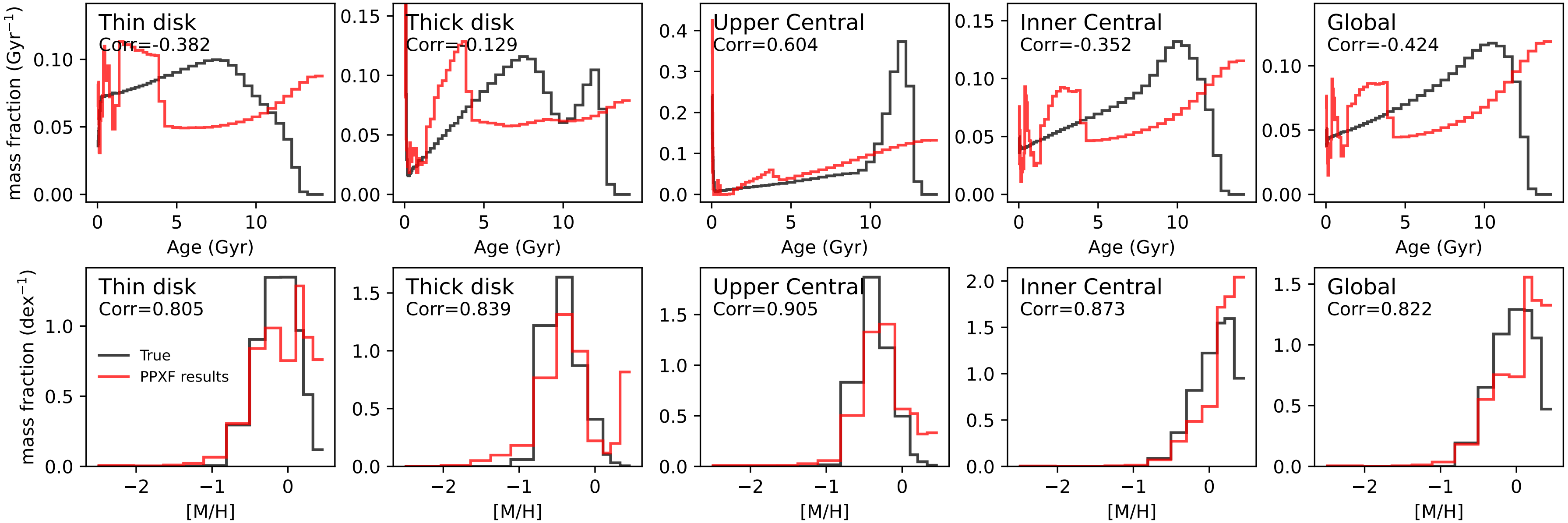

Figure 5.8: Age (top row) and [M/H] (bottom row) distributions of the thin disk, thick disk, inner central, and upper central of the mock datacube generated in Section 5.1.1. These panels are obtained by integrating the mass fraction distributions in Fig. 5.7 along [M/H] and age axis, respectively. Red lines are results from pPXF with $R = 5$. Black lines are the true values calculated using particles' properties in the mock E-Galaxia catalog. The correlation coefficient of these two lines is texted in the left top corner. Especially, the correlation of age distributions is calculated after evenly re-binning the age grid into linear space. The overall trends from pPXF are consistent with the true values for [M/H]. However, same as Fig. 5.7, the mass fraction is overestimated in $2 \sim 4$ Gyr and $12 \sim 14$ Gyr and underestimated in $5 \sim 12$ Gyr for all the components, and the coefficients indicate inconsistency for age distributions. We also find peaky features in regions less than 1 Gyr. The mass weights of the most metal-rich bin are overestimated for all the components.

Fig. 5.7 shows the mass fraction distributions of these four components, respectively. The weights are normalized to one for each panel. The left column is the results from pPXF with $R = 5$, and the right column shows the true values calculated using particles' properties in the mock stellar catalog. We only plot the weights above 0.001 $\mathrm{Gyr}^{-1}\mathrm{dex}^{-1}$. For the thin disk, true values indicate a rapid metallicity enrichment history $\sim$ 10 Gyr ago, and it slowly increases later on. While from pPXF results, mass fraction weights are dominated by two regions, which are relatively young ($2 \sim 4$ Gyr) and old ($12 \sim 14$ Gyr) stellar populations and the metallicity enrichment trend is indistinguishable. The same features are also seen in other Galaxy components. For the overdensity region of $2 \sim 4$ Gyr, we think this is due to the decrease in templates age bin size from old to young population. For the region of $12 \sim 14$ Gyr, we think this is because at the level of SNR ~ 80 pix^{-1}, the differences between old and metal-rich templates are smaller than the Gaussian noise added on the spectra, causing the degeneracy in this parameter region. More details will be discussed in Section 5.2.2. For the thick disk, true values also show the metallicity enrichment trend, but the populations born less than 1 Gyr are more metal-poor than the thin disk; The upper central shows a similar trend with the thick disk but is more dominated in the old populations, but this domination is smoothed out in the pPXF results; For the inner central, true values are showing a clearer chemical enrichment trend and there is no new population born with $[M/H] < -0.2$ dex in the young region, while the pPXF results show again two dominated regions, one at $\sim$ 2 Gyr and the other at $12 \sim 14$ Gyr. Therefore, except for the overestimation in the young and old regions, the mass fraction distributions of pPXF are generally in agreement with true values.

In Fig. 5.8, we integrate the mass fraction distributions in Fig. 5.7 along the two axes and derive age and [M/H] distributions for each component. The top panels are mass distributions as a function of age which is the definition of star formation history or star formation rate, and the bottom panels are as a function of [M/H]. Results from pPXF are in red lines and the true values are in black lines. For each panel, we calculate the correlation between these two lines to quantify their similarity. Particularly, for age distributions (top row), we first re-bin the age in an evenly spaced grid in linear space and then calculate the correlations. This is to avoid bias caused by the uneven distribution of the original age grid. For the age distributions, same as Fig. 5.7, compared to the true values, pPXF results of all the components demonstrate an overestimation of weights in the ranges of $2 \sim 4$ Gyr and $12 \sim 14$ Gyr and underestimation in the range of $4 \sim 11$ Gyr. The underestimated regions seem just to compensate the overestimated regions. And the thin disk, inner central, and global panels show a peaky feature with age $<$ 1 Gyr. For the [M/H] distributions, pPXF results are consistent with true values for most regions, as indicated by the correlation coefficients. However, weights in the metal-rich region are overestimated by pPXF. Other than that, the overall trend of results from pPXF is consistent with true values.

Therefore, we can confirm pPXF can recover similar trends of 2-D mass fraction distributions for different components, but when integrating to 1-D age distributions, pPXF results and the true values are different, and the weights are overestimated in the age of $2 \sim 4$ Gyr and $12 \sim 14$ Gyr, and the most metal-rich regions. According to correlation coefficients, [M/H] distributions are more consistent with the true values than age distributions. We will discuss this effect in detail in Section 5.2.2.

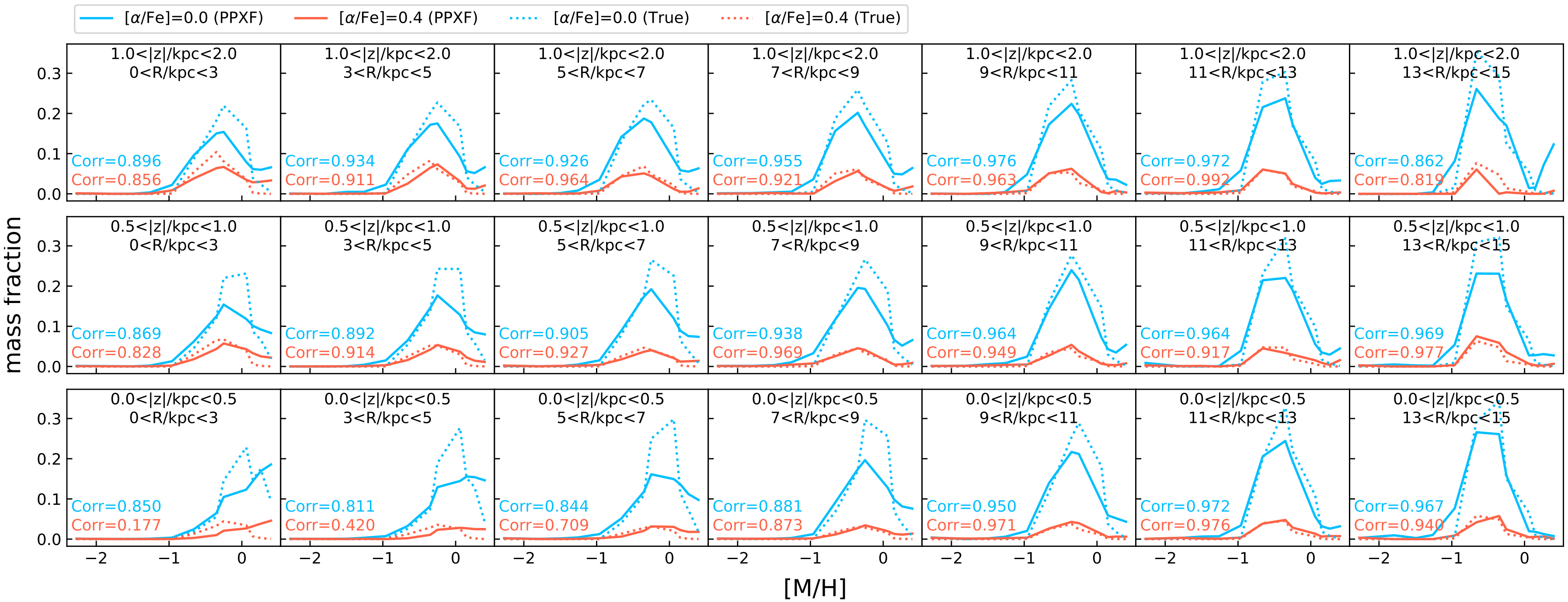

Figure 5.9: [M/H] distributions for different [α/Fe] components distributed along projected R and $|z|$ of the mock cube generated in Section 5.1.1. For each panel, the total mass fractions of true (dotted lines) and pPXF (solid lines) are normalized to 1, respectively. Blue lines represent mass fraction distributions for [α/Fe] = 0.0 dex while red lines represent the component with [α/Fe] = 0.4 dex. Solid lines represent the results from pPXF and dotted lines are the true values. This figure is similar to the [α/Fe]-[Fe/H] distributions in Hayden et al. 2015 and Scott et al. 2021. The correlation coefficient of the dotted and solid lines are texted in the same color in the left bottom panel. For most of the panels, pPXF results are consistent with the true values. The discrepancy happens on the inner central and thin disk, where both lines show a mass fraction underestimation at [M/H] $\sim$ 0.0 dex and overestimation at the most metal-rich regions.

5.1.7 Distributions of $[\alpha/\mathrm{Fe}]$-$[\mathrm{M/H}]$ along Different Galaxy Locations (R, z)

According to S21, different radial migration and kinematic rates will cause different fractions of stars in the distinct $[\alpha/\mathrm{Fe}]$-rich and $[\alpha/\mathrm{Fe}]$-poor sequences. A recent study by Scott et al. (2021) derived this distribution from an external galaxy UGC 10738 using MILES α-variable templates (Vazdekis et al., 2015) and concluded it has similar bimodality distributions to the MW. Since we know the true values from the mock stellar catalog particles, here we explore the recovery ability of pPXF on measuring the changes of this bimodality at different projected R and $|z|$. Different from Scott et al. (2021) where they compared integrated values of UGC 10738 with individual values of the MW, our integrated-to-integrated value comparison is more direct and there is no systematic bias due to different methodologies.

The results are shown in Fig. 5.9 where we separate different locations in the same way as Hayden et al. (2015). For each panel, the total mass fraction given by true (dotted lines) and pPXF (solid lines) is normalized to 1, respectively. Blue lines are mass fraction distributions for $[\alpha/\mathrm{Fe}] = 0.0$ dex while red lines represent $[\alpha/\mathrm{Fe}] = 0.4$ dex. The overall trends for $[\alpha/\mathrm{Fe}]$-rich and $[\alpha/\mathrm{Fe}]$-poor sequences from pPXF are consistent with the true values for most of the panels, which indicates both sequences are well recovered by pPXF, as also shown by correlation coefficients. However, the discrepancy happens in some regions such as the inner central and thin disk, where both the blue and red lines show a mass fraction underestimation at $[\mathrm{M/H}] \sim 0.0$ dex and overestimation at the most metal-rich regions. The latter issue is also seen in Fig. 5.8 and mentioned by Scott et al. (2021), where they think this is due to the lack of intermediate $[\alpha/\mathrm{Fe}]$ templates. However, we do not see the tail at the most metal-rich regions in the true values for both $[\alpha/\mathrm{Fe}]$ sequences, indicating it is more likely due to the degeneracy of metal-rich populations whose spectra are very similar that pPXF will obtain more uncertain results. But we also emphasize that more $[\alpha/\mathrm{Fe}]$ bins in the templates are also necessary to obtain detailed distributions of this bimodality and make it more feasible to identify the effect of different radial migration and kinematic heating efficiency on the relative fractions of these two sequences.

5.2 Discussion

5.2.1 Systematic Offsets in the Kinematics Recovery

In Fig. 5.3, we show the deviations between the kinematics map from pPXF and true values. To better illustrate the offsets of each moment, we plot their residuals as a function of true velocity dispersion σ_{true} in Fig. 5.10, in which the y-axis value is calculated by pPXF fitted value subtracting true value. Each data point represents one Voronoi bin. The blue dotted lines represent the instrumental dispersion (σ_{inst}) in MUSE spectral resolution. We also plot the red zero-line in red to guide the eyes. The first row demonstrates the results in Fig. 5.3, and it clearly shows the systematic offsets for each moment: ΔV_{LOS} increase with σ_{true}, and σ is overestimated for most of the Voronoi bins; as for h_3 and h_4, the residuals have a slight positive slope with σ_{true}.

From Fig. 5.3, we find that there are several Voronoi bins having both large h_3 and h_4

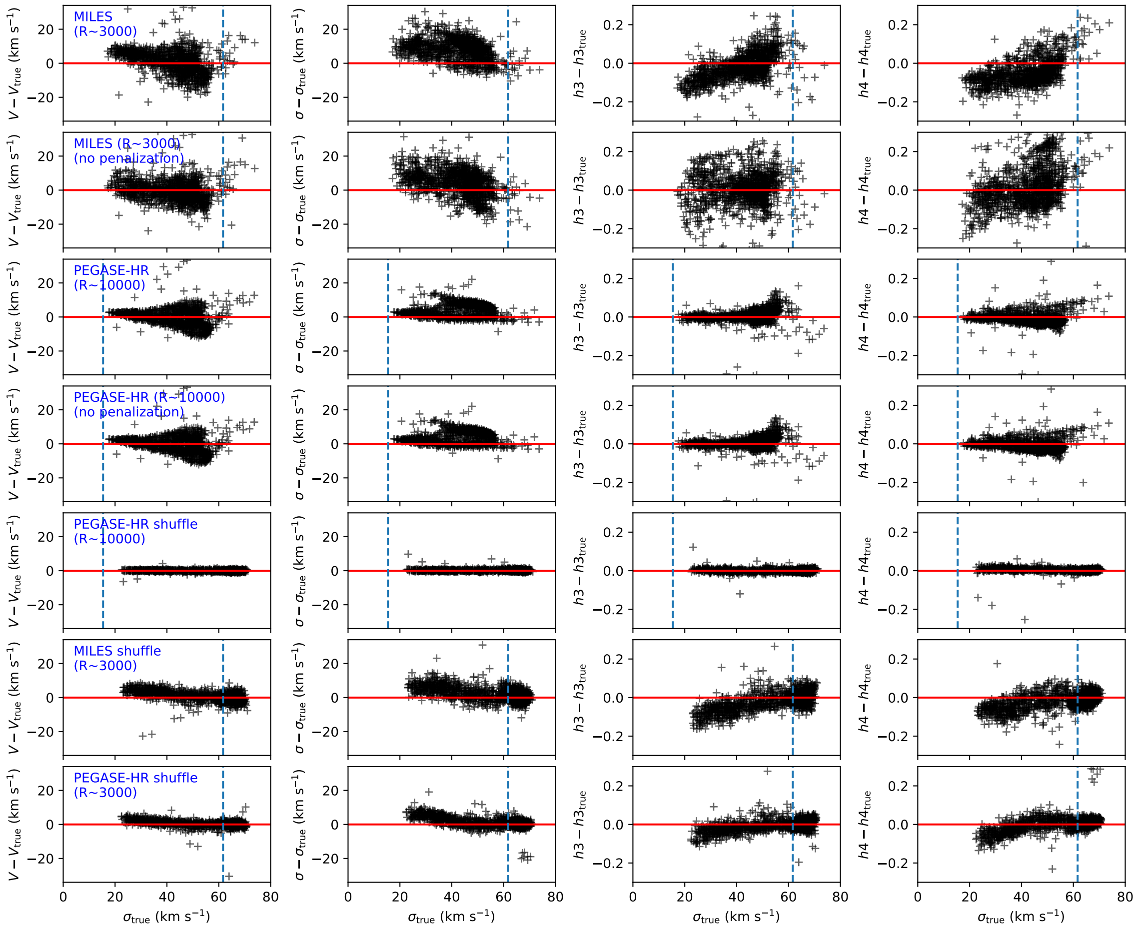

Figure 5.10: Stellar kinematics recovery of four Gauss-Hermite moments $\Delta(V_{\rm LOS}, \sigma, h_3, h_4)$ as a function of true velocity dispersion. Each data point represents a Voronoi bin. The red line represents zero to guide the eyes. The vertical dashed line indicates the instrumental dispersion $\sigma_{\rm inst}$. **First row**: Results from pPXF using MILES templates in MUSE spectral resolution ($\sigma_{\rm inst} \sim 62$ km s^{-1}), same as the residual panels in Fig. 5.3. **Second row**: Similar to the first row but results are fitted by pPXF with no penalization on h_3 and h_4. **Third row**: Results from pPXF on a mock cube similar to Section 4.2 but generated and fitted using PEGASE-HR (Le Borgne et al., 2004) templates ($\sigma_{\rm inst} \sim 16$ km s^{-1}). **Fourth row**: Similar to the third row but penalization was turned off during pPXF fitting. **Fifth row**: Similar to the second row but on a mock cube generated using particles' velocities randomly shuffled for each Voronoi bin. **Sixth row**: Similar to the first row but on a mock cube generated using particles' velocities randomly shuffled for each Voronoi bin. **Seventh row**: Similar to the fourth row but all the analysis is in MUSE resolution to eliminate the effects of different SSP templates. This figure indicates that both the spectral resolution and the relation between LOSVD and age, [M/H] affect the kinematics recovery.

values (e.g., central region and the thick disk), which can be penalized by pPXF during the fitting and led to incorrect measurements. This effect of penalization is also shown in real SAMI galaxy measurements (van de Sande et al., 2017). Therefore, we turn off the penalization of high-order moments and re-measure the kinematics, and show the results in the second row of Fig. 5.10. A spatially distributed map can be also seen in Fig. C.1 of the Appendix C. After turning off penalization, both h_3 and h_4 from pPXF are visually closer to the true values in Fig. C.1. However, the second row of Fig. 5.10 still shows similar trends of $V_{\rm LOS}$ and σ, and h_3 and h_4 are very scattered, which indicates penalization is not the cause of the systematic offsets.

One other reason could be the low spectral resolution of MILES templates. As shown in the first row of Fig. 5.10, the $\sigma_{\rm true}$ for most of the Voronoi bins are smaller than the MUSE instrumental dispersion $\sigma_{\rm inst}\sim$ 62 km s^{-1}. In this case, the recovery of kinematics will be uncertain. This has been pointed out in Fig. 3 of Cappellari (2017), and the only way to avoid it is to increase the spectral resolution of the instrument. The explanation is when $\sigma < \sigma_{\rm inst}$, during pPXF fitting, the broadening of one sharp feature is less than the distance to its nearby pixels, then the nearby pixels have minor changes and this confuses the algorithm to measure σ correctly and h_3 and h_4 will go towards zero because there are also not enough pixels to identify the skewness.

To remove the effect of low spectral resolution, we tested the kinematics recovery again by using PEGASE-HR templates (Le Borgne et al., 2004) generated by Kroupa IMF and PADOVA 1994 isochrones, which has a higher spectral resolution (FWHM = 0.55Å) than MILES. The instrumental velocity scale of PEGASE-HR is $\sigma_{\rm inst}\sim$ 15.6 km s^{-1}, which is smaller than $\sigma_{\rm true}$ of our Voronoi bins. We repeat all the procedures in Section 5.1.2 to generate new cubes using PEGASE-HR, and then apply them to the GIST pipeline to measure the kinematics. We keep the original PEGASE-HR spectral resolution throughout the whole process. The third row of Fig. 5.10 shows the kinematics recovery using PEGASE-HR (also see the kinematics map in Fig. C.2). Compared to the first row, the residuals of each panel are slightly better or clearer when using higher spectral resolution templates. However, all these four moments from pPXF still have the same systematic offsets to the true values. Most improvements are for h_3 and h_4 with less bias to zero values because higher spectral resolution helps identify skewness and kurtosis. We also show the results by turning off penalization in the fourth row (also see the kinematics map in Fig. C.3), and find the h_3 and h_4 residuals are identical. Therefore, turning off penalization helps with low-spectral-resolution kinematics measurements, but has no effect on higher-resolution spectra. The higher spectral resolution improves the kinematics recovery, but does not fix the systematics offsets.

Fundamentally, during pPXF fitting process, SSP templates are loaded and convolved with a Gauss-Hermite function by FFT to match the observed spectrum (Cappellari, 2017), then it returns one series of LOSVD moments which are the measured kinematics best fitted to the observed spectrum of each Voronoi bin. Therefore, pPXF assumes all the templates having the same kinematics ($V_{\rm LOS}$, σ, h_3, h_4) in a line-of-sight, i.e., stellar populations with different [M/H] and ages are assumed to have the same LOSVD. But in real galaxies, due to the joint process of chemical enrichment and dynamical movements of stars, populations with different ages and [M/H] are different in LOSVDs. The effect should not be severe for face-on galaxies because the line-of-sight is perpendicular to the main disk,

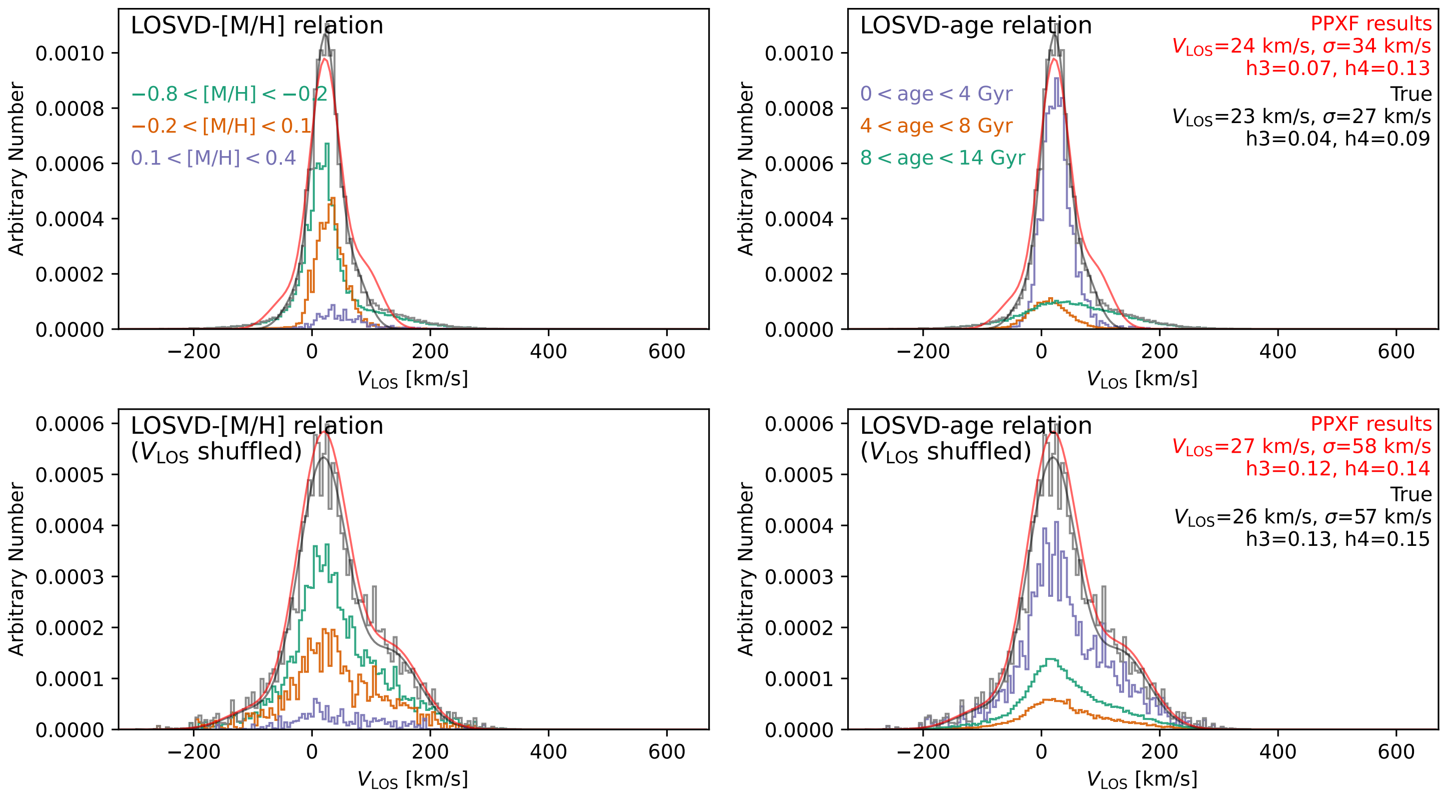

Figure 5.11: LOSVD of particles with different ages and V_{LOS} in one Voronoi bin. Particles are divided into three [M/H] and age groups, then their V_{LOS} histograms are plotted on the left and right panels in different colors, respectively. The top rows are results using the original E-Galaxia catalog. The bottom rows are results using the E-Galaxia catalog but randomly shuffle the particles' V_{LOS} in each Voronoi bin. The red lines are pPXF results for mock cubes generated and fitted using PEGASE-HR templates in PEGASE-HR spectral resolution (FWHM = 0.55Å). The black lines are the true values calculated by fitting the Gauss-Hermite function with particles' LOSVD in black histogram. The kinematic parameters for these two lines are texted on the right top corner. This figure indicates LOSVD changes with age and [M/H], which disagrees with pPXF assuming LOSVD is a constant in a Voronoi bin. After shuffling the V_{LOS}, this relationship can be artificially eliminated, and pPXF can measure kinematic parameters accurately.

so particles' age and [M/H] in this LOS should be identical. But it is non-negligible for edge-on projection because metallicity and age gradients along the disk are the strongest. Therefore, $V_{\rm LOS}$ should change with [M/H] and age, which disagrees with the pPXF assumption.

To investigate it in our mock stellar catalog, we select one Voronoi bin and split the included particles into three groups using [M/H] and age, respectively. Then, we plot the LOSVDs of these groups of particles, weighted by each particle's total flux, and show them in the first row of Fig. 5.11. In the top-left panel, particles with $-0.8 < [\mathrm{M/H}] < -0.2$ dex has the most skewed distribution, while particles with $0.1 < [\mathrm{M/H}] < 0.4$ dex shows a nearly normal distribution. In the top-right panel, particles with $8 < \mathrm{age} < 14$ Gyr have the broadest distribution, while particles with $0 < \mathrm{age} < 4$ Gyr have the narrowest and most peaked distribution. The red line and black line represent pPXF fitted and the true LOSVD for this Voronoi bin, with kinematic parameters texted in the top right corner, respectively. Compared to the true LOSVD, pPXF fitted curve in red is close to the youngest populations ($0 < \mathrm{age} < 4$ Gyr), which is reasonable because these populations dominate the light. However, the top right panel shows that pPXF tends to also care for the oldest populations by having more skewed tail their LOSVD at $V_{\rm LOS}$ around $100 \sim 200$ km s^{-1}. Even though the youngest populations dominate the spectral light, the old populations have the most features along the whole fitted wavelength region. Therefore, the differences of the skewness in the region of $V_{\rm LOS}$ at $100 \sim 200$ km s^{-1} indicate the effect caused by different spectral features having different LOSVDs. This figure strongly suggests that disregarding this variability during the pPXF fitting process could lead to biased results. Because of the relation between [α/Fe] and [M/H], stars with different [α/Fe] should also have different $V_{\rm LOS}$. Given the lack of [α/Fe] bins in the templates, we will investigate it in the future.

We made a new experiment to verify this effect on pPXF results. Firstly, we select all the particles in each Voronoi bin, then keep the other parameters but randomly shuffle their $V_{\rm LOS}$. In this way, we eliminate the relation of $V_{\rm LOS}$ with [M/H] and age. The bottom two panels of Fig. 5.11 show the LOSVDs of the same groups of particles after shuffling their $V_{\rm LOS}$. We can see that stars with different [M/H] and age now have the same LOSVD, meaning all the spectral features also have the same LOSVD. We do it for all the Voronoi bins and generate an artificially modified mock stellar catalog. Then we follow the procedures in Section 5.1.2 and generate new mock cubes using this modified catalog and PEGASE-HR templates. Next, we employ pPXF to measure the kinematics. The fifth row of Fig. 5.10 shows the kinematics recovery by using this modified catalog. It is clear that all the Gauss-Hermite moments are well recovered with all the data points aligned to the zero line. The red and black lines in the bottom row of Fig. 5.11 also indicate the correction of pPXF measurements. We also plot the kinematics maps in the same way as Fig. 5.3 in Appendix (Fig. C.4) to better demonstrate the consistency.

We also tested the kinematics recovery on a datacube generated from the velocity-shuffled catalog using MILES templates, as shown in the sixth row of Fig. 5.11 (see also Fig. C.5). In this low spectral resolution, even though we shuffled $V_{\rm LOS}$ to make different populations have the same LOSVD, there are still systematic offsets appearing in all the panels. We also demonstrate in the last row the results using PEGASE-HR templates on $V_{\rm LOS}$-shuffled catalog but in MUSE spectral resolution to remove the effects due to different templates (see kinematics maps in Fig. C.6). And the appearance of offsets indicates the effect purely due to insufficient spectral resolution.

In conclusion, all these results strongly suggest that potentially both the low spectral resolution and variant LOSVD contribute to the systematic offsets we see in residual panels of Fig. 5.3. However, shuffling V_{LOS} changes the true LOSVD (see the black line of the top and bottom rows of Fig. 5.11), which causes it not to be a one-to-one comparison. Therefore, future tests are needed to further verify our conclusions. In the future, to remove the systematic offsets of stellar kinematics measured by pPXF, one needs to use an instrument with higher spectral resolution than MUSE. This means one has to also increase the exposure time to achieve the same SNR. As we know V_{LOS} changes with [M/H] and age in the line-of-sight of real galaxy observations, it might be necessary to allow different templates to have different LOSVDs during the full spectrum fitting to obtain more accurate results. However, since kinematic and stellar populations measurements via full spectrum fitting have already been a highly-degenerate problem. Allowing variable LOSVDs will increase the degeneracy by a factor of the number of templates. In this case, the algorithm is easy to obtain incorrect results. One possible way would be simply assuming a quantitative relation between LOSVD and metallicity and age at different locations of the galaxy and that this relation can be expressed by an equation. This is equivalent to adding a prior to the pPXF fitting process to improve the accuracy of kinematics measurement without adding degeneracy. Even though this way has many limitations, it will be still useful for Milky-Way-like galaxies' analysis, and the GalCraft code can help provide this prior for galaxies in different projections.

5.2.2 Inconsistency of Mass Fraction Distributions Recovery

In pPXF, measuring mass fraction distributions of stellar populations is a highly degenerate problem, especially when one wants to obtain the star formation history of the Galaxy at different components. Even though previous studies used different templates and mostly plotted mass fractions without dividing the bin size, the increase from young to old and metal-rich populations in Fig. 5.4 and Fig. 5.8 is also seen in some components of the galaxies from Pinna et al. (2019a,b) and M21. In the mock and real data tests of Kacharov et al. (2018) (see their Fig. 5), the mass fraction distributions are also smoothed towards old and metal-rich regions. Besides, the overdensity feature of young populations at around $2 \sim 4$ Gyr is also seen in some of the above studies. In this section, we re-investigate all these tests more comprehensively using mock cubes of GalCraft and discuss the differences of mass fraction distribution from pPXF whether or not dividing the bin size and try to address potential reasons causing the inconsistency to the true values.

We first convert the age distribution x-axis of Fig. 5.8 from linear to log-scale, because MILES templates are almost evenly scaled in log age, then the width of each step is almost equal. Next, we re-plot the age distribution recovery in Fig. 5.12 and calculate the correlation coefficients using the original age grids. The first row shows age distribution in mass fraction Gyr^{-1}, which is the representation of the star formation history of each component. The thin disk, inner central, and global panels clearly show three overdensity regions, which are pointed out by three yellow boxes. The thick disk and upper central panels also have peaky features in these boxes, but not significant. The second row shows

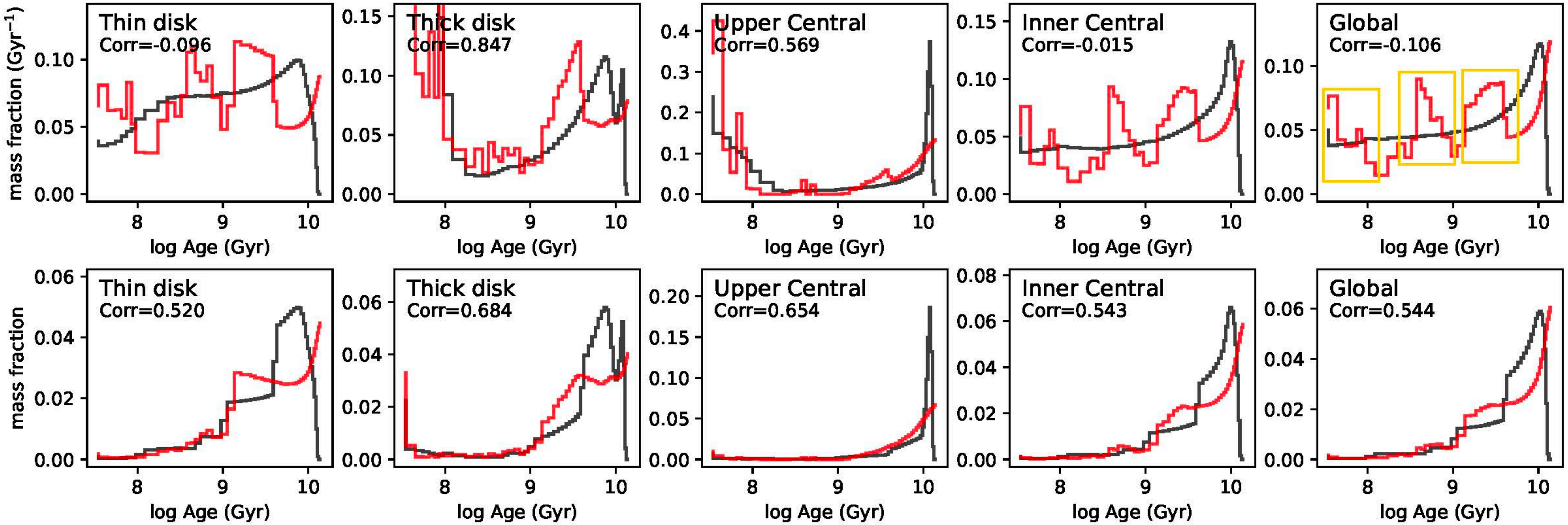

Figure 5.12: Age distribution recovery of different components of the galaxy from pPXF on the mock MUSE cube generated in Section 5.1.1. This figure is similar to Fig. 5.8 but the x-axis is in logarithmic scale. Black lines are the true values and red lines are the results from pPXF. The first row shows mass fraction Gyr^{-1}, which is the representation of star formation history, and the three yellow boxes in the "Global" panel are the three overdensity regions. The second row shows just the mass fraction, which is the direct output from pPXF. The correlation coefficients of the red and black lines are texted on the top right corner for each panel. Especially, for the age distributions, we use the original age grid to calculate correlations, which is evenly space in log age.

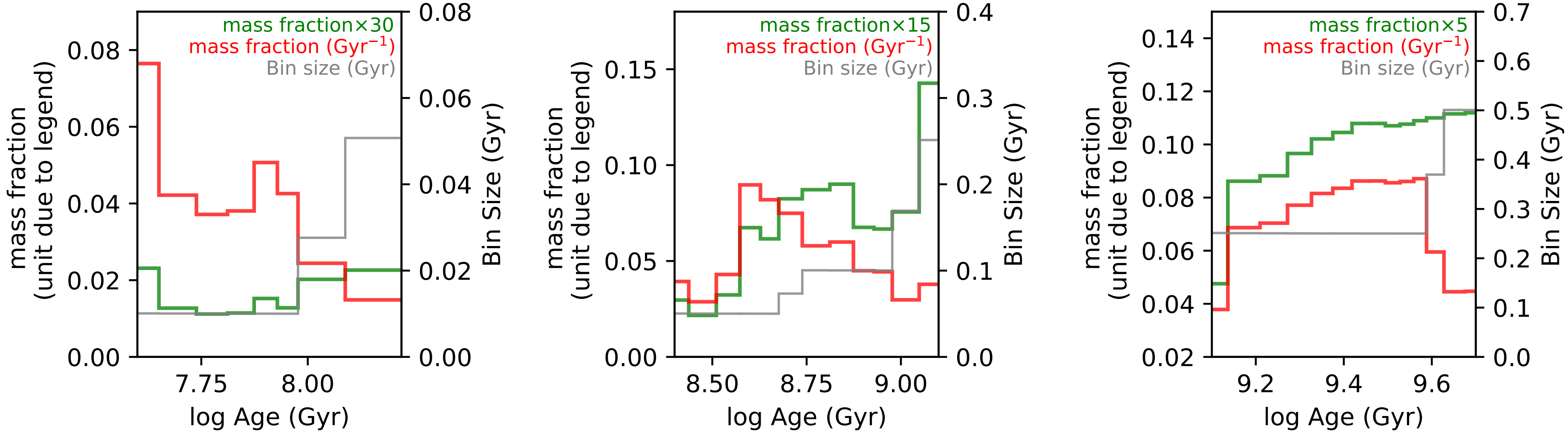

Figure 5.13: Zoom-in of the three overdensity regions pointed out in Fig. 5.12. The red and green lines are mass fraction (Gyr^{-1}) and mass fraction distributions, equivalent to the first and second row of Fig. 5.12, respectively. The mass fraction for each panel is multiplied by a factor for visualization. We also plot MILES templates' age bin size as the grey line and the scale in the right y-axis.

just the mass fraction, which is the direct output from pPXF. Opposite the first row, all the panels in the second row show relatively smooth distributions. Therefore, given the differences between the first and second rows are whether or not dividing bin size, we speculate the three overdensity regions are artifacts due to the bin size arrangements to the template grids.

To verify our speculation, we plot in Fig. 5.13 a zoom-in of the three overdensity regions pointed out in the top right panel of Fig. 5.12. The red and green lines are mass fraction (Gyr^{-1}) and mass fraction distributions, equivalent to the first and second row of Fig. 5.12, respectively. The mass fraction for each panel is multiplied by a factor for better visualization. We also plot age bin size of MILES templates as the grey line and the scale in the right y-axis. From these three panels, we find the peaky features of mass fraction (Gyr^{-1}) (red line) appear when the age bin size experiences a decrease from old to young ages, and the mass fraction distribution (green line) is still relatively smooth. Therefore, we can confirm the peaky features of star formation history from pPXF is due to the age bin size of templates. This also explains the better recovery of [M/H] distributions (bottom row of Fig. 5.8) which is due to the almost linearly spaced [M/H] grids.

Previous studies applied regularization (McDermid et al., 2015) in pPXF to obtain smooth mass fraction distributions. The regularization works as an extra term in the equation to be minimized during the optimization to damp high-frequency variations in the mass weights distribution along spectral templates grid (see details in Section 3.5 of Cappellari 2017) and lead to smooth mass weights distributions. However, the source code of pPXF indicates that this smoothness has an effect on mass weights rather than mass weights per bin size. This explains why mass fraction distributions are smooth in our results, and whether mass fraction (Gyr^{-1}) distributions are smooth depends on how the template grids are spaced. A further test by modifying pPXF and taking into account the bin size during regularization is needed to verify in the future. From now on, we only use the mass fraction without dividing the bin size to justify the recovery consistency, because this is expected to be achieved by pPXF.

Although current regularization ignores the bin size of templates grid and different regularization strategies can affect the mass fraction recovery (will be discussed in Section 5.2.3), we still see the inconsistency of mass fraction recovery for the oldest populations in the bottom row of Fig. 5.12. The main reason is that spectral templates in the oldest and metal-rich regions are very similar and indistinguishable at the current SNR level. Therefore, the pPXF algorithm gets confused when measuring mass fraction weights in these age-[M/H] regions. We also confirm in the Appendix (Fig. C.7) that shuffling particles' V_{LOS} do not improve the results significantly.

According to the reasons above, one way to increase the quality of the mass fraction recovery is to increase the SNR of the spectrum. This can be achieved by requiring a higher target SNR during the process of Voronoi binning, but with the cost of reducing the number of Voronoi bins, i.e., details in the parameter distributions. We show in Fig. 5.14 the age and [M/H] distributions at different SNRs (shown in different colors), compared to the true values shown in black lines. This figure is obtained by running the GIST pipeline on the mock MUSE cube generated in Section 5.1.1 with different target SNRs during Voronoi binning, and then the mass fraction distributions of all the Voronoi bins are added

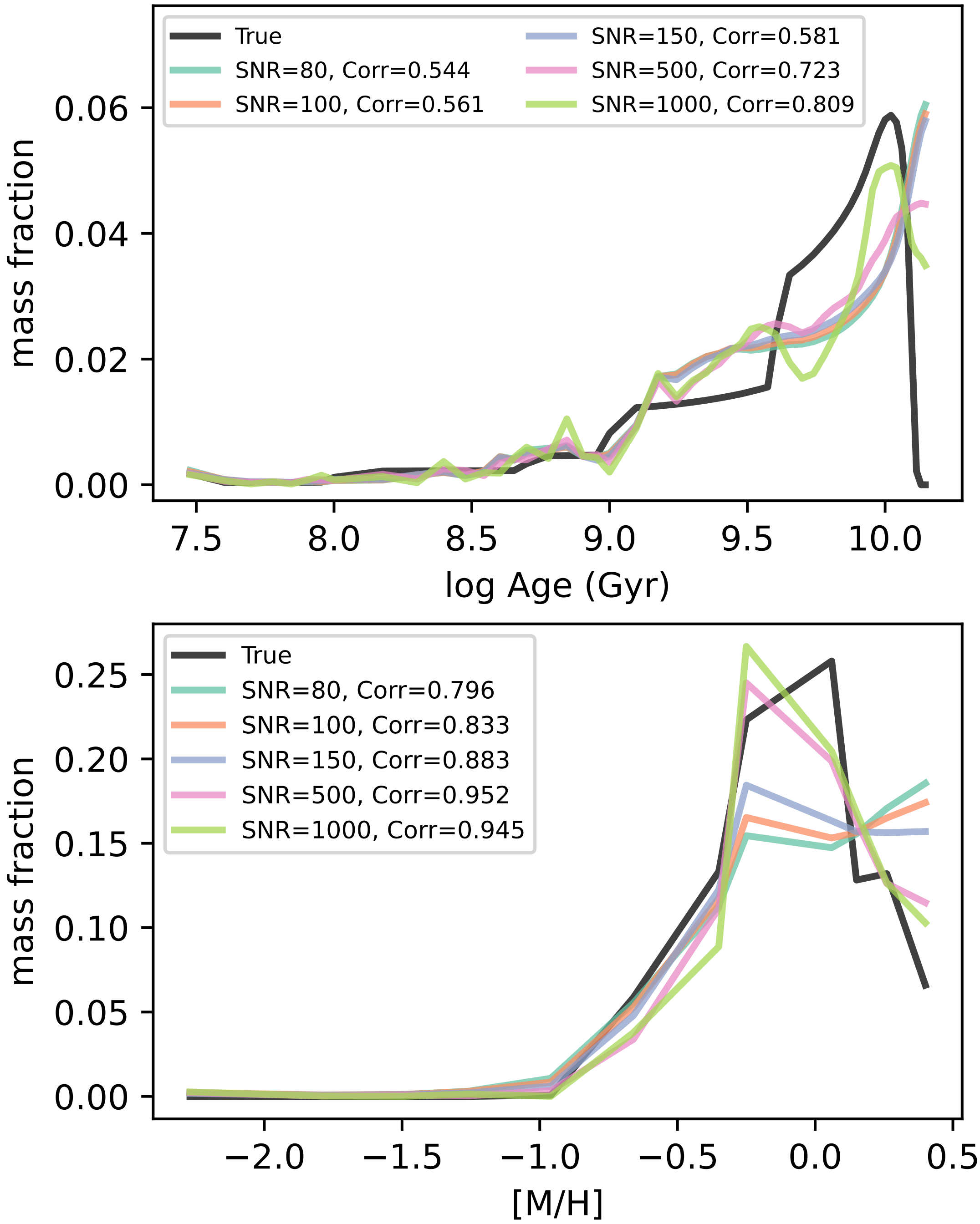

Figure 5.14: Global age (top panel) and [M/H] (bottom panel) distributions recovery at different SNRs, which are obtained by running pPXF on the mock MUSE cube generated in Section 5.1.1 with each target SNR of Voronoi bins, and then the mass fraction of all the bins are added together to represent the global mass distributions of the galaxy. Black lines are the true values. This figure indicates that increasing SNR can help obtain better mass distributions compared to the true values, especially the issue of metal-rich, old stellar population overestimation is alleviated.

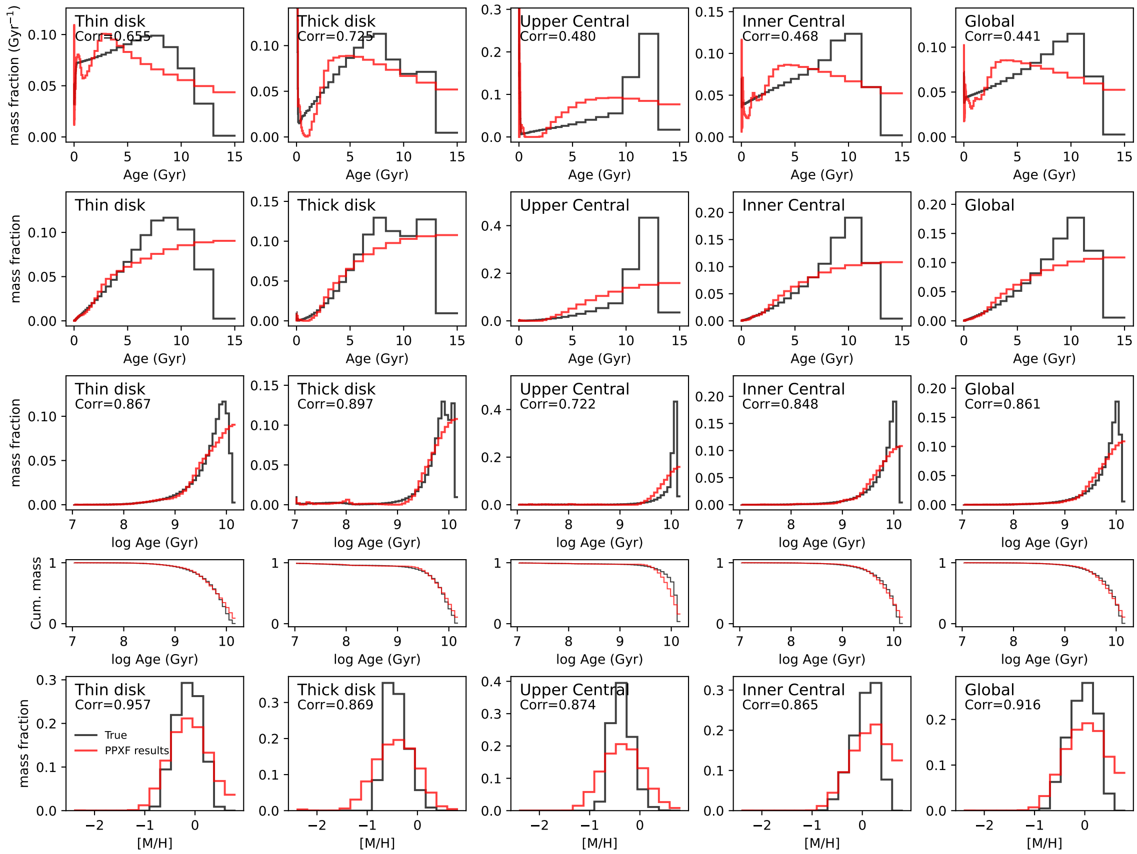

Figure 5.15: Age and [M/H] distributions of each component for the mock cubes generated using PEGASE-HR templates and degraded to MUSE resolution (FWHM = 2.65Å). The first row demonstrates mass fraction (Gyr^{-1}) changes with age (star formation history). The second row shows the mass fraction changes with age. The third row shows mass fraction changes with log age. The fourth row shows cumulative mass fraction distributions and the fifth row shows metallicity distributions.

together to represent the global mass distributions of the galaxy, and finally integrated through age and [M/H] axes, respectively. To make it a fair comparison, we make sure that the spatial pixels used to generate Voronoi bins are the same. According to the correlation coefficients, this figure indicates that with the increase of Voronoi bin SNR, the global mass fraction distributions in both panels are more consistent with the true values. Specifically, the weights of metal-rich and old stellar populations are going lower with higher SNR. Therefore, Fig. 5.14 indicates that increasing SNR can improve in SFH recovery. However, we want to note that in reality, integrated spectrum with SNR larger than 200 pix^{-1} is pointless because in this case numerical issues, spectral library uncertainty, etc become the dominant factor.

To study the effect of different templates, as shown in Fig 5.15, we re-plot age and [M/H] distributions mock cubes generated by PEGASE-HR templates, which are evenly spaced in log age. During the pPXF fitting, we use the same templates and fitting strategies (fitting wavelength region, regularization, etc.) as those in Section 5.1.2 and also degrade to MUSE spectral resolution to remove the resolution differences (even though we find increasing spectral resolution does not improve the recovery significantly after comparing to Figure C.8 of a PEGASE-HR spectral resolution version). The Voronoi bin allocation is the same as mock cubes generated by MILES templates. According to the correlation coefficients, mass fraction distribution using PEGASE-HR in Fig. 5.15 are more consistent than those using MILES in Fig. 5.8 and Fig. 5.12 for both age and [M/H]. Moreover, the mass fraction - log age panels (second row) have almost no blobs than the results using MILES (second row of Fig. 5.12), and the oldest populations have a more modest increase. Therefore, in our test, PEGASE-HR performs better during the mass fraction recovery and could provide more consistent mass fraction distributions than MILES. We think this is due to the templates' perfectly even spacing in log age, which leads to smoother input and is preferred by pPXF during regularization. However, the current limitation of PEGASE-HR templates is they only have one bin in [α/Fe], we emphasize here again the need for multiple [α/Fe]-grid templates.

5.2.3 Effect of Regularization

In this section, we focus on the effects of different regularization strategies on mass fraction distributions. There are two parameters controlling the regularization in pPXF, which are `regul` and `reg_ord`, where `regul` or R applies linear regularization to the weights during the pPXF fit and `reg_ord` controls the order of regularization. Kacharov et al. (2018) investigated mass fraction differences of mock data with different order of regularization (see their Fig. 5 and Fig. C2) and concluded that the choice of order of regularization does not affect the results over a significant level. Boecker et al. (2020) tested the mass fraction distributions recovery of the second- and third-order of regularization and compared them to the true values using stellar particles from an EAGLE simulated galaxy. They found results from third-order regularization are more consistent to the true values.

The selection of R value was briefly discussed in Section 5.1.2. Here we re-investigate this question in detail by applying different `regul` and `reg_ord` strategies on our mock cubes generated in Section 5.1.1. The results are shown in Fig. 5.16, where we plot the age distributions on the top panels and [M/H] distributions on the bottom panels for different

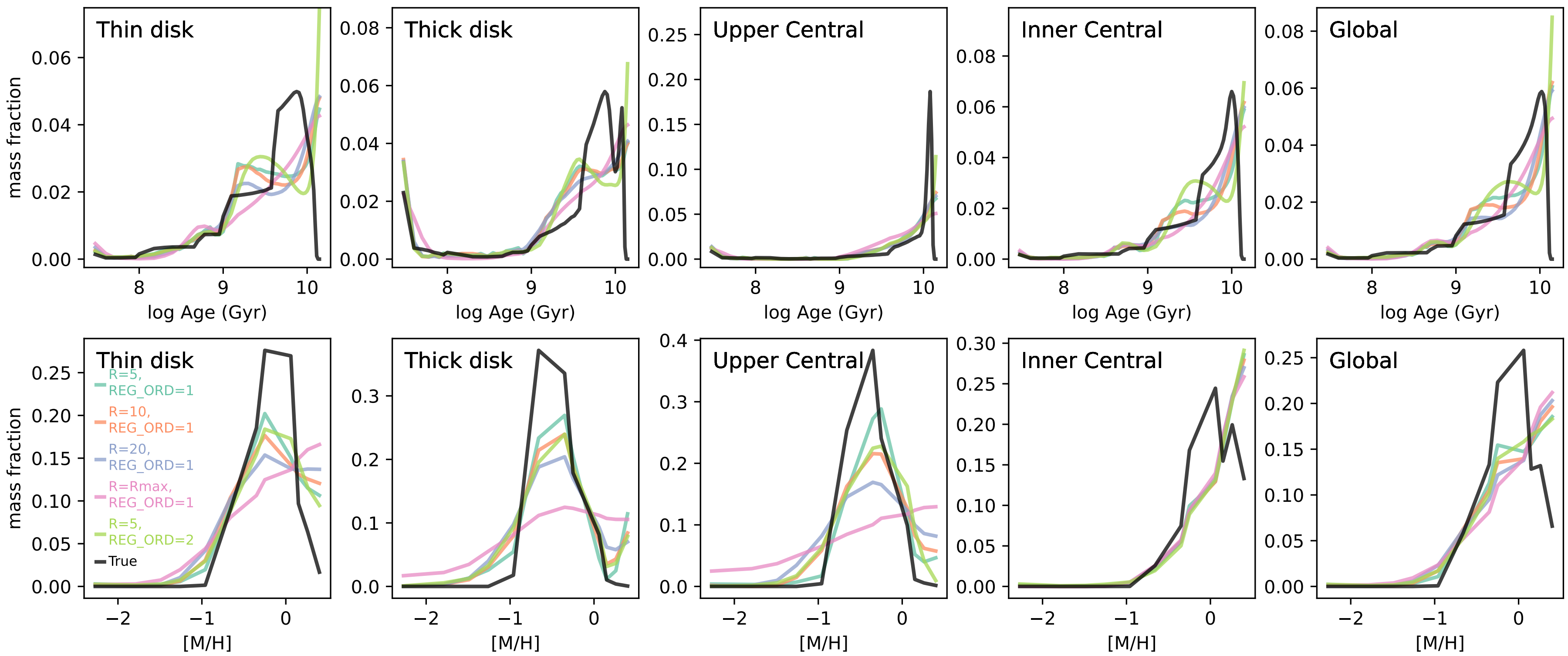

Figure 5.16: Age (top panel) and [M/H] (bottom panel) distributions recovery for different components of the galaxy by running pPXF on the mock MUSE cube generated in Section 5.1.1 with different regularization strategies (shown in different colors). Black lines are the true values. The last column shows global age and [M/H] distributions. This figure indicates that results using the first order of regularization are better than results using the second order of regularization; increasing R can help obtain smoother results but the weights will go towards metal-rich regions for all the components.

components of the galaxy. We also plot the global mass distributions in the last column. We apply five different strategies during the fitting: the first three cases use first-order regularization with $R = 5$, $R = 10$, $R = 20$, and $R = R_{max}(50 \sim 150)$, respectively; the last uses $R = 5$ and second-order regularization. Black lines are the true values. When comparing results with the same order of regularization but different R, we find results with larger R are smoother, and their weights at old populations are relatively smaller than results with smaller R. However, in the [M/H] distribution panels, weights are going more toward metal-rich populations and the distributions become less consistent with the true values for all the components. For results with the same R but different orders of regularization, we see results with `reg_ord` = 1 show better consistency with the true values than those with `reg_ord` = 2. Specifically, results with `reg_ord` = 2 have more mass fractions in the oldest populations.

Therefore, this figure indicates that for this mock cube, results using the first order of regularization are better than results using the second order of regularization; increasing R can help obtain smoother results but the weights will go towards metal-rich regions. Given we do not have the codes to perform third-order regularization, it is hard to say which order is the best. Intuitively, it seems higher-order regularization tends to have more mass fraction in the old populations. Therefore, we think the third-order regularization is preferred for EAGLE simulated galaxy (Boecker et al., 2020) because the particles' mass weights from the EAGLE simulated galaxy are naturally dominated by old and metal-rich populations, then the third-order regularization tends to fit better in this region. Our E-Galaxia mocked catalog has no particle in the metal-rich and oldest regions, so the first-order regularization is better. However, all the above discussions need further verification and we will test the third-order results in the future. For real galaxy fitting results of Pinna et al. (2019a,b) and M21, they all used second-order regularization and showed an overdensity of mass weights in the oldest, metal-rich populations for different components of the galaxies, which makes it difficult to discuss differences in the formation epoch between these components, and it is not possible to tell if the overdensity is true or just artifacts from pPXF.

5.2.4 How to interpret the galaxy evolution from pPXF results?

In this section, we discuss the interpretation of a galaxy's evolution using results from pPXF. According to our tests and discussions in the sections above, [M/H] distributions recovery is mostly robust with correlation coefficients of pPXF and true results higher than 0.9, except for the most metal-rich regions. The [M/H] results in M21 (see their Fig. 11) are dominated by the most metal-rich populations, which is similar to pPXF results of the upper central component in our test in Fig. 5.8, where the true values have a drop at solar [M/H], indicating the location where the chemical enrichment is getting flattened. However, it is different from the true values. Therefore, we infer that it is difficult to interpret at which [M/H] the chemical enrichment history is flattened from pPXF results due to the poor recovery ability of metal-rich populations. For NGC 5746 in M21, one conclusion we can derive is this galaxy is overall more metal-rich than our MW, which is reasonable because it is more massive due to the stellar mass-metallicity relation (e.g., Andrews & Martini 2013).

The star formation history recovery for all the components is inconsistent based on correlation coefficients of pPXF and true results being negative from our tests in Fig. 5.8, and we see similar features such as the overestimation in the age ranges of 2 ∼ 4 Gyr and 12 ∼ 14 Gyr in some studies of real galaxies such as Pinna et al. (2019a,b) and M21. Therefore, we infer that these features are due to the artifacts of uneven spacing of age grids of MILES in linear space rather than an indication of recent or early starburst. Even though Kacharov et al. (2018) presented a good pPXF recovery test, their comparison to the true values is a cumulative distribution in log age scale using PEGASE-HR templates, where the important regions in 2 ∼ 14 Gyr are indistinguishable. We also confirm this visualization effect in the first and fourth row of Fig. 5.15. In addition, the old and metal-rich features in Fig. 5.7 also indicate the artifact signals in previous pPXF analysis on other galaxies. Even though all the results are verified by calculating the weight errors using "bootstrap" realizations (e.g., Kacharov et al. 2018; Prichard et al. 2019 and M21), it is more likely to calculate the scale of the local minimum rather than the actual errors.

In the future, non-parametrically, it is worthwhile to test the mass fraction recovery ability using a linearly spaced templates grid or modifying pPXF to regularize the fitting in terms of mass weights per Gyr or dex. Parametrically, one can try to recover both chemical enrichment history and star formation history by taking into account chemical evolution theories, which will be helpful to remove counter-intuitive features in the mass fraction distributions. One example is the semi-analytic spectral fitting method in Zhou et al. (2022). This method applies full spectrum fitting with the predicted best-fit spectrum from chemical evolution models, which is similar to adding a prior on pPXF during the fitting. Fig. 6 of Zhou et al. (2022) showed that this method has better consistency on mass fraction recovery than pPXF, and the overestimations in metal-rich, 2 ∼ 4 and 12 ∼ 14 Gyr regions are removed successfully. Therefore, it proves a way to accurately measure chemical enrichment and SFH for future studies. One caveat is that this chemical evolution model is a close-box model, which is not applicable to galaxies dominated by frequent passive merger events (e.g. NGC 7793 in Kacharov et al. 2018). In addition, this method does not take into account the relation of chemical abundances with velocity, which are important indicators for radial migration and kinematic heating.

5.2.5 Caveats when using GalCraft

In this section, we give some caveats when using the GalCraft code. Since this code is using the mock stellar catalog from E-Galaxia which embedded the chemical evolution model of S21. The success of this model is that it can reproduce $[\alpha/\mathrm{Fe}]$-[Fe/H] bimodality which is quantitatively consistent with the APOGEE observations of the MW in Hayden et al. (2015). In reality, a more complex process is likely to exist as non-asymmetric perturbations like spiral arms, bars, interlopers, etc. The current model does not have these features but can be added in the future; In addition, S21 model suggests the $[\alpha/\mathrm{Fe}]$ bimodality is a natural result of radial migration and kinematic heating of stars, which is opposite to some clarifications by simulations which indicated that a merger is needed to contribute to the formation of bimodality (e.g., Clarke et al. 2019; Buck 2020), and this model lacks stars in the Galactic halo. Therefore, this code is not applicable for comparing the merger features or the halo distributions; This model also lacks descriptions of parameter distributions in the bulge, especially the nuclear disk in the MW center (e.g., Schultheis et al. 2021) and chemodynamics of the B/P bulge (e.g.. Wylie et al. 2021), which could be improved in the future.

5.3 Future Studies

In this study, we did not apply extinction on the mock datacube because this is the work purely on testing the kinematic and stellar populations recovery via full-spectrum fitting. In real edge-on galaxies, extinction can have an essential effect because it can obscure nearly 50% of the total light from stars in the thin disk, making SNR in this region lower. With the GalCraft code, we have the ability to test the kinematics and stellar population recovery of mock cubes with extinction. The extinction model in E-Galaxia is assumed as a double exponential distribution mainly in the thin disk, which can be integrated through the line-of-sight, and each particle will have a $E(B-V)$ value. Then, we can add the extinction effect to the spectrum by using a specific reddening curve (e.g., Cardelli et al. 1989; Calzetti et al. 2000), which is also the strategy that pPXF applies to estimate extinction. Ge et al. (2018) has already tested the effect of extinction on full-spectrum fitting results using one stellar population template, we are able to re-investigate it with a realistic E-Galaxia model and also study how different extinction models can affect the results. In addition, several non-asymmetric features should be added to the chemical evolution model like the spiral arms, bars, mergers, and halo particles to mimic a more realistic MW.

There are several improvements can be made to SSP models. The current version of GalCraft has embedded MILES and PEGASE-HR templates. In the future, we will add other empirical models such as BC03 (Bruzual & Charlot, 2003), X-Shooter (Verro et al., 2022), and theoretical models like STARBURST99 (Leitherer et al., 1999) and BPASS (Eldridge et al., 2017), which can satisfy different science goals since different models have different advantages.

In addition, even though we showed in Section 5.1.4 that the [α/Fe]-rich and [α/Fe]-poor populations can be well identified using two [α/Fe] bins, it is still essential to add more bins to obtain detailed [α/Fe] distributions. The Milky Way results in Hayden et al. (2015) have shown that stars in the thin and thick disk have [α/Fe] values about 0.0 and 0.2 dex, respectively, and there are some stars with negative [α/Fe], which is also shown in the E-Galaxia catalog. However, it is challenging to use current MILES α-variable templates to identify these two components because most of the weights are in the bin of [α/Fe] = 0, and it is also difficult to assign a spectrum to particles with negative [α/Fe]. More [α/Fe] bins will be helpful in studying [α/Fe]-[M/H] distributions for external galaxies at different components.

According to Section 5.2.1, given that spectral resolution is important in stellar kinematics measurements, it is also necessary to increase both the instruments' and stellar population models' spectral resolution. Current IFS observations such as MUSE perform well on kinematics measurements of the galaxies' central region. However, to go deeper and obtain distributions of the outer regions, especially the outer thin disk, which normally has lower dispersion, one might need to use instruments and SSP templates with a higher spectral resolution to obtain more accurate kinematics maps.

Other than increasing spectral resolution. As discussed in Section 5.2.1, a prior can be added to the pPXF fitting process which assumes the relation between LOSVD and metallicity and age at different locations of the galaxy can be expressed by a function. Then it can allow templates to have different LOSVDs and the degeneracy is not increased. The other way is to add the evolution of stellar kinematics to the semi-analytic model like the one in Zhou et al. (2022) and make it derive the chemical evolution histories along with kinematic processes directly by fitting with the integrated spectra, where the fitting process is constrained by several model parameters and more physical. We will test the feasibility of these methods in the future.

As for mass fractions, Section 5.2.2 demonstrated how the bin size can artificially create peaky features in the star formation history. In the future, it might be worthwhile to test fitting results using linearly spaced template grids or modifying pPXF to regularize the fitting in terms of mass weights per Gyr or dex. Parametrical methods like Zhou et al. (2022) can also be used to test recovery abilities.

The aim of GalCraft code is to apply the knowledge we learned from the Milky Way to other Milky-Way-like galaxies by mock the IFS observations as a bridge. Especially, we want to interpret physical processes such as radial migration, kinematic heating, and [α/Fe]-[M/H] distributions of other galaxies to answer the question of whether all the Milky Way analogues have similar processes to the MW, and if not, how much difference they can be. Therefore, we will perform a comparison of a mock Milky Way datacube with a real galaxy observation from current instruments and see the feasibility of identifying these processes from parameter distributions in the future when templates with more [α/Fe] bins are available.

5.4 Summary and Conclusions

In this work, we present the GalCraft code which uses simple-stellar population models and mock Milky Way catalog from E-Galaxia to generate a mock Milky Way datacube in the same data format as integral-field spectrograph (IFS) extragalactic observations. In this way, we eliminate the differences of conceptions in the Galactic and extragalactic studies such as the use of individual stars vs. stellar populations, number density distributions of stars' parameters vs. mass- or light-weighted distributions, results with or without projection effect, etc. The mock datacube can be put into the GIST pipeline to perform data reduction in the same way as extragalactic observations and the results can be compared directly to external galaxies to study the similarities and differences between the Milky Way and its analogues. Therefore, this code is a bridge to link the Galactic and extragalactic studies to understand the formation and evolution of disk galaxies.

The GalCraft code is flexible, allowing users to choose their preferred SSP templates, galaxy distance and inclination, spatial/spectral resolution, and field-of-view of the instrument. GalCraft is also designed to mock current existing instrumental observations such as MUSE, SAMI, or MaNGA, as well as test the performance of future instrument designs.

Using GalCraft, we tested the ability of pPXF to recover the stellar kinematics and stellar population parameters on mock MUSE datacubes at a distance of 26.5 Mpc and inclination 86°. We verified that pPXF can recover chemodynamic maps, and both the [α/Fe]-rich and $[\alpha/\mathrm{Fe}]$-poor can be identified with reasonable regularization values during the fitting. Using line-strength indices can also recover these components.

Despite the similar kinematic trends, we found that there are systematic offsets between pPXF-recovered four kinematic moments (V_{LOS}, σ, h_3, h_4) and the true values because velocity dispersion σ of most Voronoi bins are smaller than the instrument velocity dispersion σ_{inst} of MUSE spectral resolution (FWHM = 2.65Å) and V_{LOS} changes with age and [M/H] for particles in each Voronoi bin. By using higher spectral resolution templates PEGASE-HR (FWHM = 0.55Å) and randomly shuffling V_{LOS} for particles in each Voronoi bin, we can obtain consistent kinematic results compared to the true values. We indicate that it may be necessary to allow different templates to have different LOSVDs, even though this will increase the degeneracy significantly.

We found mass fraction distributions of stellar populations from pPXF using MILES templates on MILES-generated mock cubes show deviations compared to the true values, where mass weights are overestimated in regions of 2 ~ 4 Gyr, 12 ~ 14 Gyr, most metal-rich region and underestimated in regions of 5 ~ 10 Gyr. This is due to the linearly uneven bin size of MILES templates' age grids. When repeating the test by employing pPXF and PEGASE-HR on PEGASE-HR-generated mock cubes, the mass fraction distributions are better with correlation coefficients of pPXF and true results. In addition, we found using first-order regularization can obtain better mass fraction distributions than the second-order.

In the future, to obtain better parameter recovery, it is necessary to use higher spectral resolution templates with more bins in $[\alpha/\mathrm{Fe}]$, which will be essential to identify the mass-fraction of different $[\alpha/\mathrm{Fe}]$ components in more detail. Besides, there needs a test to modify pPXF to regularize mass fraction smoothness by taking into account the template grids. We will also test the effect of different extinction models on these recovery results, and see if we can detect distribution differences of galaxies with different radial migration and kinematic heating rates.

Acknowledgements

We thank Glenn Van de Ven, Yuan-Sen Ting, Zhuyun Zhuang, Richard McDermid, Scott Croom, Andy Casey, Alina Boecker, Francesca Pinna, Eric Emsellem, Jesus Falcon-Barroso and James Binney for useful discussions. ZW acknowledges the HPC service at The University of Sydney for providing HPC resources that have contributed to the research results reported in this paper.

ZW is supported by the China Scholarship Council and Australian Research Council Centre of Excellence for All Sky Astrophysics in Three Dimensions (ASTRO-3D) through project number CE170100013. MRH acknowledges support from ARC DP grant DP160103747 and ASTRO-3D. SS is funded by ASTRO-3D Research Fellowship and JBH's Laureate Fellowship from the Australian Research Council. JBH is supported by an ARC Australian Laureate Fellowship (FL140100278) and ASTRO-3D. JvdS acknowledges the support of an Australian Research Council Discovery Early Career Research Award (project number DE200100461) funded by the Australian Government.

This research has made use of Astropy[1], a community-developed core Python package for Astronomy (Astropy Collaboration et al., 2013, 2018), numpy (Harris et al., 2020), scipy (Virtanen et al., 2020), matplotlib (Hunter, 2007) and SpectRes (Carnall, 2017).

Data Availability

GalCraft source code will be publicly available in https://github.com/purmortal/galcraft. All the pPXF recovery test figures will also be available on this website.

[1] http://www.astropy.org

Chapter 6

The Impact of Extinction on Kinematics and Stellar Population Recoveries

Zixian Wang,[1,2] Sanjib Sharma,[1,2] Michael R. Hayden,[1,2] Joss Bland-Hawthorn,[1,2,3]

[1]Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW, 2006, Australia
[2]ARC Centre of Excellence for All Sky Astrophysics in Three Dimensions (ASTRO-3D)
[3]Miller Professor, Miller Institute, UC Berkeley, Berkeley CA 94720

This Chapter presents the investigations on the recovery ability of pPXF on measuring kinematics and stellar population parameters from mock cubes with extinction added. I wrote all the text and produced all the data reduction procedures and figures. Sanjib Sharma provided extinction information in the E-Galaxia generated mock stellar catalog. Joss Bland-Hawthorn, Sanjib Sharma, and Michael Hayden supervised the whole work and provided suggestions.

Dust removes more than half of the light from stellar energy distributions (SEDs), so it can have a large impact on kinematics and chemical composition measurements (Calzetti et al., 2000). Previous extragalactic studies avoided highly-extinct regions (e.g., the dust lane in the disk of edge-on galaxies) to minimize this effect. And until now, there has been no good way to quantify the effect of dust and determine how much it affects extragalactic results. With the powerful tool of GalCraft introduced in Chapter 4, we address this question by adding 3D Milky Way $E(B-V)$ distribution from Schlegel et al. (1998) on the mock datacubes. We find that this extinction model has a stronger obscuring effect in the galaxy center than real observations, causing the central region darker than the thin disk structures, and more than 90% of the total flux is contributed by the stars at the nearer distances. We also run the GIST pipeline embedded with pPXF on these extinction-added mock cubes and find that the extinction causes V_{LOS} and σ becoming smaller in the galaxy central but the V_{LOS}–h_3 anti-correlation still exists. As for the stellar population parameters, we find the inner and upper central regions demonstrate older and more α-rich population distributions which extend further to the outer disk. With the increase of the reddening factor, all these features are more dominant. We also compared our results with those from real MUSE observation of NGC 5746, and several similar features are identified to be due to extinction. This work provides valuable insights for future studies on distinguishing parameter distribution features attributed to extinction from those due to other effects, to

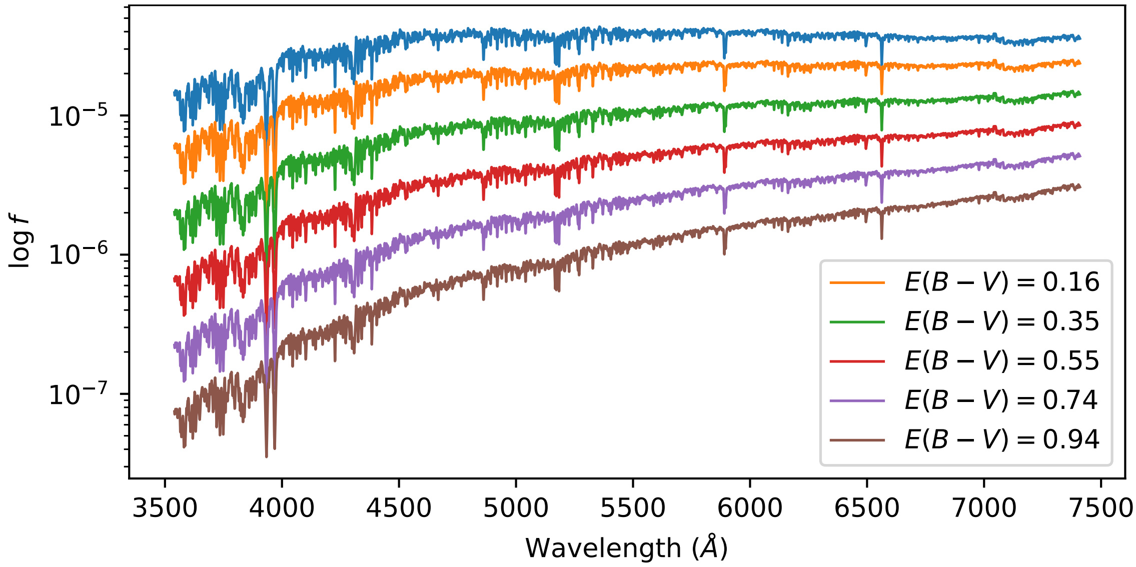

Figure 6.1: Example of an SSP spectrum (5 Gyr, [M/H] = 0.06 dex) by adding different reddening effects, as is seen in the legend. The reddening curve is calculated by the equation in Calzetti et al. (2000) with a given $E(B-V)$.

.

help better understand the evolution of galaxies.

6.1 Adding Extinction to Mock Cubes

In this section, we introduce the method used to add extinction to the GalCraft mock cubes. E-Galaxia estimates the extinction of a star by integrating a 3D dust extinction map through the line-of-sight. The general procedure is similar to that implemented in Galaxia (Sharma et al. 2011, hereafter S11), but modified to account for the fact that the center of the modelled galaxy and its dust disc is now located at a different distance and angular coordinates.

The dust model used is the same as in S11, following Bland-Hawthorn et al. (2010), it firstly models the 3D distribution of the dust by a double exponential function,

$$\rho_{\text{Dust}}(R,z) = \frac{\rho_0}{k_{\text{flare}}} \exp\left(-\frac{R-R_\odot}{h_R}\right) \times \exp\left(-\frac{\left|z-z_{\text{warp}}\right|}{k_{\text{flare}}\, h_z}\right), \tag{6.1}$$

where $z_{\text{warp}}(R,\phi)$ and $k_{\text{flare}}(R)$ represents the warp and flare of the MW thin disk as described in Section 3.2 of S11. Next, this equation was fit to the $E(B-V)$ dust map from Schlegel et al. (1998) and the best-fit parameters are h_R = 4200 pc and h_z = 88 pc derived from χ^2 minimization. We adopt a dust normalization of ρ_0 = 0.23 mag/kpc for the $E(B-V)$, which is a typical value in the solar neighbourhood. In Section 6.3.1, we will talk about several phenomena of the flux distributions of mock cubes in detail after applying this extinction model and illustrate potential limitations and future improvements.

The 3D dust distribution of Equation 6.1 is then used to generate a 3D extinction map by integrating through the line-of-sight based on the viewing point and distance given

Table 6.1: Exposure time of the bulge and disk cubes in different extinction scenarios.

Datacube	**Exposure time (s)** Bulge	Disk
no extinction	79.26	285.14
$E(B-V) \times 0.2$	537.11	1932.34
$E(B-V) \times 0.5$	1018.65	3664.80
$E(B-V) \times 1$	1647.51	5927.21
$E(B-V) \times 2$	2555.34	9193.30

by `GALPARAMS` in GalCraft configurations. The angular resolution is 400×400 pixels, and the radial directions are linearly spaced into 400 bins stretching from $d - 25$ kpc to $d + 25$ kpc, where d is the galaxy distance to the observer. For each particle in the catalog, according to its position, E-Galaxia interpolates the relation of dust with line-of-sight distance using adjacent dust grids and integrates the dust through the line-of-sight, and then transfers it into a $E(B-V)$ value. Then during the process of GalCraft, the spectrum of each particle will be multiplied by an reddening curve from Calzetti et al. (2000). An example of an SSP spectrum after adding reddening curve is demonstrated in Fig. 6.1. Finally, the mock datacubes with extinction are the summation of all the extinct spectra in each spatial bin.

6.2 Mock Data and Parameter Measurements

6.2.1 Mock Cubes Generation with Different Extinction

Same as Section 5.1.1, we generate two mock MUSE cubes towards the bulge and disk by GalCraft using the E-Galaxia mock catalog with 10^8 particles in a distance of 26.5 Mpc and inclination of 86°, which is in the same projection of NGC 5746 observed by Martig et al. (2021). All the GalCraft configurations are the same as Section 5.1.1, which means we use MILES templates with two [α/Fe] bins in Kroupa IMF and BASTI isochrones, and the same spectral interpolation method and spatial/spectral resolution. We generate mock cubes with different extinction effects by multiplying the $E(B-V)$ values of all the particles by several constant factors, which are 0.2, 0.5, 1, and 2. For comparison to the results with no dust effect, we also include the cubes generated in Section 5.1.1 for later pPXF fitting.

One key to fairly comparing results with different extinction effects is that we should ensure all these cubes have the same SNR, and then the effect of spectrum noise on the measurement is identical. Because datacubes with a larger reddening factor have a smaller flux value, their exposure time should be larger than those with a smaller reddening factor to keep similar SNR. In addition, we need to ensure that the noise added to the spectra should be larger than the sampling error (see details in Section 4.2.4).

To satisfy this requirement and make SNR as high as possible, we first estimate the MUSE error of cubes with no reddening, reddening with $E(B-V) \times 1$ and $E(B-V) \times 2$ by using Equations 5.1 and 5.2 and assuming the bulge and disk cubes having an exposure time of 1252.12 s and 4504.72 s, respectively. Next, we run GalCraft with bootstrapping mode to estimate the sampling error of cubes with $E(B-V) \times 1$ and $E(B-V) \times 2$. We also include the sampling error of non-extinction cubes estimated in Section 5.1.1. Then, we calculate the median ratio of MUSE error and sampling error along the wavelength grid of these three cases. We find that this ratio decreases with a larger reddening factor, meaning that with the same exposure time, more extinction will make the MUSE error grow faster than sampling error. Also, this MUSE and sampling error ratio for cubes with $E(B-V) \times 0.2$ and $E(B-V) \times 0.5$ can be simply interpolated. Therefore, even though 1252.12 s (as Section 5.1.1) and 4504.72 s is the maximum exposure time for non-extinction cubes, there is still enough room for cubes with extinction to increase the SNR, and the only criteria we should satisfy is the sampling error of $E(B-V) \times 2$ cubes (largest reddening factor in our cases) is smaller than MUSE error, and then adjust the exposure time of cubes with other extinction cases to make them have the same SNR. Finally, we list in Table 6.1 the exposure times of all these cubes and we use them to calculate the MUSE error for each cube and add Gaussian noise on the spectra. The bulge and disk cubes are finally stitched together for each case.

6.2.2 Galaxy Properties Measurement and Decomposition

We run the GIST pipeline (Bittner et al., 2019) with pPXF embedded to measure kinematics and stellar population parameters from these cubes, with procedures similar to Section 5.1.2. First, we apply the Voronoi tessellation software (Cappellari & Copin, 2003) to re-bin the mock cube and increase the SNR to 50 pix^{-1}, which results in 689 Voronoi bins. Next, we apply the full-spectrum fitting method pPXF (Cappellari & Emsellem, 2004; Cappellari, 2017) to measure the stellar kinematics of each Voronoi bin. Here we use the same MILES templates as when we generated the mock cubes. We use a wide wavelength range of (4700, 6800)Å to fit with the Voronoi binned spectra and remove the first and last 75Å to remove uncertain extrapolated pixels. No regularization is applied, and we include a 12th-order additive Legendre polynomial in this process rather than a fourth-order polynomial in Section 5.1.2 because the spectral continuum shape is changed by extinction effect. Tests on real observation data by van de Sande et al. (2017) indicates that the non-constant residual ($\mathrm{spec_{obs}} - \mathrm{spec_{bestfit}}$) which varies with wavelength can be removed with a polynomial larger than at least 12th order.

Next, we employ pPXF again to measure the stellar population parameters of each Voronoi bin. We use the same MILES templates, spectral resolution, fitting wavelength range, and kinematics results as in previous steps. For regularization, same as Section 5.1.2, we choose $R = 5$ to keep smooth solutions while still allowing for a variation on short timescales in the star formation history. To include the effect of extinction, we use the reddening curve from Calzetti et al. (2000) to measure $E(B-V)$ of each Voronoi bin. This curve is the same as when we add reddening on the spectrum of each particle in Section 6.1. We also tried to apply multiplicative polynomials to mimic the effect due to extinction (see figures in Appendix D) and found out it does not affect our conclusions. After obtaining mass weights of each template, we calculate the mass-weighted age, [M/H], [α/Fe], and mass fraction distributions for each Voronoi bin. We also apply the same decomposition strategies in Section 5.1.6 and separate our mock galaxy into “thin disk”, “thick disk”, “upper central”, and “inner central” regions. The decomposition

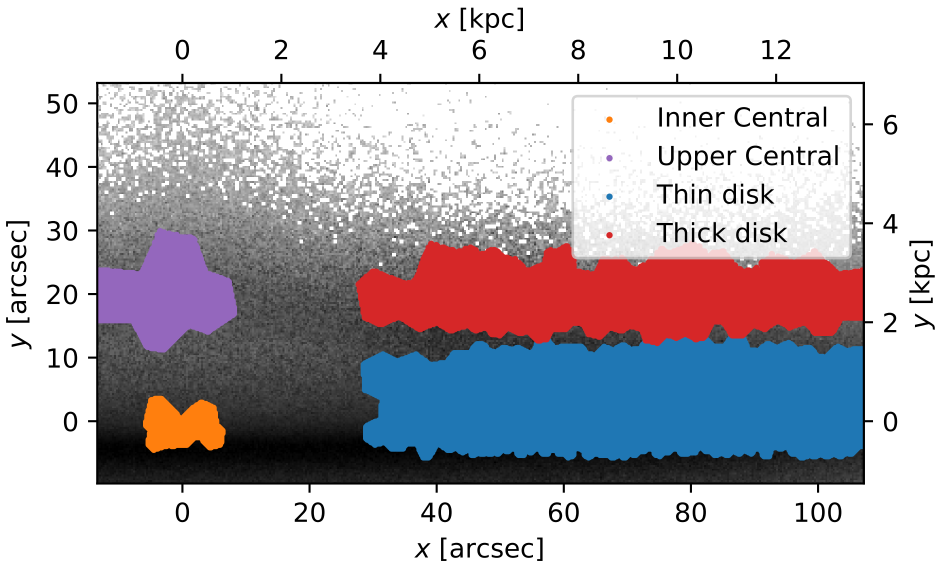

Figure 6.2: Demonstration of the projected thin disk, thick disk, upper central, and inner central regions on top of a grey-scale image of the mock datacube generated similar to Section 5.1.1 but adding extinction models.

map is shown in Fig 6.2.

6.3 Results

In this section, we will talk about the effect of extinction on seeing depth, pPXF-measured kinematics and stellar population parameter distributions, and mass fraction distributions for each component. We also compare our results with the parameter distributions of real MUSE observation on NGC 5746 (Martig et al., 2021), which has the same distance and inclination as our mock data, and discuss which of their results may be due to the extinction according to our conclusions.

6.3.1 Seeing Depth at Different Galaxy Locations

The extinction caused by dust has a significant impact on galaxy observations and properties measurements. This is mainly seen for observations of edge-on disk galaxies (e.g., Scott et al. 2021), where a black dust line is normally seen at the thin disk, hence the luminosity of the thin disk and SNR of the spectrum are reduced. As a result, the kinematics and stellar population parameters of the think disk are more difficult to estimate. Moreover, whether observations can obtain light from the inner central region of the galaxy due to the dust is still a question. And we do not know what percentage of the total flux comes from the stars in the nearer and further regions. In this section, given the power of GalCraft on generating mock IFS observations, we combine the extinction model illustrated in Section 6.1 and investigate these two questions using the mock cubes with $E(B-V)\times 1$.

Fig. 6.3 demonstrates the flux distribution of the entire edge-on Milky Way projected at 26.5 Mpc away with inclination 86° with extinction added. This figure is obtained by first

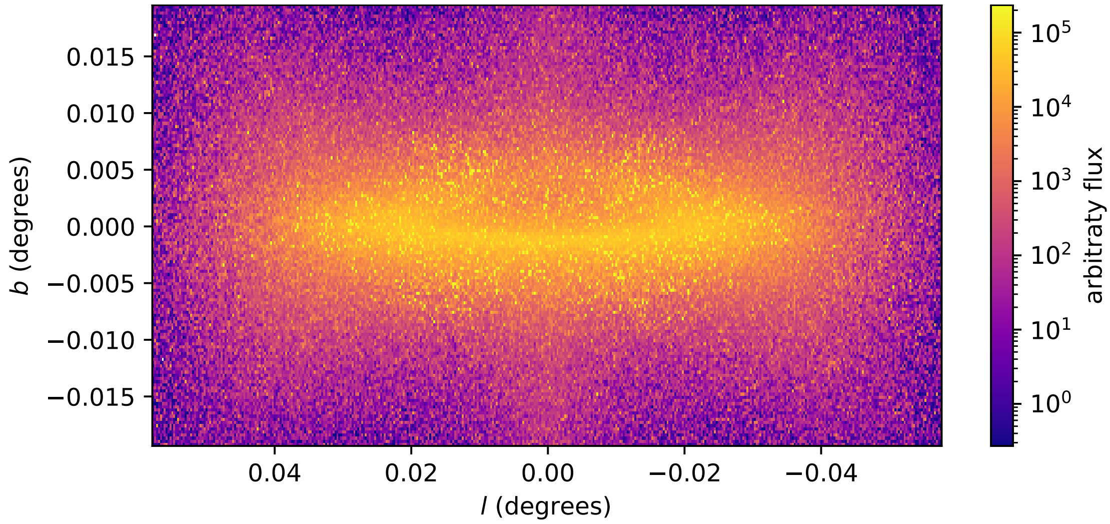

Figure 6.3: Flux distribution of the mock datacube with extinction added in the same projection as NGC 5746. We calculate the total flux by first transferring the apparent G-band magnitude of particles in E-Galaxia mock catalog to flux, and then summing them up using particles included in each spatial bin.

transferring the G-band absolute magnitude of particles in E-Galaxia mock catalog to the apparent magnitude, then adding the effect of extinction A_G by using the equation as follows:

$$A_G = 0.789 \times A_V = 0.789 \times 3.1 \times E(B - V), \tag{6.2}$$

where $E(B - V)$ is the color excess estimated by E-Galaxia. This equation is obtained from Table 3 of Wang & Chen (2019). Next, transfer the apparent magnitude into flux and sum them up for each spatial bin. We can see from Fig. 6.3 that the thin disk structure is the brightest and the central region is relatively dark, and we do not see a thin dust line similar to real observations. However, the image of NGC 5746 (Martig et al., 2021) indicates that the central region is the brightest and the thin disk is dark, and the dust seems to be confined to the thin disk region. This disagreement indicates that possibly the E-Galaxia catalog or the extinction model of Schlegel et al. (1998) is not the best representation of the Milky Way. Given that this extinction model is known to be overestimated in the Galactic plane with $A_V > 0.5$ mag by about 30% (Arce & Goodman, 1999), we also think the double-exponential dust distribution may cause a thicker dust line than observations, which leads to anomalies in the luminosity of the thin disk and the central region. Another reason is the lack of boxy/peanut bulge stellar particles in E-Galaxia catalog, which is currently an extrapolation of the exponential disk distributions, and there is no nuclear disk structure. Therefore, adding particles following these structures might increase the brightness of the central region due to the larger scale height. However, due to the small size of the Galactic bulge, we are unsure if this increase is significant, which can be re-investigated in future studies.

To study how much percentage of the total flux is coming from the stars in the nearer and further regions, we show in Fig. 6.4 the seeing depth of the GalCraft mock cubes at different percentiles. Here the mock cubes are spatially rebinned into Voronoi bins. For each bin, we first calculate the total flux of each particle by summing up the spectrum

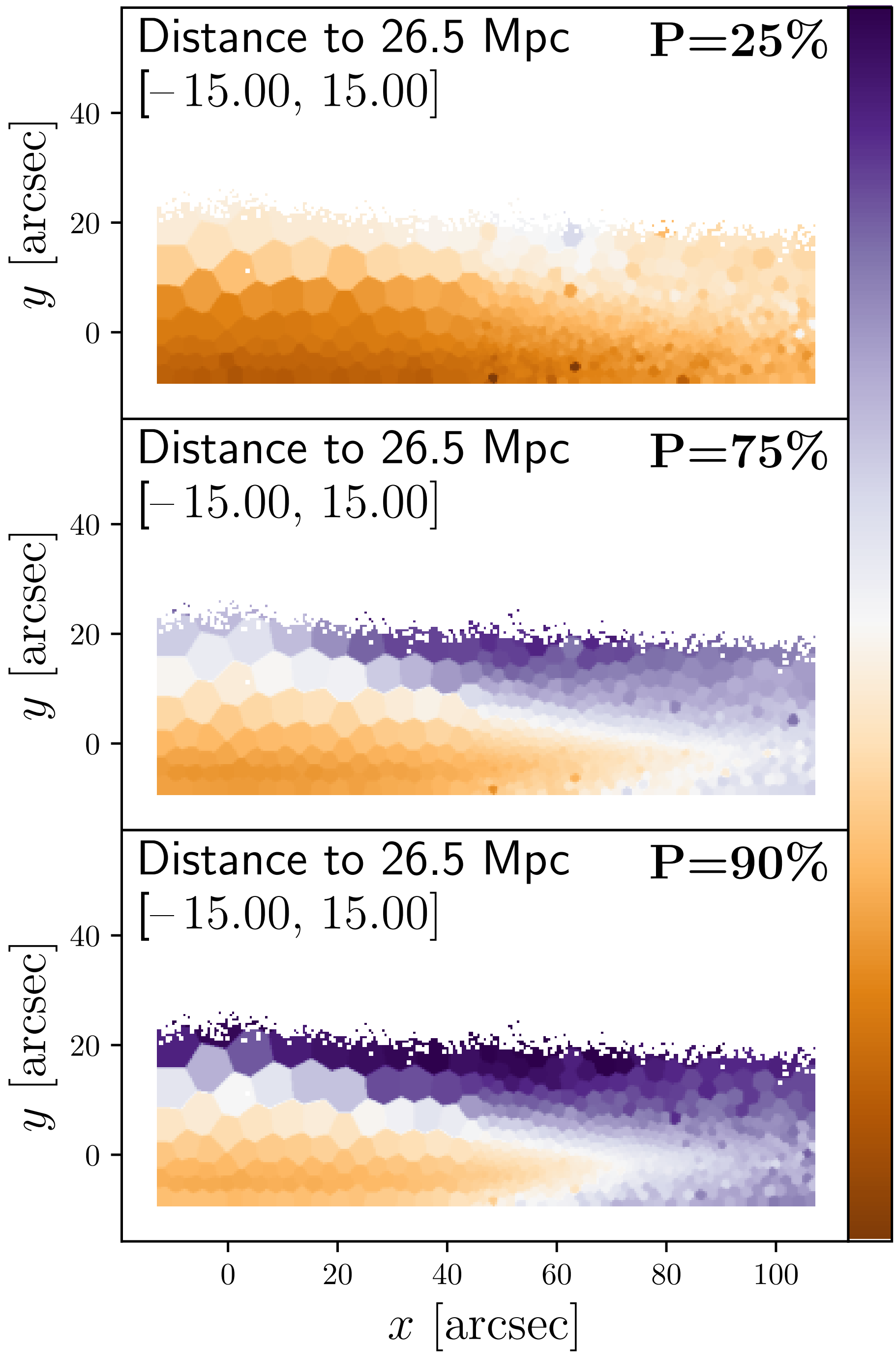

Figure 6.4: Seeing depth of the GalCraft generated mock cubes at different percentile numbers. The mock cubes are spatially rebinned into Voronoi bins. For each panel, the percentile number is in the top right corner, and the distance range to the Galactic center (26.5 Mpc) is texted in the second row of the top left corner. Brown means that the cumulative flux at the percentile comes from stars in the nearer region, white means that the cumulative flux can come from the central region, and purple means that the cumulative flux can reach the further region.

flux in the pPXF fitted wavelength region to obtain a cumulative distribution of flux as a function of stellar distance to the observer. Next, with a given percentile number, we interpolate the distance from the cumulative distribution function to represent the seeing depth of each Voronoi bin in kpc. For each panel, the percentile number is in the top right corner, and the distance range to the Galactic center (26.5 Mpc) is texted in the second row of the top left corner. Brown means that the cumulative flux at the percentile comes from stars in the nearer region to the observer ($d < 26.5$ Mpc), white means that the cumulative flux can come from the central region, and purple means that the cumulative flux can reach the further region to the observer ($d > 26.5$ Mpc). This figure indicates that the first 20% flux is contributed by stars in the nearer region for most of the Voronoi bins. For the first 75% flux, stars in the further regions for Voronoi bins at $x \sim [20, 100]$ arcsec and $y \sim [10, 20]$ arcsec can contribute, but the central region flux is from stars in the nearer region. Even for the first 90% flux, when the outer region goes further, we still do not reach the galaxy center. These results are the same as what we saw in Fig. 6.3 that the flux in the galaxy center is highly obscured by dust, whether due to the lack of bulge stars or the overestimation of $E(B-V)$. From the GalCraft generated mock cubes, we can not obtain light from stars in the galaxy center. In the outer part of the galaxy, the light from very distant stars can still contribute around 10% of the total flux. Our results could not conclude whether we can obtain light from the galaxy center in real observations, and further studies are needed to refine E-Galaxia and dust models to make the flux distributions of mock data match real observations, and then we can re-investigate this question.

6.3.2 Kinematics Recovery

Using results in pPXF by full-spectrum fitting, we show in Fig. 6.5 the stellar kinematic maps ($V_{\rm LOS}$, σ, h_3, h_4) of the mock MUSE cubes generated similar to Section 5.1.1 but adding different extinction models. The first two rows are the true values and pPXF-measured kinematics on mock cubes without extinction, which is the same as Fig 5.3. Results from the third to the last rows are from mock cubes generated in Section 6.1 with different reddening factors on the $E(B-V)$ values, respectively, as seen on the top right corner of the left panel in each row.

From this figure, we find that regions with extreme $V_{\rm LOS}$ are moving upwards from $(x, y) = (0, 0)$ with the increase of $E(B-V)$ factor because the light from the galaxy center is heavily obscured by the dust and the spectra are contributed by stars nearer to the observer, which has less absolute $V_{\rm LOS}$ than stars in the galaxy center. $V_{\rm LOS}$ in the outer thin disk regions ($x \sim [60, 100]$, $y \sim [-10, 10]$) remains unchanged because the extinction effect here is much less than those in the central region, and we can still obtain about 10% light from the further stars. σ distributions change with reddening factor for the same reasons, where it remains unchanged in the outer disk region but decreases with the increase of reddening factor in the central regions. The patterns of h_3 with different $E(B-V)$ are also inconspicuous as results with no extinction (second row) because the pPXF recovery ability on kinematics, which has been discussed in detail in Section 5.2.1. Despite this effect, the $V_{\rm LOS}$–h_3 anti-correlation still exists, where the high-h_3 regions along with the extreme low $V_{\rm LOS}$ regions move to the right with the increase of reddening factor, which indicates that the extinction has an equivalent effect on both $V_{\rm LOS}$ and h_3.

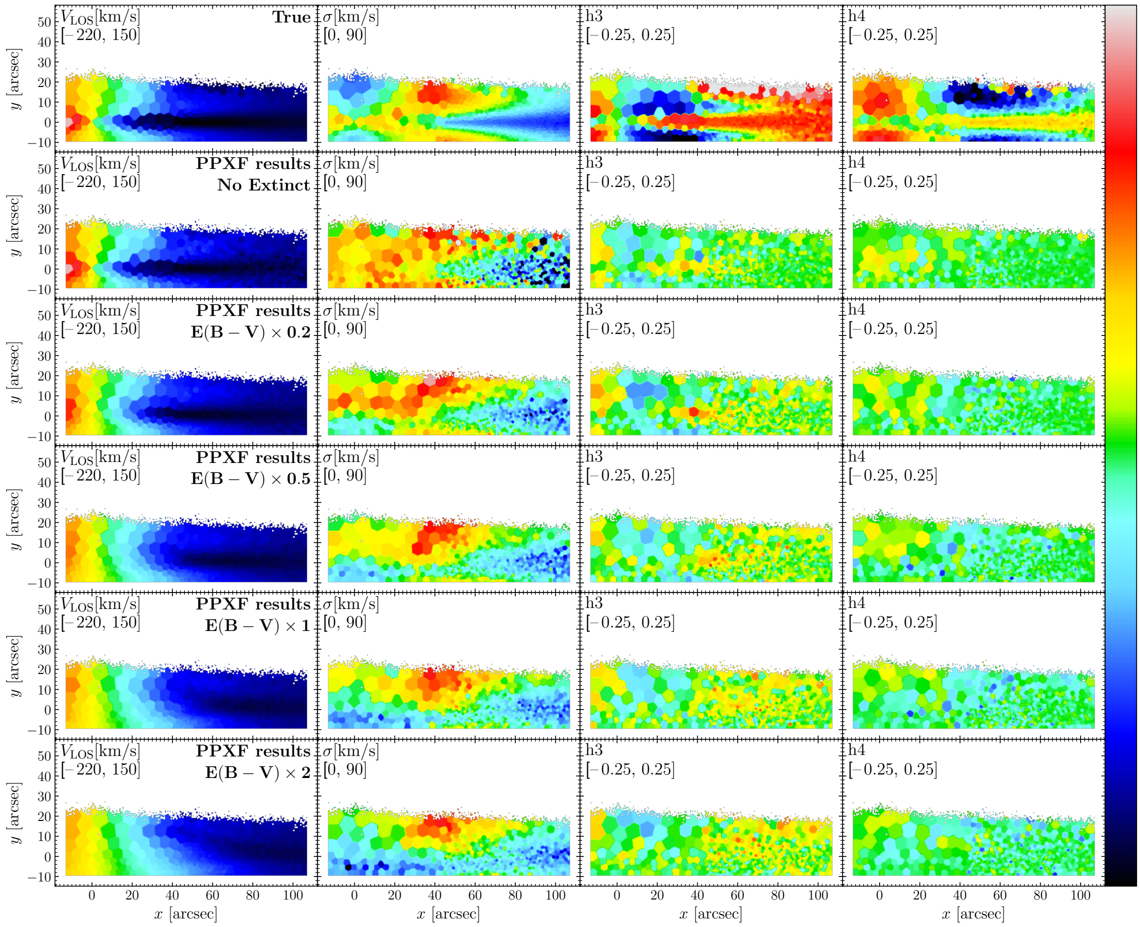

Figure 6.5: Stellar kinematic maps ($V_{\rm LOS}$, σ, h_3, h_4) of the mock MUSE cubes generated similar to Section 5.1.1 but adding different extinction models. The first two rows are the true values and pPXF-measured kinematics on mock cubes without extinction, which is the same as Fig 5.3. Results from the third to the last rows are from mock cubes generated Section 6.1 with different reddening factors on the $E(B-V)$ values, respectively, as seen on the top right corner of the left panel in each row. All the results are measured and demonstrated in the same Voronoi binning arrangements.

h_4 pattern is also inconspicuous for all the rows, but we find the high-value regions also move upwards.

In conclusion, with the increase of reddening factor, $V_{\rm LOS}$, σ and h_4 decrease in the central region but remain unchanged in the outer disk region. h_3 along with extreme low $V_{\rm LOS}$ move to the right, and the $V_{\rm LOS}$–h_3 anti-correlation still exists. These results indicate that the central region is mostly affected by extinction than the outer region, which is consistent with what we found in Section 6.3.1. In addition, we revisited the kinematics of real MUSE observations on NGC 5746 (Fig. 4 of Martig et al. 2021). According to our conclusions, we can confirm that the relatively smaller σ in the central region, and the high h_3 in the regions with $x \sim 60$ arcsec are due to the effect of extinction rather than some unusual LOSVDs caused by internal or external dynamic processes. However, their extinction effect is not as severe as in our figure.

6.3.3 Stellar Population Recovery

We show in Fig. 6.6 the stellar population maps (age, [M/H] and [α/Fe]) of the mock MUSE cubes generated similar to Section 5.1.1 but adding reddening with different factors. The first two rows are the true values and pPXF-measured mass-weighted stellar population properties on mock cubes without extinction, which is the same as Fig 5.4. Results from the third to the last rows show the stellar population properties measured by pPXF on mock cubes generated with the extinction model in Section 6.1 multiplied by different factors on the $E(B-V)$ values, respectively, as seen on the top right corner of the left panel in each row.

From this figure, we find the age distributions have a continuous change with the increase of reddening factor, where the very old region appears at $x \sim 0$ arcsec and moves upwards with the increase of reddening factor and extends towards the y-axis. The age of the outer thin disk region ($x \sim [60, 100]$, $y \sim [-10, 10]$) also has a slight increase. [α/Fe] also has similar features to the age distributions. [M/H] in the central region decreases with the increase of reddening factor and also extend towards the y-axis. To interpret these features, we first need to clarify the distribution of stellar ages, [M/H] and [α/Fe] in E-Galaxia. Our mock data have an inclination of 86°, and the ages at $(x, y) = (0, 0)$ are mainly from young stars ($<$ 6 Gyr) in the galaxy center, and old stars ($>$ 10 Gyr) with a larger scale height. Because of extinction, young stars in the galaxy center are heavily obscured, and the light in the central region is contributed mostly by old stars at a larger scale height. Therefore, the mass-weighted age in this region becomes larger with a higher reddening factor. As for the outward expansion of this region, we think it is because of the further young stars in the extended region are also heavily obscured when the reddening factor becomes large enough. [α/Fe] have similar features to age distributions because these two parameters correlate strongly. In addition, we can see a clear young and α-poor thin disk structure projected at $y \sim [-10, 0]$, $x \sim [-10, 110]$ and $y \sim [-10, 20]$, $x \sim [80, 100]$. Without extinction, this structure can not be seen. As for [M/H], metal-rich stars concentrated in the thin disk of the galaxy center are obscured by the dust. Therefore, the central mass-weighted [M/H] decreases with the increase of the reddening factor. The extension towards the y-axis is due to the obscuring of further metal-poor stars.

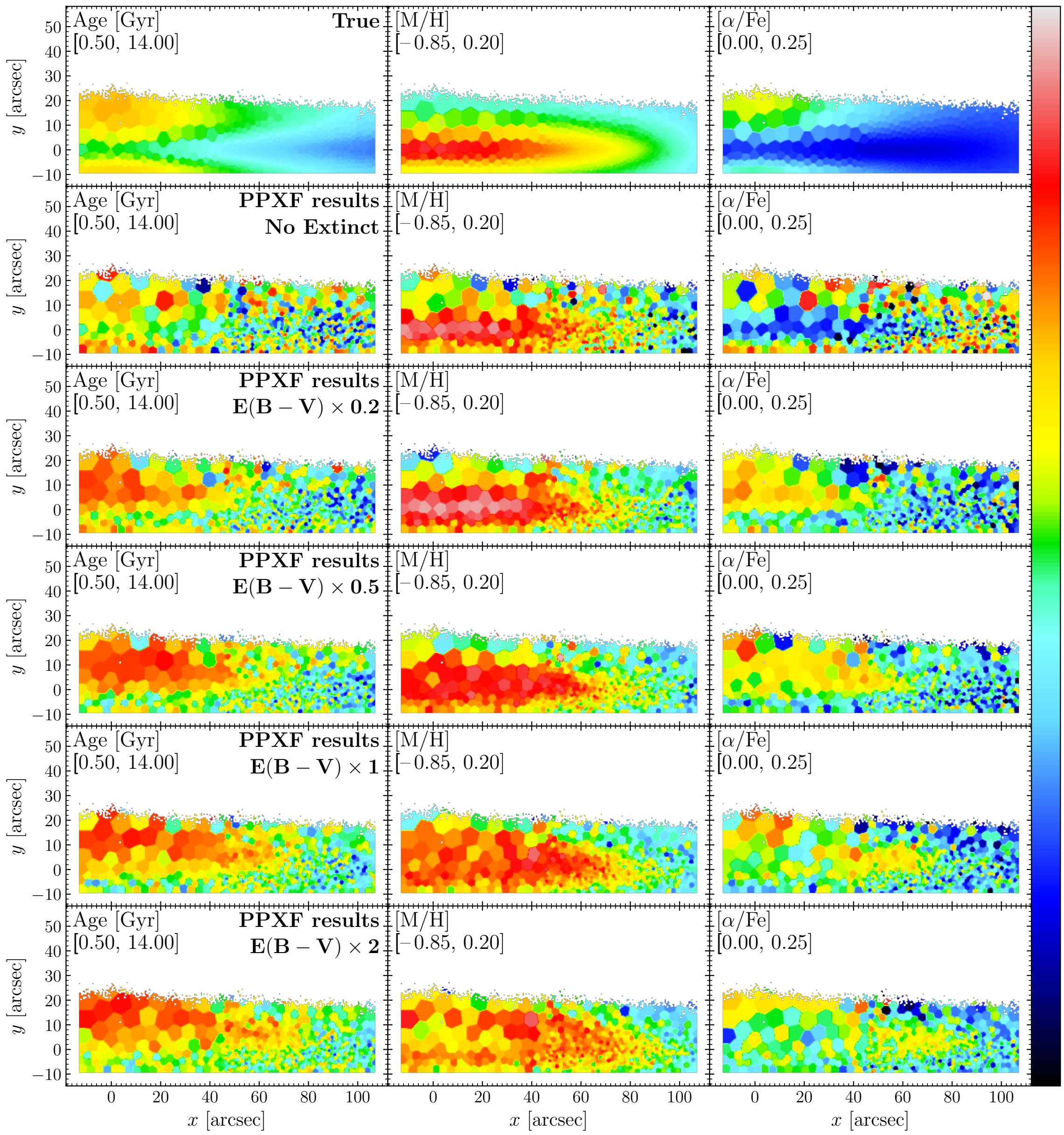

Figure 6.6: Stellar population maps (age, [M/H] and [α/Fe]) of the mock MUSE cubes generated similar to Section 5.1.1 but adding different extinction models. The first two rows are the true values and pPXF-measured mass-weighted stellar population properties on mock cubes without extinction, which is the same as Fig 5.4. Results from the third to the last rows show the stellar population properties measured by pPXF on mock cubes generated with the extinction model in Section 6.1 multiplied by different factors on the $E(B-V)$ values, respectively, as seen on the top right corner of the left panel in each row. All the results are measured and demonstrated in the same Voronoi binning arrangements.

In conclusion, due to the heavy extinction in the central region, the mass-weighted age and [α/Fe] are larger because of the dust obscuring young stars in the galaxy center, and for the same reason, these high-value regions have an extension to the outer thin disk structures. By looking at the stellar population parameter maps of NGC 5746 (Fig. 7 of Martig et al. 2021), we can confirm that the central old and α-rich distribution and the extended metal-rich regions along the thin disk can be due to the extinction.

6.3.4 Mass Fraction Distribution Recovery

We show in Fig. 6.7 the mass fraction distributions of the thin disk, thick disk, inner central, and upper central of the mock datacube generated similar to Section 5.1.1 but adding different extinction models. The first and second columns are the true values and pPXF-measured mass-weighted mass fraction distributions on mock cubes without extinction, which is the same as Fig 5.7. Results from the third to the last columns show the mass fraction distributions measured by pPXF on mock cubes generated with the extinction model in Section 6.1 multiplied by different factors on the $E(B-V)$ values, respectively, as seen on the top of each column. For the thin disk panels from left to right, we find the overdensity in $2 \sim 4$ Gyr is decreasing with the increase of reddening factor. This is due to the obscuring of young stars at further distances. The thick disk also demonstrates similar features, but the difference is the reduction in mass weights for metal-rich populations at $2 \sim 4$ Gyr. The upper central region shows a significant reduction in young populations. As discussed in Section 6.3.3, this is due to the severe obscuring of the galaxy center populations. Interestingly, this distribution is similar to that of the B/P bulge in NGC 5746 (Fig. 10 of Martig et al. 2021). The inner central has a mass distribution that more similar to the thin disk is also due to the heavy obscuring of stars in the galaxy center.

To study the effect of extinction on star formation and metallicity distribution recoveries, we show in Fig. 6.8 the age and [M/H] distributions of each component for the mock cubes generated similar to Section 5.1.1 but adding different extinction models. The true values, pPXF results without or within adding extinction models are indicated in different colors, as illustrated in the legend. Note we choose to plot mass fraction on the y-axis without dividing the bin size due to the uneven linear spacing of age, as discussed in detail in Section 5.2.2. For the star formation history (first and second row), we find that the overestimation at $2 \sim 4$ Gyr is smaller when adding extinction, but the oldest population has higher mass fractions. Interestingly, different reddening factors have similar effects on the star formation history. As for the metallicity distribution, we find with a higher reddening factor, [M/H] are moving toward metal-poor regions because young stars are obscured.

6.4 Summary and Future Studies

In this chapter, we explored the impact of extinction on the kinematics and stellar population parameters derived from pPXF. We used the $E(B-V)$ distribution map of Schlegel et al. (1998) as applied in Galaxia (Sharma et al., 2011). The flux distribution (Fig. 6.3 and Fig. 6.4) indicates a dark galaxy bulge and bright thin disk regions. For the central region, more than 90% of total flux is contributed by nearer stars ($d < 26.5$ Mpc) as further

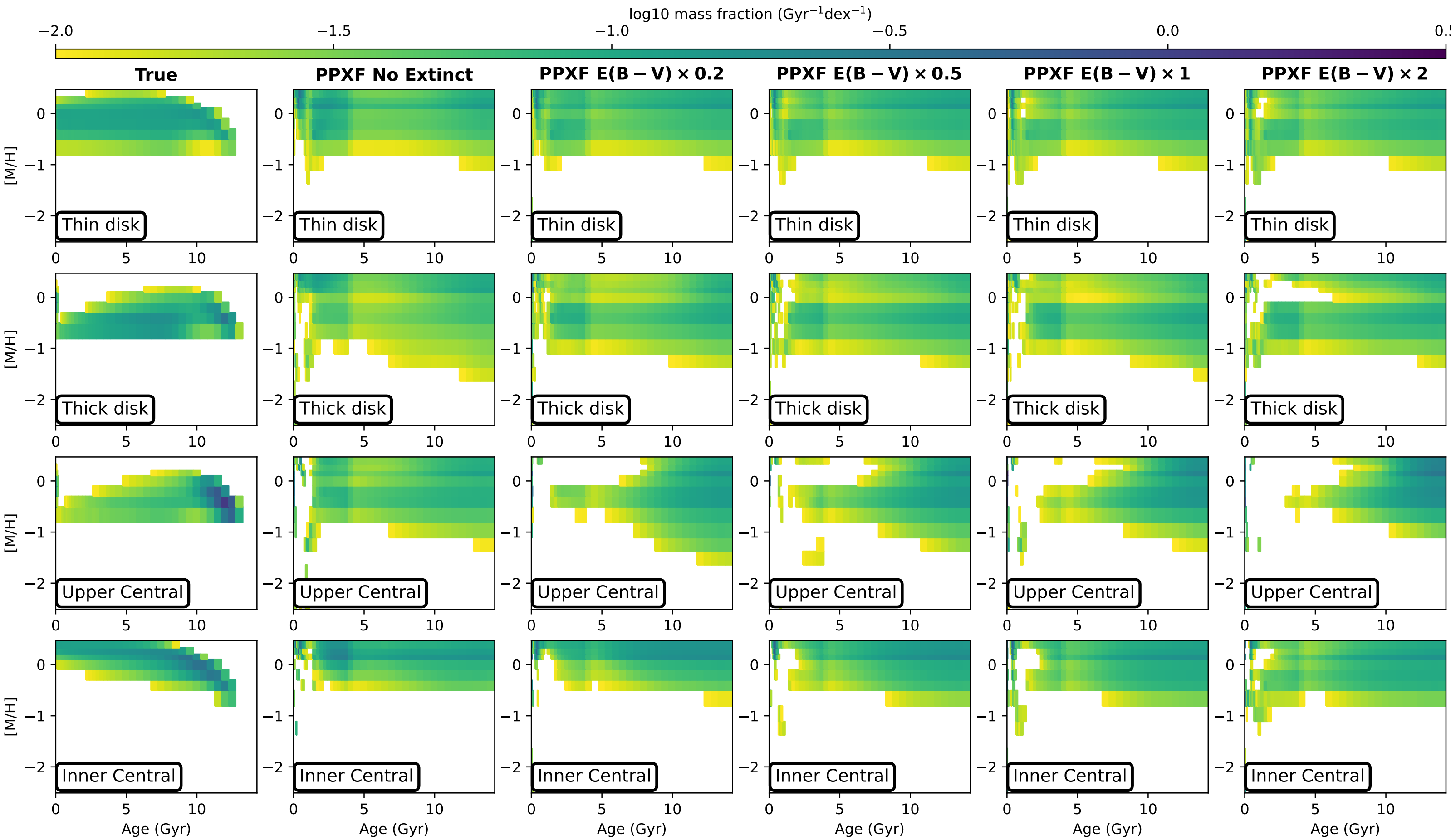

Figure 6.7: Mass fraction distributions of the thin disk, thick disk, inner central, and upper central of the mock datacube generated similar to Section 5.1.1 but adding different extinction models. The first and second columns are the true values and pPXF-measured mass-weighted mass fraction distributions on mock cubes without extinction, which is the same as Fig 5.7. Results from the third to the last columns show the mass fraction distributions measured by pPXF on mock cubes generated with the extinction model in Section 6.1 multiplied by different factors on the $E(B-V)$ values, respectively, as seen on the top of each column.

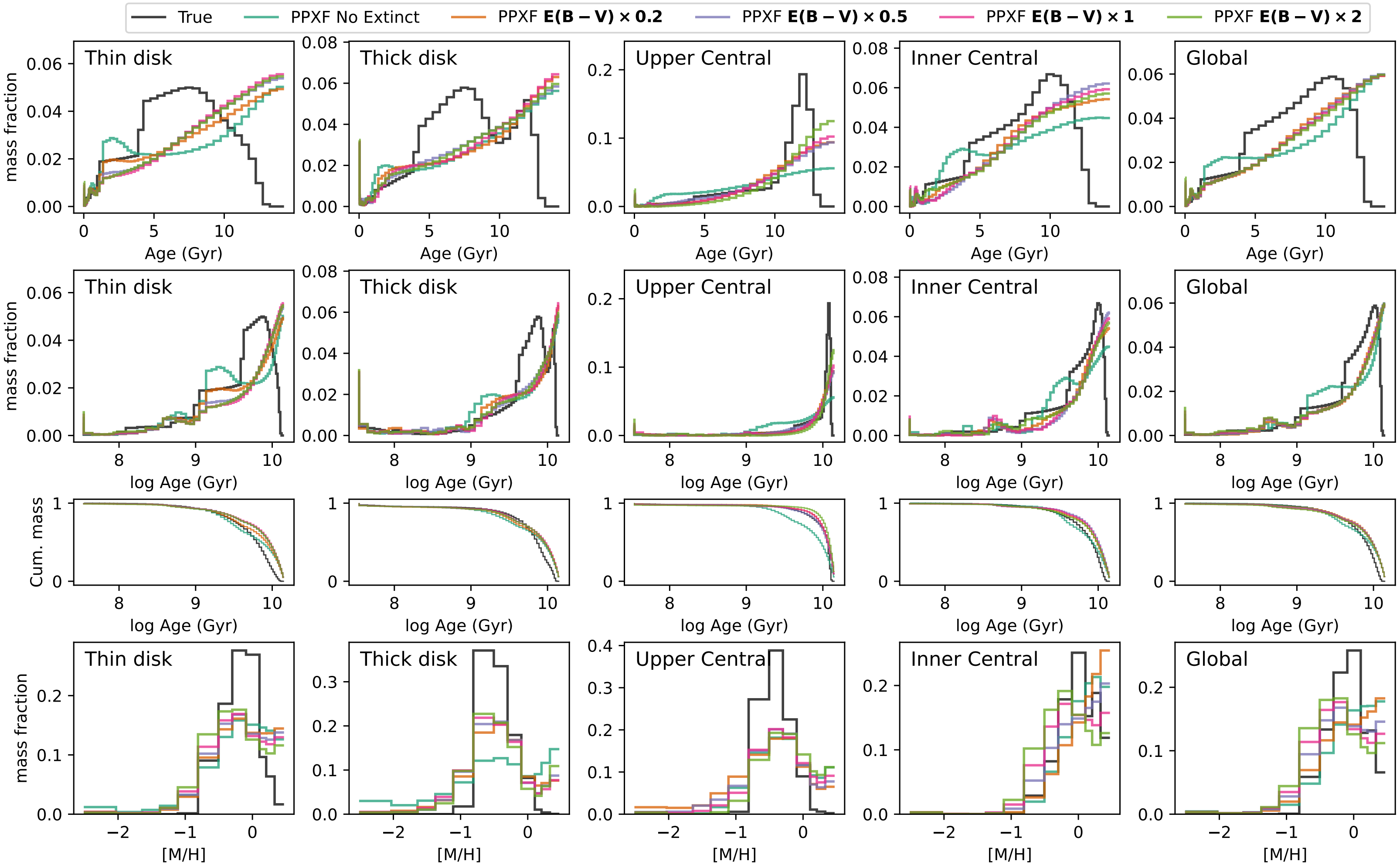

Figure 6.8: Age and [M/H] distributions of each component for the mock cubes generated similar to Section 5.1.1 but adding different extinction models. The true values, pPXF results without or within adding extinction models are indicated in different colors, as illustrated in the legend.

stars are fully extinct. This could be due to the lack of a bulge or bar component in the E-Galaxia catalog or an overestimation of the 3D $E(B-V)$ map. Future studies are needed to incorporate these stellar components and improve dust models for a more accurate flux distribution that matches real observations.

We created datacubes using the $E(B-V)$ particles in the E-Galaxia mock catalog, multiplied by different factors, to investigate the effect with different extinction models. These cubes were modified to ensure a similar SNR for each Voronoi bin before processing through the GIST pipeline to measure kinematics and stellar population parameters.

Our results show that for kinematics, severe extinction effects on central regions obscure populations in the galaxy center, leading to smaller absolute values of V_{LOS} and σ. Despite this, the V_{LOS}–h_3 anti-correlation still exists. For the stellar population, the obscuring of young and α-poor populations causes the inner and upper central regions to exhibit older and more α-rich population distributions extending further to the outer disk. As extinction increases, these effects become more dominant.

Upon comparison of these results with real MUSE observations of NGC 5746 (Martig et al., 2021), we identified several similar features, including smaller σ in the central region, high h_3 in the inner thin disk, and an older and α-rich center. Our extinction effects seem to be more severe than those from NGC 5746. These findings suggest that these features in real galaxy observations are due to extinction. Such conclusions could not be drawn in previous studies due to the lack of correct parameter distributions in real galaxies. With the power of GalCraft, we can address this question.

In the future, we will update the extinction model within GalCraft to the more recent and realistic dust model such as Bayestar19 (Green et al., 2019) and re-investigate our results. We also plan to test the parameter recovery ability by applying multiplicative Legendre polynomial and reddening curve simultaneously in the newest version of pPXF. More importantly, a B/P bulge with a bar should also be included to create more realistic Milky Way mock IFS observations.

Chapter 7

Thesis Summary and Prospects

In this thesis, I will first summarize the results in each chapter, and then highlight key future studies enabled by my work that can be undertaken to improve our understanding of the evolution of disk galaxies, as well as better link Galactic and extragalactic observations and analysis techniques.

7.1 Summary

We first applied *DD-Payne* (Xiang et al., 2019), which was developed for LAMOST spectra, on MUSE datacubes and estimated $T_{\rm eff}$, $\log g$, [Fe/H] and chemical abundances of Mg, Si, Ti, Ca for individual stars in MUSE datacubes. By comparing our MUSE results with that of LAMOST using common stars, we verified that the uncertainty of $T_{\rm eff}$, $\log g$ and other chemical abundances less than 75K and 0.15 dex and 0.1 dex, respectively. Furthermore, the [Fe/H]-[Mg/Fe] distribution estimated by our method in the bulge and inner disk region agrees with distributions from APOGEE-*Payne* in two sequences with similar center values and overall trends. This indicates the ability of our method to measure chemical abundances in regions with high stellar density.

In the future, we plan to use this method to survey nearly 100,000 individual stars contained in existing MUSE cubes and measure their stellar parameters. This sample is larger than many existing spectroscopic studies of the Milky Way and can be accomplished simply from the ESO archive of MUSE observations. This method can also be applied to MUSE observations on globular clusters to study the connection between kinematics and multiple populations. In the future, this method can also be used directly for BlueMUSE (Richard et al., 2019) and MAVIS (McDermid et al., 2020). These instruments offer higher spectral resolution and shorter wavelengths, making it possible to measure additional abundances with higher precision.

Next, we applied this method on a 3.6-hour MUSE observation towards one field in Baade's window with $A_V = 1.45$ mag and studied the chemical evolution and age distribution using MSTO-SGB stars in the Galactic bulge. Employing the method in Chapter 2 in combination with BSTEP (Sharma et al., 2018), we measured the stellar parameters ($T_{\rm eff}$, $\log g$, and chemical abundances), as well as the distance, extinction $E(B-V)$, and ages of stars in this field. The analysis resulted in 129 giant and 121 MSTO-SGB stars with $\mathrm{SNR_{MUSE}} > 80\ \mathrm{pix}^{-1}$, including 105 MSTO-SGB stars located within the bulge region with $R_{\mathrm{gc}} < 3.5$ kpc.

The age distribution exposed by this dataset suggests the presence of three groups with ages around 4 ~ 6 Gyr, 8 Gyr, and 10 ~ 12 Gyr, respectively. The [α/Fe]-[Fe/H] distribution of stars in the bulge indicates a "knee" at [Fe/H] = −0.5 dex, which agrees with Bensby et al. (2017) and the chemical evolution model of Sharma et al. (2021b). The [Fe/H] distribution can be separated into five different components, with the peaks consistent with previous studies of Bensby et al. (2017) within 1 ~ 2σ.

Despite only 3.6 hours of exposure time, this study obtained results comparable to those of previous studies. It highlights the effectiveness of MUSE in observing crowded regions like the Galactic center. Moving forward, we intend to apply for more MUSE time to observe additional low-extinction fields in the bulge. By combining stellar [Fe/H] with velocities, we aim to study the bulge chemodynamical distributions and examine how they vary across Galactic longitude and latitude. Giants can also be used to compare the distributions in the near and far sides of the Galactic center. Furthermore, the [α/Fe]-age relation at different birth radii proposed by Sharma et al. (2021b) is an extrapolation of the solar neighbourhood. Our datasets can expand this relation to the Galactic central regions and help refine chemical evolution models.

To enable a direct comparison of the Milky Way with extragalactic observations, we developed the GalCraft code which uses simple-stellar population models and a mock Milky Way catalog from E-Galaxia to generate Milky Way datacubes in the same data format as integral-field spectrograph (IFS) extragalactic observations. In this way, we eliminate the projection effect and differences in Galactic and extragalactic studies such as the use of individual stars vs. stellar populations, number density vs. mass- or light-weighted distributions, etc. The mock datacube can be analyzed using the GIST pipeline (Bittner et al., 2019) in the same way as extragalactic observations and the results can be compared directly to external galaxies to study the similarities and differences between the Milky Way and its analogues.

Using GalCraft, we tested the ability of pPXF (Cappellari & Emsellem, 2004; Cappellari, 2017) to recover the stellar kinematics and stellar population parameters on mock MUSE datacubes of the Milky Way placed at a distance of 26.5 Mpc and inclination 86°. We found that pPXF can recover chemodynamic maps, and both the [α/Fe]-rich and [α/Fe]-poor components can be identified with reasonable regularization values during the fitting. Using line-strength indices can also recover these components.

However, there are systematic offsets between the four kinematic moments (V_{LOS}, σ, h_3, h_4) measured by pPXF and the input values. This is because of two effects: first, the velocity dispersion σ of most Voronoi bins are smaller than the instrument velocity dispersion σ_{inst} of MUSE (FWHM = 2.65Å); additionally, the V_{LOS} changes with age and [M/H] for particles in each Voronoi bin, which is not considered in previous studies when using pPXF. By using higher spectral resolution templates PEGASE-HR (FWHM = 0.55Å) and randomly shuffling V_{LOS} for particles in each Voronoi bin, we can obtain consistent kinematic results compared to the true values. We indicate that it may be necessary to allow different templates to have different LOSVDs, even though this will increase the degeneracy significantly.

We found the stellar population from pPXF using MILES templates on MILES-generated mock cubes show deviations compared to the true values, where mass weights are overestimated in regions of $2 \sim 4$ Gyr, $12 \sim 14$ Gyr, the most metal-rich regions, and underestimated in the regions of $5 \sim 10$ Gyr. This might be due to the linearly uneven bin size of MILES templates' age grids. When repeating the test by employing pPXF and PEGASE-HR on PEGASE-HR-generated mock cubes, distributions are closer to the true value, with positive correlation coefficients between pPXF and the input values.

In the future, to obtain better [M/H] and [α/Fe] recoveries for stellar populations, it is necessary to use higher spectral resolution templates with more bins in [α/Fe], which will be essential to identify the mass-fraction of different [α/Fe] components or α-bimodality in more detail. Besides, we recommend changes to how pPXF determines the mass weights by considering the template grid size and spacing.

We also studied the effect of extinction on the pPXF recovered kinematics and stellar populations by adding the $E(B-V)$ distribution map of Schlegel et al. (1998) in GalCraft generated mock cubes. The flux distribution of suggests a dark galaxy bulge and bright thin disk regions. For the central region, more than 90% of the total flux is contributed by nearer stars ($d < 26.5$ Mpc) as further stars are fully extinct. This might be due to the lack of a bulge or bar component in E-Galaxia catalog or an overestimation of the 3D $E(B-V)$ map.

For the kinematics, we find due to the severe extinction effect on the central regions, populations in the galactic center are obscured and the absolute values of V_{LOS} and σ are smaller. Despite this, the V_{LOS}–h_3 anti-correlation still exists. For the stellar populations, due to the obscuring of young and α-poor populations, the inner and upper central regions demonstrate older and more α-rich population distributions which extend further to the outer disk. As extinction increases, these effects become more dominant.

We also compared these results with real MUSE observations of NGC 5746, and several similar features are identified, such as the relatively smaller σ in the central region, the high h_3 in the inner thin disk, an older and α-rich center, and the extended metal-rich regions along the thin disk. Our extinction effects seem to be more severe than those from NGC 5746. Therefore, we conclude these features in real galaxy observations are due to extinction. None of these features could be verified in previous studies due to the lack of correct parameter distributions in real galaxies. With the power of GalCraft, we can address this question.

In the future, we will update the extinction model within GalCraft to the more recent and realistic dust model such as Bayestar19 (Green et al., 2019) and re-investigate our results. With the newest version of pPXF, it can apply multiplicative Legendre polynomial and reddening curve simultaneously during the fitting. We will also test the recovery ability in this case. More importantly, a B/P bulge with a bar should also be added to mimic more realistic Milky Way mock IFS observations.

7.2 Prospects

Galactic studies can be done using individual stars. However, extending our Galactic knowledge to extragalactic galaxies requires studying spatially resolved stellar populations. We did both types of studies in this thesis by utilizing the MUSE integral-field spectrograph with its high spatial resolution and observation depth. The former by deriving stellar parameters for stars using MUSE spectra and the latter by generating and studying mock integral field MUSE cubes of Milky-Way-like disc galaxies. Hence, the MUSE instrument is the common thread linking the two different types of studies but focused on understanding the formation and evolution of disc galaxies.

With the future NASA Roman mission (Gaudi, 2022) and the possibility of *Gaia*-Infrared (Hobbs et al., 2016), a deeper photometric survey of the Galactic bar/bulge region can be carried out. Combining these with MUSE or future infrared IFS observations of the Galactic bulge, we can measure kinematics, elemental abundances and ages of hundreds of thousands of stars in the poorly explored dense inner regions of the Galaxy. This will enable us to unravel the chemodynamical evolution of the Galactic bar and the bulge, which presently is a mystery. Future fiber-fed multi-object spectroscopic surveys, see Table 1.2 for a list, will also make essential contributions. Similar studies can also be carried out on globular clusters to uncover their formation history and also to understand how they co-evolved with the Galactic halo.

The feasibility of studying individual stars in dense stellar regions using the MUSE instrument opens the way for studying individual stars in other galaxies tens of Mpcs away, e.g., the study of NGC 300 by Roth et al. (2018) with a long exposure time and better seeing conditions (< 0.6 arcsec). The future Extreme Large Telescope (Gilmozzi & Spyromilio, 2007) being planned with a very large aperture will improve the observation depth and signal-to-noise ratio of spectra from distant stars, making the so-called extragalactic archaeology feasible.

Our study to test the ability of pPXF to recover the kinematics and stellar populations of mock datacubes generated by GalCraft indicates many potential improvements. Modifying the pPXF and verifying its improvements is our next goal. If successful, we can reanalyze existing IFS observations, and arrive at better and more accurate inferences than before. GalCraft with extinction model also provides opportunities for qualitatively studying dust distributions in external galaxies.

The usage of GalCraft with future SSP models with more $[\alpha/\text{Fe}]$ bins provides a great opportunity to explore the α-bimodality at different locations of external galaxies, similar to the study done on the Milky Way (Hayden et al., 2015). With a spatially resolved chemical composition distribution, we can address the stellar radial migration and kinematic heating rate of external galaxies for the first time, and will be able to compare them to that of the Milky Way. This will become possible with the ongoing and upcoming large VLT/MUSE programs and other IFU surveys of face-on and edge-on nearby galaxies, see Table 1.3 for a detailed list, and will push the extragalactic studies into a new era.

Appendix A

The [Mg/Fe]-Temperature trend in LAMOST-*DD-Payne-A* and APOGEE-*Payne* catalogs

The [Mg/Fe]-$T_{\rm eff}$ trends was found in the *DD-Payne* model in X19 (see Figure 14), where for dwarfs with $-0.2 <$ [Fe/H] < 0.2, [Mg/Fe] measured by *DD-Payne-A* decreases with $T_{\rm eff}$. In addition to [Mg/Fe], some other abundances also show similar non-negligible trends. X19 verified that these trends are inherited from the APOGEE-*Payne* values in the training set. Since no abundance calibration was applied to APOGEE-*Payne* values (Ting et al., 2019), the [Mg/Fe]-$T_{\rm eff}$ trend should come from the neural network model *The Payne*, i.e., stellar astrophysics.

To investigate this trend, we take APOGEE-*Payne* (Ting et al., 2019) and LAMOST-*DD-Payne-A* (X19) catalogs, separate stars into giants and dwarfs and divide them into different [Fe/H] bins, then plot the [Mg/Fe]-$T_{\rm eff}$ distributions in Figure A.1. Selection criteria are applied to these two samples to discard stars with bad fittings: For APOGEE-*Payne*, excluding stars with any of the flags: "quality"="bad" and $\chi^2 > 20$; For LAMOST-*DD-Payne-A*, excluding stars with any of the flags: SNR$_{\rm g} < 30$, $\chi^2 > 20$, "qflag-chi2"="bad" and "flag-singlestar"="no". The dots and filled ranges illustrate the median values and the standard deviations at different $T_{\rm eff}$ bins. In this Figure, both the giants and dwarfs demonstrate that there is a clear $T_{\rm eff}$-[Mg/Fe] trend in all the metallicity bins. The slopes of these trends (illustrated in Table A.1) in the two catalogs are similar as well.

Another reason leading to this trend is the Galactic evolution and selection functions, i.e., stars with higher $T_{\rm eff}$ tend to be younger, and therefore have lower [Mg/Fe] than stars with lower $T_{\rm eff}$. Even though the metallicity bins in Figure A.1 are chosen to be small, it is not sufficient to conclude the [Mg/Fe]-$T_{\rm eff}$ trend is all caused by the model. To explore the proportion of this trend due to the model, we select seven open clusters from LAMOST DR5 and plot [Mg/Fe] as a function of $T_{\rm eff}$ in Figure A.2, where the name and star number are marked in each panel. The stars in the open clusters are selected by the criteria in Zhong et al. 2020. Their [Mg/Fe] and $T_{\rm eff}$ are then measured by the *DD-Payne* models. In all the panels, results from *DD-Payne-A* show clear [Mg/Fe]-$T_{\rm eff}$ trends with larger slopes than those from *DD-Payne-G*, which is -0.082 dex/1000K on average. Since open

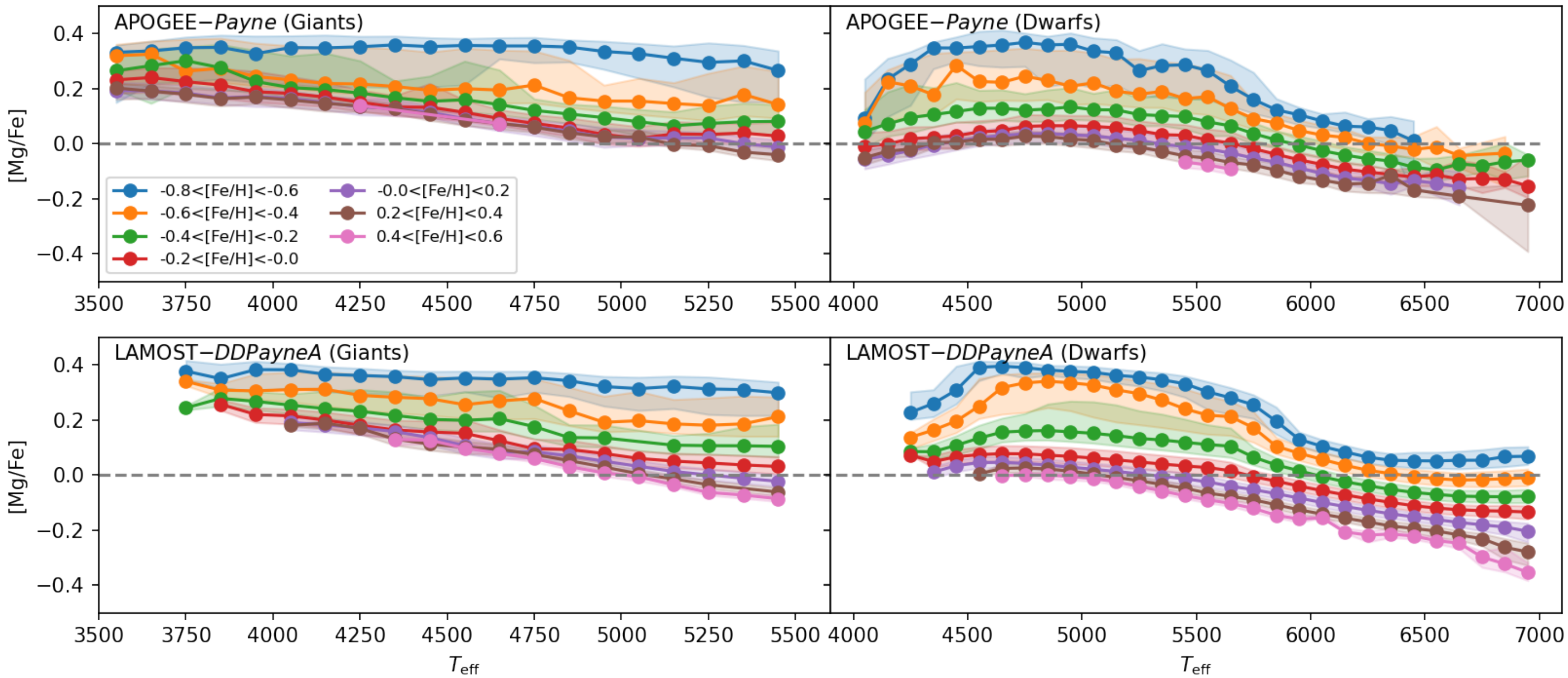

Figure A.1: Stellar [Mg/Fe] as a function of $T_{\rm eff}$ for stars in different metallicity bins. The top row shows the results from APOGEE-*Payne* (Ting et al., 2019), the the bottom row the LAMOST-*DD-Payne-A* results (X19). The dots and filled ranges illustrate the median values and standard deviations at different $T_{\rm eff}$ bins. Both of the results demonstrate that there is a $T_{\rm eff}$-[Mg/Fe] trend in all the metallicity bins. The slopes of these trends are illustrated in Table A.1. In addition, the trends in APOGEE-*Payne* and LAMOST-*DD-Payne-A* are consistent because *DD-Payne-A* is trained from labels in APOGEE-*Payne* and inherited its features.

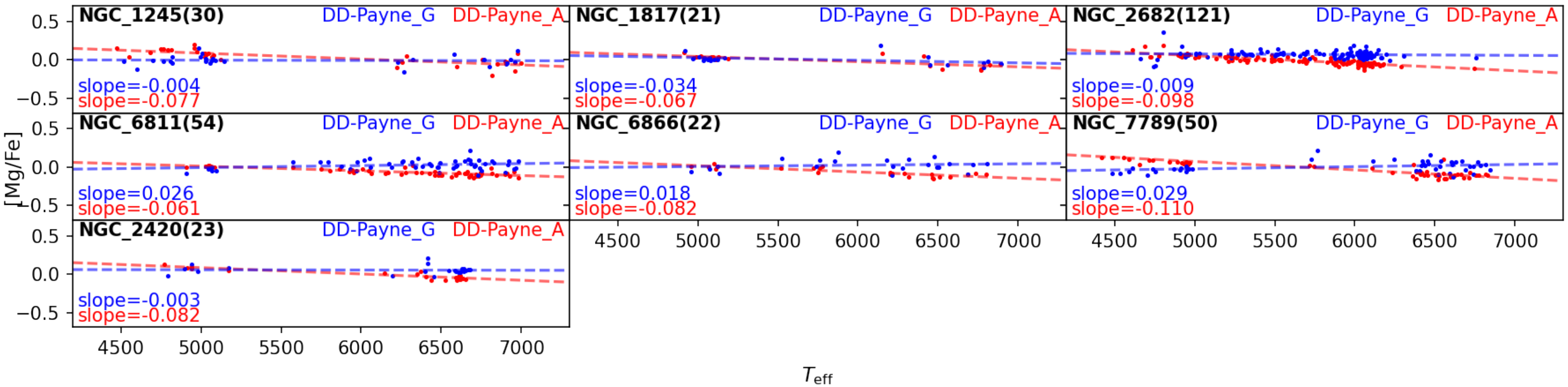

Figure A.2: Stellar [Mg/Fe] as a function of $T_{\rm eff}$ for stars in seven open clusters (the name and stellar number are marked in each panel). The stars are from LAMOST DR5 and selected by the criteria in Zhong et al. 2020. Their [Mg/Fe] and $T_{\rm eff}$ are then measured by the *DD-Payne* models. The results measured by *DD-Payne-A* show clear [Mg/Fe]-$T_{\rm eff}$ trends with larger slopes than *DD-Payne-G*. Since the metallicity and abundances of stars in open clusters are homogeneous, which means no galactic evolution or selection function applied, the trends illustrated here are caused by labels in APOGEE-*Payne*.

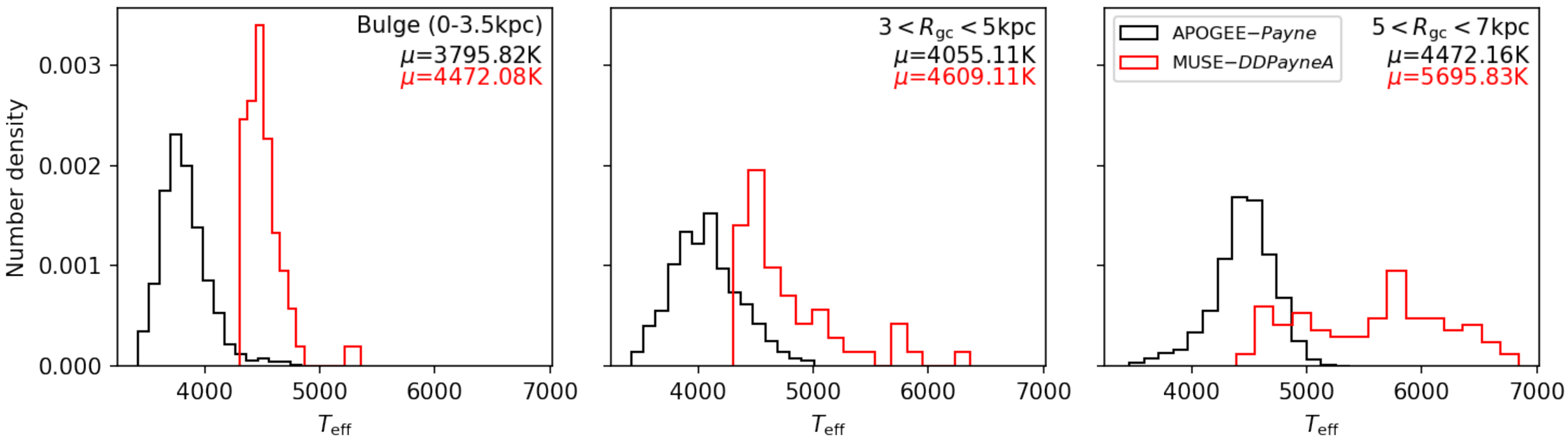

Figure A.3: T_{eff} density histograms of stars in Figure 2.12 as a function of R_{gc}. Results from APOGEE-*Payne* (Ting et al., 2019) are illustrated in black, and the red shows the MUSE-*DD-Payne-A* results. The medians (μ) are marked on each panel. This figure shows the difference of T_{eff} in the APOGEE-*Payne* and MUSE-*DD-Payne-A* samples, which can lead to a shift in [Mg/Fe].

Table A.1: Slope for the [Mg/Fe]-$T_{\rm eff}$ trend (dex/1000K) in different metallicity bins for stars in APOGEE-*Payne* and LAMOST-*DD-Payne-A*.

[Fe/H] bins	APOGEE-*Payne* giants	APOGEE-*Payne* dwarfs	LAMOST-*DD-Payne-A* giants	LAMOST-*DD-Payne-A* dwarfs
$(-0.8, -0.6)$	-0.025	-0.117	-0.044	-0.141
$(-0.6, -0.4)$	-0.086	-0.105	-0.092	-0.132
$(-0.4, -0.2)$	-0.125	-0.080	-0.111	-0.098
$(-0.2, 0.0)$	-0.128	-0.071	-0.142	-0.093
$(0.0, 0.2)$	-0.118	-0.064	-0.162	-0.104
$(0.2, 0.4)$	-0.131	-0.078	-0.186	-0.132
$(0.4, 0.6)$	-0.165	-0.130	-0.212	-0.154

clusters are found to be chemical homogeneous (Bovy, 2016; Ness et al., 2018), which means no galactic evolution or selection function is applied, the trend illustrated here is verified to be caused by the model. Combining with the slopes in Figure A.1, for stars with [Fe/H] $\sim$ 0, which is the approximate [Fe/H] for these open clusters, the average slope is -0.124 dex/1000K. Therefore, the contribution of this trend is comparable but higher from the model (stellar astrophysics).

In APOGEE-*Payne* (Ting et al., 2019), the authors chose to not calibrate the abundance trend because at the $T_{\rm eff}$ of APOGEE range (4000K-5000K), it was shown that an ab-initio fitting of the spectra will not incur much [Mg/H]-$T_{\rm eff}$ trend (see their Figure 15). However, we found that even with the ab-initio fitting, for stars with a wider $T_{\rm eff}$ range, in the [Mg/Fe] scale, the values in APOGEE-*Payne* do have a non-negligible trend, which is similar to the APOGEE pipeline ASPCAP (Holtzman et al., 2015) and requires calibration.

Therefore, when plotting the [Mg/Fe]-[Fe/H] distributions, additional calibration should be added to eliminate this impact. In Figure A.3, we plot the $T_{\rm eff}$ density histograms of stars in Figure 2.12 in different $R_{\rm gc}$ regions. Results from APOGEE-*Payne* are illustrated in black, and the red shows the MUSE-*DD-Payne-A* results. The medians (μ) are marked on each panel. This figure illustrates the difference of $T_{\rm eff}$ in the APOGEE-*Payne* and MUSE-*DD-Payne-A* samples. Stars with $5 < R_{\rm gc} < 7$ kpc in MUSE-*DD-Payne-A* are mostly dwarfs, and the other samples are mostly giants. Referring to Figure A.1, we can roughly estimate that for stars in the bulge, $3 < R_{\rm gc} < 5$ and $5 < R_{\rm gc} < 7$ kpc, the deviations of [Mg/Fe] between two samples are $\sim 0.08, 0.06$ and 0.08 dex, respectively. These are the values added to [Mg/Fe] in Figure 2.12 for calibrating.

Appendix B

Modifications of the GIST pipeline

We provide a list of modifications on the GIST pipeline (Bittner et al., 2019) to improve the flexibility of this software: We add more templates such as the original and interpolated PEGASE-HR (Le Borgne et al., 2004) to the software and allow the option to oversample the spectra when degrading to lower observation spectral resolution, which is for the same reason as we do for the GalCraft code in Section 4.2.3; We also added the option to select SSP templates with a certain age and [M/H] range in some special cases. When measuring stellar kinematics and stellar population properties, we add the options to choose `velscale_ratio`, `reg_ord`, and use real spectra noise during the fitting and normalizing the integrated spectrum and noise by the median of the spectrum, which will be important to perform the iteration for estimating R_{max}; For regularization, we add an option to estimate R_{max} following the procedures in McDermid et al. (2015) and save the stellar population results with $R = R_{max}$. We also allow further degrading to lower spectral resolution during the fitting when measuring SFH, which is mainly for the reduction procedures in M21 where they wanted to compare the mass-weighted stellar population parameters from pPXF with the SPP-equivalent parameters from line-strength indices in the same spectral resolution; We add the option to change the penalization value `bias` during pPXF fitting and measure mass weights uncertainties using bootstrapping following procedures in Kacharov et al. (2018).

Appendix C

Additional Kinematic and Stellar Population Results

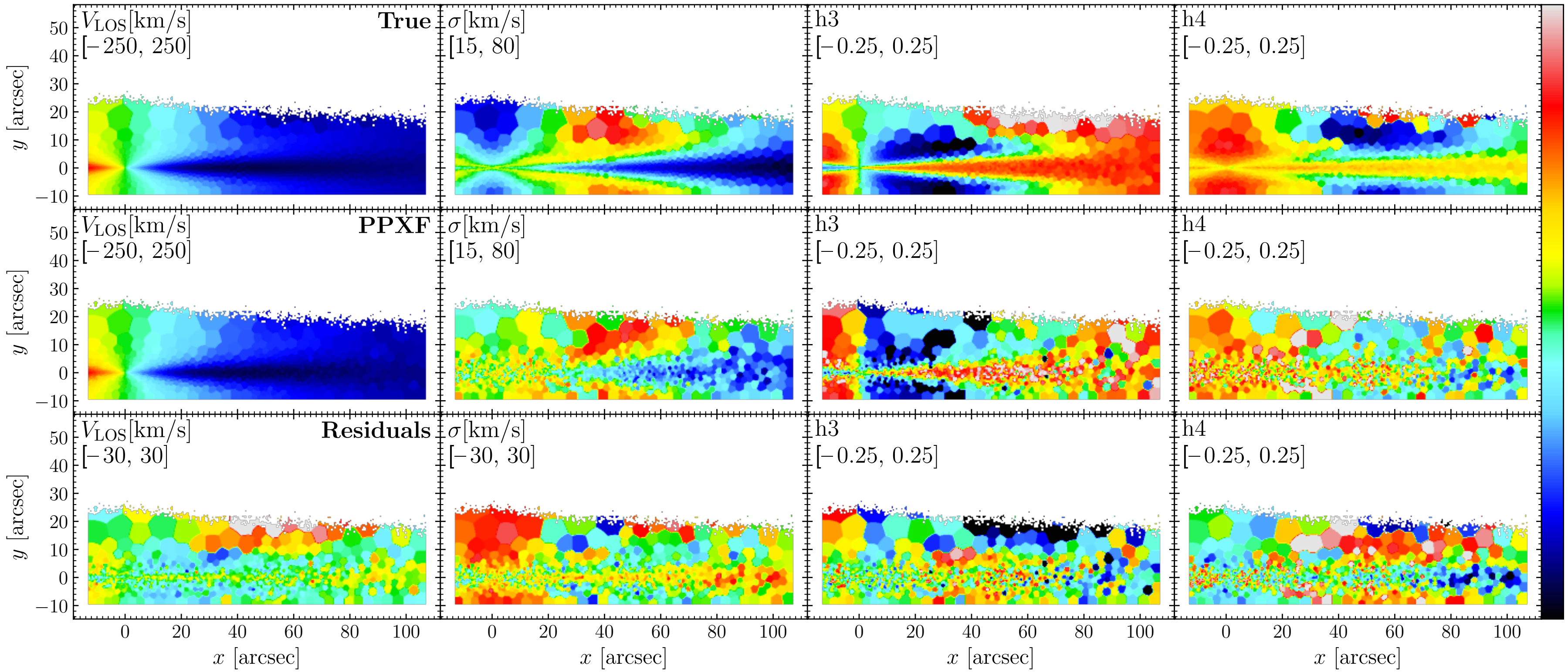

Figure C.1: Same as Fig. 5.3 but turning off the kinematics penalization during PPXF fitting.

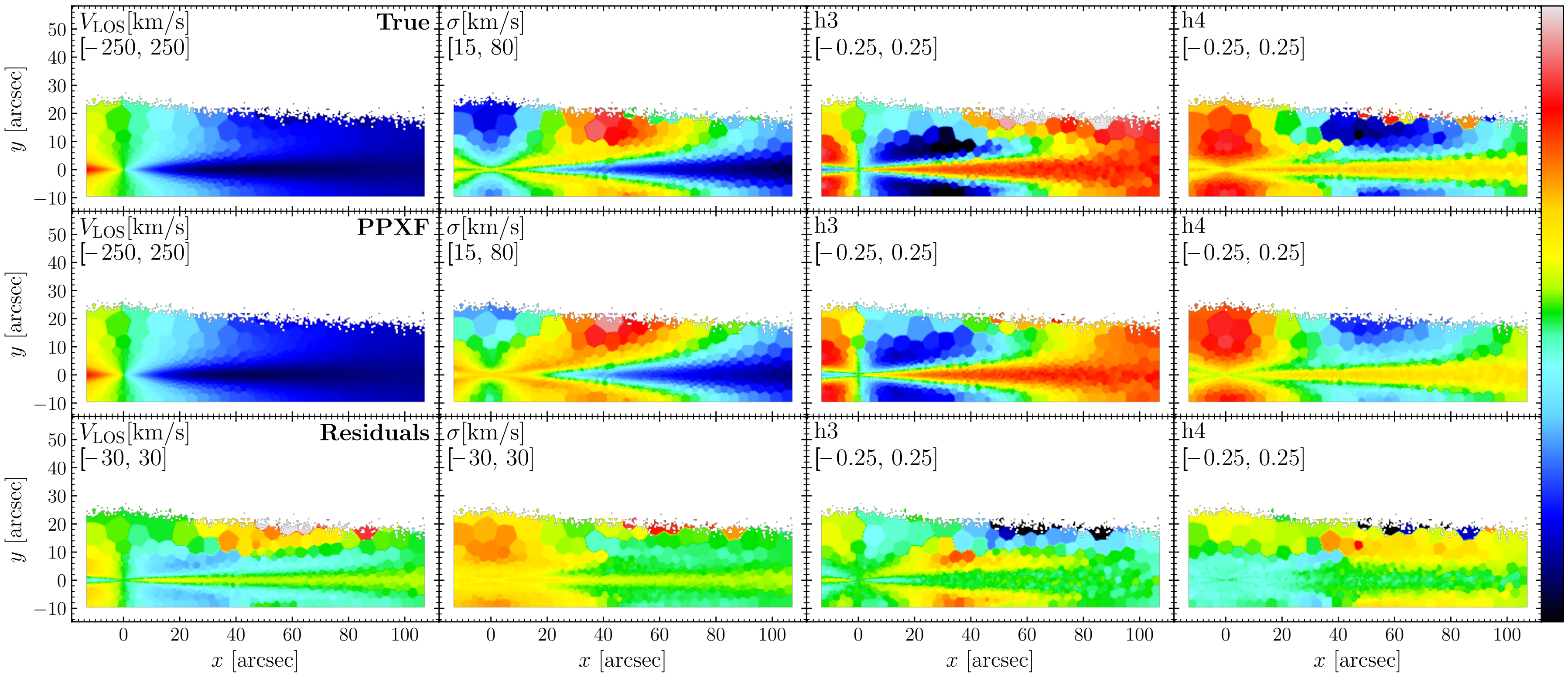

Figure C.2: Same as Fig. 5.3 but for the mock cube generated and fitted by PEGASE-HR templates in PEGASE-HR spectral resolution (FWHM = 0.55Å).

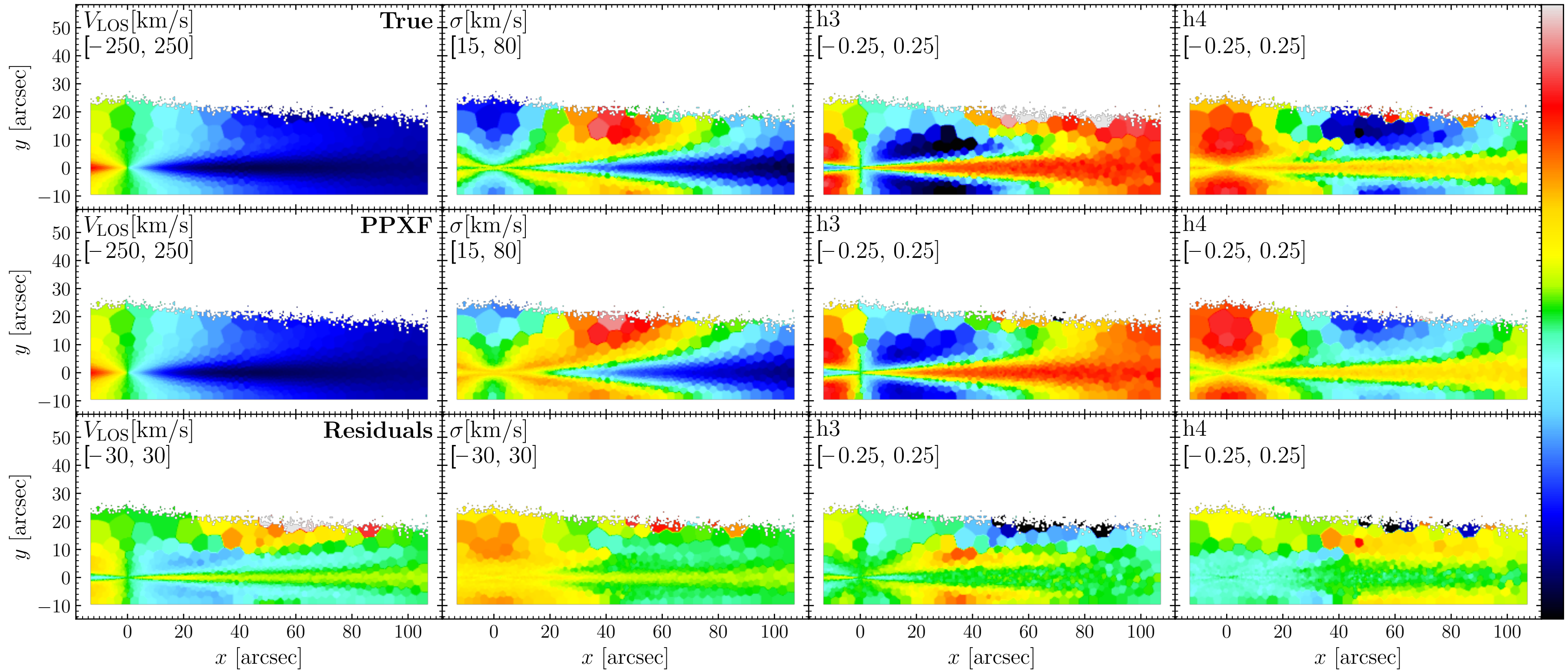

Figure C.3: Same as Fig. C.2 but turning off the kinematics penalization during pPXF fitting.

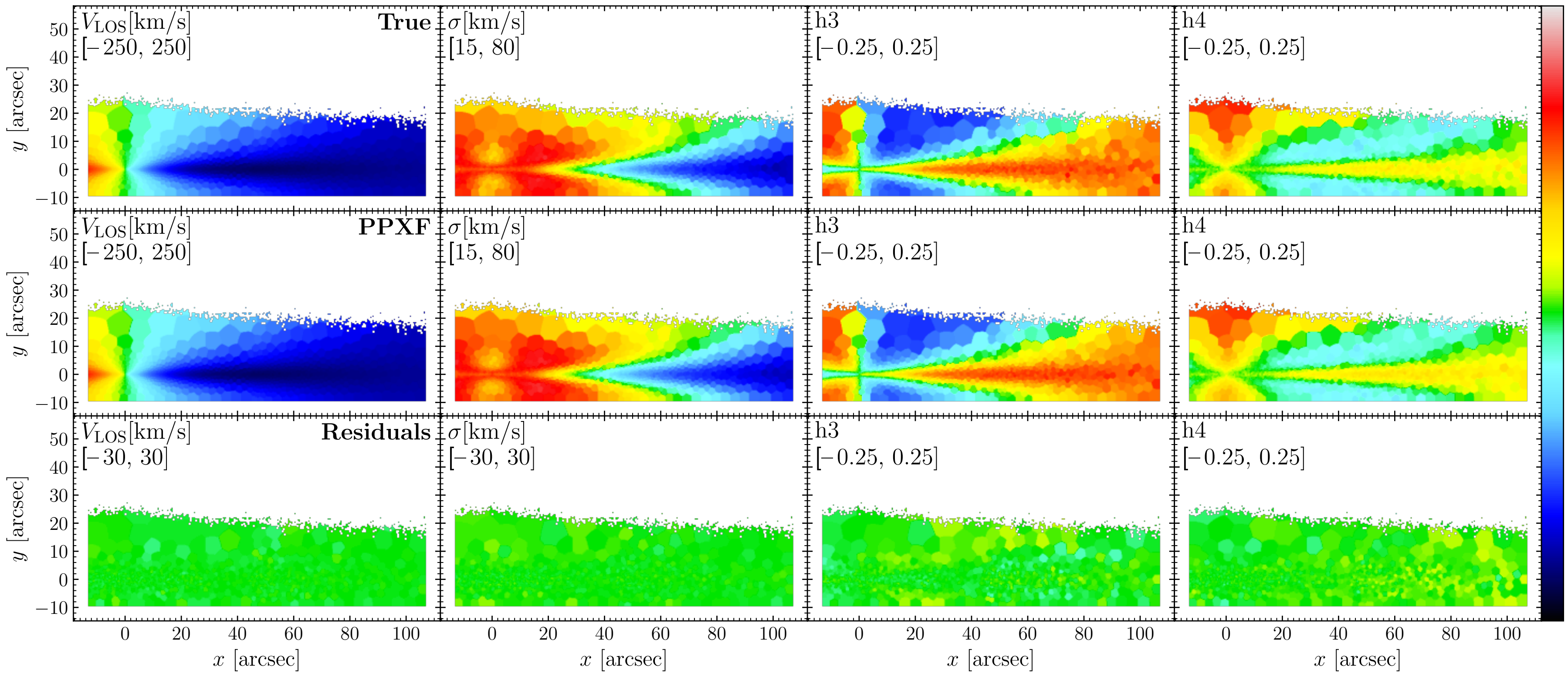

Figure C.4: Same as Fig. 5.3 but for the mock cube generated from using V_{LOS} randomly shuffled GALAXIA catalog and fitted by PEGASE-HR templates in PEGASE-HR spectral resolution (FWHM = 0.55Å).

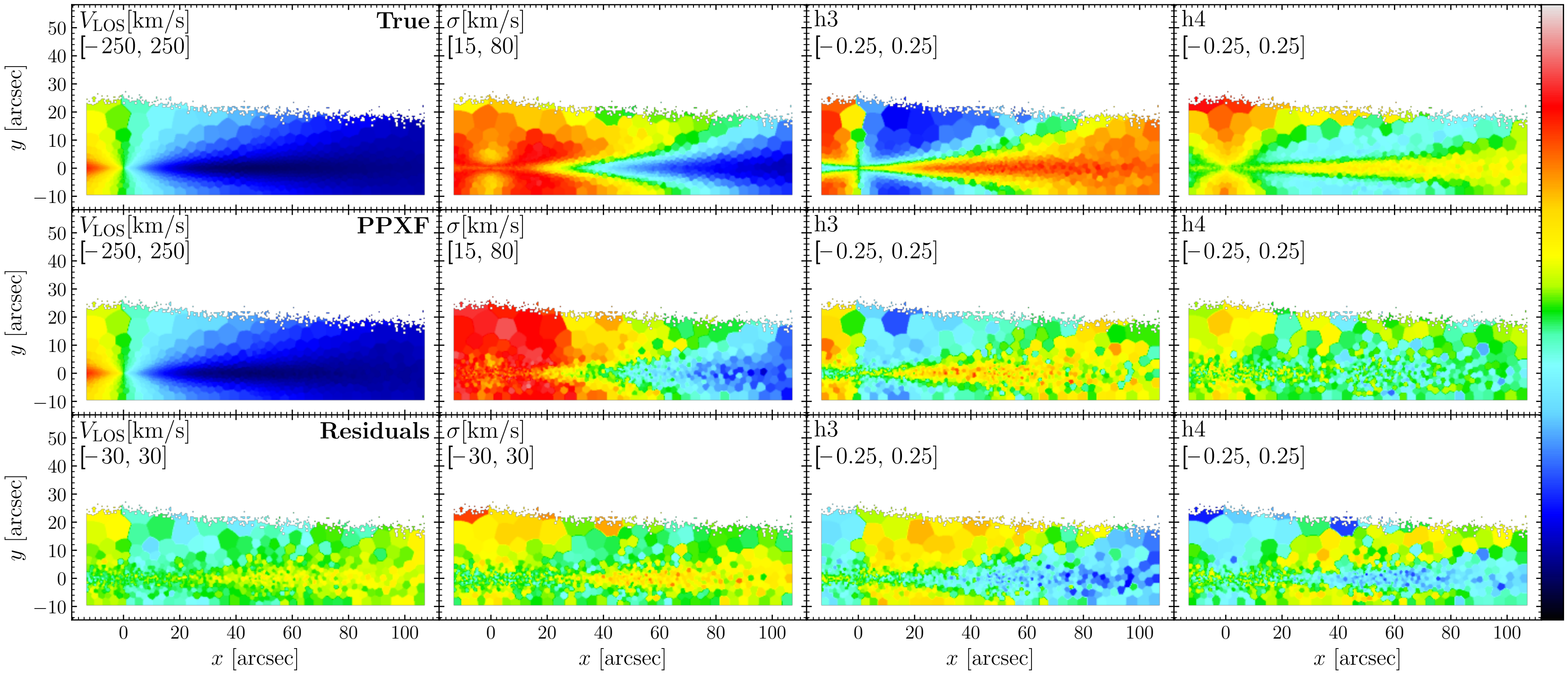

Figure C.5: Same as Fig. 5.3 but for the mock cube generated from using V_{LOS} randomly shuffled GALAXIA catalog.

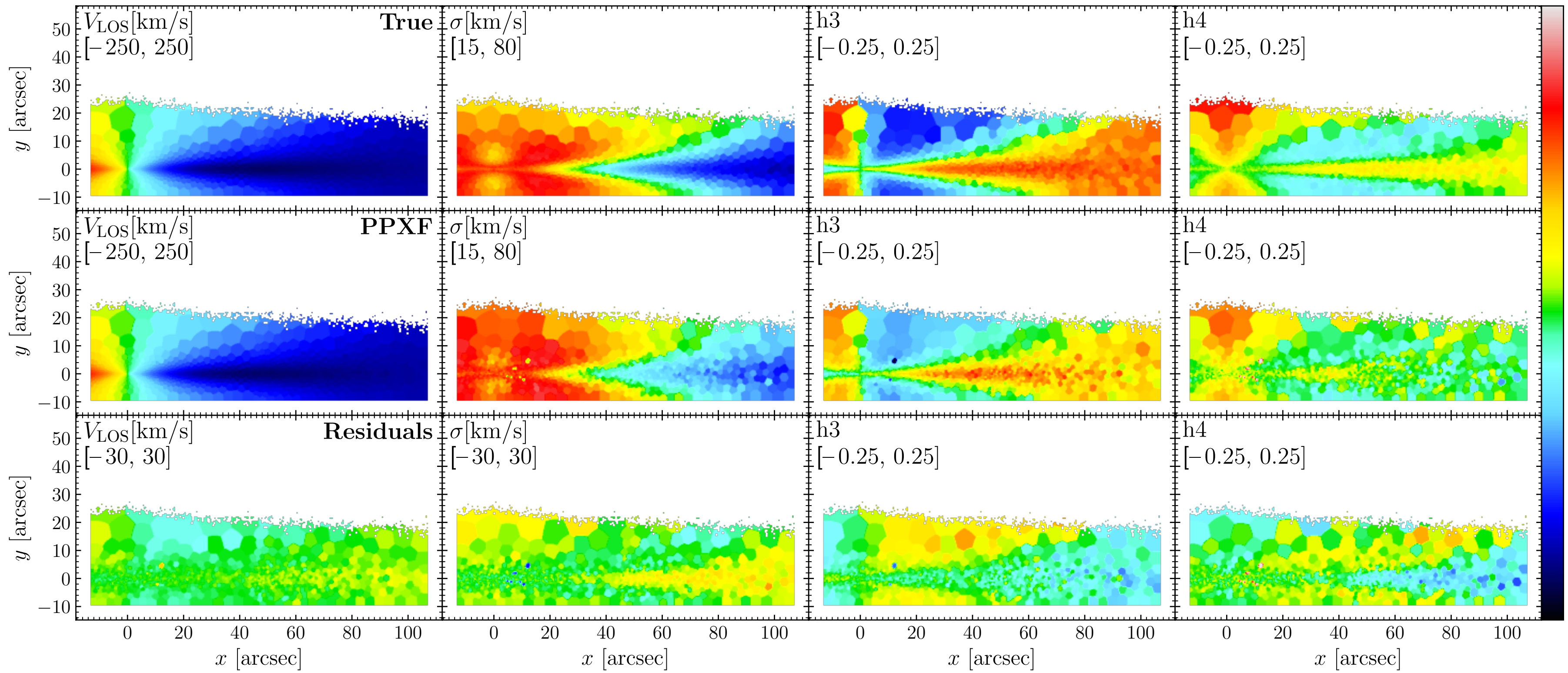

Figure C.6: Same as Fig. 5.3 but for the mock cube generated from using $V_{\rm LOS}$ randomly shuffled GALAXIA catalog and fitted by PEGASE-HR templates in MUSE spectral resolution (FWHM = 2.65Å).

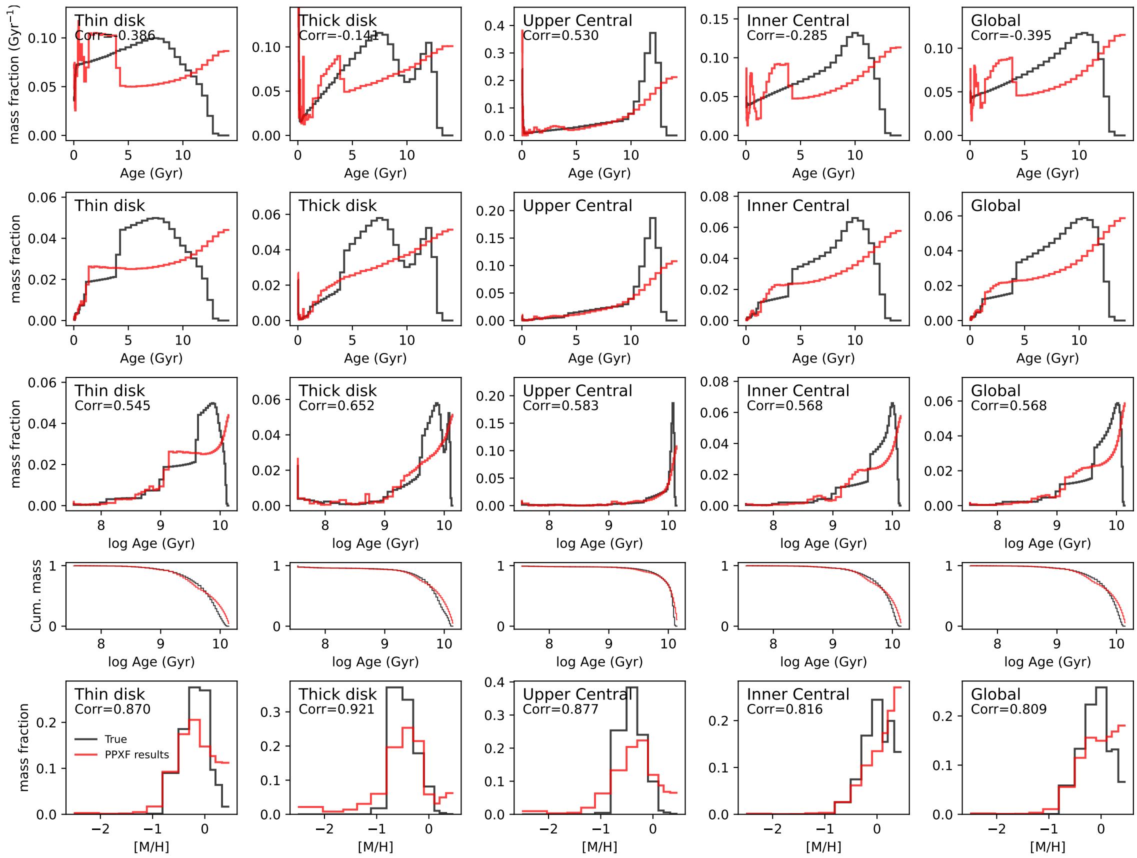

Figure C.7: Same as Fig. 5.8 and Fig. 5.12 but for the mock cube generated from using V_{LOS} randomly shuffled GALAXIA catalog using MILES templates in MUSE spectral resolution (FWHM = 2.65Å).

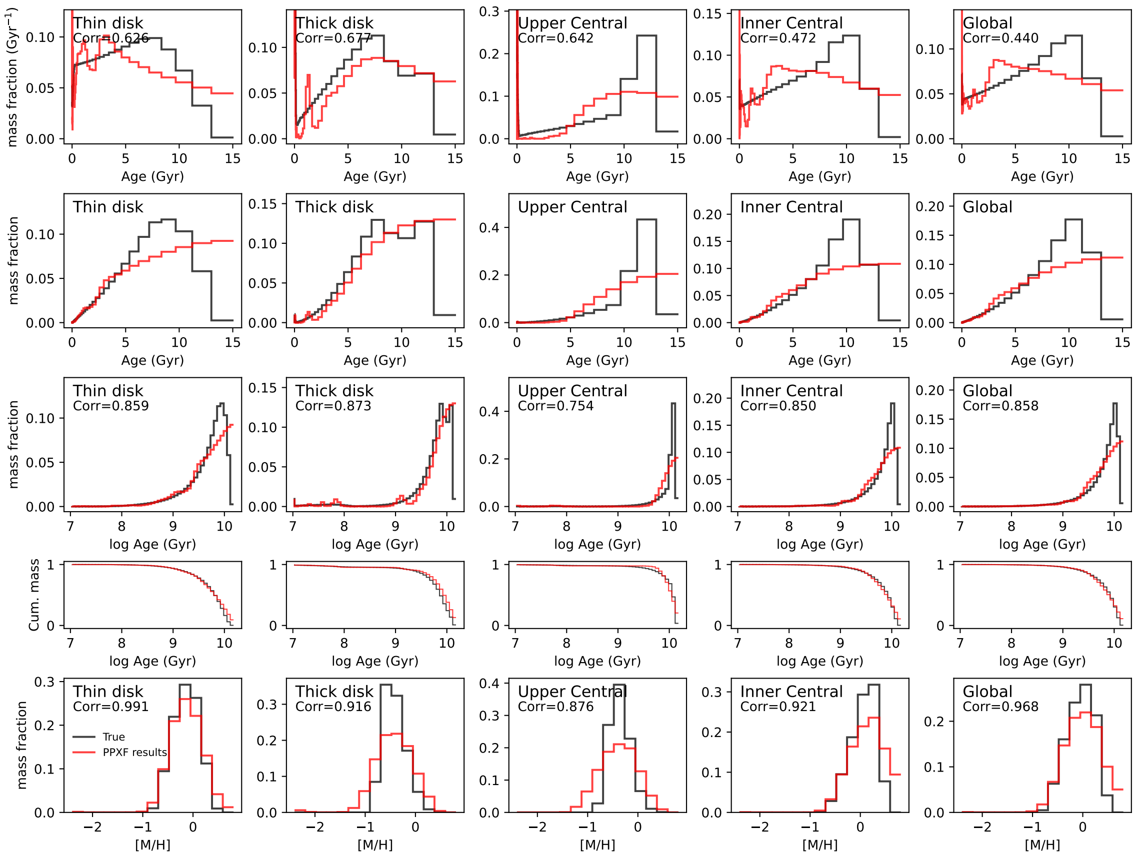

Figure C.8: Same as Fig. 5.15 but for the mock cube generated using PEGASE-HR templates in PEGASE-HR spectral resolution (FWHM = 0.55Å).

Appendix D

Additional Stellar Population Results by Applying Multiplicative Polynomials

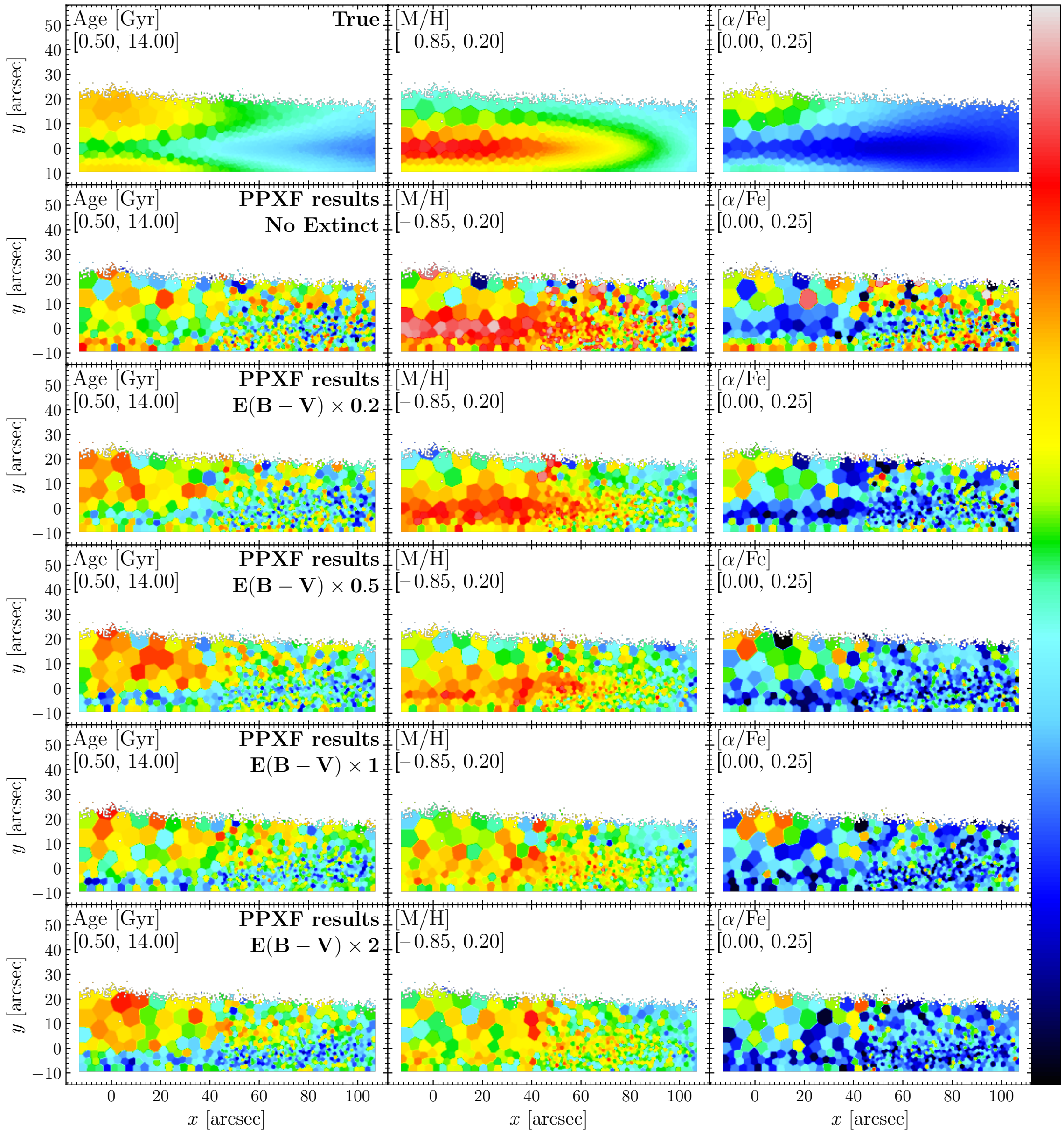

Figure D.1: Stellar population maps similar to Fig. 6.6 but obtained by applying an 8th-order multiplicative polynomial rather than a reddening curve in PPXF.

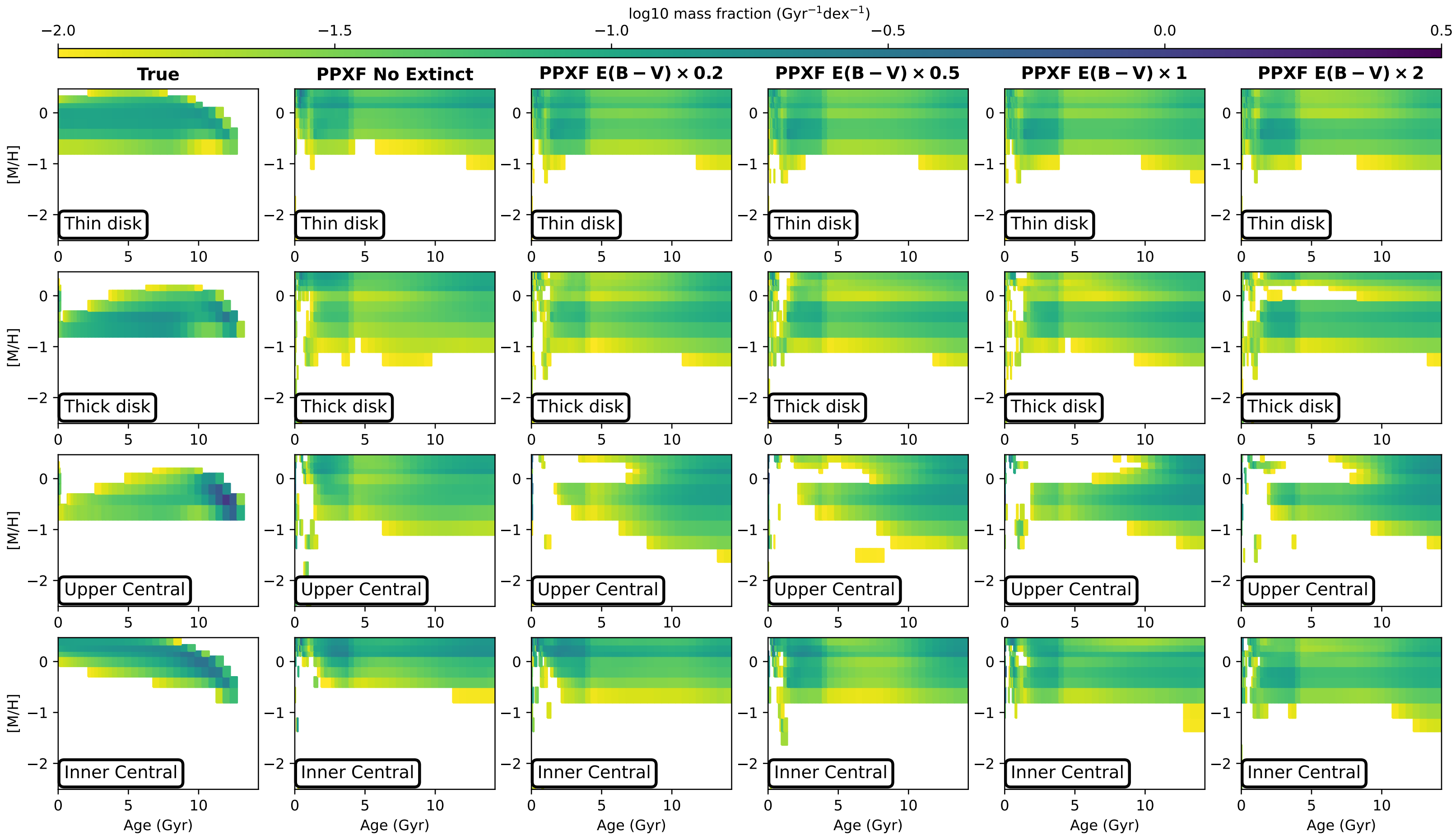

Figure D.2: Mass fraction distributions similar to Fig. 6.7 but obtained by applying an 8th-order multiplicative polynomial rather than a reddening curve in pPXF.

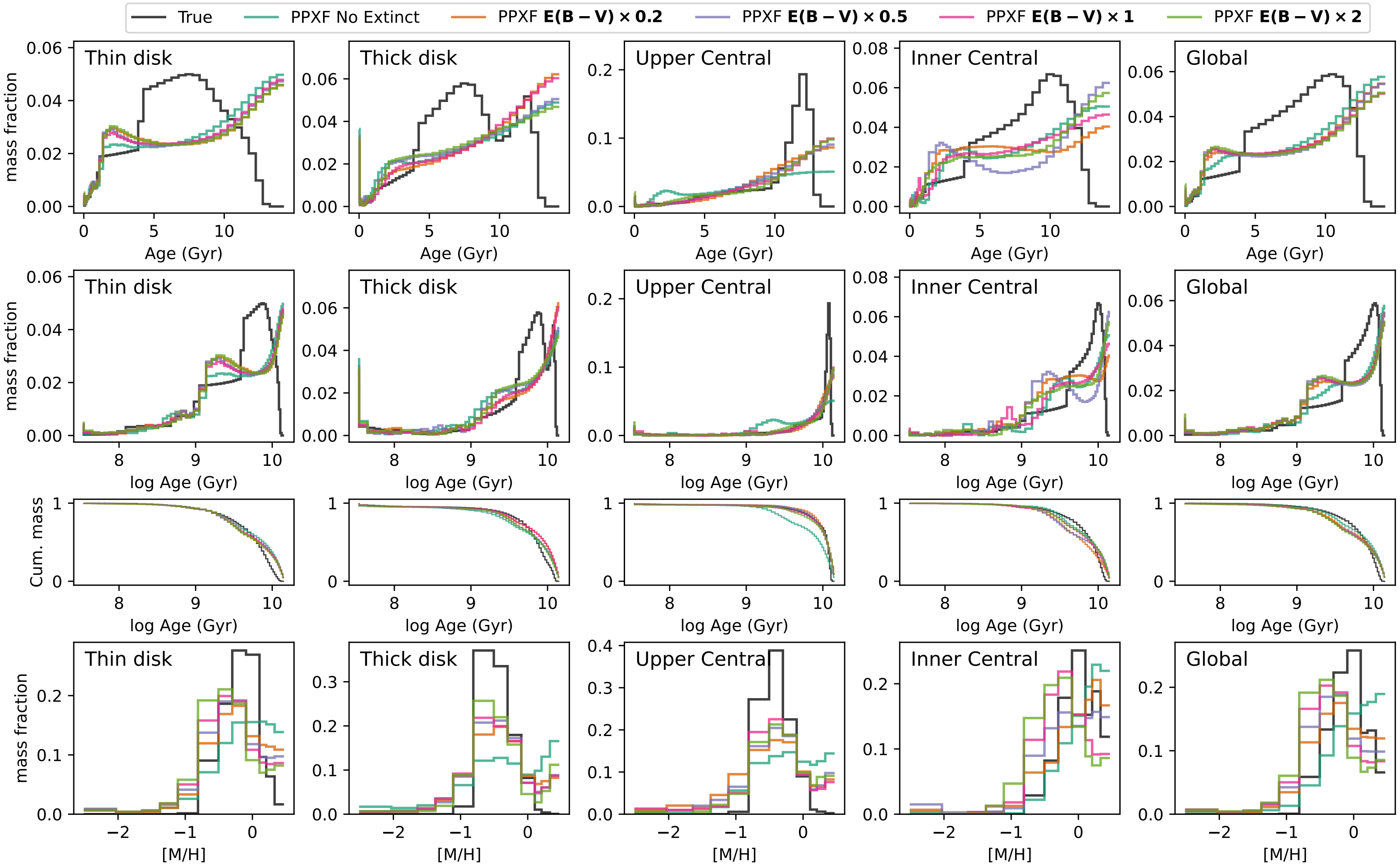

Figure D.3: Age and [M/H] distributions similar to Fig. 6.8 but obtained by applying an 8th-order multiplicative polynomial rather than a reddening curve in PPXF.

Bibliography

Abadi M. G., et al., 2003, ApJ, 597, 21

Abazajian K. N., et al., 2009, ApJS, 182, 543

Abdurro'uf et al., 2022, ApJS, 259, 35

Abolfathi B., et al., 2018, ApJS, 235, 42

Ahumada R., et al., 2020, ApJS, 249, 3

Almeida A., et al., 2023, arXiv e-prints, p. arXiv:2301.07688

Alvarez R., Plez B., 1998, A&A, 330, 1109

Andrae R., Rix H.-W., Chandra V., 2023, arXiv e-prints, p. arXiv:2302.02611

Andrews B. H., Martini P., 2013, ApJ, 765, 140

Antoja T., et al., 2018, Nature, 561, 360

Arce H. G., Goodman A. A., 1999, ApJ, 512, L135

Astropy Collaboration et al., 2013, A&A, 558, A33

Astropy Collaboration et al., 2018, AJ, 156, 123

Athanassoula E., 2005, MNRAS, 358, 1477

Babusiaux C., et al., 2010, A&A, 519, A77

Babusiaux C., et al., 2022, arXiv e-prints, p. arXiv:2206.05989

Bacon R., et al., 2001, MNRAS, 326, 23

Bacon R., et al., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III. p. 773508 (arXiv:2211.16795), doi:10.1117/12.856027

Bacon R., et al., 2014, The Messenger, 157, 13

Bacon R., et al., 2017, A&A, 608, A1

Bastian N., Lardo C., 2018, ARA&A, 56, 83

Belokurov V., et al., 2018, MNRAS, 478, 611

Bensby T., Feltzing S., Lundström I., 2003, A&A, 410, 527

Bensby T., et al., 2013, A&A, 549, A147

Bensby T., Feltzing S., Oey M. S., 2014, A&A, 562, A71

Bensby T., et al., 2017, A&A, 605, A89

Bensby T., et al., 2019, The Messenger, 175, 35

Bertelli G., et al., 1994, A&AS, 106, 275

Bialek S., et al., 2020, MNRAS, 498, 3817

Binney J., 2012, MNRAS, 426, 1328

Bittner A., et al., 2019, A&A, 628, A117

Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529

Bland-Hawthorn J., Krumholz M. R., Freeman K., 2010, ApJ, 713, 166

Bland-Hawthorn J., et al., 2011, Optics Express, 19, 2649

Bland-Hawthorn J., et al., 2019, MNRAS, 486, 1167

Boardman N., et al., 2020a, MNRAS, 491, 3672

Boardman N., et al., 2020b, MNRAS, 498, 4943

Boecker A., et al., 2020, MNRAS, 491, 823

Bonnarel F., et al., 2000, A&AS, 143, 33

Borisova E., et al., 2016, ApJ, 831, 39

Bovy J., 2016, ApJ, 817, 49

Bovy J., Rix H.-W., Hogg D. W., 2012a, ApJ, 751, 131

Bovy J., et al., 2012b, ApJ, 753, 148

Bovy J., et al., 2012c, ApJ, 755, 115

Bovy J., et al., 2016, ApJ, 823, 30

Bressan A., et al., 2012, MNRAS, 427, 127

Brough S., et al., 2011, MNRAS, 414, L80

Brunetti M., Chiappini C., Pfenniger D., 2011, A&A, 534, A75

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Bryant J. J., et al., 2011, MNRAS, 415, 2173

Bryant J. J., et al., 2015, MNRAS, 447, 2857

Bryant J. J., et al., 2020, in Evans C. J., Bryant J. J., Motohara K., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol.

11447, Ground-based and Airborne Instrumentation for Astronomy VIII. p. 1144715, doi:10.1117/12.2560309

Buck T., 2020, MNRAS, 491, 5435

Buder S., et al., 2018, MNRAS, 478, 4513

Buder S., et al., 2021, MNRAS, 506, 150

Buell J. F., 2013, MNRAS, 428, 2577

Bundy K., et al., 2015, ApJ, 798, 7

Burstein D., 1979, ApJ, 234, 829

Calzetti D., et al., 2000, ApJ, 533, 682

Cantat-Gaudin T., et al., 2018, A&A, 618, A93

Cappellari M., 2017, MNRAS, 466, 798

Cappellari M., Copin Y., 2003, MNRAS, 342, 345

Cappellari M., Emsellem E., 2004, PASP, 116, 138

Cappellari M., et al., 2011, MNRAS, 413, 813

Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245

Carini R., et al., 2020, AJ, 159, 152

Carnall A. C., 2017, arXiv e-prints, p. arXiv:1705.05165

Carretta E., et al., 2010, A&A, 516, A55

Chaboyer B., Demarque P., Sarajedini A., 1996, ApJ, 459, 558

Chabrier G., 2003, PASP, 115, 763

Chambers K. C., et al. 2017, VizieR Online Data Catalog, p. II/349

Chen B., et al., 2022, arXiv e-prints, p. arXiv:2204.11413

Chiappini C., Matteucci F., Gratton R., 1997, ApJ, 477, 765

Chiappini C., Matteucci F., Romano D., 2001, ApJ, 554, 1044

Chiappini C., et al., 2015, A&A, 576, L12

Chiappini C., et al., 2019, The Messenger, 175, 30

Christlieb N., et al., 2019, The Messenger, 175, 26

Cid Fernandes R., et al., 2005, MNRAS, 358, 363

Clarke A. J., et al., 2019, MNRAS, 484, 3476

Clarkson W., et al., 2008, ApJ, 684, 1110

Colless M., et al., 2001, MNRAS, 328, 1039

Combes F., Sanders R. H., 1981, A&A, 96, 164

Comerón S., Salo H., Knapen J. H., 2018, A&A, 610, A5

Conroy C., 2013, ARA&A, 51, 393

Conroy C., et al., 2019, ApJ, 887, 237

Content R., 1997, in Ardeberg A. L., ed., Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 2871, Optical Telescopes of Today and Tomorrow. pp 1295–1305, doi:10.1117/12.269020

Cooper A. P., et al., 2010, MNRAS, 406, 744

Courtes G., 1982, in Humphries C. M., ed., Astrophysics and Space Science Library Vol. 92, IAU Colloq. 67: Instrumentation for Astronomy with Large Optical Telescopes. p. 123, doi:10.1007/978-94-009-7787-7_16

Creevey O. L., et al., 2022, arXiv e-prints, p. arXiv:2206.05864

Croom S. M., et al., 2012, MNRAS, 421, 872

Croom S. M., et al., 2021, MNRAS, 505, 991

Daniel K. J., Wyse R. F. G., 2015, MNRAS, 447, 3576

De Angeli F., et al., 2022, arXiv e-prints, p. arXiv:2206.06143

De Silva G. M., et al., 2006, AJ, 131, 455

De Silva G. M., et al., 2007, AJ, 133, 1161

De Silva G. M., et al., 2015, MNRAS, 449, 2604

Deason A. J., et al., 2018, ApJ, 862, L1

Dékány I., et al., 2013, ApJ, 776, L19

Dekker H., et al., 2000, in Iye M., Moorwood A. F., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4008, Optical and IR Telescope Instrumentation and Detectors. pp 534–545, doi:10.1117/12.395512

Deng L.-C., et al., 2012, Research in Astronomy and Astrophysics, 12, 735

Di Matteo P., et al., 2013, A&A, 553, A102

Di Matteo P., et al., 2014, A&A, 567, A122

Edvardsson B., et al., 1993, A&A, 275, 101

Eggen O. J., 1974, PASP, 86, 162

Eisenhauer F., Raab W., 2015, ARA&A, 53, 155

Eldridge J. J., et al., 2017, PASA, 34, e058

Elmegreen B. G., 1999, ApJ, 517, 103

Elmegreen B. G., Shadmehri M., 2003, MNRAS, 338, 817

Emsellem E., et al., 2004, MNRAS, 352, 721

Emsellem E., et al., 2011, MNRAS, 414, 888

Emsellem E., et al., 2022, A&A, 659, A191

Erkal D., et al., 2019, MNRAS, 487, 2685

Erroz-Ferrer S., et al., 2019, MNRAS, 484, 5009

Evans C. J., 2008, in Revista Mexicana de Astronomia y Astrofisica Conference Series. pp 23–25 (arXiv:0704.1119), doi:10.48550/arXiv.0704.1119

Evans T. A., et al., 2020, MNRAS, 497, 4311

Fabbro S., et al., 2018, MNRAS, 475, 2978

Faber S. M., et al., 1985, ApJS, 57, 711

Falcón-Barroso J., Martig M., 2021, A&A, 646, A31

Feltzing S., Holmberg J., Hurley J. R., 2001, A&A, 377, 911

Feng Y., Krumholz M. R., 2014, Nature, 513, 523

Foreman-Mackey D., et al., 2013, PASP, 125, 306

Foster C., et al., 2021, PASA, 38, e031

Fox A. J., et al., 2014, ApJ, 787, 147

Fraser-McKelvie A., Merrifield M., Aragón-Salamanca A., 2019, MNRAS, 489, 5030

Freeman K., Bland-Hawthorn J., 2002, ARA&A, 40, 487

Freudling W., et al., 2013, A&A, 559, A96

Gadotti D. A., et al., 2019, MNRAS, 482, 506

Gaia Collaboration et al., 2016a, A&A, 595, A1

Gaia Collaboration et al., 2016b, A&A, 595, A2

Gaia Collaboration et al., 2018, A&A, 616, A1

Gaia Collaboration et al., 2021, A&A, 649, A1

Gaia Collaboration et al., 2022a, arXiv e-prints, p. arXiv:2206.05534

Gaia Collaboration et al., 2022b, arXiv e-prints, p. arXiv:2208.00211

Gaudi B. S., 2022, in Bulletin of the American Astronomical Society. p. 102.146

Ge J., et al., 2018, MNRAS, 478, 2633

Georgiev I. Y., et al., 2019, MNRAS, 484, 3356

Gerhard O. E., 1993, MNRAS, 265, 213

Gesicki K., et al., 2014, A&A, 566, A48

Gibbons S. L. J., Belokurov V., Evans N. W., 2014, MNRAS, 445, 3788

Giesers B., et al., 2018, MNRAS, 475, L15

Gilmore G., Reid N., 1983, MNRAS, 202, 1025

Gilmore G., et al., 2012, The Messenger, 147, 25

Gilmozzi R., Spyromilio J., 2007, The Messenger, 127, 11

Girardi L., et al., 2000, A&AS, 141, 371

Girardi L., et al., 2005, A&A, 436, 895

Gonzalez O. A., et al., 2011, A&A, 534, A3

Göttgens F., et al., 2019, A&A, 631, A118

Grand R. J. J., et al., 2016, MNRAS, 459, 199

Gratton R., 2020, in Bragaglia A., et al., eds, Vol. 351, Star Clusters: From the Milky Way to the Early Universe. pp 241–250 (arXiv:1908.06905), doi:10.1017/S1743921319007877

Green G. M., et al., 2019, ApJ, 887, 93

Greggio L., et al., 2012, PASP, 124, 653

Grieco V., et al., 2012, A&A, 548, A60

Guérou A., et al., 2016, A&A, 591, A143

Haffert S. Y., et al., 2019, Nature Astronomy, 3, 749

Harris C. R., et al., 2020, Nature, 585, 357

Hasselquist S., et al., 2019, ApJ, 872, 58

Hasselquist S., et al., 2020, ApJ, 901, 109

Hayden M. R., et al., 2015, ApJ, 808, 132

Hayden M. R., et al., 2022, MNRAS, 517, 5325

Haywood M., et al., 2013, A&A, 560, A109

Haywood M., et al., 2018, A&A, 618, A78

Haywood M., et al., 2019, A&A, 625, A105

Hekker S., Johnson J. A., 2019, MNRAS, 487, 4343

Helmi A., 2020, ARA&A, 58, 205

Helmi A., et al., 1999, Nature, 402, 53

Helmi A., White S. D., Springel V., 2002, Phys. Rev. D, 66, 063502

Helmi A., et al., 2017, A&A, 598, A58

Helmi A., et al., 2018, Nature, 563, 85

Helmi A., et al., 2019, The Messenger, 175, 23

Ho A. Y. Q., et al., 2017, ApJ, 836, 5

Hobbs D., et al., 2016, arXiv e-prints, p. arXiv:1609.07325

Høg E., et al., 2000, A&A, 355, L27

Holtzman J. A., et al., 2015, AJ, 150, 148

Hook I., 2009, in Science with the VLT in the ELT Era. p. 225, doi:10.1007/978-1-4020-9190-2_38

Hopkins P. F., et al., 2018, MNRAS, 480, 800

Horta D., et al., 2023, MNRAS, 520, 5671

Hunter J. D., 2007, Computing in Science and Engineering, 9, 90

Husser T.-O., et al., 2016, A&A, 588, A148

Husser T.-O., et al., 2020, A&A, 635, A114

Irwin P. G. J., et al., 2018, , 302, 426

Jin S., et al., 2023, MNRAS,

Jones D. H., et al., 2004, MNRAS, 355, 747

Jurić M., et al., 2008, ApJ, 673, 864

Kacharov N., et al., 2018, MNRAS, 480, 1973

Kafle P. R., et al., 2012, ApJ, 761, 98

Kamann S., Wisotzki L., Roth M. M., 2013, A&A, 549, A71

Kamann S., et al., 2016, A&A, 588, A149

Kamann S., et al., 2018, MNRAS, 473, 5591

Keller S. C., et al., 2007, PASA, 24, 1

Kobayashi C., Leung S.-C., Nomoto K., 2020a, ApJ, 895, 138

Kobayashi C., Karakas A. I., Lugaro M., 2020b, ApJ, 900, 179

Koleva M., et al., 2009, A&A, 501, 1269

Koppelman H., Helmi A., Veljanoski J., 2018, ApJ, 860, L11

Koppelman H. H., et al., 2019, A&A, 631, L9

Kormendy J., Bender R., 2019, ApJ, 872, 106

Krajnović D., et al., 2008, MNRAS, 390, 93

Kregel M., van der Kruit P. C., 2005, MNRAS, 358, 481

Kroupa P., 2001, MNRAS, 322, 231

Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019, ARA&A, 57, 227

Kubryk M., Prantzos N., Athanassoula E., 2015, A&A, 580, A126

Kunder A., et al., 2012, AJ, 143, 57

Kuntschner H., et al., 2006, MNRAS, 369, 497

Kuntschner H., et al., 2010, MNRAS, 408, 97

Kurucz R. L., 1970, SAO Special Report, 309

Kurucz R., 1993, SYNTHE Spectrum Synthesis Programs and Line Data. Kurucz CD-ROM No. 18. Cambridge, 18

Kurucz R. L., 2005, Memorie della Societa Astronomica Italiana Supplementi, 8, 14

Lada C. J., Lada E. A., 2003, ARA&A, 41, 57

Latour M., et al., 2019, A&A, 631, A14

Latour M., et al., 2021, A&A, 653, L8

Le Borgne D., et al., 2004, A&A, 425, 881

Leitherer C., et al., 1999, ApJS, 123, 3

Leung H. W., Bovy J., 2019, MNRAS, 483, 3255

Levan A. J., et al., 2017, ApJ, 848, L28

Li J., et al., 2016, Research in Astronomy and Astrophysics, 16, 110

Lian J., et al., 2020a, MNRAS, 497, 2371

Lian J., et al., 2020b, MNRAS, 497, 3557

Licquia T. C., Newman J. A., 2015, ApJ, 806, 96

Licquia T. C., Newman J. A., 2016, ApJ, 831, 71

Licquia T. C., Newman J. A., Brinchmann J., 2015, ApJ, 809, 96

Licquia T. C., Newman J. A., Bershady M. A., 2016, ApJ, 833, 220

Liu X. W., et al., 2014, in Feltzing S., et al., eds, Vol. 298, Setting the scene for Gaia and LAMOST. pp 310–321 (`arXiv:1306.5376`), doi:10.1017/S1743921313006510

Liu X.-W., Zhao G., Hou J.-L., 2015, Research in Astronomy and Astrophysics, 15, 1089

Liu F., et al., 2016a, MNRAS, 457, 3934

Liu F., et al., 2016b, MNRAS, 463, 696

Liu F., et al., 2019, A&A, 627, A117

Liu C., et al., 2020, arXiv e-prints, p. arXiv:2005.07210

Loebman S. R., et al., 2016, ApJ, 818, L6

Luo A. L., et al., 2015, Research in Astronomy and Astrophysics, 15, 1095

Mackereth J. T., et al., 2017, MNRAS, 471, 3057

Mackereth J. T., et al., 2018, MNRAS, 477, 5072

Majewski S. R., et al., 2017, AJ, 154, 94

Malhan K., Ibata R. A., 2019, MNRAS, 486, 2995

Mao Y.-Y., et al., 2021, ApJ, 907, 85

Marigo P., et al., 2017, ApJ, 835, 77

Martig M., et al., 2021, MNRAS, 508, 2458

Martín-Navarro I., et al., 2018, MNRAS, 475, 3700

Martínez-Lombilla C., Knapen J. H., 2019, A&A, 629, A12

Massari D., Koppelman H. H., Helmi A., 2019, A&A, 630, L4

Masseron T., Gilmore G., 2015, MNRAS, 453, 1855

Massey R., et al., 2015, MNRAS, 449, 3393

Masters K. L., et al., 2019, MNRAS, 487, 1808

McDermid R. M., et al., 2015, MNRAS, 448, 3484

McDermid R. M., et al., 2020, arXiv e-prints, p. arXiv:2009.09242

McLeod A. F., et al., 2015, MNRAS, 450, 1057

McLeod A. F., et al., 2019, MNRAS, 486, 5263

McWilliam A., Zoccali M., 2010, ApJ, 724, 1491

Mészáros S., et al., 2020, MNRAS, 492, 1641

Milone A. P., Marino A. F., 2022, Universe, 8, 359

Milone A. P., et al., 2018, MNRAS, 481, 5098

Minchev I., Chiappini C., Martig M., 2013, A&A, 558, A9

Minchev I., et al., 2014, ApJ, 781, L20

Minchev I., et al., 2015, ApJ, 804, L9

Minniti D., Zoccali M., 2008, in Bureau M., Athanassoula E., Barbuy B., eds, Vol. 245, Formation and Evolution of Galaxy Bulges. pp 323–332 (arXiv:0710.3104), doi:10.1017/S1743921308018048

Minniti D., Lucas P., VVV Team 2017, VizieR Online Data Catalog, p. II/348

Monari G., et al., 2018, A&A, 616, L9

Morrison H. L., 1993, AJ, 106, 578

Mutch S. J., Croton D. J., Poole G. B., 2011, ApJ, 736, 84

Myeong G. C., et al., 2019, MNRAS, 488, 1235

Nataf D. M., 2016, PASA, 33, e023

Nataf D. M., et al., 2010, ApJ, 721, L28

Nataf D. M., et al., 2016, MNRAS, 456, 2692

Ness M., et al., 2012, ApJ, 756, 22

Ness M., et al., 2013a, MNRAS, 430, 836

Ness M., et al., 2013b, MNRAS, 432, 2092

Ness M., et al., 2015, ApJ, 808, 16

Ness M., et al., 2018, ApJ, 853, 198

Ness M. K., et al., 2019, ApJ, 883, 177

Nidever D. L., et al., 2014, ApJ, 796, 38

Nissen P. E., 2015, A&A, 579, A52

Nordström B., et al., 2004, A&A, 418, 989

Norman C. A., Sellwood J. A., Hasan H., 1996, ApJ, 462, 114

O'Briain T., et al., 2021, ApJ, 906, 130

Ocvirk P., et al., 2006, MNRAS, 365, 46

Ortolani S., et al., 1995, Nature, 377, 701

Pasquini L., et al., 2002, The Messenger, 110, 1

Perryman M. A. C., et al., 2001, A&A, 369, 339

Pietrinferni A., et al., 2004, ApJ, 612, 168

Pietrinferni A., et al., 2006, ApJ, 642, 797

Pietrukowicz P., et al., 2015, ApJ, 811, 113

Pinna F., et al., 2019a, A&A, 623, A19

Pinna F., et al., 2019b, A&A, 625, A95

Piqueras L., et al., 2017, arXiv e-prints, p. arXiv:1710.03554

Plez B., 2012, Turbospectrum: Code for spectral synthesis, Astrophysics Source Code Library, record ascl:1205.004 (ascl:1205.004)

Poggio E., et al., 2017, A&A, 601, A115

Poggio E., et al., 2021, A&A, 651, A104

Prichard L. J., Vaughan S. P., Davies R. L., 2019, MNRAS, 488, 1679

Prugniel P., Soubiran C., 2001, A&A, 369, 1048

Prugniel P., et al., 2007, arXiv e-prints, pp astro–ph/0703658

Putman M. E., et al., 1998, Nature, 394, 752

Quillen A. C., et al., 2009, MNRAS, 397, 1599

Raha N., et al., 1991, Nature, 352, 411

Recio-Blanco A., et al., 2022, arXiv e-prints, p. arXiv:2206.05541

Richard J., et al., 2019, arXiv e-prints, p. arXiv:1906.01657

Rix H.-W., et al., 2022, ApJ, 941, 45

Robotham A. S. G., et al., 2012, MNRAS, 424, 1448

Rodríguez-Puebla A., et al., 2017, MNRAS, 470, 651

Rojas-Arriagada A., et al., 2014, A&A, 569, A103

Rojas-Arriagada A., et al., 2017, A&A, 601, A140

Rojas-Arriagada A., et al., 2019, A&A, 626, A16

Rojas-Arriagada A., et al., 2020, MNRAS, 499, 1037

Rossi E. M., et al., 2017, MNRAS, 467, 1844

Roth M. M., et al., 2018, A&A, 618, A3

Roth M. M., Weilbacher P. M., Castro N., 2019, Astronomische Nachrichten, 340, 989

Rupke D., 2018, Galaxies, 6, 138

Salaris M., Weiss A., Percival S. M., 2004, A&A, 414, 163

Salpeter E. E., 1955, ApJ, 121, 161

Sánchez S. F., et al., 2012, A&A, 538, A8

Sánchez S. F., et al., 2016, A&A, 594, A36

Sanders J. L., Binney J., 2015, MNRAS, 449, 3479

Sandford N. R., Weisz D. R., Ting Y.-S., 2020, ApJS, 249, 24

Santiago B. X., et al., 2016, A&A, 585, A42

Sarzi M., Iodice E., Fornax3D Collaboration 2022, The Messenger, 189, 9

Schaye J., et al., 2015, MNRAS, 446, 521

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

Schmidt M., 1963, ApJ, 137, 758

Schönrich R., Binney J., 2009a, MNRAS, 396, 203

Schönrich R., Binney J., 2009b, MNRAS, 399, 1145

Schultheis M., et al., 2017, A&A, 600, A14

Schultheis M., et al., 2021, A&A, 650, A191

Scott N., et al., 2021, ApJ, 913, L11

Sellwood J. A., Binney J. J., 2002, MNRAS, 336, 785

Serra P., Trager S. C., 2007, MNRAS, 374, 769

Sharma S., 2017, ARA&A, 55, 213

Sharma S., et al., 2011, ApJ, 730, 3

Sharma S., et al., 2018, MNRAS, 473, 2004

Sharma S., et al., 2021a, MNRAS, 506, 1761

Sharma S., Hayden M. R., Bland-Hawthorn J., 2021b, MNRAS, 507, 5882

Sharma S., et al., 2022, MNRAS, 510, 734

Sharp R. G., Bland-Hawthorn J., 2010, ApJ, 711, 818

Shu F. H., 1969, ApJ, 158, 505

Siegel M. H., et al., 2002, ApJ, 578, 151

Silva Aguirre V., et al., 2018, MNRAS, 475, 5487

Skrutskie M. F., et al., 2006, AJ, 131, 1163

Sneden C., et al., 2012, MOOG: LTE line analysis and spectrum synthesis, Astrophysics Source Code Library, record ascl:1202.009 (ascl:1202.009)

Sohn S. T., et al., 2018, ApJ, 862, 52

Soto M., Rich R. M., Kuijken K., 2007, ApJ, 665, L31

Soto K. T., et al., 2016, MNRAS, 458, 3210

Spitoni E., et al., 2019, A&A, 623, A60

Stanek K. Z., 1996, ApJ, 460, L37

Steinmetz M., et al., 2006, AJ, 132, 1645

Steinmetz M., et al., 2020a, AJ, 160, 82

Steinmetz M., et al., 2020b, AJ, 160, 83

Szymański M. K., et al., 2011, Acta Astron., 61, 83

Tanvir N. R., et al., 2017, ApJ, 848, L27

Ting Y.-S., Weinberg D. H., 2022, ApJ, 927, 209

Ting Y.-S., et al., 2017a, ApJ, 843, 32

Ting Y.-S., et al., 2017b, ApJ, 849, L9

Ting Y.-S., et al., 2018, ApJ, 860, 159

Ting Y.-S., et al., 2019, ApJ, 879, 69

Tissera P. B., et al., 2013, MNRAS, 432, 3391

Tojeiro R., et al., 2007, MNRAS, 381, 1252

Trager S. C., Somerville R. S., 2009, MNRAS, 395, 608

Tumlinson J., Peeples M. S., Werk J. K., 2017, ARA&A, 55, 389

Valenti J. A., Piskunov N., 1996, A&AS, 118, 595

Valenti E., et al., 2018, A&A, 616, A83

Vanderriest C., 1980, PASP, 92, 858

Vazdekis A., 1999, ApJ, 513, 224

Vazdekis A., et al., 1996, ApJS, 106, 307

Vazdekis A., et al., 2010, MNRAS, 404, 1639

Vazdekis A., et al., 2015, MNRAS, 449, 1177

Verro K., et al., 2022, A&A, 661, A50

Virtanen P., et al., 2020, Nature Methods, 17, 261

Wang S., Chen X., 2019, ApJ, 877, 116

Wang J., et al., 2011, MNRAS, 413, 1373

Wang Z., et al., 2022, MNRAS, 514, 1034

Watkins L. L., Evans N. W., An J. H., 2010, MNRAS, 406, 264

Watkins L. L., et al., 2019, ApJ, 873, 118

Wegg C., Gerhard O., 2013, MNRAS, 435, 1874

Weilbacher P. M., et al., 2012, in Radziwill N. M., Chiozzi G., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8451, Software and Cyberinfrastructure for Astronomy II. p. 84510B, doi:10.1117/12.925114

Weilbacher P. M., et al., 2014, in Manset N., Forshay P., eds, Astronomical Society of the Pacific Conference Series Vol. 485, Astronomical Data Analysis Software and Systems XXIII. p. 451 (arXiv:1507.00034), doi:10.48550/arXiv.1507.00034

Weilbacher P. M., et al., 2015, A&A, 582, A114

Weilbacher P. M., Streicher O., Palsa R., 2016, MUSE-DRP: MUSE Data Reduction Pipeline, Astrophysics Source Code Library, record ascl:1610.004 (ascl:1610.004)

Weilbacher P. M., et al., 2020, A&A, 641, A28

Wheeler A., et al., 2020, ApJ, 898, 58

Wilkinson D. M., et al., 2017, MNRAS, 472, 4297

Wu Y., et al., 2011, Research in Astronomy and Astrophysics, 11, 924

Wylie S. M., et al., 2021, A&A, 653, A143

Wyse R. F. G., Gilmore G., 2006, in Whitelock P., Dennefeld M., Leibundgut B., eds, Vol. 232, The Scientific Requirements for Extremely Large Telescopes. pp 140–148 (arXiv:astro-ph/0604130), doi:10.1017/S1743921306000391

Xiang M., Rix H.-W., 2022, Nature, 603, 599

Xiang M. S., et al., 2015, MNRAS, 448, 822

Xiang M. S., et al., 2017, MNRAS, 464, 3657

Xiang M., et al., 2019, ApJS, 245, 34

Xiang M.-S., et al., 2020, ApJ, 898, 28

Xue X. X., et al., 2008, ApJ, 684, 1143

Yan R., et al., 2016, AJ, 152, 197

Yniguez B., et al., 2014, MNRAS, 439, 73

Yoachim P., Dalcanton J. J., 2006, AJ, 131, 226

Yong D., et al., 2016, MNRAS, 459, 487

Zhang M., et al., 2021, ApJ, 922, 145

Zhang X., Green G. M., Rix H.-W., 2023, arXiv e-prints, p. arXiv:2303.03420

Zhao G., et al., 2012, Research in Astronomy and Astrophysics, 12, 723

Zhong J., et al., 2020, A&A, 640, A127

Zhou S., Merrifield M., Aragón-Salamanca A., 2022, MNRAS, 513, 5446

Zhou S., et al., 2023, MNRAS, 521, 5810

Zhu L., et al., 2022, A&A, 664, A115

Zoccali M., et al., 2003, A&A, 399, 931

Zoccali M., et al., 2014, A&A, 562, A66

Zoccali M., et al., 2017, A&A, 599, A12

Zolotov A., et al., 2009, ApJ, 702, 1058

de Jong R. S., et al., 2019, The Messenger, 175, 3

de Silva G. M., et al., 2011, MNRAS, 415, 563

van Dishoeck E. F., Blake G. A., 1998, ARA&A, 36, 317

van de Sande J., et al., 2017, ApJ, 835, 104

van de Sande J., et al., 2023, arXiv e-prints, p. arXiv:2306.00059

van den Bergh S., 1962, AJ, 67, 486

van der Marel R. P., Franx M., 1993, ApJ, 407, 525

van der Wel A., et al., 2014, ApJ, 788, 28