Title

Probing the Extremes of Galaxy Evolution With New Stellar Population Synthesis Models

Author

Villaume, Alexa M

Publication Date

2020

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

PROBING THE EXTREMES OF GALAXY EVOLUTION WITH NEW STELLAR POPULATION SYNTHESIS MODELS

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

Doctor of Philosophy

in

ASTRONOMY AND ASTROPHYSICS

by

Alexa A. Villaume

June 2020

The Dissertation of Alexa A. Villaume is approved:

Professor Jean Brodie, Chair

Professor Charlie Conroy

Professor Aaron Romanowsky

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Probing the Extremes of Galaxy Evolution with New Stellar Population Synthesis Models

by

Alexa A. Villaume

Understanding how galaxies form and evolve is a fundamental goal of modern astrophysics. It touches on critical problems at both small scales, e.g., star-formation, and large scales, e.g., the hierarchical growth of structure as expected from the cosmological paradigm. Ironically, the very nature which makes galaxy evolution so important to understand is the same property that makes it so difficult to understand. The vast range in physical scales that need to be understood and non-linear evolutionary behavior make *ab initio* models describing galaxy evolution impossible. In this thesis, I adopt the "archaeological" approach to unraveling the evolutionary histories of galaxies that exist at the extreme ends of the observed galaxy population.

I first describe the empirical stellar spectral library that forms the foundation of the stellar population synthesis models that underpin this work. The expansion in stellar parameter and wavelength coverage of this library enables the broad applicability of the models to objects that collectively populate various ends of the range of galaxy parameter space.

The first such application demonstrates that the primary correlate of initial mass function variability is not metallicity, like suggested by previous work. Then I show that the internal properties of the ancient star clusters that are used as primary fossil record to infer the histories of the most massive galaxies in the Universe are not as well-understood as previously considered. I discuss how this result both alters the interpretations of galaxy evolution based on these fossils and the implications this has on the understanding of star cluster formation and, thus, star formation. I then, with the new models for the individual clusters and a novel statistical framework for the systems of these clusters, provide new constraints on the assembly processes in a massive galaxy. Finally, I apply the new models to the spatially-resolved galaxy light of a contentious new galaxy type to show that it is indeed a galaxy with hitherto unprecedented stellar population properties.

Acknowledgments

In more or less order of appearance with respect to this thesis, I would like to thank: Ricardo for introducing me to Santa Cruz; Jean and Charlie for encouraging me to apply to graduate school and providing the opportunities that made this work possible; Aaron whose fastidiousness could always be relied on; Ashela and Eric M. who were the best hosts whenever I flew in and out of SFO; Anna R. who has always been a champion and a mentor; Ben J. whose technical expertise is an inspiration and a help; Emily, Chris, Phil, and Zach for being great friends throughout; Jieun for always being a pal when I came to Cambridge; Diana, Jaime, and Ari for the time spent working on homework together; Eric G., Tiffany, and Kat for helping make Calvin Place fun; Nacho who I had many great discussions with; the Astronomy on Tap crew who gave me something to look forward to almost every month for the last two years; to Joe and Amanda for being great friends and support; to the Astronomy & Astrophysics department on the whole for providing a unique and stimulating environment but with particular thanks to Sandy who can always be relied on for insightful discussions and tough feedback.

Most importantly, I want to thank my husband, Ben. I cannot think of a better person I would want to be under shelter-in-place orders for months with as I finish my thesis. His intelligence, kindness, and generosity are an inspiration.

Published Material

The text of this dissertation includes reprints of the following published material led by Villaume, with the permission of the listed coauthors:

Chapter 2 was published in the literature as Villaume et al. (2017b). I was responsible for the observations, data reduction, post-processing of the spectra, calibrating the SPI coefficients, quality tests on SPI, performing the analysis, writing the text, and creating the figures. I wish to thank my coauthors Charlie Conroy, Benjamin Johnson, John Rayner, Andrew W. Mann, and Pieter van Dokkum for their generous support and feedback in the course of this research, with particular thanks to Benjamin who wrote SPI. I would also like to thank M. Cushing for helpful discussion and P. Sánchez-Blázquez for generously sharing data.

Chapter 3 was published in the literature as Villaume et al. (2017a). I was responsible for running the models, performing the analysis, writing the text, and creating the figures. I wish to thank my coauthors Jean Brodie, Charlie Conroy, Aaron J. Romanowsky, and Pieter van Dokkum for their generous upport and feedback in the course of this research.

Chapter 4 was published in the literature as Villaume et al. (2019). I was responsible for running the models, performing the analysis, writing the text, and creating the figures. I wish to thank my coauthors Aaron J. Romanowsky, Jean Brodie, and Jay Strader for their generous support and feedback in the course of this research.

Chapter 5 is currently under review at the Astrophysical Journal. I was responsible for running the SPS models, testing and calibrating the HBM, performing the analysis, writing the text, and creating the figures. I wish to thank my coauthors Dan Foreman-Mackey, Aaron Romanowsky, Jean Brodie, and Jay Strader for their generous support and feedback in the course of this research, with particular thanks to Dan who taught me the ways of HBM and without whom this project would not exist. I also thank the Center for Computational Astrophysics at the Flatiron Institute for enabling this collaboration and Karl Gebhardt for generously sharing the Mitchell data.

Scientific Acknowledgments

I gratefully thank the National Science Foundation Graduate Research Fellowship and UCSC's Regent and Dissertation Year Fellowships. I want to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. I am most fortunate to have had the opportunity to conduct observations from this mountain.

For Marylou

Chapter 1

Introduction

> *Stars. Or rather, the drains of heaven–waiting. Little holes. Little centuries opening just long enough for us to slip through.*
>
> **Ocean Vuong, Immigrant Haibun**

Almost 100 years since Hubble (1936) – and innumerable observational surveys that have gone deeper, farther, broader – the general classification scheme established then between early-type galaxies (ETGs), late-type galaxies (LTGs), and irregulars is still used. That is not to say significant progress has not been made, but the question of how galaxies initially form and then evolve to what is observed today is still a fundamental question of modern astrophysics.

Galaxy evolution is related to both questions in star-formation and cosmology. In the modern ΛCDM[1] cosmological paradigm (Blumenthal et al. 1984), galaxies form within dark matter halos in the nodes of the filaments of the "cosmic web". Galaxies then provide the environments in which star-formation takes place, but are then im-

---

[1] "Cold dark matter" indicates that "dark matter" is expected to be a weakly interacting particle with low thermal velocity. "Λ" refers to the expansion of the Universe.

pacted by that process in the form of “stellar feedback”. Furthermore, galaxies within relatively massive dark matter halos experience bombardments of galaxies in lower mass dark matter halos that bring in some combination of stars and other detritus which further assemble the massive galaxy (Naab et al. 2009).

In this sense, our understanding of galaxy evolution is the keystone to the most fundamental aspects of astrophysics. It is this very nature, however, that makes galaxy evolution so difficult to understand. The processes that affect the evolution of galaxies occur over a huge range of physical scales which makes it currently impossible to have an *ab inito* formulation of galaxy evolution (Somerville & Davé 2015; Naab & Ostriker 2017).

1.1 The archaeological approach to galaxies

The present-day stellar population properties are used to provide one of the only windows into the evolutionary histories of galaxies – particularly at early epochs that are not directly observable. The stellar populations of our own Milky Way galaxy and in nearby, Local Group, galaxies are relatively easily studied and understood because we can resolve and directly observe the individual stars.

However, in more distant galaxies the stars are unresolved and stellar population synthesis (SPS) models are relied on to interpret the integrated light observations. SPS models work by creating a synthetic spectrum from the summed spectra of all the stars expected to be found in a galaxy, weighted according to the expected numbers of each type of star. This synthetic spectrum is compared to the observations with the weights and stellar types adjusted for a best-fit model. This is a standard, powerful tool for understanding galaxies with unresolved stellar populations.

SPS models are limited in their applicability by the ingredients that go into making them[2]. Most fundamentally are the stellar libraries that are used to convert stages of stellar evolution – parameterized by surface gravity ($\log g$), effective temperature (T_{eff}), and metallicity ([Fe/H]) – into spectral energy distributions (SEDs) that can be compared to observations.

One of the most intriguing results to come out of SPS models recently is the measurement of variable initial mass functions (IMFs) in the centers of massive ETGs. This remains a controversial result as it does not correspond to the extreme star-formation environments found in the Milky and because there is no theoretical underpinning to explain it (Krumholz 2014).

Part of the problem is obtaining a clearer observational picture of this phenomenon is that the types of objects that could be assessed in this way has been very limited. This is because this kind of measurement require gravity-sensitive spectral features that are mainly available in the near infrared (IR) and near-IR spectral stellar libraries available at the time only existed for solar metallicities.

Chapter 2 details my efforts to create an empirical, optical to near infrared (IR) stellar spectra library with expanded stellar parameter coverage. Then, Chapter 3, presents the results from my modeling of optical to near-IR spectroscopy of low-velocity, compact stellar systems to expand the parameter space of IMF studies.

---

[2] This problem has existed since the conception of computers. As reflected by Charles Babbage, "*On two occasions I have been asked, 'Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?'...I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.*"

1.2 Globular clusters as fossil records

The access we have to resolved stellar populations in the Galaxy is extremely powerful. The key step to understanding the initial formation, and subsequent assembly, of the galaxy is to differentiate the *in*– and *ex-situ* populations. In recent years, a renaissance of sorts has transformed our understanding of the origins of the Milky Way through a combination of heroic data collection efforts (e.g. the Gaia and APOGEE surveys Gaia Collaboration et al. 2018; Majewski et al. 2017a) and increasingly sophisticated simulations of stellar halos in a cosmological context (e.g., Latte/FIRE2, Auriga Wetzel et al. 2016; Grand et al. 2017).

Even with an increasing ability to obtain spatially-resolved spectroscopy of extragalactic systems, integrated light will still never be as good as resolved stellar populations. But, instead of focusing on the stellar populations within a galaxy, discrete tracers can be used to probe the content of a galaxy. Globular clusters (GCs) are one of the most widely used of the possible tracers as they are nearly ubiquitous around galaxies and are relatively luminous compared to galaxy starlight.

But their strength lies in their uniformly old ($>$ 10 Gyr) ages. This makes them an ideal “fossil record” for the archaeological approach as they presumably reflect the conditions of the early-universe under which they were formed. In the Milky Way, it was the GC system that provided the pre-ΛCDM evidence that the stellar halo of the Milky Way was built through the accretion of satellite galaxies (Searle & Zinn 1978). Moreover, difference in accretion history between the Milky Way and M31 is clearly seen in the different metallicity distributions of their own GC systems (Caldwell &

Romanowsky 2016). GC systems also extend far beyond the reaches of galaxy starlight, and so provide a window to the fully *ex-situ* outer halos of galaxies.

However, massive star cluster formation is still an unsolved problem. GCs are additionally subject to a range of dynamical effects that destroy an unknown fraction of the initial population (Ma et al. 2020). And, while they are generally considered to be good approximations of simple stellar populations things like the inexplicable star-by-star abundances variations seen in Milky Way and Local Group GCs (Bastian & Lardo 2017) and unexpected dynamical mass-to-light ratios (M_d/L, Strader et al. 2011a) seem to challenge that view.

This is of particular concern in extragalactic systems, where the inherent data limitations have rooted the field in assumptions. Imaging has been the dominant means to study GC systems outside the Local Group because it is an efficient way to obtain large data sets and, because of their old ages, it is assumed that the broadband colors of individual GCs can easily be converted into metallicity through "color–metallicity relations" (CMRs). In Chapter 4, I show that this is not case because the internal properties of GCs are not as well-understood as previously thought but that spectroscopy and SPS models provides a means to understand and mitigate these difficulties.

1.3 Stellar population gradients, as opposed to integrated values

SPS models provide the means to infer ages (though that is an ill-posed parameter) and chemical abundances of galaxies beyond the Local Group. The different chemical elements can be broadly categorized as Fe-peak elements (Fe), α elements (Mg, Ti, Ca, Si), and light elements (C, N). Since the production sites of these elements differ, so too do the timescales which they are produced. The Fe-peak elements are produced in Type Ia (and Type II) supernova, while the α elements are produced in Type II supernova which originate from low mass and high mass stars, respectively. Therefore, $[\alpha/\mathrm{Fe}]$[3] is used as a clock of star-formation timescales.

Thomas et al. (2005) leveraged this fact over a large sample of ETGs and determined that the centers of massive ETGs are very metal-rich and α-enhanced, implying that galaxies produced a very short, intense period of star-formation before being quenched. On the surface, this is difficult to reconcile with the large sizes and extended stellar halos seen in these galaxies and was called "antiheiarchcial" or "downsizing" when the result first came out. In principle, though, this can be explained with late assembly of these galaxies through primarily "dry" mergers [4] which preserve the old appearance of the original galaxy (Di Matteo et al. 2009).

In light of this, there has been a push to move beyond integrated stellar population values, into the spatially-resolved regime to obtain galactocentric radial gradients of stellar population parameters. This field is only recently reaching maturity in the form of the MASSIVE (Ma et al. 2014) and MaNGa (Bundy et al. 2015) surveys. The primary difficulty has been data acquisition as the surface brightness of massive ETGs drops precipitously with increasing galactocentric radius. Both those surveys rely on integral field unit (IFU) spectrographs.

---

[3] Almost always in the extragalactic context α is really just Mg but that does not mean that they all have the same origins. To quote Stan Woosely, "*Ti is a bastard.*"

[4] Mergers that do not bring in gas and so do not form stars as a result.

The most massive galaxies in the present-day Universe are brightest cluster galaxies (BCGs), an extreme end of the ETG class ($M_{\rm halo} \sim 10^{14} M_{\odot}$). They should therefore be the best environments to test the predictions of hierarchical assembly. An outstanding puzzle in this regard, is that despite the violence that underpins the evolution of these galaxies from the high-redshift Universe, they have been shown to follow tight relations among their global properties. For example, it is not clear how the very little scatter in the Fundamental Plane, which places strong constraints on the age and metallicity distributions in ETGs, is reconcilable for the merging scenario (Renzini 2006).

In Chapter 5 I model the stellar population gradients of the integrated starlight in BCG M87 and its GC system. For the GC system, I model with the metallicity gradient within a hierarchical Bayesian statistical framework. This allows me to account for the covariance between the subpopulation memberships of the individual GCs and the subpopulation properties that I am measuring as well as statistically de-project the galactocentric distances. I also use the subpopulation memberships to my physically-motivated spectral stacks of the GC system to compare to the field star population.

More recently, a new class of objects was discovered, so-called ultra-diffuse galaxies (UDGs). Their large sizes but low stellar masses pose a significant challenge for galaxy formation and evolution theories. In Chapter 6, I detail the first spatially-resolved stellar population gradients of an UDG, DF44. Finally, in Chapter 7, I summarize this thesis and discuss the future work that needs to be done.

Chapter 2

The Extended IRTF Spectral Library: Expanded coverage in metallicity, temperature, and surface gravity

2.1 Introduction

Stellar libraries, whether empirical or theoretical, are foundational to several different fields of astrophysics. In stellar population synthesis (SPS) models stellar libraries are needed to convert the stellar evolution predictions of stellar parameters: effective temperature ($T_{\rm eff}$), surface gravity (log$g$), and metallicity ([Fe/H]), into spectral energy distributions (SEDs) (Conroy 2013). Stellar spectra can be used to compute the line-of-site velocity distributions in galaxies (Cappellari & Emsellem 2004; Emsellem et al. 2004). Well-characterized stellar spectra are also a key ingredient in exoplanet studies (e.g, Newton et al. 2014; Mann et al. 2015).

The goal for any stellar library is extensive coverage in stellar parameter space and large wavelength coverage at comparable or better resolution than typical observations to which they are compared. Large wavelength coverage is important because different wavelength regimes probe different stellar populations in the integrated light of galaxies. The optical is the most well-studied wavelength regime but there is important information that is only available in other regimes. For example, the ultra-violet (UV) probes the populations of hot, massive stars (e.g., Vazdekis et al. 2016, and the references therein) and the near infrared (IR) probes the populations of cool, evolved stars such as those on the asymptotic giant branch (Athey et al. 2002; Martini et al. 2013; Villaume et al. 2015; Simonian & Martini 2016). Furthermore, the endeavor to characterize biological signatures of exoplanets must consider the effect and characteristics of the host stars, which requires UV-IR coverage (France et al. 2015). Moreover, uniform coverage in parameter space is also necessary to ensure accurate stellar population models. Unfortunately, there is still no single stellar library that covers the entire range of wavelength and parameter space.

Empirical stellar libraries have been available for over 30 years (e.g., Gunn & Stryker 1983; Pickles 1985; Diaz et al. 1989; Silva & Cornell 1992; Pickles 1998; Worthey et al. 1994; Jones 1999; Lançon & Wood 2000; Cenarro et al. 2001; Prugniel & Soubiran 2001; Le Borgne et al. 2003; Valdes et al. 2004; Sánchez-Blázquez et al.

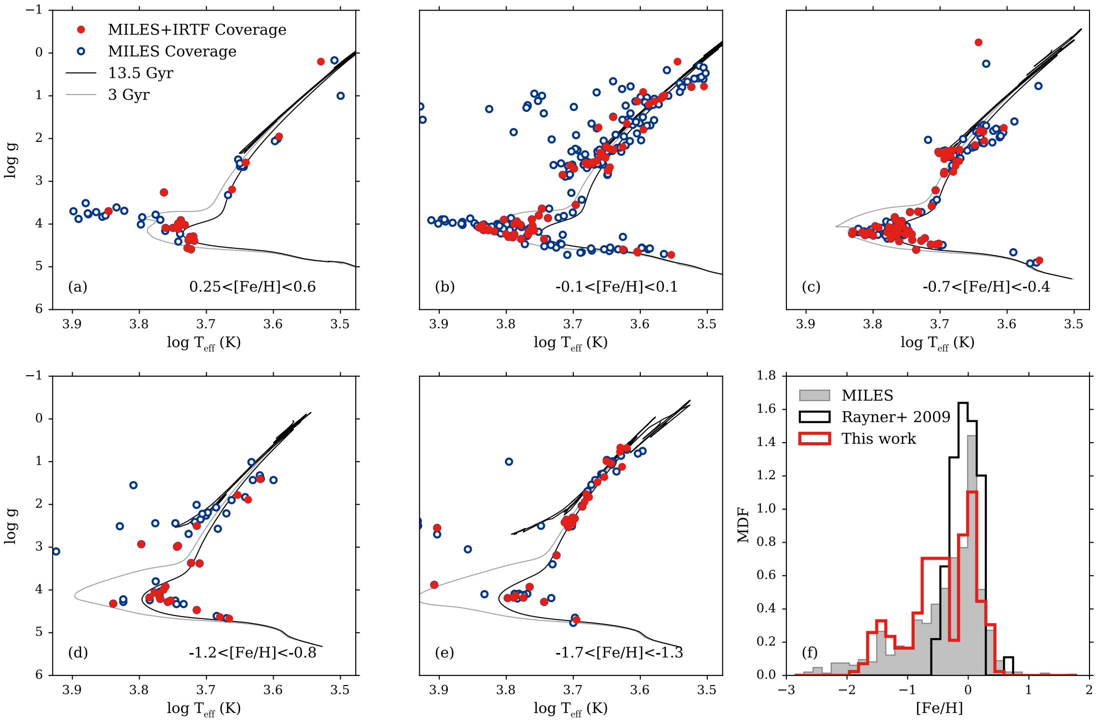

Figure 2.1: (Panels a-e) Comparison of the MILES spectral library (blue) with the library presented in this work (red). All the stars in this work were selected from the MILES library. Stars were chosen to sample isochrones spanning from 3 Gyr (grey line) to 13.5 Gyr (black line) with metallicities [Fe/H] = 0.25, 0.00, -0.50, -1.00, and -1.50. (f) Comparison of the metallicity density functions (MDFs) of the MILES spectral library (solid grey), the Rayner et al. (2009) (black line), and the library presented in this work (red line). The different libraries have different total number of stars so the MDFs were normalized such that the integral over the range of metallicity values is 1 for each library, indicating at any given metallicity value what the probability that a given star has that metallicity for each library.

2006a; Mármol-Queraltó et al. 2008; Rayner et al. 2009; Ardila et al. 2010; Sharon et al. 2010). A key limitation of existing stellar libraries is their often limited coverage in stellar parameter space. This is in some respects a fundamental problem because not all stars of interest are close enough to enable detailed observations (e.g., hot metal poor stars). Standard observational constraints such as atmospheric absorption and sky emission, which is especially prominent in the IR, flux calibration, wavelength coverage, and spectral resolution further challenge the development of comprehensive empirical stellar libraries.

Theoretical libraries offer the advantage of dense coverage in parameter space, arbitrarily high spectral resolution, and no need to correct for atmospheric absorption or flux calibration (e.g., Munari et al. 2005; Martins et al. 2005; Coelho et al. 2005). However, theoretical stellar libraries are only as good as the available atomic and molecular line lists and the approximations made in computing the models, e.g., the assumption of local thermodynamic equilibrium (LTE) and 1D plane-parallel atmospheres. The effect of non-LTE is generally taken into account where it matters most (e.g., hot stars and metal-poor stars Lanz & Hubeny 2003; Lind et al. 2012) and some 3D theoretical models (e.g, Magic et al. 2013) have begun to emerge but these techniques have not been widely adopted due to the fact that they are very computationally expensive. Theoretical stellar spectra are most unreliable for very cool stars and very hot stars (Allard et al. 1997; Martins & Coelho 2007; Bertone et al. 2008; Allard et al. 2013; Rajpurohit et al. 2014, 2016). The former is a particularly acute problem for both exoplanet and galaxy studies. All the habitable zone planes found by TESS will be around M dwarfs (France et al. 2015) and counting cool dwarf stars in the absorption lines of integrated spectra of galaxies has emerged as a way to constrain the initial mass function (IMF) (Spinrad 1962; Wing & Ford 1969; Cohen 1978; Frogel et al. 1978; Kleinmann & Hall 1986; Diaz et al. 1989; Ivanov et al. 2004; Conroy & van Dokkum 2012a) in unresolved stellar populations. The limitations of the theoretical stellar spectra in the cool star regime makes it necessary to turn to empirical stellar libraries.

Sánchez-Blázquez et al. (2006a) created a landmark empirical optical stellar library, the MILES library. The MILES library, consisting of nearly 1000 stars, covers a wide range of stellar parameter space over the wavelength range $0.35 - 0.75\mu m$. The MILES library enabled the creation of more precise SPS models which in turn facilitated a greater understanding of galaxies beyond the reach of resolved stellar population studies.

However, as stated previously, the optical window does not contain a complete picture of a stellar population. A major advance occurred with the release of the Rayner et al. (2009) Infrared Telescope Facility (IRTF) spectral library. The creation of this near-IR library was a great step forward for SPS models (e.g., Conroy & van Dokkum 2012a; Spiniello et al. 2012; Meneses-Goytia et al. 2015; Röck et al. 2016). A limitation of the IRTF stellar library is its narrow stellar parameter range, all stars being around solar-metallicity. Furthermore, there is very little overlap in the stars between the MILES and IRTF stellar libraries. The X-Shooter Spectral Library (XSL) is observing many stars with continuous spectral coverage from $0.35 - 2.5\mu m$ and will be another valuable library once complete (e.g., Chen et al. 2014).

In this work we present the Extended IRTF Library. The extension is two-fold: we expanded the metallicity coverage to $-1.7 <$[Fe/H]< 0.6 a large expansion from the just solar-metallicity coverage of the original IRTF library, and we selected all of our objects from the MILES stellar library to provide continuous coverage from the optical through near-IR. In addition to the new stellar library we also present an interpolator that uses a data-driven model created from the new stellar library. The interpolator

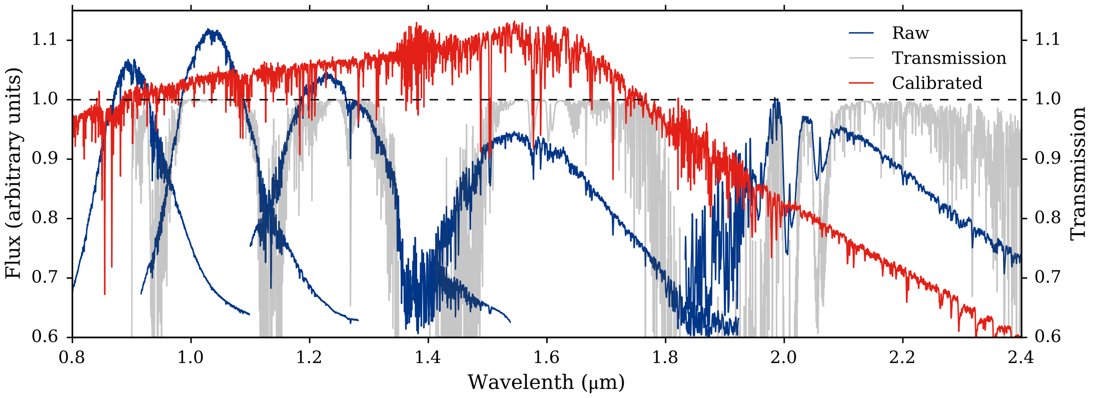

Figure 2.2: Demonstration of the telluric correction for the star BD+053080. We show the atmospheric transmission spectrum (grey), the spectrum for BD+053080 before the telluric correction (blue), and the spectrum after telluric correction and flux calibration (red).

generates a stellar spectrum as a function of $T_{\rm eff}$, log$g$, and [Fe/H]. This interpolator provides smooth variation in stellar spectra across parameter space, and is an important component in the creation of stellar population models.

The rest of this paper is organized in the following manner: Section 2 details our sample selection, observational strategy, data reduction, and characteristics of the objects and data. In Section 3 we describe our interpolator and assess its quality. In Section 4 we explore the behavior of the library and the interpolator and compare both to theoretical predictions. Finally, in Section 5 we summarize the main points of this work.

2.2 The Extended IRTF Library

2.2.1 Sample Selection

We selected our target stars from the MILES stellar library (Sánchez-Blázquez et al. 2006a), avoiding known spectroscopic binary stars, to provide continuous coverage from the optical to the near-IR. We selected our targets to span a stellar parameters that would enable the creation of stellar population models for intermediate and old population ages as determined with stellar isochrones from the MIST project (Choi et al. 2016) .

In Figure 2.1 we compare the MILES coverage in $T_{\rm eff}$-log$g$ space and the stars observed with IRTF for this work. Each panel shows the stars in a given metallicity bin where the stars from the MILES library are represented as blue circles and the stars observed for this library are represented as red circles. Each panel also displays 13.5 Gyr (black line) and 3 Gyr (grey line) MIST isochrones. Selecting targets by eye, we achieved relatively uniform coverage along the intermediate age isochrones for all but the lowest metallicity bins and the old age isochrones across all the metallicity bins. In particular, we have well sampled turn-offs for all except the intermediate-age, low-metallicity stellar tracks. The metallicity distribution of the library is summarized in panel f of Figure 2.1.

Throughout this work we use the stellar parameters determined by Prugniel et al. (2011) and Sharma et al. (2016) instead of those reported by Cenarro et al. (2007). The stellar parameters determined by Cenarro et al. (2007) were determined heterogeneously, depending on which additional literature data was available for a given

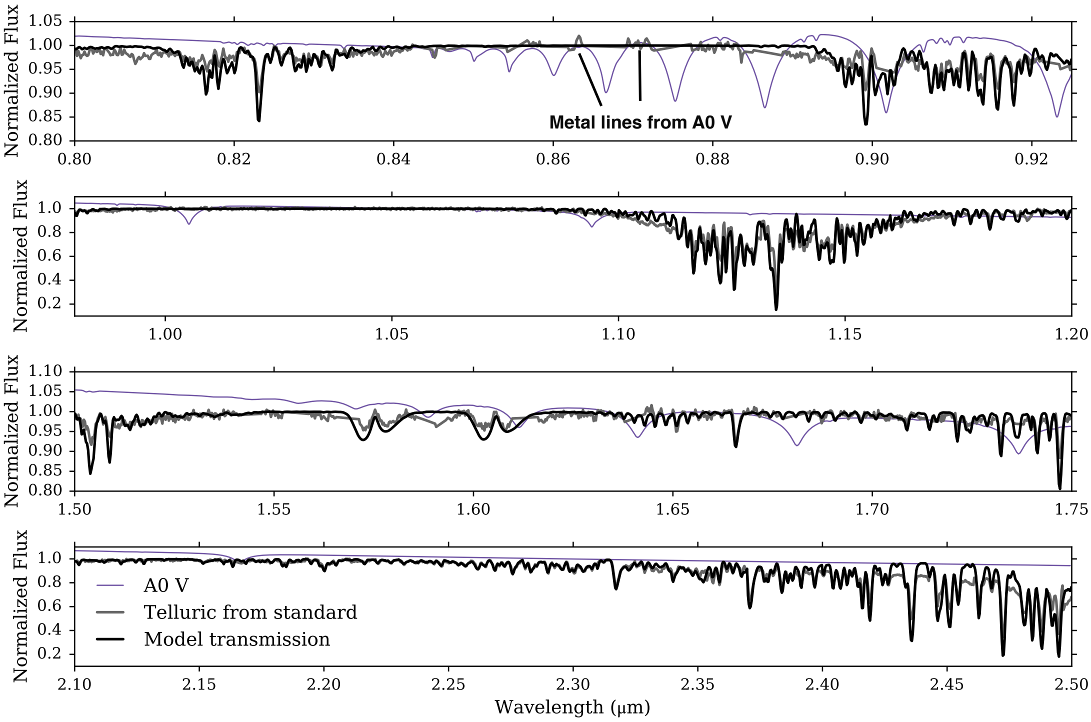

Figure 2.3: Comparison of a theoretical atmospheric transmission spectrum (black) to an example of a telluric spectrum (grey) used to telluric correct the library spectra and a theoretical spectrum of an A0 V star (purple). In regions of the A0 V spectrum containing strong hydrogen and metallic lines (top panel), the telluric spectrum includes unphysical features. We indicate a few of these features introduced by the metal lines in the A0 V stars in the top panel. Note that no attempt was made to fit the model transmission spectrum to the observations; the model is meant only to guide the eye.

star. As a result, the stellar parameters from Cenarro et al. (2007) vary in quality and trustworthiness. Prugniel et al. (2011) sought to revise the stellar parameters of the stars in the MILES library in a more robust and homogeneous way. They used ELODIE spectra to perform full-spectrum χ^2 minimization fits (Koleva et al. 2009) between the MILES spectra and templates built from the ELODIE 3.2 library (Wu et al. 2011). Sharma et al. (2016) revised the stellar parameter values of MILES library stars with $T_{\rm eff} < 4800$ K with an improved interpolator.

It is important to note that the accuracy of SPS models is predicated upon

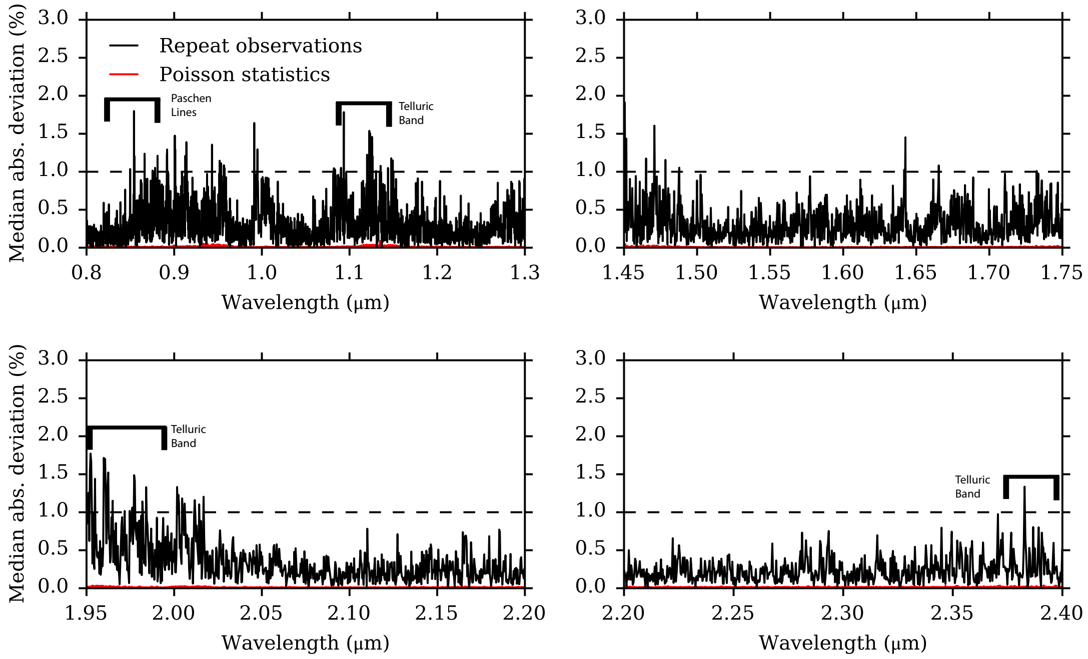

Figure 2.4: Comparison of the median absolute deviation of the Poisson and empirical uncertainty values of five library stars with repeat observations with the same standard star. We indicate 1% uncertainty with the dashed line as a guide. The Poisson uncertainty is ∼.2% throughout the wavelength range and underpredicts the empirical uncertainty (difference between the two observations) but nonetheless the median empirical uncertainty is generally $< 1\%$. The regions where the uncertainty exceeds 1% are regions contaminated by telluric absorption or where there exist prominent hydrogen absorption features in the A0 V standard stars.

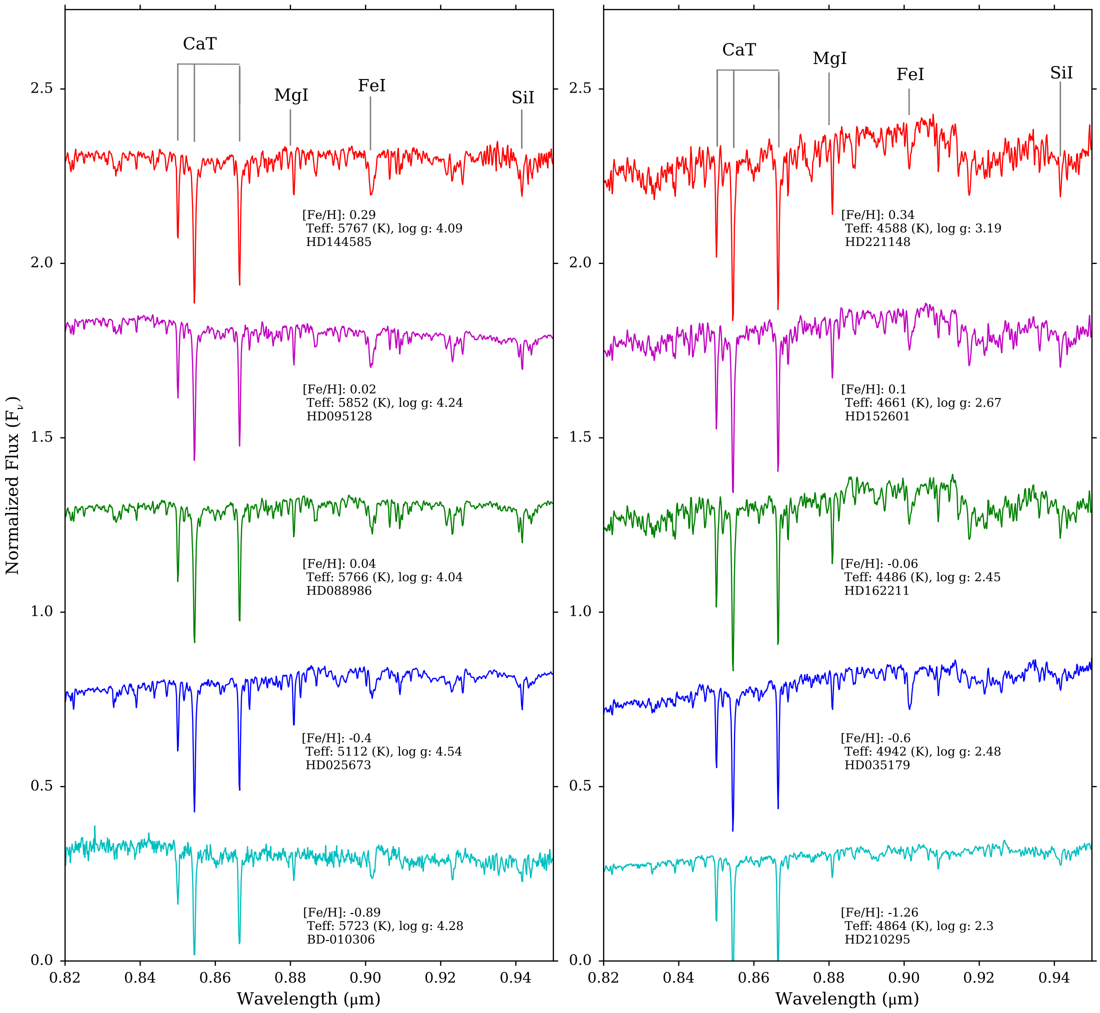

Figure 2.5: Sequence of stars on the main sequence turn-off (left) and subgiant branch (right) plotted as a function of metallicity metallicity over the *I* band (0.82-0.95μm).The spectra have been divided by the median flux value over the wavelength region shown and offset by constants.

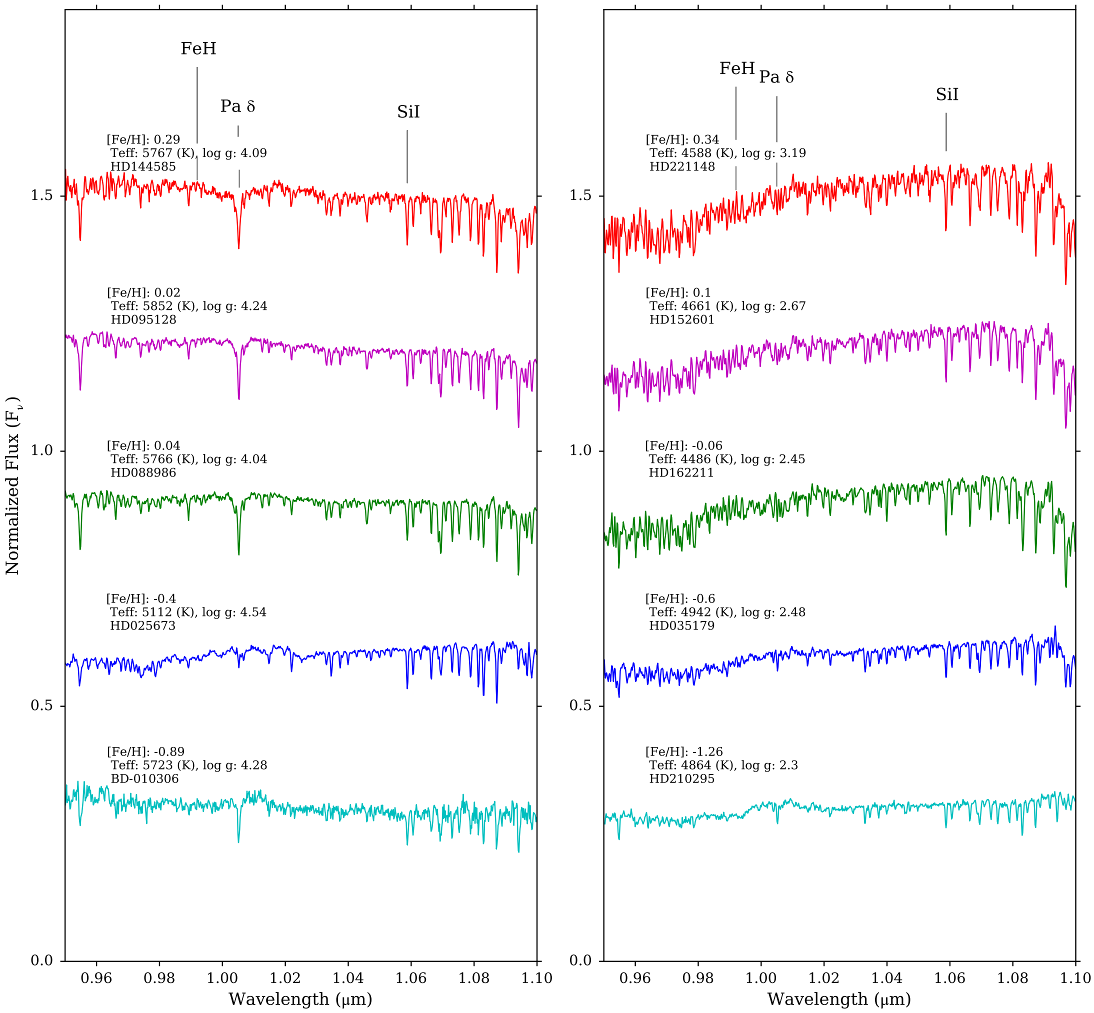

Figure 2.6: Same as Figure 2.5 except over the Y band (0.95-1.10μm).

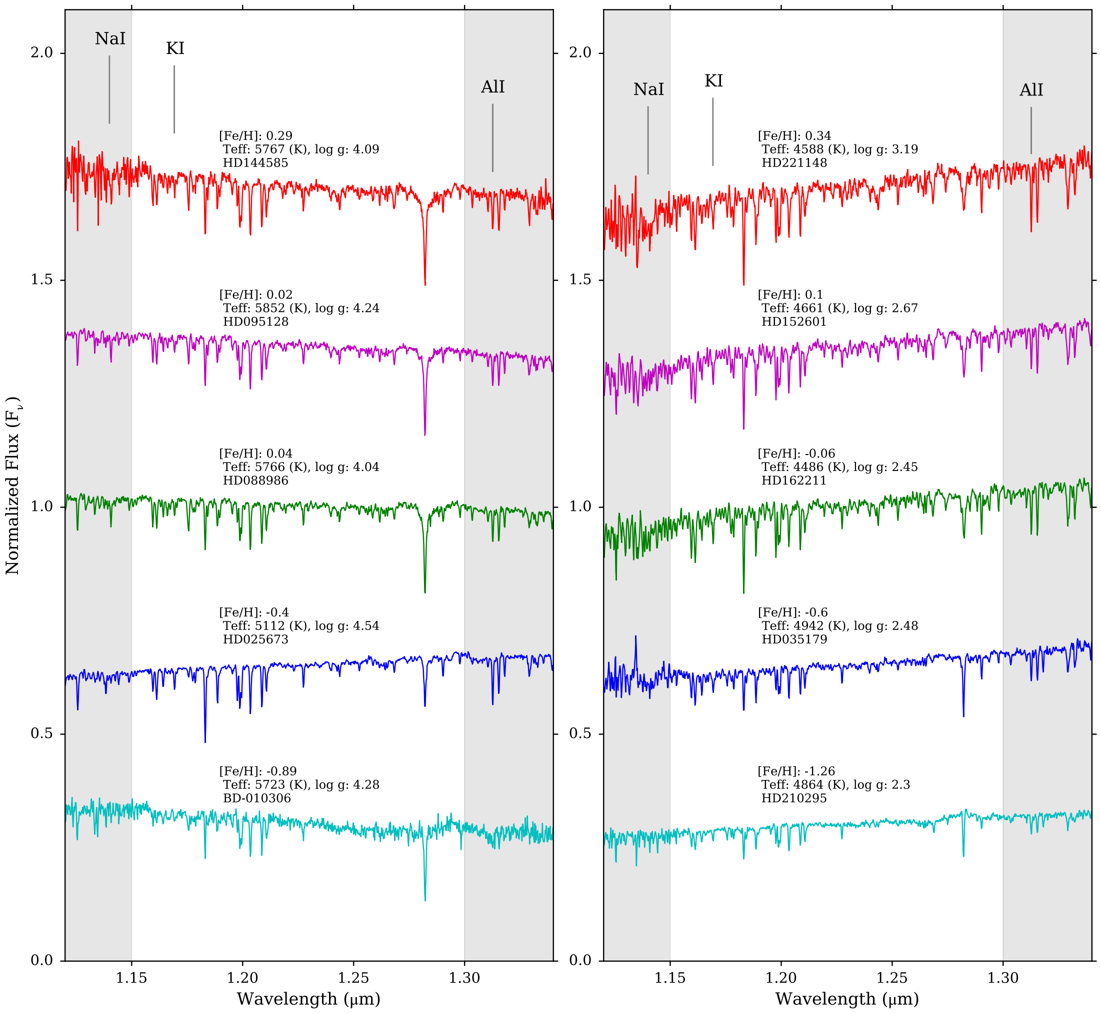

Figure 2.7: Same as Figure 2.5 except over the *J* band (1.12-1.34μm). The vertical grey bands mark regions of poor transmissivity due to telluric absorption.

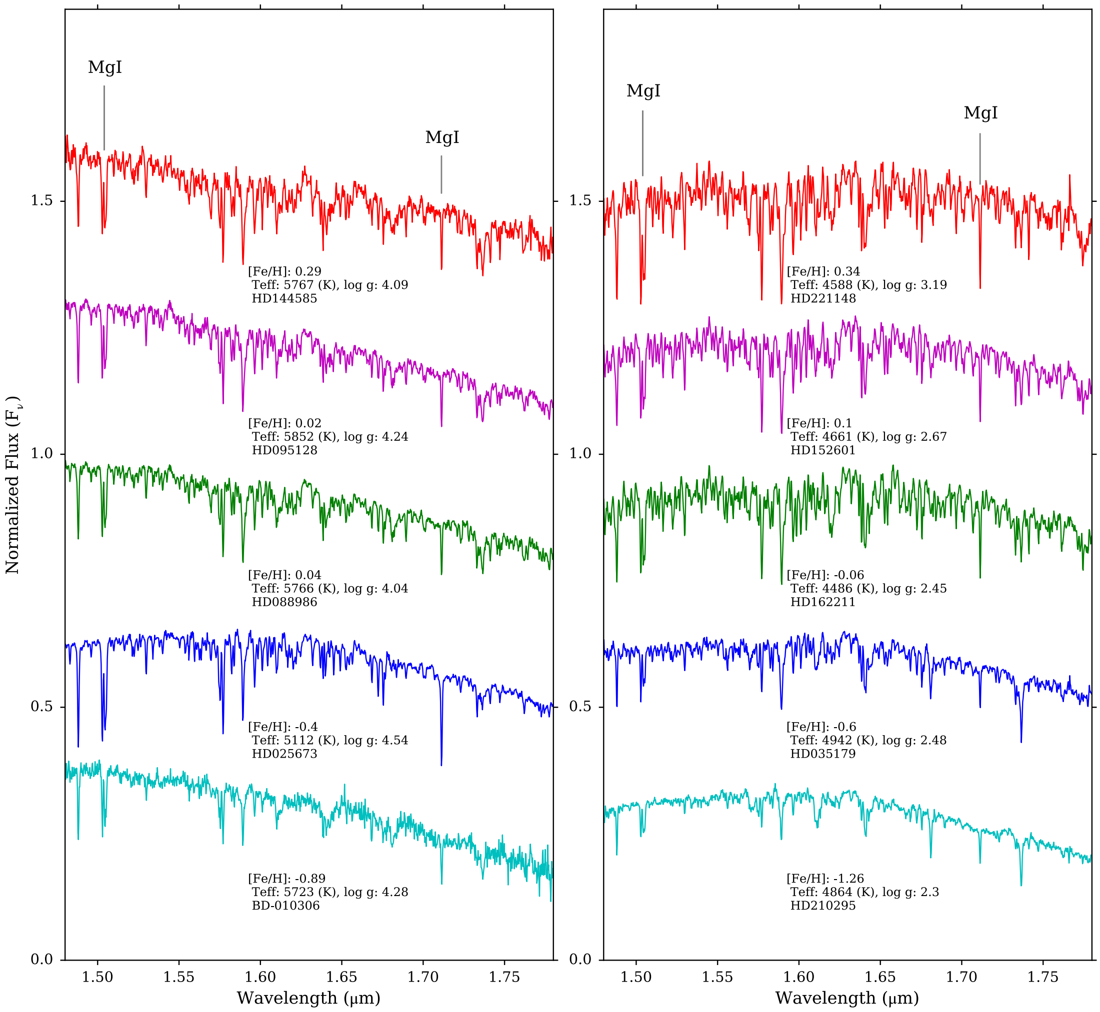

Figure 2.8: Same as Figure 2.5 except over the *H* band (1.48-1.78μm).

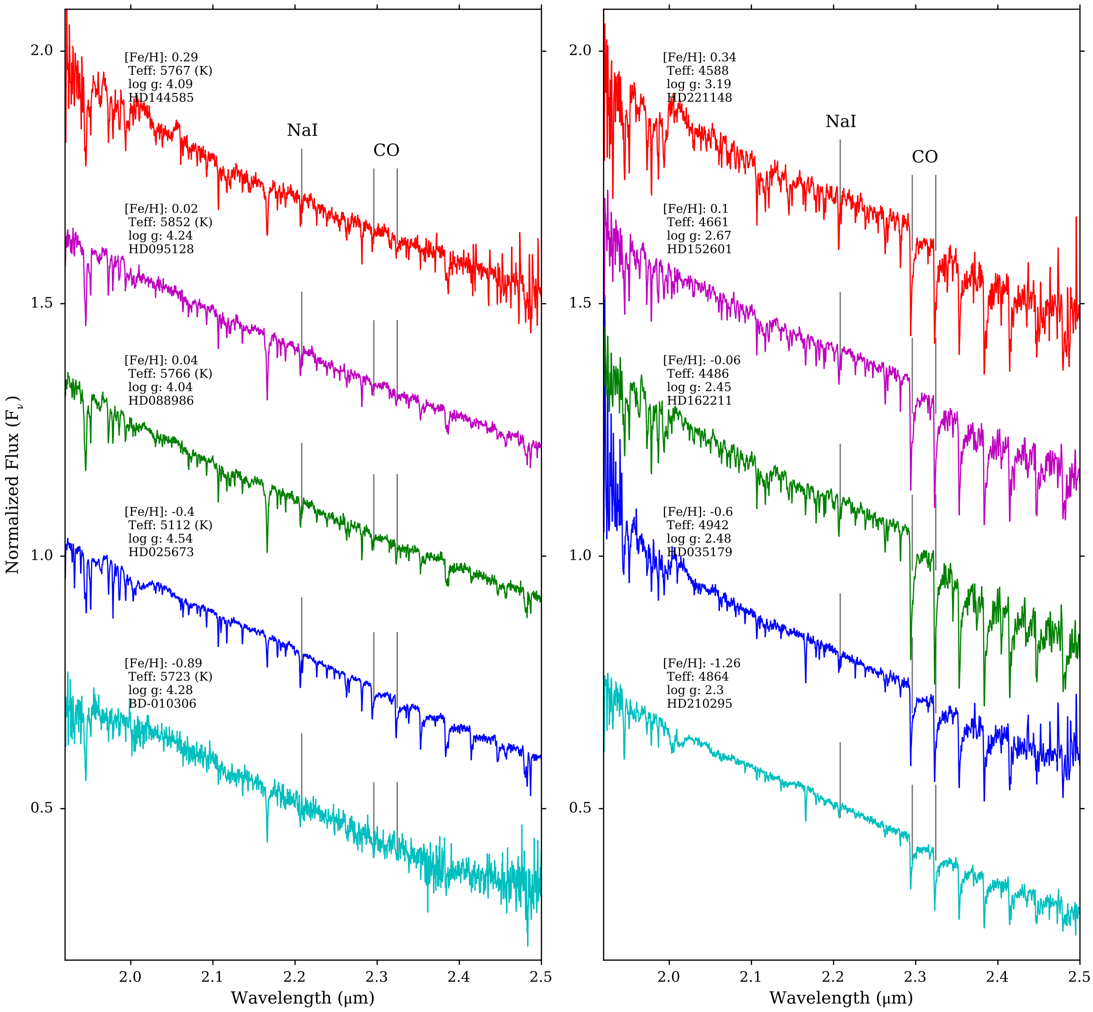

Figure 2.9: Same as Figure 2.5 except over the K band (1.92-2.50μm). The vertical grey band marks a region of poor transmissivity due to telluric absorption.

the accuracy of the stellar parameters. Percival & Salaris (2009) found that a change of 100 K in the effective temperature of stars in a stellar library can propagate into a 20% error in absolute ages estimated from stellar population models. Furthermore, derived exoplanet properties are a direct function of the stellar parameters of the host star (e.g., Mann et al. 2015). These examples underly the need for accurate stellar parameters in empirical stellar libraries.

We do not make use of the Rayner et al. (2009) library (except for comparison with re-observed stars) for two reasons. First, as we will discuss later, the detectors on SpeX have been upgraded since the IRTF library of Rayner et al. (2009). A significant component of the upgrade included a blueward extension to $0.70\mu m$. The original IRTF library had a blue cutoff of $0.8\mu m$, which meant that there was no overlap with the red cutoff of the MILES library. Second, there were few stars that were in both the MILES stellar library and the Rayner et al. (2009) library. Our requirement for continuous coverage from the optical to the near-IR for a common set of stars led us to not make use of the original IRTF library.

2.2.2 Observations

We carried out 24 nights of observations from August 2014 to June 2016 using the upgraded SpeX instrument on IRTF. For details on SpeX we refer the reader to Rayner et al. (2003). In short, SpeX is a medium-resolution, cross-dispersed spectrograph equipped with a 2048x2048 Hawaii-2RG array. All stars were observed in the short-wavelength cross-dispersed mode (SXD) using the $0''.3$ (2 pixel) slit, providing a nominal resolving power of R$\sim$2000. We observed every star at the parallactic angle.

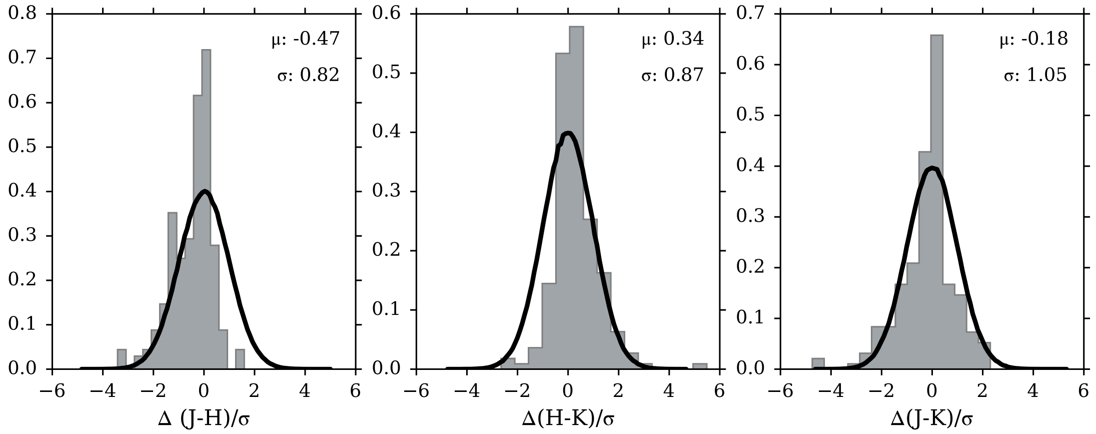

Figure 2.10: Normalized histogram of the residuals between the observed and synthesized 2MASS $J-H$, $H-K_S$, and $J-K_S$ colors for a subset of the library stars without quality issues, shape issues, or stellar parameter issues (see later discussion)divided by the uncertainty. We have indicated the mean offset, μ, and standard deviation, σ. A Gaussian distribution with $\sigma = 1$ is also shown.

The upgraded detectors extended the wavelength coverage further into the blue, giving a wavelength range from $0.7 - 2.55\mu m$. The upgraded detectors also remove the 0.06 μm gap that used to exist between the H and K bands (Rayner et al. 2009).

An A0 V star (or a star of similar spectral type) was observed either before or after each science object to correct for absorption due to the Earth's atmosphere and to flux calibrate the science object spectra. For most stars the airmass difference between the target star and the standard star was less than 0.1. Standard stars were also chosen to be located within 10 degrees of the target object whenever possible, to minimize the effects of any differential flexure in the instrument between the observations of the target and standard. Due to a shortage of standard stars near some targets, in a few cases airmass differences between the two reached 0.15.

We took a set of internal flat field and argon arc lamp exposures with each

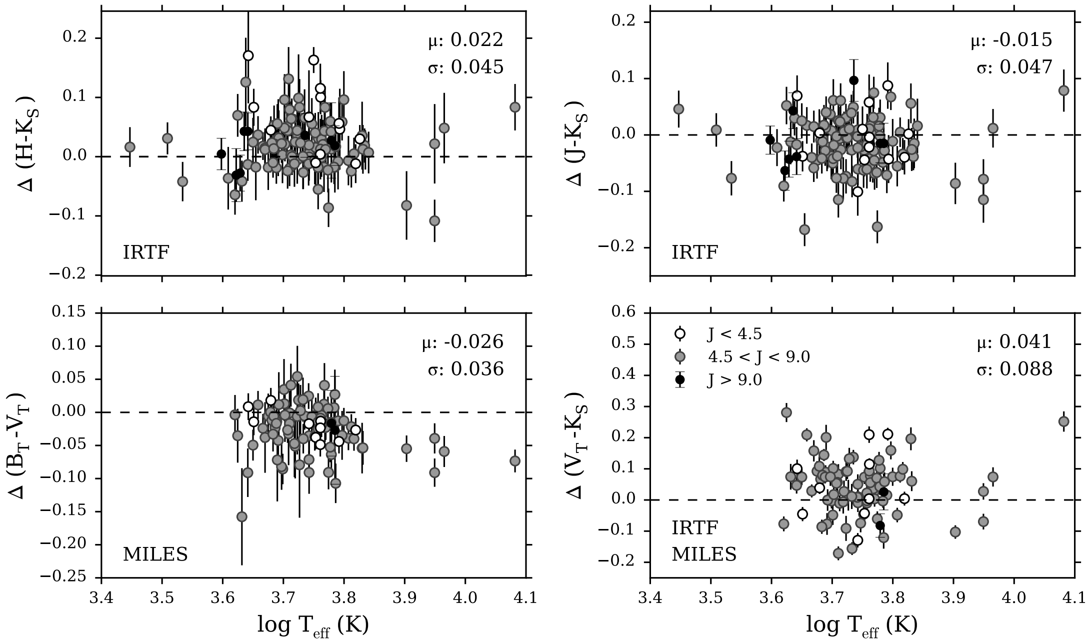

Figure 2.11: Computed residuals between the observed and synthesized 2MASS $H-K_S$, $J-K_S$, optical $B_{\rm T}-V_{\rm T}$, and $V_{\rm T}-K_S$ colors of 124 library stars with relatively good (uncertainty $\leq 3\%$) 2MASS photometry. The plotted error bars show the 2MASS and Tycho errors in magnitudes. The $V_{\rm T}-K_S$ colors are a test of how well the MILES and IRTF stars were stitched together.

target-standard pair. This procedure helps minimize the possible effects of flexure of the detector on the quality of the flat-fielding and the final wavelength solution.

2.2.3 Data Reduction

Data were reduced following the same procedure described in Rayner et al. (2009) using the facility IDL-based reduction package for SpeX, Spextool v4.0.4 (Cushing et al. 2004). Cushing et al. (2004) provides a comprehensive explanation of the data reduction process and so here we summarize and highlight the main steps.

The science images were processed by correcting for nonlinearity, subtracting the pairs of images taken at two different slit positions, and dividing the pair-subtracted images by a normalized flat field. Any residual sky signal left over from the pair subtraction is removed from the image and the spectra in individual orders were then optimally extracted. Argon lines were used to wavelength calibrate the data. For the SXD observing mode, several arc exposures are combined to increase the signal-to-noise ratio (SNR).

We used the A0 V standard star observed with each science object to correct for the telluric absorption and flux calibrate the science object using the method described by Vacca et al. (2003). In brief, the method uses a theoretical spectrum of Vega to determine the intrinsic spectrum of any A0 V star by scaling and reddening the Vega model to match the near-IR magnitudes of the observed A0 V stars and modifying the model to account for differences in line strengths, radial and rotational velocities, and spectral resolution. This method also corrects for instrument throughput and flux calibrates the spectra of the target stars.

In Figure 2.2 we show the SpeX spectrum for the library star BD+0503080 before (blue) and after (red) the telluric correction. The before and after spectra have been vertically scaled to clarify the comparison. Also shown is atmospheric transmission at Maunakea for an airmass of 1.5 and precipitable water vapor of 3mm (Lord 1992). The before spectrum has areas of very low signal and features in wavelength regions that align with the deepest parts of absorption in the atmospheric transmission spectrum. The red line demonstrates that much of the noise in this region is decreased after the telluric correction. We subsequently mask the regions in the IRTF spectra where the atmospheric absorption is most prominent, 1.32 μm-1.41μm and 1.82μm-1.94μm.

This telluric correction procedure is quite effective in removing the effects of the atmosphere. However, A0 V stars have prominent hydrogen absorption lines and weak metal features, which can complicate the telluric correction. In Figure 2.3 we show a theoretical spectrum of an A0 V star (purple) comparef to an atmospheric transmission (black, the model for the first three panels is from Lord (1992) and the model for the fourth panel is from Clough & Brown (2005)), and an example of a telluric spectrum used to correct the library stars (grey). Both the A0 V and transmission spectra have been degraded to a resolution of $R \equiv 2000$. The top panel shows the region around the Paschen break, which can be particularly difficult to fit with models at the percent level. We also see that there are features in the telluric spectrum that are likely spurious and are associated with the metal lines in A0 V stars. We have labeled several of the metal features in Figure 2.3. Our telluric correction method also relies on modeling and removing the A0 V spectrum by modifying the model spectrum of Vega to match the A0 V star. This method is better than interpolating over the hydrogen lines, as there are telluric features at those wavelengths, but there are imperfections to this process. We address the uncertainties associated with these difficulties shortly. The telluric correction appears to be well-behaved in regions away from hydrogen lines and the weak metal lines.

After the telluric correction the spectra extracted from different orders were merged into a single, continuous spectrum. In Rayner et al. (2009) an additional scale factor was needed to match the flux levels of the different orders but the updates to the SpeX detector made this step unnecessary. All wavelengths are in vacuum. Spextools provides a measure of uncertainty for each reduced spectrum and the typical formal SNR of the library is very high with a mean value ~ 500. However, the quoted SNR only includes statistical (photon counting) uncertainty, not correlated noise or any source of systemic uncertainty. Most importantly, we are interested in the uncertainty our telluric correction method introduces (see above).

To empirically measure the uncertainty of the library stars we compare five stars with repeat observations: HD007106, HD021197, HD138776, HD201891, and HD204613. For all of these observations we followed the same observing and reduction strategies as described above. The repeat observations were done with the same standard star. We computed the difference between the normalized spectra of the same star and divided by a factor of $\sqrt{2}$ (black line in Figure 2.4). The Poisson uncertainty is the inverse of the wavelength-dependent average of the SNR for each pair of target observations (red line in Figure 2.4).

While Figure 2.4 demonstrates that the Poisson uncertainty underpredicts the empirical uncertainty for most of the wavelength range, the empirical uncertainty is almost always $\sim 0.2 - 0.3\%$. The exceptions to this are in regions heavily affected by telluric absorption (upper-right and lower left panels) or where there are many broad hydrogen absorption and metal lines in the A0 V standard stars (upper-left panel). Still, even in these regions the uncertainty is 1-2%.

The empirical uncertainty presented here encompasses differences in weather and/or observing conditions, potential human error in the reduction process, and correlated noise. It does not assess the uncertainty associated with using different standard stars or using different techniques for telluric correction (e.g. van Dokkum & Conroy 2012; Kausch et al. 2015). It is not known what effect using a different telluric correction technique would have on the quality of SpeX spectra. Assessing the impact of these additional sources of uncertainty is beyond the scope of this work but will be pursued in the future.

To demonstrate the quality of the individual features, noise in the continuum after the full reduction, and qualitative metallicity trends we show a sequence of stars ordered by metallicity over the $IYJHK$ photometric bands in Figures 2.5-2.9. The left-panel shows stars roughly on the main-sequence turn-off and the right-panel shows stars roughly on the subgiant branch. We highlight a selection of prominent lines in each bandpass. We qualitatively see general trends with metallicity. However, we are not completely controlling for temperature and surface gravity effects and so it is difficult to say anything definitively about potential trends. In Section 4 we look at equivalent width trends as a function of temperature and surface gravity.

Correcting to Restframe Wavelength

We used the code `Prospector` (Johnson et al. in prep)[5] to determine line-of-sight velocities for each star, which were used to correct the final reduced spectra to the restframe. Prospector is a code for inference of physical parameters from spectroscopic data via MCMC sampling of the posterior probability distributions. To obtain estimates of the posterior velocity distribution using prospector, a spectral model is constructed for

---

[5] https://github.com/bd-j/prospector

each star by linearly interpolating the C3K theoretical spectral models[6] to the Prugniel et al. (2011) and Sharma et al. (2016) parameters, and then broadened to the resolution of IRTF. At each MCMC step this spectrum is shifted in velocity and the likelihood of the data given the redshifted model is calculated, after masking regions of strong telluric lines.

2.2.4 Flux Calibration

To de-redden the flux calibrated spectra we used the $E(B-V)$ values given in the MILES stellar library (P. Sánchez-Blázquez, private communication),

$$f_{\lambda}^{\mathrm{cor}}(\lambda) = f_{\lambda}(\lambda) \times 10^{(0.4 \times A_{\lambda})}, \tag{2.1}$$

where $f_{\lambda}(\lambda)$ is the observed spectrum and A_{λ} is the extinction law as a function of wavelength. We adopted the near-IR law given by Fitzpatrick & Massa (2007) with $R_V = 3.1$,

$$A_{\lambda} = 1.056 \times E(B-V)\lambda^{-1.84}. \tag{2.2}$$

The MILES spectra were de-reddened using the same $E(B-V)$ values.

We absolutely flux calibrated the IRTF spectra using a correction factor based on 2MASS JHK_S photometry. We computed the correction factor for each star by using the reported 2MASS magnitude, μ, and error, σ, to create a random normal distribution for each bandpass, $\mathcal{N}\left(\mu, \sigma^2\right)_{X_{\mathrm{obs}}}$, where X denotes a given bandpass. We

---

[6] This is the theoretical library described in Conroy & van Dokkum (2012a) with some minor updates to the line lists.

created a distribution of calibration factors using the following expression,

$$f(C_X) = 10^{(X_{\rm syn} - \mathcal{N}X_{\rm obs})/2.5}, \tag{2.3}$$

$X_{\rm syn}$ is the AB synthetic magnitude computed from the spectrum. The observed spectrum for each star is multiplied by a single correction factor,

$$\langle C \rangle = \frac{\sum_X w_X {\rm Peak}[f(C_X)]}{\sum_X w_X}, \tag{2.4}$$

where the weight, w_X, is the inverse variance given by the 1σ deviations from $f(C_X)$, and Peak$[f(C_X)]$ is the peak of the distribution of calibration factors. As in Rayner et al. (2009), this scaling has the effect of shifting the entire spectrum up or down so that the overall absolute flux level is correct, while simultaneously preserving the relative flux calibration of each spectral order derived from the observations and telluric correction procedures.

The publicly available MILES spectra are normalized to unity so in addition to absolutely flux calibrating the IRTF spectra we needed to absolutely flux calibrate the MILES spectra. We corrected the MILES spectra in a similar manner to the IRTF spectra using the average of the available BV photometry. We preferentially used Tycho (Høg et al. 2000b,a) $B_T V_T$ photometry and for stars where that is not available we used Johnson BV photometry taken from Simbad. However, uncertainties in the observed photometry and the different epochs at which the optical and near-IR photometry were taken means that for most stars this correction of the IRTF and MILES spectra does result in perfect agreement in the overlap region. Since we are eventually going to stitch the two spectra together we need them to be at the same flux level, and any incongruity would affect the derived bolometric luminosities and spectral shapes around $0.7\mu m$. We force the MILES and IRTF spectra by computing the percent difference of the flux in the overlap region assuming that the IRTF spectra is at the correct level. Then, to ensure that we do not have flux going into the negative, if the difference is negative we shift the IRTF spectrum by that factor. If the difference is positive we shift the MILES spectrum by that factor.

In the case where 2MASS photometry is not available (e.g. HD 134439) but there is available near-IR photometry from Simbad we used the same basic method as described above but only considered the K_S-band photometry, using the raw difference between the observed and synthetic magnitudes. For variable stars in the sample, the absolute flux calibration is only approximate since the 2MASS photometry was obtained at an earlier epoch than the SpeX spectroscopy. However, consistency between the MILES spectra and IRTF spectra is a natural check of the calibration.

Quality of Flux Calibration

We assess the quality of the spectral shape of the IRTF spectra in Figure 2.10. Here we show a histogram of the error-normalized color differences between the observed and synthesized 2MASS (Skrutskie et al. 2006) $J-H$ (left), $H-K_S$ (middle), and $J-K_S$ (right) colors. We also show in each panel of Figure 2.10 a Gaussian distribution of $\sigma = 1$. Overall the differences between observed and synthesized colors are consistent with the errors in the observed colors. There is some tension in the $J-H$ and $H-K_S$

colors but not in $J - K_S$, suggesting that there might be some modest issue with the $H-$band normalization.

We further examine the quality of the flux calibration by computing the residuals between observed and synthetic photometric colors. For this exercise we choose stars with 2MASS photometry better than 3% for the JHK_S bands and with good quality spectra and well-determined stellar parameters and later for how we determine this); 124 stars in total. In Figure 2.11 we show these residuals as a function of effective temperature for the colors $H - K_S$ (upper-right), $J - K_S$ (lower-left), $B_{\mathrm{T}} - V_{\mathrm{T}}$ (we only display stars with Tycho photometry in this plot) and $V_{\mathrm{T}} - K_S$ (lower-right). The observed photometry have been redden corrected using the same $E(B - V)$ values used to correct the spectra. Points are color-coded by J band magnitudes because at $J \sim 9$ and $J \sim 4.5$ the 2MASS observing strategy changed to avoid non-linearity and saturation for bright stars. This will affect the photometry and corresponding error bars.

For each color we find mean residuals of 0.022, −0.015, −0.026, and 0.041, respectively. The residuals for the near-IR colors are consistent with those of Rayner et al. (2009) and the $B_{\mathrm{T}} - V_{\mathrm{T}}$ residuals are consistent with the result from Sánchez-Blázquez et al. (2006a). The $V_{\mathrm{T}} - K_S$ residuals are a reflection of the overall uncertainty of the absolute flux calibration and stitching of the MILES and IRTF spectra. Given the large wavelength baseline and the complications associated with stitching together different spectral datasets, it is not surprising that this color shows the largest offsets between synthesized and observed data.

We explore the reliability of the IRTF spectra further in Figure 2.12, where we compare our new IRTF spectra to the spectra in the Rayner et al. (2009) IRTF library for four stars in common between the two libraries. In this figure we also compare to $BVJHK_S$ photometry and the MILES spectra. The bottom panel for each star shows the ratio between the old and new IRTF spectra. The agreement in the shape is excellent for HD076151 and HD036003, except at the blue end for the former where we see a small, unphysical bump in the flux at $\sim 0.8\ \mu m$. This is caused by difficulty modeling the Paschen break in the A0 V standard and occurs in a small subset (4%) of our sample. We found that the stars showing this bump were flux calibrated with standard stars that show nebulosity (labelled A0 Vn in Simbad). This was not an issue in the Rayner et al. (2009) library since the detectors did not extend to as blue wavelengths. This issue is addressed further in Section 3.

We explore the reliability of the flux calibration of the blue end of the SpeX spectra by comparing the Extended IRTF library stars with the Next Generation Spectral Library (NGSL Gregg 2001) which spans 1680-10000Å. The NGSL spectra were observed on the STIS instrument on the Hubble Space Telescope (HST) so we can use these spectra as flux standards. There are 35 Extended IRTF Library stars that are also in NGSL. To measure how consistent the flux calibration of the Extended IRTF Library is with NGSL we synthesized SDSS i and z photometry (the two passbands that span the blue end of the Extended IRTF Library spectra) for both the NGSL and Extended IRTF Library spectra for the 35 overlap stars. In Figure 2.13 we show the residuals between the NGSL i-z colors and the IRTF i-z colors. The mean residual is sub-1% with a scatter on the order of a few percent, indicating consistency between the flux calibration of the NGSL spectra and the IRTF spectra.

The shape of the spectra is important for computing bolometric luminosities and in creating stellar population models. Consistency at the level of absorption lines between the old and new IRTF spectra is also important as that will change equivalent widths and affect the accuracy of fitting stellar features using these spectra. From Figure 2.12 we can see that the differences between the Rayner et al. (2009) stellar features and the stellar features in the spectra from this work are 1-2%, much smaller than the over all shape and flux level differences.

In Figure 2.14 we show a small sample of stars in the library. They are ordered from cool (upper left corner) to hot (lower right corner). The reduced, redshift corrected, and absolutely flux-calibrated spectra for all the stars observed as part of the Extended IRTF library are available at the IRTF website `http://irtfweb.ifa.hawaii.edu/~spex/IRTF_Spectral_Library/`.

2.2.5 Combining the Extended IRTF Library with MILES

The wavelength solution for the SpeX data is not linear and so there is no constant $\Delta\lambda$ between pixels. In practice this means that the spectra for the different stars in the library are on different wavelength grids. The first step to combining the SpeX spectra with the MILES spectra is to put all the Extended IRTF Library stars on the same wavelength grid. We did this by finding the wavelength range shared by all the stars in the library, $\lambda_{\rm min} = 0.713\ \mu m$ and $\lambda_{\rm max} = 2.559\ \mu m$ with 7408 wavelength points (with spacing between the pixels as $\Delta\lambda = 2.495$Å), and interpolated every spectrum onto that grid.

With the Extended IRTF Library spectra all on a uniform wavelength grid we combine them with the MILES spectra (where we have converted the MILES wavelengths from being in air to being in vacuum) as follows: for each star, we took the weighted average between the MILES and the Extended IRTF spectra between $0.713\mu m$ and $0.743\mu m$ using a linear ramp function. The Extended IRTF Library spectra between $0.713\mu m$-$0.743\mu m$ were placed on the MILES wavelength grid. At $0.728\mu m$ the blend is half MILES and half IRTF. The final, combined spectrum is a concatenation of the MILES spectrum at $\lambda \leq 0.713\mu m$, the blended section, and the IRTF spectrum at $\lambda \geq 0.743\mu m$.

2.2.6 Bolometric Luminosity

To compute the bolometric luminosity for each star we extended the combined MILES and IRTF spectrum to 0.03μm and 20μm using the C3K stellar grid, and integrated the extended spectrum to obtain the bolometric flux. For most of the stars we used parallax based distances to convert from bolometric flux to bolometric luminosity. For the cluster stars we used cluster distances as a proxy for distance to the star (for NGC 6791, Grundahl et al. 2008; Harris 1996, 2010 edition, for the rest).

14 stars or 5% have neither parallaxes nor known distances and so the bolometric luminosity cannot be computed for them.

2.2.7 Resolution

The nominal resolution of the SpeX instrument is $R \sim 2000$. In this section we characterize the wavelength-dependent SpeX resolution, which is essential for using these spectra to model observational data. We also re-measure the MILES spectral resolution for comparison.

The spectral resolution of the MILES/IRTF data was measured by fitting theoretical stellar spectra to these data with the Prospector code. Briefly, for each wavelength regime of each star we calculate a posterior probability distribution for the line spread function (LSF) width and residual velocity, marginalized over the stellar parameters. The log-likelihood for the data given these parameters is calculated by simple χ^2 between the data and the smoothed theoretical model. This model is constructed by tri-linear interpolation of a high resolution version of the C3K theoretical stellar library in the stellar parameters log $T_{\rm eff}$, log$g$, and [Fe/H], which is then smoothed to a fixed resolution in $\Delta\lambda$(FWHM Å) using fast Fourier transforms, and interpolated to the wavelength scale of the data after shifting in velocity. This smoothing assumes a gaussian LSF with rms width $\Delta\lambda/2.355$. To minimize template mismatch, the stellar parameters log$g$, log $T_{\rm eff}$, and [Fe/H] are allowed to vary from the Prugniel et al. (2011) parameters within some tolerance. The velocity and instrumental resolution $\Delta\lambda$ (Å) have uniform priors within some reasonable range for each segment. The residual velocities are consistent with zero. This posterior probability is then sampled using MCMC techniques, specifically the ensemble sampling algorithm (Goodman & Weare 2010; Foreman-Mackey et al. 2013a). The resulting maximum a-posteriori values are reported, corrected for the resolution of the theoretical library.

These fits are done separately for several wavelength regions of the spectrum of each star; the optical spectra are split into 6 regions each 400Å wide, while 6 regions in the IRTF spectra are defined around prominent spectral features and avoiding regions of low atmospheric transmission.

In Figure 2.15 we show the median resolution (solid line) for 135 warm ($3980K < T_{\rm eff} < 6300$K) stars and the scatter (shaded region) over twelve wavelength segments spanning the full Extended IRTF spectra. In the top panel we show the wavelength segments for the part of wavelength space covered by the MILES spectra and in the bottom panel we show the wavelength segments for the wavelength space covered by the IRTF spectra. In the top panel we show a constant line at $\Delta\lambda = 2.54$(Å), the revised resolution found by Beifiori et al. (2011). The median resolution we find over the different wavelength regions of the MILES spectra is $\Delta\lambda = 2.54$(Å)± 0.19 (R $\sim$ 1970 at $0.5\mu m$), consistent with the Beifiori et al. (2011) value. In the bottom panel we show a constant line for $R \equiv \lambda/\Delta\lambda = 2000$. The median resolution we find over the different wavelength regions for the IRTF spectra is $R = 2020 \pm 230$, consistent with the quoted value for the SpeX instrument.

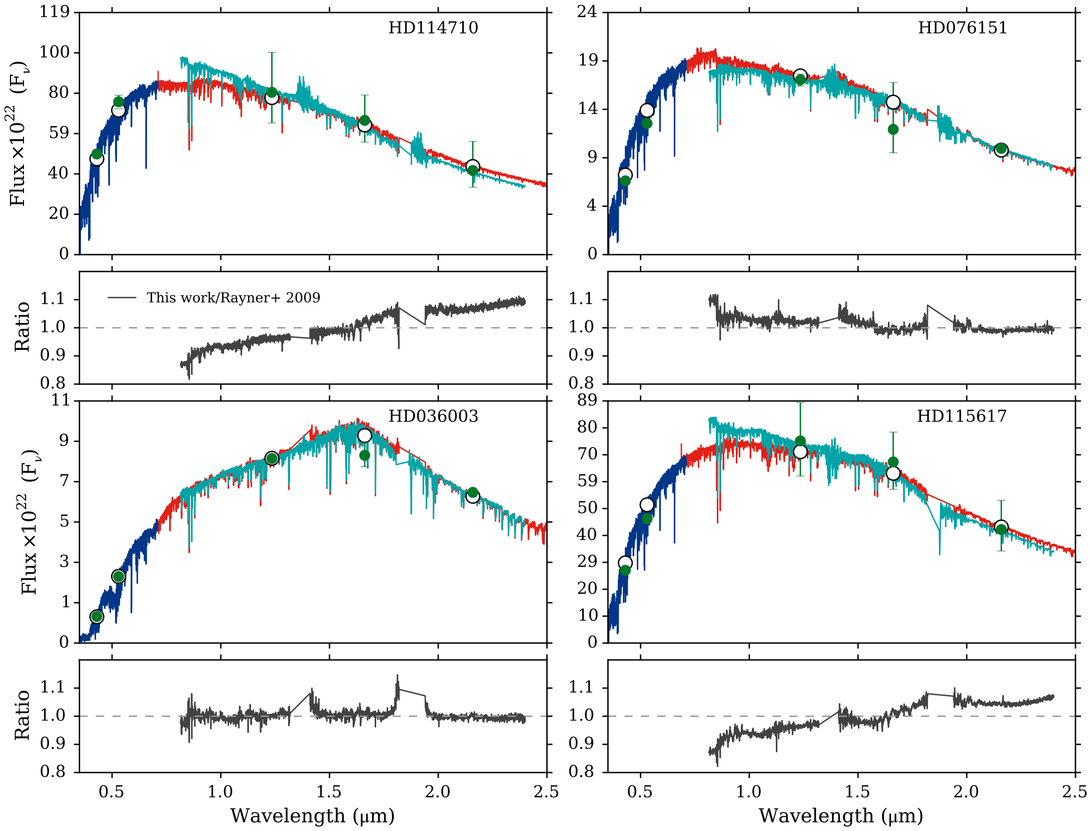

Figure 2.12: Comparison of new spectra presented in this work (red), spectra from the Rayner et al. (2009) library (turquoise), and the MILES spectra (blue) for four stars. Also plotted is the observed photometry (green circles) and synthetic photometry (open circles) derived from the spectra. The shape of the spectra presented in this work match well with the shape of the MILES spectra in the overlap region, and also with the observed JHK_S photometry. Below each spectral comparison we plot the ratio of the IRTF spectra presented in this work and the spectra from Rayner et al. (2009).

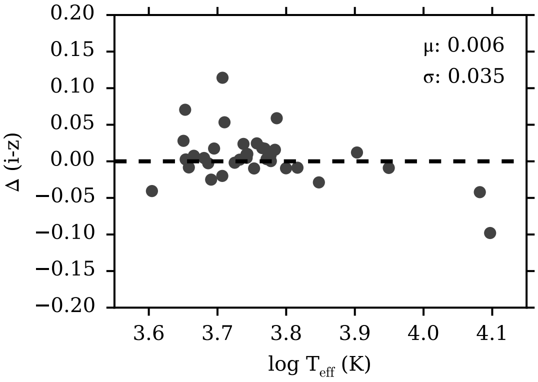

Figure 2.13: Computed residuals between the NGSL and IRTF synthetic SDSS $i - z$ colors for 35 stars that are in both libraries.

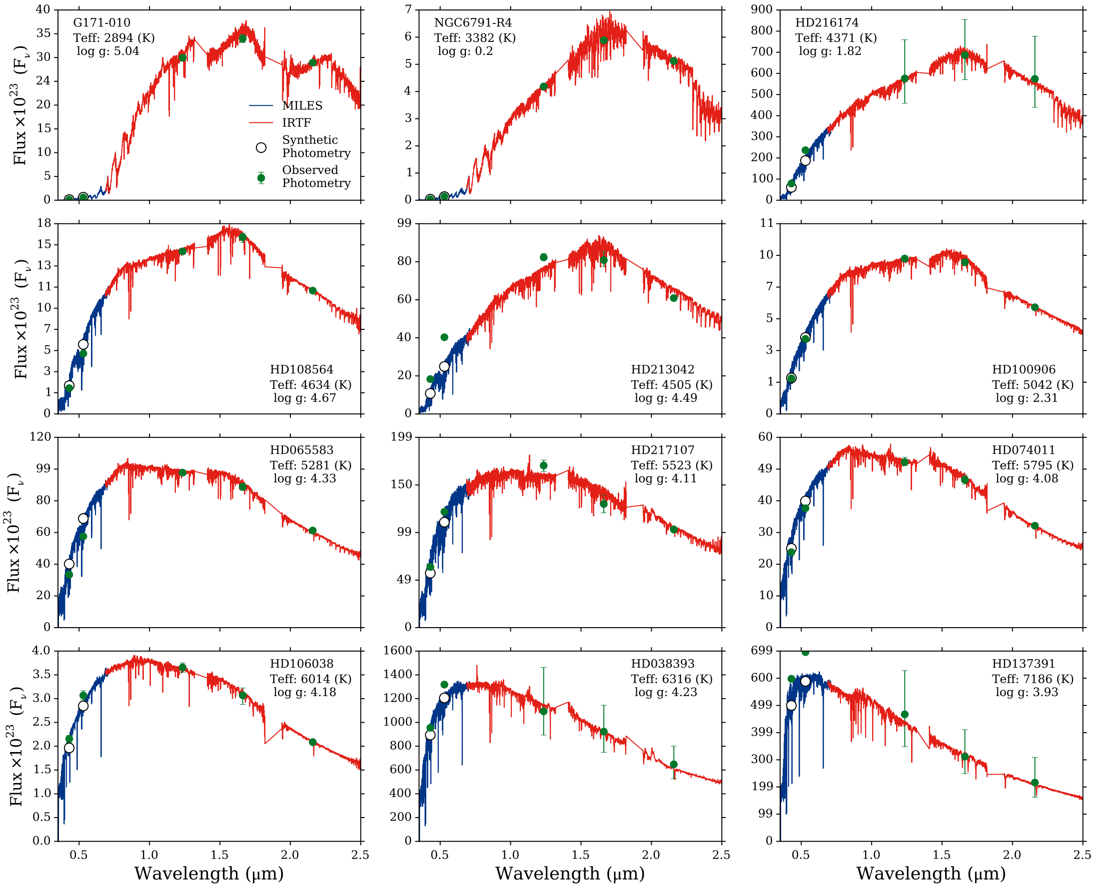

Figure 2.14: IRTF spectra (red) for a small subset of the library stars plotted with the MILES spectra (blue), observed photometry (green circles), and synthetic photometry (open circles).

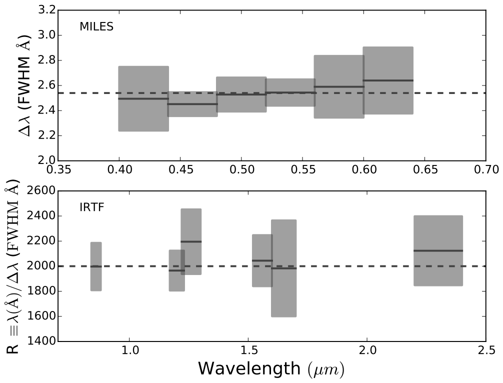

Figure 2.15: Median resolution (solid line) for 135 stars and the scatter (shaded region) over twelve wavelength segments. The top panel is the wavelength range of the MILES spectra with the resolution quoted as $\Delta\lambda$ (FWHM Å), the commonly quoted unit for MILES spectral resolution. We plot a constant line at $\Delta\lambda = 2.54$Å, the revised resolution found in Beifiori et al. (2011). The bottom panel is the wavelength range of the IRTF spectra and the resolution is quoted in $R \equiv \lambda/\Delta\lambda$. We plot a constant line at $R = 2000$, the quoted resolution for SpeX. For the MILES spectra we measure a median $\Delta\lambda = 2.54\text{Å} \pm 0.19$ and for the IRTF spectra we measure a median $R = 2020 \pm 230$.

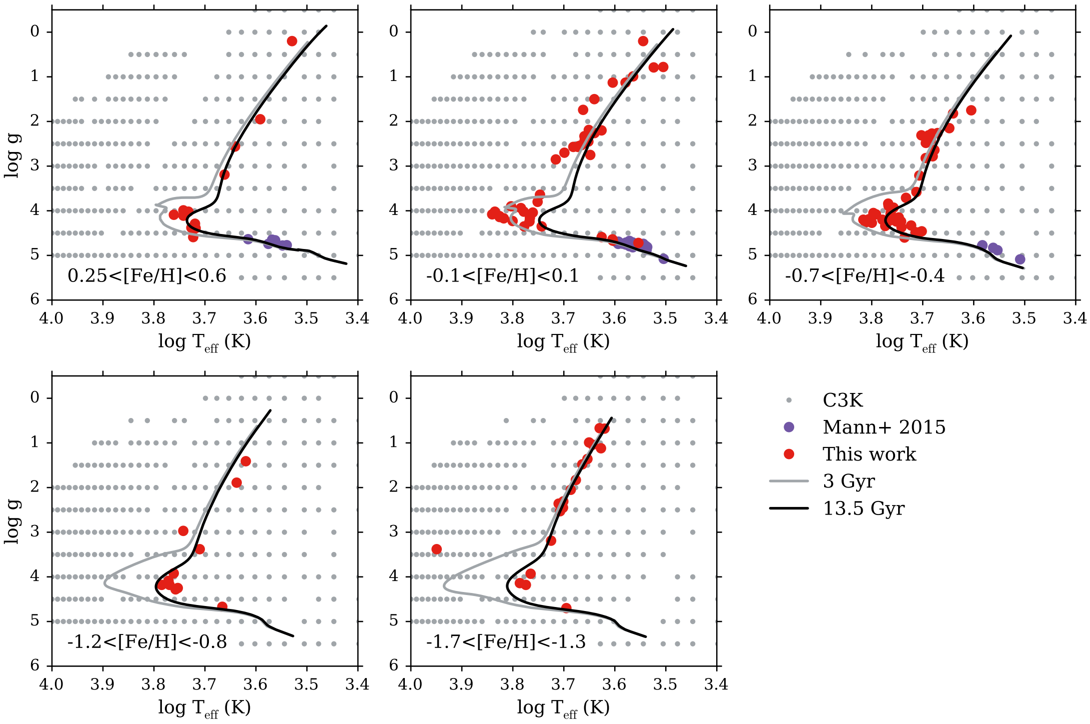

Figure 2.16: Location of all the stars included in the SPI training set in $T_{\rm eff}$ vs $\log g$ space including the stars from the Extended IRTF library (red circles), the M dwarfs from Mann et al. (2015) (purple circles), and the C3K theoretical spectra (grey dots). We only use the C3K theoretical spectra outside the convex hull determined by the empirical spectra. Isochrones are the same as in Figure 2.1.

2.3 Spectral Polynomial Interpolator

Interpolators have long been used in conjunction with stellar libraries. What began as "fitting functions" for specific indices (e.g., Gorgas et al. 1993; Worthey et al. 1994) has now evolved to full spectral interpolators (e.g., Vazdekis et al. 2003; Prugniel et al. 2011; Ness et al. 2015; Sharma et al. 2016; Dries et al. 2016). Broadly, interpolators are either "global", where a polynomial is fit to the input sample points, or "local", which essentially averages the nearby data e.g., the linear or spline interpolation is a weighted average of only the data closest in parameter space. Global interpolation is appropriate when the surface is smooth and can be approximated by a (relatively) simple function. The interpolators of Vazdekis et al. (2003) and Dries et al. (2016) are local interpolators while the interpolators described in Prugniel et al. (2011), Ness et al. (2015), and Sharma et al. (2016) are global.

In this same spirit, as a companion to the Extended IRTF library we created the Spectral Polynomial Interpolator (SPI [7]). With SPI we create a data-driven, global interpolator which we can use to retrieve a spectrum for a set of arbitrary stellar parameters. In this section we will describe the input to SPI, how we construct the model, the quality of interpolation, and example uses for SPI.

2.3.1 The Training Set

SPI works by fitting polynomial functions of $T_{\rm eff}$, log$g$, and [Fe/H] to the $L_{\rm bol}$-normalized flux of all the stars included in the training set. These fits are carried out independently of the fluxes at each wavelength. Then, for any set of stellar parameters,

---

[7] http://github.com/bd-j/spi

these polynomial functions can be used to predict or approximate the flux at a given wavelength.

The primary input into SPI is the IRTF Extended Library, limiting the sample to stars with computed bolometric luminosities (see section 2.4.1), high-quality spectra, and accurate stellar parameters (more on the latter two later;). The library stars that made the final cut to be included in the training set are shown as red points in Figure 2.16. We have augmented the SPI training set in a couple of ways, which we describe below. In total 194 Extended IRTF Library stars were included in the final SPI training set.

We include the wavelength dependent uncertainties for every star in the training set and weight the fluxes in the training sample by the corresponding uncertainty.

Additions to the Cool Dwarf Regime

We can see from Figure 2.1 that there is a paucity of cool dwarf stars in the MILES library and subsequently the library presented in this work. Having few cool dwarf stars would have hindered our ability to build an accurate polynomial model in this regime. We therefore included 76 M dwarf stars presented in Mann et al. (2015) to the SPI training set (purple points in Figure 2.16), excluding stars with low SNR. We also found that 4 stars had strong chromospheric Balmer emission lines which we removed from the training set.

The spectra from Mann et al. (2015) are a combination of SNIFS ($0.3-0.95\mu m$) and pre-detector upgrade SpeX ($0.8 - 2.4\mu m$). The resolution of the SNIFS data is coarser (R$\sim$ 1000) than the resolution of the MILES data. To put the SNIFS data on the same spectral resolution scale as the MILES data we used a high resolution (R∼10,000) version of C3K to deconvolve the SNIFS portion of the Mann et al. (2015) data to the resolution of the MILES data. We did this by producing a spectrum using the stellar parameters of a given star in the Mann et al. (2015) sample and made two models, one each for the MILES and SNIFS resolutions. We took the ratio of these two models, interpolated it onto the MILES wavelength grid and multiplied the Mann et al. (2015) spectrum by this correction factor. Visual comparison of a SNIFS and MILES star of the same stellar parameters led us to conclude that this procedure was effective.

We use the stellar parameters as presented in Mann et al. (2015) with surface gravity given by,

$$\log g = \log_{10}\left(\frac{6.6743\times10^{-8}\times M\times1.989\times10^{33}}{(R\times6.955\times10^{10})^2}\right), \tag{2.5}$$

where M and R represent the stellar mass and radius in solar units.

We note that while we use surface gravity to make the parameterization of the M dwarf stars consistent with the other regimes, surface gravity is not a commonly used metric in M dwarf research. M dwarfs do not age on a Hubble time and so surface gravity can be uniquely determined by effective temperature and metallicity.

Support for extrapolation

To preserve the quality of the interpolator at the edges of the empirical parameter space, we supplement the training set with spectra from the theoretical C3K library (grey points in Figure 2.16) degraded to the SpeX resolution. We only added C3K spec-

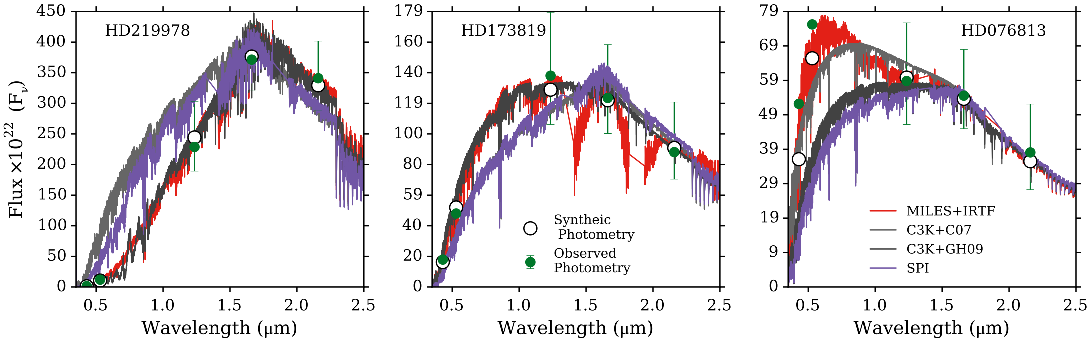

Figure 2.17: Comparison of data (spectroscopy in red and photometry as green circles) with the SPI model (purple) and theoretical spectra using two choices for effective temperature (grey lines). Synthetic photometry is shown for clarity (open symbols). For the dark grey line the temperature used was derived from the metallicity-color relations of González Hernández & Bonifacio (2009). The temperature used for the light grey line is the value from Cenarro et al. (2007). This shows the spread in T_{eff} values, and thus the spread in expected spectral shape, from different methods for each star. For the stars shown the spectral shape expected from the Prugniel et al. (2011) and Sharma et al. (2016) values is not consistent with the observed spectral shape.

tra for stellar parameters outside the convex hull (a hull being the set of planes that encloses all the training points) of the MILES stars. The C3K spectra are normalized in the same way as the observed spectra.

Our goal is for the C3K library to keep the polynomial terms "well-behaved" at the boundaries of parameter space for the empirical spectra, but we do not want the fits to be driven by the large numbers of C3K stars. We therefore weight the C3K spectra in the fits by a factor of 10^{-2} times the median SNR of the empirical spectra. This factor was chosen after considering a range of values and inspecting the resulting polynomial behavior.

2.3.2 Training the Model

The library contains a wide range of spectral types, with stellar temperature being the primary driver of the shape of the spectra. Modeling all the stars together is not feasible since it is difficult to specify a polynomial model that can encompass such a diverse set of stars. We therefore partition the library into five subsets based on temperature and surface gravity, essentially making five global interpolators that we use in conjunction with each other to span the entirety of parameter space. For each subset we define the polynomial terms we use to create the model ($t \equiv \log T_{\rm eff}$, $z \equiv$[Fe/H], $g \equiv \log g$):

- Cool Dwarfs

$$\begin{aligned}
\log F_\nu = {} & a_0 + a_1 t + a_2 z + a_3 g + a_4 z^2 + a_5 t^2 + a_6 g^2 + \\
& a_7 (t \times z) + a_8 (t \times g) + a_9 (z \times g) + a_{10} z^3 + \\
& a_{11} t^3 + a_{12} g^3 + a_{13} (t^2 \times z) + a_{14} (z^2 t) + \\
& a_{15} (g \times t^2) + a_{16} t^4 + a_{17} z^4 + a_{18} (t^2 \times z^2) + \\
& a_{19} (t^3 \times z) + a_{20} t^5 .
\end{aligned} \tag{2.6}$$

- Cool Giants

$$\begin{aligned}\log F_\nu = {} & a_0 + a_1 t + a_2 z + a_3 g + a_4 t^2 + \\ & a_5 g^2 + a_6 z^2 + a_7 (z \times g) + a_8 (t \times g) + \\ & a_8 (t \times z) + a_9 t^3 + a_{10} g^3 + a_{11} z^3 + \\ & a_{12} (t \times g \times z) + a_{13} (t \times t \times z) + \\ & a_{14} (t \times t \times g) + a_{15} (z^2 \times t) + a_{16} (z^2 \times g) + \\ & a_{17} (t \times g^2) + a_{18} (z \times g^2) + a_{18} t^4 .\end{aligned} \tag{2.7}$$

- Warm Dwarfs

$$\begin{aligned}\log F_\nu = {} & a_0 + a_1 t + a_2 z + a_3 g + a_4 t^2 + a_5 g^2 + a_6 z^2 + \\ & a_7 (t \times z) + a_8 (t \times g) + a_9 t^3 + a_{10} (t \times g^2) + a_{11} z^3 + \\ & a_{12} (t^2 \times g) + a_{13} (t^2 \times z) + a_{13} (t \times z^2) + \\ & a_{14} (t \times g \times z) + a_{15} (g \times z^2) + a_{16} t^4 + a_{17} g^4 + \\ & a_{18} (t^3 \times g) + a_{20} (z \times t^3) + a_{21} (z^2 \times t^2) + \\ & a_{21} (z^3 \times t) + a_{22} (t^2 \times g)^2) + a_{23} (z \times t^2 \times g) + \\ & a_{24} t^5 .\end{aligned} \tag{2.8}$$

- Warm Giants

$$
\begin{aligned}
\log F_\nu = a_0 + a_1 t + a_2 z + a_3 g + a_4 t^2 + a_5 g^2 + a_6 z^2 + \\
a_7(t \times z) + a_8(t \times g) + a_9(g \times z) + a_9 t^3 + a_{10} g^3 + \\
a_{11} z^3 + a_{12}(t^2 \times z) + a_{13}(t \times z^2) + a_{14}(g \times t^2) + \\
a_{15}(g^2 \times t) + a_{16} t^4 + a_{17} z^4 + a_{18}(t^2 \times z^2) + \\
a_{19}(t^2 \times g^2) + a_{20}(z^2 \times g^2) + a_{21} t^5.
\end{aligned} \tag{2.9}
$$

- Hot Stars

$$
\begin{aligned}
\log F_\nu = a_0 + a_1 t + a_2 z + a_3 g + a_4 t^2 + a_5 z^2 + a_6 g^2 + \\
a_7(t \times g) + a_8(t \times z) + a_9(g \times z) + a_{10} t^3 + a_{11} g^3 + \\
a_{12} z^3 + a_{13}(t \times g \times z) + a_{14}(t^2 \times z) + a_{15}(t^2 \times g) + \\
a_{16}(z^2 \times g) + a_{17}(t \times g^2) + a_{18}(z \times g^2) + a_{19} t^4.
\end{aligned} \tag{2.10}
$$

The maximum likelihood coefficients for each regime are determined by weighted linear least squares. The glut of polynomial terms is a classic problem in polynomial regression modeling. As we are concerned with the ability to predict spectra (which we assess later in Section 2.3.3) rather than the values of the coefficients themselves we do not make an effort to simplify the polynomial functions using, e.g., L1 regularization. Furthermore, the oscillatory behavior of extrapolations that can result from unregularized high order polynomial regression is mitigated by our use of theoretical spectra near the boundaries of the valid parameter space.

Table 2.1: Valid ranges for SPI interpolation

Regime	Training T_{eff}	log$g$	[Fe/H]	Interpolating T_{eff}	log$g$	[Fe/H]
Cool Dwarfs	1100-5500	3.5-6.5	-2.5-0.6	2500-4000	-0.5-3.5	-2.0-0.6
Cool Giants	1100-4500	-1.0-2.75	-2.5-0.6	2500-4000	>3.5	-2.0-0.6
Warm Dwarfs	3000-6500	3.0-5.75	-2.5-0.6	4000-6000	-0.5-3.5	-2.0-0.6
Warm Giants	3500-6500	-0.75-4.0	-2.5-0.6	4000-6000	>3.5	-2.0-0.6
Hot Stars	5500-12,500	2.5-5.5	-2.5-0.6	6000-12,000	<5.0	-2.0-0.6

In Table 2.1 we show the "training" ranges for T_{eff}, log$g$, and [Fe/H], i.e., the stellar parameter limits for the training set stars in each regime and the "Interpolating" ranges for T_{eff}, log$g$, and [Fe/H], i.e., the stellar parameter limits for safe interpolation. We have the same range in metallicity for the training bounds of all the stellar regimes. The ranges were determined by minimizing the rms difference between the observed and interpolated spectra in the training set. The overlap in the T_{eff} training ranges is meant to mitigate the effect of five separate, disjoint hulls having a smaller volume than the hull for all the library points. For most regimes the overlap in $\mathrm{T_{eff}}$ is 500 K but for the cool dwarf regime we extend the training sample to 5500 K. We do this to compensate for the paucity of low-metallicity cool dwarfs in the empirical library. If we did not extend the training sample to 5500 K the metallicity dependence for the cool dwarf regime would end at [Fe/H] $\sim$ -0.7, which would bias the interpolation to lower metallicities. With the extended temperature range of the training sample SPI is able to pull information from the warmer low-metallicity stars which mitigates the effects of the lack of observed low-metallicity cool dwarfs.

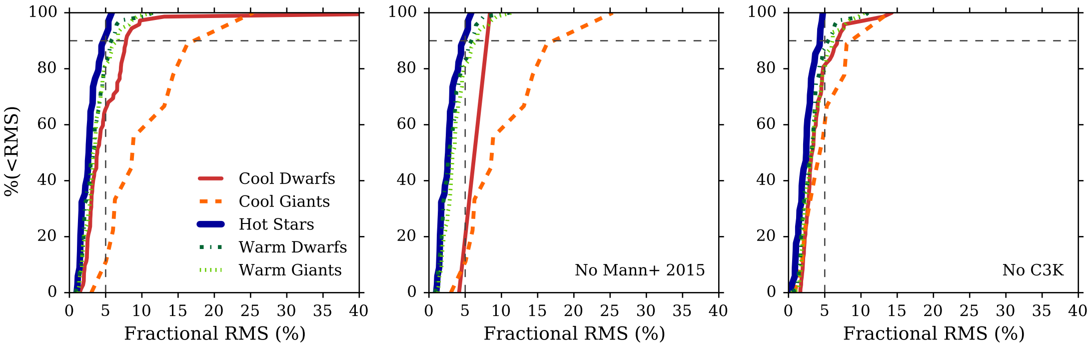

Figure 2.18: Cumulative distribution functions of the fractional rms differences between the observed spectra and interpolated spectra in the cool dwarf (red), cool giant (orange), hot (blue), warm dwarf (green), and warm giant (lime) regimes. To aid interpretation, a fractional rms difference of 5% is marked with a vertical dashed line and 90% of the sample is marked with a horizontal dashed line.

Culling the Extended IRTF Library

Only the highest quality spectra should be included in the SPI training set. Any star used in the training set needs to have accurate stellar parameters and a reliably flux-calibrated observed spectrum. Based on visual inspection we removed stars with spectra that were either of poor quality or appeared to have flux calibration issues that severely affected the shapes. We removed 51 stars following this criteria.

Furthermore, since the interpolation relies on having like-spectra grouped by like-stellar parameters, we need to ensure that our stellar parameters are accurate. SPI provides an opportunity to check the accuracy of the stellar parameters associated with each of the stars in the library. If there is significant discrepancy between the observed spectrum of the star and the interpolated spectrum given by SPI it might be indicative of an issue with the stellar parameters associated with that star. A similar technique was used by Vazdekis et al. (2010) to remove 60 stars from the MILES library in creating

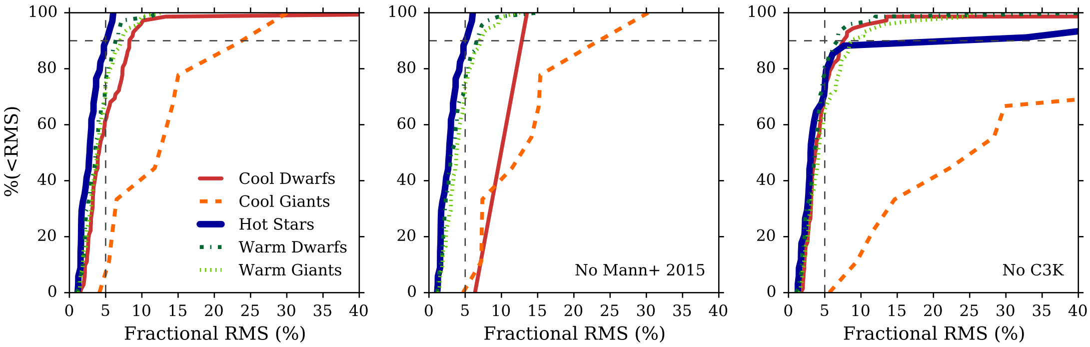

Figure 2.19: Same as Figure 2.18 but now the interpolated spectra are the result of the leave-one-out ("jack knife") test where each star in the training sample was removed from the model in turn before the interpolation. Including the C3K spectra reduces the dependence of the model on the presence of any one star, especially the cool giant stars and hot stars. The inclusion of the Mann et al. (2015) M dwarfs also helps mitigate issues in the cool dwarf regime.

stellar population models.

In Figure 2.17 we show an example of three stars where the SPI predicted spectra (purple) using the stellar parameters from Prugniel et al. (2011) and Sharma et al. (2016) are not consistent with the observed data (both spectroscopy, red line, and photometry, green circles). Consistency between the observed photometry and observed spectra for these stars suggests that the problem is not exclusively with the flux calibration but with the stellar parameters. To further emphasize this point we also show two C3K spectra corresponding to different effective temperatures in each panel. For the dark grey spectra we used the effective temperatures computed using the González Hernández & Bonifacio (2009) $J-K_S$-metallicity relations. For the light grey spectra we used the effective temperatures from Cenarro et al. (2007). For HD219978 (left) and HD173819 (middle) the overall shape of the spectra using the $J-K_S$ derived temperatures are more consistent with the observed data than SPI and the spectra corresponding to the Cenarro et al. (2007) spectra. For HD076813 (right) the spectrum corresponding to the temperature from Cenarro et al. (2007) is most consistent.

This test both demonstrates the possible range in derived stellar parameter values and that there are available stellar parameters that would better match the observed spectra and photometry and thus the problem is with the Prugniel et al. (2011) and Sharma et al. (2016) stellar parameters. These stars were flagged for having an exceptionally high rms difference between the observed and SPI predicted spectra. We flagged 9% of our observed sample as having incorrect stellar parameters and do not include them in the SPI training set. The differences in the temperatures determined using different methods for these stars are large compared to the bulk of the library stars. This indicates that in general the stellar parameters from Prugniel et al. (2011) and Sharma et al. (2016) are consistent with the observations.

This test does not mean that any issue with the data is exclusively an issue with the stellar parameters. Several of the stars flagged as potentially having incorrect stellar parameters also have flux calibration issues or other quality issues as indicated by their Quality and Shape Flags (see Table A1). For example, we still see the effects of prominent telluric absorption for HD173819 (the middle) panel which is indicated in Table A1. However, the SPI predicted spectrum is not consistent with the shape of the observed spectrum but the predictions based on the Cenarro et al. (2007) temperature and temperature from the González Hernández & Bonifacio (2009) relations are consistent with the observed shape.

Thus, we can use SPI as a way to flag stars that need more accurate stellar parameters but otherwise have good data (in the future we will use SPI to re-derive stellar parameters in an iterative approach). We removed 27 stars following this criteria and re-derived the SPI parameters with these stars removed.

2.3.3 Quality of Interpolation

We can assess the quality of the interpolation by examining how well SPI can recover the spectra of the stars in the training sample (including both the data presented in this work and the Mann et al. (2015) data). For each star we compute the fractional rms between an observed spectrum in the empirical training set and the corresponding interpolated spectra from SPI. In Figure 2.18 we show the cumulative distribution of this fractional rms. In each panel we show results separately for the cool dwarf stars (red), cool giant stars (orange), hot stars (blue), warm dwarf stars (dark green), and warm giant stars (light green). Also plotted in each panel is a horizontal line showing where 90% of the stars are placed on the cumulative distribution and a vertical line marking 5% fractional rms. In the left panel we show the distributions when using the complete training set in the model, in the middle panel we show the distributions when we exclude the Mann et al. (2015) spectra from the training set, and in the right panel we show the distributions that result when we exclude the C3K spectra from the training set. Note that the rms is insensitive to overall offsets between the two spectra. Note that SPI is not an interpolator in the strictest sense of exactly reproducing the input spectrum at the input points, and so there is no guarantee that the rms should be small.

The most important take away from Figure 2.18 is that for all the regimes the recovery of the training set spectra is very good. For cool dwarfs, warm dwarfs, warm giants, and the hot stars the recovery is better than 10% for 90% of the stars. When we exclude the Mann et al. (2015) M dwarfs from the training set the recovery for the cool dwarfs is worse. The recovery of the cool giants is worse than the other regimes and counterintuitively improves when we exclude the C3K spectra from the training set. This could be an effect of the problems theoretical spectra have in the cool giant regime (see Bertone et al. 2008).

Since SPI relies on stars with similar stellar parameters having similar stellar spectra we can look at stars where SPI fails to recreate its spectrum as stars with possible issues with the stellar parameters. Since we see in Figure 2.18 that for most of the library ($\sim 90\%$) the rms difference between the observed and interpolated spectra is $< 5\%$ we can feel confident that the stellar parameters from Prugniel et al. (2011) and Sharma et al. (2016) are internally self-consistent. We looked at Figure 2.18 separately for the MILES and IRTF spectra and found that the recovery is about the same for the majority of the stars.

Figure 2.19 is the same as Figure 2.18 but now the interpolated spectra are the result of leave-one-out validation. The leave-one-out test consists of going through all the stars in training sample, taking one out of the training sample at a time, retraining the function such that the information from that star is no longer included in the model, then comparing the SPI prediction for that spectrum to the actual spectrum.

The leave-one-out test is an assessment of how sensitive the model is to the presence of any one star and demonstrates the utility of including the C3K spectra in

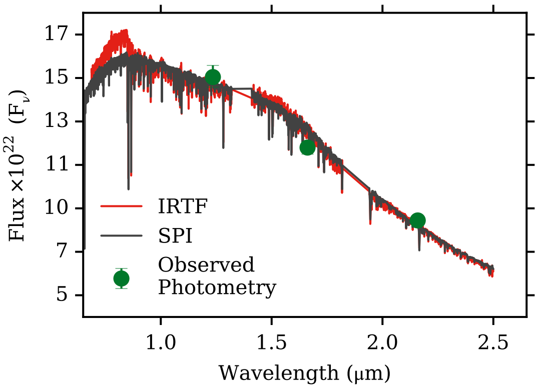

Figure 2.20: Demonstration of the “self-calibration” possibilities with SPI. The observed spectrum (red) for the star, HD004307, has an unphysical artifact at $\sim 0.8\mu m$ due to nebulosity near the standard star. Since this issue only affected a small subset, 4%, of the stellar library we can use SPI to obtain a spectrum for HD004307 and others like it without the unphysical feature. This is shown in the interpolated spectrum (black) that is largely the same as the observed spectrum but without the bump at $\sim 0.8\mu m$.

the training set. This figure shows that by including the C3K spectra in the training set we are mitigating the bias a single star can introduce in the model, especially for the cool giant stars and hot stars, where the number of empirical stars is low. We see that the Mann et al. (2015) data has a similar effect in the cool dwarf star regime. This figure demonstrates that the inclusion of the Mann et al. (2015) and C3K spectra is important for the interpolation of stars outside the confines of the training set.

2.3.4 Applications of SPI

As discussed previously, SPI can be used to flag stars with potentially inaccurate stellar parameters. More generally, stars that are outliers with respect to their SPI prediction could be used to uncover other unusual behavior such as variability or peculiar abundance patterns.

SPI also allows for “self-calibration” of the observations. By this we mean that because SPI uses the information of all the stars in the library it is possible to remove artifacts that affect a small fraction of the library stars. As we mentioned earlier, some of the stars observed for this library have an unphysical bump from standard stars with unusual spectra. Since most stars were not affected in this way we can remove this artifact by using the model to interpolate for a spectrum for the stars that are affected. In Figure 2.20 we show the observed spectrum (red) for the star HD004307 and the interpolated spectrum (black). The interpolated spectrum is consistent with the observed spectrum except for the bump seen at 0.8 μm in the observed spectrum, where the interpolated spectrum produces more sensible behavior. In future work we will use the self-calibration capabilities of SPI to re-derive the stellar parameters for the stars presented in this work.

We emphasize that any library can be used as input for SPI. The new metallicity coverage of the Extended IRTF Library allows us to interpolate to a larger metallicity range than is possible with the original IRTF library (Rayner et al. 2009). Furthermore, the library presented in this work can be augmented with, for example, NGSL. The inclusion of UV data into SPI could potentially help with modeling planet atmospheric properties.

2.4 Behavior of the Stellar Libraries

The main feature of the Extended IRTF Library is the expansion of near-IR coverage into the sub- and super-metallicity regimes. We would like to highlight various spectral features in the data and explore how these features depend on not just surface gravity and effective temperature but also on metallicity. This is also an opportunity to examine the behavior of SPI beyond its ability to simply reproduce the training set spectra. Also of interest is how the empirical trends compare with the theoretical models. To explore the behavior of SPI and C3K we computed spectra for stellar parameters along a 3 Gyr (for metallicities > -0.7) and a 13.5 Gyr (for metallicities < -0.7) MIST isochrone.

Sharp boundaries where the five different polynomial models that make up SPI join together is a concern. To ensure smoothness we took the weighted average of the different predicted fluxes for the evolutionary points (EPs) that have temperatures that are in the overlap between the cool and warm dwarf and the cool and warm giant training bounds. This means that for EPs with $\log g > 4.0$ and temperatures between 3500-4500 K we generated a flux using both the warm giant and dwarf giant models and averaged the fluxes together weighted depending on the temperature. We did the same for the EPs with $\log g < 4.0$ and temperatures between 3000-5500 K.

In this section we analyze the metallicity-dependence of key stellar features by using equivalent widths. The limitations of equivalent widths are well known – they are sensitive to the definition of the psuedocontinuum and each index is a blend of features from more than one element. Here we use equivalent widths as a way to compress the information to explore broad trends. All the equivalent widths quoted in this work are in units of Å and all wavelengths are in vacuum. We present a combination of a selection of the Lick indices defined in Table 1 Worthey et al. (1994) and Table 1 of Conroy & van Dokkum (2012a).

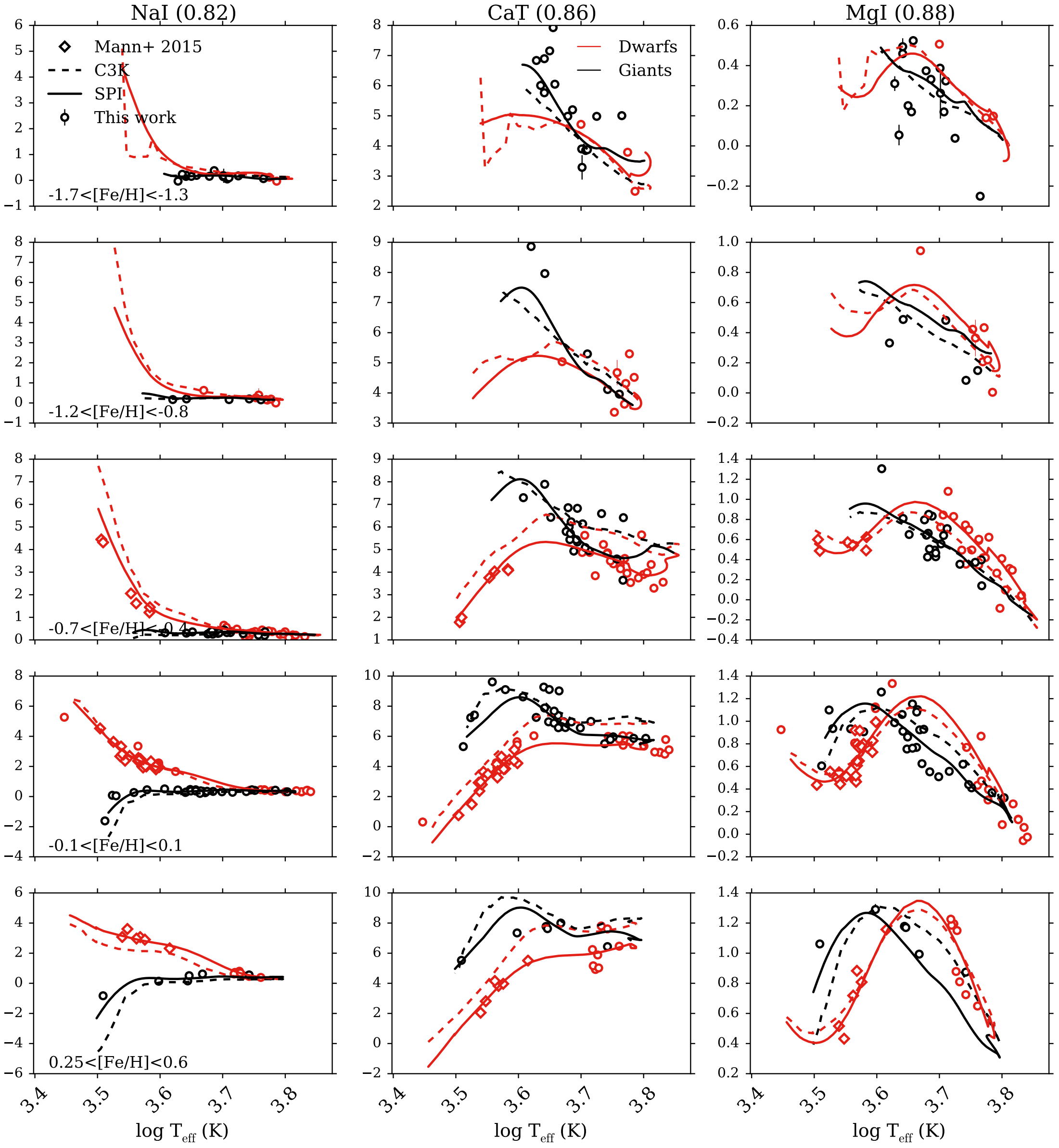

Figure 2.21: Dependence of selected spectral indices on effective temperature for dwarfs ($\log g > 4.0$, red) and giants ($\log g \leq 4.0$, black). Plotted are index strengths using the IRTF data from this work (open circles), data from Mann et al. (2015) (open diamonds), empirical prediction from SPI (solid lines), and theoretical predictions from C3K (dashed lines).

We computed equivalent widths using the following equation,

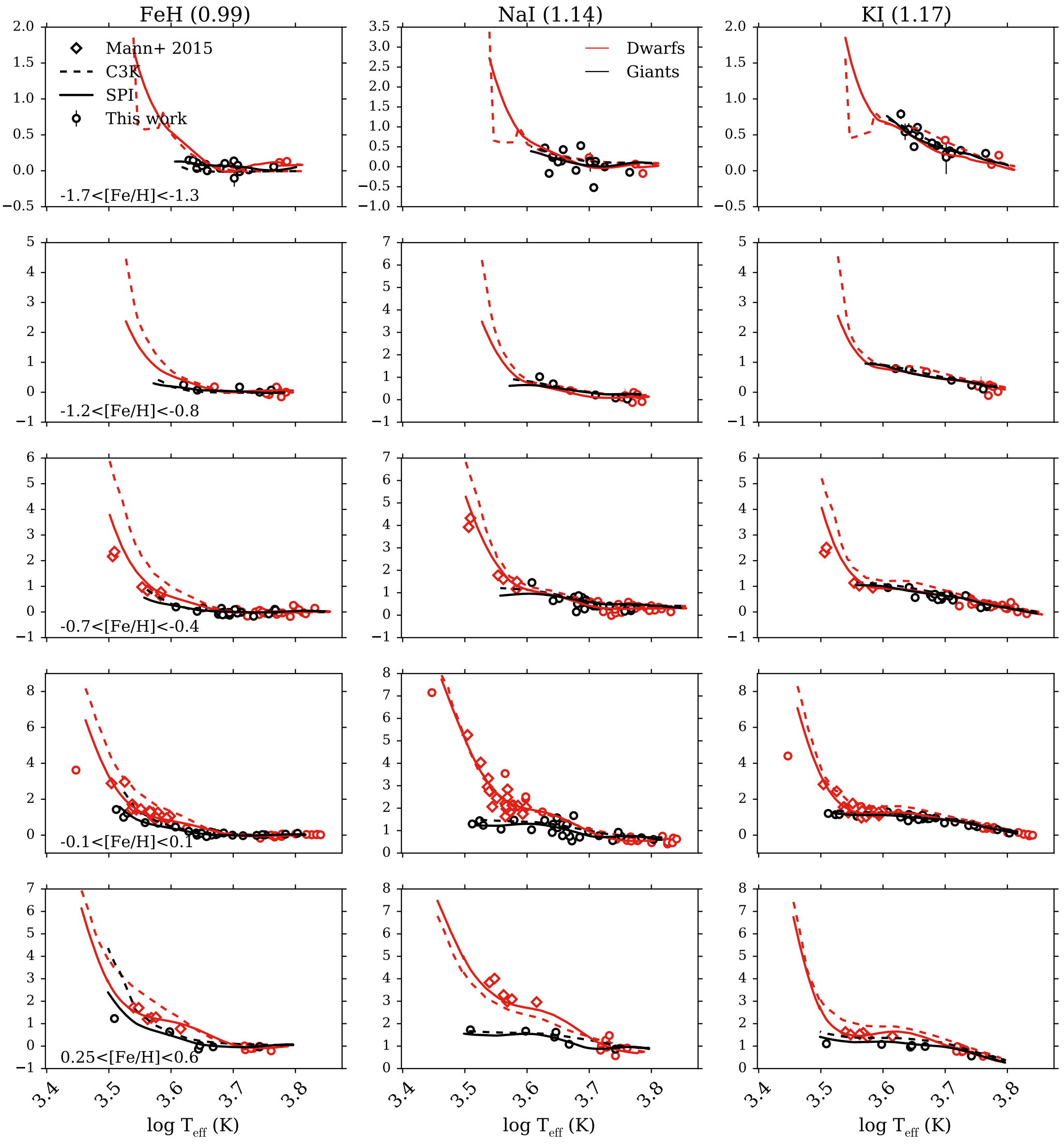

Figure 2.22: Continuation of Figure 2.24.

$$\mathrm{EW} = \lambda_2 - \lambda_1 - \int_{\lambda_1}^{\lambda_2} \frac{F_i}{c_b + (\lambda_i - \lambda_b)\left(\frac{c_r - c_b}{\lambda_r - \lambda_b}\right)} d\lambda, \qquad (2.11)$$

where λ_1 and λ_2 are the blue and red wavelength boundaries of the feature definition, λ_b and λ_r are the average wavelengths of the blue and red continuum definitions for

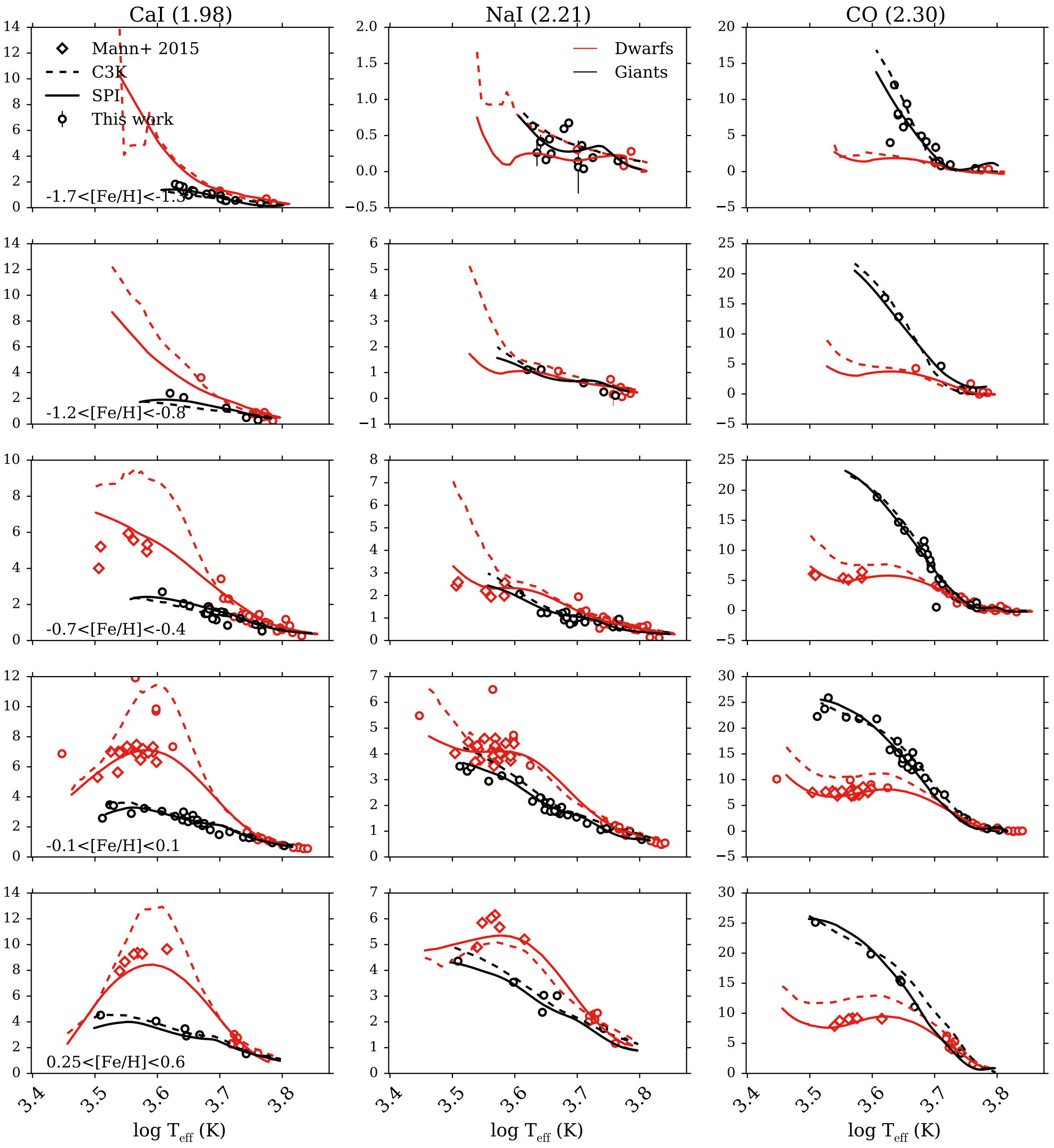

Figure 2.23: Continuation of Figure 2.21.

each feature. The blue and red continuum values c_b and c_r are the integral of the flux over the wavelength range that defines the blue and red continuum.

For the observations, we made 100 realizations of each spectrum by Monte Carlo sampling the noise. For each realization we used Equation 2.11 to compute the

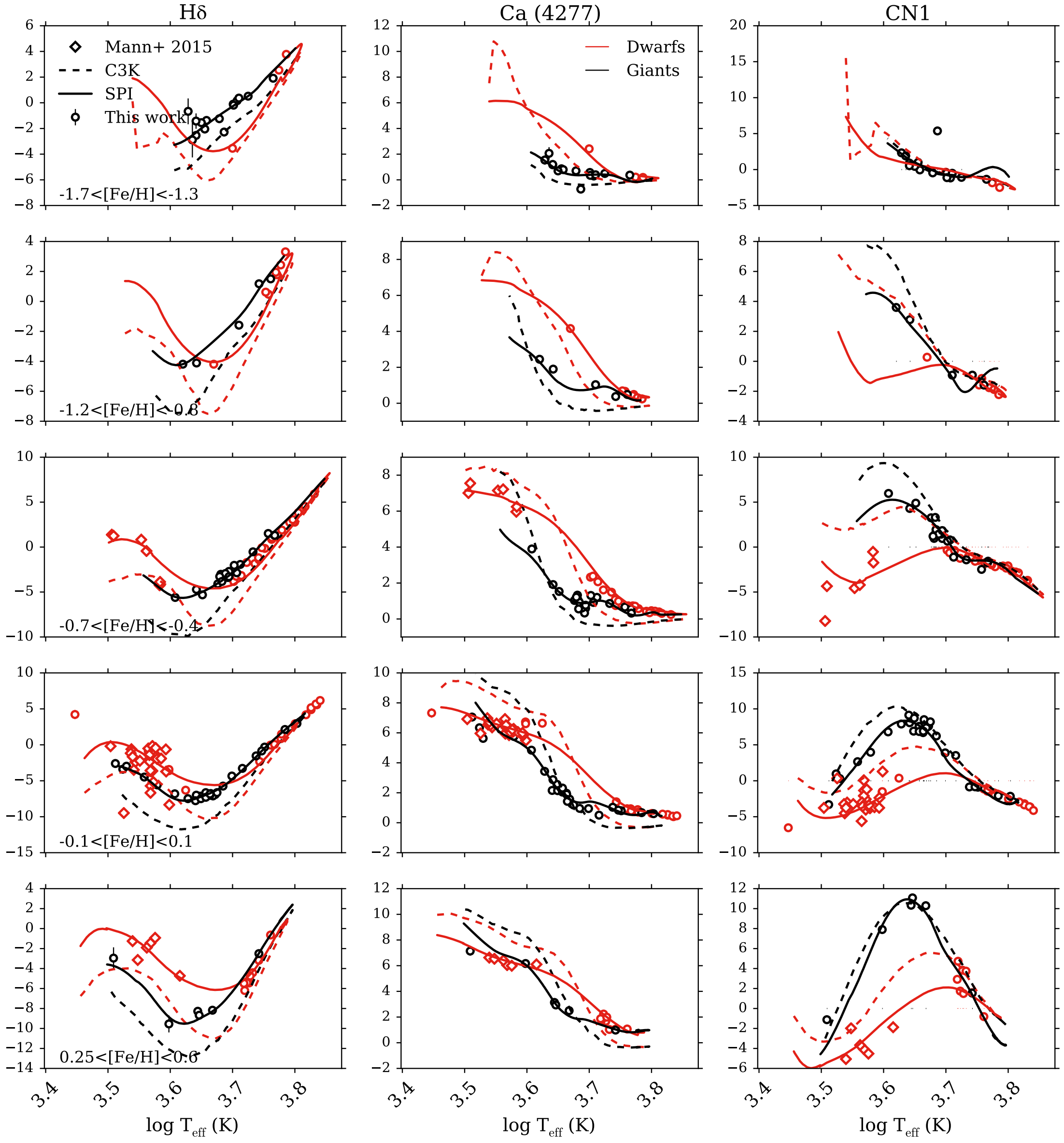

Figure 2.24: Continuation of Figure 2.21 except now using the corresponding MILES spectra for the stars in the Extended IRTF Library..

equivalent width. The errors for the equivalent widths are given by the 1σ confidence values from the distribution of equivalent widths.

In Figures 2.21-2.29 we show equivalent width strength versus effective temperature for different spectral features (columns) and different metallicity bins (rows).

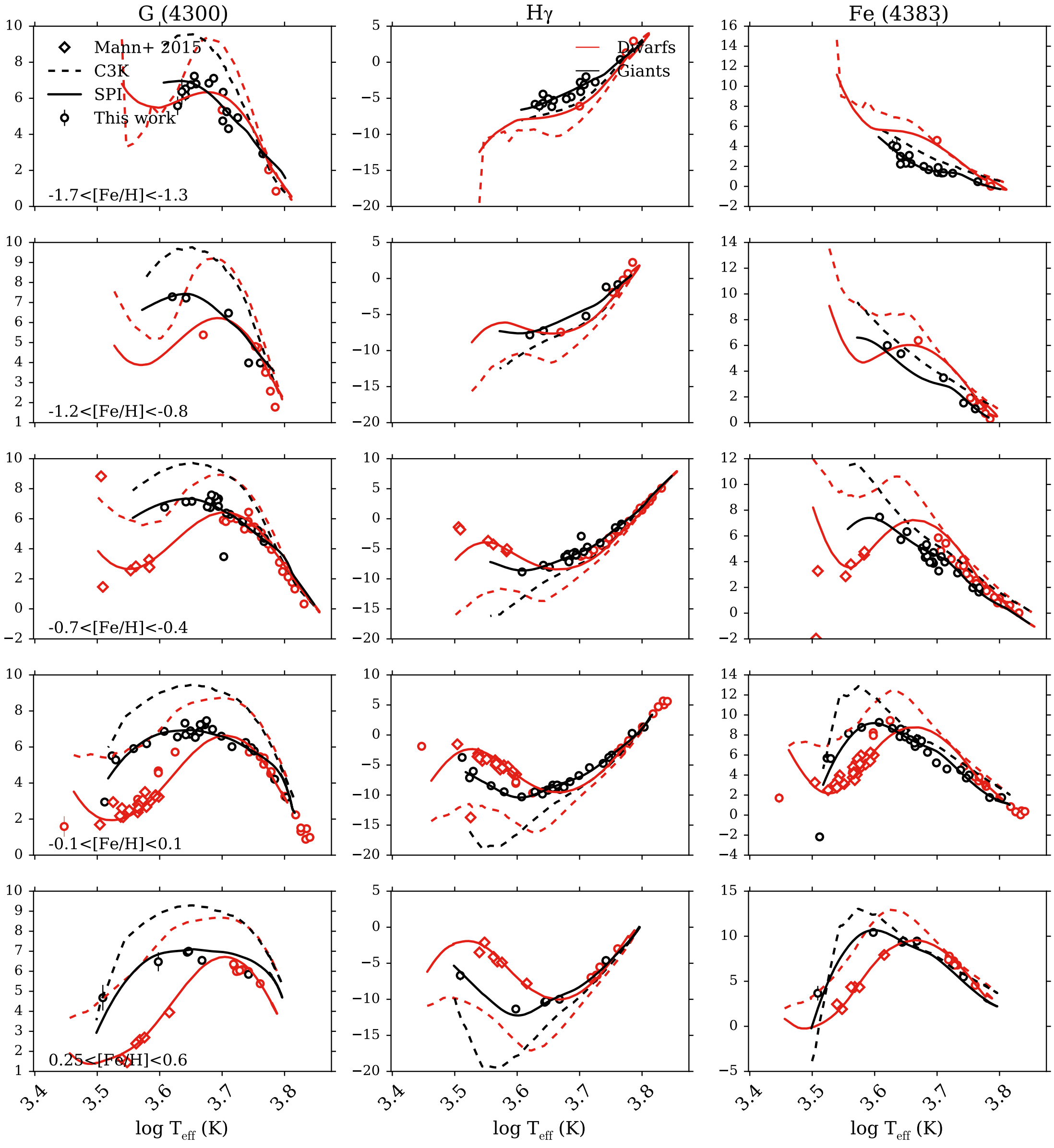

Figure 2.25: Continuation of Figure 2.24.

In every panel dwarf stars (log$g$ > 4.0, red) are distinguished from giant stars (log$g$ ≤ 4.0, black). The equivalent widths for the stars in the Extended IRTF Library are shown in the empty circles, the equivalent widths from the Mann et al. (2015) data are shown in empty diamonds, the SPI equivalent widths are shown as the solid lines, and

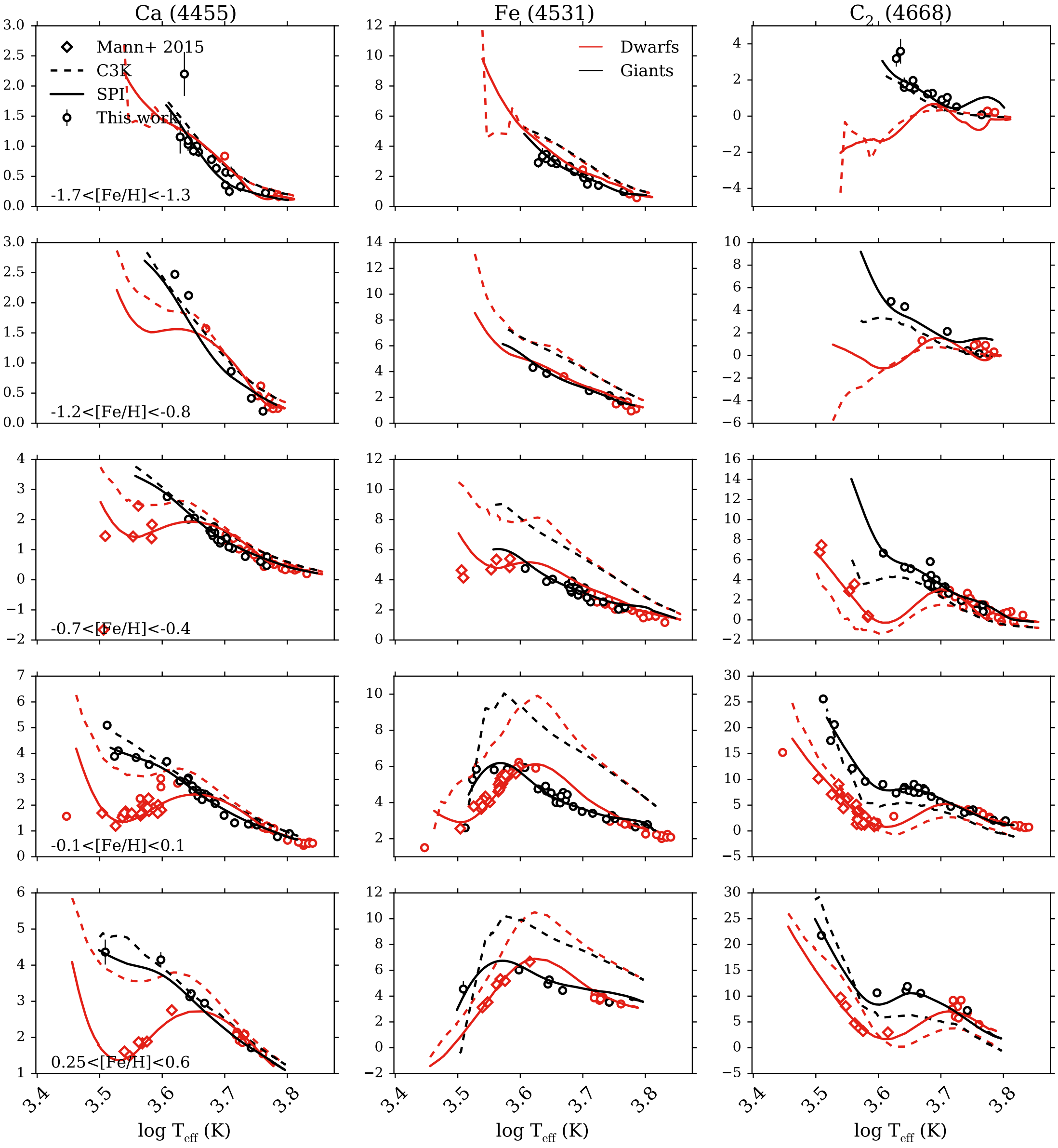

Figure 2.26: Continuation of Figure 2.24.

the C3K equivalent widths are shown in the dashed line. All data points have error bars though in most cases the error bars are smaller than the symbols. We do not expect a perfect match between the lines and the data points especially for the hot effective temperatures where there can be a range of log g for fixed T_{eff} in the data.

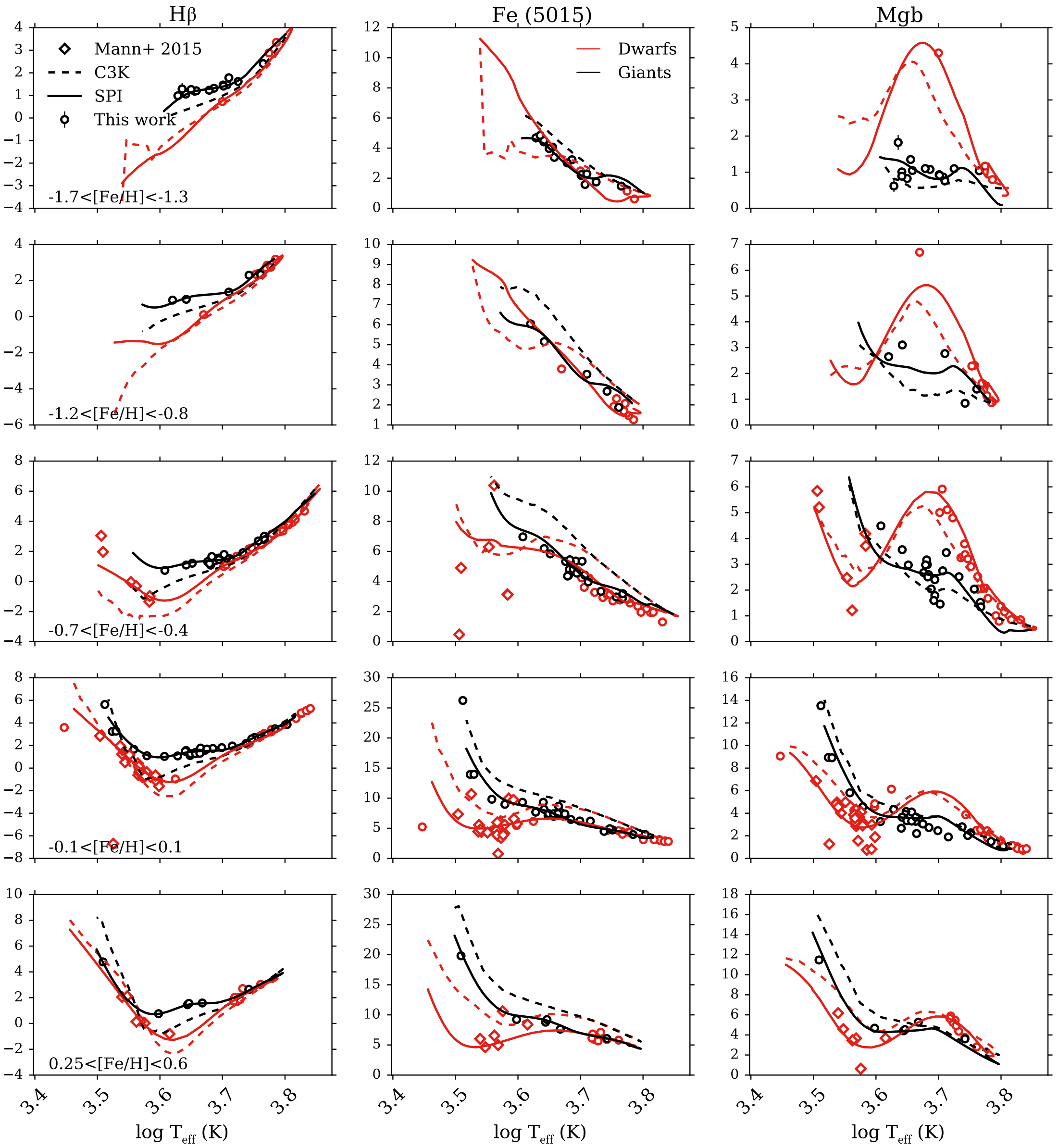

Figure 2.27: Continuation of Figure 2.24.

2.4.1 Data and empirical trends

In this section we focus on the observations and the SPI predictions. The theoretical predictions will be discussed in the following section.

Several spectral features were highlighted in Conroy & van Dokkum (2012a)

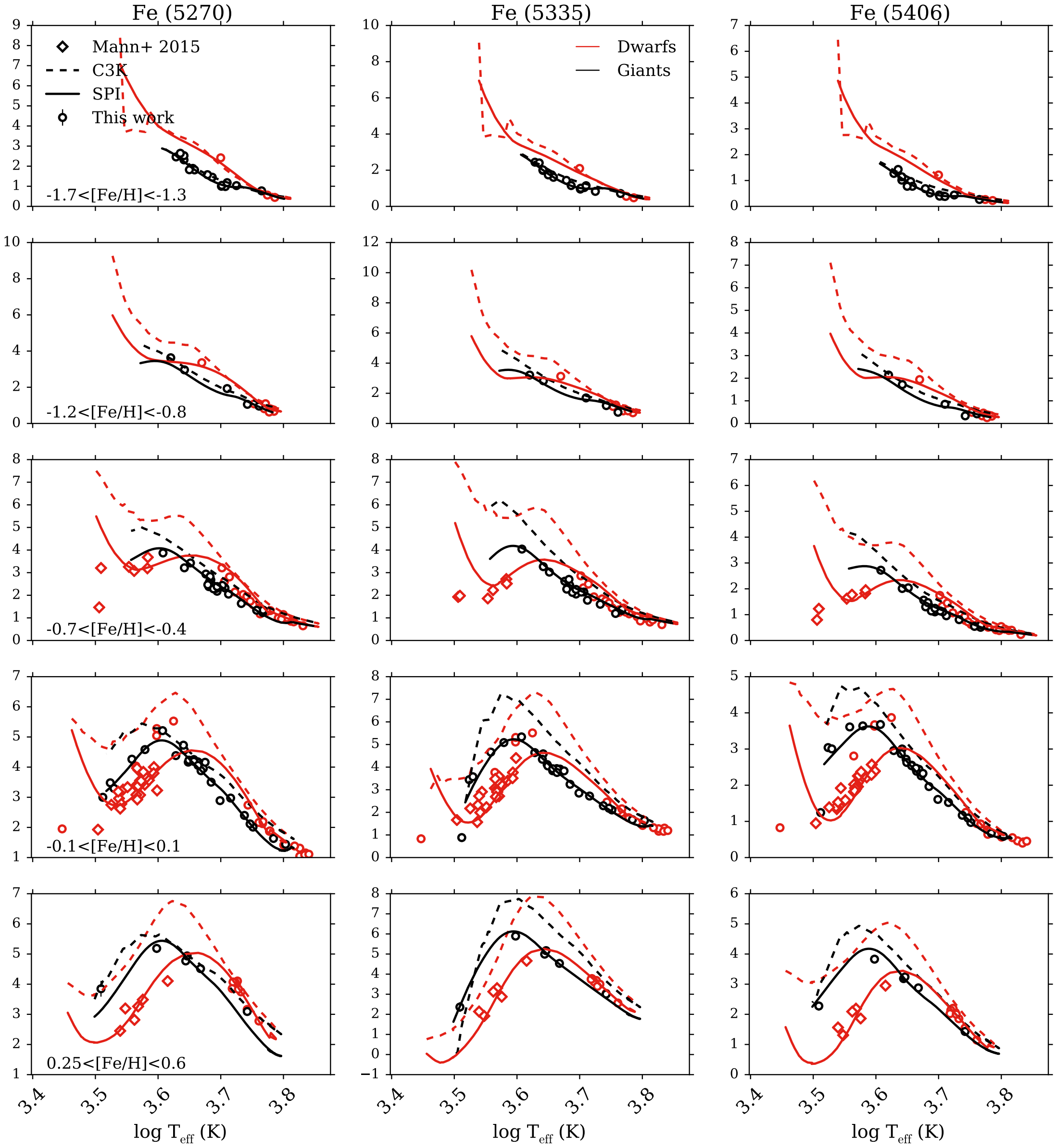

Figure 2.28: Continuation of Figure 2.24.

as means to discriminate between dwarf and giant stars: NaI0.82μm, CaT, FeH0.99μm KI1.17μm, CaI1.98μm, and CO2.30μm (see also, e.g., Spinrad 1962; Wing & Ford 1969; Cohen 1978; Frogel et al. 1978; Kleinmann & Hall 1986; Diaz et al. 1989; Ivanov et al. 2004; Rayner et al. 2009). This ability was only assessed for solar metallicity stars and

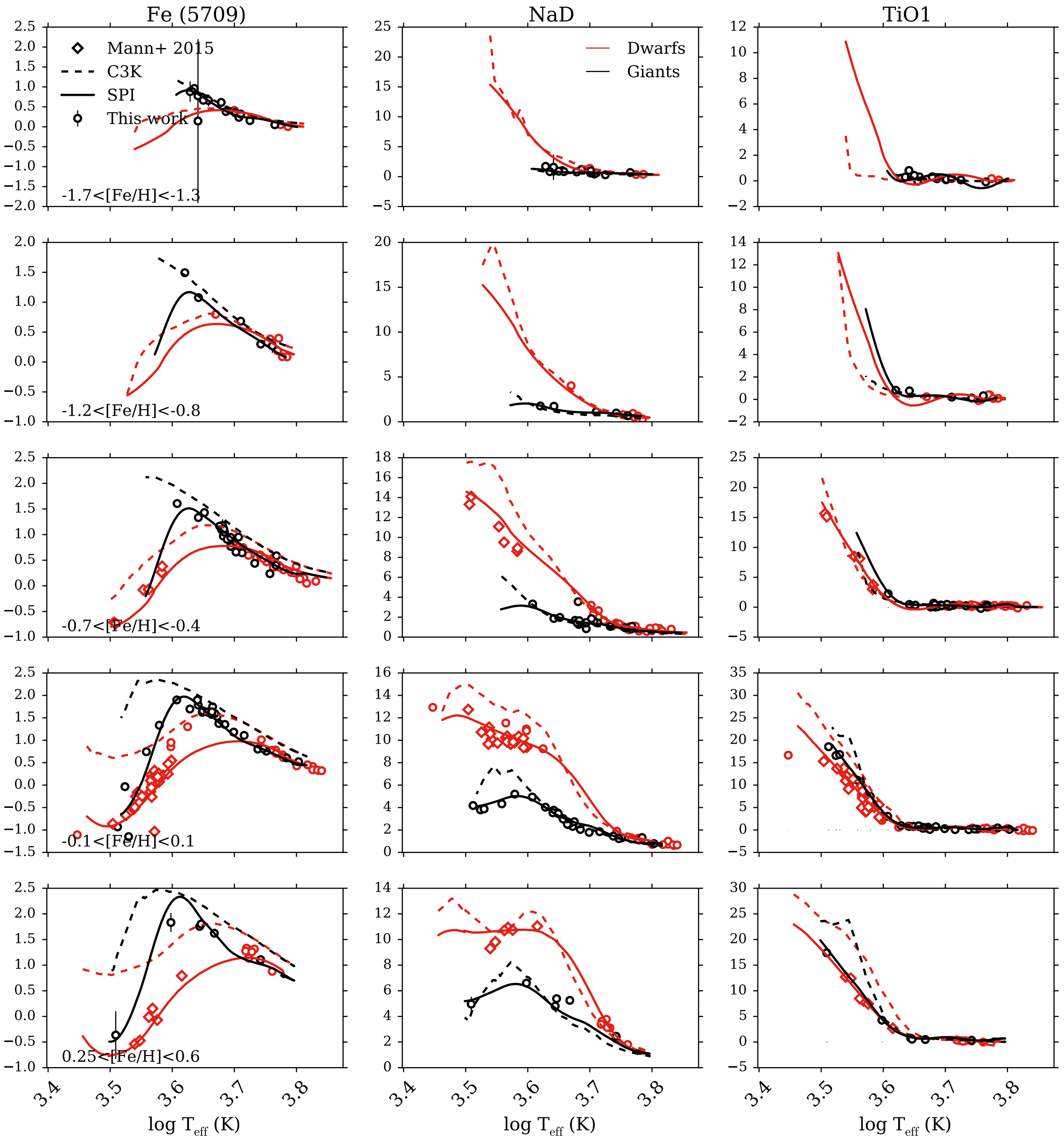

Figure 2.29: Continuation of Figure 2.21.

it is of interest to know whether this behavior holds for an extended metallicity range.

In Figures 2.21-2.23 we show the temperature dependent equivalent width trends for several near-IR spectral features separated by dwarfs and giants. For the three highest-metallicity bins we can see the clear separation between dwarf and giant stars from the data alone. In all of these features there is a metallicity dependence on the temperature range over which the separation between dwarf and giant stars occurs. All of the sodium lines, NaD, NaI0.82μm, NaI1.14μm, and NaI2.21μm have equivalent widths that increase among the dwarf stars with decreasing temperature. For the dwarf stars in the $-0.7 <$ [Fe/H] < -0.4 this is a precipitous increase for the coolest dwarf stars while at high metallicity there is a steadier increase over large range in temperature.

The CO2.30μm spectral feature remains a clear discriminator between giant and dwarf stars for the whole metallicity range and the temperature at which the separation occurs does not qualitatively change over the metallicity range. Likewise, the overall temperature dependent trend does not appear to change significantly for the KI1.17μm and CaT features. The strong FeH0.99μm feature remains *unique* to the cool dwarf stars for the entire metallicity range.

In addition to the near-IR features we show some of the Lick indices (Worthey et al. 1994) (Figures 2.24-2.29) that are the classic indicators of stellar population characteristics (e.g, age and metallicity). For some spectral features there is a noticeable difference in the temperature-dependent equivalent width trends from metallicity bin to metallicity bin (e.g., Fe4383, 4531, 5015, 5335, 5270, 5406, and 5782Å, Ca4455Å, Hβ, C4668Å). However, other spectral features ($H\gamma$, Hδ, Ca4277Å, G4300Å, and MgI0.88μm) the temperature dependent trends remain similar from metallicity bin to metallicity bin.

As we would expect from the quality assessment given in Section 2.3.3, SPI is consistent with the behavior of the data. Even where there is sparse data, e.g., the low-metallicity warm dwarf regime, SPI appears to make reasonable predictions of the behavior of all the spectral features displayed in Figures 2.21 - 2.29.

As shown in Figure 2.16 there are no dwarf stars cooler than 5000 K for $-1.7 <$ [Fe/H] < -0.8. This makes it impossible to know how accurate the SPI predicted spectra in this regime are. However, as described earlier the training and interpolating ranges are optimized so that SPI is able to use all three stellar parameter dimensions to make a prediction for this regime. That is, SPI is able to pull information from the predictions of both the hot and warm dwarf stars with $-1.7 <$ [Fe/H] < -0.8 and the higher metallicity cool dwarf stars. Indeed, the cool dwarfs in the lowest metallicity bins have temperature dependent trends that appear consistent with the trends in the higher metallicity regimes.

2.4.2 Comparison with theoretical trends

We now turn to a comparison between the theoretical C3K models and the empirical trends. Comparisons between models and observations have been made previously. Martins & Coelho (2007) compared the ATLAS9 (Castelli & Kurucz 2003), MARCS (Gustafsson et al. 2003), and PHOENIX (Brott & Hauschildt 2005) theoretical stellar libraries to various empirical stellar libraries. Bertone et al. (2008) compared the ATLAS9 library to the ELODIE catalog (Prugniel & Soubiran 2001), and Conroy & van Dokkum (2012a) compared equivalent width trends between theoretical and empirical stellar libraries. However, there are limitations to these studies that we can now overcome. Martins & Coelho (2007) and Conroy & van Dokkum (2012a) were limited to solar metallicity stars and could not reach the coolest M dwarfs. The sample from Bertone et al. (2008) spanned a wide range of metallicity but was limited to stars with $T_{\rm eff} > 4000$K and Bertone et al. (2008) did not compare specific feature strengths.

In the present case we have extended the library and the interpolator into the cool dwarf regime, and cover a much wider range in metallicity in the near-IR than was previously available. This means that we can now make explicit comparisons between the theoretical and empirical behavior in this regime.

Starting our comparison with near-IR features of the solar metallicity stars (fourth row from the top in Figures 2.21-2.23) we see that the theoretical predictions are consistent with the empirical trends for the NaI0.82μm, MgI0.88μm, NaI1.14μm, KI1.17μm, and NaI2.21μm features. There is discrepancy between the theoretical predictions and empirical trends for the CaT, FeH, CaI1.98μm, and CO2.30μm features. Although, with the exception of the warm dwarf regime in the CaI1.98μm feature, the discrepancies are relatively small offsets in strength of the feature, with an overall similar trend with effective temperature. Furthermore, the discrepancies tend to be in the dwarf stars rather than the giant stars.

We find some small differences when comparing the consistency between the empirical and theoretical trends shown in Figures 2.21-2.23 with Conroy & van Dokkum (2012a). First, the isochrones used in this work extend to cooler temperature than those used in Conroy & van Dokkum (2012a) which means we can now compare the equivalent widths of the coolest dwarfs to the theoretical libraries. For NaI0.82μm the prediction for the equivalent widths of the coolest dwarfs is inconsistent with the data. However, the prediction for NaI0.82μm feature strength in this work is completely consistent with the empirical trend. Similarly, in Conroy & van Dokkum (2012a) the theoretical prediction for the CaT was consistent with the data and here the CaT is slightly overpredicted. This most likely do the modest changes that have occurred in the theoretical libraries since Conroy & van Dokkum (2012a).

Moving onto solar metallicity stars in the optical, Figures 2.24-2.29, we find that the theoretical predictions fare worse in the optical than the near-IR. We show that in the optical the overall trends predicted by the theoretical library are consistent with the empirical trend but are often offset in predicted strength. Martins & Coelho (2007) compared the strengths of Lick indices of various empirical and theoretical libraries and we will compare the discrepancies of that work with those shown here. Here we focus on the comparison between the theoretical models from Martins et al. (2005) and the MILES library.

The C3K models underpredict the index strength for all the Balmer lines: Hδ, Hγ, and Hβ with the Hβ difference being less severe than the former two. This is consistent with the result from Martins & Coelho (2007). Furthermore, C3K overpredicts the G4300Å band which is again consistent with Martins & Coelho (2007). The theoretical predictions for the iron lines are also stronger than what is observed (with the exception of the feature at 5782 Å). However, Martins & Coelho (2007) found that the inconsistency between the theoretical and empirical predictions for the iron lines was worse at cooler temperatures which is not consistent with what we find in our comparison. Additionally we find good agreement between the C3K predictions for the line strength of MgI0.88μm and Mgb at temperatures where Martins & Coelho (2007) find disagreement in the magnesium lines at cool temperatures. While both C3K and the theoretical stellar spectra in Martins et al. (2005) use the ATLAS model atmospheres, C3K uses ATLAS12 and Martins et al. (2005) uses ATLAS9 so it is not unexpected that the comparisons in Martins & Coelho (2007) are different.

We now turn to a comparison of the metallicity-dependence of the models and observations. For the near-IR features the consistency seen at solar-metallicity holds at the other metallicities with the exception of some of the coolest stars, particularly dwarf stars, and especially at low-metallicity. For instance, the theoretical predictions for the low-metallicity cool dwarf stars is different from the empirical trend for the CaT, NaI (at 0.82μm and 1.14μm), FeH0.99μm, CaI1.98μm, and KI1.17μm features. However, there is good consistency over this same regime for the CO2.30μm and MgI0.88μm features.

The theoretical predictions for the NaD feature are consistent with the empirical trends over all the metallicity bins, albeit slightly overpredicted in the solar metallicity bin. For some of the optical features the theoretical predictions seem to improve at lower metallicities. This includes many of the iron lines, CN1, G4300Å, and the Balmer lines with the exception of the coolest dwarf stars. The theoretical predictions for the TiO band are in agreement with the empirical trends at solar metallicity but that agreement worsens at lower metallicities for the cool dwarf stars.

We emphasize that SPI is not well constrained for the low-metallicity cool dwarf regime and therefore apparent disagreement between the models and the data should not be over-interpreted. However, the pattern of discrepancy in feature strengths for the low metallicity cool dwarf stars does suggest that caution should be employed when supplementing the cool dwarf regime in empirical libraries with theoretical spectra. We will note that the stellar population models of Conroy & van Dokkum (2012a) and Lick index models (e.g., Trager et al. 2000; Thomas et al. 2003; Schiavon 2007) use theoretical models only differentially, calculating relative changes with respect to a fiducial model (the response functions) and so they are less sensitive to the absolute limitations of the models.

2.5 Summary

In this paper we presented a new spectroscopic stellar library, the Extended IRTF Library, which in its entirety consists of 284 stars covering a wide range of stellar parameter space, including both low and high metallicities. The stars were observed on the SpeX instrument and the spectra cover a wavelength range of $0.7 - 2.5\mu$m. All the stars included in this library were selected from the MILES optical library, providing continuous coverage from $0.35 - 2.5\mu$m.

In addition to the new library we have also created a Spectral Polynomial Interpolator (SPI). This is a tool that generates a data-driven model from a subset (194 out of 284) of the highest quality library stars and can be used to produce a stellar spectrum for arbitrary values for $T_{\rm eff}$, log$g$, and [Fe/H]. With the Extended IRTF Library and SPI:

- We find good agreement between observed and synthesized colors for all of the colors explored, including 2MASS H-K_S, J-K_S, Tycho B_T-V_T, and V_T-K_S. This agreement means that we are recovering the overall spectral shapes to within 1-4% percent, on average.

- We find that the empirical uncertainty of the spectra is $\sim 0.5\%$ with the exception of regions heavily contaminated by telluric absorption or regions in the A0 V standard star heavily contaminated by hydrogen absorption lines. In these cases the uncertainty is on the order of a 1-2% percent.
- We measured the wavelength-dependent SpeX instrument resolution and found the median resolution of the stars in the library to be consistent with the nominal value of $R \approx 2000$. We also independently measured the MILES resolution and found it to be consistent with the updated value from Beifiori et al. (2011).
- We find that stellar features retain their characteristic properties at non-solar metallicities. This includes the surface gravity sensitive lines such as NaI (at 0.82μm and 1.14μm), CaT, FeH0.99μm, KI1.17μm, and CO2.30μm.
- We find the theoretical predictions for the spectral features qualitatively agree with the observed trends. The C3K theoretical spectra in many cases reproduce the trends quantitatively as well, especially in the near-IR. Nonetheless, there are many features that show significant quantitative discrepancies between models and observations, especially in the optical.

Chapter 3

Initial Mass Function Variability (or not) Among Low-Velocity Dispersion, Compact Stellar Systems

3.1 Introduction

The assumption of a universal stellar initial-mass function (IMF) has been a cornerstone of stellar population and galaxy evolution studies for decades. Nevertheless, there has been much observational effort to test and challenge this assumption. The work done in nearby systems where it is possible to measure resolved star counts is extensive (see Ch. 9 in Kroupa et al. 2013, and references therein). Since the discovery of surface gravity sensitive absorption features (e.g., Wing & Ford 1969) the measurement of the IMF in systems beyond the reach of resolved star counts has been possible. In principle, these lines can measure the ratio of giant-to-dwarf stars in integrated light, which can be used as an IMF proxy (e.g., Cohen 1978; Faber & French 1980; Kroupa & Gilmore 1994).

In practice, only in recent years have the stellar population synthesis (SPS) model precision and near-infrared (near-IR) data quality reached the point where it is to possible measure the dwarf-to-giant ratio. Cenarro et al. (2003) found that age and metallicity effects alone could not explain the variations in CaT strength in a sample of early-type galaxies (ETGs) and tentatively attributed it to IMF variability. More recent work (e.g., van Dokkum & Conroy 2010; Spiniello et al. 2011; Conroy & van Dokkum 2012a; Ferreras et al. 2013; Martín-Navarro et al. 2015) has made progress on making quantitative statements about the relative number of giant and dwarfs stars. The results from SPS modeling broadly agree with investigations using gravitational lensing and kinematics (e.g., Treu et al. 2010; Cappellari et al. 2013). However, there remain inconsistencies from the different methods on an object-by-object basis (Smith 2014).

There is not yet a clear physical mechanism driving IMF variability. Metallicity has become a possibility from recent observational work (Martín-Navarro et al. 2015; van Dokkum et al. 2016) but velocity dispersion (σ) and α-element abundances also correlate with IMF variation (Conroy & van Dokkum 2012a; La Barbera et al. 2013). Furthermore, there are still unexplained complications in the emerging picture of IMF variability. Newman et al. (2017) demonstrated that even high-velocity dispersion ETGs can have MW IMFs, and, furthermore, it is not yet clear how IMF variability conforms to the expectations from chemical evolution and star-formation measurements (e.g., Martín-Navarro 2016).

Most integrated light probes of the IMF focused on ETGs and so have only looked at IMF variations in relatively narrow regions of parameter space. To better constrain IMF variations as a function of the physical characteristics of the stellar population we need to push IMF studies to the extremes of parameter space. Ultracompact dwarfs (UCDs) are extremely dense objects that can have high dynamical mass-to-light ratio values $(M/L)_{\mathrm{dyn}}$ (e.g., Mieske et al. 2013). Globular clusters (GCs) are conventionally thought to have Kroupa (2001) (MW) IMF. However, Strader et al. (2011a) found a trend of decreasing $(M/L)_{\mathrm{dyn}}$ of M31 GCs as a function of metallicity, in disagreement with the expectation from a MW IMF.

Whether UCDs and GCs actually have variable IMFs and, if so, what the shape is, is still being debated (Jeřábková et al. 2017). Dabringhausen et al. (2012) took an overabundance of X-ray binaries in a sample of Fornax UCDs as evidence that those UCDs produced more massive stars than expected from a Kroupa IMF. Marks et al. (2012) used the gas-expulsion timescale of a sample of UCDs and GCs to predict that the IMF would create more massive stars with increasing density. However, Pandya et al. (2016) analyzed 336 spectroscopically confirmed UCDs across 13 host systems and found an X-ray detection fraction of only $\sim 3\%$. Zonoozi et al. (2016) showed that the combination of a variable IMF and removal of stellar remnants could plausibly explain the $(M/L)_{\rm dyn}$ trend in the M31 GCs.

Fitting the integrated light of UCDs and GCs with SPS models is needed to obtain a more direct measurement of the IMF shape. One caveat is that GCs can be strongly influenced by dynamical evolution, i.e., mass-segregation and evaporation of low-mass stars. For the low-mass stars the "initial" mass function is not being measured, but rather the "present-day" mass function (PDF). However, this should not be a concern for high mass GCs or UCDs, the PDF is expected to closely resemble the IMF owing to long relaxation times (see eq. 17 in Portegies Zwart et al. 2010).

In this paper we present a pilot study of stellar mass-to-light ratios, $(M/L)_*$, of various compact stellar systems (CSSs): M59-UCD3 (Sandoval et al. 2015), three M31 GCs that span a large range of metallicity, and the compact elliptical (cE) M32. For the first time we fit the spectra of the individual objects with flexible SPS models that allow IMF variability.

3.2 Observations and Data

All of the objects presented in this paper were observed with LRIS (Oke et al. 1995), a dual-arm spectrograph, on the Keck I telescope on Maunakea, Hawaii.

The data for one metal-poor (MP) GC (M31-B058), two metal-rich (MR) GCs (M31-B163 and M31-B193), and M59-UCD3 were obtained on December 19–20 2014, using the instrument setup and using the same "special" long slit discussed in van Dokkum et al. (2016) ($0.7'' \times 290''$). Since the objects in this paper are bright and compact we obtained 4 300s exposures using an ABAB pattern where we dithered up

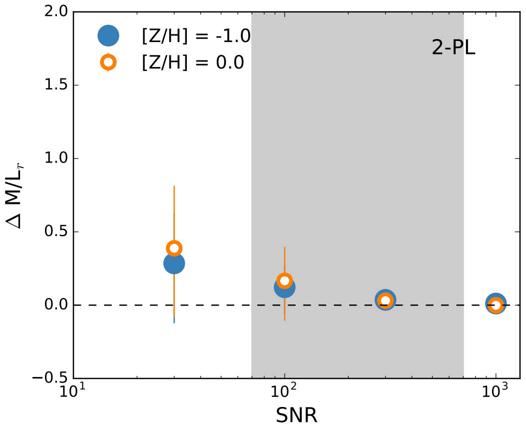

Figure 3.1: Recovery of $(\mathrm{M/L})_*$ from mock data as a function of S/N for $[Z/\mathrm{H}] = 0.0$ (orange) and $[Z/\mathrm{H}] = -1.0$ (blue) models. The circles show the median difference between the input $(\mathrm{M/L})_*$ and the inferred $(\mathrm{M/L})_*$ derived from the fits of 10 realization of mock data. A S/N of $\gtrapprox 100$ is needed to recover the M/L. The grey band shows the range of S/N values in the data.

and down the slit by $20''$.

Three exposures of 180 s were taken for M32 on January 2012. The 600 l mm^{-1} grating was used on the blue arm but the same grism as the other objects was used on the red arm. We extracted a spectrum using a square aperture of $0.8''$x$0.8''$ (≈ 3 pc).

The intrinsic resolution of the the objects in this sample is higher than the models (which are smoothed to a common resolution of $\sigma = 100$ km s^{-1}) so we broadened the spectra in our sample. To have roughly the same dispersion in the red for all objects we broadened the M32 and UCD spectra by 150 km s^{-1} and the GCs by 200 km s^{-1}.

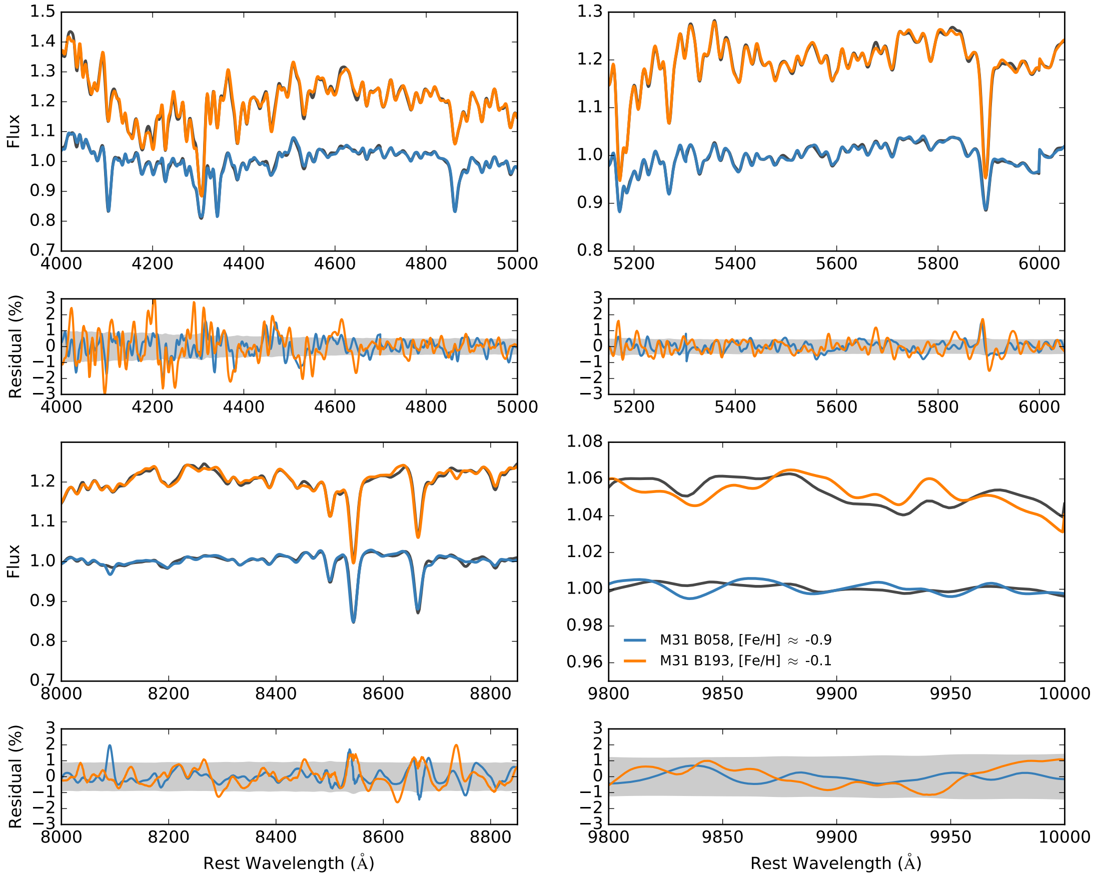

Figure 3.2: (Upper panels) Comparison of best-fit models (grey) and data in key wavelength regions for M31-B193 (metal-rich GC, orange) and M31-B058 (metal-poor GC, blue). (Lower panels) Comparison of the percentage difference between the best-fit model and data regions for M31-B193 and M31-B058. The data have been smoothed and so the pixels are highly correlated. In the grey band we show the uncertainty for one of the GCs, M31-B058, as the uncertainties are comparable. The residuals between metal-rich and metal-poor GC are also comparable.

3.3 Modeling

3.3.1 Model Overview

The methodology we use for fitting the models to data and the parameters fitted are described in detail in Conroy et al. (2018). The models described in Conroy et al. (2018) ("C2V" models) are the updated versions of the stellar population models from Conroy & van Dokkum (2012a) ("CvD" models). The most important update for this paper is the increased metallicity range provided by the Extended IRTF library (Villaume et al. 2017b) and metallicity-dependent response functions.

We explore the parameter space using a Fortran implementation of emcee (Foreman-Mackey et al. 2013a), which uses the affine-invariant ensemble sampler algorithm (Goodman & Weare 2010). We use 512 walkers, 25,000 burn-in steps, and a production run of 1,000 steps for the final posterior distributions.

We perform full-spectrum fitting. We continuum normalize the models by multiplying them by higher-order polynomials to match the continuum shape of the data.

We sample the posteriors of the following parameters: redshift and velocity dispersion, overall metallicity, a two component star formation history (two bursts with free ages and relative mass contribution), 18 individual elements, the strengths of five emission line groups, fraction of light at $1\mu m$ contributed by a hot star component, two higher order terms of the line-of-sight velocity distribution, and nuisance parameters for the data (normalization of the atmospheric transmission function, error and sky inflating terms).[8]

Additionally, we fit for the slopes of a two component power-law (break point at 0.5 $M_\odot$):

$$\xi(m) = \mathrm{d}N/\mathrm{d}m_* = \begin{cases} k_1 m^{-\alpha_1} \text{ for } 0.08 < m < 0.5, \\ k_2 m^{-\alpha_2} \text{ for } 0.5 < m < 1.0, \text{ and} \\ k_3 m^{-2.3} \text{ for } \geq 1.0. \end{cases}$$

For a MW IMF $\alpha_1 = 1.3$ and $\alpha_2 = 2.3$. The IMF above 1.0$M_\odot$ is assumed to have a Salpeter (1955) slope. The k_i's are normalization constants that ensure continuity of the IMF. The upper mass limit is 100$M_\odot$ and the low-mass cutoff, m_c, is fixed at 0.08$M_\odot$. In this paper we present our IMF results in terms of $(M/L)_*$. The mass of the stellar population is calculated from the best inferred slopes of the IMF and stellar remnants are included in the final mass calculation following Conroy et al. (2009). A stellar population is considered bottom-heavy, an overabundance of low-mass stars, if the exponents on the first two terms are larger than the MW IMF and is considered bottom-light, a paucity of low-mass stars, if they are less than those values.

3.3.2 Mock Data Demonstrations

To test our ability to recover $(M/L)_*$ from the data, we synthesize mock spectra by assuming a Salpeter IMF, adding different amounts of noise, and then use our models and fitting procedures to derive Δ M/L_*. We show Δ M/L_* for mock spectra with

---

[8] Models fitted with only a single age and excluding the emission lines made a negligible effect on the inferred parameters for the GCs.

Table 3.1: List of objects and associated physical parameters.

Object	S/N Å^{-1}	σ (km s^{-1})	[Fe/H]	Age (Gyr)	[Mg/Fe]	M/L$_V$ 2 PL	M/L$_V$ MW
M32	730*	75$^{\rm a}$	$0.15^{+0.01}_{-0.01}$	$2.98^{+0.05}_{-0.06}$	$0.02^{+0.04}_{-0.01}$	$2.4^{+0.64}_{-0.64}$	$1.63^{+0.03}_{-0.03}$
M59-UCD3	70	70$^{\rm b}$	$0.01^{+0.01}_{-0.01}$	$7.7^{+0.49}_{-0.48}$	$0.18^{+0.01}_{-0.01}$	$5.1^{+0.87}_{-1.17}$	$2.98^{+0.11}_{-0.1}$
M31-B163	100	21$^{\rm c}$	$-0.18^{+0.01}_{-0.01}$	$11.37^{+0.7}_{-0.61}$	$0.21^{+0.01}_{-0.01}$	$3.61^{+0.59}_{-0.49}$	$3.34^{+0.12}_{-0.11}$
M31-B193	250	19$^{\rm c}$	$-0.11^{+0.01}_{-0.01}$	$9.7^{+0.54}_{-0.45}$	$0.24^{+0.01}_{-0.01}$	$2.69^{+0.43}_{-0.2}$	$3.16^{+0.09}_{-0.1}$
M31-B058	120	23$^{\rm c}$	$-0.96^{+0.01}_{-0.01}$	$6.92^{+0.09}_{-0.1}$	$0.37^{+0.02}_{-0.02}$	$1.38^{+0.07}_{-0.08}$	$1.54^{+0.01}_{-0.01}$

Note. — Mean best inferred value for each parameter is shown with 1σ statistical uncertainty. Values were determined with our models and fitting procedure, as described in Section 3.1. The second to last column are the (M/L)$_*$ values where the IMF was allowed to vary as a two component power-law IMF and the last column is the (M/L)$_*$ values where the IMF was fixed to a Kroupa IMF.

[*] Although the S/N is high it was cloudy at the time of observation so there is additional uncertainty in the data not represented by Poisson statistics.

[a] Gültekin et al. (2009)

[b] Janz et al. (2016)

[c] Strader et al. (2011a)

solar, $[Z/\mathrm{H}] = 0.0$ (orange), and sub-solar, $[Z/\mathrm{H}] = -1.0$ (blue). For each S/N and metallicity value we create 10 mock spectra with fixed S/N per Å over the wavelength range $0.4 - 1.015\mu m$, a velocity dispersion of 250 km s^{-1}, and an age of 10 Gyr. The abundance patterns of the mock spectra are solar scaled (e.g., Choi et al. 2016) and the nuisance parameters are set to zero. The points shown in Figure 3.1 are the median values of the differences between the input (M/L)$_*$ and the derived (M/L)$_*$ from the inferred IMF parameters for each metallicity and S/N pair. The uncertainties shown are the median statistical uncertainties of the recovered values.

For solar metallicity the models recover (M/L)$_*$ when the S/N $\gtrapprox 100$Å^{-1}. A similar trend is also seen in the low-metallicity mock data. While not a significant difference, it is somewhat counterintuitive that the (M/L)$_*$ at the low-S/N regime is better recovered for the low-metallicity mocks. It could be that in the low-S/N regime weaker metal lines help distinguish IMF effects. Below S/N $\sim 100\text{Å}^{-1}$ there will be large uncertainty and bias in the $(\mathrm{M/L})_*$ measurement. The bias exists in the low-S/N regime because the priors become important and the truth is at the edge of the prior. The measurements are less sensitive to S/N if the true m_c is higher (see Conroy et al. 2017, for details). As discussed in Conroy et al. (2017) the S/N requirements for allowing m_c to vary is even higher than what is shown in Figure 3.1. Most of the data in this paper do not meet the S/N requirements for this type of parametrization.

3.4 Results

3.4.1 Basic Stellar Population Characteristics

In the upper panels of Figure 3.2 we compare the best-fit models (grey) and data for M31-B193 (orange), a metal-rich (MR) GC, and M31-B058 (blue), a metal-poor (MP) GC. In the lower panels we show the percentage difference between the models and data. The uncertainty for M31-B058 is shown by the grey band (the uncertainty for M31-B193 is comparable). The CvD models would not have been able to fit M31-B058 because of the limited metallicity range, but with the C2V models the residuals between MP and MR GC are comparable and small.

In Table 3.1 we show the best inferred median values for [Fe/H], mass-weighted age, [Mg/Fe], and the $(\mathrm{M/L})_*$ in Johnson V where we have and have not allowed the IMF to vary from Kroupa. Our stellar parameters are broadly consistent with previous work on these objects. From deep *HST*/ACS imaging of M32 Monachesi et al. (2012) inferred two dominant populations, one 2–5 Gyr and metal-rich and an older population,

~ 7 Gyr. Our inferred age skews young as the integrated light observations are almost certainly dominated by the young population. Monachesi et al. (2012) determined near-solar mass- and light-weighted metallicities for M32. Our inferred metallicity is slightly more metal-rich than that. Janz et al. (2016) used Lick indices on M59-UCD3 and found $[Z/\mathrm{H}] = 0.15 \pm 0.10$. Converting our value for [Fe/H] to $[Z/\mathrm{H}]$ (Trager et al. 2000) we get $[Z/\mathrm{H}] \approx 0.2$, consistent with the Janz et al. (2016) value. Furthermore, our inferred values for M59-UCD3 are consistent with those presented in Sandoval et al. (2015) with a spectrum from a different instrument and an earlier iteration of our models.

Our inferred ages for M31-B163 and B193 are consistent with the ages derved by Colucci et al. (2014). This is particularly striking since Colucci et al. (2014) worked with high-resolution data and a completely different analysis technique. The age for M31-B058 is young for a GC but is consistent with previous work in modeling integrated light of MP GCs (Graves & Schiavon 2008). In the case of M31-B058 there is a moderate blue horizontal branch that could be boosting the strength of the Balmer lines (Rich et al. 2005).

3.4.2 The IMF

For our main analysis we define the "IMF mismatch" parameter, α_{IMF}. This parameter is the ratio of $(\mathrm{M/L})_*$ where we have fitted for the IMF, to $(\mathrm{M/L})_*$ where we have assumed a MW IMF. In Figure 3.3 we show α_{IMF} plotted against [Fe/H] (left), [Mg/Fe] (middle), and velocity dispersion (σ, right) for all the objects in our sample: the M31 GCs (purple), M59-UCD3 (red), M32 (green). We supplement our data set with the ETG data from van Dokkum et al. (2016) (grey, open circle) with the same

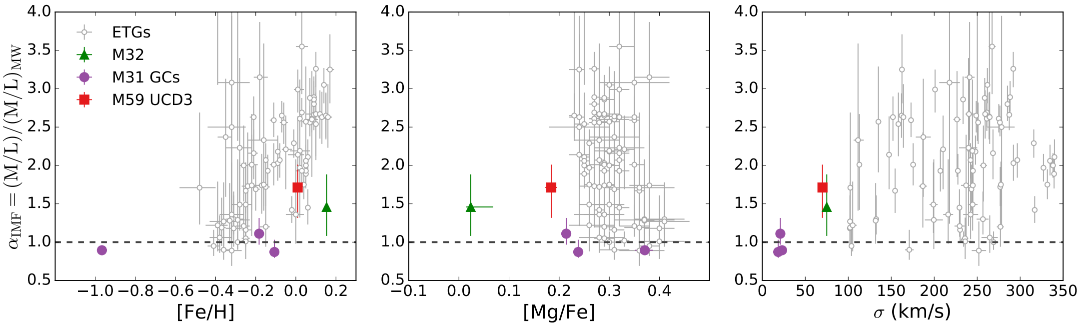

Figure 3.3: The IMF mismatch parameter plotted against [Fe/H] (left), [Mg/Fe], (middle), and σ (right) for the two component power-law IMF. Values shown are for M59-UCD3 (red squares), the M31 GCs (purple cricles), and M32 (green triangles). We show the full sample of ETG local values from van Dokkum et al. (2016) (open grey).

instrumental and model setups.

In Figure 3.4 we compare our $(M/L)_*$ measurements with available $(M/L)_{\rm dyn}$ measurements. In the left panel, we show the kernel density estimate (KDE) for [Fe/H] vs. $(M/L)_{\rm dyn}$ for M31 GCs from Strader et al. (2011a) (contours, darker color indicates higher concentration of objects) along with our $(M/L)_*$ for three GCs. Published $(M/L)_{\rm dyn}$ measurements do not currently exist for M59-UCD3. However, in the middle panel we show the KDE of [Fe/H] vs. $(M/L)_{\rm dyn}$ of the sample of UCDs from Mieske et al. (2013) (we removed objects that belong to NGC 5128 owing to suspicions of spurious σ measurements) and $(M/L)_*$ for M59-UCD3. In the right panel of Figure 3.4 we compare $(M/L)_*$ for M32 with $(M/L)_{\rm dyn}$ from van den Bosch & de Zeeuw (2010) where the grey band represents the lower and upper limits given by the uncertainty. In each panel we show metallicity-dependent $(M/L)_*$ predictions using SSPs with MW IMFs and solar-scaled abundance patterns. The ages of the SSPs were chosen to approximate the inferred ages from full-spectrum fitting.

We note the slight discrepancy in Figures 3.3 and 3.4 in how much M32 appears to deviate from a MW IMF. This is due to the fact that the MW IMF in Figure 3.3 also accounts for non-solar abundance patterns while the SSPs used to generate the orange lines in Figure 3.4 do not.

3.5 Discussion

McConnell et al. (2016) and Zieleniewski et al. (2017) computed line indices for a variety of ETGs and claimed that observed line strengths can be explained by abundance variations alone. These studies have driven debates about the extent IMF measurements are affected by the underlying abundance patterns. The M31 GCs are an excellent test bench for the models in this respect since they have similar metallicities and element enhancements as massive ETGs. If the models did conflate metallicity and abundance effects with IMF effects we would expect to find similar $(M/L)_*$ enhancements in the M31 GCs. Recovering $\alpha_{\rm IMF} \sim 1$ for the M31 GCs over a wide metallicity range is a strong validation that our models can distinguish IMF and abundance effects.

Our modeling of the M31 GCs improves upon earlier work in several important ways. Zonoozi et al. (2016) did not fit models to data and assumed a top-heavy IMF. Conroy & van Dokkum (2012a) used a stacked spectrum of MR GCs to test the CvD models while making measurements for individual clusters and include a MP GC. The lack of expected dark matter in GCs means that dynamical measurements provide tight constraints on our expectations for $(M/L)_*$. This makes the continued discrepancy between dynamical and stellar measurements on the MR end of the M31 GCs troubling.

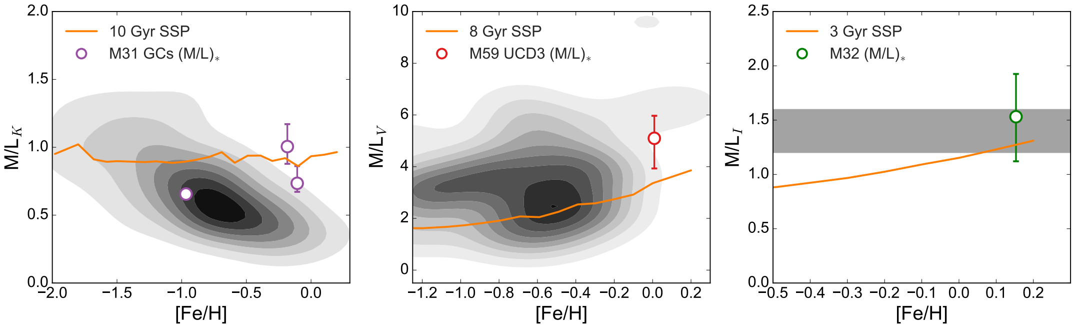

Figure 3.4: Comparison of $(\mathrm{M/L})_{\mathrm{dyn}}$ (grey) to $(\mathrm{M/L})_*$ values for M31 GCs (left, purple), M59-UCD3 (middle, red), and M32 (right, green). In each panel we show the metallicity-dependent $(\mathrm{M/L})_*$ predicted from SSPs with Kroupa IMF and solar-scaled abundance patterns. The ages of the SSPs (orange line) were chosen to approximate the inferred ages from our full-spectrum fitting. Our inferred $(\mathrm{M/L})_*$ values for M59-UCD3 and M32 are consistent with available $(\mathrm{M/L})_{\mathrm{dyn}}$ measurements. There remain inconsistencies between the dynamical and stellar measurements at high metallicity for the M31 GCs.

For the current models m_c is fixed at $0.08M_\odot$ but a higher m_c would lower the inferred $(\mathrm{M/L})_*$ values. Chabrier et al. (2014) explored the different theoretical conditions which would create a higher m_c, while there is empirical evidence that the IMF in GCs becomes flatter for $< 0.5M_\odot$ (Marks et al. 2012), which would mimic an increase in m_c. It is not out of the realm of possibility that m_c could differ from our fiducial value. However, it takes increasing m_c to $0.5\mathrm{M}_\odot$, an extreme value, to decrease $(\mathrm{M/L})_*$ by 35%, i.e., closer to the locus of the MR $(\mathrm{M/L})_{\mathrm{dyn}}$ values. It is premature to make any definitive conclusions but these preliminary results suggest that a variable IMF cannot explain the [Fe/H] vs $(\mathrm{M/L})_{\mathrm{dyn}}$ trend for the M31 GCs. Zonoozi et al. (2016) were able to achieve better agreement by making *ad hoc* adjustments to the retention rates of stellar remnants in the GCs. Follow-up work with a larger sample and more detailed physical models is required.

The mild bottom-heaviness of M59-UCD3 contrasts with the expectations of

Dabringhausen et al. (2012) and Marks et al. (2012). That is not to say that our results are in direct contradiction with either study. First, those studies are tracing the stars and we are tracing the low-mass stars. Second, It is becoming increasingly clear that UCDs as a class encompass a diverse set of objects (Janz et al. 2016). Until we have a better understanding of a more comprehensive sample of objects it is premature to make any firm conclusions about how UCDs *as a whole* behave.

For the sample presented in this work, the main feature of Figure 3.3 is that the CSSs are distinct from the main ETG sample. Though they span large [Fe/H] and [Mg/Fe] ranges, they vary much less in α_{IMF} than the ETG sample. Both M59-UCD3 and M32 have elevated α_{IMF} values but are not on the main [Fe/H]–α_{IMF} trend for massive ETGs. M59-UCD3 is in a cluster of ETG points that also deviate from the main trend. Those points originate from the central regions of just two of the galaxies in the ETG sample: NGC 1600 and NGC 2695.

The main conclusion of this work is that metallicity is not the sole driver of IMF variability (see Martín-Navarro et al. 2015; van Dokkum et al. 2016). The right panel of Figure 3.3 suggests that velocity dispersion is also associated with IMF variation. This is an important result because different theoretical frameworks will be controlled by different fundamental variables depending on the kind of physics they evoke to fragment gas clouds (see Krumholz 2014). By expanding IMF probes into the parameter space that CSSs occupy we can elucidate what these variables are.

Moreover, it is unclear how theoretical frameworks of star-formation should treat monolithically formed populations (GCs, some UCDs) as compared with populations that build up over time (some CSSs and ETGs) (see Ch. 13 in Kroupa et al. 2013). By measuring the IMFs of CSSs with the same modeling framework that we do for ETGs, we can obtain a self-consistent observational picture of how the IMF manifests in the different types of population. Currently, with our small sample, it is unclear whether the GCs have IMFs that are distinct from the UCDs and cEs (the left and middle panels of Figure 3.3) or are a part of the same continuum (right panel of Figure 3.3).

Chapter 4

New Constraints on Early-Type Galaxy Assembly from Spectroscopic Metallicities of Globular Clusters in M87

4.1 Introduction

Although ΛCDM cosmology gives us the broad framework that galaxies form hierarchically, the details of how giant early-type galaxies (ETGs) form is still a matter of debate. Areas of ongoing uncertainty include the *assembly* of ETGs such as the epoch of the last merger and what kind of progenitor galaxies now constitute the stellar halos of ETGs. In particular, while cosmological simulations point to massive progenitor satellites as building the stellar halos of present day giant ETGs (see, for example, Figure 13 of Pillepich et al. 2018), observational constraints suggest dwarf galaxies as the progenitors (Figure 2 of Forbes et al. 2015).

Globular clusters (GCs) are nearly ubiquitous around galaxies and have been determined to be old ($\sim$ 10 Gyr) in a variety of systems (see references in Brodie & Strader 2006). Those properties as well as their luminosity ($-5 < M_V < -10$) make them potentially useful tracers of galaxy formation and assembly. However, the promise of GCs in this capacity has yet to be fully realized, in part, because of our limited means to understand the present-day physical properties of GC systems.

van den Bergh (1975) first used the likely connection between a galaxy's star-formation episodes and its GC population to suggest a link between galaxy luminosity and the metallicities of its GCs. This relation was confirmed by Brodie & Huchra (1991), and subsequently the paradigm of *bimodality* has overtaken the extragalactic GC field. Bimodality was first established through optical color distributions from *Hubble Space Telescope* photometry (Gebhardt & Kissler-Patig 1999; Kundu & Whitmore 2001a; Larsen et al. 2001). Since then GC systems around ETGs are treated as composed from two subpopulations and separately track the subpopulation characteristics with host galaxy characteristics to place constraints on galaxy formation scenarios (e.g., Côté et al. 2002; Strader et al. 2005; Rhode et al. 2005; Li & Gnedin 2014). Recently though, Harris et al. (2017) presented observational evidence that the most massive ETGs, brightest cluster galaxies, can have broad unimodal distributions in addition to bimodal distributions.

GCs are thought to contain coeval stars with old ages and mostly homogenous metallicities and so broadband colors of GCs are generally considered to reflect their underlying mean metallicity. The simplicity of this logic belies the fact that there is no consensus on how broadband colors should be transformed into metallicities (parameterized as the "color–metallicity relation"). The core of almost all astronomical problems is translating observed characteristics into physically meaningful properties and understanding GC systems is no exception. We have very limited means to obtain spectroscopy – our best observational tool for deriving physical stellar population characteristics – of individual GCs around the largest elliptical galaxies. This is a result of a two-fold problem: at the distances of elliptical galaxies, GCs are faint, and the largest elliptical galaxies can host systems of tens of thousands of GCs. This means that in extragalactic work we often only have access to coarse observational characteristics of individual GCs, such as broadband photometry.

The problems associated with obtaining the metallicity distribution are illustrated through the difference between the Harris et al. (2006) and Peng et al. (2006) color–metallicity relations. Peng et al. (2006) used *HST*/ACS photometry of GCs around Virgo Cluster galaxies from Jordán et al. (2004) and metallicities gathered from the few spectroscopic studies of extragalactic GCs available at the time (Cohen et al. 1998, 2003a). Peng et al. (2006) found a color–metallicity relation with a significant break when transitioning to the blue GCs, but, crucially their relation was based almost entirely on Milky Way GCs at the metal-poor end. Harris et al. (2006) derived a linear relation between $B-I$ colors and metallicities for Milky Way GCs to interpret the broadband colors they obtained for Virgo Cluster GC systems. Peng et al. (2006) and Harris et al. (2006) reported essentially the same color distributions for the Virgo GC systems but different *metallicity* distributions.

Despite their differences, both Peng et al. (2006) and Harris et al. (2006) maintained evidence for metallicity bimodality but that paradigm was challenged by Yoon et al. (2006). Yoon et al. (2006) introduced the idea of generating synthetic color–metallicity relations to transform the overall color distributions of GC systems to metallicity distributions. They found that highly non-linear color–metallicity relations, like those that result from inclusion of helium-rich hot horizontal branch stars, can transform unimodal metallicity distributions into bimodal color distributions.

Contrary to Yoon et al. (2006) and their follow-up work (Lee et al. 2019), studies that directly model the spectroscopic observations of GCs consistently find bimodal metallicity distributions (Alves-Brito et al. 2011; Usher et al. 2012; Brodie et al. 2012). Despite the near-consensus regarding bimodality, the differences in various color–metallicity relations (see also Usher et al. 2012) highlight that there may be physical properties beyond metallicity that affect the broadband colors of GCs.

Full-spectrum stellar population synthesis (SPS) modeling provides a way to move past these problems. Modern full-spectrum models allow for variations in abundance patterns (Conroy et al. 2014) over a variety of ages (Choi et al. 2014) and metallicities (Conroy et al. 2018). In addition to fully accounting for possible variations in many stellar population parameters, we have shown that full-spectrum fitting allows us to extract information from data in a lower signal-to-noise (S/N) regime than traditional index fitting (Conroy et al. 2018).

It is exactly this last property of full-spectrum SPS models that enables us to make use of the Strader et al. (2011b) database of spectroscopy of individual GCs around M87. In this paper we present the most comprehensive and accurate compendium of metallicities for individual GCs around M87 (which we describe in Section 4.2). We use these metallicities to derive a new color–metallicity relation in Section 4.3. We discuss the implications of the new color–metallicity relation in Section 4.4.

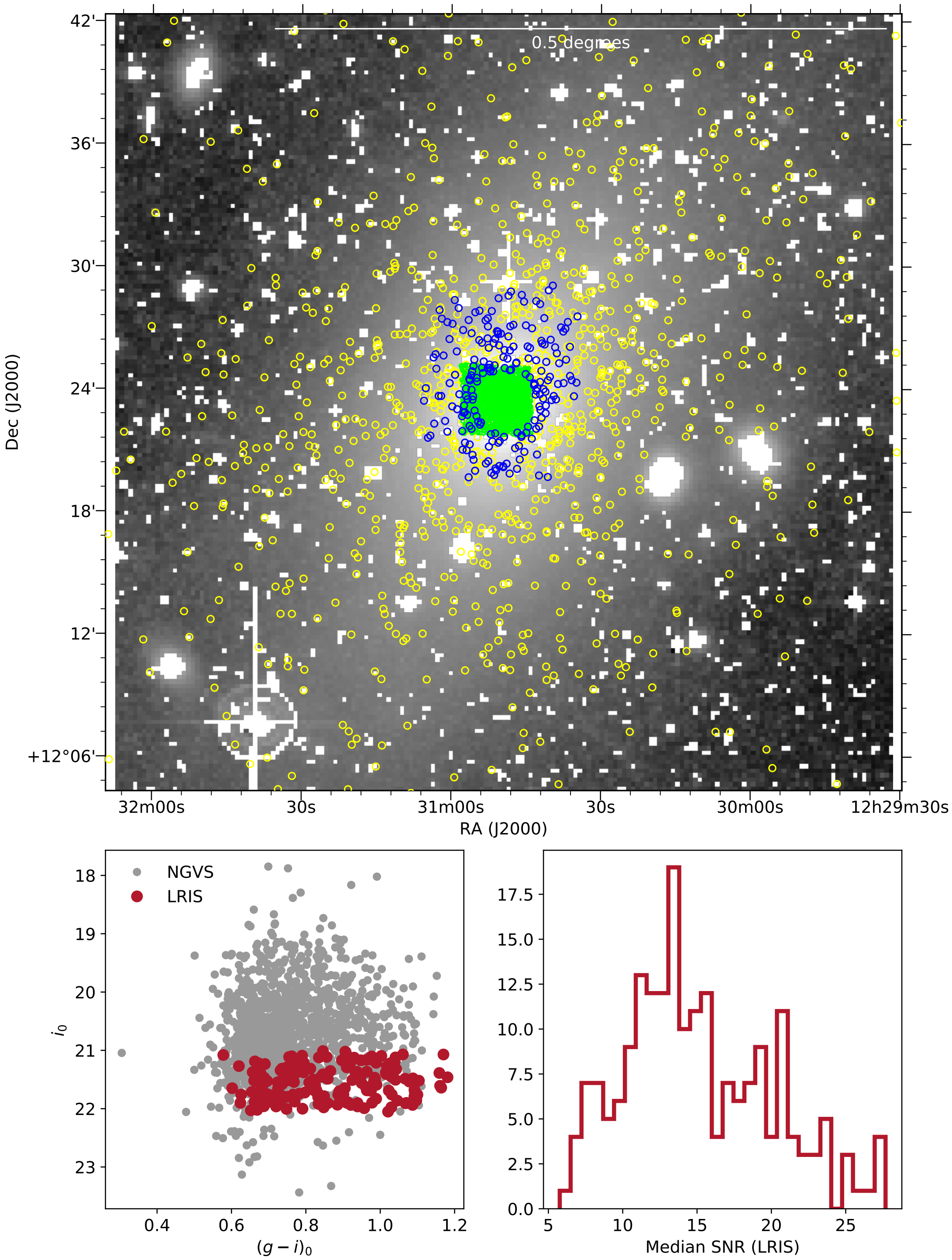

Figure 4.1: (Upper) Image of M87 from the Burrell Schmidt Deep Virgo Survey (Mihos et al. 2017) with spatial distributions of the NGVS photometry (yellow), ACSVCS photometry (green), and LRIS spectroscopy (blue), (Left) Color–magnitude diagram for the M87 GCs from the NGVS catalog from Oldham & Auger (2016a) (grey), the culled sample of the LRIS data set (red) from Strader et al. (2011b). (Right) Histogram of median signal-to-noise ratio values for the individual spectra of the LRIS sample.

4.2 Stellar Population Synthesis Modeling

We make use of the Keck/LRIS spectroscopic subsample of the dataset described in Strader et al. (2011b) ($\sim 3300 - 5600$Å). In the top panel of Figure 4.1 we show a deep image of M87 from the Burrell Schmidt Deep Virgo Survey (Mihos et al. 2017) with the NGVS photometric catalog (yellow, Oldham & Auger 2016a), the ACSVCS photometric catalog (green, Jordán et al. 2009), and the LRIS spectroscopic sample (blue, Strader et al. 2011b).

There are several features of the Strader et al. (2011b) sample that are salient to the work presented in this paper. First, the clusters in this sample were chosen to be fainter in magnitude than the obvious "blue tilt" clusters, which will help when we assess bimodality. Second, in the bottom-left panel of Figure 4.1 we compare the NGVS photometry sample with the LRIS sample in color–magnitude space. The LRIS sample spans nearly the whole color range of the M87 GC system (middle panel Figure 4.1).

This work makes use of the updated full-spectrum SPS models (`alf`) described in Conroy et al. (2018). The most relevant update of the Conroy et al. (2018) models with regards to this work is the expansion of stellar parameter coverage of the models with the Spectral Polynomial Interpolator (SPI, Villaume et al. 2017b)[9]. With SPI we used the optical MILES stellar library (Sánchez-Blázquez et al. 2006b), the Extended IRTF stellar library (E-IRTF, Villaume et al. 2017b), and a large sample of M Dwarf spectra (Mann et al. 2015) to create a data-driven model which we can use to generate stellar spectra as a function of effective temperature, surface gravity, and metallicity.

---

[9] `https://github.com/AlexaVillaume/SPI_Utils`

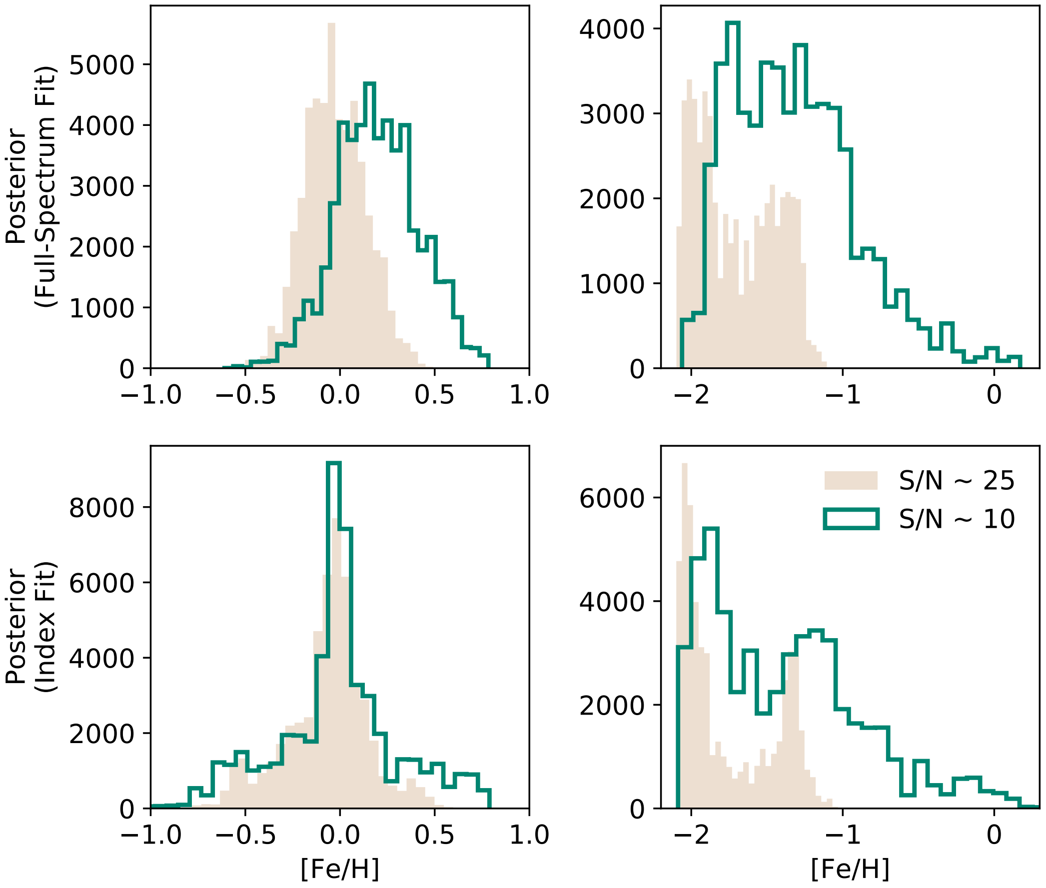

Figure 4.2: Comparing [Fe/H] posteriors for metal-rich GCs (left column) and metal-poor GCs (right column) where the spectra were fitted using the full spectrum (top row) and Lick indices (row). In all cases the full-spectrum fits provide better constraints on [Fe/H] than index fits. For the metal-rich GCs, the index fits have broader tails than the full-spectrum fits of the same GCs. For the metal-poor GCs, the index fits do not result in well-behaved posterior distributions on [Fe/H]. The GCs in this figure correspond to the GCs in Figures 4.3 and 4.4.

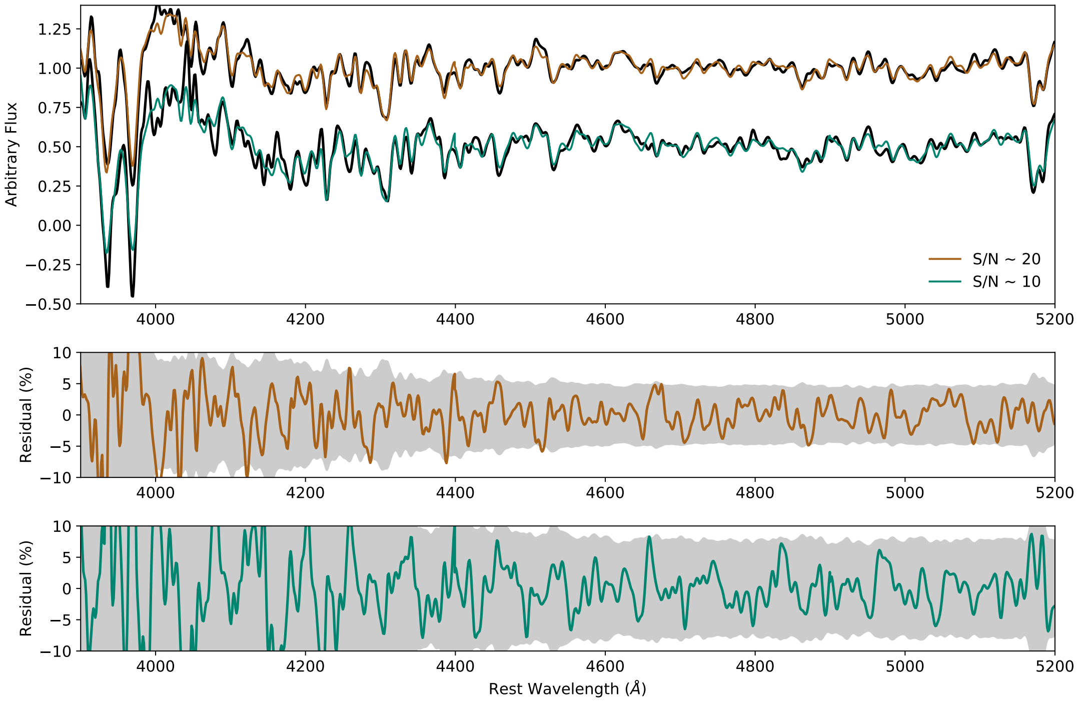

Figure 4.3: Top: Comparison of metal-rich ([Fe/H] > +0.1) LRIS spectra (black) and best-fit models for a high-S/N (H51142, $(g-z)_{\rm NGVS} = +1.38$, brown) observation and a low-S/N (H51943, $(g-z)_{\rm NGVS} = +1.33$, green). Middle: Comparison of residuals between best-fit model and data for H51075 and uncertainty of flux from the input spectrum (grey). Bottom: Same as middle panel but for H51943.

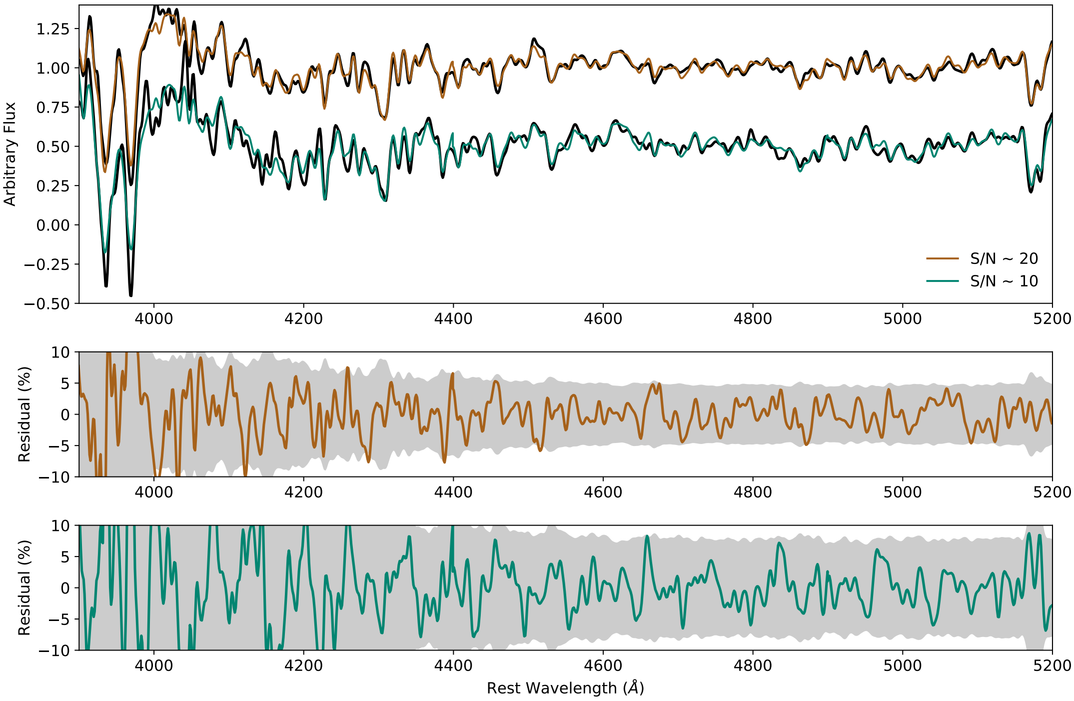

Figure 4.4: Same as for Figure 4.3 but for the metal-poor GCs (< -1.5) H38032 ($(g-z)_{\mathrm{NGVS}} = +0.70$, brown) and H42981 ($(g-z)_{\mathrm{NGVS}} = +0.69$, green).

The empirical parameter space is set by the E-IRTF and Mann et al. (2015) samples which together span $-2.0 \lesssim$ [Fe/H] $\lesssim +0.5$ and $3.9 \lesssim \log \mathrm{T_{eff}} \lesssim 3.5$. To preserve the quality of interpolation at the edges of empirical parameter space we augment the training set with a theoretical stellar library (C3K). The `alf` models allow for variable abundance patterns by differentially including theoretical element response functions. In Conroy et al. (2018) we fitted the Schiavon et al. (2005) spectroscopic sample of Milky Way GCs and compared the `alf`-inferred [Fe/H] values with a compilation of [Fe/H] values from the literature (see Roediger et al. 2014, for details). Over a range of $-2.5 \lesssim$ [Fe/H] $\lesssim +0.0$ we had nearly one-to-one consistency between the literature values and our measured [Fe/H] values from integrated light (specifically, $\mathrm{[Fe/H]_{lit}} \propto 1.06\mathrm{[Fe/H]}_{\texttt{alf}}$).

The LRIS sample is in the low signal-to-noise (S/N) regime with $\sim 5-30$ encompassing the range of the median S/N over each spectrum (bottom-right panel in Figure 4.1). In this modest S/N regime it is difficult to obtain accurate stellar population parameters (Sánchez-Blázquez et al. 2011). To obtain an accurate color–metallicity relation we need the metallicities of *individual* GCs and therefore stacking spectra is not a good option for this particular problem.

We fit objects using both full-spectrum (left) and traditional line-index methods (right). For our line-index fits we use the canonical set of Lick indices (Faber et al. 1985; Burstein et al. 1986; Worthey et al. 1994): Hδ_F, CN$_2$, Ca4227, G4300, Hγ_F, Fe4383, Fe4531, C$_2$4668, Hβ, Fe5015, Mgb, Fe5270, Fe345, and Fe5406. For the full-spectrum fits we fit in simple-mode over the wavelength regions: $3900-4400$Å, 4400 – 4900Å, 4900 – 5200Å. We smoothed the LRIS spectra to be a constant 200 km/s over the whole wavelength range.

In Figure 4.2 we demonstrate the utility of full-spectrum fitting over line-index methods. In this figure we compare [Fe/H] posteriors for metal-rich GCs (left column) and metal-poor GCs (right column) where the spectrum were fitted using the full spectrum (top row) and Lick indices (bottom row). In each panel we compare the results of high-S/N and low-S/N spectra. We demonstrate that in both the metal-rich and metal-poor cases the posteriors are better constrained when full-spectrum fitting is used. In the metal-rich case, the posterior distributions for high and low-S/N using Lick indices have larger tails than the posterior distributions from full-spectrum fitting. The real utility of the new models is shown in the low-metallicity case where the posterior distributions are more centered on a single value from full-spectrum fitting than from indices.

In Figures 4.3 and 4.4 we examine the quality of our fits for metal-rich and metal-poor GCs, respectively. In each Figure we compare the LRIS spectrum (black) with the best-fit model spectrum for a high-S/N (brown) spectrum and a low-S/N (green) spectrum in the top panel. The middle and bottom panels in each figure compare the residuals between the high-S/N spectra and low-S/N, respectively, with the flux uncertainty of each LRIS spectrum (grey band). These comparisons demonstrate that the fitting was successful as the residuals are consistent with the flux uncertainty. Even with the low-S/N spectra several spectral features are still prominent, including CaII, Hδ, Hβ, and Mgb, which are well-characterized by the best-fit model.

After we fit every spectrum we visually inspected the residuals between the observed spectrum and the best-fit model. From this inspection we identified cases where the best-fit model is clearly a poor fit to the data. We removed these clusters from our subsequent analysis, bringing our final sample to 177 GCs. Of the 23 GCs we culled from our final metallicity sample, 20 have NGVS photometry, and 15 of those are considered to be blue ($g - z < 1.0$). This suggests that it is more difficult to obtain adequate spectra of the blue and, presumably, metal-poor GCs. However, with our remaining blue GCs we are still adequately covering the metal-poor parameter space. The posteriors for the [Fe/H] values for the final sample of GCs are available at `https://github.com/AlexaVillaume/m87-gc-feh-posteriors`.

4.3 Results

4.3.1 Comparison to Previous Work

Cohen et al. (1998) previously did stellar population analysis on a spectroscopic sample of M87 GCs (Cohen & Ryzhov 1997) using indices to determine metallicity values. To aid our analysis we matched the Cohen et al. (1998) sample to the Oldham & Auger (2016a) NGVS-based photometry catalog. We matched the Cohen et al. (1998) sample to the data presented in Hanes et al. (2001), which provided right ascension and declination values for all the GCs in the Strom et al. (1981) catalog that Cohen & Ryzhov (1997) selected their targets from.

Then we used the position values to match with the Oldham & Auger (2016a) catalog with a max separation of $1''$. We dereddened the Oldham & Auger (2016a) photometry using the Fitzpatrick (1999) extinction law and extinction values taken from the Schlegel et al. (1998) dust map using the NASA/IPAC Infrared Science Archive ($A_g = 0.087$, $A_i = 0.048$, $A_z = 0.034$, $R_g = 3.793$, $R_i = 2.086$, $R_z = 1.479$).

We do not include the objects in Table 1 of Cohen & Ryzhov (1997) and not every GC in the Cohen et al. (1998) sample has NGVS photometry so we go from the full Cohen et al. (1998) sample of 150 GCs with [Fe/H] values to 101 GCs. In the left panel of Figure 4.5 we compare the normalized cumulative metallicity distributions of both the full (blue) and matched (orange) Cohen et al. (1998) sample. This comparison demonstrates that we are not biasing the Cohen et al. (1998) sample by doing the matching.

In Figure 4.6 we compare our final sample of 177 GCs to the photometry-matched Cohen et al. (1998) sample. In the left panel we compare the cumulative brightness distributions of each sample. In the middle panel we compare the NGVS $(g-z)$ colors of the two sample. In the right panel we compare the cumulative metallicity distributions of both samples. We see that $\sim 40\%$ of the objects in our sample are fainter than the faintest GC included in the Cohen et al. (1998) sample. The range of colors spanned by each sample are similar but the Cohen et al. (1998) sample has a different overall distribution than our sample. More importantly, we see that from the way the curves change from color to metallicity that the Cohen et al. (1998) color–metallicity relation will be different than ours. Furthermore, the Cohen et al. (1998) metallicities are, on the whole, lower than our metallicities. We discuss the nature of this last difference in more detail in Section 4.1.

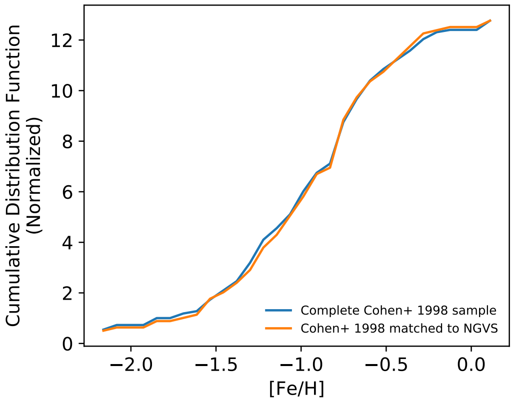

Figure 4.5: Comparing the complete sample of M87 GC metallicities from Cohen et al. (1998) (140 GCs, blue) with the sample when matched to the Oldham & Auger (2016a) photometry (101 GCs, orange) to demonstrate that we are not biasing the Cohen et al. (1998) metallicity distribution by matching to photometry.

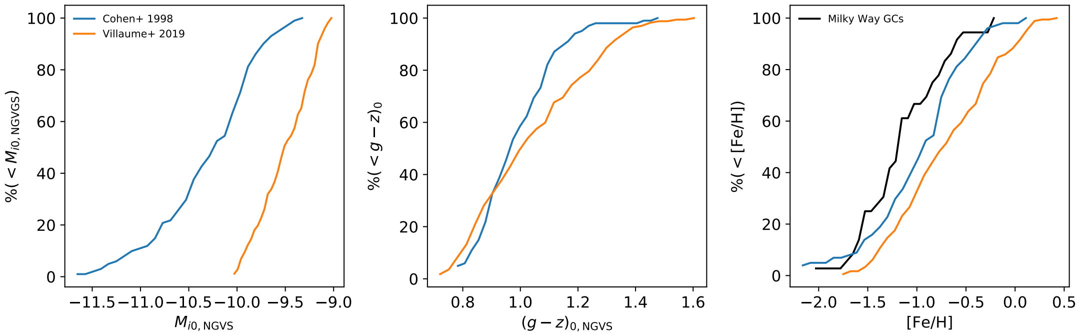

Figure 4.6: Left: Comparing the cumulative magnitude functions for the matched Cohen et al. (1998) sample and the sample from this work. Middle: Same as left but for $(g - z)$. Right: Same as left and middle but for [Fe/H]. Also in the right panel we show the distribution of [Fe/H] values for Milky Way GCs from our full-spectrum fits to the Schiavon et al. (2005) data, which shows that the Milky Way GCs are typically more metal-poor than the M87 GCs.

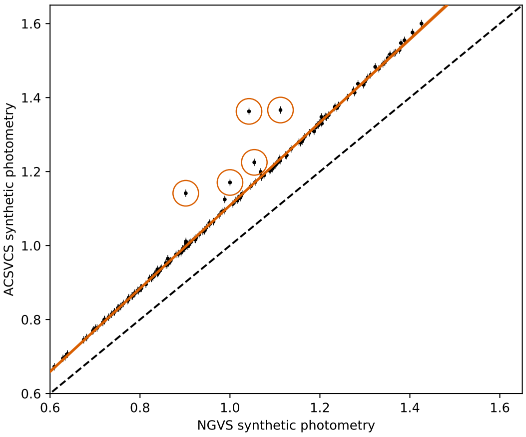

Figure 4.7: Relation between synthetic NGVS and ACSVCS photometry for the spectroscopic sample. Since the two surveys are on slightly different filter systems we present a way to transform colors between each: $(g - z)_{\mathrm{ACSVCS}} = 1.123(g - z)_{\mathrm{NGVS}} - 0.015$.

4.3.2 Updated color–metallicity Relationships

We use two photometric datasets of the M87 GC system: the Oldham & Auger (2016a) catalog of ground based photometry using the NGVS survey data (Ferrarese et al. 2012) and photometry from the ACS Virgo Cluster Survey (ACSVCS) from Jordán et al. (2009). We use the g- and z-band filters from each survey but it is important to note that the filters are not identical between the two instruments (see Figure 4.7) and so the color–metallicity relationships for the two instruments will be slightly different.

Our sample of 177 spectroscopically-derived [Fe/H] values overlaps with 172 objects from the NGVS catalog but only 37 of the GCs with spectroscopically-derived metallicities overlap with the ACSVCS catalog. To mitigate any problems that might arise from such a sparse sample we leverage the fact that the underlying `alf` models extend over a wider wavelength range than the LRIS data and are flux calibrated (see Villaume et al. 2017b; Conroy et al. 2018, for discussion).

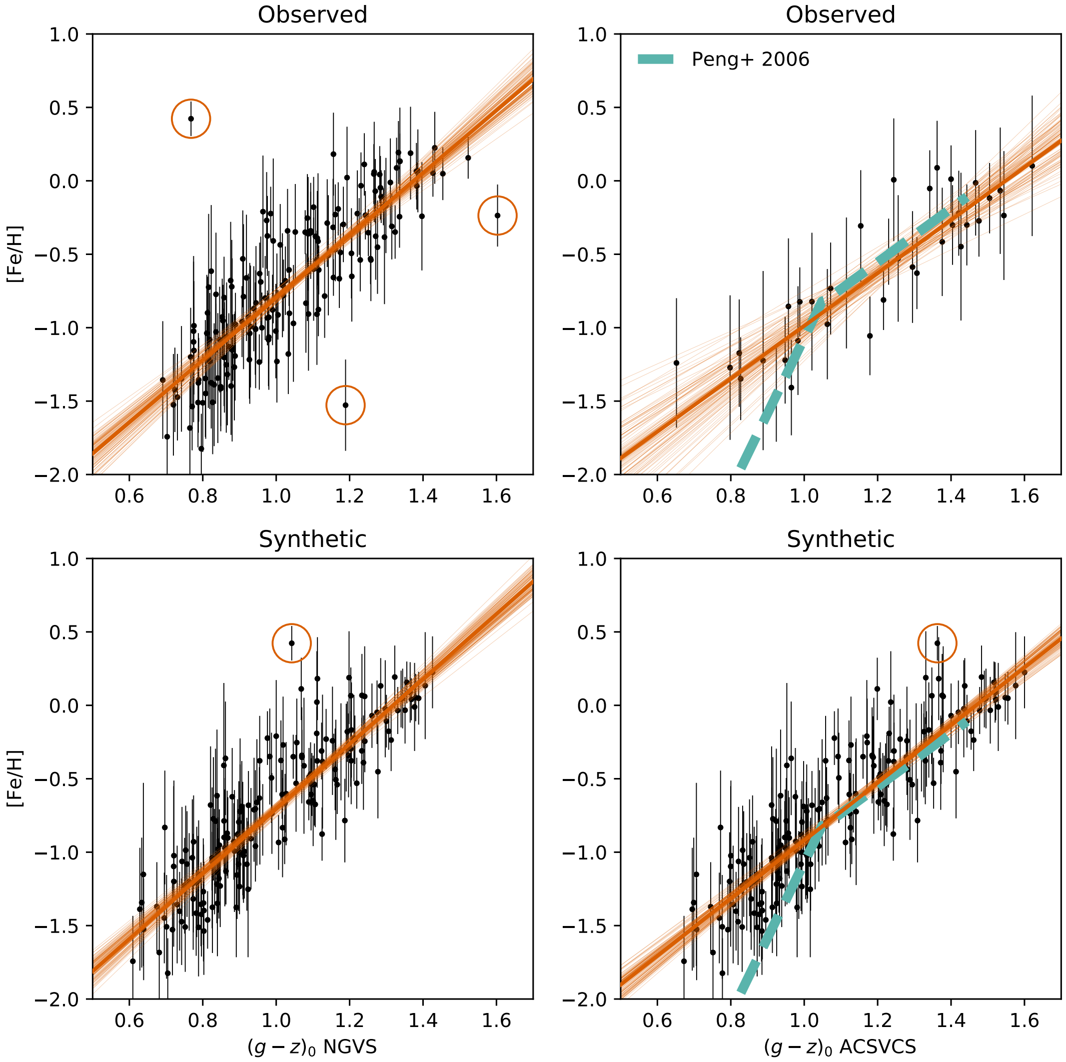

Figure 4.8: (Top-left): Color–metallicity relation using observed NGVS $g-z$ colorsfor the 172 GCs that are in both the spectroscopic sample and NGVS. (Bottom-left): Same as top-left with synthetic colors for all 177 GCs in the spectroscopic sample. (Top-right): Color–metallicity relation using observed ACSVCS colors for the 37 GCs that are in both that and the spectroscopic sample. (Bottom-right): Synthetic color–metallicity relation in the ACSVCS bands for all 177 GCs in the spectroscopic sample. In each panel we show the best-fit line and 100 samples drawn from the posterior distribution by fitting the corresponding data points with a linear model (see text for details). In the right panels we show the Peng et al. (2006) relation (dashed green). The regression algorithm detects outliers in the data which are shown in each plot by the red circles.

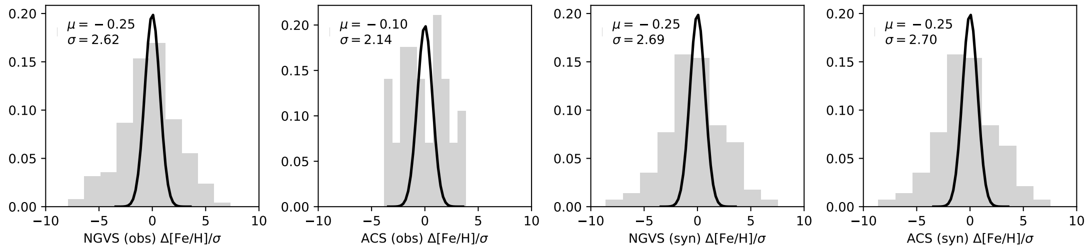

Figure 4.9: Normalized histograms of the residuals between the observed [Fe/H] values and the values predicted by the best-fit color–metallicity relations divided by the observed [Fe/H] uncertainties. We have indicated the mean offset, μ, and standard deviation, σ for the distribution of residuals. A Gaussian distribution with $\sigma = 1$ is also shown.

We used the flux-calibrated models that correspond to the inferred stellar parameters for each individual GC to compute synthetic photometry for both the ACSVCS and NGVS bandpasses. In Figure 4.7 we show the relation between the synthetic photometry using the different filter systems. We also show our best-fit line to the data (excluding the outliers marked with the open circles) so that the colors of GCs can be transformed from one system to the other. GCs identified as outliers by the regression model are marked with open circles. The outliers from this relation are just the result of numerical problems for these particular clusters in generating models over the available wavelength range. As can be seen in Figure 4.7, the overwhelming majority of the GCs follow a tight relation between the ACSVCS filter system and the NGVS system.

In Figure 4.8 we show the color–metallicity relations using the NGVS (left) and ACSVCS (right) photometry for both the observed (top) and the synthetic (bottom) $g - z$ colors. We fit all four color–metallicity relations using linear regression in a Bayesian framework with outlier pruning and uncertainty weighting (see Hogg et al. 2010, for details) and show the best-fit lines for each relation and 100 samples drawn

Table 4.1: Median values of posterior distributions of best-fit line parameters with standard deviations for each fit. We also show the standard deviation of the residual [Fe/H] distributions, $\sigma_{\mathrm{residuals}}$.

	Slope	σ_{slope}	Intercept	$\sigma_{\mathrm{intercept}}$	$\sigma_{\mathrm{residuals}}$
ACSVCS (obs)	1.79	0.25	−2.77	0.31	2.14
ACSVCS (syn)	1.96	0.08	−2.88	0.10	2.70
NGVS (obs)	2.12	0.12	−2.92	0.13	2.62
NGVS (syn)	2.20	0.10	−2.90	0.11	2.69

from the posteriors in each panel (orange lines).

We demonstrate that there is good agreement between the relations using observed and synthetic NGVS photometry. This is important because this assures us of the quality of the synthetic color–metallicity relation for the ACS photometry. The relation using the observed ACSVCS photometry has large uncertainties because of the sparsity of the sample.

Any outliers detected by the fitting algorithm are highlighted by red open circles in each panel. The regression fits do not include those points. Linearity is a good representation of the data in all four cases. We fit the data with a quadratic relation which was not statistically preferred over the linear relation in any case. In Table 4.1 we list the median and standard deviation of slope and intercept values of each relation.

In Figure 4.9 we show the normalized histograms of the residuals between the observed [Fe/H] values and the values predicted by the best-fit color–metallicity relations divided by the observed [Fe/H] uncertainties. In each panel we show a standard normal distribution and indicate in the legend the measured mean and variance of the residual distribution. The residuals have a larger variance than what is expected from a standard normal distribution. This is likely because the color–metallicity relations have genuine spread since GC systems are an amalgamation of different stellar populations.

In the right panels of Figure 4.8 we also show the Peng et al. (2006) relation. Our relation is consistent with Peng et al. (2006) for the red ($g - z > 1.0$) GCs but differs significantly for the blue GCs. We already noted in the previous section that the Cohen et al. (1998) metallicities used by Peng et al. (2006) are more metal-poor as a whole than the metallicities that we have derived for the M87 GCs. Peng et al. (2006) also supplemented their sample with Milky Way GCs.

To understand how the presence of Milky Way GCs might have affected the color–metallicity relation we look at how the Milky Way GCs compare to the M87 GCs in Figure 4.10. We generated synthetic photometry for the Milky Way GCs to obtain ACS $g - z$ colors for the clusters. We show the color–metallicity relation using both the [Fe/H] values we derived from our fits to the Schiavon et al. (2005) spectroscopy (brown circles) and [Fe/H] values compiled from various literature sources (Roediger et al. 2014, open green circles). We also show the M87 GCs (black points). We show the best-fit lines for the Milky Way GC color–metallicity relation (colored lines) and the Peng et al. (2006) relation (dashed black line). In the left panel we show the blue GCs and in the right panel we show the red GCs.

We see in Figure 4.10 that the blue Milky Way GCs have a different color–metallicity relation than the M87 GCs. The color–metallicity relations for the Milky Way GCs are closer to the Peng et al. (2006) relation, which makes sense because it is the Milky Way GCs that drive the blue end of Peng et al. (2006) relation. Moreover, Peng et al. (2006) used the Harris (1996) compilation of Milky Way GC [Fe/H] values

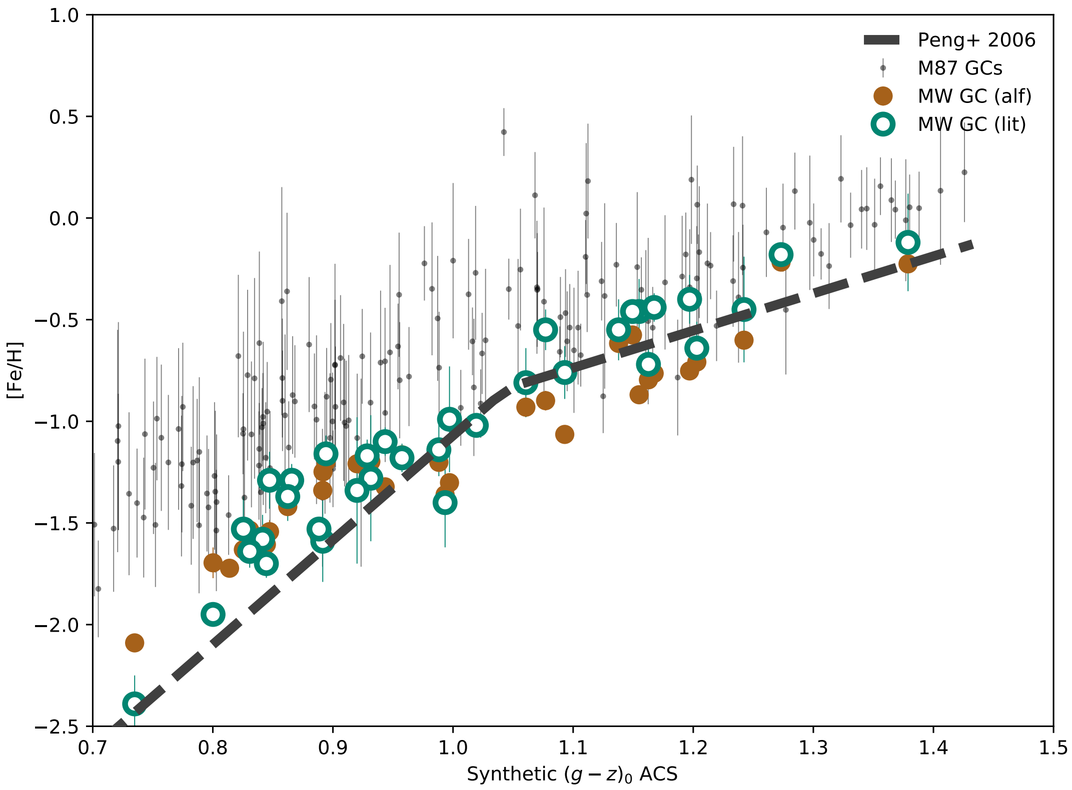

Figure 4.10: We show synthetic ACS $g - z$ color versus metallicity for the M87 clusters (black) and the Milky Way GCs. The inclusion of the MW GCs in the Peng et al. (2006) analysis explains much of the discrepancy between our color–metallicity relations.

and we show that the color–metallicity relation using [Fe/H] values from literature is even closer to the Peng et al. (2006) relation than the relation using the spectroscopically derived [Fe/H] values.

4.3.3 Metallicity Distributions

In Figure 4.11 we demonstrate the effect of our new color–metallicity relations on the derived metallicity distributions. In the left panel we compare the NGVS (yellow) and ACSVCS (green) color distributions. For NGVS we only show clusters within $R_{\rm gal} <$

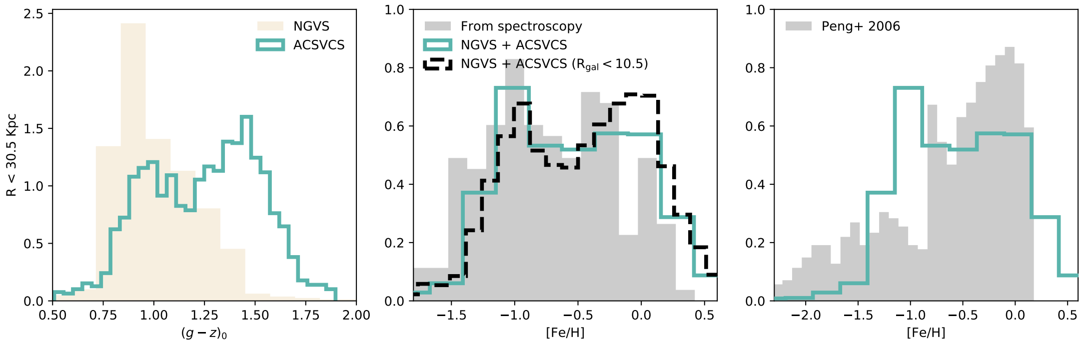

Figure 4.11: (Left): Distributions of the $(g-z)_0$ colors from NGVS (yellow) and ACSVCS (green). The ACSVCS sample is redder and more metal-rich than the NGVS sample on average because it is drawn from a more central region of the galaxy. We limited the NGVS sample to objects within $R_{\text{gal}} < 30.5$ kpc to match the footprint of the spectroscopic sample. (Middle): Comparing distributions of metallicity measured from spectroscopy (grey) and from our color–metallicity relationships including both NGVS and ACSVCS photometry where we truncate the sample to $R_{\text{gal}} < 30.5$ kpc (green) and $R_{\text{gal}} < 10.5$ kpc (black–dashed). Objects in both samples were removed from the NGVS sample. (Right): Comparing the derived metallicity distributions from NGVS and ACSVCS with the metallicity distribution derived from the Peng et al. (2006) relation applied to the ACSVCS colors. The peak of the metal-poor subpopulations are dramatically different, which will affect comparisons to models.

30.5 kpc to match the spatial extent of the spectroscopic dataset. This comparison emphasizes the effect that the spatial extent of the data has on the analysis. In GC systems around massive galaxies, it has been established that the blue GCs begin to dominate further away from the center (e.g., Harris et al. 2017). The ACSVCS sample only extends to $R_{\rm gal} \sim 13$ kpc and we see bimodality clearly in the color distribution for that sample. Meanwhile, the NGVS sample extends more than twice as far out and bimodality gets completely washed out in its color distribution.

In the middle panel we compare the spectroscopically-derived metallicity distribution (grey) with the metallicity distributions derived from the ACSVCS and NGVS samples using their respective color–metallicity relations for two galactocentric radius cut-offs: $R_{\rm gal} < 30.5$ kpc (green) and $R_{\rm gal} < 10.5$ kpc (black–dashed). We removed those GCs that are in both samples from the NGVS sample. The photometrically-derived MDF appears to be consistent to the spectroscopically-derived MDF but gives less noisy view of the MDF. The MDF where we truncate at $R_{\rm gal} = 10.5$ kpc more obviously displays bimodality than the MDF where the sample extends further out.

In the right panel we compare the metallicity distributions derived from the ACSVCS and NGVS colors to the metallicity distribution derived from applying the Peng et al. (2006) color–metallicity relation to the ACSVCS data only (grey). We see that the different color–metallicity relations lead to drastically different MDFs. The peak of the metal-poor subpopulation is more metal-rich in MDF established in this work and the dispersions of both subpopulations are very different between the different MDFs.

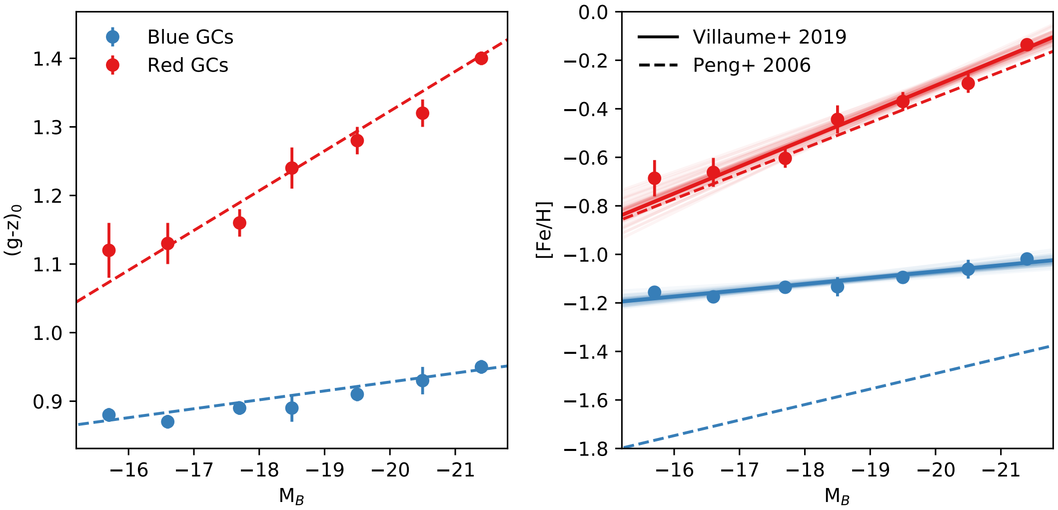

Figure 4.12: (Left) Mean values of the blue and red GC colors as a function of host galaxy luminosity in seven bins of host galaxy magnitude (see Peng et al. 2006, for details). (Right) Mean metallicities of the blue and red GC populations using the color–metallicity relation determined in this work (solid lines) and the best-fit lines from Peng et al. (2006) (dashed lines). The different color–color metallicity established in this work propagates to a dramatically different metal-poor relation.

In Figure 4.12 we see the importance of the new color–metallicity relations derived in this work. In the left panel we show the mean values of the blue and red GC populations as a function of host galaxy luminosity in seven bins of host galaxy magnitude for the Virgo Cluster galaxies included in the Peng et al. (2006) analysis. In the right panel we have used the color–metallicity relation determined in this paper to transform the mean colors established in Peng et al. (2006) into mean metallicities. We derived uncertainties for the metallicity values by doing Monte Carlo sampling of the color–metallicity relation using the color uncertainties.

We show the linear fit to the new relations in the solid lines. We show the relations Peng et al. (2006) determined as dashed lines. As we would expect from the previous results, the new relation between host galaxy luminosity and mean metallicity for the metal-rich GCs is similar to the Peng et al. (2006) result but the relation for the metal-poor GCs is shallower and more metal rich than the Peng et al. (2006) result.

4.4 Discussion

4.4.1 Which Metallicity is it Anyway?

The difference between our and the Peng et al. (2006) color–metallicity relationship is substantial for the blue GCs. We can understand this difference by examining the origin of the [Fe/H] values Peng et al. (2006) used in their analysis. First, the Milky Way GCs make up the majority of the blue GCs used in the Peng et al. (2006) sample. We demonstrated in Figure 4.6 that the Milky Way GCs are more metal-poor than the M87 GCs. In Figure 4.10 we show that, using both literature [Fe/H] values and [Fe/H] values derived from full-spectrum fitting, the Milky Way GCs have a different color–metallicity relation than the M87 GCs. The closeness of the Peng et al. (2006) relation in the blue to the Milky Way GC relation is highly suggestive that the presence of the Milky Way GCs is driving and biasing the relation in the blue for Peng et al. (2006).

Second, we show in Figure 4.6 that even though the GCs in our sample and the Cohen et al. (1998) sample span a similar color range, the Cohen et al. (1998) metallicity values are systemically lower than the metallicities we derive. There are no GCs that are shared between the Cohen et al. (1998) sample and our sample but we can understand the differences between the two by bearing in mind two related facts: the fitting functions that underlie the Worthey et al. (1994) models are not well-calibrated at high metallicities and the Cohen et al. (1998) metallicities are placed on the Zinn (1985) metallicity scale which is set by Milky Way GCs.

The former was discussed in Cohen et al. (2003a) as a serious concern. Cohen et al. (2003a) redid the [Fe/H] determinations of the M87 GCs from Cohen et al. (1998) by extrapolating the models to higher metallicity by assuming that the indices are on the damping part of the curve of growth. This affected five M87 GCs in their sample. We are, to be clear, using the Cohen et al. (1998) metallicities in this work as Peng et al. (2006) did.

For the latter, Cohen et al. (1998) noted that from their qualitative analysis of the line indices of both the Milky Way and M87 GCs, the M87 GCs have a metal-rich tail that extends to significantly higher metallicities than the Milky Way GCs, which we confirm. The relation they use to scale the Worthey et al. (1994) models to the Milky Way GCs is $[\mathrm{Fe/H}]_{\mathrm{Z}} = 0.760 \times [\mathrm{Fe/H}]_{\mathrm{W}} - 0.265$ which would lower the overall metallicity of their sample. Overall, we see that the Peng et al. (2006) relation is yoked to the Milky Way GCs in both explicit and implicit ways. The [Fe/H] values that we present in this work come from the underlying isochrones (Choi et al. 2016) and the underlying stellar library (Villaume et al. 2017b). While the stellar library consists of stars from the Milky Way, there is not a Milky Way specific trend in [Fe/H] that we need to correct like we would for α elements (Tripicco & Bell 1995).

We also find that the color–metallicity relation differs between the Milky Way and M87, especially near the blue end ($g - z \lesssim 1.0$). We defer an in-depth examination of the physical cause of this difference to later work but we speculate that it might be age differences between the two GC populations. If the M87 GCs were younger than the Milky Way GCs, they would appear bluer at the same metallicities. We used simple stellar population (SSP) synthesis models to examine how age affects color at fixed metallicity (in this case, $[\mathrm{Fe/H}] = -1.5$) and found that the metal-poor M87 GCs would have to be about 4 Gyr younger than the Milky Way GCs to explain the color difference. We also cannot rule out the possible effects that α elements or the morphology of the blue horizontal branch have on the color.

4.4.2 Bimodality

Bimodality of GC systems has been the dominant paradigm in which extragalactic GC studies have been conducted over the past 30 years. In this paper we defer quantitative analysis of the subpopulations of the GC system around M87 to a forthcoming paper on the subject. This is to more appropriately address the complexities around the topic that have been raised recently. Even with just the M87 system, consensus has yet to be reached on the number of subpopulations that make up the system (e.g., Strader et al. 2011b; Agnello et al. 2014; Oldham & Auger 2016a). With that being said, there are still some things worth pointing out.

First, Cohen et al. (1998) detected bimodality in M87 only after excluding the metal-rich tail from their analysis. Usher et al. (2012) speculated that the lack of convincing proof for bimodality from Cohen et al. (1998) was a result of their typically bright targets. Since Cohen et al. (1998) we have become aware of the blue-tilt phenomenon as well as liminal objects like ultra-compact dwarfs that could contaminate populations of bright canonical GCs (e.g., Usher et al. 2012; Villaume et al. 2017a).

Second, we take advantage of obtaining color–metallicity relations using both the NGVS and ACSVCS datasets by converting both into metallicity and combining the data sets. The ACSVCS data probe the very inner region of the M87 GC system while the NGVS data extends further out. We see that the color-converted MDF is consistent with the spectroscopically determined MDF. Furthermore, bimodality can be seen visually from the MDFs, especially when only GCs within $\mathrm{R_{gal}} < 10.5$ kpc are included.

The M87 system consists of a huge number of GCs that represent the culmination of a complex history. Previous analyses (e.g., Strader et al. 2011b; Romanowsky et al. 2012) indicated that the GCs in the outer halo behave differently and are dominated by blue/metal-poor GCs. As mentioned previously, a paper specifically addressing the subpopulations and their characteristics will follow this paper.

Third, the Yonsei Evolutionary Population Synthesis (YEPS) models have been used to argue that most bimodal color distributions reflect a truly unimodal underlying metallicity distribution because of the inclusion of hot horizontal branch stars (Yoon et al. 2006). The approach of this group is different from the one typically taken, where spectroscopic observations of individual GCs are modeled with SPS models. Instead, the YEPS group transforms the color distributions of GC systems to metallicity distributions using synthetic color–metallicity relations generated from the YEPS models.

The results from the synthetic color–metallicity relation method (Yoon et al. 2006; Lee et al. 2019) and the direct spectroscopic modeling (e.g., Alves-Brito et al. 2011;

Usher et al. 2012; Brodie et al. 2012) method continue to be at odds. The results we find in this work are consistent with other studies that have directly modeled spectroscopy of individual GCs. Beyond the final results there are few points of comparison between the two methods. However, we note that from our work with Milky Way GCs we know that the presence of hot horizontal branch stars affects our ability to measure accurate ages from integrated light but not metallicity (see Figure 15 in Conroy et al. 2018, for reference). We therefore do not have a reason to doubt our metallicity measurements for the M87 GCs, even with the possible presence of GCs with prominent hot horizontal branches.

Fourth, it is important to note that our MDFs, both from the purely spectroscopic sample and the sample converting NGVS and ACSVCS photometry, differ significantly from the MDF computed from the Peng et al. (2006) relation. The peaks and widths of the distributions are different. These quantities are crucial for making quantitative comparisons to theoretical models of GC system formation, and thus, of galaxy formation. In recent years modern theoretical galaxy formation models have emerged with the E-MOSAICS simulations (Pfeffer et al. 2018) for Milky Way-type galaxies, and alternatively with Choksi et al. (2018) specifically trying to recreate the observed properties of the Virgo Cluster galaxies. These models take divergent approaches: E-MOSAICS adds models describing the formation and evolution of star clusters into the EAGLE galaxy formation simulations, while Choksi et al. (2018) uses semi-analytic models of merger histories. They also take different approaches to the role GC destruction plays in our understanding of $z \sim 0$ GC systems. Accuracy and credible uncertainties in the physical characteristics derived from observables are crucial for moving forward with constraining galaxy formation theories based on GCs.

4.4.3 Implications for GC and Galaxy Formation

We have derived a new galaxy luminosity–GC metallicity relation separately for the blue and red GCs in the Virgo galaxies included in Peng et al. (2006) (Figure 4.12). The difference in our new color–metallicity relation is two-fold: the metal-poor GCs now correlate with galaxy luminosity *less strongly* than previously measured, and the metal-poor GCs are more metal-rich than what Peng et al. (2006) determined.

Larsen et al. (2001) were the first to assess the relationship between GC *sub-population* metallicity and galaxy luminosity with a homogeneously acquired sample. Then Strader et al. (2004) combined elliptical galaxy data from a variety of sources (Larsen et al. 2001; Kundu & Whitmore 2001a,b) with data from spiral galaxies (Harris 1996; Barmby et al. 2000) to look at just the metal-poor GCs. Most recently Peng et al. (2006) determined this relationship for the Virgo Cluster galaxies. Like Larsen et al. (2001), Peng et al. (2006) found shallower slopes for the metal-poor GCs relative to the metal-rich GCs. There is remarkable similarity between the slopes that Larsen et al. (2001), Strader et al. (2004), and Peng et al. (2006) found for the metal-poor GCs.

We already know that the difference between our relation and the relation from Peng et al. (2006) is due to the color–metallicity relation. What about the difference with Strader et al. (2004)? Strader et al. (2004) used the Barmby et al. (2000) color–metallicity relation based on a sample of M31 GCs. Barmby et al. (2000) noted that the M31 color–metallicity relation is similar to the Milky Way relation. This raises the likelihood that it is not an appropriate way to convert colors to metallicities for the early-type galaxies included in the Strader et al. (2004) sample. The similarity in slopes between Strader et al. (2004) and Peng et al. (2006) might be an artifact of the similar source of their respective color–metallicity relations.

To explain the correlation between galaxy luminosity and blue GC metallicity Strader et al. (2005) and Brodie & Strader (2006) invoked the concept of "biasing", also introduced in the context of Milky Way stellar halo assembly by Robertson et al. (2005). In short, the progenitor satellites that now constitute the stellar halos of massive galaxies were more metal-rich, at fixed mass, than present day satellites. In the light of the new, much weaker correlation, this needs to be reassessed. The new correlation could indicate that biasing is not as strong as an effect as once thought. Put another way, the new correlation suggests that metal-poor GCs formed irrespective of their host galaxies.

The change in metallicity intercept for the metal-poor GCs on this relation has implications for their formation epoch. Forbes et al. (2015) evolved the galaxy mass–GC metallicity relation through redshift to determine bulk ages of the GCs belonging to the galaxies in the SLUGGS survey (e.g., Usher et al. 2012). In their model, higher metallicities indicate younger ages and/or more massive hosts. The slopes of their metal-poor and metal-rich relations are not totally consistent with what we present in this work, but the intercepts are roughly similar. Following the logic of Forbes et al. (2015), the nearly constant slope we find for the metal-poor GCs as a function of galaxy luminosity indicates that the metal-poor GCs in the Virgo Cluster formed at nearly the same time. The correlation between the metal-rich GC [Fe/H] values and host galaxy luminosity indicates that the metal-rich GCs around the giant galaxies formed more recently than the metal-rich GCs around the dwarf galaxies. The increase in metallicity for the metal-poor GCs may also help ease the tension between simulated and observational results as discussed in the Introduction, if it indicates that the GCs formed in more massive satellites.

It is important to note the crucial underlying caveat of Figure 4.12 – that the color–metallicity relation we developed for M87 is applicable to the other Virgo Cluster galaxies in the Peng et al. (2006) analysis. This is probably not a good assumption, particularly in light of the Powalka et al. (2016) results which showed that color–color relations in the NGVS sample depend on environment, with colors on the whole becoming bluer with increased radial distance from M87 and that GCs $>$ 200 kpc from M87 have color–color distributions similar to those of the Milky Way. Unfortunately, Powalka et al. (2016) also showed that mass is not the driving factor in these differences so we cannot make a simple correction to Figure 4.12. More detailed spectroscopy of lower-mass systems in the Virgo Cluster is ultimately needed.

4.5 Summary

- We have fitted a spectroscopic sample of GCs around M87 with full-spectrum SPS models and obtained [Fe/H] for 177 GCs. We demonstrate that the metallicity values we derive are systematically higher-metallicity than previous spectroscopic studies. We attribute this difference to the limitations of the previously-used

Worthey et al. (1994) SPS models and because the previously determined metallicity values were scaled to match the Milky Way GCs, which are, as a whole, lower in metallicity than the M87 GCs.

- Using synthetic photometry from flux-calibrated stellar population models we determine a transformation between the NGVS and ACSVCS photometric systems: $(g-z)_{\mathrm{ACSVCS}} = 1.123(g-z)_{\mathrm{NGVS}} - 0.015$.

- We derived new color–metallicity relations using both NGVS and ACSVCS $g-z$ colors. Our ACSVCS color–metallicity relation differs significantly for the blue GCs from the previously published color–metallicity relation using the ACS filters. This is because we find the relation for the Milky Way GCs to be significantly different than the relation for the M87 GCs. We discuss the necessary age difference needed to explain this result, but previous work in colors of Virgo Cluster GCs suggested that there is some environmental effect on chemical abundance patterns.

- While we advocate that color–metallicity relations be confirmed with spectroscopic follow-up for individual galaxies, we assume that in this respect the Virgo cluster galaxies are similar to one another and as a result we find a shallower galaxy luminosity-GC metallicity relation for the blue GCs than previous studies. This could either indicate that progenitor satellites were less massive than previously thought, or the properties of metal-poor GCs are not as dependent on their present-day host galaxy as metal-rich GCs.

Chapter 5

Mapping the Assembly History of M87 Through Radial Variations in Chemical Abundances

5.1 Introduction

Any complete theory of galaxy evolution must explain massive early-type galaxies (ETGs). In the Local Group, theories of how the Milky Way and Andromeda (M31) galaxy formed have been benchmarked by our understanding of their chemodynamical properties. With the advent of modern surveys (e.g., *Gaia*, APOGEE; Prusti et al. 2016; Majewski et al. 2017b) our picture of the structure of the Milky Way has become more detailed and from this we have made progress in quantifying the formation and accretion histories of both galaxies (e.g., Helmi et al. 2018; Gilbert et al. 2019).

In contrast, our understanding of the structure of massive ETGs has remained relatively simple and traditionally defined by their uniform "red and dead" appearance. Without access to the resolved stellar populations in these galaxies – none exist in the Local Group – our understanding of the formation and evolution of these galaxies has necessarily relied on coarser information such as scaling relations between global properties of large samples of such galaxies (e.g., the Fundamental Plane) over cosmic time and trends of integrated stellar population parameters in present-day ETGs (Renzini 2006). However, the power of spatially-resolved stellar population parameters to falsify formation scenarios has long been recognized (e.g., the early debate of dissipationless versus dissipative collapse reviewed in Faber 1977).

An early picture of ETG formation was that they were simply major-merger products of Milky Way-type galaxies (Toomre 1977). More recently, high-redshift observations of ETG progenitor compactness have motivated the "two-phase scenario" (e.g., Bundy et al. 2009; Oser et al. 2010). In this scenario the first phase is characterized by rapid star formation, creating the core of a given galaxy that hosts "*in-situ*" stars. The second phase is characterized by size growth over an extended period of time via galaxy mergers that bring in an "*ex-situ*" population of stars into a galaxy. This framework is widely used but there is still no consensus on the details of these phases and, thus, what the present-day stellar populations of these galaxies should be like (e.g., Somerville & Davé 2015; Naab & Ostriker 2017).

The most massive ETGs, including brightest cluster galaxies (BCGs) are expected to be the most extreme examples of the second phase – highly merger dominated.

Deep imaging surveys (e.g, Burrell-Schmidt Survey, Hyper Suprime-Cam; Mihos et al. 2017; Aihara et al. 2018) have revealed complex structure in these galaxies. There has been limited ability to accurately test with observations the picture laid out by simulations. To achieve the same depth of understanding for massive ETGs as we have for the Local Group galaxies we need a way to probe their chemodynamical properties.

Because are bright, compact, ubiquitous above a certain galaxy mass, and generally thought to be uniformly old globular clusters (GCs) have an extensive history of constraining galaxy formation scenarios. Generally this has been done through scaling relations between GC system properties (e.g., mean metallicity, number) and global properties of their host galaxy. For example, it was detailed understanding of GC populations across galaxy types that provided early observational evidence contradicting the spiral-spiral merger scenario (van den Bergh 1982).

GCs have also been used as "discrete tracers", primarily to study kinematics. Another important property of GC, though, is that they are near-simple stellar populations. This means they can also be used as discrete tracers of the stellar populations of a galaxy. This has mainly been exploited through the study of the metallicity distribution functions (MDFs) of GC systems. One of the early breakthroughs in extragalactic GC studies was the discovery of pervasive bimodality in the colors of GC populations (Zepf & Ashman 1993; Larsen et al. 2001). The "red" GCs are thought to be metal-rich and formed within massive galaxies at the peak of star formation ($z \sim 2$–3), while the "blue" GCs are metal-poor objects that formed in low mass galaxies at even earlier times and were then accreted into galaxy halos (Forbes et al. 2015).

The vast majority of the work done to understand extragalactic GC systems has been done via broadband photometry (e.g., Larsen et al. 2001; Côté et al. 2004). The most extensive *spectroscopic* work to date has been through the SLUGGS Survey (Brodie et al. 2014, SAGES Legacy Unifying Globulars and GalaxieS) which obtained spectra of ~ 4000 GCs around 27 nearby ETGs out to $\sim 10R_e$ (Forbes et al. 2017)[10]. The main focus, however, has been on kinematics of the GC systems, and the important chemical abundance information has yet to be fully leveraged.

The BCG of the Virgo Cluster, M87 (NGC 4486), has one of the most extensively studied GC systems (e.g., Baum 1955; Cohen et al. 1998; Forte et al. 2012; Oldham & Auger 2016b), both photometrically (Strader et al. 2011b) and kinematically (Romanowsky et al. 2012). In Villaume et al. (2019) we obtained spectroscopic iron metallicity values for a large, representative sample of M87 GCs using full-spectrum stellar population synthesis (SPS). In this work, we expand our chemical analysis of the GC system with a new statistical framework in which to simultaneously separate individual GCs in the M87 system into subpopulations and measure the metallicity gradient of the subpopulations more accurately than previous studies. We also make physically motivated spectral stacks from the individual GC spectra as a way to measure detailed abundance patterns of the GC system. We make a comprehensive analysis of the metallicity gradients, the detailed abundance patterns, and the kinematics of the M87 GC system to qualitatively discuss the possible origins of the GC subpopulations, and thus, the assembly of M87.

We take the distance to M87 to be $D_L = 16.5$ Mpc, with effective radius

---

[10] `http://sluggs.swin.edu.au/Start.html`

$R_e = 16.0$ kpc (Kormendy et al. 2009), and $\log(M_*/M_\odot) = 11.61 \pm 0.10$ (Oldham & Auger 2016c). In Section 2 we describe the spectroscopic samples and the stellar population synthesis models we use to extract abundance information from both M87 and its GC system in a consistent way. In Section 3 we outline the statistical framework we use for the subpopulation membership identification for the individual GCs and the measurement of the [Fe/H] gradients. In Section 4 we present the results from the SPS modeling and gradients, and in Section 5 we discuss the results in the context of the formation history of M87.

5.2 Spectroscopic Data and Abundance Analysis

5.2.1 Obtaining the stellar population parameters

We model the spectra with an updated version of the absorption line fitter (`alf`, Conroy & van Dokkum 2012b; Choi et al. 2014; Conroy et al. 2014, 2018)[11] that uses the Extended IRTF Library (E-IRTF; Villaume et al. 2017b) and the MIST isochrones (Choi et al. 2016). With `alf` we can model the full continuum-normalized spectrum of integrated light for stellar ages > 1Gyr and for metallicities ~ -2.0 to $+0.25$. The full model has 36 free parameters (see Table 2 in Conroy et al. 2018). The parameter space is explored using a Markov Chain Monte Carlo algorithm (`emcee`; Foreman-Mackey et al. 2013b). In this work we use the priors as described in Conroy et al. (2018) and fix the IMF to the Kroupa (2001) form.

Theoretical elemental response functions that tabulate the effect on the spec-

---

[11] https://github.com/cconroy20/alf

trum of enhancing each individual element modeled in `alf` were computed with the ATLAS and SYNTHE programs (Kurucz 1970, 1993). For the α elements relative to Fe considered in our analysis (Mg, Si, and Ca) we correct for the underlying abundance pattern in the empirical stellar library using the [Mg/Fe] values from Milone et al. (2011) and [Ca/Fe] values from Bensby et al. (2014). We assume Si has the same library abundance pattern as Ca.

We analyze several different data sets in this work (see below). To make the different samples as homogeneous as possible we fitted over the same spectral range for every spectrum analyzed in this work: $4000 < \lambda Å < 4400$ and $4400 < \lambda Å < 5225$.

While we obtain estimates of the light-weighted age as part of the `alf` models, we do not include age in our analysis. This is because of the uncertain effect of the blue horizontal branch, particularly in the GCs, which can make the inferred ages artificially young. Our analysis of the Milky Way GC system indicates that iron metallicity can still be reliably recovered in the presence of a blue horizontal branch (see Conroy et al. 2018).

5.2.2 The globular clusters

Strader et al. (2011b) carried out a wide-field kinematic analysis of the M87 GC system using two key data sets: the Keck/LRIS sample and the MMT/Hectospec sample. In Figure 5.1 we compare the two samples in color–magnitude space. The Keck/LRIS sample (green) was selected to sample the low luminosity population over the full color range of the GC system, in contrast to previous work that targeted high-luminosity clusters that likely have different properties from the bulk of the GCs (Cohen

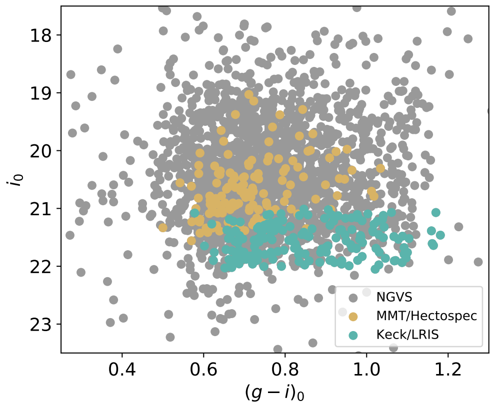

Figure 5.1: Comparing the coverage of the Keck/LRIS (green) and MMT/Hectospec (yellow) samples in color-magnitude space. Also shown is the NGVS sample (Oldham & Auger 2016b, grey). The MMT/Hectospec sample is overall more luminous than the Keck/LRIS sample and has more blue GCs than red, while the Keck/LRIS sample is evenly distributed over color space.

et al. 1998). The MMT/Hectospec sample (yellow) was selected from the higher luminosity population. The MMT/Hectospec objects were also primarily selected at large radii to aid the sky subtraction since Hectospec is a fiber instrument.

In Villaume et al. (2019) we applied full-spectrum stellar population synthesis (SPS) models to the Keck/LRIS dataset of M87 GCs (Strader et al. 2011b) to obtain estimates of iron metallicity ([Fe/H]) relative to solar. We refer the readers to the original paper for details on the modeling and validation of the [Fe/H] values for the Keck/LRIS sample. Here, we do the same analysis for the MMT/Hectospec sample. We used the square root of the summed sky spectrum and flux generated by the reduction pipeline as the uncertainty on the individual GCs. The S/N of this data set ranges from S/N $\sim 1-30\text{Å}^{-1}$ with a resolution of 5Å. The resolution of the data is higher than the native resolution of the models so we smoothed the data to 200 km/s to be consistent

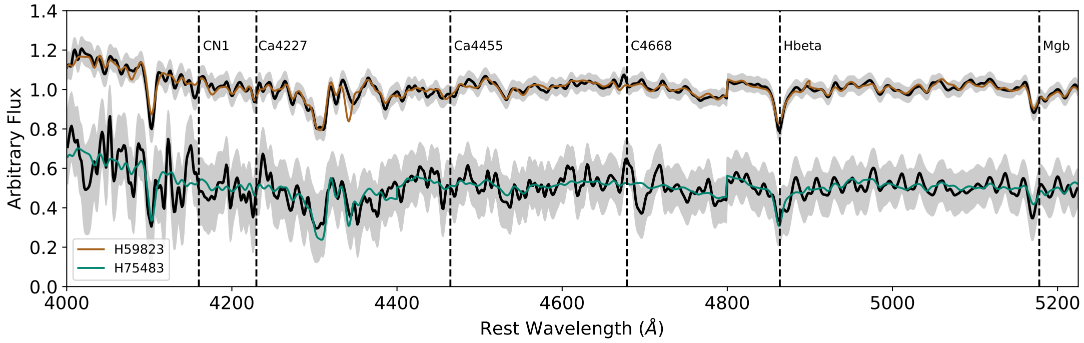

Figure 5.2: Comparison of MMT/Hectospec GC spectra (black) and best-fit models for a comparatively high-S/N spectrum (S/N$\sim$ 30, brown) and a low-S/N spectrum (S/N$\sim$ 10, green). Grey band is the uncertainty of the flux from the input spectrum. Within the uncertainties, the fits are successful.

with our previous analysis with the Keck/LRIS data.

Before we smoothed, we identified particularly bad sky lines in the spectra at $4040 < \lambda\text{Å} < 4050$, $4355 < \lambda\text{Å} < 4365$, and $5458 < \lambda\text{Å} < 5470$ and interpolated over the flux in each spectrum in those wavelength regions. We fitted 156 spectra and rejected 12 spectra from our analysis based on visual inspection of the residuals between the observed spectra and best-fit models. We show successful fits in Figure 5.2 for a comparatively high-S/N spectrum (S/N$\sim$ 30, brown) and a low-S/N spectrum (S/N$\sim$ 10, green) with spectral features of particular interest highlighted. The black line and grey band represent the data flux and uncertainty, respectively.

For the individual GCs, we focus our analysis on [Fe/H] and summarize our measurements in Table 5.1. The majority of the GC spectra do not have sufficient S/N to reliably extract more detailed abundance information. In Section 4.3 we describe how we stacked the individual GC spectra and fit the stacks with `alf`.

ID	RA	DEC	[Fe/H]	$\sigma_{[\mathrm{Fe/H}]}$	Instrument
H47487	187.73553	12.32802	−1.16	0.30	LRIS
H49585	187.67674	12.32961	−0.51	0.41	LRIS
H49328	187.71423	12.32992	−0.78	0.24	LRIS
...					
H47487	187.59446	12.02249	−0.80	0.37	Hectospec
H49585	187.52539	12.03362	−0.40	0.23	Hectospec
H49328	187.91104	12.04288	−1.17	0.39	Hectospec

Table 5.1: Table of summary statistics of the [Fe/H] measurements for the GCs included in this work. (This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

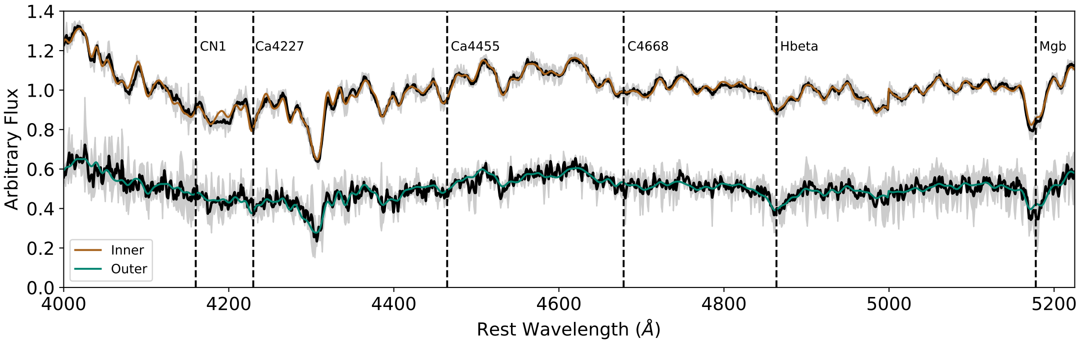

Figure 5.3: Same as Figure 5.2 but for Mitchell spectra close to the center of the galaxy ($R_{gal} = 1.32$ kpc, brown) and from the outer region ($R_{gal} = 19.5$ kpc, green).

5.2.3 The galaxy light

The stellar populations of the galaxy light of M87 itself have remarkably never been studied with spectroscopy beyond the central few kpc. Here we use data from the Mitchell (formerly VIRUS-P) integral field unit (IFU) spectrograph at McDonald Observatory (Murphy et al. 2011; spectroscopy obtained via private communication with K. Gebhardt). The signal-to-noise (S/N) of the individual spectra ranges from $\sim 20 - 50\text{Å}^{-1}$. We stacked the individual spectra in 10 bins of galactocentric radius by bootstrapping for the median of the individual spectra in a given bin. We used the 50th

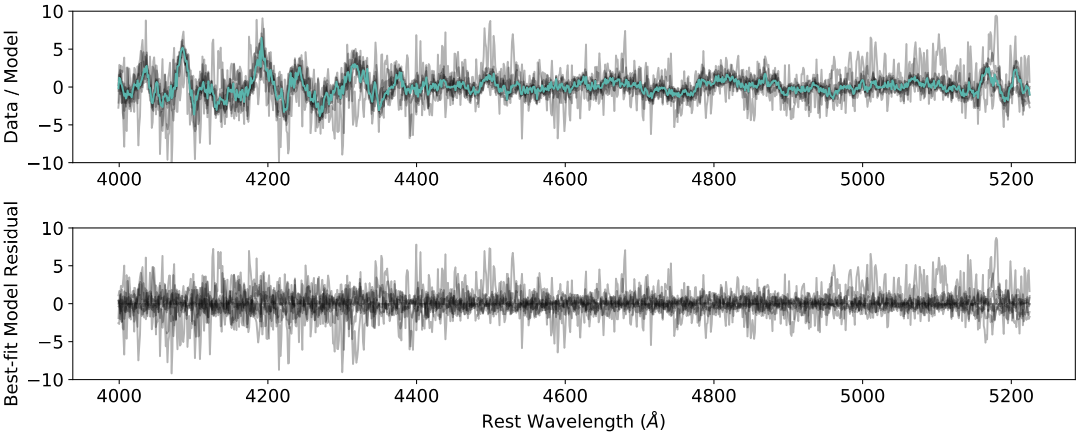

Figure 5.4: (Top) Residuals from dividing the best-fit models from the corresponding Mitchell integrated galaxy light (black) and the median residual for all spectra (green). The residuals are nearly identical for all spectra and the large wavelength-scale features are likely systematic to the models and not dependent on stellar parameters. (Bottom) Residuals after subtracting the median residual.

percentile from the resulting distribution of flux at a given wavelength as the stacked spectrum and used the average of the 16th and 84th percentiles as the uncertainties on the stacks with the S/N of the stacks ranging from $\sim 40 - 200$, with the outermost spectrum having the lowest S/N.

In Figure 5.3 we examine the quality of our fits for spectra in the inner ($\mathrm{R_{gal}} \sim$ 1.32 kpc, S/N$\sim$ 200) part of the galaxy and the outer part ($\mathrm{R_{gal}} \sim$ 19.4 kpc, S/N$\sim$ 40). We compare the Mitchell spectra (black) with the best-fit model spectrum for the inner region (brown) and the outer region spectrum (green) with selected spectral features highlighted. The grey bands are the flux uncertainty from the data. In Figure 5.4 we examine the residuals between the best-fit model and input data for all the Mitchell spectra used in this analysis. The residuals are typically small ($< 5\%$).

5.3 Characterizing Globular Cluster Systems Via Statistical Modeling

Our goal is to develop a method to measure the properties of GC systems as a way to understand the formation history of M87 and other galaxies. In this work, we are focused on measuring the metallicity gradients and abundance patterns of the M87 GC system. Obtaining a metallicity gradient might seem as simple as fitting a line to data, but a recent meta-analysis of many of the studies that have measured metallicity gradients of GC systems revealed a troubling result – different studies often get significantly different answers for the *same* GC systems (see Figure 1 in Forbes & Remus 2018). Several underlying issues could be causing an accuracy problem in these studies, which motivates us to characterize GC systems in a novel way using a hierarchical Bayesian model (HBM). In the following, we detail these issues and describe how HBMs provide a natural means to overcome them.

First, the studies included in the Forbes & Remus (2018) analysis all used a version of linear least-squares to fit the gradients of GC studies. However, linear least-squares only works if one of the dimensions of data has negligible uncertainties. These studies also assumed that the galactocentric distances of the GCs are perfectly known. This is not the case, however, since only 2D *projected* distances are known. The distances can be de-projected if the 3D density distributions of the GCs are known (e.g. McLaughlin 1999), but this is not the case for the vast majority of extragalactic GC observations. As discussed in Liu et al. (2011), using the projected distances as a substitute for true distance introduces systematic uncertainty into the measured gradients because the GCs projected into the center will, in reality, be a mix of GCs at all radii. Liu et al. (2011) estimated that this could lead to an uncertainty of $\sim 10\%$ in the measured gradients, but in reality, this depends on the degree of the true underlying slope. We must take that uncertainty into account when interpreting the measured gradients.

Second, characterizing GC systems is further complicated because these systems, especially around massive galaxies, are the aggregate of many different stellar populations. The constraints on GC system assembly and galaxy formation depends on our ability to differentiate and understand the *subpopulations* of a GC system. Broadly, GCs are separable into “metal-poor” and “metal-rich” populations. In detail, however, it is not trivial to separate the individual GCs into subpopulations.

Previous work measuring the metallicity gradients of GC systems has primarily used constant cuts on color to separate the metal-poor and metal-rich subpopulations (e.g., Harris 2009a,b; Liu et al. 2011; Hargis & Rhode 2014; Kartha et al. 2016). However, wide-field photometric surveys have demonstrated that the demographics of GC populations change with increasing distance from the center of the galaxy (e.g. Strader et al. 2011b; Harris et al. 2017), with the relative number of blue GCs typically increasing. As a result, a constant cut across the GC sample could bias the gradient measurements (see later in this section for demonstration of this effect). A few studies have attempted to mitigate this issue by separating the GC subpopulations at different radial steps (e.g., Blom et al. 2012; Usher et al. 2013). However, these studies did not measure the gradient for their full samples but only considered the peaks of the metallicity distribution functions (MDFs) when computing the gradients. Moreover, by cutting on subpopulation *and then* determining subpopulation characteristics, all these studies fail to account for the covariance between subpopulation membership assignments and whatever parameter of interest is being measured. This, again, will bias the gradient measurements.

Finally, linear least-squares is highly sensitive to the presence of outliers in a sample. The studies included in the Forbes & Remus (2018) analysis used photometric samples of GCs with colors as a proxy for stellar metallicity, except for Pastorello et al. (2015) who also had calcium triplet (CaT) determined metallicities. Without spectroscopic follow-up to confirm GC candidates in photometric surveys, any study based on such data will be affected by contaminant populations. Furthermore, the color–metallicity relations that are used to convert broadband colors of GCs into iron ([Fe/H]) metallicities have been recently called into question (Usher et al. 2012; Villaume et al. 2019). In this work, we use only spectroscopically-determined [Fe/H] measurements of the individual GCs.

HBM provides a means to address and mitigate these issues. Specifically, HBM is a natural way to fit the galactocentric metallicity gradients of GC systems for a number of reasons:

- The Bayesian framework allows us to model unobserved (latent) parameters. This means we can directly model and fit any intrinsic scatter in the metallicities as an explicit parameter and *marginalize* over the unknown 3D distribution of the GC system to mitigate the bias from the projected distances.

- We do not need to make *a priori* cuts to obtain the subpopulations. Instead, we can fit the linear metallicity gradients jointly with the subpopulation memberships, allowing us to capture the covariance between the subpopulation slopes and the subpopulation memberships. This helps us obtain more accurate subpopulation membership assignments for the individual GCs and, thus, more accurate metallicity gradients.

- Relatedly, instead of making a binary cut with the subpopulation assignments, we get probabilistic memberships. We can propagate the uncertainties on the subpopulation membership assignments throughout this work. This is especially important because we are also interested in the detailed abundance patterns of the GC subpopulations. Currently, the signal-to-noise (S/N) of the spectroscopy does not allow for reliable estimates of abundances for the majority of the individual GCs in our sample, so creating spectral stacks with reliable uncertainties is crucial.
- Moreover, with HBM, like all Bayesian methods, we produce posterior distributions for *all* the model parameters. In practice, this gives us trustworthy and interpretable uncertainties on the gradient measurements.

In short, HBM provides us with results that are more accurate and interpretable and with uncertainties better representative to reality than previous studies. In the rest of this section, we develop a method that allows this full propagation of uncertainty from the measurements to the inferences made about the subpopulation distributions and demonstrate its efficacy with mock data.

In the following Section we provide a rather pedagogical explanation of our model as a way to introduce HBM. For those already familiar with this statistical

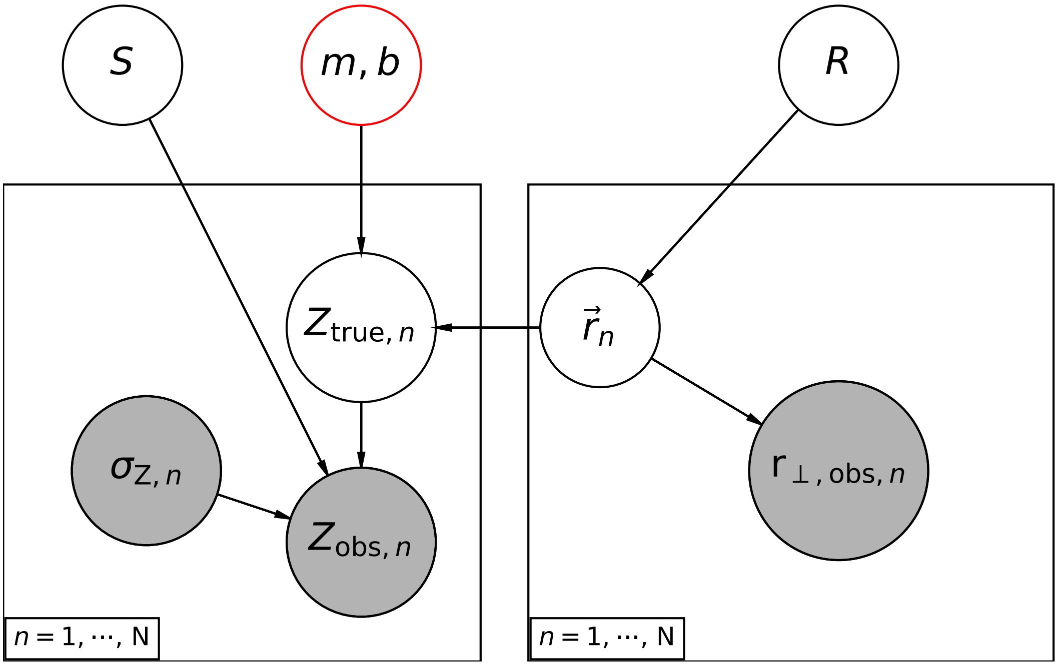

Figure 5.5: The graphical representation of our single population model that we use to factorize the joint distribution of our model. We condition on the observations (grey) to make inferences about the latent parameters (open circles) parameters of interest, the slope, m, and intercept, b (red circle). The rectangle ("plate") represents the structure of the individual parameters and data that is repeated for all of the GCs in our sample ($n = 1, ..., N$). The arrows show the direction of conditional dependence among the parameters. See Section 3.1 for details on the parameters.

technique, our full model is collectively summarized in Table 5.2, Figure 5.10, and Equation 7.

5.3.1 A model for a single population

We begin with a model for a single population of objects as a way to demonstrate some of the key reasons for using a HBM framework in a simplified setting. Bayesian inference is an application of Bayes' theorem[12],

---

[12] https://doi.org/10.1098/rstl.1763.0053

Probability density distributions	Prior	Description
$p(Z_{\mathrm{obs,n}} \vert Z_{\mathrm{true,n}}, S, \sigma_n)$	Normal($Z_{\mathrm{true,n}}$, $\sqrt{S^2 + \sigma_n^2}$)	Observations are a noisy realization of true values
$p(\sigma_n)$	Delta(σ_n)	Measured uncertainty on measurements
$p(S_c)$	log Normal(-1.0, 1.0)	Unmeasured uncertainty in gradient
$p(Z_{\mathrm{true},n} \vert m, b, \vec{r}_n)$	Deterministic($m \times \vert\vec{r}_n\vert + b$)	True values come from an underlying gradient
$p(m_c, b_c)$	Deterministic($\tan^{-1}(\theta)$, $b_\perp/\cos(\theta)$)	The covariant gradient parameters
$p(\theta)^\dagger$	Uniform(-0.5π, $+0.5\pi$)	Angle between gradient and horizontal axis
$p(b_\perp)^\dagger$	Uniform(-10, 10)	Perpendicular intercept
$p(r_{\perp,obs,n} \vert \vec{r})$	Delta($r_{\perp,\mathrm{obs}} - \sqrt{r_x^2 + r_y^2}$)	Observed distances are a realization of true 3D distance
$p(\vec{r_n} \vert R_c)$	Isotropic Normal	True 3D distance
$p(R_c)$	log Normal(10, 5)	Scale length of the subpopulation
q_n	Categorical(P_c)	Subpopulation membership identifier

Table 5.2: Table of parameters for the full hierarchical mixture model with their prior distributions and qualitative description of their purpose in the model.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \text{ if } P(B) \neq 0. \tag{5.1}$$

which is derived from an axiom of conditional probability. Bayes' Theorem is just a way to compute conditional probabilities of events while folding in prior knowledge related to that event. In practice as a tool for statistical inference, Bayes' theorem is often written in terms of parameters, θ, and data, x, and the denominator, also known as the Bayesian evidence is often dropped to yield the *unnormalized posterior density*, $p(\theta|x) \propto p(x|\theta)p(\theta)$ (Gelman et al. 2013).

The first term on right-hand side of the proportionality is the *likelihood function* and the second term is the *prior distribution*. The likelihood function describes the connection between the available data and the parameters of interest.

In this work, the data we have is $x = [r_{\perp,\mathrm{obs},n}, Z_{\mathrm{obs},n}, \sigma_{\mathrm{Z},n}]$ for each n GC and our ultimate parameters of interest are the slope, m, and intercept b (highlighted with a red node and are together in the same node to indicate their covariance) of the metallicity gradient. However, we construct our model based on the idea that the data we have correspond to true versions which introduces *latent*, i.e., unobserved, parameters to our model. That is, the observed metallicity of a GC, $Z_{\mathrm{obs,n}}$, is a noisy realization of that GC's true metallicity, $Z_{\mathrm{true,n}}$ and its observed projected distance, $r_{\perp,\mathrm{obs},n}$, is a realization of the true 3D distance, $|\vec{r}_n|$.

In the left-hand side of Figure 5.5, we show part of the graphical representation of our probabilistic model ($|\vec{r}_n|$ will be discussed in more detail later). This shows how the relationship between the observations (filled nodes) and the parameters relevant to the inference we want to make (open nodes). Within the rectangle (known as the "plate"), we show the data and parameters for the individual GCs in the sample. The parameters outside the plate are the parameters for the whole population. We will now distinguish these population parameters (the *hyperparameters*, α) from the parameters for the individual GCs (θ_n). With the introduction of the α parameters, the joint distribution we seek to constrain is $p(\alpha, \theta_n|x_n)$, to which we can apply Bayes' Theorem:,

$$p(\alpha, \theta_n|x_n) \propto p(x_n|\theta_n, \alpha)p(\theta_n, \alpha). \tag{5.2}$$

In Figure 5.5 the arrows represent the conditional dependency among the different parameters and makes clear the hierarchical nature of our model. The key point is that the data, x_n, are only conditionally dependent on the parameters θ_n and are therefore *conditionally independent* from the hyperparameters, α. That, and being able to factor $p(\theta_n, \alpha)$ to $p(\theta_n|\alpha)p(\alpha)$ gives,

$$p(\alpha, \theta_n|x_n) \propto p(x_n|\theta_n)p(\theta_n|\alpha)p(\alpha). \tag{5.3}$$

The key difference between standard Bayesian models and HBMs is that we constrain the population parameters by conditioning on the observations of the many individual GCs rather than fixing them and using them as priors.

The gradient parameters are inferred through modeling the "true" metallicity values of the individual GCs, $Z_{\mathrm{true},n}$. The $Z_{\mathrm{true},n}$ values are set deterministically by the linear relation $\mathrm{p}(Z_{\mathrm{true},n}|m, b, \vec{r_n}) = m \times |\vec{r_n}| + b$. We condition $Z_{\mathrm{true},n}$ on $Z_{\mathrm{obs},n}$

by modeling the observations as drawn from normals centered on the true values and with a standard deviation that encompasses our uncertainty on the metallicity gradient. This uncertainty is the quadrature sum of the uncertainties on the individual [Fe/H] measurements, $\sigma_{Z,n}$, and *unobserved* uncertainty for the intrinsic scatter in the radial metallicity gradient, S, such that $\sigma = \sqrt{S^2 + \sigma_{Z,n}^2}$.

With the $\vec{r}_n$ dependence for $Z_{\mathrm{true},n}$ we introduce a key advantage when using a Bayesian framework. That is, even though we do not have the line-of-sight distances, $r_{\parallel,n}$, we can make inferences on the true distances for each GC while only making weak assumptions about the population. Specifically, we model the angular distribution of GCs as isotropic and assume that the GCs are normally distributed in radius by some scale length, R, in all 3 coordinates xyz and marginalize over the angle, φ_n, between x_n and y_n to get,

$$p(\vec{r}_n|R) = \frac{r_{\perp,n}}{R^2}\exp\left[\frac{-r_{\perp,n}^2}{2R^2}\right] \times \frac{1}{\sqrt{2\pi R^2}}\exp\left[\frac{-r_{\parallel,n}^2}{2R^2}\right], \tag{5.4}$$

which we fully derive in Appendix A. As such, we model the projected distances as drawn from a Rayleigh distribution (the first term on the right-hand side of the above equation) and the line-of-sight distances as drawn from a normal distribution (the second term). This structure is graphically represented in the right-hand side of Figure 5.5.

In reality, a power-law distribution better describes the projected radial distribution of a typical GC system. In the left panel of Figure 5.6, we compare the expected quantiles from a Rayleigh distribution versus the quantiles of the projected distances for our M87 sample (open circles). This figure demonstrates the extent the mock spatial

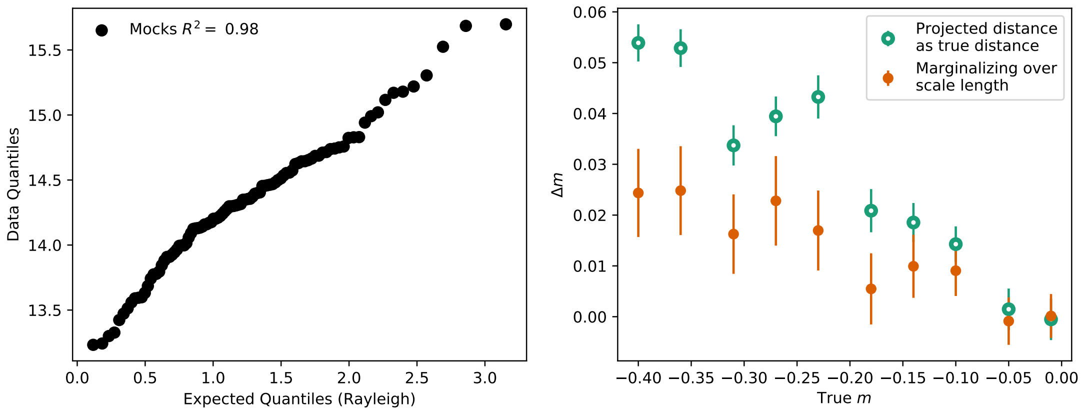

Figure 5.6: (Left) Expected quantiles of a Rayleigh distribution versus quantiles of the projected distances for the mock data (open circles). (Right) Recovery of slope as a function of true slope from weighted least-squares with projected distance as a proxy for true distance (open circles) and from our statistical framework (see text for details).

distribution deviates from the assumption of our model. If the projected distances were drawn from a Rayleigh distribution, the theoretical quantiles versus the data quantiles would be a straight line. The Rayleigh distribution is not a significant deviation from the distribution of the observed projected distances in our sample.

Figure 5.5 displays the joint probability distribution of all our parameters and data, $p(Z_{\mathrm{obs},n}, r_{\perp,\mathrm{obs},n}, Z_{\mathrm{true},n}, \vec{r}_n, R, S, m, b, \sigma_n)$. Because the arrows indicate the conditional dependence among the parameters and data we can use this to show how we can factorize the joint probability distribution of all our parameters into conditionally independent probability distributions to obtain,

$$
\begin{aligned}
p(Z_{\mathrm{obs},n}, r_{\perp,\mathrm{obs},n}, Z_{\mathrm{true},n}, \vec{r}_n, R, S, m, b, \sigma_n) \propto& \\
&\prod_{n=1}^{N} p(Z_{\mathrm{obs,n}} | Z_{\mathrm{true,n}}, S, \sigma_n) p(r_{\perp,obs,n} | \vec{r}_n) \times \\
&\prod_{n=1}^{N} p(Z_{\mathrm{true},n} | m, b, \vec{r}_n) p(\vec{r_n} | R) p(\sigma_n) \times \\
&p(S) p(R) p(m, b), \quad (5.5)
\end{aligned}
$$

To test the efficacy of this model, we generated mock data from where the coordinates are drawn from a power-law that goes as -2.5, with each data point getting randomly assigned 10% to 55% uncertainty. In the right panel of Figure 5.6 we compare how well we recover the true slope when using projected distance as a proxy for true distance and a weighted least-squares fit to get the gradients (open circles) to when how well we recover the slope when we marginalize over the scale length (closed circles). Over the range of slope values, the recoverability improves with the statistical de-projection. The difference in results between the two methods is starkest when the gradient is significant, while there is no difference in the recoverability when the gradient is consistent with being flat ($m = 0$).

5.3.2 Generalizing to multiple subpopulations

Effect of making cuts on the population

In the previous section, we demonstrated the efficacy of a Bayesian linear regression approach relative to weighted least-squares to accurately recover the gradient

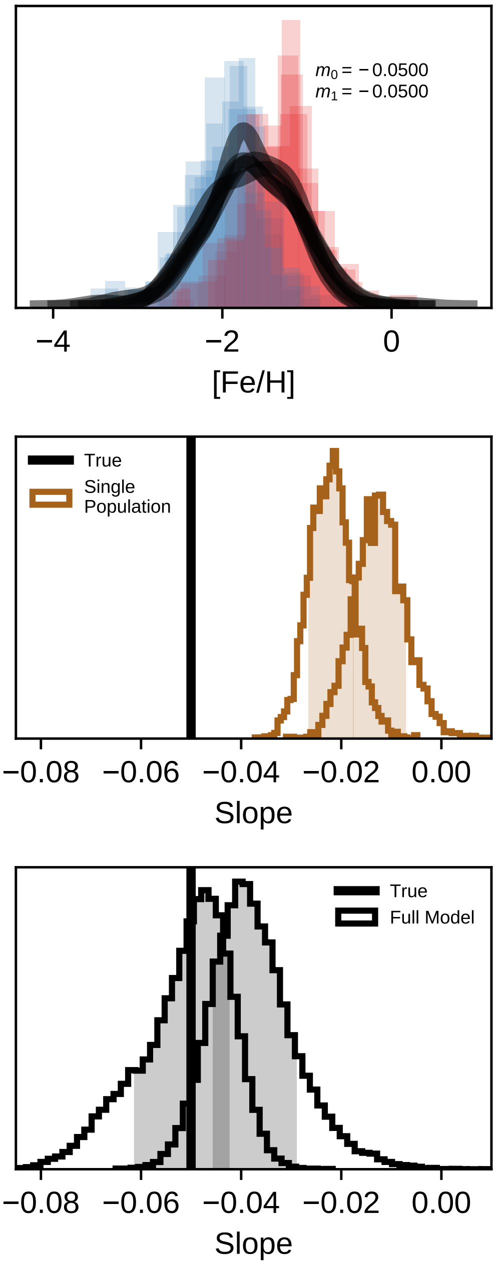

Figure 5.7: (Top) Metallicity distribution function (MDF) of 5 realizations of mock data generated from $m_0 = m_1 = -0.05$ and $b_0 = -0.4$ and $b_1 = -1.0$. Colored histograms show the true subpopulation separations and the black lines are the non-parametrically smoothed MDF of the combination of the subpopulations. (Middle) Demonstration of recovery of true slopes (black line) for when a the single population model from Section 5.3.1 is used on the subpopulations determined from a constant cut on [Fe/H] (brown) Bands show the range between the 16th and 84th percentiles for all posteriors. (Bottom) Same as middle panel but now using the full hierarchical mixture model.

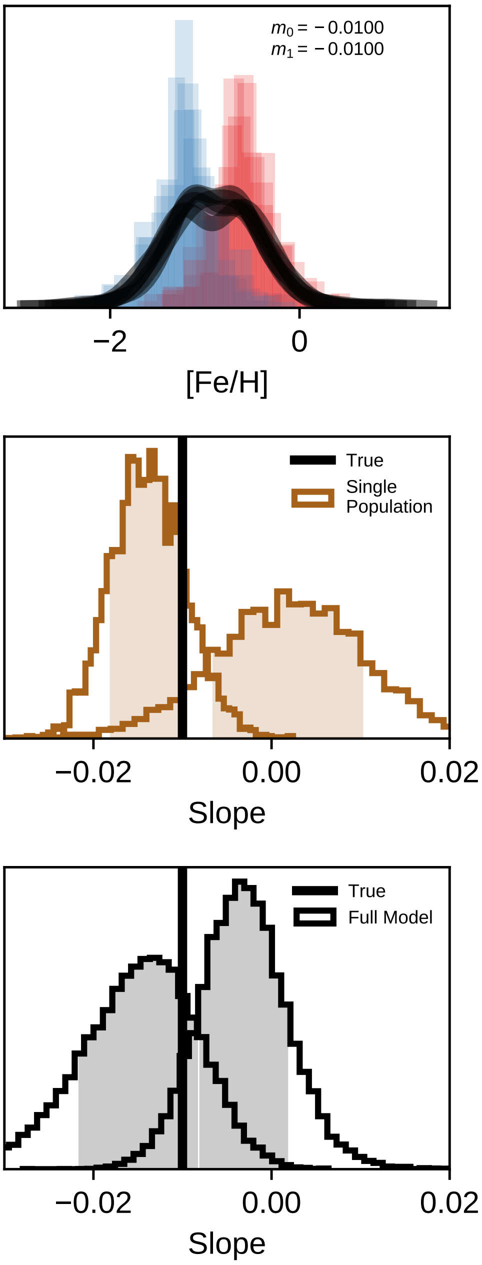

Figure 5.8: Same as Figure 5.7 but for $m_0 = m_1 = -0.01$.

parameters of a set of data points drawn from a particular line. A fundamental assumption in the method presented is the data points come from the *same population*. As previously described, however, GC systems are generally composed of subpopulations, and knowing how to separate the individual GCs of a system into the correct subpopulations is one of the most difficult steps towards characterizing GC systems.

In the top panels of Figures 5.7 to 5.9 we show three versions of mock data, all generated from power law distributions and two underlying gradients. In all versions we use $b_{\mathrm{true},0} = -0.4$ and $b_{\mathrm{true},1} = -1.0$ and a variety of slope parameters: $m_{\mathrm{true},0} = m_{\mathrm{true},1} = -0.05$ (Figure 5.7), $m_{\mathrm{true},0} = m_{\mathrm{true},1} = -0.01$ (Figure 5.8), and $m_{\mathrm{true},0} = -0.01$ and $m_{\mathrm{true},1} = -0.015$ (Figure 5.9). For each "system" we generated 5 realizations of mock data.

For each set of mock data, we made constant cuts based on the MDFs to separate the populations, mimicking what one might do if they did not have *a priori* information on the different subpopulations. We fit each realization of the subsequent subpopulations with our Bayesian linear regression model presented in Section 5.3.1, which we note is already an improvement over previously used methods, as demonstrated in Section 3.1.

In the middle panels of Figures 5.7 to 5.9 we show how effective this method is by comparing the true slope values (black line) to the median of the posteriors for each realization of the mock data (brown histograms). Even with the improvements to the linear regression outlined in Section 5.3.1 the inferred slope values are not accurate, with the inferred slopes typically being flatter than the true slopes. We therefore need to generalize our single population model to account for the covariance between the gradient parameters and the subpopulation membership assignments to more accurately estimate both.

A mixture model

Hogg et al. (2010) discussed mixture models in the context of linear regression for the purposes of outlier rejection. Separating individual GCs into subpopulations is an equivalent problem. We model the system such that a given GC has C number of subpopulations it could be assigned to through an identifier parameter q_n. Like Hogg et al. (2010), we directly marginalize over the class membership of each GC by introducing a new parameter, the prior on q_n, $P_c \in [0, 1]$ such that $\sum_{c=1}^{C} P_c = 1$. The P_c parameters are the mixture weights for each subpopulation and allow us to marginalize out the subpopulation identifiers.

In principle, we can fit for any number of subpopulations. In practice, however, throughout this work we specialize to the $C = 2$ case, that is, we model the mock and observed data as a bimodal distribution. We set a lower limit on P_c, $P_{\min} = 0.3$. Then,

$$P_c = \begin{cases} P_0 \sim \text{Uniform}(P_{\min}, 1 - P_{\min}) \\ \\ P_1 \sim 1 - P_0 \end{cases} \tag{5.6}$$

The structure otherwise remains the same as our single population model. We are able to transition our population parameters from the single population model to be a part of the mixture model because the parameters will exist for each mixture component (i.e., GC subpopulation). So we make a small adjustment to our notation: m_c, b_c, $\log R_c$, and $\log S_c$ where the subscript refers to a given subpopulation. The joint probability distribution is then given by,

$$
\begin{aligned}
p(\alpha_C, \theta_n | x_n) \propto & \left(\prod_{c=1}^{C} p(R_c, m_c b_c, S_c) \right) \times \\
& \prod_{n=1}^{N} \left(\sum_{c=1}^{C} P_c \times p(Z_{\mathrm{obs},n}, r_{\perp,\mathrm{obs},n}, Z_{\mathrm{true},n}, \vec{r}_n, R_c, S_c, m_c, b_c, \sigma_n) \right)
\end{aligned}
\tag{5.7}
$$

The third term in this equation is what we factorized in Section 3.1 for the single population model. We show the graphical representation of the final hierarchical mixture model in Figure 5.10.

We specify our model with the probabilistic programming package `PyMC3` (Salvatier J. 2016). `PyMC3` uses the Hamiltonian Monte Carlo (HMC) family of samplers. For this work in particular, we use the No-U-Turn Sampler (NUTS, Hoffman & Gelman 2011). HMC samplers are more efficient than the commonly used ensemble samplers because they do not rely on the current state to propose the next state (for an introduction to HMC see Betancourt 2017) and so they are the more appropriate choice for high-dimensional problems.

The sampling efficiency of the HMC algorithm is highly sensitive to several tuning parameters. For this work, the most important tuning parameter is the mass matrix because our model parameters are highly covariant. If the mass matrix is not well-matched to the covariance of the posterior, both the step size will need to be

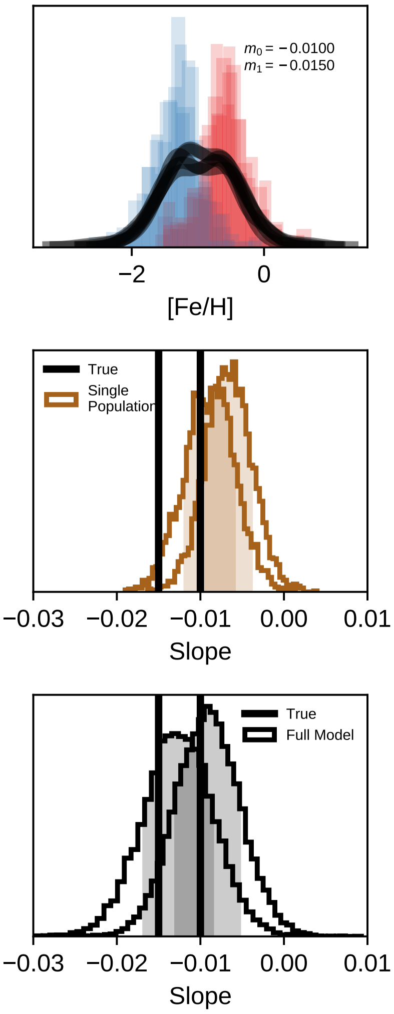

Figure 5.9: Same as Figure 5.7 but for $m_0 = -0.01$ and $m_1 = -0.015$.

decreased, and the number of steps increased, making it difficult to achieve convergence.

`PyMC3` does not have a built-in way to optimize the mass matrix. We use the `exoplanet`[13] extensions to `PyMC3` to fit for a dense mass matrix during burn-in. We find values to to initialize the sampler via several steps: first, we fit a 1D mixture of Guassians on the metallicities while taking into account the uncertainties to get an initial guess of the class membership for each observation, and then, second, we fit a linear model to the project metallicity gradients for each subpopulation to find initial guesses for the intercepts and slopes. With this approach, we obtain a converged model based on the Gelman-Rubin statistic for each parameter, $\hat{R}$, (where $\hat{R} > 1$ indicates the chains have not converged).

In the bottom panels of Figures 5.7 to 5.9 we show the inferred slope posteriors (black histograms) to demonstrate the efficacy of this method. In all cases the recovery is better when using the full model, with the biggest improvement made in the case where the two subpopulations are most well-mixed in the MDF (Figure 5.7).

The full model accurately recovers the different slopes in Figure 5.9 within 1σ uncertainty but cannot distinguish the gradients as different at a statistically significant level. This is still an improvement over existing methods, but, in the context of GC subpopulations, the ability to discern any gradient differences is essential for understanding how potentially similar the assembly histories of the different subpopulations (see Section 5 for more discussion on this). Improving the precision of the subpopulation parameters will be the subject of future work.

---

[13] `https://exoplanet.dfm.io`

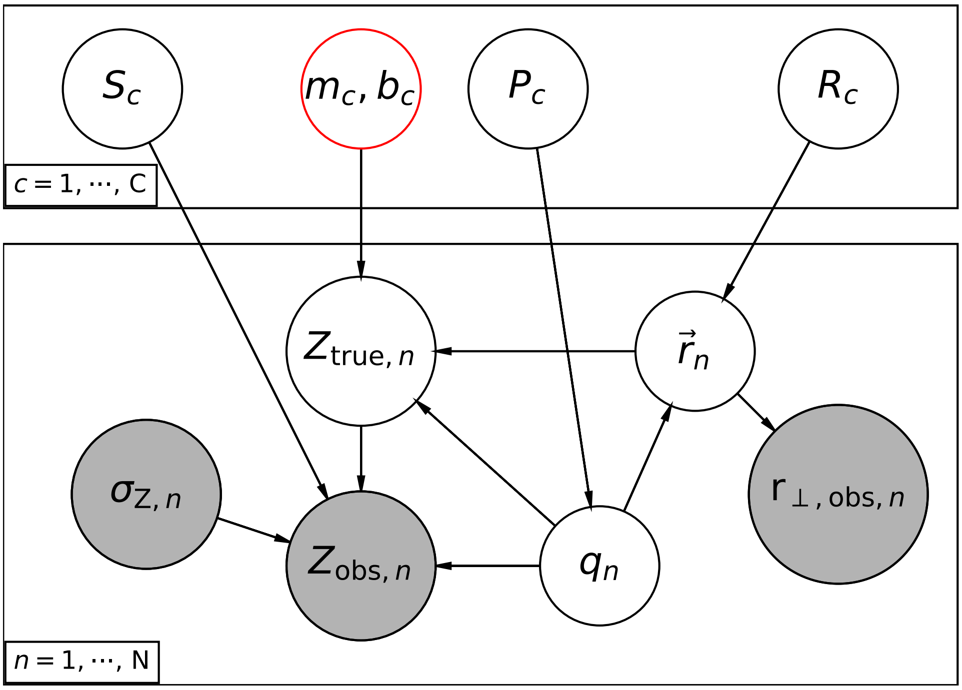

Figure 5.10: Similar to Figure 5.5 but now for our final hierarchical mixture model (see text for details). Now there is a second plate around are population parameters which indicates these parameters are determined for all subpopulations in our sample ($c = 1, ..., C$) and we have subpopulation identifiers, q_n set by the prior P_c.

5.4 Results

5.4.1 Radial metallicity gradients

In this work, we have two spectroscopic data sets for the GC system: the Keck/LRIS sample and the MMT/Hectospec sample. The former covers a radial range of $\sim 7 - 27$ kpc while the latter spans $\sim 14 - 142$ kpc. Previous kinematical analyses of GCs and planetary nebula show signs of a transition at $\sim$ 40 to 50 kpc, which may be related to a recent accretion event (Romanowsky et al. 2012; Longobardi et al. 2015; Zhang et al. 2015). Photometric surveys have also shown in M87, and other massive ETGs, that blue GCs begin to dominate at large radii. To measure the metallicity gradients, we split our sample at 40 kpc. The inner halo sample consists mostly of the Keck/LRIS data with a small fraction coming from the MMT/Hectospec data. The outer halo sample consists completely of MMT/Hectospec data.

Before discussing the results from fitting the model to the individual [Fe/H] measurements, we first empirically examine what our expectations ought to be for the gradients in Figure 5.11. In the top panel we show two metallicity distribution functions (MDF) for the Keck/LRIS sample, one for the inner part of the dataset (light green line, $r_\perp \leq 25$ kpc) and one for the outer part (dark green dashed-line, $r_\perp > 25$ kpc). There are two distinct peaks in the inner MDF, while the outer MDF has a less significant second, metal-rich peak. In the outer bin, the metal-poor peak shifts noticeably from the inner metal-poor peak.

In the bottom panel we do the same demonstration for the outer halo. Bimodality is not as clearly seen in the outer halo sample as it is in the inner halo but

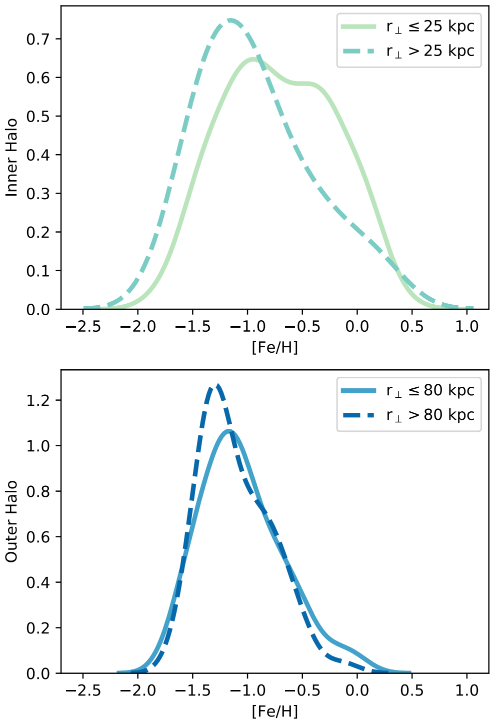

Figure 5.11: Empirical demonstration of gradients for inner (top) and outer (bottom) halos. In each panel we show the metallicity distribution function of the data set broken into two radial bins. Both the inner and outer halos show evidence of multiple subpopulations from their MDFs and a slight gradient.

there is a distinct negative shift from the main peak from the inner bin to the outer bin. The lack of clear bimodality could be a result of the MMT/Hectospec sample having far fewer red GCs than blue GCs and is consistent with the findings for other BCGs (see, for example, Harris et al. 2017).

For the modeling, we initialized the MCMC chains in the same manner as the mock data and modeled the data as composed of two subpopulations for both the inner and outer halos. In Figure 5.12 we show the posteriors of slope values for the metal-

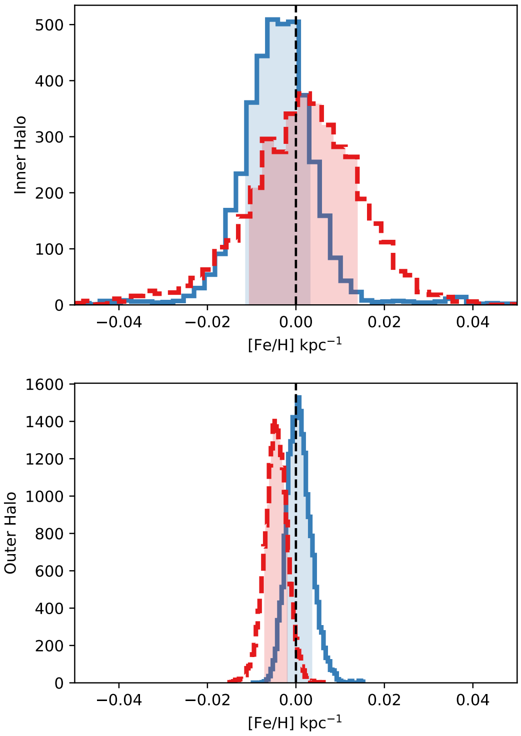

Figure 5.12: Comparing the posteriors on the slopes for the metal-poor (blue) and metal-rich (red) subpopulations for the inner (top) and outer data (bottom) halos. The 1σ uncertainty in each posterior is shown in the colored bands and a flat gradient is marked (black dashed line).

poor (blue) and metal-rich populations for the inner halo (top panel) and the outer halo (bottom panel). The 16th and 84th percentiles are marked by the colored bands. In each panel a flat gradient is marked with the black dashed line.

The uncertainty on the slope measurements is significantly larger for the inner halo than the outer halo measurements even though the [Fe/H] uncertainty is $\sim 20\%$ higher for the MMT/Hectospec data. The large uncertainty in the inner halo could be a result of the comparatively non-uniform coverage in $r_\perp$ for the inner halo sample, we get less information from each individual measurement in the inner halo than the outer

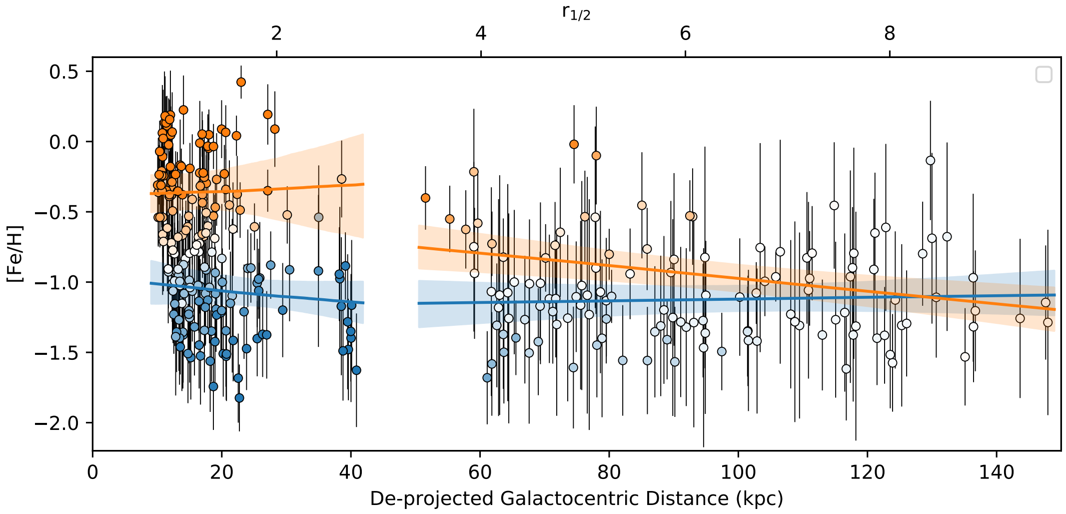

Figure 5.13: Radial metallicity gradients of the subpopulations with respect to the de-projected distances. The circles show the [Fe/H] measurements. They are colored by subpopulation assignment and the opacity of the individual points is scaled by certainty of that subpopulation assignment, with white indicating the assignment is highly uncertain. We show posterior median (solid lines) and the range encompassed by the 16th and 84th percentiles (bands) of the gradient distributions.

halo. For the inner halo, both the metal-rich and metal-poor slopes are consistent with a flat gradient and are statistically consistent with one another. In the outer halo, the metal-poor slope is consistent with a flat gradient while the metal-rich slope is slightly negative.

In Figure 5.13 we show the radial metallicity gradients of the subpopulations with respect to the de-projected distances. The circles show the [Fe/H] measurements which are colored by subpopulation assignment. The opacity of the individual points is scaled by certainty of that subpopulation assignment, with white indicating the assignment is highly uncertain. We show posterior median (solid lines) and the range encompassed by the 16th and 84th percentiles (bands) of the gradient distributions. Even though the gradient parameters are more uncertain in the inner halo, the sub-

	m	σ	b	σ	MDF 16th	MDF 50th	MDF 84th
Inner Halo							
Metal-Poor	-0.004	0.010	-0.957	0.224	-1.43	-1.15	-0.91
Metal-Rich	-0.001	0.014	-0.384	0.216	-0.64	-0.34	$+0.05$
Outer Halo							
Metal-Poor	$+0.001$	0.003	-1.196	0.287	-1.42	-1.24	-0.96
Metal-Rich	-0.005	0.003	-0.522	0.268	-0.83	-0.64	-0.45

Table 5.3: Summary of gradient parameters and MDF characteristics for the subpopulations of the M87 GC system.

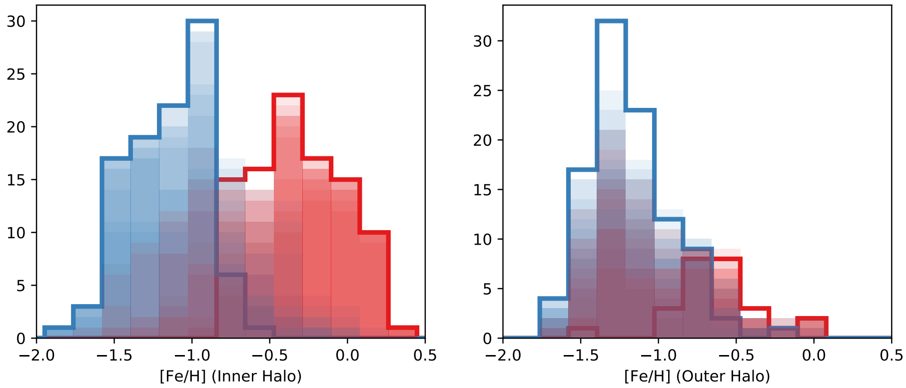

Figure 5.14: (Left) MDF for the metal-rich (red) and the metal-poor GCs (blue) in the inner halo.The solid blue and red lines show the posterior median of the subpopulation assignments of the individual GCs. The filled-in blue and red histograms represent how the uncertainty in the subpopulation assignments (see text for details) propagates to uncertainty in the MDF. (Right) Same as left but for the outer halo GCs.

population membership assignments are more certain than in the outer halo population because there are fewer metal-rich GCs and the metallicity separation between the subpopulations is smaller.

5.4.2 Characteristics of the subpopulations

In Figure 5.14 we compare the metallicity distribution functions (MDFs) for the metal-poor GCs (blue) and metal-rich GCs (red) for the inner halo GCs (left) and outer halo GCs (right). The solid lines show the result of using the posterior median of the subpopulation assignments of the individual GCs. We represent how the uncertainty in the subpopulation assignment affects the MDF by plotting the result of selecting class labels using a random number generated by the probability of the cluster-subpopulation pair for 10 random samples from the posterior (solid histograms).

To check the results of our model we compare Figure 5.11 with Figure 5.14. Figure 5.11 indicates that the two subpopulations should be of about equal size for the inner halo sample and that the metal-poor GCs would be a larger population in the outer halo sample. Even though the subpopulation membership assignments are much less certain for the outer halo sample, we see that the model assigns significantly fewer metal-rich GCs in the outer halo. This picture is overall consistent with our broad understanding that with increasing galactocentric distance there will be more metal-poor GCs relative to metal-rich.

In Table 5.3 summarize the gradient parameters and the characteristics of the MDFs for the subpopulations.

In the Milky Way GC system, the metallicity subpopulations are associated with different spatial and kinematical components of the Galaxy itself (Zinn 1985). We assess how well this pattern holds for M87 in Figure 5.15 by examining the chemodynamics of the subpopulations in [Fe/H]–radial velocity space for the inner halo (top) and outer halo (bottom). For the inner halo, our modeled metallicity subpopulations correspond to differences in the radial velocity distributions. The metal-rich subpopulation has significantly less radial velocity dispersion than the metal-poor subpopulation.

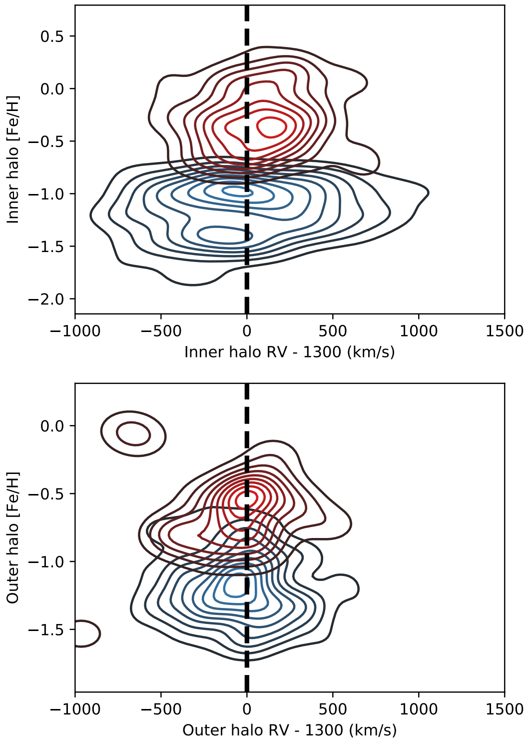

Figure 5.15: [Fe/H] vs. radial velocity for the inner halo sample (top) and the outer halo sample (bottom). In the inner halo, it is clear that our inferred subpopulation assignments correspond to genuinely distinct kinematic distributions.

Unlike the inner halo, there does not seem to be a correspondence between metallicity subpopulation and differences in the kinematic properties of the subpopulations for the outer halo. The subpopulations in the outer halo have a radial velocity dispersion similar to the inner halo metal-rich subpopulation.

5.4.3 Abundance patterns

In Figure 5.16 we show the stellar population radial profiles for M87 from our fits to the Murphy et al. (2011) sample (black circles) for [Fe/H] (upper-left) and a variety of α-elements. The M87 starlight shows a declining [Fe/H] profile and slightly positive profiles for [Mg/Fe] (upper-right), [Si/Fe] (lower-left), and [Ca/Fe] (lower-right). The [Fe/H] gradients are consistent with previous work that have studied stellar population gradients in massive ETGs (e.g. Greene et al. 2015; van Dokkum et al. 2017a; Gu et al. 2018b). These previous studies typically found a flatter [Mg/Fe] than what we present here but it is not a substantial difference. We also demonstrate the limitations inherent with integrated light with the [Fe/H] estimates from deep broadband photometry of a population of resolved stars in M87 (contours, Bird et al. 2010). This shows a metal-poor population that is not probed by the stellar population models.

To obtain abundance information for the GCs we have to stack the individual spectra since the majority of the Keck/LRIS and MMT/Hectospec spectra have too low-S/N to reliably extract abundance information. Stacking the GC spectra is made difficult by the need to separate the sample by subpopulation as it is expected the different subpopulations will have different origins and, thus, different abundance patterns. We demonstrated in Section 3 the importance of using a HBM framework for making accurate determinations of subpopulation membership of the individual GCs, which help make more physically-appropriate stacks. Additionally, we can take advantage of having made probabilistic determinations of subpopulation membership for the individual GCs. In Figure 5.14 we demonstrated how uncertainty in the subpopulation memberships propagated to the MDF of the GC system. In the same manner, we can propagate that uncertainty to our stacks and abundance information.

In the same manner we used for the Mitchell data, we made four inner halo stacks, binning by metal-rich and metal-poor and then further separating the GCs at a radius at 16 kpc, and two stacks for the outer halo only binning by metal-rich and metal-poor. We made ten iterations for each subpopulation, determined by different draws from the posterior for different subpopulation assignments for each GC (same draws that are shown in Figure 5.14). The stacked GC spectra have a typical S/N of $\sim 100 - 150/\text{Å}$.

Each version of each subpopulation stack was fitted using `alf` in the same manner as the Mitchell data. For each parameter of interest, we computed the 16th, 50th, and 84th percentiles of the posteriors for each fit. In Figure 5.16 we show the results for the metal-rich stacks (red circles) and metal-poor stacks (blue squares). The inner halo stacks are open symbols and the outer halo stacks are filled.

While we note that Figure 5.16 cannot be directly compared to Figure 5.13 because we have moved from de-projected to projected distances, broadly the [Fe/H] gradient measured from the metal-rich inner halo stacks is consistent with the flat gradient measured from the individual [Fe/H] measurements. For the inner halo metal-

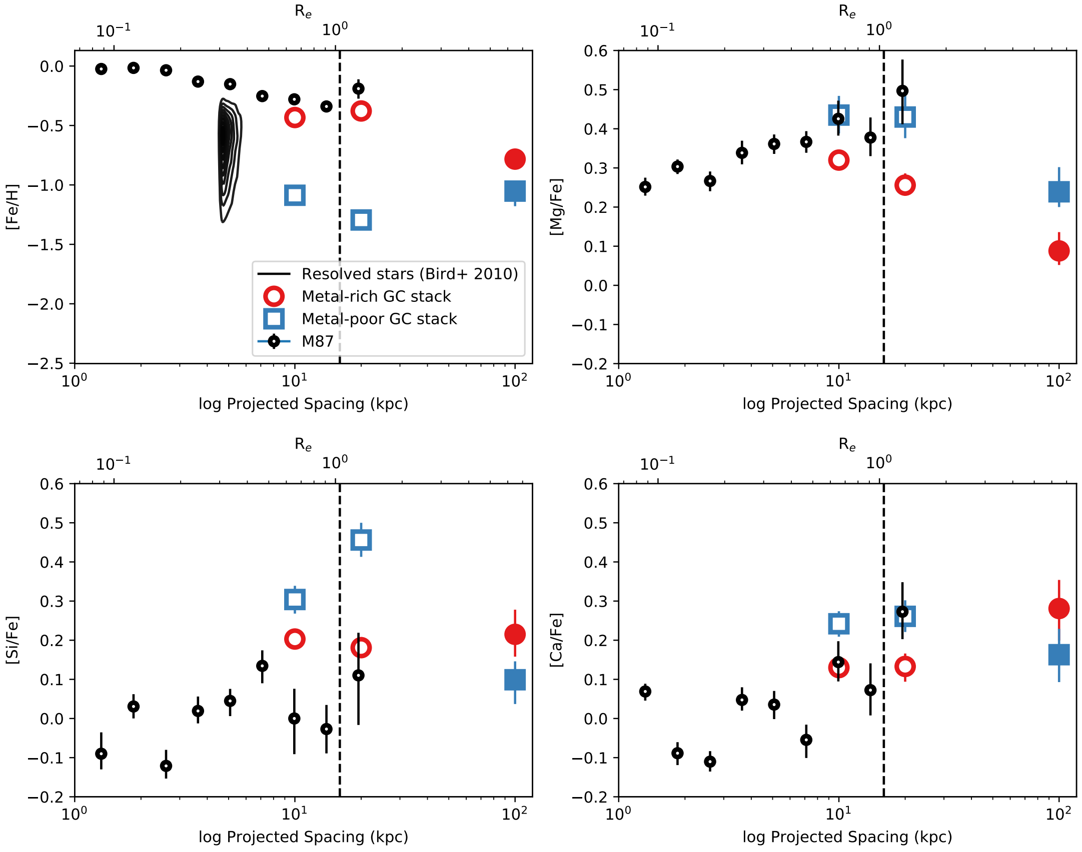

Figure 5.16: Stellar population radial profiles for [Fe/H] and a variety of α-elements as derived from full spectrum fitting to the Murphy et al. (2011) spectroscopy of the M87 galaxy light (black), the metal-rich GC stacks (red circles), and the metal-poor GC stacks (blue squares). The grey band indicates $R_{\rm gal} \leq 2.0$ kpc, i.e., the central-most region where massive ETGs display many exotic stellar population characteristics, and the dashed line indicates $\sim 1 R_e$.

poor stacks we find a slightly negative gradient from the stack measurements that differs from the flat gradient shown in Figure 5.13. The likely cause of this difference is the non-uniform sampling of the inner halo GCs in galactocentric radius. For the individual [Fe/H] measurements the MMT/Hectospec sample provides the only coverage past ~ 27 kpc and the more metal rich measurements (~ -1.0) seem to be enough to keep the gradient nearly flat. However, for the stacks there are fewer of these comparatively metal-rich GCs than metal-poor so their contribution to the stack is not as important.

The inner halo metal-rich stacks are slightly less metal rich than the M87 starlight in the same region while the metal-poor stacks are less [Fe/H]-enhanced by ~ 1 dex. The metal-rich inner halo stacks are less [Mg/Fe]-enhanced than the galaxy light while the metal-poor stacks have similar [Mg/Fe]. The metal-poor stacks are also enhanced in [Si/Fe] and [Ca/Fe]. The metal-rich stacks have similar [Ca/Fe] enhancement but are less enhanced in [Si/Fe] in this same region. Enhancement in [Mg/Fe] drops precipitously from the inner halo stacks to the outer halo stacks. Enhancement in [Si/Fe] drops for the metal-poor stacks from the inner to outer but remains stable for the metal-rich stacks. Enhancement in [Ca/Fe] remains steady between inner and outer halo for all stacks.

In the left panel of Figure 5.17 we show [Mg/Fe] vs. [Fe/H] for the GC stacks and the galaxy data (colors and symbols same as previous figure). Also in Figure 5.17 we show the abundances for the Milky Way stars (purple cloud, from the JINAbase Abohalima & Frebel 2018, see detailed references in Appendix A), stars in dwarf galaxies around the Milky Way (brown cloud, JINAbase and Bonifacio et al. 2004), and Milky

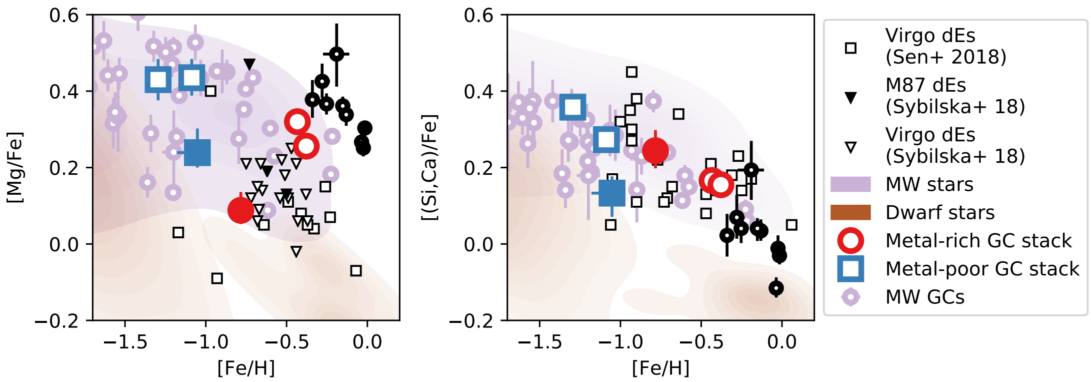

Figure 5.17: (Left) [Mg/Fe] vs. [Fe/H] for the GC stacks and M87 (symbols same as previous figure) For M87 the two measurements we have that are within 2 kpc are filled in. Also plotted is the kernel density estimate of the Milky Way field stars (purple) and the field stars from the Milky Way dwarf satellite population (brown), measurements from Virgo dwarf ellipticals (upside down triangles and squares), and integrated light measurements of Milky Way GCs (purple circles). (Right) Same as left panel but for the median of [Si/Fe] and [Ca/Fe].

Way GCs fitted from Schiavon et al. (2005) (see Villaume et al. 2019, for details on these fits). Also displayed are the abundances for dwarf ellipticals in Virgo from two different studies, Şen et al. (2018) (open squares) and Sybilska et al. (2018) (upside down triangles). For the Sybilska et al. (2018) sample we differentiate their sample into “M87 dwarf ellipticals” (closed) and “Virgo dwarf ellipticals” (open) with a cut at 300 kpc from M87 (distances from Peng et al. 2008). We were motivated by the results from Liu et al. (2016) which showed a transition in [Mg/Fe] in the dwarf elliptical population at this distance.

In the right panel of Figure 5.17 we show the median values for each object of [Si/Fe] and [Ca/Fe], except for Şen et al. (2018) who only measured [Ca/Fe]. All symbols are the same as the left panel. Even though Mg, Si, and Ca are all α elements they have different formation sites. Mg is purely a product of massive stars while Si

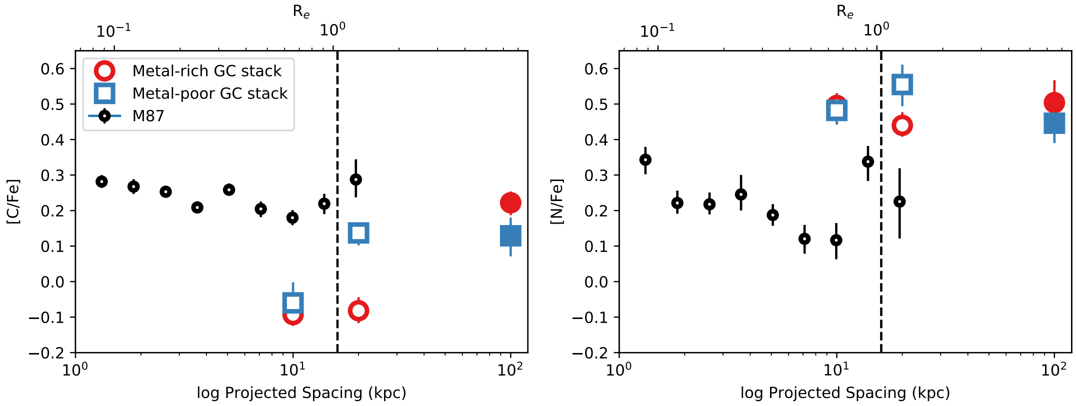

Figure 5.18: Same as Figure 5.16 but for [C/Fe] (left) and [N/Fe] (middle) (right).

and Ca can both be produced in Type Ia Sn as well (Woosley et al. 2002).

In Figure 5.18 we show a similar figure to Figure 5.16 but now for light elements: radial profiles for [C/Fe] (left) and [N/Fe] (right). For the M87 starlight we see enhanced [C/Fe] and [N/Fe] values and with negative radial gradients for both abundances. The GCs as a whole are less enhanced than the galaxy starlight for [C/Fe] but more enhanced in [N/Fe].

5.5 Discussion

5.5.1 The formation of the inner halo

Piecing together how the inner halo (< 40 kpc) formed is complicated by the fact that it is a mix of *in-* and *ex-situ* stellar populations. Decomposing the whole population into these components from observations cannot be quantitatively done with integrated galaxy starlight alone. With the GC system we have discrete tracers of near-simple stellar populations that overlap with and extend our coverage of the galaxy field star population, providing additional insight into how this region formed.

The red/metal-rich GC populations in massive ETGs have long been thought to have formed along with the original galaxy because they follow field star density profiles and kinematics (see Strader et al. 2011b, for M87 in particular). Until this work, however, a direct metallicity comparison at the same galactocentric radius has not been done. We have established that the metal-rich GCs have an average [Fe/H]~ -0.4 and [α/Fe]$\sim +0.15$ (Figures 5.16 and 5.17), similar to the galaxy field star population ([Fe/H]~ -0.3 and [α/Fe]$\sim +0.40$) over the same radial extent, indicating a common origin of the two populations.

In Figure 5.16 we show the radial gradients for [Fe/H] (upper-left panel) and various α elements measured from the integrated light of M87 (black circles). We find that the [Fe/H] gradient for the field star population is flat within the inner 2 kpc and then steepens to a negative gradient. This corresponds to rising gradients in all the α elements. This is characteristic of the populations seen in other massive galaxies (Gu et al. 2018a; Greene et al. 2013) and can be viewed as consistent with as the second phase of the "two-phase" formation framework (Oser et al. 2010) within an environmentally quenched environment (Liu et al. 2016).

This scenario would suggest that the bulk of the metal-rich GCs also came in from mergers. However, the measurements from the GC stacks indicate that the [Mg/Fe] values decline with radius for the metal-rich GCs. The population that then presumably brought in the metal-rich GCs would be diluting the [Mg/Fe] and depressing the gradient, rather than contributing to its rise. On the other hand, the metal-poor

GCs are very Mg-enhanced.

A negative gradient is expected from the kind of minor mergers that would bring metal-poor GCs into M87 while major mergers that would bring in metal-rich GCs can flatten gradients (e.g. Taylor & Kobayashi 2017). Our current measurements show that the metal-poor subpopulation gradient is skewed negative but is consistent with the metal-rich (and with a flat gradient) within the 1σ uncertainties. At first glance, this seemingly indicates that this population was affected by the same processes that flattened the metal-rich GC gradient and so it follows that the metal-poor population would have had to already be in place by the time the major mergers began.

The problem is that, according to cosmological simulations, this metal-poor population should not exist. Cosmological simulations predict that the Milky Way-mass (mass ratio$\sim$1:5) galaxies are the primary building blocks of the stellar halos of massive galaxies (Oser et al. 2012; Pillepich et al. 2018). This makes the metal-poor GCs too metal-poor to fit this framework, even though in Villaume et al. (2019) we established that the metal-poor GCs in the inner halo of M87 are ~ 0.4 dex more metal rich than previously thought. From a purely theoretical prospective, the mass ratios of the mergers suggested by this metal-poor population is not inconsistent with the what is needed to achieve the necessary size growth from high-redshift (see equation 4 in Naab et al. 2009). This tension may point to problems with the predicted populations of satellite galaxies in cosmological simulations.

An alternative explanation is that the metal-poor population is not entirely *ex-situ* It has been suggested that some of fraction of the metal-poor stars in the Milky Way halo formed *in-situ* and were kicked out (Zolotov et al. 2010). In this scenario, the *in-situ* metal-poor population would be more metal-enriched than the *ex-situ* population which could "dilute" the metallicity gradient. In the context of metal-poor GCs specifically, Mandelker et al. (2018) proposed a scenario in which cold gas flows that can form metal-poor GCs directly in the halos of massive galaxies at high-redshift. It is, however, currently unclear what the metallicity gradient of such a scenario would be but it would depend on the relation between the accretion and star formation rates.

So far, including the GC system in the analysis helps establish important benchmarks – the existence of the α-enhanced, metal-poor halo and the flat (albeit not necessarily the same) metallicity gradients of the GC subpopulations – that help us make qualitative advancements in our understanding of the assembly history of M87. However, until we can make more precise determinations of the *in-* and *ex-situ* populations we necessarily have to be agnostic towards the specifics of the *in-situ* star-formation in M87. This, however, is a particularly important process to understand because of the unexplained, exotic properties of the stellar populations in the innermost regions of the most massive ETGs. For example, the unexpected excess in ultraviolet (UV) flux within (e.g., Code & Welch 1979) and the bottom-heavy initial mass functions (IMFs) up to $1R_e$ (van Dokkum et al. 2017b). For M87, in particular, Sarzi et al. (2018) measured a bottom-heavy IMF out to ~ 4 kpc. Constraining the nature of this initial phase of ETG formation will likely clarify the star formation processes that can give rise to such characteristics.

It has been suggested that these characteristics are the result of atypical abundance patterns arising from a significant fraction of stars from dissolved GCs in the galaxy cores (Goudfrooij 2018; Chung et al. 2018). However, in Figure 5.18 we show that the metal-rich GCs do not agree with the field star population in light elements. This indicates that the GCs are altered by internal processes, as is well-known for the Milky Way GCs (Bastian & Lardo 2017). This multiple-population phenomenon has been hinted at by the UV excess in the M87 GCs (Peacock et al. 2017) but this is the first time these abundance patterns have been shown for the M87 GC population. Therefore, our results show little evidence for stars from dissolved GCs being a significant population in M87. For the time being, we have to conclude that the exotic stellar populations are a result of the initial star-formation process that formed the initial galaxy, whatever that process may be.

5.5.2 The formation of the outer halo

The previous work on radial gradients of stellar population parameters of massive ETGs only extended out to just a couple effective radii with the galaxy light (e.g., Greene et al. 2013), which, as we discussed in the previous section, is expected to consist of a mix of *in-situ* and accreted stars. The GC system provides access spectroscopically to the outer halo of M87 ($\gtrsim 40$ kpc). Stellar population parameters in the outer halo provide cleaner benchmarks for accretion predictions because we do not have to worry about a large *in-situ* population complicating our interpretation. While observations of GCs and planetary nebula indicate the presence of an intracluster component that becomes significant beyond ~ 300 kpc (Longobardi et al. 2018b,a), our work focuses on the material that is inside this radius and bound to M87.

Despite the difficulty in determining confident subpopulation membership assignments for the individual GCs, we find there do appear to be two, genuinely different populations in the outer halo. Even taking into account the subpopulation membership assignment uncertainties when creating the stacked spectra, we measure distinct metallicities for the two outer halo GC stacks. Moreover, while the [C/Fe] and [N/Fe] abundances are similar between the two metallicity populations, there are differences in the α element abundance patterns. The metal-poor stack is more Mg-enhanced than the metal-rich stack while they have similar Si- and Ca-enhancements. This is in contrast with the views of previous work that the treat the outer halo GC populations as singular (e.g., Forbes & Remus 2018).

The populations in the inner and outer halo appear to originate from different galaxy types. The subpopulations in the outer halo both have overall lower metallicities than the corresponding inner halo subpopulations. There also appear to be differences in the abundance patterns between the inner and outer halo (Figure 5.16), with the outer halo GC stacks having lower [Mg/Fe] abundance than the inner halo stacks. However, when comparing the abundance differences between the inner and outer halo GCs, we need to consider the possibility of mass effects. The outer halo sample consists of more luminous, and therefore more massive, GCs than the inner halo sample (Figure 5.1). The abundance spreads in the Milky Way GCs have been shown to correlate with luminosity (Figure 16, Carretta et al. 2010), so it would follow that the more massive outer halo GCs might in some way be impacted by this effect. We need to address whether we expect this effect to be significant.

Carretta et al. (2010) demonstrated that the correlation between luminosity and extent of abundance spreads (specifically in their case, Na–O) in GCs is driven by the extreme of the abundnace anti-correlation, not the median values (see their Figure 11). Since integrated light probes the average parameters of the stellar populations, this would suggest that the mass dependency would not be as strong when measuring integrated light. To test this we found the correlation between mass for the Milky Way GCs and [C/Fe], [N/Fe], and [Mg/Fe] as measured by `alf` from integrated light (Schiavon et al. 2005). For [Mg/Fe] and [C/Fe] we found a mass dependence of $\sim 10\%$. Even accounting for the outer halo GCs being more massive than the most massive GCs, this effect is unlikely to explain the full difference between the inner and outer halo abundances.

The entirety of this outer halo GC population appears to be *ex-situ* and, with the exception of the [Mg/Fe] value for the metal-rich stack, the outer halo stacks remain α-enhanced. This suggests that the "coordinated assembly" picture continues in some capacity into the outer halo. The metal-rich stack displays the same unusual abundance pattern among the α elements seen in the Virgo dwarf ellipticals, with standard looking [(Si,Ca)/Fe] but very depressed [Mg/Fe]. The metal-poor GCs do not display this same abundance pattern.

It is important to acknowledge the difficulty in chemical tagging in this situation. Interpreting similarities or differences between the dwarf ellipticals and the other objects is complicated because the strong radial gradients in dwarf ellipticals (Figure 1 Sybilska et al. 2018), with their nuclei often being distinct. Şen et al. (2018) used an

$R_e/8$ aperture with the nucleus included while the Sybilska et al. (2018) measurements were done by taking the luminosity-weighted average of spectra within $1R_e$. While the [Fe/H] and [Mg/Fe] measurements between the two samples are similar we still need to be careful with our analysis because neither sample measures both [Ca/Fe] and [Mg/Fe] so we are assuming the two subsamples of galaxies are similar.

We can look at other aspects of the GC system and see that they are broadly consistent with our interpretation of the stellar population parameters. In the outer halo, M87 has a V-band luminosity of $\sim 2.9 \times 10^{10} L_{\odot}$, very similar to the Milky Way (Kormendy et al. 2009; Bland-Hawthorn & Gerhard 2016). From Sersic fits to the photometric sample of M87 GCs from Strader et al. (2011b) and correcting for GC luminosity function incompleteness, we estimate there are ~ 1200 metal-rich GCs and ~ 4500 metal-poor GCs in this region. This means that the GC specific frequency of the outer halo is $S_N \sim 16$; which besides M87, the only galaxies in Virgo with such a high value are dwarfs with $M_V \sim -17$ to -16 and fainter ($M_* \sim 10^8 - 10^9 M_{\odot}$, Peng et al. 2008, Figure 12). These are also generally close to M87 (~ 200 kpc projected). They are the plausible progenitors of the outer halo, if ~ 100 of them were accreted.

These high S_N dEs could be the building blocks of the metal-poor population. The metal-rich GC system would have to come from a second, higher-mass population of dEs, perhaps similar to the dEs plotted in Figure 5.17. This scenario is in tension with the predictions from IllustrisTNG which at > 100 kpc predicts that 90% of the *ex-situ* mass is coming from progenitors with stellar masses $\gtrsim 5 \times 10^9 M_{\odot}$ (Figure 13b Pillepich et al. 2018), with the typical progenitor mass being $\sim 7 \times 10^{10} M_{\odot}$. A similar conclusion to our own was come to by Longobardi et al. (2018a) using the outer halo light color M87 to infer low-mass progenitors. Hartke et al. (2018) took that result to indicate a problem with the feedback prescription in IllustrisTNG. However, we also need to consider that in hierarchical structure formation it is possible that the accreted satellites were different than the surviving population. That is, that there *may* have been Milky Way-mass galaxies but with GC S_N more like dwarfs that no longer exist today, or possibly similar to the ultra-diffuse galaxies that have been found to have very high S_N (Peng & Lim 2016).

5.6 Summary

Using updated full-spectrum SPS models we present the first detailed stellar population analysis of the M87 GC system from spectroscopy. We applied the models to 322 GCs extending from the inner to outer halo. We use these same models to fitted IFU spectroscopy to get spatially-resolved stellar population parameters of M87 itself. This work represents initial steps in applying a chemodynamical galactic archaeology approach to M87, following what has been done in the Milky Way and M31.

We present a new statistical framework to measure the radial metallicity gradients of a multimodal GC system that accounts for the covariance between subpopulation membership assignments and the physical parameters of the subpopulations while doing a statistical de-projection of the galactocentric distances which enables much more accurate measurement of the linear gradient parameters of the GC subpopulations. This work has resulted in the following observational benchmarks to aid our understanding of massive ETG formation and assembly:

- We also confirm the presence of a metal-poor, α-enhanced inner halo in M87 that is unanticipated by current cosmological galaxy simulations.
- We show the first direct spectroscopic comparison of field stars and GCs in M87, confirming the association of field stars and red GCs.
- For both the metal-rich and metal-poor subpopulations we find flat metallicity gradients, although within the uncertainty we cannot confirm the subpopulations share the same slope.
- We measure, for the first time, the [C/Fe] and [N/Fe] abundances for the M87 GC population, which show evidence that they have the same multiple-population phenomenon seen in Local Group GCs.
- From the α abundances of the outer halo GC stacks, we find evidence for relatively recent accretion of low-mass satellites with extended star-formation histories.
- We find potential tension between our results, which suggest a dominant contribution from dwarf galaxies to the stellar halo of M87, and simulations, which point to Milky Way-mass galaxies as the progenitors of the stellar halos of M87-like galaxies.

Chapter 6

Spatially-Resolved Stellar Populations for DF44

6.1 Introduction

Low-surface brightness (LSB) galaxies were first introduced by Sandage & Binggeli (1984) with the discovery of, "a new type of very large diameter (10000 pc), low central surface brightness ($\geq$ 25 B mag/arcsec) galaxy, that comes in both early (i.e., dE) and late (i.e., Im V) types". The nature of these galaxies made them difficult to discover and characterize and so for decades only a handful of LSB galaxies were known (Impey et al. 1988; Bothun et al. 1991; Dalcanton et al. 1997). This number precipitously jumped with the detection of 47 such objects in the Coma Cluster (van Dokkum et al. 2015) using the Dragonfly Telephoto Array (Abraham & van Dokkum 2014) which was designed specifically for LSB imaging.

Knowing what to look for, Koda et al. (2015) and Yagi et al. (2016) discovered a bonanza of similar objects from archival data, increasing the number of objects, now known as ultra-diffuse galaxies (UDGs), known in the Coma Cluster into the thousands. The search for UDGs was expanded to galaxy clusters beyond the Coma (van der Burg et al. 2016, e.g.,) which has lead to more substantial increases in the number of known UDGs.

Significant resources have been put towards understanding how such objects form and how they fit into our larger conception of galaxy formation. The central tension that underlies this work is determining whether UDGs are anything "special" or not. That is, are they the "failed" massive galaxies originally speculated by van Dokkum et al. (2015) or are they simply dwarf galaxies that were puffed up either through spin (Amorisco & Loeb 2016) or bursty star-formation histories (SFHs, Di Cintio et al. 2017).

There is observational support for the latter scenario both from local analogues to UDGs (e.g., Müller et al. 2018; Collins et al. 2020) and for more distant UDGs in cluster environments (e.g., Beasley & Trujillo 2016; Toloba et al. 2018; Prole et al. 2019). However, while the original scenario of failed $\sim L_*$ is no longer considered, the extreme overabundance of globular clusters (GCs)(e.g., Peng & Lim 2016; Beasley et al. 2016; van Dokkum et al. 2016, 2017b; Lim et al. 2018) in some UDGs still lends support to scenario that some UDGs are in more massive dark matter halos than would be expected from their luminosities.

Stellar population parameters like age, stellar metallicity, and abundance pattern hold the promise to help clarify this situation. Forbes et al. (2020) put forth a toy model that relates the GC systems of UDGs to the global stellar parameters of the galaxies – UDGs that are really just dwarfs should have fewer GCs and extended SFHs reflected in low [α/Fe] ratios. While UDGs that are truly failed galaxies should have just the opposite, overabundant GC systems and intense, truncated SFHs reflected in high [α/Fe] ratios.

However, only recently have the stellar population parameters of UDGs have been made accessible because of the difficulty in obtaining even moderate signal-to-noise (S/N) spectroscopy. For three UDGs in the Coma Cluster, Gu et al. (2018b) found old ages and low metallicities which disfavor formation scenarios that predict late star formation or quenching. Ferré-Mateu et al. (2018) looked at 7 UDGs in the Coma Cluster and found more intermediate ages and small range of metallicities and α-enhancements.

The Coma Cluster UDG Dragonfly 44 (DF44) is one of the most well-studied UDGs. It has an abundant GC system and large stellar velocity dispersion (van Dokkum et al. 2016), making it a good candidate to be a bonafide failed galaxy. van Dokkum et al. (2019) presented spatially-resolved spectroscopy for the UDG DF44 which confirmed that the potential of the galaxy is dominated by dark matter. In this paper, we present the first spatially-resolved stellar population gradients of a UDG using this same data.

6.2 Data and Methods

We make use of the spectroscopic sample described in van Dokkum et al. (2019) (henceforth referred to as Paper I) and interested readers should refer to it for detailed descriptions of the observations and data reductions. Briefly, we obtained integral field unit (IFU) spectroscopy of Dragonfly 44 (DF44) with the Keck Cosmic Web Imager (KCWI) and extracted spectra in nine elliptical apertures following the isophotes of the galaxy.

KCWI enables a huge signal-to-noise (S/N) increase over other instruments. For example, the S/N of our spatially-resolved spectra ranges from $12 - 20\text{\AA}^{-1}$ which is higher than the integrated spectra of other UDGs previously studied (e.g., Gu et al. 2018b; Ferré-Mateu et al. 2018). We also make an integrated spectrum of DF44 by bootstrapping for the median of the individual spectra. We used the 50th percentile from the resulting distribution of flux at a given wavelength as the stacked spectrum and used the 16th and 84th percentiles as the uncertainties on the stacks to achieve S/N $\sim 60\text{\AA}^{-1}$.

To extract stellar population parameters from the data we use the full-spectrum SPS models (`alf`) described in Conroy et al. (2018). The most relevant update of the Conroy et al. (2018) models with regards to this work is the expansion of stellar parameter coverage of the models with the Spectral Polynomial Interpolator (SPI, Villaume et al. 2017b)[14]. With SPI we used the optical MILES stellar library (Sánchez-Blázquez et al. 2006b), the Extended IRTF stellar library (E-IRTF, Villaume et al. 2017b), and a large sample of M Dwarf spectra (Mann et al. 2015) to create a data-driven model from which we can generate stellar spectra as a function of effective temperature, surface gravity, and metallicity.

The empirical parameter space is set by the E-IRTF and Mann et al. (2015)

---

[14] https://github.com/AlexaVillaume/SPI_Utils

samples which together span $-2.0 \lesssim$ [Fe/H] $\lesssim +0.5$ and $3.9 \lesssim \log \mathrm{T_{eff}} \lesssim 3.5$. To preserve the quality of interpolation at the edges of empirical parameter space we augment the training set with a theoretical stellar library (C3K). The `alf` models allow for variable abundance patterns by differentially including theoretical element response functions.

We used the medium slicer on KCWI, yielding a spectral resolution of $R \sim 4000$ which necessitated smoothing the data to the native resolution of the `alf` models, 100 km/s, It has been previously demonstrated that the accuracy of the recovered stellar parameters is not affected by smoothing (see Appendix A in Choi et al. 2014).

Before smoothing, we interpolate over bad pixels that remain in the spectra. We test two techniques for this interpolation – a simple linear interpolation (closed blue circles) and interpolation using a kernel (open green circles). We smooth each spectrum to the desired velocity resolution by convolving a wavelength dependent Gaussian kernel with $\sigma = \sqrt{\sigma_D^2 - \sigma_I^2}$, where σ_D is the desired resolution and σ_I is the instrumental resolution (see Figure 9 of Paper I).

We fit over wavelength regions $4800 \leq \lambda\text{Å(obs)} \leq 5150$ and $5150 \leq \lambda\text{Å(obs)} \leq 5300$. Due to our limited S/N and wavelength region we use `alf` in simple mode using the standard priors (see Conroy et al. 2018, for details), except for the prior on [Mg/H]. Based on examination of the posteriors, we changed the lower prior to be [Mg/H]$= -1.0$. Due to the S/N limitations of the aperture spectra, the results we derive in this work are all derived from using the “simple” mode in `alf`.

6.3 Results and Discussion

Figure 6.1 summarizes the results from applying the SPS models to the DF44 data from left to right: stellar age, [Fe/H], and [Mg/Fe]. We show the measurement from the integrated spectrum in each panel (open star). For age and [Fe/H] we compare our integrated measurements with those from Gu et al. (2018b) (black square).

We also show the radial gradients for each parameter (black circles). We used linear regression to measure the slopes of each gradient: $m_{\rm age} \sim +0.01^{+0.08}_{-0.07}$, $m_{\rm [Fe/H]} \sim +0.07^{+0.19}_{-0.03}$, and $m_{\rm [Mg/Fe]} \sim -0.19^{-0.02}_{-0.36}$, the range on each value indicates the 16th and 84th percentile values of the posterior distribution. We show the "best-fit" line in Figure 6.1 (black) and the range between the 16th and 84th percentiles (grey band).

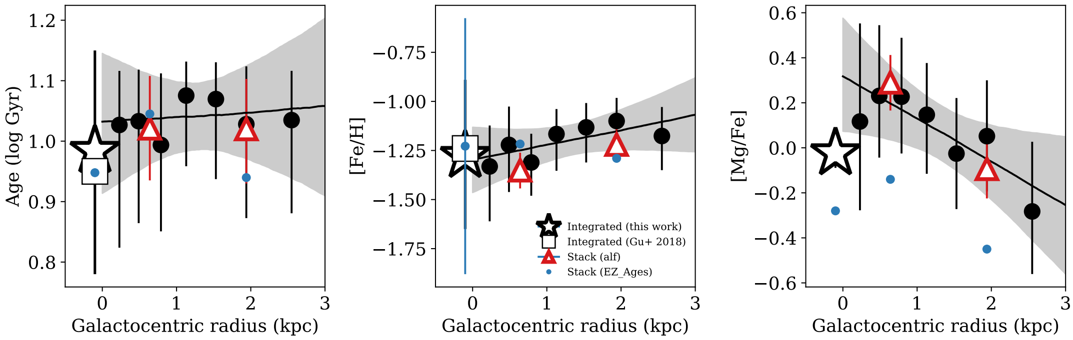

Figure 6.1: Radial profiles of stellar population parameters (black circles), measurements from the integrated spectrum (star) compared to the integrated measurements from Gu et al. (2018b), and measurements of two radial bins using both `alf` (red triangles) and `EZ_AGES` (blue circles). From left to right: stellar age, [Fe/H], and [Mg/Fe].

At low-redshift, it is more typical to find negative [Fe/H] gradients and flat-to-rising [α/Fe] gradients for massive early-type galaxies (ETGs, see Chapter 5) and dwarf spheroidals (e.g., Kirby et al. 2009). Although, the large sample in Greene et al.

(2019) revealed at least one massive ETG with a positive [Fe/H] gradient and three with negative [α/Fe] gradients. Furthermore, early work by Efstathiou & Gorgas (1985) showed declining strength in the $\mathrm{Mg_2}$ index with radius in the massive ETG NGC 5813. The negative-to-slightly positive age gradient seems typical of what is found in local dwarf galaxies, younger stars are found in the centers and older stars in the outskirts (e.g., Graus et al. 2019, and the references therein).

While not completely unprecedented, the population gradients we measure for DF44 add to its ambiguity, especially the very steep [Mg/Fe] gradient. We preform several tests to check the veracity of the gradient results. First, we created two spatial bins, the first includes the inner four apertures and the second includes the outer three apertures, achieving S/N $\sim$ 30 for each. These spatial stacks were fitted using `alf` (red triangles). We see that the measurements from the spatial stacks correspond to both the overall value and general direction of the gradients. Additionally, we fitted both the spatial stacks and the integrated spectrum with `EZ_AGES` Graves & Schiavon (2008)[15] (blue circles) to check for possible model-based systematics. For stellar age and [Fe/H] there is good agreement between `alf` and `EZ_Ages` measurements of the integrated spectrum. There are slight discrepancies in the best-fit values for some of the spatial stacks but the values are within the measurement uncertainties. For [Mg/Fe] the `EZ_Ages` measurements are overall lower than the `alf` measurements but `EZ_Ages` is not reliable for detailed abundances at [Fe/H] $\lesssim -1.0$. Moreover, the overall trend of the `EZ_Ages` measurements are consistent with the `alf` values.

---

[15] We measured the Hβ, Fe5015, Mgb, C4668, CN2, and Ca4227 indices after correcting the spectra to restframe using the `alf` measurements of recession velocity for each. For the fitting, we used the solar-scaled isochrones, an IMF exponent of 1.35, and [Ti/Fe] = 0.

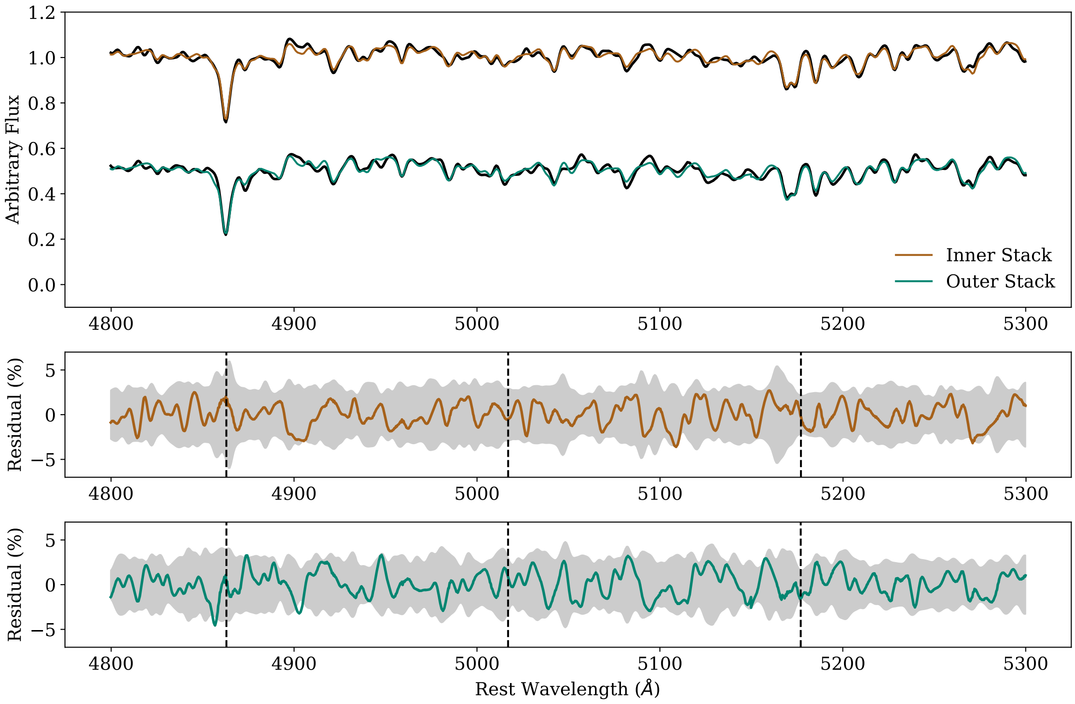

Figure 6.2: (Top) Comparison of spatial stacks (black) and best-fit models for the inner stack (brown) and outer stack (green). (Middle) Comparison of residuals between best-fit model and data for inner stack and uncertainty of flux from the input spectrum (grey). (Bottom) Same as middle panel but for outer stack. Location of Hβ, Fe5017, and Mg*b* are highlighted in middle and bottom panels (dashed lines).

Second, in the top panel of Figure 6.2, we compare the spatially stacked spectra to the best-fit inner (brown) and outer (green) `alf` models. In the middle and bottom panels, we compare the residuals between the data and best-fit spectrum for the inner and outer stacks, respectively. We also show the uncertainty in the data in each panel (grey band) to show that the best-fit model recovers the data within the uncertainty. The key spectral features Hβ, Fe5017, and Mg*b* are highlighted in the residual panels to show that these features are recovered by the model.

These tests confirm that neither S/N limitations nor model systematics are biasing the overall gradients. Additionally, in Figure 6.3 we check for any systematic problems that may reside in the data itself. Because of the low-surface brightness of UDGs, we have particular concern over contamination by the sky background and the possibility that the smoothing smears bad pixels throughout the spectrum.

In the top row of Figure 6.3, we show the inferred values for [Fe/H], Age, and [Mg/Fe] from fitting the inner most spectrum after smoothing to different resolutions. The σ-dependent differences in the inferred parameters are much smaller than the uncertainties on those parameters and there is little sensitivity to the interpolation technique. This indicates that the results from our models are insensitive to the processing we need to do to the spectra. In the bottom row we compare the uncertainty on the inferred parameters for our lowest S/N spectrum (closed black circle) and mock data generated with a similar S/N (open blue circle). If the observations had systematic uncertainty significantly affecting the spectral features we would expect the parameter uncertainties to be much larger than the parameter uncertainties for the mock data, which only include statistical uncertainty.

Finally, with regards to the potential effects of data systematics, we note the excellent agreement between the measurements presented in this work and Gu et al. (2018b). Gu et al. (2018b) also used the `alf` SPS models but their data were acquired on an entirely different telescope and instrument.

We computed a synthetic color profile from the parameters (blue) to the observed color profile (grey, van Dokkum et al. 2019) in Figure 6.4. We find that there is a small overall offset (~ 0.1) between the synthetic and observed color profiles. This is

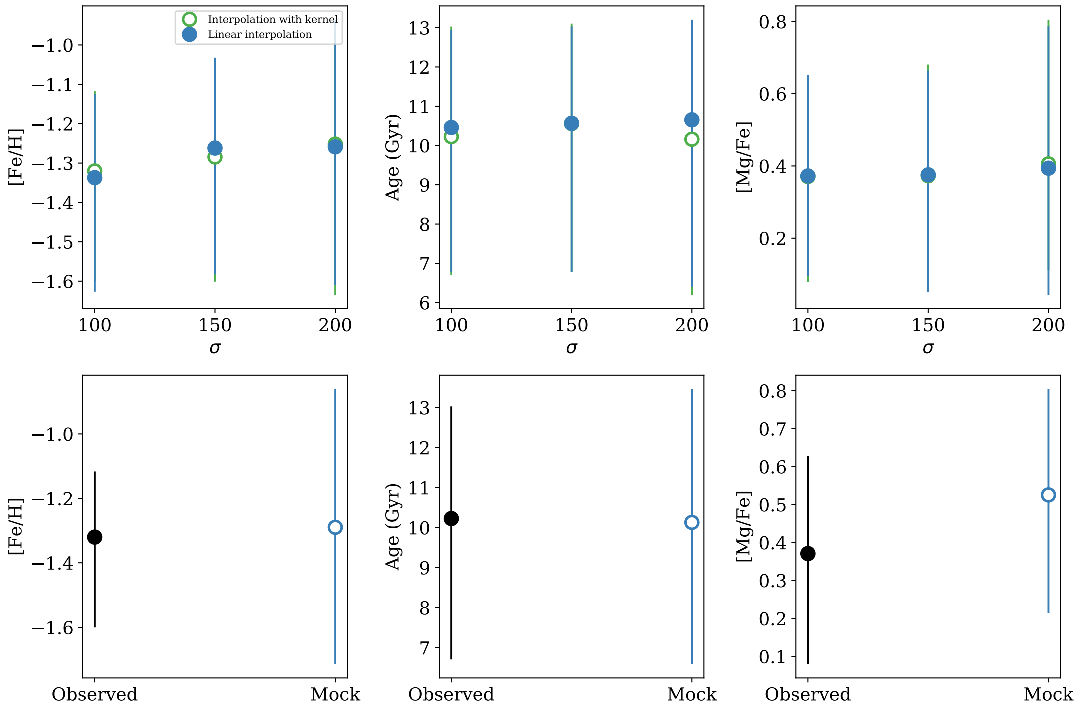

Figure 6.3: (Top row) For the Aperture 0 spectrum looking at how the level of smoothing impacts the stellar parameters when linearly interpolating over bad pixels (closed blue circles) and using a kernel to interpolate over bad pixels (open green circles). (Bottom row) Statistical uncertainty appears in mock data (open blue circles) appears to be representative of the relevant uncertainty in the observed data (closed black circles).

possibly due to issues in the isochrones for the upper giant branch (Choi et al. 2016) but synthetic colors have been generated for globular clusters in the Milky Way and M87 (see Villaume et al. 2019) without this discrepancy. Another possibility is the presence of dust in DF44. Independent work modeling the broadband spectral energy distribution of DF44 (filter to filter) indicates this with an overall dust attenuation of $A_V \sim 0.25$ (S. Laine, private communication). We show that applying this dust attenuation value to the synthetic color profile (red) resolves the discrepancy in Figure 6.4. Within the large observational uncertainties, the overall behavior of the color profiles are consistent

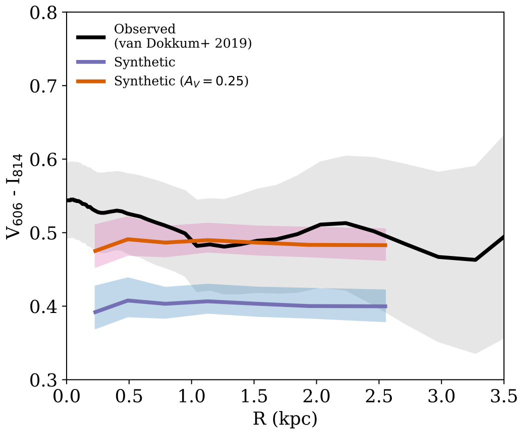

Figure 6.4: Comparison of observed color profile (grey) to synthetic color profiles without dust extinction (blue) and with dust extinction (red, see Section 6.3, for details).

with each other.

In Figure 6.5 we contextualize DF44 with other objects in [Fe/H]–[Mg/Fe] space using measurements from both the integrated spectrum (left) and the spatially-resolved spectra (right). In the left panel, we compare DF44 (open star) to other UDGs with measured [Mg/Fe] values. In this context, DF44 appears to be a middling UDG, far from the extreme [Mg/Fe] value of DGSAT I (Martín-Navarro et al. 2019) and not on any particularly end of the range spanned by the Coma UDGs measured by Ferré-Mateu et al. (2018).

However, the integrated [Mg/Fe] measurement elides the highly unusual gradient seen in the spatially-resolved measurements. In the right panel of Figure 6.5 we compare the DF44 measurements (black points) to the measurements of the integrated M87 starlight (green circles), and the measurements from the GC stacks of the inner halo M87 GC population ($R_{\rm gal} < 40$ kpc, open blue squares) and outer halo GC population ($40 < R_{\rm gal}$ kpc < 140), closed blue squares) from Chapter 5. The center of DF44

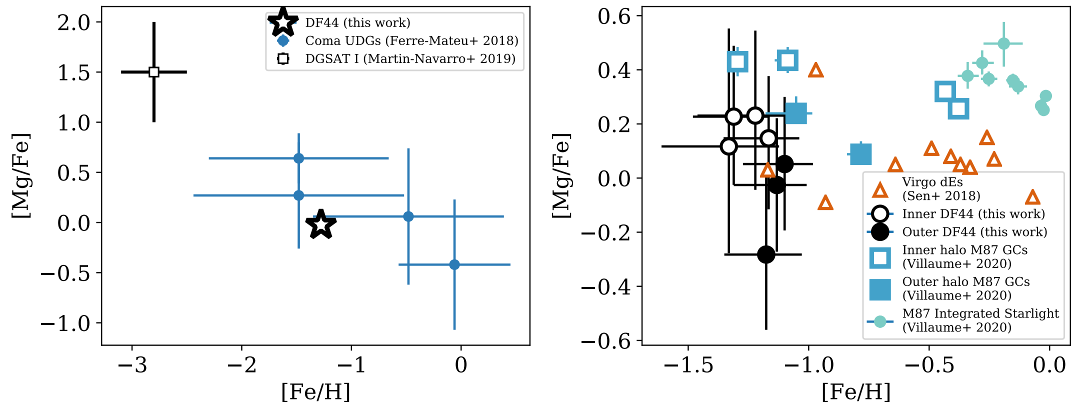

Figure 6.5: (Left) Comparing the integrated measurements of DF44 to other UDGs with [Mg/Fe] measurements including the very unusual DGSAT I (open square) and other UDGs in the Coma Cluster (blue circles). (Right) Comparing the spatially-resolved measurements of DF44 to spatially-resolved measurements from integrated starlight for M87 (green circles), spectral stacks of M87's inner halo GC population ($R_{\rm gal} < 40$ kpc, open blue squares) and outer halo GC population ($40 < R_{\rm gal}$ kpc < 140), closed blue squares), and dwarf ellipticals in Virgo (black triangles). The DF44 measurements are split between the inner sample ($R_{\rm gal} < 1.5$ kpc, open black circles) and the outer sample (closed black circles)

is consistent with the metal-poor ([Fe/H] $\lesssim -1.0$) M87 GCs in this space, while the outer region of DF44 is consistent with the values of more typical dwarf galaxies (black triangles Şen et al. 2018).

Several formation scenarios for UDGs have been predicted by simulations. Most predict that UDGs are simply puffed up dwarf galaxies from either high-spin (Amorisco & Loeb 2016; Tremmel et al. 2020) or a bursty star formation history that has continued to the present day (Di Cintio et al. 2017). A scenario that has not been predicted by simulations is that UDGs are "pure stellar halos", that is, massive galaxies that somehow failed to fulfill their potential for forming stars (Peng & Lim 2016). This scenario was established to explain the abundant GC systems around many UDGs (e.g., van Dokkum et al. 2017b; Lim et al. 2018; van Dokkum et al. 2018) since massive star cluster formation requires epochs of intense star formation. It has been suggested that "UDGs" as a class are actually composed of at least two types, with distinct origins (Forbes et al. 2020).

The flat rotation curve seen measured for DF44 (van Dokkum et al. 2019) rules out the first scenario scenario and the uniformly old ages rule out the feedback scenario. The metal-poor, α-enhanced inner 1.5 kpc of DF44 aligns well with the prediction from the pure stellar halo scenario that the initial star-formation would be intense, but truncated. The other properties of DF44 also consistent with this scenario, with a high stellar velocity dispersion ($\sigma_* \sim 33$ km/s, van Dokkum et al. 2019) and populous GC system ($N_{GC} \sim 74$, van Dokkum et al. 2017b). Moreover, the age and [Fe/H] gradients we find in this work are not typical of nearby dwarf galaxies.

However, the steep [Mg/Fe] gradient still needs to be explained. In principle, we might explain the positive [Fe/H] gradient seen in DF44 by invoking the fact that, even though it has a low stellar mass, its halo mass might be sufficient to bring in satellite galaxies of higher metallicity than expected from the mass-metallicity relation. In this scenario the [Mg/Fe] could be explained if this population had more extended star-formation histories than M87. However, with this scenario there would be an expectation of a negative age gradient since the higher metallicity, α-depressed satellites would have had more recent star-formation.

The comparison of DF44 to M87 and its GC system hints at a resolution to the tension between the predictions from cosmological simulations of the progenitor populations of stellar halos in massive galaxies and the constraints provided by GCs discussed in Villaume et al. (2019) and 5. With the caveat that the apples-to-apples comparison would be between the GC systems of DF44 and M87, the result shown here strengthens the possibility that UDGs are the progenitors of the stellar halos of massive ETGs. This, of course, would provide additional constraints on the formation histories of UDGs themselves as it would indicate that what we see now are only the surviving population. We could potentially use stellar halos of massive ETGs to predict the primordial population of UDGs.

6.4 Summary

We have presented the first spatially-resolved stellar population study for a UDG, DF44 using optical spectra from KCWI on the Keck II telescope. We summarize our results as follow:

- We measure the gradients of the stellar parameters: $m_{\log\ \mathrm{age}} \sim +0.01^{+0.08}_{-0.07}$, $m_{\mathrm{[Fe/H]}} \sim +0.07^{+0.19}_{-0.03}$, and $m_{\mathrm{[Mg/Fe]}} \sim -0.19^{-0.02}_{-0.36}$.
- The inner 1.5 kpc of DF44 has GC-like abundance patterns while the outer region is more consistent with the abundances of dwarf ellipticals. These results are a promising path to relieving the tension between the predictions from cosmological simulations of the progenitors of massive ETGs stellar halos and the constraints from GCs.
- The flat age gradient is not consistent with an accretion scenario to explain the steep [Mg/Fe] gradient which suggests that the abundance gradients were formu-

lated with the initial star-formation and assembly of the galaxy.

Chapter 7

Summary and Future Directions

What we see, we see

and seeing is changing

the light that shrivels a mountain

and leaves a man alive

Adrienne Rich, Planetarium

7.1 Summary

This thesis used full-spectrum SPS models to examine the star-formation and assembly processes at the extremes of galaxy type. Chapter 2 details the stellar library that was used to make the SPS models. The library greatly expanded the stellar parameter coverage of the full-spectrum models, which is what enabled their application to such a diverse array of stellar populations. Chapter 3 examined the behavior of the stellar M/L measurements in a sample of low-velocity, compact stellar systems where I allowed for a variable IMF in the models. These results represent the first time IMF measurements from integrated light have been extracted from these kinds of objects. They show that the IMF varies less on the whole for the compact objects than for massive ETGs from which it must be concluded that metallicity is not the sole driver of IMF variability, as has been suggested.

Chapter 4 uses the models to extract [Fe/H] measurements from a large sample of individual GCs around M87 that (i) confirm metallicity bimodality in the inner ~ 10 kpc of M87 and (ii) show that at fixed, low-metallicity the M87 GCs are bluer than the Milky Way GCs. This latter result not only has far-reaching implications on the interpretation of the many large imaging data sets of extragalactic GCs which exist, but also indicate that the internal properties of GCs are not as well-understood as perhaps hoped.

In Chapter 5, I model the metallicity gradient of the individual GCs using a hierarchical Bayesian statistical framework which allowed me to (i) account for the covariance between the subpopulation membership assignments of the individual GCs and the gradient parameters I was trying to measure for the subpopulations and (ii) statistically de-project the galactocentric distances of the GCs. This allowed me to measure the GC subpopulation gradients more accurately than previously possible. To this end, I measured remarkably flat gradients in both the metal-poor and metal-rich subpopulations, indicating that they were both affected by the same physical processes. The presence of the metal-poor subpopulation is in tension with predictions from cosmological simulations.

Finally, in Chapter 6, I present the first radial stellar population gradients for a

UDG, DF44. I find mostly flat age and [Fe/H] gradients but a steeply negative [Mg/Fe] gradient which, on the whole, leaves it ambiguous whether DF44 was affected by some accretion events or formed more or less in a monolithic collapse.

7.2 Future directions

The main thrust of this thesis is to emphasize the critical fossil record GCs hold and the importance of modeling them accurately to extract that information, both as individual objects and as systems of objects. Observations of GCs need to prioritized along with their host systems. By analyzing the GCs of M87 in tandem with its integrated galaxy starlight, I was able to establish essential benchmarks to guide galaxy evolution theories. For DF44, understanding the stellar population properties of its GC system could significantly clarify the meaning of the measured parameters for DF44 itself and UDGs as a whole (more on this below).

However, this work also opens many questions that need to be addressed. First, the internal properties of GCs, and how they may be affected by environment must be better understood. Star cluster formation is intrinsically linked to star formation processes, which in several ways is still poorly understood. My result in Chapter 4 significantly alters the interpretations of the data from photometric surveys. The differences I found in the colors of M87 and Milky Way GCs at fixed metallicity might be attributable to an age difference, with the M87 blue GCs being up to 4 Gyrs younger than the Milky Way blue GCs. Another explanation for the color difference is that the M87 GCs could also host more significant populations of blue horizontal branch stars than the Milky Way. Establishing the underlying cause of the discrepancy has the potential to change our interpretation of the origins of the GCs.

Second, there have been several proposals on how massive ETGs evolve from high-redshift (Barro et al. 2013; van Dokkum et al. 2015). However, there is still a lack of clarity about their initial star-formation processes and subsequent assembly. So, while there is a lot known about massive ETGs, the overall impression is that the whole is less than the sum of the parts. A significant source of this ambiguity is the limited means to decompose the stellar populations of these systems into *in-* and *ex-situ* populations. GC systems will be key to making this advancement, but improvements to both how individual GCs are modeled and the statistical frameworks in which systems of GCs are modeled need to be improved.

In the following, I detail the projects that I think are necessary to achieve these goals.

7.3 High priority projects

7.3.1 Breaking the degeneracy between age and horizontal branch

Age measurements for GCs from integrated light are highly uncertain, not due to limitations in S/N but because of the degeneracy caused by the presence of the blue horizontal branch. The horizontal branch is a distinct population of stars outside the main track predicted by stellar evolution physics. The physics that determine the formation of horizontal branch stars is not well understood and great diversity in horizontal branch morphology is observed in Milky Way and M31 GCs. The morphology of the horizontal branch is strongly influenced by metallicity of the GC but that alone does not explain the different morphologies seen among the GCs (Rood 1973; Dotter et al. 2010).

Consequently, horizontal branch stars that contribute to the integrated light spectra of extragalactic GCs are not directly modeled by current SPS models. This is an issue because this population also tends to be bright and blue which means their presence can make stellar populations look artificially young. This has been the primary hurdle for obtaining accurate ages from integrated light for extragalactic GCs. The effect of this can be seen in the age determinations for some of the M31 GCs in Chapter 3.

The problem is difficult but not insurmountable. First, the great strength of Bayesian modeling is the ability to marginalize over nuisance parameters. The morphology of the horizontal branch, that is, how blue or red it is, can be characterized by the peak and width of the distribution of stars in L_{bol} vs. T_{eff} space. If these parameters (μ_{HB}, σ_{HB}) can be inferred, they can be used when "accounting" for the inferred ages.

So, how to infer these parameters? By creating data-driven isochrones directly from the CMDs of Milky Way GCs. A new library of horizontal branch star spectroscopy (Villaume et al., work in progress) can be tied to the stars in the Milky Way horizontal branches and summed according to the different morphologies. This will create empirically-determined predictions of the contribution of the horizontal branch stars to the integrated spectrum of a stellar population.

With this, horizontal branch "response functions" (directly analogous to the response functions needed to obtain variable abundance variations) will be known as a function of metallicity, age, and individual chemical abundances for the Milky Way GCs. These are the characteristics the SPS models optimize for, and so the horizontal branch contribution can easily be incorporated as a parameter to be optimized for a best-fit model.

The remaining question is whether there is sufficient information content in an integrated spectrum for the model described above to be reliable. Schiavon et al. (2004) showed that the ratio between Hδ_F and Hβ is far more sensitive to horizontal branch morphology, than to age. Further information about age could be extracted if the observed spectra were flux calibrated and this was incorporated as part of the model. Currently, `alf` continuum normalizes the observed spectra before the fit. However, this is more due to uncertain effects of dust than to inherent limitations in the models. In fact, as shown in Chapter 2, the underlying empirically stellar library is exquisitely flux calibrated. Spectroscopy that spans the optical to near-IR would be particularly useful in this respect as the overall continuum shape would provide significant age information.

More accurately determining the ages and the horizontal branch morphology in extragalactic GCs will provide significant information about how formation and evolution of massive clusters may change with environment.

7.3.2 Extending the HBM to determine *in-* and *ex-situ* populations in GC systems

The study of extragalactic GC systems has been rooted in the paradigm of bimodality – the metal-poor and metal-rich subpopulations. This has greatly influenced the way the assembly of GC systems and their host galaxies have been framed. The metal-rich GCs are thought to have formed within massive galaxies at the peak of star formation, in contrast, the metal-poor objects are thought to have formed in low mass galaxies at even earlier times and then fell into galaxy halos.

The reality is almost certainly more complicated than this, with both the metal-poor and metal-rich subpopulations consisting of a mix of *in-* and *ex-situ* populations in the inner halos of massive galaxies. Being able to more precisely separate GC systems into subpopulations will help determine which is which. This will also help clarify the the origins of the GCs, and therefore the star-formation and assembly history of their present-day host galaxy.

The HBM presented in Chapter 5 does not require the system be separated into two subpopulations, that was done because of convention but also because metallicity and galactocentric distance alone are not sufficient to make finer-grained determinations. The Milky Way GC system has been shown to have an age–metallicity distribution (AMD, Leaman et al. 2013). There is currently a lot of excitement in the GC community that the AMD could be used to infer galaxy assembly histories (Muratov & Gnedin 2010; Kruijssen et al. 2019).

Breaking the age degeneracy will provide a crucial additional constraint to condition the HBM on. However, the cosmological simulations have limited ability to distinguish progenitor satellites of different masses in this space (O. Gnedin, private communication). In the Milky Way, it takes many dimensions of information (e.g., kinematics, chemistry, age) to reconstruct the assembly history (Helmi 2020) and so this should also be expected in extragalactic systems.

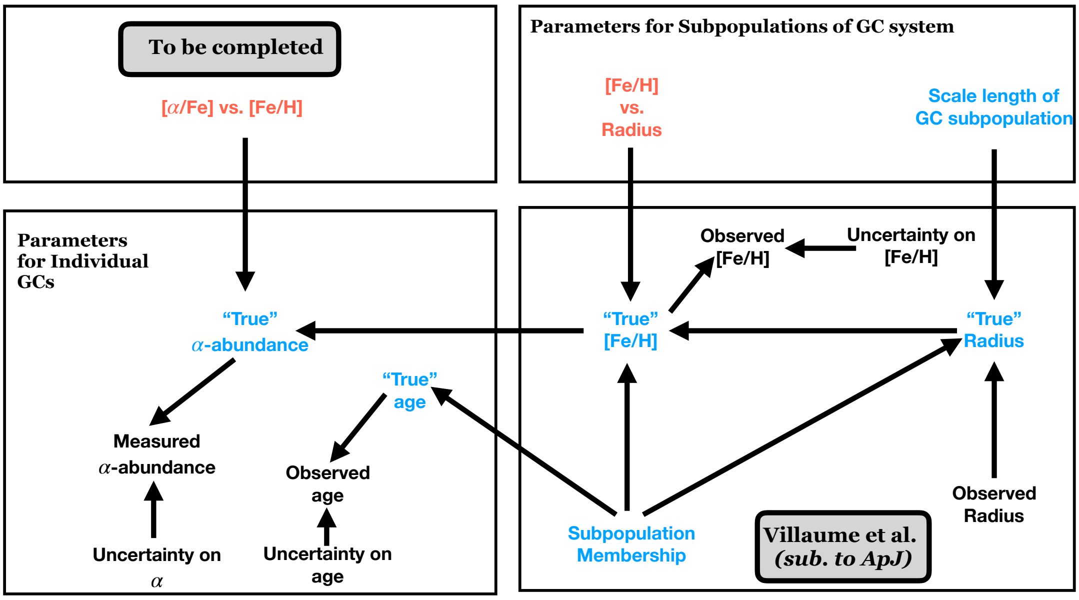

Figure 7.1: (Right column) A simplified graphical representation of the hierarchical model as presented in Chapter 5 to simultaneously model the radial metallicity gradients for the subpopulations and subpopulation membership. This figure shows the relation between the observed parameters (black), the unobserved but modelled parameters (blue), and the parameters of interest for the inference (red). (Left column) Extensions to the model needed infer α–[Fe/H] for the subpopulations and include age as a constraint for the subpopulation membership assignments.

In particular, α-elements are crucial in this task (e.g., Pritzl et al. 2005). Obtaining α-element abundances for extragalactic GCs is, however, a significant challenge. The issue is in the data itself, there are fewer and weaker spectral lines for α-element abundances than there are for measuring metallicity. This is not something that improvements to the SPS modeling can fix, it would require significant investments in telescope time to get the spectroscopy of exquisite enough quality.

A natural solution is to further extend the HBM form Chapter 5 to infer α–metallicity trends for the subpopulations. This works because the GCs in a given subpopulation will have the same origins and because α-element abundances, by their nature, will have a correlation with metallicity. In this way we can use the many, albeit noisy, measurements of α-abundance to infer the overall trends in every subpopulation. A simplified schematic of this model is shown in Figure 7.1.

Finally, a way to assess the "best" number of subpopulations in a GC system is needed. The best way to do this is move from a model where the number of subpopulations input to the model, to nonparameterized, unsupervised clustering in the form of Dirichlet Processes. The essential idea is that the number of subpopulations is also a parameter that is fit for.

More precise subpopulation separation will make all the measurements of the subpopulations, including things like metallicity gradients, more accurate and informative. For instance, I speculated in Chapter 5 that a possible reason for the flat metallicity gradient in the metal-poor subpopulation is that the accreted population is being "diluted" by an *in-situ* population that would have a flat gradient. Being able to separate what is now lumped together in the "metal-poor" or "metal-rich" subpopulations will provide actual tests of different GC formation scenarios.

7.3.3 Determining detailed stellar populations of the GC systems around UDGs

The comparison between the spatially-resolved stellar population parameters for DF44 and the GCs around M87 presented in Chapter 6 are highly suggestive of a link between stellar halos of massive ETGs and UDGs. Furthermore, comparing the stellar parameters of UDGs with their GCs could help clarify the significance of the gradients measured for DF44 (and, presumably, other UDGs in the future). With this in mind, I think the following things, more or less ordered by expected difficulty, are worthwhile to look into:

- Examine the color gradients of the GC system around DF44. Is it flat like the metallicity gradient measured for DF44 itself?
- There are now many imaging data sets of UDGs and their GC systems. Stacking the GC colors among all these systems and checking for bimodality would help constrain the formation scenarios of UDGs. The stacking could be binned along halo type: under massive, normal, over massive.
- Obtain spectroscopy of the GC system around DF44 with sufficient S/N to extract stellar population parameters from the individual GCs. This comparison would help not only to understand DF44 but also allow for an apples-to-apples comparison GCs in other systems.

Bibliography

Abohalima, A., & Frebel, A. 2018, ApJS, 238, 36

Abraham, R. G., & van Dokkum, P. G. 2014, PASP, 126, 55

Agnello, A., Evans, N. W., Romanowsky, A. J., & Brodie, J. P. 2014, MNRAS, 442, 3299

Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S8

Allard, F., Hauschildt, P. H., Alexander, D. R., & Starrfield, S. 1997, ARA&A, 35, 137

Allard, F., Homeier, D., Freytag, B., et al. 2013, Memorie della Societa Astronomica Italiana Supplementi, 24, 128

Allen, D. M., Ryan, S. G., Rossi, S., Beers, T. C., & Tsangarides, S. A. 2012, A&A, 548, A34

Alves-Brito, A., Hau, G. K. T., Forbes, D. A., et al. 2011, MNRAS, 417, 1823

Amorisco, N. C., & Loeb, A. 2016, MNRAS, 459, L51

Aoki, W., Beers, T. C., Christlieb, N., et al. 2007, ApJ, 655, 492

Aoki, W., Ito, H., & Tajitsu, A. 2012, ApJ, 751, L6

Aoki, W., Norris, J. E., Ryan, S. G., Beers, T. C., & Ando, H. 2002a, PASJ, 54, 933

—. 2002b, ApJ, 567, 1166

Aoki, W., Tominaga, N., Beers, T. C., Honda, S., & Lee, Y. S. 2014, Science, 345, 912

Aoki, W., Honda, S., Beers, T. C., et al. 2005, ApJ, 632, 611

Aoki, W., Beers, T. C., Sivarani, T., et al. 2008, ApJ, 678, 1351

Aoki, W., Beers, T. C., Lee, Y. S., et al. 2013, AJ, 145, 13

Ardila, D. R., Van Dyk, S. D., Makowiecki, W., et al. 2010, ApJS, 191, 301

Athey, A., Bregman, J., Bregman, J., Temi, P., & Sauvage, M. 2002, ApJ, 571, 272

Barbuy, B., Spite, M., Spite, F., et al. 2005, A&A, 429, 1031

Barklem, P. S., Christlieb, N., Beers, T. C., et al. 2005, A&A, 439, 129

Barmby, P., Huchra, J. P., Brodie, J. P., et al. 2000, AJ, 119, 727

Barro, G., Faber, S. M., Pérez-González, P. G., et al. 2013, ApJ, 765, 104

Bastian, N., & Lardo, C. 2017, ArXiv e-prints, arXiv:1712.01286

Baum, W. A. 1955, PASP, 67, 328

Beasley, M. A., Romanowsky, A. J., Pota, V., et al. 2016, ApJ, 819, L20

Beasley, M. A., & Trujillo, I. 2016, ApJ, 830, 23

Beifiori, A., Maraston, C., Thomas, D., & Johansson, J. 2011, A&A, 531, A109

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

Bensby, T., Adén, D., Meléndez, J., et al. 2011, A&A, 533, A134

Bertone, E., Buzzoni, A., Chávez, M., & Rodríguez-Merino, L. H. 2008, A&A, 485, 823

Betancourt, M. 2017, arXiv e-prints, arXiv:1701.02434

Bird, S., Harris, W. E., Blakeslee, J. P., & Flynn, C. 2010, A&A, 524, A71

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

Blom, C., Spitler, L. R., & Forbes, D. A. 2012, MNRAS, 420, 37

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bonifacio, P., Sbordone, L., Marconi, G., Pasquini, L., & Hill, V. 2004, A&A, 414, 503

Bothun, G. D., Impey, C. D., & Malin, D. F. 1991, ApJ, 376, 404

Brodie, J. P., & Huchra, J. P. 1991, ApJ, 379, 157

Brodie, J. P., & Strader, J. 2006, Annual Review of Astronomy and Astrophysics, 44, 193

Brodie, J. P., Usher, C., Conroy, C., et al. 2012, ApJ, 759, L33

Brodie, J. P., Romanowsky, A. J., Strader, J., et al. 2014, ApJ, 796, 52

Brott, I., & Hauschildt, P. H. 2005, in ESA Special Publication, Vol. 576, The Three-Dimensional Universe with Gaia, ed. C. Turon, K. S. O'Flaherty, & M. A. C. Perryman, 565

Bundy, K., Fukugita, M., Ellis, R. S., et al. 2009, ApJ, 697, 1369

Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7

Burstein, D., Faber, S. M., & Gonzalez, J. J. 1986, AJ, 91, 1130

Caldwell, N., & Romanowsky, A. J. 2016, ApJ, 824, 42

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

Cappellari, M., Scott, N., Alatalo, K., et al. 2013, MNRAS, 432, 1709

Carretta, E., Bragaglia, A., Gratton, R. G., et al. 2010, A&A, 516, A55

Carretta, E., Gratton, R., Cohen, J. G., Beers, T. C., & Christlieb, N. 2002, AJ, 124, 481

Castelli, F., & Kurucz, R. L. 2003, in IAU Symposium, Vol. 210, Modelling of Stellar Atmospheres, ed. N. Piskunov, W. W. Weiss, & D. F. Gray, A20

Cayrel, R., Depagne, E., Spite, M., et al. 2004, A&A, 416, 1117

Cenarro, A. J., Beasley, M. A., Strader, J., Brodie, J. P., & Forbes, D. A. 2007, AJ, 134, 391

Cenarro, A. J., Gorgas, J., Cardiel, N., et al. 2001, MNRAS, 326, 981

Cenarro, A. J., Gorgas, J., Vazdekis, A., Cardiel, N., & Peletier, R. F. 2003, MNRAS, 339, L12

Chabrier, G., Hennebelle, P., & Charlot, S. 2014, ApJ, 796, 75

Chen, Y.-P., Trager, S. C., Peletier, R. F., et al. 2014, The Messenger, 158, 30

Choi, J., Conroy, C., Moustakas, J., et al. 2014, ApJ, 792, 95

Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102

Choksi, N., Gnedin, O. Y., & Li, H. 2018, MNRAS, 480, 2343

Chung, C., Lee, Y.-W., Lim, D.-W., & Hong, S. 2018, in The Galactic Bulge at the Crossroads (GBX2018, 6

Clough, S. A., M. W. S. E. J. M. J. S. D. M. J. I. K. C. S. B., & Brown, P. D. 2005, J. Quant. Spectrosc. Radiat. Transfer, 233

Code, A. D., & Welch, G. A. 1979, ApJ, 228, 95

Coelho, P., Barbuy, B., Meléndez, J., Schiavon, R. P., & Castilho, B. V. 2005, A&A, 443, 735

Cohen, J. G. 1978, ApJ, 221, 788

Cohen, J. G., Blakeslee, J. P., & Côté, P. 2003a, ApJ, 592, 866

Cohen, J. G., Blakeslee, J. P., & Ryzhov, A. 1998, ApJ, 496, 808

Cohen, J. G., Christlieb, N., Qian, Y. Z., & Wasserburg, G. J. 2003b, ApJ, 588, 1082

Cohen, J. G., Christlieb, N., Thompson, I., et al. 2013, ApJ, 778, 56

Cohen, J. G., & Huang, W. 2009, ApJ, 701, 1053

Cohen, J. G., & Ryzhov, A. 1997, ApJ, 486, 230

Cohen, J. G., Christlieb, N., McWilliam, A., et al. 2004, ApJ, 612, 1107

Cohen, J. G., McWilliam, A., Shectman, S., et al. 2006, AJ, 132, 137

Collins, M. L. M., Tollerud, E. J., Rich, R. M., et al. 2020, MNRAS, 491, 3496

Colucci, J. E., Bernstein, R. A., & Cohen, J. G. 2014, ApJ, 797, 116

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

Conroy, C., Graves, G. J., & van Dokkum, P. G. 2014, ApJ, 780, 33

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Conroy, C., & van Dokkum, P. G. 2012a, ApJ, 760, 71

—. 2012b, ApJ, 760, 71

Conroy, C., van Dokkum, P. G., & Villaume, A. 2017, ApJ, 837, 166

Conroy, C., Villaume, A., van Dokkum, P. G., & Lind, K. 2018, ApJ, 854, 139

Côté, P., West, M. J., & Marzke, R. O. 2002, ApJ, 567, 853

Côté, P., Blakeslee, J. P., Ferrarese, L., et al. 2004, The Astrophysical Journal Supplement Series, 153, 223

Cowan, J. J., Sneden, C., Burles, S., et al. 2002, ApJ, 572, 861

Şen, Ş., Peletier, R. F., Boselli, A., et al. 2018, MNRAS, 475, 3453

Cui, W. Y., Sivarani, T., & Christlieb, N. 2013, A&A, 558, A36

Cushing, M. C., Vacca, W. D., & Rayner, J. T. 2004, PASP, 116, 362

Dabringhausen, J., Kroupa, P., Pflamm-Altenburg, J., & Mieske, S. 2012, ApJ, 747, 72

Dalcanton, J. J., Spergel, D. N., Gunn, J. E., Schmidt, M., & Schneider, D. P. 1997, AJ, 114, 635

Di Cintio, A., Brook, C. B., Dutton, A. A., et al. 2017, MNRAS, 466, L1

Di Matteo, P., Pipino, A., Lehnert, M. D., Combes, F., & Semelin, B. 2009, A&A, 499, 427

Diaz, A. I., Terlevich, E., & Terlevich, R. 1989, MNRAS, 239, 325

Dotter, A., Sarajedini, A., Anderson, J., et al. 2010, ApJ, 708, 698

Dries, M., Trager, S. C., & Koopmans, L. V. E. 2016, MNRAS, 463, 886

Efstathiou, G., & Gorgas, J. 1985, MNRAS, 215, 37P

Emsellem, E., Cappellari, M., Peletier, R. F., et al. 2004, MNRAS, 352, 721

Faber, S. M. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & D. C. Larson, Richard B. Gehret, 157

Faber, S. M., & French, H. B. 1980, ApJ, 235, 405

Faber, S. M., Friel, E. D., Burstein, D., & Gaskell, C. M. 1985, The Astrophysical Journal Supplement Series, 57, 711

Feltzing, S., Eriksson, K., Kleyna, J., & Wilkinson, M. I. 2009, A&A, 508, L1

Ferrarese, L., Côté, P., Cuilland re, J.-C., et al. 2012, The Astrophysical Journal Supplement Series, 200, 4

Ferré-Mateu, A., Alabi, A., Forbes, D. A., et al. 2018, MNRAS, 479, 4891

Ferreras, I., La Barbera, F., de La Rosa, I. G., et al. 2013, MNRAS, 429, L15

Fitzpatrick, E. L. 1999, Publications of the Astronomical Society of the Pacific, 111, 63

Fitzpatrick, E. L., & Massa, D. 2007, ApJ, 663, 320

Forbes, D. A., Alabi, A., Romanowsky, A. J., Brodie, J. P., & Arimoto, N. 2020, MNRAS, 175

Forbes, D. A., Pastorello, N., Romanowsky, A. J., et al. 2015, MNRAS, 452, 1045

Forbes, D. A., & Remus, R.-S. 2018, MNRAS, 479, 4760

Forbes, D. A., Alabi, A., Brodie, J. P., et al. 2017, AJ, 153, 114

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013a, PASP, 125, 306

—. 2013b, PASP, 125, 306

Forte, J. C., Vega, E. I., & Faifer, F. 2012, MNRAS, 421, 635

François, P., Monaco, L., Bonifacio, P., et al. 2016, A&A, 588, A7

France, K., Shkolnik, E., Linsky, J., et al. 2015, ArXiv e-prints, arXiv:1505.01840

Frebel, A., Simon, J. D., Geha, M., & Willman, B. 2010, ApJ, 708, 560

Frogel, J. A., Persson, S. E., Matthews, K., & Aaronson, M. 1978, ApJ, 220, 75

Fulbright, J. P. 2000, AJ, 120, 1841

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

Gebhardt, K., & Kissler-Patig, M. 1999, AJ, 118, 1526

Geisler, D., Smith, V. V., Wallerstein, G., Gonzalez, G., & Charbonnel, C. 2005, AJ, 129, 1428

Gelman, A., Carlin, J. B., Hal, S. S., et al. 2013, Bayesian Data Analysis

Gilbert, K. M., Kirby, E. N., Escala, I., et al. 2019, ApJ, 883, 128

Gilmore, G., Norris, J. E., Monaco, L., et al. 2013, ApJ, 763, 61

González Hernández, J. I., & Bonifacio, P. 2009, A&A, 497, 497

Goodman, J., & Weare, J. 2010, Commun. Appl. Math. Comput. Sci., 5, 65

Gorgas, J., Faber, S. M., Burstein, D., et al. 1993, ApJS, 86, 153

Goudfrooij, P. 2018, ApJ, 857, 16

Grand, R. J. J., Gómez, F. A., Marinacci, F., et al. 2017, MNRAS, 467, 179

Graus, A. S., Bullock, J. S., Fitts, A., et al. 2019, MNRAS, 490, 1186

Graves, G. J., & Schiavon, R. P. 2008, ApJS, 177, 446

Greene, J. E., Janish, R., Ma, C.-P., et al. 2015, ApJ, 807, 11

Greene, J. E., Murphy, J. D., Graves, G. J., et al. 2013, ApJ, 776, 64

Greene, J. E., Veale, M., Ma, C.-P., et al. 2019, ApJ, 874, 66

Gregg, M. 2001, A Next Generation Spectral Library of Stars, HST Proposal

Grundahl, F., Clausen, J. V., Hardis, S., & Frandsen, S. 2008, A&A, 492, 171

Gu, M., Conroy, C., & Brammer, G. 2018a, ApJ, 862, L18

Gu, M., Conroy, C., Law, D., et al. 2018b, ApJ, 859, 37

Gültekin, K., Richstone, D. O., Gebhardt, K., et al. 2009, ApJ, 698, 198

Gunn, J. E., & Stryker, L. L. 1983, ApJS, 52, 121

Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 288, Stellar Atmosphere Modeling, ed. I. Hubeny, D. Mihalas, & K. Werner, 331

Hanes, D. A., Côté, P., Bridges, T. J., et al. 2001, ApJ, 559, 812

Hansen, T., Hansen, C. J., Christlieb, N., et al. 2015, ApJ, 807, 173

Hargis, J. R., & Rhode, K. L. 2014, ApJ, 796, 62

Harris, W. E. 1996, AJ, 112, 1487

—. 2009a, ApJ, 699, 254

—. 2009b, ApJ, 703, 939

Harris, W. E., Ciccone, S. M., Eadie, G. M., et al. 2017, ApJ, 835, 101

Harris, W. E., Whitmore, B. C., Karakla, D., et al. 2006, ApJ, 636, 90

Hartke, J., Arnaboldi, M., Gerhard, O., et al. 2018, A&A, 616, A123

Helmi, A. 2020, arXiv e-prints, arXiv:2002.04340

Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, Nature, 563, 85

Hoffman, M. D., & Gelman, A. 2011, arXiv e-prints, arXiv:1111.4246

Høg, E., Fabricius, C., Makarov, V. V., et al. 2000a, A&A, 357, 367

—. 2000b, A&A, 355, L27

Hogg, D. W., Bovy, J., & Lang, D. 2010, ArXiv e-prints, arXiv:1008.4686

Hollek, J. K., Frebel, A., Placco, V. M., et al. 2015, ApJ, 814, 121

Honda, S., Aoki, W., Kajino, T., et al. 2004, ApJ, 607, 474

Howes, L. M., Casey, A. R., Asplund, M., et al. 2015, Nature, 527, 484

Howes, L. M., Asplund, M., Keller, S. C., et al. 2016, MNRAS, 460, 884

Hubble, E. P. 1936,

Impey, C., Bothun, G., & Malin, D. 1988, ApJ, 330, 634

Ishigaki, M., Chiba, M., & Aoki, W. 2010, PASJ, 62, 143

Ishigaki, M. N., Aoki, W., Arimoto, N., & Okamoto, S. 2014, A&A, 562, A146

Ivanov, V. D., Rieke, M. J., Engelbracht, C. W., et al. 2004, ApJS, 151, 387

Ivans, I. I., Simmerer, J., Sneden, C., et al. 2006, ApJ, 645, 613

Ivans, I. I., Sneden, C., James, C. R., et al. 2003, ApJ, 592, 906

Jacobson, H. R., Keller, S., Frebel, A., et al. 2015, ApJ, 807, 171

Janz, J., Norris, M. A., Forbes, D. A., et al. 2016, MNRAS, 456, 617

Jeřábková, T., Kroupa, P., Dabringhausen, J., Hilker, M., & Bekki, K. 2017, A&A, 608, A53

Ji, A. P., Frebel, A., Chiti, A., & Simon, J. D. 2016, Nature, 531, 610

Johnson, J. A. 2002, ApJS, 139, 219

Johnson, J. A., & Bolte, M. 2004, ApJ, 605, 462

Jones, L. 1999, PhD thesis. Univ. North Carol.

Jonsell, K., Barklem, P. S., Gustafsson, B., et al. 2006, A&A, 451, 651

Jonsell, K., Edvardsson, B., Gustafsson, B., et al. 2005, A&A, 440, 321

Jordán, A., Blakeslee, J. P., Peng, E. W., et al. 2004, The Astrophysical Journal Supplement Series, 154, 509

Jordán, A., Peng, E. W., Blakeslee, J. P., et al. 2009, The Astrophysical Journal Supplement Series, 180, 54

Kartha, S. S., Forbes, D. A., Alabi, A. B., et al. 2016, MNRAS, 458, 105

Kausch, W., Noll, S., Smette, A., et al. 2015, A&A, 576, A78

Kirby, E. N., Guhathakurta, P., Bolte, M., Sneden, C., & Geha, M. C. 2009, ApJ, 705, 328

Kleinmann, S. G., & Hall, D. N. B. 1986, ApJS, 62, 501

Koch, A., McWilliam, A., Grebel, E. K., Zucker, D. B., & Belokurov, V. 2008, ApJ, 688, L13

Koch, A., McWilliam, A., Preston, G. W., & Thompson, I. B. 2015, VizieR Online Data Catalog, J/A+A/587/A124

Koda, J., Yagi, M., Yamanoi, H., & Komiyama, Y. 2015, ApJ, 807, L2

Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, A&A, 501, 1269

Kormendy, J., Fisher, D. B., Cornell, M. E., & Bender, R. 2009, ApJS, 182, 216

Kroupa, P. 2001, MNRAS, 322, 231

Kroupa, P., & Gilmore, G. F. 1994, MNRAS, 269, 655

Kroupa, P., Weidner, C., Pflamm-Altenburg, J., et al. 2013, The Stellar and Sub-Stellar Initial Mass Function of Simple and Composite Populations, ed. T. D. Oswalt & G. Gilmore, Vol. 5, 115

Kruijssen, J. M. D., Pfeffer, J. L., Reina-Campos, M., Crain, R. A., & Bastian, N. 2019, MNRAS, 486, 3180

Krumholz, M. R. 2014, Phys. Rep., 539, 49

Kundu, A., & Whitmore, B. C. 2001a, AJ, 121, 2950

—. 2001b, AJ, 122, 1251

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

—. 1993,

La Barbera, F., Ferreras, I., Vazdekis, A., et al. 2013, MNRAS, 433, 3017

Lai, D. K., Bolte, M., Johnson, J. A., et al. 2008, ApJ, 681, 1524

Lançon, A., & Wood, P. R. 2000, A&AS, 146, 217

Lanz, T., & Hubeny, I. 2003, ApJS, 146, 417

Larsen, S. S., Brodie, J. P., Huchra, J. P., Forbes, D. A., & Grillmair, C. J. 2001, AJ, 121, 2974

Le Borgne, J.-F., Bruzual, G., Pelló, R., et al. 2003, A&A, 402, 433

Leaman, R., VandenBerg, D. A., & Mendel, J. T. 2013, MNRAS, 436, 122

Lee, S.-Y., Chung, C., & Yoon, S.-J. 2019, The Astrophysical Journal Supplement Series, 240, 2

Li, H., & Gnedin, O. Y. 2014, ApJ, 796, 10

Li, H.-N., Zhao, G., Christlieb, N., et al. 2015, ApJ, 798, 110

Lim, S., Peng, E. W., Côté, P., et al. 2018, ApJ, 862, 82

Lind, K., Bergemann, M., & Asplund, M. 2012, MNRAS, 427, 50

Liu, C., Peng, E. W., Jordán, A., et al. 2011, ApJ, 728, 116

Liu, Y., Peng, E. W., Blakeslee, J., et al. 2016, ApJ, 818, 179

Longobardi, A., Arnaboldi, M., Gerhard, O., & Mihos, J. C. 2015, A&A, 579, L3

Longobardi, A., Arnaboldi, M., Gerhard, O., Pulsoni, C., & Söldner-Rembold, I. 2018a, A&A, 620, A111

Longobardi, A., Peng, E. W., Côté, P., et al. 2018b, ApJ, 864, 36

Lord, S. D. 1992, NASA Technical Memorandum 103957

Ma, C.-P., Greene, J. E., McConnell, N., et al. 2014, ApJ, 795, 158

Ma, X., Grudić, M. Y., Quataert, E., et al. 2020, MNRAS, 493, 4315

Magic, Z., Collet, R., Hayek, W., & Asplund, M. 2013, A&A, 560, A8

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017a, AJ, 154, 94

—. 2017b, AJ, 154, 94

Mandelker, N., van Dokkum, P. G., Brodie, J. P., van den Bosch, F. C., & Ceverino, D. 2018, ApJ, 861, 148

Mann, A. W., Feiden, G. A., Gaidos, E., Boyajian, T., & von Braun, K. 2015, ApJ, 804, 64

Marks, M., Kroupa, P., Dabringhausen, J., & Pawlowski, M. S. 2012, MNRAS, 422, 2246

Mármol-Queraltó, E., Cardiel, N., Cenarro, A. J., et al. 2008, A&A, 489, 885

Martín-Navarro, I. 2016, MNRAS, 456, L104

Martín-Navarro, I., Vazdekis, A., La Barbera, F., et al. 2015, ApJ, 806, L31

Martín-Navarro, I., Romanowsky, A. J., Brodie, J. P., et al. 2019, MNRAS, 484, 3425

Martini, P., Dicken, D., & Storchi-Bergmann, T. 2013, ApJ, 766, 121

Martins, L. P., & Coelho, P. 2007, MNRAS, 381, 1329

Martins, L. P., González Delgado, R. M., Leitherer, C., Cerviño, M., & Hauschildt, P. 2005, MNRAS, 358, 49

Masseron, T., Johnson, J. A., Lucatello, S., et al. 2012, ApJ, 751, 14

McConnell, N. J., Lu, J. R., & Mann, A. W. 2016, ApJ, 821, 39

McLaughlin, D. E. 1999, AJ, 117, 2398

McWilliam, A., Preston, G. W., Sneden, C., & Searle, L. 1995, AJ, 109, 2757

Meneses-Goytia, S., Peletier, R. F., Trager, S. C., et al. 2015, A&A, 582, A96

Mieske, S., Frank, M. J., Baumgardt, H., et al. 2013, A&A, 558, A14

Mihos, J. C., Harding, P., Feldmeier, J. J., et al. 2017, ApJ, 834, 16

Milone, A. D. C., Sansom, A. E., & Sánchez-Blázquez, P. 2011, MNRAS, 414, 1227

Monachesi, A., Trager, S. C., Lauer, T. R., et al. 2012, ApJ, 745, 97

Müller, O., Jerjen, H., & Binggeli, B. 2018, A&A, 615, A105

Munari, U., Sordo, R., Castelli, F., & Zwitter, T. 2005, A&A, 442, 1127

Muratov, A. L., & Gnedin, O. Y. 2010, ApJ, 718, 1266

Murphy, J. D., Gebhardt, K., & Adams, J. J. 2011, ApJ, 729, 129

Naab, T., Johansson, P. H., & Ostriker, J. P. 2009, ApJ, 699, L178

Naab, T., & Ostriker, J. P. 2017, ARA&A, 55, 59

Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16

Newman, A. B., Smith, R. J., Conroy, C., Villaume, A., & van Dokkum, P. 2017, ApJ, 845, 157

Newton, E. R., Charbonneau, D., Irwin, J., et al. 2014, AJ, 147, 20

Norris, J. E., Ryan, S. G., Beers, T. C., & Deliyannis, C. P. 1997, ApJ, 485, 370

Oke, J. B., Cohen, J. G., Carr, M., et al. 1995, PASP, 107, 375

Oldham, L. J., & Auger, M. W. 2016a, MNRAS, 455, 820

—. 2016b, MNRAS, 455, 820

—. 2016c, MNRAS, 457, 421

Oser, L., Naab, T., Ostriker, J. P., & Johansson, P. H. 2012, ApJ, 744, 63

Oser, L., Ostriker, J. P., Naab, T., Johansson, P. H., & Burkert, A. 2010, ApJ, 725, 2312

Pandya, V., Mulchaey, J., & Greene, J. E. 2016, ApJ, 819, 162

Pastorello, N., Forbes, D. A., Usher, C., et al. 2015, MNRAS, 451, 2625

Peacock, M. B., Zepf, S. E., Kundu, A., & Chael, J. 2017, MNRAS, 464, 713

Peng, E. W., & Lim, S. 2016, ApJ, 822, L31

Peng, E. W., Jordán, A., Côté, P., et al. 2006, ApJ, 639, 95

—. 2008, ApJ, 681, 197

Percival, S. M., & Salaris, M. 2009, ApJ, 703, 1123

Pfeffer, J., Kruijssen, J. M. D., Crain, R. A., & Bastian, N. 2018, MNRAS, 475, 4309

Pickles, A. J. 1985, ApJS, 59, 33

—. 1998, PASP, 110, 863

Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, MNRAS, 475, 648

Placco, V. M., Beers, T. C., Ivans, I. I., et al. 2015, ApJ, 812, 109

Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, ARA&A, 48, 431

Powalka, M., Puzia, T. H., Lançon, A., et al. 2016, ApJ, 829, L5

Preston, E. F., Martins, J. S. S., & Rundle, J. B. 2001, arXiv e-prints, cond

Preston, G. W., & Sneden, C. 2000, AJ, 120, 1014

Preston, G. W., Thompson, I. B., Sneden, C., Stachowski, G., & Shectman, S. A. 2006, AJ, 132, 1714

Pritzl, B. J., Venn, K. A., & Irwin, M. 2005, AJ, 130, 2140

Prole, D. J., Hilker, M., van der Burg, R. F. J., et al. 2019, MNRAS, 484, 4865

Prugniel, P., & Soubiran, C. 2001, A&A, 369, 1048

Prugniel, P., Vauglin, I., & Koleva, M. 2011, A&A, 531, A165

Prusti, T., de Bruijne, J. H. J., Brown, A. G. A., et al. 2016, A&A, 595, A1

Rajpurohit, A. S., Reyle, C., Allard, F., et al. 2016, ArXiv e-prints, arXiv:1609.07062

Rajpurohit, A. S., Reylé, C., Allard, F., et al. 2014, A&A, 564, A90

Rayner, J. T., Cushing, M. C., & Vacca, W. D. 2009, ApJS, 185, 289

Rayner, J. T., Toomey, D. W., Onaka, P. M., et al. 2003, PASP, 115, 362

Renzini, A. 2006, ARA&A, 44, 141

Rhode, K. L., Zepf, S. E., & Santos, M. R. 2005, ApJ, 630, L21

Rich, R. M., Corsi, C. E., Cacciari, C., et al. 2005, AJ, 129, 2670

Robertson, B., Bullock, J. S., Font, A. S., Johnston, K. V., & Hernquist, L. 2005, ApJ, 632, 872

Röck, B., Vazdekis, A., Ricciardelli, E., et al. 2016, A&A, 589, A73

Roederer, I. U., Preston, G. W., Thompson, I. B., et al. 2014, AJ, 147, 136

Roederer, I. U., Sneden, C., Thompson, I. B., Preston, G. W., & Shectman, S. A. 2010, ApJ, 711, 573

Roederer, I. U., Frebel, A., Shetrone, M. D., et al. 2008, ApJ, 679, 1549

Roediger, J. C., Courteau, S., Graves, G., & Schiavon, R. P. 2014, The Astrophysical Journal Supplement Series, 210, 10

Romanowsky, A. J., Strader, J., Brodie, J. P., et al. 2012, ApJ, 748, 29

Rood, R. T. 1973, ApJ, 184, 815

Ryan, S. G., Norris, J. E., & Bessell, M. S. 1991, AJ, 102, 303

Salpeter, E. E. 1955, ApJ, 121, 161

Salvatier J., Wiecki T.V., F. C. 2016, PearJ Computer Science

Sánchez-Blázquez, P., Ocvirk, P., Gibson, B. K., Pérez, I., & Peletier, R. F. 2011, MNRAS, 415, 709

Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006a, MNRAS, 371, 703

Sánchez-Blázquez, P., Peletier, R. F., Jiménez- Vicente, J., et al. 2006b, MNRAS, 371, 703

Sandage, A., & Binggeli, B. 1984, AJ, 89, 919

Sandoval, M. A., Vo, R. P., Romanowsky, A. J., et al. 2015, ApJ, 808, L32

Sarzi, M., Spiniello, C., La Barbera, F., Krajnović, D., & van den Bosch, R. 2018, MNRAS, 478, 4084

Schiavon, R. P. 2007, ApJS, 171, 146

Schiavon, R. P., Rose, J. A., Courteau, S., & MacArthur, L. A. 2004, ApJ, 608, L33

—. 2005, The Astrophysical Journal Supplement Series, 160, 163

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Searle, L., & Zinn, R. 1978, ApJ, 225, 357

Sharma, K., Prugniel, P., & Singh, H. P. 2016, A&A, 585, A64

Sharon, C., Hillenbrand, L., Fischer, W., & Edwards, S. 2010, AJ, 139, 646

Shetrone, M., Venn, K. A., Tolstoy, E., et al. 2003, AJ, 125, 684

Shetrone, M. D., Côté, P., & Sargent, W. L. W. 2001, ApJ, 548, 592

Silva, D. R., & Cornell, M. E. 1992, ApJS, 81, 865

Simonian, G. V., & Martini, P. 2016, ArXiv e-prints, arXiv:1603.09345

Siqueira Mello, C., Hill, V., Barbuy, B., et al. 2014, A&A, 565, A93

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Skúladóttir, Á., Tolstoy, E., Salvadori, S., et al. 2015, A&A, 574, A129

Smith, R. J. 2014, MNRAS, 443, L69

Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51

Spiniello, C., Koopmans, L. V. E., Trager, S. C., Czoske, O., & Treu, T. 2011, MNRAS, 417, 3000

Spiniello, C., Trager, S. C., Koopmans, L. V. E., & Chen, Y. P. 2012, ApJ, 753, L32

Spinrad, H. 1962, ApJ, 135, 715

Strader, J., Brodie, J. P., Cenarro, A. J., Beasley, M. A., & Forbes, D. A. 2005, AJ, 130, 1315

Strader, J., Brodie, J. P., & Forbes, D. A. 2004, AJ, 127, 3431

Strader, J., Caldwell, N., & Seth, A. C. 2011a, AJ, 142, 8

Strader, J., Romanowsky, A. J., Brodie, J. P., et al. 2011b, The Astrophysical Journal Supplement Series, 197, 33

Strom, S. E., Forte, J. C., Harris, W. E., et al. 1981, ApJ, 245, 416

Sybilska, A., Kuntschner, H., van de Ven, G., et al. 2018, MNRAS, 476, 4501

Taylor, P., & Kobayashi, C. 2017, MNRAS, 471, 3856

Thomas, D., Maraston, C., & Bender, R. 2003, MNRAS, 339, 897

Thomas, D., Maraston, C., Bender, R., & Mendes de Oliveira, C. 2005, ApJ, 621, 673

Toloba, E., Lim, S., Peng, E., et al. 2018, ApJ, 856, L31

Toomre, A. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & D. C. Larson, Richard B. Gehret, 401

Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, AJ, 120, 165

Tremmel, M., Wright, A., Brooks, A., et al. 2020, in American Astronomical Society Meeting Abstracts, Vol. 52, American Astronomical Society Meeting Abstracts, 316.02

Treu, T., Auger, M. W., Koopmans, L. V. E., et al. 2010, ApJ, 709, 1195

Tripicco, M. J., & Bell, R. A. 1995, AJ, 110, 3035

Usher, C., Forbes, D. A., Spitler, L. R., et al. 2013, MNRAS, 436, 1172

Usher, C., Forbes, D. A., Brodie, J. P., et al. 2012, MNRAS, 426, 1475

Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2003, PASP, 115, 389

Valdes, F., Gupta, R., Rose, J. A., Singh, H. P., & Bell, D. J. 2004, ApJS, 152, 251

van den Bergh, S. 1975, Annual Review of Astronomy and Astrophysics, 13, 217

—. 1982, PASP, 94, 459

van den Bosch, R. C. E., & de Zeeuw, P. T. 2010, MNRAS, 401, 1770

van der Burg, R. F. J., Muzzin, A., & Hoekstra, H. 2016, A&A, 590, A20

van Dokkum, P., Conroy, C., Villaume, A., Brodie, J., & Romanowsky, A. J. 2017a, ApJ, 841, 68

van Dokkum, P., Abraham, R., Brodie, J., et al. 2016, ApJ, 828, L6

van Dokkum, P., Abraham, R., Romanowsky, A. J., et al. 2017b, ApJ, 844, L11

van Dokkum, P., Cohen, Y., Danieli, S., et al. 2018, ApJ, 856, L30

van Dokkum, P., Wasserman, A., Danieli, S., et al. 2019, ApJ, 880, 91

van Dokkum, P. G., & Conroy, C. 2010, Nature, 468, 940

—. 2012, ApJ, 760, 70

van Dokkum, P. G., Romanowsky, A. J., Abraham, R., et al. 2015, ApJ, 804, L26

Vazdekis, A., Cenarro, A. J., Gorgas, J., Cardiel, N., & Peletier, R. F. 2003, MNRAS, 340, 1317

Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, MNRAS, 463, 3409

Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., et al. 2010, MNRAS, 404, 1639

Villaume, A., Brodie, J., Conroy, C., Romanowsky, A. J., & van Dokkum, P. 2017a, ApJ, 850, L14

Villaume, A., Conroy, C., Johnson, B., et al. 2017b, ApJS, 230, 23

Villaume, A., Conroy, C., & Johnson, B. D. 2015, ApJ, 806, 82

Villaume, A., Romanowsky, A. J., Brodie, J., & Strader, J. 2019, ApJ, 879, 45

Wetzel, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, ApJ, 827, L23

Wing, R. F., & Ford, Jr., W. K. 1969, PASP, 81, 527

Woosley, S. E., Heger, A., & Weaver, T. A. 2002, Reviews of Modern Physics, 74, 1015

Worthey, G., Faber, S. M., Gonzalez, J. J., & Burstein, D. 1994, The Astrophysical Journal Supplement Series, 94, 687

Wu, Y., Singh, H. P., Prugniel, P., Gupta, R., & Koleva, M. 2011, A&A, 525, A71

Yagi, M., Koda, J., Komiyama, Y., & Yamanoi, H. 2016, ApJS, 225, 11

Yoon, S.-J., Yi, S. K., & Lee, Y.-W. 2006, Science, 311, 1129

Zacs, L., Nissen, P. E., & Schuster, W. J. 1998, A&A, 337, 216

Zepf, S. E., & Ashman, K. M. 1993, MNRAS, 264, 611

Zhang, H.-X., Peng, E. W., Côté, P., et al. 2015, ApJ, 802, 30

Zhang, L., Ishigaki, M., Aoki, W., Zhao, G., & Chiba, M. 2009, ApJ, 706, 1095

Zieleniewski, S., Houghton, R. C. W., Thatte, N., Davies, R. L., & Vaughan, S. P. 2017, MNRAS, 465, 192

Zinn, R. 1985, ApJ, 293, 424

Zolotov, A., Willman, B., Brooks, A. M., et al. 2010, ApJ, 721, 738

Zonoozi, A. H., Haghi, H., & Kroupa, P. 2016, ApJ, 826, 89

Appendix A

Derivation of Likelihood Function for True Distances

While the true 3D distance is given by $|\vec{r}| = \sqrt{x^2 + y^2 + z^2}$, what we actually observe is $r_\perp = \sqrt{x^2 + y^2}$. The xy coordinates can be written in terms of $r_\perp$,

$$
\begin{aligned}
x =& r_\perp cos\theta \\
y =& r_\perp sin\theta, \text{ and then,} \\
z =& r_\parallel \quad .
\end{aligned}
$$

Our model defines the distribution in x, y, and z as Gaussian, but it is useful to, instead, reparameterize in terms of $r_\perp$ and $r_\parallel$. In order to maintain the same density through this change of variables, we need to take the Jacobian of the transformation into account. Specifically,

$$
\begin{aligned}
\left|p(x,\, y,\, z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\right| &= \left|p(r_\perp,\, \theta,\, r_\parallel)\,\mathrm{d}r_\perp\,\mathrm{d}\theta\,\mathrm{d}r_\parallel\right| \qquad \text{(A.1)}\\
p(r_\perp,\, \theta,\, r_\parallel) &= |J|\,p(x,\, y,\, z) \qquad \text{(A.2)}
\end{aligned}
$$

where $p(x,\, y,\, z)$ is Gaussian and $|J|$ is the absolute value of the determinant of the Jacobian matrix

$$
\begin{aligned}
|J| &= \begin{vmatrix} \frac{\mathrm{d}x}{\mathrm{d}r_\perp} & \frac{\mathrm{d}y}{\mathrm{d}r_\perp} & \frac{\mathrm{d}z}{\mathrm{d}r_\perp} \\ \frac{\mathrm{d}x}{\mathrm{d}\theta} & \frac{\mathrm{d}y}{\mathrm{d}\theta} & \frac{\mathrm{d}z}{\mathrm{d}\theta} \\ \frac{\mathrm{d}x}{\mathrm{d}r_\parallel} & \frac{\mathrm{d}y}{\mathrm{d}r_\parallel} & \frac{\mathrm{d}z}{\mathrm{d}r_\parallel} \end{vmatrix} \qquad \text{(A.3)}\\
&= \begin{vmatrix} \cos\theta & \sin\theta & 0 \\ -r_\perp \sin\theta & r_\perp \cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} \qquad \text{(A.4)}\\
&= \left|r_\perp\,(\sin^2\theta + \cos^2\theta)\right| = r_\perp \quad . \qquad \text{(A.5)}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
p(r_\perp,\, \theta,\, r_\parallel) &= \frac{r_\perp}{(2\,\pi\,R^2)^{3/2}} \exp\left(-\frac{x^2+y^2+z^2}{2\,R^2}\right) \qquad \text{(A.6)}\\
&= \frac{r_\perp}{(2\,\pi\,R^2)^{3/2}} \exp\left(-\frac{r_\perp^2+r_\parallel^2}{2\,R^2}\right) \quad . \qquad \text{(A.7)}
\end{aligned}
$$

Finally, under our assumption of isotropy, we can marginalize over the position angle θ

to find

$$
\begin{aligned}
p(r_\perp,\, r_\parallel) &= \int p(r_\perp,\, \theta,\, r_\parallel)\,\mathrm{d}\theta && \text{(A.8)} \\
&= \frac{r_\perp}{\sqrt{2\,\pi}\,R^3}\exp\left(-\frac{r_\perp^2 + r_\parallel^2}{2\,R^2}\right) && \text{(A.9)} \\
&= \mathrm{Rayleigh}(r_\perp;\, R)\,\mathrm{Normal}(r_\parallel;\, 0,\, R) \quad . && \text{(A.10)}
\end{aligned}
$$

Appendix B

Individual References for JINABase data

Ref	[Fe/H]	[Mg/Fe]	[Si/Fe]	[Ca/Fe]
Allen et al. (2012)	x	x	–	–
Aoki et al. (2002b)	x	x	–	–
Aoki et al. (2002a)	x	x	x	x
Aoki et al. (2005)	x	x	x	x
Aoki et al. (2007)	x	x	–	–
Aoki et al. (2008)	x	x	–	–
Aoki et al. (2012)	x	x	–	–
Aoki et al. (2013)	x	x	–	–
Aoki et al. (2014)	x	x	–	–
Barbuy et al. (2005)	x	x	–	–

Barklem et al. (2005)	x	x	–	–
Bensby et al. (2011)	x	x	x	x
Carretta et al. (2002)	x	x	x	x
Cayrel et al. (2004)	x	x	x	x
Cohen et al. (2003b)	x	x	x	x
Cohen et al. (2004)	x	x	x	x
Cohen et al. (2006)	x	x	x	x
Cohen et al. (2013)	x	x	x	x
Cowan et al. (2002)	x	x	x	x
Cui et al. (2013)	x	x	x	x
Fulbright (2000)	x	x	x	x
Hansen et al. (2015)	x	x	–	–
Hollek et al. (2015)	x	x	–	–
Honda et al. (2004)	x	x	x	x
Howes et al. (2015)	x	x	x	x
Howes et al. (2016)	x	x	x	x
Ishigaki et al. (2010)	x	x	x	x
Ivans et al. (2003)	x	x	x	x
Ivans et al. (2006)	x	x	x	x
Jacobson et al. (2015)	x	x	x	x
Johnson (2002)	x	x	x	x
Johnson & Bolte (2004)	x	x	–	–

Jonsell et al. (2005)	x	x	x	x
Jonsell et al. (2006)	x	x	–	–
Koch et al. (2015)	x	x	x	x
Lai et al. (2008)	x	x	x	x
Li et al. (2015)	x	x	x	x
Masseron et al. (2012)	x	x	–	–
McWilliam et al. (1995)	x	x	x	x
Norris et al. (1997)	x	x	x	x
Placco et al. (2015)	x	x	x	x
Preston & Sneden (2000)	x	x	–	–
Preston et al. (2001)	x	x	–	–
Preston et al. (2006)	x	x	x	x
Roederer et al. (2008)	x	x	–	–
Roederer et al. (2010)	x	x	x	x
Roederer et al. (2014)	x	x	x	x
Ryan et al. (1991)	x	x	x	x
Siqueira Mello et al. (2014)	x	x	x	x
Zacs et al. (1998)	x	x	–	–
Zhang et al. (2009)	x	x	x	x

Table B.1: Individual references for JINAbase compilation of Milky Way field stars used in this paper.

Ref	[Fe/H]	[Mg/Fe]	[Si/Fe]	[Ca/Fe]
Cohen & Huang (2009)	x	x	x	x
Feltzing et al. (2009)	x	x	–	–
François et al. (2016)	x	x	–	–
Frebel et al. (2010)	x	x	x	x
Geisler et al. (2005)	x	x	x	x
Gilmore et al. (2013)	x	x	x	x
Ishigaki et al. (2014)	x	x	–	–
Ji et al. (2016)	x	x	x	x
Koch et al. (2008)	x	x	x	x
Shetrone et al. (2001)	x	x	x	x
Shetrone et al. (2003)	x	x	x	x
Skúladóttir et al. (2015)	x	x	x	x

Table B.2: Individual references for JINAbase compilation of dwarf galaxy field stars used in this paper.