UNIVERSITY OF CALIFORNIA
SANTA CRUZ

MEASURING ASTROPHYSICAL PARAMETERS FROM RESOLVED STELLAR POPULATIONS IN THE MILKY WAY USING APPLIED STATISTICS

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

Doctor of Philosophy

in

ASTRONOMY AND ASTROPHYSICS

by

Kevin A. McKinnon

September 2023

The Dissertation of Kevin A. McKinnon is approved:

Constance M. Rockosi, Chair

Puragra Guhathakurta

Melissa K. Ness

Brant E. Robertson

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Measuring Astrophysical Parameters from Resolved Stellar Populations in the Milky Way using Applied Statistics

by

Kevin A. McKinnon

Stars, as key building blocks of galaxies, retain information about the conditions in which they formed and can therefore be used to trace galaxy formation and evolution. Using data from the HALO7D survey, we measure chemical abundances from stellar spectra of main sequence turn-off stars in the Milky Way (MW) stellar halo. From these abundances, in combination with previously-measured velocities, we show that the chemodynamical distributions of stars along four individual lines-of-sight (LOS) are statistically different from one another, in agreement with a growing body of evidence that suggests the the MW stellar halo is not as well-mixed as often assumed.

With the goal of improving our understanding of the MW merger's history by expanding precise chemodynamics measurements to additional LOS in the Galaxy, we develop a technique for measuring precise proper motions (PMs) of stars in sparse fields by combining archival *Hubble Space Telescope* (HST) images with *Gaia* data. The resulting PMs are a median of 2.6 times more precise than *Gaia* alone in sparse *HST* images of COSMOS, and we recover PMs for the $\sim 25\%$ of sources that are too faint for *Gaia* to constrain. This technique also enables us to simulate future missions, such as the *Roman Space Telescope*. With these simulated observations, we design an

observation strategy that significantly improves parallax precision at no cost to PM precision.

Chemical tagging experiments are hindered by unexpected, non-stellar signatures in spectra, some of which can originate in the interstellar medium (ISM). To increase the scientific potential of the APOGEE spectrograph, we present a detailed accounting of light in APOGEE stellar spectra using a data-driven model of red clump stars. These near-infrared, H-band spectra are well-described by this model, though their residuals reveal a wealth of information about the intervening gas and dust in the ISM. We characterize the non-stellar light to measure as many as 84 Diffuse Interstellar Bands in the APOGEE wavelength range, ~ 74 of which were previously unknown, and show that these ISM-based features are likely impacting stellar chemical abundance measurements.

Acknowledgments

I've been very fortunate for the teachers, of all forms, in my life that have shaped me into the person I am today. This thesis is the culmination of not just my work, but of the lessons they imparted in me.

I couldn't ask for better first teachers than my parents, Teresa and Grant, even if I wasn't always the easiest student. You fight for your kids to have every opportunity to follow their dreams, and I don't even want to guess what my life would look like without your love and support. From reading to us every night, to teaching me algebra at the dinner/homework table, you always showed us the importance of getting an education, and I like to think I got the message loud and clear. To Ben and Elise, thank you for always reminding me that, "All work and no play makes Kevin a dull boy." You've fought an uphill battle to keep me from going too far off the "nerd" deep end, and I'll always be grateful for that. Astronomers are often asked when they first fell in love with space. For me, the answer is all the nights spent star gazing with my family during the summers at Lake Cowichan.

To Stephen Lane, you truly awakened a passion for science in me. When I was a kid, I wanted to have magical powers more than anything in the world, but I eventually grew up and learned that the magic I wanted didn't exist. In your physics classroom, I became entranced with a new type of magic, where the equations you taught could be used to explain the whole universe. By the time I graduated high school, you had nurtured my budding interest into a concrete goal of pursuing physics at university. Your wit and intelligence, your devotion for teaching, and your love of

physics and space make up the soil that I sprouted from.

To my PhD advisors, Connie and Raja, I really appreciate how you've guided me over the years to help me to see the bigger picture, especially when I'm lost in the trees. I didn't have much of a specific research goal in mind when I first came to Santa Cruz, but your guidance helped me take ownership in my first project. While it may have occasionally prevented immediate research progress, I'm extremely grateful that you supported my decision to pursue a Designated Emphasis in Statistics, which I believe will be the trademark of my future career. The late nights we spent observing and last minute frantic proposal writing were tough, but it was always exciting, and I thank you for the foundations you've built for me in my academic career.

To my mentors – Emily, Melissa, and Stéphane – I wouldn't be here without you. Emily, independent of Raja's family tree of students, you truly felt like my big sister in research. You always made sure to look out for me, you were more protective of my time than I was, and you taught me the ropes of being a graduate student and researcher. Melissa, our professional relationship began at a key moment when I was really doubting my place in academia. You treated me as an equal with exciting ideas to offer, and that was invaluable to me as a young scientist; it was the first time I was able to see my future in the field, and it helped me persevere in finishing my PhD at a time when it seemed insurmountable. Stéphane, you got me into this whole mess, and for that I am forever changed. When I took my first astronomy course at Queen's on a whim, I couldn't have imagined it would lead to this, but I think you knew. If you had not planted the idea of summer research in my head – and later graduate school in

general – I don't know where I would be today. You always push for your students in ways I aspire to, and I look forward to building a professional relationship together in the future.

A PhD is a long, arduous journey, and I'm not sure that I would have made it without the friends I made along the way. To my Santa Cruz family – Amanda, Brenna, Tyler, Maggie, Jeremy, Diana, and Jamie – you are incredible human beings, and I'm so fortunate that we could spend these past years together. From that first visit to UCSC, I knew I had found some of the coolest, brightest, and kindest people in the cosmos. Thank you for all the late night discussions over wine, the movie nights, the book club meetings, and the fun adventures that define my happiest memories in Santa Cruz. These days, we seem to have spread ourselves around the globe, but I know we will find excuses to meet up as often as we can.

To Avesta, there is no single person I could thank more than you for making sure that I made it all the way to the end, and your brilliant mind inspires me every day. Thank you for being my best friend, my motivation, and my sounding board. Thank you for talking through best ways to display figures, for making sure I took breaks to have fun, and for supporting me in the bleaker days of my PhD. Every time I get frustrated when preparing for a talk or writing papers (this thesis, for example), I remember to live up to your quote of "Science isn't finished until it's been communicated." I know that we are both going to miss our life in Santa Cruz, but I can't wait for our future adventures together.

Published Material

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

Chapter 2 was published in the literature as McKinnon et al. (2023). I was responsible for measuring stellar abundances, performing the chemodynamic analyses, writing the text, and creating the figures. I wish to thank all my coauthors, with particular praise to Emily C. Cunningham, Connie M. Rockosi, and Raja Guhathakurta for their feedback and (scientific and personal) support throughout the years that this project occupied.

Chapter 3 will soon be submitted to *The Astrophysical Journal*. I was responsible for creating the *BP3M* code, performing the analysis, writing the text, and creating the figures. I wish to thank all my coauthors, with specific thanks to Andrés del Pino, Mattia Libralato, Roeland P. van der Marel, and the rest of the HSTPROMO (High-resolution Space Telescope PROper MOtion) Collaboration for their expertise and assistance in my understanding of measuring precise proper motions of stars.

Chapter 4 is currently under review at *The Astrophysical Journal* and is expected to be published later this year. I was responsible for modelling the data, characterizing the residuals, performing the analysis, writing the text, and creating the figures. I wish to thank my coauthors of Melissa Ness, Connie Rockosi, and Raja Guhathakurta for their support and feedback.

Scientific Acknowledgments

Throughout my PhD, I was supported by the National Science Foundation (NSF) with Grants AST-1616540 and AST-2206328, as well as by NASA through grants associated with HST Archival Program AR-16625 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555.

I thank the various anonymous journal referees for their comments that improved the clarity of the published work presented in this thesis, as well as the Center for Computational Astrophysics Astrophysical Data Group for useful discussions. I also thank Avesta Rastan for helpful conversations that improved the presentation my scientific figures, Brian DiGiorgio for insight into `python` matrix math that led to significant reductions in computation time in Chapter 2, and Melodie Kao for code that helped implement the motion-from-parallax calculations in Chapter 3. Additionally, I owe a lot of thanks to the individuals and teams of researchers whose decades of work to characterize the optics of *HST* that have made my work possible.

Finally, I 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.

Software: `Astropy` (Astropy Collaboration et al. 2013, 2018, 2022), Bovy's `APOGEE` Code (`https://github.com/jobovy/apogee`), `corner` (Foreman-Mackey 2016), `dustmaps` (Green 2018), `emcee` (Foreman-Mackey et al. 2013), `GaiaHub` (del Pino et al.

2022), `IPython` (Pérez & Granger 2007), `jupyter` (Kluyver et al. 2016), `matplotlib` (Hunter 2007), `numpy` (Harris et al. 2020), `pandas` (Wes McKinney 2010; pandas development team 2020), Price-Jones' `SPECTRAL SPACE PYTHON` Code (`https://github.com/npricejones/spectralspace`),`scipy` (Virtanen et al. 2020), `scikit-learn` (Pedregosa et al. 2011), `spec2d` (Cooper et al. 2012)

For the teachers of all forms in my life,

Chapter 1

Introduction

In a Lambda Cold Dark Matter (λCDM) cosmology, galaxies – which live inside of dark matter halos – merge together over time to produce successively larger galaxies until present day. Stars, which reside inside of these galaxies and halos, are born in gas clouds that carry chemical and dynamical information about the conditions of their host potential well at the time the star formed; in this way, measuring stellar properties allow us to look back in time to understand the formation and evolution of galaxies. With current technology, the most complete information about resolved stellar populations is only available in the Local Group – encompassing the Milky Way (MW), Andromeda, Triangulum, the Large and Small Magellanic Clouds, and a collection of other nearby galaxies and clusters – making their study particularly valuable to constraining the connection between stars, galaxies, and cosmology.

Broadly, there are two main types of direct data from stars that this work focuses on: astrometry and spectroscopy. Stellar spectroscopy encodes information

about a star's evolutionary state (i.e. its surface temperature and pressure), the relative abundance of different chemical elements, and the velocity of that star in the direction of the line of sight (LOS). Astrometry at an instant in time allows us to measure the 2D position of a star, while multiple position measurements in time yields 2D motion on the sky as well as parallax constraints. From these two techniques, we are able to extract seven dimensional information (i.e. 3D positions, 3D velocities, and one or more dimensions of chemical abundance).

Stars born in different environments – different galaxies or even gas clouds within a galaxy – have different kinematic and chemical signatures due to changing host masses, potential well distributions, and star formation histories (e.g., Eggen et al. 1962; Searle & Zinn 1978; Bullock & Johnston 2005; Helmi 2008). While individual 7D measurements of stars may be useful for stellar physics, chemodynamical distributions from populations of stars can reveal different groups of stars that were born together and their average properties contain information about their origin. Ultimately, extreme precision chemical and dynamical measurements may allow stars to be assigned to common birth clusters, which would lead to significant improvements in our understanding of nucleosynthesis, chemical evolution, enrichment of the ISM, migration of stars, and the merger history of the MW way. Thanks to an explosion in the amount and quality of chemodynamical data for stars in the Milky Way – from RAVE (Steinmetz et al. 2006), SEGUE (Yanny et al. 2009; Rockosi et al. 2022), APOGEE (Wilson et al. 2012; Majewski et al. 2016, 2017), Gaia-ESO (Gilmore et al. 2012; Gaia Collaboration et al. 2018), GALAH (De Silva et al. 2015; Martell et al. 2017), LAMOST (Zhao et al. 2012),

H3 (Conroy et al. 2019a,b), and DESI (DESI Collaboration et al. 2016a,b), for example – we've been experiencing a revolution in our understanding of our home Galaxy. Recent work, for example, has found strong evidence for a massive merger of the MW with a progenitor known as the Gaia-Sausage-Enceladus, which occurred approximately 10 billion years ago; the stars from this alien galaxy seem to dominate the inner ~ 25 kpc of the stellar halo (Helmi et al. 2018; Belokurov et al. 2018; Haywood et al. 2018).

Not unexpectedly, the story is never so simple as just collecting data and then analysing it. Stellar light, on its journey to a telescope's detector, can pass through a combination of intervening gas and dust, the Earth's atmosphere, and then finally the telescope's optics, all of which imprint signatures on that light. Detailed accounting of all these sources and effects is important for properly attributing light to a star, and therefore the inferences we make about its 7D properties. Astronomy has recently entered its Big Data Era, but our surveys are only as good as the measurements we make from them. The uniting theme of this thesis is that we are able to extract novel, previously-obscured inferences about our Galaxy when we employ rigorous statistical methods, especially in cases where measurement precision is a limiting factor. Understanding the data we collect and how best to model it is of extreme importance, so developing our analysis techniques now has compounding dividends for the future.

1.1 Outline

The remaining text of this thesis is organized as follows. Chapter 2 performs a 7D chemodynamical analysis of the HALO7D survey using newly-measured chemical

abundances, Chapter 3 develops a Bayesian technique to measure improved positions, parallaxes, and proper motions by combining *HST* and *Gaia*, and Chapter 4 models APOGEE spectra to reveal a wealth of information about the interstellar medium.

In Chapter 2, we consider The Halo Assembly in Lambda Cold Dark Matter: Observations in 7 Dimensions (HALO7D) survey, which measures the kinematics and chemical properties of stars in the Milky Way (MW) stellar halo to learn about the formation of our Galaxy. This survey is comprised of Keck II/DEIMOS spectroscopy and Hubble Space Telescope-measured proper motions of MW halo main sequence turn-off (MSTO) stars in the four CANDELS (Grogin et al. 2011; Koekemoer et al. 2011, PIs: S. Faber, H. Ferguson) fields: COSMOS, GOODSN, GOODSS, and EGS. HALO7D consists of deep pencil-beam fields, making it complementary to other contemporary wide-field surveys. In Cunningham et al. (2019a) and Cunningham et al. (2019b), the HALO7D survey is introduced and the radial velocities and proper motions are measured. We present the [Fe/H] and [α/Fe] abundances for 113 HALO7D stars in the Galactocentric radial range of $\sim 10 - 40$ kpc along four separate pointings. Using the full 7D chemodynamical data (3D positions, 3D velocities, and abundances) of HALO7D, we measure the velocity anisotropy, β, of the halo velocity ellipsoid for each field and for different metallicity-binned subsamples. We find that two of the four fields have stars on very radial orbits, while the remaining two have stars on more isotropic orbits. Separating the stars into high, mid, and low [Fe/H] bins at -2.2 dex and -1.1 dex for each field separately, we find differences in the anisotropies between the fields and between the bins; some fields appear dominated by radial orbits in all

metallicity bins while other fields show variation in their orbits between the [Fe/H] bins. These chemodynamical differences are evidence that the HALO7D fields have different fractional contributions from the progenitors that built up the MW stellar halo. Our results highlight the additional information available on smaller spatial scales compared to results from a spherical average of the stellar halo.

In Chapter 3, we build a hierarchical Bayesian pipeline, `BP3M`, that measures positions, parallaxes, and proper motions (PM) for cross-matched sources between *Hubble Space Telescope* (*HST*) images and *Gaia*, expanding on the recent `GaiaHub` (del Pino et al. 2022) tool with the specific goal of improving motion measurements in sparse fields (e.g. $N_* < 10$ per image). Our publicly-available code measures accurate PMs that are up to 13 times more precise than *Gaia* DR3 alone (e.g. for $20.5 < G < 21$ mag and a time baseline of $\sim$ 12 years) for nearby dwarf spheroidal galaxies. Using 787 *HST* images in the sparse COSMOS field, we measure `BP3M` PMs for 2184 unique sources, 25% of which have no *Gaia* PMs (median `BP3M` PM uncertainty of 1.12 mas/yr for $G > 21$ mag). Using 775 *HST* ACS/WFC images in the F814W filter, we find that the on- and off-axis skew terms of the transformation parameters that map *HST* images onto *Gaia* can be described by a time-varying function of *HST*'s rotation, potentially reducing the complexity of the transformation fitting by a factor of 2/3. The underlying statistics of `BP3M` are general in that they apply to any two or more sets of position measurements separated by time, regardless of telescope or instrument. Leveraging this, we simulate *Roman Space Telescope* (Spergel et al. 2015) observations and find an observing strategy that produces the largest improvements on parallax precision at no cost to

the PM uncertainty.

In Chapter 4, we investigate the impact of un-modelled ISM features and other residuals on chemical tagging – the ability to identify groups of stars as having been born together based on their chemical signatures – by detailed characterization of light in APOGEE (Wilson et al. 2012; Majewski et al. 2016, 2017) stellar spectra. Using $\sim$ 55000 spectra of $\sim$ 17000 red clump stars (Bovy et al. 2014) from the APOGEE DR16 dataset (Jönsson et al. 2020), we create 2nd order polynomial models of the continuum-normalized flux as a function of stellar parameters ($T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], and Age). The model and data show good agreement within uncertainties across the APOGEE wavelength range, although many regions reveal residuals that are not in the stellar rest-frame. We show that many of these residual features – having average extrema at the level of $\sim$ 3% in stellar flux on average – can be attributed to incompletely-removed spectral lines from the Earth's atmosphere and Diffuse Interstellar Bands (DIBs) from the intervening gas and dust of the ISM. After removing most of the remaining contamination from the Earth's sky, we identify 84 (25) absorption features that have less than a 50% (5%) probability of being explained by chance alone, including all 10 previously-known DIBs in the APOGEE wavelength range (e.g. Geballe et al. 2011; Zasowski et al. 2015; Elyajouri et al. 2017; Cox et al. 2014). Because many of these features occur in the wavelength windows that APOGEE uses to measure chemical abundances, characterization and removal of this non-stellar contamination is an important step in reaching the precision required for chemical tagging experiments. Proper characterization of these features will benefit Galactic ISM science and the currently-ongoing Milky Way Mapper

program of SDSS-V (Kollmeier et al. 2017), which relies on the APOGEE spectrograph.

Finally, Chapter 5 summarizes the work in this thesis and presents some directions for future research.

Chapter 2

HALO7D III: Chemical Abundances of Milky Way Halo Stars from Medium Resolution Spectra

2.1 Introduction

Owing to their long dynamical timescales, galactic stellar halos are long-lived structures that preserve information about their origins. Within a ΛCDM cosmology, galaxies are built up by merger events, each of which can contribute stars to the halo. The positions, kinematics, and chemical properties of halo stars thus reveal a galaxy's mass assembly history and information about the dwarf galaxy progenitors that con-

tributed to it. By studying the stellar halo of our home Galaxy, we seek to place the Milky Way (MW) in its cosmological context.

The chemodynamical properties of halo stars are powerful for constraining masses, star formation rates and efficiencies, and accretion times of progenitors as well as the total mass and shape of the MW gravitational potential (e.g., Eggen et al. 1962; Searle & Zinn 1978; Bullock & Johnston 2005; Helmi 2008). While the positional clustering of stars from an accreted satellite is eventually washed out, the kinematic coherence of accreted debris persists for much longer periods of time. Stellar atmospheric chemical abundances are relatively stable over a star's lifetime, and they are a direct result of the environment in which it was formed; interstellar medium enrichment levels, star formation rates, and formation lifetimes of a galaxy all impact the chemical makeup of star-forming gas as a function of time.

As α elements (i.e. O, Ne, Mg, Si, S, Ar, Ca, Ti) can be produced at early times in core-collapse events at relatively constant rates with iron, high [α/Fe] stars tend to be formed at early times in a galaxy's star-forming life. After enough time (e.g. $\sim$ 330 Myr as measured by Maoz & Badenes 2010), and assuming sufficient star formation, Type Ia supernovae "turn on" and create much of the iron we see in the universe. This causes the [α/Fe] ratio to drop as [Fe/H] increases (Wallerstein 1962; Tinsley 1980). The mass-metallicity relation of local dwarf galaxies (Kirby et al. 2013, 2017, 2020) reveals that the metallicity distribution function (MDF) of a galaxy is tied to its mass; this arises because more massive systems have deeper potential wells that are able to retain a greater fraction of their enriched gas from supernovae. While this

relationship was determined for local dwarf galaxies that are observed today, recent work (e.g. Leethochawalit et al. 2018, 2019; Naidu et al. 2022) has explored how these mass-metallicity relations were different at earlier cosmic times. The star formation and quenching time of a progenitor system are impacted during its accretion, but the chemical properties of the accreted stars are linked to the mass, star formation rate, and formation lifetime of their birth environment; this inter-relatedness enables the inferences of progenitor properties from stellar halo chemical abundances (e.g., Lee et al. 2015; Hasselquist et al. 2021; Horta et al. 2023; Cunningham et al. 2022).

At present time, a stellar halo is comprised of stars from many different progenitor systems. Simulations of purely accreted stellar halos (e.g. Bullock & Johnston 2005; Robertson et al. 2005; Font et al. 2006a,b,c) have average stellar halo abundances driven by their merger histories (e.g. Johnston et al. 2008), with average [α/Fe] tracking accretion time of infalling dwarf galaxies and [Fe/H] tracking the mass/luminosity of those dwarfs. Recent work with state-of-the-art simulations (e.g., Horta et al. 2023, using the *Latte* suite of FIRE-2 simulations; Hopkins (2015); Hopkins et al. (2018); Wetzel et al. (2016)) have further refined our understanding of how the distributions of present-day chemodynamical observables of stellar halo stars are dictated by the accretion times and masses of disrupted dwarfs.

In addition to stars accreted from dwarf galaxy mergers, in-situ stars – those formed in the potential well of the host galaxy – can be heated onto orbits similar to accreted halo stars during merger events. As a result, the halo population is a combination of in-situ stars and those accreted during the many mergers a galaxy experiences

throughout its history. Zolotov et al. (2009), for example, finds evidence from simulations that the in-situ halo fraction of the inner regions traces how quiescent the recent merger history has been; more recent mergers tend to cause the inner halo in-situ fraction to decrease[1]. Using data from the H3 survey (Conroy et al. 2019b), Naidu et al. (2020) find that the in-situ halo fraction drops from 60% to 5% of the total halo mass when going from Galactocentric radii of 8 kpc to 20 kpc.

For this paper, "in-situ halo" refers to the progenitor high-α disk that was kinematically heated through early merger events (Nissen & Schuster 2010; Bonaca et al. 2017, 2020; Haywood et al. 2018; Di Matteo et al. 2019; Amarante et al. 2020). This is the so-called "Splash" of Belokurov et al. (2020). It is thus chemically similar to the thick disk – that is, peaked towards relatively metal-rich [Fe/H] around -0.5 dex (Naidu et al. 2020) – but consists of stars on more isotropic orbits instead of predominantly circular ones. Belokurov et al. (2020) show that their "Splash" sample has less net prograde rotation and larger scatters in all velocity components compared to their "Thick Disk" sample.

Thanks in large part to the *Gaia* survey (Gaia Collaboration et al. 2018), our current picture of the MW's formation history is becoming clearer: aside from the recent interactions with Sgr and the LMC/SMC, the MW halo seems to have had a relatively quiescent recent history. Recent work has shown that the inner ~ 25 kpc of the stellar halo is dominated by a single massive progenitor called Gaia-Sausage-Enceladus (GSE) (Belokurov et al. 2018; Helmi et al. 2018; Haywood et al. 2018). The GSE is relatively

[1]While their definition of "in-situ stars" is different than ours, the dependence on fractional contribution of in-situ halo stars as a function of merger history should be similar.

metal-rich, with a peak [Fe/H] of ~ -1.2 dex (Naidu et al. 2020), and radially biased orbits. The GSE has a net rotation, $\langle v_\phi \rangle$, that is consistent with 0 km s^{-1} (Belokurov et al. 2020), and is estimated to have come from a 4:1 mass ratio merger approximately 10 Gyr ago (Helmi et al. 2018). While the GSE and in-situ halo dominate the bulk of the inner regions of the stellar halo, there have been many other substructures identified over the last five years (e.g. Myeong et al. 2019; Koppelman et al. 2019; Yuan et al. 2020; Naidu et al. 2020; Belokurov et al. 2020). There is growing evidence that 92% to 99% of the MW stellar halo stars can be associated with one of the currently known progenitors (Naidu et al. 2020).

These detailed inventories of the MW stellar halo have helped constrain answers to questions about our Galaxy, such as the approximate formation history, fractional contributions from different progenitors, and range of progenitor properties. Many of these studies rely on *Gaia*-based parallaxes and proper motions. Because of *Gaia*'s apparent magnitude limit of $G \sim 20$ mag, this means that these samples are either focused on nearby main sequence (MS) halo stars when studying the local halo or more distant evolved stars when studying the distant halo. These evolved stars are intrinsically less numerous/spatially dense than their MS counterparts, which means that much of the distant-halo work has focused on average properties over large areas of the sky.

The Halo Assembly in Lambda Cold Dark Matter: Observations in 7 Dimensions (HALO7D; Cunningham et al. 2016, 2019a,b) is complementary to contemporary MW stellar halo surveys because it targets 3D positions, 3D velocities, and abundances of main sequence turn-off stars at moderate halo radii ($10 < r < 40$ kpc). The HALO7D

sample consists of Keck II/DEIMOS spectroscopy and HST-measured proper motions for stars in the apparent magnitude range $19 < m_v < 24.5$ mag, making it a deep complement to *Gaia*-based surveys. This paper, the third in the series, measures the 7-th and final dimension of HALO7D stars: chemical abundances. The high spatial density of MS halo stars allows us to compare average chemodynamical properties along different lines of sight (LOS). Cunningham et al. (2019b), for example, measure the halo velocity anisotropy along the four HALO7D LOS and find variations between the different fields.

In this paper, we describe the HALO7D data set in Section 4.2. We create a Bayesian spectrophotometric pipeline to measure chemical abundances ([Fe/H] and [α/Fe]) and stellar parameters ($T_{\rm eff}$, $\log g$, Age, and distance) for MSTO stars without known distances in Section 3.3 and present the resulting abundance measurements for the HALO7D sample in Section 2.4. In Section 2.5, we combine our abundances with previously-measured LOS velocities and proper motions from HALO7D (Cunningham et al. 2019a,b) to measure the variation in average chemodynamical properties – such as net halo rotation, $\langle v_\phi \rangle$, and the velocity anisotropy parameter, β – using different subsamples of the HALO7D stars. Our findings are summarized in Section 2.6. The detailed tests on the outputs of our chemical abundance pipeline are described in Appendix 2.7.1, and then validated against well-studied globular clusters in Appendix 2.7.2. We show the statistical significance of our kinematic measurements in Appendix 2.7.3.

Table 2.1: Summary of Targets in CANDELS Fields with line-of-sight velocities and proper motions from Cunningham et al. (2019a) and Cunningham et al. (2019b).

Field	l (deg)	b (deg)	Area[a] (arcmin^2)	N Halo Stars with v_{3D}	Catalog References
COSMOS	236.8	42.1	288	81	Muzzin et al. (2013); Nayyeri et al. (2017)
GOODSN	125.9	54.8	166	32	Barro et al. (2019)
GOODSS	223.6	−54.4	160	20	Guo et al. (2013)
EGS	96.4	60.4	384	66	Barro et al. (2011) Stefanon et al. (2017)
Total	–	–	998	199	

[a]The listed field area is the area covered with HST multi-epoch imaging.

2.2 Data

The HALO7D dataset consists of HST-measured proper motions and Keck II/ DEIMOS spectroscopy for 199 main sequence turn-off (MSTO) MW halo stars in four Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011, PIs: S. Faber, H. Ferguson) fields: COSMOS, GOODSN, GOODSS, and EGS. These fields are located at high Galactic latitudes – meaning they have minimal foreground contamination from the MW disk – and they are not located in regions of previously-known streams or substructure (e.g. Sagittarius). While the stars that make up the HALO7D sample lie within the same footprints as the CANDELS fields (see Figure 1 of Cunningham et al. 2019a) – and we refer to them using the same field names – it should be noted that the HALO7D dataset does not include every MSTO star found in the CANDELS fields. The first paper in this series (Cunningham et al. 2019a) presents the DEIMOS spectroscopic dataset (see their Section 2) and measures line-of-sight velocities with a hierarchical Bayesian pipeline

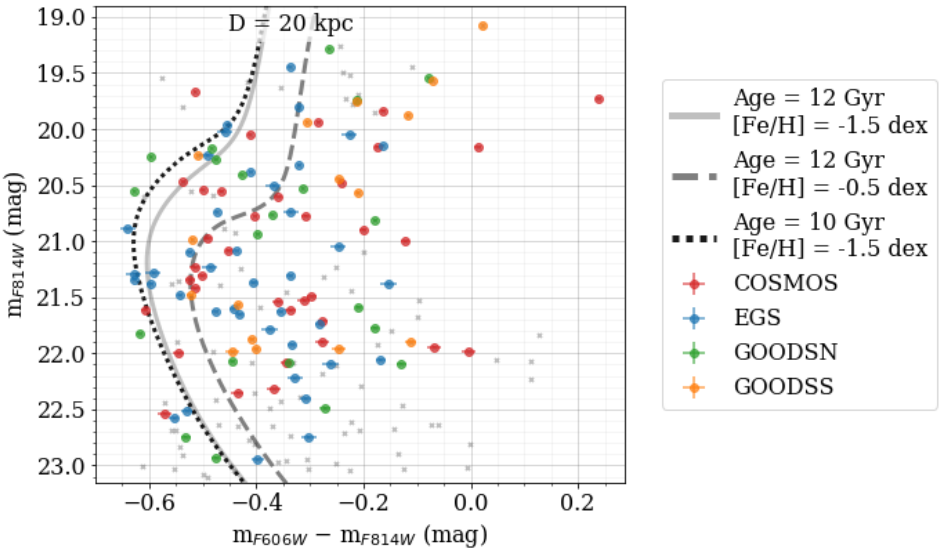

Figure 2.1: Color-magnitude diagram of the 113 HALO7D stars with measured LOS velocities and chemical abundances in HST filters (STMAG). Three MIST isochrones at a distance of 20 kpc with typical MW halo properties are shown in black/grey to guide the eye in this region of color-magnitude space; an increase (decrease) in distance causes the isochrones to move down (up) vertically to fainter (brighter) apparent magnitudes. The faint grey dots show the 86 HALO7D stars that do not have measured abundances.

called `velociraptor`. The second paper (Cunningham et al. 2019b) presents proper motions measured from multi-epoch HST imaging (see their Section 2 and Table 2 for a description of the *HST* Programs) and characterizes the halo velocity ellipsoid.

A detailed description of the HALO7D sample and the selection process are presented in the first two papers in this series. To summarize the relevant information, the fields were chosen because of their deep, multi-epoch HST astrometry and photometry, which enables proper motion measurements to much fainter magnitudes than *Gaia*. HALO7D's velocity sample consists of stars in the $19 < m_v < 24.5$ mag range and are generally blue to minimize the impact of contamination by foreground disk stars; to see the full CMD selection criteria, please see Section 2.2 and Figures 2 and 3 of Cunning-

ham et al. (2019a). HALO7D is thus complementary to previous studies because it is able to measure kinematics and chemical abundances for individual main sequence stars at halo distances with high enough spatial sampling density to measure halo properties along individual lines-of-sight instead of measuring sky-averaged properties versus Galactocentric radius.

The Keck II/DEIMOS spectroscopic observations were collected between March 2014 and April 2017 using the 600 line/mm grating with a 7200 Å central wavelength configuration and 1" slitwidth. These medium resolution spectra ($R \sim 2000$) consist of 8192 pixels, pixel spacing of ~ 0.65 Å/pixel, and covering a typical wavelength range of $\sim 5000 - 10000$ Å. Each target was typically observed for $\sim 5 - 6$ hours over the course of this time period, with a minimum of 2 and a maximum of 33 individual visits per target. The raw spectroscopic data were reduced with the `spec2d` pipeline produced by DEEP2 at UC Berkeley (Cooper et al. 2012). Table 2.1 shows a summary of the observations in each field, Figure 2.1 shows a color-magnitude diagram for the 199 stars in HALO7D with the colored points denoting the 113 stars for which we are able to measure abundances. Figure 2.2 shows cumulative histograms of the combined spectral signal-to-noise ratio (SNR) for the stars with abundances in each field, where we define the combined SNR using a quadrature sum of the SNRs of the individual observations:

$$\mathrm{SNR}_{\mathrm{combined}} = \left(\sum_{i}^{n_{obs}} \mathrm{SNR}_i^2 \right)^{1/2}.$$

The median combined spectral SNRs are 62.2, 64.8, 95.8, and 71.3 Å^{-1} for COSMOS, EGS, GOODSN, and GOODSS respectively and 67.2 Å^{-1} for the complete sample.

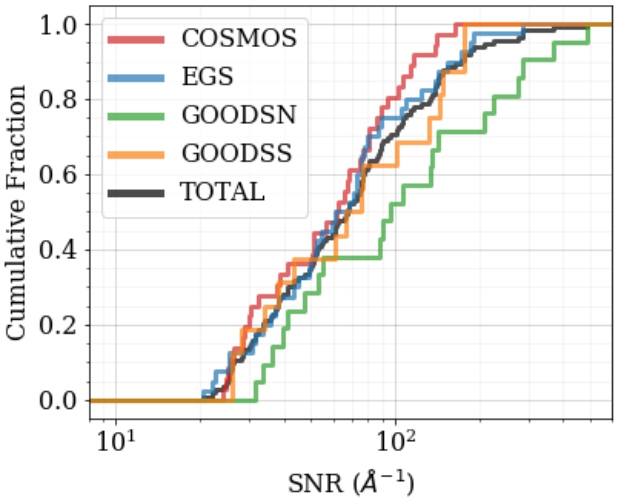

Figure 2.2: Cumulative histograms of the combined spectral signal-to-noise ratios of stars in the chemistry sample of HALO7D (colored points in Figure 2.1). The black line represents the total sample. By design, GOODSN had more observations than the other fields and thus shows higher combined SNRs. The median SNRs for each field and the total sample are found in the text.

2.3 Measuring Stellar Parameters and Abundances

We create a Bayesian pipeline that uses stellar photometry and spectroscopy to measure chemical abundances ([Fe/H], [α/Fe]), stellar parameters ($T_{\rm eff}$, $\log g$, age), and distances for our HALO7D stars. This technique relies on the library of MIST isochrones[2] (Dotter 2016; Choi et al. 2016; Paxton et al. 2011, 2013, 2015) and a set of synthetic model spectra (described by Escala et al. (2019) in the blue, $4100 - 6300$ Å, and Kirby et al. (2008); Kirby (2011) in the red, $6300 - 9100$ Å). The model spectra are generated with the MOOG spectral-synthesis software (Sneden 1973), using the ATLAS9 model atmospheres (Kurucz 1993; Kirby 2011). Each synthetic spectrum is defined by a set of ($T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe]) values. For these spectra, [α/Fe] is the

[2] https://waps.cfa.harvard.edu/MIST/index.html

total α-element abundance of a stellar atmosphere, where O, Ne, Mg, Si, S, Ar, Ca, and Ti are the α-elements we consider.

Many techniques that measure chemical abundances and stellar parameters from spectra using synthetic model spectra typically study populations of stars that have photometrically-constrained distances (Kirby et al. 2010; Escala et al. 2019, for example) or parallax measurements (Conroy et al. 2019b; Cargile et al. 2020). This means that distance-degenerate parameters, such as $\log g$, are able to be fixed and a smaller region of parameter space needs to be explored. For the HALO7D sample, however, we deal with MSTO stars with unknown distances. The absolute magnitudes of stars near the MSTO vary much more than their colors, which means that the stars have large distance uncertainties. As a result, we must explore larger regions of parameter space and consider the relationships between distance and the possible stellar parameters. To help constrain our abundance fits, we use the MIST isochrones, measured properties of the MW stellar halo, and photometric observations of the stars in our sample to create multi-dimensional, multi-modal prior distributions on the parameters of interest for each star.

2.3.1 Generating Prior Distributions of Stellar Parameters

In building our priors, we use MIST isochrones in the range of $0.1 <$ age $<$ 14.6 Gyr and $-4.0 <$ [Fe/H] < 0 dex in steps of 0.2 Gyr and 0.02 dex respectively. The current version of the publicly available MIST isochrones are only for $[\alpha/\mathrm{Fe}] = 0$ dex, so we are not able to include this parameter in the isochrone-derived prior distributions; we assume a flat prior in the range of $-0.8 < [\alpha/\mathrm{Fe}] < +1.2$ dex. These isochrones are

then assigned weights based on average properties of the MW stellar halo, such as an assumed metallicity distribution function (MDF) and age distribution. Our particular choices of parameters for the halo priors are shown in Table 2.2. The [Fe/H] distribution is chosen to have a moderately metal-rich peak and a width that is not overly constraining that is consistent with recent halo studies (e.g., Belokurov et al. 2018; Helmi et al. 2018; Mackereth et al. 2019; Conroy et al. 2019a). Similarly, the age distribution comes previous studies such as Kalirai (2012) and Bonaca et al. (2020). As we show in Appendix 2.7.2, the abundances we measure are fairly insensitive to our choice of priors in the spectral SNR range of our data.

The points within a particular isochrone each have an initial mass, $M_{F606W} - M_{F814W}$ color, M_{F814W} absolute magnitude, $T_{\rm eff}$, and $\log g$. The isochrone points are weighted by integrating a Kroupa (2001) initial mass function (IMF) over the range of masses within an isochrone to account for uneven mass spacing.

Until now, the isochrones have been weighted only by properties that are generic to a halo population, but we now create unique prior distributions for the stellar parameters of each star. We apply additional weights to the isochrone points for each star separately using the observed photometry. Each star has observed apparent HST magnitudes, m_{F606W} and m_{814W}, and corresponding uncertainties, σ_{F606W} and σ_{F814W}, which we have de-reddened using the dust maps of Schlegel et al. (1998) through the `dustmaps` package in `Python` (Green 2018).

To give the isochrone points weights based on distances/absolute magnitudes,

Table 2.2: MW halo prior distributions for abundance pipeline.

Distribution	Functional Form
$p([\mathrm{Fe/H}])$	$\mathcal{N}(-1.5\ \mathrm{dex}, 1.0\ \mathrm{dex})$
$p(\mathrm{age})$	$\mathcal{N}(12\ \mathrm{Gyr}, 2.0\ \mathrm{Gyr})$
$p([\alpha/\mathrm{Fe}])$	$\propto 1$, for $-0.8 < [\alpha/\mathrm{Fe}] < 1.2$
$p(\mathrm{mass}\|[\mathrm{Fe/H}], \mathrm{age})$	Kroupa (2001) IMF, $\propto k \left(\frac{\mathrm{mass}}{M_\odot}\right)^{-\alpha}$ with $\begin{cases} k = 25,\ \alpha = 0.3, & \mathrm{mass} < 0.08 M_\odot \\ k = 2,\ \alpha = 1.3, & \mathrm{mass} < 0.5 M_\odot \\ k = 1,\ \alpha = 2.3, & \mathrm{mass} > 0.5 M_\odot \end{cases}$
$p(\mu)$	$\propto D^3 \left(\frac{R_q}{27\mathrm{kpc}}\right)^{-\alpha}$ with $\begin{cases} \alpha = 2.3, & R_q < 27\,\mathrm{kpc} \\ \alpha = 4.6, & R_q \geq 27\,\mathrm{kpc} \end{cases}$

we use the MW stellar halo density profile of Deason et al. (2011):

$$\frac{dN}{dV} \propto \begin{cases} \left(\frac{R_q}{27\,\mathrm{kpc}}\right)^{-2.3}, & R_q < 27\,\mathrm{kpc} \\ \left(\frac{R_q}{27\,\mathrm{kpc}}\right)^{-4.6}, & R_q \geq 27\,\mathrm{kpc} \end{cases}$$

where $R_q^2 = x^2 + y^2 + (z/q)^2$ with $q = 0.59$. Accounting for volume elements and the Jacobian between distance and distance modulus, the distance modulus prior is thus $p(\mu) \propto D^3 \cdot \frac{dN}{dV}$. We marginalize over the distance modulus by taking equally spaced values in μ that correspond to distances between 0.1 kpc and 500 kpc and comparing the observed stellar magnitudes to the implied apparent magnitude of the isochrone point at a given distance modulus. Overall, the weight of a given isochrone point ends up as:

$$\begin{aligned} p(\mathrm{p}_i|[\mathrm{Fe/H}], \mathrm{age}) \propto p(\mathrm{mass}_i|[\mathrm{Fe/H}], \mathrm{age}) \cdot \sum_{j=0}^{n_\mu} [p(\mu_j) \cdot \\ \mathcal{N}(M_{i,F606W} + \mu_j | m_{F606W}, \sigma_{F606W}) \cdot \\ \mathcal{N}(M_{i,F814W} + \mu_j | m_{F814W}, \sigma_{F814W})] \end{aligned}$$

where p_i is the i-th point in an isochrone defined by $([\mathrm{Fe/H}], \mathrm{age})$ which has absolute magnitudes $(M_{i,F606W}, M_{i,F814W})$.

When comparing the MIST isochrones to HST photometry of stars in a handful of nearby globular clusters, we noticed that many of the isochrones required color offsets as large as ~ 0.02 mag to have better agreement with the data. To allow for a potential mismatch between the MIST isochrones and the halo stars, we also marginal-

ize over color offsets between -0.02 mag and 0.02 mag. This is much larger than the typical photometric uncertainty of stars in the HALO7D sample (median uncertainty in $m_{F606W} - m_{F814W}$ of 0.006 mag), so this marginalization has a relatively large effect in increasing the width of the prior distributions of stellar parameters.

For each star, the total weighting for a given isochrone point is thus a product of its isochrone [Fe/H] and Age weighting, its mass weighting, and the weighting from marginalizing over the distance modulus while comparing the photometry.

At a given color in an isochrone, there are generally three possible distances corresponding to a star being on the main sequence, the subgiant branch, or the horizontal branch. This results in our prior distributions having three local maxima: one peak for each of the possible phases in a star's evolution. Because it is generally more efficient to sample posterior distributions that are singly-peaked, we break each isochrone up into these three phases. The prior probability of a particular phase in any given isochrone is the fraction of the total weight of the isochrone in that phase. In this work, phase = 0 corresponds to MS stars, phase = 1 corresponds to the sub-giant branch, and phase = 2 corresponds to the horizontal branch[3].

Finally, for each phase of each isochrone, we generate a 3D prior distribution on ($T_{\rm eff}$, $\log g$, and M_{F814W}) by passing the weighted isochrone points to a kernel-density estimator (KDE) with a Gaussian kernel. We compute the KDE width by measuring a standard deviation and mean of each of the parameters ($T_{\rm eff}$, $\log g$, and M_{F814W}) in the set of weighted isochrones, and normalize the measurements so that

[3]In the MIST parlance, our phase = 0 is also their phase = 0, our phase = 1 is their phase = 2, and our phase = 2 is their phase = 3 and phase = 4.

each individual parameter's distribution corresponds to a unit Gaussian. The Gaussian width used in the kernel density estimator is chosen to be 0.1, such that the width is 10% of the standard deviation in each parameter. This KDE approach has the benefit of allowing combinations of $T_{\rm eff}$, $\log g$, and M_{F814W} that do not fall perfectly on the isochrone, meaning we are less reliant on the isochrones perfectly capturing the relationships between stellar parameters.

Overall, the prior probabilities of the stellar parameters are described by:

$$\begin{aligned} p(\vec{\theta_*}) =& p(T_{\rm eff}, \log g, [\rm Fe/H], [\alpha/\rm Fe], M_{814W}, \rm age, \rm phase) \\ \propto& p([\alpha/\rm Fe]) \cdot p([\rm Fe/H]) \cdot p(\rm age) \cdot p(\rm phase|[Fe/H], age) \cdot \\ & p(T_{\rm eff}, \log g, M_{814W}|[\rm Fe/H], \rm age, \rm phase) \end{aligned} \tag{2.1}$$

where $p(T_{\rm eff}, \log g, M_{814W}|[\rm Fe/H], \rm age, \rm phase)$ is calculated by evaluating the KDE generated from the isochrone points with that particular [Fe/H], age, and phase: $KDE_{\rm [Fe/H],age,phase}(T_{\rm eff}, \log g, M_{814W})$.

2.3.2 Preprocessing of Spectra

Before the spectral observations can be used in our abundance pipeline, we must shift each of the spectra to the rest frame, continuum-normalize, identify useful wavelength regions, and characterize a few quality-of-observation parameters such as the line-spread function (LSF). For our chemical abundance pipeline, we do not coadd the multiple spectral observations of a given star into a single observation. Instead, we model each observation simultaneously. We choose this approach because the spectral

observations were taken over the course of years and can have vastly different observing conditions (e.g. seeing, line spread functions, wavelength solution offsets), which causes their coadded spectrum to have washed-out/hard-to-model features. This section is quite technical, so readers who are more interested in the big-picture steps of the abundance pipeline may choose to skip ahead to Section 2.3.3 or to the results in Section 2.4.

Measuring Line Spread Functions

As a star's light passes through the atmosphere to the telescope, its light is spread out by the seeing we measure during data collection. It is adequate for our purposes to assume the resulting shape is a Gaussian with a width that is the $\mathrm{FWHM}_{\mathrm{seeing,i}}/2.355$, where $i \in \{1, \ldots, n_{obs}\}$ refers to the spectral observation number. Accounting for the pixel scale and anamorphic factor allows us to convert this into a width in units of Angstroms in the spectral dimension. Because of the 1" width of the slits in the DEIMOS mask, any light outside of this width is truncated. The star's truncated-Gaussian light is further smoothed out as it passes through the instrument and lands on the CCD. To characterize this additional amount of instrument smoothing as a function of wavelength, we use arc lamp exposures[4].

We identify peaks in the arc lamp spectrum and then fit those peaks with a top-hat function convolved by a Gaussian, where the Gaussian width is unknown and the top-hat width is set by the size of the slit. This is because the arc lamp light fully illuminates each slit, producing a top-hat shape, before it passes through

[4]We use Kr, Xe, Ar, Ne for our red arcs, and the same elements plus Hg, Cd, Zn for our blue arcs.

the instrument. In particular, we are interested in measuring the smoothing Gaussian widths. The resulting widths as a function of wavelength are approximately quadratic with a minimum near the chip gap. For each individual observation, we measure this quadratic function on the blue- and red-chips separately, giving the additional smoothing of the instrument as a function of wavelength. The final $\mathrm{LSF}_i(\lambda)$ for the i-th observation is then the seeing-defined truncated Gaussian convolved with the wavelength-dependent instrument smoothing. In cases where the seeing is very good, such that virtually all of the star's light is completely inside of the slit (e.g. $3\sigma_{\mathrm{seeing}} < \mathrm{slitwidth}/2$ which implies a seeing FWHM of ~ 0.39" for our slitwidth of 1"), the resulting $\mathrm{LSF}_i(\lambda)$ is a Gaussian with a total width that is the quadrature sum of the seeing Gaussian and the instrument smoothing. In most cases, however, the seeing is large enough relative to the slitwidth that we find it better to use the truncated-Gaussian-then-smoothed model instead. During the fitting process, we allow for the instrument smoothing function to change by a multiplicative factor between 0.5 and 1.5 to not be overly constraining. We place a prior on this multiplicative factor that is a Gaussian centered at 1 and has a width of 0.1 for each observation and allow the blue and red data to have different multiplicative factors:

$$\text{blue mult}_{\mathrm{phase},i} \sim \mathcal{N}(1.0, 0.1)$$

$$\text{red mult}_{\mathrm{phase},i} \sim \mathcal{N}(1.0, 0.1)$$

Shift to Stellar Rest Frame and Pixel Masking

Next, we shift each of the n_{obs} observations of a single star to stellar restframe using the `velociraptor`-measured LOS velocities from Cunningham et al. (2019a). Specifically, we use the median LOS velocity from each spectral observation's posterior distribution. The data wavelengths are then restricted to $4100 - 9100$ Å because that is the coverage of the synthetic model spectra we will compare to. We also mask out a few additional wavelength regions, some before shifting to restframe and others after, as listed in Table 2.3; these include telluric features, poorly-modeled (or un-modeled) features in the synthetic model spectra, and regions where the wavelength solution of the spectral reductions are unstable.

For our spectroscopic setup with DEIMOS, there are not many arc lamp lines at particularly blue wavelengths, which makes anchoring the wavelength solution difficult in this region. We find that the wavelengths on the red-chip (i.e. $\lambda > 7200$ Å) are robust and that the blue-chip wavelengths are quite reliable down to $\lambda \sim 5500$ Å, but they are often unstable below $\lambda \sim 5000$ Å. Using a handful of stars with previously-measured stellar parameters and high SNR spectra, we cross-correlate with the best-fit synthetic model spectrum at different wavelength locations to re-measure the wavelength solution and compare it to the output of `spec2d`. Because the wavelength solution we measure is using the stellar features at all wavelengths, this process is able to anchor the solution using lines at the bluest wavelength where arc lines are not available. We find that the offset as a function of wavelength is approximately linear for $\lambda > 5000$ Å on the blue-chip, with the size of the maximum offset being ~ 3 Å.

Table 2.3: Masked spectral regions.

Masked Before Shift to Rest Frame	
Name	Region Masked (Å)
Unstable Wavelength Solution[a]	< 5000
B-Band	(6864, 7020)
A-Band	(7570, 7713)
Telluric Feature	(7157, 7325)
Telluric Feature	(8123, 8356)
Telluric Feature	(8235, 8275)
Telluric Feature	> 8933
Chip Gap	± 20 at gap
Masked After Shift to Rest Frame	
Name	Region Masked (Å)
MgH[b]	(5115, 5125)
Na D1, D2[c]	(5885, 5910)
Ca I (λ6343)	(6341, 6346)
Ca I (λ6362)	(6356, 6365)
Hα	(6559.797, 6565.797)
K I (λ7665)	(7662, 7668)
V I (λ8116, λ8119),	(8113, 8123)
Poor Arcturus Model [d]	(8317, 8330)
Ca II (λ8498)	(8488.023, 8508.023)
Ca II (λ8542)	(8525.091, 8561.091)
Ca II (λ8662)	(8645.141, 8679.141)
Mg I (λ8807)	(8804.756, 8809.756)

[a] Very blue DEIMOS wavelengths tend to be extrapolations of a small number of arc lamp lines and are not easily fixed with our correction prescription described in Section 2.3.2.

[b] This molecular absorption feature is masked for stars that have more than 5% prior probability of $T_{\rm eff} < 5300$K because the synthetic models currently do not include this feature.

[c] This absorption feature is masked because of potential contamination from ISM absorption.

[d] From Kirby et al. (2008), this is a region where the model spectra were showed to poorly reproduce the spectral features of Arcturus.

With these lessons in mind, we decide to mask out wavelengths less than 5000 Å for all spectra because the wavelength offset function tends to become non-linear in this region. During the stellar parameter measuring process that we will discuss in Section 2.3.3, we fit each spectral observation with a linear wavelength offset function on the blue-chip wavelengths to correct for these issues. Ideally, we would use more terms in the wavelength offset function. However, we must balance the number of parameters being fit per observation with computation time; we find that a linear correction function is a good compromise. For spectra with SNR less than 20 Å^{-1}, we mask out wavelengths less than 5500 Å because this is where it becomes difficult to assess how well the wavelength offset has been measured. For the same reason, we drop individual spectral observations with SNR less than 3 Å^{-1}.

Continuum Normalization

We then measure an initial continuum estimate for each spectrum and for each possible stellar evolution phase, as defined in Section 2.3.1. We take the synthetic model defined by the median stellar parameters of the corresponding prior distribution, degrade it to the data quality using the $\text{LSF}_i(\lambda)$, and then divide the result from the observed spectrum. This gives an approximate continuum assuming a particular phase; we then smooth the resulting spectrum using a median boxcar of width 6.5 Å (i.e. approximately 10 pixels) to limit the impact of outliers such as poorly-removed skylines. This median-binned spectrum is then fit with a cubic B-spline whose knots are spaced by 100 Å, yielding an estimate of a given observation's continuum assuming a particular phase, $c_{\text{phase},i}(\lambda)$. This process is repeated for each spectral observation and for each phase.

Identifying Useful Wavelength Regions and Corrections to Wavelength Solutions

We also use these degraded model spectra to build prior distributions for the parameters of the linear wavelength offset function, $\Delta\lambda_i(\lambda) = m_{\lambda,i} \cdot \lambda + b_{\lambda,i}$. We cross-correlate the synthetic model with the spectral data in steps of $\sim$ 100 Å. This gives measures of the wavelength offset as a function of wavelength for a particular observation. We then fit a line to these offsets, taking the resulting fit's mean, $\vec{\theta}_{\lambda,i}$, and covariance matrix, $\boldsymbol{V}_{\lambda,i}$, to define the prior distribution as a multivariate normal distribution,

$$m_{\lambda,i}, b_{\lambda,i} \sim \mathcal{MVN}_2(\vec{\theta}_{\lambda,i}, 10\boldsymbol{V}_{\lambda,i})$$

where we inflate the covariance matrix by a factor of 10 so as not to be overly constraining.

Finally, we identify wavelength regions of the spectra that are particularly useful for measuring stellar parameters. Using the prior distributions to define a useful region of parameter space (e.g. 95% prior probability), we are able to define a grid of equally spaced $T_{\rm eff}$, $\log g$, [Fe/H], and [α/Fe] values for each phase. For each set of parameters on the prior grid, we degrade the corresponding synthetic model to the data quality and compare the model spectra to each other as well as to the continuum-normalized spectral observations. This is done for two purposes. The first reason is to determine wavelength regions that do not change significantly across the parameter grid, which means they provide the least power for our likelihood measurements. We choose to mask the 10% least useful pixels of each observation. This has the additional

benefit of speeding up computations because we consider a smaller number of pixels; in total, the masking of poorly-modeled, telluric, and low-likelihood information regions leaves between $35-55\%$ of the data pixels ($\sim 3000-4500$ of the original 8129 pixels) for the likelihood measurements. The other reason is to better define the continuum for each observation given a particular phase. For each set of parameters in our prior grid, we divide the continuum-normalized data by the current synthetic model and then fit a B-spline with 100 Å-spaced knots to the result, which defines the continuum adjustment required to have the best agreement between that model and the data. The new continuum for each phase is taken to be the initial continuum multiplied by the new continuum adjustment of the model that had the minimum χ^2 comparison. The parameters that define the best model (i.e. minimum χ^2) for each phase is where we choose to begin our search of parameter space during the fitting process.

2.3.3 Fitting Spectra with Synthetic Models

With our prior distributions and rest frame, continuum-normalized spectra in hand for each star, we begin the fitting process. The synthetic model spectra we compare to are defined by a set of $T_{\rm eff}$, $\log g$, [Fe/H], and [α/Fe] values according to the model grid in Table 2.4. In cases where we draw a set of parameters that don't lie directly on the model grid, we linearly interpolate from the nearest neighboring models using between 2 and 2^4 nearest neighbors in ($T_{\rm eff}$,$\log g$,[Fe/H],[α/Fe]) space. For each observation of a given star, the synthetic model is smoothed with the corresponding $\mathrm{LSF}_i(\lambda)$ and re-sampled onto the data wavelengths.

Table 2.4: Model spectrum grid spacing by parameter value from Escala et al. (2019, Table 4).

Parameter	Min. Value	Max. Value	Step
$T_{\rm eff}$ (K)	3500	5600	100
	5600	8000	200
$\log g$ (cm s^{-2})	0.0 ($T_{\rm eff} < 7000$ K)	5.0	0.5
	0.5 ($T_{\rm eff} > 7000$ K)	5.0	0.5
[Fe/H]	-4.5 ($T_{\rm eff} \leq 4100$ K)	0.0	0.1
	-5 ($T_{\rm eff} \leq 4100$ K)	0.0	0.1
[α/Fe]	-0.8	$+1.2$	0.1

The posterior probability for our parameters of interest is:

$$\begin{aligned} p(\vec{\theta}_*, \boldsymbol{\theta}_{spec} | \boldsymbol{F}, \boldsymbol{\Sigma}) \propto & p(\vec{\theta}_*) \cdot p(\vec{\theta}_{spec}) \cdot p(\boldsymbol{F} | \vec{\theta}_*, \boldsymbol{\theta}_{spec}, \boldsymbol{\Sigma}) \\ \propto & p(\vec{\theta}_*) \cdot \prod_{i=1}^{n_{obs}} \left[p(\vec{\theta}_{spec,i}) \cdot p(\vec{f}_i | \vec{\theta}_*, \vec{\theta}_{spec,i}, \vec{\sigma}_i) \right] \\ \propto & p(\vec{\theta}_*) \cdot \prod_{i=1}^{n_{obs}} \left[p(\vec{\theta}_{spec,i}) \cdot \prod_{j=1}^{n_{pix}} p(f_{i,j} | \vec{\theta}_*, \vec{\theta}_{spec,i}, \sigma_{i,j}) \right] \end{aligned} \tag{2.2}$$

where i corresponds to the spectral observation number, j corresponds to the pixel number within a spectrum, $\boldsymbol{F} = (\vec{f}_1, \vec{f}_2, \ldots, \vec{f}_{n_{obs}})$ are the fluxes of the measured spectra with corresponding uncertainties of $\boldsymbol{\Sigma} = (\vec{\sigma}_1, \ldots, \vec{\sigma}_{n_{obs}})$, $\boldsymbol{\theta}_{spec} = (\vec{\theta}_{spec,1}, \ldots, \vec{\theta}_{spec,n_{obs}})$ are the set of spectral parameters for all observations, and θ_* is the set of stellar parameters with $p(\theta_*)$ defined in Equation 2.1. Because the spectra for a given star are on a common wavelength array, λ_j represents the wavelength of a given pixel, and $f_{i,j}$ is the flux measured in that pixel for spectrum i, with corresponding flux uncertainty $\sigma_{i,j}$.

The $\vec{\theta}_{spec,i}$ parameters include a blue- and red-side multiplier to the $LSF_i(\lambda)$, and the slope and intercept for the linear wavelength solution correction of that obser-

vation. Thus, the prior on the spectral parameters for each observation is:

$$p(\vec{\theta}_{spec,i}) = p(\text{blue mult}_i) \cdot p(\text{red mult}_i) \cdot p(m_{\lambda,i}, b_{\lambda,i})$$

where $p(\text{blue mult}_i)$ and $p(\text{red mult}_i)$ are defined in Section 2.3.2 and $p(m_{\lambda,i}, b_{\lambda,i})$ is defined in Section 2.3.2.

The likelihood of a particular spectrum's flux measurement at a particular pixel, $p(f_{i,j}|\vec{\theta}_*, \vec{\theta}_{spec,i})$, comes from a comparison with the synthetic model that has had measured continuum applied. While the continuum we have measured in the previous section does a good job of normalizing each spectrum, we allow for one final continuum fit before evaluating the likelihood to have an optimal comparison. This is particularly important in cases where there is significant line blanketing at blue wavelengths because the original continuum definition may have removed the effects of line blanketing while trying to normalize the fluxes. To this end, for every draw of parameters, we smooth the corresponding synthetic model to the appropriate data quality, then divide it from the continuum-normalized data; we then coadd the remaining noise spectra from the different observations and fit the result with a final B-spline with 100 Å-spaced knots. This process captures any remaining large-scale variations that are required to have good agreement between a particular model and the data.

In words, the parameter measurement proceeds as follows:

1. Draw stellar parameters

   ($T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], M_{814W}, age, phase) and spectral correction parameters (blue mult$_i$, red mult$_i$, $m_{\lambda,i}$, and $b_{\lambda,i}$ for each of the i spectral observations)

using Multivariate Normal proposal distributions;

2. Read in the high-resolution synthetic model defined by the current stellar parameters;

3. For each observation, smooth the synthetic model by $\mathrm{LSF}_i(\lambda < \lambda_{\mathrm{chipgap}})\cdot$blue $\mathrm{mult}_{\mathrm{phase},i}$ on the blue side and $\mathrm{LSF}_i(\lambda > \lambda_{\mathrm{chipgap}}) \cdot$ red $\mathrm{mult}_{\mathrm{phase},i}$ on the red side to get the data-quality model fluxes;

4. Define the new wavelength vector for each observation using the current linear wavelength offset correction, $\Delta\lambda_i(\lambda) = m_{\lambda,i} \cdot \lambda + b_{\lambda,i}$, applied only to the blue side (i.e. $\Delta\lambda(\lambda > \lambda_{\mathrm{chipgap}}) = 0$);

5. Re-sample the corresponding smoothed model onto this new wavelength array, giving $\vec{m}_i$;

6. For each observation, divide the continuum-normalized observation, $\vec{f}_i/\vec{c}_{\mathrm{phase},i}$, by the re-sampled smoothed model, $\vec{m}_i$, to get a noise spectrum that is centered at 1;

7. Coadd the noise spectra together, and fit the result with a B-Spline to account for any missing continuum; this is the continuum adjustment vector $\vec{a}$;

8. For each observation, the likelihood is then

$$p(\vec{f}_i|\vec{\theta}_*, \vec{\theta}_{spec,i}) = \prod_{j=1}^{n_{pix}} \mathcal{N}(f_{i,j}|a_j \cdot c_{\mathrm{phase},i,j} \cdot m_{i,j}, \sigma_{i,j});$$

9. Measure the posterior probability for a current draw using Equation 3.4;

10. Use Metropolis-Hastings criterion to accept or reject the drawn parameters;

11. Repeat until the parameter samples have converged.

We use 500 MCMC walkers that are initialized with parameters drawn from the prior distributions. These walkers are updated to new parameter values at each iteration of the fitting process using the `emcee` package (Foreman-Mackey et al. 2013). We generally require ~ 500 iterations to reach convergence, so we choose to sample for 1500 iterations for each star. We use a conservative burn-in period that throws out the first 70% of samples, keeping the most recent 30% as our posterior samples. Because we need to consider multiple spectral correction parameters for each observation and we consider each observation separately, this process can be quite computationally expensive in both RAM and time. An example of the pipeline's output posterior distributions and model comparison are shown in Figure 2.3 for a star in the COSMOS field; in this case, we see good agreement between the synthetic model and the data[5].

One limitation of our process comes from the assumption that the flux measurements at each pixel within an observation and between observations are uncorrelated, though we know this not to be true because of our preprocessing steps. While a more rigorous fitting approach would incorporate these correlations, we use the uncorrelated assumption because it simplifies the calculations and increases computation speed.

To validate the results of our pipeline, we generate synthetic, HALO7D-like spectral observations with known stellar parameters. The results of analyzing those synthetic observations are shown in Appendix 2.7.1. The main takeaway is that the

[5]The data and model in the "Normalized Flux" panel have been smoothed slightly for the purpose of visual comparison. The data used by the chemistry pipeline and shown in the "Flux Residual" spectrum and histogram are the original, unsmoothed data.

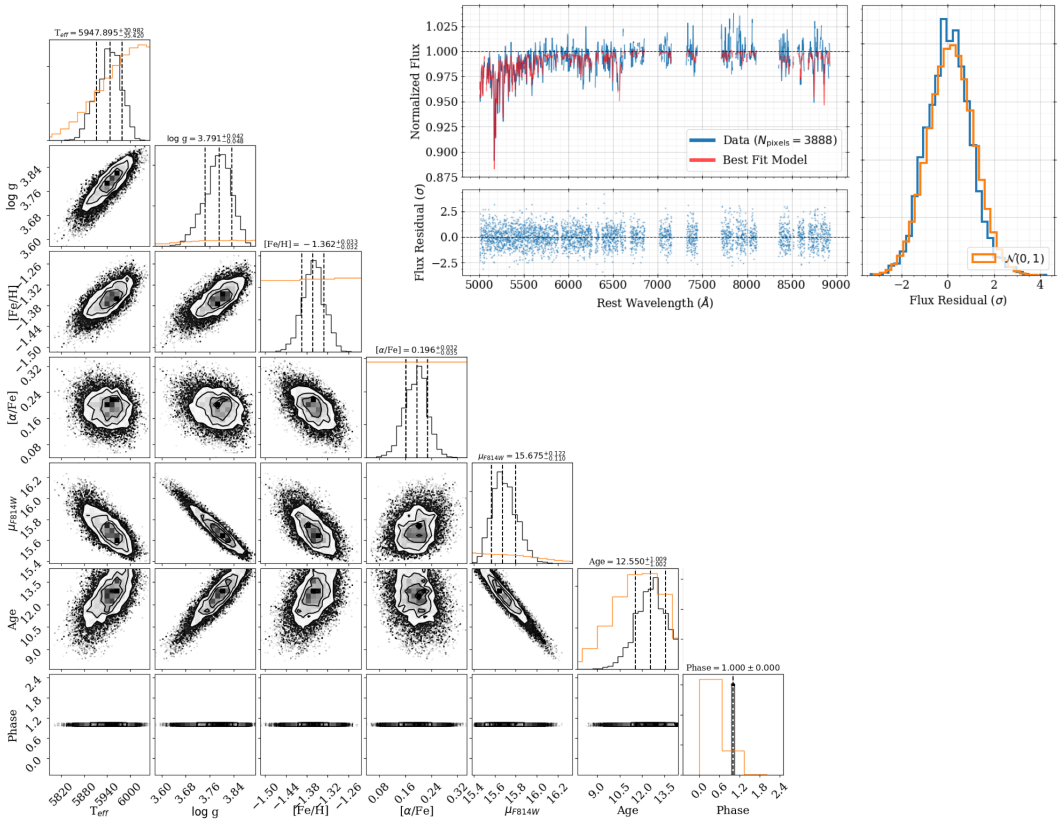

Figure 2.3: Posterior distributions and model comparison for a star in the COSMOS field. The black points, lines, and contours in the corner plot on the left show the posterior samples, and the orange histograms show 1d projections of the prior for a given parameter. The titles above each histogram show the median and the 68% confidence interval in that parameter. In the upper right corner, the star's spectrum shows close agreement with the best fit model (red line) as defined by the median parameters of the corner plot. As described in Section 3.3, various wavelength regions of the data have been masked out for the fitting process, including the central region of Hα. The uncertainty-scaled flux residuals are shown in the lower panel, and the distribution of these residuals show good agreement with the expected unit normal in the rightmost panel.

abundances agree with the known parameters for stars with combined spectral SNR> 20 Å^{-1}. For SNR< 20 Å^{-1}, we begin to see a bias in the abundances, so we omit any stars with combined spectral SNR below this limit for the following analyses. We also find that our posterior abundance distributions are slightly too narrow, requiring an inflation of the posterior covariance by a factor of 1.31^2. This factor implies that our posterior abundance distributions have widths that are approximately 1.31 times smaller than needed to explain the disagreement with the expected values; this implies that our systematic uncertainties in abundances are approximately 31% of the uncertainty reported by the pipeline. All abundance uncertainties shown or reported in this chapter have been inflated by this factor. Finally, we analyze MSTO stars in the well-studied globular clusters of M2 and M92 in Appendix 2.7.2 and show that the pipeline is able to recover results that are consistent with the literature.

2.4 Chemical Abundances of HALO7D Stars

In this section, we present our chemical abundance measurements for the HALO7D dataset. Of the 199 stars from HALO7D with measured 3D velocities (which we hereafter refer to as the "Velocity" sample), 113 had converged posterior results for their chemical abundances (hereafter the "Chemistry" sample). The numbers per field are summarized in Table 2.5 and the abundances are displayed in Figure 2.4. There are several reasons why a star did not converge in the abundance pipeline, but the two main factors are (1) the resulting posterior distribution peaking too close to the edge of the model grid in at least one of the stellar parameters (i.e. $T_{\rm eff}$, $\log g$, [Fe/H],

Table 2.5: Summary of HALO7D targets in each field for different measurements.

Field	$N_{\mathrm{v_{3D}}}$ [a]	$N_{\mathrm{abundances}}$ [b]
COSMOS	81	36
GOODSN	32	21
GOODSS	20	16
EGS	66	40
TOTAL	199	113

[a] Targets that have measured v_{LOS} and proper motions from Cunningham et al. (2019a) and Cunningham et al. (2019b).

[b] Targets that have measured [Fe/H] and [α/Fe].

and α), and (2) having a combined spectral SNR$<$ 20 Å^{-1} which makes constraining the wavelength solution difficult and generally leads to unconstrained posterior distributions. The latter reason effectively acts as a magnitude cut, which makes the faint magnitude limit $m_{F606W} = 23$ mag for EGS and GOODSN, 22.5 mag for COSMOS, and 22 mag for GOODSS, instead of the $m_{F606W} = 24.5$ mag of the original/complete HALO7D sample. As a result, we expect that COSMOS and GOODSS cover slightly nearer distances when compared to EGS and GOODSN.

Figure 2.4 shows the median [Fe/H] and [α/Fe] and corresponding posterior uncertainty for individual stars colored by the field they belong to. The errorbars are perpendicular lines that show the eigenvectors of the posterior covariance of the [Fe/H] and [α/Fe] distribution, with the length corresponding to 1-σ in those eigenvectors. The bulk of stars lie in the $-2 <$ [Fe/H] < -1 region, and the vast majority of stars are supersolar in [α/Fe] as we would expect for old, halo populations and has been seen by previous studies (e.g. SDSS with Carollo et al. 2007, 2010, H3 with Conroy et al. 2019b,a; Naidu et al. 2020, APOGEE with Helmi et al. 2018; Mackereth et al. 2019). GOODSN (in green) has abundance distributions with smaller dispersions compared to the other

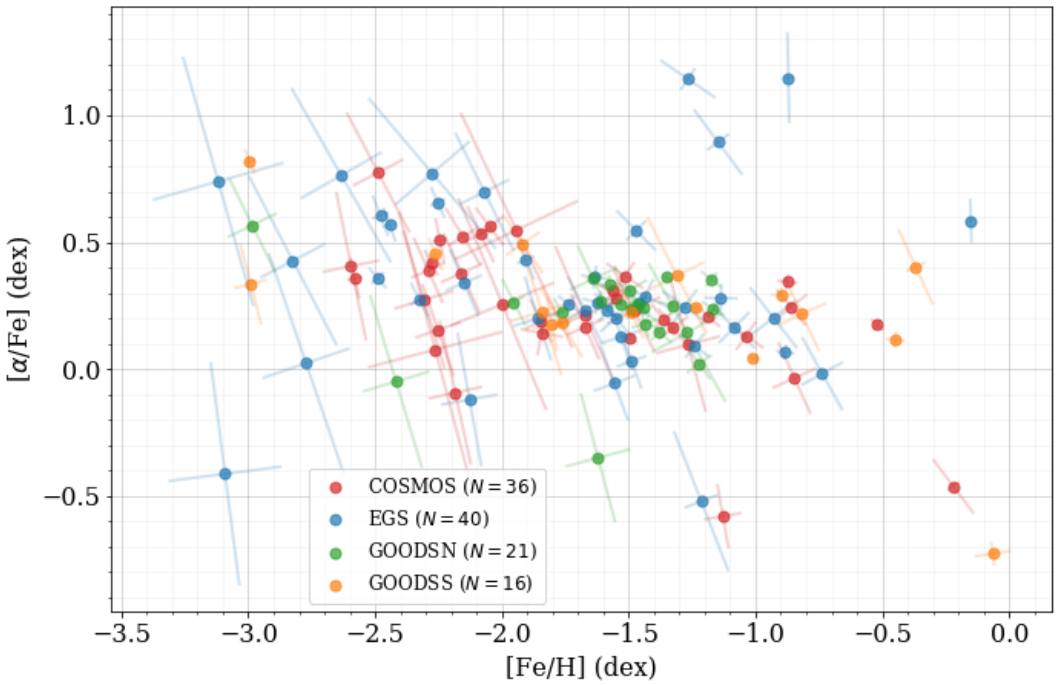

Figure 2.4: Posterior chemical abundances of HALO7D stars colored by the field to which they belong. The errorbars are perpendicular vectors that show the 1-sigma eigenvectors of the posterior covariance matrices.

fields; this is partly because GOODSN stars generally have higher average SNR spectra than the other fields (as seen in Figure 2.2), and thus have smaller posterior abundance distributions, though it could also be that the GOODSN field is less chemically diverse than the other fields. As we will discuss in Section 2.5, the kinematics and abundances of the GOODSN stars are consistent with originating almost exclusively from the GSE progenitor, whereas the other fields appear to have multiple contributions. Excluding the clustering of GOODSN's around [Fe/H] $= -1.5$ dex, the stars of each field occupy the same regions of abundance space, and there isn't an immediately obvious difference between the fields when considering chemistry alone.

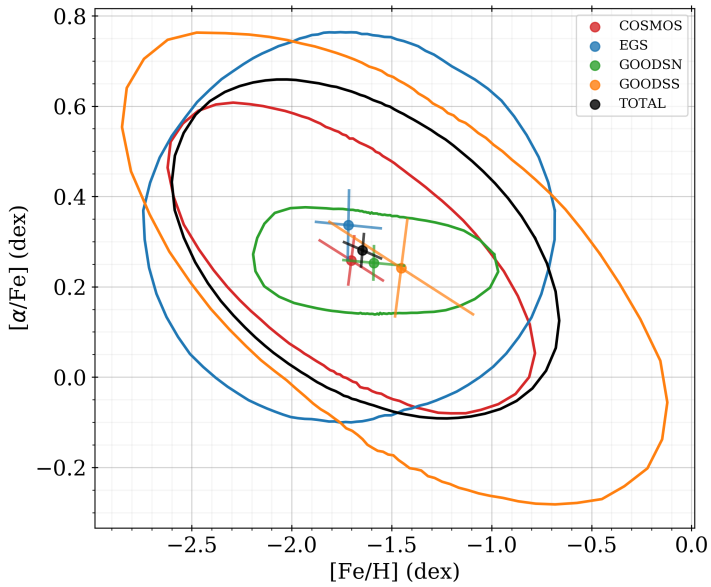

Figure 2.5: Resulting population distribution of abundances for the HALO7D fields and total Chemistry sample after fitting 2D multivariate Gaussians. The colored data points show the median abundances, and the lines coming off of those points are perpendicular 1-sigma eigenvectors showing the uncertainty in that median. The approximately elliptical shapes show the region that the model predicts contains 68% of the population data; these shapes are generated by taking samples of the population medians and covariance matrices from the 2D Gaussian fitting process.

To compare the distributions of [Fe/H] and [α/Fe] in each field more quantitatively, we fit 2D multivariate normal distributions in a hierarchical model to each field's abundances independently. The abundances in each sample are assumed to follow a 2D Gaussian distribution whose mean and covariance we are measuring, and each star's posterior abundances are assumed to be draws from that population distribution. These results are shown in Table 2.6 (labelled as the "Chemistry" sample) and Figure 2.5. The field-colored datapoints show the median of the 2D distributions, the crosses show the 1-σ uncertainties on the median, and the approximately elliptical shapes show the median

region that the model predicts to contain 68% of data. These 2D fits show that all the fields have similar median abundances and uncertainties. For the total sample, we find $\langle$[Fe/H]$\rangle = -1.65 \pm 0.06$ dex, $\sigma_{\rm [Fe/H]} = 0.64 \pm 0.05$ dex, $\langle[\alpha/{\rm Fe}]\rangle = +0.28 \pm 0.03$ dex, and $\sigma_{[\alpha/{\rm Fe}]} = 0.24 \pm 0.02$ dex. As we show in Section 2.5, the fraction of disk contamination in each field and for the total sample is relatively small, so the impact of foreground disk stars on these abundance distributions is minimal.

Recent studies, such as the H3 survey (Conroy et al. 2019b,a), have found a flat halo MDF (with respect to Galactocentric radius) that is peaked at -1.2 dex, has an approximate scatter of 0.5 dex, and has a fairly significant tail towards low [Fe/H]. The H3-measured MDF is generally more metal-rich than previous studies have measured. As Conroy et al. (2019a) point out, many of the previous studies imposed [Fe/H] cuts or targeted blue stars to minimize contamination from nearby MS stars, which biases the MDF towards the metal-poor end (e.g. Carollo et al. 2007; Xue et al. 2015; Das & Binney 2016), while H3 does not select targets based on color. The HALO7D survey also targets blue stars, which may explain why we find a more metal-poor average of -1.65 dex for our sample.

We estimate the impact of our selection function on the HALO7D results in Figure 2.6. For the "No Selection" model, we take MIST isochrones and weight them by

$$[{\rm Fe/H}] \sim \mathcal{N}(-1.2, 1.0 \text{ dex})$$

$$\text{age} \sim \mathcal{N}(12, 2 \text{ Gyr})$$

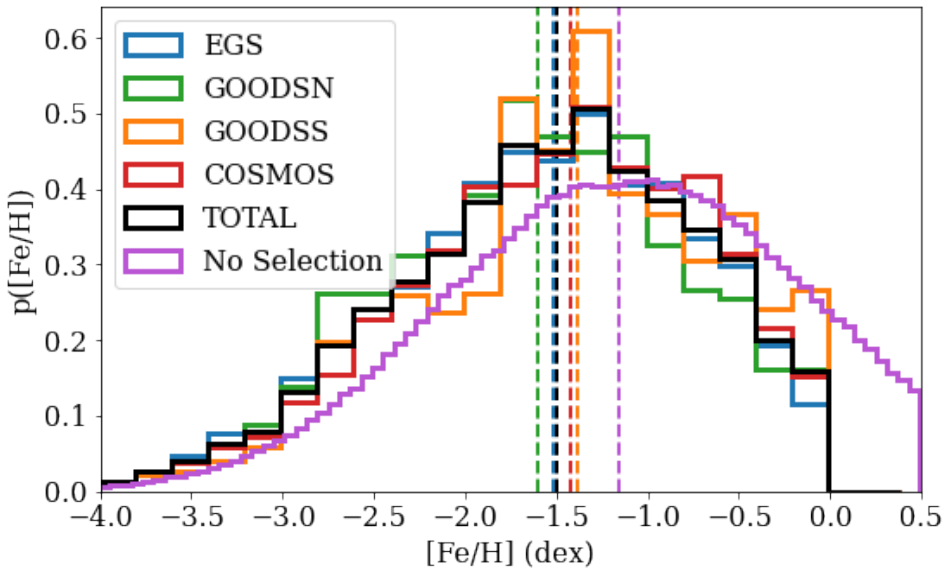

Figure 2.6: Resulting changes to the underlying MDF (purple) when the effective selection functions in each HALO7D field are used. The vertical dashed lines show the median of each distribution. HALO7D, like other surveys that target blue stars, results in a bias towards lower [Fe/H].

$$\text{mass} \sim \text{Kroupa IMF}$$

which produces the purple histogram[6]. We then take those weighted isochrones and apply the observed HALO7D selection functions for each of the fields. This includes masking points outside of the model grid in Table 2.4, as well as ensuring that the resulting distributions in m_{f606W} and m_{F814W} are exactly what is observed in the HALO7D Chemistry sample, producing the other histograms. The dashed vertical lines show the locations of each distribution's median [Fe/H]. From this, we see that the HALO7D cuts on color, magnitude, and stellar parameters bias the measured MDF

[6]While our choice of $\sigma = 1.0$ dex in the [Fe/H] distribution is larger than the scatter measured by Conroy et al. (2019a), we choose to use this simplified model as a reasonable approximation that produces a significant number of low-[Fe/H] stars.

towards lower [Fe/H] compared to the underlying model population, and the median is shifted by ~ -0.3 dex. When accounting for this bias, our median [Fe/H] of -1.65 ± 0.06 dex with a scatter of 0.64 ± 0.05 dex is consistent with the H3 mean halo metallicity of -1.2 dex.

Table 2.6: Summary of abundances for each field, and for different subsamples of the data. Measurements report the 16-th, 50-th, and 84-th percentiles. The [Fe/H] bins are split at -2.0 dex and -1.1 dex based on median posterior [Fe/H]. The missing subsamples of GOODSS Low [Fe/H], GOODSN High [Fe/H], and GOODSN LOW [Fe/H] are omitted because they have too few stars ($N \leq 3$) for a useful analysis.

Field	Sample	$N_{\rm stars}$	$\langle$[Fe/H]$\rangle$	$\sigma_{\rm [Fe/H]}$	$\langle[\alpha/{\rm Fe}]\rangle$	$\sigma_{[\alpha/{\rm Fe}]}$
COSMOS	Chemistry	36	$-1.70^{+0.11}_{-0.09}$	$0.61^{+0.08}_{-0.07}$	$+0.26 \pm 0.04$	$0.23^{+0.04}_{-0.03}$
	High [Fe/H]	6	-0.72 ± 0.15	$0.35^{+0.22}_{-0.10}$	$+0.08 \pm 0.15$	$0.34^{+0.20}_{-0.10}$
	Mid [Fe/H]	16	$-1.52^{+0.06}_{-0.07}$	$0.24^{+0.07}_{-0.05}$	$+0.17^{+0.06}_{-0.07}$	$0.20^{+0.07}_{-0.05}$
	Low [Fe/H]	14	-2.28 ± 0.05	$0.17^{+0.05}_{-0.04}$	$+0.44 \pm 0.04$	$0.09^{+0.05}_{-0.04}$
GOODSN	Chemistry	21	$-1.59^{+0.10}_{-0.09}$	$0.42^{+0.09}_{-0.07}$	$+0.26 \pm 0.02$	0.07 ± 0.02
	Mid [Fe/H]	19	-1.48 ± 0.05	$0.21^{+0.05}_{-0.03}$	$+0.25 \pm 0.02$	0.07 ± 0.02
GOODSS	Chemistry	16	$-1.43^{+0.22}_{-0.24}$	$0.91^{+0.20}_{-0.13}$	$+0.24 \pm 0.09$	$0.35^{+0.08}_{-0.06}$
	High [Fe/H]	6	-0.60 ± 0.20	$0.47^{+0.25}_{-0.15}$	$+0.05^{+0.23}_{-0.21}$	$0.52^{+0.32}_{-0.15}$
	Mid [Fe/H]	7	$-1.61^{+0.14}_{-0.13}$	$0.33^{+0.15}_{-0.09}$	$+0.26^{+0.06}_{-0.05}$	$0.08^{+0.10}_{-0.05}$
EGS	Chemistry	40	$-1.72^{+0.10}_{-0.11}$	0.67 ± 0.08	$+0.33 \pm 0.05$	0.29 ± 0.04
	High [Fe/H]	6	$-0.78^{+0.18}_{-0.21}$	$0.45^{+0.23}_{-0.14}$	$+0.33^{+0.25}_{-0.23}$	$0.54^{+0.36}_{-0.19}$
	Mid [Fe/H]	20	$-1.50^{+0.06}_{-0.05}$	$0.23^{+0.05}_{-0.04}$	$+0.30 \pm 0.07$	$0.30^{+0.08}_{-0.05}$
	Low [Fe/H]	14	$-2.45^{+0.08}_{-0.09}$	$0.30^{+0.10}_{-0.07}$	$+0.43^{+0.07}_{-0.10}$	$0.23^{+0.11}_{-0.08}$
TOTAL	Chemistry	113	-1.65 ± 0.06	$0.64^{+0.05}_{-0.04}$	$+0.28^{+0.02}_{-0.03}$	0.24 ± 0.02
	High [Fe/H]	18	-0.70 ± 0.08	$0.35^{+0.07}_{-0.06}$	$+0.15^{+0.10}_{-0.09}$	$0.39^{+0.09}_{-0.07}$
	Mid [Fe/H]	62	-1.51 ± 0.03	0.22 ± 0.02	$+0.24 \pm 0.03$	$0.19^{+0.03}_{-0.02}$
	Low [Fe/H]	33	$-2.43^{+0.05}_{-0.06}$	$0.30^{+0.05}_{-0.04}$	$+0.46 \pm 0.04$	$0.15^{+0.04}_{-0.03}$

2.5 Chemodynamics with HALO7D

2.5.1 Anisotropy Parameter, β

The power of a survey like HALO7D comes from being able to consider multiple dimensions together. We do this by computing field-averaged kinematics in the form of the anisotropy parameter:

$$\beta = 1 - \frac{\langle v_\phi^2 \rangle + \langle v_\theta^2 \rangle}{2\langle v_r^2 \rangle},$$

as was used in the analysis of Cunningham et al. (2019b). With this definition, $\beta = 1$ implies that a population is on entirely radial orbits, $\beta = 0$ is for a population with isotropic orbits, and $\beta \to -\infty$ for a population with entirely circular orbits. We calculate anisotropies, net halo rotation ($\langle v_\phi \rangle$), and the fraction of disk contamination (f_{disk}) for our four fields and the total sample using the HALO7D stars that have chemical abundance measurements. This is done by modeling the spatial densities, MDFs, age distributions, and velocity component distributions for both the disk and the halo.

We first generate a "Thick Disk" and a "Halo" absolute magnitude probability distribution for each star. As in Section 2.3.1, we do this by weighting MIST isochrones using a Kroupa IMF, an age distribution, an [Fe/H] distribution, each star's de-reddened color, and a density profile for either the Thick Disk or MW stellar halo. These absolute magnitude distributions give distributions on distances that we use in combination with the previously-measured PMs and LOS velocities to calculate the kinematics for each field. For both the disk and halo models, we use the posterior distributions in [Fe/H] and age for each star as measured from the abundance pipeline; because of this, the

Table 2.7: Disk model distributions. For measuring kinematics from the real HALO7D data, we use the posterior [Fe/H] and age distributions as measured using the chemistry pipeline. The disk's density profile distribution is chosen to be the same as used in the analysis of Cunningham et al. (2019b).

Distribution	Functional Form
$p(\text{mass}\|[\text{Fe/H}], \text{age})$	Kroupa (2001) IMF
$p(\mu = m - M)$	$\propto D^3 \exp\left(-\frac{R_D}{3\ \text{kpc}} - \frac{z}{1\ \text{kpc}}\right)$, where $R_D^2 = x^2 + y^2$
$p(\vec{M}\|[\text{Fe/H}], \text{age}, \mu) = \prod_f p(M_f\|[\text{Fe/H}], \text{age}, \mu)$	$\prod_f \mathcal{N}(M_{f,[\text{Fe/H}],\text{age}}\|m_f - \mu, \sigma_{m_f})$

Table 2.8: Velocity prior distributions for Disk and Halo model. These are the same distributions used in the analysis of Cunningham et al. (2019b).

Component	Distribution	Functional Form
Halo	$p(v_r)$	$\mathcal{N}(0, \sigma_r)$
	$p(v_\phi)$	$\mathcal{N}(\langle v_\phi \rangle, \sigma_\phi)$
	$p(v_\theta)$	$\mathcal{N}(0, \sigma_\theta)$
Disk	$p(v_{R_D})$	$\mathcal{N}(0, 45 \text{ km s}^{-1})$
	$p(v_z)$	$\mathcal{N}(0, 20 \text{ km s}^{-1})$
	$p(v_T)$	$\mathcal{SKN}(\mu = 242 \text{ km s}^{-1}, \sigma = 46.2 \text{ km s}^{-1}, a = -2)$

isochrone weighting – and therefore the distance distribution and corresponding velocity measurements – for each star depends on the abundances we've measured. For the disk isochrone weighting, we use the distributions shown in Table 2.7; the density profile is chosen to match that of Cunningham et al. (2019b). For the halo isochrone weighting, we use the same priors as in Table 2.2.

When these absolute magnitude distributions are combined with the LOS velocities and proper motion measurements of Cunningham et al. (2019a,b), we have 3D velocities and positions for each star when belonging to the disk or halo, which allow us to measure the properties of the halo's velocity ellipsoid. For the velocity components, we assume the distributions shown in Table 2.8, where $\mathcal{SKN}$ is a skew-normal distribution with shape parameter a. As before, these distributions are chosen to match those of Cunningham et al. (2019b). To transform between the observer frame and a Galactocentric one, we use $r_\odot = 8.5$ kpc, assume a circular speed of 235 km s^{-1}, and solar peculiar motion $(U, V, W) = (11.1, 12.24, 7.25)$ km s^{-1} (Schönrich et al. 2010). Our calculations use a right-handed coordinate system such that $\langle v_\phi \rangle < 0$ corresponds to prograde motion.

We fit for the unknown halo velocity components of $\langle v_\phi \rangle$, σ_{v_r}, σ_{v_θ}, σ_{v_ϕ}, which

are used to calculate the anisotropy parameter β, choosing a uniform prior on $\langle v_\phi \rangle$, and non-informative priors that are proportional to $1/\sigma_{r,\phi,\theta}$ for the dispersion components. Because each star has the possibility of belonging to the disk or the halo, we fit the population as a mixture model using f_{disk}, which is the fraction of contamination by foreground disk stars. We note that the velocity ellipsoid parameters (i.e. $\langle v_\phi \rangle$, σ_{v_r}, σ_{v_θ}, σ_{v_ϕ}) mentioned here and shown in the following tables and figures refer only to the halo population; we fix the parameters of the disk model velocity distributions for each LOS to those shown in the bottom of Table 2.7 based on the Besançon disk model (Robin et al. 2003). During the fitting process, we marginalize over the absolute magnitude distributions of each star to account for uncertainties in distance. The anisotropy results for various HALO7D subsets are listed in Table 2.9. One key takeaway is that the disk contamination is quite low in all the HALO7D fields for the chemistry sample, so we are measuring the properties of a fairly clean halo sample. The small possible number of disk stars in our sample is thus unlikely to bias the average abundances we present in Section 2.4.

Table 2.9: Summary of anisotropy, distance, and Galactocentric radius for each field. Measurements report the 16-th, 50-th, and 84-th percentiles, and we use the convention that $v_\phi < 0$ corresponds to prograde rotation. The subsamples are the same as listed in Table 2.6.

Field	Sample	$\langle v_\phi \rangle$ (km s^{-1})	σ_ϕ (km s^{-1})	σ_θ (km s^{-1})	σ_r (km s^{-1})	$\langle D \rangle$ (kpc)	D Range (kpc)	$\langle r \rangle$ (kpc)	r Range (kpc)	f_{disk}	β
COSMOS	Chemistry	-66^{+12}_{-13}	71^{+10}_{-8}	69^{+9}_{-7}	106^{+14}_{-12}	7.3 ± 0.3	$3.4 - 13.3$	$13.3^{+0.3}_{-0.2}$	$10.4 - 18.5$	$0.02^{+0.03}_{-0.02}$	$+0.36^{+0.17}_{-0.21}$
	High [Fe/H]	-86^{+56}_{-48}	96^{+57}_{-31}	44^{+22}_{-16}	98^{+49}_{-26}	$9.6^{+1.2}_{-0.7}$	$5.0 - 17.1$	$15.2^{+1.1}_{-0.6}$	$11.5 - 21.9$	$0.18^{+0.24}_{-0.13}$	$-0.13^{+0.75}_{-1.73}$
	Mid [Fe/H]	-37 ± 17	58^{+14}_{-10}	54^{+11}_{-9}	116^{+25}_{-18}	$8.6^{+0.8}_{-0.6}$	$4.6 - 13.7$	$14.4^{+0.6}_{-0.5}$	$11.2 - 18.8$	$0.04^{+0.06}_{-0.03}$	$+0.69^{+0.12}_{-0.18}$
	Low [Fe/H]	-81 ± 19	71^{+18}_{-13}	91^{+22}_{-15}	106^{+23}_{-17}	$5.7^{+0.3}_{-0.2}$	$2.4 - 8.8$	12.0 ± 0.2	$9.7 - 14.5$	$0.04^{+0.07}_{-0.03}$	$+0.07^{+0.35}_{-0.53}$
GOODSN	Chemistry	-28 ± 14	55^{+12}_{-10}	75^{+16}_{-13}	144^{+27}_{-21}	$6.4^{+0.4}_{-0.3}$	$3.8 - 17.6$	$12.2^{+0.3}_{-0.2}$	$10.4 - 22.0$	$0.10^{+0.08}_{-0.05}$	$+0.76^{+0.09}_{-0.13}$
	Mid [Fe/H]	-34 ± 14	53^{+13}_{-10}	59^{+17}_{-13}	152^{+31}_{-22}	$6.4^{+0.4}_{-0.3}$	$3.6 - 14.3$	$12.2^{+0.3}_{-0.2}$	$10.3 - 18.9$	$0.10^{+0.08}_{-0.06}$	$+0.83^{+0.07}_{-0.11}$
GOODSS	Chemistry	-68^{+14}_{-13}	47^{+13}_{-9}	136^{+31}_{-23}	121^{+28}_{-20}	$6.0^{+0.4}_{-0.3}$	$2.3 - 13.1$	$12.3^{+0.3}_{-0.2}$	$9.7 - 18.4$	$0.15^{+0.10}_{-0.07}$	$+0.11^{+0.34}_{-0.55}$
	High [Fe/H]	-96^{+28}_{-25}	46^{+38}_{-19}	223^{+126}_{-68}	124^{+64}_{-36}	8.3 ± 0.5	$1.4 - 14.6$	14.4 ± 0.4	$9.2 - 19.7$	$0.36^{+0.18}_{-0.16}$	$-1.20^{+1.42}_{-3.63}$
	Mid [Fe/H]	-36 ± 20	42^{+20}_{-14}	55^{+21}_{-13}	124^{+44}_{-28}	$7.5^{+0.8}_{-0.5}$	$4.7 - 11.6$	$13.5^{+0.6}_{-0.4}$	$11.3 - 17.1$	$0.10^{+0.14}_{-0.08}$	$+0.77^{+0.12}_{-0.24}$
EGS	Chemistry	-31 ± 13	78^{+10}_{-9}	69^{+9}_{-8}	137^{+17}_{-14}	8.7 ± 0.3	$4.6 - 17.1$	$12.5^{+0.3}_{-0.2}$	$9.9 - 19.5$	$0.04^{+0.04}_{-0.02}$	$+0.68^{+0.08}_{-0.11}$
	High [Fe/H]	$+5^{+17}_{-16}$	35^{+20}_{-14}	35^{+16}_{-10}	99^{+40}_{-24}	$11.3^{+1.0}_{-0.9}$	$7.6 - 20.9$	14.6 ± 0.8	$11.7 - 23.0$	$0.09^{+0.14}_{-0.07}$	$+0.85^{+0.09}_{-0.22}$
	Mid [Fe/H]	-32 ± 9	34^{+9}_{-7}	49^{+11}_{-8}	132^{+24}_{-19}	$8.8^{+0.4}_{-0.5}$	$5.6 - 13.9$	12.6 ± 0.3	$10.4 - 16.7$	$0.03^{+0.05}_{-0.02}$	$+0.86^{+0.05}_{-0.07}$
	Low [Fe/H]	-46^{+38}_{-35}	126^{+32}_{-23}	102^{+26}_{-19}	165^{+40}_{-28}	$5.7^{+0.3}_{-0.2}$	$2.9 - 18.5$	$10.5^{+0.2}_{-0.1}$	$9.1 - 20.8$	$0.13^{+0.11}_{-0.08}$	$+0.43^{+0.23}_{-0.38}$
TOTAL	Chemistry	-47 ± 7	70^{+6}_{-5}	80^{+6}_{-5}	125^{+9}_{-8}	$7.6^{+0.3}_{-0.2}$	$3.5 - 14.8$	12.7 ± 0.2	$10.0 - 19.0$	0.04 ± 0.02	$+0.57^{+0.07}_{-0.08}$
	High [Fe/H]	-53^{+23}_{-22}	78^{+20}_{-15}	106^{+25}_{-18}	101^{+23}_{-16}	$10.5^{+1.0}_{-0.7}$	$4.6 - 16.1$	$14.9^{+1.0}_{-0.7}$	$11.0 - 20.8$	$0.18^{+0.11}_{-0.09}$	$-0.05^{+0.41}_{-0.62}$
	Mid [Fe/H]	-35 ± 7	47 ± 5	52^{+6}_{-5}	130^{+13}_{-11}	$8.2^{+0.4}_{-0.3}$	$4.8 - 13.7$	$13.0^{+0.3}_{-0.2}$	$10.6 - 18.3$	$0.02^{+0.03}_{-0.02}$	$+0.82^{+0.04}_{-0.05}$
	Low [Fe/H]	-63 ± 17	94^{+14}_{-11}	104^{+14}_{-12}	132^{+19}_{-15}	$5.0^{+0.3}_{-0.2}$	$2.6 - 16.4$	11.1 ± 0.3	$9.2 - 19.2$	$0.05^{+0.05}_{-0.03}$	$+0.30^{+0.19}_{-0.25}$

2.5.2 Anisotropy and Abundances

Figure 2.7 reveals that the HALO7D fields have significant differences in their kinematics, as was first seen in Cunningham et al. (2019b). We note that we are considering a smaller subset of the HALO7D stars than was used in Cunningham et al. (2019b), so we will not directly compare our measured kinematics. With the chemistry sample of HALO7D, we find that GOODSN and EGS are more radially-biased than GOODSS and COSMOS.

We also notice that all of the fields have significant net halo rotation, different than the small or nearly zero net-rotation measured by Cunningham et al. (2019b) and others (e.g., Belokurov et al. 2020; Bird et al. 2021). We also note a small apparent correlation between f_{disk} and anisotropy. These may be due to distance systematics (Schönrich et al. 2011) or our choice to model the velocity ellipsoid with a single Gaussian. Lancaster et al. (2019) find the halo velocity ellipsoid is well described by two components, one non-rotating and radially-biased and a second non-rotating, more isotropic population. We use a single component as more robust choice for our smaller sample size. As a consequence, we focus our analysis on the trends in $\langle v_\phi \rangle$ between different subsamples.

To assess the statistical significance of the differences we see between the fields, we assume a null hypothesis that the fields have the same halo velocity ellipsoid parameters that we find for the total population. After generating many simulations of HALO7D-like data and measuring their anisotropies, we indeed find the differences we see between the four fields are not likely to be explained by chance alone. A description

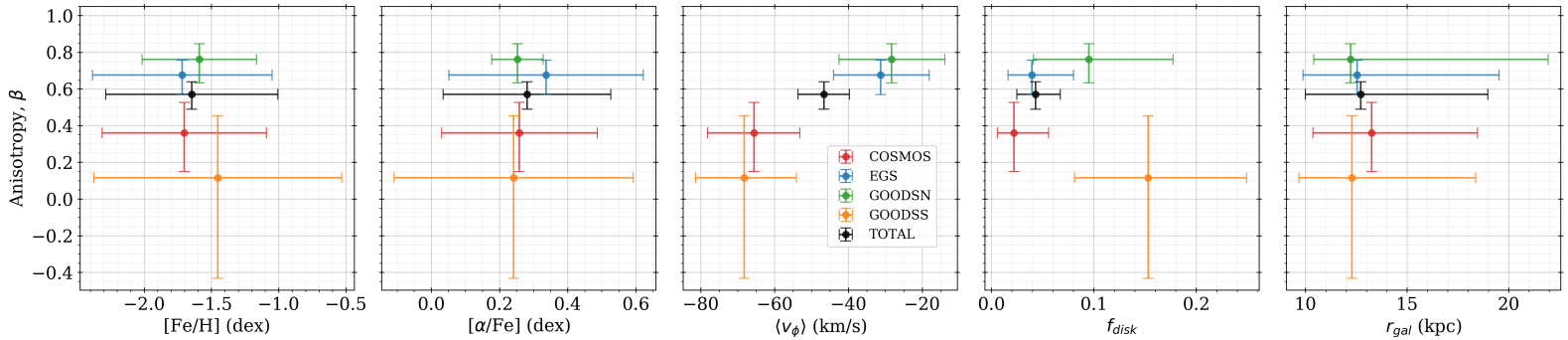

Figure 2.7: Summaries of field distributions in [Fe/H], [α/Fe], net halo rotation $\langle v_\phi \rangle$, fraction of disk contamination f_{disk}, and Galactocentric radii compared to anisotropy. The abundance distributions are those fit with a 2D multivariate normal as summarized in Figure 2.5 and explained in the text. The "errorbars" in these panels show the median and 68% confidence interval for each distribution being considered. These plots show that the fields begin to separate when we consider kinematics. As described in the text of Section 2.5.1, we model both the disk and halo velocities component distributions, so a large f_{disk} measurement does not directly cause the large prograde rotations we measure in some subsamples. EGS and GOODSN both have very radial anisotropies ($\beta \to 1$) while GOODSS is the field with the most isotropic ($\beta \sim 0$) orbits. All fields cover the same approximate radial extent. EGS and GOODSN have the smallest net rotation and the most radial anisotropies, both of which are signatures of the Gaia-Sausage-Enceladus. COSMOS and GOODSS both have large and negative net rotations and show less radial anisotropies, suggesting that these fields may have a larger contribution from the in-situ halo.

and summary of these tests are presented in Appendix 2.7.3.

These anisotropy differences are also not likely explained by differences in β as a function of average Galactocentric radius; between Galactocentric radii of 10 and 20 kpc, Loebman et al. (2018) finds an increase in median β from ~ 0.5 to ~ 0.6 using simulated MW-like galaxies. This approximate trend of $\Delta\beta/\Delta r = 0.01$ kpc^{-1} is not likely to explain the large anisotropy differences between the HALO7D fields. Similarly, various studies of different stellar populations in the MW stellar halo have found that the anisotropy profile does not change significantly in the 10 to 20 kpc radial range (e.g., Lancaster et al. 2019; Bird et al. 2019, 2021; Liu et al. 2022; Wu et al. 2022) after stars associated with Sagittarius have been removed. In these studies, however, the β measurements come from averages over the sky whereas the HALO7D sample is able to compare along different lines of sight.

Anisotropy variations at different positions in the MW stellar halo have been previously observed. Iorio & Belokurov (2021), for example, measure anisotropies of RR Lyrae in bins of Galactic Z and R (see their Figures 3 and 7). While many of their bins have high anisotropy ($\beta \sim 0.8$), there are also a handful of regions with isotropic measurements ($\beta \sim 0$). It is possible that the more-isotropic HALO7D fields may intersect with some of these regions.

One possible explanation for the variation in anisotropy among the fields is differences in fractional contribution from different MW halo progenitors. For instance, the highly radial EGS and GOODSN samples could indicate that these fields are dominated by stars from the GSE – a progenitor that is marked by stars on strongly radial

orbits with no net rotation (e.g., Belokurov et al. 2020, finds $\langle v_\phi \rangle \sim 0$ km s^{-1}) – while GOODSS and COSMOS have more significant contributions from non-GSE sources. Naidu et al. (2020), for instance, found that the halo fraction contributed by the GSE peaks at $r_{gal} \sim 20$ kpc and that the in-situ halo contribution becomes minimal around the same radius. Many other studies have found similar peaks in the fractional contribution from a radial halo population in the r_{gal} range of 10 to 20 kpc (e.g. Deason et al. 2018; Lancaster et al. 2019; Iorio & Belokurov 2021; Liu et al. 2022). Our HALO7D fields cover the range of radii where the transition between the dominance of the in-situ halo decreases and the dominance of the GSE peaks, so differences in their average kinematics may be particularly sensitive to variation in the fractional contribution from these structures. Because all the fields have similar average chemical abundance patterns, we expect that all the fields have relatively large contributions from the GSE, and that contributions from non-GSE sources likely cover the similar regions of abundance-space, such as Wukong (Naidu et al. 2020; Yuan et al. 2020), Nereus (Donlon et al. 2022; Donlon & Newberg 2023), and the in-situ halo/Splash.

2.5.3 Anisotropy in [Fe/H] Bins

To explore the relationship between kinematics and chemistry, we split the sample up into different [Fe/H] bins. Based on previous MW halo inventory studies, the high [Fe/H] bin ([Fe/H] > -1.1 dex) is where we expect the largest fraction of in-situ halo/disk stars and the mid [Fe/H] bin ($-2 <$ [Fe/H] < -1.1 dex) is where we expect stars from the GSE to dominate. The binned abundances and kinematics are summarized in Tables 2.6 and 2.9 and Figure 2.8.

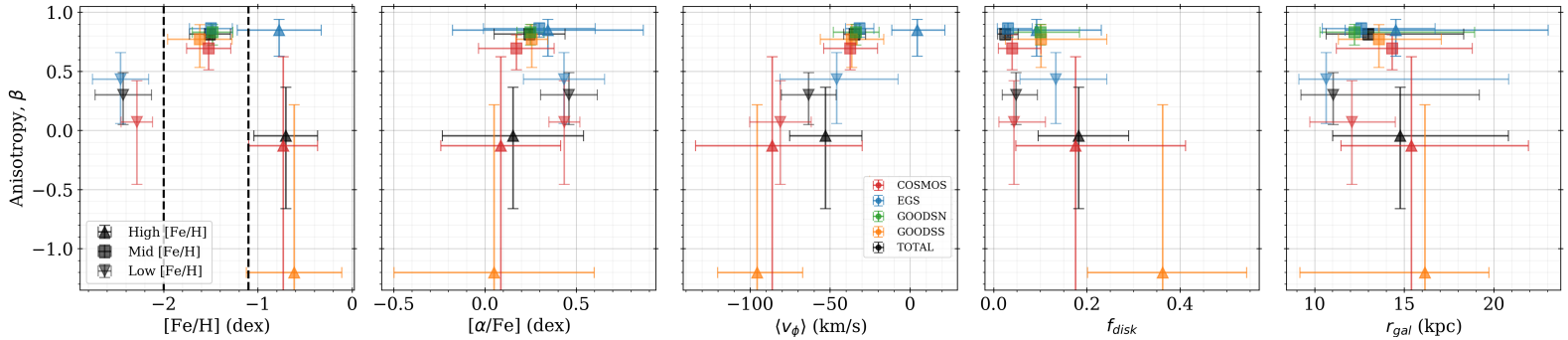

Figure 2.8: Same as Figure 2.7 but for different [Fe/H] bins that are split at −2.0 dex and −1.1 dex. Triangles that point up correspond to the high [Fe/H] bin, squares are the middle [Fe/H] bin, and triangles that point down are the low [Fe/H] bin. The missing [Fe/H] bins for GOODSN and GOODSS are omitted for having too few stars for useful calculations. All of the mid [Fe/H] bins have very radial anisotropies; the stars in this bin have metallicities and kinematics similar to the GSE. COSMOS changes in anisotropy and halo rotation more than EGS does for each of the bins. This suggests that EGS might be dominated by stars from a single progenitor while COSMOS has a higher fractional contribution from multiple progenitors. The GOODSS high [Fe/H] bin has the most circular orbits compared to the other fields, suggesting that this subsample may contain the largest fraction of in-situ halo stars.

All fields in the mid [Fe/H] bin have highly radial anisotropies and a small net halo rotation ($\langle v_\phi \rangle$), which suggests that stars in this bin are chemodynamically similar and thus may originate from a single progenitor. In the low and high [Fe/H] bins, we notice differences in the kinematics between the fields. In particular, the high [Fe/H] bin shows the largest differences in anisotropy between the fields, with EGS having the most radial stars and GOODSS having the most circular stars. As before, we test the probability of observing these results by chance alone and find that the anisotropy differences in the high [Fe/H] bin are likely statistically significant; these tests are summarized in Appendix 2.7.3. These anisotropy differences between the fields may be caused by variations in the fractional contributions of different progenitors, which changes the average kinematics we measure. These differences highlight the additional information available on smaller spatial scales that a survey like HALO7D is able to capture.

The mid [Fe/H] bin covers $-2 <$ [Fe/H] < -1.1 dex in metallicity and spans Galactocentric radii of $\sim 10 - 20$ kpc in all fields. This is the radial and metallicity range that other studies (e.g. Naidu et al. 2020) have measured the GSE to dominate the halo. These facts, combined with the low net rotation and highly radial anisotropy we measure in the mid [Fe/H] bins, are evidence that all the fields have a dominant fractional contribution from the GSE in the $-2 <$ [Fe/H] < -1.1 dex range. Because we measure a non-zero $\langle v_\phi \rangle$ of ~ -30 km s^{-1} in this bin, GSE is almost certainly not the only progenitor present in this bin; however, the relative contributions from prograde structures must be relatively consistent between the different fields to produce a similar

net rotation. The high [Fe/H] bin has the largest range of anisotropies and v_ϕ values compared to the other bins. This is the bin that we expect to contain a significant number of stars from the GSE, the in-situ halo, and a few other known MW progenitors because its higher metallicity intersects with the MDFs of multiple structures. As a consequence, this large number of progenitor options is likely causing the increased variation we see in kinematics in this bin. Like the mid [Fe/H] bin, the low [Fe/H] bin doesn't show much variation between the chemodynamics of the fields.

The results from EGS and COSMOS are particularly useful to compare because they have approximately the same number of stars in each of the [Fe/H] bins, which limits the impact that sample size has on these comparisons. EGS has anisotropies in each [Fe/H] bin that are quite radial and fairly consistent with each other, which suggests that EGS is dominated by stars from a single progenitor in all [Fe/H] bins. As explained in the previous paragraph, the origin of these stars may be the GSE, as the EGS stars show the characteristic radial bias and low net rotation previously measured for the GSE. For COSMOS, the anisotropy and net halo rotation change more noticeably between the [Fe/H] bins. At high and low [Fe/H], the COSMOS anisotropies are isotropic and the net rotations are prograde, which suggests that these bins have significant contributions from sources other than that which produced the kinematics in the mid [Fe/H] bin. Like the other fields, the COSMOS mid [Fe/H] bin has kinematics and chemistry that are consistent with the GSE. A possible origin for the high-[Fe/H], prograde stars in COSMOS is the in-situ halo, while the low-[Fe/H], prograde stars may originate from the Wukong, Nereus, and/or the metal-poor tail of the in-situ halo.

Because COSMOS is closer to the disk plane than any of the other HALO7D fields and EGS is the farthest, the larger contribution of in-situ halo stars in COSMOS and a smaller contribution in EGS is not unexpected. All the HALO7D fields are located far away from any Sagittarius Stream debris, so this is not likely to explain the difference in chemodynamics we observe between the fields.

In summary, the HALO7D fields show significant differences when considering chemistry and kinematics together. Looking at different [Fe/H] bins reveals that all the fields may have a significant fraction of stars from the GSE in the middle [Fe/H] bin of $-2 <$ [Fe/H] < -1.1 dex. EGS has similar kinematic properties in its high and low [Fe/H] bins, and these are again consistent with debris from GSE. COSMOS, on the other hand, has more variation in its kinematics across the [Fe/H] bins which suggests it has larger fractional contributions from non-GSE sources. The high [Fe/H] bin has properties consistent with a large fraction of in-situ halo stars, the mid [Fe/H] bin has properties like the GSE, and the low [Fe/H] bin has properties like the metal-poor tail of the in-situ halo, the Wukong progenitor, and/or the Nereus progenitor. This spatial non-uniformity is compelling evidence that the MW stellar halo is not uniformly mixed in its chemodynamical distributions.

2.6 Summary

We have measured [Fe/H] and [α/Fe] abundances for 113 main sequence turn-off MW stellar halo stars across four CANDELS fields in the HALO7D survey. By focusing on MSTO stars in the stellar halo at Galactocentric radii in the range $10-40$ kpc,

HALO7D is able to measure MW halo properties on smaller spatial scales than other contemporary surveys. Our abundances are combined with the previously-measured 3D positions and 3D velocities from HALO7D (Cunningham et al. 2019a,b) to measure the variation in average chemodynamical properties along each LOS; these properties include the net halo rotation, anisotropy, and average abundances. To measure our abundances for HALO7D, we have created a Bayesian pipeline that uses photometric and spectroscopic information to determine stellar parameters ($T_{\rm eff}$, $\log g$, age, and distance modulus) and chemical composition ([Fe/H], [α/Fe]) for MSTO stars without known distances (Section 3.3). Our key results include:

1. The abundance patterns in each of the HALO7D fields agree with each other. The average [Fe/H] of the full 113 star HALO7D Chemistry sample is -1.65 dex with a scatter of 0.61 dex, which is more metal-poor than some recent contemporary surveys (e.g., $\langle$[Fe/H]$\rangle \sim -1.2$ dex for Conroy et al. 2019a; Naidu et al. 2020), but this is almost certainly because of our blue selection function. The average [α/Fe] for the HALO7D Chemistry sample is $+0.28$ dex with a scatter of 0.24 dex, which is in agreement with what we would expect for a sample drawn from a population of old, metal-poor halo stars. (Section 2.4, Figures 2.4, 2.5)

2. The HALO7D fields separate in kinematic-space when we measure average properties like the anisotropy parameter β. EGS and GOODSN show more radially-biased orbits and near-zero halo rotation compared to GOODSS and COSMOS which have more isotropic orbits and fairly negative halo rotation. (Section 2.5.2, Figure 2.7)

3. Breaking the HALO7D fields into low, mid, and high [Fe/H] bins at -2.0 dex and -1.1 dex shows differences in the chemodynamic makeup of the fields. All the fields have similar anisotropy in the mid [Fe/H] bin, but the high and low [Fe/H] bins show differences between the fields. EGS has relatively similar anisotropy and net halo rotation between the [Fe/H] bins, all of which are similar to the properties of Gaia-Sausage-Encaladus. COSMOS, on the other hand, has a mid [Fe/H] bin with kinematics that are similar to the GSE, but its low and high [Fe/H] bins have kinematics that suggests these bins have larger contributions from prograde structures, such as the kicked-up disk/in-situ halo, the metal-poor progenitors of Nereus and Wukong. These chemodynamical differences between the fields, even at the same Galactocentric radii, suggest that the MW stellar halo is not uniformly mixed along different lines of sight. (Section 2.5.3, Figure 2.8)

Future work will focus on studying the full 7D chemodynamic relationship of our stars, such as using a MW potential model to estimate individual orbits, which will allow us to better quantify the relative contributions from different progenitors in each field. We are also in the process of expanding the HALO7D survey to include more LOS and more stars along each LOS, which will allow us to measure how the chemodynamical distributions change as a function of 3D position with the goal of contributing to a more complete picture of our Galaxy's accretion history.

2.7 Appendices

2.7.1 Chemical Abundance Pipeline Internal Error with Fake Stars

We test the ability of our pipeline to recover input chemical abundances across a range of stellar parameters and spectral SNR by generating fake stars with known photometry and stellar parameters. We create two types of fake stars; those with single spectral observations that model our real observations of globular cluster stars in M2 and M92 (see Appendix 2.7.2), and those with multiple spectroscopic visits that model real HALO7D data. In each case, a fake star consists of (m_{F606W}, m_{F814W}) apparent magnitudes and between one and ten separate spectral observations at a chosen SNR.

In the case of the globular cluster analog fake stars, we use the literature-defined MIST isochrone to define the photometry and stellar parameters. We do this by taking a real globular cluster star, and finding the closest point on the MIST isochrone in terms of photometry; the fake star then inherits that point's absolute magnitudes $(M_{F606W},\ M_{F814W})$ and stellar parameters $(T_{\rm eff},\ \log g)$ as well as the cluster distance, age, and [Fe/H]. The [α/Fe] is randomly drawn from a uniform distribution in the range of -0.2 dex to $+0.6$ dex. With the fake star's stellar parameters and abundances defined, we can smooth the corresponding synthetic model spectrum by the real globular cluster star's observed seeing and LSF, and then re-bin onto the observed DEIMOS wavelength array. We next multiply the smoothed, continuum-normalized fake spectrum by an estimate of the real star's continuum (as described in Section 2.3.2) to obtain DEIMOS-like observations for the fake star. Finally, the chosen SNR of the fake star's spectrum is used to apply flux noise and to define the flux uncertainty in each pixel. With the

fake star having apparent magnitudes and a spectral observation at a chosen SNR, we are able to feed it to the chemical abundance pipeline by following the process listed in Section 3.3.

For the HALO7D-like fake star analogs, we follow a similar process as with the globular cluster stars. To define the input parameters of the fake star, we use the real HALO7D star's posterior distribution from its analysis using the chemical abundance pipeline; the fake star's parameters are chosen to be the median of the posterior distribution, and the apparent magnitudes are exactly those of the HALO7D star. The corresponding synthetic model spectrum is repeated so that the fake star is given the same number of spectroscopic visits/observations as the real HALO7D star that it is based on. The seeing, LSF, and continuum of each real spectroscopic observation are used to define the smoothing and continuum of the different fake spectral visits. Next, a chosen SNR defines the resulting combined SNR of the different spectroscopic visits, where the SNRs of the individual visits are chosen to follow the same ratio of the real HALO7D star; the SNR of each spectroscopic visit sets the flux noise and the flux uncertainty in each pixel. As with HALO7D stars, the apparent magnitudes and multiple spectroscopic visits of the fake stars are processed through the chemistry pipeline.

For each real star that defines a set of stellar parameters and abundances, we create fake stars with varying SNR from 5 Å^{-1} to 200 Å^{-1}. These fake stars are processed identically to the real stars following the methods of Section 3.3 to assess the pipeline's ability to recover the input parameters as a function of spectral SNR. An example posterior distribution for one fake star with spectral SNR of 40 Å^{-1} and

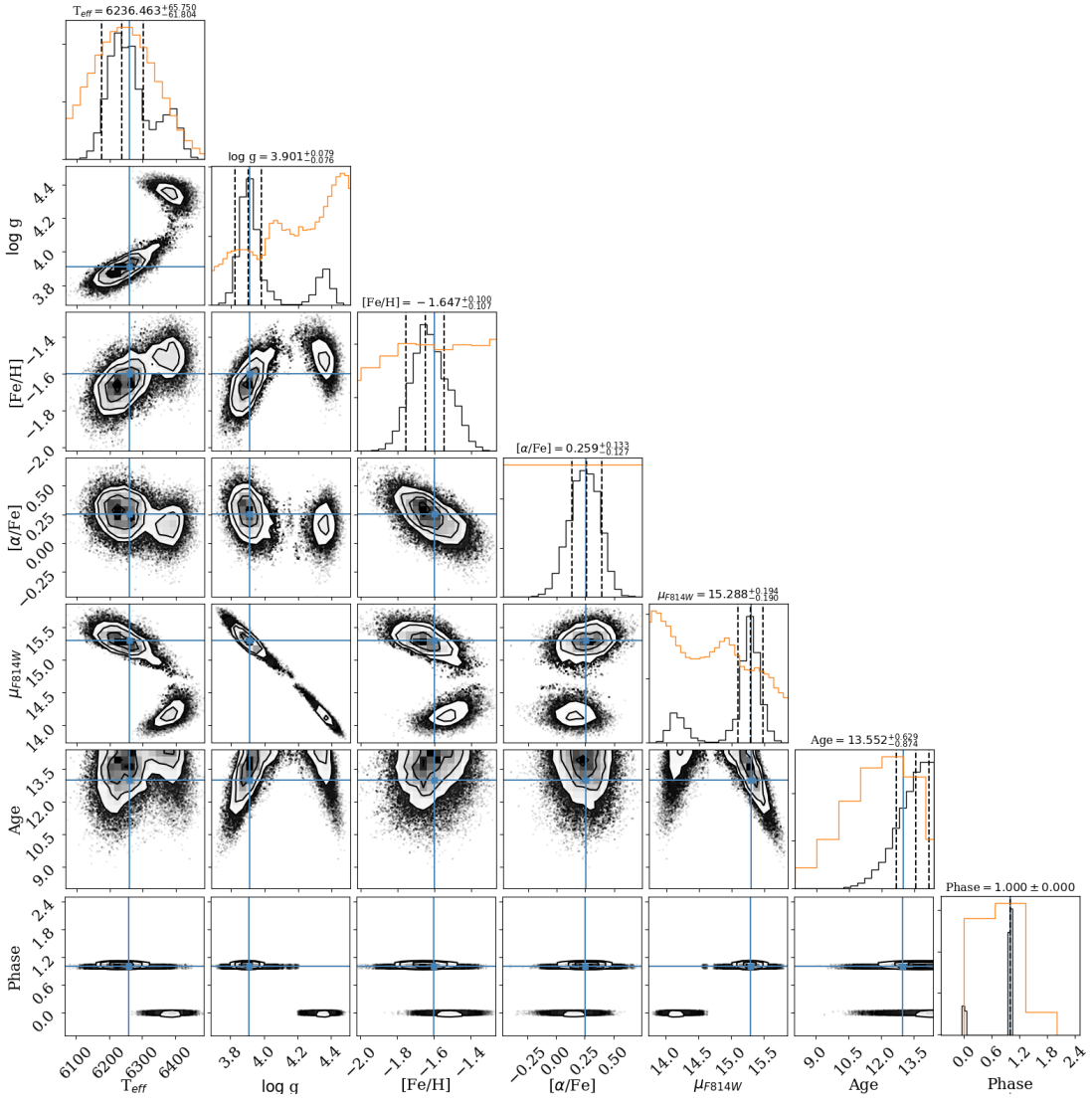

Figure 2.9: Posterior distributions for a fake star that has a single spectral observation with median SNR of 40 $Å^{-1}$ with stellar parameters that are near the MSTO of M2. The blue lines show the location of the fake star's true parameters, the black points/lines/contours show the posterior samples, and the orange histograms show 1d projections of the prior for a given parameter. The titles above each histogram show the median and the 68% confidence interval in that parameter.

properties that place it near the MSTO of M2 (i.e. distance of 11 kpc, age of 13 Gyr, [Fe/H] of -1.65 dex) is shown in Figure 2.9. The black points and lines show the posterior samples that have good agreement with the blue lines which show the values of the input parameters. In cases where the pipeline incorrectly measures $\log g$ – that is, when the star is assigned the incorrect phase, placing it on one branch when it belongs to another – we find that the chemistry is unchanged; specifically, the posterior [Fe/H] and [α/Fe] samples are in agreement for each of the possible phases. Since we are most interested in measuring abundances, we include stars with incorrectly-measured posterior phase/$\log g$.

The posterior chemical abundance distributions of the fake stars are then used to assess the internal errors of our chemistry pipeline and the reasonableness of the posterior uncertainties. The 2D abundance distance between the input truth and the posterior median is measured using the covariance of the posterior [Fe/H] and [α/Fe] samples; these distances should follow a chi-squared distribution with 2 degrees of freedom (DOF) and a scale of 1 if the posterior uncertainties are reasonable. While the distribution we measure does indeed follow a chi-square with 2 DOF, we instead measure a scale factor of 1.31, meaning that our posterior widths need to be inflated by this amount (i.e. the covariance matrix needs to be inflated by 1.31^2) to capture the pipeline's true uncertainty. All remaining stellar parameter measurements in this chapter, including Figure 2.10 and all other figures, have had their posterior uncertainties increased by this factor. After applying this inflation, we still noticed a slight bias in the [Fe/H] and [α/Fe] results for fake stars with spectral SNR < 20 Å^{-1}. As a result,

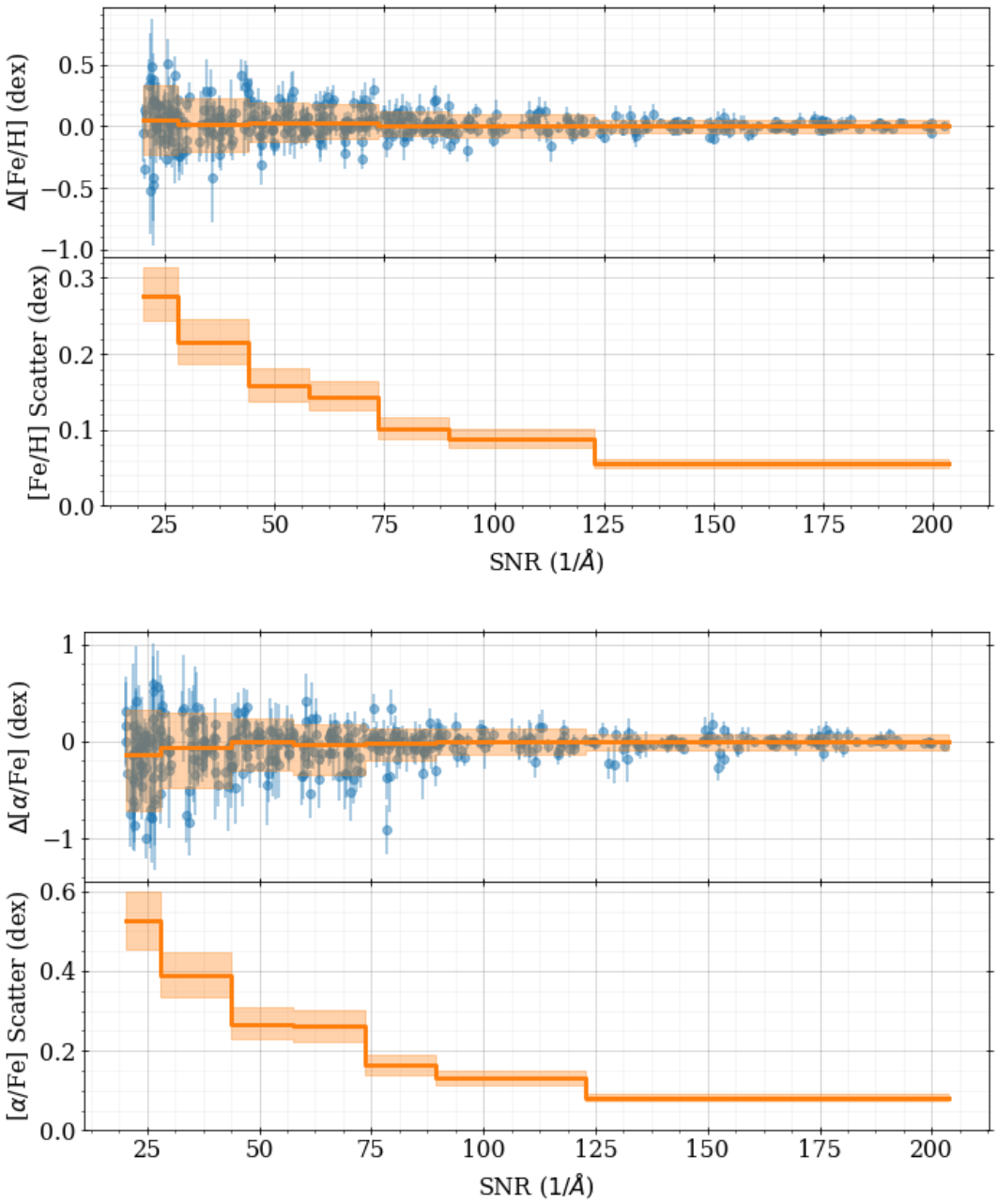

Figure 2.10: Comparison of measured posterior [Fe/H], [α/Fe] and input truth for fake spectra as a function of spectral signal-to-noise. Blue errorbar points show the 68% confidence interval of the posterior around the median. The orange lines in the top panels show the median and 68% scatter within SNR bins of 50 stars and the orange lines in the bottom panels show the median 68% width of the posterior samples (and corresponding 1-sigma uncertainty) in those same SNR bins. We remove observations with spectral SNR < 20 Å^{-1} because the abundance uncertainties begin to show systematic biases at this level.

we ignored real and fake stars with spectral SNR < 20 Å^{-1} from all analyses.

Plotting the difference between the posterior abundances and the input truth versus input spectral SNR, as in Figure 2.10, shows that the pipeline recovers useful (posterior disagreement < 0.25 dex) [Fe/H] above SNR ~ 25 Å^{-1} and [α/Fe] above SNR ~ 45 Å^{-1}. This is similar performance to other spectroscopic analysis pipelines of DEIMOS data (e.g., Escala et al. 2019; Kirby et al. 2008). The uncertainty-scaled difference between truth and posterior abundance is shown in Figure 2.11. These distributions show good agreement with the unit Gaussian (shown in orange), implying that the posterior abundances and uncertainties are reasonable. Our choice of a uniform prior on [α/Fe] that has no correlation with the other stellar parameters can be improved in the future with access to MIST isochrones that contain different values of [α/Fe]; this flat [α/Fe] prior currently plays a part in causing the relatively large scatter on our ability to recover [α/Fe] as compared to [Fe/H].

Our posterior uncertainties are relatively large compared to other pipelines because of our large prior uncertainties in $\log g$, which come from unknown distances. Many other investigations that measure stellar abundances are able to use previously-measured parallaxes or distances to tightly constrain the possible $\log g$ values, which helps to return better-measured abundances from the same spectral SNR and to push the limiting SNR to much lower values. Our necessarily diffuse distance priors cause the pipeline to consider a wider range of models and thus return more uncertain abundances.

While a given star's posterior [Fe/H] and [α/Fe] measurements are correlated (as seen in Figure 2.9, as well as our other posterior distributions), we want to en-

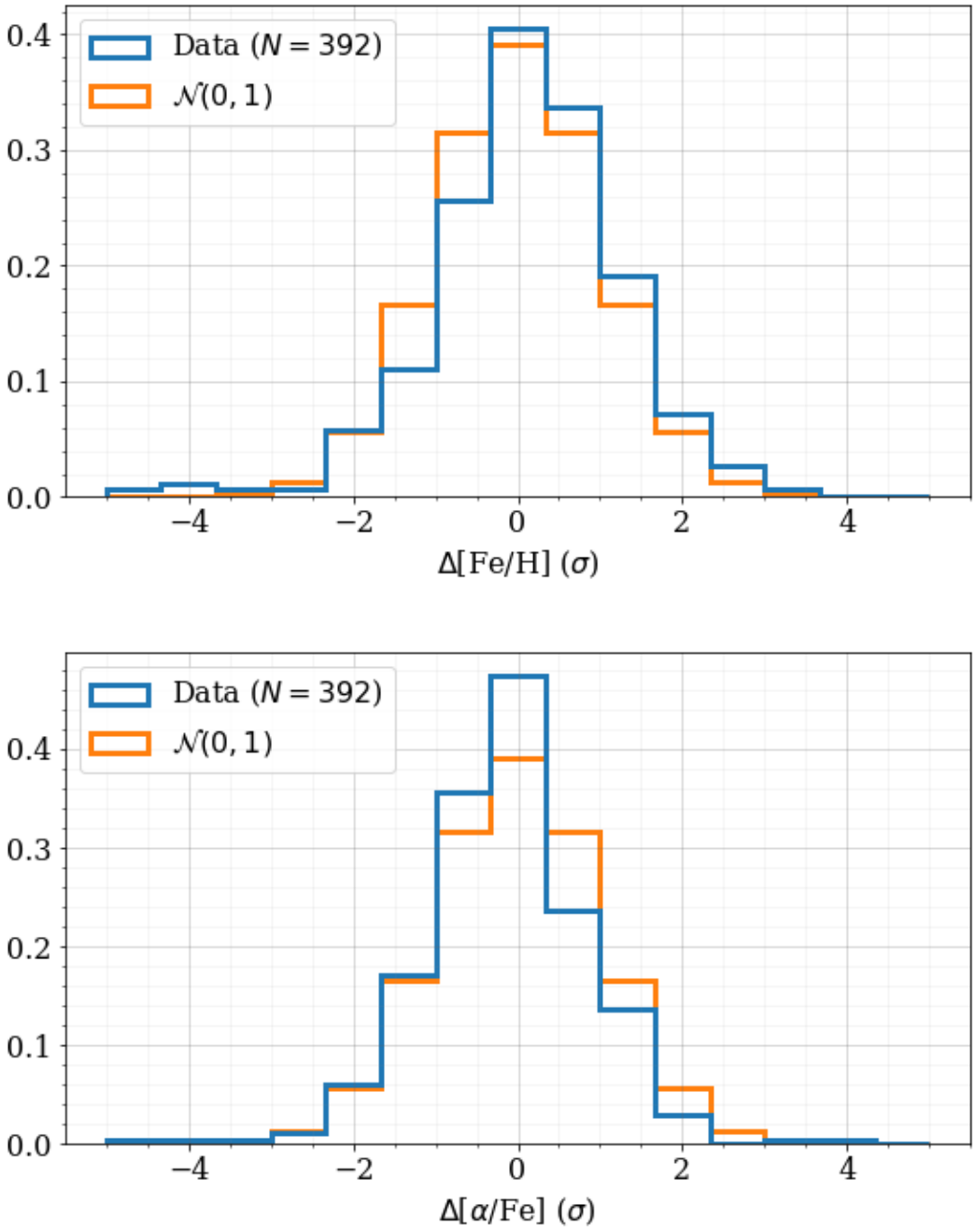

Figure 2.11: Uncertainty-scaled comparison of measured posterior [Fe/H], [α/Fe] and input truth for fake spectra. The unit Gaussian (orange line) shows that the pipeline is returning reasonable abundances and corresponding uncertainties.

sure that the pipeline doesn't produce a correlation in any population statistics (e.g. group means and spreads) measured from a collection of stars. To test this, we fit a hierarchical model using the Δ[Fe/H] and $\Delta[\alpha/\mathrm{Fe}]$[7] posterior distributions of all fake stars to measure a pooled population mean and covariance matrix of a 2D Multivariate Normal. The posterior mean of the population distribution was centered at $(\Delta[\mathrm{Fe/H}], \Delta[\alpha/\mathrm{Fe}]) = (0, 0)$ dex, corresponding spreads of $(0.02, 0.01)$ dex, and a correlation between Δ[Fe/H] and $\Delta[\alpha/\mathrm{Fe}]$ that is consistent with 0. Finally, we also perform a prior sensitivity analysis by changing the parameters of the [Fe/H] and age distributions in Table 2.2 while re-analyzing the fake stars; we find that the posterior chemical abundances are largely unchanged for fake stars with spectral SNR > 20 Å^{-1} when using reasonable choices of the [Fe/H] and age distribution parameters. Therefore, the pipeline does not induce artificial correlations in the population statistics.

2.7.2 Chemical Abundance Pipeline External Validation with Globular Clusters

In Appendix 2.7.1, we used fake stars with well-behaved data to test our chemical abundance pipeline. In this appendix, we assess the pipeline's ability to recover chemical abundances using real data from well-studied globular clusters: M2 and M92. M2 is at a heliocentric distance of 11 kpc and has an iron abundance of -1.65 dex (Harris 1996, 2010 edition), which is relatively typical of MW halo stars. M92 is at a heliocentric distance of 9 kpc and is quite metal poor with an average [Fe/H] of -2.35 dex (Harris 1996, 2010 edition), so it is a fairly extreme test of our pipeline's

[7] Here, $\Delta X = X_{\mathrm{posterior}} - X_{\mathrm{input}}$

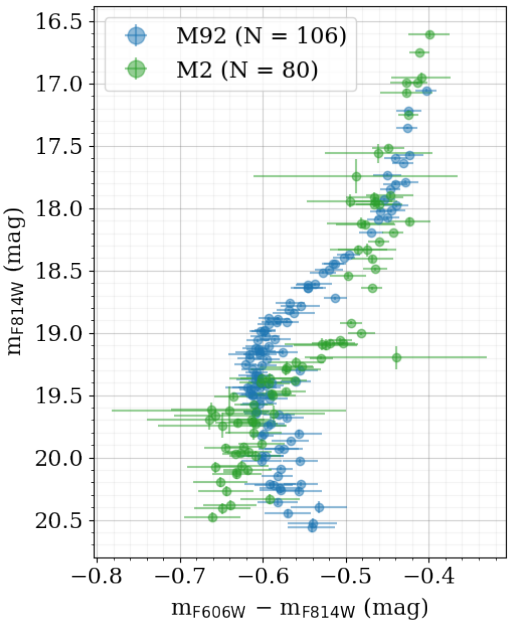

Figure 2.12: Color-Magnitude Diagram of the M2 and M92 globular cluster stars in HST filters (STMAG) used in our validation sample. The HST photometry was calculated by transforming ground-based photometry, which is why the uncertainties are quite large. Both clusters have stars that are around the MSTO.

ability to measure abundances with hot, metal-poor stars on the MSTO. A CMD of the stars used in our analysis is shown in Figure 2.12 to highlight that this sample includes many stars around the MSTO, which makes this a good test for the pipeline's accuracy for HALO7D-like data. The photometry comes from Stetson et al. (2019). The spectral observations have $20 < \mathrm{SNR} < 100$ Å and were observed with Keck II/DEIMOS in a similar configuration to the HALO7D data; see Escala et al. (2019) for details. The main differences between the DEIMOS configurations are that the GC data used a 0.8" slitwidth and a central wavelength of ~ 7500 Å instead of ~ 7200 Å.

To transform the Johnson-Cousins photometry of Stetson et al. (2019), we use the m_u magnitudes and $m_u - m_v$ colors of each cluster and align the literature MIST

isochrone on the CMD data using the appropriate filters. With this, we find the closest matching point on the isochrone to each star, which assigns an estimated mass to each star. Using this mass estimate and the same isochrone in HST filters gives each star's approximate m_{F606W} and m_{F814W} magnitudes.

We use the same prior-building process for the GC stars as for the MW halo stars. That is, the known distances, ages, and abundances of the clusters are not used; instead, we assume that all the cluster stars are distributed like MW halo stars. While this leads to larger posterior uncertainties than we could achieve using the known GC parameters, it gives a better measurement of how well the pipeline will perform with MW halo stars. These GC comparisons also function to show the insensitivity of the pipeline to our choice of priors in Table 2.2; we know the prior distributions do not describe the GC stars very well, so our ability to recover reasonable posterior abundances for stars with spectral SNR > 20 Å^{-1} is not overly dependent on our priors.

For each cluster, we measure a population average and scatter in [Fe/H] and [α/Fe] using the posterior distributions of each star. For M92, we measure an average [Fe/H] of -2.31 dex with a scatter of 0.08 dex and an average [α/Fe] of 0.16 dex with a scatter of 0.10 dex. For M2, we measure an average [Fe/H] of -1.64 dex with a scatter of 0.10 dex and an average [α/Fe] of 0.19 dex with a scatter of 0.09 dex. These means in [Fe/H] are consistent with the literature values quoted above. A significant contribution to the scatter is that we force the priors to contain MW halo-like properties instead of the properties of each cluster. This leads to stars with lower SNR spectra relying on incorrect prior information, thereby increasing the scatter.

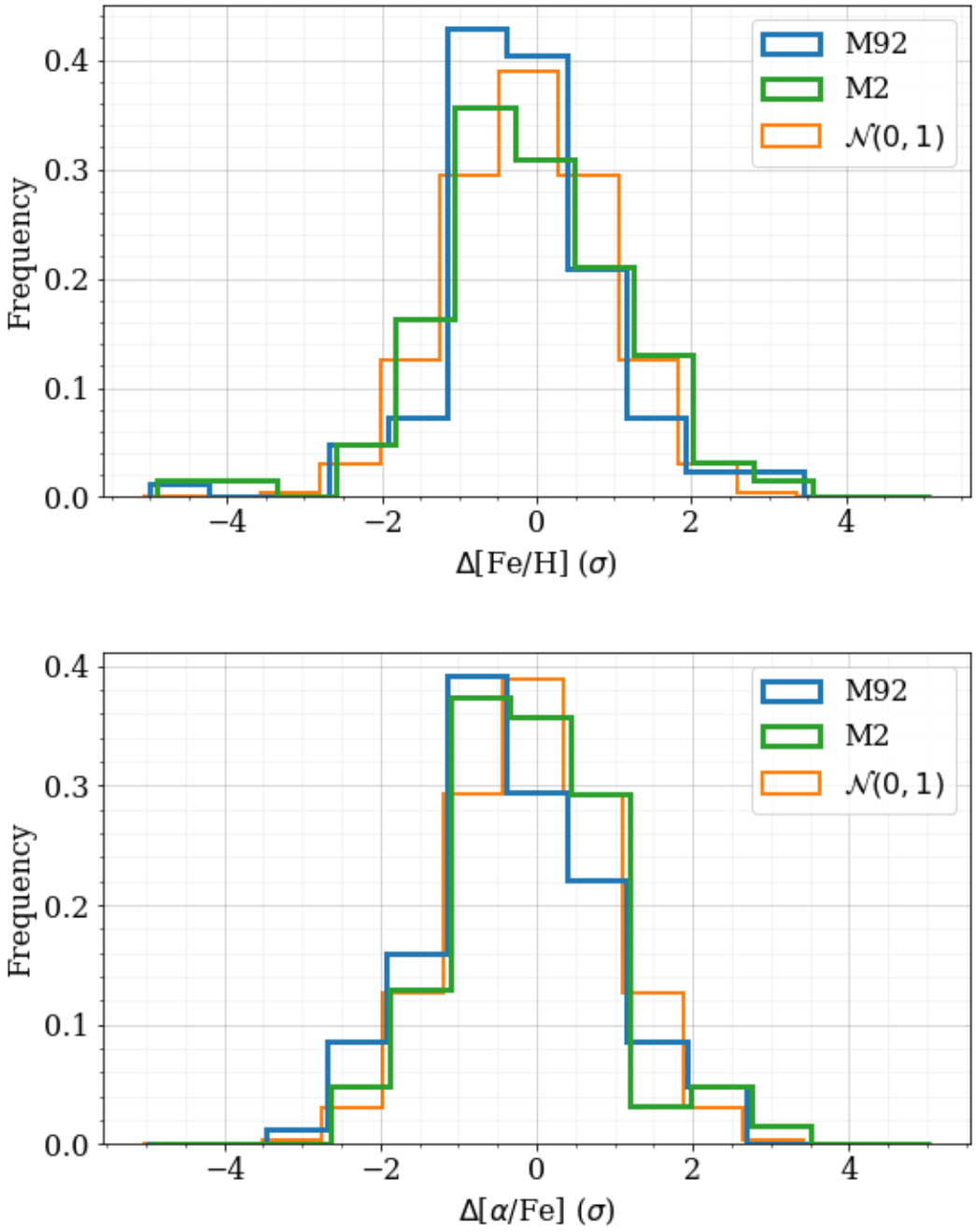

Figure 2.13: Uncertainty-scaled comparison of measured [Fe/H] and [α/Fe] with population averages for stars in M92 and M2. The green lines show that a unit Gaussian has good agreement with our results.

We are most interested in assessing whether the pipeline returns abundance measurements that are unbiased over the HALO7D range of spectral SNR and whether the posterior uncertainties are reasonable. To that end, we compare the chemical abundances to our population averages of [Fe/H] and [α/Fe] for the GCs instead of to literature values. This is shown in Figure 2.13, where the uncertainty-scaled [Fe/H] and [α/Fe] distributions for both clusters agree with the unit Gaussian. This is evidence that the pipeline is giving reasonable abundances for real stars with spectra like those of HALO7D.

In addition to not being MW halo stars, and therefore not truly following the assumptions that go into the priors, the cluster data is dissimilar to HALO7D in a few other ways. First, all the GC data consists of a single spectroscopic observation per star instead of multiple, which means that any systematic errors in an individual spectrum are more likely to impact the results (e.g. poor wavelength solution, bad skyline removal, vignetting). Second, the GC photometry comes from ground-based observations which have significantly larger uncertainties than our HALO7D photometry measured with HST; this large color uncertainty leads to significantly more diffuse priors for the GC stars as compared to HALO7D stars. Third, the GC spectra have wavelength solutions that are generally less well-behaved than HALO7D spectra because the observations were calibrated with a single set of relatively red arc lamp exposures and didn't use a second set of bluer arc lamps as was done for the HALO7D observations. Finally, the DEIMOS configuration for the GC stars is centered on $\sim$ 7500 Å instead of the $\sim$ 7200 Å used in HALO7D, which means that many of the GC spectra do not extend

to wavelengths as blue as $\sim$ 5000 Å where many of the strongest [Fe/H] and [α/Fe] features exist.

2.7.3 Statistical Significance of HALO7D Anisotropy Differences

To test the statistical significance of the differences we see in the anisotropy measurements of the four HALO7D fields, we generate simulated HALO7D-like surveys and measure their kinematics. For each HALO7D-like realization, we assume that each of the four fields has the same number of stars as we observed: 40 in EGS, 36 in COSMOS, 21 in GOODSN, and 16 in GOODSS. For the fraction of disk contamination, we again use the values measured in our HALO7D analysis, such that 2 of the EGS stars are disk contaminants, as are 2 of the GOODSN and GOODSS stars, and 1 of the COSMOS stars; the remaining 106 of the 113 stars are halo stars.

We generate realizations of HALO7D that follow the same magnitude and color limits in each of the four fields. For the disk stars, the [Fe/H]s are drawn from

$$\begin{aligned}\text{[Fe/H]} \sim & \frac{1}{6}\mathcal{SKN}(\mu = -1.05 \text{ dex}, \sigma = 0.6 \text{ dex}, a = -5) \\ & + \frac{5}{6}\mathcal{N}(\mu = -0.54 \text{ dex}, \sigma = 0.3 \text{ dex})\end{aligned} \tag{2.3}$$

the ages are drawn from

$$\text{Age} \sim \mathcal{N}(\mu = 10 \text{ Gyr}, \sigma = 2 \text{ Gyr}) \tag{2.4}$$

and the masses and distance moduli are drawn from the prior distributions shown in Table 2.7. The [Fe/H] distribution comes from an analytical approximation of the

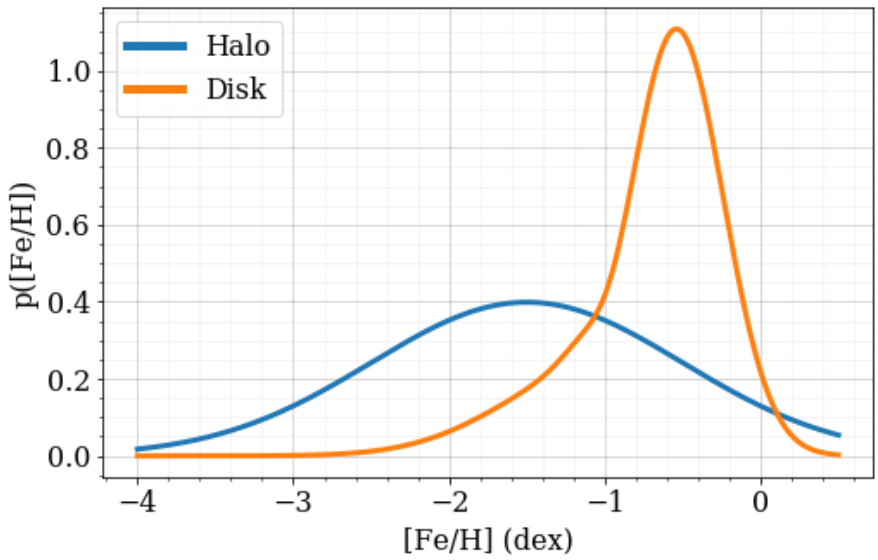

Figure 2.14: Comparison of the chosen prior Halo and Disk model MDFs. The Halo model is a simple Gaussian centered at −1.5 dex with a scatter of 1.0 dex while the Disk model is a mixture of a Gaussian and skew-normal distribution as described in Equation 2.3. The Disk's distribution comes from an analytical approximation of the 'high-alpha disk" and "metal-weak thick disk" populations of Naidu et al. (2020).

"high-alpha disk" and "metal-weak thick disk" populations of Naidu et al. (2020), the age distribution comes from another analytical approximation of the "high-alpha disk" from Bonaca et al. (2020). The resulting prior MDFs for the Halo and Disk models are compared in Figure 2.14.

The velocity components for the disk stars are drawn from the prior distributions in the bottom half of Table 2.8. For the halo stars, the [Fe/H], age, mass, and distance modulus are drawn from the prior distributions shown in Table 2.2. For the kinematics of each halo star, we assume that all four fields have the same halo velocity ellipsoid that matches the values we measure for the total HALO7D population. Specifically, $\langle v_r \rangle = \langle v_\theta \rangle = 0$ km s^{-1}, $\langle v_\phi \rangle = -47$ km s^{-1}, and $(\sigma_r, \sigma_\phi, \sigma_\theta) = (125, 70, 80)$ km s^{-1},

which implies an anisotropy of $\beta = 0.57$. The velocity components for each halo star are drawn from the distributions in the top half of Table 2.8.

With each star having a defined [Fe/H], age, mass, and distance modulus, we use the corresponding MIST isochrone to get the apparent m_{F606W} and m_{F814W} magnitudes, ensuring that the colors and magnitudes are within the limits that we observe for each HALO7D field. The distance to each star is used to transform the velocity components into proper motions and LOS velocities, which means that each fake star has a set of observables that make them comparable to the HALO7D sample. Each proper motion component is given an uncertainty equal to the median uncertainty from the HALO7D sample: $\sigma_{\mu_l \cdot \cos b} = 0.17$ mas year^{-1}, $\sigma_{\mu_b} = 0.16$ mas year^{-1}. For the LOS velocities, we use Figure 7 of Cunningham et al. (2019a), which shows the relationship between $\sigma_{V_{LOS}}$ and a star's apparent m_{606W} magnitude for the HALO7D sample; we capture this relationship as

$$\sigma_{V_{LOS}}(m_{F606W}) = (4 \times 10^{-11} \text{ km s}^{-1}) \cdot \exp\left(\frac{m_{F606W}}{0.86859 \text{ mag}}\right) + 1.5 \text{ km s}^{-1}$$

which gives LOS velocity uncertainties of 1.6 km s^{-1} for $m_{606W} = 19$ mag, 5.5 km s^{-1} for $m_{606W} = 22$ mag, and 41.5 km s^{-1} for $m_{606W} = 24.5$ mag.

As with the HALO7D sample, we generate absolute magnitude prior distributions assuming each star belongs to the Thick Disk and the Halo. With the absolute magnitude distributions giving corresponding distance distributions, we are able to follow the same steps as outlined in Section 2.5, fitting for the halo velocity components ($\langle v_\phi \rangle$, σ_r, σ_ϕ, σ_θ) and thick disk fraction, f_{disk}. We repeat this process for each real-

ization of HALO7D and measure the median β for each of the simulated fields. We keep results from the realizations that had a measured β for the total sample that is within the 68% region of the real data's total sample (i.e. only cases where the fake star total sample's median β agrees closely with the data's $\beta_{\rm TOTAL} = 0.57^{+0.07}_{-0.08}$); for our 200 realizations, we find 129 of the medians fall within this region, which agrees with the expectation from a binomial distribution with $p = 0.68$ and $n = 200$. These results are summarized in Figure 2.15, where the black points show the median β measured for each of the 129 realizations. The dashed colored lines show the input anisotropy, $\beta = 0.57$, while the shaded grey regions show the 68% region of the results we measure for the real HALO7D data. The dashed grey lines and grey Xs show the median β we measure for the real HALO7D data.

From the 129 simulated HALO7D realizations with total β in agreement with the input value, we find only 3 cases that are similar to the results we measure from the real HALO7D data. In particular, there are 3 realizations where the measured β is within the 68% shaded regions in all four fields. The fractions of medians within each field's 68% region separately are approximately 31%, 26%, 48%, and 38% for GOODSS, GOODSN, COSMOS, and EGS respectively. This works out to a 1.5% probability of seeing data in all four shaded regions at the same time from chance alone, which is similar to the 3/129 realizations we measure in this region. From this analysis, we reject the null hypothesis that the anisotropy differences we see between the four HALO7D fields are likely to be explained by statistical chance.

We repeat this analysis technique for the different [Fe/H] bins. We do this

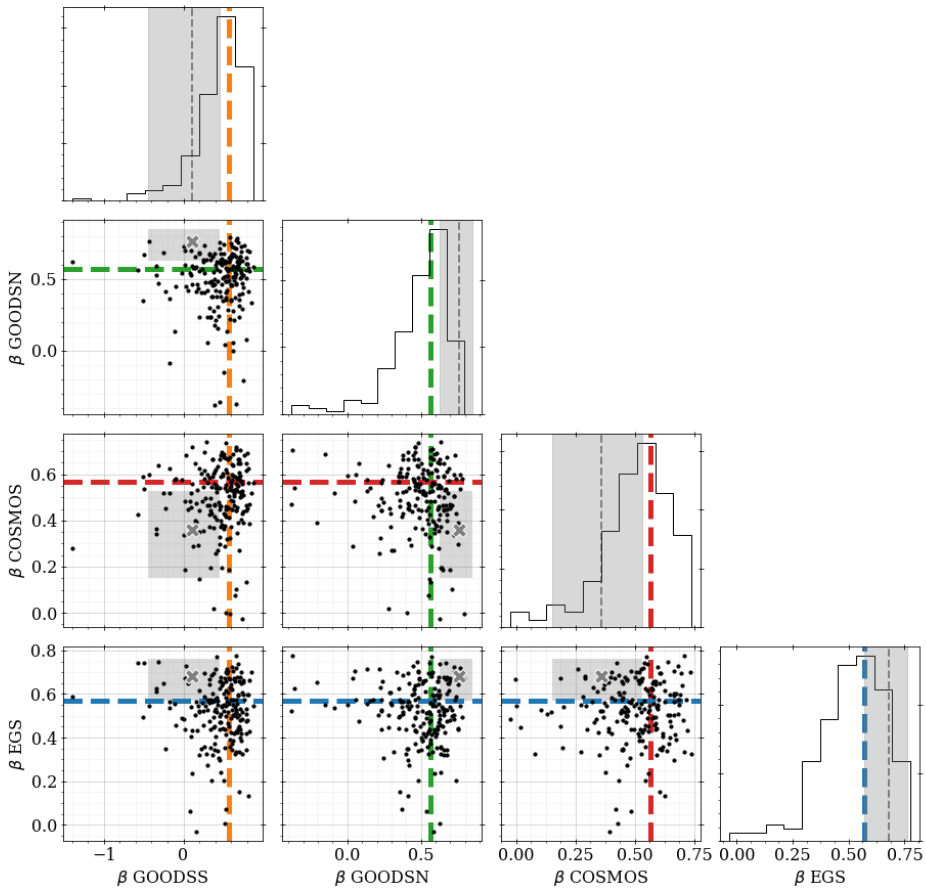

Figure 2.15: Distributions of median anisotropy β for 129 HALO7D-like realizations for each field. The black points show the posterior median β for each realization, and the histograms on the diagonal show their 1D distribution. The dashed vertical colored lines show the input β that was used to generate the data. The grey dashed lines in the histograms and the grey Xs in the scatter plots show the posterior median β we measure for each field in the real HALO7D data. The grey-shaded regions show the 68% region of the β distribution for the real data. There are only 3 black points that are within the 4D hypercube defined by the grey-shaded region. This implies that the differences in β that we measure for the real data are not likely explained by random chance.

by again assuming that the input velocity components are those measured using the total population in the real HALO7D data in that [Fe/H] bin (i.e. the values reported in Table 2.9). We again create realizations of HALO7D data, this time drawing the expected number of stars in each of the [Fe/H] bins so that the number of stars in each field and each bin matches the totals shown in Table 2.6. As before, we keep only the β medians from the realizations that agree with the 68% region of the β_{TOTAL} for that [Fe/H] bin; this works out to 140, 139, and 130 of the 200 total realizations for the high, mid, and low [Fe/H] bins respectively. For the high [Fe/H] bin case, the β medians of the realizations are shown in Figure 2.16. We find that β medians fall in all the grey shaded regions only 3.3%, 18%, and 35% of the time for the high, mid, and low [Fe/H] bins respectively; as with the real HALO7D data, we omit the GOODSS low [Fe/H] bin measurements as well as the GOODSN high and low [Fe/H] bin measurements for having too few stars. From these probabilities, we can see that the anisotropy distributions we measure in the high [Fe/H] bin aren't likely to be produced by random chance alone, meaning the GOODSS, COSMOS, and EGS fields likely have different average halo properties in this [Fe/H] bin. For the mid and low [Fe/H] bins, these differences are not as statistically significant. In the case of the mid [Fe/H] bin, this is in line with our expectations because the fields have similar β measurements in the HALO7D data, suggesting that the velocity distributions of the stars in this [Fe/H] bin are more similar between the fields. For the low [Fe/H] bin, these results tell us that the differences in anisotropy we see between COSMOS and EGS in this bin could be produced by chance alone approximately $\sim 2/5$ of the time.

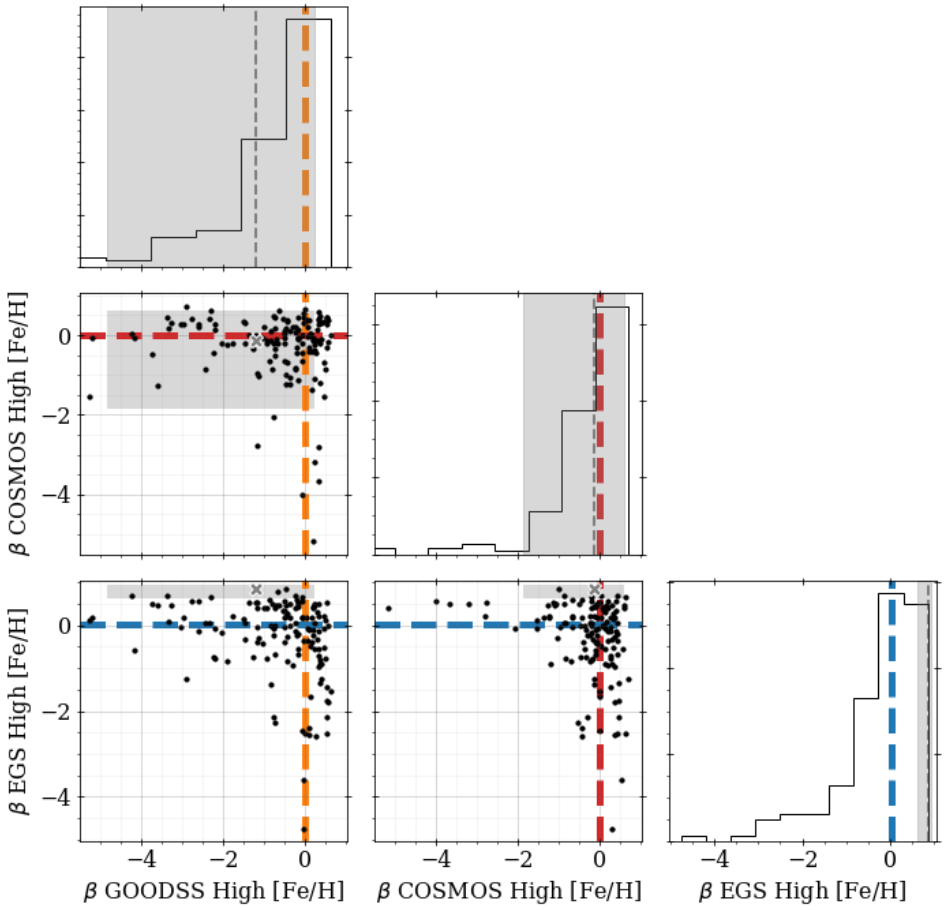

Figure 2.16: Same as Figure 2.15, but for 140 realizations in the high [Fe/H] bin. The GOODSN panels are omitted because they had too few stars in this [Fe/H] bin for a useful analysis. There are only 6 black points that are within the 3D hypercube defined by the grey-shaded region in this figure. This implies that the differences in β that we measure for the real data are not likely explained by random chance.

Chapter 3

BP3M: Bayesian Positions, Parallaxes, and Proper Motions derived from the *Hubble Space Telescope* and *Gaia* data

3.1 Introduction

High precision proper motions (PMs) of individual stars have dramatically increased our understanding of kinematics in the Local Group. In particular, the *Hubble Space Telescope* (*HST*) has a rich history of providing the precise astrometry needed

to make accurate motion measurements. This is thanks to detailed work over the last few decades to characterize and correct *HST* geometric distortions, describe point spread functions (PSFs), and define best practices for mapping images onto one another (e.g. Anderson & King 2004, 2006; Anderson 2007; Bellini & Bedin 2009; Bellini et al. 2011). As a result, the PMs measured from multi-epoch *HST* imaging have been used to study key astronomical questions, such as the relative motion of the Milky Way (MW) and M31 (Sohn et al. 2012; van der Marel et al. 2012), the mass of the MW (Sohn et al. 2018), and measuring MW stellar halo kinematics along individual sight-lines (Cunningham et al. 2019b).

More recently, the *Gaia* mission (Gaia Collaboration et al. 2018) has facilitated substantial leaps in Local Group science by providing full astrometric solutions for millions of stars. In the MW, these results include the identification of our most massive merger, the Gaia-Sausage-Enceladus (e.g Helmi et al. 2018; Belokurov et al. 2018; Mackereth et al. 2019), detailed inventories of the progenitors that built the MW (e.g. Naidu et al. 2020), and nearly-complete catalogs on the motion of nearby globular clusters (Vasiliev & Baumgardt 2021).

While both of these telescopes have and will continue to fundamentally alter our knowledge about the local universe, they both have limitations. In *HST*'s case, the small field of view and relatively long observation times mean that only a small portion of the sky has been observed at multiple epochs with significant time baselines. For *Gaia*, their all-sky catalog necessarily does not see as faint as a standard *HST* observation, and the precision of their astrometric solutions decreases fairly significantly

for their faintest magnitudes ($G \sim 21$ mag). *Gaia*'s fairly bright limiting magnitude for stars with precise PMs restricts the spatial resolution on which it can measure average kinematics in the stellar halo. Consequently, many of the previous studies cited above focus on stellar populations over large portions of the sky when measuring the kinematics of the MW stellar halo, but a growing body of work has shown the benefit of being able to resolve (chemo)dynamics on small spatial scales (e.g. Cunningham et al. 2019b; Iorio & Belokurov 2021; McKinnon et al. 2023). With better sky velocities, we can improve constraints on the formation and evolution history of our Galaxy (e.g. Cunningham et al. 2022, 2023, Apfel et al. in prep).

To reduce the effect of these limitations and to increase the astrometric constraining power of either telescope alone, recent studies have been exploring how to combine information between datasets from different telescopes. Specifically, using archival *HST* images that have long time baselines with *Gaia*, PMs that are factors of 10 to 20 times more precise than *Gaia* alone have been measured for dwarf spheroidal galaxy and globular cluster stars (Massari et al. 2017, 2018, 2020; del Pino et al. 2022), which have enabled internal velocity dispersion measurements. Techniques for cross-telescope combinations will become even more important as the field progresses further into the Big Data era of astronomy, especially as future missions come online (e.g. *Roman Space Telescope*).

Regardless of where the astrometry comes from, PMs in the MW can be more challenging to measure than PMs of distant sources. This is largely because the apparent motion on the sky can be quite large and the motion from parallax can become

significant. There are three main ways that a star can appear to move between successive images: (1) apparent offset because of statistical uncertainty on position, (2) offsets due to parallax, and (3) offsets due to proper motion. In studies of distant sources, the motion from parallax can often be ignored. In many cases with *Gaia* data, the uncertainty on position is usually small enough that it can also be ignored. For distant stars, then, all apparent motion is the result of proper motion, and the longer the time baseline between observations, the more precise and accurate the PM measurements. When the position uncertainties become large (e.g. for faint *Gaia* sources) and the parallaxes become substantial (e.g. $D < 1$ kpc), then detailed and simultaneous accounting of all three motion components becomes necessary. One additional difficulty in measuring MW PMs, especially in studies of the stellar halo, comes from the fact that many lines of sight are quite sparse; this limits the number of sources that can be used to constrain the transformation parameters that align images from multiple epochs and impacts the accuracy and precision of the resulting PMs.

While the `GaiaHub` pipeline of del Pino et al. (2022) was developed for and performs well in populated fields, a key motivation of this work was to develop a complementary pipeline that can also handle very sparse fields (e.g. $N_* < 10$). To address the challenges listed previously, we create a hierarchical Bayesian pipeline, named `BP3M` (Bayesian Positions, Parallaxes, and Proper Motions), to simultaneously measure the positions, parallaxes, and proper motions of all *Gaia* sources in an *HST* image while also finding the best mapping of *HST* images onto *Gaia*. This package is publicly available and is designed to be used in combination with the `GaiaHub` code. The underlying

statistics are general in that they apply to any two or more sets of position measurements separated by time, regardless of the telescope they come from. In this way, it is also a useful tool for planning future observations.

In this chapter, we describe a standard approach for measuring PMs in Section 3.2, which then motivates the statistics and pipeline, `BP3M`, we present in Section 3.3. We examine and validate the pipeline's outputs in Section 3.4, after which we run `BP3M` on real, sparse fields in the MW in Section 3.5. In Section 3.6, we explore the variability of the transformation parameters between *HST* and *Gaia* as a function of time. Next, we generate synthetic *Roman*-like observations in a sparse MW field to estimate the impact of combining *Roman* and *Gaia* measurements in Section 3.7. Finally, our complete set of results are summarized in Section 4.6.

3.2 Measuring Proper Motions

To measure the motion of stars, one traditionally identifies the positions of sources at two epochs, measures the best transformation parameters between those two images, and transforms the coordinates of sources in one image into the other. The final offsets can then be used to measure the relative motion of each of the sources. When mapping one image onto another, a standard approach (Anderson 2007) is to describe the position of a source in the first image, (X, Y), in coordinates of the second image (X', Y') using:

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \left[\begin{pmatrix} X \\ Y \end{pmatrix} - \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} \right] + \begin{pmatrix} W_0 \\ Z_0 \end{pmatrix} \tag{3.1}$$

where the (X_0, Y_0) and (W_0, Z_0) vectors are the center of rotation in each coordinate system. The (a, b, c, d) matrix accounts for differences in pixel scale, rotation, and skew as follows:

$$\begin{aligned} \text{pixel scale ratio} = \text{PSR} &= \sqrt{ad - bc} \\ \tan\theta &= \left(\frac{b-c}{a+d}\right) \\ \text{on axis skew} &= \frac{1}{2}(a-d) \\ \text{off axis skew} &= \frac{1}{2}(b+c) \end{aligned} \tag{3.2}$$

With this relationship, we see that there are 6 parameters that need to be fit when transforming one image onto another[8]. In general, this would only require the positions of 3 sources found in both images, but in practice, many more sources are required to measure a robust transformation solution.

To be clear, this technique measures the change in (X, Y) as a function of time, but these measurements are relative to the population of sources in the image. For instance, a collection of stars moving with the same proper motion will only have a difference in translation at two epochs; fitting for transformation parameters will then yield zero change in the relative positions of all the stars. This is still a very useful technique in the cases where the relative motion of stars is of interest, such as for measuring internal kinematics of clusters and galaxies.

To find the proper motion of the stars in an absolute reference frame, we require known anchor points that have no motion (e.g. background stationary sources

[8]We only need to fit for either (X_0, Y_0) or (W_0, Z_0) because we can fix one vector and the other will absorb those choices

like galaxies and QSOs). An alternative approach is to cross-match the stars in an image to an absolute-frame-calibrated survey to estimate the bulk velocity/average absolute velocity of the sources, which is the technique that the `GaiaHub` pipeline (del Pino et al. 2022) employs. In a high level overview, the pipeline steps are as follows:

1. Following previous techniques for *HST* data reduction (e.g. Bellini et al. 2017, 2018; Libralato et al. 2018, 2019), fit *HST* images with different point spread functions (PSFs) to identify sources and measure their positions using `hst1pass` (Anderson 2022);

2. Cross-match *HST* sources with *Gaia*;

3. Find the transformation parameters that minimize the offsets between the *HST* source positions and *Gaia* positions for each image;

4. With the best fit transformation parameters in hand, any remaining offsets are divided by the time baseline to give the relative proper motion for each sources in an image;

5. Use an average of *Gaia*-measured PMs to estimate the bulk velocity, giving proper motions in an absolute frame;

6. Inverse-variance combine the multiple `GaiaHub` PM measurements of a star to get a final PM and uncertainty.

As is shown in del Pino et al. (2022), the pipeline performs well, especially in the medium density outskirts of nearby galaxies that it was developed for. Two key assumptions are required for `GaiaHub` results to accurately describe a population (1) the

parallax motion of the sources are insignificant, and (2) there are enough sources such that their PM distribution is close to Gaussian. The second point is particularly important because the offsets between the sources in different images, which determines the transformation parameters, are minimized when fitting for the transformation. If there are a small number of sources, the individual offsets have an out-sized impact on the resulting transformation parameters; sources with high offsets between images (such as a fast-moving foreground star), can toque the transformation solution around to try to reduce what is an intrinsically large separation. Similarly, in the limiting case where an image contains 3 sources, we can find transformation parameters such that there is no remaining offset, even if the sources truly have moved. In both cases, the transformation parameters are different from the truth, and this impacts the proper motion measurements of all sources in an image.

Motivated by the PM improvements that `GaiaHub` has shown when information from *Gaia* and *HST* is combined together, we create a tool that enables *Gaia*+*HST* PM measurements in sparse fields to study the MW stellar halo. Because these sparse fields have low stellar density (e.g. < 20 *Gaia* sources per *HST* image) and can contain a significant fraction of nearby sources (e.g. foreground disk and nearby halo stars, which can have non-negligible parallax motion and relatively large sky velocities), our pipeline cannot use the same assumptions that go into `GaiaHub`. This realization is what ultimately informed our decision to model the motions and transformation parameters simultaneously, though we emphasize that the following pipeline benefits immensely from the `GaiaHub` project. Specifically, the PSF fitting to measure centroids in *HST*,

the cross-matching between *HST* and *Gaia*, and the initial measurements of the transformation parameters from `GaiaHub` are integral components to the pipeline we present in the remainder of this work.

3.2.1 COSMOS Test Sample

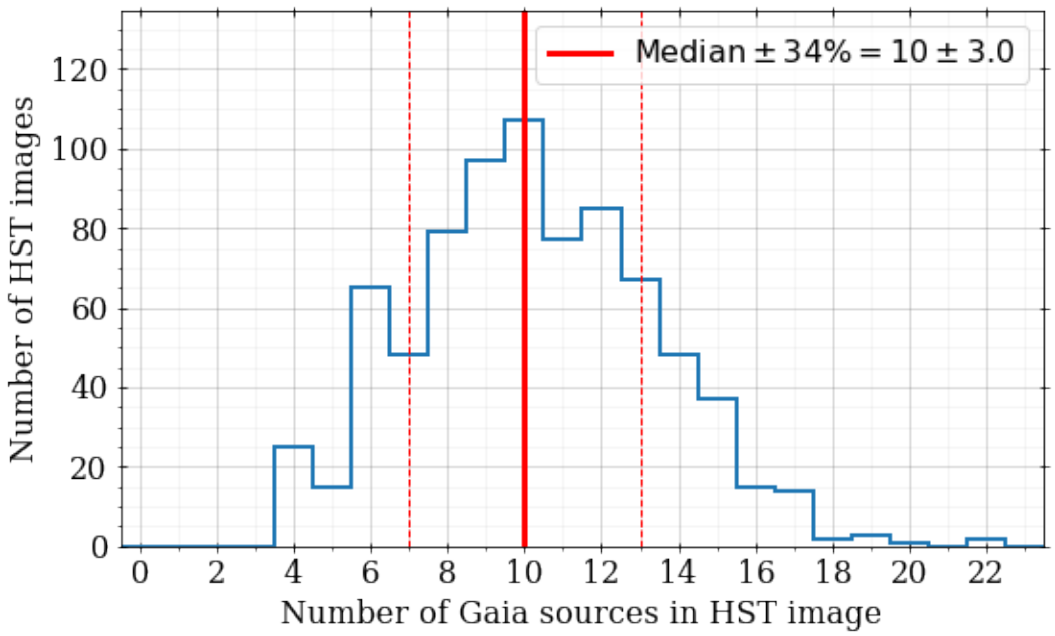

Figure 3.1: Histogram of the number of `GaiaHub` cross-matched sources between *Gaia* and *HST* for the 787 *HST* images in COSMOS within 0.5 deg of the field's center. There is a median of 10 *Gaia* sources in each *HST* frame, and a total of 2184 unique *Gaia* sources.

To test performance in sparse MW halo fields, we turn to the COSMOS field (Nayyeri et al. 2017; Muzzin et al. 2013) from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011, PIs: S. Faber, H. Ferguson). COSMOS is located at a high Galactic latitude and near the Galactic anti-center ($l = 236.8$ deg, $b = 42.1$ deg) and boasts *HST* imaging that covers

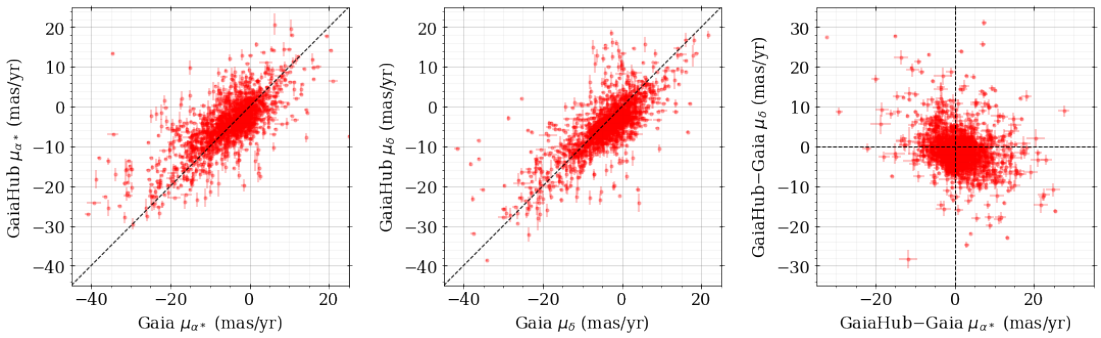

Figure 3.2: Comparison of the PMs measured by *Gaia* and `GaiaHub` for the 2184 COSMOS stars discussed in Figure 3.1. While the PMs generally fall along the one-to-one line (bashed black diagonals in the left 2 panels), the PM measurements do not agree within their uncertainties. This is largely because the `GaiaHub` pipeline was not tuned for small numbers of sources per *HST* image or for the large PMs seen in the MW stellar halo.

a large area of sky: ~ 2.25 deg^2 was imaged by ACS/WFC in 2003, and 288 arcmin^2 of that area (i.e. $\sim 3.6\%$) was imaged again in 2009. From simulated COSMOS-like data – presented in detail in Appendix 3.9.2 – PMs of halo and foreground disk stars in our magnitude range (i.e. $16 < G < 21.5$ mag) can regularly be as large as ~ 100 mas$\cdot$yr^{-1} in size.

We analyse all COSMOS *HST* images within a 0.5 deg radius of the field's center, which is $\sim 35\%$ of the total COSMOS area. Figure 3.1 shows a histogram of the number of sources matched between the *HST* frames and *Gaia* for the 787 *HST* images in our analysis. In these *HST* images, `GaiaHub` finds 2184 unique *Gaia* sources, with a median of 10 sources per image and a maximum of 23 sources per image.

A comparison of the *Gaia*- and `GaiaHub`-measured PMs for these COSMOS sources is presented in Figure 3.2. In general, the `GaiaHub` and *Gaia* PMs follow the one-to-one line, but they are quite far from agreeing within their uncertainties, as is

especially evident in the rightmost panel. Because `GaiaHub` is not tuned for sparse fields or for large PMs, it is not surprising to see these disagreements.

3.3 Bayesian Positions, Parallaxes, and Proper Motions: The `BP3M` Pipeline

We build a hierarchical Bayesian tool, `BP3M`, that measures transformation parameters to map *HST* images onto *Gaia* while also simultaneously measuring the PMs, parallaxes, and positions for the sources in an image. While it may appear computationally prohibitive to measure PMs, parallaxes, position of all stars in an image and the transformation parameters simultaneously, a few statistical tricks – namely conjugacy between likelihood and prior distributions – make this approach feasible. The pipeline is able to consider multiple *HST* images concurrently in a proper Bayesian fashion, which can significantly improve precision for the different motion components of each source. The `python`-based code of `BP3M` is publicly available on Github[9].

We use the *Gaia*-measured positions, parallaxes, and PMs and corresponding uncertainties/covariance as prior distributions to describe each source position over time. This improves the measured astrometric solution because the priors allow the measured positions of the sources to be compared to their expected positions at the epoch of each image, rather than compare the individual position measurements at each time. This approach is quite general in that it does not require identifying a clean kinematic sample or reference stars and background sources. In many cases, bright foreground stars with

[9] https://github.com/KevinMcK95/BayesianPMs

well-measured *Gaia* astrometry can serve as anchors that help define the transformation solution.

Throughout the statistics presented in this section and Appendix 3.9.1, we refer to *HST* and *Gaia* in the subscripts of the variables, but there is nothing about this math or its implementation in `BP3M` that restricts us to only these telescopes. The statistical statements are generally true for any collection of measurements with at least two positions of the same source – it applies to *HST*+*HST*, *HST*+*Roman*, *HST*+*Gaia*+*Roman*, etc – though there are a minimum of three sources per image needed to be able to measure the 6 transformation parameters.

3.3.1 The Statistics

This subsection will be quite technical in detail, so readers that are less interested in the formal statistics are encouraged to skip to Section 3.3.2.

First, we define many of the key variables and parameters in Table 3.1. While we are ultimately concerned with the proper motions in RA and Dec, it is convenient to work in a 2D plane projection (e.g. what a CCD sees) when comparing images. Because *Gaia*'s measurements don't correspond to positions in a true image, we follow `GaiaHub`'s approach of transforming the *Gaia* coordinates into XY coordinates on a pseudo image

Table 3.1: Definitions of fitting parameters. In general, H refers to an *HST* value, G refers to an *Gaia* value, an apostrophe indicates a prior measurement, j refers to the *HST* image number, and i refers to the source index.

Parameter	Description
(a_j, b_j, c_j, d_j)	Transformation matrix parameters for the j-th *HST* frame to the j-th pseudo *Gaia* frame
$(X_{0,j}, Y_{0,j})$	Center coordinate in the j-th *HST* frame
$(W_{0,j}, Z_{0,j})$	Center coordinate of j-th pseudo *Gaia* frame
$(X'_{H,i,j}, Y'_{H,i,j})$	Observed coordinate of i-th source in the j-th *HST* frame
$\sigma_{H,i,j}$	`GaiaHub`-measured pixel position uncertainty (in both x and y) the i-th source in the j-th *HST* frame
$(X'_{G,i,j}, Y'_{G,i,j})$	Observed coordinate of i-th source in the j-th pseudo *Gaia* frame
$\mathrm{PS}_{H,j}$	Pixel scale of the j-th *HST* frame
$\mathrm{PS}_{G,j}$	Pseudo pixel scale of j-th pseudo *Gaia* frame
$\vec{\theta}'^{T}_i = (\alpha'_i, \delta'_i)$	*Gaia*-measured position vector for the i-th source
$(\sigma'_{\alpha*,i}, \sigma'_{\delta,i}, \rho'_{\alpha*,\delta,i})$	*Gaia*-measured position uncertainties and correlation for the i-th source
plx'_i	*Gaia*-measured parallax for the i-th source
$\sigma'_{\mathrm{plx},i}$	*Gaia*-measured parallax uncertainty for the i-th source
$\vec{\mu}'^{T}_i = (\mu'_{\alpha*,i}, \mu'_{\delta,i})$	*Gaia*-measured PM vector for the i-th source
$(\sigma'_{\mu_{\alpha*},i}, \sigma'_{\mu_\delta,i}, \rho'_{\mu_{\alpha*},\mu_\delta,i})$	*Gaia*-measured PM uncertainties and correlation for the i-th source
$\vec{\Delta\theta}^{T} = (\Delta\alpha*_i, \Delta\delta_i)$	True position offset vector for the i-th source
plx_i	True parallax for the i-th source
$\vec{\mu}'^{T}_i = (\mu_{\alpha*,i}, \mu_{\delta,i})$	True PM vector for the i-th source
$t_G = \mathrm{J2016.0}$	Time of *Gaia* frame observation
$t_{H,j}$	Time of j-th *HST* frame observation
$\Delta t_j = t_{H,j} - t_G$	Time between *Gaia* and j-th *HST* frame (negative for older *HST* images)
$\Delta\vec{\mathrm{plx}}_{i,j}$	Offset-per-parallax vector for i-th source between *Gaia* and j-th *HST* frame

using

$$
\begin{aligned}
r_{G,i,j} &= \sin\left(\delta_{G,j,0}\right)\sin\left(\delta'_{G,i}\right) + \cos\left(\delta_{G,j,0}\right)\cos\left(\delta'_{G,j}\right)\cos\left(\alpha'_{G,i} - \alpha_{G,j,0}\right) \\
\text{rad2pix}_{G,j} &= \frac{180}{\pi}\cdot 3600 \cdot 1000 \cdot \text{PS}_{G,j}^{-1} \text{ pixels/radian} \\
X_{G,i,j} &= X_{G,j,0} - \text{rad2pix}_{G,j}\cdot\frac{\cos\left(\delta'_{G,i}\right)\sin\left(\alpha'_{G,i} - \alpha_{G,j,0}\right)}{r} \\
Y_{G,i,j} &= Y_{G,j,0} + \text{rad2pix}_{G,j}\cdot\left[\frac{\cos\left(\delta_{G,j,0}\right)\sin\left(\delta'_{G,i}\right)}{r}\right. \\
&\qquad \left. - \frac{\sin\left(\delta_{G,j,0}\right)\cos\left(\delta'_{G,i}\right)\cos\left(\alpha'_{G,i} - \alpha_{G,j,0}\right)}{r}\right]
\end{aligned}
\tag{3.3}
$$

where the $(\alpha'_{G,i}, \delta'_{G,i})$ are the *Gaia*-measured coordinates of source i are in radians, $(\alpha_{G,j,0}, \delta_{G,j,0})$ are the coordinates of the center of the *Gaia* pseudo image when considering *HST* image j, $(X_{G,j,0}, Y_{G,j,0})$ are the pixel coordinates of the center of the *Gaia* pseudo image, and the $(X_{G,i,j}, Y_{G,i,j})$ coordinates of source i in the *Gaia* pseudo image are in pixels. The chosen pixel scale for each *Gaia* pseudo image is set to the nominal *HST* pixel scale that it is paired with (e.g. 50 mas$\cdot$pixel^{-1} for *HST*'s ACS/WFC). Then, changes in the RA and Dec can be transformed to changes in XY using the Jacobian matrix:

$$
\boldsymbol{J}_{i,j} = \begin{pmatrix} \frac{dX_{G,i,j}}{d\alpha*'_{G,i}} & \frac{dX_{G,i,j}}{d\delta'_{G,i}} \\ \frac{dY_{G,i,j}}{d\alpha*'_{G,i}} & \frac{dY_{G,i,j}}{d\delta'_{G,i}} \end{pmatrix}
$$

which is usually very close to

$$
\boldsymbol{J}_j = \frac{1}{\text{PS}_{\text{G,j}}}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}
$$

when $(\Delta\alpha*, \Delta\delta)$ are in mas and $(\Delta X_{G,i,j}, \Delta Y_{G,i,j})$ are *Gaia* pseudo pixels. We note that this simplification only holds for small FOV detectors. In the cases of large PMs and large time baselines, the off-diagonal elements can start to become important, so we opt to use the more general version in our pipeline.

To set up the probability model, there are a few remaining key terms that we need to define:

$$\boldsymbol{R}_j = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}$$

which is the transformation matrix for the j-th *HST* frame to the j-th pseudo *Gaia* frame,

$$\boldsymbol{V}_{H,i,j} = \begin{pmatrix} \sigma^2_{H,i,j} & 0 \\ 0 & \sigma^2_{H,i,j} \end{pmatrix}$$

which is the `GaiaHub`-measured pixel position covariance matrix for the i-th source in the j-th *HST* frame,

$$\boldsymbol{V}_{\theta,i} = \begin{pmatrix} \sigma'^2_{\alpha*,i} & \rho'_{\alpha*,\delta,i} \cdot \sigma'_{\alpha*,i} \cdot \sigma'_{\delta,i} \\ \rho'_{\alpha*,\delta,i} \cdot \sigma'_{\alpha*,i} \cdot \sigma'_{\delta,i} & \sigma'^2_{\delta,i} \end{pmatrix}$$

which is the *Gaia*-measured position covariance matrix for the i-th source, and finally

$$\boldsymbol{V}_{\mu,i} = \begin{pmatrix} \sigma'^2_{\mu_{\alpha*},i} & \rho'_{\mu_{\alpha*},\mu_\delta,i} \cdot \sigma'_{\mu_{\alpha*},i} \cdot \sigma'_{\mu_\delta,i} \\ \rho'_{\mu_{\alpha*},\mu_\delta,i} \cdot \sigma'_{\mu_{\alpha*},i} \cdot \sigma'_{\mu_\delta,i} & \sigma'^2_{\mu_\delta,i} \end{pmatrix}$$

which is the *Gaia*-measured PM covariance matrix for the i-th source. In these equations and the ones that will follow, our convention is to show matrices (and matrices of

matrices) using a bold-faced typesetting.

The fundamental relationships between the prior *Gaia* measurements and the true parameter for source i are given by:

$$\vec{\theta}_i' = \vec{\theta}_i + \vec{\Delta\theta}_i$$

$$p(\vec{\Delta\theta}_i'|\vec{\Delta\theta}_i) = \mathcal{N}\left(\vec{\Delta\theta}_i'|\vec{\Delta\theta}_i, \boldsymbol{V}_{\theta,i}\right)$$

$$p(\mathrm{plx}_i'|\mathrm{plx}_i) = \mathcal{N}\left(\mathrm{plx}_i'|\mathrm{plx}_i, \sigma'_{\mathrm{plx},i}\right)$$

$$p(\vec{\mu}_i'|\vec{\mu}_i) = \mathcal{N}\left(\vec{\mu}_i'|\vec{\mu}_i, \boldsymbol{V}_{\mu,i}\right)$$

which says that the measured values are offset from the true values by noise dictated by their measured uncertainties. While we have explicitly included a mean for the *Gaia*-measured position offset vector $\vec{\Delta\theta}_i'$ to be as general as possible, in practice $\vec{\Delta\theta}_i' = \vec{0}$ because *Gaia* has no expected offset from the positions it reports. In the form of the probability statements above, we have assumed that there is no correlation between, for example, parallax and $\mu_{\alpha*}$, which is not exactly correct because *Gaia* has measures for these correlations. Our choices, however, make the following math easier, though it comes at the cost of some additional constraining power from the *Gaia* priors being ignored. The *Gaia*-measured correlations between these parameters are indeed included in future sections of this work when we compare the `BP3M` distributions to *Gaia*'s results. Future versions of the `BP3M` pipeline will work to incorporate these prior correlations, which will lead to even tighter posterior distributions on position, parallax, and PMs.

While it may seem nonphysical, *Gaia* observed parallax measurements can be negative (Lindegren et al. 2018), but it is important to remember that the observed

parallax values define the mean of a distribution whose width/uncertainty usually places a large amount of probability in positive parallaxes; in this way, we must treat the *Gaia* astrometric measurements as distributions and not individual points. Luri et al. (2018), for example, explain how the definitions of motion on the sky (both by *Gaia* as well as in this work) technically allow for negative observed parallax values, which is especially likely to occur when the position measurement as a function of time is relatively large compared to the size of the parallax motion (e.g. see their Section 3 and Figure 2); as a result, they remind the reader that the *Gaia* observed parallaxes should not be thought of as a direct measurement of distance, and instead, distances need to be estimated by proper statistical modelling of the information contained in the astrometric solution distributions.

For sources where there are no *Gaia*-measured parallaxes and PMs, we find it convenient to put a population/global prior on the PMs and parallaxes:

$$p(\mathrm{plx}_i|\hat{\mathrm{plx}}, \sigma_{\hat{\mathrm{plx}}}) = \mathcal{N}\left(\mathrm{plx}_i|\hat{\mathrm{plx}}, \sigma_{\hat{\mathrm{plx}}}\right)$$

$$p(\vec{\mu}_i|\hat{\mu}, \boldsymbol{V}_{\hat{\mu}}) = \mathcal{N}\left(\vec{\mu}_i|\hat{\mu}, \boldsymbol{V}_{\hat{\mu}}\right)$$

which says that there are some global distributions for the parallaxes and PMs that the true values originate from. While we are free to play with the parameters of the population distributions, we note that they do need to be Gaussian in form so that we retain the necessary conjugacy. We choose to use diffuse hyperpriors, with the goal of minimally impacting the sources with *Gaia*-measured parallaxes and PMs while offering some guidance to the sources without. In the future, the parallax global prior could

be made more constraining in cases where the distances are better known (e.g. clean populations of extra-Galactic stars) or using a magnitude-dependent parallax prior to better incorporate our understanding of the distribution of stars in the MW. Similar changes could also be made to the PM global prior. For the current version of the pipeline, which focuses on faint stars in the MW stellar halo and extra-Galactic sources, we choose the parallax prior to have a mean of 0.5 mas and width of 10.0 mas. For the PM prior, we use the sources with *Gaia*-measured PMs to estimate a mean and covariance matrix, and then we multiply that covariance matrix by a factor of 10^2 to guard against the possibility that some of the sources without *Gaia* PMs are significantly different in their PM from the other sources.

The information linking *Gaia* to *HST* image j for source i is then given by:

$$\vec{\Delta d}_{G,i,j} = \begin{pmatrix} X_{G,i,j} \\ Y_{G,i,j} \end{pmatrix} - \boldsymbol{R}_j \cdot \begin{pmatrix} X_{H,i,j} - X_{0,j} \\ Y_{G,i,j} - Y_{0,j} \end{pmatrix} - \begin{pmatrix} W_{0,j} \\ Z_{0,j} \end{pmatrix}$$

$$\vec{\Delta m}_{i,j} = \vec{\mu}_i \cdot \Delta t_j + \mathrm{plx_i} \cdot \vec{\Delta \mathrm{plx}}_{i,j} - \vec{\Delta \theta}_i$$

$$\vec{\Delta d}_{G,i,j} \sim \mathcal{N}\left(\boldsymbol{J}_{i,j} \cdot \vec{\Delta m}_{i,j}, \boldsymbol{V}_{d,i,j} = \boldsymbol{J}_{i,j} \cdot \boldsymbol{V}_{\theta,i} \cdot \boldsymbol{J}_{i,j}^T + \boldsymbol{R}_j \cdot \boldsymbol{V}_{H,i,j} \cdot \boldsymbol{R}_j^T\right)$$

which says that the measured offset between the *Gaia* and *HST* positions (after applying the transformation) in the pseudo *Gaia* image is distributed around the offset implied by the sum of the motion from PM, parallax, and uncertainty in position. We note that this relationship ignores the impact that radial motion has on changing the distance to a star between observations – and therefore the magnitude of the proper motion and parallax – though this is a safe assumption because the distance change is almost always

extremely small.

The $\Delta\vec{\text{plx}}_{i,j}$ term, which we refer to as the parallax offset vector, is a 2D vector that defines the direction and magnitude of the offset between two observations as a result of parallax motion for a source with a parallax of 1 mas; in this way, the parallax offset vector can be multiplied by any parallax value (i.e. $\text{plx}_i \cdot \Delta\vec{\text{plx}}_{i,j}$) to find the appropriate offset in $(\Delta\alpha*, \Delta\delta)$ between two observations as a result of parallax motion. To measure the $\Delta\vec{\text{plx}}_{i,j}$, we use the *HST* and *Gaia* observation times, $(t_{H,j}, t_G)$, the position of the source in *Gaia*, (α'_i, δ'_i), and built-in functions of `astropy` (Astropy Collaboration et al. 2013, 2018, 2022). Examples of the parallax motion for different positions on the sky are shown in Figure 3.3, where the *Gaia* observation time (i.e. J2016.0) is at the origin and the orbits trace out the parallax motion over the course of a year; the parallax offset vector at any time is simply the vector that connects the corresponding point on the ellipse to the origin.

With the above definitions in hand, we can use the distributions on the $\vec{\Delta d}_{G,i,j}$ vectors for all the sources in all images we are considering to construct the following likelihood distribution:

$$p(\boldsymbol{X_H}, \boldsymbol{Y_H}, \boldsymbol{X_G}, \boldsymbol{Y_G} | \vec{a}, \vec{b}, \vec{c}, \vec{d}, \vec{W}_0, \vec{Z}_0, \boldsymbol{\mu}, \mathbf{plx}, \boldsymbol{\Delta\theta}) = \prod_{i=1}^{n_*} \prod_{j=1}^{n_{im}} \mathcal{N}\left(\Delta\vec{d}_{G,i,j} | \boldsymbol{J}_{i,j} \cdot \vec{\Delta m}_{i,j}, \boldsymbol{V}_{d,i,j}\right)$$

We note here that we have purposefully chosen omit writing out the dependency on a few of the explanatory variables (e.g. Δt_j, the covariance matrices) for ease of reading the math.

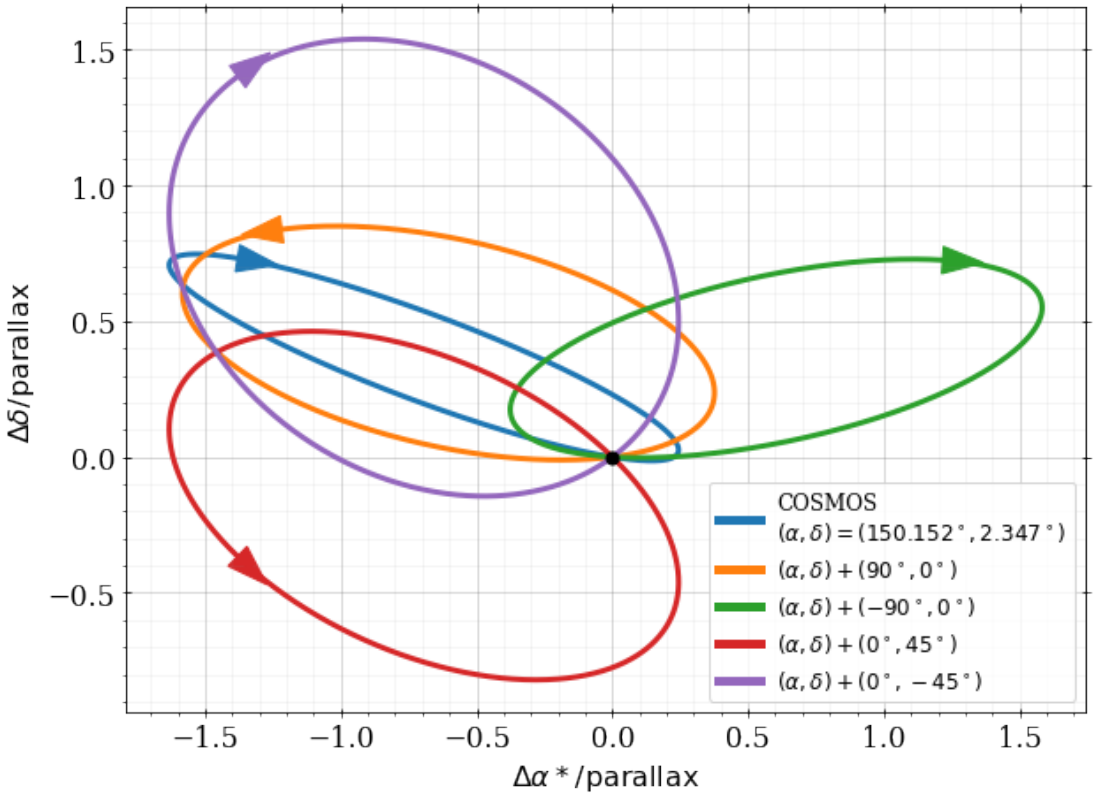

Figure 3.3: Examples of parallax in the plane of the sky for different lines-of-sight. The origin (black point) is the time that corresponds to *Gaia* observations (i.e. J2016.0), and each ellipse is the path that a star with a parallax of 1 mas sweeps out on the sky over the course of a year, with the direction of motion shown by the arrowhead.

From Bayes' Law, we arrive at the following posterior:

$$
\begin{aligned}
p\left(\vec{a},\vec{b},\vec{c},\vec{d},\vec{W}_0,\vec{Z}_0,\boldsymbol{\mu},\vec{\mathrm{plx}},\boldsymbol{\Delta\theta}|\boldsymbol{X_H},\boldsymbol{Y_H},\boldsymbol{X_G},\boldsymbol{Y_G},\hat{\mathrm{plx}},\sigma_{\hat{\mathrm{plx}}},\hat{\mu},\boldsymbol{V}_{\hat{\mu}}\right) \\
\propto \prod_{j=1}^{n_{im}} \left[p(a_j,b_j,c_j,d_j,W_{0,j},Z_{0,j})\right] \cdot \\
\prod_{i=1}^{n_*} \Big\{ p(\vec{\mu}_i'|\vec{\mu}_i) \cdot p(\vec{\mu}_i|\hat{\mu},\boldsymbol{V}_{\hat{\mu}}) \cdot p(\mathrm{plx}_i'|\mathrm{plx}_i) \cdot p(\mathrm{plx}_i|\hat{\mathrm{plx}}) \cdot \\
p(\vec{\Delta\theta}_i) \cdot \prod_{j=1}^{n_{im}} \mathcal{N}\left(\Delta\vec{d}_{G,i,j}|\boldsymbol{J}_{i,j}\cdot\vec{\Delta m}_{i,j},\boldsymbol{V}_{d,i,j}\right) \Big\}
\end{aligned}
\tag{3.4}
$$

where $p(a_j, b_j, c_j, d_j, W_{0,j}, Z_{0,j})$ is the prior distribution on the transformation parameters and the remaining distributions have been defined above. In many of our initial analyses, we chose to use a flat prior on the transformation parameters, which yielded reasonable results. However, we soon realized that there are relatively tight constraints for our expectations of the pixel scale ratio and skew terms when using *HST* ACS/WFC images, which will be discussed more in Section 3.6. To that end, when using *HST* ACS/WFC data, we choose to employ a few relatively diffuse priors on the skews, pixel scale ratio, and *HST* angle, as well as the $(W_{0,j}, Z_{0,j})$ vector:

$$
\begin{aligned}
p(\mathrm{skew}_j | \textit{HST}\ \mathrm{ACS/WFC}) &= \mathcal{N}\left(\mathrm{skew}_j | \mu = \mathrm{skew}'_j, \sigma = 5 \times 10^{-4}\right) \\
p(\mathrm{PSR}_j | \textit{HST}\ \mathrm{ACS/WFC}) &= \mathcal{N}\left(\mathrm{PSR}_j | \mu = \mathrm{PSR}'_j, \sigma = 5 \times 10^{-4}\right) \\
p(\theta_j) &= \mathcal{N}\left(\theta_j | \mu = \theta'_j, \sigma = 1^\circ\right) \\
p(W_{0,j}) &= \mathcal{N}\left(W_{0,j} | \mu = W'_{0,j}, \sigma = 10\ \mathrm{pixels}\right) \\
p(Z_{0,j}) &= \mathcal{N}\left(Z_{0,j} | \mu = Z'_{0,j}, \sigma = 10\ \mathrm{pixels}\right)
\end{aligned}
$$

where we use the same prior on both the on- and off-axis skew terms. The prior means of each distribution come from the parameters measured by the previous iteration of transformation parameter fitting; in the case of the first iteration, we use the `GaiaHub` outputs. In this way, we tell the pipeline that the transformation solution is likely nearby the previous iteration's solution, while the relatively large widths allows the values to change significantly if necessary. To transform from priors in angle, skews, and PSR,

we need to account for a transformation Jacobian:

$$\boldsymbol{J}_{\mathrm{TR,j}} = \begin{pmatrix} \frac{d \text{ on skew}_j}{da_j} & \frac{d \text{ off skew}_j}{da_j} & \frac{d\mathrm{PSR}_j}{da_j} & \frac{d\theta_j}{da_j} \\ \frac{d \text{ on skew}_j}{db_j} & \frac{d \text{ off skew}_j}{db_j} & \frac{d\mathrm{PSR}_j}{db_j} & \frac{d\theta_j}{db_j} \\ \frac{d \text{ on skew}_j}{dc_j} & \frac{d \text{ off skew}_j}{dc_j} & \frac{d\mathrm{PSR}_j}{dc_j} & \frac{d\theta_j}{dc_j} \\ \frac{d \text{ on skew}_j}{dd_j} & \frac{d \text{ off skew}_j}{dd_j} & \frac{d\mathrm{PSR}_j}{dd_j} & \frac{d\theta_j}{dd_j} \end{pmatrix}$$

such that

$$p(a_j, b_j, c_j, d_j) = p(\text{on skew}_j, \text{off skew}_j, \mathrm{PSR}_j, \theta_j) \cdot |\boldsymbol{J}_{\mathrm{TR,j}}|.$$

The detailed steps to find the posterior conditional distributions for each motion component of source i are given in Appendix 3.9.1. In the end, that work yields the following important distributions:

$$p(\mathrm{plx}_i | \vec{a}, \vec{b}, \vec{c}, \vec{d}, \vec{W}_0, \vec{Z}_0, \dots)$$

$$p(\vec{\mu}_i | \mathrm{plx}_i, \vec{a}, \vec{b}, \vec{c}, \vec{d}, \vec{W}_0, \vec{Z}_0, \dots)$$

$$p(\Delta\vec{\theta}_i | \vec{\mu}_i, \mathrm{plx}_i, \vec{a}, \vec{b}, \vec{c}, \vec{d}, \vec{W}_0, \vec{Z}_0, \dots)$$

where the ellipsis refers to the other variables that each distribution depends on. One key takeaway from the statistics in Appendix 3.9.1 is that the conjugacy between all distributions (i.e. all distributions are normally distributed/Gaussians) allows us to fairly easily combine multiple distributions to arrive at well-defined and closed-form posterior distributions.

With these posterior conditionals in hand, we can perform the following steps for each source for a given set of transformation parameters:

1. Draw samples of plx_i from $p(\mathrm{plx}_i|\dots)$;

2. Use those plx_i samples to draw samples of $\vec{\mu}_i$ from $p(\vec{\mu}_i|\mathrm{plx}_i,\dots)$;

3. Use those $(\vec{\mu}_i, \mathrm{plx}_i)$ samples to draw $\Delta\vec{\theta}_i$ samples from $p(\Delta\vec{\theta}_i|\vec{\mu}_i, \mathrm{plx}_i,\dots)$.

Once we have samples of the PMs, parallaxes, and position offsets for each source, we can calculate the posterior probability of a set of transformation parameters. We do this by marginalizing over the individual samples of the PMs, parallaxes, and position offsets using Bayes' Law such that

$$p(\vec{a},\vec{b},\vec{c},\vec{d},\vec{W}_0,\vec{Z}_0|\dots) = \frac{p(\vec{a},\vec{b},\vec{c},\vec{d},\vec{W}_0,\vec{Z}_0,\boldsymbol{\mu},\vec{\mathrm{plx}},\boldsymbol{\Delta\theta}|\dots)}{\prod_{i=1}^{n_*} p(\vec{\mu}_i,\mathrm{plx}_i,\Delta\vec{\theta}_i|\vec{a},\vec{b},\vec{c},\vec{d},\vec{W}_0,\vec{Z}_0,\dots)}.$$

which is independent of the particular values of the PMs, parallaxes, and positions of each source. This relationship is the magic that makes the simultaneous fitting of the 6 transformation parameters per image and 5 motion parameters per source feasible; because we can quickly draw the PM, parallax, and position samples directly from a known posterior distribution given a set of transformation parameters, we can efficiently calculate the posterior probability of the transformation parameters alone. These probabilities then allow us to sample from the posterior distribution on the transformation parameters using a Metropolis-Hastings MCMC algorithm. Then, for each sample of transformation parameters, we can sample from the posterior conditional distributions of parallaxes, PMs, and position offsets for each source. In this way, the uncertainty of the transformation fitting is propagated to the resulting PM, parallaxes, and positions.

3.3.2 The Pipeline

In terms of implementing the statistics in the `BP3M` pipeline, the general steps are as follows:

1. Read in the position, parallax, PM data from *Gaia*, where it exists, and the corresponding positions in the *HST* frames;
2. Use the `GaiaHub` transformation parameter values as starting guesses;
3. Using sources with *Gaia* priors on PM, estimate the global PM distribution;
4. Simultaneously fit the transformation parameters and stellar motion using MH-MCMC, using only sources with *Gaia* priors when measuring the transformation;
5. Identify outliers (e.g. bad cross-matches between *Gaia* and *HST*);
6. Re-estimate the global PM distribution using the new posteriors;
7. Repeat the fitting process using the non-outlier sources, including sources without *Gaia* priors;
8. Check if the list of outliers has changed substantially, and stop if it hasn't.

This process requires a minimum of two iterations to achieve good results, though it isn't uncommon for the outlier list to change enough that a third iteration is required. In terms of processing time, running `BP3M` on a 2016 MacBook Pro for an *HST* image with $\sim$ 200 sources takes approximate 15 minutes to complete its analysis, while an image with $\sim$ 10 sources can be analysed in around 5 minutes. Increasing the number of *HST* images being analysed together effectively multiplies the computation time by

the number of images; this is largely because we need to fit 6 additional transformation parameters for each additional image, and necessarily need to increase the number of MCMC walkers and number of MCMC iterations. Best practice is to analyse all images independently before combining them to save time on searching the transformation parameter space.

Because the *HST* images are mapped onto the global reference frame of *Gaia*, the transformation parameters that `BP3M` measures are also useful in constraining non-*Gaia* sources. Once the best set of transformation parameters between *HST* and *Gaia* have been measured, we can return to the list of all sources in the *HST* image as measured by `GaiaHub`, which can include sources much fainter than *Gaia* can see (i.e. $G > 21.5$ mag). When comparing multiple *HST* images at various epochs, these faint sources can be cross-matched with each other to recover their PMs and parallaxes to much fainter magnitudes. While the pipeline currently offers this feature, it is currently untested and unoptimized, but preliminary results suggest this approach will be fruitful. It may prove particularly useful in very sparse fields where the number of *Gaia* sources in each *HST* image is small, but the shared source list between the *HST* frames is large.

3.3.3 Caveats

One key assumption that is built in to `GaiaHub` – and therefore `BP3M` – comes from how *HST* sources are cross-matched with *Gaia*. A *HST* source is matched with a *Gaia* source if they are nearest neighbours within some angular distance of each other (e.g. `GaiaHub` default of 5 *HST* pixels, ~ 250 mas). This works quite well in medium to low density regions where the stars are likely far enough from each other with respect

to their motion between images. In high-density regions (e.g. disks of nearby galaxies like M31) or for fast moving stars, the closest pair of sources successive images aren't necessarily the correct matching. While future versions of our pipeline may address this better (e.g. re-wind the *Gaia* star's using their PMs, where they exist) to reduce this confusion, we stress that our work focuses on the low to medium density regions where the cross-matching assumptions are valid.

As a back-of-the-envelope test, the cross-matching technique is a good assumption when the average distance between stars in an HST image is twice as large as the average position change from proper motion:

$$\rho_*^{-1/2} > 2 \cdot \Delta t \cdot \bar{\mu}. \tag{3.5}$$

where ρ_* is the stellar number density in area on the sky. We can use this equation to determine different limits for choices of time baselines, densities, and average proper motion sizes. For example, with ACS/WFC's 4096×4096 pixel detector with pixel scale of 50 mas $\cdot$ pixel^{-1}, a time baseline of 15 years, and an average proper motion of 100 mas $\cdot$ yr^{-1}, we find that there would need to be greater than 4660 stars in the image for there to be a significant amount of confusion in the cross-matching; as these numbers are similar to HST images in the COSMOS field, this implies that sparse regions in the halo have very low risk of incorrect cross-matching. In denser regions, like nearby galaxies, a time-baseline of 15 years and average proper motion of 5 mas $\cdot$ yr^{-1} implies that a limiting stellar number density of 1/9 pixel^{-2} $= 4.4 \times 10^{-5}$ mas^{-2}. In practice, we would like the factor of 2 to be even larger (e.g. 5 or 10) to be safe, but this sets

the threshold. Future versions of the pipeline will likely explore improvements to the cross-matching between *HST* and *Gaia* using the posterior motion measurements from `BP3M`.

3.4 Pipeline Validation

We test the pipeline in four key ways: (1) using synthetic data where we know the input transformation parameters and stellar motions perfectly, (2) using well-studied nearby dwarf spheroidal (dSph) galaxies, (3) using cross-matches with an external QSO catalog, and (4) using cross-matches with a PM catalog in COSMOS derived from multi-epoch *HST* imaging.

3.4.1 Comparison with Synthetic Data

To understand how the pipeline performs in sparse halo fields, we first generate synthetic, COSMOS-like data. This process is described in detail in Appendix 3.9.2. In summary, the synthetic sources are a mix of foreground thick disk stars and halo stars covering the $16 < G < 21.5$ mag range with realistic *Gaia* measurements and uncertainties of position, parallax, and proper motion. Like with real *Gaia* data, sources with $G > 21$ mag have no *Gaia*-measured PMs or parallaxes. We create synthetic *HST* observations of these sources – as well as the corresponding `GaiaHub`-like outputs that `BP3M` expects to use for initial guesses of the transformation parameters – while varying the numbers of sources per *HST* image and time baselines. Finally, the synthetic images are analysed by `BP3M`. We emphasize that creating synthetic data

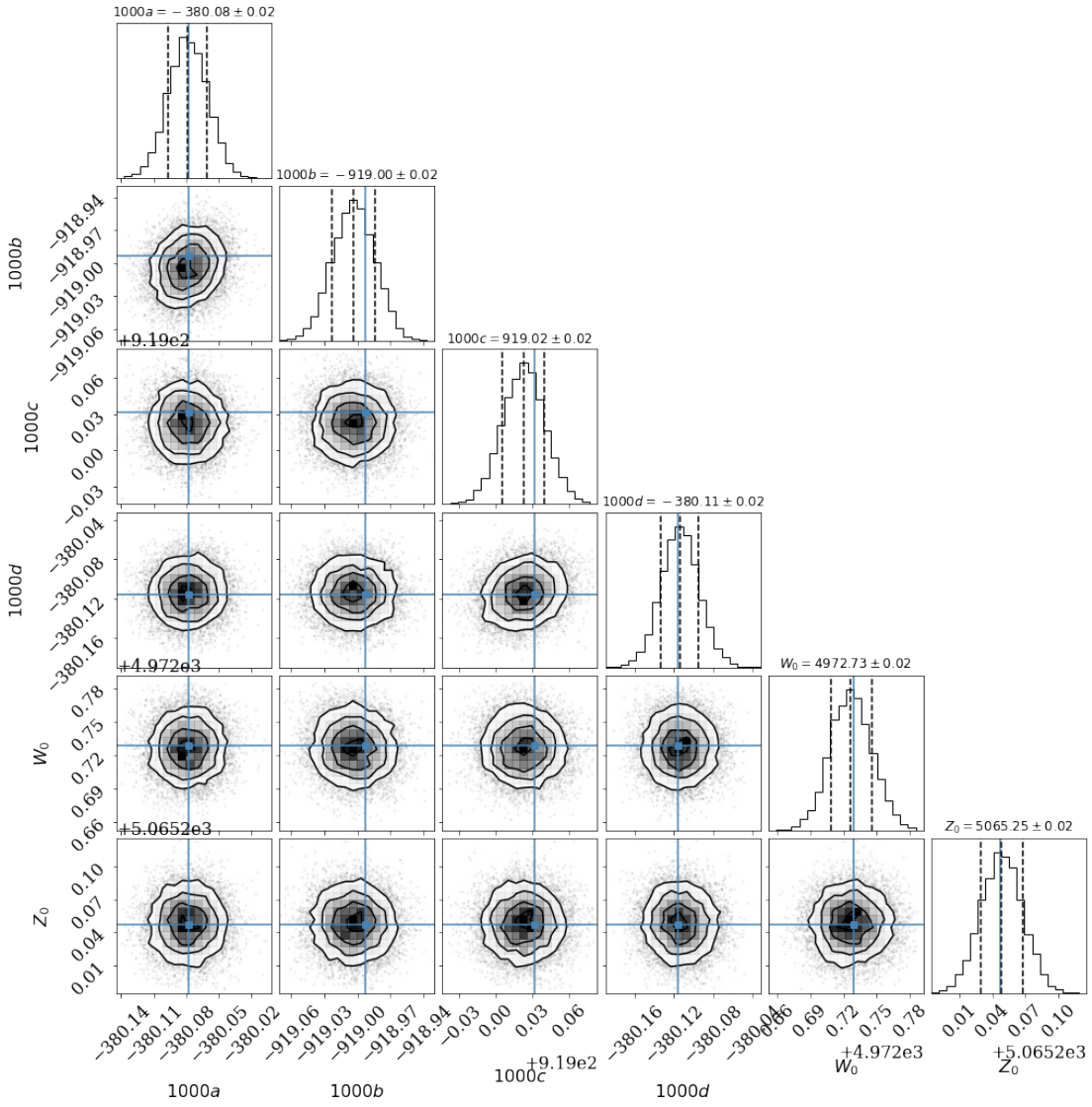

Figure 3.4: Corner plot of the posterior samples (black points, lines, and histograms) of the transformation parameters when fitting an image with 200 stars with a 15 year time-baseline. The blue lines show the locations of the input parameters used to create the synthetic data, and the values of the transformation matrix (a, b, c, d) have been multiplied by 1000 for clarity. The chosen input parameters are representative of real transformation solutions measured by `GaiaHub` for *HST* ACS/WFC images.

with different configurations (e.g. transformation parameters, time baselines, stellar velocities and distance distributions) is quite straightforward using our technique, and it is not necessarily restricted to only *HST*-like observations. Our method is a useful avenue for estimating the impact that future observations or telescopes can have on stellar motion measurements as well as designing best practices. We expand upon this further in the case of *Roman* in Section 3.7.

While the exciting comparison with real data will follow in Section 3.4.2, we begin by testing the pipeline on synthetic data to see how well we can recover the input transformation parameters and stellar motion (i.e. position, parallax, and PM) used to generate the synthetic *HST* image. Figure 3.4 shows the posterior distributions on the transformation parameters that `BP3M` measures for one synthetic *HST* image that has 200 sources and a 15 year time baseline from *Gaia*; the black points and histograms show the posterior draws, while the blue lines show the true values used to generate the image. The posterior distribution and the input values agree very strongly with each other. The chosen input transformation parameters are representative of real transformation solutions as measured by `GaiaHub`.

We repeat this process of generating synthetic images while changing the number of sources and the time baseline, and then measure how far the posterior distribution is from the truth. Specifically, we use the posterior samples of the transformation parameters (e.g. the black points in Figure 3.4) to define a 6D posterior mean vector, $\vec{v}$, and corresponding 6×6 covariance matrix, $\boldsymbol{V}_{\vec{v}}$. The 6D distance between the posterior

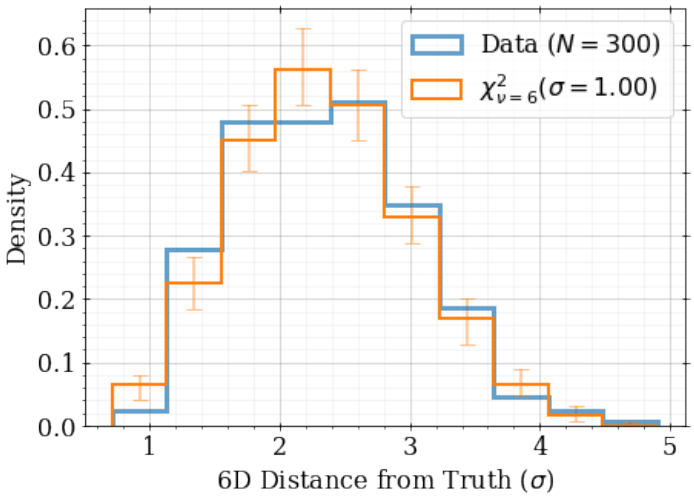

Figure 3.5: 6D distance of the transformation parameters (i.e. the parameters in Figure 3.4) from their true values for 300 realizations of synthetic *HST* images with time baselines from 5 to 15 years and number of sources from 5 to 200. The blue histogram shows the data, while the orange curve is the expected distribution; the agreement between these histograms show that the `BP3M` pipeline does a good job of recovering the input transformation parameters within their uncertainties.

mean vector and the truth vector, $\vec{v}'$, is then defined as

$$D = \left((\vec{v} - \vec{v}')^T \cdot \boldsymbol{V}_{\vec{v}}^{-1} \cdot (\vec{v} - \vec{v}')\right)^{1/2}. \tag{3.6}$$

This is analogous to the 1D case of dividing the absolute difference between a mean value and the truth by the uncertainty. Likewise, the units of D can be thought of as the number of σ between the truth and mean, and we will use this definition (for varying dimensions of $\vec{v}$) when studying residuals throughout this work. When we analyse 300 synthetic *HST* images with time baselines between 5 and 15 years from *Gaia* and 5 to 200 sources per image, we get the blue histogram in Figure 3.5; the orange distribution shows the expected outcome – namely, a χ^2 distribution with 6 degrees

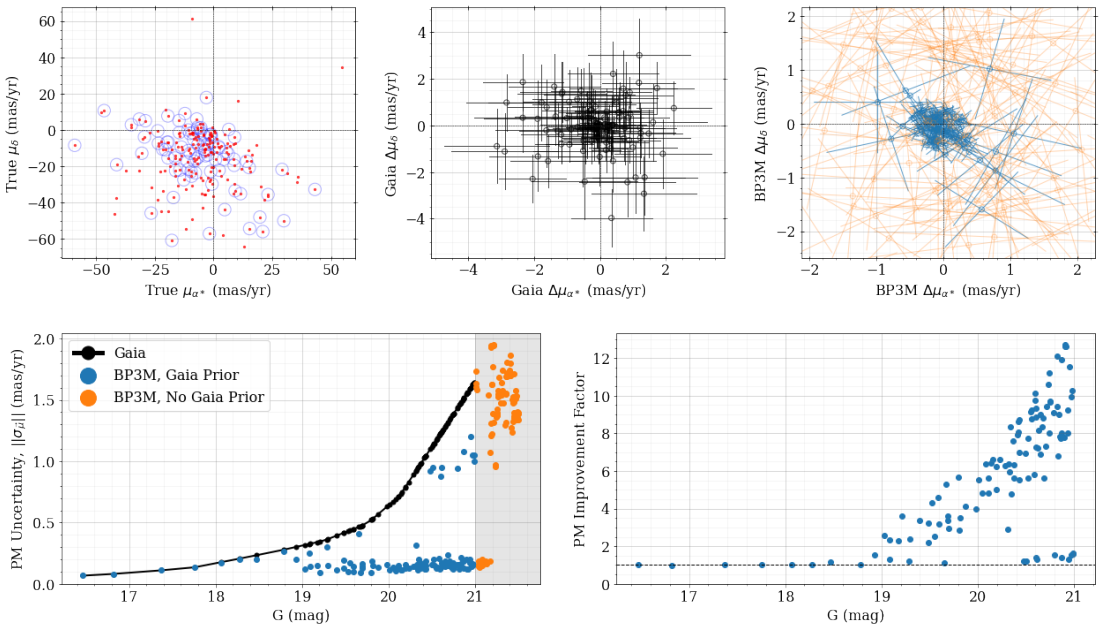

Figure 3.6: Comparison of posterior PMs for 200 synthetic COSMOS-like stars with a 15 year time baseline from *Gaia*. The upper left panel shows the true PM values (red points), and the dark blue outline circles highlight PMs that are too faint for *Gaia* to have measurements. The upper middle and right panels show the offsets of the measured PMs from the truth for both the synthetic *Gaia* data and the `BP3M` posteriors; in the `BP3M` panel, the blue points are star that have *Gaia*-measured PMs and parallaxes, while the orange points are those without *Gaia*-measured priors. The bottom two panels compare the PM vector uncertainty size between *Gaia* and `BP3M`, with the right panel showing the division of the black points by the blue points from the left panel.

of freedom and a scale of 1.0 – which agrees well with the measured outputs. This result shows that the pipeline does a good job of recovering the input transformation parameters within the posterior uncertainties.

Next, we analyse how well the pipeline is able to recover the true motions of the sources in each synthetic image. An example of this is shown in Figure 3.6 for the same synthetic image that is considered in Figure 3.4. When showing `BP3M` results, we color the sources by whether there are *Gaia*-measured parallaxes and PMs to be used as priors; blue points designate sources with ("Gaia Prior") and orange points

designate sources without ("No Gaia Prior"). To be precise, all of the sources have *Gaia*-measured priors on their position at J2016.0, but not all sources have *Gaia* priors on their parallaxes and PMs.

The first key takeaway from this figure is that the `BP3M` PMs are clustered closer to the true values (i.e. the origin in the upper middle and right panels) when compared to *Gaia* alone; the scale of the `BP3M` panel is a factor ~ 2 times smaller than that of of the *Gaia* panel. Next, the bottom panels show that the size of the PM uncertainty has decreased significantly when *HST* information is combined with *Gaia* using `BP3M`; the left panel compares the PM uncertainties, while the right panel shows how much smaller the `BP3M` PM uncertainties are compared to the *Gaia* PM uncertainties as a function of magnitude. As a whole, this figure shows that combining the *HST* and *Gaia* data not only increases PM precision, it also increases accuracy. To be clear, `BP3M` is not simply shrinking the *Gaia* PM distribution around the same mean as a result of an increased time baseline and number of images, it is truly improving the astrometric solution of each source.

The particular pattern of PM uncertainties (i.e. a lower branch around 0.2 mas$\cdot$ yr^{-1} and an upper branch around 1.5 mas$\cdot$yr^{-1}) is a consequence of our choices in modelling the synthetic data, which is detailed in Appendix 3.9.2. As a brief summary, the *HST* position uncertainties are based on real `GaiaHub`-analysed sources as a function of magnitude in COSMOS. At a given magnitude, some of the sources have well-measured positions, while others are less constrained, which leads to the PM uncertainty bifurcation shown in Figure 3.6.

When considering the uncertainty on a vector, there are a few different approached one can take. In this work, we are concerned with 2D vectors (e.g. positions, PMs), 3D vectors (e.g. parallax plus 2D PM), and 5D vectors (i.e. 2D position, parallax, and 2D PM) of the different motion components. Because the different components of these vectors can have differing units, and because there can be substantial correlations in the covariance matrices we measure, we choose not to use the standard metric of the quadrature sum of the individual uncertainties of each component. Instead, when we compare vectors that have associated covariance matrices, we are interested in how much the size of that covariance matrix has changed, which includes the effect of correlation. We define the size of a vector's uncertainty to be the determinant of the covariance matrix to the $1/2d$ power, where d is the vector's number of dimensions:

$$||\sigma_{\vec{v}}|| = |\boldsymbol{V}_{\vec{v}}|^{1/2d} \tag{3.7}$$

where $\boldsymbol{V}_{\vec{v}}$ is the covariance matrix of $\vec{v}$. With this definition, the area/volume created by the vector's covariance matrix, i.e. the determinant $|\boldsymbol{V}_{\vec{v}}|$, is equal to the area/volume of a purely diagonal covariance matrix:

$$\boldsymbol{V}_{||\sigma_{\vec{v}}||} = \begin{pmatrix} ||\sigma_{\vec{v}}||^2 & 0 & \dots & 0 \\ 0 & ||\sigma_{\vec{v}}||^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & ||\sigma_{\vec{v}}||^2 \end{pmatrix}.$$

In this case, the resulting uncertainty size allows for the correlations between parameters

to impact the certainty about the vector's position, which serve to shrink the volume defined by that covariance matrix. For the remainder of this work, where we compare the uncertainty sizes of different vectors between *Gaia* and `BP3M`, we will be using the vector uncertainty size definition of Equation 3.7.

As an illustrative example, a highly-correlated measurement might have large uncertainties in all of the individual vector components, but the probability distribution implied by its covariance matrix covers only a small volume of parameter space owing to the high correlation between the components. Here, the quadrature sum of individual uncertainties would yield a large result, implying we know little about the true value of the vector, whereas the vector uncertainty size in Equation 3.7 would return a small value. This small value tells us that the volume of possible values that the vector can occupy is quite small because of the relationship between the dimensions. For a 2D vector $\vec{v}^T = (x, y)$, the vector uncertainty size is given by:

$$\langle \sigma_{\vec{v}} \rangle = \left(\sigma_x^2 \cdot \sigma_y^2 \cdot \left[1 - \rho_{x,y}^2 \right] \right)^{1/4}$$

where σ_x and σ_y are the corresponding uncertainty in each of the vector components and $\rho_{x,y}$ is the correlation coefficient between x and y.

To account for differences that individual realizations might have on the posterior `BP3M` PMs, we repeat the measurements of Figure 3.6; that is, we create 5 separate synthetic *HST* images that have 200 sources and a time baseline of 15 years. For the different realizations, the list of randomly chosen sources all come from the same synthetic catalog of COSMOS-like data, but each realization will have a slightly different set of

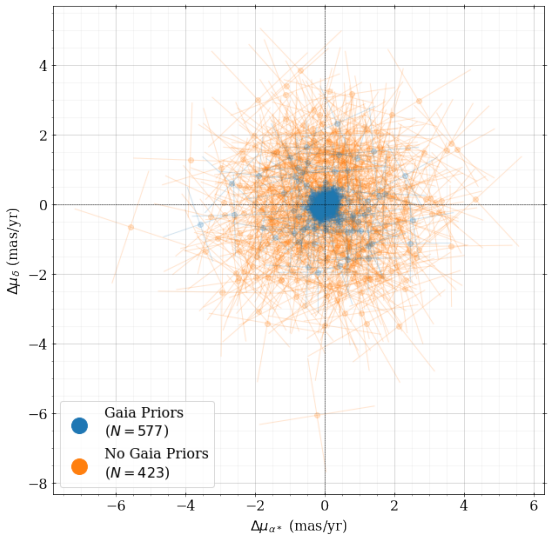

Figure 3.7: Comparison of `BP3M`-measured posterior PM's with the truth for 5 sets of synthetic *HST* images, each with 200 sources and a 15 year time baseline from *Gaia*. The points are colored by whether the `BP3M` sources have *Gaia* parallax and PM priors (blue), or are too faint to have *Gaia* parallax and PM priors (orange). The uncertainty lines are the eigenvalues of the posterior covariance matrix, scaled so that each line corresponds to 68% probability. All the points are clustered around (0,0), implying that `BP3M` recovers trustworthy PMs.

true PMs, parallaxes, positions, magnitudes, and numbers of sources with/without *Gaia* priors. The differences of the posterior PMs from the truth for these $5 \times 200 = 1000$ sources are shown in Figure 3.7, where the sources with *Gaia*-measured parallaxes and PMs are in blue and the converse are in orange. As expected for a well-behaved pipeline, the difference distribution clusters around the origin, and we see that the sources with *Gaia* priors on their motion have smaller posterior uncertainties than those without. Of course, the PMs are only 2 of the 5 dimensions of the vector measured for each source, so we also compare the 5D vector (2D position, parallax, 2D PM) to the truth, again using the distance definition of Equation 3.6; this distribution for the 1000 stars in Figure 3.7 is given as the blue histogram in Figure 3.8, and it shows remarkable agreement with

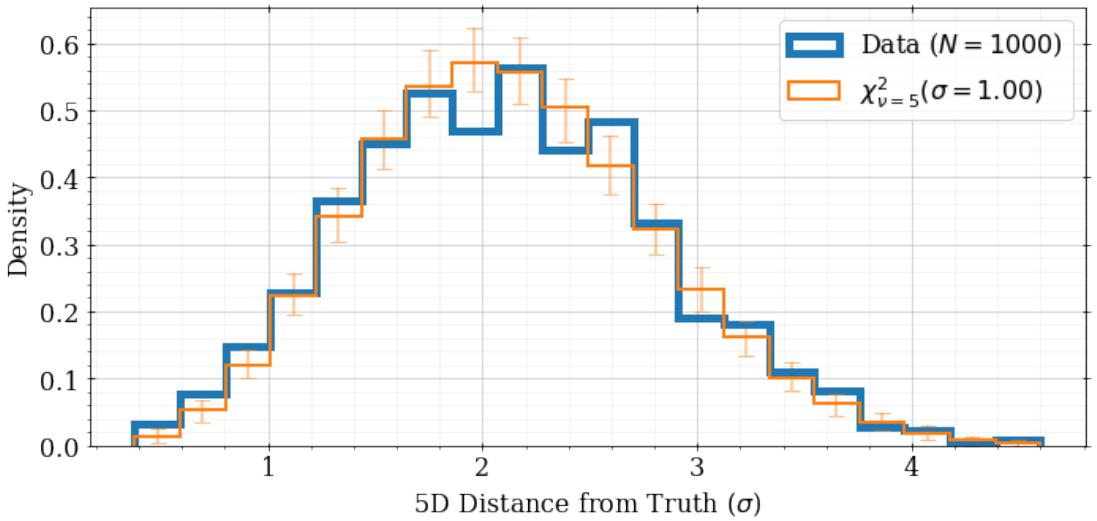

Figure 3.8: Comparison of uncertainty-scaled distance of the 5D posterior vectors (i.e. 2D position, parallax, and 2D proper motion) from the truth. The blue histogram represents the `BP3M` data for the 1000 synthetic stars in Figure 3.7, and the orange histogram is the expected distribution (χ^2 with 5 degrees of freedom and a scale of 1.0). The agreement between these two curves is evidence that the pipeline is recovering good posterior 5D vectors with realistic uncertainties.

the expected χ^2 distribution with 5 degrees of freedom and a scale of 1. These figures show that the pipeline recovers the true positions, parallaxes, and proper motions of all sources within their posterior uncertainties.

Finally, we explore the impact that the number of sources in an image have on the PM improvement factor when comparing `BP3M` to *Gaia*. An example analysis of a synthetic *HST* image with 10 sources and a 15 year time baseline is given in Figure 3.9. While the pipeline again finds good agreement between the PMs and the truth, the median PM improvement factor over *Gaia* alone is only 1.39 for the sources in this synthetic image, which is much smaller than the median factor of 8.3 seen for the sources in the $20.5 < G < 21$ mag range of Figure 3.6. This is largely

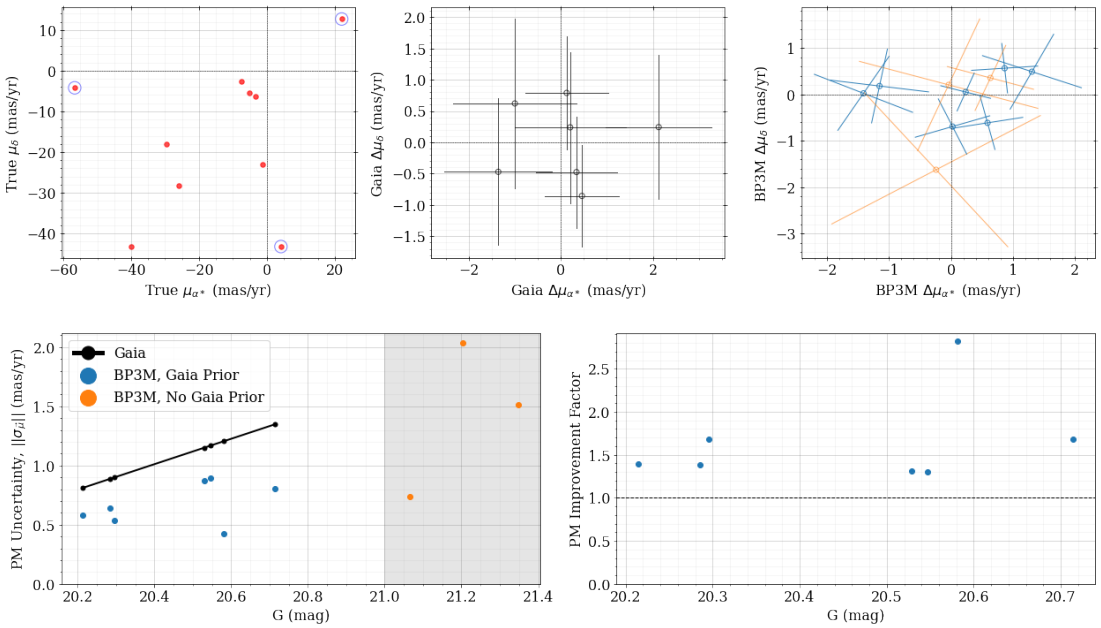

Figure 3.9: Same as Figure 3.6, but with 10 synthetic COSMOS-like stars. The PM improvement factor is less significant than in the bottom right panel of Figure 3.6, owing to the larger uncertainty in the transformation parameter fitting because of the smaller number of sources.

because the smaller number of sources aren't able to place as strong a constraint on the transformation parameters' distribution, and the larger transformation parameter distribution propagates to the PM distributions of individual sources.

We leave out example figures exploring the effect that a changing time baseline has on the posterior PMs because it follows expectations exactly; namely, for the same position uncertainty between two images, a smaller time baselines leads to larger PM uncertainties (e.g. Equation 2 of del Pino et al. 2022).

Our synthetic analysis also allows us to measure the improvement factors in the parallaxes and 2D position vectors of the sources. These results show, not unexpectedly, that the precision of these measurements is strongly tied (but not limited) to

the magnitude of the source, the time baseline between images, the number of sources per image, and the position on the sky's effect on the shape and orientation of the parallax motion. Because we cannot change time baselines of archival *HST* images, we will explore observation strategies for future missions in Section 3.7 that allow `BP3M` to maximally improve 5D precision.

That said, there are a few general lessons about trying to improve parallax and position precision that we will list here. First, it is difficult to improve on the *Gaia*-measured parallax or position vectors when using only one *HST* image (i.e. one *HST* in combination with *Gaia*). When multiple *HST* images with common targets are analysed together, then the parallaxes and positions begin to show improvement. In particular, multiple *HST* images taken very nearby in time (e.g. the multiple exposures taken to guard against cosmic rays) are useful for improving the position uncertainty, but don't offer much information to the parallax.

Next, if the time baselines from *Gaia* are taken in multiples of a year, then the parallax orbit (e.g. see Figure 3.3) is sampled at the same time, which results in no additional information about the parallax motion. Instead, if the observations are spaced around the parallax orbit to have maximum offset from *Gaia* (and each other when considering multiple *HST* exposures), then the parallax precision can be improved.

Finally, measurements that are closely spaced in time (i.e. within the same year, but still some time separation) are able to break the degeneracy between the PMs and parallaxes because the PMs – at least for low sky velocity stars – do not amount to much motion in one year, whereas the parallax motion can be significant. Conversely,

a long time baseline, regardless of the offset within each year, is able to escape the largest impact of the PM-parallax degeneracy because the PM motion is generally a more significant contribution to the total offset; even if even if the parallax motion is completely attributed to PM motion, the large time baseline serves to reduce its effect.

3.4.2 Comparison with Nearby Dwarf Spheroidals

We analyse *HST* images using `BP3M` in nearby dSph galaxies to see how the posterior PMs and parallaxes compare to *Gaia*. This serves as a test with real data, as well as more proof that `BP3M` is improving the astrometric solutions for the sources it measures. We also show how analysing multiple *HST* images together impact the resulting PM and parallax precision.

A comparison of the *Gaia* and `BP3M` PMs is presented in Figure 3.10 for the Fornax dwarf spheroidal. For the `BP3M` PMs, 6 *HST* images at the same epoch (time baseline from *Gaia* of 12.8 years) with a total of 198 unique sources are analysed together. While this figure is similar to Figure 3.6, the top panels are slightly different in that they show PM measurements instead of offsets from known PMs of synthetic data. Sources without *Gaia*-measured parallaxes and PMs are colored orange orange in the `BP3M` panel, which show larger uncertainties than the sources with *Gaia* priors in blue. The axes of the *Gaia* and `BP3M` panels have been set to cover the same range of values for easier visual comparison. The upper right panel shows the difference between the *Gaia* and `BP3M` PMs, with uncertainties that are the same as shown in the *Gaia* panel; that is, we do not combine the `BP3M` and *Gaia* uncertainties because the *Gaia* uncertainties have been used to inform the `BP3M` distributions. As before, the lower

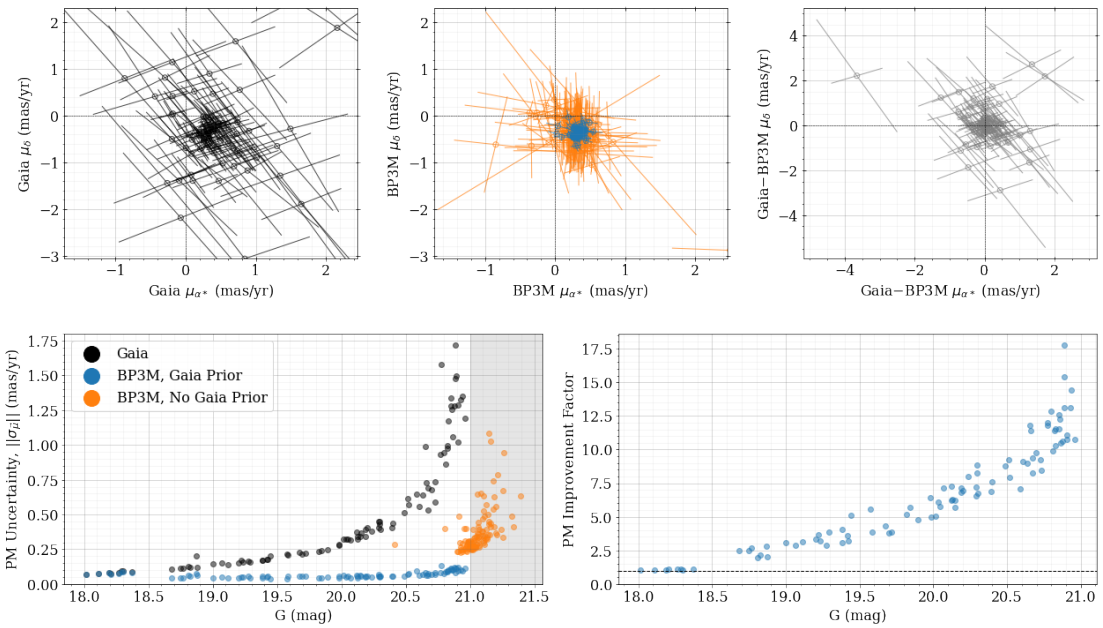

Figure 3.10: Comparison of PMs using 6 *HST* images of Fornax dSph analysed concurrently; these *HST* images were all taken using ACS/WFC in the F775W filter at the same epoch and same RA and Dec. The exposure time of each image is ~ 250 sec and the time offset from *Gaia* is 12.8 years. The panels are similar to Figure 3.6, except the upper panels show measured PM values instead of offsets from known PMs of synthetic data; the upper left panel shows *Gaia* measurements, the middle panel shows `BP3M` posteriors (colored by whether *Gaia*-measured PMs exist for each source), and the right panel shows a comparison between the *Gaia* and `BP3M` measurements (with *Gaia* PM uncertainties). The *Gaia* and `BP3M` panels covers the same range of PM values to visually show the improved clustering in the `BP3M` panel.

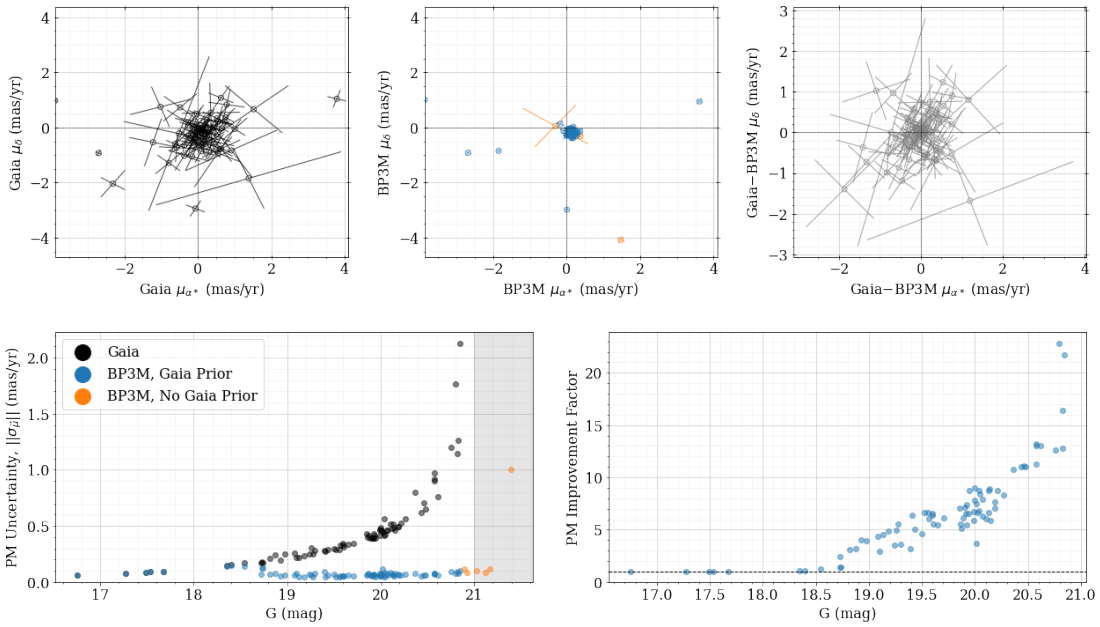

Figure 3.11: Same as Figure 3.10, but analyzing 10 *HST* images of Draco concurrently. These *HST* images were all taken using ACS/WFC in the F606W filter at the same RA and Dec, with an average exposure time of $\sim$ 215 sec. The images span three epochs, with time offsets from *Gaia* of 11.2, 9.2, and 2.2 years.

panels compare the PM uncertainties. Figure 3.11 shows another PM comparison, this time using 10 *HST* images of the Draco dwarf spheroidal, taken at 3 different epochs (baselines of 11.2, 9.2, and 2.2 years from *Gaia*).

First, these figures show a significant tightening of the PMs when *Gaia* and *HST* are combined using `BP3M`. This suggests that the increase in PM precision is not at the cost of PM accuracy. We emphasize that the sample of stars used in the `BP3M` analysis were not cleaned in any way to focus only on galaxy members, highlighting the general nature of our pipeline; for `BP3M` foreground bright stars are often useful anchors for the transformation fitting owing to their well-measured *Gaia* positions, parallaxes, and PMs.

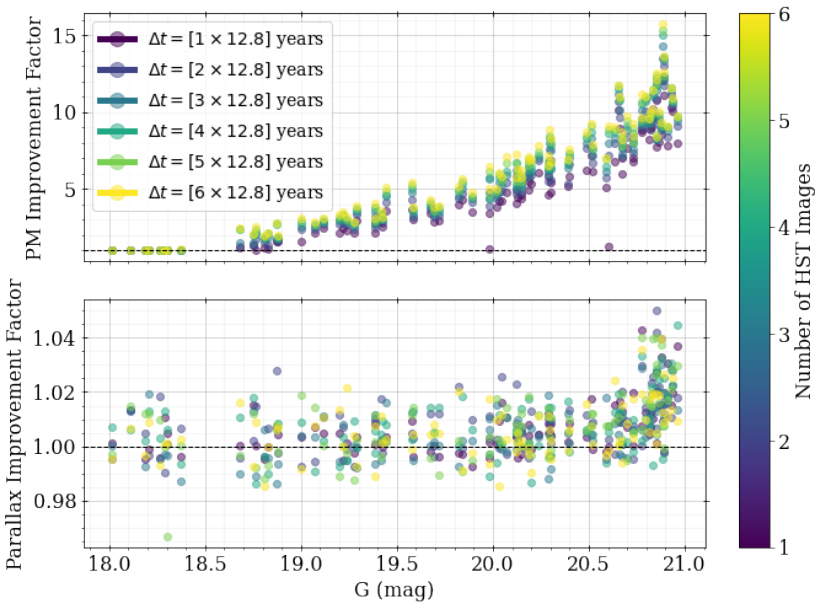

Figure 3.12: Comparison of parallax and proper motion improvement factor to *Gaia* (i.e. size of *Gaia* parallax or PM uncertainty divided by size of `BP3M` parallax or PM uncertainty) for Fornax dSph as a function of the number of *HST* images used. These measurements concern the same data presented in Figure 3.10. While the most significant improvement happens when combining a single *HST* image with *Gaia* (i.e. factor of ~ 10 for $G = 21$ mag), there is still a $\sim 25\%$ improvement between considering one *HST* image and considering six. There is also a very slight improvement on the parallax uncertainties at the faintest magnitudes ($\sim 2\% - 4\%$), though the *HST* observations occurring at the same approximate time make it difficult to improve the parallax precision.

We next repeat the `BP3M` analysis while varying the number of *HST* images considered together to explore the effect this has on the posterior PMs and parallaxes. The resulting improvement factors as a function of magnitude are given in Figures 3.12 and 3.13 for Fornax and Draco respectively, with the legend displaying the time baselines from *Gaia* for the different sets of *HST* images. In the case of Fornax, the most significant improvement in PM occurs when analysing a single *HST* with *Gaia* (i.e. a

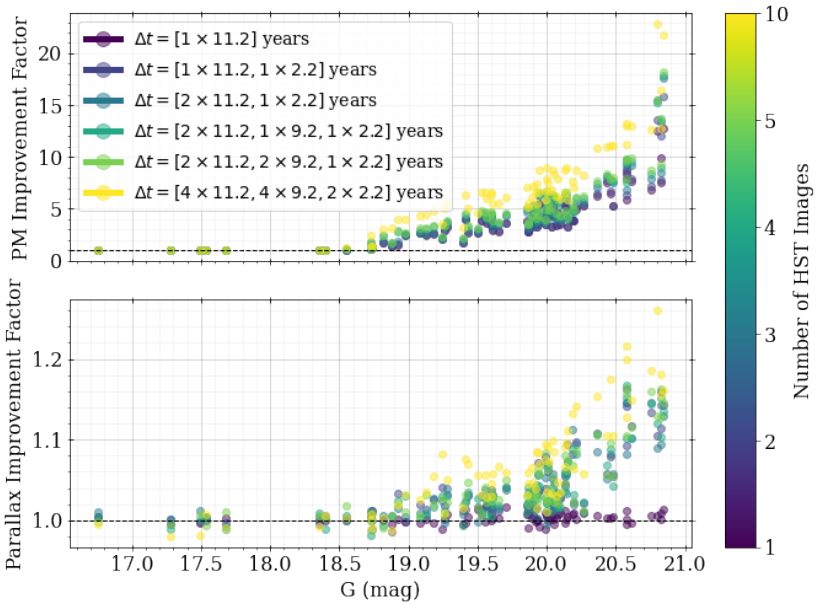

Figure 3.13: Same as Figure 3.12, but for Draco instead of Fornax. These measurements concern the same data presented in Figure 3.11. Note that the number of images jumps from 5 to 10 on the colorbar. In contract to the Fornax results, there is a more-noticeable improvement in the parallax uncertainties at the faintest magnitudes when the number of images increases. This is likely a result of the *HST* images occuring at different epochs.

median PM precision increase by a factor of ~ 8.6 for $20.5 < G < 21$ mag), though there is still an additional improvement when the remaining five *HST* images are brought in ($\sim 14\%$ improvement for for $20.5 < G < 21$ mag). The parallax uncertainties, however, do not improve very significantly for any of the images, though there is a slight, $\sim 1\%$, precision increase at the faintest magnitudes. This is likely because the *HST* images are taken at a single epoch and the time baseline is quite close to a one year multiple of *Gaia*'s observation date (i.e. only offset by 20% of a year); this means that the parallax orbit is being sampled repeatedly in a small area at the same approximate time, which

does not lead to significant increases in parallax precision.

For Draco, increasing the number of *HST* images does substantially improve the PMs for each image added; considering 10 *HST* images together with *Gaia* improves the PM precision by a median factor of ~ 13 for $20.5 < G < 21$ mag compared to a median factor of ~ 7.8 when only one *HST* image is used. The parallaxes also see noticeable change in precision as a function of *HST* images analysed. While the time offsets from *Gaia* are again only $\sim 20\%$ of a year, the *HST* images are spread over three epochs. In the case of Fornax, we effectively have measurements of positions for each source at two different times (i.e. a single epoch of *HST* images plus *Gaia*) while the Draco data have measurements of positions at four different times; for the Draco stars then, there is smaller set of parallax and PM combinations that can explain those four position measurements per source than if there were only two position measurements. It is likely that the improvement in the parallax precision also contributes to the increase in PM precision because the degeneracy between the different types of apparent motion become less entangled. However, the time offsets of the *HST* images from *Gaia* and from each other are not optimally spaced in the parallax orbit (i.e. near the apocenters), resulting in a median parallax precision increase by a factor of ~ 1.18 for the faintest magnitudes ($G > 20.5$ mag).

To confirm that the `BP3M` PMs are a true improvement in the astrometric solution compared to *Gaia* alone, we measure internal kinematics. Specifically, we fit a

2D Bayesian population model of the form:

$$
\begin{aligned}
\hat{\mu}^T &= (\hat{\mu}_{\alpha*}, \hat{\mu}_\delta) \\
\mathbf{\Sigma} &= \begin{pmatrix} \sigma^2_{\mu_{\alpha*}} & \rho \cdot \sigma_{\mu_{\alpha*}} \cdot \sigma_{\mu_\delta} \\ \rho \cdot \sigma_{\mu_{\alpha*}} \cdot \sigma_{\mu_\delta} & \sigma^2_{\mu_\delta} \end{pmatrix} \\
p(\sigma_{\mu_{\alpha*}}) &\propto 1 \\
p(\sigma_{\mu_\delta}) &\propto 1 \\
p(\rho) &\propto 1 \\
p(\hat{\mu}|\mathbf{\Sigma}) &\propto 1 \\
\vec{\mu}_i &\sim \mathcal{N}\left(\hat{\mu}, \mathbf{\Sigma}\right), \\
\vec{\mu}'_i &\sim \mathcal{N}\left(\vec{\mu}_i, \boldsymbol{V}_{\vec{\mu_i}}\right),
\end{aligned}
\tag{3.8}
$$

where $\hat{\mu}$ and $\mathbf{\Sigma}$ are the population mean and covariance matrix we wish to measure, $\vec{\mu}_i$ is the true PM of star i, $\vec{\mu}'_i$ is the observed PM of star i with covariance matrix of $\boldsymbol{V}_{\vec{\mu_i}}$, and we have chosen to use flat priors on the components of the $\mathbf{\Sigma}$ matrix. This model allows us to capture the population distribution that describes the data while also accounting for the 2D uncertainties in the individual PM measurements. We fit for the population mean and covariance matrix when using both the *Gaia* PMs and the `BP3M` PMs. We restrict the analysis to stars within 2σ of 2D distance of the median PM of the population to ensure a relatively clean kinematic sample. The result of this fitting process for both Fornax and Draco are summarized in Table 3.2.

In general, there is good agreement between the measured population distribution parameters, though the uncertainties on the `BP3M`-measured population statistics

Table 3.2: Result of Bayesian population fit to the PMs for Dwarf Spheroidals.

Name	PM Measurement	N_*	$\mu_{\alpha*}$ (mas $\cdot$ yr^{-1})	μ_δ (mas $\cdot$ yr^{-1})	$\sigma_{\alpha*}$ (mas $\cdot$ yr^{-1})	σ_δ (mas $\cdot$ yr^{-1})	ρ
Fornax	*Gaia*	45	0.331 ± 0.020	-0.379 ± 0.032	$0.043^{+0.024}_{-0.020}$	$0.091^{+0.035}_{-0.033}$	$0.273^{+0.430}_{-0.540}$
	BP3M	74	0.328 ± 0.007	-0.367 ± 0.009	$0.033^{+0.008}_{-0.007}$	0.033 ± 0.011	$0.725^{+0.174}_{-0.285}$
Draco	*Gaia*	54	$0.088^{+0.041}_{-0.040}$	$-0.156^{+0.039}_{-0.041}$	$0.125^{+0.056}_{-0.052}$	$0.086^{+0.059}_{-0.044}$	$0.351^{+0.416}_{-0.608}$
	BP3M	51	0.135 ± 0.009	-0.203 ± 0.010	$0.020^{+0.011}_{-0.010}$	$0.030^{+0.014}_{-0.013}$	$-0.205^{+0.563}_{-0.455}$

are significantly smaller (factor of $\sim 3-4$ in the population mean and scatter) on account of the improved PM uncertainties. These results suggest that `BP3M` is not only shrinking the PM uncertainties, it is also truly moving the *Gaia* PMs closer towards their true values.

3.4.3 Comparison with QSOs

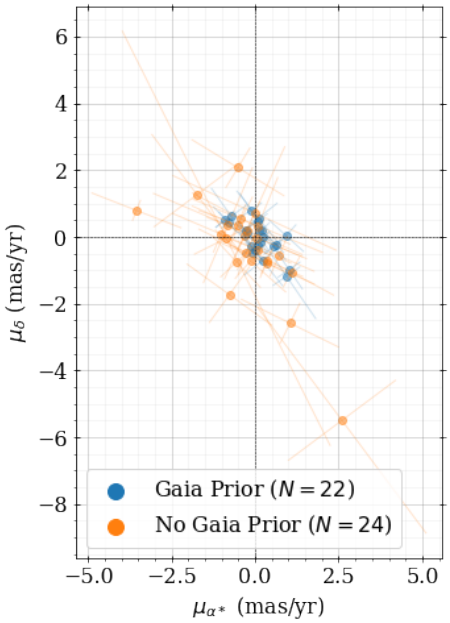

Figure 3.14: Proper motions of the 46 sources in the COSMOS field that are nearest to a previously-identified QSO. The PMs are colored by whether *Gaia*-measured PMs and parallaxes were available as priors. The proper motion uncertainties are such that all of these proper motion measurements are consistent with stationary.

As an additional method of validating the `BP3M` posterior PMs and their uncertainties, we cross-match the ~ 2000 unique COSMOS sources mentioned in Section 3.2 with the MILLIQUAS QSO catalog[10] (Flesch 2023). There are 46 sources in the QSO

[10] https://heasarc.gsfc.nasa.gov/W3Browse/all/milliquas.html

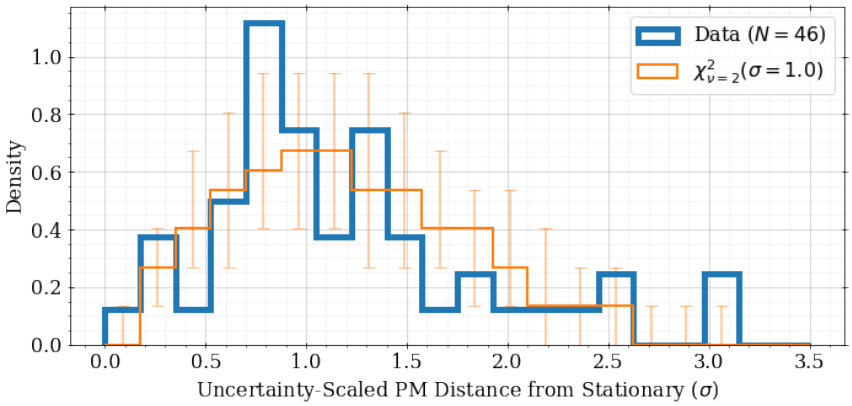

Figure 3.15: QSO proper motion distance from stationary, scaled by their uncertainties. The orange histogram shows the expectation if the uncertainties explain all the non-zero sizes of the PMs, with errorbars giving the 68% region on the density in each bin for a sample size of $N = 46$; the two distributions agree quite well by eye, suggesting that the posterior PM uncertainties are reasonable and trustworthy.

list that have COSMOS sources within 10 mas, so we run the *HST* images that contain these sources through the `BP3M` pipeline, with the resulting posterior PMs shown in Figure 3.14. As expected, the PMs are closely distributed around the origin.

Next, we use the PM uncertainties to measure the 2D distances of the PMs from stationary, and then compare these measurements to the expected distribution; the result of this process is shown in Figure 3.15. The expected distribution (orange histogram) for an equivalent sample size agrees quite well with the measured one (blue histogram). This is additional and external evidence that the `BP3M` posterior PM distributions are reasonable and trustworthy.

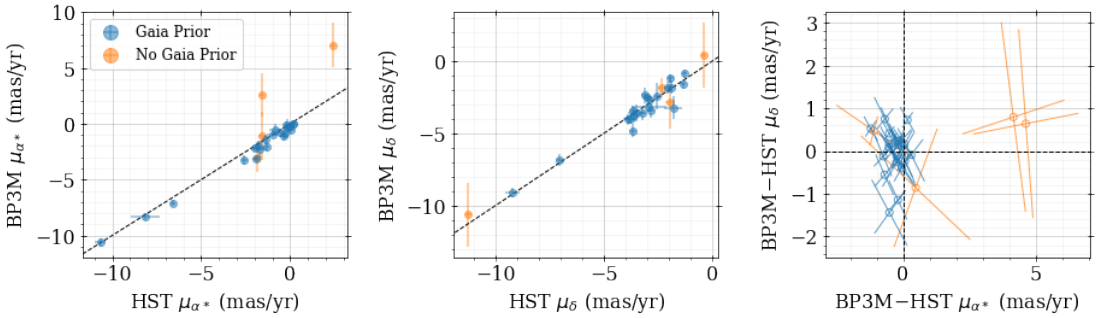

Figure 3.16: Posterior `BP3M` PMs for 25 sources in the COSMOS field that are shared with the HALO7D survey. The "HST" designation in the axis labels refer to the PMs of Cunningham et al. (2019b), which use *HST* images alone to measure PMs.

3.4.4 Comparison with *HST*-measured PMs

Like in Section 3.4.3, we identify sources in our COSMOS sample that are shared with an external survey; in this case, we cross-match with the COSMOS sample of the HALO7D survey (Cunningham et al. 2019a,b; McKinnon et al. 2023). The HALO7D PMs are described in Cunningham et al. (2019b); to summarize, the PMs are measured using multiple epochs of *HST* imaging with the absolute reference frame being defined by registration with background galaxies. These *HST*+*HST* PMs are therefore a result of traditional PM measurement techniques. We compare the 25 sources that are shared between our sample and HALO7D in Figure 3.16, which shows the `BP3M` PMs have strong agreement with their *HST*-only counterparts. As in the other comparisons we have presented, the distribution of 2D differences between the PM measurements agree statistically with our expectations, providing the final piece of validation for the `BP3M` outputs.

3.5 Proper Motions in Sparse Fields: COSMOS

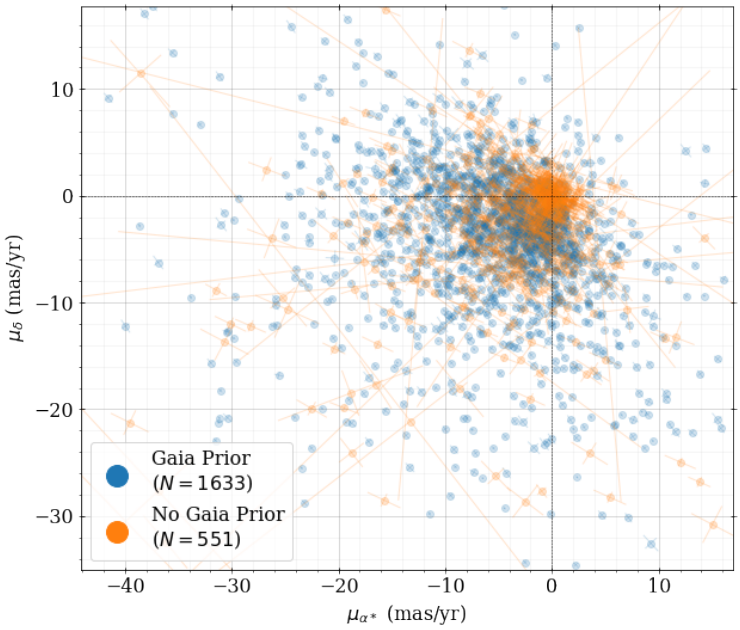

Figure 3.17: Posterior `BP3M` PMs for the 2184 unique sources in the COSMOS field (i.e. same sources referred to in Figure 3.1). While each source may fall in multiple HST image, we only show the PMs measured from analysing a single HST image with Gaia. The PMs are colored by whether *Gaia*-measured PMs and parallaxes were available as priors. The PM uncertainties of the sources with *Gaia* priors (blue points) are, on average, much smaller than those without *Gaia* priors (orange points).

Using the *HST* images of the COSMOS field discussed in Section 3.2.1, we run `BP3M` on all images within 0.5° of the field's center. This includes all regions of COSMOS with multi-epoch *HST* imaging (near the center), as well as single-epoch-only *HST* regions. The posterior PMs we measure for the 2184 unique sources that are found in common between the *HST* images and *Gaia* are shown in Figure 3.17, where the sources are colored by whether or not they have *Gaia*-measured PMs and parallaxes. As

a quick sanity check, this PM distribution visually agrees with our expectations based on the synthetic COSMOS data created using previously-measured velocity distributions for the MW thick disk and stellar halo[11]. While each source may appear in multiple *HST* frames, we only show the results from analysing each *HST* image separately, and the displayed PM is from the posterior distribution with the smallest uncertainty.

We also compare the `BP3M` PMs and uncertainties to the corresponding *Gaia* PMs in Figure 3.18, where we again see excellent agreement between the two samples (top panels). The middle and bottom panels show the improvement on the PM uncertainty that `BP3M` measures by combining *HST* and *Gaia*, especially for the $\sim 25\%$ of sources that had no *Gaia*-measured PMs; the median PM uncertainty for sources without *Gaia* priors is found to be ~ 1.12 mas$\cdot$yr^{-1}. While many of the brightest magnitude sources – those most likely to be at or near saturation in the *HST* images – show no improvement on *Gaia*'s PMs, the faintest magnitude sources ($20.25 < G < 20.75$ mag) have a median PM improvement factor of 2.62. As mentioned for the small-source-count case in Figure 3.9, the less extreme improvements in PM we see in sparse fields are likely the result of larger uncertainty in the transformation parameters, which propagates to the PM uncertainties of all sources in an image.

The `BP3M` transformation parameters could likely be pinned down much more precisely by identifying stationary sources in each image. For example, by finding the centroids in each *HST* image that are clearly associated with extended sources and background galaxies, then updating the priors in parallax and PM to a narrow distri-

[11] see the 2D PM distribution in Appendix 3.9.2, but note that the points are colored based on belonging to the halo or thick disk, not whether there are *Gaia* priors on parallax and PM.

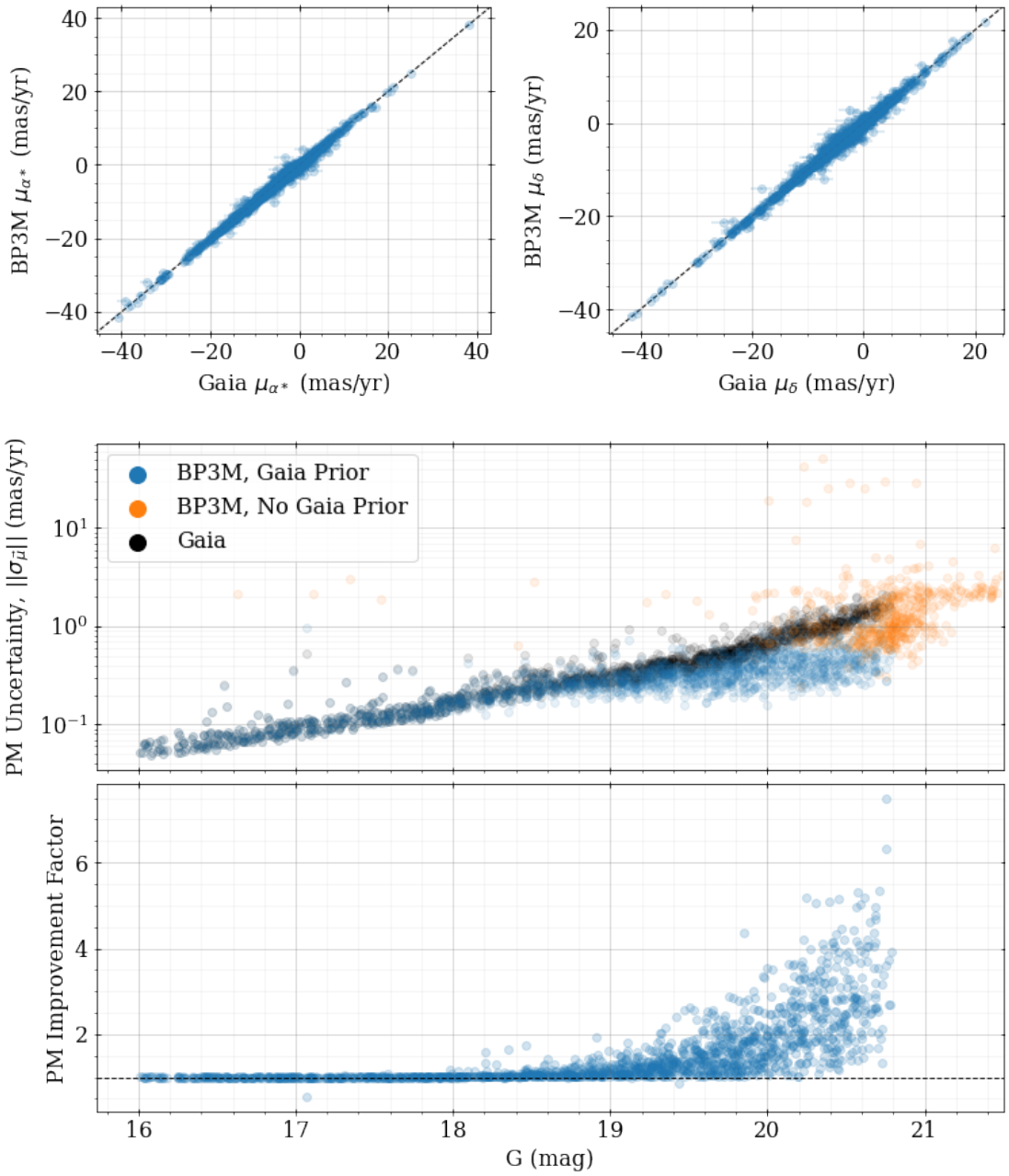

Figure 3.18: Comparison of posterior PMs and corresponding uncertainties between *Gaia* and `BP3M` for the same COSMOS sources as Figure 3.17. The different PM values agree with each other within their uncertainties across a range of PM sizes. For many of the bright sources ($G < 19$ mag), their *HST* detection has a large position uncertainty, which leads to no improvement over the Gaia-measured PMs. For fainter targets ($G > 19$ mag), where the *Gaia* PM uncertainty increases and the *HST* positions are better measured, there is a significant improvement in the PM uncertainty. Approximately 25% of the COSMOS sources have no *Gaia*-measured PM, and these sources have a median PM uncertainty of ~ 1.12 mas $\cdot$ yr^{-1}.

Table 3.3: Median uncertainty in proper motion for the COSMOS stars in *Gaia*, BP3M with *Gaia* priors, and BP3M without *Gaia* priors in the $20 < G < 21.5$ mag range. While the median uncertainties in the *Gaia* sample are smaller than the medians from the BP3M without *Gaia* priors sample, the former has significantly fewer stars than the latter and does not reach as faint. As shown in the middle panel of Figure 3.18, the *Gaia* COSMOS data extend only to $G \sim 20.75$ mag, while the BP3M sample extends to $G \sim 21.5$ mag.

Sample	Number of Sources	Median $\|\|\sigma_{\vec{\mu}_i}\|\|$ (mas/yr)
Gaia	375	1.01
BP3M, Gaia Prior	375	0.43
BP3M, No Gaia Prior	534	1.10

bution around $\vec{0}$, the transformation solution can be improved by these anchor points. This would reduce the posterior width of the transformation parameter distributions for each image, especially for particularly sparse images, which would also reduce the posterior widths in the PMs for all of the sources in an *HST* image. Exploring the impact that known stationary targets can have on the BP3M PMs will be the focus of future work.

The median PM uncertainties in the $20 < G < 21.5$ mag range are summarized for the COSMOS sample in Table 3.3. First, the BP3M sample that has *Gaia* parallax and PM priors has a significant increase in precision compared to *Gaia* alone. Second, the *Gaia* COSMOS sample ($20 < G < 20.75$ mag) has a similar median PM uncertainty to the BP3M sample without *Gaia* priors ($20.75 < G < 21.5$ mag), but the latter has contains significantly more stars. Effectively, BP3M substantially increases the number of stars in a single LOS with *Gaia*-quality PMs at the faintest magnitudes.

An typical MW stellar halo isochrone (i.e. Age = 12 Gyr and [Fe/H] = -1.2 dex, see Appendix 3.9.2) has a main sequence turn-off (MSTO) around an ab-

solute magnitude of $M_G \approx 3.9$ mag. For MSTO stars with apparent magnitudes in the $20 < G < 20.75$ mag range, this translates to a distance range of $\sim 17 - 23$ kpc. Using the median PM uncertainties in Table 3.3, this would imply an uncertainty in velocity on the sky in the range of $\sim 79 - 112$ km $\cdot$ s^{-1} for the *Gaia*-alone sample versus a $\sim 33 - 48$ km $\cdot$ s^{-1} range for the `BP3M` sample with *Gaia* priors. For the `BP3M` sample without *Gaia* priors, the apparent magnitude range of $20.75 < G < 21.5$ mag translates to a distance range of $\sim 23 - 33$ kpc, and results in a sky velocity uncertainty range of $\sim 122 - 172$ km $\cdot$ s^{-1} when using the median PM uncertainty for this subsample.

Like with the dwarf spheroidal data, many of these COSMOS sources appear in multiple *HST* images. Future work will analyse these images concurrently, allowing for the best possible PM measurements of the COSMOS sample. This may also likely yield improvements on the parallax precision, especially for sources in the multi-epoch *HST* regions of the field. Finally, there are many *HST* sources that are too faint to have *Gaia* counterparts, but they nonetheless appear in multiple *HST* frames. For these sources, future versions of `BP3M` can use the *HST*-to-*Gaia* transformation parameters for each image to identify the positions of these sources in a global reference frame as a function of time, and thus determine their best fit PMs and parallaxes.

3.6 Time Stability of Transformation Parameters

While the offset and rotation terms of the transformation parameters will always need to be measured for every image, we explore the possibility of describing the pixel scale ratio and the skew terms as a function of time and *HST* orienta-

Table 3.4: Number of *HST* images analysed with `BP3M` in different nearby galaxies, clusters, and MW halo fields in different filters of ACS/WFC.

Name	Number of *HST* Images			
	F555W	F606W	F775W	F814W
COSMOS	0	0	0	640
Fornax dSph	0	0	6	0
Draco dSph	0	38	0	0
Sculptor dSph	0	0	11	0
47 Tucanae	10	17	2	22
Arp2	0	0	0	6
E3	0	5	0	0
IC10	0	0	0	6
IC1613	0	0	0	24
Leo A	0	0	0	12
Leo I	0	0	0	21
Leo T	0	35	0	12
NGC2419	0	0	0	6
NGC205	0	1	0	0
Pal1	0	6	0	6
Pal2	0	0	0	5
SAG DIG	0	4	0	9
Terzan8	0	2	0	6
TOTAL	10	108	19	775

tion. The transformation parameter measurements come from `BP3M` analyses of various nearby galaxies, clusters, and patches of the MW stellar halo using different filters of ACS/WFC. The breakdown of number of *HST* frames per field and per filter is summarized in Table 3.4.

In particular, the PSR is expected to change as a function of *HST*'s oribt; specifically, a "breathing" of the telescope as it expands and contracts from changes in solar radiation as it passes in and out of the Earth's shadow. We attempt to quantify this breathing by fitting a time-varying sinusoid to the PSRs we measure from 775 $F814W$ ACS/WFC images, the result of which is shown in Figure 3.19. While we recover the nominal *HST* orbit time of ~ 95 minutes as the best fit-period, we do not

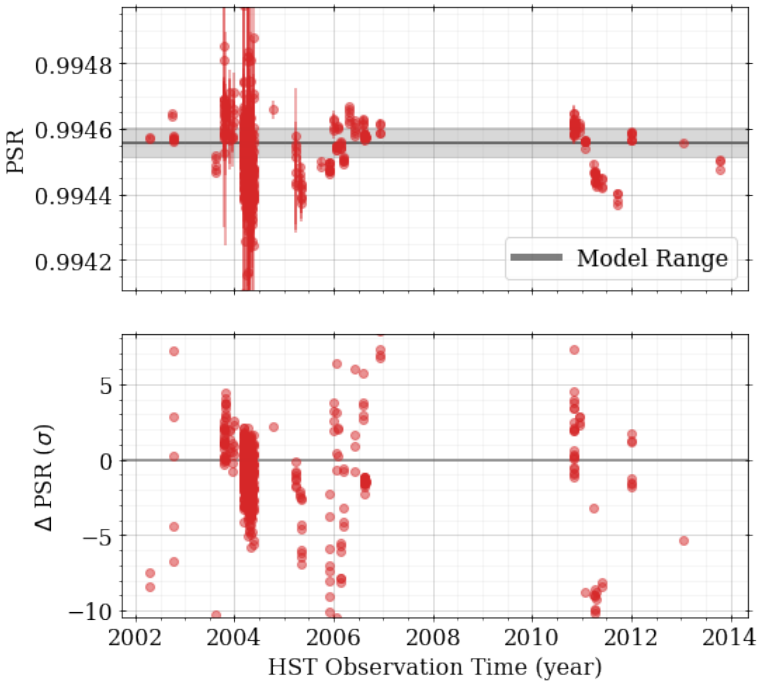

Figure 3.19: PSR measurements from 775 F814W filter images taken with of ACS/WFC as a function of observation time and a comparison of the data with a time-varying sinusoid of the form $A \cdot \sin(t \cdot 2\pi/P - \phi) + H$. The model has a best fit period of ~ 95 min so only the best fit height and amplitude are displayed with the grey line and shaded region. From the uncertainty scaled residuals in lower panel, we see that the data are not well-described by this model.

see good agreement between the data and the model; the residuals suggest that the PSR uncertainties would need to be a factor of ~ 2.6 times larger to explain the disagreement. Allowing for more complicated models, we find that the best function to describe the PSR measurements versus time is of the form

$$\mathrm{PSR}(\theta, t) = \mathrm{PSR}_{0^\circ,2000} + m_{\mathrm{PSR}} \cdot (t - 2000) + (A_{2000} + m_A \cdot (t - 2000)) \cdot \sin\theta \quad (3.9)$$

where θ refers to the same *HST* angle in Equation 3.2, $\mathrm{PSR}_{0^\circ,2000}$ is the pixel scale ratio at *HST* angle of $\theta = 0^\circ$ in the year 2000, A_{2000} is the sinusoid amplitude in the

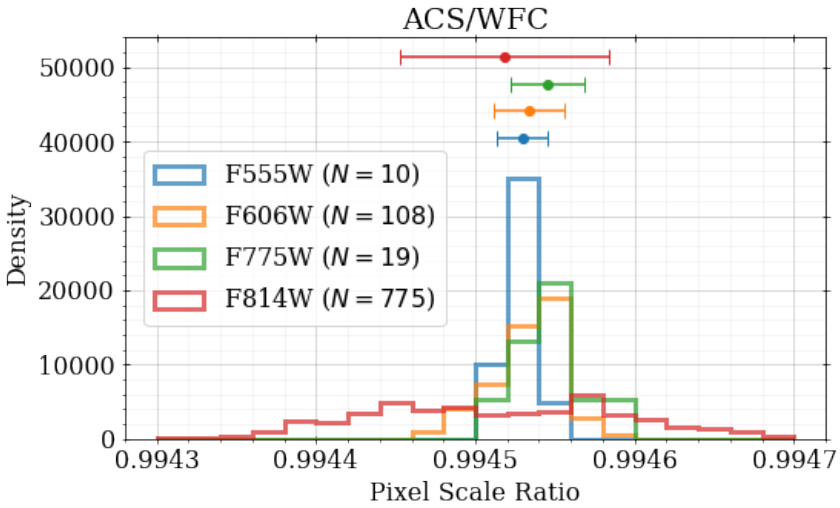

Figure 3.20: Histogram of the measured PSRs from different filters of ACS/WFC. Above each histogram is the population mean and scatter for that filter as listed in Table 3.5.

year 2000, and $(m_{\rm PSR}, m_A)$ are the slopes of the PSR and amplitude with time. This functional form allows for the sinusoid amplitude to change linearly with time, as well as for the vertical offset to change linearly with time. When we fit this model to the data, we find that the PSR uncertainties would need to be 1.5 to 2 times as large as they currently are, which is an improvement over the "breathing" model but is still unsatisfactory in terms of describing the data. This may be in line with recent findings; Hoffmann & Kozhurina-Platais (2020), for example, measure the PSR when mapping 47 Tucanae *HST* images onto *Gaia* DR2 (see their Figure 3), which seems to show a trend that is more complicated than a single sinusoid or linear relationship as a function of time.

From our initial tests, it appears that the best approach for measuring transformation parameters is to fit the PSR independently in each image, though the relatively

Table 3.5: Result of Bayesian population fit to the Pixel Scale Ratios in different filters of ACS/WFC.

Filter	N_{images}	Mean	Scatter
F555W	10	0.994530	1.6×10^{-5}
		$\pm 5 \times 10^{-6}$	$\pm 4 \times 10^{-6}$
F606W	108	0.994533	2.2×10^{-5}
		$\pm 2 \times 10^{-6}$	$\pm 1 \times 10^{-6}$
F775W	19	0.994546	2.3×10^{-5}
		$\pm 5 \times 10^{-6}$	$\pm 4 \times 10^{-6}$
F814W	775	0.994518	6.5×10^{-5}
		$\pm 2 \times 10^{-6}$	$\pm 2 \times 10^{-6}$

narrow range of values indicate that we could benefit by using a fairly strong prior on this parameter. The PSR population statistics (i.e. mean and scatter) that describe each of the filters that we have measurements for are presented in Table 3.5 and Figure 3.20; these population statistics are measured using a 1D analog of the model in Equation 3.8.

We next explore the on- and off-axis skew terms. Beginning with some set of transformation parameters for some orientation of the telescope

$$\boldsymbol{R} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{3.10}$$

a set of on- and off-axis skew values are defined by the on and off diagonal elements respectively, as given by Equation 3.2. It is fairly trivial to show that applying a rotation (remembering that θ is positive in our definition when clockwise) to $\boldsymbol{R}$, i.e.

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \cdot \boldsymbol{R} \tag{3.11}$$

causes a change in the skew terms such that the new set of skews $(\mathrm{on}', \mathrm{off}')$ are related to the previous set of skews, $(\mathrm{on}, \mathrm{off})$, by that same rotation matrix:

$$\begin{pmatrix} \mathrm{on}' \\ \mathrm{off}' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} \mathrm{on} \\ \mathrm{off} \end{pmatrix}. \tag{3.12}$$

If the skew terms aren't changing significantly over time, then Equation 3.12 implies that we should be able to fit for some $(\mathrm{on}_0, \mathrm{off}_0)$ vector that every other set of skew measurements are just some rotation away from. To be careful, we assume that the $(\mathrm{on}_0, \mathrm{off}_0)$ vector is different for the various *HST* instruments/detectors and filter and instrument; in practice, we only have dense enough sampling in time and *HST* rotation to apply this analysis to the $F814W$ ACS/WFC images in our sample.

We fit for the $(\mathrm{on}_0, \mathrm{off}_0)$ vector using the same 775 $F814W$ ACS/WFC images that were used to measure the PSR population statistics, and a comparison of the model and data are shown in Figure 3.21. While the model does a good job of tracing the on and off skews as a function of rotation, the (uncertainty-scaled) residuals are quite large, and it appears as though there may be a trend in the residuals as a function of observation time (3rd panel from the top). The uncertainty-scaled 2D distance of the data from the Equation 3.12 model is shown in Figure 3.22, where the orange histogram shows the expected distribution if the model described the data within their uncertainties, and the green histogram is the best fit distribution. Only residuals with 2D distances less than 5σ are included to enable good agreement between the shape of the best fit distribution and the data histogram, which removes 57 relatively large

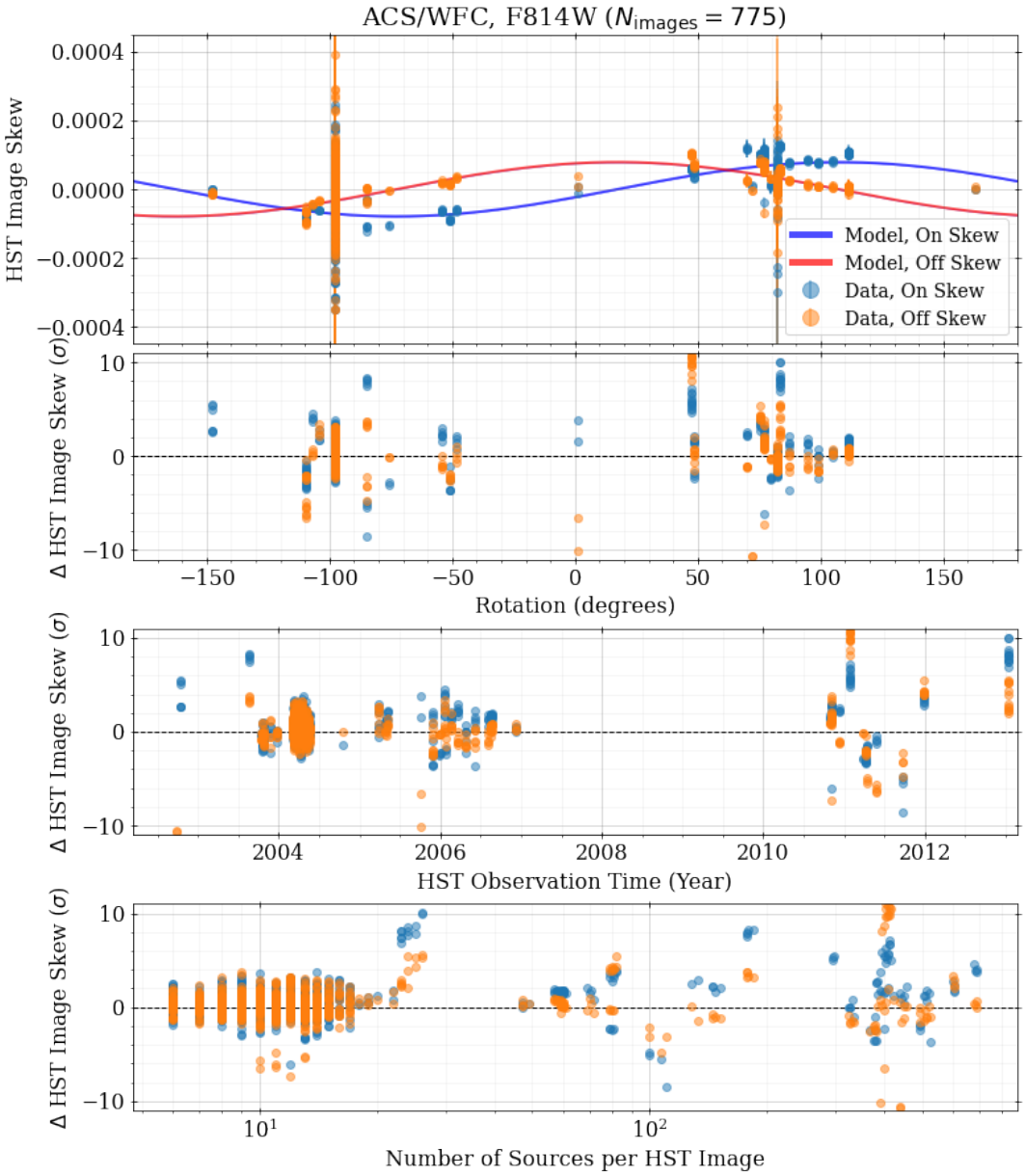

Figure 3.21: Best fit to the on and off skew measurements of F814W filter images taken with of ACS/WFC as a function of *HST* rotation as described by Equation 3.12. While there is generally good agreement between the best fit model and data, there are some points that disagree, and there are potentially trends in the residuals with the number of stars in each image and the time the observation was taken.

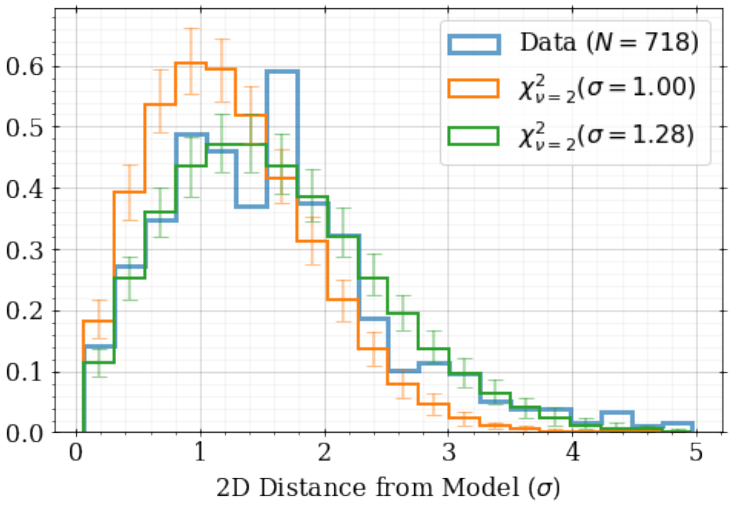

Figure 3.22: 2D residual distance from Figure 3.21 with the expected distribution (orange) and the best fit one (green). We have removed 57 of the 775 measurements where the residual 2D distance is greater than 5σ. The blue data distribution does not match the expectation, instead, its best fit curve implies that the skew term distribution width needs to be a factor of 1.28 larger.

outliers. The best fit distribution suggests that the on and off axis uncertainties would need to be a factor of 1.28 times larger to account for the differences between the model and data.

Taking the skew model one step further, we introduce time evolution by allowing the $(\text{on}_0, \text{off}_0)$ vector to change linearly with time:

$$\begin{pmatrix} \text{on}_{(\theta,t)} \\ \text{off}_{(\theta,t)} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} m_{\text{on}} \cdot (t - 2000) + \text{on}_{(0^\circ,2000)} \\ m_{\text{off}} \cdot (t - 2000) + \text{off}_{(0^\circ,2000)} \end{pmatrix} \tag{3.13}$$

where $(\text{on}_{(0^\circ,2000)}, \text{off}_{(0^\circ,2000)})$ are the skew values at 0° in the year 2000, and (m_{on}, m_{off}) are the time slopes of the on and off skew terms respectively. The residuals of this model are shown in Figure 3.23. Now, much of the residual trend with observation time has

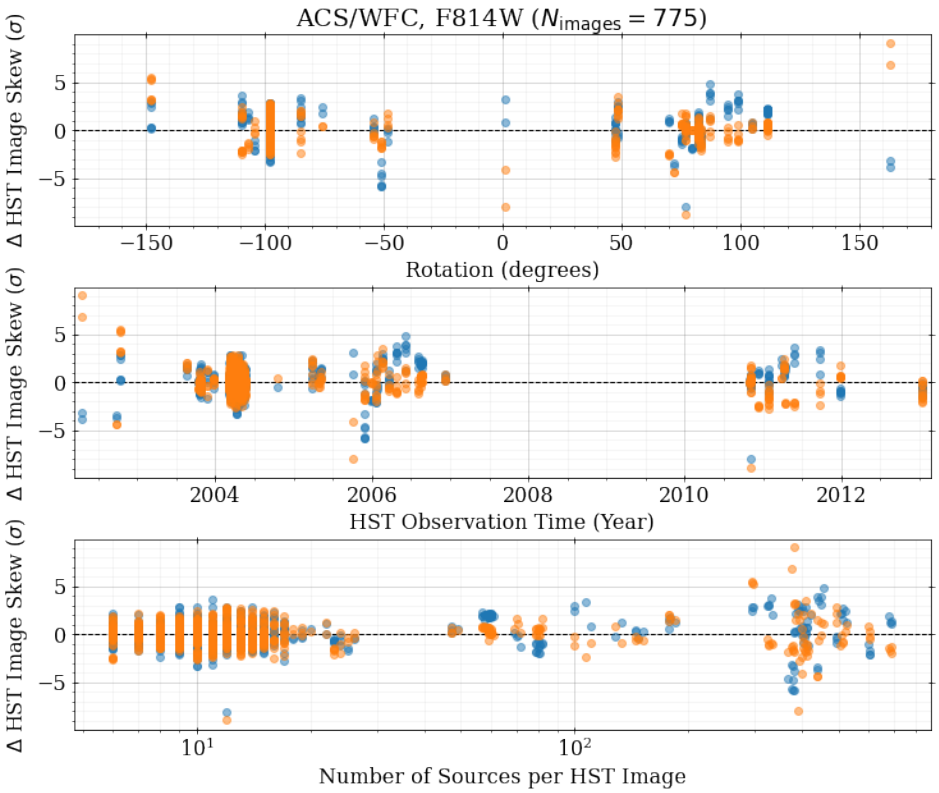

Figure 3.23: Same as the lower three residual panels of Figure 3.21, but now using the time-varying model of Equation 3.13. The residuals show that the model does a better job of describing the skew terms over time (middle panel) compared to the time-independent model of Equation 3.12.

been removed (middle panel). Figure 3.24 shows the 2D distance residuals between the model and the data; the best fit residual distribution is smaller – requiring the skew uncertainties to be only 1.15 times larger than measured – and only 13 skew measurements are removed for having distances greater than 5σ. Given a model that perfectly describes the data within their uncertainties, we would expect that 0 of the skew measurements would be removed and the best fit residual distribution would have

a width of 1.

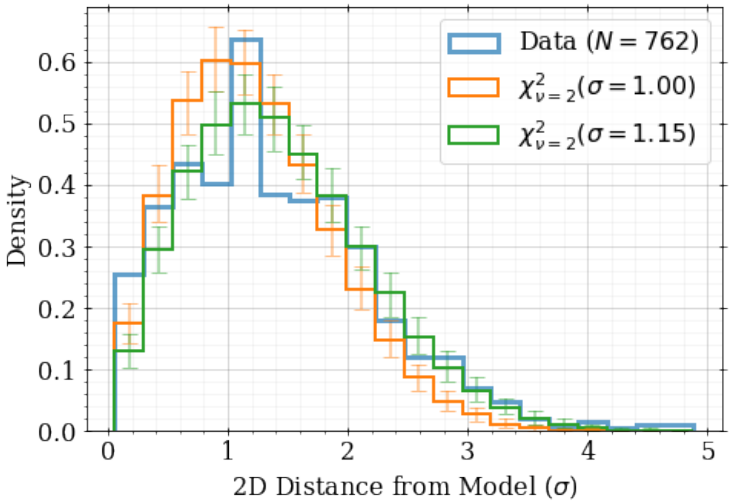

Figure 3.24: Same as Figure 3.22, but now using the residuals of Figure 3.23 from the time-varying model of Equation 3.13. We have removed 13 of the 775 measurements where the residual 2D distance is greater than 5σ. While the blue data distribution does not match the expectation (best fit σ is greater than 1), its shape has better agreement with the best fit curve and has a smaller width than the results from the non-time varying residuals of Figure 3.22.

We repeat this skew fitting process for increasing polynomial orders of the time term in the $(\text{on}_0, \text{off}_0)$ vector, keeping track of the number of measurements of the 775 total within 5σ of the model as well as the Uncertainty Scale factor. The results of this are summarized in Table 3.6. In the highest order cases, there are still a small number of measurements that are outliers, and the best models imply that the uncertainty on the skew terms should be inflated by 7% (i.e. posterior covariance multiplied by a factor of 1.07^2).

These results are quite promising in that they suggest the skew terms can almost completely be described as a function of *HST* rotation and time. In this case,

Table 3.6: Results of fitting different time-varying models to the on- and off-axis skew terms using 775 $F814W$ ACS/WFC images. The first column is the order of time-variation in the $(\text{on}_0, \text{off}_0)$ vector; the 0th and 1st time order entries correspond to the models in Equation 3.12 and 3.13 respectively. The second column is the number of measurements, of the 775 total, that are within 5σ of the current best fit model, which are used to estimate the best fit Uncertainty Scale (e.g. the σ values of the green distributions in Figures 3.22 and 3.24) in the third column. A model that completely describes the data would have 775 in the second column and 1.00 in the third column.

Time-Variation Order	$N_{\text{images}} < 5\sigma$	Uncertainty Scale
0	718	1.28
1	762	1.15
2	770	1.10
3	769	1.07
4	769	1.07
5	771	1.07

2 of the 6 transformation parameters are completely determined – or, can have an extremely strong prior placed on them – which could significantly aid in our ability to measure the remaining parameters and therefore the stellar motions. Future work should focus on getting more complete time and *HST* rotation sampling to better constrain these relationships, as well as to make similar measurements using the additional *HST* instrument, detector, and filter combinations.

3.7 Modelling Future Missions

The mathematics presented in Section 3.3 are quite general in that they apply to any two or more sets of position measurements of stars. In this way, we are able to model the effect that future missions (e.g. *Roman*) will have on PM measurements for different time baselines, telescope orientations, and transformation parameters. While it is not required to consider synthetic *Roman* measurements in combination with *Gaia*,

we focus on comparisons with *Gaia* to show the improvement that *Roman* will bring in PMs, and emphasize how careful planning on the observation times allow for improved constraints on parallax measurements.

When it comes to making synthetic, COSMOS-like *Roman* observations, we follow the same process as with the synthetic *HST* images explained in Appendix 3.9.2. The main differences are that we use the appropriate pixel scale (i.e. 110 mas $\cdot$ pixel^{-1}) and (i.e. 18 CCDs with 4096 $\times$ 4096 pixels arranged in a 6 $\times$ 3 layout) detector dimensions as stated by the *Roman* specifications (Spergel et al. 2015). For the position uncertainty of centroids in *Roman* images, we assume an average uncertainty of 0.01 pixels, which is approximately what can be measured with *HST* (e.g. Bellini et al. 2011). While we recognize that the position accuracy changes as a strong function of signal-to-noise/magnitude, we adopt a single value instead of modelling uncertainty versus magnitude to show a reasonable but optimistic set of measurements. With the currently planned launch of late-2026/early-2027, we choose to set the first observation time as 2027, with followup observations near the middle of the mission lifetime (i.e. 2029), and again at the end (i.e. 2032).

Comparing the FOV of *HST*'s ACS/WFC to that of the *Roman* detector, the novel observatory will see an area of sky that is 89.5 times larger than *HST*. However, each 4096 $\times$ 4096 CCD in *Roman*, which sees 4.84 times more sky than *HST*'s ACS/WFC, will likely need to be fit independently for its transformation parameters (e.g. small rotational offsets and differences in skew and focus). Because there are a median of 10 sources in common between *HST* and *Gaia* in COSMOS, we assume that

an equivalent observation with *Roman* would have 48 sources in common with *Gaia* per CCD. We generate synthetic images at the 3 *Roman* epochs, allowing for some offsets in time to test different strategies to improve the parallax measurements. Specifically, we make the observations occur at

1. All *Roman* observations at year-multiples of the *Gaia* observation time,
   $\Delta\ t = [11.0, 13.0, 16.0]$ years;

2. All *Roman* observations at half-a-year offset from the *Gaia* date,
   $\Delta\ t = [11.5, 13.5, 16.5]$ years;

3. The first *Roman* observations at the parallax orbit apocenter farthest from *Gaia*, then offset by a quarter year for the other epochs, $\Delta\ t = [11.39, 13.64, 16.14]$ years;

4. The first *Roman* observations at the parallax orbit apocenter farthest from *Gaia*, then alternating offsets by half a year for the other epochs,
   $\Delta\ t = [11.39, 13.89, 16.39]$ years.

The resulting position, parallax, PM uncertainties from analysing these *Roman* observations concurrently is given in Figure 3.25, where the two bottom-most panels show the 3D (i.e. parallax, PM) and 5D (i.e. position, parallax, PM) vector uncertainties. This figure reveals many important lessons about the impact that relatively small time offsets of observations within a year have on the stellar measurements.

First, our particular choices in position uncertainty and time baselines have resulted in minimal improvement in the position uncertainty of the sources (top panel). In the second panel from the top, however, the different observing strategies lead to vast

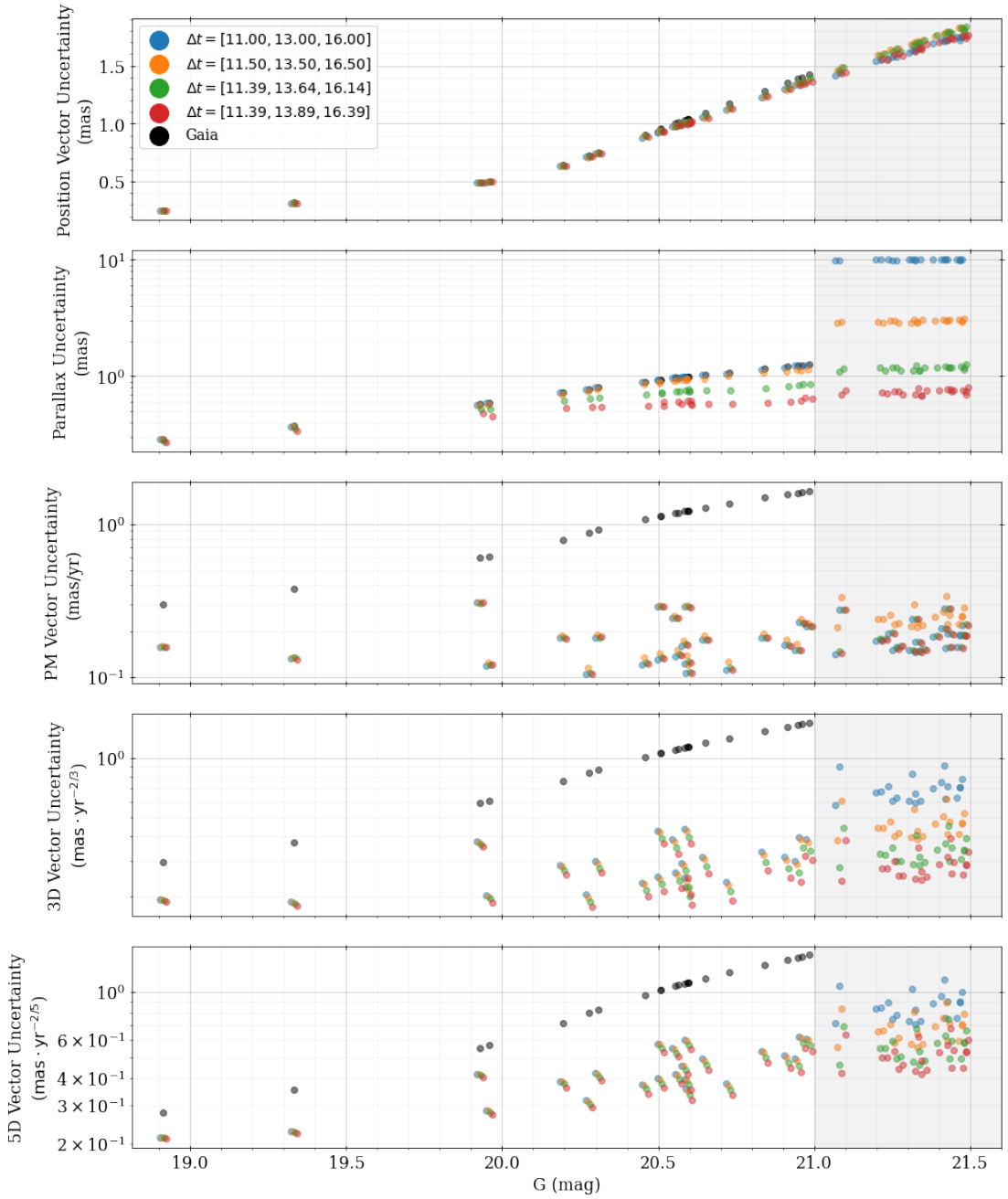

Figure 3.25: Comparison of the uncertainties on position, parallax, and proper motion for 48 synthetic *Roman*-observed stars in the COSMOS field. The points are colored by the time baseline offsets from *Gaia* for *Roman* observations at 2027, 2029, and 2032. Note that some of the y-axes are in log scale while others are linear. The two bottom-most panels correspond to the 3D vector of parallax and PM, and the 5D vector of position, parallax, and PM; the units of these axes are a result of our definition of vector uncertainty in Equation 3.7.

differences in the resulting parallax precision. Focusing on the third panel shows that all of the different observing strategies result in final PM uncertainties that are almost exactly the same, suggesting that achieving better parallaxes and positions has minimal cost to the PM precision. We emphasize, however, that this is only after analysing all 3 epochs together; when considering a single epoch at a time, sources without *Gaia* priors (i.e. $G > 21$ mag) do indeed have significantly larger PM uncertainties if there is an offset from the *Gaia* date as a result of the degeneracy between PM and parallax motions. While the *Roman* science goals are not defined by obtaining best possible PMs and parallaxes, our analysis highlights the difference that careful survey planning can have on the resulting astrometric precision.

Of the observing strategies we consider (including many that are not represented in Figure 3.25), the best choice for improving positions and parallaxes was to use an alternating-parallax-orbit-apocenters approach, which is the 4th option in the list above and the red points in the Figure. This strategy suggests that the parallax is most improved – and therefore the degeneracy between PM and parallax motions most disentangled – when we use the average parallax orbit of the sources in a given LOS (e.g. the blue orbit in Figure 3.3 for COSMOS) to identify the points in time that are most distant from each other. Then, we first observe at the time that is most offset from the *Gaia* observation time, and subsequent observations alternate by half a year such that they jump back and forth between the two apocenters. This approach also appears to have a large improvement on the position uncertainty, likely caused by the repeated sampling of the source's position at the same time in the parallax orbit. Originally, we

had expected that observing the farthest apocenter from *Gaia* with following observations a quarter of a year ahead and behind it (i.e. list option 3 and the green points in Figure 3.25) would better trace the parallax orbit, thereby recovering the most precise parallaxes, but this is not the case.

One way to visualize why the alternating apocenter approach gives the best improvement comes from thinking about the lines that connects the positions of the stars at successive measurements. These slopes/angles of these lines are generally defined by the PM of each star, but there are also offsets from the parallax motion. To measure the largest effect from parallax, we desire that those lines have the largest differences in their angles with respect to the previous measurement in time, which generally occurs a half year interval apart. The largest possible offset happens when we alternate between the two apocenters of the parallax orbit.

While we are not actually fitting for these lines/angles with `BP3M`, the information they represent is contained in the expected source position as a function of time. It is therefore a useful way to visualize the differences between the future position of the source produced by the possible combinations of the PM and parallax implied by the *Gaia* prior. Plotting these angles versus time for the different possible tracks of a source reveals that some points in time have a larger range of angle values than others. It is those times when the certainty about the source's position is least constrained, so having a measurement at that time does a lot to improve the understanding of the source's motion. As expected, the largest scatter in the angles occurs about half-a-year offset from whatever starting time we consider, and the largest possible separation

occurs when alternating between the parallax orbit apocenters. As time increases, the relative impact of the parallax motion on the total observed motion decreases, assuming the source has some non-zero PM, so the best parallax constraints are achieved when measuring the position multiple times over the course of the same year.

We recognize that our search of the best observing strategy has been non-exhaustive, so there may be an even better option that we have yet to consider. Future work will focus on identifying the best parallax-improvement approach when using 3 epochs by considering all combinations of time offsets (e.g. say 100 different offsets for each epoch, and then analyses of all possible combinations).

3.8 Summary

We have presented the statistics that describe, in general, how positions of sources from two or more sets of images (from any telescope or instrument) should be combined in a Bayesian hierarchical framework to measure transformation parameters between the images, as well as distributions on the positions, parallaxes, and proper motions for the individual sources. We use these statistics to created `BP3M` – a pipeline that builds off of `GaiaHub` – to combine archival *HST* and *Gaia* data to measure reliable positions, parallaxes, and PMs in the absolute reference frame of *Gaia*. We have tested this pipeline using a combination of synthetic *HST*-like images of the MW stellar halo, nearby dwarf galaxies, and catalogs of QSOs, showing in all cases that the `BP3M` PMs are both accurate and precise, especially for magnitudes where *Gaia* has large PM uncertainties ($19 < G < 21$ mag) or no PM measurements at all ($21 < G < 21.5$ mag).

Our key results include:

1. For time baselines of 10 to 13 years, the `BP3M` pipeline is able to measure PMs of stars in nearby dwarf spheroidals with uncertainties that are, on average, 8 to 13 times more precise than *Gaia* alone for $20.5 < G < 21$ mag (Section 3.4.2, Figures 3.10 and 3.11);

2. For the sparse fields of COSMOS (median 10 sources shared with *Gaia* per *HST* image), we measure a median PM improvement factor of 2.62 for sources with *Gaia*-measured PMs, and find a median PM uncertainty of 1.12 mas $\cdot$ yr^{-1} for the $\sim 25\%$ of sources without *Gaia*-measured PMs (Section 3.5, Figure 3.18);

3. Using the `BP3M`-measured transformation parameters for *HST* ACS/WFC images, we find that pixel scale ratio is best left as a free parameter when measuring transformations, but that the on- and off-axis skew terms may be explained as a function of *HST*'s angle and time (Section 3.6, Figures 3.20, 3.21, 3.23, Tables 3.5 and 3.6);

4. After generating synthetic *Roman*-like observations in a COSMOS-like field, we find an observing strategy that significantly increases parallax and position precision at no cost to PM precision (Section 3.7, Figure 3.25).

Broadly speaking, future work for this project can be broken into two parts: (1) improvements to the `BP3M` pipeline, and (2) analyses of `BP3M` results. For the first type of next steps, we plan to:

- Speed up computation time, especially when analysing multiple *HST* images to-

gether;

- Allow for cross-matching between multiple *HST* images for sources that are too faint for *Gaia* (i.e. $G > 21.5$ mag) to measure *HST*+*HST* PMs;

- Include the correlation terms between the PMs, parallaxes, and positions from *Gaia* (i.e. correlation between parallax and $\mu_{\alpha*}$) for increased posterior precision;

- Use the *Gaia*-measured distributions on PMs, parallaxes, and positions (where they exist) to assign probabilities to cross-match options in the *HST* images, so that `BP3M` can also improve on the cross matching, which might be particularly important for more dense regions.

For applications of the `BP3M` pipeline, our future endeavours will include:

- Expand our COSMOS analysis to include the entire ~ 2 square degree region, and combine different sets of *HST* images to extract the best possible `BP3M` PMs and parallaxes;

- Use the `BP3M` COSMOS PMs and parallaxes to constrain the MW disk and halo kinematics with unprecedented accuracy along a single LOS (e.g. similar to the analysis of Cunningham et al. 2019b);

- Analysis of large number of *HST* images that are well spaced in time and use different rotation angles (e.g. 47 Tucanae) to better constrain the pixel scale ratio and skew terms as a function of time and *HST* rotation for different instruments, detector, and filter combinations;

- Exhaustive search of time offsets between 3 epochs of *Roman*-like observations to determine the strategy that maximized the parallax precision.

3.9 Appendices

3.9.1 Motion Statistics

Here, we present the detailed accounting needed to derive the posterior conditional distributions in proper motion, parallax, and position for a given source. Using the posterior distribution of Equation 3.4, we can write out the posterior full conditional on the motion components (i.e. $\vec{\mu}_i, \mathrm{plx}_i, \Delta\vec{\theta}_i$) for source i:

$$\begin{aligned} p\left(\vec{\mu}_i, \mathrm{plx}_i, \Delta\vec{\theta}_i | \vec{a}, \vec{b}, \vec{c}, \vec{d}, \vec{W}_0, \vec{Z}_0, \dots\right) \propto & p\left(\mathrm{plx}_i' | \mathrm{plx}_i\right) \cdot p\left(\mathrm{plx}_i | \hat{\mathrm{plx}}, \sigma_{\hat{\mathrm{plx}}}\right) \cdot p\left(\vec{\mu}_i' | \vec{\mu}_i\right) \cdot \\ & p\left(\vec{\mu}_i | \hat{\mu}, \boldsymbol{V}_{\hat{\mu}}\right) \cdot p\left(\Delta\vec{\theta}_i\right) \cdot \\ & \prod_{j=1}^{n_{im}} \mathcal{N}\left(\Delta\vec{d}_{ij} | \boldsymbol{J}_{ij} \cdot \Delta\vec{m}_{ij}, \boldsymbol{V}_{d,ij} | \vec{a}, \vec{b}, \vec{c}, \vec{d}, \vec{W}_0, \vec{Z}_0\right) \end{aligned}$$

Because of its lack of coefficients, it is easiest to solve for the posterior full conditional on $\Delta\vec{\theta}_i$ first:

$$\begin{aligned} \left(\Delta\vec{\theta}_i | \vec{\mu}_i, \mathrm{plx}_i, \dots\right) \sim \mathcal{N}\Bigg(& \boldsymbol{\Sigma}_{\theta,i} = \left[\boldsymbol{V}_{\theta,i}^{-1} + \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij}\right]^{-1}, \\ & \vec{\mu}_{\theta,i} = \boldsymbol{\Sigma}_{\theta,i} \cdot \Bigg[\boldsymbol{V}_{\theta,i}^{-1} \cdot \Delta\vec{\theta}_i' + \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \\ & \left(\mathbf{plx_i} \cdot \Delta\vec{\mathrm{plx}}_{ij} + \boldsymbol{\Delta t_j} \cdot \vec{\mu}_i - \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij}\right)\Bigg]\Bigg) \end{aligned}$$

where we use bold-faced versions of scalars to represent the corresponding identity

matrix version of that number, such as

$$\mathbf{plx}_i = \mathrm{plx}_i \cdot \boldsymbol{I}_{2\times 2}.$$

We have also included a term for the *Gaia*-measured position offset vector $\vec{\Delta\theta}_i'$ to be as general as possible, but this vector is just $\vec{0}$ because *Gaia* has no measured offset from the positions it reports.

With these results in hand, and using

$$p(\vec{\mu}_i, \mathrm{plx}_i|\dots) = \frac{p(\vec{\mu}_i, \mathrm{plx}_i, \vec{\Delta\theta}_i|\dots)}{p(\vec{\Delta\theta}_i|\vec{\mu}_i, \mathrm{plx}_i, \dots)},$$

we can plug in $\vec{\Delta\theta}_i = \vec{\mu}_{\theta,i}$ in to $p\left(\vec{\mu}_i, \mathrm{plx}_i, \vec{\Delta\theta}_i|\dots\right)$ to get $p(\vec{\mu}_i, \mathrm{plx}_i|\dots)$. This works out to:

$$\begin{aligned}
p(\vec{\mu}_i, \mathrm{plx}_i|\dots) \propto & \exp\left(-\frac{1}{2\sigma^2_{\mathrm{plx},i}}\left(\mathrm{plx}_i - \mathrm{plx}_i'\right)^2\right) \cdot \exp\left(-\frac{1}{2\sigma^2_{\hat{\mathrm{plx}}}}\left(\mathrm{plx}_i - \hat{\mathrm{plx}}\right)^2\right) \cdot \\
& \exp\left(-\frac{1}{2}\left(\vec{\mu}_i - \vec{\mu}_i'\right)^T \cdot \boldsymbol{V}^{-1}_{\mu,i} \cdot \left(\vec{\mu}_i - \vec{\mu}_i'\right)\right) \cdot \\
& \exp\left(-\frac{1}{2}\left(\vec{\mu}_i - \hat{\mu}\right)^T \cdot \boldsymbol{V}^{-1}_{\hat{\mu}} \cdot \left(\vec{\mu}_i - \hat{\mu}\right)\right) \cdot \\
& \exp\left(-\frac{1}{2}\left(\vec{\mu}_{\theta,i} - \vec{\Delta\theta}_i{}'\right)^T \cdot \boldsymbol{V}^{-1}_{\theta,i} \cdot \left(\vec{\mu}_{\theta,i} - \vec{\Delta\theta}_i{}'\right)\right) \cdot \\
& \exp\left(-\frac{1}{2}\sum_{j=1}^{n_{im}}\left(\mathbf{plx_i} \cdot \Delta\vec{\mathrm{plx}}_{ij} + \boldsymbol{\Delta t_j} \cdot \vec{\mu}_i - \vec{\mu}_{\theta,i} - \boldsymbol{J}^{-1}_{ij} \cdot \Delta\vec{d}_{ij}\right)^T \cdot \boldsymbol{J}^T_{ij} \cdot \boldsymbol{V}^{-1}_{d,ij} \cdot \boldsymbol{J}_{ij} \cdot \right. \\
& \qquad \left.\left(\mathbf{plx_i} \cdot \Delta\vec{\mathrm{plx}}_{ij} + \boldsymbol{\Delta t_j} \cdot \vec{\mu}_i - \vec{\mu}_{\theta,i} - \boldsymbol{J}^{-1}_{ij} \cdot \Delta\vec{d}_{ij}\right)\right)
\end{aligned}$$

The next easiest term to solve for is $\vec{\mu}_i$. If we rewrite $\vec{\mu}_{\theta,i}$ to be in terms of $\vec{\mu}_i$,

then we get:

$$\boldsymbol{A}_{\mu,i} = \left[\boldsymbol{\Sigma}_{\theta,i} \cdot \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^{T} \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Delta t_j\right]$$

$$\vec{B}_{\mu,i} = \boldsymbol{\Sigma}_{\theta,i} \cdot \left[\sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^{T} \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij} - \boldsymbol{V}_{\theta,i}^{-1} \cdot \Delta\vec{\theta}_i' - \mathrm{plx}_i \cdot \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^{T} \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Delta\vec{\mathrm{plx}}_{ij}\right]$$

$$\vec{\mu}_{\theta,i} = \boldsymbol{A}_{\mu,i} \cdot \vec{\mu}_i - \vec{B}_{\mu,i}.$$

Then we have:

$$\mathcal{N}\left(\vec{\mu}_{\theta,i} - \Delta\vec{\theta}_i{}', \boldsymbol{V}_{\theta,i}\right)$$
$$\Longrightarrow \mathcal{N}\left(\vec{\mu}_i | \boldsymbol{\Sigma}_{\mu,\theta,i} = \left[\boldsymbol{A}_{\mu,i}^{T} \cdot \boldsymbol{V}_{\theta,i}^{-1} \cdot \boldsymbol{A}_{\mu,i}\right]^{-1}, \vec{\mu}_{\mu,\theta,i} = \boldsymbol{A}_{\mu,i}^{-1} \cdot \left[\Delta\vec{\theta}_i{}' + \vec{B}_{\mu,i}\right]\right)$$

Similarly, we can define:

$$\boldsymbol{C}_{\mu,ij} = \boldsymbol{\Delta t_j} - \boldsymbol{A}_{\mu,i}$$

$$\vec{D_{\mu,ij}} = \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij} - \vec{B}_{\mu,i} - \mathbf{plx_i} \cdot \Delta\vec{\mathrm{plx}}_{ij}$$

$$\Delta\vec{m}_{ij} - \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij} = \boldsymbol{C}_{\mu,ij} \cdot \vec{\mu}_i - \vec{D}_{\mu,ij}$$

so that we have:

$$\mathcal{N}\left(\Delta\vec{m}_{ij} - \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij}, \left[\boldsymbol{J}_{ij}^{T} \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij}\right]^{-1}\right)$$
$$\Longrightarrow \mathcal{N}\left(\vec{\mu}_i | \boldsymbol{\Sigma}_{\mu,d,ij} = \left[\boldsymbol{C}_{\mu,ij}^{T} \cdot \boldsymbol{J}_{ij}^{T} \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \boldsymbol{C}_{\mu,ij}\right]^{-1}, \vec{\mu}_{\mu,d,ij} = \boldsymbol{C}_{\mu,ij}^{-1} \cdot \vec{D}_{\mu,ij}\right)$$

and we find that the posterior full conditional on $\vec{\mu}_i$ is given by:

$$
\begin{aligned}
p(\vec{\mu}_i|\mathrm{plx}_i \dots) \propto & \mathcal{N}\left(\vec{\mu}_i' - \vec{\mu}_i, \boldsymbol{V}_{\mu,i}\right) \cdot \mathcal{N}\left(\vec{\mu}_i - \hat{\mu}, \boldsymbol{V}_{\hat{\boldsymbol{\mu}}}\right) \cdot \\
& \mathcal{N}\left(\vec{\mu}_i | \boldsymbol{\Sigma}_{\mu,\theta,i} = \left[\boldsymbol{A}_{\mu,i}^T \cdot \boldsymbol{V}_{\theta,i}^{-1} \cdot \boldsymbol{A}_{\mu,i}\right]^{-1}, \vec{\mu}_{\mu,\theta,i} = \boldsymbol{A}_{\mu,i}^{-1} \cdot \left[\Delta\vec{\theta}_i{}' + \vec{B}_{\mu,i}\right]\right) \cdot \\
& \prod_{j=1}^{n_{im}} \mathcal{N}\left(\vec{\mu}_i | \boldsymbol{\Sigma}_{\mu,d,ij} = \left[\boldsymbol{C}_{\mu,ij}^T \cdot \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \boldsymbol{C}_{\mu,ij}\right]^{-1}, \vec{\mu}_{\mu,d,ij} = \boldsymbol{C}_{\mu,ij}^{-1} \cdot \vec{D}_{\mu,ij}\right) \\
= & \mathcal{N}\left(\boldsymbol{\Sigma}_{\mu,i} = \left[\boldsymbol{V}_{\mu,i}^{-1} + \hat{\boldsymbol{V}}_{\boldsymbol{\mu}}{}^{-1} + \boldsymbol{\Sigma}_{\mu,\theta,i}^{-1} + \sum_{j=1}^{n_{im}} \boldsymbol{\Sigma}_{\mu,d,ij}^{-1}\right]^{-1}, \right. \\
& \vec{\mu}_{\mu,i} = \boldsymbol{\Sigma}_{\mu,i} \cdot \left[\boldsymbol{V}_{\mu,i}^{-1} \cdot \vec{\mu}_i' + \boldsymbol{V}_{\hat{\boldsymbol{\mu}}}{}^{-1} \cdot \hat{\mu} + \boldsymbol{\Sigma}_{\mu,\theta,i}^{-1} \cdot \vec{\mu}_{\mu,\theta,i} \right. \\
& \left. \left. + \sum_{j=1}^{n_{im}} \boldsymbol{\Sigma}_{\mu,d,ij}^{-1} \cdot \boldsymbol{C}_{\mu,ij}^{-1} \cdot \vec{D}_{\mu,ij}\right]\right)
\end{aligned}
$$

Using the same arguments as before, we can plug $\vec{\mu}_i = \vec{\mu}_{\mu,i}$ into $p\left(\vec{\mu}_i, \mathrm{plx}_i | \dots\right)$ to get $p\left(\mathrm{plx}_i | \dots\right)$:

$$
\begin{aligned}
p(\mathrm{plx}_i | \dots) \propto & \exp\left(-\frac{1}{2\sigma_{\mathrm{plx},i}^2}\left(\mathrm{plx}_i - \mathrm{plx}_i'\right)^2 - \frac{1}{2\sigma_{\hat{\mathrm{plx}}}^2}\left(\mathrm{plx}_i - \hat{\mathrm{plx}}\right)^2\right) \cdot \\
& \exp\left(-\frac{1}{2}\left(\vec{\mu}_{\mu,i} - \vec{\mu}_i'\right)^T \cdot \boldsymbol{V}_{\mu,i}^{-1} \cdot \left(\vec{\mu}_{\mu,i} - \vec{\mu}_i'\right) - \frac{1}{2}\left(\vec{\mu}_{\mu,i} - \hat{\mu}\right)^T \cdot \boldsymbol{V}_{\hat{\mu}}^{-1} \cdot \left(\vec{\mu}_{\mu,i} - \hat{\mu}\right)\right) \cdot \\
& \exp\left(-\frac{1}{2}\left(\vec{\mu}_{\theta,i} - \Delta\vec{\theta}_i{}'\right)^T \cdot \boldsymbol{V}_{\theta,i}^{-1} \cdot \left(\vec{\mu}_{\theta,i} - \Delta\vec{\theta}_i{}'\right)\right) \cdot \\
& \prod_{j=1}^{n_{im}} \exp\left(-\frac{1}{2}\left(\mathbf{plx_i} \cdot \Delta\vec{\mathrm{plx}}_{ij} + \boldsymbol{\Delta t_j} \cdot \mu_{\mu,i} - \vec{\mu}_{\theta,i} - \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij}\right)^T \cdot \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \right. \\
& \left. \left(\mathbf{plx_i} \cdot \Delta\vec{\mathrm{plx}}_{ij} + \boldsymbol{\Delta t_j} \cdot \mu_{\mu,i} - \vec{\mu}_{\theta,i} - \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij}\right)\right).
\end{aligned}
$$

We can then rewrite the definition of $\vec{\mu}_{\mu,i}$ to be more helpful:

$$\begin{aligned}\vec{\mu}_{\mu,i} = \boldsymbol{\Sigma}_{\mu,i} \cdot \Big[\boldsymbol{V}_{\mu,i}^{-1} \cdot \vec{\mu}_i' + \boldsymbol{A}_{\mu,i}^T \cdot \boldsymbol{V}_{\theta,i}^{-1} \cdot \Big[\vec{\Delta\theta}_i' - \boldsymbol{\Sigma}_{\theta,i} \cdot \Big\{ \boldsymbol{V}_{\theta,i}^{-1} \cdot \vec{\Delta\theta}_i' \\ + \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Big(\mathbf{plx_i} \cdot \Delta\vec{\text{plx}}_i - \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} \Big) \Big\} \Big] \\ + \sum_{j=1}^{n_{im}} \boldsymbol{C}_{\mu,ij}^T \cdot \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Big[\boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} + \boldsymbol{\Sigma}_{\theta,i} \cdot \boldsymbol{V}_{\theta,i}^{-1} \cdot \vec{\Delta\theta}_i' - \mathbf{plx_i} \cdot \Delta\vec{\text{plx}}_{ij} \\ + \boldsymbol{\Sigma}_{\theta,i} \cdot \sum_{j=1}^{n_{im}} \Big(\boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Big\{ \mathbf{plx_i} \cdot \Delta\vec{\text{plx}}_{ij} - \boldsymbol{J}_{ij}^{-1} \vec{\Delta d}_{ij} \Big\} \Big) \Big] \Big] \Big] .\end{aligned}$$

Next, we rewrite $\vec{B}_{\mu,i}$ with plx_i in mind to get:

$$\begin{aligned}\vec{B}_{\mu,i} &= \boldsymbol{\Sigma}_{\theta,i} \cdot \Bigg[\sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} - \boldsymbol{V}_{\theta,i}^{-1} \cdot \vec{\Delta\theta}_i' \\ &\qquad - \text{plx}_i \cdot \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Delta\vec{\text{plx}}_{ij} \Bigg] \\ &= \boldsymbol{\Sigma}_{\theta,i} \cdot \Bigg[\sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} - \boldsymbol{V}_{\theta,i}^{-1} \cdot \vec{\Delta\theta}_i' \Bigg] \\ &\qquad - \text{plx}_i \cdot \boldsymbol{\Sigma}_{\theta,i} \cdot \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Delta\vec{\text{plx}}_{ij} \\ &= \text{plx}_i \cdot \vec{A}_{\text{plx},\mu,i} - \vec{B}_{\text{plx},\mu,i}\end{aligned}$$

where

$$\vec{A}_{\text{plx},\mu,i} = -\boldsymbol{\Sigma}_{\theta,i} \cdot \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Delta\vec{\text{plx}}_{ij}$$

$$\vec{B}_{\text{plx},\mu,i} = \boldsymbol{\Sigma}_{\theta,i} \cdot \Bigg[\boldsymbol{V}_{\theta,i}^{-1} \cdot \vec{\Delta\theta}_i' - \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} \Bigg] ,$$

which gets us to:

$$\begin{aligned}
\vec{\mu}_{\mu,i} = \boldsymbol{\Sigma}_{\mu,i} \cdot \Big[& \boldsymbol{V}_{\mu,i}^{-1} \cdot \vec{\mu}_i' + \boldsymbol{V}_{\hat{\boldsymbol{\mu}}}{}^{-1} \cdot \hat{\mu} + \boldsymbol{\Sigma}_{\mu,\theta,i}^{-1} \cdot \boldsymbol{A}_{\mu,i}^{-1} \cdot \left(\Delta\vec{\theta}_i{}' + \mathrm{plx}_i \cdot A_{\vec{\mathrm{plx}},\mu,i} - B_{\vec{\mathrm{plx}},\mu,i} \right) \\
& - \sum_{j=1}^{n_{im}} \boldsymbol{\Sigma}_{\mu,d,ij}^{-1} \cdot \boldsymbol{C}_{\mu,ij}^{-1} \cdot \left(\mathrm{plx}_i \cdot A_{\vec{\mathrm{plx}},\mu,i} - B_{\vec{\mathrm{plx}},\mu,i} + \mathrm{plx}_i \cdot \Delta\vec{\mathrm{plx}}_{ij} - \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij} \right) \Big] \\
= \mathrm{plx}_i \cdot \boldsymbol{\Sigma}_{\mu,i} \cdot & \left[\boldsymbol{\Sigma}_{\mu,\theta,i}^{-1} \cdot \boldsymbol{A}_{\mu,i}^{-1} \cdot A_{\vec{\mathrm{plx}},\mu,i} - \sum_{j=1}^{n_{im}} \boldsymbol{\Sigma}_{\mu,d,ij}^{-1} \cdot \boldsymbol{C}_{\mu,ij}^{-1} \cdot \left(A_{\vec{\mathrm{plx}},\mu,i} + \Delta\vec{\mathrm{plx}}_{ij} \right) \right] \\
+ \boldsymbol{\Sigma}_{\mu,i} \cdot \Big[& \boldsymbol{V}_{\mu,i}^{-1} \cdot \vec{\mu}_i' + \boldsymbol{V}_{\hat{\boldsymbol{\mu}}}{}^{-1} \cdot \hat{\mu} + \boldsymbol{\Sigma}_{\mu,\theta,i}^{-1} \cdot \boldsymbol{A}_{\mu,i}^{-1} \cdot \left(\Delta\vec{\theta}_i{}' - B_{\vec{\mathrm{plx}},\mu,i} \right) \\
& + \sum_{j=1}^{n_{im}} \boldsymbol{\Sigma}_{\mu,d,ij}^{-1} \cdot \boldsymbol{C}_{\mu,ij}^{-1} \cdot \left(B_{\vec{\mathrm{plx}},\mu,i} + \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij} \right) \Big] \\
= \mathrm{plx}_i \cdot \vec{C}_{\mathrm{plx},\mu,i} - \vec{D}_{\mathrm{plx},\mu,i} &
\end{aligned}$$

where

$$\vec{C}_{\mathrm{plx},\mu,i} = \boldsymbol{\Sigma}_{\mu,i} \cdot \left[\boldsymbol{\Sigma}_{\mu,\theta,i}^{-1} \cdot \boldsymbol{A}_{\mu,i}^{-1} \cdot A_{\vec{\mathrm{plx}},\mu,i} - \sum_{j=1}^{n_{im}} \boldsymbol{\Sigma}_{\mu,d,ij}^{-1} \cdot \boldsymbol{C}_{\mu,ij}^{-1} \cdot \left(A_{\vec{\mathrm{plx}},\mu,i} + \Delta\vec{\mathrm{plx}}_{ij} \right) \right]$$

$$\begin{aligned}
\vec{D}_{\mathrm{plx},\mu,i} = - \boldsymbol{\Sigma}_{\mu,i} \cdot \Big[& \boldsymbol{V}_{\mu,i}^{-1} \cdot \vec{\mu}_i' + \boldsymbol{V}_{\hat{\boldsymbol{\mu}}}{}^{-1} \cdot \hat{\mu} + \boldsymbol{\Sigma}_{\mu,\theta,i}^{-1} \cdot \boldsymbol{A}_{\mu,i}^{-1} \cdot \left(\Delta\vec{\theta}_i{}' - B_{\vec{\mathrm{plx}},\mu,i} \right) \\
& + \sum_{j=1}^{n_{im}} \boldsymbol{\Sigma}_{\mu,d,ij}^{-1} \cdot \boldsymbol{C}_{\mu,ij}^{-1} \cdot \left(B_{\vec{\mathrm{plx}},\mu,i} + \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij} \right) \Big] .
\end{aligned}$$

Similarly, we can rewrite $\vec{\mu}_{\theta,i}$ as:

$$\begin{aligned}
\vec{\mu}_{\theta,i} = \mathbf{\Sigma}_{\theta,i} \cdot \Big[& \boldsymbol{V}_{\theta,i}^{-1} \cdot \vec{\Delta\theta}_i' \\
& + \sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Big(\mathrm{plx}_i \cdot \Delta\vec{\mathrm{plx}}_{ij} - \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} \\
& \qquad + \boldsymbol{\Delta t_j} \cdot \Big(\mathrm{plx}_i \cdot \vec{C}_{\mathrm{plx},\mu,i} - \vec{D}_{\mathrm{plx},\mu,i} \Big) \Big) \Big] \\
= \mathrm{plx}_i \cdot \mathbf{\Sigma}_{\theta,i} \cdot & \left[\sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Big(\Delta\vec{\mathrm{plx}}_{ij} + \Delta t_j \cdot \vec{C}_{\mathrm{plx},\mu,i} \Big) \right] \\
& - \mathbf{\Sigma}_{\theta,i} \cdot \left[\sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Big(\Delta t_j \cdot \vec{D}_{\mathrm{plx},\mu,i} + \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} \Big) - \boldsymbol{V}_{\theta,i}^{-1} \cdot \vec{\Delta\theta}_i' \right] \\
= \mathrm{plx}_i \cdot \vec{E}_{\mathrm{plx},\theta,i} & - \vec{F}_{\mathrm{plx},\theta,i}
\end{aligned}$$

where

$$\vec{E}_{\mathrm{plx},\theta,i} = \mathbf{\Sigma}_{\theta,i} \cdot \left[\sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Big(\Delta\vec{\mathrm{plx}}_{ij} + \Delta t_j \cdot \vec{C}_{\mathrm{plx},\mu,i} \Big) \right]$$

$$\vec{F}_{\mathrm{plx},\theta,i} = \mathbf{\Sigma}_{\theta,i} \cdot \left[\sum_{j=1}^{n_{im}} \boldsymbol{J}_{ij}^T \cdot \boldsymbol{V}_{d,ij}^{-1} \cdot \boldsymbol{J}_{ij} \cdot \Big(\Delta t_j \cdot \vec{D}_{\mathrm{plx},\mu,i} + \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} \Big) - \boldsymbol{V}_{\theta,i}^{-1} \cdot \vec{\Delta\theta}_i' \right].$$

We also rewrite $\vec{\Delta m}_{ij} - \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij}$ as:

$$\begin{aligned}
\vec{\Delta m}_{ij} - \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} &= \mathrm{plx_i} \cdot \Delta\vec{\mathrm{plx}}_{ij} + \Delta t_j \cdot \vec{\mu}_i - \vec{\mu}_{\theta,i} - \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} \\
&= \mathrm{plx_i} \cdot \Delta\vec{\mathrm{plx}}_{ij} + \Delta t_j \cdot \Big(\mathrm{plx}_i \cdot \vec{C}_{\mathrm{plx},\mu,i} - \vec{D}_{\mathrm{plx},\mu,i} \Big) \\
&\qquad - \Big(\mathrm{plx}_i \cdot \vec{E}_{\mathrm{plx},\theta,i} - \vec{F}_{\mathrm{plx},\theta,i} \Big) - \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} \\
&= \mathrm{plx_i} \cdot \Big[\Delta\vec{\mathrm{plx}}_{ij} + \Delta t_j \cdot \vec{C}_{\mathrm{plx},\mu,i} - \vec{E}_{\mathrm{plx},\theta,i} \Big] \\
&\qquad - \Big[\Delta t_j \cdot \vec{D}_{\mathrm{plx},\mu,i} - \vec{F}_{\mathrm{plx},\theta,i} + \boldsymbol{J}_{ij}^{-1} \cdot \vec{\Delta d}_{ij} \Big] \\
&= \mathrm{plx_i} \cdot G_{\mathrm{plx},d,ij} - \vec{H}_{\mathrm{plx},d,ij}
\end{aligned}$$

where:

$$\vec{G}_{\text{plx},d,ij} = \left[\Delta\vec{\text{plx}}_{ij} + \Delta t_j \cdot \vec{C_{\text{plx},\mu,i}} - \vec{E_{\text{plx},\theta,i}}\right]$$

$$\vec{H}_{\text{plx},d,ij} = \left[\Delta t_j \cdot \vec{D_{\text{plx},\mu,i}} - \vec{F_{\text{plx},\theta,i}} + \boldsymbol{J}_{ij}^{-1} \cdot \Delta\vec{d}_{ij}\right].$$

Finally, we can make use of the following relationship to deal with the multivariate Gaussians that contain the plx_i factors:

$$\mathcal{N}\left(a \cdot \vec{X} - \vec{Y}, \boldsymbol{V}\right) \implies \mathcal{N}\left(a|\sigma^2 = \left[\vec{X}^T \cdot \boldsymbol{V}^{-1} \cdot \vec{X}\right], \mu = \sigma^2 \cdot \left[\vec{X}^T \cdot \boldsymbol{V}^{-1} \cdot \vec{Y}\right]\right)$$

to arrive at:

$$
\begin{aligned}
p(\mathrm{plx}_i|\dots) \propto & \mathcal{N}\left(\mathrm{plx}_i|\hat{\sigma}^2_{\mathrm{plx}}, \hat{\mathrm{plx}}\right) \cdot \mathcal{N}\left(\mathrm{plx}_i|\sigma^2_{\mathrm{plx},i}, \mathrm{plx}'_i\right) \cdot \\
& \mathcal{N}\left(\mathrm{plx}_i|\sigma^2_{\mathrm{plx},\mu,i} = \left[\vec{C}^T_{\mathrm{plx},\mu,i} \cdot \boldsymbol{V}^{-1}_{\mu,i} \cdot \vec{C}_{\mathrm{plx},\mu,i}\right]^{-1},\right. \\
& \qquad \left. \mu_{\mathrm{plx},\mu,i} = \sigma^2_{\mathrm{plx},\mu,i} \cdot \left[\vec{C}^T_{\mathrm{plx},\mu,i} \cdot \boldsymbol{V}^{-1}_{\mu,i} \cdot \left(\vec{D}_{\mathrm{plx},\mu,i} + \vec{\mu}'_i\right)\right]\right) \cdot \\
& \mathcal{N}\left(\mathrm{plx}_i|\sigma^2_{\mathrm{plx},\hat{\mu},i} = \left[\vec{C}^T_{\mathrm{plx},\mu,i} \cdot \boldsymbol{V}_{\hat{\boldsymbol{\mu}}}{}^{-1} \cdot \vec{C}_{\mathrm{plx},\mu,i}\right]^{-1},\right. \\
& \qquad \left. \mu_{\mathrm{plx},\hat{\mu},i} = \sigma^2_{\mathrm{plx},\hat{\mu},i} \cdot \left[\vec{C}^T_{\mathrm{plx},\mu,i} \cdot \boldsymbol{V}_{\hat{\boldsymbol{\mu}}}{}^{-1} \cdot \left(\vec{D}_{\mathrm{plx},\mu,i} + \hat{\mu}\right)\right]\right) \cdot \\
& \mathcal{N}\left(\mathrm{plx}_i|\sigma^2_{\mathrm{plx},\theta,i} = \left[\vec{E}^T_{\mathrm{plx},\theta,i} \cdot \boldsymbol{V}^{-1}_{\theta,i} \cdot \vec{E}_{\mathrm{plx},\theta,i}\right]^{-1},\right. \\
& \qquad \left. \mu_{\mathrm{plx},\theta,i} = \sigma^2_{\mathrm{plx},\theta,i} \cdot \left[\vec{E}^T_{\mathrm{plx},\theta,i} \cdot \boldsymbol{V}^{-1}_{\theta,i} \cdot \left(\vec{F}_{\mathrm{plx},\theta,i} + \Delta\vec{\theta}_i{}'\right)\right]\right) \cdot \\
& \prod_{j=1}^{n_{im}} \mathcal{N}\left(\mathrm{plx}_i|\sigma^2_{\mathrm{plx},d,ij} = \left[\vec{G}^T_{\mathrm{plx},d,ij} \cdot \boldsymbol{J}^T_{ij} \cdot \boldsymbol{V}^{-1}_{d,ij} \cdot \boldsymbol{J}_{ij} \cdot \vec{G}_{\mathrm{plx},d,ij}\right]^{-1},\right. \\
& \qquad \left. \mu_{\mathrm{plx},d,ij} = \sigma^2_{\mathrm{plx},d,ij} \cdot \left[\vec{G}^T_{\mathrm{plx},d,ij} \cdot \boldsymbol{J}^T_{ij} \cdot \boldsymbol{V}^{-1}_{d,ij} \cdot \boldsymbol{J}_{ij} \cdot \vec{H}_{\mathrm{plx},d,ij}\right]\right) \\
= & \mathcal{N}\left(\mathrm{plx}_i|\sigma_{\hat{\mathrm{plx}},i}{}^2 = \left[\hat{\sigma}^{-2}_{\mathrm{plx}} + \sigma^{-2}_{\mathrm{plx},i} + \sigma^{-2}_{\mathrm{plx},\theta,i} + \sigma^{-2}_{\mathrm{plx},\mu,i} + \sigma^{-2}_{\mathrm{plx},\hat{\mu},i} + \sum_{j=1}^{n_{im}} \sigma^{-2}_{\mathrm{plx},d,ij}\right]^{-1},\right. \\
& \qquad \mu_{\hat{\mathrm{plx}},i} = \sigma_{\hat{\mathrm{plx}},i}{}^2 \cdot \left[\hat{\sigma}^{-2}_{\mathrm{plx}} \cdot \hat{\mathrm{plx}} + \sigma^{-2}_{\mathrm{plx},i} \cdot \mathrm{plx}'_i + \sigma^{-2}_{\mathrm{plx},\theta,i} \cdot \mu_{\mathrm{plx},\theta,i}\right. \\
& \qquad\qquad + \sigma^{-2}_{\mathrm{plx},\mu,i} \cdot \mu_{\mathrm{plx},\mu,i} + \sigma^{-2}_{\mathrm{plx},\hat{\mu},i} \cdot \mu_{\mathrm{plx},\hat{\mu},i} \\
& \qquad\qquad \left.\left. + \sum_{j=1}^{n_{im}} \sigma^{-2}_{\mathrm{plx},d,ij} \cdot \mu_{\mathrm{plx},d,ij}\right]\right)
\end{aligned}
$$

thereby completing the set of posterior full conditionals, all of which are Gaussians, that we need to quickly sample PMs, parallaxes, and position offsets for source i given a set of transformation parameters.

3.9.2 Generating Synthetic, COSMOS-like Data

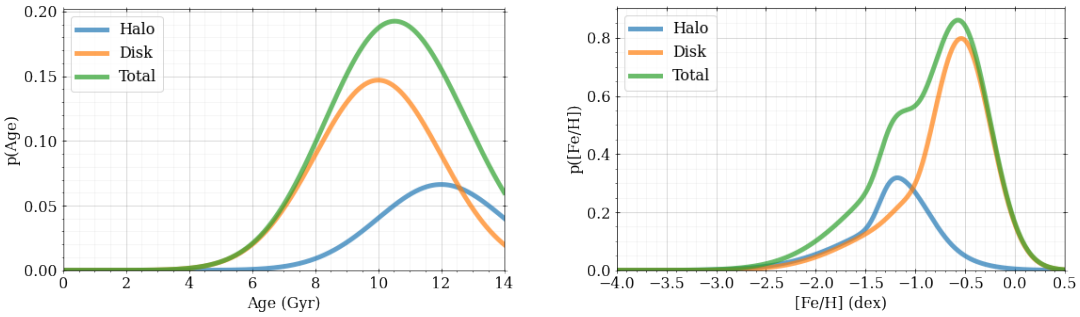

Figure 3.26: Distributions on age (left panel) and [Fe/H] (right panel) for the MW halo (blue) and thick disk (orange) populations used to create synthetic stars. The halo population makes up 28% of the total population (green), while the disk is the other 72%.

Following a similar approach to McKinnon et al. (2023), we define mass, age, [Fe/H], and distance distributions for both a MW thick disk and a MW halo population. We also define distributions for the Galactocentric 3D velocities for the halo and thick disk. The distributions for each parameter that go into generating a synthetic star are summarized in Table 3.7. The particular choices for the velocity distributions come from the MW halo velocity ellipsoid measurements of Cunningham et al. (2019b), as do the distance modulus distributions. For the age distributions, we use an analytical approximation of the age distributions presented in Bonaca et al. (2020). The [Fe/H] distributions are again analytical approximations of the results presented in Conroy et al. (2019a) and Naidu et al. (2020). The age and [Fe/H] distributions are displayed in Figure 3.26, where the distributions have been scaled by their contribution to the total halo population. After consulting a Besançon model (Robin et al. 2003) along a COSMOS-like LOS in the $16 < G < 21.5$ mag range, we choose to set the fraction of

Table 3.7: Distributions used to generate synthetic data of halo and thick disk MW stars in the COSMOS field. $\mathcal{SKN}$ indicates a skew-normal distribution.

Component	Distribution	Functional Form
Halo	$p([\mathrm{Fe/H}])$	$0.406 \cdot \mathcal{SKN}(\mu = -1.35 \text{ dex}, \sigma = 0.5 \text{ dex}, a = 5)$ $+0.594 \cdot \mathcal{SKN}(\mu = -0.9 \text{ dex}, \sigma = 0.7 \text{ dex}, a = -3)$
	$p(\mathrm{age})$	$\mathcal{N}(\mu = 12 \text{ Gyr}, \sigma = 2 \text{ Gyr})$
	$p(\mathrm{mass}\|[\mathrm{Fe/H}], \mathrm{age})$	Kroupa (2001) IMF, $\propto k \left(\frac{\mathrm{mass}}{M_\odot}\right)^{-\alpha}$ with $\begin{cases} k = 25,\ \alpha = 0.3, & \mathrm{mass} < 0.08 M_\odot \\ k = 2,\ \alpha = 1.3, & \mathrm{mass} < 0.5 M_\odot \\ k = 1,\ \alpha = 2.3, & \mathrm{mass} > 0.5 M_\odot \end{cases}$
	$p(\text{distance moludus})$	$\propto D^3 \left(\frac{R_q}{27\mathrm{kpc}}\right)^{-\alpha}$ with $\begin{cases} \alpha = 2.3, & R_q < 27\,\mathrm{kpc} \\ \alpha = 4.6, & R_q \geq 27\,\mathrm{kpc} \end{cases}$
	$p(v_r)$	$\mathcal{N}(\mu = 0 \text{ km} \cdot \text{s}^{-1}, \sigma = 130 \text{ km} \cdot \text{s}^{-1})$
	$p(v_\phi)$	$\mathcal{N}(\mu = 0 \text{ km} \cdot \text{s}^{-1}, \sigma = 70 \text{ km} \cdot \text{s}^{-1})$
	$p(v_\theta)$	$\mathcal{N}(\mu = 0 \text{ km} \cdot \text{s}^{-1}, \sigma = 70 \text{ km} \cdot \text{s}^{-1})$
Disk	$p([\mathrm{Fe/H}])$	$\frac{1}{6}\mathcal{SKN}(\mu = -1.05 \text{ dex}, \sigma = 0.6 \text{ dex}, a = -5)$ $+\frac{5}{6}\mathcal{N}(\mu = -0.54 \text{ dex}, \sigma = 0.3 \text{ dex})$
	$p(\mathrm{Age})$	$\mathcal{N}(\mu = 10 \text{ Gyr}, \sigma = 2 \text{ Gyr})$
	$p(\mathrm{mass}\|[\mathrm{Fe/H}], \mathrm{age})$	Kroupa (2001) IMF
	$p(\text{distance moludus})$	$\propto D^3 \exp\left(-\frac{R_D}{3 \text{ kpc}} - \frac{z}{1 \text{ kpc}}\right)$, where $R_D^2 = x^2 + y^2$
	$p(v_{R_D})$	$\mathcal{N}(\mu = 0 \text{ km} \cdot \text{s}^{-1}, \sigma = 45 \text{ km} \cdot \text{s}^{-1})$
	$p(v_z)$	$\mathcal{N}(\mu = 0 \text{ km} \cdot \text{s}^{-1}, \sigma = 20 \text{ km} \cdot \text{s}^{-1})$
	$p(v_T)$	$\mathcal{SKN}(\mu = 242 \text{ km} \cdot \text{s}^{-1}, \sigma = 46.2 \text{ km} \cdot \text{s}^{-1}, a = -2)$

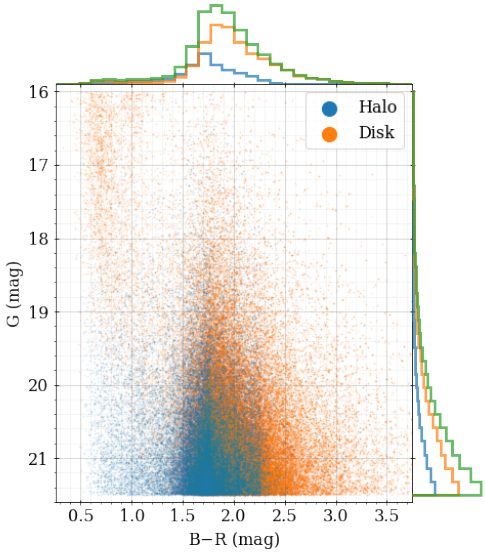

Figure 3.27: Color-magnitude diagram of synthetic COSMOS-like stars, colored by whether the stars belong to the MW halo (blue) or thick disk (orange). The histograms along the right and top edges include the total population (green). The halo population makes up 28% of the total population, while the disk is the other 72%.

the number of observed stars that belong to the halo (i.e. "halo fraction") to 28%.

Using these distributions and the halo fraction, we draw stellar parameters (i.e. mass, age, and [Fe/H]) for each synthetic star, and then interpolate to that point using the MIST isochrones[12] (Dotter 2016; Choi et al. 2016; Paxton et al. 2011, 2013, 2015). This interpolated point yields the color and absolute magnitude of each synthetic star. After we draw distances, we can measure the apparent magnitudes for each star, and then perform rejection sampling to get a sample of stars that fall within a particular range in magnitude and/or color. A color-magnitude diagram is shown in Figure 3.27 for the COSMOS-like stars we generate, with histograms along the top and

[12] https://waps.cfa.harvard.edu/MIST/index.html

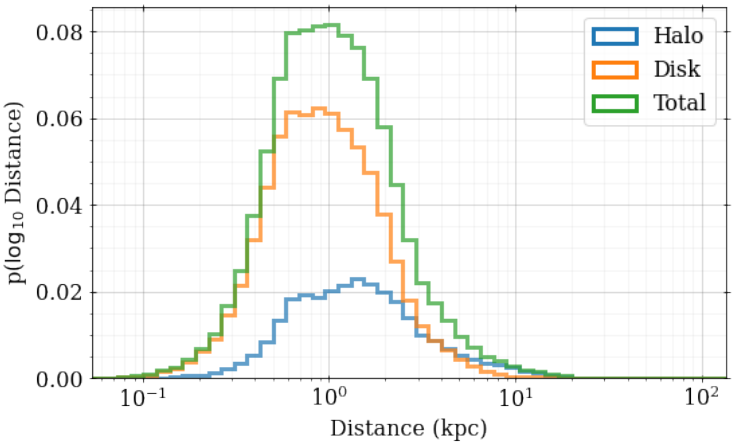

Figure 3.28: Distance distributions of synthetic COSMOS-like stars, colored by whether the stars belong to the MW halo (blue) or thick disk (orange).

right edges showing the distribution of the halo (blue), thick disk (orange), and total (green) populations. A histogram of the LOS distance to each synthetic star is given in Figure 3.28.

Next, for each synthetic star, we draw the 3D velocity components based on whether that star belongs to the halo or the thick disk. Using the stellar position on the sky and distance allow us to transform those Galactocentric velocities into observable velocities[13]. This yields the proper motions shown in Figure 3.29.

To create outputs that the `BP3M` pipeline expects, the next step is to create synthetic *HST* images and corresponding *Gaia* measurements. For synthetic *Gaia* uncertainties on position, parallax, and proper motion, we look at real *Gaia*-measured stars within 1 degree of the COSMOS field center; the resulting uncertainties in each dimension are shown in Figure 3.30 as a function of magnitude (blue points) with a

[13]To transform between the observer frame and a Galactocentric one, we use $r_\odot = 8.1$ kpc, assume a circular speed of 235 km s^{-1}, and solar peculiar motion $(U, V, W) = (11.1, 12.24, 7.25)$ km s^{-1} (Schönrich et al. 2011).

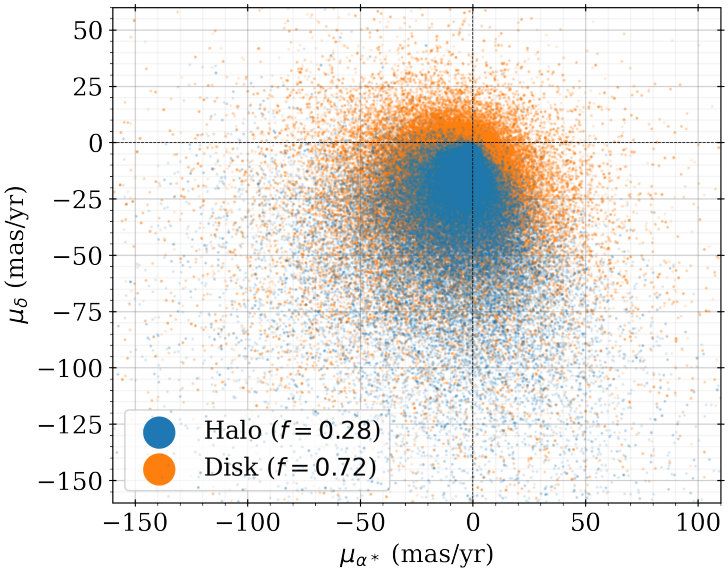

Figure 3.29: Proper motions of synthetic COSMOS-like stars, colored by whether the stars belong to the MW halo (blue) or thick disk (orange).

median-binned line overlaid (orange). In cases where the data do not extend as faint as needed, we linearly extrapolate from the median binning.

We use a similar approach when it comes to modeling the uncertainties in the *HST* image positions. Specifically, we use the $\sim$ 2000 real COSMOS stars from Section 3.5 to look at the `GaiaHub`-measured *HST* position uncertainties as a function of magnitude, and this is shown in Figure 3.31. When we go to assign *HST* position uncertainties to the synthetic stars, we identify the 10 nearest real COSMOS stars in G magnitude and randomly select one of those stars to bequeath its position uncertainty to the synthetic one.

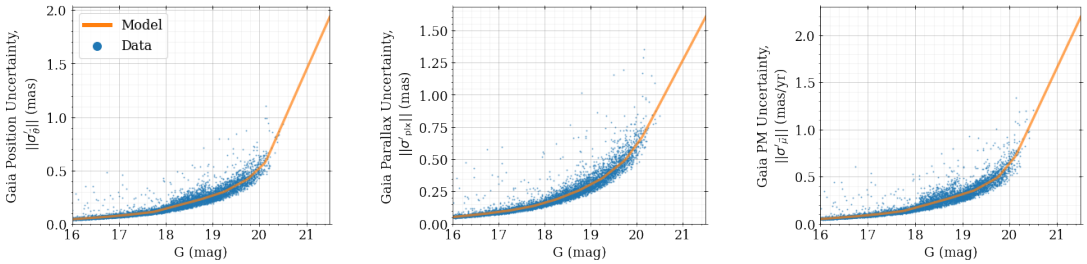

Figure 3.30: *Gaia* Position, parallax, and proper motion uncertainties as a function of magnitude from all *Gaia* sources within 1 degree of the COSMOS field center. The blue points are the *Gaia*-measured values and the orange lines are a median-binning of those data, with a linear extrapolation where no *Gaia* measurements exist.

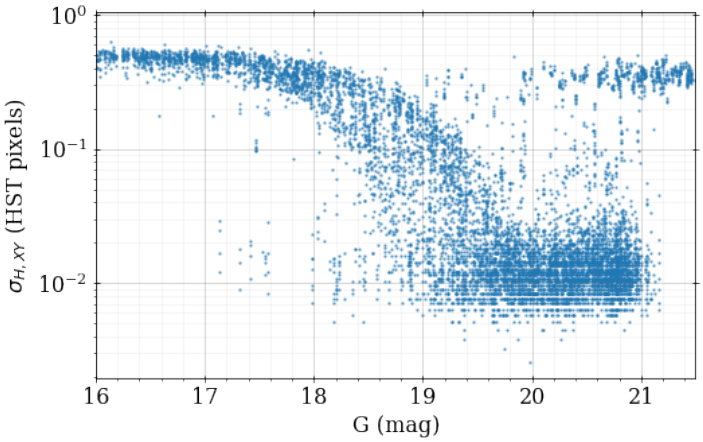

Figure 3.31: Centroid position uncertainty in *HST* pixels as a function of magnitude as measured by `GaiaHub` for the $\sim$ 2000 real COSMOS stars in Figure 3.17, where the y-axis is in log scale. In general, the brighter magnitude sources have larger *HST* position uncertainties because they are more likely to be saturated in the COSMOS exposures. At the faintest magnitudes, there are some sources with large *HST* position uncertainties because their PSF-fitting was not as well-constrained as some other faint sources.

With all of the true properties of the synthetic star defined, we can create *Gaia*-like measurements of the position, parallax, and uncertainty by applying some noise corresponding to the uncertainties we've defined. Those true positions and motions can then be played forward or backward in time until reaching the correct epoch of the synthetic *HST* image. After choosing the transformation parameters that map the synthetic *HST* image onto the synthetic *Gaia* data, we can apply the correct *HST* position noise, and then save the outputs in the same `.csv` files that `BP3M` expects as inputs. The final step of the process is to run `BP3M` on the newly-generated synthetic *HST* images.

Chapter 4

Data-driven Discovery of Diffuse Interstellar Bands with APOGEE Spectra

4.1 Introduction

Stellar spectra capture the parameters of a star's evolutionary state and record the chemical composition of the material in which it formed. Small samples of high resolution stellar spectra have been used to describe the individual element abundance distributions of the Milky Way (MW) in the local neighbourhood (e.g. Edvardsson et al. 1993; Feltzing & Gustafsson 1998; Prochaska et al. 2000; Bensby et al. 2003). With the advent of large surveys – such as RAVE (Steinmetz et al. 2006), SEGUE (Yanny et al. 2009; Rockosi et al. 2022), APOGEE (Wilson et al. 2012; Majewski et al. 2016, 2017),

Gaia-ESO (Gilmore et al. 2012), GALAH (De Silva et al. 2015; Martell et al. 2017), LAMOST (Zhao et al. 2012), and H3 (Conroy et al. 2019b) – has come the ability to map abundances across the disk, bulge, and halo of our Galaxy (e.g. Bergemann et al. 2014; Rojas-Arriagada et al. 2014; Nidever et al. 2014; Hayden et al. 2015; Buder et al. 2019, 2022; Wylie et al. 2021; Eilers et al. 2022). These large data ensembles have also enabled new, statistically-motivated questions to be tackled about topics such as the underlying dimensionality of individual abundance distributions and the information content of stellar spectra (e.g. Mitschang et al. 2014; Ting et al. 2015; Price-Jones & Bovy 2018; Ness et al. 2019, 2022; Ting et al. 2019; Feeney et al. 2021; Weinberg et al. 2022; Griffith et al. 2022). The answers to these questions are key to understanding the origin of individual elements and the utility of those elements to reconstruct the assembly history of the MW.

Chemical tagging – the ability to distinguish co-natal stars based on chemical abundances derived from spectra – is one of the foundational ideas of stellar surveys. Understanding the conditions that create particular populations of stars informs our stellar physics models and puts constraints on models of galaxy formation and evolution. In theory, stars that are born together were formed from the same gas cloud and thus share a chemical signature in their atmospheres. In practice, the level of precision required for chemical tagging is not currently feasible (< 0.02 dex; Ness et al. 2019).

The difficulties around chemical tagging become even more severe if there are unknown or unmodeled features in a spectrum, especially if those features impact wavelength regions used for measuring chemical abundances. In the visible and infrared

(IR) regimes, the largest and most obvious source of non-stellar signal comes from the Earth's atmosphere. Because of detailed measurements of the night sky's effects as well as knowing the rest-frame that spectral features are produced in, astronomers are able to account for and remove the bulk of Earth's atmosphere's signature. However, many spectra suffer from imperfect skyline and telluric removal, which leaves residual features capable of confusing spectral analysis pipelines.

Another (often ignored) source of contamination comes from intervening dust and gas along the line-of-sight (LOS) to a star. Due to the velocity offset between gas/dust clouds and stars, spectral features from the Interstellar Medium (ISM) can appear at different wavelength locations in a set of observations at different LOS in the Galaxy. This issue is complicated further when the identification or central wavelength of an ISM-based feature is unknown or poorly constrained. Without a complete and accurate model of a star's light, it is often difficult to know a priori whether a particular residual feature is caused by non-stellar sources or is simply unknown physics/missing chemical species in the model.

One common detection and characterization method for diffuse interstellar bands (DIBs) is to measure a feature's presence in multiple spectra of different stars and then to show correlations between ISM properties (e.g. extinction from dust) and that feature's strength. Efforts to detect, characterize, and map these DIBs have historically been focused on the optical regime, though a growing number of studies have been exploring the near-IR (e.g. Joblin et al. 1990; Geballe et al. 2011; Cox et al. 2014; Zasowski et al. 2015; Elyajouri et al. 2016, 2017; Tchernyshyov & Peek 2017; Tchernyshyov

Table 4.1: Most precise measurements of rest-frame wavelengths for currently-known DIBs that fall inside of the wavelength regions covered by the APOGEE spectrograph.[a] DIBs that fall between the wavelength coverage of the three APOGEE detectors have been omitted.

λ_0 (Å)	Reference
15225 ± 10	Geballe et al. (2011)
15272.42 ± 0.04	Zasowski et al. (2015)
15616.13 ± 0.07	Elyajouri et al. (2017)
15651.38 ± 0.07	Elyajouri et al. (2017)
15671.82 ± 0.03	Elyajouri et al. (2017)
15990 ± 10	Geballe et al. (2011)
16231.1 ± 0.5	Cox et al. (2014)
16571.5 ± 0.5	Cox et al. (2014)
16582.5 ± 0.5	Cox et al. (2014)
16592.5 ± 0.5	Cox et al. (2014)

[a] The Cox et al. (2014) values have been converted from their reported Air wavelengths to Vacuum.

et al. 2018; Smoker et al. 2023). For instance, the ten currently-known DIBs that fall in the near-IR H-band ($1.51-1.7$ μm) wavelengths seen by the APOGEE spectrograph are summarized in Table 4.1, which is in stark contrast to the thousands of known optical DIBs. It is particularly important to understand sources of IR features as this regime is able to peer through the dusty regions of our Galaxy's disk.

Astronomy's burgeoning "Big Data Era" has facilitated the development of novel data-driven approaches to understanding stellar spectra that are less reliant on underlying physical models. A few successful techniques to characterize stellar light include using deep learning (Leung & Bovy 2019), polynomial models of stellar labels (e.g. The Cannon; Ness et al. 2015), and non-Gaussian Processes (e.g. Feeney et al. 2021). One significant benefit of data-motivated models is that they can describe stellar features – and correlations between features – in spectra that are currently unknown to

physics-based models. Additionally, the data models do not rely on many of the simplifying assumptions that are common in synthetic models (e.g. local thermal equilibrium, 1D radial stellar models and atmospheres). Finally, data-driven models are especially useful when the physics is not well constrained, such as in the low-density environments of the ISM that are currently impossible to recreate on Earth. As a star's light passes through intervening gas and dust on its way to our telescopes, it is imprinted with ISM signatures from many chemical species whose identities and properties are generally unknown. Detailed characterization of all the signatures in a spectrum are therefore important in disentangling the origin of various spectral features – which furthers the science goals of both abundance measurements and ISM studies.

Currently, the Milky Way Mapper (MWM) program of SDSS-V is using the APOGEE spectrograph to collect millions of stellar across all regions of the MW to understand its formation history and the physics of it stars (Kollmeier et al. 2017); any improvements in the APOGEE reduction pipeline will therefore have compounding effects on the MWM science goals. Finally, better constraints on near-IR DIBs can be studied with tomography techniques (e.g. Tchernyshyov & Peek 2017; Tchernyshyov et al. 2018) to develop a more complete picture of our Galaxy's ISM.

In this chapter, we describe the APOGEE spectra and stellar parameters used in our analysis in Section 4.2 and then present a data-driven model of those spectra in Section 4.3. In Section 4.4, we study the spectral residuals to show that DIBs, tellurics, and skylines are responsible for many of the relatively-large remaining features. We remove the Earth-based residuals to detect and characterize the remaining DIB features

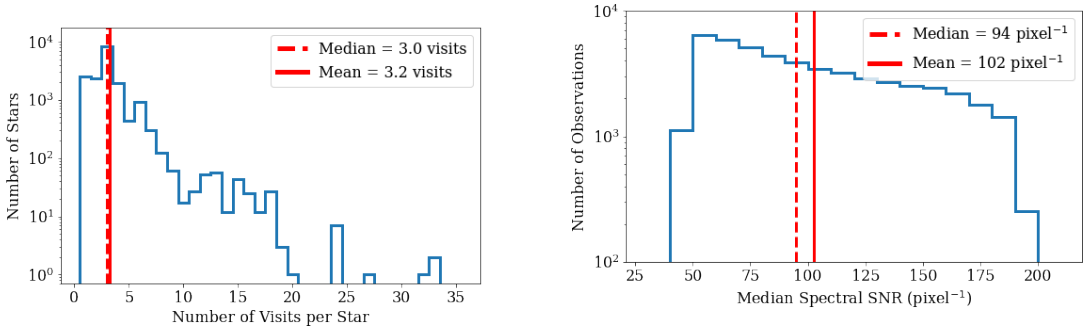

Figure 4.1: **Left:** Histograms of the number of spectroscopic visits per star. The bin width has been set to 1. The median and mean number of visits per star are shown with vertical red lines. There is a total of 55028 individual visit spectra for the 17104 RC stars in our sample. **Right:** Histogram of the median spectral SNR for each visit used in our analysis, with mean and median shown with vertical red lines.

in Section 4.5. Finally, we summarize our results in Section 4.6.

4.2 Data

This work makes extensive use of stellar spectra, abundances, and parameters of Red Clump (RC) stars in the MW as measured by the APOGEE spectrograph (Wilson et al. 2012; Majewski et al. 2016, 2017) on the Sloan Telescope at the Apache Point Observatory as a component of the Sloan Digital Sky Survey (SDSS; York et al. 2000; Eisenstein et al. 2011; Blanton et al. 2017). The RC sample – defined by Bovy et al. (2014) using stellar parameters and simulated stellar evolution – boasts high spectral signal-to-noise ratios (SNR) as well as precise stellar parameters and abundances. These properties make the RC sample an ideal population for data-driven modelling and for studying non-stellar residuals.

The APOGEE spectra cover the H-band ($\sim 15000 - 17000$ Å) with high reso-

lution ($R \sim 22500$) and fine pixel spacing ($\Delta\lambda \sim 0.2\ \mathring{A} \cdot \text{pixel}^{-1}$). The publicly-available spectral data are given in the rest-frame of each star. Our analysis uses the individual visit spectra instead of the combined visit spectra to account for changes in the LOS velocity – and, therefore, the location of non-stellar features – of each observation. Distributions of the number of visits per star and the median spectral SNR of the individual visit spectra are given in Figure 4.1.

The spectra were analysed by the APOGEE Stellar Parameter and Chemical Abundances (ASPCAP García Pérez et al. 2016) pipeline. We use the ASPCAP T_{eff}, $\log g$, [Fe/H], and [α/Fe] measurements and uncertainties from Data Release 16 (DR16; Jönsson et al. 2020), while stellar ages come from the catalogue of Sit & Ness (2020). As exemplified in Figure 4.2, the stars in our sample occupy a relatively narrow range in stellar parameters and abundances. After noticing a minor secondary peak near $([\text{Fe/H}], [\alpha/\text{Fe}]) = (-0.6, +0.2)$ dex in the [α/Fe] versus [Fe/H] panel, we removed stars with abundances above the red line to ensure that our modelling only focuses on a single chemical population.

For the individual visit spectra in our sample, we remove pixels that have SNR < 50 pixel^{-1}. We also set the maximum pixel SNR to 200 pixel^{-1} as recommended by ASPCAP, which suggests that the "uncertainty floor floor is at the level of 0.5%"[14]. Because of known superpersistence issues in the blue detector, we mask out spectral observations where the fiber number is ≤ 100; this selection removes approximately 6500 RC stars that do not have a single observation with a fiber number greater than 100. Finally, following the approach of Price-Jones & Bovy (2018), we remove data

[14] see https://www.sdss4.org/dr16/irspec/spectra/

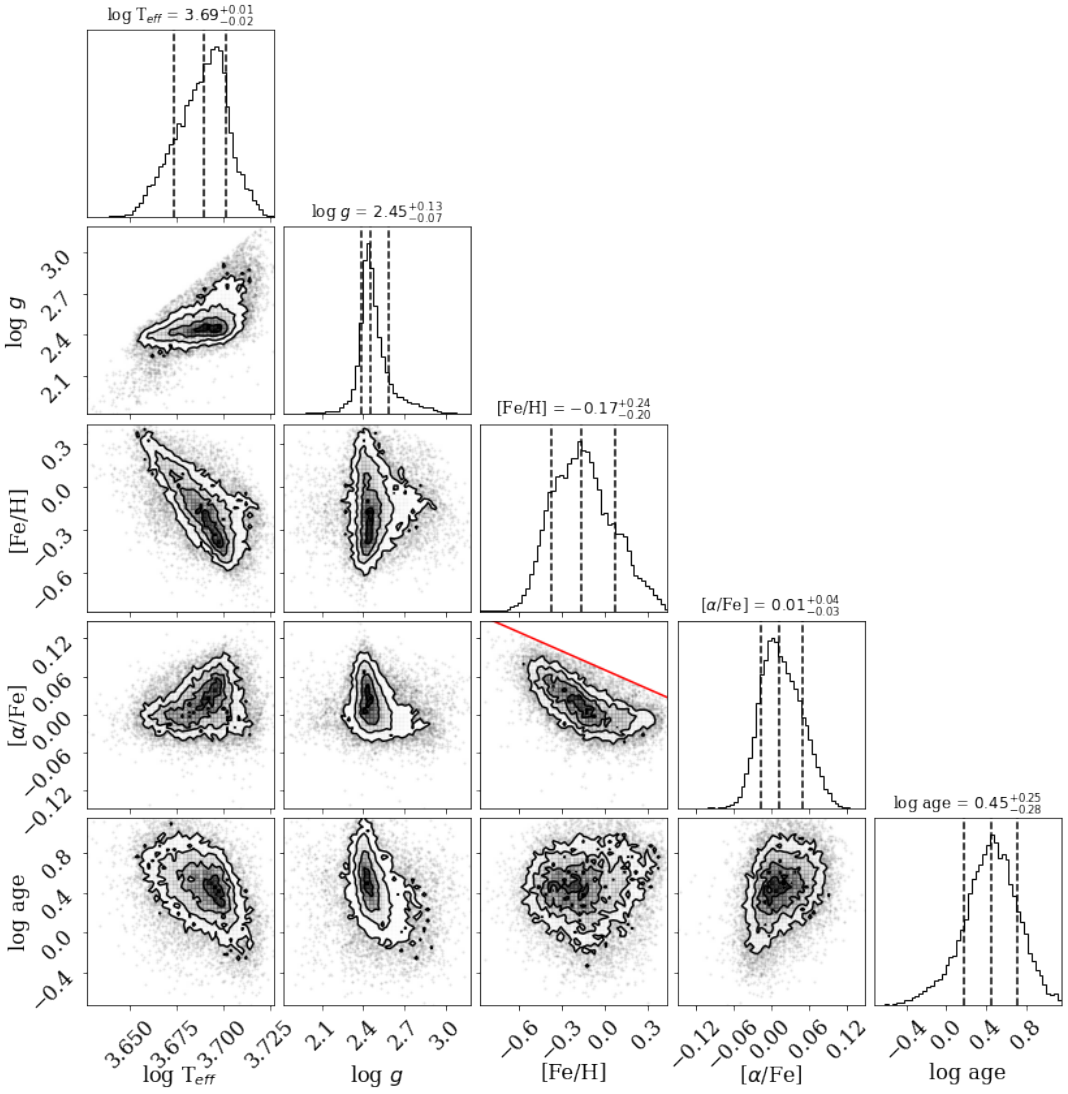

Figure 4.2: Distribution of the stellar parameters for the 17104 RC stars in our sample. The red line in the [Fe/H] versus [α/Fe] plot near the center of the figure shows where we mask out approximately 2600 stars that lie above this relationship; we do this to keep the 2D abundance distribution singly-peaked.

Table 4.2: Pixel bitmasking of spectra[a]

Bitmask	Name
0	Pixel marked as BAD in bad pixel mask
1	Pixel marked as cosmic ray in ap3d
2	Pixel marked as saturated in ap3d
3	Pixel marked as unfixable in ap3d
4	Pixel marked as bad as determined from dark frame
5	Pixel marked as bad as determined from flat frame
6	Pixel set to have very high error (not used)
7	No sky available for this pixel from sky fibers
12	Pixel falls near skyline

[a]This is a subset of Table 1 in Price-Jones & Bovy (2018), which masks use of the bitmask definitions of Holtzman et al. (2015)

from pixels that have any of the bitmask flags listed in Table 4.2.

These APOGEE individual visit spectra in the stellar rest-frame, along with the Sit & Ness (2020) ages and ASPCAP parameters and abundances, are used in combination to build data-driven models in Section 4.3, which are the basis of our analysis.

4.3 Modelling RC Spectra

4.3.1 Preprocessing Spectra

First, we continuum normalize the individual visit spectra using iterative B-spline fitting. At each iteration, the B-spline is defined by 50 Å-spaced knots. For the first iteration, all flux measurements in a spectrum are used to define the initial spline. For subsequent iterations, the new spline is measured using only flux values that are within 3σ (in flux uncertainty) for fluxes below the spline or 5σ for fluxes above the

spline. This masking of fluxes is done to ensure that strong absorption features do not overly impact the continuum measurement.

We iterate the spline fitting up to 100 times per spectrum, but stop iterating if the current iteration's spline is very similar to the previous one:

$$\sum_i \frac{|c'_{i,j,k} - c_{i,j,k}|}{c_{i,j,k}} < 1 \times 10^{-5}$$

where $c_{i,j,k}$ is the continuum spline value at pixel i from the previous iteration and $c'_{i,j,k}$ is the current iteration's continuum spline value at pixel i for observation/visit number k of star j. This condition is such that the summed absolute fractional change is smaller than 0.001%, which tends to occur when the subset of fluxes being masked hasn't changed from one iteration to another. This process usually only takes a handful of iterations (i.e. ≤ 5), and virtually all of the spectra converge on a spline well before the 100 maximum iterations.

To capture any remaining continuum, we repeat this continuum-fitting process after the first model fit to define a continuum-adjustment B-spline. We divide each individual visit spectrum by the previously-defined continuum spline and the best-fit model to get an approximate noise spectrum that may still have some continuum trends in it. We then use the same iterative B-spline fitting process as above, but using 25 Å-spaced knots and a 3σ threshold above and below for masking. The finer-spaced knots and the narrower threshold are because the residual spectrum ideally only consists of noise and any trends larger than $\sim$ 20 Å likely arise from an incomplete initial continuum removal.

The continuum-normalized flux in pixel i for spectral observation k of star j is then $y_{i,j,k} = f_{i,j,k}/c_{i,j,k}$ with corresponding uncertainty of $\sigma_{i,j,k} = \sigma_{f,i,j,k}/c_{i,j,k}$, where $f_{i,j,k}$ and $\sigma_{f,i,j,k}$ are the raw flux and uncertainty values. The continuum-normalized fluxes are then used as the inputs for the data-driven modelling.

4.3.2 Modelling Flux using Stellar Labels

Following the approach of Ness et al. (2016), we define the continuum-normalized flux in each pixel to be a 2nd-order polynomial of stellar labels (see also Price-Jones & Bovy 2018). We note that a key difference between these bodies of work is the stellar labels used in the fitting: Price-Jones & Bovy (2018) uses $T_{\rm eff}$, $\log g$, and [Fe/H], Ness et al. (2016) uses the same, but also includes [α/Fe] and mass, while our model uses $T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], and age. The vector of stellar labels in our analysis for star

j is therefore given as

$$\vec{x}_j = \begin{pmatrix} T_{\mathrm{eff},j}^2 \\ T_{\mathrm{eff},j} \times \log g_j \\ T_{\mathrm{eff},j} \times [\mathrm{Fe/H}]_j \\ T_{\mathrm{eff},j} \times [\alpha/\mathrm{Fe}]_j \\ T_{\mathrm{eff},j} \times \mathrm{Age}_j \\ T_{\mathrm{eff},j} \\ \log g_j^2 \\ \vdots \\ [\mathrm{Fe/H}]_j^2 \\ \vdots \\ [\alpha/\mathrm{Fe}]_j^2 \\ \vdots \\ \mathrm{Age}_j^2 \\ \vdots \\ 1 \end{pmatrix} \tag{4.1}$$

such that the vector contains all the parameters to the first and second powers, cross terms, and a constant. Our model of the continuum-normalized fluxes is defined as

$$y_{i,j,k} = \vec{x}_j^T \cdot \vec{\theta}_i + \varepsilon_{i,j,k} \tag{4.2}$$

where $\vec{\theta}_i$ are the coefficients for the stellar label terms in Equation 4.1 for pixel i and

$$\varepsilon_{i,j,k} \sim \mathcal{N}\left(0, \sigma_{i,j,k}\right) \tag{4.3}$$

describes the noise as a result of the uncertainty in a pixel's flux. This functional form implies that the data likelihood at pixel i is

$$\text{likelihood}_i = \prod_{j}^{n_*} \prod_{k}^{n_{\text{obs},j}} \mathcal{N}\left(y_{i,j,k} | \vec{x}_j^T \cdot \vec{\theta}_i, \sigma_{i,j,k}\right). \tag{4.4}$$

We then see that the likelihood at a given pixel is maximized when

$$\sum_{j}^{n_*} \sum_{k}^{n_{\text{obs},j}} \left(\frac{y_{i,j,k} - \vec{x}_j^T \cdot \vec{\theta}_i}{\sigma_{i,j,k}} \right)^2 \tag{4.5}$$

is minimized, which occurs when

$$\hat{\theta}_i = \left[\sum_{j}^{n_*} \sum_{k}^{n_{\text{obs},j}} \frac{1}{\sigma_{i,j,k}^2} \vec{x}_j \cdot \vec{x}_j^T \right]^{-1} \cdot \left[\sum_{j}^{n_*} \sum_{k}^{n_{\text{obs},j}} \frac{y_{i,j,k}}{\sigma_{i,j,k}^2} \vec{x}_j \right], \tag{4.6}$$

which defines the best-fit coefficient vector at pixel i for a set of normalized fluxes, uncertainties, and stellar labels.

To propagate the uncertainties on the stellar parameters to the model coefficients, we repeatedly draw realizations of the stellar parameters for each star and remeasure the best-fit coefficients at each pixel. After 500 iterations, we take the median of the best-fit coefficients to be the coefficients of final model. The residual flux

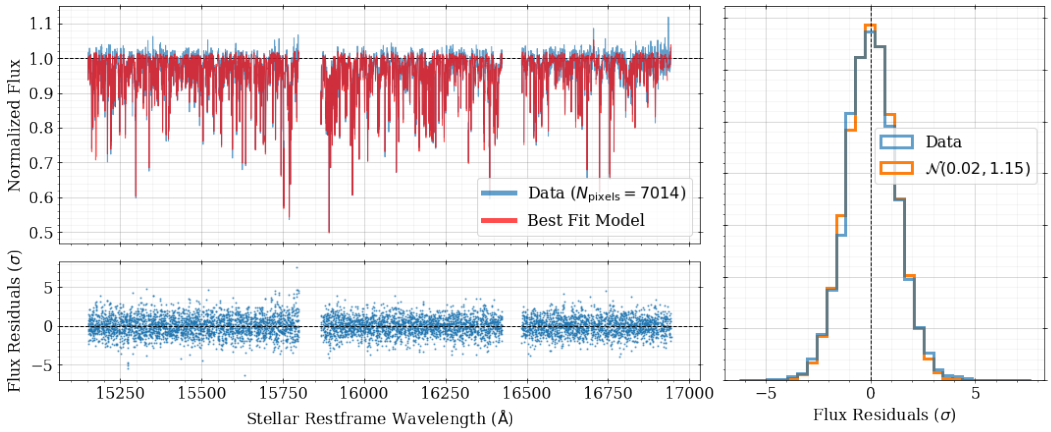

Figure 4.3: Comparison of best fit model (red line) and normalized flux (blue line) for one RC star with median SNR of 105 pixel^{-1}. The residuals in the lower left panel and the right panel have been scaled by their corresponding uncertainties. The orange histogram in the right panel is the best fit normal distribution, which shows that this star has good agreement with the data (mean near 0, standard deviation near 1) and that the difference between the data and model are larger than expected by the flux uncertainties at the level of $\sim 15\%$. The gaps in the data around 15850 Å and 16450 Å correspond to the wavelength gaps between the three APOGEE CCDs.

for pixel i of observation k of star j is defined to be

$$r_{i,j,k,l} = y_{i,j,k} - \vec{x}_{j,l}^{T} \cdot \hat{\theta}_{i,l}, \tag{4.7}$$

for realization l of the stellar parameters draws, $\vec{x}_{j,l}$, which yields a best-fit coefficient measurement of $\hat{\theta}_{i,l}$. To propagate the uncertainty on the stellar labels – and therefore the uncertainty on the best-fit coefficients – to the residuals, we also repeatedly measure the residual flux values for the 500 realizations, giving samples of $r_{i,j,k,l}$. The final residual flux, $\hat{r}_{i,j,k}$, is taken to be the median of these realizations, with an uncertainty

that is given by

$$\sigma^2_{r,i,j,k} = \sigma^2_{i,j,k} + \mathrm{var}\left(r_{i,j,k,1}, \ldots, r_{i,j,k,500}\right), \tag{4.8}$$

where $\mathrm{var}\left(r_{i,j,k,1}, \ldots, r_{i,j,k,500}\right)$ is the variance of the 500 residual measurements at a given pixel for a given spectrum. In words, the residual uncertainty is a quadrature sum of the normalized flux uncertainty and the propagation of the best-fit coefficient uncertainty.

A comparison of the best fit model and data for a single observation of one star in our sample is shown in Figure 4.3. The uncertainty-scaled residual distribution for this star is close to the expected unit Gaussian distribution (mean of 0, standard deviation of 1), which shows that the model does a good job of capturing the information in the spectrum. If we assume that the fluxes and uncertainties reported by APOGEE accurately describe the data, then the best-fit Gaussian to the residual distribution (orange histogram) having a width of $\sim 1.15\sigma$ implies that there is information in this star's spectrum that the model does not capture at the level of $\sim$ 15% of the flux uncertainties.

If we instead look at the uncertainty-scaled residual distribution across all observations and all pixels, we get the distribution shown in Figure 4.4. As before, the distribution is quite similar to the unit Gaussian, which suggests that the model performs well across all the spectral observations in our sample. Again, that the width is greater than 1σ reveals that there may be more information in the spectra than the model is able to describe.

Using the residual histogram and best-fit Gaussian distribution in Figure 4.4,

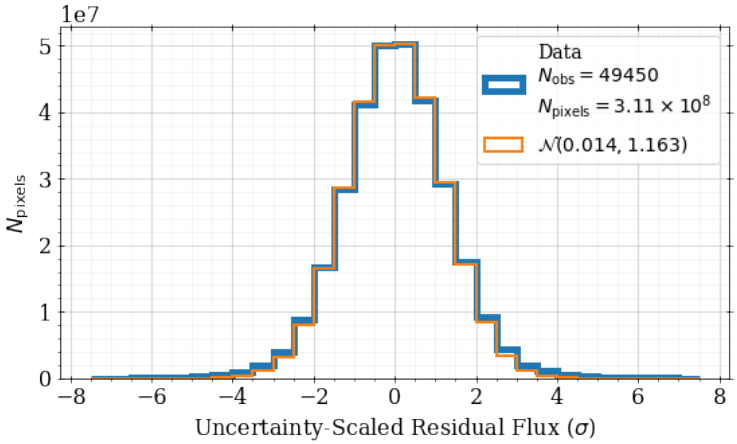

Figure 4.4: Distribution of uncertainty-scaled residuals for $\sim 3 \times 10^8$ pixels from all $\sim 5 \times 10^5$ spectral observations of ~ 17000 stars in our sample (blue histogram). The orange histogram is the best fit Gaussian, which shows that the residuals are centered near zero and have a standard deviation near 1, as would be expected if the model and flux uncertainties perfectly describe the observed fluxes. Because the best fit Gaussian width is ~ 1.16, the model doesn't perfectly describe the data within their uncertainties, as is expected if there is non-modelled information remaining.

we vertically align the distributions so that they have the same height at their peaks so that we can quantify their difference in the wings (i.e. excess in the data over the expectation). First, we find that the data excess in the wings corresponds to $\sim 3.3\%$ of the total pixels. Next, we draw many realizations of residual flux measurements from each bin in the blue data histogram where the data counts exceed the expected distribution. The results of this process are summarized in Figure 4.5, where the blue curve shows the median distribution of the size of flux residual in excess of the expectation. The blue point above the histogram shows the median and 68% region of the distribution, and the grey histograms show individual realizations of residual flux draws. In summary, the residuals that are not explained by the model have a median size of $\sim 2.9\%$ of the stellar flux, with a 68% region covering 2% to 5% of the stellar flux.

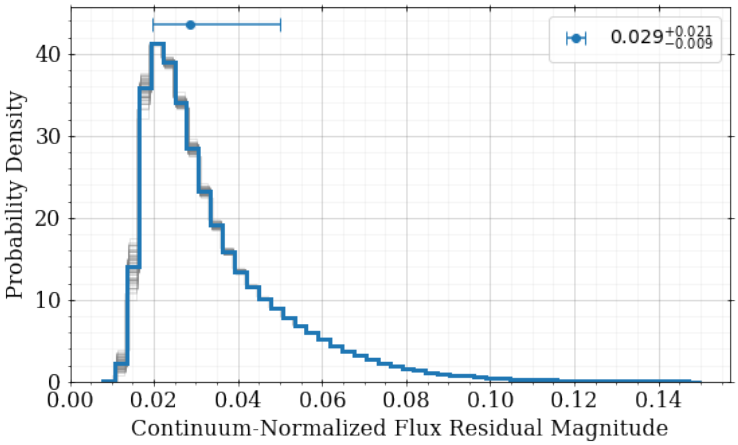

Figure 4.5: Distribution on the size of residual fluxes that are in excess of the best-fit Gaussian in Figure 4.4. The grey histograms show different realizations of measuring the excess, while the blue histogram shows the median of these realizations. The blue point at the top of the figure summarizes the median and 68% region of the distributions, revealing that the non-Gaussian excess of residuals is $\sim 3\%$ of the stellar flux on average.

4.4 Structure in Residuals

To further explore the residuals, we begin by looking for trends in the flux residuals as a function of wavelength (in stellar rest-frame) and other explanatory variables. In many APOGEE pixels, we find that the model almost perfectly describes the data; for instance, the pixel summarized in the left panels of Figure 4.6 shows a residual distribution (scaled by the residual uncertainties) that agrees very well with the unit Gaussian. When the stars are split up into different heliocentric velocity ($V_{\rm HELIO}$) bins – as is shown in the left middle panel – we see little difference between the residual distributions, and no obvious trend in those distributions' medians or standard deviations (left bottom panels).

On the other hand, there are some pixels where the residual distributions are

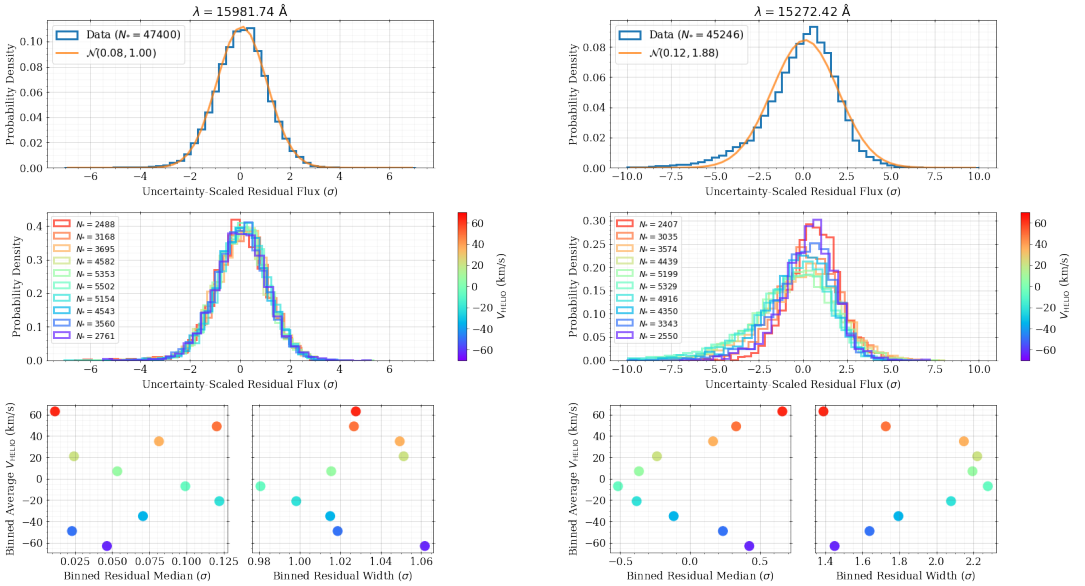

Figure 4.6: Residual fluxes scaled by their uncertainty (i.e. [normalized flux−model]/uncertainty) for all the stars in two different APOGEE pixels. The wavelength of each pixel, in stellar rest-frame, is listed above the top panel in columns of figures. By highlighting a single pixel, we focus on the residual information from all of the spectra at a particular stellar rest-frame wavelength. **Left:** A well-modelled pixel, where the uncertainty-scaled flux residuals (blue in top panel) follow the expected unit normal (orange in top panel). In the middle panel, the residuals are binned by heliocentric velocity (number of spectra in each bin given by N_* in the legend), and we see no differences between the velocity bins. The bottom two panels show the median and standard deviations of the velocity-binned histograms, again showing no noticeable trend in the residuals with velocity. **Right:** A pixel where the residuals do not follow a unit Gaussian, or any Gaussian for that matter. After binning by heliocentric velocity in the middle panel, and plotting the binned medians and standard deviations in the bottom two panels, we see a noticeable trend in the median and width of the residual distribution with velocity.

distinctly non-Gaussian and have widths that are much larger than 1σ. An example of this is shown in the right panels of Figure 4.6 for a pixel near the central wavelength of the strongest known DIB in the APOGEE wavelength range (e.g. $\lambda_0 = 15272.42$ Å, from Zasowski et al. 2015). Breaking the stars up into $V_{\rm HELIO}$ bins in the middle right panel, we now see significant differences between the residual distributions' shapes as well as their medians and widths (right bottom panels. In general, the extreme velocity bins have more positive residual medians and smaller standard deviations while the moderate velocity bins have negative residual medians and larger widths.

We next look at the trends in the residuals with heliocentric velocity across neighbouring pixels at different wavelength cutouts. First, we sort the stars by their heliocentric velocity and then smooth the residuals (using a combination of inverse variance weighting and Gaussian smoothing based on $V_{\rm HELIO}$ at each pixel for each residual spectrum) to produce Figure 4.7. Each of the 18 panels show the smoothed residuals using 21 APOGEE pixels (i.e. width of ~ 4 Å) centered on a particular feature, with the central wavelength of that feature listed on the right edge of the cutout. These regions correspond to 15 elemental features used by ASPCAP to measure abundances (these are the same central wavelengths as used by Feeney et al. 2021), 1 region we identify as having minimal visible features (i.e. continuum), 1 region where we notice a large amount of residual Earth-based contamination, and 1 region around the strongest DIB in the APOGEE wavelength range.

The y-axis of each panel in Figure 4.7 corresponds to wavelength relative to the central wavelength label on the right of each plot, with blue wavelengths at the top

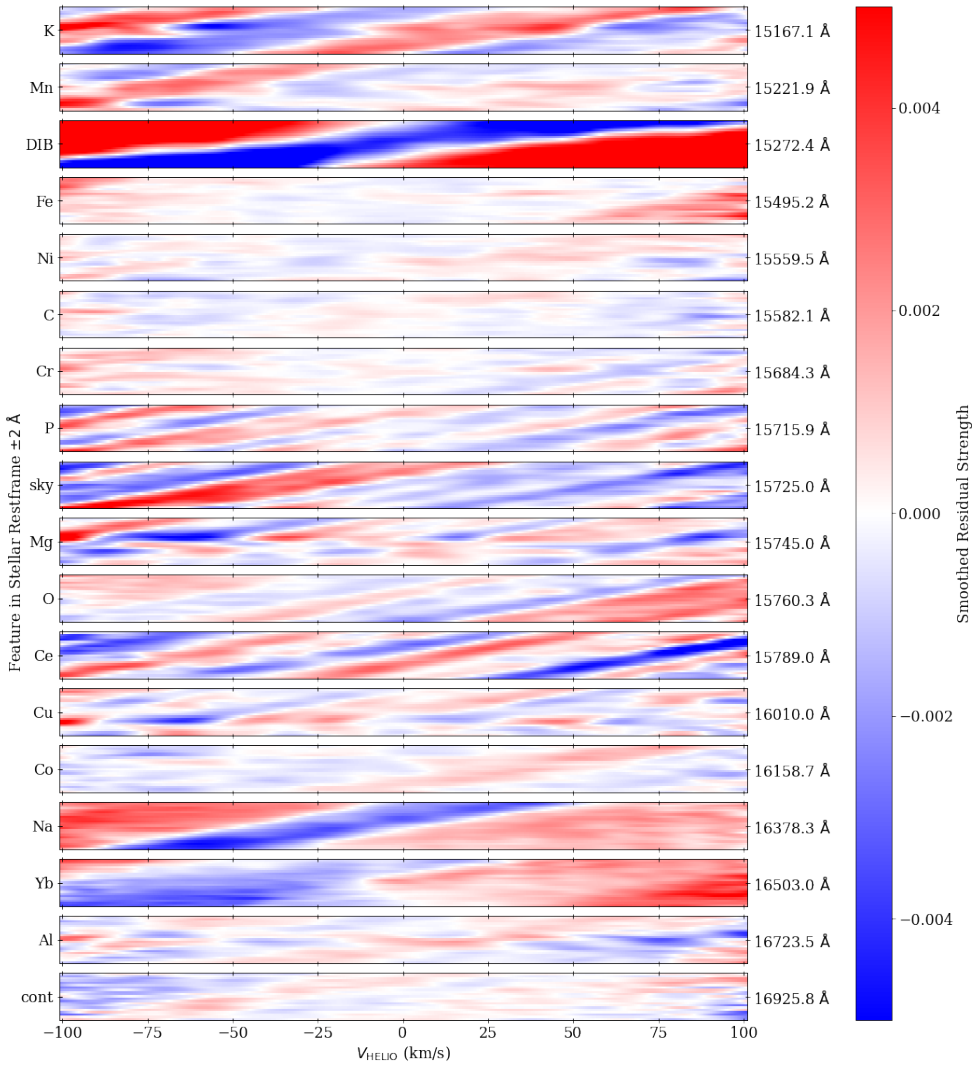

Figure 4.7: Smoothed residual spectra of all RC stars with $|V_{\mathrm{HELIO}}| < 100$ km/s, sorted by the star's V_{HELIO}. Each column corresponds to a residual spectrum and each row corresponds to a pixel/wavelength, with bluest wavelengths at the top and reddest wavelengths at the bottom. This image focuses on 18 different features (wavelengths listed on the right edge of the figure) with each panel consisting of 21 pixels; these cutouts include 15 elemental features, 1 region we've identified as being mostly continuum, 1 region of high telluric contamination, and 1 region near the strongest known DIB in APOGEE (15272 Å feature). The residual spectra have been Gaussian-smoothed in the horizontal direction with a kernel width of 5 km/s. Some elemental regions (e.g. the DIB region, 3rd from the top) show a residual minimum that moves through the element window as a function of heliocentric velocity; this can be explained as an unmodelled residual feature that is not in the same reference frame as the star.

and red wavelengths at the bottom. The x-axis is the sorted list of heliocentric velocities of the stars such that a vertical column in this figure corresponds to a single residual spectrum of a particular observation with that velocity.

By eye, some of the wavelength cutouts (e.g. Fe, Ni, C, Cr, continuum) show small amplitude in smoothed residuals and not much correlation with heliocentric velocity. Other cutouts (e.g. K, DIB, P, sky, Na, Yb) show relatively strong trends in residuals with $V_{\rm HELIO}$.

Focusing on the DIB panel in particular, we see a residual minimum (i.e. an absorption feature) move through the cutout as a function of heliocentric velocity. This is because the strong DIB feature, while present in many of the APOGEE spectra, appears at different wavelength locations in the stellar rest-frame spectrum as a result of the offset between the velocity of the star and the velocity of the DIB-producing source. The wavelength of a DIB feature in a star's rest-frame spectrum is given by:

$$\lambda_{\rm rest,*} = \lambda_{\rm rest,DIB} \cdot \left(1 + \frac{v_* + \Delta v}{c}\right) \cdot \left(1 + \frac{v_*}{c}\right)^{-1} \tag{4.9}$$

where $\lambda_{\rm rest,DIB}$ is the wavelength of the DIB in its own rest-frame, v_* is the star's heliocentric velocity, and Δv is the LOS velocity offset between the DIB source and the star. Because DIB features will show up at different wavelength locations in the stellar rest-frame spectra, our model is not able to describe the feature and thus leaves DIB residuals behind. Similarly, features from the Earth's sky are also likely to show up as residuals because they too appear at different wavelengths in the stellar rest-frame

spectra:

$$\lambda_{\mathrm{rest},*} = \lambda_{\mathrm{rest,Earth}} \cdot \left(1 + \frac{v_*}{c}\right)^{-1}. \tag{4.10}$$

For the Earth-based features, sorting by heliocentric velocity means that the features will occur at similar wavelength locations, which is why they appear strengthened in the velocity-sorted, smoothed residual panels. Comparing the features in the sky and the DIB panels, we also notice a difference in the horizontal width of the features; DIB features are known to be quite broad, so it agrees with expectations that the DIB panel residual has a visually larger width than, for example, the width of the local maximum diagonal stripe in the left half of the sky panel. Based on this difference, the narrow width of features in the P and Ce panels suggest these residuals are likely caused by the Earth instead of DIBs, while the large width of the residual minima stripes in the K and Na panels may originate from DIBs.

The particular pattern seen in the DIB panel of Figure 4.7 is a result of the average RC heliocentric velocity having a correlation with the average ISM heliocentric velocity in the MW disk. Because the RC stars in our sample are generally quite old (average of ~ 2.8 Gyr, from Figure 4.2), they have had a longer time to kinematically decouple from the gaseous disk they were likely born in. In the Galactocentric cylindrical radius range of our stellar sample – 95% of the APOGEE RC sample is within $6.4 < R < 13.4$ kpc with a median at 10.1 kpc – the gaseous disk has been seen to have relatively flat rotation at ~ 220 km/s (e.g., from the HII measurements of Brand & Blitz 1993). Using RC stars from APOGEE and GALAH, Khanna et al. (2019) measure rotation velocities that are also relatively constant ($210 < V_\phi < 230$ km/s) in the $6 < R < 11$ kpc region

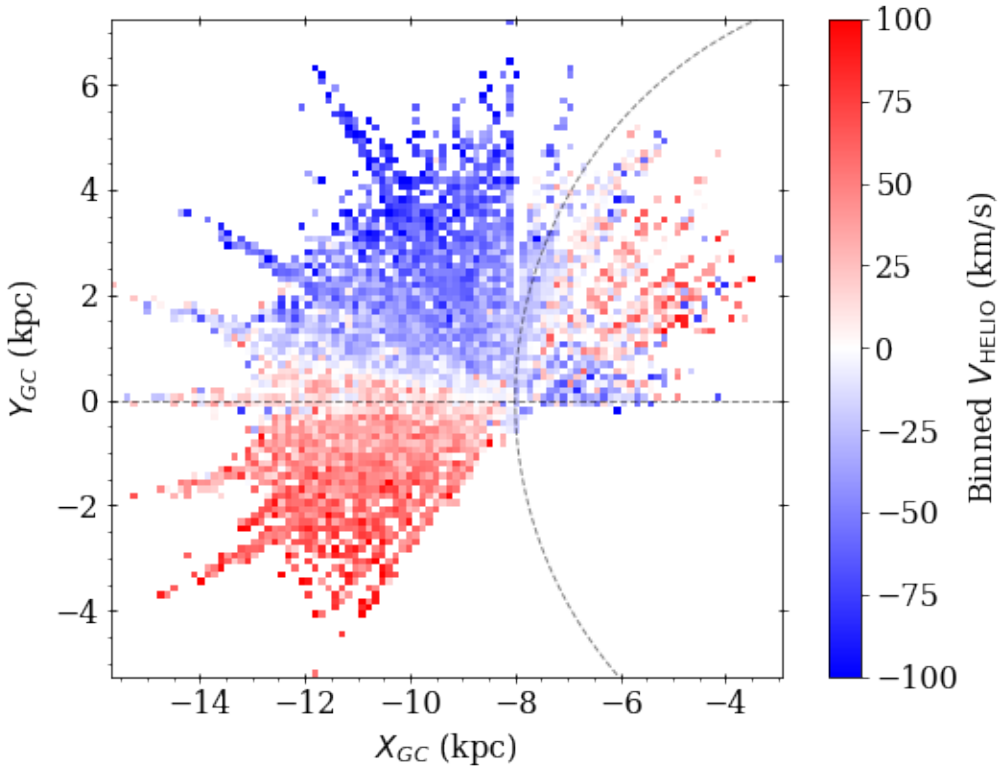

Figure 4.8: Binned heliocentric velocity in the Galactic disk for the RC stars in the APOGEE DR16 sample with $|Z| < 0.75$ kpc. The dashed circle of radius 8 kpc and the horizontal dashed line intersect at the approximate location of the sun.

for $|Z| < 0.75$ kpc. In particular, their $R > 9$ kpc measurements – where approximately 75% of our APOGEE RC stars fall – shows remarkably stable $V_\phi \sim 215$ km/s. This suggests that RC stars can be thought of as belonging to a rigidly rotating disk that is spinning ~ 5 km/s slower than the gaseous disk.

When we sort the DIB-based features by heliocentric velocity, we are largely sorting by azimuthal angle, as can be seen in Figure 4.8. If we assume both the RC stars and the ISM are described by rigid-rotating disks, then there is an average stellar V_{HELIO} and an average DIB V_{HELIO} for each bin in azimuthal angle; it is the relationship

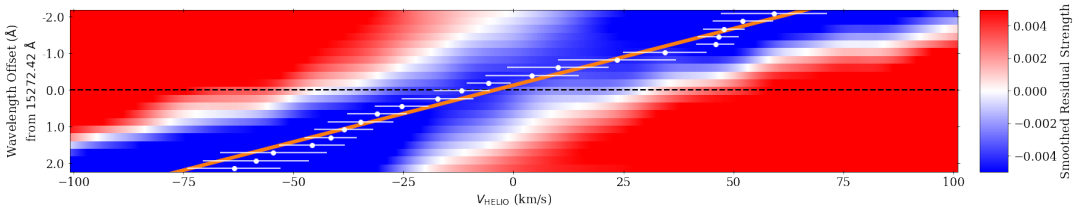

Figure 4.9: Tracing the 15272 Å DIB feature location to estimate the velocity offset between DIB sources and RC stars. The smoothed residual data are the same as the 3rd-from-top panel of Figure 4.7. The white points showing the approximate location of the residual minimum at each pixel in stellar rest-frame, and the orange line shows the best fit line to these data. Using the known rest-frame wavelength of this DIB feature, the orange line is used to infer the $V_{\rm HELIO}$ of the DIB source as a function of stellar $V_{\rm HELIO}$.

between the average stellar $V_{\rm HELIO}$ and average DIB $V_{\rm HELIO}$ that produces the pattern in the DIB panel. We can explore this relationship explicitly by tracing the location of the DIB minimum as a function of stellar rest-frame wavelength. This is done in Figure 4.9, where the orange best fit line to the DIB minimum location and the known rest wavelength of the DIB feature allow us to estimate the DIB source velocity as a function of stellar $V_{\rm HELIO}$.

We next compare the relationship we measure for gas $V_{\rm HELIO}$ as a function of stellar $V_{\rm HELIO}$ to our expectations. To that end, we assume the RC stars all belong to a rigidly rotating disk with $V_\phi \sim 215$ km/s. Similarly, we assume there is also a gaseous disk rotating at $V_\phi \sim 220$ km/s. For each RC star in Figure 4.8, we use the star's Galactic position to obtain the expected velocity vector from the rotating stellar disk, which is then transformed into a heliocentric velocity after accounting for the sun's position and motion[15]. We follow a similar procedure for generating the simulated

[15]We use the same values for solar position and motion as

$V_{\rm HELIO}$ of the gas, but we now need to assume a distance from the sun to the DIB source along each LOS. To be agnositic of the exact DIB source distance, we choose a fraction of the total LOS distance to each star, and use those new implied Galactic coordinates and the gas disk to measure $V_{\rm HELIO}$ for the intervening gas. We repeat this process for distance fractions of 100%, 50%, 10%, and 1% of the total distance to each star.

A comparison of the velocity difference between simulated gas and simulated stars is shown in Figure 4.10, where the data points are colored by the different fractional distances of the gas. The orange line is the result of using the best fit line in Figure 4.9 with the known rest-frame wavelength of the DIB to measure heliocentric velocity of the DIB as a function of stellar $V_{\rm HELIO}$; to be clear, the orange line in Figure 4.10 is not a fit to the simulated data. We see good agreement between the orange line and the simulated cases when the DIB sources are, on average, between 1% and 50% of the distance to the stars. This implies that the velocity offset function we measure from the location of the DIB minimum in the stellar rest-frame is caused by a source that is in the foreground of the stars, as we would expect for an ISM-based absorption feature.

Additionally, we see that the gas velocity offset in the simulated data has a relatively tight correlation with stellar velocity. This is ultimately what causes the pattern we see in the DIB panel of the smoothed residual in Figure 4.7; for the RC sample, stellar velocity has an approximately linear relationship with the gas offset

Khanna et al. (2019):

$$(X, Y, Z)_\odot = (-8, 0, 0) \text{ kpc}$$

and

$$(U, V, W)_\odot = (11.1, 241.92, 7.25) \text{ km/s}.$$

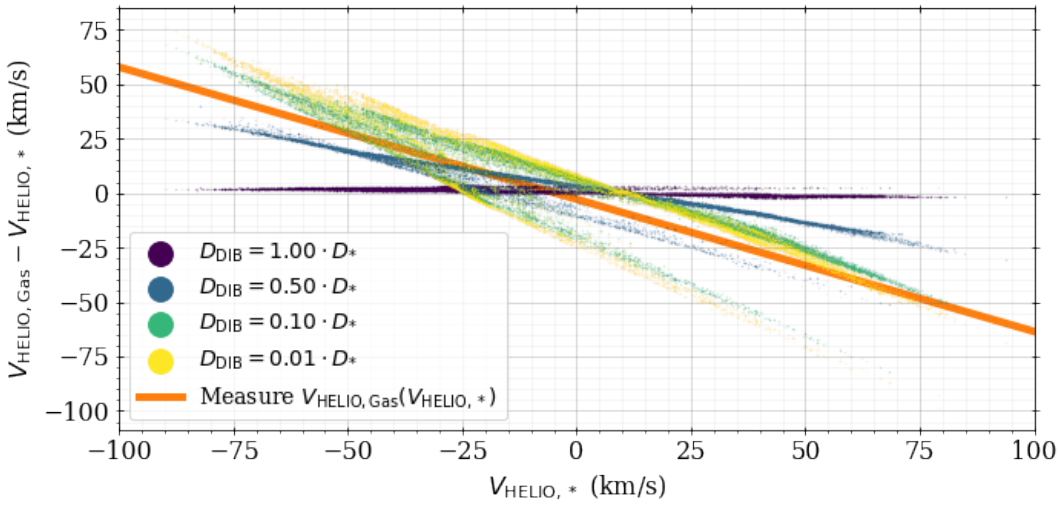

Figure 4.10: Simulated comparison of the difference in heliocentric velocity between gas and stars when the intervening gas is at different fractions of the stellar LOS distance. The simulated stellar disk is rotating at 215 km/s while the simulated gaseous disk is rotating at 220 km/s, and measurements are generated for each RC star in Figure 4.8. The orange line is the result of using the known DIB rest-frame wavelength with the orange line in Figure 4.9 to estimate DIB velocity as a function of stellar velocity. The agreement between the orange line and the results of placing the DIB sources between 1% and 50% of the distance to each star suggests that the DIB absorption feature is being produced by intervening material, as we would expect. The tight relationship between $\Delta V_{\mathrm{HELIO}}$ and V_{HELIO} explains why binning in stellar heliocentric velocity for the RC sample leads to increased DIB residual strength in Figure 4.7.

velocity, so the wavelength location of a DIB in the stellar rest-frame, Equation 4.9, effectively becomes a function of stellar velocity alone. This explains why sorting by V_{HELIO} amplifies DIB signals and causes their location to move smoothly in wavelength.

By comparing the apparent strength of residual spectral features with various parameters, we found a slight correlation in residual strength with spectral SNR. In particular, residual spectra with lower median SNR show higher levels of contamination, especially in wavelengths regions where tellurics and skylines are known to occur. We shift the residual spectra to the observer reference frame using each observa-

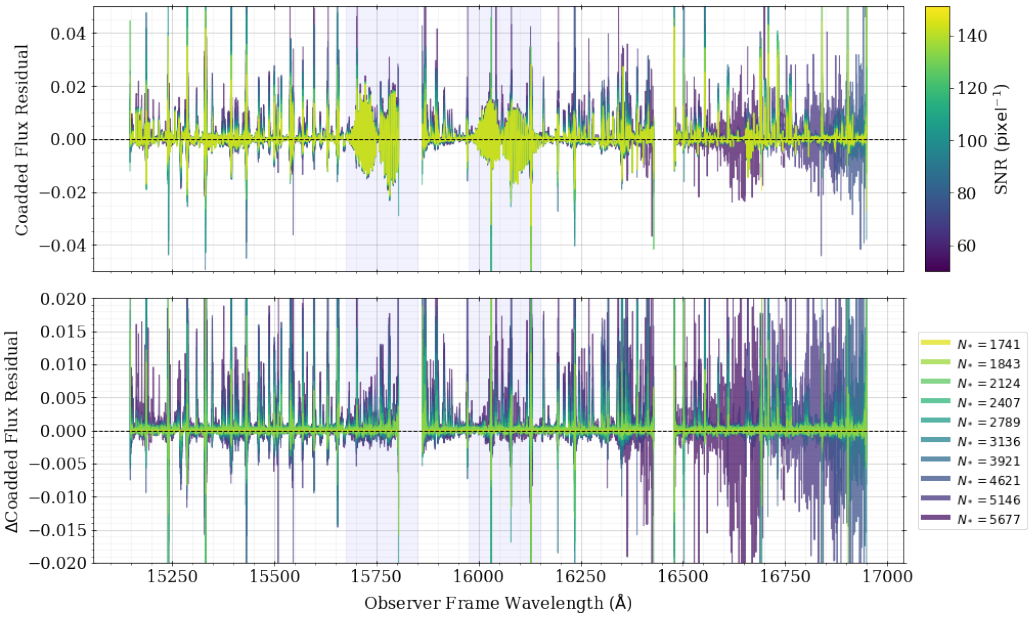

Figure 4.11: Inverse-variance coadded residual spectra in the observer reference frame in different spectral SNR bins. The blue shaded regions highlight where the strongest telluric features (caused by CO_2) occur in APOGEE. The top panel shows the raw coadded residual spectra, and each combined spectrum has a similar median SNR of ~ 550 pixel^{-1}; this panel demonstrates that the residual spikes of the lower SNR bins are larger than those of the high SNR bins. The bottom panel takes the coadded spectra from the top panel and subtracts off the combined spectrum of highest SNR bin (yellowest line in top panel) to better highlight the difference between the spectra of the SNR bins.

tion's LOS velocity to allow for Earth-based features to align in wavelength. We then inverse-variance combine residual spectra in different spectral SNR bins, which yields the coadded spectra shown in Figure 4.11; the combined uncertainty in each of the SNR bins are approximately the same, with median combined SNRs of ~ 550 pixel^{-1}. The top panel shows the coadded residual flux in 10 different SNR bins, and this panel reveals that the extrema of the low SNR bins are generally larger in magnitude than the

extrema of the higher SNR bins. The bottom panel is a similar plot, but now the spectrum of the highest SNR bin has been subtracted from all the spectra of the other SNR bins to highlight the change in residuals as a function of SNR. The faint blue regions in both panels corresponds to the strongest telluric absorption region in APOGEE (e.g. see Figure 19 of Nidever et al. 2015).

In even the highest SNR bins, the residual features are commonly on the order of $1\% - 2\%$ of the stellar flux, and as high as $4\% - 5\%$ in the extreme cases. The structure we observe in our residual spectra reveals that a complicated combination of Earth-based features and DIBs are common across the APOGEE wavelength range. These trends of residuals with spectral SNR and heliocentric velocity highlighted in this work are what prompted the authors of Ness et al. (2022) to include SNR and $V_{\rm HELIO}$ as explanatory variables in their regression fitting to measure the relationship between abundances and stellar evolutionary state.

We posit that the non-stellar residuals seen in Figure 4.7 may be responsible for the larger-than-expected scatter in the some of abundance distributions measured by ASPCAP (e.g. Zasowski et al. 2019; Jönsson et al. 2020); some elemental abundances (e.g. Na) are determined by flux measurements at a small number of pixels, so the presence of DIBs at those wavelengths may have a relatively large impact. If we attribute a large portion of the residual excess in Figure 4.5 to non-stellar features, then it wouldn't be unexpected to see a large effect from features with local minima that are $\sim 3\%$ of the stellar flux.

4.5 Removing Skylines and Tellurics to Identify Diffuse Interstellar Bands

To identify the DIBs present in the residual spectra, we first need to characterize and remove the more common contamination from Earth-based features. One advantage of isolating the Earth-based residuals versus the DIBs is that we are easily able to shift to the rest-frame of the skyline and telluric residuals.

Because almost all the spectra have residual Earth-based features, but not all spectra have DIB contamination, it is useful to split the residual spectra up into two groups: those with and without the strongest DIB feature. Once we have these two groups, we can study the Earth-based contamination using the low-DIB-strength group. Those results can then be used to remove the Earth-based contamination from the high-DIB-strength group.

To accomplish this, we need to rank the spectra by their DIB strength. We fit an inverted Gaussian to the strong DIB feature that is present around 15272 Å DIB in each residual spectrum, after shifting to the heliocentric frame using the $V_{\rm HELIO}$ for each observation; an example of this is show in Figure 4.12. This allows us to measure a LOS velocity and equivalent width for the strong DIB in each residual spectrum. We then use the EW measurements to define a low- and a high-DIB-strength group; residual spectra that failed to fit an absorption feature near 15272 Å DIB were automatically assigned to the low-DIB-strength group, and then a threshold of EW $= 1 \times 10^{-2}$ Å

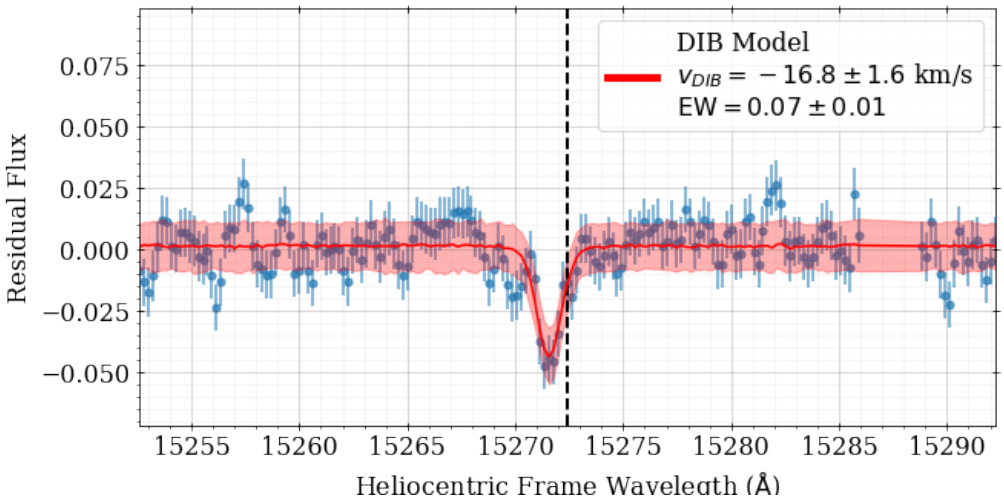

Figure 4.12: Fitting the 15272 Å DIB feature with an inverted Gaussian to the residuals of the spectrum in Figure 4.3. The blue data points show the residual data, the red line shows the best fit inverted Gaussian model with a 68% error envelope, and the dashed vertical black line marks the rest-frame wavelength of the 15272 Å DIB feature.

was chosen to divide the spectra with DIB detections to ensure an approximately equal number of spectra ($\sim$ 21000) in the low- and high-DIB-strength groups. For the residual spectra in the high-DIB-strength group, the 15272 Å DIB EW is compared to the K-band reddening in Figure 4.13 with data points colored by the EW bins used in our later analysis.

For the low-DIB-strength group, we fit a 2nd-order polynomial model to the observer-frame residual fluxes at each pixel using a process similar to what is described in Section 4.3.2. In this case, however, the stellar label vector in Equation 4.1 is replaced

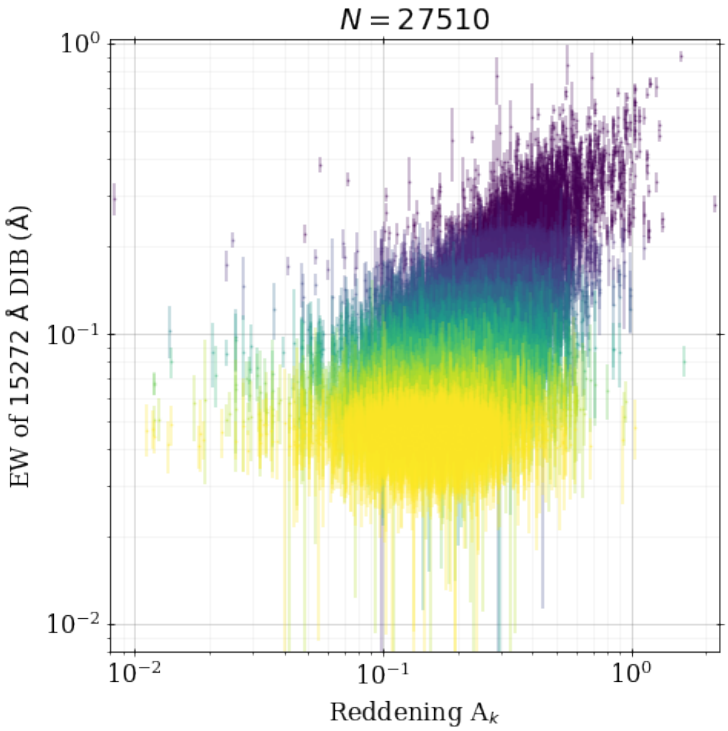

Figure 4.13: Equivalent width strength of the 15272 Å DIB feature versus K-band reddening for the spectra in our sample with well-measured 15272 Å DIB features. The different colors correspond to the 10 different EW bins used later in combining the residuals. There are approximately 2100 residual spectra per bin.

by

$$\vec{x}_j = \begin{pmatrix} \mathrm{SNR}_j^2 \\ \mathrm{SNR}_j \\ 1 \end{pmatrix} \tag{4.11}$$

and the data we are fitting are the residuals in the observer frame. This model enables us to quantify the predictive power of SNR alone to describe the spectral residuals that are not captured by our five label model from Section 4.3.2.

Examples of the resulting best-fit model for different SNR bins are shown in

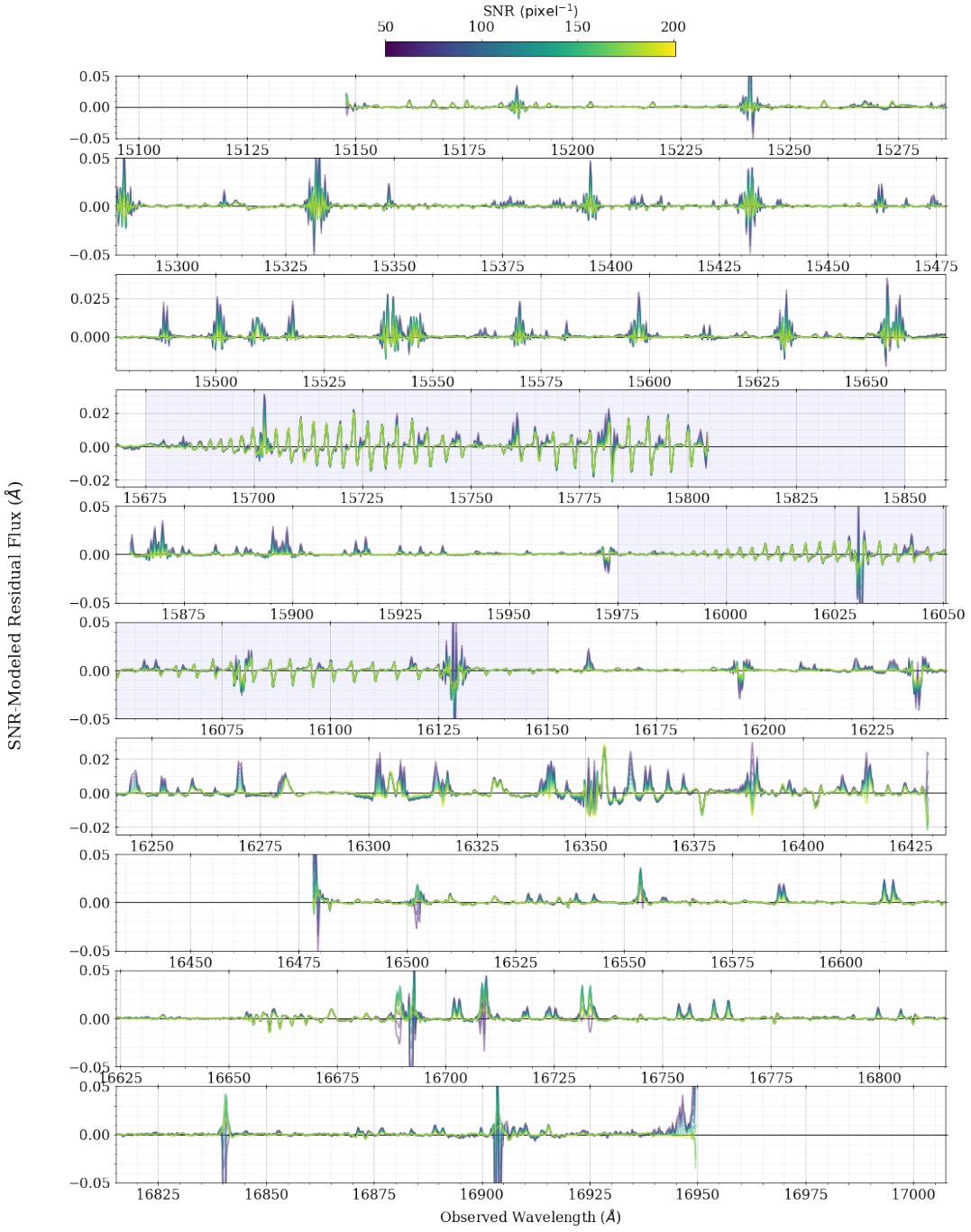

Figure 4.14: SNR-modeled residual spectra in the observer rest-frame. The colors of the spectra in this figure correspond to the SNR values in the colorbar above the top panel, and not the EW bins in Figure 4.13. The blue shaded regions highlight where the strongest telluric features (caused by CO_2) occur in APOGEE.

Figure 4.14. Overall, this model captures much of the behaviour we see in the coadded residuals of Figure 4.11. There are a few regions that stand out visually (e.g. in the blue shaded regions where CO_2 telluric features are particularly strong) that have a saw-toothed shape to the residuals; this is a shape that can be created by subtracting two Gaussians with the same width but a slight offset in mean. We suggest that these Earth-based residuals are caused by a wavelength mismatch between the raw spectra and the sky models being subtracted.

As shown in Nidever et al. (2015), the APOGEE telluric and sky removal performs very well in most cases. Holtzman et al. (2018) further shows improvement in APOGEE's Earth-feature removal; however, they also point out that the APOGEE reduction pipeline's approach is to flag and ignore pixels near particularly strong telluric and sky lines instead of a more computationally expensive approach that might achieve smaller residuals. Of particular interest, Figure 3 of Holtzman et al. (2018) reveals the same saw-tooth residuals that we see in our Figure 4.14, with their figure focusing on the $15680 - 15800$ Å region that is sensitive to tellurics from CO_2. While the APOGEE reduction process works well for their science goals, our analysis has shown that detailed accounting of Earth's atmosphere's light can reveal previously-hidden information.

By looking at the wavelengths and shapes of the Earth-based features in Figure 4.14, we can see that some of the smoothed residual patterns in Figure 4.7 are likely explained by the night sky. The Ce panel of Figure 4.7, for example, shows a repeating alternation of residual fluxes from low to high, which is visually dissimilar to the DIB panel which shows a single feature moving through the cutout window. The Ce panel's

central wavelength of 15789.1 Å places it in a region of saw-toothed features in Figure 4.14, so this mottled pattern that appears after sorting by heliocentric velocity is likely the result of these saw-tooth shapes constructively and destructively combining.

While this SNR-dependent Earth-residual model performs well in describing the observer-frame residuals, we choose to not use this model in our subsequent DIB analysis without extensive further testing. Instead, we remove skyline and telluric features by pairing up residual spectra in the low- and high-DIB-strength groups based on spectral SNR. That is, for each high-DIB-strength residual spectrum, we identify a low-DIB-strength residual spectrum with median spectral SNR that is within 5 pixel^{-1} of the high-DIB-strength spectrum's SNR. We ensure that each low-DIB-strength residual spectrum is used only once.

Once each high-DIB-strength residual spectrum has a corresponding low-DIB-strength residual spectrum, we subtract each pair and add their uncertainties in quadrature in the observer frame. This has the effect of removing the Earth-based residuals from the high-DIB-strength residual spectra while leaving the possible DIB features relatively untouched. The spectra are then shifted from the observer frame to the 15272 Å DIB rest-frame using the velocity measured in the inverted-Gaussian fitting for the high-DIB-strength spectrum in the pair.

The next step is to break the spectra up into different bins based on the strength of the 15272 Å DIB. We define 10 DIB-strength bins such that approximately 2100 spectra are in each bin; these are shown by the different colors in Figure 4.13. We then combine the residual spectra in these bins using population fitting to measure a

population mean and uncertainty/standard deviation at each pixel. The precise statistics used for this combination are explained in Appendix 4.7.1, but the key takeaway is that the population fitting takes into account both the uncertainty in each residual measurement as well as the spread in those measurements to give back realistic means and uncertainties at each pixel.

The results of this population-fitting process are shown in Figure 4.15, where the colors denote the same DIB-strength bins as in Figure 4.13. The vertical red lines show the locations of the previously-known DIBs in Table 4.1; by eye, the increasing DIB-strength bins show increasing depth of features at most of these wavelengths. For the 15225 Å feature, the large literature uncertainty in central wavelength may suggest that the true central wavelength of that DIB should correspond to one of the features near 15250 Å instead. We also notice additional absorption features (e.g. slightly redward of 15700 Å) that have shapes similar to the previously-known DIBs.

We recognize that all the DIB features are likely not in the same rest-frame as the 15272 Å DIB, as they might be produced by different intervening clouds along the sight-line to each star and therefore have different relative LOS velocities. However, the velocity difference between DIB sources along a single LOS aren't likely to be large enough to shift the wavelength by more than a few APOGEE pixels. For instance, the vast majority of the 15272 Å DIB velocities we measure are in the $\pm$30 km/s range, so at $\sim$ 16000 Å, a 10 km/s velocity offset would manifest as a $\sim$ 0.5 Å offset in wavelength. When we combine spectra in each DIB-strength bin, we may slightly reduce the signal of DIB features from other sources because of a combination of wavelength offsets and

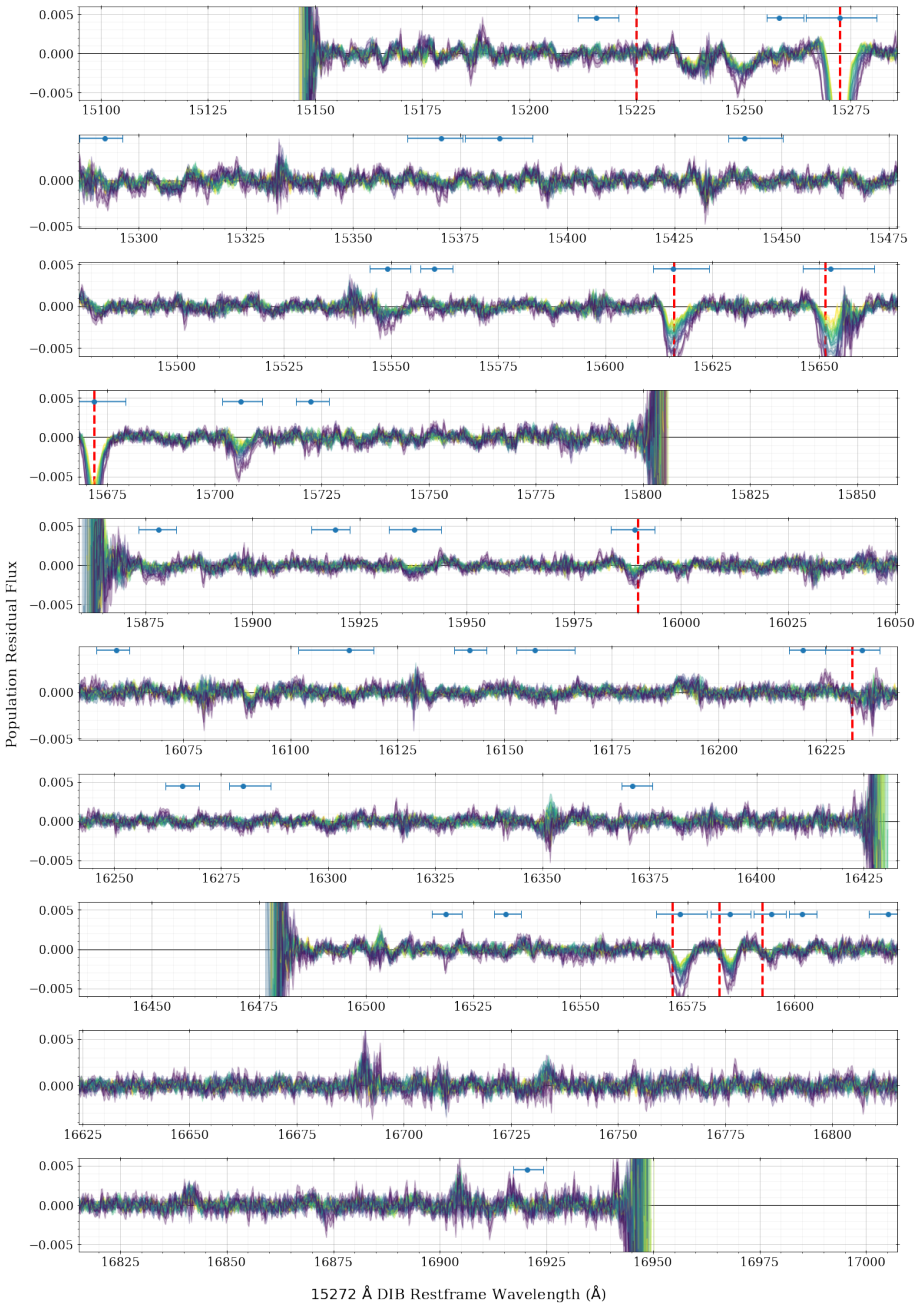

Figure 4.15: Coadded residual spectra in the 15272 Å DIB rest-frame colored by the 15272 Å DIB strength as shown in Figure 4.13. The vertical red lines show the rest-frame wavelengths of the previously-known DIB features in Table 4.1. There are many newly discovered features in emission and absorption that scale as a function of EW bin, making them new candidate DIB features. 35 of the possible DIBs we detect – which we will later measure to have significant correlation with the 15272 Å DIB strength – are shown by the blue points above the spectra with widths showing the region we use as a cutout when measuring equivalent widths.

their strength not correlating with the 15272 Å strength, but we still expect to detect their presence. Our decision to use the 15272 Å DIB rest-frame will, however, preserve the strength of features that are being produced by the 15272 Å DIB source, so this technique is particularly useful for identifying DIBs that correlate with the strongest-DIB. For future work where extremely precise central wavelengths are needed or the goal is to find every possible DIB in the APOGEE wavelength range, one approach would be to cross-correlate the residual spectra using different wavelength cutouts to find the velocities that amplify the signals of all DIBs.

To identify possible locations of DIBs, we smooth the highest DIB-strength residual spectrum (i.e. the darkest purple line in Figure 4.15) with a Gaussian kernel width of 5 pixels; after some visual vetting, this yields 133 local maxima and minima as possible DIBs in emission and absorption. Of the possible DIBs, we highlight the locations of 35 features in Figure 4.15 that we will ultimately find are likely produced by the same source that produces the 15272 Å DIB; we choose to not show all 133 possible features for visual clarity. As expected, this step finds features nearby to all of the previously-known DIBs. Many of the local extrema we find are likely spurious detections, but we choose to err on the side of testing too many extrema versus applying a more complicated thresholding at the detection step. For each local extremum, we then step redward and blueward from that wavelength in the smoothed spectrum until the slope is near 0 to define a useful wavelength region around each feature. We then manually check these regions to confirm that they cover the entirety of an apparent feature, tweaking the boundaries where necessary. This defines the wavelength cutouts

we will later use for measuring the EW strength of the possible DIBs and for quantifying DIB detection probabilities. The list of possible features agree well with the features we find when we smooth a population-combined spectrum using all $\sim$ 21000 residual spectral pairs; we choose to use the feature locations from the highest DIB-strength spectrum instead of the total population because the signal from true DIBs will be strongest in the high DIB-strength bin.

As mentioned in the last few paragraphs, the central wavelengths we measure in the 15272 Å DIB rest-frame for DIBs that are produced by a different source may be offset from their true value by a few pixels, but this is still an improvement on the wavelength location for a few of the previously-known DIBs in Table 4.1. Additionally, because the spectra in each DIB-strength bin are a combination of $\sim$ 2100 individual residual spectra, a systematic velocity offset between different DIB sources would be required to produce a significant shift in the central wavelength location. This implies that our central wavelength measurements are likely within a handful of pixels of the true rest-frame wavelength of each DIB feature. To be conservative, we estimate the uncertainty on the central wavelengths of the possible DIBs to be $\sim$ 1 Å.

We next measure the equivalent width of the local extrema in each of the DIB-strength bins of Figure 4.15. Using the wavelength cutouts defined above, we empirically integrate the flux. Specifically, we use the 5 pixels at the edges of each cutout (3 inside, 2 outside of the cutout) to define an average flux and wavelength on the blue and red edge; these blue and red average fluxes define a line for each DIB-strength bin that we use at the local continuum measurement. After subtracting off the local continuum line

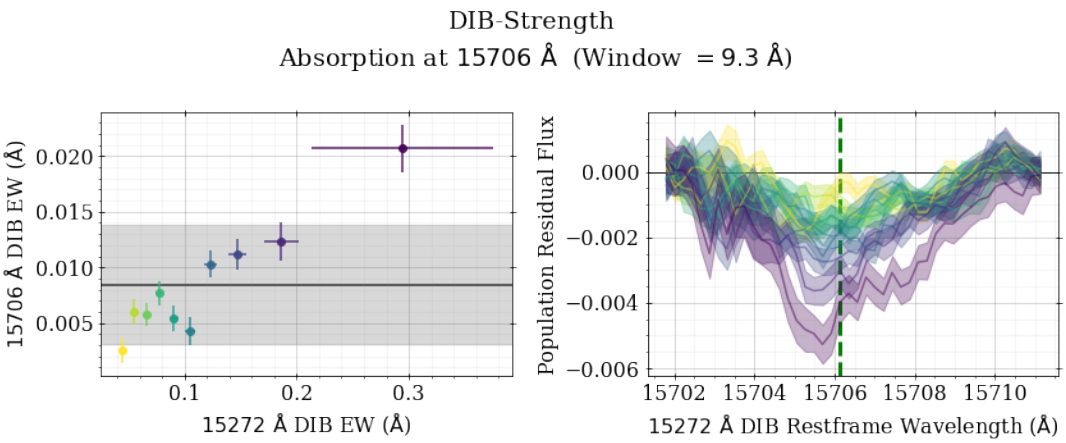

Figure 4.16: Equivalent width strength of a previously-unknown DIB feature near 15706 Å for the different EW bins shown in Figure 4.13. The "Window" in the title defines the width of the wavelength cutout, and the dashed vertical green line in the right panel shows the location of the local minimum identified in the smoothed highest DIB-strength residual spectrum. The black horizontal line in the left panel and the corresponding grey shaded region are the population mean and width of the current DIB EW measurements (i.e. population fit to the y-axis values). This particular feature's EW shows a fairly strong correlation with the strength of the 15272 Å DIB.

for each bin, we perform a trapezoid numerical integration to measure the area of the feature. To propagate uncertainties in the flux, we repeated sample flux measurements – including the fluxes used to measure the local continuum line – to get a distribution on the empirical EW for each bin. The EW measurements we report are the median and standard deviation of those realizations. An example of the resulting EW measurements is shown in the left panel of Figure 4.16 for the possible DIB near 15706 Å, with a cutout of the feature shown in the right panel for the different DIB-strength bins.

To characterize how much the EW measurements change between the bins, we perform a population fit to the EW measurements of the possible DIB feature (e.g. the y-axis values of the left panel); this population fit again follows the approach detailed in Appendix 4.7.1, and an example of the resulting medians on the population mean

and width are shown with a black line and grey region in the left panel of Figure 4.16. We are particularly interested in the population width because a large width indicates that the EW measurements between the bins are quite different.

Of course, each possible DIB feature may only be a spurious detection; that is, a feature may only show up in the highest DIB-strength bin by random chance alone. To determine the significance of a detection, we repeat the binning, combining, and EW-measuring process, but this time, we bin randomly. To be clear, instead of using the strong-DIB-strength ranking to bin our spectra, we assign spectra randomly into 10 different bins, with the same $\sim$ 2100 spectra per bin. Using random binning, we expect that the features in each bin will be quite similar to each other, as will their EW measurements.

To quantify exactly how strong a detection is (i.e. its significance above randomness), we also measure the population width of the EW measurements from random binning. In particular, we are interested in how the population width distributions compare between the random binning and the binning by the strong-DIB-strength cases. To account for the variation in the random binning that might occur by chance alone, we repeat the random binning 5 times. The resulting population width distributions of the populations agree quite well between the realizations for each possible feature. This implies that our results aren't overly sensitive to the particular choice in random binning. To be careful, however, we average the population width distributions from the 5 random realizations together. Measurements from this point on that refer to "random" come from the "averaged random" results.

Recognizing that strong-DIB-strength might not be the only/best parameter to explain the change in a feature's strength between bins, we explore a few additional binning options: [Fe/H], median spectral SNR, and A_K reddening. Respectively, these can be thought of as testing for features that are the result of residual chemical information that the model did not capture, remaining Earth-based contamination or other SNR effects, and DIBs that correlate with the amount of LOS dust but don't originate from the 15272 Å DIB source. As before, we sort by a given parameter and then bin the spectra so that there are approximately 2100 spectra per bin. In every case, we measure EWs in each bin at each possible feature, and then use those EWs to measure a population width distribution, which are then compared to the results from random binning. An example of these population width distributions for the different sorting parameters is shown in the left panel of Figure 4.17; in the right panel, samples from the random distribution have been subtracted from samples in the other distributions to get a distribution of $\sigma_{\mathrm{SORT}} - \sigma_{\mathrm{RANDOM}}$.

We can then directly integrate these population width difference distributions to measure the probability that the population width for a given sorting parameter is greater than the population width from random sorting. In cases where none of the probabilities are greater than 50%, we decide that randomness alone likely produced the observed feature. We also use the medians of these distributions to determine which sorting parameter yields the greatest difference from random, implying that a particular feature is best explained by that parameter. Because the relationship between each feature's EW measurements and a given sorting parameter is potentially complicated (i.e.

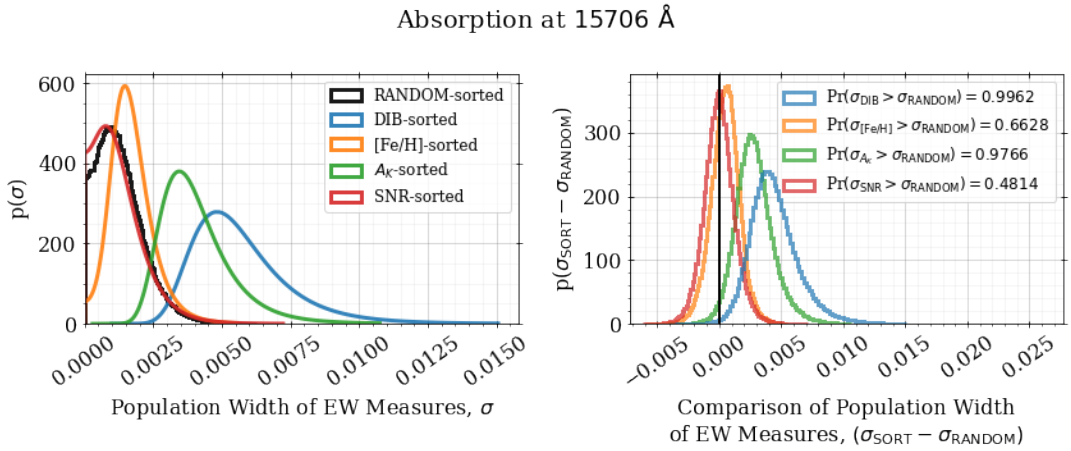

Figure 4.17: Comparison of the population width distributions of EW measurements around the possible DIB absorption feature shown in Figure 4.16 when sorted and binned by different parameters. In each case, we measure the EW of the possible DIB feature from the combined residual spectra in different parameter bins, and then measure the population width of the EW measurements (e.g. the shaded grey region of Figure 4.16). **Left:** Distributions on the EW population widths for different parameter bins; the black histogram is the resulting average of 5 realizations of random sorting and binning. **Right:** Distribution on the difference between the population widths in the left panel and the random population width. In this case, the largest population width is achieved after binning by DIB-strength, suggesting that the absorption feature in Figure 4.16 is best correlated with 15272 Å DIB-strength.

not a simple functional form), we argue that our approach is a cautious, statistically-motivated way of making detections without assuming a parameterized relationship. We are simply answering the question: "Which of the sorting parameters produces the largest difference in EW measurements between the bins?" If a feature has its best sorting from DIB-strength, it may suggest that feature is produced by the same source as the 15272 Å DIB. If the best sorting is A_K, it may be that feature is still truly a DIB, but that it has a different source than the 15272 Å DIB.

We summarize the probabilities from all the sorting parameters for all 133 possible features in Figure 4.18. At our best estimate of the wavelength for each feature,

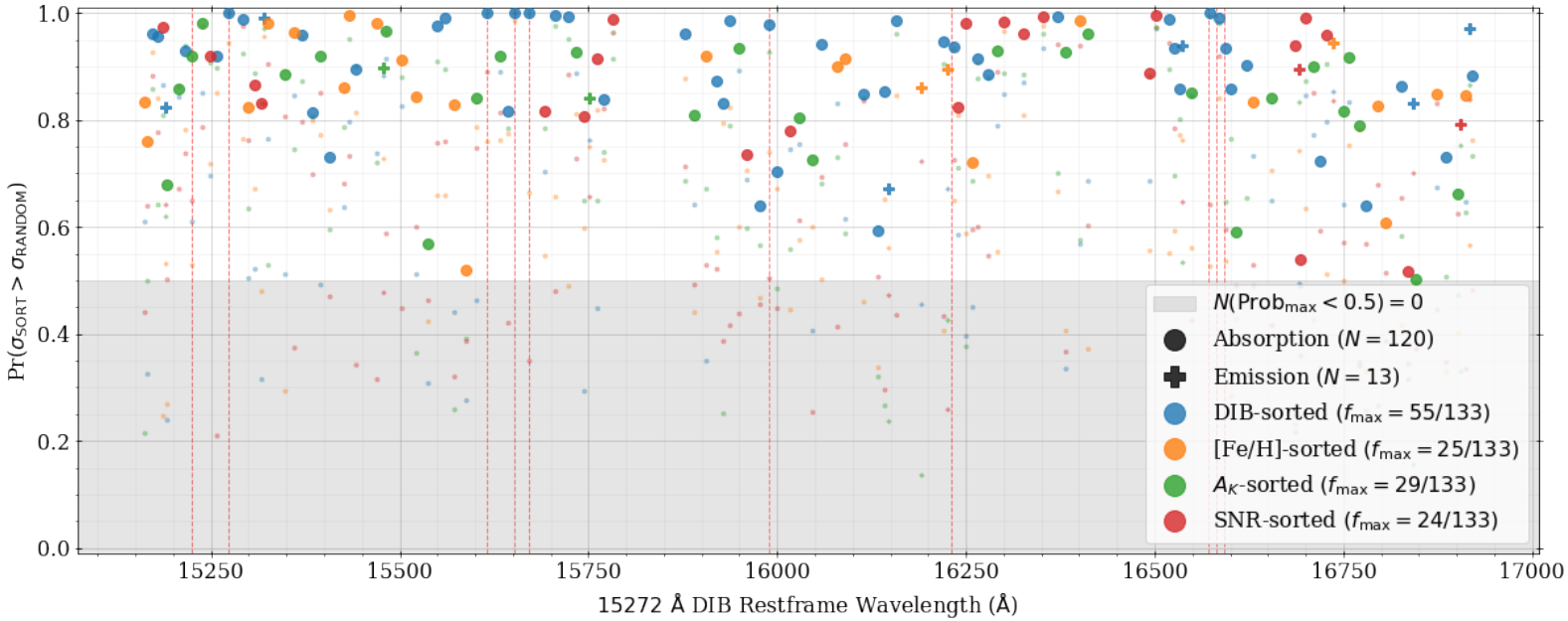

Figure 4.18: Summary of the probabilities that a particular sorting parameter's population width of EW measurements is greater than random chance can explain (e.g. the legend values in the right panel of Figure 4.17) for all possible DIB features. The possible DIB wavelengths are our best estimates of each feature's central wavelength; the exact locations of the features that are best described by A_K or DIB-strength sorting are given provided in a later table. There are 0 possible DIB features where the maximum probability of all sorting parameters is less than 50%, meaning that none of the features are best explained by chance alone. At each possible DIB wavelength, the four sorting parameter's probabilities are plotted in a vertical line; the sorting parameter with the maximum median population width is enlarged compared to the others. Of the 133 possible DIB features, 55 show the largest difference from random when binned by 15272 Å DIB strength, another 29 features are potentially DIBs produced by sources other than the 15272 Å DIB, 24 features may originate from SNR effects, and 25 features are best explained by [Fe/H] binning which may be a result of the model imperfectly describing the data.

a vertical line of 4 points are plotted to show the probability above random for each of the different sorting parameters; the point with the highest median population width is enlarged compared to the others, provided that the maximum probability above random is greater than 50% (i.e. there is only one enlarged point per feature). These probabilities reveals that 0 of the features are best explained by random chance alone (i.e. spurious detections), while the other 55, 25, 29, and 24 features correlate best with the 15272 Å DIB strength, [Fe/H], A_K, and spectral SNR respectively. Comparing to the previously-measured DIBs (vertical red lines), we recover $> 93\%$ detection probabilities in each case. Apart from the 15225 Å DIB – which finds best sorting using A_K – all remaining previously-measured DIBs in Table 4.1 have best sorting from DIB-strength. These results may suggest that most of the previously-known DIBs in APOGEE are produced by the same source/species along each LOS while the remaining DIB has a different origin.

While we may not have perfectly captured the central wavelength or cutout region around each of the possible features, we emphasize that our process is likely biased towards calling a true DIB spurious than the other way around. A more rigorously-defined cutout region and fitting of the local continuum would likely yield stronger detection signals than we report, suggesting there may be even more DIB features than we discover. Highly-detailed analyses in the future may reveal DIBs that are not included in our results, and possible DIBs we classify as having non-ISM origin may be found to have significant signal.

In general, the average probability above random for detected features is higher

Table 4.3: Number of the 133 possible DIBs with best sorting parameter detection probabilities above different thresholds.

Probability	Number of Detected DIBs by SORT Parameter				Number below
Threshold	DIB-strength	[Fe/H]	A_K	SNR	Threshold
0.50	55	25	29	24	0
0.68	51	23	24	22	13
0.95	22	5	3	9	94
0.99	11	1	0	3	118

for the DIB-strength sorting compared to the other sorting parameters (e.g. mean $\mathrm{Pr_{DIB}} = 0.89$ versus mean $\mathrm{Pr_{A_K}} = 0.83$). This is likely a consequence of our choice to bin in the 15272 Å DIB rest-frame; as discussed previously, DIBs from other sources with slightly different rest-frames will experience a reduced signal, so their difference from random is not as strong. Table 4.3 lists the number of features in each bin with detection probabilities (i.e. probability above random) above various thresholds. This is equivalent to asking how many of the enlarged points persist above changing heights of the shaded grey region in Figure 4.18. We notice that the number of detected features in the A_K sorting drops off faster than the number of features in the DIB-strength bin, and this again is likely caused by binning in the 15272 Å DIB rest-frame.

Linking the resulting DIBs back to the features seen in Figure 4.7, many of the element windows occur near newly-detected DIBs. The Na panel, for instance, is near a detected feature at 16382.1 Å that is best explained by A_K sorting with 93% probability above random. This feature likely explains the diagonal streak in the smoothed residuals of the Na panel. It may also be responsible for increasing the scatter in Na abundances that ASPCAP reports, emphasizing the need to account for DIBs in future stellar abundance pipelines.

To summarize the patterns in the $V_{\rm HELIO}$-sorted, smoothed residuals, the Fe, Ni, C, and Cr panels show minimal residual features suggesting these windows are free of significant non-stellar light. Like the DIB panel, the K, Mn, Co, Na, and Yb panels have nearby detected DIB features that likely explain their broad diagonal stripes. The P, Mg, Ce, Cu, Al, and even the continuum panels have narrower diagonal stripes like the sky panel and have no significant nearby detected DIBs, suggesting that these patterns are almost exclusively explained by Earth-based residuals. The O panel is unique in that it shows a mix of both DIB signal (i.e. broad emission near the right edge) and telluric residuals (i.e. narrower diagonals near the middle).

As an initial step towards identifying the chemical species that producing each DIB, we compare the central wavelength locations to known hydrogen transitions in the APOGEE wavelength range. The newly-detected DIB in Figure 4.16, for instance, is within 1 Å of the Brackett series $n = 15$ to $n = 4$ transition ($\lambda_0 = 15705.0$ Å), suggesting that atomic hydrogen in the ISM is a probable source of this feature. That we are able to detect features at many known hydrogen wavelengths without a priori searching there bodes well for our general methods.

A complete summary of the 84/133 DIBs whose best sorting parameter is either DIB-strength or A_K is listed in Table 4.4. Eight of the features are in emission and the remaining seventy-six are in absorption. All wavelengths are given in the 15272 Å DIB rest-frame, and the Wavelength Range column gives the window we used to measure the EWs; the continuum wavelengths are defined by the 5 pixels nearest to the values in the Wavelength Range column. By summing the probabilities above random, we

expect that 73.3 of the 84 DIB detections are truly DIBs. For features that have central wavelengths within 2 Å of a known hydrogen recombination line, we give the name of the series and level. Many of these newly-discovered DIBs occur in the same wavelength regions that were obscured by incomplete skyline and telluric removal; this, combined with their sub-percent sizes, explains why our analysis has been able to reveal these features for the first time.

Table 4.4: Summary of the 84 DIB features (76 in absorption, 8 in emission) where the best sorting parameter is either A_K or 15272 Å DIB-strength. Only features where $\Pr(\sigma_{\rm SORT} > \sigma_{\rm RANDOM}) > 0.5$ are included.

λ_0[a]	Hydrogen	Feature	Wavelength Range[c]	$\Pr(\sigma_{\rm SORT} > \sigma_{\rm RANDOM})$		Best Sort	EW[d](mÅ) in bin			
(Å)	Transition[b]	Type	(Å)	DIB	A_K	Parameter	min DIB	max DIB	min A_K	max A_K
15172.08		Absorption	15169.6 – 15175.4	0.9607	0.8572	DIB	-0.8 ± 0.9	5.6 ± 1.5	-0.8 ± 0.8	4.6 ± 1.5
15178.79		Absorption	15176.1 – 15181.3	0.9565	0.6423	DIB	-1.4 ± 0.8	4.3 ± 1.4	0.3 ± 0.7	2.1 ± 1.3
15189.07		Emission	15186.8 – 15192.2	0.8240	0.6193	DIB	-3.4 ± 0.8	-10.0 ± 1.6	-3.6 ± 0.8	-6.1 ± 1.4
15191.80		Absorption	15189.5 – 15194.5	0.2387	0.6785	A_K	2.3 ± 0.7	3.7 ± 1.3	1.4 ± 0.7	0.8 ± 1.2
15206.50		Absorption	15203.1 – 15210.1	0.8433	0.8584	A_K	3.0 ± 0.6	4.9 ± 1.3	4.4 ± 0.7	1.8 ± 1.3
15215.54		Absorption	15211.3 – 15220.8	0.9289	0.9387	DIB	-1.8 ± 0.9	7.8 ± 1.8	-0.9 ± 1.0	4.3 ± 1.8
15224.37*		Absorption	15221.8 – 15226.7	0.6093	0.9189	A_K	-0.0 ± 0.5	1.2 ± 1.0	-1.5 ± 0.5	2.1 ± 0.9
15237.62		Absorption	15233.2 – 15245.0	0.8508	0.9808	A_K	16.0 ± 1.0	23.0 ± 2.4	13.2 ± 1.3	10.2 ± 2.2
15258.27		Absorption	15255.3 – 15264.0	0.9202	0.8982	DIB	2.4 ± 0.8	6.8 ± 1.5	6.0 ± 0.8	0.5 ± 1.3
15272.40*		Absorption	15264.6 – 15281.0	0.9999	0.9999	DIB	59.4 ± 1.5	303.0 ± 2.8	75.3 ± 1.5	215.5 ± 2.6
15292.03		Absorption	15286.1 – 15296.3	0.9883	0.9240	DIB	5.8 ± 1.1	15.7 ± 2.2	3.5 ± 1.1	12.3 ± 2.0
15319.31		Emission	15315.7 – 15325.2	0.9906	0.9707	DIB	-0.2 ± 1.0	-14.7 ± 1.8	2.7 ± 1.0	-8.5 ± 1.8
15347.49	Br18	Absorption	15344.5 – 15351.5	0.5131	0.8855	A_K	-0.2 ± 0.6	3.7 ± 1.1	-0.4 ± 0.6	5.1 ± 1.2

continued ...

. . . continued

λ_0[a] (Å)	Hydrogen Transition[b]	Feature Type	Wavelength Range[c] (Å)	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) DIB	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) A_K	Best Sort Parameter	EW[d](mÅ) in bin min DIB	EW[d](mÅ) in bin max DIB	EW[d](mÅ) in bin min A_K	EW[d](mÅ) in bin max A_K
15370.61		Absorption	15362.8 – 15375.5	0.9578	0.8730	DIB	3.4 ± 1.1	14.9 ± 2.3	5.0 ± 1.1	9.1 ± 1.9
15384.21		Absorption	15376.1 – 15391.9	0.8128	0.8068	DIB	5.3 ± 1.4	9.8 ± 2.8	1.8 ± 1.4	2.7 ± 2.7
15394.84		Absorption	15392.1 – 15397.8	0.4916	0.9186	A_K	2.6 ± 0.7	5.5 ± 1.3	4.3 ± 0.7	3.2 ± 1.3
15407.18		Absorption	15404.6 – 15410.6	0.7304	0.5966	DIB	1.3 ± 0.6	4.0 ± 1.1	1.2 ± 0.6	2.5 ± 1.0
15441.28	Br17	Absorption	15437.6 – 15450.2	0.8960	0.8995	DIB	-1.3 ± 1.0	10.0 ± 2.0	3.4 ± 1.0	5.3 ± 1.8
15477.59		Emission	15469.9 – 15481.2	0.8836	0.8983	A_K	-5.9 ± 1.0	-15.9 ± 2.0	-7.5 ± 1.0	-12.2 ± 1.7
15481.22		Absorption	15477.8 – 15485.7	0.9155	0.9664	A_K	0.2 ± 0.7	8.3 ± 1.3	6.1 ± 0.8	3.4 ± 1.4
15537.79		Absorption	15532.2 – 15541.2	0.3082	0.5685	A_K	0.1 ± 1.5	3.1 ± 2.7	0.8 ± 1.4	2.3 ± 2.8
15549.17		Absorption	15545.1 – 15554.5	0.9757	0.9101	DIB	3.4 ± 1.0	16.6 ± 1.8	6.0 ± 1.0	7.4 ± 1.8
15560.13	Br16	Absorption	15556.9 – 15564.4	0.9910	0.8949	DIB	-2.4 ± 0.7	8.2 ± 1.2	0.3 ± 0.7	-0.2 ± 1.1
15601.89		Absorption	15599.5 – 15604.9	0.4632	0.8414	A_K	-0.6 ± 0.6	2.5 ± 1.2	1.1 ± 0.6	-0.0 ± 1.0
15615.91*		Absorption	15611.2 – 15624.3	0.9999	0.9999	DIB	10.0 ± 1.0	44.0 ± 2.2	16.1 ± 1.1	42.4 ± 1.7
15633.18		Absorption	15627.4 – 15637.7	0.8891	0.9183	A_K	3.1 ± 0.9	8.1 ± 1.6	2.4 ± 0.9	5.3 ± 1.6
15643.55		Absorption	15641.0 – 15646.4	0.8174	0.7844	DIB	0.5 ± 0.5	3.9 ± 0.8	-0.1 ± 0.5	1.4 ± 0.8
15652.63*		Absorption	15646.1 – 15662.8	0.9999	0.9999	DIB	14.2 ± 1.6	76.6 ± 2.8	10.9 ± 1.5	62.0 ± 2.5

continued . . .

. . . continued

λ_0[a] (Å)	Hydrogen Transition[b]	Feature Type	Wavelength Range[c] (Å)	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) DIB	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) A_K	Best Sort Parameter	EW[d](mÅ) in bin min DIB	EW[d](mÅ) in bin max DIB	EW[d](mÅ) in bin min A_K	EW[d](mÅ) in bin max A_K
15671.89*		Absorption	15663.7 – 15679.2	0.9999	0.9999	DIB	21.7 ± 1.2	76.0 ± 2.4	26.6 ± 1.3	59.3 ± 2.3
15706.13	Br15	Absorption	15701.8 – 15711.1	0.9953	0.9764	DIB	2.5 ± 1.2	20.7 ± 2.1	3.2 ± 1.0	14.0 ± 2.0
15722.41		Absorption	15718.9 – 15726.8	0.9921	0.9520	DIB	-1.7 ± 0.9	7.6 ± 1.5	-0.8 ± 0.9	5.7 ± 1.4
15733.93		Absorption	15728.1 – 15738.9	0.8863	0.9278	A_K	1.6 ± 1.2	10.6 ± 2.3	0.9 ± 1.4	12.3 ± 2.1
15751.55		Emission	15746.8 – 15760.3	0.7631	0.8419	A_K	-1.9 ± 1.2	-10.4 ± 2.5	-6.7 ± 1.3	-8.9 ± 2.3
15769.62		Absorption	15767.0 – 15772.5	0.8378	0.7406	DIB	1.5 ± 0.7	3.2 ± 1.2	2.3 ± 0.8	4.1 ± 1.2
15878.05		Absorption	15873.5 – 15882.2	0.9611	0.6855	DIB	1.3 ± 0.9	13.3 ± 1.6	3.5 ± 0.9	9.4 ± 1.7
15889.24		Absorption	15886.6 – 15894.3	0.6431	0.8079	A_K	2.6 ± 0.8	5.6 ± 1.5	5.6 ± 0.9	3.5 ± 1.4
15919.35		Absorption	15913.9 – 15922.6	0.8737	0.5808	DIB	-1.4 ± 0.8	5.1 ± 1.4	1.0 ± 0.7	0.5 ± 1.4
15927.93		Absorption	15925.1 – 15930.8	0.8308	0.2518	DIB	0.1 ± 0.5	3.0 ± 0.9	0.3 ± 0.5	2.0 ± 0.9
15937.83		Absorption	15931.9 – 15944.0	0.9847	0.7884	DIB	4.1 ± 1.0	16.5 ± 1.8	6.4 ± 0.9	10.3 ± 1.7
15949.95		Absorption	15944.7 – 15954.4	0.8917	0.9338	A_K	1.6 ± 0.8	8.5 ± 1.4	5.7 ± 0.8	6.4 ± 1.2
15977.29		Absorption	15974.4 – 15981.5	0.6406	0.5648	DIB	0.3 ± 0.7	4.0 ± 1.4	1.3 ± 0.7	2.5 ± 1.2
15989.22*		Absorption	15983.7 – 15993.9	0.9771	0.9274	DIB	2.8 ± 0.9	14.3 ± 1.6	7.7 ± 1.0	8.2 ± 1.4
15999.38		Absorption	15995.0 – 16004.0	0.7028	0.4841	DIB	0.3 ± 0.8	5.6 ± 1.2	3.4 ± 0.7	4.5 ± 1.3

continued . . .

. . . continued

λ_0[a] (Å)	Hydrogen Transition[b]	Feature Type	Wavelength Range[c] (Å)	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) DIB	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) A_K	Best Sort Parameter	EW[d](mÅ) in bin min DIB	EW[d](mÅ) in bin max DIB	EW[d](mÅ) in bin min A_K	EW[d](mÅ) in bin max A_K
16029.47		Absorption	16025.7 – 16043.2	0.7538	0.8039	A_K	-3.9 ± 2.1	12.0 ± 3.9	10.5 ± 2.0	2.5 ± 3.7
16046.53		Absorption	16044.3 – 16049.2	0.4056	0.7245	A_K	-0.0 ± 0.8	3.7 ± 1.8	1.2 ± 0.7	0.8 ± 1.6
16059.40		Absorption	16054.7 – 16062.5	0.9423	0.6806	DIB	1.5 ± 0.8	4.7 ± 1.6	1.1 ± 0.8	5.8 ± 1.5
16113.62	Br13	Absorption	16101.8 – 16119.4	0.8491	0.8396	DIB	3.3 ± 1.8	17.1 ± 2.9	6.3 ± 1.7	15.4 ± 2.9
16132.78		Absorption	16130.5 – 16135.5	0.5925	0.3200	DIB	0.5 ± 0.6	3.2 ± 1.1	1.0 ± 0.6	2.7 ± 1.1
16141.92		Absorption	16138.4 – 16145.7	0.8523	0.2659	DIB	2.0 ± 0.6	3.4 ± 1.2	1.5 ± 0.6	1.1 ± 1.0
16147.94		Emission	16142.1 – 16157.3	0.6719	0.2360	DIB	-4.0 ± 1.3	-12.2 ± 2.3	-5.4 ± 1.2	-8.3 ± 2.0
16157.09		Absorption	16152.9 – 16166.5	0.9850	0.9613	DIB	-2.3 ± 1.2	14.0 ± 2.1	-0.2 ± 1.2	6.8 ± 2.2
16219.72		Absorption	16216.4 – 16224.9	0.9462	0.9046	DIB	0.9 ± 0.7	8.0 ± 1.5	1.1 ± 0.7	6.5 ± 1.4
16233.39		Absorption	16224.9 – 16237.6	0.9373	0.9335	DIB	4.3 ± 1.4	18.2 ± 3.2	7.8 ± 1.3	14.9 ± 2.5
16265.94		Absorption	16262.1 – 16270.0	0.9133	0.8788	DIB	2.4 ± 0.7	7.6 ± 1.3	3.3 ± 0.7	4.7 ± 1.2
16280.11		Absorption	16277.0 – 16286.6	0.8855	0.8796	DIB	1.4 ± 0.8	8.2 ± 1.5	-0.3 ± 0.8	4.4 ± 1.4
16290.68		Absorption	16288.0 – 16294.7	0.7890	0.9301	A_K	1.3 ± 0.6	4.6 ± 1.2	1.9 ± 0.6	4.4 ± 1.1
16371.00		Absorption	16368.5 – 16375.8	0.9926	0.7010	DIB	2.4 ± 0.7	9.1 ± 1.4	1.3 ± 0.7	4.7 ± 1.3
16382.09		Absorption	16378.9 – 16385.0	0.3357	0.9270	A_K	0.6 ± 0.7	3.7 ± 1.3	0.0 ± 0.6	2.7 ± 1.2

continued . . .

. . . continued

λ_0[a] (Å)	Hydrogen Transition[b]	Feature Type	Wavelength Range[c] (Å)	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) DIB	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) A_K	Best Sort Parameter	EW[d](mÅ) in bin min DIB	EW[d](mÅ) in bin max DIB	EW[d](mÅ) in bin min A_K	EW[d](mÅ) in bin max A_K
16411.31	Br12	Absorption	16406.3 – 16414.5	0.6866	0.9621	A_K	0.0 ± 1.0	8.6 ± 2.2	-5.1 ± 1.0	-0.3 ± 1.9
16518.68		Absorption	16515.5 – 16522.3	0.9870	0.8953	DIB	-0.6 ± 0.7	8.1 ± 1.3	-0.2 ± 0.7	4.1 ± 1.1
16525.52		Absorption	16522.8 – 16528.5	0.9339	0.6967	DIB	1.0 ± 0.6	4.2 ± 1.1	0.4 ± 0.6	1.6 ± 1.0
16532.60		Absorption	16529.9 – 16536.3	0.8580	0.8020	DIB	1.5 ± 0.7	3.5 ± 1.3	2.4 ± 0.7	-2.0 ± 1.1
16536.49		Emission	16532.4 – 16550.2	0.9392	0.5342	DIB	-3.3 ± 1.6	-19.3 ± 3.1	-8.1 ± 1.7	-2.9 ± 3.2
16549.51		Absorption	16541.7 – 16554.5	0.8462	0.8512	A_K	-0.3 ± 1.5	5.7 ± 2.7	3.2 ± 1.5	8.5 ± 2.5
16573.31*		Absorption	16567.8 – 16579.5	0.9998	0.9919	DIB	5.9 ± 1.1	30.3 ± 2.1	7.1 ± 1.0	22.9 ± 1.9
16584.99		Absorption	16580.4 – 16589.8	0.9906	0.9814	DIB	8.7 ± 0.9	18.3 ± 1.8	3.5 ± 0.9	14.7 ± 1.6
16594.62		Absorption	16590.5 – 16598.1	0.9346	0.6475	DIB	2.9 ± 0.7	9.0 ± 1.5	1.3 ± 0.7	4.5 ± 1.3
16601.73		Absorption	16598.7 – 16605.2	0.8588	0.7641	DIB	0.3 ± 0.6	5.3 ± 1.2	0.1 ± 0.6	2.6 ± 1.2
16608.38		Absorption	16605.8 – 16611.4	0.3847	0.5903	A_K	1.4 ± 0.6	2.2 ± 1.4	1.6 ± 0.6	4.2 ± 1.1
16621.92		Absorption	16617.3 – 16627.2	0.9017	0.6553	DIB	-1.4 ± 0.9	7.8 ± 2.0	0.9 ± 0.9	-0.3 ± 1.8
16655.25		Absorption	16645.8 – 16661.2	0.6496	0.8407	A_K	1.2 ± 1.6	7.2 ± 3.5	-2.4 ± 1.7	11.8 ± 2.6
16711.03		Absorption	16708.3 – 16715.2	0.7723	0.9000	A_K	1.0 ± 1.0	6.9 ± 2.0	1.5 ± 1.0	1.3 ± 2.0
16718.42		Absorption	16716.1 – 16722.8	0.7223	0.3377	DIB	0.2 ± 0.8	3.9 ± 1.7	1.0 ± 0.8	1.0 ± 1.4

continued . . .

. . . continued

λ_0[a] (Å)	Hydrogen Transition[b]	Feature Type	Wavelength Range[c] (Å)	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) DIB	Pr ($\sigma_{\rm SORT} > \sigma_{\rm RANDOM}$) A_K	Best Sort Parameter	EW[d](mÅ) in bin min DIB	EW[d](mÅ) in bin max DIB	EW[d](mÅ) in bin min A_K	EW[d](mÅ) in bin max A_K
16750.09		Absorption	16739.7 – 16753.3	0.6424	0.8165	A_K	0.9 ± 1.5	12.5 ± 3.3	3.7 ± 1.5	11.0 ± 2.9
16757.04		Absorption	16753.8 – 16762.1	0.8341	0.9163	A_K	0.2 ± 1.2	9.1 ± 2.2	2.9 ± 1.0	0.8 ± 2.0
16770.47		Absorption	16764.7 – 16777.7	0.3466	0.7885	A_K	5.1 ± 1.4	10.0 ± 3.2	5.4 ± 1.5	4.8 ± 3.0
16780.44		Absorption	16778.1 – 16783.0	0.6389	0.2577	DIB	0.9 ± 0.6	2.2 ± 1.2	0.7 ± 0.6	0.4 ± 1.1
16826.17		Absorption	16823.4 – 16828.7	0.8639	0.5821	DIB	-0.5 ± 0.6	2.5 ± 1.5	0.0 ± 0.6	0.9 ± 1.2
16842.68		Emission	16835.5 – 16846.4	0.8304	0.1570	DIB	-2.4 ± 1.3	-8.6 ± 2.9	-5.7 ± 1.3	-2.1 ± 2.7
16845.94		Absorption	16843.2 – 16849.0	0.4582	0.5016	A_K	-0.8 ± 0.8	5.6 ± 1.7	2.7 ± 0.8	1.8 ± 1.5
16885.78		Absorption	16881.1 – 16888.8	0.7304	0.5061	DIB	2.1 ± 1.0	3.7 ± 2.4	0.3 ± 1.0	3.0 ± 1.8
16901.42		Absorption	16894.7 – 16905.4	0.4721	0.6623	A_K	6.6 ± 2.0	2.8 ± 5.5	1.6 ± 2.0	-0.5 ± 3.8
16916.84		Emission	16911.7 – 16921.8	0.9711	0.8659	DIB	-2.0 ± 1.3	-14.2 ± 3.4	-4.8 ± 1.3	-9.4 ± 2.9

16920.58	Absorption	16917.3 – 16924.3	0.8827	0.7339	DIB	1.8 ± 0.9	8.2 ± 2.4	2.5 ± 0.9	2.6 ± 2.0

[a] Wavelength, in the 15272 Å DIB rest-frame, of the local extremum found by smoothing the highest DIB-strength spectrum in Figure 4.15 with a 5 pixel Gaussian kernel.

[b] Hydrogen transition wavelengths that are within 2 Å of the feature's central wavelength.

[c] Region, in the 15272 Å DIB rest-frame, that we measure the equivalent width from. The wavelengths used to estimate the local continuum are taken to be the 5 pixels nearest to the Wavelength Range values.

[d] Equivalent width measured when binning by A_K or the 15272 Å DIB strength, in the minimum and maximum bins.

[*] Nearest detected DIB to the previously-known DIBs in Table 4.1.

4.6 Summary

We have created data-driven models of RC stellar spectra using $\sim 5.5 \times 10^5$ individual observations of $\sim 1.7 \times 10^5$ stars from the APOGEE dataset. The modeling uses five parameters – $T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], and age – to predict the spectra, and the resulting models agree quite well with the data (residual Gaussian mean near 0 and width of $\sim 1.16\sigma$) across the APOGEE wavelength range. This implies that these five labels are sufficient to explain the majority of information present in the spectra. Consequently, there is not a substantial amount of residual information that may be leveraged for pursuits such as chemical tagging. Though the residuals of the data-minus-model are relatively small ($\sim 3\%$ of stellar flux on average), it is very important to understand and isolate their astrophysical origin. We discover that there are many pixels where the residuals suggest that a significant number of non-stellar features are also present in the stellar spectra. We identify which of these features are likely Earth-based and which are likely Diffuse Interstellar Bands. We find 84 possible DIBs in APOGEE spectra that have less than 50% probability of appearing by chance alone, including all previously-discovered DIBs in this wavelength region.

Our key results include:

1. The residuals of our data compared to the model show correlations with heliocentric velocity, which we show is evidence that many of these residual features are not in the stellar rest-frame, such as skylines, tellurics, and DIBs. These residual features appear at the level of 3% of the stellar flux on average (Section 4.3.2, Figure 4.5; Section 4.4, Figures 4.6, 4.7, and 4.11);

2. The size of Earth-based residuals appear anti-correlated with spectral SNR. The shape of these residuals suggest that they may be removed by correcting for a wavelength offset between the sky model and raw observations (Section 4.5, Figure 4.14);

3. After removing the Earth-based residuals, we combine residual spectra in the rest-frame of the strongest DIB in the APOGEE wavelength range (λ_0 = 15272 Å). We detect 84 DIB features in absorption (including all of previously-measured DIBs) and emission that show highest correlation in strength with either K-band reddening (A_K) or EW strength of the 15272 Å DIB feature (Section 4.5, Figures 4.15 and 4.18, Tables 4.3 and 4.4).

Future work based on our results will focus on measuring the impact that unaccounted-for DIB and Earth-based features have on ASPCAP-measured abundances. It would also be worthwhile to compare the DIB wavelengths to lines produced by chemical species (i.e. more than just atomic hydrogen) that are known components of the ISM, as well as to correlate the new DIB strengths with additional tracers of ISM density. Another useful undertaking that would advance this work is detailed joint modelling of the DIBs, skylines, tellurics, and stellar spectra in APOGEE. A complete understanding of all sources of spectral features is necessary for chemical tagging experiments, and this will have the added benefit of improving our understanding of the ISM.

4.7 Appendix

4.7.1 Spectral Combinations

Our analysis, like many studies involving spectroscopy, requires the combination of multiple spectral observations. The standard approach would be to use an inverse-variance weighted combination using the fluxes and corresponding uncertainties at each pixel, but this results in a combined uncertainty that is often overly constraining and which becomes smaller for any increase in the number of observations. Instead, we argue that fitting a population-level distribution is generally the better approach when combining spectra, particularly in cases where the uncertainty on the combined measurement is important. This technique returns population-level means and variances that incorporate the uncertainties in each individual measurement as well as the dispersion in those measurements.

As an illustrative example, two flux measurements of the same star in one pixel might differ from one another by an amount larger than is described by their uncertainties; this case is displayed in Figure 4.19. Combining these measurements using inverse-variance weighting (black line) produces a mean that is, again, statistically far away from either measurement and has a high confidence (small uncertainty). If we instead fit a population-level distribution (red line), the resulting population width we measure is much larger than the inverse-variance width so that the population distribution can capture the large distance between the data[16].

For another example, consider set of N normalized flux measurements that

[16] While this is a helpful visual example, we should be cautious about performing population fits on a very small number of measurements.

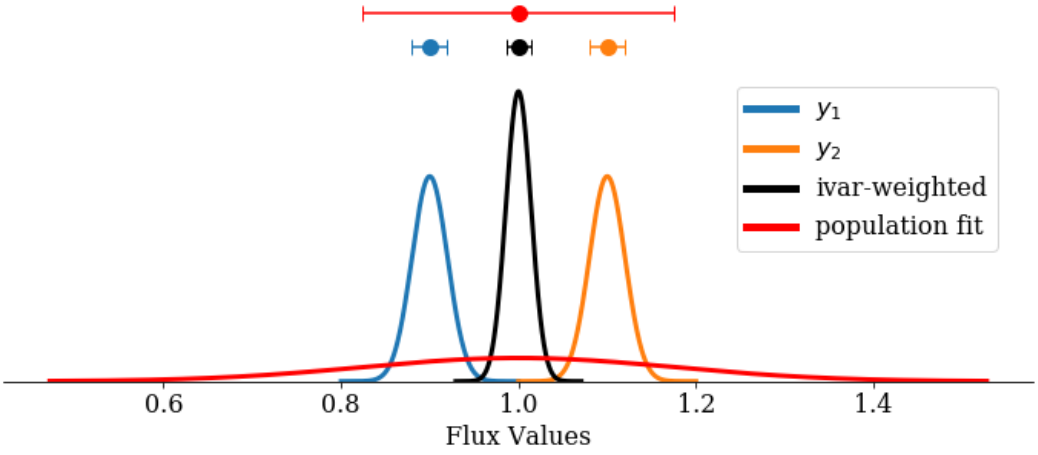

Figure 4.19: Example of the differences between the inverse-variance weighted combination of multiple data measurements and Bayesian population-level distribution fitting. The lines show normal distributions that define each set of measurements, and the points above the distributions show the mean and standard deviation of each distribution. The blue and orange lines correspond to the two data measurements while the other two lines correspond to different methods of combining those data.

all have SNR of 10 pixel^{-1}. The inverse-variance weighted combination will return an uncertainty of $\left(10/\sqrt{N}\right)$ pixel^{-1}, which will clearly decrease in size as we add more measurements, regardless of how similar or disparate those measurements may be to one another. Using population fitting instead, the distribution of the population width becomes narrower and can converge on an underlying true width, assuming one exists. This is illustrated in Figure 4.20.

For this chapter's analysis of residuals[17], we consider the measured/observed residual flux at pixel i for star j to be $r_{i,j}$ with corresponding uncertainty $\sigma_{r,i,j}$. Then, we define the following hierarchical statistical model that describes the relationship between

[17]NOTE: for this example, we are assuming that each star has one observation for clarity of the math, but the approach is easily expanded to include multiple observations of the same star.

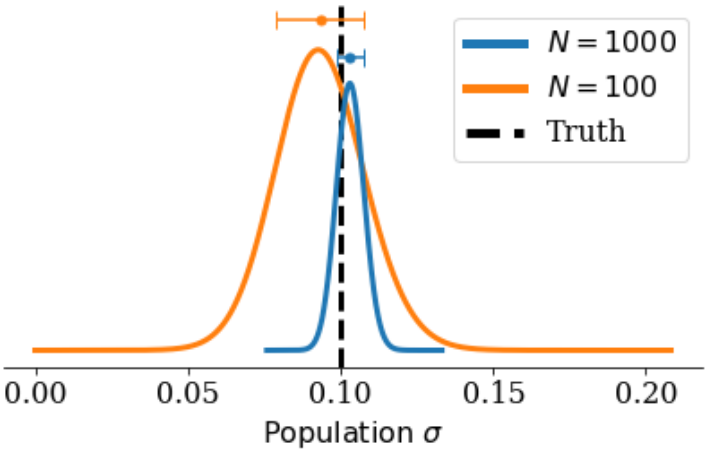

Figure 4.20: Illustration of the impact that the population size has on the distribution of the population width/uncertainty as measured from Bayesian population fitting. The vertical dashed black line shows the input true population width that was used to generate the data, and the points above the distributions show the median and 68% region for each distribution.

the individual measurements, their uncertainties, and the population distribution:

$$p\left(\hat{\sigma}_{r,i}\right) \propto 1$$

$$p\left(\hat{r}_i|\hat{\sigma}_{r,i}\right) \propto 1$$

$$\left(r'_{i,j}|\hat{r}_i, \hat{\sigma}_{r,i}\right) \sim \mathcal{N}\left(\hat{r}_i, \hat{\sigma}^2_{r,i}\right)$$

$$\left(r_{i,j}|r'_{i,j}, \sigma_{r,i,j}\right) \sim \mathcal{N}\left(r'_{i,j}, \sigma^2_{r,i,j}\right)$$

where $\hat{r}_i$ is the population mean of the residual fluxes in pixel i, $\hat{\sigma}_{r,i}$ is the population width/uncertainty/standard deviation of the residuals in pixel i, and $r'_{i,j}$ is the true residual flux for star j in pixel i. We have chosen flat priors on the population parameters, though these could be changed to other distributions if there was good reason for

it (e.g. Gaussian for $p\left(\hat{r}_i|\hat{\sigma}_{r,i}\right)$). The full posterior distribution is then:

$$\begin{aligned} &p(\hat{r}_i,\hat{\sigma}^2_{r,i},r'_{i,1},\ldots,r'_{i,n_*}|r_{i,1},\sigma^2_{i,1},\ldots,r_{i,n_*},\sigma^2_{i,n_*}) \\ &\quad\propto p\left(\hat{r}_i,\sigma_{r,i}\right)\cdot\prod_j^{n_*}\mathcal{N}\left(r'_{i,j}|\hat{r}_i,\hat{\sigma}^2_{r,i}\right)\cdot\mathcal{N}\left(r_{i,j}|r'_{i,j},\sigma^2_{r,i,j}\right) \end{aligned} \tag{4.12}$$

We next see that the posterior full conditional on $r'_{i,j}$ is given by:

$$\begin{aligned} p(r'_{i,j}|\hat{r}_i,\ldots) &\propto \mathcal{N}\left(r'_{i,j}|\hat{r}_i,\hat{\sigma}^2_{r,i}\right)\cdot\mathcal{N}\left(r'_{i,j}|r_{i,j},\sigma^2_{r,i,j}\right) \\ &= \mathcal{N}\left(r'_{i,j}|\mu_{r',i,j},\sigma^2_{r',i,j}\right) \end{aligned} \tag{4.13}$$

where

$$\sigma^2_{r',i,j} = \left[\hat{\sigma}^{-2}_{r,i}+\sigma^{-2}_{r,i,j}\right]^{-1}$$

and

$$\mu_{r',i,j} = \sigma^2_{r',i,j}\cdot\left[\hat{\sigma}^{-2}_{r,i}\cdot\hat{r}_i+\sigma^{-2}_{r,i,j}\cdot r_{i,j}\right].$$

We can then use these results to integrate over $r'_{i,j}$ in the full posterior distribution of Equation 4.12 to find the marginal posterior of $(\hat{r}_i|\hat{\sigma}_{r,i},\text{data})$:

$$p\left(\hat{r}_i|\hat{\sigma}_{r,i},\text{data}\right) = \mathcal{N}\left(\mu_{\hat{r},i},\sigma^2_{\hat{r},i}\right) \tag{4.14}$$

where

$$\sigma^2_{\hat{r},i} = \left[\sum_j^{n_*}\left(\hat{\sigma}^2_{r,i}+\sigma^2_{r,i,j}\right)^{-1}\right]^{-1}$$

and

$$\mu_{\hat{r},i} = \sigma^2_{\hat{r},i} \cdot \left[\sum_j^{n_*} \left(\hat{\sigma}^2_{r,i} + \sigma^2_{r,i,j} \right)^{-1} \cdot r_{i,j} \right].$$

Using Bayes' Law, we can find the marginal posterior for $p\left(\hat{\sigma}_{r,i}|\text{data}\right)$ as:

$$p(\hat{\sigma}_{r,i}|\text{data}) \propto p(\hat{\sigma}_{r,i}) \cdot \hat{\sigma}_{r,i}^{1/2} \cdot \prod_j^{n_*} \left(\hat{\sigma}^2_{r,i} + \sigma^2_{r,i,j} \right)^{-1/2} \cdot \exp\left(-\frac{\left(r_{i,j} - \hat{r}_i\right)^2}{2\left(\hat{\sigma}^2_{r,i} + \sigma^2_{r,i,j}\right)^2} \right) \quad (4.15)$$

With the functional forms of the distributions in Equation 4.14 and 4.15 in hand, we are able to draw samples of $(\hat{r}_i, \hat{\sigma}_{r,i}|\text{data})$ fairly quickly. First, we evaluate $p(\hat{\sigma}_{r,i}|\text{data})$ for a reasonable range and number of $\hat{\sigma}_{r,i}$ values, then we use those probabilities to draw $\hat{\sigma}_{r,i}$ samples. Next, we use those $\hat{\sigma}_{r,i}$ samples to draw samples from the $(\hat{r}_i|\hat{\sigma}_{\hat{r},i}, \text{data})$ Gaussian distribution, which is an relatively easy and efficient step. Once we've repeated this process a sufficient number of times, we can take the median of the $(\hat{r}_i, \hat{\sigma}_{r,i}|\text{data})$ samples as our best estimate of the underlying population mean and width. Sometimes, like in the main text of this work, we are also interested in the distribution on $(\hat{r}_i, \hat{\sigma}_{r,i}|\text{data})$ itself, and this technique allows us to study this distribution.

Chapter 5

Summary and Future Directions

The content of this thesis has focused on making precise and statistically-rigorous astrophysical measurements using resolved stellar populations in the Milky Way. In Chapter 2, we show that the 7D chemodynamical distributions along four different MW halo LOS are statistically different using the HALO7D survey. In Chapter 3, we build a hierarchical Bayesian pipeline that is able to measure precise positions, parallaxes, and proper motions of stars, even in sparse fields, by leveraging the information from both *HST* and *Gaia*. In Chapter 4, we build a model of APOGEE spectra that does a good job of describing the data, and in doing so find that a significant amount of the residual features in the spectra can be explained by previously-unknown signatures from the ISM.

More specifically, Chapter 2 presents [Fe/H] and [α/Fe] abundances for 113 main sequence turn-off MW stellar halo stars across four fields in the HALO7D survey. Combining these abundances with HALO7D catalog 3D velocities and distance

constraints, we measure net halo rotation and anisotropy in each LOS as a function of chemical abundance. Recent work has discovered the Gaia-Sausage-Enceladus (GSE; e.g. Helmi et al. 2018; Belokurov et al. 2018; Haywood et al. 2018), which contributes up $\sim 70\%$ of the halo stars at a Galactocentric radius of 20 kpc (Naidu et al. 2020) and is in line with theoretical predictions that the inner regions of the Milky Way stellar halo should be dominated by a small number of massive progenitors (e.g. Robertson et al. 2005). Our results agree with GSE debris dominating the inner halo, but variations we measure in the average kinematics in different [Fe/H] bins for the HALO7D fields suggest that the fractional contribution of various MW progenitors is not the same in each LOS. This supports a growing body of evidence (e.g. Iorio & Belokurov 2021) that finds the inner regions of the MW stellar halo are not uniformly mixed.

In Chapter 3, we combine *HST* and *Gaia* data to measure precise stellar motion, even in sparse fields (e.g. $N_* < 10$ stars per *HST* image). We present the general statistics that allow 2D positions of sources from any two or more images to be combined in a consistent, Bayesian way. This technique simultaneously measures the transformation parameters that allow images to map onto one another as well as finding the posterior distributions on positions, parallaxes, and proper motion for each source. To apply these techniques, we build a `python`-based program, `BP3M`, that combines archival *HST* images with *Gaia* data. This pipeline is able to measure PMs that are a $\sim$ 13 times more precise for faint sources ($20.5 < G < 21$ mag) than *Gaia* alone when using *HST* images of nearby dwarf spheroidals with time baselines of $\sim 10 - 13$ years. `BP3M` also measures improved PMs for $\sim$ 2000 sources in the sparse COSMOS field, 25% of

which are too faint for *Gaia* to measure PMs; the median PM precision improvement for faint sources ($20.25 < G < 20.75$ mag) is a factor of 2.6, and the median PM uncertainty for sources without *Gaia* PMs is 1.12 mas/yr. By analysing ~ 1000 *HST* images, we show that the on- and off-axis skew terms of the standard 6 parameter transformation equation are likely described as a function of *HST* rotation and observation time, thereby reducing the number of fitting parameters per image by a factor of 2/3. In the final section of this chapter, we use simulated *Roman* observations to find a survey strategy that is able to significantly improve parallax precision at no cost to the PM precision.

Finally, in Chapter 4 we identify the sources of key residuals in APOGEE stellar spectra whose currently-unmodelled presence contain important information for ISM science as well as chemical tagging experiments. We use a clean sample of $\sim 50,000$ Red Clump stellar spectra from the APOGEE survey to build a polynomial model of stellar labels. Studying the residuals of this model compared to the data, we notice that many of the resulting features can be attributed to DIBs from the ISM and incompletely-removed sky lines or tellurics. Through careful removal of Earth-based residuals, we isolate and characterize 84 possible DIBs, which amounts to an ~ 8 times increase in the number of detected DIBs in the APOGEE wavelength region. A handful of these newly-detected DIBs are shown to impact the wavelengths that APOGEE uses to measure abundances, potentially explaining a larger-than-expected scatter in some APOGEE abundance distributions.

Expanding upon the work presented in this thesis will continue to improve our understanding of stellar physics, the ISM, and the formation and evolution of the MW.

Some possible next steps, including some that are currently underway, are:

- Apply the `BP3M` pipeline to a large set of *HST* images that are well-sampled in time and rotation angle across all *HST* filters (e.g. using the 47 Tucanae test bed). This would help pin down which of the parameters in the transformation solution are able to be defined by functional form instead of a free parameter. This would significantly simplify the problem of mapping one image onto another, which would have a huge benefit to sparse fields where there are only a handful of sources from which to learn the transformation.

- Improved simulation of future telescope missions for use with `BP3M` to design best-practice strategies that enable all measurements from different surveys to maximally benefit one another. In specific cases – for example, in measuring proper motions in M31's disk (Dey et al. 2023) – detailed planning with a `BP3M`-like technique can find an optimal approach that meets or exceeds the science goals.

- Build a model of the APOGEE DIBs that can be used to clean up the stellar spectra, then repeat the ASPCAP abundance measurements to identify the impact of non-modelled DIBs. While this isn't likely to reach chemical tagging limits of precision, the cleaned spectra and stellar labels may be a step in creating models whose residuals are able to constrain the intrinsic spread in abundances for co-natal stars.

- Previously, the lack of multi-epoch *HST* regions in the MW halo restricted a HALO7D-like survey to a small number of LOS. With `BP3M` functioning in sparse fields, any archival *HST* images with large enough time baselines in the MW

halo are accessible, making a HALO7D Extension survey feasible. In current work using simulations of MW-like stellar halos (Apfel et al. inc. McKinnon, in prep.), observed chemodynamical distributions in a handful of LOS are enough to distinguish which simulated halo, and therefore merger history, produced those data.

As the field continues down the path of Big Data Astronomy, we must remember that our choices in analysis techniques plays a vital role in the amount and quality of information we can learn.

Bibliography

Amarante, J. A. S., Smith, M. C., & Boeche, C. 2020, MNRAS, 492, 3816

Anderson, J. 2007, Variation of the Distortion Solution, Instrument Science Report ACS 2007-08, 12 pages

—. 2022, One-Pass HST Photometry with hst1pass, Instrument Science Report WFC3 2022-5, 55 pages

Anderson, J., & King, I. R. 2004, Multi-filter PSFs and Distortion Corrections for the HRC, Instrument Science Report ACS 2004-15, 51 pages

—. 2006, PSFs, Photometry, and Astronomy for the ACS/WFC, Instrument Science Report ACS 2006-01, 34 pages

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, apj, 935, 167

Barro, G., Pérez-González, P. G., Gallego, J., et al. 2011, ApJS, 193, 13

Barro, G., Pérez-González, P. G., Cava, A., et al. 2019, ApJS, 243, 22

Bellini, A., Anderson, J., & Bedin, L. R. 2011, PASP, 123, 622

Bellini, A., Anderson, J., Bedin, L. R., et al. 2017, ApJ, 842, 6

Bellini, A., & Bedin, L. R. 2009, PASP, 121, 1419

Bellini, A., Libralato, M., Bedin, L. R., et al. 2018, ApJ, 853, 86

Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, MNRAS, 478, 611

Belokurov, V., Sanders, J. L., Fattahi, A., et al. 2020, MNRAS, 494, 3880

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

Bergemann, M., Ruchti, G. R., Serenelli, A., et al. 2014, A&A, 565, A89

Bird, S. A., Xue, X.-X., Liu, C., et al. 2019, AJ, 157, 104

—. 2021, ApJ, 919, 66

Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28

Bonaca, A., Conroy, C., Wetzel, A., Hopkins, P. F., & Kereš, D. 2017, ApJ, 845, 101

Bonaca, A., Conroy, C., Cargile, P. A., et al. 2020, ApJ, 897, L18

Bovy, J., Nidever, D. L., Rix, H.-W., et al. 2014, ApJ, 790, 127

Brand, J., & Blitz, L. 1993, A&A, 275, 67

Buder, S., Lind, K., Ness, M. K., et al. 2019, A&A, 624, A19

—. 2022, MNRAS, 510, 2407

Bullock, J. S., & Johnston, K. V. 2005, ApJ, 635, 931

Cargile, P. A., Conroy, C., Johnson, B. D., et al. 2020, ApJ, 900, 28

Carollo, D., Beers, T. C., Lee, Y. S., et al. 2007, Nature, 450, 1020

Carollo, D., Beers, T. C., Chiba, M., et al. 2010, ApJ, 712, 692

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

Conroy, C., Naidu, R. P., Zaritsky, D., et al. 2019a, ApJ, 887, 237

Conroy, C., Bonaca, A., Cargile, P., et al. 2019b, ApJ, 883, 107

Cooper, M. C., Newman, J. A., Davis, M., Finkbeiner, D. P., & Gerke, B. F. 2012, spec2d: DEEP2 DEIMOS Spectral Pipeline, Astrophysics Source Code Library, record ascl:1203.003, ascl:1203.003

Cox, N. L. J., Cami, J., Kaper, L., et al. 2014, A&A, 569, A117

Cunningham, E. C., Deason, A. J., Rockosi, C. M., et al. 2019a, ApJ, 876, 124

Cunningham, E. C., Hunt, J. A. S., Price-Whelan, A. M., et al. 2023, arXiv e-prints, arXiv:2307.08730

Cunningham, E. C., Deason, A. J., Guhathakurta, P., et al. 2016, ApJ, 820, 18

Cunningham, E. C., Deason, A. J., Sanderson, R. E., et al. 2019b, ApJ, 879, 120

Cunningham, E. C., Sanderson, R. E., Johnston, K. V., et al. 2022, ApJ, 934, 172

Das, P., & Binney, J. 2016, MNRAS, 460, 1725

De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, MNRAS, 449, 2604

Deason, A. J., Belokurov, V., & Evans, N. W. 2011, MNRAS, 416, 2903

Deason, A. J., Belokurov, V., & Koposov, S. E. 2018, ApJ, 852, 118

del Pino, A., Libralato, M., van der Marel, R. P., et al. 2022, ApJ, 933, 76

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016a, arXiv e-prints, arXiv:1611.00036

—. 2016b, arXiv e-prints, arXiv:1611.00037

Dey, A., Najita, J., Filion, C., et al. 2023, arXiv e-prints, arXiv:2306.12302

Di Matteo, P., Haywood, M., Lehnert, M. D., et al. 2019, A&A, 632, A4

Donlon, Thomas, I., Newberg, H. J., Kim, B., & Lépine, S. 2022, ApJ, 932, L16

Donlon, T., & Newberg, H. J. 2023, ApJ, 944, 169

Dotter, A. 2016, ApJS, 222, 8

Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, A&A, 275, 101

Eggen, O. J., Lynden-Bell, D., & Sandage, A. R. 1962, ApJ, 136, 748

Eilers, A.-C., Hogg, D. W., Rix, H.-W., et al. 2022, ApJ, 928, 23

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72

Elyajouri, M., Lallement, R., Monreal-Ibero, A., Capitanio, L., & Cox, N. L. J. 2017, A&A, 600, A129

Elyajouri, M., Monreal-Ibero, A., Remy, Q., & Lallement, R. 2016, ApJS, 225, 19

Escala, I., Kirby, E. N., Gilbert, K. M., Cunningham, E. C., & Wojno, J. 2019, ApJ, 878, 42

Feeney, S. M., Wandelt, B. D., & Ness, M. K. 2021, MNRAS, 501, 3258

Feltzing, S., & Gustafsson, B. 1998, A&AS, 129, 237

Flesch, E. W. 2023, arXiv e-prints, arXiv:2308.01505

Font, A. S., Johnston, K. V., Bullock, J. S., & Robertson, B. E. 2006a, ApJ, 638, 585

—. 2006b, ApJ, 646, 886

Font, A. S., Johnston, K. V., Guhathakurta, P., Majewski, S. R., & Rich, R. M. 2006c, AJ, 131, 1436

Foreman-Mackey, D. 2016, The Journal of Open Source Software, 1, 24

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

Gaia Collaboration, Helmi, A., van Leeuwen, F., et al. 2018, A&A, 616, A12

García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, AJ, 151, 144

Geballe, T. R., Najarro, F., Figer, D. F., Schlegelmilch, B. W., & de La Fuente, D. 2011, Nature, 479, 200

Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25

Green, G. 2018, The Journal of Open Source Software, 3, 695

Griffith, E. J., Weinberg, D. H., Buder, S., et al. 2022, ApJ, 931, 23

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35

Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357

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

Hasselquist, S., Hayes, C. R., Lian, J., et al. 2021, ApJ, 923, 172

Hayden, M. R., Bovy, J., Holtzman, J. A., et al. 2015, ApJ, 808, 132

Haywood, M., Di Matteo, P., Lehnert, M. D., et al. 2018, ApJ, 863, 113

Helmi, A. 2008, A&A Rev., 15, 145

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

Hoffmann, S. L., & Kozhurina-Platais, V. 2020, Validation of New ACS/WFC Geometric Distortion Reference Files Derived Using Gaia Data Release 2, Instrument Science Report ACS 2020-9, 17 pages

Holtzman, J. A., Shetrone, M., Johnson, J. A., et al. 2015, AJ, 150, 148

Holtzman, J. A., Hasselquist, S., Shetrone, M., et al. 2018, AJ, 156, 125

Hopkins, P. F. 2015, MNRAS, 450, 53

Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, MNRAS, 480, 800

Horta, D., Cunningham, E. C., Sanderson, R. E., et al. 2023, ApJ, 943, 158

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

Iorio, G., & Belokurov, V. 2021, MNRAS, 502, 5686

Joblin, C., Maillard, J. P., D'Hendecourt, L., & Léger, A. 1990, Nature, 346, 729

Johnston, K. V., Bullock, J. S., Sharma, S., et al. 2008, ApJ, 689, 936

Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, AJ, 160, 120

Kalirai, J. S. 2012, Nature, 486, 90

Khanna, S., Sharma, S., Bland-Hawthorn, J., et al. 2019, MNRAS, 482, 4215

Kirby, E. N. 2011, PASP, 123, 531

Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, ApJ, 779, 102

Kirby, E. N., Gilbert, K. M., Escala, I., et al. 2020, AJ, 159, 46

Kirby, E. N., Guhathakurta, P., & Sneden, C. 2008, ApJ, 682, 1217

Kirby, E. N., Rizzi, L., Held, E. V., et al. 2017, ApJ, 834, 9

Kirby, E. N., Guhathakurta, P., Simon, J. D., et al. 2010, ApJS, 191, 352

Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmidt, IOS Press, 87 – 90

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36

Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv e-prints, arXiv:1711.03234

Koppelman, H. H., Helmi, A., Massari, D., Price-Whelan, A. M., & Starkenburg, T. K. 2019, A&A, 631, L9

Kroupa, P. 2001, MNRAS, 322, 231

Kurucz, R. L. 1993, Physica Scripta Volume T, 47, 110

Lancaster, L., Koposov, S. E., Belokurov, V., Evans, N. W., & Deason, A. J. 2019, MNRAS, 486, 378

Lee, D. M., Johnston, K. V., Sen, B., & Jessop, W. 2015, ApJ, 802, 48

Leethochawalit, N., Kirby, E. N., Ellis, R. S., Moran, S. M., & Treu, T. 2019, ApJ, 885, 100

Leethochawalit, N., Kirby, E. N., Moran, S. M., Ellis, R. S., & Treu, T. 2018, ApJ, 856, 15

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

Libralato, M., Bellini, A., Piotto, G., et al. 2019, ApJ, 873, 109

Libralato, M., Bellini, A., van der Marel, R. P., et al. 2018, ApJ, 861, 99

Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2

Liu, G., Huang, Y., Bird, S. A., et al. 2022, MNRAS, 517, 2787

Loebman, S. R., Valluri, M., Hattori, K., et al. 2018, ApJ, 853, 196

Luri, X., Brown, A. G. A., Sarro, L. M., et al. 2018, A&A, 616, A9

Mackereth, J. T., Schiavon, R. P., Pfeffer, J., et al. 2019, MNRAS, 482, 3426

Majewski, S. R., APOGEE Team, & APOGEE-2 Team. 2016, Astronomische Nachrichten, 337, 863

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

Maoz, D., & Badenes, C. 2010, MNRAS, 407, 1314

Martell, S. L., Sharma, S., Buder, S., et al. 2017, MNRAS, 465, 3203

Massari, D., Breddels, M. A., Helmi, A., et al. 2018, Nature Astronomy, 2, 156

Massari, D., Helmi, A., Mucciarelli, A., et al. 2020, A&A, 633, A36

Massari, D., Posti, L., Helmi, A., Fiorentino, G., & Tolstoy, E. 2017, A&A, 598, L9

McKinnon, K. A., Cunningham, E. C., Rockosi, C. M., et al. 2023, arXiv e-prints, arXiv:2302.07293

Mitschang, A. W., De Silva, G., Zucker, D. B., et al. 2014, MNRAS, 438, 2753

Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJS, 206, 8

Myeong, G. C., Vasiliev, E., Iorio, G., Evans, N. W., & Belokurov, V. 2019, MNRAS, 488, 1235

Naidu, R. P., Conroy, C., Bonaca, A., et al. 2020, ApJ, 901, 48

—. 2022, arXiv e-prints, arXiv:2204.09057

Nayyeri, H., Hemmati, S., Mobasher, B., et al. 2017, ApJS, 228, 7

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

Ness, M., Hogg, D. W., Rix, H. W., et al. 2016, ApJ, 823, 114

Ness, M. K., Johnston, K. V., Blancato, K., et al. 2019, ApJ, 883, 177

Ness, M. K., Wheeler, A. J., McKinnon, K., et al. 2022, ApJ, 926, 144

Nidever, D. L., Bovy, J., Bird, J. C., et al. 2014, ApJ, 796, 38

Nidever, D. L., Holtzman, J. A., Allende Prieto, C., et al. 2015, AJ, 150, 173

Nissen, P. E., & Schuster, W. J. 2010, A&A, 511, L10

pandas development team, T. 2020, pandas-dev/pandas: Pandas

Paxton, B., Bildsten, L., Dotter, A., et al. 2011, ApJS, 192, 3

Paxton, B., Cantiello, M., Arras, P., et al. 2013, ApJS, 208, 4

Paxton, B., Marchant, P., Schwab, J., et al. 2015, ApJS, 220, 15

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning

Research, 12, 2825

Pérez, F., & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21

Price-Jones, N., & Bovy, J. 2018, MNRAS, 475, 1410

Prochaska, J. X., Naumov, S. O., Carney, B. W., McWilliam, A., & Wolfe, A. M. 2000, AJ, 120, 2513

Robertson, B., Bullock, J. S., Font, A. S., Johnston, K. V., & Hernquist, L. 2005, ApJ, 632, 872

Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523

Rockosi, C. M., Lee, Y. S., Morrison, H. L., et al. 2022, ApJS, 259, 60

Rojas-Arriagada, A., Recio-Blanco, A., Hill, V., et al. 2014, A&A, 569, A103

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schönrich, R., Asplund, M., & Casagrande, L. 2011, MNRAS, 415, 3807

Schönrich, R., Binney, J., & Dehnen, W. 2010, MNRAS, 403, 1829

Searle, L., & Zinn, R. 1978, ApJ, 225, 357

Sit, T., & Ness, M. K. 2020, ApJ, 900, 4

Smoker, J. V., Müller, A., Monreal Ibero, A., et al. 2023, A&A, 672, A181

Sneden, C. A. 1973, PhD thesis, University of Texas, Austin

Sohn, S. T., Anderson, J., & van der Marel, R. P. 2012, ApJ, 753, 7

Sohn, S. T., Watkins, L. L., Fardal, M. A., et al. 2018, ApJ, 862, 52

Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757

Stefanon, M., Yan, H., Mobasher, B., et al. 2017, ApJS, 229, 32

Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645

Stetson, P. B., Pancino, E., Zocchi, A., Sanna, N., & Monelli, M. 2019, MNRAS, 485, 3042

Tchernyshyov, K., & Peek, J. E. G. 2017, AJ, 153, 8

Tchernyshyov, K., Peek, J. E. G., & Zasowski, G. 2018, AJ, 156, 248

Ting, Y.-S., Conroy, C., & Goodman, A. 2015, ApJ, 807, 104

Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2019, ApJ, 879, 69

Tinsley, B. M. 1980, Fund. Cosmic Phys., 5, 287

van der Marel, R. P., Fardal, M., Besla, G., et al. 2012, ApJ, 753, 8

Vasiliev, E., & Baumgardt, H. 2021, MNRAS, 505, 5978

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261

Wallerstein, G. 1962, ApJS, 6, 407

Weinberg, D. H., Holtzman, J. A., Johnson, J. A., et al. 2022, ApJS, 260, 32

Wes McKinney. 2010, in Proceedings of the 9th Python in Science Conference, ed. Stéfan van der Walt & Jarrod Millman, 56 – 61

Wetzel, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, ApJ, 827, L23

Wilson, J. C., Hearty, F., Skrutskie, M. F., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84460H

Wu, W., Zhao, G., Xue, X.-X., Bird, S. A., & Yang, C. 2022, ApJ, 924, 23

Wylie, S. M., Gerhard, O. E., Ness, M. K., et al. 2021, A&A, 653, A143

Xue, X.-X., Rix, H.-W., Ma, Z., et al. 2015, ApJ, 809, 144

Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377

York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579

Yuan, Z., Chang, J., Beers, T. C., & Huang, Y. 2020, ApJ, 898, L37

Zasowski, G., Ménard, B., Bizyaev, D., et al. 2015, ApJ, 798, 35

Zasowski, G., Schultheis, M., Hasselquist, S., et al. 2019, ApJ, 870, 138

Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, Research in Astronomy and Astrophysics, 12, 723

Zolotov, A., Willman, B., Brooks, A. M., et al. 2009, ApJ, 702, 1058