Magellan/M2FS and MMT/Hectochelle Spectroscopy of Dwarf Galaxies and Faint Star Clusters within the Galactic Halo

Citation for published version:
Walker, MG, Caldwell, N, Mateo, M, Olszewski, EW, Pace, AB, John, IBIII, Koposov, SE & Roederer, IU 2023, 'Magellan/M2FS and MMT/Hectochelle Spectroscopy of Dwarf Galaxies and Faint Star Clusters within the Galactic Halo', *Astrophysical Journal Supplement*, vol. 268, no. 1, 19, pp. 1-40. https://doi.org/10.3847/1538-4365/acdd79

Digital Object Identifier (DOI):
10.3847/1538-4365/acdd79

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Astrophysical Journal Supplement

Draft version June 12, 2023
Typeset using LaTeX twocolumn style in AASTeX631

Magellan/M2FS and MMT/Hectochelle Spectroscopy of Dwarf Galaxies and Faint Star Clusters within the Galactic Halo*

Matthew G. Walker,[1] Nelson Caldwell,[2] Mario Mateo,[3] Edward W. Olszewski,[4] Andrew B. Pace,[1] John I. Bailey, III,[5] Sergey E. Koposov,[6,7,8,1] and Ian U. Roederer[3,9]

[1] McWilliams Center for Cosmology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA
[2] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS-15, Cambridge, MA 02138, USA
[3] Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA
[4] Steward Observatory, The University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA
[5] Department of Physics, University of California Santa Barbara, Santa Barbara, CA 93016, USA
[6] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[7] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[8] Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[9] Joint Institute for Nuclear Astrophysics—Center for the Evolution of the Elements (JINA-CEE), USA

ABSTRACT

We present spectroscopic data for 16369 stellar targets within and/or toward 38 dwarf spheroidal galaxies and faint star clusters within the Milky Way halo environment. All spectra come from observations with the multi-object, fiber-fed echelle spectrographs M2FS at the Magellan/Clay telescope or Hectochelle at the MMT, reaching a typical limiting magnitude $G \lesssim 21$. Data products include processed spectra from all observations and catalogs listing estimates—derived from template model fitting—of line-of-sight velocity (median uncertainty 1.1 km s^{-1}) effective temperature (234 K), (base-10 logarithm of) surface gravity (0.52 dex in cgs units), [Fe/H] (0.38 dex) and [Mg/Fe] (0.24 dex) abundance ratios. The sample contains multi-epoch measurements for 3720 sources, with up to 15 epochs per source, enabling studies of intrinsic spectroscopic variability. The sample contains 6078 likely red giant stars (based on surface gravity), and 4494 likely members (based on line-of-sight velocity and *Gaia*-measured proper motion) of the target systems. The number of member stars per individual target system ranges from a few, for the faintest systems, to ~ 850 for the most luminous. For most systems, our new samples extend over wider fields than have previously been observed; of the likely members in our samples, 823 lie beyond $2\times$ the projected halflight radius of their host system, and 42 lie beyond $5R_{\rm half}$.

1. INTRODUCTION

The Galactic halo teems with stellar substructure. This local environment provides our clearest window onto the processes of galaxy formation and the nature of dark matter. The hierarchy of surviving Halo substructures stretches from the smallest scales, where diffuse star clusters overlap in luminosity with the faintest, most primitive dwarf galaxies (Gilmore et al. 2007; Martin et al. 2008), to the readily-visible and star-forming Magellanic Clouds. All of these objects are in various stages of dissolution within the Galactic Halo, where ghosts of their earlier-infalling cousins remain detectable by their stellar-orbital configurations and chemical composition (e.g., Belokurov et al. 2018; Helmi et al. 2018; Naidu et al. 2020).

Known Halo substructures exhibit a wide range of properties that reveal details of their own formation, internal structure, and chemical evolution (Helmi 2020). The abundance and systemic motions of Halo substructures can be used to trace and characterize the Galaxy's extended dark matter halo. The internal kinematics of individual substructures—dwarf galaxies, stellar streams and stellar overdensities—trace dark matter on the smallest scales where it is known to exist (Aaronson 1983; Mateo et al. 1993; Willman et al. 2011). The

Corresponding author: Matthew G. Walker
mgwalker@cmu.edu

* This paper presents data gathered with the Magellan Telescopes at Las Campanas Observatory, Chile, and the MMT Observatory, a joint facility of the Smithsonian Institution, and the University of Arizona.

chemical abundance patterns of constituent stars reflect the processes at work in the earliest stages of cosmic star formation (Tolstoy et al. 2009; Weisz & Boylan-Kolchin 2017).

Over the past several decades, spectroscopic studies of individual stars within the Milky Way's surviving satellites have developed in fits and starts. Early campaigns used 2-4m class telescopes to target red giant candidates in the Milky Way's ~ 10 'classical' dwarf spheroidal companions, building line-of-sight velocity samples for a few to several tens of member stars per system (e.g., Aaronson 1983; Olszewski & Aaronson 1985; Mateo et al. 1991, 1993; Hargreaves et al. 1994b,a, 1996; Olszewski et al. 1995). With the advent of multi-object fiber spectrographs, samples grew to ~ 100 members per system (e.g., Kleyna et al. 2002; Wilkinson et al. 2004). Ultimately, multi-object spectrographs at 6-10m class telescopes enabled samples not only of line-of-sight velocity, but also chemical composition for several hundreds to a few thousand members per system (e.g., Tolstoy et al. 2004; Koch et al. 2006; Battaglia et al. 2006; Koch et al. 2007b,a; Walker, Mateo & Olszewski 2009; Kirby et al. 2010).; for a few bright confirmed member stars, higher-resolution followup could then measure detailed abundance patterns (e.g., Shetrone et al. 2001; Letarte et al. 2009; Aoki et al. 2009; Cohen & Huang 2009; Lucchesi et al. 2020) Meanwhile, the same instrumentation provided samples reaching a few to tens of members per each of the low-luminosity ($M_V \gtrsim -6$), 'ultra-faint' Milky Way satellites that were revealed by, e.g., the Sloan Digital Sky Survey, Pan-STARRs and the Dark Energy Survey (e.g., Kleyna et al. 2005; Muñoz et al. 2006; Martin et al. 2007; Simon & Geha 2007; Koposov et al. 2011).

These observational datasets have delivered a wealth of information about the systemic motions and internal chemo-dynamical properties of the Milky Way satellite population; for review articles, see Mateo (1998); Tolstoy et al. (2009); McConnachie (2012); Simon (2019); Battaglia & Nipoti (2022); Belokurov & Evans (2022). However, the available datasets leave room for substantial improvement. First there is the obvious statistical improvement that would come with even larger samples and higher (spectroscopic) resolution. There is also the systematic improvement that would come with expanded spatial and temporal sampling. Due to finite field sizes, the oldest, lowest-metallicity, most weakly bound outermost member stars are under-represented in nearly all existing spectroscopic samples of dwarf galaxies. While recent observational campaigns are beginning to focus on outer regions (e.g. Waller et al. 2023; Sestito et al. 2023; Tolstoy et al. 2023), most measurements of stellar velocity and metallicity distributions, as well as formation histories, remain biased toward central values where stellar populations skew younger, kinematically colder and more chemically evolved (Tolstoy et al. 2009). Furthermore, the lack of multi-epoch observations for most stars precludes knowledge of intrinsic variability, limiting the accuracy with which, e.g., intrinsic velocity distributions (and hence dynamical masses) can be inferred (e.g., McConnachie et al. 2009).

With the goal of overcoming these and other limitations, we are using wide-field, high-resolution, multi-object spectrographs at the 6.5m MMT and Magellan telescopes in the northern and southern hemispheres, respectively, to conduct a spectroscopic campaign that targets the known dwarf galaxies and faint star clusters within the Galactic halo. Compared to previous efforts, our current observations provide higher spectroscopic resolution, wider spatial coverage and/or multi-epoch temporal coverage. Here we describe the observations, data processing and quality, and release processed spectra and data catalogs from our ongoing programs.

2. OBSERVATIONS

We present results from spectroscopic observations of 38 **dwarf galaxies and star clusters within the Galactic Halo**, conducted over portions of more than 200 clear nights during the years 2005 – 2022. All observations use multi-object, fiber-fed echelle spectrographs at one of two telescopes. We observe northern targets using the Hectochelle spectrograph (Szentgyorgyi 2006) at the 6.5m MMT Observatory in Arizona, and southern targets using the M2FS spectrograph (Mateo et al. 2012) at the 6.5m Magellan/Clay telescope at Las Campanas Observatory in Chile.

2.1. *Target Selection*

The quantity and quality of imaging data available for selecting spectroscopic targets have evolved dramatically over the course of our observations. Our earliest spectroscopic targets were chosen based on our own two-filter photometry, which was limited to relatively central regions of the most luminous dwarf galaxies (e.g., Mateo et al. 2008). Others use more recent data sets from observational campaigns—e.g., the PRISTINE survey (Starkenburg et al. 2017)—that target individual systems. In one case—M2FS observations of the Reticulum II dwarf galaxy—we received targeting coordinates directly from the Dark Energy Survey's Milky Way working group, which had just discovered Reticulum II based on their then-proprietary photometric catalogs (The DES Collaboration et al. 2015). Most recently, we select targets based on public data from large sky surveys—e.g., SDSS (Ahn et al. 2012), PanSTARRs (Flewelling et al. 2020), DES (Abbott et al. 2021)—that provide multi-color photometry and, with the *Gaia* mission, precise and time-dependent astrometry over wide fields (Gaia Collaboration et al. 2016, 2022). In the special case of recent observations of star clusters at low Galactic latitude, we select targets based entirely on photometry and astrometry from *Gaia* (Pace et al. 2023).

One consequence of this progress is that our spectroscopic targeting criteria are heterogeneous, varying not only from system to system, but also across different fields and/or different epochs within a given system. Thus we cannot provide a rigorous and consistent selection function that accounts for the sampling that produced the spectroscopic data sets presented herein. Instead, here we describe our general approach to selecting spectroscopic targets, and how that approach has evolved in response to advances in imaging surveys. In any case, our data products include coordinates of all observed spectroscopic targets regardless of data quality, allowing users to infer effective selection functions where necessary.

For nearly all of the stellar systems studied here, the member stars that are sufficiently bright for spectroscopy (magnitude $G \lesssim 21$) are post-main-sequence stars on the red giant, subgiant and horizontal branches. At distances ranging from tens to hundreds of kpc, stars at these evolutionary stages have broad-band colors and magnitudes that are similar to those of late-type dwarf stars in the Galactic foreground. Our general strategy for target selection is first to use available photometry to identify these sequences of evolved stars along the line of sight to the system of interest, then to use additional information (e.g., parallax and proper motion), where available, to filter out likely foreground contaminants.

More specifically, since proper motion data became available with *Gaia's* second data release (Gaia Collaboration et al. 2018c), we select spectroscopic targets according to the following procedure. First, we use wide-field survey photometry (e.g., SDSS, DES, PanSTARRS, etc.) to identify red giant, horizontal and subgiant branch candidates as likely point sources (based on survey-specific criteria, e.g., requiring TYPE=6 for SDSS photometry, |wavg_spread_model_r| < 0.003 for DES data) having g-band magnitudes and $g-r$ colors within δ magnitudes of a best-fitting (by eye) theoretical isochrone (Dotter 2016). The tolerance $\delta = \sqrt{\delta_{\rm err}^2+\delta_{\rm min}^2}$ is set by the observational error, $\delta_{\rm err}$, associated with the photometric color, and a minimum tolerance that takes a typical value of $\delta_{\rm min} = 0.2$ mag. Next we identify the photometrically-selected stars for which *Gaia* measures a parallax that is unresolved (parallax angle is smaller than 3 times its observational error), and a proper motion that is consistent, given observational errors, with the systemic mean (e.g., Gaia Collaboration et al. 2018a; Pace & Li 2019). Given the list of prospective spectroscopic targets that pass these photometric and astrometric filters, we select randomly from those that lie within the available field of view of a given telescope pointing. For systems that extend beyond a single telescope pointing, our choice of pointing is based on competing interests in 1) observing large numbers of high-probability member stars, which favors central fields, 2) fairly sampling across the target system, which requires outer fields where member stars can be scarce, and 3) obtaining sufficient repeat measurements to gauge observational errors and intrinsic variability. Finally, we note that photometric and/or astrometric filter tolerances can be adjusted based on target density in order to make use of available fibers.

Figures 22 and 23 of the Appendix display sky positions, color-magnitude diagrams (CMDs), proper motion coordinates and our own measurements of metallicity, [Fe/H], vs. (heliocentric) line-of-sight velocity, $V_{\rm LOS}$, for spectroscopic targets toward each Galactic satellite that we observe. As discussed above, our actual target selection used a variety of different photometric data sets; however, for uniformity of presentation the plotted CMDs all use *Gaia's* G-band photometry and integrated BP-RP spectra, with extinction corrections applied according to the procedure described by Gaia Collaboration et al. (2018b). Overplotted in the CMDs are theoretical isochrones (Dotter 2016; Morton 2015) computed for old (age=10 Gyr) stellar populations and published values of metallicity for each object (e.g., McConnachie 2012). Ellipses in the sky maps have semi-major axes $a = 2R_{\rm half}/\sqrt{1-\epsilon}$, where $R_{\rm half}$ is the projected halflight radius and $\epsilon \equiv 1-b/a$ is the measured ellipticity. In the proper motion and [Fe/H] vs. $V_{\rm LOS}$ panels, dashed lines indicate previously-published values for systemic mean proper motions and velocities (Pace et al. 2022), where available.

2.2. *Magellan/M2FS*

The Michigan/Magellan Fiber System (M2FS; Mateo et al. 2012) is a fiber-fed, double spectrograph operating at the f/11 Nasmyth port of the Magellan/Clay 6.5 m telescope at Las Campanas Observatory, Chile. Each of the two M2FS spectrographs receives light from up to 128 fibers. A wide field corrector provides good image quality over a field of diameter 30 arcmin. Fibers can operate at wavelengths between 3700 - 9500 Å, have entrance apertures of diameter 1.2 arcsec, and tolerate center-to-center target separations as small as 12 arcsec. M2FS observers plug fibers by hand into masks that are machined at the Carnegie Observatories machine shop. Depending on choice of diffraction grating and order-blocking filters, M2FS offers a wide range of observing configurations, with spectral resolution ranging from $\mathcal{R} \sim [0.2-34] \times 10^3$ and wavelength coverage ranging from tens to thousands of Å.

For the vast majority of M2FS observations reported here, we use the high-resolution ('HiRes' hereafter) grating with both spectrographs, and with filters selected to pass light over a single order at $5130 \lesssim \lambda \lesssim 5190$ Å. The most prominent feature in this region is the Mg I 'b' triplet, with rest wavelengths of 5167.32 Å, 5172.68 Å, and 5183.60 Å. This region also contains many iron lines that enable a direct measurement of iron abundance. With these choices, we acquire single-order spectra for up to 256 sources per pointing, with resolving power $\mathcal{R} \sim 24,000$. We bin the detector at 2×2 pixels^2, giving plate scale ~ 0.065 Å/pixel over the useful wavelength range.

For a small fraction of M2FS observations reported here, we use an alternative configuration that has at least one of the two spectrographs using a medium-resolution (henceforth 'MedRes') grating that gives resolving power $\mathcal{R} \sim 7000$. In order to cover the Mg triplet region, we use an order-blocking filter that passes light over the range 5115–5300 Å. Using the same 2×2 binning that we use with the HiRes grating, the MedRes observations have plate scale ~ 0.2 Å/pixel over the useful wavelength range.

During a typical observing night with M2FS, we take 100-200 zero-second 'exposures' in order to measure the bias levels of the detectors in both spectrographs. We take between 3-10 exposures of the (scattered) solar spectrum during evening and/or morning twilight. For a typical science field, we expose for 1-3 hours, broken into 2-5 sub-exposures. Of the 256 available fibers, we assign ~ 30 to regions of blank sky. Immediately before and after science exposures, and often between sub-exposures, we acquire calibration spectra of an LED source and then a ThArNe arc lamp, both of which are located at the secondary cage and illuminate the fibers at the focal surface. During daylight hours, we acquire sequences of hour-long 'dark' exposures with both spectrographs' shutters closed.

Table 1 lists the instrument configuration, central field coordinates, date, total exposure time and number of targets for all M2FS science fields observed for our program thus far. Including repeat observations, we have observed a total of 92 science fields with M2FS for this program—74 with both spectrographs using the HiRes grating, 1 with both using the MedRes grating, and 17 with one spectrograph in HiRes mode and the other in MedRes mode—for a total science exposure time of 0.68 megaseconds (Ms). We obtain acceptable M2FS HiRes spectroscopic measurements for ∼6.6k unique sources within 18 different target systems, and we obtain acceptable M2FS MedRes measurements for ∼82 unique sources within 5 different systems. For ∼1.4k M2FS sources we have (up to 15 per source) multiple independent measurements.

2.3. *MMT/Hectochelle*

Hectochelle is a fiber-fed echelle spectrograph at the f/5 focal surface of the MMT Observatory on Mt. Hopkins, Arizona, United States (Szentgyorgyi 2006). Hectochelle's optical fibers have entrance apertures of diameter 1.5 arcsec, and are positioned robotically, allowing simultaneous observation of up to 240 distinct sources. A wide field corrector, coupled with an atmospheric dispersion compensator, gives a field of view of diameter 1 degree. Hectochelle spectra consist of a single diffraction order spanning ~ 150 Å, with resolving power $\mathcal{R} \sim 32{,}000$ at wavelength $\lambda \sim 5200$ Å. We use Hectochelle's 'RV31' order-blocking filter, which isolates the wavelength range 5150–5300 Å. We bin the detector by factors of 2 and 3 in the spectral and spatial dimensions, respectively, giving plate scale ~ 0.10 Å/pixel.

Our observing strategy with Hectochelle is similar to the one described above for M2FS. On a typical night we acquire ~ 100 zero-second bias 'exposures', plus exposures of the scattered solar spectrum during evening and/or morning twilight. As with M2FS, for a given science field we acquire between 2-5 sub-exposures totalling 1-3 hours of integration time. Before and after science exposures we acquire spectra of a ThAr arc lamp. Either before or after science exposures, we acquire the spectrum of a quartz lamp. The observatory staff acquires dark exposures regularly during daylight hours.

Table 2 lists the same information as Table 1, but for Hectochelle observations. With Hectochelle we have observed a total of 92 (including repeat observations) science fields for a total science exposure time of 1.42 Ms. We obtain acceptable measurements for ∼9.7k unique sources within within 21 target systems. For ∼2.4k sources we have (up to 13 per source) multiple independent measurements.

3. PROCESSING OF RAW SPECTRA

Table 1: Log of M2FS Observations of Galactic Halo Objects (abbreviated—see electronic version for full table)

Instrument	Field Center		UT date[a]	UT start[b]	Exp. Time	$N_{\rm exp}$	$N_{\rm target}$	Object
	α_{2000} [deg.]	δ_{2000} [deg.]			[sec.]			
M2FS HiRes	153.028333	−001.754667	2014-02-24	08:44:35	8900	5	218	Sextans
M2FS HiRes	100.746667	−050.848333	2014-02-25	03:01:36	5400	3	214	Carina
M2FS HiRes	100.610000	−051.082361	2014-02-25	05:04:05	5700	3	214	Carina
M2FS HiRes	153.684583	−001.500944	2014-02-26	08:00:05	6600	5	216	Sextans
M2FS HiRes	153.685000	−001.501000	2014-02-27	07:13:52	3600	3	216	Sextans
M2FS HiRes	153.292917	−001.604694	2014-02-28	07:46:28	3600	3	216	Sextans
M2FS HiRes	100.399167	−050.947139	2014-12-15	06:28:27	8100	3	218	Carina
M2FS HiRes	100.835417	−051.099611	2014-12-19	08:05:29	4500	3	207	Carina
M2FS HiRes	099.959583	−050.786417	2014-12-21	07:56:08	5400	3	187	Carina
M2FS HiRes	099.961667	−050.784722	2014-12-23	08:01:05	3600	3	177	Carina
M2FS HiRes	053.810000	−054.075444	2015-02-19	02:24:12	7200	3	186	Reticulum II
M2FS HiRes	153.292500	−001.601639	2015-02-22	06:50:05	7200	4	214	Sextans
M2FS HiRes/MedRes	343.062917	−058.493583	2015-07-18	09:40:45	9000	5	137	Tucana II

[a]YYYY-MM-DD format

[b]Universal time at start of first exposure; HH:MM:SS format

Table 2: Log of MMT/Hectochelle Observations of Galactic Halo Objects (abbreviated—see electronic version for full table)

Instrument	Field Center		UT date[a]	UT start[b]	Exp. Time	$N_{\rm exp}$	$N_{\rm target}$	Object
	α_{2000} [deg.]	δ_{2000} [deg.]			[sec.]			
Hectochelle	152.064708	+012.349136	2005-04-01	05:06:30	9079	3	143	Leo I
Hectochelle	152.064708	+012.349136	2005-04-02	06:13:04	14400	4	143	Leo I
Hectochelle	259.425000	+058.049972	2005-04-02	10:52:57	8700	3	132	Draco
Hectochelle	152.166792	+012.274975	2006-04-20	05:53:41	7500	3	135	Leo I
Hectochelle	152.107875	+012.309992	2006-04-24	05:07:09	8100	3	135	Leo I
Hectochelle	168.355875	+022.149333	2006-04-25	05:12:38	8100	3	114	Leo II
Hectochelle	260.958333	+057.870000	2006-04-25	08:12:45	4846	5	107	Draco
Hectochelle	210.005667	+014.483664	2006-05-08	04:24:44	5400	3	191	Bootes I
Hectochelle	260.102667	+057.885250	2007-02-23	12:27:48	5400	3	120	Draco
Hectochelle	257.091792	+057.877306	2007-02-26	10:03:25	5400	3	139	Draco
Hectochelle	262.915167	+058.382108	2007-02-26	12:22:21	7200	4	145	Draco
Hectochelle	152.765458	−001.052389	2007-02-27	09:57:01	8400	4	203	Sextans
Hectochelle	259.407542	+057.775056	2007-02-27	12:05:59	5400	3	89	Draco

[a]YYYY-MM-DD format

[b]Universal time at start of first exposure; HH:MM:SS format

All MMT/Hectochelle spectra are processed using the standard TDC pipeline[1] which is written in IDL. Briefly, the four channels from two CCDs are corrected for bias and merged. Cosmic rays are then detected and interpolated over, and individual exposures are coadded. Dark structure is subtracted depending on the exposure time, and spectra are extracted in the manner described in the next section.

The remainder of this section describes the set of Python-based modules that we have written for end-to-end processing of Magellan/M2FS spectra. Where applicable and convenient, we incorporate modules that are publicly available as part of the Astropy software package (Astropy Collaboration et al. 2013, 2018, 2022).

3.1. *Overscan/bias/dark/gain corrections and uncertainties*

We begin by using the Astropy-affiliated package 'ccdproc' (Craig et al. 2017) to perform standard corrections for overscan, bias, dark current and gain. We apply all of these corrections independently to images from each of the two M2FS channels and, for a given channel, to each of the 1024×1028 (plus 128×128 overscan) image sections read out via each detector's four independent amplifiers. 'ccdproc' replicates the tasks performed by the original IRAF (Tody 1986) package of the same name, but also calculates and stores a 2D array containing an estimate of the variance at each pixel.

For each amplifier on each detector and for each M2FS run individually, we generate an image of the master bias level, denoted B, by averaging (after iteratively discarding 3σ outliers at the pixel level) $\gtrsim 100$ zero-second (overscan-corrected) exposures. We generate an image of the master dark current rate, denoted D, by averaging (again with iterative 3σ outlier pixel rejection) the ≈ 250 3600-second dark exposures (after performing overscan correction and subtracting the run-dependent master bias image) taken over all M2FS runs[2] involving observations presented here. For all individual exposures of interest, we then use 'ccdproc' to perform overscan correction to obtain an image of raw counts, denoted C, and then to subtract estimates of the master bias and dark counts. Finally, 'ccdproc' applies the appropriate gain correction (typically $g \approx 0.68$ e$^-$/ADU) to obtain an estimate of counts in units of electrons:

$$\hat{N} = g(\hat{C} - \hat{B} - t_{\rm exp}\hat{D}), \quad (1)$$

where $t_{\rm exp}$ is exposure time and we adopt the convention under which $\hat{X}$ denotes the estimate of X.

The variance estimated by 'ccdproc' is by default the sum of the estimated gain-corrected count, $\hat{N}$, and the square of the read noise. One problem with this estimate is that for weak signals, read noise can dominate such that $\hat{N}$ and hence the estimated variance can be negative. Another problem with weak signals is that—even in the absence of read noise—the observed count skews toward values smaller than the expected count, and hence the variance, of a Poisson distribution. For example, for expected counts of $1, 10, 100$, random draws from Poisson distributions will be smaller than the expectation value with probability 0.37, 0.46, 0.49, respectively, and larger than the expectation value with probability 0.26, 0.42, 0.47.

Illustrating these problems and our ad hoc solution, Figure 1 depicts the mean value, from 10^6 trials over a range of input signals, of $\chi^2_1 \equiv (S - S_{\rm in})^2/\hat{\sigma}^2_S$, where S is a simulated observation, $\hat{\sigma}^2_S$ is an estimate of its variance, and $S_{\rm in}$ is the known input signal. In each trial, the simulated observation is $S = S_0 + \epsilon$, where[3] $S_0 \sim \mathcal{P}_k(S_{\rm in})$ is drawn from the Poisson distribution with expected value equal to the input signal, and $\epsilon \sim \mathcal{N}_x(0, \delta^2)$ is drawn from the normal distribution with mean 0 and variance δ^2. In our simulation, we set δ equal to the typical M2FS read noise of $\sigma_r = 2.6\,e^-$, and we assume that any additional noise associated with, e.g., empirical estimation of bias and dark levels is negligible.

The black curve in Figure 1 indicates the mean values of χ^2_1 that are calculated using the 'true' variance, $\hat{\sigma}^2_S = {\rm Var}(S) = S_{\rm in} + \delta^2$. As expected, use of the true variance gives mean χ^2_1 values of unity; unfortunately, the true variance is inaccessible to the observer who does not know the input signal.

The blue curve in Figure 1 shows the result of estimating the variance as the observationally-accessible—and commonly used—quantity $\hat{\sigma}^2_S = \max(S + \delta^2, \delta^2)$. The mean χ^2_1 asymptotes to unity only at $N_{\rm in} \gtrsim 100$. For $\delta \lesssim S_{\rm in} \lesssim 100$, the aforementioned bias toward $S < S_{\rm in}$ gives mean $\chi^2_1 > 1$ as the true variance is underestimated. For the smallest signals, $S_{\rm in} \lesssim \delta$, $\chi^2_1 < 1$ as the max operation causes the true variance to be overestimated on average.

[1] https://lweb.cfa.harvard.edu/mmti/hectospec/hecto_pipe_report.pdf

[2] We do not generate a new master dark frame for each run because a given run permits the acquisition of only a few long dark exposures.

[3] We use $\mathcal{P}_k(\lambda)$ to denote the Poisson distribution of number of occurrences, k, with expected value λ, and we use $\mathcal{N}_x(\mu, \sigma^2)$ to denote the normal distribution of random variable x, with expected value μ and variance σ^2.

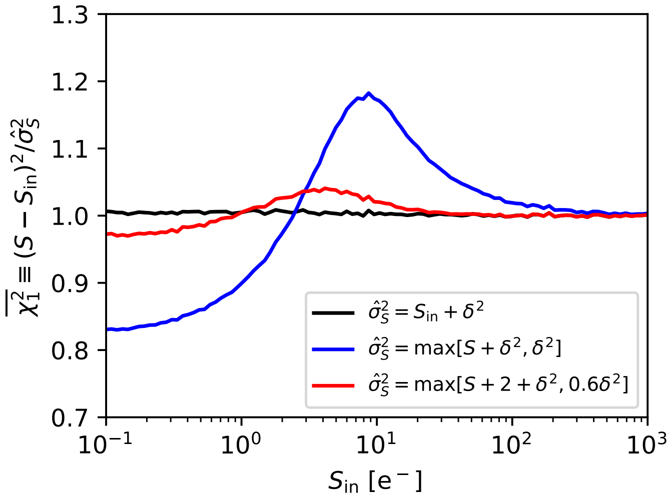

Figure 1: Mean value of $\chi_1^2 \equiv (S - S_{\rm in})^2/\hat{\sigma}_N^2$ as a function of expected signal $S_{\rm in}$, from 10^6 realizations at each input signal, assuming read noise $\sigma_r = 2.6e^-$; see Section 3.1 for details. Curves show results for different estimators of the variance, $\hat{\sigma}^2$; black uses the true variance, blue uses a commonly-used estimator, red uses the estimator we use for real M2FS data (Equation 2).

The red curve in Figure 1 shows the result of taking the variance to be

$$\hat{\sigma}_S^2 = \max(S + 2 + \delta^2, 0.6\delta^2), \qquad (2)$$

a formula that we found, via experiment, to bring mean values of χ_1^2 closer to unity at all input signals specifically when $\delta \approx 2.6e^-$; for other values of the Gaussian noise, the constants in Equation 2 would need to be re-determined.

Based on the above experiment, we use Equation 2 to estimate the variance at each pixel of the real M2FS images. In our application to real data, we take the Poisson component to be $S = \hat{N} + t_{\rm exp}\hat{D}$, the sum of estimated source (including background) and dark counts, and the Gaussian component to be $\delta^2 = \hat{\sigma}_r^2 + \sigma_{\hat{B}}^2 + t_{\rm exp}^2\sigma_{\hat{D}}^2$, the sum of contributions from the estimated read noise and noise associated with empirical estimates of bias and dark count levels.

The estimated M2FS read noise is typically $\hat{\sigma}_r \approx 2.6$ e$^-$, as calculated from the mean standard deviation over all pixels within individual images contributing to the master bias frames. The master dark frame indicates a mean dark current rate of $\hat{D} \approx 2.0$ e$^-$ hour^{-1}. The run-dependent master bias frames and the global master dark frame have typical uncertainties of $\sigma_{\hat{B}} \approx 0.15$ e$^-$ and $\sigma_{\hat{D}} \approx 0.25$ e$^-$ hour^{-1}, respectively, calculated as the standard deviations over the individual calibration frames divided by the square root of the number of calibration frames, and converted to units of electrons.

We reiterate that our application of Equation 2 represents an ad hoc solution to the problem of estimating variances of pixel counts directly from the data. It is tuned specifically to produce $\chi_1^2 \approx 1$ at $N_{\rm in} \lesssim 100$ electrons, given M2FS-like read noise; at other levels of read noise the form of Equation 2 would need to be re-determined. We note that there exist alternative solutions; e.g., Guy et al. (2022) develop a full model of the CCD image in order to estimate the variance at each pixel.

Finally, for each channel we stitch together the four independently-processed sections read by each amplifier in order to obtain a single image of size 2048 (columns) $\times$2056 (rows) square pixels. Figure 2 displays examples of the stitched frames obtained for four types of exposures, with illumination by: LED (top-left), twilight sky (top-right), Thorium-Argon-Neon lamp (bottom-left), and target stars (bottom-right). Single-order spectra appear as horizontal bands, each spanning 5130–5190 Å over columns 300–1400. Signals outside this range are contributed by light from adjacent orders, which we discard (see below).

3.2. *Identification and Tracing of Spectral Apertures*

M2FS disperses light approximately along the direction parallel to rows in the stitched images, henceforth called the x direction, where x is a continuous variable along the discrete 'column' axis (see Figure 2). Adjacent spectra are offset approximately along the 'row' axis, which we represent with continuous variable y. In order to identify and trace spectral apertures, we follow procedures similar to those performed by IRAF's 'apall' package. For each science field, we operate on the corresponding stitched LED frame (top-left panel of Figure 2), as it contains sufficient counts to identify and trace most spectral apertures. For calibration exposures of standard stars or of twilight sky, counts are sufficiently high that we can operate directly on the stitched standard and twilight frames themselves. The top-right panel of Figure 2 displays the raw image obtained from an exposure taken during evening twilight.

We begin by bundling the central 20 columns (columns 1013–1032), effectively combining them by storing their mean count as a function of row number (y value). Figure 3 displays a characteristic example of this function, which resembles an emission-line spectrum; but of course here the 'lines' are central aperture illumination profiles. We use the astropy.modeling package to fit a Chebyshev polynomial that represents the 'continuum' of this psuedo-spectrum. After subtracting the best-

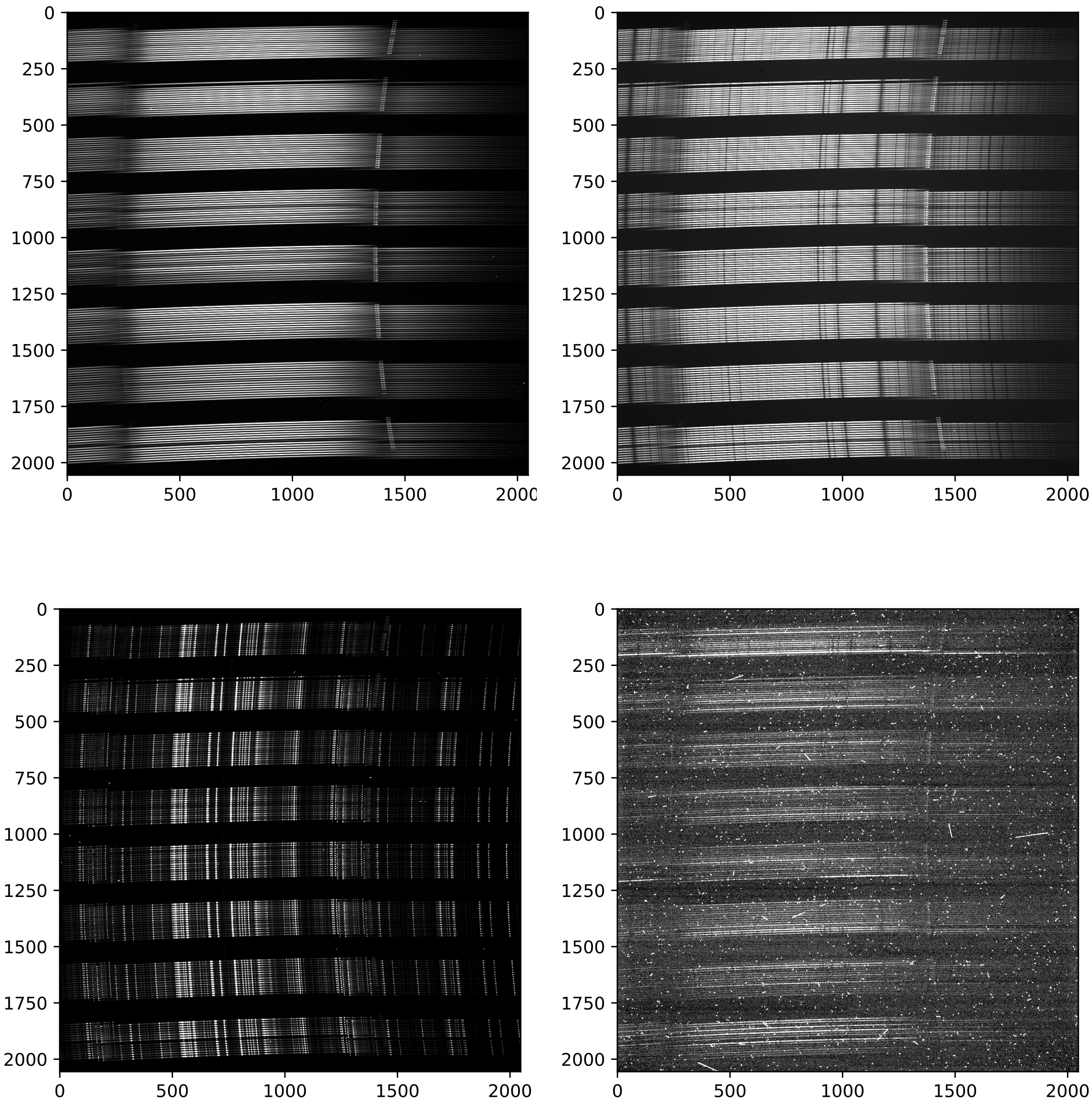

Figure 2: Examples of raw M2FS (HiRes configuration) images obtained during exposures of a calibration LED source (top left), evening twilight (top right), ThArNe arc lamp (bottom left) and a science field (bottom right). Single-order spectra appear as horizontal bands, each spanning 5130–5190 Å over columns $\sim$ 300–1400 (signal outside this column range is contributed by light from adjacent orders and is not used). The separation into eight groups of 16 apertures reflects the physical bundling of fiber ends at the spectrograph.

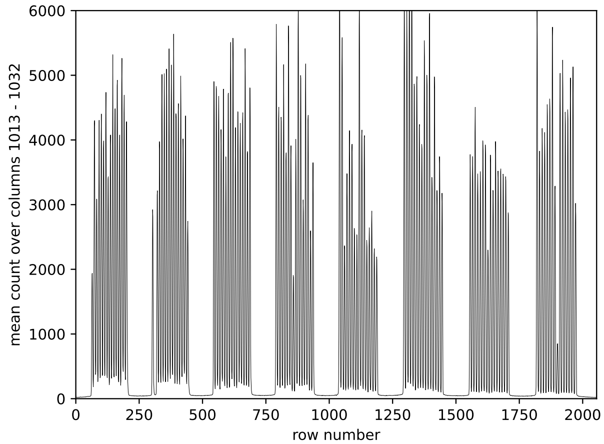

Figure 3: Identification of spectral apertures in an example M2FS frame. Plotted for each row on the detector is the mean count recorded in columns $1013-1032$ (the middle 20 columns). Local maxima signify the centers of spectral apertures.

fitting model continuum, we use the find_lines_derivative function from the astropy.specutils package to identify aperture centers as local maxima.

We use these centers to initialize Gaussian fits (again via the astropy.modeling package) to the 'continuum'-subtracted pseudo-spectrum, restricting our fits to the 10 rows around the centers returned by the find_lines_derivative function, but re-fitting those centers under the Gaussian model. The fitted Gaussian functions then represent the aperture illumination profiles across the center of the stitched image. We repeat the this process for all 102 bundles of 20 (non-overlapping) consecutive columns, allowing us to quantify how the centers and widths of aperture profiles vary with column number across the stitched image.

Next we inspect, by eye, the pseudo-spectrum along the central column bundle, as well as the Gaussian fit to each aperture profile along that central bundle. Since the apertures in both 'blue' and 'red' channels are known to follow a regular pattern of eight approximately evenly-spaced groups, with each group containing 16 approximately evenly-spaced apertures (see Figure 2), we can delete any obviously spurious aperture detections and insert artificial placeholders to represent (for book-keeping purposes) apertures corresponding to unassigned and/or broken fibers.

Then, for each visually-confirmed aperture, we trace the full 2D shape by 'marching' from the center to each edge of the useful region (columns $\sim 300-1400$; see Figure 2) in the stitched image. We begin at the position whose x coordinate is the median column number of the central column bundle, and whose y coordinate is the fitted center of the aperture profile in that bundle. We then find the fitted aperture center in the adjacent bundle that has the smallest deviation in its y coordinate. If the deviation has absolute value smaller than some threshold (we use 1.5 pixels), we step to a new position whose x coordinate is the median column number of that adjacent bundle, and whose y value is the fitted center of that bundle's aperture profile. We proceed in this manner either to the edge of the useful region in the image, or until three consecutive column bundles have no aperture whose center deviates from the current y coordinate by less than the specified threshold. We then return to the center column and march, in the same manner, toward the opposite edge. We thus obtain a list of (x, y) positions that sample the aperture's 2D trace pattern. To these data we fit and store a 4^{th}-order polynomial function, iteratively rejecting 3σ outliers. We also fit and store 4^{th}-order polynomials to the stored amplitudes and standard deviations; these two functions then characterize the aperture profile as a function of x.

3.3. *Correction for Variations in Pixel Sensitivity*

M2FS does not have an internal lamp that uniformly illuminates the detectors; all incident light travels through the fibers. In order to correct for random variations in pixel sensitivity within a given aperture, we use the previously-fit (Section 3.2) polynomials that represent center, amplitude and standard deviation of the LED aperture profile, all as functions of x, to generate a model 2D aperture image. At a given column within the aperture, we evaluate the fitted polynomials to specify the parameters (center, amplitude, standard deviation) of the Gaussian aperture profile model. We integrate that model to estimate the expected count within each pixel along the column, including all rows whose centers are within 3 aperture profile standard deviations of the aperture center. Repeating this procedure at each column, we obtain a pixelated model of the two-dimensional LED spectrum.

Dividing the actual 2D LED spectrum by this model, we obtain the equivalent of a normalized 'flat field' spectrum. After repeating for each aperture, we divide the normalized 'flat field' frame into each individual stitched image whose random variations in pixel sensitivity we wish to correct (these include science, twilight, and arc-lamp exposures).

3.4. *Correction for Scattered Light*

Having applied flat-field corrections to the stitched (science and calibration) images, next we estimate and remove scattered light. We first use the corresponding LED exposures (or bright standard star and/or twilight exposures) to mask the regions corresponding to the identified and traced spectral apertures. Specifically, we mask all pixels whose centers lie more than 3 standard deviations away from the center of the nearest aperture trace pattern, where the center and scale length are obtained by evaluating the polynomial functions fit to the aperture trace and aperture profile, respectively, at the pixel's x coordinate (Section 3.2).

Returning to the frame of interest, we then fit a 2D 4^{th}-order polynomial to the unmasked pixels, iteratively rejecting 3σ outliers, in order to estimate the contribution from scattered light. We remove scattered light by subtracting this function from the frame of interest.

3.5. *Extraction of 1D spectra*

In order to extract 1D spectra from each aperture, we collapse each column within the aperture into a single pixel regardless of the aperture trace pattern, thereby preserving independence between adjacent columns. This strategy would be optimal in the case that the spectral dispersion axis is exactly parallel to the detector's x axis. In reality the spectral apertures have nonzero curvature (Section 3.2); our procedure therefore results in some degradation of spectral resolution.

Let $\hat{N}(X,Y)$ and $\hat{\sigma}^2(X,Y)$ be the estimated count (in electrons) and estimated variance (in electrons^2), respectively, at discrete pixel (X,Y). Let $f(x,y)$ be the function that generates the 2D image of the spectrum—i.e., $f(x,y)\,dx\,dy$ is the expected count within area element $dx\,dy$ on the detector. Physically, the function $f(x,y)$ is set by the intrinsic source (plus background) spectrum, the spectral resolution and the geometry of both aperture and detector. We assume that, perpendicular to the spectral dispersion direction (i.e. the 'spatial' direction, taken to be along the y axis), the signal decays according to the Gaussian aperture profile whose parameters we evalaute from our polynomial fits described in Section 3.2, such that $f(y|x) = \mathcal{N}(y_0(x), \sigma^2(x))$, where $y_0(x)$ is the center of the aperture profile at dispersion coordinate x, and $\sigma(x)$ is the standard deviation.[4]

[4] Any functional dependence of y_0 on x violates our starting assumption that the spectra are parallel to the x axis; however, in practice the spectra are approximately aligned such that the dependence is weak.

Under this model, the predicted count at discrete pixel (X,Y) is

$$\begin{aligned} N_{\text{mod}}(X,Y) &= \int_{X_1}^{X_2} \mathrm{d}x \int_{Y_1}^{Y_2} \mathrm{d}y\, f(x,y) \\ &= \int_{X_1}^{X_2} \mathrm{d}x \int_{Y_1}^{Y_2} \mathrm{d}y\, f(x)\, f(y|x)) \\ &= \int_{X_1}^{X_2} \mathrm{d}x\, f(x) \int_{Y_1}^{Y_2} \mathrm{d}y\, \mathcal{N}(y_0(x), \sigma^2(x)) \\ &= N_{\text{mod}}(X) \int_{Y_1}^{Y_2} \mathcal{N}(y_0(X), \sigma^2(X)), \end{aligned} \tag{3}$$

where (X_1, Y_1) and (X_2, Y_2) are corners across the diagonal of the pixel. The count $N_{\text{mod}}(X)$ is, by definition, the expectation value of $f(x) = \int p(x,y)\,dy$ at column X.

Within a given aperture, we take each of the N_{row} pixels in column X to be drawn independently from a normal distribution with mean predicted by Equation 3 and variance equal to the estimated value, $\hat{\sigma}^2(X,Y)$. We define

$$\chi^2 \equiv \sum_{i=1}^{N_{\text{row}}} \frac{\left[\hat{N}(X,Y_i) - N_{\text{mod}}(X,Y_i)\right]^2}{\hat{\sigma}^2(X,Y_i)} \tag{4}$$

Minimizing χ^2 with respect to $N_{\text{mod}}(X)$, we recover the 'optimal' estimator of Horne (1986),

$$\hat{N}(X) = \frac{\sum_{i=1}^{N_{\text{pix}}} \frac{\hat{N}(X,Y_i) I_i}{\hat{\sigma}^2(X,Y_i)}}{\sum_{i=1}^{N} \frac{I_i^2}{\hat{\sigma}^2(X,Y_i)}} \tag{5}$$

where $I_i \equiv \int_{Y_{1_i}}^{Y_{2_i}} \mathrm{d}y\, \mathcal{N}\big(y0(X), \sigma^2(X)\big)$. Given the data in the 2D image, and the Gaussian aperture profile parameters fit to spectral apertures in the LED frame (Section 3.2), the estimator in Equation 5 is fully specified. For all science and calibration frames, we use Equation 5 to extract 1D spectra at every column of every aperture. We propagate the estimated variance as

$$\hat{\sigma}^2[\hat{N}(X)] = \left(\sum_{i=1}^{N_{\text{pix}}} \frac{I_i^2}{\hat{\sigma}^2(X,Y_i)}\right)^{-1}. \tag{6}$$

3.6. *Wavelength Calibration*

We calibrate wavelengths using the 1D spectra extracted from exposures of the illuminated arc lamp containing Thorium, Argon and Neon ('ThArNe') gases. At the outset, for each individually-extracted 1D ThArNe spectrum, we use a 5^{th}-order polynomial to fit and subtract the continuum component, iteratively rejecting outliers at more than 5σ below the fit or more than

1σ above (the asymmetry effectively rejects pixels that sample emission features). To the continuum-subtracted spectrum, we then use the 'find_lines_derivative' function from the astropy.specutils package to find emission features and estimate their centers in pixel space. Within the ten pixels around the center of each identified emission line, we fit a Gaussian function and store the best-fitting center, standard deviation and amplitude. The standard deviation quantifies the local spectral resolution.

Next we manually identify emission lines in a single 1D extracted ThArNe spectrum (i.e., the spectrum obtained in a single aperture), which thereafter serves as a template for automatically identifying emission lines in all other ThArNe spectra in all apertures in all ThArNe exposures acquired using the same M2FS configuration. Operating on the template spectrum only, we use NOIRLab's thorium-argon spectral atlas (Palmer & Engleman 1983) to visually identify individual emission lines interactively by eye. We store the atlas wavelength and pixel coordinate (from the Gaussian fit described above) of each line center.

Since we retain the pixelation native to the detector along the x (column) axis, we expect that the wavelength/pixel relationship will be unique for each aperture and—given small temporal changes in the aperture trace pattern—unique for each exposure. The next task, then, is to transfer our mapping of emission lines in the template ThArNe spectrum automatically to all individual non-template ThArNe spectra. For a given non-template spectrum, we begin by fitting a polynomial function that effectively distorts the template's pixel scale to bring the template's emission lines into alignment with those of the non-template spectrum. That is, letting $T(X)$ and $F(X)$ denote continuum-subtracted template and non-template ThArNe counts as functions of pixel number X, we find the order-m polynomial $P_m(x) = c_0 + c_1\big(\frac{x-x_0}{x_s}\big) + c_2\big(\frac{x-x_0}{x_s}\big)^2 + \ldots + c_m\big(\frac{x-x_0}{x_s}\big)^m$ that minimizes the sum of squared residuals $\sum_{i=1}^{N_{\rm pix}}(F(X_i) - A_1 I(X_i))^2$, where $x_0 \equiv 0.5\big(X_{max} + X_{min}\big)$ is the midpoint of the template spectrum, $x_s \equiv 0.5\big(X_{max} - X_{min}\big)$ is half the range of the template spectrum, and $I(X)$ is the linear interpolation of $T\big(A_2 + x(1 + P_m(x))\big)$ at X. We adopt $m = 4$; free parameters include the five polynomial coefficients and constants A_1, A_2.

We use the best-fitting model to transform the pixel coordinates of known emission lines in the template ThArNe spectrum to pixel coordinates in the non-template spectrum. To each emission line in the non-template spectrum, we assign the atlas wavelength of the nearest line in the transformed template spectrum, tolerating coordinate mismatches of ≤ 2 pixels. We then conduct the following iterative procedure: 1) Using only the matched features (which typically number between 25-40 per non-template spectrum), we fit a 5th-order polynomial to the atlas wavelength as a function of pixel coordinate at the line center; 2) using this updated wavelength/pixel relation for the non-template spectrum, we assign atlas wavelengths to any as-yet unidentified emission lines in the non-template spectrum if their central wavelengths match those of as-yet unused template lines within a tolerance of ≤ 0.05 Å. After iterating up to 10 times, we save for each non-template ThArNe spectrum the pixel coordinates at atlas wavelengths of the identified emission lines, coefficients of the final polynomial fit to the wavelength/pixel relation, the number of emission features used in the wavelength/pixel fit, and the rms of residuals to the fit. For the HiRes (resp. MedRes) configuration, over 34698 (3248) non-template ThArNe spectra, the mean rms residual, after excluding those below the 1st percentile and those above the 99th, is 0.009 Å (0.023 Å), with standard deviation 0.001 Å (0.003 Å).

The next step is to use the wavelength/pixel relations obtained for the ThArNe spectra to estimate wavelengths at all pixels of each individual science frame exposure. Typically we obtain ThArNe calibration frames before and after each set of science exposures for a given target field, sometimes with an additional ThArNe exposure taken in between individual science exposures. These sequences let us quantify systematic shifts in the wavelength/pixel relationship that we expect to be due to flexure of the detector hardware and its sensitivity to temperature (as measured within the spectrograph cell) changes. Using one science field's set of ThArNe exposures as an example (observed with the HiRes grating), Figure 4 displays, across both detector arrays, the slopes $d\lambda/dT$ that we fit to the wavelength/temperature relation at the location of each identified ThArNe emission line. We detect smooth variation across both detectors, with slope ranging from ~ 0 to ~ 0.04 Å/K.

In order to compensate for these systematic drifts of the wavelength/pixel relation, for every pixel in the set of ThArNe exposures corresponding to a given science field, we fit a linear model for pixel wavelength as a function of time. Individual wavelengths are weighted by the inverse-square of the rms residual with respect to the fitted wavelength/pixel relation. For the time coordinate, we use the time at the exposure midpoint. At every pixel of a given science exposure, we then assign the wavelength obtained by evaluating the linear wavelength/time function at the temporal midpoint of the science exposure. In cases where multiple ThArNe exposures are not available for monitoring tempera-

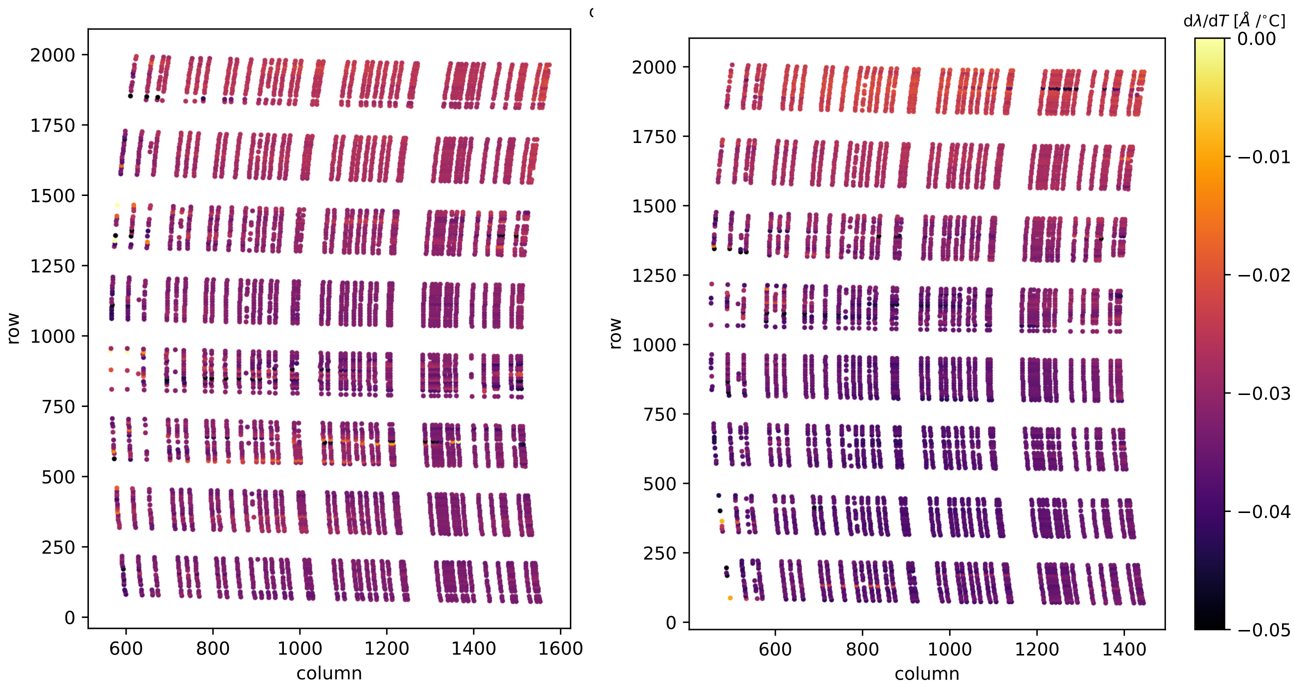

Figure 4: Change in wavelength per change in temperature (as measured at the detector), from emission lines observed in calibration exposures acquired immediately before and immediately after M2FS (HiRes configuration) observations of one example science field. Left/right panels show results for the blue and red channel, respectively.

ture and/or time dependence of the wavelength solution, we flag the corresponding catalog entries accordingly (see Section 5). The catalogs contain a column 'n_wav_calibrations' that states the number of independent ThArNe exposures used in the wavelength calibration. Two other columns, 'temp_min' and 'temp_max', give the minimum and maximum spectrograph temperature, across the science sub-exposures. For the 221 spectra where n_wav_cal=1 and temp_max−temp_min $>$ 1 K (42 of which yield measurements passing our crude quality-control filter based on velocity uncertainty), we set flag wav_cal_flag=True in the M2FS catalogs.

We do not apply heliocentric corrections to the calibrated wavelengths, which therefore include Doppler shifts due to the line-of-sight component of the observatory's velocity with respect to the barycentric rest frame. Instead we apply heliocentric corrections directly to the line-of-sight velocities estimated using the observed wavelengths (Section 4.1.2).

3.7. *Identification and masking of cosmic rays*

It is at this point that we identify and mask pixels in the extracted, wavelength-calibrated 1D science spectra that are affected by cosmic rays. To each science spectrum, we first fit the continuum level using a 4th-order polynomial, iteratively rejecting outliers more than 2σ below or 3σ above the fit, where σ is the root-mean-square value of residuals. We then flag as a likely cosmic ray signal any pixel value that exceeds the fitted continuum level by more than 5σ. In subsequent analysis, we mask these as well as the four nearest pixels. While this procedure will similarly mask bona fide emission features, we expect emission lines to be largely absent from the targeted stellar spectra over the observed spectral region.

3.8. *Correction for variations in fiber throughput*

We use the entire set of twilight exposures acquired during the observing run to estimate relative throughput as functions of fiber and wavelength. We begin by averaging, on a pixel-by-pixel basis within each aperture, the (3–10) 1D spectra from individual exposures during a given twilight sequence (i.e., the set of exposures taken during a given evening/morning twilight). When computing the mean, we weight the count in each pixel by its inverse variance. We then combine these nightly weighted-mean twilight spectra across all twilight observations within a given run, taking a new weighted mean count on a pixel-by-pixel basis within each aperture. This second averaging is unique to each science exposure, as the count in each pixel is weighted by the inverse-squared difference in time between the midpoint of the nightly twilight sequence and the midpoint of the science exposure.

In order to estimate the relative fiber throughputs that pertain to a given science exposure, we operate on the corresponding run-averaged twilight frame, where the dominant spectral features are solar absorption lines. Within each aperture, we fit a 4th-order polynomial to the mean twilight count as a function of wavelength, iteratively rejecting outliers more than 3σ above and more than 1σ below the fit in order to isolate the continuum component. For each pixel of a given science spectrum, we evaluate the twilight-continuum polynomials from all apertures at the wavelength of the pixel in the science spectrum. We then apply a wavelength-dependent throughput correction by dividing the count at each pixel by the ratio of the aperture's twilight continuum level to the median level across all apertures.

3.9. *Sky subtraction*

A typical M2FS observation allocates 20–40 fibers to regions of blank sky, split approximately evenly among the two spectrographs. For each field and spectrograph, we combine the throughput-corrected sky spectra to obtain a median sky spectrum, and then subtract the mean sky spectrum from the all throughput-corrected spectra for all targets observed with that spectrograph.

When combining individual sky spectra to obtain a single median spectrum, we must again contend with the fact that the wavelength/pixel relation is unique to each individual spectrum. Following Koposov et al. (2011), we interpolate all individual sky spectra onto a common wavelength grid that oversamples, with ten times the number of pixels, the original spectrum. We then store the median sky spectrum in the oversampled space, and record the variance at each pixel as $2.198\pi\mathrm{MAD}^2/(2N_{\mathrm{sky}})$, where N_{sky} is the number of individual sky spectra and MAD is the median absolute deviation (Koposov et al. 2011). From each individual science spectrum, we then interpolate the median sky (and variance) spectrum onto the pixel scale of the science spectrum, letting us then perform the sky subtraction directly on a pixel by pixel basis.

3.10. *Stacking subexposures*

The final step of our M2FS image processing is to combine, on an aperture-by-aperture basis, the spectra obtained in multiple exposures. For a given aperture, we combine spectra from multiple exposures by taking the weighted mean (sky-subtracted) count at each pixel. One drawback of this stacking on a pixel-by-pixel basis is that it can exacerbate the effect of temperature changes inside the spectrograph, which tend to cause the aperture trace pattern and wavelength/pixel relation to drift (Section 3.6. In order to compensate for this effect and assign wavelengths to individual pixels in the stacked spectra, we follow the same procedure described in Section 3.6, where we evaluate for each pixel the linear wavelength vs. time relation determined from ThArNe exposures. For each pixel in the stacked spectrum, we adopt as the time coordinate the mean midpoint of the individual science exposures, weighted by the inverse variance of the sky-subtracted count.

3.11. *Products*

All processed M2FS spectra are available for download from the Zenodo database[5]. For each frame of (up to) 128 spectra obtained on one of the spectrograph channels, a fits file contains the pixel wavelengths (as calibrated to the observatory rest frame—i.e., not shifted to the heliocentric frame), the sky-subtracted counts and their variances, the sky spectrum that was subtracted, the pixel mask, and the best-fitting model spectrum (Section 4), plus various observational details (e.g., date, time and spectrograph temperature of each individual exposure) and random samples from posterior probability distribution functions for model parameters inferred during analysis of the spectra (Section 4).

Figures 5, 6 and 7 display examples of fully processed spectra acquired with M2FS HiRes, M2FS MedRes and Hectochelle, respectively. Source magnitude increases from top to bottom. Left-hand and right-hand panels show spectra from stars measured to have weak and strong surface gravity, respectively, distinguishing the likely red giant stars within Galactic halo structures from dwarf stars in the Galactic foreground. **Subpanels display residuals with respect to the best-fitting model spectra (Section 4.1.2), normalized by the propagated uncertainty in the observed count. In the top two panels, hash marks identify wavelengths of known FeI, FeII and MgI absorption features that are listed in the database maintained by the Virtual Atomic and Molecular Data Centre (VALDC) Consortium, provided by the BASS2000 website**[6]

4. ANALYSIS OF MAGELLAN/M2FS AND MMT/HECTOCHELLE SPECTRA

We analyze each individual processed spectrum by fitting a model that is derived from a library of syn-

[5] DOI: 10.5281/zenodo.7837922

[6] https://doi.org/10.25935/9TXJ-F095.

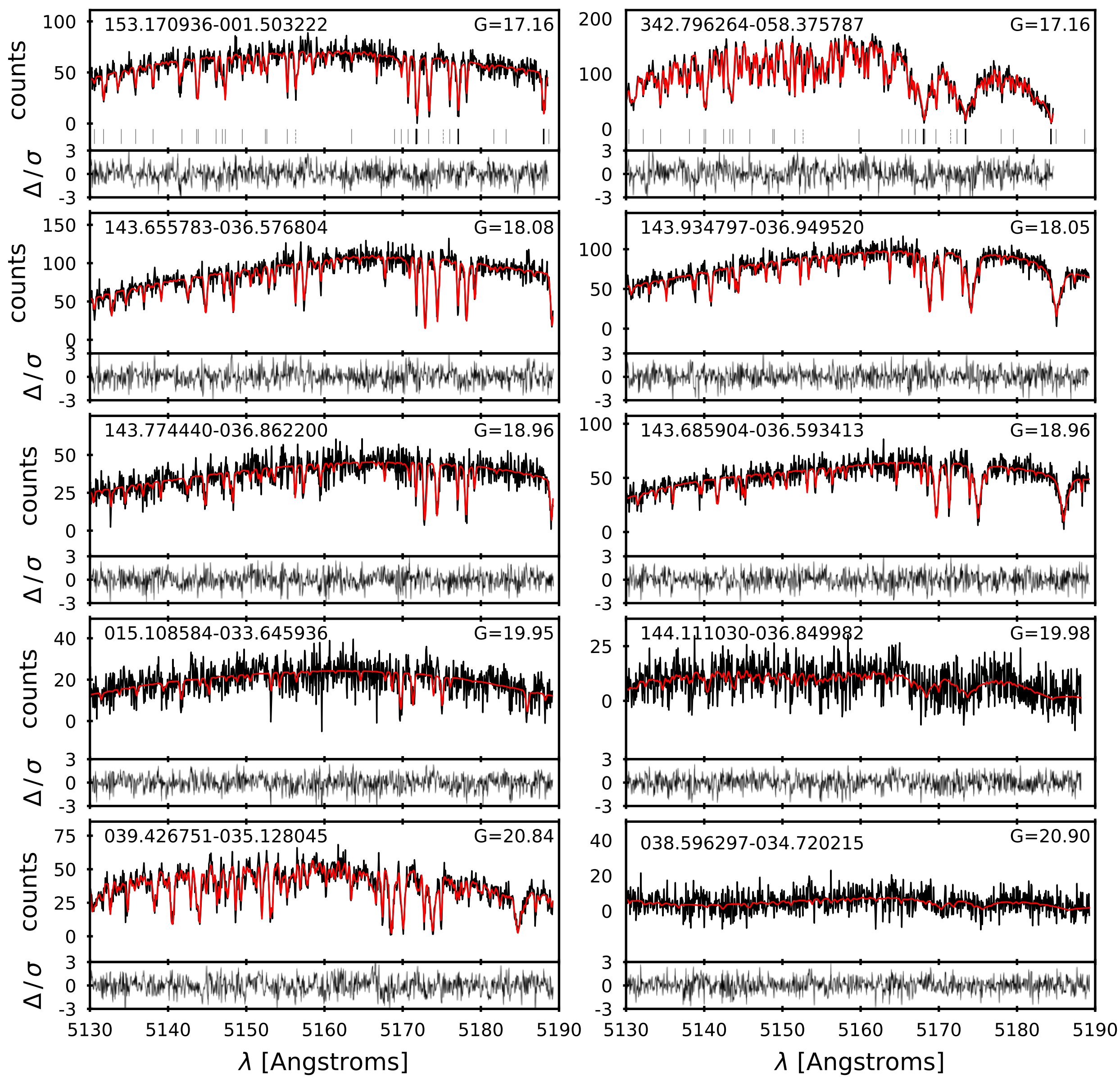

Figure 5: Examples of Magellan/M2FS HiRes spectra (black, main panels; sky-subtracted counts are scaled to the dimensions of the plotting window), which for our observing configuration cover 5125–5190 Å at resolution $\mathcal{R} \approx 24,000$. Text indicates Gaia ID and Gaia G-band magnitude. Overplotted (red) are best-fitting model spectra. **Smaller panels display normalized (by the count error propagated through the processing pipeline) residual with respect to the best fit. In the top panels, hash marks identify wavelengths (redshifted to match the observed spectrum) of known FeI (solid grey), FeII (broken grey) and MgI (solid black) lines.** Left-hand (resp. right-hand) panels depict spectra for likely red giant (dwarf) stars, with surface gravity measured to be $\log g < 1$ ($\log g > 4$).

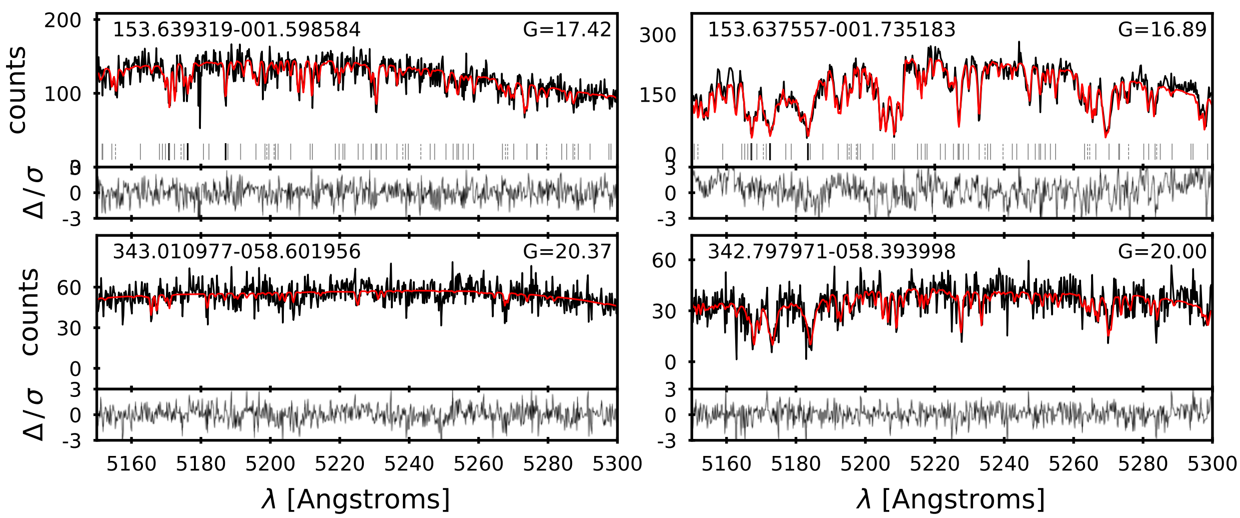

Figure 6: Same as Figure 5, but for example Magellan/M2FS MedRes spectra, which span 5115–5300 Å at $\mathcal{R} \approx 7000$.

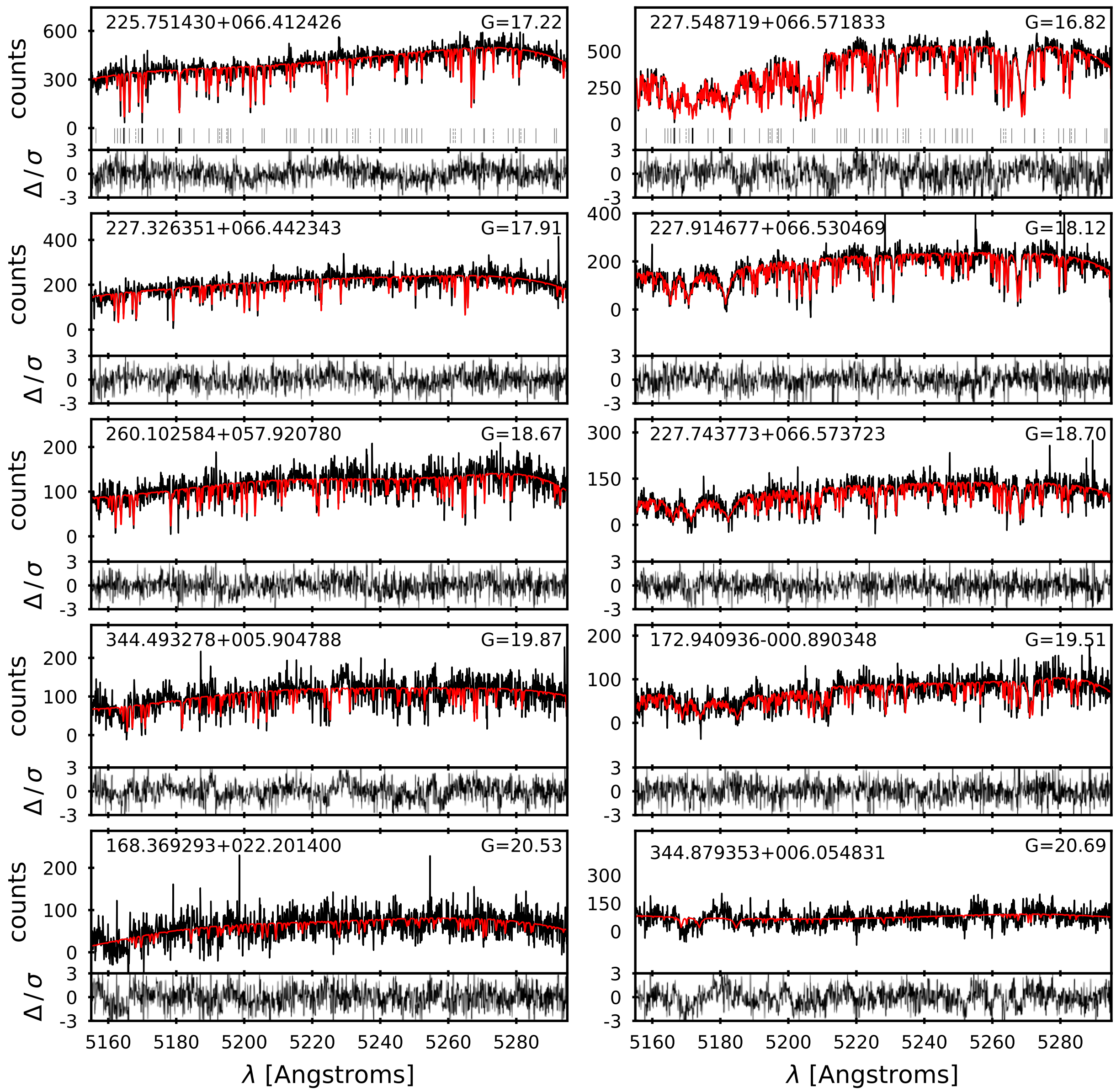

Figure 7: Same as Figure 5, but for example MMT/Hectochelle spectra, which span 5150–5300 Å at $\mathcal{R} \approx 32,000$. The larger numbers of counts (cf. Figures 5 and 6) reflect the fact that the Hectochelle pipeline calculates the sum of counts across sub-exposures, while the M2FS pipeline calculates the average.

thetic template spectra. The procedure is similar to others previously deployed for modeling stellar spectra (e.g., Koleva et al. 2009; Koposov et al. 2011; Walker, Olszewski & Mateo 2015; Li et al. 2019). Continuum-normalized synthetic spectra are computed over a grid of stellar-atmospheric parameters that has dimensions $T_{\rm eff}$, $\log g$, [Fe/H], [Mg/Fe]. An additional grid dimension extends over a parameter, $\sigma_{\rm LSF}$, that sets the spectral line spread function and thus the resolving power ($\mathcal{R} \approx \lambda/(2.355\sigma_{\rm LSF})$). Given proposed values for these parameters, we generate a model spectrum by combining (via kernel smoothing) the surrounding templates within the multi-dimensional grid space, multiplying by a flexible continuum model and adjusting template wavelengths to account for source redshift as well as any low-order corrections to the wavelength/pixel relation. We use this model spectrum to evaluate the likelihood of the observed spectrum. We use the likelihood evaluations to perform Bayesian inference, ultimately obtaining a random sample from the posterior probability distribution function (PDF) in model parameter space.

We provide details of our analysis procedure below. In most respects our procedure is identical to the one described by Walker, Olszewski & Mateo (2015) and subsequently followed by Walker et al. (2015); Spencer et al. (2017, 2018); Buttry et al. (2022); Pace et al. (2021). However, our current implementation differs in one significant way. In previous work, we adopted synthetic template spectra originally used to analyze spectra from the Sloan Digital Sky Survey's SEGUE project (Lee et al. 2008), which implicitly assumed the abundance ratio of α elements to Fe to be a fixed function of [Fe/H]. Now we use a new set of synthetic template spectra (Section 4.1.1) that we have computed over a range of [Mg/Fe], with the value of [Mg/Fe] no longer dependent on [Fe/H].

4.1. *Modeling*

Given a continnum-normalized, zero-redshift template spectrum, $T_\theta(\lambda)$, corresponding to parameters $\theta \equiv (T_{\rm eff},\log g,[{\rm Fe/H}],[{\rm Mg/Fe}],\sigma_{\rm LSF})$, we compute a model stellar spectrum according to

$$M(\lambda) = P_l(\lambda)\, T_\theta\bigg(\lambda\big[1+z+Q_m(\lambda)\big]\bigg), \qquad (7)$$

where

$$P_l(\lambda) \equiv p_0 + p_1\left[\frac{\lambda-\lambda_0}{\lambda_s}\right] + p_2\left[\frac{(\lambda-\lambda_0)}{\lambda_s}\right]^2 + \ldots + p_l\left[\frac{(\lambda-\lambda_0)}{\lambda_s}\right]^l \qquad (8)$$

is an order-l polynomial that represents a smooth continuum component. In Equation 7, rest wavelengths of the template spectrum are modified according to source redshift (in the observatory rest frame), $z \approx V_{\rm LOS}/c$, and an order-m polynomial,

$$Q_m(\lambda) \equiv \frac{q_1}{c}\left[\frac{\lambda-\lambda_0}{\lambda_s}\right] + \frac{q_2}{c}\left[\frac{(\lambda-\lambda_0)}{\lambda_s}\right]^2 +\ldots + \frac{q_m}{c}\left[\frac{(\lambda-\lambda_0)}{\lambda_s}\right]^m, \qquad (9)$$

that can apply non-linear corrections to the wavelength/pixel relation. Note that we omit from $Q_m(\lambda)$ a zero$^{\rm th}$-order term, as it would be entirely degenerate with source redshift in Equation 7. We examine zero-point redshift errors via direct comparison to external data sets (Section 4.3).

We choose $l = 5$ and $m = 2$, which provide sufficient flexibility to fit the continuum shape and to accommodate low-order corrections to the wavelength solution. We adopt scale parameters $\lambda_0 = \frac{1}{2}(\lambda_{\rm max} + \lambda_{\rm min})$ and $\lambda_s = \frac{1}{2}(\lambda_{\rm max}-\lambda_{\rm min})$ Å, such that $-1 \leq (\lambda-\lambda_0)/\lambda_s \leq +1$ over the entire range of observed wavelengths. For M2FS HiRes we use the range $\lambda_{\rm min} = 5127$ Å to $\lambda_{\rm max} = 5190$ Å. For M2FS MedRes and Hectochelle we use the range $\lambda_{\rm min} = 5155$ Å to $\lambda_{\rm max} = 5295$ Å.

4.1.1. *Template Spectra*

We present a new high-resolution grid of template spectra spanning $5050 \leq \lambda \leq 5350$ Å around the Mg I 'b' triplet. It is sampled at $\Delta\lambda = 0.05$ Å intervals, yielding a resolving power of $\mathcal{R} \approx 104{,}000$. We generate these template spectra using a recent version (2017) of the MOOG line analysis code (Sneden 1973; Sobeck et al. 2011). We interpolate model atmospheres from the ATLAS9 grid (Castelli & Kurucz 2004).

We generate line lists for the synthesis using the LINEMAKE code[7] (Placco et al. 2021). LINEMAKE creates an initial list of lines drawn from the Kurucz (2011) line compendia. It subsequentely updates the transition probabilities, hyperfine splitting structure, and isotope shifts for lines with recent laboratory analysis (e.g., Lawler et al. 2009, 2017). LINEMAKE also incorporates recent laboratory work on molecules, including CH, CN, C_2, and MgH in this spectral range (Hinkle et al. 2013; Masseron et al. 2014; Ram et al. 2014; Sneden et al. 2014). The initial list includes more than 39,000 lines. We remove the weakest lines, ones contributing less than 0.5% to the line-to-continuum opacity ratio, in a synthetic spectrum for a cool, metal-rich red giant ($T_{\rm eff}$ = 4000 K, $\log g$ = 0.0, [Fe/H] = +0.5). These lines contribute negligible absorption to stars that are warmer

[7] https://github.com/vmplacco/linemake

and/or more metal poor. The final line list contains 17,884 lines.

As a proof of concept, we compare a small region of synthetic spectra generated using these tools with the observed spectra of the Sun and Arcturus (Kurucz et al. 1984; Hinkle et al. 2000) in Figure 8. We adopt the Holweger & Müller (1974) empirical model atmosphere for the Sun, and we adopt the Ramírez & Allende Prieto (2011) model atmosphere parameters for Arcturus ($T_{\rm eff}$ = 4286 K, $\log g$ = 1.66, microturbulence velocity parameter ($v_{\rm t}$) = 1.74 km s^{-1}, and [Fe/H] = −0.52). We also adopt [Mg/Fe] = +0.37, [Si/Fe] = +0.33, and [Ti/Fe] = +0.24 in our synthesis of the Arcturus spectrum. We have empirically adjusted a small fraction (≈0.4%) of the $\log(gf)$ values in our final linelist to better reproduce the 300 Å region of interest for the Solar and Arcturus spectra. The overwhelming majority (75 of 77) of these changes are to lines without modern laboratory work, and most are relatively weak and thus will have negligible impact on the fitting of metal-poor stellar spectra. The median absolute deviations for these regions of the Solar and Arcturus spectra are 1.2% and 3.8%, respectively, demonstrating the general reliability of our method.

We synthesize a grid spanning $3900 \leq T_{\rm eff} \leq 7500$ K in intervals of 100 K, $0.0 \leq \log g \leq 5.0$ [cgs] in intervals of 0.25 dex, $-4.0 \leq$ [Fe/H] $\leq +1.0$ in intervals of 0.25 dex, and $-1.0 \leq$ [Mg/Fe] $\leq +1.4$ in intervals of 0.20 dex. A few regions near the edge of this grid are excluded because they represent non-physical combinations of parameters or they extend beyond the ATLAS9 grid. ATLAS9 models with α enhancement are adopted when [Mg/Fe] $\geq +0.1$. The microturbulence velocity parameter is adopted as a function of $\log g$: $v_{\rm t}$ = 1.0 km s^{-1} for dwarfs ($\log g \geq 4.0$), $v_{\rm t}$ = 2.0 km s^{-1} for giants ($\log g \leq 1.0$), and varying linearly between these two points. The macroturbulence velocity is assumed to be 3.0 km s^{-1} for dwarfs and subgiants ($\log g \geq 3.0$), 8.0 km s^{-1} at $\log g = 0.0$, and varying linearly between these two points. We adopt the Solar values for carbon (^{12}C/^{13}C = 89/1) and magnesium (^{24}Mg/^{25}Mg/^{26}Mg = 79/10/11) isotope ratios. Our final grid contains a total of 186071 model spectra**, all of which we make publicly available at the Zenodo database (DOI: 10.5281/zenodo.7837922)**.

We account for the finite spectral resolution of M2FS (resolving power $\mathcal{R} \approx 24,000$ in our chosen configuration) and Hectochelle ($\mathcal{R} \approx 32,000$) by broadening each template spectrum via Gaussian kernel smoothing. We repeat for six different values of smoothing bandwidths: for modeling M2FS 'HiRes' and Hectochelle spectra we use $\sigma_{\rm LSF}$= 0.06 Å , 0.09 Å, and 0.12 Å (resolving power $\mathcal{R} \approx 37,000$, 24,000, and 18,000, respectively, at λ = 5200 Å). For modeling M2FS 'MedRes' spectra we use $\sigma_{\rm LSF}$= 0.20 Å , 0.30 Å, and 0.40 Å (resolving power $\mathcal{R} \approx 11,000$, 7,400, and 5,500, respectively.

Thus we obtain a library of 'raw' synthetic stellar template spectra that discretely samples over a regular grid spanning a finite, 5-dimensional volume. We denote as $T_{\theta_0}(\lambda)$ the raw template corresponding to grid point $\theta_0 \equiv (T_{\rm eff},\ \log g,\ [{\rm Fe/H}],\ [{\rm Mg/Fe}],\ {\rm and}\ \sigma_{\rm LSF})$. In order to evaluate models at arbitrary location (i.e., not necessarily at grid points), we combine the $2^5 = 32$ surrounding raw templates via five-dimensional Gaussian kernel smoothing:

$$T_\theta\left(\lambda\right) = \frac{\displaystyle\sum_{i=1}^{32} T_{\theta_{0,i}}(\lambda)\, K_H\big(\theta_{0,i} - \theta\big)}{\displaystyle\sum_{i=1}^{32} K_H\big(\theta_{0,i} - \theta\big)}, \qquad (10)$$

where $K_H(\mathbf{x}) \equiv \exp\big[-\frac{1}{2}\mathbf{x}^T\mathbf{H}^{-1}\mathbf{x}\big]$, and we adopt diagonal bandwidth matrix $\mathbf{H} = {\rm diag}(\mathbf{h} \circ \mathbf{h})$, with $\mathbf{h} = (300\,{\rm K}, 0.5, 0.25, 0.2, 0.03\mathring{A})$ so that the smoothing bandwidth in each dimension equals the grid spacing. We note that, as a result of this nearest-neighbor smoothing, $T_\theta(\lambda)$ is not strictly a continuous function of θ and does not necessarily equal $T_{\theta_0}(\lambda)$ when evaluated at grid points. Nevertheless, tests with mock spectra generated directly from templates indicate reliable recovery of input parameters (Walker, Olszewski & Mateo 2015).

4.1.2. *Inference*

We estimate model parameters via Bayesian inference. Given observed spectrum S, the model specified by free parameter vector θ has posterior probability distribution

$$P(\theta|S) = \frac{P(S|\theta)\,P(\theta)}{P(S)}, \qquad (11)$$

where $P(S|\theta)$ is the conditional probability, given the model (or 'likelihood'), of obtaining the observed spectrum, $P(\theta)$ is the prior probability distribution function for model parameters, and

$$P(S) \equiv \int P(S|\theta)\,P(\theta)\,d\theta \qquad (12)$$

is the marginal likelihood. Assuming independence among the counts at all $N_{\rm pix}$ pixels, the spectrum has likelihood

$$P(S|\theta) = \prod_{i=1}^{N_{\rm pix}} \mathcal{N}_{S_i}\big(M(\lambda_i), \sigma_i^2\big), \qquad (13)$$

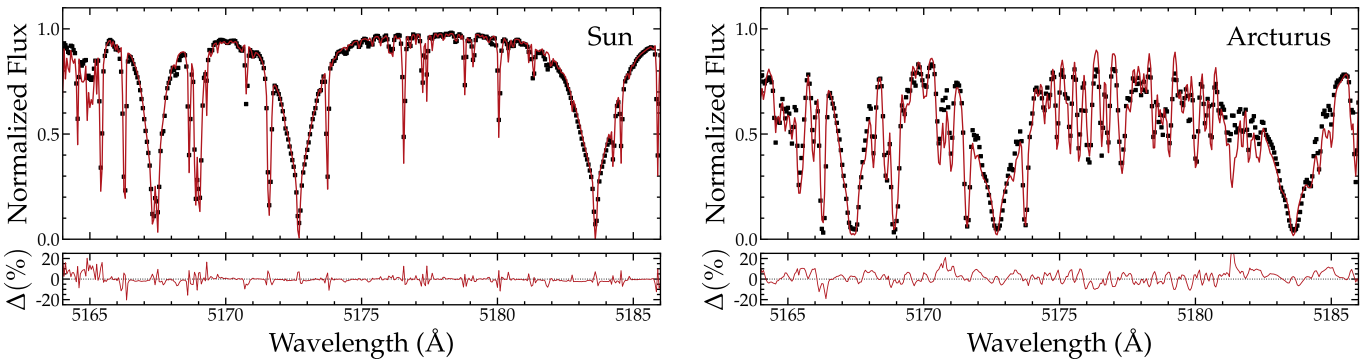

Figure 8: Validation of template spectra. Top panels compare observed Solar (left) and Arcturus (right) spectra with synthetic spectra generated using the same tools as for our template grid. Data points represent the observed spectra, resampled to the resolution of our models, $\Delta\lambda = 0.05$ Å, which are shown by the red lines. Bottom panels illustrate the differences in percent.

where

$$\mathcal{N}_{S_i}(M(\lambda_i), \sigma_i^2) \equiv \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2}\frac{\big(S_i - M(\lambda_i)\big)^2}{\sigma_i^2}\right] \tag{14}$$

is the normal distribution, with mean $M(\lambda_i)$ equal to the model prediction (Equation 7) for the count at the wavelength assigned to pixel i in the observed spectrum, and variance

$$\sigma_i^2 \equiv s_1 \hat{\sigma}_{S_i}^2 + s_2^2 \tag{15}$$

allows for a linear correction to the variance originally estimated for the observed count. In practice, given the fixed and discrete wavelength sampling of our template spectra, we evaluate (the logarithm of) Equation 13 after performing a linear interpolation of $M(\lambda)$ onto the wavelengths assigned to pixels in the observed spectrum.

Our model contains 16 free parameters. Table 3 lists each parameter, along with the range over which the priors that we adopt are uniform and nonzero.

We use the software package MultiNest (Feroz & Hobson 2008; Feroz et al. 2009) to perform the inference. MultiNest implements a nested sampling algorithm (Skilling 2004) explicitly to compute the integral in Equation 12. As part of this procedure it obtains a random sample from the posterior PDF (Equation 11). These samples, for all of our M2FS and Hectochelle spectra, are provided along with the spectra at https://cmu.box.com/v/m2fs-hectochelle.

For convenience and simplicity of downstream analysis, we use simple statistics to summarize the full posterior PDFs. Specifically, we use MultiNest's random sampling of the PDF to estimate the mean, standard deviation, skew and kurtosis of the marginal (1D) posterior PDF for each model parameter.

4.2. *Internal Validation*

In previous work we have used the summary statistics for posterior PDFs to define quality control filters. For example, Walker, Olszewski & Mateo (2015) discard any observation for which the sampled marginal PDF for $V_{\rm LOS}$ has standard deviation > 5 km s^{-1}, and/or skew and/or kurtosis deviating by more than one from the Gaussian value of zero. However, we find that the skew/kurtosis filters are approximately redundant with the cut on standard deviation alone. Therefore, in the analysis that follows, by default we discard only those observations for which the sampled marginal PDF for $V_{\rm LOS}$ has standard deviation > 5 km s^{-1}. Of course, other users might reasonably choose other quality control criteria, depending on scientific goals.

In our M2FS HiRes (M2FS MedRes, Hectochelle) sample, 8983 (189, 13328) spectra yield measurements that pass our simple quality-control filter. These spectra come from 6609 (82, 9678) unique sources, with 1330 (33, 2357) sources having multiple independent measurements. Figure 9 displays the distribution of number of independent measurements per star. As the number of independent measurements increases, the number of stars having that number of measurements declines approximately as a power law, with the the M2FS sample containing stars having as many as 16 measurements, and the Hectochelle sample containing stars having as many as 14 measurements. In the M2FS sample, all stars having more than 10 measurements come from repeated observations of the Tucana II dwarf galaxy.

We use the stars with repeat observations to fit models that specify the observational error associated with each measurement of each physical model parameter ($V_{\rm LOS}$,$T_{\rm eff}$,$\log g$,[Fe/H],[Mg/Fe]). For a given physical parameter, denoted here generically as X, we consider all pairs of independent measurements, X_1 and X_2, of the same sources. Following Li et al. (2019), we as-

Table 3: Free parameters and priors of Spectral Model

parameter	prior	description
$V_{\rm LOS}/({\rm km\ s}^{-1})$	uniform between $-500, +500$	line-of-sight velocity
$T_{\rm eff}/{\rm K}$	uniform between $3900, 7500$	effective temperature
$\log g$	uniform between $0, 5$	base-10 logarithm of surface gravity, cgs units
[Fe/H]	uniform between $-4.0, +0.5$	iron abundance
[Mg/Fe]	uniform between $-0.8, +1.0$	magnesium abundance
p_0	uniform between[a] $-\max[S(\lambda)], +\max[S(\lambda)]$	polynomial coefficient (continuum; eq 8)
p_1	uniform between $-\max[S(\lambda)], +\max[S(\lambda)]$	polynomial coefficient (continuum; eq 8)
p_2	uniform between $-\max[S(\lambda)], +\max[S(\lambda)]$	polynomial coefficient (continuum; eq 8)
p_3	uniform between $-\max[S(\lambda)], +\max[S(\lambda)]$	polynomial coefficient (continuum; eq 8)
p_4	uniform between $-\max[S(\lambda)], +\max[S(\lambda)]$	polynomial coefficient (continuum; eq 8)
p_5	uniform between $-\max[S(\lambda)], +\max[S(\lambda)]$	polynomial coefficient (continuum; eq 8)
$q_1/({\rm km\ s}^{-1})$	uniform between $-10, +10$	polynomial coefficient (wavelength solution; eq. 9)
$q_2/({\rm km\ s}^{-1})$	uniform between $-10, +10$	polynomial coefficient (wavelength solution; eq. 9)
$\sigma_{\rm LSF}/{\rm Å}$	uniform between $0.06, 0.12$ (M2FS HiRes, Hectochelle)	bandwidth of Gaussian kernel to broaden line spread function
$\sigma_{\rm LSF}/{\rm Å}$	uniform between $0.2, 0.4$ (M2FS MedRes)	bandwidth of Gaussian kernel to broaden line spread function
$\log_{10} s_1$	uniform between $-1, +6$	rescales observational errors (eq. 15)
$\log_{10} s_2$	uniform between $-2, +2$	adds to observational errors (eq. 15)

[a] $\max[S(\lambda)]$ is the maximum value (discounting pixels flagged as cosmic rays) of the sky-subtracted spectrum.

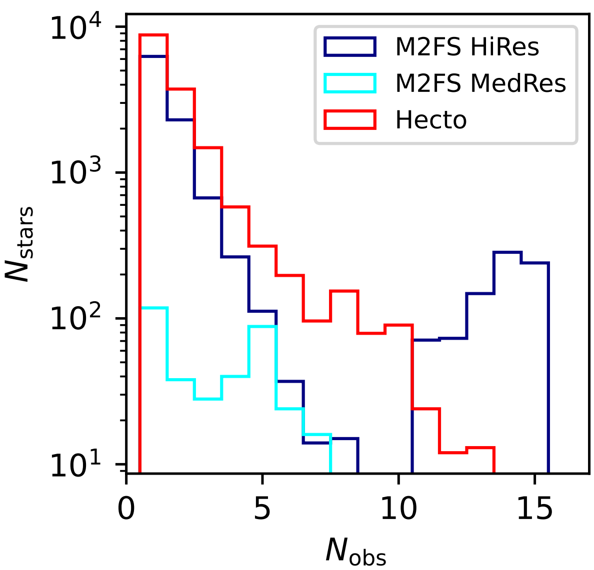

Figure 9: Distribution of number of independent measurements (having line-of-sight velocity error < 5 km s^{-1}), for M2FS HiRes (blue), M2FS MedRes (cyan) and Hectochelle (red) samples. The bump in the M2FS HiRes sample at $N_{\rm obs} > 10$ is contributed entirely by repeated observations of the Tucana II dwarf galaxy.

sume that deviations $\Delta X \equiv X_1 - X_2$ are distributed as a mixture of two Gaussian distributions. The first has variance set by formal observational errors; the second, which allows for 'outlier' measurements—including spurious measurements and/or cases of true variability, as with velocities measured for stars in binary systems—has constant variance $\sigma^2_{\rm out}$ that is unrelated to formal observational errors. That is, given zero-point offset $\mu_{\Delta X}$, variance $\sigma^2_{\Delta X} \equiv \sigma^2_{X_1} + \sigma^2_{X_2}$ that is set by formal observational errors σ_{X_1} and σ_{X_2}, outlier variance $\sigma^2_{\rm out}$ and outlier fraction $f_{\rm out}$, the deviation between measurements 1 and 2 of a common source has probability

$$P(\Delta X | \mu_{\Delta X}, \sigma^2_{\Delta X}, f_{\rm out}, \sigma_{\rm out}) = (1 - f_{\rm out})\mathcal{N}(\mu_{\Delta X}, \sigma^2_{\Delta X}) + f_{\rm out}\mathcal{N}(0, \sigma^2_{\rm out}). \quad (16)$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. We assume $\mu_{\Delta X} = 0$ when comparing measurements from the same instrument, as in this section, but not when comparing measurements from different instruments, as in Section 4.3. We model the formal random errors as linear (in quadrature) functions of the standard deviations, denoted $\sigma_{X_{1,\rm MN}}$ and $\sigma_{X_{2,\rm MN}}$, obtained directly from MultiNest's random sampling of the marginal (1D) posterior PDF for parameter X. That is, we assume

$$\sigma^2_{X_1} = s^2 + k^2\sigma^2_{X_{1,\rm MN}},$$
$$\sigma^2_{X_2} = s^2 + k^2\sigma^2_{X_{2,\rm MN}}, \quad (17)$$

and similar for all pairs of measurements obtained for common sources that deviate by amounts smaller than a threshold, $|\Delta X|_{\rm out}$. We assume that deviations larger than $|\Delta X|_{\rm out}$ are are contributed by spurious measurements, which we then exclude from our analysis (but not from the catalogs presented below). We take $|\Delta V_{\rm LOS}|_{\rm out} = 100$ km s^{-1}, $|\Delta T_{\rm eff}|_{\rm out} = 2000$ K, $|\Delta \log {\rm g}|_{\rm out} = 2.5$ dex, $|\Delta[{\rm Fe/H}]|_{\rm out} = 2.5$ dex and $|\Delta[{\rm Mg/Fe}]|_{\rm out} = 1.0$ dex. We assume that a single value of the error 'floor', s, and a single value of scaling parameter, k, hold across the entire sample obtained with a given telescope/instrument. The total set of deviations, over all $N_{\rm pair}$ pairs of measurements, has likelihood

$$\prod_{i=1}^{N_{\rm pair}} \frac{P(\Delta X_i|\mu_{\Delta X},\sigma_{X_1},\sigma_{X_2},f_{\rm out},\sigma_{\rm out})}{\int_{-|\Delta X|_{\rm out}}^{+|\Delta X|_{\rm out}} P(\Delta X_i|\mu_{\Delta X}\sigma_{X_1},\sigma_{X_2},f_{\rm out},\sigma_{\rm out})\,d(\Delta X)}. \quad (18)$$

We consider all pairs of measurements that both satisfy our crude quality-control criterion (velocity error ≤ 5 km s^{-1}) for common sources, excluding measurements from sources listed in Gaia's (DR3) catalog of RR Lyrae variables (see Section 4.4). This selection gives $N_{\rm pair} = 6830$ for M2FS HiRes, $N_{\rm pair} = 259$ for M2FS MedRes, and $N_{\rm pair} = 6301$ for Hectochelle. We use MultiNest to perform the inference. For each of the five physical parameters we infer from spectra, Table 4 lists the prior for each of the four parameters of our error model, as well as the mean and standard deviation of the marginal posterior PDF.

For M2FS HiRes (MedRes), we infer error 'floors' of $s_{V_{\rm LOS}} = 0.57 \pm 0.01$ km s^{-1} (0.59 ± 0.78 km s^{-1}), $s_{T_{\rm eff}} = 58.59 \pm 2.13$ K (10.16 ± 19.39 K), $s_{\rm logg} = 0.12 \pm 0.02$ (0.03 ± 0.02), $s_{\rm [Fe/H]} = 0.06 \pm 0.00$ (0.12 ± 0.08), $s_{\rm [Mg/Fe]} = 0.04 \pm 0.01$ (0.03 ± 0.02). For Hectochelle, the floors are all lower, presumably as a benefit of wider spectral coverage, with $s_{V_{\rm LOS}} = 0.39 \pm 0.01$ km s^{-1}, $s_{T_{\rm eff}} = 0.61 \pm 1.00$ K, $s_{\rm logg} = 0.03 \pm 0.01$, $s_{\rm [Fe/H]} = 0.01 \pm 0.00$, $s_{\rm [Mg/Fe]} = 0.01 \pm 0.00$. The inferred scaling parameters are scattered around unity, in several cases (including the velocity measurements for M2FS HiRes and Hectochelle) consistent with a value of unity within the 99% credible interval. The outlier fraction tends to comprise $\lesssim 10\%$ of the samples, except for the measurements of [Fe/H] and [Mg/Fe], where the outlier fractions reach $\sim 20-50\%$. Analyses of chemical abundance distributions may therefore benefit from stricter sample selection criteria than our fiducial one that is based solely on the formal error in $V_{\rm LOS}$.

Figure 10 shows distributions of pair-wise measurement deviations normalized by combined measurement errors, with the combined measurement error calculated from the standard deviations of the posterior PDF originally sampled by MultiNest, $\sigma_{\Delta X_{\rm MN}} = \sqrt{\sigma^2_{X_{1,\rm MN}} + \sigma^2_{X_{2,\rm MN}}}$ (black histograms), and from the formal errors returned by the best-fitting error model, $\sigma_{\Delta X} = \sqrt{\sigma^2_{X_1} + \sigma^2_{X_2}}$ (red histograms). By design, the latter are generally closer to the standard normal distribution (solid black curves). In our data catalogs, the columns 'X_error' list the errors for observable 'X' after performing the adjustment of Equation 17, with mean values of error model parameters listed in Table 4. The columns 'X_error_raw' list the pre-adjusted values obtained directly from the posterior sampled by MultiNest.

4.3. *External Comparisons*

We compare our M2FS and Hectochelle catalogs directly to each other and to large spectroscopic data sets that are previously published and/or in progress. Our primary goal is to detect and quantify systematic differences, e.g., zero-point offsets. The top panels of Figures 11 and 12 compare velocities and stellar-atmospheric parameters, respectively, that we measure with M2FS HiRes, M2FS MedRes and Hectochelle, for all stars that appear in at least two instrument-specific samples. In both figures, the bottom three rows of panels compare our M2FS and Hectochelle measurements to those from external catalogs by Walker, Mateo & Olszewski (2009, 'W09' hereafter), Kirby et al. (2010, 'K10' hereafter), the Sloan Digital Sky Survey's APOGEE project (Abdurro'uf et al. 2022, DR17), and the Hectochelle in the Halo at High Resolution Survey (Conroy et al. 2019, 'H3' hereafter).

W09's catalog includes 8855 line-of-sight velocities measured for 7103 unique sources toward the dwarf spheroidal galaxies Carina, Fornax, Sculptor and Sextans. The W09 spectra were acquired using the Michigan-MIKE Fiber System (Walker et al. 2007), a precursor to M2FS at Magellan that operated at similar spectral resolution over a similar spectral range. The W09 catalog has 1440 sources in common with our current M2FS HiRes sample, 10 sources in common with our M2FS MedRes sample, and 194 sources in common with our Hectochelle sample. While W09 measure spectroscopic indices for iron and magnesium absorption features, they do not measure the set of stellar-atmospheric parameters that we have in our current samples. Our comparisons to W09's catalog are therefore limited to line-of-sight velocities.

K10 measure $T_{\rm eff}$, $\log g$, [Fe/H] and [Mg/Fe] for ~ 3000 stars in eight of the Milky Way's dSph satellites. The K10 catalog has 115 (all HiRes mode) and 326 stars in common with our M2FS and Hectochelle samples, respectively. The K10 spectra have resolving power $\mathcal{R} \sim 6500$ near the calcium triplet at $\lambda \sim 8500$ Å, probing a different wavelength range at lower resolution than the other catalogs considered here. In contrast to our estimates of $T_{\rm eff}$ and $\log g$, which rely entirely on information contained in the spectrum, K10 incorporate stellar photometry into their estimate of $T_{\rm eff}$ and use photometry alone to estimate $\log g$. The K10 catalog does not list measurements of $V_{\rm LOS}$; therefore, our comparisons

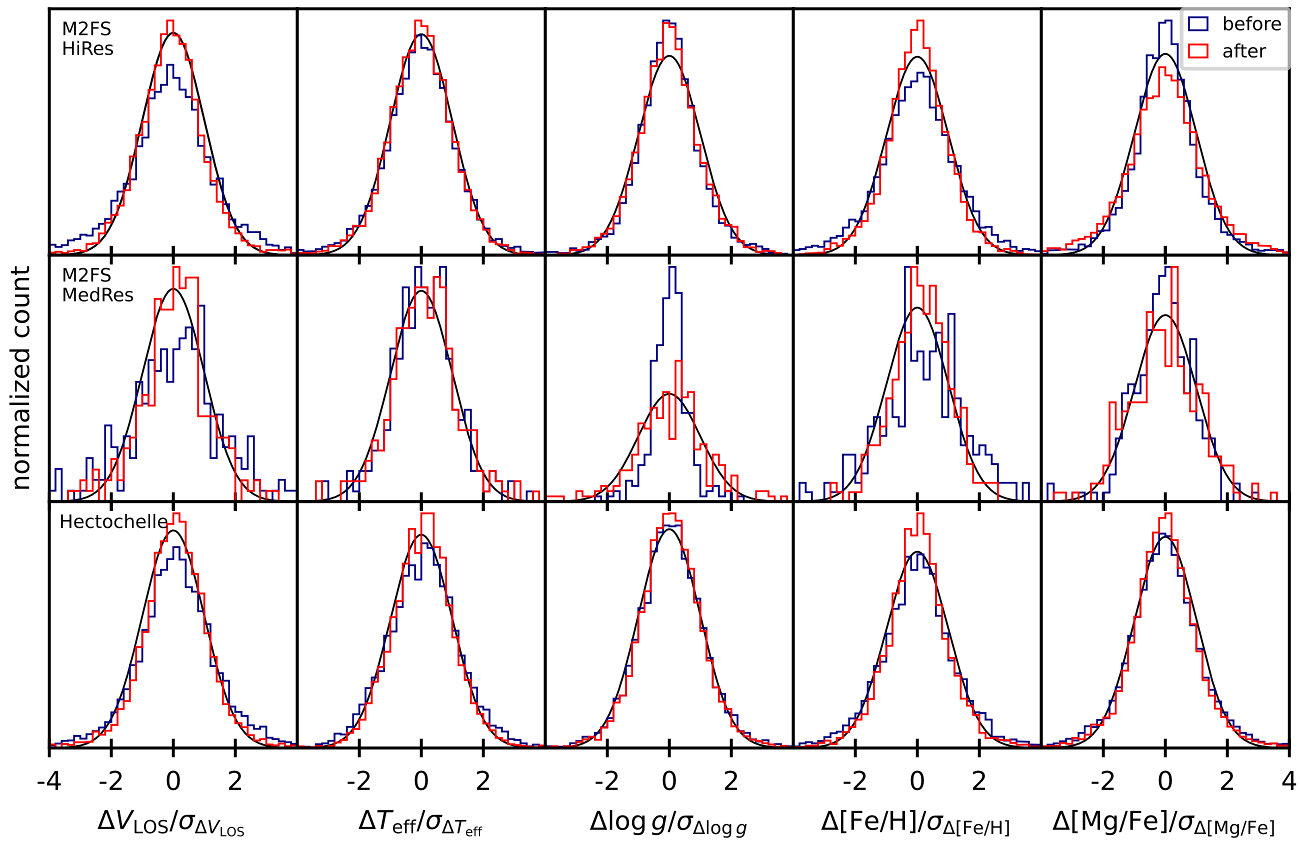

Figure 10: Internal validation of formal uncertainties. For each of the spectroscopic observables $V_{\rm LOS}$, $T_{\rm eff}$, $\log g$, [Fe/H], [Mg/Fe], panels indicate distributions of pair-wise deviations between independent measurements of the same target, normalized by the combined error in both measurements. Errors are calculated using the standard deviation taken directly from the marginal posterior PDFs returned by MultiNest (blue 'before' histograms), and using the formal errors obtained by fitting the error model described in Section 4.2 (red 'after' histograms). Individual panels in the top, middle and bottom rows show results from 6830 pairs of M2FS HiRes observations, 259 pairs of M2FS MedRes observations, and 6301 pairs of Hectochelle observations, respectively. In all panels, the solid black curve is the standard normal distribution.

to K10's catalog are limited to stellar-atmospheric parameters.

The APOGEE catalog, from the 17th data release of the Sloan Digital Sky Survey (DR 17; Abdurro'uf et al. 2022), includes line-of-sight velocities and stellar-atmospheric parameters measured from high-resolution ($\mathcal{R} \sim 22,500$ over $\sim 1.5-1.7$ microns in wavelength) spectra obtained for $\sim 650,000$ stars in the Milky Way and a few of its dwarf galaxy satellites. We select all sources from the APOGEE DR17 'allstar' catalog for which the APOGEE Stellar Parameters and Abundances Pipeline (ASPCAP) returns measurements for all of $T_{\rm eff}$, $\log g$, [Fe/H] and [Mg/Fe] (ASPCAP lists separate measurements of [Mg/Fe] and [α/Fe]; we use only the former for purposes of direct comparison). We then discard any sources for which the 'RV_FLAG' bitmask has the 'RV_SUSPECT' bit set, and/or the 'ASPCAPFLAG' bitmask has the 'STAR_WARN' bit set. After applying these filters and then removing stars for which we measure [Fe/H]< -2.5 (i.e., below the minimum metallicity of APOGEE's template spectra), there are 117 APOGEE stars in common with our M2FS HiRes sample, 2 stars in common with our M2FS MedRes sample, and 94 in common with our Hectochelle sample. For a given star, we take the mean APOGEE velocity as given by the 'VHELIO_AVG' parameter, with observational error given by 'VERR'.

Finally, the H3 Survey (Conroy et al. 2019) is ongoing, using the same MMT/Hectochelle configuration that we do. H3 is designed to map the Galactic stellar halo, tar-

Table 4: Summary of posterior PDFs for parameters of ther model used to adjust observational errors (Section 4.2).

quantity	s (floor)	k (multiplier)	$f_{\rm out}$ (outlier fraction)	$\sigma_{\rm out}$ (outlier std. dev.)
M2FS HiRes				
$V_{\rm los}$	$0.57 \pm 0.01\,{\rm km\ s^{-1}}$	0.86 ± 0.02	0.10 ± 0.00	$24.30 \pm 0.69\,{\rm km\ s^{-1}}$
$T_{\rm eff}$	58.59 ± 2.13 K	0.91 ± 0.01	0.08 ± 0.01	424.43 ± 24.79 K
$\log_{10}[g]$	0.12 ± 0.02	0.86 ± 0.04	0.13 ± 0.06	0.61 ± 0.38
[Fe/H]	0.06 ± 0.00	1.18 ± 0.02	0.17 ± 0.02	0.29 ± 0.02
[Mg/Fe]	0.04 ± 0.01	0.74 ± 0.03	0.48 ± 0.01	0.34 ± 0.01
M2FS MedRes				
$V_{\rm los}$	$0.59 \pm 0.78\,{\rm km\ s^{-1}}$	1.38 ± 0.14	0.12 ± 0.03	$116.61 \pm 170.64\,{\rm km\ s^{-1}}$
$T_{\rm eff}$	10.16 ± 19.39 K	1.00 ± 0.10	0.18 ± 0.10	507.14 ± 519.59 K
$\log_{10}[g]$	0.03 ± 0.02	0.46 ± 0.05	0.09 ± 0.05	4.24 ± 2.43
[Fe/H]	0.12 ± 0.08	1.34 ± 0.15	0.19 ± 0.12	0.51 ± 0.55
[Mg/Fe]	0.03 ± 0.02	0.80 ± 0.07	0.19 ± 0.10	0.97 ± 1.50
MMT/Hectochelle				
$V_{\rm los}$	$0.39 \pm 0.01\,{\rm km\ s^{-1}}$	0.94 ± 0.02	0.06 ± 0.00	$27.64 \pm 1.14\ {\rm km\ s^{-1}}$
$T_{\rm eff}$	0.61 ± 1.00 K	1.17 ± 0.01	0.06 ± 0.01	216.25 ± 26.55 K
$\log_{10}[g]$	0.03 ± 0.01	1.05 ± 0.02	0.10 ± 0.02	0.42 ± 0.04
[Fe/H]	0.01 ± 0.00	1.20 ± 0.02	0.16 ± 0.01	0.37 ± 0.02
[Mg/Fe]	0.01 ± 0.00	1.11 ± 0.02	0.23 ± 0.02	0.38 ± 0.02

geting $\sim 2\times 10^5$ halo stars down to a magnitude limit of $r \lesssim 18$. H3's and our spectra are acquired and modeled independently, but processed using the same CfA pipeline discussed at the beginning of Section 3. The H3 team models individual spectra using the software package MINESweeper (Cargile et al. 2020), which simultaneously fits isochrone models to stellar magnitudes measured from broad-band photometry. H3's incorporation of photometric information provides additional power to constrain stellar-atmospheric parameters, while also giving capability to infer spectro-photometric distances to individual sources.

The faint end of H3's sample overlaps only slightly with the bright end of ours, leaving relatively few stars common to both surveys. In order to provide a more meaningful basis for comparison, the H3 team applied their MINESweeper analysis directly to our Hectochelle spectra from four different fields in the Sextans dSph galaxy (P. Cargile, private communication). While not part of the actual H3 survey, this comparison 'H3' sample contains 77 sources common to our M2FS HiRes sample, 1 sources common to our M2FS MedRes sample, and 767 sources common to our Hectochelle sample.

For a given observable quantity, X, we infer zero-point offsets for each of the above catalogs simultaneously. We begin by constructing vectors of deviations, $\Delta X \equiv X_1 - X_2$, and corresponding errors, $\sigma_{\Delta X} = \sqrt{\sigma^2_{X,1}+\sigma^2_{X_2}}$, for all pairs of sources common to different catalogs '1' and '2'. We loop over all possible combinations of catalogs, such that a star appearing at least once in all six catalogs will have 10 pairs of measurements[8]; within a given catalog, multiple measurements of the same source are replaced by the inverse-variance-weighted mean value.

We assume that, for a given observable, X, the pairwise deviations, ΔX, follow a Gaussian distribution, with standard deviation $\sigma_{\Delta X}$ and pair-dependent mean, $\mu_{\Delta X} = \overline{\Delta X_1} - \overline{\Delta X_2}$, that is specified by the differ-

[8] Recall that the W09 catalog lacks stellar-atmospheric parameters and the K10 catalog lacks line-of-sight velocities, so measurements of a given quantity for a given star can appear in up to five different catalogs.

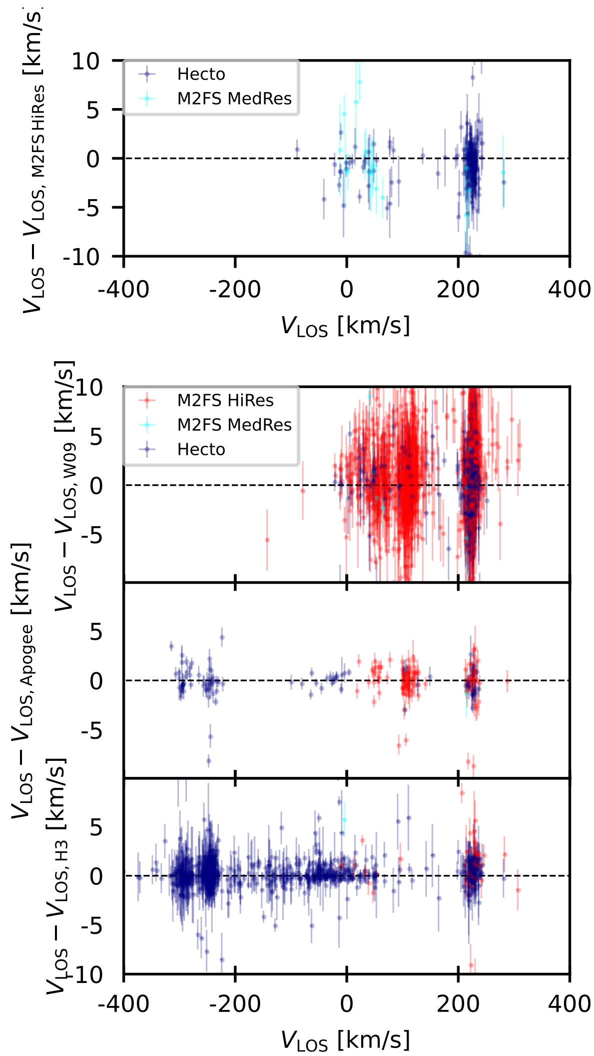

Figure 11: *Top row:* Difference between line-of-sight velocities measured by M2FS HiRes and either M2FS MedRes (cyan) or Hectochelle (blue), for stars common to both samples. *Bottom three rows:* Differences between line-of-sight velocities we measure using M2FS HiRes (red), M2FS MedRes (cyan) or Hectochelle (blue) and those measured in external surveys by Walker, Mateo & Olszewski (2009, second row), APOGEE (DR17; Abdurro'uf et al. 2022, third row), the H3 survey (Conroy et al. 2019, fourth row).

ence in mean offsets (from some standard zero point) of catalogs 1 and 2. This model and the corresponding likelihood function can be specified by Equations 16 and 18, respectively, only now with the outlier fraction assumed to be $f_{\rm out} = 0$ and the catalogued observational errors taken at face value ($s = 0$, $k = 1$). In order to guard against catastrophic outliers, as described in Section 4.2, we discard pairs with deviations in excess of $|\Delta V_{\rm LOS}|_{\rm out} = 100$ km s^{-1}, $|\Delta T_{\rm eff}|_{\rm out} = 2000$ K, $|\Delta \log {\rm g}|_{\rm out} = 2.5$ dex, $|\Delta{\rm [Fe/H]}|_{\rm out} = 2.5$ dex and $|\Delta{\rm [Mg/Fe]}|_{\rm out} = 1.0$ dex. Four free parameters specify zero-point offsets: $\overline{\Delta X}_{\rm M2FS}$, $\overline{\Delta X}_{\rm Hecto}$, $\overline{\Delta X}_{\rm H3}$, and $\overline{\Delta X}_{\rm W09}$ (if $X = V_{\rm LOS}$), $\overline{\Delta X}_{\rm K10}$ (if $X = T_{\rm eff}$, $\log g$, [Fe/H] or [Mg/Fe]).

Given the APOGEE catalog's size and widespread use across different sub-fields, we choose that catalog to define the absolute zero point, assuming $\overline{\Delta X}_{\rm Apo} = 0$ for all X. Table 5 lists offsets, relative to the APOGEE zero point, that we infer (again via MultiNest, as in Section 4.2) for each observable and each catalog. Positive offsets, $\overline{\Delta X} > 0$, imply that a catalog's zero point is more positive than APOGEE's. For each catalog named in Column 1, Columns 2–7 identify the number of pairs of sources in common with each of the other individual catalogs.

Examining the results for our M2FS and Hectochelle samples, we find that, whereas the Hectochelle sample shows little velocity offset with respect to APOGEE ($\overline{\Delta V_{\rm LOS}}_{\rm Hecto} = -0.14 \pm 0.05$ km s^{-1}), the M2FS HiRes sample is systematically offset by $\overline{\Delta V_{\rm LOS}}_{\rm M2FS,HiRes} = 0.47 \pm 0.05$ km s^{-1}. The M2FS MedRes sample shows no significant offset, with $\overline{\Delta V_{\rm LOS}}_{\rm M2FS,MedRes} = 0.07 \pm 0.44$ km s^{-1}, but with a large uncertainty reflecting the fact that the M2FS MedRes sample has relatively few stars in common with the other samples. For most stellar-atmospheric parameters, both M2FS HiRes and Hectochelle samples show statistically significant offsets from APOGEE. The offsets in surface gravity ($\overline{\Delta \log g}_{\rm M2FS\,HiRes} = -0.53 \pm 0.02$ and $\overline{\Delta \log g}_{\rm Hecto} = -0.49 \pm 0.01$) and metallicity ($\overline{\Delta {\rm [Fe/H]}}_{\rm M2FS\,HiRes} = -0.26 \pm 0.01$ and $\overline{\Delta{\rm [Fe/H]}}_{\rm Hecto} = -0.26 \pm 0.01$) are similar for both samples, while the difference in temperature offsets ($\overline{\Delta T_{\rm eff}}_{\rm M2FS\,HiRes} = -141 \pm 5$ K and $\overline{\Delta T_{\rm eff}}_{\rm Hecto} = -221 \pm 4$ K) likely reflects the different wavelength coverage of the different instruments/configurations. However, the smaller temperature offset of the H3 sample ($\overline{\Delta T_{\rm eff}}_{\rm H3} = -69 \pm 4$ K), which uses the same Hectochelle configuration that we do, also implicates differences in analysis procedure as a source of systematic error. Finally, while our Hectochelle sample shows good agreement with APOGEE in terms of the magnesium abundance ($\overline{\Delta{\rm [Mg/Fe]}}_{\rm Hecto} = -0.01 \pm 0.01$), the M2FS sample is offset by $\overline{\Delta{\rm [Mg/Fe]}}_{\rm M2FS} = 0.19 \pm 0.01$.

Perhaps most eye-catching among the external comparisons are those involving surface gravity in the H3 catalog (bottom row, second column of Figure 12). The H3 surface gravities are multi-modal at $\log g \gtrsim 2$. This feature is likely real (i.e., reflecting a true multi-modality

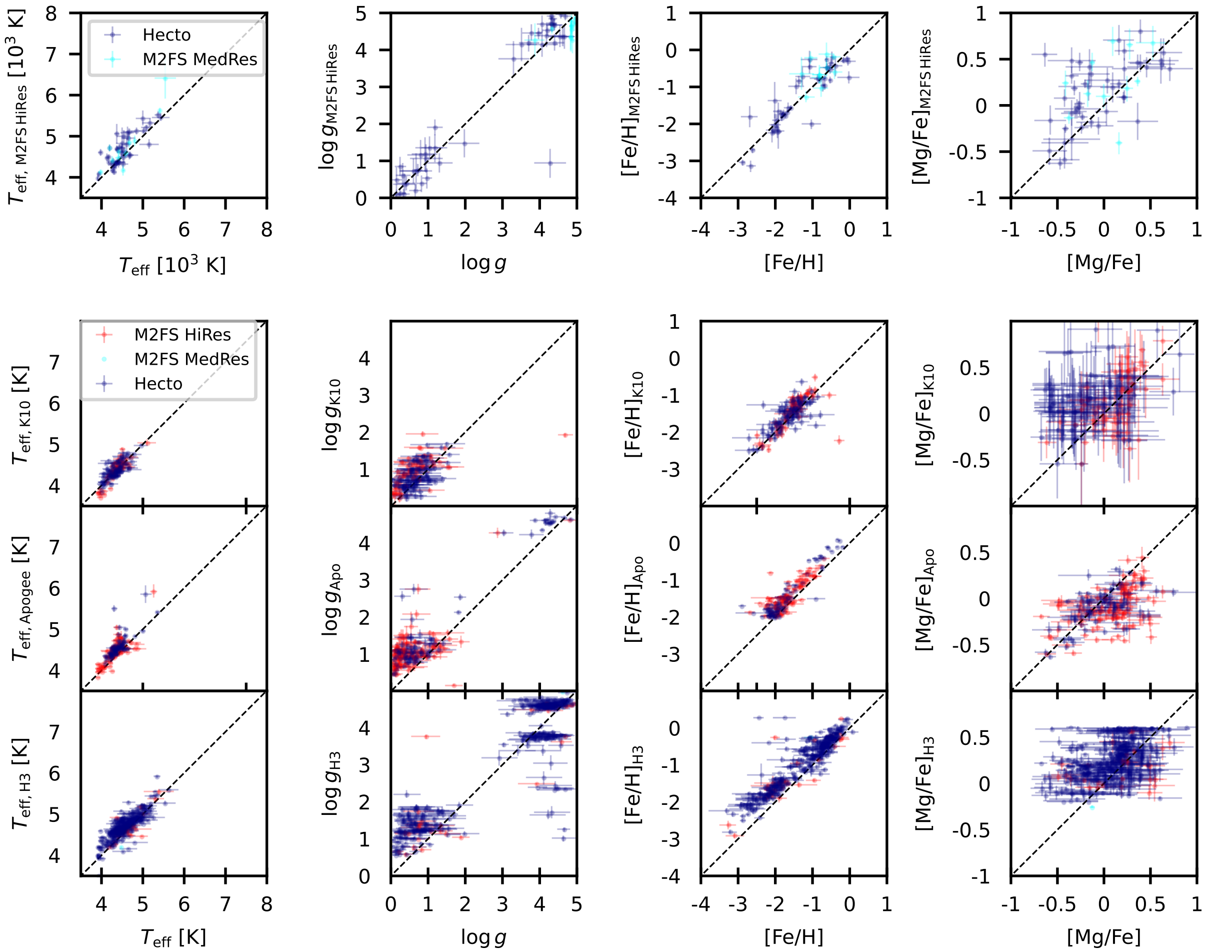

Figure 12: *Top row:* Comparison of stellar-atmospheric parameters measured **(before applying zero-point adjustments)** by M2FS HiRes and either M2FS MedRes (cyan) or Hectochelle (blue), for stars common to both samples. *Bottom three rows:* Comparison of parameters that we measure **(before applying zero-point adjustments)** using M2FS (red) and Hectochelle (blue) to those measured in external surveys by Kirby et al. (2010, second row), APOGEE (DR17; Abdurro'uf et al. 2022, third row), and the H3 survey (Conroy et al. 2019, fourth row). In all panels of the bottom three rows, the quantity plotted along the horizontal axis is the measurement from M2FS (red) and/or Hectochelle (blue).

among the observed high-gravity stars) and detectable here because of H3's simultaneous fitting of isochrone and spectral models. The modes at $\log g \sim 4.5$, $\log g \sim 3.7$ and $\log g \sim 2.5$ correspond to the main sequence, sub-giant and horizontal branches, respectively, all of which are confined to distinct ranges of surface gravity in isochrone space. H3's fitting of isochrone models to broad-band photometry effectively requires these evolutionary stage to be separated, giving rise to the observed multi-modality in $\log g$ space.

The primary lesson we take from all of these external comparisons is that zero-point offsets among *all* of the independent datasets are common at the level of a few $\times 0.1$ km s^{-1} in line-of-sight velocity, ~ 100 K in effective temperature, and a few $\times 0.1$ dex in surface gravity, metallicity and magnesium abundance. Offsets of these magnitudes are perhaps not surprising, given the variety of spectral resolutions, wavelength ranges and analysis techniques employed. We acknowledge that our

Table 5: Zero-point offsets (with respect to APOGEE DR17) inferred for M2FS, Hectochelle and external data sets (Section 4.3).

Sample[a]	N_1	N_2	N_3	N_4	N_5	N_6	N_7	$\overline{\Delta V_{\rm LOS}}$[b]	$\overline{\Delta T_{\rm eff}}$	$\overline{\Delta \log {\rm g}}$	$\overline{\Delta[{\rm Fe/H}]}$	$\overline{\Delta[{\rm Mg/Fe}]}$
M2FS HiRes	—	180	26	1440	115	77	117	0.47 ± 0.05	-141 ± 5	-0.53 ± 0.02	-0.26 ± 0.01	0.19 ± 0.01
M2FS MedRes		—	4	10	0	1	2	0.07 ± 0.44	-323 ± 16	-0.40 ± 0.05	-0.59 ± 0.06	0.09 ± 0.06
Hectochelle			—	194	326	767	94	-0.14 ± 0.05	-221 ± 4	-0.49 ± 0.01	-0.26 ± 0.01	-0.01 ± 0.01
W09				—	···	91	281	-0.19 ± 0.05	···	···	···	···
K10					—	25	75	···	-172 ± 3	-0.26 ± 0.01	-0.18 ± 0.01	0.24 ± 0.02
H3[c]						—	22	-0.31 ± 0.05	-69 ± 4	-0.20 ± 0.01	-0.01 ± 0.01	0.23 ± 0.01

[a] Samples: 1=M2FS HiRes; 2=M2FS MedRes; 3=Hectochelle; 4=Walker, Mateo & Olszewski (2009); 5=Kirby et al. (2010); 6=H3; 7=APOGEE DR17

[b] A value $\overline{\Delta X} \equiv \overline{X - X_7} > 0$ implies a zero point that is more positive than that of the APOGEE catalog.

[c] The 'H3' sample that we use here is from the H3 team's analysis of a subset of ~ 750 spectra from our program.

M2FS+Hectochelle results for individual stars are susceptible to systematic errors at these levels.

In the M2FS (HiRes and MedRes) and Hectochelle catalogs presented below, we subtract from each individual measurement of $V_{\rm LOS}$, $T_{\rm eff}$, $\log g$, [Fe/H] and [Mg/Fe] the zero-point offset listed in Table 5, such that the catalogs are effectively shifted to the APOGEE zero point. Table columns labeled 'X' list values of observable 'X' after shifting to the Apogee zero point. Columns labeled 'X_raw' list the original values—i.e., before applying the zero-point correction.

After applying the zero-point corrections, we compare our current M2FS and Hectochelle catalogs to measurements that we have previously published for subsets of the current samples—including stellar targets in the dwarf galaxies Draco, Reticulum II, Tucana II, Grus I, Crater II, Leo II, Ursa Minor, Hydrus I and Fornax (Walker, Olszewski & Mateo 2015; Walker et al. 2015, 2016; Caldwell et al. 2017; Spencer et al. 2017, 2018; Koposov et al. 2018; Pace et al. 2021). Despite using the same raw M2FS+Hectochelle spectra, the previously-published measurements can differ systematically from current ones even before applying zero-point corrections, as they are derived using an entirely different library of synthetic template spectra. Specifically, the previously-published measurements are based not on the library we introduce in Section 4.1.1, but instead on a library that was designed originally for use with the SDSS Segue Stellar Parameter Pipeline ('SSPP' Lee et al. 2008). The SSPP library is computed over a fixed grid in $T_{\rm eff}$ $\log g$ and [Fe/H], and assumes a monotonic relationship between α-element abundance and [Fe/H]. Experimenting with three independent libraries of synthetic template spectra, Walker, Olszewski & Mateo (2015) observed library-dependent zero-point offsets as large as $\overline{\Delta V_{\rm LOS}} \sim 0.5$ km s^{-1}, $\overline{\Delta T_{\rm eff}} \sim 300$ K, $\overline{\Delta \log {\rm g}} \sim 0.7$ dex and $\overline{\Delta[{\rm Fe/H}]} \sim 0.5$ dex.

The previously published M2FS HiRes, M2FS MedRes and Hectochelle data sets contain 1265, 33 and 3008 sources, respectively, from our current samples. Comparing these measurements directly to the current ones, we find that the previously-published M2FS HiRes (M2FS MedRes) measurements are offset from current (raw, i.e., before applying an offset to the APOGEE zero point) values by $\overline{\Delta V_{\rm LOS}} = -0.47 \pm 0.03$ km s^{-1} (-2.19 ± 0.63 km s^{-1}), $\overline{\Delta T_{\rm eff}} = 168 \pm 3$ K (123 ± 36 K), $\overline{\Delta \log {\rm g}} = 0.45 \pm 0.01$ dex (0.15 ± 0.07 dex) and $\overline{\Delta[{\rm Fe/H}]} = 0.21 \pm 0.01$ dex (0.20 ± 0.07 dex), where positive values imply that the current measurements are, on average, larger than the previously-published ones. The previously-published Hectochelle measurements show offsets of similar magnitude, with $\overline{\Delta V_{\rm LOS}} = 0.68 \pm 0.01$ km s^{-1}, $\overline{\Delta T_{\rm eff}} = -180 \pm 1$ K, $\overline{\Delta \log {\rm g}} = -0.24 \pm 0.00$ dex and $\overline{\Delta[{\rm Fe/H}]} = -0.18 \pm 0.00$ dex. We notice that these offsets with respect to current values are similar to, or smaller than, the zero-point shifts that were applied to raw measurements in the previously-published work (see Walker, Olszewski & Mateo 2015; Walker et al. 2015 for details). Those shifts were determined empirically, based on observed offsets between known solar values and values measured from high-S/N spectra acquired during twilight exposures. Specifically, the previously-published M2FS measurements include zero-point shifts (i.e., quantities that were added to raw measurements) of $\Delta V_{\rm LOS} = 0$ km s^{-1}, $\Delta T_{\rm eff} = -69$ K, $\Delta \log g = -0.09$ dex, $\Delta[{\rm Fe/H}] = +0.20$ dex, while the previously-published Hectochelle measurements include shifts of $\Delta V_{\rm LOS} = -0.81$ km s^{-1}, $\Delta T_{\rm eff} = +303$ K,

$\Delta \log g = +0.63$ dex, Δ[Fe/H] = +0.48 dex. Based on these direct comparisons, then, we find that our switch to the new template library (described in Section 4.1.1), followed by our new zero-point calibration based on external comparisons, results in relatively small offsets from previous values.

Finally, after having applied the zero-point calibration as discussed above, we compare our measurements of [Fe/H] and [Mg/Fe] directly to previously-published abundance measurements derived from high-resolution spectra acquired for relatively small samples of individual stars in dSph galaxies. The external samples come from observations with the HIRES spectrograph at the Keck Telescopes (Shetrone et al. 2001; Fulbright et al. 2004; Cohen & Huang 2009, 2010; Frebel et al. 2010), the High Dispersion Spectrograph at the Subaru Telescope (Sadakane et al. 2004; Aoki et al. 2009), the UVES (Shetrone et al. 2003; Norris et al. 2010; Tafelmeyer et al. 2010; Lucchesi et al. 2020) and X-Shooter (Starkenburg et al. 2013) spectrographs at the Very Large Telescope, and the MIKE spectrograph at Magellan (Simon et al. 2015). Figure 13 displays the comparisons.

Comparing [Fe/H] metallicities (left panel of Figure 13) we find generally good agreement with the high-resolution studies. The bulk of measurements are consistent with a small offset such that our values may be systematically metal-rich by ~ 0.1 dex, **with no significant dependence on additional stellar-atmospheric parameters like $T_{\rm eff}$ or $\log g$.** At the very metal-poor end, however, our measurements for two stars (both in the Sculptor dSph galaxy) with previously-published values [Fe/H] $\lesssim -4$ (Tafelmeyer et al. 2010; Simon et al. 2015) both come in at [Fe/H] $\gtrsim -3$ in our work, disagreeing with the previous measurements at the $\sim 2\sigma$ level. One of these stars, Scl07-50, has been identified (based on the previous measurements) as the most metal-poor star known in an external galaxy (Tafelmeyer et al. 2010). It is potentially concerning that our measurements do not reproduce this result. However, we note that our library of template spectra includes only metallicities [Fe/H] ≥ -4, and that our applied offsets of $\overline{\Delta\mathrm{[Fe/H]}}$ imply that the minimum metallicity that we can in principle measure is [Fe/H] ~ -3.75. The relatively large uncertainties on our measurements of these two stars imply that the mean values will be correspondingly larger than this minimum. We expect, therefore, that users may choose to apply stricter quality-control filters (e.g., a threshold in formal uncertainty) when analyzing chemical abundances, especially when working near the limits of our metallicity scale.

Comparing [Mg/Fe] abundances (right panel of Figure 13, we see what is perhaps the opposite problem, as our template library extends to lower [Mg/Fe] than is allowed in some previous studies. At solar and higher values of [Mg/Fe] we find generally good agreement with the results of previous high-resolution studies. At sub-solar abundance, our measurements of [Mg/Fe] tend to be lower than those previously reported. Again, we expect that users may want to tighten quality-control filters when analyzing chemical abundances; we note that requiring our measurement of [Fe/H] to have uncertainty smaller than 0.5 dex would remove from the comparison sample all but one of the stars for which we measure [Mg/Fe] to be sub-solar.

We perform one additional external cross-check on our metallicity measurements, fitting our spectral models (Section 4) to archival Hectochelle spectra acquired during observations of globular and open star clusters. These observations, performed by other investigators (including the H3 team), used the same spectrograph configuration and processing pipeline that we employ for our own Hectochelle spectra. Figure 14 displays histograms of [Fe/H] that we obtain for each of the clusters M3, M13, M67, M71, M92, M107, which span a range of $-2.2 \lesssim$[Fe/H] $\lesssim 0$ in metallicity. For each cluster we keep only stars for which our measurements have velocity error < 5 km s^{-1}, metallicity error < 0.5 dex, and — in order to reduce contamination from non-member sources — $\log g < 3$ and $V_{\rm LOS}$ within 10 km s^{-1} of the systemic mean tabulated by Harris (1996), except for M67, for which we adopt the spectroscopic mean velocity and metallicity measured by Pace et al. (2008). Figure 14 shows the resulting distributions of [Fe/H] observed toward each cluster, with clear peaks associated with cluster members. We find good agreement with the previously-published mean metallicities, giving confidence that our calibrated zero-point is accurate.

4.4. *Anomalous Sources*

Our target selection filters (Section 2) are designed to isolate primarily red giant stars in the Galactic halo substructures of interest, with contamination contributed mainly by dwarf stars in the Galactic foreground. Our spectral templates are designed to fit individual stars within the limited range of stellar-atmospheric parameters identified in Section 4.1.1, which can accommodate the vast majority of selected targets. Nevertheless, we expect our target selection filters to admit various kinds of anomalous sources for which our templates may provide relatively poor fits—e.g., carbon-enhanced stars, unresolved galaxies and quasars.

In order to identify anomalous sources systematically, first we look for cases where the observed spectrum, $S(\lambda)$, exhibits relatively large residuals with respect to

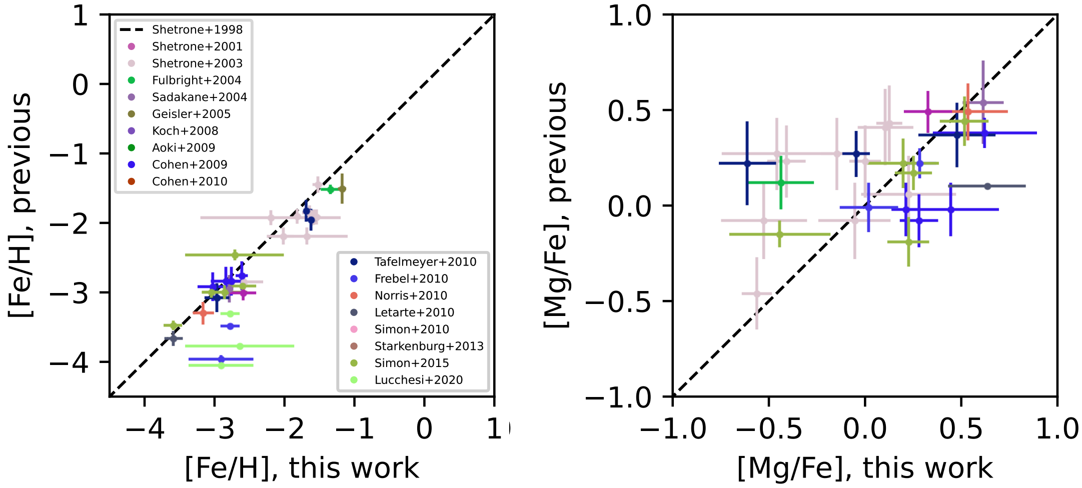

Figure 13: Comparison of current M2FS+Hectochelle measurements of [Fe/H] (left) and [Mg/Fe] (right) to previously-published values derived from high-resolution spectra.

the best-fitting model spectrum, $M(\lambda)$. For each individual spectrum in our M2FS (top) and Hectochelle (bottom) samples, the top two panels of Figure 15 plot the mean value of $\chi^2 \equiv \sum_{i=1}^{N_{\rm pix}}(S_i - M(\lambda_i))^2/{\rm Var}[S_i]$ as a function of the median S/N ratio, where the mean and median are evaluated over all $N_{\rm pix}$ unmasked pixels. The variance spectrum, Var$[S]$, is the original one, uncorrected by the linear re-scaling parameters inferred as part of the spectral fit (see Equation 15), as the re-scaled variance will be inflated to compensate for template mismatch. For both M2FS and Hectochelle, we find that the mean value of χ^2 is approximately constant at median S/N$\lesssim$ 10, with characteristic values of $\chi^2/$pix ~ 1.0 for M2FS and $\chi^2/$pix ~ 1.5 for Hectochelle, suggesting that the uncertainties in pixel counts estimated by the Hectochelle pipeline tend to be under-estimated by $\sim 20\%$. We reiterate that, by design, our linear re-scaling of the raw variances (Equation 15) brings the typical values to $\chi^2/$pix ~ 1.

Figure 15 also reveals that mean χ^2 values rise steadily at S/N ratios $\gtrsim$ 10. One contribution to this behavior comes from the fact that our polynomial model for the continuum spectrum is fixed at order $l = 5$ (Section 4), limiting ability to fit details of the continuum structure that become apparent only at high S/N. In order to flag anomalous spectra despite the steady rise in χ^2 with S/N ratio, we identify outliers above the smooth S/N-dependent curves drawn in both panels of Figure 15. The curves are broken power laws of the form $\chi^2/$pix $= a_1(1 + (S/N)/a_2)^3$, with $(a_1, a_2) = (1.2, 25)$ for M2FS and $(4.0, 75)$ for Hectochelle. For all anomalous spectra identified in this way, we set the flag chi2_flag=True in the data catalogs (Section 5). We identify 60 such anomalous M2FS HiRes spectra, 114 anomalous M2FS MedRes spectra and 131 anomalous Hectochelle spectra having median S/N ratio ≥ 1 per pixel. For sources having at least one observation that passed our quality-control filter, we set the flag 'any_chi2_flag=True' if the spectrum from any of the individual accepted observations has chi2_flag=True. There are 44 such sources in our M2FS HiRes catalog (not necessarily the same as those that have S/N$\geq$ 1), 28 in our M2FS MedRes catalog and 41 in our Hectochelle catalog.

Figure 16 displays representative examples of these anomalous spectra, some types of which have already been identified and discussed in previous M2FS papers by Walker, Olszewski & Mateo (2015); Song et al. (2019, 2021). The top two M2FS spectra (left-hand panels) and the top Hectochelle spectrum (right-hand panels) are from stars showing various levels of carbon enhancement, with the Swan (1857) C_2 bandhead clearly visible near 5165 Å. The second (from top) Hectochelle spectrum is dominated by emission lines, presumably from a distant star-forming galaxy; a few tens of similar spectra are among the χ^2 outliers in our Hectochelle sample but not, due to our masking of strong emission-like features (Section 3.7), in our M2FS sample. The third (from top) row of spectra are from cool M dwarf stars, with the TiO bandhead visible near 5170 Å. The bottom row of spectra are from known quasars, previously measured

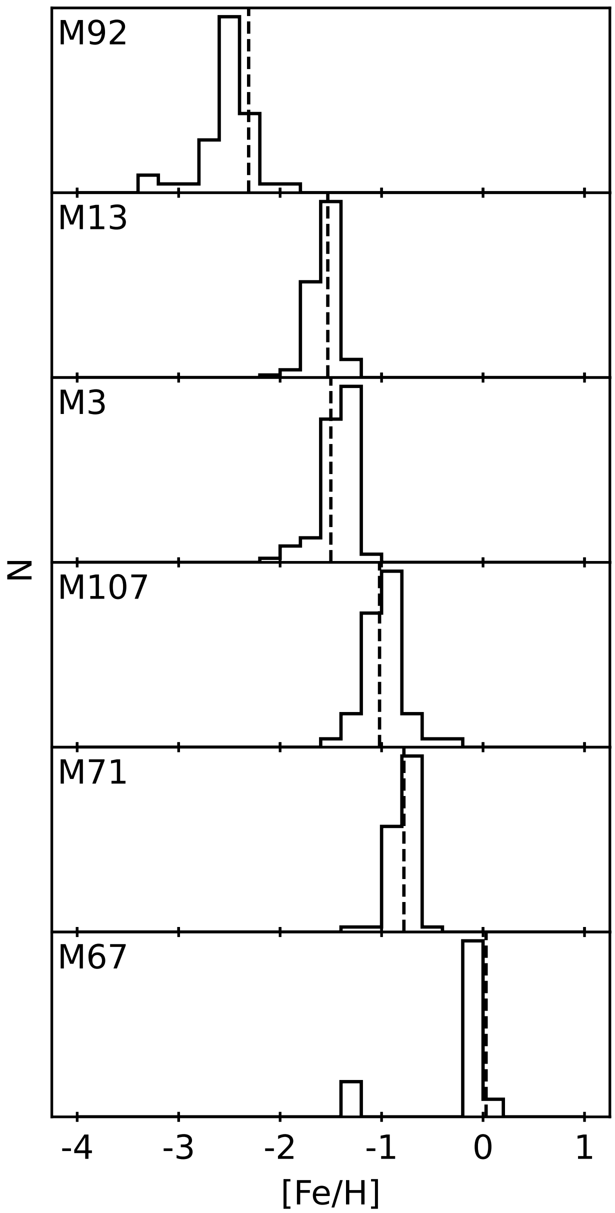

Figure 14: Histograms of metallicities we infer from archival Hectochelle observations of targets in the star clusters (top to bottom) M92, M13, M3, M107, M71 and M67. In each panel, the the dashed vertical line indicates the metallicity tabulated by Harris (1996), except for the metallicity of M67, which we adopt from Pace et al. (2008).

to have redshifts of $z \sim 3.7$ (Boutsia et al. 2021, left) and $z \sim 3.4$ (Pâris et al. 2014).

Following Song et al. (2021), we obtain a cleaner sample of carbon stars by comparing the median flux across the bandpass 5160–5167 Å, denoted W_{5163} to the me-

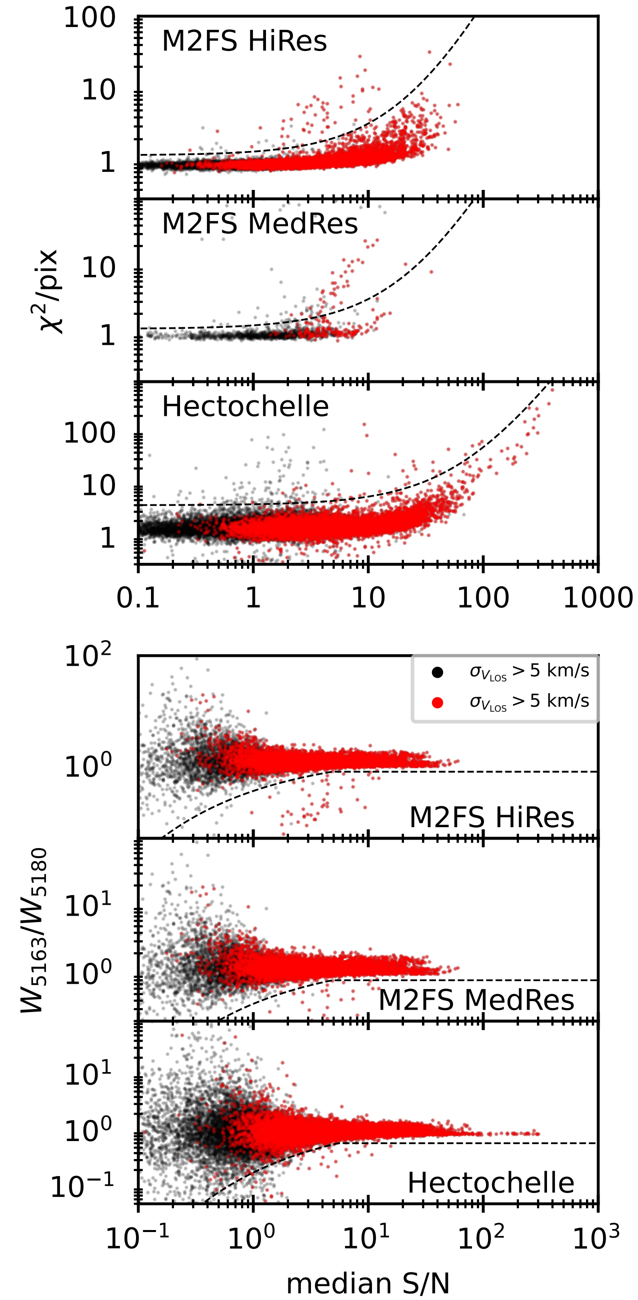

Figure 15: *Top three panels:* χ^2 per pixel vs median S/N ratio per pixel, from the best-fitting model for each individual spectrum obtained with M2FS HiRes, M2FS MedRes and Hectochelle. Red points identify observations that pass our crude quality-control filter, with raw velocity error ≤ 5 km s^{-1}. Outliers having χ^2/pix above the dashed curves tend to correspond to anomalous sources, primarily carbon stars, background galaxies and quasars. *Bottom three panels:* Ratio of median flux in the $5160 - 5167$ Å bandpass to the median flux in the $5176 - 5183$ Å bandpass, vs. median S/N ratio. Outliers having flux ratios below the dashed curves are flagged in our data catalogs as likely carbon stars.

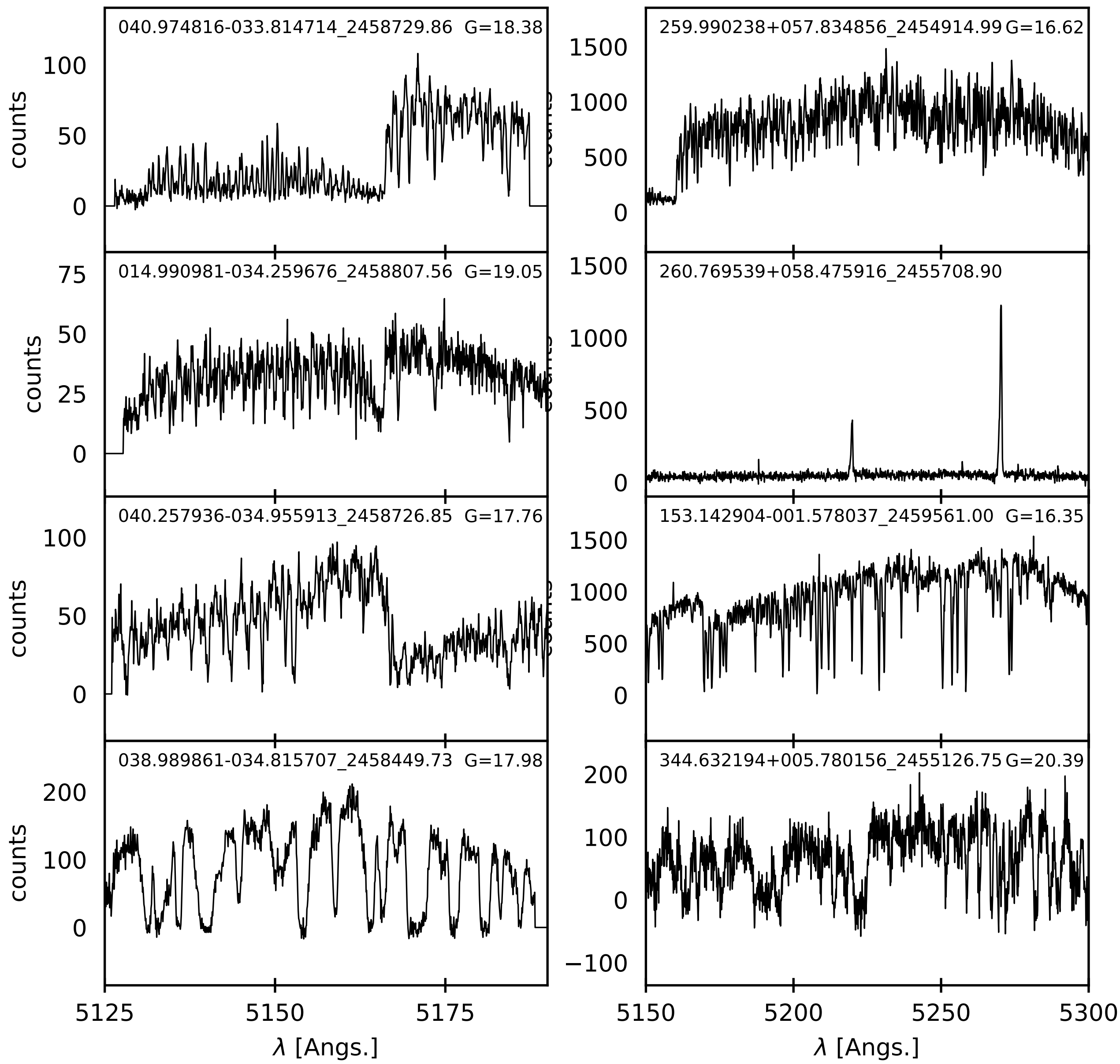

Figure 16: Examples of M2FS HiRes (left) and Hectochelle (right) spectra from anomalous sources, with text indicating celestial coordinates, HJD of observation and Gaia G-band magnitude (if available). The top two M2FS spectra, and the top Hectochelle spectrum, come from stars showing various levels of carbon enhancement, with the prominent Swan (1857) C_2 bandhead near 5165 Å. The second (from top) Hectochelle spectrum is dominated by emission lines from an extragalactic source. Spectra in the third row are from cool M giant stars, with the TiO bandhead apparent near 5170 Å. Spectra in the bottom row are from known quasars, at redshift $z \sim 3.7$ (Boutsia et al. 2021, left) and $z \sim 3.4$ (Pâris et al. 2014, right).

dian flux across 5176–5183 Å, denoted W_{5180}. The bottom two panels of Figure 15 plot the ratio W_{5163}/W_{5180} as a function of median S/N ratio. We identify as candidate carbon stars those sources for which the flux ratio falls below the curves drawn in the bottom two panels of Figure 15; spectra that satisfy this criterion have flag carbon_flag=True in the data catalogs (Section 5). The M2FS HiRes sample contains 37 sources that have at least one spectrum that is flagged as carbon enhanced and has S/N≥ 1; the M2FS MedRes sampled contains 1 such sources and the Hectochelle sample contains 144 such sources. For sources having at least one observation that passed our quality-control filter, we set the flag 'any_carbon_flag'=True if the spectrum from any of the individual accepted observations has carbon_flag=True. There are 37 such sources in our M2FS HiRes catalog, 0 in our MedRes catalog and 88 in our Hectochelle catalog.

Our samples also contain sources that the Gaia (DR3) database flags as photometrically variable ('phot_variable_flag='VARIABLE') in the main source catalog, and/or lists in dedicated variability tables for active galactic nuclei (variability table 'vari_agn') or RR Lyrae ('vari_rrlyrae'). Our spectroscopic catalogs list for each source the value of Gaia's phot_variable_flag, and also sets flags gaia_agn=True, gaia_rrl=True if the source appears in the corresponding variability tables. Considering only those having at least one spectrum with S/N≥ 1, our M2FS HiRes, M2FS MedRes and Hectochelle samples contain 551, 3 and 764 sources, respectively, that Gaia flags as photometric variables in the main source catalog, with 75, 1 and 363 sources appearing in Gaia's dedicated AGN table. For all but 3, 0 and 6 of these sources, our M2FS HiRes, M2FS MedRes and Hectochelle observations do not yield measurements that pass our quality-control criteria.

Finally, considering only those sources having at least one M2FS HiRes, M2FS MedRes or Hectochelle observation that passed our quality control filter, 292, 0 and 40, respectively, are listed in Gaia's dedicated RR Lyrae table. While we can obtain good fits to the spectra of RR Lyrae, our repeat measurements detect the intrinsic line-of-sight velocity variability of these pulsating stars. For each of our sources that have multiple spectroscopic measurements that pass our quality-control filter, histograms in Figure 17 show distributions of the ratio of the weighted standard deviation (about the weighted mean) of the measured $V_{\rm LOS}$, $T_{\rm eff}$, $\log g$, [Fe/H]and [Mg/Fe] to the weighted mean error. This ratio is a measure of intrinsic variability of the source. Red (blue) histograms represent sources that are (are not) listed in Gaia's (DR3) RR Lyrae variability table (vari_rrlyrae). The ratios for RRL stars generally track those of the non-RRLs for the atmospheric parameters $T_{\rm eff}$, $\log g$, [Fe/H] and [Mg/Fe]. For $V_{\rm LOS}$, however (leftmost panel of Figure 17), the RRLs exhibit dramatically larger scatter than the non-RRLs, directly reflecting the rates at which the pulsating stars expand and contract. Users who are interested in the observed stars as dynamical tracers will need to take into account this source of intrinsic velocity variability.

Of course, there are sources of intrinsic variability other than pulsation—e.g., binary star systems—for which we do not necessarily have a diagnostic classification *a priori*. For all stars having multiple independent measurements that pass our quality-control filter, we identify sources exhibiting potentially intrinsic variability as those for which the ratio of weighted standard deviation to weighted mean error exceeds a value of 3, regardless of whether the source is classified as RRL. In our data catalogs (Section 5), we set the flag 'X_variable_flag'=True for such cases, where X can be any of the observables 'vlos', 'teff', 'logg', 'feh', 'mgfe'.

5. M2FS+HECTOCHELLE DATASET

We provide complete data catalogs for our M2FS HiRes, M2FS MedRes and Hectochelle samples. The catalogs are stored as binary tables in standard '.fits' format, and are available electronically at both the Journal website and the Zenodo database (DOI: 10.5281/zenodo.7837922). Table 6 lists and briefly explains each of the columns listed in these catalogs. Most users will need to be mindful of the 'obs' and/or 'good_obs' columns, which indicate for a given star the chronologically-ordered observation number. A star having only one observation will have 'obs=1', but for stars observed multiple times, the first observation will have 'obs=1', the second will have 'obs=2', etc. The 'good_obs' parameter works the same way, but counts only those observations that pass our crude quality-control filter (velocity error $\sigma_{V_{\rm LOS}} \leq 5$ km s^{-1}); all measurements for stars having zero 'good' measurements will have good_obs=0. This information can be used in tandem with the (inverse variance-weighted) mean parameter estimates that are computed over all 'good' observations of a given star, and listed for each individual-epoch measurement ('good' or otherwise) of the star. So, for example, a user who wants only the mean parameter estimates for each star (as opposed to individual-epoch measurements) can select the mean values (e.g., vlos_mean, teff_mean, logg_mean, feh_mean, mgfe_mean, with associated errors vlos_mean_error, teff_mean_error, logg_mean_error, feh_mean_error, mgfe_mean_error) listed for only observations with good_obs=1.

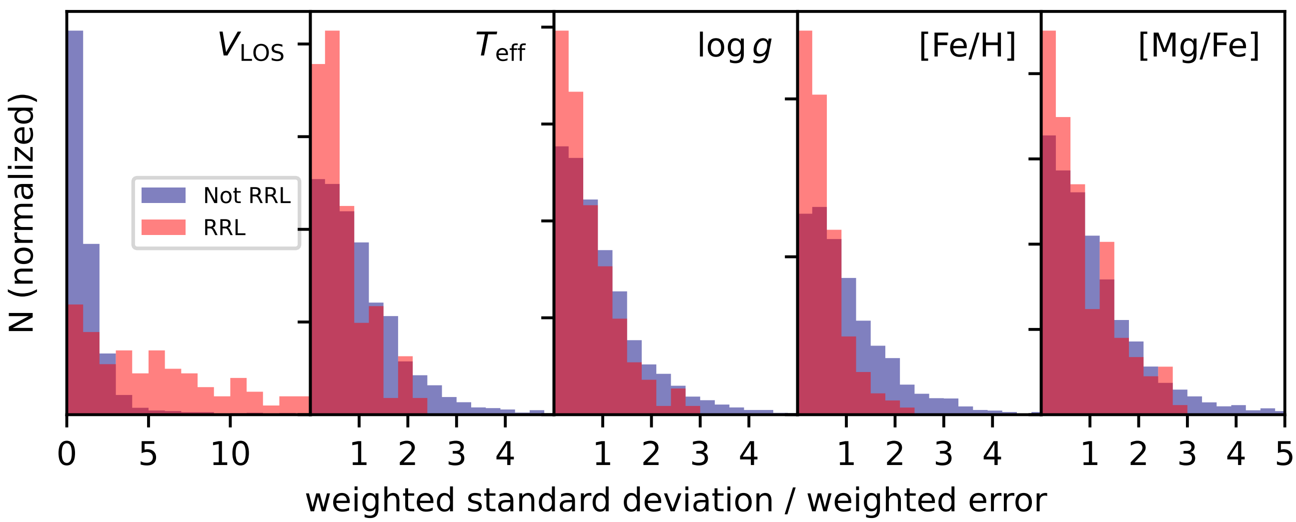

Figure 17: Distributions of the ratio of weighted standard deviation to weighted mean uncertainty of spectroscopically-measured parameters, for stars having multiple independent measurements passing our quality-control filter. Red (resp. blue) histograms correspond to sources that do (do not) appear in Gaia's (DR3) RR Lyrae catalog. The left-most panel demonstrates the intrinsic variability of V_{LOS} for RRL stars.

The Zenodo database (DOI: 10.5281/zenodo.7837922) also makes available all of the individual (extracted, 1D, wavelength-calibrated) spectra produced by our processing pipeline. The spectra are provided in multi-extension .fits files. A given file contains all (up to 128 for M2FS, up to 240 for Hectochelle) spectra obtained on a given data frame. In the .fits catalogs discussed above, the 'fits_filename' and 'fits_index' columns specify the filename and array index where the processed spectrum can be found. Along with the spectra, these multi-extension fits files provide the central wavelength, variance, best-fitting model, mean sky level, and (bad pixel) mask status at each pixel.

Table 6. Columns in electronic data catalogs

column name	description
instrument	Instrument used to acquire spectrum ('Hectochelle', 'M2FS_HiRes' or 'M2FS_MedRes')
target_system	Name of target system (name of dwarf galaxy, star cluster, etc.)
obs_id	unique identifier for this observation (R.A._Dec_HJD)
exptime	exposure time (s)
gaia_source_id	source ID in Gaia (DR3) catalog, if available
gaia_gmag	Gaia (DR3) G magnitude, if available
gaia_bpmag	Gaia (DR3) BP magnitude, if available
gaia_rpmag	Gaia (DR3) RP magnitude, if available
gaia_siggmag	Gaia (DR3) error in gaia_gmag
gaia_sigbpmag	Gaia (DR3) error in gaia_bpmag
gaia_sigrpmag	Gaia (DR3) error in gaia_rpmag
gaia_gmag_dered	Gaia (DR3) G magnitude, de-reddened
gaia_bpmag_dered	Gaia (DR3) BP magnitude, de-reddened
gaia_rpmag_dered	Gaia (DR3) RP magnitude, de-reddened
gaia_pmra	Gaia (DR3) proper motion, right ascension component, if available (mas yr^{-1})
gaia_pmdec	Gaia (DR3) proper motion, declination component, if available (mas yr^{-1})
gaia_sigpmra	Gaia (DR3) error in gaia_pmra (mas yr^{-1})
gaia_sigpmdec	Gaia (DR3) error in gaia_pmdec (mas yr^{-1})
gaia_parallax	Gaia (DR3) parallax, if available (mas)
gaia_sigparallax	Gaia (DR3) error in gaia_parallax (mas)
ra	Right Ascension (J2000)
dec	Declination (J2000)
ra_dec_source	source catalog from which ra_deg and dec_deg are adopted (Gaia DR3 if available)
hjd	Heliocentric Julian Date of spectroscopic observation (days)
sn_ratio	median signal-to-noise ratio per pixel
vlos_raw	mean of posterior PDF for V_{LOS} (km s^{-1}; solar rest frame), without shift to APOGEE zero point
vlos_raw_error	standard deviation of posterior PDF for V_{LOS} (km s^{-1}), as sampled by MultiNest
vlos_raw_skew	skewness of posterior PDF for V_{LOS}, as sampled by MultiNest
vlos_raw_kurtosis	kurtosis of posterior PDF for V_{LOS}, as sampled by MultiNest
vlos	vlos_raw, shifted to APOGEE zero point
vlos_error	error in vlos_raw and vlos (km s^{-1}), after applying adjustment in Section 4.2
teff_raw	mean of posterior PDF for T_{eff} (K), without shift to APOGEE zero point
teff_raw_error	standard deviation of posterior PDF for T_{eff} (K), as sampled by MultiNest

Table 6 *continued*

Table 6 *(continued)*

column name	description
teff_raw_skew	skewness of posterior PDF for $T_{\rm eff}$, as sampled by MultiNest
teff_raw_kurtosis	kurtosis of posterior PDF for $T_{\rm eff}$, as sampled by MultiNest
teff	teff_raw, shifted to APOGEE zero point
teff_error	error in teff_raw and teff (K), after applying adjustment in Section 4.2
logg_raw	mean of posterior PDF for $\log g$(cgs units), without shift to APOGEE zero point
logg_raw_error	standard deviation of posterior PDF for $\log g$ (cgs units), as sampled by MultiNest
logg_raw_skew	skewness of posterior PDF for $\log g$, as sampled by MultiNest
logg_raw_kurtosis	kurtosis of posterior PDF for $\log g$, as sampled by MultiNest
logg	logg_raw, shifted to APOGEE zero point
logg_error	error in logg_raw and logg, after applying adjustment in Section 4.2 (cgs units)
feh_raw	mean of posterior PDF for [Fe/H] without shift to APOGEE zero point
feh_raw_error	standard deviation of posterior PDF for [Fe/H], as sampled by MultiNest
feh_raw_skew	skewness of posterior PDF for [Fe/H], as sampled by MultiNest
feh_raw_kurtosis	kurtosis of posterior PDF for [Fe/H], as sampled by MultiNest
feh	feh_raw, shifted to APOGEE zero point
feh_error	error in feh_raw and feh, after applying adjustment in Section 4.2
mgfe_raw	mean of posterior PDF for [Mg/Fe] without shift to APOGEE zero point
mgfe_raw_error	standard deviation of posterior PDF for [Mg/Fe], as sampled by MultiNest
mgfe_raw_skew	skewness of posterior PDF for [Mg/Fe], as sampled by MultiNest
mgfe_raw_kurtosis	kurtosis of posterior PDF for [Mg/Fe], as sampled by MultiNest
mgfe	mgfe_raw, shifted to APOGEE zero point
mgfe_error	error in mgfe_raw and mgfe, after applying adjustment in Section 4.2
smooth_raw	bandwidth $\sigma_{\rm LSF}$ (Angstroms), of Gaussian smoothing kernel applied to template spectra
smooth_raw_error	standard deviation of posterior PDF for $\sigma_{\rm LSF}$ (Angstroms), as sampled by MultiNest
smooth_raw_skew	skewness of posterior PDF for $\sigma_{\rm LSF}$, as sampled by MultiNest
smooth_raw_kurtosis	kurtosis of posterior PDF for $\sigma_{\rm LSF}$, as sampled by MultiNest
logs1_raw	base-10 logarithm of error re-scaling parameter s_1 (Equation 15)
logs1_raw_error	standard deviation of posterior PDF for $\log_{10} s_1$, as sampled by MultiNest
logs1_raw_skew	skewness of posterior PDF for $\log_{10} s_1$, as sampled by MultiNest
logs1_raw_kurtosis	kurtosis of posterior PDF for $\log_{10} s_1$, as sampled by MultiNest
logs2_raw	base-10 logarithm of error floor parameter s_2 (Equation 15)
logs2_raw_error	standard deviation of posterior PDF for $\log_{10} s_2$, as sampled by MultiNest
logs2_raw_skew	skewness of posterior PDF for $\log_{10} s_2$, as sampled by MultiNest
logs2_raw_kurtosis	kurtosis of posterior PDF for $\log_{10} s_2$, as sampled by MultiNest
median_sky	median count of sky spectrum that was subtracted
standard_deviation_median_sky	standard deviation of median_sky, over spectra acquired in same observation
filter_name	name of filter used for observation
chi2	χ^2 for best-fitting model spectrum, using original variance spectrum
chi2_rescaled	χ^2 for best-fitting model spectrum, using re-scaled variance spectrum from Equation 15
npix	number of (unmasked) pixels included in spectrum fit
w5163	median (sky-subtracted) counts over spectral range $5160-5167$ Å

Table 6 *continued*

Table 6 *(continued)*

column name	description
w5180	median (sky-subtracted) counts over spectral range 5176 − 5183 Å
vhelio_correction	heliocentric correction that was applied (added) to $V_{\rm LOS}$ after spectrum model fitting (km s^{-1})
fits_filename	name of multi-extension fits file containing processed spectrum
fits_index	index containing the spectrum of this source (in multi-extension fits frame)
obs	(chronological) observation number for this source
n_obs	total number of observations of this source
good_obs	(chronological) observation number for this source, after quality control filter
good_n_obs	total number of observations of this source, after quality control filter
vlos_raw_mean	(inverse-variance) weighted mean of vlos_raw (km s^{-1}; solar rest frame) over good_n_obs observations
vlos_mean	vlos_raw_mean, shifted to APOGEE zero point (km s^{-1})
vlos_mean_error	error in vlos_raw_mean and vlos_mean (km s^{-1})
vlos_mean_scatter	(inverse-variance) weighted standard deviation of $V_{\rm LOS}$ (km s^{-1}) over good_n_obs observations
teff_raw_mean	(inverse-variance) weighted mean of teff_raw (K) over good_n_obs observations
teff_mean	teff_raw_mean, shifted to APOGEE zero point (K)
teff_mean_error	error in teff_raw_mean and teff_mean (K)
teff_mean_scatter	(inverse-variance) weighted standard deviation of $T_{\rm eff}$ (K) over good_n_obs observations
logg_raw_mean	(inverse-variance) weighted mean of logg_raw over good_n_obs observations
logg_mean	logg_raw_mean, shifted to APOGEE zero point
logg_mean_error	error in logg_raw_mean and logg_mean
logg_mean_scatter	(inverse-variance) weighted standard deviation of $\log g$ over good_n_obs observations
feh_raw_mean	(inverse-variance) weighted mean of feh_raw over good_n_obs observations
feh_mean	feh_raw_mean, shifted to APOGEE zero point
feh_mean_error	error in feh_raw_mean and feh_mean
feh_mean_scatter	(inverse-variance) weighted standard deviation of [Fe/H] over good_n_obs observations
mgfe_raw_mean	(inverse-variance) weighted mean of mgfe_raw over good_n_obs observations
mgfe_mean	mgfe_raw_mean, shifted to APOGEE zero point
mgfe_mean_error	error in mgfe_raw_mean and mgfe_mean
mgfe_mean_scatter	(inverse-variance) weighted standard deviation of [Mg/Fe] over good_n_obs observations
n_wav_cal	(M2FS only) number of ThArNe calibration frames used for wavelength calibration
temp_min	(M2FS only) minimum temperature (°C) recorded at detector during science sub-exposures
temp_max	(M2FS only) maximum temperature (°C) recorded at detector during science exposures
wav_cal_flag	(M2FS only) True if n_wav_cal=1 and temp_max−temp_min ≥ 1 °C
chi2_flag	True if chi2 is above curve in top panels of Figure 15
carbon_flag	True if flux ratio W_{5163}/W_{5180} is below curve in bottom panels of Figure 15
any_chi2_flag	True if any observations contributing to mean have chi2_flag=True
any_carbon_flag	True if any observations contributing to mean have carbon_flag=True
vlos_variable_flag	True if vlos_mean_scatter ≥ 3 vlos_mean_error
teff_variable_flag	True if teff_mean_scatter ≥ 3 teff_mean_error
logg_variable_flag	True if logg_mean_scatter ≥ 3 logg_mean_error
feh_variable_flag	True if feh_mean_scatter ≥ 3 feh_mean_error
mgfe_variable_flag	True if mgfe_mean_scatter ≥ 3 mgfe_mean_error

Table 6 *continued*

Table 6 *(continued)*

column name	description
gaia_phot_variable_flag	Gaia (DR3) phot_variable_flag
gaia_rrl	True if source is listed in Gaia DR3 variability RR Lyrae table (vari_rrlyrae)
gaia_agn	True if source is listed in Gaia DR3 variability AGN catalog (vari_agn)

We now present some of the macroscopic properties of the M2FS+Hectochelle dataset. Figure 18 provides a comprehensive view of chemo-dynamical structure within the Galactic Halo, plotting metallicity against line-of-sight velocity for the entire sample (using inverse-variance-weighted mean values for stars with multiple good measurements), with marker color coded according to surface gravity. Red giants within dwarf galaxies are conspicuous as bluer ($\log g \lesssim 3$) points that tend to have lower mean metallicity ([Fe/H] $\lesssim -1.5$) and cluster into narrower velocity distributions (velocity dispersion $\lesssim 10$ km s^{-1}) than do foreground stars, which tend to be late-type dwarfs ($\log g \gtrsim 4$) contributed by the Galactic disk. Visually dominating population of substructures traced by red giants are the classical dwarf spheroidals Ursa Minor ($V_{\rm LOS} \sim -250$ km s^{-1}), Draco ($V_{\rm LOS} \sim -290$ km s^{-1}), Fornax ($V_{\rm LOS} \sim +55$ km s^{-1}), Leo II ($V_{\rm LOS} \sim +80$ km s^{-1}), Sculptor ($V_{\rm LOS} \sim +110$ km s^{-1}), Carina/Sextans (both at $V_{\rm LOS} \sim 220$ km s^{-1}) and Leo I ($V_{\rm LOS} \sim +280$ km s^{-1}). Many less luminous Halo substructures are present in our sample, but are less obvious against the foreground populations. Figures 22 and 23 display the [Fe/H] vs $V_{\rm LOS}$ scatterplots for individual systems.

Figure 19 plots [Mg/Fe] against [Fe/H] for our M2FS+Hectochelle sample. For clarity, we display only the 8189 stars for which observational errors in $\log g$, [Fe/H] and [Mg/Fe] are all ≤ 0.5 dex. The red giant sample (bluer points), dominated by Halo substructures, is clearly offset toward lower metallicity than the foreground Galactic stellar populations. Also apparent, although blurred somewhat by the inclusion of all targeted systems simultaneously, is the characteristic 'knee' (near [Fe/H] ~ -2), where [Mg/Fe] declines toward higher metallicities because stars have formed from gas pre-enriched by Type-Ia supernovae.

Figure 20 plots surface gravity against effective temperature, with marker color indicating [Fe/H]. Again, for clarity, we display only stars for which errors in $\log g$, [Fe/H] and [Mg/Fe] are all ≤ 0.5 dex. Overplotted are MESA isochrones (Morton 2015; Dotter 2016), calculated for age = 10 Gyr and a range of stellar metallicity. Reassuringly, low-gravity stars within our sample clearly populate the red giant branch expected for low-metallicity stars ($-3 \lesssim$ [Fe/H] $\lesssim -1$). Higher-gravity stars populate regions near the main sequence expected for the higher-metallicity stars contributed by the Galactic foreground.

We note the presence in Figures 18 and 19 of ~ 10 sources that are measured to have extremely low metallicity ([Fe/H]$\lesssim -3.6$), high surface gravity ($\log g \gtrsim 4.5$) and approximately solar [Mg/Fe]. Figure 21 of the Appendix displays spectra from each of these sources, with best-fitting models overplotted. We find that most of these spectra exhibit the broad absorption features characteristic of AGN, suggesting that our measurements for these sources are spurious. However, none of the sources are listed in Gaia's AGN variability table.

We do not attempt here to evaluate the population (e.g., dwarf galaxy vs Galactic foreground) membership status of individual stars within our sample. The reason is that a star's probability of membership to a specific population, given the star's observed properties, depends fundamentally on the model invoked to describe the ensemble of populations. We hope and anticipate that our dataset will be used to evaluate a large variety of models. Therefore we leave to the user any evaluation of membership status for individual stars.

Instead we use our spectroscopic measurements to give rough indications of the mixtures of stellar populations that are present within our samples. As is evident in Figure 20, our measurements of surface gravity can effectively distinguish red giants from dwarf stars. While red giant status correlates strongly with membership within most of the dwarf galaxies and Halo substructures targeted by our program, systems at distances $\lesssim 50$ kpc can have observed targets on the sub-giant branch at $\log g \gtrsim 3$. Moreover, we expect red giant samples to include contamination from bona fide red giants within the Galactic halo. Thus the number of observed red giants is a useful but imperfect proxy for the number of observed member stars within the targeted systems.

Therefore, in order to summarize the contents of our spectroscopic samples, we count not just the number of red giant sources, but also the number of sources that have both $V_{\rm LOS}$ and proper motion consistent with mem-

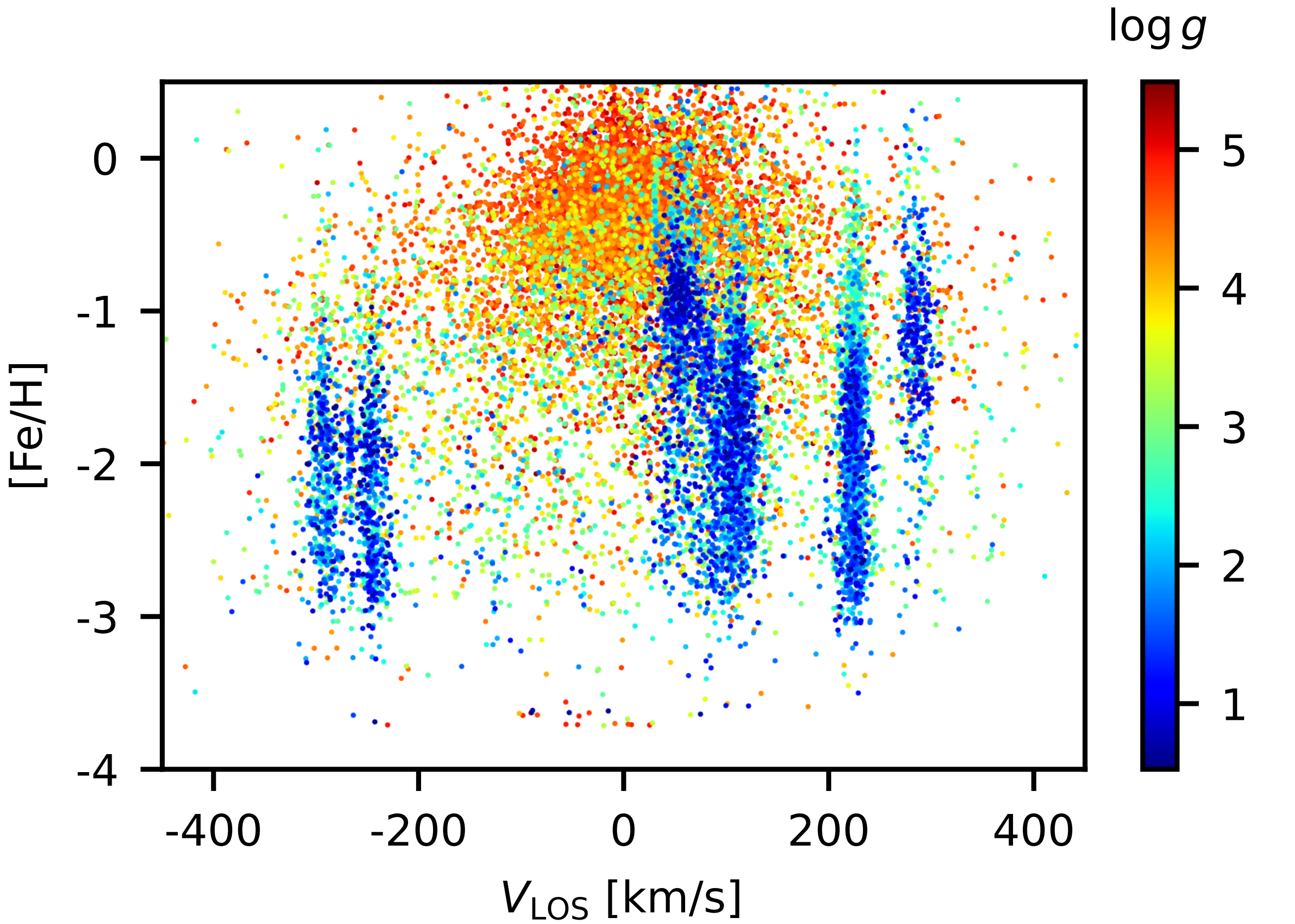

Figure 18: Chemo-dynamic substructure within the Milky Way Halo: Metallicity vs. line-of-sight velocity, from Magellan/M2FS and MMT/Hectochelle spectra acquired for 16369 stars observed toward 38 Galactic Halo objects. Marker color indicates spectroscopically-estimated surface gravity. Given our color/magnitude criteria for spectroscopic target selection, redder marker colors tend to identify dwarf stars in the Galactic disk, while bluer marker colors indicate giant stars in the Galactic halo and its substructures. The halo objects that are most obvious here are the 'classical' dwarf spheroidal galaxies Draco ($V_{\rm LOS} \sim -290$ km s^{-1}), Ursa Minor (-250 km s^{-1}), Fornax ($+55$ km s^{-1}), Leo II ($+80$ km s^{-1}), Sculptor ($+110$ km s^{-1}), Carina/Sextans (both at $+220$ km s^{-1}) and Leo I ($+280$ km s^{-1}).

bership (regardless of $\log g$). Text in Figures 22 and 23 lists numbers of individual sources observed (denoted $N_{\rm obs}$) with at least one 'good' measurement that passes our crude quality-control filter, the number of likely giant stars (denoted $N_{\rm giant}$), identified as sources measured to have $\log g \lesssim 3$, and the number of sources that have $V_{\rm LOS}$ and Gaia=measured (DR3) proper motion to be within 3σ of the previously-measured systemic mean values (denoted $N_{\rm mem}$). For the $V_{\rm LOS}$ criterion, we define σ to be quadrature sum of formal uncertainties in our measurement of $V_{\rm LOS}$ for the source, the measurement of the systemic mean velocity, and the (previously-measured) systemic velocity dispersion. We take the previously-published mean values from the compilation by Pace et al. (2022). For the proper motion criterion, we take σ to be the propagated uncertainty in the separation (in 2D proper motion space, neglecting covariance between the components) between the source and previously-measured systemic mean proper motions (Pace et al. 2022).

Our samples contain several hundred members in each of the Milky Way's eight 'classical' dSph satellites, ranging from ~ 200 in Leo II to ~ 850 in Carina and Sculptor. In the less luminous satellites and star clusters, likely member samples range from zero to a few tens. These samples extend to larger galactocentric radii than most previously-published counterparts. We count 823 (124, 64, 42) likely members projected farther than 2 (3, 4, 5) projected halflight radii from the center of their host galaxy, providing information about the stellar pop-

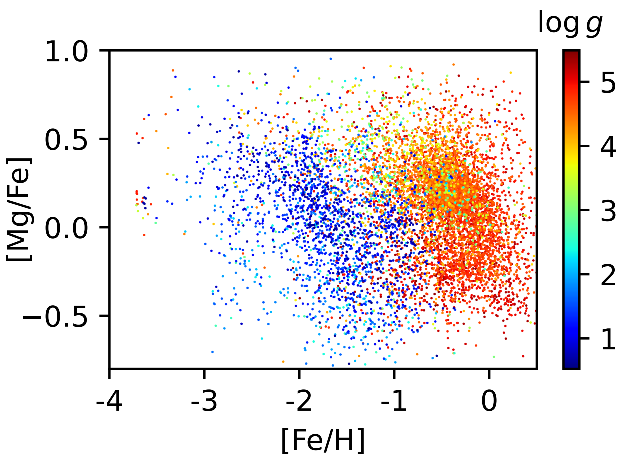

Figure 19: Magnesium abudance vs. metallicity, for the 8189 stars in our M2FS+Hectochelle dataset that have observational errors ≤ 0.5 in each of $\log g$, [Fe/H] and [Mg/Fe]. Marker color indicates spectroscopically-estimated surface gravity.

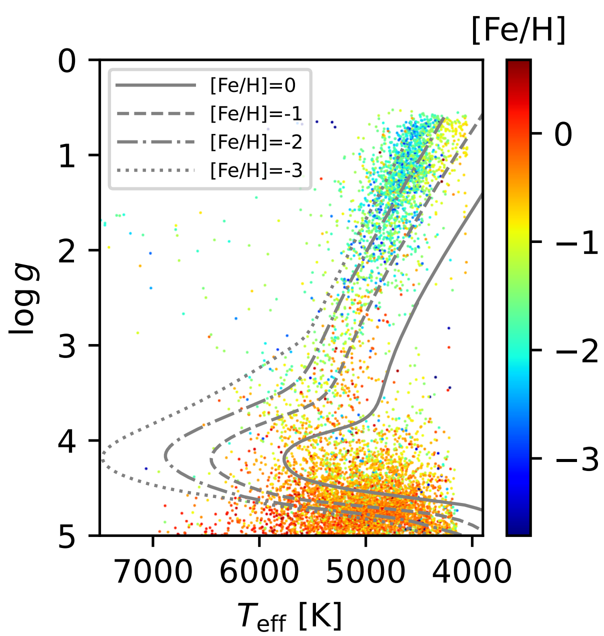

Figure 20: Surface gravity vs. effective temperature (both estimated spectroscopically), for the 8189 stars in our M2FS+Hectochelle dataset that have observational errors ≤ 0.5 in each of $\log g$, [Fe/H] and [Mg/Fe]. Marker color indicates spectroscopically-estimated metallicity. Overplotted, for comparison, are theoretical isochrones (Morton 2015; Dotter 2016) calculated for age = 10 Gyr and a range of [Fe/H].

ulations and dynamical state in the outer parts of these systems.

6. SUMMARY

We have presented new spectroscopic data and catalogs of new measurements of spectroscopic parameters for 16369 unique sources toward 38 target systems. The sample includes repeat (multi-epoch) measurements for 3720 sources, with as many as 15 epochs per source. We have calibrated internal errors and used external data sets to calibrate zero points for each physical parameter. We have defined criteria for identifying anomalous sources that should be handled carefully in subsequent analysis. Using simple but crude diagnostic criteria, we estimate that the sample includes $\sim$ 6078 red giant stars and $\sim$ 4494 members of the target systems, in some many cases pushing the available samples beyond several halflight radii. Data products include catalogs of measured stellar parameters and all processed and calibrated spectra.

ACKNOWLEDGEMENTS

We thank the anonymous referee, whose feedback and suggestions improved the manuscript. We thank David Nidever and Paul Cristofari for providing helpful tests and feedback about the M2FS processing pipeline. We thank Charlie Conroy, Phill Cargile and the H3 team for providing comparison measurements using methodology from the H3 survey, and for providing spectra from H3's observations of globular clusters.

E.O. wants to remember Jill Bechtold here.

M.G.W. acknowledges support from National Science Foundation (NSF) grants AST-1813881, AST-1909584 and AST-2206046. M.M. acknowledges support from NSF grants AST-0923160, AST-1312997, AST-1815403 and AST-2205847. E.O. acknowledges support from NSF grants AST-1815767, AST-1313006, AST-0807498. Nelson Caldwell acknowledges support from NSF grant AST-1812461. I.U.R. acknowledges support from NSF grants AST-1815403, AST-2205847, and PHYS-1430152 (Physics Frontier Center/JINA-CEE). A.B.P. acknowledges support from NSF grant AST-1813881.

This work has made use of NASA's Astrophysics Data System Bibliographic Services. This paper made use of the Whole Sky Database (WSDB), created by Sergey Koposov and maintained at the Institute of Astronomy Cambridge by Sergey Koposov, Vasily Belokurov and Wyn Evans with financial support from the Science & Technology Facilities Council (STFC) and the European Research Council (ERC). This work made use of Astropy:[9] a community-developed core Python package and an ecosystem of tools and resources for astronomy.

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www. cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

Figures 5-7 use atomic line identifications from the Virtual Atomic and Molecular Data Centre (VAMDC) Consortium (Dubernet et al. 2016), provided by the BASS2000 website.

This work has made use of data from the Sloan Digital Sky Survey IV. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions.

SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is www.sdss4.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics — Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

This project used public archival data from the Dark Energy Survey (DES). Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology FacilitiesCouncil of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the OzDES Membership Consortium, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

This project has made use of public data from the Pan-STARRS1 survey. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astron-

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

omy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

APPENDIX

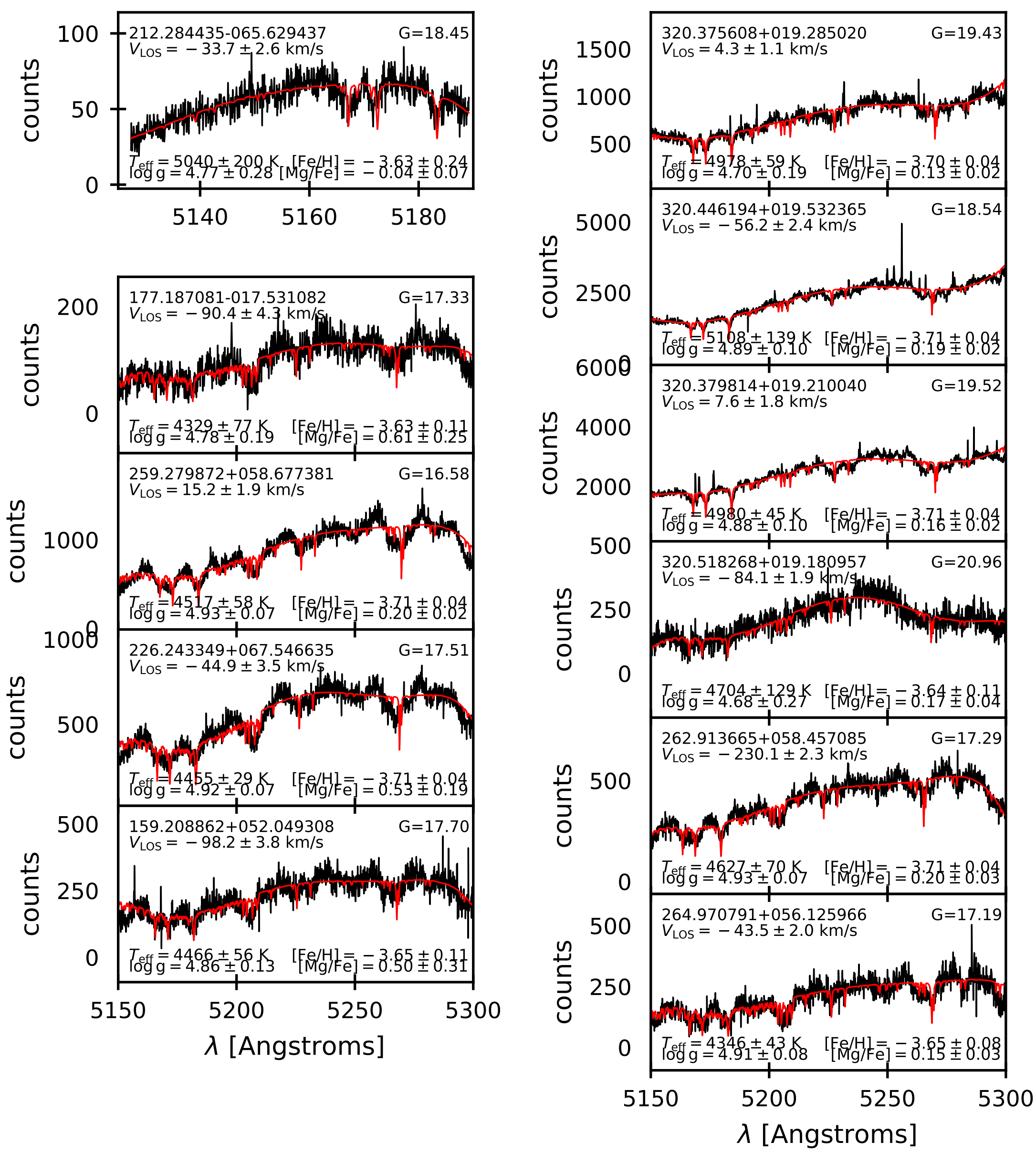

Figure 21: Examples of M2FS (top left) and Hectochelle (all other panels) spectra corresponding to spurious measurements of extremely low metallicity ([Fe/H] $\lesssim -3.6$), high surface gravity ($\log g \gtrsim 4.5$) and alpha-enhanced [Mg/Fe]. Over-plotted in red are best-fitting models, which tend to find absorption features but fail to reproduce their broadness. Text indicates target coordinates, Gaia G-band magnitude, and values of spectroscopically-inferred parameters.

REFERENCES

2106 Aaronson, M. 1983, ApJL, 266, L11

2107 Abbott, T. M. C., Adamów, M., Aguena, M., et al. 2021,
2108 ApJS, 255, 20

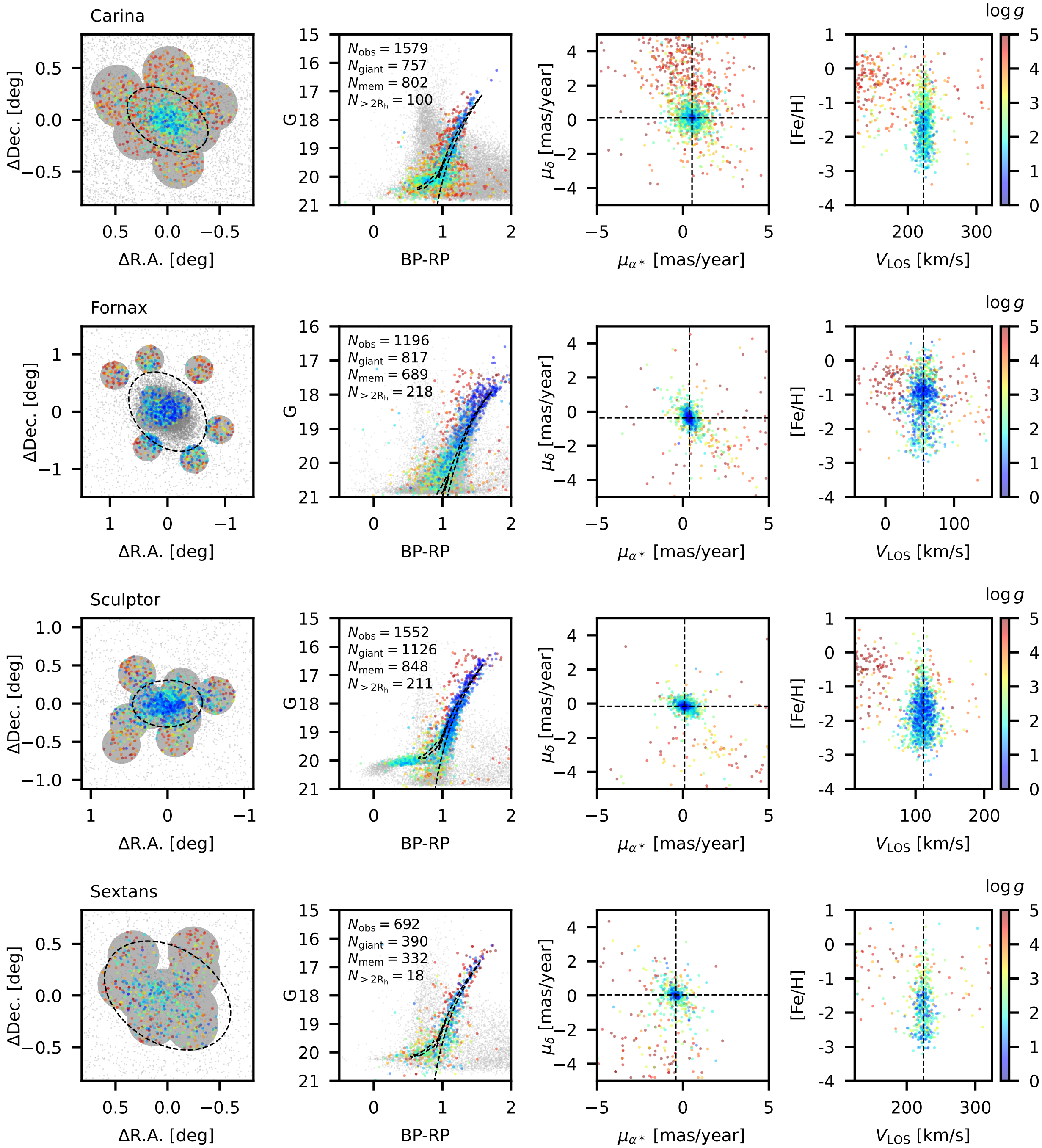

Figure 22: Sky maps, color-magnitude diagrams (CMDs; middle), proper motion coordinates and spectroscopic [Fe/H] vs $V_{\rm LOS}$, from *Gaia* photometry/astrometry and Magellan/M2FS spectroscopy of point sources toward Galactic satellites. Colored points indicate sources for which we report spectroscopic measurements, with bluer colors identifying likely red giant stars belonging to the satellites. In sky maps, dashed ellipses have semi-major axis $a = 2R_{\rm half}/\sqrt{1-\epsilon}$, where $R_{\rm half}$ is the projected halflight radius and $\epsilon \equiv 1 - b/a$ is the measured ellipticity, both adopted from the compilation by Pace et al. (2022). In CMDs, gray points indicate unobserved point sources within 1° of the satellite center; in sky maps, gray points indicate unobserved sources within $\delta = \max(0.15, \sqrt{\sigma_{\rm G}^2 + \sigma_{\rm BP}^2 + \sigma_{\rm RP}^2})$ magnitudes of the theoretical isochrone (Morton 2015; Dotter 2016) overplotted in the corresponding CMD (chosen for typical age =10 Gyr and according to previously published mean metallicity). Dashed lines indicate previously measured mean systemic proper motions and line-of-sight velocities.

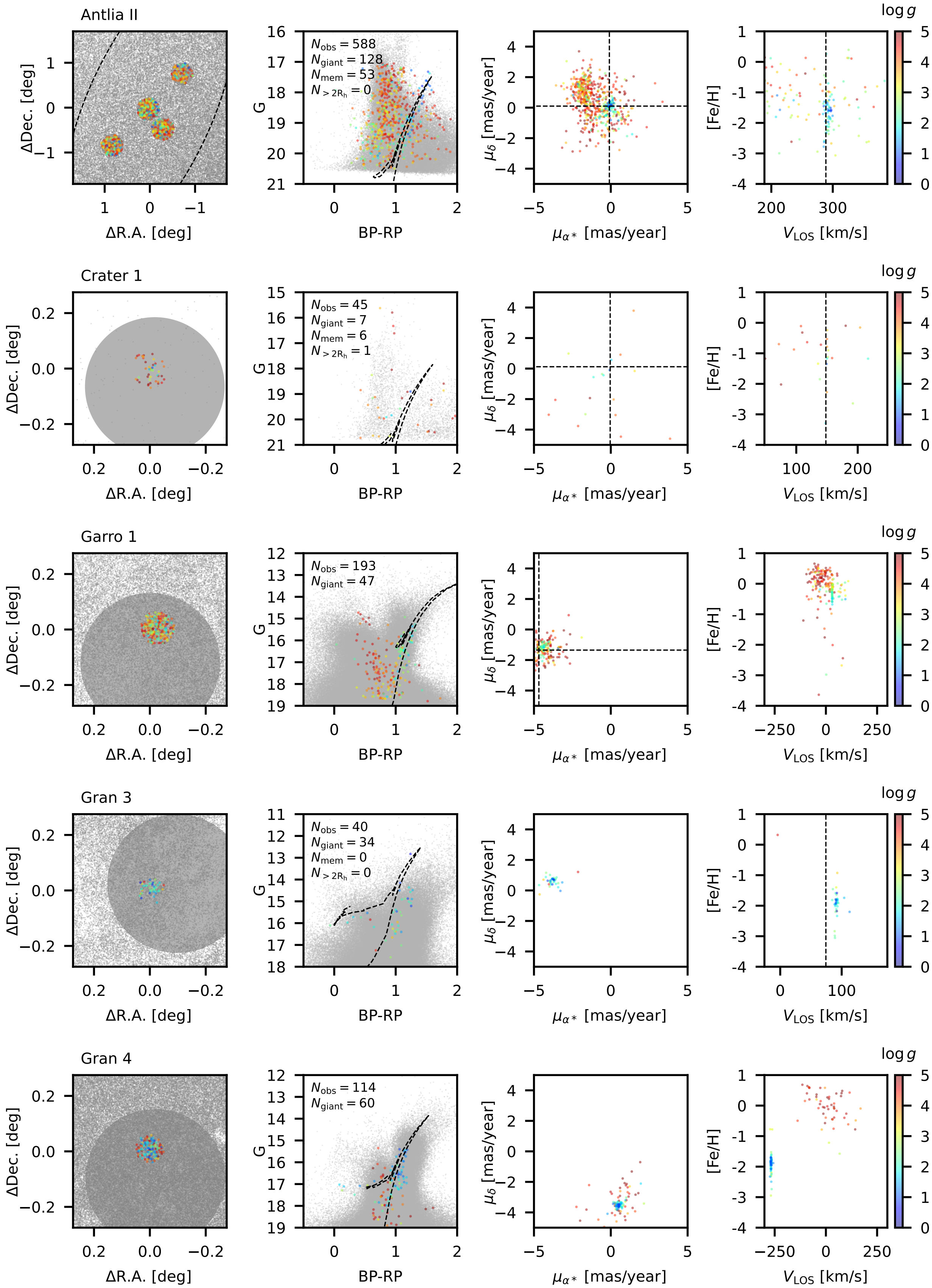

Figure 22 (Continued) :

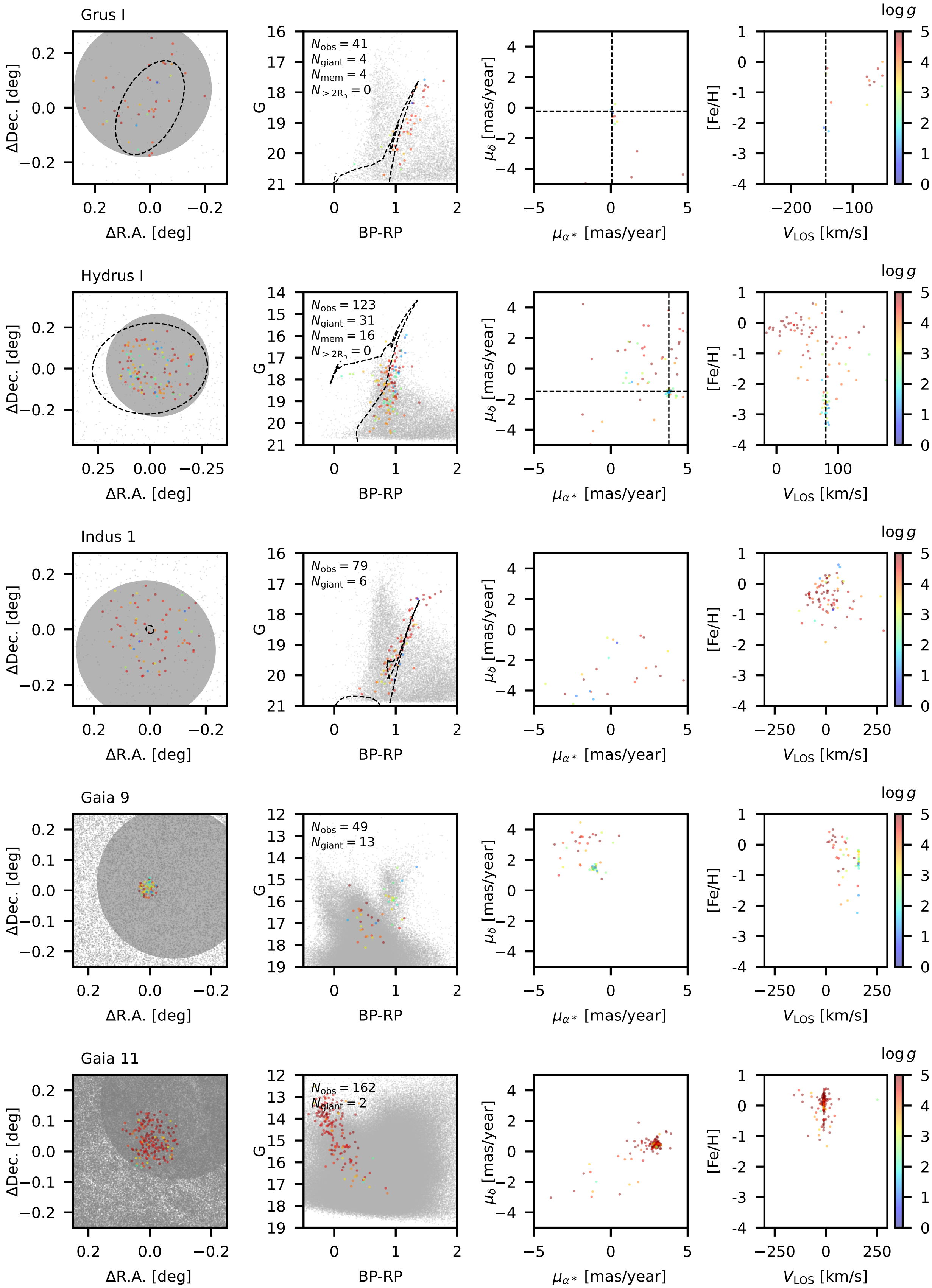

Figure 22 (Continued) :

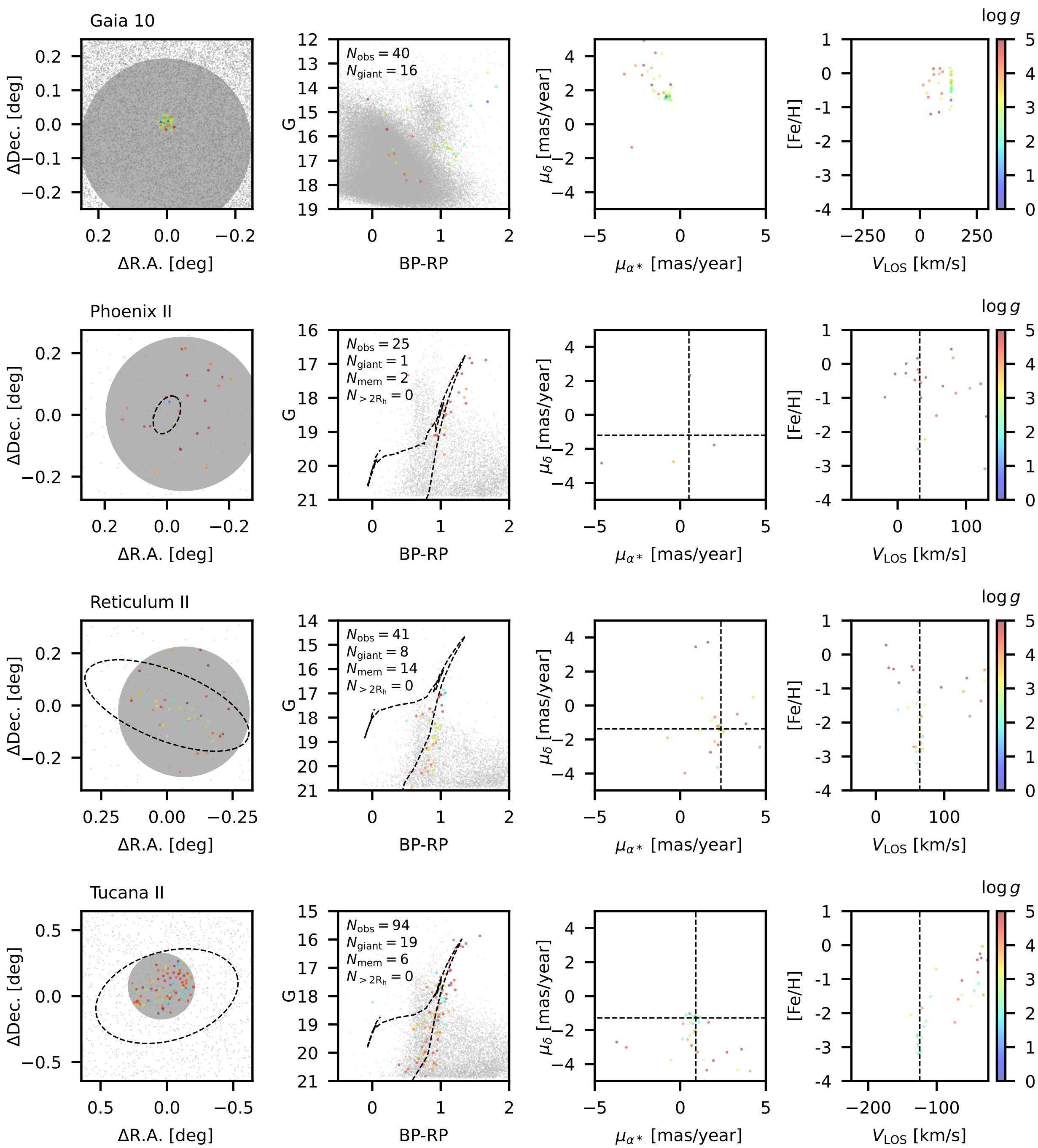

Figure 22 (Continued) :

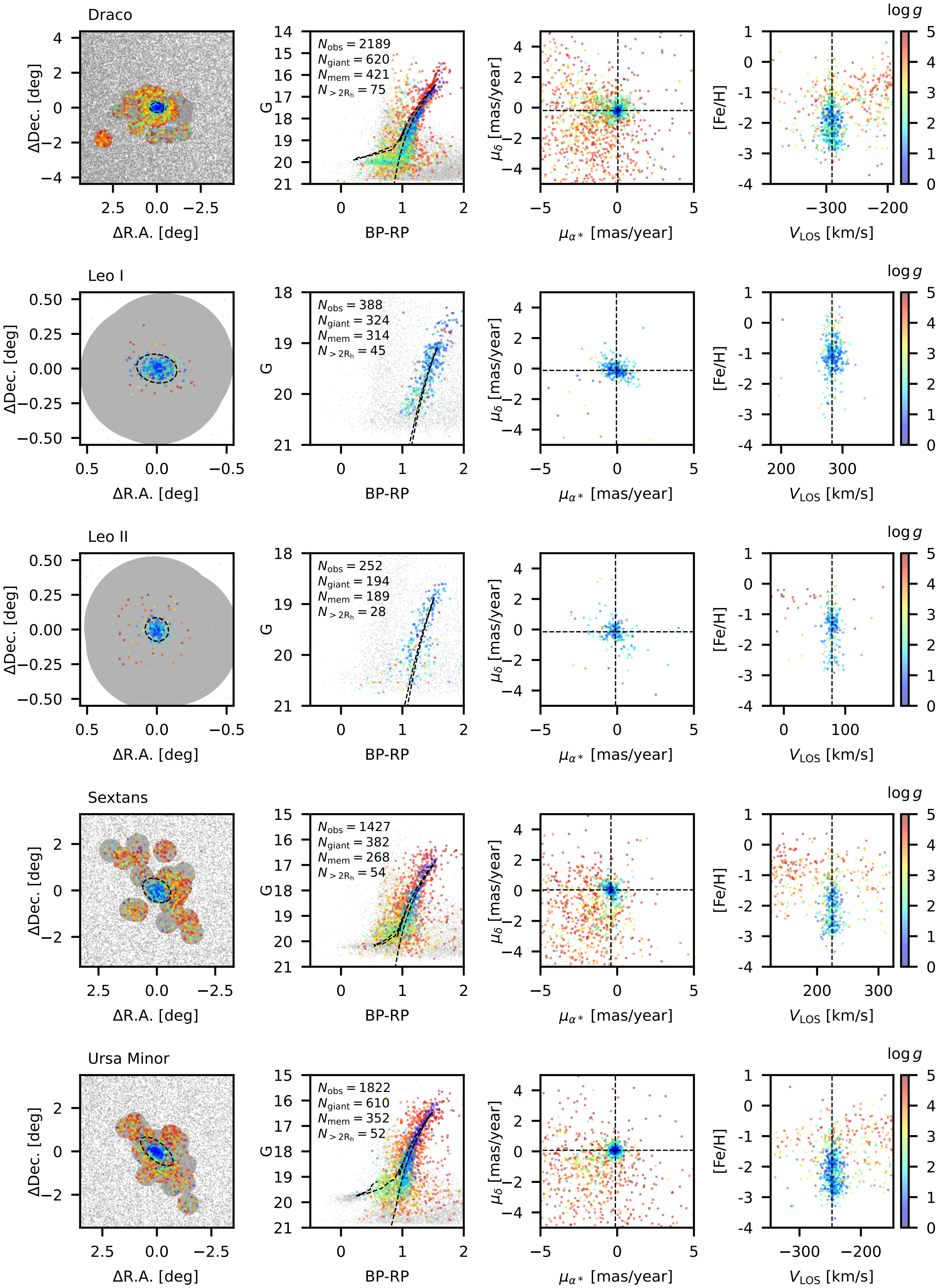

Figure 23: Same as Figure 22, but for Galactic satellites observed with MMT/Hectochelle.

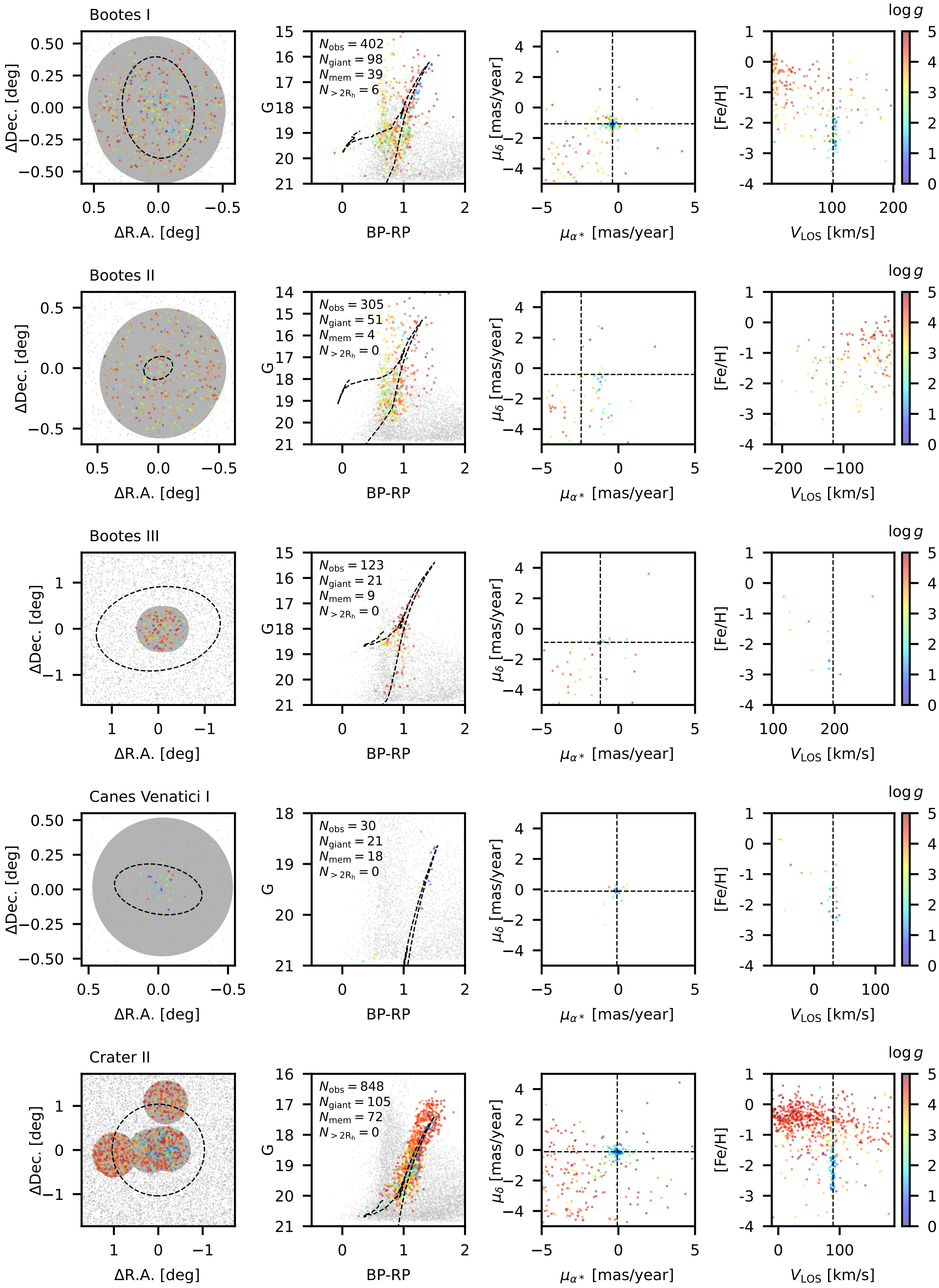

Figure 23 (Continued) :

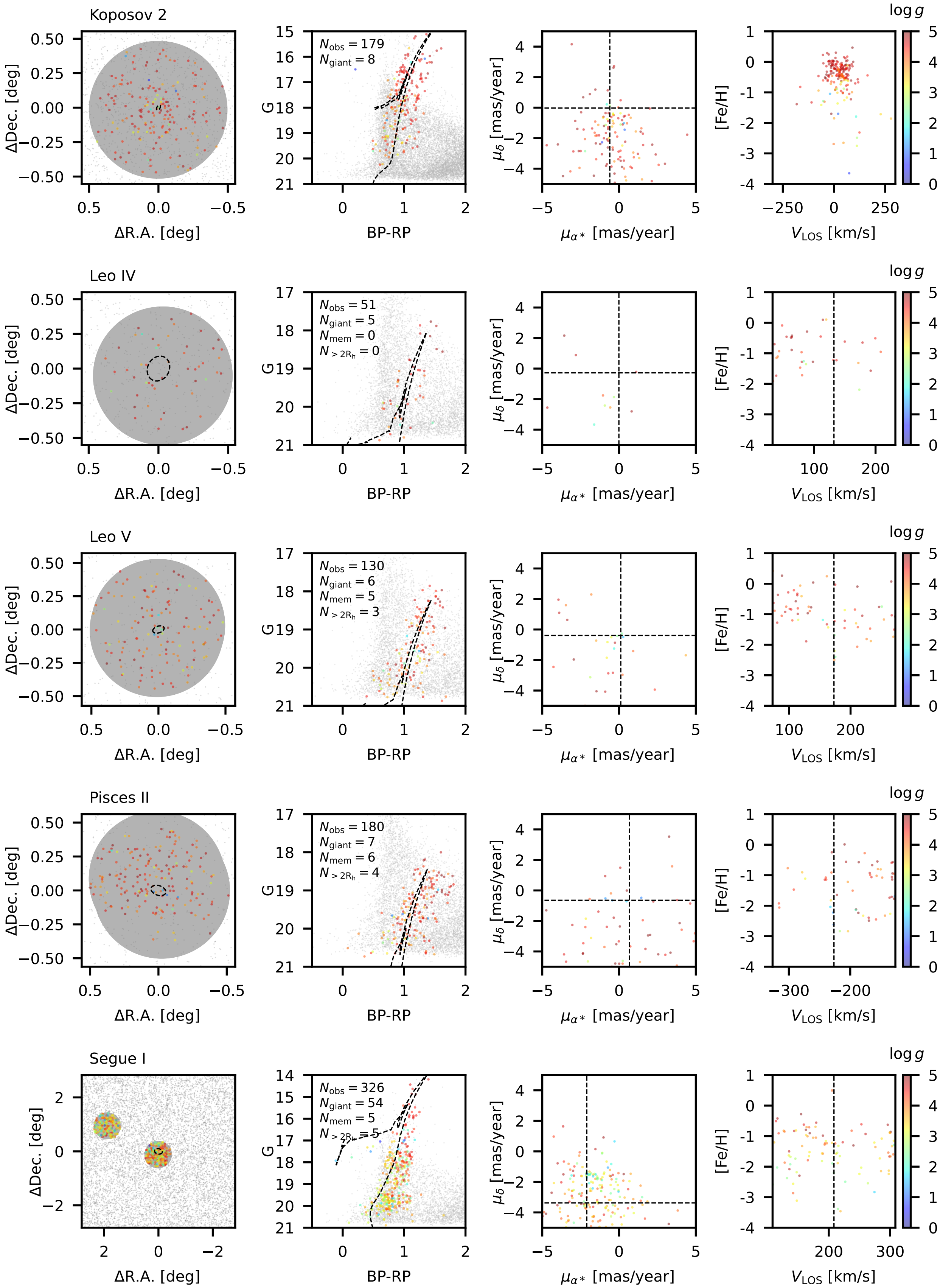

Figure 23 (Continued) :

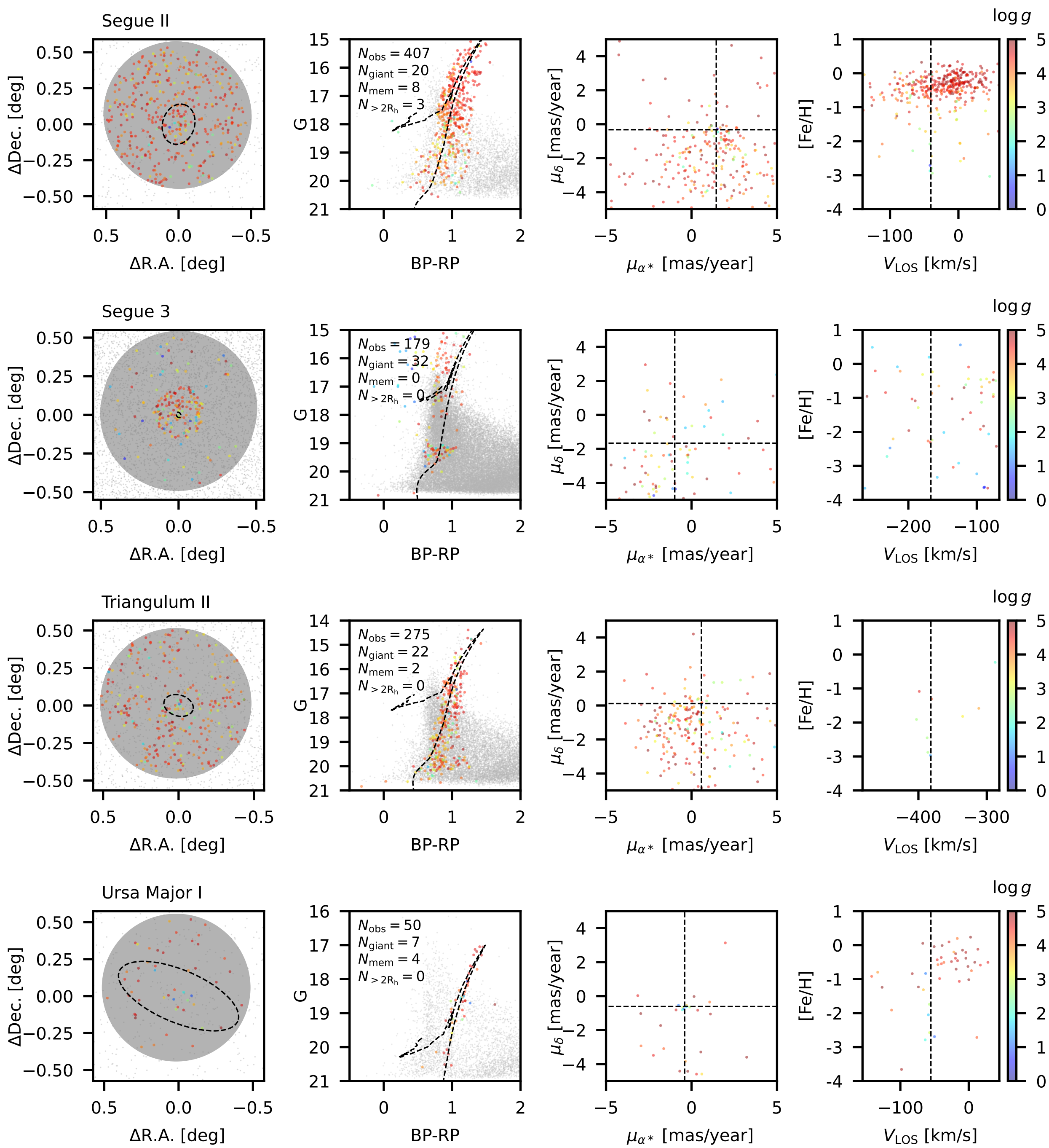

Figure 23 (Continued) :

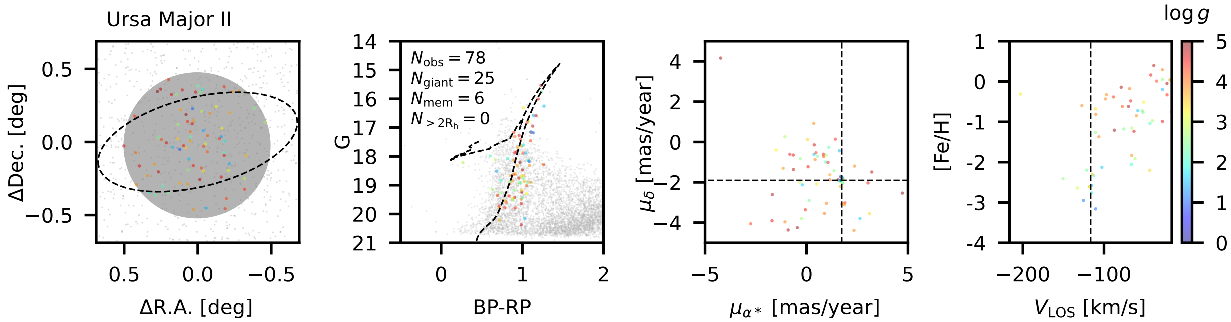

Figure 23 (Continued) :

Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21

Aoki, W., Arimoto, N., Sadakane, K., et al. 2009, A&A, 502, 569

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

Battaglia, G., & Nipoti, C. 2022, Nature Astronomy, 6, 659

Battaglia et al. 2006, A&A, 459, 423

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

Belokurov, V., & Evans, N. W. 2022, Nature Astronomy, 6, 911

Boutsia, K., Grazian, A., Fontanot, F., et al. 2021, ApJ, 912, 111

Buttry, R., Pace, A. B., Koposov, S. E., et al. 2022, MNRAS, 514, 1706

Caldwell, N., Walker, M. G., Mateo, M., et al. 2017, ApJ, 839, 20

Cargile, P. A., Conroy, C., Johnson, B. D., et al. 2020, ApJ, 900, 28

Castelli, F., & Kurucz, R. L. 2004, ArXiv e-prints, arXiv:astro-ph/0405087

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

—. 2010, ApJ, 719, 931

Conroy, C., Bonaca, A., Cargile, P., et al. 2019, ApJ, 883, 107

Craig, M., Crawford, S., Seifert, M., et al. 2017, astropy/ccdproc: v1.3.0.post1, doi:10.5281/zenodo.1069648

Dotter, A. 2016, ApJS, 222, 8

Dubernet, M. L., Antony, B. K., Ba, Y. A., et al. 2016, Journal of Physics B: Atomic, Molecular and Optical Physics, 49, 074003

Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601

Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, ApJS, 251, 7

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

Fulbright, J. P., Rich, R. M., & Castro, S. 2004, ApJ, 612, 447

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1

Gaia Collaboration, Helmi, A., van Leeuwen, F., et al. 2018a, A&A, 616, A12

Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018b, A&A, 616, A10

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

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2022, arXiv e-prints, arXiv:2208.00211

Gilmore, G., Wilkinson, M. I., Wyse, R. F. G., et al. 2007, ApJ, 663, 948

Guy, J., Bailey, S., Kremin, A., et al. 2022, arXiv e-prints, arXiv:2209.14482

Hargreaves, J. C., Gilmore, G., Irwin, M. J., & Carter, D. 1994a, MNRAS, 269, 957

—. 1994b, MNRAS, 271, 693

—. 1996, MNRAS, 282, 305

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

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

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

Hinkle, K., Wallace, L., Valenti, J., & Harmer, D. 2000, Visible and Near Infrared Atlas of the Arcturus Spectrum 3727-9300 A

Hinkle, K. H., Wallace, L., Ram, R. S., et al. 2013, ApJS, 207, 26
Holweger, H., & Müller, E. A. 1974, SoPh, 39, 19
Horne, K. 1986, PASP, 98, 609
Kirby, E. N., Guhathakurta, P., Simon, J. D., et al. 2010, ApJS, 191, 352
Kleyna, J., Wilkinson, M. I., Evans, N. W., Gilmore, G., & Frayn, C. 2002, MNRAS, 330, 792
Kleyna, J. T., Wilkinson, M. I., Evans, N. W., & Gilmore, G. 2005, ApJL, 630, L141
Koch, A., Grebel, E. K., Wyse, R. F. G., et al. 2006, AJ, 131, 895
Koch, A., Kleyna, J. T., Wilkinson, M. I., et al. 2007a, AJ, 134, 566
Koch, A., Wilkinson, M. I., Kleyna, J. T., et al. 2007b, ApJ, 657, 241
Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, A&A, 501, 1269
Koposov, S. E., Gilmore, G., Walker, M. G., et al. 2011, ApJ, 736, 146
Koposov, S. E., Walker, M. G., Belokurov, V., et al. 2018, MNRAS, 479, 5343
Kurucz, R. L. 2011, Canadian Journal of Physics, 89, 417
Kurucz, R. L., Furenlid, I., Brault, J., & Testerman, L. 1984, Solar flux atlas from 296 to 1300 nm
Lawler, J. E., Sneden, C., Cowan, J. J., Den Hartog, E. A., & Wood, M. P. 2017, Canadian Journal of Physics, 95, 783
Lawler, J. E., Sneden, C., Cowan, J. J., Ivans, I. I., & Den Hartog, E. A. 2009, ApJS, 182, 51
Lee et al. 2008, AJ, 136, 2050
Letarte, B., Chapman, S. C., Collins, M., et al. 2009, ArXiv e-prints, arXiv:0901.0820
Li, T. S., Koposov, S. E., Zucker, D. B., et al. 2019, MNRAS, 490, 3508
Lucchesi, R., Lardo, C., Primas, F., et al. 2020, A&A, 644, A75
Martin, N. F., de Jong, J. T. A., & Rix, H.-W. 2008, ApJ, 684, 1075
Martin, N. F., Ibata, R. A., Chapman, S. C., Irwin, M., & Lewis, G. F. 2007, MNRAS, 380, 281
Masseron, T., Plez, B., Van Eck, S., et al. 2014, A&A, 571, A47
Mateo, M., Bailey, J. I., Crane, J., 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, 84464Y
Mateo, M., Olszewski, E., Welch, D. L., Fischer, P., & Kunkel, W. 1991, AJ, 102, 914
Mateo, M., Olszewski, E. W., Pryor, C., Welch, D. L., & Fischer, P. 1993, AJ, 105, 510
Mateo, M., Olszewski, E. W., & Walker, M. G. 2008, ApJ, 675, 201
Mateo, M. L. 1998, ARA&A, 36, 435
McConnachie, A. W. 2012, AJ, 144, 4
McConnachie, A. W., Irwin, M. J., Ibata, R. A., et al. 2009, Nature, 461, 66
Morton, T. D. 2015, isochrones: Stellar model grid package, Astrophysics Source Code Library, record ascl:1503.010, ascl:1503.010
Muñoz, R. R., Carlin, J. L., Frinchaboy, P. M., et al. 2006, ApJL, 650, L51
Naidu, R. P., Conroy, C., Bonaca, A., et al. 2020, ApJ, 901, 48
Norris, J. E., Yong, D., Gilmore, G., & Wyse, R. F. G. 2010, ApJ, 711, 350
Olszewski, E. W., & Aaronson, M. 1985, AJ, 90, 2221
Olszewski, E. W., Aaronson, M., & Hill, J. M. 1995, AJ, 110, 2120
Pace, A. B., Erkal, D., & Li, T. S. 2022, ApJ, 940, 136
Pace, A. B., & Li, T. S. 2019, ApJ, 875, 77
Pace, A. B., Walker, M. G., Koposov, S. E., et al. 2021, ApJ, 923, 77
Pace, A. B., Koposov, S. E., Walker, M. G., et al. 2023, arXiv e-prints, arXiv:2304.06904
Pace, G., Pasquini, L., & François, P. 2008, A&A, 489, 403
Palmer, B. A., & Engleman, R. 1983, Atlas of the Thorium spectrum
Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, A&A, 563, A54
Placco, V. M., Sneden, C., Roederer, I. U., et al. 2021, Research Notes of the American Astronomical Society, 5, 92
Ram, R. S., Brooke, J. S. A., Bernath, P. F., Sneden, C., & Lucatello, S. 2014, ApJS, 211, 5
Ramírez, I., & Allende Prieto, C. 2011, ApJ, 743, 135
Sadakane, K., Arimoto, N., Ikuta, C., et al. 2004, PASJ, 56, 1041
Sestito, F., Roediger, J., Navarro, J. F., et al. 2023, MNRAS, 523, 123
Shetrone, M., Venn, K. A., Tolstoy, E., et al. 2003, AJ, 125, 684
Shetrone, M. D., Côté, P., & Sargent, W. L. W. 2001, ApJ, 548, 592
Simon, J. D. 2019, ARA&A, 57, 375
Simon, J. D., & Geha, M. 2007, ApJ, 670, 313
Simon, J. D., Jacobson, H. R., Frebel, A., et al. 2015, ApJ, 802, 93

Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405

Sneden, C., Lucatello, S., Ram, R. S., Brooke, J. S. A., & Bernath, P. 2014, ApJS, 214, 26

Sneden, C. A. 1973, PhD thesis, The University of Texas at Austin.

Sobeck, J. S., Kraft, R. P., Sneden, C., et al. 2011, AJ, 141, 175

Song, Y.-Y., Mateo, M., Bailey, J. I., et al. 2021, MNRAS, 504, 4160

Song, Y.-Y., Mateo, M., Mackey, A. D., et al. 2019, MNRAS, 490, 385

Spencer, M. E., Mateo, M., Olszewski, E. W., et al. 2018, AJ, 156, 257

Spencer, M. E., Mateo, M., Walker, M. G., et al. 2017, AJ, 153, 254

Starkenburg, E., Hill, V., Tolstoy, E., et al. 2013, A&A, 549, A88

Starkenburg, E., Martin, N., Youakim, K., et al. 2017, MNRAS, 471, 2587

Swan, W. 1857, Transactions of the Royal Society of Edinburgh, 21, 411429

Szentgyorgyi, A. 2006, New Astronomy Review, 50, 326

Tafelmeyer, M., Jablonka, P., Hill, V., et al. 2010, ArXiv:1008.3721, arXiv:1008.3721

The DES Collaboration, Bechtol, K., Drlica-Wagner, A., et al. 2015, ArXiv:1503.02584, arXiv:1503.02584

Tody, D. 1986, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 627, Proc. SPIE, ed. D. L. Crawford, 733

Tolstoy, E., Hill, V., & Tosi, M. 2009, ARA&A, 47, 371

Tolstoy, E., Skúladóttir, Á., Battaglia, G., et al. 2023, arXiv e-prints, arXiv:2304.11980

Tolstoy et al. 2004, ApJL, 617, L119

Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2015, ApJ, 808, 108

—. 2007, ApJS, 171, 389

—. 2016, ApJ, 819, 53

Walker, Mateo & Olszewski. 2009, AJ, 137, 3100

Walker, Olszewski & Mateo. 2015, (accepted for publication in MNRAS), 0, 0

Waller, F., Venn, K. A., Sestito, F., et al. 2023, MNRAS, 519, 1349

Weisz, D. R., & Boylan-Kolchin, M. 2017, MNRAS, 469, L83

Wilkinson, M. I., Kleyna, J. T., Evans, N. W., et al. 2004, ApJL, 611, L21

Willman, B., Geha, M., Strader, J., et al. 2011, AJ, 142, 128