



# A Comprehensive Study of Open Cluster Chemical Homogeneity Using APOGEE and Milky Way Mapper Abundances

Amaya Sinha<sup>1</sup> , Gail Zasowski<sup>1</sup> , Peter Frinchaboy<sup>2</sup> , Katia Cunha<sup>3,4</sup> , Diogo Souto<sup>5</sup> , Jamie Tayar<sup>6</sup> , and Keivan Stassun<sup>7</sup>

<sup>1</sup> Department of Physics & Astronomy, University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, USA

<sup>2</sup> Department of Physics & Astronomy, Texas Christian University, Fort Worth, TX 76129, USA

<sup>3</sup> Steward Observatory, University of Arizona, Tucson, AZ 85721, USA

<sup>4</sup> Observatório Nacional, Rua General José Cristino, 77, 20921-400 São Cristóvão, Rio de Janeiro, RJ, Brazil

<sup>5</sup> Departamento de Física, Universidade Federal de Sergipe, Av. Marcelo Deda Chagas, Cep 49.107-230, São Cristóvão, SE, Brazil

<sup>6</sup> Department of Astronomy, University of Florida, Bryant Space Science Center, Stadium Rd., Gainesville, FL 32611, USA

<sup>7</sup> Department of Physics and Astronomy, Vanderbilt University, VU Station 1807, Nashville, TN 37235, USA

Received 2024 June 26; revised 2024 August 9; accepted 2024 August 26; published 2024 October 28

## Abstract

Stars in an open cluster are assumed to have formed from a broadly homogeneous distribution of gas, implying that they should be chemically homogeneous. Quantifying the level to which open clusters are chemically homogeneous can therefore tell us about interstellar medium pollution and gas mixing in progenitor molecular clouds. Using Sloan Digital Sky Survey (SDSS)-V Milky Way Mapper and SDSS-IV Apache Point Observatory Galaxy Evolution Experiment DR17 abundances, we test this assumption by quantifying intrinsic chemical scatter in up to 20 different chemical abundances across 26 Milky Way open clusters. We find that we can place  $3\sigma$  upper limits on open cluster homogeneity within 0.02 dex or less in the majority of elements, while for neutron capture elements, as well as those elements having weak lines, we place limits on their homogeneity within 0.2 dex. Finally, we find that giant stars in open clusters are  $\sim 0.01$  dex more homogeneous than a matched sample of field stars.

*Unified Astronomy Thesaurus concepts:* Open star clusters (1160); Stellar astronomy (1583); Star clusters (1567); Galactic archaeology (2178)

*Materials only available in the [online version of record](#): machine-readable tables*

## 1. Introduction

Immediately after the Big Bang, the only elements in the Universe were hydrogen, helium, and trace lithium. It took the formation of stars and galaxies to populate the Universe with the rest of the periodic table. Therefore understanding where and how stars produce and disperse heavy elements is essential to understanding the enrichment of the Universe. However, many questions regarding the chemical enrichment of the Universe still remain unanswered. Specifically, there is still much uncertainty on how well mixed giant molecular clouds are or how heavy elements get from their production sites into stars (F. Matteucci & P. Francois 1989; C. Chiappini et al. 2001; E. Spitoni & F. Matteucci 2011; D. H. Weinberg et al. 2019). Fortunately, with a few exceptions, the surface abundances of stars are fossil records of the gas composition from the molecular cloud in which they formed. As a result, we can use the present-day chemistry of stars to learn about the chemistry of the Milky Way in the past.

In the age of large astronomical surveys such as GALAH (G. M. De Silva et al. 2015; A. Sheinis et al. 2015; J. Kos et al. 2017; S. Buder et al. 2021; T. Zwitter et al. 2021), LAMOST (G. Zhao et al. 2012), RAVE (M. Steinmetz et al. 2020a, 2020b), Apache Point Observatory Galaxy Evolution Experiment (APOGEE; S. R. Majewski et al. 2017), and Gaia (Gaia Collaboration et al. 2016; D. Katz et al. 2023), we can now

probe the chemistry of stars in the Milky Way on the scale of  $\sim 0.1$  dex or smaller in multiple elements across different nucleosynthetic families, allowing us to trace different chemical enrichment pathways. Furthermore, we can now study the chemistry of the Milky Way at multiple different scales, from the simplest population in conatal binaries (K. Hawkins et al. 2020) to large populations of dispersed field stars (M. K. Ness et al. 2022).

Stars in an open cluster (OC) are assumed to have formed from a broadly homogeneous distribution of gas at the same time, implying that they should all have the same age and be at the same distance (C. J. Lada & E. A. Lada 2003). Using the chemistry of OC stars, we can infer the chemistry of the gas available at that point in the Milky Way's history, in particular within the thin disk. In the past, it has been suggested that by using assumptions of chemical homogeneity from simple stellar populations like OCs, it would be possible to reconstruct a dissolved cluster purely by its members' chemistry (K. Freeman & J. Bland-Hawthorn 2002). This technique, known as chemical tagging, has been a strong motivator for studies of cluster chemistry.

While many studies support this assumption of OC chemical homogeneity (G. M. De Silva et al. 2007; J. Bovy 2016; M. Ness et al. 2018; C. M. Cheng et al. 2021), there has been work showing that at least some clusters are chemically inhomogeneous (M. Ness et al. 2018). For example, D. Geisler et al. (2012) argued that NGC 6791 may not be chemically homogeneous, due to the presence of a potential Na–O anticorrelation, a relationship most commonly found in globular clusters (R. G. Gratton et al. 2001). This would be



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

an exciting result as NGC 6791 is already unique in the Milky Way, as both the most massive and the most metal-rich OC. However other studies, such as K. Cunha et al. (2015), have shown that it is chemically homogeneous within measurement uncertainties. For a more detailed discussion on previous studies of OC chemical homogeneity, see Section 5.

There are reasons why OCs could be heterogeneous in specific elements. Slow neutron-capture element abundances, such as Sr, Ba, and Zr, can change over a star's lifetime as it enters the asymptotic giant branch phase of its evolution (V. V. Smith & D. L. Lambert 1990). Dredge-up, which occurs in stars on the giant branch, causes the star's convective envelope to expand, and it eventually gets deep enough to pull CNO-cycled elements to the surface, thereby altering the surface abundances we measure (e.g., A. Bressan et al. 2012; N. Lagarde et al. 2012; J. Choi et al. 2016). NGC 6705 is an interesting OC regarding this effect, as it has been observed to also be enhanced in Na due to dredge-up (V. Loaliza-Tacuri et al. 2023).

The surface abundances of elements such as Mg can be affected by effects like mass transfer and atomic diffusion (G. Michaud et al. 2015; D. Souto et al. 2019). However, the latter only weakly impacts the upper giant branch. Lastly, mass transfer (e.g., C. Abate et al. 2013; N. Bastian et al. 2013; K. E. Milliman et al. 2015) and pollution events such as planetary engulfment (e.g., G. Laughlin & F. C. Adams 1997; M. H. Pinsonneault et al. 2001; J. K. Carlberg et al. 2010) can also alter a star's surface abundances.

However, if an OC was measured to have nonzero chemical scatter even accounting for these factors, that could point to interesting and understudied physics that may have occurred during the formation of the OC. Simulations have shown that turbulent mixing during cloud assembly naturally produces a stellar abundance scatter that is  $\sim 5$  times smaller than that in the natal gas; Y. Feng & M. R. Krumholz (2014) suggest that this mixing could explain the observed chemical homogeneity of stars forming from the same molecular cloud. This is supported by recent work by B. Bhattarai et al. (2024) who find OCs in FIRE-2 simulations (P. F. Hopkins et al. 2014, 2018; A. R. Wetzel et al. 2016) to have chemical scatter within 0.02 dex on average.

However, chemical inhomogeneity in real clouds could be due to effects not fully captured by simulations, related to internal turbulence and gas mixing within the progenitor molecular cloud or pollution events such as core collapse supernovae (CCSNe) that occurred earlier in the cluster's lifetime (e.g., L. W. Looney et al. 2006; M. R. Krumholz et al. 2019).

Quantifying the level of chemical homogeneity in OCs across a broad set of elements from various nucleosynthetic families would provide the basis for understanding the physics of early OC formation.

This work aims to constrain the chemical homogeneity in a large set of abundances and clusters to disentangle the causes of those chemical variances. The structure of the paper is as follows: Section 2 outlines the survey data, verification of the abundance uncertainties, and determination of the cluster membership. Section 3 details the methodology and calculation of the intrinsic scatter within each  $[X/Fe]$  across the final cluster sample. Section 4 presents the results of our work, and Section 5 compares our results to previous findings.

## 2. Data

### 2.1. SDSS

#### 2.1.1. SDSS-V/MWM

The abundances and radial velocities (RVs) we use are primarily drawn from the Milky Way Mapper (MWM; J. A. Johnson 2024, in preparation), a component of the fifth generation of the Sloan Digital Sky Survey (SDSS-V; J. A. Kollmeier et al. 2017; J. Kollmeier 2024, in preparation). We use data from Internal Product Launch 3 (IPL-3), which will form the basis for SDSS DR19 (K. Hawkins 2024, in preparation). This data set builds off of the observing strategies and survey goals outlined in SDSS Data Release 18 and includes observations of over one million targets (A. Almeida et al. 2023).

SDSS-V/MWM uses two telescopes: the Sloan Foundation Telescope at Apache Point Observatory (APO; J. E. Gunn et al. 2006) and the duPont Telescope at Las Campanas Observatory (LCO; I. S. Bowen & A. H. J. Vaughan 1973). Both are outfitted with nearly identical custom-built 300 fiber APOGEE spectrographs (J. C. Wilson et al. 2019), which reach a resolution of  $R \sim 22,500$ , spanning the range of wavelengths between 1.51 and 1.70  $\mu\text{m}$ . Unlike in SDSS-IV, which used a plug-plate system, SDSS-V now uses robotic fiber positioners (R. W. Pogge et al. 2020), which benefited from the adoption of a three-element corrector for the Sloan telescope at APO (R. H. Barkhouser et al. 2022).

Within IPL-3, three different data pipelines were used to analyze the data taken from APO and LCO: *The Payne* (Y.-S. Ting et al. 2019), *The Cannon* (M. Ness et al. 2015), and the APOGEE Stellar Parameters and Abundances Pipeline (ASPCAP; A. E. García Pérez et al. 2016). Both *The Payne* and *The Cannon* are label-transfer methods that determine stellar labels from spectra after being trained on a set of spectra with known labels. They are differentiated by the fact that *The Cannon* is a data-driven model which requires no information about stellar models. Rather, measurements from *The Cannon* inherit information from the models from its training data. *The Payne* incorporates physical models directly into its analysis. While these data sets are similar in many aspects, comparing the limits derived through each of them will provide a stronger constraint on the true homogeneity of the OCs in our sample.

#### 2.1.2. SDSS-IV/APOGEE

In addition to IPL-3, we use abundances and RVs from the 17th and final data release (DR17; Abdurro'uf et al. 2022) of SDSS-IV's (M. R. Blanton et al. 2017) APOGEE (S. R. Majewski et al. 2017), which contains over 700,000 stars. The initial targeting strategy for APOGEE-1 and APOGEE-2 are outlined in G. Zasowski et al. (2013) and G. Zasowski et al. (2017), respectively, and the final targeting for APOGEE-2 is outlined in R. L. Beaton et al. (2021) for APOGEE-2 north, and F. A. Santana et al. (2021) for APOGEE-2S.

The details for the APOGEE data reduction pipeline are described in D. L. Nidever et al. (2015), and the details for ASPCAP are found in A. E. García Pérez et al. (2016). The description for the updates to these pipelines for DR17 is included in J. A. Holtzman et al. (2024, in preparation). The MARCS model atmospheres and interpolation methodology used in APOGEE are described by H. Jönsson et al. (2020) and B. Gustafsson et al. (2008). The line lists used for DR17 are

outlined in V. V. Smith et al. (2021), and the spectral fitting used for ASPCAP is described in C. Allende Prieto et al. (2006). The details describing the APOGEE spectral grids can be found in Y. Osorio et al. (2020), and I. Hubeny & T. Lanz (2017), and lastly, the details for Turbospectrum can be found in B. Plez (2012).

We also include abundances from the BACCHUS Analysis of Weak Lines in APOGEE Spectra (BAWLAS; C. R. Hayes et al. 2022), a value-added catalog (VAC) in DR17. This VAC provided abundances for several chemical species having weak and blended lines that cannot be reliably analyzed using ASPCAP. This sample consists of high signal-to-noise ratio ( $\text{SNR} > 150$ ) red giant stars with no flags in either STARFLAG or ASPCAPFLAG and analyzed using the BACCHUS code (T. Masseron et al. 2016), which measures line-by-line elemental abundances from on-the-fly spectral synthesis. High-quality measurements are stacked to create a sample of elemental abundances for elements with weak or blended lines. We use the BAWLAS VAC abundances and uncertainties for the following elements: Na, P, S, V, Cu, Ce, and Nd.

Two separate uncertainties are reported for each BAWLAS abundance measurement. One is the  $X_{\text{FE\_ERR\_MEAS}}$  describing the measured uncertainty from the combined spectra using the same methodology as ASPCAP. The other is  $X_{\text{FE\_ERR\_EMP}}$ , which describes the uncertainty derived from the spectral lines themselves. Here to remain consistent in our analysis we use  $X_{\text{FE\_ERR\_MEAS}}$  as it is the closest to the uncertainty calculation method that we verify in Section 2.3.

The calculation of the abundances in the BAWLAS catalog is outlined fully in C. R. Hayes et al. (2022). Each spectrum has an associated  $X_{\text{SPECTRA\_FLAG}}$ , with values 0, 1, or 9. Both 0 and 9 indicate either suspicion with the final fit or total failure, and only 1 indicates that the spectral fit is trustworthy. To ensure the highest quality sample, we require all the stars used in this study to have measurements with  $X_{\text{SPECTRA\_FLAG}} = 1$ .

### 2.1.3. Quality Cuts

We limit our sample in APOGEE DR17 to stars with  $\text{VERR} < 0.1 \text{ km s}^{-1}$  and  $|\text{VHELIO\_AVG}| < 5000 \text{ km s}^{-1}$  to ensure our stars have reliable RVs. We also restrict our sample to stars with  $\text{VSCATTER} < 1 \text{ km s}^{-1}$ , to remove potential binaries within our sample (C. Badenes et al. 2018; A. M. Price-Whelan et al. 2020). Here  $\text{VHELIO\_AVG}$  refers to the average RV derived from individual RVs that are drawn from cross correlation of individual spectra with a combined spectrum.  $\text{VERR}$  refers to the uncertainty on that RV, and  $\text{VSCATTER}$  refers to the scatter of individual visit RVs around the average. To ensure the sample has reliable measurements we enforce an  $\text{SNR} > 50$ . These limits are identical between DR17 and IPL-3. We also limit our sample to stars between 3000 and 6500 K, and  $-1 \leq [\text{Fe}/\text{H}] \leq 1$ .

We also exclude from our sample any stars that have  $[\text{X}/\text{Fe}]$  flags in more than two elements. Lastly, we only sample from stars with  $\text{LOGG} \leq 3.5$  in order to ensure that every star we study is a member of the giant branch. These requirements result in a sample of 305,201 stars. When using IPL-3 we use the corresponding columns and limits, with the exception of  $\text{VSCATTER}$  which is not included in IPL-3.

In DR17's allStar file, we enforce quality cuts on the STARFLAG and ASPCAPFLAG columns, the details of which are included at [https://data.sdss.org/datamodel/files/APOGEE\\_ASPCAP/APRED\\_VERS/ASPCAP\\_VERS/allStar.html](https://data.sdss.org/datamodel/files/APOGEE_ASPCAP/APRED_VERS/ASPCAP_VERS/allStar.html). The

details of the APOGEE bitmasks are located at <https://www.sdss4.org/dr17/algorithms/bitmasks/>. Within the ASPCAP-FLAG column we enforced the following requirements:

1. BITMASK 23; STAR\_BAD == 0; BAD overall for star: set if any of TEFF, LOGG, CHI2, COLOTE, ROTATION, SN error are set, or any parameter is near grid edge (GRIDEDGE\_BAD is set in any PARAMFLAG).
2. BITMASK 19; METALS\_BAD == 0; FERRE failed to return a value for metals.
3. BITMASK 20; ALPHAFE\_BAD == 0; Elemental abundance from window differs  $> 0.5$  dex from parameter abundance for  $[\alpha/\text{Fe}]$ .

and within the STARFLAG we made the following cuts:

1. BITMASK 2; BRIGHT\_NEIGHBOR == 0; star has a neighbor more than 10 times brighter.
2. BITMASK 3; VERY\_BRIGHT\_NEIGHBOR == 0; star has a neighbor more than 100 times brighter.
3. BITMASK 17; SUSPECT\_BROAD\_LINES == 0; WARNING: cross-correlation peak with template significantly broader than autocorrelation of template.

The IPL-3 ASPCAP allStar file does not have a published ASPCAPFLAG column. As such, we require that all the stars in our sample have  $\text{flag\_bad} == \text{False}$  and  $\text{flag\_warn} == \text{False}$ . For all three IPL-3 pipelines we use we enforce a  $\text{result\_flags} == 0$  requirement. Where possible we enforce a  $\text{x\_h\_flag} == 0$  requirement within each pipeline. Within the three IPL-3 allStar files we use, we enforce these quality cuts on the  $\text{spectrum\_flag}$  column, which has bits that correspond to DR17's STARFLAG column.

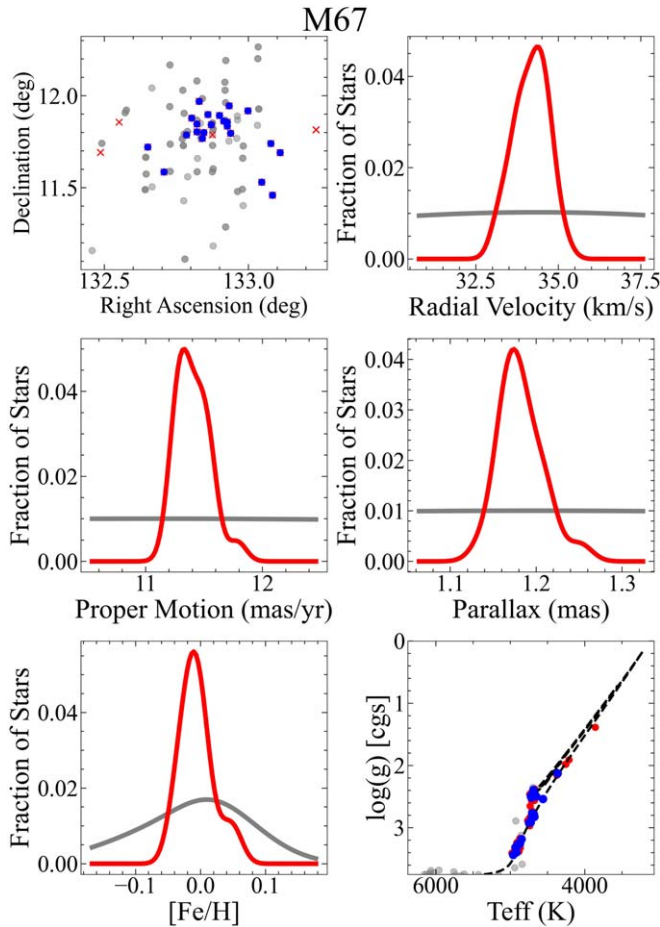
## 2.2. Cluster Membership

We start from the catalog of cluster members published in T. Cantat-Gaudin et al. (2018), hereafter **CG18**, which contains membership information for over 200,000 stars across  $\sim 2000$  OCs using Gaia DR2 (Gaia Collaboration et al. 2018). Due to the more recent availability of Gaia DR3 (Gaia Collaboration et al. 2021, 2023), we first match the stars identified as cluster members in **CG18** to their DR3 kinematics and positions. **CG18** used two spatial (R.A.  $\alpha$ , decl.  $\beta$ ) and two kinematic (proper motion–R.A.  $\delta_{\alpha*}$ , proper motion–decl.  $\delta_{\beta*}$ ) parameters, as well as parallax  $\varpi$ , as inputs into an unsupervised machine learning algorithm to determine cluster membership. We limit the initial cluster membership candidacy to stars from **CG18** within three cluster radii of their cluster centers. Using each cluster's distribution in RV, we find that a minimum probability cut of  $P \geq 0.5$  in the **CG18** catalog maximized the overlap in membership within DR17 and IPL-3 while also minimizing contamination from noncluster members. Of the 2019 OCs identified in **CG18**, only 145 clusters have any members within both DR17 and IPL-3.

### 2.2.1. Kinematic Selection

To further ensure that the clusters identified had minimal contamination, we use a kernel density estimator with a variable bandwidth, following Silverman's rule, to measure the dispersion in four dimensions (RV,  $\delta_{\alpha*}$ ,  $\delta_{\beta*}$ , and  $\varpi$ ). We reject stars further than two standard deviations from the cluster median. For an example of this methodology, see Figure 1, which shows the final distributions after these cuts in M67. The





**Figure 1.** Cluster membership plot for M67. Top left: the distribution of stars on the sky in R.A. and decl., with cluster members colored in blue, cluster member candidates that did not pass Section 2.2.1’s kinematic and parallax cuts in red, and field stars in gray. Top right: the KDE distribution of radial velocities of the cluster members compared to nearby field stars. Middle left: the distribution of the combined proper motion of the cluster members compared to nearby field stars. Middle right: the distribution of parallaxes of the cluster members compared to nearby field stars. Bottom left: the distribution of  $[\text{Fe}/\text{H}]$  of the cluster members compared to nearby field stars; note that this was not used to prune membership. Bottom right: The Kiel diagram of cluster members compared to surrounding stars. We also plot the MIST isochrone (J. Choi et al. 2016; A. Dotter 2016) for this cluster’s age and metallicity in black.

kinematic selection plots for all clusters in our sample can be found in Appendix B.

We plot the  $[\text{Fe}/\text{H}]$  distribution of each cluster against the distribution of nearby noncluster members between 2 and 5 cluster radii, as published in CG18. We visually inspect the H-R diagrams for the cluster members compared to the annulus to ensure no contamination, as each cluster should follow a single distinct isochrone. We use MIST isochrones (J. Choi et al. 2016; A. Dotter 2016), generated using ages from T. Cantat-Gaudin et al. (2020), and the median cluster  $[\text{Fe}/\text{H}]$ . Lastly, we only select clusters with  $N \geq 6$  members, resulting in a final sample set of 26 OCs. Determination of this minimum limit and the final sample size is related to material in Section 3.2.1.

### 2.2.2. Final Cluster Sample

The distribution of the final cluster sample in radius, age, and  $[\text{Mg}/\text{Fe}]$  versus  $[\text{Fe}/\text{H}]$  can be seen in Figure 2.

We calculate the positional, kinematic, and orbital information for each cluster in our final cluster sample. Using Astropy SkyCoords (Astropy Collaboration et al. 2013, 2018, 2022) we calculate the Cartesian  $X, Y, Z$  galactocentric coordinates for all the clusters in our sample, as well as the galactocentric radius.

We integrate each cluster’s orbit to measure its  $Z_{\text{max}}$ , eccentricity, and guiding radius using Galpy, with the MWPotential2014 gravitational potential (J. Bovy 2015; J. T. Mackereth & J. Bovy 2018). While this potential lacks a bar, no member of our sample is close enough to the Galactic center to produce a noticeable difference in the final measured eccentricity, guiding radius, or maximum height above the Galactic plane. While there have been measured effects on these parameters due to the Milky Way’s spiral arms, as seen in R. Carrera et al. (2022), as we do not use these parameters in the final science results, we do not include spiral arms in our potential.

We use a Monte Carlo method with  $N = 100$  iterations to estimate uncertainties on these parameters. We measure 3D space velocity dispersion as a proxy for cluster mass and using the methodology outlined in D. H. Weinberg et al. (2019), we quantify the ratio of nucleosynthetic enrichment within each cluster from CCSNe and Type Ia supernovae.

Lastly, we include age estimates on all our clusters from T. Cantat-Gaudin et al. (2020). These ages were derived using an artificial neural network trained on a sample of reference clusters. For further description see T. Cantat-Gaudin et al. (2020). At this stage we also flag red clump stars by eye within each cluster. All the cluster parameters are included in Table 1.

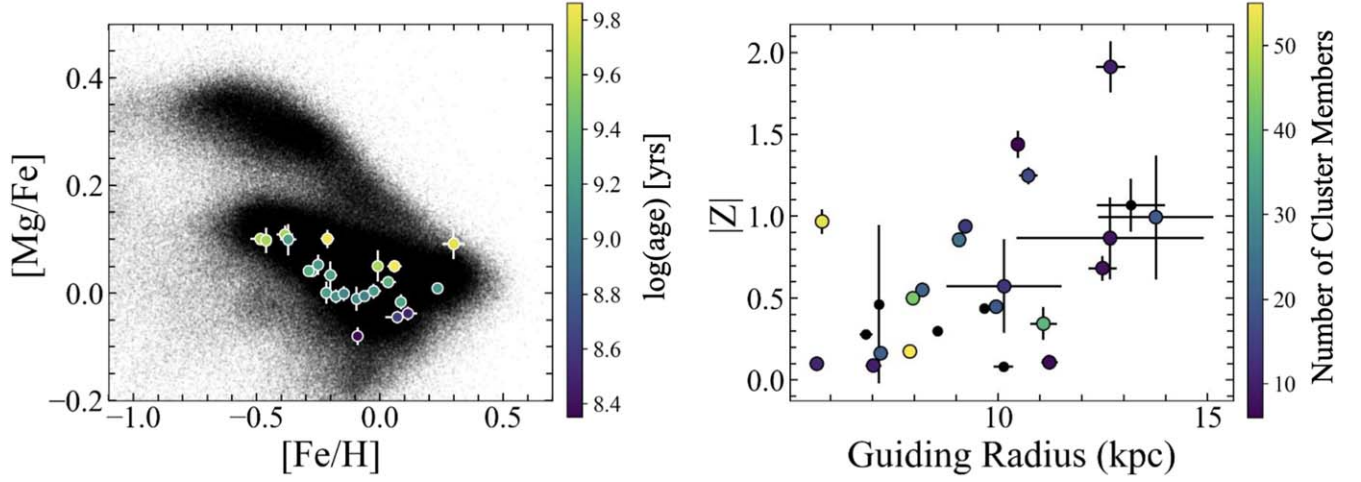
### 2.3. Verification of Abundances and Uncertainties

The values of intrinsic scatter in OC abundances are on the order of  $\sim 0.01$  dex (e.g., J. Bovy 2016; F. Liu et al. 2016; M. Ness et al. 2018; V. J. Poovelil et al. 2020), as are the abundance uncertainties within DR17. Therefore, verifying that the uncertainties on the abundances we are studying were not underestimated or overestimated was necessary. The method we use is the same as V. J. Poovelil et al. (2020). Since each star has some intrinsic true abundance measurement, repeated observations of the same star should result in a normal distribution around that value, where the width of the distribution is due only to the measurement uncertainties. Therefore we can use stars that have multiple visits in different APOGEE fields to measure the true uncertainty, and compare it to reported values within DR17 and IPL-3. To quantify our multiple-visit empirical uncertainty, we use Equation (1) from V. J. Poovelil et al. (2020):

$$e_{[\text{X}/\text{Fe}],k} = \frac{\sqrt{\pi}}{2} \text{median}(|[\text{X}/\text{Fe}]_j - [\text{X}/\text{Fe}]_i|), \quad (1)$$

where  $[\text{X}/\text{Fe}]_j$  and  $[\text{X}/\text{Fe}]_i$  are the abundance measurements from the same star’s  $i$ th and  $j$ th visits.  $e_{[\text{X}/\text{Fe}],k}$  is the resulting  $[\text{X}/\text{Fe}]$  uncertainty after median stacking the pairwise measurements in the  $k$ th bin.

We group stars by SNR in bins spanning 50–70, 70–100, 100–130, 130–200, and greater than 200, as shown in Figure 3. Within each SNR range, stars are binned by  $T_{\text{eff}}$  and  $[\text{M}/\text{H}]$ , where  $\Delta[\text{M}/\text{H}] = 0.2$  dex and  $\Delta T_{\text{eff}} = 200$  K. We use the Kolmogorov–Smirnov statistical test to ensure the distribution of abundance differences in each bin is consistent with a normal distribution, and flag those that are not to ensure they do not contaminate our sample. Bins that are not consistent

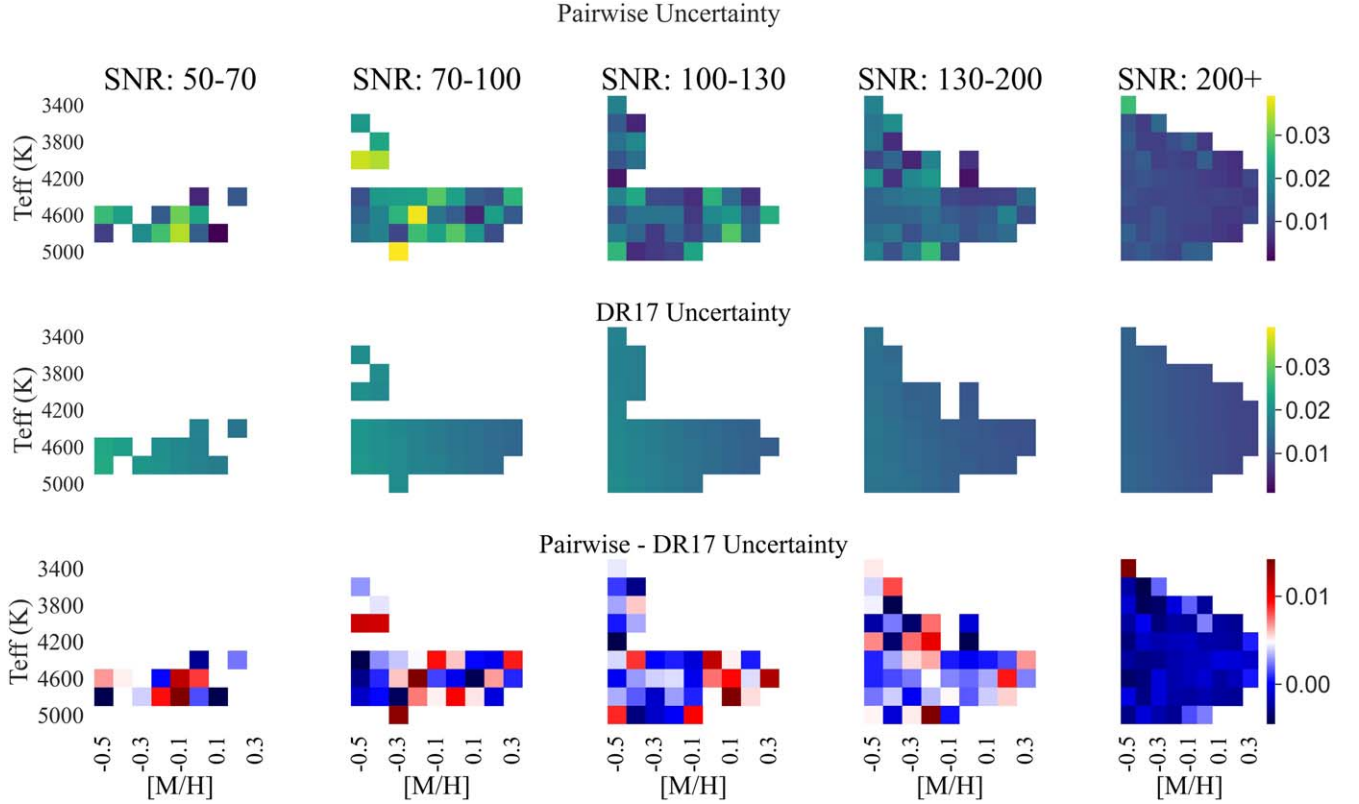


**Figure 2.** Left: a plot of the  $[Mg/Fe]$  vs.  $[Fe/H]$  plane where the full giant-star sample in APOGEE DR17 (subject to the same quality cuts as in Section 2.1.3) is shown in black, and the clusters are colored by  $\log(\text{age})$ . Right: the guiding radius and galactocentric height of all the clusters in our sample, colored by the number of member stars as described in Section 2.2.1.

**Table 1**  
List of Columns in the Table of Cluster Parameters

Column Name	Units	Column Description
cluster	...	Cluster name
Nmems	...	Number of identified cluster members identified in DR17 that passed all kinematic and quality cuts
RA	deg	Central R.A. for all the cluster members, from the median of Gaia DR3 coordinates
DEC	deg	Central decl. for all the cluster members, from the median of Gaia DR3 coordinates
parallax	mas	Median parallax of cluster members from Gaia DR3
parallax_error	mas	Standard deviation parallax of cluster members from Gaia DR3
pmRA	mas yr <sup>-1</sup>	Median R.A. proper motion of cluster members from Gaia DR3
pmRA_error	mas yr <sup>-1</sup>	Standard deviation of R.A. proper motion of cluster members from Gaia DR3
pmDEC	mas yr <sup>-1</sup>	Median decl. proper motion of cluster members from Gaia DR3
pmDEC_error	mas yr <sup>-1</sup>	Standard deviation of decl. proper motion of cluster members from Gaia DR3
RV	km s <sup>-1</sup>	Median RV of cluster members
RV_sigma	km s <sup>-1</sup>	Standard deviation of RV of cluster members
RV_sigma_error	km s <sup>-1</sup>	Uncertainty on RV_sigma calculated by bootstrapping the cluster
TV	km s <sup>-1</sup>	Median tangential velocity, calculated by $v_{\text{tangential}} = 4.74\mu\varpi^{-1}$ , where $\mu$ is the total proper motion
TV_sigma	km s <sup>-1</sup>	Standard deviation in tangential velocity of cluster members
TV_sigma_error	km s <sup>-1</sup>	Uncertainty on TV_sigma calculated by bootstrapping the cluster
SV	km s <sup>-1</sup>	Median 3D space velocity calculated by $v_{3D} = \sqrt{v_{\text{radial}}^2 + v_{\text{tangential}}^2}$
SV_sigma	km s <sup>-1</sup>	Standard deviation in 3D space velocity of cluster members
SV_sigma_error	km s <sup>-1</sup>	Uncertainty on SV_sigma calculated by bootstrapping the cluster
Rgc	kpc	Median galactocentric radius from Galpy integration
Rgc_error	kpc	Uncertainty on galactocentric radius from Galpy integration
Rguiding	kpc	Median guiding radius from Galpy integration
Rguiding_error	kpc	Uncertainty on guiding radius from Galpy integration
Zmax	kpc	Median maximum distance from galactic midplane from Galpy integration
Zmax_error	kpc	Uncertainty on the maximum distance from galactic midplane from Galpy integration
e	...	Median eccentricity from Galpy integration
e_error	...	Uncertainty on eccentricity from Galpy integration
X	kpc	Median $x$ galactocentric coordinate using Cartesian coordinates
X_error	kpc	Uncertainty on $x$ galactocentric distance using Cartesian coordinates
Y	kpc	Median $y$ galactocentric coordinate using Cartesian coordinates
Y_error	kpc	Uncertainty on $y$ galactocentric coordinate using Cartesian coordinates
Z	kpc	Median $z$ galactocentric coordinate using Cartesian coordinates
Z_error	kpc	Uncertainty on $z$ galactocentric coordinate using Cartesian coordinates
SNe_ratio	...	Ratio of enrichment in the cluster from Type Ia versus CCSNe
SNe_ratio_erro	...	Uncertainty on the ratio of enrichment in the cluster from Type Ia versus CCSNe
log_age	dex	Log age in years of cluster, from T. Cantat-Gaudin et al. (2020)

(This table is available in its entirety in machine-readable form in the [online article](#).)



**Figure 3.** Top: [Mg/Fe] uncertainties calculated using stars with multiple visits. Middle: [Mg/Fe] uncertainties calculated by ASPCAP for DR17. Bottom: the subtracted difference between the uncertainties calculated using multiple visits and the uncertainties from ASPCAP.

with a normal distribution have two main causes. First, some are at the edge of the [M/H] parameter space, in particular at low metallicities where ASPCAP is less reliable. Second, some bins have poor completion, with less than 10 measurements.

Within those bins where the empirical uncertainties are well measured and normally distributed, we find very good agreement with the native pipeline uncertainties; thus for the rest of this work, we adopt those pipeline uncertainties directly from DR17 and IPL-3.

### 3. Measuring Chemical Scatter

#### 3.1. Paired-stars Method

Our primary method of determining the intrinsic scatter uses the difference in abundances between stars close to one another on the H-R diagram:  $\Delta T_{\text{eff}} < 100$  K and  $\Delta \log g < 0.1$  dex. These limits are chosen because the maximum induced-abundance offset between pair members due to the systematics discussed in Section 3.2.2 is on the order  $\sim 0.001$  dex.

We measure the intrinsic dispersion within each pair using Equation (2), where  $e_1$  and  $e_2$  are the abundance uncertainties of each star in the pair,  $|\Delta[X/\text{Fe}]|$  is the absolute value of the difference in abundance measurements between the pair, and  $\sigma$  is the inferred intrinsic dispersion

$$\sigma = \sqrt{\frac{\pi}{2} |\Delta[X/\text{Fe}]|^2 - (e_1^2 + e_2^2)}. \quad (2)$$

We use a Monte Carlo method with  $N = 100$  iterations to vary the abundance measurement of each star within the pair to estimate a final uncertainty on the measured pairwise scatter. Within each cluster, we separate the red clump and red giant

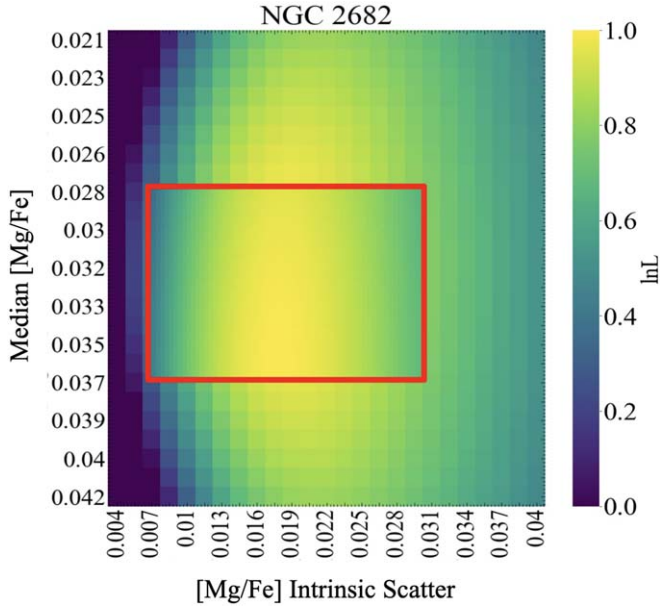
branch stars as to not induce any scatter from potential evolutionary effects. Within each subsample, we sort the stars into pairs and then measure the intrinsic scatter between them. Finally, we take the median scatter of all the pairs within an OC as the true intrinsic scatter of the cluster. This method allows for extremely precise results, with final uncertainties on the order of  $\sim 0.001$  in most elements.

To determine the number of pairs needed for a reliable measurement with this method, we use a synthetic “cluster” of points, with “true” and “observed” [X/H] abundances, and the same temperature and  $\log(g)$  distribution as our real clusters. The true abundances for the synthetic stars reflect a given intrinsic scatter for the synthetic cluster. The observed abundances are generated by perturbing each true abundance by a random value drawn from a normal distribution with  $\sigma$  set to the uncertainty of a real APOGEE star with the same temperature,  $\log(g)$ , and metallicity.

We then pair the stars as described in Section 3.1 and perform the intrinsic scatter measurement described using  $N = 3$  to  $N = 20$  pairs. We find that the measured cluster scatters are noisy and have both systematic offsets and larger uncertainties up until  $N = 8$ , at which point the difference between the true and measured scatter does not change at larger  $N$ . Thus, we require a minimum of eight stellar pairs for clusters using this method.

Due to the restrictions outlined above regarding the separation of the pairs in  $T_{\text{eff}}$  and  $\log g$ , as well as the minimum number of pairs required, this method can only be applied to 15 of the 26 clusters studied in this paper. However due to the significantly higher precision of these values as compared to those derived using the method outlined





**Figure 4.** Outside red: the initial likelihood surface over which we initially test to determine a best guess for  $\sigma_{[X/Fe]}$ . Within red: using the findings from the initial run, we more finely sample the likelihood surface in order to measure the final intrinsic scatter and uncertainty.

in Section 3.2, within these 15 clusters we only publish results from this method.

### 3.2. Maximum Likelihood Estimator

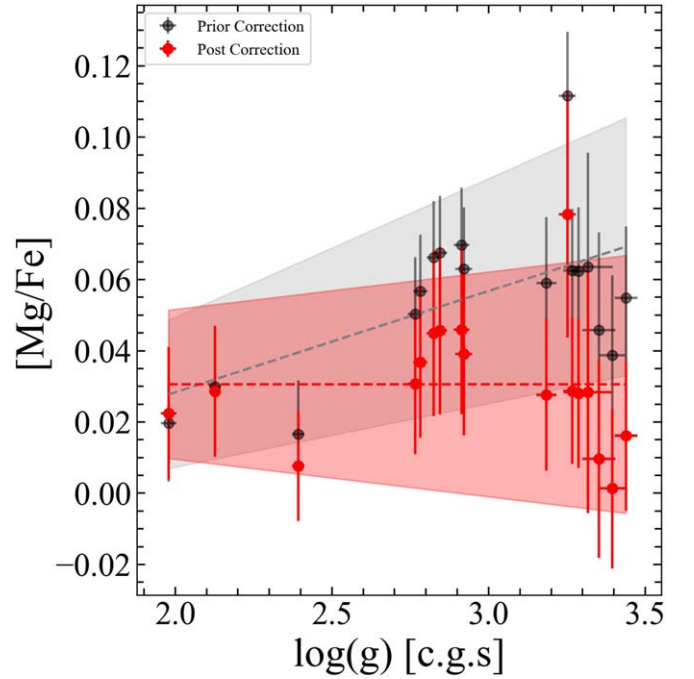
We adopt a second method in the form of a maximum likelihood estimator (MLE) to calculate the intrinsic scatter across our sample. This method produces larger uncertainties than our pairwise method, but it also has fewer sampling restrictions and can be applied to a larger set of OCs. The form of the MLE is shown in Equation (3) below:

$$\ln L = \prod_{i=1}^{\infty} \frac{1}{\sqrt{2\pi}(\sigma_{[X/Fe]}^2 + e_i^2)^{1/2}} \exp\left(\frac{-(x_i - \mu_{[X/Fe]})^2}{2(\sigma_{[X/Fe]}^2 + e_i^2)}\right). \quad (3)$$

In Equation (3),  $\sigma_{[X/Fe]}$  is the intrinsic scatter being tested, and  $e_i$  is the uncertainty on the  $[X/Fe]$  measurement for the  $i$ th star in that cluster. We sample a narrow range of mean  $[X/Fe]$  ( $\mu$ ), where  $\Delta\mu = 0.05$  dex around the calculated mean of the cluster members, with an initial range of 0.1 dex for intrinsic scatter and spacing of 0.003 dex. We then do a second iteration with finer spacing in the intrinsic scatter dimension, with spacing on the order of  $10^{-4}$  dex, centered on the likeliest value from the coarser grid. An example of this is shown in Figure 4. We calculate a variance from the Fisher information matrix, and from that we calculate the uncertainty the intrinsic scatter. We apply the MLE method to all 26 clusters in our sample. Of these 26 clusters in our sample, 10 use the MLE intrinsic scatters for their final measurement.

#### 3.2.1. MLE Corrections for Small Samples

Based on tests with synthetic data, we find that the measured intrinsic scatter is unreliable when the number of cluster stars ( $N$ ) is low, with a consistent systematic bias at  $N < 9$  in the



**Figure 5.** Uncalibrated surface gravity vs.  $[Mg/Fe]$  for M67. The  $[Mg/Fe]$  values from DR17 are shown in black, and have a clear linear relationship with  $\log(g)$ . The adjusted abundances used in the calculation of intrinsic scatter are shown in red.

measured scatter. This is due to the fact that the MLE is a biased estimator and at small sample sizes has a negative bias that causes it to underestimate the true parameter value, which can be accounted for with a multiplicative factor (J. Thorley & C. Schwarz 2018):

$$\Delta\sigma_{\text{true}} = \sqrt{\frac{N}{N-1}} \sigma_{\text{MLE}}. \quad (4)$$

This correction is applied to clusters with  $6 \leq N < 9$  members. Below six members, the measured scatters are too unpredictably noisy. This small sample size correction results in an additional six clusters added to our sample, creating the final sample of 26 OCs.

#### 3.2.2. Systematics

As discussed in A. E. García Pérez et al. (2016), the methodology for measuring a star's  $[X/Fe]$  involves a multiparameter fit that includes fitting the observed stellar spectrum to synthetic models. First, a global fit to the spectrum is done to determine the best-fit values for temperature, surface gravity, microturbulent velocity, and  $[M/H]$ . Holding these parameters constant, individual elemental abundances are extracted from narrow spectral windows.

To test for any systematic trend between the global stellar parameters and elemental abundances, we quantify the slope of the uncalibrated  $\log(g)$  versus  $[X/Fe]$  within all the clusters in each pipeline, approximating it as a linear relationship, as seen in Figure 5. We exclude red clump stars as they are further along their evolutionary track than red giant stars, and potentially have slightly different surface abundances due to evolutionary effects or internal systematics. As a result, including them in the slope measurement could artificially drive any measurements of chemical homogeneity. However, these red clump stars are adjusted afterwards along with the

other giants in their cluster. This process ensures a uniform abundance correction across the entire cluster. From this, we find that the existing slopes in the cluster sample are nonzero, with a median slope of  $\sim 0.02 \text{ dex dex}^{-1}$ .

This systematic bias is present in DR17, and in The Cannon and ASPCAP allStar files from IPL-3. However it is not present in the IPL-3 allStar release analyzed using The Payne. We adjust the measured  $[X/\text{Fe}]$  for each cluster star using the following equation:

$$[X/\text{Fe}]_{i,\text{corrected}} = [X/\text{Fe}]_i - m \log g_i + \text{ZP}, \quad (5)$$

where each  $i$ th index is a star in a cluster and  $m$  is the best-fit slope of  $[X/\text{Fe}]$  and  $\log(g)$  within a specific cluster. We set the zero-point (ZP) of the cluster  $[X/\text{Fe}]$  after the correction using the abundances of the stars on the giant branch below the red clump to ensure that the median cluster  $[X/\text{Fe}]$  is reflective of its true value. The fitting uncertainties are propagated to uncertainties on the correction, which are then added in quadrature with the existing abundance uncertainties.

#### 4. Results

Within one standard deviation, the only abundance that showed evidence of inhomogeneity, or consistent nonzero intrinsic scatter, across multiple clusters was  $[\text{M}/\text{H}]$ . This is because  $[\text{M}/\text{H}]$  has small uncertainties compared to other  $[X/\text{Fe}]$  measurements ( $\sim 0.008 \text{ dex}$  as compared to  $\sim 0.015 \text{ dex}$  for the rest of the DR17 and IPL-3 abundances and  $\geq 0.03$  for the BAWLAS abundances). Given that the  $[\text{M}/\text{H}]$  uncertainties are smaller but not underreported, as we verified in Section 2, it potentially implies that we could detect the existence of inhomogeneities that were then masked by larger uncertainties in the other abundances.

However, within three standard deviations (a 99.7% confidence interval), none of the clusters show measurable inhomogeneities in any of the measured elements. Furthermore, the scale of the limits derived using the paired-stars method are also on the scale of  $\sim 0.001 \text{ dex}$  in many elements. And using that method we find no measurement of scatter across in any of the elements and clusters. As a result, we are confident that in the majority of elements we can constrain the homogeneity of the OCs to less than  $0.01 \text{ dex}$  in a 99.7% confidence interval.

All the measured quantities for each  $[X/\text{Fe}]$  are presented in Table 2. We show that across DR17 and all MWM pipelines, we do not find any elements that show consistent chemical inhomogeneity across our cluster sample, and in Figure 6 we show that we do not find any clusters with consistent scatter across their abundance samples. We only show the results from the IPL-3 ASPCAP data here because they include values for weak-lined elements. The literature comparison plots using DR17, IPL-3 Cannon, and IPL-3 Payne releases are comparable and are shown in Appendix A.

The upper limits on the intrinsic scatter measured in elements included in the BAWLAS catalog are higher than the limits on intrinsic scatter placed on the more well-measured elements, such as Mg or Ni. The reasons for this are twofold: First, while many of the clusters studied did have enough stars to measure an intrinsic scatter, the number of stars that contained BAWLAS abundances within each cluster was smaller than the number of stars used to calculate  $\alpha$  and iron-peak elements. Second, the associated uncertainties for the

**Table 2**  
List of  $[X/\text{Fe}]$  Columns in the Table of Cluster Parameters for IPL-3 ASPCAP

Column Name	Column Description
cluster	Cluster name
Pipeline	Which pipeline (ASPCAP, Cannon, Payne) was used to derive abundances
Method	Whether the paired-stars ( <code>pair</code> ) or MLE method (MLE) was used
<code>x_fe</code>	Median $[X/\text{Fe}]$ of the cluster
<code>x_fe_cluster_disp</code>	Dispersion in $[X/\text{Fe}]$ within the cluster
<code>x_fe_sigma</code>	$[X/\text{Fe}]$ intrinsic MLE scatter
<code>x_fe_sigma_uncertainty</code>	$[X/\text{Fe}]$ intrinsic pairwise scatter uncertainty
<code>x_fe_sigma_upper_limit</code>	$[X/\text{Fe}]$ intrinsic MLE scatter 3 uncertainty upper limit

**Note.** All columns have units of dex, and all X/Fe measurements are scaled to solar values. The abundances measured in these columns are Mg, Ca, Si, S, P, Ni, Cr, Ti, V, Mn, Co, Na, K, Al, V, Mn, Co, Cu, CNO, Nd, and Ce. In addition we provide columns for  $[\text{Fe}/\text{H}]$  and  $[\text{M}/\text{H}]$ .

(This table is available in its entirety in machine-readable form in the [online article](#).)

weak-lined elements were appreciably larger ( $0.03\text{--}0.08 \text{ dex}$ ) than the ones included in DR17 and MWM ( $0.01\text{--}0.04 \text{ dex}$ ).

#### 5. Discussion

##### 5.1. Comparison to Milky Way Field Stars

To quantify the difference in chemical homogeneity between our OCs and the Milky Way field, we create a matched field-star (MFS) sample that mirrors our existing cluster sample.

For each of the 15 clusters in our study with enough members to apply the pairwise method, we match each star within the cluster to a field star within  $2\sigma^8$  in Galactocentric radius,  $[\text{M}/\text{H}]$ ,  $[\alpha/\text{M}]$ ,  $T_{\text{eff}}$ , and  $\log g$ . Here we consider the parameter uncertainties of both the cluster star as well as any candidate field stars. We use each star's  $[\text{M}/\text{H}]$  and  $[\alpha/\text{M}]$  value from prior to the stellar parameter correction. We then measure the intrinsic scatter in each of the MFS samples, replicating the methodology outlined in Section 3.1. Finally, we compare the difference in intrinsic scatter between our MFS sample and the OC sample (Figure 7).

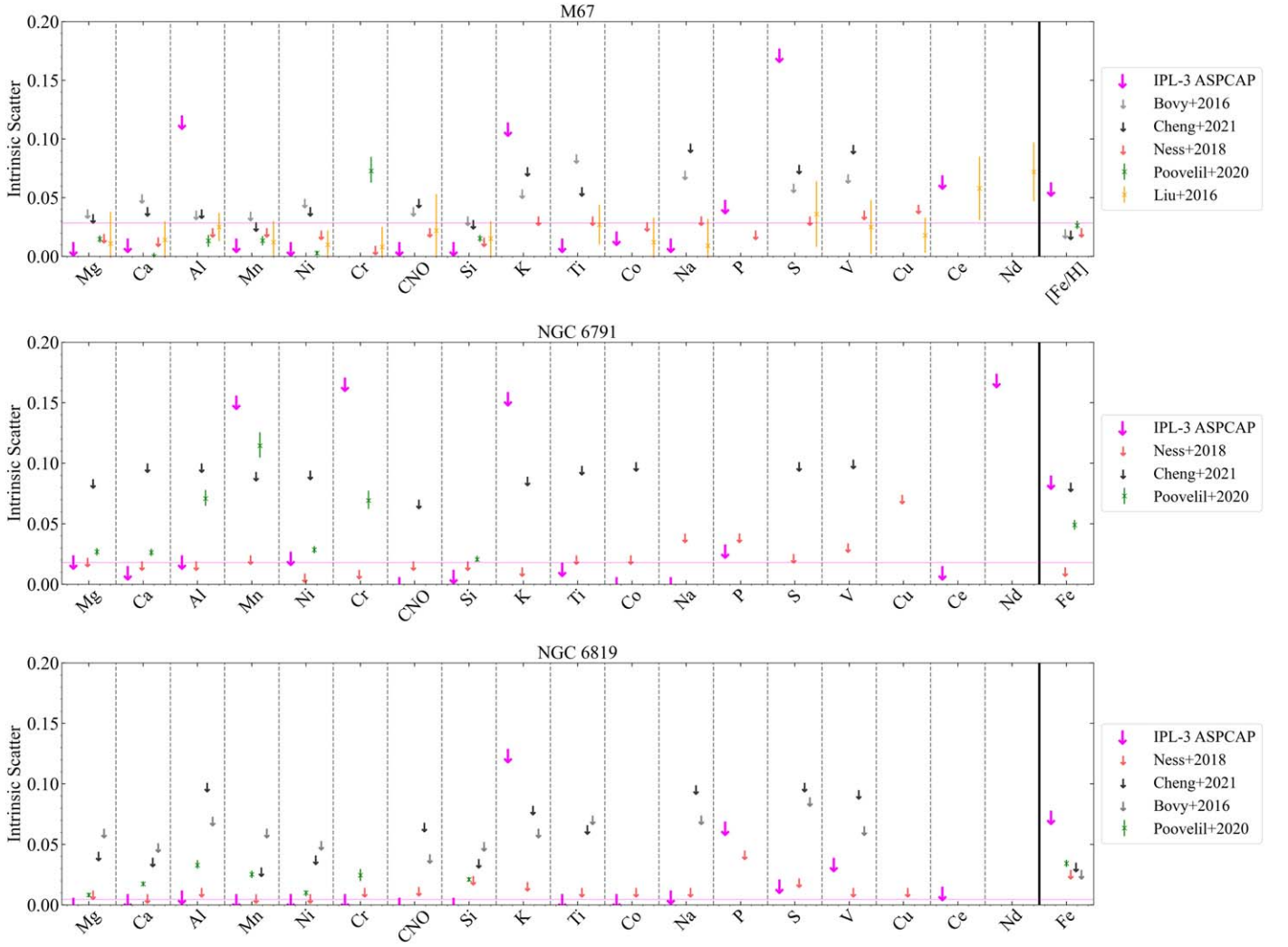
We find that on average, across all abundances, the MFS samples have  $+0.012^{+0.02}_{-0.01} \text{ dex}$  more intrinsic scatter than the OCs in our sample. This is in strong agreement with M. K. Ness et al. (2022), which states that stars in the Milky Way are chemically similar (within  $\sim 0.01\text{--}0.02 \text{ dex}$ ) when given a fixed Galactocentric radius,  $[\text{M}/\text{H}]$ , and  $[\alpha/\text{Fe}]$ . The median difference between OC intrinsic scatter and field-star intrinsic scatter ( $\Delta\sigma_{[X/\text{Fe}]}$ ) for each nucleosynthetic family is given below:

1.  $\alpha$ -elements (Mg, Si, Ca, Ti, P, S):  $0.002 \text{ dex}$ .
2. Iron-peak elements (Cr, Mn, Fe, Co, Ni, V):  $0.012 \text{ dex}$ .
3. Odd-Z (Na, Al, K):  $0.023 \text{ dex}$ .
4. Neutron capture (Nd, Ce):  $0.02 \text{ dex}$ .

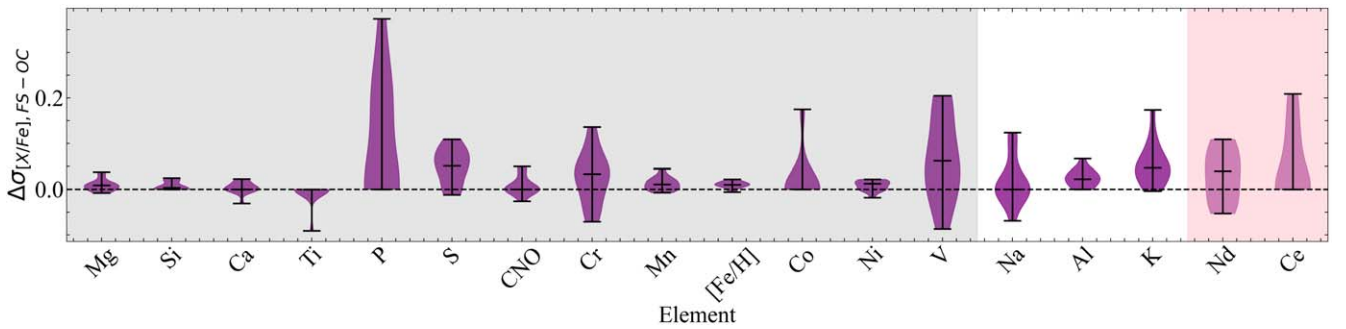
Due to our selection criteria for the field-star sample, we expect similar intrinsic scatter in the  $\alpha$  and iron-peak elements. Interestingly, two of the odd-Z elements, Al and K, both have

<sup>8</sup> We tested the effects of using  $1\sigma$  and  $3\sigma$  to match stars and found no difference in our conclusions.





**Figure 6.** The 99.7% upper limits using abundances from IPL-3 ASPCAP in M67, NGC 6791, and NGC 6819. The median 99.7 percentile upper limit for each cluster is shown as a magenta horizontal line. We compare our findings to upper limits and measurements from previous literature. The upper limits are broadly in agreement with other studies for well-measured elements in APOGEE and Milky Way Mapper.



**Figure 7.** A violin plot comparing the difference in element intrinsic scatter between open clusters and matched samples of field stars as shown in Section 5.1. Here the elements that belong to the  $\alpha$ , CNO, and iron-peak families are shaded in gray as we expect them to be similar due to our selection criteria. Within these elements P and V have larger limits due to them being weak-lined elements from the BAWLAS catalog. While broadly we find that the field-star samples have more scatter than their OC counterparts, we do not find any conclusive difference in any element between the two samples.

nonzero scatter in the majority of our MFS samples despite being measured as homogeneous in our OCs. While in  $[\text{Na}/\text{Fe}]$  the distribution in  $\Delta\sigma_{[\text{X}/\text{Fe}]}$  is roughly symmetric, Na is an element with weak lines in APOGEE’s wavelength range and was included in our sample with the BAWLAS catalog. As a result, not only does it have comparatively larger uncertainties than the other odd-Z elements, but it has poorer completion as

well, as only a subset of high-SNR stars in our study have BAWLAS abundances. This implies that odd-Z elements may be a useful tool in differentiating otherwise chemically similar populations. However, given that the distribution for all of these elements is consistent with zero in at least a subset of clusters and field-star comparisons, more precise limits are needed to accurately test this.

Neutron-capture elements also show slightly larger differences in scatter than their field-star counterparts. C. Manea et al. (2023) showed that neutron-capture elements have more discriminatory power in distinguishing “doppelgänger stars.” What we find potentially corroborates that, but we also show that for coeval and conatal stars within an OC, the expectation of chemical homogeneity within neutron-capture elements is broadly comparable to that of elements from other well-studied and well-measured nucleosynthetic families. In a field-star sample with a high expectation of chemical similarity to an OC, as shown by the relative lack of difference in measured  $\alpha$ -element intrinsic scatter, there is a  $0.02 \pm 0.02$  dex difference in measured scatter for neutron-capture elements. Due to this, it seems possible that neutron-capture elements could be useful in distinguishing otherwise highly similar stellar populations but significantly more precise limits would be required to accomplish that goal.

In both cases of odd- $Z$  and neutron-capture elements, the differences between each OC and their respective MFS are consistent with zero within three uncertainties. More precise limits are required to make any conclusive statements on their effectiveness in distinguishing conatal and coeval stellar populations.

### 5.2. Previous Studies of Chemical Homogeneity

There have been numerous studies focusing on measuring the chemical homogeneity of star clusters; however, most of these studies have been focused on larger and more complicated globular clusters. Within OCs, different studies have found a wide range of limits on inhomogeneity. This is further complicated by the fact that not every study investigates the same set of abundances, nor is every analysis method comparable to one another. One of the differences between this study and many others is that their published limits on homogeneity are either 68% and 95% limits; the values we publish and show in Figure 6 and Appendix A are all 99.7% limits.

Given the number of studies done on M67, NGC 6791, and NGC 6819, they are shown in Figure 6, while the literature comparisons for the remaining clusters are located in Appendix A. The figures comparing the intrinsic scatter measured using DR17, Cannon, and Payne abundances for all 26 OCs are also shown in Appendix A.

We compare our results to four previous studies (J. Bovy 2016; F. Liu et al. 2016; M. Ness et al. 2018; V. J. Poovelil et al. 2020). We find that within the  $\alpha$  and iron-peak elements, which are well measured in APOGEE and MWM, the upper limits derived in this study are consistent with previous findings. In the weak-lined elements such as V, Cu, Ce, and Nd, there is more variance. But even within those elements, in many clusters we find comparable limits to previous works.

It is worth outlining the differences in the analyses and sample sets between these different studies. V. J. Poovelil et al. (2020) is the most similar to ours in terms of sample size, analysis method, and data set (using APOGEE DR16; R. Ahumada et al. 2020). While they measure upper limits that are far less constraining than ours, they also measure *lower* limits that strongly imply the existence of true intrinsic scatter. However, the stellar parameter systematic that we found in DR17 and MWM is also present in DR16 but was unaccounted for in V. J. Poovelil et al. (2020)’s final results. Therefore, it is possible that the measured scatters in V. J. Poovelil et al. (2020) are impacted by a relationship

between stellar parameters and abundance measurements. Our uncorrected  $\sigma_{[X/Fe]}$  values (not shown in this paper) are in strong agreement with the ones published in V. J. Poovelil et al. (2020), which lends evidence to this hypothesis.

F. Liu et al. (2016) uses high-resolution spectroscopy ( $R \sim 50,000$ ) to study two solar twins in M67. This makes it less likely that their final abundances are driven by systematics due to stellar atmospheres or poor line fitting, as the stars are in very similar parts of parameter space. This method is similar, but not identical, to the pairwise method of deriving intrinsic scatter outlined in Section 3.1. They derive abundances for a total of 26 elements as well as  $[Fe/H]$ , with an average measurement uncertainty of  $e_{[X/Fe]} \leq 0.02$  dex. As a result, within their sample they were able to very tightly constrain the difference in  $[X/Fe]$  between the two stars, showing that for elements with  $Z < 30$ , there was an average difference of  $< 0.06$  dex, and for the neutron-capture elements there was an average difference of  $< 0.05$  dex. It is encouraging that in the strong-lined elements included in our study, we place similar upper limits on abundance scatter.

M. Ness et al. (2018) uses APOGEE DR13 (F. D. Albareti et al. 2017) abundances and spectra to measure intrinsic dispersions within a set of seven OCs, six of which are included in this study. However, their methodology in constraining scatter was noticeably different than this work. Using The Cannon (M. Ness et al. 2015), they derive abundances and uncertainties in 20 different elements from APOGEE spectra and a training set of OC stars in APOGEE DR13. M. Ness et al. (2018) notes that while their Cannon abundance measurements were broadly comparable to ASP-CAP’s, the uncertainties are between 20% and 50% smaller. Beyond that, using a chi-squared fit they determined that the uncertainties, a quadrature sum of formal abundance uncertainty and cross-validation uncertainty, are overestimated given the widths of calculated abundance distributions. Therefore, they derive a scaling factor to correct the uncertainties to match the theoretically predicated value; however, this methodology also introduces the risk of artificially downscaling the measured limits on the intrinsic scatter. However, it is also worth noting that in the majority of elements, the values they publish are comparable to the ones derived in this work.

J. Bovy (2016) studied the abundance spread of 15 different elements in three clusters, all of which are included in this study: M67, NGC 2420, and NGC 6819. The data came from APOGEE DR12 (S. Alam et al. 2015), with cluster membership from S. Mészáros et al. (2013). However, the key difference between their study and this one is the methodology. J. Bovy (2016) made the assumption that in the absence of any intrinsic chemical scatter, the main driver for variation in the photometric and spectroscopic attributes of OC stars is their mass, which can be modeled as a one-dimensional sequence. They correct for any systematic variations in the spectra driven by mass, using  $T_{\text{eff}}$  as a proxy. They then use detailed forward modeling of the spectra and approximate Bayesian computation to measure the intrinsic scatter as well as upper limits. Overall J. Bovy (2016) found no indication of chemical inhomogeneity in any of the three clusters they studied; the upper limits they derived are largely in agreement with the ones calculated using the MLE method in this work. However, using the paired-stars method we can constrain tighter upper limits in the majority of elements. Similar to studies discussed previously, the limits

placed on the BAWLAS neutron-capture elements by J. Bovy (2016) are lower than ours.

Finally, C. M. Cheng et al. (2021) use spectroscopic data from APOGEE DR14 (B. Abolfathi et al. 2018) to measure intrinsic scatter in M67, NGC 6791, and NGC 6819 in 15 different elemental abundances. The analysis method is very similar to the one outlined in J. Bovy (2016), though there are a few differences—notably, that they use DR14 instead of DR12, which includes several differences in the line lists (detailed in J. A. Holtzman et al. 2018). Furthermore, unlike J. Bovy (2016), they use spectroscopic effective temperatures in their one-dimensional model as opposed to photometric effective temperatures. C. M. Cheng et al. (2021) found the clusters to be chemically homogeneous, placing upper limits comparable to this study across its sample of elements. While they measured fewer abundances than this work, in the three clusters studied the upper limits they derive are similar to J. Bovy (2016).

Thus, within the majority of the elements included in DR17 and IPL-3, such as the  $\alpha$ -elements and iron-peak elements, the upper limits we calculate here are in agreement with what has been previously found across all the clusters studied. This lends credibility to the limits placed on the numerous clusters studied in this work that did not have previously derived limits in the literature. However, the limits derived in this study for the neutron-capture elements are larger than what has been previously found in any of the literature. This is likely driven by the comparatively large uncertainties on the elements (0.03–0.08 dex).

## 6. Conclusion

The purpose of this study was to quantify the level of chemical homogeneity within the largest sample to date of Milky Way OCs for a broad set of elements. Using SDSS-V MWM IPL-3 abundances and Gaia DR3 kinematics, we identify a sample of 26 OCs with large enough membership to measure the intrinsic scatter in up to 20 elements. Using the abundance differences between paired stars along the H-R diagram, as well as a MLE, we then measure the intrinsic scatter within each element for each cluster. We find the following:

1. We assemble a sample of 26 OCs across a broad range of metallicity, age, mass, and Galactic radii. Within a 99.7% confidence interval, we do not find any evidence of intrinsic scatter on the giant branch or in the red clump in any element across all the OCs in our sample.
2. Within the majority of abundances included in APOGEE DR17 and MWM IPL-3, we constrain the chemical homogeneity to  $\leq 0.02$  dex within a 99.7% confidence interval, and within  $\leq 0.2$  dex for the weak-lined elements, such as those included in the DR17 BAWLAS catalog. Our limits are consistent with those in the literature for well-studied elements and clusters, and we add roughly a dozen clusters to this literature sample. Given the limited data set in some of the elements, we recommend follow-up measurements to better quantify their upper limits.
3. When compared to a sample of field stars with similar Galactocentric radii,  $[\alpha/M]$ , and  $[M/H]$ , we find our OCs to be more chemically homogeneous, with an average difference of  $\sim 0.012$  dex between the two samples. This corroborates previous findings that the dimensionality of chemical enrichment of the Milky Way is low, and can

likely be explained through a few processes. In the future this could be useful in placing constraints on radial mixing and azimuthal variations within the Milky Way.

4. We identify surface-gravity-dependent abundance shifts within APOGEE DR17 and MWM IPL-3 (corrected for in this analysis). This systematic needs to be accounted for in similar future work. We also find that the abundance uncertainties within both APOGEE and MWM are accurately estimated.
5. These findings have implications for attempts to implement chemical tagging, especially strong chemical tagging, specifically showing that within the light elements alone it is not possible to confidently separate field stars and conatal stars given similar stellar parameters and Galactic radii. The tightest abundance variation constraints in OCs may also help set limits on the rate of binary interactions and planetary engulfment in different environments.

## Acknowledgments

We thank the anonymous referee for the helpful and insightful comments. This material is based upon work supported by the National Science Foundation under grant No. AST-2206542. A.S. also acknowledges support from the University of Utah’s J. Irvin and Norma K. Swigart First-Year Summer Graduate Research Fellowship. P.F. acknowledges support from NSF grant AST-2206541, and K.C. acknowledges support from NSF grant AST-2206543.

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](http://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, Observatório 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.

Funding for the Sloan Digital Sky Survey V has been provided by the Alfred P. Sloan Foundation, the Heising-Simons Foundation, the National Science Foundation, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High Performance Computing at



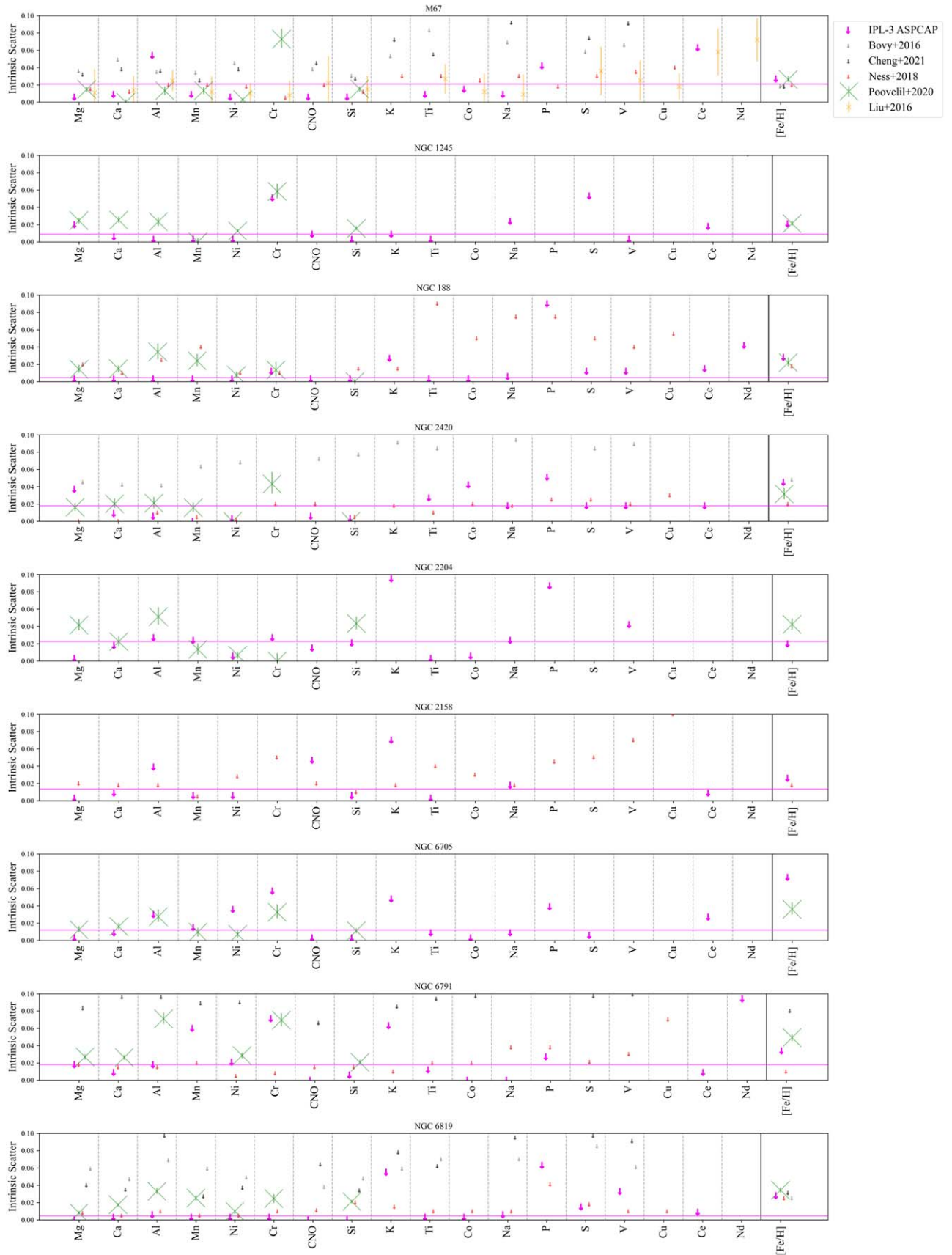
the University of Utah. The SDSS website is [www.sdss.org](http://www.sdss.org). SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration, including the Carnegie Institution for Science, Chilean National Time Allocation Committee (CNTAC) ratified researchers, the Gotham Participation Group, Harvard University, Heidelberg University, The Johns Hopkins University, L'Ecole polytechnique fédérale de Lausanne (EPFL), Leibniz-Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Extraterrestrische Physik (MPE), Nanjing University, National Astronomical Observatories of China (NAOC), New Mexico State University, The Ohio State University, Pennsylvania State University, Smithsonian Astrophysical Observatory, Space Telescope Science Institute (STScI), the Stellar

Astrophysics Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Illinois at Urbana-Champaign, University of Toronto, University of Utah, University of Virginia, and Yale University.

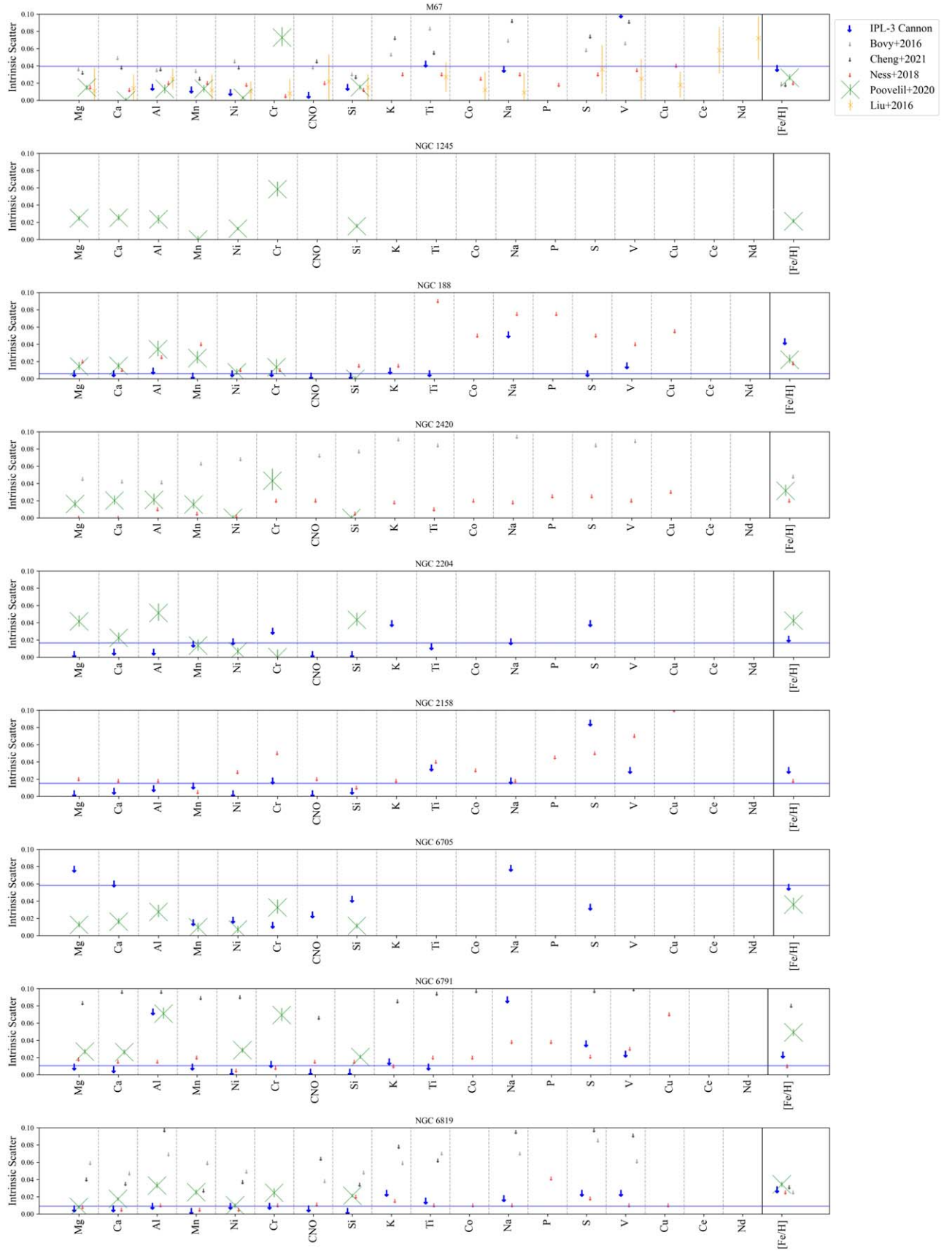
*Software:* Astropy (Astropy Collaboration et al. 2013, 2018, 2022), SciPy (P. Virtanen et al. 2020), Matplotlib (J. D. Hunter 2007), NumPy (C. R. Harris et al. 2020), pandas (W. McKinney 2010; pandas development team 2020), and seaborn (M. L. Waskom 2021).

## Appendix A

Here we show the literature comparison plots (Figures 8–11) for the well-studied OCs in our sample, as discussed in Section 5.

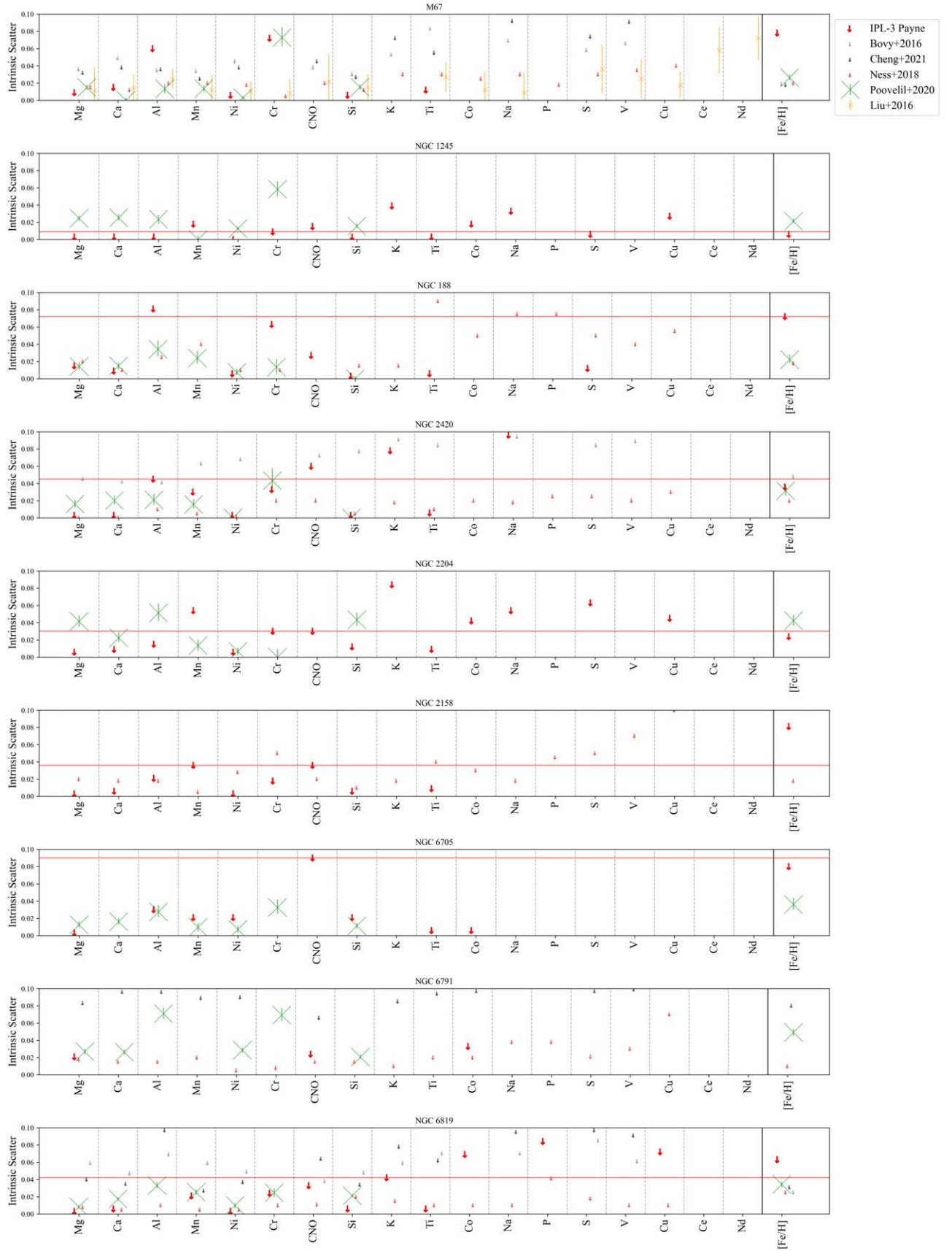


**Figure 8.** The literature comparison plots for the well-studied OCs in our sample using IPL-3 ASPCAP (colored in magenta). The horizontal line indicates the median 99.7% confidence limit for that cluster's homogeneity.

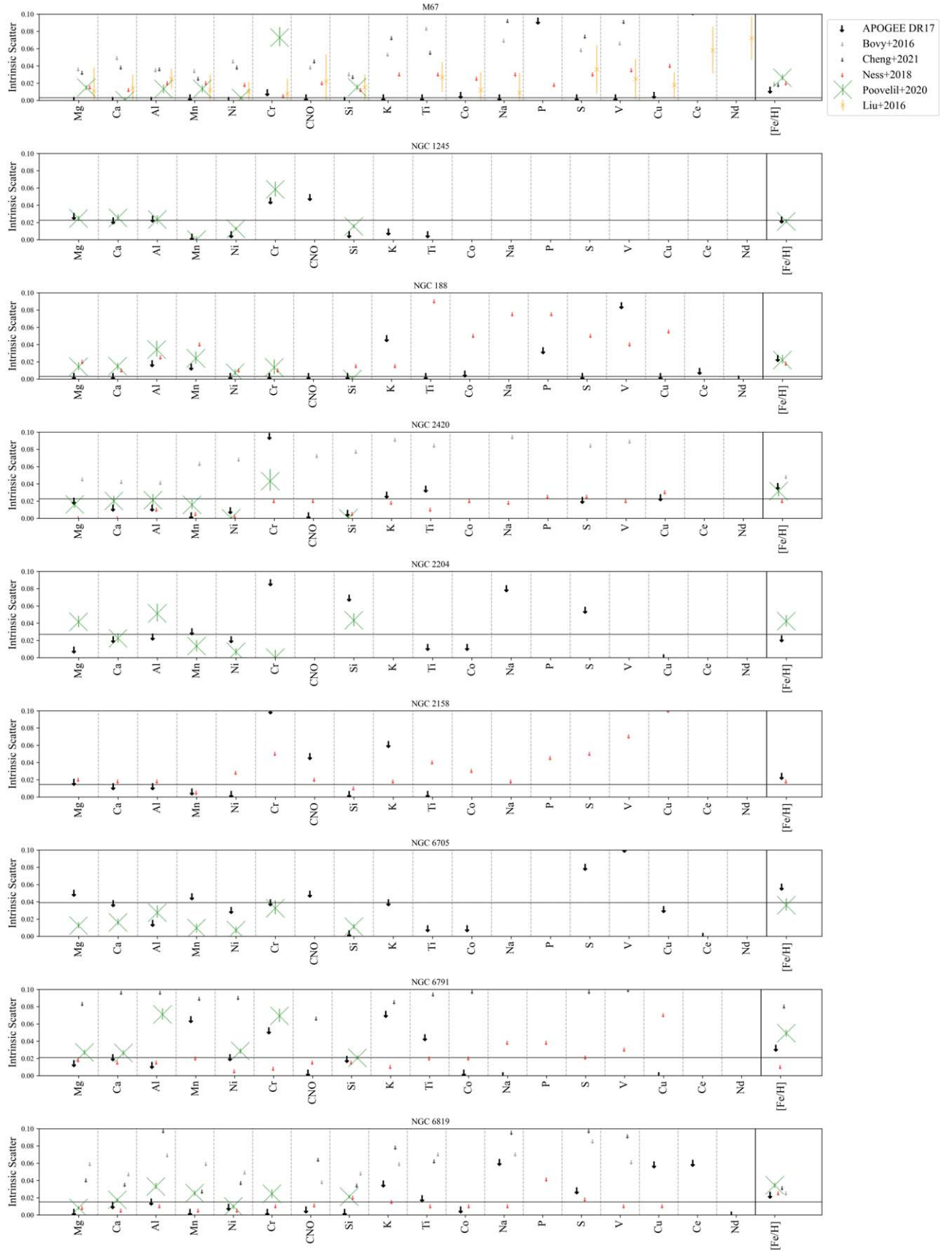


**Figure 9.** The literature comparison plots for the well-studied OCs in our sample using IPL-3 Cannon (colored in blue). The horizontal line indicates the median 99.7% confidence limit for that cluster's homogeneity.





**Figure 10.** The literature comparison plots for the well-studied OCs in our sample using IPL-3 Payne (colored in red). The horizontal line indicates the median 99.7% confidence limit for that cluster's homogeneity.



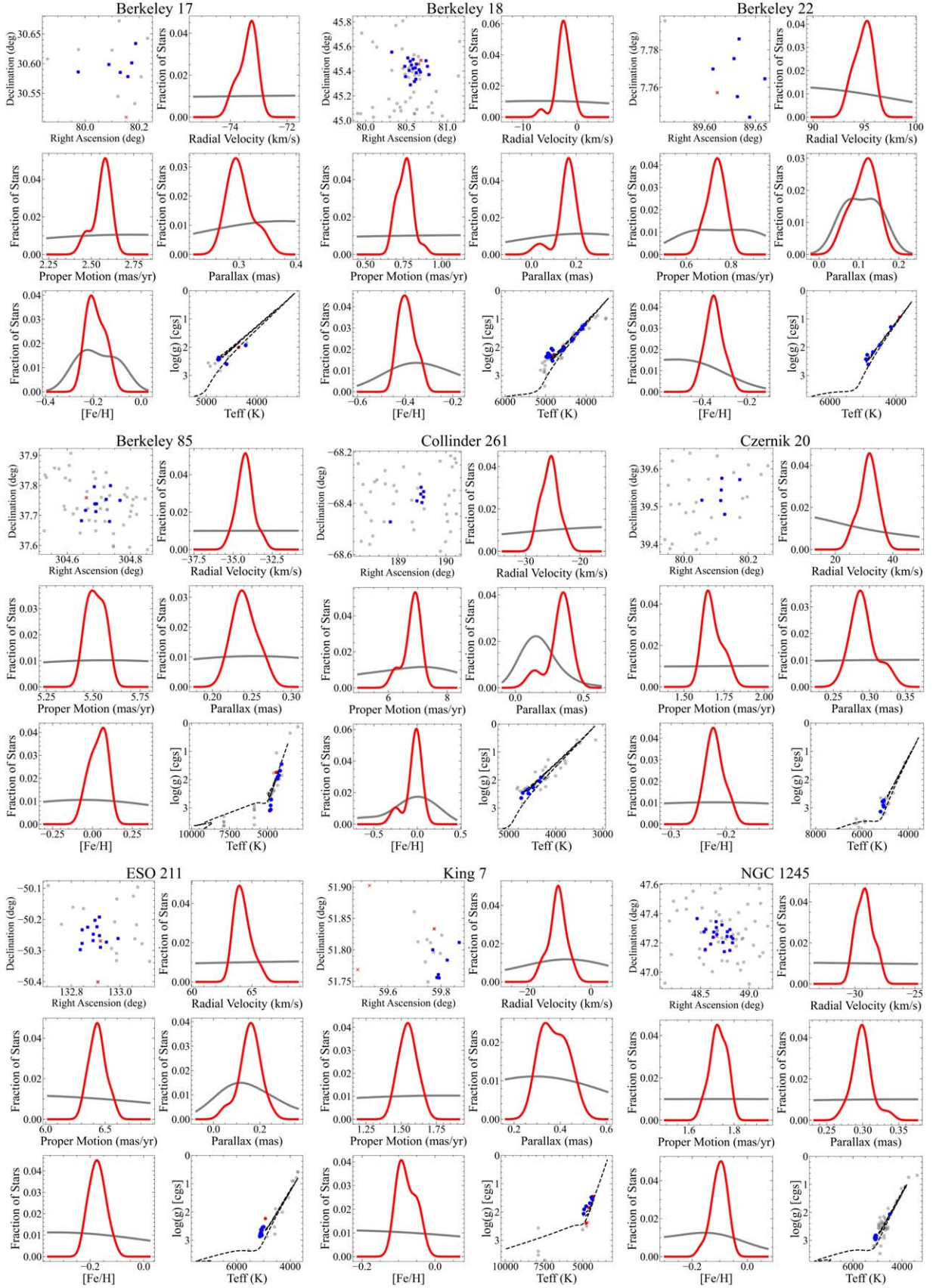
**Figure 11.** The literature comparison plots for the well-studied OCs in our sample using APOGEE DR17 (colored in black). The horizontal line indicates the median 99.7% confidence limit for that cluster's homogeneity.

## **Appendix B**

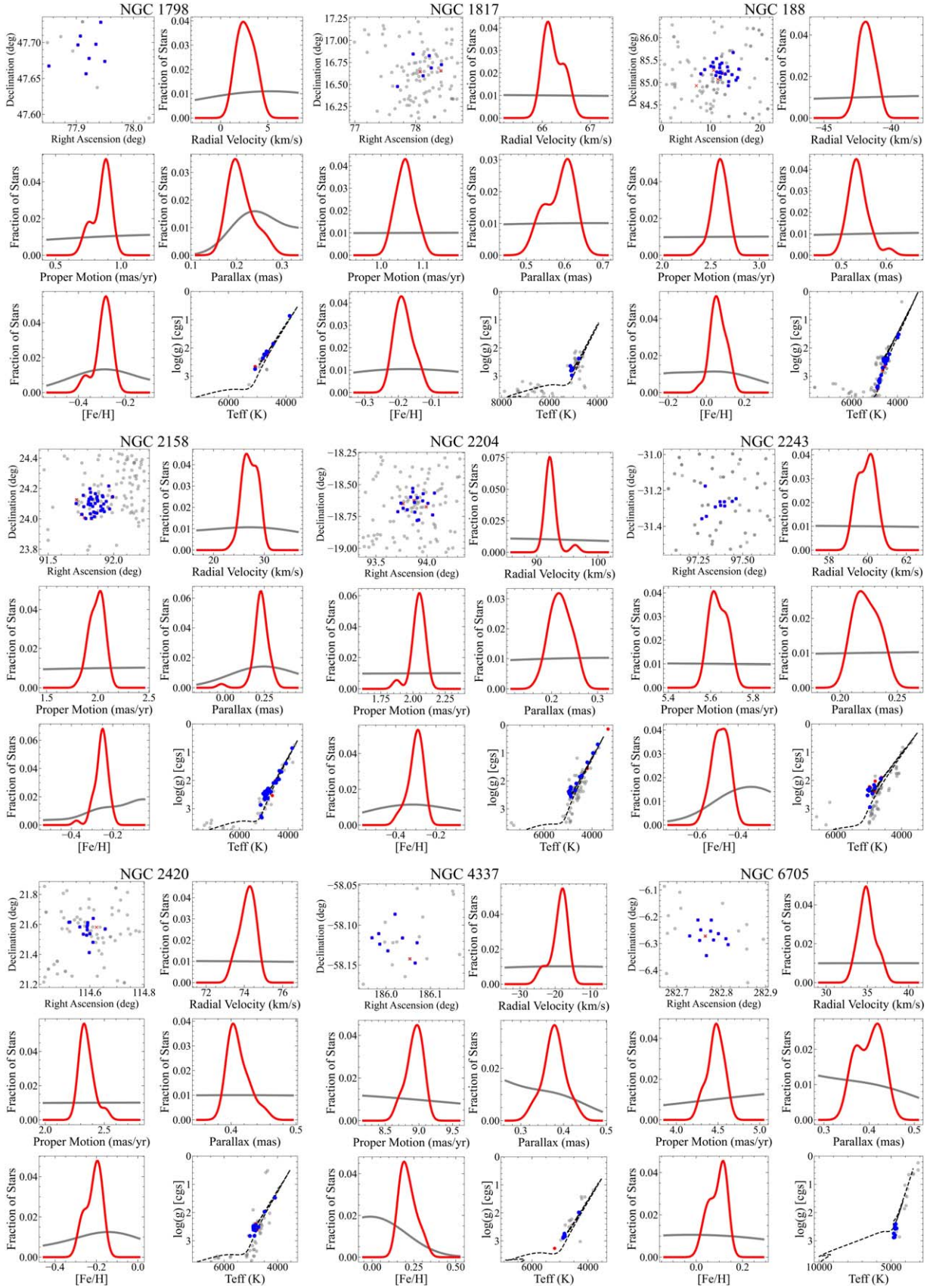
### **Open Cluster Sample Membership**

Here we show the plots (Figures [12–14](#)) outlining the kinematic selection of our OC sample, as described in Section [2.2.1](#).



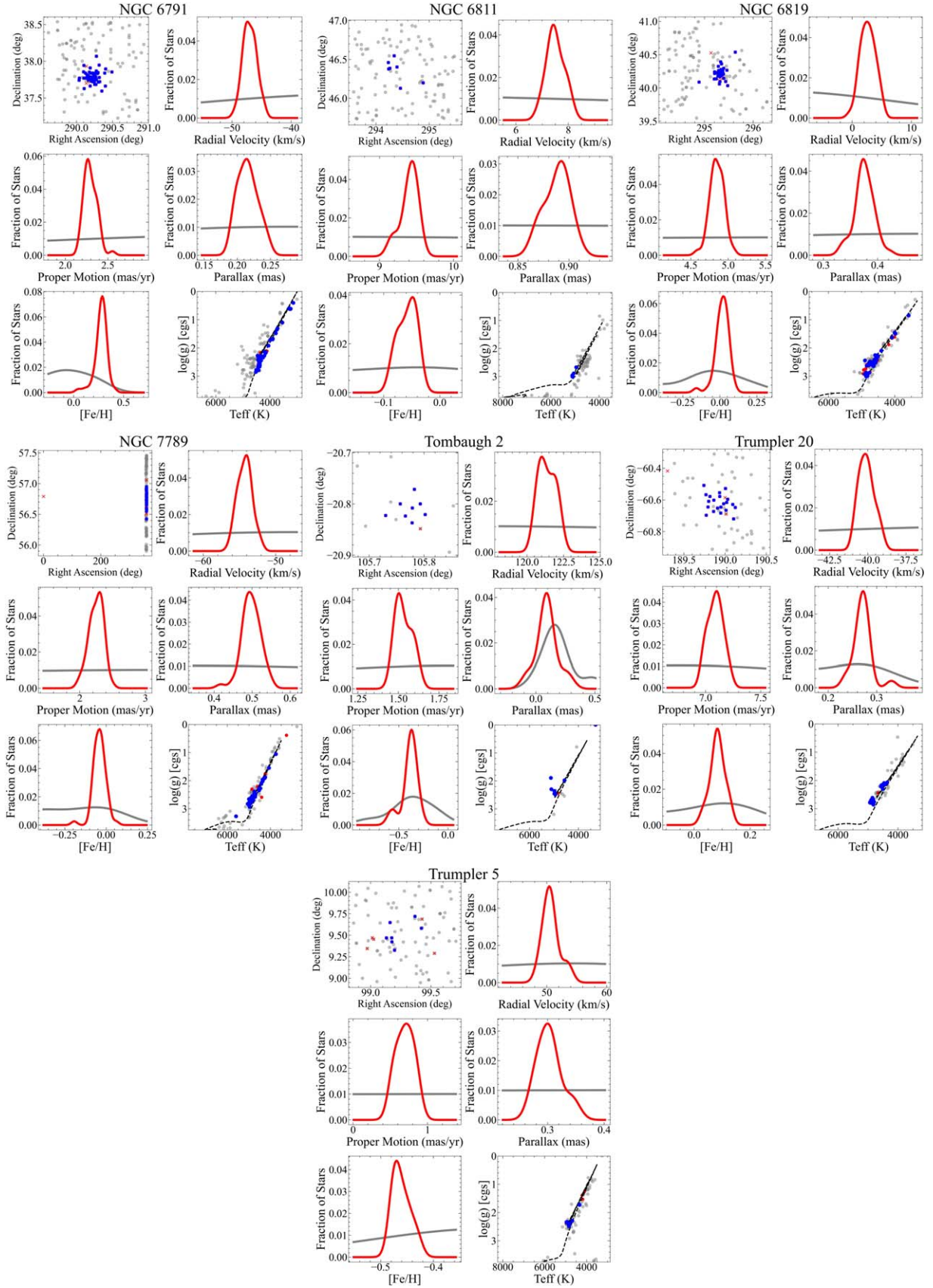


**Figure 12.** These figure follows the same layout as Figure 1 in Section 2.2.1.



**Figure 13.** This figure follows the same layout as Figure 1 in Section 2.2.1.





**Figure 14.** This figure follows the same layout as Figure 1 in Section 2.2.1.

## ORCID iDs

Amaya Sinha  <https://orcid.org/0009-0005-0182-7186>  
 Gail Zasowski  <https://orcid.org/0000-0001-6761-9359>  
 Peter Frinchaboy  <https://orcid.org/0000-0002-0740-8346>  
 Katia Cunha  <https://orcid.org/0000-0001-6476-0576>  
 Diogo Souto  <https://orcid.org/0000-0002-7883-5425>  
 Jamie Tayar  <https://orcid.org/0000-0002-4818-7885>  
 Keivan Stassun  <https://orcid.org/0000-0002-3481-9052>

## References

- Abate, C., Pols, O. R., Izzard, R. G., Mohamed, S. S., & de Mink, S. E. 2013, *A&A*, **552**, A26
- Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, *ApJS*, **259**, 35
- Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, *ApJS*, **235**, 42
- Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, *ApJS*, **249**, 3
- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, **219**, 12
- Albareti, F. D., Allende Prieto, C., Almeida, A., et al. 2017, *ApJS*, **233**, 25
- Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, *ApJ*, **636**, 804
- Almeida, A., Anderson, S. F., Argudo-Fernández, M., et al. 2023, *ApJS*, **267**, 44
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, **935**, 167
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Badenes, C., Mazzola, C., Thompson, T. A., et al. 2018, *ApJ*, **854**, 147
- Barkhouser, R. H., Smee, S. A., Hammond, R. P., et al. 2022, *Proc. SPIE*, **12182**, 121823O
- Bastian, N., Lamers, H. J. G. L. M., de Mink, S. E., et al. 2013, *MNRAS*, **436**, 2398
- Beaton, R. L., Oelkers, R. J., Hayes, C. R., et al. 2021, *AJ*, **162**, 302
- Bhattarai, B., Loebman, S. R., Ness, M. K., et al. 2024, [arXiv:2408.02228](https://arxiv.org/abs/2408.02228)
- Blanton, M. R., Bershad, M. A., Abolfathi, B., et al. 2017, *AJ*, **154**, 28
- Bovy, J. 2015, *ApJS*, **216**, 29
- Bovy, J. 2016, *ApJ*, **817**, 49
- Bowen, I. S., & Vaughan, A. H. J. 1973, *ApOpt*, **12**, 1430
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, **427**, 127
- Buder, S., Sharma, S., Kos, J., et al. 2021, *MNRAS*, **506**, 150
- Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, *A&A*, **640**, A1
- Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, *A&A*, **618**, A93
- Carlberg, J. K., Smith, V. V., Cunha, K., Majewski, S. R., & Rood, R. T. 2010, *ApJL*, **723**, L103
- Carrera, R., Casamiquela, L., Carbajo-Hijarrubia, J., et al. 2022, *A&A*, **658**, A14
- Cheng, C. M., Price-Jones, N., & Bovy, J. 2021, *MNRAS*, **506**, 5573
- Chiappini, C., Matteucci, F., & Romano, D. 2001, *ApJ*, **554**, 1044
- Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, **823**, 102
- Cunha, K., Smith, V. V., Johnson, J. A., et al. 2015, *ApJL*, **798**, L41
- De Silva, G. M., Freeman, K. C., Asplund, M., et al. 2007, *AJ*, **133**, 1161
- De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, *MNRAS*, **449**, 2604
- Dotter, A. 2016, *ApJS*, **222**, 8
- Feng, Y., & Krumholz, M. R. 2014, *Natur*, **513**, 523
- Freeman, K., & Bland-Hawthorn, J. 2002, *ARA&A*, **40**, 487
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, **616**, A1
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, *A&A*, **649**, A1
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, **595**, A1
- Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, *A&A*, **674**, A1
- García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, *AJ*, **151**, 144
- Geisler, D., Villanova, S., Carraro, G., et al. 2012, *ApJL*, **756**, L40
- Gratton, R. G., Bonifacio, P., Bragaglia, A., et al. 2001, *A&A*, **369**, 87
- Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, *AJ*, **131**, 2332
- Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, *A&A*, **486**, 951
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hawkins, K., Lucey, M., Ting, Y.-S., et al. 2020, *MNRAS*, **492**, 1164
- Hayes, C. R., Masseron, T., Sobeck, J., et al. 2022, *ApJS*, **262**, 34
- Holtzman, J. A., Hasselquist, S., Shetrone, M., et al. 2018, *AJ*, **156**, 125
- Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, *MNRAS*, **445**, 581
- Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, *MNRAS*, **480**, 800
- Hubeny, I., & Lanz, T. 2017, [arXiv:1706.01859](https://arxiv.org/abs/1706.01859)
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, *AJ*, **160**, 120
- Katz, D., Sartoretti, P., Guerrier, A., et al. 2023, *A&A*, **674**, A5
- Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, [arXiv:1711.03234](https://arxiv.org/abs/1711.03234)
- Kos, J., Lin, J., Zwitter, T., et al. 2017, *MNRAS*, **464**, 1259
- Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, *ARA&A*, **57**, 227
- Lada, C. J., & Lada, E. A. 2003, *ARA&A*, **41**, 57
- Lagarde, N., Decressin, T., Charbonnel, C., et al. 2012, *A&A*, **543**, A108
- Laughlin, G., & Adams, F. C. 1997, *ApJL*, **491**, L51
- Liu, F., Asplund, M., Yong, D., et al. 2016, *MNRAS*, **463**, 696
- Loaiza-Tacuri, V., Cunha, K., Souto, D., et al. 2023, *MNRAS*, **526**, 2378
- Looney, L. W., Tobin, J. J., & Fields, B. D. 2006, *ApJ*, **652**, 1755
- Mackereith, J. T., & Bovy, J. 2018, *PASP*, **130**, 114501
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, **154**, 94
- Manea, C., Hawkins, K., Ness, M. K., et al. 2024, *ApJ*, **972**, 69
- Masseron, T., Merle, T., & Hawkins, K. 2016, BACCHUS: Brussels Automatic Code for Characterizing High accuracy Spectra, Astrophysics Source Code Library, [ascl:1605.004](https://ascl.net/1605.004)
- Matteucci, F., & Francois, P. 1989, *MNRAS*, **239**, 885
- McKinney, W. 2010, in *Proc. 9th Python in Science Conf.*, ed. S. van der Walt & J. Millman (Austin, TX: SciPy), 56
- Mészáros, S., Holtzman, J., García Pérez, A. E., et al. 2013, *AJ*, **146**, 133
- Michaud, G., Alecian, G., & Richer, J. 2015, *Atomic Diffusion in Stars* (Cham: Springer)
- Milliman, K. E., Mathieu, R. D., & Schuler, S. C. 2015, *AJ*, **150**, 84
- Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., & Zasowski, G. 2015, *ApJ*, **808**, 16
- Ness, M., Rix, H. W., Hogg, D. W., et al. 2018, *ApJ*, **853**, 198
- Ness, M. K., Wheeler, A. J., McKinnon, K., et al. 2022, *ApJ*, **926**, 144
- Nidever, D. L., Holtzman, J. A., Allende Prieto, C., et al. 2015, *AJ*, **150**, 173
- Osorio, Y., Allende Prieto, C., Hubeny, I., Mészáros, S., & Shetrone, M. 2020, *A&A*, **637**, A80
- pandas development team 2020, pandas-dev/pandas: Pandas, v2.2.2, Zenodo, [doi:10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134)
- Pinsonneault, M. H., DePoy, D. L., & Coffee, M. 2001, *ApJL*, **556**, L59
- Plez, B. 2012, Turbospectrum: Code for Spectral Synthesis, Astrophysics Source Code Library, [ascl:1205.004](https://ascl.net/1205.004)
- Pogge, R. W., Derwent, M. A., O'Brien, T. P., et al. 2020, *Proc. SPIE*, **11447**, 1144781
- Poovelil, V. J., Zasowski, G., Hasselquist, S., et al. 2020, *ApJ*, **903**, 55
- Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., et al. 2020, *ApJ*, **895**, 2
- Santana, F. A., Beaton, R. L., Covey, K. R., et al. 2021, *AJ*, **162**, 303
- Sheinis, A., Anguiano, B., Asplund, M., et al. 2015, *JATIS*, **1**, 035002
- Smith, V. V., Bizyaev, D., Cunha, K., et al. 2021, *AJ*, **161**, 254
- Smith, V. V., & Lambert, D. L. 1990, *ApJS*, **72**, 387
- Souto, D., Allende Prieto, C., Cunha, K., et al. 2019, *ApJ*, **874**, 97
- Spitoni, E., & Matteucci, F. 2011, *A&A*, **531**, A72
- Steinmetz, M., Guiglion, G., McMillan, P. J., et al. 2020a, *AJ*, **160**, 83
- Steinmetz, M., Matijević, G., Enke, H., et al. 2020b, *AJ*, **160**, 82
- Thorley, J., & Schwarz, C. 2018, *JOSS*, **3**, 1082
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2019, *ApJ*, **879**, 69
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Waskom, M. L. 2021, *JOSS*, **6**, 3021
- Weinberg, D. H., Holtzman, J. A., Hasselquist, S., et al. 2019, *ApJ*, **874**, 102
- Wetzel, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, *ApJL*, **827**, L23
- Wilson, J. C., Hearty, F. R., Skrutskie, M. F., et al. 2019, *PASP*, **131**, 055001
- Zasowski, G., Cohen, R. E., Chojnowski, S. D., et al. 2017, *AJ*, **154**, 198
- Zasowski, G., Johnson, J. A., Frinchaboy, P. M., et al. 2013, *AJ*, **146**, 81
- Zhao, G., Zhao, Y., Chu, Y., Jing, Y., & Deng, L. 2012, [arXiv:1206.3569](https://arxiv.org/abs/1206.3569)
- Zwitter, T., Kos, J., Buder, S., et al. 2021, *MNRAS*, **508**, 4202