

To appear on arXiv on 13 Sep 2016 (submission in 1 days)

The *unRAVE* catalog

Andrew R. Casey

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA

Some combination of: Harry Enke, Kenneth C. Freeman, Gerry Gilmore, Keith Hawkins, David W. Hogg, Georges Kordopatis, Gal Matijevic, Melissa Ness, Jason Sanders, Matthias Steinmetz, Hans Walter-Rix

Luca Casagrande, Andrea Kunder, Paul McMillan, Alessandro Siviero, Jennifer Wojno, Tomaž Zwitter

Joss Bland-Hawthorn, Brad K. Gibson, Arnaud Siebert, Olivier Bienayme, Julio F. Navarro, Ulisse Munari, Warren Reid, Rosemary Wyse, Amina Helmi

and the *RAVE* collaboration

ABSTRACT

The Milky Way is a powerful laboratory for understanding galaxy formation, as it can provide orbits, ages, atmospheric parameters, as well as chemical abundances for vast sets of individual stars. These inferences require both astrometric and spectroscopic data. Indeed, in order to be maximally useful for chemo-dynamic studies, the Tycho-Gaia Astrometric Solution (*TGAS*) sample in the first *Gaia* data release requires a spectroscopic complement that includes radial velocities, stellar parameters, and elemental abundances. Among existing spectroscopic samples, the RAdial Velocity Experiment (*RAVE*) survey has the largest overlap with *TGAS*: up to 292,036 stars. Here we present a data-driven re-analysis of *RAVE* spectra, using an implementation of *The Cannon*, that yields more precise and accurate stellar parameters and abundances than previous *RAVE* data releases. We constructed our model using high-fidelity *APOGEE* stellar parameters and abundances for red giant branch stars that overlap with *RAVE*, and stellar parameters from the *K2/Ecliptic Pole Input Catalog* for main-sequence and sub-giant branch stars. We derive, and validate, improved effective temperature T_{eff} , surface gravity $\log g$, and chemical abundances of up to seven elements (O, Mg, Al, Si, Ca, Fe, Ni) for 434,470 stars. The typical precision in chemical abundances is X.XX dex for red giant branch stars. The *unRAVE* catalog presented here, in conjunction with *TGAS*, represents the most powerful data set for chemo-orbital analysis at the dawn of broadly available *Gaia* data.

Subject headings:

1. Introduction

The Milky Way is considered to be our best laboratory for understanding galaxy formation and evolution. This premise hinges on the ability to precisely measure the astrometry and chemistry for (many) individual stars, and to use those data to infer the structure, kinematics, and chemical enrichment of the Galaxy (e.g., Nordström et al. 2004; Schlaufman et al. 2009; Deason et al. 2011; Casagrande et al. 2011; Ness et al. 2012, 2013a,b; Casey et al. 2012, 2013, 2014a,b; Boeche et al. 2013; Kordopatis et al. 2015; Bovy et al. 2016). However, these quantities are not known for even 1% of stars in the Milky Way. Stellar distances are famously imprecise (e.g., van Leeuwen 2007; Jofré et al. 2015; Mädler et al. 2016), proper motions can be plagued by unquantified systematics from the first epoch observations (e.g., Casey & Schlaufman 2015), and stellar spectroscopists frequently report significantly different chemical abundance patterns from the same spectrum (Smiljanic et al. 2014). The impact these issues have on scientific inferences cannot be understated. Imperfect astrometry or chemistry limits understanding in a number of sub-fields in astrophysics, including the properties of exoplanet host stars, the formation (and destruction) of stars and clusters, as well as the study of stellar populations and Galactic structure, to name a few.

The *Gaia* mission represents a critical step forward in understanding the Galaxy. *Gaia* is primarily an astrometric mission, and will provide precise positions, parallaxes and proper motions for more than 10^9 stars in its final data release in 2022. While this is a sample size about four orders of magnitude larger than its predecessor *Hipparcos*, both astrometry and chemistry are required to fully characterize the formation and evolution of our Galaxy. *Gaia* will also provide radial velocities, stellar parameters and chemical abundances for a subset of brighter stars, but these measurements will not be available in the first few data releases. Until those abundances are available, astronomers seeking to simultaneously use chemical and dynamical information are reliant on ground-based spectroscopic surveys to complement the available *Gaia* astrometry.

The first *Gaia* data release will include the Tycho-Gaia Astrometric Solution (hereafter *TGAS*; Michalik et al. 2015a,b): positions, proper motions, and parallaxes for approximately two million stars in the Tycho-2 (Høg et al. 2000) catalog. After cross-matching all major stellar spectroscopic surveys¹, we found that the RAdial Velocity Experiment (*RAVE*) survey is expected to have the largest overlap with the first *Gaia* data release: up to 292,036 stars. We used the *Gaia* universe model snapshot (Robin et al. 2012) to estimate the precision in parallax and proper motions that could be available in the first *Gaia* data release (DR1) for stars in those overlap samples. Comparing the expected precision to what is currently available, we further found that the *RAVE* survey will benefit most from *Gaia* DR1: 63% of stars in the *RAVE*-*Gaia* DR1 overlap sample ($\approx 182,862$ stars) are expected to improve with the first *Gaia* data release, and 47% of stars ($\approx 137,211$ stars) are likely to have better proper motions. Although the *Gaia* universe model assumes end-of-mission

¹Specifically we cross-matched the Tycho-2 catalog against the *APOGEE* (Zasowski et al. 2013), *Gaia-ESO* (Gilmore et al. 2012; Randich et al. 2013), *GALAH* (De Silva et al. 2015), *LAMOST* (Cui et al. 2012), and *RAVE* (Steinmetz et al. 2006) surveys

uncertainties — and does not account for systematics in the first data release — this calculation still provides intuition for the relative improvement the first *Gaia* data release can make to ground-based surveys. The expected improvements for *RAVE* motivated us to examine what chemical abundances were available from those data, and to evaluate whether we could enable new chemo-dynamic studies by contributing to the existing set of chemical abundances.

We briefly describe the *RAVE* data in Section 2, before explaining our methods in Section 3. In Section 4 we outline a number of validation experiments, including: internal sanity checks, comparisons with literature samples, and investigations to ensure our results are consistent with expectations from astrophysics. We discuss the implications of these comparisons in Section 5, and conclude with instructions for how to access our results electronically.

2. Data

RAVE is a magnitude-limited stellar spectroscopic survey of the (nearby) Milky Way, principally designed to measure radial velocities for up to 10^6 stars. Observations were conducted on the 1.2 m UK Schmidt telescope at the Australian Astronomical Observatory² from 2003–2013. A large 5.7 degree field-of-view and robotic fibre positioner made for very efficient observing: spectra for up to 150 targets could be simultaneously acquired. When observations concluded in April 2013, at least 520,782 useful spectra had been collected of more than 457,589 unique objects.

The target selection for *RAVE* is based on the I -band apparent magnitude, $9 < I < 12$, with a weak $J - K_s > 0.5$ cut near the disk and bulge (Wojno et al. 2016). The I band was used for the target selection because it has a good overlap with the wavelength range that *RAVE* operates in: 8410–8795 Å. The resolution and wavelength coverage is comparable to the Radial Velocity Spectrometer on board the *Gaia* space telescope (Munari et al. 2005; Kordopatis et al. 2011; Recio-Blanco et al. 2016), and the wavelength range overlaps with one of the key setups used for the ground-based high-resolution *Gaia-ESO* survey (Gilmore et al. 2012; Randich et al. 2013). The region includes the Ca II near infrared triplet lines — strong transitions that are dominated by pressure broadening — which are visible even in metal-poor stars or spectra with very low signal-to-noise (S/N) ratios. Atomic transitions of light-, α -, and Fe-peak elements are also present, allowing for detailed chemical abundance studies.

The exposure times for *RAVE* observations were optimised to obtain radial velocities for as many stars as possible. Detailed chemical abundances were always an important science goal of the survey, but this was a secondary objective. For this reason the distribution of S/N ratios in *RAVE* spectra is considerably lower than other stellar spectroscopic surveys where chemical abundances are the primary motivation. The *RAVE* spectra have an effective resolution $\mathcal{R} \approx 7,500$ and the distribution of S/N ratios peaks at ≈ 50 pixel⁻¹. For comparison, the *GALAH* survey (De Silva et

²Formerly the Anglo-Australian Observatory

al. 2015) — which was specifically constructed for detailed chemical abundance analyses — includes a wavelength range about 2.5 times larger at resolution $\mathcal{R} \approx 28,000$, and yet the *GALAH* project still targets for $S/N \gtrsim 100$ per resolution element.

Despite the relatively low resolution and S/N of the spectra compared to other surveys, the *RAVE* data releases have provided excellent radial velocities, stellar astrophysical parameters (T_{eff} , $\log g$), as well as individual chemical abundances (Steinmetz et al. 2006; Zwitter et al. 2008; Siebert et al. 2011; Boeche et al. 2011; Kordopatis et al. 2013; Kunder et al. 2016). In this work we make use of spectra that has been reprocessed for the fifth *RAVE* data release. These re-processing steps include: a detailed re-reduction of all the original data frames, with flux variances propagated at every step; an updated continuum-normalization procedure; as well as revised determinations of stellar radial velocities and morphological classifications. At the end of this processing for each survey observation we were provided with: rest-frame wavelengths (without resampling), continuum-normalized fluxes, 1σ uncertainties in the continuum-normalized flux values, as well as relevant metadata for each observation. We refer the reader to the official fifth data release paper of the *RAVE* survey, as presented by Kunder et al. (2016), for more details of this re-processing.

Given the high-quality of the normalization performed by the *RAVE* team, we chose not to re-normalize the spectra. Our tests demonstrated that the procedure outlined in Kunder et al. (2016) is sufficient for our analysis procedure. Therefore, there is a limited number of pre-processing steps that we performed before starting our analysis. First, we calculated inverse variance arrays from the 1σ uncertainties provided, and then we re-sampled the flux and inverse variance arrays onto a common rest-wavelength map for all stars. Depending on the fibre used and the stellar radial velocity, the range of rest-frame wavelength values varied for each star. Given that fluxes were unavailable in the edge pixels for most stars, we excluded pixels outside of the rest wavelength range $8423.2 \text{ \AA} \leq \lambda \leq 8777.6 \text{ \AA}$. This corresponds to about 30 pixels excluded on either side of the common wavelength array, leaving us with 945 pixels per spectrum for science.

3. Method

We chose to adopt a data-driven model for this analysis, rather than the physics-based models used in the *RAVE* data releases. Specifically, we will use an implementation of *The Cannon* (Ness et al. 2015, 2016). Although this choice complicated the construction of our model (e.g., see Section 3.2), a data-driven approach makes use of all available information in the spectrum and lowers the S/N ratio at which systematic effects begin to dominate. In other words, in the low S/N regime, a well-constructed data-driven model will yield more precise *labels* (e.g., stellar parameters and chemical abundances) than most physics-driven models³. This is particularly relevant for the low-resolution *RAVE* data analysed here, because about half of the spectra have $S/N \lesssim 50 \text{ pixel}^{-1}$.

³However, see Casey (2016)

There are two main analysis steps when using *The Cannon*: the *training* step and the *test* step. We describe these stages in the context of our model in the following section, and a more thorough introduction can be found in Ness et al. (2015). We make the following explicit assumptions about the RAVE spectra and *The Cannon*:

- We assume that any fibre- and time-dependent variations in spectral resolution in the RAVE spectra are negligible.
- The RAVE noise variances are approximately correct, independent between pixels, and normally distributed.
- We assume that the normalization procedure employed by the RAVE pipeline is invariant with respect to the labels we seek to measure (e.g., T_{eff} , $\log g$, or [Fe/H]), and invariant with respect to the S/N ratio. In other words, we assume that the normalization procedure does not produce different results for high S/N spectra compared to low S/N spectra, nor does the normalization procedure vary non-linearly with respect to stellar parameters (e.g., [Fe/H]).
- We assume that stars with similar labels (T_{eff} , $\log g$, and abundances) have similar spectra.
- A stellar spectrum is a smooth function of the label values for that star, and we assume that the function is smooth enough within a sub-space of the labels (e.g., the giant branch or the main-sequence) that it can be reasonably approximated with a low-order polynomial in label space.
- The training set (Section 3.2) has mean accurate labels. Note that we do not assume that *every* label in the training is accurate – we can afford to have a small fraction of inaccurate labels (e.g., a few obvious misclassifications in the training set are affordable).
- We assume that the training data are similar (in spectra) to the test data where they overlap in label space, and we assume that the training data spans enough of the label space to capture the variation in the test-set spectra.

3.1. The model

Given our assumptions, the model we adopt is

$$y_{jn} = \mathbf{v}(\ell_n) \cdot \boldsymbol{\theta}_j + e_{jn} , \quad (1)$$

where y_{jn} is the pseudo-continuum-normalized flux for star n at wavelength pixel j , $\mathbf{v}(\ell_n)$ is the vectorizing function that takes as input the K labels ℓ_n for star n and outputs functions of those labels as a vector of length $D > K$, $\boldsymbol{\theta}_j$ is a vector of length D of parameters influencing the model at wavelength pixel j , and e_{jn} is the residual (noise). Here we will only consider vectorizing functions

with second-order polynomial expansions (e.g., T_{eff}^2 , see Sections 3.3-3.5). The noise values e_{jn} can be considered to be drawn from a Gaussian distribution with zero mean and variance $\sigma_{jn}^2 + s_j^2$, where σ_{jn}^2 is the variance in flux y_{jn} and s_j^2 describes the excess variance at the j -th wavelength pixel.

At the *training* step we fix the K -lists of labels for the n training set stars. At each wavelength pixel j , we then find the parameters $\boldsymbol{\theta}_j$ and s_j^2 by optimizing the penalized likelihood function

$$\boldsymbol{\theta}_j, s_j^2 \leftarrow \underset{\boldsymbol{\theta}, s}{\operatorname{argmin}} \left[\sum_{n=0}^{N-1} \frac{[y_{jn} - \mathbf{v}(\ell_n) \cdot \boldsymbol{\theta}]^2}{\sigma_{jn}^2 + s^2} + \sum_{n=0}^{N-1} \ln(\sigma_{jn}^2 + s^2) + \Lambda Q(\boldsymbol{\theta}) \right] , \quad (2)$$

where Λ is a regularization parameter which we will heuristically set in later sections, and $Q(\boldsymbol{\theta})$ is a L1 regularizing function that encourages $\boldsymbol{\theta}$ values to take on zero values without breaking convexity (Casey et al. 2016):

$$Q(\boldsymbol{\theta}) = \sum_{d=1}^{D-1} |\theta_d| . \quad (3)$$

Note that the d subscript here is zero-indexed; the function $Q(\boldsymbol{\theta})$ does not act on the (first) θ_0 coefficient, as this is a ‘pivot point’ (mean flux value) that we do not expect to diminish with increasing regularization (e.g., see equation 5). In practice we first fix $s_j^2 = 0$ to make equation 2 a convex optimization problem, then we optimize for $\boldsymbol{\theta}_j$, before solving for s_j^2 .

The *test step* is where we fix the parameters $\boldsymbol{\theta}_j, s_j^2$ at all wavelength pixels j , and optimize the K -list of labels ℓ_m for the m -th test set star. Here the objective function is:

$$\ell_m \leftarrow \underset{\ell}{\operatorname{argmin}} \left[\sum_{j=0}^{J-1} \frac{[y_{jm} - \mathbf{v}(\ell) \cdot \boldsymbol{\theta}_j]^2}{\sigma_{jm}^2 + s_j^2} \right] . \quad (4)$$

After optimizing equation 4 for the m -th star we store the covariance matrix Σ_m for the labels ℓ_m , which provides us with the formal errors on ℓ_m . The formal errors are expected to be under-estimated, and in Section 4 we judge the veracity of these errors through validation experiments.

3.2. The training set

We sought to construct a training set of stars across the main-sequence, the sub-giant branch, and the red giant branch. We required stars with precisely measured effective temperature T_{eff} , surface gravity $\log g$, and elemental abundances of O, Mg, Si, Ca, Al, Fe, and Ni. This proved to be difficult because the magnitude range of *RAVE* does not overlap substantially with high-resolution spectroscopic surveys. The fourth internal data release of the *Gaia-ESO* survey includes

giant and main-sequence stars, but only 142 overlap with *RAVE*, which is too small to be a useful training set for our purposes. The thirteenth data release from the *Sloan Digital Sky Survey* (SDSS Collaboration et al. 2016) includes labels for *APOGEE* stars on the giant branch and (uncalibrated values for) the main-sequence, but our tests indicated that the *APOGEE* main-sequence labels suffered from significant systematic effects. A flat, then ‘up-turning’ main-sequence is present, and the metallicity gradient trends in the opposite direction with respect to $\log g$ on the main-sequence (i.e., metal-poor stars incorrectly sit above an isochrone in a classical Hertzsprung-Russell diagram). If we consider lower-resolution studies as potential training sets, there are 2,369 stars that overlap with *LAMOST* — of which 2,213 have positive S/N ratios in the *g*-band (**snrg**). However, the labels are expectedly less precise given the lower resolution, there are no elemental abundances available for the main-sequence stars⁴, and the *LAMOST* lower main-sequence suffers from the same systematic effects seen in *APOGEE* data.

These constraints forced us to construct a heterogeneous training set. Given previous successes in transferring high S/N ratio labels from *APOGEE* (Ness et al. 2015, 2016; Ho et al. 2016; Casey et al. 2016), we chose to use the 1,355 stars in the *APOGEE*—*RAVE* overlap sample for giant star labels in the training set. Of these, about 900 are giants according to *APOGEE*. From this sample we selected stars to have: determinations in all abundances of interest ($[X/H] > -5$ for O, Mg, Al, Si, Ca, Fe, and Ni); S/N ratios of >200 in *APOGEE* and >25 in *RAVE*; and we further required that *ASPCAP* did not report any peculiar flags (**ASPCAPFLAG = 0**). These restrictions left us with 536 stars along the giant branch, with metallicities ranging from $[Fe/H] = -1.79$ to 0.26. Intermediate tests with globular cluster members showed that the metallicity range of the training set needed to extend at least below $[Fe/H] \lesssim -2$ in order for our catalog to be practically useful. Without additional metal-poor stars, the lowest metallicity labels reported by our model would be around $[Fe/H] \approx -2$, even for well studied stars with $[Fe/H] \sim -4$ (e.g., CD 38-245). For this reason we supplemented our sample of *APOGEE* giant stars with 176 known metal-poor giant stars observed by *RAVE*. The effective temperature T_{eff} , surface gravity $\log g$ and iron abundance $[Fe/H]$ labels were adopted from Fulbright et al. (2010); Ruchti et al. (2011). The elemental abundances for O, Mg, Al, Si, Ca, and Ni were assumed to follow typical trends of Galactic chemical evolution, such that we assumed $[Mg/Fe] = +0.4$, $[O/Fe] = +0.4$, $[Al/Fe] = -0.5$, $[Ca/Fe] = +0.4$, $[Si/Fe] = +0.4$, and $[Ni/Fe] = -0.25$. This decision is made solely to ensure that our overall metallicity scale extends that of the *RAVE* survey, down to $[Fe/H] \sim -4$. In other words, we do not recommend the use of individual abundance labels at $[Fe/H] \sim -4$; in practice the abundances of most of these elements are largely unrecoverable in the *RAVE* wavelength region at $[Fe/H] \sim -4$. We discuss this issue in more detail in Section 5.

Assembling a suitable training set for the main-sequence and sub-giant branch was less trivial. There are no spectroscopic studies that extend the range of stellar types we are interested in

⁴Abundance information is available for *LAMOST* stars from Ho et al. (2016), but that sample contains only giant stars

(e.g., FGKM-type stars), and which have a large enough sample size overlapping with *RAVE*. Moreover, most of the spectroscopic studies we considered also showed a flat lower main-sequence, a systematic consequence of the analysis method adopted (see Bensby et al. 2014, for discussion on this issue). For these reasons we chose to make use of the *K2/EPIC* catalog (Huber et al. 2016) for the training set labels on the main-sequence and sub-giant branch. The *K2/EPIC* catalog follows from the successful *Kepler* input catalog (Brown et al. 2011), and provides probabilistic stellar classifications for 138,600 stars in the *K2* fields based on the astrometric, asteroseismic, photometric, and spectroscopic information available for every star. There are 4,611 stars that overlap between *K2/EPIC* and *RAVE*.

K2/EPIC differs from the *Kepler* input catalog because *K2/EPIC* does not benefit from having narrow-band *DDO*₅₁ photometry in order to aid dwarf/giant classification. Despite this limitation the labels in the *K2/EPIC* catalog have already been shown to be accurate and trustworthy (Huber et al. 2016). However, when the posteriors are wide (i.e., the quoted confidence intervals are large) due to limited information available, it is possible that a star has been misclassified. This is most prevalent for sub-giants, where Huber et al. (2016) note that $\approx 55 - 70\%$ of sub-giants are misclassified as dwarfs. The probability of misclassification is usually quantified in the uncertainties given for each star; most dwarfs that have a higher possibility of being sub-giants have large confidence intervals. Therefore, requiring low uncertainties will decrease the total sample size, but in practice it removes most misclassifications. The situation is far more favourable for dwarfs and giants. Only 1 – 4% of giant stars are misclassified as dwarfs, and about 7% of dwarfs are misclassified as giants. To summarise, the *K2/EPIC* labels with narrow confidence intervals are usually of high fidelity, and given that we have spectra, we can identify any spurious misclassifications.

We sought to have a small overlap between our giant and main-sequence star training sets. Most of our giant training set is encapsulated within $0 < \log g < 3.5$, however there is a sparse sampling of stars reaching to $\log g \approx 4$. We required $\log g > 3.5$ for the *K2/EPIC* main-sequence/sub-giant star training set, allowing for ≈ 0.5 dex of overlap between the two training sets. We further employed the following quality constraints: the upper and lower confidence intervals in T_{eff} must be below 150 K; the upper and lower confidence intervals in $\log g$ must be less than 0.15 dex; the *S/N* of the *RAVE* spectra must exceed 30 pixel⁻¹; and $T_{\text{eff}} \leq 6750$ K. Unfortunately these strict constraints removed most metal-poor stars, which we later found to cause the test labels to have under-predicted abundances for dwarfs of low metallicity. For this reason we relaxed (ignored) those quality constraints for stars with $[\text{Fe}/\text{H}] < -1$, and included an additional 12 turn-off stars with $-1.6 \gtrsim [\text{Fe}/\text{H}] \gtrsim -2.1$ from Ruchti et al. (2011). After training a model based on main-sequence and giant stars (Section 3.1), we found we could identify misclassifications by leave-one-out cross-validation. However, we chose not to do this because the number of likely misclassifications in the training set was negligible ($\approx 1\%$), and the improvement in main-sequence test set labels was minimal. The distilled sample of the *RAVE*–*K2/EPIC* overlap catalog contains 595 stars (583 of 4,611 from *K2/EPIC*). The full training set for each model (see next sections) is shown in Figure 1.

3.3. A 3-label model (T_{eff} , $\log g$, [Fe/H]) for all stars

We have constructed a justified training set for stars across the main-sequence, sub-giant, and red giant branch. However the lack of overlap between *RAVE* and other works have resulted in a somewhat peculiar situation. Detailed abundances are available from *APOGEE* for all giant stars in our sample, but only imprecise metallicities are available from *K2/EPIC* for stars on the main-sequence and the sub-giant branch. Here we will construct a simple model for *all* stars that only makes use of three labels (T_{eff} , $\log g$, [Fe/H]), before we outline how we derive abundances for giant branch stars. The complexity for this model will be quadratic (T_{eff}^2 is the highest term), where the vectorizer $\mathbf{v}(\ell_n)$ expands as,

$$\mathbf{v}(\ell_n) \rightarrow [1, T_{\text{eff},n}, \log g_n, [\text{Fe}/\text{H}]_n, T_{\text{eff},n}^2, \log g_n T_{\text{eff},n}, [\text{Fe}/\text{H}]_n T_{\text{eff},n}, \log g_n^2, [\text{Fe}/\text{H}]_n \log g_n, [\text{Fe}/\text{H}]_n^2] \quad (5)$$

such that $\mathbf{v}(\ell)$ produces the design matrix:

$$\mathbf{v}(\ell) \rightarrow \begin{bmatrix} \mathbf{v}(\ell_0) \\ \vdots \\ \mathbf{v}(\ell_{N-1}) \end{bmatrix} . \quad (6)$$

We used no regularization ($\Lambda = 0$) for this model. After training the model we treated all 520,782 spectra as test set objects. In the left-hand panel of Figure 2 we show the effective temperature T_{eff} and surface gravity $\log g$ for all spectra. The main-sequence and red giant branch are clearly visible. However, the details of stellar evolution are no longer present: the sub-giant branch is not discernible, and there are a number of systematic artefacts (over-densities) present in label space. These artefacts disappear when we require additional quality constraints (e.g., no peculiar morphological classifications), but the complexity of the Hertzsprung-Russell diagram is still not present. Thus, we concluded that while this model could be useful for deriving stellar classifications (e.g., F2-type giant), the labels are too imprecise.

We chose to adopt separate models for the main-sequence and the red giant branch rather than switch to a single model with higher complexity. This choice allowed us to derive stellar parameters for stars on the main-sequence and sub-giant branch, as well as detailed elemental abundances for red giant branch stars. However, adopting two separate models introduces the challenge of how to combine the results from two models, or how to assign one star as ‘belonging’ to a single model. In Section 3.6 we describe how we will use the 3-label model of the main-sequence and giant branch (introduced in this section) to discriminate between results from a 3-label main-sequence model in Section 3.4 and a 9-label giant star model in Section 3.5.

3.4. A 3-label model (T_{eff} , $\log g$, [Fe/H]) for unevolved stars

We constructed a three-label quadratic model using only main-sequence and sub-giant stars. In order to set the regularization hyperparameter Λ for this model, we trained 30 models with

different regularization strengths, spaced evenly in logarithmic steps between $\Lambda = 10^{-3}$ to $\Lambda = 10^3$. We then performed leave-one-out cross-validation for each model. Specifically, for each star in the training set: we removed the star; trained the model; and then inferred labels from the removed star as if it was a test object. We also performed leave-one-out cross-validation on an unregularized ($\Lambda = 0$) model, which we will use as the basis for comparison. For the unregularized case, we calculated the bias and root-mean-square (RMS) deviation between: the training set labels, and the labels we derive by cross-validation, where one star is removed at a time and the model is re-trained. We repeated this calculation of bias and RMS deviation for all 30 models with different regularization strengths Λ .

We show the *percentage difference* in the RMS deviation of the labels with respect to the unregularized model in Figure 3. The upper and lower envelope represent the boundaries across all labels, showing that with increasing regularization, the RMS decreased in *all* labels. We found similar improvements in the biases, however these were already minimal in the unregularized case. The improvement in RMS reaches a minimum value near $\Lambda = 35.6$ ($\approx 10^{1.5}$), where we achieve RMS deviations that are about 10% better than the unregularized case. Based on this improvement we set $\Lambda = 35.6$ for this model. At this regularization strength, the bias and RMS values found by leave-one-out cross validation are, respectively: 38 K and 256 K for T_{eff} , 0.05 dex and 0.29 dex for $\log g$, with 0.03 dex and 0.17 dex for [Fe/H].

We inferred labels for all *RAVE* spectra using this main-sequence/sub-giant star model. The results for the survey sample are shown in the centre panel of Figure 2. The increased density of Solar-type stars is consistent with *RAVE* observing stars in the local neighbourhood, and the high number of turn-off and main-sequence stars relative to the sub-giant branch is expected from the relative lifetimes of these evolutionary phases. An over-density of stars near the base of the giant branch is also present. This artefact is due to having giant stars in the test set, but not in the training set, and the model is (poorly) extrapolating outside the convex hull of the training set.

3.5. A 9-label model for detailed abundances of giant stars

The red giant branch stars in our training set have stellar parameters (T_{eff} , $\log g$) and up to 15 elemental abundances from the ASPCAP (García Pérez et al. 2016). A subset of these elements have atomic transitions in the *RAVE* wavelength region: O I, Mg I, Al I, Si I, Ca II, Ti I, Fe I, and Ni I. However, we excluded [Ti/H] from our abundance list because of systematics in the ASPCAP [Ti/H] abundances (Holtzman et al. 2015; Hawkins et al. 2016). Therefore we are left with nine labels in our giant star model: T_{eff} , $\log g$, and seven elemental abundances.

Similar to Sections 3.3 and 3.4, we used a quadratic vectorizer for the giant star model. Here the terms are expanded in the same way as equation 5, only with nine labels instead of three. We set the regularization hyperparameter Λ in the same way described in Section 3.4, using the same 30 trials of Λ . The results are shown in Figure 3, where again the enveloped region represents the

minimum and maximum change in RMS label deviation with respect to the unregularized case. At the point of maximum improvement near $\Lambda = 0.13$ ($\approx 10^{-0.9}$), the RMS in all nine labels has decreased by up to 30%, with all labels showing an improvement $> 5\%$, and the mean improvement over all labels is about 10%. Near $\Lambda \approx 10^{-0.9}$ to $10^{-0.3}$, the regularization also produces a sparser matrix of $\boldsymbol{\theta}$, with $\approx 20\%$ more terms (mostly cross-terms) having zero-valued entries. Based on the increased model sparsity and decreasing RMS deviation in the labels, we adopt $\Lambda = 0.57$ ($10^{-0.25}$) for the giant star model. The bias in labels from a regularized model with $\Lambda = 0.57$ is negligible: -0.3 K in T_{eff} , and <0.007 dex in magnitude for $\log g$ and all seven elements. The RMS at this regularization strength is 69 K in T_{eff} , 0.18 dex in $\log g$, and varies between 0.07–0.09 dex depending on the elemental abundance. We inferred labels for all 520,782 *RAVE* spectra using this model, and these results are summarized in the right panel of Figure 2. The red clump is clearly visible and in the expected location, without requiring any post-analysis calibration. However, artefacts due to dwarf stars in the test set are also present.

3.6. Joining the models

We have derived labels for all 520,782 *RAVE* spectra using the three models described in previous sections. The results from the joint model — that includes the main-sequence, sub-giant and red giant branch — shows that a single 3-label quadratic model is too simple for the *RAVE* spectral range. The other models have problems, too: unrealistic over-densities in label space show that the main-sequence model and the giant model make very poor extrapolations for stars outside their respective training sets. For these reasons we were forced to exclude or severely penalize incorrect results from both models. We emphasize that the choices here are entirely heuristic, and depart from interpreting *The Cannon* output labels as the maxima of individual likelihood functions. Each model produces estimates of the labels for a given star, and we use those estimates to produce a unified estimate, but this joint estimate is calculated by disregarding the probabilistic attributes of individual estimates.

Before attempting to join the results from different models, we excluded results in either model that had a reduced $\chi_r^2 > 3$. We further discarded stars with labels that are outside the extent of the training set. Specifically for the results from the giant model we (conservatively) excluded stars with derived $\log g > 3.5$, and for the results from the main-sequence model we excluded sub-giant stars ($\log g < 4$ and $T_{\text{eff}} < 5000$ K) that were outside the two-dimensional ($T_{\text{eff}}, \log g$) convex hull of the training set used for the main-sequence model. Unfortunately these restrictions did not remove all spurious results. The reason for this can be explained with an example: consider that our giant star model was trained with only giant stars but tested with both giant stars and dwarf stars. Some classes of stars (e.g., metal-poor dwarfs) can project into a region of label space that would suggest it is a giant (e.g., a clump star). These objects could have relatively low χ_r^2 values (e.g., $\chi_r^2 < 3$) and in this example, they would appear as bona fide red clump stars. These incorrect projections are extrapolation errors in high dimensions that project to ‘normal’ parts of the label space in two

dimensions. For these reasons we also made use of the joint model in Section 3.3 to inform whether we should adopt results from: the red giant branch model; the main-sequence/sub-giant model; or a linear combination of the two.

In Figure 4 we show the differences in effective temperature T_{eff} and surface gravity $\log g$ between: the main-sequence model and the joint model; and the differences between the red giant branch model and the joint model. We have scaled the differences in T_{eff} and $\log g$ to make the central peak near $(0, 0)$ to be approximately isotropic by setting $\delta_{T_{\text{eff}}} = 90$ K for the main-sequence model, $\delta_{T_{\text{eff}}} = 50$ K for the giant model, and $\delta_{\log g} = 0.15$ dex for both models. It is important to note that these δ values do not represent any kind of intrinsic uncertainty or precision: they are merely normalization factors. Adopting substantially different scaling factors would produce in clear inconsistencies in our results (e.g., sub-giants being misclassified as giants). Therefore we chose these factors empirically to make the distributions in Figure 4 approximately isotropic, and to some extent, comparable. In Figure 4 the stars within the peak at $(0, 0)$ represent objects where the joint model and the comparison model both report similar labels. The artefacts seen in the Hertzsprung-Russell diagrams in Figure 2 are also present in Figure 4 as over-densities far away from the central peak. Therefore, we can adopt the scaled distance in labels T_{eff} and $\log g$ from the joint model to the main-sequence model d_{ms} ,

$$d_{ms} = \left(\frac{T_{\text{eff},ms} - T_{\text{eff},joint}}{\delta_{T_{\text{eff},ms}}} \right)^2 + \left(\frac{\log g_{ms} - \log g_{joint}}{\delta_{\log g,ms}} \right)^2 . \quad (7)$$

However when adopting a similar distance metric for the giant model, we found that the giant model and the joint model produced similar label estimates (with low χ^2_r values) for metal-poor ($[\text{Fe}/\text{H}] < -1$) dwarf stars near the turn-off. The similarity in estimated labels occurred in T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$, implying that we could not use the difference between $[\text{Fe}/\text{H}]_{\text{giant}} - [\text{Fe}/\text{H}]_{\text{joint}}$ as a term that would benefit the main-sequence model. As a result, our joint estimate for metal-poor dwarf stars that overlap with high-resolution studies (Reddy et al. 2003, 2006; Bensby et al. 2014; Valenti & Fischer 2005), would appear as metal-poor giant stars, because the giant star model would have a much smaller distance metric than the main-sequence model. It is reasonable to assert that this issue may be complicated by the increased complexity, or flexibility, of the giant star model relative to the main-sequence model. In any case, we found that by empirically penalizing the giant model distance metric by absolute metallicity,

$$d_{giant} = \left(\frac{T_{\text{eff},giant} - T_{\text{eff},joint}}{\delta_{T_{\text{eff},giant}}} \right)^2 + \left(\frac{\log g_{giant} - \log g_{joint}}{\delta_{\log g,giant}} \right)^2 + |[\text{Fe}/\text{H}]_{\text{joint}}| , \quad (8)$$

the weighting of dwarfs was no longer a function of joint metallicity, and this issue was resolved without introducing new artefacts. Using these distance metrics we then derived the weights,

$$w_{ms} = \frac{1}{d_{ms}^2} \quad \text{and} \quad w_{giant} = \frac{1}{d_{giant}^2} , \quad (9)$$

and produce the weighted labels $\hat{\ell}$:

$$\hat{\ell} = \frac{w_{ms} \ell_{ms} + w_{giant} \ell_{giant}}{w_{ms} + w_{giant}} . \quad (10)$$

We calculate weighted errors of $\hat{\ell}$ in the same manner. In Figure 5 we show the mean relative weight $w_{ms}/(w_{ms} + w_{giant})$ within each two-dimensional bin of T_{eff} and $\log g$. For giant stars the relative weight of the main-sequence model is zero, and vice-versa for main-sequence stars. The relative weights smoothly transition from 0 to 1 on the sub-giant branch near $\log g \approx 3.5$, in the training set overlap region of both models. For abundance labels in the giant model that are not in the main-sequence model (e.g., [O/H], [Mg/H]), we only report abundances for objects if $w_{ms} < 0.05$.

The weighted T_{eff} and $\log g$ values for stars meeting different S/N constraints are shown in Figure 6, both in logarithmic density and mean metallicity. The artefacts from individual models are no longer apparent, and the rich structure of the Hertzsprung-Russell is visible. Despite this, there are a number of caveats introduced by the choices we have made to combine estimates from multiple models. We discuss these issues in detail in Section 5.

4. Validation experiments

In addition to the cross-validation tests that we have previously described, we have conducted a number of internal and external validation experiments to test the validity of our results. We will begin by describing internal validation tests based on repeat observations, before evaluating our accuracy based on high-resolution literature comparisons.

4.1. Multi-epoch observations

The RAVE survey performed repeat observations for 43,918 stars with time intervals ranging from a few hours to up to four years. This timing was constructed to be quasi-logarithmic such that spectroscopic binaries could be optimally identified. Most of the stars that were observed multiple times were only observed twice, with thirteen visits being the maximum number of observations for any target. These repeat observations allow us to quantify the level of (in)correctness in our formal errors. For every star with multiple visits we constructed a high S/N stacked spectrum for that star, weighted by the inverse variances in the individual visit spectra.

We treated the stacked spectra as normal survey stars. We inferred labels using all models and joined the results as per Section 3.6. The labels inferred from the stacked spectra served as the basis of comparison for all individual visits to that object. Figure 7 shows the difference in labels between the comparison spectrum and a repeat visit, normalized by their formal errors summed in quadrature (e.g., $\Delta \log g / \sqrt{\sigma_{\log g, \text{stacked}}^2 + \sigma_{\log g, \text{visit}}^2}$). If our measurements were unbiased by S/N

and the formal errors were representative, these values should be normally distributed with a zero mean and unit variance. **Are they? (Probably not!). How much should we inflate our formal errors, and should this be as a function of S/N? Note that in all Figures and tables, we show these inflated errors only, as we gauge them to be more representative.**

In Figure 8 we look at the difference as a function of SNR

4.2. External validation

4.2.1. Comparison with RAVE DR4

As an initial point of comparison, we cross-matched our results against the official fourth *RAVE* data release (Figure 9). In order to provide a fair comparison, we only show stars that meet a number of quality flags in *both* samples. Our constraints require that the S/N ratio exceeds 10 pixel⁻¹, and $\chi_r^2 < 3$. For this comparison we further required that: the QK flag from Kordopatis et al. (2013) is zero, indicating no problems were reported by the pipeline; $T_{\text{eff},DR4} > 4000$ K; the error in radial velocity ϵ_{HRV} is < 8 km s⁻¹; and the three principal morphological flags c1, c2, c3, from Matijević et al. (2012) all indicate ‘n’ for a normal FGK-type star. There is good agreement in T_{eff} , with a bias and RMS of just 4 K and 240 K, respectively. The offset in $\log g$ on the giant branch between this study and Kordopatis et al. (2013) has been noted in other studies (e.g., *APOGEE*), and this has issue has been resolved in the fifth *RAVE* data release by correcting $\log g$ values with a calibration sample consisting of asteroseismic targets and the *Gaia* benchmark stars. There is also a slight discrepancy in the $\log g$ values along the main-sequence, where our work tends to taper down towards higher $\log g$ values at cooler temperatures, and the *RAVE* DR4 sample tends to have a slightly flatter lower main-sequence. This difference is not likely to have a very significant effect on the detailed abundance or spectrophotometric distance determinations between these studies (Binney et al. 2014).

4.2.2. Comparisons with Reddy, Bensby, and Valenti & Fischer

We searched the literature for studies that overlap with *RAVE*, and which base their analysis on high-resolution, high S/N spectra. We found four notable studies with a sufficient level of overlap: the disk studies by Reddy et al. (2003, 2006) and Bensby et al. (2014), as well as the Valenti & Fischer (2005) work on exoplanet host star candidates. These studies perform a careful (manual; expert) analysis using extremely high-resolution, high S/N spectra, and make use of *Hipparcos* parallaxes where possible. Most of the stars in these samples are main-sequence or sub-giant stars. Therefore, these works constitute an excellent comparison to evaluate the accuracy of our results on the main-sequence and sub-giant branch.

We compare the stellar parameters (T_{eff} , $\log g$, [Fe/H]) of stars that overlap these literature

studies in Figure 10. We only include stars with $\chi_r^2 < 3$ and $S/N > 10 \text{ pixel}^{-1}$, although the latter cut removed only few stars because the average S/N in the *RAVE* spectra for these stars is relatively high ($\gtrsim 50 \text{ pixel}^{-1}$). When considering the relative information content available in *RAVE* (945 pixels in the near infrared with $\mathcal{R} \approx 7,500$) compared to these literature studies that use *Hipparcos* parallaxes where possible, and base their inferences on spectra with resolving power \mathcal{R} between 40,000 to 110,000, and S/N ratios exceeding 150, we consider the agreement to be very satisfactory. Indeed, given the metallicity precision available in the *unRAVE* catalog, these results will likely be useful for future studies based on exoplanet host star properties (e.g., *TESS*). At present, however, there are just ≈ 30 stars in *RAVE* that overlap with the compilations of exoplanet host star properties listed at exoplanets.org and exoplanets.eu.

4.2.3. Comparison with the *Gaia-ESO* survey

There are 142 stars that overlap between *RAVE* and the fourth internal data release of the *Gaia-ESO* survey. These are a mix of main-sequence, sub-giant and red giant branch stars. About half (67) of the sample were acquired with the UVES instrument, the other with the GIRAFFE spectrograph, and the S/N of the *Gaia-ESO* spectra peaks at $\approx 140 \text{ pixel}^{-1}$. Despite most of these stars having relatively low S/N ratios in *RAVE* ($\approx 25 \text{ pixel}^{-1}$), there is good agreement in with *Gaia-ESO* and the *unRAVE* stellar parameters (Figure 12). The RMS in effective temperature, surface gravity and metallicity is 233 K, 0.37 dex, and 0.17 dex, respectively.

Based on this comparison, we find no evidence for a systematic offset between stars on the main-sequence and those on the giant branch. Specifically, we do not find any dwarf stars that are misclassified as sub-giants, as we found for some metal-poor stars in Section 4.2.2. In addition we find no evidence for a systematic offset in metallicity (or individual abundances) between giant stars and main-sequence stars. This is a crucial observation, as the metallicities for stars in our main-sequence training set have a principally different source to those on the giant branch. We cannot make these same inferences based on other surveys, like *APOGEE*, because (1) *APOGEE* stars formed part of the training set, and (2) they do not include main-sequence stars. Even if we found good agreement between *K2/EPIC* and *APOGEE* metallicities, this would not be informative, because *APOGEE* is the source of metallicity for many stars on the giant branch in the *K2/EPIC* sample. Therefore, although this is a qualitative comparison only, it is reassuring that there is no obvious systematic difference between the metallicities of main-sequence and giant branch stars.

The metallicity agreement between this work and *Gaia-ESO* extends down to low metallicity, near $[\text{Fe}/\text{H}] \approx -1.5$. The scatter increases for the four stars in the overlap sample with $[\text{Fe}/\text{H}] < -1$, however these spectra have low S/N ratios, which is reflected in the larger uncertainties reported for these metallicities.

The fourth internal data release of the *Gaia-ESO* includes detailed chemical abundances of up to 45 species (≈ 32 elements at different ionization stages). This provides us with an independent

validation for our detailed abundances on the giant branch. These comparisons are shown for six elemental abundances in Figure 13, where markers are colored by the S/N of the *RAVE* spectrum. The number of stars available in each abundance comparison varies due to what is available in the *Gaia-ESO* data release, which is itself a function of the instrument used, the spectral type, and other factors. The absolute bias for individual elements varies from as low as 0.01 dex ([Al/H]) to as high as 0.22 dex ([Si/H]), where we over-estimate [Si/H] abundances relative to the *Gaia-ESO* survey. The large bias in [Si/H] is likely a consequence of an offset between [Si/H] abundances in *Gaia-ESO* and *APOGEE*, the source of our training set for giant star abundances. The RMS deviation in each label is small for stars with $[X/H] \sim -0.5$, before increasing at lower metallicities. If we consider all stars, the smallest abundance RMS we see with respect to *Gaia-ESO* is 0.18 dex for [Ca/H]. The increasing RMS at low metallicity is likely a consequence of multiple factors, namely: inaccurate abundance labels for metal-poor stars (Section 3.2); only weak, blended lines being available in *RAVE*, which cease to be visible in hot and/or metal-poor stars; and to a lesser extent, low S/N ratios for those particular stars being compared. Unfortunately not all of these factors are represented by the quoted errors in each label. For these reasons, although it affects only a small number of stars, we recommend caution when using individual abundances for very metal-poor giant stars in our sample.

4.3. Astrophysical validation

4.3.1. Globular clusters

After verifying that our atmospheric parameters and abundances are comparable with other high-resolution studies, here we verify that the abundances we find are consistent with expectations from astrophysics. One example is globular clusters, which show anti-correlations in light element abundances (e.g., Norris & Da Costa 1995; Carretta et al. 2009, and references therein). In the *RAVE* survey, Anguiano et al. (2015) identified 70 stars with positions and radial velocities that are consistent with being members of globular clusters: 49 stars belonging to NGC 5139 (ω Centauri), 11 members of the retrograde globular cluster NGC 3201, and 10 members of NGC 362. In addition to these systems, 12 stars thought to belong to NGC 1851 were found in the *RAVE* data by Kunder et al. (2014), and a further 10 stars in NGC 6752. We refer the reader to those studies for details regarding the membership selection.

In Figure 14 we show the effective temperature T_{eff} and surface gravity $\log g$ for pre-defined members of globular clusters in *RAVE*. The top panels indicate measurements made in this work, and for comparison purposes we have included the results from the fourth *RAVE* data release. We show representative *PARSEC* isochrones (Bressan et al. 2012) in all panels, where the isochrone ages and metallicities are adopted from Kunder et al. (2016); Marín-Franch et al. (2009), and the Harris (1996, accessed 6 September 2016) online catalog of globular cluster properties. The globular cluster with the most number of *RAVE* members is NGC 5139 (ω -Centauri), where we

find a significant metallicity spread consistent with high-resolution studies (Marino et al. 2011; Carretta et al. 2009, 2013). We find the mean metallicity of prescribed ω -Centauri members to be $[\text{Fe}/\text{H}] = -0.85$. However, it is clear that the membership criteria could be improved with our revised metallicities and detailed chemical abundances. Indeed, this demonstrates that our abundance labels can be used to identify globular cluster members that are now tidally disrupted (Anguiano et al. 2016; Kuzma et al. 2016; Navin et al. 2016), even for stars with low S/N ratios.

4.3.2. Open clusters

The *RAVE* survey observed \sim 160 highly probable members of several open clusters: 78 stars in the Pleiades, 26 in Hyades, 30 stars in the Solar-metallicity cluster M67 and 13 stars in IC4651. Details on the open cluster membership and selection can be found in Kunder et al. (2016). We show the effective temperature T_{eff} and surface gravity $\log g$ for these stars in Figure 15. The isochrones are sourced from Bressan et al. (2012), with cluster properties adopted from Kunder et al. (2016) and Kharchenko et al. (2013).

We find good agreement between our atmospheric parameters (top panels) and the isochrones shown. The position of the red clump in IC4651 and M67 are perfectly matched to the isochrone, and the Hyades main-sequence is in good agreement down to $T_{\text{eff}} \approx 4000$ K. Similarly, we find consistency with the literature and our metallicity scale. We find the mean metallicity of M67 stars to be $[\text{Fe}/\text{H}] = -0.02 \pm 0.17$ dex (30 stars), in excellent agreement with the expected $[\text{Fe}/\text{H}] = 0.00$ value (not accounting for atomic diffusion). We further find the Hyades mean metallicity to be $[\text{Fe}/\text{H}] = 0.07 \pm 0.09$ dex, consistent with Paulson et al. (2003): $[\text{Fe}/\text{H}] \approx 0.13$. For IC4651 from 13 stars we find a mean $[\text{Fe}/\text{H}] = 0.15 \pm 0.12$ dex, matching the high-resolution, high S/N study of Pasquini et al. (2004), where they find $[\text{Fe}/\text{H}] = 0.10 \pm 0.03$ dex. Finally, from 78 stars we find the mean metallicity of the Pleiades to be $[\text{Fe}/\text{H}] = -0.02 \pm 0.12$ dex, in agreement with the $[\text{Fe}/\text{H}] = -0.034 \pm 0.024$ dex measurement reported by Friel & Boesgaard (1990).

The most pronounced discrepancy is seen in the main-sequence of the Pleiades. Like Kordopatis et al. (2013), we find lower $\log g$ values than what the isochrone shown predicts. **Hawkins: use correct isochrone from Bouy to see if this issue is resolved.**

5. Discussion

We have derived atmospheric parameters (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$) for all *RAVE* spectra, as well as detailed abundances for red giant branch stars, based on an independent re-analysis of *RAVE* spectra. Our catalog provides an excellent complement to the current *RAVE* data release (Kunder et al. 2016). When combined with the TGAS sample, these results may amount to a powerful compendium for chemo-dynamic studies of the Milky Way. However, our analysis has caveats. Inferences based on these results should recognize those caveats, and acknowledge that these results

are subject to our explicit assumptions, some of which are provably incorrect.

For practical purposes we adopted separate models: one for the giant branch and one for the main-sequence. A third model was used to derive relative weights for which results to use. The relative weighting we have used does not have any formal interpretation as a likelihood or belief (in any sense): it was introduced for practical reasons to identify systematic errors and combine results for multiple models. Because the relative weights have no formal interpretation, it is reasonable to consider this method is as *ad hoc* as any other approach. Therefore, the relative weighting has no warranty to be (formally) correct, and may introduce inconsistencies or systematic errors rather than minimizing them. Evidence of this is already present in our comparisons to ‘gold standard’ works (Bensby et al. 2014; Reddy et al. 2003, 2006; Valenti & Fischer 2005), where metal-poor dwarf stars appear as metal-poor sub-giants in our combined results, but have reasonable labels (by any available statistic) as dwarfs in the main-sequence model. Clearly, the relative distance from the joint model to the giant model is slightly favoured (over the main-sequence model) for metal-poor dwarf stars. In this sense, our approach towards combining labels from multiple models is sub-optimal.

If we only consider the results from individual models, there are a number of cautionary remarks that stem from the construction of the training set. The labels for red giant branch stars primarily come from *APOGEE*, where previous successes with *The Cannon* have demonstrated that *APOGEE* labels based on high S/N data can be of high fidelity (Ness et al. 2015, 2016; Ho et al. 2016; Casey et al. 2016). However, the lack of metal-poor stars in the *APOGEE/RAVE* overlap sample produced a tapering-off in the test set — where *no* stars had reliably reported metallicities below that of the training set — which forced us to construct a heterogeneous training set. The metal-poor stars included in this sample are from high-resolution studies (Fulbright et al. 2010; Ruchti et al. 2011), but it is unknown if the stellar parameters are of high fidelity (e.g., we have limited quality metrics available), and there is no guarantee that the stellar parameters *or* abundances are on the same scale as *APOGEE* (and good reasons to believe they will not be; see Smiljanic et al. 2014).

If the metallicities of metal-poor giant stars are on the same scale as the *APOGEE* abundances, there is a larger issue in verifying that the main-sequence metallicities and giant branch metallicities are on the same scale. The training set for the main-sequence stars includes metallicities from a variety of sources — including *LAMOST*, and the fourth *RAVE* data release. Even on expectation value, there is no straightforward manner to ensure that these two models produce metallicities on the same scale. We see no systematic offset in metallicities of dwarf and giant stars that overlap between *RAVE* and the *Gaia-ESO* survey, suggesting that if there is a systematic offset, it must be small. Nevertheless, these are only verification checks based on < 1% of the data, and there is currently insufficient data for us to prove both models are on the same abundance scale.

For some of the most metal-poor giant stars in *RAVE* we *know* the abundances are not on the same scale as *APOGEE*, because we were forced to adopt abundances for specific elements when they were unavailable. In that sense, our assumption that the training set contains accurate

labels is clearly untrue (for *at least* some stars). Although we sought to adopt the mean level of Galactic chemical enrichment at the overall metallicity, this is not a representative abundance. Even if that is the *mean* enrichment at that Galactic metallicity, there is no requirement for zero abundance spread. More fundamentally, we are incorrectly assuming that the element *must* be detectable in the photosphere of the star. There may be no transition that is detectable in that star, even with zero noise, because it is too weak to have any effect on the spectrum. In the most optimistic case, this can be considered to be forcing the model to make use of correlated information between abundances. In a more representative (pessimistic) case, we are simply invoking what all abundances should be at low metallicity.

This choice is reflected in the abundances of the test set. While we do recover trustworthy metallicities for ultra metal-poor ($[\text{Fe}/\text{H}] \lesssim -4$) stars like CD-38 245, the individual abundances for all extremely metal-poor stars aggregate (in $[\text{X}/\text{Fe}]$ space; Figure 17) at the assumed abundances for the metal-poor stars in our sample. Thus, while the overall metallicities appear reliable, the individual abundances for extremely metal-poor stars in the test set cannot be considered trustworthy in any sense. For this reason we have updated the electronic catalog to discard these results, and included a quality control flag to indicate the result was discarded (see catalog README file for specifics related to the data structure and content).

In Section 3 we assumed that any fibre- or time-dependent variations in the *RAVE* spectra are negligible. This is provably incorrect. Indeed, Kordopatis et al. (2013) note that the effective resolution of *RAVE* spectra varies from $6,500 < \mathcal{R} < 8,500$, and that the effective resolution is a function of temperature variations, fibre-to-fibre variations, and thus position on the CCD (Steinmetz et al. 2006). For this reason we ought to expect our derived stellar parameters or abundances to be correlated with the fibre number, or over time. If significant, the trend could produce systematically offset stellar abundances solely due to the fibre number. Kordopatis et al. (2013) conclude that resolution-based effects on the *RAVE* stellar parameters should be a second-order effect. We have not seen evidence of these resolution-based correlations in our results, however, we have only performed cursory (non-exhaustive) experiments to investigate this issue.

assumptions that are probably wrong: fringing in high S/N spectra produces issues with the normalization and thus the spectrum. KCF to comment more on this

The fibre fringing is another example of how test spectra that are unusual from the training set can be projected as peculiar artefacts in the label space. In other words, unusual spectra can appear as ‘clumps’ in regions of parameter space that we could consider as being normal (e.g., Solar-type stars). We addressed this issue for the main-sequence and giant models by using a joint model to calculate relative weights. However, spectra that are unusual from the joint model training set can still project as systematic artefacts.

Indeed, there are two known artefacts in our data that are relevant to this discussion. The first is a small over-density at the base of the giant branch, which is likely a consequence of joining the 9-label and 3-label models. The second has an astrophysical origin: there are no hot stars

($T_{\text{eff}} > 8000$ K) present in our training set, yet there are many in the *RAVE* survey. However, the *RAVE* pre-processing pipeline (**SPARV**) performs template matching against grids of cool *and* hot stars, and therefore we can use those results to identify hot stars. In Figure 18 we show our derived effective temperatures T_{eff} and surface gravities $\log g$, where each hexagonal bin is colored by the *maximum* temperature reported by **SPARV** for any star in that bin. Hot stars dominate at just above the turn-off point in our results.

this following paragraph needs rewording Based on a thorough interrogation of our results, we have adopted two boolean columns to signify whether our labels meet specific quality constraints: one marked **QC** to indicate reliable stellar parameters (T_{eff} , $\log g$, [Fe/H]), and **QC_ABUNDANCE** to indicate reliable chemical abundances. For stellar atmospheric parameters to meet our quality constraints, we require: $\chi^2_r < 3$; $S/N > 10 \text{ pixel}^{-1}$; and **TeffS** < 8000 K in order to remove hot stars that are outside the training set. We specified the quality constraint in chemical abundances to further require that the iron abundance $[\text{Fe}/\text{H}] > -1.5$, since this approximately corresponds to the point where individual abundances will become undetectable in *RAVE* spectra, and the individual abundance labels will be driven by our constraint to place them on the mean galactic chemical evolution trend. These quality constraints should not be taken verbatim: judicious use of the morphological classifications (Matijević et al. 2012) or metadata from the pre-processing pipelines is encouraged.

We have presented the largest collection of stellar abundances for stars in the first *Gaia* data release. In total more than X.XX million abundances are available, with effective temperature T_{eff} , surface gravity $\log g$ and metallicity available for XXXXXX stars. Despite the caveats and limitations discussed here, our validation experiments and comparisons with high-resolution spectroscopic studies suggests that our results have sufficient accuracy and precision to be useful for exoplanet host stars studies (c.f., Valenti & Fischer 2005), investigations of the Milky Way disk (c.f., Reddy et al. 2003, 2006; Bensby et al. 2014), and a multitude of chemo-dynamic studies. When combined with positions and 3D velocities from *Gaia*, we are optimistic that the *unRAVE* catalog will help advance understanding of the Milky Way’s formation and evolution.

Access the results electronically

Source code for this project is available at <https://github.com/AnnieJumpCannon/RAVE>, and this document was compiled from revision hash 50c0c62 in that repository. Inferred labels, associated errors, and relevant metadata are available electronically at **ZENODO-URL** (?). These results are also available through the *RAVE* DR5 database (Kunder et al. 2016). Please note that it is a condition of using these results that the fifth *RAVE* data release by Kunder et al. (2016) must also be cited, as the work presented here would not have been possible without the tireless efforts of the entire *RAVE* collaboration, past and present.

We thank Jonathan Bird (Vanderbilt), Sven Buder (MPIA), Sergey Koposov (Cambridge), and Sofia Randich (INAF). This research made use of: NASA’s Astrophysics Data System Bibliographic Services; Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013); and *TOPCAT* (Taylor 2005). This work was partly supported by the European Union FP7 programme through ERC grant number 320360. A. R. C. thanks King’s College of the University of Cambridge for their support. Funding for K. H. has been provided through the Simons Foundation Society of Fellows and the Marshall Scholarship. Funding for RAVE has been provided by: the Australian Astronomical Observatory; the Leibniz-Institut fuer Astrophysik Potsdam (AIP); the Australian National University; the Australian Research Council; the French National Research Agency; the German Research Foundation (SPP 1177 and SFB 881); the European Research Council (ERC-StG 240271 Galactica); the Istituto Nazionale di Astrofisica at Padova; The Johns Hopkins University; the National Science Foundation of the USA (AST-0908326); the W. M. Keck foundation; the Macquarie University; the Netherlands Research School for Astronomy; the Natural Sciences and Engineering Research Council of Canada; the Slovenian Research Agency; the Swiss National Science Foundation; the Science & Technology Facilities Council of the UK; Opticon; Strasbourg Observatory; and the Universities of Groningen, Heidelberg and Sydney. The RAVE web site is <https://www.rave-survey.org>.

REFERENCES

- Anguiano, B., Zucker, D. B., Scholz, R.-D., et al. 2015, MNRAS, 451, 1229
Anguiano, B., De Silva, G. M., Freeman, K., et al. 2016, MNRAS, 457, 2078
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, Astronomy & Astrophysics, 558, AA33
Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71
Binney, J., Burnett, B., Kordopatis, G., et al. 2014, MNRAS, 437, 351
Boeche, C., Siebert, A., Williams, M., et al. 2011, AJ, 142, 193
Boeche, C., Chiappini, C., Minchev, I., et al. 2013, A&A, 553, A19
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Bovy, J., Rix, H.-W., Schlafly, E. F., et al. 2016, ApJ, 823, 30
Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, AJ, 142, 112
Bouy, H., Bertin, E., Sarro, L. M., et al. 2015, A&A, 577, A148
Carretta, E., Bragaglia, A., Gratton, R. G., et al. 2009, A&A, 505, 117

- Carretta, E., Bragaglia, A., Gratton, R. G., et al. 2013, *A&A*, 557, A138
- Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, *A&A*, 530, A138
- Casey, A. R., Keller, S. C., & Da Costa, G. 2012, *AJ*, 143, 88
- Casey, A. R., Da Costa, G., Keller, S. C., & Maunder, E. 2013, *ApJ*, 764, 39
- Casey, A. R., Keller, S. C., Da Costa, G., Frebel, A., & Maunder, E. 2014, *ApJ*, 784, 19
- Casey, A. R., Keller, S. C., Alves-Brito, A., et al. 2014, *MNRAS*, 443, 828
- Casey, A. R., & Schlaufman, K. C. 2015, *ApJ*, 809, 110
- Casey, A. R. 2016, *ApJS*, 223, 8
- Casey, A. R., Hogg, D. W., Ness, M., et al. 2016, arXiv:1603.03040
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *Research in Astronomy and Astrophysics*, 12, 1197
- De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, *MNRAS*, 449, 2604
- Deason, A. J., Belokurov, V., & Evans, N. W. 2011, *MNRAS*, 416, 2903
- Friel, E. D., & Boesgaard, A. M. 1990, *ApJ*, 351, 480
- Fulbright, J. P., Wyse, R. F. G., Ruchti, G. R., et al. 2010, *ApJ*, 724, L104
- García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, *AJ*, 151, 144
- Gilmore, G., Randich, S., Asplund, M., et al. 2012, *The Messenger*, 147, 25
- Harris, W. E. 1996, *AJ*, 112, 1487
- Hawkins, K., Masseron, T., Jofre, P., et al. 2016, arXiv:1604.08800
- Holtzman, J. A., Shetrone, M., Johnson, J. A., et al. 2015, *AJ*, 150, 148
- Ho, A. Y. Q., Ness, M. K., Hogg, D. W., et al. 2016, arXiv:1602.00303
- Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, *A&A*, 355, L27
- Huber, D., Bryson, S. T., Haas, M. R., et al. 2016, *ApJS*, 224, 2
- Jofré, P., Mädler, T., Gilmore, G., et al. 2015, *MNRAS*, 453, 1428
- Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2013, *A&A*, 558, A53

- Kordopatis, G., Recio-Blanco, A., de Laverny, P., et al. 2011, *A&A*, 535, A106
- Kordopatis, G., Gilmore, G., Steinmetz, M., et al. 2013, *AJ*, 146, 134
- Kordopatis, G., Binney, J., Gilmore, G., et al. 2015, *MNRAS*, 447, 3526
- Kunder, A., et al. 2016, submitted
- Kunder, A., Bono, G., Piffl, T., et al. 2014, *A&A*, 572, A30
- Kuzma, P. B., Da Costa, G. S., Mackey, A. D., & Roderick, T. A. 2016, *MNRAS*, 461, 3639
- Mädler, T., Jofré, P., Gilmore, G., et al. 2016, arXiv:1606.03015
- Matijević, G., Zwitter, T., Bienaymé, O., et al. 2012, *ApJS*, 200, 14
- Marino, A. F., Milone, A. P., Piotto, G., et al. 2011, *ApJ*, 731, 64
- Marín-Franch, A., Aparicio, A., Piotto, G., et al. 2009, *ApJ*, 694, 1498
- Michalik, D., Lindegren, L., & Hobbs, D. 2015, *A&A*, 574, A115
- Michalik, D., Lindegren, L., Hobbs, D., & Butkevich, A. G. 2015, *A&A*, 583, A68
- Munari, U., Sordo, R., Castelli, F., & Zwitter, T. 2005, *A&A*, 442, 1127
- Navin, C. A., Martell, S. L., & Zucker, D. B. 2016, arXiv:1606.06430
- Ness, M., Freeman, K., Athanassoula, E., et al. 2012, *ApJ*, 756, 22
- Ness, M., Freeman, K., Athanassoula, E., et al. 2013, *MNRAS*, 430, 836
- Ness, M., Freeman, K., Athanassoula, E., et al. 2013, *MNRAS*, 432, 2092
- Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, *ApJ*, 808, 16
- Ness, M., Hogg, D. W., Rix, H.-W., et al. 2016, *ApJ*, 823, 114
- Nordström, B., Mayor, M., Andersen, J., et al. 2004, *A&A*, 418, 989
- Norris, J. E., & Da Costa, G. S. 1995, *ApJ*, 447, 680
- Pasquini, L., Randich, S., Zoccali, M., et al. 2004, *A&A*, 424, 951
- Paulson, D. B., Sneden, C., & Cochran, W. D. 2003, *AJ*, 125, 3185
- Randich, S., Gilmore, G., & Gaia-ESO Consortium 2013, *The Messenger*, 154, 47
- Recio-Blanco, A., de Laverny, P., Allende Prieto, C., et al. 2016, *A&A*, 585, A93
- Reddy, B. E., Tomkin, J., Lambert, D. L., & Allende Prieto, C. 2003, *MNRAS*, 340, 304

- Reddy, B. E., Lambert, D. L., & Allende Prieto, C. 2006, MNRAS, 367, 1329
- Robin, A. C., Luri, X., Reylé, C., et al. 2012, A&A, 543, A100
- Ruchti, G. R., Fulbright, J. P., Wyse, R. F. G., et al. 2011, ApJ, 743, 107
- Schlaufman, K. C., Rockosi, C. M., Allende Prieto, C., et al. 2009, ApJ, 703, 2177
- Siebert, A., Williams, M. E. K., Siviero, A., et al. 2011, AJ, 141, 187
- SDSS Collaboration, Albareti, F. D., Allende Prieto, C., et al. 2016, arXiv:1608.02013
- Smiljanic, R., Korn, A. J., Bergemann, M., et al. 2014, A&A, 570, A122
- Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645
- Taylor, M. B. 2005, Astronomical Data Analysis Software and Systems XIV, 347, 29
- Valenti, J. A., & Fischer, D. A. 2005, ApJS, 159, 141
- van Leeuwen, F. 2007, A&A, 474, 653
- Wojno, J., et al. 2016, submitted
- Zasowski, G., Johnson, J. A., Frinchaboy, P. M., et al. 2013, AJ, 146, 81
- Zwitter, T., Siebert, A., Munari, U., et al. 2008, AJ, 136, 421

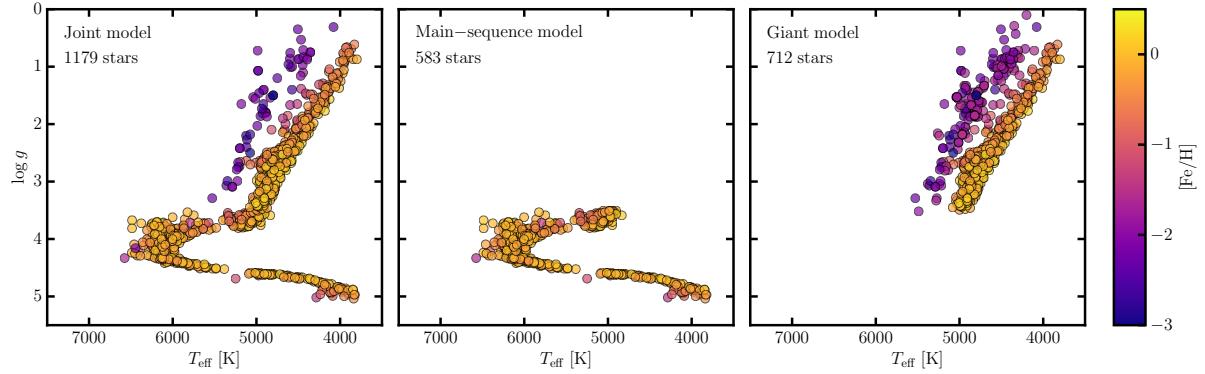


Fig. 1.— Effective temperature T_{eff} and surface gravity $\log g$ for all stars in the training sets. Stars are colored by their metallicity $[\text{Fe}/\text{H}]$, and the three panels show stars in the joint model (left panel; Section 3.3), the main-sequence star model (middle panel; Section 3.4), and the giant star model (right panel; Section 3.5).

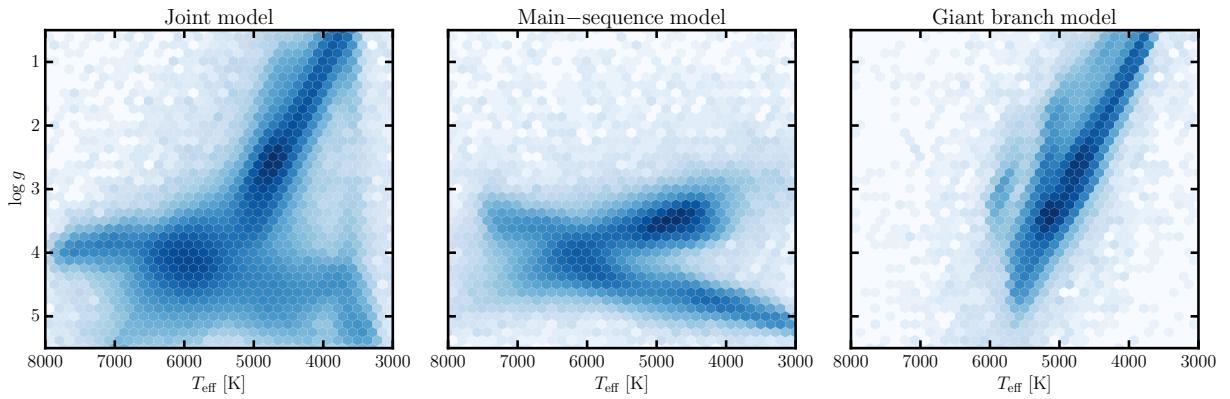


Fig. 2.— The logarithmic density of effective temperature T_{eff} and surface gravity $\log g$ for all 520,782 *RAVE* spectra, as derived using the joint model (left panel; Section 3.3), the main-sequence star model (center panel; Section 3.4), and the giant star model (right panel; Section 3.5). These panels demonstrate how a single quadratic model is insufficient for all *RAVE* stars (left panel), and illustrate some of the systematic artefacts that can result from testing on stars outside of the training set (center and right panel). These panels do not represent our final results, which are shown in Figure 6.

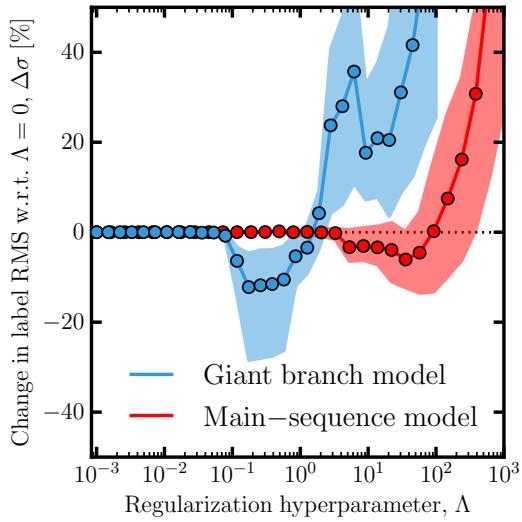


Fig. 3.— The percentage change in RMS deviation between inferred and training labels at different regularization strengths. The RMS values were calculated by leave-one-out cross-validation, and are shown with respect to an unregularized model ($\Lambda = 0$). The points and solid line indicate the mean improvement across all labels. The filled area represents the minimum and maximum improvements over all labels. With increasing regularization strength, there is a minimum in the RMS deviation over all labels, which is where we set Λ for each model (see text for details).

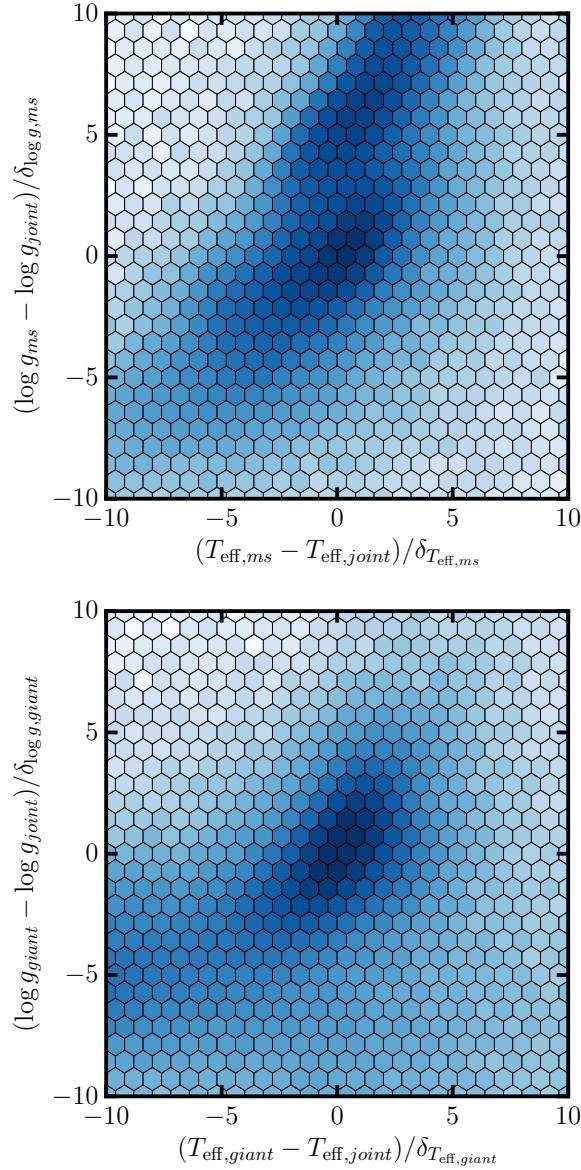


Fig. 4.— The normalized differences in effective temperature T_{eff} and surface gravity $\log g$ between the main-sequence model and the joint model (top panel), and the giant model and the joint model (bottom panel). The density scaling is logarithmic, and the differences in T_{eff} and $\log g$ are scaled to make them approximately isotropic (see text for details). The peak at $(0,0)$ represents good agreement between the joint model and comparison model, whereas the over-densities elsewhere are a consequence of testing the model on stars very different to the training set (e.g., dwarf stars tested on a model trained with only giant stars).

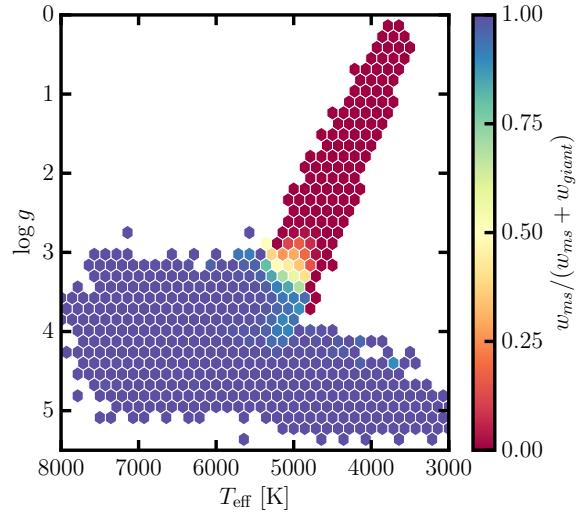


Fig. 5.— The mean relative main-sequence model weight $w_{ms}/(w_{ms} + w_{giant})$ at each hexagonal bin of weighted effective temperature T_{eff} and surface gravity $\log g$. The relative weighting illustrates how only results from the main-sequence model are adopted for unevolved stars, and there is a gradual transition to using results from the giant model, before only results from the giant model are used for evolved stars.

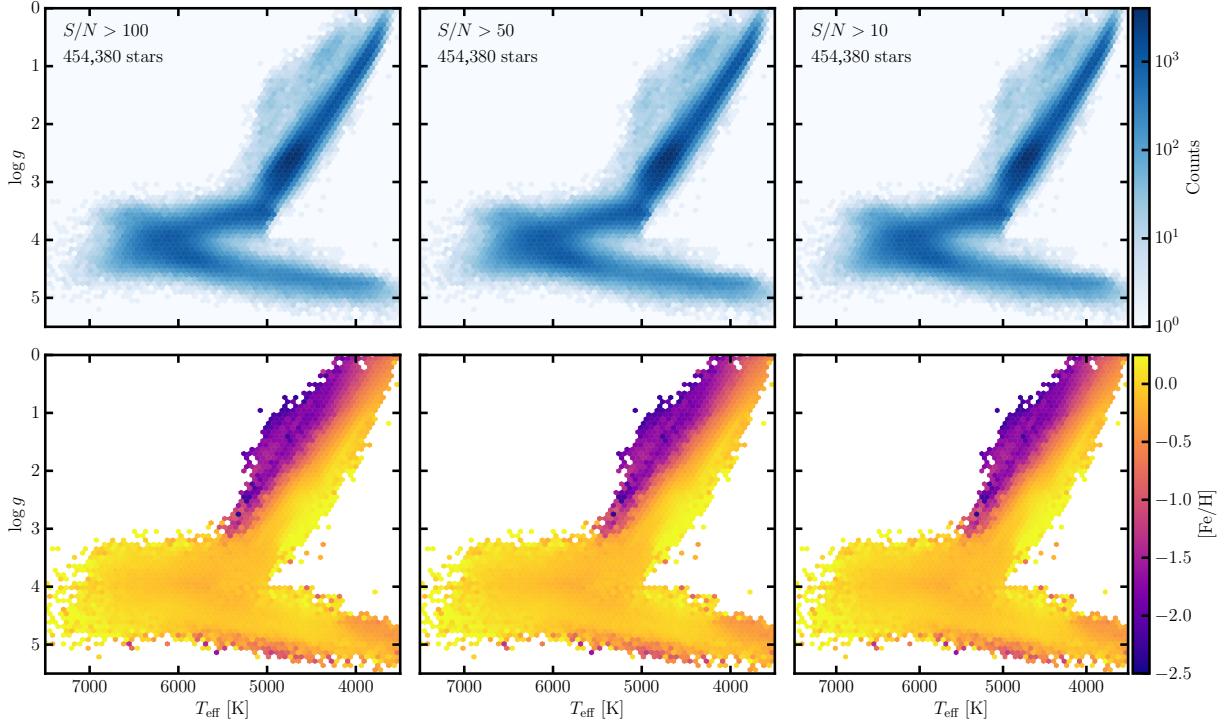


Fig. 6.— The effective temperature T_{eff} and surface gravity $\log g$ for RAVE stars after combining labels from the main-sequence and giant star models. Only results with $\chi^2_r < 3$ and S/N ratios greater than the quoted limit in each panel are shown. The top three panels show logarithmic density, and bins in the bottom three panels are colored by the median metallicity in each bin.

Fig. 7.— Distribution of differences in label estimates from multiple visits, divided by their formal errors summed in quadrature. If our measurements were unbiased and the formal errors were representative, these distribution should be normally distributed with zero mean and unit variance.

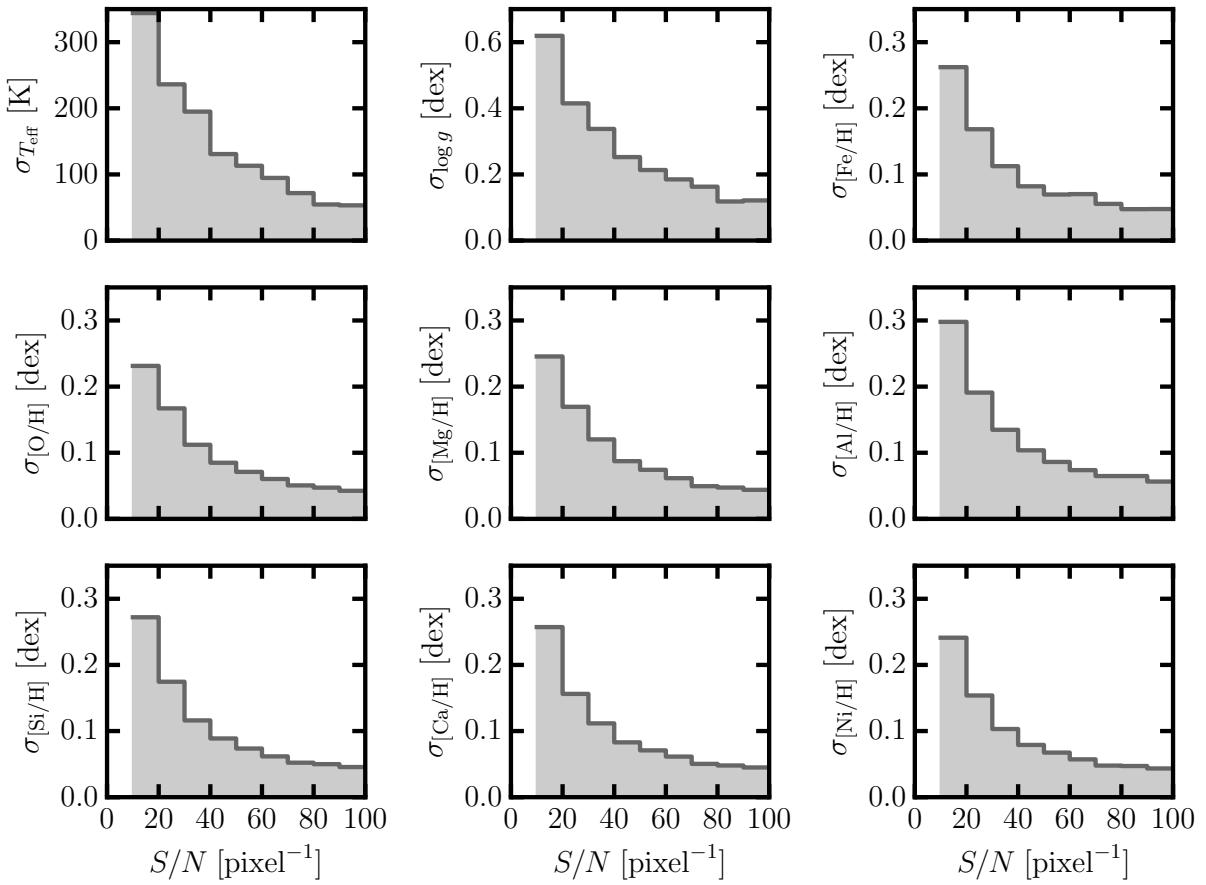


Fig. 8.— The RMS of the test set labels as a function of S/N ratio for repeated stars in the test set.

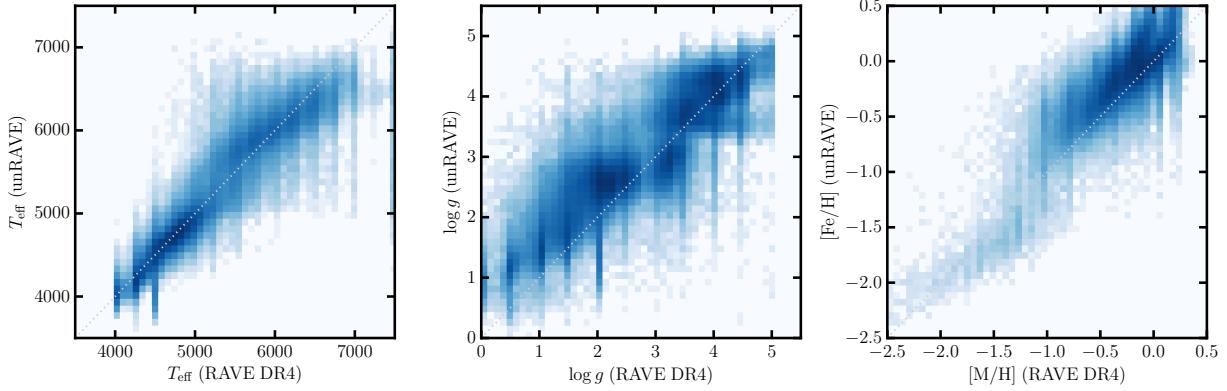


Fig. 9.— Stellar parameter (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$) comparison between the fourth *RAVE* data release (Kordopatis et al. 2013) and this work. Here we show the ‘calibrated’ metallicity (column `c_M_H_K`) from the *RAVE* survey. Only stars meeting quality constraints in *both* studies are shown (see text for details).

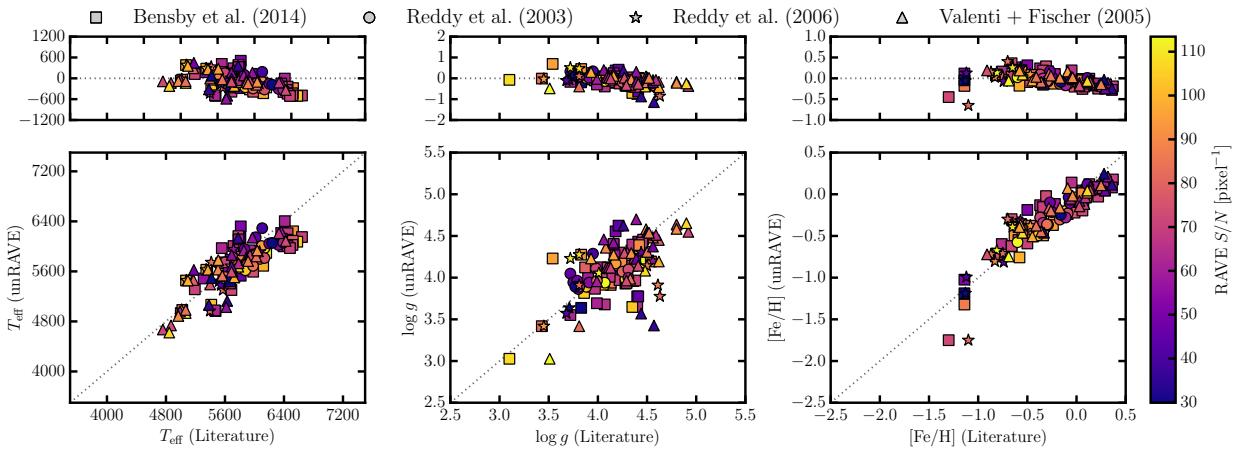


Fig. 10.— Stellar parameter (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$) comparisons for stars in common between this work and ‘gold standard’ studies that use high-resolution, high S/N spectra and *Hipparcos* parallaxes where available: Bensby et al. (2014); Reddy et al. (2003, 2006); Valenti & Fischer (2005). Stars are colored by the S/N of the *RAVE* spectra, and only stars with $\chi^2_r < 3$ and S/N > 10 are shown.

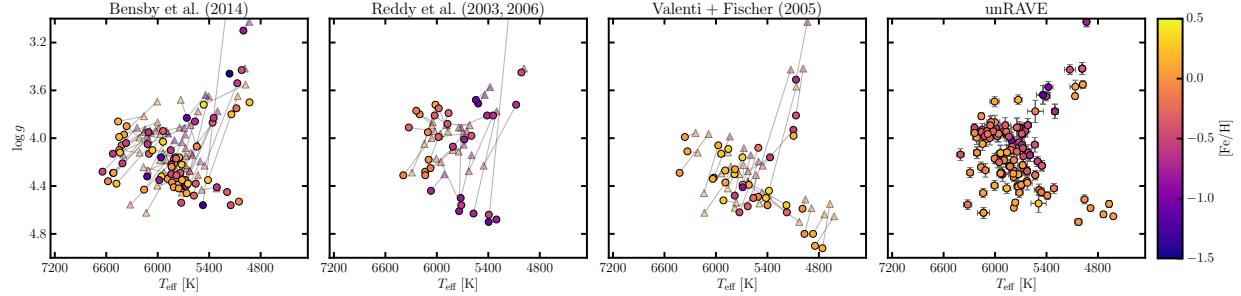


Fig. 11.— Hertsprung-Russell diagrams of stars in common between this work and that of Bensby et al. (2014); Reddy et al. (2003, 2006); Valenti & Fischer (2005). Stars are colored by the metallicity of each study. Circles indicate literature markers in the first three panels, and the linked triangles indicate *unRAVE* parameters for the same object. This figure illustrates the good overall agreement in the shape of the turnoff and sub-giant branch, yet also demonstrates how some literature metal-poor ($[{\rm Fe}/{\rm H}] \lesssim -1$) dwarfs near the turn-off appear as metal-poor sub-giants in *unRAVE*.

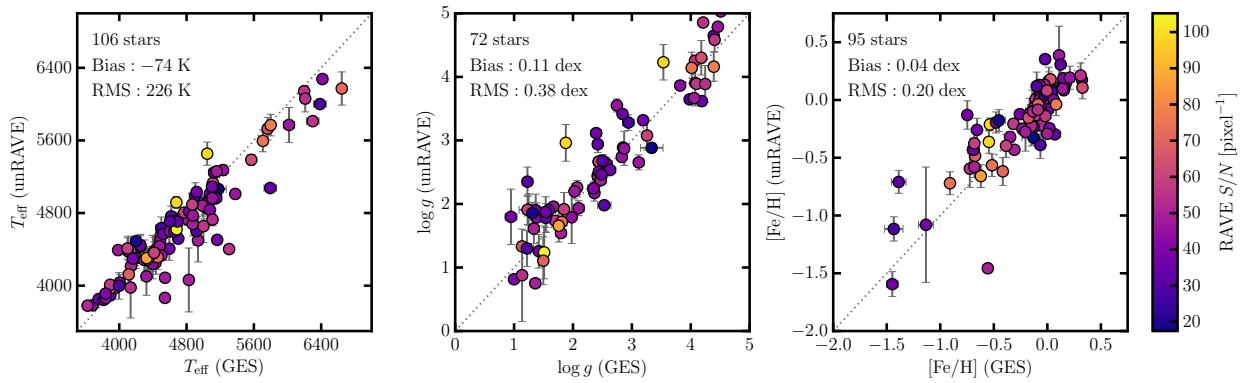


Fig. 12.— Stellar parameter (T_{eff} , $\log g$, $[{\rm Fe}/{\rm H}]$) comparison between the fourth internal data release from the *Gaia-ESO* survey, and this work. The number of stars in each panel are shown, as well as the bias and RMS deviation in each label. Stars are colored by the S/N of the *RAVE* spectra, and only stars with $\chi_r^2 < 3$ and S/N > 10 are shown. Most of the *Gaia-ESO/RAVE* overlap stars have relatively low S/N ratios in *RAVE*, near $\approx 30 \text{ pixel}^{-1}$.

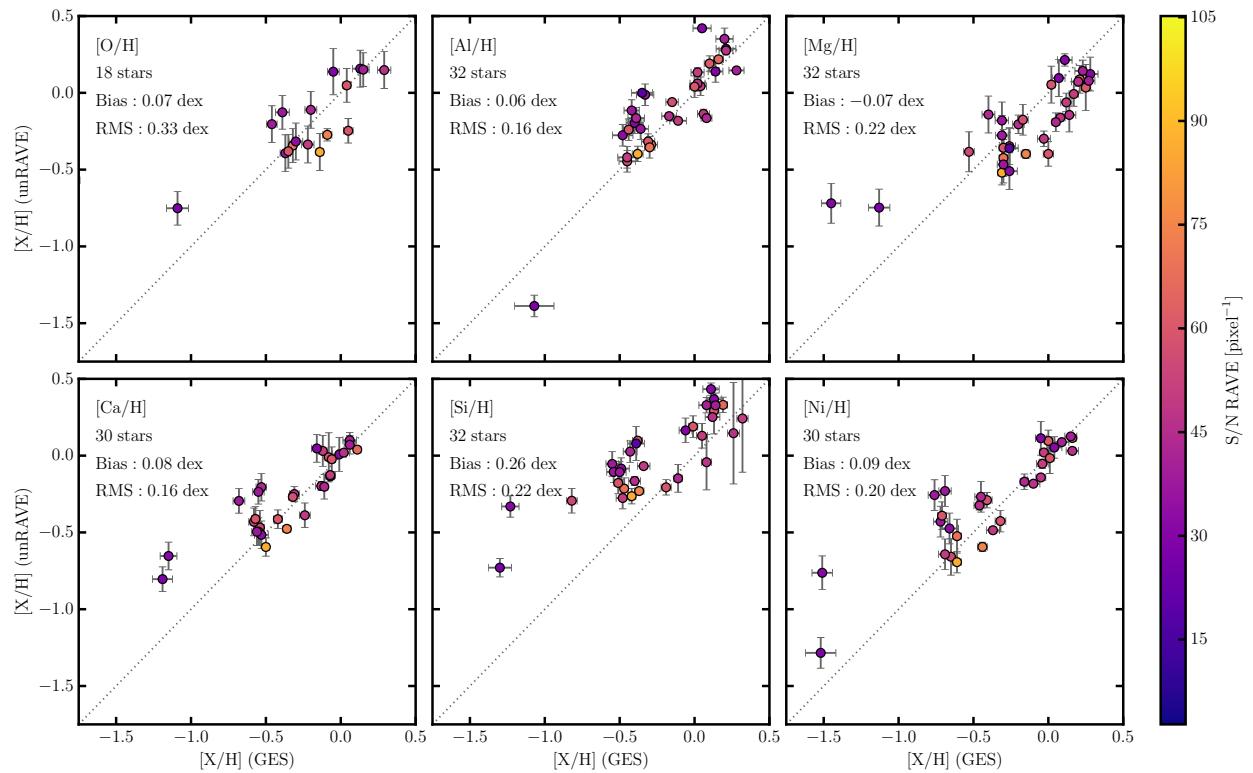


Fig. 13.— Detailed chemical abundances in the fourth internal data release from the *Gaia-ESO* survey compared to this work. The number of stars shown in each panel is indicated, and the bias and RMS deviations are shown. Stars are colored by the S/N of the *RAVE* spectra, and only stars with $\chi^2_r < 3$ and $S/N > 10$ are shown.

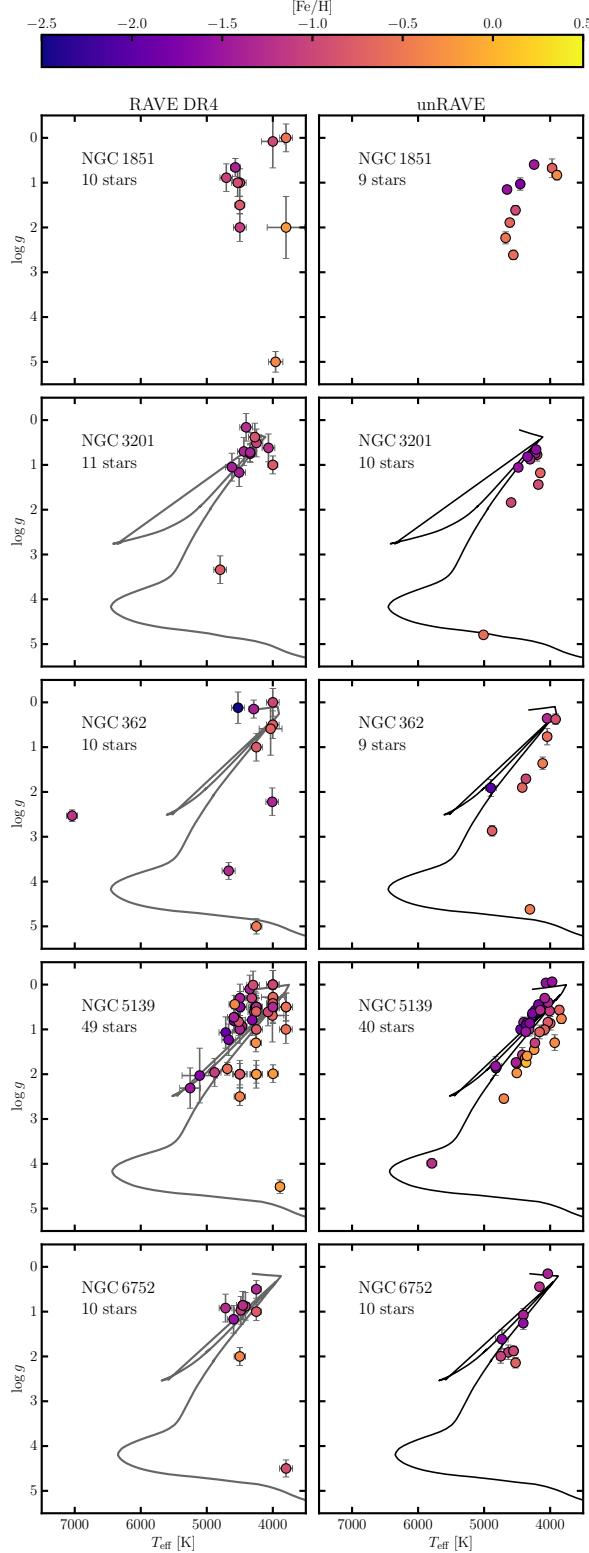


Fig. 14.— Effective temperature T_{eff} and surface gravity $\log g$ for globular cluster members identified by Kunder et al. (2014); Anguiano et al. (2015). Left-hand panels indicate results from the fourth *RAVE* data release (Kordopatis et al. 2013), and the right-hand panels show results from this work. A representative isochrone is shown for each cluster (Bressan et al. 2012).

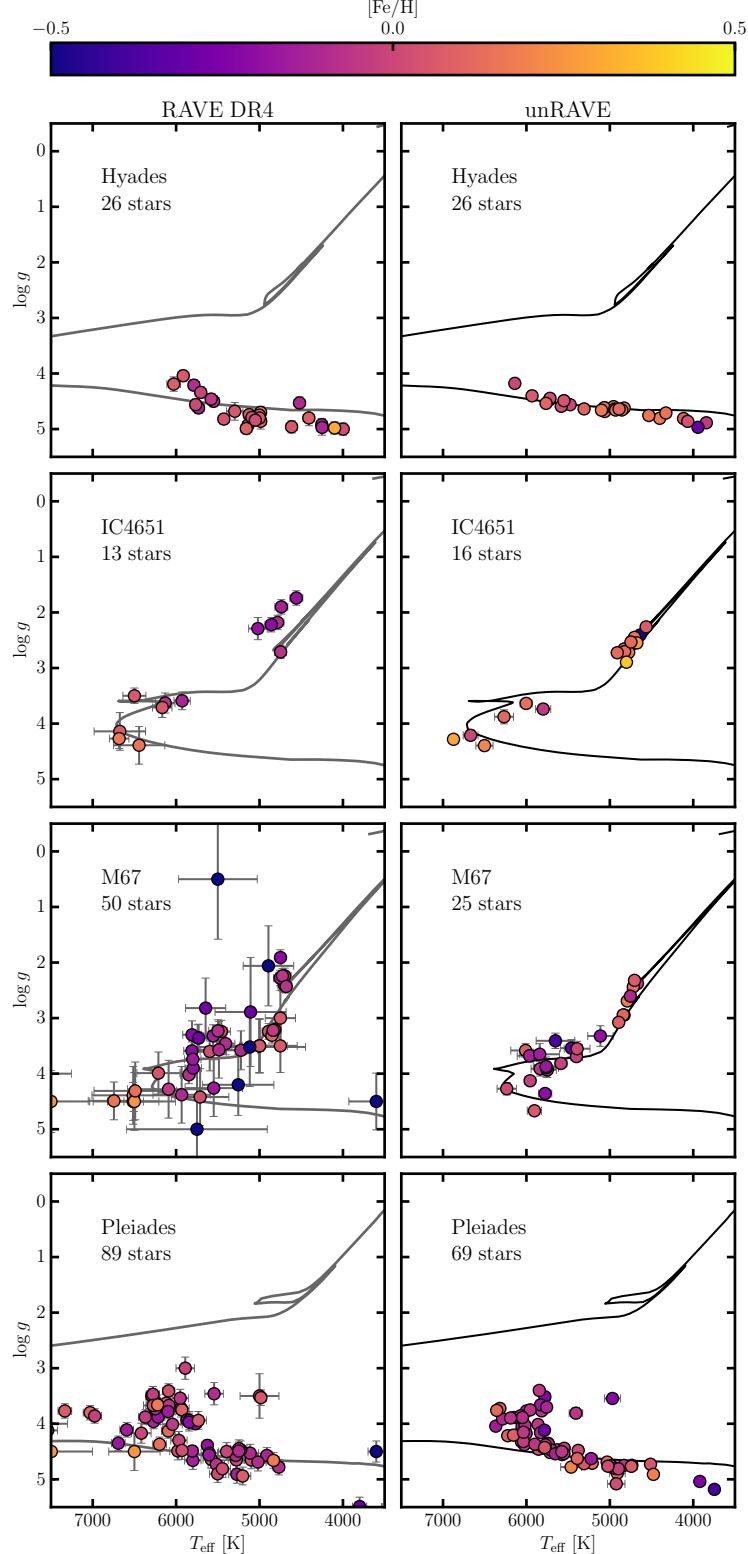


Fig. 15.— Effective temperature T_{eff} and surface gravity $\log g$ for open cluster members identified by (Kunder et al. 2016). Left-hand panels indicate results from the fourth *RAVE* data release (Kordopatis et al. 2013), and the right-hand panels show results from this work. A representative isochrone is shown for each cluster (Bressan et al. 2012).

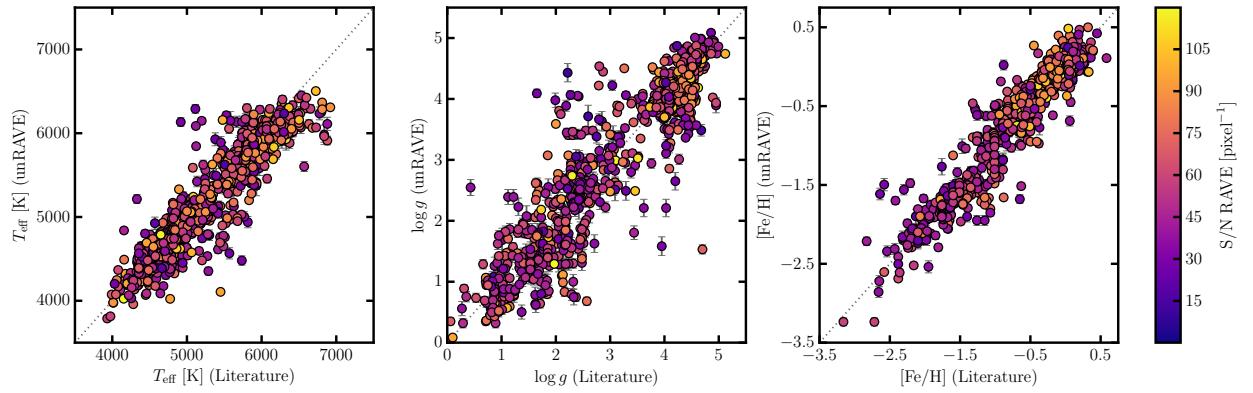


Fig. 16.— Stellar parameter (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$) comparison with the literature calibration sources used by Kordopatis et al. (2013) and Kunder et al. (2016). Stars are colored by the S/N of the RAVE spectra. Note that this comparison is for illustrative purposes only: it is not an indication of independent agreement with the literature because some metal-poor stars in this literature sample were used in the construction of our training set (see text for details).

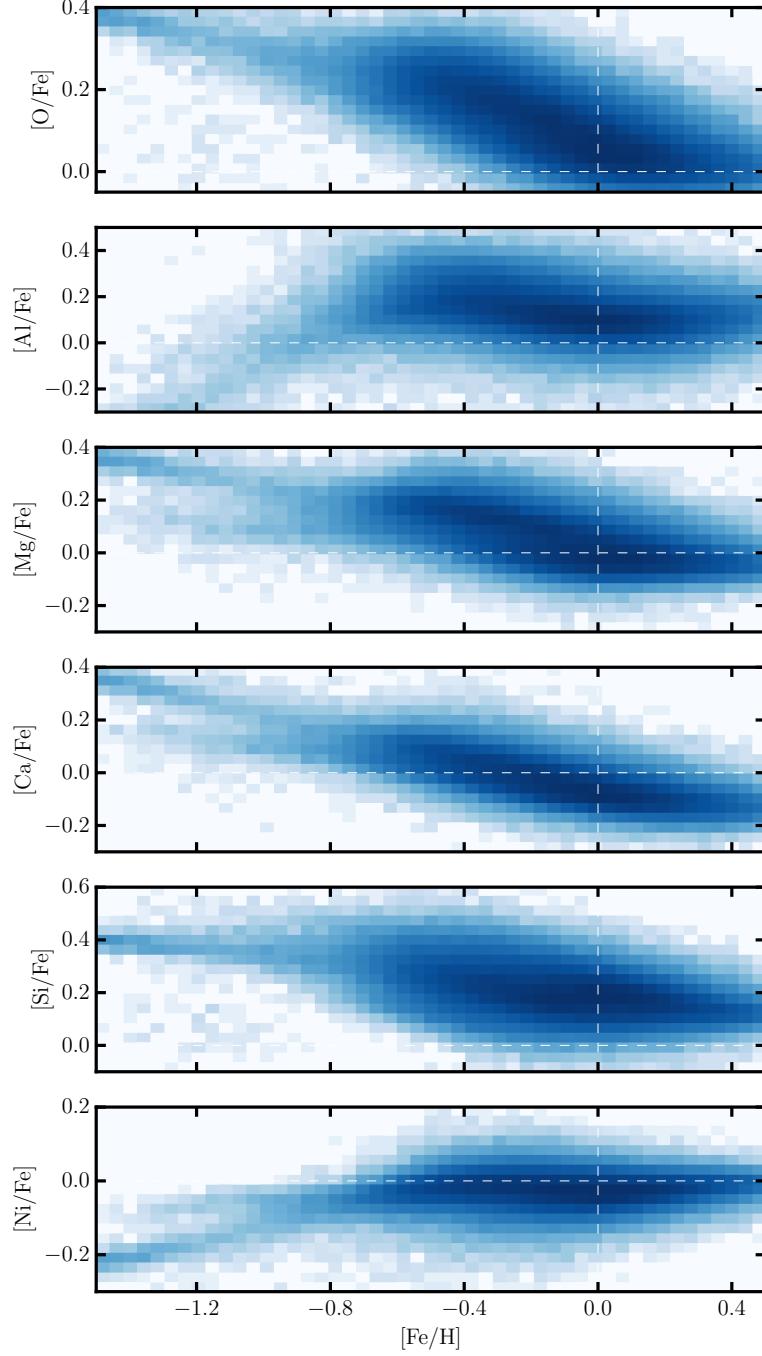


Fig. 17.— Detailed chemical abundances ($[X/Fe]$) for giant stars in *unRAVE* with respect to $[Fe/H]$, showing the Galactic chemical evolution derived for each element. Note that the y-axes limits vary for each panel, however for clarity we show the scaled-Solar position by dashed lines, and have common tick mark spacing on the y-axis for all panels. Only stars with $S/N > 10$ and $\chi^2_r < 3$ are shown.

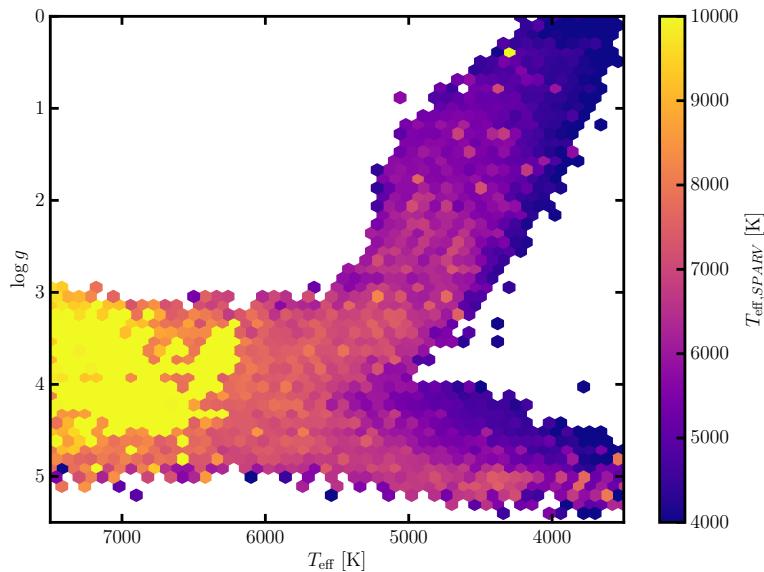


Fig. 18.— Effective temperatures T_{eff} and surface gravities $\log g$ from this work, where each hexagonal bin is colored by the *maximum* temperature for any star in that bin, as reported by the *RAVE* pre-processing pipeline **SPARV**. Hot stars are not included in our training set, and appear above the turnoff in our labels. For this reason, we supply a boolean quality control flag **QC** that flags stars with $T_{\text{eff}} > 8000$ K (as reported by **SPARV**), and other results that are known to be erroneous.