Draft version October 10, 2020
Typeset using LaTeX **twocolumn** style in AASTeX63

Searching for Low-mass Population III Stars Disguised as White Dwarfs

Vedant Chandra [1] and Kevin C. Schlaufman [1]

[1] *Department of Physics & Astronomy, Johns Hopkins University, 3400 N Charles St, Baltimore, MD 21218, USA*

ABSTRACT

It is uncertain whether or not metal-free low-mass stars ever existed. While limits on the number density of metal-free stars with $M_* \approx 0.8\ M_\odot$ have been derived using Sloan Digital Sky Survey (SDSS) data, little is known about the occurrence of metal-free stars at lower masses. In the absence of reliable parallaxes, the spectra of metal-poor main sequence (MPMS) stars with $M_* \lesssim 0.8\ M_\odot$ can easily be confused with cool white dwarfs. To resolve this ambiguity, we present a classifier that differentiates between MPMS stars and white dwarfs based on photometry and/or spectroscopy without the use of parallax information. We build and train our classifier using state-of-the-art theoretical spectra and evaluate it on existing SDSS-based classifications for objects with reliable Gaia DR2 parallaxes. We then apply our classifier to a large catalog of objects with SDSS photometry and spectroscopy to search for MPMS candidates. We discover several previously unknown candidate extremely metal-poor (EMP) stars and recover numerous confirmed EMP stars already in the literature. We conclude that archival SDSS spectroscopy has already been exhaustively searched for EMP stars. We predict that the lowest-mass metal-free stars will have redder optical-to-infrared colors than cool white dwarfs at constant effective temperature due to surface gravity-dependent collision-induced absorption from molecular hydrogen. We suggest that the application of our classifier to data produced by next-generation spectroscopic surveys will set stronger constraints on the number density of low-mass Population III stars in the Milky Way.

Keywords: Chemically peculiar stars (226); Low mass stars (2050); Population II stars (1284); Population III stars (1285); Sky surveys (1464); White dwarf stars (1799)

1. INTRODUCTION

The very first stars in the Universe, Population III stars, are theorized to have formed in the absence of metals around 100 Myr after the Big Bang (e.g., Bromm 2013). It was once believed that Population III stars only formed at high stellar masses $M_* \gtrsim 100\ M_\odot$ (e.g., Bromm et al. 1999; Abel et al. 2002). However, recent work has shown that fragmentation in the accretion disks around massive Pop III protostars could potentially form pristine stars at much lower masses (e.g., Stacy et al. 2010; Hartwig et al. 2015). If this is true, then numerous metal-free low-mass stars may survive to this day in the local universe.

To date, no Pop III star has been directly observed. Observational searches have instead found numerous metal-poor (Population II) stars, whose chemical abundances can be used to infer the properties of Pop III stars (e.g., Hartwig et al. 2015; Placco et al. 2016; Hansen et al. 2020), as well as the early chemical evolution of the Milky Way (e.g., Beers & Christlieb 2005; Frebel & Norris 2015). The vast majority of metal-poor stars studied to date are at least as luminous than the main sequence turnoff (Yong et al. 2013; Cohen et al. 2013; Roederer et al. 2014). Aside from early examples turned up in studies of high proper motion stars (e.g., Ryan & Norris 1991a,b; Ryan et al. 1991), metal-poor main sequence (MPMS) stars have been largely ignored in favor of their evolved counterparts. Assuming the Kroupa (2001, 2002) initial mass function (IMF), the luminosity function of metal-poor stars indicates that for every metal-poor turnoff star there are around 20 MPMS stars with $T_{\rm eff} \lesssim 5000$ K in the same volume (Bressan et al. 2012). As a result, only a small fraction of the Milky Way's metal-poor stellar population has been characterized to date.

One reason for this lack of attention is that the optical spectra of metal-poor (and metal-free) main se-

Corresponding author: Vedant Chandra
vchandra@jhu.edu

quence stars are virtually indistinguishable from the spectra of more numerous cool hydrogen-atmosphere white dwarfs (WDs). The high surface gravity of white dwarfs causes metals to rapidly sink out of their atmospheres via gravitational settling (Schatzman 1948). Therefore, the chemical composition (and consequently spectroscopic presentation) of most white dwarf atmospheres is very similar to that of metal-free stars. Additionally, a large fraction of white dwarf atmospheres are externally contaminated with metals (e.g., Gänsicke et al. 2012; Koester et al. 2014), usually due to ongoing accretion from circumstellar disks. Hence, the absence of metal lines is insufficient to separate white dwarfs from metal-poor stars.

The traditional method to distinguish white dwarfs from MPMS stars has been to visually identify MPMS stars on the basis of their narrower Balmer absorption lines at a given photometric color (Kepler et al. 2019). However, this process is very difficult for large datasets and prone to human error. This problem has historically been made worse by theoretical uncertainties inherent in modeling cool white dwarf atmospheres (Kepler et al. 2019). Recent advances have improved the fidelity of cool ($T_{\rm eff} \lesssim 5000$ K) white dwarf models (Blouin et al. 2018a,b), facilitating a theoretical comparison between MPMS and WD stars at low temperature ranges where insufficient observational data exists today.

There is a strong need for an automated way to search for metal-poor main-sequence candidates in current and upcoming large-scale spectroscopic surveys like the Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST; Cui et al. 2012), the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016), the WHT Enhanced Area Velocity Explorer (WEAVE; Dalton et al. 2012) and the Sloan Digital Sky Survey V (SDSS-V; Kollmeier et al. 2017). These surveys will produce tens of thousands of spectra with prominent Hydrogen lines and no strong metal features, rendering a visual comparison to differentiate MPMS and WD stars completely infeasible.

In this paper, we develop an automated framework to find and identify metal-poor main sequence stars in large spectroscopic surveys. We focus on the problem of differentiating MPMS stars from spectroscopically-similar white dwarfs. We present a simple classification algorithm to differentiate metal-poor main-sequence candidates from white dwarfs on the basis of spectrophotometric observables, fitted on synthetic data from theoretical models. We test and validate our methods on stars from the Sloan Digital Sky Survey (SDSS; York et al. 2000; Blanton et al. 2017) with parallax-confirmed classifications. Finally, we perform a search for EMP candidates in a large dataset of SDSS spectra. We host our classifier online[1], along with synthetic training data for a variety of photometric systems.

In Section 2 we describe state-of-the-art theoretical models for both MPMS and WD stars, and introduce the dataset of stars from SDSS we use to validate and test our methods. We discuss feature extraction from spectroscopic and photometric observables in Section 3 and outline our classification algorithm. Finally, we validate our classifier in Section 4 and perform a search for interesting metal-poor candidates in the SDSS dataset.

2. DATA

2.1. *Synthetic Data*

Our goal is to differentiate MPMS stars from white dwarfs on the basis of spectroscopy and photometry alone. To do so, we first generate a synthetic dataset of stars using the latest theoretical models. We define a grid of temperatures spanning 4000 K $< T_{\rm eff} <$ 8000 K in 40 K steps. At each temperature, we generate white dwarf spectra with $7 < \log g < 9$ and MPMS spectra with $3.5 < \log g < 5.5$, in steps of 0.5 dex. This results in a synthetic dataset of 500 stars that we use to train our classifier.

For the white dwarfs, our grid of synthetic spectra is derived from atmospheric models described in Blouin et al. (2018a,b). These models assume pure-hydrogen atmospheres, no magnetic field, and no metals present in the photosphere. Blouin et al. (2018a,b, 2019) describe their various developments over past models, especially in the cool ($T_{\rm eff} \lesssim 5000$ K) regime. These include an improved equation of state, and improved radiative opacities. For metal-poor main sequence stars, we use synthetic spectra computed with the PHOENIX code (Husser et al. 2013). Their parameter space extends across a range of temperatures and surface gravities, and includes metallicities $-4.0 \leq$ [Fe/H] $\leq +1.0$ and alpha abundances $-0.2 \leq [\alpha/{\rm Fe}] \leq +1.2$. In this work we assume a default solar α abundance, and use the [Fe/H] $= -4.0$ models as our theoretical MPMS spectra.

For each class of spectroscopic model (WD and MPMS) we tri-linearly interpolate the logarithm of the flux with respect to effective temperature, surface gravity, and the logarithm of wavelength. Some sample synthetic spectra are illustrated in Figure 1. All synthetic spectra are convolved with an instrumental resolution of 1.5 Å, comparable to the resolution of large-scale spectroscopic surveys like SDSS.

[1] https://github.com/vedantchandra/mpms

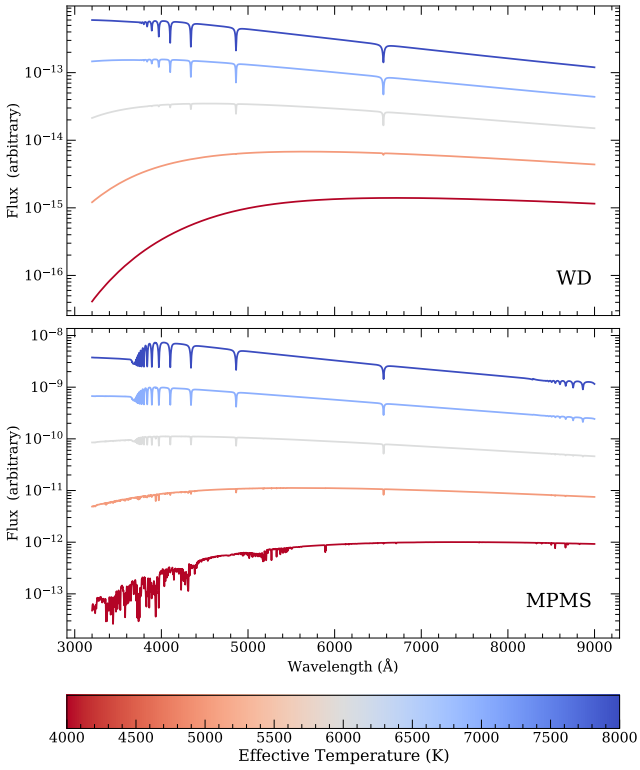

Figure 1. Sample of synthetic WD and MPMS spectra from the Blouin et al. (2018a) and PHOENIX libraries respectively. The displayed spectra span $4000 \leq T_{\rm eff} \leq 8000$ K, with $\log g = 8$ for the WDs and $\log g = 4.5$ for the MPMS stars. For the MPMS models we assume [Fe/H] $= -4.0$ with solar-scaled abundances. The log-fluxes are vertically displaced with arbitrary scaling constants for visual clarity. A very low-mass Population III star would be spectroscopically indistinguishable from the $T_{\rm eff} = 4000$ K white dwarf in the top panel.

2.2. *SDSS Data*

We evaluate the methods developed in this work using spectro-photometry from the SDSS and parallaxes from the Gaia space observatory (Mignard et al. 2018). In search of white dwarf candidates, Kepler et al. (2019) selected a sample of 37 053 SDSS spectra that consists of a mixture of white dwarfs and subdwarf A (sdA) stars (i.e., likely metal-poor main-sequence stars). Amongst the white dwarfs, over 78% were classified as having hydrogen-rich atmospheres (DA) on the basis of broad Balmer absorption lines with no other strong spectral features. Kepler et al. (2019) visually differentiated DA white dwarfs and subdwarf A stars under the assumption that DA white dwarfs have broader Balmer lines at a given photometric color, since their high surface gravities contribute to increased pressure broadening.

We begin by restricting our sample to only those stars identified by Kepler et al. (2019) as `'DA'`, `'sdA'`, or

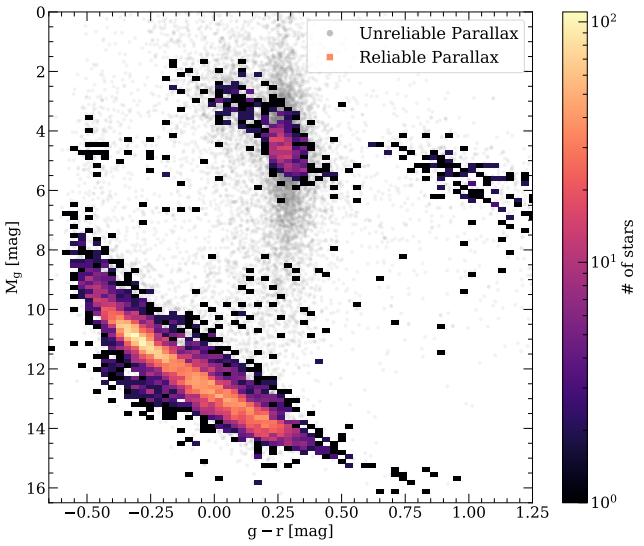

Figure 2. Color–magnitude diagram of the stars in our SDSS spectroscopic sample that includes subdwarf A stars and white dwarfs. The stars with unreliable parallaxes ($\sim$ 75% of the sample) are not included in the histogram and are plotted in gray. There is a clear separation between the white dwarf track on the bottom-left and the more luminous main sequence stars at the top.

`'sdA/F'` – this limits our sample to spectra with prominent hydrogen Balmer lines and no strong metal features, and removes other white dwarf spectral types like helium-rich WDs (DB). This leaves us with 14 522 stars that are spectroscopically classified as either hydrogen-rich white dwarfs or subdwarfs. The prevailing theory is that most of the subdwarf contaminants are metal-poor A/F type stars (Brown et al. 2017; Pelisoli et al. 2018b,a). We mention for completeness that a very small fraction of the sdA contaminants could also be so-called extremely low-mass (ELM) white dwarfs (Kosakowski et al. 2020).

As mentioned in Section 1, the visual method of identifying subdwarfs from white dwarfs is prone to error, and is especially uncertain in the temperature regimes where sdA and DA spectra are visually similar. Therefore, Kepler et al. (2019) appealed to parallaxes from the Gaia space observatory Data Release 2 (Luri et al. 2018; Mignard et al. 2018) to distinguish subdwarf A stars and DAs. At a given color, white dwarfs are far less luminous than subdwarfs due to their smaller radii, and hence a good parallax measurement provides an accurate classification via absolute magnitudes. However, only $\sim$ 15% of the stars in this sample have reliable parallaxes ($\pi/\sigma_\pi > 10$), and therefore the classification between sdA stars and WDs cannot be confirmed for a large part of the sample.

When validating our methods, we use only those stars for which a definite classification can be made. We apply the following Gaia quality cuts for the validation sample of stars:

```
parallax_over_error > 10
AND visibility_periods_used > 8
AND astrometric_sigma5d_max < 1
```

This results in a validation sample of 1807 stars with reliable subdwarf/white dwarf classifications, of which 65% are classified as DA white dwarfs and the remaining 35% are classified as subdwarfs. Figure 2 illustrates this validation sample on a Sloan color–magnitude diagram with Gaia parallaxes.

3. METHODS

In this section we describe automated methods to differentiate metal-poor main-sequence stars from white dwarfs using spectroscopic and photometric observables alone, without appealing to accurate parallax measurements. We develop and train these methods using the synthetic data described in Section 2.1.

3.1. *Hydrogen Balmer Lines*

Both MPMS and WD stars exhibit strong hydrogen Balmer lines on their spectra. Historically, for temperatures $T_{\rm eff} \gtrsim 6500$, MPMS stars have been differentiated from WDs on the basis of the Balmer lines (in the absence of a good parallax). The general rule of thumb is that MPMS stars have narrower Balmer lines (Kepler et al. 2019) at a given temperature. In this section we quantify this difference into a rigorous selection function by utilizing information across several Balmer lines at once.

We fit each Balmer absorption line – Hα, Hβ, Hγ, and Hδ – with a Voigt profile, a convolution of Gaussian and Lorentzian profiles that well–approximates the pressure-broadened wings of the Balmer lines (Tremblay & Bergeron 2009). From this fitted profile we derive two summary statistics for each line – the full-width at half-maximum (FWHM, in Angstroms), and the line amplitude (minimum of the profile, in continuum-normalized flux units). Therefore, for each spectrum we derive 8 line summaries in total which together quantify the phase space of the Balmer lines. Figure 3 illustrates two 2D 'slices' through this phase space, plotting width versus amplitude for a Hα (top) Hδ (bottom). In observational surveys, most white dwarfs have $T_{\rm eff} \gtrsim 6000$ K and most main sequence subdwarfs have temperatures $T_{\rm eff} \lesssim 8000$ K. Examining these temperature ranges on the respective curves, most subdwarfs clearly have much narrower (lower FWHM) $H\alpha$ than white dwarfs, matching the expected rule-of-thumb.

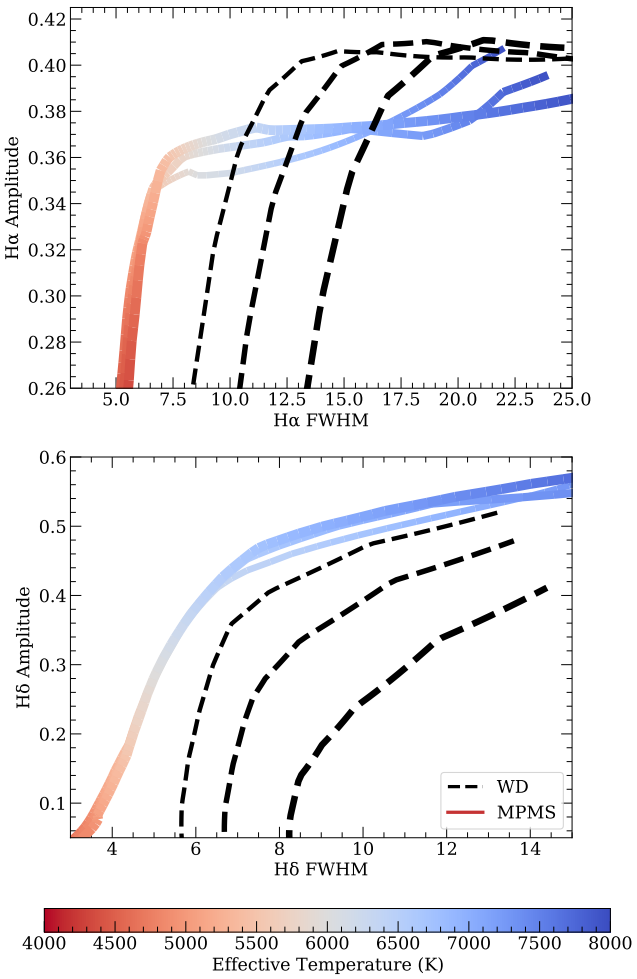

Figure 3. Visualization of two 'slices' of the Balmer line phase space for MPMS (solid lines colored by effective temperature) and WD (dashed lines) stars. Line thickness indicates the magnitude of surface gravity ($\log g$ = 4, 4.5, 5 for MPMS models and $\log g$ = 7.5, 8, 8.5 for WD models).

These plots are representative of how the relations change as a function of $T_{\rm eff}$ and $\log g$. Above $T_{\rm eff} \gtrsim 6000$ K the MPMS and WD populations are visually differentiable on many of similar 2D slices of the Balmer phase space. Below $T_{\rm eff} \lesssim 6000$ K, the higher-order Balmer lines on white dwarf spectra start to rapidly lose their intensity as thermal energy fails to populate the $n = 2$ atomic state. In this range, the combined information from several Balmer lines is helpful in identifying white dwarfs. At $T_{\rm eff} \lesssim 5000$ K, the absorption lines for both MPMS and WD stars begin to disappear altogether, resulting in a pure-continuum spectrum (Blouin et al. 2019). At these temperatures, there is no spectroscopic information available. We discuss this special case of the lowest-mass MPMS stars further in Section 4.4.

We combine the information from the first four Balmer lines to differentiate MPMS stars from WD stars. We compute the Balmer line summaries for the dataset of synthetic spectra described Section 2.1 as described above. This results in a vector of 8 Balmer features per spectrum, and a 'label' determining whether the spectrum is a white dwarf or MPMS star.

3.2. *Photometric Colors*

We compute photometric magnitudes from the WD and MPMS synthetic spectra using the PYPHOT utility (Fouesneau 2020). We integrate the apparent fluxes using the SDSS *ugriz* passbands to derive synthetic absolute magnitudes. In principle this methodology can be extended to any other photometric system. However, the majority of stars observed in future surveys will likely have *ugriz* measurements from SDSS or SkyMapper (Keller et al. 2007), or at the very least *griz* measurements from Pan-STARRS (Chambers et al. 2016). We focus on SDSS *ugriz* and *griz* colors in this paper since we validate our method on SDSS data. In the software tool accompanying this work[1], we allow the user to provide photometry from any of the following facilities: SDSS, Pan-STARRS, SkyMapper (Keller et al. 2007), DECam (Flaugher et al. 2015), and the Vera Rubin Observatory (Ivezic et al. 2019).

Figure 4 illustrates the comparison between MPMS and WD stars in Sloan color–color space. The $u-g$ color measures the Balmer jump, a sensitive probe of surface gravity, and hence it cleanly differentiates MPMS and WD stars at a given $g-r$ color (Mihalas 1967; Bessell 2007). However at lower temperatures, the difference becomes less pronounced. Therefore the range of temperatures within which this color–color diagram is adequate to visually differentiate MPMS and WD stars is limited to $T_{\rm eff} \gtrsim 6000$.

In this study, we don't restrict ourselves to any 2-D color-color diagram. Rather, we consider the entire 10-dimensional *ugriz* color space at once, computing all unique pairwise differences of the photometric bands ($u-g$, $u-r$, $u-i$, $u-z$, $g-r$, $g-i$, $g-z$, $r-i$, $r-z$, and $i-z$). Some photometric surveys like Pan-STARRS (Chambers et al. 2016) do not measure magnitudes in the u-band, so for comparison we additionally consider the 6-dimensional *griz* color space as well. We compute photomety for both cases using the same grid of synthetic spectra described in Section 2.

3.3. *Logistic Regression Classifier*

We are now armed with a consistent synthetic grid of spectroscopic and photometric observables for both WD and MPMS stars. To map these spectroscopic and photometric features to the stellar type, we use a logistic regression classifier. A logistic regression is a model that assumes a linear relationship between the input features and the log-odds of a binary (Bernoulli) random variable being 1 (or 'true'). Given input features $\vec{X} = [x_0, x_1, x_2, ...]$, fitting a logistic regression involves solving for the coefficients $\vec{\beta} = [\beta_0, \beta_1, \beta_2, ...]$ such that

$$\vec{\beta}^{\intercal}\vec{X} = \log\left(\frac{p}{1-p}\right) \quad (1)$$

where p is the probability that the Bernoulli random variable is 1 (or 'true'). The model is fitted on data with known labels (where p is known to be either zero or one), and the fitted coefficients $\vec{\beta}$ are subsequently used to estimate $\hat{p}$ for new input data. There is no closed-form solution to determine $\vec{\beta}$, and therefore an iterative gradient descent algorithm is used to find the optimal coefficients.

We define the underlying Bernoulli variable to have value 0 if the star is a white dwarf and 1 if it is a metal-poor main-sequence star. The associated probability in the model can therefore be defined as $p = P_{\rm MPMS}$, the probability that a given star is a metal-poor main-sequence star as opposed to a white dwarf. We demonstrate 3 possible input configurations: the Balmer line summary statistics (8 features), *ugriz* photometric col-

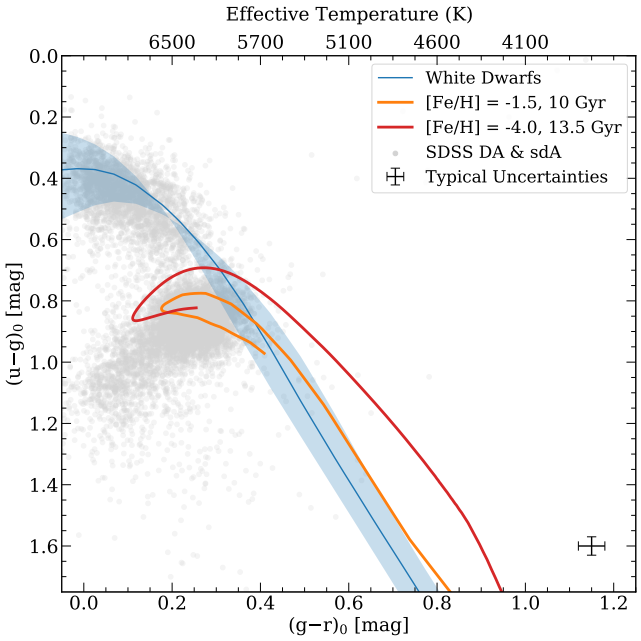

Figure 4. Synthetic color-color diagram using SDSS passbands. The blue line indicates white dwarfs with $\log g = 8$, and the shaded region indicates $7 < \log g < 9$. The red and orange lines are isochrones from MIST. For reference, we overlay all stars with strong hydrogen lines from the SDSS sample of Kepler et al. (2019) as gray points.

ors (10 features), and *griz* photometric colors (6 features). We also consider a combined classifier that uses *ugriz* colors and Balmer features simultaneously.

When training our logistic regression models, we randomly sample and reserve 5% of the stars in the dataset of synthetic observables, leaving them out of the training step. We evaluate the model on these unseen samples and confirm that it correctly predicts their stellar type close to 100% of the time. This validation step is simply a sanity check to ensure the logistic regression is working as expected on noiseless synthetic data – we further validate the model on SDSS spectro-photometry of stars with reliable parallaxes (and therefore, confident stellar types) in Section 4.

In this work we employ a logistic regression model due to its simple form and ease of interpretation. The classification probability $P_{\rm MPMS}$ returned by the logistic regression is well-calibrated by default (Yu et al. 2011), providing an accurate confidence in the classification. We find that using a more sophisticated classification algorithm like a random forest or support vector machine needlessly increase the complexity of the method without much increase in accuracy. Regardless, we make note of them here in case future studies wish to employ a different classification algorithm. We caution that most other algorithms require an external 'calibration function' to transform the returned classification probabilities to statistically meaningful values (Niculescu-Mizil & Caruana 2005).

3.4. *Equivalent Widths of Metal Lines*

After differentiating MPMS stars from white dwarfs, the next important step is to select the most promising and extremely metal-poor candidates among them. In order to select the best MPMS candidates, we require a method to approximately quantify stellar metallicity from the spectrum. The equivalent width is a ubiquitous measurement of the intensity of absorption lines that is relatively independent of the line shape and instrument resolution. On high-resolution spectra, the equivalent width is usually computed on the datapoints themselves – however our SDSS spectra are mid-resolution and often have intermediate signal-to-noise ratio.

Therefore, we add an initial step of fitting a noiseless line profile to the spectroscopic features, and then computing the equivalent width of the fitted profile. To each absorption line under consideration, we fit a pseudo-Voigt profile (a convex linear combination of a Gaussian and a Lorentzian profile) using a nonlinear least-squares algorithm (Newville et al. 2014). We compute the equivalent width (EW) of the fitted profile computed on the original wavelength grid, using numerical quadrature integration. For our selection in this study, we fit and consider the Ca II K (3934.8 Å) line and Ca II triplet (8500.4 Å, 8544.4 Å, 8664.5 Å), which are prominent tracers of metallicty even on mid-resolution spectra.

4. RESULTS

4.1. *Validation Stars with Reliable Parallaxes*

As described in Section 2, we have a sample of 1807 stars with strong hydrogen features and reliable Gaia parallaxes, enabling a definite classification between white dwarfs and MPMS stars. We assume that Kepler et al. (2019) correctly classified these stars, given that they utilized this parallax information as well, and visually inspect the color–magnitude diagram to ensure that this holds true. We take these classifications as the 'ground truth' data classes. We use the SDSS spectrophotometry for these validation stars to compute Balmer features and *ugriz* colors, and run our logistic regression classifier (Section 3.3) on these features, which produces a probability that each star is an MPMS star.

One way to evaluate our classifier is the receiver operating characteristic (ROC) curve (Fawcett 2006). This curve maps the true-positive and false-positive rates of the classifier as a function of the acceptance threshold, the output probability above which a 'positive' classification is made. It provides a broad overview of the sensitivity (probability of detection) and specificity (probability of false alarm). Another way to evaluate our classifier is the precision-recall (PR) curve (Powers 2011). This is particularly useful in our application, since precision (the probability that an MPMS star is correctly identified as such by our classifier) is arguably the most important metric. Additionally, the PR curve is more agnostic to a class imbalance in the dataset – in our sample, there are around two times as many white dwarfs as MPMS stars. Together, these diagnostic curves provide a holistic overview of the performance of our classifications.

We illustrate our validation ROC curves in Figure 5. We evaluate our logistic regression classifier on solely the 8 Balmer line summaries, the 10 *ugriz* colors, and 6 *griz* colors, as well as a combination of the Balmer features and *ugriz* colors. As expected, in terms of observables the Balmer spectroscopic features provide the most discriminating power. Including the $u-g$ color greatly improves the quality of classification compared to using *griz* alone, since it captures $\log g$ information via the Balmer jump (Mihalas 1967; Bessell 2007). In terms of precision, the combined spectro-photometric classifier far outperforms the solely spectroscopic classifier. This implies that a degeneracy in the spectroscopic features (with respect to the classification problem at hand)

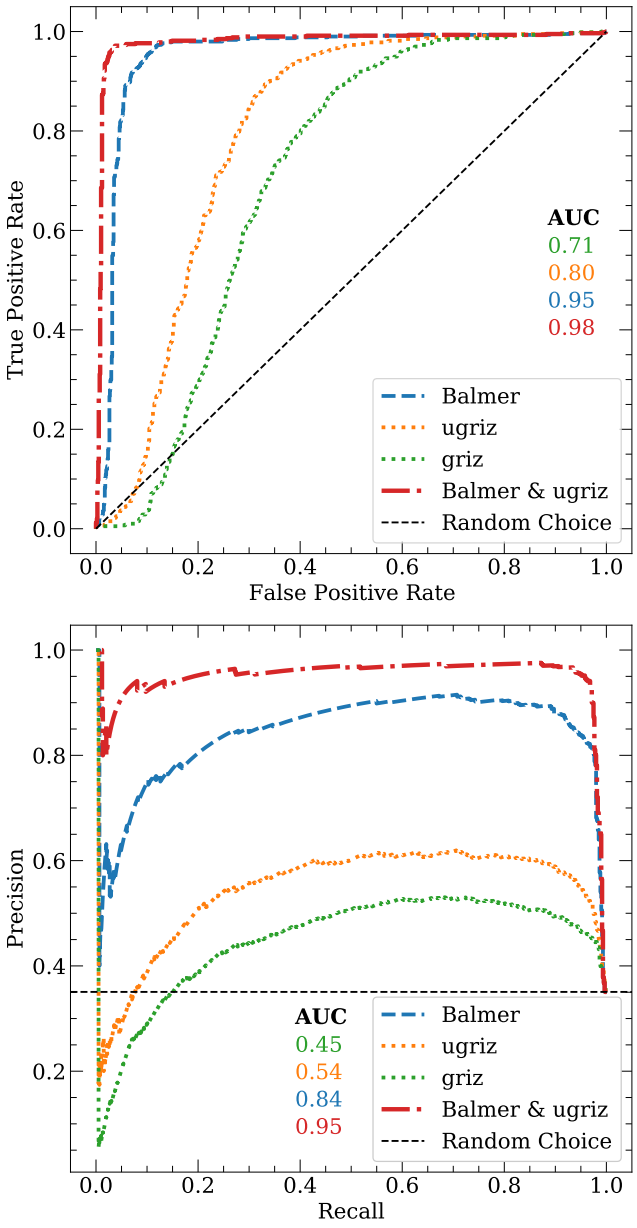

Figure 5. Receiver operating characteristic (ROC) and precision-recall (PR) curves for the pure-spectroscopic, pure-photometric, and combined logistic regression classifiers respectively, trained on theoretical models and validated on SDSS spectra with reliable parallaxes.

can be overcome by including photometric information. That natural interpretation is that the photometric colors provide a strong constraint on temperature, breaking the $T_{\rm eff}$-$\log g$ degeneracy of the absorption line features (Blanco-Cuaresma 2019). We conclude that our spectro-photometric classifier is suitably accurate and precise to identify metal-poor main sequence stars from the SDSS dataset.

4.2. *Classifying Stars with Unreliable Parallaxes*

With a reliable classifier in hand, we turn to the remaining $\sim$ 12 000 stars from Kepler et al. (2019) with strong hydrogen features and unreliable parallaxes. We fit the hydrogen Balmer lines of these spectra to derive line summaries, and apply our combined logistic regression classifier to the Balmer features and *ugriz* colors of these stars. This provides a probability measure $P_{\rm MPMS}$ of each star being an MPMS star as opposed to a white dwarf. As expected, the distribution of probabilities is bimodal – most stars have probabilities close to zero or one, indicating a strong confidence in the binary classification. We select a relatively liberal decision threshold of 0.5, marking all stars with $P_{\rm MPMS} > 0.5$ as MPMS candidates.

Comparing these classifications to those made by visual inspection in (Kepler et al. 2019), we find agreement for $\approx 95\%$ of the stars. The 5% of cases where our classifier disagrees with Kepler et al. (2019) primarily consist of stars with bad data, usually spectra with extremely low S/N. Overall, there is broad agreement between our computerized classification and the visual inspection of Kepler et al. (2019). We reiterate that in future spectroscopic surveys, where the number of spectra will range in the hundreds of thousands, the computerized approach will certainly be preferred.

4.3. *EMP Candidate Selection*

We emulate future searches for EMP candidates in upcoming surveys like SDSS-V and DESI by mining the stars in our SDSS dataset for EMP candidates. Several studies have already searched this SDSS data for metal-poor candidates (e.g., Bonifacio et al. 2012; Aoki et al. 2013), so we expect the best candidates to already be identified in the literature. We select all stars with $P_{\rm MPMS} > 0.5$ according to our combined clasifier as possible MPMS candidates, and also apply a selection cut on color of $g - r > 0.25$. Below this color, the calcium absorption lines become subdued due to higher temperatures rather than low metallicity, so we apply this cut to minimize false positives. This leaves us with ≈ 9000 MPMS spectra for which we compute equivalent widths as described in Section 3.4. Next, we perform a blind search for potentially metal-poor candidates on the basis of the equivalent widths of the Calcium absorption lines. We select stars with the lowest equivalent widths within a range of colors and visually inspect their spectra, examining the Ca II K (3934.8 Å), Na I (5895.6 Å), and Mg I (5176.7 Å) lines in particular.

Finally, we search the literature for each of our selected candidates using the SIMBAD database (Wenger et al. 2000). Our selection turns up dozens of previously-

Gaia DR2 Source ID	g_0 (mag)	$(g-r)_0$ (mag)	EW$_{\rm Ca\ II\ K}$ (Å)
2580454937423535744	17.2	0.27	0.61
923068053360267136	16.4	0.3	0.88
2757064989066940928	17.5	0.21	0.84
2572124320071981696	18.0	0.29	0.66
4427787686456496128	17.0	0.33	0.94
2540804903154179840	16.8	0.23	0.97
1041117669033018112	17.0	0.27	0.88
926917477929764608	16.6	0.25	0.87
2294041061855802496	17.4	0.18	0.97
881205690025926400	17.4	0.27	0.97
1543322422520153472	16.9	0.27	0.97
2823062869583476224	17.8	0.22	0.96
4448895182975654784	17.8	0.26	0.97
1214202544663262848	17.8	0.28	0.95
3092216989574463616*	17.1	0.25	0.51
290930261314166528*	15.8	0.33	0.98
2779958921396953088*	16.9	0.29	0.71
1276882477044162688*	16.6	0.31	0.92
4190837398756490112*	17.0	0.18	0.42
2782111559005651584†	15.5	0.27	0.82
1184737183522291712†	16.2	0.27	0.87
3740963179636227968†	16.6	0.33	0.66
3976087728282022272‡	16.9	0.31	0.95
3890626773968983296§	16.9	0.41	0.43

Table 1. Top: EMP candidates identified in our study with no previous record in the literature, ranked by subjective interest. Bottom: EMP stars identified by our selection that do have a prior bibliographic record and a measured [Fe/H] < -3.0 in SIMBAD. The symbol indicates the reference for the metallicity measurement: *Aoki et al. (2013), †Bonifacio et al. (2012), ‡Yoon et al. (2016), §Caffau et al. (2011).

known metal-poor stars with confirmed metallicity [Fe/H] < -3.0. A few of our selected candidates have no prior reference in the literature. We tabulate the known and unknown candidates that are the most apparently metal-poor Table 1. Although we find some new candidates, we conclude that the SDSS data has already been well-mined for extremely metal-poor stars. Regardless, this serves as a comprehensive validation of our methodology to search for metal-poor stars in future spectroscopic surveys.

4.4. *The Least-massive Primordial Stars*

The lowest mass primordial-composition object that can burn hydrogen should have $M_* \approx 0.1\ M_\odot$. Stellar models therefore predict a lower limit effective temperature $T_{\rm eff} \approx 3600$ K for metal-free stars (Burrows et al. 1993; Saumon et al. 1994). At these temperatures, the optical spectra of the coolest primordial stars will be indistinguishable the optical spectra of white dwarfs, since both will have pure-continuum spectra with no Balmer or metal lines.

However, collision-induced absorption (CIA) from molecular hydrogen will produce surface gravity-dependent continuum features in the near-infrared to infrared portion of the spectrum (Saumon et al. 1994; Blouin et al. 2017, 2019). At a given temperature (or $g-r$ color), a white dwarf will have more CIA than an MPMS star due to its higher surface gravity, and will consequently be relatively fainter in photometric bands like JHK and even iz. Therefore, a photometric color from the optical to the infrared (e.g., $V-J$ or $r-z$) will be bluer for a white dwarf compared to an MPMS star at the same temperature. This effect is larger than typical photometric uncertainties in ground-based surveys like SDSS, Pan-STARRS, and SkyMapper (Saumon et al. 1994; Blouin et al. 2017).

Uncertainties in the theoretical modelling of CIA make it challenging to precisely quantify this discrepancy and build a predictive model like the one presented in this work for warmer stars. Regardless, the photometric signature of CIA should certainly be detectable via infrared photometry and consequently differentiate the coolest pristine main-sequence stars from white dwarfs without parallax information. Future work on CIA opacities will enable a more quantitative comparison.

5. DISCUSSION

In this work, we present an automated framework to differentiate metal-poor main sequence stars from white dwarfs in future large-scale spectroscopic surveys, in the absence of reliable parallax measurement. We train a classification algorithm on synthetic data from theoretical models, and validate it using a large dataset of stars from the Sloan Digital Sky Survey with definitive classifications. We further apply our classifier to SDSS stars with unreliable parallaxes, and perform a search to demonstrate its utility in identifying extremely metal-poor candidates. Our code is hosted online for reproducibility, along with a simple software tool in PYTHON that enables anyone to use our classifier with their data[1].

It remains an open question whether metal-free and low-mass primordial stars exist in the local universe. A promising technique to probe the hypothesis of surviving Pop III stars (apart from awaiting serendipitous direct detections) is the use of galactic archaeology to constrain the primordial initial mass function. Hartwig et al. (2015) propose that a total sample size of $\sim 10^7$

halo stars is required to rule out Pop III survivors at the 95% confidence level. However, this is likely an overestimate since it only considers blind surveys – in practice, searching for extremely metal-poor (EMP) and ultra metal-poor (UMP) stars is far more efficient. Magg et al. (2019) instead propose to use the occurrence of EMP and UMP stars to constrain the existence of Pop III survivors. The largest uncertainty in this method, by far, is estimating of the total number of EMP and UMP stars in the Milky Way halo, and therefore a comprehensive search for such stars is needed.

Future large-scale spectroscopic surveys like DESI, WEAVE, and SDSS-V will observe over ten million new stars, with potentially thousands of extremely metal-poor candidates among them. The methods presented in this work will facilitate the rapid and reliable identification of metal-poor stars, providing new constraints on the uncertain existence of surviving pristine stars from the primordial Universe.

ACKNOWLEDGMENTS

We are grateful to Simon Blouin for insightful comments and suggestions, and for providing his latest grid of model spectra for cool white dwarfs. We thank JJ Hermes for helpful conversations and feedback. Vedant Chandra is gratefully supported by the Institute for Data Intensive Engineering & Science (IDIES) at Johns Hopkins University.

Based on observations from the Sloan Digital Sky Survey. Funding for the Sloan Digital Sky Survey IV (Blanton et al. 2017; Ahumada et al. 2020) has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. This work has made use of data from the European Space Agency (ESA) mission Gaia (Prusti et al. 2016; Mignard et al. 2018), processed by the Gaia Data Processing and Analysis Consortium (DPAC). This research has made use of the SIMBAD database (Wenger et al. 2000), operated at CDS, Strasbourg, France. This research has made use of NASA's Astrophysics Data System (ADS).

Software: numpy (Harris et al. 2020), scipy (Virtanen et al. 2020), matplotlib (Hunter 2007), astropy (Robitaille et al. 2013; Price-Whelan et al. 2018), scikit-learn (Pedregosa et al. 2011), lmfit (Newville et al. 2014)

Facilities: Sloan, Gaia

REFERENCES

Abel, T., Bryan, G. L., & Norman, M. L. 2002, Science, 295, 93, doi: 10.1126/science.295.5552.93

Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, The Astrophysical Journal Supplement Series, 249, 3, doi: 10.3847/1538-4365/ab929e

Aoki, W., Beers, T. C., Lee, Y. S., et al. 2013, Astronomical Journal, 145, doi: 10.1088/0004-6256/145/1/13

Beers, T. C., & Christlieb, N. 2005, ARA&A, 43, 531, doi: 10.1146/annurev.astro.42.053102.134057

Bessell, M. S. 2007, PASP, 119, 605, doi: 10.1086/519981

Blanco-Cuaresma, S. 2019, MNRAS, 486, 2075, doi: 10.1093/mnras/stz549

Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, The Astronomical Journal, 154, 28, doi: 10.3847/1538-3881/aa7567

Blouin, S., Dufour, P., & Allard, N. F. 2018a, The Astrophysical Journal, 863, 184, doi: 10.3847/1538-4357/aad4a9

Blouin, S., Dufour, P., Allard, N. F., & Kilic, M. 2018b, The Astrophysical Journal, 867, 161, doi: 10.3847/1538-4357/aae53a

Blouin, S., Dufour, P., Thibeault, C., & Allard, N. F. 2019, The Astrophysical Journal, 878, 63, doi: 10.3847/1538-4357/ab1f82

Blouin, S., Kowalski, P. M., & Dufour, P. 2017, The Astrophysical Journal, 848, 36, doi: 10.3847/1538-4357/aa8ad6

Bonifacio, P., Sbordone, L., Caffau, E., et al. 2012, Astronomy and Astrophysics, 542, 1, doi: 10.1051/0004-6361/201219004

Bressan, A., Marigo, P., Girardi, L., et al. 2012, Monthly Notices of the Royal Astronomical Society, 427, 127, doi: 10.1111/j.1365-2966.2012.21948.x

Bromm, V. 2013, Reports on Progress in Physics, 76, 112901, doi: 10.1088/0034-4885/76/11/112901

Bromm, V., Coppi, P. S., & Larson, R. B. 1999, ApJL, 527, L5, doi: 10.1086/312385

Brown, W. R., Kilic, M., & Gianninas, A. 2017, The Astrophysical Journal, 839, 23, doi: 10.3847/1538-4357/aa67e4

Burrows, A., Hubbard, W. B., Saumon, D., & Lunine, J. I. 1993, ApJ, 406, 158, doi: 10.1086/172427

Caffau, E., Bonifacio, P., François, P., et al. 2011, Nature, 477, 67, doi: 10.1038/nature10377

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560. https://arxiv.org/abs/1612.05560

Cohen, J. G., Christlieb, N., Thompson, I., et al. 2013, ApJ, 778, 56, doi: 10.1088/0004-637X/778/1/56

Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, Research in Astronomy and Astrophysics, 12, 1197, doi: 10.1088/1674-4527/12/9/003

Dalton, G., Trager, S. C., Abrams, D. C., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, 84460P, doi: 10.1117/12.925950

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv e-prints, arXiv:1611.00036. https://arxiv.org/abs/1611.00036

Fawcett, T. 2006, Pattern Recognition Letters, 27, 861 , doi: https://doi.org/10.1016/j.patrec.2005.10.010

Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150, doi: 10.1088/0004-6256/150/5/150

Fouesneau, M. 2020, pyphot – A tool for computing photometry from spectra, GitHub. https://github.com/mfouesneau/pyphot

Frebel, A., & Norris, J. E. 2015, ARA&A, 53, 631, doi: 10.1146/annurev-astro-082214-122423

Gänsicke, B. T., Koester, D., Farihi, J., et al. 2012, MNRAS, 424, 333, doi: 10.1111/j.1365-2966.2012.21201.x

Hansen, C. J., Koch, A., Mashonkina, L., et al. 2020, Astronomy & Astrophysics. http://arxiv.org/abs/2009.11876

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2

Hartwig, T., Bromm, V., Klessen, R. S., & Glover, S. C. 2015, Monthly Notices of the Royal Astronomical Society, 447, 3892, doi: 10.1093/mnras/stu2740

Hartwig, T., Clark, P. C., Glover, S. C. O., Klessen, R. S., & Sasaki, M. 2015, ApJ, 799, 114, doi: 10.1088/0004-637X/799/2/114

Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Husser, T. O., Wende-Von Berg, S., Dreizler, S., et al. 2013, Astronomy and Astrophysics, 553, 1, doi: 10.1051/0004-6361/201219058

Ivezic, Z., Kahn, S. M., Tyson, J. A., et al. 2019, The Astrophysical Journal, 873, 111, doi: 10.3847/1538-4357/ab042c

Keller, S. C., Schmidt, B. P., Bessell, M. S., et al. 2007, Publications of the Astronomical Society of Australia, 24, 1, doi: 10.1071/AS07001

Kepler, S. O., Pelisoli, I., Koester, D., et al. 2019, Monthly Notices of the Royal Astronomical Society, 486, 2169, doi: 10.1093/mnras/stz960

Koester, D., Gänsicke, B. T., & Farihi, J. 2014, Astronomy and Astrophysics, 566, doi: 10.1051/0004-6361/201423691

Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, in Bulletin of the American Astronomical Society, 274. http://arxiv.org/abs/1711.03234

Kosakowski, A., Kilic, M., Brown, W. R., & Gianninas, A. 2020, The Astrophysical Journal, 894, 53, doi: 10.3847/1538-4357/ab8300

Kroupa, P. 2001, Monthly Notices of the Royal Astronomical Society, 322, 231, doi: 10.1046/j.1365-8711.2001.04022.x

—. 2002, Science, 295, 82, doi: 10.1126/science.1067524

Luri, X., Brown, A. G., Sarro, L. M., et al. 2018, Astronomy and Astrophysics, 616, A9, doi: 10.1051/0004-6361/201832964

Magg, M., Klessen, R. S., Glover, S. C., & Li, H. 2019, Monthly Notices of the Royal Astronomical Society, 487, 486, doi: 10.1093/mnras/stz1210

Mignard, F., Klioner, S. A., Lindegren, L., et al. 2018, Astronomy & Astrophysics, 616, A14, doi: 10.1051/0004-6361/201832916

Mihalas, D. 1967, ApJ, 149, 169, doi: 10.1086/149239

Newville, M., Ingargiola, A., Stensitzki, T., & Allen, D. B. 2014, Zenodo, , doi: 10.5281/ZENODO.11813

Niculescu-Mizil, A., & Caruana, R. 2005, in Proceedings of the 22nd International Conference on Machine Learning, ICML '05 (New York, NY, USA: Association for Computing Machinery), 625–632, doi: 10.1145/1102351.1102430

Pedregosa, F., Michel, V., Grisel, O., et al. 2011, Journal of Machine Learning Research, 12, 2825. http://scikit-learn.sourceforge.net.

Pelisoli, I., Kepler, S. O., & Koester, D. 2018a, Monthly Notices of the Royal Astronomical Society, 475, 2480, doi: 10.1093/mnras/sty011

Pelisoli, I., Kepler, S. O., Koester, D., et al. 2018b, Monthly Notices of the Royal Astronomical Society, 478, 867, doi: 10.1093/MNRAS/STY1101

Placco, V. M., Frebel, A., Beers, T. C., et al. 2016, ApJ, 833, 21, doi: 10.3847/0004-637X/833/1/21

Powers, D. 2011, Journal of Machine Learning Technologies, 2, 37

Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, The Astronomical Journal, 156, 123, doi: 10.3847/1538-3881/aabc4f

Prusti, T., De Bruijne, J. H., Brown, A. G., et al. 2016, Astronomy and Astrophysics, 595, doi: 10.1051/0004-6361/201629272

Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, Astronomy and Astrophysics, 558, 1, doi: 10.1051/0004-6361/201322068

Roederer, I. U., Preston, G. W., Thompson, I. B., et al. 2014, AJ, 147, 136, doi: 10.1088/0004-6256/147/6/136

Ryan, S. G., & Norris, J. E. 1991a, AJ, 101, 1835, doi: 10.1086/115811

—. 1991b, AJ, 101, 1865, doi: 10.1086/115812

Ryan, S. G., Norris, J. E., & Bessell, M. S. 1991, AJ, 102, 303, doi: 10.1086/115878

Saumon, D., Bergeron, P., Lunine, J. I., Hubbard, W. B., & Burrows, A. 1994, ApJ, 424, 333, doi: 10.1086/173892

Schatzman, E. 1948, Nature, 161, 61, doi: 10.1038/161061b0

Stacy, A., Greif, T. H., & Bromm, V. 2010, MNRAS, 403, 45, doi: 10.1111/j.1365-2966.2009.16113.x

Tremblay, P. E., & Bergeron, P. 2009, Astrophysical Journal, 696, 1755, doi: 10.1088/0004-637X/696/2/1755

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2

Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9, doi: 10.1051/aas:2000332

Yong, D., Norris, J. E., Bessell, M. S., et al. 2013, ApJ, 762, 26, doi: 10.1088/0004-637X/762/1/26

Yoon, J., Beers, T. C., Placco, V. M., et al. 2016, The Astrophysical Journal, 833, 20, doi: 10.3847/0004-637x/833/1/20

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, The Astronomical Journal, doi: 10.1086/301513

Yu, H.-F., Huang, F.-L., & Lin, C.-J. 2011, Machine Learning, 85, 41, doi: 10.1007/s10994-010-5221-8