Draft version October 15, 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 and Astronomy*
Johns Hopkins University
3400 N Charles St
Baltimore, MD 21218, USA

(Received October 15, 2020)

Submitted to AAS Journals

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 first generation of stars formed in the Universe was made of only the stable products of Big Bang nucleosynthesis: hydrogen, helium, and a tiny amount of lithium. These Population III stars are predicted to start forming around 100 Myr after the Big Bang (e.g., Bromm 2013; Glover 2013; Greif 2015). The earliest Population III star formation calculations suggested that inefficient cooling would require large Jeans masses and therefore that Population III stars would form with a characteristic stellar mass $M_* \sim 100\ M_\odot$ (e.g., Silk 1983; Tegmark et al. 1997; Bromm et al. 1999, 2002; Abel et al. 2000, 2002). However, more recent simulations have shown that fragmentation in the accretion disks around massive Population III protostars could potentially form pristine stars at much lower masses (e.g., Stacy et al. 2010, 2012, 2016; Clark et al. 2011a,b; Greif et al. 2011, 2012; Stacy & Bromm 2013, 2014; Dopcke et al. 2013; Riaz et al. 2018). While it is theoretically uncertain if these fragments survive or merge with the more massive protostar growing at the center of their parent accretion disk (e.g., Hirano & Bromm 2017), there is at least circumstantial observational evidence to suggest that they might survive (Schlaufman et al. 2018). If

Corresponding author: Vedant Chandra
vchandra@jhu.edu

these fragments do avoid merging, then metal-free low-mass stars might persist to the present day in the local Universe.

While no Population III star has been directly observed to date, observational searches have instead found numerous metal-poor stars with [Fe/H] $\lesssim -3$. The chemical abundances of these extreme Population II stars can be used to infer the properties of Population III stars (e.g., Hartwig et al. 2015; Placco et al. 2016; Fraser et al. 2017; 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 in detail to date are at least as luminous as 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 PARSEC evolutionary tracks (Bressan et al. 2012; Chen et al. 2014, 2015) and a Kroupa (2001, 2002) initial mass function (IMF), the luminosity function for stars with [Fe/H] ≈ -2.2 indicates that for every metal-poor turnoff star with $T_{\rm eff} \gtrsim 6000$ K there are about 20 MPMS stars with $T_{\rm eff} \lesssim 5000$ K in the same volume. As a result, only a small fraction of the Milky Way's metal-poor stellar population has been characterized to date.

One reason for the lack of attention paid to MPMS stars is that the optical spectra of metal-poor (and metal-free) main sequence stars are virtually indistinguishable from the optical spectra of more numerous cool hydrogen-atmosphere white dwarfs (WD). While a significant fraction of white dwarf photospheres are externally seeded with metals by ongoing accretion from circumstellar disks (e.g., Gänsicke et al. 2012; Koester et al. 2014), the high surface gravities of white dwarfs cause metals to rapidly sink out of their photospheres via gravitational settling (Schatzman 1948). The chemical compositions and consequently spectra of most white dwarf photospheres are therefore very similar to those of metal-free stars.

The traditional method to distinguish MPMS stars from white dwarfs 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). This process is prone to human error and very difficult for large datasets. These problems have historically been aggravated by imperfect models for the atmospheres of cool white dwarfs (Kepler et al. 2019). Recent advances have improved the fidelity of atmosphere models for white dwarfs with $T_{\rm eff} \lesssim 5000$ K (Blouin et al. 2018a,b), facilitating a theoretical comparison between MPMS star and white dwarf spectra at cool temperatures where observational data are lacking.

Tens of thousands of spectra with weak or non-existent metal lines produced by either cool white dwarfs or metal-poor/metal-free main sequence stars will be collected by ongoing and next-generation spectroscopic facilities/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 fifth phase of the Sloan Digital Sky Survey (SDSS-V; Kollmeier et al. 2017), and the WHT Enhanced Area Velocity Explorer (WEAVE; Dalton et al. 2012). The volume of data produced by these ongoing and future surveys will make it impossible to rely on the traditional human classification of MPMS stars and white dwarfs. This implies that there is a need for an automated way to differentiate between MPMS stars and white dwarfs.

In this paper, we develop an automated framework to identify MPMS stars in large spectroscopic surveys with a focus on the problem of differentiating MPMS stars from spectroscopically similar white dwarfs. We describe in Section 2 state-of-the-art theoretical spectra for both MPMS stars and white dwarfs. We then introduce the Sloan Digital Sky Survey (SDSS) photometric and spectroscopic data for MPMS stars and white dwarfs securely classified using Gaia Data Release 2 (DR2) parallaxes that we use to validate and test our methods. We outline our classification algorithm in Section 3, including our extraction of features from photometric and spectroscopic observables and the validation of our classifier. In Section 4 we perform a search for candidate extremely metal-poor stars in a sample of possible MPMS stars with SDSS photometry and spectroscopy. We review our results and their implications in Section 5. We conclude by summarizing our findings in Section 6.

2. DATA

Our goal is to differentiate MPMS stars from white dwarfs on the basis of spectroscopy and photometry alone. While there is enough observational data for MPMS stars and white dwarfs with $T_{\rm eff} \gtrsim 6000$ K, the relative faintness of MPMS stars and white dwarfs with $T_{\rm eff} \lesssim 6000$ K has resulted in few examples in the SDSS spectroscopic archive. Even though these cool MPMS stars and white dwarfs were largely out of reach of the Sloan Foundation 2.5 m Telescope, spectroscopic surveys using 4 m-class telescopes like DESI will target many cool MPMS stars and white dwarfs. It is therefore necessary for us to use both theoretical spectra for

cool MPMS stars and white dwarfs and empirical spectra for warmer MPMS stars and white dwarfs to train and validate our classifier. We describe those data in the following two subsections.

2.1. *Theoretical Spectra*

We first construct grids of theoretical spectra for white dwarfs and MPMS stars using state-of-the-art models. We consider temperatures in the range 4000 K $\leq T_{\rm eff} \leq$ 8000 K in 40 K steps. For each temperature, we generate white dwarf spectra with $7 \leq \log g \leq 9$ and MPMS star spectra with $3.5 \leq \log g \leq 5.5$ in steps of 0.5 dex. This procedure results in grids of 505 theoretical spectra that we use to train our classifier. For the white dwarfs, our grid of theoretical spectra is derived from atmospheric models described in Blouin et al. (2018a,b). These models assume pure-hydrogen atmospheres, no magnetic fields, and no metals present in the photosphere. Blouin et al. (2018a,b, 2019) describe the improved equation of state and radiative opacities featured in this new generation of models that are especially important for cool white dwarfs with $T_{\rm eff} \lesssim 5000$ K. For MPMS stars, we use theoretical spectra computed with the PHOENIX code (Husser et al. 2013). Those models cover a broad range of temperatures and surface gravities and includes metallicities $-4.0 \leq$ [Fe/H] $\leq +1.0$ and α abundances $-0.2 \leq$ [α/Fe] $\leq +1.2$. In our theoretical grid of MPMS stars spectra, we use the [Fe/H] $= -4.0$ spectra with solar α abundances.

For both grids of spectra, we tri-linearly interpolate the logarithm of the fluxes with respect to effective temperature, surface gravity, and the logarithm of wavelength. We convolve all theoretical spectra to match an instrumental resolution of 1.5 Å, comparable to the resolution of large-scale spectroscopic surveys like the SDSS, LAMOST, and DESI. We illustrate some sample theoretical spectra in Figure 1.

2.2. *SDSS Data*

Our empirical data for warmer MPMS stars and white dwarfs come mostly from SDSS DR16 (Ahumada et al. 2020). They were collected during the first four phases of the SDSS (York et al. 2000; Eisenstein et al. 2011; Blanton et al. 2017), including its Sloan Extension for Galactic Understanding and Exploration (SEGUE), Baryon Oscillation Spectroscopic Survey (BOSS), and extended BOSS (eBOSS) programs (Yanny et al. 2009; Dawson et al. 2013, 2016). The data were collected using the Sloan Foundation 2.5 m Telescope and its imager and optical spectrographs (Gunn et al. 1998, 2006; Doi et al. 2010; Smee et al. 2013) then placed on the SDSS photometric system using the methods described in Fukugita et al. (1996), Smith et al. (2002), and Padmanabhan et al. (2008). We also make use of Gaia DR2 astrometry (Gaia Collaboration et al. 2016, 2018; Salgado et al. 2017; Arenou et al. 2018; Lindegren et al. 2018; Luri et al. 2018; Marrese et al. 2019). We de-redden the SDSS *ugriz* magnitudes using the `mwdust` utility (Bovy et al. 2016) and the combined dust maps of Drimmel et al. (2003), Marshall et al. (2006), and Green et al. (2019).

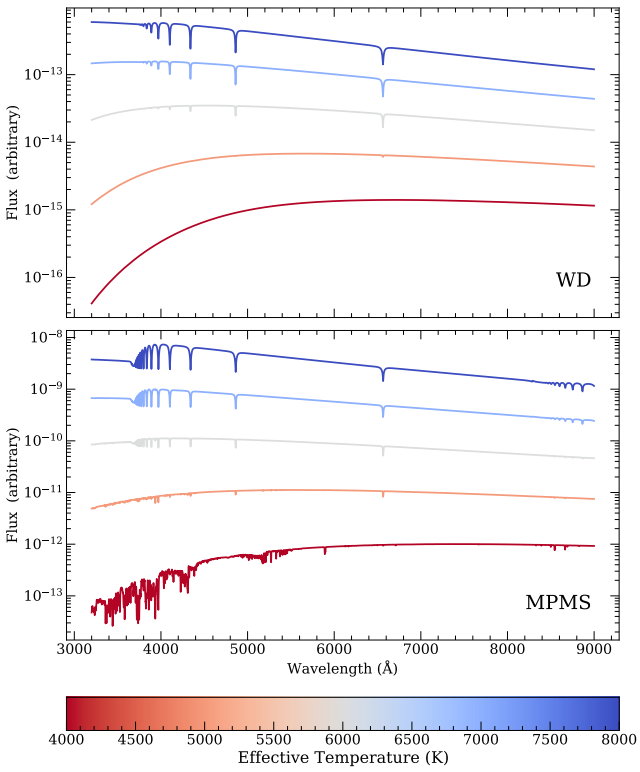

Figure 1. Sample theoretical white dwarf and MPMS star spectra from the Blouin et al. (2018a) and PHOENIX (Husser et al. 2013) libraries, respectively. We plot spectra spanning the range 4000 K $\leq T_{\rm eff} \leq$ 8000 K assuming $\log g = 8$ and $\log g = 4.5$ for the white dwarf and MPMS star spectra, respectively. For the MPMS star spectra, we assume [Fe/H] $= -4.0$ with scaled solar abundances. We offset the fluxes vertically for visual clarity. The optical spectrum of a very low-mass Population III star would be spectroscopically indistinguishable from the $T_{\rm eff} = 4000$ K white dwarf in the top panel.

We focus on the MPMS star and white dwarf classifications provided by Kepler et al. (2019). Those authors examined 500,000 SDSS spectra plausibly produced by white dwarfs. They provide classifications for 37,053 spectra based on 11 criteria, of which 15,716 were classified as DA white dwarfs and 15,855 were classified as subdwarf A or sdA stars (i.e., likely MPMS stars). Among the white dwarfs, over 78% were classified as DA white dwarfs with hydrogen-rich atmospheres on the basis of broad Balmer absorption lines with no other strong

spectral features. Kepler et al. (2019) visually differentiated these DA white dwarfs from subdwarf A stars based on the fact that DA white dwarfs have broader Balmer lines at a given photometric color due to their high surface gravities and increased pressure broadening. Together these two classes comprise more than 85% of the classifications provided in Kepler et al. (2019).

We use in our analysis only those white dwarfs or MPMS stars classified by Kepler et al. (2019) as members of their "DA", "sdA", or "sdA/F" classes. This limits our sample to 14,522 spectra with prominent Balmer lines and no strong metal features. It also removes other white dwarf spectral types like helium-rich DB white dwarfs. While most of the stars classified as "sdA" or "sdA/F" stars are likely MPMS stars (e.g., Brown et al. 2017; Pelisoli et al. 2018b,a), a very small fraction could also be extremely low-mass (ELM) white dwarfs (e.g., Kosakowski et al. 2020).

As we argued above, visually separating MPMS stars from white dwarf is prone to error, especially at cool temperatures when MPMS star and white dwarf spectra are visually similar. For that reason, Kepler et al. (2019) appealed to Gaia DR2 parallaxes to distinguish MPMS stars and DA white dwarfs. At a given color, white dwarfs are far less luminous than MPMS stars due to their smaller radii. A high-quality parallax measurement therefore provides accurate classifications via absolute magnitudes. However, only about 15% of the 14,522 stars in this sample have reliable parallaxes (i.e., $\pi/\sigma_\pi > 10$). As a result, the MPMS star and white dwarf classification cannot be confirmed for most of this sample. Consequently, when validating our methods we only use those stars for which a definite classification can be made based on a high-quality Gaia DR2 parallax. We require (1) `parallax_over_error > 10`, (2) `visibility_periods_used > 8`, and (3) `astrometric_sigma5d_max < 1`. This results in an empirical validation sample of 1807 stars with reliable MPMS star/white dwarf classifications. Of these, 65% are classified as DA white dwarfs and the remaining 35% are classified as MPMS stars. We illustrate this empirical validation sample in the Sloan color–magnitude diagram depicted in Figure 2.

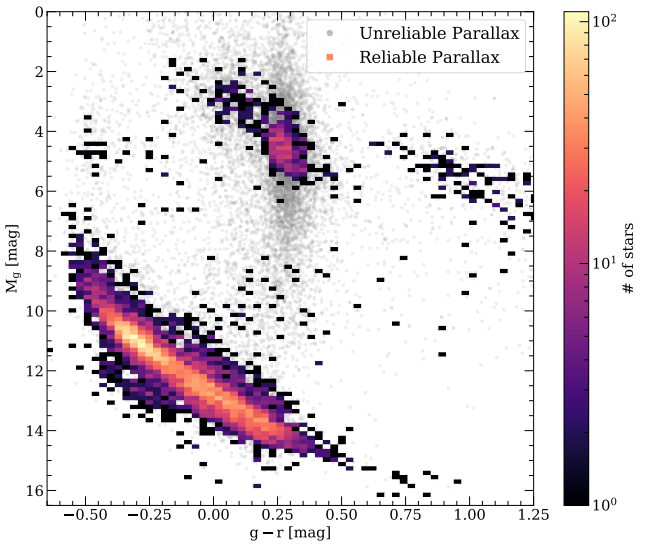

Figure 2. Color–magnitude diagram of the stars classified by Kepler et al. (2019) as white dwarfs or subdwarf A stars based on SDSS spectroscopy. Stars with unreliable parallaxes—75% of the sample—are not included in the 2D histogram and are plotted in gray. There is a clear separation between the white dwarf track on the bottom left and the more luminous stellar main sequence at the top. While it is trivial to discriminate between white dwarfs and MPMS stars when parallaxes are available, most of the Kepler et al. (2019) sample lacks reliable parallaxes in Gaia DR2.

3. METHODS

Even though visual classifications have been sufficient to separate MPMS stars from white dwarfs in the past, the volume of data expected to be produced by ongoing and next generation spectroscopic surveys will make the traditional approach impractical. We therefore use the grids of theoretical spectra described in Section 2.1 to build an automated process to differentiate MPMS stars from white dwarfs using spectroscopic and photometric observables alone without appealing to accurate parallax measurements. We then validate our classifier using the empirical labels in Section 2.2.

3.1. *Balmer Lines*

Both relatively warm MPMS stars and white dwarfs exhibit strong Balmer lines in their spectra. On this basis, in the absence of high-quality parallaxes MPMS stars have usually been differentiated from white dwarfs based on Balmer lines for temperatures $T_{\rm eff} \gtrsim 6000$ K. As we described above, MPMS stars have narrower Balmer lines than white dwarfs at a given temperature (Kepler et al. 2019). We now quantify this difference across several Balmer lines at once and use it to produce a rigorous selection function.

For each spectrum in the theoretical grids described in Section 2.1 we fit each of Balmer absorption lines Hα, Hβ, Hγ, and Hδ with a Voigt profile[1]. From each profile we derive two summary statistics: the full-width at half-maximum (FWHM) in angstroms and the min-

[1] A Voigt profile is a convolution of Gaussian and Lorentzian profiles that well-approximates the pressure-broadened wings of the Balmer lines (e.g., Tremblay & Bergeron 2009)

imum of the profile in continuum-normalized flux units that we define as the line amplitude. For each spectrum we therefore derive eight line statistics in total which together quantify the phase space of the Balmer lines. This results in a vector of eight Balmer features per spectrum plus a "label" identifying it as the spectrum a MPMS star or white dwarf. Our classifier then makes use of all eight features to differentiate MPMS stars from white dwarfs.

We illustrate in Figure 3 two 2D "slices" through the Balmer line phase space that demonstrate why Balmer line features are useful for classification. Because of the faintness of cooler white dwarfs, most white dwarfs targeted by the SDSS have $T_{\rm eff} \gtrsim 6000$ K. As a consequence of the low space density of massive stars in high Galactic latitude fields, most MPMS stars targeted by the SDSS have temperatures $T_{\rm eff} \lesssim 8000$ K. In the overlap region between these two temperature ranges, the theoretical spectra confirm the expectation that MPMS stars have much narrower (i.e., smaller FWHM) Hα profiles than white dwarfs.

Figure 3 is representative of the relationship between Balmer line features as a function of $T_{\rm eff}$ and $\log g$. Above $T_{\rm eff} \approx 6000$ K, MPMS star and white dwarf spectra are visually differentiable on many 2D slices through Balmer line phase space. Below $T_{\rm eff} \approx 6000$ K, the higher-order Balmer lines in white dwarf spectra start to rapidly lose their intensities as collisions fail to populate the $n = 2$ energy levels of hydrogen atoms. In the range $5000 \text{ K} \lesssim T_{\rm eff} \lesssim 6000$ K, combining statistics from several Balmer lines can still separate MPMS stars from white dwarfs. Below $T_{\rm eff} \approx 5000$ K, the Balmer lines for both metal-free low-mass stars and white dwarfs begin to disappear altogether, resulting in pure-continuum spectra (Blouin et al. 2019). In other words, below $T_{\rm eff} \approx 5000$ K there is no way to use optical spectra to distinguish metal-free low-mass stars from white dwarfs. We focus on the special case of metal-free low-mass stars further in Section 4.

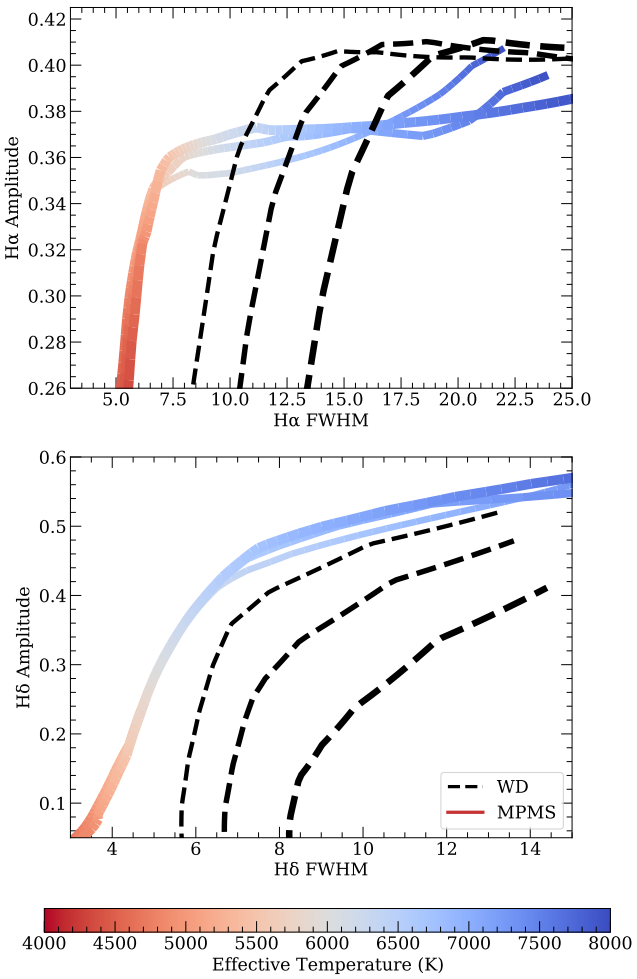

Figure 3. Two "slices" of the Balmer line phase space for MPMS stars (solid lines colored by effective temperature) and white dwarfs (dashed lines). The thickness of each line is proportional to surface gravity $\log g = 4$, 4.5, 5 for MPMS star models and $\log g = 7.5$, 8, 8.5 for white dwarf models.

3.2. *Synthetic Photometry*

In addition to the Balmer features we extract from each theoretical spectrum, we also calculate synthetic photometry in several photometric systems using the `pyphot` utility (Fouesneau 2020). We calculate synthetic absolute magnitudes in the SDSS *ugriz*, Pan-STARRS *grizy* (Chambers et al. 2016), SkyMapper *uvgriz* (Bessell et al. 2011), DECam *ugrizY* (Flaugher et al. 2015), and Vera Rubin Observatory *ugrizy* (Ivezic et al. 2019) systems. While the software accompanying this paper can be used with any of these photometric systems, we focus on SDSS photometry colors in this paper since we validate our method on SDSS data.

We plot in Figure 4 a comparison between MPMS stars and white dwarfs in Sloan color–color space. Sloan $u - g$ color is strongly affected by the Balmer jump, a sensitive probe of surface gravity. A $u - g$ versus $g - r$ color–color plot can therefore be used to cleanly differentiate relatively warm MPMS stars and white dwarfs. At lower temperatures the difference becomes less pronounced, so this color–color diagram can only differentiate MPMS stars and white dwarfs when $T_{\rm eff} \gtrsim 6000$. While Figure 4 depicts one well known photometric difference between MPMS stars and white dwarfs, we consider the complete color–color spaces defined by each of the SDSS, Pan-STARRS, SkyMapper, DECam, and Rubin Observatory photometric systems, computing all

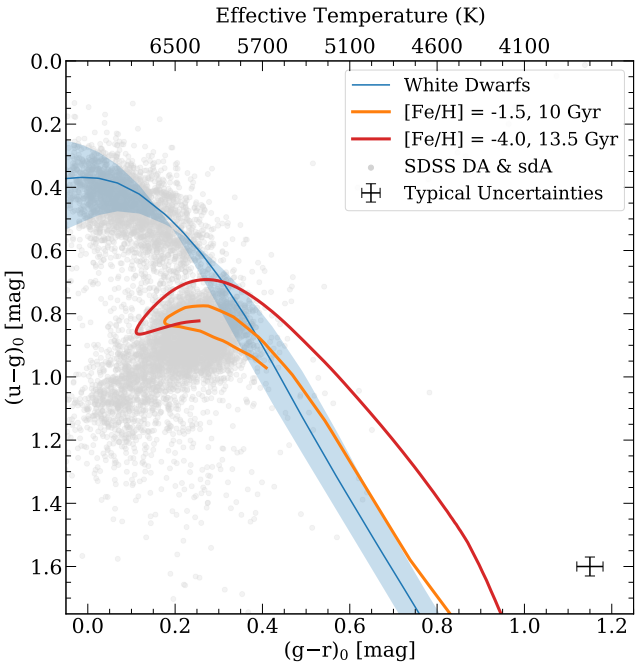

Figure 4. Synthetic color–color diagram using SDSS bands. The blue line corresponds to white dwarfs with $\log g = 8$ while the shaded region indicates the locus of white dwarfs with surface gravities in the range $7 < \log g < 9$. We also plot MIST isochrones for MPMS stars with [Fe/H] $= -1.5$ and age 10 Gyr (orange) and [Fe/H] $= -4.0$ and age 13.5 Gyr (red). We overlay as gray points all stars classified as "DA", "sdA", or "sdA/F" objects by Kepler et al. (2019).

unique pairwise differences of the magnitudes in each photometric band.

3.3. *Logistic Regression Classifier*

To map the spectroscopic and photometric features described above to a MPMS star or white dwarf classification, 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 (i.e., Bernoulli) random variable taking on the value 1 (or "true"). Given input features $\mathbf{x} = [x_0, x_1, x_2, \ldots]$, fitting a logistic regression involves solving for the coefficients $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \ldots]$ such that

$$\boldsymbol{\beta}^\intercal \mathbf{x} = \log\left(\frac{p}{1-p}\right), \tag{1}$$

where p is the probability that the Bernoulli random variable is 1 (or "true"). A logistic regression model is typical fit to data with known labels (i.e., where p is known to be either zero or one), and the coefficients $\boldsymbol{\beta}$ are subsequently used to estimate $\hat{p}$ for new input data. There is no closed-form solution to determine $\boldsymbol{\beta}$, so an iterative gradient descent algorithm is often used to find the optimal coefficients.

We define the underlying Bernoulli variable to have value 0 if the an object is a confirmed white dwarf and 1 if an object is a confirmed 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 object is a MPMS star as opposed to a white dwarf. We demonstrate three possible input configurations: the Balmer line summary statistics (eight features), *ugriz* photometric colors (ten features), and *griz* photometric colors (six features). We also consider a combined classifier that uses *ugriz* colors and Balmer features simultaneously.

We use a logistic regression model due to its simplicity and ease of interpretation. The classification probability $P_{\rm MPMS}$ returned by a logistic regression is well-calibrated by default (Yu et al. 2011), providing confidence in the classification. We found that using a more complex classification algorithm like a random forest or support vector machine increased the complexity of the method without much increase in accuracy. In addition, these other algorithms require an external "calibration function" to transform the returned classification probabilities to statistically meaningful values (Niculescu-Mizil & Caruana 2005). This step is unnecessary for the logistic regression.

When training our logistic regression models, we reserve 5% of the stars in our grid of theoretical models as a synthetic validation set. We evaluate our logistic regression model on these unseen synthetic validation data and confirm that the model correctly predicts their classifications virtually 100% of the time. This validation step ensures that the logistic regression is working as expected on noiseless synthetic data. We will further validate our logistic regression model with SDSS photometry and spectroscopy for objects with secure parallax-based classifications in Section 3.4. Our classifier and synthetic training data for the SDSS, Pan-STARRS, SkyMapper, DECam, and Rubin Observatory photometric systems are publicly available.[2]

3.4. *Model Validation with Objects that have Secure Empirical Classifications*

We now further validate our classifier with the empirical validation sample of 1807 objects described in Section 2.2. These objects were classified by Kepler et al. (2019) based on spectroscopic features with Gaia DR2-confirmed classifications as either MPMS stars or white dwarfs. We confirm that the Kepler et al. (2019) classifications for these objects are accurate based on our

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

inspection of their locations in a color–magnitude diagram. We use the available SDSS photometry and spectroscopy for this empirical validation sample to compute Balmer features and *ugriz* colors. We then run the logistic regression classifier described in Section 3.3 on these features to predict based solely on SDSS photometry and spectroscopy the probability that each object is a MPMS star.

Receiver operating characteristic (ROC) curves are one way to graphically evaluate classifiers. ROC curves describe the relationship between the true-positive and false-positive rates of a classifier as a function of its acceptance threshold (i.e., the output probability above which a "positive" classification is made). ROC curves provide a high-level summary of the sensitivity (i.e., probability of detection) and specificity (i.e., probability of false alarm). Precision-recall (PR) curves provide an alternative graphical classifier diagnostic. They are particularly useful for our application, as precision (i.e., the probability that a MPMS star is accurately classified by our classifier) is the most important metric for our scientific goal. PR curves are also more agnostic to the possible existence of class imbalance in a dataset, as is the case in our empirical validation sample in which there are twice as many white dwarfs as MPMS stars. Together, ROC and PR curves provide an overview of the performance of a classifier.

We illustrate in Figure 5 ROC and PR curves for the application of our classifiers to our empirical validation data. We evaluate our logistic regression classifier in several scenarios: only the eight Balmer-line-summary-based features, only the ten *ugriz* colors, only the six *griz* colors, and both the Balmer-line-summary-based features and *ugriz* colors. As expected, the Balmer spectroscopic features provide the most discriminating power. The inclusion of $u-g$ color greatly improves the quality of classifications compared to *griz* colors alone, as $u-g$ color is sensitive to $\log g$ via the Balmer break. In terms of precision, the classifier using both photometric and spectroscopic features far outperforms the spectroscopy-only classifier. This implies that a degeneracy in spectroscopic features in MPMS star and white dwarf spectra can be overcome by including photometric features. 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. We conclude that our combined photometric and spectroscopic classifier has the accuracy and precision necessary to separate MPMS stars from white dwarfs using SDSS data or equivalent LAMOST or DESI data.

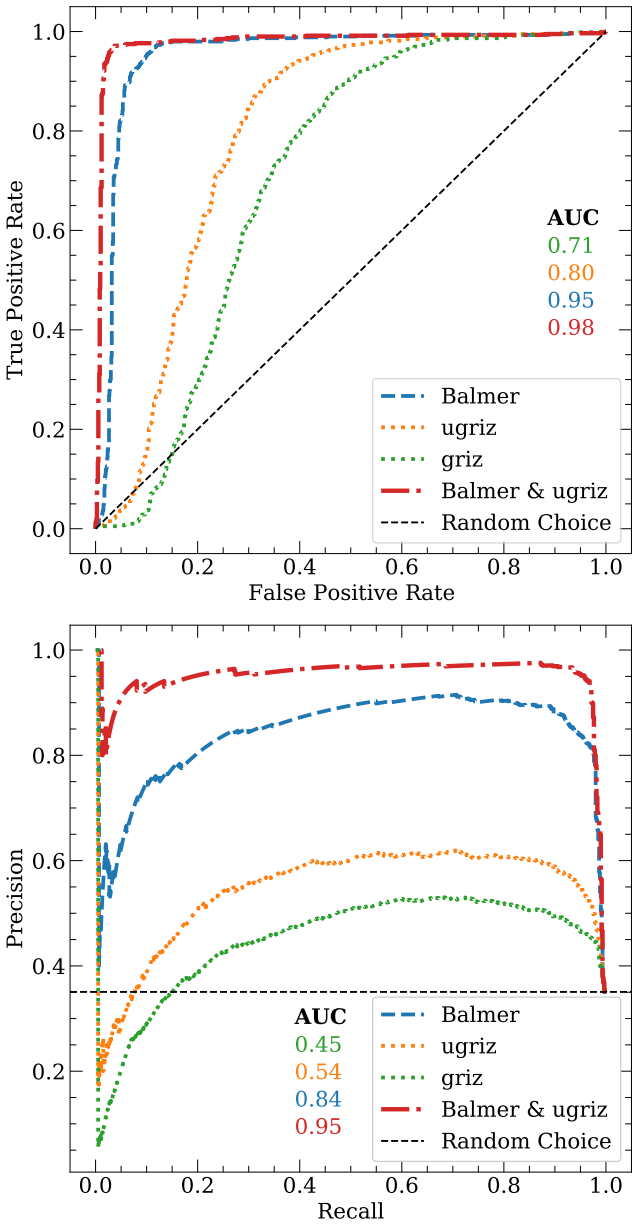

Figure 5. Receiver operating characteristic (ROC) and precision-recall (PR) curves for the photometry-only, spectroscopy-only, and combined logistic regression classifiers trained on theoretical models and validated with classifications confirmed by high-quality Gaia DR2 parallaxes. The classifier that uses both spectroscopic and photometric features achieves the best performance, followed closely by the classifier that uses only spectroscopic features. Classifiers that use only photometric features achieve acceptable performance but lag behind classifiers that use spectroscopic features.

3.5. *Equivalent Widths of Metal Lines*

After differentiating MPMS stars from white dwarfs, another important task is to select the most metal-poor candidates among the predicted MPMS stars. In order to select the most metal-poor MPMS star candidates, we implement a simple method to approximately quantify stellar metallicity from MPMS star spectra using equivalent widths.

Equivalent widths are ubiquitous measurements of the intensities of absorption lines in stellar spectra that are relatively independent of line shape and instrumental resolution. While equivalent widths are usually calculated from high signal-to-noise ratio (S/N), high-resolution spectra, the SDSS spectra that we have on hand have much lower resolution and only intermediate S/N. We therefore fit noiseless line profiles to the spectroscopic features and then compute the equivalent width of the fitted profile. In particular, we fit pseudo-Voigt profiles (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 of a fitted profile computed on the original wavelength grid using numerical quadrature integration. We follow this equivalent width procedure for four metal lines that are present even in moderate-resolution MPMS star spectra: the Ca II K line (3934.8 Å) and the near-infrared Ca II triplet (8500.4 Å, 8544.4 Å, and 8664.5 Å).

4. RESULTS

With the accuracy and precision of our classifier now established, we can now classify the remaining 12,000 MPMS star/white dwarf candidates from Kepler et al. (2019) with SDSS photometry and spectroscopy but without high-quality Gaia DR2 parallaxes. We fit the Balmer lines of these spectra to derive the line summaries described in Section 3.1. We then apply our combined logistic regression classifier to the Balmer features and *ugriz* colors of these stars. The result of this process is a prediction of the probability $P_{\rm MPMS}$ that a spectrum was produced by a MPMS star as opposed to a white dwarf. Reassuringly, the distribution of probabilities is bimodal: most stars have probabilities close to zero or one. This distribution provides even more confidence that our classifier is both precise and accurate. We decide on decision threshold of 0.5 and classify all objects with $P_{\rm MPMS} > 0.5$ as MPMS star candidates.

We find that approximately 95% of our logistic regression-based classifications agree with the visual inspection-based classifications reported in (Kepler et al. 2019). The 5% of cases where our classifier disagrees with Kepler et al. (2019) are due to objects with extremely low S/N spectra. Spectroscopic surveys like DESI will produce hundreds of thousands of spectra plausibly produced by MPMS stars or white dwarfs, and only our automated classification approach will be able to scale to that data volume.

We emulate future searches for candidate extremely metal-poor (EMP) stars in upcoming surveys like SDSS-V and DESI by mining the stars in our SDSS dataset for EMP candidates. Numerous studies have already searched these SDSS data for metal-poor stars (e.g., Allende Prieto et al. 2006, 2008, 2014, 2015; Caffau et al. 2011a,b; Bonifacio et al. 2012, 2015; Aoki et al. 2013; Frebel et al. 2015; Placco et al. 2015; Aguado et al. 2016, 2017a,b, 2018a,b; Yoon et al. 2016; Carbon et al. 2017), so our expectation is that many of the most promising candidates have already been discovered. We select all stars from the 12,000 star subsample of the Kepler et al. (2019) study referenced above with $P_{\rm MPMS} > 0.5$ according to our combined classifier as possible MPMS candidates. We also focus on objects with $g - r > 0.25$ to minimize the occurrence of false positives. Below this color, the calcium absorption lines we measure become subdued due to higher temperatures rather than low metallicity. This leaves us with 9,000 MPMS star spectra for which we compute equivalent widths as described in Section 3.5. We perform a blind search for potentially EMP star candidates on the basis of the equivalent widths of the calcium absorption lines. We selected stars with the lowest equivalent widths within a range of colors and visually inspected their spectra, examining the Ca II K (3934.8 Å), Na I (5895.6 Å), and Mg I (5176.7 Å) lines in particular. We report our most promising candidates in Table 1.

In addition, we use SIMBAD to search for literature references to our candidate MPMS stars (Wenger et al. 2000). While several of our MPMS star candidates have no prior reference in the literature, we find that our selection turns up dozens of previously-known metal-poor stars with confirmed metallicities [Fe/H] < -3.0. We therefore conclude that SDSS data have already been exhaustively searched for extremely metal-poor stars. Nevertheless, this serves as a comprehensive validation of our methodology to search for metal-poor stars in future spectroscopic surveys.

While the existence of metal-free low-mass stars is still debated, we assert that it will be possible to uniquely classify a featureless optical spectrum plausibly produced be either a metal-free low-mass star or a cool white dwarf based on the optical-to-infrared colors of the object. 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 on the ef-

Table 1. Extremely Metal-poor Candidates Identified in This Study

Gaia DR2 Source ID	R.A. (deg)	Decl. (deg)	g_0 (mag)	$(g-r)_0$ (mag)	EW[Ca II K] (Å)	Reference
923068053360267136	119.6018	43.61185	16.4	0.3	0.88	
2572124320071981696	26.58228	8.55885	18.0	0.29	0.66	
4427787686456496128	236.01893	4.71759	17.0	0.27	0.94	
2434705943788272128	356.69676	-10.31683	17.8	0.37	0.76	
1041117669033018112	133.67586	60.40062	17.0	0.22	0.88	
926917477929764608	114.89642	44.57836	16.6	0.25	0.87	
881205690025926400	120.85979	32.86292	17.4	0.27	0.97	
1543322422520153472	191.67915	48.74803	16.9	0.27	0.97	
2823062869583476224	356.01949	19.72158	17.8	0.22	0.96	
4448895182975654784	252.11282	12.5559	17.8	0.26	0.97	
1214202544663262848	230.83424	21.46138	17.8	0.28	0.95	
2757064989066940928	354.90132	6.74032	17.5	0.21	0.84	
2540804903154179840	7.4275	-2.62331	16.8	0.23	0.97	
2294041061855802496	280.00989	78.60719	17.4	0.18	0.97	
3092216989574463616	126.33875	4.05948	17.1	0.24	0.51	Aoki et al. (2013)
290930261314166528	25.15092	23.74941	15.8	0.33	0.98	Aoki et al. (2013)
2779958921396953088	6.95615	14.07167	16.9	0.29	0.71	Aoki et al. (2013)
1276882477044162688	230.50864	30.92395	16.6	0.31	0.92	Aoki et al. (2013)
4190837398756490112	301.30624	-10.75094	17.0	0.18	0.42	Aoki et al. (2013)
2782111559005651584	10.12171	16.07109	15.5	0.27	0.82	Bonifacio et al. (2012)
1184737183522291712	221.66916	12.8215	16.2	0.27	0.87	Bonifacio et al. (2012)
3740963179636227968	207.34546	14.12687	16.6	0.33	0.66	Bonifacio et al. (2012)
3976087728282022272	175.84753	20.34943	16.9	0.31	0.95	Yoon et al. (2016)
3890626773968983296	157.31307	17.49107	16.9	0.41	0.43	Caffau et al. (2011a)

NOTE—We indicate SIMBAD references for stars with confirmed [Fe/H] < -3.0 inferences.

fective temperature of metal-free stars of $T_{\rm eff} \approx 3600$ K (Burrows et al. 1993; Saumon et al. 1994). At these temperatures, the optical spectra of the coolest primordial stars will be indistinguishable from the optical spectra of white dwarfs because 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 red optical and infrared (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 a MPMS star due to its higher surface gravity. White dwarfs will consequently be fainter than metal-free low-mass stars in the red optical izy and infrared bands like JHK. Therefore, at constant temperature photometric colors comparing optical to red optical or infrared bands (e.g., $r-z$, $V-J$, etc.) will be bluer for white dwarfs than metal-free low-mass stars. This effect is expected to be larger than the typical photometric uncertainties in ground-based surveys like the SDSS, Pan-STARRS, SkyMapper, or DES (Saumon et al. 1994; Blouin et al. 2017).

While future work on CIA opacities will enable a more quantitative comparison, current uncertainties in the theoretical modeling of CIA make it challenging to precisely quantify the expected difference between metal-free low-mass stars and white dwarfs, and thereby build a predictive model like the one we presented for warmer stars. Nevertheless, we predict that the photometric signature of CIA should be detectable via infrared photometry and consequently able to differentiate between metal-free low-mass stars and white dwarfs even without parallax information.

5. DISCUSSION

It remains an open question whether or not metal-free low-mass stars exist in the local Universe. Beyond awaiting serendipitous direct detections, Galac-

tic/stellar archaeology provide a promising technique to probe the hypothesis of surviving Population III stars and 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 Population III survivors at the 95% confidence level. However, this is likely an overestimate since it only considers blind surveys. In practice, searching for EMP and ultra metal-poor (UMP) stars is far more efficient. Magg et al. (2019) propose instead to use the occurrence of EMP and UMP stars to constrain the existence of Population III survivors. The largest uncertainty in this method by far is the total number of EMP and UMP stars in the Milky Way's stellar halo. As a result, a comprehensive search for such stars is an important need.

6. CONCLUSION

While it is trivial to separate metal-poor main sequence stars and white dwarfs when high-quality parallaxes are available, even post-Gaia many objects with DESI or Sloan spectroscopy will lack reliable parallaxes. For that reason, we developed a classifier capable of separating metal-poor main sequence stars from cool white dwarfs using a range of photometric and spectroscopic features. We trained and validated our classifier using theoretical spectra and synthetic photometry. We also validated the classifier using objects securely classified as metal-poor main sequence stars or white dwarfs based on Sloan spectroscopy and high-quality Gaia DR2 parallaxes. We then applied our classifier to a sample of candidate metal-poor main sequence stars and white dwarfs with visual classifications and confirmed that our automated approach reproduces the human classifications. We make our classifier and its underlying source code publicly available for reproducibility and application to future survey data. For the stars classified as metal-poor main sequence stars, we executed a search for extremely metal-poor stars. We uncovered 14 previously unidentified candidates and flagged dozens of already confirmed extremely metal-poor stars. We predict that the application of our classifier to future DESI and SDSS-V spectroscopy will discover thousands of metal-poor main sequence stars. We further predict that even in the absence of high-quality parallaxes, any candidate metal-free low-mass stars identified by their featureless optical spectra can be separated from cool white dwarfs due to their redder optical-to-infrared colors. These red colors are produced by relatively weak collision-induced absorption from molecular hydrogen in the relatively low-surface gravity photospheres of metal-free low-mass stars. The methods presented in this work will facilitate the rapid and reliable identification of MPMS stars, providing improved 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 supported by the Institute for Data Intensive Engineering & Science (IDIES) at Johns Hopkins University.

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). Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is www.sdss.org. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics — Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/. SDSS-

III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institut für Astronomie (MPIA Heidelberg), the Max-Planck-Institut für Astrophysik (MPA Garching), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington. This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

Facilities: Gaia, Sloan

Software: `numpy` (Harris et al. 2020), `scipy` (Virtanen et al. 2020), `matplotlib` (Hunter 2007), `astropy` (Robitaille et al. 2013), `scikit-learn` (Pedregosa et al. 2011), `lmfit` (Newville et al. 2014), `mwdust` (Bovy et al. 2016)

REFERENCES

Abel, T., Bryan, G. L., & Norman, M. L. 2000, ApJ, 540, 39, doi: 10.1086/309295

—. 2002, Science, 295, 93, doi: 10.1126/science.295.5552.93

Aguado, D. S., Allende Prieto, C., González Hernández, J. I., et al. 2016, A&A, 593, A10, doi: 10.1051/0004-6361/201628371

Aguado, D. S., Allende Prieto, C., González Hernández, J. I., & Rebolo, R. 2018a, ApJL, 854, L34, doi: 10.3847/2041-8213/aaadb8

Aguado, D. S., Allende Prieto, C., González Hernández, J. I., Rebolo, R., & Caffau, E. 2017a, A&A, 604, A9, doi: 10.1051/0004-6361/201731320

Aguado, D. S., González Hernández, J. I., Allende Prieto, C., & Rebolo, R. 2017b, A&A, 605, A40, doi: 10.1051/0004-6361/201730654

—. 2018b, ApJL, 852, L20, doi: 10.3847/2041-8213/aaa23a

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

Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, ApJ, 636, 804, doi: 10.1086/498131

Allende Prieto, C., Sivarani, T., Beers, T. C., et al. 2008, AJ, 136, 2070, doi: 10.1088/0004-6256/136/5/2070

Allende Prieto, C., Fernández-Alvar, E., Schlesinger, K. J., et al. 2014, A&A, 568, A7, doi: 10.1051/0004-6361/201424053

Allende Prieto, C., Fernández-Alvar, E., Aguado, D. S., et al. 2015, A&A, 579, A98, doi: 10.1051/0004-6361/201525904

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

Arenou, F., Luri, X., Babusiaux, C., et al. 2018, A&A, 616, A17, doi: 10.1051/0004-6361/201833234

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

Bessell, M., Bloxham, G., Schmidt, B., et al. 2011, PASP, 123, 789, doi: 10.1086/660849

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

Bonifacio, P., Caffau, E., Spite, M., et al. 2015, A&A, 579, A28, doi: 10.1051/0004-6361/201425266

Bovy, J., Rix, H.-w., Green, G. M., Schlafly, E. F., & Finkbeiner, D. P. 2016, The Astrophysical Journal, 818, 130, doi: 10.3847/0004-637x/818/2/130

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

—. 2002, ApJ, 564, 23, doi: 10.1086/323947

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. 2011a, Nature, 477, 67, doi: 10.1038/nature10377

—. 2011b, A&A, 534, A4, doi: 10.1051/0004-6361/201117530

Carbon, D. F., Henze, C., & Nelson, B. C. 2017, ApJS, 228, 19, doi: 10.3847/1538-4365/228/2/19

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

Chen, Y., Bressan, A., Girardi, L., et al. 2015, MNRAS, 452, 1068, doi: 10.1093/mnras/stv1281

Chen, Y., Girardi, L., Bressan, A., et al. 2014, MNRAS, 444, 2525, doi: 10.1093/mnras/stu1605

Clark, P. C., Glover, S. C. O., Klessen, R. S., & Bromm, V. 2011a, ApJ, 727, 110, doi: 10.1088/0004-637X/727/2/110

Clark, P. C., Glover, S. C. O., Smith, R. J., et al. 2011b, Science, 331, 1040, doi: 10.1126/science.1198027

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

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10, doi: 10.1088/0004-6256/145/1/10

Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44, doi: 10.3847/0004-6256/151/2/44

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

Doi, M., Tanaka, M., Fukugita, M., et al. 2010, AJ, 139, 1628, doi: 10.1088/0004-6256/139/4/1628

Dopcke, G., Glover, S. C. O., Clark, P. C., & Klessen, R. S. 2013, ApJ, 766, 103, doi: 10.1088/0004-637X/766/2/103

Drimmel, R., Cabrera-Lavers, A., & López-Corredoira, M. 2003, Astronomy and Astrophysics, 409, 205, doi: 10.1051/0004-6361:20031070

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72, doi: 10.1088/0004-6256/142/3/72

Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, Astronomical Journal, 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

Fraser, M., Casey, A. R., Gilmore, G., Heger, A., & Chan, C. 2017, MNRAS, 468, 418, doi: 10.1093/mnras/stx480

Frebel, A., Chiti, A., Ji, A. P., Jacobson, H. R., & Placco, V. M. 2015, ApJL, 810, L27, doi: 10.1088/2041-8205/810/2/L27

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

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748, doi: 10.1086/117915

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1, doi: 10.1051/0004-6361/201629272

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1, doi: 10.1051/0004-6361/201833051

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

Glover, S. 2013, Astrophysics and Space Science Library, Vol. 396, The First Stars (Springer), 103, doi: 10.1007/978-3-642-32362-1_3

Green, G. M., Schla, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 93, doi: 10.3847/1538-4357/ab5362

Greif, T. H. 2015, Computational Astrophysics and Cosmology, 2, 3, doi: 10.1186/s40668-014-0006-2

Greif, T. H., Bromm, V., Clark, P. C., et al. 2012, MNRAS, 424, 399, doi: 10.1111/j.1365-2966.2012.21212.x

Greif, T. H., Springel, V., White, S. D. M., et al. 2011, ApJ, 737, 75, doi: 10.1088/0004-637X/737/2/75

Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040, doi: 10.1086/300645

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332, doi: 10.1086/500975

Hansen, C. J., Koch, A., Mashonkina, L., et al. 2020, Astronomy & Astrophysics. https://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

Hirano, S., & Bromm, V. 2017, MNRAS, 470, 898, doi: 10.1093/mnras/stx1220

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

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. https://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

Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, Astronomy and Astrophysics, 616, doi: 10.1051/0004-6361/201832727

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

Marrese, P. M., Marinoni, S., Fabrizio, M., & Altavilla, G. 2019, A&A, 621, A144, doi: 10.1051/0004-6361/201834142

Marshall, D. J., Robin, A. C., Reylé, C., Schultheis, M., & Picaud, S. 2006, Astronomy & Astrophysics, 651, 635

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

Padmanabhan, N., Schlegel, D. J., Finkbeiner, D. P., et al. 2008, ApJ, 674, 1217, doi: 10.1086/524677

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., Lee, Y. S., et al. 2015, ApJ, 809, 136, doi: 10.1088/0004-637X/809/2/136

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

Riaz, R., Bovino, S., Vanaverbeke, S., & Schleicher, D. R. G. 2018, MNRAS, 479, 667, doi: 10.1093/mnras/sty1635

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

Salgado, J., González-Núñez, J., Gutiérrez-Sánchez, R., et al. 2017, Astronomy and Computing, 21, 22, doi: 10.1016/j.ascom.2017.08.002

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

Schlaufman, K. C., Thompson, I. B., & Casey, A. R. 2018, The Astrophysical Journal, 867, 98, doi: 10.3847/1538-4357/aadd97

Silk, J. 1983, MNRAS, 205, 705, doi: 10.1093/mnras/205.3.705

Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, AJ, 146, 32, doi: 10.1088/0004-6256/146/2/32

Smith, J. A., Tucker, D. L., Kent, S., et al. 2002, AJ, 123, 2121, doi: 10.1086/339311

Stacy, A., & Bromm, V. 2013, MNRAS, 433, 1094, doi: 10.1093/mnras/stt789

—. 2014, ApJ, 785, 73, doi: 10.1088/0004-637X/785/1/73

Stacy, A., Bromm, V., & Lee, A. T. 2016, MNRAS, 462, 1307, doi: 10.1093/mnras/stw1728

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

—. 2012, MNRAS, 422, 290, doi: 10.1111/j.1365-2966.2012.20605.x

Tegmark, M., Silk, J., Rees, M. J., et al. 1997, ApJ, 474, 1, doi: 10.1086/303434

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

Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377, doi: 10.1088/0004-6256/137/5/4377

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