Temperatures and Metallicities of M Dwarfs in the APOGEE Survey

Jessica Birký, David W. Hogg, Andrew W. Mann, and Adam Burgasser

1 Center for Astrophysics and Space Science, University of California San Diego, La Jolla, CA 92093, USA; jbirky@uw.edu
2 Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
3 Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA
4 Center for Data Science, New York University, 60 Fifth Ave, New York, NY 10011, USA
5 Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Ave, New York, NY 10010, USA
6 Department of Astronomy, Columbia University, 550 West 120th Street, New York, NY 10027, USA
7 Department of Physics and Astronomy, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Received 2019 August 23; revised 2020 January 13; accepted 2020 January 15; published 2020 March 23

Abstract

M dwarfs have enormous potential for our understanding of structure and formation on both Galactic and exoplanetary scales through their properties and compositions. However, current atmosphere models have limited ability to reproduce spectral features in stars at the coolest temperatures ($T_{\text{eff}} < 4200$ K) and to fully exploit the information content of current and upcoming large-scale spectroscopic surveys. Here we present a catalog of spectroscopic temperatures, metallicities, and spectral types for 5875 M dwarfs in the Apache Point Observatory Galactic Evolution Experiment (APOGEE) and Gaia-DR2 surveys using The Cannon: a flexible, data-driven spectral-modeling and parameter-inference framework demonstrated to estimate stellar-parameter labels ($T_{\text{eff}}$, log g, [Fe/H], and detailed abundances) to high precision. Using a training sample of 87 M dwarfs with optically derived labels spanning $2860 < T_{\text{eff}} < 4130$ K calibrated with bolometric temperatures, and $-0.5 < [\text{Fe/H}] < 0.5$ dex calibrated with FGK binary metallicities, we train a two-parameter model with predictive accuracy (in cross-validation) to 77 K and 0.09 dex respectively. We also train a one-dimensional spectral classification model using 51 M dwarfs with Sloan Digital Sky Survey optical spectral types ranging from M0 to M6, to predictive accuracy of 0.7 types. We find Cannon temperatures to be in agreement to within 60 K compared to a subsample of 1702 sources to color-derived temperatures, and Cannon metallicities to be in agreement to within 0.08 dex metallicity compared to a subsample of 15 FGK+M or M+M binaries. Finally, our comparison between Cannon and APOGEE pipeline (ASPCAP DR14) labels finds that ASPCAP is systematically biased toward reporting higher temperatures and lower metallicities for M dwarfs.

Unified Astronomy Thesaurus concepts: M dwarf stars (982); Stellar abundances (1577); Surveys (1671); Astronomy data analysis (1858); High resolution spectroscopy (2096)

Supporting material: machine-readable tables

1. Introduction

Low-mass stars, with masses $M_\star < 0.7 M_\odot$ and effective temperatures $T_{\text{eff}} < 4000$ K, are by far the most ubiquitous type of star, comprising ~70% of the Galaxy’s population by number (Bochanski et al. 2010). With nuclear fusion timescales $\tau > 10^{11}$ yr (Laughlin et al. 1997), the chemical compositions of the M-dwarf population trace the nucleosynthetic processes and interstellar mixing of heavy elements from many generations of shorter-lived, high-mass stars, and are a unique probe for piecing together Galactic structure and evolution (Bochanski et al. 2010; Woolf & West 2012).

Additionally, the low masses of M dwarfs make for easier detection of planets by variability in radial velocity (Trifonov et al. 2018), high ratios of planet-to-star radii make for easier detection of exoplanet transits in observations of light curves (Nutzman & Charbonneau 2008), and shorter orbital periods (for a fixed stellar insolation flux) allow for discovery of new planets in less observation time than for more massive stars. For these reasons, M dwarfs are primary candidates for exoplanet searches, including by the NASA Kepler (e.g., Dressing & Charbonneau 2015) and Transiting Exoplanet Survey Satellite (e.g., Muirhead et al. 2018) missions. As a result, detailed and precise knowledge of M-dwarf chemical compositions has become key to constraining the properties, formation scenarios, and atmospheric conditions of potentially habitable exoplanets observable with the James Webb Space Telescope (Clampin 2008).

Advances in instrumentation and the implementation of several spectroscopic surveys in the past decade, such as the Sloan Digital Sky Survey (SDSS; Eisenstein et al. 2011; Blanton et al. 2017) and the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Zhao et al. 2012), have dramatically increased the sample of known M dwarfs (West et al. 2011; Guo et al. 2015) with spectroscopic catalogs of over 70,000 sources, enabling studies of M-dwarf abundances on a Galactic scale. The Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2015) survey, as part of the SDSS III/IV mission, has introduced the largest sample of M dwarfs observed with high-resolution spectroscopy (Deshpande et al. 2013). APOGEE pipeline measurements of $T_{\text{eff}}$ and [Fe/H] (García Pérez et al. 2016) for M dwarfs have been determined to precisions of 100 K and 0.18 dex down to $T_{\text{eff}} \sim 3550$ K using atmosphere models (Schmidt et al. 2016).
Elemental abundance measurements from high-resolution spectra of F, G, and K stars have achieved extremely high precision (down to 0.01–0.03 dex; Nissen & Gustafsson 2018) enabled by improvements in atmosphere models including realistic assumptions of 3D local thermodynamic equilibrium (Asplund 2005), and differential abundance techniques using equivalent widths (Bedell et al. 2014). However, the determination of precise metallicities for M dwarfs has remained a long-standing challenge due to the formation of diatomic and triatomic molecules at M-dwarf temperatures, with absorption from TiO and VO in the optical, H₂O and CO in the infrared, and hydrides (FeH, CaH, CrH, MgH, etc.) present in the spectra of the latest spectral types (Allard et al. 1997). Atmospheric models often fail to reproduce these spectral features (e.g., Mann et al. 2013b) because of incomplete line lists and opacities. The presence of millions of weak, blended transitions, and the absence of a clear continuum, contribute to making it difficult to deconvolve individual features and extract line strengths from equivalent widths. The combination of these effects limits our ability to explore the information content of high-resolution spectra using traditional methods.

A number of studies focused on improving predictions of M-dwarf metallicity have used systems of M dwarfs in common proper motion with an FGK star and strong, isolated lines in the spectra of the M dwarf (e.g., Rojas-Ayala et al. 2010; Terrien et al. 2012; Neves et al. 2014; Newton et al. 2014; Lindgren et al. 2016) to develop precise empirical relations (as good as ∼0.07 dex). However these metallicity calibrations do not take advantage of the full wavelength coverage available, nor information about the overall spectral shape often used to determine T eff and spectral type. Furthermore, earlier calibrations are generally based on moderate-resolution data (with some exceptions: Neves et al. 2014; Lindgren et al. 2016) that fail to utilize the greater spectral information provided by APOGEE’s resolution.

In this work we build a data-driven model for M-dwarf APOGEE spectra with The Cannon (Ness et al. 2015; Casey et al. 2016; Ho et al. 2017a; Behmard et al. 2019)—a fully empirical model that employs no line lists or radiative transfer models. The Cannon is a generative model that parameterizes the flux at each pixel of a spectrum in terms of a set of stellar labels (a flexible number of parameters chosen by the user; described in more detail in Section 3). The model in this sense is used to transfer labels from spectra for which we know parameters to those for which we do not. This data-driven approach effectively circumvents the challenges of physically modeling the atmosphere of a star (and common issues associated such as incomplete line lists or opacities), provided that we have a subset of spectra in the data set with known (and very accurately measured) reference labels possibly measured from other data.

The data-driven approach of The Cannon is ideal in certain cases: if stellar labels are known for a small number of stars but there are spectra taken for many more; if it is computationally expensive to obtain labels for a star, and there are many stars that need labels; or if there are spectral models or techniques that work in one wavelength range or resolution but not in another. Existing methods to model M-dwarf spectra in the near infrared at high resolution are computationally expensive, and often calibrated over a narrow range of T eff and/or metallicity. The Cannon thus fills this niche: it does not require the use of specific lines or opacity information that may be missing from the models; instead it allows us to determine labels from a lot of low-level metallicity information present in thousands of lines, and as we demonstrate, it does so with very good precision.

Here we take M-dwarf labels from samples of well-characterized stars that are present in the SDSS-IV APOGEE sample, and use those labels to train a model and label all of the M dwarfs observed by SDSS-IV APOGEE. One set of labels are physical parameters (effective temperatures and metallicities), the other set are spectral types.

This paper is organized as follows: in Section 2 we describe the technical specifications of the data from the APOGEE and Gaia surveys, as well as previous studies of M dwarfs in APOGEE. Section 3 describes our model implementation using The Cannon framework, and Section 4 describes our sample selection and derivation of training parameters. In Section 5 we present our experimental results, evaluate the predictive accuracy of our models, apply our model to a selected test sample of nearly 6000 sources, and examine the validity of our parameters against color–temperature relations and metallicities of binary pairs. Finally, in Section 6 we discuss model performance, future improvements, and implications of our results.

2. Data

The APOGEE survey is a high-resolution (R ~ 22,500), H-band (1.5–1.7 μm), multi-epoch survey that has observed over 250,000 stellar spectra up to its fourteenth data release (DR14; Abolfathi et al. 2017). Fundamental parameters for each of these stars are estimated by the APOGEE Stellar Parameter and Chemical Abundances Pipeline (ASPCAP; García Pérez et al. 2016), which employs a χ² fitting procedure using the FERRE code to fit radiative transfer models and determine atmospheric parameters, 15 chemical abundances, and microturbulence parameters (Mészáros et al. 2012). The pipeline uses MARCS plane-parallel/spherical models (Gustafsson et al. 2008) for low temperatures (2800 K < T eff < 3500 K), and ATLAS9 plane-parallel models (Castelli & Kurucz 2003) for higher temperatures (T eff ≥ 3500 K).

APOGEE is primarily designed to target bright stellar populations, particularly red giants, with dereddened photometry and color cutoffs of 7 ≤ H ≤ 13.8 and [J − K]0 ≥ 0.5 (Zasowski et al. 2013), with the objective of studying Galactic composition and evolution. However, numerous cool, main-sequence sources have also been observed either as targets proposed by the APOGEE M-dwarf ancillary survey (~1200 sources; Deshpande et al. 2013) or serendipitously.

A number of studies out of the M-dwarf ancillary survey have already been conducted to measure reliable fundamental atmospheric parameters and make kinematic measurements using spectral synthesis of atmospheric model grids. These studies include Deshpande et al. (2013) and Gilhool et al. (2018), which have studied the radial and rotational kinematics for 700+ sources; Souto et al. (2017, 2018), which have modeled three exoplanet-hosting M dwarfs (Kepler-138, Kepler-168, and Ross-128), determining T eff/log g/metallicity and 13 elemental abundances; Rajpurohit et al. (2018), which tested BT-Settl (Allard et al. 2012) and MARCS (Gustafsson et al. 2008) model grids on 45 M dwarfs to estimate T eff/log g/metallicity; and Skinner et al. (2018), which identified and measured mass ratios and radial velocities for 44 M-dwarf spectroscopic binaries. This work complements...
existing studies by producing a model-independent catalog of spectroscopic temperatures and metallicities to test against model predictions for the entire APOGEE M-dwarf sample, which we quantify to contain at least 10,000 sources to date (DR14).

The ASPCAP pipeline releases several types of data files, with various levels of processing: ap1D (the raw one-dimensional spectra for individual visits), apVisit (the individual visit spectra with telluric subtraction), apStar (the co-added apVisit spectra), and aspcapStar, which contains the pseudo-continuum-normalized, rest-frame-shifted, co-added spectrum of all observed epochs (see García Pérez et al. 2016 for a complete description of the pipeline). We use the last data set for our study. In previous work it has been recommended to use an alternative pseudo-continuum normalization (Ness et al. 2015), but we did not find obvious issues with the normalization in our analysis, so we retain the survey pipeline outputs.

3. Method

The Cannon is a regression model that relies on two assumptions: first, that sources with identical labels have near-identical flux at each wavelength pixel; and second, that the expected flux at each pixel varies continuously with change in label.

Inferring the label of a star with such a model requires two steps: first, the training step, in which a generative model describing the probability density function of the flux is constructed at each pixel from the set of spectra with known reference labels; and second, the test step, in which the model is applied to determine the labels of a spectrum.

Following the procedure of Ness et al. (2015) and Ho et al. (2017a), we adopt a simple linear model that assumes that the flux at each pixel of the spectrum can be parameterized as a function of a label vector \( \ell \) and coefficient vector \( \theta \). For each star \( n \), at wavelength pixel \( \lambda \), we assume that the measured flux for a star at a given pixel is the sum of the coefficient and label product, and observational noise:

\[
f_n \lambda = \theta^T \cdot \ell_n + N_n.
\]

Here we use the noise model \( N_n = [s^2 + \sigma^2_n] \xi_n \), where the bracketed term is the root mean square of the intrinsic scatter of the model at each pixel \( s_n \) and the uncertainty due to instrumental effects \( \sigma_n \), which is then multiplied by a Gaussian random number \( \xi_n \sim N(0, 1) \). Equation (1) corresponds to the single-pixel log-likelihood function

\[
\ln p(f_n | \theta^T, \ell_n, s^2) = -\frac{1}{2} \left[ \frac{f_n - \theta^T \cdot \ell_n}{s^2} \right]^2 - \frac{1}{2} \ln(s^2 + \sigma^2_n),
\]

which gives the probability density function of the measured flux, given the labels, coefficients, and scatters.

We apply a quadratic parameterization of the model such that the label vectors for the two models are all combinations of reference labels up to second order:

\[
\ell_n = [1, \text{SPT}, \text{SPT}^2]
\]

\[
\ell_n = [1, \text{TEff}, \text{[Fe/H]}, \text{TEff}^2, \text{[Fe/H]}^2, \text{[Fe/H]}^2].
\]

Equation (3) is the label vector for the spectral type model, and Equation (4) is the label vector for the physical parameter model; the first element “1” is included to allow flexibility for a linear offset to the model. We find that a second-order parameterization is sufficient for reproducing the flux of each spectrum to 1% accuracy, as discussed further in Section 5.1.

The training step consists of optimizing the likelihood function (Equation (2)) for the coefficient vector and scatter \( (\theta, s) \) given the fixed label vector \( (\ell_n) \) constructed from the reference labels. The test step consists of optimizing the likelihood function for the labels at fixed \( \theta \) and \( s \), obtained in the training step (see Ness et al. 2015 for further description).

In the training step, the regression is designed to predict spectral pixels given labels, by learning zeroth, first, and second derivatives of the data with respect to the labels. In the test step, the regression is designed to predict labels given the spectral derivatives.

4. Sample Selection

The Cannon model can in principle be trained on any physical or empirical labels available beyond those that typically parameterize theoretical atmospheric models (TEff, log g, [M/Na], etc.), such as additional physical parameters (e.g., mass/age Ho et al. 2017b) or empirical proxies for physical parameters (e.g., spectral types, colors, magnitudes), giving a wide range of flexibility to the model. However, choosing a training sample with high-quality labels is critical to its performance. Limitations of The Cannon include that test (output) labels are only accurate if the training labels are accurate, and only precise if the training labels are measured consistently across the training sample. It is also critical to have a training sample with the dynamic range to span the entire parameter space of interest, because The Cannon does not extrapolate well outside the parameter space of the training sample. Finally, The Cannon assumes that the dependence of the spectrum on labels is continuous and smooth—and in this implementation is well approximated by quadratic functions. If that is not true, there will be features that The Cannon cannot reproduce.

For the purpose of this study, we have constructed two different training samples: first a one-dimensional spectral type model, and second, a two-dimensional physical parameter model, which describes the temperature and metallicity. The choice of training labels, the dimensionality of our data set, and requirements for a good training set are discussed further in Section 6.

4.1. Spectral Type Training Sample

The spectral type training sample consists of 51 sources, spanning M0–M9 cross-matched from the catalog of West et al. (2011, hereafter W11) that contains 78,841 M dwarfs from SDSS. For each source in the catalog, spectral types were determined both through an automated routine for comparing spectral type templates to data using The Hammer (Covey et al. 2007) and by visual inspection to a reported accuracy of \( \pm 1 \) type. A spectral sequence of spectra from the training sample spanning M0–M9 is shown in Figure 1.

4.2. Physical Parameter Training Sample

The physical parameter training sample consists of 87 sources with reference labels distributed over 2859 K < \( T_{\text{eff}} \) < 4131 K, and \(-0.48 < \text{[Fe/H]} < 0.49\) dex, 41 of which are drawn from Mann et al. (2015, hereafter M15), and 46 of which
Figure 1. Spectral sequence of dwarfs in training set M0–M9; separate plots show three detector chips of the APOGEE spectrum (black) and the best-fit Cannon trained model (red) with highlighted regions sensitive to spectral type identified in Deshpande et al. (2013).
are part of a previously unpublished extension sample to M15, analyzed using similar data and identical techniques to M15. The major difference in the extension sample is that its sources had lower-quality or no parallaxes (prior to Gaia data) and hence were omitted from the M15 study and were less vetted for binarity than the M15 sample (however, all sources in the training sample were visually inspected by color–magnitude position for binarity before addition).

The M15 catalog in total contains 183 sources and the extension sample another 500 stars. Both samples were primarily selected from the proper-motion-selected CONCHSHELL (Gaidos et al. 2013) M-dwarf catalog. All targets have low-resolution optical spectra from the SNIFS spectrograph (Lantz et al. 2004) and infrared spectra taken with the SpeX Spectrograph (Rayner et al. 2003), which have been combined to estimate largely empirical bolometric fluxes. Effective temperatures have been estimated by comparing the SNIFS spectra to BT-Settl atmospheric models (Allard et al. 2011). A subsample of 29 sources with measured angular diameters from long-baseline optical interferometry (Boyajian et al. 2012) are used to calibrate the model comparison, including masking of spectral regions poorly reproduced by the model spectra (Mann et al. 2013b). Based on the difference between assigned $T_{\text{eff}}$ values and those from angular diameters, absolute uncertainty on $T_{\text{eff}}$ is estimated to be 60 K, although the relative uncertainty is likely a factor of $\approx 2$ better.

Iron abundances ([Fe/H]) are assigned to the physical parameter sample based on the strength of metal-sensitive lines in the near-infrared SpeX spectra (Rojas-Ayala et al. 2010) using the calibration from Mann et al. (2013a). The relation between these lines and an absolute [Fe/H] scale is calibrated using wide binaries containing an F-, G-, or K-type primary and an M-dwarf companion, under the assumption that binaries formed from the same molecular cloud and therefore have the same metallicity (Bonfiils et al. 2005). Uncertainties are estimated to be $\approx 0.08$ dex based on irreducible scatter in the empirical relation between selected lines and the assigned [Fe/H] from the primary star. As with $T_{\text{eff}}$, relative errors on [Fe/H] are smaller, estimated to be 0.04–0.06 dex over most of the temperature and metallicity range considered here.

We note that surface gravity is not included as a training label. The reason for this is that for main-sequence M dwarfs, the parameter is almost entirely redundant with metallicity. The properties of M dwarfs, unlike those of their more massive counterparts, do not change measurably over the age of the universe after arriving at the zero-age main sequence. Hence perfect knowledge of abundances and the parameter is almost entirely redundant with metallicity. The label. The reason for this is that for main-sequence M dwarfs, the temperature and metallicity range considered here.

5. Experiments and Results

5.1. Temperature/Metallicity Model

For the physical parameter model, we trained The Cannon on 87 M dwarfs with two-dimensional temperature/metallicity labels, to a precision of 77 K/0.09 dex as estimated by the cross-validation scatter, similar to the uncertainties on the original training sample of 60 K/0.08 dex. We note for this model that five of out 87 sources show possible rotational line broadening identified by visual inspection (as indicated by the red circles in Figure 2), while the remaining sources show no obvious broadening. We note that these broadened sources have high $\chi^2$ values (those sources with $\chi^2 > 80,000$ in Figure 4), and that the labels for these five sources are biased by an average of +65 K and $-0.08$ dex. However, removing them from the training sample does not significantly change the overall scatter and bias of the model. For the model overall, the cross-validation bias is $+4$ K/+0.008 dex with the rapid rotators included in the training set, and $+5$ K/0.01 dex when they are excluded. Hence we do not remove them from the training sample.

To assess the validity of our model’s labels we used a leave-one-out cross-validation (LOOCV) test, in which we train a model on all sources but $n$, then apply the $N-1$ source-trained model to obtain the labels for star $n$. Precision (scatter) and bias of the model for each test are calculated as the standard deviation and mean of the difference in training and test (or LOOCV) labels respectively (Figure 2). Since the LOOCV test both evaluates how well the model reproduces the training values and penalizes the model for overfitting, we adopt the LOOCV scatter as the estimate of the model’s precision. The
Table 1  
Metallcities of APOGEE M Dwarfs in Wide Binaries

| M Dwarf Gaia ID | Companion Gaia ID | M Dwarf [M/H] | M Dwarf [Fe/H] | Companion [Fe/H] | Companion Source | Separation (au) |
|----------------|------------------|---------------|---------------|-----------------|-----------------|----------------|
| 2MASS ID | Temperature (K) | Metallicity (dex) | Model Fit χ² | Training | Test | LOOCV | Training | Test | LOOCV |
| 2MA00182565 | +4401222 | 4.59542 | 44.02278 | 3603 | 3538 | 3525 | −0.3 | −0.28 | −0.26 | 15304 |
| 2MA00182549 | +4401376 | 4.60779 | 44.02734 | 3218 | 3528 | 3560 | −0.3 | −0.29 | −0.29 | 21273 |
| 2MA00285391 | +5022330 | 7.22488 | 50.37588 | 3207 | 3190 | 3192 | 0.11 | 0.04 | 0.01 | 31790 |
| 2MA00401001 | +0308050 | 10.04169 | 3.13473 | 3725 | 3777 | 3772 | 0.04 | 0.12 | 0.14 | 9830 |
| 2MA00580115 | +3919111 | 14.50482 | 39.31977 | 3157 | 3100 | 3107 | −0.07 | −0.02 | 0.01 | 13337 |
| 2MA01235242 | +1638384 | 20.8559 | 16.64401 | 3272 | 3225 | 3223 | 0.1 | 0.02 | −0.01 | 34797 |
| 2MA02001278 | +1303112 | 30.05402 | 13.05196 | 3080 | 3059 | 3077 | −0.16 | −0.26 | −0.28 | 40739 |
| 2MA0261535 | +0652191 | 39.06358 | 6.87167 | 3264 | 3241 | 3243 | −0.12 | −0.28 | −0.28 | 48445 |
| 2MA03044335 | +6144097 | 46.18104 | 61.73583 | 3500 | 3466 | 3466 | −0.12 | −0.26 | −0.28 | 40739 |
| 2MA03553688 | +5212429 | 58.90373 | 52.2414 | 3435 | 3386 | 3375 | −0.35 | −0.28 | −0.2 | 53870 |
| 2MA04128800 | +5236421 | 63.24499 | 52.61165 | 3100 | 3183 | 3204 | −0.04 | −0.01 | 0.0 | 46007 |
| 2MA04310001 | +3647548 | 67.75003 | 36.79855 | 3419 | 3371 | 3369 | 0.08 | 0.11 | 0.1 | 20602 |
| 2MA05220205 | +3031097 | 80.58561 | 30.51933 | 3389 | 3423 | 3431 | 0.28 | 0.18 | 0.13 | 28029 |
| 2MA05312734 | −0340536 | 82.86417 | −3.67722 | 3801 | 3814 | 3849 | 0.49 | 0.5 | 0.14 | 17842 |
| 2MA05430737 | +5329239 | 85.37804 | 53.48987 | 3765 | 3751 | 3753 | 0.19 | 0.15 | 0.14 | 33868 |
| 2MA05428097 | +1229252 | 85.53833 | 12.48956 | 3250 | 3220 | 3229 | −0.22 | −0.33 | −0.29 | 33868 |
| 2MA06003051 | +0242236 | 90.01458 | 2.70657 | 3214 | 3170 | 3170 | 0.07 | 0.03 | 0.01 | 21044 |
| 2MA06011106 | +5955508 | 90.2961 | 59.59713 | 3340 | 3259 | 3255 | −0.09 | −0.05 | −0.04 | 23329 |
| 2MA06112610 | +1032599 | 92.8588 | 10.54998 | 3636 | 3720 | 3706 | −0.39 | −0.43 | −0.4 | 36898 |
| 2MA06554902 | +3316058 | 103.70411 | 33.26823 | 3448 | 3368 | 3363 | −0.02 | 0.03 | 0.04 | 19412 |
| 2MA07171706 | −0501031 | 109.32108 | −5.01754 | 3193 | 3175 | 3188 | −0.09 | −0.15 | −0.13 | 103000 |
| 2MA07224500 | +0531329 | 111.852 | 5.2263 | 3317 | 3279 | 3279 | −0.11 | −0.18 | −0.17 | 23615 |
| 2MA07444018 | +0333008 | 116.16744 | 3.55225 | 3217 | 3174 | 3167 | 0.23 | 0.25 | 0.17 | 34181 |
| 2MA08031949 | +5250387 | 120.83131 | 52.84402 | 3508 | 3617 | 3620 | −0.26 | −0.23 | −0.23 | 17491 |

Note. Reported uncertainties are ±60 K/0.08 dex for the M15 training sample labels, and ±77 K/0.09 dex for The Cannon model based on the cross-validation scatter.

This table is available in its entirety in machine-readable form.

Table 2  
Cannon Results for Temperature/Metallicity Model

Note. Reported uncertainties are ±60 K/0.08 dex for the M15 training sample labels, and ±77 K/0.09 dex for The Cannon model based on the cross-validation scatter.

set of training, test, and cross-validated labels for each training source is reported in Tables 1 and 2.

Another mode of analysis we can utilize with The Cannon is how the derivative of the model changes with respect to given training parameters, which makes our model interpretable for discovering or verifying atomic or molecular lines with strong dependence on different physical parameters. The top two panels of Figures 12 and 13 show two example spectra and...
Table 3

| Species | Wavelength | Derivative | $\sigma_{\text{jackknife}}$ | Significance |
|---------|------------|------------|-------------------|-------------|
| FeH     | 16107.127  | -0.002     | 0.001             | -2.78       |
| FeH     | 16245.463  | -0.005     | 0.002             | -3.214      |
| FeH     | 16271.519  | -0.002     | 0.001             | -2.013      |
| FeH     | 16284.562  | -0.006     | 0.001             | -5.807      |
| FeH     | 16377.066  | -0.005     | 0.001             | -4.554      |
| FeH     | 16741.489  | -0.002     | 0.001             | -2.873      |
| FeH     | 16812.415  | -0.004     | 0.001             | -4.241      |
| FeH     | 16813.808  | -0.006     | 0.001             | -3.954      |
| FeH     | 16922.405  | -0.003     | 0.001             | -3.867      |
| FeH     | 16934.801  | -0.004     | 0.001             | -3.591      |
| OH      | 15278.267  | 0.006      | 0.003             | 2.47        |
| OH      | 15280.8    | 0.004      | 0.002             | 2.007       |
| OH      | 15391.186  | 0.006      | 0.003             | 2.127       |
| OH      | 15407.355  | 0.01       | 0.002             | 4.836       |
| OH      | 15409.271  | 0.008      | 0.002             | 4.683       |
| OH      | 15505.797  | 0.009      | 0.003             | 3.511       |
| OH      | 15560.305  | 0.008      | 0.002             | 3.865       |
| OH      | 15565.895  | 0.005      | 0.002             | 2.505       |
| OH      | 15568.691  | 0.007      | 0.001             | 4.929       |
| OH      | 15572.133  | 0.008      | 0.002             | 4.393       |
| OH      | 16052.7    | 0.008      | 0.002             | 3.517       |
| OH      | 16065.362  | 0.006      | 0.002             | 3.7         |
| OH      | 16061.796  | 0.007      | 0.003             | 2.502       |
| OH      | 16065.124  | 0.009      | 0.002             | 4.214       |
| OH      | 16069.564  | 0.008      | 0.002             | 3.275       |
| OH      | 16074.227  | 0.007      | 0.002             | 3.338       |
| OH      | 16190.345  | 0.007      | 0.003             | 2.498       |
| OH      | 16192.134  | 0.008      | 0.002             | 4.012       |
| OH      | 16203.995  | 0.006      | 0.002             | 3.039       |
| OH      | 16207.129  | 0.004      | 0.002             | 2.196       |
| OH      | 16352.196  | 0.007      | 0.002             | 3.609       |
| OH      | 16354.682  | 0.011      | 0.003             | 4.167       |
| OH      | 16364.626  | 0.007      | 0.003             | 2.661       |
| OH      | 16368.244  | 0.008      | 0.002             | 4.303       |
| OH      | 16581.281  | 0.008      | 0.002             | 4.41        |
| OH      | 16581.74   | 0.004      | 0.001             | 4.842       |
| OH      | 16866.621  | 0.005      | 0.001             | 3.169       |
| OH      | 16871.982  | 0.012      | 0.002             | 4.892       |
| OH      | 16878.976  | 0.006      | 0.002             | 3.807       |
| OH      | 16884.573  | 0.005      | 0.002             | 3.066       |
| OH      | 16886.206  | 0.006      | 0.001             | 5.131       |
| OH      | 16895.074  | 0.005      | 0.002             | 2.197       |
| Fe I    | 15207.509  | -0.007     | 0.003             | -2.622      |
| Fe I    | 15648.478  | -0.006     | 0.002             | -3.273      |
| Fe I    | 15692.643  | -0.008     | 0.002             | -4.056      |
| Si I    | 15960.044  | -0.009     | 0.004             | -2.532      |
| Si I    | 16094.893  | -0.014     | 0.003             | -4.368      |
| Si I    | 16680.77   | -0.01      | 0.003             | -3.475      |
| K I     | 15162.824  | -0.006     | 0.002             | -2.294      |
| Ti I    | 15715.641  | 0.006      | 0.002             | 2.894       |
| Ti I    | 16634.973  | 0.003      | 0.001             | 4.224       |
| V I     | 15923.924  | 0.003      | 0.001             | 3.6         |
| Mn I    | 15159.054  | -0.006     | 0.002             | -3.123      |

Note. “Significance” quantifies the number of standard deviations that the derivative of the line is significant to (with negative corresponding to negative derivatives).

(This table is available in machine-readable form.)

The spectra contain roughly 8000 pixels, so we might expect the $\chi^2$ values to be close to 8000 in magnitude, but they are much higher. This discrepancy follows from the fact that, while the spectral model is good at the level of a few per cent, the signal-to-noise ratio of a typical spectrum is more than 100. That is, the $\chi^2$ values do show that the model is not good in the frequentist sense; it is only good at the level of a few per cent.

5.2. Spectral Type Model

We trained The Cannon on 51 M dwarfs in the range M0 $-$ M9 with a one-dimensional spectral type label, and obtained a precision of $\pm 0.9$ spectral types, similar to the uncertainty of the original training label of $\pm 1$ spectral type. We note, however, that the training sample is distributed heavily toward sources of earlier type, with a median spectral type of 3 and only one M8 and one M9 source. As seen in Figure 3, the model performs poorly at reproducing spectral types $>$M8,
which confirms that The Cannon does not extrapolate well to labels outside the training sample space. Because of this skew for late-type sources, we report our spectral type model to be precise to ±0.7 spectral types for the range M0−M6. Repeating the analysis of Section 5.1, Figure 3 shows LOOCV test for the labels reported in Table 5, and Figure 14 shows the derivative of model flux with varying spectral type.

### 5.3. Test Sample

Out of the total APOGEE DR14 catalog of 258,475 sources, we selected 254,478 sources that were in the cross-match of Gaia-DR2 (Gaia Collaboration et al. 2018) and applied Gaia color–magnitude cuts of $1 < BP − RP < 6$ and $7.5 < M_g < 20$ for sources with only positive parallaxes ($\pi > 0$), yielding a sample of 14,828 sources. From there we applied additional selection criteria, described below, to identify a sample of single, main-sequence M stars, with minimal contamination from reddened K dwarfs, pre-main-sequence stars, and binaries:

1. **Quality of fit cut**: We apply a Cannon model $\chi^2$ cut of less than 100,000, chosen to remove badly fit sources (such as fast rotators) but include $\chi^2$ values close to the distribution of the training sample (Figure 4).
2. **Color–magnitude cuts**: Using Gaia and photometry from the Two Micron All Sky Survey (2MASS) we apply the additional color–magnitude selections shown in Figure 5 to remove sources above the main sequence (which are likely pre-main-sequence, reddened K dwarfs and/or multiples), and subdwarfs below the main sequence.
3. **Model extrapolation cuts**: Because The Cannon does not perform well in extrapolated regions of parameter space, we select only sources inside the range of our training sample with ASPCAP parameters $2800 < T_{\text{eff}} < 4100$ K and $−0.5 < [\text{M/H}] < 0.5$, and with Cannon parameters of $2850 < T_{\text{eff}} < 4150$ K, $−0.5 < [\text{Fe/H}] < 0.5$, and $0 < \text{SPT} < 9$. 

| Designation | R.A. (deg) | Decl. (deg) | Training | Test | LOOCV |
|-------------|------------|-------------|----------|------|-------|
| 2MA0114974+0115158 | 47.957262 | 1.254404 | 1 | 0.9 | 1.4 |
| 2MA0122509+0021585 | 48.104563 | 0.366251 | 4 | 3.9 | 3.9 |
| 2MA0323963+0012102 | 55.66513 | 0.202859 | 5 | 5.0 | 5.1 |
| 2MA04262170+1800009 | 66.590421 | 18.000265 | 3 | 3.6 | 3.7 |
| 2MA09152918+4407461 | 138.871602 | 44.129498 | 3 | 2.4 | 2.4 |
| 2MA09183649+2207022 | 139.652051 | 22.117298 | 3 | 2.8 | 2.7 |
| 2MA09332262+2749021 | 143.344279 | 27.817253 | 6 | 6.5 | 6.6 |
| 2MA09373349+5534057 | 144.389577 | 55.568275 | 1 | 2.0 | 2.3 |
| 2MA10313413+3441535 | 157.892222 | 34.698212 | 8 | 7.5 | 6.5 |
| 2MA11194647+0820356 | 169.943658 | 8.343246 | 6 | 6.1 | 5.9 |
| 2MA11203609+0704135 | 170.1504 | 7.070432 | 4 | 3.4 | 3.1 |
| 2MA11570299+2028436 | 179.262465 | 20.4788 | 0 | 1.4 | 1.8 |
| 2MA12203634+2505351 | 185.15143 | 25.093107 | 2 | 2.4 | 2.5 |
| 2MA12212701-0030560 | 185.362567 | −0.515566 | 4 | 3.6 | 3.5 |
| 2MA12423245-0646077 | 190.635249 | −6.788287 | 3 | 3.1 | 3.0 |
| 2MA12464541-0312524 | 191.689212 | −3.214578 | 3 | 3.3 | 3.5 |
| 2MA12471099+1109566 | 191.795795 | 11.165737 | 4 | 4.0 | 4.0 |
| 2MA12492657-0312032 | 192.360749 | −3.200903 | 4 | 4.1 | 4.1 |
| 2MA12503440+4309482 | 192.643353 | 43.163414 | 3 | 3.1 | 3.1 |
| 2MA12532816+1240586 | 193.159002 | 12.682945 | 3 | 3.1 | 3.1 |
| 2MA12552414+4150425 | 193.839213 | 41.845161 | 3 | 3.1 | 3.1 |
| 2MA12564117+4233175 | 194.171558 | 42.554871 | 2 | 2.7 | 2.8 |
| 2MA13032161-4220407 | 195.840051 | 42.344654 | 4 | 3.9 | 3.9 |
| 2MA13415860-1852278 | 205.494169 | 18.874393 | 5 | 5.9 | 6.0 |
| 2MA13442970+5625445 | 206.123779 | 56.429039 | 5 | 5.9 | 6.0 |

The Astrophysical Journal, 892:31 (18pp), 2020 March 20

Birky et al.

**Note.** Reported uncertainties are ±1 spectral types for the W11 training sample labels, and ±0.7 spectral types for types ≤M6 for The Cannon model based on the cross-validation scatter.

(This table is available in its entirety in machine-readable form.)
4. Astrometric cut: Using the Gaia renormalized unit weight error (RUWE)—a metric for evaluating the fit of the astrometric solution described in the additional release notes (Lindegren 2018)—we apply a cut of RUWE < 1.2 to remove sources with high astrometric error or noise, such as binaries (see Figure 6).

5. Binary cut: To remove further contamination from binary sources, we applied an additional color–magnitude cut on sources above the main sequence, which we visibly selected for in Figure 6.

The top, middle, and bottom panels of Figure 7 show before and after selection of the sources in Gaia color–magnitude space, colored by temperatures, metallicities, and spectral types determined by The Cannon, with their respective training samples overplotted in orange. Each plot shows the expected gradient: temperature increases with decreasing color, spectral subtype increases with increasing color, and the metallicity gradient is largely perpendicular to the main-sequence branch. We also note that applying our model requires very little computational demand: the time to train and test a model on all 14,828 sources was two minutes on a 2.7 GHz Intel core i7 laptop.

Table 6 outlines the parameters included in the test sample catalog, which can be downloaded from the online journal. Included are two versions of the catalog: the first containing all 14,828 sources before selection, and the second containing the 5875 sources kept after making selections 1–5 described in this section.

5.4. Temperature Validation

As a validation test of our derived temperatures, we perform a comparison between several color–temperature relations from the literature, which use combinations of 2MASS and visual-band photometries to predict temperatures (similarly to the
To obtain visual-band magnitudes for a set of sources, we evaluated ASPCAP temperatures by Schmidt et al. (2016). To obtain visual-band magnitudes for a set of sources, we cross-matched our M-dwarf final sample with the catalog of >50,000 high-confidence, widely separated binaries identified by Gaia-DR2 presented in El-Badry & Rix (2018). In total we found 216 of the APOGEE M dwarfs to have binary pairs (46 FGK+M, 155 M+M, and 15 WD+M). Out of the 155 M+M pairs, eight contained both pairs in APOGEE. Cross-matching the list of FGK+M dwarf companions with several catalogs/surveys with measured stellar metallicities, we found an additional seven sources with FGK metallicities from LAMOST (Zhao et al. 2012) and APOGEE (ASPCAP). The metallicity measurements for the 15 M-dwarf binaries and their companions are given in Table 1 and shown in Figure 9, and the overall scatter is 0.08 dex—an improvement over the scatter of ASPCAP metallicities, which is 0.15 dex for these 15 sources. The internal consistency of the two models (the scatter of the eight M+M pairs both in APOGEE) is 0.06 dex for Cannon and 0.12 dex for ASPCAP.

As expected, the Toomre diagram in Figure 10 shows that higher-metallicity sources in the sample are concentrated in low-velocity space corresponding roughly to the thin-disk population; while the thick-disk population contains a slightly higher concentration of lower-metallicity sources. Separating the two populations into separate histograms (also shown in Figure 10), we find that thick-disk sources are marginally more metal-poor than thin-disk sources, with the mean ± standard deviation of [Fe/H] = 0.00 ± 0.17 dex for the thin-disk distribution, and [Fe/H] = −0.14 ± 0.19 dex for the thick-disk distribution. Metallicities of the two populations from ASPCAP show a similar distribution, with the mean ± standard deviation being [M/H] = −0.16 ± 0.16 dex for thin-disk sources and [M/H] = −0.23 ± 0.17 dex for thick-disk sources.

Figure 11 shows that ASPCAP metallicities are systematically lower than Cannon metallicities. We further find that the bias is temperature-dependent: at the highest temperatures ($T_{\text{eff}} > 3600$ K) ASPCAP and Cannon metallicities are consistent to a scatter of 0.05–0.06 dex and offset by an average of −0.12–0.15 dex, while at the lowest temperatures ($T_{\text{eff}} < 3200$ K) ASPCAP and Cannon are consistent to a scatter of −0.13 dex and offset by an average of −0.3 dex.

## 5.5 Metallicity Validation

As a check of the reliability of our test sample metallicity, we cross-matched our M-dwarf final sample with the catalog of >50,000 high-confidence, widely separated binaries identified by Gaia-DR2 presented in El-Badry & Rix (2018). In total we found 216 of the APOGEE M dwarfs to have binary pairs (46 FGK+M, 155 M+M, and 15 WD+M). Out of the 155 M+M pairs, eight contained both pairs in APOGEE. Cross-matching the list of FGK+M dwarf companions with several catalogs/surveys with measured stellar metallicities, we found an additional seven sources with FGK metallicities from LAMOST (Zhao et al. 2012) and APOGEE (ASPCAP). The metallicity measurements for the 15 M-dwarf binaries and their companions are given in Table 1 and shown in Figure 9, and the overall scatter is 0.08 dex—an improvement over the scatter of ASPCAP metallicities, which is 0.15 dex for these 15 sources. The internal consistency of the two models (the scatter of the eight M+M pairs both in APOGEE) is 0.06 dex for Cannon and 0.12 dex for ASPCAP.

As expected, the Toomre diagram in Figure 10 shows that higher-metallicity sources in the sample are concentrated in low-velocity space corresponding roughly to the thin-disk population; while the thick-disk population contains a slightly higher concentration of lower-metallicity sources. Separating the two populations into separate histograms (also shown in Figure 10), we find that thick-disk sources are marginally more metal-poor than thin-disk sources, with the mean ± standard deviation of [Fe/H] = 0.00 ± 0.17 dex for the thin-disk distribution, and [Fe/H] = −0.14 ± 0.19 dex for the thick-disk distribution. Metallicities of the two populations from ASPCAP show a similar distribution, with the mean ± standard deviation being [M/H] = −0.16 ± 0.16 dex for thin-disk sources and [M/H] = −0.23 ± 0.17 dex for thick-disk sources.

Figure 11 shows that ASPCAP metallicities are systematically lower than Cannon metallicities. We further find that the bias is temperature-dependent: at the highest temperatures ($T_{\text{eff}} > 3600$ K) ASPCAP and Cannon metallicities are consistent to a scatter of 0.05–0.06 dex and offset by an average of −0.12–0.15 dex, while at the lowest temperatures ($T_{\text{eff}} < 3200$ K) ASPCAP and Cannon are consistent to a scatter of −0.13 dex and offset by an average of −0.3 dex.

## 6. Discussion

We trained a data-driven model (The Cannon; Ness et al. 2015) to deliver high-quality atmospheric parameters ($T_{\text{eff}}$ and [Fe/H]) for M-type dwarf stars from high-resolution infrared spectra from APOGEE. This work was motivated by the problem that M dwarfs stars are difficult to model physically; the data are better than the models in important senses. Indeed we find that our data-driven model is both accurate in the data domain (as a spectral synthesis model) and precise in the latent domain (as a tool for deriving physical parameters). This accuracy and precision is consistent with previous work with The Cannon (Ness et al. 2015, 2018; Casey et al. 2016; Ho et al. 2017a), but here extends to a new regime in spectral type ($T_{\text{eff}}$). The primary result of this work is that we have compiled a catalog of 5875 M dwarfs with Cannon temperatures,
metallicities, spectral types, and six-dimensional kinematics. These data are provided in Table 6.

While The Cannon achieves excellent precision at predicting labels and reproducing spectral features, the accuracy of labels it produces is limited by the accuracy, relative precision, size, dynamic range, and representation of the training sample. That is, being a supervised method, The Cannon is never any better in a mean (bias) sense than the input training data, although it can be better in a precision or variance sense. The catalog we have produced is a label transfer from parameters provided in our input data (M15) and it implicitly adopts all the biases and issues from those input data. It is also limited to the stellar-parameter domain of that input catalog. That said, this work provides an external validation of the M15 stellar parameters.

The model we have developed does have limitations, however. For example, it delivers chi-squared goodness-of-fit measures that are large; the model is not technically an accurate

Figure 7. Full sample of 14,828 M dwarfs colored by Cannon labels before selection (left) and final sample selection of 5875 M dwarfs after applying selection criteria described in Section 5.3 (right), to reduce contamination from sources that are not similar to the training sample (not single, main-sequence M stars, such as pre-main-sequence, spectroscopic binaries, and K dwarfs). Overplotted with orange triangles are the M15 and W11 training samples, for their respective Cannon test labels. Temperature gradient increases with decreasing color, spectral subtype increases with increasing color, and metallicity gradient increases perpendicularly up from the main-sequence branch as expected. Deviations from these gradients seen at the upper boundary of the main sequence are likely remaining contamination from the binary sequence.
description of the spectra, especially when the spectra are observed at signal-to-noise levels above 100. The model does not include some known physical and instrumental effects, such as line broadening from rotation or convection (for example, Behmard et al. 2019), or binarity and the superposition of multiple stellar spectra (as in, say, El-Badry et al. 2018). The model also does not include any adjustments for instrumental variations, such as the small but significant variations of APOGEE resolution with spectrograph fiber number (as included in Ness et al. 2018).

The APOGEE instrument was designed to be sensitive to more than a dozen individual elemental abundances in stellar spectra. So the M-dwarf spectra analyzed here contain individual elemental abundance information that we have ignored. Exploitation of that information requires a better training set of M dwarfs than we have at present, but is an important goal for the future with these data.

While a detailed analysis of atmospheric model limitations is beyond the scope of this paper, our results provide an avenue to compare the metallicity scale for FGK stars to the less well-understood metallicity scale for M dwarfs. These results find that atmospheric metallicities are systematically metal-poor biased compared to Cannon-based metallicities trained on sources with metallicities calibrated to those of FGK companions. At the high-temperature end ($T_{\text{eff}} > 3600$ K), the ASPCAP metallicity bias is $-0.12 \pm 0.15$ dex with a scatter of $0.05 \pm 0.06$ dex relative to Cannon metallicities, and it increases to a bias of $-0.3$ dex and scatter of 0.13 dex at the low-temperature end ($T_{\text{eff}} < 3200$ K) (Figure 11). We suspect that this metal-poor bias, while not explored to a great extent in this work, is due to the line lists of the models—an effect in which the optimizer of the pipeline may be lowering the continuum level and metallicity of the fit to compensate for the missing lines or opacities. We also note that this analysis was

| Column          | Unit | Description                                                                 |
|-----------------|------|-----------------------------------------------------------------------------|
| APOGEE_ID       |      | APOGEE 2MASS designation                                                    |
| GAIA_ID         |      | Gaia identification number                                                 |
| T eff           | K    | Mann-trained Cannon effective temperature                                   |
| F e/H           | dex  | Mann-trained Cannon [Fe/H]                                                  |
| SPT             |      | West-trained Cannon spectral subtype (M0–M9)                               |
| T eff, APOGEE   | K    | ASPCAP pipeline effective temperature                                        |
| M H, APOGEE     | dex  | ASPCAP pipeline [M/H]                                                       |
| CHI_MANN        |      | $\chi^2$ fit of Mann-trained Cannon model                                 |
| CHI_WEST        |      | $\chi^2$ fit of West-trained Cannon model                                  |
| J, MAG          | mag  | APOGEE J-band photometry                                                   |
| H, MAG          | mag  | APOGEE H-band photometry                                                   |
| K, MAG          | mag  | APOGEE K-band photometry                                                   |
| B P, MAG        | mag  | Gaia BP-band photometry                                                    |
| R P, MAG        | mag  | Gaia RP-band photometry                                                    |
| G, MAG          | mag  | Gaia G-band photometry                                                    |
| J, ABS          | mag  | APOGEE J-band absolute magnitude                                            |
| H, ABS          | mag  | APOGEE H-band absolute magnitude                                            |
| K, ABS          | mag  | APOGEE K-band absolute magnitude                                            |
| G, ABS          | mag  | Gaia G-band absolute magnitude                                             |
| B P, RP         | mag  | Gaia BP – RP color                                                         |
| RA              | deg  | APOGEE right ascension angle                                                |
| DEC             | deg  | APOGEE decl. angle                                                          |
| PMRA            | mas yr$^{-1}$ | Gaia right ascension proper motion                               |
| PMRA, ERR       | mas yr$^{-1}$ | Gaia right ascension proper motion uncertainty                           |
| PMDEC           | mas yr$^{-1}$ | Gaia decl. proper motion                                                  |
| PMDEC, ERR      | mas yr$^{-1}$ | Gaia decl. proper motion uncertainty                                        |
| PLX             | mas  | Gaia parallax                                                               |
| PLX, ERR        | mas  | Gaia parallax uncertainty                                                  |
| DIST            | kpc  | Distance (1/parallax)                                                       |
| RV, APOGEE      | km s$^{-1}$ | APOGEE radial velocity                                                    |
| RV, APOGEE, ERR | km s$^{-1}$ | APOGEE radial velocity uncertainty                                         |
| RV, GAIA        | km s$^{-1}$ | Gaia radial velocity                                                       |
| RV, GAIA, ERR   | km s$^{-1}$ | Gaia radial velocity uncertainty                                           |
| Vx              | km s$^{-1}$ | Cartesian x velocity in Galactocentric coordinates                        |
| Vy              | km s$^{-1}$ | Cartesian y velocity in Galactocentric coordinates                        |
| Vz              | km s$^{-1}$ | Cartesian z velocity in Galactocentric coordinates                        |
| X               | kpc  | Cartesian x position in Galactocentric coordinates                        |
| Y               | kpc  | Cartesian y position in Galactocentric coordinates                       |
| Z               | kpc  | Cartesian z position in Galactocentric coordinates                       |

SNR | Signal-to-noise ratio of the APOGEE spectrum |

(This table is available in its entirety in machine-readable form.)
completed using data from Data Release 14 of APOGEE, which did include molecular lines from FeH in the pipeline at the time, which become numerous and strong for $T_{\text{eff}} \lesssim 3600$ K (Souto et al. 2017). Further analysis would need to be done to quantify the metallicity improvement for M dwarfs in future data releases of APOGEE, and determine whether the metallicity bias is found in other model grids (besides the ATLAS/MARCS models used by the ASPCAP pipeline), and whether the effect is present at other wavelengths.
Given that physics-based spectral models of M dwarfs have issues, one of the possible future values of the data-driven model shown here is that it is highly interpretable: it contains within it first and second derivatives of the spectral expectation with respect to the atmospheric parameters. We show some of these derivatives in Figures 12, 13, and 14 and deliver relevant...
Figure 12. Top two panels of each plot: APOGEE spectra (black), overlaid by the Mann-trained Cannon model for two sources of varying temperatures, and similar metallicities. Third panel of each plot: Derivative of The Cannon model with respect to temperature, taken at the median training temperature, $T_{\text{eff}} = 3463$ K; an error estimate computed using a jackknife statistic at each pixel is marked in red, making it possible to distinguish which features vary significantly with change in spectral type, and which are likely due to noise.
Figure 13. Top two panels of each plot: APOGEE spectra (black), overlaid by the Mann-trained Cannon model for two sources of varying metallicities and similar temperatures. Third panel: derivative of The Cannon model with respect to metallicity, taken at the median training metallicity, $[\text{Fe/H}] = -0.03$ dex; the jackknife computed error at each pixel is shown in red.
data in Tables 3 and 4. These tables summarize spectral features in the APOGEE bandpass that are found to be strong temperature and metallicity indicators. In the long run, this is the primary value of data-driven models for astronomy: to provide physical insights that drive physical understandings. It is our hope that The Cannon, and models like it, will lead to new and improved physical models which will, in turn, put The Cannon out of business.

We would like to acknowledge Hans Walter Rix (MPIA), Derek Homeier (Universität Heidelberg), Wolfgang Bradner (MPIA), Anna-Christina Eilers (MPIA), Melissa Ness (Columbia), Kevin Covey (WWU), Diogo Souto (Observatário Nacional/MCTI), Keivan Stassun (Vanderbilt), Katia Cunha (NOAO), Anthony Brown (Leiden), Aida Behmard (Caltech), and Christopher Theissen (UCSD) for constructive discussions in the process of this project, as well as Bertrand Goldman (MPIA) and those who have supported the internship program at the Max Planck Institute für Astronomie for providing J.B. with funding and hospitality. This project was partially supported by the US National Aeronautics and Space Administration (NASA grant NNX12AI50G), the US National Science Foundation (NSF grant AST-1517237), and the Moore–Sloan Data Science Environment at NYU. A.B. acknowledges funding support from the National Science Foundation under award No. AST-1517177. This work is supported by the SDSS Faculty and Student Team (FAST) initiative.

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

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: SDSS-IV (APOGEE), Gaia.

Software: Astropy (Astropy Collaboration et al. 2013, 2018), matplotlib (Hunter 2007), numpy (van der Walt et al. 2011), scipy (Jones et al. 2001), Topcat (Taylor 2005), The Cannon (Ness et al. 2015).

ORCID iDs

Jessica Birky @ https://orcid.org/0000-0002-7961-6881
David W. Hogg @ https://orcid.org/0000-0003-2866-9403
Andrew W. Mann @ https://orcid.org/0000-0003-3654-1602
Adam Burgasser @ https://orcid.org/0000-0002-6523-9536

References

Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2017, ApJ, 235, 42
Allard, F., Hauschildt, P. H., Alexander, D. R., & Starrfield, S. 1997, ARA&A, 35, 137
