DATA-DRIVEN SPECTRAL MODELS FOR APOGEE M DWARFS

Jessica Birky[1], David W. Hogg[2,3,4] and Adam Burgasser[1]

[1] UC San Diego, San Diego, CA; [2] New York University, NYC, NY; [3] Max Planck Institut für Astronomie, Heidelberg Germany; [4] Flatiron Institute, NYC, NY

ABSTRACT: **The Cannon** [1,2] is a flexible, **data-driven spectral modeling** and parameter inference framework, demonstrated on high-resolution Apache Point Galactic Evolution Experiment (**APOGEE**; λ/Δλ~22,500, 1.5-1.7μm) spectra of giant stars to estimate stellar labels (Teff, logg, [Fe/H], and chemical abundances) to precisions higher than the model-grid pipeline. The lack of reliable stellar parameters reported by the APOGEE pipeline for temperatures less than ~3550K [4], motivates the extension of this approach to M dwarf stars. Using a training set of **51 M dwarfs with spectral types ranging M0-M9** obtained from SDSS optical spectra, we demonstrate that The Cannon can **infer spectral types to a precision of 0.6 types**. We then use **30 M dwarfs ranging 3072 < Teff < 4131K, and -0.48 < [Fe/H] < 0.49** to train a two-parameter model **precise to 44K and 0.05 dex** respectively. Additionally we discuss the extension of a model to other labels, and the scientific objectives a data-driven pipeline could enable.

DATA-DRIVEN APPROACH

MODEL INPUT

MODEL ASSUMPTIONS:

1. Sources with identical labels have near-identical flux at each pixel.
2. Expected flux at each pixel varies continuosly with change in label.

INPUT: Set of ***training sources*** w/ known ***reference labels***, and a label vector.

TRAINING STEP

GENERATIVE MODEL:

$\mathbf{f}^{L}_{n\lambda}$ = predicted flux for pixel λ

$\boldsymbol{\theta}_{\lambda}$ = set of model coefficients for pixel λ

ℓ_{n} = label vector (in this case quadratic)

$$f^{L}_{n\lambda} = g(l^{A}_{n}|\theta_{\lambda}) + \text{noise}$$

$$f^{L}_{n\lambda} = \theta^{T}_{\lambda} \cdot l^{A}_{n} + \underbrace{[s^{2}_{\lambda} + \sigma^{2}_{n\lambda}]\varepsilon_{n\lambda}}_{\text{noise}}$$

Solving for the coefficients and scatter:

$$\ln p(f_{n\lambda}\,|\,\boldsymbol{\theta}^{T}_{\lambda}, \boldsymbol{\ell}_{n}, s^{2}_{\lambda}) = -\frac{1}{2}\frac{[f_{n\lambda} - \boldsymbol{\theta}^{T}_{\lambda}\cdot\boldsymbol{\ell}_{n}]^{2}}{s^{2}_{\lambda}+\sigma^{2}_{n\lambda}} - \frac{1}{2}\ln(s^{2}_{\lambda}+\sigma^{2}_{n\lambda})$$

$$\boldsymbol{\theta}_{\lambda}, s_{\lambda} \leftarrow \underset{\boldsymbol{\theta}_{\lambda}, s_{\lambda}}{\text{argmax}} \sum_{n=1}^{N} \ln p(f_{n\lambda}\,|\,\boldsymbol{\theta}^{T}_{\lambda}, \boldsymbol{\ell}_{n}, s^{2}_{\lambda})$$

TEST STEP

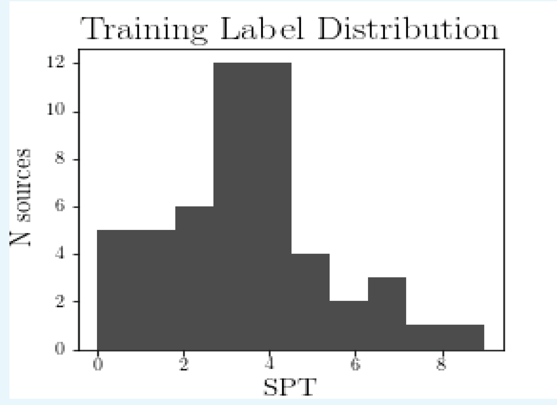
INFERING LABELS: optimize labels for each star, using $\boldsymbol{\theta}_{\lambda}$ & $\mathbf{s}_{\lambda}$ from training step:

$$\{\ell_{mk}\} \leftarrow \underset{\{\ell_{mk}\}}{\text{argmax}} \sum_{\lambda=1}^{N_{\text{pix}}} \ln p(f_{m\lambda}\,|\,\boldsymbol{\theta}^{T}_{\lambda}, \boldsymbol{\ell}_{m}, s^{2}_{\lambda})$$

VALIDATION: Model consistency tested by self-test (training vs test labels), and leave-one-out cross validation.

(See [1] - Ness et al. 2015)

SPECTRAL TYPE MODEL



TRAINING SAMPLE: West et al. 2011
51 sources, M0-M9

REFERENCE LABELS:
SPT from SDSS optical spectra using The Hammer.

QUADRATIC MODEL:

$$l_n \equiv [1, SPT, SPT^2]$$

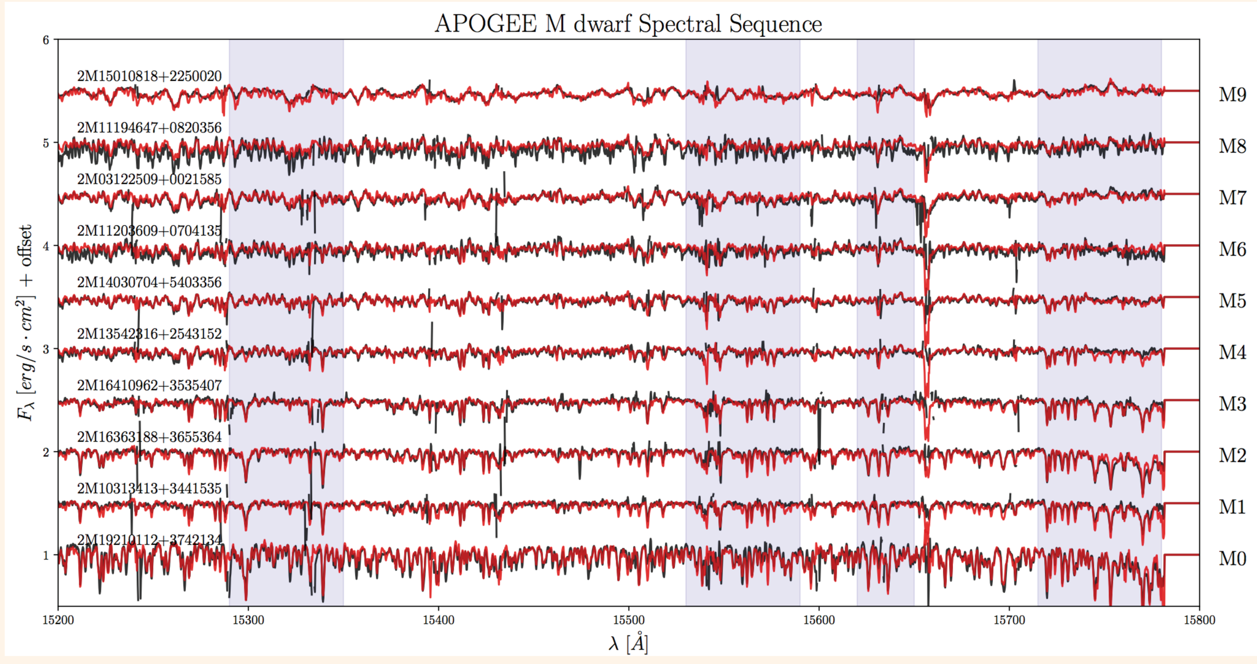
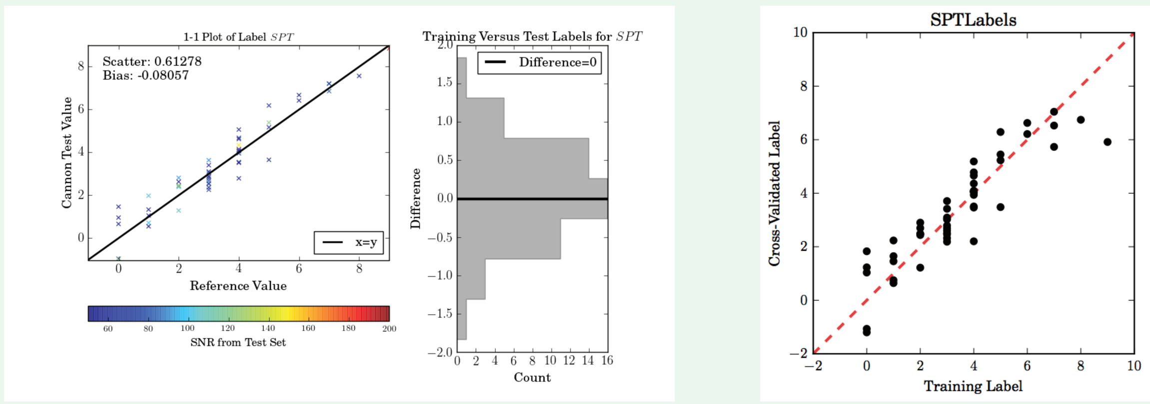
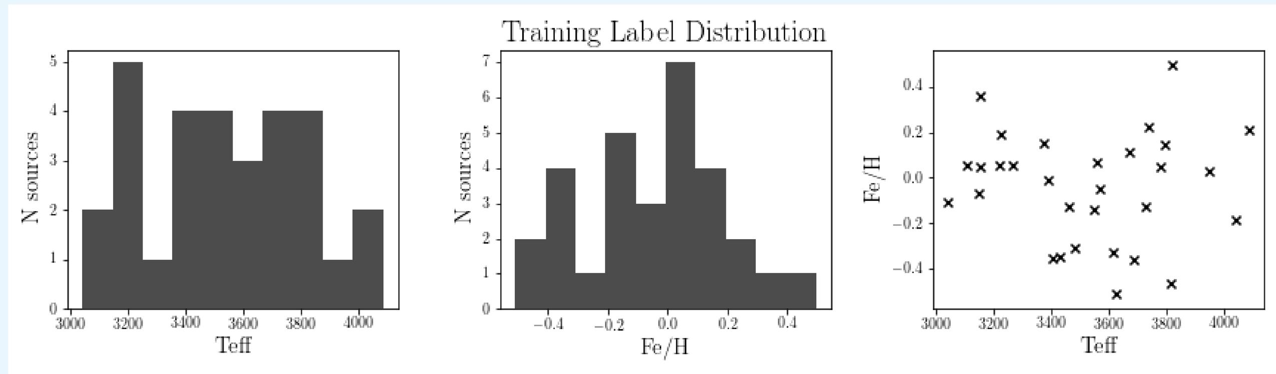


Figure 1: Spectral sequence of dwarfs in training set M0-M9; chip 1 of APOGEE spectrum with highlighted spectral type sensitive regions identified in [3].



Figures 4-5: Label self-test *(left)* and cross validation *(right)* for **West-trained model**.

TEMPERATURE/METALLCITY MODEL

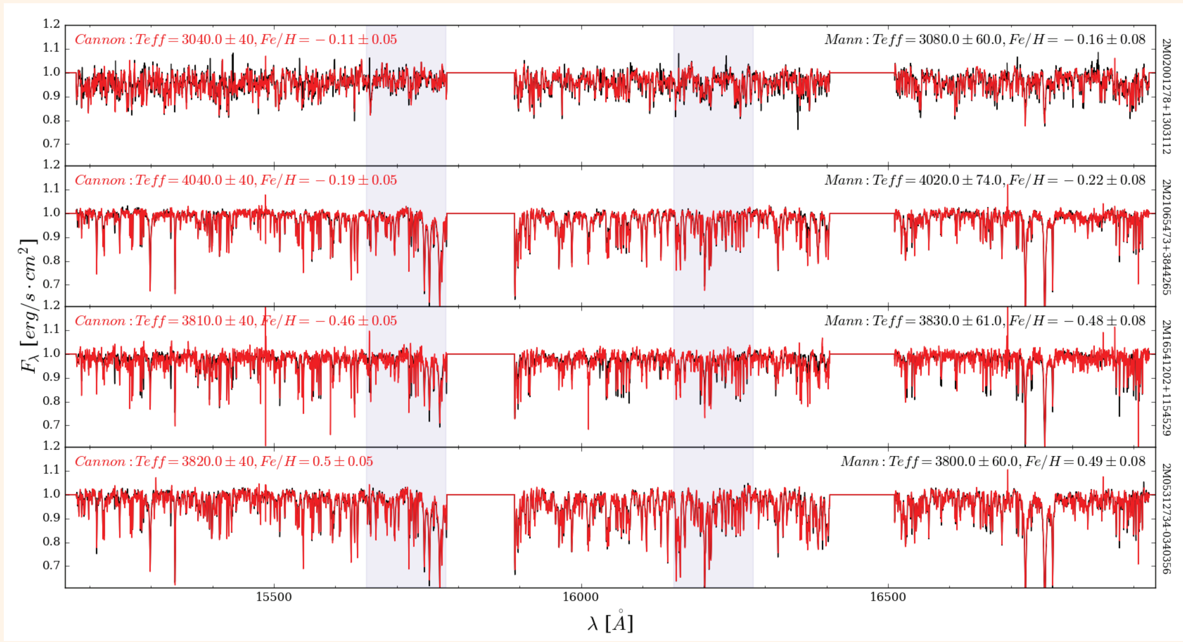


TRAINING SAMPLE: Mann et al. 2015
30 sources
3072 < Teff < 4131K, -0.48 < [Fe/H] < 0.49

REFERENCE LABELS: Teff - interferometry calibration; [Fe/H] - FGK pair calibration.

QUADRATIC MODEL:

$$l_n \equiv [1, T_{\text{eff}}, [Fe/H], T^2_{\text{eff}}, T_{\text{eff}}\cdot[Fe/H], [Fe/H]^2]$$



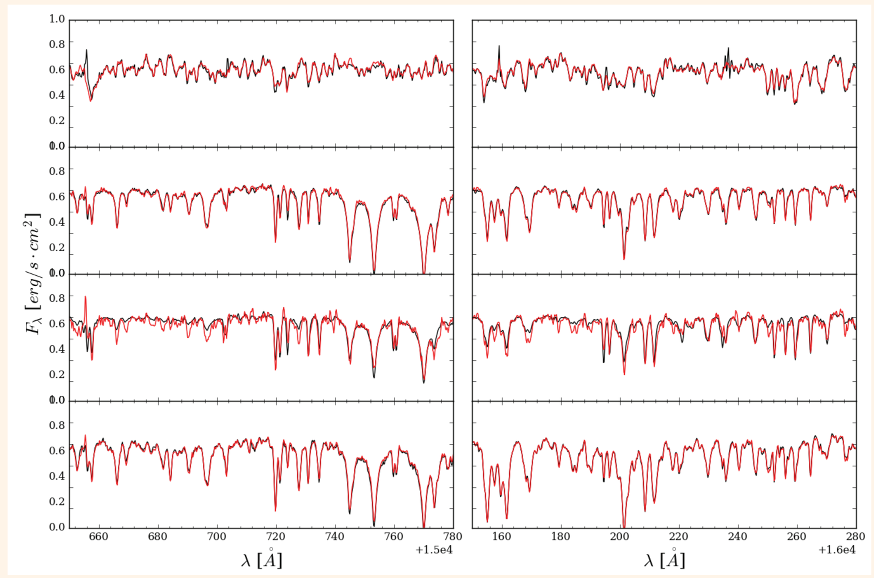
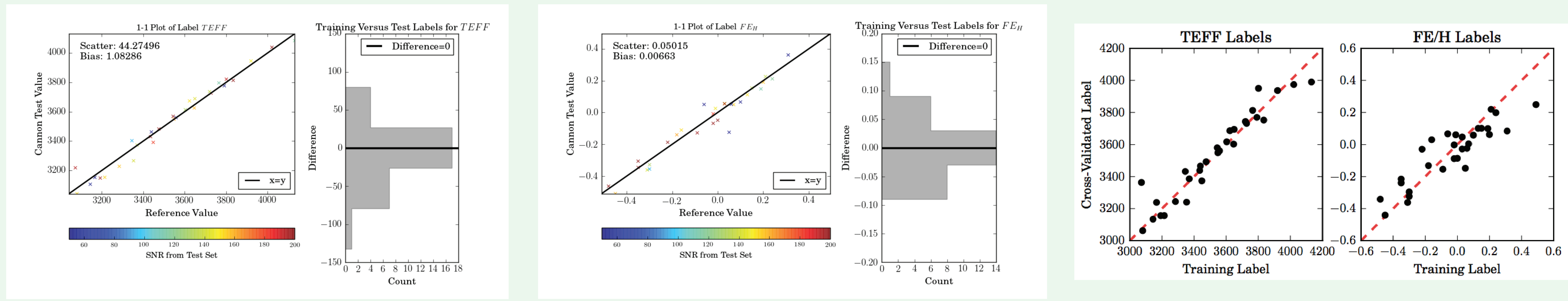


Figure 2: Top two plots: Mann-trained model for varying temperatures; bottom two plots: Mann-trained model for varying metallicities.
Figure 3: Zoomed in plot of two regions highlighted in Figure 2.



Figures 6-8: Label self-test *(left two)* and cross validation *(right)* for **Mann-trained model**.

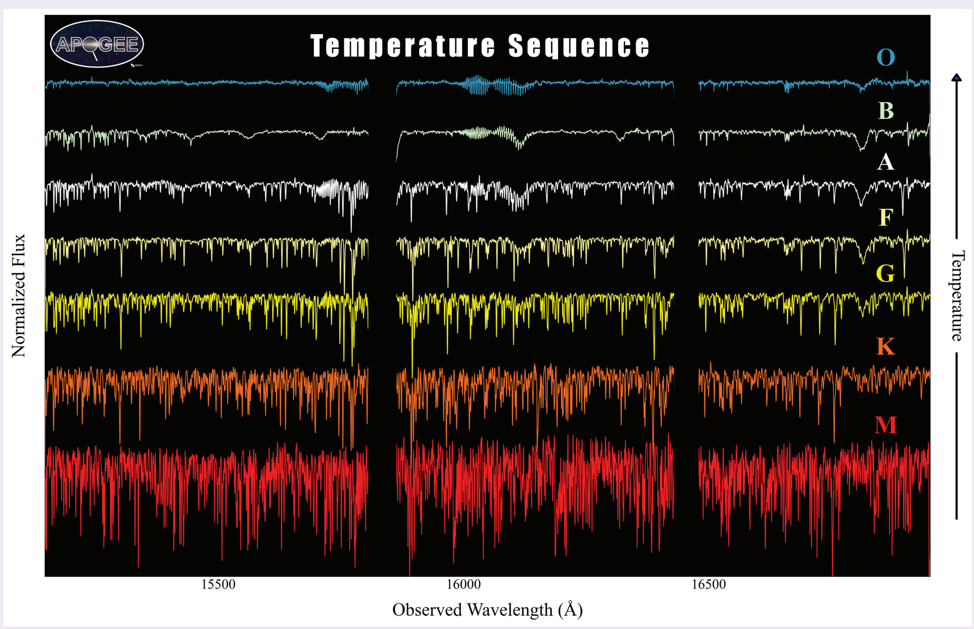
THE APOGEE SURVEY

SPECIFICATIONS:
R~22,500, 1.5-1.7 μm; Targeted mainly at giant sources with the objective of studying galactic structure [11].

M DWARF CHALLENGE:
ASPCAP pipeline fails to deliver reliable parameters for sources cooler than ~3550K [4].

Numerous overlapping features present in sources this cool make it infeasible to use equivalent width methods [5].

Spectral synthesis with precomputed model grids has produced some stellar parameter estimates of the warmer sources (>M5) [5,10]



DISCUSSION

MODEL UNCERTAINTIES: Uncertainties are given by the scatter of the model and are more precise than reported training set uncertainties (1 SPT for West, and 60K/0.08dex for Mann).

Label	Scatter
SPT	0.61 types
Teff	44 K
[Fe/H]	0.05 dex

TEST OF REFERENCE LABEL QUALITY: The fact that The Cannon derives such precise values for both the West-trained model and the Mann-trained model, is perhaps a good validation that the reference parameters are very accurate.

OTHER TRAINING SETS: Several other training sets were tested in this project (which included testing logg & color magnitude labels), however the two other most sizable training sets from Rajpurohit [10] and from Simbad (each ~45 sources) were not very consistent and had uncertainties >140K in Teff and >0.2 dex in [Fe/H].

MODEL CAVEATS: Particular: Training sets are relatively small (30 and 51 sources) and are weak in the very low mass range (only 1 M8 and M9 in SPT model; no sources <M5 in Teff/[Fe/H] model); **General:** No fitting for vsini or LSF broadening; assume reference labels are very accurate; assume all stars have same lineshapes.

WHY DATA-DRIVEN MODELS?

WORK WITH THE CANNON:
The Cannon has been used on APOGEE giants to infer stellar parameters (Teff, logg, [Fe/H]) [1] and 15 elemental abundances [2] to higher precisions than ASPCAP.

SOLUTION: Data-driven models take away the challenge of directly infering labels from a survey [3]--instead we ***transfer labels*** from another (more accurate/easier-to-model) survey.

BENEFITS: Fast computation time; flexible model labels (can train on any parameters that you have reference labels for); flexible label vectors (can specify degree of polynomial). Enables systematic search for lines/features that vary strongly with change in parameter.

Accurate training parameters + very precise label transfer = high quality label inference!

FUTURE WORK

IMPROVE THE MODEL: Expand training sets by either (1) obtaining more reference labels for other APOGEE M dwarfs (expanding Mann's sample or observing sources w/ SpeX/NIR-SPEC), or (2) obtaining APOGEE spectra for more known M dwarfs. Also, construct a training set with more reference labels (logg, abundances).

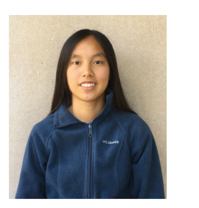
M DWARF PIPELINE FOR APOGEE: Identify, classify and label all of the (probably 1000s of) M dwarfs in the APOGEE survey, which do not have reliable parameters from the ASPCAP pipeline--will require training additional Cannon models to descriminate M stars from hotter stars, and dwarfs from giants.

SCIENTIFIC GOALS: strong match to features => precise radial velocity measurements (look at rv variations over multiple epochs, and velocity distributions in galaxy); chemical abundance analysis; line analysis and comparison to theoretical models (i.e. BT-Settl, PHOENIX).

REFERENCES:

[1] Ness et al. 2015, ApJ, 808, 16
[2] Casey et al. 2016, ApJ, 840, 1
[3] Ho et al. 2016, ApJ, 808, 1
[4] Schmidt et al. 2016, MNRAS, 460, 2611
[5] Deshpande et al. 2013, AJ, 146, 156
[6] Mann A. W., et al. 2015, ApJ, 804, 64
[7] West et. al. 2011, AJ, 141, 97
[8] Burgasser et. al. 2014, ASICS, 10, 1-10
[9] Souto et al. 2017, ApJ, 835, 239
[10] Rajpurohit et al. 2017, A&A, arXiv:1708.06211
[11] Majewski et. al. 2015, AJ, arXiv:1509.05420

UC San Diego



Email: jbirky@ucsd.edu
https://github.com/jbirky/