The Astrophysical Journal Supplement Series, 246:9 (14pp), 2020 January https://doi.org/10.3847/1538-4365/ab55ef

Deriving the Stellar Labels of LAMOST Spectra with the Stellar LAbel Machine (SLAM)

Bo Zhang[1,2], Chao Liu[1], and Li-Cai Deng[1]
[1] Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China; bozhang@nao.cas.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
Received 2019 August 22; revised 2019 October 5; accepted 2019 October 19; published 2020 January 6

Abstract

The LAMOST survey has provided 9 million spectra in its Data Release 5 (DR5) at $R \sim 1800$. Extracting precise stellar labels is crucial for such a large sample. In this paper, we report the implementation of the Stellar LAbel Machine (SLAM), which is a data-driven method based on support vector regression (SVR), a robust nonlinear regression technique. Thanks to the capability to model highly nonlinear problems with SVR, SLAM can generally derive stellar labels over a wide range of spectral types. This gives it a unique capability compared to other popular data-driven methods. To illustrate this capability, we test the performance of SLAM on stars ranging from $T_{\rm eff} \sim 4000$ to $\sim$8000 K trained on LAMOST spectra and stellar labels. At g-band signal-to-noise ratio (S/N_g) higher than 100, the random uncertainties of $T_{\rm eff}$, log g, and [Fe/H] are 50 K, 0.09 dex, and 0.07 dex, respectively. We then set up another SLAM model trained by APOGEE and LAMOST common stars to demonstrate its capability of dealing with high dimensional problems. The spectra are from LAMOST DR5 and the stellar labels of the training set are from APOGEE DR15, including $T_{\rm eff}$, log g, [M/H], [α/M], [C/M], and [N/M]. The cross-validated scatters at $S/N_g \sim 100$ are 49 K, 0.10 dex, 0.037 dex, 0.026 dex, 0.058 dex, and 0.106 dex for these parameters, respectively. This performance is at the same level as other up-to-date data-driven models. As a byproduct, we also provide the latest catalog of $\sim$1 million LAMOST DR5 K giant stars with SLAM-predicted stellar labels in this work.

Unified Astronomy Thesaurus concepts: Astronomical methods (1043); Astronomy data analysis (1858); Bayesian statistics (1900); Stellar abundances (1577); Chemical abundances (224); Fundamental parameters of stars (555); Catalogs (205); Surveys (1671)

Supporting material: FITS file

1. Introduction

As large spectroscopic surveys, e.g., Sloan Digital Sky Survey (SDSS)/SEGUE (Beers et al. 2006), RAVE (Steinmetz et al. 2006), SDSS/APOGEE (Majewski 2012), LAMOST (Deng et al. 2012), *Gaia*-ESO (Gilmore et al. 2012), and GALAH (Freeman 2012) proceed, deriving the stellar labels (or stellar parameters) is of extreme importance. In particular, such large surveys often observe stars covering a large range of spectral types. LAMOST, for instance, has observed stars from O-type to M-type (Liu et al. 2019; Zhong et al. 2019). This requires that the stellar label estimator must be able to deal with stellar samples over a large range of spectral types.

Stellar labels are usually determined by comparing an observed spectrum to a stellar spectral library (either a precomputed synthetic or empirical stellar spectral library). The idea of data-driven methods (The Cannon, Ness et al. 2015) is proposed for its capability to set up the mappings from stellar labels to spectra with a training set and use them to predict stellar labels for the observed spectra. It has not only been proved competitive to ASPCAP in the APOGEE case (Ness et al. 2015) but has also demonstrated the capability of predicting stellar labels from the low-resolution spectra of LAMOST K giant stars (Ho et al. 2017b, 2017a). Casey et al. (2017) and Buder et al. (2018) extended the application of The Cannon to main-sequence stars in the analysis of the RAVE and GALAH data, respectively. Based on quadratic models, improvements such as regularization have been made to make The Cannon capable of predicting stellar labels more precisely (Casey et al. 2016, 2017).

In the training stage, with a training set, a Cannon-like method uses regression methods to build a generative model of the spectral flux at a given wavelength as a function of stellar labels, i.e.,

$$F(\lambda) = f_\lambda(T_{\rm eff}, \log g, [{\rm X/H}], \ldots), \quad (1)$$

where $F(\lambda)$ is the normalized spectral flux at wavelength λ, f_λ is the assumed form of spectral flux at λ, and $T_{\rm eff}$, log g, and [X/H] are the stellar effective temperature, surface gravity, and elemental abundances, respectively. In Ness et al. (2015), f_λ is adopted as a quadratic function whose coefficients are optimized in the training process in order to fit the training set. A more general case is discussed by Rix et al. (2016). In the test step, the stellar labels are determined by operating the generative model to search for a model spectrum that best fits the observed one.

The idea of data-driven methods is important. However, a better form of f_λ is needed when modeling spectra that cover a wide range of spectral types. For instance, at around some strong atomic lines, fluxes can dramatically change in highly nonlinear ways with $T_{\rm eff}$ and log g. In the left/right panel of Figure 1, we show the trends of normalized synthetic fluxes from PHOENIX library (Husser et al. 2013) at around Mg b/Hα. It is clearly seen that a quadratic function is no longer sufficient to associate the stellar labels with spectral fluxes when $T_{\rm eff}$ changes from 3000 to 15000 K. This is also shown in Ting et al. (2019).

One possible solution, the Payne (Ting et al. 2017b, 2019), is based on neural networks (NN). It is fascinating because in the

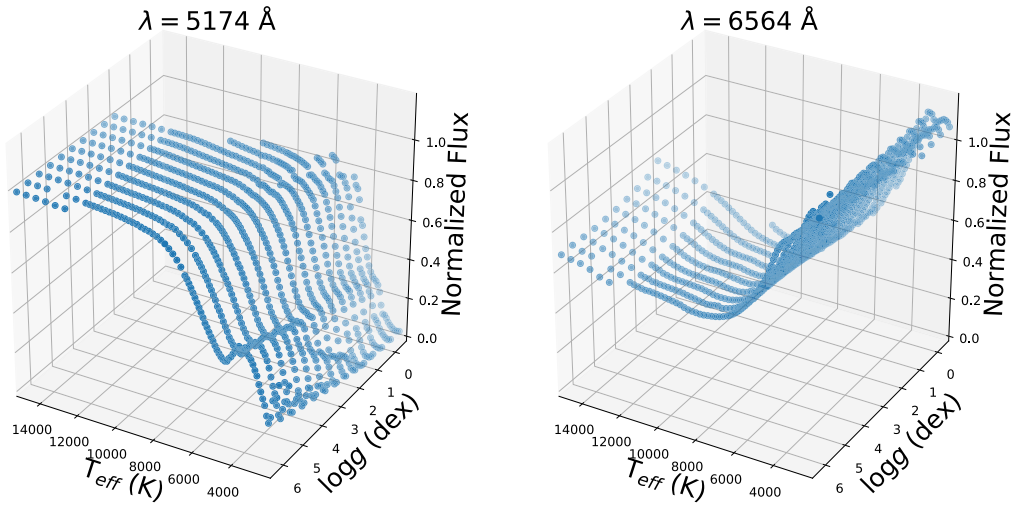

Figure 1. Examples of how spectral flux changes with two primary stellar labels, i.e., $T_{\rm eff}$ and $\log g$, at two fixed wavelengths. Blue dots in left and right panels are the flux values of normalized PHOENIX spectra with [Fe/H] = 0.0 at $\lambda = 5174$ Å (Mg b) and 6564 Å (Hα), respectively. Since these two pixels are around the spectral lines, which are extremely important in deriving stellar labels, a model's fitting performance for such data is crucial for stellar label prediction.

training stage the cost function of the Payne is regularized by a synthetic gradient. However, an NN-based method may suffer from the "when-to-stop" problem because the learning curve would not tell one when the NN is optimized (neither overfitting nor underfitting). As a consequence, the optimization of these kinds of methods depends on the expertise and experience of the users.

The support vector regression (SVR; Smola & Schölkopf 2004; Chang & Lin 2011), which is well-known for its robustness in modeling noisy data, is not a newcomer in the field of spectral data analysis. It is often used to build the mapping from stellar spectra to stellar labels in previous works (Li et al. 2014; Liu et al. 2014; Bu & Pan 2015; Lu & Li 2015). In this paper, following the idea of data-driven approaches, we present an alternative implementation, the Stellar LAbel Machine (SLAM), by using SVR to build a generative model of spectra, which automatically adjusts the model complexity according to data, and robustly extract as much information as possible from stellar spectra. Section 2 gives a brief description of SLAM and Section 3 assesses the performance of SLAM using the LAMOST DR5. In Section 4, we predict stellar labels for LAMOST DR5 K giant stars using SLAM with APOGEE DR15 stellar labels as the training data. Then we present the resulting catalog of more than a million red giant stars with precise stellar labels. We discuss the advantages and disadvantages of SLAM in Section 5 and present the coefficients of dependence (CODs) in Section 6. At last, we conclude this paper in Section 7.

2. The Construction of SLAM

In principle, SLAM consists of three steps.

1. The first step is data preprocessing. This includes spectra normalization and training data standardization, e.g., rescale both stellar labels and spectral fluxes so that their mean is 0 and variance is 1.
2. The second step is to train the SVR model at each wavelength pixel using the training set.
3. And the last is to predict stellar labels for observed spectra.

The details are described below.

2.1. Preprocessing

This step is to map all the normalized spectral fluxes and the stellar labels of the training set in standardized space (with zero mean and unity variance). It is necessary for most machine-learning methods, including SVR, to avoid issues due to the different scales in different dimensions of the input data.

After correcting its radial velocity (RV), each stellar spectrum in the training set is normalized by dividing its pseudocontinuum. In SLAM, we first use a smoothing spline (de Boor 1978) to smooth the whole spectrum and then exclude those pixels deviates from the smoothed spectrum by a distance larger than a threshold, e.g., 1.5 times the standard deviation of the residuals in the wavelength bin. The pseudocontinuum is then estimated by smoothing the reserved pixels in the spectrum. The softness of the smoothing spline, the distance threshold and the width of the wavelength bins can be adjusted using experience and also according to the properties of the spectral data in hand. Then all stellar spectra are resampled to the same wavelength grid. Assuming that we have m stellar spectra in the training set and each spectrum has n pixels, let $F_{i,j}$ be the jth pixel of the ith normalized stellar spectrum in the training set, then we have

$$\mu_i = \frac{1}{m}\sum_{i=1}^{m} F_{i,j} \tag{2}$$

and

$$s_i = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(F_{i,j}-\mu_i)^2}\,. \tag{3}$$

$F_{i,j}$ is then standardized via

$$f_{i,j} = \frac{F_{i,j}-\mu_i}{s_i}. \tag{4}$$

Stellar labels are also standardized in the same way. When the stellar labels are estimated for the observed spectrum in the prediction process, they will be rescaled back to physical units.

It is noted that bad pixels are quite common in spectroscopic surveys due to sky subtraction, cosmic rays, and problems that occur in data reduction. These bad pixels contain no information about stars and their errors cannot be estimated so that they should be excluded in our analysis. Usually, they flagged in the released spectral data products by setting flux error to a very large number or assigning a special flag. In particular, in our analysis, the LAMOST spectra provide `OR_MASK` for every pixel in a spectrum, which equals to zero when no problems occur in any single exposure and otherwise equals to a nonzero integer depending on the kind of problem it suffers from (see http://dr5.lamost.org/doc/data-production-description for more information). We exclude those bad pixels with nonzero `OR_MASK` values by assigning zero weights in the final spectral fitting.

2.2. Training

SVR is a robust nonlinear regression method and has been used in many astronomical studies (Liu et al. 2012, 2015), particularly in spectral data analysis (Li et al. 2014; Liu et al. 2014; Bu & Pan 2015; Lu & Li 2015). A more complete description of SVR can be found in Smola & Schölkopf (2004). Since SLAM is implemented with `python`, we adopt the python wrapper of LIBSVM[3] in the `scikit-learn` (Pedregosa et al. 2012) package for convenience.

There are two free hyperparameters, C and ϵ, that represent the penalty level and tube radius, respectively, in the SVR algorithm. We adopt the radial basis function as the kernel (RBF kernel, $K(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\gamma||\boldsymbol{x} - \boldsymbol{x}'||)$) in SVR. As a consequence, an additional hyperparameter γ, which indicates the width of the RBF kernel, also needs to be determined.

The choice of the hyperparameters, C, ϵ, and γ, sets the complexity of the SVR model. For example, a large C penalizes outliers heavily so that the regression will probably be very curved to pass through as many data points as possible, while a small C tells SVR to ignore the outliers and follow a smooth trend of the data. In SLAM, the best values of the hyperparameters are not freely controlled but are automatically determined by the training set itself. In other words, it is the training set itself, not the user, that determines the adopted model (SVR) complexity pixel-by-pixel.

2.2.1. Adaptive Model Complexity

As described above, the jth pixel in the training set has a mean of 0 and a variance of 1. Let $\boldsymbol{\theta}_i$ denote the stellar label vector of the ith star in the training set (i.e., a vector consisting of $T_{\rm eff}$, $\log g$ and elemental abundances) and $f_j(\boldsymbol{\theta}_i)$ be the jth pixel of the model output spectrum corresponding to the input stellar label vector $\boldsymbol{\theta}_i$. Once the model is trained via a specific set of hyperparameters, we are able to evaluate the mean squared error (MSE) and median deviation (MD) separately defined as

$$\mathrm{MSE}_j = \frac{1}{m}\sum_{i=1}^{m}[f_j(\boldsymbol{\theta}_i) - f_{i,j}]^2 \tag{5}$$

and

$$\mathrm{MD}_j = \frac{1}{m}\sum_{i=1}^{m}[f_j(\boldsymbol{\theta}_i) - f_{i,j}]. \tag{6}$$

MSE quantifies the scatter of the regression model and MD quantifies the bias. For the worst regression model, i.e., a constant model, $\mathrm{MSE} = 1$ because it turns out to be the variance of $f_{i,j}$ according to Equation (5). Theoretically, the smaller MSE is, the better the fitting is. However, if we train SVR models directly on the whole training set and pursue the model that minimizes both MSE_j and MD_j, we probably get an overfitted model which gives us $\mathrm{MD}_j = \mathrm{MSE}_j = 0$ by interpolating data. In practice, to avoid the overfitting problem, we use the k-fold cross-validated MSE (CV MSE) and k-fold cross-validated MD (CV MD) instead, i.e., evaluate the Equations (5) and (6) via though the k-fold cross-validation technique. In this process, the training set is randomly split out into k subsets (usually 5 to 10), and the $f_j(\boldsymbol{\theta}_i)$ is predicted by an SVR model trained on the other $k-1$ subsets of the training set. After looping over all subsets, we calculate the MSE_j and MD_j based on these predicted fluxes in cross-validation and the true fluxes in the training set. To distinguish them from normal MSE_j and MD_j without cross-validation, we name them CV MSE_j and CV MD_j, respectively.

Because overfitting is generally avoided through such a cross-validation technique, we can use the CV MSE_j and CV MD_j to reasonably assess how well the SVR with such model complexity can reproduce the spectral flux of the jth pixel in the training set. In particular, CV MD_j is usually very small once a proper model complexity is adopted. Therefore, the best model complexity (the best set of hyperparameters) can be determined by searching for the lowest CV MSE_j after looping over all sets of hyperparameters specified. Finally, we train SVR with the chosen best set of hyperparameters on the whole training set for this pixel. The MSE_j and MD_j of this final model are calculated directly based on the whole training set to quantify the scatter and bias of the SVR with the best model complexity. The final MSE_j is also adopted as the model error in the later processes. By doing so pixel-by-pixel, we guarantee the best model complexity for each pixel. As a comparison, the final MSE and MD of SLAM are compared to The Cannon in Section 3 to show the improvements introduced by this adaptive model complexity. In the Appendix A we use mock data to show how to choose the best hyperparameters from a grid more comprehensively.

2.3. Prediction

With the Bayesian formula, the posterior probability density function of stellar labels given an observed spectrum is shown as the following,

$$p(\boldsymbol{\theta}|\boldsymbol{f}_{\rm obs}) \propto p(\boldsymbol{\theta}) \prod_{j=1}^{n} p(f_{j,\rm obs}|\boldsymbol{\theta}), \tag{7}$$

where $\boldsymbol{\theta}$ is the stellar label vector, $\boldsymbol{f}_{\rm obs}$ is the observed spectrum vector, $f_{j,\rm obs}$ is the jth pixel of the normalized observed spectral flux, $p(f_{j,\rm obs}|\boldsymbol{\theta})$ is the likelihood of the spectral flux $f_{j,\rm obs}$ given $\boldsymbol{\theta}$, and $p(\boldsymbol{\theta})$ is the prior of $\boldsymbol{\theta}$. The estimation of stellar labels can be easily done by maximizing the posterior probability $p(\boldsymbol{\theta}|\boldsymbol{f}_{\rm obs})$. Although it is important to set a proper prior of

[3] A multiprogramming language package to solve the support vector machine problems, including SVR regression provided by Chang & Lin (2011).

stellar parameters from an external source (e.g., the Galactic model, parallax, proper motions), we adopt a uniform prior in this paper for simplicity in SLAM. A prior can be added depending on the specific scientific scenario in future works.

Adopting a Gaussian likelihood, the logarithmic form of Equation (7) becomes

$$\ln p(\boldsymbol{\theta}|\boldsymbol{f}_{\rm obs}) = -\frac{1}{2}\sum_{j=1}^{n} \times \left\{\frac{[f_{j,\rm obs} - f_j(\boldsymbol{\theta})]^2}{\sigma_{j,\rm obs}^2 + \sigma_j(\boldsymbol{\theta})^2} + \ln[2\pi(\sigma_{j,\rm obs}^2 + \sigma_j(\boldsymbol{\theta})^2)]\right\}, \quad (8)$$

where $f_{j,\rm obs}$ is the jth pixel of the observed spectrum, $f_j(\boldsymbol{\theta})$ is the output spectral flux given the stellar label vector $\boldsymbol{\theta}$, $\sigma_{j,\rm obs}$ is the uncertainty of the jth pixel of the observed spectrum, and $\sigma_j(\boldsymbol{\theta})$ is the uncertainty of the jth pixel of the output spectrum corresponding to the stellar labels $\boldsymbol{\theta}$. In practice, $\sigma_j(\boldsymbol{\theta})$ is roughly replaced with CV MSE$_j$, which is independent of $\boldsymbol{\theta}$.

Using the Markov Chain Monte Carlo technique to sample the posterior function Equation (8) for millions of spectra is not practical due to the computational cost, especially when the number of dimensions is large. Therefore, we adopt the maximum likelihood estimation method with the Levenberg–Marquardt (Moré 1978) least-squares optimizer in this work. The initial guess of $\boldsymbol{\theta}$ is determined by comparing the $\boldsymbol{f}_{\rm obs}$ to the training set and picking the one with maximum likelihood. The outputs are the stellar labels that maximize the likelihood function and the corresponding covariance matrix. The convergence status is also part of the output, and stars will be marked out if SLAM does not converge within the maximum number of iterations.

2.3.1. Uncertainty

The output covariance matrix of SLAM is converted from the Hessian matrix produced in the `scipy.optimize.least_squares` method in the SciPy package (Virtanen et al. 2019). We refer to Moré (1978) for how the Hessian matrix is calculated. The diagonal elements of the covariance matrix are considered as the formal errors for the corresponding stellar labels, hereafter we call them SLAM errors.

When making predictions for a data set whose true stellar labels are known, we are able to calculate the cross-validated scatter (CV scatter) and cross-validated bias (CV bias), which are considered as the standard deviation and mean deviation, respectively. Namely,

$$\text{CV bias} = \frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{\theta}_{i,\rm SLAM} - \boldsymbol{\theta}_i) \quad (9)$$

and

$$\text{CV scatter} = \frac{1}{m}\sqrt{\sum_{i=1}^{m}(\boldsymbol{\theta}_{i,\rm SLAM} - \boldsymbol{\theta}_i)^2}. \quad (10)$$

Note that the CV scatter/bias are statistics of stellar labels, while the CV MSE/MD described above are statistics of stellar spectra. In principle, a good data-driven method has a very small CV bias and CV scatter. To investigate the precision of a data-driven method, the CV scatter should be used because the CV scatter quantifies the precision, while SLAM error represents the internal uncertainty of the optimization method.

3. Tests on LAMOST DR5

3.1. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) Survey

The LAMOST telescope, also called the Guo Shou Jing Telescope, is a 4 m reflecting Schmidt telescope with a 20 square degree field of view, on which 4000 fibers are installed. The spectral resolution is $R \sim 1800$ covering all optical wavelengths (Cui et al. 2012; Deng et al. 2012; Liu et al. 2014, 2015). The r-band apparent magnitude of the survey covers from 9 to 17.8 mag. In this work, we use the LAMOST Data Release 5 (DR5), which includes observations from September in 2011 to June in 2017. The LAMOST DR5 provides ∼9 million spectra among which ∼5 million have stellar parameters estimated by LASP (Wu et al. 2011, 2014). We use this data set to investigate the performance of SLAM on dealing with a wide range of $T_{\rm eff}$ and compare it with The Cannon.

3.2. Training

The stellar labels of LAMOST AFGK stars used here are estimated with the LAMOST Stellar Parameter pipeline (LASP; Wu et al. 2011, 2014) and can be downloaded from http://dr5.lamost.org/. We first select the samples with reliable stellar parameters using the following empirical criteria:

1. $4000 < T_{\rm eff}/{\rm K} < 8000$,
2. $0.5 < \log g/{\rm dex} < 5.5$,
3. $-2.5 < {\rm [Fe/H]}/{\rm dex} < 1.0$,
4. $\Delta(T_{\rm eff})/{\rm K} < 200$,
5. $\Delta(\log g)/{\rm dex} < 0.1$,
6. $\Delta({\rm [Fe/H]})/{\rm dex} < 0.1$.

Then, we randomly select 5000 training stars with a high g-band signal-to-noise ratio ($120 < {\rm S/N}_g < 140$) among them, where ${\rm S/N}_g$ refers to the average S/N of the LAMOST spectrum at g-band. To validate the model at different S/Ns, we separate ${\rm S/N}_g$ into six intervals, i.e., (5, 20), (20, 40), (40, 60), (60, 80), (80, 100), and (100, 120) and randomly select 5000 stars in each of the ${\rm S/N}_g$ intervals as the test sets.

All spectra are shifted to the rest frame using the LAMOST DR5 RV and resampled to 1.0 Å step wavelength grid from 3900 to 5800 Å. And those with more than 50 bad pixels are excluded. The reason for this exclusion is that for these spectra we are not certain about whether their pseudocontinuum estimated is consistent with other spectra or not. In the training process, we set the grid of hyperparameters to be $\epsilon = 0.05$, $C = [10, 100]$ and $\gamma = [0.1, 0.01]$. Each pixel is trained with SVR and set with the hyperparameters based on 5-fold cross-validation. ϵ, though one of the three hyperparameters, is simply fixed because it represents the tube radius of the SVR outside of which the SVR regard data as outliers and ignores them. The role of ϵ is very like the tolerance of the regression function. Taking this training set as an example, the typical standard deviation of the normalized flux is about 0.02. Therefore, the tube radius stands for $0.02 \times 0.05 = 0.001$ in normalized flux (recall that the normalized flux is standardized to have a variance of unity and SVR works in the standardized space). So the typical "S/N" ratio of a spectrum predicted by the SVR model is around $1/0.001 = 1000$. In other words, the SVR could reproduce spectra at an "S/N" ratio at 1000. Decreasing ϵ to an even lower value to raise this S/N ratio is

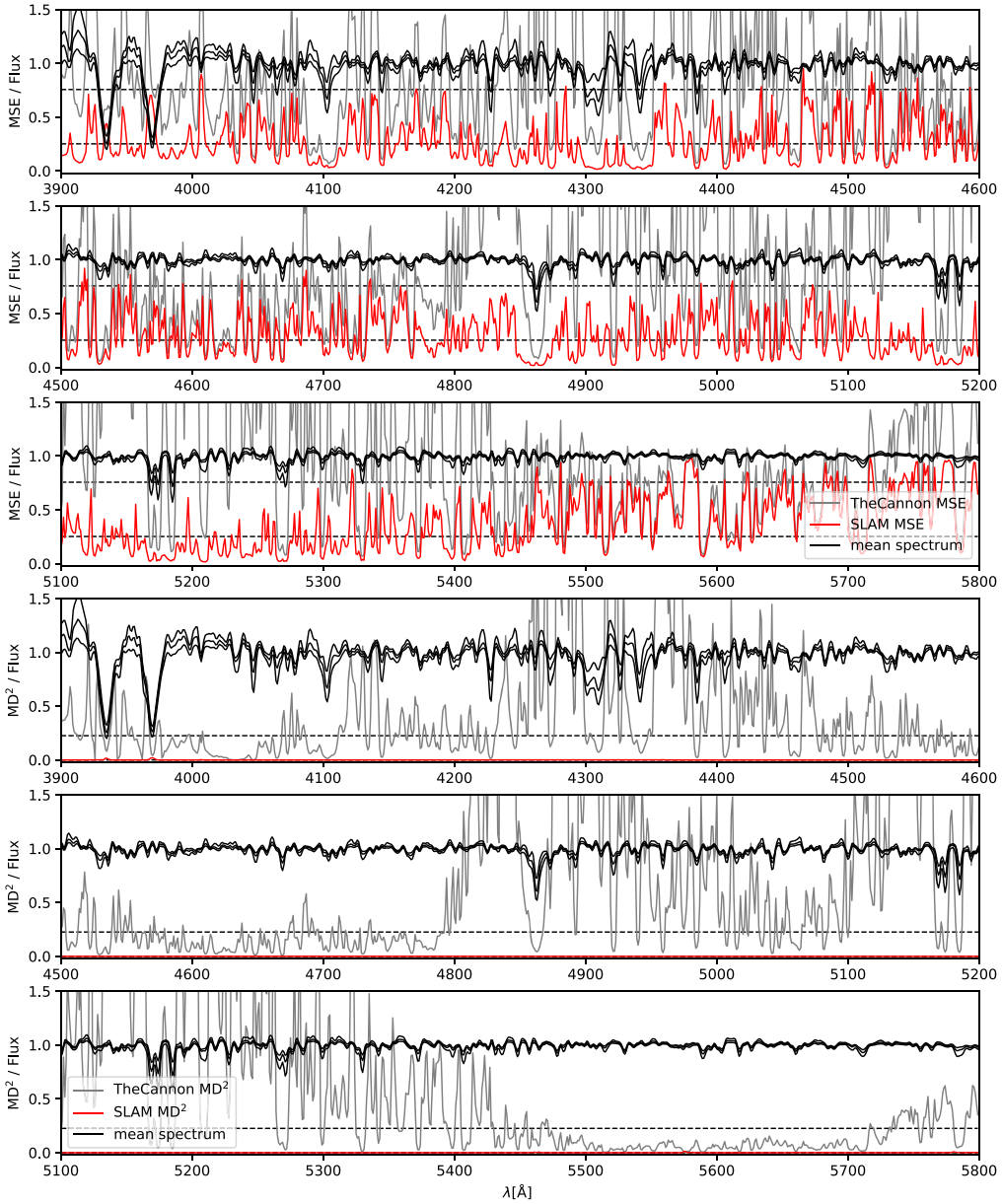

Figure 2. This figure shows the comparison of MSE and MD^2 of SLAM and The Cannon. The black lines always represent the 16, 50, and 84 percentile values of each pixel of the spectra of the training set. The red lines in the upper (lower) three panels show the MSE (MD^2) from SLAM, while the gray lines show similar quantities derived from The Cannon. The black dashed lines are the mean level of the MSE/MD^2.

not necessary as in our test we typically work at $S/N \sim 100$, while increasing ϵ could make our training more computationally expensive and easily get overfitted. For the C and γ, it could be inferred that setting them too large or too small is unnecessary in the standardized space from Appendix A. As a test, we try to make the grid of C and γ very sparse and see how good the results could be.

We also train The Cannon with the same training set for comparison and plot both the training MSE of SLAM and The Cannon in Figure 2. The black lines show the 16, 50, and 84 percentiles of the training spectra. The median one represents the "typical" spectrum of the training sample. In the upper three panels, the red and gray lines show the CV MSE of SLAM and The Cannon, respectively, while in the lower three panels the red and gray lines denote the CV MD squared of SLAM and The Cannon, respectively. The two black dashed lines represent the median values of the red and gray lines. We found that SLAM can make the CV MSE much lower than that of The Cannon at almost all wavelengths. The reason is that the quadratic model adopted by The Cannon is insufficiently flexible to model spectra in such wide ranges of stellar labels like $4000 < T_{\rm eff}/{\rm K} < 8000$. The CV MD^2 of SLAM also shows many improvements compared to that of The Cannon.

3.3. Prediction

In the first row of Figure 3, we show the $T_{\rm eff}$–$\log g$ distribution of the training sample stars with $5 < S/N_g < 20$ in the first panel and the SLAM-predicted $T_{\rm eff}$ and $\log g$ in the second panel. In the third, fourth, and fifth panels, we show the diagonal plot of the $T_{\rm eff}$, $\log g$, and [Fe/H], respectively, to compare the estimates from SLAM with the originals of LAMOST. From the second to the last row, we show similar plots for stars with $20 < S/N_g < 40$, $40 < S/N_g < 60$,

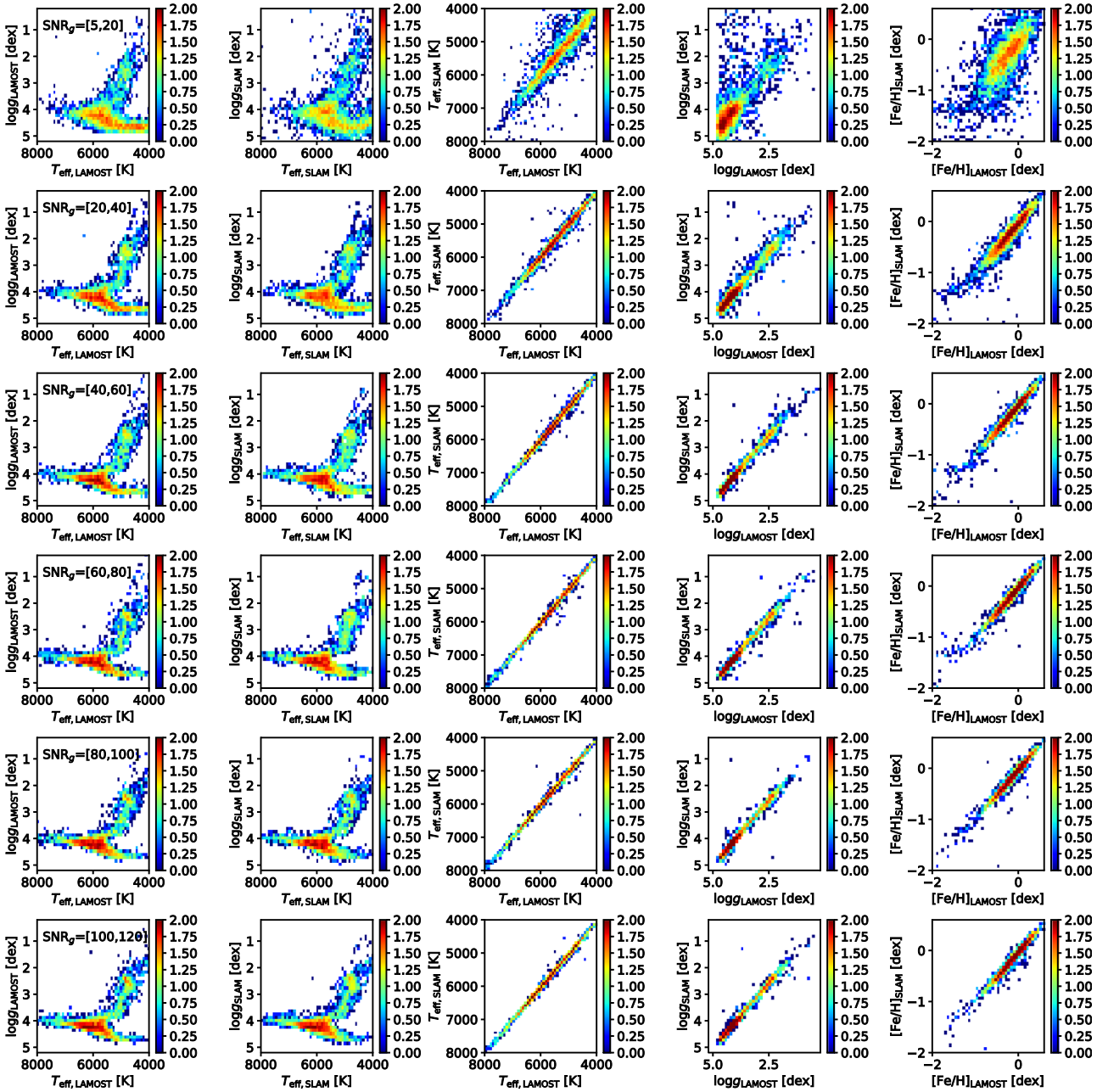

Figure 3. This figure shows the distributions of the predicted stellar labels at different ranges of S/N$_g$. The six rows from top to bottom correspond to six different S/N$_g$ intervals. In each row, the first panel shows the diagram of LAMOST DR5 $T_{\rm eff}$-log g which are regarded as the true values. The second panel shows the similar $T_{\rm eff}$ and log g diagram with values derived from SLAM. The third, fourth, and the last panels show the SLAM-derived stellar labels against the corresponding LAMOST values. In all panels, color indicates the sample counts on a logarithmic scale.

$60 < {\rm S/N}_g < 80$, $80 < {\rm S/N}_g < 100$, and $100 < {\rm S/N}_g < 120$, respectively. As S/N increases, the SLAM-predicted values become more and more consistent with the true values.

In Figure 4, we show the SLAM errors, CV scatter, and CV bias at various S/N$_g$. Note that the SLAM errors are very small compared to CV scatter. For stars with ${\rm S/N}_g > 100$, the SLAM errors for $T_{\rm eff}$, log g, and [Fe/H] are smaller than 10 K, 0.03 dex, and 0.02 dex, respectively. We also show the simulated error values for a solar-like star and a K giant star at different S/N$_g$ using the ATLAS9 synthetic spectra (Castelli & Kurucz 2003). Although the SLAM errors of the observed spectra are very small, they are much larger than the simulated values, which can be regarded as the lower limits of errors.

On the other hand, the CV scatters are larger than the SLAM errors. At the high S/N$_g$ end, the CV scatters of $T_{\rm eff}$, log g, and [Fe/H] are ∼50 K, 0.10 dex, and 0.07 dex, respectively. These values are very similar to the values reported in Ho et al. (2017a). However, it is worth noting that our sample is distributed in a wider range than that studied by Ho et al. (2017a), who only considered red giants with low effective temperatures. In general, the hot and warm stars may suffer from larger uncertainties of stellar parameter estimates than the cool stars (Liu et al. 2012).

The reason that the SLAM errors are substantially smaller than the CV scatters is because we assumed that both the spectral fluxes and the stellar labels in the training set are infinitely accurate. When we model fluxes as functions of

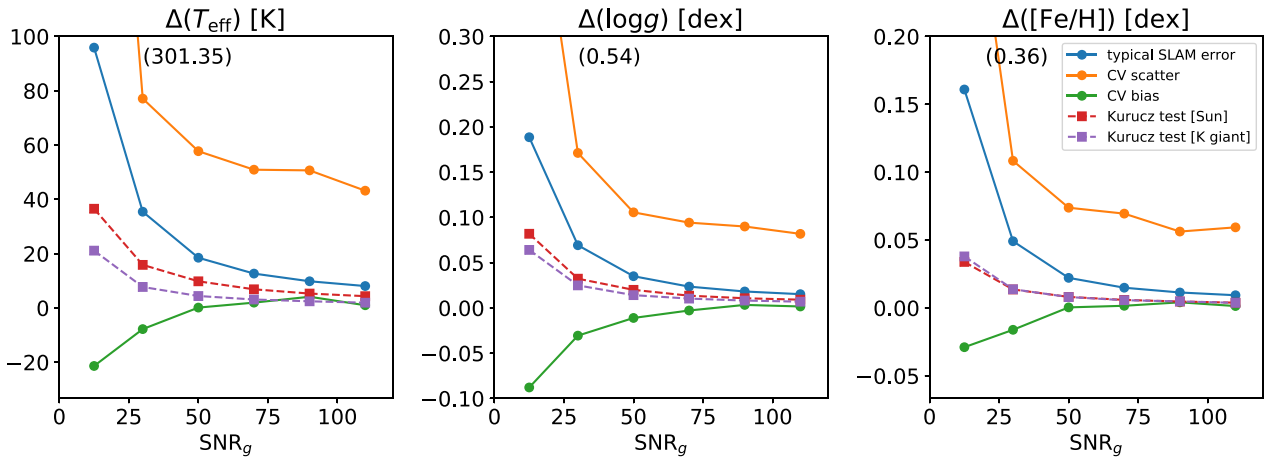

Figure 4. This figure shows how the errors of stellar labels change with S/N_g. In all panels, the blue curves represent the SLAM errors (formal errors). The red and purple curves represent the formal errors of tests for synthetic spectra of solar-like and K giant stars, respectively, selected from the Kurucz ATLAS9 model. The orange and green curves represent the CV scatter and bias, respectively. The first orange points at $S/N_g = 12.5$ (corresponding to the $5 < S/N_g < 20$ bin) are located beyond the figure, thus we mark the their values in brackets at the top of the panels. Clearly, all of them decrease as S/N_g increases. At $S/N_g > 100$, the typical CV scatters of $T_{\rm eff}$, $\log g$, and [Fe/H] are about 50K, 0.10 dex, and 0.07 dex, respectively.

stellar labels, the observed fluxes of the training stars are composed of noise, i.e., $\boldsymbol{f}_{\rm obs} = \boldsymbol{f}(\boldsymbol{\theta}) + \epsilon$, where ϵ denotes noise. Meanwhile, the errors in stellar labels are also not taken into account in the model. If we train our model with different training samples, the predicted stellar labels would be different due to the different errors implied in the training set. This difference should be larger than the SLAM error which is internal.

Another reason is that the errors of stellar labels in the validation sample also exist. This can increase the CV scatter to some extent. For instance, if the stellar labels of the validating sample have errors of 30 K in $T_{\rm eff}$, it is impossible to decrease CV scatter to under 30 K.

Therefore, to assess the performance of a data-drive method, CV scatter is the fair quantity rather than the SLAM error (or the internal error of the method), since the former has taken into account the uncertainties contributed by the training set.

3.4. A Comparison to the LASP

SLAM is different from LASP (Wu et al. 2011, 2014) in many different aspects. Since LASP, which uses Ulyss (Koleva et al. 2009) to predict stellar labels, builds a polynomial model of spectral flux on the ELODIE spectral library (Prugniel et al. 2007), SLAM offers several advantages over it. The first is that SLAM offers more flexibility and adaptive model complexity taking advantage of the RBF kernel. Second, we made SLAM open source and users can choose whatever they want as the training set rather than sticking to ELODIE. The third is that SLAM can generally provide uncertainty estimates of stellar labels by applying the relationship between the CV scatters and S/N ratios. Last, but the most important difference, is that SLAM is able to extend to more stellar labels, e.g., [α/Fe] and other element abundances, which is impossible with LASP currently.

4. Predict Stellar Labels for LAMOST Spectra Based on APOGEE DR15

4.1. The APOGEE Survey

The APOGEE survey provides high-resolution ($R \sim 22{,}500$) H-band (15200–16800 Å) spectra (Majewski et al. 2017). APOGEE DR15 comprises >270,000 high S/N spectra. Its pipeline, ASPCAP (García Pérez et al. 2016), produces estimates of the basic stellar labels, abundances, and microturbulence. In this section, we use the APOGEE DR15 (Holtzman et al. 2018) stellar labels in the training set to set up the SLAM model and predict stellar labels for the LAMOST DR5 low-resolution spectra.

4.2. Training and Test Set

We first select our training set from the 86,552 common stars between APOGEE DR15 and LAMOST DR5 by adopting the following criteria:

1. the S/N of the APOGEE spectra $S/N_{\rm APOGEE} > 100$,
2. $S/N_g > 40$ for LAMOST spectra,
3. the ASPCAP stellar label flag `ASPCAPFLAG` $= 0$,
4. the ASPCAP effective temperature $3000 < T_{\rm eff,APOGEE}/{\rm K} < 5500$,
5. the ASPCAP surface gravity $-1 < \log g_{\rm APOGEE}/{\rm dex} < 5$,
6. the ASPCAP overall metallicity $-4.0 < {\rm [M/H]/dex} < 2.0$,
7. the ASPCAP α-elemental abundance $-1.0 < [\alpha/{\rm M}]/{\rm dex} < 1.0$,
8. the ASPCAP carbon abundance $-0.4 < {\rm [C/Fe]/dex} < 1.0$,
9. the ASPCAP nitrogen abundance $-0.5 < {\rm [N/Fe]/dex} < 1.0$, and
10. the difference of the corresponding LAMOST and APOGEE effective temperature $|T_{\rm eff,APOGEE} - T_{\rm eff,LAMOST}|/{\rm K} < 800$.

The purpose of the last criterion is to give a loose condition on the consistency between the stellar labels provided by LAMOST and APOGEE, so that the selected stars have

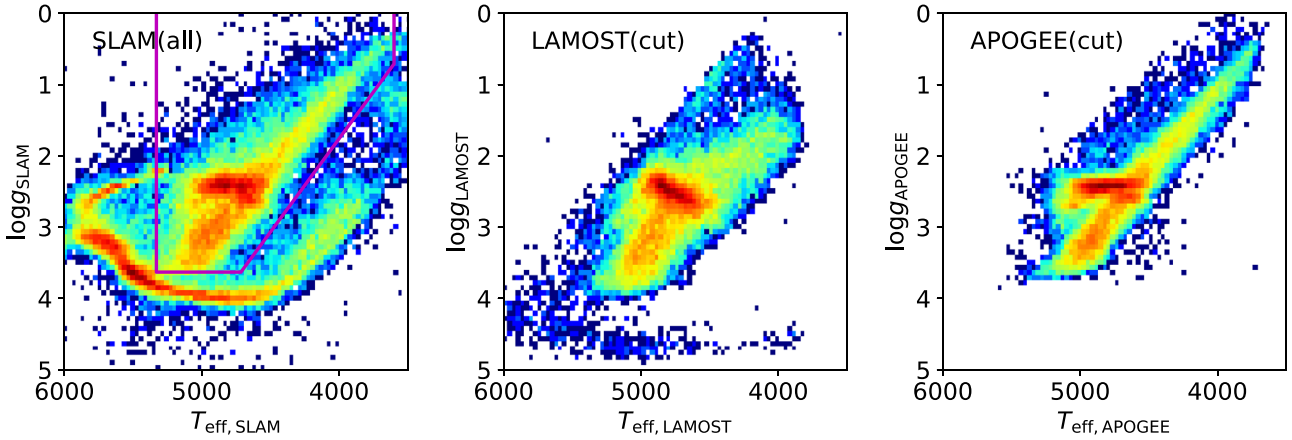

Figure 5. Left panel shows the distribution of SLAM-predicted $T_{\rm eff}$ and $\log g$ of all converged LAMOST DR5 stars. The pink solid polygon represents the selected area for K giant stars. The middle/right panel shows LAMOST/APOGEE $T_{\rm eff}$-$\log g$ diagrams for the sample located in the pink polygon.

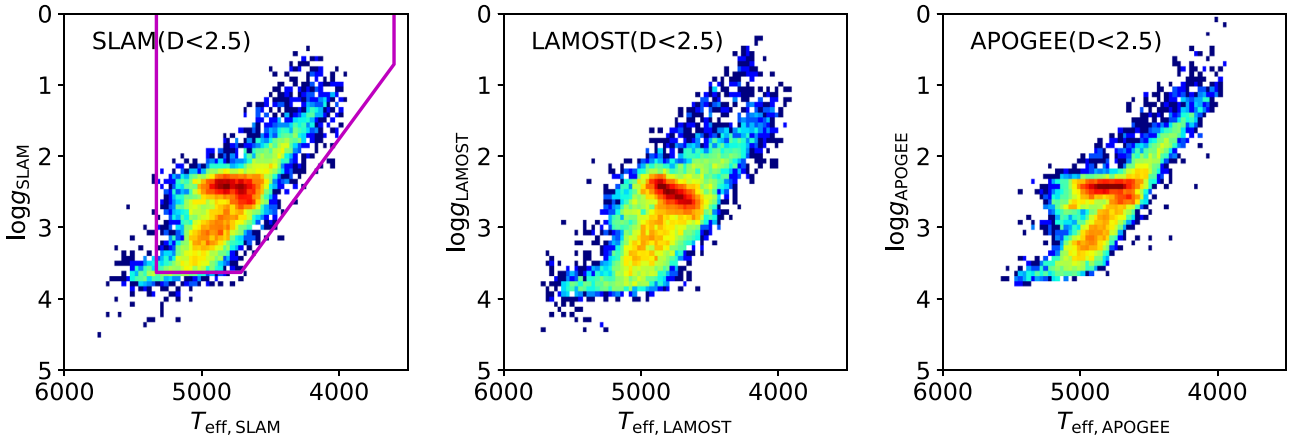

Figure 6. Left panel shows the distribution of SLAM-predicted $T_{\rm eff}$ and $\log g$ of LAMOST DR5 stars with $D < 2.5$. The pink solid polygon is the same as that in Figure 5. The middle/right panel shows the LAMOST/APOGEE $T_{\rm eff}$ and $\log g$ for the same samples.

reliable stellar label values. With these criteria, we obtain 17,703 common stars with reliable stellar labels.

Then we exclude the LAMOST spectra containing more than 50 bad pixels and obtain 17,623 stars. The grid of hyperparameters C and γ is set to be uniform in logarithmic scale, i.e., $C = 10^{[0.,0.5,1.,1.5,2.]}$ and $\gamma = 10^{[-3.,-2.5,-2.,-1.5,-1.]}$, while ϵ is fixed at 0.05. We use a leave-1/10-out training process to exclude the stars whose stellar labels deviate from the training set values by more than four times the standard deviation stellar label residuals in any dimension. Finally, our training sample contains 17,175 stars and then we train SLAM on this data set. We use eight-fold cross-validation to find the best-fit hyperparameters and conduct the training process. Then we apply the tuned SLAM model to all 8171,443 stars (`class = STAR` in LAMOST catalog) in LAMOST DR5. SLAM successfully converges for 5,378,550 stars.

In the LAMOST–APOGEE common samples (86,552), SLAM converged for 57,703 of them and derived their stellar labels. In the left panel of Figure 5, we show their distribution in the $T_{\rm eff}$–$\log g$ plane. The SLAM-predicted stellar labels have a red giant branch and a stripe, which looks like a distorted main sequence on which most stars do not have APOGEE stellar labels. The stellar labels of the objects located in the stripe are unreliable because the stellar labels are too far away from the stellar label range of our training set. We cannot apply the simple parameter cut described in Liu et al. (2014) because the stripe turns upward at $T_{\rm eff} < 4500$ K. Therefore, we empirically set up a polygon (shown in pink in the figure) for the selection. The sample stars located in the pink solid polygon are selected as the K giant stars with reliable stellar labels. We show the corresponding LAMOST (APOGEE) stellar labels of the samples located in the polygon in the middle (right) panel.

To assess the completeness of our cut, we select stars with criteria similar to those listed at the very beginning of this subsection but ignore the constraints depending on LAMOST stellar labels. Then we have ∼22,000 stars with good APOGEE stellar labels left. We check whether these known K giant stars are selected by the empirical polygon and find that the polygon cut loses ∼500 K giants, which gives completeness of about 97% for the K giant stars.

We also calculated the "label-distance" D defined by Ho et al. (2017b), i.e.,

$$\begin{aligned} D = &(T_{\rm eff,SLAM} - T_{\rm eff,LAMOST})^2/(100\ {\rm K})^2 \\ &+ (\log g_{\rm SLAM} - \log g_{\rm LAMOST})^2/(0.2\ {\rm dex})^2 \\ &+ ([{\rm M/H}]_{\rm SLAM} - [{\rm Fe/H}]_{\rm LAMOST})^2/(0.1\ {\rm dex})^2. \end{aligned} \quad (11)$$

To be consistent with Ho et al. (2017b), we consider stars with $D < 2.5$ as K giant stars. In the left, middle, and right panels of Figure 6, we show plots similar to Figure 5. The distribution of the SLAM-predicted stellar labels is quite similar to the sample

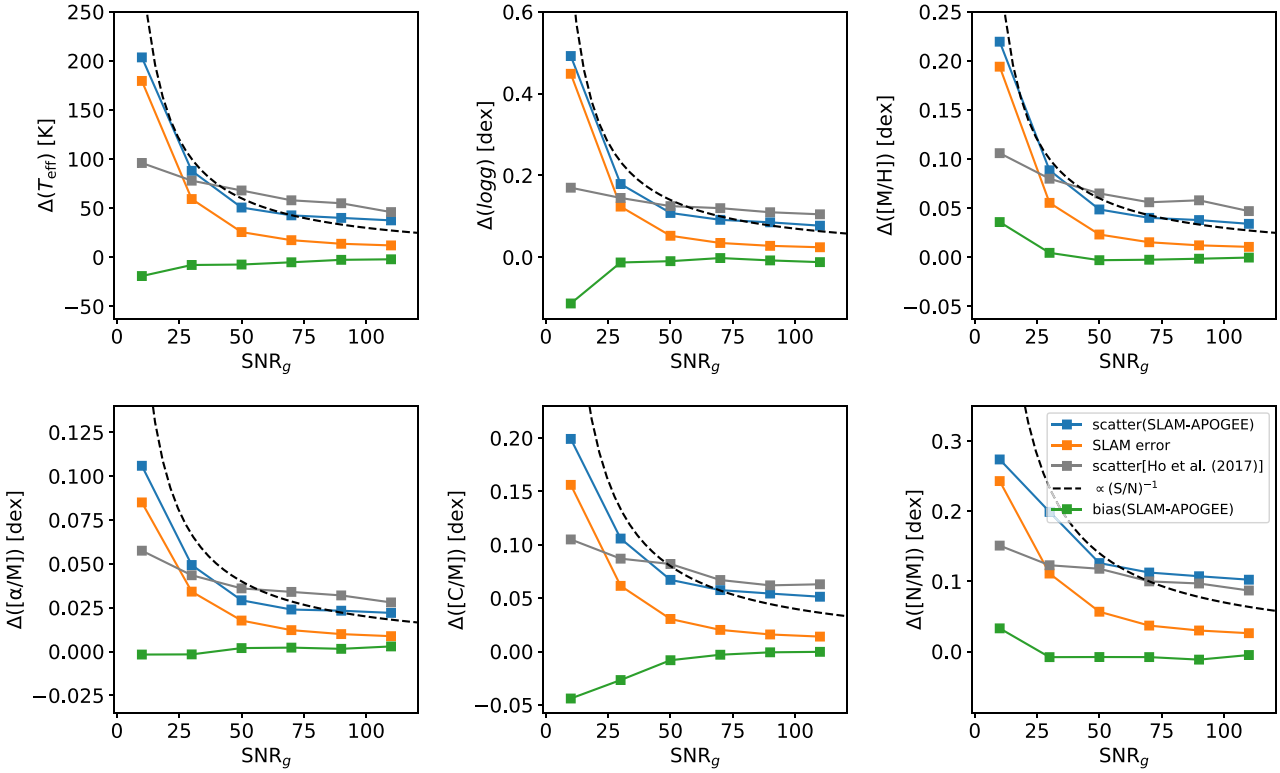

Figure 7. This figure shows the comparison of the CV scatters of stellar labels between SLAM(blue) and The Cannon (gray). SLAM errors and bias are also shown with green and orange lines, respectively. The inverse signal-to-noise ratio trends are also superposed with a black dashed line.

selected using the polygon cut. To select K giant samples with reliable stellar labels, we suggest using either the polygon cut, the "label-distance" method, or a combination of them.

4.3. Performance

Figure 7 shows the CV scatter of the SLAM-predicted stellar labels for the LAMOST–APOGEE common stars at different S/N intervals. As S/N$_g$ increases, the CV scatters decrease rapidly (shown by the blue line) as expected. At the high S/N$_g$ end, the CV scatters of estimated stellar labels are 49 K, 0.10 dex, 0.037 dex, 0.026 dex, 0.058 dex, and 0.106 dex for $T_{\rm eff}$, $\log g$, [M/H], [α/M], [C/M], and [N/M], respectively. These values are quite similar to those reported in Ting et al. (2017a). Compared to CV scatters, the biases are only as large as one-fourth of the scatters at most and thus do not contribute a lot in the total uncertainties. SLAM errors are again much smaller than the corresponding CV scatters.

We found that, although the CV scatters are smaller than that in Ho et al. (2017a) at the high S/N$_g$ end, they are much larger at the low S/N$_g$ end. One probable source of this difference is that in our result, only the 3900 to 5800 Å part of the LAMOST spectra are used. And we did not utilize information from photometry. Our CV scatters are more similar to the inverse S/N$_g$ trend, which is more realistic for a general test sample. According to the correlations between the CV scatters and S/N$_g$, we suggest that the carbon and nitrogen abundances derived by SLAM can only be used for stars with S/N$_g$ > 40. We also noticed that the bias is significant at the low S/N$_g$ end. We will try to overcome this problem in our future work.

In Figure 8, we show the diagonal plot of our stellar labels against corresponding APOGEE stellar labels for the subsample with S/N$_g$ > 100. It is seen that the SLAM-derived $T_{\rm eff}$, $\log g$, [M/H], [α/M] [C/M], and [N/M] agree well with the APOGEE values.

In Figure 9, we show the comparison of color-coded median [C/N] in the [M/H]–[α/M] plane. In the left column, the top and bottom panels show logarithmic counts and median [C/N] of SLAM results, respectively, in the [M/H]–[α/M] plane using the LAMOST–APOGEE common stars. All the stellar labels in the two panels are from SLAM. The middle column shows similar plots but using APOGEE stellar labels. In the right column, the two panels show the SLAM results for all LAMOST K giant stars with S/N$_g$ > 100. It is seen that the predicted [M/H], [α/M], and [C/N] are similar to the training data. Meanwhile, [C/N] is in the reasonable range at the high-density region.

In the final catalog, the output errors of stellar labels are approximated from S/N$_g$ using the empirical function, $a\exp(-b\times {\rm S/N}_g)+c$. The best-fit coefficients in the empirical functions are listed in Table 1. The whole catalog of the 5 million converged stars is hosted at the China-VO Paper Data website http://paperdata.china-vo.org/. An example of it is shown in Table 2.

5. Discussion

Although the performance of SLAM has been well illustrated in Sections 3 and 4, several challenges and issues, most of which quite commonly appear in data-driven approaches, are worth discussing here.

5.1. Preprocessing

In the preprocessing step, SLAM and other data-driven methods operate with RV-corrected and normalized spectra. Consequently, uncertainties in these processes propagate to the

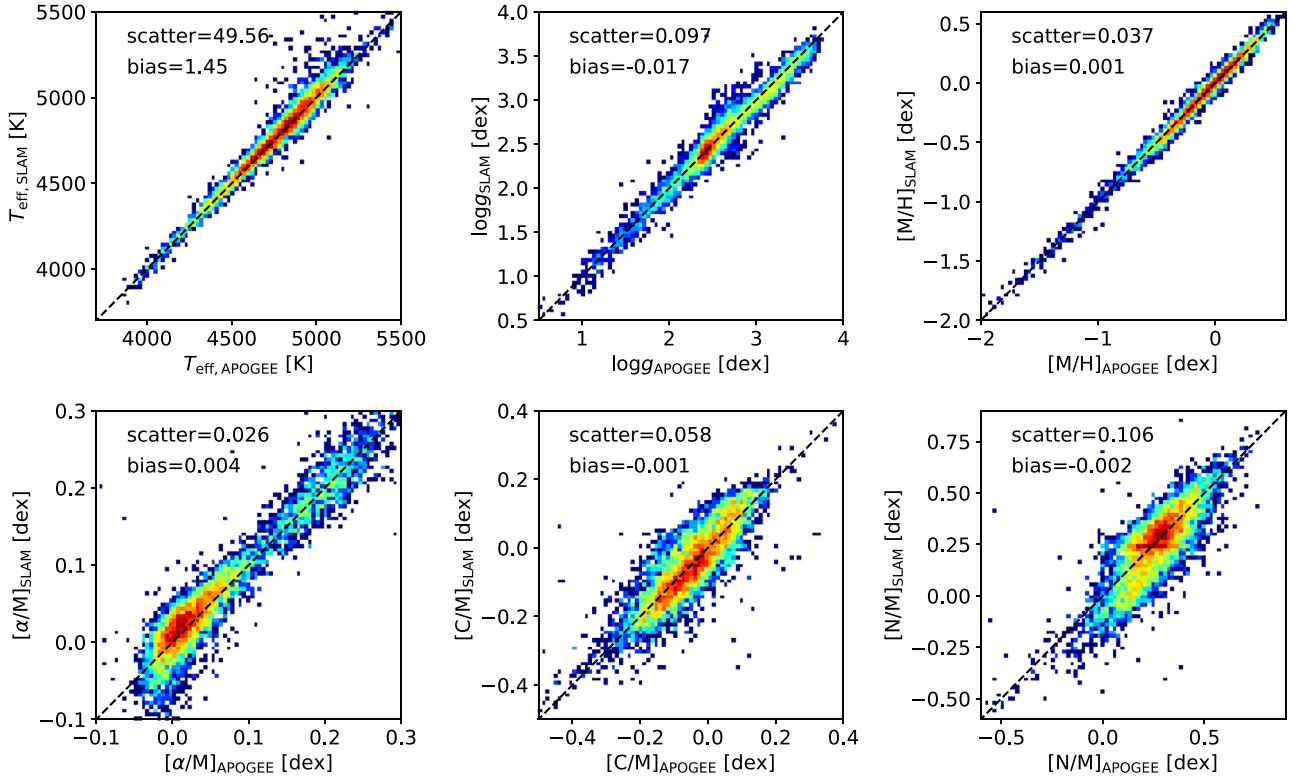

Figure 8. This figure shows the diagonal plots of the six stellar labels (effective temperature, surface gravity, metallicity, α-element abundance, carbon abundance, and nitrogen abundance) for the LAMOST–APOGEE common stars with $S/N_g > 100$.

final results. However, it is extremely difficult, if it is possible, to automatically determine the proper and consistent pseudo-continuum in the normalization process for various types of stars. Hence, the normalization process induces a certain amount of uncertainties in the normalized spectra, especially for the late-type stars ($T_{\rm eff} < 4500$ K).

In low-resolution spectra, the blending of spectral lines and molecular bands, such as the g band, also increases the uncertainties of the normalized spectra. Weak lines could be overwhelmed by the inconsistency of the normalization. In some cases, the inconsistent normalization may lead to the failure of the stellar label estimation in the data-driven methods.

Although for K giant stars, the normalization preprocessing in both SLAM and The Cannon seem adequate and may not affect the final performance, we should be cautious about this issue, especially when the normalization may induce a variation/deviation larger than the typical training precision. We refer the readers to Jofré et al. (2019) for a review of the various sources of uncertainties induced in transforming spectra to stellar labels including normalization.

5.2. Training

In the training step, the most important issue is the limited coverage of the parameter space of the training sample. This is also described in Section 5.5 in Ness et al. (2015). Therefore, the selection of a training set is crucial. Once some types of input stars are not included in the training set, the program will not derive meaningful stellar labels.

The second issue is the imbalance of the training sample. Usually, very few stars are located near the edge of the parameter space. For example, extremely hot/cool stars and very metal-rich/metal-pool stars are rare. Their spectra are very different from those of normal stars and thus play more important roles in the training process. These stars are anchors that define the edge of the parameter space. However, their small numbers may not effectively leverage the objective function compared to the majority of the normal stars.

The third issue arises in the flux model. The flux model of SLAM does not make use of the uncertainties of the stellar labels in the training set. This leads to the underestimation of uncertainties of both the spectra and stellar labels in the model. To take into account the stellar label errors in the training set, one possible solution is to cross-validate the training samples and get different models using different subsets of the training set, and derive the deviations of predictions using these models. However, so far it is difficult for us to conduct such a complicated training process due to the high computational expenses.

5.3. The Computational Cost of SLAM

Although SVR is a powerful tool, its computational cost and storage requirements increase rapidly with the number of training vectors (Pedregosa et al. 2012). The complexity of the problem solved with LIBSVM scales between $\mathcal{O}(n_{\rm features} \times n_{\rm samples}^2)$ and $\mathcal{O}(n_{\rm features} \times n_{\rm samples}^3)$ (Chang & Lin 2011), which means that adding more stars in the training set is more difficult than adding more stellar labels. A fiducial cost of SLAM is that in our experiment on transferring stellar labels from APOGEE DR15 to LAMOST DR5 in Section 4, the training takes about 1 day using an Intel Xeon CPU E5-2690v4 (2.60 GHz). The training cost is also proportional to the number of pixels and the size of the hyperparameter grid that is tried. For prediction, it takes <1 minutes to predict the six

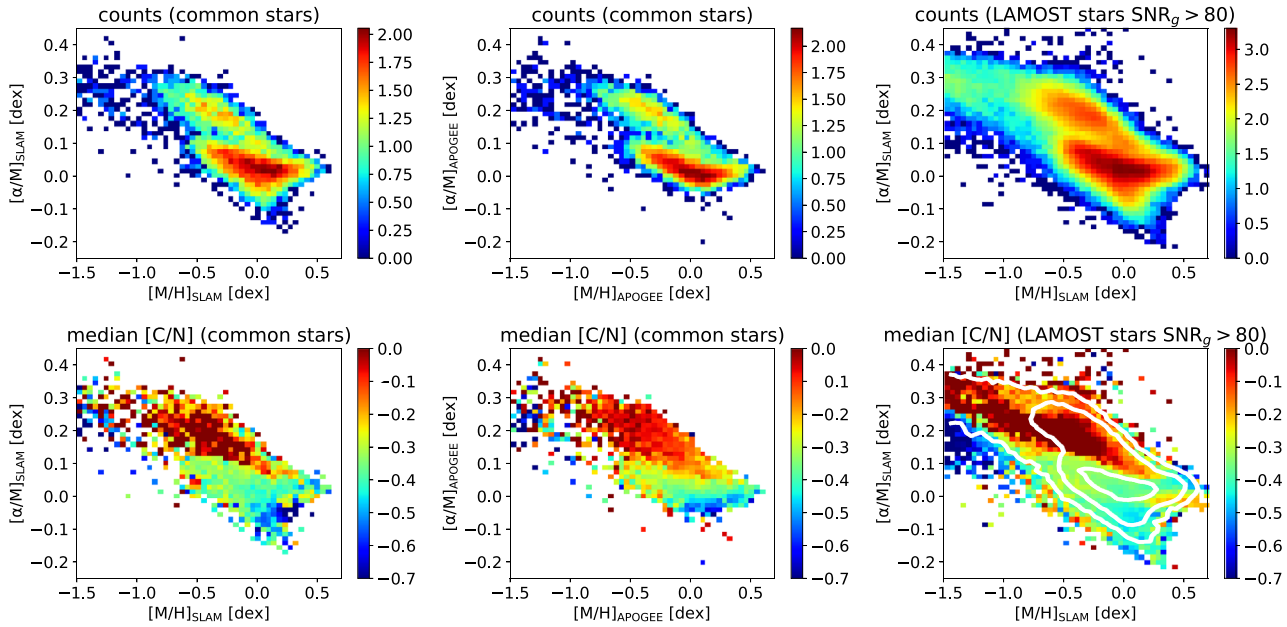

Figure 9. Top left panel shows the distribution of $[\alpha/M]$–$[M/H]$ plane for the SLAM-derived labels for the LAMOST–APOGEE common stars. The top middle panel shows a similar plot but with APOGEE parameters. The top right panel shows a plot similar to that in the top left panel for all the LAMOST K giant stars with $S/N_g > 80$. The bottom panels show the distributions of median $[C/N]$ in the $[\alpha/M]$–$[M/H]$ plane. Similar to the top panels, from left to right are the SLAM-derived stellar labels for the LAMOST–APOGEE common stars, the APOGEE labels for the common stars, and the SLAM-predicted labels for all LAMOST K giant stars with $S/N_g > 80$. In the bottom right panel, white contours of counts are superposed.

Table 1
The Fitting Coefficients a, b, and c for Each Stellar Label

Stellar Label	a	b	c
$T_{\rm eff}$/K	204.8	0.056	38.8
$\log g$/dex	0.592	0.063	0.069
[M/H]/dex	0.431	0.073	0.029
$[\alpha/M]$/dex	0.090	0.049	0.019
[C/M]/dex	0.152	0.043	0.040
[N/M]/dex	0.152	0.031	0.072

stellar labels for a spectrum with modest S/N. For those with a very low S/N ratio, it sometimes does not converge so that it typically takes much longer. Therefore, users should be cautious of the computational expense when using SLAM to derive a large number of stellar labels.

6. Learning from Data: Coefficients of Dependences

In this section, we present the CODs, which enable us to better understand why machine-learning methods generally agree with our experience in traditional spectroscopy.

As described in Section 2, the worst regression model, i.e., the constant model, has MSE = 1 in standardized space. Any better model should reduce the MSE of this pixel to a value far below 1. We denote this MSE as $\rm MSE_{full}$. Then, $1 - \rm MSE_{full}$ can be considered as a proper measure of the fraction of the variation of the pixel being explained by the model. We define $1 - \rm MSE_{full}$ as the full coefficient of dependency (the full COD) of stellar labels, i.e.,

$$\rm COD_{full} = 1 - MSE_{full}. \quad (12)$$

The maximum and minimum values of $\rm COD_{full}$ are 1 and 0, respectively. The larger $\rm COD_{full}$ is, the better the model is.

Let $\boldsymbol{L}$ denote the collection of stellar labels ($T_{\rm eff}$, $\log g$, and etc.), and let l denote one specific stellar label in $\boldsymbol{L}$. To derive the contribution of each stellar label in $\rm COD_{full}$, we did a Leave-One-Label-Out training. For example, to quantify the contribution of l, we remove l from $\boldsymbol{L}$ and train SVR on the other stellar labels. We write the obtained MSE in this case as $\rm MSE_{\it l}$. In principle, $\rm MSE_{\it l}$ is equal to or larger than $\rm MSE_{full}$ because the model ignores the variation of the spectra driven by the stellar label l. The difference, $\rm MSE_{\it l} - MSE_{full}$, measures the loss due to excluding stellar label l in the model. We then define $\rm COD_{\it l}$ as

$$\rm COD(\it l) = COD_{full} \times \frac{MSE_{\it l} - MSE_{full}}{\sum_{\it l \in \boldsymbol{L}} (MSE_{\it l} - MSE_{full})}. \quad (13)$$

By definition, $\sum_l \rm COD(\it l) = COD_{full}$.

We derive the CODs of $T_{\rm eff}$, $\log g$, and [Fe/H] for the training samples used in Section 3 (covering $T_{\rm eff}$ from 4000 to 8000 K) and show them in Figure 10. For most of the spectra, the COD spectra are amazingly consistent with the empirical knowledge about which spectral lines are sensitive to which stellar labels. The blue, orange, and green filled regions represent the COD ($T_{\rm eff}$), COD ($\log g$), and COD ([Fe/H]), respectively. The most significant features are around the Balmer lines. At Hδ, Hγ, and Hβ, COD ($T_{\rm eff}$) is very large and dominant, while COD ([Fe/H]) and COD ($\log g$) are small, meaning that these pixels depend mainly on $T_{\rm eff}$ rather than $\log g$ and [Fe/H]. Across the whole spectrum, the Balmer lines are the most prominent features sensitive to effective temperature. The line centers of Balmer lines appear to be slightly different from the line wings, which reflect different mechanism information of line centers.

From Figure 10, it seems that the most information regarding $\log g$ (for K giant stars) comes from the ~4200 Å region and the Mg I triplet at around λ5175. We can find that most of the dependence on $\log g$ comes from the doublets, triplets, and line wings. The pixels located at the wings of the three lines of the Mg I triplet show high dependence on $\log g$. This behavior is

Table 2
An Example of the Catalog of the 5 Million Converged LAMOST DR5 K Giant Stars with SLAM-derived Stellar Labels

LAMOST obsid	S/N_g	R.A.	Decl.	$T_{\rm eff}$	$\log g$	[M/H]	[α/M]	[C/M]	[N/M]
		(deg)	(deg)	(K)	(dex)	(dex)	(dex)	(dex)	(dex)
101008	332.36874	−1.95577	20	5262.5	3.30	0.0350	0.2063	−0.0615	0.2708
101009	332.20666	−1.86865	12	5672.1	4.05	−0.3786	0.1140	−0.1829	0.2263
101016	332.34873	−2.13610	31	4571.9	3.89	0.2541	0.0874	−0.1105	0.1043
101017	332.44442	−1.92405	23	5277.8	3.12	−0.6337	0.2501	−0.0415	0.4757
101020	332.22238	−1.98763	23	5210.4	3.03	−1.6938	0.3061	−0.0436	−0.3604
101021	332.35138	−1.80933	42	5469.3	2.75	−1.6935	0.3750	0.1408	−0.5156
101023	332.50637	−2.01690	17	5229.9	3.82	−0.4062	0.0468	−0.2554	−0.0016
101026	331.55123	−1.68436	38	5380.3	3.70	0.2258	−0.0303	−0.1946	0.2326
101027	331.76831	−1.91449	42	5577.1	3.54	−0.8770	0.2438	0.0322	−0.0597
101029	331.77234	−1.83052	47	5572.1	3.60	−0.7497	0.2368	0.0299	−0.0473

$\sigma(T_{\rm eff})$	$\sigma(\log g)$	σ([M/H])	σ([α/M])	σ([C/M])	σ([N/M])	rmse	Kgiant(cut)	D(Ho2017)	in APOGEE
(K)	(dex)	(dex)	(dex)	(dex)	(dex)		bool		bool
102.9	0.23	0.1245	0.0516	0.1029	0.1516	0.068	True	39.26	False
138.4	0.33	0.1971	0.0668	0.1281	0.1740	0.077	False	1.43	False
72.8	0.15	0.0714	0.0377	0.0788	0.1278	0.051	False	41.81	False
91.5	0.20	0.1034	0.0465	0.0942	0.1434	0.050	True	47.49	False
93.8	0.21	0.1076	0.0476	0.0960	0.1451	0.048	True	99.94	False
57.2	0.11	0.0481	0.0299	0.0643	0.1114	0.029	False	130.25	False
115.8	0.27	0.1500	0.0573	0.1124	0.1603	0.067	False	30.37	False
62.7	0.12	0.0559	0.0327	0.0697	0.1177	0.031	False	7.10	False
57.8	0.11	0.0491	0.0302	0.0650	0.1122	0.023	False	9.77	False
53.0	0.10	0.0427	0.0276	0.0600	0.1060	0.025	False	8.44	False

Note. Column 1 is the LAMOST IDs of the objects, column 2 is the g-band signal-to-noise ratio of the spectra, columns 3–4 are the sky coordinates of the objects, columns 5–10 are the SLAM-predicted stellar labels and column 11–16 are the corresponding errors, column 17 is the rms deviation between the observed and fitted spectra, column 18 is the index of our selection of K giant stars, column 19 is the label distances, and Column 20 is the APOGEE observation flag (observed by APOGEE if true).
(This table is available in its entirety in FITS format.)

largely different compared to the COD ($T_{\rm eff}$) and the COD ($\log g$).

The COD([Fe/H]) is largely coincident with the positions of metal lines such as the Fe λ5709 and Fe λ5782. In our experience, the Ca II K and H lines are good proxies of metallicity. However, because of the inverse variance of the LAMOST spectra in the very blue part of the spectrum ($\lambda \sim 3936$ Å and 3970 Å) are frequently marked as bad pixels, many of the Ca II H and K lines are unavailable. Therefore, the COD([Fe/H]) does not show strong dependence at the Ca II H and K lines.

The picture gives us a good interpretation of how machine-learning algorithms learn from the data and help humans understand the data. Although the CODs are very similar to the gradient $\left(\frac{\partial f}{\partial l}\right)$ which is also shown in other works such as Ness et al. (2015), they are different. The gradient is essentially the first-order partial derivative, so it reflects the local dependence of the fluxes on stellar labels only in the first-order. In contrast, the CODs measure the global dependence and do not rely on the specific analytic models to map the stellar labels to the spectral fluxes.

7. Conclusions

Following the idea of the data-driven methods, we present the SLAM, an SVR-base method, in this work. Taking advantage of the nonparametric nature of SVR, SLAM is able to fit a multidimensional and highly nonlinear relationship between the fluxes and stellar labels, which is very different from The Cannon.

We validate our method with LAMOST DR5 to investigate the performance and precision of the predicted labels. The cross-validated scatters of $T_{\rm eff}$, $\log g$, and [Fe/H] at high S/N_g ($\sim$100) are 50 K, 0.09 dex, and 0.07 dex, respectively.

We also use our method to predict stellar labels of LAMOST DR5 K giant stars with the training labels from APOGEE DR15. The performance assessment indicates that SLAM is moderately better than The Cannon. The cross-validated scatters at high S/N_g end are 49 K, 0.10 dex, 0.037 dex, 0.026 dex, 0.058 dex, and 0.106 dex for $T_{\rm eff}$, $\log g$, [M/H], [α/M], [C/M], and [N/M], respectively. We provide a downloadable catalog composed of SLAM-derived $T_{\rm eff}$, $\log g$, [M/H], [α/M], [C/M], and [N/M] for more than a million LAMOST K giant stars.

The authors thank the referee for kindly providing many constructive comments.

This work is supported by National Key R&D Program of China No. 2019YFA0405501. C.L. is thankful for the National Natural Science Foundation of China (NSFC) grant No. 11835057.

Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated

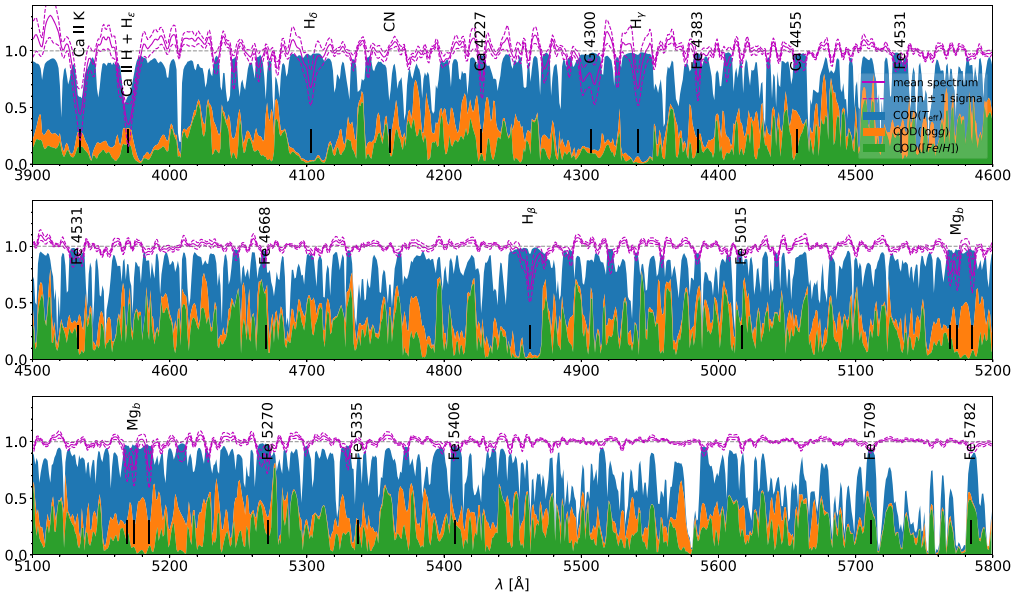

Figure 10. CODs from the training set with LAMOST spectra and stellar labels. The pink solid and dashed lines are the 50, 16, and 84 percentiles of normalized spectral fluxes. The blue, orange, and green filled regions represent COD($T_{\rm eff}$), COD($\log g$), and COD([Fe/H]), respectively.

and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

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, 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.

Software: scikit-learn (Pedregosa et al. 2012), astropy (Astropy Collaboration et al. 2013, 2018), IPython (Perez & Granger 2007), Scipy (Virtanen et al.2019).

Appendix A
How to Choose the Best Hyperparameters

Generally, there are two kinds of pixels, i.e., spectral line pixels and continuum pixels. The former kind contains much information on stellar labels while the latter contains almost no information. In this section, we show how the best hyperparameters are chosen in these two cases.

We simulate the first case in the upper row of Figure 11. We use $\mathcal{N}(\mu, \sigma^2)$ to denote the normal distribution with a mean of μ and variance of σ^2. The x data follows $\mathcal{N}(0, 1)$ and $y = \sin 2x +$ noise, where the noise here follows $\mathcal{N}(0, 0.16)$. In the first three panels of the upper row, we show how the fitting performance changes when varying one hyperparameter, i.e., C, ϵ, and γ, respectively. We also superposed a quadratic model in the gray dashed–dotted line. We can infer from these three panels that C (the penalty level) and γ (the width of the Gaussian kernel or the softness of the SVR model) are more important relative to ϵ (the tube radius). In the fourth panel, we show the color-coded tenfold CV MSE as a function of C and γ. At $\log\gamma \sim 0$ and $0 < \log C < 4$, the CV MSE reaches the minimum. At $\log\gamma \sim 4$ and $\log C \sim 4$, where the SVR has a high penalty for outliers and is extremely soft, the CV MSE is even larger than 1, which means overfitting occurs. Clearly, we are able to determine the best set of hyperparameters by choosing the one with the lowest CV MSE in this diagram.

To simulate the latter case, we make both x and y follow $\mathcal{N}(0, 1)$. In the lower row of Figure 11, we show similar plots. In this case, it is seen in the last panel that the CV MSE is around 1 for most of the combinations of hyperparameters. And again overfitting arises at large C and large γ. SLAM chooses the set of hyperparameters with the lowest CV MSE, which prevents the model from overfitting.

Appendix B
The Source Code of SLAM

The source code of SLAM is available on GitHub https://github.com/hypergravity/astroslam under an MIT License and the latest version is archived in Zenodo (Zhang 2019). It can

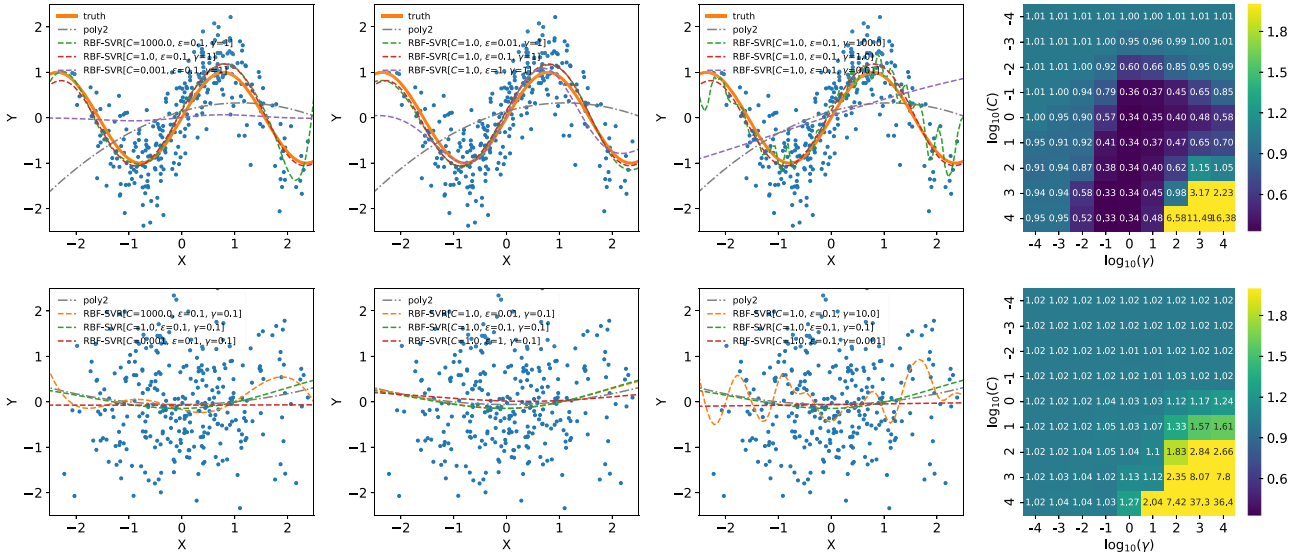

Figure 11. Two examples of how the best hyperparameters are chosen for SVR.

also be directly installed by running the following command in a terminal, `pip install astroslam`.

ORCID iDs

Bo Zhang https://orcid.org/0000-0002-6434-7201
Chao Liu https://orcid.org/0000-0002-1802-6917
Li-Cai Deng https://orcid.org/0000-0001-9073-9914

References

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Beers, T. C., Lee, Y., Sivarani, T., et al. 2006, MmSAI, 77, 1171
Bu, Y., & Pan, J. 2015, MNRAS, 447, 256
Buder, S., Asplund, M., Duong, L., et al. 2018, MNRAS, 478, 4513
Casey, A. R., Hogg, D. W., Ness, M., et al. 2016, arXiv:1603.03040
Casey, A. R., Hawkins, K., Hogg, D. W., et al. 2017, ApJ, 840, 59
Castelli, F., & Kurucz, R. L. 2003, in IAU Symp. 210, Modelling of Stellar Atmospheres, ed. N. Piskunov, W. W. Weiss, & D. F. Gray (Cambridge: Cambridge Univ. Press), A20
Chang, C.-C., & Lin, C.-J. 2011, ITMTT, 2, 1
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197
de Boor, C. 1978, Applied Mathematical Sciences (New York: Springer)
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RAA, 12, 735
Freeman, K. C. 2012, in ASP Conf. Proc. 458, Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way, ed. W. Aoki et al. (San Francisco, CA: ASP), 393
García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, AJ, 151, 144
Gilmore, G., Randich, S., Asplund, M., et al. 2012, Msngr, 147, 25
Ho, A. Y. Q., Rix, H.-W., Ness, M. K., et al. 2017a, ApJ, 841, 40
Ho, A. Y. Q., Ness, M., Hogg, D. W., et al. 2017b, ApJ, 836, 5
Holtzman, J. A., Hasselquist, S., Shetrone, M., et al. 2018, AJ, 156, 125
Husser, T.-O., Wende-von Berg, S., Dreizler, S., et al. 2013, A&A, 553, A6
Jofré, P., Heiter, U., & Soubiran, C. 2019, ARA&A, 57, 571
Koleva, M., Prugniel, P., Bouchard, A., et al. 2009, A&A, 501, 1269
Li, X., Wu, Q. M. J., Luo, A., et al. 2014, ApJ, 790, 105
Liu, C., Bailer-Jones, C. A. L., Sordo, R., et al. 2012, MNRAS, 426, 2463
Liu, C., Deng, L.-C., Carlin, J. L., et al. 2014b, ApJ, 790, 110
Liu, C., Fang, M., Wu, Y., et al. 2015b, ApJ, 807, 4
Liu, C.-X., Zhang, P.-A., & Lu, Y. 2014b, RAA, 14, 423
Liu, X.-W., Zhao, G., & Hou, J.-L. 2015b, RAA, 15, 1089
Liu, X.-W., Yuan, H.-B., Huo, Z.-Y., et al. 2014c, in IAU Symp. 298, Setting the scene for Gaia and LAMOST, ed. S. Feltzing (Cambridge: Cambridge Univ. Press), 310
Liu, Z., Cui, W., Liu, C., et al. 2019, ApJS, 241, 32
Lu, Y., & Li, X. 2015, MNRAS, 452, 1394
Majewski, S. R. 2012, AAS, 219, 205.06
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Moré, J. J. 1978, LNM, 630, 105
Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2012, arXiv:1201.0490
Perez, F., & Granger, B. E. 2007, CSE, 9, 21
Prugniel, P., Soubiran, C., Koleva, M., et al. 2007, arXiv:astro-ph/0703658
Rix, H.-W., Ting, Y.-S., Conroy, C., et al. 2016, ApJ, 826, L25
Smola, A. J., & Schölkopf, B. 2004, Stat. Comput., 14, 199
Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645
Ting, Y.-S., Conroy, C., Rix, H.-W., et al. 2019, ApJ, 879, 69
Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2017a, ApJ, 843, 32
Ting, Y.-S., Rix, H.-W., Conroy, C., Ho, A. Y. Q., & Lin, J. 2017b, ApJ, 849, L9
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2019, arXiv:1907.10121
Wu, Y., Du, B., Luo, A., Zhao, Y., & Yuan, H. 2014, in IAU Symp. 306, Statistical Challenges in 21st Century Cosmology, ed. A. Heavens, J.-L. Starck, & A. Krone-Martins (Cambridge: Cambridge Univ. Press), 340
Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, RAA, 11, 924
Zhang, B. 2019, SLAM: Stellar LAbel Machine v1.2019.1005.0, Zenodo, doi:10.5281/zenodo.3461503
Zhong, J., Li, J., Carlin, J. L., et al. 2019, ApJS, 244, 8