Research in Astronomy and Astrophysics, 23:055013 (12pp), 2023 May https://doi.org/10.1088/1674-4527/acc3d4

The Stellar Spectra Factory (SSF) Based on SLAM

Wei Ji[1,2], Chao Liu[1,2,3], and Bo Zhang[1]
[1] Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, CAS, Beijing 100101, China; jiwei@nao.cas.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Institute for Frontier in Astronomy and Astrophysics, Beijing Normal University, Beijing 100875, China
Received 2023 January 10; revised 2023 February 28; accepted 2023 March 3; published 2023 April 26

Abstract

An empirical stellar spectral library with large coverage of stellar parameters is essential for stellar population synthesis and studies of stellar evolution. In this work, we present Stellar Spectra Factory (SSF), a tool to generate empirical-based stellar spectra from arbitrary stellar atmospheric parameters. The relative flux-calibrated empirical spectra can be predicted by SSF given arbitrary effective temperature, surface gravity, and metallicity. SSF constructs the interpolation approach based on the Stellar LAbel Machine, using ATLAS-A library, which contains spectra covering from O type to M type, as the training data set. SSF is composed of four data-driven sub-models to predict empirical stellar spectra. Sub-model SSF-N can generate spectra from A to K type and some M giant stars, covering $3700 < T_{\rm eff} < 8700$ K, $0 < \log g <$ dex, and $-1.5 < [\rm M/H] < 0.5$ dex. Sub-model SSF-gM is mainly used to predict M giant spectra with $3520 < T_{\rm eff} < 4000$ K and $-1.5 < [\rm M/H] < 0.4$ dex. Sub-model SSF-dM is for generating M dwarf spectra with $3295 < T_{\rm eff} < 4040$ K, $-1.0 < [\rm M/H] < 0.1$ dex. Sub-model SSF-B can predict B-type spectra with $9000 < T_{\rm eff} < 24{,}000$ K and $-5.2 < M_G < 1.5$ mag. The accuracy of the predicted spectra is validated by comparing the flux of predicted spectra to those with same stellar parameters selected from the known spectral libraries, MILES and MaStar. The averaged difference of flux over optical wavelength between the predicted spectra and the corresponding ones in MILES and MaStar is less than 5%. More verification is conducted between the magnitudes calculated from the integration of the predicted spectra and the observations in PS1 and APASS bands with the same stellar parameters. No significant systematic difference is found between the predicted spectra and the photometric observations. The uncertainty is 0.08 mag in the *r* band for SSF-gM when comparing with the stars with the same stellar parameters selected from PS1. The uncertainty becomes 0.31 mag in the *i* band for SSF-dM when comparing with the stars with the same stellar parameters selected from APASS.

Key words: methods: statistical – techniques: spectroscopic – stars: atmospheres – stars: general

1. Introduction

An empirical spectral library with complete stellar parameter coverage and flux calibration is of great significance for stellar population synthesis (Bruzual & Charlot 2003; Vazdekis et al. 2010, 2012; Maraston & Strömbäck 2011), which can be used to model stellar populations of galaxies (Le Borgne et al. 2004; Prugniel et al. 2007). In addition, an empirical spectral library can provide reference for classifying stars and determining stellar atmospheric parameters (Prugniel & Soubiran 2001; Koleva et al. 2009). An automatic stellar parameterization and spectral classification tools, nowadays, are necessary for large spectroscopic surveys, such as LAMOST (Deng et al. 2012), SDSS/APOGEE (Majewski 2012), GALAH (Freeman 2012) etc.

Compared with theoretical spectral library, the spectra in an empirical library can better reflect the real information of stars so that it can also be used to constrain the theoretical stellar model. The empirical stellar spectral libraries are widely used, such as Pickles (Pickles 1985, 1998), ELODIE (Prugniel & Soubiran 2001), STELIB (Le Borgne et al. 2003), UVES-POP (Bagnulo et al. 2003), INDO-US (Valdes et al. 2004), MILES (Sánchez-Blázquez et al. 2006) etc. The main shortcomings for these libraries is the poor coverage of parameters or the limited wavelength range by the capacity of observation instruments.

In recent years, newly released experiential spectral libraries, e.g., Du et al. (2019), MaStar (Yan et al. 2019), and ATLAS (Ji et al. 2023), are based on large spectroscopic surveys and hence cover broader parameter space and more stellar types. Compared with Du et al. (2019) and MaStar, ATLAS covers larger parameter space and spectral types. In addition, the spectra of ATLAS calibrate in absolute manner, while spectra of MaStar calibrate the absolute fluxes that are normalized to 5450 Å.

Other than the coverage of parameter space, all current stellar libraries provide discrete stellar spectra with more or less randomly distributed stellar parameters (as a comparison, the synthetic libraries usually align their stellar parameters with a regular parameter grid). To establish a synthetic stellar population, one always needs to develop an approach to interpolate spectra from the grid of data. As an empirical spectral library, a good interpolation may play important role in the population synthesis or in the stellar parameterization.

On the other hand, data-driven approaches for stellar parameterization from stellar spectra was well developed in these years. The Cannon (Ness et al. 2015), Stellar LAbel Machine (SLAM) (Zhang et al. 2020) etc. provide a forward modeling framework based on empirical stellar spectra. Such technique can take the place of the conventional interpolation to provide an alternative but effective way to obtain a spectrum with arbitrarily given stellar parameters. In this work, we apply the technique of forward-modeling developed by SLAM to provide a stellar spectra factory (SSF), that is, to produce an empirical stellar spectrum given an arbitrary set of stellar parameters as input. The SSF can be used as a software package for many potential usage, such as stellar population synthesis, stellar classification, and parameterization.

This paper is organized as follows. We give a brief description of SLAM in Section 2, and we briefly describe the four training sets for SLAM to derive the four sub-models of SSF in Section 3. In Section 4, we assess the performance of the four sub-models in details. We list some caveats when using SSF in Section 4.2. Finally, we draw conclusions of this paper in Section 5.

2. Method

The key of SSF is to set up a *function* that maps a set of arbitrary stellar parameters to a stellar spectrum with somehow calibrated flux. SLAM is adopted as the function. SLAM is originally designed to estimate atmospheric parameters from the stellar spectra (Guo et al. 2021; Li et al. 2021). As a data-driven method, SLAM (Zhang et al. 2020) well combines the empirical stellar spectra with a wide range of atmospheric parameters and produce a forward model to generate spectrum in response to a set of given stellar parameters. Because the mapping from stellar parameters to the spectrum is highly nonlinear, SLAM selects a machine-learning algorithm to handle it. Support Vector Regression (SVR; Smola & Schölkopf 2004; Chang & Lin 2011) is adopted as the default algorithm to build the mapping. It is a robust nonlinear regression method used in many astronomical research and analysis (Liu et al. 2012, 2014; Li et al. 2014; Bu & Pan 2015; Lu & Li 2015).

Support Vector Machine (SVM) is a kind of nonlinear classifier that classifies data in a binary way according to a supervised learning process. As an extension of SVM, SVR slightly modifies SVM so that it can solve regression problems.

As SLAM is based on SVR and adopts the radial basis function (RBF) as the kernel in SVR, it has three hyperparameters. C and ϵ represent penalty level and tube radius, respectively, while γ represents the width of the kernel. These three hyperparameters are all needed to be determined using the empirical spectra as the training data.

We apply SLAM to build SSF in three steps:

1. *Pre-processing*. As the first step, we pre-process the training sample so that its spectral fluxes and parameters are in the standardized space. Unlike the other SLAM applications that normalize the training spectra to remove the continua, we normalize the spectra to the sum of fluxes to keep the information of the relative flux calibration. Then, we standardize the training spectral fluxes at each wavelength pixel. The stellar labels, i.e., the stellar atmospheric parameters, are also standardized so that their mean is 0 and standard deviation is 1.[4] The standardized procedure is taken from SLAM's internal algorithm.
2. *Data training*. We then train the SVR model with the standardized labels and fluxes at each wavelength pixel to obtain the best-fit hyperparameters for each wavelength pixel. To achieve this, we first set the grid of hyperparameters to be $\epsilon = 0.05$, $C = [0.01,\ 100]$ and γ = [0.01, 100] and then run SLAM to search for the hyperparameters based on a k-fold cross-validation procedure, i.e., we separate the training data into k groups, then we train the SVR model using $k - 1$ groups and predict the spectra for the rest group. For each test, we measure the performance of the model with the mean squared error (MSE) at the jth pixel, which is defined as
$$\mathrm{MSE}_j = \frac{1}{m}\sum_{i=1}^{m}[f_j(\boldsymbol{\theta}_i|\phi) - f_{i,j}]^2, \tag{1}$$
where $\boldsymbol{\theta}_i$ is the input stellar label vector of the ith star of the test data. $f_j(\boldsymbol{\theta}_i|\boldsymbol{\phi})$ denotes the predicted jth pixel of the test spectrum corresponding to $\boldsymbol{\theta}_i$ given the hyperparameter of $\boldsymbol{\phi} = (\epsilon, C, \gamma)$. $f_{i,j}$ denotes the flux of the jth pixel of the ith spectrum of the test data. After k run by selecting different group as the test data, the averaged MSE_j of the k-fold cross-validation can be used to find the best-fit hyperparameters. We finally adopt the hyperparameters that can minimize the averaged MSE_j for the jth wavelength pixel.
3. *Spectra prediction*. The last step is to predict the spectrum corresponding to a set of given stellar labels with the trained SVR model with best-fit hyperparameters.

3. The Training Data Set

To provide a reliable and precise SSF, one needs a suitable empirical stellar spectral library to be acted as the training data

[4] The standardization is calculated as $x_{\mathrm{std}} = (x - \bar{x})/\sigma_x$, where $\bar{x}$ and σ_x are the mean and standard deviation, respectively, of the random variable x.

Table 1
The Coverage of Spectral Type and Parameters of the SSF Packages

Library Name	Coverage of Spectral Type	Coverage of Parameters
SSF-N	A, F, G, K, and early-M type (part of M giant)	$T_{\rm eff}$: 3700 to 8700 K; log g: 0 to 6 dex; [M/H]: −1.5 to 0.5 dex.
SSF-gM	M giant	$T_{\rm eff}$: 3520 to 4000 K; [M/H]: −1.5 to 0.4 dex; giant: 1
SSF-dM	M dwarf	$T_{\rm eff}$: 3295 to 4040 K; [M/H]: −1.0 to 0.1 dex; dwarf: 0
SSF-B	B type	$T_{\rm eff}$: 9000 to 24,000 K; M_G: −5.2 to 1.5 dex

set of the SLAM. Some broadly used empirical library, e.g., MILES, ELODIE, BC03 and etc. (Prugniel & Soubiran 2001; Bruzual & Charlot 2003; Sánchez-Blázquez et al. 2006) can only provide limited stellar spectra. Therefore, we select ATLAS-A library, which is a set of the ATLAS library (Ji et al. 2023) and covers a wide range of spectral types, as the training data. ATLAS is an empirical stellar spectra library with resolution of $R \sim 1800$ and wavelength coverage from 3800 to 8700 Å.

ATLAS-A contains 5342 spectra with effective temperature ($T_{\rm eff}$), surface gravity ($\log g$), and metallicity ([M/H]) and 242 spectra with only the effective temperature and surface gravity. It covers spectral types from O to M and also includes some special types of stars, such as A supergiant, blue horizontal-branch, and Carbon stars. The parameter coverage of ATLAS-A is from $T_{\rm eff} \sim 3000$ to 50,000 K, from $\log g \sim 0$ to 6 dex, and from $[{\rm M/H}] \sim -1.5$ to 0.5 dex. ATLAS-A also provides the coverage of absolute magnitude of G-band from −9.5 to 12 mag. The spectra in ATLAS-A are absolutely calibrated in flux with wavelength range of 3800–8700 Å. Because not all spectra in ATLAS-A have all three atmospheric parameters measured, we separate them into four groups and set up for SLAM sub-models.

3.1. Training Data for SSF-N

We first select spectra with $T_{\rm eff}$, $\log g$ and [M/H] from ATLAS-A as the training set for SSF-N, i.e., the SSF for normal stars. The ranges of parameters of SSF-N are $3700 < T_{\rm eff} < 8700$ K, $0 < \log g < 6$ dex, and $-1.5 < [{\rm M/H}] < +0.5$ dex. To ensure that the parameters are roughly evenly distributed, we assign the samples into small grids in the $T_{\rm eff}$, $\log g$, and [M/H] space with size of $\Delta T_{\rm eff} = 100$ K, $\Delta \log g = 0.2$ dex, and $\Delta[{\rm M/H}] = 0.05$ dex. In each grid, we select 1 to 3 spectra with highest signal-to-noise ratio in the g-band. We finally select 3609 spectra as the training spectra, covering from A to early M-type stars.

3.2. Training Data for SSF-gM

We select M giant spectra only with $T_{\rm eff}$ and [M/H] as the training data for SSF-gM. The [M/H] of M giants is measured by SLAM (Qiu et al. 2023). This training set contains 249 spectra. The range of $T_{\rm eff}$ is from 3520 to 4000 K and that of [M/H] is from −1.5 to 0.4 dex.

3.3. Training Data for SSF-dM

SSF-dM is trained with M dwarf spectra with only $T_{\rm eff}$ and [M/H]. The [M/H] for M dwarfs is adopted from Li et al. (2021). This training set contains 217 spectra. The range of $T_{\rm eff}$ is from 3295 to 4040 K and that of [M/H] is from −1.0 to 0.1 dex.

3.4. Training Data for SSF-B

Because some parameters, such as metallicity, of B-type stars are not given, we use $T_{\rm eff}$ and absolute magnitude (M_G) as input parameters and 30 B type spectra as input spectra for SSF-B. The range of $T_{\rm eff}$ is from 9000 to 24,000 K and M_G is from −5.2 to 1.5 mag.

4. Results

According to the different training sets, we finally establish the SSF, which is composed of four sub-models for different spectral types: SSF-N, SSF-gM, SSF-dM and SSF-B. The coverage of spectral types and parameters for the four sub-models are listed in Table 1. All four sub-models can be downloaded from the website: https://nadc.china-vo.org/res/r101182/ and we also provide a *README* which can be downloaded from the website: https://github.com/hypergravity/SFF to introduce how to use the models. The source code of SLAM can be downloaded on GitHub https://github.com/hypergravity/astroslam under an MIT License.

4.1. Validation

In order to assess the accuracy of the predicted spectra, we verify them from two aspects. The first is to verify the flux continua of predicted spectra by comparing with the spectra corresponding to the same parameters in the known stellar empirical spectra library. The second is to verify the accuracy of the photometry calculated from the predicted spectra by comparing with the observed magnitude with same stellar parameters. It is noted that for different training sample sets, we adopt different verification methods.

4.1.1. Validation of the Spectra

In order to verify the predicted spectra, we make a few comparisons with spectra. First, we select the normal spectral type spectra with stellar labels from the known libraries. Then, we input the stellar labels of these selected spectra to SSF to derive the corresponding predicted spectra.

We select the spectra from MILES and MaStar libraries to verify SSF-N and SSF-gM, while we only select spectra from MILES library for the verification of SSF-dM, since MILES contains few M-type stars.

First, we predict the spectra from SSF using the stellar labels of the spectra in the known libraries. Then, since MaStar does not correct the extinction, we need to add extinction effect back to the predicted spectra before comparison. This step can be conducted as

$$F_{\rm ex}(\lambda) = F_{\rm pre}(\lambda)10^{-A_V R(\lambda)}, \tag{2}$$

where $F_{\rm ex}(\lambda)$ represents the flux of predicted spectra with extinction effect added back, $F_{\rm pre}(\lambda)$ denotes the flux of predicted spectra, A_V is V band extinction obtained from spectral energy distribution (SED) fitting of the photometries of the selected stars from MaStar and $R(\lambda)$ is the extinction law adopted from Cardelli et al. (1989).

Finally, we normalize both the predicted spectra and the spectra from the known libraries using the sum of flux of the spectra over all range of wavelength. Since that the spectra in MILES are originally normalized to the median flux in the range of 5000–5050 Å while spectra with MaStar are originally normalized to the flux at 5450 Å, we perform a further normalization of the predicted spectra before comparison. When comparing the predicted spectra with MILES spectra, we further normalize both spectra by the median flux in the range of 5000–5050 Å. When comparing the predicted spectra with MaStar spectra, we further normalize both spectra using the flux at 5450 Å. We use two steps of normalization to adopt a more accurate flux comparison for that the accuracy of comparison with MaStar can be improved by 19% if we add one more normalization at 5450 Å after the first normalization.

In order to show the comparison more clearly, we quantified the comparison results by calculating the flux ratio over the wavelength. At each wavelength pixel, we calculate the 15th, 50th, and 85th percentiles. The difference between 15th and 50th percentiles and between 50th and 85th percentiles in each wavelength pixel are calculated by:

$$\Delta F_{15-50}(\lambda) = \frac{F_{15}(\lambda) - F_{50}(\lambda)}{F_{50}(\lambda)}$$
$$\Delta F_{50-85}(\lambda) = \frac{F_{50}(\lambda) - F_{85}(\lambda)}{F_{50}(\lambda)}, \tag{3}$$

where $\Delta F_{15-50}(\lambda)$ and $\Delta F_{50-85}(\lambda)$ denotes the difference of flux ratio between 15th and 50th percentiles and between 50th and 85th percentiles, respectively. $F_{15}(\lambda)$, $F_{50}(\lambda)$ and $F_{85}(\lambda)$ represents the value of the flux ratio of multiple spectra at 15th, 50th, and 85th percentiles, respectively, at each wavelength pixel. Notice that the fluxes of strong absorption lines are excluded in the calculation.

We use 532 spectra from MILES library and 237 spectra from MaStar library for the comparison with SSF-N. We use five and four spectra selected from MILES library for the comparison with SSF-gM and SSF-dM, respectively. Figure 1 shows the distribution of the flux ratio between the predicted spectra from SSF-N and the corresponding selected MILES spectra with same stellar parameters. The color codes the effective temperatures of individual stars in the upper panel. The pink lines in the bottom panel denote the 15th, 50th, and 85th percentiles of the flux ratios. The mean values over wavelength of $\Delta F_{15-50}(\lambda)$ and $\Delta F_{50-85}(\lambda)$ are 3.7% and 4.2%, respectively.

Figure 2 shows the distribution of the flux ratio between the predicted spectra from SSF-N and corresponding spectra with same stellar parameters selected from MaStar. The mean values over the wavelength of the flux ratios $\Delta F_{15-50}(\lambda)$ and $\Delta F_{50-85}(\lambda)$ are 4.4% and 5.0%.

Since either SSF-gM or SSF-dM contains only one spectral type of stars (M giant and M dwarf spectra, respectively), we selected five M giants and four M dwarf stars in MILES library for the comparison. Figures 3 and 4 show the flux ratios between the predicted spectra by SSF-gM and SSF-dM and the corresponding MILES spectra. It is noted that some MILES spectra of cooler M type stars is not consistent with their originally published stellar parameters. Therefore, we adopt the $T_{\rm eff}$ and metallicity from Sharma et al. (2016) as the atmospheric parameters to extract spectra from SSF-gM or SSF-dM. The mean values over wavelength of $\Delta F_{15-50}(\lambda)$ and $\Delta F_{50-85}(\lambda)$ are 3.0% and 3.6%, respectively, for SSF-gM, while the values become 2.7% and 2.9%, respectively, for SSF-dM.

We also find four common M giants in both MaStar library and ATLAS-A. We obtain the predicted spectra from SSF-gM for these stars using the stellar parameters given by MaStars. The flux ratios of these spectra are shown in Figure 5. The mean values of $\Delta F_{15-50}(\lambda)$ and $\Delta F_{50-85}(\lambda)$ are 2.7% and 2.3%, respectively. The dispersion of the flux ratio becomes larger at wavelength range smaller than 4500 Å in Figure 5.

We find the largest uncertainty occurs in the wavelength range of 3800–4000 Å, which is illustrated in Figure 5. To see these differences more clearly, we plot a few typical sample spectra in Figure 6. The blue lines denote the predicted spectra and the red lines denote the spectra with same stellar parameters selected from MaStar spectra. All the spectra are normalized to the flux at 5450 Å. At 3800–4000 Å, the large uncertainty shown in Figure 5 is likely due to the lower fluxes at blue wavelength for these cold stars.

For SSF-dM, we find the largest uncertainty occur in a wide range of wavelength, as shown in Figure 4. Similarly,

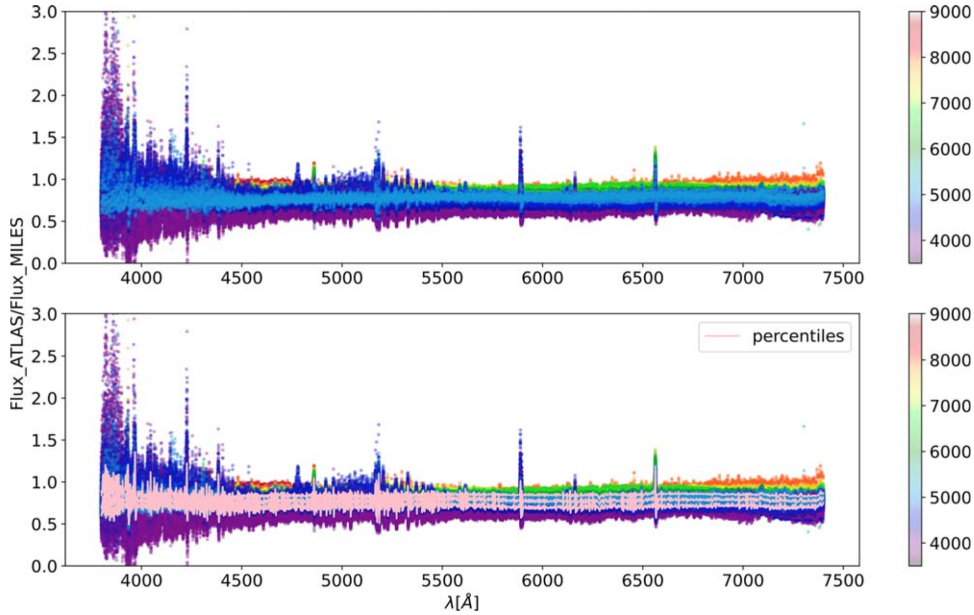

Figure 1. Flux ratio between predicted spectra from SSF-N and the corresponding MILES spectra with the same stellar parameters. In the upper panel, colors represent the effective temperatures of the individual spectrum. In the bottom panel, the pink thick lines indicate the 15th, 50th, and 85th percentiles of the flux ratio at each wavelength pixel.

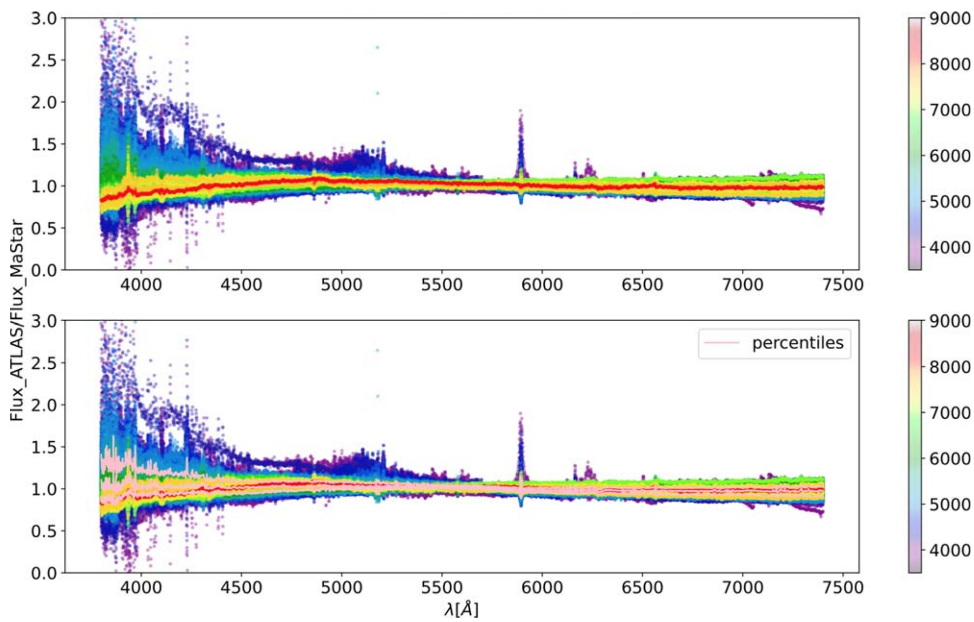

Figure 2. Flux ratio between predicted spectra from SSF-N and the corresponding MaStar spectra. The symbols and colors are same as in Figure 1.

we plot a comparison some predicted sample spectra with the MILES spectra with same stellar parameters in Figure 7. The spectra shown in the figure are normalized to the median flux in the range of 5000–5050 Å. We find that the largest difference in wavelength range of 6750–6850 Å mainly appear in the M dwarf spectra with T_{eff} less than 3700 K. It may be cause that either the absorption lines for atmospheric oxygen molecule in MILES is not cleanly reduced or they have been over-subtracted in the process of skylight subtraction of LAMOST data pipeline. To further investigate

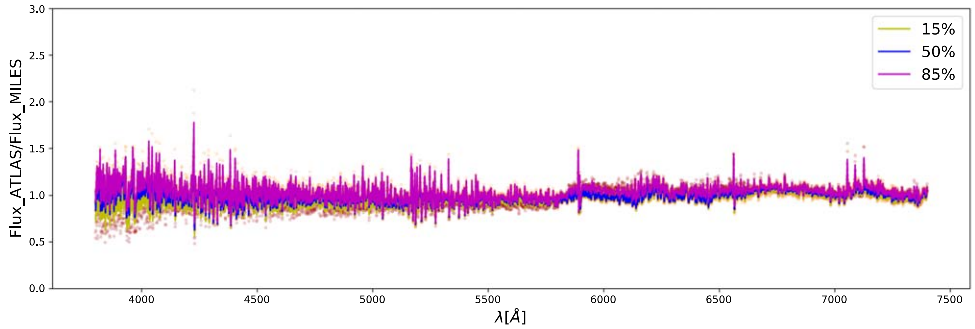

Figure 3. Flux ratio between predicted spectra from SSF-gM and the corresponding MILES spectra. The yellow, blue, and purple thick lines indicate the 15th, 50th, and 85th percentiles of the flux ratio at each wavelength pixel, respectively.

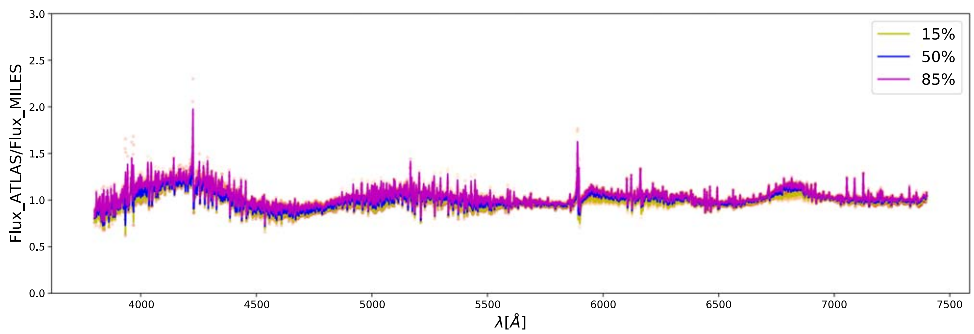

Figure 4. Flux ratio between predicted spectra from SSF-dM and the corresponding MILES spectra. The symbols and colors are same as in Figure 3.

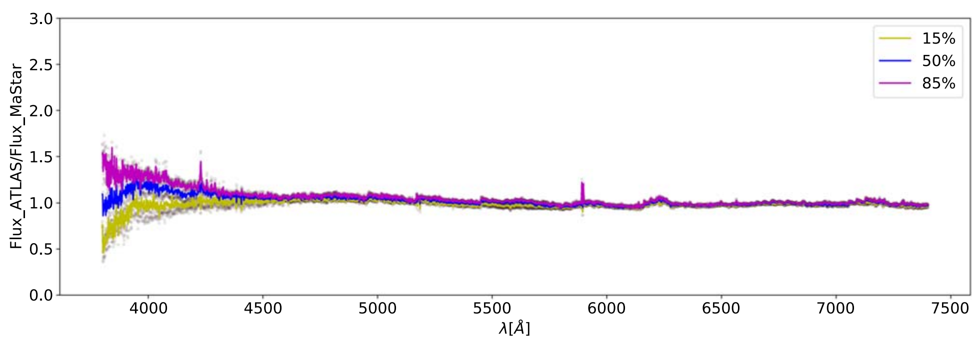

Figure 5. Flux ratio between predicted spectra from SSF-gM and the corresponding MaStar spectra. The symbols and colors are same as in Figure 3.

this difference, we compare some sample M dwarf spectra from MaStar spectral library with the predicted spectra using same stellar parameters. As shown in Figure 8, the predicted spectra are well consistent with the MaStar spectra. This means that it is likely the MILES that does not cleanly subtract the atmospheric O_2 lines during data reduction. Except the uncertainty at 6750–6850 Å, Figures 7 and 8 show that the wave-like uncertainties are mainly due to (a) the larger difference of continuum between the predicted and MILES/MaStar spectra or (b) the large uncertainties in stellar parameters for these cold stars.

4.1.2. Validation by the Observed Photometric Measurement

To further verify the accuracy of the predicted spectra, we also compare the values of the observed magnitude with the corresponding ones that calculated from the predicted spectra. In accordance with the steps in Section 4.1.1, we first select the

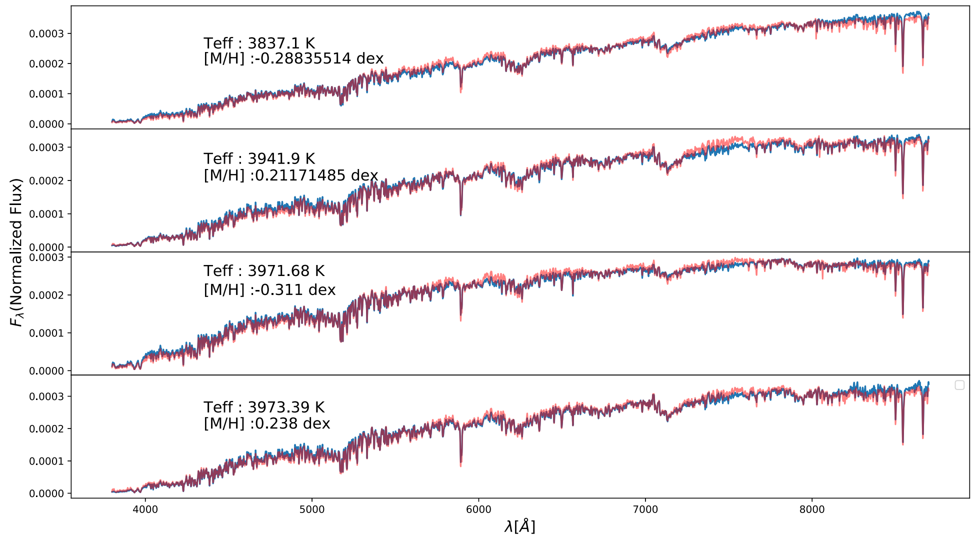

Figure 6. This figure shows some samples of the direct comparison between the MaStar spectra and the corresponding predicted spectra from SSF-gM. The blue and red indicate the predicted spectra from SSF and the spectra from MaStar, respectively. We label the different parameters in the panels for each spectra.

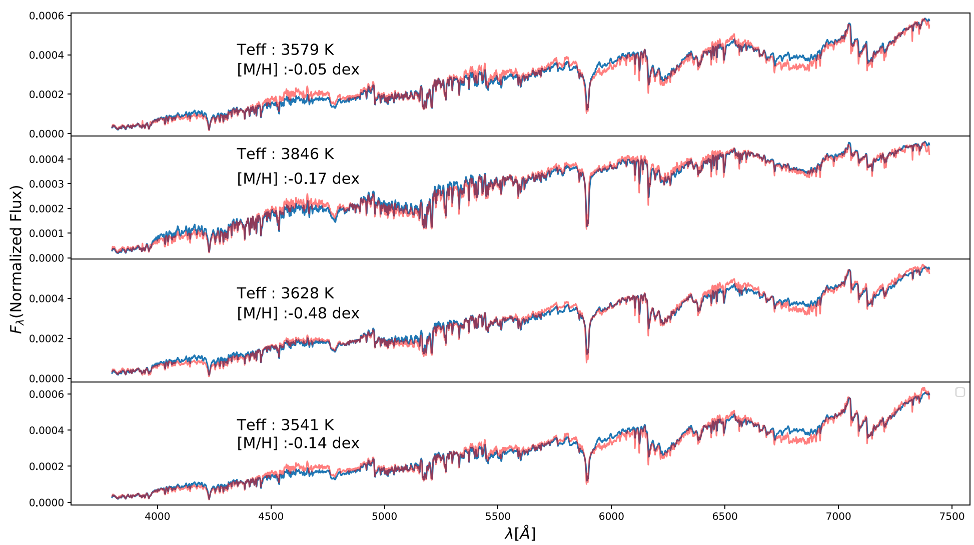

Figure 7. This figure shows some samples of the direct comparison between the MILES spectra and the corresponding predicted spectra from SSF-dM. The blue and red indicate the predicted spectra from SSF and the spectra from MILES, respectively. We label different parameters in the panels for each spectrum.

samples with stellar labels falling in the parameter space of the training set from LAMOST DR5 data. Then, we apply the SVR model to find the corresponding predicted spectra using the stellar parameters provided by LAMOST catalog. We add extinction, which is derived from ATLAS-A catalog using SED fitting, back to the predicted spectra by Equation (2) and integrate the fluxes of the predicted spectra in each photometric band by

$$F_i = \int F_{\rm ex}(\lambda) T_i(\lambda) d\lambda, \tag{4}$$

where i denotes the name of band, F_i is the derived flux of the given band from the predicted spectrum, and $T_i(\lambda)$ represents the transmission curve of the band i. The F_i is then converted into AB magnitude by

$$m_i = -2.5 \log \frac{F_i \times c}{\lambda^2} - 48.6, \tag{5}$$

where m_i represents the AB magnitude, λ is the effective wavelength for the given photometric band and c is the speed of light.

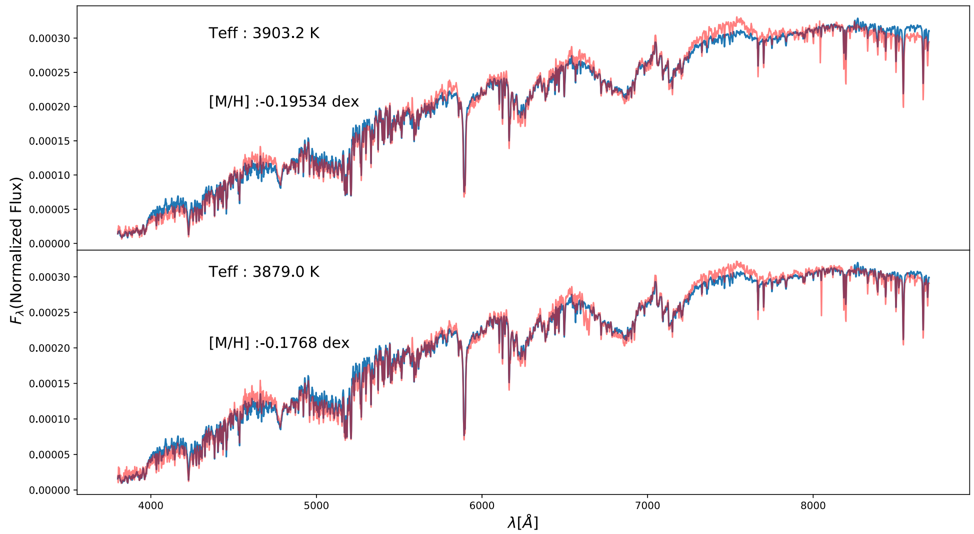

Figure 8. This figure shows some samples of the direct comparison between the MaStar spectra and the corresponding predicted spectra from SSF-dM. The blue and red indicate the predicted spectra from SSF and the spectra from MaStar, respectively. We label different parameters in the panels for each spectrum.

Because that the fluxes of ATLAS spectrum are calibrated according to g, r, and i bands in PS1 (The Panoramic Survey Telescope and Rapid Response System-1 Chambers et al. 2016) and APASS (The American Association of Variable Star Observers Photometric All-Sky Survey Henden & Munari 2014; Henden et al. 2016), we also select the photometric values of the same bands in these two photometric systems for comparison.

Since our predicted spectrum is not absolutely flux-calibrated, in order to obtain accurate magnitude value in photometric bands, we need to calculate the absolute flux of spectrum with extinction by

$$F_{\mathrm{cal}}(\lambda) = k \times F_{\mathrm{ex}}(\lambda), \tag{6}$$

where $F_{\mathrm{cal}}(\lambda)$ denotes the flux with extinction and distance that to calculate the observed magnitude and k is coefficient between $F_{\mathrm{cal}}(\lambda)$ and $F_{\mathrm{ex}}(\lambda)$.

In order to get k, we first calculate the magnitude in the g band for $F_{\mathrm{ex}}(\lambda)$ and then we calculate the difference between the calculated magnitude and the observed magnitude in the g band. According to the Pogson formula, we get k by

$$\begin{aligned} \Delta g &= g_{\mathrm{pre}} - g_{\mathrm{obs}} \\ k &= 10^{-2.5\times\Delta g} \end{aligned} \tag{7}$$

where g_{pre} represents the magnitude for the predicted spectrum with extinction effect added and g_{obs} represents the observed magnitude in the g band.

Since we use the g band to obtain k, the value of magnitude calculated from $F_{\mathrm{cal}}(\lambda)$ in the g band must be equal to the observed magnitude. Therefore, we only choose r and i bands in PS1 and APASS photometric systems for comparison.

For the validation of SSF-N, we generate two sets of samples with atmospheric parameters and spectra published by LAMOST DR5. One has been observed in PS1 with accurate (but not saturated) g, r, and i bands and the other has APASS g, r, and i bands. We randomly select 726 objects with clear photometry information and with representative observed parameter ranges in SSF-N. We finally obtain 484 samples with PS1 photometry and 242 samples with APASS photometry. Both samples are sent to SSF-N to obtain the predicted spectra by using their corresponding T_{eff}, $\log g$, and [M/H] from LAMOST as the input stellar labels.

Figures 9 and 10 display the distributions of the difference of magnitudes between the predicted spectral fluxes and observed magnitudes in r and i bands. In Figure 9, by comparing with the corresponding magnitudes from PS1, the systematic shifts of the predicted spectra are 0.01 and 0.01 mag in r and i band, respectively, with the random uncertainties of $<$0.04 mag, which is slightly larger than the typical error of PS1 photometry. In Figure 10, by comparing with APASS magnitudes, the systematic biases of the predicted spectra are -0.05 and -0.11 mag in r and i bands, respectively, with the random uncertainties less than 0.31 mag. Comparing with the typical error of APASS magnitude, which is around $0.1 \sim 0.2$ mag, the uncertainty is not surprisingly large. The two figures show that the performance of the fluxes of spectra is poor in the i band, which is mainly contributed by the late type stars with significant molecular absorption bands in red wavelength. Nevertheless, these figures show that the prediction of spectra according to the stellar labels can robustly reproduce the photometry of stars with slightly larger uncertainty than observed photometry.

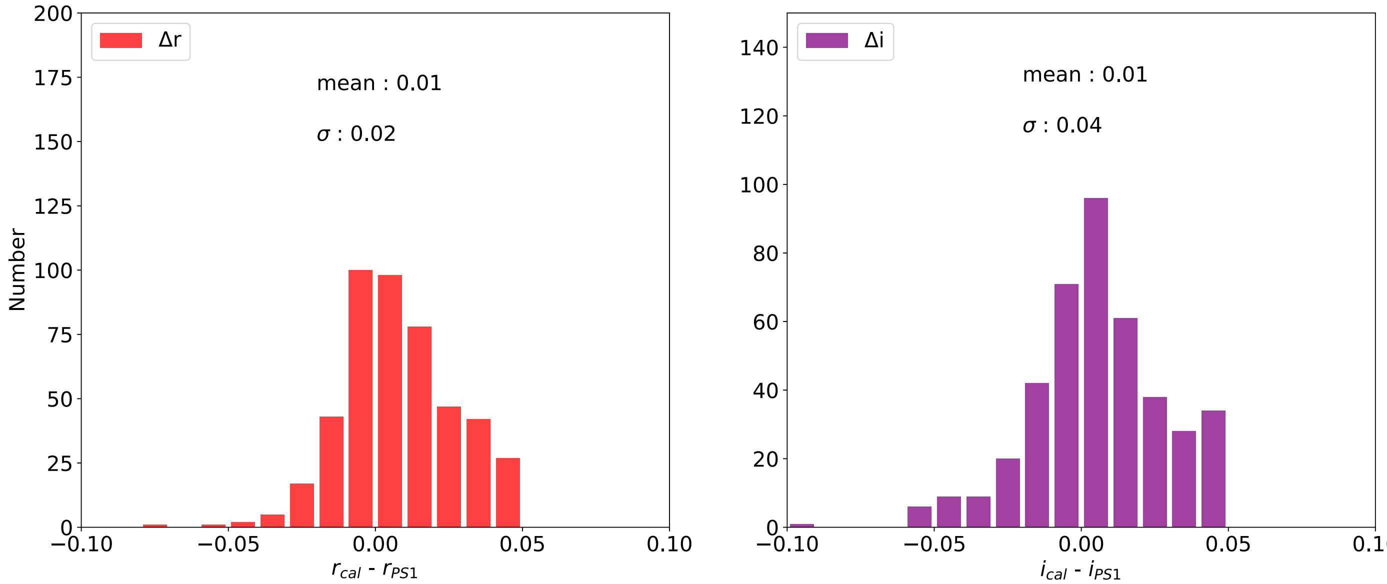

Figure 9. From left to right, the panels display the distributions of the difference between the observed and the SSF-N predicted magnitudes in PS1 r and i bands, respectively. The vertical axes are the star numbers.

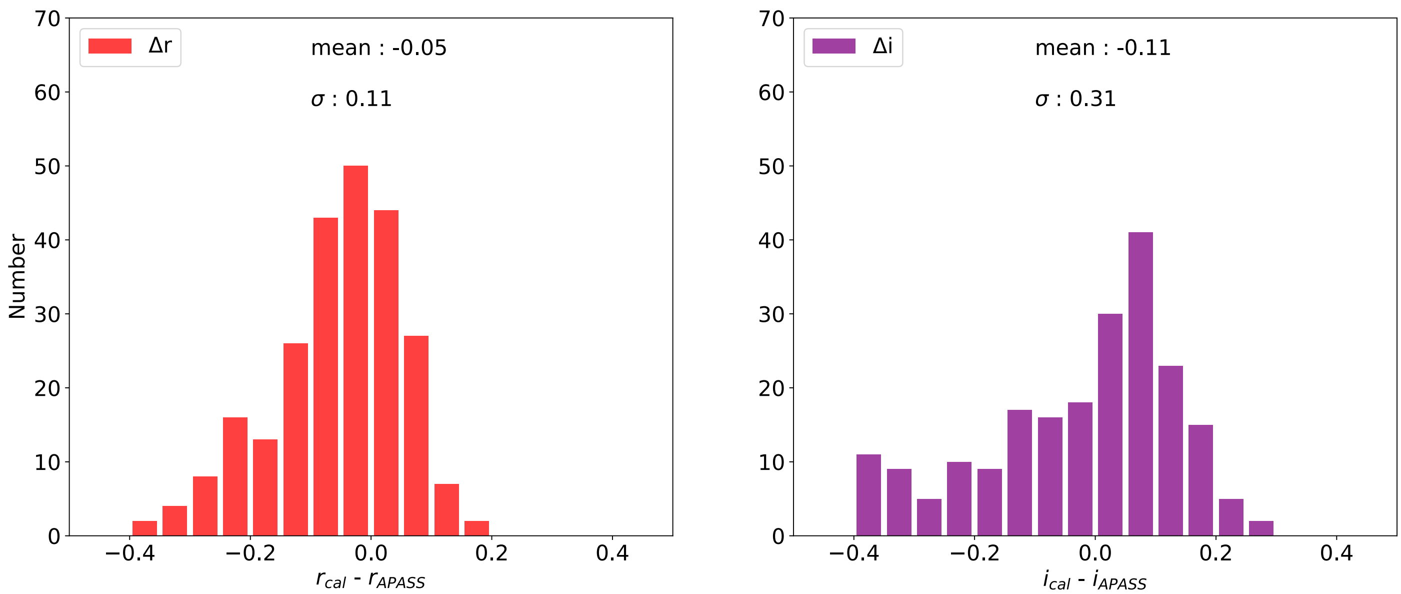

Figure 10. These panels are similar to Figure 9, but are compared with APASS r and i bands.

Since the number of M-type training stars is relatively small, we select 106 samples with photometric information in APASS g, r, and i bands from LAMOST DR5 data for the validation of SSF-gM and 108 samples with photometric information in PS1 g, r, and i bands from LAMOST DR5 data for the validation of SSF-dM. Figure 11 displays the distribution of the difference between the observed and the SSF-gM predicted magnitudes in APASS r and i bands, while Figure 12 displays the distribution of the difference between the observed and the SSF-dM predicted magnitudes in PS1 r and i bands. It is seen that the systematic bias are -0.169 and -0.256 mag in the comparison with APASS r and i band, respectively, with the random uncertainties of less than 0.18 mag in Figure 11. The systematic shifts are -0.05 and -0.07 mag in the comparison with PS1 r and i band, respectively, with the random uncertainties of less than 0.08 mag in Figure 12. The bias in APASS i band shown

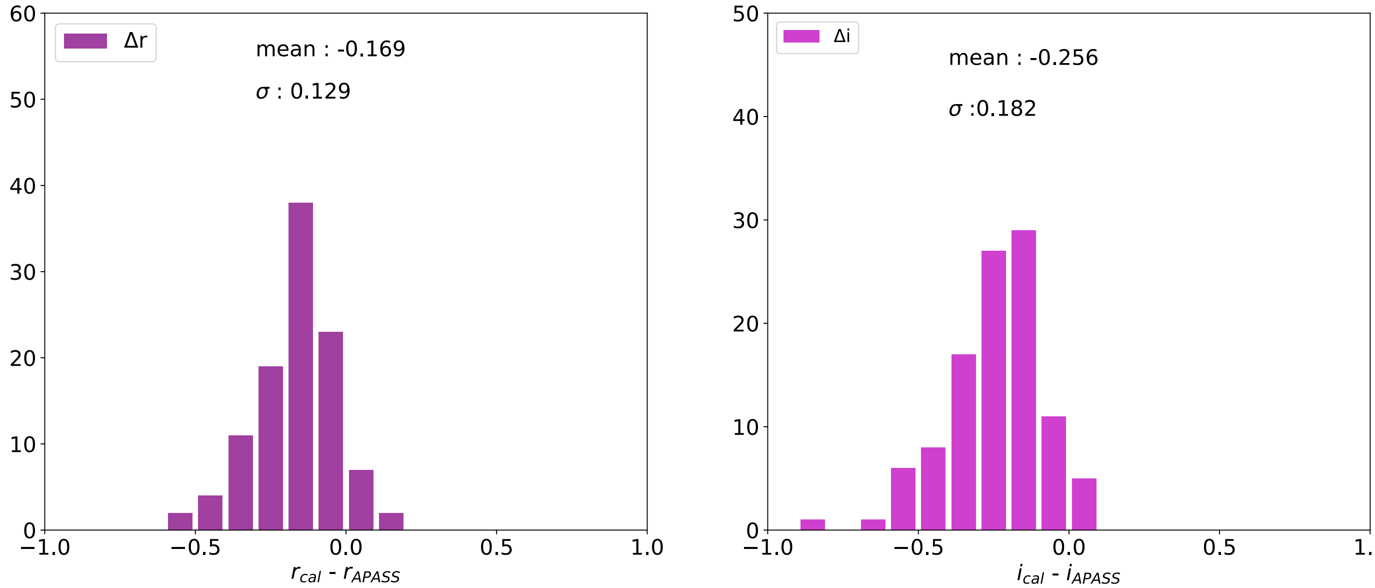

Figure 11. The histograms display the distributions of the difference between the observed and the SSF-gM predicted magnitudes in APASS *r* (left panel) and *i* (right panel) bands.

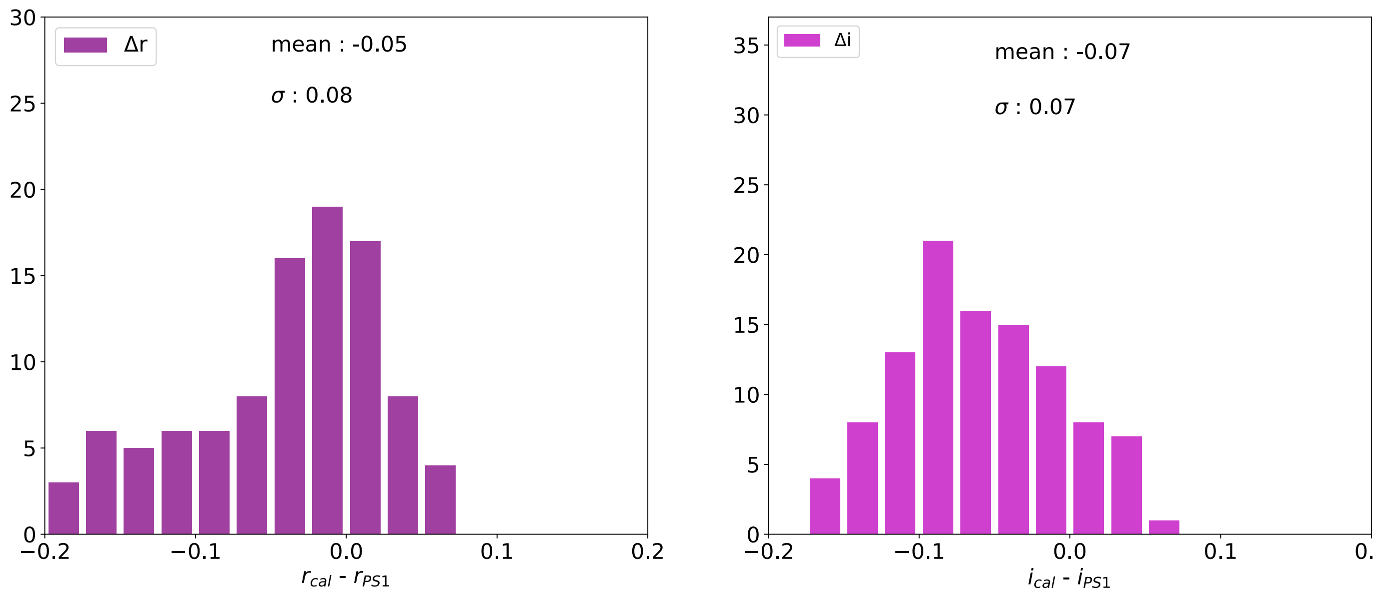

Figure 12. These panels are similar to Figure 9, but for SSF-dM. From left to right, the panels show the distributions of the difference between the observed and the predicted magnitudes in PS1 *r* and *i* bands, respectively.

in Figure 11 is slightly large, which is probably due to the fact that the parameters of the training samples do not sufficiently fill in with the parameter space and the photometry of the red stars may not be precise in APASS.

We select 14 spectra with input parameters in the training samples of SSF-B to evaluate the difference between the observed magnitude and the corresponding magnitude from the spectra predicted by SSF-B. Figure 13 shows the final comparison result. The biases in the figure are −0.001 and −0.003 mag in PS1 *r* and *i* band, respectively with the uncertainties are 0.018 and 0.015 mag, respectively.

4.2. Caveats

There are a few caveats need to be noticed when using SSF.

1. With the exception of SSF-N, which adopts three parameters ($T_{\rm eff}$, $\log g$ and [M/H]) as stellar labels, the rest of the libraries basically contain only two. However,

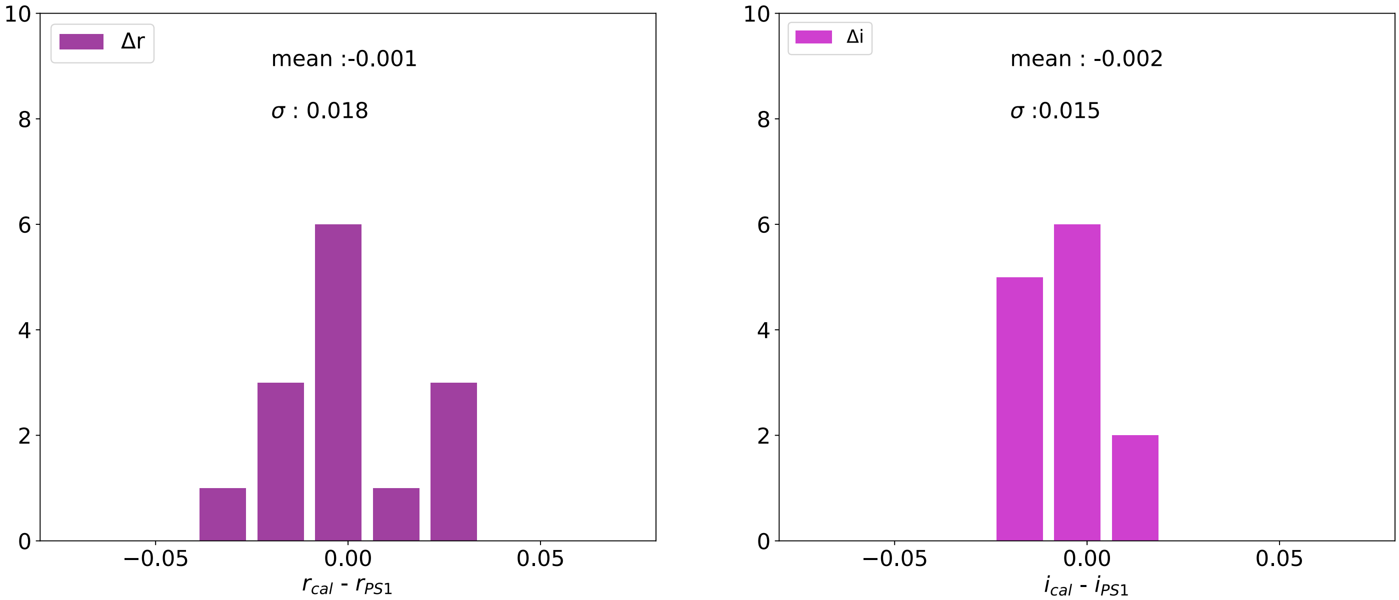

Figure 13. These panels are similar to Figure 9, but for SSF-B. From left to right, the panels show the distributions of the difference between the observed and the predicted magnitudes in PS1 r and i bands, respectively.

to minimize the change of the SLAM code, we adopt a constant value 1 and 0 as the fixed third parameter for SSF-gM and SSF-dM, respectively. The SSF-N covers the spectral types from A to early-M, while the M type stars here are only some of the early type M giants. And the rest three libraries are only contain one specific class of stellar spectra, so users need to pay attention before applying these libraries.

2. According to the stellar parameters coverage in the training samples, the spectral prediction at the edge of the training sample may not be accurate. This mainly causes that there are very few stars being located near the edge of the training sample.

5. Conclusions

Based on ATLAS-A, an empirical stellar spectral library, we use SLAM to build SSF, a tool that can get arbitrary empirical spectra of stars given arbitrary stellar labels. According to the parameters ranges, we obtain four sub-models, SSF-N, -gM, -dM, and -B, with training data sets of different types of stars.

SSF-N mainly covers the parameter space such that $T_{\rm eff}$ from 3700 to 8700 K, log g from 0 to 6 dex and [M/H] from −1.5 to 0.5 dex. SSF-gM covers M giant stars with $T_{\rm eff}$ from 3520 to 4000 K and [M/H] from −1.5 to 0.5 dex. SSF-dM covers M dwarf stars with $T_{\rm eff}$ from 3295 to 4040 K and [M/H] from −1.0 to 0.1 dex. SSF-B mainly covers the hot stars with $T_{\rm eff}$ from 9000 to 24,000 K and M_G from −5.2 to 1.5 mag. These spectral libraries can not only provide help for stellar population research, but also can be used to estimate stellar parameters for all types of spectra.

We validate the four libraries by comparing with the known empirical libraries: MILES and MaStar. The photometry can be obtained by integrating the predicted spectra and is compared with the observed photometric information in g, r and i bands either from PS1 or from APASS. The predicted spectra well agree with the corresponding spectra from the known libraries and the magnitudes calculated from predicted spectra are also well consistent with the corresponding observed ones.

Acknowledgments

This work is supported by the National Key R&D Program of China No. 2019YFA0405500. C.L. thanks the National Natural Science Foundation of China (NSFC) with grant No. 11835057. Guo Shou Jing 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 and managed by the National Astronomical Observatories, the Chinese Academy of Sciences.

References

Bagnulo, S., Jehin, E., Ledoux, C., et al. 2003, Msngr, 114, 10
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bu, Y., & Pan, J. 2015, MNRAS, 447, 256
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560

Chang, C.-C., & Lin, C.-J. 2011, ACM Transactions on Intelligent Systems and Technology, 2, 1
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RAA, 12, 735
Du, B., Luo, A.-L., Zuo, F., et al. 2019, ApJS, 240, 10
Freeman, K. C. 2012, Galactic Archaeology: Near-field Cosmology and the Formation of the Milky Way, 458, 393
Guo, Y., Zhang, B., Liu, C., et al. 2021, ApJS, 257, 54
Henden, A. A., Templeton, M., Terrell, D., et al. 2016, yCat, II/336
Ji, W., Liu, C., Deng, l., et al. 2023, ApJS, 265, 61
Koleva, M., Prugniel, P., Bouchard, A., et al. 2009, A&A, 501, 1269
Le Borgne, D., Rocca-Volmerange, B., Prugniel, P., et al. 2004, A&A, 425, 881
Le Borgne, J.-F., Bruzual, G., Pelló, R., et al. 2003, A&A, 402, 433
Li, J., Liu, C., Zhang, B., et al. 2021, ApJS, 253, 45
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. 2014, ApJ, 790, 110
Lu, Y., & Li, X. 2015, MNRAS, 452, 1394
Majewski, S. R. 2012, AAS, 219, 205.06
Maraston, C., & Strömbäck, G. 2011, MNRAS, 418, 2785
Ness, M., Hogg, D. W., Rix, H.-W., et al. 2015, ApJ, 808, 16
Pickles, A. J. 1985, ApJS, 59, 33
Pickles, A. J. 1998, PASP, 110, 863
Prugniel, P., Koleva, M., Ocvirk, P., et al. 2007, Stellar Populations as Building Blocks of Galaxies, 241, 68
Prugniel, P., & Soubiran, C. 2001, A&A, 369, 1048
Qiu, D., Tian, H., Li, J., et al. 2023, arXiv:2303.08344
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Sharma, K., Prugniel, P., & Singh, H. P. 2016, A&A, 585, A64
Smola, A. J., & Schölkopf, B. 2004, Statistics and Computing, 14, 199
Valdes, F., Gupta, R., Rose, J. A., et al. 2004, ApJS, 152, 251
Vazdekis, A., Ricciardelli, E., Cenarro, A. J., et al. 2012, MNRAS, 424, 157
Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., et al. 2010, MNRAS, 404, 1639
Yan, R., Chen, Y., Lazarz, D., et al. 2019, ApJ, 883, 175
Zhang, B., Liu, C., & Deng, L.-C. 2020, ApJS, 246, 9