On the Construction of New Stellar Classification Templates Library for LAMOST Spectra Analysis Pipeline

Peng Wei[1,2], Ali Luo*[1,3], Yinbi Li[1] , Liangping Tu [1,4], Fengfei Wang[1], Jiannan Zhang[1], Xiaoyan Chen [1], Wen Hou[1,2], Xiao Kong[1,2],Yue Wu [1],Fang Zuo [1]

Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100012, China

lal@nao.cas.cn weipeng@nao.cas.cn

Jingchang Pan[3],Bin Jiang [3],Jie Liu[3], Zhenping Yi[1,2,3]

School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, 264209, China

Yongheng Zhao[1], Jianjun Chen[1,2], Bing Du[1], Yanxin Guo[1], Juanjuan Ren[1,2], Yihan Song[1], Mengxin Wang [1], Kefei Wu[1],Haifeng Yang[1,2,5],

Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100012, China

Ge Jin[6]

University of Science and Technology of China, Hefei 230026, China

Received ______________________; accepted ______________________

[1]Key Laboratory of Optical Astronomy,National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100012, China

[2]University of Chinese Academy of Sciences, Beijing, 100049, China

[3]School of Mechanical, Electrical and Information Engineering, Shandong University,Weihai, 264209, China

[4]School of Science, Liaoning University of Science and Technology, Anshan, 144051, China

[5]School of Computer Science and Technology, Taiyuan University of Science and Technology,Taiyuan 030024, China

[6]University of Science and Technology of China, Hefei 230026, China

ABSTRACT

The LAMOST spectra analysis pipeline, so called 1D pipeline, aims to classify and measure the spectra observed in LAMOST survey. Through this pipeline, the observed stellar spectra are classified into different sub-classes by matching with templates spectra. Consequently, the performance of the stellar classification greatly depends on the quality of template spectra. In this paper, we construct a new LAMOST stellar spectral classification template library, which is supposed to improve the precision and credibility of the present LAMOST stellar classification. About one million spectra are selected from LAMOST Data Release one (DR1) to construct the new stellar templates, and they are gathered in 233 groups by two criteria: I) pseudo g-r colors obtained by convolving the LAMOST spectra with the SDSS *ugriz* filter response curve. II) the stellar subclass given by the LAMOST pipeline. In each group, the template spectra are constructed with three steps. I) Outliers are excluded using the Local Outlier Probabilities (LoOP) algorithm, and then the Principal Component Analysis(PCA) method is applied to the remaining spectra of each group. About 5% outliers are ruled out from one million spectra. II) All remaining spectrum are reconstructed using by the first principal components of each group. III) The weighted average spectrum is made as the template spectrum in each group. Through previous three steps, we initially obtain 216 stellar template spectra. We visually inspect all template spectra, and 29 spectra are abandoned due to low spectral quality. Furthermore, the MK classification for the remaining 187 template spectrum is manually determined by comparing with three template libraries. Meanwhile, 10 template spectra whose subclass is difficult to determine are abandoned. Finally, we obtain a new template library containing 183 LAMOST template spectra with 61 different MK classes by combing with current library.

Subject headings: methods: data analysis, methods: statistical, surveys

1. Introduction

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) is a special reflecting Schmidt telescope with an effective aperture of 3.6-4.9 m, a focal length of 20 m and a field of view (FOV) of 5° (Cui et al. 2012). In virtue of its unique design, LAMOST can observe 4000 spectra simultaneously in a single exposure. Consequently, the LAMOST has a great potential to efficiently survey a large volume of space for stars and galaxies (Zhao et al. 2012).

The LAMOST data are processed by data processing softwares written specifically for the LAMOST Spectral Survey. The LAMOST spectra analysis pipeline (also called 1D pipeline) (Luo & Zhao 2001; Luo et al. 2004, 2008; Wang et al. 2010; Luo et al. 2012) is one of these softwares to produce and analyze LAMOST spectra. The pipeline performs χ^2 fits of the spectra to templates in wavelength space, fitting spectra with linear combinations of eigen-spectra and low-order polynomials. Through the pipeline, the observed stellar spectra are classified into different sub-classes. Consequently, the performance of the stellar classification greatly depends on the quality of templates.

Considering the similarity of the LAMOST stellar spectra with other survey spectra, spectra selected from SDSS and MILES (Falcón-Barroso et al. 2011) are used as templates for stellar classification in current LAMOST spectra analysis pipeline. The current LAMOST stellar classification template library contains 56 templates, which includes 35 templates constructed from a set of SDSS spectra (Wang et al. 2010) and 21 templates selected out from MILES library (Falcón-Barroso et al. 2011). These template spectra cover nearly all common types of stellar spectra. Matching with these templates, the LAMOST stellar spectra are classified into different stellar sub-classes. Although the majority of the LAMOST stellar spectra are correctly classified using current library, there are some significant differences between these spectra. Firstly, the LAMOST , SDSS and MILES spectra have different resolutions, which are 1800 , 2000 and 2000 respectively. Secondly, different instrumental designs bring about different effects on the spectra. In addition, the processes of spectrum extraction, wavelength calibration and flux calibration (Bai 2012) are also different. Considering these issues which can not be ignored, it is necessary to construct a new template library based on the LAMOST spectra.

In this paper, we describe in detail the construction of the new LAMOST stellar classification template library. The paper is organized as follows. Section 2 describe in detail the construction process of the LAMOST stellar template library. The results and discussions are given in section 3. A brief summary is given in section 4.

2. The construction of the templates library

2.1. The Spectra From LAMOST Data Release One (DR1)

The first data release (DR1) of LAMOST survey contains the spectra in the pilot survey and the first year of general survey. The pilot survey of LAMOST was launched on Oct 2011, and ended on June 2012.The first year of the LAMOST regular survey began from September 2012 and ended on June 2013. The DR1 totally contains 2,204,860 spectra, including 717,660 spectra of pilot survey and 1,487,200 spectra of regular survey. The sky coverage of LAMOST DR1 is shown in Fig 1.

There are totally 1,946,429 stellar spectra in LAMOST DR1 and 1,173,928 spectra with SNR>10. The spectral resolution R is about 1800 around g band with a 2/3 silt width (Wang et al. 2013), and the wavelength coverage is from 3700 Å to 9100 Å. Considering the effect of interstellar dust extinction on the spectra and the over density of stars, a mount of 855,583 spectra in direction of the Galactic Anti Center and M31(Liu et al. 2013) whose plate name in the catalogue starts with 'GAC' or 'M31' are excluded. Finally, 1,090,846 stellar spectra are left, which are used for the construction of template library.

2.2. Spectra Grouping

We gather the left 1,090,846 spectra in 233 different groups to construct different kinds of templates by two criteria, including the proposed $g^* - r^*$ color and the subclass given by the LAMOST 1D pipeline.

2.2.1. The pseudo g-r color

LAMOST is a spectroscopic survey oriented telescope, and its photometric data are from catalogs of other surveys. Besides, the flux calibration of LAMOST spectra is relative (Bai 2012). Therefore, accurate and uniform colors can not be obtained for LAMOST spectra. In order to solve this problem, we propose a pseudo g-r color (hereafter g^*-r^*) obtained by convolving each observed spectra with the SDSS *ugriz* filter response curves. The calculation method is described in detail as follows.

1. Suppose that the sampling points of g and r filter response curves are P_g, P_r respectively , and the response values are C_g, C_r.

2. Interpolate the flux of the observed spectra in the points of P_g and P_r to get F_g and F_r.

3. Get the pseudo color g^*-r^*:

$$g^* - r^* = -2.5 * log\frac{F_g \bigotimes C_g}{\sum C_g} + 2.5 * log\frac{F_r \bigotimes C_r}{\sum C_r} \tag{1}$$

The g-r color is a indicator of stellar surface effective temperature ($Teff$) (Lee et al. 2008; Ivezić et al. 2008), so we select objects with SDSS *ugriz* magnitudes and signal to noise ratio (SNR) larger than 20 to check whether the $g^* - r^*$ color is consistent with T_{eff}. The average value and standard deviation of g^*-r^* for different spectral types (O, B, A, F, G, K, M-type) are shown in Figure 2, and we can see that the g^*-r^* color varies obviously for each spectral type.

For these selected objects, the relationship between g-r color and g^*-r^* is shown in Fig.3. There is an obvious linear relationship between the two colors, and the derived best-fit expression is shown in formula 2:

$$g - r = 0.807 * (g^* - r^*) + 0.655 \tag{2}$$

Ivezić et al. (2008) derived a relation between the T_{eff} and the color g-r in the range of $-0.3 < g - r < 1.3$ for SDSS spectra:

$$Log_{10}(T_{eff}/K) = 0.0283 * (g - r)^3 + 0.0488 * (g - r)^2 - 0.316 * (g - r) + 3.882 \tag{3}$$

Thus, we are able to derive a expression shown in the formula 4 between effective temperature T_{eff} and our pseudo color g^*-r^* using the formula 3 and the formula 2 as:

$$Log_{10}(T_{eff}/K) = 0.0283 * (g^* - r^*)^3 + 0.0318 * (g^* - r^*)^2 - 0.203 * (g^* - r^*) + 3.696 \tag{4}$$

The LAMOST Stellar Parameter pipeline (LASP, see Wu et al. (2011)) provides effective T_{eff}, $logg$ and $[Fe/H]$ for 1,085,404 A,F, G and K-type spectra. The relationship between g^*-r^* color and T_{eff} is shown in Figure 4 and the derived polynomial expression is as shown in formula 5:

$$Log_{10}(T_{eff}/K) = 0.0432 * (g^* - r^*)^3 + 0.0107 * (g^* - r^*)^2 - 0.165 * (g^* - r^*) + 3.746 \tag{5}$$

As shown in Fig.4, the formulas 4 and 5 nearly coincide with each other in the range of T_{eff} [5500K,7000K]. Thus, the defined g^*-r^* color is also be a indicator of $Teff$, which is used to divide the selected spectra into different groups.

2.2.2. *Group dividing criteria*

In order to construct different kinds of LAMOST templates, we divide these spectra into 233 different groups by the proposed g^*-r^* color and the stellar subclass classified by the LAMOST 1D pipeline using the current template library.

As discussed above, the proposed g^*-r^* color is a good indicator of T_{eff}. Therefore, we select these spectra with g^*-r^* in the range[-1.5,2.0], and divide them into 175 groups with 0.02 mag width interval. These groups are marked with group-id from 1 to 175, and the the number distribution is shown in Figure 5.

In addition, other 58 groups are formed by the subclass given by the current LAMOST 1D pipeline. These groups are marked with group-id from 176 to 233. Yi et al. (2013) presented a spectroscopic catalog of 67,082 M dwarfs from the LAMOST Pilot Survey, and we put these spectra into groups with group-id from 220 to 229. Zhao et al. (2013) presented a spectroscopically identified catalog of 70 DA white dwarfs (WDs), and Zhang et al. (2013) identified other 230 DA white dwarfs, We combine these two catalogs and put them into group 233. Jiang et al. (2013) reported the identification of 10 cataclysmic variables, and we allocate a group id 204 for these spectra.

2.3. The construction of spectral templates library

2.3.1. *Excluding outliers using Local Outlier Probabilities (LoOP)*

To construct template spectra, 233 different groups are formed by gathering similar spectra following two criteria. Although the spectra in the same group are very similar with each other, there are still some outliers existing in each group for many reasons, including the effect of instellar extinction on the continuum, strong noises, existence of unusual spectral features and other issues. Obviously, these outliers should be excluded to generate much purer spectra for constructing the template library.

In our work, the Local Outlier Probabilities (LoOP, see Kriegel et al. (2009) for the detailed description) method is used to exclude these outliers. LoOP is a local density based method that uses statistical concepts to output the final score. The LoOP score represents the probability that a particular point is a local density outlier.

2.3.2. The spectra reconstruction using Principal Component Analysis (PCA)

The Principal Component Analysis (PCA, see Jolliffe (2002)) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible).

As a feasible tool, Principal Component Analysis (PCA) has been applied in the classification and outlier detection of spectra (Whitney 1983; Bailer-Jones et al. 1998; Yip et al. 2004; Tu et al. 2009, 2010; Almeida & Prieto 2013; Wei et al. 2013; Jiang et al. 2013) by reducing the dimensionality of the original spectral data to very few components. PCA are also able to successfully reconstruct the original spectra by using the first few components (Singh et al. 1998). In our work, PCA is used to reconstruct the original spectra to improve the similarity of spectra in the same group.

2.3.3. The steps to construct the template library

We use the following ten steps to construct the new stellar template library. Note that, the number in each bracket is the number of remaining spectra after this step, and the initial number of spectra to construct the template library is 1,090,846.

1. For the groups with more than 5,000 spectra, only first 5,000 spectra with the largest SNR are selected.[525,723]

2. Remove the readshift of each spectrum, unify their wavelength to 3800Å-9000Å with fixed step 1Å (the amount of all sampling points is N=5201) and get the unified flux F.

3. Exclude these spectra existing $F \leq 0$, and normalize the flux F of the remaining spectra as follows [489,137]:
$$F_i = \frac{F_i}{\sqrt{\sum_{j=1}^{N} F_i^2}} \tag{6}$$

4. Calculate LoOP for each group.

5. These spectra with $LoOP \geq 0.4$ are excluded. [415,381]

6. Apply the PCA method to the remaining spectra in each group to obtain a feature matrix T and the corresponding eigen values λ.

7. Select the first k-th principal components (eigen spectra) while the variance contribution rate μ
$$\mu = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i} > \theta \tag{7}$$
where θ is a fixed given threshold (0.99 is used in our work), and k is set to 2 when $k = 1$.

8. Reconstruct each remaining spectra using obtained first k principal components.

9. Calculate LoOP of remaining reconstructed spectra in each group again, and exclude these spectra with $LoOP \geq 0.2$. [367,248]

10. Take the SNR weighted average spectrum as the template spectrum in each group.

2.3.4. Determining stellar spectral subclass for template spectra

Following above ten steps, the template spectra are successfully constructed in 216 groups (nearly 92%) while other 17 groups fail mainly due to the lack of enough spectra with high quality. After matching with these template spectra, each observed spectrum in the LAMOST survey will be classfied as a stellar spectral subtype given by the best matched template spectrum. Therefore, it is also an important step to determine subclass for these constructed template spectra. In order to get better stellar spectral subclass, we use following two steps to classify these template spectra.

1. First, each template spectrum are matching with three template libraries, and the first four closest spectra in each library are chosen. That is to say, there are 12 different spectra from three different libraries for each template spectrum in our library. The three libraries mentioned above are described as follows.

   - Danks & Dennefeld (1994) presented spectra for MK standards in the wavelength range 5800Å-10200Å. The stars cover the normal spectral types from O to M and luminosity types I, III, and V. The projected slit width along the dispersion is about 4Å and the resolution R is about 1200. Two wavelength ranges [7500Å,7700Å] and [6800Å,7000Å] are masked to get rid of the strong telluric lines left in the spectra. We decrease the resolution R of our templates to 1200 by convolving a gaussian function. All template spectra and standard spectra are unified into the wavelength range [6100Å,9000Å] with a fixed step 4Å.

- Bolton et al. (2012) described the detail of the pipeline for SDSS III, and published the template used. For stellar spectral classification, 123 templates which are created from the full database of Indo-U.S. spectra are provided. Each spectrum are labeled a MK class by matching with POLLUX database. The resolution R of these 123 spectra is about 2000, and the wavelength coverage is from 3500Å to 11200Å. These spectra are unified into the wavelength range [3800Å,9000Å] with a fixed step 1Å similar with the spectra in our template library.
- As mentioned before, the current template library of stellar classification in LAMOST spectra analysis pipeline contains 63 spectra, their resolution R is about 2000, and the wavelength coverage is from 3800Å to 9200Å. These spectra are unified into the wavelength range [3800Å,9000Å] with a fixed step 1Å similar with the spectra in our library.

2. Visual inspection is carried out after automatic matching with spectra libraries. Each template spectrum is visually inspected by checking the matching results with 12 chosen spectra from three libraries described above. And then each spectrum is labeled a MK class given by the visually chosen best matched spectra. Meanwhile, those template spectra with bad data or low SNR are excluded. We develop a web site to finish this process. As shown on the web page[1], the best 12 matched templates in three libraries can be selected to plot together with our constructed template. By comparing the details of these spectra, the best matched template is picked out and

[1]`http://sciwiki.lamost.org/lamost_sctl/v1/tsl_groupdetail.php?id=59`, and 59 can be replaced with other group id

our constructed template is assigned with the subclass of this template.

Finally, there are 177 template spectra remaining with 55 different MK classes in the template library.

3. The finally combined template library used in the LAMOST spectra analysis pipeline

The new stellar template library consists of the newly constructed templates and six templates in current pipeline, and there are 183 spectra and 61 different MK classes left in the combined template library.

3.1. Comparison with the template library in current LAMOST spectra analysis pipeline

As shown in Table 1, our constructed templates replace most templates of the current library. We describe in detail the finally combined library as follows.

- *O-type and B-type Spectra* Due to the lack of enough O-type and B-type spectra in LAMOST DR1, no O-type and B-type templates are successfully constructed and we remain the four O-type and B-type templates in current library. Therefore, there are four types of spectra.

- *A-type Spectra* After visual inspection of all constructed templates, there are 49 spectra with 13 different MK classes left in the combined template library. These template covers two types of A-type giant star, and one is A0-type (early A-type) giant while another one is A7-type (late A-type) giant. For A-type dwarf stars, there are more than one templates of each subclass. Meanwhile, other types of A-type templates from MILES are abandoned due to the wavelength coverage.

- *F, G and K-type Spectra* As the most common types of spectra in the survey, F, G and K-type spectra account for more than 88 percent of the total. The 11 F, G and K-type templates in current library are totally replaced by newly constructed templates. There are 84 templates and 26 subclasses, which significantly exceeds the amount of the templates and the subclasses in current library. What's more, there are more than one templates for the vast majority of the subclasses. These templates cover nearly all subclasses for F, G and K-type spectra.

- *M-type Spectra* There are 122,718 M-type spectra and current pipeline classified these spectra into 12 subclasses. The two templates M2V and M0V are abandoned. Finally, there are 28 M-type templates with 10 different subclasses. For early M-type subclasses, there are more than five template spectra while there are only one spectra for each subclass in current library.

- *Other-type Spectra* There are five relatively rare types of stellar spectra in current library, and three templates are newly reconstructed. Meanwhile, two new types (Double star for the spectra with two obvious components and CV for the spectra cataclysmic variable stars) are added. That's to say, there are 7 templates with 7 relatively rare types. Here, we note that *detailed works are needed to produce more complete and purer samples of these rare types of stars, not taking our classification results as the only criterion.*

Considering that these newly constructed spectra are constructed from a healthy sum of observed spectra from LAMOST DR1, they are similar to the spectra observed in LAMOST survey. Consequently, we can say the new stellar template library can effectively improve the precision and credibility of the stellar classification.

3.2. Comparison with Bolton et al. (2012)

The template library used in SDSS DR 9 contains 123 stellar subclasses (Bolton et al. 2012). These template spectra are individual spectra in Indo-U.S. database. As shown in Table 2, our finally combined library contains more templates for A, F, G, K and M-type spectra but less subclasses than Bolton et al. (2012). For B-type spectra, there are less templates and subclasses in final library than Bolton et al. (2012). From SDSS casjobs, we get that there are 80 subclasses which contain less than 500 spectra in SDSS DR9 and there are only 12,897 spectra (about 1.66% in all 773,275 spectra) with these subclasses. Besides, only 1,062 spectra with SNR> 10 (about 0.22% in all 475,694 spectra with SNR> 10) and the average SNR of these spectra is about 4.93 which is much less the one of all spectra. In other words, the other 43 stellar subclasses contains most spectra especially these spectra with high SNR. Moreover, these spectra are different from LAMOST spectra in spectral resolution, instrumental effect and processing process. Hence, we do not add these templates in Bolton et al. (2012) into our final template library.

3.3. Examples

Here we choose four typical templates(see Fig 6 for the template spectra) to discuss the construction process and the constructed template spectra in detail. The main information of these templates is shown in Table 3. Their template-id are 69, 10, 183, 159, constructed in group 59, 179 ,204 and 222, and the MK classes are F5, A1IV, CV and M2 respectively.

3.3.1. Template #69 Subclass:F5

The template #69 is constructed from the group #59, which contains the spectra in the color g^*-r^* range [-0.34,-0.32]. There are totally 18,534 spectra and the first 5,000 spectra with the highest SNR are chosen. Among these spectra, 4,024 spectra are used to get the principal components. As shown in Fig 7, the variance of the first principal component exceeds more than 99% due to the high similarity of the spectra in the group. Consequently, the reconstructed spectra using first two principal components are nearly similar to the original spectra (see Fig 8). After excluding outliers, 3,048 spectra are left to construct the template spectrum, and the stellar spectral subclass is finally classified as 'F5V'. As shown in Fig 9, the template spectrum is close to the F3V/F5V type spectrum in Bolton et al. (2012). The zoomed figure shows that the strong features such as the Blamer lines and Ca lines of these two spectra coincide with each other well.

3.3.2. Template #10 Subclass: A1IV

The template #10 is constructed from the group #179, which contains the spectra classified as 'A1IV' by current LAMOST spectra analysis pipeline. There are totally 653 spectra in this group. Among these spectra, 517 spectra are used to get the principal components which are used in the spectra reconstruction. Similar with group 59, the variance of the first principal component also exceeds more than 99% (see Fig 10). There are not as many spectra as in group 59 and a fraction of spectra are not well reconstructed (as shown in Fig 11). In spite of this, the template spectrum is well constructed after excluding these badly reconstructed spectra. After excluding outliers, 381 spectra are left to construct the template spectrum. The stellar spectral subclass is finally labeled as 'A1IV'. The template spectra of group 179 is shown in Fig 12, and we can see that our constructed template is similar with even a little better than the 'A1IV' type template in the current library.

3.3.3. Template #183 Subclass: CV

The template #183 is constructed from the group #204, which contains the spectra classified as 'CV' by current LAMOST spectra analysis pipeline. There are totally 27 spectra in this group. Among these spectra, 23 spectra are used to get the principal components. Compared with normal stars, the spectra of CV stars are these with strong hydrogen Balmer and helium emission lines that typically signify ongoing accretion. As shown in Fig 13, the first two principal components show obvious and strong emission lines and the sum of the variances of these two principal components exceeds more than 99%. Compared to stars misclassified as 'CV', the spectra of CV stars are almost faultlessly reconstructed (see Fig 14). And then these misclassified spectra are excluded in the next following steps. After excluding outliers, 17 spectra are left to construct the template spectrum and these spectra are all real CV star. The stellar spectral subclass is finally labeled as 'CV', and the template spectrum shows strong hydrogen Balmer and helium emission lines which coincides the feature of CV stars.

3.3.4. Template #159 Subclass:M2

The template #159 is constructed from the group #222, which contains the spectra classified as 'M2' by current LAMOST spectra analysis pipeline and the spectra identified as 'M2' in Yi et al. (2013). There are totally 17,231 spectra in this group. Due to the existence of wavelength points with $flux \leq 0$, only 325 spectra are selected from the first 5,000 spectra with the highest SNR. In spite of this, the template spectrum is also well constructed. As shown in Fig 15, the sum of the variances of the first two principal components exceeds more than 99% of the total variance of the original data. The selected spectra shown in Fig 16 are well reconstructed and the obvious bad features are removed from the spectra.

The stellar spectral subclass is finally ladled as 'M2'. In order to check the quality of our M2 type template spectrum, we compare it with currently used M2 type template in the LAMOST spectra analysis pipeline and M2 type template spectrum presented by Bochanski et al. (2007). From Fig 17, we can infer that our constructed M2-type spectrum is similar with these two spectra.

3.4. Remaining problems

As discussed above, our constructed template spectra can be used in the stellar classification in LAMOST survey. However, there are still some problems needing to solve.

1. We notice that most of our template spectra are main sequence stars. In order to construct the template library which contains as many typs of spectra as possible, such as K-type giants, DC and DZ white dwarfs (Si et al. 2013), we need to add these rare spectra into our template library.

2. At present, we use three libraries to classify our template spectra. However, how to label them better is still a problem.

3. In addition, there are some outliers excluded in each group while constructing the templates. It is also worth of studying these objects and finding rare types even new types of star.

4. Summary

In order to improve the precision and credibility of the stellar classification, a new LAMOST stellar spectral classification templates library is constructed. We select about 750,0000 stellar spectra from LAMOST Data Release One (DR1) and gather them in 233 different groups by proposed pseudo g-r colors and the subclass given by current LAMOST spectra analysis pipeline. Through the proposed steps, including excluding outliers using LoOP, spectral PCA reconstruction etc., the weighted average spectra are constructed as the template spectra in the groups. Afterwards, each template spectrum is assigned with a MK type by comparing with three libraries and visual inspection. Finally, the new stellate classification templates library LAMOST spectra analysis pipeline consists of 176 spectra and 71 different MK classes, and it has been used for new version of LAMOST spectra analysis Pipeline and is published on the website [2]. With the survey progress, more and more spectra will be gathered, and making a complete template library is our goal.

The authors would like to thank the anonymous referees for their constructive comments. This work is supported by the National Natural Science Foundation of China (Grant No.61202315, 11078013, 11203045, 11303036 and 11303061).

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

[2]`http://sciwiki.lamost.org/lamost_sctl/v1`

REFERENCES

Almeida, J. S., & Prieto, C. A. 2013, ApJ, 763, 50

Bai, Z. 2012, Proceedings of the International Astronomical Union, 8, 189

Bailer-Jones, C. A., Irwin, M., & Hippel, T. V. 1998, MNRAS, 298, 361

Bochanski, J. J., West, A. A., Hawley, S. L., & Covey, K. R. 2007, AJ, 133, 531

Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144

Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197

Danks, A. C., & Dennefeld, M. 1994, PASP, 382–396

Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95

Jiang, B., Luo, A., Zhao, Y., & Wei, P. 2013, MNRAS, 430, 986

Jolliffe, I. T. 2002, Principal component analysis (Springer verlag)

Kriegel, H.-P., Kröger, P., Schubert, E., & Zimek, A. 2009, in Proceedings of the 18th ACM conference on Information and knowledge management, 1649–1652 (ACM)

Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008, AJ, 136, 2022

Liu, X.-W., Yuan, H.-B., Huo, Z.-Y., et al. 2013, arXiv preprint arXiv:1306.5376

Luo, A.-L., Wu, Y., Zhao, J., & Zhao, G. 2008, in Proc. of SPIE Vol, vol. 7019, 701935–1

Luo, A.-L., Zhang, H.-T., Zhao, Y.-H., et al. 2012, RAA, 12, 1243

Luo, A.-L., Zhang, Y.-X., & Zhao, Y.-H. 2004, in Astronomical Telescopes and Instrumentation, 756–764 (International Society for Optics and Photonics)

Luo, A.-L., & Zhao, Y.-H. 2001, Chinese Journal of Astronomy and Astrophysics, 1, 563

Luo, A.-L., et al. 2013, in preparation

Si, J., Luo, A., Zhang, J., et al. 2013, arXiv preprint arXiv:1309.1883

Singh, H. P., Gulati, R. K., & Gupta, R. 1998, MNRAS, 295, 312

Tu, L., Luo, A., Wu, F., & Zhao, Y. 2010, Science China Physics, Mechanics and Astronomy, 53, 1928

Tu, L.-P., Luo, A.-L., Wu, F.-C., Wu, C., & Zhao, Y.-H. 2009, RAA, 9, 635

Željko Ivezić, Sesar, B., Jurić, M., et al. 2008, ApJ, 684, 287

Wang, F., Luo, A., & Zhao, Y. 2010, in SPIE Astronomical Telescopes and Instrumentation: Observational Frontiers of Astronomy for the New Decade, 774031–774031 (International Society for Optics and Photonics)

Wang, F., Zhang, H., Luo, A., et al. 2013, Science in China G: Physics and Astronomy, 56, 1833

Wei, P., Luo, A., Li, Y., et al. 2013, MNRAS, 431, 1800

Whitney, C. 1983, Astronomy and Astrophysics Supplement Series, 51, 443

Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, RAA, 11, 924

Yi, Z., Luo, A., Song, Y., et al. 2013, arXiv preprint arXiv:1306.4540

Yip, C., Connolly, A., Berk, D. V., et al. 2004, AJ, 128, 2603

Zhang, Y.-Y., Deng, L.-C., Liu, C., et al. 2013, AJ, 146, 34

Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, RAA, 12, 723

Zhao, J., Luo, A., Oswalt, T., & Zhao, G. 2013, AJ, 145, 169

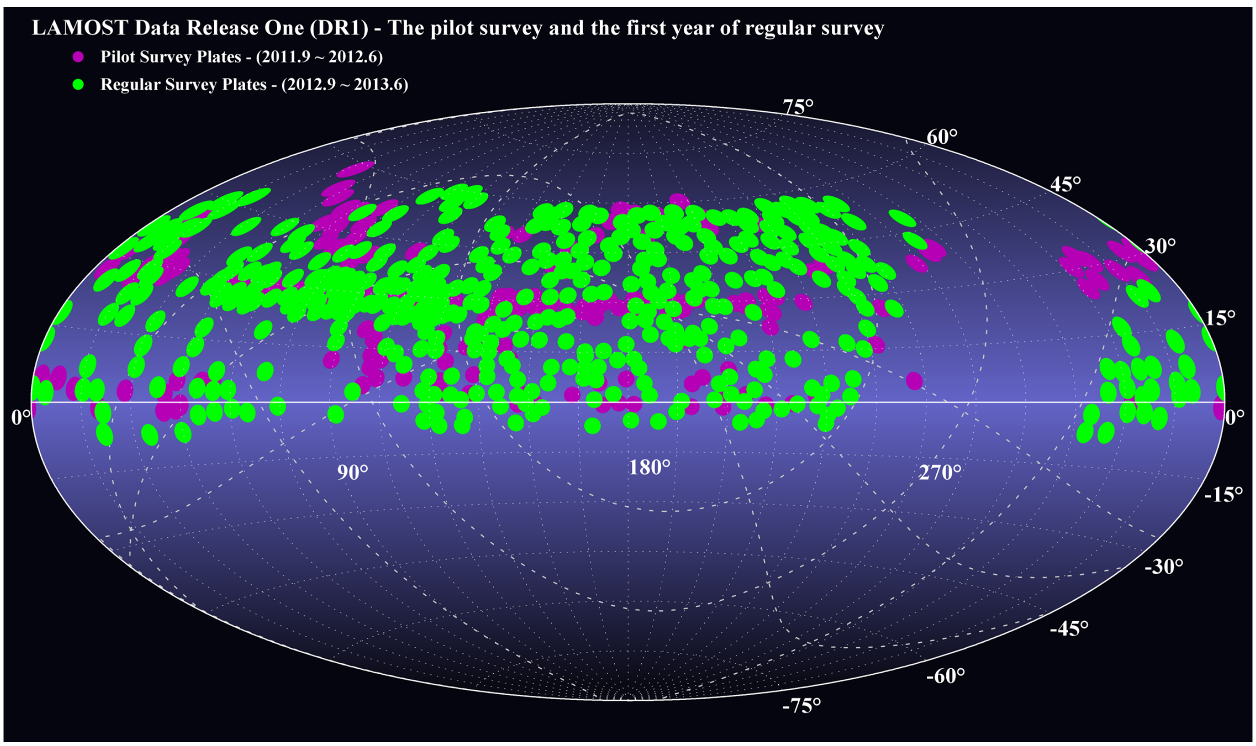

Fig. 1.— The LAMOST DR1 skycoverage (`http://data.lamost.org/u/img/dr1-full.png`)

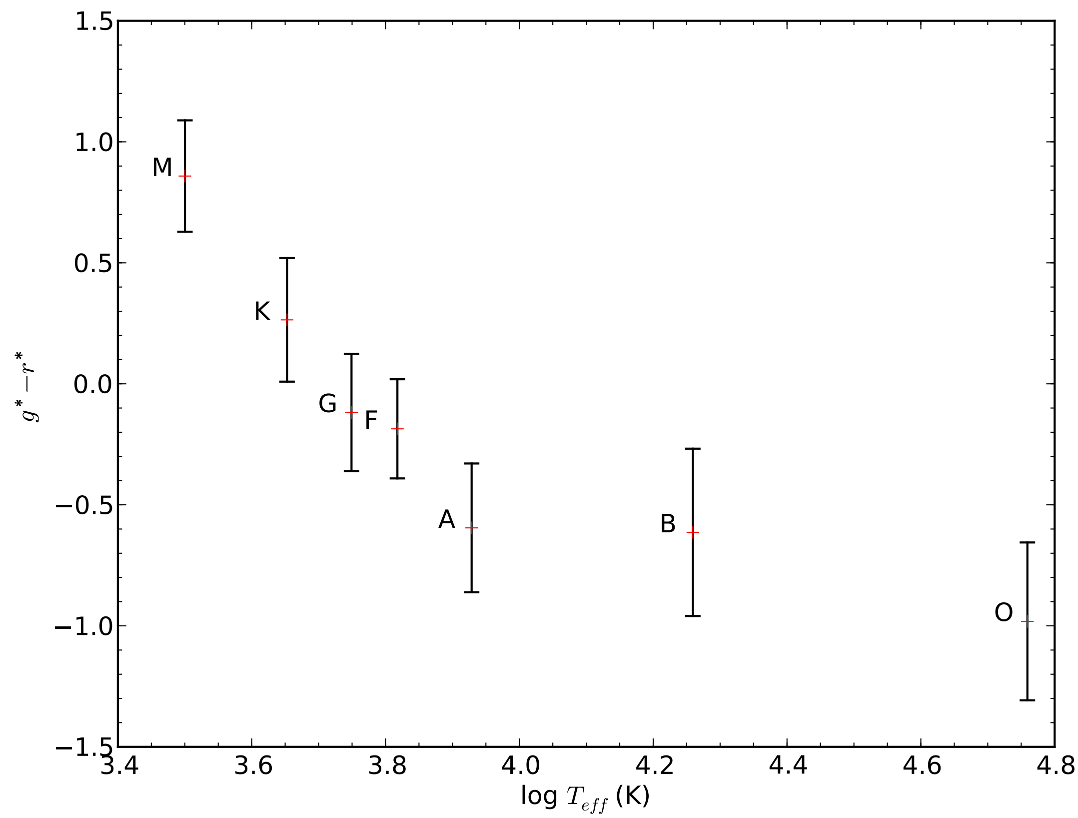

Fig. 2.— The average value and standard deviation of g^*-r^* for each spectral type. The X value of the error bar is the median effective temperature in theory. The y value of the center of each error-bar is the average g^*-r^* color and the half length is the standard deviation of g^*-r^*.

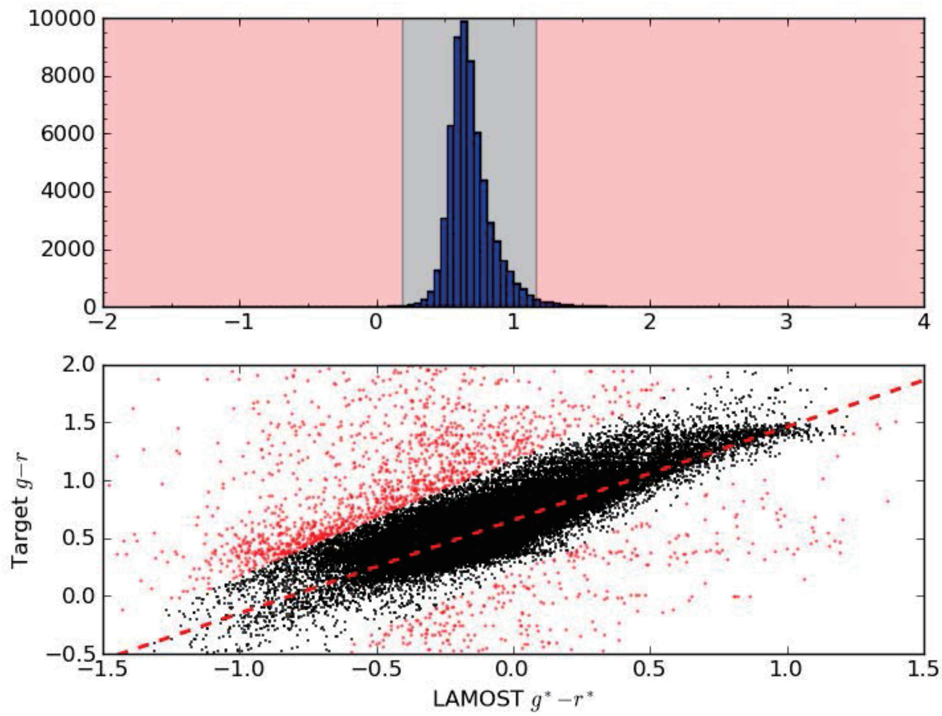

Fig. 3.— The relationship between g-r color and g^*-r^*. The X value is the g^*-r^* color and the Y value is the g-r color in the target catalogue. The red line is derived best-fit expression shown in the formula 3, and the red points are excluded outliers while deriving the expression.

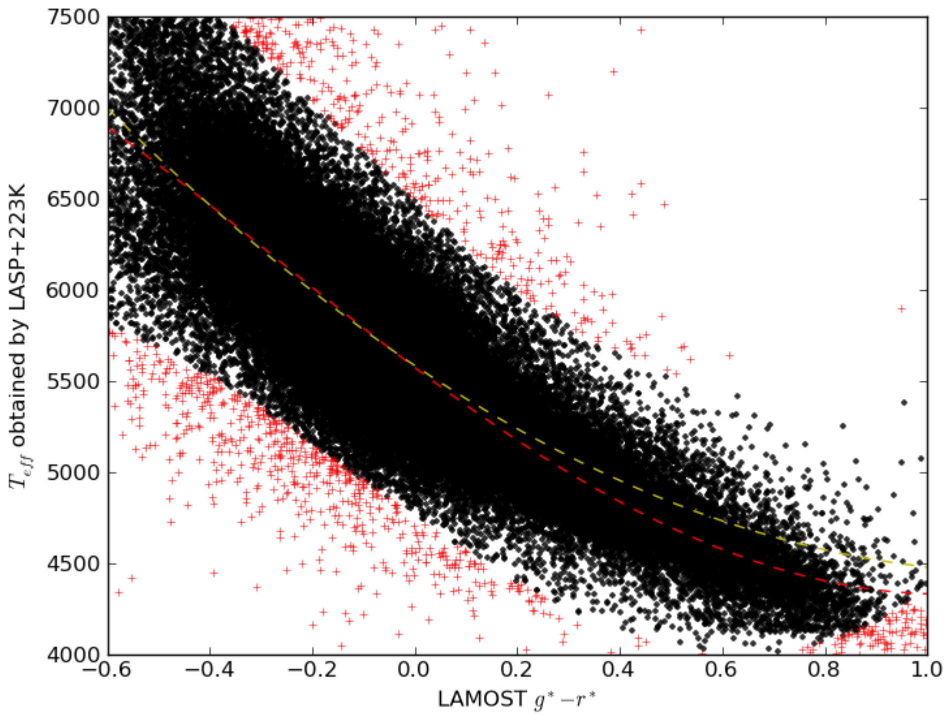

Fig. 4.— The relationship between the g^*-r^* color and the T_{eff}. The T_{eff} is added by 223K to decrease the system inconsistency between the SSPP and the LASP (Wu et al. 2011). The yellow line is the expression as formula 4, the red line is derived best-fit expression shown in formula 5, and the points in red are excluded outliers while deriving the expression.

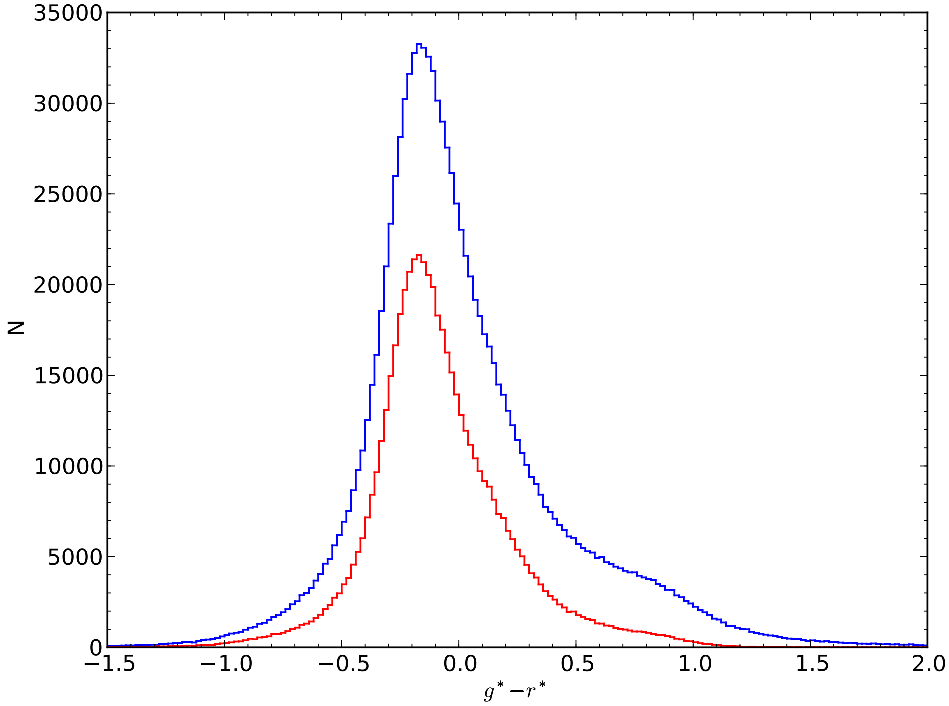

Fig. 5.— The number distribution of spectra in each g^*-r^* bin. The blue line is the distribution of all spectra while the red line is the distribution of the spectra with $SNR > 10$.

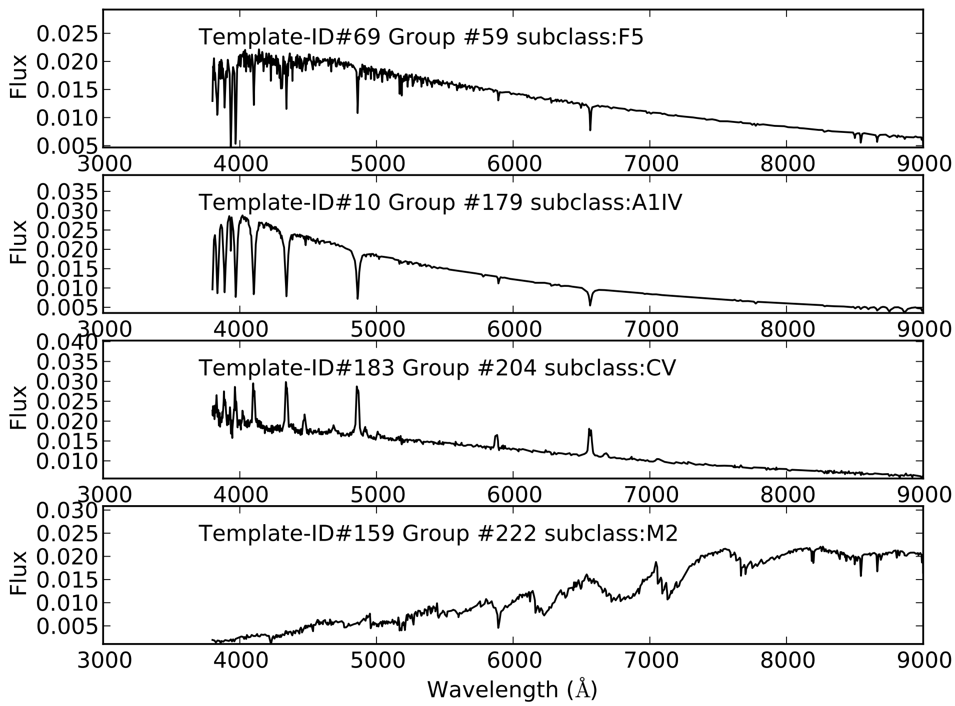

Fig. 6.— The template spectra #69, #10, #183 and #159.

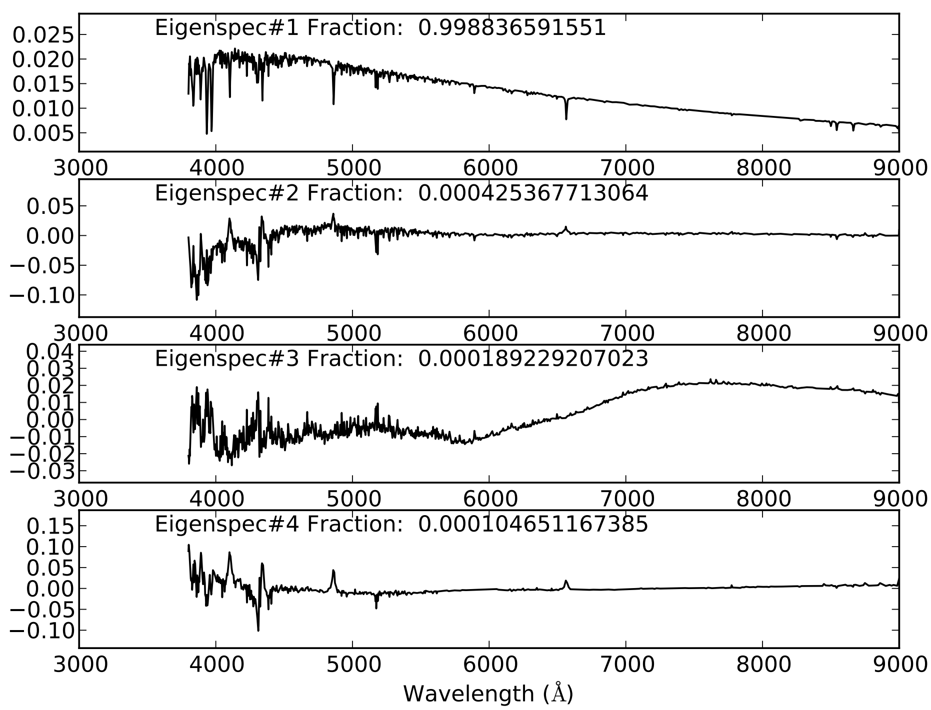

Fig. 7.— The first four eigen spectra (principal components) of group #59.

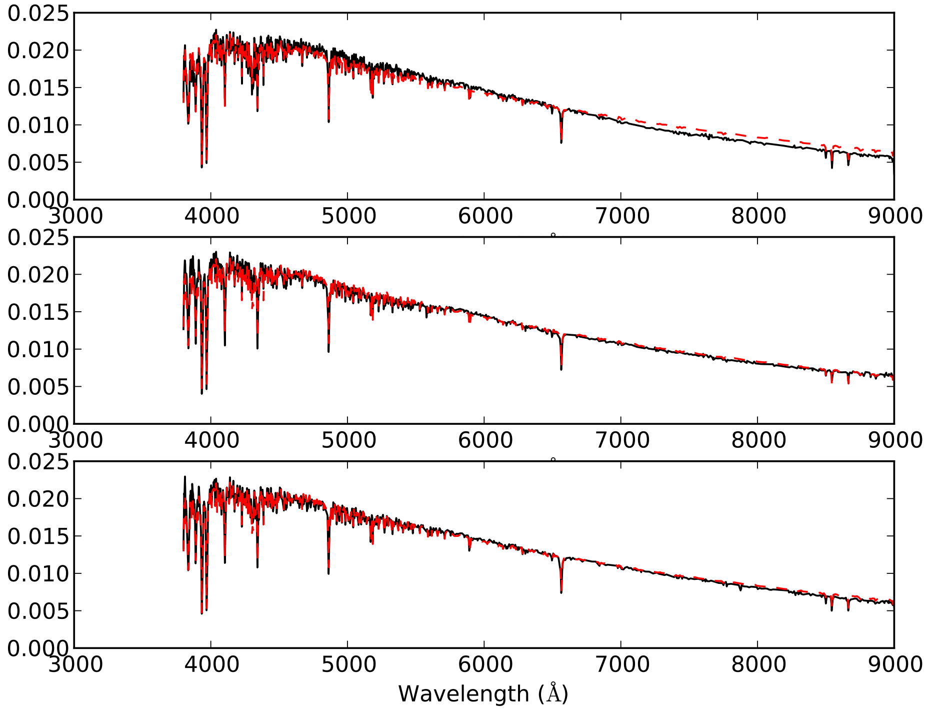

Fig. 8.— Three examples of reconstructed spectra in group #59. The black lines are the original spectra and the red lines are reconstructed spectra.

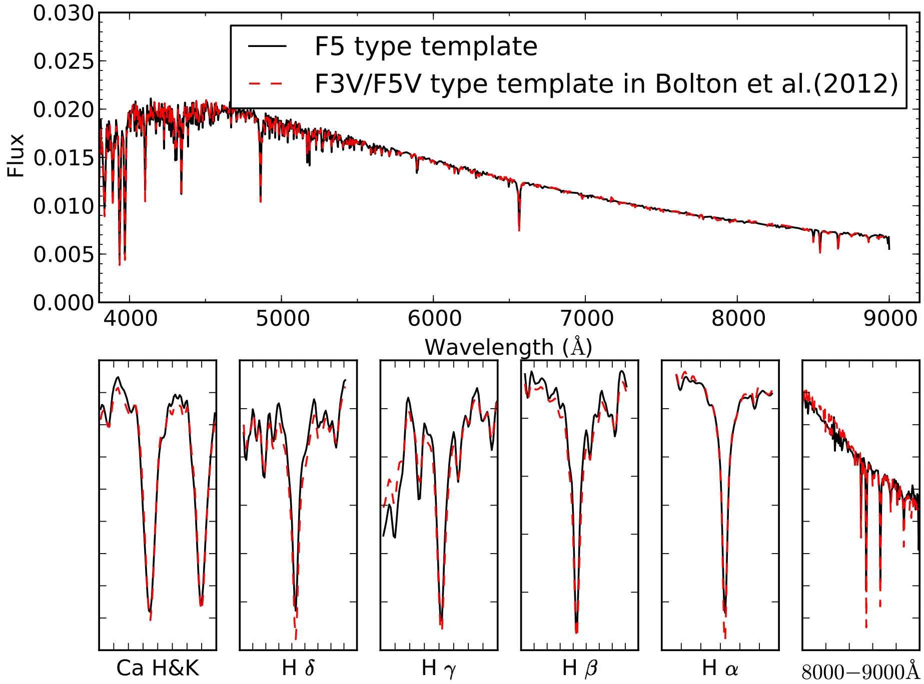

Fig. 9.— The comparison of the template spectrum #69 with F3V/F5V type spectrum in Bolton et al. (2012). The black line is the spectrum constructed in our work. The red one is the closest spectrum in Bolton et al. (2012).

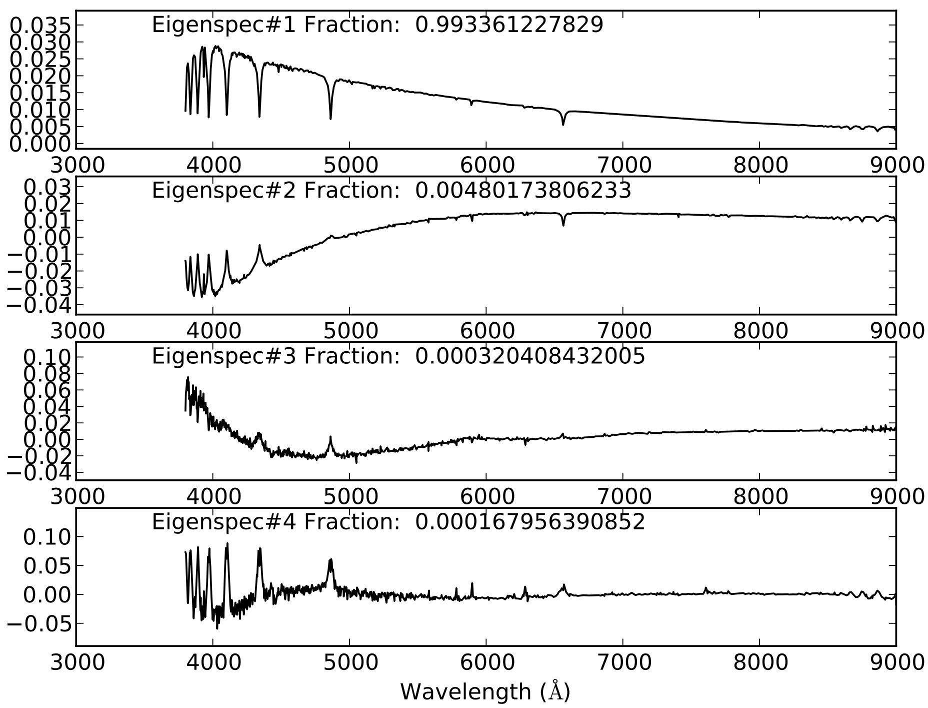

Fig. 10.— The first four eigen spectra (principal components) of group #179.

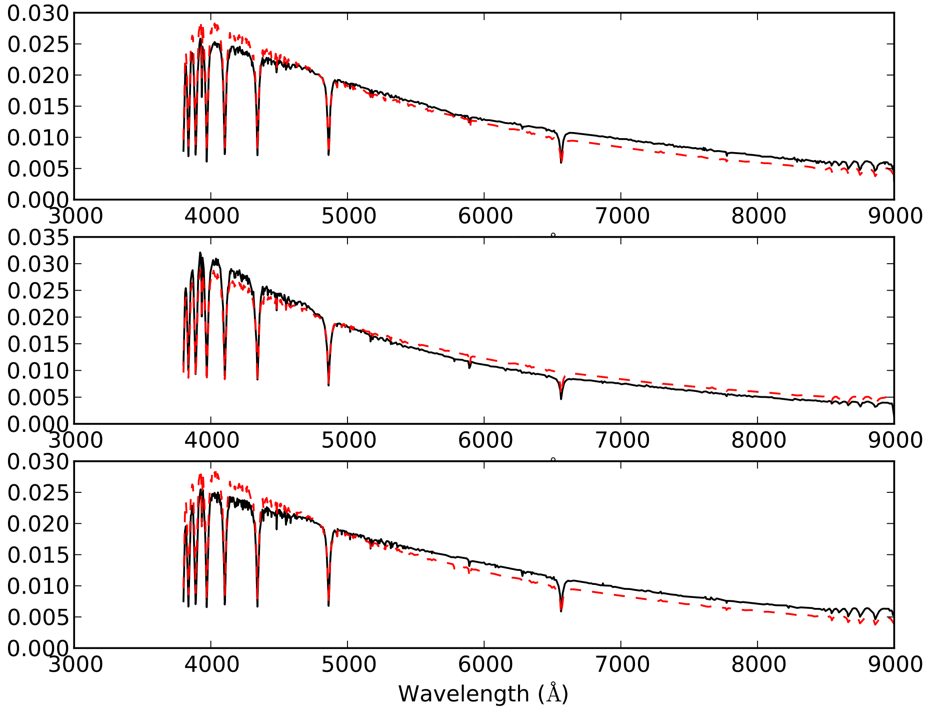

Fig. 11.— Three examples of reconstructed spectra in group #179. The black lines are the original spectra and the red lines are reconstructed spectra.

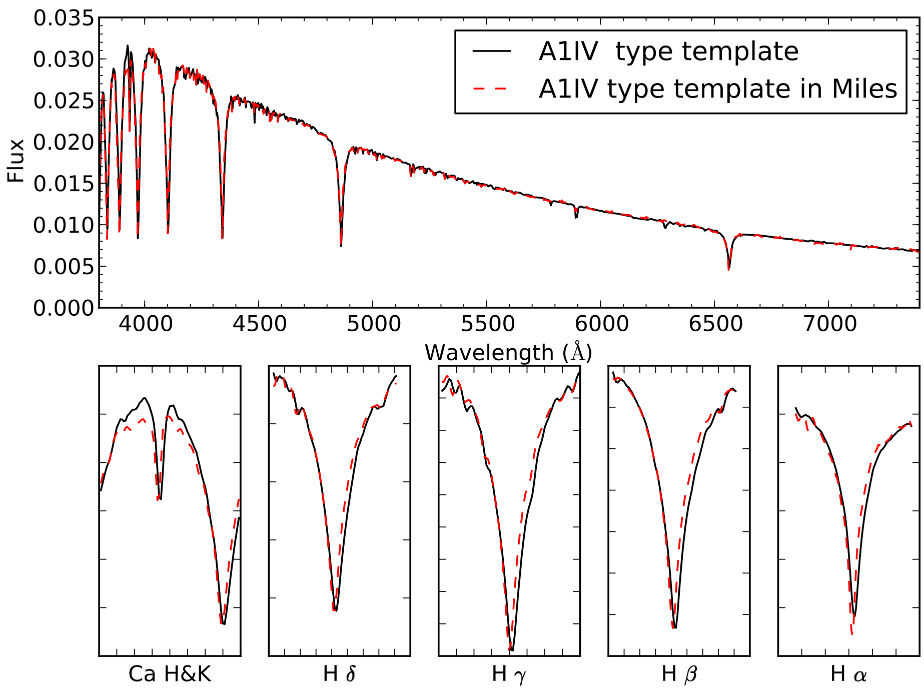

Fig. 12.— The comparison of the template spectrum #10 with A1IV type template in current library. The black line is the spectrum constructed in our work. The red one is the closest spectrum in current library.

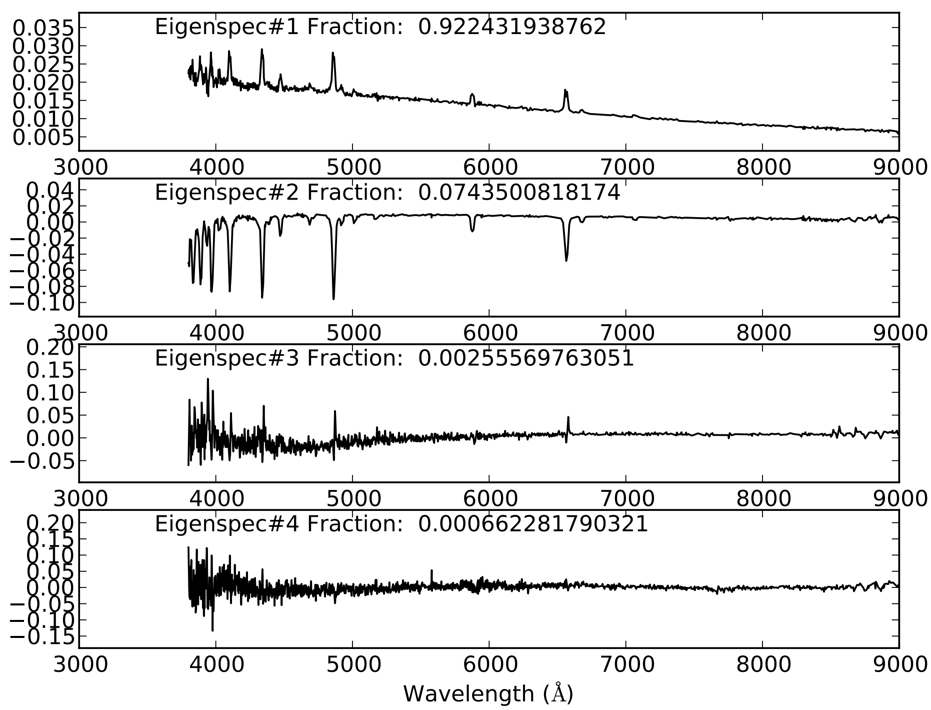

Fig. 13.— The first four eigen spectra (principal components) of group #204. Note that the strong lines in eigen spectra#2 are emission lines not absorption lines.

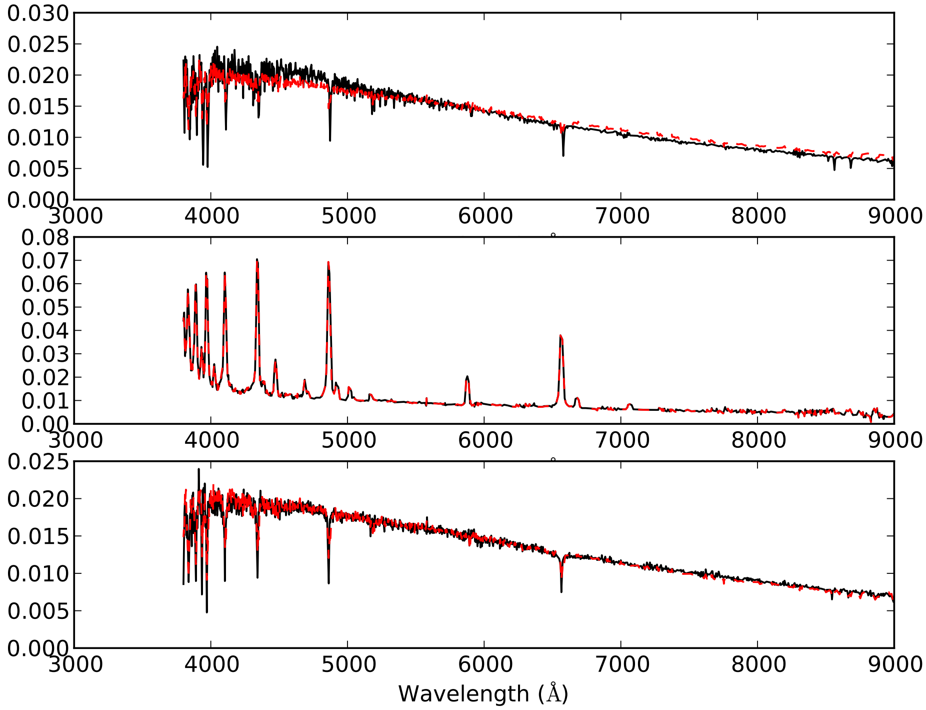

Fig. 14.— Three examples of reconstructed spectra in group #204. The black lines are the original spectra and the red lines are reconstructed spectra.

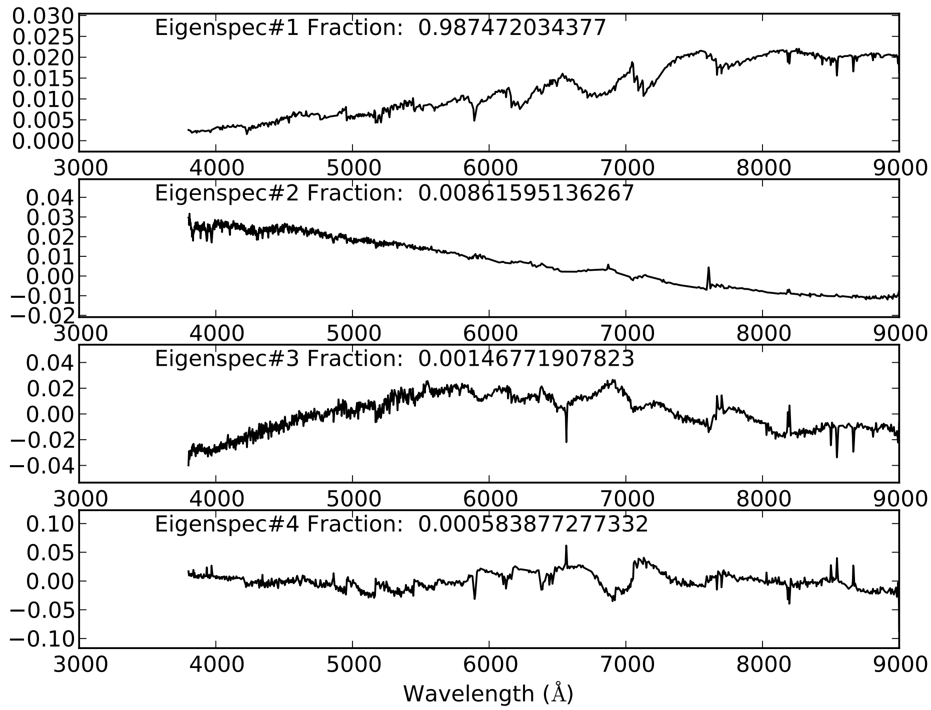

Fig. 15.— The first four eigen spectra (principal components) of group #222.

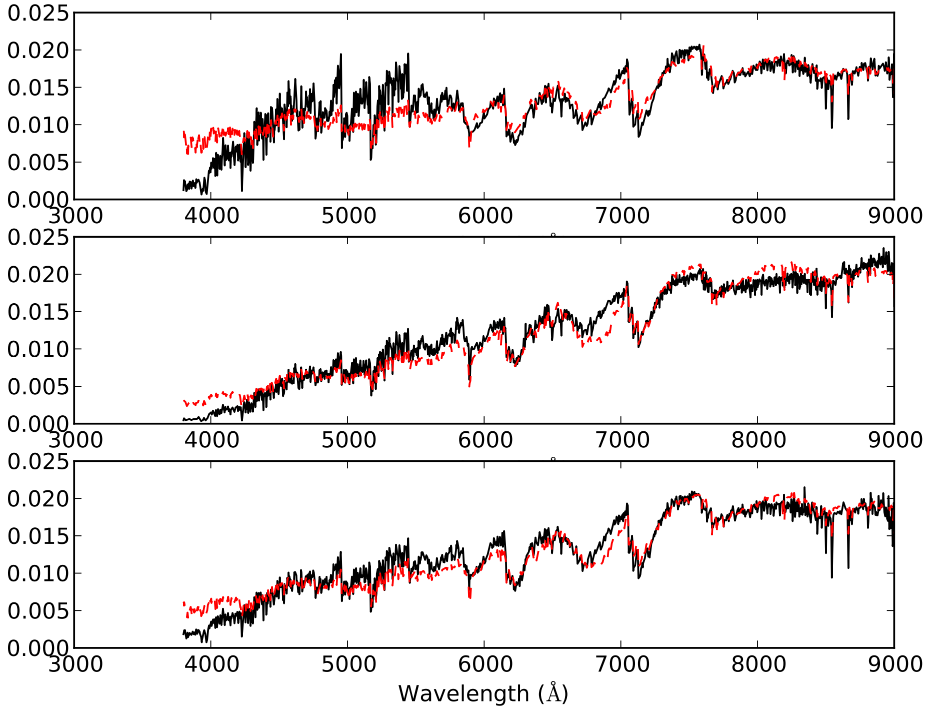

Fig. 16.— Three examples of reconstructed spectra in group #222.

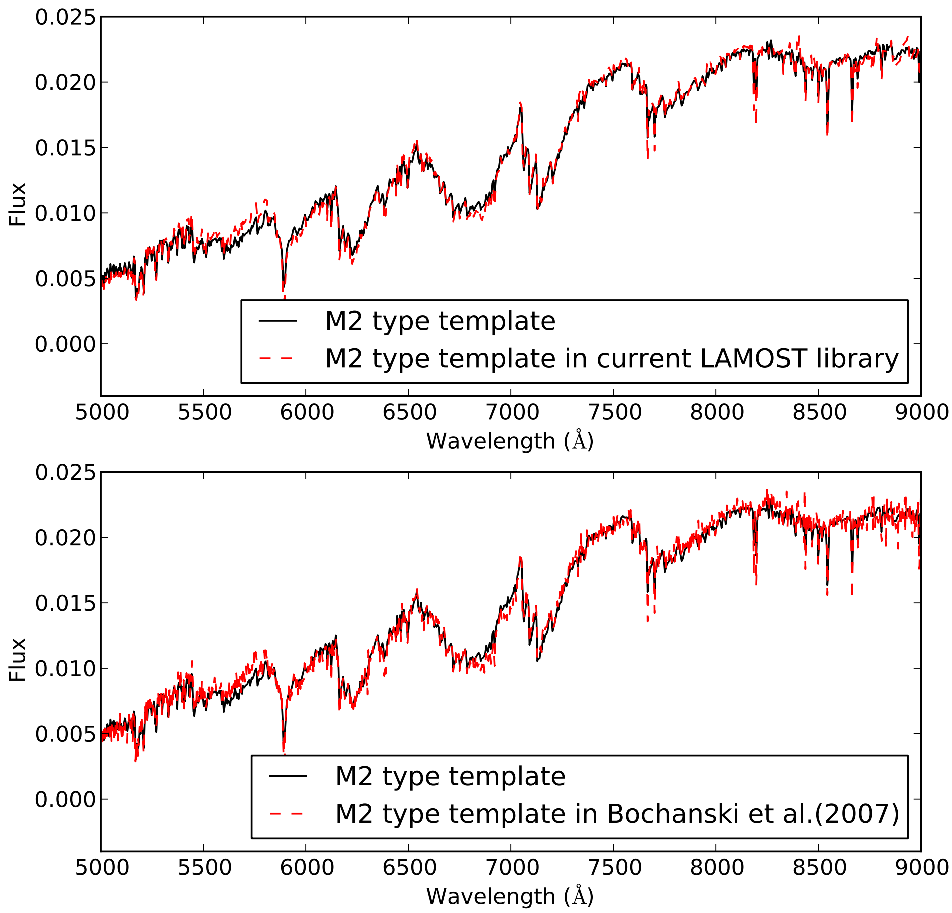

Fig. 17.— The comparison of the M2 type template spectrum #159 with M2 in current LAMOST library and Bochanski et al. (2007). The black lines are the M2 type spectrum constructed in our work. The red ones are the M2 type template in current LAMOST library and Bochanski et al. (2007).

Table 1. The main inforamtion of the finally combined template library

Spectra Type	Subclass ID	Subclass	C1	C2	Template IDs
O	1	O	0	1	1
	2	OB	0	1	2
B	3	B6	0	1	3
	4	B9	0	1	4
A	5	A0III	1	1	5
	6	A1IV	1	1	6
	7	A1V	4	4	7-10
	8	A2IV	3	3	11-13
	9	A2V	18	18	14-31
	10	A3IV	6	6	32-37
	11	A3V	2	2	38-39
	12	A5V	3	3	40-42
	13	A6IV	2	2	43-44
	14	A6V	3	3	45-47
	15	A7III	1	1	48
	16	A7IV	2	2	49-50
	17	A7V	2	2	51-52
	18	A9V	1	1	53
F	19	F0	4	4	54-57
	20	F2	2	2	58-59
	21	F3	2	2	60-61
	22	F4	1	1	62
	23	F5	7	7	63-69
	24	F6	4	4	70-73
	25	F7	2	2	74-75
	26	F8	1	1	76
	27	F9	2	2	77-78
G	28	G0	2	2	79-80
	29	G1	1	1	81
	30	G2	3	3	82-4
	31	G3	3	3	85-87
	32	G4	2	2	88-89

Table 1—Continued

Spectra Type	Subclass ID	Subclass	C1	C2	Template IDs
	33	G5	2	2	90-91
	34	G6	1	1	92=
	35	G7	3	3	93-95
	36	G8	4	4	96-99
	37	G9	3	3	100-102
K	38	K0	3	3	103-105
	39	K1	4	4	106-109
	40	K2	2	2	110-111
	41	K3	4	4	112-115
	42	K4	4	4	116-119
	43	K5	8	8	120-127
	44	K7	11	11	128-138
M	45	M0	10	10	139-148
	46	M1	6	6	149-154
	47	M2	5	5	155-159
	48	M3	7	7	160-166
	49	M4	1	1	167
	50	M5	1	1	168
	51	M6	2	2	169
	52	M7	2	2	171-172
	53	M8	2	2	173-174
	54	M9	2	2	175-176
Other	55	Carbon	1	1	177
	56	Carbon_lines	1	1	178
	57	WD	1	1	179
	58	CarbonWD	0	1	180
	59	WDmagnetic	0	1	181
	60	Binary	1	1	182
	61	CV	1	1	183

Note. — C1 is the amount of templates with corresponding subclass in our constructed library. C2 is the amount of templates with corresponding subclass in the finally combined library.

Table 2. Comparison with Bolton et al. (2012)

Spectra type	A1	A2	A3
O	3	2	2
B	30	2	2
A	19	14	49
F	19	9	25
G	15	10	23
K	19	7	36
M	12	10	28
Others	6	7	7

Note. — A1 is the Amount of templates in Bolton et al. (2012). A2 is amount of subclasses in final library and A3 is the amount of templates in final library.

Table 3. The main information of template #69, #10, #183, #159

Template ID	Group ID	All spectra	Used spectra	Subclass	Subclass1	Subclass2	Subclass3
69	59	18534	3048	F5	F3V/F5V	F1V	F5
10	179	653	381	A1IV	A4V/A1V	A2V	A1V
183	204	27	13	CV	-	-	-
159	222	17231	325	M2	M1/M0	M1.5V/M3V	M2/M1

Note. — Subclass is the finally labeled MK class. Subclass2 is the best fit MK class with Danks & Dennefeld (1994). Subclass3 is the best fit MK class with Luo et al. (2013).