Research in Astronomy and Astrophysics manuscript no.
(LATEX: sample.tex; printed on October 10, 2013; 23:25)

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], Jingchang Pan[3], Fengfei Wang[1], Jiannan Zhang[1], Liangping Tu[1,4], Bin Jiang[3], Yongheng Zhao[1], Jianjun Chen[1,2], Xiaoyan Chen[1], Bing Du[1], Wen Hou[1,2], Ge Jin[6], Xiao Kong[1,2], Jie Liu[3], Juanjuan Ren[1,2], Yihan Song[1], Yue Wu[1], Haifeng Yang[1,2,5] and Zhenping Yi[1,2,3]

[1] Key Laboratory of Optical Astronomy,National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100012, China *lal@nao.cas.cn weipeng@nao.cas.cn*

[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 is one of LAMOST softwares to produce and analyze the final spectra and its aim is to classify and measure the spectra observed in the survey. Through the pipeline, the observed stellar spectra are classified into different sub-classes by matching with templates spectra. Consequently, the performance of the stellar classification is greatly influenced by the quality of templates. A new LAMOST stellar spectral classification templates library is constructed, which is supposed to improve the precision and credibility of the 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 spectra with the SDSS *ugriz* filter response curve II) the subclass given by the pipeline. In each group, the template spectra are constructed within three steps: I) Outliers are excluded using Local Outlier Probabilities (LoOP) algorithnm, 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 spectra are made as the template spectra in the groups. And we initially obtain stellar tempalte spectra in 216 groups. All template spectra are visually inspected, and 52 spectra are abadoned due to low spectral quality.

Furthermore, the MK classification for each template spectrum is manually determined by comparing with three libraries of label-known templates with known MK class. Meanwhile, some unlabeled or wrongly labeled spectra are relabeled or abandoned. And we finally obtain 164 new template spectra with 65 different MK classes. The template library is composed by the spectra left and the first version contains 164 spectra and 65 different MK classes.

Key words: 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.

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 final 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 similiarity of the LAMOST stellar spectra with other survey spectra, some 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 library contains 36 stellar subclasses plus 20 subclasses specially for A-type star. The former 36 templates are constructed from a set of SDSS spectra (Wang et al. 2010). while the latter 20 A-type template spectra are picked out from MILES library . These template spectra cover nearly all common types of stellar spectra in the survey. Matching with these templates, the LAMOST stellar spectra are classified as 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 resolutoins, 1800 , 2000 and 2000 respectively. Secondly, different instrumental designs bring about differnt effects on the spectra. In addition, the processes of spectrum extraction, wavelength calribration and flux calribration (Bai 2012) are also different. Considering these issues which can not be ignored, it is very necessary to construct a new template library based on the spectra observed and processed by LAMOST.

In this paper, we described in detail the construction of the new LAMOST stellar classification template library. The paper is organized as follows: Section 2 detailedly describe 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.

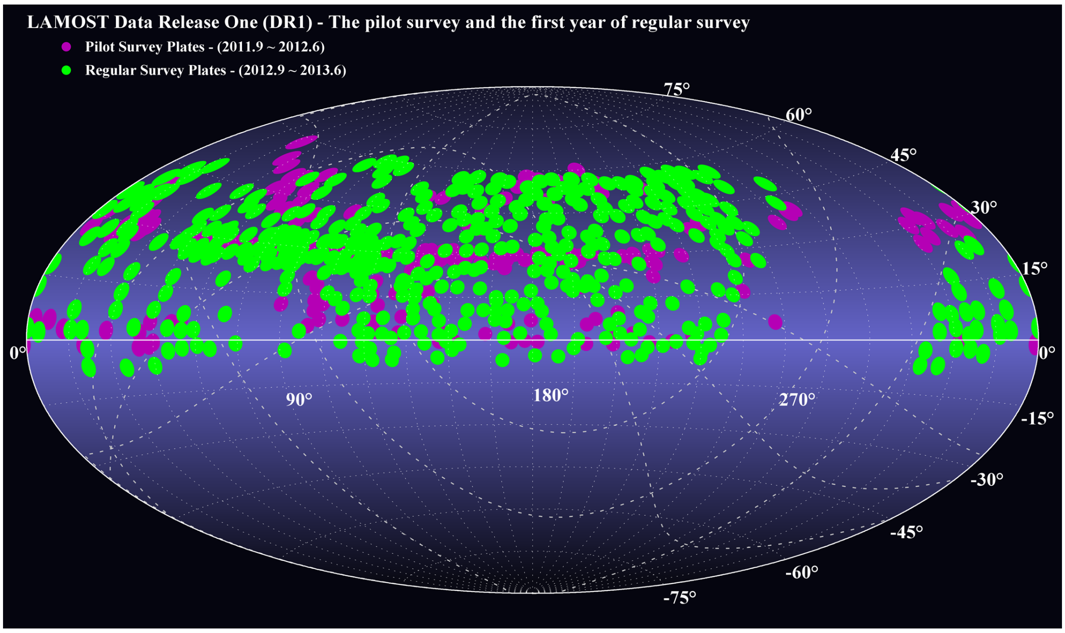
LAMOST Data Release One (DR1) - The pilot survey and the first year of regular survey
● Pilot Survey Plates - (2011.9 ~ 2012.6)
● Regular Survey Plates - (2012.9 ~ 2013.6)
75°
60°
45°
30°
15°
0°
0°
90°
180°
270°
-15°
-30°
-45°
-60°
-75°

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

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. To extract spectra from raw observation data, the raw data have been reduced with LAMOST 2D pipeline (Bai 2012) including bias subtraction, cosmic-ray removal, spectral trace and extraction, flat-fielding, wavelength calibration sky subtraction, and combination. Then the 1D pipeline (Wang et al. 2010; Luo et al. 2012) gives spectral type and redshift (radial velocity for stellar spectra). Considering the effect of interstellar dust extinction on the spectra and the closeness of stars, a mount of 855,583 spectra in the Galactic Anti Center and M31(Liu et al. 2013) whose plate name in the catalogue starts with 'GAC' or 'M31' are excluded. And then 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: the proposed $g^* - r^*$ color and the subclass labeled by the pipeline.

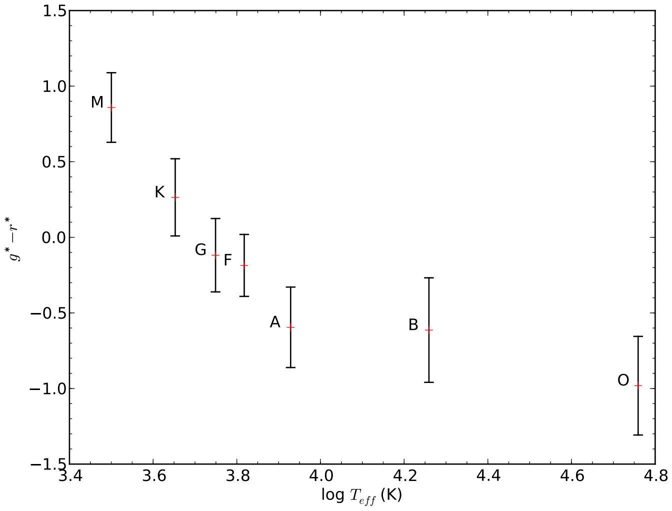

Fig. 2 The average value and standard deviation of g^*-r^* for each spectral type. The X value of the error bar is the median effective temperature in theory. The y value of the center of each error-bar is the average g^*-r^* color and the half length is the standard deviation of g^*-r^*.

2.2.1 The pseudo g-r color

LAMOST is a spectroscopic survey oriented telescope and its photometric data are from different catalogs of other surveys. Meanwhile, the flux calibration of LAMOST spectra is relative (Bai 2012). Therefore, accurate and uniform colors can not be obtained for LAMOST spectra. Inorder to slove 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^*:

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

The g-r color is a very good indicator of stellar surface effective temperature (Teff) (Lee et al. 2008; Željko 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 relationship between the $g^* - r^*$ and the Teff exists. The average value and standard deviation of g^*-r^* for different spectral types (O, B, A, F, G, K, M-type) are calculated. And as shown in Figure 2, 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 a obvious linear relationship between the two colors, and the derived best-fit expression is shown in formula 2:

$$g - r = 0.807 * (g^* - r^*) + 0.655 \qquad (2)$$

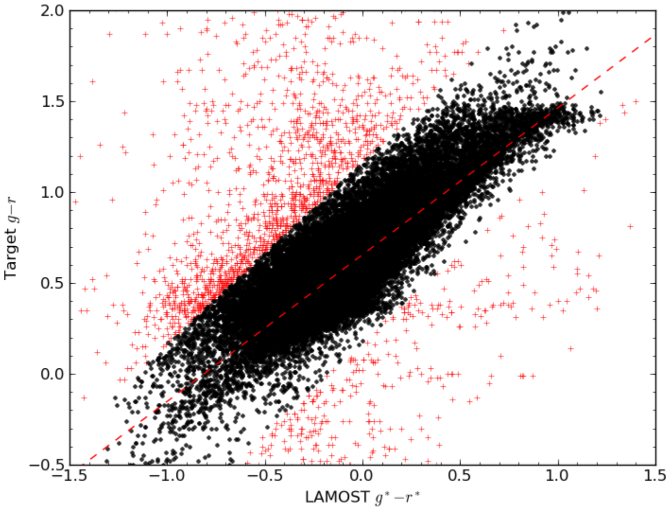

Fig. 3 The relationship between g-r color and g^*-r^*. The X values are the g^*-r^* colors and the Y values are the g-r colors in the target catalogue. The red line is derived best-fit expression as formula 3, and the red points are excluded outliers while deriving the expression.

Željko Ivezić et al. (2008) derived a relation between the Teff and the color g-r in the range of $-0.3 < g - r < 1.3$ For SDSS spectra:

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

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

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

For 5,220,138 A,F, G and K-type spectra, their effective temperatures , surface gravities and metallicities determined by the LAMOST Stellar Parameter pipeline (LASP, see Wu et al. (2011)) are provided. 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:

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

As shown in Fig.4, the formulas 4 and 5 nearly coincide with each other in the range of Teff [5500K,7000K]. Thus, the defined g^*-r^* color is also be a good indicator of Teff, which is used as a criterion of dividing groups of the slected spectra.

2.2.2 *Group dividing criteria*

To construct different kinds of templates, we gather these spectra in 233 different groups by the proposed g^*-r^* color and the stellar subclass classified by the pipeline using the current template library.

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

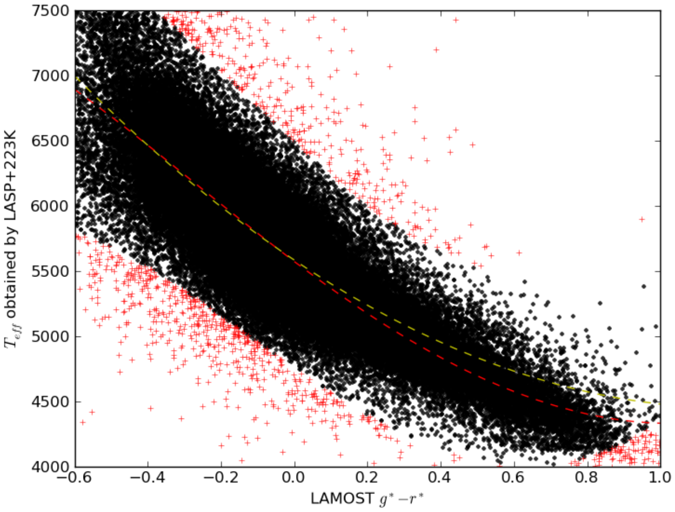

Fig. 4 The relationship between the g^*-r^* color and the T_{eff}. The T_{eff} is added by 223K to decrease the system inconsistency between SSPP and LASP (Wu et al. 2011). The yellow line is the expression as formula 4. The red line is derived best-fit expression as formula 5, and the points in red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

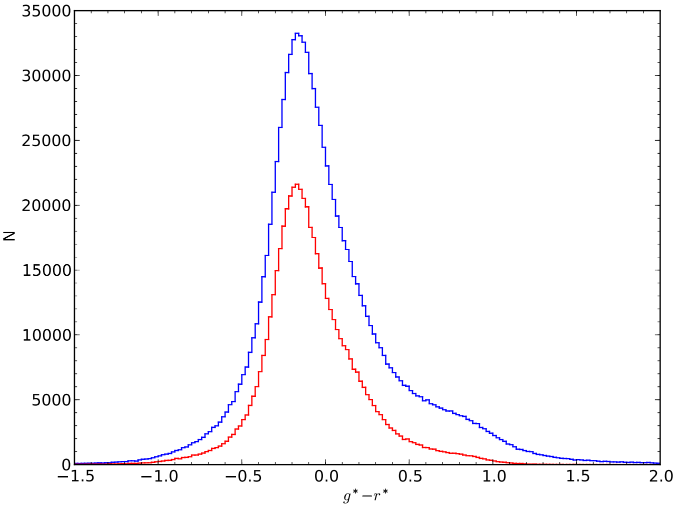

Fig. 5 The number distribution of spectra in each g^*-r^* bin. The blue line is the distribution of all spectra while the red line is the distribution of the spectra with $SNR > 10$.

In addition, other 60 groups are formed by the subclass labeled by current pipeline. After the automated processing of LAMOST spectra analysis pipeline and visual inspection, there are 60 different stellar subclasses in our selected spectra. 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 LAMOST Pilot Survey, and we mark these spectra with group id from 220 to 229. Zhao et al. (2013) presented a spectroscopically identified catalog of 70 DA white dwarfs (WDs). Meanwhile, Zhang et al. (2013) identified 230 other DA white dwarfs, and we

Table 1 The number distribution of different subclasses

Group ID	Subclass	Amount	Group ID	Subclass	Amount	Group ID	Subclass	Amount
176	A0	94	196	A9	4	216	K3	77164
177	A0I	27	197	A9V	2170	217	K5	73045
178	A0III	407	198	B	15	218	K7	45839
179	A1IV	653	199	B0	1	219	K9	4
180	A1V	527	200	B9	443	220	M0	19152
181	A2I	10	201	Binary	170	221	M1	19953
182	A2IV	1692	202	Carbon	178	222	M2	17243
183	A2V	5761	203	CarbonWD	6	223	M3	9749
184	A3I	37	204	CV	27	224	M4	3860
185	A3IV	2084	205	EM	63	225	M5	855
186	A3V	2080	206	F0	27808	226	M6	412
187	A4III	926	207	F2	44192	227	M7	259
188	A4V	773	208	F5	119328	228	M8	47
189	A5	26	209	F9	292830	229	M9	66
190	A5I	213	210	G0	47697	230	O	79
191	A5V	1253	211	G2	92229	231	OB	16
192	A6IV	1240	212	G5	81202	232	WD	535
193	A6V	322	213	G7	3650	233	WD Magnetic	14
194	A7III	4033	214	K0	1998			
195	A7V	647	215	K1	85218			

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. The distribution of these groups is as shown in Table 1.

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 similiar spectra following two criteria. Although the spectra in the same group are very similiar with each other, there are still some outliers existing in each group for many reasons, including the effect of instellar extinction on the continiuum, strong noises, existence of unusual spectral features and other issues. Obviously, these outliers should be excluded to generate much purer spectra for construction of template spectra.

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 viable tool, Principal Component Analysis (PCA) has been applied in the classification of spectra (Whitney 1983; Bailer-Jones et al. 1998; Yip et al. 2004; Almeida & Prieto 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 for spectra analysis pipeline (note that the number in an bracket is the number of remaining spectra after this step):

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 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 remaining spectra F 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 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). 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 Labeling 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 lackness 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 label stellar spectral subclass for these constructed template spectra. To get better stellar spectral subclass, we use following two steps to label these template spectra.

1. First, each template spectrum are matching with three libraries and the first four closest spectra in each library are chosen. That is to say, there are 12 differnt spectra from three differnt libraries for each template spectrum in our library. The three libraries used 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 of our templates to R 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 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 the library .
   - As introduced before, the current library used for stellar classification in LAMOST spectra analysis pipeline contains 36 classes plus 20 subclasses specially for A-type star. The resolution R of these 56 spectra 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 the library .
2. Visual inspection is carried out after automatic matching with spectra library. 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 best matched spectra visually chosen . Meanwhile, those template spectra with bad data or low SNR are excluded. Finally, there are 164 spectra and 65 different MK classes are left in the template library. And these spectra are publicly available on the web site[1].

2.4 The new template library used in the LAMOST spectra analysis pipeline

A new template library used in the LAMOST spectra analysis pipeline is fromed by combining the newly constructed templates and templates in current pipeline. These subclasses whose template spectra have been newly constructed are removed from the current library, and the template spectra of other subclasses are added into newly constructed template library in ourwork. The current library has been used in the new new version of LAMOST spectra analysis pipeline for spectra after data release one.

[1] http://sciwiki.lamost.org/lamost_sctl/v1

Table 2 The main information of groups 59,179, 204 and 222

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

Notes: Subclass is the finally labeled MK class. Subclass1 is the best fit Mk class with Bolton et al. (2012). Subclass2 is the best fit Mk class with Danks & Dennefeld (1994). Subclass3 is the best fit Mk class with Luo et al. (2013).

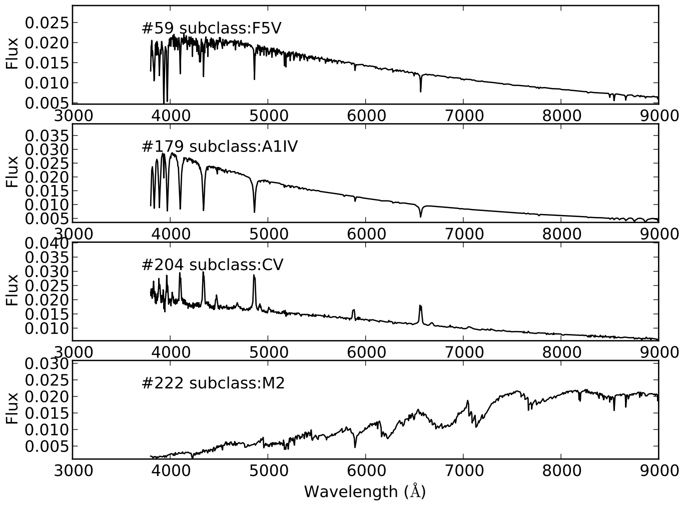

Fig. 6 The template spectra of groups 59,179, 204 and 222.

3 RESULTS AND DISCUSSIONS

3.1 Examples

Here we choose four typical groups (Group 59,179 ,204 and 222) to discuss the construction process and the constructed template spectra in detail. The main information of these groups is shown in Table 2. The MK classes are F5V, A1IV, CV and M2 respectively. The finally constructed template spectra of these groups are shown in Fig.6.

Group:59 This group 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 which are used in the spectra reconstruction. 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). Afetr excluding outliers, 3,048 spectra are left to construct the template spectrum. The stellar spectral subclass is finally labled as 'F5V'. As shown in Fig 9, the template spectrum is close to the F3V/F5V type spectrum in Bolton et al. (2012).

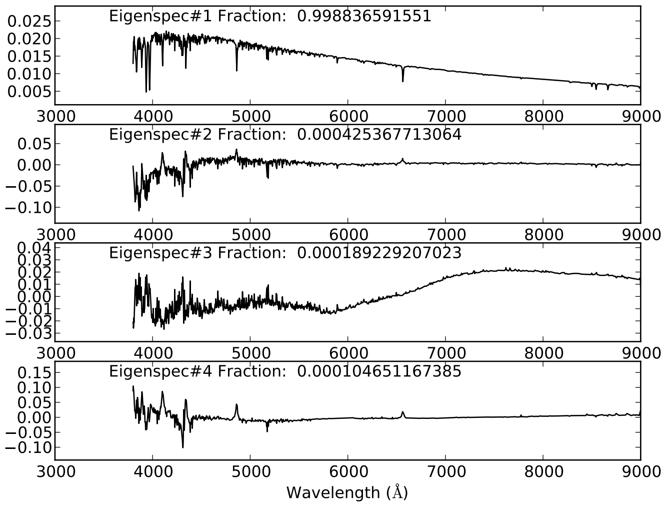

Fig. 7 The first four eigen spectra (principal components) of group 59.

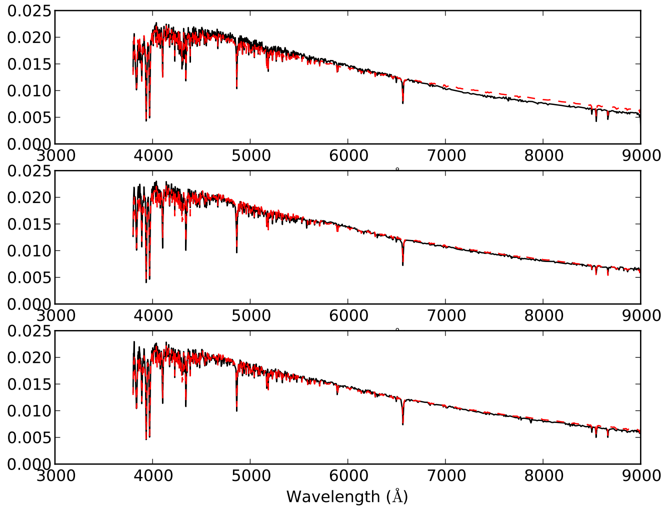

Fig. 8 Three examples of reconstructed spectra in group 59. The black lines are the original spectra and the red lines are reconstructed spectra.

Group:179 This group 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. Afetr excluding outliers, 381 spectra are left to construct the template spectrum. The stellar spectral subclass is finally labled as ‘A1IV’,

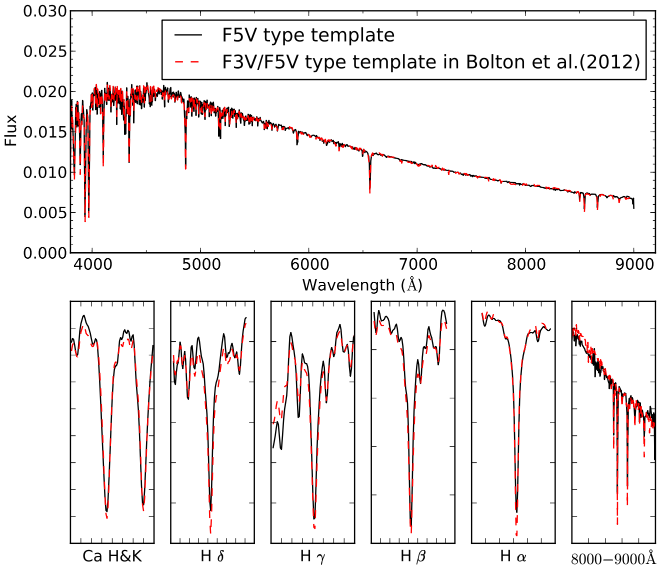

Fig. 9 The comparison of the template spectrum in group 59 with F3V/F5V type spectrum in Bolton et al. (2012). The black line is the spectrum constructed in our work. The red one is the closest spectrum in Bolton et al. (2012).

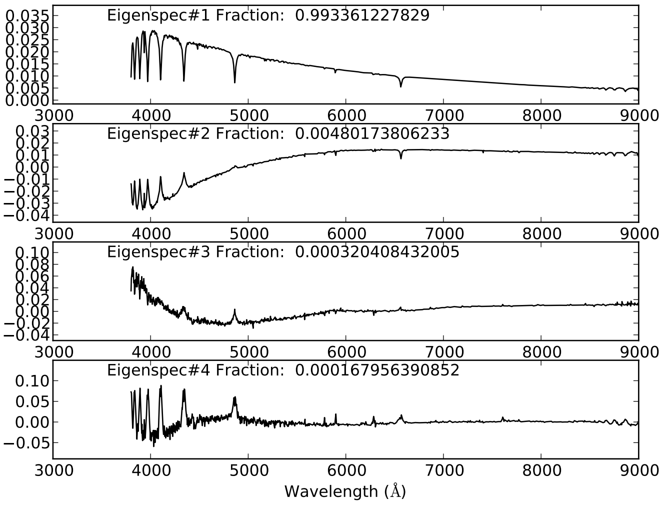

Fig. 10 The first four eigen spectra (principal components) of group 179.

which coincides with the group selection criteria. The template spectra of group 179 is shown in Fig 12, and we cans see that the SNR of the template is a little larger than the template in current library.

Group:204 This group 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 which are used in the spectra reconstruction. Compared with normal stars, the spectra of CV

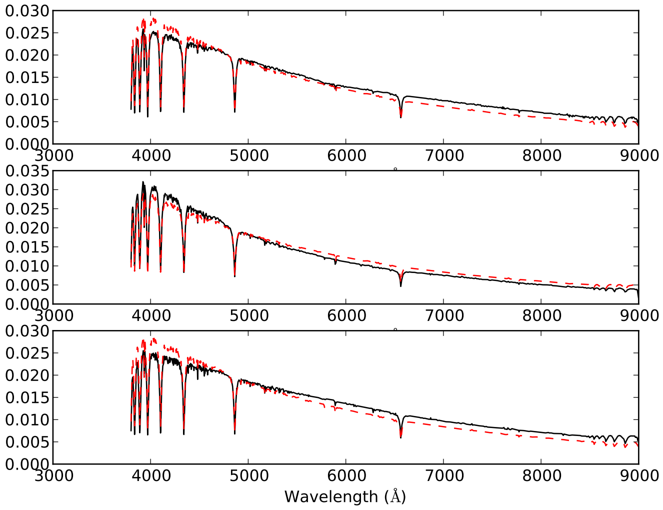

Fig. 11 Three examples of reconstructed spectra in group 179. The black lines are the original spectra and the red lines are reconstructed spectra.

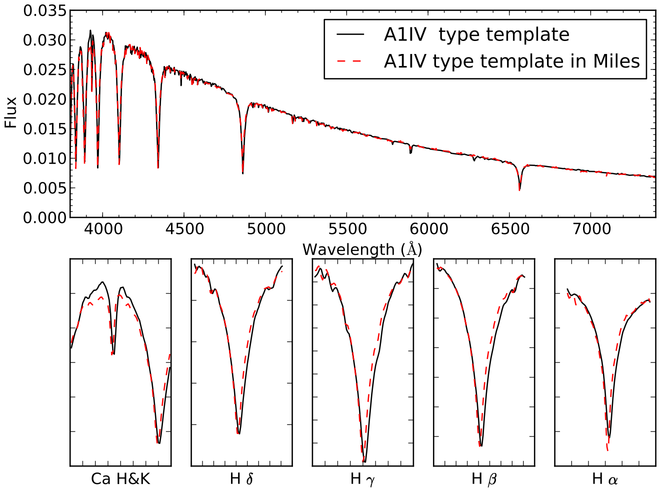

Fig. 12 The comparison of the template spectrum in group 59 with A1IV type template in current library. The black line is the spectrum constructed in our work. The red one is the closest spectrum in current library.

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 normal 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. Afetr 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 labled as ‘CV’, which coincides with the group selection criteria.

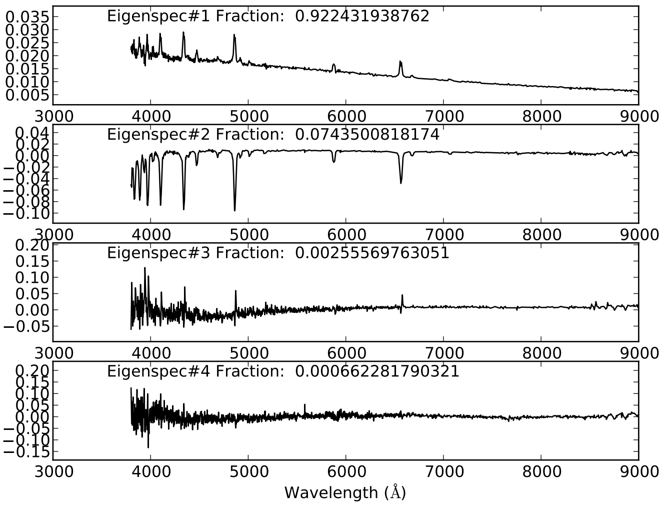

Fig. 13 The first four eigen spectra (principal components) of group 204. Note that the strong lines in eigen spectra#2 are emission lines not absorption lines.

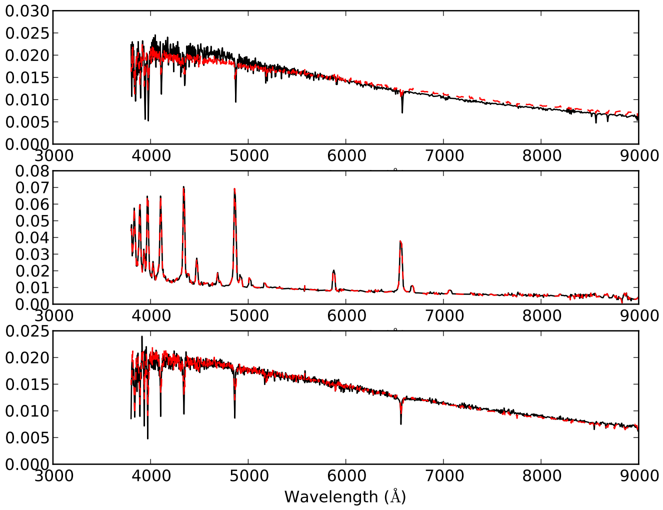

Fig. 14 Three examples of reconstructed spectra in group 204. The black lines are the original spectra and the red lines are reconstructed spectra.

Group:222

This group contains the spectra classified as 'M2' by current LAMOST spectra analysis pipeline.

There are totally 17,231 spectra in this group. Due to the existence of wavelength points with $flux \leq 0$, a large amount of spectra are excluded and 325 spectra are selected from the first 5,000 spectra with the highest SNR.

The spectrum is labeled as 'M2' following the group selection criteria.

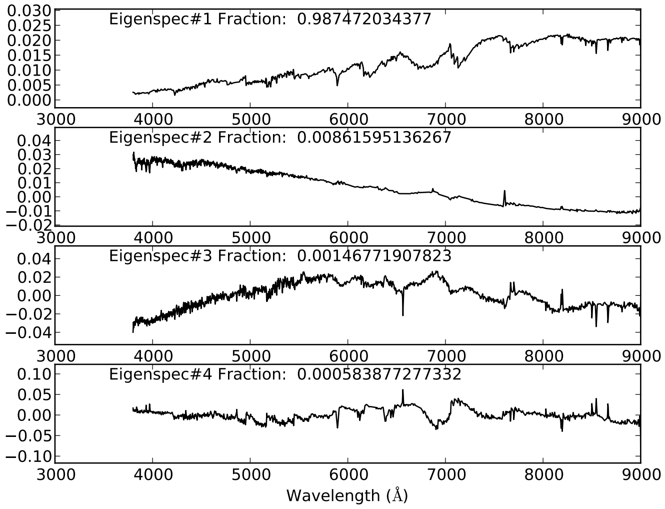

Fig. 15 The first four eigen spectra (principal components) of group 222.

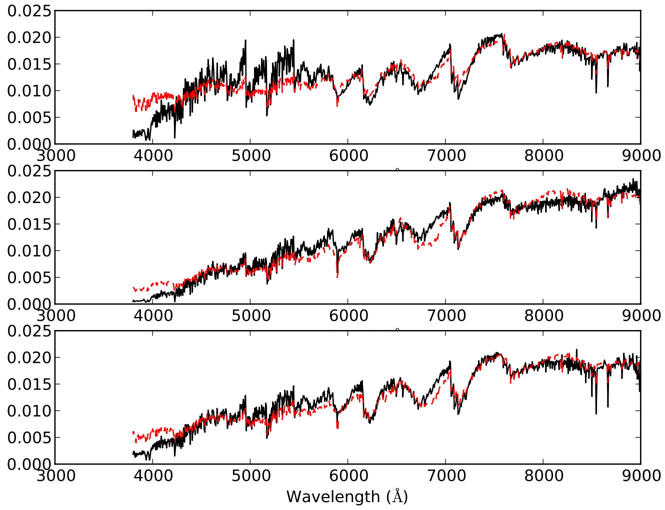

Fig. 16 Three examples of reconstructed spectra in group 222.

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 are not well reconstructed in the blue arm (as shown in Fig 16). In spite of this, the template spectrum is also well constructed.

To check the quality, we choose the M2-type template spectrum in the current template library (Wang et al. 2010) and compare it with the template spectrum of group 222 (see Fig 17 upper panel). Bochanski et al. (2007) presented template spectra of low-mass (M0-L0) dwarfs derived from over 4000 Sloan Digital Sky Survey spectra. We choose the M2-type template spectrum and alsoe compare it with the template spectrum of group 222 (see Fig 17 bottom panel). As shown in Fig 17, we can infer that our constructed M2-type spectrum is a little better than these two spectra.

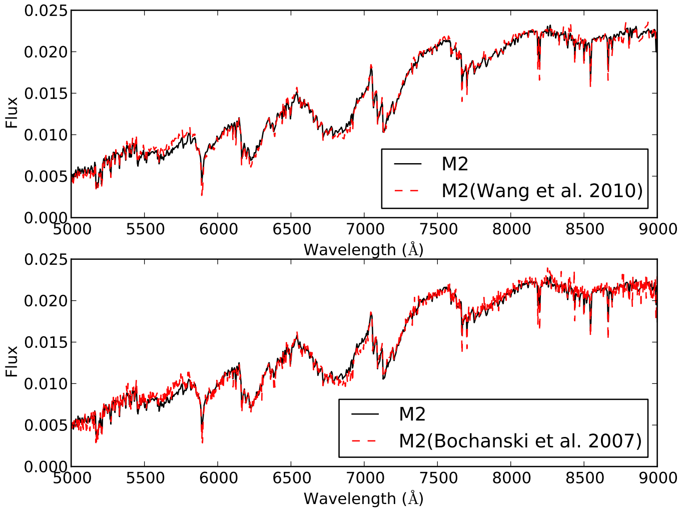

Fig. 17 The comparison of the template spectrum in group 59 with M2 in Bochanski et al. (2007). The black line is the spectrum constructed in our work. The red one is the closest spectrum in Bochanski et al. (2007).

3.2 Discussions

3.2.1 Comparison with McGurk et al. (2010)

McGurk et al. (2010) applied PCA to about 100,000 SEGUE spectra by dividing all spectra into 55 different bins. For each bin, the first four eigenspectra are published and the first one is a high SNR mean spectra. For the template spectra in all groups, to check the difference, we use similar method as MK class labeling to find three closest mean spectra in McGurk et al. (2010). As our color range is wider than McGurk et al. (2010), not all template spectra are similar with these in McGurk et al. (2010). As shown in Table 3, the groups with groups id from 38 to 99 (totally 60 groups, not including group 71) cover nearly all mean spectra constructed by McGurk et al. (2010). The finally subclasses are from A3 to K3, which coincides with sayings in McGurk et al. (2010).

3.2.2 Comparison with current templates of LAMOST spectra analysis pipeline and Bolton et al. (2012)

As supplements to current templates, our constructed templates replace most spectra in the current library including the A-type stellar spectra. From the comparisons discussed in section 3.1, we can infer that newly constructed template spectra are a little better than current ones.

There are 123 stellar subclasses in Bolton et al. (2012). These template spectra are individual spectra in Indo-U.S. database. We notice that there are 80 subclasses which contain less than 500 spectra in SDSS DR9. There are totally 12,897 spectra (about 1.66% in all 773,275 spectra) and 1,062 spectra (about 0.22% in all 475694 spectra) with SNR$>$ 10. Meanwhile, the average SNR of these spectra is about 4.93 which is much less the one of all spectra. In other words, there are mainly 43 stellar subclasses containing most spectra especially these spectra with high SNR.

Table 3 The comparison with McGurk et al. (2010)

Group ID	ESID1	ESID2	ESID3	Group ID	ESID1	ESID2	ESID3	Group ID	ESID1	ESID2	ESID3
38	1	2	3	59	19	20	18	80	39	38	40
39	1	2	3	60	19	20	21	81	40	41	39
40	2	3	1	61	20	23	22	82	41	40	42
41	3	2	4	62	24	23	25	83	42	41	43
42	4	5	3	63	25	24	26	84	43	42	44
43	5	6	7	64	26	25	27	85	44	43	45
44	6	7	5	65	27	26	28	86	45	44	46
45	7	8	6	66	28	27	29	87	45	46	44
46	8	9	7	67	28	29	27	88	46	47	45
47	9	10	11	68	29	30	28	89	47	48	46
48	11	10	12	69	30	29	31	90	48	47	49
49	12	11	10	70	30	31	29	91	49	48	50
50	12	13	11	71	-	-	-	92	50	49	51
51	13	14	12	72	32	33	31	93	51	50	52
52	14	15	13	73	33	32	34	94	52	51	53
53	15	14	16	74	34	35	33	95	53	52	54
54	16	15	17	75	35	34	36	96	53	54	52
55	16	17	15	76	35	36	37	97	54	55	53
56	17	18	16	77	36	37	35	98	55	54	53
57	18	17	19	78	37	38	36	99	55	54	53
58	18	19	20	79	38	37	39				

Notes: The ESID1, ESID2 and ESID3 are bin ID of the first three closest eigen spectra in McGurk et al. (2010) respectively.

Considering the differences between LAMOST spectra with spectra in other survey, these newly constructed spectra are more reliable and more similar to the spectra observed in LAMOST survey. That is because these template spectra are constructed from a healthy sum of spectra from LAMOST DR1.

3.2.3 Remaining problems

The result shows that 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 types 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 label our template spectra. However, how to label them better is still a remaining 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 labeled by current LAMOST spectra analysis pipeline. Through the proposed contruction 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 labeled with a MK type by comparing with three libraris and visual inspection, and some low-quality spectra are excluded afetr visual inspection . Meanwhile, some unlabeled or wrongly labeled spectra are relabeled or abandoned. Finally, the new stellat classification templates library LAMOST spectra analysis pipeline consists of 164 spectra and 65 different MK classes. The new templates library has been used for new version of LAMOST Spectra Analysis Pipeline and is published on the website [2].

Acknowledgements 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.

References

Almeida, J. S., & Prieto, C. A. 2013, ApJ, 763, 50 8

Bai, Z. 2012, Proceedings of the International Astronomical Union, 8, 189 2, 3, 4

Bailer-Jones, C. A., Irwin, M., & Hippel, T. V. 1998, MNRAS, 298, 361 8

Bochanski, J. J., West, A. A., Hawley, S. L., & Covey, K. R. 2007, AJ, 133, 531 15, 16

Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144 9, 10, 12, 16

Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197 2

Danks, A. C., & Dennefeld, M. 1994, PASP, 382–396 9, 10

Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95 2

Jiang, B., Luo, A., Zhao, Y., & Wei, P. 2013, MNRAS, 430, 986 7

Jolliffe, I. T. 2002, Principal component analysis (Springer verlag) 7

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) 7

Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008, AJ, 136, 2022 4

Liu, X.-W., Yuan, H.-B., Huo, Z.-Y., et al. 2013, arXiv preprint arXiv:1306.5376 3

Luo, A.-L., Wu, Y., Zhao, J., & Zhao, G. 2008, in Proc. of SPIE Vol, vol. 7019, 701935–1 2

Luo, A.-L., Zhang, H.-T., Zhao, Y.-H., et al. 2012, RAA, 12, 1243 2, 3

Luo, A.-L., Zhang, Y.-X., & Zhao, Y.-H. 2004, in Astronomical Telescopes and Instrumentation, 756–764 (International Society for Optics and Photonics) 2

Luo, A.-L., & Zhao, Y.-H. 2001, Chinese Journal of Astronomy and Astrophysics, 1, 563 2

Luo, A.-L., et al. 2013, in preparation 10

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

McGurk, R. C., Kimball, A. E., & Željko Ivezić 2010, AJ, 139, 1261 16, 17

Si, J., Luo, A., Zhang, J., et al. 2013, arXiv preprint arXiv:1309.1883 17

Singh, H. P., Gulati, R. K., & Gupta, R. 1998, Monthly Notices of the Royal Astronomical Society, 295, 312 8

Željko Ivezić, Sesar, B., Jurić, M., et al. 2008, ApJ, 684, 287 4

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) 2, 3, 15

Wang, F., Zhang, H., Luo, A.-L., et al. 2013, arXiv preprint arXiv:1306.1600 3

Whitney, C. 1983, Astronomy and Astrophysics Supplement Series, 51, 443 8

Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, RAA, 11, 924 5, 6

Yi, Z., Luo, A., Song, Y., et al. 2013, arXiv preprint arXiv:1306.4540 6

Yip, C., Connolly, A., Berk, D. V., et al. 2004, AJ, 128, 2603 8

Zhang, Y.-Y., Deng, L.-C., Liu, C., et al. 2013, AJ, 146, 34 6

Zhao, J., Luo, A., Oswalt, T., & Zhao, G. 2013, AJ, 145, 169 6