

## The first data release (DR1) of the LAMOST regular survey

This content has been downloaded from IOPscience. Please scroll down to see the full text.

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

### Download details:

IP Address: 98.209.175.166

This content was downloaded on 05/10/2016 at 11:47

Please note that [terms and conditions apply](#).

You may also be interested in:

[Spectral classification of stars based on LAMOST spectra](#)

Chao Liu, Wen-Yuan Cui, Bo Zhang et al.

[Preface: The LAMOST Galactic surveys and early results](#)

Xiao-Wei Liu, Gang Zhao and Jin-Liang Hou

[A method of measuring the  \$\[\alpha/\text{Fe}\]\$  ratios from the spectra of the LAMOST survey](#)

Ji Li, Chen Han, Mao-Sheng Xiang et al.

[Validation of LAMOST stellar parameters with the PASTEL catalog](#)

Hua Gao, Hua-Wei Zhang, Mao-Sheng Xiang et al.

[Identifying Carbon stars from the LAMOST pilot survey with the efficient manifold ranking algorithm](#)

Jian-Min Si, Yin-Bi Li, A-Li Luo et al.

[M Dwarf catalog of LAMOST general survey data release one](#)

Yan-Xin Guo, Zhen-Ping Yi, A-Li Luo et al.

[19 low mass hypervelocity star candidates from the first data release of the LAMOST survey](#)

Yin-Bi Li, A-Li Luo, Gang Zhao et al.

## The first data release (DR1) of the LAMOST regular survey

A-Li Luo<sup>1</sup>, Yong-Heng Zhao<sup>1</sup>, Gang Zhao<sup>1</sup>, Li-Cai Deng<sup>1</sup>, Xiao-Wei Liu<sup>5</sup>, Yi-Peng Jing<sup>4</sup>, Gang Wang<sup>1</sup>, Hao-Tong Zhang<sup>1</sup>, Jian-Rong Shi<sup>1</sup>, Xiang-Qun Cui<sup>2</sup>, Yao-Quan Chu<sup>3</sup>, Guo-Ping Li<sup>2</sup>, Zhong-Rui Bai<sup>1</sup>, Yue Wu<sup>1</sup>, Yan Cai<sup>1</sup>, Shu-Yun Cao<sup>1</sup>, Zi-Huang Cao<sup>1</sup>, Jeffrey L. Carlin<sup>7,8</sup>, Hai-Yuan Chen<sup>2</sup>, Jian-Jun Chen<sup>1</sup>, Kun-Xin Chen<sup>2</sup>, Li Chen<sup>4</sup>, Xue-Lei Chen<sup>1</sup>, Xiao-Yan Chen<sup>1</sup>, Ying Chen<sup>1</sup>, Norbert Christlieb<sup>9</sup>, Jia-Ru Chu<sup>3</sup>, Chen-Zhou Cui<sup>1</sup>, Yi-Qiao Dong<sup>1</sup>, Bing Du<sup>1</sup>, Dong-Wei Fan<sup>1</sup>, Lei Feng<sup>1</sup>, Jian-Ning Fu<sup>6</sup>, Peng Gao<sup>1</sup>, Xue-Fei Gong<sup>2</sup>, Bo-Zhong Gu<sup>2</sup>, Yan-Xin Guo<sup>1</sup>, Zhan-Wen Han<sup>10</sup>, Bo-Liang He<sup>1</sup>, Jin-Liang Hou<sup>4</sup>, Yong-Hui Hou<sup>2</sup>, Wen Hou<sup>1</sup>, Hong-Zhuan Hu<sup>3</sup>, Ning-Sheng Hu<sup>2</sup>, Zhong-Wen Hu<sup>2</sup>, Zhi-Ying Huo<sup>1</sup>, Lei Jia<sup>1</sup>, Fang-Hua Jiang<sup>2</sup>, Xiang Jiang<sup>2</sup>, Zhi-Bo Jiang<sup>2</sup>, Ge Jin<sup>3</sup>, Xiao Kong<sup>1</sup>, Xu Kong<sup>3</sup>, Ya-Juan Lei<sup>1</sup>, Ai-Hua Li<sup>2</sup>, Chang-Hua Li<sup>1</sup>, Guang-Wei Li<sup>1</sup>, Hai-Ning Li<sup>1</sup>, Jian Li<sup>1</sup>, Qi Li<sup>1</sup>, Shuang Li<sup>1</sup>, Sha-Sha Li<sup>1</sup>, Xin-Nan Li<sup>2</sup>, Yan Li<sup>4</sup>, Yin-Bi Li<sup>1</sup>, Ye-Ping Li<sup>2</sup>, Yuan Liang<sup>1</sup>, Chien-Cheng Lin<sup>4</sup>, Chao Liu<sup>1</sup>, Gen-Rong Liu<sup>2</sup>, Guan-Qun Liu<sup>2</sup>, Zhi-Gang Liu<sup>3</sup>, Wen-Zhi Lu<sup>2</sup>, Yu Luo<sup>1</sup>, Yin-Dun Mao<sup>4</sup>, Heidi Newberg<sup>7</sup>, Ji-Jun Ni<sup>2</sup>, Zhao-Xiang Qi<sup>4</sup>, Yong-Jun Qi<sup>2</sup>, Shi-Yin Shen<sup>4</sup>, Huo-Ming Shi<sup>1</sup>, Jing Song<sup>1</sup>, Yi-Han Song<sup>1</sup>, Ding-Qiang Su<sup>2</sup>, Hong-Jun Su<sup>1</sup>, Zheng-Hong Tang<sup>4</sup>, Qing-Sheng Tao<sup>2</sup>, Yuan Tian<sup>1</sup>, Dan Wang<sup>1</sup>, Da-Qi Wang<sup>1</sup>, Feng-Fei Wang<sup>1</sup>, Guo-Min Wang<sup>2</sup>, Hai Wang<sup>2</sup>, Hong-Chi Wang<sup>11</sup>, Jian Wang<sup>3</sup>, Jia-Ning Wang<sup>2</sup>, Jian-Ling Wang<sup>1</sup>, Jian-Ping Wang<sup>3</sup>, Jun-Xian Wang<sup>3</sup>, Lei Wang<sup>2</sup>, Meng-Xin Wang<sup>1</sup>, Shou-Guan Wang<sup>1</sup>, Shu-Qing Wang<sup>1</sup>, Xia Wang<sup>1</sup>, Ya-Nan Wang<sup>2</sup>, You Wang<sup>2</sup>, Yue-Fei Wang<sup>2</sup>, You-Fen Wang<sup>1</sup>, Peng Wei<sup>1</sup>, Ming-Zhi Wei<sup>1</sup>, Hong Wu<sup>1</sup>, Ke-Fei Wu<sup>1</sup>, Xue-Bing Wu<sup>5</sup>, Yu-Zhong Wu<sup>1</sup>, Xiao-Zheng Xing<sup>3</sup>, Ling-Zhe Xu<sup>2</sup>, Xin-Qi Xu<sup>2</sup>, Yan Xu<sup>1</sup>, Tai-Sheng Yan<sup>1</sup>, De-Hua Yang<sup>2</sup>, Hai-Feng Yang<sup>1</sup>, Hui-Qin Yang<sup>1</sup>, Ming Yang<sup>1</sup>, Zheng-Qiu Yao<sup>2</sup>, Yong Yu<sup>4</sup>, Hui Yuan<sup>1</sup>, Hai-Bo Yuan<sup>5</sup>, Hai-Long Yuan<sup>1</sup>, Wei-Min Yuan<sup>1</sup>, Chao Zhai<sup>3</sup>, En-Peng Zhang<sup>1</sup>, Hua-Wei Zhang<sup>5</sup>, Jian-Nan Zhang<sup>1</sup>, Li-Pin Zhang<sup>2</sup>, Wei Zhang<sup>1</sup>, Yong Zhang<sup>2</sup>, Yan-Xia Zhang<sup>1</sup>, Zheng-Chao Zhang<sup>2</sup>, Ming Zhao<sup>4</sup>, Fang Zhou<sup>2</sup>, Xu Zhou<sup>1</sup>, Jie Zhu<sup>2</sup>, Yong-Tian Zhu<sup>2</sup>, Si-Cheng Zou<sup>1</sup> and Fang Zuo<sup>1</sup>

<sup>1</sup> National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China; [lal@nao.cas.cn](mailto:lal@nao.cas.cn)

<sup>2</sup> Nanjing Institute of Astronomical Optics & Technology, National Astronomical Observatories, Chinese Academy of Sciences, Nanjing 210042, China

<sup>3</sup> University of Science and Technology of China, Hefei 230026, China

<sup>4</sup> Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China

<sup>5</sup> Department of Astronomy, Peking University, Beijing 100871, China

<sup>6</sup> Department of Astronomy, Beijing Normal University, Beijing 100875, China

- <sup>7</sup> Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
- <sup>8</sup> Department of Physics and Astronomy, Earlham College, Richmond, IN 47374, USA
- <sup>9</sup> Universität Heidelberg, Zentrum für Astronomie, Landessternwarte, Königstuhl 12, 69117 Heidelberg, Germany
- <sup>10</sup> Yunnan Observatories, Chinese Academy of Sciences, Kunming 650216, China
- <sup>11</sup> Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210008, China

Received 2015 January 27; accepted 2015 June 1

**Abstract** The Large sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) general survey is a spectroscopic survey that will eventually cover approximately half of the celestial sphere and collect 10 million spectra of stars, galaxies and QSOs. Objects in both the pilot survey and the first year regular survey are included in the LAMOST DR1. The pilot survey started in October 2011 and ended in June 2012, and the data have been released to the public as the LAMOST Pilot Data Release in August 2012. The regular survey started in September 2012, and completed its first year of operation in June 2013. The LAMOST DR1 includes a total of 1202 plates containing 2 955 336 spectra, of which 1 790 879 spectra have observed signal-to-noise ratio (SNR)  $\geq 10$ . All data with SNR  $\geq 2$  are formally released as LAMOST DR1 under the LAMOST data policy. This data release contains a total of 2 204 696 spectra, of which 1 944 329 are stellar spectra, 12 082 are galaxy spectra and 5017 are quasars. The DR1 not only includes spectra, but also three stellar catalogs with measured parameters: late A,FGK-type stars with high quality spectra (1 061 918 entries), A-type stars (100 073 entries), and M-type stars (121 522 entries). This paper introduces the survey design, the observational and instrumental limitations, data reduction and analysis, and some caveats. A description of the FITS structure of spectral files and parameter catalogs is also provided.

**Key words:** techniques: spectroscopic survey — data release — catalog

## 1 INTRODUCTION

The Large sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST), with an effective aperture of 4 m and size of 6.67 m for the main mirror, is a telescope located at the Xinglong Observatory northeast of Beijing, China. The telescope is characterized by both a large field of view and large aperture. A total of 4000 fibers mounted on its focal plane make the LAMOST facility a powerful tool with a high spectral acquisition rate (Deng et al. 2012). With the capability of its large field of view and strong multiplexing ability, the LAMOST telescope is dedicated to a spectral survey of celestial objects over the entire available northern sky. After first light in 2008, commissioning observations were conducted for two years. In spite of the ongoing optimization in performance of the LAMOST system, with related limitations in observations (e.g., dome seeing and fiber positioning accuracy), and data processing software, the commissioning data have contributed some preliminary results on quasars (Huo et al. 2010; Wu et al. 2010a,b), planetary nebulae (Yuan et al. 2010) and nearby galaxies (Zou et al. 2011).

In order to inspect the real instrumental performance and assess the feasibility of the science goals, the science working groups for LAMOST devised the pilot survey, which was initiated on 24 October 2011. In the pilot survey, both Galactic and extragalactic surveys were conducted – these were named LAMOST Experiment for Galactic Understanding and Exploration (LEGUE) and

LAMOST ExtraGalactic Survey (LEGAS), respectively (Zhao et al. 2012; Liu et al. 2015). LEGUE is made up of three parts, which focus on the Galactic Anti-center, the disk at longitudes away from the Galactic Anti-center, and the halo. Some scientific results from the pilot survey include catalogs of white dwarf-main sequence binaries (Ren et al. 2013), DA white dwarfs (Zhang et al. 2013; Zhao et al. 2013), M dwarfs (Yi et al. 2014), and identification of velocity substructure in the Galactic disk (Carlin et al. 2013). Moreover, the LEGAS extragalactic portion of the survey studies galaxy redshifts, stellar populations of galaxies, and QSOs. Scientific results obtained include the identification of QSOs in the background of M31/M33 (Huo et al. 2013), a study of the radio-loud quasar LAMOST J1131+3114 (Shi et al. 2014a), and a search for double-peaked, narrow emission-line galaxies (Shi et al. 2014b). Additionally, a catalog of newly identified QSOs from DR1 was compiled (Ai et al. 2015, in preparation). The LAMOST pilot survey ended in June 2012, and included nine full moon cycles covering all good observing seasons at the site. The pilot survey data release includes 717 660 spectra, in which there are 648 820 stars, 2 723 galaxies and 621 quasars, and a stellar parameter catalog containing entries for 373 481 stars (Luo et al. 2012).

From the experience gained by the pilot survey, the LAMOST science committee fully analyzed the observations, assessing the magnitude limit, characteristics of the telescope, and data success rate. A five-year general survey of LAMOST was designed, which would mainly focus on the Galactic survey of stars, yet still include a significant number of extragalactic sources.

As the only resolved classical disk galaxy, the Milky Way serves as a laboratory to study the formation and evolution of galaxies. However, there have been no large-scale Galactic spectroscopic surveys with a uniform sampling over all stellar populations, due to the time consuming nature of observing a large fraction of its  $\sim 20$  billion members spread over the entire sky. Thus our knowledge on the detailed structure of our Galaxy lags behind the understanding of the large-scale structure of the universe. Several notable achievements have been made by modern large-scale surveys such as the Sloan Digital Sky Survey (SDSS); however, because SDSS is mainly focused on extragalactic targets, its spectroscopic observations have sampled only a small fraction of the Milky Way's stellar components. The powerful spectral acquisition rate of LAMOST enables a homogeneous and (statistically) complete Galactic survey.

The design of the survey will be shown in Section 2, including fields and sample selection. The observation conditions and limitations due the instruments will be discussed in Section 3. Data reduction, spectral analysis and stellar parameter pipelines used for LAMOST will be introduced in Section 4. The data products of Data Release 1 (DR1) will be introduced in Section 5, as well as caveats. Finally, the data access and use policy are described in Section 6.

## 2 DESIGN OF THE SURVEY

### 2.1 LAMOST Experiment for Galactic Understanding and Exploration (LEGUE)

As a major component of the LAMOST project, the LEGUE spectroscopic survey of the Galactic halo will observe  $\sim 5.8$  million objects with  $r < 16.8$  mag at  $|b| > 30^\circ$ , mostly located in the sky area that also has SDSS photometric coverage (Yang et al. 2012; Deng et al. 2012; Zhang et al. 2012). Considering the time required for faint objects which should be observed in dark nights, two stripes of medium-bright (M) plates with high priority have been proposed. One is the area at Dec. =  $30^\circ$  in the Northern Galactic Cap (NGC), and the other is at Dec. =  $5^\circ$  in the Southern Galactic Cap (SGC).

Another unique component of LEGUE is the LAMOST Spectroscopic Survey of the Galactic Anti-center (LSS-GAC), which aims to survey a significant volume over a continuous sky area of  $\sim 3400$  deg<sup>2</sup>, covering Galactic longitudes  $150^\circ \leq \ell \leq 210^\circ$  and latitudes  $|b| \leq 30^\circ$  (Yuan et al. 2014b). The input catalog is selected from photometric catalogs of the Xuyi Schmidt Telescope Photometric Survey of the Galactic Anti-center (XSTPS-GAC) (Liu et al. 2014), with most targets selected with uniform probability. The catalog will include a statistically complete sample of about

3 million stars with magnitudes in the range  $14.0 < r < 17.8$  (and as deep as 18.5 mag for some fields), with rare objects exhibiting extreme colors targeted with high priority. A detailed description of LSS–GAC is presented by Yuan et al. (2014b).

For the Galactic disk survey, eight plates are chosen near the Galactic plane ( $|b| \leq 20^\circ$ ) and are nearly uniformly distributed along Galactic longitude, as well as in the region  $0^\circ < \alpha < 67^\circ$  and  $42^\circ < \delta < 59^\circ$  (Chen et al. 2012). This survey concentrates on open clusters (OCs), and is expected to obtain about 8.9 million stars down to a limiting magnitude of  $r \sim 16$  mag. The highest observational priority is given to known or potential members of OCs, and a total of 194 OCs have high priority. The remainder of the disk targets in the  $11.3 < I_{\text{mag}} < 16.3$  range have been uniformly selected over  $I$  and  $B-I$  color-magnitude space. Objects with very high proper motion ( $7 \text{ mas yr}^{-1}$ ) are removed from the target catalogs. Positions, proper motions and magnitudes are taken from the PPMXL (Roeser et al. 2010) catalog.

## 2.2 LAMOST ExtraGalactic Survey (LEGAS)

The extragalactic survey LEGAS consists of two main parts – a galaxy survey and a QSO survey. For galaxies, it is split into two regions (again, to avoid the Galactic plane). One is in the NGC region, and shares the same footprint as the SDSS Legacy Survey. This survey mainly aims to observe those SDSS main sample galaxies (with  $r < 17.75$ ) that were not spectroscopically observed by SDSS due to fiber collisions (Stoughton et al. 2002). Such “missed” SDSS galaxies are named the complementary galaxy sample in the LAMOST survey. There are 68 722 such targets, which are given highest priority for fiber assignment on plates when possible. Because the number density of galaxies is very low, the remaining fibers on LEGAS plates are filled by LEGUE halo survey stars. The other region of the galaxy survey in LAMOST is the SGC. The LEGAS team has a long-term strategic goal of performing a galaxy survey in the SGC footprint of LAMOST ( $b < -30^\circ$  and  $\delta > -10^\circ$ ), which aims to take spectra of all galaxies with  $r < 18$ , and a sample of blue galaxies down to  $r < 18.8$ . Since this desired galaxy survey in the SGC requires much deeper observations than the main LAMOST stellar sample (currently, mostly limited to  $r < 16.8$ ), only a few dark nights per observing season are available for these galaxies. Therefore, the galaxy survey in SGC now focuses on a strip, in the range  $45^\circ < \alpha < 60^\circ$  and  $0.5^\circ < \delta < 9.5^\circ$ . Besides these two well-defined samples of galaxies, some bright-infrared galaxies are chosen as extra observational targets which are selected from infrared surveys such as IRAS, WISE and *Herschel*.

The efficiency of identifying quasars with redshifts between  $2.2 < z < 3$  is low in SDSS (Schneider et al. 2010), because quasars with such redshifts usually have optical colors that are similar to stars and are thus mostly ignored by the SDSS quasar candidate selection algorithm. As permitted by the observational conditions and instrument standards of the LAMOST telescope, the LAMOST quasar survey aims to find as many quasars as possible. We selected quasar targets from optical data within the SDSS photometric footprint using data mining methods such as support vector machines (SVMs) (Peng et al. 2012; Zhang & Zhao 2003), distribution of quasars and stars in different color spaces including  $Y-K$  vs.  $g-z$  (Wu et al. 2012) and  $z-W1$  vs.  $g-z$  (Wu & Jia 2010), and the extreme deconvolution method (Bovy et al. 2011). The targets selected by multiple methods are given much higher priority in observation. The core sample is contributed from the above methods using survey data from SDSS, UKIDSS and WISE. The bonus sample comes from other survey data, such as X-ray bands from ROSAT, XMM and *Chandra*, and radio bands from FIRST and NVSS. In this case, QSO targets are not limited to the LEGAS plates, and are given high priority in the LEGUE plates as well.

## 2.3 The Special Fields

### 2.3.1 M31-M33

Targets in the M31-M33 region include known and candidate planetary nebulae (PNe), HII regions, supergiants and globular clusters in M31 and M33, plus known and candidate background quasars and foreground Galactic stars selected from the XSTPS-GAC photometric catalogs. These stars are selected, assigned and observed in the same way as for the LSS-GAC main survey. The selection criteria of various types of targets and candidates of interest are presented elsewhere (e.g., Yuan et al. 2010 for PNe; Huo et al. 2010 for quasars). The central stars for the M31-M33 survey are selected similarly to the LSS-GAC main survey in the (RA, Dec) plane. Two groups of central stars are independently chosen to cover the area around M31-M33 ( $0^\circ \leq \text{RA} \leq 30^\circ$ ,  $25^\circ \leq \text{Dec} \leq 50^\circ$ ). Each group contains 41 central stars.

### 2.3.2 The LAMOST Complete Spectroscopic Survey of Pointing Area at the SGC

A LAMOST Key Project named the LAMOST Complete Spectroscopic Survey of Pointing Area (LCSSPA) at SGC is being conducted (Yang et al. 2015), which is designed to have repeated spectroscopic observations of all sources (Galactic and extragalactic) in two 20 deg<sup>2</sup> fields in the SGC. The central coordinates of the fields are  $(\alpha, \delta) = (37.88150939^\circ, 3.43934500^\circ)$  and  $(21.525988792^\circ, -2.200949833^\circ)$ , respectively. Targets in this field mainly consist of stars, galaxies, *u*-band variables and QSOs. The stars and galaxies are selected from SDSS Data Release Nine (DR9) using *r*-band *psf* magnitudes and *Petrosian* magnitudes between  $14.0 < r < 18.1$  mag, respectively. The *u*-band variable sources are selected as objects for which the difference in magnitudes between the SDSS and the SGC *u*-band Sky Survey (SCUSS) is larger than 0.2 mag; these targets are a magnitude-limited sample down to *u* = 19.0 mag. The QSOs are selected by the same methods discussed above. The highest priority is assigned to *u*-band variables and QSOs, with galaxies having the second priority level, and stars having the lowest priority.

### 2.3.3 Kepler fields

Additionally, a total of 14 *Kepler* fields with a fixed field of view (FoV) of 105 square degrees in the constellations Lyra and Cygnus are also selected according to the prioritized list provided by the LK project (De Cat et al. 2015, in preparation). The prioritized lists are mainly based on the type of targets and brightness. According to the type of object, these targets are divided into four priority levels: standard targets (~250, MK secondary standard stars), KASC targets (~7000, selected by the *Kepler* Asteroseismic Science Consortium, or KASC), planet targets (~150 000, selected by the *Kepler* planet search group, Batalha et al. 2010), and extra targets (~1 000 000 other targets in the *Kepler* FoV from the KIC 10). For these, a higher observation priority is given to fainter objects.

## 2.4 Plate Design

### 2.4.1 Plate definitions and their magnitudes

The LAMOST plates are divided into four modes, very bright (VB), bright (B), medium-bright (M) and faint (F) according to the magnitude range of stars to be observed. In DR1, the VB mode contains 2693 plates, and is used to observe stars brighter than  $r = 14$  mag over the entire sky observable by LAMOST ( $-10^\circ \leq \text{Dec} \leq 60^\circ$ ) in order to make use of bright nights or nights with unfavorable conditions (e.g., poor seeing or low atmospheric transparency). In the XSTPS-GAC footprint, stars are selected as potential targets from the XSTPS-GAC and 2MASS catalogs with magnitudes  $r \leq 14.0$  mag and  $9.0 \leq J \leq 12.5$ , respectively. Stars outside of the XSTPS-GAC sky

area are selected as potential targets from the PPMXL (Roeser et al. 2010) catalog with magnitudes  $10.0 \leq b_1 \leq 15.0$ ,  $10.0 \leq b_2 \leq 15.0$ ,  $9.0 \leq r_1 \leq 14.0$ ,  $9.0 \leq r_2 \leq 14.0$ , or  $8.5 \leq i \leq 13.5$ , and from the 2MASS catalog with magnitude  $9.0 \leq J \leq 12.5$ . The “B” (bright) observing mode contains 751 plates in DR1 that observes stars with  $r$  magnitudes between between 14 and 16.8. There are 692 and 655 plates assigned to the M ( $16.8 \leq r \leq 17.8$ ) and F ( $r > 17.8$ ) designations, respectively.

#### 2.4.2 Tiling method

Merging all of the science goals described above, the science committee of LAMOST designed the survey strategies and input catalogs. Based on these, the LAMOST team developed a software package named Sky Strategy System (SSS) to produce observational plates that most efficiently use the multi-fiber abilities of LAMOST (Cui et al. 2012). The maximum density algorithm, density gradient algorithm and the mean-shift algorithm are used for most sky areas, with the exception of LSS-GAC, which uses separate target selection criteria (Yuan et al. 2014b).

Given a sky area, the maximum density method can find the tile (or plate) with the maximum object count by making a uniform tiling. By traversing all the tiles covering a sky area, a tile with the maximum object count will be found and then a new observational plate will be configured at that position. The only factor that affects the speed and accuracy of this algorithm is the target distribution in the sky area. The denser the covering is, the higher the accuracy of finding the maximum tile becomes and the more efficient the method is, and vice-versa. The density gradient method is essentially iterative. Based on a uniform covering, this method makes an estimate of the distribution of objects and moves the tile in the direction that is expected to have a higher density of objects. The basic idea of the mean-shift algorithm is also iterative. As long as the objects are distributed unevenly in areas of the sky, a density gradient exists between different positions. Vectors are used to describe the position of targets in the tile, and the sum of vectors is set along the direction of the density gradient. By iteratively moving the field center along the sum of vectors and calculating the new vector summation, a new peak density location is obtained. In order to maintain the goal of uniformity in the remaining area and to avoid neglecting some of the peak positions in areas where density varies in complicated ways, the algorithm will try to find a position with a higher global density and produce a coverage of the whole available sky area before the next iteration (Ren et al. 2007).

For each plate, one bright star (the so called central star, brighter than 8 mag) must be placed in the center of the plate to feed a Shack-Hartmann detector that drives the active optics system, and four stars (brighter than 17 mag) must be placed in the fields of each of the four guiding CCD cameras. The bright stars located in the center were selected from the Hipparcos catalog (Perryman et al. 1997), with binaries excluded, and at the time of observation, the focal plane is rotated a bit to confirm that the proper guiding stars are inside the FoV of the four guiding CCD cameras. Guide stars were selected from the Guide Star Catalog (GSC), Version 2.3 (Lasker et al. 2008). Central stars will be excluded if there are no corresponding guide stars. Since in some areas (e.g., at low Galactic latitudes) there are many more available central stars than needed to fully cover that part of the sky, a step by step search is applied to reject redundant bright stars and find the tiling that maximizes sky coverage in that area (Yuan et al. 2012).

#### 2.4.3 Fiber allocation scheme

Once the central position of a plate is decided by the tiling method, each object in the input catalog will be assigned a priority according to its scientific requirements. Five flux standards and 20 sky fibers are reserved in each spectrograph for data reduction; therefore, these five flux standards are allocated with the highest priority. All of the scientific targets are assigned in order of priority;

however, if more than two targets have the same priority in the same fiber cell, the closest target will be chosen to avoid collision with other fibers. On the other hand, if a target can be reached by different fibers at the same time, then it will be assigned to the fiber within its fiber cell that contains the minimum number of possible targets (Yuan et al. 2014a). A total of 20 sky fibers will eventually be assigned to fibers for which no targets were originally assigned. If the number of sky fibers allocated to a given spectrograph is less than 20, the lowest priority fibers will be reassigned to the sky until the 20 sky fibers are allocated.

### 3 LIMITATIONS ON THE OBSERVATIONS AND PERFORMANCE OF THE TELESCOPE

#### 3.1 Seeing

One of the significant factors affecting the observations is seeing, which consists of the native site seeing as well as dome seeing. The seeing depends on the season, and is worse in a windy winter, but exhibits huge variations. The peak in the distribution of the nightly site seeing is at around 3 arcsec, and 85% of the seeing measurements are better than 4 arcsec (Yao et al. 2012). The non-classical dome of the LAMOST telescope seemingly affects the seeing measured directly at the focal plane, with the measurements showing seeing of about 3.5 arcsec on average. Compared to the site seeing, it is dome seeing that contributes the most in limiting the capabilities of the system (especially the limiting magnitude). Considering a fiber diameter of 3 arcsec, an average increase of 1 arcsec in seeing (at the focal plane) will lead to a 0.4-magnitude decrease in flux collected by each fiber.

#### 3.2 Fiber Positioning and Efficiency of Usage

The precision of fiber positioning is an important factor affecting the efficiency of the survey. Through repeated observations and calculations, the accuracy of fiber positioning has been determined to be less than 1.5 arcsec on average (Chen et al. 2014), which has less effect on the fiber magnitude than the seeing. Another limitation is provided by the fixed fiber density of around 200 per  $\text{deg}^2$ . This limits sampling in both dense and sparse fields in different sky areas. For dense fields, objects are too crowded to resolve (especially near objects such as stellar clusters). By contrast, the fiber density results in a considerable waste of fibers (which are allocated to the sky) for sparse fields in areas with high Galactic latitude.

#### 3.3 The Performance of the Instruments

The performance of the instruments plays the same important role in the quantity and quality of spectra observed as the procedures used in data reduction. We briefly describe it as follows.

*Overhead* During the observations, the overhead time will reduce the observational efficiency of the survey to some extent. Specifically, the telescope needs to be confocal, guide stars need to be located, and the fibers need to be repositioned when a new plate will be observed. Including the CCD readout time, the total overhead prior to observation of a field takes about 20 minutes.

*Throughput* The throughput is a major indicator we use to measure the efficiency of the telescope. The whole throughput takes the contributions of the spectrograph, fibers, CCDs, optical system, etc., into account. In general, the throughputs of the red and blue channels are 4% and 2% respectively (Cui et al. 2012). During the pilot survey and the first year regular survey, the throughput of LAMOST was relatively low, and also showed a large dispersion among the 4000 fibers. Spectrograph No. 4 (out of a total of 16 spectrographs) has the highest throughput of those located near the center of the focal plane; median throughput of the valid fibers connected to spectrograph No. 4 during the pilot survey was about 1.0% and 2.0% for the blue and red bands respectively, and improved to 1.2% and 3.0% for the blue and red bands during the first year regular survey.



Spectrograph Nos. 3, 5, 8, 9 and 15 are located in the middle-ring of the focal plane; their throughputs are about 80%–90% of that of spectrograph No. 4. Spectrograph Nos. 1, 2, 6, 7, 10, 11, 12, 13, 14 and 16 are located in the outer-ring of the focal plane, and have measured throughputs that are about 50%–80% of that of spectrograph No. 4. Among the 250 fibers accommodated by each spectrograph, the difference in the throughput can change by a factor of two.

*Stability of instrument profile* Variations in emission lines from arc lamps have been monitored during a continuous 300 day period in order to measure the stability of the instrument profile. From Figure 1, it can be seen that the profile of each fiber changes little with time, but that there are differences among the profiles between the fibers.

*Flux calibration.* The LAMOST project is designed as a spectroscopic survey without a component that can acquire corresponding photometry data. Therefore, absolute flux calibration cannot be done without the photometric data (Song et al. 2012). Moreover, not enough first-level standard stars are available over the large FoV observed by LAMOST that can be used for relative flux calibration. Thus we choose some stars with high quality spectra as standard stars. The uncertainty in the shape of the continuum in these standard stars will thus contribute to uncertainty in the LAMOST flux calibration.

### 3.4 Sky Subtraction and Flat Fielding

According to the requirements set for LAMOST observations,  $\sim 10 - 15\%$  of the fibers for each plate are assigned for modeling the sky. This fiber assignment strategy works well for fields at high Galactic latitude, providing sky-subtraction that is consistent with results from the SDSS survey. However, the proportion of fibers that sample “clean” sky background is not adequate to build a supersky in observations near the Galactic disk, especially during bright nights. The emission lines of the night sky spectra are not subtracted satisfactorily for a number of spectra under these conditions. Differences in the fiber-to-fiber profile make the subtraction of sky emission lines more complicated. Moreover, the sky continua are more difficult to remove than narrow sky emission lines, such that there will always be at least some sky residual remaining in reduced spectra. This in turn presents a challenge we have to deal with when determining (stellar) parameters from spectra.

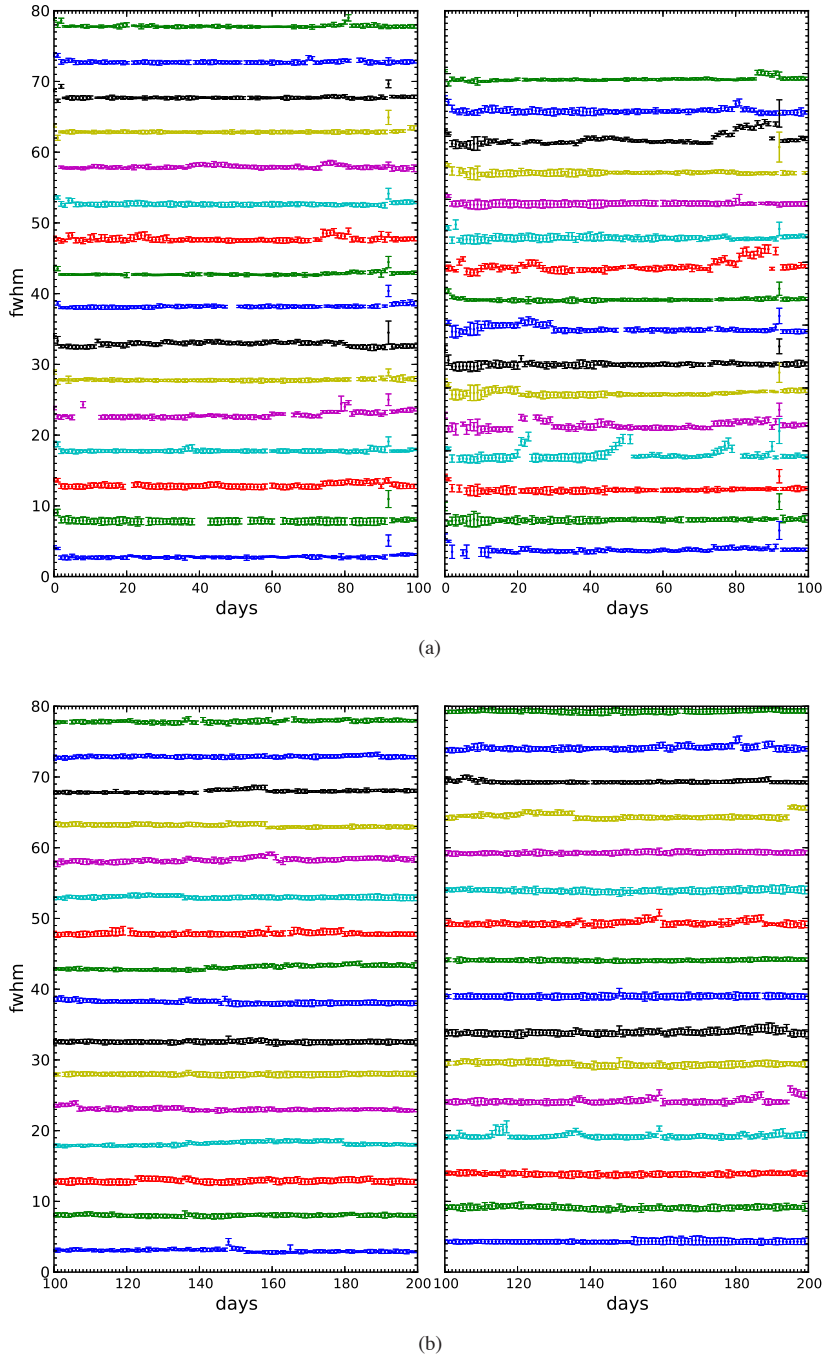
It is not easy to find a light source with a five-degree FoV to use as a flat field. We therefore separately correct the flat field for each spectrograph, since each spectrograph covers only about a  $1^\circ$  FoV. Twilight flats and screen flats are used. Additionally, the vignetting and the change in aperture caused by the telescope pointing also complicate the correction of the flat field. The movement of fibers may lead to focal ratio degradation caused by changes in fiber tension, which will cause differences in the efficiency of fibers.

## 4 DATA REDUCTION

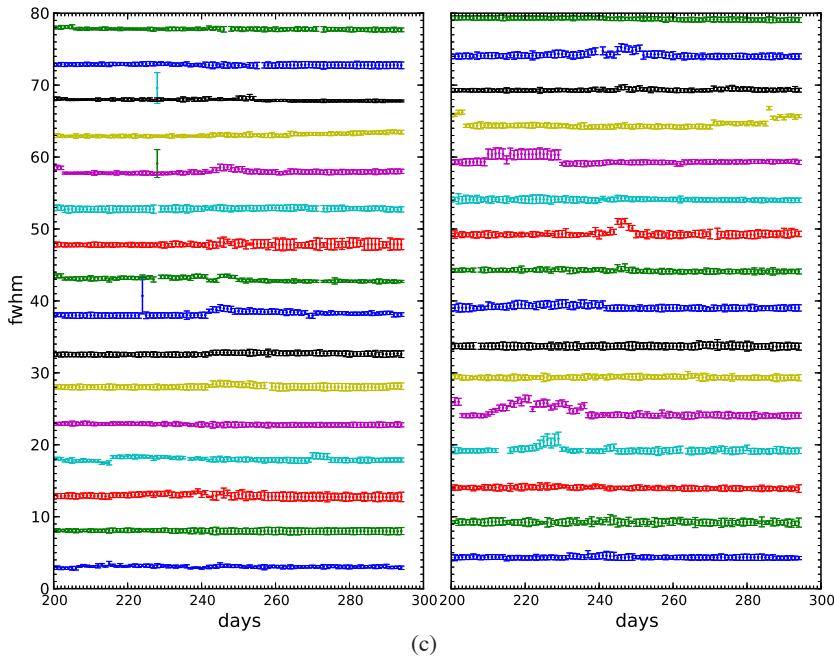
### 4.1 Statistics of Data Quality

#### 4.1.1 Resolution and re-sampling

The theoretical resolution of the 16 LAMOST spectrographs is  $R = 1000$  (as determined by the gratings), and the designed wavelength coverage is 365–900 nm. In practice, the resolution reaches  $R > 1500$  by narrowing the slit at the output end of the fibers to  $2/3$  the fiber width. Since the spectral images are sampled on CCDs, to avoid confusion between resolution and sampling for the final re-binned spectrum, we explain the difference as follows. On the original CCD image, 4K pixels are used for recording blue or red wavelength regions ranging from 370 nm to 590 nm or from 570 nm to 900 nm respectively, which means each Å has been sampled in two CCD pixels in the raw data. When the blue and red channels are combined together, each spectrum is re-binned for easier calculation of radial velocity. The combined spectra are re-sampled in constant-velocity pixels, with



**Fig. 1** Variations in arc lines during 300 continuous days. The left panels show the full width at half maximum (FWHM) of arc lines in the blue channel for the first 100 days, the second 100 days and the last 100 days, from top to bottom respectively, and the right panels are the same observations in the red channel.



**Fig. 1** — *Continued.*

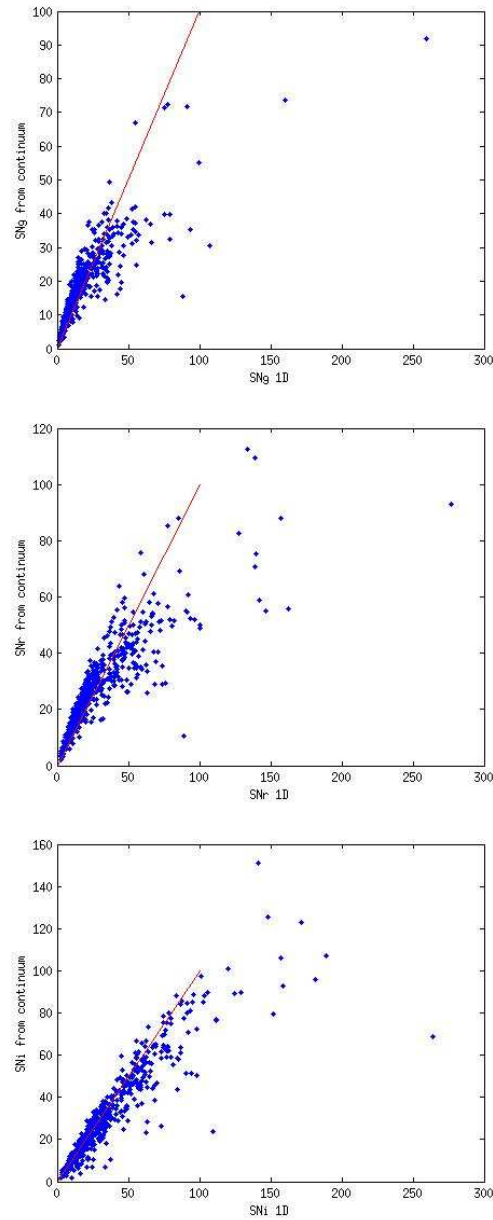
a pixel scale of  $69 \text{ km s}^{-1}$ , which means the difference in wavelength between two adjacent points is  $\Delta \log(\lambda) = 0.0001$ . For convenience, the re-binned data interval is still referred to as a ‘pixel.’

#### 4.1.2 The definition of signal to noise ratio

The signal to noise ratio (SNR) is always used as a quality indicator, and is the ratio of flux to noise in some bandpass. Generally, SNR is an average value in a wavelength band and predominantly represents the continuum flux level. SNR hereafter indicates the SNR *per ‘pixel’*. In the LAMOST data products, SNR is defined by using the concept of inverse-variance, which is included in the spectral FITS files. It is easy to obtain the SNR for each pixel by calculating the product between the flux and its inverse-variance. The mean SNR in a wavelength band can be averaged from the SNR of each pixel in the band range. Another way to obtain SNR is to calculate the ratio between the continuum and the  $1\sigma$  standard deviation. These two definitions of SNR are shown to provide consistent results when applied to LAMOST data (Fig. 2).

#### 4.1.3 Data quality

The data quality is determined by SNR, and a successful observational target should have  $\text{SNR} > 10$  in  $g$ - or  $i$ -band. Table 1 lists the statistics on the rate of successful spectral acquisition. This success rate improved from the pilot survey to the regular survey, and the rate on VB plates reached 70% even though these targets were observed in bright nights.



**Fig. 2** The two definitions of SNR are compared for a sample of LAMOST spectra. From top to bottom, the panels show SNR in the  $g$ ,  $r$  and  $i$  bands respectively.

#### 4.2 2D Pipeline

Raw CCD data were reduced by the LAMOST data reduction software which is known as the LAMOST 2D pipeline. The procedures used by the 2D pipeline, similar to those of SDSS (Stoughton

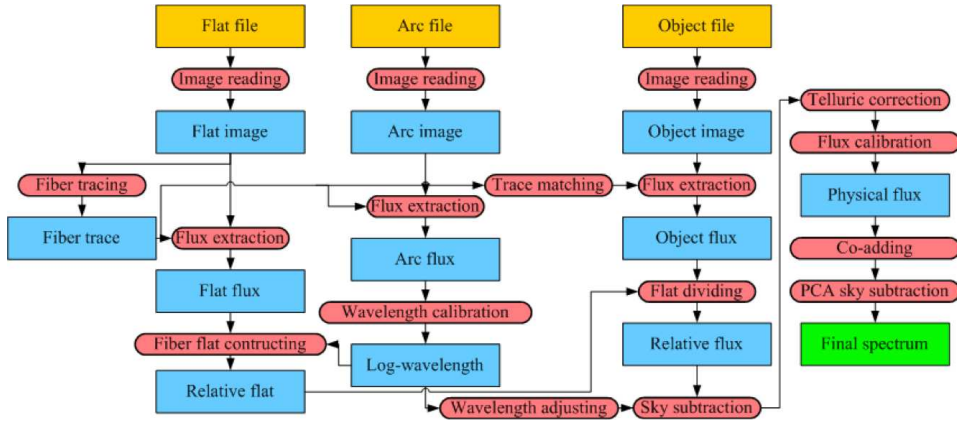
**Table 1** The Statistics on Rate of Successful Spectral Acquisition

Plate Type	Number of Plates	Number of Spectrographs	Rate of Successful Spectral Acquisition
GAC	190	2592	318854/626554 = 51%
M31	41	511	54950/138974 = 40%
VB	343	4405	552245/964080 = 57%
HD	347	5124	479076/776830 = 62%
EG	71	1057	102675/210303 = 49%
F M B	202	2805	268154/628069 = 43%
<i>Kepler</i>	8	104	15333/30499 = 50%
Total	1202	16598	1791287/3375309 = 53%

Notes: (1) The last column is the rate of successful spectral acquisition, which means the ratio of the number of spectra with  $\text{SNR}_g$  or  $\text{SNR}_r \geq 10$  to the number of fibers that were assigned to targets. (2) Plate types are named from the pilot survey and the first year of the regular survey. The details are described as follows: *GAC*: Galactic Anti-center plates. *VB*: Very bright stars with  $r \leq 14$  in the northern sky ( $-10^\circ \leq \text{Dec} \leq 60^\circ$ ). *HD*: Halo and disk plates (Sep. 2012–Jun. 2013). *EG*: Extragalactic plates. *F M B*: Halo and disk plates (Oct. 2011–Jun. 2012). F, M and B mean faint, medium and bright, respectively.

et al. 2002), aim to extract spectra from CCD images and calibrate them. The main tasks of the 2D pipeline include dark and bias subtraction, flat field correction, spectral extraction, sky subtraction, wavelength calibration, merging sub-exposures and combining wavelength bands. The processes incorporated in the 2D pipeline are listed in the following (see the flowchart in Figure 3).

- (1) Bias and dark counts are subtracted from each raw image. During the pilot survey, some interior light source affected the CCD that was acquiring images, so that we needed dark frames to eliminate this stray light. In the regular survey, this light source was identified and removed, so that no dark subtraction is needed.
- (2) Flat field corrections are mainly achieved using twilight flat fields. However, white-screen flat fields were used on some days. LAMOST spectrographs are regarded as 16 different individual instrument systems for flat correction, so that the flat for each spectrograph is independent.
- (3) The flat-field spectra are traced for each fiber, and the centroid of the row position of each fiber is fitted by a polynomial function. A good fitting of each spectral track overcomes interference between fibers. However, the fitting will fail at the end of the wavelength coverage range when the SNR becomes low.
- (4) A spectrum is extracted along the track fitted by step 3 using a weighted aperture integration method; flux on 15 pixels is integrated to provide a single flux for 1D spectra.
- (5) The sky background is modeled and subtracted by a fixed proportion of the fibers on each plate that is reserved for pointing toward blank sky. The method of B-spline fitting is applied to make a supersky for each sub-exposure and then principal component analysis (PCA) is used to re-correct the supersky after combining in this procedure. Additionally, the relative strength of sky emission lines is used to correct differences in efficiency between sky fibers and object fibers, since these lines exist in both sky spectra and object spectra.
- (6) Wavelength calibration is accomplished by using arc lines. The arc lamp spectra are extracted, and centroids of the lines are measured, to which we fit a Legendre polynomial as a function to describe the relationship between wavelengths and pixels, with a fifth-order polynomial in the blue branch and sixth-order in the red. A vacuum wavelength scale is applied for wavelength calibration, slightly adjusted to match the known positions of certain sky lines, and then corrected to the heliocentric frame.
- (7) The telluric absorption in four wavelength regions (6850–6960 Å, 7150–7350 Å, 7560–7720 Å, 8100–8240 Å) is removed.



**Fig. 3** Flowchart illustrating the LAMOST 2D pipeline processing.

- (8) The spectra are flux-calibrated by matching the selected flux standard stars with their matching templates of F-type stars. First, 250 spectra from one spectrograph are sorted in descending order of SNR. Stars with bad SNR or saturated fibers are abandoned. After normalizing the continuum of each spectrum, atmospheric parameters are obtained by matching a template using the Kurucz model. Then stars with  $6000 \text{ K} \leq T_{\text{eff}} \leq 7000 \text{ K}$  are selected as standard stars, after visual inspection to confirm their correct identification. If the number of standard stars is less than three, then the effective temperature will be extended from 5500–7500 K to obtain more standard stars.
- (9) For each object in all the exposures, the red and blue spectra are combined by stacking the points with corresponding wavelengths using a B-spline function, with inverse-variance weighting. Outliers due to cosmic rays are removed and masked, then errors in the fluxes are estimated.

#### 4.2.1 Uncertainties in the wavelength calibration

During the wavelength calibration, errors are most likely to be introduced when finding centroids of the emission lines of the arc lamp. In these emission lines, not all of the lines are single – some observed lines are actually blends of two or three lines. Therefore, the fitting quality of a unimodal function for these mixtures may have large uncertainties. Statistically, the average error of wavelength calibration for the arc lamp is less than  $0.02 \text{ \AA}$ , corresponding to about  $1 \text{ km s}^{-1}$ . Generally, the calibration accuracy of blue spectra is better than that of red spectra, but it degrades below  $4000 \text{ \AA}$ .

#### 4.2.2 Uncertainties in the relative flux calibration

If there are no standard stars observed on one of the spectrographs (for a given plate), combined object spectra spanning the entire wavelength range cannot be produced. This is because matching the blue- and red-channel spectra relies on matching their flux in the wavelength region where they overlap.

The reddening calibration is not considered in selection of standard stars. The extinction is negligible in areas with high Galactic latitude. However, reddening will severely affect flux calibration

in areas with low Galactic latitude, biasing the continuum shapes of the resulting spectra in these high-extinction regions.

The flux of calibrated spectra from different exposures for each object may be not on the same scale. Hence before combining them, we adjust them onto a uniform scale, and then combine the spectra from different exposures by the method of B-spline fitting, which is similar to the SDSS pipeline.

### 4.3 1D Pipeline

The LAMOST 1D pipeline works on spectral type classification and measurement of radial velocity (RV) for stars or redshift for galaxies and QSOs. By using a cross-correlation method, the pipeline recognizes the spectral classes and simultaneously determines the redshifts or radial velocities from the best fit correlation function. In this procedure, a low order polynomial is added while matching the observed spectra with templates, since the physical continuum can be difficult to detect because of the low resolution, reddening or uncertainties in the flux calibration. Currently we employ the same templates as SDSS. However, the templates will be updated with LAMOST spectra for reductions after DR1. Finally, the pipeline produces four primary classifications, namely STAR, GALAXY, QSO and UNKNOWN.

Measuring redshifts and recognizing galaxy and QSO spectra are difficult tasks since the pipeline does not work as well as for stellar classification. An additional independent pipeline was designed for galaxy recognition and redshift measurement after the initial 1D pipeline was run. It automatically identifies galaxies from the objects with confidence of classification below 80 or a target type as “galaxy” in the input catalog. The redshifts of galaxies are measured through line centers; a Gaussian function with sigma of 1.5 times the wavelength step is applied to the spectra to eliminate noise before line centers are measured. The continua, smoothed by a median filter, are divided out to complete the normalization. Those data points that exceed  $2\sigma$  of a normalized spectrum are selected as candidate emission lines, and a set of Gaussian functions is used to fit the lines. All the line centers are matched with line lists spaced by steps of 0.0005 in redshift ( $z$ ). If most of the lines are matched successfully with heavily weighted lines such as  $H_\alpha$ , OII,  $H_\beta$ , OIII or NII for emission line galaxies, or NaD, Mgb, CaII H or CaII K for absorption galaxies, the spectrum is classified as a galaxy, and the corresponding  $z$  is the raw redshift of the spectrum. To detect mistakes in the redshift measurement, this redshift is compared with the results of the 1D pipeline, and the final redshifts are decided by experts. For QSOs, the recognitions and measurements highly depend on visual inspection; a QSO catalog of DR1 detections will be released by Ai et al. (2015, in preparation).

### 4.4 LAMOST Stellar Parameter Pipeline - LASP

LAMOST has carried out the pilot survey and the first year general survey, producing more than 2.2 million spectra, among which nearly 2 million are stellar spectra. Through statistical analysis of the physical properties related to stars in this large spectroscopic database spanning wide sky coverage, LAMOST will definitely enhance and increase our understanding of the structure, chemistry, and kinematics of the Galactic disk and the halo as well as its origin and evolution history. Efficient software that accurately derives the fundamental stellar atmospheric parameters plays a key role in scientific exploration. For this purpose, we developed and implemented the LAMOST stellar parameter pipeline (LASP, version 1.0).

#### 4.4.1 Input data and strategy for LASP

LASP has been used to automatically determine the stellar parameters (effective temperature:  $T_{\text{eff}}$ , surface gravity:  $\log g$ , metallicity:  $[\text{Fe}/\text{H}]$ , and radial velocity:  $V_r$ ) for late A and FGK type stellar

observations from the LAMOST survey. The input data for LASP are the output FITS spectra produced by the 2D and 1D pipeline. For DR1, the FITS file data version is v2.6.4 (according to status of the 2D and 1D pipelines, for DR2 the version will be upgraded to v2.7.5 after some improvement), with a resolving power of  $R \sim 1800$  covering wavelengths 3800 – 9000 Å. Observing nights for the survey are generally divided into three categories: dark nights (eight nights before and after the new moon), test nights (three nights around the full moon, used for instrumental testing), and bright nights (the remaining nights in a lunar cycle). The current criteria required for stars to be processed by LASP are: *final\_class* is STAR, *final\_subclass* is spectral type late A, F, G or K, and *g*-band SNR of  $SNR_g \geq 15$  and  $SNR_g \geq 6$  for the bright and dark nights, respectively. The *final\_class* and *final\_subclass* information is retrieved from the 1D pipeline results.

#### 4.4.2 Methods used by LASP

The LASP consecutively adopts two methods — Correlation Function Initial (CFI) value and UlySS (Wu et al. 2011a) — to determine stellar parameters. The CFI method is used to produce a set of coarse measurements that serve as initial guesses for ULYSS, which then generates the final values that are output by LASP.

1. *CFI method*: Initial stellar parameter estimates are provided by a method called CFI (Du et al. 2012). The core algorithm that is part of the CFI method is as follows: provided the observed flux vector is  $O$ , and the synthetic flux vector is  $S$ , the theoretical best-fit case is  $\cos\langle O, S \rangle = 1$ . The algorithm searches for the best-fit by maximizing  $\cos\langle O, S \rangle$  as functions of  $T_{\text{eff}}$ ,  $\log g$  and  $[\text{Fe}/\text{H}]$ , where  $\cos\langle O, S \rangle = \frac{O \cdot S}{|O| \times |S|}$ , which is also referred to as the correlation coefficient. Practically, CFI first explores temperature in the grid of synthetic spectra to determine the  $T_{\text{eff}}$  value of the input spectrum. After  $T_{\text{eff}}$  is fixed, the code searches for the  $[\text{Fe}/\text{H}]$  in metallicity space in the grid, and then with fixed  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$ , it searches for the  $\log g$ . According to maps of the correlation coefficient distribution for the three parameters, the reliability of  $T_{\text{eff}}$  estimation is relatively more credible, the  $[\text{Fe}/\text{H}]$  is less credible, and it is hard for the algorithm to determine a  $\log g$  solution. Thus the biweight mean of several  $\log g$  estimations with several correlation coefficient values is provided as the output.

For the production of the grid of synthetic spectra used by our CFI algorithm, we employed Kurucz spectrum synthesis code based on the ATLAS9 stellar atmosphere models provided by Castelli & Kurucz (2003). This grid is smoothed to the resolution of the LAMOST spectrograph ( $R \sim 1800$ ) and the pseudo-continuum of the grid is manually obtained by an experienced spectral analyzer. The analyzer manually marks the points which may be on the pseudo-continuum based on spectral features. Then spline curves are employed to fit the spectral pseudo-continuum. Finally, the points can be adjusted until the pseudo-continuum meets the requirements. The CFI's adopted synthetic library contains 8903 spectra, in a grid with stellar parameter coverage of:  $3600 \text{ K} \leq T_{\text{eff}} \leq 7500 \text{ K}$  in steps of 200 K,  $0.0 \text{ dex} \leq \log g \leq 5.0 \text{ dex}$  in steps of 0.25 dex, and  $-3.0 \leq [\text{Fe}/\text{H}] \leq 0.4$  in steps of 0.2 dex.

2. *ULySS method*: The ULYSS (Université de Lyon Spectroscopic analysis Software, Koleva et al. 2009; Wu et al. 2011a, available at <http://ulyss.univ-lyon1.fr/>) package is employed by LASP to derive the stellar parameters via minimizing the squared difference between the observations and the model. The model is adjusted to the same resolution and sampling as the observation, and the fit is performed in pixel space. This method determines all free parameters within a single fit in order to properly handle the degeneracy between some parameters, e.g., temperature and metallicity. Wu et al. (2011a) have already described the  $\chi^2$  minimization of ULYSS. The parametric model is

$$\text{Obs}(\lambda) = P_n(\lambda) \times [\text{TGM}(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}], \lambda) \otimes G(v_{\text{sys}}, \sigma)],$$

where  $\lambda$  is the logarithm of the wavelength,  $\text{Obs}(\lambda)$  is the observed stellar spectrum sampled in  $\log \lambda$ ,  $P_n(\lambda)$  a series of Legendre polynomials of degree  $n$ , and  $G(v_{\text{sys}}, \sigma)$  is a Gaussian broaden-



ing function characterized by residual velocity  $v_{\text{sys}}$  and dispersion  $\sigma$ . The  $\sigma$  encompasses both the instrumental broadening and the effects of rotation. The multiplicative polynomial is used to absorb errors in the flux calibration, Galactic extinction or any other source affecting the shape of the spectrum. It replaces the prior rectification or normalization by the pseudo-continuum that other methods require. The parameters of the TGM are evaluated together with those of the line-of-sight velocity distribution (LOSVD) using a Levenberg-Marquardt (Marquardt 1963; Moré et al. 1980) routine (hereafter LM). The (linear) coefficients of the  $P_n(\lambda)$  polynomials are determined by ordinary least-squares at each evaluation of the function minimized by the LM routine. The gaps and bad pixels (e.g. emission lines, cosmic rays, bad sky subtraction, telluric lines or bad ones due to instrument errors, etc.) are automatically rejected during the fitting process by ULYSS, using a clipping algorithm that is iteratively applied to the residuals of the fit.

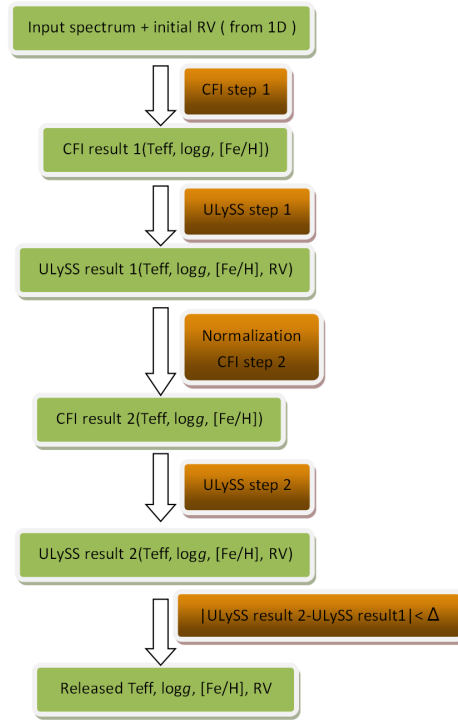
The non-linear TGM function is an interpolator that can return a spectrum for a given set of atmospheric parameters ( $T_{\text{eff}}$ ,  $\log g$ , [Fe/H]) by making an interpolation over the whole ELODIE library (Prugniel & Soubiran 2001, 2004; Prugniel et al. 2007). The interpolator (available at <http://ulyss.univ-lyon1.fr/models.html>) consists of polynomial expansions of each wavelength element in powers of  $\log T_{\text{eff}}$ ,  $\log g$ , [Fe/H] and  $f(\sigma)$  (a function of the rotational broadening parameterized by  $\sigma$ , the standard deviation of a Gaussian). Three sets of polynomials are defined for three temperature ranges (roughly matching OBA, FGK and M types) with an important overlap between each other where they are linearly interpolated. For the FGK and M polynomials, 26 terms are adopted; for OBA, 19 terms are used. The coefficients of these polynomials have been fitted over the  $\sim 2000$  spectra in the library, and the choice of the terms in  $T_{\text{eff}}$  limits and of weights were fine tuned to minimize the residuals between observations and the interpolated spectra. This interpolator built on the ELODIE library (version 3.2) provides valid inverted atmospheric parameters covering 3100  $\sim$  59 000 K in  $T_{\text{eff}}$ , 0.00  $\sim$  5.00 dex in  $\log g$  and  $-2.80 \sim 1.00$  dex in [Fe/H], from O to M type stars. The derived intrinsic external accuracies of ULYSS for high quality AFGK stellar spectra are 43 K, 0.13 dex and 0.05 dex for  $T_{\text{eff}}$ ,  $\log g$  and [Fe/H] respectively (Wu et al. 2011a). The performances of ULYSS applied to observations from the LAMOST commissioning period were already illustrated in Wu et al. (2011b).

#### 4.4.3 Processing procedure used by LASP

LASP operates in two stages to measure the stellar parameters. Since the LAMOST spectral flux calibration is relative (because there is no valid complete system for photometric standard stars), in the first stage we measure the original spectrum, and in the second stage we use the normalized spectrum. For the medium resolution of the LAMOST survey spectra, the blue band spectra contain most of the physical features necessary to constrain the stellar parameters. To avoid the effects of low instrumental response near both wavelength edges, as well as for optimizing the computing time and using minimal storage space, CFI selects wavelengths in the range 3850 – 5500 Å, and ULYSS selects those in the range 4400 – 5700 Å (in the 1st stage) and 4400 – 6800 Å (in the 2nd stage).

In each stage, we first utilize CFI to get a set of initial coarse estimates. After this, we use CFI results as starting guesses for ULYSS to obtain more accurate and credible measurements. Figure 4 displays a general flowchart for LASP.

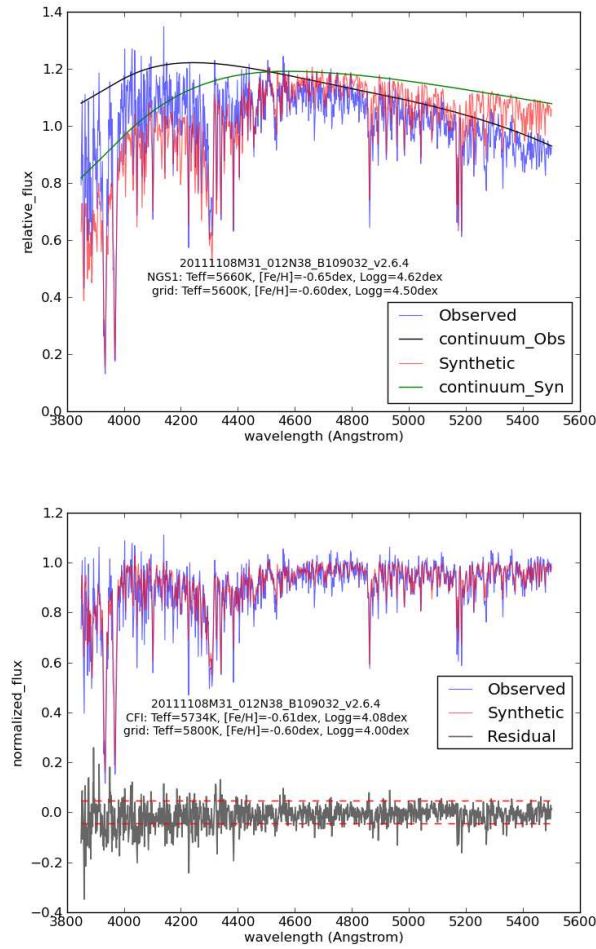
As CFI runs its core algorithm to determine the parameters individually one after another, it determines the best match between the observation and the synthetic grid spectra, both of which have been normalized. In the first stage of CFI, it uses the RV estimated by the 1D pipeline to shift the observed spectrum to the rest frame. Theoretically, observed spectra could be approximated by synthetic spectra multiplied by low order polynomials. To quickly get an initial guess of the effective temperature for CFI as a prior, the code uses the NGS1 grid (Lee et al. 2008). The NGS1 grid of synthetic spectra was generated from the Kurucz NEWODF models, which employ solar relative abundances from Grevesse & Sauval (1998). More details about generation of NGS1 synthetic



**Fig. 4** Flow chart illustrating the parameter estimation process of LAMP.

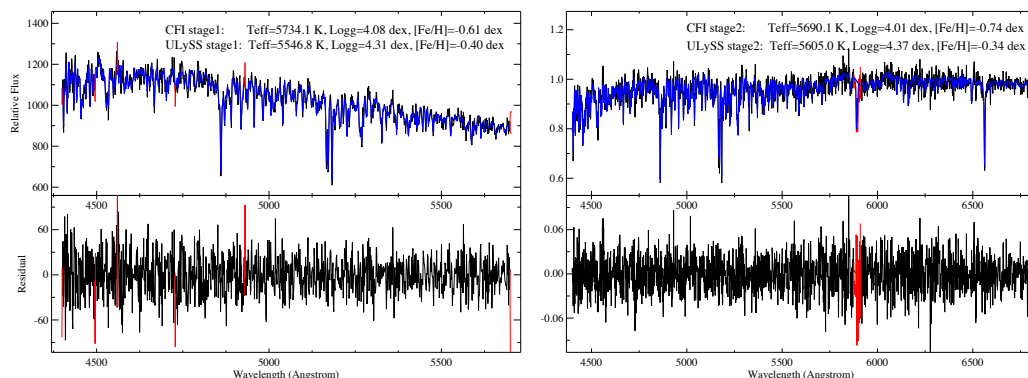
spectra were described in Lee et al. (2008). These synthetic spectra were convolved to  $R = 1000$ , then sampled into  $1.67 \text{ \AA}$  per pixel to be the NGS1 grid spectra. In order to speed up the analysis, only the spectral range of  $4500 \text{ \AA}$  to  $5500 \text{ \AA}$  was used. The synthetic spectra have been normalized by iteratively using a 9th order polynomial, rejecting points that are more than  $1\sigma$  below and  $4\sigma$  above the fitted function. The NGS1 grid covers  $4000 \text{ K} \leq T_{\text{eff}} \leq 9750 \text{ K}$  in steps of  $250 \text{ K}$ ,  $0.0 \text{ dex} \leq \log g \leq 5.0 \text{ dex}$  in steps of  $0.25 \text{ dex}$ , and  $-4.0 \leq [\text{Fe}/\text{H}] \leq +0.5 \text{ dex}$  in steps of  $0.25 \text{ dex}$  for the three parameters respectively. The observed LAMOST spectra are also processed in the same fashion over the same wavelength range, then we search for the best-fit using the CFI core algorithm to get an initial guess of  $T_{\text{eff}}$ . Then with this  $T_{\text{eff}}$ , CFI uses its core algorithm to search for the most similar synthetic spectra (with the already known pseudo-continuum, which could be subtracted before fitting) in the grid described before in Sec. 4.4.2. We use a 5th order polynomial to absorb the difference between the normalized observation and the normalized synthetic spectra, and finally get an estimation of the three parameters one after another. The grid described in Section 4.4.2 is more dense and has a bigger volume than the one used by NGS1. In the second stage of CFI, it takes the ULySS first stage results ( $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ ,  $\text{RV}$ ), which provide a more accurate  $\text{RV}$  to correct to the rest frame, and removes the need to use the NGS1 grid, since we can use the  $T_{\text{eff}}$  given by the 1st result from ULySS. Since the spectra processed in the second stage are normalized, CFI directly matches the spectra without a prior step of removing the pseudo-continuum. Figure 5 shows an example of the CFI spectral fitting in the blue wavelength range.

*Normalization* procedure between the first and the second stage. Before stellar spectra are parameterized, the pseudo-continuum should be divided out (or normalized). For the input observation, its analogous synthetic spectrum can be found in the synthetic grid (Sect. 4.4.2) according to the parameters produced by the ULySS 1st stage results. Since the pseudo-continuum of the grid spec-



**Fig. 5** An example of normalization through the CFI grid. *Top*: The pseudo-continuum derived from the corresponding parameter grid. *Bottom*: Colors show normalized flux of the blue wavelength range (*blue*) and the synthetic grid spectrum (*red*).

tra has been derived previously, the pseudo-continuum of the observed spectrum can be calculated through the analogous synthetic spectrum multiplied by a polynomial. In order to obtain a continuum fit over the whole range, the wavelength range is divided into four pieces ( $\leq 5600 \text{ \AA}$ ,  $5400\text{--}6400 \text{ \AA}$ ,  $6200\text{--}8000 \text{ \AA}$ ,  $\geq 7800 \text{ \AA}$ ) and a 5th order polynomial is applied to each of these ranges. After this piecewise fitting, the pseudo-continuum of the entire wavelength range is fitted by an 11th order polynomial, then the original 1D spectrum is normalized and delivered for the second stage. We performed experiments with this normalization technique on well calibrated SDSS/SEGUE (York et al. 2000; Yanny et al. 2009) survey spectra, and found that the difference between the determined parameters given by ULySS before and after normalization is negligible. However, when we applied this method to the LAMOST observations, the differences are noticeable. This difference is mainly



**Fig. 6** Fits of a LAMOST stellar spectrum with ULYSS (based on the ELODIE interpolator). This spectrum shows a star with `specid=20111108M31_012N38_B109032_v2.6.4` and  $g$ -band SNR=17.51. The top panels represent the relative flux distribution – the observation is in black and the model is in blue. Red lines represent the masked region. The bottom panels show the residuals between the observation and the best-fitted interpolated spectrum. The left panels show the original spectrum fitting, and right panels show the normalized spectrum fitting.

due to the unstable relative flux calibration of the spectra mentioned in Sec. 4.2.2; sometimes the shape of the flux-calibrated spectrum is quite strange, for example showing very steep regions or having several big fake humps or other spurious features, which obviously cannot be fitted by a fixed order polynomial. On the other hand, this simple intervening normalization procedure can potentially alter the shape of the original physical lines. So besides the  $g$ -band SNR, the difference in the ULYSS parameters between first and second stage results, to some extent, helps reflect the quality of the spectral data. If the two stages of ULYSS measurements are similar to each other, ULYSS determinations given by the normalized spectra are provided in the final publicly released stellar parameter catalog for late A and FGK-type spectra. Figure 6 illustrates one fitting example of ULYSS applied to LAMOST observations.

For observations with no prior knowledge, in order to prevent the ULYSS algorithm from becoming trapped in a local minimum, we use a grid of initial guesses that sample all of the parameter space. For example, take the nodes of the guessing grid as  $T_{\text{eff}} = [4000, 7000, 15000]$  K,  $\log g = [1.8, 3.8]$  dex, and  $[\text{Fe}/\text{H}] = [-1, 0]$  dex as adopted in the work of Wu et al. (2011b). The final solution (absolute minimum) is the best values obtained from those different guesses. For late A and FGK type spectra, by using a wide guessing grid, the global solution can statistically almost always be found, with only several cases where the solution was a local minimum occurring out of ten thousand fitted stars. Also, for some spectra, the solution simply expressed by three physical parameters may not be unique.  $\chi^2$  maps together with the convergence maps that are tools provided as part of the ULYSS package (see Wu et al. 2011a) can be used to visualize the parameter space and identify the degeneracies and local minima. We can see that when applying the above computational grid, each observation has to be fitted 12 times. For a survey project with a massive data flow, this is quite time consuming, which is why we use CFI before ULYSS to get single initial guess nodes. This effectively helps reduce the computational requirements, saving time and storage space for LASP by using fewer starting guesses in ULYSS.

Although CFI is just used to give a coarse estimation, for most cases the results used for the initial guesses in ULYSS are acceptable ( $T_{\text{eff}}$  is most reliable). In some cases, due to its low accuracy related to either the algorithm (for example, parameters which are not physically possible or the

spectral SNR is poor) or from poor parameter coverage of the synthetic spectral grid, some of the determinations given by CFI are not credible. These biases are obvious for very hot and cold stars, and also for wide regions in the  $\log g$  and  $[\text{Fe}/\text{H}]$  spaces. In practice, if any of the CFI parameters reside around the edges of its grid coverage, or in an unphysical region, or outside of the grid, ULYSS does not take its results, but instead either uses a moderate guess like (7000 K, 3.0 dex, 0.0 dex) or roughly estimates the starting value for  $T_{\text{eff}}$  from the subclass indicated by the 1D pipeline.

Note that for LAMOST DR1, if the star is processed by LASP (i.e., it has  $T_{\text{eff}}$ ,  $\log g$  and  $[\text{Fe}/\text{H}]$  values), the final released RV is determined by ULYSS. If the spectrum was not processed by LASP, the RV is inherited by the 1D pipeline, and for late M-type stars, the RV is from a separate module (inside the 1D pipeline) which is specially tailored for cooler stars (Guo et al. 2015).

#### 4.4.4 Why iterate once

As is known from the normalization described in Sec. 4.4.3, the quality of the calculated pseudo-continuum depends on: (1) the credibility of prior known stellar atmospheric parameters (stage1: CFI gets values from the NGS1 grid, stage2: get the ULYSS result derived from the normalized spectrum); (2) the accuracy of matching the most similar synthetic spectrum; (3) the density of the synthetic grid. The three aspects mentioned may introduce large discrepancies between the adopted synthetic spectrum and the observation. Therefore, if the quality of observation is not good, the pseudo-continuum of the observed spectrum based on the synthetic spectrum is unacceptable, and thus will lead to inconsistency between stellar parameters determined by ULYSS when using normalized and unnormalized spectra. Likewise, if the two stellar parameter estimations are consistent with each other, it can be concluded that the parameters are reliable and the pseudo-continuum of the observed spectrum calculated based on the synthetic one is acceptable. A reasonable degree of approximation can be measured statistically. In brief, if the two results approximately agree with each other, the stellar atmospheric parameters estimated using normalized spectra are adopted.

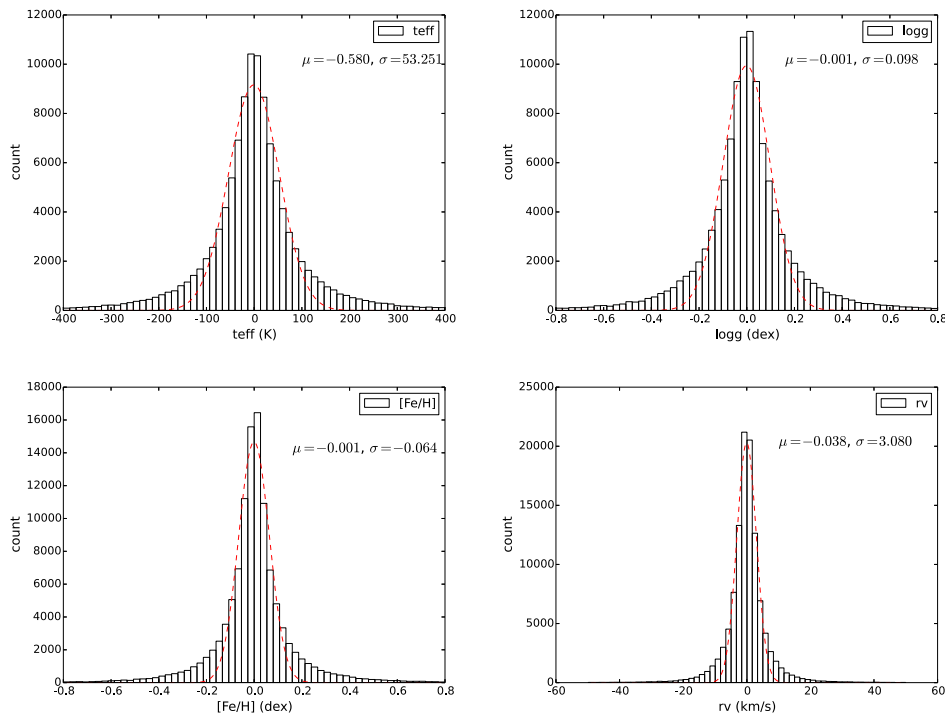
#### 4.4.5 Errors in the stellar atmospheric parameters produced by LASP

Since reliable random errors for each wavelength are unavailable for LAMOST spectra, like in the work of Wu et al. (2011a), when using ULYSS, we assumed a constant noise for all wavelength data points. We estimated an upper limit for the internal errors on the derived parameters by assuming  $\chi^2 = 1$ . We performed the fit with an arbitrary value of SNR and re-scaled the internal computed errors returned by ULYSS by multiplying them by  $\chi^2$ . The computed internal errors are small, likely for reasons related to internal degeneracies between the parameters. We used statistics from comparing with common targets observed by both LAMOST and SDSS/SEGUE (Wu et al. 2014) to estimate the external error. We then rescaled the errors using the ratio of the differences between LAMOST and SDSS parameters to the formal errors, assuming that the uncertainties of each series are equivalent. Currently the external errors determined in this way, which are released in the parameter catalog, are larger than the real measurement errors.

#### 4.4.6 Internal comparison

In total, about 18% of the targets in LAMOST have been observed more than one time. These multiple exposures can be used to estimate the internal errors on the parameters. We selected 111 000 targets with two observations to estimate the internal errors.

Figure 7 demonstrates the internal error distribution (i.e., the difference between parameters measured for the same star with different spectra) including a comparison of  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$  and RV.



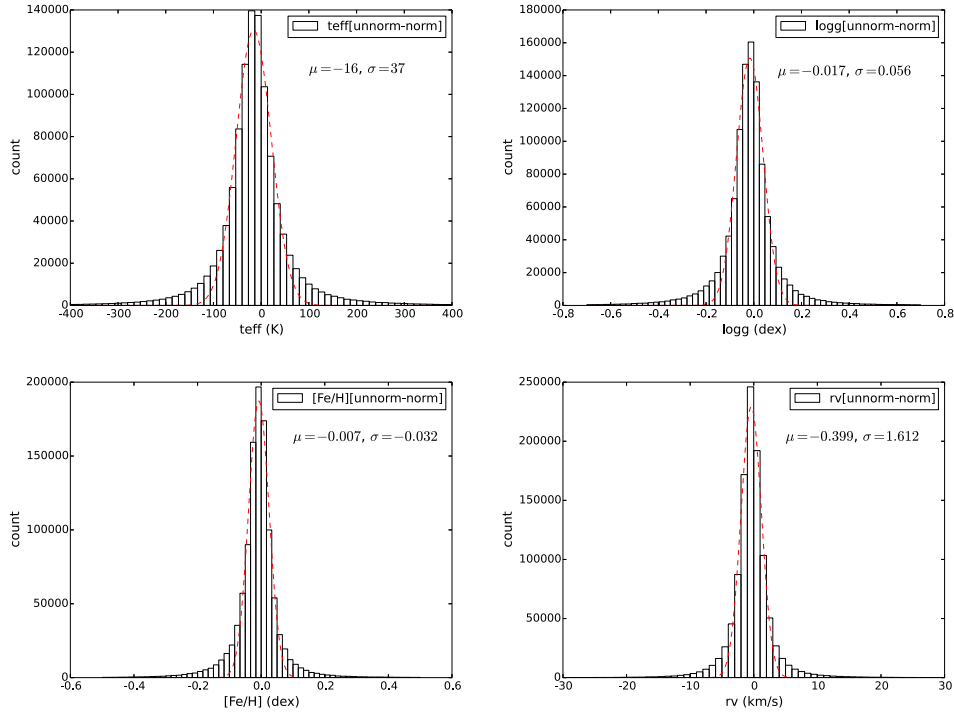
**Fig. 7** Histograms comparing differences in measured stellar parameters for pairs of observations of the same star. The panels show the internal consistency of  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$  and RV for 111 000 targets.

To some extent, another way to estimate the internal error is by comparing the parameters calculated through normalized spectra and unnormalized spectra. Figure 8 shows the distribution of the difference between the parameters measured using normalized and unnormalized spectra.

#### 4.4.7 External errors

We compared the common targets that have both LAMOST DR1 LASP measurements and high resolution spectral results (Bruntt et al. 2012; Thygesen et al. 2012; Molenda-Żakowicz et al. 2013; Dong et al. 2014). For this comparison,  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$  come from the high resolution spectra, while  $\log g$  comes from the *Kepler* (Borucki et al. 2010) asteroseismic estimation. The comparison between LAMOST DR1 and high-resolution data is shown in Figure 9, along with statistical fits characterizing the quality of agreement. Figure 10 shows a similar comparison for the common targets between LAMOST (LASP results) and SDSS DR9.

There are several independent works comparing parameters of DR1 with other reliable external databases. For example, Gao et al. (2015, in preparation) compare DR1 parameters with the PASTEL catalogue. Results from the LAMOST Stellar Parameter Pipeline at Peking University (LSP3) (Xiang et al. 2015), an independent parameter pipeline developed at Peking University, show very good consistency with LAMOST/LASP values. All of these comparisons are summarized in Table 2.



**Fig. 8** Same as in Fig. 7, but now showing the difference between parameters measured using normalized and unnormalized spectra.

**Table 2** External Comparison with DR1

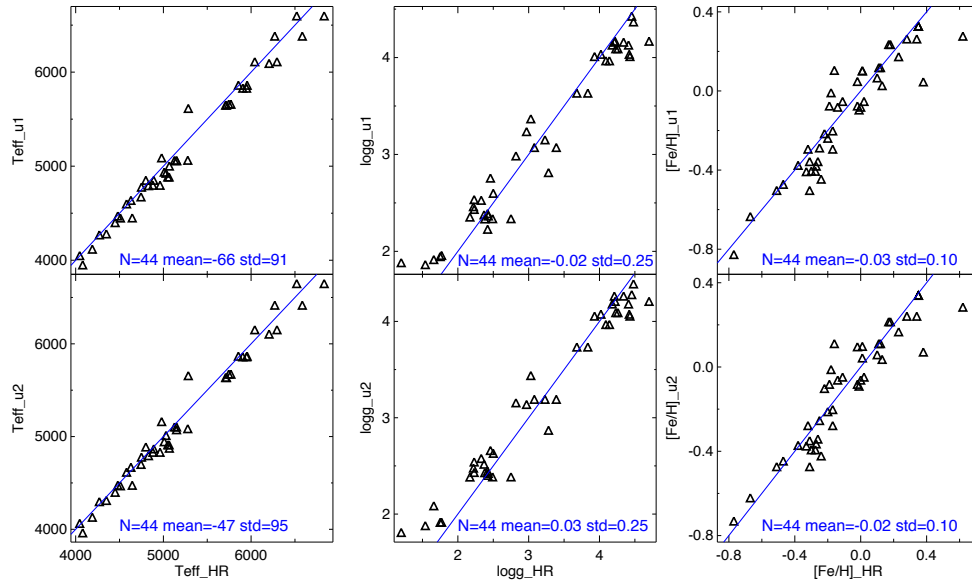
	$T_{\text{eff}}$ error	$\log g$ error	[Fe/H] error	RV error
High resolution	$-47 \pm 95$ K	$0.03 \pm 0.25$ dex	$-0.02 \pm 0.1$ dex	–
SDSS DR9	$-91 \pm 111$ K	$0.16 \pm 0.22$ dex	$0.04 \pm 0.15$ dex	$-7.2 \pm 6.6$ km s $^{-1}$
Gao et al. 2015	$11 \pm 110$ K	$0.03 \pm 0.19$ dex	$0.01 \pm 0.11$ dex	$-3.78 \pm 4.91$ km s $^{-1}$
LSP3	$7 \pm 133$ K	$0.02 \pm 0.22$ dex	$0.06 \pm 0.14$ dex	$1.2 \pm 3.1$ km s $^{-1}$

Huang & Zhang (private communication) also compared the parameters given by DR1 with those from APOGEE, as shown in Figure 11. Besides these, Huang et al. (2015) have compared 499 radial velocities (from DR1) with RVs of the same targets from MMT+Hectospec, which has an RV accuracy of  $\sim 2.5$  km s $^{-1}$  (Drout et al. 2009, 2012). The result is  $V_r(\text{DR1}) - V_r(\text{MMT}) = -3.76 \pm 6.65$  km s $^{-1}$  (see Huang et al. 2015).

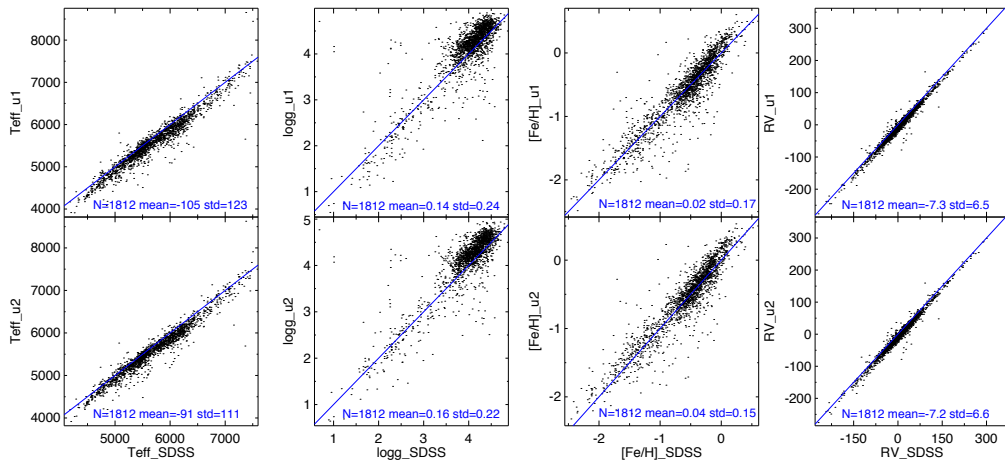
## 5 DATA PRODUCTS

LAMOST DR1 consists of 2 204 860 spectra of stars, galaxies, quasars and unknown objects. The footprint on the sky is shown in Figure 12, and the distributions of stellar parameters for all stars in RV vs. [Fe/H] and  $\log g$  vs.  $T_{\text{eff}}$  are shown in Figure 13. DR1 includes the following:

1. *Spectra*. — In general, there are 2 204 860 flux- and wavelength-calibrated, sky-subtracted spectra released publicly in DR1. DR1 includes 1 944 329 stars, 12 082 galaxies, 5017 quasars and 243 355 unknown objects. These spectra cover the range 3690–9100 Å with a resolution of 1800.

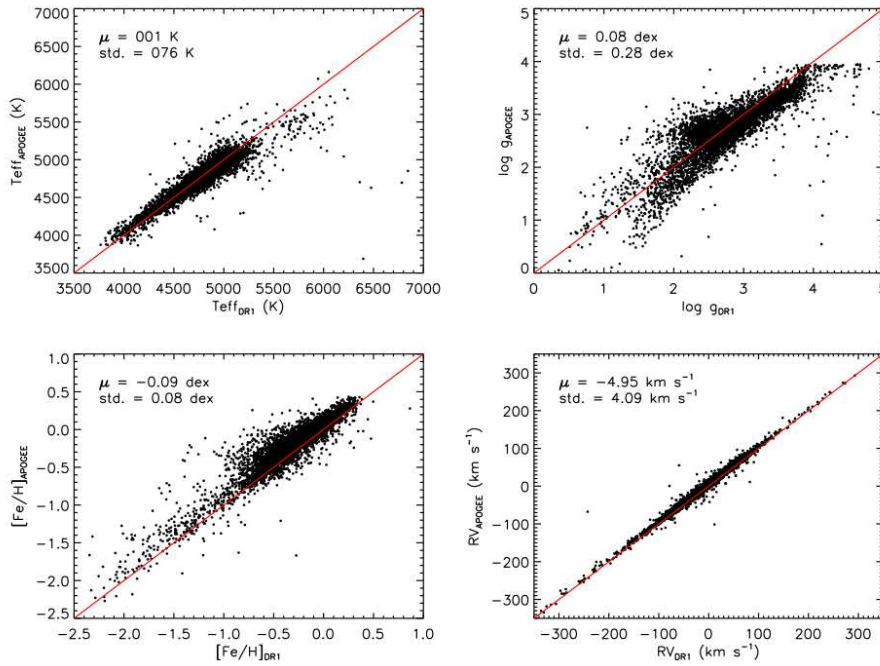


**Fig. 9** Comparison of the three parameters measured by LASP with LAMOST DR1 spectra and results from high resolution (HR) spectra (or *Kepler* asteroseismology, in the case of  $\log g$ ) for the same stars.  $u_1$  means the ULYSS determination from the original spectra and  $u_2$  means the ULYSS determination from normalized spectra.

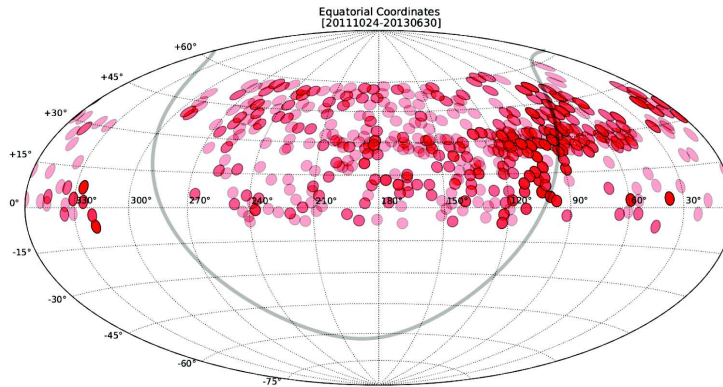


**Fig. 10** Comparison of the three parameters between LASP and SDSS DR9. The convention is the same as described in Fig. 9.



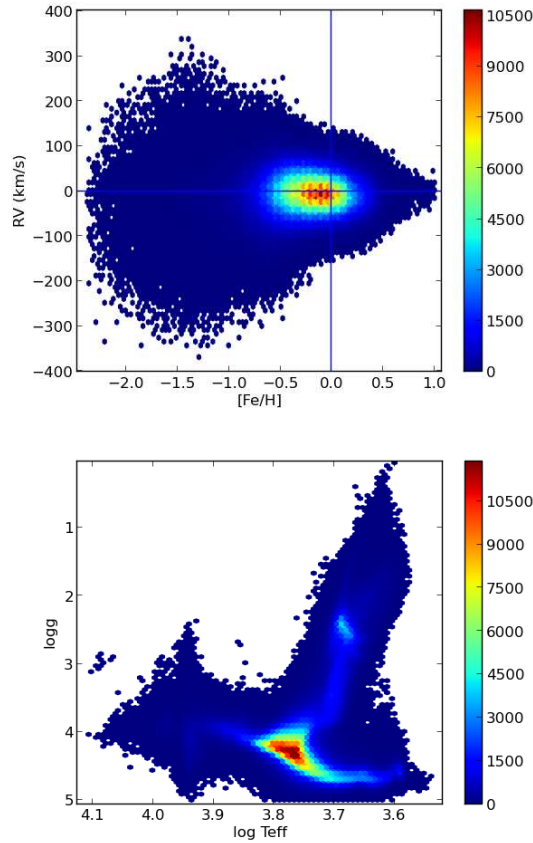


**Fig. 11** Comparison of stellar parameters for stars in common between LAMOST DR1 and APOGEE.



**Fig. 12** The footprint of LAMOST DR1 is shown in equatorial coordinates, with red circles representing LAMOST plates. A darker color means more repeated observations.

*2. Spectroscopic Parameter Catalogs.* — In this data release, four spectroscopic parameter catalogs are included: the DR1 general catalog, the A type star catalog, the late A, F, G and K type stellar parameter catalog, and the M star catalog. All of these include 30 basic pieces of information about each spectrum; e.g., RA, Dec, SNR, magnitude, spectral type and redshift, among others. In addition,



**Fig. 13** The distribution of stars in LAMOST DR1 in the planes of RV vs. [Fe/H] (*top*) and  $\log g$  vs.  $T_{\text{eff}}$  (*bottom*).

the A type star catalog also publishes four Balmer line indices and four spectral line widths at 20% below the local continuum; the late A, F, G and K type star catalog provides effective temperature, surface gravity and metallicity; and the M star catalog additionally contains a flag to indicate if a star shows magnetic activity.

## 5.1 FITS File Description

### 5.1.1 FITS file designation

The file names of the released FITS files are in the form ‘spec-MMMMM-YYYY\_spXX-FFF.fits’, where ‘MMMMM’ represents the local modified Julian day (MJD), ‘YYYY’ is the plan identity string (PLANID), ‘XX’ indicates the spectrograph number (between 1 and 16), and ‘FFF’ shows the fiber number (between 1 and 255). In addition, we provide a LAMOST designation for an object via the DESIG keyword that is formed as ‘LAMOST JHHMMSS.ss+DDMMSS.ss’, where ‘HHMMSS.ss’ is RA in units of HMS and ‘+DDMMSS.ss’ is Dec in units of DMS.

### 5.1.2 Primary FITS header

We reorganize keywords of the primary FITS header in DR1 and divided them into eight groups including mandatory keywords, file information keywords, telescope parameter keywords, observation parameter keywords, spectrograph parameter keywords, weather condition keywords, data reduction parameter keywords and spectral analysis results. See details in the appendix (provided in the online version of this paper).

### 5.1.3 Primary data array

The primary data array has five rows and NAXIS1 (a keyword explained in the online version) columns. The five rows are Flux, Inverse Variance, Continuum-Subtracted Flux, Andmask and Ormask respectively. It is noted that the ‘inverse variance’ of the uncertainties (one over sigma-squared) can be used to estimate the signal-to-noise ratio of each pixel ( $flux * (inverse\ variance)^{0.5}$ ), and ‘Andmask’/‘Ormask’ are mask flags of six quality situations for each pixel. The details of these two masks are listed in Table A.9 in the online version.

## 5.2 LAMOST Catalogs

Four LAMOST DR1 catalogs are released and available at the website <http://dr1.lamost.org/>, including the LAMOST general catalog, the A type star catalog, the late A, F, G and K type star catalog and the M star catalog. Detailed information is listed in tables (Table A.11, Table A.12 and Table A.13 in the online version) for each catalog.

The uncertainties of measurements in DR1 are not included since the error estimation is unreliable and needs to be improved; however, the uncertainties will be released in DR2.

### 5.2.1 LAMOST general catalog

The LAMOST general catalog not only includes 717 660 spectra from the LAMOST pilot survey, comprised of 648 820 stellar spectra, 2723 galaxy spectra, 621 quasar spectra and 65 496 unknown spectra, but also 1 487 200 spectra from the LAMOST regular survey, which includes 1 295 586 stellar spectra, 9359 galaxy spectra, 4396 quasar spectra and 177 859 unknown spectra. In this catalog, there are 1 221 538 spectra with SNR in the *g*-band larger than 10, 1 839 808 spectra with SNR in *i*-band larger than 10, and 1 198 410 spectra with SNR in both *g*- and *i*-bands larger than 10.

All parameters from this catalog are listed in Table A.10 (online version). For galaxies and quasars, ‘z’ provides their redshifts, and their ‘z’ is set to –9999 if their redshift is unable to be estimated. For stars, ‘rv’ is their heliocentric radial velocity, and the ‘rv’ is set to –9999 if their rv is unable to be estimated. For objects with unknown type, their ‘z’ and ‘rv’ are both set to –9999. In addition, for a galaxy or a quasar, the value of ‘rv’ is empty, and for a star, the value of ‘z’ is null.

On the website <http://dr1.lamost.org/>, we provide two formats with the same contents for the LAMOST general catalog; one catalog is in .fits format and the other is .csv format.

### 5.2.2 A type star catalog

There are 100 073 A-type stellar spectra in this catalog from both the pilot and regular surveys.

Table A.11 (online version) shows all fields associated with this catalog. The ‘Class’ field presents the two-dimensional spectral classification results, and nearly all A type stars have luminosity class provided by the LAMOST analysis pipeline. Balmer line indices are included in ‘Halpha\_Indice’, ‘Hbeta\_Indice’, ‘Hgama\_Indice’, and ‘Hdelta\_Indice’ fields, and ‘Halpha\_D0.2’, ‘Hbeta\_D0.2’, ‘Hgama\_D0.2’, and ‘Hdelta\_D0.2’ are the widths at 20% below the local continuum of these four Balmer lines. However, these eight fields are temporarily empty, and will be released

in September of 2015. The A type star catalog in .csv format can be downloaded from the website (<http://dr1.lamost.org>).

### 5.2.3 AFGK high quality stellar parameter catalog

The AFGK high quality stellar parameter catalog includes 61 686 A type stars, 548 214 F type stars, 253 275 G type stars and 222 229 K type stars. These spectra are selected with  $\text{SNR}_g \geq 6$  from dark-night observations and  $\text{SNR}_g \geq 20$  from bright nights. The effective temperatures ( $T_{\text{eff}}$ ), surface gravities ( $\log g$ ) and metallicities (Fe/H) are determined by LASP. The A type stars in this catalog are those from the A type star catalog with high SNR spectra. These data are available from the website <http://dr1.lamost.org>.

### 5.2.4 M star catalog

The M star catalog contains 121 522 spectra, and their parameters are listed in Table A.13. The basic parameters are provided like those in other catalogs, but the magnetic activity flag ‘magact’ is also included in this catalog. The flag will be assigned to be 1 if stars have magnetic activity, 0 if there is no magnetic activity and -9999 if the magnetic activity could not be determined due to low SNR. The specific catalog for M dwarfs is given in (Guo et al. 2015). The M star catalog in .csv format can be obtained from the website <http://dr1.lamost.org>.

## 5.3 Caveats

### 5.3.1 Radial velocity error

The DR1 pipeline adopted local Beijing Time rather than UTC during the correction from geocentric coordinates to heliocentric coordinates. Thus an error in radial velocity is introduced, with a mean value of  $0.36 \text{ km s}^{-1}$ . In addition, the RV errors of some A type stars reach  $80 \text{ km s}^{-1}$  because their wavelength scale was incorrectly calibrated with air wavelengths instead of vacuum wavelengths. These are listed on the DR1 website.

### 5.3.2 Spectral line errors coming from a combination of sub-exposures

The LAMOST spectra released in DR1 are combined from several sub-exposures. If one of the three sub-exposures is problematic, spectral lines may be erroneously scaled (clipped) during the B-spline fitting by using the wrong data points. There are 66 838 spectra affected in DR1, which are listed on the DR1 website.

### 5.3.3 Offset from very bright targets

Some coordinates of bright objects are artificially shifted to avoid saturated exposures. However, we list the original coordinates in the DR1 website.

### 5.3.4 Mistaken spectra due to a problem with the fiber

Problems were found for tens of thousands of spectra in DR1 that are due to fibers or positioners. These were discovered after the data release, and include several thousand spectra that have stellar parameters  $T_{\text{eff}}$ ,  $\log g$  and [Fe/H]. A separate list of these problematic spectra is provided on the DR1 website along with their IDs, and will be excluded from DR2.

## 6 DATA ACCESS AND USE

The LAMOST DR1 can be accessed at the LAMOST Data Release web portal, <http://dr1.lamost.org>, which has been developed under a collaboration with the Chinese Virtual Observatory (China-VO). Web form based database search and data retrieval are available for all interested users. Advanced features are provided for registered users, for example batch search, bulk retrieval and personal space.

For registered users, lists of identifiers or coordinates supported by CDS SIMBAD<sup>1</sup> and VizieR<sup>2</sup>, or lists of LAMOST identifiers are acceptable to upload and use to search the database in batch mode. SQL-like queries are also supported. Virtual Observatory (VO) tools, such as Aladin<sup>3</sup> and SpecView<sup>4</sup>, and tools developed specifically for LAMOST are integrated into the web portal. Search results and spectra can be interactively displayed and analyzed. Large results and datasets can be retrieved in several ways, including FTP, wget, cloud storage and VO data access interfaces.

Based on the CDS X-Match Service<sup>5</sup>, cross identification with dozens of well-known catalogs is also provided as an advanced feature of the data release portal.

According to the LAMOST data policy, any publication making use of LAMOST data should cite the standard acknowledgement as follows ‘The 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 National Astronomical Observatories, Chinese Academy of Sciences.’

**Acknowledgements** The work was funded by the National Basic Research Program of China (973 Program, 2014CB845700) and the National Natural Science Foundation of China (Grant Nos.11390371). The 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 National Astronomical Observatories, Chinese Academy of Sciences. The authors deeply thank Prof. Georges Comte for valuable discussions.

## References

- Batalha, N. M., Borucki, W. J., Koch, D. G., et al. 2010, *ApJ*, 713, L109  
 Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, 327, 977  
 Bovy, J., Hénnaoui, J. F., Hogg, D. W., et al. 2011, *ApJ*, 729, 141  
 Bruntt, H., Basu, S., Smalley, B., et al. 2012, *MNRAS*, 423, 122  
 Carlin, J. L., DeLaunay, J., Newberg, H. J., et al. 2013, *ApJ*, 777, L5  
 Chen, J.-J., Bai, Z.-R., Luo, A.-L., & Zhao, Y.-H. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 9149, 1  
 Chen, L., Hou, J.-L., Yu, J.-C., et al. 2012, *RAA (Research in Astronomy and Astrophysics)*, 12, 805  
 Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *RAA (Research in Astronomy and Astrophysics)*, 12, 1197  
 Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, *RAA (Research in Astronomy and Astrophysics)*, 12, 735

<sup>1</sup> <http://simbad.u-strasbg.fr/simbad/>

<sup>2</sup> <http://vizier.u-strasbg.fr/viz-bin/VizieR>

<sup>3</sup> <http://aladin.u-strasbg.fr/aladin.gml>

<sup>4</sup> [http://www.stsci.edu/institute/software\\_hardware/specview/](http://www.stsci.edu/institute/software_hardware/specview/)

<sup>5</sup> <http://cdsxmatch.u-strasbg.fr/xmatch>

- Dong, S., Zheng, Z., Zhu, Z., et al. 2014, *ApJ*, 789, L3
- Drout, M. R., Massey, P., & Meynet, G. 2012, *ApJ*, 750, 97
- Drout, M. R., Massey, P., Meynet, G., Tokarz, S., & Caldwell, N. 2009, *ApJ*, 703, 441
- Du, B., Luo, A., Zhang, J., Wu, Y., & Wang, F. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 8451, 37
- Guo, Y. X., Yi, Z. P., Luo, A. L., et al. 2015, *RAA (Research in Astronomy and Astrophysics)*, 15, 1182
- Huang, Y., Liu, X. W., Zhang, H. W., et al. 2015, *RAA (Research in Astronomy and Astrophysics)*, 15, 1240
- Huo, Z.-Y., Liu, X.-W., Yuan, H.-B., et al. 2010, *RAA (Research in Astronomy and Astrophysics)*, 10, 612
- Huo, Z.-Y., Liu, X.-W., Xiang, M.-S., et al. 2013, *AJ*, 145, 159
- Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, *A&A*, 501, 1269
- Lasker, B. M., Lattanzi, M. G., McLean, B. J., et al. 2008, *AJ*, 136, 735
- 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. 2014, in *IAU Symposium*, 298, eds. S. Feltzing, G. Zhao, N. A. Walton, & P. Whitelock, 310
- Liu, X. W., Zhao, G., & Hou, J. L. 2015, *RAA (Research in Astronomy and Astrophysics)*, 15, 1089
- Luo, A.-L., Zhang, H.-T., Zhao, Y.-H., et al. 2012, *RAA (Research in Astronomy and Astrophysics)*, 12, 1243
- Marquardt, D. W. 1963, *Journal of the Society for Industrial & Applied Mathematics*, 11, 431
- Molenda-Żakowicz, J., Sousa, S. G., Frasca, A., et al. 2013, *MNRAS*, 434, 1422
- Moré, J. J., Garbow, B. S., & Hillstom, K. E. 1980, *User guide for MINPACK-1*, Tech. rep., CM-P00068642
- Peng, N., Zhang, Y., Zhao, Y., & Wu, X.-b. 2012, *MNRAS*, 425, 2599
- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, *A&A*, 323, L49
- Prugniel, P., & Soubiran, C. 2001, *A&A*, 369, 1048
- Prugniel, P., & Soubiran, C. 2004, *astro-ph/0409214*
- Prugniel, P., Soubiran, C., Koleva, M., & Le Borgne, D. 2007, *astro-ph/0703658*
- Ren, J., Luo, A., Li, Y., et al. 2013, *AJ*, 146, 82
- Ren, J., Wang, J., Li, K., F. H., et al. 2007, *Acta Astronomica Sinica*, 48, 500
- Roeser, S., Demleitner, M., & Schilbach, E. 2010, *AJ*, 139, 2440
- Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, *VizieR Online Data Catalog*, 7260, 0
- Shi, Z. X., Comte, G., Luo, A. L., et al. 2014a, *A&A*, 564, A89
- Shi, Z.-X., Luo, A.-L., Comte, G., et al. 2014b, *RAA (Research in Astronomy and Astrophysics)*, 14, 1234
- Song, Y.-H., Luo, A.-L., Comte, G., et al. 2012, *RAA (Research in Astronomy and Astrophysics)*, 12, 453
- Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, *AJ*, 123, 485
- Thygesen, A. O., Frandsen, S., Bruntt, H., et al. 2012, *A&A*, 543, A160
- Wu, X.-B., & Jia, Z. 2010, *MNRAS*, 406, 1583
- Wu, X.-B., Chen, Z.-Y., Jia, Z.-D., et al. 2010a, *RAA (Research in Astronomy and Astrophysics)*, 10, 737
- Wu, X.-B., Jia, Z.-D., Chen, Z.-Y., et al. 2010b, *RAA (Research in Astronomy and Astrophysics)*, 10, 745
- Wu, X.-B., Hao, G., Jia, Z., Zhang, Y., & Peng, N. 2012, *AJ*, 144, 49
- Wu, Y., Singh, H. P., Prugniel, P., Gupta, R., & Koleva, M. 2011a, *A&A*, 525, A71
- Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011b, *RAA (Research in Astronomy and Astrophysics)*, 11, 924
- Wu, Y., Luo, A., Du, B., & Guo, Y. 2014, in *IAU Symposium*, 298, eds. S. Feltzing, G. Zhao, N. A. Walton, & P. Whitelock, 445
- Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2015, *MNRAS*, 448, 822
- Yang, F., Carlin, J. L., Liu, C., et al. 2012, *RAA (Research in Astronomy and Astrophysics)*, 12, 781
- Yang, M., et al. 2015, *MNRAS*, in press
- Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, *AJ*, 137, 4377
- Yao, S., Liu, C., Zhang, H.-T., et al. 2012, *RAA (Research in Astronomy and Astrophysics)*, 12, 772
- Yi, Z., Luo, A., Song, Y., et al. 2014, *AJ*, 147, 33

- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579
- Yuan, H.-B., Liu, X.-W., Huo, Z.-Y., et al. 2010, *RAA (Research in Astronomy and Astrophysics)*, 10, 599
- Yuan, H., Zhang, H., Lei, Y., & Dong, Y. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 8448, 2
- Yuan, H., Zhang, H., Zhang, Y., Lei, Y., & Dong, Y. 2014a, in *IAU Symposium*, 298, eds. S. Feltzing, G. Zhao, N. A. Walton, & P. Whitelock, 452
- Yuan, H., Liu, X., Huo, Z., et al. 2014b, arXiv:1412.6628
- Zhang, Y., & Zhao, Y. 2003, *PASP*, 115, 1006
- Zhang, Y.-Y., Carlin, J. L., Yang, F., et al. 2012, *RAA (Research in Astronomy and Astrophysics)*, 12, 792
- 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 (Research in Astronomy and Astrophysics)*, 12, 723
- Zhao, J. K., Luo, A. L., Oswalt, T. D., & Zhao, G. 2013, *AJ*, 145, 169
- Zou, H., Yang, Y.-B., Zhang, T.-M., et al. 2011, *RAA (Research in Astronomy and Astrophysics)*, 11, 1093

**Appendix A: TABLE LIST**

There are nine tables for FITS files, these values must be written in a fixed format, including:

Mandatory Keywords	(Table A.1)
File Information Keywords	(Table A.2)
Telescope Parameter Keywords	(Table A.3)
Observation Parameter Keywords	(Table A.4)
Spectrograph Parameter Keywords	(Table A.5)
Weather Condition Keywords	(Table A.6)
Data Reduction Parameter Keywords	(Table A.7)
Spectral Analysis Result Keywords	(Table A.8)
The significance of six bits of ‘Andmask’ and ‘Ormask’	(Table A.9)

There are another four tables,

The general catalog	(Table A.10)
A-type star catalog	(Table A.11)
Late A, FG K-type catalog	(Table A.12)
M-type star catalog	(Table A.13)

**Table A.1** Mandatory Keywords

Keywords	Value	Comment
SIMPLE	T	Primary Header created by MWRFITS v1.8
BITPIX	-32	
NAXIS	2	
NAXIS1	3909	
NAXIS2	5	
EXTEND	T	Extensions may be present

Notes: **1. SIMPLE** It is required to be the first keyword in the primary header of all FITS files. The value field should contain a logical constant with the value T if the file conforms to this standard. This keyword is mandatory for the primary header and is not permitted in extension headers. A value of F signifies that the file does not conform to this standard. **2. BITPIX** The value field should contain an integer, and it should specify the number of bits that represent a data value. A value of -32 represents IEEE single precision floating point. **3. NAXIS** The value field should contain a non-negative integer no greater than 999, representing the number of axes in the associated data array. A value of zero signifies that no data follow the header in the HDU. **4. NAXIS1 and NAXIS2** The value field of these two indexed keywords should contain a non-negative integer, representing the number of elements along axis n of a data array. The NAXIS1 keyword represents the number of the wavelength array, i.e., the column number of the primary data array, and the NAXIS2 keyword indicates the row number of the primary data array. **5. EXTEND** The value field should contain a logical value indicating whether the FITS file is allowed to contain conforming extensions following the primary HDU. This keyword may only appear in the primary header and must not appear in an extension header. If the value field is T then there may be conforming extensions in the FITS file following the primary HDU. This keyword is only advisory, so its presence with a value T does not require that the FITS file contains extensions, nor does the absence of this keyword necessarily imply that the file does not contain extensions.



**Table A.2** File Information Keywords

Keywords	Value	Comment
FILENAME	'spec-55859-F5902_sp01-001.fits'	
AUTHOR	'LAMOST Pipeline'	Who compiled the information
N_EXTEN	1	The extension number
EXTEN0	'Flux Inverse Subcontinuum Andmask Ormask'	
ORIGIN	'NAOC-LAMOST'	Organization
DATE	'2013-07-17T10:23:12'	(UTC)

Notes: **1. FILENAME** The value field should contain a character string giving the name of this FITS file. Take the 'spec-55859-F5902\_sp01-001.fits' as an example, '55859' is the local modified Julian day, 'F5902' is the plan ID, 'sp01' is the spectrograph ID, and '001' is the Fiber ID. **2. AUTHOR** This keyword contains a string constant 'LAMOST Pipeline', which represents the author who produced this file. **3. N\_EXTEN** The value field should contain an integer giving the extension number of a FITS file. **4. EXTEN0** This keyword contains a string constant 'Flux Inverse Subcontinuum Andmask Ormask' explaining each row of the primary data array in a primary HDU. **5. ORIGIN** This ORIGIN keyword contains a string constant 'NAOC-LAMOST', which indicates the organization responsible for this FITS file. 'NAOC' represents the abbreviation of National Astronomical Observatories, Chinese Academy of Sciences. **6. DATE** The value field should contain a character string giving the UTC time when this FITS file is created.

**Table A.3** Telescope Parameter Keywords

Keywords	Value	Comment
TELESCOP	'LAMOST'	GuoShouJing Telescope
LONGITUD	117.580	Longitude of site [°]
LATITUDE	40.3900	Latitude of site [°]
FOCUS	19964	Telescope focus [mm]
CAMPRO	'NEWCAM'	Camera program name
CAMVER	'v2.0'	Camera program version

Notes: **1. TELESCOP** This keyword contains a string constant 'LAMOST' giving the name of our telescope. **2. LONGITUD** The keyword contains a floating-point constant, which provides the longitude of Xinglong station where LAMOST is located. **3. LATITUDE** The keyword contains a floating-point constant, which provide the latitude of Xinglong station. **4. FOCUS** The keyword gives the telescope focus, and its unit is millimeter. **5. CAMPRO** The value field contains a string constant 'NEWCAM', which shows the name of the camera. **6. CAMVER** The value field contains a character string 'v2.0', which gives the present version of the program controlling the camera.

**Table A.4** Observation Parameter Keywords

Keywords	Value	Comment
DATE-OBS	'2013-06-03T16:47:08.38'	The observation median UTC
DATE-BEG	'2013-06-04T00:13:52.0'	The observation start local time
DATE-END	'2013-06-04T01:21:17.0'	The observation end local time
MJD	56447	Local Modified Julian Day
MJDLIST	'56447-56447'	Local Modified Julian Day list
MJMLIST	'81283693-81283731'	Local Modified Julian Minute list
PLANID	'HD174640N254456F01'	Plan ID in use
RA	266.506521500	Right ascension of object [°]
DEC	23.7254148000	Declination of object [°]
DESIG	'LAMOST J174601.56+234331.4'	Designation of LAMOST target
FIBERID	80	Fiber ID of Object
CELL_ID	'E0818 '	Fiber Unit ID on the focal plane
X_VALUE	51.4330635340	X coordinate of object on the focal plane [mm]
Y_VALUE	703.039784897	Y coordinate of object on the focal plane [mm]
OBJNAME	'230388178985208'	Name of object
OBJTYPE	'Star'	Object type from input catalog
OBJSOURC	'LEGUE.LCH'	Name of input catalog
FIBERTYP	'Obj'	Fiber type of object
MAGTYPE	'gri'	Magnitude type of object
MAG1	17.94	Mag1 of object [mag]
MAG2	17.47	Mag2 of object [mag]
MAG3	17.24	Mag3 of object [mag]
MAG4	99.00	Mag4 of object [mag]
MAG5	99.00	Mag5 of object [mag]
MAG6	99.00	Mag6 of object [mag]
MAG7	99.00	Mag7 of object [mag]
OBS_TYPE	'OBJ'	The type of target (OBJ, FLAT, ARC or BIAS)
OBSCOMM	'Science'	Science, Test
RADECSYS	'FK5'	Equatorial coordinate system
EQUINOX	2000.00	Equinox in years

Notes: **1. DATE-OBS** The value field should contain a character string, which gives the median UTC of multiple exposures. **2. DATE-BEG** The value field should contain a character string giving the start of the observation according to Beijing Time. **3. DATE-END** The value field should contain a character string, which provides the end of the observation according to Beijing Time. **4. MJD** The value field should be a non-negative integer giving the local modified Julian day. **5. MJDLIST** The value field should contain a character string, which shows a list of the local modified Julian day of n exposures. **6. MJMLIST** The value field should contain a character string, which shows a list of local modified Julian minute of n exposures. **7. PLANID** The value field should contain a character string providing the plan ID of the target. **8. RA** The value field should contain a non-negative real floating-point number, which gives the right ascension of the target. **9. DEC** The value field should contain a non-negative real floating-point number, which gives the declination of the target. **10. DESIG** The value field should contain a character string, which indicates the name of the LAMOST target. Like the name of the SDSS target, numbers after the character 'J' and before '+' represent RA in the unit of HMS, and numbers after the character '+' are DEC in the unit of DMS. **11. FIBERID** The value field should contain a non-negative integer between 1 and 250, which shows the fiber ID and should be used together with the spectrograph ID. **12. CELL\_ID** The value field should contain a character string, which gives the fiber unit ID on the focal plane. The LAMOST focal plane is divided into four quadrants named 'EFGH' respectively; the first character of this keyword represents the quadrant number, the first two numbers after the first character are the row numbers in this quadrant, and the next two numbers the column numbers. **13. X\_VALUE and Y\_VALUE** Their value field should contain two real floating-point numbers, which give X and Y coordinates of the target on the focal plane. **14. OBJNAME** The value field should contain a character string, giving the name ID of the object that is determined by the RA, DEC and HTM method. **15. OBJTYPE** The value field should contain a character string giving the class of objects in input catalogs. **16. OBJSOURC** The value field should contain a character string which shows the name of the organization or person who submitted the input catalog. **17. FIBERTYP** The value field should contain a character string, giving the type of fiber assigned to this target. This keyword has five values, i.e., Obj, Sky, F-std, Unused, PosErr and Dead. Obj means the fiber is assigned to an object, including a star, galaxy and so on. Sky indicates the fiber is allocated to take skylight. F-std shows the fiber is used to take the light of a flux calibration standard star. Unused, PosErr and Dead mean the fiber is not used, goes to a wrong position, or does not work respectively. **18. MAGTYPE** The value field should contain a character string, which shows the magnitude type of a target. **19. MAG1, MAG2, MAG3, MAG4, MAG5, MAG6 and MAG7** The value field should contain a real floating-point numbers between 0 and 100, giving the associated magnitudes of MAGTYPE keyword. For example, if MAGTYPE keyword is 'ugrizjh', the MAG1, MAG2, MAG3, MAG4, MAG5, MAG6 and MAG7 keywords provide the magnitudes of u, g, r, i, z, j and h filter respectively. **20. OBS\_TYPE** The value field should contain a character string giving the type of observation targets, which include object, flat, bias and arc lamp. **21. OBSCOMM** The value field should contain a character string constant representing the purpose of the observation, which includes observations used for science researches and different kinds of tests. **22. RADECSYS** The value field should contain a character string giving the equatorial coordinate system based on the J2000 position. **23. EQUINOX** The value field should contain a real floating-point number giving the standard epoch used at present.

**Table A.5** Spectrograph Parameter Keywords

Keywords	Value	Comment
SPID	1	Spectrograph ID
SPRA	266.640	Average RA of this spectrograph [ $^{\circ}$ ]
SPDEC	23.7144	Average DEC of this spectrograph [ $^{\circ}$ ]
SLIT_MOD	'x2/3 '	Slit mode, x1 ,x2/3 or x1/2

Notes: **1. SPID** The value field should contain a non-integer number between 1 and 16, which provides the spectrograph ID. **2. SPRA and SPDEC** The value field of these two keywords should contain two real floating-point numbers, which are the averages of RA and DEC for all objects in each spectrograph. **3. SLIT\_MOD** The value field should contain a character string giving the mode of the slit, which includes x1, x2/3 and x1/2. At present, only mode x2/3 is available, which corresponds to a spectral resolution of 1800.

**Table A.6** Weather Condition Keywords

Keywords	Value	Comment
TEMPCCDB	-101.60	The temperature of the blue CCD [ $^{\circ}$ C]
TEMPCCDR	-103.90	The temperature of the red CCD [ $^{\circ}$ C]
SEEING	3.60	Seeing during exposure ["]
MOONPHA	24.51	Moon phase for a 29.53 day period [d]
TEMP_AIR	20.35	Temperature outside dome [ $^{\circ}$ C]
DEWPOINT	3.11	[ $^{\circ}$ C]
DUST	' '	Reservation
HUMIDITY	0.32	
WINDD	87.00	Wind direction [ $^{\circ}$ ]
WINDS	2.30	Wind speed [ $\text{m s}^{-1}$ ]

Notes: **1. TEMPCCDB** The value field should contain a real floating-point number, which provides the temperature of the blue CCD. The unit 'C degree' represents degree Centigrade. **2. TEMPCCDR** The value field should contain a real floating-point number, which provides the temperature of the red CCD. The unit 'C degree' represents degree Centigrade. **3. SEEING** The value field should contain a real floating-point number giving the seeing during an exposure, which is calculated by manually measuring the full width at half maximum of the guide star image. **4. MOONPHA** The value field should contain a real floating-point number giving the moon phase. **5. TEMP\_AIR** The value field should contain a real floating-point number giving the temperature outside the dome, which is measured by and automatic weather instrument. The unit 'C degree' represents degree Centigrade. **6. DEWPOINT** The value field should contain a real floating-point number giving the dew-point temperature, which is also measured by the automatic weather instrument. The unit 'C degree' represents degree centigrade. **7. DUST** The value of this keyword is temporarily empty at present, because the dust measuring instrument is now being debugged, and we will write this parameter into the FITS header when the problems are resolved. **8. HUMIDITY** The value field should contain a real floating-point number between 0 and 1, which gives humidity in the air. **9. WINDD** The value field should contain a real floating-point number which records the instantaneous wind direction when the exposure starts, and the direction of north is defined as 0 degrees. **10. WINDS** The value field should contain a real floating-point number which records the instantaneous wind speed when the exposure starts, and the wind direction and speed are also measured by an automatic weather instrument.

**Table A.7** Data Reduction Parameter Keywords (1)

Keywords	Value	Comment
LAMPLIST	'lamphgcdne.dat'	Arc lamp emission line list
SKYLIST	'skylines.dat'	Sky emission line list
EXPID01	'01b-20130604001350-4-81283693'	ID string for exposure 1
EXPID02	'01b-20130604005112-5-81283731'	ID string for exposure 2
EXPID03	'01r-20130604001348-4-81283693'	ID string for exposure 3
EXPID04	'01r-20130604005109-5-81283731'	ID string for exposure 4
NEXP	2	Number of valid exposures
NEXP_B	2	Number of valid blue exposures
NEXP_R	2	Number of valid red exposures
EXPT_B	3600.00 [s]	Blue exposure duration time
EXPT_R	3600.00 [s]	Red exposure duration time
EXPTIME	3600.00 [s]	Minimum exposure time for all cameras
BESTEXP	81283693	MJM of the best exposure
SCAMEAN	2.11 [ADU]	Mean level of scattered light
EXTRACT	'aperture'	Extraction method
SFLATTEN	T	Super flat has been applied
PCASKYSB	T	PCA sky-subtraction has been applied
NSKIES	22	Sky fiber number
SKYCHI2	1.91072039307	Mean $\chi^2$ of sky-subtraction
SCHI2MIN	1.51493247984	Minimum $\chi^2$ of sky-subtraction
SCHI2MAX	2.31929337901	Maximum $\chi^2$ of sky-subtraction

Notes: **LAMPLIST** The value field should contain a character string giving the file name of the arc lamp emission line list, which is used in the process of wavelength calibration. **SKYLIST** The value field should contain a character string giving the file name of the sky emission line list, which is used in the process of sky subtraction. **EXPID01**, **EXPID02**, **EXPID03**, **EXPID04** The value field of these four keywords contains four character strings giving the ID strings of the first, second, third and fourth exposures respectively, which are analogous to the file names of LAMOST raw data for the four exposures. It should be noted that, if the exposure times of a target are  $n$ , there should be  $n$  associated keywords here, which are EXPID01, EXPID02, ..., EXPID $n$  respectively. **NEXP**, **NEXP\_B**, **NEXP\_R** The value field of these three keywords should contain three non-negative integers, which provide numbers of exposures, and numbers of valid blue and red exposures respectively. **EXPT\_B**, **EXPT\_R** The value fields of these two keywords should contain two real floating-point numbers, which give the exposure durations of the blue and red CCD. **EXPTIME** The value field should contain a real floating-point, which gives the minimum of blue and red total exposures times. **BESTEXP** The value field should contain an integer, which gives the MJM of an exposure with the maximum signal and noise ratio in  $n$  exposures. **SCAMEAN** The value field should contain a real floating-point value giving the mean level of scattered light, which is the average flux of regions where there is no fiber and is at the left and right edge of a two dimensional spectral image. **EXTRACT** The value field should contain a character string, which indicates the method of spectral extraction. In the LAMOST spectral reduction pipeline, only the aperture method is applied to spectral extraction. **SFLATTEN** The value of this keyword should be Boolean, which represents whether or not use the super flat is used. In the LAMOST spectral reduction pipeline, the super flat is used to make the fiber-to-fiber relative efficiency around 1, and it can be estimated through spline fitting the flat flux of all fibers in a spectrograph. **PCASKYSB** The value of this keyword should be Boolean, which represents whether or not use the PCA method is used to subtract sky light. In the LAMOST spectral reduction pipeline, the PCA method is used to subtract sky light in the wavelength range larger than 7200 Å. **NSKIES** The value field should contain an integer, which shows the number of the sky fiber in a spectrograph. **SKYCHI2** The value field should contain a real floating-point, which gives the mean chi square of sky-subtraction. In the process of LAMOST spectral reduction, a super sky is obtained by spline fitting  $m$  sky spectra. Thus, the chi square between the super sky and each sky spectrum in an exposure, and the average chi square of  $m$  sky spectra can also be obtained. Assuming there are  $n$  exposures, there will be  $2n$  average chi squares value because of  $n$  blue spectra and  $n$  red spectra, and this keyword will be evaluated by calculating the mean value of these  $2n$  average chi square values. **SCHI2MIN** The value field should contain a real floating-point value, which gives the minimum chi square of sky-subtraction. As mentioned above, there will be  $2n$  average chi square values assuming  $n$  exposures, and this keyword will be the maximum of these chi squares. **SCHI2MAX** The value field should contain a real floating-point, which gives the maximum chi square of sky-subtraction. As mentioned above, there will be  $2n$  average chi square values assuming  $n$  exposures, and this keyword will be the minimum of these chi square values.

**Table A.7** Data Reduction Parameter Keywords (2)

Keywords	Value	Comment
NSTD	2	Number of (good) standard stars
FSTAR	'149-182 '	Fiber ID of flux standard stars
FCBY	'catalog '	Standard stars origin (auto, manual or catalog)
HELIO	T	Heliocentric correction
HELIO_RV	-3.99846411542	[km s <sup>-1</sup> ] Heliocentric correction
VACUUM	T	Wavelengths are in a vacuum
NWORDER	2	Number of linear-log10 coefficients
WFITTYPE	'LOG-LINEAR'	Linear-log10 dispersion
COEFF0	3.56820	Central wavelength (log10) of first pixel
COEFF1	0.000100000	log10 dispersion per pixel
WAT0_001	'system=linear'	
WAT1_001	wtype=linear label=Wavelength units=Angstroms	
CRVAL1	3.56820	Central wavelength (log10) of first pixel
CD1_1	0.000100000	log10 dispersion per pixel
CRPIX1	1	Starting pixel (1-indexed)
CTYPE1	'LINEAR'	
DC-FLAG	1	Log-linear flag

Notes: **NSTD** The value field should contain a non-negative integer, which shows the number of flux standard stars with good spectra quality. **FSTAR** The value field should contain a character string giving the fiber identity numbers of flux standard stars, which are separated by the symbol '-'. **FCBY** The value field should contain a character string giving the selection methods of flux standard stars, which include auto, manual and catalog. Auto indicates the standard stars are selected by the LAMOST reduction pipeline, manual means they are picked out by experienced staffs, and catalog indicates the standard stars are provided by the input catalog. **HELIO** The value of this keyword should be Boolean, which indicates whether or not to perform the heliocentric correction. **HELIO\_RV** The value field should contain a real floating-point number, which gives the radial velocity used to carry out the heliocentric correction. **VACUUM** The value of this keyword should be Boolean, which denotes whether or not the LAMOST spectrum is converted to the vacuum wavelength. **NWORDER** The value of this keyword should contain an integer, which gives the number of linear-log10 coefficients. **WFITTYPE** The value field should contain a character string giving the linear-log10 dispersion. **COEFF0** The value field should contain a real floating-point number, which provides the central wavelength (log10) of the first pixel. **COEFF1** The value field should contain a real floating-point number giving the log10 dispersion per pixel. **WAT0\_001** The value field contains a character string. **WAT1\_001** The value field contains a character string. **CRVAL1** The value field should contain a real floating-point number, which gives the coordinate of the reference pixel provided by the CRPIX1 keyword. **CD1\_1** The value field should contain a real floating-point number giving the dispersion of per pixel. **CRPIX1** The value of this keyword should contain an integer, which sets the reference pixel location on pixel axis. **CTYPE1** The value field should contain a character string, which will have the value 'LINEAR' that defines the wavelength axes to be linear. **DC-FLAG** The value of this keyword should be Boolean, where a value of 0 defines a linear sampling of the dispersion and a value of 1 defines a logarithmic sampling of the dispersion.

**Table A.8** Spectral Analysis Result Keywords

Keywords	Value	Comment
VERSPIPE	'v2.6.4'	Version of Pipeline
CLASS	'STAR'	Class of object
SUBCLASS	'F5'	Subclass of object
Z	-0.000314280000000	Redshift of object
Z_ERR	','	Redshift error of object
SN_U	1.26	SNR of u filter
SN_G	4.68	SNR of g filter
SN_R	9.62	SNR of r filter
SN_I	11.24	SNR of i filter
SN_Z	7.44	SNR of z filter

Notes: **VERSPIPE** The value field should contain a character string constant, which provides the version of the LAMOST pipeline that is used in spectral processing and analysis. In this data release, the value of **VERSPIPE** is 'v2.6.4'. It should be noted that, 'v2.6' is the version of the spectral reduction pipeline, 'v4' is the version of the spectral analysis pipeline, and 'v2.6.4' combines these two versions together.

**CLASS** The value field should contain a character string that provides the classification result determined by the LAMOST spectral analysis pipeline, which includes 'STAR', 'GALAXY', 'QSO' or 'Unknown'.

**SUBCLASS** The value field should contain a character string, which gives sub-classification result for stars. This keyword provides a more detailed spectral type for late A, F, G, K and M stars, and spectral and photometric type for A type stars. For galaxies, quasars or unknown type objects, this keyword is set to 'Non' now.

**Z** The value field should contain a real floating-point number providing the redshift of the target, which is mainly determined by the LAMOST spectral analysis pipeline. For the case that redshift is unable to calculate by the pipeline, it will be manually determined through measuring the shifts in some spectral line centers. If the quality of a spectrum is poor, or it is classified as 'Unknown', its redshift is artificially set to -9999.0.

**Z\_ERR** The value field should contain a real floating-point number, which gives the redshift error of a target. At present, the values of **Z\_ERR** keywords are not yet published, and we will release redshift errors at the next data release in September of 2015.

**SN\_U, SN\_G, SN\_R, SN\_I and SN\_Z** Keywords The value fields of these five keywords should contain five real floating-point numbers, which give the signal to noise ratio (SNR) of u, g, r, i and z bands. Using the center wavelength and bandwidth, we can obtain the wavelength range for each SDSS band, and then the SNR in each band is the median value at each pixel in this band.

**Table A.9** The significance of six bits of 'Andmask' and 'Ormask'

Bit	Keyword	Comments
1	BADCCD	bad pixel on CCD
2	BADPROFILE	bad profile extraction
3	NOSKY no	sky information at this wavelength
4	BRIGHTSKY	sky level too high
5	BADCENTER	fiber trace of the CCD
6	NODATA	no good data

Notes: (1) The 'andmask' information is a decimal integer determined by a six-bit binary number, which represents six situations respectively that are listed in this table. The associated bit of 'andmask' will be set to 1, if the case always appears in each exposure. Like 'andmask', 'ormask' information is also a decimal integer determined by a six-bit binary number. The difference is that each bit of 'ormask' will be set to 1 if the related case happens in any exposure. (2) For a spectrum, if you want to check which case occurred during the reduction process, you should convert the decimal 'Andmask' and 'Ormask' to the six-bit binary number, and a case must have happened in all exposures (at least one exposure) if the associated bit is 1 in binary 'Andmask' ('Ormask').

**Table A.10** LAMOST General Catalog

Field (unit)	Type	Comment
designation	varchar	Target Designation
specObjId	long integer	Unique Spectral ID
obsdate	char	Target Observation Date
mjd	char	Local Modified Julian Day
planid	char	Plan ID
spid	integer	Spectrograph ID
fiberid	integer	Fiber ID
objra (°)	float	Right Ascension
objdec (°)	float	Declination
dataversion	char	Data Version
snru	float	Signal Noise Ratio of u filter
snrg	float	Signal Noise Ratio of g filter
snrr	float	Signal Noise Ratio of r filter
snri	float	Signal Noise Ratio of i filter
snrz	float	Signal Noise Ratio of z filter
objtype	varchar	Object Type
class	varchar	Spectral Type
subclass	varchar	Stellar Sub-Class
magtype	varchar	Target Magnitude Type
mag1 (mag)	float	Associated Magnitude 1
mag2 (mag)	float	Associated Magnitude 2
mag3 (mag)	float	Associated Magnitude 3
mag4 (mag)	float	Associated Magnitude 4
mag5 (mag)	float	Associated Magnitude 5
mag6 (mag)	float	Associated Magnitude 6
mag7 (mag)	float	Associated Magnitude 7
tsource	varchar	Organization or person who submitted the input to the catalog
tfrom	varchar	Input catalog submitted by an organization or a person
fibertype	varchar	Fiber Type of target [Obj, Sky, F-std, Unused, PosErr, Dead]
z	float	Redshift
rv (km s <sup>-1</sup> )	float	Heliocentric Radial Velocity

**Table A.11** A Type Star Catalog

Field (unit)	Type	Comment
designation	varchar	Target Designation
specObjId	long integer	Unique Spectral ID
obsdate	char	Target Observation Date
mjd	char	Local Modified Julian Day
planid	char	Plan ID
spid	integer	Spectrograph ID
fiberid	integer	Fiber ID
objra (°)	float	Right Ascension
objdec (°)	float	Declination
dataversion	char	Data Version
snru	float	Signal Noise Ratio of u filter
snrg	float	Signal Noise Ratio of g filter
snrr	float	Signal Noise Ratio of r filter
snri	float	Signal Noise Ratio of i filter
snrz	float	Signal Noise Ratio of z filter
objtype	varchar	Object Type
class	varchar	Stellar Class
subclass	varchar	Stellar Sub-Class
magtype	varchar	Target Magnitude Type
mag1 (mag)	float	Associated Magnitude 1
mag2 (mag)	float	Associated Magnitude 2
mag3 (mag)	float	Associated Magnitude 3
mag4 (mag)	float	Associated Magnitude 4
mag5 (mag)	float	Associated Magnitude 5
mag6 (mag)	float	Associated Magnitude 6
mag7 (mag)	float	Associated Magnitude 7
tsource	varchar	Organization or person who submitted the input to the catalog
tfrom	varchar	Input catalog submitted by an organization or a person
fibertype	varchar	Fiber Type of target [Obj, Sky, F-std, Unused, PosErr, Dead]
rv (km s <sup>-1</sup> )	float	Heliocentric Radial Velocity
Halpha_Indice (Å)	float	Line Index of Halpha line
Hbeta_Indice (Å)	float	Line Index of Hbeta line
Hgamma_Indice (Å)	float	Line Index of Hgamma line
Hdelta_Indice (Å)	float	Line Index of Hdelta line
Halpha_D0.2 (Å)	float	Width at 20% below the local continuum of H <sub>α</sub> line
Hbeta_D0.2 (Å)	float	Width at 20% below the local continuum of H <sub>β</sub> line
Hgamma_D0.2 (Å)	float	Width at 20% below the local continuum of H <sub>γ</sub> line
Hdelta_D0.2 (Å)	float	Width at 20% below the local continuum of H <sub>δ</sub> line



**Table A.12** A, F, G and K Type Star Catalog

Field (unit)	Type	Comment
designation	varchar	Target Designation
specObjId	long integer	Unique Spectral ID
obsdate	char	Target Observation Date
mjd	char	Local Modified Julian Day
planid	char	Plan ID
spid	integer	Spectrograph ID
fiberid	integer	Fiber ID
objra (°)	float	Right Ascension
objdec (°)	float	Declination
dataversion	char	Data Version
snru	float	Signal Noise Ratio of u filter
snrg	float	Signal Noise Ratio of g filter
snrr	float	Signal Noise Ratio of r filter
snri	float	Signal Noise Ratio of i filter
snrz	float	Signal Noise Ratio of z filter
objtype	varchar	Object Type
class	varchar	Stellar Class
subclass	varchar	Stellar Sub-Class
magtype	varchar	Target Magnitude Type
mag1 (mag)	float	Associated Magnitude 1
mag2 (mag)	float	Associated Magnitude 2
mag3 (mag)	float	Associated Magnitude 3
mag4 (mag)	float	Associated Magnitude 4
mag5 (mag)	float	Associated Magnitude 5
mag6 (mag)	float	Associated Magnitude 6
mag7 (mag)	float	Associated Magnitude 7
tsource	varchar	Organization or person who submitted the input to the catalog
tfrom	varchar	Input catalog submitted by an organization or a person
fibertype	varchar	Fiber Type of target [Obj, Sky, F-std, Unused, PosErr, Dead]
rv (km s <sup>-1</sup> )	float	Heliocentric Radial Velocity
teff (K)	float	Effective Temperature
log (dex)	float	Surface Gravity
feh (dex)	float	Metallicity

**Table A.13** M-type Star Catalog

Field (unit)	Type	Comment
designation	varchar	Target Designation
specObjId	long integer	Unique Spectra ID
obsdate	float	Target Observation Date
mjd	char	Local Modified Julian Day
planid	char	Plan ID
spid	integer	Spectrograph ID
fiberid	integer	Fiber ID
objra (°)	float	Right Ascension
objdec (°)	float	Declination
dataversion	char	Data Version
snru	float	Signal Noise Ratio of u filter
snrg	float	Signal Noise Ratio of g filter
snrr	float	Signal Noise Ratio of r filter
snri	float	Signal Noise Ratio of i filter
snrz	float	Signal Noise Ratio of z filter
objtype	varchar	Object Type
class	varchar	Stellar Class
subclass	varchar	Stellar Sub-Class
magtype	varchar	Target Magnitude Type
mag1 (mag)	float	Associated Magnitude 1
mag2 (mag)	float	Associated Magnitude 2
mag3 (mag)	float	Associated Magnitude 3
mag4 (mag)	float	Associated Magnitude 4
mag5 (mag)	float	Associated Magnitude 5
mag6 (mag)	float	Associated Magnitude 6
mag7 (mag)	float	Associated Magnitude 7
tsource	varchar	Organization or person who submitted the input to the catalog
tfrom	varchar	Input catalog submitted by an organization or a person
fibertype	varchar	Fiber Type of target [Obj, Sky, F-std, Unused, PosErr, Dead]
rv (km s <sup>-1</sup> )	float	Heliocentric Radial Velocity
magact	integer	Magnetic Activity