Research in Astron. Astrophys. **2012** Vol. **12** No. **9**, 1243–1246
http://www.raa-journal.org *http://www.iop.org/journals/raa*

Research in Astronomy and Astrophysics

Data release of the LAMOST pilot survey

A-Li Luo[1], Hao-Tong Zhang[1], Yong-Heng Zhao[1], Gang Zhao[1], Xiang-Qun Cui[2], Guo-Ping Li[2], Yao-Quan Chu[3], Jian-Rong Shi[1], Gang Wang[1], Jian-Nan Zhang[1], Zhong-Rui Bai[1], Xiao-Yan Chen[1], Feng-Fei Wang[1], Yan-Xin Guo[1], Jian-Jun Chen[1], Bing Du[1], Xiao Kong[1], Ya-Juan Lei[1], Yin-Bi Li[1], Yi-Han Song[1], Yue Wu[1], Yan-Xia Zhang[1], Xin-Lin Zhou[1], Fang Zuo[1], Peng Du[1], Lin He[1], Wen Hou[1], Yi-Qiao Dong[1], Jian Li[1], Guang-Wei Li[1], Shuang Li[1], Jing Song[1], Yuan Tian[1], Meng-Xin Wang[1], Ke-Fei Wu[1], Hui-Qin Yang[1], Hai-Long Yuan[1], Shu-Yun Cao[1], Hai-Yuan Chen[2], Kun-Xin Chen[2], Ying Chen[1], Jia-Ru Chu[3], Lei Feng[1], Xue-Fei Gong[2], Bo-Zhong Gu[2], Yong-Hui Hou[2], Zhi-Ying Huo[1], Hong-Zhuan Hu[3], Ning-Sheng Hu[2], Zhong-Wen Hu[2], Lei Jia[1], Fang-Hua Jiang[2], Xiang Jiang[2], Zi-Bo Jiang[2], Ge Jin[3], Ai-Hua Li[2], Qi Li[1], Xin-Nan Li[2], Yan Li[4], Ye-Ping Li[2], Gen-Rong Liu[2], Guan-Qun Liu[2], Zhi-Gang Liu[3], Qi-Shuai Lu[2], Wen-Zhi Lu[2], Yu Luo[1], Yin-Dun Mao[4], Li Men[1], Ji-Jun Ni[2], Yong-Jun Qi[2], Zhao-Xiang Qi[4], Huo-Ming Shi[1], Ding-Qiang Su[2], Shi-Wei Sun[1], Hong-Jun Su[1], Zheng-Hong Tang[4], Qing-Sheng Tao[2], Liang-Ping Tu[1], Da-Qing Wang[1], Dan Wang[1], Guo-Min Wang[2], Hai Wang[2], Jia-Ning Wang[2], Jian Wang[3], Jian-Ling Wang[1], Jian-Ping Wang[3], Lei Wang[2], Shou-Guan Wang[1], Shu-Qing Wang[1], Ya-Nan Wang[2], You Wang[2], Yue-Fei Wang[2], Ming-Zhi Wei[1], Xiang-Xiang Xue[1], Xiao-Zheng Xing[3], Ling-Zhe Xu[2], Xin-Qi Xu[2], Yan Xu[1], De-Hua Yang[2], Shi-Hai Yang[2], Zheng-Qiu Yao[2], Yong Yu[4], Hui Yuan[1], Chao Zhai[3], En-Peng Zhang[1], Jing Zhang[5], Li-Ping Zhang[2], Wei Zhang[1], Yong Zhang[2], Zhen-Chao Zhang[2], Ming Zhao[4], Fang Zhou[2], Yong-Tian Zhu[2], Jie Zhu[2] and Si-Cheng Zou[1]

[1] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China; *lal@lamost.org*
[2] Nanjing Institute of Astronomical Optics and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Nanjing 210042, China
[3] University of Science and Technology of China, Hefei 230026, China
[4] Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China
[5] Institute of Architecture Design and Research, Chinese Academy of Sciences, Beijing 100190, China

Received 2012 August 14; accepted 2012 August 17

Abstract This paper describes the data release of the LAMOST pilot survey, which includes data reduction, calibration, spectral analysis, data products and data access. The accuracy of the released data and the information about the FITS headers of spectra are also introduced. The released data set includes 319 000 spectra and a catalog of these objects.

Key words: techniques: spectroscopic survey — data reduction — data release

1 INTRODUCTION

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST, also called the Guo Shou Jing Telescope) has the capability of taking 4000 spectra with resolution ($R = 1800$) simultaneously in a single exposure (Cui et al. 2012). This project plans to start a scientific spectroscopic survey of over 10 million objects in September 2012. This effort will provide useful data for the study of stellar astrophysics and the structure of the Galaxy, and extragalactic astrophysics and cosmology (Zhao et al. 2012). In order to check the real instrumental performance and assess the feasibility of the science goals, the science working groups of LAMOST came up with the pilot survey of LAMOST, which was conducted before the regular spectral survey (Deng et al. 2012). The pilot survey was launched on 2011 Oct 24, and ended in June 2012, which includes nine full Moon cycles and basically covers all good observing seasons in a year at the site (Yao et al. 2012). Each Moon cycle of the pilot survey was divided into three parts, i.e. dark nights, bright nights and grey nights. In principle, dark nights should be used for faint objects with magnitude in the r (or V) band in the range [14.5 $\sim$ 19.5], and bright nights should be for sources with magnitude in the r band in [11.5 $\sim$ 16.5]. The exposure time for dark nights was 3×30 minutes, and that for bright nights was of order 3×10 minutes (depending on the distribution of the brightness of the sources). The pilot survey obtained spectra of stars in the Milky Way, which included fainter objects on dark nights (Yang et al. 2012; Carlin et al. 2012), brighter objects on bright night (Zhang et al. 2012), objects in the disk of the Galaxy with low latitude (Chen et al. 2012) and objects in the region of the Galactic Anti-Center. It also targets extragalactic objects located in two regions, i.e., the South Galactic Cap and the North Galactic Cap. The footprint of LAMOST is shown in Figure 1.

Raw CCD data were reduced and analyzed by the LAMOST data reduction system, which includes the 2D pipeline, the 1D pipeline, the stellar parameter pipeline, and the data mining pipeline. In Section 2, the 2D and 1D pipelines are outlined because they are related to this data release. Calibration is also described in Section 2. About one million spectra were taken during the pilot survey, but because the system was still experimental and being tuned during this period, about half of the data are of very low quality. In this early data release, we have removed the low quality data which are not useful to the public. The criteria of the data cut, format of the spectra, and FITS information about each spectrum are introduced in Section 3. Data access is presented in Section 4.

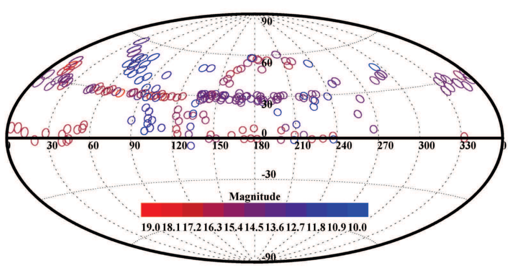

Fig. 1 The footprint of the LAMOST pilot survey.

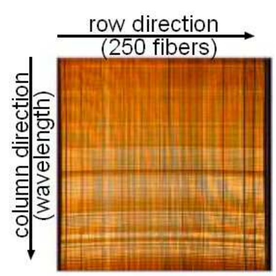

Fig. 2 Raw data on the CCD.

2 DATA REDUCTION AND CALIBRATION

A spectrum of a light beam from fibers (arranged along a 144 mm height slit from each spectrograph) is imaged onto 32 cameras using 4096×4096 EEV CCD chips, each with a 12 μm square pixel size (Wei & Stover 1996) (Fig. 2). The LAMOST 2D pipeline separately reduces the data from each CCD chip and of each exposure, and then it combines the results from different exposures. A lot of procedures from the spectro2d pipeline of SDSS are used (Stoughton et al. 2012). Bias and dark frames are subtracted from each raw image. Then the flat-field spectra are traced for each fiber, and the centroid of the row position of each fiber is fitted by a polynomial. The profile of one fiber in a row direction is assumed to be a Gaussian-like profile, which is characterized by $y = a\exp[(x-b)^d/dc^d]$. There are 250 profiles in each image, and the coefficients are calculated and used for other exposures. The arc lamp spectra are then extracted, and centroids of the lines are measured, to which we fit a Legendre polynomial as a function between wavelengths and pixels, to fifth-order in blue branch and sixth-order in red. The flat-fields are then extracted, wavelength-calibrated, normalized, and then combined to form a superflat for each image. For each fiber, the extracted spectrum is divided by the superflat to form the fiber flat. Variations between fibers are thus removed with the flat field. For each science image, the flux of each fiber is extracted by the coefficients from flat field spectra. The flux values are then flat fielded by dividing by the fiber flats. A vacuum wavelength scale is applied for wavelength calibration, slightly adjusting to match the known positions of certain sky lines, and correcting to the heliocentric frame. The wavelength calibrations are accurate to 10 km s^{-1} or better. More than 20 fibers are used to model the background of a specific region of the sky, called a supersky, and for each fiber in that region, the supersky background is subtracted. Telluric absorption in four wavelength regions is removed. Next, the spectra are flux-calibrated by matching the selected standard flux stars with their templates for types A or F. Finally, for each object in the individual exposures, the spectra from the red and blue spectrograph are combined by stacking the points with corresponding wavelengths using a B-spline function, with inverse-variance weighting. Outliers due to cosmic rays are rejected and masked, then errors in the fluxes estimated. The combined spectra are resampled in constant-velocity pixels, with a pixel scale of 69 km s^{-1}, which means the wavelength difference between two adjacent points is shown as $\Delta\log(\lambda) = 0.0001$.

The LAMOST 1D pipeline is designed to classify spectra from the output of the 2D pipeline and to measure their redshift (or radial velocity). A reliable technique, which was developed by Glazebrook et al. (1998), for automatically determining the redshift of a galaxy, is becoming increasingly important. This method generalizes the cross-correlation approach by replacing the individual templates with a simultaneous linear combination of orthogonal templates. This effectively eliminates the mismatch between templates and data and has the potential to provide the possibility of good error estimation. The method, which is called "PCAZ," is based upon the use of principal component analysis to make the general linear problem amenable to efficient computation. This method proves to be more robust than independent cross-correlation in the low S/N regime and has greater potential for very high success rates in upcoming surveys. It is also appropriate for QSOs and stellar spectra. Our 1D pipeline is based on the specBS pipeline which is used for analysis of SDSS spectra. This analysis code carries out χ^2 fits of the spectra to templates in wavelength space (in the spirit of Glazebrook et al. 1998), fitting spectra with linear combinations of eigen-spectra and low-order polynomials. A set of SDSS spectra with carefully chosen zero points of a variety of spectral types, which also contain galaxies, QSOs and stars (categorized in terms of subtypes), is used to construct our templates.

The accuracy of our method, like any other analysis, is affected by the SNR, wavelength calibration, and flux calibration. Relatively high accuracy redshift and correct spectral type are obtained from good quality spectra. For spectra with low quality, our template matching method cannot guarantee the best classification or redshift estimate, but we can give an indicator which presents the confidence of the redshift measurement. Meanwhile, the SNR is a direct index of spectral quality. The Pilot Survey data release needs a strategy for identifying the final spectral type and final redshift of LAMOST spectra for astronomical users. For the spectra with S/N > 5, the 1D pipeline gives spectral type and redshift. For LAMOST spectra with low confidence in measurement or low SNR, human checking is also applied to data quality. A flag named "specflag" can be used to explain the status of the final redshift. After comparing the classification results of the 1D pipeline with spectral types found from SDSS DR8, our pipeline shows a good performance for the high quality part of the spectra, and a correct classification rate of 96% is obtained from spectra with a threshold of 10% for ZCONF and 10 for the r-band SNR. For correctly classified spectra, we compare the redshift estimate with SDSS. A precision in redshift of 0.0001 for galaxies and 0.01 for QSOs, as well as precision of 13 km s^{-1} for stellar spectra, shows the capability of the code.

3 DATA PRODUCTS FROM THE PILOT SURVEY

LAMOST 1D FITS files are named in the form spec-MMMMM-YYYY_spXX-fff.fit, where MMMMM is the modified Julian date (mjd), YYYY is the plan identity number (id), XX is the spectrograph id number, and fff is the fiber id number. LAMOST 1D FITS files are approximately flux calibrated, wavelength calibrated, sky-subtracted spectra, with errors and mask arrays, and with information gained from the 1D pipeline such as redshift and identification. For a single object, each LAMOST 1D FITS file includes the classification and redshift determination, as well as the spectrum, summing over all of its exposures of a given mapped plate. Each file contains the following HDUs: Primary HDU image: spectrum, inverse variance ($1/\sigma^2$) for the flux, continuum-subtracted spectrum, and-mask, and or-mask. The spectra are binned on a log-linear scale and are given in vacuum wavelengths in the heliocentric frame, with flux density given in relative units. The first row is the spectrum, the second row is the inverse variance of the spectrum (standard deviation, in the same units as the spectrum), the third row is the continuum subtracted spectrum, the forth row is the and-mask array (all of its exposures are bad pixels), and the fifth row is the or-mask array (at least one of its exposures is a bad pixel). HDU1 records 1D results.

To publish relatively high quality spectra from all the data, the first selection criterion is the SNR. After cutting data points with SNR$\leq$10 (in the g and r bands), 493 000 spectra are kept. As is well known, the SNR of a spectrum is correlated with magnitude if exposure time or other observing conditions (for example, seeing) are given. We can get the theoretical relation between SNR and magnitude, and compare it with a fitting relationship from each of the 4000 spectra taken at the same time. Then, the relationship is fixed, and those spectra with SNR deviating far from the relationship should be cut. After this cutting, 476 000 spectra are kept. Another point we should consider is about the pointing direction of the fibers, that is, if a given fiber is pointing to the exact target we want to observe. We repeatedly calculated the precision of each fiber based on the SNR - magnitude relation, and discovered that ‘good’ fibers have a repeated precision higher than 95%. We cut those spectra observed to be ‘bad’ fibers, and 319 000 spectra are left as the released data.

4 DATA USE AND STANDARD CITATION TO A LAMOST PAPER

Any users can access the LAMOST pilot data through LAMOST’s official web site: *http://www.lamost.org/*. One can download FITS files of spectra. Data are indexed by a catalog which is on the page for DATA ACCESS of the LAMOST website. The data can also be linked from the RAA website: *http://www.raa-journal.org/*. The following two papers regarding LAMOST’s technical specifications and data release would be appropriate references for your study. For the Data Release of the LAMOST Pilot survey, please refer to Luo et al. (2012, RAA, 12, 1243) in your research paper. For an overview of the LAMOST Telescope, fiber positioning, and spectrographs, please refer to Cui et al. (2012 RAA, 12, 1197) in your research paper.

References

Carlin, J. L., Lépine, S., Newberg, H. J., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 755

Chen, L., Hou, J. L., Yu, J. C., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 805

Cui, X., Zhao, Y., Chu, Y., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 1197

Deng, L., Newberg, H. J., Liu, C., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 735

Glazebrook, K., Offer, A. R., & Deeley, K. 1998, ApJ, 492, 98

Stoughton, C., Lupton, R. H., et al. 2002, AJ, 123, 485

Wei, M., & Stover R. J. 1996, Solid State Sensor Arrays and CCD Cameras (San Jose), 226

Yang, F., Carlin, J. L., Liu, C., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 781

Yao, S., Liu, C., Zhang, H., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 772

Zhang, Y., Carlin, J. L., Yang, F., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 792

Zhao, G., Zhao, Y. H., Chu, Y. Q., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 723