RAA 2017 Vol. **17** No. **9**, 91 (14pp) doi: 10.1088/1674–4527/17/9/91

http://www.raa-journal.org *http://iopscience.iop.org/raa*

***R**esearch in*
***A**stronomy and*
***A**strophysics*

Sky subtraction for LAMOST

Zhong-Rui Bai[1,2], Hao-Tong Zhang[2], Hai-Long Yuan[2], Guang-Wei Li[2], Jian-Jun Chen[2], Ya-Juan Lei[2], Hui-Qin Yang[2], Yi-Qiao Dong[2], Gang Wang[2] and Yong-Heng Zhao[1,2]

[1] University of Chinese Academy of Sciences, Beijing 100049, China
[2] Key Lab for Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China; *htzhang@bao.ac.cn*

Received 2017 February 8; accepted 2017 April 25

Abstract Sky subtraction is a key technique in data reduction of multi-fiber spectra. Knowledge of characteristics related to the instrument is necessary to determine the method adopted in sky subtraction. In this study, we describe the sky subtraction method designed for the Large sky Area Multi-Object fiber Spectroscopic Telescope (LAMOST) survey. The method has been integrated into the LAMOST 2D Pipeline v2.6 and applied to data from LAMOST DR3 and later. For LAMOST, calibration using sky emission lines is used to alleviate the position-dependent (and thus time-dependent) $\sim 4\%$ fiber throughput uncertainty and small wavelength instability (0.1 Å) during observation. Sky subtraction using principal component analysis (PCA) further reduces 25% of the sky line residual from OH lines in the red part of LAMOST spectra after the master sky spectrum, which is derived from a B-spline fit of 20 sky fibers in each spectrograph. Using this approach, values are adjusted by a sky emission line and subtracted from each fiber. Further analysis shows that our wavelength calibration accuracy is about 4.5 km s^{-1}, and the averages of residuals after sky subtraction are about 3% for sky emission lines and 3% for the continuum region. The relative sky subtraction residuals vary with moonlight background brightness, and can reach as low as 1.5% for regions that have sky emission lines during a dark night. Tests on F stars with both similar sky emission line strength and similar object continuum intensity show that the sky emission line residual of LAMOST is smaller than that of the SDSS survey.

Key words: techniques: spectroscopic — methods: data analysis — instrumentation: spectrographs

1 INTRODUCTION

Multi-object spectroscopy with optical fibers, which has been a major advance for astronomical observation due to its ability to simultaneously observe many more objects than traditional long slit spectroscopy, has been routinely carried out over the last three decades. Unlike slit or multi-slit systems, the sky spectrum close to an object cannot be sampled adjacent to the object both on the CCD and in the focal plane in multi-object fiber spectroscopy. This aspect makes both the observation and the data reduction strategies of multi-fiber spectroscopy differ from those of slit spectroscopy. The standard procedure of sky subtraction applied in multi-object fiber spectroscopy involves using a subset of fibers (sky fibers) to measure the sky background simultaneously with object fibers (Wyse & Gilmore 1992; Watson et al. 1998). A master sky spectrum is constructed from the sky fibers and then subtracted from each object+sky spectrum.

In practice, accuracy for sky subtraction in the range 1%–2% is considered good (Elston & Barden 1989; Cuby & Mignoli 1994). Limitations come from various reasons, including focal-ratio degradation of the fibers, internal scattered light, variation of the sky, telecentricity effects (Wynne 1993), cross talk from adjacent fibers and poor determination of fiber transmittance (Elston & Barden 1989; Watson et al. 1998).

A number of astronomers have explored techniques to improve sky subtraction. Observational strategies such as beam-switching (Barden et al. 1993; Puech et al. 2014; Rodrigues et al. 2012) and nod-and-shuffle (N+S, Glazebrook & Bland-Hawthorn 2001; Sharp & Parkinson 2010) can help eliminate the throughput difference between fibers and obtain higher sky subtraction

accuracy, but the extra cost associated with exposure time or CCD space is unavoidable. For the standard observation mode, strong night sky emission lines are often used to calibrate the relative transmission of fibers to an accuracy of sky subtraction better than 2% (e.g. Lissandrini et al. 1994). Principal component analysis (PCA) is another well-established technique that has been applied in sky subtraction for fiber spectroscopy in the last 10 years after its first demonstration by Kurtz & Mink (2000). Wild & Hewett (2005) presented a technique to remove residual OH features based on applying PCA to the residuals of sky subtracted sky spectra in the Sloan Digital Sky Survey (SDSS) Data Release 2 (DR2) and achieved a dramatic improvement in the quality of a large fraction of SDSS spectra, particularly for fainter objects such as high-redshift quasars. Sharp & Parkinson (2010) demonstrated that PCA is more efficient than the N+S technique for observations in a sky limited regime with durations of 10-100 h. Soto et al. (2016) introduced ZAP, an approach to sky subtraction based on PCA, which is likely to be a useful tool for substantially improving sky subtraction accuracy.

In this paper, we describe the sky subtraction technique used for the Guo Shou Jing Telescope (a.k.a. LAMOST, Cui et al. 2012). The technique has been integrated into the LAMOST 2-dimensional (2D) Pipeline v2.6 and applied to LAMOST DR3 and later. The telescope and instrument characteristics involved in sky subtraction are introduced in Section 2. The sky subtraction methodology for LAMOST is presented in Section 3. The accuracy of sky subtraction is analyzed in Section 4. Some discussions and conclusions are provided in Sections 5 and 6, respectively.

2 LAMOST

2.1 LAMOST Instruments and Observation

LAMOST is a special Schmidt telescope which allows both a large aperture (effective aperture of 3.6 m–4.9 m) and a wide field of view (FoV) of 5° (Cui et al. 2012). 4000 optical fibers are accommodated on the focal plane, each of which is 320 microns in diameter, equivalent to 3.3 arcseconds in the sky. Each fiber is driven by a fiber positioning unit containing two stepper motors, by which all fibers can be positioned simultaneously in less than 10 minutes. The fibers are grouped into 16 spectrographs, in each of which the light beam is split into a blue arm (370–590 nm) and a red arm (570–900 nm) by a dichroic mirror, then registered by a 4k×4k CCD camera in each arm. The spectral resolution is about 2.5 Å in the blue arm and 4 Å in the red arm.

To optimize observing efficiency and mitigate fiber cross talk, targets in the LAMOST survey are grouped into bright, medium and faint cases according to their r-band magnitude. The on-site astronomer decides which one to execute based on the Moon phase and weather conditions. Fainter cases are always observed on darker nights with better weather conditions. Multiple exposures, usually three, are taken to obtain enough Signal to Noise Ratio (S/N) and to remove cosmic rays. The typical exposure times of one sub-exposure for bright, medium and faint cases are 600, 900 and 1800 seconds, respectively. Twilight flats are taken at the zenith both in the evening and morning for correcting the instrumental differences between fibers, and three frames, each from both a mercury-cadmium arc lamp for blue and a neon-argon arc lamp for red, are taken at the beginning, middle and end of an observational night, respectively.

The 5° diameter FoV, or 20 square degrees, is divided into 16 pieces, and each piece (250 fibers, 1.25 square degrees) is fed into one spectrograph, as shown in Figure 1. On dark nights, the night sky shows stable gradients on scales of degrees (Wyse & Gilmore 1992), so the spatial variation of sky inside one spectrograph is insignificant. 20 of 250 fibers, distributed homogeneously both on the sky and on the CCD, are dedicated to sampling the sky spectra. The traditional sky subtraction method is performed separately, spectrograph by spectrograph. A master sky spectrum is constructed from those sky fibers, as described in the following sections.

2.2 Dark Night Sky Spectrum at the LAMOST Site

A typical dark night sky spectrum observed by LAMOST is shown in Figure 2. Except for a few emission lines coming from artificial light pollution, most of the distinctive features in the night-sky spectrum, including the continuum, absorption lines and most of the emission lines, are due to natural processes.

The continuum of the night-sky is contributed by zodiacal light, starlight, extragalactic light, and reflected solar light (Benn & Ellison 1998). Airglow, emitted by various processes in atoms and molecules in the upper atmosphere, produces [OI] 5577 Å, 6300 Å, 6363 Å lines, the O band at 8600–8700 Å, NaD 5890–5896 Å, as well as the OH bands in red and IR, known as the Meinel bands (Meinel 1950). Light pollution mostly comes from street lamps including mercury lamps and sodium lamps. For LAMOST, mercury streetlights produce strong narrow lines at 3651 Å, 3663 Å, 4047 Å, 4358 Å, 5461 Å, 5770 Å and 5791 Å and sodium lamps contribute to NaD lines while other Na lines are relatively weak.

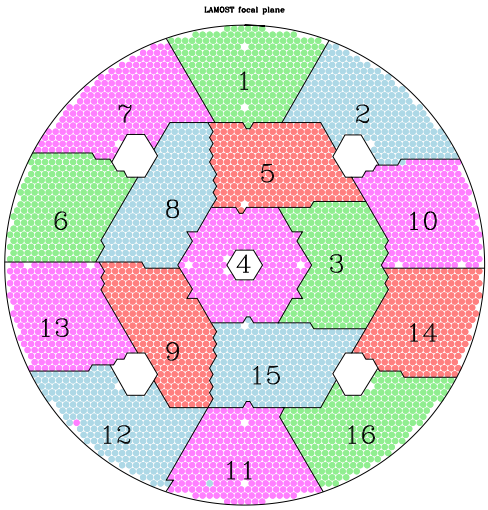

Fig. 1 LAMOST fiber division scheme. The fibers on the focal plane are divided to 16 regions, and each feeds into one spectrograph. Colored circles indicate the positions of the fibers and solid lines delineate the borders of spectrographs. The IDs of the spectrographs are marked in the corresponding regions.

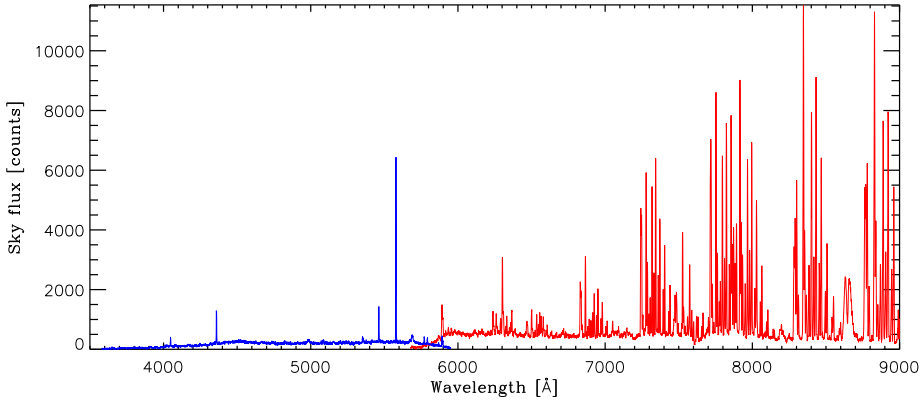

Fig. 2 Moonless night sky spectra taken by LAMOST on 2015 September 12. Spectra of the blue and red arms are indicated by colors. The x-axis is the calibrated wavelengths directly measured from CCD pixels.

2.3 LAMOST Instrumental Effects

2.3.1 Instrumental differences

To subtract the sky spectrum with sky fibers, accurate calibration of relative instrumental differences from the mirror of the telescope to the CCD pixel is very important. Generally, those differences include vignetting of the telescope; throughput difference between fibers which may be caused either by an intrinsic difference due to various reasons (e.g. fiber length, polishing of the end face, etc) or by misalignments between a fiber and the optical axis of the input light beam (Wyse & Gilmore 1992); and vignetting of the spectrograph and pixel to pixel differences in the CCD. Usually, the correction is achieved by using a uniformly illuminated flat field, either the twilight sky or a screen at the telescope pupil. But for LAMOST, a specially designed reflecting Schmidt telescope with very large FoV, these differences cannot be corrected directly by either kind of flat field.

As shown in Xue & Shi (2008), since the aperture of LAMOST changes with the direction of telescope pointing, the effect of vignetting is as large as 30% across the LAMOST field or 10% for spectrographs at the edge of the LAMOST field, depending on the target declination and hour angle. Considering the fact that twilight is only considered to be homogenous within several degrees around the zenith and there is very limited observation time during twilight, no twilight flat can compensate this position dependent effect. It is also impossible to build a large flat screen in the dome at the telescope pupil, which is a 4.5-meter Schmidt reflector, and illu-

minate it uniformly at the same pointing as observation. So, a twilight flat is only taken to correct the instrumental response along the wavelength direction and measure instrument differences, which are relatively stable with time and position.

On the other hand, twisting, bending and stress of fibers can change the focal ratio degeneration of the light beam at the fiber output end, leading to a change in the fiber throughput. Unfortunately, LAMOST suffers such an effect when the fiber positioner repositions a fiber.

Figure 3 shows the result of a test in which the fiber throughput changed with fiber positioning on 2013 January 19. During the test, the telescope pointing direction and focal plane position were fixed; a flat screen was illuminated by incandescent lamps in front of the focal plane; flat field exposures were taken for two sets of fiber positions in turn, so that neighboring exposures represented different fiber positions and every second exposure had the same fiber position. A total of 14 exposures were taken for each fiber position. Since the difference in position of each fiber between neighboring exposures was relatively small, the difference in flat field brightness between the two positions of the same fiber can be ignored. The only reason that causes a difference between neighboring exposures of the same fiber would be stress put on the fiber by the fiber positioner. Statistics related to the flux ratio between each pair of neighboring exposures and the flux ratio between exposures using the same fiber position show that the uncertainty in throughput caused by the fiber positioner is about 4.8%, much larger than the uncertainty from Poisson noise (0.2%), which will leave large sky subtraction residuals if not corrected properly.

As described above, differences in fiber-to-fiber throughput depend on the telescope pointing direction and the positions of fibers, which cannot be corrected by a twilight flat field. Currently, the only possible solution is calibrating with strong sky emission lines that go through the same light path as the target, such as [OI] 5577 Å in the blue arm and some OH lines in the red arm.

2.3.2 Image shift

Variations in ambient temperature and flexure of the components (if moving with the telescope) will induce instability in the spectrograph. Thus the image shifts in both the spatial and dispersion directions will lead to trace and wavelength calibration errors, and finally bad sky subtraction. For LAMOST, the image shift is mainly caused by temperature variation and loss of liquid nitrogen, which varies the weight on a CCD camera. As shown in Table 1, most of the shift in the wavelength direction during the whole night is smaller than 0.1 Å, but there are certain spectrographs (e.g. spectrograph 16) that are larger. Something apparent from the table is that the shift is not homogenous during the night. An addition to taking arc lamp images between each exposure is using sky emission lines, which is a cheaper but still robust solution to calibrate the shift.

3 METHODOLOGY

Sky subtraction is one of the final steps in the LAMOST 2D data reduction pipeline, but it is dependent on the quality of the previous steps. The spectra are extracted from raw science data using a fiber trace obtained from the flat field frame. The initial wavelength solution is obtained by an arc lamp frame and the initial fiber-to-fiber transmittances are estimated by twilight flat field spectra. The sky emission lines are used to fine tune the wavelength solution and the fiber-to-fiber transmittances. After that, master sky spectra are created from the sky sampling fibers, then subtracted from the object spectra. The object spectra are flux-calibrated, and different exposures are co-added and interpolated to a logarithmically-spaced wavelength scale, $\Delta \lg \lambda = 10^{-4}$. Finally, PCA sky subtraction is performed on the co-added spectra in the wavelength range of 7200–9000 Å, where most sky emission lines lie.

This paper will focus on the sky subtraction, skipping other steps like flux extraction, arc lamp wavelength calibration, flat-fielding, flux calibration and spectra co-addition, which will be described in detail in a forthcoming paper (Bai et al., in preparation). We start from the extracted flux, assuming that the initial wavelength solution and initial fiber transmittances have been performed.

3.1 Identification of Sky Emission Lines

Sky emission lines can be easily identified after an initial wavelength calibration with the arc lamp. In the blue arm, sky emission lines are relatively sparse and street light lines are too weak to offer reliable calibration, so only the strong airglow line [OI] 5577 Å is used. However in the red arm, there are bunches of strong emission lines, such as OH bands. It is not easy to identify single lines in this region, yet after a careful comparison between the observed spectra and those in literature (Osterbrock et al. 1996, 1997), 13 single lines and six doublets are selected, as listed in Table 2. For the doublets, the intensities of the two lines are similar and the separation of the pair is less than 0.5 Å, so that they can be treated as single

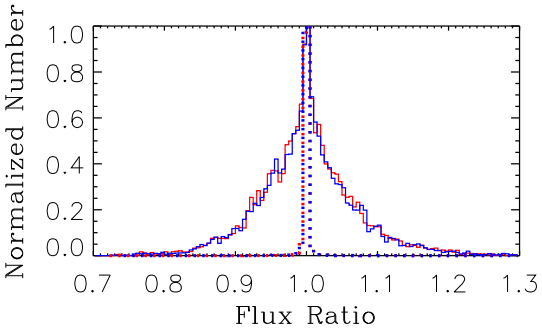

Fig. 3 Flux ratio of dome flat fields. The solid lines show the flat field flux ratio of different fiber positions, while the dotted lines are those of fibers at the same position. The blue and red lines are from the red and blue arms, respectively. σ values of the Gaussian fitting for the solid lines are 4.77% and 4.83% for the blue and red arms, respectively, while those for the dot-dashed lines are 0.280% and 0.283%.

Table 1 Wavelength Shift between Arc Lamp Exposures on 2017 Jan 1

Spectrograph	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
Blue 0–1	–0.010	0.057	–0.064	0.020	0.016	0.013	0.018	–0.104	0.046	0.100	–0.067	0.036	0.004	0.046	0.064	–0.032
Blue 0–2	–0.033	0.081	–0.089	–0.007	0.033	0.023	0.005	–0.104	0.044	0.152	–0.074	0.046	0.013	0.032	0.114	–0.021
Red 0–1	0.003	0.116	–0.010	–0.026	0.038	–0.084	0.083	–0.094	–0.017	0.097	–0.108	0.031	–0.012	0.068	–0.028	–0.166
Red 0–2	0.009	0.164	0.011	–0.040	0.046	–0.127	0.118	–0.023	–0.068	0.140	–0.121	0.043	–0.026	0.063	0.028	–0.183

Notes: "Blue" and "Red" denote the blue and red parts of the spectrograph; 0–1 is the wavelength shift (in Å) between the arc lamp exposure taken at the beginning of the night observation (evening) and that in the middle of the observation (midnight); 0–2 is the difference between the arc lamp exposure at the beginning and that at the end of observation (morning).

lines under LAMOST resolution ($\sim$ 4 Å). Hence, only the average wavelengths of these two lines are adopted in the table.

For the ith selected line with wavelength λ_i in an individual fiber, the profile is fitted with a Sérsic function and a linear background within ± 8 Å around the line center λ_i

$$f(\lambda) = \alpha e^{-\frac{|\lambda-\lambda_i|^\delta}{\delta\gamma^\delta}} + a\lambda + b, \quad (1)$$

where $f(\lambda)$ is the flux corrected by the twilight flat, α and δ are parameters of the Sérsic function and $a\lambda + b$ is the fitted background continuum. The intensity of the line is the sum of the continuum subtracted segment

$$F_i = \int_{\lambda_i-8}^{\lambda_i+8} \Big[f(\lambda) - a\lambda - b\Big] d\lambda. \quad (2)$$

As noticed by Bai et al. (2017) and Li et al. (2015), some of the LAMOST emission line profile cannot be perfectly fitted by a Sérsic function due to optical aberration and distortion. So, the Sérsic function is only used to derive an accurate wing of the emission line and the background function, which are not taken into account in Equation (2).

Table 2 Sky Lines used to Measure the Image Shift and Scale Factor

λ (Å)	Source	Pattern
5577.334	OI	single
6300.304	OI	single
6363.780	OI	single
6863.955	OI	single
6923.220	OH 7-2 P1	single
7316.282	OH 8-3 P1	single
7340.885	OH 8-3 P1	single
7369.366	OH 8-3 P2	blend 7369.248 7369.483
7401.858	OH 8-3 P2	blend 7401.688 7402.029
7750.640	OH 9-4 Q1	single
7794.112	OH 9-4 P1	single
7821.503	OH 9-4 P1	single
7993.332	OH 5-1 P1	single
8025.810	OH 5-1 P2	blend 8025.668 8025.952
8399.170	OH 6-2 P1	single
8465.358	OH 6-2 P2	blend 8465.208 8465.509
8885.850	OH 7-3 P1	single
8943.395	OH 7-3 P2	single
8958.084	OH 7-3 P2	blend 8957.922 8958.246
9001.346	OH 7-3 P2	blend 9001.115 9001.577

3.2 Wavelength Calibration with Sky Lines

Since the image shift varies slightly with fibers, the wavelength solution is corrected fiber by fiber. The wavelength

shift of a sky line is defined as the difference between the literature wavelength $\hat{\lambda}_i$ and the initial line center λ_i

$$\Delta\lambda_i = \hat{\lambda}_i - \lambda_i. \quad (3)$$

Then $\Delta\lambda_i$ values are fitted with a linear function of λ_i

$$\Delta\lambda_i = m\lambda_i + n, \quad (4)$$

and the coefficients m and n are derived by solving the above functions with the least squares method.

Finally, the updated wavelength solution is obtained by

$$\lambda' = \lambda + m\lambda + n, \quad (5)$$

where λ' is the updated wavelength. For the blue arm, m is set to be zero since only [OI] 5577 Å is used. An example of the wavelength correction for the red arm is shown in Figure 4.

3.3 Fiber Transmittance Correction

As described in Section 2.3.1, the relative fiber throughput varies with telescope pointing and fiber position. This could be calibrated using the intensity of sky emission line F, calculated in Equation (2). For the ith line in a fiber, the scale factor $s(i)$ can be calculated by

$$s(i) = \frac{F(i)}{\langle F(i)\rangle}, \quad (6)$$

where $F(i)$ is the line flux obtained by Equation (2) and $\langle F(i)\rangle$ is the median of $F(i)$ over all fibers.

For the blue arm, the scale factor is the relative scale of the [OI] 5577 Å line. For the red arm, the median of $s(i)$ over 19 lines listed in Table 2 is adopted as the final scale factor. The flat field corrected spectra are then divided by the scale factor.

Figure 5 shows an example of the distribution of the scale factors on 2016 October 26. The scale factors in both the blue and red arms show a similar non-Gaussian distribution with standard deviation of about 0.058, while the ratio of the scale factor of the blue arm to the red one is close to a Gaussian distribution with standard deviation of 0.028. These indicate that the uncertainty in fiber throughput induced by the effect of telescope vignetting and the fiber positioner is about 5.8%, while the accuracy of scale factor correction is about 2.8%.

As measured from the data, the uncertainty of sky emission line intensity is about 1.5% for [OI] 5577 Å and any single OH line. Considering that s_r is derived from the median of 19 OH lines, the uncertainty of the median is approximately $1.5\%/\sqrt{19}\approx 0.35\%$. From the twilight flat fields observed on two consecutive days, the uncertainty of the twilight sky flat is about 1.6%. In total, the synthetic uncertainty of s_b/s_r is about $\sqrt{1.5^2 + 0.35^2 + 2\times 1.6^2} \approx 2.7\%$, consistent with the accuracy of scale factor correction.

3.4 Master Sky

After the wavelength and fiber transmittance are fine tuned by sky emission lines, a master sky spectrum is created from the sky fibers in the same spectrograph, using a B-spline fitting procedure similar to the SDSS 2D pipeline (see SDSS data reduction pipeline 'idlspec2d', Bolton & Burles 2007). Spectra of sky fibers are treated as fluxes in discrete pixels; pixels from different sky spectra are aligned together in order of their wavelength. The master spectrum is fitted in 2D using a cubic B-spline function in the wavelength direction, and allowing the B-spline coefficients to vary with fiber number. Bad pixels are rejected during the fitting. The B-spline function is then interpolated back to each fiber, obtaining the final sky spectrum, which will be subtracted from the object spectrum. Figure 6 shows an example of the master sky spectrum and the sky subtraction residuals.

3.5 PCA Sky Subtraction

After the master sky is subtracted, each spectrum is flux calibrated, then different exposures are combined and interpolated to a logarithmically-spaced wavelength scale, i.e. $\Delta \log\lambda = 10^{-4}$. PCA is performed on the combined spectra in the range 7200–9000 Å, where the OH sky emission lines are dominant. For each spectrograph, about 20 sky subtracted sky spectra are used to generate the components of PCA. Both the sky and object spectra are first continuum subtracted using a rolling median filter to remove large-scale structures, then the eigenvectors and eigenvalues of PCA are derived from the ∼20 sky residual spectra. For each spectrum, a projection coefficient is calculated for each eigenvector, and the sum of the 20 most-significant principal eigenvectors weighted by the projection coefficients is adopted to derive the sky residual spectrum. This residual part is then removed from the spectrum and the median filtered continuum is added back. Details about PCA sky subtraction can be found in Wild & Hewett (2005) and Sharp & Parkinson (2010).

4 ACCURACY OF SKY SUBTRACTION

The sky subtraction routine described here is part of the LAMOST 2D Pipeline v2.6 and has been applied in all LAMOST data later than DR 3. As the sky subtraction is

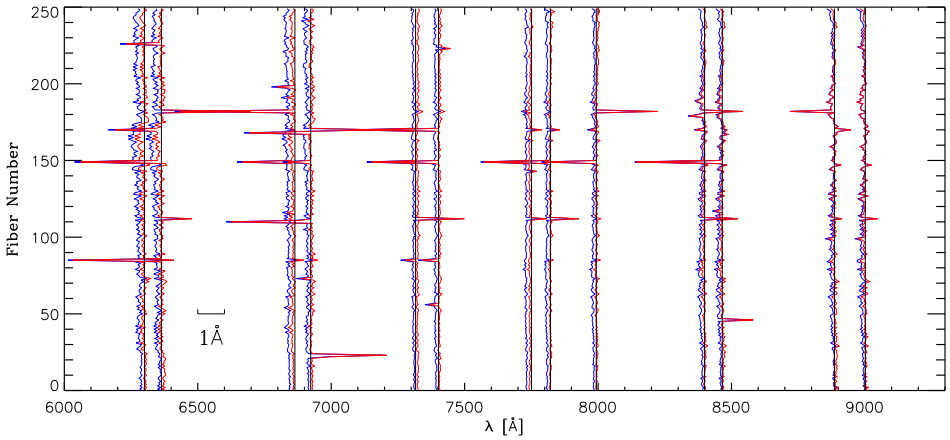

Fig. 4 An example of the wavelength corrections in a single frame. The blue and red lines show measured wavelengths of the sky emission lines based on the arc lamp solution and the fine tuned wavelength of the sky emission line, respectively; the black lines are wavelengths in the literature. The distance of the blue and red lines relative to the black ones are exaggerated by 100 times to show details. The average difference between the initial wavelength solution and the corrected wavelength is 0.13 ± 0.07 Å, which is typical for the correction. Large residuals seen in the plot are caused by bad fibers or bad pixels.

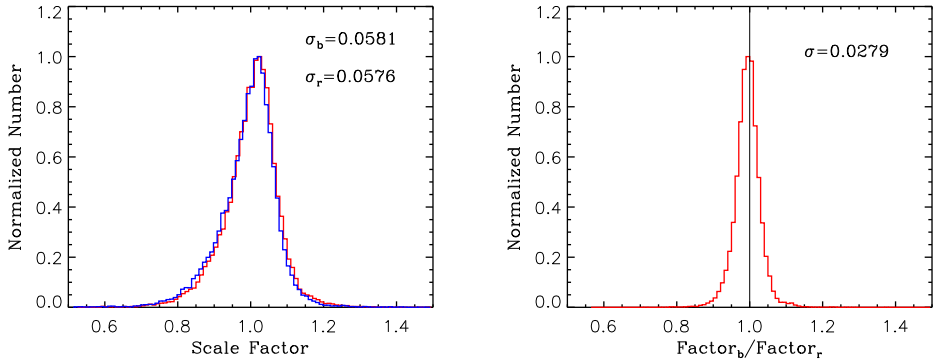

Fig. 5 Left: Distribution of the scale factors of the blue and red arms, with color representing each. Right: Distribution of the ratio of the blue factors to red ones. The standard deviations are as indicated.

performed in two stages (i.e. the master sky stage and the PCA stage), the accuracies of the two stages are analyzed separately.

4.1 Wavelength Calibration Accuracy

The typical error of sky line calibration within one observation is about 0.07 Å, as shown in Figure 4. It is not clear whether our calibration is stable between observations, so it is necessary to test it by measuring the radial velocity (RV) variations of stars. There are quite a lot of stars with multiple observations from more than one day in the LAMOST database and their RVs can be used to indicate the stability of our wavelength calibration. A search of repeated observations of F, G and K stars with S/N over 20 in LAMOST DR3 results in 689 897 spectra of 301 106 stars. Every two observations of the same star is defined as a "pair," so there will be C_n^2 pairs for a star with n observations. For each pair, ΔRV is defined as $|v_1 - v_2|$, where v_1 and v_2 are the measured RVs of the two observations, respectively. The distribution of the RV difference vs. the number of observations is shown in Figure 7. The standard deviation of the Gaussian fit for the core of the ΔRV distribution is $4.47\,\mathrm{km\,s^{-1}}$, consistent with typical wavelength calibration uncertainty (0.07 Å).

4.2 Sky Subtraction Accuracy of a Single Frame

A forthright way to estimate the sky subtraction accuracy is to measure the residual part of the sky subtracted sky spectra. The absolute sky residual is dominated by shot noise from the original sky flux. A relative sky residual

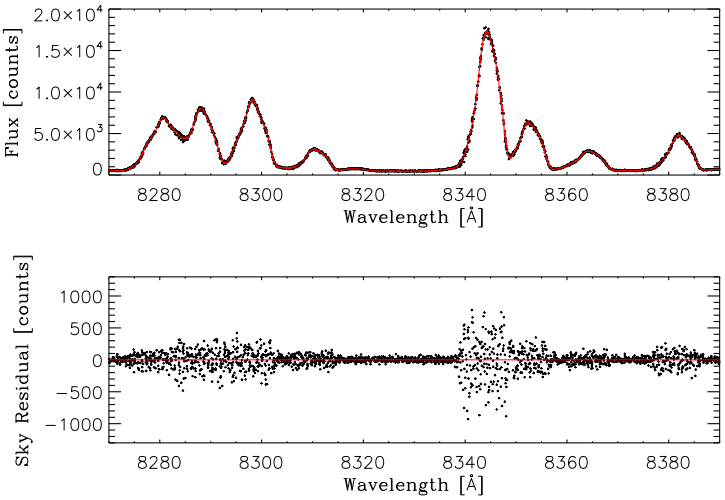

Fig. 6 An example of sky subtraction for spectra taken on 2016 Dec. 31. Upper: Fluxes from 20 sky fibers attached to one spectrograph are corrected by sky emission lines. The black crosses show all 20 sky spectra and the solid red line indicates the best fitted master sky. The discrepancy of different sky fibers is too small to be discriminated on the current scale. Lower: The residuals of sky subtracted sky spectra.

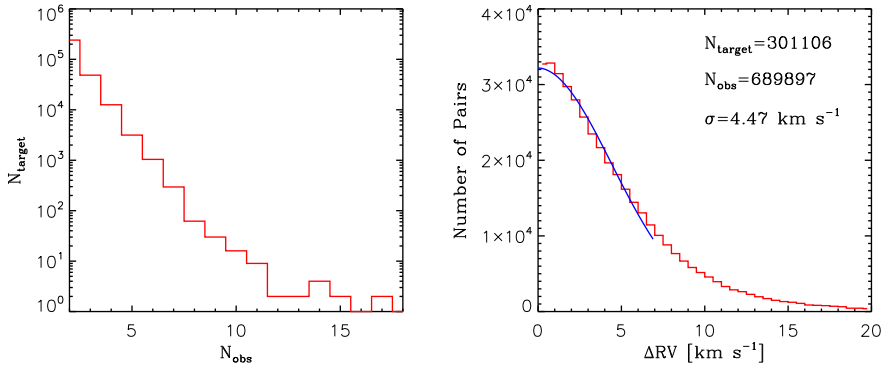

Fig. 7 *Left*: A histogram of repeated observations in the LAMOST database. The x-axis indicates the number of repeated observations. *Right*: Distribution of RV difference for repeated observations; the blue curve is a Gaussian fit to the "core" of the distribution.

is defined as the ratio of the absolute sky residual to the sky flux

$$r_s(\lambda) = \frac{f_r(\lambda)}{f(\lambda)}, \tag{7}$$

where $f_r(\lambda)$ is the absolute residual of the sky spectra and $f(\lambda)$ is the original sky flux.

For sky emission lines, since the typical FWHM of any single sky line is 3–4 Å, pixels within ±3 Å around the line center are used to calculate the residuals of sky emission lines. In the blue arm, [OI] 5577 Å is measured, while in the red arm 10 strong lines including 7714 Å, 7750 Å, 7794 Å, 7821 Å, 7853 Å, 7913 Å, 7964 Å, 7993 Å, 8025 Å and 8062 Å are adopted.

For sky continuum, the flux in an individual pixel is much lower than that of the emission lines. To depress shot noise, instead of using counts from individual pixels in Equation (7), the averages of the continuum in the 5470–5560 Å and 6000–6200 Å regions (see Fig. 2) are adopted for the blue and red arms, respectively

$$r_s(\lambda) = \frac{\overline{f_r(\lambda)}}{\overline{f(\lambda)}}. \tag{8}$$

As an example, Figure 8 shows the distribution of relative sky residuals on 2016 September 20. The dispersion of the residuals, which can be estimated by σ of the Gaussian fitting of the histogram, is an indicator of the sky subtraction accuracy. As can be seen in Figure 8, the histogram of the [OI] 5577 emission line residuals can be

well described by a Gaussian function except at the tail of the histogram, which is mostly caused by the profile difference between the master sky spectrum and the optical aberration distorted profile of individual fibers. In the OH line region, the scatter is larger, since not all the 10 lines selected to measure the residual part are from the line list in Table 2 that determines the scale factor. The residual of non single lines will also be affected by neighboring lines. The residuals in the continuum are consistent with corresponding emission lines, indicating that the scale factor works well in the continuum region.

The LAMOST survey is carried out on both bright and dark nights. On moonlit nights, the relative strength of a sky emission line to the sky continuum is weaker than that on darker nights, due to the increase of background brightness. To investigate how the accuracy of sky subtraction depends on moonlight, we have computed σ for all the LAMOST sky subtracted sky spectra taken on each night before 2016 December 31 and the results are shown in Figure 9.

There is no obvious systematic tendency for σ of sky continuum residuals on Moon phase. The average values of σ for the blue and red sky continuum residuals are comparable to those in Figure 8, but with much larger scatter than σ of the emission lines. There are three reasons for the large scatter. The first is that the exposure time varies between 10 and 30 minutes, so both the continuum and the sky emission line used to calibrate the continuum can vary by as much as 3; thus the relative noise can change by $\sqrt{3 \times 2} = 2.45$. The second is that both the sky continuum and the emission line change from time to time. The third is that sky continuum does not necessarily vary in the same way as the sky emission line, especially on bright nights, though the difference should be small inside an individual spectrograph (see Sect. 5.1). A check of points with large scatter on dark and grey nights in the upper panel of Figure 9 shows that those are from observations with short exposures, where both the sky continuum and emission lines are relatively weak. The large scatter of continuum residuals from bright nights follows a similar trend of emission lines in the corresponding Moon phase, which is caused by the short exposure time and dramatic variation of the moonlight background, thus increasing the relative uncertainty of the sky emission lines.

The relative residuals of sky emission lines are much smaller on dark nights than the continuum, with a median level of 1.5% for [OI] 5577 and 2.2% for OH lines, but obviously increasing when the Moon phase is between the 9th and 21st day; up to 3% for [OI] 5577 and 4.5% for OH lines on full Moon nights.

Considering that the [OI] 5577 emission line is less affected by neighboring lines than OH lines in the red arm and the region used to calculate scatter in the blue arm is the region around [OI] 5577 itself, it is easy to understand why the blue arm has smaller residuals.

4.3 Advantages of PCA

Because the PCA sky subtraction focuses on the OH emission lines in the range of 7200–9000 Å, the residuals of these lines decrease significantly. Three examples with different S/N are shown in Figure 10.

To quantify the contribution of PCA subtraction, the spectra of 7522 F-type stars observed on 2016 Feb 20 are analyzed. For F stars, there are less features in the wavelength range 7720–8100 Å. On the contrary, there is an abundance of sky emission lines in the same region, so the smoothness of the continuum could be used to evaluate the quality of sky subtraction. Smoothness is defined as

$$a_t = \mathrm{RMS}\Big(\frac{f_r - c}{c}\Big), \tag{9}$$

where f_r is the co-added object spectra, c is the pseudo-continuum derived by a rolling median filter and RMS is short for Root Mean Square. Figure 11 shows the ratio of the RMS a_t after PCA sky subtraction to that before PCA sky subtraction. All of the stars show smaller RMS after PCA sky subtraction. The peak occurs at about 75%, which indicates that PCA improves the sky subtraction of OH lines by about 25%. Incorporating the results of Figure 9, the median of the final sky subtraction residuals from OH lines on dark and the bright nights can reach as low as 1.7% and 3.4%, respectively.

4.4 Comparison with SDSS

SDSS (York et al. 2000), similar to LAMOST in resolution and wavelength coverage, is the most successful low resolution multi-fiber spectroscopic observing program to date. SDSS has higher system efficiency than LAMOST, 18% and 20% in the blue and red arms, respectively (Stoughton et al. 2002). Considering the differences in efficiency, signals from the same object in the two surveys are different, so it is hard to tell which survey shows a smaller relative sky subtraction residual to the object continuum. Instead of comparing coincident targets in both surveys, comparing targets with both similar strengths of sky emission lines and similar counts of object flux is more reasonable. As shown in the top left panel of Figure 12, about 11 000 F stars are selected from both SDSS DR10 and LAMOST DR3 with similar inte-

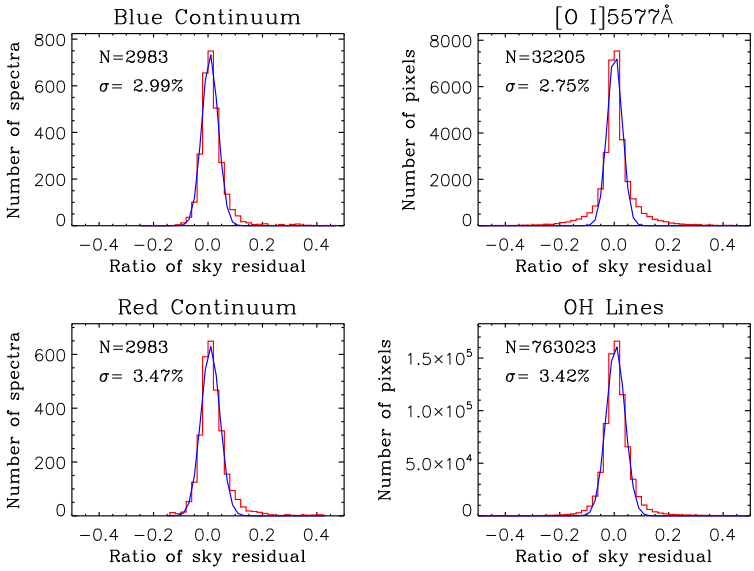

Fig. 8 Distributions of residuals from sky subtracted sky spectra taken on 2016 Sep. 20. The left column shows residuals from the average continuum (see Equation 8) in 5470–5560 Å and 6000–6200 Å for the blue arm (*top*) and red arm (*bottom*). The right column displays distributions of residuals from individual pixels of sky emission lines. Each histogram is fitted by a Gaussian function, as shown by the blue curve in each plot. The total number of pixels and σ_r of the fitting are marked in the plots.

gral strength of [OI] 5577 Å and similar object continuum intensity. Checking the same sample in the red arm shows that the OH sky emission lines in LAMOST are much stronger than those in SDSS, so we only keep those LAMOST spectra with strengths of OH 7913 Å comparable to those of SDSS, and reject about 7000 spectra with stronger OH lines, as shown in the top right panel of Figure 12.

When the intensities of sky emission lines are similar in the blue part of the two surveys, the sky continuum of LAMOST is brighter than that of SDSS, thus the S/N of the LAMOST stellar continuum in 5500-5560 Å is lower than that of SDSS. However in the red part, the sky continuum intensity is similar to that of SDSS, so the S/N of the two surveys is similar. As can be noticed in the middle row of Figure 12, the reason why the S/N of the two surveys departs at the high object flux region is as follows.

The S/N of both LAMOST and SDSS spectra is defined as

$$\mathrm{S/N} = \mathrm{flux} * \sqrt{\mathrm{invvar}}\,, \tag{10}$$

where 'invvar' is the inverse variance, which is accumulated during data processing. In SDSS image processing, it is defined as

$$\mathrm{invvar} = \frac{1}{\sqrt{f + (g * n_{\mathrm{rd}})^2 + 10^{-4} f^2}}, \tag{11}$$

where f is the object flux count, g is the gain and n_{rd} is the read-out noise. The third term in the denominator is just to limit the S/N below 100. But in LAMOST image processing, the inverse variance is simply

$$\mathrm{invvar} = \frac{1}{\sqrt{f + (g * n_{\mathrm{rd}})^2}}. \tag{12}$$

The difference in estimation makes SDSS spectra have lower S/N than LAMOST at the same flux, especially when the flux is high.

According to Equation (9), the residuals in the emission line region of the blue (5570–5585 Å around [OI] 5577) and red (7700–8100 Å in the region of OH lines) parts are calculated respectively for each star. The residuals vs. their object continuum intensities are plotted in the bottom panel of Figure 12. LAMOST spectra show smaller sky emission line subtraction residuals than SDSS in both the blue and red parts of the spectra, but the red part is more significant than the blue one. As indicated in the middle left panel of Figure 12, in the blue part, the S/N of the LAMOST spectra is lower than that of SDSS, so the actual effect of LAMOST sky subtraction should be even better than SDSS if they are in the same situation. The two reasons that LAMOST performs better than SDSS in sky subtraction may be: LAMOST sky sampling fibers are denser than those of SDSS, both on the sky (20 fibers per 1.25 square degrees vs. 32 fibers per 7 square degrees) and in the CCD image (20 per 250 fibers vs. 32 per 640 fibers), so the sky spectra are better represented in LAMOST; the PCA sky subtraction, which contributes a lot to reducing the sky subtraction residuals in the red part of LAMOST spectra, is not used

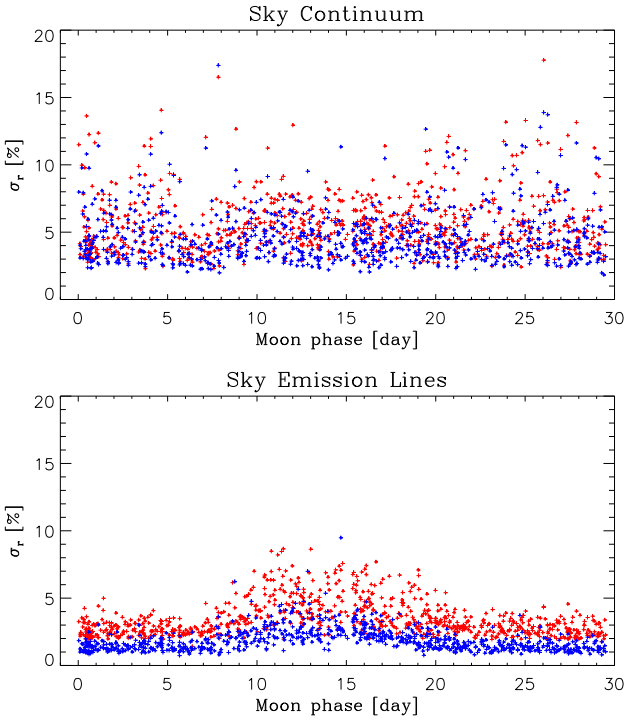

Fig. 9 Relative sky subtraction residual σ_r vs. Moon phase. Upper panel: σ_r of the average sky continuum. Lower panel: σ_r of individual pixels associated with sky emission lines. The blue and red symbols represent the blue and red arms, respectively. The x-axis is date of the Moon phase, where 0 denotes New Moon and 14.7 denotes Full Moon.

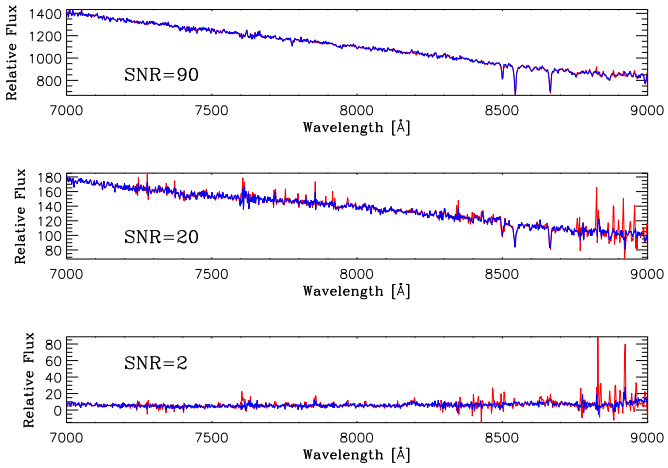

Fig. 10 Three example spectra with different S/N in the 7200 Å–9000 Å band. In each panel, the master sky subtracted spectrum is shown in red and the PCA sky subtracted spectrum is in blue.

for the SDSS spectra. Many other details in software and hardware may also explain the different performances of the two surveys, but it is hard to make a comprehensive comparison.

5 DISCUSSION

The final sky subtraction accuracy is affected by many factors in the data reduction process. Most of them can be corrected by the above schemes, but there are certain problems that cannot be solved currently, such as the moonlight background and variation of the PSF.

5.1 Sky Subtraction in a Moonlit Night

The sky emission line calibration of fiber throughput is based on the assumption that sky background is homoge-

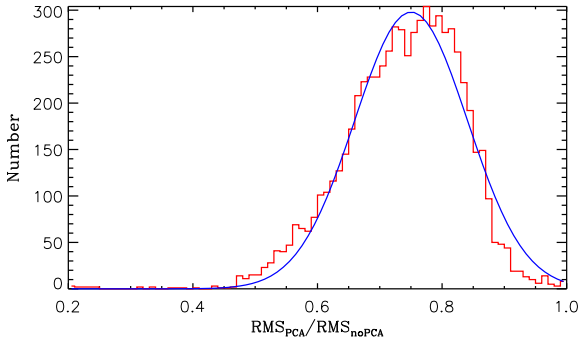

Fig. 11 Distribution of the ratio of sky subtraction residuals for 7522 F-type stars after PCA sky subtraction to that before PCA sky subtraction. The red line is the histogram and the blue line is the Gaussian fitting of the histogram.

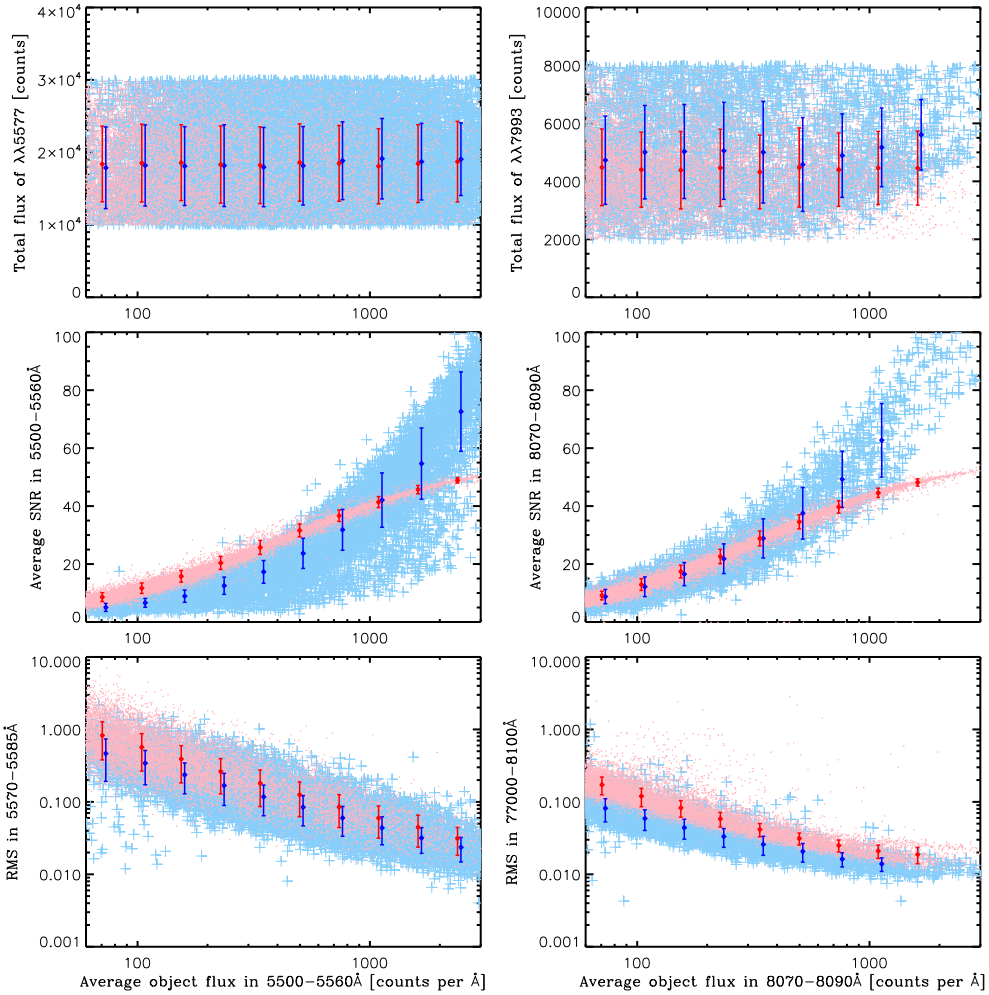

Fig. 12 Comparison of sky emission line subtraction residuals for SDSS and LAMOST F type stars. The blue symbols signify LAMOST DR3 and red symbols represent spectra of SDSS DR10 in all panels. The stars are selected with similar [OI] 5577 Å integral intensity and similar object continuum flux. The object flux, indicated by the x-axis of all panels, is calculated as the average counts per angstrom of star continuum in 5500–5560 Å and 8070–8090 Å for the blue and red parts, respectively. Top: the star continuum intensity vs. integral flux of sky emission line, [OI] 5577 Å and OH 7913 Å for the blue and red parts, respectively. Middle: the star continuum intensity vs. S/Ns in the blue and red parts, respectively. Bottom: the star continuum intensity vs. relative residuals of the sky emission lines, [OI] 5577 Å and OH lines in 7700–8100 Å for the blue and red parts, respectively. Note that both x- and y-axes are in log scale, so the error bars are not symmetric.

nous and the intensity of sky continuum is proportional to that of the emission line. Generally, the assumption is good on a dark night within a one degree FoV, which is comparable to the FoV of a single LAMOST spectrograph. On a moonlit night, moonlight scattered by the atmosphere produces a background gradient, which must be taken into account during sky subtraction.

The sky brightness at the LAMOST site is about $V = 17\,\mathrm{mag\,arcsec^{-2}}$ at a point in the lunar cycle of about 14 days (Yao et al. 2013), which is close to a Full Moon. Considering the size of a LAMOST fiber, $3.3''$ in diameter, moonlight dominates the sky background with $V = 14.85$ mag. Moreover, the faintest targets that can be observed on a bright night are about 14 mag, 2.2 times the background brightness. The typical flux for a 10 min exposure of a $V = 14$ mag star is about 11 000 counts $\mathrm{pix^{-1}}$, while the sky background is about 5000 counts $\mathrm{pix^{-1}}$, which means 2% of the sky background residual will lead to a 1% uncertainty in the object flux. Thus a 4% sky background gradient across the spectrograph field will lead to a 1% systematic bias in sky subtraction, which is acceptable when sky subtraction residuals are generally larger than 2%, as in Figure 9. The sky brightness gradient depends strongly on the angular distance between the target field and the Moon. If the sky background gradient inside a single spectrograph is limited to less than 4%, then we can define a secure distance to the Moon, beyond which the current sky subtraction scheme is still suitable.

With a Moon night sky brightness model, Yao et al. (2013) calculated the typical sky brightness distribution for the LAMOST site. According to the model, the secure angular distances for sky brightness gradients of 2%, 3% and 4% are derived respectively for different Moon phases, as shown in Figure 13. The angular distance needs to be more than 33° to obtain a gradient of less than 4%. Most LAMOST observations, both regular surveys and test observations during Full Moon nights, conform to this situation. A better sky subtraction scheme including the moonlight brightness model will be future work for the LAMOST 2D pipeline.

5.2 Optical PSF Variations

Sky subtraction with the master sky spectrum requires that the shapes of the PSFs in different fibers are similar, which is usually not satisfied due to imperfect spectrograph optics such as optical aberration and distortion, or due to the irregular shape of a fiber output end (e.g. due to the poor coupling between the fiber alignment and the slit). The sky subtraction residuals caused by profile differences in master sky subtraction cannot be completely removed by the PCA sky subtraction. To further improve sky subtraction, future works like careful re-alignment of the fiber and the slit, and introducing a spectral extraction method with a 2D de-convolution algorithm (e.g. Bolton & Schlegel 2010; Li et al. 2015) will be necessary.

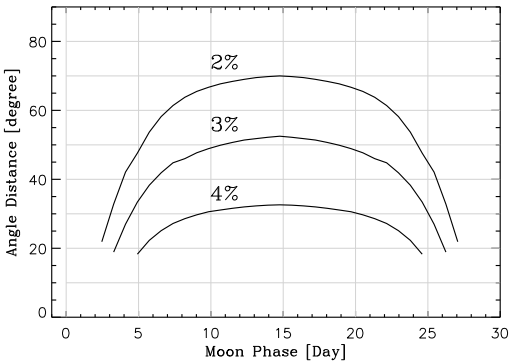

Fig. 13 The curve of angular distance between the LAMOST field center and the Moon vs. the Moon phase. The curves represent the limit of angular distance beyond which the moonlight brightness gradient inside an individual spectrograph is less than the given value marked above each curve. The x-axis is the day of a Moon cycle, where 0 denotes New Moon and 14.7 denotes Full Moon.

6 CONCLUSIONS

Sky subtraction, related to almost every step of the data reduction process, is the most important indicator of performance for a 2D multi-fiber data reduction pipeline. The key algorithms and the results of sky subtraction in the LAMOST 2D pipeline are demonstrated here. Due to special characteristics of the LAMOST telescope (variable vignetting with telescope pointing) and defects in hardware manufacturing and installation (force added by fiber positioner to the fibers), the throughput of LAMOST varies with telescope pointing and fiber position, which leads to the failure of traditional flat field correction methods, leaving the sky emission line throughput calibration as the currently most reliable method. Sky emission lines are also used to fine tune wavelength shifts by about 0.1 Å caused by instability of the spectrograph. After subtraction by the fine tuned master sky spectra, the red part of the object spectra is processed by PCA sky subtraction to further remove the emission line residual part of the OH band. Overall, the wavelength calibration accuracy is about 4.5 km $\mathrm{s^{-1}}$ according to RV measurements of repeated observations of the same stars. On dark nights, the median of the sky residual is about 1.5% and 1.7% for [OI] 5577 and OH lines, respectively. On moonlit nights, the residuals rise to 3% and 3.4% for

[OI] 5577 and OH lines respectively, due to decrease in the exposure time. For the sky continuum, typical relative residuals are about 3% (Fig. 8). As pointed out in Section 5.1, a systematic bias in sky subtraction could be limited to within 1% if the distance between a target and the Moon is larger than 33°. Our final sky subtracted spectra show smaller residuals in the sky emission line region than SDSS spectra from the analysis of F-type stars.

Acknowledgements Thanks for the advice of the reviewers. This work is supported by the National Natural Science Foundation of China (NSFC) (Grant No. 11503054), NSFC Key Program (Grant No. 11333004) and the National Key Basic Research Program of China (Program 973; Grant No. 2014CB845700).

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.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III website is *http://www.sdss3.org/*.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

References

Bai, Z., Zhang, H., Yuan, H., et al. 2017, PASP, 129, 024004

Barden, S. C., Elston, R., Armandroff, T., & Pryor, C. P. 1993, in Astronomical Society of the Pacific Conference Series, 37, Fiber Optics in Astronomy II, ed. P. M. Gray, 223

Benn, C., & Ellison, S. 1998, La Palma Tech. Note, 115

Bolton, A. S., & Burles, S. 2007, New Journal of Physics, 9, 443

Bolton, A. S., & Schlegel, D. J. 2010, PASP, 122, 248

Cuby, J.-G., & Mignoli, M. 1994, in Proc. SPIE, 2198, Instrumentation in Astronomy VIII, ed. D. L. Crawford & E. R. Craine, 98

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

Elston, R., & Barden, S. 1989, NOAO Newsletter, 9, 21

Glazebrook, K., & Bland-Hawthorn, J. 2001, PASP, 113, 197

Kurtz, M. J., & Mink, D. J. 2000, ApJ, 533, L183

Li, G., Zhang, H., & Bai, Z. 2015, PASP, 127, 552

Lissandrini, C., Cristiani, S., & La Franca, F. 1994, PASP, 106, 1157

Meinel, A. B. 1950, ApJ, 111, 433

Osterbrock, D. E., Fulbright, J. P., Martel, A. R., et al. 1996, PASP, 108, 277

Osterbrock, D. E., Fulbright, J. P., & Bida, T. A. 1997, PASP, 109, 614

Puech, M., Rodrigues, M., Yang, Y., et al. 2014, in Proc. SPIE, 9147, Ground-based and Airborne Instrumentation for Astronomy V, 91476L

Rodrigues, M., Cirasuolo, M., Hammer, F., et al. 2012, in Proc. SPIE, 8450, Modern Technologies in Space- and Ground-based Telescopes and Instrumentation II, 84503H

Sharp, R., & Parkinson, H. 2010, MNRAS, 408, 2495

Soto, K. T., Lilly, S. J., Bacon, R., Richard, J., & Conseil, S. 2016, MNRAS, 458, 3210

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

Watson, F., Offer, A. R., Lewis, I. J., Bailey, J. A., & Glazebrook, K. 1998, in Astronomical Society of the Pacific Conference Series, 152, Fiber Optics in Astronomy III, ed. S. Arribas, E. Mediavilla, & F. Watson, 50

Wild, V., & Hewett, P. C. 2005, MNRAS, 358, 1083

Wynne, C. G. 1993, MNRAS, 260, 307

Wyse, R. F. G., & Gilmore, G. 1992, MNRAS, 257, 1

Xue, Y., & Shi, H.-M. 2008, ChJAA (Chin. J. Astron. Astrophys.), 8, 580

Yao, S., Zhang, H.-T., Yuan, H.-L., et al. 2013, RAA (Research in Astronomy and Astrophysics), 13, 1255

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579