

# A Sky-Subtraction Algorithm for LAMOST Using Two-Dimensional Sky-Background Modeling

J. Zhu<sup>A</sup> and Z. Ye<sup>A,B</sup>

<sup>A</sup>Institute of Statistical Signal Processing, Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China

<sup>B</sup>Corresponding author. Email: [yezf@ustc.edu.cn](mailto:yezf@ustc.edu.cn)

**Abstract:** A novel algorithm is proposed for the sky subtraction of the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) based on two-dimensional sky background modeling. Different from the standard fiber spectrum data processing techniques, a two-dimensional sky background model can be obtained with the new algorithm and sky subtraction can now be performed as an earlier step, before the spectrum extraction. In this study, experiments are performed on simulated data based on the LAMOST project to analyze the accuracy and the effectiveness. The results show that the proposed algorithm can give a more effective sky subtraction than the method that is currently used for LAMOST.

**Keywords:** instrumentation: spectrographs — methods: data analysis — techniques: spectroscopic — telescopes

*Received 2011 December 5, accepted 2012 January 13, published online 2012 February 10*

## 1 Introduction

Compared with the traditional slit spectroscopy, multi-object spectroscopic observations can simultaneously observe a great number of high-quality celestial spectra. Although large-field, multi-object spectroscopic observation has been in use for decades, sky subtraction is still an important and difficult process which effects the accuracy and quality of the obtained original object spectrum, due to the limited number of sky sampling fibers and the variable large-field sky background.

Astronomers have made a lot of effort in this research field and have extensively studied this technology and spectral processing methods (Elston & Barden 1989; Wyse & Gilmore 1992; Wynne 1993; Cuby & Mignoli 1994). The results show that the accuracy of sky subtraction for multi-object-fiber spectroscopic observations is influenced by many factors, such as the stray light inside the spectrometer, the changes of sky in the observation field, telescope efficiency, pollution of adjacent fibers, fiber transmission and the accuracy of wavelength calibration.

Depending on the software and equipment, Zhang, Chu & Chen (2002) made some discussion and appropriate improvements about the stray light, fiber transmission, wavelength calibration and flux calibration in order to improve the effectiveness of sky subtraction, by using the data obtained by the multi-object fiber spectrometer of SAO 6-m telescope. Parry and Carrasco (1990) verified that the signal-to-noise ratio of a spectrum after sky subtraction depends not only on the response of the equipment to the sky and object spectra, but also on the

sub-pixel sampling in the wavelength orientation. Their results show that high-precision wavelength calibration can improve the accuracy of sky subtraction. An improved algorithm of wavelength calibration (Abazajian et al. 2004) was proposed to eliminate the effects of OH emission lines through calibration at the sub-pixel level. Algorithms based on beam switching (Watson 1987) and the sky-chopping technique (Barden et al. 1993) were proposed for increasing the accuracy of sky subtraction, but their disadvantage is the expense of observing efficiency.

By considering the stray light from the spectrometer during the sky subtraction process, the precision of sky subtraction has effectively improved (Wyse & Gilmore 1992). An optimal subtraction algorithm was proposed for infrared fiber spectroscopy based on point spread function (PSF) matching (Causi & De Luca 2005), and it can compensate for both imperfect wavelength calibration residuals and fiber-to-fiber PSF variations. Use of the Principal Components Analysis (PCA) approach for sky subtraction was first demonstrated by Kurtz and Mink (2000) and then improved by Wild and Hewett (2005), using data from the Sloan Digital Sky Survey (SDSS). The application of PCA eliminates large amounts of the residual OH emission lines in the spectra, and greatly improves the precision of sky subtraction. The PCA approach was applied to over 200 h of on-sky observations for the AAOmega system of the Anglo-Australian Telescope (Sharp & Parkinson 2010), who made some discussions about the sky subtraction at the Poisson limit with fiber-optic multi-object spectroscopy.

**Table 1.** The main parameters of spectra observed by LAMOST<sup>a</sup>

Parameter	Blue channel	Red channel
Wavelength range (Å)	3700–5900	5700–9000
Number of points <sup>b</sup>	4096	4096
Dispersion (Å pixel <sup>-1</sup> )	0.57	0.85
Resolution	1000	1000
FWHM <sup>c</sup> (pixel)	8	8
Separation <sup>d</sup> (pixel)	15	15

<sup>a</sup>Low-resolution mode, full slit.<sup>b</sup>The number of pixels along the wavelength orientation.<sup>c</sup>The full width at half maximum of spectra.<sup>d</sup>The spacing of two adjacent fiber centers along the spatial orientation.

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) is a meridian reflecting Schmidt telescope with 4-m aperture and 5-deg field of view (Zhu et al. 2006; Cui 2009). The aim of LAMOST is to observe the spectra of 107 galaxies and 105 quasars to a limiting magnitude of 20.5. The telescope is equipped with 4000 fibers, which are averagely distributed to 16 multi-object fiber spectrographs. The main parameters of the spectra observed by LAMOST (low resolution mode, full slit) are listed in Table 1.

In each spectrograph, there are 250 fibers, 240 of which are used for observing the targets, while the other 10 s are used for observing the sky. The sky sampling fibers are limited and the system has high complexity, which present a great challenge to the sky subtraction. The sky subtraction method adopted by the SDSS, which is also a multi-object spectroscopic observation, is based on B-spline curve fitting. The research of Bai, Zhang & Ye (2008) focused on the red channel of spectra (6700–9180 Å), and analyzed all the sky spectra in one observation based on the PCA method. It is more effectual and provides a good solution for the sky subtraction in LAMOST.

In this paper, a novel sky subtraction algorithm is proposed for LAMOST. Two-dimensional sky background modeling is acquired based on basis-spline (B-spline) curve fitting and sky subtraction is performed as an earlier step, before the spectrum extraction. The experimental data are simulated based on the parameters of LAMOST and presented to demonstrate the effectiveness of the proposed method.

In Section 2, the proposed sky subtraction algorithm is described. Experiments and results are given in Section 3. The summary is presented in the last section of this article.

## 2 The Proposed Sky-Subtraction Algorithm

### 2.1 B-Spline Curve Fitting

The B-spline function is a widely used curve-fitting technology for numerical analysis. B-splines function can be evaluated in a numerically stable way by the de Boor algorithm (de Boor 1978). Based on research into Bezier curve fitting, the B-spline function was introduced

as a basic method for mathematical shape description (Gordon & Riesenfeld 1974). The B-spline model was often used in the processing of fiber spectrum data, such as bias subtraction, wavelength calibration, spectrum extraction and sky subtraction (Bolton & Burles 2007).

The advantage of B-spline curve fitting is its capability to model flexibly. As long as the number and the locations of spline nodes have been chosen, we can construct a special curve containing line segment, sharp points, tangent and other special circumstances, which is very beneficial to simulate the characteristics of sky.

The B-spline function of one-dimensional sky is described by

$$sky(x) = \sum_{k=0}^n P_k B_{k,d}(x). \quad (1)$$

where  $x_{\min} \leq x \leq x_{\max}$  and  $2 \leq d \leq n + 1$ .  $P_k$  denotes the fitting parameters which are also the control points of B-spline function, which can be figured out by least-square method and Cholesky decomposition.  $B_{k,d}(x)$  are the B-spline basic functions of degree  $d$ , which can be obtained by the de Boor–Cox recursion formula and are described as follows:

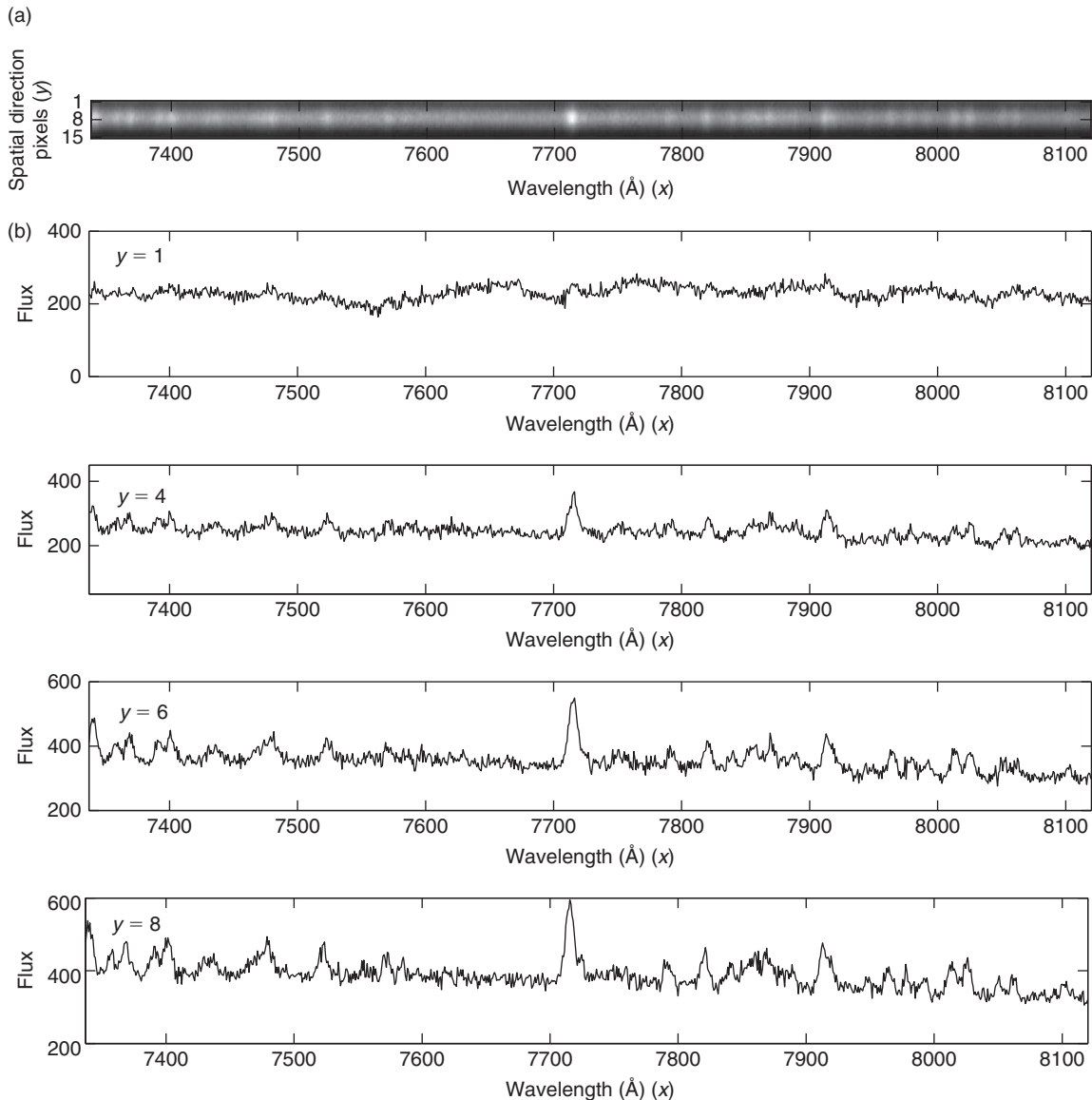
$$B_{k,1}(x) = \begin{cases} 1, & x_k \leq x \leq x_{k+1} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

and

$$B_{k,d}(x) = \frac{x - x_k}{x_{k+d-1} - x_k} B_{k,d-1}(x) + \frac{x_{k+d} - x}{x_{k+d} - x_{k+1}} B_{k+1,d-1}(x), \quad (3)$$

for  $2 \leq d \leq n + 1$ , where  $x_k$  are the coordinates of spline nodes, and  $d$  denotes degree of the polynomial. These formulas show that it needs  $(d + 1)$  nodes ( $x_k, \dots, x_{k+d}$ ) to determine the  $k$ th  $d$ -degree basic spline function  $B_{k,d}(x)$ . The B-spline curve can influence the local character of a curve by changing the number and location of the control points.

The currently widely used sky subtraction method is based on the B-spline curve fitting, which is also adopted by the SDSS and LAMOST. After preprocessing, such as wavelength calibration, spectrum extraction and efficiency-correction of the fiber, the fluxes of the ten sky fibers are reordered according to the wavelength. Each fiber has different dispersion curves. Therefore, we can obtain a sky spectrum of which the sampling rate has increased ten times. Then the B-spline curve fitting method is applied to obtain an overall sky spectrum. For each object spectrum, sky-subtraction processing can be executed by subtracting the overall sky spectrum. The advantage of this method is that if a sky fiber has anomalous flux at a certain wavelength, this point will be rejected during the process of the fitting.



**Figure 1** A section of two-dimensional sky spectrum before the spectrum extraction and its slices. (a) A section of two-dimensional sky spectrum; (b) the slices of two-dimensional sky when the spatial component  $y = 1, 4, 6$  and  $8$  respectively.

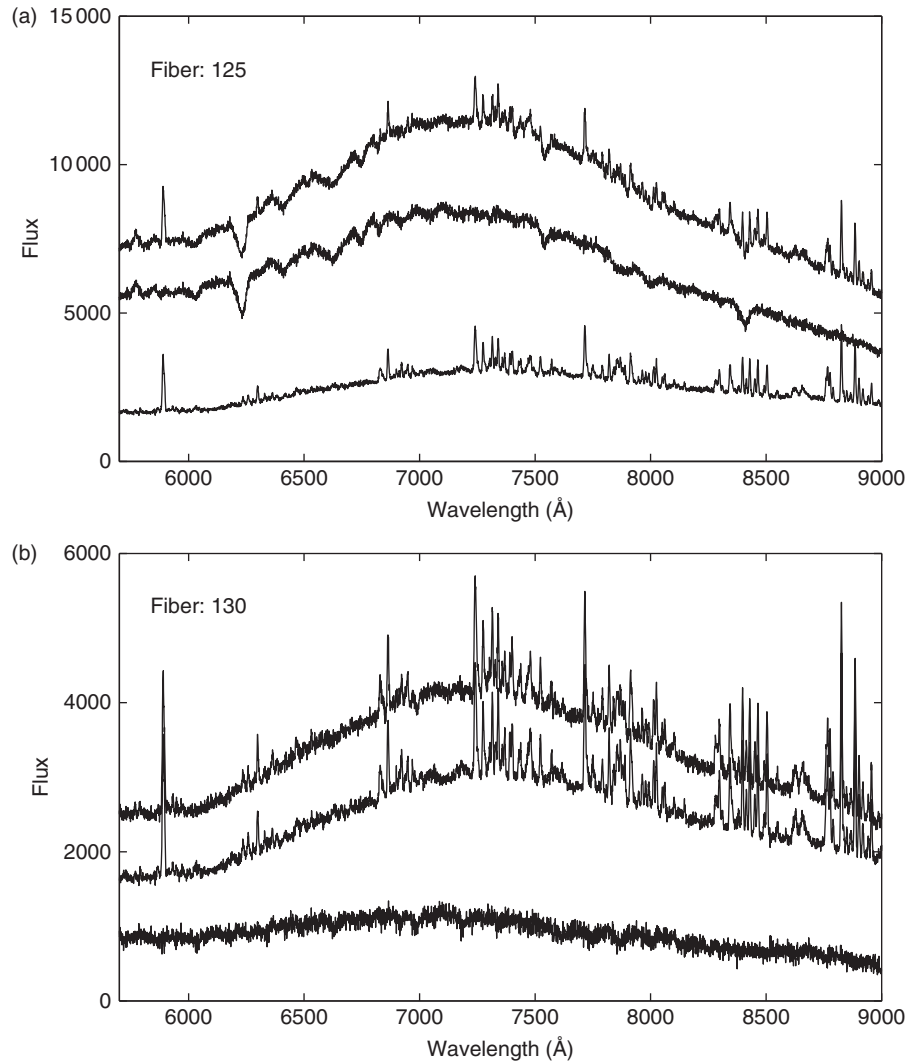
## 2.2 Sky Subtraction Using Two-Dimensional Sky-Background Modeling

The two-dimensional sky before the spectrum extraction and its slices are illustrated in Figure 1. For this two-dimensional sky spectrum, there are 15 pixels in the spatial direction ( $y = 1, \dots, 15$ ), and the full width at half-maximum ( $FWHM$ ) is approximately 8 pixels.

From these slices shown in Figure 1, we can see that the characteristics of flux are not exactly the same for different slices because of the noise. These characteristics can be adopted to describe and model the sky more accurately. Modeling the two-dimensional sky spectrum and performing the sky subtraction operation *before* the spectrum extraction can improve the accuracy of the sky subtraction, as more information about the sky is taken advantage of and the influence caused by the spectrum-extraction processing is avoided.

The flow of the two-dimensional fiber spectrum data processing was adjusted slightly. After preprocessing the observed spectral data, including bias subtraction, cosmic-ray rejection, wavelength calibration, efficiency correction of the fiber, and so on, we can obtain 250 two-dimensional spectra for each spectrograph which have not yet undergone spectral extraction. 10 of them are two-dimensional skylight spectra and other 240 are two-dimensional object spectra. Each sky fiber has a different dispersion curve and we process the fiber along the wavelength orientation.

For each two-dimensional sky spectrum  $sky_i(x, y)$  ( $i = 1, \dots, 10$ ), there are several one-dimensional slices along the wavelength orientation for the different spatial components  $y$ . Each slice has 4096 pixels and each pixel corresponds to a wavelength value according to the dispersion curves. For the slices with the same spatial



**Figure 2** Two groups of spectra obtained by the proposed method in the red channel. (a) Top: the initial spectrum (fiber 125); middle: the spectrum after sky subtraction; bottom: the reconstructed sky spectrum. (b) Top: the initial spectrum (fiber 130); middle: the reconstructed sky spectrum; bottom: the spectrum after sky subtraction.

component  $y$  of the ten observed sky spectra, all flux of sky fibers are reordered according to the wavelength. Hence, a new two-dimensional sky spectrum  $sky'_y(x, y)$  is obtained, of which there are 15 pixels in spatial orientation, 40 960 wavelength values in wavelength orientation, and the sampling rate has increased ten times.

Then, for each slice with different spatial components  $y$  of this new two-dimensional sky spectrum  $sky'_y(x, y)$ , the B-spline curve fitting method is applied to obtain a one-dimensional overall sky spectrum  $sky''_y(x)$ . The B-spline function is described as follow

$$sky''_y(x) = \sum_{k=0}^n P_{k,y} B_{k,d}(x), \quad x_{\min} \leq x \leq x_{\max} \quad (4)$$

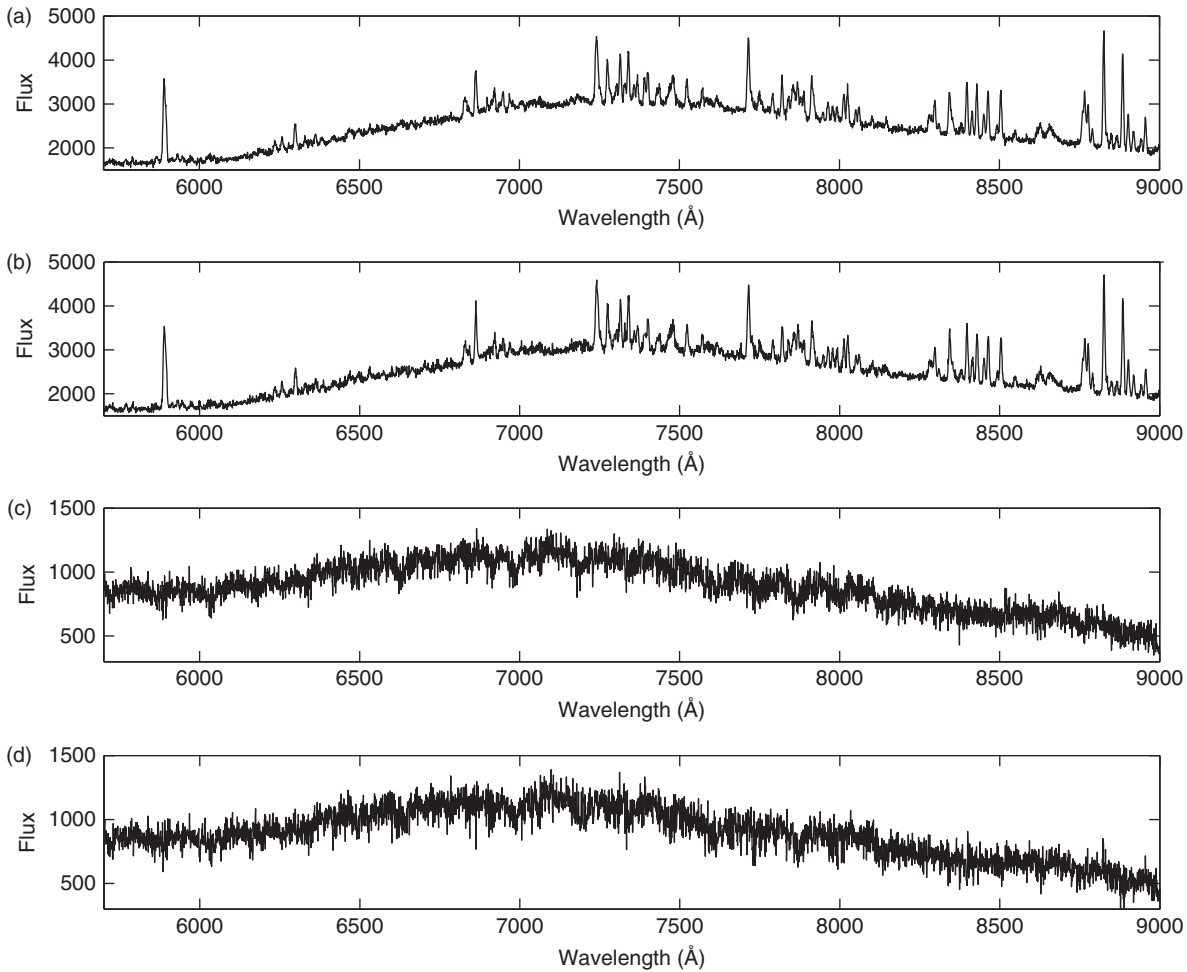
where  $x$  is the value of wavelength,  $x_{\min}$ ,  $x_{\max}$  are the minimum and maximum wavelength value respectively.  $P_{k,y}$  denotes the fitting parameters for different spatial components  $y$ .

The two-dimensional overall sky spectrum  $sky''(x, y)$  can be acquired by combining these one-dimensional overall sky spectra  $sky''_y(x)$ . After processing, the new sky subtraction processing can be extracted by subtracting the two-dimensional overall sky spectrum from each two-dimensional object spectrum. The spectrum extraction and other processing can be performed subsequently.

### 2.3 Discussion

With the design of LAMOST, there are only 10 sky fibers per spectrograph. The sky sampling fibers are actually limited, which present a great challenge to the sky subtraction for LAMOST. For each slice  $sky'_y(x)$  with different spatial components  $y$  of the two-dimensional sky spectrum  $sky'_y(x, y)$ , it consists of the sky signal  $s_y(x)$  and noise  $n_y(x)$ ,

$$sky'_y(x) = s_y(x) + n_y(x). \quad (5)$$



**Figure 3** The results of the proposed method and the LAMOST reduction in the red channel (fiber 130). (a) The reconstructed sky spectrum of the proposed method. (b) The reconstructed sky spectrum of the LAMOST reduction. (c) The spectrum after sky subtraction of the proposed method. (d) The spectrum after sky subtraction of the LAMOST reduction.

In the currently used approach for LAMOST, sky subtraction is performed after the spectrum extraction. Taking the weighted aperture extraction method as an example, the final sky spectrum can be described as

$$\begin{aligned}\hat{S}_1(x) &= BS \left[ \sum_{y=1}^{15} w_y s_k y'_y(x) \right] \\ &= BS \left[ \sum_{y=1}^{15} w_y s_y(x) + \sum_{y=1}^{15} w_y n_y(x) \right],\end{aligned}\quad (6)$$

where  $BS[\cdot]$  denotes the process of the B-spline curve fitting, and  $w_y$  denotes the weight of the aperture extraction method.

In our approach, the sky subtraction is performed before the spectrum extraction. The final sky spectrum can be described as

$$\begin{aligned}\hat{S}_2(x) &= \sum_{y=1}^{15} w_y BS[s_k y'_y(x)] \\ &= \sum_{y=1}^{15} w_y BS[s_y(x) + n_y(x)].\end{aligned}\quad (7)$$

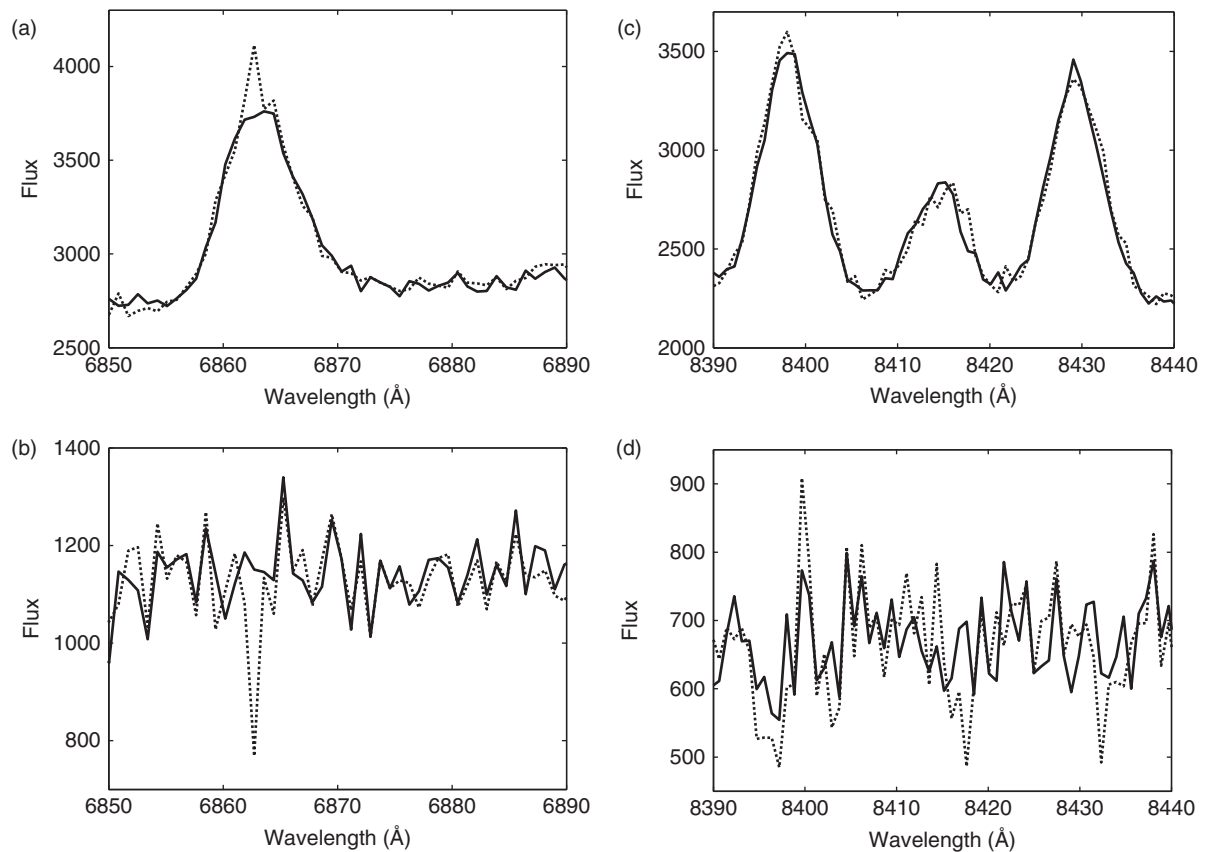
For different spatial components  $y$ , the B-spline fitting parameter  $P_{k,y}$  is different. The noise affects the effect of the B-spline curve fitting. When we perform the B-spline fitting of Equation 6, the noise,

$$n(x) = \sum_{y=1}^{15} w_y n_y(x),\quad (8)$$

is the accumulation of each spatial component. In contrast, there are 15 different implementations of the B-spline fitting in Equation 7, and for each fitting, the noise  $n_y(x)$  is weaker than from Equation 8. Therefore, with our approach, the effect of each B-spline fitting can be more accurate, and the accuracy of the sky subtraction can be improved.

### 3 Results and Analysis

In this section, the model of the simulated data is based on the LAMOST project. The data model is designed strictly in accordance with the characteristics of LAMOST. The PSF is a Gaussian model and the FWHM is approximately 8 pixels. The noise of the simulated signals is mainly



**Figure 4** Comparison between the proposed method and the LAMOST reduction in the red channel (fiber 130). The continuous lines denote the results of proposed method and the dotted lines denote the LAMOST reduction. (a) and (c) are the details of reconstructed sky spectrum; (b) and (d) are the corresponding details of the spectrum after sky subtraction.

Poisson distributed. We simulated 250 fibers, 10 of which are used for the sky. The spacing of two adjacent fiber centers is about 15 pixels. The sky spectrum is composed of the continuum and emission lines. The emission lines are simulated by using the existing data while the continuum is calculated according to the empirical formula.

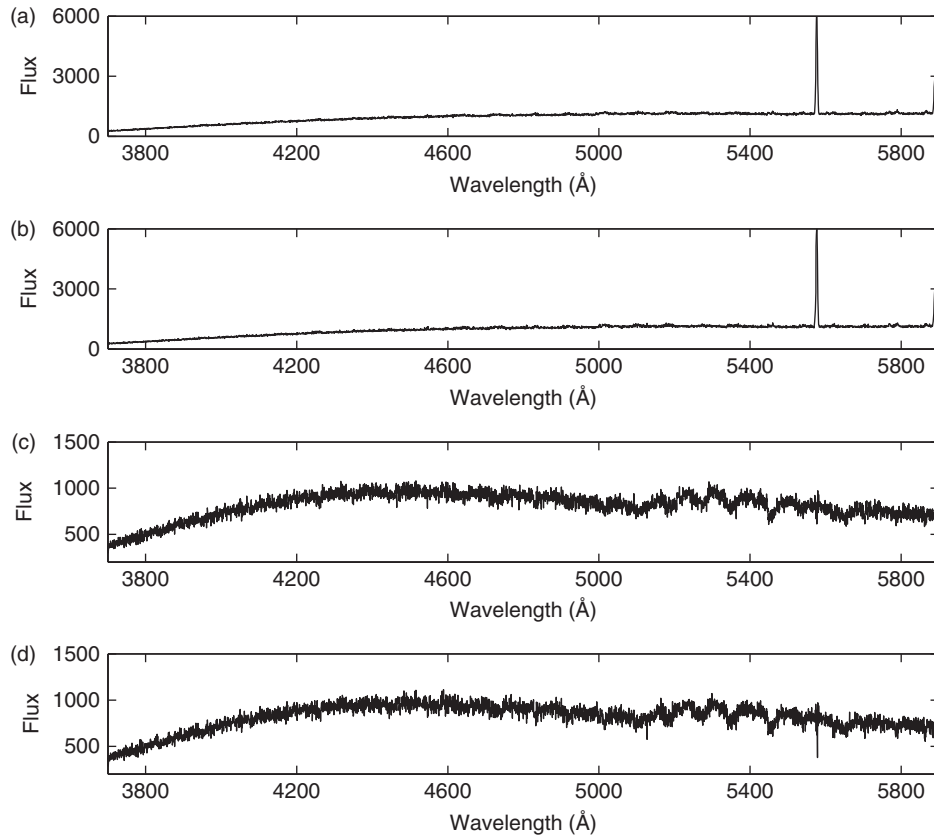
The experiments and analysis in this article are focused on the red channel because the sky emission lines in the red channel are concentrated and strong, and the sky residual errors in this wave band are usually large and difficult to eliminate. In the wavelength orientation of the fiber spectrum, there are 4096 pixels and the wavelength is in a range of 5672.58 Å to 9084.53 Å for the red channel and 3624.98 Å to 5947.67 Å for the blue channel.

Before the experiments, the spectrum data need to be preprocessed first, such as wavelength calibration, determining the centers of the fiber profiles, efficiency correction of the fiber, and so on. All flux of the ten sky fibers are reordered according to the wavelength to obtain a new two-dimensional sky spectrum. The cubic B-spline curve fitting is adopted for the experiments. For each slice with different spatial components of this spectrum, there are 40960 wavelength values which can be used as the control points to figure out the fitting parameters  $P_{k,y}$  for the B-spline curve fitting. The spectrum extraction is performed subsequently to obtain the one-dimensional spectra.

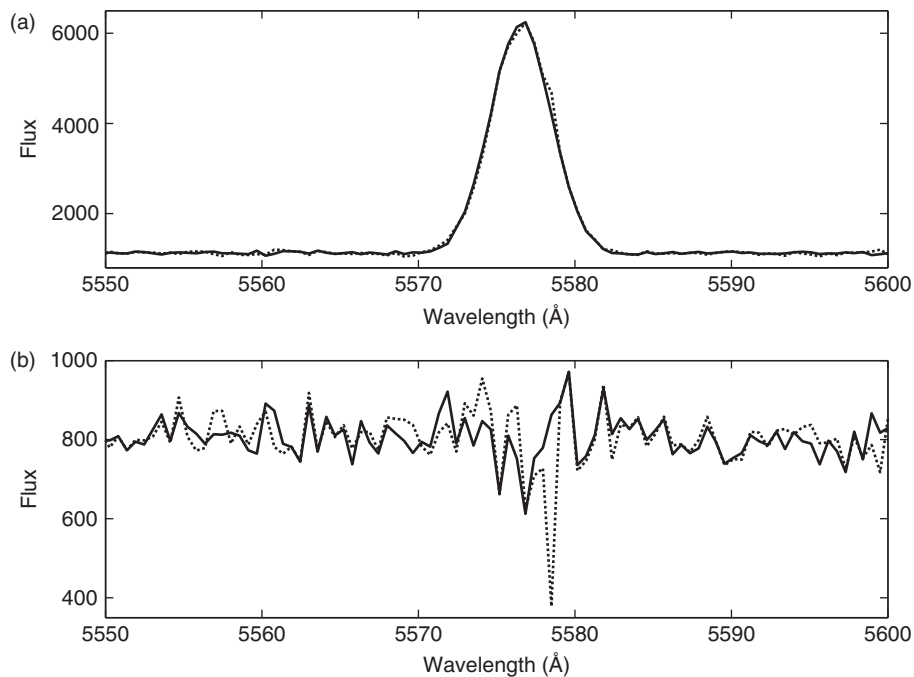
The widely used sky subtraction method based on the B-spline curve fitting, which is also adopted by the LAMOST and SDSS, is applied as the contrast experiment. The obtained one-dimensional object spectrum, which has 4096 pixels after the proposed sky subtraction processing and spectrum extraction, is used to demonstrate the experimental results and compare with the LAMOST reduction results.

Two groups of the results obtained by the proposed method are shown in Figure 2. They are one-dimensional spectra which have been performed spectrum extraction processing. We can see that the sky components in object spectra can be obtained effectively by the proposed method, and only a small residual amount of sky is left in the object spectra after the sky subtraction.

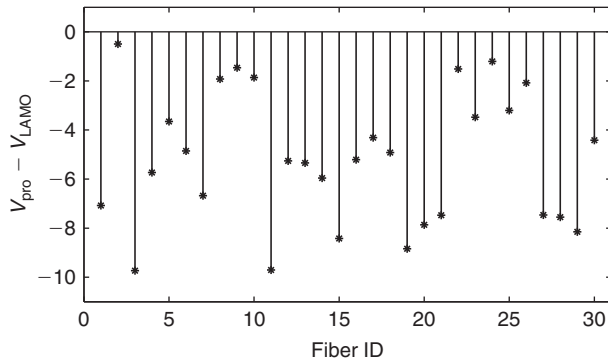
Two groups of contrast experiments between the proposed method and the LAMOST method are displayed in Figures 3–6. Figure 3 displays the comparison in the red channel (130th fiber), and Figure 4 shows some details of this comparison. Figure 5 displays the comparison in the blue channel (130th fiber), and Figure 6 shows the detail of this comparison. In Figure 4 and Figure 6, the continuous lines denote the results of proposed method and the dotted lines denote the LAMOST reduction. The magnitudes and number of the residual spikes in the spectrum obtained by the proposed method are smaller than that in the spectrum obtained by the LAMOST method. This



**Figure 5** The results of the proposed method and the LAMOST reduction in the blue channel (fiber 130). (a) The reconstructed sky spectrum of the proposed method. (b) The reconstructed sky spectrum of the LAMOST reduction. (c) The spectrum after sky subtraction of the proposed method. (d) The spectrum after sky subtraction of the LAMOST reduction.



**Figure 6** Comparison between the proposed method and the LAMOST reduction in the blue channel (fiber 130). The continuous lines denote the results of proposed method and the dotted lines denote the LAMOST reduction. (a) is the detail of reconstructed sky spectrum; (b) is the corresponding detail of the spectrum after sky subtraction.



**Figure 7** The differences between the variance of results obtained by the proposed method and the variance of the LAMOST reduction results (i.e.  $V_{\text{pro}} - V_{\text{LAMO}}$ ).

demonstrates that the proposed method gives a more effective sky subtraction than the LAMOST method.

Because the original object spectrum is smooth at the wave band without any characteristic lines, the performance of the sky subtraction algorithm can be evaluated by analyzing the magnitudes and number of the residual spikes in the spectrum. Therefore, we can adopt the variance to evaluate the effectiveness of the proposed sky subtraction algorithm in this section. The variance can reflect the smoothness of the spectrum: a smaller variance indicates a smoother spectrum, and a more thorough sky subtraction processing.

The variance is defined by

$$V = \frac{1}{N} \sum_{k=1}^N \left[ \text{obj}(k) - \frac{1}{N} \sum_{k=1}^N \text{obj}(k) \right]^2, \quad (9)$$

where  $\text{obj}(k)$  denote the flux of the obtained object spectrum and  $N$  is the length of the spectrum.

The differences between the variance of results obtained by the proposed method and the variance of the LAMOST results (i.e.  $V_{\text{pro}} - V_{\text{LAMO}}$ ) are illustrated in Figure 7. The figure displays the differences of the variances for 30 object-fiber spectra. As we can see from Figure 7, the proposed method has smaller variances than the LAMOST reduction pipeline. A smaller variance indicates a smoother spectrum, and a more thorough sky subtraction processing.

## 4 Conclusion

A novel sky subtraction method is proposed based on two-dimensional sky background modeling and B-spline curve fitting for LAMOST. The two-dimensional sky spectrum before the spectrum extraction has a lot of information and characteristics that can be applied for modeling an overall two-dimensional sky background model. Sky subtraction can now be performed as an earlier step, just before the spectrum extraction, that can improve the accuracy of the sky subtraction by taking advantage of more information about the sky and avoiding the influence caused by the spectrum extraction processing. Experiments based on the LAMOST project are presented to demonstrate the effectiveness of the proposed method. From the experimental results, we can see that the proposed method can obtain the original object spectrum properly and give a more effective sky subtraction than the currently used method for LAMOST.

## Acknowledgments

The authors acknowledge support from the National Natural Science Foundation of China (NSFC 11078016).

## References

- Abazajian, K., et al., 2004, *ApJ*, 128, 502
- Bai, Z. R., Zhang, L. & Ye, Z. F., 2008, *ChA&A*, 32, 109
- Barden, S. C., Elston, R., Armandroff, T. & Pryor, C. P., 1993, *ASPC*, 37, 223
- Bolton, A. S. & Burles, S., 2007, *NJPh*, 9, 443
- Causi, G. L. & De Luca, M., 2005, *NA*, 11, 81
- Cuby, J. & Mignoli, M., 1994, *SPIE*, 2198, 98
- Cui, X. Q., 2009, *AAS*, 213, 226
- de Boor, C., 1978, *A Practical Guide to Splines* (New York: Springer)
- Elston, R. & Barden, S., 1989, *NOAO Newsletter*, 19, 21
- Gordon, W. & Riesenfeld, R., 1974, in *Computer Aided Geometric Design*, ed. R. Barnhill & R. Riesenfeld (New York: Academic Press), 95
- Kurtz, M. J. & Mink, D. J., 2000, *ApJL*, 533, L183
- Parry, I. & Carrasco, E., 1990, *SPIE*, 1235, 702
- Sharp, R. & Parkinson, H., 2010, *MNRAS*, 408, 2495
- Watson, F. G., 1987, PhD Thesis, University of Edinburgh
- Wild, V. & Hewett, P. C., 2005, *MNRAS*, 358, 1083
- Wynne, C., 1993, *MNRAS*, 260, 307
- Wyse, R. & Gilmore, G., 1992, *MNRAS*, 257, 1
- Zhang, H., Chu, Y. & Chen, J., 2002, *ChA&A*, 26, 97
- Zhu, Y. T., Hu, Z. W., Zhang, Q. F., Wang, L. & Wang, J. N., 2006, *SPIE*, 6269, 20