

A New Method of Wavelength Calibration for LAMOST by Combining Short- and Long-Exposure Spectral Lines

G. H. Ye^A, J. Zhu^A, and Z. F. Ye^{A,B}

^AInstitute of Statistical Signal Processing, Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, P. R. China

^BCorresponding author. Email: yezf@ustc.edu.cn

Abstract: In wavelength calibration using arc lines, the normal approach is to use the strongest unsaturated lines, leaving weak lines unused. A new method is proposed in this paper, which not only utilizes the strong spectral lines, but also makes most use of weak spectral lines. In order to validate the effectiveness of the method we propose, experiments are performed on simulated spectra. Firstly, two kinds of spectra are generated: one with a short exposure and another with a long exposure. Secondly, calibration lines are chosen from the short exposure and long exposure spectra separately according to some rules. Thirdly, the initial wavelength calibration is completed by using the selected short-exposure lines. Fourthly, the approximate centroids of the selected long-exposure lines are obtained by utilizing the result of the initial wavelength calibration. These are then adjusted iteratively to obtain the centroids. Finally, the selected lines from the short- and long-exposures are combined to obtain the final wavelength calibration. Compared with traditional calibration methods which only use short exposures and strong lines, the proposed method is shown to be more accurate.

Keywords: instrumentation: spectrographs — methods: data analysis — techniques: spectroscopic — lines: identification

Received 2011 September 13, accepted 2011 December 11, published online 2012 January 16

1 Introduction

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST), is a quasi-meridian reflecting Schmidt telescope with 4-m aperture, 20-m focal length and 5-deg field of view (Su et al. 1998). The light of celestial objects is reflected by the Schmidt corrector MA and again reflected by the spherical mirror MB, then collected by optical fibers installed at the focal plane. The focal plane can accommodate as many as 4000 optical fibers, which are distributed, on average, to 16 multi-object fiber spectrographs. In each spectrograph, the spectra from fibers are divided into blue (3700–5900 Å) and red (5700–9000 Å) channels by a dichroic filter, and finally focused onto two Charge Coupled Devices (CCDs), respectively. The main parameters of spectra observed by LAMOST (low resolution mode, full slit) are listed in Table 1. In Table 1 ‘Number of Points’ refers to the number of pixels along the dispersion direction and ‘Separation’ refers to the spatial distance between adjacent spectra on the CCD detector. A more detailed description of the instrument can be found in Zhu et al. (2006).

Wavelength calibration is a very important step in the reduction procedure. It aims to establish a relationship between the CCD pixel position and wavelength. Some calibration methods have been reported, such as the linear

regression method (Lo & Fountain 2006), the interpolation method (Bai et al. 2004), and the polynomial fitting method (Sanchez 2006; van Geffen et al. 2003). The accuracy of wavelength calibration has a direct influence on the subsequent data reductions, such as sky subtraction and redshift determination. It is therefore necessary to calibrate the wavelength with the highest possible accuracy.

Mostly, arc spectra are selected for wavelength calibration. The arc lamp, such as the mercury vapor and neon lamp used in LAMOST, is placed in front of the focal plane some distance away. The longer the exposure time is, the stronger the intensities of the spectra will be.

Usually, the lines with comparatively strong intensities are chosen for wavelength calibration. The weak lines are not used because they have poor signal-to-noise (S/N) ratio. To some extent, this method works well, but does not provide the best possible wavelength calibration for LAMOST. Lines of high and moderate intensities are usually poorly distributed throughout the whole wavelength range, and in some regions there is an insufficient number of these lines, though many weak lines may be present. The precision of the calibration can be improved by using the weak lines. The weak lines can be measured by using a longer exposure to improve the S/N ratio.

Some wavelength calibration methods (Qin et al. 2010; Wang, Qin & Ye 2010; Xue et al. 2009) have been

Table 1. The main parameters of spectra observed by LAMOST^a

Parameter	Blue channel	Red channel
Wavelength range (Å)	3700–5900	5700–9000
Number of points	4096	4096
Dispersion (Å per pixel)	0.57	0.85
FWHM (pixel)	8	8
Separation (pixel)	15	15

^aLow-resolution mode, full slit.

researched for LAMOST. In this paper we will present another new method. As the calibration using intensity modelling (Balona 2010), the proposed method also eliminates the necessity of different lamps. In practice, it utilizes only one kind of arc lamp. By controlling the exposure time, we can obtain spectra of different intensities. In our case we are interested in a typical exposure, as used in traditional calibration methods. We shall refer to such an exposure as a ‘short exposure’ since we will be also a long exposure to increase the S/N ratio of the weak lines. Most arc lines in the long-exposure spectra have high S/N ratio, so the weak lines blended with noise in the short-exposure spectra now can be used in the long-exposure spectra. By combining these two exposures, the proposed method can utilize more arc lines, and the relationship between CCD pixel position and wavelength can also be better defined. As the spectra from fibers are separated into blue and red channels, wavelength calibration is performed on the two channels separately.

The rest of the paper is organized as follows. Section 2 describes the generation of simulated spectra, and explains how to choose the calibration lines. The new wavelength calibration method will be described in Section 3. Section 4 gives the experimental results and analysis. Section 5 concludes the paper.

2 Simulated Spectra

2.1 Generation of Simulated Spectra

In this paper, we take the red channel for example, and the blue channel can be performed in the same way.

We simulate the spectra based on a first approximation to the real dispersive curves and the known laboratory wavelengths. In this paper, the first approximation to the real dispersive curves are regarded as the standardized dispersive curves, given by a previous experienced wavelength calibration. The laboratory wavelengths are listed in Table 2.

Given the dispersive curves and the laboratory wavelengths, we can simulate the arc spectra. Firstly, the pixel positions corresponding to the laboratory wavelengths are calculated using a nominal dispersion curve. However, the calculated pixel positions are generally non-integers, so we select the two nearest pixels close to the calculated values. The intensities at these two pixels are obtained from the relative intensities of laboratory wavelengths

Table 2. The spectral lines of the arc lamp in the red channel

Wavelength (Å)	Intensity ^a	Wavelength (Å)	Intensity ^a
5769.60	1.5	7514.65	1.0
5790.66	1.5	7535.77	7.0
5852.49	4.0	7544.04	3.0
5881.90	2.0	7635.11	10.0
5944.83	3.0	7723.76	2.0
5975.53	0.2	7724.21	2.0
5974.63	0.2	7948.18	2.5
6023.00	1.0	8006.16	1.0
6074.34	3.0	8014.79	2.0
6096.16	5.0	8103.69	3.0
6143.06	8.0	8115.31	13.0
6163.60	3.0	8118.55	4.0
6217.28	2.0	8128.91	1.0
6266.50	4.0	8136.41	1.0
6304.79	1.5	8264.52	3.0
6334.43	6.0	8300.33	3.0
6382.99	7.0	8377.61	22.0
6402.25	21.0	8408.21	2.0
6506.53	9.0	8417.16	3.0
6532.88	4.0	8418.43	3.0
6598.95	1.2	8424.65	3.0
6678.28	4.0	8495.46	10.0
6678.15	4.0	8521.44	1.0
6717.04	3.5	8591.26	4.0
6929.47	6.0	8634.65	4.0
7024.05	0.1	8654.38	8.0
7032.41	6.0	8679.49	1.0
7173.94	2.0	8681.92	1.0
7245.17	6.0	8780.62	2.5
7383.98	1.0	8783.75	2.5
7438.90	6.0	8853.87	4.0
7488.87	0.7	8865.31	0.5
7503.87	1.0	8865.76	1.0

^aThe intensities in the table are theoretic relative values.

allotting proportionally. The allotment equations are defined as follows:

$$IN_{x_1} = \frac{\lambda_{x_2} - \lambda_x}{\lambda_{x_2} - \lambda_{x_1}} \times IN_x \tag{1}$$

and

$$IN_{x_2} = \frac{\lambda_x - \lambda_{x_1}}{\lambda_{x_2} - \lambda_{x_1}} \times IN_x, \tag{2}$$

where x is the non-integer pixel position corresponding to the laboratory wavelength λ_x , x_1 and x_2 are the two nearest pixels close to calculated value x ($x_1 < x < x_2$). λ_{x_1} and λ_{x_2} are the wavelengths corresponding to x_1 and x_2 ($\lambda_{x_1} < \lambda_x < \lambda_{x_2}$), they can be obtained from the standardized dispersion curve. IN_{x_1} , IN_x and IN_{x_2} are the relative intensities corresponding to λ_{x_1} , λ_x and λ_{x_2} separately.

Finally, we convolve the intensity at each pixel with a Gaussian kernel to create the simulated arc lines.

The arc line intensities will increase as the exposure time increases. In order to obtain the simulated spectra for the short exposure and the long exposure, we need to

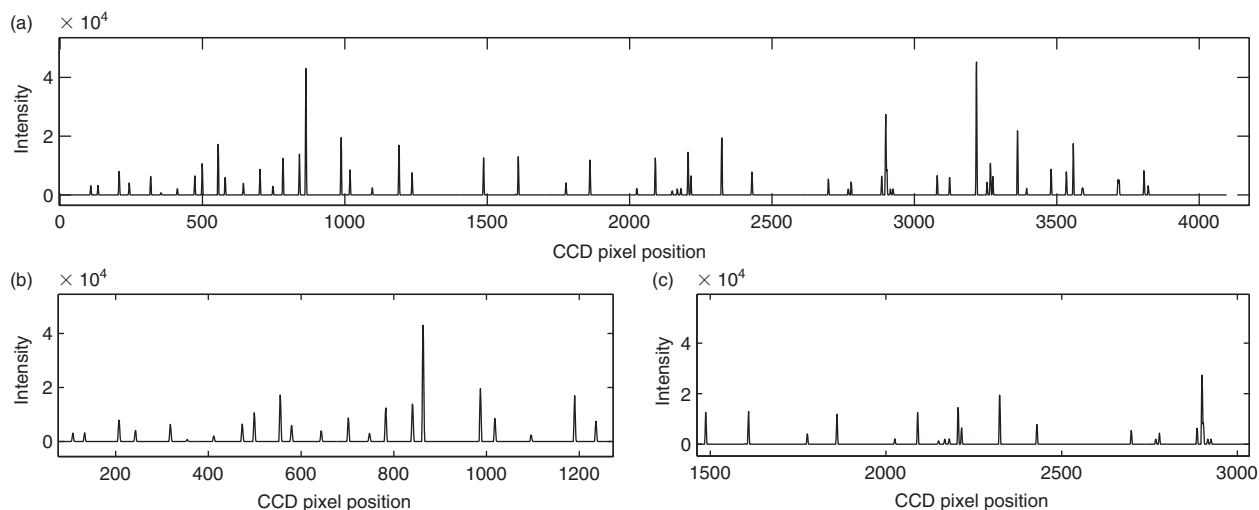


Figure 1 The short-exposure spectrum of the 111th fiber in the red channel. (a) is the entire spectrum, while (b) and (c) are parts of the spectrum in some small ranges.

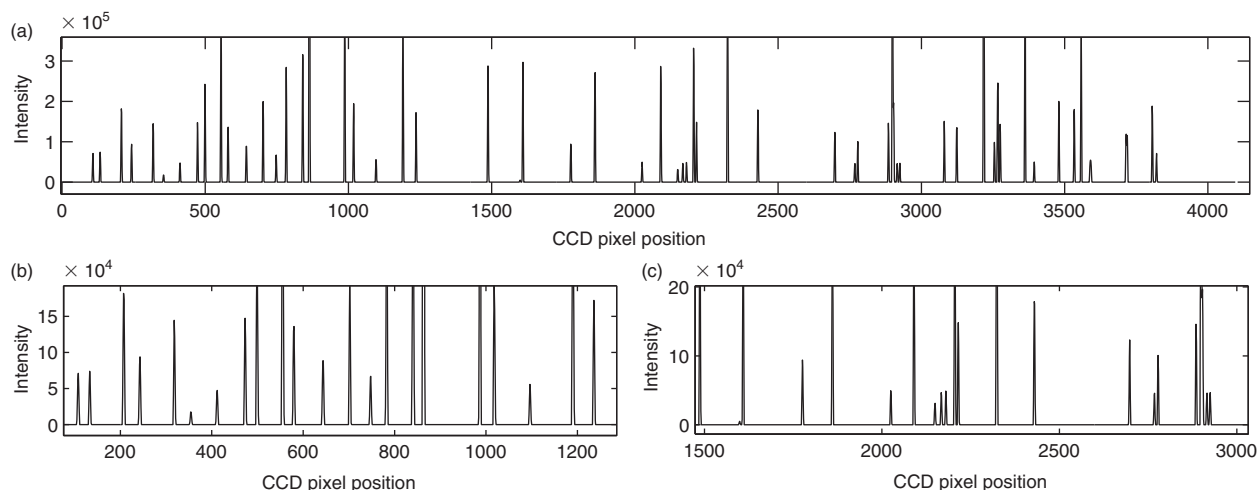


Figure 2 The long-exposure spectrum of the 111th fiber in the red channel. (a) shows the entire spectrum, while (b) and (c) show parts of the spectrum in some small ranges.

multiply the relative intensities listed in Table 2 by two different factors. One factor is used to simulate the short-exposure spectra, and the other for the long-exposure spectra. The factors are related to the exposure-time. The longer the exposure time is, the larger the factor will be. The calculated intensities are utilized as the final relative intensities.

Finally, we obtain the short-exposure and long-exposure spectra for each channel. In order to compare the difference between these two spectra, we select the spectra of the 111th fiber in the red channel as an illustration. The short exposure spectrum and the long-exposure spectrum are shown in Figure 1 and Figure 2, respectively. They are noise-free spectra.

From these two figures, we know that the arc line intensities of the long-exposure spectrum are larger than those of the short-exposure spectrum. Compared with the weak lines with low S/N ratio in the short-exposure

spectrum, the same lines in the long-exposure spectrum have high S/N ratio. Therefore, these lines in the long-exposure spectrum can be used to improve the wavelength calibration.

2.2 Choosing Calibration Lines

Before choosing which arc lines are to be used for the calibration, some considerations need to be borne in mind. Firstly, if the lines are blended they need to be rejected. Secondly, some laboratory wavelengths are nearly the same, in which case the corresponding spectral lines nearly coincide with each other. These lines, which can be determined from the wavelength table, should not be used.

However, the actual spectra contain noise, which restricts the lines that can be used for calibration. In the short-exposure spectra, the lines with comparatively strong intensities have high S/N ratio, while other lines

Table 3. The selected calibration lines in the red channel

Wavelength (Å)	Wavelength (Å)	Wavelength (Å)
5852.49	7245.17	5769.60 ^a
5944.83	7438.90	5790.66 ^a
6074.34	7535.77	5881.90 ^a
6096.16	7635.11	6023.00 ^a
6143.06	7948.18	6217.28 ^a
6163.60	8103.69	6304.79 ^a
6266.50	8264.52	6598.95 ^a
6334.43	8300.33	7173.94 ^a
6382.99	8377.61	7383.98 ^a
6402.25	8495.46	7488.87 ^a
6506.53	8591.26	8014.79 ^a
6532.88	8634.65	8408.21 ^a
6717.04	8654.38	8521.44 ^a
6929.47	8853.87	
7032.41		

^aThis label denotes the line is the LE calibration line, while without this label, the line is the SE calibration line.

are almost invisible due to noise. In the long-exposure spectra, lines which are almost invisible in the short-exposure spectra are easily distinguished from noise, while the intensities of some strong lines in the short-exposure spectra may be saturated and should be rejected. It is therefore necessary to combine these two kinds of spectra to perfect calibration.

According to these guidelines, we selected 29 calibration lines from the short-exposure (SE) spectrum and a different set of 13 calibration lines from the long-exposure (LE) spectrum in the red channel. The 29 SE calibration lines and 13 LE calibration lines are listed in Table 3. In the same way, we can select SE and LE calibration lines for the blue channel.

3 Wavelength Calibration

3.1 The Wavelength-Calibration Algorithm

The initial wavelength calibration (a traditional calibration using the SE calibration lines) affects subsequent steps in our procedure. For example, the centroids of the LE calibration lines are calculated using the initial calibration from the SE exposure, while the final calibration combines lines from the SE calibration and the LE calibration. This means that it is important to use a reliable method for the initial calibration. We chose the wavelength calibration method used in the Sloan Digital Sky Survey (SDSS) (Burles, Finkbeiner & Schlegel 2008).

Our initial wavelength calibration algorithm is as follows.

1. The approximate centroids of the calibration lines in the selected spectrum are located manually. These approximate values are used as starting values and adjusted iteratively to obtain the final centroids. In our case, we began the procedure on the spectrum from fiber 111.
2. Using the centroids of the calibration lines in the selected spectrum, the centroids of the calibration lines

in the spectra of other fibers can be located. Because the distortion of contiguous spectra is slight, the centroids of the calibration lines in the selected spectrum could be considered as the approximate centroids of the calibration lines in the adjacent spectra. Then the approximate centroids are adjusted iteratively. Adopting this method, the centroids of the calibration lines in the other spectra can be located sequentially along the spatial direction.

3. After locating the centroids of calibration lines, we can fit the relationship between CCD pixel position and wavelength by a fifth-order polynomial. The function can be described as follows:

$$f_j(x) = \sum_{i=0}^5 a_{i,j} x^i, \quad j = 1, 2, \dots, 250, \quad (3)$$

where $a_{i,j}$ is the i th coefficient of the polynomial for the j th fiber, x is the CCD pixel position, and $f_j(x)$ is the fitted wavelength corresponding to x in the j th fiber.

We choose the fifth-order polynomial and then use the least-squares method to obtain the coefficients. Given the pixel positions and wavelengths, (x_i, λ_i) ($i = 1, 2, \dots, N$), the sum of the squares of the differences between $f(x_i)$ and λ_i is

$$err = \sum_{i=1}^N [f(x_i) - \lambda_i]^2, \quad (4)$$

where N is the number of the selected lines, x_i is the centroid of the i th line, $f(x_i)$ is the calibrated wavelength of the i th line, and λ_i is the corresponding laboratory wavelength.

The chosen polynomial should minimize the sum of squares of the differences.

4. Standardized dispersion curves are required before the calibration and they are used to generate the simulated spectra. Two formulas are used to evaluate the accuracy of the fitting:

$$ME_j = \frac{1}{4096} \sum_{i=0}^{4095} |f_j(i) - s_j(i)| \quad (5)$$

and

$$MSE_j = \frac{1}{4096} \sum_{i=0}^{4095} |f_j(i) - s_j(i)|^2 \quad (6)$$

where i is the CCD pixel position, $f_j(i)$ and $s_j(i)$ are the wavelengths corresponding to i in the j th fiber, they are determined by the experimental dispersive curves and the standardized dispersive curves, respectively. ME_j and MSE_j represent the mean value and mean square value of absolute errors corresponding to the j th fiber respectively. The smaller ME and MSE are, the more accurate the calibration will be.

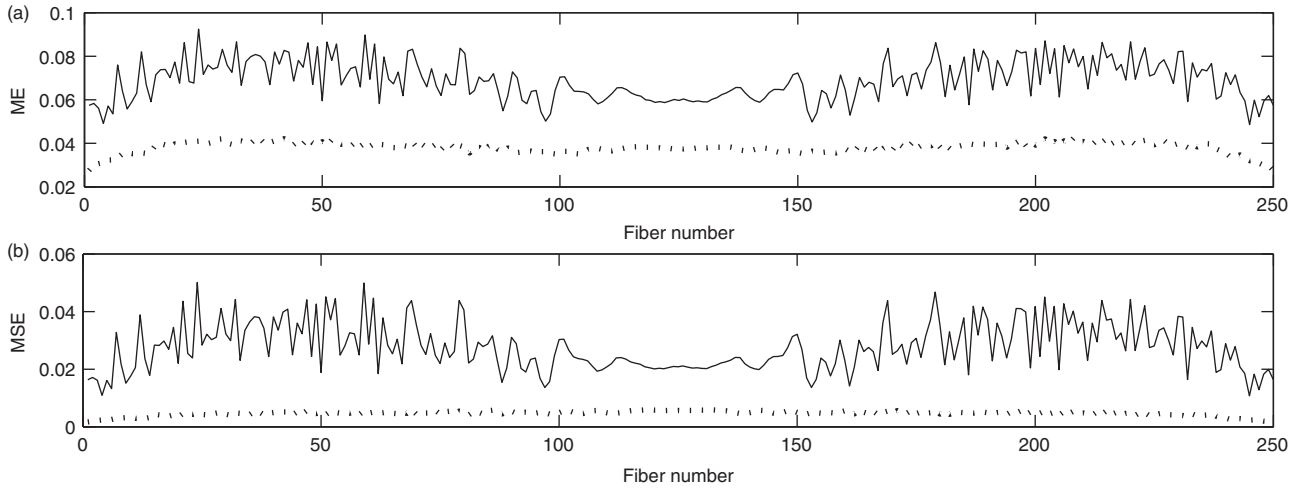


Figure 3 (a) and (b) show the mean value and mean square value of the wavelength absolute errors corresponding to each fiber in the red channel, respectively. The real lines present the result of the short-exposure calibration, while the dotted lines present the result of our proposed method.

3.2 The Initial Wavelength Calibration

In this step, the calibration lines are chosen from the short-exposure spectrum. They are listed in Table 3. Given the short-exposure spectra, the centroids of the SE calibration lines in the spectrum of each fiber are located according to step 1 and step 2 of the algorithm described in last subsection. After the centroids have been obtained, the calibration can be performed by using a fifth-order polynomial.

There are 250 fibers and 29 selected SE lines in the red channel, so a 250×29 matrix is used to store the centroids of the calibration arc lines. The coefficients of the fifth-order polynomial are recorded in a 250×6 matrix.

3.3 Proposed Wavelength Calibration

The proposed method combines the SE calibration lines and the LE calibration lines to perform calibration. Before this procedure, we should locate the centroids of the LE calibration lines first.

3.3.1 Locating the Centroids of the LE Calibration Lines

The initial functions of the relationship between CCD pixel position and wavelength have been set up by the initial calibration. They are monotone and increasing. Given the pixel position, we can calculate its corresponding wavelength, and given the wavelength, we can also locate its corresponding pixel position. As the wavelengths of the chosen LE calibration lines are listed in Table 3, the approximate centroids of the LE calibration lines can be obtained by the inverse functions of $f_j(x_i)$ as defined in Equation 3, i.e.

$$x_{i,j} = f_j^{-1}(\lambda_i) \quad i = 1, 2, \dots, M \quad j = 1, 2, \dots, 250, \quad (7)$$

where M is the number of selected LE lines, λ_i is the laboratory wavelength of the i th LE line, $f_j^{-1}(\lambda_i)$

is the inverse function of $f_j(x_i)$, $x_{i,j}$ is the approximate centroid corresponding to λ_i in the j th fiber.

After obtaining the approximate centroids, the final centroids are obtained by successive approximations until convergence is obtained. There are 250 fibers and 13 LE calibration lines in the red channel, so the centroids are stored in a 250×13 matrix.

3.3.2 The Ultimate Wavelength Calibration

In the previous steps, we obtained two matrices, a 250×29 matrix and a 250×13 matrix, which contain the centroids of SE and LE calibration lines, respectively. It is required to combine them into a single matrix.

First of all, a 250×42 matrix is defined. The centroids of the calibration lines in a given fiber are stored in a given row, while the given column contains the centroids corresponding to a given calibration line. This matrix thus contains the centroids of all the calibration lines. Then the centroids in the same row and their relevant wavelengths are fitted by a fifth-order polynomial, generating a new 250×6 coefficient matrix.

4 Results and Analysis

To test the performance of the proposed wavelength calibration method, experiments were performed on the red and blue channels, respectively. Before starting the experiments, the short-exposure simulated spectra and the long-exposure simulated spectra were generated separately. The SE and LE calibration lines were chosen according to the criteria discussed in Section 2. We then performed the calibration using the algorithm described in Section 3. The traditional calibration for LAMOST usually uses short-exposure spectra exposed for normal time. In the proposed method, we utilized not only the short-exposure spectra, but also the long-exposure spectra. The traditional calibration only by SE calibration lines was used as the comparison method in the experiments.

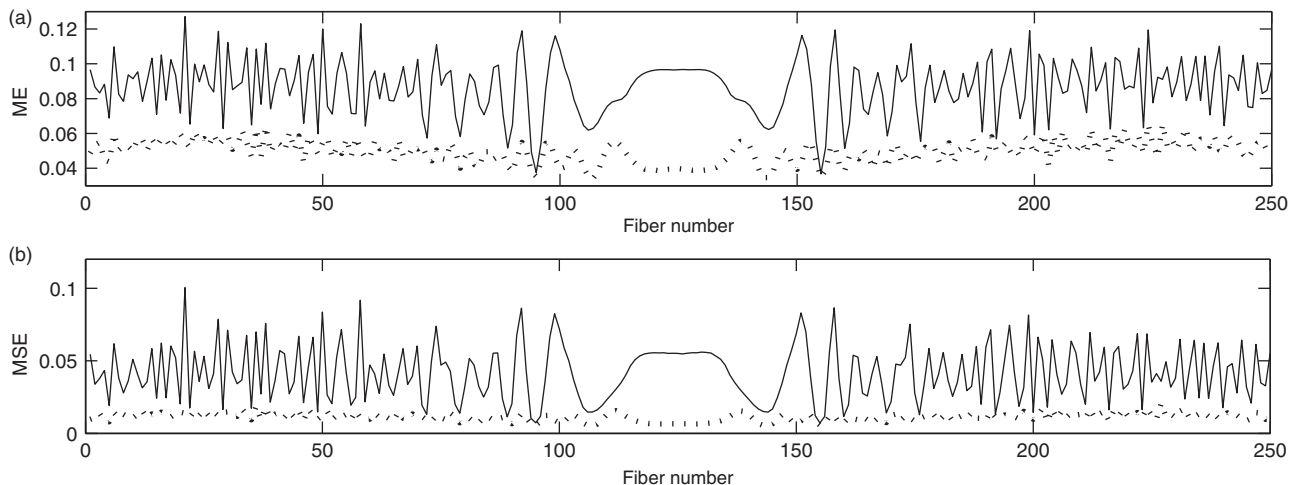


Figure 4 (a) and (b) show the mean value and mean square value of the wavelength absolute errors corresponding to each fiber in the blue channel, respectively. The real lines present the result of the short-exposure calibration, while the dotted lines present the result of our proposed method.

After calibration, we obtained the result, the new dispersion curves. The dispersion curves set up the relationship between the CCD pixel position and the wavelength. The new dispersion curves obtained in the experiments were compared with the standardized dispersion curves to get the wavelength absolute errors corresponding to each pixel. The smaller the differences between them are, the more accuracy the calibration will be. Here, we used the mean value of the wavelength absolute errors in the whole pixel range corresponding to each fiber and the mean value of the wavelength absolute errors squared as defined in Equations 5 and 6 to test the effect of the calibration. For the red channel, the results of the two methods are shown in Figure 3.

In the same way, we selected 10 SE calibration lines and 10 LE calibration lines from the spectra in the blue channel. The results of the two kinds of calibration are shown in Figure 4.

The experiments on the two channels show that the combining calibration can achieve better results than the traditional calibration. Therefore, the method proposed here is suitable for LAMOST wavelength calibration.

5 Conclusion

A new arc wavelength calibration method is presented in the paper. Compared with the traditional calibration, the

proposed method utilizes not only the short-exposure spectra, but also the long-exposure spectra. Some weak lines with low S/N ratio in the short-exposure spectra become usable in the long-exposure spectra, thus the new method is able to use more arc lines to perform calibration. The results of experiments show that the proposed method leads to a calibration of higher precision.

Acknowledgments

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

References

- Bai, L., Liao, N. F., Li, Z. J. & Yang, W. P., 2004, *ChOpL*, 2, 174
- Balona, L. A., 2010, *MNRAS*, 409, 1601
- Burles, S., Finkbeiner, D. & Schlegel, D., 2008, available at http://das.sdss.org/software/idlspec2d/v5_3_12/pro/spec2d/
- Lo, E. & Fountain, A. W., 2006, *SPIE*, 6233, 62330L
- Qin, H. Q., Zhu, J., Zhu, Z. Q., Ye, Z. F. & Luo, A. L., 2010, *PASA*, 27, 265
- Sanchez, S. F., 2006, *AN*, 327, 850
- Su, D. Q., Cui, Y. Q., Wang, Y. N. & Yao, Z. Y., 1998, *SPIE*, 3352, 76
- Van Geffen, J. H. G. M. & Van Oss, R. F., 2003, *ApOpt*, 42, 2739
- Wang, S., Qin, H. Q. & Ye, Z. F., 2010, *EA*, 28, 195
- Xue, X. L. & Ye, Z. F., 2009, *AR&T*, 6, 181
- Zhu, Y. T., Hu, Z. W., Zhang, Q. F., Wang, L. & Wang, J. N., 2006, *SPIE*, 6269, 62690M