

## BFOSC Long-Slit Spectroscopy Automated Processing Software Postprint

**Authors:** Song Deyang, Junbo Zhang, Wang Liang, Wang Liang

**Date:** 2023-11-24T00:00:00+00:00

### Abstract

In astronomical observations, long-slit spectrographs are commonly used spectral observation instruments. The long-slit spectroscopy mode of the Beijing Faint Object Spectrograph and Camera (BFOSC), installed on the 2.16 m telescope at the Xinglong Observatory of the National Astronomical Observatories, is widely used but lacks dedicated data processing software. Typically, users process their observations interactively using IRAF (Image Reduction and Analysis Facility) software. This paper introduces an automated data processing software specifically designed for BFOSC long-slit spectra; the data processing procedure includes bias and flat-field correction, background correction, spectral extraction, wavelength calibration, and other steps. Results from processing actual observational data are presented and compared with those obtained using IRAF software. The software achieves fully automated operation while completing all essential steps required for spectral processing, employs an optimized spectral extraction method, can automatically remove most cosmic rays, and ensures high-precision output results.

### Full Text

## BFOSC Long-slit Spectrum Automatic Extraction Pipeline

**Song Deyang<sup>1,2,3</sup>, Zhang Junbo<sup>4</sup>, Wang Liang<sup>1,2,3,\*</sup>**

<sup>1</sup>Nanjing Institute of Astronomical Optics & Technology, Chinese Academy of Sciences, Nanjing 210042, China; Email: [lwang@niaot.ac.cn](mailto:lwang@niaot.ac.cn)

<sup>2</sup>CAS Key Laboratory of Astronomical Optics & Technology, Nanjing Institute of Astronomical Optics & Technology, Nanjing 210042, China

<sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>4</sup>National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

<sup>5</sup>Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

---

## Abstract

In astronomical observations, the long-slit spectrograph is a commonly used spectral observation instrument. The long-slit spectral mode of the Beijing Faint Object Spectrograph and Camera (BFOSC) mounted on the 2.16 m telescope at the Xinglong Observatory of the National Astronomical Observatories, Chinese Academy of Sciences, is widely used, but it is not equipped with dedicated data processing software. After observations, users typically process the data interactively using IRAF software. This paper introduces an automatic data processing software specifically designed for BFOSC long-slit spectra. The data processing procedure includes bias and flat-field correction, background subtraction, spectrum extraction, and wavelength calibration. We present the results of processing observed data and compare them with those obtained using IRAF software. This software achieves automated operation while completing all necessary steps for spectral processing, employs an optimal extraction method, automatically removes most cosmic rays, and ensures high-precision output results.

**Keywords:** long-slit spectrum; optimized extraction method; automatic pipeline

---

## 1. Introduction

The 2.16 m telescope at the Xinglong Observatory of the National Astronomical Observatories is China's first independently developed large astronomical telescope. It is equipped with the Beijing Faint Object Spectrograph and Camera (BFOSC), which supports multiple observation modes, among which long-slit spectroscopy is the most widely used. BFOSC can switch between different slit widths, gratings, and filter combinations. According to publicly available statistics, the 2.16 m telescope allocated 20-30% of its observation time to BFOSC long-slit spectroscopy between 2016 and 2021. While photometric mode data can be processed using programs developed by the Observatory, no dedicated software exists for long-slit spectroscopy. Observers typically use IRAF software interactively to process raw spectral images, but IRAF is not specifically designed for BFOSC and lacks optimization for long-slit spectroscopy features. Moreover, the National Optical Astronomy Observatory (NOAO) announced in 2020 that it would no longer officially support IRAF.

To address this gap, we have developed an automatic spectral data processing software for BFOSC that enables batch data processing. The software incorporates an optimal extraction method and automatically masks cosmic rays,

significantly reducing user workload.

## 2. System Framework

The program is written in Python 3 and utilizes several open-source third-party libraries: Astropy for file reading and storage, and Numpy, Scipy, and Matplotlib for scientific computing and visualization. The software is released as open-source at [www.zenodo.org/record/7866030#.ZEi9uM5ByUk](http://www.zenodo.org/record/7866030#.ZEi9uM5ByUk).

### 2.1 Program Structure and Workflow

The program mainly consists of three components: data reading, data combination, and spectral extraction. Data reading includes scanning to generate observation logs and reading data from each step. Data combination includes bias and flat-field merging and correction, as well as merging of calibration lamp red and blue ends. Spectral extraction includes bias correction, image distortion correction, background subtraction, wavelength calibration, and one-dimensional spectrum extraction. The specific workflow is shown in [Figure 1: see original paper].

Users need to input the path to the raw data files. The program automatically executes the necessary steps, stores intermediate results, visualizes key steps, and produces final data products.

### 2.2 Observation Log Generation

During observations with the 2.16 m telescope, observers often record electronic observation logs. However, to avoid format inconsistencies, the program scans the computer directory containing the raw data, identifies all FITS format files, and regenerates a machine-readable formatted observation log based on each file's header information combined with the user's recorded log. The log contains information such as filename, observation target name, and observation start time.

### 2.3 Bias Correction

Bias refers to the image obtained when a CCD is powered on without photon input. Typically, observers obtain a set of bias images by performing multiple zero-second exposures with the camera shutter closed before or after the observation. These images are combined to obtain bias values and suppress readout noise errors. The program provides two combination modes: mean and median, with mean mode as the default. After combination, a master bias image is generated and stored separately as a FITS file. Subsequent data processing simply requires subtracting this master bias to complete bias correction.

## 2.4 Target Source Location

The program determines the position of the target's spectrum on the image using the following method: starting from the left side of the spectral image, a cross-section is taken every 10 pixels along the x-axis. The highest point of the cross-section is used as the initial position, and data within  $\pm 10$  pixels are fitted with a Gaussian function to obtain the central position and full width at half maximum (FWHM). This process is repeated, using the fitted center position as the new initial position until the fitting parameters converge. The converged center position is taken as the spectral center position curve. Next, a third-order polynomial function is fitted to the relationship between all y-direction center positions and their x-coordinates to obtain the target source's spectral center position curve. The program also provides a manual mode where users can mark the approximate position of the target source on a pop-up graphical interface, and the program will then locate the accurate spectral profile near the marked position.

## 2.5 Flat Field Correction

Even with uniform light input, different pixels of a CCD have slightly different responses to photons. During observation, multiple exposures are generally taken and combined to obtain a high signal-to-noise ratio and suppress readout noise. The program uses the Savitzky-Golay algorithm to fill bad pixel data with cubic spline interpolation and obtains a master flat-field image. Each target spectral image is divided by the master flat-field image to complete flat-field correction.

## 2.6 Field Distortion Correction

Due to optical distortion in the CCD camera, the image of the slit on the spectral image is somewhat curved. The program selects the one-dimensional spectrum at the middle position along the y-axis as the reference spectrum. In both upward and downward directions along the y-axis, the same one-dimensional extraction operation is performed, and the cross-correlation function is calculated between the obtained spectrum and the reference spectrum:

$$CCF(k) = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_{i+k} - \bar{y})]}{n \cdot SD_x \cdot SD_y}$$

where  $x$  and  $y$  are the spectra to be corrected and the reference spectrum respectively,  $n$  is the number of spectral points,  $k$  is the relative pixel offset, and  $SD_x$  and  $SD_y$  are the standard deviations of the two spectra. The pixel difference corresponding to the maximum value of the cross-correlation function is the actual pixel offset. A correction function is obtained from these offsets, and a corrected image is generated. [Figure 3: see original paper] shows that after distortion correction, the wavelength calibration accuracy of the one-dimensional

spectrum can be effectively improved and systematic errors avoided.

## 2.7 Background Subtraction

In actual observations, the obtained spectrum includes not only light from the target source but also sky background and internal instrument background. The target star's spectrum contains these components as well. Therefore, to obtain the true flux of the target source, background subtraction must be performed on the image data. The program typically controls the target star to fall near the middle of the y-axis. It automatically selects rows on the CCD image that are 340-500 pixels away from the target star, where the background light is relatively uniform. During selection, for rows with intensity exceeding the mean by more than 3 times the standard deviation (which may contain spectra of other objects near the target, especially common in dense star fields), the program masks these rows when calculating background values to avoid obtaining significantly overestimated background flux. Subtracting the background spectrum from the pixel values near the target source yields the spectrum of the target object.

## 2.8 Wavelength Calibration

The purpose of wavelength calibration is to establish a mapping between pixel position and wavelength on the image. The program includes automatic wavelength identification with the following procedure: (1) Read the pre-calibrated template spectral data provided for the mode; (2) Calculate the cross-correlation function between the spectral data to be processed and the template spectral data to obtain the offset between the two spectral positions; (3) Read the template pixel position data, add the offset to obtain new pixel position data, then take one-dimensional spectra of  $\pm 5$  pixels around the new pixel position, fit with a generalized Gaussian function  $A \exp[-(x - x_0)^2/2\sigma^2] + B$  to obtain the peak position corresponding pixel value, and finally fit with a fifth-order polynomial function to obtain the wavelength-pixel correspondence function, saving all wavelength-pixel data pairs.

Due to the characteristics of the calibration lamps used by BFOSC, their intensities vary significantly. If the exposure time is too long, the red end saturates, while if too short, the blue end lines lack sufficient intensity. We typically take one long exposure and one short exposure as the red and blue end wavelength calibration images respectively. In the example presented here, the exposure times are 30 s and 300 s. Before processing, this method truncates and merges the two sets of spectral data to generate a new calibration spectral image. During processing, the cross-correlation function is calculated separately for the red and blue ends.

## 2.9 Optimal Extraction Method

The optimal extraction method uses weighted summation to include as few noisy pixels as possible while ensuring sufficient flux, which can significantly improve

the signal-to-noise ratio for faint sources. The optimal extraction process mainly consists of two steps: (1) Obtain the precise spatial profile of the target star. This step first defaults to a spatial profile width of 10 pixels, with each pixel having the same weight in the profile direction, then calculates the variance of the spectral data and corrects the weight and spatial profile according to the variance. This process is iterated 2-3 times to finally obtain an accurate spatial profile; (2) Extract the one-dimensional spectrum of the scientific target based on the profile obtained in the previous step, calculate the difference with the one-dimensional spectrum obtained by ordinary extraction method, mask pixel points with deviations greater than 5 times the standard deviation (considered to be pixels affected by cosmic rays), and repeat the above steps until convergence. This yields the final one-dimensional spectrum of the scientific target with cosmic rays masked.

### 3. Results and Comparison

To verify the reliability of the program, we processed the same set of spectral data of SN2023eoc using both IRAF and our software, and plotted the one-dimensional spectra. The blue curve shows the IRAF processing result, while the yellow curve shows the result from our automated software. The two methods produce basically consistent results, but our software effectively masks cosmic rays automatically. In low-flux regions, the signal-to-noise ratio is significantly better than the IRAF result. IRAF requires manual masking of cosmic rays for each spectrum during processing, whereas our method saves processing time and provides higher precision.

[Figure 8: see original paper] Comparison of results. The blue curve is the IRAF result, and the yellow curve is the result of this pipeline.

### 4. Conclusion

This paper systematically introduces a long-slit spectral data processing software specifically designed for the BFOSC spectrograph on the 2.16 m telescope. The main procedures include bias and flat-field merging and correction, target star location, field distortion correction, background subtraction, wavelength calibration, and optimal extraction. The software is easy to use, requiring only the input path to the raw data. The processing is transparent and visual, with each step generating corresponding images and saving data for reuse. It can batch-process spectral data automatically while automatically removing cosmic rays, producing better results than IRAF and offering high practical value.

### Acknowledgments

We thank all the staff of the Xinglong 2.16 m telescope for their support.

## References

- [1] Fan Z, Wang H, Jiang X, et al. The Xinglong 2.16-m telescope scientific projects and current instruments. Publications of the Astronomical Society of the Pacific.
- [2] Huang L, Wu H, Li H. BFOSC Operating Manual [EB/OL]. 2012-04-15/2023-05-06. [http://astrocloud.china-vo.org/s/static/u/doc/BFOSCmanualv2.1\\_{chinese}.doc](http://astrocloud.china-vo.org/s/static/u/doc/BFOSCmanualv2.1_{chinese}.doc).
- [3] Li H. BFOSC Operating Manual [EB/OL]. 2012-04-15/2023-05-06. [http://astrocloud.china-vo.org/s/static/u/doc/BFOSCmanualv2.1\\_{chinese}.doc](http://astrocloud.china-vo.org/s/static/u/doc/BFOSCmanualv2.1_{chinese}.doc).
- [4] Xinglong Observatory, National Astronomical Observatories, Chinese Academy of Sciences. About the telescope and instrument [EB/OL]. 2012-04-15/2023-05-06. [http://www.xinglong-naoc.cn/html/gcyq/216/detail\\_{18}.html](http://www.xinglong-naoc.cn/html/gcyq/216/detail_{18}.html).
- [5] Zheng J, Jiang L. An automatic light-curve extraction pipeline. *Astronomical Research & Technology*, 83-89.
- [6] Perley D. Fully automated reduction of long-slit spectroscopy with the low resolution imaging spectrometer at the Keck Observatory. Publications of the Astronomical Society of the Pacific.
- [7] Tody D. The IRAF data reduction and analysis system. *IRAF/DAOPHOT/Proceedings of SPIE*. 1986.
- [8] Savitzky A, Golay M. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 1627-1639.
- [9] Horne K. An optimal extraction algorithm for CCD spectroscopy. Publications of the Astronomical Society of the Pacific.

---

**TABLE:1** BFOSC Grating Parameters [1-3]

| Rec.    | Lin.     | Disp. ( $\text{nm} \cdot \text{mm}^{-1}$ ) | Disp. ( $\text{nm} \cdot \text{pix}^{-1}$ ) | Wav. Range |
|---------|----------|--|---|------------|
| 13.9    | 19.8     | 0.208                                      | 0.297                                       | 330-660    |
| 19.9    | 39.2     | 0.298                                      | 0.132                                       | 360-870    |
| 29.5    | 83.7     | 0.142                                      | 0.120                                       | 520-1000   |
| 0.588   | 0.442    | 1.256                                      | 330-545                                     | 387-676    |
| 580-828 | 330-1000 | 360-960                                    | 520-1000                                    |            |

---

**FIGURE:1** Pipeline flow chart

**FIGURE:2** The central positions and full-width half-maximum of the spectral profiles: (a) central position and FWHM; (b) average profile; (c) real target source position; (d) profile center fitting

**FIGURE:3** FeAr spectra at different positions along the slit direction: (a) after distortion correction; (b) before distortion correction

**FIGURE:4** (a) Original 2D image of the target source; (b) CCD image areas used to determine the background level

**FIGURE:5** Peak positions of the identified FeAr lines

**FIGURE:6** Peak positions of the identified FeAr lines: (a) polynomial fitted wavelength ( $y$ ) vs. pixel ( $x$ ) relationship, blue dots represent calibration lines; (b) fitting residuals

**FIGURE:7** Peak positions of the identified FeAr lines

**FIGURE:8** Comparison of results: The blue curve is the IRAF result, and the yellow curve is the result of this pipeline

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv — Machine translation. Verify with original.*