Automated Wavelength Calibration for the CSST Integral Field Spectrograph*

Zhang Xuchen   Xu Shuyan

2026-08-18
ChinaRxiv

Abstract

The Integral Field Spectrograph (IFS) is an important scientific instrument at the back end of the Chinese Space Survey Telescope (CSST), used to carry out spectral observations with high spatial resolution. In both ground testing and future on-orbit operation, wavelength calibration of the IFS is required to obtain accurate detection results and high-quality astronomical observation data. The IFS has two observational bands, ultraviolet and visible light, among which the visible band contains 32 data channels. The mercury-argon lamp used for calibration will produce nearly 30 spectral lines in each channel, making the traditional manual calibration method time-consuming and labor-intensive. To improve calibration efficiency, an automated wavelength calibration method needs to be developed. To address this problem, connected-component analysis is first used to automatically extract spectra from the two-dimensional IFS data, obtaining one-dimensional spectra for all channels; then, particle swarm optimization combined with Gaussian fitting for peak finding is used to automatically identify characteristic spectral lines in the spectra, and the relevant algorithm parameters are provided; finally, a fifth-order polynomial is used to fit the dispersion curve. This method is applied to wavelength calibration of all 32 channels in the visible band of the IFS, and the calibration accuracy (root mean square of fitting residuals) is better than 0.35 Å for all channels, satisfying the accuracy requirements and realizing an automated workflow for IFS wavelength calibration.

Full Text

doi: 10.15940/j.cnki.0001-5245.2026.04.001

Automated Wavelength Calibration for the CSST Integral Field Spectrograph*

Zhang Xuchen[1,2] Xu Shuyan[1†]

(1 Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033)
(2 University of Chinese Academy of Sciences, Beijing 100049)

Abstract The Integral Field Spectrograph (IFS) is an important scientific instrument at the back end of the Chinese Space Survey Telescope (CSST), used to carry out spectroscopic observations with high spatial resolution. In ground testing and future in-orbit operation, wavelength calibration must be performed for the IFS in order to obtain accurate detection results and high-quality astronomical observational data. The IFS has two observational bands, ultraviolet and visible. The visible band contains 32 data channels; the mercury lamp used for calibration will produce nearly 30 spectral lines in each channel. Traditional manual calibration methods are time-consuming and labor-intensive.

To improve calibration efficiency, an automated wavelength-calibration method needs to be developed. To address this problem, a connected-domain analysis method is first used to automatically extract spectra from the two-dimensional IFS data, obtaining one-dimensional spectra for all channels. Then, a method combining particle swarm optimization with Gaussian fitting for peak searching is used to automatically identify the characteristic spectral lines in the spectra, and the relevant parameters of the algorithm are given. Finally, a fifth-order polynomial is used to fit the dispersion curve. Using this method, wavelength calibration was performed for all 32 channels in the visible band of the IFS. The calibration accuracy (root-mean-square value of the fitting residuals) is better than 0.35 Å, satisfying the accuracy requirement and realizing an automated workflow for IFS wavelength calibration.

Keywords Instruments: spectrographs; techniques: spectroscopy; methods: data analysis

CLC number: P111; **Document code:** A

1 Introduction

The Chinese Space Survey Telescope (CSST) is an off-axis three-mirror anastigmatic telescope with an aperture of 2 m and an f-number of 14. It will mainly be used to carry out large-scale multicolor imaging and slitless spectroscopic sky surveys[1-3]. The back-end instruments carried by the telescope include a multicolor imaging and slitless spectroscopic survey camera, a multichannel imager, an Integral Field Spectrograph (IFS), a spectrograph for exoplanet imaging, and a terahertz spectrometer. Among these, the IFS can acquire spectral information for every spatial-resolution element in a two-dimensional image; it has relatively high spatial resolution and is a powerful tool for studying neighboring galaxies and their physical processes[1].

Wavelength calibration[4-6] is an important step in the data processing of various spectrographs. Using a calibration light source (generally a line lamp) to produce emission lines of specific wavelengths, and combining these with the positions at which the spectral lines are imaged on the detector, a correspondence between detector pixels and spectral wavelengths is established; this is the dispersion curve of the spectrograph. For the IFS in CSST, wavelength calibration is particularly important. On the one hand, in ground testing of the IFS, its bandpass range and spectral resolution must be tested to ensure that the design specifications are met, and wavelength calibration is a necessary prerequisite for carrying out these test items. On the other hand, after the IFS is put into operation, wavelength calibration is also required to provide a basis for further analysis of celestial spectra and for astrophysical research.

Identification of the characteristic spectral lines of the calibration lamp is an important step in wavelength calibration. In traditional wavelength-calibration work, spectral-line identification is performed manually: the positions of characteristic lines are found in a one-dimensional spectrum, and then the wavelengths of those characteristic lines are found by comparison with a line list. This not only requires the operator

Received original manuscript: 2025-08-28; received revised manuscript: 2025-11-04

* Supported by the National Natural Science Foundation of China (61905241)

† xusy@ciomp.ac.cn

···possess a certain level of experience, which makes the calibration work more laborious. The visible-light band of the IFS has a total of 32 channels. The spectral data in each channel must be wavelength-calibrated separately. The light source used for calibration is a mercury–argon lamp, which produces nearly 30 spectral lines in each spectrum that must be identified one by one. Clearly, an inefficient manual identification method is not suitable for IFS wavelength calibration, and it is necessary to develop an automated spectral-line identification method. In addition, the one-dimensional calibration-lamp spectrum required for IFS wavelength calibration must be extracted from the original two-dimensional image data. To automate the entire calibration procedure, it is necessary to develop an automatic spectrum-extraction method corresponding to the IFS data format.

In response to the needs of ground testing and scientific applications of the integral-field-of-view spectrograph for the Survey Space Telescope, this paper establishes an automated processing workflow for wavelength-calibration data. First, based on connected-component analysis, automated spectral extraction is performed on the two-dimensional IFS image data. Then, using an iterative approximation method based on prior values, a template spectrum of the mercury–argon lamp is generated by a particle swarm optimization algorithm; the template spectrum is then used to match the positions and wavelengths of the calibration lines in the measured spectra, thereby realizing automated spectral-line identification. Finally, the method presented in this paper is applied to IFS ground-test data, and the calibration results are analyzed.

2 Automated Spectral Extraction Based on Connected-Component Analysis

Connected component analysis (CCA)[7] is a technique widely used in digital image processing. By selecting an appropriate threshold to binarize a digital image, the portions with larger numerical values (light intensity) in the original image appear as connected regions. CCA can be used to extract feature information such as the area and centroid coordinates of these regions. In this section, CCA is used to process the two-dimensional IFS data. The characteristic information of the connected regions of spectral lines is used to determine the distribution region of the spectral data in each channel, and the one-dimensional spectral data in the dispersion direction are extracted, thereby realizing automated spectral extraction for the IFS.

Fig. 1 2D image of IFS data

Figure 1: Fig. 1 2D image of IFS data

The two-dimensional IFS data image after preprocessing such as bias correction is shown in Fig. 1 (the IFS has two detectors for the ultraviolet and visible bands; this paper processes the data from the visible-band detector, and this will not be stated separately below). In the figure, the horizontal direction is the spatial direction, and the vertical direction is the dispersive direction. In the central region of the image, 32 channels of data arranged along the spatial direction can be seen; from left to right, they are channels 1–32. There are 3072 pixels in the dispersive direction. The wavelength increases with increasing pixel coordinate; that is, the lower end of the image is the short-wavelength end, and the upper end is the long-wavelength end.

Fig. 1 2D image of IFS data

Because of the influence of instrument efficiency, the light intensities differ among channels. If CCA is applied directly to the full-frame two-dimensional image, it is difficult to determine a unified threshold for image binarization. Therefore, the approximate boundary of each channel' s data should first be determined, and the data of each channel should be extracted for separate processing. Determining the boundary of each channel means identifying the pixel coordinates of the channel boundary in the spatial direction, i.e., the initial localization of the channel data.

The imaging regions of the IFS channels on the detector are essentially stable. After the initial localization is completed, the approximate boundary coordinates obtained can be repeatedly used in subsequent work. The approximate channel boundaries can be manually marked on the two-dimensional image, or they can be determined using IFS data under white-light illumination (the interior of the channels is uniformly illuminated); in a single row of data along the spatial direction, the points where the light intensity changes most sharply are the channel-boundary positions. The initial localization results for the 16 odd-numbered channels are shown in Table 1. The table gives the pixel positions of the left and right boundaries of each channel in the spatial direction. Taking the average of the pixel positions of the two boundaries gives the center position of each channel. Taking the average of the center positions of adjacent odd-numbered channels gives the approximate center positions of the first 15 even-numbered channels. The center position of channel 32 is obtained by adding the average width of a single channel to the center position of channel 31.

After the approximate center positions of the channels are obtained, several columns of data are taken on both the left and right sides of each center to obtain the two-dimensional data of a single channel. Here, the width of each channel is taken to be 120 columns, which is significantly larger than the average

Fig. 2 The original image (a) and binary image (b) of IFS channel 5

Figure 2: Fig. 2 The original image (a) and binary image (b) of IFS channel 5

channel width (80 columns), to ensure that the data of the entire channel are included. At this point, the extracted region contains a considerable amount of background data, so further analysis is needed to determine the positions of the calibration-lamp spectra. Binarization and CCA are performed separately on the two-dimensional data of each channel. Taking IFS channel 5 as an example, Fig. 2(a) shows the original channel data image, in which the bright lines are mercury–argon lamp emission lines. Because the channel contains many relatively strong···

emission lines. The range for setting the binarization threshold is relatively loose; generally, taking 10% of the maximum flux value within the channel is sufficient. The binarized image is shown in Fig. 2(b), where the regions containing spectral lines of relatively high intensity become thin, strip-like connected components.

Table 1 The rough edge positions of IFS odd channels

Channel	Left edge /pixel	Right edge /pixel	Channel	Left edge /pixel	Right edge /pixel
1	1384	1463	17	3081	3169
3	1620	1675	19	3293	3381
5	1807	1891	21	3506	3593
7	2031	2106	23	3720	3806
9	2232	2318	25	3930	4016
11	2443	2527	27	4144	4230
13	2656	2741	29	4361	4445
15	2872	2959	31	4572	4655

Fig. 2 The original image (a) and binary image (b) of IFS channel 5

In the CCA of this study, the most important connected-component features are area and centroid position. First, according to the order of area, the five connected components with the largest areas are selected; their feature information is shown in Table 2. The centroid coordinates of these five connected components in the spatial direction are averaged and used as the central position of the spectral line within the channel. Taking these average coordinates as centers, 80 columns of data are extracted from the two-dimensional channel data; each column is a one-dimensional spectrum of a calibration lamp distributed along the dispersion direction, which can be used for subsequent wavelength calibration. A one-dimensional spectrum in IFS channel 5 is shown in Fig. 3.

Table 2 Properties of the connected components in IFS channel 5

Fig. 3 The 1D spectrum in IFS channel 5

Figure 3: Fig. 3 The 1D spectrum in IFS channel 5

Area/pixel^2	Centroid/pixel
459	57.90
365	56.54
331	56.75
325	55.94
296	57.00

3 Automated spectral-line identification based on particle swarm optimization

The key step in wavelength calibration is identifying the spectral lines in the spectrum, that is, matching the pixel positions of the spectral lines in the dispersion direction with their wavelength values. Achieving automated spectral-line identification is the core problem to be solved in this paper. In the initial wavelength calibration work for the Large Sky Area Multi-Object Fiber Spectroscopy Telescope (LAMOST) in China' s Guo Shoujing Telescope, there is a class of iterative approximation calibration methods based on prior values[^8-9]. According to the prior information of the spectrograph, a rough relationship between wavelength and pixel is preset, namely a prior dispersion curve. Using the calibration-lamp line list and the dispersion curve, the calibration-lamp spectrum under this dispersion curve can be generated, namely a template spectrum. An appropriate optimization algorithm is then used to iteratively optimize the dispersion curve so that the template spectrum approaches the measured spectrum, ultimately obtaining an accurate dispersion curve as the calibration result. This method realizes automated wavelength calibration; however, to achieve relatively high calibration accuracy, relatively complete prior information for the spectrograph is required. At the same time, in order to rapidly optimize the dispersion curve to the optimal result, the requirements on the performance of the optimization algorithm are also relatively high.

The method in this paper uses an iterative approximation calibration based on prior values

As the basis of the method, only the most basic spectrometer design parameters and the calibration-lamp line list are used. A particle swarm algorithm is employed to optimize the template spectrum; however, the optimization result is not used directly as the calibration result. Instead, through comparison between the template spectrum and the measured spectrum, the positions and wavelengths of the spectral lines in the measured spectrum are automatically matched, thereby fitting the dispersion curve and realizing automated wavelength calibration.

图 3　IFS 第 5 通道一维光谱

Fig. 3　The 1D spectrum in IFS channel 5

3.1 Application of the Particle Swarm Algorithm in Template-Spectrum Optimization

Particle Swarm Optimization (PSO) is a stochastic optimization method proposed by Clerc et al. in 2002[^10]. Inspired by the foraging behavior of flocks of birds, it uses the continuous updating of the positions and velocities of a swarm of initial particles (random solutions) to iteratively search for the optimal solution of a function. In each iteration, a particle evaluates the quality of its own solution using the fitness function, and updates its own velocity and position by referring to the individual best position and the global best position, until a solution satisfying the convergence criterion is found. The PSO algorithm has the advantages of simple parameter settings, fast convergence, strong global search capability, and relatively easy programming implementation, making it particularly suitable for engineering applications.

In this study, the particle swarm algorithm is used to optimize the coefficients of the prior dispersion curve, so as to obtain a template spectrum sufficiently close to the measured spectrum. The spectral distribution of the IFS is very close to linear, and the prior dispersion curve can be directly set as a simple linear function; thus the coefficients to be optimized are only two: the 0th-order coefficient (intercept) and the 1st-order coefficient (slope). According to the IFS design parameters, the spectral coverage in the visible band is 6000 Å to 10000 Å, and the number of pixels in the dispersion direction is 3076. Based on these parameters, the prior dispersion coefficients to be optimized are set as follows: the 0th-order coefficient is the starting value of the spectral coverage, 6000; the 1st-order coefficient is the ratio of the spectral coverage interval (4000) to the number of pixels in the dispersion direction, approximately 1.3. No other complicated prior information is needed.

Using the dispersion curve and the calibration-lamp line list, the template spectrum of the calibration lamp can be generated[^8]. IFS wavelength calibration uses a mercury–argon lamp. The line list within the instrumental spectral range is shown in Table 3; the wavelength values and relative-intensity data in the table are from the National Institute of Standards and Technology (NIST). In addition, the table also indicates the elemental spectrum to which each emission line belongs and the corresponding ionization state of the element.

In Table 3, lines with relatively weak intensities have been removed to ensure that all lines in the line list can be detected by the IFS and maintain good profiles. This is important for subsequent processing.

In the particle swarm algorithm, each iteration generates multiple sets of dispersion-curve coefficients, and then generates the template spectra corresponding to these coefficients. The correlation coefficient between the template spectrum and the measured spectrum obtained from extraction is calculated[^8] and is used as the fitness function of the algorithm, i.e., the evaluation parameter for the optimization effect of the algorithm.

During the optimization process, updating the particle velocity is a key step. Its basic formula is as follows[^11]:

$$v_i^{k+1} = \omega v_i^k + c_1 r_1 \left(p_i^k - x_i^k\right) + c_2 r_2 \left(p_g^k - x_i^k\right), \qquad (1)$$

where k is the iteration number, ω is the inertia weight, c_1 and c_2 are learning factors, r_1 and r_2 are random numbers in $[0,1]$, x_i^k is the position of particle i at the k-th iteration, v_i^k is the velocity of particle i at the k-th iteration, p_i^k is the position of the individual optimal solution of particle i after k iterations, and p_g^k is the position of the global optimal solution of all particles after k iterations. v_i^k represents the current velocity state of the particle; the self-cognition component $r_1 \left(p_i^k - x_i^k\right)$ is the vector difference between the current particle position and the particle' s own best position, indicating that the source of the velocity-update component for the next iteration is its own experience; the social-experience component $r_2 \left(p_g^k - x_i^k\right)$ is the vector difference between the current particle position and the global best position of all particles, indicating that the source of the velocity-update component for the next iteration is the sharing of information among particles.

The velocity-update formula determines the particle' s state in the next iteration.

position, i.e., the random solution found by the search. Therefore, correctly selecting the inertia weight and learning factors in the formula plays a decisive role in the rapid convergence of the algorithm. For different practical problems, the inertia weight and learning factors have different selection methods.

Table 3 Emission lines of the Hg-Ar lamp for IFS wavelength calibration

Spectrum	Wavelength/Å	Relative intensity	Spectrum	Wavelength/Å	Relative intensity
Hg I	5771.283	1000	Ar I	8008.479	600
Hg I	5792.354	1000	Ar I	8017.110	700
Ar I	6967.451	300	Ar I	8106.043	600
Ar I	7069.267	300	Ar I	8117.664	1000
Hg I	7083.954	150	Ar I	8266.919	300
Ar I	7275.045	60	Ar I	8410.648	400
Ar I	7386.121	300	Ar I	8427.091	600
Ar I	7506.045	600	Ar I	8523.913	400
Ar I	7516.831	400	Ar I	8670.458	130
Ar I	7637.320	700	Ar I	9125.613	1000
Ar I	7726.001	400	Ar I	9227.174	400
Ar I	7726.447	300	Ar I	9660.587	700

Spectrum	Wavelength/Å	Relative intensity	Spectrum	Wavelength/Å	Relative intensity
Ar I	7950.481	600	Hg I	10142.7	1000

Through repeated experiments, we determined algorithm parameters that yield good performance, as shown in Table 4. The table includes parameters applicable to the IFS template-spectrum optimization problem, such as the inertia weight, learning factors, number of particles, and initial estimates of the coefficients to be optimized. Among them, the maximum velocity (Max v) limits the maximum amount of change of the coefficient to be optimized in each iteration. In this study, a dynamically varying inertia weight factor was adopted[1], so that the inertia weight factor gradually decreases as the number of iterations increases, thereby enhancing the local search capability of the particle swarm.

To test the optimization performance of the algorithm, 500 rounds of optimization experiments were carried out separately using the spectral data from IFS channels 5 and 6, and the fitness value after each round of optimization was recorded. To take into account both the optimization success rate and operating efficiency of the algorithm, the fitness-value threshold in each round of optimization was set to 0.8; that is, when the cross-correlation value between the template spectrum and the measured spectrum is greater than 0.8, the optimization is considered successful, and the dispersion coefficients at that time are output to generate the corresponding template spectrum. The histogram of the fitness values obtained in each round of optimization is shown in Figure 4. The results show that, for this problem, the particle swarm algorithm with the parameters shown in Table 4 has a relatively high optimization success rate; the average success rates for single optimizations of channels 5 and 6 are 96.6% and 99.8%, respectively. In the experiments, an Intel Core i7 processor was used, and the average running time for a single optimization was about 5 s, indicating relatively high operating efficiency.

Table 4 Parameters of PSO in this particular research

Parameter	Value
Max iterations	300
Particle amount	150
Max ω	3.0
Min ω	0.4
c_1	1.8
c_2	0.04
0 order coefficient	6000
1 order coefficient	1.3
Max v (0 order)	30

[1] Reference marker visible on the page.

Histogram of the cross-correlation coefficient of IFS channel 5 (a) and channel 6 (b)

Figure 4: Histogram of the cross-correlation coefficient of IFS channel 5 (a) and channel 6 (b)

Parameter	Value
Max v (1 order)	0.05

Figure 4 Histogram of the cross-correlation coefficient of IFS channel 5 (a) and channel 6 (b)

Comparing the successfully optimized template spectra with the measured spectra, the results are shown in Figure 5. Figure 5(a) and Figure 5(c) are the measured spectra of channels 5 and 6, respectively, while Figure 5(b) and Figure 5(d) are obtained by plotting together the template spectra for which all cross-correlation values in the 500 rounds of optimization for the two channels are greater than 0.8. These template spectra have different cross-correlation values relative to the measured spectra; that is, they have not been optimized to the optimum. However, all the template spectra almost completely overlap and are very similar to the measured spectra, and all can be used in the subsequent spectral-line matching work.

3.2 Matching and Profile Fitting of Calibration Spectral Lines

After optimization, we obtained template spectra that are sufficiently close to the measured spectra. In this study, these template spectra will be used to segment the measured spectra, and peak searching will be performed on the measured spectral data blocks. The precise center coordinates of the spectral-line profiles will then be fitted, thereby realizing the matching of spectral-line pixel positions and wavelengths, i.e., automatic identification of spectral lines.

In the optimized template spectra, the wavelength and pixel position of each spectral line are determined, and the positional distributions of these spectral lines are very close to those in the measured spectra, with at most a deviation of 1 to 2 pixels. First, according to the distribution of spectral lines, the entire template spectrum is divided into several data blocks, each containing several spectral lines that are close to one another. Then, according to the pixel positions of the boundaries of each data block, the data at the corresponding positions are extracted from the measured spectrum to obtain the measured spectral data blocks to be processed. Such segmentation is mainly based on two considerations. First, for spectral lines that are close to one another, there is cross-contamination of flux between them; when fitting the spectral-line profiles,

Fig. 5

Figure 5: Fig. 5

an appropriate number of Gaussian functions should be selected for simultaneous fitting according to the number of spectral lines within the data block. Second, the IFS calibration-lamp spectra to be processed are not strictly line spectra, but instead have a certain continuous flux distribution. Within small data blocks of several tens of pixels, the slowly varying continuous flux distribution can be regarded as a constant or a linear function, facilitating subsequent spectral-line profile fitting.

Generally speaking, the measured spectral data block should contain the same number of peaks as the number of spectral lines in the corresponding template data block, and they should correspond one-to-one. However, the spectral-line list used to generate the template spectrum is incomplete; weak calibration-lamp spectral lines may not appear in the template spectrum because they were removed from the line list, but they may appear as peaks in the measured spectrum. At the same time, some spectral lines of unknown origin or strong noise may also appear in the measured spectrum. These factors can cause the number of peaks in the measured spectral data block to be greater than the number of spectral lines in the template spectrum. In addition, owing to the limited spectral resolution of the IFS, multiple spectral lines with very close wavelengths in the line list may overlap, causing the number of peaks in the measured spectrum to be smaller than the number of spectral lines in the template spectrum. To solve this problem, peak searching must be performed on the measured spectral data block, and the characteristic spectral lines in the calibration-lamp line list (Table 3) are selected for fitting of the dispersion curve.

Fig. 5 Measured spectrum (a), (c) and model spectrum (b), (d) of IFS channels 5 and 6

At present, commonly used spectral peak-searching methods mainly include the derivative method[^13], fitting method[^14], symmetric zero-area method[^15], and continuous wavelet transform method[^16], etc. This paper adopts the fitting method for peak searching: a Gaussian function is used to fit the characteristic spectral-line profiles in the measured spectral data block[^17], obtaining the precise center position of each spectral line. The fitting formula is as follows:

$$\sum_{j=1}^{n} a_j e^{-\frac{(x-b_j)^2}{2s_j^2}} + d, \qquad (2)$$

where n is the number of superposed Gaussian functions, which is equal to the number of spectral lines in the corresponding model spectral data block; a_j, b_j, and s_j are respectively the amplitude, center position, and standard deviation of the j-th Gaussian function, and the constant term d is used to fit the continuous background flux distribution. During fitting, the pixel positions of the spectral lines in the model spectrum, the flux values at the corresponding positions in the measured spectrum, and the mean flux values at the two ends of the measured spectral data block are used respectively as the initial values of the parameters b_j, a_j, and d, in order to obtain accurate fitting results; the initial value of the parameter s_j can be set to 1.

After the fitting is completed, every spectral line (wavelength) in the line list corresponds to a pixel position in the measured spectrum. Taking the processing result of one spectrum in channel 5 as an example, Fig. 6 shows the measured data (circles) and profile fitting results (solid curves) for some spectral lines, and marks the wavelength of each spectral line. The spectral lines shown in the figure are all lines that are close enough to each other to produce partial overlap or have relatively weak intensity, and they are distributed basically uniformly over the entire spectrum and are representative. The correspondence between the wavelengths of all calibration spectral lines and pixels is shown in Table 5. Using these corresponding points, the dispersion curve can be fitted and wavelength calibration completed. Compared with the line list used to generate the model spectrum (Table 3, with a total of 26 spectral lines), Table 5 used for fitting the dispersion curve contains only 24 spectral lines. This is because, in the measured IFS spectrum, the emission lines of argon at wavelengths 7726.0 Å and 7726.4 Å overlap. In the actual data processing, it was found that the line-center positions obtained by fitting the data block containing these two spectral lines have relatively large errors, which reduces the precision of the wavelength calibration. Since the characteristic spectral lines of the mercury-argon lamp are relatively abundant, these two spectral lines are not used when fitting the dispersion curve. In addition, due to differences in the spectral ranges of the channels, the two spectral lines at 5771.3 Å and 5792.4 Å located at one end of the spectrum appear only in the spectra of odd-numbered channels. Therefore, wavelength calibration of spectra in odd-numbered channels generally uses 24 spectral lines, whereas even-numbered channels use 22 spectral lines.

4 Automated Wavelength-Calibration Results for Measured Data

Using the automated wavelength-calibration method presented in this paper, the IFS ground-based

The test data were processed, yielding wavelength calibration results for 32 channels in the visible-light band. For each channel, 80 columns of spectral data were taken and calibrated separately, for a total of 2560 columns. Before calibration, the data in each column were combined with one column of data on each side, making three columns in total, and then processed to increase the signal-to-noise ratio and improve calibration accuracy. Owing to the influence of instrumental efficiency, some channels contained weak spectral data; these spectra were strongly affected by noise, their line-intensity distributions were

Fig. 6 Measured data and profile fitting result of spectral lines

Figure 6: Fig. 6 Measured data and profile fitting result of spectral lines

abnormal, and the mutual-correlation value with the template spectrum could not be optimized to the set threshold (0.8). Among the test images used in the experiment, after removing these abnormal data, a total of 2539 columns of valid data remained.

Figure 6 Measured data and profile fitting results of spectral lines

Fig. 6 Measured data and profile fitting result of spectral lines

In the experiment, a fifth-order polynomial was used to fit the dispersion curve. After obtaining the dispersion curve, the pixel positions of the spectral lines were substituted into it to calculate the fitted values of the spectral-line wavelengths. The fitted wavelength values were compared with the theoretical values in the line list to obtain the fitting residuals, and the root mean square value (Root Mean Square value, RMS) of the fitting residuals for all calibration spectral lines was taken as the evaluation parameter for calibration accuracy. The RMS error distribution of the wavelength calibration results for the 2539 columns of spectra is shown in Fig. 7.

As can be seen from Fig. 7, the wavelength-calibration errors of the vast majority of spectral data are less than 0.3 Å. Reference [4] points out that the accuracy requirement for wavelength calibration of a spectrograph is generally 10% of the spectral resolution. For IFS, this upper limit of accuracy is 0.35 Å. Among the 2539 columns of valid data, a total of 5 columns of spectral data have calibration errors exceeding the accuracy requirement; these spectra come respectively from channel 11, channel 22, channel 25, and channel 30. By analyzing the fitting residuals of individual calibration spectral lines, it was found that the maximum fitting residuals in these spectral calibration results all came from the spectral line with wavelength 8106.0 Å. The two spectral lines at 8106.0 Å and 8117.7 Å are close together in the spectrum and partially overlap, and their intensities differ greatly; the fitted centroid position of the weaker 8106.0 Å spectral line ...

there is a certain error. For spectra with relatively large calibration errors, the spectral line with wavelength 8106.0 Å was removed from the fitting points of the dispersion curve, and wavelength calibration was carried out again. The calibration results meet the precision requirements. A comparison of the calibration errors of the 5 spectra before and after is shown in Table 6; the first column of the table indicates the channel number to which each spectrum belongs.

Table 5 Position and wavelength of the lines used for calibration (channel 5)

Fig. 7 Wavelength calibration error (RMS)

Figure 7: Fig. 7 Wavelength calibration error (RMS)

Position /pixel	Wavelength /Å	Position /pixel	Wavelength /Å
18.300	5771.283	1455.776	8017.110
31.768	5792.354	1512.743	8106.043
783.954	6967.451	1520.150	8117.664
849.144	7069.267	1615.666	8266.919
858.655	7083.954	1707.701	8410.648
980.877	7275.045	1718.155	8427.091
1051.909	7386.121	1780.140	8523.913
1128.664	7506.045	1873.994	8670.458
1135.603	7516.831	2165.241	9125.613
1212.728	7637.320	2230.299	9227.174
1413.144	7950.481	2507.507	9660.587
1450.312	8008.479	2816.129	10142.700

Fig. 7 Wavelength calibration error (RMS)

The mean value of the spectral calibration error in each channel was calculated separately, and the results are shown in Table 7. It can be seen that there are differences in wavelength calibration precision among different channels; among them, the average calibration errors of channels 1 to 5, channel 31, and channel 32 are about 0.07 Å, indicating relatively high calibration precision.

Table 6 Wavelength calibration error of 5 spectra

Spectrum number	Calibration error /Å (original)	Calibration error /Å (corrected)
1 (CH11)	0.941	0.133
2 (CH22)	0.496	0.066
3 (CH25)	0.415	0.085
4 (CH30)	0.890	0.168
5 (CH30)	0.576	0.166

Table 7 The average calibration error of single channel

Channel number	Calibration error /Å	Channel number	Calibration error /Å
1	0.067	2	0.064
3	0.068	4	0.067

Channel number	Calibration error /Å	Channel number	Calibration error /Å
5	0.071	6	0.087
7	0.087	8	0.101
9	0.109	10	0.119
11	0.131	12	0.120
13	0.112	14	0.111
15	0.123	16	0.120
17	0.138	18	0.115
19	0.119	20	0.108
21	0.150	22	0.132
23	0.122	24	0.127
25	0.124	26	0.092
27	0.094	28	0.094
29	0.093	30	0.117
31	0.072	32	0.064

To visually display the wavelength-calibration results, the fitting result for the dispersion curve of one spectrum in IFS channel 5 (Fig. 8) and the fitting residual plot (Fig. 9) were plotted.

Fig. 8 Dispersion curve fitting result

Visible labels in the figure: "Fitting points" ; "Dispersion curve" ; vertical axis "Wavelength/Å" ; horizontal axis "The Dispersive Direction/pixel" .

Fig. 9 Residual error of fitting points

Visible labels in the figure: "Residual error of fitting points" ; "Mean residual error of all points"; "RMS=0.0569 Å"; vertical axis "Residual error/Å"; horizontal axis "The Dispersive Direction/pixel" .

5 Conclusion

This paper introduced the automated wavelength-calibration process for the integral-field spectrograph of the China Space Station Telescope Survey Space Telescope. The paper described in detail the connected-domain analysis processing method for IFS two-dimensional data, obtained the distribution region of the calibration-lamp spectrum on the two-dimensional image, and extracted the one-dimensional spectrum in the dispersion direction. It focused on the automatic spectral-line identification method based on the particle swarm optimization algorithm, gave the algorithm parameters applicable to IFS template-spectrum optimization, and realized automatic matching between the wavelengths of calibration spectral lines and pixel positions by using the template spectrum and Gaussian fitting. Finally, wavelength-calibration experiments were carried out using IFS ground-test data. A fifth-order polynomial was used to fit the dispersion curve, yielding wavelength-calibration results for all 32 IFS channels, with a total of 2,539 columns of valid spectral data. The RMS values of the wavelength-fitting residuals of the calibration spectral lines were used to evaluate the calibration accuracy. After correction, the calibration accuracy of all spectra was better than 0.35 Å, meeting the requirements for IFS wavelength-calibration accuracy. This work realizes automated wavelength calibration for all channels in the visible band of IFS, without relying on manual identification of spectral lines, thereby improving calibration efficiency. It provides a useful approach for developing efficient wavelength-calibration methods for IFS ground testing and on-orbit operation. Meanwhile, the automatic spectral-line identification method used in this paper, which optimizes the template spectrum and combines Gaussian fitting for peak searching, can also be applied to wavelength calibration for other similar spectrographs.

Acknowledgments We thank Associate Researcher Bian Guohao of the Changchun Institute of Optics for his support of this work; we thank the reviewers for their valuable comments and the editorial office for their assistance.

References

[1] Zhan Hu. *Chinese Science Bulletin*, 2021, 66: 1290

[2] Yuan H B, Deng D S, Sun Y. *RAA*, 2021, 21: 074

[3] Wang Wanhao, Liu Huigen, Zhou Jilin. *Acta Astronomica Sinica*, 2023, 64: 45

[4] Song J, Ren B C, Tang Y Y, et al. *Electronics*, 2024, 13: 4131

[5] Qin H Q, Zhu J, Zhu Z Q, et al. *PASA*, 2010, 27: 265

[6] Labiano A, Argyriou I, Marquez J, et al. *A&A*, 2021, 656: A57

[7] Qu Liguo, Chen Guohao, Hu Jun, et al. *Acta Electronica Sinica*, 2022, 50: 1521

[8] Wang Sheng, Zhu Zhangqin, Zhu Jia, et al. *Acta Astronomica Sinica*, 2011, 52: 413

[9] Xue Xuming, Ye Zhongfu. *Astronomical Research & Technology*, 2009, 6: 181

[10] Clerc M, Kennedy J. *ITEC*, 2002, 6: 58

[11] Lian Xiaoqin, Liu Qian, Chen Yanming, et al. *Spectroscopy and Spectral Analysis*, 2021, 41: 1452

[12] Pi Qingying, Ye Hongtao. *Journal of Guangxi University of Science and Technology*, 2016, 27: 26

[13] Wu Leichao, Li Jiangkun, Zhang Yayan, et al. *Nuclear Techniques*, 2024, 47: 080401

[14] Guo Xiaotong, Yang Qiuxiang. *Electronic Design Engineering*, 2018, 26: 161

[15] Wang Hai, Huang Ning, He Ze, et al. *Acta Optica Sinica*, 2024, 44: 0330001

[16] Zhang Z M, Tong X, Peng Y, et al. *Ana*, 2015, 140: 7955

[17] Li Huan, Hao Lei. *Progress in Astronomy*, 2017, 35: 255

Automated Wavelength Calibration of Space Integral Field Spectrograph of CSST

ZHANG Xu-chen[1,2] XU Shu-yan[1]

(1 Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033)
(2 University of Chinese Academy of Sciences, Beijing 100049)

Abstract The Integral Field Spectrograph (IFS) is an important scientific instrument on the Chinese Space Survey Telescope (CSST), and is designed for high spatial resolution spectroscopic observations. During both ground testing and future in-orbit operation, wavelength calibration for IFS is always required for precise test results and high-quality astronomical data. IFS has two observation bands, ultraviolet and visible, with the visible band containing 32 data channels, each of which has about 30 lines emitted by the Hg-Ar lamp used for calibration. Traditional manual calibration methods will be time-consuming and strenuous. To improve the efficiency of wavelength calibration, an automated wavelength calibration method must be developed. First, the 1D spectra of all IFS channels are automatically extracted from the 2D data using Connected Component Analysis (CCA). Then, Particle Swarm Optimization (PSO) and Gaussian fitting are applied to automatically identify the emission lines, meanwhile the parameters of PSO are given. Finally, the dispersion curve is fitted using a fifth-order polynomial. Implementing this method to all 32 channels in IFS visible band, the resulting calibration accuracy (root-mean-square value of residual error) is better than 0.35 Å, meeting the accuracy requirement. This demonstrates that the automated wavelength calibration of IFS is achieved.

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

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