

Baseline Correction of Fourier Transform Infrared Spectroscopy Signals Based on Cubic Splines

Miaomiao Zhu¹, Bo Yu^{1,*}, Zhiwen Yao²

¹College of Science, China Three Gorges University, Yichang 443002, Hubei, China.

²China Shipbuilding Industry Group Co. Ltd Alphapec Instrument (Hubei), Yichang 443005, Hubei, China.

How to cite this paper: Miaomiao Zhu, Bo Yu, Zhiwen Yao. (2024) Baseline Correction of Fourier Transform Infrared Spectroscopy Signals Based on Cubic Splines. *Journal of Applied Mathematics and Computation*, 8(4), 342-350.
DOI: 10.26855/jamc.2024.12.010

Received: November 25, 2024

Accepted: December 23, 2024

Published: January 21, 2025

***Corresponding author:** Bo Yu, College of Science, China Three Gorges University, Yichang 443002, Hubei, China.

Abstract

In order to solve the problem of baseline drift caused by the influence of environment, temperature, noise, and other factors in Fourier infrared spectrum data, we propose a baseline correction method based on the cubic spline, which is compared with other existing baseline correction techniques. Since baseline correction is a pre-processing step for subsequent substance composition identification, we can compare the baseline correction effect with the accuracy of substance composition identification. In order to exclude the influence of other factors on the accuracy of substance component identification, we will use the same method in subsequent noise reduction treatment and substance composition identification to effectively compare various baseline correction techniques. Furthermore, we compare different kinds of baseline correction method from two other aspects of the spectrum and root mean square error. The numerical results show that the proposed method has strong comparability with the existing methods in the spectral map and root mean square error, keeping the best performance in substance composition identification.

Keywords

Fourier-transform infrared spectroscopy; baseline correction; substance composition identification; comparative analysis

1. Introduction

Fourier transform infrared spectroscopy (FTIR) is increasingly used in chemistry, biomedicine, and other fields due to its unique analytical capabilities. The technique possesses powerful characterization [1], fast analytical capabilities [2], and easy handling procedures [3], and provides detailed information about the molecular structure of the sample. It has significant advantages in the identification and quantification of substance components.

When using FTIR spectroscopy for the identification of target substances, we first sample and obtain spectral data, which are then compared with the data in the database to determine the substance composition. However, the data acquisition process is subject to system noise caused by factors such as instrument aging and changes in the temperature of the light source, this can lead to a drift in the baseline of the spectrogram, which further can result in distortion in the measurement of spectral peak heights and peak areas, and thus will have an impact on the accuracy of target substance identification [4, 5]. In order to solve this problem, baseline correction of spectral data obtained from sampling is necessary. Because of the complexity of spectral data, the baseline correction techniques are difficult to meet the needs of modern spectroscopy, and researchers need to explore more efficient baseline correction techniques.

In the field of spectral data processing, a variety of baseline corrections have been widely studied and applied. These include Improved Iterative Polynomial Fitting (IIPF) [6, 7], Adaptive Iterative Reweighting Penalty Least Squares (AIRPLS) [8-10], Asymmetrically Reweighted Penalized Least Squares (ARPLS) [11, 12], and Iterative Discrete Wavelet Decomposition (IDWD) [13]. Each of them has its specific application scenarios with different advantages and

disadvantages. The IIPF is known for its great adaptability and flexibility to handle different types of data with different degrees of baseline drift. Specifically, the IIPF combines the algorithm of identifying spectral peaks with the algorithm of polynomial fitting. After identifying the spectral peaks, the spectral peaks that are closer to each other are merged, which reduces the number of points of baseline correction to some extent and makes the number of iterations decrease, thus effectively improving the speed of operation. However, its computation is large, and the selection of the parameter of the order of polynomials is difficult [14, 15]. On the basis of penalized least squares [16, 17], AIRPLS uses iterative re-weighting to change the weights to approximate the baseline, which has a strong noise immunity and a wide range of applications. However, in the region that is slightly smaller than the fitted baseline, this may underestimate the position of the baseline in the region with no spectral peaks [18]. The ARPLS, also based on the theory of penalized least squares, is able to adapt to different datasets and problems by choosing different weight functions and penalty terms, thus improving the prediction accuracy of the model. Since the choice of weights has a great impact on the results, it remains a challenge to choose the weights. Finally, the IDWD determines the optimal decomposition layer by selecting wavelet bases to obtain the approximation coefficients and then uses an iterative process to obtain the baseline. The baseline obtained by this process can remove baseline drift caused by the influence of noise, however, it suffers from a lot of application difficulties due to the need to select different thresholds according to different data in the iterative process. In summary, in order to improve the accuracy and efficiency of baseline correction, each of these baseline correction methods has its own merits but needs to be selected according to the specific data characteristics and application requirements.

Through further analysis of FTIR spectral data, we find that the core difficulty is how to deal with the local minimum points. Because the instrument is affected by noise, temperature, and humidity in the process of data collection, there will be many small oscillations and minimum points in the collected data. If all of these small values are taken out and put into the baseline generation process, on the one hand, it will lead to noise and other contaminated data in the baseline, and on the other hand, it will involve a large amount of computation. On the contrary, some of the information in these minimum points does represent the characteristics of the target material. If they are completely removed, it will make the identification of the target object more difficult. Therefore, a definition of "effective minimum point" is proposed in this paper. All effective minimum points are calculated according to amplitude. Finally, we fit the baseline by using cubic spline interpolation. The baseline data obtained through these steps are used for substance composition identification, and the results show the high efficiency of the proposed baseline correction algorithm.

The main idea and algorithm of the proposed method are presented in the second section. In the third section, we compare the numerical results obtained by the proposed method with those of other methods from three aspects: spectral spectrum, substance composition identification accuracy, and root mean square error. Finally, the fourth section gives a brief summary of this paper.

2. The Idea of the Proposed Algorithm

Since the quality of samples affects the accuracy of the analytical results of FTIR spectroscopy directly, the samples need to be strictly screened before analyses are performed, where the selected samples should be representative and not contaminated. In the process of collection of data, we first set the parameters of the spectrometer and then record the spectral data of the sample. In the raw spectral data, there is a baseline drift, which may affect the accuracy of substance composition identification. Therefore, before substance composition identification, we need to perform baseline correction on the raw spectral data. Through this step, we can effectively eliminate baseline drift and make the spectral data smoother and clearer.

To ensure the feasibility of the proposed method, we carried out multiple measurements and collected 30 sets of test data from 11 samples such as ibuprofen, butazone, and salicylic acid in the database which contains 475 substances including tetrachloroethane, tetramethylbenzene, and butanediol. All these data are performed a baseline correction and analyzed carefully. The corresponding numerical results will be presented in the next section. Before introducing the algorithm for baseline correction in this paper, we first give the definition of effective minima and the process of cubic spline interpolation.

2.1 Effective Minimum Points

The acquired spectral data are first stored as an $n \times 2$ matrix X , whose first column represents the horizontal coordinates of the spectral signal, denoted as vector $x = [x_1, \dots, x_n]^T$ and the second column represents the vertical coordinates of the spectral signal, denoted as vector $y = [y_1, \dots, y_n]^T$. In order to obtain the minimum point to determine the baseline, we first symmetrically extend the vector y by one point outward such that $y_{n+1} = y_{n-1}$, and the newly generated vector is denoted as $\tilde{y} = [y_1, \dots, y_{n+1}]^T$, and then we compute the first-order difference on the sequence \tilde{y} in the form $\tilde{y}'_i = \tilde{y}_{i+1} - \tilde{y}_i$,

($i=1, \dots, n$), and generate a new vector $\tilde{y}' = [\tilde{y}'_1, \dots, \tilde{y}'_n]^T$. If the absolute value of y'_i is less than a predefined smaller threshold, this means that the sequence \tilde{y} reaches a local extreme on the i -th component. To further find the minima, we then perform a first-order difference on \tilde{y}' to generate another vector $\tilde{y}'' = [\tilde{y}''_1, \dots, \tilde{y}''_{n-1}]^T$. If \tilde{y}''_i is greater than zero, then \tilde{y} has reached a local minima at the i -th data point, and the corresponding x_i is the minima point.

After determining the minima and minima points, we calculate the mean value of the minima \bar{y} and use \bar{y} as a threshold. If the minimum \tilde{y}_i is smaller than this threshold, we define it as an effective minima, and the corresponding x_i is called an effective minima points. Otherwise, if the value of \tilde{y}_i is greater than or equal to the threshold, it is defined as an ineffective minima and its corresponding x_i is called an ineffective minima point. Through the definition of effective minima point, we can classify the minima points more carefully and eliminate the ineffective minima points. Thus, we can identify the feature points with practical significance more accurately, which is convenient to use cubic spline interpolation to fit the baseline.

2.2 Cubic Spline Interpolation

After obtaining the effective minima point, in order to connect these data points smoothly to form a new baseline, we choose cubic spline interpolation since it is efficient and practical. Although higher-order spline interpolation can theoretically provide finer fitting results, it is more complicated in computation. In contrast, since cubic spline has the form of a piecewise cubic polynomial, it can not only achieve a similar fitting accuracy as that of higher-order spline interpolation but also be particularly convenient to calculate. In view of the advantages of cubic spline interpolation in computational accuracy and complexity, this paper decides to use cubic spline interpolation to construct an accurate baseline from the effective minima of obtained spectral data.

To perform the cubic spline interpolation, we first take the obtained effective minima as the node coordinates (x_j, y_j) , and in order to ensure the continuity and integrity of the interpolation, we include the start and end points of the raw data into the effective minima points and place them at the first and last nodes of the interpolation process. To ensure that the interpolated output is consistent with the length of the raw data sequence, we divide the interval consisting of the start point and the end point evenly, resulting in a uniformly spaced series of sample points between the two specified endpoints.

After we add the start point and end point of the raw data to the effective minima point, we get $t+1$ data nodes with $r = x_0 < x_1 < \dots < x_t = R$, these nodes are partitioned into t intervals. Below we give the cubic polynomial expression for the j -th interval [19]:

$$s_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3, \tag{1}$$

and its first derivative expression:

$$s'_j(x) = b_j + 2c_j(x - x_j) + 3d_j(x - x_j)^2, \tag{2}$$

and the second derivative expression:

$$s''_j(x) = 2c_j + 6d_j(x - x_j). \tag{3}$$

Next, we are going to find the specific coefficient in equation (1), namely, a_j, b_j, c_j and d_j . Based on the continuity of the cubic spline function itself and its first and second order derivatives, we know that the following equation holds at the node point:

$$s_j(x_{j+1}) = s_{j+1}(x_{j+1}), \quad s'_j(x_{j+1}) = s'_{j+1}(x_{j+1}), \quad s''_j(x_{j+1}) = s''_{j+1}(x_{j+1}), \quad (j = 0, 1, \dots, t-1).$$

Since the curve endpoints pass through all data points, it can be obtained that:

$$s_j(x_j) = y_j, \quad (j = 0, 1, \dots, t). \tag{4}$$

Let $m_j = s''_j(x_j)$ and the step length be $h_j = x_{j+1} - x_j$. From Eqs. (1), (2), (3) and (4), we get that:

$$\begin{cases} a_j = y_j \\ b_j = \frac{y_{j+1} - y_j}{h_j} - \frac{2m_j h_j + m_{j+1} h_j}{6} \\ c_j = \frac{m_j}{2} \\ d_j = \frac{m_{j+1} - m_j}{6h_j} \end{cases}, \quad j = 0, 1, \dots, t-1. \quad (5)$$

Substituting $x = x_{j+1}$ into $s'_j(x)$ and $s'_{j+1}(x)$ gives the following equation:

$$\begin{cases} s'_j(x_{j+1}) = b_j + 2c_j h_j + 3d_j h_j^2 \\ s'_{j+1}(x_{j+1}) = b_{j+1} \end{cases}, \quad j = 0, 1, \dots, t-2. \quad (6)$$

From Eqs. (5), (6) and $s'_j(x_{j+1}) = s'_{j+1}(x_{j+1})$, we obtain that:

$$h_j m_j + 2(h_j + h_{j+1})m_{j+1} + h_{j+1}m_{j+2} = 6 \left(\frac{y_{j+2} - y_{j+1}}{h_{j+1}} - \frac{y_{j+1} - y_j}{h_j} \right), \quad j = 0, 1, \dots, t-2. \quad (7)$$

From the range of values of j , we can see that Eq. (7) gives a total of $t-1$ equations, but there are $t+1$ unknowns m_j , $j = 0, 1, 2, \dots, t$. So to find all m_j s, we add two boundary conditions in the following way:

$$s_0'''(x_1) = s_1'''(x_1), \quad s_{t-2}'''(x_{t-1}) = s_{t-1}'''(x_{t-1}). \quad (8)$$

From Eq. (3), we get $s_j'''(x) = 6d_j$. In combination with Eq. (8), we have that:

$$\begin{cases} 6d_0 = 6d_1 \\ 6d_{t-2} = 6d_{t-1} \end{cases}. \quad (9)$$

Substituting d_j in Eq. (5) into Eq. (9) yields that:

$$\begin{cases} h_1(m_1 - m_0) = h_0(m_2 - m_1) \\ h_{t-1}(m_{t-1} - m_{t-2}) = h_{t-2}(m_t - m_{t-1}) \end{cases}. \quad (10)$$

The combination of Eq. (7) with Eq. (10) can be expressed in the following matrix form:

$$\begin{pmatrix} -h_1 & h_0+h_1 & -h_0 & 0 & \dots & 0 \\ h_0 & 2(h_0+h_1) & h_1 & 0 & \dots & 0 \\ 0 & h_1 & 2(h_1+h_2) & h_2 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & h_{t-2} & 2(h_{t-1}+h_{t-2}) & h_{t-1} \\ 0 & \dots & \dots & -h_{t-1} & h_{t-2}+h_{t-1} & -h_{t-2} \end{pmatrix} \begin{pmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_t \end{pmatrix} = 6 \begin{pmatrix} 0 \\ \frac{y_2 - y_1}{h_1} - \frac{y_1 - y_0}{h_0} \\ \frac{y_3 - y_2}{h_2} - \frac{y_2 - y_1}{h_1} \\ \vdots \\ \frac{y_t - y_{t-1}}{h_{t-1}} - \frac{y_{t-1} - y_{t-2}}{h_{t-2}} \\ 0 \end{pmatrix}. \quad (11)$$

Solving Eq. (11) we can find the values of all m_j s, and then substitute them into Eq. (5) we can get the coefficients a_j , b_j , c_j and d_j , thus, the expression of the polynomial function of cubic spline interpolation can be obtained.

Finally, the newly generated sample points can be substituted into the cubic spline interpolation polynomial function and its corresponding function value can be computed. Thus, a new baseline is obtained with accurate interpolation results over the entire interval while the same length as the raw data series is maintained.

2.3 Algorithm Description

In the previous theoretical analysis, we used the cubic spline interpolation principle to derive a piecewise polynomial expression, which constitutes the mathematical model of baseline fitting. In order to apply this expression to practical data and generate a smooth and continuous curve to fit the baseline, we will give a detailed step-by-step description of the baseline fitting algorithm as follows:

Algorithm: A baseline correction algorithm based on cubic spline interpolation

Input: Signal matrix X to be processed;

Output: Processed signal "baseline" y_{new} and corrected signal \tilde{s} ;

Algorithm:

- (1) Extract the first and second columns from the signal matrix X to be processed and store them as vectors x and vector y ;
- (2) Take the divided difference of the spectral signal to find the abscissa x_{min} and ordinate y_{min} of the minima point;
- (3) Calculate the absolute value of the ordinate y_{min} of all minima points, and find the mean value \bar{y} , take this mean value \bar{y} as the threshold, select the ordinate of the minima point that is smaller than the \bar{y} , record their horizontal and vertical coordinates (x_i, y_i) . These points are the effective minima points;
- (4) Add the start point and end point of the raw signal to the coordinates of the effective minima point as the interpolation point. We then sample evenly between the start point and the end point of this set of interpolation points to generate a series of sampling points equally distributed between the two endpoints, which ensures that a smooth curve can be obtained when these points are interpolated;
- (5) The expression for the cubic spline interpolation function is computed, and then the newly generated point x_{new} is substituted into the expression to compute the interpolation point y_{new} that corresponds to x_{new} . This interpolation point is referred to as the "baseline" of the signal being processed;
- (6) Finally, the raw signal's vertical coordinate y is subtracted from the fitted "baseline" y_{new} to obtain the corrected signal ($\tilde{s} = y - y_{new}$).

3. Results and Discussion

In the previous section, we give the definition of effective minima and detailed steps for generating a smooth baseline using cubic spline interpolation. In this section, we will perform some numerical experiments to verify the effectiveness of the method proposed in this paper.

3.1 Experimental Data Collection

We sampled 30 groups of samples using alphapec instrument portable Fourier transform infrared spectrometer ALPHAPEC5010 (spectral range: $500\text{ cm}^{-1} \sim 5000\text{ cm}^{-1}$, resolution up to 1 cm^{-1} , signal-to-noise ratio: 45000:1, transmittance accuracy better than 0.5%T) and used them as test data for spectral data.

3.2 Assessment Methods

In order to comprehensively assess the effectiveness of different baseline correction methods, we use three aspects as judging criteria: direct spectrogram comparison, quantitative analysis of substance recognition accuracy, and the industry's widely recognized root mean square error (RMSE). In this section, we will analyze the effect of baseline correction under each of the three criteria.

3.2.1 Spectrum Comparison

In order to validate the optimal baseline correction method, we conducted rigorous experiments using five methods on 30 sets of test data. Only two representative sets of spectra are given in this subsection to visually show the significant changes between the raw and corrected spectra. Firstly, let us focus on the test data set of salicylic acid, whose raw spectral and fitted baseline (Figure 1) clearly reveal the problem of baseline drift. After baseline correction by different methods, we get Figure 2. Comparing Figure 1 with Figure 2, we can see that the problem of baseline drift is solved after the baseline correction. Secondly, let's turn to the set of test data for coumarin (Figure 3), we can see that the raw spectral also suffers from baseline drift problems. When the baseline is corrected using different methods (Figure 4), the spectra become more stable to some degree. This further confirms the usefulness of the baseline correction when dealing with complicated spectra data.

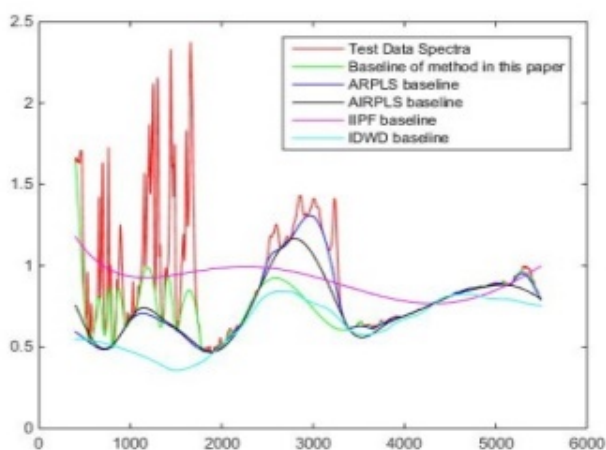


Figure 1. Raw spectra of salicylic acid test data and baselines obtained by fitting with different methods.

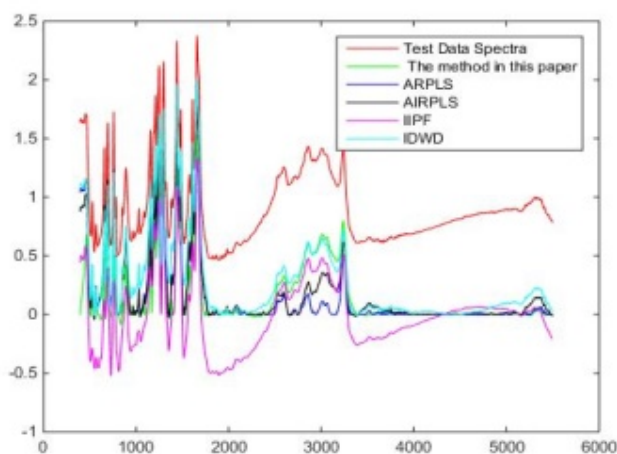


Figure 2. Raw spectra of salicylic acid test data and spectra after baseline correction using different methods.

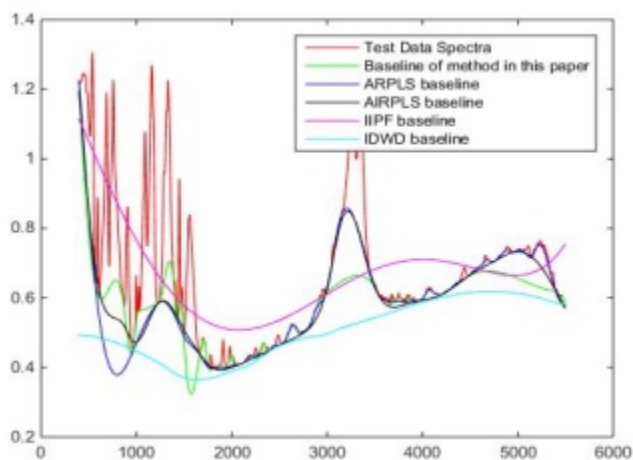


Figure 3. Raw spectra of coumarin test data and baselines obtained by fitting with different methods.

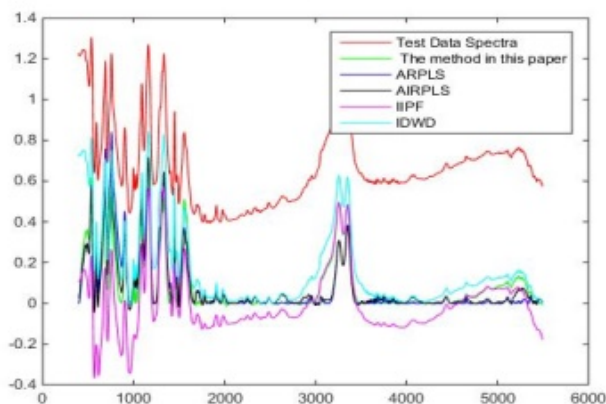


Figure 4. Raw spectra of coumarin test data and spectra after baseline correction using different methods.

By looking closely at Figures 1-4, we can see that ARPLS and AIRPLS mitigate the baseline drift of the raw spectral to some extent. However, in the vicinity of the characteristic peaks, the spectral peaks of the baseline corrected spectra are attenuated. IIPF suffers from the disadvantage of overfitting when fitting the baseline. IDWD eliminates the effect of the baseline to some extent, but the peak is enhanced due to the presence of noise. Therefore, when baseline correction is applied to the raw spectra, using the four baseline correction methods mentioned in the introduction, the spectra show a certain degree of distortion, which directly affects the subsequent qualitative and quantitative analyses.

Compared with the existing methods, the proposed baseline correction method effectively avoids the above problems. The proposed method not only ensures that the positions and areas of the spectral peaks remain unchanged but also significantly enhances the clarity of the characteristic peaks, thus effectively eliminating the baseline drift phenomenon. This improvement has important implications for the accuracy of subsequent qualitative analyses.

3.2.2 Substance Composition Identification Accuracy

We first apply five different baseline correction methods to the same test data. After the baseline correction, we use the processed spectra data for the recognition of substance components. Secondly, the recognized results are compared with the information on the substances stored in the database. Finally, the accuracy rates of the substance composition recognition of the different methods are obtained, and the baseline correction effects of the different methods are compared through the accuracy rates. The database contains a total of 475 substances such as indene, trichloroethane, ibuprofen, butylene glycol, bilirubin, methyl acrylate, quinone, propylene diacetate, benzene tetracarboxylic acid, phenylacetic acid. In the process of identifying the composition of substances, in order to avoid the influence of other factors, we maintain the same noise reduction treatment and give the same similarity definition standard. Next, we give in Table 1 the experimental results for the identification of 30 substance components and calculate the corresponding accuracies, which clearly show the identification effect of each baseline correction method.

Table 1. Number and percentage of substance identifications

Method	AIRPLS	IDWD	IIPF	ARPLS	This Paper Gives the Method
Number of recognitions	1	10	17	18	30
Percentage	3.33%	33.3%	56.67%	60%	100%

From the data in the table, we can observe that the accuracy of substance composition identification after baseline correction using AIRPLS was the lowest at 3.33%. The accuracy of substance composition identification after baseline correction using IDWD is still relatively low although it reaches 33.3%. The baseline correction results using IIPF showed a certain accuracy, and the accuracy of substance composition identification reached 56.67%. The baseline corrected accuracy of ARPLS was the most accurate of the four methods, at 60%, but was still subject to some space to be improved. In contrast, the accuracy of substance composition identification after baseline correction of the proposed method reaches 100%, which is much higher than the above four methods. This reflects the significant advantage of the method in the aspect of material composition identification. Therefore, the proposed method effectively reduces the negative effects caused by noise and other interfering factors, indicating that it has good results after the baseline correction of spectral data.

3.2.3 Root Mean Square Error (RMSE)

In the industry, RMSE is often used to evaluate the effect of baseline correction, which is an important measure of the degree of difference between predicted and actual values [20]. The RMSE takes into account the prediction error for each data point, which allows for a better assessment of the fitting accuracy of the baseline correction model. Therefore, it can be used to evaluate the baseline correction effect. It should be noted that the smaller the RMSE, the better the baseline correction effect. Specifically, the RMSE is defined as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y(i) - s(i))^2}, \quad (12)$$

where y is the raw spectral of the test data, S is the spectral after baseline correction, and N is the number of samples.

We selected 4 representative sets of test data from 30 sets of test data. The RMSE for the baseline correction of the four sets of test data using the five methods described above is shown in the following table.

Table 2. RMSE for four substances for five methods

Name of substance Method	sarcosine	salicylate	benzenesulfonamide	coumarin
	IIPF	0.020114	0.013314	0.009681
AIRPLS	0.018303	0.011048	0.007517	0.014955
ARPLS	0.018290	0.011049	0.007742	0.014884
IDWD	0.015673	0.011056	0.007740	0.015028
The proposed method	0.018262	0.011025	0.007736	0.014872

By analyzing the data in Table 2, we can clearly find that IIPF performed best when processing coumarin, but the worst when processing the other three samples, indicating that its stability was insufficient. AIRPLS and IDWD have similar problems, but they are more stable than IIPF. Although ARPLS was more stable in the treatment of these four samples, its results were all in the middle and late streams, so the overall performance was not excellent. The proposed method performed best in treating salicylic acid and performed equally well in treating the other three samples, which were all ranked second. Therefore, the proposed method not only has good stability but also are comparable with other existing methods in RSME values.

In summary, through comparative analysis and practicality assessment, we can conclude that: the method proposed in this paper can not only effectively reduce the baseline error in removing the baseline of Fourier transform infrared spectroscopy, but also show high practicability and reliability in practical application.

4. Conclusion

In this paper, a new baseline correction method is introduced, and it was also analyzed through three criteria to compare with the existing four methods. From the study of the two judging criteria of spectrogram and substance composition identification accuracy, we found that the baseline correction method proposed in this paper is significantly better than other methods in improving the clarity of the spectrogram and accurately identifying substance compositions. In addition, based on the RSME criterion, the proposed method exhibits greater stability. The results under these three criteria show that the proposed method has a decent practicality and superiority in baseline correction.

Funding

This paper is supported by the Natural Science Foundation of China Three Gorges University Horizontal Program (HZ2023215).

References

- [1] Peng JT, Peng SL, Xie Q, et al. Baseline correction combined partial least squares algorithm and its application in on-line Fourier transform infrared quantitative analysis. *Analytica Chimica Acta*. 2011;690(2):162-168.

- [2] Chen NP. Review on Identification of Wine by Fourier Transform Infrared Spectroscopy. *Modern Chemical Research*. 2019;(06):22-24.
- [3] Chen YY, Zou CN, Mastalerz M, et al. Applications of Micro-Fourier Transform Infrared Spectroscopy (FTIR) in the Geological Sciences—A Review. *International Journal of Molecular Sciences*. 2015;16(12):30223-30250.
- [4] Jiang A. *Research on the Qualitative Analysis Method of Complex Mixtures Based on Infrared Spectroscopy*. Shanghai: Chinese Academy of Sciences. 2012.
- [5] Wang L, Li DM, Qian HJ, et al. Baseline drift and calibration methods in NIR analysis. *Analysis Laboratory*. 2016;35(10):1203-1208.
- [6] Hu YG, Zhao ZY, Wang G. Baseline correction and background elimination using wavelet transforms. *Journal of Huazhong University of Science and Technology (Nature Science Edition)*. 2011;39(06):36-40.
- [7] Ma H, Wang ZB, Zhang JL, et al. Infrared Spectrum Baseline Correction Method Based on Improved Iterative Polynomial Fitting. *Laser Technology*. 2013;37(02):223-226.
- [8] Jiang XY, Li FS, Wang QY, et al. Comparison of the Spectral Baseline Correction Based on the Penalized least squares algorithm. *Transactions of the Chinese Society for Agricultural Machinery*. 2021;52(08):205-212.
- [9] Zhang ZM, Chen S, Liang YZ, et al. Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*. 2010;135(5):1138-1146.
- [10] Peng J, Peng S, Xie Q, et al. Baseline correction combined partial least squares algorithm and its application in on-line Fourier transform infrared quantitative analysis. *Analytica Chimica Acta*. 2011;690(2):162-168.
- [11] Baek SJ, Park A, Ahn YJ, et al. Baseline correction using asymmetrically reweighted penalized least squares smooting. *Analyst*. 2015;140(1):250-257.
- [12] Zhang F, Tang X, Tong A, et al. Baseline correction for infrared spectra using adaptive smoothness parameter penalized least squares method. *Spectroscopy Letters*. 2020;53(3):222-233.
- [13] Zhao FK, Xu XM, Lv LY, et al. A Background Removing Method for Spectrum Signal Based on Iterative Wavelet Transform. *Journal of Instrumental Analysis*. 2019;38(10):1275-1279.
- [14] Ning ZQ, Liu JX, Wu Y, et al. Infrared Spectrum BASELINE Correction Method Based on Improved Iterative Polynomial Fitting. *Laser & Optoelectronics*. 2020;57(03):255-261.
- [15] Gan F, Ruan G, Mo J. Baseline correction by improved iterative polynomial fitting with automatic threshold. *Chemometrics and Intelligent Laboratory Systems*. 2006;82(1):59-65.
- [16] Ma Z, Ma E, Xiong FB, et al. Dynamic Moving Savitzky-Golay Smoothing Algorithm for the Raman Spectrum Automatic Baseline Correction. *Applied Laser*. 2017;37(04):614-618.
- [17] Zhao H, Chen YX, Xu XD, et al. Baseline Correction for Raman Spectra Based on Locally Symmetric Reweighted Penalized Least Squares. *Chinese Journal of Lasers*. 2018;45(12):280-291.
- [18] Guo Y, Wang W. Baseline correction for Raman spectra using a spectral estimation-based asymmetrically reweighted penalized least squares method. *Applied Optics*. 2023;62(18):4766-4776.
- [19] Nie L, Tao M, Zhai ZS, et al. Undersampled modified algorithm for cylindrical near-field measurement based on cubic spline interpolation. *Journal of Hubei University of Technology*. 2017;32(01):63-67.
- [20] Chat T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 2014;7(3):1247-1250.