Hindawi Journal of Spectroscopy Volume 2020, Article ID 8920732, 14 pages https://doi.org/10.1155/2020/8920732



Research Article

Rapid Identification of Atmospheric Gaseous Pollutants Using Fourier-Transform Infrared Spectroscopy Combined with Independent Component Analysis

Wei Ju, 1 Changhua Lu, 1,2 Chun Liu, 3 Weiwei Jiang, 1 Yujun Zhang, 2 and Feng Hong

¹School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, Anhui 230009, China ²Anhui Institute of Optics Fine Mechanics, Chinese Academy of Sciences, Hefei, Anhui 230031, China ³School of Electrical Engineering and Automation, Hefei University of Technology, Hefei, Anhui 230009, China

Correspondence should be addressed to Chun Liu; dqlch03@hfut.edu.cn

Received 3 September 2019; Revised 7 November 2019; Accepted 15 November 2019; Published 22 January 2020

Academic Editor: Alessandra Durazzo

Copyright © 2020 Wei Ju et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fourier-transform infrared (FTIR) spectroscopy is a rapid and nondestructive technology for monitoring atmospheric quality. The identification of each component from the FTIR spectra is a prerequisite for the accurate quantitative analysis of gaseous pollutants. Due to the overlap of different gas absorption peaks and the interference of water vapor in the actual measurement, the existing identification methods of gas spectra have drawbacks of low identification rate and the inability to carry out real-time online analysis in atmospheric quality monitoring. In this work, independent component analysis (ICA) is applied to the spectral separation of heavily overlapped spectra of gaseous pollutants. The proposed method is validated by the analysis of mixture spectra obtained in laboratory and actual atmospheric spectra collected from stationary source. The average time consumption of separation process is less than 0.2 seconds, and the identification rate of experimental gases is up to 100%, as shown by the results of peak searching and the analysis of the correction coefficient between the separated spectra and the standard spectra database. The identification results of actual atmospheric spectra demonstrated that the proposed method can effectively identify the gaseous pollutants whose concentration changes in the measured spectra, and it is a promising qualitative spectral analysis tool that can shorten the identification time, as well as increase the identification rate. Therefore, this method can be a useful alternative to traditional qualitative identification methods for real-time online atmospheric pollutant detection.

1. Introduction

Air is fundamental for living organisms on the Earth, and the quality of the atmospheric environment is closely related to human activities. With the development of modern industrialization and the improvement of human living standards, combustion of fossil fuels and emission of automobile exhaust cause severe atmospheric pollution, which has raised more concern in recent years. Typical gaseous pollutants include nitrogen oxides (NO_X), sulfur dioxide (SO_X) [1], volatile organic compounds (VOCs) [2], and ozone (O₃) [3]. These gaseous pollutants exist as trace gases in the atmosphere but have detrimental effects on human health and the environment [4]. The health effects include cancer, increased hospitalizations, and reduced life

expectancy from heart and lung diseases. For example, even short exposure to SO_2 can cause serious respiratory illnesses, such as asthma and bronchitis [5]. Moreover, two-thirds of VOCs are carcinogenic and harmful to the central nervous system and hematopoietic system, especially the immune system [6]. The environmental effects are mainly reflected in the greenhouse effect, ozone layer destruction, and acid rain [7]. With atmospheric pollution–caused incidents become increasingly serious, local governments are attaching more attention to atmospheric pollution, and laws and regulations aiming at controlling this have been promulgated. Despite this, some industrial chemical parks still secretly discharge toxic and high-concentration exhaust gases directly into the air from time to time. Therefore, the establishment and improvement of a real-time continuous gaseous pollutant

monitoring system that can confirm the source and concentration of specific pollutants are essential for the control of atmospheric pollution.

The advanced online atmospheric monitoring technologies, include gas chromatography-mass spectrometry (GC-MS) [8], proton transfer reaction-mass spectrometer (PTR-MS) [9], and chemical ionization reaction mass spectrometry (CIRMS) [10, 11]. Among these, GC-MS is most widely used in commercial monitoring instruments, and it can rapidly identify and quantify pollutants from atmospheric samples [12]. PTR-MS and CIRMS are improvements of mass spectrometry and have the advantages of high sensitivity and continuous detection. Although these technologies have been proven to be able to detect gaseous pollutants accurately, their implementation is complex and requires reagent assistance. Being reagent free, nondestructive, and providing continuous and simultaneous multicomponent measurement, optical detection has become a new trend in the development of atmospheric monitoring [13].

Fourier-transform infrared (FTIR) spectroscopy is a powerful spectral detection technology that has been recommended by the US Environment Protect Agency for monitoring air pollutants. FTIR spectra are composed of absorption peaks generated from infrared radiation absorption during the vibration transition of asymmetric dipole moment polyatomic molecules, and a wide variety of gaseous pollutants can be measured by FTIR technology due to their physical structures. FTIR has high sensitivity, permitting the detection of changes in gas concentration at the ppb (parts per billion, volume concentration) level. Openpath FTIR have been applied to the online detection of trace gas emissions from forest fires [14] and soil during the spring thaw [15], and from traffic emissions along highways [16] and in harbors [17]. The identification of each component in the FTIR spectrum is crucial for accurate quantitative analysis. In the early stage of FTIR spectrum identification, library searching (LS) methods [18] were adopted when the spectral components were single or the absorption peaks rarely overlapped. Lavine et al. [19] used pattern recognition methods as the search prefilters for library searching and made the identification more precise and faster. Due to the strong abilities of curve fitting and anti-interference of the artificial neural network (ANN), a variety of ANN-based FTIR spectrum identification algorithms have been proposed. Yang and Griffiths [20] encoded reference FTIR spectra of five alcohols as prototype vectors for the Hopfield network and identified each of the five alcohols correctly from more than 100 spectra of different compounds through this network. At present, the most common methods used in practice combine the calibration spectra from the spectral database and chemometrics methods for quantitative as well as qualitative analyses. Griffith [21] calculated the calibration spectra from the high-resolution transmission molecular absorption (HITRAN) database and retrieved the gas information from FTIR spectra by the nonlinear least squares algorithm. Traditional component identification method of open-path FTIR spectra is an artificial identification, which needs a professional spectroscopist to determine the possible components that contained in the observed spectrum, then

subtracting the corresponding components from the observed spectrum until a relatively flat baseline is produced [22]. These qualitative identification methods and their improvements can effectively identify spectra in specific occasions; however, in the case of serious spectral overlap, small sample size, absence of reference spectra, and high real-time requirements, it is difficult to accurately recognize trace gases from FTIR spectra.

Independent component analysis (ICA) is a higher-order statistical model-based signal processing method that can separate statistically independent source signals from a set of observed signals without any prior knowledge [23]. ICA has been widely applied to blind source separation (BSS) fields such as speech separation [24], biomedical signals analysis [25], and facial recognition [26, 27]. As a novel unmixing method that can recover the original information from observed overlapping mixtures, ICA has attracted great interest in spectral analytical chemistry [28]. Pasadakis and Kardamakis [29] used ICA to identify constituents in commercial gasoline from its FTIR spectra. Al-Mbaideen and Benaissa [30] investigated ICA combined with multiple linear regression to determine the concentration of glucose from its near-infrared spectra. Yu et al. [31] utilized ICA and wavelet analysis to discriminate the seriously overlapped three-dimensional fluorescence spectra of water pollutants. Monakhova et al. [32] presented the effectiveness of four different ICA deconvolution algorithms for overlapped NMR spectra of eight component mixtures of foods and related products. The studies showed the accuracy and robustness of ICA method in recovering overlapped spectra. However, due to the complexity of atmospheric gaseous compositions and water vapor interference in the actual measured spectra, there are few studies where ICA is applied to the component identification of atmospheric gaseous pollutant spectra.

2. Materials and Methods

In this study, we conducted identification experiments on the spectra collected both in the laboratory and in the actual atmosphere. The infrared spectra were collected by extraction FTIR gas spectrometer independently developed by Anhui Institute of Optics and Fine Mechanics CAS, with a multipass gas cell of a 10 m optical path inside. The spectrometer has a measuring bandwidth of $500\,\mathrm{cm}^{-1}$ – $5000\,\mathrm{cm}^{-1}$ and spectral resolution of $1\,\mathrm{cm}^{-1}$. The output spectrum of the spectrometer is the average value after 16 scans.

2.1. Materials. Methane (CH₄), ethanol (C₂H₅OH), ethyne (C₂H₂), and ethylene (C₂H₄) are common air pollution gases. Because of their low cost, simple preparation, and seriously overlapping infrared absorption bands, we selected these four gases as the experimental gases. The infrared spectra of experimental gases from standard spectral database is shown in Figure 1; as we can see from the figure, the absorption bands of experimental gases are overlapped heavily in the waveband of $860-1500\,\mathrm{cm}^{-1}$ and $2800-3200\,\mathrm{cm}^{-1}$. The pure gases were produced by Hefei

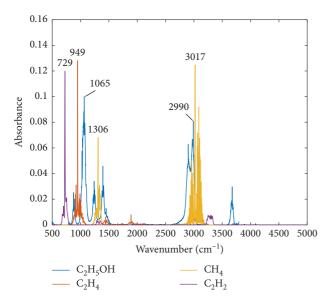


FIGURE 1: Infrared spectra of four gases with $1\,\mathrm{cm}^{-1}$ spectral resolution.

Ningte Gas Company and separately stored in 4L sealed air cylinders. The initial concentrations of the four gases were 65 ppm (CH₄), 207 ppm (C₂H₅OH), 506 ppm (C₂H₂), and 609 ppm (C₂H₄). We obtained target gases of different concentrations through a four-channel high-precision gas distribution system independently developed by Hefei Institute of Physical Science, Chinese Academy of Sciences (CAS). The gas distribution platform can simultaneously distribute four gases and get the required concentration by adjusting the volume ratio of the auxiliary gas (nitrogen) and the pure gas. In total, we got 12 samples of mixed gases, the concentration changes of four gases in different samples are mutually independence, this is a necessary condition for ICA separation. In order to simulate the composition of pollutants in the atmosphere, the concentrations of the gases were controlled below 60 ppm.

2.2. FTIR Spectra Acquisition of Experimental Gases. Before spectra acquisition, the gas storage cylinders were placed in a ventilation room to prevent injury caused by dangerous gas leakage, the experimental system and the connecting pipes were completely sealed, and a tightness test was performed. The four gases and nitrogen were configured to the concentrations required by the software built into the gas distribution platform. The spectra were recorded when the gas flow and barometric pressure were stable. The experimental setup for gas mixture spectra acquisition is shown in Figure 2. Spectra obtained from the spectrometer include not only the absorption information of target gases but also random noise; in this experiment, Savitzky-Golay smoothing filter was used for raw spectra denoising. Figure 3 is the denoised spectra of 12 groups mixed experimental gases.

In addition to the absorption peaks of experimental gases, water vapor and carbon dioxide also have strong absorption bands in raw spectra. H₂O (g) and CO₂ produce

serious interference in the analysis of atmospheric infrared spectra. As we can see from Figure 3, H_2O (g) has strong and wide absorption peaks in the range of $1350~\rm cm^{-1}$ – $1800~\rm cm^{-1}$ and $3500~\rm cm^{-1}$ – $3900~\rm cm^{-1}$, CO_2 gets strong absorption peaks in the range of $2280~\rm cm^{-1}$ – $2385~\rm cm^{-1}$. When using FTIR to analyze gaseous components, the common method to avoid the interference of H_2O (g) is to analyze the gaseous components in the "infrared atmospheric window" that is far away from the H_2O (g) absorption wavebands. Based on this idea, we remove the strong absorption bands of H_2O (g) and CO_2 from the spectra and connect the rest of the waveband together. The elimination of interference of water vapor and carbon dioxide before implementing the identification procedure can greatly reduce the complexity of analysis.

2.3. FTIR Spectra Acquisition of Actual Atmospheric. The atmospheric spectra were measured on the roof of student apartment (3 floors) of the Hefei University of Technology; this area is located near the center of the city. There is no large-scale chemical industry source, and the atmospheric gaseous pollutants mainly come from the exhaust gas of motor vehicles. Vehicle exhaust contains a variety of toxic and harmful gases, including CO, CO2, and VOCs like benzene, alkenes, and aromatic hydrocarbon. Because these VOCs all contain hydrocarbon bond functional groups, they have obvious absorption peaks in the range of 2800 cm⁻¹-3200 cm⁻¹, and these absorption peaks are similar in spectral profile; hence, it is difficult for traditional identification methods to effectively distinguish the specific component information in the overlapped spectrum. In this study, spectral data in the region from 2800 cm⁻¹ to $3200\,\mathrm{cm}^{-1}$ were used for components identification. Figure 4 is one of the measured atmospheric spectra.

Figure 5 shows 12 measured spectra in the range of 2800 cm⁻¹ to 3200 cm⁻¹. We can see that the actual atmospheric spectra have a relatively small variation in absorption amplitude, and it is difficult to distinguish the specific gases from the absorption peaks except methane.

2.4. Background Spectra Elimination. The background spectrum is caused by the response of infrared detector to external radiation, and it is a relatively slow change process, which can be considered as a slow change baseline superimposed on the absorption spectrum. In FTIR applications, the background spectrum can be obtained by measuring the spectrum without the absorption of the target component. It is easy to achieve this under laboratory conditions, but it is difficult in pollution gases monitoring by open-path FTIR. In order to suppress the effect of background spectra, Monte Carlo simulation is performed to get the new spectral samples from experimental spectra. Two spectra are randomly selected, and one is divided from the other to produce a new spectrum. For the 12 groups of mixed gases spectra obtained in the laboratory, we get a total of 66 different spectra. However, 12 spectra are enough to form a sample set. Finally, 12 spectra of experimental gases and 12 spectra of actual atmosphere are obtained by the background spectra elimination method. Unlike the background removal in the 4

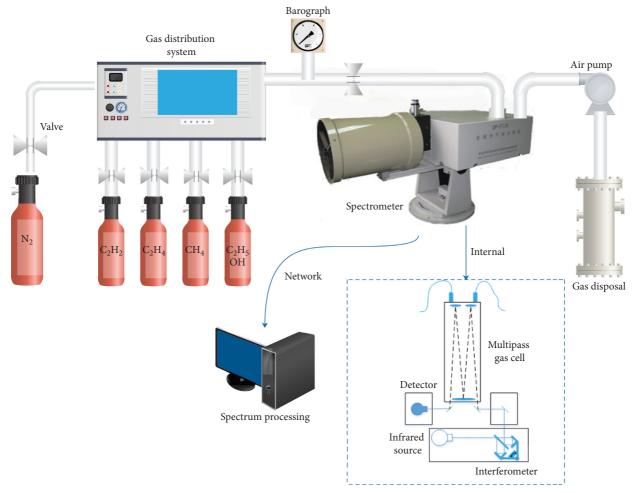


Figure 2: Experimental setup for mixture spectra acquisition.

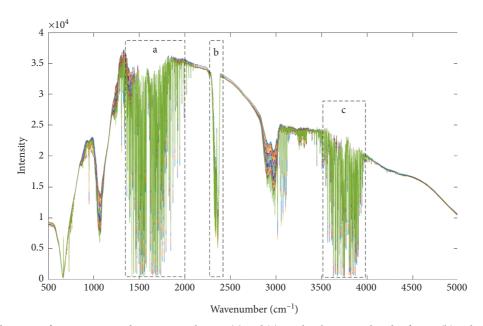


FIGURE 3: Denoised spectra of 12 groups mixed experimental gases. (a) and (c) are the absorption bands of H_2O . (b) is the absorption band of CO_2 .

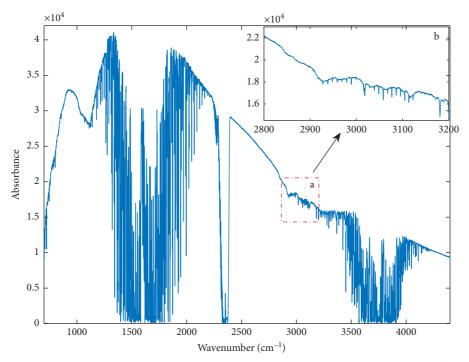


FIGURE 4: Denoised spectrum of actual atmosphere. (a) The spectrum in the waveband of 2800 cm⁻¹ to 3200 cm⁻¹. (b) The magnification of (a).

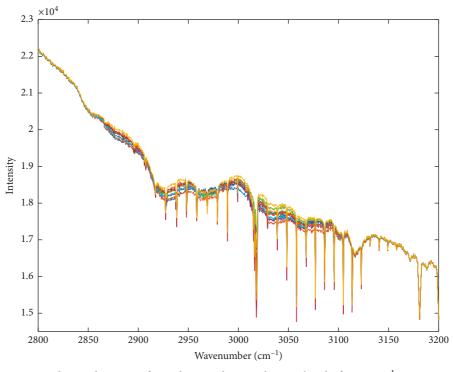


Figure 5: 12 denoised spectra of actual atmosphere in the waveband of $2800\,\mathrm{cm}^{-1}$ to $3200\,\mathrm{cm}^{-1}$.

ordinary FTIR process, we are not seeking the pure background, and the new spectra not only eliminate the non-linear part of the background but also keep the useful information of target gases.

2.5. ICA. ICA is a powerful signal processing tool, which can restore the original source signal from a set of observed mixed signal; it aims at finding a separation matrix that can transform the observed signal into a linear combination of

mutually independent components, which are close to the source signal.

The noise-free ICA model can be expressed by the following linear relationship:

$$X = AS, (1)$$

where $\mathbf{X} = (x_1, x_2, \dots, x_m)^T$ denotes the observed m-dimensional signal vector, $\mathbf{S} = (s_1, s_2, \dots, s_n)^T$ is an n-dimensional source component vector, and A denotes the unknown mixing matrix. ICA attempts to find the optimal separation matrix \mathbf{W} to make the components of \mathbf{Y} as independent as possible.

$$\mathbf{Y} = \mathbf{M}\mathbf{X} = \mathbf{W}^T \mathbf{X}.\tag{2}$$

Because the source signals and mixing information are unknown, the implementation of ICA requires two important restrictions [33]:

- (1) The components of the source signal need to be statistically independent.
- (2) At most one component of the source signal obeys the Gaussian distribution.

The concentration and existence of different gaseous pollutants in the atmosphere rarely appear in proportion to each other, and the spectral signal is not subject to the Gaussian distribution. Therefore, ICA can be used in the separation of spectral signals of atmospheric gaseous pollutants.

Among various kinds of ICA approaches, fixed-point ICA (FastICA) has become one of the most classical algorithms due to its fast operation speed and high accuracy. FastICA was proposed by Hyvärinen and Oja [34] in 1997, and it adopted a fixed-point iteration scheme to make the separation process 10 to 100 times faster than traditional ICA optimization algorithms. The optimized algorithm of traditional ICA is based on stochastic gradient methods, which can be achieved by neural network learning [35]. The main problems of neural network learning are slow convergence and high dependence on the selection of learning rate parameters. In contrast, FastICA gets a cubic convergence speed and does not rely on any custom parameters. In this study, FastICA was used as the ICA approach for rapid spectra separation process.

3. Results and Discussion

3.1. Establishment of the Linear Model. ICA can efficiently separate original signals from the mixed system, which is linear. We need to make sure that the system is under the assumption of a linear mixture model before we do the analysis. The Beer–Lambert law and absorbance additivity describe the relationship between the absorption intensity of multicomponent spectra and the concentration of each light-absorbing substance. The basic principles can be described as follows:

$$A_{\text{total}}(\nu) = a_1(\nu)c_1l + a_2(\nu)c_2l + \dots + a_m(\nu)c_ml,$$
 (3)

where $A_{\text{total}}(\nu)$ is the total absorption intensity of the spectrum at a certain wavelength ν , $a_i(\nu)$ is the absorption

coefficient, c_i is the concentration of component i at wavelength ν , and l is the optical path length. According to the absorbance additivity, the absorbance of multicomponent spectrum is the sum of absorbance of each component, and the change of concentration of the component will only cause the change of absorbance amplitude but not the position of absorption peaks and the spectral profile. From the above analysis, we can know that the absorption intensity of atmospheric spectra conforms to the linear model, and the ICA algorithm can be used to separate the spectra of mixed gases.

In Section 2, we obtained the transmittance spectra from FTIR spectrometer. The transmittance information should be converted into absorbance information so that the separation algorithm can analyze the experimental data. The Beer–Lambert law also states the relationship between transmittance $T(\nu)$ and absorbance $A(\nu)$:

$$A(\nu) = \log_{10} \frac{1}{T(\nu)}. (4)$$

Thus, we get the absorption spectra by taking the logarithm of the reciprocal transmittance data. The calculated 12 absorption spectra of experimental gases are shown in Figure 6, which will be used as the input of the separation algorithm after the discrete wavebands are connected to each other. The calculated 12 absorption spectra of actual atmosphere in waveband of 2800 cm⁻¹–3200 cm⁻¹ are shown in Figure 7, which can be directly used as the input of the separation algorithm.

3.2. Identification Results of Experimental Gases Spectra. All the calculation was performed on Matlab R2016 (The Math Works, Natick, MA, USA). We divided the 12 calculated absorption spectral data of experimental gases into three groups, and ICA decomposition was carried out on each group to get the separated spectra; the component identification results were obtained by the analysis of separated spectra.

Figure 8 shows the separated spectra of three groups. The separated spectra are restored to the original waveband positions.

The identification of separated spectra is achieved through peak searching and comparing the similarity among separated spectra and the standard spectral database. Table 1 lists the infrared absorption characteristics of the four experimental gases from standard spectra database.

Because the spectrum may produce wavenumber drift in the actual measurement process, it is necessary to set an offset of peak position for peak searching. Generally, the wavenumber drift will not be greater than $1 \, \mathrm{cm}^{-1}$, and therefore, we set the peak position offset to $1 \, \mathrm{cm}^{-1}$.

Table 2 shows the peak searching results of the separated spectra; the peak positions are obtained from the maximum value of absorption bands. By sorting the absorption peaks, the positions of the first two peaks are listed in the table. The gas (or gases) that the spectrum may correspond to is discriminated by the position of the strongest peak, and the sub-strong peak is used to assist this discrimination.

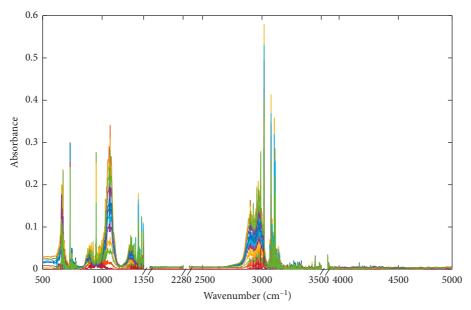


FIGURE 6: The spectral samples for ICA separation.

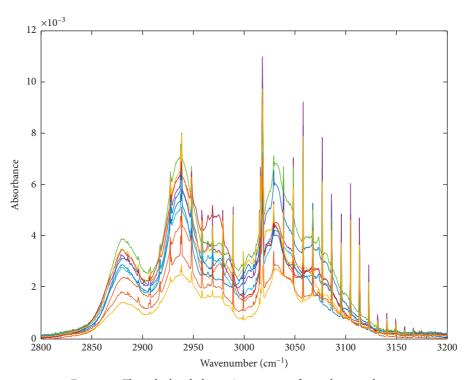


Figure 7: The calculated absorption spectra of actual atmosphere.

The peak searching results show that most of the possible gases can be recognized by the strongest absorption peak of the separated spectrum, especially the spectrum with a single absorption peak like C_2H_2 and CH_4 . In the first separated spectrum of group (b), there are two absorption peaks with a similar amplitude, corresponding to two possible gases. In this case, the comparison of spectra is necessary for accurate identification.

Based on the results of peak searching, the final identification is completed by analyzing the similarity between the possible components and the corresponding standard spectrum. The correlation coefficient is a measure of the closeness degree between two column vectors and is used here for spectra similarity discrimination. The correlation coefficient r between the standard absorbance spectrum x and the spectrum y to be identified is as follows:

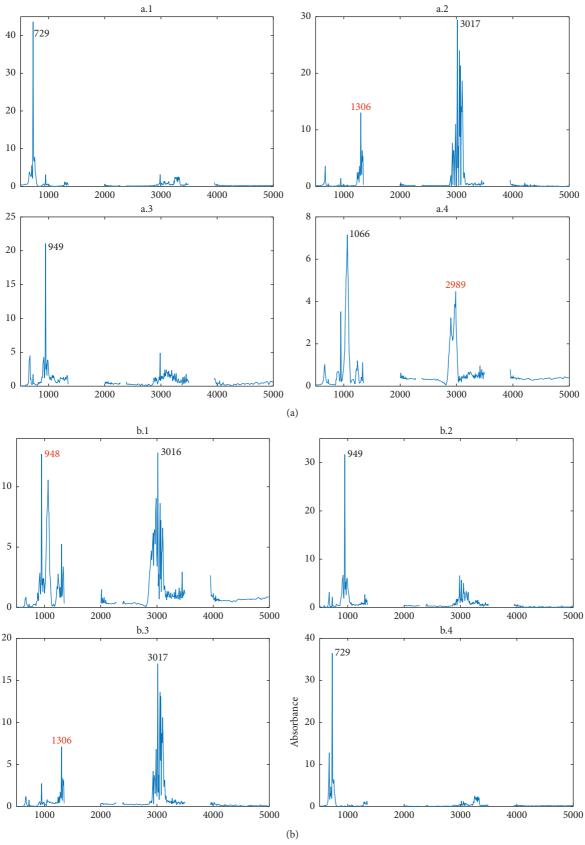


FIGURE 8: Continued.

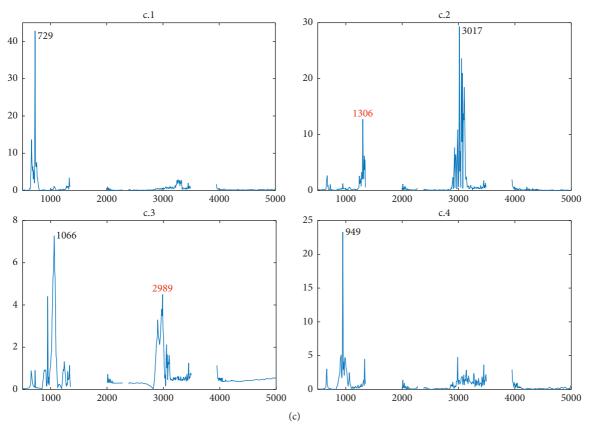


FIGURE 8: Separated results of three groups of experimental gases spectra. X axis is wavenumber (cm $^{-1}$), and Y axis is calculated absorbance; black number is the position (cm $^{-1}$) of first absorption peak in the single separated spectrum, and red number is the second absorption peak.

Table 1: Infrared absorption characteristics of the four gases.

Name	Formula	CAS no.	Absorption band (cm ⁻¹)	Absorption peak (cm ⁻¹)	Absorption band (cm ⁻¹)	Absorption peak (cm ⁻¹)
Acetylene	C_2H_2	74-86-2	705-760	729.342	_	_
Ethylene	C_2H_4	74-85-1	860-1040	949.533	_	_
Ethanol	C_2H_5OH	115-10-6	1050-1240	1065.243	2760-3040	2990.539
Methane	CH_4	74-82-8	1190-1400	1306.492	2870-3200	3017.271

TABLE 2: Results of peak searching.

No. of input spectra	No. of separated spectra	Position of the first peak (cm ⁻¹)	Position of the second peak (cm ⁻¹)	Potential corresponding gas	
	1	729	_	C_2H_2	
1	2	3017	1306	CH_4	
1	3	949	_	C_2H_4	
	4	1066	2989	C_2H_5OH	
	1	3016	948	C ₂ H ₅ OH, C ₂ H ₄	
2	2	949	_	C_2H_4	
2	3	3017	1306	CH_4	
	4	729	_	C_2H_2	
	1	729	_	C_2H_2	
3	2	3017	1306	CH_4	
3	3	1066	2989	C_2H_5OH	
	4	949	-	C_2H_4	

$$r = \frac{\sum [(x_i - \overline{x}) * (y_i - \overline{y})]}{\sqrt{\sum (x_i - \overline{x})^2} * \sqrt{\sum (y_i - \overline{y})^2}}, \quad i = 1, 2, 3, \dots$$
 (5)

where N is the number of data points in the spectra, and \overline{x} and \overline{y} are the mean values of x and y. $-1 \le r \le 1$ can be obtained from the definition. According to the previous study [36], the two spectra match well when r > 0.9. A further criterion is required for the spectral identification when $r \le 0.8$. The calculated results of correlation coefficients are presented in Table 3.

All values of r are greater than 0.8, which means that the separated spectra match well with the standard spectra. The highest correlation is observed with CH₄, whose r values are all greater than 0.98. Meanwhile, the lowest correlation is observed with C_2H_5OH , whose r values are in the range of 0.8259–0.9213. This may be because the absorption bands of CH₄ do not overlap with the absorption bands of H₂O and CO_2 , whereas the absorption bands of C_2H_5OH are complex and part of them overlap with the absorption band of H_2O , which is removed before the separation. The first spectrum of Figure 8(b) has two corresponding gases according to peak searching, and the spectrum can be identified by calculating the correlation coefficient between this spectrum and the standard spectrum of the two possible gases. The r value is 0.4126 for CH₄ and 0.8259 for C₂H₅OH; therefore, it is easy to judge that the spectrum is strongly related to C₂H₅OH. The peak searching and correlation coefficients calculated results show that all the four experimental gases can be correctly recognized in three groups of experiment, and the identification rate of the mixed gases spectra is 100%.

From the provided figures, we can see that the separation results are not completely consistent with the source spectra in the standard database, and the separated spectra may have redundant absorption peaks related to interfering gases. This may be attributed to the incomplete removal of $\rm H_2O$ absorption wavebands. Although redundant absorption peaks exist, their amplitudes are small, and thus, do not affect recognition results.

3.3. Identification Results of Actual Atmospheric Spectra. The number of sources, also known as the number of independent components (ICs), needs to be evaluated before ICA decomposition. The number of observed signals, as the input of ICA, should be equal to or greater than the actual number of components in the mixture; otherwise, the ICA results will have error. There are many researches about how to evaluate the number of components in chemical mixture [37-39]. The number of components exist in the actual atmosphere can be evaluated by the singular value decomposition of the denoised spectral matrix, and the number of components is equal with the number of singular values that are higher than a certain threshold. Because the singular vector contains more useful information than the singular value, it is more accurate to judge the source number through the singular vector. Therefore, the specific source number of the atmosphere spectra can be determined comparing the singular vector waveform after

TABLE 3: Correlation coefficients between separated spectra and standard spectra.

	C_2H_2	CH_4	C_2H_4	C ₂ H ₅ OH
Group 1	0.9531	0.9858	0.9381	0.9115
Group 2	0.9487	0.9885	0.9374	0.8259
Group 3	0.9413	0.9812	0.9401	0.9206

decomposition. Figure 9 is the first eight singular vectors obtained from singular value decomposition of the atmospheric spectral matrix; we can see that the first five singular vectors have obvious waveforms, but from the sixth singular vector, they fluctuate slightly near zero. According to the result of singular value decomposition, the number of components in atmospheric spectra can be determined to be 5.

We selected 10 spectra from the 12 calculated absorption spectra data of actual atmosphere and divided the 10 spectra into two groups; ICA decomposition was carried out on each group to get the separated spectra; Figure 10 shows the separated spectra of two groups.

Because the specific gases in the actual atmosphere are unknown, the component identification of atmospheric spectra can be achieved through calculating and comparing the correlation coefficient between the separated spectra and all the gas spectra with absorption peaks in the waveband of $2800 \, \mathrm{cm}^{-1}$ – $3200 \, \mathrm{cm}^{-1}$. The gaseous pollutants in the measurement site were mainly vehicle exhaust, so we compared the separated spectra with the standard spectra of common VOCs in vehicle exhaust, and recorded the gas type corresponding to the maximum value of correlation coefficient. Table 4 lists identification results of separated spectra.

According to the identification results from the correlation coefficient between the separated spectra and the standard spectra, it can be concluded that the five gases existed in the atmospheric were o-xylene (C_8H_{10}), methane (CH₄), toluene (C₇H₈), p-xylene (C₈H₁₀), and n-hexane (C_6H_{14}) , respectively. Compared with the analysis results of build-in software of spectrometer, which are based on the classical least squares algorithm, all of the five gases existed in the atmospheric spectra. At the same time, according to the software analysis result, there were traces of benzene and styrene in the atmosphere, which were not identified by the ICA separation. The reason for this may be that the concentration of the two gases had not changed in the selected spectral samples, so the absorption peaks of the two gases were eliminated by the division operation during the background spectra elimination process.

From the separation results, we can see that ICA algorithms can separate single gas spectrum from heavily overlapped spectra; it can transform the spectra set into independent absorption peaks. The results also show that the sequence and amplitude of the results obtained by the ICA algorithm are uncertain.

3.4. Efficiency of ICA Spectra Separation. There are total 6831 spectral data points in the target waveband of experimental gas spectra, and the number is 830 in actual atmospheric

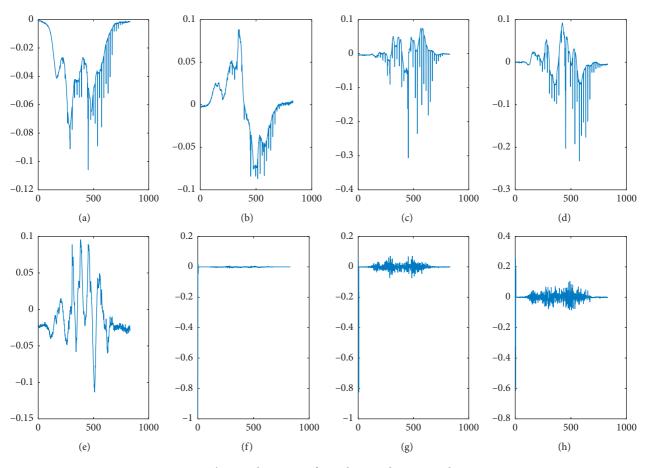


Figure 9: The singular vector of actual atmosphere spectral matrix.

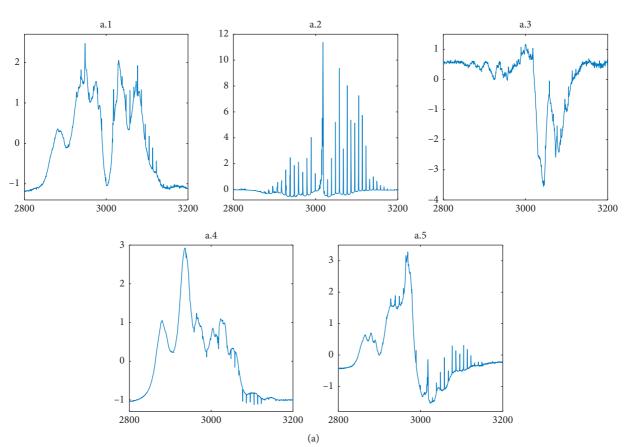


FIGURE 10: Continued.

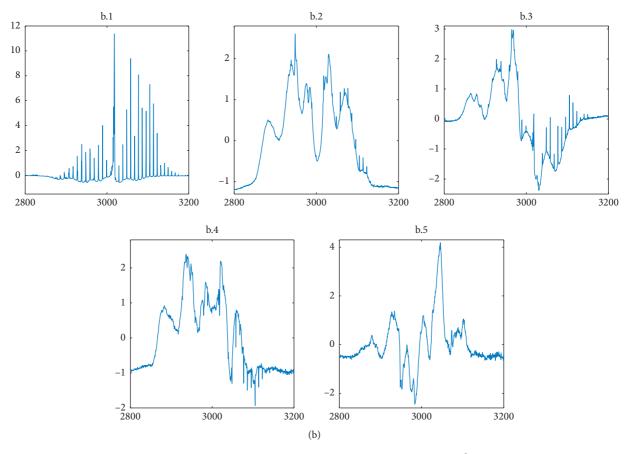


FIGURE 10: Separated results of two groups of actual atmospheric spectra. *X* axis is wavenumber (cm⁻¹), and *Y* axis is calculated absorbance.

Group 1 Group 2 3 5 5 Separated spectrum 1 2 4 1 2 3 4 C_8H_{10} C_7H_8 C_8H_{10} C_8H_{10} C_8H_{10} CH_4 C_6H_{14} CH_4 C_6H_{14} C_7H_8 Possible gas p-xylene o-xylene Methane Toluene p-xylene *n*-hexane Methane n-hexane o-xylene Toluene Correlated coefficient 0.9331 0.9199 0.9562 0.8856 0.7482 0.9644 -0.81230.9161 0.9752 0.8151

TABLE 4: Identification results of separated spectra.

spectra. Table 5 lists the number of iterations and running time of FastICA algorithm for spectral separation.

Table 5 shows operation efficiency of FastICA algorithm. The average running time of ICA separation is less than 0.2 seconds, and it is conducive to the application of signal separation algorithm in online monitoring with high real-time requirements. The identification results of the separated spectra demonstrate that the ICA algorithm is a remarkable tool in the qualitative identification of heavily overlapped mixed gases FTIR spectra. It can rapidly retrieve the qualitative information from complex multicomponent spectra.

Although the ICA algorithm has good performance on extracting the source signal, the separated signals have uncertain scaling factors. Besides, the absorbance of gas is not in direct proportion to its concentrations in practice. Therefore, establishing a quantitative spectral model by multiple linear regression is necessary for obtaining the concentration information of each pollutant. The common

quantitative spectral methods include partial least squares (PLS), classical least squares (CLS), stepwise regression analysis (SRA), and support vector machine (SVR).

It is noteworthy that only several spectral samples are needed to complete the whole identification process by the separation algorithm. Conversely, the neural network algorithm and other identification methods like CLS require multiple samples for modeling. Thus, the proposed method is more suitable in practice when reference substances are not available or are expensive. Moreover, the spectral separation process does not depend on the existing spectral database. When the separated spectra are not included in the spectral database, they may belong to new species. We can thus enrich the spectral database by adding the spectra of new species into the existing database.

The identification results of actual atmospheric spectra show that the ICA method can effectively identify the gaseous pollutants whose concentration changes in the

Spectra type	No. of groups	Data points	Iteration number	Average	Running time (s)	Average (s)
	1		872	630	0.2631	0.1727
Laboratory obtained	2	6831	580		0.1603	
•	3		436		0.0947	
A -t1 -t	1	830	121	112	0.0314	0.02954
Actual atmosphere	2		103		0.0276	

TABLE 5: The number of iterations and running time of FastICA algorithm.

measured spectra. Because the absorption peaks of gas spectrum with constant concentration would be eliminated by the background spectrum elimination method used in this study, we need to focus on the background spectrum elimination method, which can eliminate the background spectrum while retaining all target gas information in the following study, and the selection of proper waveband for ICA separation can reduce the difficulty of spectra identification.

4. Conclusion

Because the absorption peaks of different gas infrared spectra overlap extensively, the traditional atmospheric spectral detection methods have difficulties rapidly and accurately distinguishing specific gas types from overlapping spectra when monitoring atmospheric gaseous pollutants. In this study, we use the source information recovery ability of ICA algorithm to separate single gas spectra from seriously overlapped, multicomponent FTIR spectra. A series of pretreatments are presented to combat the noise and interference produced in the spectral acquisition process, and experiments on laboratory obtained spectra and actual atmospheric spectra verify the effectiveness of ICA algorithm on mixed spectra separation. The proposed method can improve the identification rate as well as the identification speed and enrich the spectral library compared with traditional gaseous spectral identification methods. The whole separation process only needs several mixed spectral samples, and it can make the online atmospheric spectral monitoring and other linear system analysis with high real-time requirements simpler and more effective.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no potential conflicts of interest related to this manuscript.

Acknowledgments

This work was supported by the National Key Scientific Instrument and Equipment Development Project of China (no. 2013YQ220643) and The National 863 Program of China (no. 5702014AA06A503).

References

- [1] N. A. Krotkov, C. A. McLinden, C. Li et al., "Aura OMI observations of regional SO₂ and NO₂ pollution changes from 2005 to 2015," *Atmospheric Chemistry and Physics*, vol. 16, no. 7, pp. 4605–4629, 2016.
- [2] Z. Hong, M. Li, H. Wang et al., "Characteristics of atmospheric volatile organic compounds (VOCs) at a mountainous forest site and two urban sites in the southeast of China," *Science of the Total Environment*, vol. 657, pp. 1491–1500, 2019.
- [3] Y. O. Khaniabadi, M. Daryanoosh, P. Sicard et al., "Chronic obstructive pulmonary diseases related to outdoor PM₁₀, O₃, SO₂, and NO₂ in a heavily polluted megacity of Iran," *Environmental Science and Pollution Research*, vol. 25, no. 18, pp. 17726–17734, 2018.
- [4] World Health Organization, Geneva, Switzerland, http://www.who.int/cancer/prevention/en/.
- [5] P. S. Kanaroglou, M. D. Adams, P. F. De Luca, D. Corr, and N. Sohel, "Estimation of sulfur dioxide air pollution concentrations with a spatial autoregressive model," *Atmospheric Environment*, vol. 79, pp. 421–427, 2013.
- [6] M. B. Hadley, R. Vedanthan, and V. Fuster, "Air pollution and cardiovascular disease: a window of opportunity," *Nature Reviews Cardiology*, vol. 15, no. 4, pp. 193-194, 2018.
- [7] J. L. Reyna, M. V. Chester, S. Ahn, and A. M. Fraser, "Improving the accuracy of vehicle emissions profiles for urban transportation greenhouse gas and air pollution inventories," Environmental Science & Technology, vol. 49, no. 1, pp. 369–376, 2015.
- [8] J.-B. Sanchez, F. Berger, M. Fromm, and M.-H. Nadal, "A selective gas detection micro-device for monitoring the volatile organic compounds pollution," *Sensors and Actuators B: Chemical*, vol. 119, no. 1, pp. 227–233, 2006.
- [9] J. de Gouw and C. Warneke, "Measurements of volatile organic compounds in the earth's atmosphere using protontransfer-reaction mass spectrometry," *Mass Spectrometry Reviews*, vol. 26, no. 2, pp. 223–257, 2007.
- [10] R. S. Blake, K. P. Wyche, A. M. Ellis, and P. S. Monks, "Chemical ionization reaction time-of-flight mass spectrometry: multi-reagent analysis for determination of trace gas composition," *International Journal of Mass Spectrometry*, vol. 254, no. 1-2, pp. 85–93, 2006.
- [11] A. R. Koss, C. Warneke, B. Yuan, M. M. Coggon, P. R. Veres, and J. A. de Gouw, "Evaluation of NO⁺ reagent ion chemistry for online measurements of atmospheric volatile organic compounds," *Atmospheric Measurement Techniques*, vol. 9, no. 7, pp. 2909–2925, 2016.
- [12] M. R. Belanger, J. R. Pare, R. Turpin, J. Schaefer, and C. W. Chuang, "Evaluation of microwave-assisted process technology for HAPSITE's headspace analysis of volatile organic compounds (VOCs)," *Journal of Hazardous Materials*, vol. 145, no. 1-2, pp. 336–338, 2007.

[13] J. R. Hodgkinson and P. Tatam, "Optical gas sensing: a review," Measurement Science and Technology, vol. 24, no. 1, Article ID 012004, 2013.

- [14] S. K. Akagi, I. R. Burling, A. Mendoza et al., "Field measurements of trace gases emitted by prescribed fires in southeastern US pine forests using an open-path FTIR system," *Atmospheric Chemistry and Physics*, vol. 14, no. 1, pp. 199–215, 2014.
- [15] T. K. Flesch, V. S. Baron, J. D. Wilson, D. W. T. Griffith, J. A. Basarab, and P. J. Carlson, "Agricultural gas emissions during the spring thaw: applying a new measurement technique," *Agricultural and Forest Meteorology*, vol. 221, pp. 111–121, 2016.
- [16] Y. You, R. M. Staebler, S. G. Moussa et al., "Long-path measurements of pollutants and micrometeorology over highway 401 in Toronto," *Atmospheric Chemistry and Physics*, vol. 17, no. 22, pp. 14119–14143, 2017.
- [17] A. Wiacek, L. Li, K. Tobin, and M. Mitchell, "Characterization of trace gas emissions at an intermediate port," *Atmospheric Chemistry and Physics*, vol. 18, no. 19, pp. 13787–13812, 2018.
- [18] M. F. Delaney, F. V. Warren, and J. R. Hallowell, "Quantitative evaluation of library searching performance," *Analytical Chemistry*, vol. 55, no. 12, pp. 1925–1929, 1983.
- [19] B. K. Lavine, K. Nuguru, N. Mirjankar, and J. Workman, "Pattern recognition assisted infrared library searching," *Applied Spectroscopy*, vol. 66, no. 8, pp. 917–925, 2012.
- [20] H. Yang and P. R. Griffiths, "Encoding FT-IR spectra in a Hopfield network and its application to compound identification in open-path FT-IR measurements," *Analytical Chemistry*, vol. 71, no. 16, pp. 3356–3364, 1999.
- [21] D. W. T. Griffith, "Synthetic calibration and quantitative analysis of gas-phase FT-IR spectra," *Applied Spectroscopy*, vol. 50, no. 1, pp. 59–70, 1996.
- [22] Z. Bacsik, J. Mink, and G. Keresztury, "FTIR spectroscopy of the atmosphere. I. Principles and methods," *Applied Spectroscopy Reviews*, vol. 39, no. 3, pp. 295–363, 2004.
- [23] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [24] J.-W. Cho and H.-M. Park, "Independent vector analysis followed by HMM-based feature enhancement for robust speech recognition," *Signal Processing*, vol. 120, pp. 200–208, 2016.
- [25] V. D. Calhoun, J. Y. Liu, and T. Adali, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *NeuroImage*, vol. 45, no. 1, pp. 163–172, 2009.
- [26] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002.
- [27] M. K. Bhowmik, P. Saha, A. Singha, D. Bhattacharjee, and P. Dutta, "Enhancement of robustness of face recognition system through reduced gaussianity in Log-ICA," *Expert* Systems with Applications, vol. 116, pp. 96–107, 2019.
- [28] G. Wang, Q. Ding, and Z. Hou, "Independent component analysis and its applications in signal processing for analytical chemistry," *TrAC Trends in Analytical Chemistry*, vol. 27, no. 4, pp. 368–376, 2008.
- [29] N. Pasadakis and A. A. Kardamakis, "Identifying constituents in commercial gasoline using Fourier transform-infrared spectroscopy and independent component analysis," *Analytica Chimica Acta*, vol. 578, no. 2, pp. 250–255, 2006.
- [30] A. Al-Mbaideen and M. Benaissa, "Determination of glucose concentration from NIR spectra using independent

- component regression," Chemometrics and Intelligent Laboratory Systems, vol. 105, no. 1, pp. 131–135, 2011.
- [31] X. Yu, Y. Zhang, G. Yin et al., "Discrimination of three dimensional fluorescence spectra based on wavelet analysis and independent component analysis," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 124, pp. 52–58, 2014.
- [32] Y. B. Monakhova, A. M. Tsikin, T. Kuballa, D. W. Lachenmeier, and S. P. Mushtakova, "Independent component analysis (ICA) algorithms for improved spectral deconvolution of overlapped signals in ¹H NMR analysis: application to foods and related products," *Magnetic Reso*nance in Chemistry, vol. 52, no. 5, pp. 231–240, 2014.
- [33] P. Comon, C. Jutten, and J. Herault, "Blind separation of sources, part II: problems statement," Signal Processing, vol. 24, no. 1, pp. 11–20, 1991.
- [34] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.
- [35] C. Jutten and J. Herault, "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.
- [36] M. Mecozzi, M. Pietroletti, and Y. B. Monakhova, "FTIR spectroscopy supported by statistical techniques for the structural characterization of plastic debris in the marine environment: application to monitoring studies," *Marine Pollution Bulletin*, vol. 106, no. 1-2, pp. 155–161, 2016.
- [37] A.-L. Xia, H.-L. Wu, Y. Zhang, S.-H. Zhu, Q.-J. Han, and R.-Q. Yu, "A novel efficient way to estimate the chemical rank of high-way data arrays," *Analytica Chimica Acta*, vol. 598, no. 1, pp. 1–11, 2007.
- [38] K. Wongravee, T. Parnklang, P. Pienpinijtham et al., "Chemometric analysis of spectroscopic data on shape evolution of silver nanoparticles induced by hydrogen peroxide," *Phys. Chem. Chem. Phys.*, vol. 15, no. 12, pp. 4183–4189, 2013.
- [39] Y.-J. Liu, G. Postma, H.-L. Wu et al., "Angle distribution of loading subspace (ADLS) for estimating chemical rank in multivariate analysis: applications in spectroscopy and chromatography," *Talanta*, vol. 194, pp. 90–97, 2019.

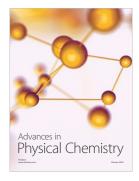
















Submit your manuscripts at www.hindawi.com





