

Characteristic wavelength selection of volatile organic compounds infrared spectra based on improved interval partial least squares

Wei Ju^{*,‡}, Changhua Lu^{*,†}, Yujun Zhang[†], Weiwei Jiang^{*},
Jizhou Wang^{*}, Yi Bing Lu[†] and Feng Hong^{*}

**School of Computer and Information*

Hefei University of Technology, Hefei, Anhui 230009, P. R. China

†Anhui Institute of Optics Fine Mechanics

Chinese Academy of Sciences, Hefei, Anhui 230031, P. R. China

‡juwei@mail.hfut.edu.cn

Received 24 October 2018

Accepted 29 December 2018

Published 31 January 2019

As important components of air pollutant, volatile organic compounds (VOCs) can cause great harm to environment and human body. The concentration change of VOCs should be focused on in real-time environment monitoring system. In order to solve the problem of wavelength redundancy in full spectrum partial least squares (PLS) modeling for VOCs concentration analysis, a new method based on improved interval PLS (iPLS) integrated with Monte-Carlo sampling, called iPLS-MC method, was proposed to select optimal characteristic wavelengths of VOCs spectra. This method uses iPLS modeling to preselect the characteristic wavebands of the spectra and generates random wavelength combinations from the selected wavebands by Monte-Carlo sampling. The wavelength combination with the best prediction result in regression model is selected as the characteristic wavelengths of the spectrum. Different wavelength selection methods were built, respectively, on Fourier transform infrared (FTIR) spectra of ethylene and ethanol gas at different concentrations obtained in the laboratory. When the interval number of iPLS model is set to 30 and the Monte-Carlo sampling runs 1000 times, the characteristic wavelengths selected by iPLS-MC method can reduce from 8916 to 10, which occupies only 0.22% of the full spectrum wavelengths. While the RMSECV and correlation coefficient (Rc) for ethylene are 0.2977 and 0.9999 ppm, and those for ethanol gas are 0.2977 ppm and 0.9999. The experimental results show that the iPLS-MC method can select the optimal characteristic wavelengths of VOCs FTIR spectra stably and effectively, and the prediction performance of the regression model can be significantly improved and simplified by using characteristic wavelengths.

Keywords: Ambient air monitoring; Fourier transform infrared spectra analysis; variable selection; interval partial least square; Monte-Carlo sampling.

[‡]Corresponding author.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

1. Introduction

Volatile organic compounds (VOCs) are defined as any organic compounds that can produce oxidants by reactions with nitrogen oxides in sunlight which would accelerate the photochemical reactions of the atmosphere.¹ The members of VOCs are more than 300 types which all have the common features of high vapor pressure, low boiling point and strong reactivity.² The major species of VOCs along with their sources and influences are listed in Table 1. We can see from the table that most VOCs can cause serious damage to human health and the environment. The main impacts on human bodies are inhalation of or exposure to the VOCs would induce

various acute and chronic health effects such as central nervous system impairment, cancer, skin and sensory irritation.³ VOCs also make detrimental effects to the ambient air because they are the crucial precursors for photochemical ozone that is the main inducement of global greenhouse effect,⁴ moreover, VOCs play a key role in the formation of secondary organic aerosols which is the principal component of PM_{2.5}.⁵ The sources of VOCs vary from countries to areas, and they are emitted from both anthropogenic and natural sources. In the developed areas, the major sources of VOCs are automobile exhausts which contribute 25% of the total VOCs concentration, other sources include

Table 1. Sources and influence of typical VOCs.⁸

Categories	Representatives	Sources	Health impacts	Environment influence
Alkane	Methane ⁹	Part-combusted gas	Breathing irritation	Greenhouse gas
	Butane ¹⁰	Fuel additive,	Memory loss	Marine ecological damage
	Hexane	Refrigerant solvent Gasoline consumption	Anesthetic	Haze
Alcohol	Methanol ¹¹	Antiseptics petrochemical derivative	Skin and eyes illness	Large amounts of waste water and gas during production
	Ethanol Isopropyl alcohol	Cosmetics Pharmaceuticals,	Breathing irritation Central nervous system impairment	Photochemical smog
Aldehyde Esters	Formaldehyde	Building and decorative materials	Throat, eyes and skin illness	Predecessor of ozone
	Acetaldehyde	Cosmetics decomposition Tobacco smoking Chemical production	Carcinogen Central nervous system damage Dizziness	Detrimental to vegetation
Olefins	Ethylene ¹²	Petrochemical derivative	Carcinogen	Photochemical ozone
	Isoprene	Product of perfumes and pharmaceutical	Anesthetic	Water and soil pollution
BTEX ¹²	Propylene	Adhesives	Vertigo	
	Benzene	Petroleum products	Carcinogen mutagenic hazards	Photochemical smog Damage the ozone layer
	Toluene Ethylbenzene xylenes	Part-combusted liquid fuels Adhesives Production of paints	Tiredness Sleepiness	Marine ecological damage
Cl-VOCs (chlorinated VOCs)	Carbon tetrachloride	Solvent Insecticide	Acute toxicity	Damage of the ozone layer
	Chlorobenzene	Dye industry	Dizziness	Cause greenhouse gas effects
	Tetrachloroethane	Pharmaceutical	Nervous and kidney damage	
	Trichloroethylene	Adhesives Sewage disposal		
Ketones	Acetone ¹³	Paint	Skin inflammation	Oxygen depletion in aquatic system
	Cyclohexanone	Cleaning agent	Corneal damage	
		Diluent Adhesives	Narcosis Nausea	

industrial emissions, petroleum refining and storage, solvents evaporation, fossil fuels combustion and biogenic emissions. Natural sources of VOCs are mainly originated from vegetation and ocean emissions,⁶ which occupy a small proportion of the TVOC (total VOCs) emissions in the city areas. With the accelerated development of industry and the rapid expansion of transportation networks, VOCs have become significant atmospheric pollutants in urban areas of China.⁷ As the atmospheric pollution incidents caused by VOCs become increasingly serious and prominent, it's necessary to establish a real-time continuous VOCs monitoring system to confirm the information of specific VOCs and control the emission of them.

Classical VOC analytical methods include chemical methods, liquid chromatography (LC),¹⁴ gas chromatography-mass spectrometry (GC-MS),¹⁵ gas chromatography-flame ionization detector (GC-FID),¹⁶ photo-ionization detection (PID)¹⁷ and so on. With the continuous progress of sensors and computer technology, various online real-time VOCs detection technologies and corresponding portable instruments have been developed. Commercial VOC online monitoring instruments are mostly based on GS-MS technology, for example, the Hazardous Air Pollutants on Site (HAPSITE)¹⁸ designed by Inficon company is a portable GC-MS unit which can identify and quantify VOCs from environmental sampling. Proton transfer reaction-mass spectrometer (PTR-MS) is also a well-established approach for monitoring VOCs in recent years with its feature of high sensitivity and fast time response. Joost *et al.*¹⁹ used PTR-MS to detect the air composition in aged forest-fire and urban plumes. Cui *et al.*²⁰ measured continuous VOCs information by PTR-MS instrument in urban roadside of Hong Kong. Chemical ionization reaction mass spectrometry (CIRMS)²¹ as an extension of PTR-MS has been applied in real-time atmospheric research, it uses various chemical ionization reagents in the ionization process which makes it a more versatile technique than PTR-MS. Koss *et al.*²² used NO^+ as CIRMS reagent ion to achieve fast online measurement of trace atmospheric VOCs.

These online real-time detecting methods are based on chemical or physical reaction between gases and reagents. The detection results are accurate, but they have the disadvantages of being time-consuming, complicated and some require toxic

reagents for preprocessing. Reagent-free and rapid continuous measurement will be the new trend of VOC detection.

Due to the progress of modern physics, especially the development of surface physics, optics and electronics, infrared spectroscopy has made considerable development during the past decades. As a nondestructive analysis tool, spectroscopy has been increasingly applied in fields like environment, food, pharmaceutical, agriculture and so on with its advantages of wide range, multi-component analysis and continuous real-time monitoring. The spectroscopic methods include nondispersion infrared (NDIR), differential absorption lidar (DIAL), differential optical absorption spectroscopy (DOAS), tunable diode laser absorption spectroscopy (TDLAS) and Fourier transform infrared (FTIR) spectroscopy.²³ Among these methods, FTIR is recommended by the US Environment Protect Agency as the specified VOC online detection method because of its rapid, highly sensitive and simultaneous detection of multiple components.⁵ FTIR spectrum is generated from infrared radiation absorption during the vibration transition by polyatomic molecules which have asymmetric dipole moment, and the spectrum consists of absorption peaks associated with functional groups (C-H, O-H, N-H, etc.) that can reflect the composition and concentration information of substances. Since the main components in the air, nitrogen (78% of dry air), oxygen (21%) and argon (1%) are transparent to infrared radiation due to their symmetrical form of molecules, the FTIR could detect trace VOCs without the interference of these components.

Spectroscopic method can identify the components and their amounts in unknown substances by building qualitative and quantitative calibration models of the acquired spectra. In conventional spectra calibration models, full spectrum modeling is often used in order to not lose any information from the spectrum. Full spectrum usually consists of thousands of variables, which not only contain the information of target components but also many other redundant information, such as noise disturbance and interference components. While modeling with redundant variables would affect the sensitivity and predict the accuracy of calibration model. Meanwhile, the number of variables is usually much larger than the number of available sample spectra, which make the analysis of full spectrum extremely difficult with the common modeling

method. Therefore, variable selection method which selects the most informative variables instead of using full spectrum is crucial in the spectroscopy analysis.²⁴ The researches about variable selection, which is also called characteristic wavelength selection on infrared spectrum, have achieved good performance in many fields. Miaw *et al.*²⁵ used iPLS method to quantify the FTIR spectra of adulterated fruits syrup. Durand *et al.*²⁶ adopted genetic algorithm (GA) integrated with PLS modeling for near infrared (NIR) spectrum quantitative prediction of cotton content in cotton-viscose textile samples. Li *et al.*²⁷ investigated the Monte-Carlo uninformative variable elimination (MC-UVE) combined with successive projections algorithm (SPA) to select the most effective variables from NIR spectrum of pears to determine the soluble solid content and firmness of the pears. Han *et al.*²⁸ presented an ensemble of Monte-Carlo uninformative variable elimination (EMCUVE) in multivariate calibration of spectra data. Fan *et al.*²⁹ used improved Monte-Carlo sampling, named competitive adaptive reweighted sampling (CARS), to select the characteristic wavelengths of vinegar spectrum in order to determine the total acid of the vinegar. The results of these researches show that variable selection can enhance the prediction ability efficiently and simplify the identify model by using selected variables. However, there were few researches on characteristic wavelengths selection of VOCs gas spectrum.

The objective of this paper was to find the optimal characteristic wavelengths of VOCs infrared spectrum by the variable selection method, as the selected characteristic wavelengths can make the calibration model of the spectrum more effective and simpler. In addition, this study aimed at selecting as few wavelengths as possible to identify the concentration information of gas spectrum. To achieve these goals, the specific works are as follows: (1) presented the acquisition and the pretreatment method of experimental materials, described the feature of general VOCs spectrum; (2) introduced the principle of the proposed wavelengths selection method and its feasibility in selecting characteristic wavelengths of VOCs spectrum; (3) compared with the regression results from full-spectrum modeling and improved iPLS modeling, the advantages of proposed iPLS-MC modeling were analyzed from the perspective of wavelength numbers, Root Mean Square Error of Cross-Validation (RMSECV), Root Mean Square Error of Prediction (RMSEP),

Related coefficient of calibration set (R_c) and Related coefficient of prediction set (R_p).

2. Materials and Methods

2.1. Sample preparation

Ethylene (molecular formula C_2H_4), as a representative gas of VOCs, is one of the world's most productive petrochemical derivative, which is mainly emitted from the engine exhaustion, thermal power plant and food industry. Excessive inhalation of ethylene can cause anesthetic disease, ethylene also causes photochemical smog and increases the ground-level ozone.³⁰ Ethanol (molecular formula C_2H_5OH) gas is one of the typical VOCs as well, it is the most common monohydric alcohol, and has been widely used in medical and health service, chemical industry, food industry and agriculture.³¹ According to the carcinogens list published by the World Health Organization, ethanol is a risk factor for many cancer types including cancer of the pharynx, liver and breast. As a flammable gas, ethanol tends to evaporate into vapor which can form an explosive mixture with air. Because of the advantages of easy preparation, low price and low concentration of the gas doing little harm to human body, ethylene and ethanol gas are suitable for laboratory research. Therefore, we choose ethylene and ethanol gas as the experimental materials for this study. The gases for experiments are produced by Hefei Ningte Gas Company and stored in 4 L sealed air cylinders, respectively. The original concentration of ethylene is 2005 ppm (parts per million, volume concentration) and the original concentration of ethanol is 2007 ppm. The experimental gases were mixed with the auxiliary gas nitrogen to get different concentrations. Gas distribution platform adopts the 4-channel high precision gas distribution system independently developed by Hefei Institutes of Physical Science, Chinese Academy of Sciences. The precision of the distribution system is 0.1% of the original gas concentration, with error range of 0.5–1‰. In total, we got 60 groups of ethylene gas with concentrations ranging from 60.15 to 178.445 ppm, with the interval between each group to be 2.005 ppm, and 60 groups of ethanol gas with concentrations ranging from 60.21 to 178.623 ppm, with the interval between each group to be 2.007 ppm. Because of the existence of distribution system error, the accuracy of the

concentration is within a range of $\pm 5\text{--}10\%$ of calculated concentration.

2.2. Spectra acquisition

The infrared spectra were collected by FTIR multicomponent gas analyzer system independently developed by Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences. The analyzer adopts FTIR spectroscopy technology and gets a multiple reflection gas cell with the optical path of 10 m inside. The analyzer's measuring bandwidth covers from 700 to 5000 cm^{-1} with a resolution of 1 cm^{-1} and it scans 16 times to obtain the average value as the output spectrum. The experiment system was completely airtight and all the pipes went through vacuum test before starting the experiment. At the beginning of the experiment, the gas cylinder was put in the ventilation room, and the air flow was kept free. The flow rate of gas in each cylinder was controlled by software parameters of gas distribution platform. The experimental gas and nitrogen were proportionally configured to the required concentration. When the gas flow was kept steady and the barometric pressure in the gas cell was close to one bar pressure, we checked the viewing screen of the analyzer to see whether the spectrum was stable, then recorded the spectrum data and its corresponding concentration when it was stable.

The raw spectra of all 60 groups of C_2H_4 were shown in Fig. 1. According to the HITRAN (high-resolution transmission molecular absorption database), the infrared spectrum of C_2H_4 has three strong absorption peaks, located in $2848\text{--}2972\text{ cm}^{-1}$, $1463\text{--}1473\text{ cm}^{-1}$ and $920\text{--}980\text{ cm}^{-1}$, respectively. The three absorption peaks correspond to C–H stretching vibration band, $-\text{CH}_2-$ scissor vibration band and trans olefin vibration band.

The raw spectra of all 60 groups of $\text{C}_2\text{H}_5\text{OH}$ are shown in Fig. 2. According to the database from PNNL (Pacific Northwest National Laboratory), the infrared spectrum of $\text{C}_2\text{H}_5\text{OH}$ has three strong absorption peaks located in $3280\text{--}3425\text{ cm}^{-1}$, $2848\text{--}2972\text{ cm}^{-1}$, $1050\text{--}1090\text{ cm}^{-1}$, respectively. The three absorption peaks correspond to $-\text{OH}$ stretching vibration band, C–H stretching vibration band and saturated alcohol's C–O stretching vibration band.

As we can see from Figs. 1 and 2, in addition to the absorption peaks of target components, there are many other interference absorption peaks which are primarily generated by $\text{H}_2\text{O}(\text{g})$ and CO_2 , as well as the spectral response of the device. The bending vibration peak of $-\text{OH}-$ in $\text{H}_2\text{O}(\text{g})$ appears in the range of $1500\text{--}1800\text{ cm}^{-1}$, the stretching vibration peak appears in the range of $3500\text{--}3950\text{ cm}^{-1}$, and the antisymmetric stretching vibration peak of CO_2 appears near 2359 cm^{-1} . Therefore, it's vital to select the characteristic wavelengths of target

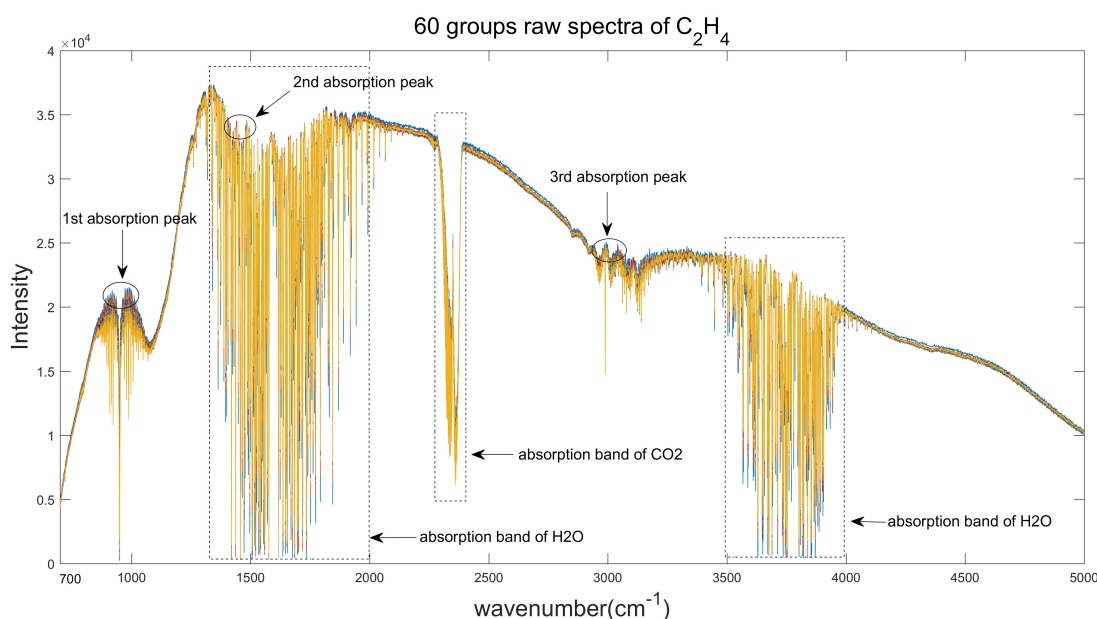
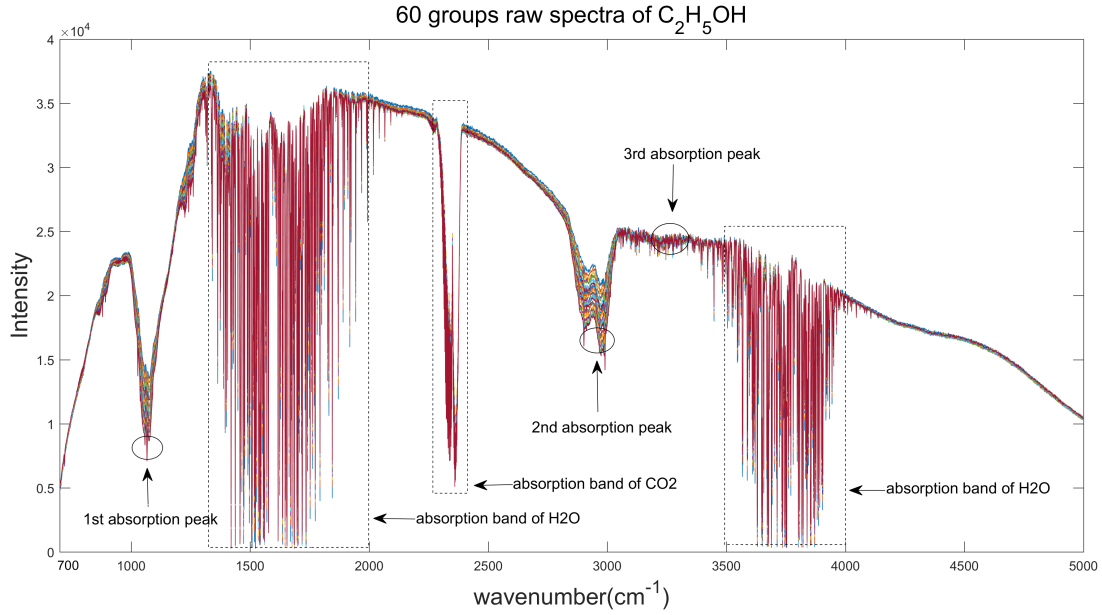


Fig. 1. Raw spectra of C_2H_4 .

Fig. 2. Raw spectra of C_2H_5OH .

components before establish the quantitative and qualitative calibration models.

2.3. Preparation of spectrum

2.3.1. Lifting wavelet spectrum denoising

In the process of spectrum acquisition, the spectrum is usually affected by the random noises from acquisition equipment or transmission path. The pretreatment of original spectrum is an essential procedure before spectrum analysis. In this paper, lifting wavelet transform (LWT) denoising method was introduced for pretreatment of the original spectrum.³² LWT is an improvement of wavelet algorithm, and it uses multiplication operation instead of convolution operation in traditional wavelet transform. The lifting scheme is divided into three stages: split, predict and update to realize its spatial transformation.

(1) **Split:** The input signal S_i is divided into two mutually disjoint subsets $even_{i-1}$ and odd_{i-1} :

$$even_{i-1} = S[2n], odd_{i-1} = S[2n - 1]. \quad (1)$$

(2) **Predict:** On the basis of maintaining the correlation of the original data, the predictor P was used to treat $even_{i-1}$ and the predicted value $P(even_{i-1})$ of odd_{i-1} was obtained. The actual value of odd_{i-1} was subtracted from its

predicted value to get the wavelet coefficient d_{i-1} :

$$d_{i-1} = odd_{i-1} - P(even_{i-1}). \quad (2)$$

(3) **Update:** The purpose of update is to preserve some global characteristics of the original signal set in subset S_{i-1} . Construct the update operator U , update $even_{i-1}$ with wavelet coefficient d_{i-1} , and obtain the low-frequency coefficient S_{i-1} :

$$S_{i-1} = even_{i-1} + U(d_{i-1}). \quad (3)$$

After n times of decomposition of the above three steps, the original signal expression is $\{S_{i-n}, d_{i-n}, d_{i-n+1}, \dots, d_{i-1}\}$, where S_{i-n} represents the low-frequency part of the signal, $\{d_{i-n}, d_{i-n+1}, \dots, d_{i-1}\}$ represents the high-frequency part of the signal. Lifting wavelet simplifies the filtering process into three basic steps, and each step of decomposition is reversible. The reconstruction process of the lifting method is the inverse process of decomposition process. The decomposition and reconstruction structure of LWT is shown in Fig. 3.

It can be seen from above description that the lifting scheme does not need other data except the output of the previous lifting steps, so that the new data flow can replace the old data flow at each point, and the lifting wavelet coefficient can be obtained by repeating the lifting filter banks. The

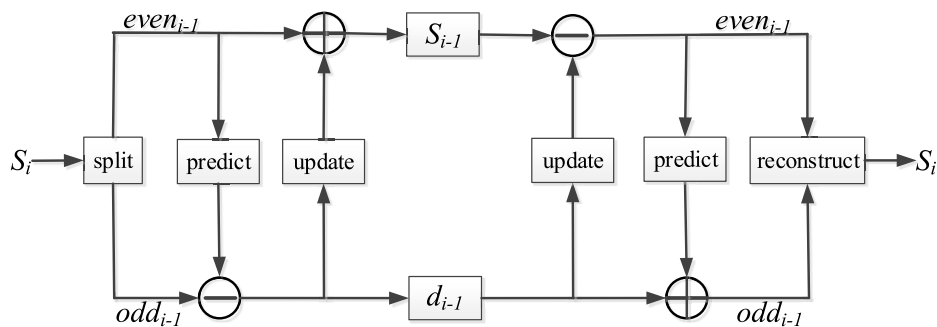


Fig. 3. Decomposition and reconstruction of the lifting scheme.

lifting wavelet not only inherits the multi-resolution property of classical wavelet but also has simple structure, low computational complexity and easy hardware implementation. For online real-time measurement system, LWT denoising has the advantage of high running speed, so it is suitable for VOC online monitoring.

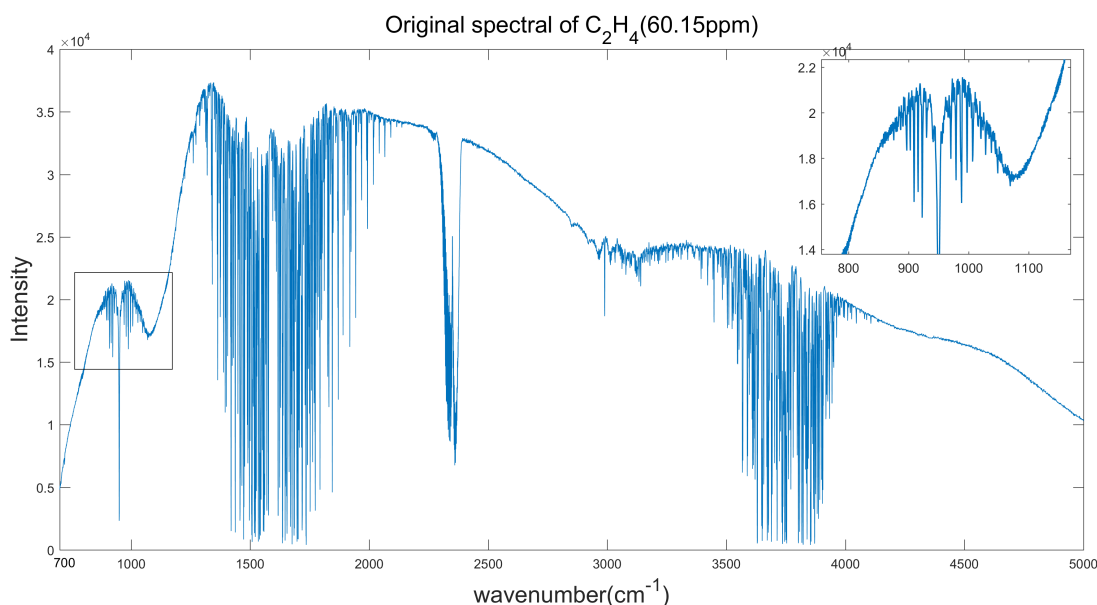
We take the preprocessing of C_2H_4 spectrum as an example. Figure 4 is the original spectrum of C_2H_4 with the concentration of 60.15 ppm, and Fig. 5 is the denoised spectrum by 6-layer “db4” wavelet function lifting scheme denoising.

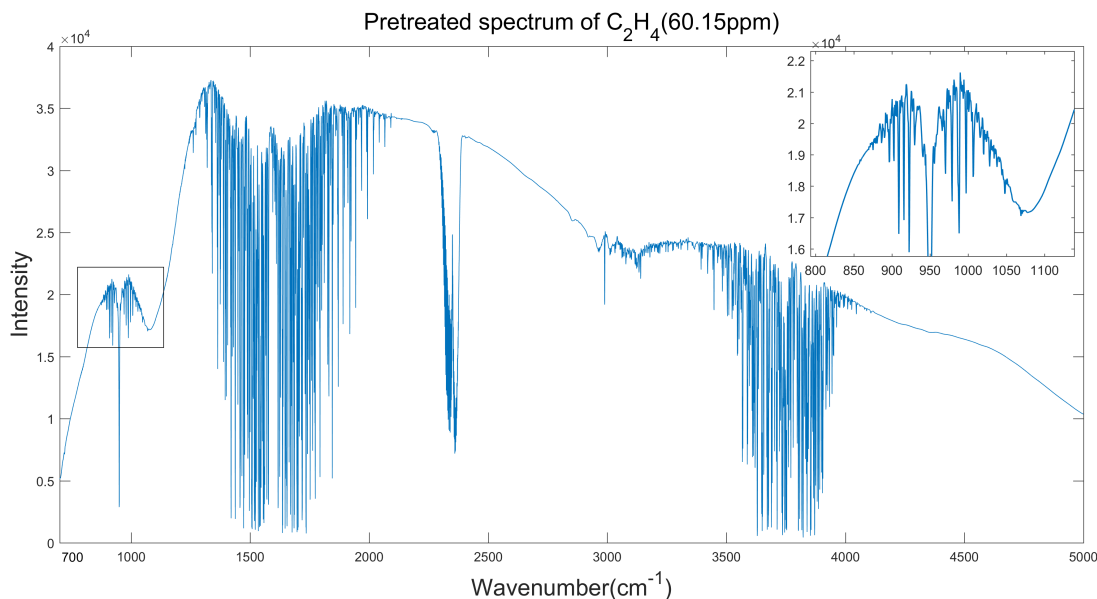
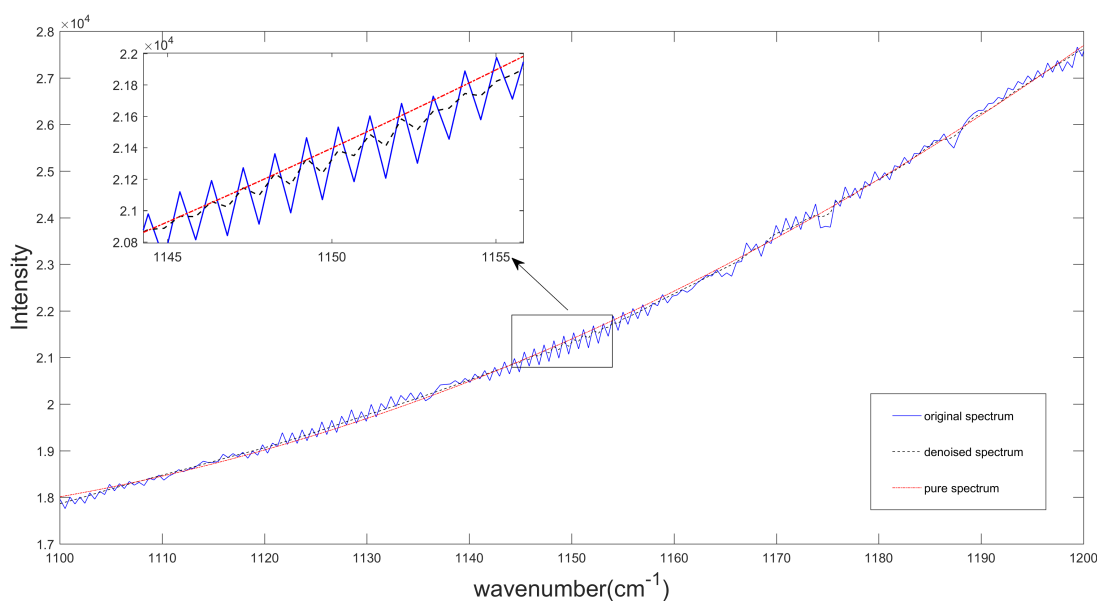
2.3.2. Comparison of denoising effect

When the resolution of the spectrometer is high enough, the infrared spectrum in an ideal state has a narrow enough absorption peak in the infrared

absorption waveband, but showing a smooth curve in the waveband without infrared absorption. In the measured spectrum, due to the hardware limitation of spectrometer and the influence of noise, the spectrum in the waveband without infrared absorption presents a sawtooth shape. According to this property, the noninfrared absorption waveband was selected to fit the smooth curve. This smooth curve was taken as the pure infrared spectrum in the noise-free environment, and the signal-to-noise ratios (SNRs) of the original spectrum and the spectrum after lifting wavelet denoising were calculated.

Figure 6 takes the waveband of $1100\text{--}1200\text{ cm}^{-1}$ as the analysis object, as this band has no infrared material absorption in this experiment. In the figure, the solid line is the original spectrum, the dotted line is the spectrum after lifting wavelet

Fig. 4. Original spectrum of C_2H_4 (60.15 ppm).

Fig. 5. Pretreated spectrum of C₂H₄ (60.15 ppm).Fig. 6. Spectrum of C₂H₄ (60.15 ppm) in the waveband of 1100–1200 cm⁻¹.

denoising, and the point solid line is the pure spectrum fitted. The calculation formula of SNR is shown in Formula (4), and the calculation results are listed in Table 2.

$$\text{SNR}(\text{dB}) = 10 \log(P_{\text{signal}}/P_{\text{noise}}). \quad (4)$$

The original spectra of 60 groups with different concentrations ranging from 60.15 to 178.445 ppm were divided into calibration set and validation set

by the Kennard-Stone method with a ratio of 2:1, the sizes of each set are 40 and 20. Perform the same divided process to the denoised spectra. The PLS models were built based on original spectra and pretreated spectra, respectively, the results are summarized in Table 2.

From Figs. 4 and 5, we can see clearly that the pretreatment of spectrum has not only reduced the noise interference but also made the absorption

Table 2. Comparison of original and denoised spectra.

	SNR (dB)	Number of LVs	Calibration sets		Prediction sets	
			Rc	RMSECV (ppm)	Rp	RMSEP (ppm)
Original spectra	25.4690	3	0.9568	2.7685	0.9454	2.9731
Pretreated spectra	35.1569	3	0.9737	2.5827	0.9629	2.7349

peaks more obvious. It can be seen from Table 2 that the spectral SNR increases from 25.4690 to 35.1569 db after lifting wavelet denoising, that is, increased by 38.03%. For the two spectra calibration models, both RMSECV and RMSEP are less than 3 ppm, and both Rc and Rp are over 0.94, which indicates that the PLS model has a good ability to identify the concentration of VOCs. The calibration result shows that the PLS modeling of pretreated spectra has a better accuracy and predictive ability than the modeling of original spectra.

2.4. Primary methods

2.4.1. PLS

PLS is an innovative multivariate statistical analysis method proposed by Svante Wold in 1983. On the basis of traditional regression model which directly uses predictors (independent variables) and responses (dependent variables), PLS also investigates the multiple correlations among variables by transferring predictors and responses into a set of independent factors named latent variables (LVs). LVs can describe the maximum covariance between predictors and responses.^{33,34} In the quantitative identification model of VOC spectra, the predictor is the spectral absorption matrix (X), and the response is the gas concentration matrix (Y). At the beginning of modeling, the spectra set is divided into calibration set and prediction set with a certain ratio. The calibration set is used to establish the calibration model, and the validation parameters of calibration set are RMSECV and Rc. RMSECV is used to calculate the errors on test split using a cross validation scheme. It can measure the goodness of fit between known data and the calibration model. The number of LVs is determined by the minimum value of RMSECV. The prediction set is used to verify the predictive ability of the calibration model, and the validation parameters are RMSEP and Rp.

2.4.2. Internal partial least squares

Interval partial least squares (iPLS) is an improved development of PLS. It was proposed by Nørgaard in 2000 and has become one of the most commonly used chemometrics method for selecting characteristic variables in recent years.³⁵ The principle of this method is to divide the whole spectrum equidistant into several subintervals, then establish the independent PLS model in each subinterval with different numbers of LVs. The optimal number of LVs in each subinterval is also determined by the minimum RMSECV from cross-validation. By comparing the model of each subintervals and the global model of full-spectrum, the model with the minimum RMSECV value is selected as the optimal subinterval. IPLS has advantages of removing the redundant information and simplifying calculation, but also has the disadvantage of losing useful characteristic information because of the single interval modeling.

2.4.3. Monte-Carlo sampling

Monte-Carlo sampling is also called statistical simulation method, it is a numerical method guided by the theory of probability statistics. The feature of Monte-Carlo sampling is that the more the experiments are, the more accurate the results will be. With the development of modern computer technology-Monte Carlo sampling is simple and fast in practical application.³⁶ In the characteristic wavelength selection of spectrum, Monte-Carlo sampling uses the different combinations of random wavelengths to carry out multiple linear regress (MLR) modeling, and calculates its regression coefficient of each combination. The final results of characteristic wavelength selection are determined by the regression coefficient. Although the calculation is simple and fast, Monte-Carlo sampling has the disadvantage that if there are too many variables to be measured, the calculation amount will be very large.

2.4.4. Proposed method

The features of VOCs infrared spectrum include the following: (1) the spectrum usually contains several absorption peaks, and the location and intensity of absorption peaks are related to the type and quantity of gas molecules which contain different functional groups; (2) infrared radiation absorption changes rapidly with wavelengths, and the spectrum has maximal values at some certain wavelengths.³⁷ According to these features, we propose a new characteristic wavelength selecting method for VOC gas infrared spectrum called iPLS-MC algorithm. This method uses the iPLS method to select subintervals with RMSECV value less than that of full-spectrum PLS as the characteristic wavebands, then combines these wavebands into a new spectrum combination. Monte-Carlo sampling is used to select random wavelengths from the new spectrum combination, then carry out PLS modeling to calculate the prediction error of different random wavelength combinations. Monte-Carlo operations are repeated many times to ensure the stability and reliability of the results. Finally, the wavelength combination with the minimum prediction error is selected as the characteristic wavelength of the spectrum. This method makes use of the advantages of both iPLS and Monte-Carlo sampling while avoiding their disadvantages. iPLS can select several characteristic spectrum bands with optimal predictive

ability to avoid the possibility of losing useful information during single modeling. Monte-Carlo sampling is applied to the new spectrum combination that has already been screened to avoid the disadvantage of massive computation, so that Monte-Carlo sampling can accurately select the optimal characteristic wavelengths of the VOCs infrared spectrum.

3. Experiment Results and Discussion

3.1. Model of iPLS

For the iPLS model, the number of intervals has significant impact on the performance of the model. If the interval number is too small, it may degenerate into full spectrum PLS model, while if the interval number is too big, the amount of computation will be increased. In this study, the full spectra are divided equally into 20 and 30 intervals, respectively. Figures 7 and 8 show the iPLS models of two gas spectra with the intervals of 20, and the results of two models are listed in Tables 3 and 4. In each figure, the gray bars represent the RMSECV value of each subinterval, and the dotted line expresses the RMSECV value of full spectrum (global) model. The results that are listed in each table include the selected interval numbers and their corresponding wavenumber range, each subinterval's optimal LVs number, calibration set's RMSECV and R_c value, prediction set's RMSEP

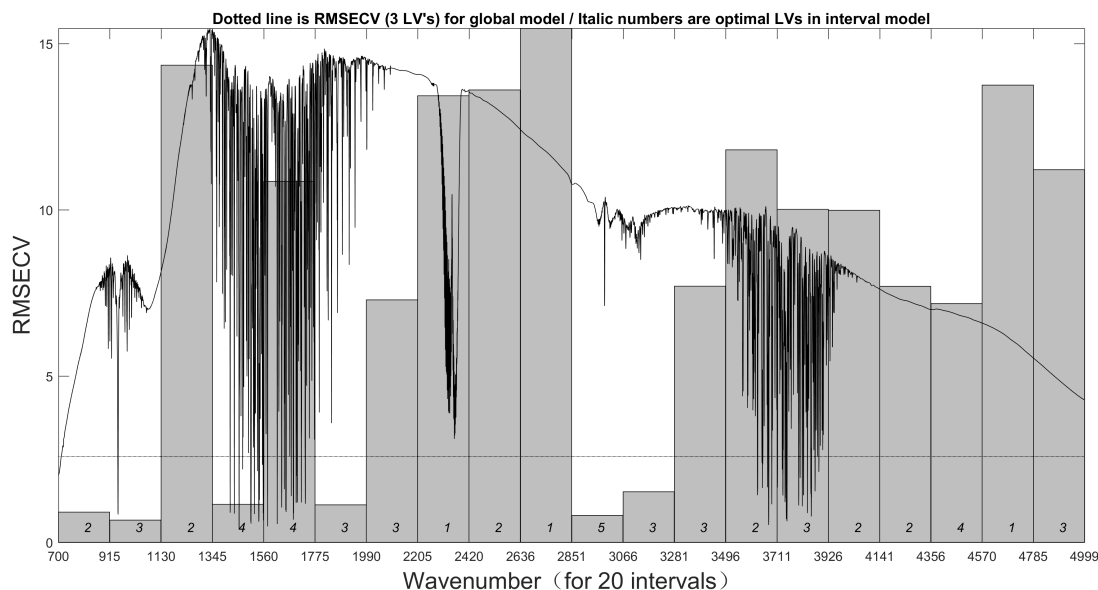


Fig. 7. 20 intervals iPLS model of C_2H_4 spectra.

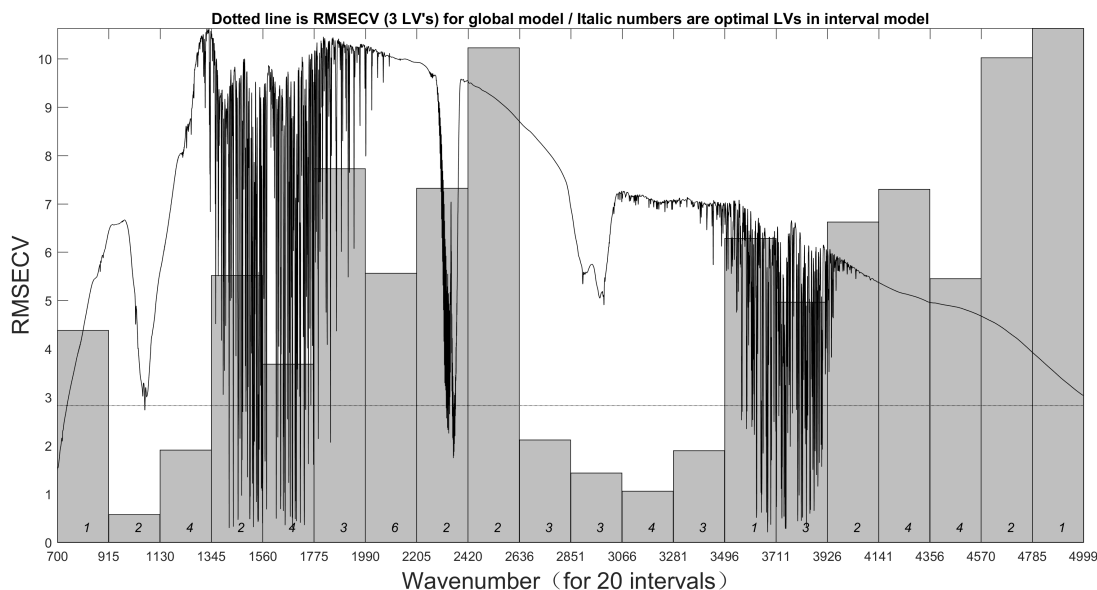

 Fig. 8. 20 intervals iPLS model of C_2H_5OH spectra.

 Table 3. Wavebands selected from 20 intervals iPLS model of C_2H_4 spectra.

Interval number	Wavenumber range (cm^{-1})	Number of LVs	Calibration set		Prediction set	
			RMSECV (ppm)	Rc	RMSEP (ppm)	Rp
1	700–915	2	0.9074	0.9977	0.9739	0.9951
2	915–1130	3	0.6662	0.9989	0.7043	0.9982
4	1345–1560	4	1.1394	0.9951	1.0484	0.9935
6	1775–1990	3	1.1251	0.9974	1.2010	0.9964
11	2851–3066	5	0.8077	0.9983	0.8205	0.9981
12	3066–3281	3	1.5176	0.9958	1.5176	0.9940
Full spectra	700–4999	3	2.5827	0.9737	2.7349	0.9629

and R_p values. Under the same experimental conditions, the number of iPLS intervals was set to 30, and the experimental results of iPLS for the two gas spectra are listed in Tables 5 and 6.

The two figures above can provide an overall view of the relevant information in different spectral subintervals. It helps us to focus on informative wavebands and remove other interference regions.

 Table 4. Wavebands selected from 20 intervals iPLS model of C_2H_5OH spectra.

Interval number	Wavenumber range (cm^{-1})	Number of LVs	Calibration set		Prediction set	
			RMSECV (ppm)	Rc	RMSEP (ppm)	Rp
2	915–1130	2	0.6726	0.9987	0.7043	0.9981
3	1130–1345	4	1.9062	0.9928	2.2734	0.9859
10	2636–2851	3	2.1153	0.9814	2.5671	0.9817
11	2851–3066	3	1.4317	0.9985	1.8502	0.9902
12	3066–3281	4	1.0554	0.9990	1.2167	0.9953
13	3281–3496	3	1.8934	0.9976	1.9631	0.9925
Full spectra	700–4999	3	3.0184	0.9625	3.2361	0.9594

Table 5. Wavebands selected from 30 intervals iPLS model of C₂H₄ spectra.

Interval number	Wavenumber range (cm ⁻¹)	Number of LVs	Calibration set		Prediction set	
			RMSECV (ppm)	Rc	RMSEP (ppm)	Rp
2	843–987	3	0.6294	0.9990	0.6458	0.9989
3	987–1131	2	0.8144	0.9983	0.8216	0.9983
5	1275–1418	10	1.2473	0.9951	1.2487	0.9960
6	1418–1562	4	1.5858	0.9913	2.1040	0.9892
9	1848–1992	4	1.1492	0.9969	1.2135	0.9964
16	2851–2994	2	0.9305	0.9974	1.0895	0.9969
17	2994–3138	2	2.4827	0.9821	2.5700	0.9804
18	3138–3281	5	2.4636	0.9836	2.5808	0.9813
Full spectra	700–4999	3	2.5827	0.9737	2.7349	0.9629

Table 6. Wavebands selected from 30 intervals iPLS model of C₂H₅OH spectra.

Interval number	Wavenumber range (cm ⁻¹)	Number of LVs	Calibration set		Prediction set	
			RMSECV (ppm)	Rc	RMSEP (ppm)	Rp
2	843–987	3	2.0877	0.9895	2.3583	0.9816
3	987–1131	3	0.5219	0.9996	0.7126	0.9983
15	2708–2851	3	2.0447	0.9901	2.6912	0.9878
16	2851–2994	3	1.1211	0.9974	1.0895	0.9969
17	2994–3138	4	1.0516	0.9978	1.5065	0.9904
18	3138–3281	3	1.9206	0.9927	2.4185	0.9813
19	3281–3424	4	1.7698	0.9921	1.9219	0.9878
Full spectra	700–4999	3	3.0184	0.9625	3.2361	0.9594

The predictive ability of each interval varies greatly, thereby we can easily compare the prediction performance of subinterval models and global model from the figures.

For the common iPLS model, the subinterval which has the smallest RMSECV value would be selected as the characteristic waveband. The VOC FTIR spectra usually have several absorption bands, so the operation of iPLS model may lose some useful parts of the spectra. In this paper, we select the subintervals in which the RMSECV value is smaller than the global model RMSECV value as the characteristic wavebands, and the corresponding subintervals of each iPLS models are listed in each table. There are six wavebands selected from the 20 intervals iPLS model of C₂H₄ spectra, 30% of the total wavelengths. Also, eight wavebands are selected from the 30 intervals iPLS model of C₂H₄ spectra, 27% of the total wavelengths. As for C₂H₅OH spectra, there are six and seven wavebands selected from the 20 and 30 intervals iPLS model,

respectively, 30% and 23% of the total wavelengths. Compared with the information from the standard infrared spectra database, the selected subintervals located in the positions coincide with the absorption peaks in the database. The results indicate that the strategy based on iPLS model could be feasible to select the characteristic regions of the VOC infrared spectra.

3.2. Model of Monte-Carlo sampling

In Sec. 3.1, iPLS model was performed for acquiring characteristic wavebands of VOCs spectra, and the selected wavebands were prepared for the next operation of Monte-Carlo sampling. A certain number of wavelengths were selected randomly from the selected wavebands by Monte-Carlo sampling and the PLS model was established based on these selected wavelengths. We repeated the Monte-Carlo sampling and PLS modeling many times to ensure the stability and reliability of the

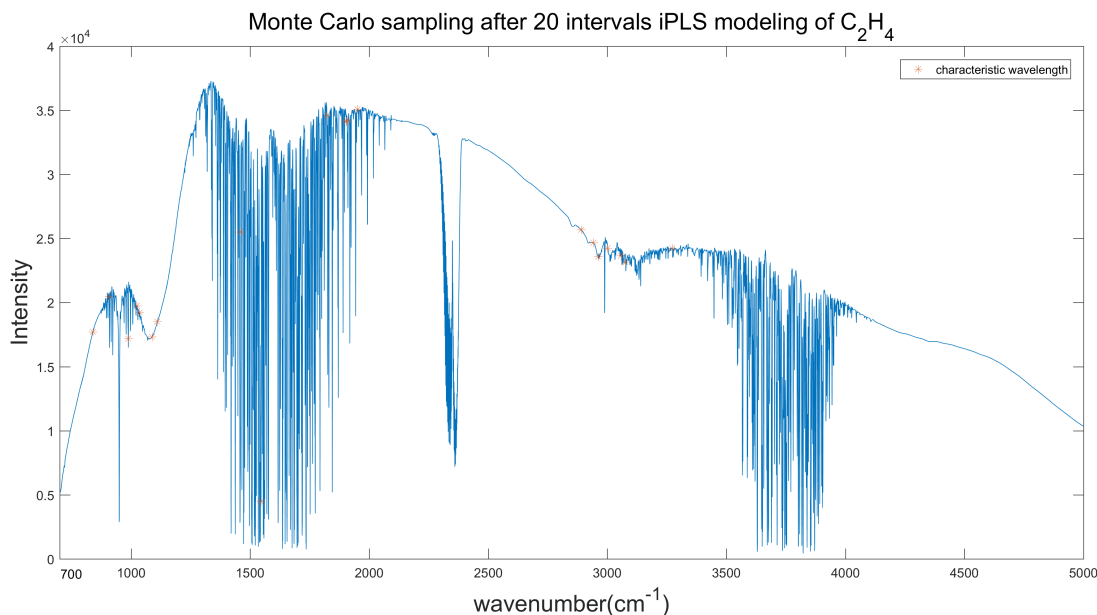


Fig. 9. Wavelengths selected by iPLS (20 intervals)-MC model of C_2H_4 spectra.

results. The prediction error was calculated separately corresponding to wavelengths set in each operation. Finally, the wavelengths set with the minimum prediction error were adopted as the optimal characteristic wavelength combinations of the spectrum.

The wavebands selected from 20 intervals iPLS model of C_2H_4 spectrum have totally 1290 wavelengths. In this study, we selected 20 wavelengths

from the preselected wavebands by Monte-Carlo sampling, the sampling process was repeated 1000 times to get the minimum RMSECV. Through several experiments, the results of RMSECV were stable at about 0.3 ppm. The results of iPLS-MC model are shown in Fig. 9, with the characteristic wavelengths denoted by red asterisk. The 20 wavelengths selected from 30 intervals iPLS model of C_2H_4 spectra are shown in Fig. 10.

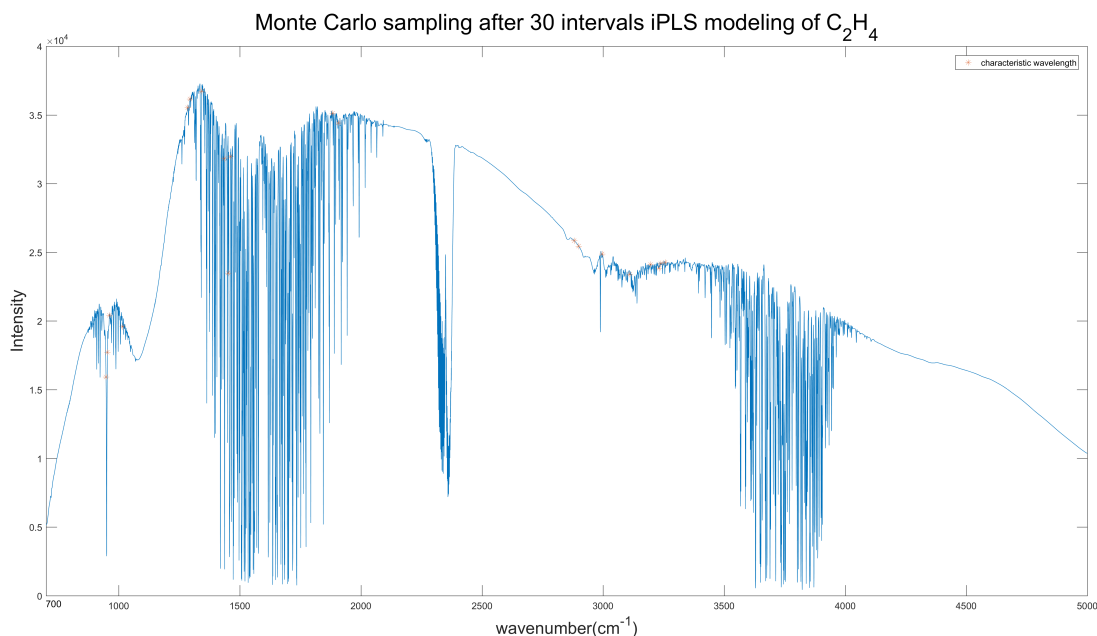


Fig. 10. Wavelengths selected by iPLS (30 intervals)-MC model of C_2H_4 spectra.

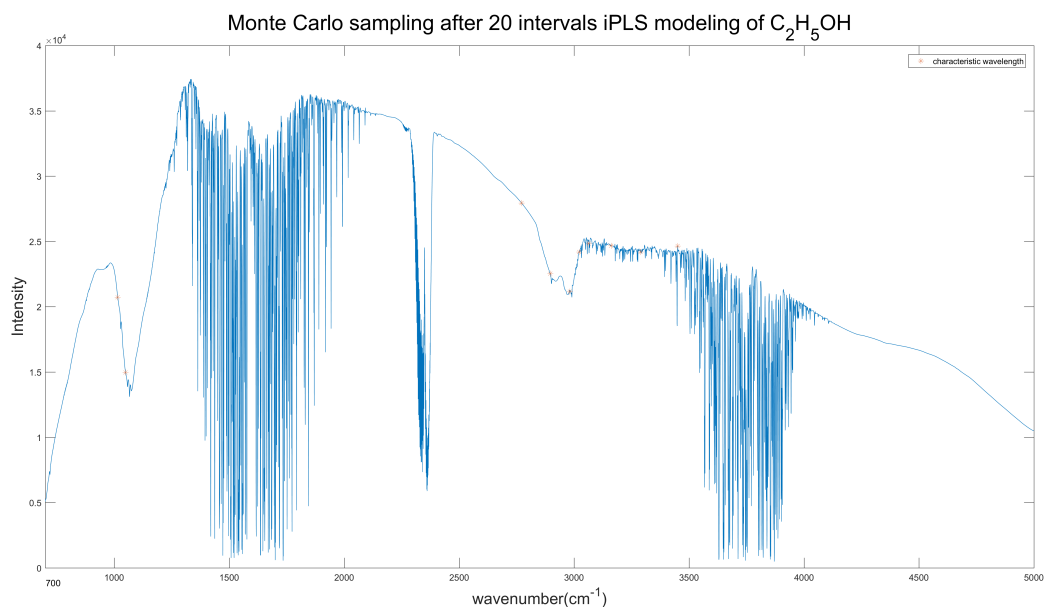


Fig. 11. Wavelengths selected by iPLS (20 intervals)-MC model of C_2H_5OH spectra.

For C_2H_5OH spectra, we selected 10 wavelengths each time by Monte-Carlo sampling and repeated 1000 times for the finally result. The 10 wavelengths selected from 20 intervals and 30 intervals iPLS model of C_2H_5OH spectra are shown in Figs. 11 and 12, respectively. The specific wavelength positions of each figure are listed in Table 7.

The results listed in Table 7 indicate that the Monte-Carlo sampling can select the characteristic

wavelengths from gas spectra efficiently, and the values of RMSECV and R_c are much smaller than the corresponding values of full spectra PLS model. In the experiment of C_2H_4 spectra modeling, we set the number of the characteristic wavelengths to 20 which is about 1.6% of the number of iPLS wavebands combination. The final results indicate that the proposed method gets excellent prediction effects. On the basis of the experiment of C_2H_4 , we

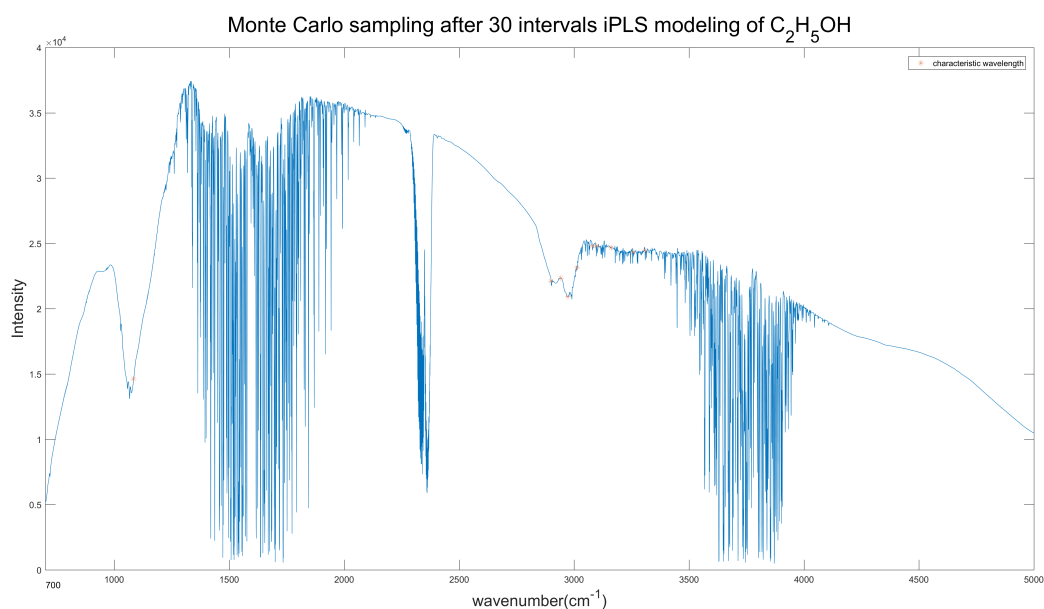


Fig. 12. Wavelengths selected by iPLS (30 intervals)-MC model of C_2H_5OH spectra.

Table 7. Results of iPLS-MC model.

Spectrum type	Ipls interval	Wavelength number	Wavelength position (cm ⁻¹)	RMSECV (ppm)	Rc
C ₂ H ₄	20	20	839, 907, 987, 1025, 1035, 1087, 1111, 14601, 546, 1823, 1905, 1907, 1951, 2891, 2942, 2963, 3002, 3057, 3077, 3273	0.3001	0.9999
	30	20	947, 953, 961, 1016, 1285, 1293, 1343, 1435, 1453, 1461, 1882, 1907, 2881, 2899, 2997, 3110, 3196, 3231, 3238, 3256	0.3364	0.9998
C ₂ H ₅ OH	20	10	1015, 1049, 2771, 2896, 2981, 3023, 3064, 3161, 3290, 3450	0.2977	0.9999
	30	10	1085, 2899, 2942, 2973, 3014, 3079, 3098, 3161, 3257, 3308	0.3109	0.9999

try to reduce the number of characteristic wavelengths number to 10 which only accounts for 0.7% of the C₂H₅OH spectrum wavebands selected by iPLS model, and the iPLS-MC modeling of C₂H₅OH get better prediction results than the experiment of C₂H₄ spectra. The reason is that in combinatorial mathematics, when the total sample points are determined, the smaller the number of points to be sampled is, the greater the probability of all possible sample combinations to be obtained will be. The VOC infrared spectrum usually contains several absorption peaks which can explain the feature of the certain gas, so the number of characteristic wavelengths can reduce to a small number corresponding to the number of absorption peaks.

Comparing different models on two VOC spectra, it is obvious that the iPLS-MC model can select the informative wavelengths effectively and accurately. The accuracy of prediction has been greatly improved.

3.3. Comparison of selected wavelengths by a different method

The performance of each model was evaluated by the prediction set randomly selected from the acquired spectral samples. Both prediction sets of two gases contain 24 samples with different concentrations. The scatter plots in Figs. 13 and 14 intuitively show the correlation between predicted and actual

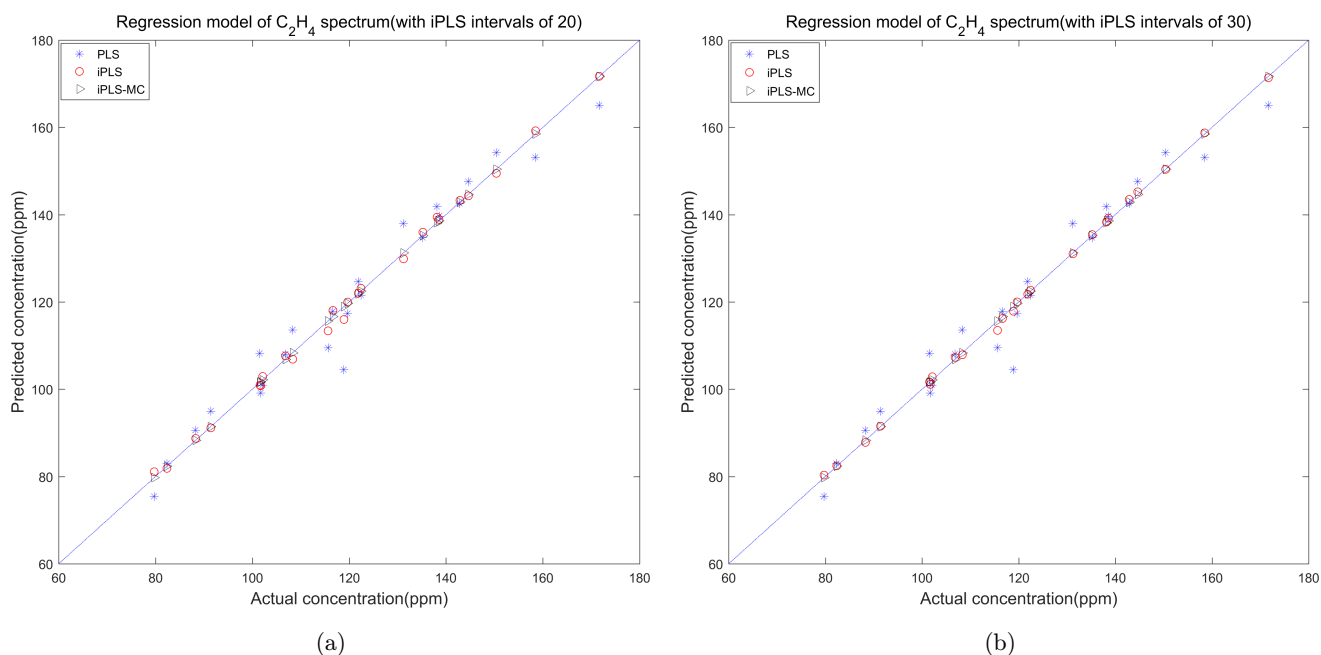


Fig. 13. Relationship between the predicted and actual concentration of C₂H₄ (a) iPLS intervals of 20 and (b) iPLS intervals of 30.

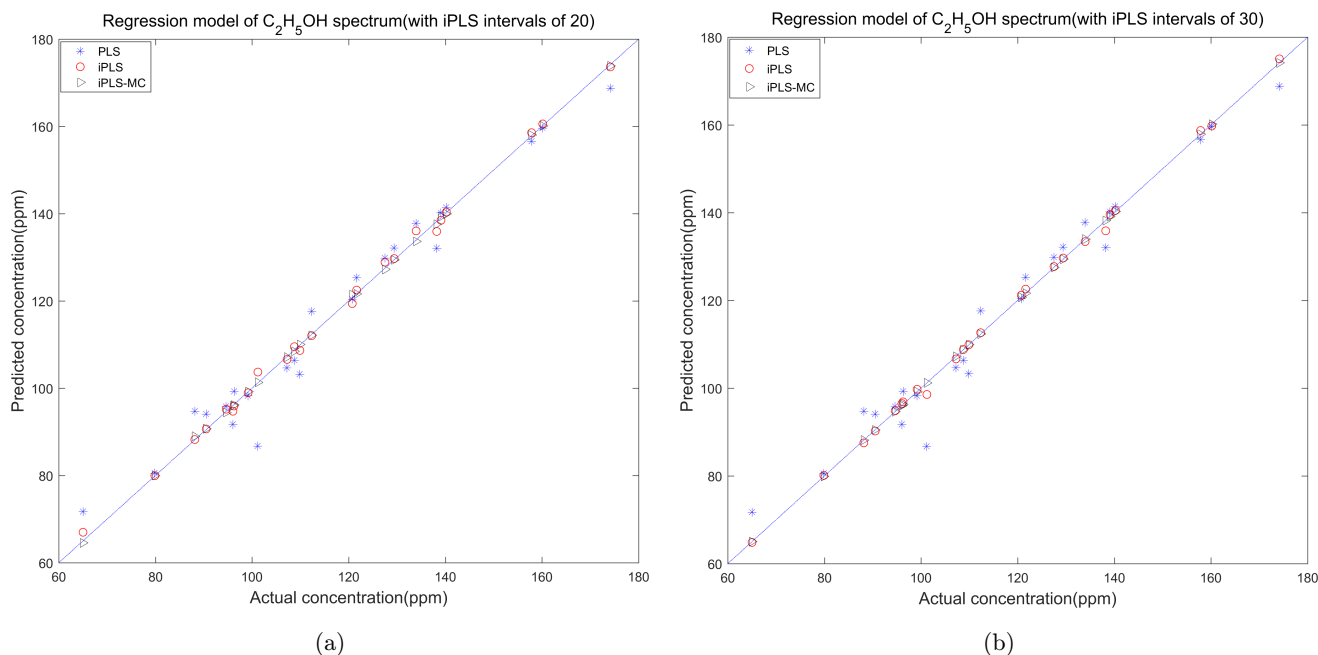


Fig. 14. Relationship between the predicted and actual concentration of C_2H_5OH (a) iPLS intervals of 20 and (b) iPLS intervals of 30.

concentrations from different modeling methods. In each figure, the star points represent the identification result of full spectrum PLS model, and the circle points and triangle points represent the results of iPLS model and iPLS-MC model, respectively. As we can see from the figures, the prediction errors of iPLS model and iPLS-MC model evidently decreased compared with the result of full spectrum PLS model. Especially in the iPLS-MC model, the predicted and actual concentrations are highly correlated.

Table 8 summarized the parameters and experiment results of three models performed on the quantitative analysis of VOC spectra. Compared with the preliminary number of full spectrum ($700\text{--}4999\text{ cm}^{-1}$, 4300 wavelengths), the characteristic wavebands selected from the iPLS model were

215 (20 intervals) and 143 (30 intervals), accounting for 5% and 3.3% of the preliminary number. Meanwhile, the RMSECV and RMSEP values reduced from 3 to about 0.6 ppm, and the R_c and R_p values increased from 0.96 to 0.998. The characteristic wavelengths selected from iPLS-MC model are 20 and 10, accounting for 0.4% and 0.2% of the preliminary number, and the RMSECV and RMSEP values reduced dramatically to 0.3 ppm, simultaneously the R_c and R_p values increased to 0.9999. All the parameters indicated the stability of the model which can get a satisfactory prediction performance in selecting characteristic wavelengths of VOC spectrum.

The prediction results of the quantitative identification models were significantly improved when employing characteristic wavelength selection, and the models were also simplified by using only a small

Table 8. Results of each model for two gas spectra.

Type parameters	C_2H_4					C_2H_5OH				
	number	RMSECV	R_c	RMSEP	R_p	number	RMSECV	R_c	RMSEP	R_p
PLS	4300	2.5827	0.9737	2.7349	0.9629	4300	3.0184	0.9625	3.2361	0.9594
iPLS	215	0.6662	0.9989	0.7043	0.9982	215	0.6726	0.9987	0.7043	0.9981
	143	0.6294	0.9990	0.6458	0.9989	143	0.5219	0.9996	0.7126	0.9983
iPLS-MC	20	0.3001	0.9999	0.3118	0.9999	10	0.2977	0.9999	0.3201	0.9999
	20	0.2846	0.9998	0.3007	0.9999	10	0.2904	0.9999	0.3193	0.9999

part of informative wavelengths. The results experimentally proved the necessity to perform characteristic wavelength selection before building a calibration model.

4. Conclusion

In this work, an improved characteristic wavelength selection method called iPLS-MC model was proposed to identify the concentration information of VOCs from its FTIR spectrum. As the spectra of different gases vary greatly, it's necessary to select the characteristic wavebands or wavelengths of gas spectrum before establishing the calibration models. First, we systematically collected the FTIR spectra of ethylene and ethanol in laboratory, then used the obtained spectra to build full-spectrum PLS model, iPLS model and iPLS-MC model, respectively. At last, we compared the experiment results of each model and got the following conclusions: (1) FTIR as a noncontact spectral technology integrated with PLS modeling can effectively detect the concentration information of VOCs. (2) Both iPLS method and iPLS-MC method can select the characteristic wavebands or wavelengths corresponding to the gas spectral absorption peak regions, and the prediction effects of models which used the selected wavebands or wavelengths were better than that of full-spectrum modeling. (3) iPLS-MC method improves the single selection of iPLS, it combines multiple wavebands screened by iPLS modeling and selects wavelengths from those wavebands by Monte-Carlo sampling to get the optimal wavelength combination. The performance of the regression model built by characteristic wavelengths from iPLS-MC model was superior in contrast with that by the iPLS model. In the iPLS-MC model, less than 0.4% of the total spectral wavelengths had been selected, the correlation coefficient of prediction set (R_p) by the selected wavelengths reached 0.9999. The results demonstrate that iPLS-MC method is a useful wavelength reduction tool for VOC spectra to retain useful information.

In this paper, for the first time, we put forward the variable selection method to optimize the VOC FTIR spectrum, meanwhile, we proposed a new characteristic wavelength selection method named iPLS-MC which achieved high accuracy prediction performance in laboratory conditions. The proposed characteristic wavelength selection method provides a new thought to make the online atmospheric

environment pollution monitoring and other spectroscopic analysis simpler and more effective. It is noteworthy that the experiment was carried out with VOC gas as the carrier because VOC gas detection is a key issue in the prevention and control of atmospheric environmental pollution at present. This method is applicable not only to VOC gas but also to any other asymmetric polyatomic gas with infrared spectrum absorption, such as CO_2 , SO_2 , SO_3 , etc. In order to improve the applicability of the model, different kinds of asymmetric polyatomic gases should be tested in this model.

Moreover, when the gas mixtures with characteristic wavebands overlapped in the same spectral region, the results should be discussed in three cases: (1) The concentration of one gas is known, while the concentrations of other mixed gases change little, this method can accurately select the characteristic wavelengths of the gas with the known concentration. As shown in Fig. 1, one absorption peak of C_2H_4 overlaps with water vapor's absorption peak in the waveband near 1500 cm^{-1} , the proposed method can select the characteristic wavelengths of C_2H_4 accurately. (2) The concentration of one gas is known, the concentrations of other mixed gases change in the same proportion to the known gas, this method can select the characteristic wavelengths in the overlapped region, but the determination of gas type requires assistance of other selected absorption peaks. (3) The concentration of the mixed gases varies in different proportions, the linear superposition of absorption peaks in the same wavelengths results in mismatches with concentration information, errors or failure of fitting may occur when spectral PLS is calibrated quantitatively. The new method cannot work in this situation because it based on PLS model.

The number of the wavelengths that participated in the Monte-Carlo sampling is not determined in this study, how to choose the optimal number of characteristic wavelengths should be studied in the future research.

Conflict of Interest

The authors declare that there is no potential conflict of interest related to this manuscript.

Acknowledgements

The authors are grateful to Prof. Minguang Gao for his technical assistance. This work was

supported by National Key Scientific Instrument and Equipment Development Project of China, Grant Nos. 2013YQ220643, the National 863 Program of China, Grant Nos. 2014AA06A503.

References

1. A. K. Boulamanti, C. J. Philippopoulos, "Photocatalytic degradation of methyl tertbutyl ether in the gas-phase: A kinetic study," *J. Hazard. Mater.* **160**, 83–87 (2008).
2. R. Atkinson, J. Arey, "Atmospheric degradation of volatile organic compounds," *Chem. Rev.* **103**, 4605–4638 (2003).
3. World Health Organization, Geneva, Switzerland, <http://www.who.int/cancer/prevention/en/>.
4. J. Lelieveld, T. M. Butler, J. N. Crowley, "Atmospheric oxidation capacity sustained by a tropical forest," *Nature* **452**, 737–740 (2008).
5. M. C. Wu, S. L. Kuo, J. C. Lin, C. M. Ma, "Studies on certain physical properties of modified smectite nanocatalysts," *Appl. Surf. Sci.* **257**, 5641–5646 (2011).
6. B. Derstroff, I. Huser, E. Bourtsoukidis, "Volatile organic compounds (VOCs) in photochemically aged air from the eastern and western Mediterranean," *Atmos. Chem. Phys.* **17**, 9547–9566 (2017).
7. B. Barletta, S. Meinardi, F. S. Rowland, "Volatile organic compounds in 43 Chinese cities," *Atmos. Environ.* **39**, 5979–5990 (2005).
8. P. Patnaik, *A Comprehensive Guide to the Hazardous Properties of Chemical Substances*, John Wiley & Sons, US (2007).
9. H. Lee, D. H. Lee, "Synergistic effect of non-thermal plasma–catalysis hybrid system on methane complete oxidation over Pd-based catalysts," *Chem. Eng. J.* **259**, 761–770 (2015).
10. M. S. Gandhi, Y. S. Mok, "Effect of various parameters for butane decomposition under ambient temperature in a dielectric barrier discharge non-thermal plasma reactor," *J. Taiwan. Inst. Chem. E* **44**, 786–794 (2013).
11. X. Y. Zhang, B. Gao, A. E. Creamer, "Adsorption of VOCs onto engineered carbon materials: A review," *J. Hazard. Mater.* **338**, 102–123 (2017).
12. S. K. Veerapandian, C. Leys, "Abatement of VOCs using packed bed non-thermal plasma reactors: A review," *Catalysts*. **7**, 113–145 (2017).
13. Y. K. Feng, W. L. Wang, Y. C. Wang, "Experimental study of destruction of acetone in exhaust gas using microwave-induced metal discharge," *Sci. Total. Environ.* **645**, 788–795 (2018).
14. W. M. Park, J. B. Park, J. Roh, S. H. Hwang, "Levels of formaldehyde and TVOCs and influential factors of 100 underground station environments from 2013 to 2015," *Hum. Ecol. Risk Assess.* **24**, 1030–1042 (2017).
15. A. J. Buczynska, A. Krata, M. Stranger, "Atmospheric BTEX-concentrations in an area with intensive street traffic," *Atmos. Environ.* **43**, 311–318 (2009).
16. L. K. Sahu, D. Pal, R. Yadav, "Aromatic VOCs at major road junctions of a metropolis in India: Measurements using TD-GC-FID and PTR-TOF-MS instruments," *Aerosol. Air. Qual. Res.* **16**, 2405–2420 (2016).
17. J. F. Liu, Y. J. Mu, Y. J. Zhang, "Atmospheric levels of BTEX compounds during the 2008 Olympic Games in the urban area of Beijing," *Sci. Total. Environ.* **408**, 109–116 (2009).
18. M. R. Belanger, J. R. Pare, R. Turpin, "Evaluation of Microwave-Assisted Process technology for HAPSITE's headspace analysis of volatile organic compounds (VOCs)," *J. Hazard. Mater.* **145**, 336–338 (2007).
19. G. Joost, W. Carsten, "Measurements of volatile organic compounds in the earths atmosphere using proton-transfer-reaction mass spectrometry," *Mass. Spectrom. Rev.* **26**, 223–257 (2007).
20. L. Cui, Z. Zhang, Y. Huang, "Measuring OVOCs and VOCs by PTR-MS in an urban roadside microenvironment of Hong Kong: Relative humidity and temperature dependence, and field intercomparisons," *Atmos. Meas. Tech.* **9**, 5763–5779 (2016).
21. R. S. Blake, K. P. Wyche, A. M. Ellis, "Chemical ionization reaction time-of-flight mass spectrometry: Multi-reagent analysis for determination of trace gas composition," *Int. J. Mass. Spectrom.* **254**, 85–93 (2006).
22. A. R. Koss, C. Warneke, B. Yuan, "Evaluation of NO+ reagent ion chemistry for online measurements of atmospheric volatile organic compounds," *Atmos. Meas. Tech.* **9**, 2909–2925 (2016).
23. J. Hodgkinson, R. P. Tatam, "Optical gas sensing: A review," *Meas. Sci. Technol.* **24**, 1–95 (2013).
24. X. B. Zhou, J. W. Zhao, "Variables selection methods in near-infrared spectroscopy," *Anal. Chim. Acta.* **667**, 14–32 (2010).
25. C. S. Miaw, C. Assis, "Determination of main fruits in adulterated nectars by ATR-FTIR spectroscopy combined with multivariate calibration and variable selection methods," *Food. Chem.* **254**, 272–280 (2018).
26. A. Durand, O. Devos, C. Ruckebusch, J. P. Huvenne, "Genetic algorithm optimisation combined with partial least squares regression and mutual information variable selection procedures in near-infrared quantitative analysis of cotton-viscose textiles," *Anal. Chim. Acta.* **595**(1–2), 72 (2007).

27. J. B. Li, C. J. Zhao, W. Q. Huang, "A combination algorithm for variable selection to determine soluble solid content and firmness of pears," *Anal. Meth.* **6**, 2170–2180 (2014).
28. Q. J. Han, H. L. Wu, C. B. Cai, L. Xu, R. Q. Yu, "An ensemble of Monte-Carlo uninformative variable elimination for wavelength selection," *Anal. Chimica. Acta.* **612**, 121–125 (2008).
29. W. Fan, Y. Shan, G. Y. Li, "Application of competitive adaptive reweighted sampling method to determine effective wavelengths for prediction of total acid of vinegar," *Food. Anal. Meth.* **5**, 585–590 (2012).
30. J. J. Li, C. Y. Ma, "Efficient elimination of trace ethylene over nano-gold catalyst under ambient conditions," *Environ. Sci. Technol.* **42**, 8947–8951 (2008).
31. M. Mofijur, M. G. Rasul, J. Hyde, "Role of biofuel and their binary (diesel-biodiesel) and ternary (ethanol-biodiesel-diesel) blends on internal combustion engines emission reduction," *Renew. Sust. Energ. Rev.* **53**, 265–278 (2016).
32. W. Ju, C. H. Lu, Y. J. Zhang, "Open-path fourier transform infrared spectrum de-noising based on improved threshold lifting wavelet transform and adaptive filter," *Spectrosc. Spect. Anal.* **38**, 1684–1690 (2018).
33. X. Chu, W. Wang, "Identifying camellia oil adulteration with selected vegetable oils by characteristic near-infrared spectral regions," *J. Innov. Opt. Health. Sci.* **11**, 1850006-1–1850006-12 (2018).
34. X. Y. Niu, F. Shen, "Analysis of sugars in Chinese rice wine by fourier transformnear-infrared spectroscopy with partial least-squares regression," *J. Agric. Food. Chem.* **56**, 7271–7278 (2008).
35. A. S. L. Nørgaard, J. Wagner, J. P. Nielsen, "Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy," *Appl. Spectrosc.* **54**, 413–419 (2000).
36. C. O. Hendren, A. R. Badireddy, E. Casman, M. R. Wiesner, "Modeling nanomaterial fate in wastewater treatment: Monte-Carlo simulation of silver nanoparticles (nano-Ag)," *Sci. Total. Environ.* **449**, 418–425 (2013).
37. G. Herzberg, *Molecular Spectra and Molecular Structure*, D. Van Nostrand Company (1986).