**CHINESE JOURNAL OF ANALYTICAL CHEMISTRY**
Volume 43, Issue 2, February 2015
Online English edition of the Chinese language journal



**Cite this article as:** Chin J Anal Chem, 2015, 43(2), 226–232.

**RESEARCH PAPER**

# Qualitative Analyses of Open-Path Fourier Transform Spectra

**YU Lian-Lian, SHAO Li-Min***

Department of Chemistry, University of Science and Technology of China, Hefei 230026, China

**Abstract:** Open-path Fourier transform infrared (OP/FT-IR) spectra were arranged into data matrices and then analyzed with target factor analysis (TFA) and band-target entropy minimization (BTEM), respectively, aiming to reconstruct the spectra of interested molecules. Five sessions of continuous OP/FT-IR monitoring were carried out around farms. For each session, the spectra were arranged row-wisely in the order of measuring time, which yielded five matrices for data analysis. The analyses results showed that both TFA and BTEM could reconstruct spectral features of target molecules from the spectra data matrix, but performance of the two methods differed slightly. TFA can retrieve spectral features of target molecules in the presence of interferences, and the reconstructed spectrum is similar to corresponding reference. BTEM can implement such spectral retrieval without the reference spectrum. This paper presents not only the application of BTEM method to qualitative analyses of OP/FT-IR spectra, but also a thorough comparison between TFA and BTEM, which is helpful to qualitative analysis of complex multi-component systems.

**Key Words:** Open-path Fourier transform infrared spectrometry; Band target entropy minimization; Target factor analysis

## 1 Introduction

Open-path Fourier transform infrared (OP/FT-IR) spectrometry is an effective and widely applied tool for atmospheric measurements[1–4], with the advantages of reliable hardware, easy deployment, free of sampling process, and unattended monitoring[5–8]. A significant disadvantage of OP/FT-IR is complex spectra that are difficult to analyze. OP/FT-IR spectra contain intense vibration-rotation lines of atmospheric water vapor and carbon dioxide, and slow baseline variations, all of which seriously affect the analysis of trace components. Furthermore, most qualitative analysis in OP/FT-IR spectroscopy is based on visual inspection or spectra library searching, both require that spectral features of the interested molecule be recognizable. Such requirement is hardly met in the presence of aforementioned interferences, and the performance of qualitative analysis degrades. In this paper, target factor analysis (TFA) and band-target entropy minimization (BTEM) were used to extract obscured spectral information of $NH_3$, $C_2H_5OH$ and $O_3$ from OP/FT-IR spectra, and compared the performance of the two methods.

TFA is a self-modeling technique, and widely applied to data analyses in chromatography[9], reaction mechanics and kinetics[10,11], spectroscopic analysis[12], medical and pharmaceutical research[13,14], and food science field[15].

BTEM method is a self-modeling curve resolution technique for reconstructing the spectrum of a pure component from a series of multi-component spectra. One significant strength of BTEM is that it neither requires a priori knowledge of the interested molecule nor relies on statistical tests[16]. BTEM was used for analyses of various liquid and solid phase reaction systems[17], spectral reconstructions of FT-IR[18], Raman[19], NMR[20], MS[21] and so on. However, the application of BTEM to the analysis of OP/FT-IR spectroscopic data has not yet been reported.

Widjaja *et al*[22] compared BTEM and other self-modeling curve resolution techniques, such as interactive principal component analysis (IPCA), orthogonal projection approach-alternating least squares (OPA-ALS), etc. But no comparative study of BTEM and TFA has been reported yet. In this paper,

Received 15 September 2014; accepted 24 October 2014
*Corresponding author. Email: lshao@ustc.edu.cn
This work is sponsored by the National Natural Science Foundation of China (Nos. 21175123), the Program for New Century Excellent Talents in University of China (No. NCET-11-0878), the Program for Changjiang Scholars and Innovative Research Team in University of China (PCSIRT), and the Fundamental Research Funds for the Central Universities of China (No. wk2060190040).

DOI: 10.1016/S1872-2040(15)60804-0

we performed a detailed comparison between BTEM and TFA in qualitative analyses of OP/FT-IR spectroscopic data. The results showed both TFA and BTEM could yield comparable spectrum to the reference for either major or minor components, and the reconstructed spectra of the two methods for the same component are fairly similar. For BTEM, it is noteworthy that the performance depends on the choice of the objective function in its algorithm. This study not only applied BTEM method to qualitative analyses of OP/FT-IR spectra, but also compared the results obtained from TFA and BTME, providing guide and references to qualitative analysis of complex multi-component systems.

## 2 Experimental

### 2.1 Instrument and OP/FT-IR measurement

OP/FT-IR measurements were carried out on and around a dairy and a hog farm in southern Idaho in a cooperative project with the United States Department of Agriculture (USDA) for monitoring gaseous emissions.

The OP/FT-IR spectrometer was manufactured by MDA Corp. (Atlanta, GA), and incorporated with a Bomem Michelson 100 interferometer (Bomem, Canada), a 31.5-cm telescope to expand the IR beam being passed to a cube-corner array retroreflector, and a Sterling engine-cooled mercury cadmium telluride (MCT) detector.

### 2.2 Spectral data processing

OP/FT-IR interferograms were measured at a nominal resolution of 1 $cm^{-1}$, and corrected for nonlinear response of the MCT detector[23]. All spectra for analyses were computed with a zero-filling factor of 8 and Norton-Beer medium apodization. All manipulation of spectra and data processing were done using MATLAB software (The MathWorks Inc., Natick MA).

Five successive monitoring were performed. In each session, OP/FT-IR spectra within 750 and 1250 $cm^{-1}$ were arranged row-wisely to form a data matrix. The consequent five data matrices contain 1177, 600, 811, 332 and 1184 spectra, respectively.

## 3 Theory and calculation

$m$ absorbance spectra were generated from a continuous OP/FT-IR monitoring session and each spectrum had $n$ points. An $m$-by-$n$ matrix, $\boldsymbol{D}_{m\times n}$, was obtained by arranging those spectra in a row-wise manner. When Beer's law is obeyed, $\boldsymbol{D}_{m\times n}$ is the product of a concentration matrix $\boldsymbol{C}_{m\times p}$ and an absorptivity matrix $\mathbf{S}_{n\times p}$.

$$\boldsymbol{D}_{m\times n} = \boldsymbol{C}_{m\times p}\boldsymbol{S}_{n\times p}^{\mathrm{T}} \quad (1)$$

where, $p$ is the number of pure compounds, and superscript T indicates the transpose of the matrix.

Performing principal component analysis (PCA) to $\boldsymbol{D}_{m\times n}$ yields.

$$\boldsymbol{D}_{m\times n} = \boldsymbol{U}_{m\times q}\boldsymbol{V}_{n\times q}^{\mathrm{T}} + \boldsymbol{R}_{m\times n} \quad (2)$$

where, $q$ is the number of principal components and theoretically equal to $p$ in Eq. (1). $\boldsymbol{U}$ and $\boldsymbol{V}$ are the score and loading matrices that contain principal components and eigenvectors, respectively. $\boldsymbol{R}$ is the residual matrix. By eliminating $\boldsymbol{R}$, the following equation can be derived:

$$\boldsymbol{C}_{m\times p}\boldsymbol{S}_{n\times p}^{\mathrm{T}} = \boldsymbol{U}_{m\times q}\boldsymbol{V}_{n\times q}^{\mathrm{T}} \quad (3)$$

Equation (3) shows that $\boldsymbol{V}$ contains all spectral information in $\boldsymbol{S}$, but in a different form, which is why eigenvectors are also called abstract spectra. Eigenvectors span the same space as real spectra do, so any spectrum in $\boldsymbol{S}$, say $\boldsymbol{S}_{n\times 1}$, can be expressed as a unique linear combination of eigenvectors,

$$\boldsymbol{S}_{n\times 1} = \boldsymbol{V}_{n\times q}\boldsymbol{T}_{q\times 1} \quad (4)$$

where, $\boldsymbol{T}_{q\times 1}$ is a rotation vector.

### 3.1 Target factor analysis

TFA is a chemical factor analysis method proposed by Malinowski[24]. By testing whether or not the chemical information of the interested compound, i.e. the target, is present in the measured data matrix, the presence or absence of the target in the sample is confirmed.

Equation (4) is the basis of TFA. Rotation vector $\boldsymbol{T}$ is obtained through least-square regression:

$$\boldsymbol{T}_{q\times 1} = (\boldsymbol{V}_{n\times q}^{\mathrm{T}}\boldsymbol{V}_{n\times q})^{-1}\boldsymbol{V}_{n\times q}^{\mathrm{T}}\boldsymbol{S}_{n\times 1} \quad (5)$$

where, $\boldsymbol{S}_{n\times 1}$ is a reference spectrum of the target. After rotation vector $\boldsymbol{T}$ is calculated, the reconstructed spectrum is readily obtained with Eq.(4). Finally, the reconstructed spectrum is compared to the reference, and if the similarity is sufficiently high, the presence of the target in the sample is confirmed.

### 3.2 Band target entropy minimization

Band target entropy minimization (BTEM) method is a self-modeling curve resolution method that incorporates the concept of entropy minimization[22]. In BTEM, certain spectral features in eigenvectors are inspected and ascribed to a target molecule, and then an objective function is employed to retain such spectral features in the reconstructed spectrum to a great extent and to minimize interferences of other components at the same time. In order to optimize the objective function, simulated annealing method (SA) was used. After the objective function is optimized, the corresponding rotation vector $\boldsymbol{T}$ is considered to be the most suitable one, and with it the reconstructed spectrum is calculated through Eq.(4). In SA, the initial value of the rotation vector $\boldsymbol{T}$ is randomly generated, resulting in some different reconstructed spectra after multiple runs.

The objective function in BTEM algorithm is shown as Eq.(6)[22]:

$$G = H(\boldsymbol{T}) + P(\boldsymbol{T}) \quad (6)$$

where, $H(\boldsymbol{T})$ is the Shannon entropy function, $P(\boldsymbol{T})$ is the penalty function. Shannon simplified the spectra by minimizing entropy function in their study[25]. Shannon Entropy reflects the discrete degree of spectral information across the whole spectrum. Therefore, the minimization of entropy would localize spectral features in a spectrum with smooth baseline, and the spectrum is simplified to the greatest extent. Penalty function $P$ ensures non-negativity of both the spectrum and the concentration of the target. The closer the penalty function is to zero, the more optimal the objective function will be [26].

## 4 Results and discussion

### 4.1 Determination of the number of principal components

Figure 1 shows the first ten eigenvectors from the PCA analysis of matrix $\boldsymbol{D}_{1177\times2075}$ of data set #1. It was found that the first eight eigenvectors exhibited clear spectral features, and the rest appeared to be noise. Therefore, this data matrix was supposed to have eight principal components, so eight eigenvectors were used in subsequent analysis. For the other four data sets, the numbers of principal components were determined in the same way as 6, 10, 6 and 9, respectively.

### 4.2 Results of TFA and BTEM

As shown in Fig.1, EV1 and EV2 were primarily absorbance of water vapor; EV3 showed two characteristic bands at 900–1000 cm$^{-1}$ labeled as i and j; EV8 showed a characteristic band in 1000–1100 cm$^{-1}$ labeled as k. Wavenumber ranges of these characteristic bands are given in the caption of Fig.1.

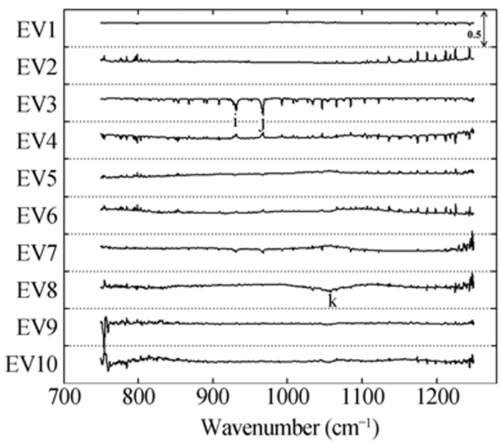


Fig.1 The first ten eigenvectors calculated for date set #1

i, j, and k represent target bands used in BTEM while reconstructing spectra, respectively. The wavenumber ranges are 931–933 cm$^{-1}$, 967–968 cm$^{-1}$, 1100–1102 cm$^{-1}$

Characteristic bands i and j in Fig.1 were used separately in BTEM to construct spectra of possible target molecules. The results showed that both reconstructed spectra were similar to the reference spectrum of $NH_3$, and the correlation coefficients between the reconstructed spectra and reference were 0.9394 and 0.9408, respectively. Therefore, the presence of $NH_3$ in air was confirmed when spectra of data set #1 were measured. It is noteworthy that spectra of $NH_3$ were successfully reconstructed in spite of significant interferences from water vapor in 1150–1250 cm$^{-1}$. The two reconstructed spectra by using characteristic bands i and j were found to be very similar with similar correlation coefficients. Therefore, band was sufficient in reconstructing a good spectrum of $NH_3$, and band j was chosen in the analyses of other data sets

In order to confirm the presence of other components in data set #1, TFA was used for spectral reconstruction with the reference spectra of $C_2H_5OH$ and $O_3$. Both components have a characteristic absorption band within 1000–1100 cm$^{-1}$. The reconstructed spectra from TFA with $C_2H_5OH$ and $O_3$ as the target respectively were similar to the reference spectrum of $C_2H_5OH$, as shown in Fig.2B and Fig.2C. Therefore, $C_2H_5OH$ was present in air and $O_3$ was absent. The reconstructed spectra from BTEM by using characteristic band $k$ were also given in Fig.2, and the presence of $C_2H_5OH$ was further confirmed. The correlation coefficients between the reconstructed spectra and the reference of $C_2H_5OH$ were calculated to be 0.9476 (TFA) and 0.9058 (BTEM). The correlation coefficients between the reconstructed spectra and the reference of $O_3$ were also calculated as 0.8328 (TFA) and 0.7682 (BTEM). From the above results, the presence of $C_2H_5OH$ in data set #1 could be confirmed.

Although both TFA and BTEM confirmed the presence of $C_2H_5OH$, the reconstructed spectra from the two methods were different. As shown in Fig.2, the spectrum reconstructed by TFA had less noise and water vapor absorption in the range of 750–760 cm$^{-1}$ and 1150–1250 cm$^{-1}$ than that reconstructed by BTEM. The spectra reconstructed by BTEM contained weak but clear information of $NH_3$, eg. two absorption peaks in 900–980 cm$^{-1}$ range. All the findings above indicated a better performance of TFA over BTEM. The main reason for this was that TFA made use of the reference spectrum of $C_2H_5OH$, while BTEM had no reference at all. Another reason was that the spectral information of $NH_3$ was fairly strong in this data set, and its effect was considerably difficult to eliminate in BTEM.

For further comparison between TFA and BTEM, the rotation vectors obtained by both methods were put into the objective function of BTEM, and values of the objective function were obtained as listed in Table 1.

According to Table 1, the objective function value G of TFA was smaller than that of BTEM. The main reason was that the Shannon entropy value of $H$ by TFA method was smaller, which was due to the fact that the spectral reconstruct-

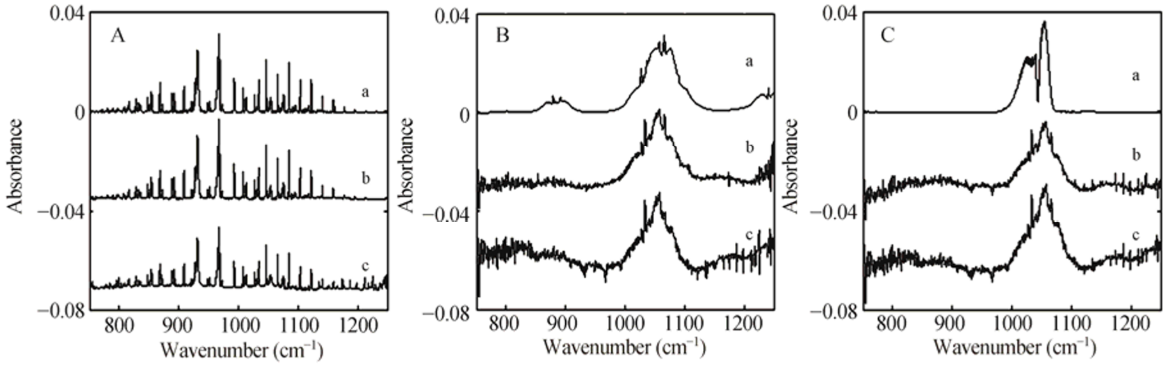


Fig.2 Analysis result of data set 1

A, B and C represent the reconstructed spectrum in the data set 1. a, b and c correspond to the reference spectra, the reconstructed spectra by TFA and BTEM, respectively

Table 1 Values in the objective function of BTEM with rotation vectors obtained by TFA and BTEM

| Value | Ammonia | | Ethanol | |
|---|---|---|---|---|
| | Target factor analysis | Band-target entropy minimization | Target factor Analysis | Band-target entropy minimization |
| *P* | 0.0191 | 0.0029 | 0.0064 | 0.00014 |
| *H* | 1.7154 | 1.9622 | 2.0159 | 2.0309 |
| *G* | 1.7345 | 1.9651 | 2.0223 | 2.0310 |

tion by TFA could make use of information from reference spectrum of target molecule. As to BTEM, no additional information could be used in spectral reconstruction, which was more likely to include interferences into the reconstructed spectrum, and the objective function yields a large value. This conclusion was in accordance with the results of some spectra in Fig.2A and Fig.2B. In addition, values of penalty function *P* of the two methods were close to zero, and thus had little impact on the objective function value *G*. Therefore, neither spectral information nor concentrations had serious non-negativity during the process of spectral reconstruction.

The reconstructed spectra from other four data sets are shown in Fig.3. As shown in Fig.3A, the spectrum of $NH_3$ was successfully reconstructed from other data sets except for data set #2 in which spectra were measured far away from the farm with negligible $NH_3$ in air.

Spectra of $C_2H_5OH$ were successfully reconstructed from data sets #2 and #4, as shown in Fig.3B. For data set #2, spectral features of $C_2H_5OH$ were clear, but the interference from $NH_3$ was obvious too. In this data set, the spectral information of $NH_3$ was actually too weak to be retrieved, so its presence in the reconstructed spectrum of $C_2H_5OH$ meaned the rather low concentration of $C_2H_5OH$. For data set #4, the reconstructed spectrum of $C_2H_5OH$ was contaminated only by interference of water vapour, resulting from the relatively high concentration of $C_2H_5OH$ and strong IR absorption. For data set #3, the reconstructed spectrum by TFA showed spectral features of $O_3$, although the reference used was $C_2H_5OH$. Therefore, $O_3$ was definitely present in data set #3 although no $O_3$ was observed in original spectrum. The correlation coefficients between the reconstructed spectra and the reference of $C_2H_5OH$ were calculated to be 0.8108 (TFA) and 0.7535 (BTEM), further proving the absence of $C_2H_5OH$.

Spectra of $O_3$ were successfully reconstructed from data sets #3 and #5 as shown in Fig.3C. The reconstructed spectra by both methods showed clear spectral features of $O_3$, but interferences were obvious too, which was probably due to weak spectral information of $O_3$ in the two data sets. For data set #3, neither TFA nor BTEM could reconstruct meaningful spectra, so $O_3$ was absent in this data set. Interesting results were obtained with data set #4. With the intention to retrieve spectral information of $O_3$, we actually had that of $C_2H_5OH$.

In Fig.3, B-c and C-c are reconstructed spectra from the same data set by BTEM using same band. The two spectra have the same spectral features of $O_3$, but with minor differences. It is because that in BTEM the initial value of rotation vector **T** was random, so the reconstructed spectra from multiple runs could not be identical. Nonetheless, interested spectral features in the reconstructed spectrum were sufficiently retained, which ensured the reliability of BTEM.

Besides visual comparison, the similarity between reconstructed and reference spectra was also evaluated using correlation coefficient, as listed in Table 2. It was found that correlation coefficient was an effective measure of similarity between two spectra, and thus could help confirm the presence or absence of the interested molecule. However, for data set #5, the correlation coefficient for $O_3$ was 0.7861 (TFA), while the spectral feature of $O_3$ was recognizable in the reconstructed spectra as shown in Fig.3C, which was due to the large noise in the spectra and the severe interferences from $NH_3$ and water vapor. So the correlation coefficient for $O_3$ was less than 0.85 despite the presence of characteristic absorption of $O_3$ in the reconstructed spectra. Therefore, it is more reliable to use both correlation coefficient and visual inspection in qualitative analysis.

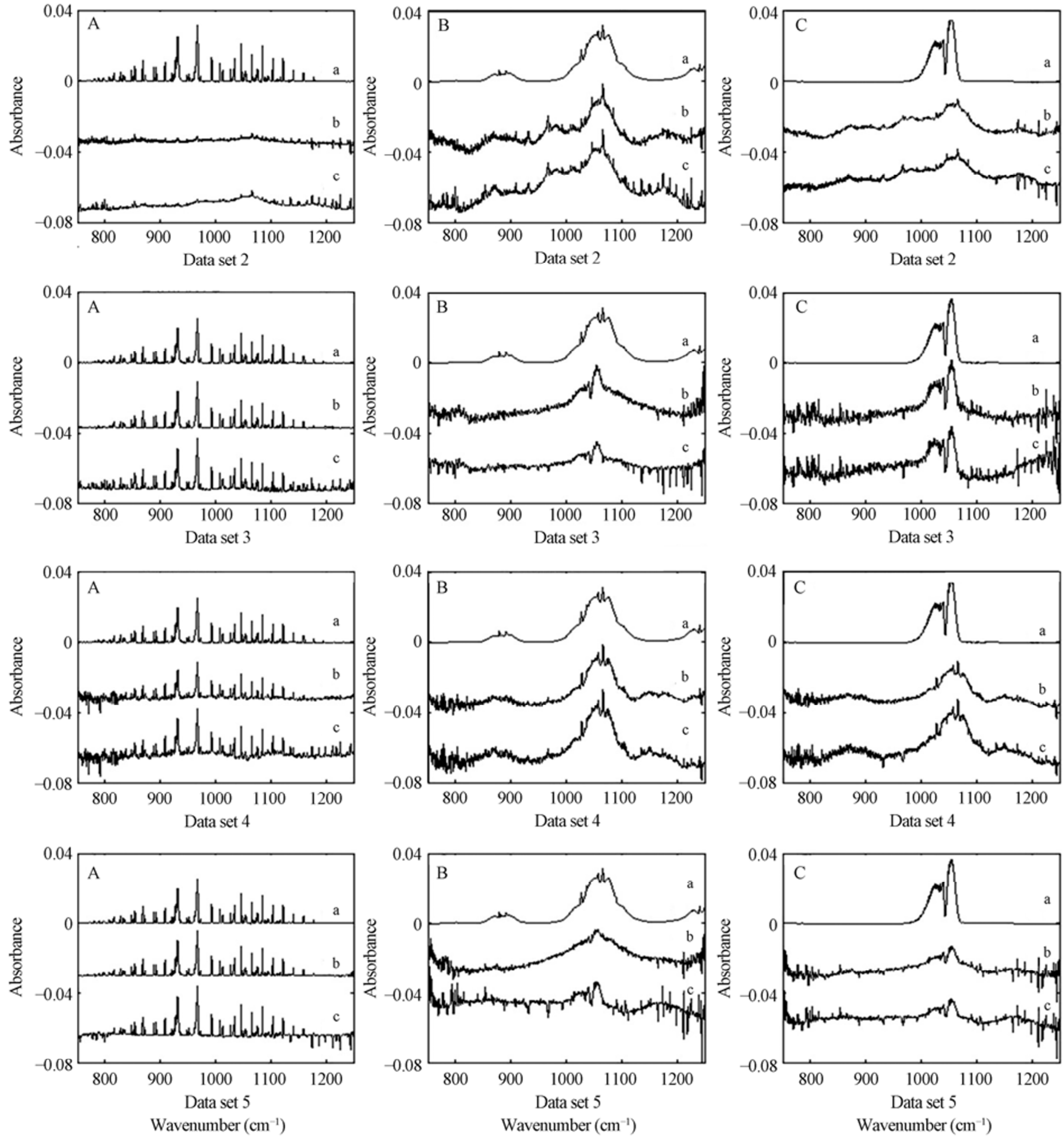


Fig.3 Analysis results of data sets #2 to #5

A, B and C represent the reconstructed spectra in the data sets 2–5; a, b and c correspond to the reference spectra, the reconstructed spectrum by TFA and BTEM, respectively

Table 2 Correlation coefficients between reconstructed and reference spectra of target molecules

| Data set | Ozone | | Ethanol | | Ammonia | |
|---|---|---|---|---|---|---|
| | Target factor analysis | Band-target entropy minimization | Target factor analysis | Band-target entropy minimization | Target factor analysis | Band-target entropy minimization |
| 1 | 0.8328 | 0.7682 | 0.9476 | 0.9058 | 0.9859 | 0.9408 |
| 2 | 0.6640 | 0.6324 | 0.8666 | 0.8531 | 0.3349 | 0.3271 |
| 3 | 0.9129 | 0.8995 | 0.8437 | 0.7689 | 0.9956 | 0.9775 |
| 4 | 0.7086 | 0.6727 | 0.9541 | 0.8835 | 0.8853 | 0.8832 |
| 5 | 0.7861 | 0.7075 | 0.8108 | 0.7535 | 0.9933 | 0.9805 |

## 5 Conclusions

TFA has been used as a main method with proved performance in qualitative analyses of OP/FT-IR spectroscopy, but there has been no report on the use of BTEM in qualitative analyses of OP/FT-IR spectroscopy. When it applied to qualitative analyses of OP/FT-IR spectroscopic data, BTEM does not require a prior information of the target, as opposed to TFA, BTEM does not require a prior information of the target. This work provides a new confirmative method in qualitative analyses of OP/FT-IR spectroscopy of multi-component systems.

## References


[1] Liu W Q, Cui Z C, Dong F Z. *Optoelectronic Technol. Inform.* , **2002**, 15: 1–12

[2] Sun X Y, Li Y, Wang J D. *Spectroscopy and Spectra Analysis*, **2003,** 23: 739–741

[3] Xu L, Liu J G, Lu Y H, Liu W Q. *Spectroscopy and Spectra Analysis*, **2007**, 27: 889–891

[4] Wang Y, Zhang Y C, Li N, Zhao X H. *Opto-Electronic Engineering*, **2008**, 35: 73–81

[5] Newman A R. *Anal. Chem*., **1997**, 69: 43A–47A

[6] Walter W T. *Proc. SPIE.,* **2004**, 5270: 144–150 (Environmental Monitoring and Remediation III)

[7] Childers J W, Thompson E L, Harris D B, Kirchgessner D A, Clayton M, Natschke D F, Phillips W J. *Atmos. Environ.,* **2001**, 35: 1923–1936

[8] Wiseman J M, Ifa D R, Song Q Y, Cooks R G. *Angew. Chem. Int. Ed.*, **2006**, 45: 7188–7192

[9] van Zomeren P V, Metting H J, Coenegracht P M J, de Jong G J. *J. Chromatogr. A*, **2005**, 1096: 165–176

[10] Carvalho A R, Wattoom J, Zhu L, Brereton R G. *Analyst,* **2006**, 131: 90–97

[11] Day J P R, Campbell R A, Russell O P, Bain C D. *J. Phys. Chem. C*, **2007**, 111: 8757–8774

[12] Kuzmanovski I, Trpkovska M, Soptrajanov B, Stefov V. *Vib. Spectrosc*., **1999**, 19: 249–253

[13] Tetteh J, Mader K T, Andanson J M, McAuley W J, Lane M E, Hadgraft J, Kazarian S G, Mitchell J C. *Anal. Chim. Acta*., **2009**, 642: 246–256

[14] Kauffmana J F, Dellibovi M, Cunninghamb C R. *J. Pharm. Biomed. Anal*., **2007**, 43: 39–48

[15] Monfreda M, Gobbi L, Grippa A. *Food. Chem.,* **2014**, 145: 584–592

[16] Widjaja E, Li C Z, Garland M. *J. Comb. Chem.*, **2009**, 11: 261–266

[17] Widjaja E, Lee W L, Loo SCJ. *Anal. Chem.*, **2010,** 8: 1277–1282

[18] Tjahjono A M, Feng G, Hermanto M W. *RSC Adv.*, **2014**, 4: 22194–22204

[19] Cheng S, Li C, Guo L. *Vib. Spectrosc.*, **2014**, 70: 110~114

[20] Guo L F, Wiesmath A, Sprenger P, Garland M. *Anal. Chem*., **2005**, 77: 1655–1662

[21] Zhang H J, Garland M, Zeng Y Z, Wu P. *J. Am. Soc. Mass. Spectrom.*, **2003**, 14: 1295–1305

[22] Widjaja E, Li C Z, Chew W, Garland M. *Anal. Chem*., **2003**, 75: 4499–4507

[23] Shao L, Griffiths P R. *Anal. Chem*., **2008**, 80: 5219–5224

[24] Malinowski E R. *Factor Analysis in Chemistry, 3rd ed.* John Wiley and Sons: New York, **2002**

[25] Shannon C E. *E. Bell. Sys. Tech. J.*, **1948**, 27: 379–623

[26] Widjaja E, Garland M. *J. Comput. Chem.*, **2002**, 23: 911–919