Anal. Chem. **2007,** *79,* 2118−2124

Obtaining Qualitative Information on Trace Species in Continuous Open-Path Fourier Transform Spectroscopic Measurements Using Target Factor Analysis and Related Techniques

Limin Shao† and Peter R. Griffiths*

Department of Chemistry, University of Idaho, Moscow, Idaho 83844-2343

We report the application of target factor analysis (TFA) to the identification of trace analytes in open-path Fourier transform infrared (OP/FT-IR) spectra. Results showed that for components such as methane and ammonia, for which the path-integrated concentration was greater than ~100 ppm·m, TFA yielded results that closely match corresponding reference spectra. Furthermore, it was shown that the rotation of certain eigenvectors allowed the presence of trace analytes to be recognized when the SNR of these molecules in individual spectra was below 1.0, even though no prior knowledge that a particular molecule was present in the atmosphere being monitored was available. The presence of the analyte was confirmed by conventional TFA. The presence of ethanol and ozone was observed in OP/FT-IR spectra measured at certain locations near dairy and hog farms in this way. No band that could be assigned to either analyte could be observed in any of the spectra in the data set. The application of TFA to OP/FT-IR spectra had the advantage that no preprocessing, e.g., compensation of water line absorption or baseline correction, is needed.

Open-path Fourier transform infrared (OP/FT-IR) spectrometry is a fast, sensitive, noninvasive, and rugged technique for continuous atmospheric analysis.[1−3] However, OP/FT-IR spectra are inevitably degraded by vibration−rotation lines from atmospheric water vapor and carbon dioxide and by slow baseline variations. The sharp vibration−rotation water lines and slow baseline variations often dominate the spectrum and present a serious inference to the identification and quantification of trace components, no matter whether the analyte spectra are composed of broad or narrow bands.

Most qualitative methods in OP/FT-IR spectroscopy are based on visual inspection or, more rarely, spectral library searching.[4−8] In either case, the performance is largely determined by either the operator's knowledge of spectroscopy or the algorithm being able to recognize the features of a target compound in the measured spectrum with high confidence. In the presence of the interferences mentioned above, it is not surprising that ambiguous, or even false, results are found. In one of the more successful investigations of the automated identification of molecules in OP/FT-IR spectra, Yang and Griffiths[9,10] applied several types of artificial neural networks to the qualitative analysis of OP/FT-IR spectra. They showed that each of five alcohols with similar spectra could be successfully identified in noisy spectra without baseline correction. However, for a given molecule to be recognized, its bands had to be present in the spectrum with a signal-to-noise ratio (SNR) of at least 5 (where noise is measured as the peak-to-peak value in a spectral region where no atmospheric components absorb). For trace components, especially when the SNR is <5, confidence in the identification is low. It is probably true to say that the serious spectral interferences impair the potential of OP/FT-IR spectroscopy and that is one reason why the popularity of this technique is relatively low.

In this paper, we show that the presence of a given compound in a series of spectra may be recognized with high confidence through the application of a technique known as target factor analysis (TFA) or target transform factor analysis (TTFA). TFA is a self-modeling technique that rotates abstract eigenvectors obtained by principal component analysis (PCA) into physically significant vectors, such as spectra or concentration profiles.[11−14] TFA can be operated in an iterative mode, resulting in the iterative target transformation factor analysis (ITTFA). Applications of multicomponent analysis by TFA and ITTFA have been reported in such fields as kinetic processes and high-performance liquid chromatography (HPLC) with diode array detection (DAD). For

* Corresponding author. E-mail: pgriff@uidaho.edu.

† Present address: Department of Chemistry, University of Science and Technology of China, Hefei, Anhui, 230026 P.R. China.

(1) Newman, A. R. *Anal. Chem.* **1997**, *69*, 43A−47A.
(2) Walter, W. T. *Proc. SPIE* **2004**, *5270*, 144−150 *(Environmental Monitoring and Remediation III)*.
(3) 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.
(4) Luinge, H. J. *Vib. Spectrosc.* **1990**, *1*, 3−18.
(5) Tanabe, K.; Tamura, T.; Uesaka, H. *Appl. Spectrosc.* **1992**, *46*, 807−810.
(6) Jarvis, T. D.; Kalivas, J. H. *Anal. Chim. Acta* **1993**, *272*, 53−59.
(7) Mittermayr, C. R.; Drouen, A. C. J. H.; Otto, M.; Grasserbauer, M. *Anal. Chim. Acta* **1994**, *294*, 227−242.
(8) Luinge, H. J.; Leussink, E. D.; Visser, T. *Anal. Chim. Acta* **1997**, *345*, 173−184.
(9) Yang, H.; Griffiths, P. R. *Anal. Chem.* **1999**, *71*, 751−761.
(10) Yang, H.; Griffiths, P. R. *Anal. Chem.* **1999**, *71*, 3356−3364.
(11) Malinowski, E. R. *Factor Analysis in Chemistry*, 3rd ed.; John Wiley and Sons: New York, 2002.
(12) Maeder, M. *Anal. Chem.* **1987**, *59*, 527−530.
(13) Lorber, A. *Anal. Chem.* **1984**, *56*, 1004−1010.
(14) Lorber, A.; Kowalski, B. R. *Anal. Chem.* **1989**, *61*, 1169−1171.

10.1021/ac062042x CCC: $37.00
Published on Web 01/24/2007

example, McCue and Malinowski[15] successfully identified individual components from poorly resolved HPLC-DAD data. Gemperline[16] applied constraints of non-negativity and unimodality to the implementation of TFA of HPLC-DAD data and obtained elution profiles of overlapped components. van Zomeren et al.[17] found by augmenting data from high-performance liquid chromatography and micellar electrokinetic chromatography coupled with diode array detection, ITTFA yielded significantly better estimates of the chromatographic and electrophoretic profiles and spectra than those without augmentation. Tam and Chau[18] used PCA and TFA to study the reaction mechanisms and found PCA and TFA were very useful for investigating first-order consecutive reactions even when the spectra were quite noisy. Carvalho et al.[19] combined ITTFA with kinetic models to investigate the course of a reaction. They showed that problems with ITTFA associated with convergence could be overcome, leading the algorithm to converge in the right direction.

A few applications to infrared spectroscopic analysis have also been reported. Liang et al.[20] employed ITTFA to analyze the IR spectra collected during the formation of polyurethane foam and successfully obtained component spectra and concentration profiles despite a very large data set, seriously overlapping bands, and lack of a priori knowledge of the intermediates and products. Kuzmanovski et al.[21] used TFA to analyze IR spectra of human urinary calculi and found the ambiguity and uncertainty in interpretation was considerably reduced in comparison to manual operation.

In this paper, we report the application of TFA to qualitative OP/FT-IR analysis. Results showed that, for major components, TFA yielded results that closely match corresponding reference spectra. For minor components with low signal intensity, the strong interferences make their spectral features imperceptible. Nevertheless, the spectral information that is concealed in individual spectra could be uncovered by TFA and allowed confident qualification. In this investigation, we also observed information in eigenvectors when the SNR of these components in individual spectra was below 1.0, even though we had no prior knowledge that a particular molecule was present at the particular area being monitored. The application of TFA to OP/FT-IR spectra had the benefit that no preprocessing, e.g., compensation of water line absorption or baseline correction, was needed.

THEORY

Throughout this paper, boldface lower- and upper-case letters denote vectors and matrices, respectively. All vectors are column vectors, the transpose of which are row vectors, indicated with superscript t. The subscript is the matrix size.

Consider that m absorbance spectra were generated from a continuous OP/FT-IR monitoring session and that each spectrum has n points. By arranging those spectra in a row-wise manner, we obtain an m-by-n matrix, $\mathbf{D}$. When Beer's law is obeyed (at least approximately), there exists a bilinear model,

$$\mathbf{D}_{m\times n} = \mathbf{C}_{m\times p}(\mathbf{S}_{n\times p})^{\mathrm{t}} \tag{1}$$

where p is the number of compounds and $\mathbf{C}$ and $\mathbf{S}$ are the concentration and spectral matrices, respectively. Each column vector of $\mathbf{C}$ or $\mathbf{S}$ contains the concentrations over the measurement time, or the complete spectrum of a certain compound.

Performing PCA on $\mathbf{D}$ yields

$$\mathbf{D}_{m\times n} = \mathbf{U}_{m\times q}(\mathbf{V}_{n\times q})^{\mathrm{t}} + \mathbf{R}_{m\times n} \tag{2}$$

where q is the number of principal components and in theory equals p in eq 1. In this equation, $\mathbf{U}$ and $\mathbf{V}$ are the score and loading matrices (or principal component and eigenvector matrices), respectively. Matrix $\mathbf{R}$ is the residual and contains information on less significant components such as measurement error or noise. If $\mathbf{R}$ in eq 2 is negligible, the following equation can be derived

$$\mathbf{C}_{m\times p}(\mathbf{S}_{n\times p})^{\mathrm{t}} = \mathbf{U}_{m\times q}(\mathbf{V}_{n\times q})^{\mathrm{t}} \tag{3}$$

Equation 3 shows that results from PCA, $\mathbf{U}$ and $\mathbf{V}$, are correlated with the concentration and spectral information that is in matrices $\mathbf{C}$ and $\mathbf{S}$, respectively. In practice, $\mathbf{V}$ contains all the real spectral information in $\mathbf{S}$, but in a different form, which is why eigenvectors (column vectors in $\mathbf{V}$) are also called abstract spectra. Eigenvectors are mutually orthogonal, whereas vectors in $\mathbf{S}$ are linearly independent because they are real spectra of different compounds. These facts manifest that eigenvectors span the same space as the real spectra. Therefore, any spectral vector in $\mathbf{S}$, say $\mathbf{s}_{n\times 1}$, can be expressed as a unique linear combination of eigenvectors by multiplying $\mathbf{V}$ with a rotation vector, $\mathbf{r}$

$$\mathbf{s}_{n\times 1} = \mathbf{V}_{n\times q}\mathbf{r}_{p\times 1} \tag{4}$$

The rotation vector $\mathbf{r}$ can be obtained through the least-squares operation,

$$\mathbf{r}_{p\times 1} = (\mathbf{V}^{\mathrm{t}}_{n\times q}\mathbf{V}_{n\times q})^{-1}\mathbf{V}^{\mathrm{t}}\mathbf{s}_{n\times 1} \tag{5}$$

Equation 5 is the core of target factor analysis.[11] In TFA, the first step is to select a target, $\mathbf{s}_{\mathrm{target}}$, then calculate the rotation vector with eq 5, and finally predict the target, $\mathbf{s}_{\mathrm{predicted}}$, with eq 4. The presence of the target in the spectrum is confirmed or rejected by whether or not $\mathbf{s}_{\mathrm{predicted}}$ equals $\mathbf{s}_{\mathrm{target}}$ within the error allowance limit.

The above theory and procedure also hold for the concentration matrix $\mathbf{C}$ and score matrix $\mathbf{U}$, which means that the real concentration profile of the target compound can be obtained. In this case, TFA is a quantitative method, but only when some knowledge of the profile is available (to construct the target).[12]

It can be found that, in TFA, acquiring the rotation vector plays a key role, which is not a problem if the target is available. However, when that is not the case, an artificial target should be

(15) McCue, M.; Malinowski, E. R. *Appl. Spectrosc.* **1983**, *37*, 463−469.
(16) Gemperline, P. J. *Anal. Chem.* **1986**, *58*, 2656−2663.
(17) van Zomeren, P. V.; Metting, H. J.; Coenegracht, P. M. J.; de Jong, G. J. *J. Chromatogr., A* **2005**, *1096*, 165−176.
(18) Tam, K. Y.; Chau, F. T. *Chemom. Intell. Lab. Syst.* **1994**, *25*, 25−42.
(19) Carvalho, A. R.; Wattoom, J.; Zhu, L.; Brereton, R. G. *Analyst* **2006**, *131*, 90−97.
(20) Liang, X.; Andrews, J. E.; de Haseth, J. A. *Anal. Chem.* **1996**, *68*, 378−385.
(21) Kuzmanovski, I.; Trpkovska, M.; Soptrajanov, B.; Stefov, V. *Vib. Spectrosc.* **1999**, *19*, 249−253.

Table 1. Data Sets from Continuous OP/FT-IR Monitoring Sessions

data set	location	date	duration (min)	optical path length (m)	no. of spectra
1	dairy pen	March 16, 2005	798	486	700
2	open land[a]	January 26, 2005	1069	372	950
3	hog lagoon	February 2, 2005	380	138	332
4	hog lagoon	June 14, 2004	742	142	600
5	dairy lagoon	January 28, 2005	1357	306	1177
6	dairy compost	January 29, 2005	931	542	811
7	dairy lagoon	March 19, 2005	1374	258	1184

[a] The beam path was 230 m to the west of the pen of a dairy farm, with a westerly wind.

constructed, and parts of this work show our approach to tackling this problem.

EXPERIMENTAL SECTION

OP/FT-IR measurements were carried out in June and July 2004 and January, March, and June 2005, on and around a dairy and a hog farm in southern Idaho in a cooperative project for monitoring gaseous emissions with the Northwest Irrigation and Soil Research Laboratory (NWISL) of the United States Department of Agriculture (USDA). The OP/FT-IR spectrometer was manufactured by MDA Corp. (Atlanta, GA), and incorporated a Bomem Michelson 100 interferometer, a 31.5-cm telescope, a cube-corner array retroreflector, and a Sterling engine-cooled mercury cadmium telluride detector. Instrument control and data acquisition were performed with GRAMS 7.00 (Thermo Galactic, Salem, NH). The distance between the telescope and retroreflector was usually between 100 and 200 m, for a total path length of 200−400 m; however, occasionally longer path lengths were necessary. Every OP/FT-IR interferogram was measured by coadding 16 interferograms at a nominal resolution of 1 cm^{-1}. All spectra for the analysis were computed with a zero-filling factor of 2 (two data points per resolution element) and Norton-Beer "medium" apodization.

In this investigation, absorbance spectra were obtained by calculating the ratio of the long-path (200−400 m) single-beam spectrum to a short-path (~2 m) single-beam background spectrum. The instrument response profile is removed by this procedure, but absorption by the analyte and atmospheric H_2O and CO_2 is not. Since strong interference by H_2O and CO_2 is always seen in the spectra, the analysis can only be performed in regions of the spectrum where these species absorb weakly (the so-called atmospheric windows.) In our work, two spectral windows, from 1250 to 750 cm^{-1} and from 3200 to 2850 cm^{-1}, were used. Spectra in a given continuous monitoring session were measured at intervals of ~70 s. Successive absorbance spectra in a particular set were assembled in the order in which they were measured as rows of the matrix for TFA.

Reference spectra of ethanol and ozone are from a commercial library published by Infrared Analysis Inc. Reference spectra of NH_3 and CH_4 were provided by courtesy of Pamela M. Chu at the National Institute of Standards and Technology (NIST).

Quantitative analysis was carried out by using partial least-squares (PLS) regression in this investigation. Details are given in ref 22. All manipulation of spectra and data processing was done using MATLAB 7.0.1 (The MathWorks Inc., Natick MA) on a Windows 2000 platform.

RESULTS AND DISCUSSION

Seven data sets were prepared from OP/FT-IR measurements that had been carried out at several locations with various path lengths and under various conditions of temperature and relative humidity. These data sets reflect a range of atmospheric compositions; for example, the air over the dairy pen is rich in ammonia and methane, whereas lagoon or compost air could be quite complex because of aerobic or anaerobic fermentation. Other information on these data sets is given in Table 1. Both visual inspection and calculation of the correlation coefficient, R^2, were employed to confirm the presence or absence of a component. A value of $R^2 = 0.85$ was adopted as the threshold value of the correlation coefficient.

TFA for Specific Molecules. On and around animal farms, the concentration of ammonia and methane is usually higher than the global average and the SNR of strong lines in the spectrum is usually greater than 10. The reference spectra of NH_3 and CH_4 were used as the targets in our first attempt to perform TFA on the OP/FT-IR data sets. The eigenvectors for data set 1 calculated by the procedure described above are shown in Figure 1. It can be seen that the first eigenvector (EV1) contains largely information due to a baseline offset, while the features in EV2 and EV3 are readily assigned to NH_3 and H_2O, respectively. Spectral features assignable to both NH_3 and H_2O can be seen in EV4, while EV5 appears to show the effect of a nonlinear baseline drift.

Figure 2 shows the results from the TFA of data sets 1 (Figure 2a) and 2 (Figure 2b) with NH_3 being the target. For data set 1, the concentration of NH_3 was quite high while for data set 2 it was much lower. The particular spectrum shown as the lowest trace in Figure 1a was chosen from data set 1 because the path-integrated concentration of NH_3 was the average value for this data set, ~217 ppm·m. The predicted spectrum exhibits high similarity to the reference spectrum, which is not surprising in light of the high path-integrated concentration of NH_3 in every spectrum in this set. Such direct evidence is not available for data set 2, since the average path-integrated concentration of NH_3 was ~1 ppm·m. Even for the spectrum in this data set that exhibits

(22) Shao, L.; Pollard, M. J.; Griffiths, P. R.; Westermann, D. T.; Bjorneberg, D. L. Rejection Criteria for Open-Path Fourier Transform Infrared Spectrometry during Continuous Atmospheric Monitoring. *Vibrational Spectrosc.* **2007**, *43*, 78−85.

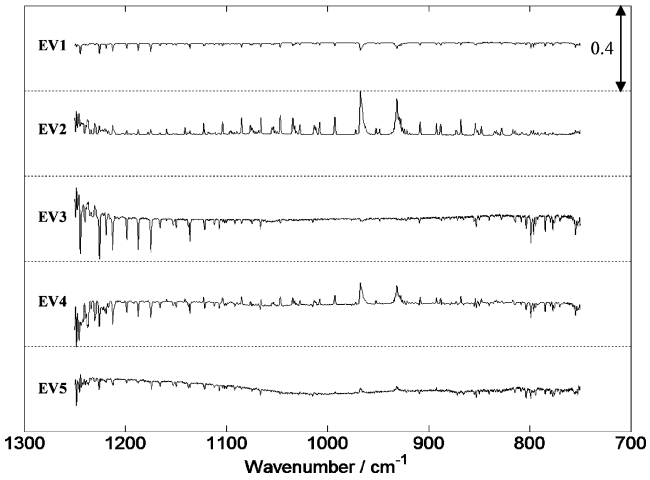

Figure 1. Five eigenvectors calculated for data set 1. Eigenvectors were displaced for clarity with the double-headed arrow indicating the scale. All eigenvectors in this figure (as well as Figures 4 and 8) are plotted on the same scale (full scale for each panel, 0.4). The eigenvectors are unitary, i.e., square root of the sum of squares is 1. This method of normalization makes it possible to plot all eigenvectors on the same scale for comparison.

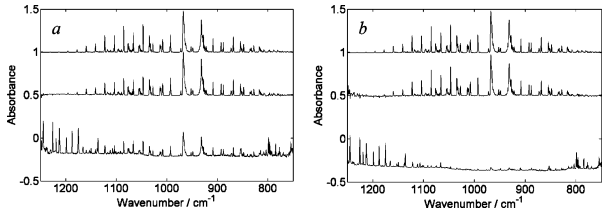

Figure 2. Target factor analyses of two data sets, with ammonia being the target. (a) Data set 1, for which the concentration of NH_3 is high; (b) data set 2, for which the concentration of NH_3 is low; 4 and 5 factors were used for (a) and (b), respectively. In each panel, the top spectrum is the reference spectrum of NH_3 at 1-cm^{-1} resolution, the middle spectrum is the result of TFA of the appropriate data set, and the lowest spectrum is a typical measured OP/FT-IR spectrum from each data set. The reference and predicted spectra were displaced for clarity. EV2 in Figure 1 appears to be more noisy than the middle spectrum in (a) because, when NH_3 is the target, the result of TFA is the combination of EV1 through EV4 for data set 1; thus other information, including water lines and noise, is well compensated and can hardly be seen in the predicted spectrum of NH_3.

the highest path-integrated concentration of NH_3 (16 ppm·m, which translates to an average concentration along the path of 40 ppbv), which is the spectrum shown in Figure 1b, spectral features of NH_3 are barely recognizable. Nevertheless, TFA yielded satisfactory results for both data sets. The measured spectra in Figure 1 exhibit significant baseline drifts and relatively high absorption of water vapor between 1250 and 1150 cm^{-1}. However, TFA minimizes these interferences. The correlation coefficients were calculated to be 0.98 and 0.98 for data sets 1 and 2, respectively, both of which are higher than the empirical threshold. Therefore, the presence of NH_3 in both data sets was further confirmed.

Figure 3 shows the results from TFA of data sets 1 and 2 with CH_4 as the target. From PLS regression, the average path-integrated concentrations of CH_4 of the two data sets were 1680 (data set 1) and 650 ppm·m (data set 2). The two OP/FT-IR spectra in Figure 3 (shown as bottom traces) are representative of the two data sets with the same path-integrated concentrations as the average for the particular set. As shown in the figure, the result obtained by TFA was almost of the quality of the refrence spectrum, whether the spectral features of CH_4 was obvious (Figure 3a) or less so (Figure 3b), although the relative intensity of the predicted lines in the R branch of CH_4 is less than the lines in the P or R branches. Many of the atmospheric water lines in the region between 3200 and 3000 cm^{-1} are quite intense and do not follow Beer's law under the conditions of the measurements, which possibly explains the low relative intensity of the R branch lines in the calculated spectrum of CH_4. Nevertheless, the correlation coefficients also provided confirmation, with the values of R^2 for data sets 1 and 2 being 0.91 and 0.95, respectively.

TFA without a Target. In the previous section, TFA was shown to perform well when the spectrum of the target molecule was available. However, in many cases, the identity of every molecule present in the air at a concentration of greater than 1 ppbv is unknown. In this case, the researcher does not have a target and, hence, cannot perform conventional TFA. Similarly, exhaustive library search, such as attempts to select the spectrum of every molecule that could possibly be present as a target, is also impractical, since qualification using TFA is not completely automatic. Data set 3 represents a case for which an unexpected molecule contributed to the spectrum.

PCA was first carried out for this data set. The first 6 eigenvectors from PCA of this data set are shown in Figure 4. For the first 5 eigenvectors, spectral features of water vapor can be seen in the region from 1250 to 1100 cm^{-1}. Spectral features of NH_3 are seen weakly in the third eigenvector and strongly in the fifth and sixth eigenvectors; indeed, with NH_3 as the target, TFA yielded as convincing a result on this data set as those of data sets 1 and 2. Careful inspection of the fourth eigenvector revealed that the segment from 1100 to 1000 cm^{-1} appeared to contain a relatively broad unexplained spectral feature that could not be assigned to a small molecule such as H_2O, NH_3, CH_4, N_2O, or CO_2 that would commonly be found in the atmosphere around a hog farm as these molecules would have resolvable rotational fine structure.

Confirmation of the identity of this molecule was impossible by conventional TFA because there was no target. In this situation, instead of finding a putative rotation vector that would predict the spectrum of the unidentified molecule, we tried to obtain another type of rotation vector that could retain the spectral features of the unidentified molecule between 1100 and 1000 cm^{-1}, but would eliminate most interfering spectral lines of water vapor and ammonia. This approach is illustrated in the bottom part of Figure 5, while the top part of this figure represents a conventional TFA procedure, which is inapplicable in this case.

In Figure 5, $\mathbf{V}$ represents the eigenvector matrix, $\mathbf{t}$ is the spectrum of the unidentified molecule, and $\mathbf{r}$ is the corresponding rotation vector. Matrix $\hat{\mathbf{V}}$ comprises two parts from $\mathbf{V}$, namely, the data from 1250 to 1100 cm^{-1} (where the strongest spectral features are due to atmospheric water vapor) and from 970 to 920 cm^{-1} (which mainly contains ammonia lines.) Since we wished to eliminate the spectral information due to water and ammonia, a zero target, $\mathbf{t}_0$, was constructed (values are 1×10^{-6} in practice), i.e., the $\hat{t}$ data, result of TFA, in the regions from 1250 to 1100 and 970 to 920 cm^{-1} are made as close as possible to zero in a

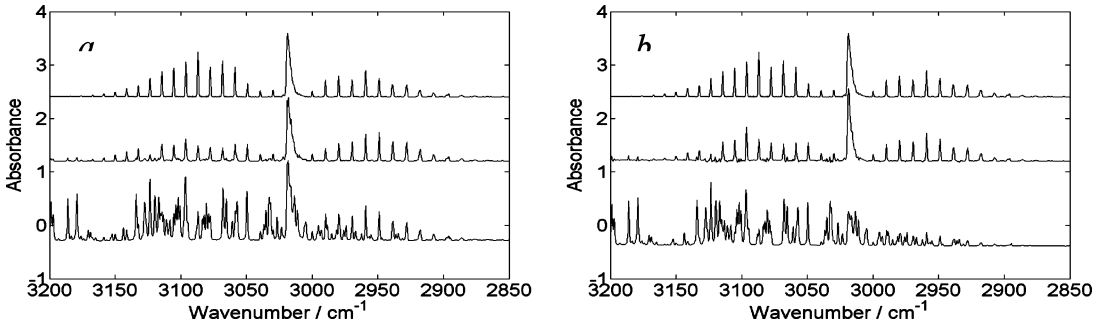

Figure 3. Target factor analyses of data sets 1 and 2 with methane being the target. Traces in (a) and (b) from top to the bottom are reference spectrum of methane, results obtained by TFA, and measured spectra for comparison. Both the reference and predicted spectra were displaced for clarity.

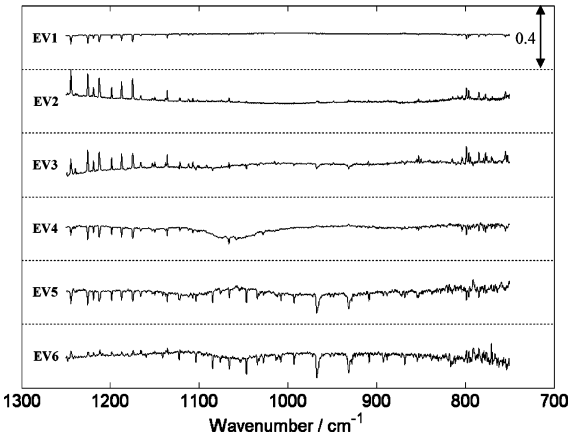

Figure 4. First 6 eigenvectors from PCA of data set 3. Eigenvectors were displaced for clarity with the double-headed arrow indicating the scale.

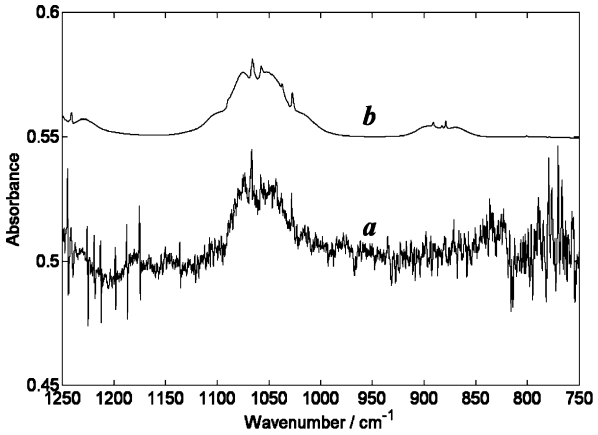

Figure 6. (a) Result (×500 000) of rotating the first 6 eigenvectors in Figure 4 using a rotation vector obtained with eq 6. (b) The reference spectrum of ethanol. Trace b was displaced for clarity.

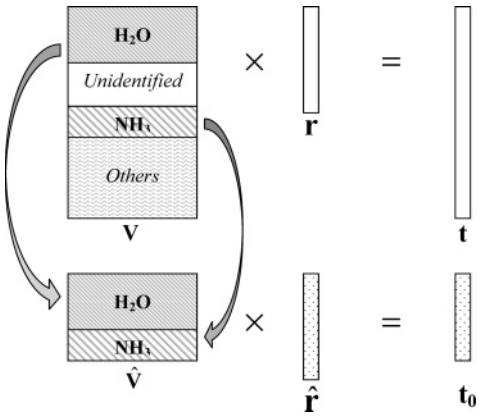

Figure 5. Approach to find a rotation vector that retains the information of the unidentified molecule and eliminates that of water vapor and ammonia.

least-squares manner, so that the absorption lines due to H_2O and NH_3 are eliminated to the greatest possible extent. During this process, the number of elements in $\mathbf{t}_0$ is set equal to number of row vectors in $\mathbf{V}$. In a way similar to eq 5, the corresponding rotation vector, $\hat{\mathbf{r}}$, was obtained as

$$\hat{\mathbf{r}} = (\hat{\mathbf{V}}^t\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}^t\mathbf{t}_0 \qquad (6)$$

By replacing $\mathbf{r}$ in the conventional TFA process in Figure 5 with $\hat{\mathbf{r}}$, a predicted spectrum, $\hat{\mathbf{t}}$, of the unidentified compound is obtained.

After applying the approach to data set 3, a predicted spectrum was obtained, shown as the lower trace in Figure 6. Even though the spectrum is noisy, the previously unknown molecule was readily identified as ethanol, the reference spectrum of which is shown as the upper trace in Figure 6. With this information, a conventional TFA was performed with the reference spectrum of ethanol being the target using 6 factors, and the existence of ethanol was very convincing, as shown in Figure 7. The spacing of the rotational lines in the spectrum of ethanol is less than 1 cm^{-1}, so that the lines are not resolved when the spectrum is measured at a resolution of 1 cm^{-1}. Without this confirmatory procedure, the contour of this weak, broad band could be easily misassigned to a variation in the baseline of the spectrum and the presence of ethanol missed. It may also be noted that the relatively strong water vapor line at ~1066 cm^{-1} absorbs at approximately the same wavenumber as the Q branch of ethanol, again making it very difficult to identify ethanol at trace levels directly from the set of OP/FT-IR spectra. In fact, for data set 3, no visual evidence of ethanol was found directly from any of the 332 spectra.

Using a similar procedure, ethanol was also found to be present in the spectra of data sets 4 and 5, for which the measurements were taken over a hog lagoon and a dairy lagoon. Ethanol was not the only molecule present at trace levels that could be identified by the procedure described above. Figure 8 shows the first 10 eigenvectors from the PCA of data set 6. As before, the spectral features of H_2O and NH_3 are seen in the first 4 eigenvectors, and the presence of ammonia is strongly indicated

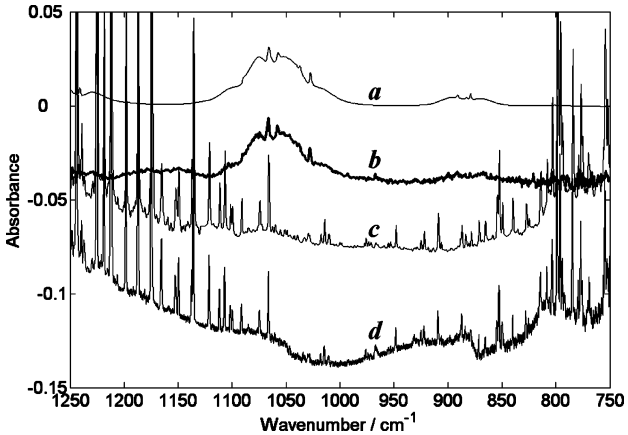

Figure 7. (a) Reference spectrum of ethanol; (b) result of TFA; (c) reference spectrum of water vapor;[1] and (d) one of the spectra measured during continuous monitoring. Traces b and c were displaced for clarity. (The reference spectrum shown in (c) was obtained by taking the reference spectrum of H_2O from the collection published by Infrared Analysis, Inc. The path-integrated concentration of water vapor for this spectrum is 1000 ppm·m. The region between 1250 and 750 cm^{-1} was multiplied by 1000 so that the strong lines between 1800 and 1400 cm^{-1} are at the appropriate intensity. This scale expansion accounts for the poor baseline in this spectrum.)

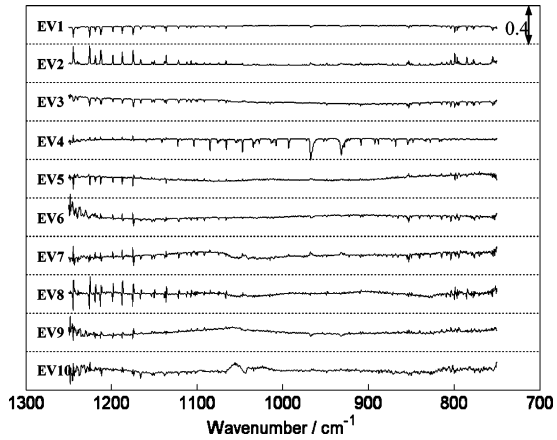

Figure 8. First 10 eigenvectors from PCA of the OP/FT-IR data measured over a dairy farm compost heap. Eigenvectors were displaced for clarity, with the double-headed arrow indicating the scale.

in EV4. However, EV10 appears to show the presence of a spectral feature of an unidentified molecule. By employing the same approach as in the case of data set 3, we obtained a spectrum for the unidentified molecule, shown as Figure 10, allowing it to be identified as ozone. A conventional TFA was then carried out with reference spectrum of ozone as the target; the results of this procedure are shown in Figure 10, which clearly confirms the presence of ozone, despite the dominant H_2O and NH_3 lines shown in Figure 10c.

The reference spectrum, Figure 10a, shows that, at a resolution of 1 cm^{-1}, the rotational lines in the spectrum of ozone are not resolved. In OP/FT-IR spectra, this weak, broad band contour is completely lost in the jungle of sharp absorption lines of water and ammonia and the variations in the baseline, which are rarely monotonic. Thus, as for the previous case of ethanol, no spectral features of ozone can be directly seen in any of the spectra in data sets 6 and 7.

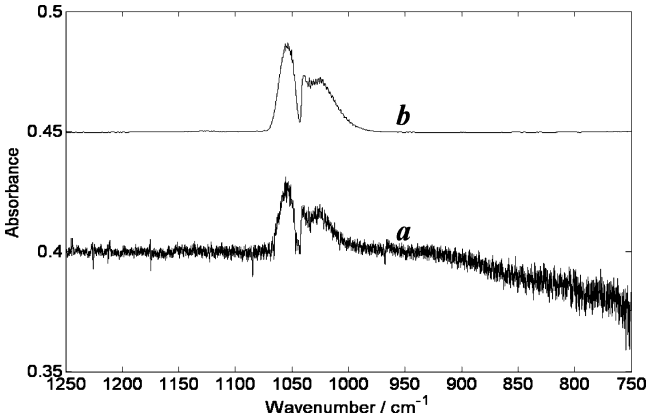

Figure 9. (a) Result (×400 000) after rotating the first 10 eigenvectors in Figure 8 by using a rotation vector obtained as described in eq 6. (b) Reference spectrum of ozone. Trace b was displaced for clarity.

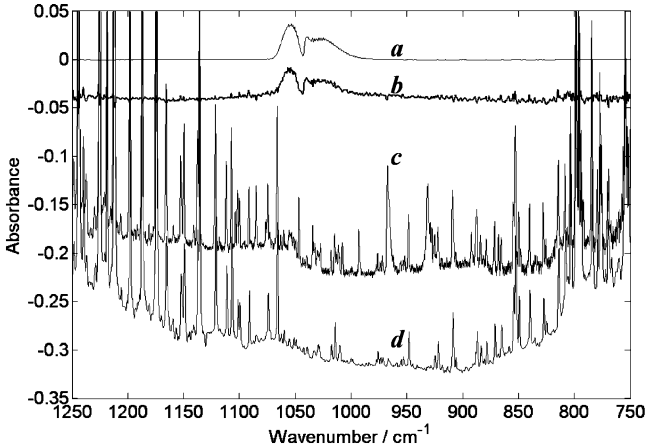

Figure 10. (a) Reference spectrum of ozone; (b) result of TFA; (c) typical spectrum measuried during continuous monitoring; (d) Reference spectrum of water vapor. Traces b and d were displaced for clarity.

Table 2. Results of TFA Using Correlation Coefficient and Visual Inspection for the Qualitative Determination*

data set	ethanol[a]	ozone[a]	factors in TFA
3	0.95 (+)	0.71 (−)	6
4	0.87 (+)	0.66 (−)	6
5	0.95 (+)	0.83 (−)	8
6	0.84 (−)	0.91 (+)	10
7	0.81 (−)	0.79 (+)	9

[a] A plus or minus symbol in parentheses indicates confirmation of the presence or absence, respectively, of detection by visual inspection of the appropriately rotated eigenvectors.

The results from analyzing data sets 3−7 are summarized in Table 2. It was found that setting the threshold value of the correlation coefficient at 0.85 was as effective as visual inspection for confirming the presence of a trace component, except in the case of the detection of ozone from data set 7. By comparing the predicted spectrum with the reference spectrum, the false negative of data set 7 was found to be primarily caused by the effect of noise, even though the spectral features of ozone are readily visible in the predicted spectrum. A detection criterion that incorporated

some spectral features might ultimately prove to be more effective. Further work on this topic is in progress. In analyzing data sets 3−7, ethanol was exclusively determined in the measurements over lagoons and is probably derived from the anaerobic fermentation of animal products in the pond. The presence of ozone is manifested in spectra measured respectively over a compost heap when the average temperature was 2.9 °C (σ = 3.1 °C) and the lagoon of a dairy farm when the average temperature was less than 6.2 °C (σ = 3.7 °C).

One of the goals of this project was to investigate the feasibility of detecting the nitrogen oxides, N_2O, NO, and NO_2. Nitrous oxide is readily detectable in OP/FT-IR spectra, and nitric oxide is usually present at concentrations below the limit of detection. Nitrogen dioxide is very difficult to detect by OP/FT-IR spectroscopy because all of its strong bands fall in spectral regions where water vapor absorbs so strongly that corresponding absorbance data are invalid. However, ozone is formed in the troposphere by the photolysis of nitrogen dioxide.[23]

$$NO_2 + h\nu(\lambda < 400\ \text{nm}) \rightarrow NO + O(^3P)$$

$$O(^3P) + O_2 \rightarrow O_3$$

Thus, it is possible that the presence of NO_2 can be detected indirectly by OP/FT-IR through the presence of ozone.

CONCLUSION

TFA proved to be a useful tool to identify major and minor components in OP/FT-IR spectroscopy. The presence of a target is confirmed or rejected by comparing the output and the reference of the target. TFA exhibits strong resistance to interferences. For OP/FT-IR spectra, water vapor lines are usually intense and baseline drifts frequently occur, especially under bad weather conditions. Nonetheless, neither problem prevented convincing evidence for the presence of trace compounds being obtained through the application of TFA. It should also be noted that, in this investigation, except rejecting invalid measurements with the procedure described in ref 22, no preprocessing, such as baseline correction or noise reduction, was carried out prior to TFA. TFA was used purely as a qualitative technique in this investigation, and the intensity of the eigenvectors does not relate to the concentration of the target. Thus, the scales of the traces in Figures 6b and 9b are only for the sake of comparison and have no relationship to the actual concentrations of ethanol of ozone.

TFA is a powerful technique for analyzing complex multicomponent systems. Application of this technique allows the presence of individual target components to be studied without requiring any knowledge of the concentration of the other components. Moreover, even when a priori knowledge of the presence of a particular analyte is unavailable, useful information can still be obtained, as was the case for ethanol and ozone in our investigation.

The results of this study indicated that TFA can yield results at a high confidence level for major, minor, even trace components despite significant difference in signal strengths. In fact, for the analyses of ethanol and ozone described above, the spectral features were completely obscured by absorption of atmospheric water vapor and carbon dioxide, noise, the presence of major and minor components, and baseline variations. By combining all the spectral data in a matrix, rather than examining individual spectra, TFA allows the limit of detection of OP/FT-IR spectroscopy (and, by extension, other types of spectroscopy) to be reduced.

Finally, results in this investigation demonstrated that, when a target exists, the predicted spectrum from TFA is largely free of the interferences present in the raw spectra. Therefore, spectral searching programs could allow the feasibility of the automatic identification of trace components to be achieved by TFA.

ACKNOWLEDGMENT

This work was funded under Cooperative Agreement 58-5368-3-269 with the United States Department of Agriculture, Agricultural Research Service, Northwest Irrigation and Soils Research Laboratory, Kimberly, ID. The cooperation of the owners of the farms at which these measurements were taken, and the suggestion by Matthew Pollard that ozone is formed by photolysis of nitrogen dioxide, is gratefully appreciated.

Received for review November 1, 2006. Accepted December 19, 2006.

AC062042X

(23) Finlayson-Pitts, B. J.; Pitts, J. N., Jr. *Chemistry of the Upper and Lower Atmosphere: Theory, Experiments and Applications;* Academic Press: New York, 2000; p 180.