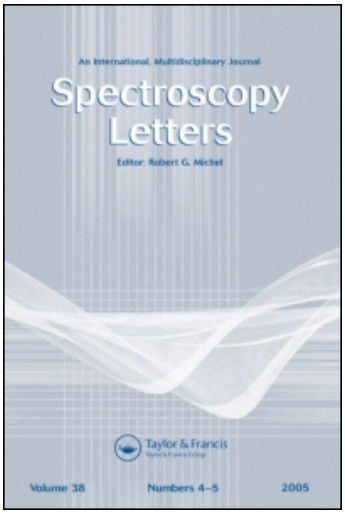

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Simultaneous Determination of Multicomponents in Air Toxic Organic Compounds with Partial Least-Squares Method Using FTIR Spectroscopy

Binghe Gu[a]; Junde Wang[a]
[a] Laboratory of Advanced Spectroscopy Nanjing, University of Science & Technology, Nanjing, P. R. China

To cite this Article Gu, Binghe and Wang, Junde(1998) 'Simultaneous Determination of Multicomponents in Air Toxic Organic Compounds with Partial Least-Squares Method Using FTIR Spectroscopy', Spectroscopy Letters, 31: 5, 1053 — 1063
To link to this Article: DOI: 10.1080/00387019808003283
URL: http://dx.doi.org/10.1080/00387019808003283

SPECTROSCOPY LETTERS, 31(5), 1053-1063 (1998)

SIMULTANEOUS DETERMINATION OF MULTICOMPONENTS IN AIR TOXIC ORGANIC COMPOUNDS WITH PARTIAL LEAST-SQUARES METHOD USING FTIR SPECTROSCOPY*

Keywords: Multicomponents Analysis; Organic Compounds Analysis; FTIR; Partial Least-Squares Method

Binghe Gu and **Junde Wang****
Laboratory of Advanced Spectroscopy
Nanjing University of Science & Technology
Nanjing 210014, P.R.China

Abstract

The multicomponent analysis with Partial Least-Squares Method (PLS) using FTIR spectroscopy has been studied in this paper. In order to ensure the reliability of the prediction results, a criterion *SO* and *SA* is created in the PLS algorithm to conclude the similarity between the prediction samples and the calibration ones. An example for the simultaneous determination of Ethylbenzene, Styrene, o-Xylene, m-Xylene and p-Xylene is supplied. The experimental designs of the

* The project was supported by National Nutural Science Foundation of China and Ph.D Foundation of National Education Committee.

** Author to whom correspondence should be sent.

both the calibration and prediction samples are discussed as well as the significant factor number. With proper multivariate calibration conditions, the average mean relative error (MRE %) is 0.25 % and the average relative standard deviations (RSD%) for Ethylbenzene, Styrene, o-Xylene, m-Xylene and p-Xylene are 0.19, 0.15, 0.06, 0.44 and 0.13 %, respectively.

I. INTRODUCTION

Multivariate calibration[1] can be regarded as a general tool to solve selectivity problems in multichannel instruments, e.g., overlapping spectra in mixture samples. The partial least squares method (PLS)[2,3] is usually optimal or close to optimal compared to other multivariate calibration methods. The use of PLS in FTIR spectroscopy[4-10]offers a combination of little or no sample preparation with high speed quantitative analysis, which has many appliance in the last 20 years. However, most of the literatures cited above have an obvious characteristic, i.e., the number of the components in the prediction samples is not more than four, and none of them check the prediction samples to ensure the similarity between the prediction samples and the calibration samples. Our efforts in this paper are trying to predict five air toxic organic compounds whose spectra are strongly overlapped in the mid-infrared region bands. We use a criterion *SO* and *SA* to conclude whether the prediction samples are similar to the calibration samples before the prediction step. The program of PLS is written in Q-basic 4.5, the data sets of the concentrations and absorbances are mean-centered but not scaled. Through designing the proper calibration samples and selecting the measured wavenumber bands, we got good results for the simultaneous determination of the five-components in the ten "unknown" prediction samples.

II. THEORY

The PLS method includes three parts, i.e., the calibration step, the diagnostic step and the prediction step. In the calibration step, the absorbance matrix X_{nxm} and the concentration matrix Y_{nxp} are decomposed into their scores and loadings according to iterative NIPALS algorithm:

$$X_{nxm} = T_{nxs} B_{sxm} + E_{nxm}$$

$$Y_{nxp} = U_{nxs} V_{sxp} + F_{nxp}$$

where T_{nxs} and U_{nxs} represent the scores matrix of X_{nxm} and Y_{nxp}, respectively. B_{sxm} and V_{sxp} are the loading matrix of X_{nxm} and Y_{nxp}, respectively. The s is the significant factor number. The E_{nxm} and F_{nxp} represent the unique variation in X_{nxm} and Y_{nxp} that is not explained by the s factor solution. Then, the scores U_{nxs} from Y_{nxp} are used as the regressions for the scores T_{nxs} from X_{nxm}:

$$U_{nxs} = T_{nxs} D_{sxs}$$

where D_{sxs} is the regression coefficient matrix.

In the diagnostic step, we calculate the fit residual e for a new sample x:

$$e = x - x\, B'_{mxs}\, B_{sxm}$$

x is the vector of new prediction sample, and B'_{mxs} is the transposed matrix of the matrix B_{sxm}. Therefore, we get the value of SO and SA:

$$SO = \|e\| / (p - s)$$

$$SA = S_s^2 (E_{n \times m}) . f$$

where $\|e\|$ is the norm of the vector e, and $S_s^2 (E_{n \times m})$ is the residual standard deviation of the matrix X_{nxm}. f is the possibility value when the free dimensions are $(p\text{-}s)$ and $(p\text{-}s)(n\text{-}s\text{-}1)$, and the confidence value is 0.90. If the new prediction sample is similar to the calibration samples, then SO is less than SA, vice versa.

In the prediction step, the absorbance matrix $X_{n'xm}$ is decomposed using the calibration model and the concentration matrix $Y_{n'xp}$ is formed:

$$X_{n'xm} = T_{n'xs} B_{sxm}$$
$$U_{n'xs} = T_{n'xs} D_{sxs}$$
$$Y_{n'xp} = U_{n'xs} V_{sxp}$$

where $X_{n'xm}$ is the prediction samples matrix and $Y_{n'xp}$ is the solution desired. Please refer to the literatures[2,3] for the specific algorithm.

III. EXPERIMENTAL

A Nicolet 170SX FTIR Spectrometer equipped with a globar light source, TGS detector with KBr window and a KBr beamsplitter was used. The data were collected by using 32 co-added interferograms in the spectral region from 4000 to 400 cm^{-1} at 2 cm^{-1} spectral resolution. The spectral resolution of 2 cm^{-1} can both provide the required information and not require the use of excessive computative time [11]. The data system was a Nicolet 1280 data station equipped with 4096 byte of RAM and a 24 Mbyte winchester disk driver.

IV. RESULTS AND DISCUSSION

4.1 The property of the five-components in the mixture

The air toxic organic compounds chosen in this study are Ethylbenzene, Styrene, o-Xylene, m-Xylene and p-Xylene. They have two distinctive characteristics: 1) The compounds are similar to one another with nearly indistinguishable spectral feature; 2) The five compounds have close boiling points, which nearly rules out the possibility to separate each component to measure their concentrations. The characteristic absorption bands and the boiling points of the five organic compounds are listed in Table 1. The FTIR spectra of these components at the concentrations of 100 ppm and their mixture where the concentrations of each component are 100 ppm can be seen from Fig 1. From Fig 1, we can see the situations of strong overlap among the five organic compounds.

Table 1. The characteristic absorption bands and boiling point of the five organic compounds

	Ethylbenzene	Styrene	o-Xylene	m-Xylene	p-Xylene
bp[a] (°C)	136	145-146	142	138-139	138
spectral bands[b]	3160-2850	3160-2850	3168-2800	3158-2825	3173-2839
(cm^{-1})	850-660	1100-660	820-660	820-650	852-650

Note : a) boiling point

b) the characteristic absorption bands

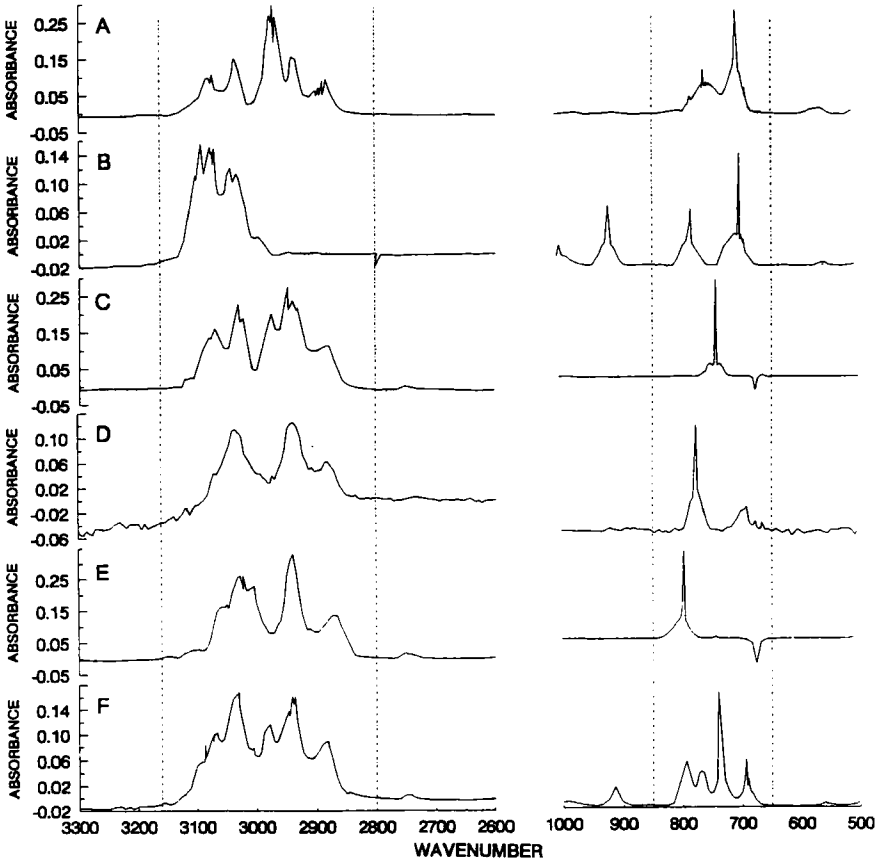

Fig 1. The spectra of the five organic compounds and their mixture.

Note : A. Ethylbenzene; B. Styrene; C. o-Xylene;
D. m-Xylene; E. p-Xylene; F. Their mixture.

4.2 The experimental design of the calibration and prediction samples

The composites of each calibration samples are designed according to the $L_{16}(4^5)$ orthogonal experimental design in that the number of the components in the mixture is five and all the concentrations of each component in the mixture vary from 5 to 100 ppm. The four levels are 5, 20, 50 and 100 ppm, respectively. That is to say, it is a five factor four level experimental design. As for the prediction samples, the concentrations of each component are chosen a little arbitrarily. We mainly consider two aspects: 1) the concentrations of each component in the prediction samples are varied from 5 to 100 ppm; 2) the composites of each prediction sample are determined to make the spectra of components interfere one another at their characteristic absorption bands. The concentrations of each component for each prediction sample are listed in Table 2.

4.3 The choice of the absorption bands for the measured wavenumbers

Since PLS is a full-spectrum calibration method, the spectral regions in 3160-2800 and 850-655 cm^{-1}, or the spectral region only in 850-655 cm^{-1} are used for the multivariate calibration. The spectral limits are selected in order to cut off most of the contributions from water vapor(4000-3170 and 2140-1230 cm^{-1}) and carbon dioxide (2400-2230 and 735-613 cm^{-1}). The interval of the measured wavenumbers is 15 cm^{-1}.

4.4 The significant factor number

Generally, the significant factor number is equal to or larger than the number of the components in the calibration samples. It is also influenced by the spectral and concentration noises, random baseline shift and chemical interaction between the components, etc. The significant factor number can be determined by many methods, such as

Table 2. The composite of the prediction samples ppm

No.	Ethylbenzene	Styrene	o-Xylene	m-Xylene	p-Xylene
1	8.0	15.0	70.0	15.0	15.0
2	30.0	40.0	60.0	70.0	20.0
3	70.0	8.0	15.0	30.0	40.0
4	40.0	35.0	45.0	35.0	50.0
5	15.0	10.0	20.0	18.0	16.0
6	35.0	30.0	28.0	25.0	26.0
7	8.0	9.0	7.0	6.0	5.0
8	60.0	65.0	64.0	63.0	58.0
9	75.0	70.0	73.0	74.0	80.0
10	60.0	50.0	40.0	30.0	20.0

validation sample determination, cross-validation[12] and manual designation. In this paper the significant factor number varies from 5 to 7 when the validation samples determination method is adapted.

4.5 The prediction results obtained at the spectral band of 3160-2800 and 850-655 cm^{-1}

As the above parameters are chosen, we simulated 10 groups of "unknown" samples to check the performance of the PLS in FTIR spectroscopy. The prediction results are listed in Table 3. From Table 3, the mean relative error (MRE %) for each prediction sample is little, averagely 0.25 %. The mean relative standard deviations (RSD %) for Ethylbenzene, Styrene, o-Xylene, m-Xylene and p-Xylene in all the prediction samples are 0.19, 0.15, 0.06, 0.44 and 0.13 % , respectively. Their average relative standard deviation is 0.19 %. The values of *SO* are less than those of *SA* for all the prediction samples, and it ensures the good prediction results and the reliability of the results.

Table 3. The determination results for the prediction samples at the spetral bands of 3160-2800 and 850-655 cm^{-1}

	Ethylbenzene			Styrene			o-Xylene			m-Xylene			p-Xylene			*SO*	*SA*	
No.	$C_{real.}$	$C_{meas.}$	RE%	$C_{real.}$	$C_{meas.}$	RE%	$C_{real.}$	$C_{meas.}$	RE%	$C_{real.}$	$C_{meas.}$	RE%	$C_{real.}$	$C_{meas.}$	RE%	($x10^7$)	($x10^7$)	MRE%
1	8.00	7.97	-0.39	15.00	15.01	0.09	70.00	69.98	-0.04	15.00	15.12	0.79	15.00	14.96	-0.27	9.0	296	0.32
2	30.00	30.12	0.41	40.00	39.99	-0.02	60.00	59.97	-0.05	70.00	69.99	-0.01	20.00	19.99	-0.01	14.2	296	0.10
3	70.00	69.98	-0.03	8.00	7.99	-0.16	15.00	14.99	-0.09	30.00	30.10	0.34	40.00	39.97	-0.08	13.6	296	0.14
4	40.00	40.09	0.22	35.00	34.98	-0.06	45.00	44.99	-0.03	35.00	35.01	0.01	50.00	49.97	-0.06	13.8	296	0.08
5	15.00	14.90	-0.69	10.00	10.03	0.33	20.00	20.03	0.16	18.00	17.98	-0.13	16.00	16.05	0.28	15.2	296	0.32
6	35.00	34.97	-0.09	30.00	30.04	0.13	28.00	28.02	0.06	25.00	25.12	0.48	26.00	26.03	0.10	15.1	296	0.17
7	8.00	8.01	0.05	9.00	9.08	0.89	7.00	7.04	0.56	6.00	5.89	-1.83	5.00	5.04	0.80	16.3	296	0.83
8	60.00	60.11	0.19	65.00	65.01	0.01	64.00	64.02	0.03	63.00	62.77	-0.36	58.00	58.07	0.11	14.4	296	0.14
9	75.00	75.07	0.09	70.00	69.89	-0.16	73.00	72.98	-0.03	74.00	74.32	0.43	80.00	79.97	-0.03	13.6	296	0.15
10	60.00	59.92	-0.14	50.00	50.04	0.07	40.00	40.03	0.06	30.00	29.76	-0.81	20.00	20.07	0.34	12.8	296	0.28
RSD%		0.19			0.15			0.06			0.44			0.13				
Average								0.19										0.25

Table 4. The determination results for the prediction samples at the spetral band of 850-655 cm^{-1}

	Ethylbenzene			Styrene			o-Xylene			m-Xylene			p-Xylene			*SO*	*SA*	
No.	$C_{real.}$	$C_{meas.}$	RE%	$C_{real.}$	$C_{meas.}$	RE%	$C_{real.}$	$C_{meas.}$	RE%	$C_{real.}$	$C_{meas.}$	RE%	$C_{real.}$	$C_{meas.}$	RE%	($x10^7$)	($x10^7$)	MRE%
1	8.00	8.40	4.97	15.00	14.81	-1.27	70.00	69.94	-0.09	15.00	15.29	1.90	15.00	15.06	0.42	105	202	1.73
2	30.00	30.19	0.62	40.00	39.84	-0.41	60.00	59.94	-0.10	70.00	70.06	0.09	20.00	20.00	0.00	154	202	0.24
3	70.00	69.12	-1.26	8.00	8.10	1.25	15.00	15.07	0.48	30.00	30.30	0.99	40.00	39.95	-0.14	107	202	0.82
4	40.00	39.65	-0.87	35.00	35.10	0.28	45.00	45.02	0.05	35.00	34.85	-0.42	50.00	50.00	0.00	89	202	0.32
5	15.00	14.77	-1.51	10.00	10.05	0.47	20.00	20.06	0.28	18.00	17.92	-0.46	16.00	16.08	0.84	101	202	0.64
6	35.00	34.60	-1.14	30.00	30.05	0.16	28.00	28.06	0.22	25.00	25.71	2.83	26.00	26.06	0.24	110	202	0.94
7	8.00	8.26	3.27	9.00	8.93	-0.77	7.00	7.02	0.29	6.00	5.81	-3.24	5.00	5.14	2.78	61	202	2.07
8	60.00	60.21	0.35	65.00	64.91	-0.14	64.00	64.00	0.00	63.00	62.56	-0.70	58.00	58.11	0.19	129	202	0.28
9	75.00	75.59	0.79	70.00	69.82	-0.26	73.00	72.93	-0.10	74.00	73.94	-0.09	80.00	79.82	-0.22	178	202	0.29
10	60.00	59.68	-0.53	50.00	50.13	0.25	40.00	40.03	0.07	30.00	29.80	-0.66	20.00	19.92	-0.38	117	202	0.38
RSD%		1.08			0.37			0.12			0.85			0.82				
Average								0.54										0.77

4.6 The prediction results obtained in the spectral band of 850-655 cm^{-1}

In this section, no condition except the band of the measured wavenumbers is different from that in 4.5. The results can be seen from Table 4. The average MRE% for all the prediction sample is 0.77%, and the mean RSD % for Ethylbenzene, Styrene, o-Xylene, m-Xylene and p-Xylene in all the prediction samples are 1.08, 0.37, 0.12, 0.85 and 0.28 %, respectively. Their average relative standard deviation is 0.54 %. Compared with the results from Section 4.5, the prediction results in this section are obviously worse. The selected bands in this Section contains less informations which are not enough to correlate the concentrations with the absorbances , so the results are a little worse.

V. CONCLUSION

The concentration analysis of five-component mixture with PLS used in FTIR spectroscopy has been introduced. The parameters *SO* and *SA* used in the PLS method is a good criterion created by us for concluding the similarity between the prediction and calibration samples to ensure the reliability of the prediction results. Better results are obtained when the measured spectral bands in 3160-2800 and 850-655 cm^{-1} are selected. The results show that PLS with a good criterion used in FTIR spectroscopy is a powerful tool for the simultaneous determination of the concentrations of even a complex mixture containing similar compounds with strongly overlapped absorption bands one another.

REFERENCES

1. Martens H., Naes T. Multivariate Calibration, Chichester, UK: John Wiley and Sons , 1989.
2. Geladi P. and Kowalski B.R., Anal. Chim. Acta, 1986; 185:1.

3. Geladi P. and Kowalski B.R., Anal. Chim. Acta, 1986; 185:19.
4. Ritter L. and Draper C., Spectrosc. Lines, 1987; 8:15.
5. Seasholtz M.B., Archibald D.D., Lorber A. and Kowalski B.R., Appl. Spectrosc., 1989; 43:1067.
6. Lindberg W., Person J.A., Wold S. Anal.Chem., 1983; 55:643.
7. Haaland D.M. and Thomas E.V., Anal. Chem., 1988; 60:1193.
8. Haaland D.M., Anal. Chem., 1988; 60:1208.
9. Donanue S.M., Brown C.W., Caputo B. and Modell M.D., Anal. Chem., 1988; 60:1873.
10. Guzman M., Ruzicka J., Christian G.D. and Shelley P. Vibrational Spectroscopy, 1991; 2:1.
11. Walter H., Fianigan D., Appl. Opt., 1975; 14:1423.
12. Wold S., Technometrics, 1987; 20: 397.

Date Received:February 16, 1998
Date Accepted: March 27, 1998