Membrane Introduction Mass Spectrometry Combined with an Orthogonal Partial-Least Squares Calibration Model for Mixture Analysis

Min LI,^{*,**} Lu ZHANG,^{*†} Xiaolong YAO,^{*,**} and Xingyu JIANG^{*,**}

^{*}State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

^{**}University of the Chinese Academy of Sciences, Beijing 100049, China

The emerging membrane introduction mass spectrometry technique has been successfully used to detect benzene, toluene, ethyl benzene and xylene (BTEX), while overlapped spectra have unfortunately hindered its further application to the analysis of mixtures. Multivariate calibration, an efficient method to analyze mixtures, has been widely applied. In this paper, we compared univariate and multivariate analyses for quantification of the individual components of mixture samples. The results showed that the univariate analysis creates poor models with regression coefficients of 0.912, 0.867, 0.440 and 0.351 for BTEX, respectively. For multivariate analysis, a comparison to the partial-least squares (PLS) model shows that the orthogonal partial-least squares (OPLS) regression exhibits an optimal performance with regression coefficients of 0.995, 0.999, 0.980 and 0.976, favorable calibration parameters (RMSEC and RMSECV) and a favorable validation parameter (RMSEP). Furthermore, the OPLS exhibits a good recovery of 73.86 – 122.20% and relative standard deviation (RSD) of the repeatability of 1.14 – 4.87%. Thus, MIMS coupled with the OPLS regression provides an optimal approach for a quantitative BTEX mixture analysis in monitoring and predicting water pollution.

Keywords Membrane introduction mass spectrometry, univariate/multivariate mixture analysis, orthogonal partial-least squares

(Received December 27, 2016; Accepted July 5, 2017; Published November 10, 2017)

Introduction

Membrane introduction mass spectrometry (MIMS) is known as an analytic method for the separation of analytes, usually volatile organic compounds (VOCs) and semi-volatile organic compounds (SVOCs),¹ from a sample matrix that includes water,^{2,3} air,² and solids.⁴ Due to its preconcentration, universality, few or no requirements for pretreatments and solvents as well as its rapid analysis speed and low detection level, MIMS has been widely accepted and used in applications, such as monitoring of reactors,⁵ industrial processes⁶ and fundamental studies.⁷ Importantly, considering these advantages, not only has this method been highly recommended for *in situ* and online monitoring,⁸ but it could be miniaturized and adapted for portable applications, enabling its use in field monitoring and emergency detection.^{9–12}

In a typical MIMS configuration, it is common to impose a 70 eV ionization energy with electron impact (EI), likely resulting in molecular ions and a large number of fragment ions. The molecular ions are responsible for the quantitative determination, while the fragment ions provide structural information.¹³ However, an interference phenomenon may exist due to the presence of such a large number of fragment ions in

the form of overlapping mass spectrum, especially in the analysis of mixtures. For example, Potyrailo *et al.*¹⁴ used MIMS to measure the selectivity (*i.e.*, the fraction of lost analytic signal due to interference) of benzene, toluene, ethyl benzene and xylene (BTEX), and the results (88.3, 47.5, 14.1 and 15.7%, respectively) indicated that there is a small spectral overlap for benzene, a moderate overlap for toluene and a strong overlap for ethyl benzene and xylene. These overlaps, or interferences, may hinder the application of MIMS to the analysis of mixtures or environmental samples. Chemometric methods, particularly multivariate calibration, may play a significant role in solving this problem,¹⁵ which has been also utilized in a multivariate study of the MIMS.¹⁶

A partial-least squares (PLS) regression is a multivariate calibration method. It is a combination of principal component analysis (PCA) and multiple linear regression (MLR).^{17,18} PLS regression can overcome multiple correlations of independent variables or dependent variables to enhance the robustness of a model.¹⁹ Due to its advantageous strengths, PLS regression is currently considered the most powerful multivariate analysis technique,²⁰ and is used widely in many fields, including economics and the social and natural sciences, especially spectroscopy,²¹ chromatography,²² and mass spectrometry²³ analyses.

Despite the high capability of the PLS regression method obtaining a desirable model, abstracted principal components contain comprehensive effects of all aspects, resulting in increasing complexity in the model construction procedure due

[†] To whom correspondence should be addressed.
E-mail: luzhang@niglas.ac.cn

to the increasing amount of unrelated information.²⁴ In real environmental samples, a variety of chemical components are present, some of which may influence the performance of the membrane, thereby interfering with the monitoring signal. For example, a high concentration of hydrophobic compounds can induce the swelling of the membrane, and a high salt concentration is likely to cause signal deviation.²⁵ To avoid the interferences from unrelated information, such as background noise, baseline shift and signal distortion, a derivative of PLS regression, termed orthogonal partial least squares (OPLS), was proposed by Trygg and Wold,²⁶ improving the calibration model interpretation,²⁷ robustness and predictive ability.²⁴ The advantages of the OPLS regression are reflected by the exclusion of uncorrelated variation from the Y-block and the maximum retention of the related effective information.²⁴

BTEX have been identified as a class of major contaminants in industrial wastewater and groundwater,²⁸ necessitating the quantitative and qualitative analysis of BTEX, especially for high frequency online usage. Consequently, in this paper, we aim to formulate a quick quantitative analysis method using the MIMS mass spectrum based on an OPLS regression model, compare this method to the traditional univariate calibration and PLS regression, and subsequently assess the generalization of the obtained optimal calibration model through the substitution of prediction sets. Quality assurance and quality control (QA & QC) are also performed by the recovery assessment of a series of standard and field samples with BTEX additions. It should be noted that this work presents the advanced use of a combination of MIMS and OPLS for the analysis mixtures, which will greatly benefit the construction of an early-warning mechanism of water pollution through online monitoring and the emergent detection of an environmental contamination accident through *in situ* detection.

Experimental

Reagents and chemicals

The stock solutions at 10 ppm are prepared with pure solutions of each analyte by diluting the reagent grade solvents in deionized water and stored at 4°C. Each concentration for the experiment is dispensed from the diluted stock solutions. The sample dataset is shown below (Table 1), with samples 1 to 24 acting as the training dataset for the model, and samples P1, P2 and P3 are used as standards for the external validation. According to the American Society for Testing and Materials (ASTM) guidelines, accurate model building preferably includes at least 24 samples for the spectra analysis.²⁹ To get insight of the reliability and applicability in a real water matrix, three test samples (YR1, YR2 and YR3) were prepared with water from the Yangtze River (0.035 – 0.067 ppb), while the other three samples (PF1, PF2 and PF3) were made from waste water of a petrochemical factory in Nanjing (0.51 – 85.20 ppb). All additions maintained the final samples at the measureable range of the method. Each sample was poured into a 10-mL open glass tube for 5 – 6 min of testing.

Apparatus

The assembled MIMS device incorporated a Pfeiffer Vacuum compact quadrupole mass spectrometer (QMG 220 M1, PrismaPlus) and vacuum equipment (HiPace 300 Turbo Pumps) from Bay Instruments, Easton, MD, USA, with a flow-through capillary membrane inlet. The internal detection system of the instrument is coupled with electron ionization (70 eV), a Faraday cup and a secondary electron multiplier (SEM). We

Table 1 Calibration and prediction datasets used for the BTEX mixture analysis

Sample	Concentration, ppb			
	Benzene	Toluene	Ethyl benzene	Xylene
1	300	650	150	220
2	190	170	200	190
3	180	110	200	185
4	250	460	170	210
5	180	130	200	190
6	170	75	200	185
7	180	120	200	190
8	240	400	170	210
9	220	320	180	200
10	180	150	200	185
11	190	185	195	190
12	230	320	180	200
13	180	155	200	190
14	155	30	210	180
15	155	20	210	180
16	220	320	180	200
17	250	440	170	210
18	160	65	200	185
19	200	220	190	195
20	200	220	190	190
21	155	10	210	180
22	170	95	200	185
23	155	25	210	180
24	150	15	210	180
P1	180	130	200	185
P2	165	60	200	185
P3	150	25	205	180
YR1	200	50	180	220
YR2	180	70	190	210
YR3	190	210	170	200
PF1	150	60	160	190
PF2	260	500	210	180
PF3	250	620	180	220

chose polydimethylsiloxane (PDMS) as the membrane material with a thickness of 250 µm to separate BTEX from water working at 25°C in this study. Water samples were pumped through the stainless steel with a peristaltic pump (Longer, China) at an inlet speed of 1.5 mL/min.

Chemometric data analysis

The quadrupole analyzer is scanned from m/z 50 to 120 at a rate of 200 ms/amu. The output of the mass spectra from 10 and 20 circles of each mixture is used for the data analysis and modeling calibration. Univariate calibration acts as a comparison group for the mixture analysis. Univariate calibration is based on selecting unique m/z in the selected ions mode (SIM) and then performing a least squares regression of the detector response *vs.* the known concentration. Selected predictors for benzene, toluene, ethyl benzene and xylene are 78, 92, 65 and 106 m/z , respectively, which were also used by Susan *et al.*¹⁵ PLS regression is a process that abstracts the main components from the X-block and Y-block simultaneously and then creates a multivariate linear model.³⁰ OPLS systematically separates the X-block into the following two parts: the first is related to the Y-block, while the second is orthogonal to the Y-block.²⁶ The multiple linear regression analysis for the univariate calibration is accomplished using SPSS 20.0 with a Pearson test, while PLS and OPLS regressions are performed directly using SIMCA-P 13.0, Umetrics.

Fig. 1 Correlation plot of the actual concentrations and signal values of 24 samples from the univariate calibration model for the BTEX mixture: (a) ■, benzene (m/z 78, $R^2 = 0.912$); (b) ●, toluene (m/z 92, $R^2 = 0.867$); (c) ▲, ethyl benzene (m/z 65, $R^2 = 0.440$); (d) ◆, xylene (m/z 106, $R^2 = 0.351$). The black line is the regression line.

Results and Discussion

Univariate calibration modeling

The univariate linear regression model of benzene shows a significant linear relationship with a regression coefficient of 0.912 ($p < 0.01$) (Fig. 1a). Additionally, a linear relationship exists between the toluene concentration and its signal intensity, even though the linearship ($R^2 = 0.867$, $p < 0.01$) is not as good as that for benzene (Fig. 1b). However, the calibration model only has 44.0 and 35.1% interpretability for ethyl benzene and xylene, respectively (Figs. 1c and 1d). Presumably, there is some interference between these components. According to the standard spectra of BTEX obtained from NIST chemistry webBook,³¹ this interference may arise from the small intensities of m/z 78 for benzene and the moderate intensities of m/z 92 generated by the other components. In addition, taking a similar spectrum of ethyl benzene and xylene into account, a simple univariate analysis may fail to distinguish the individual signal response for each of them.

PLS regression fitting

For the fitting, there are two components contributing to the model, resulting in R^2Y and Q^2Y of 0.987 and 0.98, respectively (Table 2). R^2Y represents how well the model fits the data, while Q^2Y indicates the predictive ability of this model.²³ Generally, the root-mean square error (RMSE) is used to evaluate the outcomes of the modeling, which is calculated using Eqs. (1) and (2),

Table 2 Multivariate calibration parameters for the BTEX analysis

		RMSEC ^a	RMSECV ^b	RMSEP ^c	F	p	R^2Y	Q^2Y
PLS	B	3.33	3.58	2.7	278.73	1.39×10^{-16}	0.987	0.980
	T	4.45	5.06	3.17	448.23	1.63×10^{-18}		
	E	7.79	9.32	16.33	118.50	3.73×10^{-13}		
	X	34.73	40.80	17.52	165.30	1.77×10^{-14}		
OPLS	B	1.71	1.88	2.37	343.64	1.97×10^{-17}	0.987	0.979
	T	2.21	2.83	2.88	643.85	5.41×10^{-20}		
	E	2.74	4.37	15.63	139.29	8.53×10^{-14}		
	X	6.36	14.11	16.36	160.30	2.35×10^{-14}		

a. Root-mean-square error of calibration.

b. Root-mean-square error of cross-validation.

c. Root-mean-square error of prediction.

$$\text{PRESS} = \sum_i^n (c_i^* - c_i)^2 \quad (1)$$

$$\text{RMSE} = \sqrt{\text{PRESS}/n} \quad (2)$$

where PRESS denotes the predicted residual error sum of squares, n is the number of predictors, c_i^* is the predicted dependent variable (*i.e.*, predicted concentration of BTEX) and c_i is the known dependent variable (*i.e.*, actual concentration).

The calculated root-mean-square errors of the calibration (RMSEC) and cross-validation (RMSECV) evaluated according to these equations are listed in Table 2. Generally, a satisfactory

Fig. 2 Correlation plot of the actual and fitted concentrations of 24 samples from the PLS model for the BTEX mixture: (a) ■, benzene ($R^2 = 0.957$); (b) ●, toluene ($R^2 = 0.956$); (c) ▲, ethyl benzene ($R^2 = 0.919$); (d) ◆, xylene ($R^2 = 0.910$). The black line is the regression line.

model exhibits small RMSEC and RMSECV values; thus, such small values provide evidence that a good PLS model is generated in this work. The linear regression coefficients of the constituents in BTEX mixture are 0.957, 0.956, 0.919 and 0.910 (Fig. 2), which are obviously much better than those obtained from the univariate regression model.

OPLS regression fitting

In the OPLS model, there is one component related to the principal component and one orthogonal component. The level of fitting (R^2Y) is 0.987, while the predictive ability (Q^2Y) is 0.979 (Table 2). Although these two indicators are basically the same as those obtained for the PLS model, there is an improvement in the regression coefficients of the OPLS calibration model compared to those obtained for the PLS, with the obtained values of 0.995, 0.999, 0.980 and 0.976 (Fig. 3) and a decrease ranging from 48.6 to 81.7% in the RMSEC and from 44.1 to 65.4% in the RMSECV values (Table 2). In addition, the F and p values reflect the reliability in the cross-validation analysis of variance, considered as a diagnostic method to provide general information for the assessment of a model,³² showing a better confidence coefficient in the OPLS model except xylene. Generally, not only a better model-fitting level (*i.e.* the regression coefficients) but also the obtained improvements of several common parameters can favor the achieved performance in a better way. In OPLS, continuous variable data is divided into predictive and orthogonal information and the Y-orthogonal variation in X is exempted from the calibration sample, and this model can apply well in any situation in which the Y-variable may be related to X in an

additive manner,³³ thus, the OPLS model might also do a good job under the condition of a real-world sample containing some structured noise (Y-variable), while in PLS, the noise likely complicates this model particularly when many components are present in the model.³³ On the other hand, there seems to be an underfitting of data in the PLS due to the greater volume of generated principal components. Comprehensively, the OPLS regression results indicate that it outperforms the PLS model in the calibration modeling process.

External validation

To validate the reliability and applicability of the obtained calibration models, the prediction dataset, namely P1, P2 and P3, is commonly used to conduct the external validation procedure. The indicator for this procedure is the root-mean-square error of prediction (RMSEP). A small value of RMSEP value indicates a high predictive ability.²⁴ As shown by the data in Table 2, OPLS shows a relatively lower RMSEP, demonstrating a robust predictive ability. Despite the relatively poor performance of the ethyl benzene and xylene calibration model caused by the obscured contribution of similar spectra, the use of OPLS is still advantageous in the quantification of ethyl benzene and xylene. An unambiguously linear relationship exists in these three test samples, with the PLS and OPLS regression coefficients both equal to 0.997 ($p < 0.01$) (Fig. 4). However, the fitting lines for both PLS and OPLS diverge obviously from the expected line with a slope of 1, demonstrating the presence of a large deviation at lower concentrations. It seems that both methods have poor predictability in low-concentration BTEX analysis. This assumption is consistent

Fig. 3 Correlation plot of the actual and fitted concentrations of 24 samples from the OPLS model for the BTEX mixture: (a) ■, benzene ($R^2 = 0.995$); (b) ●, toluene ($R^2 = 0.999$); (c) ▲, ethyl benzene ($R^2 = 0.980$); (d) ◆, xylene ($R^2 = 0.976$). The black line is the regression line.

Fig. 4 Correlation plots of the actual and predicted concentrations of the P1, P2 and P3 samples obtained using the PLS model (a, slope = 0.885, $R^2 = 0.997$) and OPLS (b, slope = 0.910, $R^2 = 0.997$) for the BTEX mixture: ■, P1; ●, P2; ▲, P3. The solid line is the regression line, while the dashed line is the expected line with the slope equaling 1.

with the estimated recoveries. Taking the results of the OPLS in Table 3 as an illustration, when the OPLS method is used for detecting lower concentrations (YR1, YR2 and PF1), the recoveries reach 152.74, 136.23 and 141.14%, implying the presence of overestimation of analytical concentrations. However, the recoveries of the PLS and OPLS method are within the acceptable range of 73.97 – 122.44 and 73.86 – 122.20%, respectively, which are close to those obtained in

research on BTEX using GC methods,^{34–37} regardless of the recoveries of lower concentrations. Thus, it might limit the application of this method in detecting compounds in low concentrations (lower than 100 ppb in this case), such as low concentrations of benzene due to its low maximum contaminant level (MCL) of 5 ppb according to the drinking water standards (2004 edition) of the United States Environmental Protection Agency, while the MCL of toluene, ethyl benzene and xylene is

Table 3 Results of the OPLS tests for the recovery of BTEX in the Yangtze River and industrial sewage

	Benzene, %	Toluene, %	Ethyl benzene, %	Xylene, %
YR1	83.42	152.74	113.39	83.65
YR2	91.19	136.23	104.17	90.06
YR3	103.65	96.87	113.34	96.00
PF1	94.40	141.14	96.32	115.19
PF2	117.99	122.20	73.86	110.85
PF3	119.71	89.11	100.74	93.79

1, 0.7 and 10 ppm, respectively. Furthermore, the relative standard deviation of the repeatability of this method is calculated as 1.14 – 4.87%.

Therefore, based on the multivariate method, it can provide important information in a timely manner such as for the high-concentration leakage of BTEX from a factory or a shipwreck. In other words, it facilitates the development of an online MIMS monitoring system or an *in situ* MIMS emergent detecting system. As for the limitation present at low concentrations, this might be minimized by taking the diffusion time of analytes into account when establishing a model, while the complexity resulting from the ongoing increase in the number of analytes might be minimized by expanding the dataset of a model as well as decreasing the unimportant variables through optimizing working conditions such as ionization energy.

Conclusions

In the present study, multivariate calibration is demonstrated to be obviously superior to the comparatively poor results from the univariate analysis with the obtained regression coefficients of 0.912, 0.867, 0.440 and 0.351 for benzene, toluene, ethyl benzene and xylene, respectively (Fig. 1). Further exploration of the multivariate calibration methods, PLS regression and its extension, the OPLS regression, are applied to membrane introduction mass spectrometry for a mixture analysis (BTEX in this case). Both techniques could overcome the interferences between variables and then show good behaviors with all regression coefficient values above 0.900 (Figs. 2 and 3). In our analysis, compared to the PLS regression model, the OPLS model reduces the dimensions of the model, leading to an easier interpretation and more optimal performance parameters. Therefore, MIMS in conjunction with the OPLS regression analysis, which is considered an optimization of the PLS-MIMS, provides a viable, robust and improved approach for the on-line analysis of mixtures, working as the “informer” of unexpected environmental pollution to facilitate researchers in their efforts to determine the next treatment.

References

1. T. M. Allen, M. E. Cisper, C. W. Wilkerson, and P. H. Hemberger, *Int. J. Mass Spectrom.*, **2001**, 212, 197.
2. K. Demeestere, J. Dewulf, B. D. Witte, and H. V. Langenhove, *J. Chromatogr. A*, **2007**, 1153, 130.
3. E. T. Krogh and C. G. Gill, *Int. J. Mass Spectrom.*, **2014**, 49, 1205.
4. R. A. Ketola, A. Smolander, J. Kiuru, T. Kotiaho, and V. Kitunen, *Boreal Environ. Res.*, **2011**, 16, 36.
5. D. G. Mulat, A. J. Ward, A. P. S. Adamsen, N. V. Voigt, J. L. Nielsen, and A. Feilberg, *Environ. Sci. Technol.*, **2014**, 48, 2505.
6. P. J. Strong and D. J. Gapes, *Waste Manage.*, **2012**, 32, 1669.
7. Y. Zhao, Y. Xia, C. Ti, J. Shan, B. Li, L. Xia, and X. Yan, *Environ. Sci. Technol.*, **2015**, 49, 1427.
8. T. Zavřel, H. Knoop, R. Steuer, R. J. Patrik, J. Črvený, and M. Trtílek, *Bioresour. Technol.*, **2016**, 202, 142.
9. G. H. Kristensen, M. M. Klausen, V. A. Hansen, and F. R. Lauritsen, *Rapid Commun. Mass Spectrom.*, **2010**, 24, 30.
10. L. Zhang, C. Li, B. Cao, S. Shao, and X. Ding, *Anal. Instrum.*, **2010**, 33, 1.
11. P. M. Mach, E. M. McBride, Z. J. Sasiene, R. B. Katie, K. K. Shelia, C. W. Kenneth, and F. V. Guido, *Anal. Chem.*, **2015**, 87, 11501.
12. R. J. Bell, N. G. Davey, M. Martinsen, C. H. Christian, T. K. Erik, and G. G. Christopher, *J. Am. Soc. Mass Spectrom.*, **2015**, 26, 212.
13. Private Communication, Y. Z. Ouyang (Central South University), **2010**.
14. R. A. Potyralo, *Proc. SPIE Int. Soc. Opt. Eng.*, **1999**, 3856, 181.
15. S. K. Ohorodnik, R. E. Shaffer, J. H. Callahan, and S. L. Rose Pehrsson, *Anal. Chem.*, **1997**, 69, 4721.
16. H. Abdi, “*Encyclopedia for Research Methods for the Social Sciences*”, ed. M. Lewis-Beck, A. Bryman, and T. Futing, **2003**, Thousand Oaks, CA, 792.
17. P. Geladi and B. R. Kowalski, *Anal. Chim. Acta*, **1986**, 185, 1.
18. H. W. Wang, *National Defense Ind.*, **1999**, 1.
19. H. E. Martens and T. Naes, “*Multivariate Calibration*”, **1989**, John Wiley and Sons, New York, Chichester, Brisbane, Toronto.
20. D. Wu, Y. He, and S. Feng, *Anal. Chim. Acta*, **2008**, 610, 232.
21. M. Daszykowski, Y. V. Heyden, and B. Walczak, *J. Chromatogr. A*, **2007**, 1176, 12.
22. T. T. Zou, J. Y. He, and Q. X. Qi, *Food Ferment. Ind.*, **2016**, 42, 179.
23. R. A. Ketola, R. T. Short, and R. J. Bell, “*Comprehensive Sampling and Sample Preparation*”, ed. J. Pawliszyn, **2012**, Vol. 2, Chap. 24, Elsevier, Pittsburgh, 497.
24. J. Trygg and S. Wold, *J. Chem.*, **2002**, 16, 119.
25. S. Wiklund, E. Johansson, L. Sjoestrom, E. J. Mellerowicz, U. Edlund, J. P. Shockcor, J. Gottfries, T. Moritz, and J. Trygg, *Anal. Chem.*, **2008**, 80, 115.
26. G. O. Yahaya, *J. Membr. Sci.*, **2008**, 319, 82.
27. J. Workman, J. Brown, and J. P. Coates, *Spectroscopy*, **1996**, 11, 24.
28. S. Wold, M. Sjöstrom, and L. Eriksson, *Chemom. Intell. Lab. Syst.*, **2001**, 58, 109.
29. J. Chen, W. Wang, S. Lv, P. Yin, X. Zhao, X. Lu, F. Zhang, and G. Xu, *Anal. Chim. Acta*, **2009**, 650, 3.
30. NIST chemistry webBook, <http://webbook.nist.gov/chemistry>.
31. E. Lennart, T. Johan, and W. Svante, *J. Chemom.*, **2008**, 22, 594.
32. L. Eriksson, M. Toft, E. Johansson, and S. Wold, *J. Chemom.*, **2006**, 20, 352.
33. A. Zhang, H. Wang, L. Weng, J. Hou, and Y. Fang, *Phys. Test. Chem. Anal., Part B*, **2014**, 50, 1104.
34. L. Kang, J. Yang, and R. Li, *Chin. J. Health Lab. Technol.*, **2009**, 19, 2528.
35. G. Hu, J. Li, L. Yuan, X. Xia, H. Wang, and X. Shi, *Environ. Monit. Chin.*, **2011**, 27, 30.