Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1086 (2005) 160–164
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Deconvolution of overlapped peaks based on the exponentially modified Gaussian model in comprehensive two-dimensional gas chromatography

Hongwei Kong, Fen Ye, Xin Lu, Lei Guo, Jing Tian, Guowang Xu *

National Chromatographic R&A Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China

Available online 24 June 2005

Abstract

Comprehensive two-dimensional gas chromatography (GC × GC) has attracted much attention for the analysis of complex samples. Even with a large peak capacity in GC × GC, peak overlapping is often met. In this paper, a new method was developed to resolve overlapped peaks based on the mass conservation and the exponentially modified Gaussian (EMG) model. Linear relationships between the calculated σ, τ of primary peaks with the corresponding retention time (t_R) were obtained, and the correlation coefficients were over 0.99. Based on such relationships, the elution profile of each compound in overlapped peaks could be simulated, even for the peak never separated on the second-dimension. The proposed method has proven to offer more accurate peak area than the general data processing method.

Keywords: Gas chromatography; Comprehensive two-dimensional gas chromatography; EMG; Curve fitting; Overlapped peak; Deconvolution

1. Introduction

Since the first emergence in 1991, comprehensive two-dimensional gas chromatography (GC × GC) [1] has obtained much attention in separation science, and has been successfully applied into the analysis of petrochemicals, essential oils, pesticides, cigarette smoke, food, and so on [2–7].

Several approaches are reported to perform peak quantification in GC × GC. The most common one integrates all individual second-dimension peaks by means of conventional integration algorithms, and then sums all peak areas belonging to one compound [8–10]. For another method, firstly a so-called base plane is subtracted, and subsequently three-dimensional peak volumes are calculated by means of imaging procedures [11]. Although the peak capacity of GC × GC is high, the peak overlapping in two-dimension (2D) separation is very possible, especially for highly complex samples.

The chemometric methods, like the generalized rank annihilation method (GRAM), have been used to resolve and quantify severely overlapped GC × GC peaks [12–14]. However, a bilinear structure of the data is generally required [14–16], strictly speaking, which is not completely correct when using temperature-programmed conditions for the second-dimension separation in GC × GC. Due to the modulation cycle may start at an indeterminate offset from the injection, the profile of the peaks cannot be completely reproducible. The GRAM is also not suitable for GC × GC with non-synchronized modulation mode like KT-2001 from Zoex. Some other methods have also been used to resolve overlapping chromatograms in 1D-GC, LC/DAD or GC/MS, like curve fitting [17], wavelet analysis [18], orthogonal projection resolution [19], and so on.

In this paper, an overlapped peak deconvolution method was developed based on the mass conservation and the EMG model. According to the raw secondary dimension peaks and relationships between σ, τ and t_R, the primary peak was simulated, and the elution profile of each compound in overlapping peaks can be deconvoluted. By which, more correct areas can be obtained than common data processing method.

2. Theory

In GC × GC, the primary peak emerging from first column is transferred to a secondary column via a modulating

* Corresponding author. Fax: +86 411 84379559.
E-mail address: dicp402@mail.dlptt.ln.cn (G. Xu).

0021-9673/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2005.05.103

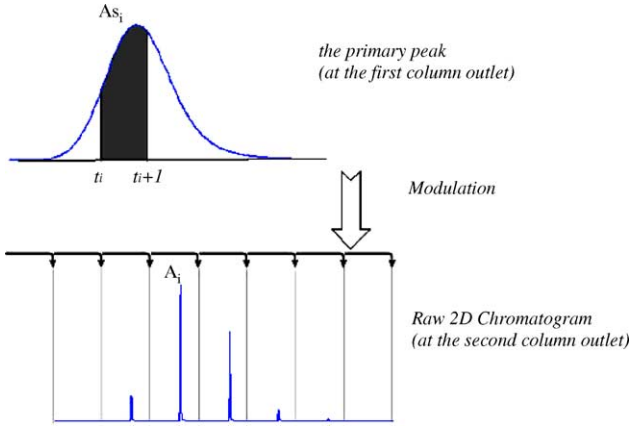

Fig. 1. Translation of primary peak to modulated secondary peaks.

interface with the period t(s) to produce a series of narrower peaks. Based on the principle of mass conservation, the areas of modulated and non-modulated peaks are the same.

$$A_{\text{all}} = \sum_{i=0}^{n} A_i \tag{1}$$

where A_i is the area of raw 2D peak in ith period at the second column outlet, n the modulated times of the primary peak. A_i should be equal to As_i that was the area of ith fraction at the first column outlet (as shown in Fig. 1).

$$\text{As}_i = A_i \tag{2}$$

Compared with the strict Gaussian function, other peak functions can be used to more correctly describe the chromatography peak [20]. The model of peak shape used in this article is the exponentially modified Gaussian (EMG) function, which is an accepted model for skewed peaks [21–22],

$$h(t) = \frac{A}{\tau} \exp\left(\frac{1}{2}\left(\frac{\sigma}{\tau}\right)^2 - \frac{t - t_{\text{G}}}{\tau}\right) \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{y^2}{2}\right) \mathrm{d}y \tag{3}$$

where A is the peak area, t_{G} the central position of a Gaussian constituent, σ the one standard deviation of a Gaussian constituent, τ is the time constant for exponential decay.

Base on the EMG model, the area of the fraction (As_i) can be calculated by Eq. (4),

$$\text{As}_i = \int_{t_i}^{t_{i+1}} h(t)\,\mathrm{d}t \tag{4}$$

and the key to get As_i is to know the peak parameters (t_{G}, σ, τ) of the primary peak.

According to the Eq. (2), As_i should be equal to A_i. Therefore, the deviation (Δ) can be used to evaluate whether the peak parameters of the primary peak is correct,

$$\Delta = \sum_{i=0}^{n} |\text{As}_i - A_i| \tag{5}$$

When the peak parameters (t_{G}, σ, τ) of a primary peak fit to fact very well, Δ reaches the minimum value.

In a given chromatogram, the ranges of peak parameters can be estimated very easily. Because the common PC computer has a very high calculation speed, we can obtain the Δ values in total possible ranges of σ, τ and t_{G} in a short time. When the Δ is the minimum, the corresponding σ, τ and t_{G} values are those we want to determine. Based on our experience, Δ is less than 0.6% of total peak area when we have the correct peak parameter values.

Because there are three unknown variables, Eq. (5) is suitable to calculate the peak parameters only when the primary peak is modulated at least three times [23,24].

In order to deconvolute the overlapped peaks, four peak parameters ($t_{\text{G}}, \sigma, \tau, A$) of the primary peak of each compound should be known. Previously, the relationships among σ, τ and t_{R} have been explored [17],

$$\sigma = a_1 + b_1 t_{\text{R}} \tag{6}$$

$$\tau = a_2 + b_2 t_{\text{R}} \tag{7}$$

where a_1, b_1, a_2, b_2 are dynamically related coefficients that can be obtained by the regression of several well resolved peaks. And the retention time of simulated peaks (t_{R}) can be got by the following equation,

$$t_{\text{R}} = \frac{t_{\text{G}} - \tau}{2} \tag{8}$$

Therefore, only two parameters, the t_{G} and area (A) for each peak need be defined, and can be obtained by following method.

We suppose that the compounds in the overlapped peak are partly overlapped. So the number of compounds can be obtained by common workstation software such as the Chemstation from Agilent or GC-Image from Zoex. For a given t_{G}, σ, τ and t_{R} can be calculated from Eqs. (6) to (8). Based on the principle of mass conservation, the deviation (Δ) can be used to evaluate the peak parameters of each compound,

$$\Delta = \sum_{i=0}^{n} |A_i - \sum_{j=0}^{k} (\text{As}_j)_i| \tag{9}$$

where k is the component number in the overlapping peaks. In a similar method in a single peak simulation mentioned above, in the possible ranges of peak parameters, Δ values are calculated, and the minimum value Δ_{min} is searched. If the Δ_{min} value was greater than 1% of total area, it can be judged that the possible compound number in the overlapped peak is more than the supposed one. We should add one to the peak number, and continue the calculation of Δ values in the total possible ranges of peak parameters to find the Δ_{min} value. This process of the iterative operation is continued until the deviation Δ_{min} was less than 1% of the total area. When such a procedure is finished, the retention time and the area of each compound are defined, the elution profiles are simulated based on Eq. (3), and the overlapped peak can be deconvoluted.

3. Experimental

3.1. Apparatus

An Agilent 6890 GC (Agilent Technologies, Wilmington, DE) fitted with a conventional split/splitless injector and a flame ionization detector (FID) with 100 Hz data collection rate was used as the detector.

A liquid-nitrogen-cooled jet modulator (KT-2001), supplied by Zoex Corp. (Lincoln, NE) was used for chromatographic modulation. Description of theoretic and operational characteristics of such a modulator is available from literatures [25]. In our experiment, the modulation period was set at 3 s.

Data acquisition and raw 2D peak integration were performed by using a Chemstation software from Agilent.

3.2. GC columns and conditions

The column set used for GC × GC was a DB-5 column (8 m × 0.1 mm i.d. × 0.25 μm) directly coupled to a DB-17ht column (0.8 m × 0.1 mm i.d. × 0.1 μm) using glass press-fit. Both columns were supplied by J&W Scientific (Folsom, CA, USA).

Helium was used as the carrier gas, and supplied to the column at 250 kPa (140° C) under constant-flow model. Temperatures of the injector and detector were set at 280° C. The injection volume was 0.01 μl, and the split ratio was 50.

3.3. Materials

The standard samples, pyridine, 2-methylpyridine, 4-methylpyridine, 2,6-dimethylpyridine, 2-ethylpyridine, 2,5-dimethylpyridine, 2,4,6-trimethylpyridine and 2,3,5-trimethylpyridine were A.R. and from Sigma–Aldrich (Milwakaukee, WI, USA).

4. Results and discussion

In GC × GC, the primary peak was eluted from first column, and divided into a series of narrower peaks by modulation. An part of raw second-dimension chromatogram at an isothermal experiment (140° C) was shown in Fig. 2A. It can be seen that the 2,3,5-trimethylpyridine was modulated five times so that five raw second-dimension peaks were obtained.

The areas of raw individual second-dimension peaks can be obtained by the Agilent Chemstation software. According to Eq. (1), the area of the primary peak of 2,3,5-trimethylpyridine can be calculated by summing the areas of five raw second-dimension peaks, which is equal to 602.2 pA s.

The peak parameters were searched in the total peak parameter ranges as described above. When a set of parameters were defined, the area of each fraction and the deviation can be calculated by Eq. (1), (2), (4) and (5). It was observed

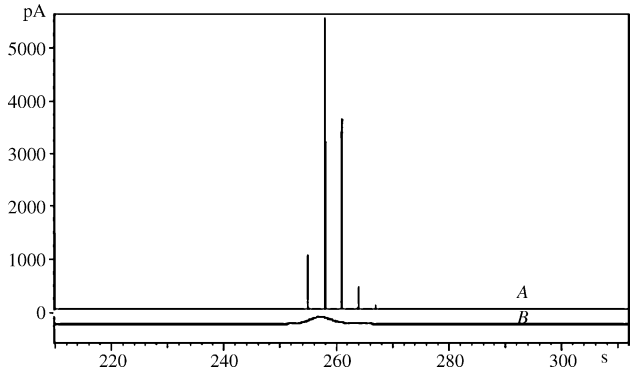

Fig. 2. The simulated result for 2,3,5-trimethylpyridine. (A) The raw second-dimension chromatogram from the Chemstation. (B) The chromatogram for the simulated primary peak.

Table 1
Comparison of the raw second dimension peak area (A_i) with simulated result (As_i)

Period	1	2	3	4	5	Δ
A_i	56.8	319.6	196	27	3.8	
As_i	56.8	319.6	194.4	27.7	2.8	3.3

that when $t_G = 256.15$s, $\sigma = 1.58$ and $\tau = 1.265$, the Δ was the minimum, and equal to 3.3 pA s, or 0.548% of total area (shown in Table 1). Based on Eq. (3), the primary peak chromatogram can be simulated, and shown in Fig. 2B.

Injection and modulation operations are not synchronized in the modulator KT-2001 (Zoex). So t_R and A_i of the raw second-dimension peaks are not very reproducible, as shown in Fig. 3A. However, with our developed method, the simulated primary peaks were coincided very well (showed in Fig. 3B), the retention time and peak width of the primary peak can be obtained accurately.

Fig. 4A shows a raw second-dimension chromatogram of eight components, in which six compounds were completely separated at second column outlet, five of them were modulated for more than three times. The parameters (t_G, σ, τ) of each primary peak from these five compounds can be calculated according to Eq. (1), (2), (4) and (5), and the simulated chromatogram of primary peaks were shown in Fig. 4B.

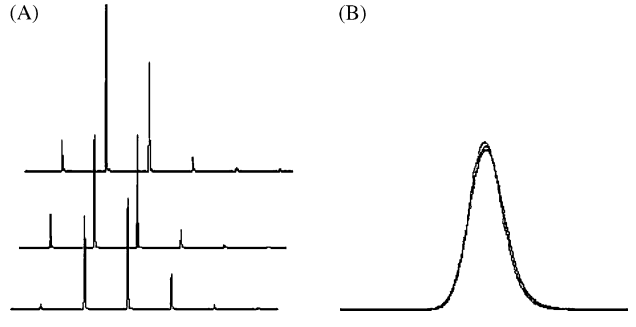

Fig. 3. (A) Overlay of three raw second-dimension chromatograms for a same sample at the same operational conditions. (B) Overlay of three chromatograms from the simulated primary peaks.

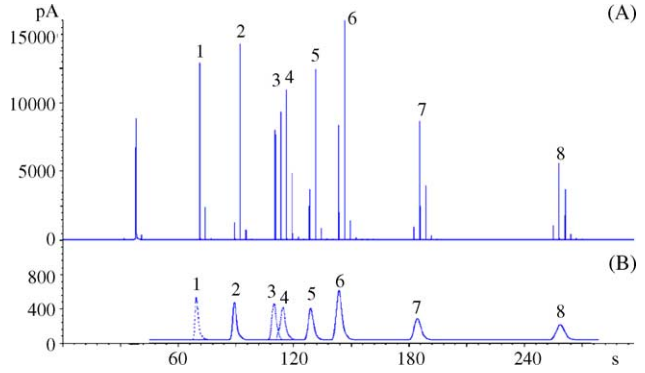

Fig. 4. (A) The raw second-dimension chromatograms. (B) The simulated chromatogram of primary peaks. Peaks 1, 3 and 4 (dashed line) were drawn based on σ and τ calculated from Eqs. (6) and (7). Peak identification: (1) pyridine; (2) 2-methylpyridine; (3) 4-methylpyridine; (4) 2,6-dimethylpyridine; (5) 2-ethylpyridine; (6) 2,5-dimethylpyridine; (7) 2,4,6-trimethylpyridine; (8) 2,3,5-trimethylpyridine.

According to Eq. (8), the retention time can be calculated from t_G and τ. In our study, quite good linear relationships, respectively, between the calculated σ or τ with the corresponding retention time of primary peaks were obtained with the correlation coefficients over 0.99, as shown at Fig. 5.

It could be observed from Fig. 4A that 4-methylpyridine and 2,6-dimethylpyridine were still totally overlapped in the raw second-dimension chromatogram, and the total area was 1465.5 pA s. With common data processing method in the Chemstation from Agilent, only one peak was detected. Therefore it was supposed that only one compound in this peak. Then the peak deconvolution operation was performed. The results revealed that the peak area was 930.5 pA s, t_G was 115.6 s, and the Δ_{min} was 562.77 pA s, more than 1% of total area. Therefore, it can be judged that there should be more than one compound in this peak. With the supposed compound number increased to two, the final calculated deviation Δ_{min} of iterative result was 11.3 pA s, 0.7% of the total area. Therefore, the iterative calculation was finished. The fit results were as follows: 4-methyl pyridine (area = 788.5 pA s, t_G = 112.6 s), 2,6-dimethyl pyridine (area = 677 pA s, t_G = 117.9 s). Further detailed results were

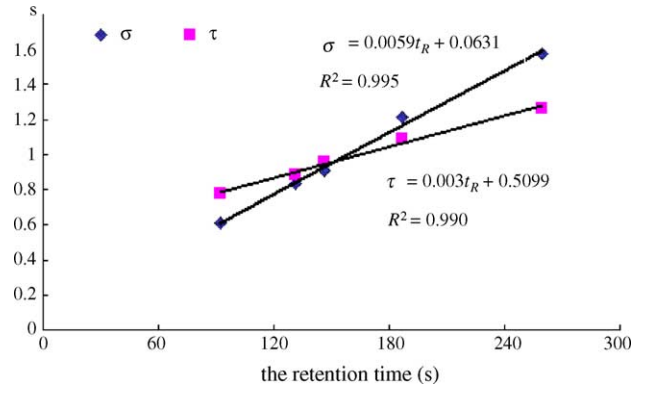

Fig. 5. Relationship between σ, τ and the corresponding retention time on the first dimension.

Table 2
Comparison of the raw second dimension peak area (A_i) with simulated result ($\sum As_i$)

Period	1	2	3	4	5	Δ
A_i	360.9	409.2	487.7	199	8.7	
$\sum As_i$	360.77	414.16	488.26	195.59	6.431	11.33

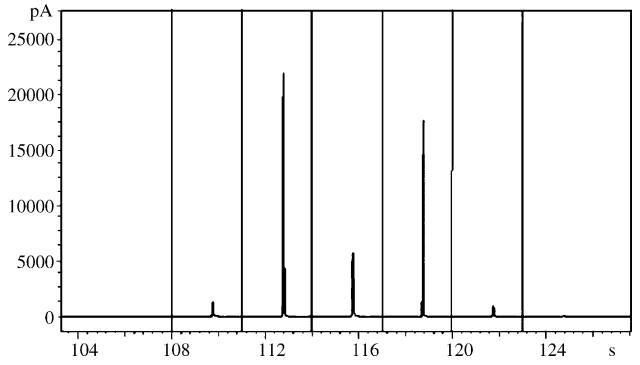

Fig. 6. An amplified raw second-dimension chromatogram from an overlapped peak of 4-methylpyridine, 2,6-dimethylpyridine on first-dimension. Sample and operational conditions were the same as those in Fig. 4.

shown in Table 2. In addition, based on Fig. 5, σ and τ can be calculated for peaks 1, 3 and 4. Therefore, the peak profiles for these three compounds could be drawn according to Eq. (3).

Fig. 6 gave another chromatogram of 4-methylpyridine and 2,6-dimethylpyridine that was different from Fig. 4 because of the un-synchronized injection and modulation. It was found that two compounds were partly overlapped on second-dimension. Their total area was 2251.4 pA s. The simulated result was obtained by the same method as mentioned above: 4-methyl pyridine (area = 1209.4 pA s, t_G = 113.1), 2,6-dimethyl pyridine (area = 1042 pA s, t_G = 118.2). According to GC-Image Ver 1.0 (Zoex), with dividing peaks at valleys in 2D, the peak integral results were as follows: 4-methyl pyridine (area = 1264.5 pA s), 2,6-dimethyl pyridine (area = 985.1 pA s), and the peak area ratio was 1.284.

To verify the above data, the same mixture was separated at 100 °C, a baseline separation of 4-methylpyridine and 2,6-dimethylpyridine was obtained. It was found that the ratio ($Vol_{4\text{-methyl pyridine}}/Vol_{2,6\text{-dimethyl pyridine}}$) was 1.158. The ratios obtained from two simulated results mentioned above were 1.164 and 1.160, respectively, which demonstrated that the simulated results were quite close to the experimental real value, and much better than that from GC-Image.

5. Conclusions

A simulation method was introduced to resolve overlapped peaks in GC × GC. Based on the principle of mass conservation and the EMG model, the primary peak can be simulated by the raw secondary dimension peaks. The simulation primary peaks were coincided very well, even for

the un-synchronized injection and modulation. Therefore, the retention time and peak width of the primary peak can be obtained correctly. Linear relationships between the calculated σ and τ of primary peaks with the corresponding retention times were obtained, and the correlation coefficients were over 0.99. In addition, the deviation of calculated result between the total second peak area and the simulated peak area was used to judge how many compounds were in the overlapped peaks. By the iterative calculation, the elution profile of each compound in overlapping peaks can be simulated even for unresolved peak on second-dimension. The proposed method will lead to more accurate peak area than general data processions.

Acknowledgements

This work has been supported by the National Basic Research Program of China (No.2003CB716003) of State Ministry of Science and Technology of China, the foundation (No. 20425516) for Distinguished Young Scholars from National Natural Science Foundation of China and the Knowledge Innovation Program (K2002A12, K2003A16) of the Chinese Academy of Sciences.

References

[1] Z. Liu, J.B. Phillips, J. Chromatogr. Sci. 29 (1991) 227.
[2] J. Dalluge, J. Beens, U.A.Th. Brinkman, J. Chromatogr. A 1000 (2003) 69.
[3] G.W. Xu, F. Ye, H.W. Kong, X. Lu, X.J. Zhao, Chin. Chromatogr. 19 (2001) 132.
[4] R. Hua, Y. Li, W. Liu, J. Zheng, H.B. Wei, J.H. Wang, et al., J. Chromatogr. A 1019 (2003) 101.
[5] R. Shellie, L.-L. Xie, P.J. Marriot, J. Chromatogr. A 968 (2002) 161.
[6] J. Dalluge, M. van Rijn, J. Beens, R.J.J. Vreuls, U.A.Th. Brinkman, J. Chromatogr. A 965 (2002) 207.
[7] J. Beens, J. Dalluge, M. Adahchour, R.J.J. Vreuls, U.A.Th. Brinkman, J. Microcol. Sep. 13 (2001) 134.
[8] P. Korytar, P.E.G. Leonards, J. de Boer, U.A.Th. Brinkman, J. Chromatogr. A 958 (2002) 203.
[9] J. Dalluge, M. van Rijn, J. Beens, R.J.J. Vreuls, U.A.Th. Brinkman, J. Chromatogr. A 965 (2002) 207.
[10] J. Beens, H. Boelens, R. Tijssen, J. Blomberg, J. High Resolut. Chromatogr. 21 (1998) 47.
[11] S.E. Reichenbach, M. Ni, D. Zhang, E.B. S Ledford, Proceedings of the 25th International Symposium of Capillary Chromatography, Riva del Garda, Italy, 2002.
[12] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Anal. Chem. 70 (1998) 2796.
[13] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 73 (2001) 5833.
[14] K.J. Johnson, B.J. Prazen, R.K. Olund, R.E. Synovec, J. Sep. Sci. 25 (2002) 297.
[15] A.E. Sinha, J.L. Hope, B.J. Prazen, C.G. Fraga, E.J. Nilsson, R.E. Synovec, J. Chromatogr. A 1056 (2004) 145.
[16] R.E. Synovec, B.J. Prazen, K.J. Johnson, C.G. Fraga, C.A. Bruckner, in: P.R. Brown, E. Grushka (Eds.), Advances in Chromatography, vol. 42, Marcel Dekker Inc., New York, 2003, p. 1.
[17] H. Li, B. Lin, P. Lu, Sci. Sinica B (Chin.) 5 (1986) 449.
[18] X. Shao, W. Cai, P. Sun, et al., Anal. Chem. 69 (1997) 722.
[19] X. Shao, W. Cai, Chem. J. Chin. Univ. 18 (1997) 49.
[20] J. Li, J. Chromatogr. A 952 (2002) 63.
[21] E. Grushka, Anal. Chem. 44 (1972) 1733.
[22] W.E. Barber, P.W. Carr, Anal. Chem. 53 (1981) 1939.
[23] R.E. Murphy, M.R. Schure, J.P. Foley, Anal. Chem. 70 (1998) 1585.
[24] J.V. Seeley, J. Chromatogr. A 962 (2002) 21.
[25] E.B. Ledford, J.B. Philips, US Patent 6,007,602 (1999).