Journal of Chromatography A, 1218 (2011) 3042–3047

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Resolving batch chromatographic overlapping peaks of flavoring essence using stepwise key spectrum selection

Haohui Kong[a], Cuilin Chen[a], Limei Peng[b], Feng Gan[b,*]

[a] Technology Centre, China Tobacco Guangdong Industrial Limited Corporation, Guangzhou 510145, China
[b] School of Chemistry and Chemical Engineering, Sun Yat-sen University, Guangzhou 510275, China

ARTICLE INFO

Article history:
Received 7 October 2010
Received in revised form 2 March 2011
Accepted 2 March 2011
Available online 10 March 2011

Keywords:
Batch analysis
Multivariate curve resolution
Stepwise key spectrum selection
Tobacco flavoring essence

ABSTRACT

Stepwise key spectrum selection (SKSS) was introduced to resolve batch overlapping peaks from gas chromatography–mass spectrometry (GC–MS) analysis of ten batch tobacco flavoring samples in different storage times. Resolution was implemented on a software platform that embedded the SKSS method. The data from GC–MS analysis of the samples were saved and prepared in ASCII files and then were inputted into the software platform for visual inspections. The data segment with overlapping peaks was precut for subsequent analysis. Spectral background in the data was removed using a linear fitting of the baseline. Four components in the overlapping peaks were automatically detected by the SKSS method. The resolution of the concentration profiles and spectra of the four components was conducted by setting only one parameter, the negative area ratio, as 0.01. The fixed size moving window evolving factor analysis and evolving factor analysis were applied to validate the resolved concentration profiles. The resolved mass spectra were validated by the searched standard through library search at the pure component regions revealed by the resolved concentration profiles. The results showed that the SKSS method could be a simple but powerful tool in resolving batch chromatographic overlapping peaks.

1. Introduction

Adding flavors to products is an important step for many commodities such as tobacco. Flavoring materials are usually sprayed on tobacco leaves to provide aromatic flavors, improve the taste and mask unpleasant smells. The development and manufacture of flavoring materials are important especially when there are requirements of low "tar" cigarette [1,2], which contain a low level of nicotine-free dry particulate matter content in blended cigarettes. Thus, the analysis of flavoring materials becomes essential in a tobacco plant, which challenges the laboratory technicians and instruments. In most cases, GC–MS works well in analyzing the components of the flavoring materials [3–8]. However, overlapping peaks are inevitable especially when there are many daily batch samples to be analyzed.

There are two possible solutions for the overlapping peaks. One is to adjust analytical instrumental conditions to improve the separation, but such improvements require a substantial investment in equipment and method development time. Alternatively, multivariate curve resolution (MCR) methods can be applied to the chromatographic data to resolve the overlapping peaks into separate components. Some MCR methods, such as the evolving factor analysis (EFA) [9,10], window factor analysis (WFA) [11] and heuristic evolving latent projection (HELP) [12,13] have proven to be useful tools to analyze the peaks. These methods mainly depend on component window information. For example, the HELP method determines the selective region and the zero concentration region using fixed size moving window evolving factor analysis (FSMWEFA) [14]. It usually needs four parameters to mark the regions. After the resolution of a component, the component stripping procedure [15] is used to subtract the contribution of the component from initial data. These steps will be repeated until all the components are resolved. The users of the method have to spend a lot of time in determining proper parameters to obtain good results.

Some resolution methods that can resolve overlapping peaks automatically had been developed. For example, Malinowski developed an automatic strategy based on WFA [16]. Manne and Grande [17] proposed their automatic resolution method based on elementary matrix transformations. Gan et al. [18] developed their stepwise key spectrum selection (SKSS) by reselecting key spectra in resolution process. Their method can generate the key spectra and determine the numbers of overlapped components automatically. Furthermore, there is only one parameter, the negative area

* Corresponding author. Tel.: +86 20 31725310.
E-mail address: cesgf@mail.sysu.edu.cn (F. Gan).

0021-9673/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.03.002

ratio, needed to be determined by users. All of these methods greatly improve the efficiency in analyzing chromatographic data. They also show potential applications in a routine analysis.

In our present work, a batch of tobacco flavoring materials needs to be investigated. The flavoring materials were fermented and stored for different periods of time. The objective was to control the quality of the flavoring materials by monitoring the concentration changes of the components in a storage period. There are two data analysis problems that require a solution. First, the overlapping peaks need to be resolved to obtain complete information about the components. Second, large data sets need to be analyzed in a short time. Thus, an effective and efficient MCR method was required.

In this paper, we report a new application of the SKSS to this problem. There are two reasons that make us choose the SKSS. First, there is only one parameter to be adjusted in the resolution process so it is very convenient for users who have little knowledge on chemometrics. Second, a software platform which embedded the SKSS method was available. We also show that a library search of the overlapped components can be implemented with the help of the resolved concentration profiles.

2. Experimental

2.1. Materials

Tobacco flavoring materials were a mixture of flavors and additives. The materials were added to cut tobacco prior to cigarette manufacture. The flavoring materials of different formulas were prepared in our technique center and then stored in a room for further ferment. In a period of ten months, one sample of the materials was put in a refrigerator at a constant temperature of 5 °C, one for each month. Finally, all the samples were taken out from the refrigerator for analysis. In this paper, we only discuss the samples from formula A.

HPLC grade dichloromethane of (CNW Technologies GmbH, Germany) and analytical grade anhydrous sodium sulfate (Fuchen Chemical Reagent Plant, Tianjin, China) were used in this study. The anhydrous sodium sulfate was heated at 650 °C for 5 h before it was used to dehydrate the dichloroethane extracts of the tobacco flavor samples.

2.2. Extraction of volatile components

The volatile and semi-volatile components in a tobacco flavoring sample were extracted by the dichloromethane. 800 μL of sample was added into a clear glass vial. 1600 μL of dichloromethane was added into the vial and was shaken for 10 min to dissolve the sample quickly. After standing for 10 min, two phases were formed in the vial. The lower layer with a volume of 800 μL was placed into another vial, and then dehydrated by the anhydrous sodium sulfate for 30 min. Finally, the dehydrated extract was sealed in an amber vial and stored in a refrigerator of 5 °C before being analyzed using GC–MS.

2.3. Instruments

A Hewlett-Packard 6890 gas chromatograph interfaced with a Hewlett-Packard mass selective detective detector 5973N (Agilent Technologies, USA) and HP Enhanced ChemStation software (D.03.00.611) were applied to analyze the samples.

2.4. Detection of volatile components

The GC–MS was used to obtain the chromatograms of the extracts. The separation was performed on a HP-5MS capillary column (60 m × 0.25 mm × 0.25 μm). The column temperature was held at 40 °C for 5 min, and then programmed from 40 °C to 150 °C at the rate of 4 °C/min, then at the rate of 6 °C/min to the final temperature of 280 °C and held for 15 min. Inlet temperature was maintained at 280 °C throughout the analysis. Helium was the carrier gas at a constant flow-rate of 1 mL/min. The injection volume was 1 μL with split ratio of 10:1. Five repeat measurements were made on a sample at the same conditions.

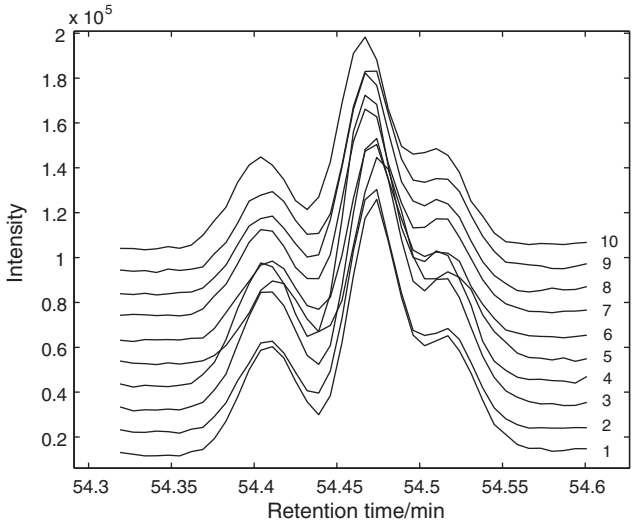

Fig. 1. The average total ion chromatography of the overlapping peaks from ten tobacco flavoring samples with different storage time. Each TIC curve was a mean of five repeat measurements. The numbers 1, 2, 3, . . ., 10 in the figure represents the corresponding storage time of 1, 2, 3, . . ., 10 months.

Electron impact (EI^+) mass spectra were recorded in the full scan mode with 0.2 s/scan velocity. The ionization energy was set at 70 eV. The mass range from 40 to 350 Da was recorded. The interface temperature was 280 °C; MS source temperature was 230 °C and the MS quadrupole temperature was 150 °C.

2.5. Data handling

The measured GC–MS data of a sample were exported into an ASCII file using the Enhanced ChemStation System. When unnecessary information in the file was deleted, a two-way evolving data matrix was left in the file. Each row of the two-way data was a mass spectrum at a retention time. The data were imported into the software platform, which was developed using C++, for primary visual inspection. The part that embraced overlapping peaks was intercepted for resolution. Detailed steps are shown in the following Section 3.1. An EFA program was coded in MATLAB 7.0. Identification of the flavor components in the overlapping peaks was performed by searching the National Institute of Standards and Technology (NIST05) and WILEY07 databases installed on the HP Enhanced ChemStation system.

3. Results and discussion

3.1. Resolution of overlapping peaks using SKSS method

The total ionic chromatogram (TIC) showed that there were about 135 peaks in the flavoring sample. Although substantial efforts were made to improve the separation, overlapping peaks were still inevitable. We took the overlapping peaks at the retention time of 54.30–54.60 min as an example to show the resolution process using the SKSS. Fig. 1 shows ten TICs of the overlapping peaks from ten flavoring samples in a storage period of ten months. To dis-

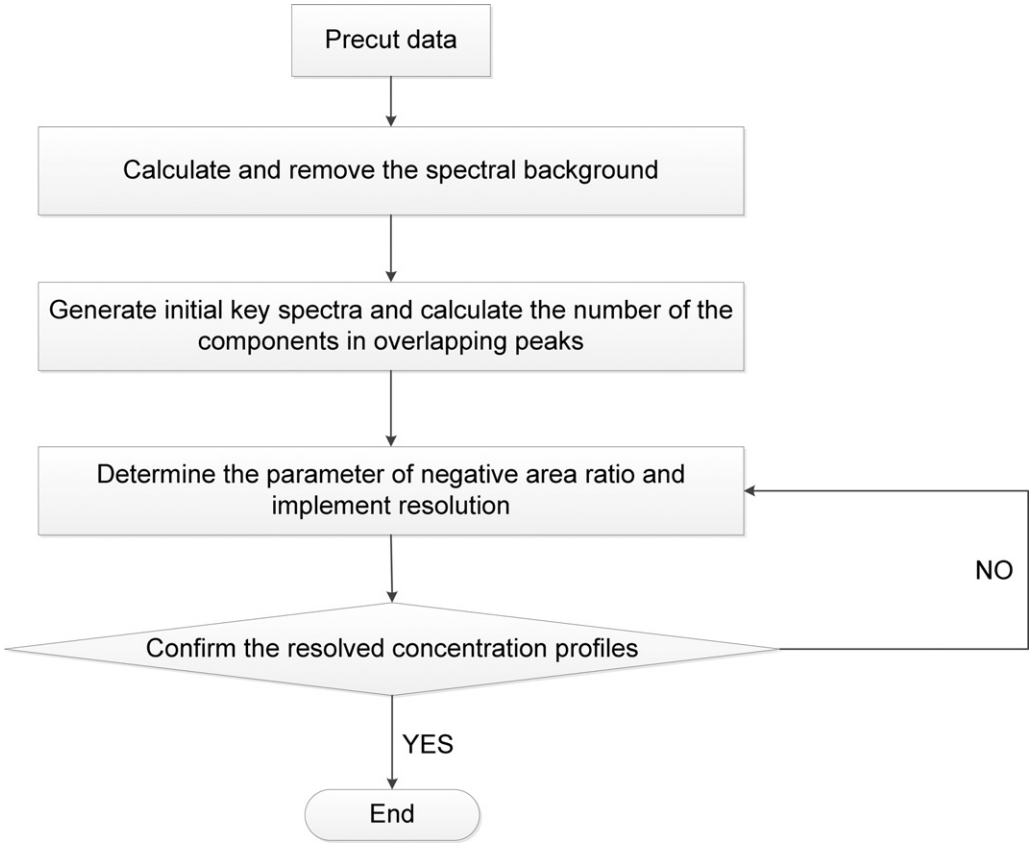

Fig. 2. Flow chart of the SKSS method on the software.

tinguish the time periods of the samples, we shifted the positions of the TICs and numbered them.

It is to be noted that there was little room for further separation of the overlapped peaks because the maximum temperature of the column was reached and other analytical conditions were in their best. We gave up the attempt in experimental approaches and resorted to MCR method. We thought that implementing a mathematical separation of the overlapping peaks as a supplement was more economical and effective.

Fig. 2 shows a flow chart of the SKSS method. It reveals that the inner mechanism of the SKSS is quite simple. The practical procedures of the SKSS executed on the software are demonstrated in Fig. 3. Three steps were involved in the procedures and they were removing background, determining the number of the components and implementing resolution. Fig. 3A shows the step for removing background. It is usually assumed that the background comes from spectral background [19–21]. To confirm the background, one should compare the correlation coefficient of the two spectra obtained before and after the overlapping peaks. If the coefficient is close to 1.0, the spectral background is confirmed and a linear fitting of the baseline is made and then removed from the two-way data. Choosing the two spectra needs to be done by users themselves. It involves repeat determination of background regions before and after the overlapping peaks and calculation of the correlation coefficient. The two spectra were obtained until the coefficient reached maximum. In this work, the background regions were the first and last five data points of the two-way data. The calculated coefficient was 0.96. Fortunately, the step is quite straight and will not cause any difficulty with the aid of the software platform.

The background corrected data are shown in Fig. 3B. A FSMWEFA plot was shown in the meantime to reveal the evolving images of the components. It also played the role of confirming the resolved concentration profiles. In the step, a set of initial key spectra was automatically generated [16]. The number of the components in the overlapping peaks was determined from the initial key spectra and the corresponding indices are shown in Fig. 3C. The maximum point in the upper plot of Fig. 3C shows that there were four components in the peaks. The determination of the component number was based on an algorithm that calculates the similarities and differences between the initial key spectra. The details of the algorithm were presented clearly in Ref. [18]. The locations of the picked key spectra were shown on the right upper site of Fig. 3C. This information was helpful for experienced users to control the resolution procedures. Under most circumstances, the SKSS works well in determining the component number automatically.

The last step implemented final resolution of the overlapping peaks. As it had been illustrated in the SKSS method, key spectra will introduce significant negative in some of the concentration profiles [16]. A parameter named negative area ratio had been defined and used to describe the area ratio of the negative part to the whole peak. The value of the ratio should be assigned carefully. Otherwise, distorted concentration profiles occur. One can evaluate the shapes of the resolved concentration profiles based on basic chromatographic theory. The negative area ratio was usually set in a range from 0.01 to 0.001. It means that the resolution error will not exceed 1.0%. In our work, the parameter was 0.01. The resolved concentration profiles and spectra are shown in Fig. 3D.

In a practical resolution process, one can try different values for the ratio and implement resolution. By comparing the resolved concentration profiles shown in Fig. 3D and the FSMWEFA plot shown in Fig. 3B, one can decide whether to keep the results or not. It is to be noted that the SKSS is relatively robust to the ratio. When a value was assigned to the ratio, it can also be applied to other data measured under the same analytical conditions of the same sample.

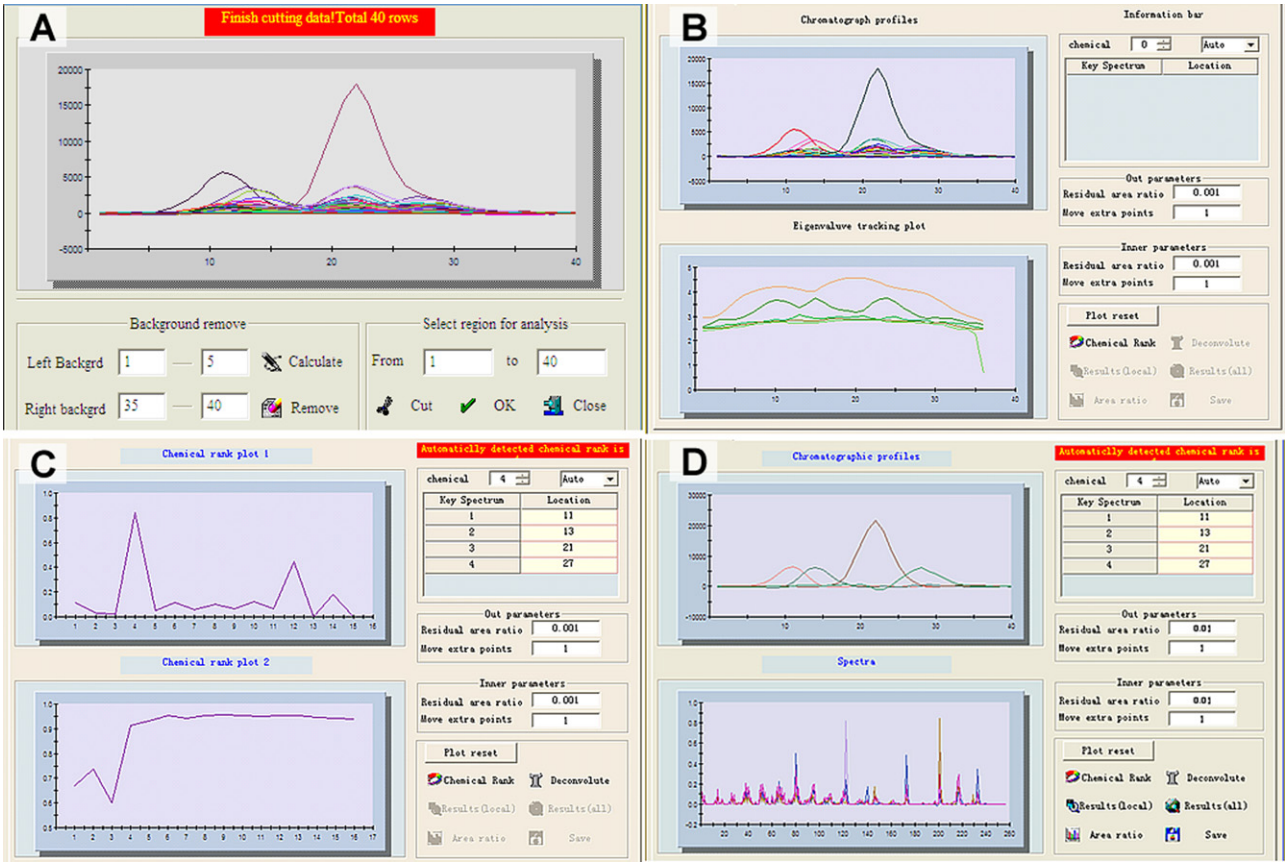

Fig. 3. Demonstration of the general procedures of the SKSS method to resolve the overlapping peaks on the software. (A) Shows the plot of the two-way data of the overlapping peaks precut from 1 to 40 rows at retention time direction. Background calculation based on the selected data was executed. The regions at first and last five data points were taken as background ranges. Background removal was implemented by clicking the bottom "remove". (B) Generates initial key spectra and calculates the number of components in the overlapped peaks by click the bottom "chemical rank". A FSMWEFA plot was shown simultaneously to show the evolving processes of the components. (C) Two chemical rank plots were shown. Four components were determined from the plots. The key spectrum locations were shown at locations 11, 13, 21 and 27 in the retention direction. (D) Determines the parameter of the residual area ratio and implements resolution. This process can be implemented several times to find the best value for the negative area ratio. The resolved concentration profiles can be compared with the FSMWEFA plot for validation.

3.2. Validation of the resolved results

In addition to the FSMWEFA plot, there are two other ways to validate the resolved results. One is to use another plot that describes the evolving process of the components such as the EFA; and the other is to resolve the data using another resolution method such as the HELP.

Although the EFA can also be applied in the resolution of overlapping chromatographic peaks [9,10], it actually provides a perfect tool to reveal the elution sequence of the components. The EFA plot is based on the forward and backward calculations of the eigenvalues of a series of sub-matrices along retention time direction. Then the appearing points and disappearing points of the components can be determined by the variations of the eigenvalues. So, the EFA is a powerful tool in validating the resolved concentration profiles.

Fig. 4 shows the resolved concentration profiles using the SKSS and EFA. From the figure, the resolved concentration profiles were in accordance with the EFA plot. We also checked the disappearing points of the resolved concentration profiles using the backward EFA plot and they confirmed well too. For the brevity of the figure, we did not show the backward EFA plot in Fig. 4. From the figure, we conclude that our resolution results were acceptable.

We did not embed the EFA plot in our software platform for two reasons. First, the calculation of EFA is time consuming with the increase of the data size. Second, the rank deficiency caused by spectral similarity of the components will distort the shape of the EFA. The FSMWEFA avoids above problems. It calculates a small data set of fixed size along the retention time direction. It can also reveal the elution information of the components.

A better explanation of the FSMWEFA is shown in Fig. 5. The FSMWEFA plot is also based on the calculations of the eigenvalues of a series of fixed size sub-matrices along the retention time direction. The eigenvalues show the appearing points and disappearing points of the components. Thus, it can also play the role of validating the resolution results. However, as the calculation of FSMWEFA makes a shift in retention time direction, adjusting the real appearing points and disappearing points is needed by users. A good strategy is to combine both the EFA and FSMWEFA to reveal the evolving information of the components.

In the meantime, we also resolved the overlapping peaks using HELP method because it is a part of our software platform. The concentration profiles calculated by the two methods were the same on the whole. However, the HELP method was apparently a time consuming one. There were totally 16 parameters to be inputted in the whole resolution process of the two-way data, four for each component. So, the HELP method might not be a good choice if there is huge data set to be analyzed.

The identification of the components from the overlapped peaks was a problem because the resolved spectra could not be put into ChemStation system for library search. However, this

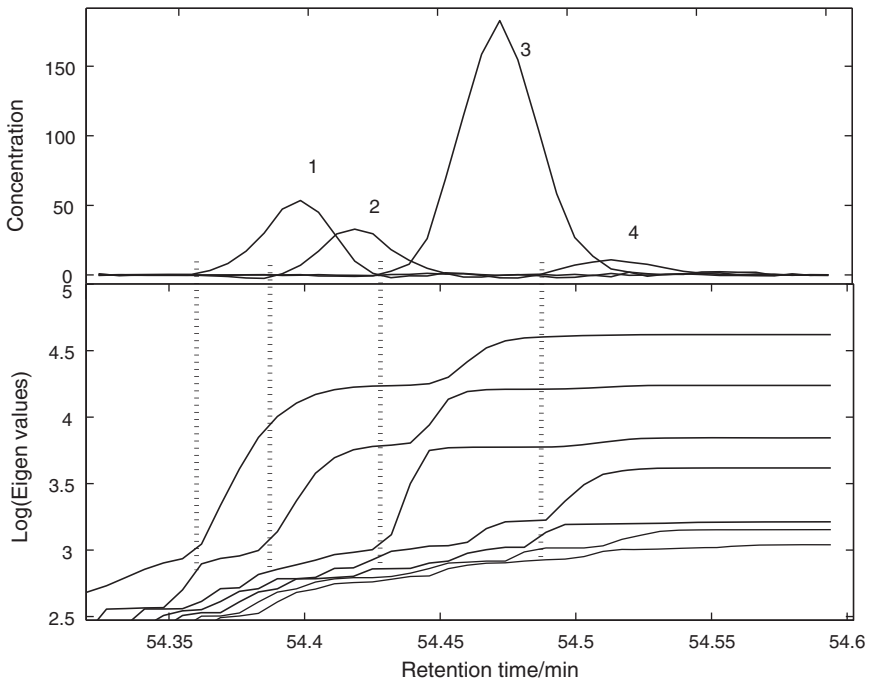

Fig. 4. The resolved concentration profiles and the forward EFA plot of the overlapping peaks. Upper part: the resolved concentration profiles of the four components using the SKSS. The numbers of 1, 2, 3 and 4 represent the eluting order of the resolved components. Lower part: the forward EFA plot indicates the elution of the components along the retention time. The dashed straight lines mark the appearing points of the components and are at 54.36 min, 54.38 min, 54.43 min and 54.49 min, respectively.

problem was partly solved by using the resolved concentration profiles. As shown in Fig. 4, the pure component regions were obtained from this plot. For example, the pure component region of the first component was 54.35–54.39 min. Thus, identification of the component could be completed by picking a mass spectrum at 54.39 min and then searching it in the NIST05 and WILEY07 database. Library search results showed that the first component was methyl 8.alpha.-isopimaran- 18-oate. Similarly, mass spectra were picked at 54.43 min, 54.47 min and 54.53 min for other three components, respectively. They turned out to be abietic acid, dihydro-, methyl ester, methyl pimar-7-en-18-oate and methyl pimar-7-e, respectively. These substances were the main aroma constitutes of a special essence.

It must be noted that library search usually gives tentative results. Further identification of the components using retention induces or standards is preferred. However, as the components do not escape our ken on the added flavoring materials, further identification of the components was not continued.

In summary, our resolved results can be validated by other methods such as EFA, FSMWEFA and HELP method. It is practicable to implement library search using the mass spectra obtained in pure component regions.

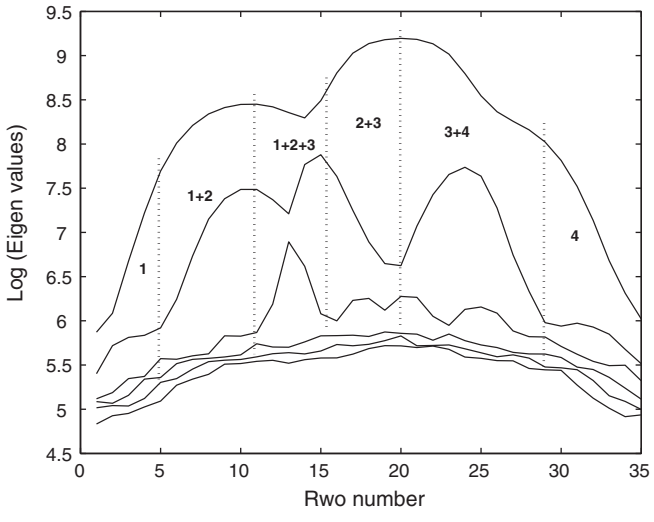

Fig. 5. FSMWEFA plot for the resolved overlapping peaks. The numbers 1, 2, 3 and 4 represent the eluting order of the four components, respectively. Dashed lines divided the eluting regions of the four components and the "+" shows overlapping regions of two or more components. The regions, 1, 1 + 2, 1 + 2 + 3, 2 + 3, 3 + 4 and 4, indicated the region of the first pure component, the overlapping region of the first and second components, the overlapping region of the first, second and third components, the overlapping region of the second and third components, the overlapping region of the third and fourth components, and the region of the fourth pure component, respectively.

Table 1
Resolved percentage peak area ratios (%) of the four components from ten batch samples.

Samples	Component 1	Component 2	Component 3	Component 4
A1	18.2	10.1	67.9	3.89
A2	17.5	11.5	67.4	3.57
A3	18.1	11.1	66.7	4.17
A4	17.0	12.3	66.6	4.08
A5	18.0	9.89	68.3	3.84
A6	17.7	10.3	68.1	3.93
A7	16.8	11.2	68.2	3.84
A8	16.0	11.9	68.2	3.97
A9	18.4	10.0	67.2	4.40
A10	18.1	11.4	66.4	4.10
Mean	17.6	11.0	67.5	3.98
RSD (%)	4.4	7.7	1.1	5.7

A1 to A10 represent ten flavoring samples from formula A in a period of ten months.
Component 1 is methyl 8.alpha.-isopimaran- 18-oate.
Component 2 is abietic acid, dihydro-, methyl ester.
Component 3 is methyl pimar-7-en-18-oate.
Component 4 is methyl pimar-7-e.
RSD is the relative standard deviation.

3.3. Batch analysis of overlapping peaks

Fig. 1 shows the overlapping parts of the average TICs from ten tobacco flavoring samples stored in a period of ten months. Our purpose was to investigate the change in their concentration. However, calculating the peak areas of the components is usually impossible because of the overlapping peaks. The SKSS solved this problem easily. The peak area ratios of the resolved components in ten storage periods were calculated and are shown in Table 1. From the table, we can make a clear investigation of the peaks areas variation of the four components in ten storage periods. For example, the peaks areas of the components 1 and 3 showed relative standard deviations (RSD%) below 5.0%. It meant that these two components were relatively stable in the storage periods. Component 2 and component 4 showed a higher RSD% than 5.0%. The highest RSD% of component 2 means that it could be one of the key components and more attention should be paid to monitor the concentration change of the component in the analysis work.

4. Conclusion

Many multivariate curve resolution methods have been proved to be valid in the analysis of chromatographic data. However, the complicated theories of these methods preclude them from being routine tools because well-trained chemometrics operators were needed. The SKSS method showed its advantage in coping with this problem. There is still a need for further study on how to reduce the users' involvement in resolution process. Finding an easier way to evaluate the validity of the resolution results is another topic that needs future study.

Acknowledgements

Financial support for this work was provided by China Tobacco Guangdong Industrial Limited Corporation (No. I05XMQK[2008]017), National Natural Science Foundation of China (No. 20875106), and Guangdong Natural Science Found Committee (No. 9151027501000003).

References

[1] R.R. Baker, J.R. Pereira da Silva, G. Smith, Food Chem. Toxicol. 42 (2004) 39.
[2] J.J. Yu, Cigarette Technology, Chinese Agriculture Press, Beijing, China, 2003.
[3] S.B. Stanfill, D.L. Ashley, J. Chromatogr. A 858 (1999) 79.
[4] X. Zhu, Y. Gao, Z. Chen, Q. Su, Chromatographia 69 (2009) 735.
[5] T.J. Clark, J.E. Bunch, J. Agric. Food Chem. 45 (1997) 844.
[6] S.M. Rocha, E. Coelho, J. Zrostlíková, I. Delgadillo, M.A. Coimbra, J. Chromatogr. A 1161 (2007) 292.
[7] S. Zhu, X. Lu, K. Ji, K. Guo, Y. Li, C. Wu, G. Xu, Anal. Chim. Acta 597 (2007) 340.
[8] P.J. Marriott, R. Shellie, C. Cornwell, J. Chromatogr. A 936 (2001) 1.
[9] M. Maeder, Anal. Chem. 59 (1987) 527.
[10] H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbuhler, Talanta 33 (1986) 943.
[11] E.R. Malinowski, J. Chemom. 6 (1992) 29.
[12] O.M. Kvalheim, Y.Z. Liang, Anal. Chem. 64 (1992) 936.
[13] Y.Z. Liang, O.M. Kvalheim, H.R. Keller, D.L. Massart, P. Kiechle, F. Erni, Anal. Chem. 64 (1992) 946.
[14] H.R. Keller, D.L. Massart, Anal. Chem. Acta 246 (1991) 379.
[15] Y. Liang, O. Kvalheim, Chemom. Intell. Lab. Syst. 20 (1993) 115.
[16] E.R. Malinowski, J. Chemom. 10 (1996) 273.
[17] R. Manne, B.V. Grande, Chemom. Intell. Lab. Syst. 50 (2000) 35.
[18] F. Gan, Q.S. Xu, Y.Z. Liang, Analyst 126 (2001) 161.
[19] Y.Z. Liang, O.M. Kvalheim, Fresenius J. Anal. Chem. 370 (2009) 694.
[20] F. Gong, Y.Z. Liang, H. Cui, F.T. Chau, B.T.P. Chan, J. Chromatogr. A 909 (2001) 237.
[21] H.L. Shen, R. Manne, Q.S. Xu, D.Z. Chen, Y.Z. Liang, Chemom. Intell. Lab. Syst. 45 (1999) 323.