Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1029 (2004) 173–183
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Correction of retention time shifts for chromatographic fingerprints of herbal medicines

Fan Gong [a,b], Yi-Zeng Liang [a,*], Ying-Sing Fung [b], Foo-Tim Chau [c]

[a] *Research Center of Modernization of Chinese Herbal Medicines, Institute of Chemometrics & Intelligent Analytical Instruments, College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, China*
[b] *Department of Chemistry, Hong Kong University, Pokfulam Road, Hong Kong, China*
[c] *Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China*

Received 10 July 2003; received in revised form 11 December 2003; accepted 17 December 2003

Abstract

In this study, the combination of chemometric resolution and cubic spline data interpolation was investigated as a method to correct the retention time shifts for chromatographic fingerprints of herbal medicines obtained by high-performance liquid chromatography–diode array detection (HPLC–DAD). With the help of the resolution approaches in chemometrics, it was easy to identify the purity of chromatographic peak clusters and then resolve the two-dimensional response matrix into chromatograms and spectra of pure chemical components so as to select multiple mark compounds involved in chromatographic fingerprints. With these mark components determined, the retention time shifts of chromatographic fingerprints might be then corrected effectively. After this correction, the cubic spline interpolation technique was then used to reconstruct new chromatographic fingerprints. The results in this work showed that, the purity identification of the chromatographic peak clusters together with the resolution of overlapping peaks into pure chromatograms and spectra by means of chemometric approaches could provide the sufficient chromatographic and spectral information for selecting multiple mark compounds to correct the retention time shifts. The cubic spline data interpolation technique was user-friendly to the reconstruction of new chromatographic fingerprints with correction. The successful application to the simulated and real chromatographic fingerprints of two *Cortex cinnamomi*, fifty *Rhizoma chuanxiong*, ten *Radix angelicae* and seventeen *Herba menthae* samples from different sources demonstrated the reliability and applicability of the approach investigated in this work. Pattern recognition based on principal component analysis for identifying inhomogenity in chromatographic fingerprints from real herbal medicines could further interpret it.

Keywords: Fingerprinting; Retention time shift; Chemometrics; Spline interpolation; Pharmaceutical analysis

1. Introduction

Nowadays, the construction of chromatographic fingerprints becomes one of the most powerful approaches for quality control of herbal medicines and others [1–4]. A chromatographic fingerprint is, in practice, a chromatographic profile representing multiple chemical components with characteristics from herbal medicines investigated so as to identify and assess the stability of the chemical constituents observed by chromatography [4–6]. However, since the herbal medicines with a large number of chemical components are very complex and the chromatographic instruments and experimental conditions are difficult to be reproducible, the construction of chromatographic fingerprints of herbal medicines for quality control is not a trivial work in fact.

Previously, chromatographic fingerprints were generally constructed with the retention time-peak area data matrices including only selected peaks whether the identity of the peaks were known or not. However, as pointed out by Nielsen et al., “It can be very difficult to select an optimal set integration parameters for chromatograms obtained from analysis of complex samples which easily contain more than 100 peaks. Furthermore, the selection and extraction of peaks to include in data analysis is difficult, partly subjective and a large amount of the data in the chromatograms are discarded. The disadvantages of peak detection and integration, and of the introduction of a subjective peak

* Corresponding author. Tel.: +86-731-883-0831; fax: +86-731-882-5637.
E-mail address: yzliang@public.cs.hn.cn (Y.-Z. Liang).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2003.12.049

selection can be avoided by using all collected data points in the chemometric analysis. A further advantage is that the peak shape can be included in data analysis." [7]. Thus, chromatographic fingerprints of herbal medicines for quality control are constructed with the entire chromatographic data containing the complete qualitative and quantitative information but not the selected retention time-peak area data in China [4–6]. However, when one deals with several chromatographic fingerprints jointly with all data points obtained, some types of variation sources are inevitably encountered from one chromatogram to another [7–13]. Under this situation, some unacceptable results, one of which is imposed by the retention time shifts, will be produced. Unfortunately, how the retention time shifts are caused by the variation sources is very complex. It might be due to (1) the degradation of the stationary phase, especially, the low stability of silica and silica-based supports at high pH values and the collapse of C_{18} bonded phase because of a highly polar mobile phase; (2) minor changes in mobile phase composition caused by temperature and pressure fluctuations, variations in flow-rate and gradient dispersion; (3) some problems involved in the detectors, for example, a wavelength shift in the UV spectrometer, a spectroscopic intensity variation and the misalignment of the monochromator; (4) the column overloading on account of the great injected amount or some components with high concentration; (5) the possible interaction between analytes; (6) other unknown shifts in the instrument. If these disadvantageous cases are in existence during the chromatographic runs, the retention times will be subsequently shifted.

When such retention time shifts occur in chromatographic fingerprints, it is great difficult to conduct data processing, for example, the construction of common chromatographic models of all samples investigated, the similarity comparison between chromatographic fingerprints and pattern recognition based on principal component analysis (PCA) since multivariate analysis with entire chromatographic profiles as input data is very sensitive to even minute variations [11]. As a result, in order to make up a consistent data, it is urgently necessary to detect the retention time shifts of chromatographic fingerprints and then the chromatographic profiles should be adjusted along the retention time direction by means of synchronizing the retention times of the chromatographic peaks from the same components.

During the past decades, several kinds of useful approaches have been developed for peak synchronization in chromatographic profiles [1,7–13]. Some of them corrected the retention time shifts by making internal standards added or marker peaks coincide in all chromatograms under study [1,8–10]. Among them, the technique in Ref. [10] was recently developed. At its first step, chromatographic peaks were identified by setting a retention time window in both of the sample and target chromatograms, and then a list of their retention times was generated. Here, the window meant the maximum shift to be considered. Clearly, the choice of the window size was critical to this technique.

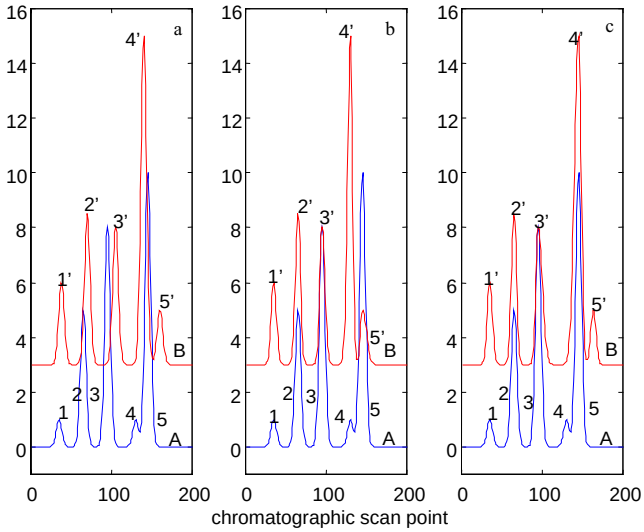

Fig. 1. An example elucidating the relationship between the correlation coefficient and the concentration profiles of chemical components.

However, if the retention time shifts are very serious for complex systems like herbal medicines, the selection of the optimal peak-matching windows might not be a trivial task. In Refs. [7,11–13], some objective functions on the correlation between the target and sample chromatograms were optimized and then the sample chromatographic profiles were aligned with the target. In general, the methods must be very efficient and elegant if the samples investigated are quite similar in the concentration profiles of the chemical components. However, if the concentration profiles change greatly for the complex samples such as herbal medicines from the different producing places and/or from the various harvest seasons, wrong results might be obtained by simply seeking the optimal correlation coefficient between the chromatograms. It is because that the correlation coefficient is influenced greatly by the big peaks in the chromatographic profiles. In this case, the maximal correlation coefficient does not certainly represent the best correction for the retention time shifts. Here, Fig. 1 shows such an example, in which the components 1′, 2′, 3′, 4′ and 5′ in the sample chromatogram (B) correspond to the components 1, 2, 3, 4 and 5 in the target (A), respectively. If the alignment is conducted simply based upon the maximal correlation coefficient between A and B, the 4′ and 5 peaks will be synchronized incorrectly as shown in Fig. 1c with the correlation coefficient of 0.7672. However, if the data is generated from hyphenated instruments like high-performance liquid chromatography–diode array detection (HPLC–DAD), the spectral information will be very useful for featuring the chromatographic peaks. With the help of the spectral information, Fig. 1b with the correlation coefficient of 0.5753 is obtained (The 4′ and 4, 5′ and 5 peaks are synchronized, respectively). Thus, it seems necessary to develop some novel procedures to correct the retention time shifts by making full use of the chromato-

graphic and spectral information provided by hyphenated chromatography.

In this study, chemical components from herbal medicines are qualitatively and quantitatively determined with HPLC–DAD. The combined approach of chemometric resolution with cubic spline interpolation was then applied to correct the retention time shifts for chromatographic fingerprints obtained. First, the purity of some chromatographic peak clusters with high signal to noise is identified and the two-dimensional response matrices are resolved into pure chromatograms and spectra of chemical components with the resolution methods in chemometrics. Next, multiple mark compounds involved in chromatographic fingerprints are selected based on their chromatographic profiles and spectra obtained. With these mark components determined, the retention time shifts are corrected with linear interpolation and new chromatographic fingerprints are then reconstructed by use of the cubic spline data interpolation technique. The proposed approaches in this work are successfully applied to both the simulated and real chromatographic fingerprints of herbal medicines covering two *Cortex cinnamomi*, fifty *Rhizoma chuanxiong*, ten *Radix angelicae* and seventeen *Herba menthae* samples from different sources. In order to further interpret the reliability and applicability of the present approach, pattern recognition based on PCA is used to identify inhomogenity in the chromatographic fingerprints from these real herbal medicines.

2. Theory

2.1. Selection of mark compounds with chemometric resolution methods

Fig. 2a shows a simulated chromatographic fingerprint (Fingerprint1) with ten peaks generated by HPLC–DAD. Here, Fingerprint1 is the total chromatogram whose each point is the summed UV intensity of the spectrum at that chromatographic time. The mesh graph of its corresponding two-dimensional data matrix is shown in Fig. 2b. Fig. 2c and d represent the pure chromatograms and spectra, respectively. If the response matrix is denoted by $\boldsymbol{A}$, then:

$$\boldsymbol{A}_{m\times n} = \boldsymbol{C}\boldsymbol{S}' + \boldsymbol{E} = \sum \boldsymbol{c}_k \boldsymbol{s}'_k + \boldsymbol{E} \quad (k = 1, 2, \ldots, N) \quad (1)$$

or

$$\boldsymbol{A}_{m\times n} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_i, \ldots, \boldsymbol{a}_n] \quad (i = 1, 2, \ldots, n) \quad (2)$$

where $\boldsymbol{A}_{m\times n}$ denotes an UV absorbance matrix with N components of m chromatographic points at n wavelength points ($N = 10$, $m = 450$ and $n = 100$ here). $\boldsymbol{C}$ (see Fig. 2c) and $\boldsymbol{S}$ (see Fig. 2d) are the pure chromatographic and spectral matrices, respectively. $\boldsymbol{E}$ represents the noise. $\boldsymbol{S}'$ is the transpose of the matrix $\boldsymbol{S}$. $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_i, \ldots, \boldsymbol{a}_n$ are column vectors of $\boldsymbol{A}_{m\times n}$. According to Eqs. (1) and (2), each column of $\boldsymbol{A}_{m\times n}$ represents a chromatographic fingerprint

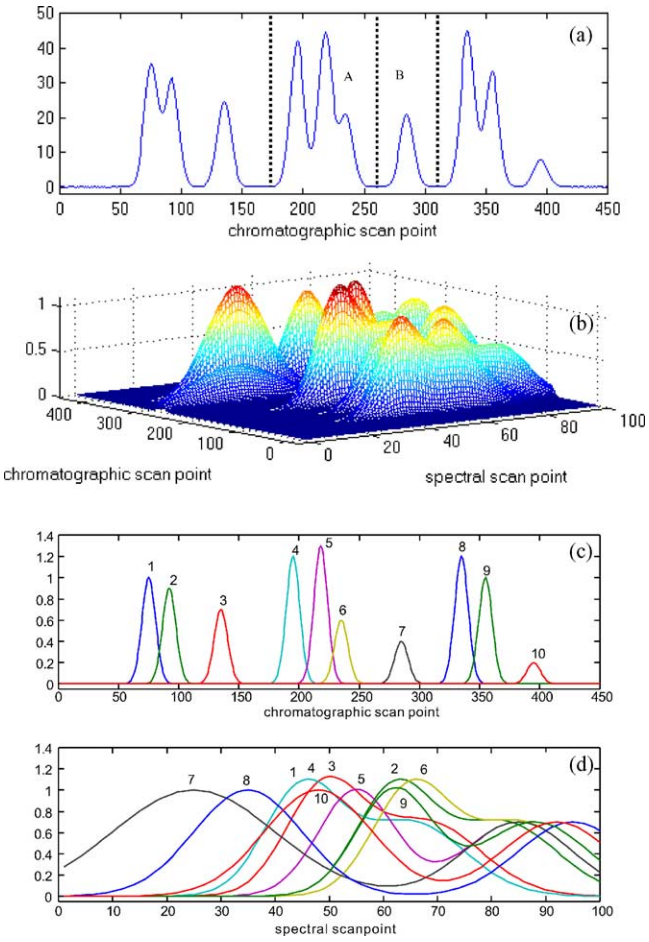

Fig. 2. Simulated Fingerprint1 (a), mesh graph (b), pure chromatograms (c) and UV spectra (d).

at one single wavelength. The total chromatogram ($\boldsymbol{p}$) can be expressed with the following Eq.:

$$\boldsymbol{p} = \sum \boldsymbol{a}_i = \boldsymbol{a}_1 + \boldsymbol{a}_2 + \cdots + \boldsymbol{a}_i + \cdots + \boldsymbol{a}_n \quad (i = 1, 2, \ldots, n) \quad (3)$$

Since the total chromatogram is generated from the two-dimensional data martrix, several useful chemometric resolution techniques, for example, heuristic evolving latent projections (HELP), evolving factor analysis (EFA), windows factor analysis (WFA), subwindow factor analysis (SFA) and orthogonal projection resolution (OPR), can be used to identify the purity of chromatographic peaks and resolve the data matrix into chromatograms and spectra of the pure chemical constituents [14–22]. As the details of the approaches could be found in Refs. [14–22], only a brief description is given in this study. Here, the peak clusters denoted by A and B in Fig. 2a are taken as the examples.

With the help of the fixed size moving window evolving factor analysis (FSWMEFA) or so-called eigenstructure tracking analysis in chemometrics [14–22], the fixed size window method (FSWM) plots of the A and B clusters can be obtained (See Fig. 3). In the FSWM plot, the noise

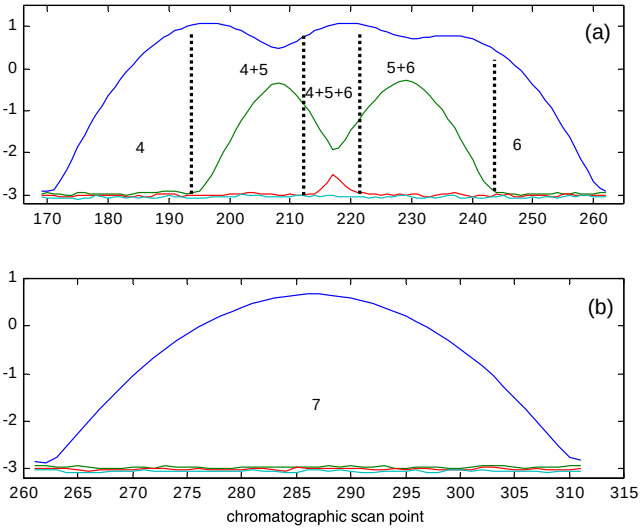

Fig. 3. FSWM plots of A (a) and B (b) peak clusters.

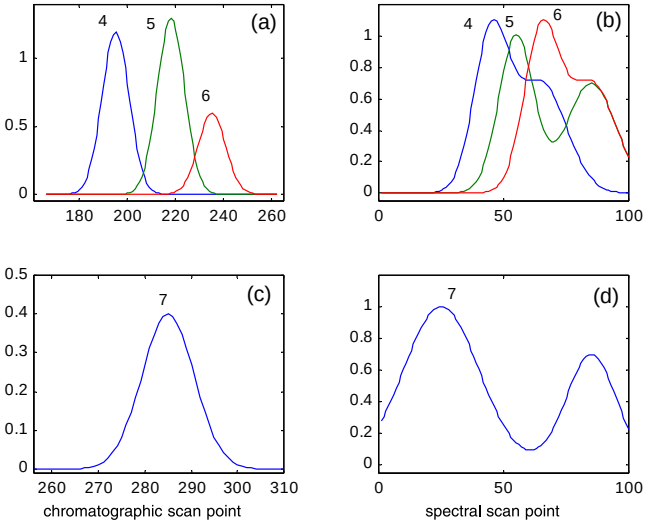

Fig. 4. Resolved chromatograms (a and c) and spectra (b and d) of A and B peak clusters.

level is characterized by eigenvalue curves which have similar numerical values and appears together at the bottom. Eigenvalue curves higher than the noise level represent the presence of new components. If a chemical system investigated contains only one component, only one eigenvalue curve higher than the noise level appears itself in the FSWM plot. Seen from Fig. 3, there are three and one eigenvalue curve(s) higher than the noise level in Fig. 3a and b, respectively. Thus, A is an overlapping peak cluster with three components and B is just a pure one.

On the other hand, the selective region and zero-concentration region of each component involved could be also determined with FSWM plots. In Fig. 3, the region i indicates the elution region of the ith pure component while the regions $i + j$ and/or $i + j + k$ represent the overlapping region of the ith, jth and/or kth components ($i, j, k = 4, 5, \ldots, 7$). With the above-obtained selective information and zero-concentration regions of A and B peak clusters, the corresponding two-dimensional data matrix might be easily resolved into pure chromatograms and UV spectra with local full rank analysis [14–23]. Fig. 4 represents the resolved chromatographic profiles and spectra. The top points of the pure chromatographic profiles resolved are taken as the retention time points. Here, they are 195, 218, 235 and 285 for the 4th to 7th components, respectively.

Fig. 5a is another simulated chromatographic fingerprint (Fingerprint2) generated by HPLC–DAD. Similar to Fingerprint1, it is also the total chromatogram from a two-dimensional data matrix. The mesh plot, the pure chromatograms and spectra are shown in Fig. 5b–d, respectively. Likely, the purity of all peak clusters is identified at first and then the response data matrix is resolved into pure chromatograms and UV spectra of all components involved. In comparison with the chromatographic profiles and spectra obtained, it is interesting that seven components are the same but the others are different in Fingerprint1 and

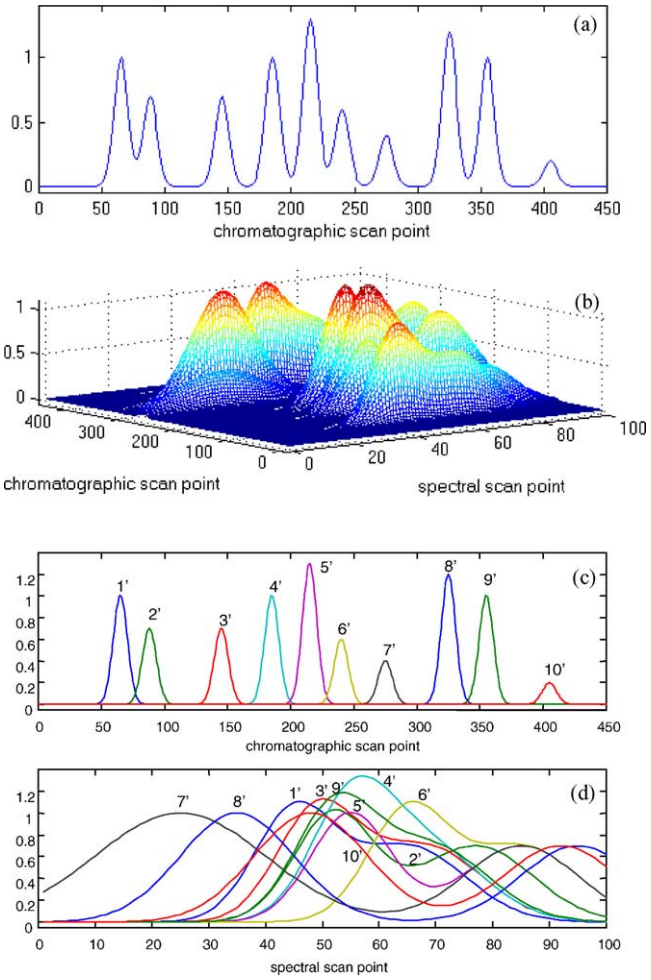

Fig. 5. Simulated Fingerprint2 (a), mesh graph (b), pure chromatograms (c) and UV spectra (d).

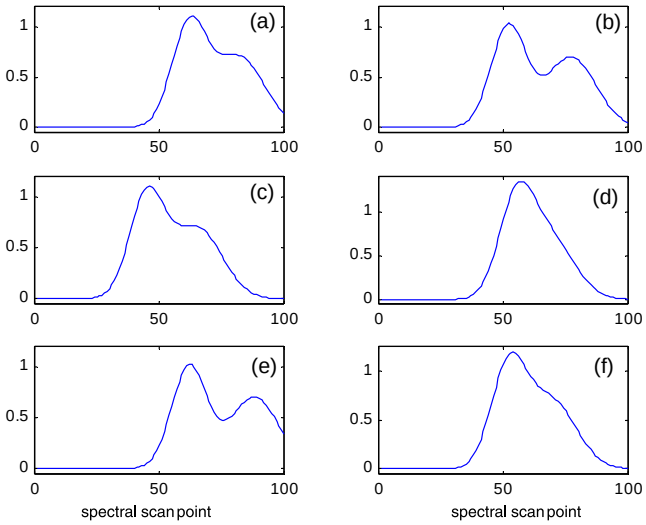

Fig. 6. UV spectra of 2 (a), 4 (c), 9 (e) components in Fig. 1c, 2′ (b), 4′ (d) and 9′ (f) components in Fig. 4c.

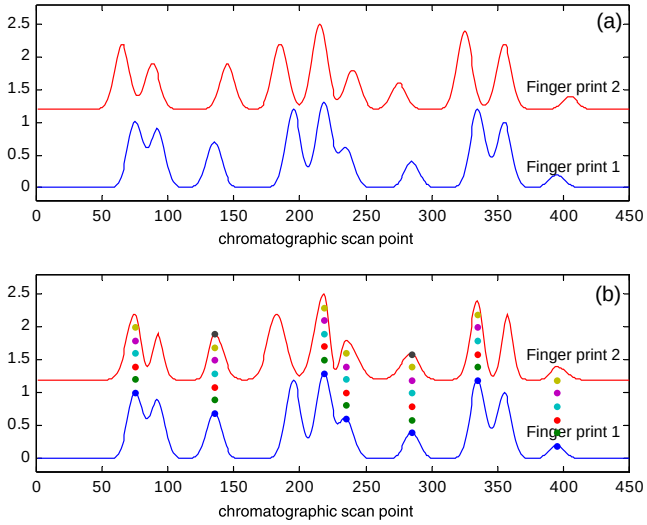

Fig. 7. Fingerprint1 and Fingerprint2 before (a) and after (b) correcting the retention time shifts.

Fingerprint2 (2, 4 and 9 components in Fig. 2c are different from 2′, 4′ and 9′ components in Fig. 5c. Their UV spectra are shown in Fig. 6). On the other hand, the retention time points are not identical even for the same components involved in Fingerprint1 and Fingerprint2 (see Table 1) suggesting the existence of the retention time shifts. As a result, in order to make up a consistent data set from Fingerprint1 and Fingerprint2, the retention time shifts should be corrected effectively.

From above, with the help of chemometric resolution methods, the chromatograms and spectra of all pure components in Fingerprint1 and Fingerprint2 could be obtained easily. Since seven of the components are involved in both fingerprints under study, they might be then selected as the mark compounds for correcting the retention time shifts. In general, the markers selected with high signal to noise should span within the whole chromatographic scan region but not concentrate on parts of the region. Moreover, if a component showing a distinct peak exists in all the chromatographic fingerprints investigated, it is reasonable to select it as one of the markers.

2.2. Correction of retention time shifts and reconstruction of chromatographic fingerprints with cubic spline interpolation

After the marker compounds and their retention time points are determined with the resolved chromatograms and spectra above obtained, the retention time shifts might be corrected now. If Fingerprint1 is taken as the target and Fingerprint2 is then aligned with it, the retention time points of the mark compounds in Fingerprint1 might be selected as the fix points and the time shifts for all points between these fix points in Fingerprint2 might be calculated by linear interpolation. Thus, new chromatographic scan points of Fingerprint2 with correction could be obtained.

Since the new chromatographic scan points of Fingerprint2 are calculated with linear interpolation, these new points might be not integer. For example, the 68th point of Fingerprint2 is adjusted to 77.25 not 77 or 78. Clearly, the non-integer should be converted into the integral data as the chromatographic scan points are integer. In this study, in order to conduct this conversion, the cubic spline data interpolation technique is used to reconstruct Fingerprint2. The new one obtained is taken as the real chromatographic fingerprint of Fingerprint2 after correcting the retention time shifts. Fig. 7 shows Fingerprint1 and Fingerprint2 with and without correction. Seen from Fig. 7b, the marker components selected from Fingerprint1 and Fingerprint2 have been synchronized after Fingerprint2 is aligned with Fingerprint1.

In sum, the steps of the present method could be described as the following:

(1) Select some peak clusters with high signal to noise, identify the purity and then resolve the two-dimensional data matrix into pure chromatograms $\boldsymbol{C}$ and spectra $\boldsymbol{S}$.

Table 1
Retention time points of 10 chemical components in chromatographic fingerprints simulated

	Retention time point									
Fingerprint1	75	92[a]	135	195[a]	218	235	285	335	355[a]	395
Fingerprint2	65	88[a]	145	185[a]	215	240	275	325	355[a]	405

[a] Different components between Fingerprint1 and Fingerprint2.

(2) In comparison with $\boldsymbol{C}$ and $\boldsymbol{S}$, select multiple marker compounds. Let $\boldsymbol{t}$**_mark** and ${}^{l}\boldsymbol{t}$**_mark** are the retention time points of the markers in the target and the lth sample fingerprint, respectively.

$$\begin{aligned} \boldsymbol{t}\text{**_mark**} &= [t_{1,1}t_{1,2}\dots t_{1,i}\dots t_{1,nn}] \quad (i = 1, 2, \dots, nn.\ \text{Here},\ nn = 7) \\ {}^{l}\boldsymbol{t}\text{**_mark**} &= [t_{l,1}t_{l,2}\dots t_{l,i}\dots t_{l,nn}] \quad (l = 1, 2, \dots, mm\ \text{and}\ i = 1, 2, \dots, nn.\ \text{Here},\ mm = 2\ \text{and}\ nn = 7) \end{aligned} \tag{4}$$

(3) With $\boldsymbol{t}$**_mark** and ${}^{l}\boldsymbol{t}$**_mark**, linear interpolation is conducted on the lth sample fingerprint to obtain the new chromatographic scan points ${}^{l}\boldsymbol{t}$**_new**. If its original scan points are denoted by ${}^{l}\boldsymbol{t}$**_orig** and $t_{l,i} \geq {}^{l}\boldsymbol{t}$**_orig** $(k) > t_{l,i-1}$, then:

$${}^{l}\boldsymbol{t}\text{**_new**}\,(k) = ({}^{l}\boldsymbol{t}\text{**_orig**}\,(k) - t_{l,i-1}) \times \frac{t_{1,i} - t_{1,i-1}}{t_{l,i} - t_{l,i-1}} + t_{1,i-1} \quad (k = 1, 2, \dots, m.\ \text{Here},\ m = 450) \tag{5}$$

However, if ${}^{l}\boldsymbol{t}$**_orig** $(k) \leq t_{l,1}$ or ${}^{l}\boldsymbol{t}$**_orig** $(k) \geq t_{l,nn}$, linear interpolation is conducted within the regions between the starting scan point and $t_{l,1}$ or $t_{l,nn}$ and the ending scan point, respectively. Here, the starting and ending scan points are 1 and 450, respectively.

(4) Reconstruct a new chromatographic fingerprint (${}^{l}\boldsymbol{p}$**_new**) with ${}^{l}\boldsymbol{t}$**_new** by means of the cubic spline data interpolation technique.

3. Experimental

3.1. Simulated chromatographic fingerprint

For the simulated two-dimensional data matrices, each chromatographic peak (450 points) and UV spectrum (100 points) are constructed with one and two Gauss peak(s), respectively. The chromatographic fingerprint is the total chromatogram whose each point is the summed UV intensity of the spectrum at that chromatographic time.

3.2. Materials and reagents

Cortex cinnamomi, *Rhizoma chuanxiong*, *Radix angelicae* and *Herba menthae* samples were purchased from several pharmaceutical stores, companies and collected from different producing areas in the mainland and in Hong Kong, PR China.

CH_3OH, H_3PO_4, K_2HPO_4, CH_3CN, isoproanol (C_3H_8O) and citric acid ($C_6H_8O_7$) were of analytical grade. Double distilled water was used. Standard compounds of cinnamyl alcohol, cinnamaldehyde, cinnamyl acetate and ferulic acid are purchased from Lancaster Synthesis Ltd. (UK) Company, Sigma, Aldrich and National Institute for the Control of Pharmaceutical and Biological Products of China.

3.3. Instruments

RT-80 pulverizer made in Taiwan, 5810 centrifuge from Eppendorf Company in Germany, Ultra Turrax T25 basic stirrer from IKA Company in Malaysia, CQ250 ultrasonic cleaner made in Shanghai, Hewlett-Packard HP-1100 HPLC coupled with a G1315A diode array detector and Hewlett-Packard 5890 Series II gas chromatograph paired with a 5972 Series mass-selective detector were employed in this study.

3.4. Extraction of herbal medicines

All of the herbal materials were dried for about 60 min under 30 °C at first. Then, about 0.5 g amount of dried and pre-pulverized herbal materials were extracted using an Ultra Turrax T25basic stirrer (11 000 rpm) with 30 ml of CH_3OH for 2 min. After centrifugation for about 20 min, the upper solution was filtered through a glass filter covered with a filter paper. Next, the solution was evaporated under reduced pressure to about 1 ml and then diluted with methanol to 5 ml in a volumetric flask. A 1 ml volume of this solution was then filtered through a Millipore filtration unit type HV 0.45 μm. A 20 and 2.5 μl volume of this solution were injected into the HPLC and GC–MS systems, respectively.

3.5. Chromatographic procedure

3.5.1. Determination of herbal medicines with HPLC–DAD

(1) Column: LiChrosorb RP-18 (Hewlett-Packard, 200 mm × 4.6 mm i.d.).
(2) Mobile phase: At the starting time, the mobile phase was composed of A (CH_3OH) and B ($H_2O + K_2HPO_4 + H_3PO_4$ with pH ≈ 3) in the ratio of 10:90 (v/v). Then, this mixed mobile phase was linearly gradient to A–B (100:0, v/v) after 60 min.
(3) Flow rate: 1.0 ml/min.
(4) Column temperature: 25 °C.
(5) Wavelength scanning range and step: 190–400 nm, 2 nm per step.

3.5.2. Determination of herbal medicines with GC–MS

(1) Column: HP-5MS column (Crosslinked 5% PH MS siloxane, 30 m × 0.25 mm i.d., 0.25 μm film thickness).
(2) Column temperature: maintained at 80 °C for 2 min at first, and then programmed from 80 to 230 °C at the rate 5 °C/min.
(3) Inlet temperature: 230 °C.
(4) Carrier gas and flow-rate: helium, 1.0 ml/min.

(5) Ionization mode and energy: electron impact (EI^+), 70 eV.
(6) Scan range and velocity: 30–400 μ, 2.025 scan/s.
(7) Ionization source temperature: 280 °C.

3.6. Data analysis

Data analyses were performed on a Pentium-based IBM-compatible personal computer. All programs were coded in MATLAB 5.1 for windows.

4. Results and discussion

4.1. Selection of mark compounds in chromatographic fingerprints of real herbal medicines with chemometric resolution methods

Fig. 8 shows real chromatographic fingerprints (Cfingerprin1 and Cfingerprin2) of *Cortex cinnamomi* from two different sources. Similar to the simulated data, Cfingerprin1 and Cfingerprin2 are also the total chromatograms from the response data matrices. Here, the peak clusters denoted by C, D and E in Fig. 8a are taken as examples to explain how to select the mark compounds in real chromatographic fingerprints with chemometric resolution approaches.

Fig. 9 represents C, D and E peak clusters. Their FSWM plots are shown in Fig. 10, Seen from Fig. 10a and c directly, two components existing in C while there is only one in E. However, there are possibly two components in D from Fig. 10b. In fact, D is also a pure peak of a single component. In Fig. 10b, the heteroscedastic noise, which results from a strong absorbance response, has a great effect on the

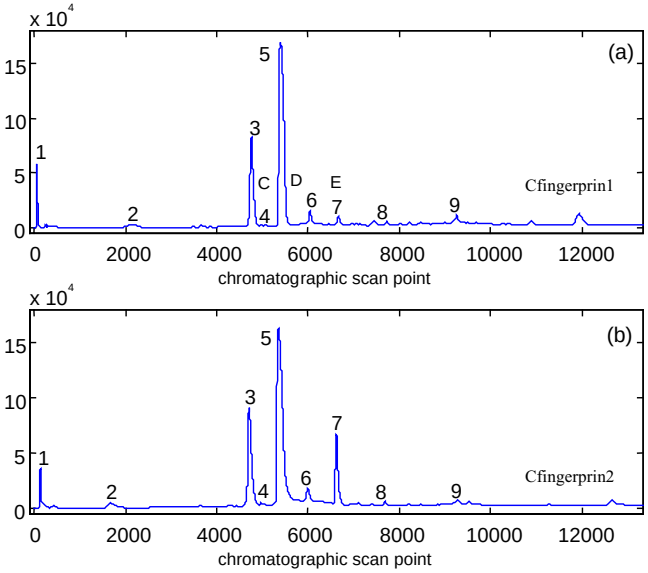

Fig. 8. Real chromatographic fingerprints of *Cortex cinnamomi* from two different sources.

FSWM graph of the D peak cluster [23]. With the help of the chemometric resolution techniques, the resolved chromatographic profiles and UV spectra are shown in Fig. 11. With the UV spectra obtained, the 4, 5 and 7 components might be qualitatively determined. They are cinnamyl alcohol ($C_9H_{10}O$), cinnamaldehyde (C_9H_8O) and cinnamyl acetate ($C_{11}H_{12}O_2$), respectively. In order to further determine these components, GC–MS analysis has been also employed in this work (see Section 3.5.2). Their mass spectra are shown in Fig. 12a–c, respectively. All these components might be pharmacologically active and marker compounds existing in *Cortex cinnamomi*.

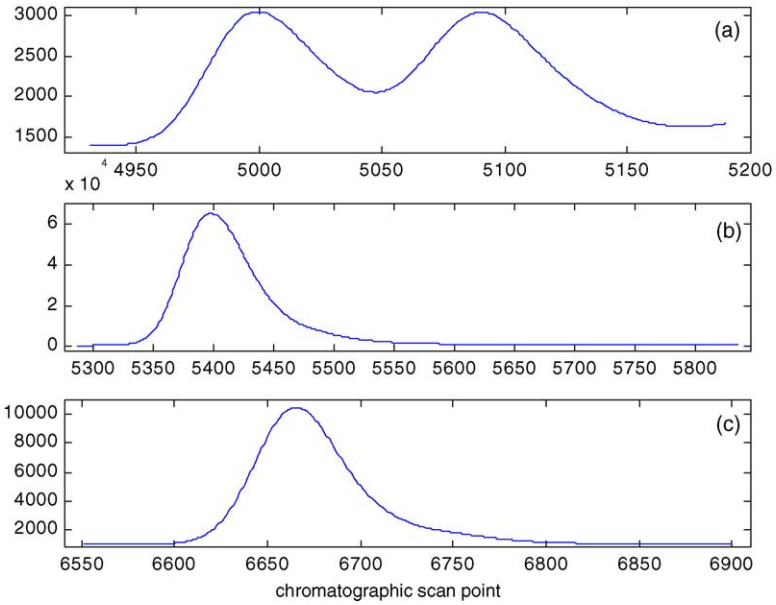

Fig. 9. C, D and E peak clusters in Fig. 8a.

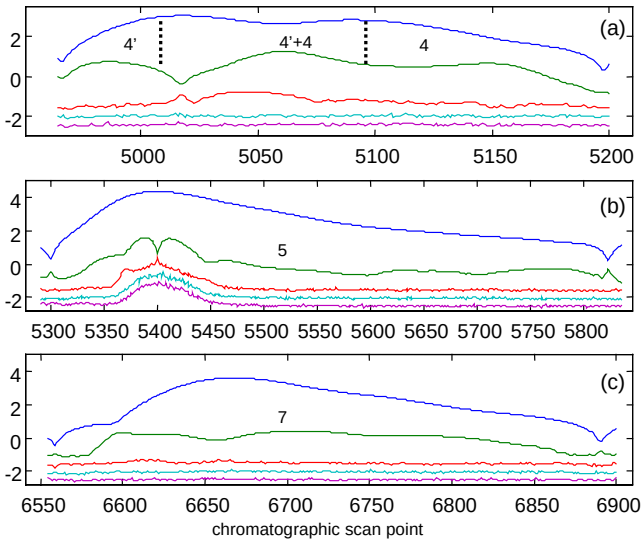

Fig. 10. FSWM plots for C (a), D (b) and E (c) peak clusters.

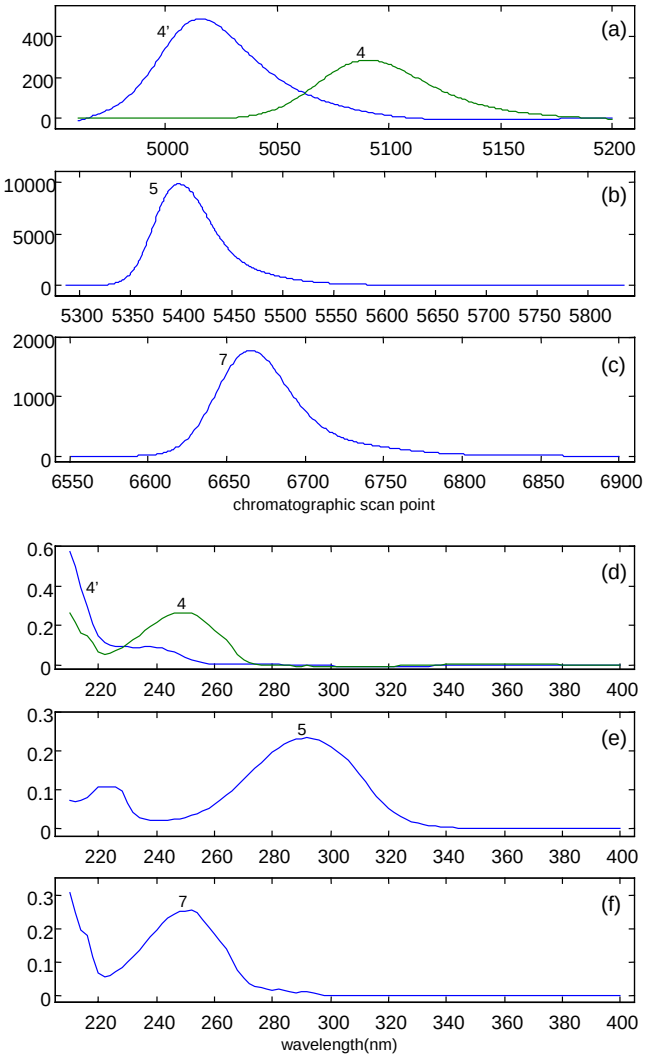

Fig. 11. Resolved chromatograms (a–c) and spectra (d–f) of C, D and E peak clusters.

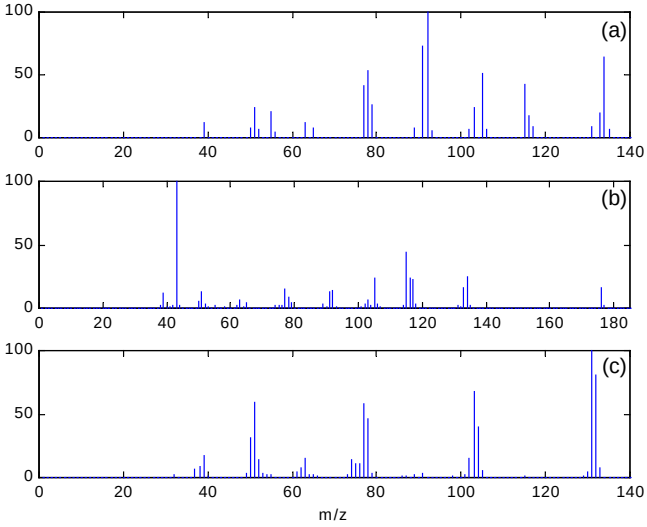

Fig. 12. Mass spectra of cinnamyl alcohol (a), cinnamaldehyde (b) and cinnamyl acetate (c).

In the same way as described above, the peak purity identification and the resolution into pure chromatograms and UV spectra of components involved in other peaks clusters from these two real chromatographic fingerprints are also conducted (see Fig. 8). For the sake of the brevity of this paper, details are not described here.

With the resolved chromatograms and UV spectra obtained, nine peaks are selected from the chromatographic fingerprints investigated. These peaks selected might represent the mark compounds which exist in both Cfingerprin1 and Cfingerprin2 (see Fig. 8). Their retention time points are listed in Table 2. Seen from Table 2, retention time shifts are surely in existence.

4.2. Correction of retention time shifts and reconstruction of chromatographic fingerprints of herbal medicines with cubic spline interpolation

From above, nine marker compounds are selected and their retention time points are determined. The correction of the retention time shifts could be conducted. In this study, Cfingerprint1 is regarded as the target and Cfingerprint2 is aligned with it. As pointed out in Section 2.2, if the chromatographic scan points of Cfingerprint2 are less than 132 (the retention time point of the first marker compound in Cfingerprint2), the retention time shifts for these points will be calculated by linear interpolation between 1 and 58 (the retention time point of the first marker compound in the target). Similar to it, the chromatographic scan points of Cfingerprint2 between 9271 (the retention time point of the last marker compound in Cfingerprint2) and its last scan point (13,343 here) are adjusted with linear interpolation between 9253 (the retention time point of the last marker compound in the target) and 13,343. However, as for the points between two adjacent fix points of Cfingerprint2 such as between

Table 2
Retention time points of nine marker compounds in the real chromatographic fingerprints of *Cortex Cinnamomi*

	Retention time point								
	1	2	3	4	5	6	7	8	9
Cfingerprint1	58	2162	4756	5090	5398	6039	6665	7716	9253
Cfingerprint2	132	1668	4710	5041	5353	5992	6620	7687	9271

5041 and 5353, the retention time shifts are corrected between the corresponding points in the target like 5090 and 5398. After the data treatment like it, new chromatographic scan points of Cfingerprint2 are obtained. Note, most of these new points are not integer.

With the new chromatographic scan points of Cfingerprint2 obtained above, a new chromatographic fingerprint of Cfingerprint2 with the correction of the retention time shifts might be reconstructed by use of cubic spline interpolation. Fig. 13 shows Cfingerprint1 and Cfingerppprint2 of real *Cortex cinnamomi* samples before and after correcting the retention time shifts. Seen from Fig. 13b, Cfingerprint2 is aligned with Cfingerprint1 effectively.

Likely, several chromatographic fingerprints from other herbal medicines are aligned with the correction of the retention time shifts. The herbal medicines investigated cover 50 *Rhizoma chuanxiong*, 10 *Radix angelicae* and 17 *Herba menthae* samples from different sources in this work. For all these herbs, the ending chromatographic scan point is 13343. Fig. 14 represents the chromatographic fingerprints from one *Cortex cinnamomi*, *Rhizoma chuanxiong*, *Radix angelicae* and *Herba menthae* before and after correcting the retention time shifts, respectively.

4.3. Pattern recognition based on principal component analysis before and after correcting retention time shifts of chromatographic fingerprints from real herbal medicines

In order to further assess the reliability and applicability of the approach proposed, this study has attempted to classify the real herbal medicines investigated (79 samples altogether) with chromatographic fingerprints obtained by means of pattern recognition based on principal component analysis (PCA). The profile matrix is constructed by assembling the chromatographic fingerprints as row vectors. The first two principal components PC1 and PC2 are used to provide a convenient visual aid for identifying inhomogenity in the data sets.

Fig. 15 shows the principle component projection plot of PC1 to PC2 before correcting the retention time shifts of 79 chromatographic fingerprints investigated. Clearly, it is very difficult to differentiate *Cortex cinnamomi*, *Rhizoma chuanxiong*, *Radix angelicae* and *Herba menthae* on account of the widely scattering projection points (see Fig. 15). However, as can be visually seen from Fig. 14, the differences are surely in existence between the chromatographic fingerprints from these real herbal medicines. Thus, directly conducting

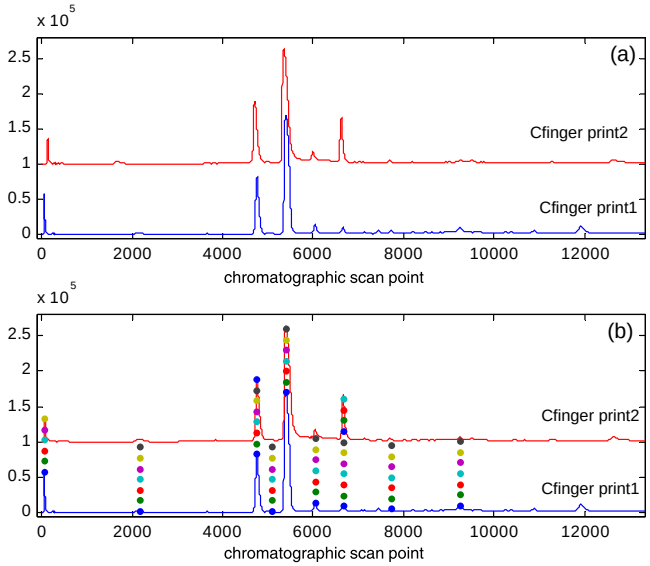

Fig. 13. Real chromatographic fingerprints of *Cortex cinnamomi* before (a) and after (b) correcting the retention time shifts.

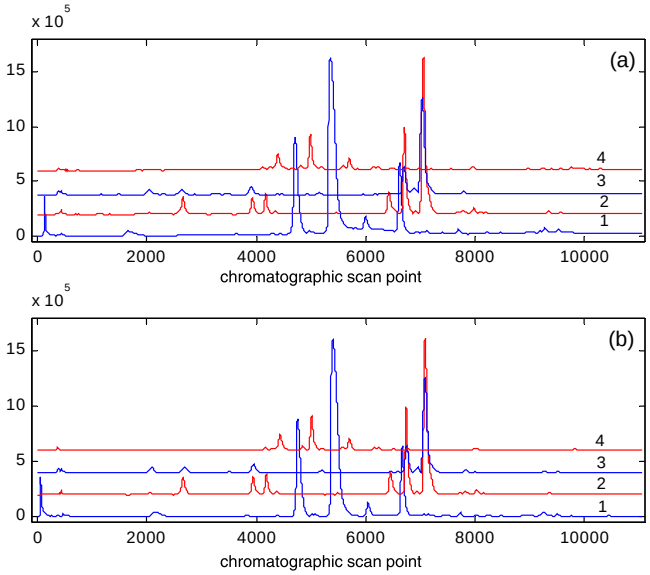

Fig. 14. Chromatographic fingerprints from one *Cortex cinnamomi* (1), *Rhizoma chuanxiong* (2), *Radix angelicae* (3) and *Herba menthae* (4) before (a) and after (b) correcting the retention time shifts.

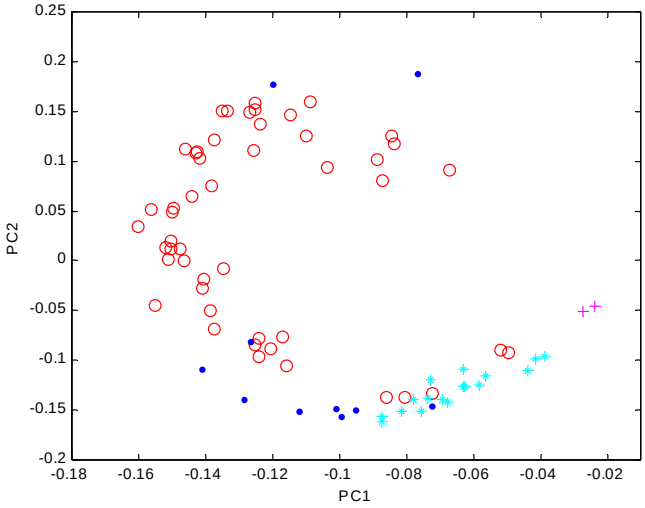

Fig. 15. Pattern recognition based on PCA before correcting the retention time shifts of chromatographic fingerprints from 79 herbal medicines ("+" *Cortex cinnamomi*, "○" *Rhizoma chuanxiong*, "●" *Radix angelicae*, "✳" *Herba menthae*. The variation explained by PC1 and PC2 is 22.42%).

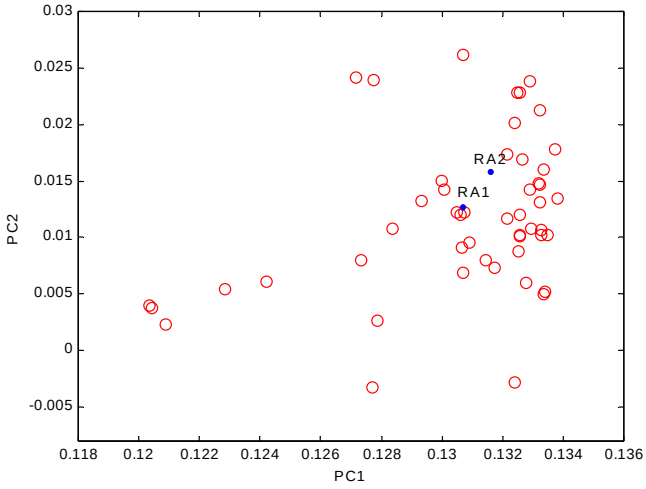

Fig. 17. Enlarged plot of the *Rhizoma chuanxiong* cluster in Fig. 16 ("○" *Rhizoma chuanxiong*, "●" *Radix angelicae*).

PCA for the original chromatographic fingerprints seems unreasonable and unreliable.

On the other hand, if the retention time shifts are corrected for the chromatographic fingerprints under study, the PC1 to PC2 plot could reveal the segregation of the projection points into four principal clusters (encircled) corresponding to four kinds of herbal medicines (see Fig. 16). In comparison with Figs. 15 and 16, the retention time shifts have a great impact on the classification and should be corrected before PCA.

Unfortunately, there are still two *Radix angelicae* samples (RA1 and RA2) which are classified incorrectly into the cluster of *Rhizoma chuanxiong* (see Fig. 17). In fact, the chromatographic fingerprints of RA1 and RA2 seem to be more similar to *Rhizoma chuanxiong* than other *Radix angelicae* due to the high similarity of a large peak cluster between 7006 and 7350 scan points in the chromatographic fingerprints (a compound named bultylidene dihydro-phthalide is involved in this region). If the mean and/or median chromatograms are used to represent the common chromatographic fingerprints from the other eight *Radix angelicae* or fifty *Rhizoma chuanxiong* samples, the similar indices obtained by $\cos\alpha$ (α is the angle between two fingerpint vectors) are listed in Table 3. Seen from Table 3, the chromatographic fingerprints of RA1 and RA2 surely match well with fifty *Rhizoma chuanxiong* other than eight *Radix angelicae*. Fig. 18 represents the chromatographic fingerprints of RA1(1) and RA2(2), the mean and median chromatograms of other 8 *Radix angelicae* (3,4) and 50 *Rhizoma chuanxiong* (5,6). From it, we can see that the simple application of pattern recognition based on PCA is sometimes not satisfactory to the classification of complicated systems. Several data pretreatment processes and even other approaches on the cluster analysis should be used to meet the requirements. Our future work will further discuss it in detail.

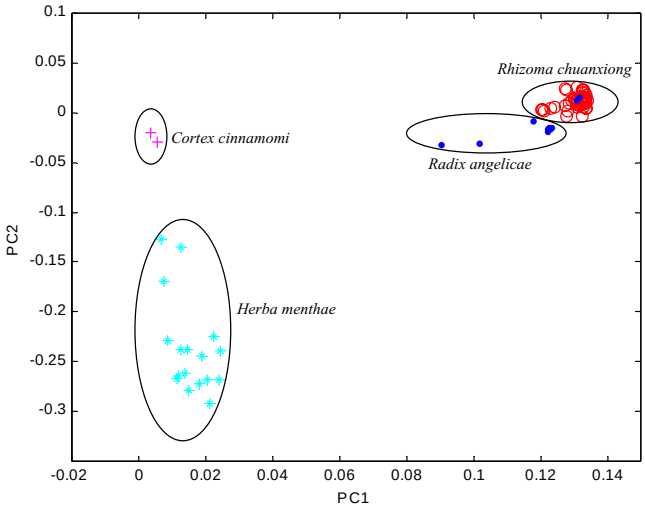

Fig. 16. Pattern recognition based on PCA after correcting the retention time shifts of chromatographic fingerprints from 79 herbal medicines ("+" *Cortex cinnamomi*, "○" *Rhizoma chuanxiong*, "●" *Radix angelicae*, "✳" *Herba menthae*. The variation explained by PC1 and PC2 is 35.09%).

Table 3
Similar index between RA1, RA2 and eight *Radix angelicae* and fifty *Rhizoma chuanxiong*

	RA1		RA2	
	1	2	3	4
Mean	0.8554	0.9634	0.8734	0.9727
Median	0.8674	0.9574	0.8757	0.9629

Note: 1 and 3, 2 and 4: the similar indices between RA1, RA2 and the mean or median chromatogram of chromatographic fingerprints from 8 *Radix angelicae* and 50 *Rhizoma chuanxiong*, respectively.

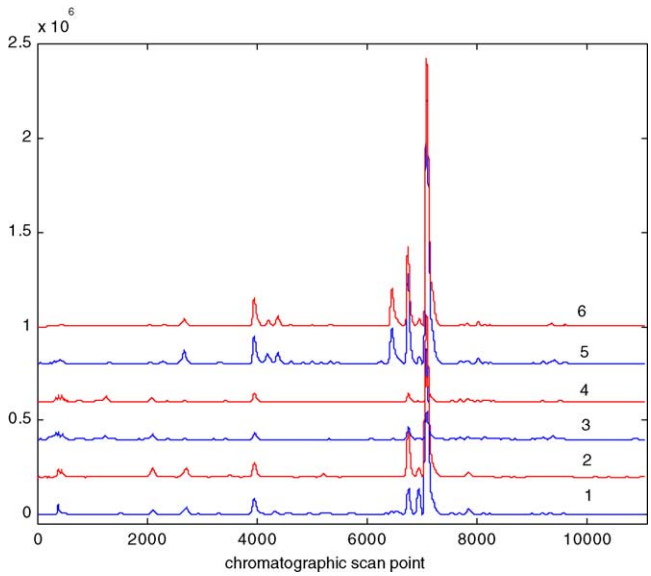

Fig. 18. Chromatographic fingerprints of RA1(1) and RA2(2), the mean and median chromatograms of other 8 *Radix angelicae* (3,4) and 50 *Rhizoma chuanxiong* (5,6).

5. Conclusion

As the retention time shifts always exist during the chromatographic runs, it is absolutely necessary to correct the retention time shifts of chromatographic fingerprints from herbal medicines for quality control. In this study, the combined approach of chemometric resolution with cubic spline data interpolation is applied to selecting marker compounds, correcting the retention time shifts and then reconstructing chromatographic fingerprints with correction. The chromatographic fingerprints from both the simulation and real herbal medicines demonstrate that the present method gives a reasonable and effective alignment for correcting the retention time shifts of chromatographic fingerprints. Especially, since the chromatographic fingerprints are produced by hyphenated chromatography, the full chromatographic and spectral information could be used to select the marker compounds in the chromatographic profiles investigated. It just benefits from the advantages of the two-dimensional data matrix.

Clearly, the proposed method is also essentially suitable for chromatographic fingerprints generated from other complex real systems like foods and other types of hyphenated chromatography such as GC–MS, GC–IR, CE–DAD, HPLC–MS.

Acknowledgements

This research work was financially supported by National Natural Science Foundation of China (Grants Nos. 20175036 and 20235020), University Grants Council of Hong Kong SAR via the Area of Excellence Project “Chinese Medicine Research and Further Development” (AoE/B-101) in Hong Kong, Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry and Youth Science Foundation of Central South University of China.

References

[1] W.J. Welsh, W.K. Lin, S.H. Tersigni, E. Collantes, R. Duta, M.S. Carey, W.L. Zielinski, J. Brower, J.A. Spencer, T.P. Layloff, Anal. Chem. 68 (1996) 3473.
[2] A. Hasler, O. Sticher, J. Chromatogr. 605 (1992) 41.
[3] N.J. Lazarowych, P. Pekos, Drug Inform. J. 32 (1998) 497.
[4] State Drug Administration of China, Chinese Trad. Pat. Med. 22 (2000) 671.
[5] F. Gong, Y.Z. Liang, P.S. Xie, F.T. Chau, J. Chromatogr. A 1002 (2003) 25.
[6] P.S. Xie, Trad. Chinese Drug Res. Clin. Pharm. 12 (2001) 141.
[7] N.P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17.
[8] J.A. Pino, J.E. McMurry, P.C. Jurs, B.K. Lavine, A.M. Harper, Anal. Chem. 57 (1985) 295.
[9] M.E. Parrish, B.W. Good, F.S. Hsu, F.W. Hatch, D.M. Ennis, D.R. Douglas, J.H. Shelton, D.C. Watson, Anal. Chem. 53 (1981) 826.
[10] K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, J. Chromatogr. A 996 (2003) 141.
[11] G. Malmquist, R. Danielsson, J. Chromatogr. A 687 (1994) 71.
[12] D. Bylund, R. Danielsson, G. Malmquist, K.E. Markides, J. Chromatogr. A 961 (2002) 237.
[13] M.D. Hämäläinen, Y.Z. Liang, O.M. Kvalheim, R. Andersson, Anal. Chim. Acta 271 (1993) 101.
[14] O.M. Kvalheim, Y. Z Liang, Anal. Chem. 64 (1992) 936.
[15] Y.Z. Liang, O.M. Kvalheim, H.R. Keller, D.L. Massart, P. Kiechle, F. Erni, Anal. Chem. 64 (1992) 946.
[16] M. Maeder, Anal. Chem. 59 (1987) 527.
[17] E.R. Malinowski, J. Chemom. 6 (1992) 29.
[18] W. Den, E.R. Malinowski, J. Chemom. 7 (1993) 89.
[19] F. Gong, Y.Z. Liang, F.T. Chau, J. Sep. Sci. 26 (2003) 112.
[20] F. Gong, Y.Z. Liang, Q.S. Xu, F.T. Chau, A.K.M. Leung, J. Chromatogr. A 905 (2001) 193.
[21] F. Gong, Y.Z. Liang, H. Cui, F.T. Chau, B.T.P. Chan, J. Chromatogr. A 909 (2001) 237.
[22] A.K.M. Leung, F. Gong, Y.Z. Liang, F.T. Chau, Anal. Lett. 33 (2000) 3195.
[23] Y.Z. Liang, White, Grey, Black Multicomponent Systems and Their Chemometric Algorithms, Hunan Publishing House of Science and Technology, Changsha, 1996, pp. 172.