

---

## Research Article

---

# Batch-to-Batch Quality Consistency Evaluation of Botanical Drug Products Using Multivariate Statistical Analysis of the Chromatographic Fingerprint

Haoshu Xiong,<sup>1</sup> Lawrence X. Yu,<sup>2</sup> and Haibin Qu<sup>1,3</sup>

Received 23 October 2012; accepted 9 April 2013; published online 2 May 2013

**Abstract.** Botanical drug products have batch-to-batch quality variability due to botanical raw materials and the current manufacturing process. The rational evaluation and control of product quality consistency are essential to ensure the efficacy and safety. Chromatographic fingerprinting is an important and widely used tool to characterize the chemical composition of botanical drug products. Multivariate statistical analysis has showed its efficacy and applicability in the quality evaluation of many kinds of industrial products. In this paper, the combined use of multivariate statistical analysis and chromatographic fingerprinting is presented here to evaluate batch-to-batch quality consistency of botanical drug products. A typical botanical drug product in China, Shenmai injection, was selected as the example to demonstrate the feasibility of this approach. The high-performance liquid chromatographic fingerprint data of historical batches were collected from a traditional Chinese medicine manufacturing factory. Characteristic peaks were weighted by their variability among production batches. A principal component analysis model was established after outliers were modified or removed. Multivariate (Hotelling  $T^2$  and DModX) control charts were finally successfully applied to evaluate the quality consistency. The results suggest useful applications for a combination of multivariate statistical analysis with chromatographic fingerprinting in batch-to-batch quality consistency evaluation for the manufacture of botanical drug products.

**KEY WORDS:** botanical drug products; chromatographic fingerprint; manufacturing process; multivariate statistical analysis; quality consistency.

## INTRODUCTION

Botanical drugs have been used in health care around the world for thousands of years (1,2). During the recent decades, use of botanical drugs has tremendously expanded and has gained popularity in both developing and developed countries (3). How to evaluate the efficacy and safety of botanical drugs has always been the intensely focused issue all over the world. While Janet Woodcock, Director for the US Food and Drug Administration (FDA)'s Center for Drug Evaluation and Research, defined a high-quality drug product as a product reproducibly delivering the therapeutic benefit promised in the label, reproducible quality of botanical drug products ensures their reproducible efficacy and safety (4). However, during the manufacturing process of botanical drug products, natural

variability of botanical raw materials, together with the current manufacturing process, can lead to batch-to-batch product quality variability (5,6). First, the quality of botanical raw materials is influenced by many factors such as cultivation position, climate, harvest time, and storage conditions. The chemical composition and biological activity of botanical raw materials often are not completely characterized. Based on quality testing on botanical raw materials, the quality consistency of products may not be achieved (7). Moreover, during multiple process operation procedures (e.g., heating, precipitating, adding acids, and bases), in-process materials may have complex chemical and physical changes. Due to the lack of process knowledge, the relationships among raw material properties, process parameters, and product quality attributes have not been fully understood. Accordingly, it is difficult to establish in-process control strategies of detecting and adjusting process abnormal variations. As a result, quality consistency evaluation should be necessary for quality assurance of finished products. Quality consistency evaluation of botanical drug products is much more difficult than that of synthetic drugs. Botanical drug products comprise hundreds of compounds, and their chemical composition is not always well characterized. Multiple compounds often work synergistically in delivering therapeutic effects, and the active constituents or biological markers are difficult to be identified (8–10).

---

Opinions expressed in this manuscript are those of the authors and do not necessarily reflect the views or policies of the FDA.

<sup>1</sup> Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China.

<sup>2</sup> Office of Pharmaceutical Science, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring Maryland 20993, USA.

<sup>3</sup> To whom correspondence should be addressed. (e-mail: qhbb@zju.edu.cn)

Chromatographic fingerprint, with the capability of characterizing the integrative chemical composition of botanical drug products, provides a reproducible, analytical approach for batch-to-batch quality consistency evaluation (11–14). Since 1991, World Health Organization (WHO) has published a series of guidelines for the evaluation of safety and efficacy as well as quality control of herbal medicinal products (3,15–17). WHO has stated that if the identification of an active principle is not possible, it should be sufficient to identify a characteristic substance or mixture of substances (e.g., “chromatographic fingerprint”) to ensure consistent quality of herbal finished products (15). Until now, the FDA and European Medicines Agency have all accepted the use of chromatographic fingerprint for quality consistency evaluation of botanical drug products (7,18). The State Food and Drug Administration (SFDA) of China has officially required all traditional Chinese medicinal (TCM) injections to be standardized by chromatographic fingerprint since 2004 (19). The Chinese pharmacopeia commission has published software, Similarity Evaluation System for Chromatographic Fingerprint of TCM, for the use in fingerprint similarity analysis. Until now, similarity analysis is still the only standardized fingerprint analysis method in China, and TCM manufacturers have widely taken the similarity indexes as the quality consistency criteria of their products. During this time, similarity analysis has been widely reported in literature as the quality consistency evaluation tool of both botanical drug products and raw materials (2,13,20–37).

Using similarity analysis, the quality consistency evaluation of a botanical drug product is achieved through comparing key signatures (chromatographic peaks) of its fingerprint against a reference fingerprint. The fingerprint of a product manufactured under a standardized process can be adopted as a reference. However, perfectly standardized products are rarely available, since operation variability and disturbances often happen during the manufacturing process. Therefore, the reference fingerprint is usually derived by calculating the mean or median fingerprint of a set of samples (2,24). However, similarity analysis has been called into question as a criterion for quality consistency (2,25–27,30–33). First, the representative of one reference fingerprint for reflecting the variability of botanical drug products is somewhat insufficient (25). Moreover, the determination of similarity threshold is rather subjective, frequently set to ensure the correct classification or discrimination of the maximal number of samples (2,25–27). As a result, more robust statistical methods are needed to overcome some of the concerns of the current approach. Multivariate statistical analysis has received considerable attention and has been widely applied in quality control of industrial manufacture across many sectors, including chemically synthesized medicines (38,39). Using multivariate statistical analysis on fingerprint data, the common-cause variations among products are presented by a statistical model established on a set of production batches. The quality variations of products can be detected through simultaneously monitoring the variations of multiple peaks and their correlated relationships using the statistical outputs of multivariate modeling, such as Hotelling  $T^2$  and DModX (40,41).

The another concern about similarity analysis is the mathematical properties of the two frequently adapted similarity indexes, i.e., correlation coefficient and vector cosine (25,28,29). The weight of each peak, for the similarity calculation, is proportional to its area under the peak. Hence, the contribution to values of similarity indexes mainly stems from

the major peaks, inducing the severe ignorance of small peaks (30–33). In fact, each peak has variability among production batches, which has a direct impact on the batch-to-batch product quality variability. Thus, more attention needs to be paid to peak variability. Since the peak area variability is not necessarily correlated with the size of peak areas, the weight distribution among peaks when calculating correlation coefficient or vector cosine is questionable. In this study, we introduced a weighting algorithm to distribute the weights of peaks according to their variability, prior to statistical modeling.

In this study, multivariate statistical analysis was combined with chromatographic fingerprinting to achieve the aim of batch-to-batch quality consistency evaluation of botanical drug products. Shenmai injection is one kind of widely used botanical drug products in China, used to treat coronary atherosclerotic cardiopathy and viral myocarditis and to raise tumor patient's immunity (34,42). Shenmai injection is adopted as a typical example for the methodology demonstration. It is produced from two botanical raw materials, Radix Ginseng and Radix Ophiopogonis. According to the state drug standards published by SFDA of China in 2011 (43), the quality of Shenmai injection is assured based on the content determination of ginsenosides and the high-performance liquid chromatographic (HPLC) fingerprint similarity analysis. Because the power of statistical methodology depends on the quality and quantity of the data, a data set with sufficient information content (usually meaning a large amount of observations) should be collected to ensure the representative of the batch-to-batch common-cause variation (39). In this study, HPLC fingerprint data of 272 historical batches during 2 years from a TCM manufacturer were collected. First, each peak in these fingerprints was standardized and weighted according to its variability. Next, outliers were modified or removed. Multivariate statistical analysis of the fingerprints was performed, and Hotelling  $T^2$  and DModX statistics with their control limits were used to evaluate quality consistency of Shenmai injection.

## MATERIALS AND METHODS

### Sample Collection

Chromatographic fingerprint data were provided by the TCM manufacturing factory, Chiatai Qingchunbao Pharmaceutical Co. Ltd. (Hangzhou, Zhejiang, China). Data were acquired from 272 batches manufactured during March, 2011 and June, 2012. Immediately after each batch was completed, one sample was collected and measured by the quality control department. Finally, 272 fingerprints were acquired in total.

### Acquirement of HPLC Fingerprint

The HPLC fingerprint method in the state drug standard of Shenmai injection published by SFDA in 2011 was used (43). The analysis was performed on an Agilent 1200 system (Agilent, Palo Alto, CA, USA) equipped with an auto-sampler, a vacuum degasser, a quaternary pump, a column oven, and a photodiode array detector. All modules and data collection were controlled by Agilent Chemstation software. A Waters symmetry shield<sup>TM</sup> RP18 column (4.6×250 mm, 5.0  $\mu$ m) and a Hanbon guard column (Hanbon Co. Ltd., Huaiyin, Jiangsu, China) were used. The mobile phase consisted of (A) water and (B) acetonitrile and was consecutively programmed as the following gradient with a flow rate of 1.0 mL/min: 0–30 min, 0–10% B; 30–40 min, 10–23%

B; 40–50 min, 23% B; 50–85 min, 23–60% B; and 85–95 min, 60–100% B. The column temperature was maintained at 30°C. The detection wavelength was 203 nm.

Accurately weighted reference standard compounds, ginsenoside Rg<sub>1</sub>, Re, and Rb<sub>1</sub>, were dissolved in 20% aqueous acetonitrile (v/v) to prepare the mixed standard solution. The concentrations of ginsenoside Rg<sub>1</sub>, Re, and Rb<sub>1</sub> in the solution were 0.10, 0.08, and 0.20 mg/mL, respectively. Shenmai injection was directly injected into the HPLC analysis without the needs of any sample preparation. The injection volume of both the standard solution and sample solution was 10 μL. Prior to measurement of each new batch sample, the standard solution was measured.

### Preprocessing of Fingerprint Data

The peak areas of  $K$  characteristic peaks in the fingerprints of  $N$  batches were used to construct the fingerprint data matrix  $X$  ( $N \times K$ ). An appropriate transformation procedure applied on raw data is important, since it has a large influence on the importance of each peak in the model (30). This can be achieved by the combination of standardization and weighting, which give specific weight to each peak according to its importance on modeling. Each peak has its variability among production batches, and the variability needs to be controlled to ensure the product quality consistency. Therefore, the importance of each peak is related to its variability. The peaks with a larger variability among normal batches will be given a lower weight, permitted to a wide tolerance range of variation. On the contrary, we can apply a stricter limitation range of variation on the peaks, which contribute a smaller variability among batches. That is to say, the weight of each peak must be inversely proportional to the relative standard deviation (RSD) of the peak areas.

To achieve this, the data matrix is first standardized to put the equivalent weight on each peak. Standardization is usually calculated as follows:

$$xs_{n,k} = \frac{x_{n,k} - \text{Average}_k}{\text{SD}_k} \quad (1)$$

where  $xs_{n,k}$  is the standardized value of the area under the  $k$ th peak in the  $n$ th observation,  $x_{n,k}$  is the raw peak area value, and  $\text{Average}_k$  and  $\text{SD}_k$  are the mean and standard deviation of the  $k$ th peak among  $N$  observations, respectively. The RSD value of the  $k$ th peak among  $N$  observations is calculated as follows:

$$\text{RSD}_k = (\text{Average}_k / \text{SD}_k \times 100)\% \quad (2)$$

In large datasets, there are often disruptive outliers with significant deviations from the other samples. When multiple outliers exist, the mean value is biased toward the outliers, and

the standard deviation is inflated. To reduce the effect of multiple outliers, robust scaling was adopted here as suggested (40,44).  $\text{Average}_k$  is represented by the median in place of the mean.  $\text{SD}_k$  is replaced with median absolute deviation (MAD) from the median, as follows:

$$S_{\text{MAD}} = 1.4826 \text{ median}_n \{|x_{n,k} - x_{\text{median}}|\}, \quad n = 1, 2, 3, \dots, N \quad (3)$$

where  $x_{\text{median}}$  is the median of the  $k$ th peak among  $N$  observations and the constant 1.4826 is required to make  $S_{\text{MAD}}$  an unbiased estimate of the standard deviation for Gaussian data.

Next, each peak is weighted according to its RSD, as follows:

$$xw_{n,k} = xs_{n,k} / \text{RSD}_k \quad (4)$$

Hence, the standardization and weighting treatments are both conducted in a robust way.

### Winsorizing and Removal of Outliers

As mentioned above, disruptive outliers often exist in large datasets. They may be introduced during sample pretreatment, by column and detector anomalies, badly transcribed data, or operator introduced reasons. Outliers will severely distort the data covariance structure and induce inaccurate modeling. Therefore, these serious outliers need to be modified or eliminated from the dataset prior to model development. Winsorizing is one simple but effective approach for outlier modification (45–47). First, each variable is sorted by its numerical values. Then a small percentage (typically between 1% and 5%) of the extreme values at the two ends of the sorted array is modified. In this paper, each variable of the dataset was 95% winsorized. That is, any variable value above the 97.5th percentile was set equal to the 97.5th percentile, and any variable value below 2.5th percentile was set equal to the 2.5th percentile. Thus, a total of 5% of the data were modified.

Winsorizing can merely modify those outliers with extreme variable values. After that, principal component analysis (PCA) is conducted on the modified data. In the score or Hotelling  $T^2$  plots of PCA, some observations will fall out of the control limits, and this kind of outliers is directly removed from further modeling (45–47).

### Multivariate Statistical Analysis of Fingerprint Data

Principal component analysis was conducted on the preprocessed fingerprint data with outlier adjustment and removal. The first  $A$  principal components (PCs) were selected

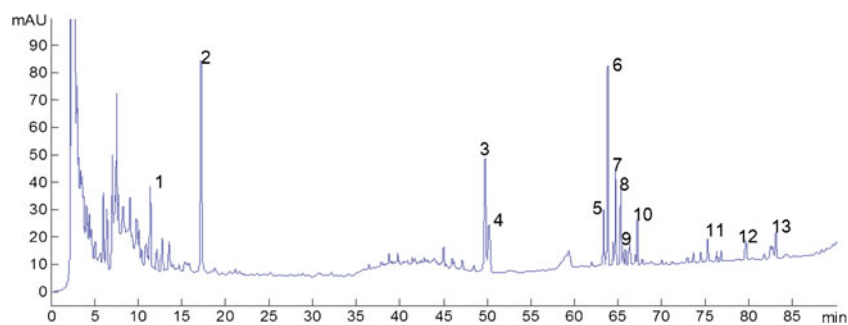


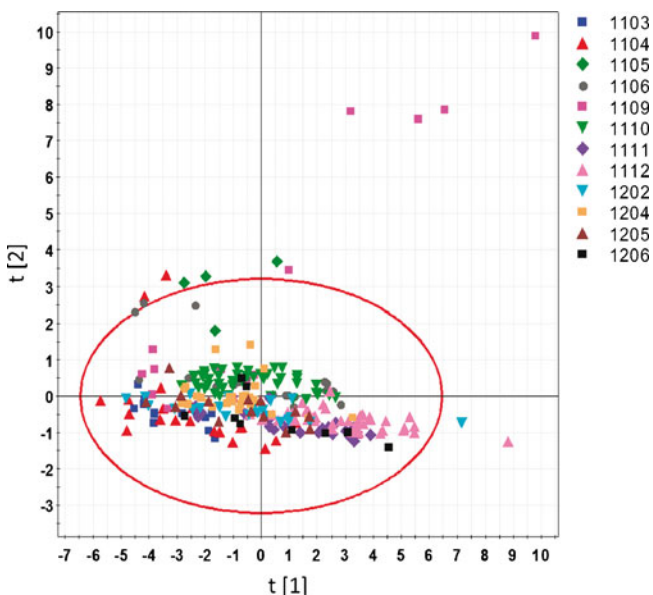
Fig. 1. Typical fingerprint of Shenmai injection with 13 characteristic peaks marked

**Table I.** Calculation Results of the Median Values,  $S_{MAD}$ , and RSDs of 13 Characteristic Peaks Among 272 Batches

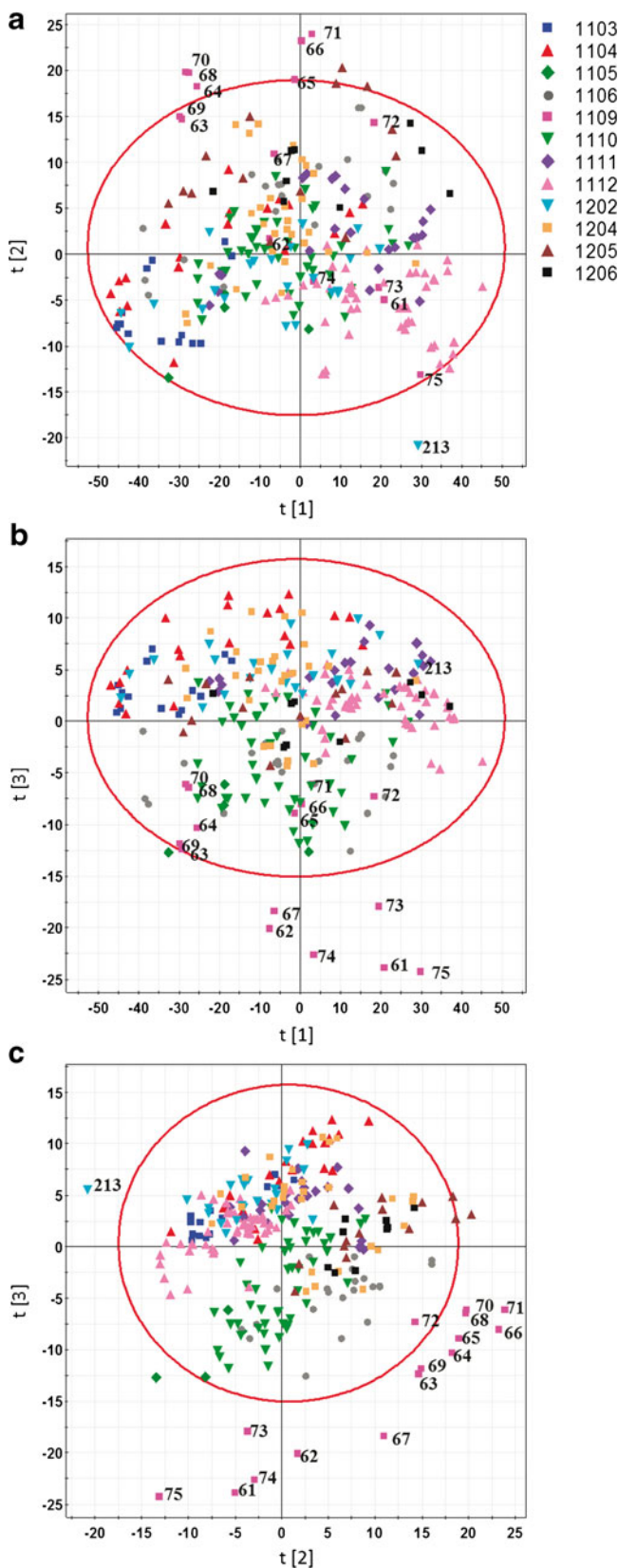
Peak no.	Retention time (min)	Median	$S_{MAD}$	RSD (%)
1	10.94	283.19	33.83	11.94
2	16.91	898.89	92.72	10.31
3	49.71	605.02	45.75	7.56
4	50.17	283.51	20.59	7.26
5	63.36	194.15	28.81	14.83
6	63.83	558.91	75.59	13.52
7	64.72	296.29	37.23	12.56
8	65.35	222.88	33.54	15.04
9	66.36	115.73	18.88	16.32
10	67.26	138.41	23.70	17.12
11	75.34	95.13	19.53	20.52
12	79.73	92.70	20.80	22.44
13	83.22	106.08	19.06	17.96

It clearly shows that peaks 1 and 2, peaks 3 and 4, peaks 5–7, peaks 8–10, and peaks 11–13 have similar RSD values with each other *RSD* relative standard deviation

through cross validation. Based on the PCA model, two kinds of multivariate control charts, Hotelling  $T^2$  and DModX, were established to monitor the data variability. The selected  $A$  PCs explain the main variability of the data. Hotelling  $T^2$  is calculated to summarize  $A$  PCs, and thus, the overall variation of fingerprint data can be monitored using this single control chart. The residuals that cannot be explained by the  $A$  PCs are represented using the DModX statistic. We take 95% as the control limits of the two statistics, and the two statistic values of the batches were then compared to the control limits. If Hotelling  $T^2$  and DModX of a batch both fall within the control limits, the batch is considered in control. If one or both of them are out of the limits, the batch deviates from the common-cause variation. When Hotelling  $T^2$  exceeds the control limit, the deviation of this observation within the

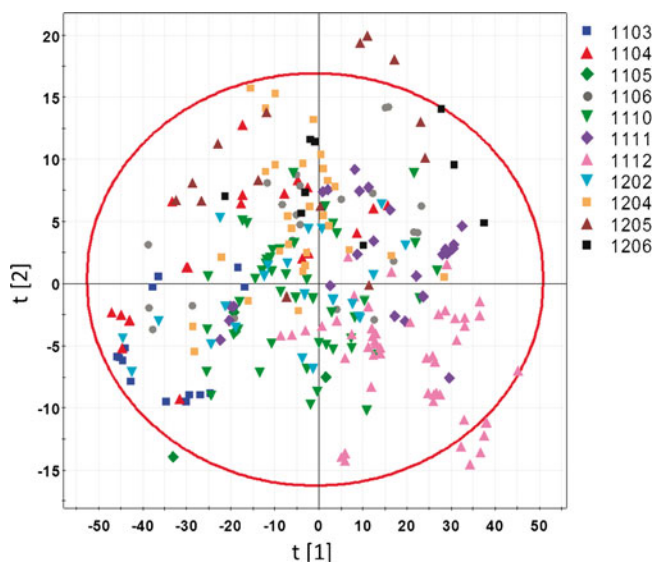


**Fig. 2.** PC1 and PC2 score plot of 272 batches before the winsorizing procedure. The red ellipse represents Hotelling  $T^2$  95% control limit. Observations are colored according to the year and month of manufacture from March, 2011 (1,103) to June, 2012 (1,206)



**Fig. 3.** Score plot on PC1 versus PC2 (a), PC1 versus PC3 (b), and PC2 versus PC3 (c) of 272 batches after the winsorizing procedure. The red ellipse represents Hotelling  $T^2$  95% control limit. Observations are colored according to the year and month of manufacture from March, 2011 (1,103) to June, 2012 (1,206). The batches 61–75 and 213 are marked by their batch number





**Fig. 4.** PC1 and PC2 score plot of 256 batches. The red ellipse represents Hotelling  $T^2$  95% control limit. Observations are colored according to the year and month of manufacture from March, 2011 (1,103) to June, 2012 (1,206)

model from the center point is further than modeling batches. When DModX is above the control limit, the covariance structure of the data has changed, suggesting occurrence of a new phenomenon not captured by the established model. Furthermore, the contribution plots can be utilized to determine which chromatographic peaks contribute most to the quality defects. More details on multivariate control charts can be found in (38–41).

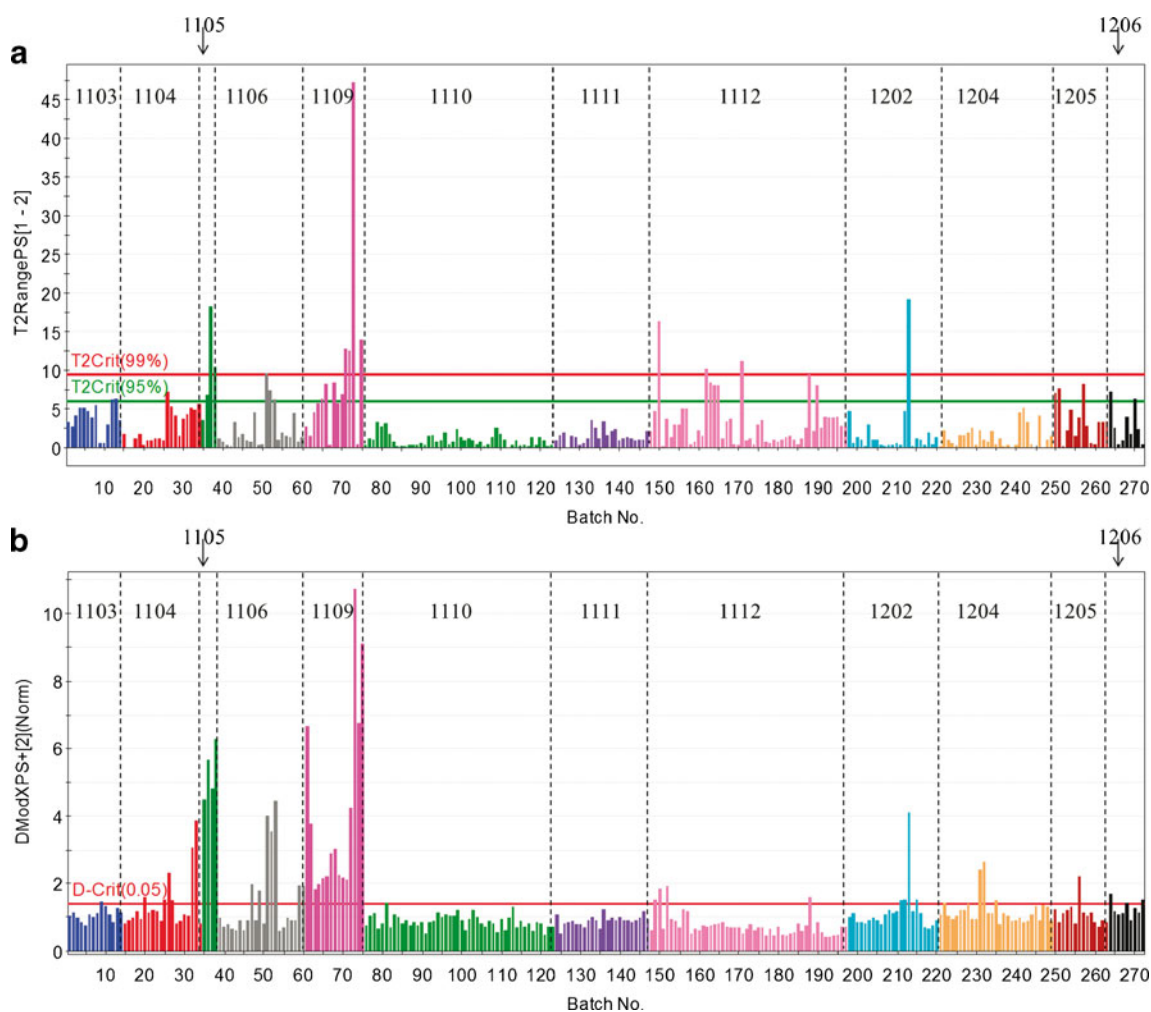
### Software

Fingerprint data preprocessing was conducted using Matlab 7.5.0 (Mathworks Inc., Natick, MA, USA). Multivariate data analysis was performed using SIMCA-P + 12.0 (Umetrics, MKS Instruments Inc., Sweden).

## RESULTS AND DISCUSSIONS

### HPLC Fingerprint of Shenmai Injection

A typical fingerprint of Shenmai injection is presented in Fig. 1. There are some common principles for choosing



**Fig. 5.** Hotelling  $T^2$  (a) and DModX (b) control charts of the 272 batches. The bars represent batches and are colored according to the year and month of manufacture from March, 2011 (1,103) to June, 2012 (1,206)

**Table II.** Monitoring Results of Hotelling  $T^2$  and DModX Control Charts

The statistics out of control	Batch no.
Hotelling $T^2$ and DModX	26, 36–38, 51–53, 66, 68, 70–73, 75, 150, 188, 213, 264
Hotelling $T^2$	12, 13, 162–165, 171, 190, 250, 251, 257, 270
DModX	9, 20, 25–27, 32, 33, 35, 47, 49, 59–65, 67, 69, 74, 81, 149, 152, 211, 212, 215, 222, 231, 232, 256, 268, 272

characteristic peaks in TCM chromatographic fingerprints (22,34–36), and we used the following principles: (1) with a relatively large peak area, (2) a well-resolved peak, (3) with the stable baseline to accurately integrate the peak, and (4) common among most batches. Finally, 13 characteristic chromatographic peaks of the 272 batches were used to construct the fingerprint data. The calculation results of the median values,  $S_{MAD}$ , and RSDs of the 13 peaks among the 272 batches are listed in Table I. The peaks with adjacent retention times have the similar RSD values, although their peak areas are not close to each other. It clearly shows that the RSD values are not correlated with the size of peak areas. Large peaks do not always owe low RSDs, while little peaks do not always owe high RSDs.

### Winsorizing and Removal of Outliers

After the preprocessing step of fingerprint data, the outliers were modified or further removed. Figures 2 and 3a show the first 2 PC scores of PCA of 272 batches before and after the winsorizing procedure, respectively. Before winsorizing, there did exist some serious outliers far outside the Hotelling  $T^2$  95% ellipse, and all of them were effectively eliminated by the winsorizing procedure.

After winsorizing, the first 3 PCs, explaining 87.6% of X variance, were selected through 7-fold cross validation. Some

outliers remained and can be observed in the first 3 PC score plots (Fig. 3). All batches manufactured in September, 2011 (batches 61–75) are outside the Hotelling  $T^2$  95% ellipse, especially in the plot of PC2 versus PC3. The reason for these outliers is coincidental with changes that were taking place at the factory. The factory had been installing the automation system on its production equipments during July to August, 2011. When the equipments with automation system were put into use at first in September, some degree of increased variability of process conditions was expected because of system setting adjustments and insufficient manual experience. In the additions, the batch 213 was also far from the Hotelling  $T^2$  ellipse. Hence, these 16 batches were excluded from statistical modeling.

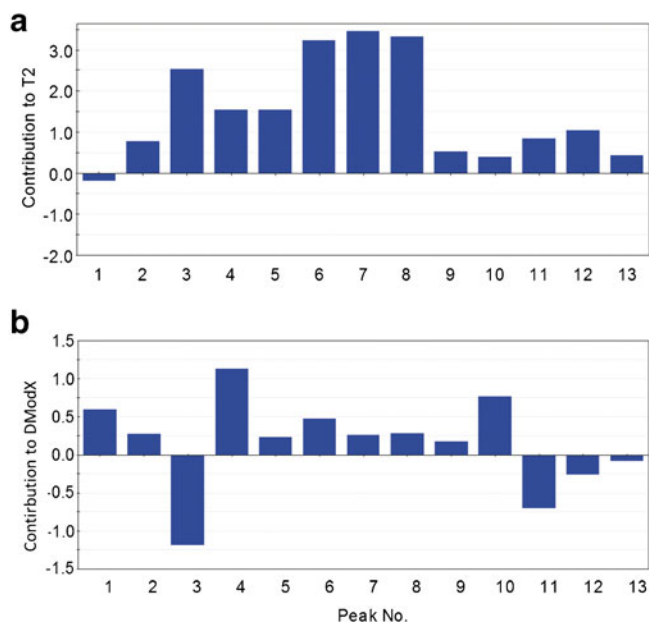
### Multivariate Statistical Analysis on Fingerprint Data

After the removal of outliers, PCA was conducted on the remaining 256 batches, and the first 2 PCs were selected through 7-fold cross validation. From the score plot (Fig. 4), it can be observed that most batches are in the Hotelling  $T^2$  95% ellipse, and as few as nine batches are outside but very close to the ellipse. Finally, the PCA model was established on these 256 batches, and the first 2 PCs totally explained 82.9% of data variance, suggesting a good fitness to the common-cause variation of the fingerprint data.

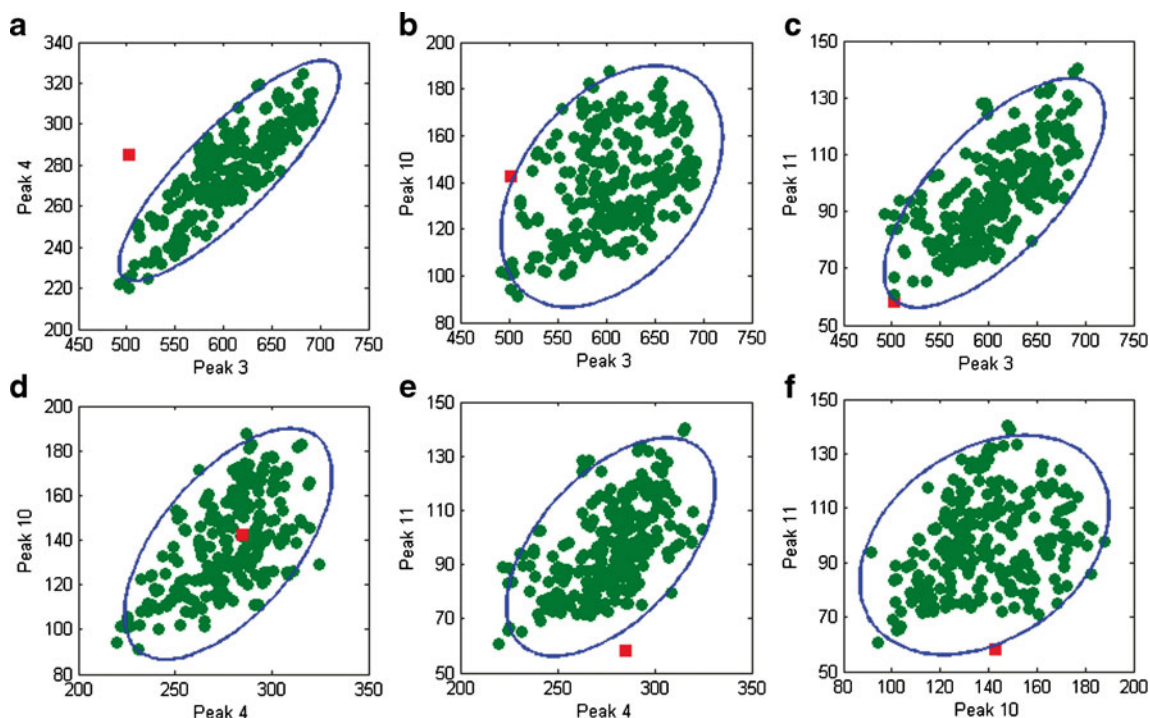
In a next step, the Hotelling  $T^2$  and DModX statistics were used to monitor the variations of fingerprints. The out-of-bounds Hotelling  $T^2$  suggests the extreme values of peak areas, while the out-of-bounds DModX reflects the change of correlation relationship between peaks, namely the area ratios. These two statistics are two complementary statistics and quite necessary to be observed at the same time. The Hotelling  $T^2$  and DModX control charts of all the 272 batches are plotted in Fig. 5. The monitoring results are summarized in Table II. A total of 18 batches exceed both the control limits of Hotelling  $T^2$  and DModX. Besides, a total of 12 batches are above the Hotelling  $T^2$  control limit, while DModX values of other 32 batches are out of control. Below are examples to demonstrate the three different situations of beyond control.

#### Batch 163

Hotelling  $T^2$  of batch 163 exceeds the control limit, while DModX is still under control (Fig. 5). From the contribution plot to Hotelling  $T^2$  (Fig. 6a), the abnormal variation of this batch is found mainly in peaks 3–8. These peaks owe much larger peak areas than the average levels among the 256 batches.



**Fig. 6.** Contribution to Hotelling  $T^2$  of batch 163 (a) and to DModX of batch 63 (b)

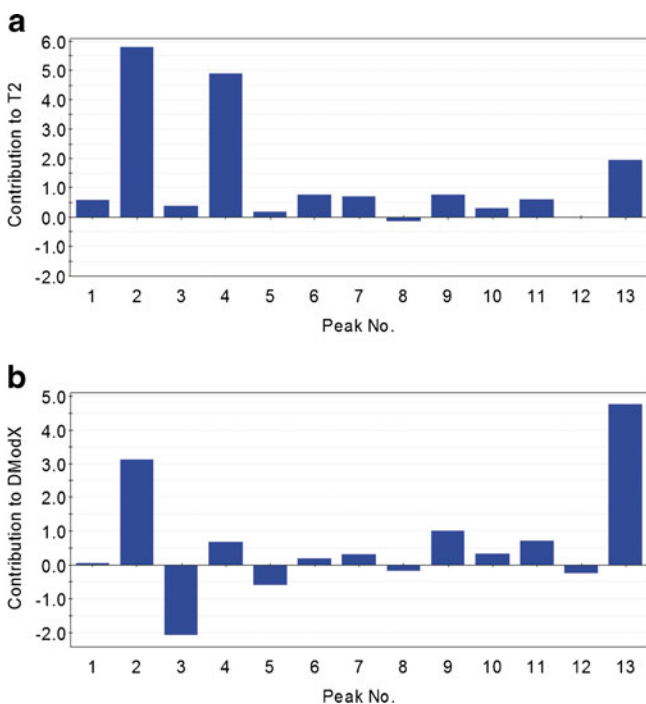


**Fig. 7.** Bivariate plots of peak 3 versus peak 4 (a), peak 3 versus peak 10 (b), peak 3 versus peak 11 (c), peak 4 versus peak 10 (d), peak 4 versus peak 11 (e), and peak 10 versus peak 11 (f). The green dots and red boxes denote the 256 batches for modeling and batch 63, respectively. The blue ellipses denote the 95% confidence limits

#### Batch 63

DModX of batch 63 is above the control limit, while Hotelling  $T^2$  below the control limit (Fig. 5). From the contribution to DModX (Fig. 6b), the abnormal variation of this batch is found mainly in peaks 3 and 4, followed by peak 10, 11, and 1, etc. These peaks together lead to the abnormal

area ratios among peaks. The immediate investigation on these peaks is quite helpful to enhance the understanding. Figure 7 shows the bivariate plots of any two peaks among peaks 3, 4, 10, and 11. Seen from the Fig. 7a, although the areas of peaks 3 and 4 are completely in the respective normal distribution ranges, the observation is still outside the 95% ellipse, showing the abnormal peak area ratio of peak 3 versus peak 4. The same situation can also be observed in the Fig. 7b, c, e, f.



**Fig. 8.** Contribution to Hotelling  $T^2$  (a) and DModX (b) of batch 72

#### Batch 72

Both Hotelling  $T^2$  and DModX of batch 72 are outside the control limit (Fig. 5). The complementary information is acquired in the contribution plots to Hotelling  $T^2$  and DModX. From contribution to Hotelling  $T^2$  (Fig. 8a), the abnormal variation was reflected mainly in the abnormal area values of peaks 2, 4, and 13. It was also noticed that peak 3 together with peaks 2 and 13 induce to the change of the area ratios (Fig. 8b).

The three examples above showed the three different situations of quality inconsistency by multivariate statistical analysis and emphasized the utility of the combination use of Hotelling  $T^2$  and DModX. Based on our current research, more explanations on the nature of abnormal batches is difficult. The establishment of the relationships among input material attributes, process parameters, and product quality attributes is the key for diagnosing the reasons of abnormal quality variations. However, it is still a challenge in the botanical drug industry in China. For a long time in the history, the manufacture is mainly based on experience, and the quality control relies on the quality testing of finished products. The systematic accumulation of detailed process data is inadequate, and the effects of process variables on

the product quality are not well understood scientifically. The identification of abnormal variability in this study is only the first step for strengthening quality control of botanical drug products and provides a basis for better process understanding.

## CONCLUSIONS

Multivariate statistical analysis combined with chromatographic fingerprint is presented in this paper to evaluate the batch-to-batch quality consistency of botanical drug products. The Hotelling  $T^2$  and DModX statistics have demonstrated themselves to be an effective statistical criterion for monitoring quality variations of the application example in this paper. Since the quality properties of botanical drug products are usually multivariate, the method of multivariate statistical analysis may also be applicable to other kinds of quality data such as the mass spectrometry, nuclear magnetic resonance, and near-infrared spectral data. This paper offers promising applications of multivariate statistical analysis in batch monitoring and evaluation for botanical drug products.

## ACKNOWLEDGMENTS

This work was financially supported by the China International Science and Technology Cooperation Project (no. 2010DFB33630).

## REFERENCES

- Liang XM, Jin Y, Wang YP, Jin GW, Fu Q, Xiao YS. Qualitative and quantitative analysis in quality control of traditional Chinese medicines. *J Chromatogr A*. 2009;1216:2033–44.
- Tistaert C, Djaegher B, Heyden YV. Chromatographic separation techniques and data handling methods for herbal fingerprints: a review. *Anal Chim Acta*. 2011;690:148–61.
- World Health Organization. General guidelines for methodologies on research and evaluation of traditional medicine. Geneva: World Health Organization; 2000. [www.paho.org/spanish/ad/thse/ev/PM-WHOTraditional-medicines-research-evaluation.pdf](http://www.paho.org/spanish/ad/thse/ev/PM-WHOTraditional-medicines-research-evaluation.pdf). Accessed 25 Feb 2013
- Woodcock J. The concept of pharmaceutical quality. *Am Pharm Rev*. 2004;7:10–5.
- Qu HB, Cheng YY, Wang YS. Some engineering problems on developing production industry of modern traditional Chinese medicine. *Chin J Chin Mat Med*. 2003;28:904–6.
- Xiong HS, Gong XC, Qu HB. Monitoring batch-to-batch reproducibility of liquid–liquid extraction process using in-line near-infrared spectroscopy combined with multivariate analysis. *J Pharm Biomed Anal*. 2012;70:178–87.
- Food and Drug Administration. Guidance for industry: botanical drug products. Rockville: Food and Drug Administration; 2004. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070491.pdf>. Accessed 19 Oct 2012
- Xie PS, Chen S, Liang YZ, Wang X, Tian R, Upton R. Chromatographic fingerprint analysis—a rational approach for quality assessment of traditional Chinese herbal medicine. *J Chromatogr A*. 2006;1112:171–80.
- Xie PS, Leung AY. Understanding the traditional aspect of Chinese medicine in order to achieve meaningful quality control of Chinese materia medica. *J Chromatogr A*. 2009;1216:1933–40.
- Wangner H, Ulrich-Merzenich G. Synergy research: approaching a new generation of phytopharmaceuticals. *Phytomedicine*. 2009;16:97–110.
- Zhou JL, Qi LW, Li P. Quality control of Chinese herbal medicines with chromatographic fingerprint. *Chin J Chromatogr*. 2008;26:153–9.
- Jiang Y, David B, Tu PF, Barbin Y. Recent analytical approaches in quality control of traditional Chinese medicines—a review. *Anal Chim Acta*. 2010;657:9–18.
- Liang YZ, Xie PS, Chau FT. Chromatographic fingerprinting and related chemometric techniques for quality control of traditional Chinese medicines. *J Sep Sci*. 2010;33:410–21.
- Fan XH, Cheng YY, Ye ZL, Lin RC, Qian ZZ. Multiple chromatographic fingerprinting and its application to the quality control of herbal medicines. *Anal Chim Acta*. 2006;555:217–24.
- World Health Organization. Guidelines for the assessment of herbal medicines. Geneva: World Health Organization; 1991. [http://apps.who.int/iris/bitstream/10665/58865/1/WHO\\_TRM\\_91.4.pdf](http://apps.who.int/iris/bitstream/10665/58865/1/WHO_TRM_91.4.pdf). Accessed 25 Feb 2013
- Regional Office for the Western Pacific, World Health Organization. Research guidelines for evaluating the safety and efficacy of herbal medicines. Manila: Regional Office for the Western Pacific, World Health Organization; 1993.
- World Health Organization. Quality control methods for medicinal plant materials. Geneva: World Health Organization; 1998. <http://apps.who.int/iris/bitstream/10665/41986/1/9241545100.pdf>. Accessed 25 Feb 2013
- European Medicines Agency. Guideline on declaration of herbal substances and herbal preparations in herbal medicinal products/traditional herbal medicinal products. London: European Medicines Agency; 2010. [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500003283.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003283.pdf). Accessed 19 Oct 2012
- State Food and Drug Administration of China. Technical requirements for the development of fingerprints of TCM injections. Beijing: State Food and Drug Administration of China; 2000. <http://www.sda.gov.cn/WS01/CL0237/15768.html>. Accessed 19 Oct 2012
- Chen C, Zhang H, Xiao W, Yong ZP, Bai N. High-performance liquid chromatographic fingerprint analysis for different origins of sea buckthorn berries. *J Chromatogr A*. 2007;1154:250–9.
- Xie BG, Gong T, Tang MH, Mi DF, Zhang X, Liu J, *et al*. An approach based on HPLC-fingerprint and chemometrics to quality consistency evaluation of Liuwei Dihuang Pills produced by different manufacturers. *J Pharm Biomed Anal*. 2008;48:1261–6.
- Xie Y, Jiang ZH, Zhou H, Cai X, Wong YF, Liu ZQ, *et al*. Combinative method using HPLC quantitative and qualitative analyses for quality consistency assessment of a herbal medicinal preparation. *J Pharm Biomed Anal*. 2007;43:204–12.
- Cao Y, Wang L, Yu X, Ye J. Development of the chromatographic fingerprint of herbal preparations Shuang–Huang–Lian oral liquid. *J Pharm Biomed Anal*. 2006;41:845–56.
- Gong F, Wang BT, Chau FT, Liang YZ. Data preprocessing for chromatographic fingerprint of herbal medicine with chemometric approaches. *Anal Lett*. 2005;38:2475–92.
- Liang YZ, Xie PS, Chan K. Quality control of herbal medicines. *J Chromatogr B*. 2004;812:53–70.
- Fang KT, Liang YZ, Yin XL, Chan K, Lu GH. Critical value determination on similarity of fingerprints. *Chemom Intell Lab Sys*. 2006;82:236–40.
- Gan F, Ye RY. New approach on similarity analysis of chromatographic fingerprint of herbal medicine. *J Chromatogr A*. 2006;1104:100–5.
- Alaerts G, Erps JV, Pieters S, Dumarey M, Nederkassel AMV, Goodarzi M, *et al*. Similarity analyses of chromatographic fingerprints as tools for identification and quality control of green tea. *J Chromatogr B*. 2012;910:61–70.
- Yang LW, Wu DH, Tang X, Peng W, Wang XR, Ma W, *et al*. Fingerprint quality control of Tianjihuang by high-performance liquid chromatography–photodiode array detection. *J Chromatogr A*. 2005;1070:35–42.
- Xu CJ, Liang YZ, Chau FT, Heyden YV. Pretreatments of chromatographic fingerprints for quality control of herbal medicines. *J Chromatogr A*. 2006;1134:253–9.
- Gu RM, Tu HY, Sun H. Discussions on the chromatographic fingerprint similarity calculation methods of traditional Chinese medicine. *Chin Trad Pat Med*. 2009;31:988–90.
- Nie L, Cao J, Luo GA, Wang YM. Comparison of different methods for evaluating the similarity of the fingerprints of traditional Chinese medicines. *Chin Trad Pat Med*. 2005;27:249–52.



33. Tian RT, Xie PS. Study on the standardization of similarity evaluation method of chromatographic fingerprints (part 1). *Trad Chin Drug Res Clin Pharm.* 2006;17:40–3.
34. Cheng YY, Chen MJ, Tong WD. An approach to comparative analysis of chromatographic fingerprints for the quality of botanical drugs. *J Chem Inf Comput Sci.* 2003;43:1068–76.
35. Xu SJ, Yang L, Tian RT, Wang ZT, Liu ZL, Xie PS, *et al.* Species differentiation and quality assessment of *Radix Paeoniae Rubra* (Chi-shao) by means of high-performance liquid chromatographic fingerprint. *J Chromatogr A.* 2009;1216:2163–8.
36. Tian RT, Xie PS, Liu HP. Evaluation of traditional Chinese herbal medicine: *Chaihu* (*Bupleuri Radix*) by both high-performance liquid chromatographic and high-performance thin-layer chromatographic fingerprint and chemometric analysis. *J Chromatogr A.* 2009;1216:2150–5.
37. Yang H, Zhao CX, Wang XM, Liang YZ, Zeng YX, Wu H, *et al.* Chromatographic fingerprint investigation for quality evaluation and control of Shengui hair-growth tincture. *Planta Med.* 2010;76:372–7.
38. Kourti T. Process analytical technology beyond real-time analyzers: the role of multivariate analysis. *Crit Rev Anal Chem.* 2006;36:257–78.
39. Miletic I, Quinn S, Dudzic M, Vaculik V, Champagne M. An industrial perspective on implementing on-line applications of multivariate statistics. *J Process Control.* 2004;14:821–36.
40. Chiang LH, Colegrove LF. Industrial implementation of on-line multivariate quality control. *Chem Intell Lab Syst.* 2007;88:143–53.
41. Kourti T. Application of latent variable methods to process control and multivariate statistical process control in industry. *Int J Adapt Control Signal Process.* 2005;19:213–46.
42. Zhang HJ, Wu YJ, Cheng YY. Analysis of 'SHENMAI' injection by HPLC/MS/MS. *J Pharm Biomed Anal.* 2003;31:175–83.
43. State Food and Drug Administration of China. State drug standards (no. WS3-B-3428-98-2010Z). *Drug Standards Chin.* 2011;12:158–60.
44. Chiang LH, Pell RJ, Seasholtz MB. Exploring process data with the use of robust outlier detection algorithms. *J Process Control.* 2003;13:437–49.
45. Kettaneh N, Berglund A, Wold S. PCA and PLS with very large data sets. *Comput Stat Data An.* 2005;48:69–85.
46. Wold S, Berglund A, Kettaneh N. New and old trends in chemometrics. How to deal with the increasing data volumes in R&D&P (research, development and production)—with examples from pharmaceutical research and process modeling. *J Chemometrics.* 2002;16:377–86.
47. Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C, Wold S. Multi- and megavariate data analysis. Part I. Basic principles and applications. Multi- and megavariate data analysis. Part II. Advanced applications and method extensions. 2nd ed. Sweden: Umetrics Academy; 2006.