Journal of Chromatography A, xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Simple automatic strategy for background drift correction in chromatographic data analysis

Hai-Yan Fu [a,*], He-Dong Li [a], Yong-Jie Yu [a,b,c,*], Bing Wang [b], Peng Lu [b], Hua-Peng Cui [b], Ping-Ping Liu [b,*], Yuan-Bin She [d]

[a] School of Pharmaceutical Sciences, South-Central University for Nationalities, Wuhan 430074, China
[b] Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China
[c] Ningxia Engineering andTechnology Research Center for Modernization of Hui Medicine, Ningxia Medical University, Yinchuan 750004, China
[d] College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China

ARTICLE INFO

Article history:
Received 13 January 2016
Received in revised form 13 March 2016
Accepted 17 April 2016
Available online xxx

Keywords:
Background drift correction
Complex sample analysis
Quality control
Metabolic profiling
LC-QTOF

ABSTRACT

Chromatographic background drift correction, which influences peak detection and time shift alignment results, is a critical stage in chromatographic data analysis. In this study, an automatic background drift correction methodology was developed. Local minimum values in a chromatogram were initially detected and organized as a new baseline vector. Iterative optimization was then employed to recognize outliers, which belong to the chromatographic peaks, in this vector, and update the outliers in the baseline until convergence. The optimized baseline vector was finally expanded into the original chromatogram, and linear interpolation was employed to estimate background drift in the chromatogram. The principle underlying the proposed method was confirmed using a complex gas chromatographic dataset. Finally, the proposed approach was applied to eliminate background drift in liquid chromatography quadrupole time-of-flight samples used in the metabolic study of *Escherichia coli* samples. The proposed method was comparable with three classical techniques: morphological weighted penalized least squares, moving window minimum value strategy and background drift correction by orthogonal subspace projection. The proposed method allows almost automatic implementation of background drift correction, which is convenient for practical use.

1. Introduction

Chromatographic instruments are widely employed for complex sample analysis in various scientific fields, including metabolomics, proteomics, and quality control of natural products [1–5]. In particular, the obtained chromatogram in the quality control of natural products and metabolic analysis may involve complexity because of the large amount of compounds eluted in finite time [6]. The traditional chromatographic data analysis strategy that inspects useful chemical information by visualization is a challenging and time-consuming task. Therefore, several data mining technologies that focus on peak picking and time-shift alignment have been developed to completely use the information in the chromatographic data and improve data analysis [7–11]. However, only few studies have been conducted on the influences of chromatographic background drift. The majority of current baseline correction methods have been developed for spectroscopic data analysis, such as NMR spectroscopy [12,13], which has not been demonstrated for chromatographic data analysis.

Chromatographic background drift correction is a critical stage for chromatographic data analysis because the background drift in the chromatogram can seriously affect the peak picking results and lead to invalid time shift alignment [6], which may provide an unacceptable conclusion. Several techniques [6,12–29] have been developed for background drift correction. These methods can be classified into penalized least squares based methods [13,14,18,24,27,28] and moving window based methods [12,21,23]. The widely known penalized least squares based method is represented by asymmetrical least squares (ALS), which estimates background drift with a penalty smooth strategy. Zhang et al. [24] improved the performance of ALS by introducing an adaptive iteratively reweighted strategy, namely, adaptive iteratively reweighted penalized least squares (airPLS).

* Corresponding authors at: School of Pharmaceutical Sciences, South-Central University for Nationalities, Wuhan 430074, China; Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China.
E-mail addresses: yongjie.yu@163.com (Y.-J. Yu), Liu_pingping2012@163.com (P.-P. Liu).

http://dx.doi.org/10.1016/j.chroma.2016.04.054
0021-9673/

The performance of ALS and airPLS largely depends on the weights assigned to useful signal parts. These methods can provide reasonable solutions for cases in which useful signals, such as peaks, in the chromatogram are properly weighted. To eliminate the influence of chromatographic peaks, Bao et al. [13] proposed an iterative technique based on peak detection to remove baselines from the NMR data. Li et al. [25] developed the morphological weighted penalized least squares (MPLS) method. Baek et al. [14] employed a logistic function to provide weights for penalized least squares. Among the successful applications of these parametric methods, the penalized parameter must be pre-optimized before use, which may be time consuming and tedious [30].

The moving window based methods [12,26] are mainly used for baseline correction in spectroscopic data analysis, such as infrared spectroscopy and Raman spectroscopy. Friedrichs [26] employed a median value in the window coupled with Gaussian filter to estimate baselines in NMR spectroscopy. Golotivin and Williams [12] subsequently developed baseline correction based on the removal of underlying signal points by comparing the difference of the maximum and minimum points with the pre-estimated instrumental noise. The efficiency of these methods depends on the window size used for data analysis. A large window may not remove the influence of background drift, whereas a small window may eliminate certain chromatographic peaks.

Yaroshchyk and Eberhardt [23] developed the moving window minimum value (MWMV) strategy for baseline correction in NMR data. The principle of MWMV is similar to that of Friedrichs' method. The minimum values in the window were initially obtained, and a moving averaging was employed by smoothing these values to estimate the underlying baseline. The performance of the MWMV still depends on the window size. Theoretically, MWMV is suitable for background drift correction of chromatograms with an acceptable baseline separation of chromatographic peaks. However, in extremely complex sample analysis, chromatographic peaks often co-elute, making MWMV lose certain useful information after baseline correction. Nevertheless, MWMV provides a new strategy for chromatographic background drift correction.

Given the numerous applications of hyphenated chromatographic techniques, such as liquid chromatography–mass spectrometry (LC–MS), in complex sample analysis, background drifts in 2D chromatographic datasets have received considerable attention [5,31–39]. Kuligowski et al. [35] employed orthogonal projection for background drift correction by taking full advantage of the bilinear structure of the hyphenated chromatographic dataset. Orthogonal subspace projection is discussed in more detail in our previous work [39]. Recently, Harrington's research group [37,38] used a similar strategy for baseline correction of hyphenated datasets. The principles of these methods are mathematically equivalent, and the only difference lies in the construction of a projection matrix of the background spectra. In complex sample analysis, the spectral profiles of blank samples or background may vary across samples, thereby presenting a challenge for these orthogonal subspace projection-based methods. Meanwhile, the signal-to-noise ratio (SNR) may significantly decrease when the spectral profiles of the components seriously overlap with those of the blank backgrounds [39]. These methods were not employed to eliminate background drift in liquid chromatography quadrupole time-of-flight (LC-QTOF) samples. Improved methods for background drift correction in high resolution LC–MS should be urgently developed [40].

In this work, a novel automatic background drift correction method was developed using local minimum values coupled with robust statistical analysis (LMV-RSA). The local minimum values in a chromatogram were immediately selected after data collection and then organized as a new baseline vector. A robust statistical strategy was employed to detect the outlier data points that may correspond to the unseparated peaks; these points were subsequently estimated using linear interpolation to obtain a new baseline vector. This procedure was repeated until convergence. The obtained baseline vector was finally expanded into the original chromatogram, and linear interpolation was employed to estimate the background drift in the chromatogram. The principle of the proposed method was completely investigated using a simulated chromatographic dataset. The new method was subsequently demonstrated using a complex gas chromatographic dataset. The performance of the proposed method was compared with two classic baseline correction techniques, MPLS and MWMV. The proposed method was employed for background drift correction in a LC-QTOF metabolic dataset for the first time to eliminate background drift; a comparison with orthogonal subspace projection was also provided.

2. Methodology

The objective of chromatographic background drift correction is to estimate baselines without removing useful information corresponding to chromatographic peaks. The development of an automatic strategy to identify the starting positions of chromatographic peaks and instrumental noise in the chromatogram entails difficulty. However, Yaroshchyk and Eberhardt [23] indicated that the majority of the minimum values in the chromatogram are not affected by chromatographic peaks; thus, the minimum values were used in our proposed method.

Fig. 1 depicts the workflow of LMV-RSA, which consists of three stages, namely, initialization (stage 1), iterative optimization (stage 2), and estimation of background drift (stage 3). The succeeding sections elaborate on each stage.

2.1. Initialization stage

The local minimum values were initially detected and their positions in the original chromatogram were recorded. A data point c_i was recognized as the local minimum for as long as the following equation was satisfied:

$$c_{i-1} > c_i \text{ and} c_i < c_{i+1} \quad (1)$$

where c_i is the ith data point in the chromatogram and c_{i-1} and c_{i+1} are the $i-1$th and the $i+1$th data points in the chromatogram, respectively. Fig. 2A presents the extracted local minimum values, which were organized as a new minimum vector (Fig. 2B). The data points in the new minimum vector can be classified into chromatographic peak points (minimum points in the peak elution range) and instrumental noise points. The chromatographic peak points can be more readily recognized from noise. The following part aims to separate these chromatographic peak points from instrumental noise.

2.2. Iterative optimization stage to eliminate outliers

As depicted in Fig. 1, the iterative optimization stage consists of two independent parts. This first part (left part in Stage 2 of Fig. 1) is based on a moving window strategy and is the core principle of our method, whereas the second part (right part in Stage 2 of Fig. 1) is complementary. The estimated local minimum vector is a combination of results from two parts. In this work, the moving window strategy is discussed in more detail.

To detect outliers (chromatographic peak points) in the minimum vector ($\boldsymbol{x}$), the noise level in the minimum vector was first

Please cite this article in press as: H.-Y. Fu, et al., Simple automatic strategy for background drift correction in chromatographic data analysis, J. Chromatogr. A (2016), http://dx.doi.org/10.1016/j.chroma.2016.04.054