Journal of Chromatography A, 1254 (2012) 51–61

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

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

Effect of background correction on peak detection and quantification in online comprehensive two-dimensional liquid chromatography using diode array detection

Robert C. Allen [a], Mallory G. John [a], Sarah C. Rutan [a,*], Marcelo R. Filgueira [b,c], Peter W. Carr [b]

[a] Department of Chemistry, Virginia Commonwealth University, Richmond, VA 23284-2006, USA
[b] Department of Chemistry, University of Minnesota, Minneapolis, MN 55455, USA
[c] Div Quim Analit, Fac Ciencias Exactas, Univ Nacl La Plata, 47 y 115, La Plata RA-1900, Argentina

ARTICLE INFO

Article history:
Received 25 April 2012
Received in revised form 5 July 2012
Accepted 6 July 2012
Available online 20 July 2012

Keywords:
Comprehensive two-dimensional chromatography
Two-dimensional peak detection
Two-dimensional background correction
Two-dimensional quantification

ABSTRACT

A singular value decomposition-based background correction (SVD-BC) technique is proposed for the reduction of background contributions in online comprehensive two-dimensional liquid chromatography (LC × LC) data. The SVD-BC technique was compared to simply subtracting a blank chromatogram from a sample chromatogram and to a previously reported background correction technique for one dimensional chromatography, which uses an asymmetric weighted least squares (AWLS) approach. AWLS was the only background correction technique to completely remove the background artifacts from the samples as evaluated by visual inspection. However, the SVD-BC technique greatly reduced or eliminated the background artifacts as well and preserved the peak intensity better than AWLS. The loss in peak intensity by AWLS resulted in lower peak counts at the detection thresholds established using standards samples. However, the SVD-BC technique was found to introduce noise which led to detection of false peaks at the lower detection thresholds. As a result, the AWLS technique gave more precise peak counts than the SVD-BC technique, particularly at the lower detection thresholds. While the AWLS technique resulted in more consistent percent residual standard deviation values, a statistical improvement in peak quantification after background correction was not found regardless of the background correction technique used.

1. Introduction

With recent advances in comprehensive two-dimensional separations, as well as the inherent complexity (*i.e.*, multidimensionality) of the data, comes the need for software and algorithms with the ability to characterize the separation system and generate reproducible quantitative results. System performance is typically measured by evaluating the effective utilization of the separation space [1,2]. One way to do this is to count the apparent number of peaks observed for a given sample [2]. A comparison of peak counts for different separation conditions (*i.e.*, different stationary phase and mobile phase combinations) yields performance information about the different systems. For complex biological samples, this comparison can be a daunting task, due to the high degree of peak overlap in both chromatographic dimensions. Furthermore, to preserve the first dimension's chromatographic resolution, each peak appearing in first dimension must be sampled several times thereby resulting in the same species appearing in several successive second dimension chromatograms [3]. Therefore, to accurately count the number of peaks present in the final two dimensional (2D) chromatogram, the peaks corresponding to the same species need to be properly identified and merged to yield a single 2D peak. Two methods have recently been proposed for detecting second dimension peaks and then merging the consecutive peaks into a single 2D peak.

The first method was proposed by Peters et al. [4]. It relies on the use of the first derivative to determine where a second dimension peak begins and ends and the second derivative to determine the retention time of the peak. After each second dimension chromatogram has been analyzed, the technique then merges the peaks based on two requirements; unimodality and overlap. The unimodality criterion assures that each 2D peak will only exhibit a single maximum. The overlap criterion assures that the second dimension peaks being merged possess a minimum degree of overlap in their start and stop times. Stevenson et al. modified this method by squaring the normalized signal of each second dimension chromatogram [5]. Squaring each second dimension peak increases its size, reduces the noise, and improves resolution. Consequently, the technique is able to detect peaks at lower intensities than is otherwise possible.

* Corresponding author. Tel.: +1 804 828 7517; fax: +1 804 828 8599.
E-mail address: srutan@vcu.edu (S.C. Rutan).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chroma.2012.07.034

The second method was proposed by Reichenbach et al. [6,7]. It translates a two-dimensional chromatogram at a single wavelength into a pixelated picture. Peak detection is then carried out using the drain algorithm [8], a modified version of the inverted watershed algorithm [9]. Unlike the derivative method, which is applied in a single dimension, the retention times of second dimension peaks are determined based on the relative intensities of the pixels in two dimensional space. A collection of pixels is determined to be a peak (also known as a blob), if a sufficient number of pixels (the minimum number being defined by the user) whose sum of intensities exceed a user defined minimum volume around the maximum pixel. Unimodality is imposed during this process by ensuring that the selected pixels around a maximum pixel do not contain any localized maxima. Assuming that shifting between second dimension chromatograms is corrected prior to peak detection, Latha et al. determined that this procedure correctly merges the peaks appearing in the consecutive second dimension chromatograms [8].

Subsequent to peak detection, peak quantification is carried out. One of the most common methods for quantifying peaks is the simple summation of the areas of the consecutive second dimension peaks belonging to a single analyte [10–12]. While not explicitly implemented in the work of Stevenson et al. [5], this is obviously possible once the peak boundaries are determined, provided the background contributions to the signal are accounted for appropriately. As long as the baseline is linear, this is straightforward. In the drain algorithm, the intensities of the pixels assigned to each parent peak are summed to quantify the peak in question. However, an issue arises with the drain algorithm that can affect quantification. As Vivó-Truyols and Janssen [13] and Bailey et al. [14] point out, the watershed algorithm allows for non-continuity across a second dimension chromatogram, illustrated by Fig. 2 in [13]. The non-continuity is a result of the watershed algorithm selecting non-adjacent parts of a second dimension chromatogram on the edge of a merged first dimension peak. While this non-continuity does not affect peak detection, non-continuity may impact the quantification of the merged first dimension peaks.

The performance of both peak detection and peak quantification algorithms can be severely compromised by the presence of irregular background contributions. Thus, a critical step in developing reliable peak detection and quantification algorithms is often the effective removal of background signals. Background noise is a more serious problem with ultra-fast LC × LC with diode array detection (DAD) than with comprehensive two-dimensional gas chromatography (GC × GC) with TOF-MS due to huge baseline signals generated in optical detectors by the very fast and large changes in the refractive index of the effluent during gradient elution LC × LC [15].

There are two main features of the background that must be addressed when correcting ultra-fast LC × LC chromatograms. First, there is a gradually increasing ridge that appears at the beginning of the second dimension chromatograms over the course of the first dimension gradient; it is caused by the difference in the composition of the first dimension effluent and the initial composition of the second dimension gradient. We call this the injection ridge (see Fig. 1). This ridge increases in size as the first dimension effluent composition becomes increasingly different from the initial composition of the second dimension gradient. Second, there is another ridge of constant height that appears at the end of the second dimension gradient; this is due to the rapid change in the second dimension eluent composition back to its initial value. We call this the re-equilibration ridge (see Fig. 1). These features appear at all wavelengths and are due to changes in the refractive index in the detector cell which form a dynamic lens in the detector cell [16,17], leading to an increase in the apparent absorbance. The magnitude of these features depends on the gradient time in the second dimension [18]; at the higher flow rates used in ultra-fast LC × LC-DAD, these features are greatly increased.

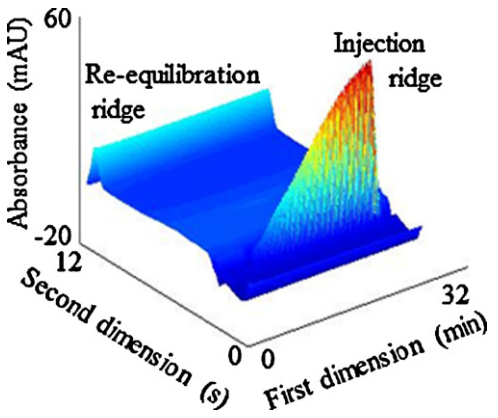

Fig. 1. LC × LC-DAD chromatogram of a blank sample at 220 nm.

Several approaches to background removal for comprehensive two-dimensional chromatograms have been described in the literature. Reichenbach et al. [19] performed a background removal on digital images of GC × GC-FID data. This was done by estimating the background level for each second dimension chromatogram across the chromatographic image based on structural and statistical properties of the GC × GC-FID data, and subsequent subtraction of the background level left a chromatogram in which the peaks were above the non-zero mean background [19]. Zhang et al. [20] suggested a chemometric technique for subtracting the background drift from a trilinear data set. However, this technique assumes that the data are completely trilinear, which is not necessarily the case for LC × LC-DAD data due to shifting between consecutive second dimension chromatograms. In addition, due to the typically large number of compounds found in complex ultra-fast LC × LC-DAD samples compared to the experimental data used by Zhang et al., the determination of the correct number of components required for trilinear decomposition of the sample is difficult. Porter et al. [21] used a similar chemometric background correction approach based on a combination of multivariate curve resolution-alternating least squares (MCR-ALS) and parallel factor analysis (PARAFAC), but this approach could only be used on small sections of the chromatogram.

The goal of this work is to compare the effectiveness of several background correction strategies, specifically with respect to the accuracy and precision of peak detection and peak quantification. We also propose a new technique for bilinear background correction for the three-dimensional data sets produced by a LC × LC-DAD system, which we call singular value decomposition-based background correction (SVD-BC). This is achieved by carrying out SVD on a blank LC × LC-DAD run to create a model for the background contribution. This blank signature is fit to the sample chromatogram over a wavelength range where no analytes are expected to absorb. This technique is similar to the asymmetric weighted least squares (AWLS) approach developed by Boelens et al. [22]. The primary differences between the two techniques are that the proposed technique is non-iterative and assumes that all of the spectral information at the higher wavelengths is due to background signals. A simple background subtraction technique is also investigated. Upon removal of the background signal, two automated peak detection methods (Stevenson et al. and Reichenbach et al.) were used to determine whether or not peak detectability was compromised or improved after application of the background removal algorithm. The peaks were also detected manually to serve as a reference method. The background correction techniques were evaluated based on their ability to remove the injection and re-equilibration

ridges while preserving the manual peak count. In addition, the impact of background correction on the quantification of selected standard replicate peaks was examined.

2. Materials and methods

All calculations and data analysis were carried out using Matlab version 7.12.0.635 (R2011a, Mathworks, Inc., Natick, MA) on a Lenovo Win 7 PC laptop with an Intel Core i5-2410 M @ 2.30 GHz and 6.00 GB of RAM except where noted.

2.1. Chemicals

Chromatographic grade water was from Sigma–Aldrich (St. Louis, MO) and acetonitrile was obtained from J.T. Baker (Phillipsburg, NJ, USA). Reagent grade perchloric acid was purchased from Mallinkrodt Baker (Paris, KY, USA). All materials were used as received. All mobile phases were prepared gravimetrically (±0.01 g) and used without any further filtration.

2.2. Samples and LC × LC-DAD instrumentation

The data consisted of two sets of samples: five replicates of a standard mixture containing 22 analytes [23] and eleven replicates of an extraction of maize seeds [24]. Four replicates of a blank sample were also provided for use in generating the background correction models. The mobile phases used for both the first and second dimensions were a 10 mM perchloric acid solution in water for mobile phase A and neat acetonitrile for mobile phase B. A 4.6 mm × 100 mm Zorbax SB-C3 column packed with 3.5 μm particles was used in the first dimension. The first dimension gradient was 0% B at 0 min, 56% B at 24.5 min, 0% B at 24.51 min, with a total analysis time of 32 min. The first dimension flow rate was 0.5 mL/min and the column was maintained at a temperature of 40 °C. A 2.1 mm × 33 mm in-house packed column with ZirChrom CARB 3.0 μm particles was used in the second dimension. The second dimension gradient was 0% B at 0 min, 100% B at 0.15 min, 0% B at 0.16 min, with a total analysis time of 0.2 min. The second dimension flow rate was 3 mL/min and the column was maintained at a temperature of 110 °C. The instrument configuration was the same as previously reported by Filgueira et al. in the split mode [23].

2.3. Background correction techniques

The standard mixture and maize replicates were examined using the three background correction techniques and without background correction (treated as a control group). The first technique (Direct) subtracted the average blank from each of the "unknown" samples (standard mixture and maize). The average blank was generated by averaging the chromatograms of the four blank samples for each wavelength collected.

The second technique was the SVD-BC method based on a doubly truncated **V** matrix ($\tilde{\tilde{\mathbf{V}}}$) from SVD, Eq. (1) [25].

$$\mathbf{X} = \mathbf{USV}^T \tag{1}$$

where **X** is an unfolded blank sample of dimensions *I* (number of collected time points) by *J* (number of collected wavelengths), **U** is the score matrix normalized to length one (*i.e.*, abstract chromatograms), **S** is a square matrix with the diagonal containing the singular values (the larger the singular value for a given component, the greater the impact of that component on **X**), and **V** is the loading matrix (*i.e.*, abstract spectra). The double truncation is accomplished by using only selected rows from the **V** matrix (440 nm to the highest wavelength collected) and only the first *N* columns. Before the procedure used to determine *N* will be described, the remainder of the SVD-BC technique will be explained. To estimate the contribution of the background to the sample chromatogram, a truncated $\tilde{\mathbf{X}}_{\text{sample}}$ matrix (consisting of *I* rows and the columns corresponding from 440 nm to the highest wavelength collected) was fit to the doubly truncated $\tilde{\tilde{\mathbf{V}}}$ matrix, as follows:

$$\mathbf{B} = \tilde{\mathbf{X}}_{\text{sample}}(\tilde{\tilde{\mathbf{V}}}^T)^{\dagger} \tag{2}$$

where **B** is a regression matrix of weighting coefficients for the contributions of the abstract spectra to the data set, and the superscript † indicates the pseudoinverse of the $\tilde{\tilde{\mathbf{V}}}^T$ matrix. To generate the background corrected sample matrix, the following calculation is carried out:

$$\mathbf{X}_{\text{sample,bc}} = \mathbf{X}_{\text{sample}} - \mathbf{B}\tilde{\mathbf{V}}^T \tag{3}$$

where $\mathbf{X}_{\text{sample,bc}}$ indicates the background corrected chromatogram and $\tilde{\mathbf{V}}^T$ is the singly truncated (to contain the first *N* columns) **V** matrix from Eq. (1). The number of components, *N* was determined by carrying out the background correction method described above for a blank chromatogram, with the number of components of interest set at one, and incremented until the standard deviation of the resulting background corrected chromatogram reached its minimum standard deviation. This procedure was carried out using the four blank samples. The rationalization for this approach was that a reduction in the background fluctuations would result in a decrease in the standard deviation of the background signal. *N* was determined to be three for three of the blanks; the third blank only required two components to reach its minimum standard deviation. An augmented blank matrix was then created by stacking the four blanks (so that the dimensions were 4*I* by *J*), and SVD was then carried out on this blank. The resulting $\tilde{\mathbf{V}}^T$ matrix was then used as the basis for the background correction of the standard mixture and maize replicates.

The third technique, AWLS, was developed by Boelens et al. [22]. This technique uses the same $\tilde{\mathbf{V}}^T$ matrix as the proposed background correction method. However, instead of truncating the $\tilde{\mathbf{V}}^T$ matrix to contain only the longer wavelengths, the AWLS fitting method generates a chromatographic model of the background based on the entire wavelength range. Using an asymmetric system of weights (positive values were given a weight of 0.001 for this study), fitting is performed on the sample until only positive intensities remain.

2.4. Peak detection methods

Prior to analysis, the standard mixture and maize chromatograms were cropped from 3 min to 28 min in the first dimension and 1.75 s to 10.9375 s in the second dimension. This cropping scheme was based on the retention time of the dead volume marker, the end of the second dimension gradient, and the end of the first dimension gradient. Two automated peak detection methods were used in this study. The first was the derivative based detection algorithm of Stevenson et al. (distributed by Hearn Scientific software, Melbourne, VIC, Australia) and implemented in Mathematica 8 (Wolfram Research, Champaign, IL) [5]. The cropped data sets were imported from Matlab into Mathematica by reshaping the two-dimensional chromatogram at a single wavelength (220 nm) into a three-column matrix. The first column contained the first dimension time scale (in increments of 0.2 min), the second column contained the second dimension time scale (in increments of 0.0125 min), and the third column contained the absorbance intensities (in mAU).

Once the selected files were imported into Mathematica, the parameters for the analysis were adjusted for each individual chromatogram. Parameter names (Capitalized) are as specified in the

Table 1
Thresholds (mAU) used for peak detection of the maize replicates.

Method	High	Low
Derivative based detection		
No background correction	6.700	5.925
Direct	3.540	2.800
SVD-BC	5.180	2.550
Asymmetric	2.120	1.400

Method	High without baseline correction	High with baseline correction	Low without baseline correction	Low with baseline correction
Drain algorithm detection				
No background correction	6.740	4.620	5.925	2.725
Direct	4.240	4.580	2.800	2.775
SVD-BC	5.220	4.640	2.600	2.625
Asymmetric	2.220	2.200	1.650	1.525

Stevenson Mathematica code. The cut time (CutTime) was set to −1 in order for the entire dataset to be analyzed; the dead time (VoidTime) was set to 0.01; the threshold criteria was set to 2, indicating that the threshold value is equivalent to the value of the detection threshold (thrh2), which was manually determined for each replicate. Due to a shifting baseline as each replicate was injected throughout the series, the threshold was set separately for each individual standard replicate. Two detection thresholds were established for each standard replicate. The higher threshold was based on detecting 21 visually confirmed peaks present in all of the standard replicates. The lower threshold was based on detecting 23 visually confirmed peaks present in four of the five standard replicates. Two low intensity peaks were not visually detected in the first standard replicate. The thresholds used for the maize replicates were the mean high and low thresholds from the standard replicates, as shown in Table 1. The peak maximum was determined from the first derivative *x*-intercept (PeakMethod), and was set to 1; the first derivative threshold (Thrfd) was set to 0.01, and the second derivative threshold (Thrsd) was set to 0.01 also. The peak region overlap threshold (ThrOV) was set to 0.5; this deviates from the suggested setting according to Peters et al. [4]. This change in overlap criteria was necessary to ensure that some of the lower intensity peaks in the standard replicates were not combined into a 2D peak. The peak maximum profile (PeakMaxProfile) was set to 1, which analyzes the peak maximum profile (plot of first dimension retention time as a function of detector response for all first dimension peaks) and separates peaks if there are valleys in the profile. The (PeakMaxInterpolation) was set to 1 to extrapolate the data points between the points in the (PeakMaxProfile) from the two dimensional data set and again looked for valleys between peaks in the first dimension.

The second automated peak detection algorithm used in this paper was developed by Reichenbach et al. [6] and is implemented in the LCImage software. The version of LCImage (LC Image, LLC, Lincoln, NE) used in this study was vR2.2. The cropping of the chromatograms and the selection of appropriate thresholds followed the same procedure as previously described for the derivative method. In addition to selecting appropriate thresholds the following parameters were established for peak detection; the sigma of the Gaussian smoothing used by LCImage was set to 0.1 in the first dimension and 1.0 in the second dimension; the minimum peak area was set to 15 (*i.e.*, the number of pixels required for the detected "blob" to be recorded); the minimum peak volume was 0 (*i.e.*, the sum of the detected pixels); and the minimum peak reference was set to 'absolute' (detection was based on set mAU thresholds). The same parameters were applied to both the standards and maize replicates.

In addition to counting the number of detected peaks after background correction, LCImage's built-in baseline correction option was also utilized. The baseline correction option within the LCImage software works as follows [19]: first, each second dimension chromatogram is divided into halves, called strides. Second, the baseline correction option determines the five pixels with the lowest intensity in each stride. Third, the mean and median of the two pixels adjacent to each of the chosen five pixels are calculated to generate the local baseline intensity. Fourth, the pixels that fall within the selected baseline intensity range are identified in each stride. Fifth, interpolation is performed between each pixel using a piecewise cubic spline and the resulting interpolated image is subtracted from the chromatogram.

As a comparison to the automated peak detection methods, manual peak counting was carried out on three of the maize sample replicates (the first, seventh, and eleventh respectively) without background correction being performed. The first and second dimension retention times of each second dimension peak was determined by visually inspecting each second dimension chromatogram. The visual inspection was carried out on 1.25 s increments of the 12 s second dimension chromatogram. This increment was chosen to reduce the influence of strongly absorbing peaks on the visual plot of the data. Unlike the automated peak detection methods, the starts and stops of each second dimension peak were not determined during manual peak counting. Instead,

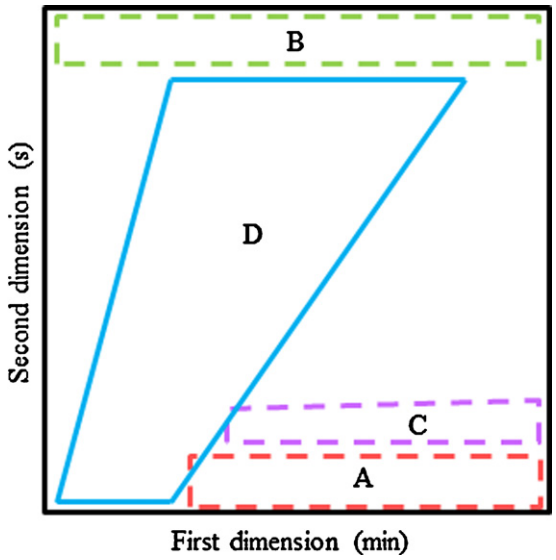

Fig. 2. Schematic diagram of the system of peak classification used in this paper. (A) Peaks detected in the region of the injection ridge. (B) Peaks detected in the region of the re-equilibration ridge. (C) Peaks generated by the background correction process. (D) Region where "real" peaks are presumed to exist.

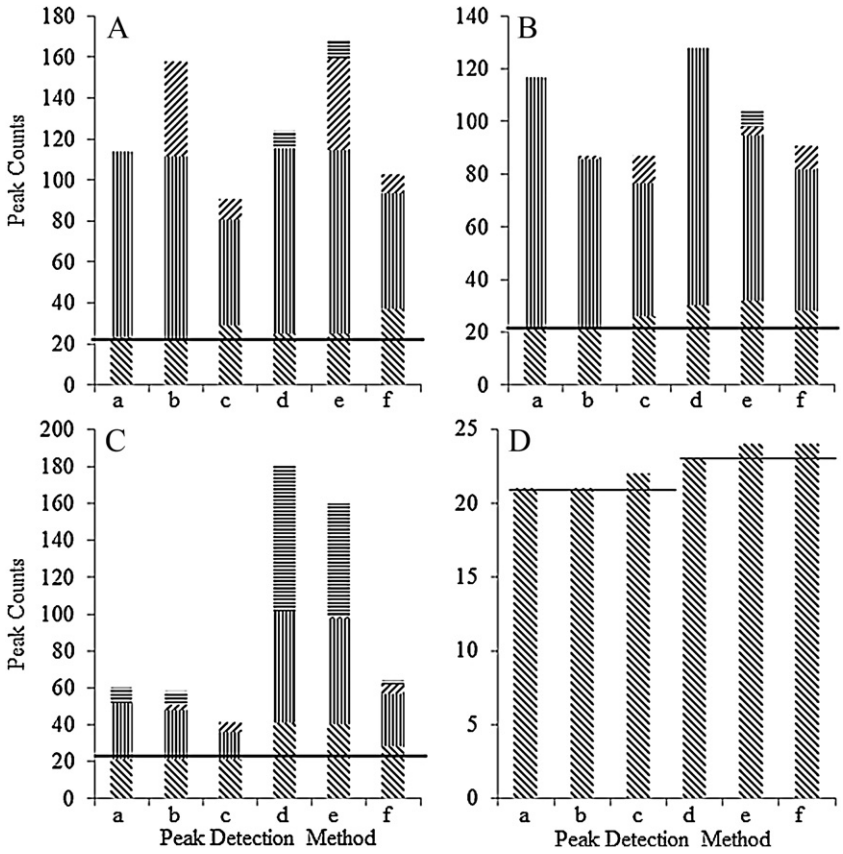

Fig. 3. The average number of peaks determined in each region shown in Fig. 2 for the standard samples. Various peak detection methods are denoted as: derivative detection—high threshold (a); drain algorithm detection without baseline correction—high threshold (b); drain algorithm based detection with baseline correction—high threshold (c). (d–f) are analogous to (a–c), except using the low threshold. Panel (A) no background correction, panels (B–D) after background correction using the Direct (B), SVD-BC (C), and AWLS (D) techniques respectively. The bar patterns correspond to: ||| (peaks located in region A of Fig. 2), /// (peaks located in region B of Fig. 2), == (peaks located in region C of Fig. 2), \\\ (peaks located in region D of Fig. 2). The lines in (A–D) are the true peak totals, 23 and 21, upon which the detection thresholds were based, for the high and low detection threshold respectively.

the prospective second dimension peaks were merged based on the unimodality criterion and a maximum allowable degree of retention time shift between consecutive second dimension maxima. Since the degree of peak overlap is not easily determined for manual peak counting, a new constraint was used to merge prospective second dimension peaks. This constraint was based on limiting the maximum allowed retention time drift between prospective second dimension peaks to a user defined tolerance. To gauge the impact of the constraint on the merging process, three tolerance values (0.0625s, 0.125 s, and 0.1875s) were used.

2.5. Adjusting peak counts

After the total number of peaks was determined, the chromatogram was divided into regions, as shown in Fig. 2. The peaks that were close to the injection (Fig. 2, regions A and C) and re-equilibration (Fig. 2, region B) ridges were removed from the peak total. Removal of the peak counts in these regions resulted in a peak count that could reasonably be expected to represent true chromatographic peaks (Fig. 2, region D).

2.6. Quantification methods

Six well resolved peaks were selected for quantification in the standard samples. A window surrounding each peak was chosen so that the peak of interest was fully present in the window (including any tailing); only one peak was present in each window. Quantification was accomplished by both manual integration assuming a linear baseline and by locating the respective “blob” in the “blob” detection table of LCImage. Stevenson et al. used a power function in their peak detection procedure [5]; therefore quantification was not performed using their technique. Manual integration was accomplished by plotting the sequence of second dimension chromatograms for the peak of interest, manually drawing the baseline of the peaks present in the sequential chromatograms, and summing up the area of the peaks after removal of the baseline.

3. Results and discussion

3.1. Standard replicate peak counts

The average peak counts (after being rounded to the nearest integer) given by the two detection methods before background correction and after background correction are illustrated in Fig. 3. The impact of the background correction techniques on each peak count is described below. The different classes of peaks mentioned in Fig. 3 are defined in Fig. 2, where region A includes peaks detected from the injection ridge, B includes peaks detected from the re-equilibration ridge, C includes peaks generated during background correction (artifact peaks), and D includes the estimated number of true peaks in the sample (real and artifact peaks).

3.1.1. Analysis of total peak counts

The average total number of peaks found for the five standard samples is shown in Fig. 3 as the cumulative height of all bars. The actual peak counts used to generate Fig. 3 are shown in

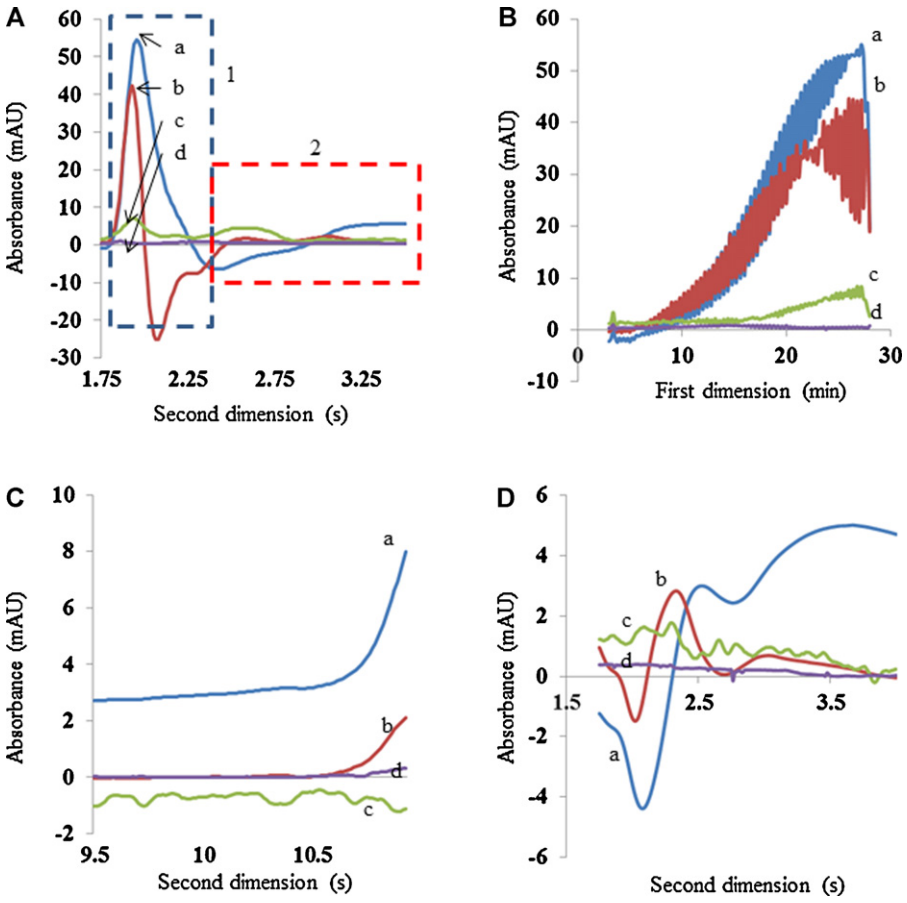

Fig. 4. Impact of the various background correction techniques on the background artifacts (data shown for the third standard replicate at 220 nm): (A) the injection ridge (box 1) and artifact peaks (box 2) from 1.75 to 3.5 s in the second dimension and at 25.6 min in the first dimension. Curve a is the raw signal (before background correction) and curves b–d show the corrected signals upon use of the Direct, SVD-BC, and AWLS techniques respectively; (B) the injection ridge from 3 to 28 min in the first dimension at 1.9375 s in the second dimension. The letters a–d have the same meaning as in A; (C) the re-equilibration ridge from 9.5 to 10.9 s in the second dimension at 25.6 min in the first dimension. The letters a–d have the same meaning as in A; (D) the solvent peaks from 1.75 to 4. s at 3.8 min (immediately after dead marker) in the first dimension. The letters a–d have the same meaning as in A.

Supplementary Table 1. With no background or baseline correction, the drain algorithm detected more peaks (Fig. 3A bars c and d) than did the derivative approach (Fig. 3A bars a and b). This is due to how the drain algorithm determines the presence of a peak. As shown in Fig. 9 of Ref. [8], this algorithm determines that a given pixel is a maximum when it is surrounded by enough smaller pixels to satisfy the minimum area requirement. After a pixel has been so identified, the volume of the pixels is determined by summing their intensities. If the calculated volume exceeds the minimum peak volume requirement, the maximum and the surrounding pixels are labeled as a peak. Clearly more peaks will be found without background correction than with background correction. However when the LCImage baseline correction is used, the volume of the detected peaks is reduced thereby preventing the drain algorithm from considering some smaller detected maxima as peaks. This results in a reduced total number of detected peaks (see Fig. 3A bars e and f). As expected the number of peaks detected increased upon decreasing the detection threshold (Fig. 3A bars b, d and f relative to bars a, c and e) although the increase varied with detection method. Except for the case of the drain algorithm without baseline correction (Fig. 3B bars c and d), the Direct technique produced peak totals very similar to the peak totals obtained when background correction was not performed. The primary difference between columns c and d in Fig. 3A and B is due to a large reduction in the B region peaks, which will be explained in more detail in Section 3.1.3. At the higher detection threshold level the SVD-BC technique (see Fig. 3C) also found fewer peaks when baseline correction was used than when not used (see Fig. 3A). However, when the lower detection threshold was used (Fig. 3C bars b and d), the number of peaks found did not differ much from the number found without baseline correction. Use of baseline correction in LCImage reduced the peak totals primarily by decreasing the number of artifact peaks; a further explanation will be given in Section 3.1.4. Finally, the AWLS technique gave peak totals approximately equal to the presumed true peak totals.

3.1.2. *Analysis of region A peaks*

The injection ridge (shown in Figs. 1 and 2, region A) caused by the solvent mismatch between the first and second dimensions was still largely present after background correction (see Fig. 4A (box 1) and B), by the Direct and to a lesser extent the SVD-BC technique. Consistent peak counts were obtained before background correction and after background correction by the Direct and AWLS techniques (see bars with vertical hatch marks in Fig. 3A, B and D) regardless of detection threshold or method. The number of injection ridge peaks detected with the SVD-BC technique changed when the threshold was lowered (see bars with vertical hatch marks in Fig. 3C). The SVD-BC technique was able to reduce the injection ridge so that the portion of the ridge occurring early in the first dimension was not detected at the higher threshold, Fig. 4B. However, because the SVD-BC technique was not able to

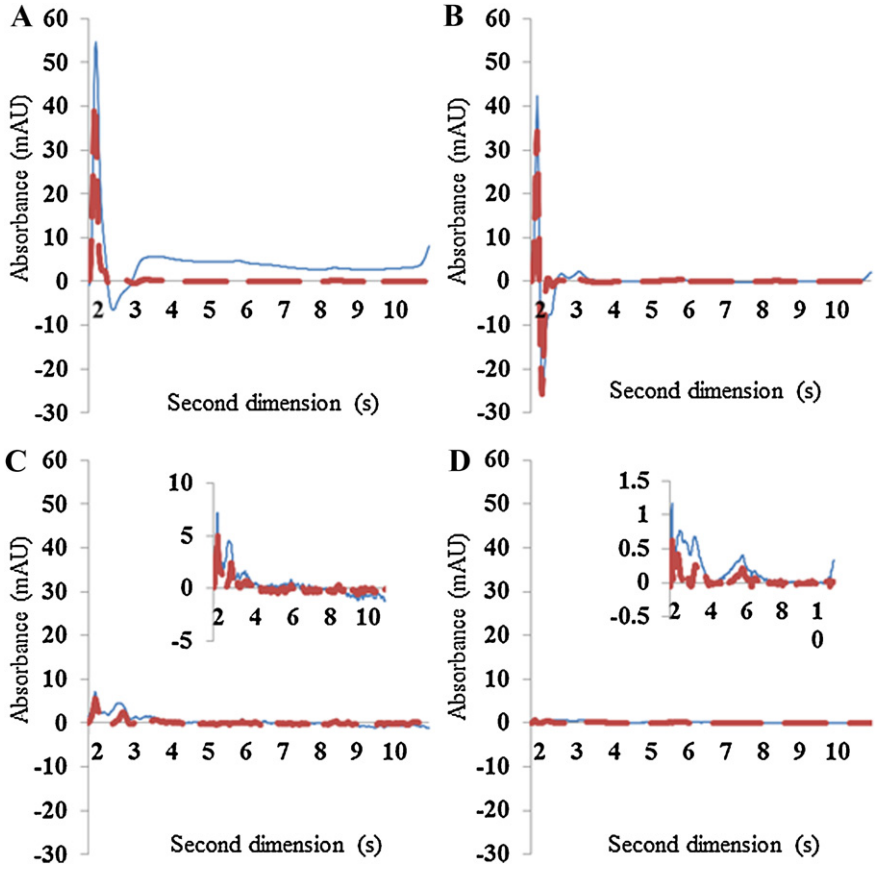

Fig. 5. Impact of the LC Image baseline correction options of the second dimension chromatogram from the third standard replicate at 220 nm and 25.6 min. The thin solid line chromatogram is before baseline correction and the thick dashed line chromatogram is after baseline correction. (A) No background correction (B–D); background removal by the Direct, SVD-BC, and AWLS techniques respectively.

completely remove the injection ridge, the lower threshold gave more peaks.

The use of the baseline correction option in LCImage did reduce the number of detected peaks by the drain algorithm. The effect of the baseline correction option on a selected second dimension blank chromatogram is shown in Fig. 5. For each of the techniques examined, the baseline correction option reduced the overall signal intensity requiring the high and low detection thresholds to be lowered, as shown in Table 1.

3.1.3. Analysis of region B peaks

Fig. 4C compares the ability of the three background correction techniques to remove the re-equilibration ridge. Detection of peaks in this region differed depending on the detection technique used. No peaks were found in this region by the derivative approach. In contrast, the drain algorithm detected peaks in this region both before and after background correction. As explained in Section 3.1.1, the drain algorithm determines that a peak is present when a sufficient number of pixels are located adjacent to a given maximum and the sum of the intensities of the pixels are greater than a given volume. Two factors then contribute to peak detection in this region. First, the minimum volume was set to zero for this study. While this ensured that all small peaks are detected, it also increases the chance that false peaks are counted. Second, since the background correction techniques did not completely remove the re-equilibration ridge, an adequate number of pixels with intensities greater than the threshold remained to satisfy the area requirement. In addition, the use of the baseline correction option found in LCImage increased the number of peaks detected in this region. The increased detection of peaks after baseline correction in LCImage was due to an apparent "wraparound" effect (a 2D peak is located on both the top and bottom of the chromatogram simultaneously) in the detection software. It is therefore possible, that the detected peaks are due to the injection ridge and not the re-equilibration ridge. It can be seen that the SVD-BC and AWLS methods reduced the background in this region the most, but a significant amount of noise was introduced with the SVD-BC method.

3.1.4. Analysis of region C peaks

Fig. 4A (box 2) shows that artifact peaks can be introduced by the background correction procedure. As shown by the bars with horizontal hatch marks in Fig. 3, the largest number of artifact peaks was introduced by the SVD-BC technique, Fig. 3C; this is true for both detection methods. A possible explanation is that the chromatographic model generated during the SVD-BC technique did not adequately represent the background at the shorter wavelengths. This result indicates that that background models for the Direct and AWLS techniques are better than the SVD-BC techniques at shorter wavelengths. Curiously, the drain algorithm detected a smaller number of artifact peaks for the SVD-BC technique. The number of pixels surrounding the maxima is probably too small to satisfy the minimum area requirement. The drain algorithm did detect artifact peaks after application of the Direct technique but the use of the baseline correction option almost completely eliminated of the artifact peaks. The impact of the baseline correction option on the chromatograms is shown in Fig. 5. The baseline correction option was able to generate an almost zero-baseline immediately after the injection ridge in region C. Also, the presence of the bar with the horizontal hatch marks in Fig. 3A for columns d and e was due to a

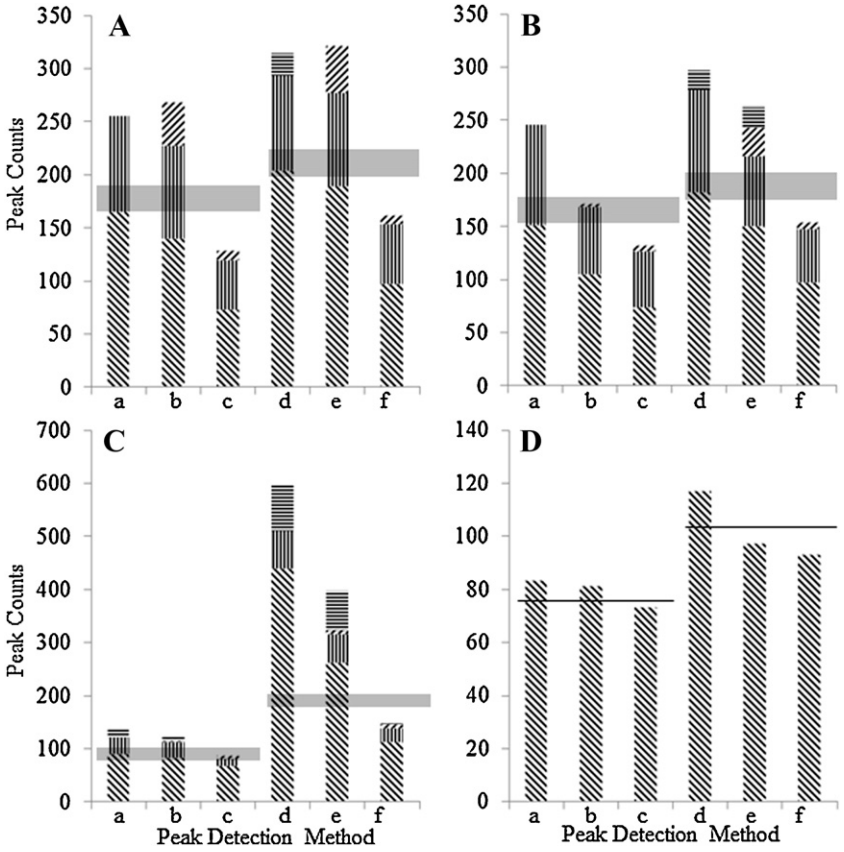

Fig. 6. The average number of peaks determined for the maize sample in each region shown in Fig. 2 after background correction for (A) no background correction, (B–D) background correction using the Direct, SVD-BC, and AWLS techniques respectively. The patterns and letters (a–f) have the same meaning as previously described in Fig. 3. The shaded boxes in (A–D) represent the range of estimated true peak totals obtained by manual peak detection and using a 0.125 s retention time drift tolerance.

single standard replicate. The low detection threshold used for this replicate was the lowest. As such, additional peaks would have been detected in the other replicates had the low detection thresholds been lowered further. These peaks are not real however. Instead they are a result of the poor baseline correction immediately after the injection ridge, illustrated by curve a) in Fig. 4A box 2.

3.1.5. Analysis of region D peaks

In addition to real solute peaks, two types of artifact peaks were observed in the D region of Fig. 2, depending on the background correction and detection method used. Although these peaks are clearly artifacts, we have decided to include them in the estimated true peak totals since this region of the chromatogram is occupied by real peaks in the maize samples. The first type is due to solvent peaks, Fig. 5D. Since the solvent peaks are not completely reproducible between second dimension injections, the Direct and SVD-BC background correction techniques overcompensate in this region of the 2D chromatogram and generate the artifactual peaks. Likewise, the LCImage baseline correction option is unable to fully model the background in this region and leads to artifact peaks being generated. The second type is due to the inability of the background correction techniques, primarily for SVD-BC but also AWLS to a lesser degree, to account for peak tailing. The most probable reason is the lack of incorporation of lower wavelengths into the background model.

The black lines in Fig. 3 represent the true estimated peak counts for the standards replicates. By looking at the \\\ bars, we can compare the ability of the background correction techniques to detect the actual peaks, once the artifact peaks are accounted for. Prior to background correction, the derivative based approach and LCImage before baseline correction detect approximately the correct number of peaks at the high threshold. However, when the detection threshold is lowered, the detection methods pick up the solvent artifact peaks mentioned previously. After baseline correction is applied, LCImage further inflates the number of detected artifact peaks in this region due to overcompensation. Of the three background correction techniques, only AWLS produces accurate peak counts at both detection thresholds. However, the LCImage baseline correction option once again introduces artifact peaks into the estimated true peak totals.

3.2. Maize samples

The true number of peaks present in the maize sample is not known. Therefore, visual inspection and peak counting was performed to establish a reference number of 2D peaks. First, a total of 441 ± 14 individual second dimension peaks were observed (measured for the 1st, 6th and 11th maize replicates). Next, peaks in consecutive second dimension chromatograms were merged to form 2D peaks. The total number of peaks detected depended on the retention time shift tolerance, as shown in Table 2. Instead of manually counting each time after background correction, the retention times before background correction were used to estimate the number of peaks above a given threshold. The Direct technique was found to retain approximately the same number of peaks as if background correction was not performed. While the AWLS technique resulted in the lowest peak count, this can be explained by the signal loss at the wavelength used for detection, illustrated in Fig. 7. Use of the AWLS technique resulted in a mean $25 \pm 2\%$ loss in intensity compared to the SVD-BC technique. This

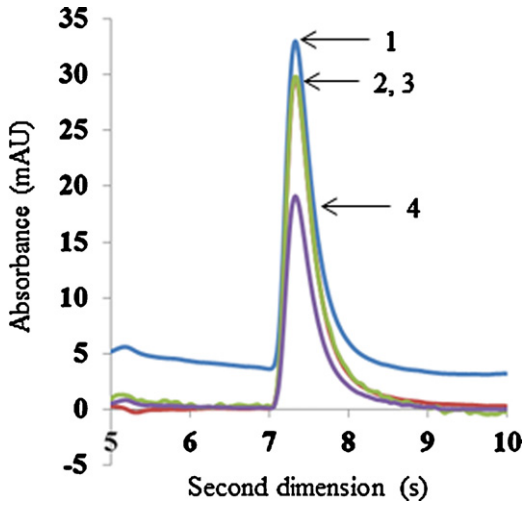

Fig. 7. The impact of the AWLS technique on signal intensity of a selected peak. The AWLS technique was found to retain on average 75 ± 2% of the net peak height compared to the SVD-BC method. A second dimension chromatogram of a peak from the first standard replicate at 220 nm and 15.4 min is shown. The blue line (labeled as 1) was not background corrected. The red line (labeled as 2) was background corrected by the Direct method. The green line (labeled as 3) was background corrected by the SVD-BC method. The purple line (labeled as 4) was background corrected by the AWLS method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

intensity loss was found to be systematic across a given peak. The reason why the SVD-BC technique resulted in a lower peak count is not immediately clear. A possible reason is that some of the peaks determined by manual detection were not real peaks but instead were due to fluctuations in the background. The peak counts at the lower threshold, however, were approximately the same as the Direct technique. The same trends were also seen when the automated peak detection methods were used, Fig. 6. The actual peak counts used to generate Fig. 6 are shown in Supplementary Table 2. The patterns of the counts in the different regions are similar to those observed for the standard replicates, Fig. 3.

In addition, none of the manual peak counts, Table 2, were the same as previously reported by Filgueira et al. [23]. The closest peak counts (prior to performing background correction) were from the largest allowed retention time drift. However, since this degree of retention time shifting was not observed in the standards replicates, the largest retention time drift tolerance will not be considered for comparison to the automated peak detection methods. Instead, the second column comprising the average retention time shift of the standards peaks with the addition of the upper bound of the standard deviation (0.125 s) will be considered as the true peak

Table 2
Average "peak" manual counts for the maize replicates.

Method	0.0625 s retention time drift tolerance	0.1250 s retention time drift tolerance	0.1875 s retention time drift tolerance
High threshold			
No background correction	214 ± 11	175 ± 10	164 ± 9
Direct	191 ± 8	160 ± 9	153 ± 7
SVD-BC	102 ± 9	92 ± 6	89 ± 7
AWLS	85 ± 3	76 ± 2	75 ± 1
Low threshold			
No background correction	256 ± 11	206 ± 11	193 ± 8
Direct	226 ± 10	184 ± 10	172 ± 7
SVD-BC	227 ± 11	187 ± 7	167 ± 18
AWLS	116 ± 6	103 ± 4	102 ± 4

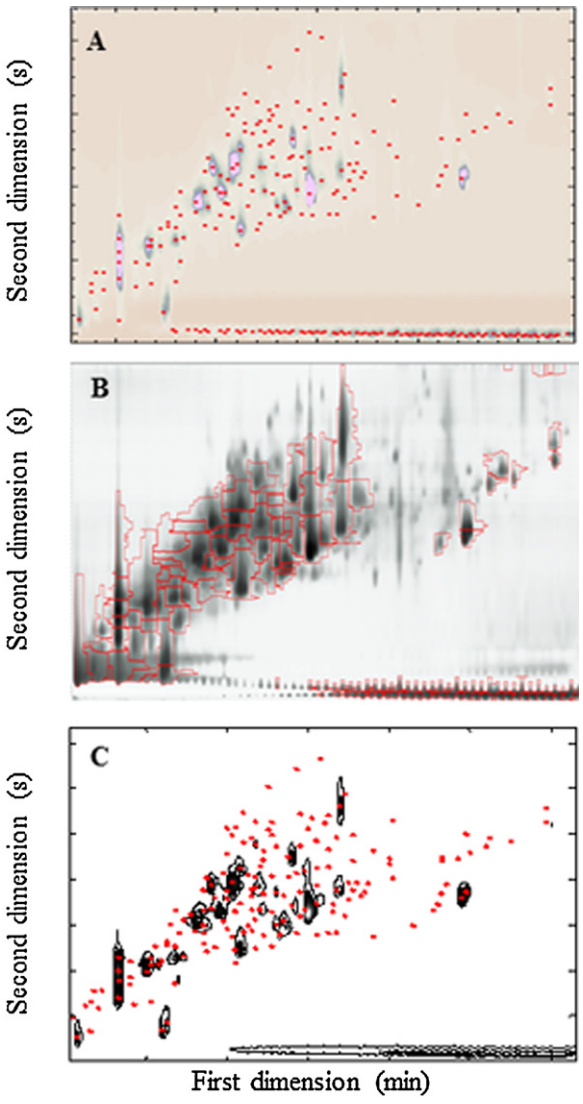

Fig. 8. Plots of the sixth maize replicate at 220 nm. Red circles or outlines indicate where a merged peak was detected at the higher threshold in Table 1. (A) Derivative approach using Stevenson's method [5]. (B) Drain algorithm with baseline correction using LC Image. (C) Manual peak detection using Matlab with a 0.0625 s retention time drift allowed between each second dimension chromatogram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

count for the maize sample. The transparent gray bars in Fig. 6 are the estimated true peak totals obtained from manual peak detection using the 0.125 s retention time drift tolerance. The \\\ bars from Fig. 6 can then be directly compared to the manual estimated peak totals. The Direct and AWLS techniques produced the closest estimated true peak totals to the peak counts found using a retention time drift tolerance of 0.125 s. However, as has been previously shown, AWLS produces a lower peak total compared to Direct. Also, after using the Direct technique, the derivative based detection method produced higher estimated true peak totals. The drain algorithm produced much lower estimated true peak totals, especially after baseline correction.

Fig. 8 illustrates the positions of the 2D peaks for the sixth maize replicate. The retention times of the 2D peaks are shown as circles in Fig. 8A and C. The boxes present in Fig. 8B are the regions designated to each 2D peak. Fig. 8A, generated using the code from Stevenson et al., is visually comparable to the manually detected peak positions, Fig. 8C. Some of the 2D peaks detected by the manual and derivative approaches were not detected by LC Image, Fig. 8B, after

Table 3
Percent relative standard deviations of six fully resolved peaks present in the standard samples.

Method	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6
Manual integration						
No background correction	1.08	1.96	1.46	1.90	1.62	2.09
Direct	0.66	2.17	1.60	2.20	1.67	1.37
SVD-BC	0.92	2.37	1.55	2.96	2.94	2.85
AWLS	2.09	4.18	1.71	2.92	0.39	2.24
Drain algorithm without baseline correction						
No background correction	14.34	7.74	6.16	4.18	3.00	3.32
Direct	14.89	22.46	10.70	11.59	4.80	6.71
SVD-BC	25.98	14.23	7.85	8.86	2.32	6.93
AWLS	1.95	6.78	1.76	4.69	1.27	1.76
Drain algorithm with baseline correction						
No background correction	1.45	2.28	1.59	2.37	1.31	1.76
Direct	1.38	2.12	1.33	1.62	1.43	1.20
SVD-BC	2.70	2.93	1.34	2.79	1.67	1.67
AWLS	2.18	3.58	4.38	3.95	1.54	1.42

baseline correction. These missing peaks are possibly due to two limitations of the method. Since LC Image treats the chromatogram like an image, the drain algorithm is unable to detect peaks that are co-eluting which do not possess an apparent maximum on the shoulder of a larger peak when looking at the entire chromatogram. In order to detect co-eluting peaks, the user is required to manually select suspected overlapped peaks and apply a built in deconvolution algorithm. This option was not tested during this study. Also, unlike the manual drift constraint and the derivative overlap criterion, a setting does not exist in LC Image to determine the degree of merging between consecutive second dimension peaks.

3.3. Quantification

The % RSDs of six selected peaks from the standard replicates are shown in Table 3. A series of Levene's tests to compare the variances [26] were conducted to determine if there is a statistical difference whether or not background correction was performed. No statistical differences between before and after background correction were found if manual integration was used as the quantification method. Likewise, statistical differences before and after background correction were not found if LCImage was used, regardless of whether the baseline correction option was employed. However, statistical differences were found when the variances from LCImage before baseline correction were compared to the variances from manual integration and LC Image after baseline correction. The likely reason for this statistical difference is once again the lack of a non-zero baseline after background correction. Since LCImage treats the chromatogram as a series of pixels, the magnitude of background noise from each pixel is included with that of the peak upon conducting quantification. An additional comparison between the variances from manual integration and LCImage after baseline correction did not result in a statistical difference. So, while the scale of the % RSDs are lower for the manual integration, a statistical improvement cannot be found if LCImage after baseline correction is used instead.

4. Conclusions

After background correction, none of the background correction techniques were able to provide a completely zero baseline. While the baseline was not zero, the AWLS technique did achieve a complete elimination of the injection and re-equilibration ridges. However, due to the loss of signal intensity, the reader is warned that peak detection may be compromised with the AWLS technique. As a result of the loss of signal intensity, a large reduction in the estimated true peak totals was observed for the AWLS. This reduction may prove problematic when attempting to assess the quality of the separation as smaller peaks may not be detected. In terms of quantification, the AWLS technique was found to produce the most consistent % RSD values regardless of the quantification approach used.

In comparison to AWLS, the Direct technique did not completely remove the injection and re-equilibration ridges. Unlike the SVD-BC technique, the Direct technique introduced a smaller number of artifact peaks which are easily accounted for based on their location in the chromatogram. The Direct technique also resulted in a larger peak count compared to the AWLS technique due to a lack of signal loss. This preservation of signal intensity allows for the detection of lower intensity peaks that may not be possible using the AWLS technique. The Direct technique can be improved if the baseline correction option found in LC Image is incorporated into the derivative based-detection method. This would allow for a further decrease in the threshold and possibly in the number of artifact peaks detected. However, the Direct technique is greatly dependent on the ability of the system to reproduce the baseline between samples.

The proposed SVD-BC technique is not recommended for background correction. While, SVD-BC did reduce or eliminate the background ridges, an unacceptable number of noise peaks were introduced with sufficient height to be detected at the lower detection threshold.

In terms of which detection method is preferable, we recommend the use of the derivative based technique to allow for the detection of co-eluting compounds that are not detected by the LCImage software. We also recommend a modification to allow for the user to define the maximum degree of retention time drift between consecutive second dimension peaks. However, we recommend the LC Image software with the baseline correction option implemented for quantification due to the ease of use.

Acknowledgements

The authors acknowledge financial support from NIH-GM-54585-13 (PWC & SCR) and NSF-CHE-0911330 (PWC &SCR), an ANPCyT-UNLP fellowship from Argentina (MF) and an Altria research fellowship (RCA) used to produce the results presented in this paper. The authors also would like to acknowledge Dr. Steven Reichenbach for providing to us a copy of his LC Image software and Dr. Paul Stevenson for his peak counting Mathematica script. We also wish to acknowledge funding from the Agilent Foundation and the gifts of columns from Agilent Technologies.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chroma.2012.07.034.

References

[1] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, Anal. Chem. 77 (2005) 62426.
[2] H. Gu, Y. Huang, M. Filgueira, P.W. Carr, J. Chromatogr. A 1218 (2011) 6675.
[3] R.E. Murphy, M.R. Schure, J.P. Foley, Anal. Chem. 70 (1998) 1585.
[4] S. Peters, G. Vivo-Truyols, P.J. Marriott, P.J. Schoenmakers, J. Chromatogr. A 1156 (2007) 14.
[5] P.G. Stevenson, M. Mnatsakanyan, G. Guiochon, R.A. Shalliker, Analyst 135 (2010) 1541.
[6] S.E. Reichenbach, M. Ni, V. Kottapalli, A. Visvanathan, Chemom. Intell. Lab. Syst. 71 (2004) 107.
[7] S.E. Reichenbach, V. Kottapalli, M. Ni, A. Visvanathan, J. Chromatogr. A 1071 (2005) 263.
[8] I. Latha, S.E. Reichenbach, Q. Tao, J. Chromatogr. A 1218 (2011) 6792.
[9] L. Vincent, P. Soille, IEEE PAMI 13 (1991) 583.
[10] D. Thekkudan, S.C. Rutan, P.W. Carr, J. Chromatogr. A 1217 (2010) 4313.
[11] O. Amador-Munoz, P.J. Marriott, J. Chromatogr. A 1184 (2008) 323.
[12] W.C. Siegler, B.D. Fitz, J.C. Hoggard, R.E. Synovec, Anal. Chem. 83 (2011) 5190.
[13] G. Vivo-Truyols, H.-G. Janssen, J. Chromatogr. A 1217 (2010) 1375.
[14] H.P. Bailey, P.W. Carr, S.C. Rutan, J. Chromatogr. A 1217 (2011) 4313.
[15] S.E. Reichenbach, P.W. Carr, D.R. Stoll, Q. Tao, J. Chromatogr. A 1216 (2009) 3458.
[16] C.E. Evans, J.G. Shabushnig, V.L. McGuffin, J. Chromatogr. 459 (1988) 119.
[17] C.E. Evans, V.L. McGuffin, J. Chromatogr. 503 (1990) 127.
[18] D.O. Hancock, C.N. Renn, R.E. Synovec, Anal. Chem. 62 (1990) 2441.
[19] S.E. Reichenbach, M. Ni, D. Zhang, E.B. Ledford Jr., J. Chromatogr. A 985 (2003) 47.
[20] Y. Zhang, H. Wu, A. Xia, L. Hu, H. Zou, R. Yu, J. Chromatogr. A 1167 (2007) 178.
[21] S.E.G. Porter, D.R. Stoll, S.C. Rutan, P.W. Carr, J.D. Cohen, Anal. Chem. 78 (2006) 5559.
[22] H.F.M. Boelens, R.J. Dijkstra, P.H.C. Eilers, F. Fitzpatrick, J.A. Westerhuis, J. Chromatogr. A 1057 (2004) 21.
[23] M.R. Filgueira, Y. Huang, K. Witt, C. Castells, P.W. Carr, Anal. Chem. 83 (2011) 9531.
[24] Y. Huang, H. Gu, M.R. Filgueira, P.W. Carr, J. Chromatogr. A 1218 (2011) 2984.
[25] R.W. Hendler, R.I. Shrager, J. Biochem. Biophys. Methods 28 (1994) 1.
[26] M.B. Brown, A.B. Forsythe, J. Am. Stat. Assoc. 69 (1974) 364.