Comparative analysis of peak-detection techniques for comprehensive two-dimensional chromatography

Indu Latha [a], Stephen E. Reichenbach [a,*], Qingping Tao [b]

[a] Computer Science and Engineering Department, University of Nebraska-Lincoln, Lincoln, NE 68588-0115, USA
[b] GC Image LLC, P.O. Box 57403, Lincoln, NE 68505-7403, USA

ARTICLE INFO

Article history:
Received 18 March 2011
Received in revised form 12 July 2011
Accepted 15 July 2011
Available online 26 July 2011

Keywords:
Two-dimensional chromatography
Comprehensive two-dimensional gas chromatography (GC×GC)
Chemometrics
Peak detection
Watershed algorithm
Two-step peak detection

ABSTRACT

Comprehensive two-dimensional gas chromatography (GC×GC) is a powerful technology for separating complex samples. The typical goal of GC×GC peak detection is to aggregate data points of analyte peaks based on their retention times and intensities. Two techniques commonly used for two-dimensional peak detection are the two-step algorithm and the watershed algorithm. A recent study [4] compared the performance of the two-step and watershed algorithms for GC×GC data with retention-time shifts in the second-column separations. In that analysis, the peak retention-time shifts were corrected while applying the two-step algorithm but the watershed algorithm was applied without shift correction. The results indicated that the watershed algorithm has a higher probability of erroneously splitting a single two-dimensional peak than the two-step approach. This paper reconsiders the analysis by comparing peak-detection performance for resolved peaks after correcting retention-time shifts for both the two-step and watershed algorithms. Simulations with wide-ranging conditions indicate that when shift correction is employed with both algorithms, the watershed algorithm detects resolved peaks with greater accuracy than the two-step method.

1. Introduction

Comprehensive two-dimensional gas chromatography (GC×GC) is a powerful technology for separating and analyzing compounds in complex samples. GC×GC data is processed to detect peaks and identify the associated compounds present in a sample. Typically, the goal of peak detection is to separately aggregate the data points belonging to each analyte peak. GC×GC peak detection popularly is performed by one of two approaches: the two-step algorithm [1] and the watershed algorithm [2,3]. In the two-step algorithm, traditional one-dimensional (1D) peak detection is employed on each secondary chromatogram, then detected 1D peaks are merged to form two-dimensional (2D) peaks. The watershed algorithm performs peak detection by operating on 2D neighborhoods, *i.e.*, in both retention-time dimensions simultaneously.

A recent study by Vivó-Truyols and Janssen [4] analyzed the effects of second-column retention-time shifts on the performance of 2D peak-detection techniques. Retention-time shift in consecutive secondary chromatograms in GC×GC may occur due to factors such as rapid temperature or pressure changes [5]. Rapid chromatographic changes can induce shifts that complicate data processing because all chromatographic peaks of a compound are expected to have the same retention-time. If the chromatography is rapidly varying (and cannot be improved to yield data without rapid peak shifts), then data processing and peak detection should account for retention-time shifts. Skov et al. [6] examined the nature and theory of retention-time shifts in GC×GC and described a method for shift correction based on cross-correlation for individual mass channels in adjacent secondary chromatograms. In the two-step algorithm, 1D peak merging typically accounts for retention-time shifts. For the watershed algorithm, retention-time shifts can be determined (*e.g.*, with cross-correlation [6]) and then corrected either by aligning the data before peak detection or equivalently by adjusting the 2D neighborhoods of the watershed algorithm to account for shifts.

In their analysis, Vivó-Truyols and Janssen [4] compared the effects of retention-time shifts on both the two-step and watershed peak-detection algorithms. For the two-step approach, 1D peak merging accounted for retention-time shifts, whereas no shift corrections were made with the watershed algorithm. Their results indicated that watershed algorithm failed at a higher rate due to uncorrected second-dimension retention-time shifts. However, their comparison of peak-detection performance was confounded by accounting for retention-time shift in the two-step algorithm but not accounting for retention-time shift in the watershed algorithm.

* Corresponding author. Tel.: +1 402 472 5007; fax: +1 402 472 7767.
E-mail addresses: ilatha@cse.unl.edu (I. Latha), reich@unl.edu (S.E. Reichenbach), qtao@gcimage.com (Q. Tao).
URLs: http://cse.unl.edu/ reich (S.E. Reichenbach), http://www.gcimage.com (Q. Tao).

0021-9673/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.07.052

This paper re-evaluates the performance of the two-step and watershed algorithms for retention-time shifts in the second-column separations when both employ shift correction. Simulated data is used to rigorously evaluate the peak-detection algorithms under controlled conditions with different levels of noise, retention-time shifts, and peak widths. The retention-time shifts in each simulated peak are corrected prior to both peak-detection algorithms for an unbiased comparison. These experiments demonstrate that when retention-time shift is corrected for both algorithms, the watershed algorithm detects peaks more accurately, over a wide range of conditions.

2. Peak detection in comprehensive two-dimensional chromatography

2.1. Two-step peak detection algorithm

The two-step algorithm builds on peak-detection methods used in traditional gas chromatography. In the first step, 1D peaks in each secondary chromatogram are detected by a 1D peak-detection algorithm. In the second step, a peak merging algorithm determines which of the detected 1D peaks should be merged to form 2D peaks [7].

The two-step algorithm typically uses two criteria to determine which 1D peaks should be merged: the overlap and unimodality criteria. The overlap criterion compares two adjacent 1D peaks in consecutive second-dimension chromatograms and determines if the peaks can be merged based on their retention-time overlap. Vivó-Truyols and Janssen [4] check the difference in second-dimension retention-time of two candidate 1D peaks and merge them if the difference is smaller than a predetermined threshold [8]. Setting that threshold allows the overlap criterion to account for retention-time shifts between secondary chromatograms. The unimodality criterion ensures that 2D peaks have only one local maximum. Initially, a single 1D peak in a secondary chromatogram is made part of a 2D peak. Then, 1D peaks in adjacent second-dimension chromatograms are added to the 2D peak if they satisfy the overlap and unimodality criteria [7].

2.2. Watershed algorithm

The watershed algorithm, originally used in image segmentation [9], was adapted for 2D chromatographic peak detection by Reichenbach et al. [2,3]. Conceptually, the algorithm initiates detection at the apex of a peak and iteratively adds all smaller neighbors until no more smaller points border the peak [10]. The watershed algorithm can be implemented with a priority queue to sort all data points. The largest data point is extracted and labeled first. This is followed by the next largest point in the queue. Each point drawn out of the queue is compared with its neighbors. If the neighbors are of equal or larger value, the extracted point is given the same label as its largest neighbor. However, if the data point is larger than its neighbors, it is given a new label to indicate that it is part of another peak. This procedure is repeated until the queue is empty.

Retention-time shifts can be corrected either prior to peak detection or in the watershed algorithm itself. For example, before peak detection, cross-correlation can be used to determine the alignment for shifted second-dimension chromatograms [6]. Cross-correlation measures the similarity between 1D chromatograms for possible shifts (within a specified range) and indicates the required shift correction. Then, prior to peak detection, shift correction can be applied to the data to align the secondary chromatograms. Alternatively, shifts can be corrected as a part of the watershed algorithm by using the shift correction (*e.g.*, as identified by cross-correlation) to adjust the 2D neighborhood for a point. Fig. 1(a) shows the standard 3×3 neighborhood (dark background) around a data point with value 1306 (gray background) at the center of a peak that has a skew of −1. In Fig. 1(b), the retention-time skew is corrected by shifting the 2D neighborhood by −1.

330	207	8
521	827	81
329	1306	328
83	824	519
7	208	331

a

330	207	8
521	827	81
329	1306	328
83	824	519
7	208	331

b

Fig. 1. Retention-time shift correction of −1 applied to a 2D neighborhood (indicated by dark gray background) in the watershed algorithm: (a) unshifted and (b) shifted.

3. Simulation experiments

3.1. Overview of simulation

In the experiments described here, simulated 2D chromatographic peaks are used to rigorously compare both peak-detection algorithms under controlled conditions for different levels of noise, retention-time shifts, and peak widths. A single, resolved peak is simulated by interval sampling a 2D Gaussian function with parametric retention-time peak widths. Fig. 2(a) displays an example 2D peak as a series of 1D peaks. Each 1D second-column peak models a 1D chromatogram at a characteristic retention-time for the first chromatographic column.

To experimentally evaluate the effects of retention-time shifts in the secondary chromatograms, a skew is introduced in the 2D peak that shifts each 1D chromatogram as shown in Fig. 2(b). The skew reduces the overlap of peak regions belonging to two consecutive 1D peaks in the 2D chromatogram. This simulates second-column retention-time shifts in comprehensive 2D chromatography, for instance, retention-time shifts due to an extreme thermal gradient that induces rapid chromatographic changes. The expression for parametric skew is:

$$\gamma(i) = (i - \mu_x)s \tag{1}$$

where $\gamma(i)$ is the shift introduced in each secondary chromatogram, i is the position of each data point along the x-dimension (the first-column separation), μ_x is the peak apex in the x-dimension, and s is the skew parameter that controls the shift.

The peak model is a Gaussian function (normalized to have unit integrated volume) subject to second-dimension skew:

$$f[i,j] = \frac{1}{2\pi\sigma_x\sigma_y} \int_{i-(1/2)}^{i+(1/2)} \int_{j-(1/2)}^{j+(1/2)} \times e^{-(((x-\mu_x)^2/2\sigma_x^2)+((y-\mu_y+\gamma(i))^2/2\sigma_y^2))}\, dx\, dy \tag{2}$$

where $f[i,j]$ is the measured signal intensity of the chromatographic peak at position $[i,j]$, i and j are the retention-time indices of the data array along the x and y dimensions respectively, μ_x and μ_y represent the peak's retention-time apex in the x and y dimensions respectively, $\gamma(i)$ is the shift in each secondary chromatogram, and σ_x and σ_y parameterize the width of the peak along the x and y dimensions respectively. The units for skew and peak widths are the data array index intervals (*i.e.*, the data sampling intervals of retention times).

The array size that is required to contain the simulated peak is determined by the peak-width standard deviations, σ_x and σ_y, and the skew parameter, s:

$M = \lceil 9\sigma_x + 2 \rceil$

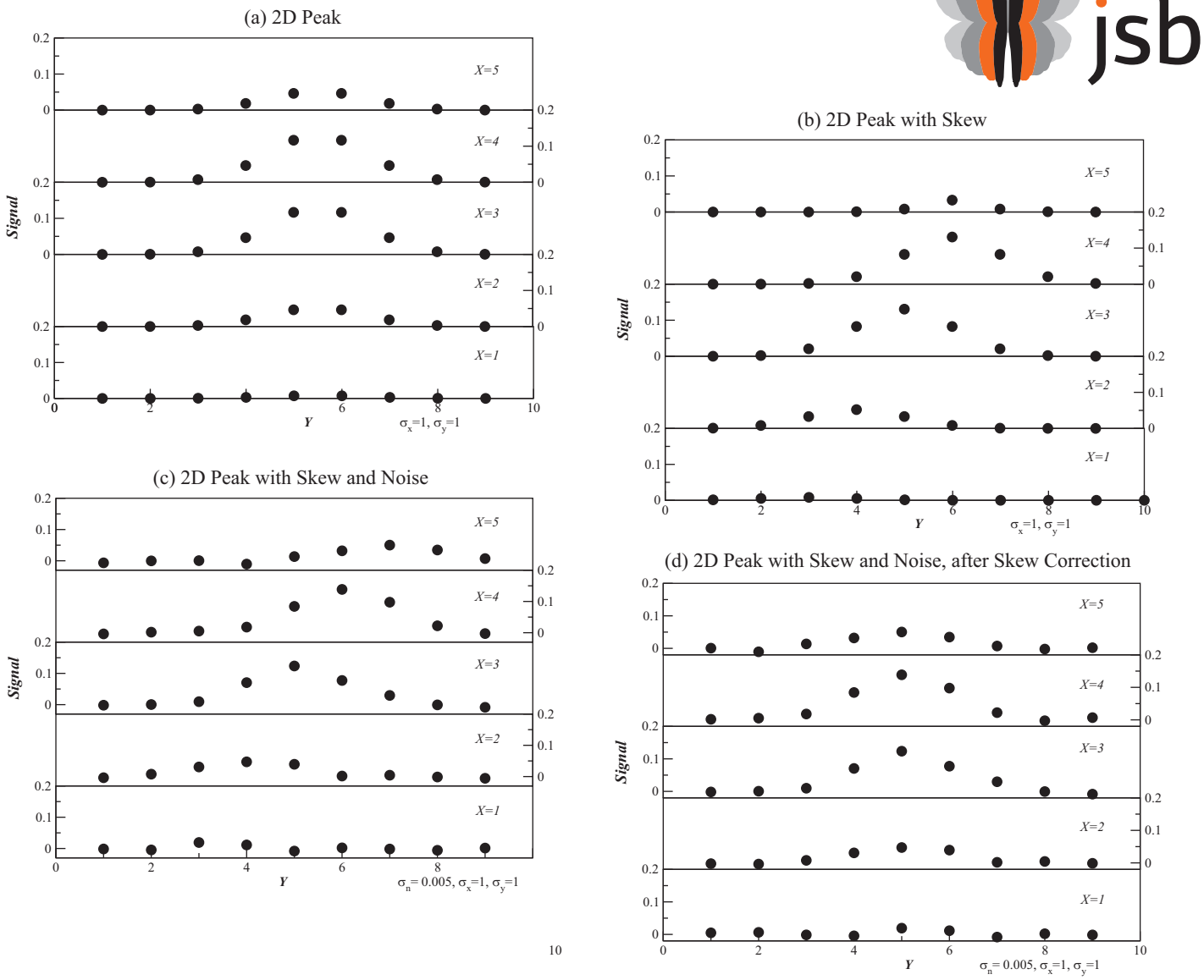

Fig. 2. Simulation of the input signal: (a) slices of a sampled 2D peak displaying each secondary 1D peak, (b) shifted slices incorporating a skew in the 2D peak, (c) slices of the skewed 2D peak with random Gaussian noise, (d) slices of the skewed 2D peak with noise after skew correction. The standard deviation of noise is σ_n, the peak-width standard deviation in the x-dimension is σ_x, and the peak-width standard deviation in the y-dimension is σ_y: (a) 2D peak, (b) 2D peak with skew, (c) 2D peak with skew and noise and (d) 2D peak with skew and noise, after skew correction.

$$N = \lceil 9\sigma_y + s(M-1) + 2 \rceil \tag{4}$$

where M is the number of columns in the array or the array size along the x-dimension and N is the number of rows in the array or the array size along the y-dimension. The size of the array along the y-dimension is a function of s, M, and σ_y, to accommodate the retention-time shifts in the secondary chromatograms. The peak is centered in the array by setting (μ_x, μ_y) to $(M/2, N/2)$. For computational convenience, the array is padded with zeros along its boundary.

Normally distributed random noise with parametric standard deviation is added to generate 2D data as shown in Fig. 2(c):

$$g[i,j] = f[i,j] + \sigma_n G_{i,j} \tag{5}$$

where $g[i, j]$ is the intensity of the noisy chromatographic data at position $[i, j]$, $f[i, j]$ is the shifted peak given by Eq. (2), σ_n is the parametric standard deviation of noise, and $G_{i,j}$ is a random number from a normal distribution. The unit for noise is the total response (*i.e.*, volume) of the two-dimensional Gaussian peak.

3.2. Peak detection

Peak detection is performed on the noisy 2D chromatographic peak given by Eq. (5). In the analysis, cross-correlation and shift correction are performed prior to peak detection by both methods. Cross-correlation identifies the shift correction required to align the 1D columns by comparing the 1D chromatograms for each relative retention-time shift within the prescribed range. The shift correction is determined by the maximum cross-correlation between 1D chromatograms. A sampled 2D peak after shift correction is shown in Fig. 2(d).

The simulation involves only one peak, so the peak-detection algorithms are configured to use two labels indicating if a data point is or is not in the analyte peak:

$$l[i,j] = \begin{cases} 1 & \text{if in peak,} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where, $l[i, j]$ is the label assigned for the data point at $[i, j]$.

3.2.1. Two-step algorithm

The two-step algorithm, as outlined by Peters et al. [7], is per-formed on the shift-corrected 2D peak, *e.g.*, as shown in Fig. 2(d).

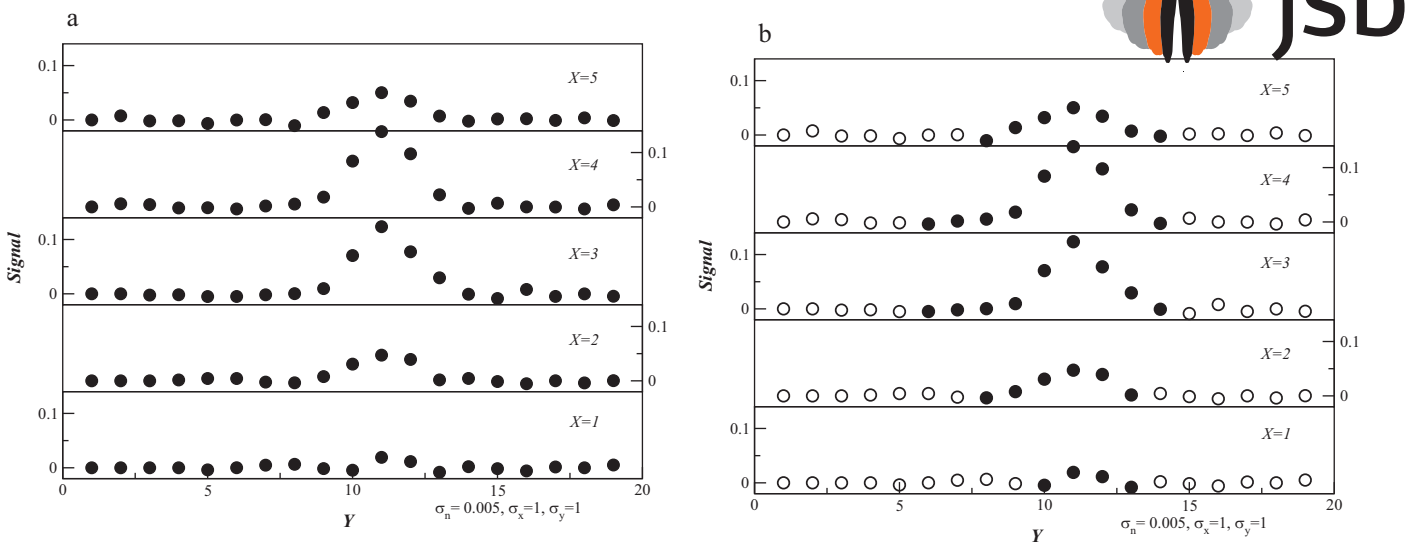

Fig. 3. Two-step peak detection applied to each 1D slice of the 2D signal followed by peak merging. The detected peak is marked by filled circles. (a) Input to two-step peak detection algorithm. (b) Peak detection by two-step algorithm.

In the first step, 1D peak detection is performed on each secondary chromatogram. With reference to [7], $Thr_0 = 0$ and $Thr_1 = 0$ so that every 1D peak is detected minima to minima. Next, the largest peak apex in the 2D chromatogram is identified. In the other secondary chromatograms (which have been shift-corrected), the overlap criterion selects the peak that overlaps (in the second-column retention time) the apex of the largest 2D peak. In any secondary chromatogram that a minimum point between two 1D peaks coincides (in the second-column retention time) with the apex of the largest 2D peak, the 1D peak with its apex closer to the largest 2D peak apex is selected and if the two 1D peaks are at the same distance from the largest peak apex, the peak with the largest apex is selected. With reference to [7], Thr_{OV}=0, so at least one data point must overlap. (These parametric settings for the two-step algorithm are the least restrictive and so maximize performance in the simulation experiments.) The unimodality criterion compares the 1D peak apex of consecutive 1D peaks to ensure the presence of a single peak maximum in the 2D peak. An example of a 2D peak detected by the two-step algorithm is shown in Fig. 3. The filled circles indicate the data points of the peak detected by the two-step method (*i.e.*, labeled '1') and the open circles are data points that are not part of the detected peak (*i.e.*, labeled '0'). The detected peak satisfies both the overlap and unimodality criteria.

3.2.2. *Watershed algorithm*

The watershed algorithm also is performed on the same shift-corrected 2D peak, *e.g.*, as shown in Fig. 2(d). The algorithm starts at the peak apex. The largest data point in the peak is identified and labeled. Then, in order by intensity, each successive point in the 2D matrix is given the label of its largest neighbor. Fig. 4 shows the peak detected by the watershed algorithm for the example 2D peak, with data points in the detected peak marked by filled circles and data points not in the detected peak marked with open circles.

3.3. *Experimental setup*

Four parameters are varied in the simulation experiments: (*a*) noise standard deviation, σ_n, from 0.0001 to 0.01; (*b*) peak-width

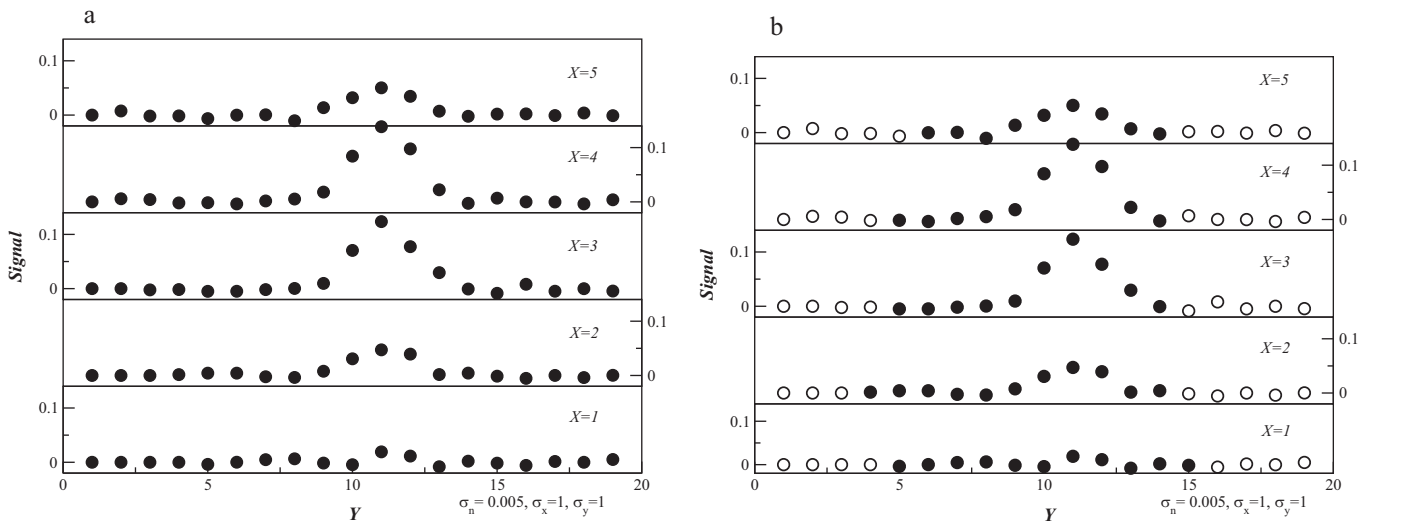

Fig. 4. The watershed algorithm applied to 2D data for peak detection. The detected peak is marked by filled circles. (a) Input to watershed algorithm. (b) Peak detection by watershed algorithm.

Table 1
Results for the watershed (WS) and two-step (2-step) peak-detection algorithms for peaks with various x-dimension peak-width standard deviation, σ_x; y-dimension peak-width standard deviation, σ_y; and noise standard deviation, σ_n. In each case, the expected mean is 1.000. For each algorithm, each test with different parameter values was repeated 1000 times with random noise.

Skew	Peak		Noise	Array size	Signal		WS				2-Step				Signif.
s	σ_x	σ_y	σ_n		Mean	S.D.	Mean	S.D.	Error	Failed	Mean	S.D.	Error	Failed	$(1-p)$
−1	1.00	1	0.0001	11×21	1.0000	0.0013	0.9997	0.0010	−0.0003	0	0.9995	0.0009	−0.0005	0	1.0000
−1	1.00	1	0.0005	11×21	0.9999	0.0061	0.9984	0.0045	−0.0015	0	0.9974	0.0040	−0.0024	0	1.0000
−1	1.00	1	0.0010	11×21	1.0006	0.0124	0.9974	0.0082	−0.0032	0	0.9949	0.0073	−0.0057	0	1.0000
−1	1.00	1	0.0050	11×21	1.0021	0.0597	0.9846	0.0368	−0.0174	0	0.9761	0.0328	−0.0260	0	1.0000
−1	1.00	1	0.0100	11×21	1.0063	0.1227	0.9649	0.0706	−0.0415	0	0.9451	0.0598	−0.0612	0	1.0000
−1	0.25	1	0.0001	5×15	1.0000	0.0006	0.9999	0.0006	−0.0001	0	0.9999	0.0005	−0.0001	0	0.0000
−1	0.50	1	0.0001	7×17	1.0000	0.0008	0.9998	0.0007	−0.0001	0	0.9997	0.0007	−0.0003	0	0.9986
−1	1.00	1	0.0001	11×21	1.0000	0.0013	0.9997	0.0010	−0.0003	0	0.9995	0.0009	−0.0005	0	1.0000
−1	2.00	1	0.0001	20×30	1.0000	0.0020	0.9994	0.0012	−0.0006	0	0.9989	0.0011	−0.0011	0	1.0000
−1	1.00	1	0.0001	11×21	1.0000	0.0013	0.9997	0.0010	−0.0003	0	0.9995	0.0009	−0.0005	0	1.0000
−1	1.00	2	0.0001	11×30	1.0000	0.0015	0.9996	0.0012	−0.0003	0	0.9987	0.0013	−0.0013	0	1.0000
−1	1.00	4	0.0001	11×48	1.0000	0.0019	0.9990	0.0015	−0.0010	0	0.9904	0.0045	−0.0096	0	1.0000
−1	1.00	8	0.0001	11×84	1.0000	0.0028	0.9951	0.0028	−0.0048	39	0.9093	0.0182	−0.0906	341	1.0000

standard deviation along the x-dimension, σ_x, from 0.25 to 2; (c) peak-width standard deviation along the y-dimension, σ_y, from 1 to 8; and (d) skew, s, from −8 to −1. (The differing ranges for the peak widths in the two dimensions reflect the typical practice that peaks are more highly sampled in the second-column separation than the first.) For each value assigned to σ_n, σ_x, σ_y, and s, the experiment is conducted $T = 1000$ times to observe the performance of the peak-detection algorithms for different input signals. The simulated 2D peak volume mean and standard deviation are calculated as:

$$SignalMean,\ \mu_s = \frac{1}{T}\sum_t\sum_{i,j} g_t[i,j] \tag{7}$$

$$SignalStandardDeviation,\ \sigma_s = \sqrt{\frac{1}{T}\sum_t\sum_{i,j} g_t[i,j]^2 - \mu_s^2} \tag{8}$$

where T is the number of test cases executed for each algorithm and $g_t[i,j]$ is the data point at $[i,j]$ in test case t.

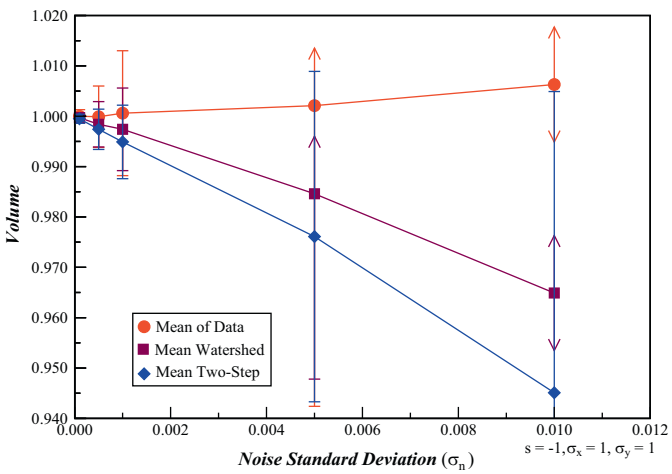

Fig. 5. Comparison of the peak volume means and standard deviations for the input signal, watershed, and two-step method as a function of the noise standard deviation, σ_n. Performance of peak detection algorithms as a function of noise.

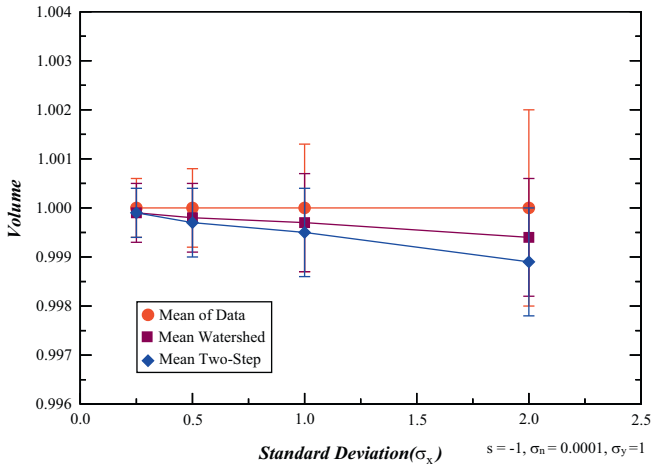

Fig. 6. Comparison of the peak volume mean and standard deviation for the input signal, watershed, and two-step method as a function of first-column peak width, σ_x. Performance of peak detection algorithms as a function of peak width, σ_x.

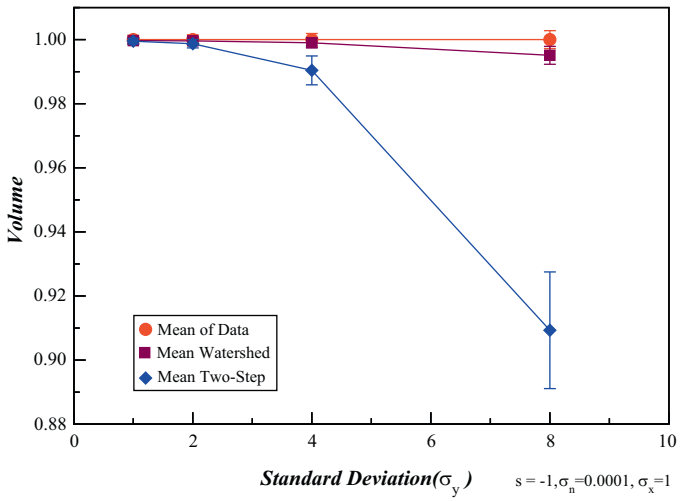

Fig. 7. Comparison of the peak volume mean and standard deviation for the input signal, watershed, and two-step method as a function of second-column peak width, σ_y. Performance of peak detection algorithms as a function of peak width, σ_y.

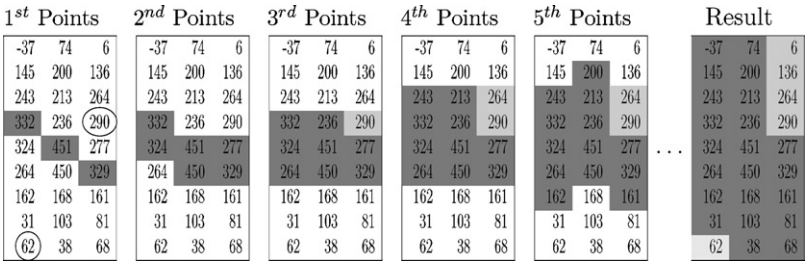

Fig. 8. In two-step peak detection, each column is a secondary chromatogram that undergoes 1D peak detection. Points included in the main peak are shown in dark gray and points not included are shown in light gray. The order that points are detected is shown in sequence from left to right.

For the two-step and watershed algorithms, the detected peak volume mean and standard deviation are computed as:

$$Mean,\ \mu = \frac{1}{T}\sum_{t}\sum_{i,j}(g_t[i,j]\, l_t[i,j]) \tag{9}$$

$$StandardDeviation,\ \sigma = \sqrt{\frac{1}{T}\sum_{t}\sum_{i,j}(g_t[i,j]\, l_t[i,j])^2 - \mu^2} \tag{10}$$

where $l_t[i,j]$ is the label assigned for the data point $g_t[i,j]$.

4. Results

Table 1 compares results of both peak-detection algorithms for various parameter values, with 1000 test cases for each set of parameter values. In all cases shown in Table 1, the skew was −1, but, with skew correction, results for other levels of skew are not materially different. (Complete experimental test results for the simulations are reported as Supplementary Data.)

Table 1 lists the volume means and standard deviations for the signal and both peak-detection algorithms with different noise levels and peak widths. Table 1 also shows the mean errors for both peak-detection algorithms, *i.e.*, the difference between the mean volume computed for the signal and the mean volume of the detected peak, and the number of failed test cases for the watershed and two-step algorithms. If the detected peak does not include all data points within one standard deviation from the peak apex in both dimensions, then the detection for the test case is counted as a failure and is not included in computing the peak volume mean and standard deviation. Table 1 displays only one such case. In that case, the two-step algorithm has a larger number of failures than the watershed algorithm. The errors for the two peak-detection methods are compared in the following paragraphs and the accompanying figures.

Table 1 also reports tests of statistical significance. The peak volumes for the two peak-detection methods in each test case were compared using independent-samples *t*-tests. As seen in Table 1, which reports one minus the *p*-values computed from the Student's *t*-distribution rounded to the nearest one-thousandth, the difference in detected peak volume by the two methods is statistically significant in most cases. In only one case, in which the first-dimension peak-width is narrow ($\sigma_x = 0.25$) and there is little noise ($\sigma_x = 0.0001$), is the difference not significant. These tests suggest that the better results of the watershed algorithm are significant and almost certainly would be observed in repeated experiments.

Table 1 and Fig. 5 show the means and standard deviations of the peak volume for the input signal, watershed, and two-step algorithm for various noise levels (and unit peak-width standard deviations). As indicated by the difference between the mean signal and the means of the peak-detection methods (shown with points in Fig. 5), the watershed algorithm performs peak detection with greater accuracy and the difference in performance increases with increasing noise. Both methods underestimate peak volume and the underestimation increases with increasing noise. The underestimation is due to over-segmentation (discussed after these results), where a single chromatographic peak is detected as multiple peaks. As indicated by the standard deviations (shown with error bars in Fig. 5), the two-step algorithm has greater precision; however, it is notable that peak volumes for both methods have smaller standard deviations than the signal. The reduced standard deviations are related to the underestimation, *e.g.*, always detecting the peak volume to be zero would have zero standard deviation.

Table 1 and Fig. 6 compare the means and standard deviations of the detected peak volume for various *x*-dimension peak widths (with constant noise and unit peak-width standard deviation σ_y). Both algorithms perform peak detection with similar accuracy and precision for narrow peaks, but as the peak width increases, the watershed algorithm has better accuracy.

Similar but more dramatic results are seen in Table 1 and Fig. 7 when the peak width along the *y*-dimension increases (with constant noise and unit peak-width standard deviation σ_x). The watershed algorithm has much better accuracy and precision than the two-step algorithm for the widest peaks. Also, for the widest peaks, the two-step algorithm had many more failed detections (341/1000) compared to the watershed algorithm (39/1000).

In the simulation, as the peak width along either or both dimensions increases, the signal-to-noise ratio (SNR), *i.e.*, the ratio of the mean to the standard deviation of the signal, decreases due to the larger number of data points, fixed noise level, and constant volume under the peak. Decreased SNR increases over-segmentation (*i.e.*,

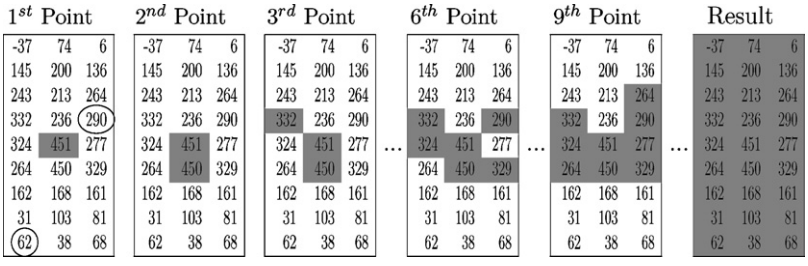

Fig. 9. The watershed algorithm labels data points (shown in dark gray) in intensity-order in the 2D chromatogram. The order that points are detected is shown in sequence from left to right.

peak splits) with both peak-detection algorithms. In the two-step algorithm, each point is compared only to its two neighbors in the 1D chromatogram for the presence in a peak. If noise changes either neighbor sufficiently, the peak is over-segmented. In the experiments, the watershed algorithm performs better than the two-step approach because it considers the 2D neighborhood of each data point to identify the data points in a 2D peak, thereby, reducing the effect of noise.

Figs. 8 and 9 illustrate peak detection for the two-step and the watershed algorithms, respectively, with a simulated peak after skew correction. The effect of noise on the signal can be observed at the circled data points, which have slightly larger values than their neighbors in the secondary chromatogram.

Fig. 8 illustrates the two-step peak-detection process in three 1D chromatograms. Initially, the largest data point in each secondary chromatogram is labeled. This is followed by labeling adjacent points in the 1D chromatograms and then merging the 1D peaks as described in Section 2.1. The circled data points are larger than their neighbors in the secondary chromatogram, hence they are not included in the 1D peaks. So, the chromatogram is split into multiple peaks even though there is only one peak in the 2D chromatogram.

Fig. 9 illustrates the watershed detection process on the 2D matrix. In each step, data points are evaluated in intensity-order to determine the peak label. Initially, the largest point is labeled. Then, the next largest data point in the 2D matrix is labeled and the process continues until all points in the matrix are labeled. For the circled data points, there is a neighbor in the peak with a larger value. The circled points adopt the label of their largest neighbor and thus all data points are labeled correctly as a single peak.

The simulation experiments demonstrate that with shift correction applied for both peak-detection algorithms, the watershed algorithm achieves more accurate peak detection than the two-step approach for varying noise levels, peak widths, and shifts. And, as both noise and peak widths increase, the two-step algorithm has more failed detections than the watershed algorithm. Varying the second-column retention-time shift did not materially impact these results because shift correction is performed for both algorithms. Detailed simulation results, including experimental results for larger retention-time shifts, are provided as Supplementary Data.

5. Conclusion

A study by Vivó-Truyols and Janssen [4] discussed the probability of failure of the watershed algorithm for GC×GC data with varying second-dimension retention-time shifts. Their analysis compared the two-step and watershed peak-detection algorithms without accounting for the retention-time shifts in the watershed algorithm, whereas the shift was accounted for in the two-step algorithm. This caused the watershed algorithm to have a larger probability of failure than the two-step approach. This paper analyzes the performance of the two algorithms when correction for retention-time shifts is performed for both algorithms.

A series of simulation experiments evaluated both peak-detection techniques for varying levels of noise, peak widths, and retention-time shifts. The watershed algorithm performed better than the two-step approach when skew correction is applied for both the methods. Neither the two-step algorithm nor the watershed algorithm ensures successful peak detection under all conditions. As the peak width and noise increase, both techniques detect peaks less accurately, even with shift correction.

The performance of both the two-step and watershed algorithms could be improved by noise suppression, *e.g.*, smoothing before peak detection. Various noise suppression techniques can be used with both peak-detection techniques. Noise suppression was not used in these experiments because the method(s) for doing so would be a confounding variable in comparing the performance of the peak detection methods. And, although noise suppression can attenuate the effect of noise, the progressive effect of increasing noise on the performance of peak-detection algorithms cannot be eliminated. Similarly, the issue of coelutions, although important for peak detection, was not considered herein, as neither the two-step nor watershed algorithm incorporates a solution for coelution. Other techniques have been developed for unmixing coeluted peaks, *e.g.*, [11–21].

Acknowledgements

This research was funded in part by the Nebraska Center for Energy Sciences Research at the University of Nebraska – Lincoln and the U.S. National Science Foundation under Award Number IIP-1013180.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2011.07.052.

References

[1] J. Beens, H. Boelens, R. Tijssen, J. Blomberg, J. High Resolut. Chromatogr. 21 (1998) 47.
[2] Q. Song, A. Savant, S. Reichenbach, in: Applications of Digital Image Processing, SPIE, 1999, p. 2.
[3] S. Reichenbach, M. Ni, V.V.A. Kottapalli, Chemom. Intell. Lab. Syst. 71 (2004) 107.
[4] G. Vivó-Truyols, H.-G. Janssen, J. Chromatogr. A 1217 (2010) 1375.
[5] L. Ramos, J. Sanz, in: L. Ramos (Ed.), Comprehensive Two-dimensional Gas Chromatography, Elsevier, 2009, p. 283.
[6] T. Skov, J.C. Hoggard, R. Bro, R.E. Synovec, J. Chromatogr. A 1216 (2009) 4020.
[7] S. Peters, G. Vivó-Truyols, P. Marriott, P. Schoenmakers, J. Chromatogr. A 1156 (2007) 14.
[8] E.J.C. van der Klift, G. Vivó-Truyols, F.W. Claassen, F.L. van Holthoon, T.A. van Beek, J. Chromatogr. A 1178 (2008) 43.
[9] S. Beucher, Scanning Microsc. Int. 6 (1992) 299.
[10] S. Reichenbach, in: L. Ramos (Ed.), Comprehensive Two-dimensional Gas Chromatography, Elsevier, 2009, p. 77.
[11] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Anal. Chem. 70 (1998) 2796.
[12] B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski, J. Microcolumn Sep. 11 (1999) 97.
[13] C.G. Fraga, B.J. Prazen, R.E. Synovec, J. High Resolut. Chromatogr. 23 (2000) 215.
[14] C.G. Fraga, C.A. Bruckner, R.E. Synovec, Anal. Chem. 73 (2001) 675.
[15] B.J. Prazen, K.J. Johnson, A. Weber, R.E. Synovec, Anal. Chem. 73 (2001) 5677.
[16] A.E. Sinha, C.G. Fraga, B.J. Prazen, R.E. Synovec, J. Chromatogr. A 1027 (2004) 269.
[17] A.E. Sinha, J.L. Hope, B.J. Prazen, C.G. Fraga, E.J. Nilsson, R.E. Synovec, J. Chromatogr. A 1056 (2004) 145.
[18] C.G. Fraga, C.A. Corley, J. Chromatogr. A 1096 (2005) 40.
[19] H. Kong, F. Ye, X. Lu, L. Guo, J. Tian, G. Xu, J. Chromatogr. A 1086 (2005) 160.
[20] J.C. Hoggard, R.E. Synovec, Anal. Chem. 79 (2007) 1611.
[21] Z.-D. Zeng, S.-T. Chin, H.M. Hugel, P.J. Marriott, J. Chromatogr. A 1218 (2011) 2301.