#S08

Quantification in Comprehensive Two-Dimensional Liquid Chromatography

Stephen E. Reichenbach†

Computer Science and Engineering Department, University of Nebraska—Lincoln, Lincoln, Nebraska 68588-0115

This correspondence corrects the description in a recent paper by Mondello et al., "Quantification in Comprehensive Two-Dimensional Liquid Chromatography" [Mondello, L.; Herrero, M.; Kumm, T.; Dugo, P.; Cortes, H.; Dugo, G. *Anal. Chem.*, 2008, *80*, 5418−5424], of previous research on peak integration. This correspondence also shows that the peak integration method proposed in that paper is equivalent to, but is less efficient than, simply summing the data values.

A recent paper, "Quantification in Comprehensive Two-Dimensional Liquid Chromatography",[1] proposed a method for integrating peaks. In discussing previous work on two-dimensional peak integration, the authors, Mondello et al., stated that the two-dimensional peak integration method described by Reichenbach et al.[2,3] utilizes "interpolated 2D plots" and concluded that "such integration methodologies may not produce correct results."[4] Actually, Reichenbach et al. integrated peaks by summing intensity values without interpolation, which is most efficient. This correspondence corrects that description and examines the effect of interpolation on peak integration.

PREVIOUS WORK BY REICHENBACH ET AL.

Mondello et al.[1] wrote "[Our proposed] procedure is based purely on chromatographic properties and differs from previously reported applications, where 2D peaks are integrated as images or considering peak volumes.[2,3,5] However, such integration methodologies may not produce correct results, because the programs employed represent bidimensional chromatograms as interpolated 2D plots, with the aim of showing three-dimensional images."

This description of the cited work is incorrect. In the cited research and in the software developed from it,[6] peak integration is performed by summing the values of the data points in the peak. In the cited paper, Reichenbach et al. describe the computation for integration as the "sum"[7] and the software documentation describes the computation as "the total of [the peak's] values or total response".[8] In both the research and the software, peak integration is performed as a summation of the data values, without interpolation.

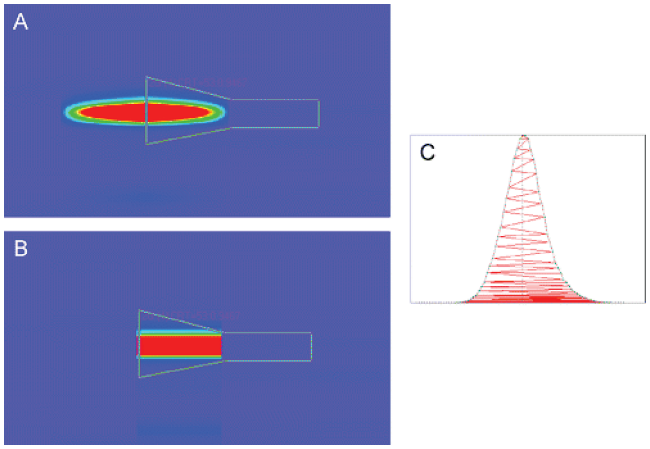

Figure 1. This figure, from Mondello et al., appeared with the caption: "Integration area of a given peak when it is shown with (A) or without graphical interpolation (B), and example of the integration process for a second-dimension peak (C)."[13] Actually, both images A and B are interpolated.

This misunderstanding may stem from the digital-image terminology used by Reichenbach et al. and from confusion of the operation of storing two-dimensional chromatographic data as a two-dimensional image with the operations that are used to view images. Two-dimensional chromatograms typically are ac-quired sequentially, in a one-dimensional array. When two-dimensional chromatographic data is stored as a two-dimensional image, the value(s) generated by the detector at each data point (e.g., an intensity value produced by a flame ionization detector or an array of intensities produced by a mass spectrometer) is/are stored successively as pixels in a two-dimensional array, with a column of pixels for each modulation cycle. In this operation, there is a one-to-one identity between the data points of the

† Corresponding author. E-mail: reich@cse.unl.edu.

(1) Mondello, L.; Herrero, M.; Kumm, T.; Dugo, P.; Cortes, H.; Dugo, G. *Anal. Chem.* **2008**, *80*, 5418–5424.
(2) Reichenbach, S. E.; Kottapalli, V.; Ni, M.; Visvanathan, A. *J. Chromatogr., A* **2004**, *1071*, 263–269.
(3) Reichenbach, S. E.; Ni, M.; Kottapalli, V.; Visvanathan, A. *Chemom. Intell. Lab. Syst.* **2004**, *71*, 107–120.
(4) Reference 1, p 5420.
(5) Amador-Muñoz, O.; Marriott, P. J. *J. Chromatogr., A* **2008**, *1184*, 323–340.
(6) GC Image, LLC, GC Image software, http://www.gcimage.com, 2008.
(7) Reference 3, p 115.
(8) GC Image, LLC, Peak Detection and Analysis, in GC Image Users' Guide,

10.1021/ac900047z CCC: $40.75 © 2009 American Chemical Society
Published on Web 05/15/2009

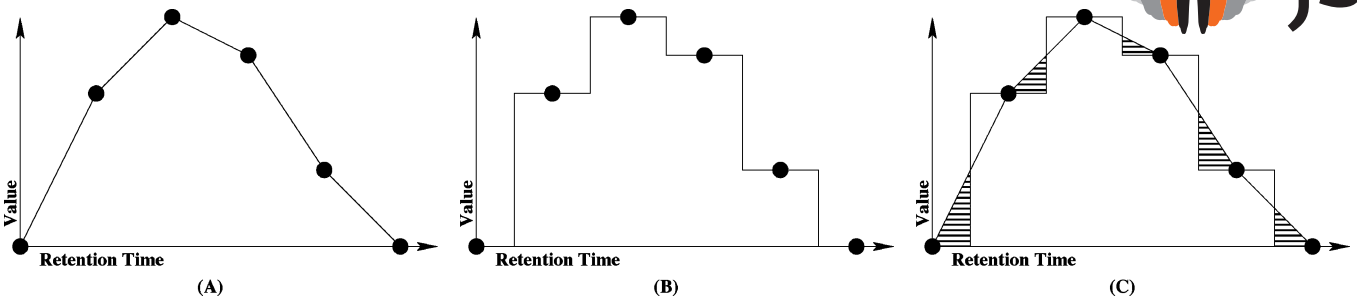

Figure 2. Interpolation of the same data points (●) forming a one-dimensional chromatographic peak with (A) linear interpolation and (B) nearest-neighbor interpolation. In part C, the overlay of the two interpolated profiles illustrates that the integrated areas of the interpolated profiles are equal.

detector output and the pixels of the image, i.e., each single pixel has the value(s) produced by the chromatograph at a single data point. After the two-dimensional chromatographic data is stored as a two-dimensional image, viewing the image may involve operations such as panning, scrolling, and zooming. These operations may require interpolation and resampling to produce a shifted or scaled visualization of the image, but the underlying digital image with the chromatographic data-point values is not changed.

INTERPOLATION

In their Figure 1, reproduced here, Mondello et al.[1] purported to show a two-dimensional peak (A) with interpolation and (B) without interpolation. However, both images clearly were created by interpolating discrete data points for visualization, because data points alone are not visible as an image. An image is visible only when the two-dimensional space between the points is filled, e.g., by interpolation.[9] In Figure 1A, the data points evidently were interpolated using a higher-order interpolator compared to the image in Figure 1B, perhaps a bilinear interpolator. In Figure 1B, it is evident that the data points were interpolated using nearest-neighbor interpolation to visualize the two-dimensional image, effectively creating a uniformly colored rectangular tile for each data value.

Although the image visualizations in Figure 1 appear different because of the different methods of interpolation, it is important to note that the data values for both are the same. That the data values are the same is more clearly seen in one dimension. Figure 2 illustrates a one-dimensional chromatographic peak with data points (shown with ●) interpolated by two popular interpolation methods: (1) Linear interpolation, in Figure 2A, fills between data points with connecting lines. (2) Nearest-neighbor interpolation, in Figure 2B, fills between data points with a constant value determined by the nearest data point.

The resulting profiles are visually different, but the data values (shown with ●) which are used for quantitative processing are the same. Two-dimensional interpolation methods (e.g., bilinear or nearest-neighbor interpolation in parts A and B of Figure 1) work similarly.

PEAK INTEGRATION

In discussing Figure 1, Mondello et al.[1] claimed that errors could result from integrating the interpolated peak in Figure 1A, compared with what they called the "noninterpolated"[10] peak in Figure 1B. However, as shown here, if the interpolation function has a weighting function that integrates to 1, the result of peak integration is the same. In this section, the equality for peak integration with interpolation is first illustrated with a one-dimensional example and then proven for two dimensions.

Figure 2C overlays the peak profiles for linear interpolation from Figure 2A and nearest-neighbor interpolation from Figure 2B. For each region under the profile for linear interpolation that is outside the profile for nearest-neighbor interpolation (the regions filled with horizontal lines), there is an equally sized region above the profile for linear interpolation that is inside the profile for nearest-neighbor interpolation. As can be seen, the total integrated areas under the two different profiles are equal.

This equality can be proven for any two-dimensional interpola-tion method that has a weighting function $h(x,y)$ that integrates to 1:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\,dx\,dy = 1 \qquad (1)$$

Both nearest-neighbor interpolation and bilinear interpolation have weighting functions satisfying eq 1. The interpolation of discrete data points (e.g., those in a peak) $p[m,n]$, $m = 1 \ldots M$, $n = 1 \ldots N$, is the result of convolution of the data points with the interpolator weighting function:[9]

$$p(x,y) = \sum_{m=1}^{M}\sum_{n=1}^{N} p[m,n]h(x-m,y-n) \qquad (2)$$

The integration of the interpolated peak can be simplified as

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\,dx\,dy &= \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{m=1}^{M}\sum_{n=1}^{N} p[m,n]h(x-m,y-n)\,dx\,dy &= \\ \sum_{m=1}^{M}\sum_{n=1}^{N} p[m,n]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-m,y-n)\,dx\,dy &= \\ \sum_{m=1}^{M}\sum_{n=1}^{N} p[m,n], &\end{aligned} \qquad (3)$$

which is just the sum of the data-point values. This proof also holds for one dimension.

(9) Burger, W.; Burge, M. J. *Digital Image Processing: An Algorithmic Introduc-*

(10) Reference 1, p 5420.

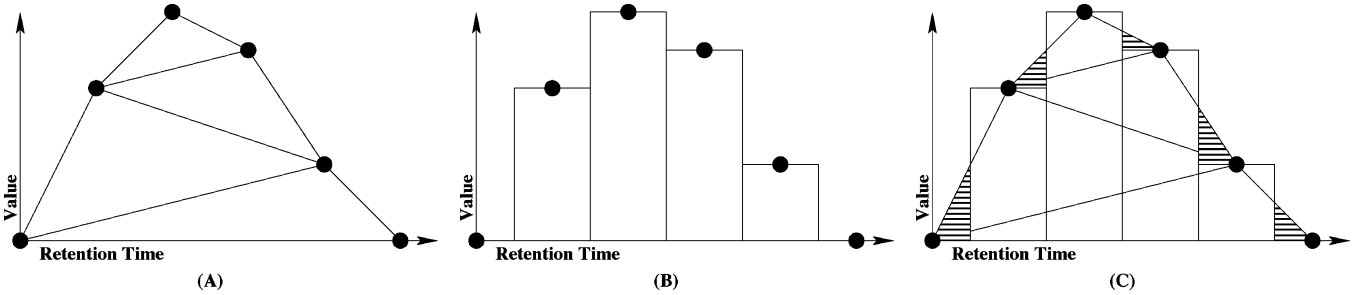

Figure 3. Integration areas of a one-dimensional peak using (A) the triangulation method proposed by Mondello et al.[1] and (B) the method of summing data values for unit-width rectangles. Data points are indicated by ●. As illustrated in part C, both methods in parts A and B yield the same total integration area, but the method of summing in part B is computationally more efficient.

So, any interpolator with a weighting function that integrates to 1 (including nearest-neighbor interpolation and bilinear interpolation) yields the same peak integration value, which can be computed by simply summing the data values. It is possible that interpolation with a weighting function that has a radius larger than one sample interval could introduce overlap between nearby peaks. Both nearest-neighbor and linear interpolation have a radius of less than one sample, so one data point on the baseline between peaks is necessary and sufficient for proper integration without unmixing/deconvolution.

EFFICIENCY

Mondello et al. described a method for peak integration which they characterized as a "novel approach".[1] Figure 3A illustrates their method of peak integration by triangularization. In their approach, each peak is sequentially triangulated from the data points (shown with ●) on alternating sides of the apex down to the baseline. As can be seen, the profile created by their approach is the profile from linear interpolation (Figure 2A).

Mondello et al. compute the peak integration as follows: "The triangle areas are then determined and then summed."[11] This is an inefficient approach that yields a result identical to simply summing the data values. Equation 3 proves this equality, and Figure 3B shows graphically that the integration can be computed by simply summing the data values. In Figure 3B, each data point determines the height of a rectangle with unit width, taking the intersample time as the horizontal unit of measure. (This profile is the same as for nearest-neighbor interpolation, in Figure 2B.) Then, the total area of the rectangles can be computed by summing the data-point values. Figure 3C illustrates that the total area of the rectangles and total area of the triangles are equal. For each region under the triangulated peak profile that is outside the rectangles (the regions filled with horizontal lines), there is an equally sized region above the triangulated peak profile that is inside the rectangles. So, the triangularization approach proposed by Mondello et al. yields the same result as simply summing the values; it is just computationally less efficient than summing the values as described by Reichenbach et al.[2,3]

SOFTWARE FOR TWO-DIMENSIONAL CHROMATOGRAPHY

Mondello et al. wrote of "the lack of dedicated software capable of identifying and quantifying more precisely a peak taken from a two-dimensional (2D) plot."[12] Software for comprehensive two-dimensional gas and liquid chromatography (GC × GC and LC × LC) from GC Image[6] (Lincoln, NE) is capable of identifying and precisely quantifying two-dimensional peaks, using the methods described by Reichenbach.[3] For example, the calibration correlation (R^2) values for the sample data provided with the GC Image software ranges from 0.999 84 to 0.999 99.[6] It is noted here that Professor Reichenbach, the author of this correspondence, is the founder of GC Image, LLC, a spinoff from the University of Nebraska, and that Professor Mondello is a founder of a commercial spinoff from the University of Messina.

SUMMARY

This correspondence recognizes that Mondello et al. have contributed to the developing literature on LC × LC technologies but corrects their description of previous work on peak integration and proves that interpolation (as it is normally implemented) does not change the result of peak integration. Peak integration is most efficiently implemented by simply summing data-point values, as described by Reichenbach et al.[2,3]

Received for review January 8, 2009. Accepted May 1, 2009.

AC900047Z

(11) Reference 1, p 5421.
(12) Reference 1, p 5419.
(13) Reference 1, p 5420.