

Chromophoreasy, an Excel-Based Program for Detection and Integration of Peaks from Chromatographic and Electromigration Techniques

Fernando A. S. Vaz, Leandra N. O. Neves, Rafael Marques, Renata T. Sato and Marcone A. L. Oliveira*

*Departamento de Química, Universidade Federal de Juiz de Fora,
36036-330 Juiz de Fora-MG, Brazil*

This paper describes the development, evaluation, features and applications of Chromophoreasy, an alternative Excel-based program for recognition and integration of chromatographic and electrophoretic peaks. The proposed recognition is made according to parameters adjustable by the analyst, such as time range, noise smoothing window size and slope/curvature sensitivity. During integration, retention/migration time, area, height, half-height width, plate numbers, asymmetry factor, US Pharmacopeia tailing factor, resolution and statistical moments are determined. A chromatogram/electropherogram is plotted along with the found baselines. The effect of peak shape (heights and symmetries) and baseline slope over accuracy was evaluated and the precision of recognition/integration was investigated under several simulated conditions, with varied signal-to-noise levels, smoothing modes and smoothing window sizes. Data from liquid and gas chromatography, capillary electrophoresis and electrochromatography techniques with refractive index, flame ionization, capacitively coupled contactless conductivity (lab-made) and ultraviolet absorbance detections, respectively, were treated, illustrating the broad applicability of the proposed program for standard and sample analysis. Statistically similar results were obtained, when compared with other commercial software, showing it to be a simple, practical and reliable tool for general use in the separation area.

Keywords: peak recognition, peak integration, chromatography, capillary electrophoresis, Excel macros

Introduction

Separation science clearly occupies a prominent position in analytical chemistry. Several advantages such as sensitivity of detection modes, selectivity and efficiency of separation columns and short analysis times, have led the chromatographic and electromigration techniques to this high level.¹⁻³ However, the dispersion of the analyte molecules during their continuous and differential motion along the separation system is one of the main unavoidable separation characteristics, so that the analyte registration should be as representative as possible. Since the presence of an analyte is observed through appearance of a chromatographic/electrophoretic peak, whose height and area are sensitive to concentration,⁴ typical responses such as efficiency, resolution, signal-to-noise ratio, analysis time, and symmetry, among others, are completely dependent

on accurate measurements made on that peak, during optimization^{5,6} and validation^{7,8} of a method.

Since the allocation (even if automated) of peak boundaries as well as the subsequent peak integration are not necessarily accurate and precise due to the presence of noise, for instance,⁹ the errors associated with these procedures may propagate up to the final result of an analysis. This issue becomes more critical with increasing demand for faster analysis and narrower peaks, which mobilizes the developers of algorithms for chromatographic data treatment. Therefore, various peak recognition methods¹⁰⁻¹⁵ and mathematical models for deconvolution of overlapping peaks and integration in noisy and complex systems^{16,17} have been developed and evaluated. However, a deficiency in the availability of programs dedicated to chromatographic or electrophoretic data processing, which are simple, practical and accessible to researchers, students and specialized laboratories, is still noticed. Furthermore, from our knowledge, a program that allows one to find the

*e-mail: nando.simas.vaz@gmail.com

details of calculations used during the peak recognition and integration does not exist. Thus, the way programs perform these procedures may be not possible to understand or control.

In this context, this paper describes the development, evaluation, features and some applications of Chromophoreasy, an alternative program proposed for recognition and integration of chromatographic and electrophoretic peaks, developed in Visual Basic for Applications (VBA), which operates in the familiar Microsoft Excel (2010) environment. In fact, Excel has been successfully employed in relevant works, such as interface for analysis of liquid chromatography (LC)-mass spectrometry (MS) metabolomics data,¹⁸ retention prediction and separation optimization in LC,¹⁹ and simulation of chromatographic runs under column and detector viewpoints.²⁰ Chromophoreasy was tested in several simulated and experimental chromatograms, electrochromatograms and electropherograms. Some of these applications are shown here and some results were compared with those obtained from Agilent ChemStation and Shimadzu GC Solution softwares.

Experimental

Since several experimental analyses through different separation techniques were carried out, the description of chemicals, reagents and instrumental conditions is large. Thus, this section only describes basic elements that allow one to understand the purpose of the present work. Further details are available in Supplementary Information under Experimental section.

Samples

For demonstration of program use on real samples, a commercial milk sample was treated according to literature reports²¹ and biodiesel samples were obtained from transesterification reactions.²²

Instrumental

The LC system used was a Breeze Modular high performance liquid chromatograph (HPLC) (Waters, Milford, USA) equipped with refractive index detector for modular systems and controlled by Waters Breeze High Performance LC software. The method conditions were based on previous work.²¹ The experiments involving gas chromatography (GC) analysis were performed in a GC 2010-Plus gas chromatograph equipment (Shimadzu, Kyoto, Japan), equipped with a flame ionization detector

(FID) and controlled by Shimadzu GC Solution software (V. 2.32.00). The method conditions were based on literature.²³ For capillary zone electrophoresis (CZE) analysis of organic acids standards, an Agilent 1600 capillary electrophoresis (CE) system (HP^{3d} CE, Palo Alto, USA) equipped with a diode array detector (DAD) and controlled by HP ChemStation software (rev A.06.01) was used. The running conditions were based on a recent work.²⁴ The experiments involving CZE analysis of lactose and lactulose standards based on literature reports²⁵ were performed in an Agilent 7100 CE system controlled by Agilent ChemStation software (rev. B.04.03) and equipped with a DAD and a lab-made capacitively coupled contactless conductivity detector (C⁴D).²⁶ For analysis of polycyclic aromatic hydrocarbons (PAH) standards by capillary electrochromatography (CEC) based on a previous work,²⁷ the Agilent 7100 CE system was also used.

Chromatogram simulation

Simulated chromatograms were generated through a peak model based on exponentially modified Gaussian (EMG) function expressed as:²⁸

$$f(t) = \begin{cases} 0, & \text{if } t < t_R - \frac{s(4-a^2)}{2a} \\ h \times \exp\left\{\left(\frac{4}{a^2} - 1\right) \times \left[\ln\left(1 + \frac{2a(t-t_R)}{s(4-a^2)}\right) - \frac{2a(t-t_R)}{s(4-a^2)}\right]\right\} & \end{cases} \quad (1)$$

where t_R is the retention time, h is the peak height, a is an asymmetry term ($-2 < a < 2$, $a \neq 0$) and s is the standard deviation (SD). This simple EMG function allows the generation of peaks with apex coordinates exactly known, which is useful for the evaluation of the integrator performance in the determination of “experimental” t_R and h . Fronting and tailing peaks can be obtained through negative and positive a values, respectively. Symmetrical Gaussian curves are obtained if a is approximately equal to zero. For generation of chromatograms with multiple peaks, each peak was generated independently and then, summed to each other ($f_1(t) + f_2(t) \dots$). The noise with Gaussian distribution was calculated with a function that returns the inverse of the normal cumulative distribution with random probability, average equal to zero and SD equal to one. The noises were multiplied by a scaling factor and added to the chromatograms.

Theory and method for peak detection and integration

Several VBA macros, required for peak recognition, peaks extraction to distinct spreadsheets, peak integration,

fine adjustments of the peak boundaries, splitting coeluted peaks for distinct spreadsheets, grouping the integration results, plotting the chromatogram with baselines, and operation of a command box, were developed. All available functions were inserted into the command box, which can be opened in any (Excel) chromatogram file via keyboard shortcut. However, this section only describes some relevant characteristics that may affect the results discussed in this work. Readers interested in using and viewing other functions and aspects of the program are encouraged to see more details provided in Supplementary Information under Additional tools sub-section.

Smoothing function

To evaluate the efficiency of peak detection, when signal-to-noise ratio (S/N) is not sufficiently elevated, two types of smoothing by convolution were employed and tested. One of them is a moving average smoothing (MA) with possibility to set the amount of points to be averaged (window). Thus, a chromatogram point (S_i) is replaced by an average ($\bar{S}_{i(MA)}$), according to equation given by:

$$\bar{S}_{i(MA)} = \frac{S_{i-n} + \dots + S_{i-1} + S_i + S_{i+1} + \dots + S_{i+n}}{2n+1} \quad (2)$$

where n , set by the analyst, is an integer greater than zero (if $n = 0$, there is no smoothing) and less than a third of the total number of chromatogram points. The second, Savitzky-Golay (SG) smoothing method,²⁹ is equivalent to a polynomial fitting, but the original points are just multiplied by specific integer numbers. The following equation was employed:

$$S_{i(PM)} = \frac{-21(S_{i-4} + S_{i+4}) + 14(S_{i-3} + S_{i+3}) + 39(S_{i-2} + S_{i+2}) + 54(S_{i-1} + S_{i+1}) + 59S_i}{231} \quad (3)$$

The result of this equation is exactly the same if the time value (t_i), in the middle of a set of nine points, is substituted in a third-order polynomial model fitted to these points by least squares. The advantage of equation 3 over fitting models is the computational simplicity that leads to a much shorter processing time. The use of SG smoothing requires that the chromatogram have a constant sampling rate. Depending on the selected smooth mode and window size (for the MA smoothing), among other factors discussed below, the behavior of peak recognition may vary severely, which actually allows this program to be employed to several chromatographic/electromigration techniques.

Thresholds calculation

A suitable way to distinguish the peaks from the (noisy) baseline, i.e., to recognize peaks, is through first

(slope) and second (curvature) derivative analyses on every chromatogram segment (or a time range only, set by user). The behavior of a derivative curve is more predictable than the chromatogram signal itself, making the peak recognition, in general, more reproducible. Moreover, the simultaneous analysis of first derivatives (FD) and SD prevents that peak apexes, shoulders and valleys between two overlapped peaks are wrongly interpreted as a peak boundary.

Once the smoothing parameters are defined and chromatogram is smoothed (or not), FD are calculated and a median for these derivatives (M_{FD}) is obtained. If the chromatogram has a representative baseline, beside the noise and some peaks, M_{FD} may represent a robust value which is probably the “main” baseline slope. For drifting baselines, for instance, M_{FD} is not necessarily zero, which makes this step important to be executed. Deviations between each FD_i and M_{FD} are calculated and a new median (M_D) is obtained from this set of results. If chromatogram has no peak and noise, M_D should be zero. Otherwise, the presence of peaks and noise raises M_D . Finally, M_{FD} and M_D are associated to provide the threshold range (T_{FD}) as:

$$T_{FD} = M_{FD} \pm 5M_D / \text{Sens} \quad (4)$$

where Sens is a sensitivity factor set (and eventually optimized) by the analyst. As Sens increases, the range of superior and inferior thresholds (T_S-T_I) decreases and, thus, the peak recognition mechanism gets more sensitive. The empirical number 5 in this equation gives suitable threshold ranges along with Sens. The same consideration viewed in this section is made for calculation of superior and inferior thresholds for second derivatives, i.e., $T_{S(SD)}$ and $T_{I(SD)}$, respectively.

If signal (S) is processed with MA smoothing or it is not smoothed, FD_i and SD_i are calculated as:

$$FD_i = (S_{i+1} - S_{i-1}) / (t_{i+1} - t_{i-1}) \quad (5)$$

$$SD_i = (FD_{i+1} - FD_{i-1}) / (t_{i+1} - t_{i-1}) \quad (6)$$

where t is time. In the case of using SG smoothing method, the derivatives are based on the fitted third-order polynomial, calculated directly from original signal as:²⁹

$$FD_i = \frac{86S_{i-4} - 142S_{i-3} - 193S_{i-2} - 126S_{i-1} + 129S_{i+1} + 193S_{i+2} + 142S_{i+3} - 86S_{i+4}}{1188(t_{i+1} - t_i)} \quad (7)$$

$$SD_i = \frac{28(S_{i-4} + S_{i+4}) + 7(S_{i-3} + S_{i+3}) - 8(S_{i-2} + S_{i+2}) - 17(S_{i-1} + S_{i+1}) - 20S_i}{462(t_{i+1} - t_i)^2} \quad (8)$$

where the terms “ $t_{i+1} - t_i$ ” in the denominators were inserted to make these derivatives dimensionally equivalent to equations 5 and 6.

Peak searching

Once $T_{S(FD)}$, $T_{I(FD)}$, $T_{S(SD)}$ and $T_{I(SD)}$ are obtained, the peak scanning procedure can be done in a matrix containing first and second derivatives smoothed by MA with a three-point window. While derivative values are less than T_S and greater than T_I , no peak starting is found. When both FD_i and SD_i are greater than respective T_S , a positive peak starting is defined with the time coordinate of the point $i - 2$ (“-2” compensates a shift made by the MA window position, ensuring that the peak baseline will touch the correct chromatogram point). When both FD_i and SD_i are less than T_I , a negative peak start is defined. The algorithm for peak detection is summarized in Figure 1.

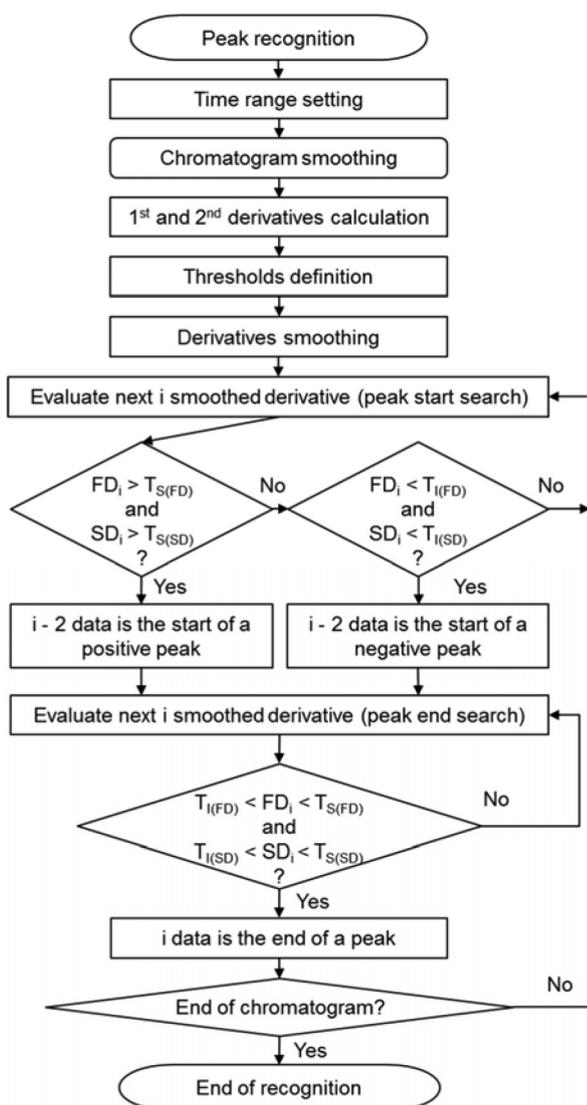


Figure 1. Flowchart of the peak recognition algorithm.

Scanning continues from next i point, now looking for a peak end, i.e., when both FD_i and SD_i are inside threshold ranges. In that condition, the time coordinate

of i point is defined as a peak end. It is important to stress that “defining” a peak start/end is not the same as “registering” it for integration purposes, which depends on the option of peak type from the command box, set previously by the analyst. For instance, if the user selected only positive peaks to be detected/integrated, the program will “define” and process both kinds of peaks, but will only “register” positive ones. This process prevents that the apex of a negative peak is interpreted as the start of a positive peak and *vice versa*.

Integration

The first step of the integration process consists in the construction of a straight baseline fitted to the points of the peak limits (Figure 2a). Parameters, such as retention/migration time (t_R), height (h), area (A), half-height width ($w_{0.5}$), plate numbers (N), asymmetry (As) and US Pharmacopeia tailing factor (Tf) are calculated from the adjusted peak, obtained through the subtraction of baseline from original signal (Figure 2b). t_R and h are, respectively, the x - and y -coordinates of the parabola maximum (Figure 2c), obtained through the first derivative of its equation, which is determined with the highest three points of the peak apex.³⁰ The same reasoning is reflected on negative peaks, considering the lowest point of the parabola. The $w_{0.5}$ value is the horizontal distance between two x -coordinates of the adjusted peak, whose y -coordinates are $h/2$. Because the coordinates $(x, h/2)$ probably do not exist in the (discrete) peak data, an interpolation between the two points nearest to $(x, h/2)$ is made at both sides of the peak (solid red lines in the peak of Figure 2b). In the case of partially coeluted peaks $w_{0.5}$ is estimated as following: since $w_{0.5}$ corresponds to 2.355σ (where σ is the standard deviation of a Gaussian curve) and A/h corresponds to 2.507σ , i.e., the width at 45.6% of height,³⁰ $w_{0.5} = (2.355\sigma) / (2.507\sigma) \times A/h = 0.93937A/h$. The plate number is calculated by $N = 5.54t_R^2 / w_{0.5}^2$, useful for symmetrical peaks. The asymmetry is calculated by $As = b_{0.1} / a_{0.1}$ and tailing factor by $Tf = (a_{0.05} + b_{0.05}) / 2a_{0.05}$, where a is the front half-width and b is the back half-width of the peak measured at $0.1h$ and $0.05h$ from the leading or trailing edge of the peak to the t_R (Figure 2d).

Next, A is calculated as the sum of increments by the trapezoidal rule:

$$A = \sum_{i=1}^n \left[(t_{i+1} - t) \times \left(\frac{S_i + S_{i+1}}{2} \right) \right] \quad (9)$$

where t is the time coordinate and S is the signal of a peak containing n points. Finally, a chart containing the peak and baseline is plotted in the worksheet.

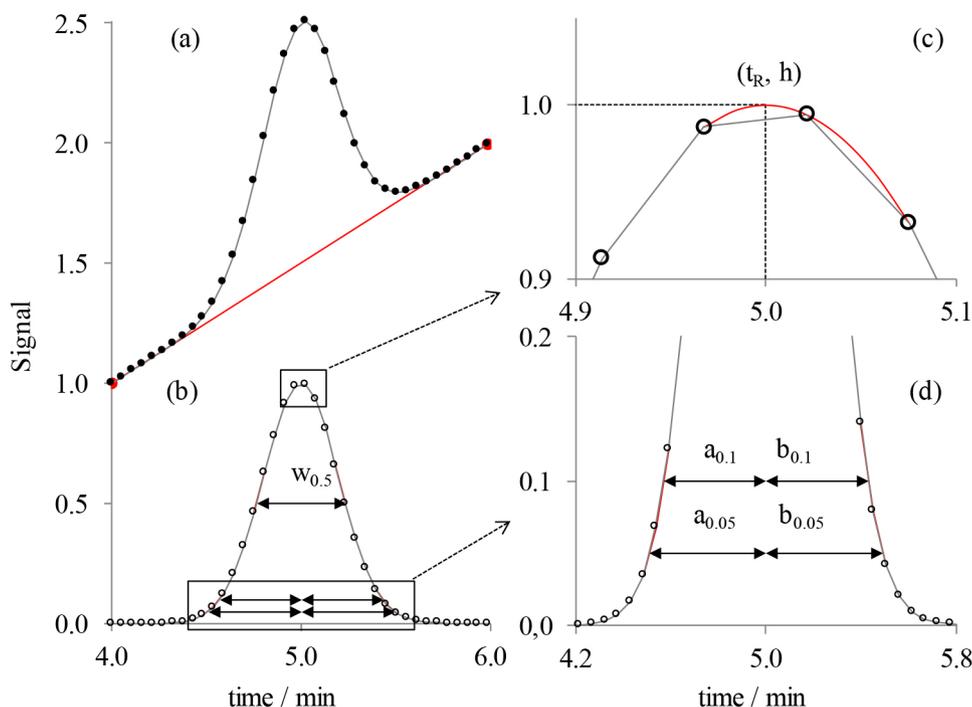


Figure 2. (a) Simulated peak (original signal) with sloped baseline (red line, $y = 0.5t - 1$); (b) adjusted peak (baseline subtracted), from which t_R , h , A , $w_{0.5}$, N , A_s and T_f are obtained; (c) apex of the adjusted peak in detail, showing the parabola fitted to the three highest points; (d) base of the peak in detail, showing red straight lines used to obtain a and b , at $0.1h$ and $0.05h$, for A_s and T_f calculations.

It is important to stress that the calculation of statistical moments of a peak is also possible. In this optional step (useful for evaluating with accuracy any type of peak shape), zeroth moment (m_0) up to fourth central moment (m_4), and N are obtained. If there are two or more peaks in a chromatogram, the resolutions between adjacent peaks are calculated in a further step as $R_{i,i+1} = 1.175 \times (t_{Ri+1} - t_{Ri}) / (w_{0.5i} + w_{0.5,i+1})$.

Results and Discussion

Effect of peak shape on accuracy

The peak recognition and integration accuracy were tested in a simulated chromatogram with 14 min at a sampling rate of 2.0 Hz, without noise, containing positive, negative, symmetric, fronting and tailing peaks with heights varying in six magnitude orders (from 1 to 100000 units, Supplementary Information Figure S5). The program was submitted to detect the six peaks automatically (with no smoothing). Since t_R and h of peaks simulated through equation 1 are known, relative errors between reference and obtained values can be calculated, as shown in Table 1. The a values (second column) were used in equation 1 to provide the resulting A_s obtained from the integrator (values greater than 1 indicate tailing; less than 1 are fronting; and equal to 1 are symmetric Gaussian peaks).

It is possible to see that the results obtained from asymmetric peaks are less accurate than that obtained for the symmetric peak (peak 6). However, even the highest relative error found (peak 1) is still small. Besides, positive and negative peaks with same dimensions, e.g., 3 and 4, show equivalent results. The peak (true) height also has some effect on the accuracy of the measured height, since the relative error tends to decrease for taller peaks. Finally, the same chromatogram was simulated with sloped baselines ($y = 1000t$ and $y = -1000t$) and the integration results including all available parameters were identical (therefore, not shown) to those obtained when baseline is horizontal. These results indicate that the peak shape and slope of the chromatogram baseline do not seriously affect the detection and integration of the peaks.

Effect of noise over threshold ranges

Figure 3a shows a simulated chromatogram segment from 2 to 4 min (0.5 Hz), without noise, with a single symmetric Gaussian peak at 3.0 min, $s = 1.5$, $h = 1.0$ and a sloped baseline ($y = -t$). Figures 3b and 3c show, respectively, FD and SD from this peak. In Figure 3d, the same peak was plotted, but with noise ($SD = 0.01$) sufficient to provide an $S/N = 100$. Figures 3e and 3f show FD and SD. These peaks were submitted to recognition through the proposed algorithm, with sensitivity set to 3.0 and without smoothing.

Table 1. Relative errors for retention time (t_R) and height (h) of a simulated chromatogram with different peak shapes

Peak	a^a	As^b	Reference t_R^c	Obtained t_R^d	Error / %	Reference h^c	Obtained h^d	Error / %
1	-1.00	0.441	2	1.9999	-7.5×10^{-3}	1	0.99988	-1.2×10^{-2}
2	-0.50	0.692	4	3.9999	-1.5×10^{-3}	-10	-9.99993	-7.5×10^{-4}
3	1.00	2.267	6	6.0002	2.6×10^{-3}	1000	1000.002	1.5×10^{-4}
4	1.00	2.267	8	8.0002	1.9×10^{-3}	-1000	-1000.002	1.5×10^{-4}
5	1.25	3.081	10	10.0002	2.4×10^{-3}	10000	10000.05	4.6×10^{-4}
6	0.0001	1.000	12	12.0000	6.2×10^{-8}	-100000	-100000.0	1.6×10^{-7}

^aAsymmetry term (a) from equation 1; ^basymmetry factor ($b_{0,1} / a_{0,1}$) obtained from integrator; ^creference retention time (t_R) and height (h) values used in equation 1; ^dobtained t_R and h values.

As sensitivity is constant (in this case), the distance between red horizontal lines (threshold ranges), are only controlled by M_D , from equation 4. The M_D of FD for the left peak (Figure 3b) is 0.09 and for the right peak (Figure 3e) is 0.27 (three times higher). The only contribution for the M_D in left peak is the signal variation, but for the right peak, there is also a noise contribution, leading the threshold range to adapt automatically for each case. The same reasoning is valid for SD. In this case, M_D for the left peak (Figure 3c) is 1.8 and for the right peak (Figure 3f) is 5.7 (again, about three times higher). It is important to keep in mind that the peak window (between vertical dashed arrows) is defined by

the algorithm as the time range where both FD and SD are outside threshold ranges. Outside the peak window, FD and SD are necessarily inside threshold ranges. Thus, when no noise is present, making threshold ranges smaller, even small changes in the slope and curvature are perceived, resulting in a larger peak baseline. Otherwise, these small changes are confounded with noise, leading to a shorter baseline.

Effect of noise over recognition precision

To study the S / N effect over the recognition and, consequently, the precision of integration results, symmetrical

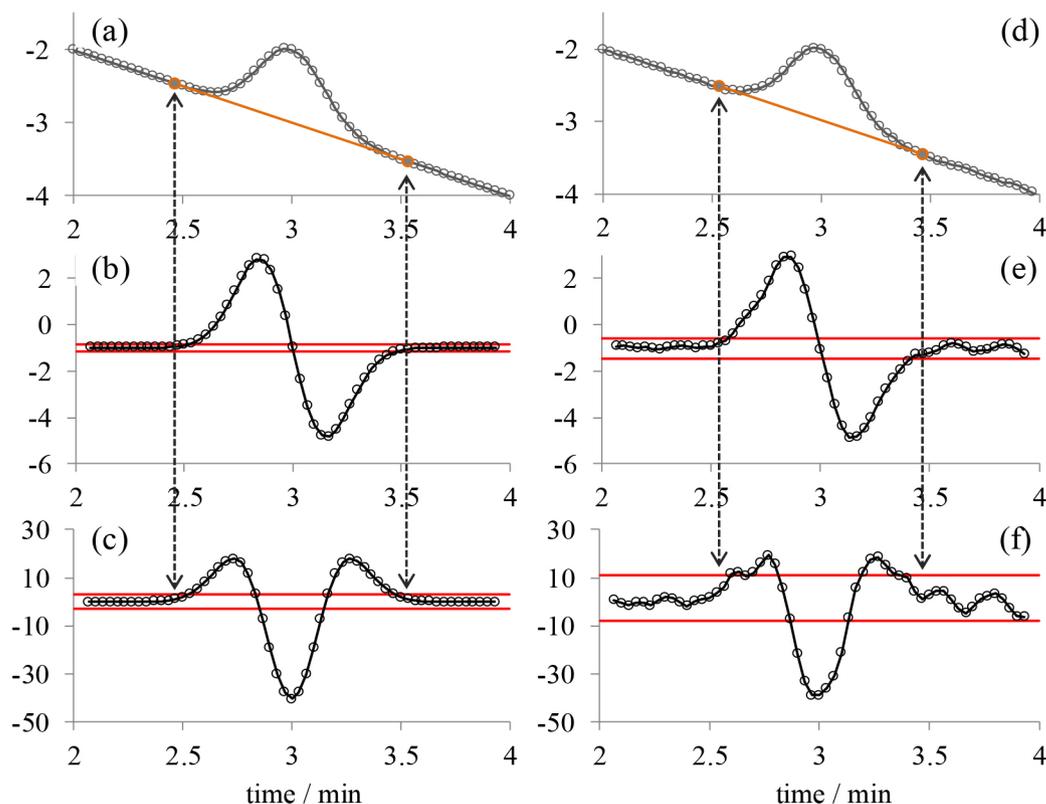


Figure 3. (a) Gaussian peak with negative slope baseline ($y = -t$); (b) first and (c) second derivatives of the Gaussian curve. Red horizontal lines define the threshold ranges, outside of which the peak window (between vertical arrows) is obtained. Sensitivity set to 3.0. Idem for (d), (e), and (f), except for the noise added ($SD = 0.01$, $S / N = 100$).

($a = 0$), tailing ($a = 1$) and fronting ($a = -1$) peaks with Gaussian height ($h = 0.3989$) and standard deviation ($s = 1$) were simulated with six levels of noise (five replicates for each level) at a sampling rate of $1 / 15$ Hz (to provide about 30 to 40 data points to draw a peak with ca. 10 min of base width). The chromatograms were submitted to peak recognition with default sensitivity (3.0) and subsequent integration. Figure 4 shows the relative standard deviations (RSD) of t_R , A , h , $w_{0.5}$ and N plotted against S / N levels. Both available smoothing modes were used.

The main effect when S / N increases is the overall decrease of RSD for all parameters, as expected, since the integration depends on the correct allocation of peak baseline. As the position of starting time (t_s) and ending time (t_E) gets more hidden in the noise, more uncertainty arises. No significant difference is observed when comparing

MA smoothing (used for the left graphs) with the SG method (right graphs). While tailing peaks (Figure 4b) are severely affected when S / N is low, Gaussian (Figure 4a) and fronting (Figure 4c) peaks provide RSD lower than 14%. The less affected parameter on all conditions is clearly t_R . In fact, even the highest noises added to the peak contributed less than 1% to its RSD. The RSDs of other parameters are more grouped, suggesting equivalent influences of S / N . Because N depends on t_R and $w_{0.5}$, its RSD is a bit larger than RSD of $w_{0.5}$, providing a similar profile in all S / N range.

Effect of smoothing over recognition

Figure 5a shows a chromatogram with three peaks ($h = 0.399$; $s = 0.5$; $a_1 = 1$; $a_2 = 0.00001$; $a_3 = -1$; sampling rate = $2 / 15$ Hz and $S / N = 10$), smoothed by SG and MA

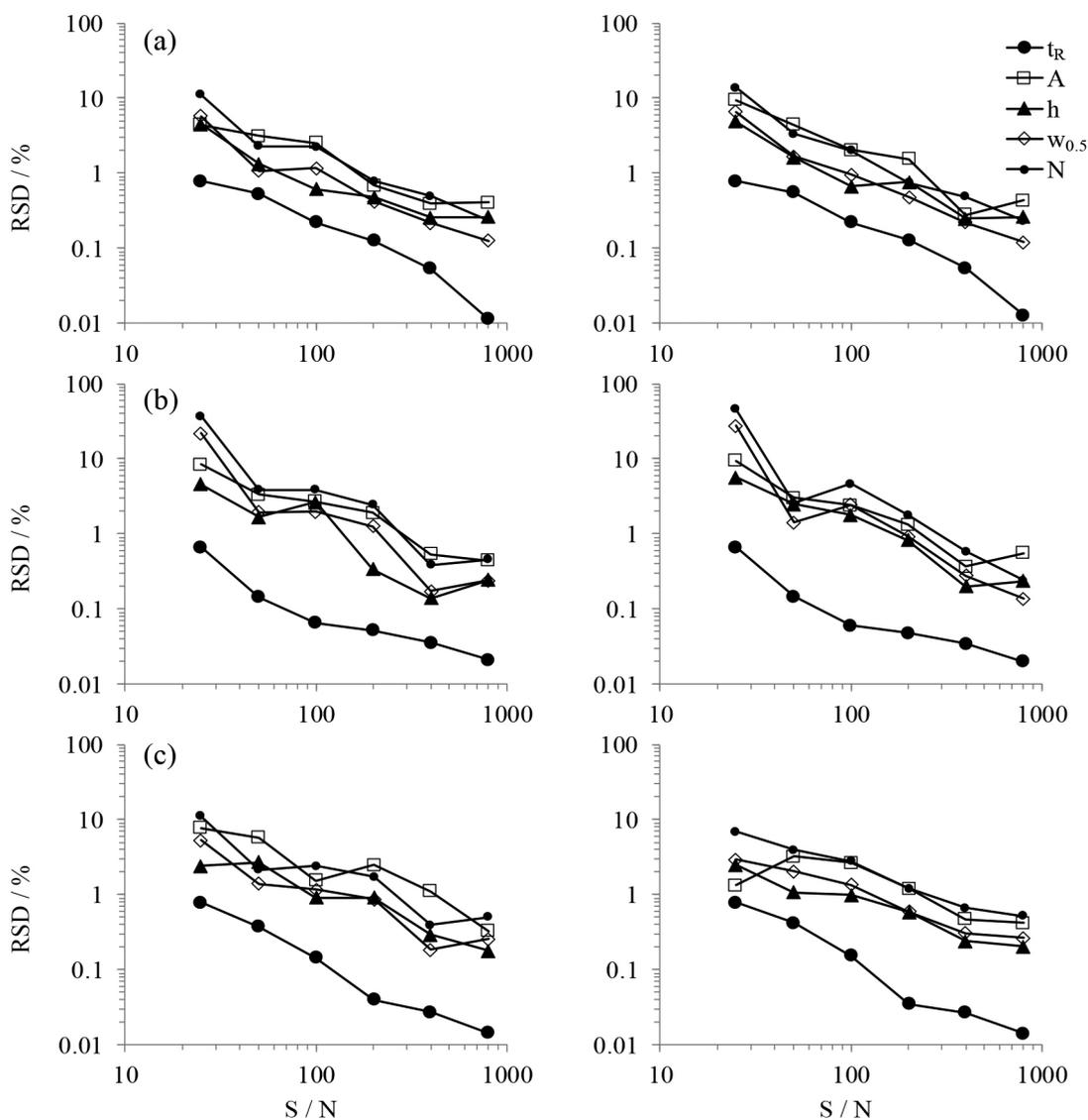


Figure 4. Precision of recognition of simulated (a) Gaussian, (b) tailing and (c) fronting peaks as a function of signal-to-noise ratio (true h / SD of baseline). Left graphs were obtained through moving average smoothing and the right ones through Savitzky-Golay polynomial smoothing.

(with several window sizes) methods. The chromatogram of Figure 5b has $S/N = 50$.

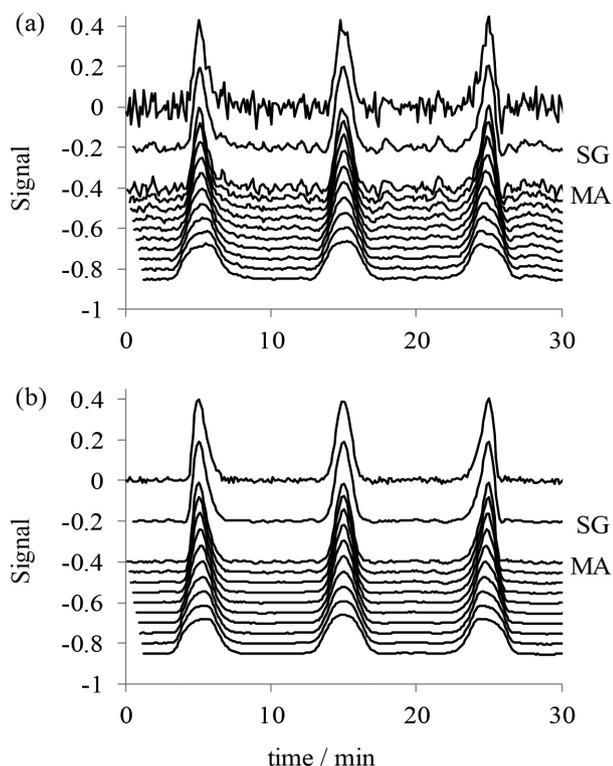


Figure 5. Simulated chromatograms with (a) $S/N = 10$ and (b) $S/N = 50$. The first chromatogram of each graph (top) is the original data; the second was obtained through Savitzky-Golay (SG) polynomial smoothing, and the ten last were obtained through moving average (MA) smoothing, with increasing window size (3-21 points, odd numbers).

The peaks from the chromatogram (Figure 5a) could not be detected without smoothing. With the SG smoothing, the first peak (tailing) was poorly detected (one fragment detected), while the second (Gaussian) was better defined

and the third (fronting) was not detected. This recognition profile was similar to that made after MA smoothing with 3 points window (first MA curve). In fact, these smoothed curves are similar. From 5 to 15 (next six MA curves) points, MA smoothing led to detection of all peaks. With more than 15 points (last three MA curves), the shape of the peaks is damaged and recognition fails (two fragments *per* peak detected). For the chromatogram (Figure 5b), the first two peaks were recognized directly on original data. With SG and MA (with up to 17 points) smoothing modes, the three peaks were normally detected. When the number of points of MA smoothing window increases the peak height decreases and the width enlarges. As a result, the found peak limits may not be appropriated.

The SG method provided a smoothing with greater fidelity to the original peaks, but with lower capacity to reduce the baseline noise. On the other hand, the MA smoothing was more efficient for reducing noise, although the peak shape is damaged with height and asymmetry losses, when higher number of points are used. Thus, if SG smoothing is not adequate to improve recognition, the window size of MS smoothing should be as small as possible to detect satisfactorily all desired peaks.

Liquid chromatography data treatment

To illustrate the applicability of the proposed program to several real situations, some separation techniques were used and the chromatograms were submitted to recognition and integration. Figure 6 shows the recognition of lactulose (1) and lactose (2) peaks in the presence of impurities separated by LC of a commercial milk sample. The position of the baselines indicates the automatic recognition profile with MA smoothing (9 points) and sensitivity set to 9. With these

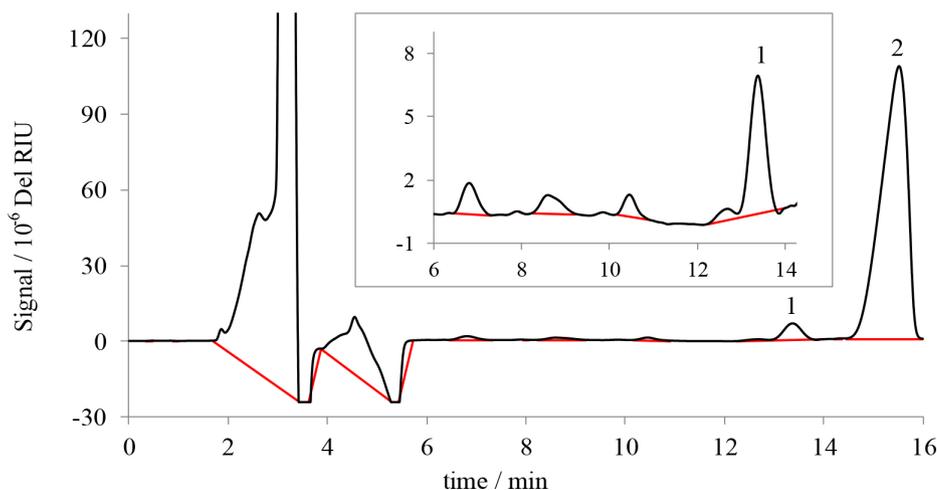


Figure 6. LC chromatogram of a milk sample for separation of lactulose (inset, $1-1.22 \text{ mmol L}^{-1}$) and lactose ($2-143 \text{ mmol L}^{-1}$), with index refraction detection. The baselines (red) indicate the automatic recognition profile with MA smoothing (9 points) and sensitivity set to 9.

settings, impurity peaks in time range between 6.00 and 14.25 min were also detected (Figure 6 inset).

Capillary zone electrophoresis data treatment

Figure 7 shows data obtained from a lab-made C⁴D, where lactulose (1) and lactose (2) were partially separated by CZE. The inset shows the recognition profile of these peaks, which were split through drop-line mode in a further step. Although noise is not apparent, an MA smooth (9-point window) was applied in this electropherogram, in order to avoid excess of peak fragments detection. Nevertheless, abrupt variations in electropherogram profile, e.g., system peak in 4 min and electroosmotic flow (EOF) signal are still detected as peaks and thus, should be ignored.

A more detailed study was made through separation of a mixture of ten organic acids standards by CZE (Figure 8), under indirect UV detection (220 nm), in order to compare the recognition and integration results of the present program with ChemStation. Table 2 shows the comparisons between t_s , t_E , t_R , A , h and $w_{0.5}$ through paired t -test. Chromophoreasy was set with SG smoothing with default sensitivity and ChemStation was configured to detect negative peaks with slope sensitivity set to 100 and

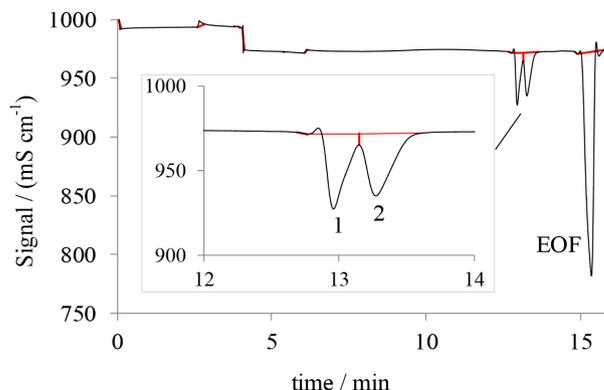


Figure 7. Electropherogram showing partial separation of lactulose (1-1.22 mmol L⁻¹) and lactose (2-143 mmol L⁻¹) by CZE, with C⁴D. The baselines indicate the peak recognition profile with MA smoothing (9-point window) and sensitivity set to 2. Electroosmotic flow signal (EOF).

peak width set to 0.02 in the time events. To match the units and signals of the programs, the areas of Chromophoreasy (initially given in mAU min) were multiplied by -60 s min^{-1} and the heights multiplied by -1 . The limits of the pyruvic acid peak (2) were manually adjusted in both programs.

A systematic difference on t_s between the programs can be observed in Table 2, evidenced by the elevated $t_{\text{calculated}}$ value (4.453). t_s values obtained by ChemStation

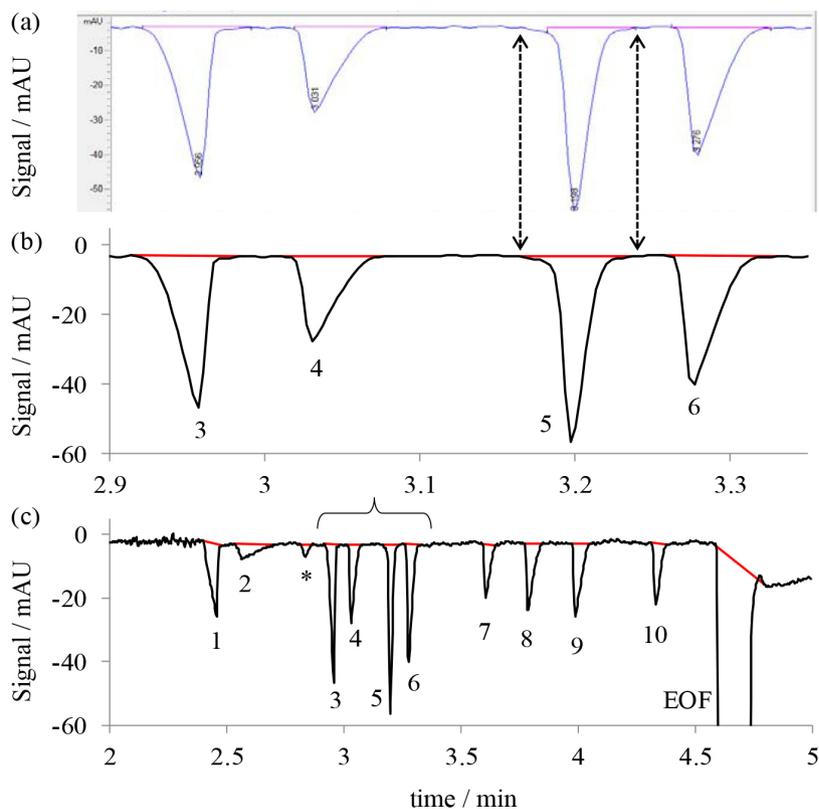


Figure 8. Electropherogram of organic acids (1 mmol L⁻¹ each): oxalic (1), pyruvic (2), tartaric (3), citric (4), formic (5), malic (6), lactic (7), succinic (8), aspartic (9) and acetic (10); *unidentified peak; electroosmotic flow (EOF). (a) ChemStation and (b) Chromophoreasy stretched views of peaks 3-6 from (c) electropherogram of all analytes. Dashed arrows indicate the Chromophoreasy integration limits of peak 5.

Table 2. Comparison of integration results of Chromophoreasy and ChemStation for the electropherogram of Figure 8

Peak ^a	t _s / min	t _E / min	t _R / min	A / (mAU s)	h / mAU	w _{0.5} / min
Chromophoreasy						
1	2.39	2.49	2.46	48.6	23.3	0.0356
2	2.53	2.71	2.57	23.4	4.8	0.0608
*	2.81	2.87	2.83	5.3	3.9	0.0221
3	2.91	2.99	2.96	52.3	43.9	0.0185
4	3.01	3.08	3.03	34.6	24.8	0.0216
5	3.16	3.24	3.20	54.2	53.9	0.0152
6	3.25	3.33	3.28	51.0	37.1	0.0217
7	3.58	3.66	3.61	26.0	17.3	0.0244
8	3.76	3.86	3.79	39.6	21.7	0.0289
9	3.96	4.06	3.99	45.9	23.4	0.0314
10	4.31	4.39	4.33	31.7	19.5	0.0258
EOF	4.58	4.81	4.67	2326.4	346.5	0.1207
ChemStation						
1	2.40	2.52	2.46	47.5	23.2	0.0269
2	2.54	2.69	2.57	19.8	4.5	0.0725
*	2.81	2.88	2.83	6.6	4.2	0.0233
3	2.92	2.99	2.96	52.8	44.0	0.0170
4	3.02	3.08	3.03	34.7	24.9	0.0192
5	3.18	3.24	3.20	53.7	54.0	0.0156
6	3.26	3.32	3.28	50.5	37.0	0.0199
7	3.59	3.66	3.61	29.0	17.8	0.0219
8	3.77	3.85	3.79	41.1	22.0	0.0258
9	3.98	4.05	3.99	47.1	23.6	0.0283
10	4.31	4.39	4.33	32.0	19.4	0.0243
EOF	4.58	4.78	4.67	2345.2	347.6	0.1187
Comparison ^b						
t _{calculated}	4.453	0.454	1.227	1.078	1.555	0.809

^aPeaks: organic acids: oxalic (1), pyruvic (2), tartaric (3), citric (4), formic (5), malic (6), lactic (7), succinic (8), aspartic (9) and acetic (10); unidentified peak (*); and electroosmotic flow signal (EOF); ^bif $t_{\text{calculated}} < t_{(11;0.05/2)} = 2.201$, there are no significant differences between two set of data in a column (at a 95% confidence level). t_s: starting time; t_E: ending time; t_R: retention time; A: area; h: height; w_{0.5}: half-height width.

integration are slightly higher (delayed) than those obtained by Chromophoreasy, leading to this result. In fact, this difference can be observed in Figures 8a and 8b, evidenced by dashed arrows on peak 5. Left arrow points to t_s detected from Chromophoreasy, which is located at a different position related to t_s from ChemStation. However, no differences between t_E from programs are evident (as the right arrow shows). This observation is valid for all peaks in this electropherogram. Other parameters showed statistically similar behaviors. Figure 8c shows the electropherogram used for this comparison. The S / N varied from 10.5 (peak *) to 143.6 (peak 5).

Capillary electrochromatography data treatment

Figure 9 shows an electrochromatogram of a PAH standard analysis. As there is no unique suitable wavelength for detection of all analytes simultaneously, 220 and 250 nm data were collected. MA smoothing and sensitivity set to

2 were the best choice for both wavelengths. The group of peaks 2 and 3 (at 220 nm) and the group 4 and 5 (at 220 and 250 nm) were split with drop-line mode. This electrochromatogram is a good example of the program application in drifting baseline. As stated earlier, the slope of the chromatogram baseline does not seem to have affected the recognition of the peaks.

Gas chromatography data treatment

Five different biodiesel samples were analyzed in GC equipment with FID (Figure 10). In this case, Chromophoreasy was set to detect peaks from 0 to 6 min, with SG polynomial smoothing and sensitivity set to 1. In GC Solution software peaks were detected and integrated automatically. Table 3 shows the results of comparisons made through paired *t*-test between Chromophoreasy and GC Solution for A and h. To match the units, the areas of Chromophoreasy (given in V min) were multiplied by 60 s min⁻¹.

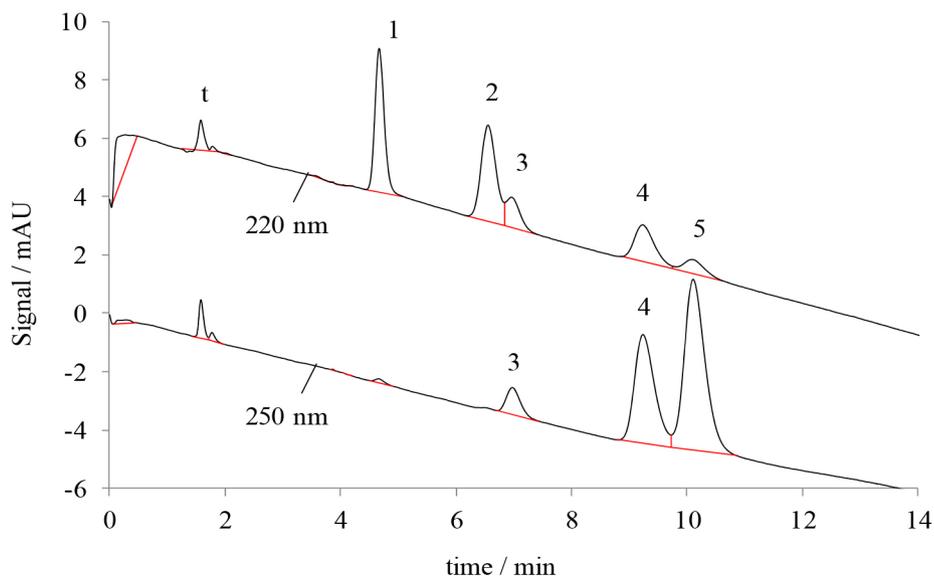


Figure 9. Electrochromatogram of PAH (1 mmol L^{-1} each): naphthalene (1), acenaphthene (2), fluorene (3), phenanthrene (4) and anthracene (5). Thiourea (t) was used as flow marker. The 220 nm data were plotted 4 mAU higher for a better view.

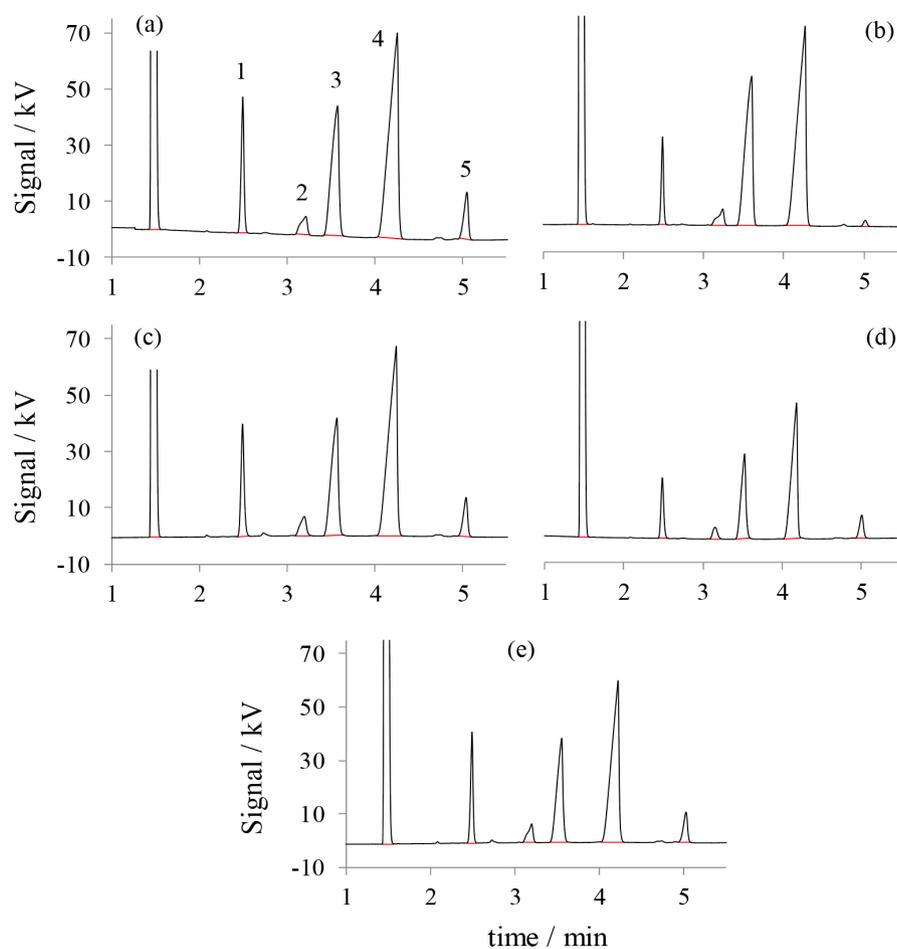


Figure 10. Chromatograms of biodiesel samples from different sources: basic catalysis of (a) soybean oil; (b) sunflower oil; (c) food frying oils; (d) acid catalysis of soybean oil; (e) acid pre-treatment followed by basic catalysis of food frying oil. Analytes are fatty acids methyl esters: methyl palmitate (1), methyl stearate (2), methyl oleate (3), methyl linoleate (4), and methyl linolenate (5). The sequence of analytes is the same for all graphs.

Table 3. Comparison of areas (A) and heights (h) from Chromophoreasy and GC Solution for chromatograms of Figure 10

Peak ^a	A in samples ^b / (kV s)					h in samples / kV				
	A	B	C	D	E	A	B	C	D	E
Chromophoreasy										
1	108.7	63.4	99.9	52.1	83.5	48.4	31.7	39.8	21.3	41.5
2	34.7	30.8	32.4	16.5	29.0	6.6	5.9	6.7	4.1	7.0
3	267.4	342.9	227.2	124.9	187.9	46.3	53.8	41.6	30.0	38.9
4	514.0	512.8	445.0	237.8	361.2	73.2	71.9	67.2	47.7	60.4
5	59.2	6.5	47.8	27.8	38.1	16.8	2.2	13.8	8.1	11.3
GC Solution										
1	109.4	64.3	100.4	52.7	84.1	47.9	31.0	39.4	21.3	41.3
2	36.3	31.6	33.3	17.5	29.5	6.7	5.9	6.8	4.2	7.0
3	271.3	343.6	229.9	127.1	189.8	46.4	53.6	41.7	29.9	38.7
4	520.6	515.6	449.4	243.3	364.6	72.8	71.8	67.1	47.9	60.1
5	60.8	6.9	48.9	29.2	38.3	16.8	2.2	13.8	8.2	11.3
Comparison ^c										
$t_{\text{calculated}}$	2.725	2.535	2.732	2.412	2.225	0.863	1.525	0.636	1.166	2.363

^aFatty acids methyl esters: methyl palmitate (1), methyl stearate (2), methyl oleate (3), methyl linoleate (4) and methyl linolenate (5); ^bbiodiesel samples obtained from different sources: basic catalysis of soybean (A), sunflower (B) and food frying (C) oils, acid catalysis of soybean oil (D), acid pre-treatment followed by basic catalysis of food frying oil (E); ^cif $t_{\text{calculated}} < t_{(4;0.05/2)} = 2.776$, there are no significant differences between two set of data in a column (at a 95% confidence level). A: area; h: height.

In Table 3, a small difference between areas from Chromophoreasy and GC Solution can be observed. In fact, t_s and t_E obtained in these programs (not shown) are slightly different, probably because of differences in peak recognition algorithms. Thus, GC Solution areas are a little greater than the Chromophoreasy ones. However, results are still statistically similar, with all $t_{\text{calculated}}$ lower than $t_{(4;0.05/2)}$ (2.776).

Conclusions

In this work, the development, evaluation, features and applications of an alternative program for recognition and integration of chromatographic and electrophoretic peaks in the familiar environment of Excel were demonstrated. The possibility to adjust parameters, such as time range, sensitivity of threshold ranges and chromatogram smoothing, developed to increase the efficiency of peak detection, allowed the program application for several experimental situations, including analysis of standards and samples through LC, GC, CZE and CEC techniques with various types of detectors, in addition to simulated chromatograms. The results are easily upgradeable in cases where adjustments are necessary. Some data and formulas can be found in the cells where they were generated, so one can view how the recognition and integration were obtained. This feature may be useful for academic purposes, for instance.

The use of the several functions, such as peak recognition, extraction, integration, results grouping and plotting chromatogram with baseline, which could be performed separately or through a single command, made the data treatment more practical and provided high throughput. Therefore, the study of the effect of peak shape, chromatogram noise, smoothing modes over accuracy and precision was made with low time consumption. Finally, the proposed program showed to be a reliable tool, providing statistically similar results when compared with other commercial software used, meeting the proposal addressed in this paper.

Supplementary Information

Supplementary data (Experimental section, additional tools, recognition efficiency, figures), Chromophoreasy program and electropherogram sample are available free of charge at <http://jbcs.sbq.org.br> as PDF file.

Acknowledgments

The authors wish to acknowledge Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES: PNPd 23038.007000/2011-70), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq: 471288/2013-6 and 302432/2014-0) and Fundação de

Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG: CEX-PPM 00398-13) for fellowships and financial support.

References

1. Righetti, P. G.; *J. Chromatogr. A* **2005**, *1079*, 24.
2. Meinert, C.; Meierhenrich, U. J.; *Angew. Chem., Int. Ed.* **2012**, *51*, 10460.
3. Núñez, O.; Gallart-Ayala, H.; Martins, C. P. B.; Lucci, P.; *J. Chromatogr. A* **2012**, *1228*, 298.
4. Li, J.; *Anal. Chim. Acta* **1999**, *388*, 187.
5. Mostafa, A.; Edwards, M.; Górecki, T.; *J. Chromatogr. A* **2012**, *1255*, 38.
6. Ferreira, S. L. C.; Bruns, R. E.; da Silva, E. G. P.; dos Santos, W. N. L.; Quintella, C. M.; David, J. M.; de Andrade, J. B.; Breitreit, M. C.; Jardim, I. C. S. F.; Neto, B. B.; *J. Chromatogr. A* **2007**, *1158*, 2.
7. Thompson, M.; Ellison, S. L. R.; Wood, R.; *Pure Appl. Chem.* **2002**, *74*, 835.
8. Ribani, M.; Bottoli, C. B. G.; Collins, C. H.; Jardim, I. C. S. F.; Melo, L. F. C.; *Quim. Nova* **2004**, *27*, 771.
9. Dyson, N.; *J. Chromatogr. A* **1999**, *842*, 321.
10. Zhang, J.; Gonzalez, E.; Hestilow, T.; Haskins, W.; Huang, Y.; *Curr. Genomics* **2009**, *10*, 388.
11. Vivó-Truyols, G.; Torres-Lapasió, J. R.; van Niderkassel, A. M.; Vander Heyden, Y.; Massart, D. L.; *J. Chromatogr. A* **2005**, *1096*, 133.
12. Peters, S.; Vivó-Truyols, G.; Marriott, P. J.; Schoenmakers, P. J.; *J. Chromatogr. A* **2007**, *1156*, 14.
13. Fong, S. S.; Rearden, P.; Kanchagar, C.; Sassetti, C.; Trevejo, J.; Brereton, R. G.; *Anal. Chem.* **2011**, *83*, 1537.
14. Yu, Y.-J.; Xia, Q.-L.; Wang, S.; Wang, B.; Xie, F.-W.; Zhang, X.-B.; Ma, Y.-M.; Wu, H.-L.; *J. Chromatogr. A* **2014**, *1359*, 262.
15. Wang, X.; Zhao, Y.; Sun, P.; Ji, M.; Bao, M.; *Anal. Methods* **2015**, *7*, 2670.
16. Marco, V. B. D.; Bombi, G. G.; *J. Chromatogr. A* **2001**, *931*, 1.
17. Vivó-Truyols, G.; Torres-Lapasió, J. R.; van Niderkassel, A. M.; Vander Heyden, Y.; Massart, D. L.; *J. Chromatogr. A* **2005**, *1096*, 146.
18. Creek, D. J.; Jankevics, A.; Burgess, K. E. V.; Breitling, R.; Barrett, M. P.; *Bioinformatics* **2012**, *28*, 1048.
19. Fasoula, S.; Zisi, C.; Gika, H.; Pappa-Louisi, A.; Nikitas, P.; *J. Chromatogr. A* **2015**, *1395*, 109.
20. Kadjo, A.; Dasgupta, P. K.; *Anal. Chim. Acta* **2013**, *773*, 1.
21. Chavez-Servin, J. L.; Castellote, A. I.; Lopez-Sabater, M. C.; *J. Chromatogr. A* **2004**, *1043*, 211.
22. Ferrari, R. A.; Oliveira, V. S.; Scabio, A.; *Quim. Nova* **2005**, *28*, 19.
23. Delmonte, P.; Kia, A.-R. F.; Kramer, J. K. G.; Mossoba, M. M.; Sidisky, L.; Rader, J. I.; *J. Chromatogr. A* **2011**, *1218*, 545.
24. Vaz, F. A. S.; da Silva, P. A.; Passos, L. P.; Heller, M.; Mücke, G. A.; Costa, A. C. O.; de Oliveira, M. A. L.; *Phytochem. Anal.* **2012**, *23*, 569.
25. Soga, T.; Serwe, M.; *Food Chem.* **2000**, *69*, 339.
26. Fracassi da Silva, J. A.; do Lago, C. L.; *Anal. Chem.* **1998**, *70*, 4339.
27. Vaz, F. A. S.; Moutinho, A. D.; Mendonça, J. P. R. F.; Araújo, R. T.; Ribeiro, S. J. L.; Polachini, F. C.; Messaddeq, Y.; Oliveira, M. A. L.; *Microchem. J.* **2012**, *100*, 21.
28. Pápai, Z.; Pap, T. L.; *J. Chromatogr. A* **2002**, *953*, 31.
29. Savitzky, A.; Golay, M. J. E.; *Anal. Chem.* **1964**, *36*, 1627.
30. Dyson, N.; *Chromatographic Integration Methods*, 2nd ed.; RSC: Loughborough, 1998.

Submitted: January 21, 2016

Published online: March 15, 2016