Journal of Chromatography A, 1217 (2010) 1375–1385

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Probability of failure of the watershed algorithm for peak detection in comprehensive two-dimensional chromatography

Gabriel Vivó-Truyols [a,*], Hans-Gerd Janssen [a,b]

[a] Analytical-Chemistry Group, van't Hoff Institute for Molecular Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands
[b] Unilever Research and Development Vlaardingen, Advanced Measurement and Data Modelling, P.O. Box 114, 3130 AC Vlaardingen, The Netherlands

ARTICLE INFO

Article history:
Received 29 September 2009
Received in revised form
15 December 2009
Accepted 22 December 2009
Available online 4 January 2010

Keywords:
Two-dimensional chromatography
Watershed algorithm
Peak detection

ABSTRACT

The watershed algorithm is the most common method used for peak detection and integration in two-dimensional chromatography. However, the retention time variability in the second dimension may render the algorithm to fail. A study calculating the probabilities of failure of the watershed algorithm was performed. The main objective was to calculate the maximum second-dimension retention time variability, $\Delta^2 t_{R,crit}$, above which the algorithm fails. Several models to calculate $\Delta^2 t_{R,crit}$ were developed and evaluated: (a) exact model; (b) simplified model and (c) simple-modified model. Model (c) gave the best performance and allowed to deduce an analytical expression for the probability of failure of the watershed algorithm as a function of experimental $\Delta^2 t_R$, modulation time and peak width in the first and second dimensions. It could be demonstrated that the probability of failure of the watershed algorithm under normal conditions in GC × GC is around 15–20%. Small changes of $\Delta^2 t_R$, modulation time and/or peak width in the first and second dimension could induce subtle changes in the probability of failure of the watershed algorithm. Theoretical equations were verified with experimental results from a diesel sample injected in GC × GC and were found to be in good agreement with the experiments.

1. Introduction

The growing complexity of the data generated by modern liquid chromatography (LC) and gas chromatography (GC) systems requires the development of new data analysis algorithms. The algorithms to be applied depend on the application, and range from base-line treatment to chromatogram alignment methods. In most of the applications, peak detection (and peak integration) is one of the key steps in the analysis process. Peak detection might be troublesome when complex chromatograms are being analysed, with peak numbers easily exceeding the thousands.

In one-dimensional chromatography with single-channel detection, peak-detection methods are almost fully developed. They are based on detecting a raise of the signal coming from the detector and applying the condition of unimodality (i.e., the signal should have only one maximum). Two main families of peak-detection methods have been developed [1]: those that make use of derivatives, and those that make use of matched filters. When a multi-channel detector is used (e.g., MS), new possibilities for peak detection are possible. Different algorithms have been developed in order to make use of the relational information provided by the existence of more than one detection channel. In particular, the advent of the "-omics" disciplines has stimulated the development of a significant quantity of statistical tools, including novel methods for peak detection in chromatography. For a review, see [2–4].

Peak-detection methods in comprehensive two-dimensional chromatography are less advanced. This is mainly due to the fact that these techniques are not completely mature yet. Adapting the peak-detection algorithms developed for hyphenated techniques to two-dimensional chromatography is not straightforward for two reasons. First, in two-dimensional chromatography, the condition of unimodality holds for both dimensions (a chromatographic peak has only a single retention time in both the first and the second dimension). This condition is normally not met in multi-channel detection. Second, a modulation cycle in comprehensive two-dimensional chromatography is normally several orders of magnitude longer than the detector's sampling rate. This makes a chromatogram in two dimensions to appear undersampled in the first dimension as opposed to the highly sampled chromatogram obtained with multi-channel detection. One should note that this second condition does not apply when the two-dimensional separation is performed in space (such as in 2D-PAGE electrophoresis, or two-dimensional thin layer chromatography). Opposed to separations in space, LC × LC or GC × GC are two-dimensional chromatographic methods in time. In these methods, a peak is analysed only a limited number of times by a (fast) second dimension during its elution in the (slow) first dimension, hence the low sampling rate in the first dimension. This article

* Corresponding author. Tel.: +31 205256531.
E-mail address: g.vivotruyols@uva.nl (G. Vivó-Truyols).

0021-9673/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2009.12.063

is devoted only to time-driven two-dimensional separations (i.e., GC × GC or LC × LC), so it is not applicable to spatially separated chromatograms (e.g., 2D-PAGE).

So far only a limited number of peak-detection methods for (time-driven) two-dimensional chromatography has been described in the literature [5–7]. Only two main families of methods are available, those based on the watershed algorithm [8], and those based on an extension of the one-dimensional peak-detection algorithms [9,10]. The main difference between the two families of methods relies in the fact that watershed-algorithm based methods make use of the true two-dimensional image generated in two-dimensional chromatography, whereas the extended one-dimensional algorithms are based on the analysis of the one-dimensional raw signal arising from the detector. The watershed algorithm was originally developed to delimitate single catchment areas of geographic zones [11] (see Section 2.2.1 for a detailed explanation), and has been adapted to peak detection in two-dimensional chromatography by Reichenbach et al. [8]. Methods of the second family are normally based on a two-step procedure. In a first step, one-dimensional peak-detection algorithms are applied to the raw, one-dimensional signal. In a second step, the previously detected peaks are then "merged" after it has been decided that they belong to the same modulated compound.

As it will be demonstrated in this paper, one of the main drawbacks of the watershed algorithm is its intolerance to second-dimension retention time variability. This intolerance may bring the algorithm to fail, splitting a two-dimensional peak into two peaks (or two catchment areas), when there is only a single two-dimensional peak. Unfortunately, second-dimension retention time variability is unavoidable, and hence so is failure of the watershed algorithm. In this article, a study is performed to predict in which situations the watershed algorithm will fail. A model for time-driven two-dimensional chromatographic peaks is developed. The model (applicable to both LC × LC and GC × GC) is used to calculate which combination of values for second-dimension retention time variability, first- and second-dimension peak width, modulation time and peak phase are not tolerated. An experimental study is performed in GC × GC to compare data calculated using the model (and its approximations) with experiments.

2. Theory

2.1. Peak model for two-dimensional chromatography

Let us suppose a two-dimensional chromatographic peak with known first- and second-dimension retention times (^1t_R and ^2t_R) and known first- and second-dimension peak widths ($^1\sigma$ and $^2\sigma$). The raw signal from the two-dimensional chromatograph (prior to any manipulation, including "folding" the data into a two-dimensional data table) is represented in Fig. 1. This signal can be modelled as a sum of sub-peaks:

$$y(t) = \sum_{i=-\infty}^{\infty} a_i y_i = \sum_{i=-\infty}^{\infty} a_i \frac{1}{{}^2\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-t_i}{{}^2\sigma}\right)^2\right] \quad (1)$$

where the subindex $i = -\infty, \ldots, -2, -1, 0, 1, 2, \ldots, \infty$ corresponds to the sub-peaks resulting from the modulated fractions of the first-dimension peak injected into the second dimension, a_i is the relative abundance of the ith modulated peak (see Eq. (2)), y_i represents the equation for the ith sub-peak, t_i is the retention time where the peak is represented (see Eq. (3)), and $^2\sigma$ is the peak width (measured as the standard deviation) of the sub-peak in the second dimension. Note that this model has two underlying assumptions. First, it assumes a constant value of $^2\sigma$ for the different sub-peaks. Second, it assumes a Gaussian, symmetric peak model. Both assumptions are not strictly true in practice, but these assumptions are not significant for our computations. As the quantity of material injected into the second dimension corresponds to the fraction of the first-dimension peak contained between $t_i - m/2$ and $t_i + m/2$ (i.e., one modulation period), a_i can be defined as:

$$a_i = \frac{A}{{}^1\sigma\sqrt{2\pi}} \int_{t_i-(m/2)}^{t_i+(m/2)} \exp\left[-\frac{1}{2}\left(\frac{t}{{}^1\sigma}\right)^2\right] dt \quad (2)$$

where m is the modulation time and A is a factor expressing the total abundance of the compound. Note that the expression inside the integral corresponds to an unmodulated peak arisen in the first dimension, and the integral limits correspond to the fraction of this peak contained between $t_i - m/2$ and $t_i + m/2$. This is a condition for the two-dimensional chromatography to be comprehensive. In practice, it may be possible to inject in the second dimension only part of the sample eluted from the first. In this case, as long as this split of the first-dimension eluent is constant along the elution, Eq. (2) is still valid (only parameter A has to be corrected). For simplicity, the first-dimension peak is centred around $t = 0$, but in practice

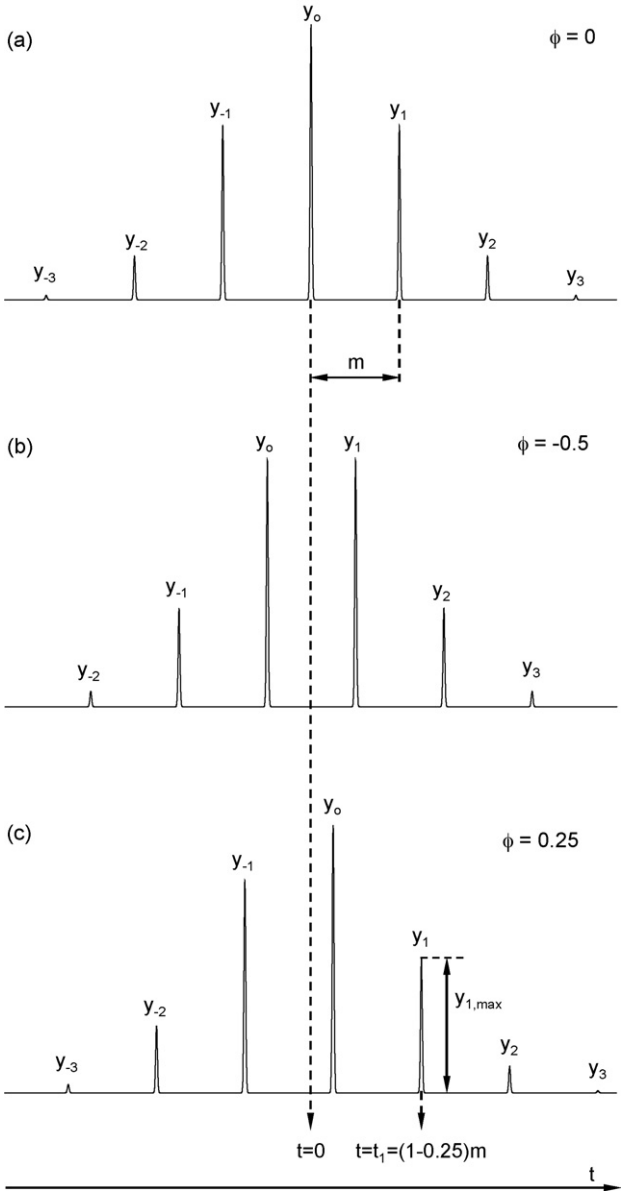

Fig. 1. Schematic representation of a modulated peak following Eqs. (1)–(4), with different values of ϕ ((a) $\phi = 0$; (b) $\phi = -0.5$; (c) $\phi = 0.25$). Only sub-peaks $i = -2$, -1, 0, 1, and 2 are represented. The value of the modulation time (m) is overlaid.

the peak is found at $t = {}^1t_R + {}^2t_R + m$. t_i is represented as:

$$t_i = (i + \phi)m \tag{3}$$

with ϕ being the modulation phase. ϕ may vary between $-1/2$ and $1/2$: the peak is "in phase" when $\phi = 0$ and "out of phase" when $\phi = -1/2$ or $1/2$. Rearranging Eqs. (1)–(3), the following expression is found for a collection of sub-peaks:

$$y(t) = \frac{A}{{}^1\sigma^2\sigma 2\pi} \sum_{i=-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{t-(i+\phi)m}{{}^2\sigma}\right)^2\right] \times \int_{m(i+\phi-(1/2))}^{m(i+\phi+(1/2))} \exp\left[-\frac{1}{2}\left(\frac{t}{{}^1\sigma}\right)^2\right] dt \tag{4}$$

One should note that, with this model, values of $i < 0$ correspond to sub-peaks that show an increasing peak height with time, whereas $i > 0$ correspond to sub-peaks with decreasing height over time. The sub-peak $i = 0$ corresponds to the sub-peak having the maximum peak height, except for the special case of $\phi = 0.5$ or $\phi = -0.5$ (where there appear two peaks with exactly the same height).

In the following computations it is useful to have an expression for a single sub-peak, $y_i(t)$. Such an expression can be straightforwardly derived by eliminating the sum in Eq. (4):

$$y_i(t) = \frac{A}{{}^1\sigma^2\sigma 2\pi} \exp\left[-\frac{1}{2}\left(\frac{t-(i+\phi)m}{{}^2\sigma}\right)^2\right] \times \int_{m(i+\phi-(1/2))}^{m(i+\phi+(1/2))} \exp\left[-\frac{1}{2}\left(\frac{t}{{}^1\sigma}\right)^2\right] dt \tag{5}$$

The chromatographic definition of a two-dimensional peak, according to this model, should be a collection of sub-peaks that (i) show the same second-dimension retention time (within some degree of tolerance due to instrument variability) and (ii) show only one maximum when the height of the sub-peaks is monitored.

2.2. Peak-detection methods for two-dimensional chromatography

2.2.1. The watershed algorithm

The so-called watershed algorithm (sometimes also called watershed transform) has been extensively used in many areas of image analysis [11]. In the context of two-dimensional chromatography, the algorithm is applied to the inverse of the image, hence a two-dimensional chromatographic peak (i.e., a mountain) appears like a negative peak (i.e., a basin). The algorithm works simulating a flood of the surface to its minima, and detecting the different basins that can be separated. The method has been applied to peak detection in two-dimensional gas chromatography [8], and 2D-gel electrophoresis [12]. Without pre-treatment, however, the watershed algorithm results in oversegmentation of the image, since noise disturbances tend to be assimilated as chromatographic peaks. Therefore, some kind of noise removal is normally applied prior to the watershed transform [8].

One disadvantage of the watershed algorithm is that it does not impose the condition of continuity for every sub-peak. This is illustrated in Fig. 2, where a peak from a GC × GC analysis of a diesel sample is detected with the watershed algorithm, and depicted in detail (for more details about the sample, see Section 3). The boxes overlaid on the contour plot indicate the regions that the watershed algorithm has detected as belonging to the peak. For simplicity, let us restrict our comments to the signal eluting at ${}^1t_R = 608$ s (marked with a large arrow). As can be seen, several interrupted regions (labelled from a to f) belong to the same two-dimensional peak.

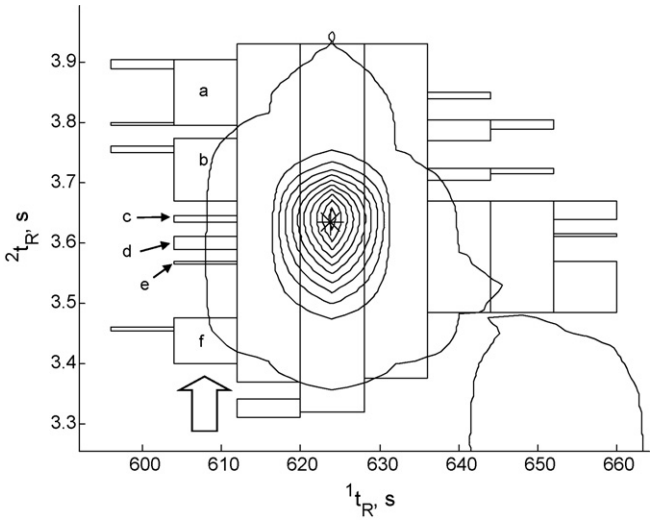

Fig. 2. Representation of a peak from the diesel sample (Section 3.1) detected with the watershed algorithm. Contour lines correspond to the FID signal intensity. Rectangles correspond to the peak region detected by the watershed algorithm. For explanation about the regions marked from a to f, see text. The symbol * corresponds to the position of the peak maximum detected by the watershed algorithm.

This would mean that the substance in question "appears" and "disappears" several times during the course of the elution. This is impossible in chromatography, where every sub-peak should be eluted in a continuous way.

2.2.2. Peak-detection methods based on merging one-dimensional peaks

Several methods have been described in the literature that detect peaks in two-dimensional chromatography using one-dimensional peak-detection methods. The general idea of these methods follows a two-step procedure. First, peaks are detected in a one-dimensional form, using the raw signal arising from the detector. This would yield, when applied to the signal in Fig. 1a, to the detection of seven peaks. This step makes use of the classical peak-detection algorithms from one-dimensional chromatography [1], and avoids the drawback of discontinuity of sub-peaks that the watershed algorithm has (explained in previous section, and Fig. 2). In a second step, a collection of criteria is applied to decide whether a collection of one-dimensional peaks belongs to the same (modulated) peak and should be "merged" in a single two-dimensional peak or not. The criteria applied may vary with the different versions, but all are based on peak profile similarities. In short, when the seven peaks detected in the first step (Fig. 1a) are similar (e.g., because they elute at the same second-dimension retention time), the integration algorithm will decide to merge them. In a previous work, we published an algorithm that measures the overlap of peak regions of adjacent peaks to decide whether peaks have to be merged [10]. A simplified versions of the algorithm, published later [13], was used in this article. The algorithm takes into account the differences between second-dimension retention times as the criterion to merge one-dimensional peak and uses the unimodality condition.

2.2.3. Appearance of saddle points in two-dimensional chromatograms

In two-dimensional gas chromatography, the watershed algorithm detects a two-dimensional chromatographic peak when a group of one-dimenisonal peaks appears like a single peak (or "mountain"). This means that, if a saddle point is detected within a single peak, the watershed algorithm splits the peak in two (like in geographical studies, a saddle point splits a single catchment

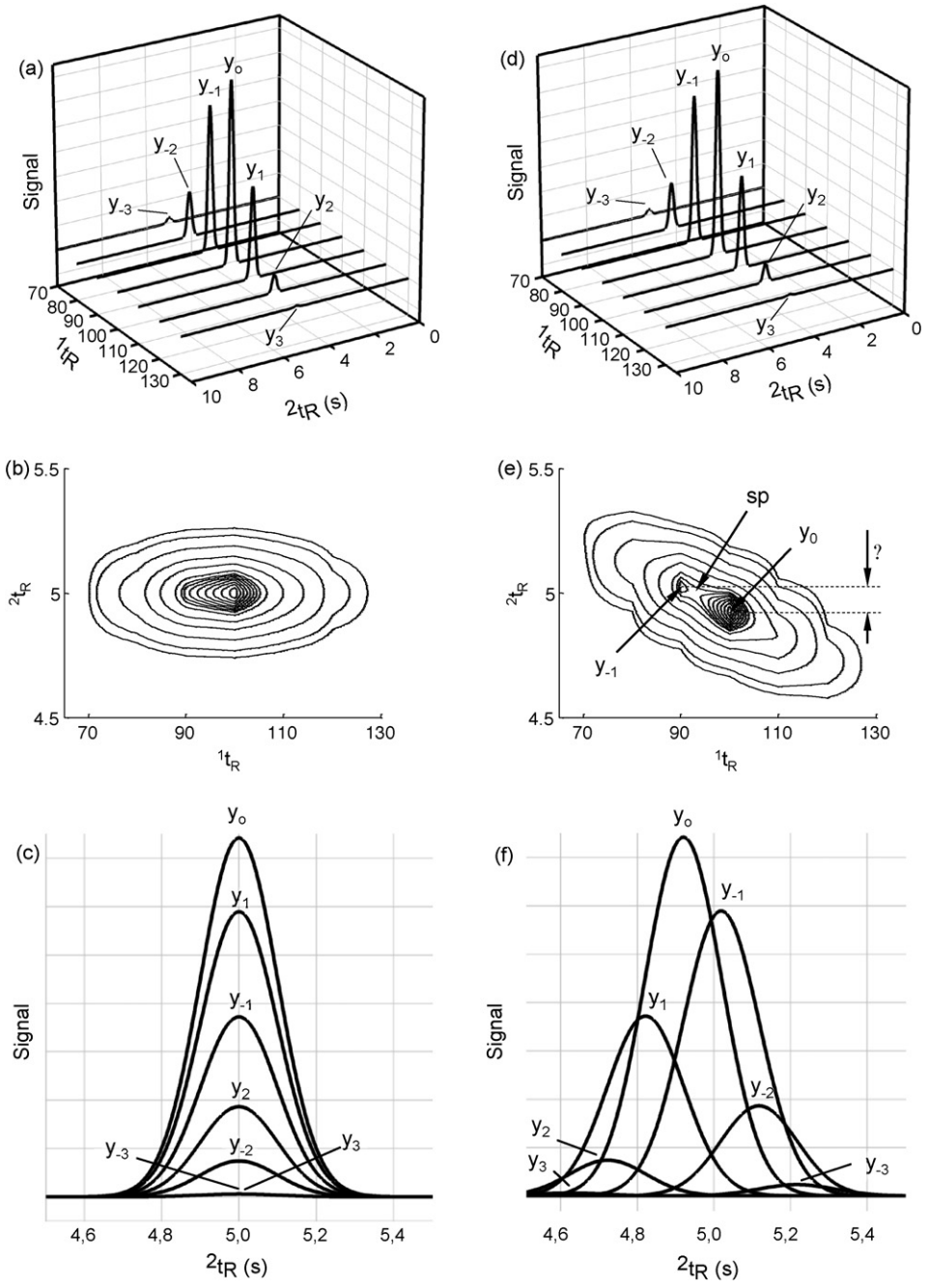

Fig. 3. Rearrangement of the signal corresponding to Fig. 1c in a matrix, represented in different ways: (a) with perspective; (b) bird-eye view; (c) from a point of view parallel to the ^1t_R axis. (a–c) Depict the situation without second-dimension retention time variability. In (d–f), a constant decrease of δ is introduced in the second-dimension retention times for subsequent peaks (Eq. (6)). The saddle point (sp) is overlaid in part e.

area in two). Without second-dimension retention time variability, the watershed definition and the chromatographic definition of a two-dimensional peak coincide. However, shifts in the second-dimension retention times can make a saddle point to appear, resulting in the segmentation of a single two-dimensional peak. To illustrate this effect, let us consider first folding the raw signal depicted in Fig. 1c to form a two-dimensional data table that collects at each column the signal corresponding to one modulation time (see Ref. [10] for details). A graphical, three-dimensional view of this table is depicted in Fig. 3a. Fig. 3b represents a contour plot (a bird-eye view) of this situation. Fig. 3c shows the same 3D plot rotated in such a way that the point of view is parallel to the ^1t_R axis. As can be seen, in Fig. 3c the variability in second-dimension retention times is absent, hence the retention times of the peak maxima of the sub-peaks are identical.

Let us suppose, however, that there is indeed a variation in the second-dimension retention times. This might be due to either (unavoidable) instrumental variability – inducing a random variation – or to the effect of a temperature gradient in the second-dimension column – inducing a constant drift. A combination of both effects is often the case in practice. Fig. 3d–f represents the same situation as in Fig. 3a–c, but a constant drift in second-dimension retention time is observed. This is artificially introduced by using $t_i = (i + \phi)m + i\delta$ as the retention time for each sub-peak,

instead of $t_i = (i+\phi)m$ (as it was defined in Eq. (3)):

$$y(t) = \frac{A}{{}^1\sigma^2\sigma 2\pi}\sum_{i=-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\frac{t-(i+\phi)m+i\delta}{{}^2\sigma}\right)^2\right] \times \int_{m(i+\phi-(1/2))}^{m(i+\phi+(1/2))}\exp\left[-\frac{1}{2}\left(\frac{t}{{}^1\sigma}\right)^2\right]dt \quad (6)$$

The δ parameter is the delay in the second-dimension retention time between two consecutive sub-peaks, see Fig. 3e. A negative value of δ (as it happens in Fig. 3e) means that the sub-peak is ahead with respect to its preceding sub-peak. One should note that we have restricted the plot to a constant drift for clarity, but all equations developed in this section are applicable to any kind of shift (random, constant or a combination of both). As can be seen (Fig. 3e), a saddle point appears, preventing the watershed algorithm to find the collection of (one-dimensional) sub-peaks that belong to the same two-dimensional peak. The saddle point appears because the maximum of one of the sub-peaks is above the profile of the previous sub-peak (in Fig. 3f the maximum of sub-peak y_1 is above sub-peak y_0 for certain regions). A saddle point may appear in situations with some second-dimension retention variability, which is in fact unavoidable from the experimental point of view. However, a small second-dimension retention time variability do not cause a saddle point to appear. The purpose of this section is to discover when the second-dimension retention time variability results in the appearance of a saddle point.

Fig. 4 represents the situation depicted in Fig. 3d with more detail. Only sub-peaks y_1 and y_0 are represented for clarity. As can be seen, there is a maximum difference (Δt_{crit}) for ${}^2t_{0,max}$ and ${}^2t_{1,max}$. Situations in which the difference between these two retention times is above this threshold will induce a saddle point in the three-dimensional surface, making the watershed algorithm to fail. Expressed more accurately, the watershed algorithm will fail when:

(i) the quantity $|{}^2t_{1,max} - {}^2t_{0,max}|$ is higher than Δt_{crit} for $-1/2 \le \phi \le 0$,
(ii) the quantity $|{}^2t_{1,max} - {}^2t_{0,max}|$ is higher than Δt_{crit} for $0 \le \phi \le 1/2$.

Examples of situation (i) and (ii) are depicted in Fig. 4a and b, respectively. One should note that we have only considered the sub-peaks corresponding to $i = 0$ and its neighbours ($i = -1$ and $i = 1$) and not the other sub-peaks. This is because these sub-peaks are always more similar in height (see Fig. 1) than other combinations of adjacent sub-peaks, and therefore they are the primary cause of saddle-point appearance.

2.2.4. Analytical expression of maximum second-dimension retention time variation between consecutive sub-peaks to avoid watershed-algorithm failure

In this section we deduce the analytical solution for Δt_{crit}, the maximum retention time variation between two consecutive sub-peaks belonging to the same compound that avoids a failure of the watershed algorithm. From Eq. (5) one can find an expression for the maximum height of the ith sub-peak, $y_{i,max}$, which is found at $t = (i+\phi)m$:

$$y_{i,max} = \frac{A}{{}^1\sigma^2\sigma 2\pi}\int_{m(i+\phi-(1/2))}^{m(i+\phi+(1/2))}\exp\left[-\frac{1}{2}\left(\frac{t}{{}^1\sigma}\right)^2\right]dt \quad (7)$$

For case (i), we want to find the value of t for peak y_0 at which the peak height equals $y_{1,max}$. Taking into account the expression

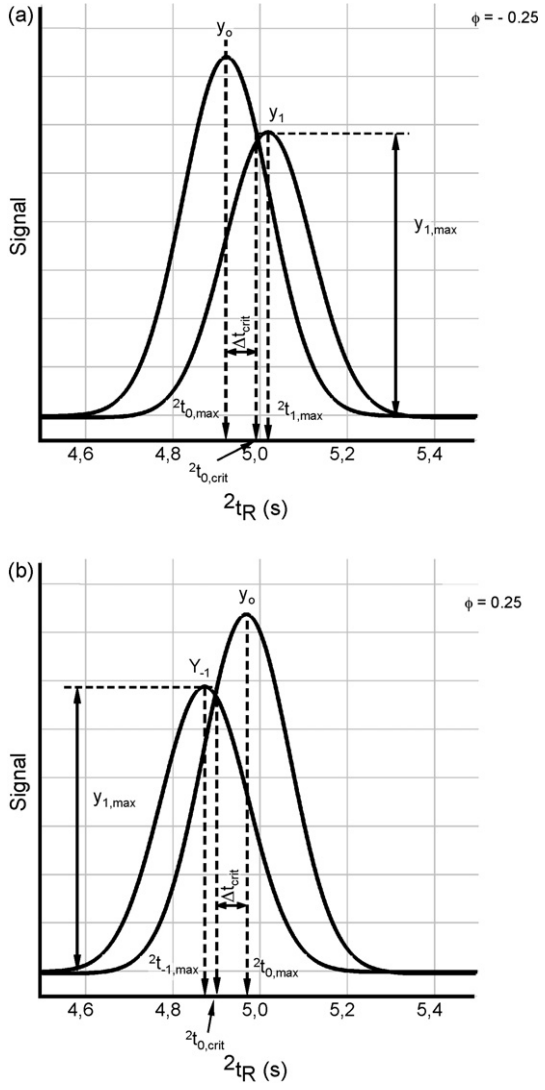

Fig. 4. (a) Representation of the same situation depicted in Fig. 3f. (b) Corresponds to the same situation as in (a), but with a value of ϕ of 0.25. For details about the rest of the parameters, see text (Section 2.2.4). For simplicity, only the two highest sub-peaks have been represented in each case.

of y_0 in Eq. (5) and equalling it to $y_{1,max}$ (Eq. (7)) yields:

$$y_0(t) = \frac{A}{{}^1\sigma^2\sigma 2\pi}\exp\left[-\frac{1}{2}\left(\frac{t-\phi m}{{}^2\sigma}\right)^2\right] \times \int_{m(\phi-(1/2))}^{m(\phi+(1/2))}\exp\left[-\frac{1}{2}\left(\frac{t}{{}^1\sigma}\right)^2\right]dt = y_{1,max} = \frac{A}{{}^1\sigma^2\sigma 2\pi}\int_{m((1/2)+\phi)}^{m((3/2)+\phi)}\exp\left[-\frac{1}{2}\left(\frac{t}{{}^1\sigma}\right)^2\right]dt \quad (8)$$

Solving the previous equation for t yields an expression for $t_{0,crit}$:

$$t_{0,crit} = \phi m + {}^2\sigma\sqrt{-2\ln\frac{\int_{m((1/2+\phi))}^{m((3/2)+\phi)}\exp\left[-(1/2)(t/{}^1\sigma)^2\right]dt}{\int_{m(\phi-(1/2))}^{m(\phi+(1/2))}\exp\left[-(1/2)(t/{}^1\sigma)^2\right]dt}} \quad (9)$$

Note that in Fig. 4a the time quantities have been related to the second-dimension retention time. However, as within the modulation period second-dimension retention times represent a shifted axis of the absolute retention time, t, Δt_{crit} can be expressed as

$\Delta t_{crit} = t_{0,crit} - t_{0,max}$ instead of $\Delta t_{crit} = {}^2t_{0,crit} - {}^2t_{0,max}$ (as can be seen in the Fig. 4a). As $t_{0,max} = \phi m$, then:

$$\Delta t_{crit} = {}^2\sigma \sqrt{-2 \ln \frac{\int_{m(\phi+(1/2))}^{m(\phi+(3/2))} \exp\left[-(1/2)(t/{}^1\sigma)^2\right] dt}{\int_{m(\phi-(1/2))}^{m(\phi+(1/2))} \exp\left[-(1/2)(t/{}^1\sigma)^2\right] dt}} \quad (10)$$

For case (ii) ($0 \le \phi \le 1/2$), we want to find the value of t for peak y_0 at which the peak height equals $y_{-1,max}$. This yields the following expression:

$$\Delta t_{crit} = {}^2\sigma \sqrt{-2 \ln \frac{\int_{m(\phi-(1/2))}^{m(\phi-(3/2))} \exp\left[-(1/2)(t/{}^1\sigma)^2\right] dt}{\int_{m(\phi-(1/2))}^{m(\phi+(1/2))} \exp\left[-(1/2)(t/{}^1\sigma)^2\right] dt}} \quad (11)$$

In this case we have considered $\Delta t_{crit} = t_{0,max} - t_{0,crit}$.

It is convenient to express Eqs. (10) and (11) as a function of the error function, erf. Making use of the property of $\mathrm{erf}(-b) = -\mathrm{erf}(b)$, Eqs. (10) and (11) yield:

$$\Delta t_{crit} = {}^2\sigma \sqrt{-2 \ln \frac{\mathrm{erf}(a(|\phi| - (1/2))) - \mathrm{erf}(a(|\phi| - (3/2)))}{\mathrm{erf}(a(|\phi| + (1/2))) - \mathrm{erf}(a(|\phi| - (1/2)))}} \quad (12)$$

where $a = m/({}^1\sigma\sqrt{2})$. The validity of Eq. (12) was tested (see Appendix A.1). Eq. (12) can be approximated (see Appendixes A.2 and A.3) to the more following expression:

$$\Delta t_{crit} = \frac{{}^2\sigma}{{}^1\sigma} m \sqrt{1 - 2|\phi|} \left(\alpha + \beta \frac{m}{{}^1\sigma}\right) \quad (13)$$

were $\alpha = 1.037$ and $\beta = -0.094$ are empirical parameters.

2.2.5. Probability of failure of the watershed algorithm

Eq. (13) can be used to calculate the probability of failure of the watershed algorithm for a given chromatogram. To this aim, the value of ϕ from the previous equation is taken apart (this operation would not be possible if the equation depends on the error function, as happens with Eq. (12)). This yields an expression for ϕ_{crit}:

$$\phi_{crit} = \frac{1}{2}\left(1 - \left(\frac{\Delta t_{crit}}{\alpha + \beta(m/{}^1\sigma)} \frac{{}^1\sigma}{{}^2\sigma m}\right)^2\right) \quad (14)$$

As larger absolute values of ϕ mean smaller differences between y_0 and $y_{\pm1}$ (see Fig. 1) one can state that the watershed algorithm fails for all situations in which $|\phi| \ge \phi_{crit}$. Next we assume that, in two-dimensional chromatography, the collection of two-dimensional peaks will show a flat distribution of ϕ values. In other words, there is no preference for a given ϕ value in a complex chromatogram. This was checked in the experimental data in Section 4.3 (results not shown). Therefore, we can easily calculate the probability of failure of the watershed algorithm at a given ϕ_{crit} value:

$$P_{fail} = \frac{(1/2) - \phi_{crit}}{(1/2)} \quad (15)$$

where the numerator of the equation is the number of failure cases, and the denominator the number of possible cases. Substituting Eq. (14) in Eq. (15) yields:

$$P_{fail} = \left(\frac{\Delta t_{crit}}{\alpha + \beta(m/{}^1\sigma)} \frac{{}^1\sigma}{{}^2\sigma m}\right)^2 = (\Delta t_{crit})^2 \left(\frac{1}{\alpha + \beta q} p\right)^2 \quad (16)$$

where we have defined $p = {}^1\sigma/({}^2\sigma m)$ and $q = m/{}^1\sigma$.

3. Experimental

3.1. Reagents, samples and apparatus

The diesel sample was obtained from a local gas station. The sample was analysed on an Agilent 6890N gas chromatograph equipped with a LECO quad-jet GC × GC modulator (LECO Corp., St. Joseph, MI, USA), a flame ionization detector (FID), electronic pressure controller and separate first- and second-dimension ovens. The first-dimension separation column was a 10-m long, 180-μm inner diameter fused silica capillary column coated with a 0.2 μm RTX-5 stationary phase (Restek, Bellefonte, PA, USA). The second-dimension column was a 1.1 m, 100 μm internal diameter column with a 0.1 μm DB-17 stationary phase (J&W, Folson, CA, USA). The sample was diluted with n-heptane to a final ratio of 1:1 diesel/n-heptane. The injector temperature was set to 250 °C, the split ratio to 150:1. The initial oven temperature was set to 40 °C for the primary oven and to 60 °C for the secondary oven with an isothermal hold of 5 min. The system was operated at a constant inlet pressure of 195 kPa. The temperature program had a ramp of 1.5 °C/min from 40 °C to 260 °C for the primary oven and 60–280 °C for the secondary oven. Data collection for the FID was performed at 200 Hz. The modulation period was 8 s.

3.2. Software

The chromatogram was exported as *.cdf format from the LECO ChromaTOF 3.22 software (LECO Corp., St. Joseph, MI, USA), and imported into MATLAB 7.7 (The Mathworks, Natick, MA, USA). Home-built routines were written in MATLAB for further data treatment, including chromatogram folding and two-dimensional peak detection (see Section 3.3).

3.3. Measuring 2D peak features

In Section 4.2, a GC × GC chromatogram of a diesel sample was studied. The peak-detection algorithm described in Section 2.2.2 was applied to the data. Further computations were performed to the detected peaks to calculate other peak features: ${}^1t_R$, ${}^2t_R$, ${}^1\sigma$, ${}^2\sigma$ and $\Delta^2 t_R$. The way these peak features are calculated is explained below.

The first-dimension retention time, ${}^1t_R$, is calculated using the definition of the first moment of a distribution, applied to the collection of sub-peaks:

$${}^1t_R = \frac{\sum_{j=1}^{np} a_j {}^1t_{R,j}}{\sum_{j=1}^{np} a_j} \quad (17)$$

where np is the number of sub-peaks that have been merged in a two-dimensional peak, a_j is the peak area of the jth sub-peak, and ${}^1t_{R,j}$ is the first-dimension retention time of the jth sub-peak, calculated using the first moment as explained in [1].

The second-dimension peak retention, ${}^2t_R$, is computed in a similar way:

$${}^2t_R = \frac{\sum_{j=1}^{np} a_j {}^2t_{R,j}}{\sum_{j=1}^{np} a_j} \quad (18)$$

Standard deviation in the first dimension is calculated as follows:

$${}^1\sigma = \sqrt{\frac{\sum_{j=1}^{np} a_j ({}^1t_{R,j} - {}^1t_R)^2}{\sum_{j=1}^{np} a_j}} \quad (19)$$

Finally, the standard deviation in the second dimension was calculated by pooling the variances of all the sub-peaks in a single

value:

$$^2\sigma = \sqrt{\frac{\sum_{j=1}^{np} a_j (^2\sigma_j)^2}{\sum_{j=1}^{np} a_j}} \quad (20)$$

where $^2\sigma_j$ corresponds to the standard deviation in the second dimension of the jth sub-peak.

4. Results and discussion

4.1. Probabilities of failure of the watershed algorithm vs. repeatability of second-dimension retention times

Fig. 5 illustrates in a single plot the probabilities of failure of the watershed algorithm (P_{fail}) against the repeatability of the second-dimension retention times of consecutive sub-peaks (Δt_{crit}) using Eq. (16). In two-dimensional gas chromatography, typical values of $^1\sigma/^2\sigma$ are around 100, with modulation values around $m = 10$ s. This yields to a value of p of 10 s^{-1} (i.e., long-dash curves in Fig. 5). The $m/^1\sigma$ ratio is inversely proportional to the number of cuts per peak. Assuming the peak is $4^1\sigma$ broad in the first dimension, $q = 0.5$ means eight cuts per peak, $q = 1$ means four cuts per peak, $q = 2$ means two cuts per peak and $q = 1$ means an average of one cut per peak (which rarely occurs in practice, since with this conditions there is still a great chance for the peak to appear in more than one modulation cycle). A typical value of $q = 1$ (four cuts per peak) and $p = 10\ s^{-1}$ brings us to the curve labelled with "a" in Fig. 5. This curve implies that there is around 20% of probability of failure of the watershed algorithm with chromatograms showing a variability of 0.05 s in the second-dimension retention time between consecutive sub-peaks. Clearly, a probability of failure of 20% cannot be neglected, specially if one considers that a variability of 0.05 s in ^2t_R is not rare in two-dimensional chromatography. Even more disappointing is the 80% probability of failure that could be expected with the same values of p and q, but with chromatograms showing a variability of 0.1 s in the second dimension between consecutive peaks (which is not rare with sharp temperature gradients in the second dimension).

It is instructive to inspect the effect on the probability of failure when $\Delta^2 t_R$, $^1\sigma$, $^2\sigma$ or m change, according to Eq. (16). Moreover, this can help the user to decrease the failure of the watershed algorithm, moving to regions of Fig. 5 that yield less probability of failure. The probability of failure decreases quadratically when the second-dimension retention time repeatability decreases. Hence, special care should be taken with $\Delta^2 t_R$. For example, a modulator yielding half the instrumental variability in second-dimension retention times would decrease the probability of failure of the watershed algorithm by a factor of four. Making the peaks sharper in the first dimension (decreasing $^1\sigma$) decreases the probability of failure, since p is decreased and q is increased. Another way to decrease the probability of failure could be via an increase of $^2\sigma$ (which decreases p). Finally, an increase of the modulation time (m) also decreases the probability of failure, since the value of p decreases and the value of q increases.

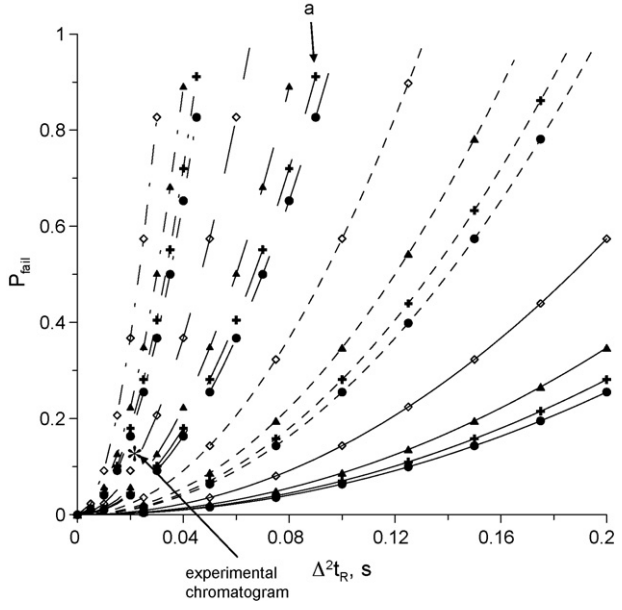

Fig. 5. Probabilities of failure of the watershed algorithm according to Eq. (16) vs. $\Delta^2 t_R$. Lines: solid, $p = 2.5$; short dash, $p = 5$; long dash, $p = 10$; dash–dot–dash, $p = 20$. Symbols: ✻ $q = 0.5$; ✚ $q = 1$; ▲ $q = 2$; ◇ $q = 4$.

4.2. Characterisation of peaks from diesel sample

In order to check empirically the probability of failure of the watershed algorithm, the real two-dimensional GC × GC chromatogram of a diesel sample was studied. The purpose of the study was to characterise a sample (considering the values of $^1\sigma$, $^2\sigma$), and getting an estimation of the experimental variability of retention times in the second dimension for the two-dimensional peaks. Then, the probabilities of failure of the watershed algorithm via Eq. (12) could be calculated, and compared with the estimation given in Eq. (16).

To this end, we used the peak-detection algorithm described in Section 2.2.2 as a reference method. The algorithm was applied to the sample, with a height-rejection threshold of 500 and a threshold of $\Delta^2 t_R$ of 0.1 s. The height-rejection threshold was set high enough to make sure that no noise was detected as a peak, but not too high to avoid rejecting the majority of peaks. The value of $\Delta^2 t_R$ was carefully selected by visual inspection of the chromatogram. This parameter may have an impact on the results, since wrongly merged peaks (being too far apart in second-dimension retention times) would bring us to the wrong conclusion that the variation between second-dimension retention times is too high. However, with a value of $\Delta^2 t_R$ of 0.10, this effect could be neglected, since the majority of the peaks showed an experimental variability below 0.10, and the number of two-dimensional peaks having high values of $\Delta^2 t_R$ (that could be wrongly detected) is not representative.

One should note that it is impossible to fully check the performance of the watershed algorithm with a real sample, since there is no method available able to decide correctly whether a collection of one-dimensional sub-peaks belong to the same two-dimensional peak. In other words, it is impossible to know *a priori* which of the cases yielding two peaks with the watershed are correctly split (because the situation truly represents two peaks) or they are wrongly split because of the appearance of the saddle point.

The application of the peak-detection algorithm described above detected more than 2100 two-dimensional peaks. The two-dimensional peaks showing a single sub-peak were first discarded, as $^1\sigma$ could not be calculated, yielding a subset of around 1100 peaks. For this subset of peaks, the distributions of values of $^1\sigma$, $^2\sigma$ and $\Delta^2 t_R$ were calculated and depicted in Fig. 6. As can be seen, the values of $^1\sigma$ are typically around 100 times larger than $^2\sigma$, which follows previous theoretical considerations [14]. The high number of peaks having a value of $^1\sigma$ between 2 and 4 s comes from the large population of peaks having only two sub-peaks.

The probabilities of failure of the watershed algorithm in this chromatogram were calculated as follows. First, the correlation between $^1\sigma$ and $^2\sigma$ was checked by visually inspecting a plot of $^1\sigma$ vs. $^2\sigma$. As the correlation was totally inexistent (results not shown), the joint distribution of $^1\sigma$ and $^2\sigma$ could be calculated just multiplying the probability associated for each combination of $^1\sigma$ and $^2\sigma$ intervals following the bar diagrams in Fig. 6. For example, the probability of finding a peak with $^1\sigma$ between 6 and 8 s and a $^2\sigma$

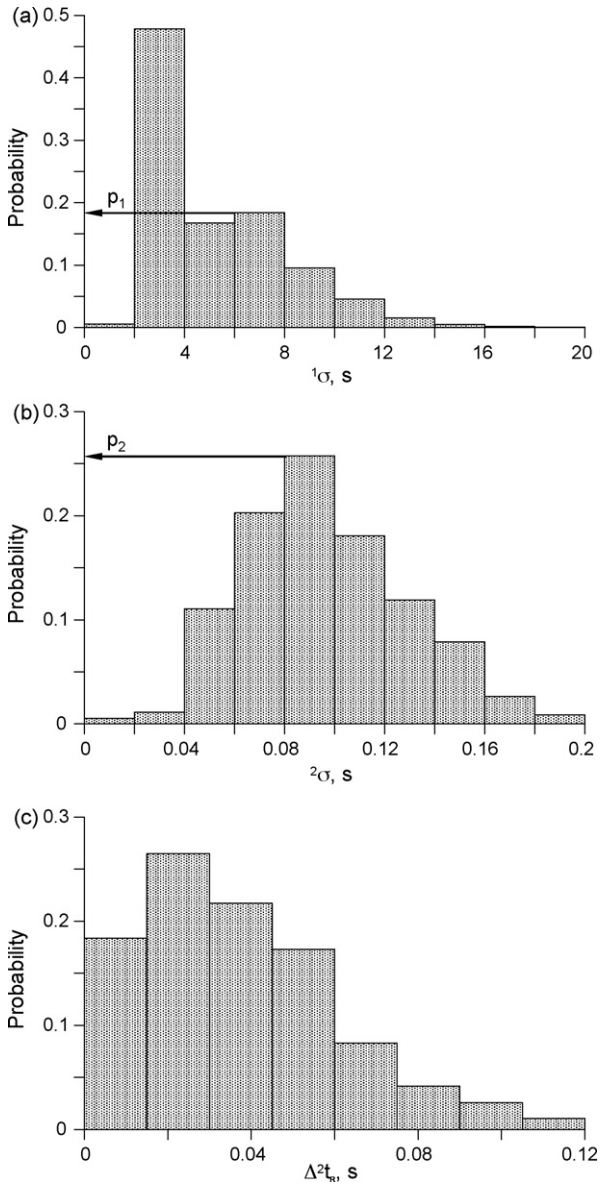

Fig. 6. Distribution of values of $\Delta^2 t_R$, $^1\sigma$, and $^2\sigma$ for the detected peaks of the diesel sample studied in this work (see Section 4.2). Peaks were detected following the algorithm explained in Section 2.2.2. Peaks that had only one sub-peak are not depicted, since $^1\sigma$ could not be calculated.

between 0.08 and 0.1 s was found multiplying the values of probability associated to both intervals in Fig. 6 (labelled with p_1 and p_2 in the figure). This was done for all $^1\sigma$ and $^2\sigma$ combinations. In a next step, for each of these combinations, a representative value of $^1\sigma$ and $^2\sigma$ was calculated taking into account the centre of each interval (e.g., in the example above, a value of $^1\sigma$ of 7 s and a value of $^2\sigma$ of 0.09 s was considered). In a next step, at each $^1\sigma$ and $^2\sigma$ combination, Eq. (12) was applied to calculate Δt_{crit} for a collection of ϕ values, with ϕ ranging from $-1/2$ to 1/2. For each of these ϕ values, the probability of having a $\Delta^2 t_R$ higher than Δt_{crit}, $p_{1\sigma,\,2\sigma,\phi}$, was calculated considering the experimental distribution of $\Delta^2 t_R$ in Fig. 6c. The probability of fail of the watershed algorithm at each $^1\sigma$ and $^2\sigma$ was calculated summing all $p_{1\sigma,\,2\sigma,\phi}$ over the whole range of ϕ values. Finally, the overall probability of failure was calculated associating the probability of fail at each $^1\sigma$ and $^2\sigma$ combination with the probability of having this combination (computed above). The sum of all these probabilities yields a value around 16%. This value is in accordance with Fig. 5. If one considers the mean values of $^1\sigma$ (around 5 s) and $^2\sigma$ (around 0.09 s) and the value of m (8 s), the values of p and q are around 7 s^{-1} and 1.6, respectively. This would produce a curve between the dash–dot–dash and the long-dash lines in Fig. 5. As the most probable variability between second-dimension retention times is around 0.02 s (see Fig. 6c), the position depicted with the * symbol (labelled in Fig. 5) represents the current experimental situation, with a probability of failure around 15% (which is in accordance with the detailed computation above). Fig. 5 allows us to consider how robust the watershed algorithm is against relatively small variations in $\Delta^2 t_R$. In situations with a steep curve, as it happens with the current p and q values, a small variability between second-dimension times yields to a considerable increase of the probabilities of failure. For example, a value of $\Delta^2 t_R$ of 0.04 s (which is still quite reasonable considering Fig. 6c), yields to the disappointing probability of failure of around 50%.

4.3. Comparison of peak-detection results obtained with watershed algorithm and a conventional two-dimensional peak-detection method

Studying the probabilities of failure of the watershed algorithm of a real chromatogram is not enough. An experimental test is needed to compare the calculated failure probability with the actual number of times that the watershed algorithm results in the wrong segmentation of a two-dimensional peak.

To simplify the notation, we will use W-algorithm to refer to the watershed algorithm and C-algorithm for the algorithm described in Section 2.2.2. The W-algorithm was applied to the same data used in Section 4.2 and compared with the results obtained with the C-algorithm. The two algorithms produced a different peak list, showing peak areas and peak regions that do not correspond completely. This is because of differences in the way the two algorithms calculate the two-dimensional peak boundaries. For example, the region corresponding to a single two-dimensional peak obtained with the C-algorithm could be covered by two (or more) peaks with the W-algorithm. Additionally, each of these two-dimensional peaks (detected with the W-algorithm) described a region that could be covered by more than one peak with the C-algorithm. To avoid these confusing situations, some conditions were applied to reduce the two original peak lists. Two conditions were applied. As for the first condition, only peaks that share 80% or more of the detected peak area and have a one-to-one correspondence between both peak-list algorithms were accepted in the pruned peak list. As for the second condition, only peaks that have a one-to-two or two-to-one correspondence in at least 80% of the area were accepted in the peak lists (i.e., 80% or more of the area of the peak is explained by two peaks in the other peak list or vice versa). Finally, as in Section 4.2, peaks containing only one sub-peak were eliminated from the list.

The study showed that 9% of the (pruned) peak list obtained with the C-algorithm were single two-dimensional peaks that corresponded with two peaks detected with the W-algorithm. On the other hand, 3% of the (pruned) peak list obtained with the W-algorithm represent single two-dimensional peaks that corresponded with two peaks with the C-algorithm.

Therefore, taking the C-algorithm as a reference, the W-algorithm fails splitting the peaks when they are supposed to be merged in 9% of the cases, and it merges the peaks when they are supposed to be split in 3% of the cases. The figure of 9% is in accordance with the computations performed in the previous section (16% of failure). The differences should be interpreted taking into account that the reference method is not a perfect method. The possibilities of the watershed algorithm yielding wrongly merged peaks have not been studied in this work, and hence the figure of 3% cannot be compared with any other theoretical computation.

As mentioned before, one should take into account that it is not possible to know *a priori* whether a collection of peaks corresponds

to a real two-dimensional peak. In this work, the C-algorithm has been taken as a reference. However, the interpretation of the results of this section should be taken with caution.

5. Conclusions

Peak-detection methods in comprehensive two-dimensional chromatography are not fully developed. Different families of methods exist: those based on the so-called watershed transform and those based on merging detected peaks that have previously been detected in the one-dimensional signal. None of the methods is perfect. An important drawback of the watershed algorithm is that it does not impose the condition of continuity for a peak, which implies that the signal of a substance may "appear" and "disappear" several times during the course of its elution. Second, it is somewhat intolerant to variability in the second-dimension retention time. It can split a true single peak in two when the retention times in the second dimension are not fully coincident. This is because the variability in the second dimension may produce the appearance of a saddle point in the two-dimensional chromatographic image.

Peak-model equations were constructed and an analytical model was solved to give the value of maximum second-dimension retention time variability ($\Delta^2 t_R$) that avoids the appearance of the saddle point (and, therefore, the failure of the watershed algorithm). The maximum variability depends on the phase of the two-dimensional peak ϕ, the modulation time m, and the peak width in the first and second dimension ($^1\sigma$, and $^2\sigma$). Approximations to the exact equation are possible, allowing to derive an analytical model that describes the probabilities of failure of the watershed algorithm as a function of $\Delta^2 t_R$, m, $^1\sigma$, and $^2\sigma$. The probability of failure was found to decrease with a decrease of $\Delta^2 t_R$, a decrease of $^1\sigma$, an increase of $^2\sigma$ and an increase of m.

The validity of the approach presented here was verified based on experimental data from the analysis of a diesel sample analysed with GC × GC. The example studied gave around 16% probability of error for the watershed algorithm. It was found that actually 9% of the peaks were split (compared to the conventional peak-detection algorithm). The latter figure, however, should be taken with care. A full experimental verification of this figure is impossible since there is no reference method to contrast the watershed-algorithm performance.

Probabilities of failure of the watershed algorithm are around 20% in normal situations in GC × GC. However, some care should be taken about robustness of each situation. In some chromatographic configurations, small variations in $\Delta^2 t_R$, m, $^1\sigma$, or $^2\sigma$ can lead to a sharp increment on the probabilities of failure of the watershed algorithm.

Acknowledgements

LECO Instrumente GmbH (Mönchengladbach, Germany) is gratefully acknowledged for the loan of the comprehensive GC × GC instrument. Adnan Kavka is acknowledged for doing the experimental work.

Appendix A.

A.1. Empirical confirmation of Eq. (12)

In order to test the performance of Eq. (12), an experimental design with computer-generated signals was carried out. Chromatograms containing a single two-dimensional peak were computed following Eq. (6). In a second step, the one-dimensional strip of data was "folded" into a two-dimensional matrix, each column representing the signal collected within a modulation cycle. The resulting data was submitted to peak detection with the watershed algorithm. The purpose of the experiment was to test the maximum variability in ^2t_R that the watershed algorithm could support without splitting a two-dimensional peak into two peaks. To introduce this variability, Eq. (6) was used instead of Eq. (4), in the same way as it was done to represent the situation in Fig. 3d–f.

Table 1
Experimental design to test the performance of Eq. (12).

Parameter	Min value	Max value	Step
$^1\sigma$	2	18	2
$^2\sigma$	0.02	0.18	0.02
ϕ	−0.5	0.5	0.05–0.02[a]
m	6	10	2

[a] Depends on the situations.

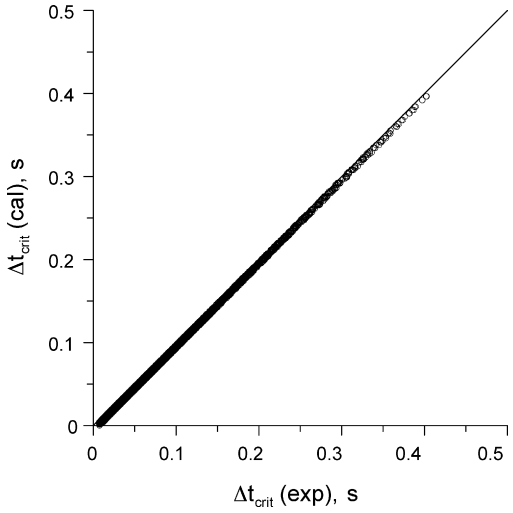

Fig. A1. Performance of Eq. (12). Experimental points are calculated as shown in Appendix A.1.

The experimental design covered different combinations of $^1\sigma$, $^2\sigma$, m, ϕ, as defined in Table 1, resulting in around 10,000 situations. For each situation (i.e., each combination of $^1\sigma$, $^2\sigma$, m, ϕ values), the maximum value of δ that could support the application of the watershed algorithm without resulting in a split peak was found. To this aim, a simple, non-linear search algorithm was programmed. The algorithm was automatically testing different values of δ, generating the two-dimensional chromatogram with Eq. (6), and decreasing the value of δ when the application of the watershed resulted in a split peak, and increasing it otherwise. With this method, the critical value of δ could be approximated with a precision up to 1e−7 s at a relatively low computational cost. Note that the critical value of δ is equivalent to the experimental value of Δt_{crit}.

At each situation, the predicted value of Δt_{crit} using Eq. (12) was also calculated. Predicted values were compared with experimental ones. Fig. A1a depicts this comparison. As can be seen, the performance of Eq. (12) is excellent for all situations, which confirms that the model is correct. Therefore, it is clear that the reason for failure of the watershed algorithm is indeed the appearance of the saddle point due to retention time variability.

A.2. Approximation of Eq. (12) using trapezoidal integration

Eq. (2) can be approximated by:

$$a_i = \frac{A}{{}^1\sigma\sqrt{2\pi}} \int_{t_i-(m/2)}^{t_i+(m/2)} \exp\left[-\frac{1}{2}\left(\frac{t}{{}^1\sigma}\right)^2\right] dt$$

$$\approx \frac{A}{{}^1\sigma\sqrt{2\pi}} \frac{m}{{}^1\sigma} \exp\left[-\frac{1}{2}\left(\frac{t_i}{{}^1\sigma}\right)^2\right] \quad \text{(A1)}$$

where we have integrated the Gaussian curve between a and b making use of the trapezoidal rule:

$$\int_b^a \exp\left[-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2\right] dt \approx \frac{a-b}{\sigma} \exp\left[-\frac{1}{2}\left(\frac{(a+b)/2}{\sigma}\right)^2\right] \quad \text{(A2)}$$

As the trapezoidal rule is used, the lower the $(a-b)/\sigma$ ratio, the more exact the approximation becomes. Therefore, the lower the $m/{}^1\sigma$ ratio, the more exact the approximation of Eq. (A1).

Considering this approximation, Eq. (4) yields:

$$y(t) \approx \frac{A}{{}^1\sigma^2\sigma 2\pi}\frac{m}{{}^1\sigma}\sum_{i=-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\left(\frac{t-(i+\phi)m}{{}^2\sigma}\right)^2+\left(\frac{(i+\phi)m}{{}^1\sigma}\right)^2\right)\right] \quad \text{(A3)}$$

The equivalent equation for the approximated Eq. (5) for a single sub-peak becomes:

$$y_i(t) \approx \frac{A}{{}^1\sigma^2\sigma 2\pi}\frac{m}{{}^1\sigma}\exp\left[-\frac{1}{2}\left(\left(\frac{t-(i+\phi)m}{{}^2\sigma}\right)^2+\left(\frac{(i+\phi)m}{{}^1\sigma}\right)^2\right)\right] \quad \text{(A4)}$$

With the use of the approximated model, Eq. (12) can be strongly simplified. If Eq. (A3) is used instead of Eq. (4), $y_{i,\max}$ (Eq. (7)) simplifies to:

$$y_{i,\max} \approx \frac{A}{{}^1\sigma^2\sigma 2\pi}\frac{m}{{}^1\sigma}\exp\left[-\frac{1}{2}\left(\frac{(i+\phi)m}{{}^1\sigma}\right)^2\right] \quad \text{(A5)}$$

For case (i), Eq. (8) is simplified to:

$$y_0(t) \approx \frac{A}{{}^1\sigma^2\sigma 2\pi}\frac{m}{{}^1\sigma}\exp\left[-\frac{1}{2}\left(\left(\frac{t-\phi m}{{}^2\sigma}\right)^2+\left(\frac{\phi m}{{}^1\sigma}\right)^2\right)\right] = y_{1,\max}$$
$$\approx \frac{A}{{}^1\sigma^2\sigma 2\pi}\frac{m}{{}^1\sigma}\exp\left[-\frac{1}{2}\left(\frac{(1+\phi)m}{{}^1\sigma}\right)^2\right] \quad \text{(A6)}$$

which yields

$$\left(\frac{t-\phi m}{{}^2\sigma}\right)^2+\left(\frac{\phi m}{{}^1\sigma}\right)^2 \approx \left(\frac{(1+\phi)m}{{}^1\sigma}\right)^2 \quad \text{(A7)}$$

and therefore

$$\Delta t_{\text{crit}} \approx \frac{{}^2\sigma}{{}^1\sigma} m\sqrt{1+2\phi} \quad \text{(A8)}$$

For case (ii), a similar route can be followed yielding:

$$\Delta t_{\text{crit}} \approx \frac{{}^2\sigma}{{}^1\sigma} m\sqrt{1-2\phi} \quad \text{(A9)}$$

As the function is symmetric, Eqs. (A8) and (A9) can be rearranged to a single equation:

$$\Delta t_{\text{crit}} \approx \frac{{}^2\sigma}{{}^1\sigma} m\sqrt{1-2|\phi|} \quad \text{(A10)}$$

The performance of the approximated model (Eq. (A10)) was tested using the same experimental design described in Section A.1. As can be seen (Fig. A2), unfortunately a notable bias remains: some situations yield to Δt_{crit} values significantly lower than the predicted model. One can conclude that the simplified model is optimistic in calculating the maximum δ value supported by the watershed algorithm. Experimental values of Δt_{crit} are always below the predicted ones.

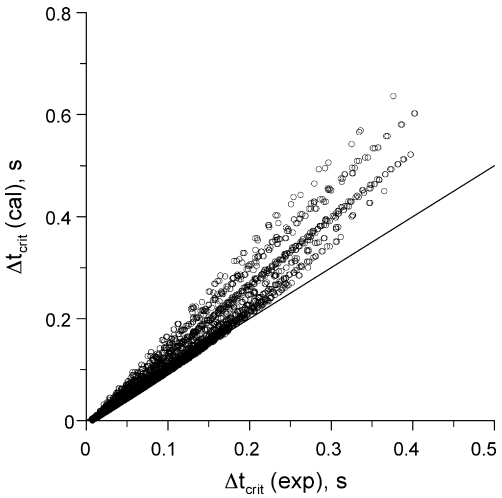

Fig. A2. Performance of Eq. (A10). Experimental points are calculated as shown in Appendix A.1.

A.3. Approximation of Eq. (12) using simulated data

The experimental values depicted in Fig. A2 were used to correct Eq. (A10). This equation can be re-written taking into account the deviation (ε) between the calculated and the experimental values:

$$\Delta t_{\text{crit}} = \frac{{}^2\sigma}{{}^1\sigma} m\sqrt{1-2|\phi|} + \varepsilon \quad \text{(A11)}$$

Deviations of the equation occur because of the approximation performed to calculate the integral in Eq. (A2). Fortunately, ε is approximately linear against $({}^2\sigma/{}^1\sigma)m\sqrt{1-2|\phi|}$ when ${}^1\sigma$ and m are constant. Fig. A3a illustrates this for two different pair values of ${}^1\sigma$ and m. The fact that the error depends on m and ${}^1\sigma$ is in accordance with Section A2, were it has been stated that the error introduced by computing the integral should depend on the $m/{}^1\sigma$ ratio. A straight line could be fitted to all the ε vs. $({}^2\sigma/{}^1\sigma)m\sqrt{1-2|\phi|}$ points sharing the same ${}^1\sigma$ and m values:

$$\varepsilon = a + b\left(\frac{{}^2\sigma}{{}^1\sigma} m\sqrt{1-2|\phi|}\right) \quad \text{(A12)}$$

This operation was performed for all ${}^1\sigma$ and m pair. The fitted values of a were, in all cases, non-significant. A list of different b values was obtained for each ${}^1\sigma$ and m pair. In a next step, the different values of b were represented against the $m/{}^1\sigma$ ratio (Fig. A3b). Fortunately again, the plot follows approximately a linear trend, and therefore a straight line could be fitted:

$$b = \alpha + \beta\left(\frac{{}^1\sigma}{m}\right) \quad \text{(A13)}$$

The experimental values of α and β were $\alpha = 1.037$ and $\beta = -0.094$. Rearranging Eqs. (A11)–(A13), and neglecting the value of a in Eq. (A12) yields:

$$\Delta t_{\text{crit}} = \frac{{}^2\sigma}{{}^1\sigma} m\sqrt{1-2|\phi|}\left(\alpha + \beta\frac{m}{{}^1\sigma}\right) \quad \text{(A14)}$$

The accuracy of this equation is checked in Fig. A4. As can be seen, the equation is approximately as accurate as Eq. (12), solving then the problems associated with Eq. (A10). Note that Eq. (A14) simplifies into (A10) when $\alpha = 1$ and $\beta = 0$.

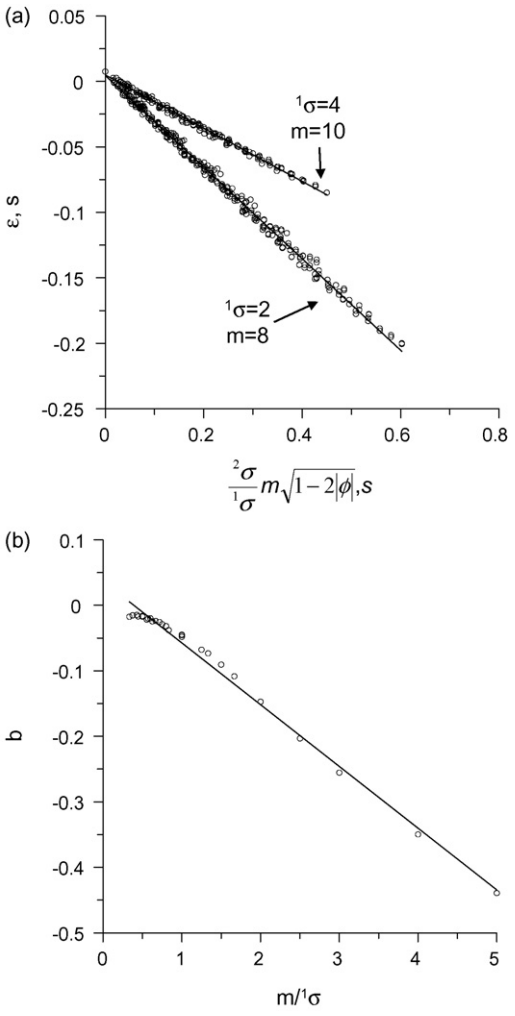

Fig. A3. (a) Representation of ε (Eq. (A11)) against $(^2\sigma/^1\sigma)m\sqrt{1-2\left|\phi\right|}$ for two cases with $^1\sigma$ and m constant (values of $^1\sigma$ and m are overlaid). (b) Linear fitting of the $m/^1\sigma$ against the b value (Eq. (A12)).

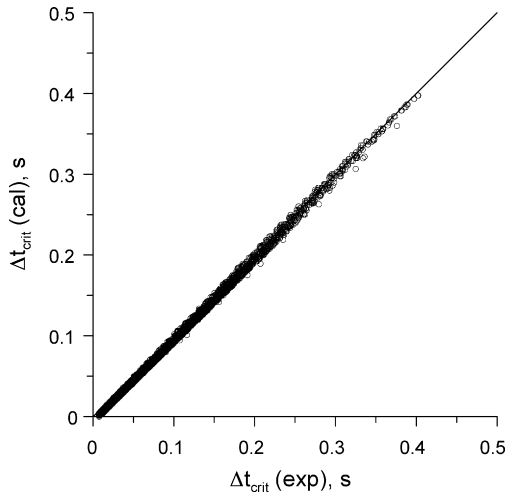

Fig. A4. Performance of Eq. (A14) with ν values of $\alpha = 1.037$ and $\beta = -0.094$. Experimental points are calculated as shown in Appendix A.1.

References

[1] A. Felinger, Data Analysis and Signal Processing in Chromatography, Elsevier, Amsterdam, 1998 (Chapter 8).
[2] J. Zhang, E. Gonzalez, T. Hestilow, W. Haskins, Y. Huang, Curr. Genomics 10 (2009) 388.
[3] R. Matthiesen, Proteomics 7 (2007) 2815.
[4] M. Dakna, Z. He, W.C. Yu, H. Mischak, W. Kolch, J. Chromatogr. B 877 (2009) 1250.
[5] I. Francois, K. Sandra, P. Sandra, Anal. Chim. Acta 641 (2009) 14.
[6] L. Mondello, M. Herrero, T. Kumm, P. Dugo, H. Cortes, G. Dugo, Anal. Chem. 80 (2008) 5418.
[7] O. Amador-Muñoz, P.J. Marriott, J. Chromatogr. A 1184 (2008) 323.
[8] S.E. Reichenbach, V. Kottapalli, M. Ni, A. Visvanathan, J. Chromatogr. A 1071 (2005) 263.
[9] J. Beens, H. Boelens, R. Tijssen, J. Blomberg, J. High Resolut. Chromatogr. 21 (1998) 47.
[10] S. Peters, G. Vivó-Truyols, P.J. Marriott, P.J. Schoenmakers, J. Chromatogr. A 1156 (2007) 14.
[11] F. Meyer, Signal Process. 38 (1994) 113.
[12] M. Daszykowski, I. Stanimirova, A. Bodzon-Kulakowska, J. Silberring, G. Lubec, B. Walczak, J. Chromatogr. A 1158 (2007) 306.
[13] 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.
[14] A. van der Horst, P.J. Schoenmakers, J. Chromatogr. A 1000 (2003) 693.