Chromatographic profiling: From samples to information
Peters, S.

Citation for published version (APA):
Peters, S. (2013). Chromatographic profiling: From samples to information

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 4

Development of an algorithm for peak detection in comprehensive two-dimensional chromatography

This chapter has been published as:

Summary
A method for peak detection in two-dimensional chromatography is presented. It is based on a traditional peak detection method, developed for one-dimensional chromatography. In a second step, a decision tree is applied to decide which one-dimensional peaks are originated from the same compound and have to be 'merged' into one two-dimensional peak. To this end, different features of the peaks (second-dimension peak regions and second-dimension retention times) are compared and different criteria (common peak regions, retention time differences, unimodality in the first dimension) are applied. The user controls this decision tree by establishing several options and “switches”. Recommendations for the set of options and switches are given. The algorithm was tested with GC×GC chromatograms obtained for a commercial air-freshener sample, detecting and merging the modulated peaks belonging to the same compound. A utility that calculates and sums peak areas from merged peaks is added to facilitate automated quantification. Although the algorithm was developed for GC×GC, its application to comprehensive two-dimensional liquid chromatography (LC×LC) data should at most require minor modifications.
4.1 Introduction

Comprehensive two-dimensional chromatography has proven to be a very powerful tool for the analysis of complex samples. Two columns, which separate the sample according to different properties of the analytes, are connected in series. In two-dimensional gas chromatography, the interface between the two columns is most commonly a (cryotrapping) modulator that traps the effluent of the first column during a given modulation period, focuses it, and then injects it as a narrow band into the second column. During the elution of a compound (peak) from the first dimension, several short-time analyses are performed in the second dimension. This requires the second-dimension analysis to be around two orders of magnitude faster than the first-dimension analysis [1]. To achieve this requirement in practice, the length and characteristic diameter (column diameter in case of open-tubular columns and particle diameter in case of packed columns) of the second-dimension column are usually smaller than the one of the first-dimension column.

In comprehensive two-dimensional gas chromatography (GC×GC), the column set usually comprises a non-polar and a polar column, which implies firstly a separation based predominantly on boiling-point and secondly a separation according to polarities. The term ‘orthogonal dimensions’ can be used if the separations (retention times) in both columns are fully independent from each other [2,3]. In truly orthogonal systems, a significant increase in peak capacity is obtained in comparison with traditional (one-dimensional) chromatography, since the individual peak capacities of the two columns can be multiplied [2,4,5]. In orthogonal systems co-eluting peaks in one dimension are statistically more likely to be separated in the other dimension. Due to the increase in peak capacity, more peaks can be detected in the two–dimensional space and more-complex samples can be analysed.
When using contemporary, high-resolution analytical instruments, the translation of data into useful information demands the use of computer-assisted data processing. This demand does not only exist in the chromatographic field, but also in other disciplines of analytical chemistry [6]. When progressing from one-dimensional to two-dimensional chromatography, the quantity and complexity of the data are invariably increased, making computerised data processing even more necessary. Unfortunately, methods for data handling in comprehensive two-dimensional chromatography are still at the basic stages of their development [7]. Ideally, such methods should include [8] (i) visualisation of multidimensional data (conversion and plotting of the matrix), (ii) data pre-processing (removing artefacts and baseline correction), (iii) peak detection, and (iv) peak quantification. There are few software packages commercially available that perform all these tasks [9]. Most packages involve only step (i) and to the authors’ knowledge, only Twee-D [7] and GC Image [8] have been designed to perform all four steps.

The focus of this paper is on the algorithms for peak detection, i.e. step (iii) above. This step implicitly requires a definition of what constitutes a two-dimensional chromatographic peak, as well as a robust algorithm able to demarcate it in the two-dimensional chromatogram. The only method found in the literature to perform this task was the watershed algorithm, adapted for peak detection by Reichenbach and co-workers [8].

In this paper, a novel approach for peak detection in comprehensive two-dimensional chromatography is presented. The method is based on conventional peak-detection algorithms for one-dimensional chromatography. This is logical because comprehensive chromatography involves sequential fast second-dimension analyses, each of which resembles a single chromatographic run. The one-dimensional approach is then adapted for use with comprehensive two-dimensional chromatography. Although the algorithm was developed for GC×GC, its application to comprehensive two-dimensional liquid chromatography (LC×LC) data should at most require minor modifications.
4.2 Theory

4.2.1 Overview

A general flow chart of the peak detection algorithm is depicted in Figure 4.1. The algorithm includes several steps: (i) data folding, (ii) 1D peak detection, (iii) two-dimensional peak merging and (iv), quantification of the generated 2D peaks. In this thesis, 1D and 2D are the abbreviations of one-dimensional and two-dimensional respectively. The primary data generated by a GC×GC or LC×LC experiment are the same as in one-dimensional chromatography: a two-column matrix is obtained, with the first column representing the time and the second the corresponding signal. This is simply the time-response stream of the detector, which is acquired independently of the modulation operation of the two-dimensional separation system, Since the modulation period and the sampling frequency are known, the raw data matrix can be converted and re-grouped to form a two-dimensional matrix (step i above). The column indices of the matrix represent the direction of the first dimension (1D) and the row indices represent the signal in each independent second-dimension (2D) analysis. The matrix resulting from this “data folding” can be plotted to yield a typical contour diagram (or colour diagram).

Figure 4.1 Flow chart of the algorithm developed.
One should note that the time axis and the matrix can be created directly if the number of experimental data points per unit modulation time is constant. This is possible if the inverse of the sampling frequency is an exact multiple of the modulation period. Otherwise, interpolation is necessary.

In this chapter, each signal set that corresponds to a single injection in the second dimension column is considered as equivalent to a one-dimensional chromatogram. As stated above, this is a single column of the data matrix. In each such chromatogram the individual peaks are detected (step ii). As usually, several individual peaks in consecutive one-dimensional chromatograms should be due to the same compound (ideally at least four) a “peak-merging algorithm” is applied. Basically, in this step the algorithm decides which one-dimensional peaks belong to the same compound (step iii). All steps will be explained in detail in the following sections.

The 2D peaks obtained after applying step i through iii can then be quantified in the fourth step, if required. Each 1D peak belonging to a 2D peak cluster is integrated independently and the areas and/or heights of all peaks within a cluster are summed automatically.

4.2.2 Peak detection in 1D chromatograms

Once the data have been arranged in a consistent two-dimensional matrix as explained in the previous section, a one-dimensional peak detection algorithm is applied to all 1D chromatograms (i.e. to each row of the data matrix). This is done by considering the information in each 1D chromatogram independently. Accordingly, a 1D peak is defined here in the same manner as in one-dimensional chromatography.

The algorithm for peak detection has been explained elsewhere and only a general overview is given here. For further details, see ref. [10]. The process is depicted in Figure 4.2. For each 1D chromatogram, the first- and the second-order derivatives are computed using the Savitzky-Golay algorithm [11]. Figure 4.2A shows the
analysis (original signal, first- and the second-order derivatives) performed on the 1D chromatogram at $t_R = 4.47$ min, between $t_R = 1.67$ and 2.3 s. To allow a comparison of the three plots on the same scale, the intensity of the signal is plotted in pA, the first derivative ($\text{pA.s}^{-1}$) is divided by a factor of 1,000 and the signal of the second derivative ($\text{pA.s}^{-2}$) is divided by a factor of 400,000.

**Figures 4.2** Measurements for 1D peak detection (A) and schematic "box-dot" representation in a two-dimensional chromatographic map (B). In Figure 4.2A, the original signal (black line), first- (red line) and second-derivatives (blue line) are shown. Characteristics of this one-dimensional peak are indicated: label (1) maximum intensity of the peak, label (2) second-dimension retention time, label (3) peak start and label (4) peak end. Figure 4.2 B shows the translation of the detected 1D peaks to the 2D chromatogram. The 2D peak’s starting point, end point and maximum are also indicated.
From the original signal and the first- and the second-order derivatives, several properties of the chromatographic peak can be defined, namely (1) the maximum peak height, (2) the peak retention time, (3) the peak starting-point and (4) the peak end-point (see Fig. 4.2A). For detecting the peak start and the peak end, two thresholds are defined by the user, \( i.e. \) Thr\(_0\) and Thr\(_1\). The first threshold, Thr\(_0\), acts as a limit for low-concentration compounds. Only peaks with a peak height exceeding Thr\(_0\) are considered. Note that Thr\(_0\) is set as a value above the baseline (\( e.g. \) 10 pA above baseline) and not as an absolute value. Thr\(_1\) is used for determining points (3) and (4), which are defined by inspecting the first derivative. The peak starting-point (3) is the first point in the first derivative that is above Thr\(_1\). The peak end-point (4) is the last point of the first derivative that is above \( -\)Thr\(_1\). One should note that Thr\(_1\) is especially required if tailing peaks are to be determined. If no significant tailing is expected, and the chromatogram could be properly baseline-corrected, Thr\(_1\) can be set to zero, and the peak start and the peak end is defined when the first derivative changes its sign. However, to the author’s experience, Thr\(_1\) should be set slightly above zero. In practice, the first derivative often does not reach zero (but only approaches it) and a low threshold value accounts for this observation.

The peak region of each individual peak in one second-dimensional chromatogram is defined using the peak start and the peak end. The bottom chromatogram in Figure 4.2 depicts a schematic representation of 1D peaks in the two-dimensional space. The peak regions are depicted as rectangles, and the peak maxima are indicated by a dot. For convenience and simplicity, this representation will be used in the remainder of this thesis.

4.2.3 Merging 1D peaks into 2D peaks

After detecting the 1D peaks present in each individual one-dimensional chromatogram, an algorithm that merges the 1D peaks into a single '2D cluster' is applied. Such a 2D cluster is defined in this chapter as a collection of 1D peaks in consecutive 1D chromatograms that have been considered to belong to the same 2D
peak, and therefore are merged or associated to form a single object. In this definition, the difference between a 2D cluster and a 2D peak is the capacity of the former to grow and thereby incorporate more 1D candidate peaks (in further steps of the algorithm). When a 2D cluster cannot be extended with more 1D peaks, the 2D peak is defined. The merging rules must be defined based on objective criteria. The user has control over some parameters that govern these criteria. Depending on the complexity of the sample and the nature of the analyte (e.g., the user may expect more or less tailing peaks), these parameters have to be modified in order to optimise the result. The peak-merging algorithm follows the scheme presented in Fig. 4.3. As can be seen, several criteria are applied consecutively. The algorithm begins by considering the first 1D peak in the first second-dimension chromatogram as the first 2D cluster (with only one element). In a second step, all 1D peaks found in the next second-dimension chromatogram are considered as candidates for merging. Two criteria are applied to discard (or accept) the merger of these 1D peaks with the first 2D cluster, namely the overlap criterion and the unimodality criterion. After all these steps, there can be zero, one or more peaks left that are acceptable candidates for merging. If more than one peak meets the overlap and unimodality criteria, then the candidate peak with the second-dimension retention time closest to the previous 1D peak (merged in the 2D cluster) is selected for merging. In case only one candidate peak is left, this peak is merged with the first peak. If no peak is found as a merger candidate for the 2D cluster, then this is considered as a complete 2D peak. If in contrast a merging candidate has been selected, the 2D cluster is enlarged and the algorithm will consider the peaks found in the following second-dimension chromatogram for more merger candidates. When a new 1D peak is not merged with a previous cluster, it is considered to be the start of a new 2D cluster.
The overlap criterion

This criterion examines the degree of overlap of the peak regions of two 1D peaks in consecutive second-dimension chromatograms. We define 1D peak A as the last 1D peak of the existing 2D cluster and 1D peak B as the candidate 1D peak to merge.
The criterion compares the peak region (see Section 4.2.2 for the definition of peak region) of 1D peak A and 1D peak B, as pictured schematically in Fig. 4.4.

![Diagram of peak regions](image)

**Figure 4.4** Schematic representation of peak regions of adjacent 1D peaks for different cases. The "box-dot" representation of Fig. 4.2 is used to depict peaks.

To simplify the visualisation, the peak regions are represented as rectangles with a dot as the peak’s maxima as defined in Figure 4.2 (Section 4.2.2). A ratio of overlap is calculated by dividing the length of the region where the two peaks overlap by the width of 1D peak A in the second dimension (length of the peak region). Depending on the adjacent peak regions considered, five different situations can be distinguished (Fig. 4.4):

a. Both peaks start and end at the same 2D retention times
b. Peak A starts later than peak B and it also ends later
c. Peak B starts later than peak A and it also ends later
d. Peak B starts later than peak A, but it ends earlier
e. Peak A starts later than peak B, but it ends earlier

Note that detected peak widths in the second dimension can also vary in adjacent chromatograms, based on the relative amounts of solute introduced into the 2D column (and not due to polarity changes). The ratio of overlap, OV, is calculated
according to \( OV = \frac{b}{a} \) with \( b \) being the peak region of the candidate peak (1D peak B) that is overlapped with the peak region of the 1D peak A and \( a \) being the peak region of peak A. OV can also be given as a percentage, \( \text{i.e. } OV \% = OV \times 100 \). A threshold, \( \text{Thr}_{OV} \), is then selected by the user. If OV is greater than \( \text{Thr}_{OV} \), the candidate 1D peak B is accepted to be subjected to the next criteria; if not, this candidate peak is rejected and the algorithm proceeds to the next candidate 1D peak.

Cases (d) and (e) are special cases as in each case one of the peaks is incorporated in the peak region of the other peak. In both cases, the candidate peak B is always accepted.

In GC×GC, the peak regions of adjacent 1D peaks should in principle obey the trend as pictured in Fig. 4.4B due to the increase in temperature between successive second-dimension chromatograms. In practice, however, the peak regions may show all variations described in Fig. 4.4. This is because a comprehensive GC system cannot be ideally controlled. The data treatment should be able to deal with acceptable variations. There are several operational reasons that can result in variations of the second dimension retention times:

- concentration-dependent retention (and peak shape) due to non-linear distribution isotherms
- non-instantaneous re-injection of the focussed peak into the second-dimension column
- non-consistent cooling of the cryotrap zone/ heating of the zone.
- inaccurate timing control of each modulated release
- non-uniform temperature programme or GC oven heating zones

**The unimodality criterion**

The 1D peaks that met the criterion explained in the previous section are then tested against the next criterion (unimodality). To apply this criterion, the signal obtained by plotting the intensities at the maxima of a 2D cluster (\( \text{i.e.} \), the "peak-
maxima profile") has to be considered. One should note that this peak-maxima profile represents the chromatographic peak profile in the first dimension. Therefore, it should only show one maximum. In other words, the peak should be unimodal. In Figure 4.5, the peak-maxima profiles of the two circled 2D clusters in Fig. 4.5A are depicted in Figs. 4.5B. In these figures, the pink dots shown in Fig. 4.5A are shown with the first-dimension retention time ($t_{R1}$) still on the horizontal axis, but the y-axis now represents the signal intensity. The blue dots represent interpolated points, the origin of which is explained below.

**Figure 4.5** GC×GC-FID chromatogram of air-freshener sample (A) and peak-maxima profiles of labelled clusters (B top and bottom). In B, purple dots represent the experimental peak-maxima profile, whereas interpolated points are represented as blue dots. The top figure in B corresponds to cluster 1 in A; the bottom figure corresponds to cluster 2.

Due to the limited number of modulation cycles per one-dimensional run, the number of data points available in the first dimension is limited (i.e. the data density is low, typically 4 or 5 points per first-dimension peak), which results in a poor representation of the "peak-maxima profile". Fortunately, the data contained in the second dimension can be used to supplement the information on the first-
dimension peak. In the second dimension the data acquisition rate is much (two orders of magnitude or more) higher than in the first dimension, so that more information is obtained in this direction. Part of this information is used to enrich the information in the peak-maxima profile.

We want to interpolate the signal between peak maxima labelled as “j” and “k” in Fig. 4.5A. First- and second-dimension retention times for those two peaks are $t_{R1} = 4.53$ and $t_{R1} = 4.60$ min, and $t_{R2} = 0.04$ and $t_{R2} = 0.039$ min (corresponding to 2.4 and 2.34 s), respectively. Taking into account the data-acquisition frequency at which the signal is collected, six points can be found between 0.039 and 0.04 min in the y-axis. However, as peaks “j” and “k” are found in consecutive second-dimension injections, no points are found in between 4.53 and 4.60 min (x-axis). Information of the extra points in the y-axis (obtained at both $t_{R1} = 4.53$ and $t_{R1} = 4.60$) is used to interpolate the signal in the x-axis. Suppose for example that we want to interpolate the signal at $t_{R1}=4.532$ min, which corresponds to $t_{R2}=0.0394$ min if one draws a line between “j” and “k”. Consider $h_j$ and $h_k$ the intensities at $t_{R1} = 4.53$ and $t_{R1} = 4.60$, respectively, both obtained at the same second-dimensional retention time ($t_{R2}=0.0394$). The interpolated peak height at $t_{R1} = 4.532$ min is calculated as follows:

$$h_{int} = \frac{h_j (1/d_j) + h_k (1/d_k)}{(1/d_j) + (1/d_k)} = \frac{h_j d_k + h_k d_j}{d_j + d_k} \quad \text{Eq. 1}$$

where $d_j$ and $d_k$ are the time differences (in x-axis units) of the interpolated point to points j and k, respectively (in this case this yields $d_j = 4.532-4.53$ and $d_k=4.60-4.532$). If this operation is repeated for all the points collected in the second-dimension axis (6 points in our example), an interpolated signal (blue points in Figs. 4.5B) is obtained.

The unimodality criterion works by monitoring the peak-maxima profile. With increasing first-dimension retention times, if a maximum has already been detected (labelled b Fig. 4.5B, top), only candidate peaks with decreasing intensities can be accepted. If in this situation the intensity of the maximum of the
candidate peak is greater, the previous 2D cluster is considered complete and a new cluster is started. It can be seen in the figure that, because of the unimodality criterion, the 2D peak is split at $t_R \approx 4.05$ min (between 1D peaks c and d). Whether 1D peak d is considered as the final peak of the first 2D cluster or as the first peak of the next 2D cluster depends on the difference in the second dimension retention times ($t_R^2$). This part will be explained below (Closest maximum). The comparison of results when either the peak-maxima profile or the interpolated peak-maxima profile is considered will be discussed in Section 4.4.

**Closest maximum**

After applying the overlap and unimodality criterion, there may still be more than one 1D peak which qualifies to be merged with an existing 2D cluster. This is because more than one peak has been detected at the same first dimension retention time. In that case, the second-dimension retention times ($t_R^2$) of the 1D peak candidates are compared. The candidate closest to the last 1D peak in the 2D cluster will be selected.

### 4.2.4 Direction of merging

The procedure and criteria described in Sections 4.2.3 can be applied in two ways: *i.e.* (i) from the left to the right in the chromatogram (in the direction of increasing first-dimension retention times) or (ii) from the right to the left in the chromatogram (decreasing first-dimension retention times). Case (ii) means a slight modification of the flow-chart depicted in Fig. 4.3: the "initial peak" is the last peak in the list of 1D peaks, and the peaks considered for candidate peaks in the next 1D chromatogram are found making $C_{next}=C_{ini}-1$. It is possible that the results obtained in case (i) differ from the results obtained in case (ii). As there is no compelling reason to follow either direction, we should take both results into account. Therefore, the algorithm presented in this work performs peak clustering in both directions - methods (i) and (ii). In a second step, the resulting peak
clusters are compared, and the best results from both methods (i) and (ii) are selected. This is done in the following way. Suppose that a certain clustering (configuration A) has been obtained using method (i, increasing $^1t_R$) and a different result (configuration B) has been obtained using method (ii, decreasing $^1t_R$). It can be demonstrated that the chromatogram can be divided into independent regions. A “region” in a chromatogram is defined as a group of 2D peaks in which a change from configuration A to configuration B does not require a change in the configuration of the 2D clusters found in other regions of the chromatogram. Therefore, either configuration A or B can be selected in each region independently, without affecting the remaining results of other regions.

The algorithm selects the best configuration (A or B) in each region. This “best configuration” for each region is selected in such a way that the differences in the second-dimension retention times for merged peaks within each 2D peak are minimal. This is calculated as follows: in the first step, the mean of the differences in second dimension retention times are calculated for each 2D peak within an independent region. In a second step, the mean of these mean differences is calculated. For this computation, only those 2D peaks that include more than one 1D peak are considered. The same value is calculated for both configurations: left to right and right to left. The configuration giving rise to the lowest retention time differences is considered to be the "best configuration" for this particular region.

The algorithm applies this criterion to all regions. Normally, the final result is a mixture of A and B configurations.

Fig. 4.6 depicts an example of the result obtained by mixing both configurations for a sample of air-freshener (see ‘4.3 Experimental’ for sample description and parameter settings).
Two enlarged zones are depicted in a separate figure. Figs. 4.7A and B depict the enlarged zone of region 1. Fig. 4.7A shows the merged peaks according to the right-to-left method (decreasing $t_R$) and 4.7B shows the merged peaks according to the left-to-right method (increasing $t_R$). Similarly, Figs. 4.7C and D depict the other enlarged zone (region 2) in Fig. 4.6 (Fig. 4.7C corresponds to the right-to-left method and Fig. 4.7D corresponds to the left-to-right method). In region 1, the differences in second-dimension retention times of the merged peaks were lower in the left-to-right method, so this configuration (Fig. 4.7B) was selected. In region 2, the right-to-left method (Fig. 4.7C) proved better.
Figure 4.7 Results of the right-to-left (A and C) and left-to-right (B and D) merging criteria when applied to region 1 (A and B) and region 2 (C and D) of the chromatogram depicted in Figure 4.6.
4.2.5 Quantification of the generated 2D peaks

The generated software also includes an integration tool, which allows the user to quantify the target analyte(s). As explained in the previous sections, each 2D peak contains several 1D peaks. These 1D peaks are integrated in a first step according to the trapezoidal method and then summed to yield the overall area of the 2D peak.

**Figure 4.8** (B) shows the intensity versus the second-dimension retention time at $t_R = 6.4$ min in Fig. A. Two methods of integration methods for overlapping peaks are shown in Fig. C and D.

The peak area of a 1D peak can be calculated in two ways, as depicted in Figs. 4.8C and D. Fig. 4.8B is the second-dimension chromatogram at $t_R = 6.4$ min of the chromatogram of the air-freshener sample shown in Figure 4.8A. As can be
seen, the two peaks are not well resolved. The area for peak 1 in Fig. 4.8B can be calculated by integrating the area under the curve from the peak start to the point where the two partly overlapping peaks are split (Fig. 4.8C). The alternative method is to subtract the area under the line connecting the peak start and the peak valley (named \(a\) in Fig. 4.8D) from the total area under the curve as calculated in the first method. Both methods can also be applied to peak 2.

For the purpose of further testing the methods, both ways of calculating the peak areas were applied to a standard mixture of \(n\)-alkanes, ranging from C14 to C22 and to an air-freshener sample, to which the alkane mixture was added to yield concentrations of 200 mg/L for each alkane. For all standards and samples, C16 was chosen as an internal standard. The results obtained for \(n\)-alkanes C19 to C22 were compared with the result obtained by ChemStation (Agilent) (see Section 4.4).

### 4.3 Experimental

#### 4.3.1 Chemicals

An air-freshener was bought locally and used in this study. The air freshener was diluted 1:100 in hexane and injected directly.

The alkane standard mixture (Sigma Aldrich) containing \(n\)-alkanes ranging from C16 to C22 was prepared by diluting a stock solution, yielding concentrations ranging from 50 mg/L to 1000 mg/L.

The air-freshener was spiked with the alkane mixture, yielding concentrations for each \(n\)-alkane of 200 mg/L (sample A) or 500 mg/L (sample B) and then diluted 1:100 in \(n\)-hexane.

#### 4.3.2 Instrumentation

The instrument used in this study was an Agilent Model 6890 gas chromatograph (Agilent Technologies, Burwood, Australia) with a flame-ionisation detection (FID) system (operated at 100 Hz data-acquisition frequency). The GC was
retrofitted with a longitudinally modulated cryogenic system (Everest Model, Chromatography Concepts, Doncaster, Australia). Both instrument and method were the same for the air-freshener sample and the analysis of the alkane standard mixture.

The column set for the air-freshener sample included a low-polarity BPX5 first-dimension column (30 m x 0.25 mm; 0.25 µm film thickness) and a more polar BPX50 second-dimension column (1 m x 0.1 mm; 0.1 µm film thickness). Both columns were from SGE International (Ringwood, Australia).

A modulation period of 4 s was applied for all analyses and the cryogenic trap was set to -10°C.

The GC system was operated under temperature-programmed conditions from 90 °C to 250°C (15 min) at 3 °C/min and hydrogen was used as the carrier gas at a flow rate of 1 mL/min. A split injection (20:1) was applied with an injection volume of 1 µL.

4.3.3 Software

The instrument was controlled by the ChemStation software Rev.A.09.01 [1206] (Agilent, Waldbronn, Germany) and data treatment was performed by home-built routines, written in MATLAB 6.5 (The Mathworks, Natick, MA, USA). The raw data obtained by the GCxGC system can be exported from the instrument software in comma-separated values format (*.csv) and then processed with the developed program.

4.4 Results and discussion

The performance of the algorithm was tested on chromatograms obtained for the air-freshener sample. The optimal parameter settings for the sample are described in Table 4.1. Setting the right parameters requires some chromatographic experience or knowledge about the sample or target analyte(s).
Concerning peak detection (Section 4.2.2), the algorithm proved able to detect the individual peaks. The number of detected peaks varied with Thr$_0$. The lower Thr$_0$, the more peaks were detected. For this sample, setting Thr$_0 = 10$ pA gave a reasonable result. If it was chosen lower, too much noise was detected as peaks. As in conventional, one-dimensional chromatography, setting the correct threshold can sometimes be problematic, especially if peaks with low intensities have to be detected. In these cases, a trial-and-error approach for optimising Thr$_0$ is recommended. The user has the opportunity to inspect the overlapped second dimension chromatograms. In these chromatograms, the signal intensities are plotted against the first dimension retention time. This type of plot simplifies setting the right Thr$_0$ for peak detection. An alternative can be to define Thr$_0$ as a function of the noise. This alternative can be useful if the user have no prior knowledge on the intensities of the peaks of interest. As this was not the case, Thr$_0$ and Thr$_1$ were selected manually by the user. Thr$_1$ was set at 2 pA.s$^{-1}$, which is a reasonable value when no strong peak tailing is expected. For our sample, setting Thr$_1$ to either 2 pA.s$^{-1}$ or 0.2 pA.s$^{-1}$ did not significantly change the result of the merging procedure. However, setting it equal to zero is not recommended, because the peak regions become too large and the merging process may be jeopardised.

Concerning the ratios of overlap, different Thr$_{OV}$ values (see Section 4.2.3) were tested for the air-freshener sample. It was found that a slight variation in Thr$_{OV}$ did not significantly change the merging results. For example, no changes were found when Thr$_{OV}$ was varied between 10% and 30%. Fig. 4.9 compares the results found with different Thr$_{OV}$ values. Differences were significant when Thr$_{OV}$ was set to 80% (B) instead of 20% (A); peaks were not correctly merged with Thr$_{OV}$

### Table 4.1 Optimal parameter for the air-freshener sample.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr$_0$</td>
<td>10</td>
</tr>
<tr>
<td>Thr$_1$</td>
<td>2</td>
</tr>
<tr>
<td>Ratio of overlap, Thr$_{OV}$</td>
<td>0.2</td>
</tr>
<tr>
<td>Unimodality option</td>
<td>1</td>
</tr>
</tbody>
</table>
set at 80%. In the authors' experience, a good recommendation is to use a ThrOV value of about 20%. Nevertheless, considering the experimental conditions, this 2D peak is quite broad and one should inject a standard to verify the merging of the correct 1D peaks. Obviously, the lower the extent of overlap required merging peaks, the more individual peaks are merged into 2D clusters.

Figure 4.9 Resulting GC×GC-FID chromatograms if different thresholds of overlap (ThrOV) are chosen: The left figure (A) corresponds to ThrOV = 20% and the right figure (B) to ThrOV = 80%.

The unimodality criterion (see Section 4.2.3) was tested and the results obtained from using both sub-options (either considering interpolated peak-maxima or not) were compared. Clearly, including the interpolated points will result in splitting more 2D clusters, as the chance is greater that an increase in the intensities of the peak-maxima profile will be found. Completely disregarding the unimodality option may be considered if the target peak is distorted; the use of the unimodality option might result in splitting a grossly deformed peak. The peak-maxima profile of a poorly shaped 2D cluster may contain some variation in the intensities of the peak-maxima profile; however, knowing that one 2D cluster is described requires the program to have the capability to not split this peak. Fig. 4.10 shows the different merging results for two groups of 2D clusters using a threshold of overlap of 20%. In Fig. 4.10A, no unimodality option was applied and
consequently all individual peaks in consecutive chromatograms were merged. The chromatographer, however, expects the two peaks of group 1 to be split. In Figs. 4.10B and C the unimodality option was applied, selecting either of the sub-options (disregarding the interpolated points – Fig. 4.10B – and including the interpolated points – Fig. 4.10C). If only the intensities of the maxima are considered, some information may be lost and candidate peaks can be incorrectly merged (false mergers of different compounds). On the other hand, interpolated points might not describe the real first-dimension chromatogram correctly and this may result in false splitting of peaks that are due to the same compound. As for the compounds in Fig. 4.10, it was found that considering the unimodality and disregarding interpolation gave the best results.

**Figure 4.10** Comparison of the results obtained using the peak merging algorithm with different options of unimodality. A, unimodality option disabled; B, unimodality option enabled, no interpolation; C, unimodality option enabled and interpolation used between data points.
In our experience, this is the case for most samples and we propose selecting the unimodality option without interpolation points as a first option. The merging results depicted in Fig. 4.10B and C can be explained by inspecting the peak-maxima profiles with and without including the interpolated points as illustrated in Fig. 4.5. Fig. 4.5B (top) shows the peak-maxima profile (purple dots) and the interpolated points between the maxima (blue dots) of the first group of the 2D cluster circled in Fig. 4.5A. For these two 2D clusters it does not matter whether the interpolated points are considered or not, as they do not add an extra minimum to the profile. Hence, the algorithm (as shown in Fig. 4.10B) splits the two 2D clusters due to the application of the unimodality criterion, independent of whether the interpolated points are included or not. For the second 2D cluster circled in Fig. 4.5A, however, different merging results are obtained when applying the unimodality option with and without including the interpolated points. This can be explained by inspecting the peak-maxima profile of that 2D cluster (Fig. 4.5B, bottom); the interpolated points clearly show a minimum between the peak maxima j and k and therefore, the 2D peak that is connected in Fig. 4.10B is split in Fig. 4.10C.

As discussed in Section 4.2.3, the second-dimension retention times of the 1D peaks merged to form this 2D peak may vary irregularly and, hence, each 2D peak may contain slightly varying second-dimension retention times. In practice, only one value for $^2t_R$ is given, usually that of the 2D peak maximum. If the $^2t_R$ values of a 2D peak are similar or equal to the modulation period, the variation of the second dimension retention times may cause the 2D peak to be split and appear at the top and at the bottom of the 2D chromatogram (i.e. in the contour plot). The algorithm will still detect and merge the 1D peaks correctly. However, to simplify visualisation, a shift of the chromatogram by, for example, half of the modulation period will re-combine the split peak and allow an easier assignment and possible quantification of the 2D peak.
When plotting the original data (string of chromatograms) versus time, the 1D peaks belonging to the same 2D cluster must occur at regular intervals equal to the modulation period and must show the same peak shape. An additional criterion to obtain the correct merging result may be to only allow a certain variation from this difference in retention time (e.g. one standard deviation of the peak width) or to test for the similarity of the peak shapes of adjacent (merger-candidate) peaks.

The different quantification methods as described in Section 4.2.5 were applied to the alkane standard mixture and to the two air-freshener samples spiked with \(n\)-alkanes (at a concentration of 200 mg/L and 500 mg/L), using C16 as internal standard (see Section 4.3 for further description of samples A and B). C21 was excluded in all calculations, since the quantification result differs significantly from the rest. The reason for that is not important at this stage as the aim of this study was only to introduce the quantification algorithm and compare it with the results obtained by the commonly used ChemStation software. ChemStation shows the unfolded chromatogram with the modulated 1D peaks. For quantification, the analyst first must decide which 1D peaks correspond to the same 2D peak and then (manually) integrate and sum those 1D peaks.

The quantification result using the first method (Fig. 4.8C) for both samples are given in Tables 4.2 and 4.3. This method was chosen since no significant overlap of the \(n\)-alkanes with other components in the sample was expected. The values were calculated using calibration standards ranging from 50 mg/L to 500 mg/L. The corresponding regression factors are also given.

Table 4.2 Quantification results for samples A and B obtained by the developed algorithm.

<table>
<thead>
<tr>
<th></th>
<th>Concentration sample A, inj 1</th>
<th>Concentration sample A, inj 2</th>
<th>Concentration sample B, inj 1</th>
<th>Concentration sample B, inj 2</th>
<th>Regression factor (R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C17</td>
<td>202.52</td>
<td>202.05</td>
<td>508.75</td>
<td>508.41</td>
<td>0.9998</td>
</tr>
<tr>
<td>C18</td>
<td>202.40</td>
<td>202.02</td>
<td>508.55</td>
<td>507.91</td>
<td>0.9998</td>
</tr>
<tr>
<td>C19</td>
<td>202.35</td>
<td>201.71</td>
<td>500.31</td>
<td>499.37</td>
<td>0.9986</td>
</tr>
<tr>
<td>C20</td>
<td>204.13</td>
<td>203.16</td>
<td>509.41</td>
<td>509.41</td>
<td>0.9995</td>
</tr>
<tr>
<td>C22</td>
<td>202.24</td>
<td>201.55</td>
<td>503.23</td>
<td>503.23</td>
<td>0.9993</td>
</tr>
</tbody>
</table>
Table 4.3 Quantification results for samples A and B obtained by ChemStation.

<table>
<thead>
<tr>
<th></th>
<th>Concentration sample A, inj 1</th>
<th>Concentration sample A, inj 2</th>
<th>Concentration sample B, inj 1</th>
<th>Concentration sample B, inj 2</th>
<th>Regression factor $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C17</td>
<td>212.47</td>
<td>212.05</td>
<td>523.72</td>
<td>523.34</td>
<td>0.9999</td>
</tr>
<tr>
<td>C18</td>
<td>203.61</td>
<td>203.44</td>
<td>511.00</td>
<td>508.45</td>
<td>0.9998</td>
</tr>
<tr>
<td>C19</td>
<td>202.08</td>
<td>201.67</td>
<td>504.73</td>
<td>504.25</td>
<td>0.9993</td>
</tr>
<tr>
<td>C20</td>
<td>203.74</td>
<td>203.13</td>
<td>508.07</td>
<td>506.73</td>
<td>0.9993</td>
</tr>
<tr>
<td>C22</td>
<td>201.99</td>
<td>200.99</td>
<td>502.57</td>
<td>501.10</td>
<td>0.9992</td>
</tr>
</tbody>
</table>

Table 4.4 Relative standard deviations (RSD) in percentage for sample A and B calculated for both algorithms applied (our algorithm and ChemStation).

<table>
<thead>
<tr>
<th></th>
<th>RSD-value [%], ChemStation software</th>
<th>RSD-value [%], developed algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample A</td>
<td>2.16</td>
<td>0.35</td>
</tr>
<tr>
<td>Sample B</td>
<td>1.66</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Table 4.2 shows the results obtained by quantifying the samples using the developed algorithm and Table 4.3 shows the results obtained by using the ChemStation software. Both methods give comparable results and have RSD values lower than 3% with the developed algorithm yielding slightly lower values (see Table 4.4). The relative standard deviations were calculated by computing first the average of both injections, and then calculating the standard deviation over all alkanes for each sample (A and B).

4.5 Conclusions

Algorithms to detect peaks and quantify peaks in comprehensive two-dimensional gas chromatography are of great importance for the successful proliferation and implementation of this family of techniques. In the present study, an algorithm based on conventional peak-detection routines for one-dimensional chromatography was adapted to detect and construct peaks in two-dimensional chromatography. In comprehensive two-dimensional chromatography, the modulation process divides the chromatographic peak in the first dimension into a collection of peaks in consecutive second-dimension chromatograms. The kernel
of the algorithm is to decide when the peaks detected in these fast one-dimensional chromatograms originate from the same compound, and therefore should be merged into a single object (a 2D peak). Obviously, problems may arise when closely-eluting peaks are to be assigned. This is especially troublesome when there is no valley (and only a shoulder) between the two co-eluting peaks in both separation dimensions. This is because the method presented here makes use of the first derivatives. In this case, the injection of standards might help and/or peak deconvolution procedures may be applied. The latter is especially (or particularly) recommended when a multichannel detector (e.g. MS) is present. However, as this study was performed using a flame-ionisation detection system, deconvolution techniques were not discussed.

The two thresholds used for 1D peak detection are manually selected. If chosen appropriately, they will result in the correct description of the 1D peaks. Broad ranges of the two criteria were found applicable in the present work, which implies that there is a great deal of tolerance in selecting appropriate values.

Several criteria have been developed for peak merging, viz. the overlap ratio and the unimodality criterion. The first criterion compares the peak regions in the second dimension that are shared for adjacent (consecutive) peaks. The higher the degree of overlap, the greater is the possibility that both peaks belong to the same compound. The second criterion is based on the fact that the peak-maxima profile in the first dimension should only show one maximum. Both criteria are controlled by (user-modified) options. The threshold of overlap (Thr_{OV}) and the unimodality criterion (that can be applied with different sub-options) have been demonstrated to be suitable criteria for correctly merging 1D peaks into a 2D cluster. For the sample studied here all compounds were clustered correctly. It could be shown that only little trial-and-error optimisation will usually lead to the correct threshold values and therefore the correct determination of the sample. Nevertheless, chromatographic experience, knowledge of the sample and target analyte(s) and good resolution will simplify this optimisation procedure.
The software also allows variations in the merging procedure, and it allows the user to manually merge and split peaks if required.
It could also be shown that the developed algorithm is able to quantify 2D peaks correctly, yielding RSD-values that are at least as good as those obtained with the commonly applied ChemStation software. In addition, the routine does not require any user input and is therefore much faster and easier to handle. This is especially the case if many target peaks are to be quantified.

References