A Novel Method for Peak Number Estimation in Chromatographic Peak Clusters

[a]J. HROUZEK, [b]J. KRUPČÍK*, [c]M. ČEPPAN, [a]Š. HATRÍK, and [d]P. A. LECLERCQ

[a]*HERMES PLUS, SK-826 03 Bratislava*

[b]*Department of Analytical Chemistry, Faculty of Chemical Technology, Slovak University of Technology, SK-812 37 Bratislava*

[c]*Department of Printing Art and Applied Photochemistry, Faculty of Chemical Technology, Slovak University of Technology, SK-812 37 Bratislava*

[d]*Department of Chemical Engineering, Eindhoven University of Technology, The Netherlands*

Received 25 October 1999

A procedure for the estimation of the number of peaks in a chromatographic peak cluster is proposed. This procedure is based on the dependence of the sum of squared errors (differences) between real and calculated chromatograms on the considered number of overlapped peaks. Starting with the number of apparent peaks, chromatograms are deconvoluted successively with increasing numbers of peaks possibly present. The sum of squared errors decreases with increasing numbers of peaks up to a point beyond which the function becomes constant. This point corresponds to the best estimation of the real number of peaks present in a cluster. The procedure was successfully used for the peak number estimation in model and real chromatograms.

In spite of considerable developments in chromatography, sufficient separation is often problematic. It is common experience that even high-resolution separations achieved in some systems are in many instances not sufficient to handle very complex mixtures, especially those of natural origin. The problem is rendered worse by the fact that the distribution of the retention of the sample components does not frequently follow regular patterns, but is generally more or less random [1, 2]. As a consequence, the analytes cause overlapped peak clusters in the chromatogram.

Overlapped peaks can be resolved by multidimensional systems combining two or more single separation systems, thus providing a much larger peak capacity. Another way is to include a computer algorithm into the system which allows mathematical resolution of overlapped peaks. In general, deconvolution procedures require sufficiently good estimations of initial parameters.

It is a problem to determine initial peak parameters for a deconvolution procedure for an overlapped peak cluster, particularly if the number of peaks present in the selected part of the chromatogram is not known. The minimum possible number of peaks in a chromatogram can relatively easy be found as the number of individual peaks. For each apparent peak two inflection points can be found in the chromatogram. Procedures for the estimation of the number of apparent peaks and corresponding peak characteristics are based on this fact. At first, maxima and adjacent inflection points can be used for the determination of retention times, peak heights, and peak widths [3]. Secondly, minima and remaining inflection points can be considered for peaks with no apparent maxima in the chromatogram.

The number of apparent peaks is usually lower than the total number of peaks present in the chromatogram. This is caused by strong peak overlapping, where the peaks differ only slightly in retention time. As a consequence no inflection points are observed for such peaks and for the estimation of the total number of peaks in the chromatogram another procedure should be used.

Peak widths from analytes eluting closely together do not differ substantially in chromatography both in an isocratic and/or a gradient mode. The dependence of the peak width on the retention time can be used for the prediction of the widths of overlapped peaks at isocratic elution. In the optimum gradient mode peak widths are almost equal all over the chromatogram. If

*The author to whom the correspondence should be addressed.

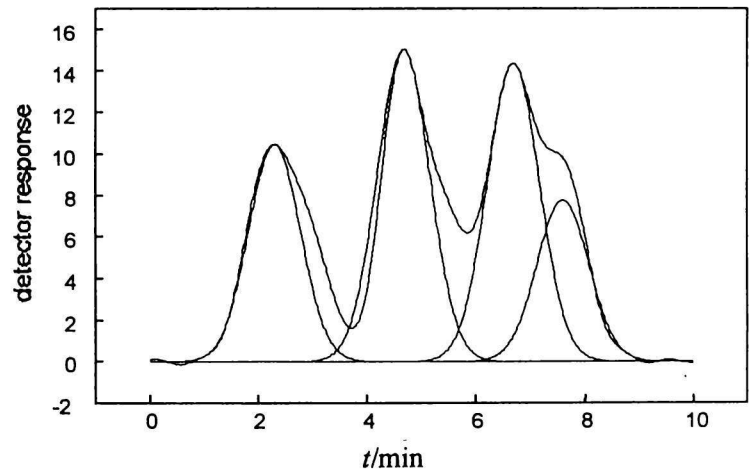

Fig. 1. Filtered and deconvoluted chromatogram after application of the apparent peak number estimation procedure.

the peaks are closely together, it can be assumed that the peak widths are not significantly different (both in isocratic as well as gradient mode) or even that they are equal. Apparent peaks with a width significantly wider than the predicted peak width can be considered as fused peaks. The number of peaks present can therefore be increased in this instance.

This knowledge can be used as a starting point for the deconvolution of a chromatogram or part thereof. The deconvolution procedure can then be repeated for increasing numbers of peaks which might be presumed to be present in the chromatogram or in the peak cluster. The number of peaks in a real chromatogram can be estimated from the dependence of the sum of squared errors between the experimental and deconvoluted chromatograms on the number of peaks considered.

The aim of this paper is to present a procedure for peak number estimation in a chromatographic peak cluster obtained at the gradient conditions and its verification in model and real chromatograms.

EXPERIMENTAL

Gas chromatographic experiments were performed with an HP 5890 Series II gas chromatograph (Hewlett —Packard), equipped with a split injector and electron capture detector. The chromatograms were recorded by an HP 3396 Series II integrator using an "area slice" width of 0.05 s and producing a digital signal.

The computer program written in Pascal runs on any IBM AT compatible computers.

RESULTS AND DISCUSSION

First, an appropriate model chromatogram was used in order to check the procedure for peak number estimation at gradient elution mode. The model chromatogram was composed of several randomly generated Gaussian peaks as shown in Fig. 1. Characteristics of the peaks are listed in Table 1. The peak widths are, however, slightly different due to expected differences in diffusion coefficients for molecules of various classes of compounds. Noise with normal distribution was added to the simulated chromatogram (Fig. 2).

Table 1. Retention Characteristics for Randomly Generated Peaks in the Model Chromatogram

Peak	Retention time / min	Peak height / units	Peak width at base/min
1	2.158	9.123	2.440
2	2.925	5.500	2.376
3	4.616	13.845	2.064
4	5.398	6.845	2.220
5	6.657	13.981	2.504
6	7.669	8.603	2.196

Prior to starting the procedure, Fourier filtering of raw signal data was used [4]. This filtering allows the removal of most of the higher frequency noise without distortion of the useful chromatographic signal. Next, the inflection points, maxima and minima were located in the smoothed chromatogram. An automated procedure was developed for apparent peak parameter estimation based on the following assumptions:

1. For anyone single peak two inflection points are present in the chromatogram.

2. The peak maximum lies between its inflection points.

3. The peak maximum lies close to a local maximum in the chromatogram investigated.

After chromatogram smoothing, the deconvolution procedure described in Ref. [5] was applied for the estimated apparent number of peaks (shown in Fig. 1).

The sum of squares $f(\boldsymbol{x})$ is minimized in the de-

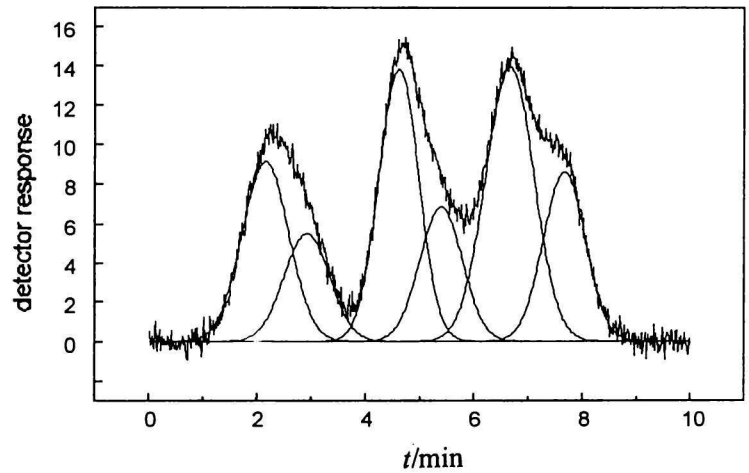

Fig. 2. Simulated chromatogram containing six randomly generated Gaussian peaks and corresponding peaks found by the peak number estimation procedure.

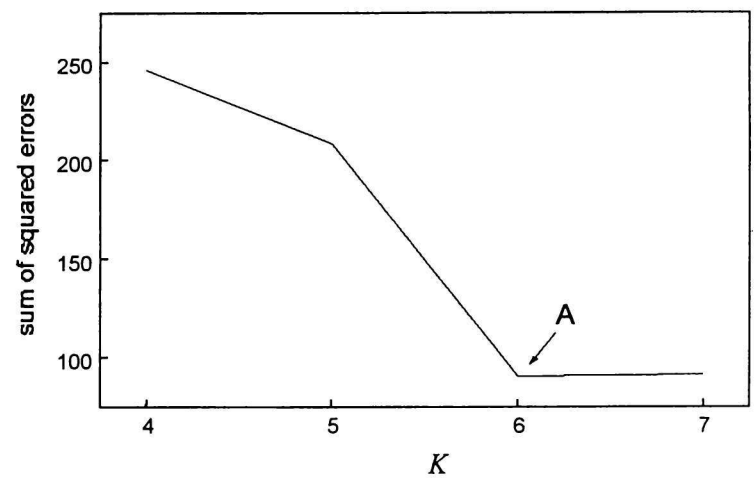

Fig. 3. The dependence of the sum of squared errors on the number of peaks (K). See the text for details.

convolution procedure as follows

$$f(\boldsymbol{x}) = \sum_{i=1}^{N}[Y(\boldsymbol{x}, t_i) + B(t_i) - Y_{\text{exp}}(\boldsymbol{x}, t_i)]^2$$
$$Y(\boldsymbol{x}, t_i) = \sum_{j=1}^{K} h_j(\boldsymbol{x}, t_i) \qquad (1)$$

Eqn (1) is written for a chromatogram consisting of N data points and K overlapping peaks where $Y_{\text{exp}}(t_i)$ is the experimental signal recorded at the time t_i, $Y(\boldsymbol{x}, t_i)$ is the sum of j individual peak signals $h(\boldsymbol{x}, t_i)$, and $B(t_i)$ is the baseline computed at time t_i. The $\boldsymbol{x}$ is a vector of parameters being optimized. As an approximation of the peak shape, different symmetrical or nonsymmetrical functions of h can be used [3—8]. A combination of those functions can be used as a best approximation of real chromatograms.

If $f(\boldsymbol{x})$ is applied to the whole chromatogram, details on local peak clusters are difficult to obtain. Therefore, $f(\boldsymbol{x})$ should be applied in a window of a limited number of data points (typically a few times over the local peak widths). This window can subsequently be shifted through the whole chromatogram.

As the model chromatogram was constructed for the gradient (linear-temperature programmed) mode, the average peak width obtained from the apparent peak estimation procedure was used. If a local peak is significantly wider than the predicted peak width, this peak can be supposed to consist of two peaks. Consequently, the number of overlapping peaks is increased by one, and the deconvolution procedure is repeated. As the number of peaks in the chromatogram increases, the resulting sum of squared errors will decrease. This dependence is not linear. In Fig. 3 there is the point A beyond which the sum of squared errors does not significantly decrease with increasing number of peaks supposed to be present. Hence the K value corresponding to point A on the curve is the best estimation of the real number of overlapped peaks. The original chromatogram and the six peaks found in the peak number estimation procedure are shown in Fig. 2.

To illustrate the performance of the peak number estimation procedure in reality, the chromatogram obtained by temperature-programmed capillary gas chromatographic separation of PCBs shown in Fig. 4

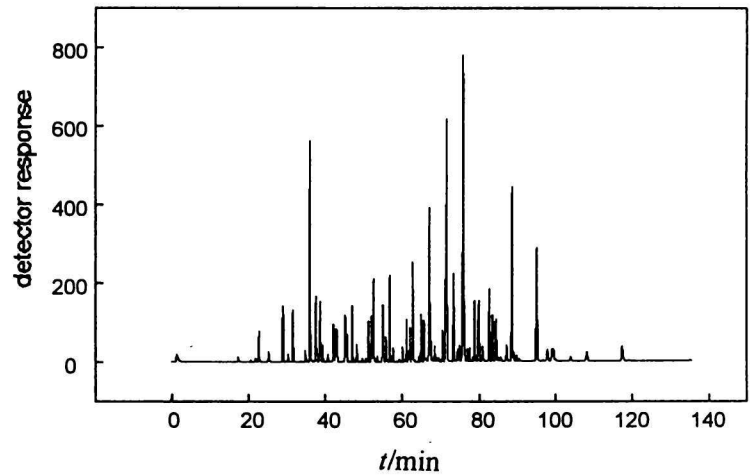

Fig. 4. Gas chromatographic separation of 100 ng of Aroclor 1260. Separation was performed in a SPB-5 capillary column (CHROMPACK International, Middelsburg, The Netherlands) operated with gradient temperature mode (from 60 °C (1 min) to 320 °C with gradient 2.5 °C min^{-1}).

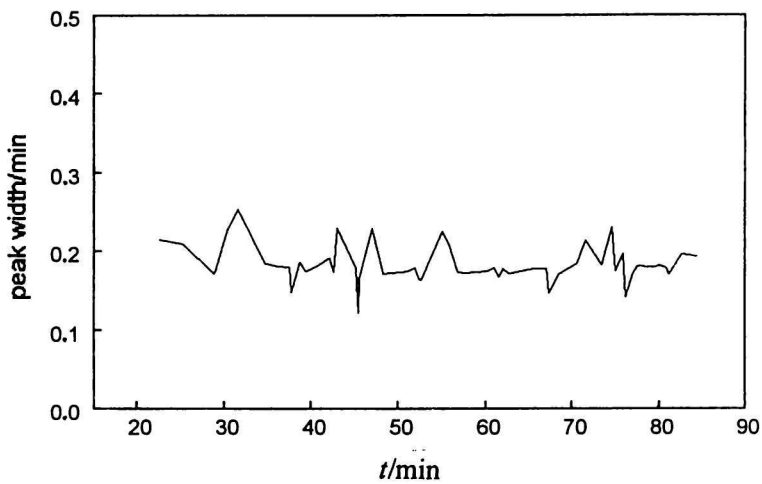

Fig. 5. Dependence of peak widths on retention time found from chromatogram in Fig. 4.

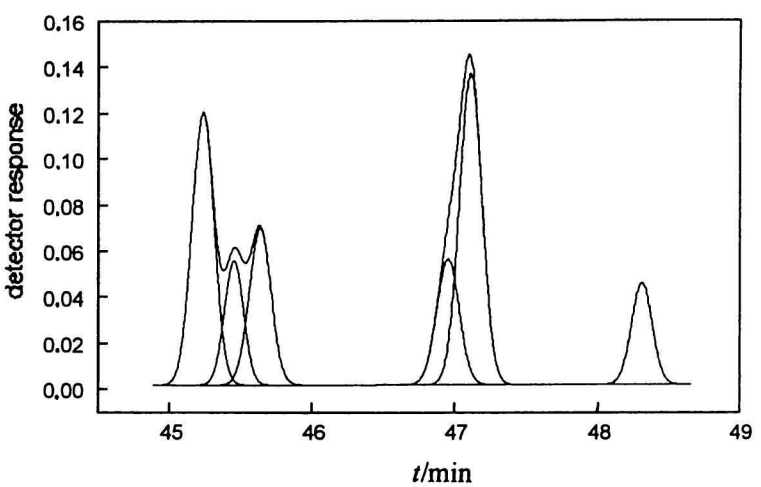

Fig. 6. Section of the PCB chromatogram shown in Fig. 5, after application of the peak number estimation procedure.

is being investigated. For this chromatogram the dependence of the peak widths at half height on the retention time was constructed (Fig. 5). As can be seen there are some peaks which are wider than might be expected from this picture.

An interval containing several peaks was selected, as shown in Fig. 6. In the middle of this interval, at 47 min, there is one wider peak with only two inflection points, indicating the presence of a single peak in this region. After application of the peak number esti-

mation procedure, however, it was found that there is one more peak in this region. The results and the calculated peak profiles in the overlapped peak clusters are shown in Fig. 6.

CONCLUSION

Using the peak number estimation method described, three situations may occur:

1. All of the overlapped peaks in a cluster are apparent. This means that two inflection points for each peak in the chromatogram can be determined, and all the peaks and retention parameters can be found. The peak number estimation method can then be used for confirmation of the known number of peaks.

2. Peaks are overlapped but the inflection points for each single peak are not recognizable. However, the retention times of the peaks under investigation are sufficiently different. In this instance, the proposed method can be used for the determination of peak characteristics.

3. When many peaks are heavily overlapped at nearly the same retention times, it is not practical to apply the mathematical procedure to all of the peaks present. Hence, the estimated number of peaks is therefore lower than the real number of peaks.

REFERENCES

1. Hrouzek, J., Krupčík, J., Čeppan, M., and Hatrík, Š., *Proceedings of the 16th International Symposium on Capillary Chromatography.* Vol. *I*, p. 125. (Sandra, P. and Devos, G., Editors.) Hüthig, Heidelberg, 1994.
2. Martin, M. *9th International Symposium on Advances and Applications of Chromatography in Industry*, Abstracts, p. 16. Bratislava, 1993.
3. Van Rijswick, M. H. J., *Chromatographia 7*, 491 (1974).
4. Hevesi, T., *Dissertation.* Faculty of Chemical Technology, Slovak University of Technology, Bratislava, 1993.
5. Hatrík, Š. and Hrouzek, J., *Chem. Pap. 48*, 376 (1994).
6. Jeansonne, M. S. and Foley, J. P., *J. Chromatogr. Sci. 29*, 258 (1991).
7. Hester, R. D., Vaidya, R. A., and Dickerson, J. P., *J. Chromatogr. 462*, 3 (1989).
8. Schoenmakers, P. J., *Optimization of Chromatographic Selectivity.* Elsevier, Amsterdam, 1986.