ICESD 2013: 19-20 January 2013, DUBAI, UAE

An Empirical and Semi Blind Algorithm for Resolving Overlapped Peaks in Chromatography: Application to the Analysis of Environmental Samples

Mauro Mecozzi[a*] and Michele Scarpiniti[b]

[a]*Laboratory of Chemometrics and Environmental Application, ISPRA, Via di Casalotti 300, 00166, Rome, Italy*
[b]*Department of Information Engineering, Electronics and Telecommunications (DIET), 'Sapienza' University of Rome, Via Eudossiana 18, 00184 Rome, Italy*

Abstract

In this paper we describe a new algorithm for enhancing the resolution (i.e. deconvolution) of chromatographic peaks strongly overlapped among them, often observed in the analysis of complex environmental samples. The main characteristic of this algorithm is that it does not require an "a priori" knowledge of the statistic moments (i.e. width, retention time and height) of the peak to be quantified so that it is considered an empirical and semi-blind (ESB) algorithm. The efficiency of the ESB algorithm has been verified for synthetic overlapped peaks, for the gas chromatographic (GC) analysis of hydrocarbons from marine sediments and for the planar chromatographic analysis (TLC) of carbohydrates in marine organic matter samples. In all the examined cases, standard errors lower than 25% and quadratic (R^2) correlation coefficients higher than 0.85 were obtained, showing the comparability of the ESB algorithm with other deconvolution methods reported in the related scientific literature.

Keywords: Chromatographic analysis, Peak resolution, Environmental monitoring.

1. Introduction

Chromatographic analysis is widely diffused in all the applications of environmental monitoring due to the

---

* Corresponding author. Tel.: +39 06 61570454; fax: +39 06 61561906
E-mail addresses: mauro.mecozzi@isprambiente.it,

possibility of analyzing complex matter samples in automated way. Anyway due to this complexity, it is often necessary to improve the resolution of chromatographic peaks when they result strongly overlapped among them so to enhance the analytical accuracy of chemicals (both natural and pollutant) to be determined. Figure 1 reports the example of strongly overlapped peaks in the case of the gas chromatographic analysis of hydrocarbons from marine sediments.

The scientific literature reports a lot of mathematic approaches for enhancing the resolution of overlapped chromatographic peaks. Most of these approaches belong to the so called curve fitting methods [1] where the resolution of the chromatographic peaks is improved according to different equations of statistical distributions and nonlinear least squares fitting methods [2].
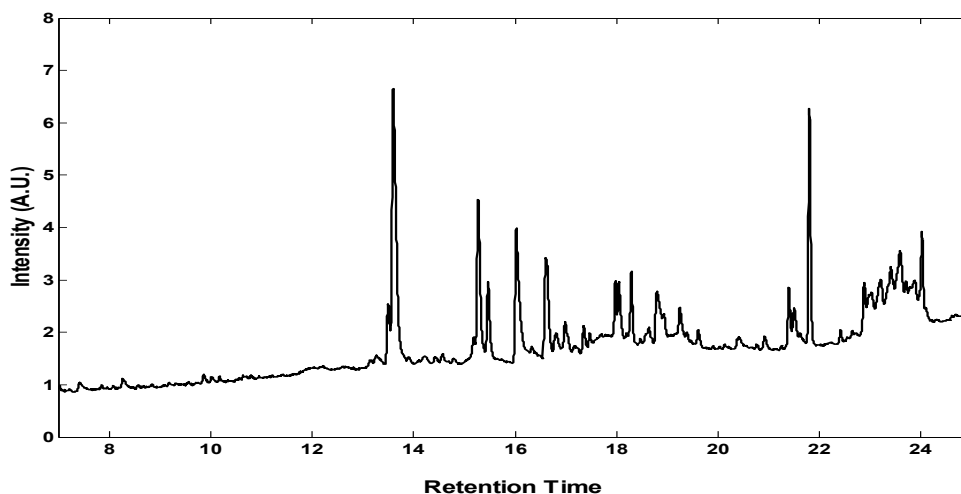

Fig. 1. Example of hydrocarbon extracts from an Antarctic sediment sample; several poorly resolved and overlapped peaks are present after twelve minutes of elution

These techniques generally allow the resolution of overlapped peaks with high analytical accuracy but in any case the users have to predetermine the statistical moments of any peak (i.e. peak height, peak position, peak width at half height) with in addition, the distortion (i.e. asymmetry) factor of the peak caused by its interactions with the stationary phase of the chromatographic support and with the other compounds present in the sample [2, 3, 4]. The proposed ESB algorithm differs from other previous algorithms because it does not requires any predetermination of the statistical moments of the chromatographic peaks. Conversely, it takes into account the time series signals of the overlapped peaks only to perform the peak restoration following an empirical approach and a Lorentian distribution [2].

2. Basic theory of the proposed algorithm

A chromatographic peak can be described as an n-point time series $y = f(t)$ with y is signal intensity and t_i values are the acquisition times following a Lorentzian distribution reported in the Eq. (1)

$$y(t_i) = \frac{h_0}{1 + \frac{(t_i - t_0)^2}{\sigma^2}} \qquad for\ i = 1,\ 2,\ ....,\ n \tag{1}$$

where h_o and t_o are the maximum peak height and retention time (i.e. peak position) respectively and σ is the peak width at half height. In presence of overlapped signals the values of h_o and t_o of any peak are shifted and in addition σ is not a constant value any more due to several experimental factors causing peak asymmetry [2, 3, 4].
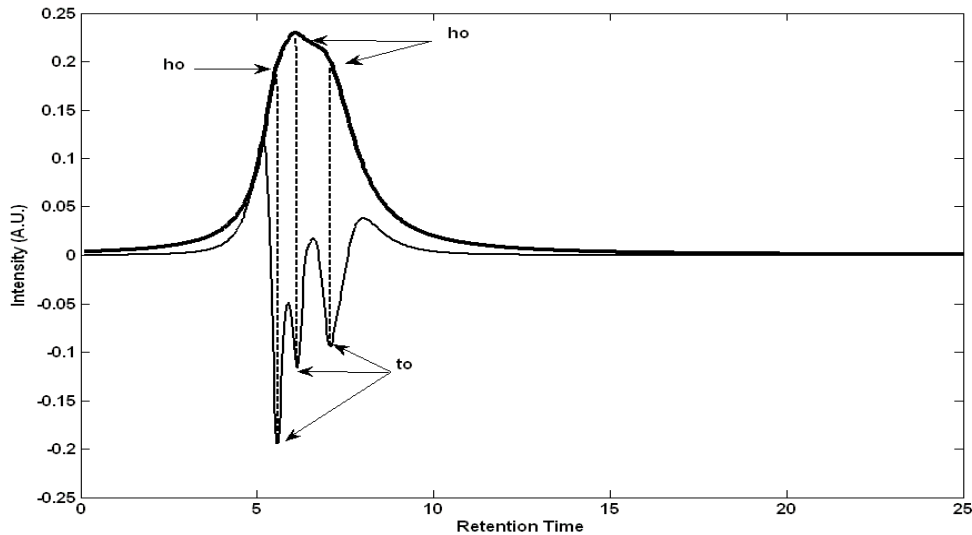

Fig. 2. Example of second derivative (thin line) of the original time series signals consisting of three overlapped peaks (bold line). The dashed lines show the determination by interpolation of the corresponding h_o value of any peak after the identification of t_o.

In the proposed algorithm, the second order derivative of the overlapped signals is determined according to the Savitzki Golay differentiation [5] so to identify the real t_o position of any peak, shifted by the presence of the other peaks. Then this t_o value is interpolated in the $y = f(t)$ series of overlapped signals to determine an approximated (i.e. empirical) h_{oe} value so that $h_{oe} \approx h_o$ (Figure 2).

Torres-Lapacio and co-workers [6] showed that the asymmetry of a Gaussian peak can be described according to

$$\sigma\left(t_i - t_0\right) = \frac{\left|t_i - t_0\right|}{\sqrt{-2\ln\dfrac{y\left(t_i\right)}{h_0}}} \qquad (2)$$

where the h_o and t_o value have the same meaning of Eq. 1 and σ is a point by point $(t_i\text{-}t_o)$ function describing the peak. Anyway, different regions of a given peak do not have the same significance for the calculation of the $\sigma(t_i\text{-}t_o)$ function [6] so we proposes a further development of the Eq. 2. This consists of determining the robust estimation (i.e. σ_m median value) of the time series described by Eq. 2, taking into account the most significant portion of peak corresponding to the range of peak between h_o and $h_o/2$. The Eq. 3 describes the determination of the robust σ estimation as the most probable value of a non Gaussian distribution [7] as

$$\sigma_m = \text{median}\left\{\sigma\left(t_i - t_0\right)\right\} = \frac{\left|t_i - t_0\right|}{\sqrt{-2\ln\dfrac{y\left(t_i\right)}{h_0}}} \quad \text{for} \quad \frac{h_0}{2} \le y\left(t_i\right) \le h_0 \tag{3}$$

Figure 3 shows an example of this approach leading to σ_m estimation. After this step, σ_m and the previously determined t_o and h_o values are used for the restoration of any peak according to the Lorentzian distribution of Eq. 1.
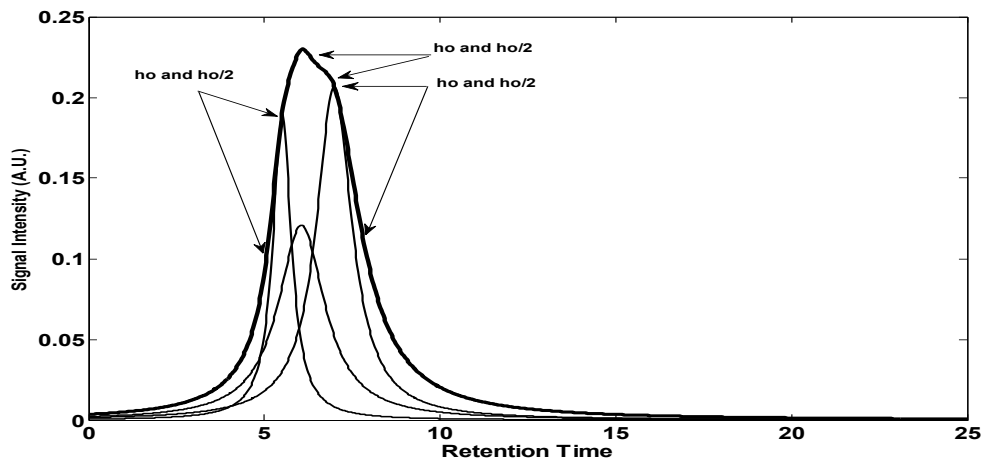

Fig. 3. Example of deconvolution of three overlapped peaks according the steps described in Fig. 2 and Eqs. 1, 2 and 3.

3. Methods and materials

Synthetic overlapped peaks were generated by means of the spreadsheet Excel for Windows according to a Lorentian distribution. GC hydrocarbon and TLC carbohydrates analysis from marine sediments were performed as reported in previous studies [8, 9]. The application of the proposed algorithm was performed by means of in house routines written in Matlab (MathWorks Inc., Natik, WI, USA) language.

4. Results and discussion

Table 1 reports the results for the quantitative estimation of the deconvolved peak area for synthetic chromatographic peaks. An acceptable peak resolution (R_S) must be higher than 1.5 [10] so that having our peaks R_S always lower than 0.8, the relative errors observed between the original area of synthetic peaks and the values after deconvolution, always lower than 25%, are a promising result. These results are comparable with those obtained by curve fitting and non linear least square methods for synthetic peaks [1].

For further validation of the ESB algorithm, we also considered real environmental samples. The TLC chromatograms of monosaccharides show strongly overlapped peaks, giving problems for their quantification (Figure 4). An analogous problem is observed in the GC determination of PAHs extracted from marine sediments where some hydrocarbons show overlapped peaks (Figure 5). Here the reliability of the algorithm has been tested using a different approach with respect to that of Table 1. In the standard samples of PAHs and monosaccharides, overlapped peaks were submitted to deconvolution by the ESB algorithm, then the area of the resolved peaks were used for the instrumental calibration with the known concentrations of any

typology of chemicals.

Table 2 reporting the results of this approach, shows again the reliability of the ESB algorithm, already shown by Table 1. Using R^2 as tool for testing the goodness of calibration [11] we can see values higher than 0.85 confirming the good performances of the ESB algorithm. In addition, the root mean square error of calibration (RMSEC) values obtained for standard solution having concentration ranges between 1 and 20 mg g μg^{-1} confirm the reliability of the instrumental calibration performed taking into account peak area values determined by the ESB algorithm.

Table 1. Example of the application of the ESB algorithm for two and three synthetic overlapped peak systems. "OA" original area, "RA" resolved area, "Err(%)" relative error percent. All the synthetic peaks have Rs < 0.8.

	Two Peaks				**Three Peaks**		
	OA	**RA**	**Err (%)**		**OA**	**RA**	**Err (%)**
Peak 1	5.843	4.974	14.9	Peak 1	5.732	4.702	17.9
Peak 2	7.853	8.821	12.3	Peak 2	5.275	6.544	19.4
				Peak 3	7.416	7.176	3.35
Peak 1	5.843	6.025	3.11				
Peak 2	8.895	9.336	4.96	Peak 1	5.843	6.63	3.64
				Peak 2	7.853	8.301	5.7
Peak 1	8.532	6.791	20.4	Peak 3	12.44	12.201	1.93
Peak 2	12.541	13.61	8.51				
				Peak 1	5.843	5.811	0.56
Peak 1	7.617	5.94	22	Peak 2	7.853	7.11	9.61
Peak 2	8.532	10.61	24	Peak 3	8.532	9.311	9.12
Peak 1	5.843	5.181	11.41	Peak 1	5.732	4.7541	17.1
Peak 2	7.621	8.635	14.4	Peak 2	5.275	5.598	5.77
				Peak 3	8.847	9.511	6.89
Peak 1	5.275	4.224	19.2				
Peak 2	8.847	9.154	3.47				
				Peak 1	5.732	4.495	21.6
Peak 1	5.732	4.393	23.3	Peak 2	8.847	9.07	2.52
Peak 2	5.275	5.71	7.63	Peak 3	7.417	8.704	17.4

Table 2. Results of calibration for the gas chromatographic analysis of overlapped PAHS and for planar chromatographic analysis of overlapped monosaccharides.

PAHs			**Monosaccharides**		
	R²	**RMSEC**		**R²**	**RMSEC**
Benz[a]anthracene	0.851	3.4	Glucose	0.91	1.1
Chrysene	0.997	3.2	Galactose	0.89	1.5
Fluoranthene	0.994	3.5	Fucose	0.92	1.1
Pyrene	0.953	3.3	Ramnose	0.91	1.3

Further studies are in progress to improve the reliability ESB algorithm to resolve overlapped peaks using other types of statistical equations describing the shape of chromatographic peaks.
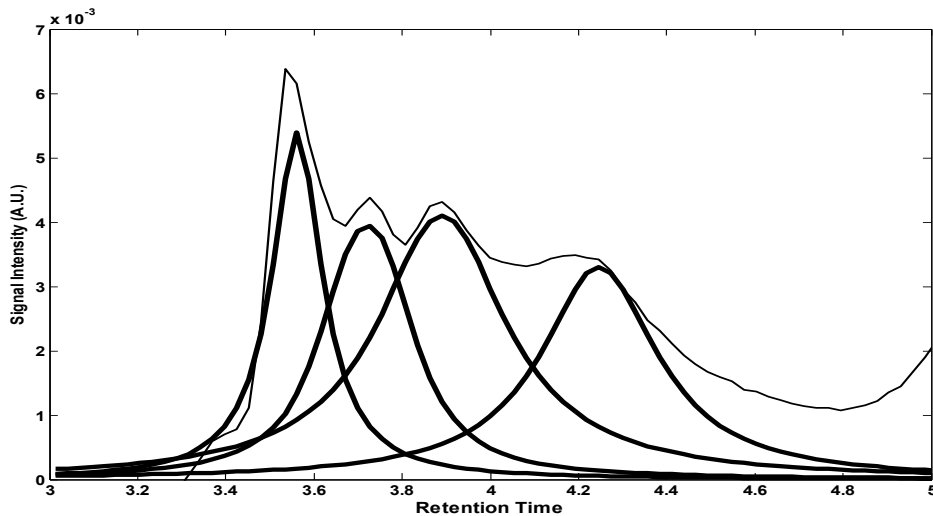

Fig. 4. Example of deconvolution performed by the ESBM algorithm for the resolution of TLC peaks of monosaccharides. See Table 2 for the results of the instrumental calibration
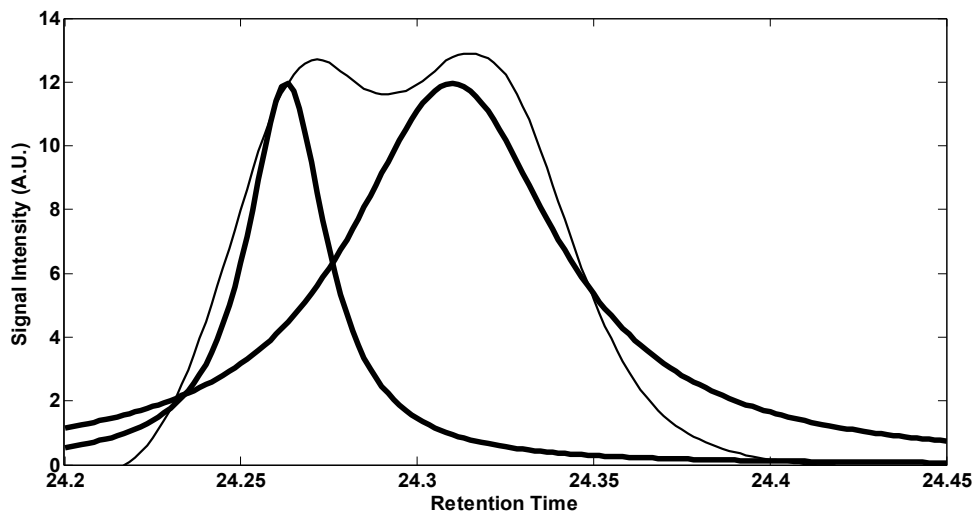

Fig. 5. Example of deconvolution performed by the ESBAM algorithm for the resolution of overlapped gaschromatographic peaks of benza[a]anthracene and chrysene from marine sediment samples.

Acknowledgements

References

[1] Reh E. Peak-shape analysis for unresolved peaks in chromatography: comparison of algorithms. Trends Anal. Chem., 1995, **14**, 1-5

[2] Nikitas P., Pappas-Louisi A., Papageourgu, A. On the equations describing chromatographic peaks and the problem of the deconvolution of overlapped peaks. *J. Chomatogr. A*., 2001, **912**, 13-29

[3] Papai Z., Pap T.L. Determination of chromatographic peak parameters by non-linear curve fitting using statistical moments. *Analyst*, 2002, **127**, 494-8.

[4] Feliger A. Estimation of chromatographic peak shape parameters in Fourier domain. *Talanta*, 2011, **83**, 1074-8

[5] Savitzki A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem., 1964, **36**, 1627-38

[6] Torres-Lapasio J.R., Baeza-Baeza, J.J., Garcia-alvarez.Coque, M.C. A model for the description simulation and deconvolution of skewed chromatographic peaks. *Anal Chem*., **1997**, 68, 3822-31

[7] Miller J.M., Miller J.N. Statistics for analytical chemistry, Hollis Horwood Series, London, UK, 1989,

[8] Pietroletti M.. Mattiello S., Moscato F., Oteri F., Mecozzi M. One step ultrasound extraction and purification method for the analysis of hydrocarbons in marine sediments: application to the monitoring of Italian costs. *Chromatographia*, 2012, **75**, 961-71

[9] Mecozzi M., Papai Z. Application of curve fitting in thin-layer chromatography with flame ionization detection analysis of the carbohydrate fraction in marine mucilage and sediments of the Italian seas. *J. Chrom. Sci*., 2004, **42**, 268-74

[10] http://www.chemistry.adelaide.edu.au/external/soc-rel/content/resolut1.htm