Available online at www.sciencedirect.com
SCIENCE DIRECT®
Talanta 66 (2005) 1073–1087
Talanta
www.elsevier.com/locate/talanta

Non-parametric linear regression of discrete Fourier transform convoluted chromatographic peak responses in non-ideal conditions

Mohamed A. Korany *, Ossama T. Fahmy, Hoda Mahgoub, Hadir M. Maher

Department of Pharmaceutical Analytical Chemistry, Faculty of Pharmacy, University of Alexandria, El-Mesala, Alexandria 21521, Egypt

Received 13 August 2004; received in revised form 2 January 2005; accepted 11 January 2005
Available online 10 February 2005

Abstract

This manuscript discusses the application of chemometrics to the handling of HPLC response data using a model mixture containing ascorbic acid, paracetamol and guaiphenesin. Derivative treatment of chromatographic response data followed by convolution of the resulting derivative curves using 8-points $\sin x_i$ polynomials (discrete Fourier functions) was found beneficial in eliminating different types of interferences. This was successfully applied to handle some of the most common chromatographic problems and non-ideal conditions, namely: very low analyte concentrations, overlapping chromatographic peaks and baseline drift. For example, a significant change in the correlation coefficient of guaiphenesin, in case of baseline drift, went from 0.9978 to 0.9998 on applying normal conventional peak area and first derivative under Fourier functions methods, respectively. It also compares the application of Theil's method, a non-parametric regression method, in handling the response data, with the least squares parametric regression method, which is considered the de facto standard method used for regression. Theil's method was found to be superior to the method of least squares as it assumes that errors could occur in both *x*- and *y*-directions and they might not be normally distributed. In addition, it could effectively circumvent any outlier data points.

Keywords: Chemometrics; HPLC; Derivative; Fourier transform; Convolution; Parametric regression; Non-parametric regression; Theil's method

1. Introduction

1.1. Background

The optimization of a given chromatographic separation, by definition, aims at obtaining well-resolved peaks within a reasonable run time that could be determined quantitatively. Such goal is often hindered by difficulties in establishing the identities of components, and by the lack of well-resolved chromatographic peaks necessary for quantitation.

One of the first steps is to identify the number of components in a chromatogram of closely eluted peaks. This is important both for purity establishment and when analysing complex mixtures. In the chromatograms of samples from natural sources, products of complex reactions, or closely related metabolites, certain regions may be very crowded. The problem of identifying all the components is particularly severe when standards are not available for all components, when some components are present in small amounts and when the number of the co-eluting components have identical or closely similar spectral characteristics, as in the case of stereo-isomers. Chlorophyll degradation is a well-studied example [1].

If spectral data are available for species eluting as poorly resolved chromatographic peaks, the problem of peak identification and peak purity evaluation is then reduced to a relatively straightforward problem of curve resolution· Fitting the spectrum observed under a particular chromatographic peak with a model consisting of the spectra obtained for all candidate species provides a simple route to qualitative analysis of completely resolved peaks, since only one of the model components will be present at a non-zero value in the fitting· The same can be done for poorly resolved chromatographic peaks where two or more components overlap. Quantitative analysis, therefore, is possible for any chromatographic peak

* Corresponding author. Tel.: +20 3 4871668; fax: +20 3 4873273.
E-mail address: makorany@yahoo.com (M.A. Korany).

0039-9140/$ – see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.talanta.2005.01.008

that is defined by a spectrum for which a model exists, even when the chromatographic resolution is far from ideal. Estimates of peak purity may also be obtained for unresolved chromatographic responses [2].

Chemometric techniques have been widely applied for exploring complex chromatograms, e.g., application of photodiode array (PDA) detection HPLC of chlorophyll a allomers [1,3]. Deconvolution of chromatograms into orthogonal polynomials has been applied for characterizing the quality of separatin in patterns of strongly overlapping peaks, e.g., sulpha drugs analyzed by reversed phase HPLC column [4]. Analysis of enantiomers giving partially overlapping peaks was also made possible by using different chemometric treatments of chromatographic ultra-violet signals, e.g., quantification of pseudo-ephedrine enantiomers [5]. Deconvolution of partially overlapping tailing peaks in diode array HPLC using purity ratios has been also applied [6]. Semi-automatic deconvolution of chromatographic data to give pure spectral and chromatographic profiles has been found beneficial in the quantitative determination of some pharmaceutical mixtures [7]. Exploratory chemometric analysis has been used in the classification of pharmaceutical substances based on chromatographic data [8] and for the purpose of pharmaceutical fingerprinting to distinguish among same-product manufacturers [9].

The present study deals with applying the derivative technique by itself or followed by convolution using Fourier functions to eliminate some problems due to non-ideal chromatographic conditions.

This study was carried out using various model mixtures, containing ascorbic acid (ASC), paracetamol (PAR) and guaiphenesin (GUA), in four different cases. These cases are; a three component mixture with well-resolved peaks (ideal case, I), a three component mixture with very low drug concentrations (case II), a three component mixture with overlapping peaks (case III) and finally a single component (GUA) with baseline drift (case IV).

The chromatographic response data was treated by applying the derivative technique (*D* method) alone and the derivative technique followed by convolution using discrete Fourier functions (*D*/FF method).

The study also presents a comparison between two statistical regression methods for handling data; parametric and non-parametric regression methods [10]. The application of the parametric (least squares) regression method assumes that the data being examined follow normal (Gaussian) distribution. However, the non-parametric regression methods can handle data that may not be normally distributed. Since the central limit theorem of the parametric regression is not really valid for the very small data sets frequently used in analytical work, this makes it of interest to apply non-parametric regression approaches to fitting a straight line to a set of points, the simplest of the non-parametric regression methods is Theil's "incomplete" method, so called to distinguish it from another more complex procedure developed by the same author (the "complete" Theil's method) [10].

The non-parametric evaluation of the chromatographic response has been applied for the immunoassay of antigens [11], environmental analysis [12], for data audits to improve the analytical performance [13] and in many aspects of quantitation [14]. Non-parametric regression of data to a straight line has been extended to by-pass any outlier problems with subsequent refitting of the regression line [15,16].

Application of derivative techniques to spectrophotometric data has become a well-established analytical method [17–20]. The elimination of interference by the use of derivative techniques depends on the fact that the first derivative of a constant function is zero and that of a linear function is a constant. Consequently, a first derivative would eliminate constant interferences and a second derivative would eliminate linear interferences.

2. Theory

2.1. Derivative technique (D method)

The application of this method depends on the fact that the chromatographic response (R) is a function of time (τ), thus:

$$D1 = \frac{\mathrm{d}R}{\mathrm{d}\tau} \tag{1}$$

and

$$D2 = \frac{\mathrm{d}^2R}{\mathrm{d}\tau^2} \tag{2}$$

where $D1$ and $D2$ are first and second derivative, respectively.

2.2. Derivative technique followed by convolution using Fourier functions (D–FF method)

The basis of harmonic analysis is that a given function, for example, $D1$ or $D2$ curves of a chromatographic peak, $f(\tau)$ can be expanded in terms of the Fourier series [21–23]. If $(n+1)$ is an odd number, the expansion is

$$f(\tau) = a_0 + a_1\cos x + a_2\cos 2x + \cdots + a_{(n/2)}\cos\left(\frac{n}{2}\right)x + b_1\sin x + b_2\sin 2x + \cdots + b_{(n/2)}\sin\left(\frac{n}{2}\right)x \tag{3}$$

or if $(n+1)$ is an even number then:

$$f(\tau) = a_0 + a_1\cos x + a_2\cos 2x + \cdots + a_{(n+1)/2}\cos\left(\frac{n+1}{2}\right)x + b_1\sin x + b_2\sin 2x + \cdots + b_{(n+1)/2}\sin\left(\frac{n+1}{2}\right)x \tag{4}$$

The calculation of the coefficients $a_1, a_2, a_3 \cdots a_j$ and $b_1, b_2, b_3 \cdots b_j$ is simplified since the trigonometric functions are mutually orthogonal.

Table 1
Chromatographic conditions used for cases I–IV

Case	Elution mode	Mobile phase composition			Run time[a] (min)	Detection wavelength (nm)
		MeOH (%)	Aqueous phase (%)	pH of the buffer		
Ideal case (I)	Isocratic	50	50[b]	6.2	5	225
Very low conc. (II)	Isocratic	50	50[b]	6.2	5	225
Overlapping peaks (III)	Isocratic	88	12[b]	5.8	5	225
Baseline drift (IV)	Linear gradient					
	Start	30	70[c]	–	10	235
	End	70	30[c]			

[a] 1 ml/min flow rate, ambient temperature (in all cases).
[b] 0.05 M phosphate buffer.
[c] 0.25% (v/v) TEA.

Any coefficient t_j, can be calculated from a set of response data measured at equally spaced time intervals, by the following summation, in which x takes values from 0 to $2\pi - [2\pi/(n+1)]$, at intervals of $2\pi/(n+1)$:

$$t_j = \frac{\sum f(\tau)_i T x_i}{\sum (T x_i)^2} \quad (5)$$

where T represents cosine or sine.

The Fourier function coefficients, t_j are proportional to $f(\tau)$. That is:

$$t_j = \alpha_j c \quad (6)$$

where α is a constant and c is the concentration of the analyte.

3. Experimental

3.1. Instrumentation

The chromatographic system consisted of Series 200 Vacuum Degasser (Perkin-Elmer™, USA), Series 200 LC pump (Perkin-Elmer™, USA), Series 200 variable-wavelength UV–vis detector (Perkin-Elmer™, USA) and Series 200 autosampler fitted with a 200 μl sample loop (Perkin-Elmer™, USA). A Perkin-Elmer™ Chromatography Interface 600 Series Link was used. HPLC separations were performed on a Spheri-5 RP stainless-steel C-18 analytical column (250 mm × 4.6 mm) packed with 5 μm diameter particles (Brownlee™ column). Data were processed using TotalChrom Workstation Chromatography Software (Perkin-Elmer™, USA) on an IBM-compatible PC connected to a Laser printer. The digital chromatographic response data were transferred to a personal computer for subsequent processing using Microsoft Excel 2000 (Microsoft Corp., Richmond, VA, USA).

3.2. Materials and reagents

Ascorbic acid (ASC), paracetamol (PAC) and guaiphenesin (GUA) were kindly supplied by Pharco Pharmaceuticals (Alexandria, Egypt). All solvents were of either HPLC or analytical grade, namely: methanol (Panreac Co.,

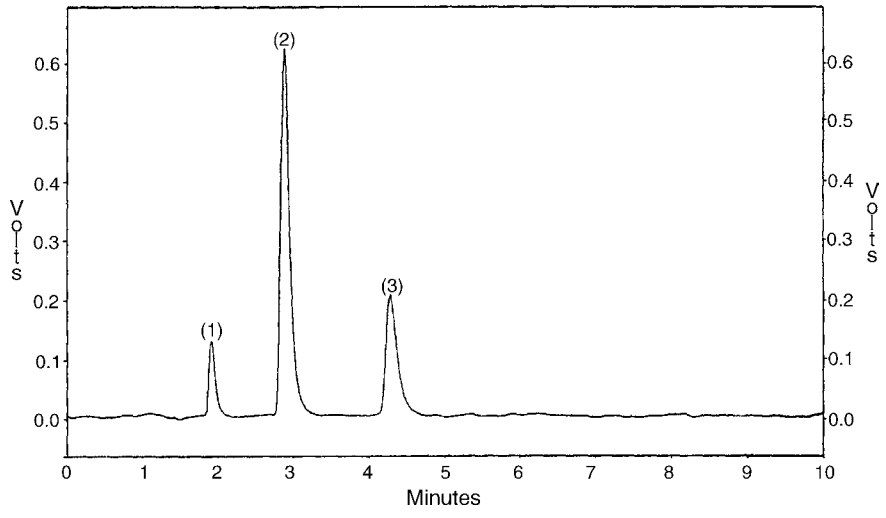

Fig. 1. A chromatogram of a synthetic mixture of 10.0 μg ml^{-1} ASC (1), 40.0 μg ml^{-1} PAR (2), and 15.0 μg ml^{-1} GUA (3) with optimum conditions (ideal case, I).

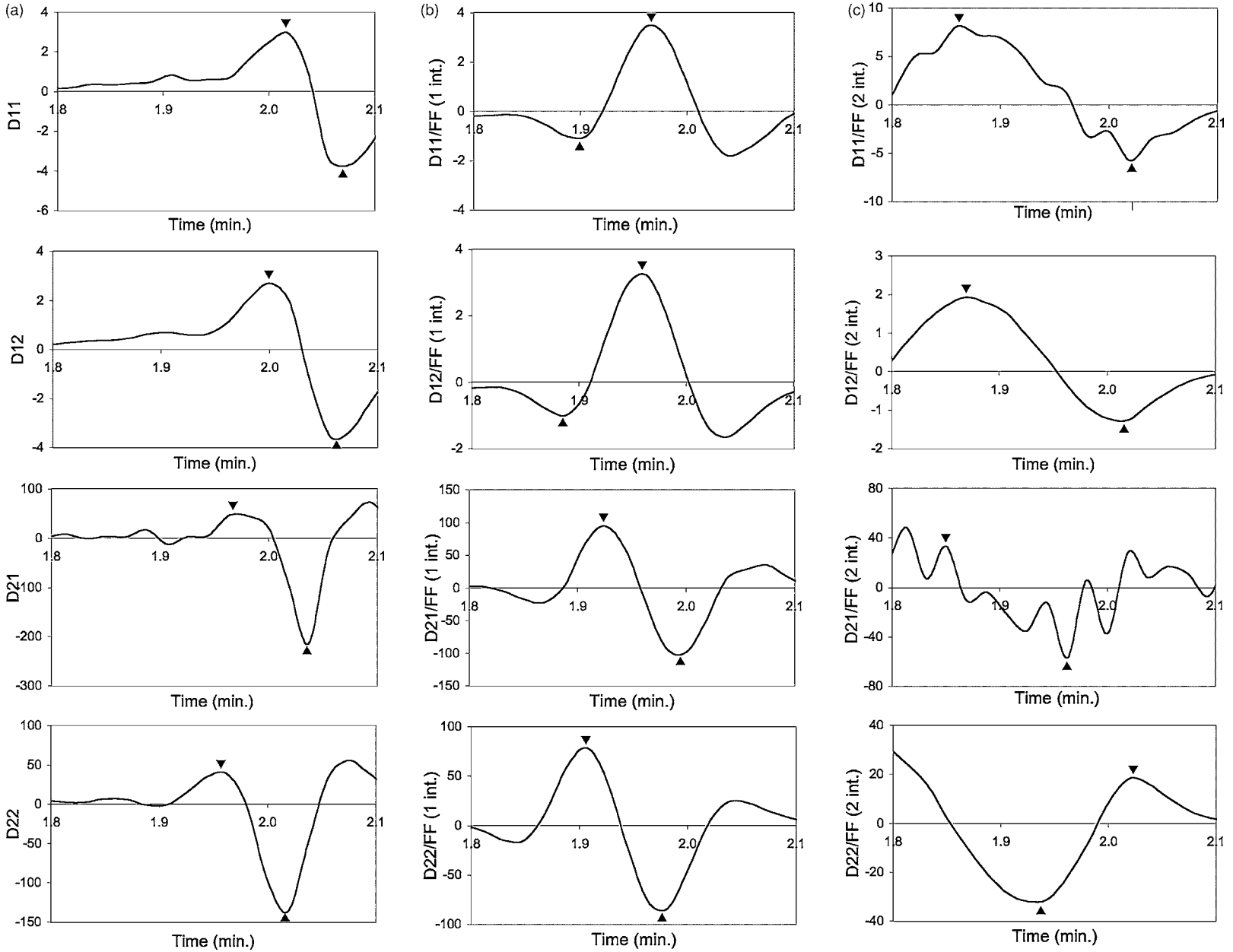

Fig. 2. Derivative curves (a) of 40.0 μg ml^{-1} ASC (▲), and their corresponding convoluted Fourier functions curves at one interval (b) and two intervals (c) in the ideal case, I.

EU), orthophosphoric acid (BDH, Poole, UK) and triethylamine (TEA) (BDH, Poole, UK). The water for HPLC was double glass distilled.

3.3. Chromatographic conditions

In all cases, the mobile phase consisted of methanol and an aqueous phase, which was either phosphate buffer (0.05 M Na_2HPO_4 in water, for cases I, II, III) or TEA solution (0.25%, v/v TEA in water, for case IV). The pH of the mobile phase was adjusted to the required value by dropwise addition of either 0.1 M H_3PO_4 or 0.1 M NaOH solutions, for cases I, II, III. The used chromatographic conditions are summarized in Table 1.

3.4. Preparation of stock and standard solutions

Stock solutions were prepared by dissolving ASC, PAR and GUA in methanol to obtain concentrations of 150.0, 100.0, and 100.0 μg ml^{-1}, respectively. These stock solutions were further diluted with mobile phase to obtain working standard solutions of suitable concentrations (corresponding to the selected linearity range of each compound, from 4.0 to 100.0 μg ml^{-1} for ASC, 1.0 to 60.0 for PAR and 2.0 to 75.0 for GUA, respectively).

3.5. Assay of mixtures containing ASC, PAR and GUA (ideal case, I)

Accurate volumes of each of ASC, PAR and GUA stock solutions were transferred into 10-ml volumetric flasks and diluted to volume with the mobile phase of case I (Table 1) to prepare five standard mixtures within the concentration range of each compound (Section 3.4). Triplicate 20-μl injections were made of each mixture solution and were chromatographed under the conditions described above for case I. For every concentration of each compound, the peak area and peak height were recorded. For each chromatogram (Fig. 1), the response readings at 0.02 min interval (1.6–2.2 min for ASC, 2.8–3.1 for PAR and 4.1–4.5 min for GUA, respectively) were recorded.

The response data were processed using Excel software. Derivative technique (D method) was first applied. For each compound, first ($D1$) and second ($D2$) derivative data at one (0.02 min) and two (0.04 min) time intervals were calculated for each of the five concentrations ($D11$, $D12$, $D21$ and $D22$, respectively). The derivative values (peak-to-peak) were measured at the corresponding time range for each compound. By using the previously obtained first and second derivative data, the Fourier function coefficient, t, was calculated for each drug using the following equation expressed for eight equally spaced time intervals as follows:

$$t = \frac{\{(0)D_0 + (-0.707)D_1 + (+1)D_2 + (+0.707)D_3 + (0)D_4 + (-0.707)D_5 + (-1)D_6 + (-0.707)D_7\}}{4} \quad (7)$$

where D_0 to D_7 stand for eight derivative values; at one or two time intervals.

The numbers in brackets are values of the selected Fourier function [21]. Thus convolutions of the four types of derivative data were made using discrete Fourier functions of 8-points $\sin x_i$ polynomials (D/FF method) at one (0.02 min) and two (0.04 min) time intervals to get convoluted first derivative curves; $D11$/FF and $D12$/FF and convoluted second derivative curves; $D21$/FF and $D22$/FF at one and two time intervals. The convoluted derivative data (peak-to-peak) were measured at the corresponding time range for each compound. As an example, Fig. 2 shows D and D/FF curves for ASC (case I).

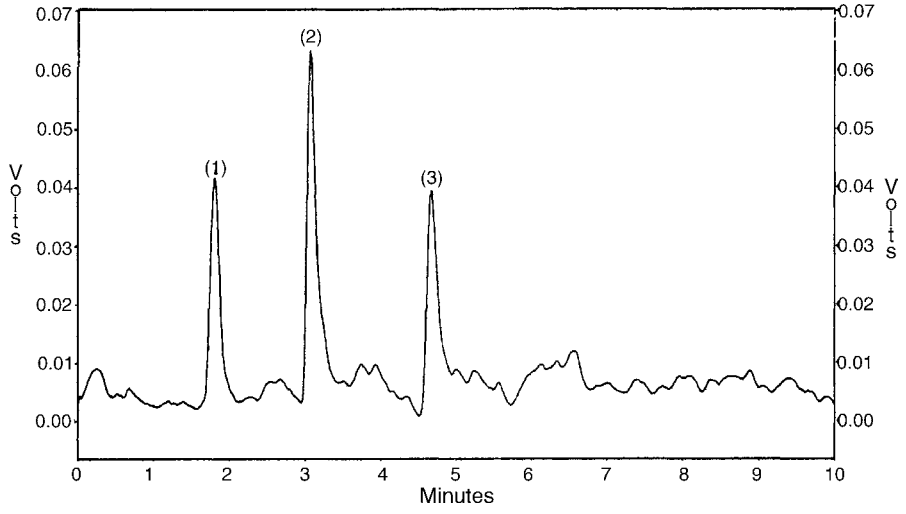

Fig. 3. A chromatogram of a synthetic mixture of 3.0 μg ml^{-1} ASC (1), 0.7 μg ml^{-1} PAR (2), and 1.0 μg ml^{-1} GUA (3) with minor concentrations (case II).

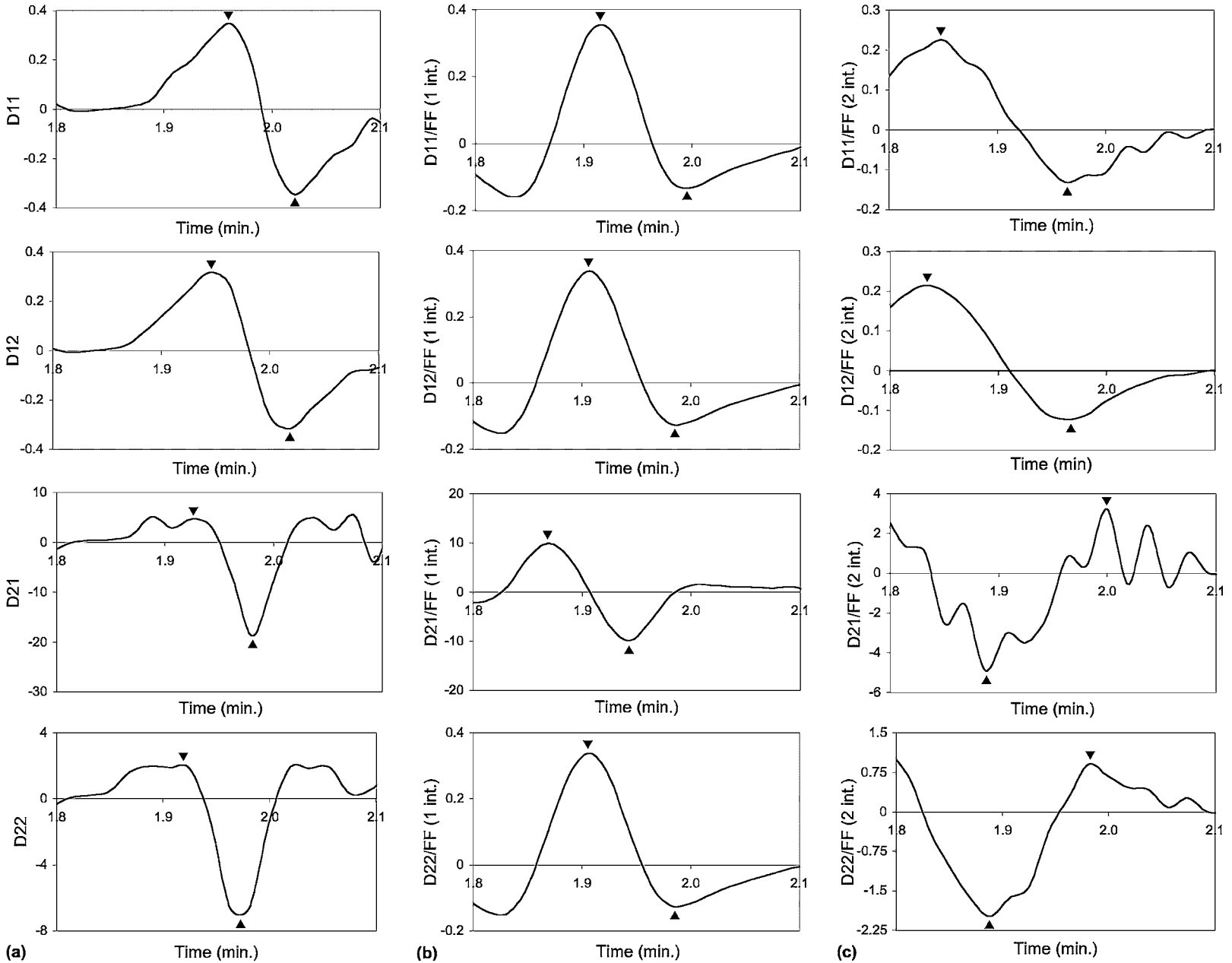

Fig. 4. Derivative curves (a) of 3.0 μg ml^{-1} ASC (▲), and their corresponding convoluted Fourier functions curves at one interval (b) and two intervals (c) in the minor concentrations case, II.

3.6. Assay of mixtures containing very low concentrations of ASC, PAR and GUA (case II)

Accurate volumes of ASC, PAR and GUA stock solutions were transferred into 10-ml volumetric flasks and diluted to volume with the mobile phase of case II (Table 1) to prepare five standard mixtures of very low concentrations (ranging from limit of detection to limit of quantitation for each compound, 2.0–4.0 μg ml^{-1} for ASC, 0.2–1.0 μg ml^{-1} for PAR and 0.5–2.0 μg ml^{-1} for GUA, respectively). Triplicate 20-μl injections were made for each mixture solution and chromatographed under the conditions described above for case II. The chromatograms obtained (Fig. 3) were then chemometrically treated using derivative (*D* method) then derivative/Fourier functions (*D*/FF method) as described under Section 3.5. As an example, Fig. 4 contains *D* and *D*/FF curves for ASC (case II).

3.7. Assay of mixtures of ASC, PAR and GUA with overlapping peaks (case III)

Accurate volumes of ASC, PAR and GUA stock solutions were transferred into a series of 10-ml volumetric flasks and diluted to volume with the mobile phase of case III to prepare five standard mixtures within the concentration range of each compound as in case I. Each mixture solution was chromatographed under the conditions described above for case III (Table 1) to get five standard chromatograms of the three overlapping peaks (Fig. 5).

For each compound, the peak area and peak height were recorded for each concentration. The response data were recorded at 0.02 min intervals over the range of 2.19–2.80 min. The three overlapping peaks were then chemometrically treated using derivative (*D* method) then derivative/Fourier functions (*D*/FF method) as described under Section 3.5, where derivative (*D*) and convoluted derivative (*D*/FF) values (peak-to-peak) were measured over the range of 2.19–2.39, 2.39–2.59 and 2.59–2.79 min for ASC, PAR and GUA, respectively. Fig. 6 shows *D* and *D*/FF curves for the three compounds (case III).

3.8. Assay of GUA in case of baseline shift (case IV)

Accurate volumes of GUA stock solutions were transferred into 10-ml volumetric flasks and diluted to volume with methanol to prepare five standard solutions within the previously mentioned linearity range GUA as in case I. These standard solutions were chromatographed (Fig. 7) using the gradient system for case IV (Table 1).

For each concentration, the peak area and peak height were recorded then chemometric treatment of GUA peaks were proceeded over the range of 8.06–8.56 min exactly as described under Section 3.5. *D* and *D*/FF curves for GUA (case IV) are illustrated in Fig. 8.

4. Results and discussion

Derivative technique (*D* method) and derivative technique followed by convolution using discrete Fourier functions (*D*/FF method) were successfully applied to treat chromatographic signals in order to eliminate interferences in a model mixture containing ASC, PAR and GUA. Different cases of induced chromatographic problems and non-ideal conditions were studied, namely: very low drug concentrations (case II), overlapping spectra (case III) or baseline drift (case IV). Such drug combination has been selected as a representative example as it is commercially available under different brand names from various pharmaceutical companies and in mixtures with other pharmaceutically active ingredients in many cough-cold medications.

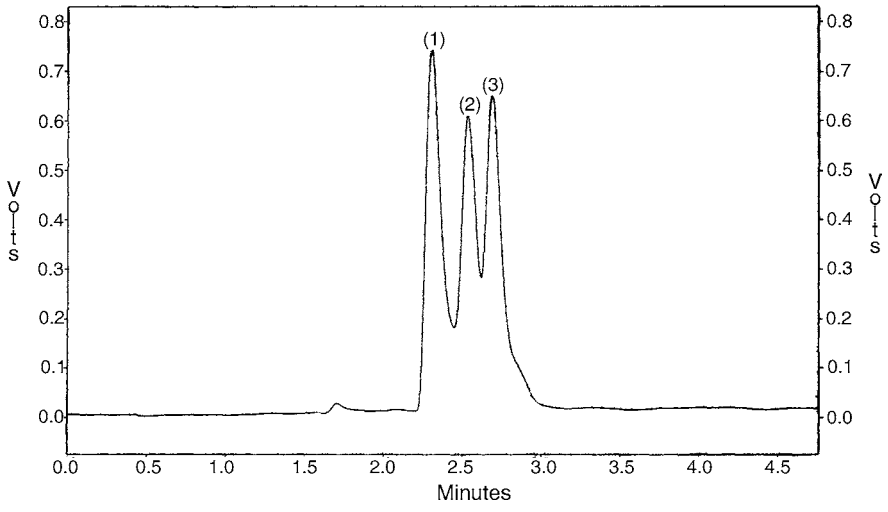

Fig. 5. A chromatogram of a synthetic mixture of 70.0 μg ml^{-1} ASC (1), 40.0 μg ml^{-1} PAR (2), and 45.0 μg ml^{-1} GUA (3) with overlapping peaks (case III).

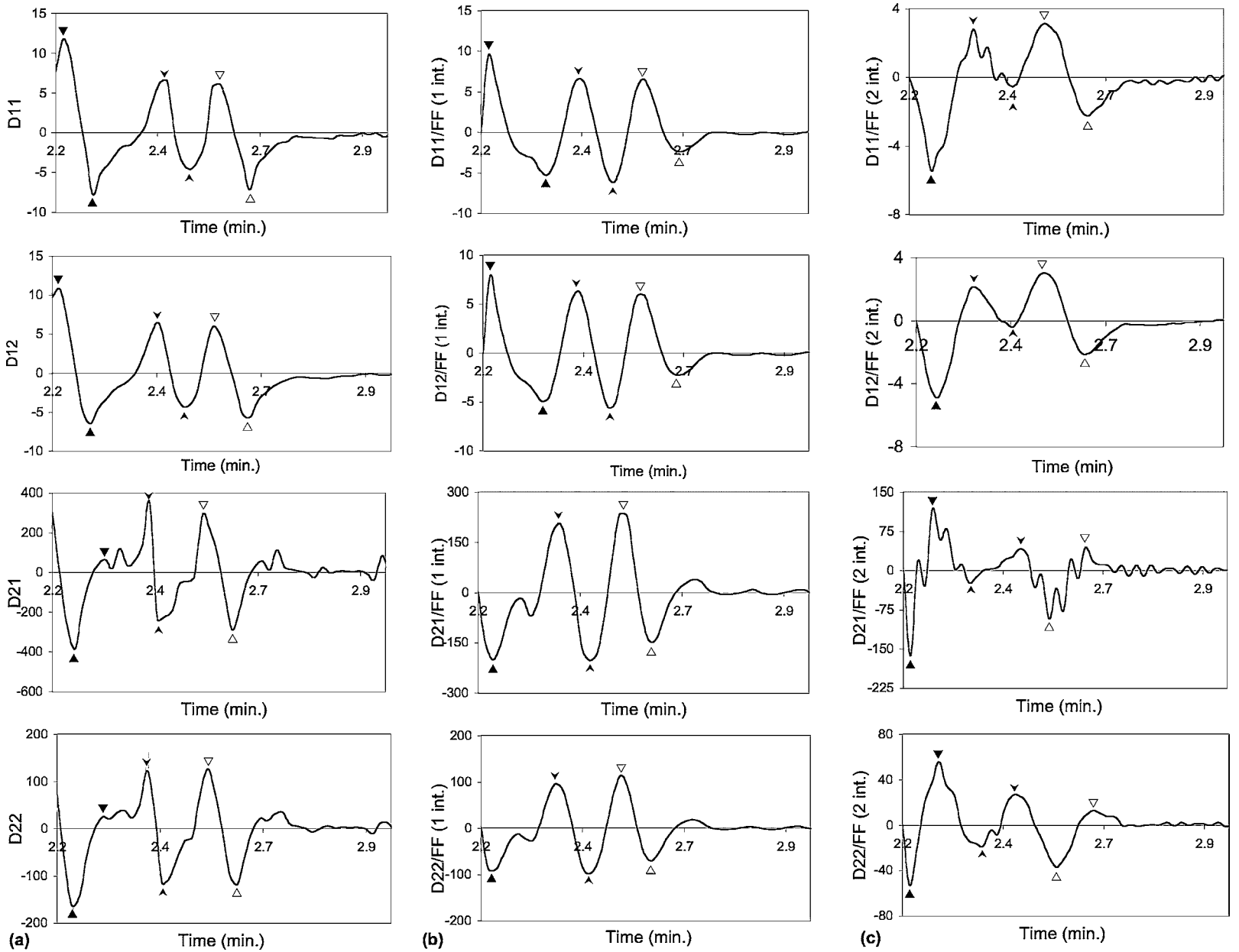

Fig. 6. Derivative curves (a) of a synthetic mixture of 50.0 μg ml^{-1} ASC (▲), 30.0 μg ml^{-1} PAR (⮝), and 50.0 μg ml^{-1} GUA (△), and their corresponding convoluted Fourier functions curves at one interval (b) and two intervals (c) in case of overlapping peaks (case III).

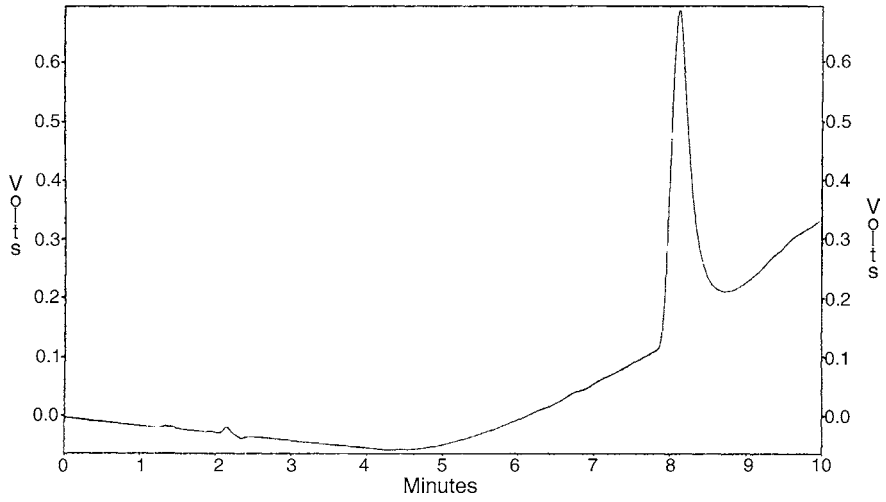

Fig. 7. A chromatogram of 40.0 μg ml^{-1} GUA with baseline drift (case IV).

4.1. Choice of chromatographic conditions

Mobile phase composition was dependant on the intended outcome to be studied. In all cases, excluding case I, it was meant to be a non-ideal chromatogram. Varying the percentage of methanol, the pH and the inclusion of TEA were used to achieve such an imperfect outcome.

4.1.1. Effect of methanol content in the mobile phase

The mobile phase used was 0.05 M phosphate buffer mixed with various proportions of methanol and adjusted to pH 6.2. The mixture of standards was thus injected and run with mobile phases of different composition. Fig. 9 shows the retention times obtained for the three compounds as a function of methanol percentage in the mobile phase. As can be seen, 50% methanol provided optimum resolution with the most symmetric and well-defined peaks. At lower methanol concentrations, separation occurred but with band broadening and increased retention times for GUA peaks. Increasing methanol concentration to 80% or more led to loss of resolution and peak overlapping.

4.1.2. Effect of pH

The influence of the pH of the mobile phase was studied by using 50:50 methanol: 0.05 M phosphate buffer at various pH values between 3.2 and 6.8 (adjusted using orthophosphoric acid or sodium hydroxide). These solutions were used as the mobile phase for a mixture of the three analytes. Fig. 10 illustrates the effect of the mobile phase pH on the retention times of the three compounds. The pH had a marked effect mainly on the retention time of ASC. This was expected since it was the ionizable drug of the three.

4.2. Application of derivative technique (D method) to chromatographic response data

Derivative calculations were applied to response data of the chromatographed mixtures of the previously mentioned four cases. Derivative methods can be applied when direct measurement exhibits some kind of interference. Constant interferences could be eliminated by calculating the first derivative ($D1$), while second derivative ($D2$) can eliminate any linear interferences. For each case, the $D1$ and $D2$ values at one and two time intervals of each of the three compounds were correlated to the concentration as in Tables 2–5 for cases I, II, III and IV, respectively.

4.3. Application of Fourier functions to derivative data (D/FF method)

For each case, the first and second derivative curves were convoluted using 8-points $\sin x_i$ polynomials at one and at two time intervals then the optimum convoluted $D11$/FF, $D12$/FF, $D21$/FF and $D22$/FF values selected for each of the three compounds were related to concentration.

Since convolution using Fourier functions corrects all types of interferences except for linear interference, application of Fourier functions on derivative data would eventually lead to removal of all types of interference producing high degree of purity of analytical peaks. This would be beneficial in cases where high incidence of interferences could be found whether from background noise, as in the case of very low concentrations (case II), from other mixture components as in the case of overlapping peaks where good resolution of the mixture components could not be achieved (case III), or where sample matrix contribution is significant in cases of gradient elution using a mobile phase containing non-transparent reagent at the wavelength of detection leading to a drift in the baseline (case IV).

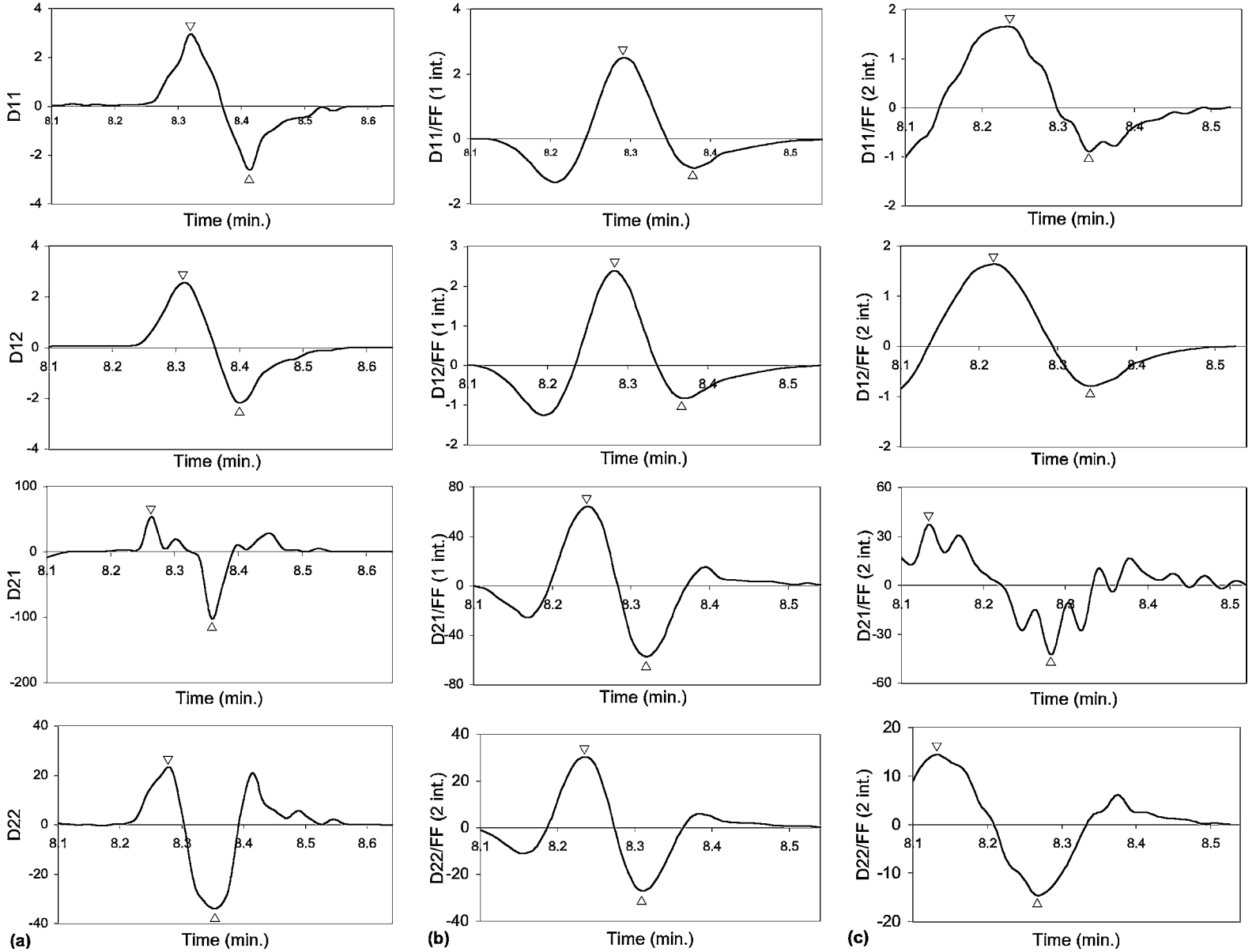

Fig. 8. Derivative curves (a) of 15.0 μg ml^{-1} GUA (△), and their corresponding convoluted Fourier functions curves at one interval (b) and two intervals (c) in the minor concentrations case, IV.

Table 2
Parametric linear regression and statistical parameters for the determination of ASC, PAR and GUA by the proposed HPLC method (ideal case)

	Ascorbic acid		Paracetamol		Guaiaphenisin	
	r^a	F^b	r^a	F^b	r^a	F^b
Direct measurement						
Peak area	0.9992	1832.175	0.9995	3239	0.9992	1818.961
Peak height	0.9975	602.851	0.9990	1172	0.9991	1800.689
Derivative technique (*D* method)						
First derivative (*D*1)						
*D*11	0.9994	2381.092	0.9998	7236.300	0.9994	2718.601
*D*12	0.9994	2552.501	0.9998	7843.422	0.9995	3200.215
Second derivative (*D*2)						
*D*21	0.9989	875.080	0.9983	577.602	0.9985	652.622
*D*22	0.9992	1954.863	0.9992	1325.001	0.9994	1578.851
Derivative under Fourier functions (*D*/FF method)						
First derivative under Fourier functions (*D*1/FF)						
*D*11/FF (1 int.)	0.9998	7976.211	0.9999	13564.572	0.9998	7794.711
*D*12/FF (1 int.)	0.9999	12787.911	0.9999	30230.611	0.9999	15093.801
*D*11/FF (2 int.)	0.9998	7206.798	0.9997	4483.390	0.9998	7412.501
*D*12/FF (2 int.)	0.9996	3958.300	0.9996	3398.481	0.9995	3154.511
Second derivative under Fourier functions (*D*2/FF)						
*D*21/FF (1 int.)	0.9999	10800.601	0.9996	3669.670	0.9999	10062.901
*D*22/FF (1 int.)	0.9998	6886.620	0.9997	4454.743	0.9998	8505.822
*D*21/FF (2 int.)	0.9997	5606.100	0.9993	2252.371	0.9997	4615.770
*D*22/FF (2 int.)	0.9996	4143.200	0.9995	2796.290	0.9996	4199.900

[a] Correlation coefficient.
[b] Variance ratio, equals the mean of squares due to regression divided by the mean of squares about regression (due to residuals).

Table 3
Parametric linear regression and statistical parameters for the determination of ASC, PAR and GUA by the proposed HPLC method in case of very low concentrations

	Ascorbic acid		Paracetamol		Guaiaphenisin	
	r^a	F^b	r^a	F^b	r^a	F^b
Direct measurement						
Peak area	0.9901	149.101	0.9907	159.345	0.9930	213.117
Peak height	0.9916	176.911	0.9625	37.808	0.9897	143.169
Derivative technique (*D* method)						
First derivative (*D*1)						
*D*11	0.9990	1526.951	0.9986	1085.339	0.9991	1636.915
*D*12	0.9989	1383.640	0.9984	972.638	0.9990	1510.043
Second derivative (*D*2)						
*D*21	0.9964	414.552	0.9977	652.264	0.9973	556.187
*D*22	0.9990	1486.650	0.9986	1118.647	0.9988	1246.581
Derivative under Fourier functions (*D*/FF method)						
First derivative under Fourier functions (*D*1/FF)						
*D*11/FF (1 int.)	0.9993	2081.261	0.9994	2712.691	0.9994	2307.285
*D*12/FF (1 int.)	0.9992	1854.122	0.9994	2581.395	0.9994	2425.947
*D*11/FF (2 int.)	0.9991	1658.492	0.9992	1817.047	0.9993	2112.058
*D*12/FF (2 int.)	0.9992	1827.661	0.9991	1741.814	0.9991	1756.511
Second derivative under Fourier functions (*D*2/FF)						
*D*21/FF (1 int.)	0.9990	1558.945	0.9991	1583.455	0.9992	1765.826
*D*22/FF (1 int.)	0.9992	1991.888	0.9993	2046.828	0.9992	1943.041
*D*21/FF (2 int.)	0.9991	1665.254	0.9992	1806.790	0.9992	1881.320
*D*22/FF (2 int.)	0.9992	1846.622	0.9992	1939.188	0.9991	1636.183

[a] Correlation coefficient.
[b] Variance ratio, equals the mean of squares due to regression divided by the mean of squares about regression (due to residuals).

Table 4
Parametric linear regression and statistical parameters for the determination of ASC, PAR and GUA by the proposed HPLC method in case of overlapping peaks

	Ascorbic acid		Paracetamol		Guaiaphenisin	
	r^a	F^b	r^a	F^b	r^a	F^b
Direct measurement						
Peak area	0.9902	150.917	0.9933	221.029	0.9954	423.563
Peak height	0.9858	103.209	0.9858	103.279	0.9898	144.497
Derivative technique (*D* method)						
First derivative (*D*1)						
*D*11	0.9994	2510.881	0.9995	2739.693	0.9996	3581.103
*D*12	0.9994	2484.088	0.9994	2601.503	0.9994	2397.579
Second derivative (*D*2)						
*D*21	0.9990	1562.214	0.9990	1570.551	0.9991	1700.909
*D*22	0.9992	1484.272	0.9994	2526.954	0.9992	1828.325
Derivative under Fourier functions (*D*/FF method)						
First derivative under Fourier functions (*D*1/FF)						
*D*11/FF (1 int.)	0.9999	22621.890	0.9999	17499.620	0.9999	18989.810
*D*12/FF (1 int.)	0.9998	8721.637	0.9998	7016.276	0.9996	3682.042
*D*11/FF (2 int.)	0.9998	7721.781	0.9996	3730.141	0.9996	4034.264
*D*12/FF (2 int.)	0.9996	4183.814	0.9996	4094.742	0.9994	2701.197
Second derivative under Fourier functions (*D*2/FF)						
*D*21/FF (1 int.)	0.9998	6577.934	0.9997	4482.751	0.9996	3935.172
*D*22/FF (1 int.)	0.9997	5182.002	0.9997	5785.880	0.9998	7002.221
*D*21/FF (2 int.)	0.9996	3694.139	0.9994	2324.217	0.9995	2739.357
*D*22/FF (2 int.)	0.9997	4721.673	0.9996	3331.207	0.9998	8952.405

[a] Correlation coefficient.
[b] Variance ratio, equals the mean of squares due to regression divided by the mean of squares about regression (due to residuals).

Table 5
Parametric linear regression and statistical parameters for the determination of GUA by the proposed HPLC method in case of baseline drift

	Guaiaphenisin	
	r^a	F^b
Direct measurement		
Peak area	0.9990	1561.759
Peak height	0.9978	674.424
Derivative technique (*D* method)		
First derivative (*D*1)		
*D*11	0.9993	2043.707
*D*12	0.9992	1904.519
Second derivative (*D*2)		
*D*21	0.9989	1354.755
*D*22	0.9990	1564.031
Derivative under Fourier functions (*D*/FF method)		
First derivative under Fourier functions (*D*1/FF)		
*D*11/FF (1 int.)	0.9998	10425.961
*D*12/FF (1 int.)	0.9996	4111.341
*D*11/FF (2 int.)	0.9996	3572.654
*D*12/FF (2 int.)	0.9994	2655.697
Second derivative under Fourier functions (*D*2/FF)		
*D*21/FF (1 int.)	0.9997	4919.478
*D*22/FF (1 int.)	0.9996	4011.603
*D*21/FF (2 int.)	0.9993	2290.423
*D*22/FF (2 int.)	0.9995	3262.978

[a] Correlation coefficient.
[b] Variance ratio, equals the mean of squares due to regression divided by the mean of squares about regression (due to residuals).

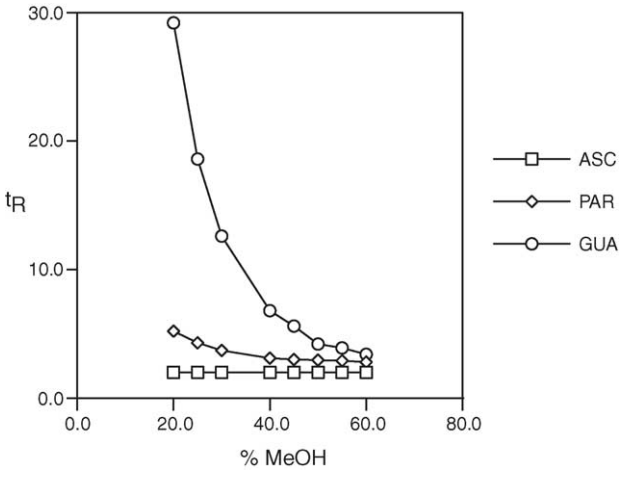

Fig. 9. Variation of the retention times of a synthetic mixture of ASC, PAR and GUA as a function of the percentage of methanol in the mobile phase.

4.4. Methods validation

4.4.1. Calibration graphs and statistical data

Under the previously described chromatographic conditions for each of the four cases, the graphs obtained by plotting derivative and convoluted derivative data versus concentration for each of the three compounds, show various degrees of linearity and were compared to those obtained using the peak area and peak height as the response signals. Using the method of least squares, regression equations, correlation coefficients (r) and standard deviation of residuals ($S_{y/x}$) were calculated. Variance ratio (F) was also determined (Tables 2–5).

The value of the correlation coefficient (r) indicates the degree of goodness of linearity of the calibration graph. Standard deviation of residuals ($S_{y/x}$) is also called standard error of estimate since it estimates random errors in the y-direction. ($S_{y/x}$) is a measure of the extent of deviation of the found (measured) y-values from the calculated ones. The smaller the ($S_{y/x}$), the closer the points are to the linear regression line. It was noticed that regression lines with high (r) values showed also low ($S_{y/x}$) values. For equal degrees of freedom, increase in the variance ratio (F-values) means increase in the mean of squares due to regression and decrease in the mean of squares due to residuals. The greater the mean of squares due to regression, the more the steepness of the regression line is. The smaller the mean of squares due to residuals, the less the scatter of the experimental points around the regression line. Consequently, regression lines with high F-values (low significance F) are much better than those with lower ones. Good regression lines show high values for both (r) and (F) values [23].

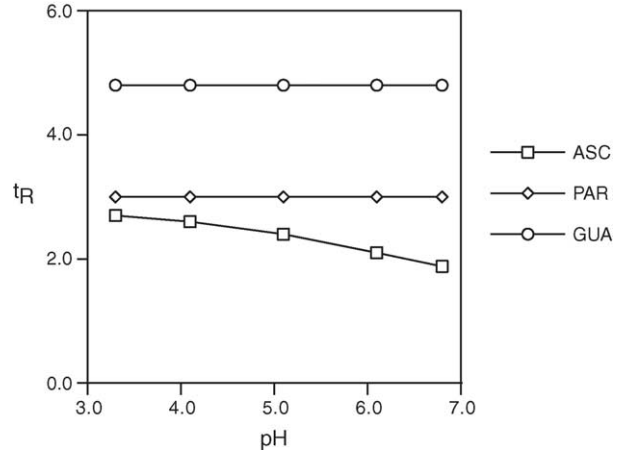

Fig. 10. Variation of the retention times of a synthetic mixture of ASC, PAR and GUA as a function of the mobile phase pH.

Concerning all the previous regression and statistical parameters, regression lines obtained using different types of linearities were compared.

For the ideal case (case I), for all three drugs, derivative values (D) produced better results than either peak area or peak height· However, convoluted derivative values (D/FF) showed the best regression lines especially for convoluted derivative values at one interval: $D$21/FF (1 int.), $D$12/FF (1 int.) gave high (r) values up to 0.9999 for each of the three compounds.

The influence of derivative (D method) and convoluted derivative (D/FF method) in improving the regression and statistical parameters was obvious in cases where sources of interference may exist· In case of very low concentration (case II, Table 3), peaks overlap (case III, Table 4) or baseline drift (case IV, Table 5), convolution of derivative data produced calibration graphs with much better regression and statistical parameters than those obtained using peak area or peak height. For case II, $D$11/FF (1 int.) and $D$12/FF (1 int.) produce the best results, with (r) values up to 0.9994 (for PAR and GUA) or 0.9993 (for ASC). For case III, $D$11/FF (1 int.), $D$12/FF (1 int.), $D$21/FF (1 int.) and $D$22/FF (1 int.) produce the best results, with (r) values up to 0.9999 for each of the three compounds. With case IV, $D$11/FF (1 int.) and $D$21/FF (1 int.) give (r) values up to 0.9998 for GUA.

4.4.2. Application of non-parametric regression methods

In parametric statistics, the arithmetic mean or average was used as the “measure of central tendency” or “measure of location.” This is logic enough when “symmetrical” normal distribution is assumed, but in non-parametric statistics, the median is usually used instead [10]. Determining the median of a set of experimental results usually requires little or no calculation· Moreover, in many cases it may be a more realistic measure of central tendency than the arithmetic mean.

Concerning the linear regression methods, the assumption of normally distributed y-direction errors was emphasized, and the complexity of some of the calculation methods was apparent. This complexity is largely overcome by using modern calculators or personal computers, and there are some rapid approximation methods for fitting straight lines to experimental data. There is still an interest to non-parametric approaches to fitting a straight line to a set of points. Of the

Table 6
Comparison between parametric (P) and non-parametric (NP) regression models for the determination of GUA by the proposed HPLC method in case of baseline shift

	Guaiaphenisin					
	$\|a\|$[a]		$\|b\|$[b]		Percentage change in $\|a\|$[c]	Percentage change in $\|b\|$[d]
	P	NP	P	NP		
Direct measurement						
Peak area	349334	113204	6081506	6107260	−67.5943	0.4235
Peak height	16697	2160	340178	348453	−87.0635	2.4326
Derivative technique (*D* method)						
First derivative (*D*1)						
*D*11	0.3885	0.5938	29.8879	29.8750	52.8443	−0.0432
*D*12	0.1937	0.1752	30.5143	30.8752	−9.5509	1.1827
Second derivative (*D*2)						
*D*21	0.3793	0.1251	30.1528	30.6252	−67.0182	1.5667
*D*22	0.7166	0.4375	29.8523	30.4375	−38.9478	1.9603
Derivative under Fourier functions (*D*/FF method)						
First derivative under Fourier functions (*D*1/FF)						
*D*11/FF (1 int.)	0.6390	0.7187	29.3177	29.5625	12.4726	0.8349
*D*12/FF (1 int.)	0.6051	0.3437	29.7403	29.5625	−43.1995	−0.5978
*D*11/FF (2 int.)	1.6438	1.3438	26.4002	26.4375	−18.2504	0.1413
*D*12/FF (2 int.)	0.3246	0.1521	28.9562	29.2501	−53.1423	1.0150
Second derivative under Fourier functions (*D*2/FF)						
*D*21/FF (1 int.)	0.7943	0.6872	30.3869	30.6255	−13.4836	0.7852
*D*22/FF (1 int.)	0.4806	0.7187	31.7729	31.8125	49.5422	0.1246
*D*21/FF (2 int.)	1.5331	0.5002	27.9328	29.0000	−67.3733	3.8206
*D*22/FF (2 int.)	0.7349	0.1252	30.6771	30.1251	−82.9637	−1.7994

[a] Modulus of intercept.
[b] Modulus of slope.
[c] Percentage change in $|a|$ means percentage change in $|a|$ of NP vs. $|a|$ of P = [($|a|$ of NP − $|a|$ of P)/$|a|$ of P] × 100.
[d] Percentage change in $|b|$ means percentage change in $|b|$ of NP vs. $|b|$ of P = [($|b|$ of NP − $|b|$ of P)/$|a|$ of P] × 100.

several methods available, perhaps the simplest is Theil's "incomplete method" that was applied to our data [10].

Theil's method determines the slope of a regression line as the median of the slopes calculated from selected pairs of points: the intercept of the line is the median of the intercept values calculated from the slopes and the coordinates of the individual points [10].

For all of the previously mentioned types of linearity, and for each of the four cases, the response data were handled using Theil's method. The best-fit straight line obtained using Theil's method was compared with the least squares best fit line calculated using the parametric regression method. Table 6 represents an example illustrating that the non-parametric regression model could be considered superior

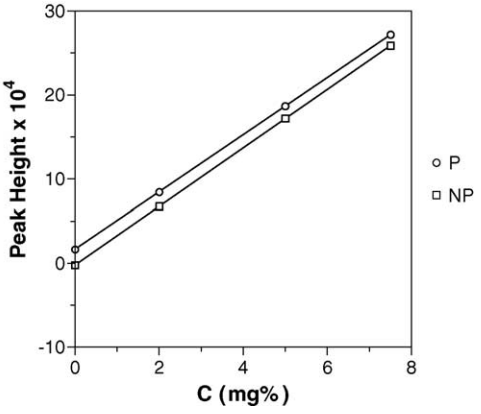

Fig. 11. Regression lines calculated by Theil's method, non-parametric (NP), and by the least squares method, parametric (P), for the determination of GUA using peak height in case of baseline drift (case IV).

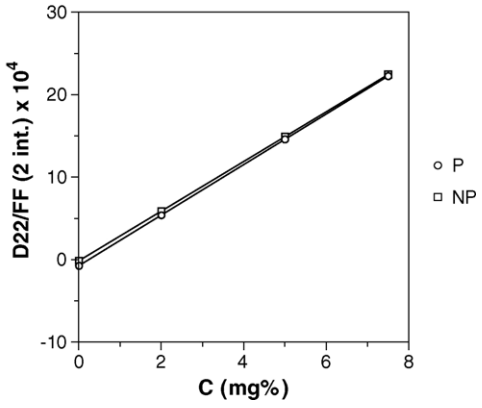

Fig. 12. Regression lines calculated by Theil's method, non-parametric (NP), and by the least squares method, parametric (P), for the determination of GUA using D22/FF (2 int.) in case of baseline drift (case IV).

over the parametric one and this was proved by calculating the percentage change in the intercept and slope. From this table, it can be seen that the percentage change in the intercept when applying the non-parametric relative to the parametric regression models was from −9.55 to −87.06% and that, in most cases, the intercept decreases almost to nil when applying the non-parametric regression model. However, the percentage change in slope was virtually negligible (from −0.04 to +3.82%).

Theil's method has three distinct advantages over the least squares method: first, it does not assume that all the errors are in the *y*-direction; second, it does not assume that either the *x*- or *y*-direction errors are normally distributed; and third it is not affected by the presence of outlying results. Generally, an outlier value does not affect the Theil's calculation at all since it does not affect the median estimate of the slope or intercept. In the least squares calculation, however, the outlying point carries as much weight as the other points. This leads to the fact that, the least squares line passes closer to the outlier than the non-parametric line does [10]. Figs. 11 and 12 show examples illustrating this fact. Since regression methods that are relatively unaffected by outlying data are necessary to provide unbiased estimates of the slope for data conforming to the straight line function, the non-parametric Theil's method is highly beneficial in this respect.

5. Conclusion

Chemometric techniques could be successfully applied for handling complex chromatograms. This is highly needed in cases where sources of interference could dramatically affect the chromatographic response, e.g., very low concentrations, overlapping peaks or baseline drift. Derivative treatment of the chromatographic response data followed by application of Fourier functions on the resulting derivative data gives improved quantitation of the chromatographic signals.

Non-parametric regression of the response data using Theil's method is highly advantageous over the usual least squares method. "Theil's method" could be used in cases where there are both *x*- and *y*-direction errors assuming that the errors are not normally distributed. It also has effectively circumvented the outlier problem.

References

[1] K. Kavianpour, R.G. Brereton, Analyst (Cambridge, UK) 123 (1998) 2035–2042.
[2] M. Redmond, S.D. Brown, H.R. Wilk, Anal. Lett. 22 (1989) 963–979.
[3] K.D. Zissis, S. Dunkerley, R.G. Brereton, Analyst (Cambridge, UK) 124 (1999) 971–979.
[4] H.J.G. Debets, A.W. Wijnsma, D.A. Doornbos, H.C. Smit, Anal. Chim. Acta 1 (1985) 17133–17143.
[5] J. Verdu-Andres, R. Herraez-Hernandez, P. Campins-Falco, J. Chromatogr. A 930 (2001) 95–107.
[6] R.G. Brereton, S. Dunkerley, Analyst (Cambridge, UK) 124 (1999) 705–711.
[7] F. Gan, X.N. Li, Y.Z. Liang, Fenxi Huaxue 28 (2000) 833–836, Thru Anal. Abstr. (CD edition).
[8] A. Detroyer, V. Schoonjans, F. Questier, Y. Vander-Heyden, A.P. Borosy, Q. Guo, D.L. Massart, J. Chromatogr. A 897 (2000) 23–36.
[9] W.J. Welsh, W. Lin, S.H. Tersigni, E. Collantes, R. Duta, M.S. Carey, W.L. Zielinski, J. Brower, J.A. Spencer, T.P. Layloff, Anal. Chem. 68 (1996) 3473–3482.
[10] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, fourth ed., Prentice Hall, Harlow, England, 2000, pp. 170–172.
[11] Y. Kurtulmus, T. Tanyalcin, G. Bozkaya, O. Gunduz, O. Cerci, F.Z. Kutay, B. Ersoz, Accred. Qual. Assur. 6 (2001) 424–426.
[12] L. Paksy, Mikrochim. Acta 123 (1996) 197–205.
[13] G.E. Rayment, R.O. Miller, E. Sulaeman, Commun. Soil Sci. Plant Anal. 31 (2000) 1513–1530.
[14] G. Lippi, C. Brentegani, C. Mazzi, C. Recchi, O. Ruzzenente, G. Guidi, Eur. J. Clin. Chem. Clin. Biochem. 35 (1997) 877–880.
[15] M.O. Moen, K.J. Griffin, A.H. Kalantar, Anal. Chim. Acta 277 (1993) 477–487.
[16] J.S.O. Odonde, J. Autom. Chem. 14 (1992) 25–27.
[17] M.A. Korany, A.M. Wahbi, M.A. Elsayed, S. Mandour, Farmaco 39 (1984) 243–252.
[18] M.A. Korany, A.M. Wahbi, S. Mandour, M.A. Elsayed, Anal. Lett. 18 (1985) 21–34.
[19] K. Kovacs-Hadady, I.T. Kiss, M. Kiss, K. Barna-Katona, Analyst (London) 113 (1988) 569–571.
[20] A. El-Gindy, M.A. Korany, M.F. Bedair, J. Pharm. Biomed. Anal. 17 (1998) 1357–1370.
[21] A.M. Wahbi, H. Abdine, M.A. Korany, Pharmazie 33 (1978) 278–282.
[22] P. Armitage, G. Berry, Statistical Methods in Medical Research, third ed., Blackwell Scientific Publications, Oxford, England, 1994, pp. 283–285.
[23] M.A. Korany, M.A. Elsayed, M.M. Bedair, H. Mahgoub, E.A. Korany, Talanta 37 (1990) 1183–1188.