ELSEVIER

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

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.
© 2005 Published by Elsevier B.V.

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).

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

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

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.