ELSEVIER

Journal of Chromatography A, 758 (1997) 1–18

JOURNAL OF CHROMATOGRAPHY A

# Simulations of liquid chromatography–diode array detector data including instrumental artefacts for the evaluation of mixture analysis techniques

J.A. Gilliard[a,*], C. Ritter[b]

[a]*Quality Assistance SPRL, Technoparc de Thudinie, 1, B-6536 Donstiennes, Belgium*

[b]*Institut de Statistique, Université Catholique de Louvain, Voie du Roman Pays, 34, B-1348 Louvain-la-Neuve, Belgium*

Received 18 March 1996; revised 31 May 1996; accepted 31 July 1996

**Abstract**

The application of mixture analysis techniques to liquid chromatography–diode array detector (LC–DAD) data can be hampered by instrumental artefacts. If not properly accounted for, these artefacts may create major problems in the interpretation of the results by introducing additional factors in the analysis that can be difficult to distinguish from minor chemical components. In this paper, we show how the main sources of instrumental errors can be identified and how their effect can be mimicked by a specifically designed simulation procedure. Such realistic simulations are then successfully used to anticipate the expected response for an homogeneous peak even under non-ideal conditions, and hence to facilitate unequivocal interpretation of results.

*Keywords:* Artefacts; Detection, LC; Diode array detector; Mixture analysis; Computer simulation; Simulation techniques; Alprazolam; Triazolam; Hydrochlorthiazide

## 1. Introduction

One important problem in analytical chemistry is the determination of the number of components in a sample. In the pharmaceutical industry for example, the composition and purity of production streams need to be monitored routinely. Liquid chromatography (LC) is often used in this context and its potential, especially for purity control, has increased considerably with the introduction of the multichannel diode array detector (DAD). The data set generated by a LC–DAD system is called a spectrochromatogram and can be described as a matrix in which the rows are the absorption spectra measured at regular time intervals and the columns are the chromatograms measured at different wavelengths. In order to extract useful information from a spectrochromatogram a number of techniques based on factor analysis [1–12] have been proposed. The ultimate goal of all these techniques as stated by Delaney [13] is "to determine the number of components in an overlapping chromatographic peak as well as the spectrum and the concentration profile of each compound, without assumption regarding peak shape, location, or identity". In practice however, these methods are principally used as qualitative tool and for peak purity [12], and several techniques have been specially developed to achieve a more accurate peak purity control [14–16].

The key step in all factor analysis approaches

*Corresponding author.

0021-9673/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved
*PII* S0021-9673(96)00676-0

is principal components analysis (PCA) in which a reduced multivariate space is defined. All these techniques are based on the assumption that the data measured with a LC–DAD system follow a bilinear structure, meaning that each compound present in the sample can be expressed as the outer product of a chromatogram and a pure spectrum. Under this structure, the number of base factors required to describe the variance of the measurements within the experimental noise equals the number of compounds in the mixture. The problem is then to extract the relevant or primary factors from the non relevant factors summarising noise. Most methods for estimating the number of primary factors (i.e., the rank of the investigated matrix) use information from the eigenvalues. Based on the assumption that spectral information is broad band (low frequency) whereas noise is narrow band (high frequency), methods based on the shape of the eigenvectors have also been proposed [17–19]. Following this assumption, primary eigenvectors can be distinguished from noise eigenvectors by their smooth appearance.

In practice however, non-ideal response behaviour is observed for UV detectors causing deviations from the hypothetical bilinear structure. Generally incriminated sources of artefacts include non-zero or sloping baseline, diode-array scan time, non-linear response due to polychromatic radiation and stray light, and heteroscedasticity. Consequently, extra primary factors are extracted by eigenvalue based rank estimation methods and the corresponding eigenvectors display a more or less smooth appearance [20]. Therefore, the task becomes more complex since one has to distinguish between signals due to instrumental artefacts and signals originating from chemical components.

There are three general strategies for dealing with non-ideal response behaviour: to correct for it, to work at conditions at which none is observed and to learn how to cope with it. First, mathematical corrections have been proposed for the effect of the baseline [21–23], the scan rate [23] and for heteroscedasticity [24,25]. For other important artefacts, such as polychromatic radiation, good correction still needs to be found and a complete suppression is currently not feasible. Second, both noise and deviations from linearity are greater at high absorbance. Their effect can therefore be minimised by keeping the maximum absorbance low [26–28]. This leads however to a loss of detection power. Moreover, to define a general ceiling for the maximum absorbance is difficult, since it would depend not only on the characteristics of the detector but also on the spectral shape of the investigated compounds [29]. Spectral regions of fine structure or of high slope/absorbance ratio may effectively show non-linearities at absorbances as low as 100 mAU [30]. Finally, most of the disturbances is expected under the region of the peak of higher intensity, i.e., in the neighbourhood of the apex of the peak. Therefore, for techniques which assess peak purity as a function of the elution time, such as error analysis in multicomponent analysis or window evolving factor analysis, several authors recommend that any peak impurity assignment aligned with the apex of the investigated peak should be considered with caution [27,28,30]. Extending this thought, Gerritsen et al. showed that the structure of the extra factors induced by the non-linear response behaviour show large similarities with different LC–DAD systems [20]. However, these authors did not suggest an interpretation of the observed eigenvector structures that could be useful in order to discriminate between base and extra primary factors. In fact, only the heuristic evolving latent projections (HELP) method allows an albeit limited diagnosis between real impurity detection and artefacts by comparing the shape of the loadings extracted from the chromatographic regions previously identified as selective [31].

Our approach uses the ideas of correction and diagnosis but in the opposite direction. In this paper, we study in detail, the effects that small impurities and known major non-idealities have on the results of singular value decompositions and of window evolving factor analyses. We use this knowledge to construct a realistic simulation scheme that generates an artificial spectrochromatogram including non-idealities which could have been observed if the substance had been pure. The realistic simulation can then be compared to the actual spectrochromatogram for purity control. We demonstrate this with a real example.

## 2. Theory

### 2.1. Singular value decomposition

A matrix **D** containing bilinear data from a LC–DAD system can be expressed as:

$$\mathbf{D} = \mathbf{XCY}^{T} + \mathbf{E} \quad (1)$$

where **X** contains the pure chromatographic profiles at unit concentrations, **Y** contains the pure spectra at unit concentrations, the diagonal matrix **C** contains the concentration values of each compound and **E** contains the noise. In mixture analysis techniques based on factor analysis, the measurement space is described by a small number of orthogonal base vectors. These base vectors can be calculated by applying a singular value decomposition (SVD) to the data matrix. According to SVD any data matrix with $r$ rows and $c$ columns can be decomposed such that:

$$\mathbf{D} = \mathbf{USV}^{T} \quad (2)$$

Assuming that $c$ is less or equal than $r$, **U** is an $rxc$ orthonormal matrix with base vectors of the column space also called loadings, **S** is a $cxc$ diagonal matrix with the singular values and **V** is a $cxc$ orthonormal matrix with the base vectors of the row space also called scores. The base vectors in **U** and **V** can also be calculated as the eigenvectors of $\mathbf{DD}^{T}$ and $\mathbf{D}^{T}\mathbf{D}$.

### 2.2. Malinowski's F-test

Malinowski has introduced an $F$-test for the determination of the number of primary factors [32]. In this test the variance of a target vector is compared with the variance of the noise-describing eigenvectors (the "null" vectors). Malinowski's statistic uses "reduced eigenvalues" which are defined as the eigenvalues divided by the degrees of freedom. The $F$-test applied is given in Eq. (3):

$$F(1,s-n) = \frac{\lambda_n/(r-n+1)(c-n+1)}{\sum_{j=n+1}^{s} \lambda_j^0 / \sum_{j=n+1}^{s} (r-j+1)(c-j+1)} \quad (3)$$

where $r$ and $c$ are respectively the number of rows and columns in the investigated matrix; $s$ is the minimum of $r$ and $c$; $n$ is the number of the target vector; $\lambda$ is the corresponding eigenvalue; and $\lambda^0$ is the error ("null ") eigenvalue.

### 2.3. Window evolving factor analysis (WEFA)

Window evolving factor analysis [15,33] is a relatively recent and powerful technique for checking peak purity in LC–DAD, derived from the evolving factor analysis (EFA) method proposed by Maeder [3]. In WEFA, one performs a series of principal components analyses on a moving subset of the original data matrix containing a fixed number of spectra. Each successive window of spectra is reduced to a set of eigenvalues whose logarithms are plotted as a function of elution time. The number of eigenvalues is equal to the number of spectra in the window. When the chromatogram is on the baseline, all the eigenvalues are about the same magnitude and small. When a single component peak elutes, one eigenvalue grows large while the others remain small and about the same size. In the case of a two component peak, two eigenvalues will rise above the rest, and so on.

## 3. Experimental

### 3.1. Apparatus

The data evaluated for this article were acquired using a Beckman Gold LC–DAD system (Beckman Instruments, Fullerton, CA, USA) equipped with a fixed 4-nm optical slit. The spectra covering a wavelength range of 200 to 350 nm were recorded at time intervals of 1 s with a digital resolution of 1 nm. In order to evaluate the influence of the scan time of the diode array, scan rates of 32, 64 and 96 Hz were used. After collection, the data files were converted to ASCII files using the GRAMS-Convert software (Galactic Inc., Salem, NH, USA).

### 3.2. Reagents and samples

HPLC-grade acetonitrile was purchased from Labscan (Dublin, Ireland) and water was obtained from

an in-house water purification system (Milli-Q, Millipore, Milford, MA, USA). For the study of homogeneous peaks, the samples consisted of alprazolam (Upjohn, Puurs, Belgium) or hydrochlorthiazide (Ciba-Geigy, Brussels, Belgium) dissolved in the corresponding mobile phase. The mixtures consisted of alprazolam dissolved in the mobile phase where triazolam (Upjohn) was added in different concentrations. The concentration was adjusted to give a maximum absorbance of about 0.08 AU to observe a bilinear behaviour, and of about 0.5 AU to study the artefacts.

### 3.3. Procedure

Chromatographic runs were carried out under isocratic conditions and at ambient temperature on a 250×4 mm I.D. RP-18 column (Merck, Lichrospher RP-select B, 5 μm). For the alprazolam–triazolam pair the mobile phase consisted of 45–60% acetonitrile–55–40% water (v/v) and permitted selection of the the chromatographic resolution between the two analytes as desired in the range of $R_s = 0.1$ to 1.0. For hydrochlorthiazide, the mobile phase consisted of acetonitrile–water (25:75, v/v). The flow-rate was 0.5, 1 or 2 ml $min^{-1}$.

## 4. Results and discussion

The spectra of the analytes investigated in this study are given in Fig. 1a–c. The difference of the spectra of the two closely related benzodiazepines alprazolam and triazolam used as a model mixture, is also displayed (Fig. 1c).

### 4.1. Ideal bilinear behaviour

Creating "real" bilinear behaviour is not trivial. As pointed out by Excoffier et al. [30] DAD response can be non-linear at absorbances as low as 100 mAU under normal chromatographic conditions. The absorbance ceiling to apply in order to avoid a non-ideal behaviour will depend on the detector's characteristics, on the shape of the spectrum of the analyte in the spectral range investigated [29] and on the nature of the mobile phase used, since the absorbance that is critical in this regard is the total absorbance which includes eluent absorbance even if balanced [27].

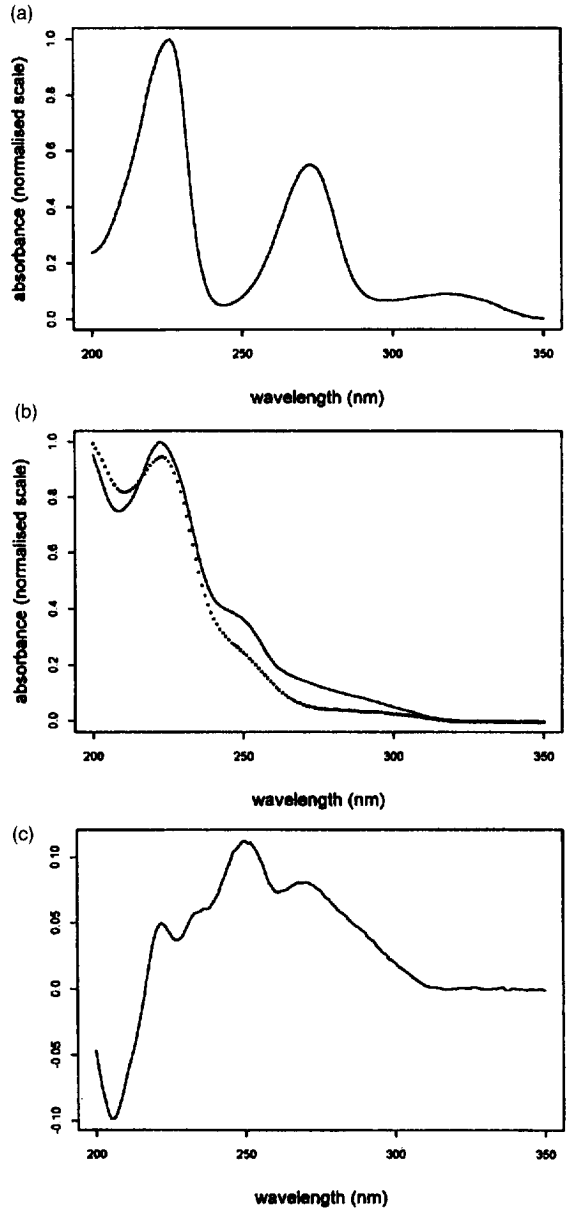


Fig. 1. Normalised spectra of (a) hydrochlorthiazide, (b) alprazolam (solid line) and triazolam (dotted line), (correlation coefficient = 0.991) and (c) difference between the spectra displayed in (b).

In order to facilitate the observation of a bilinear behaviour, alprazolam, a drug compound showing a

relatively smooth spectrum in the spectral range investigated (200–350 nm), was chosen (Fig. 1b). The pure and impure samples with a concentration adjusted to give a maximum absorbance of about 100 mAU were eluted in presence of a mobile phase highly transparent in the wavelength range used and consisting of a mixture acetonitrile–water. Fig. 2a shows the loading and score profiles of the first three eigenvectors calculated for a pure peak of alprazolam measured in these conditions. Characteristically, the scores of the first eigenvector display the shape of the chromatographic profile of the pure compound, while the corresponding loadings adopt the shape of its spectrum. The scores and loadings of the higher order eigenvectors show essentially a random fluctuation and account for the noise showing that as expected, the effect of the non-linearities was sufficiently low in these conditions to be masked by the noise and to remain unnoticed. If an impurity coelutes closely (Fig. 2b and c), the first eigenvector is an average of the mixture spectra (loadings) or elution profiles (scores), while the second one reflects the difference between the various elements and can be interpreted as the result of a subtraction of the two pure component spectra (compare Fig. 1c with the second loadings profile in Fig. 2b and c), or profiles. In the case of an unresolved minor compound in the presence of a major one, the scores of the second eigenvector will then adopt an intermediate profile between the shape of a peak (for high values of $R_s$, Fig. 2b) and the first derivative of a peak (for lower values of $R_s$, Fig. 2c).

Note that with these experimental conditions and after background correction as explained further, the number of primary factors was properly estimated by the Malinowski's $F$-test (Table 1) and a correct peak purity assignment was obtained in WEFA (Fig. 2a–c).

### 4.2. Baseline effect

As pointed out by Maeder and Zilan [34], sloping baselines and constant baselines above or below zero resemble additional chemical components (fourth column in Table 1). Fig. 3 shows the loadings and scores profiles of the first three eigenvectors calculated in case of a non-zero baseline (Fig. 3a) and a sloping baseline (Fig. 3b). Fig. 3a corresponds to the same data set as in Fig. 2a but prior to background correction. Since the effect of a non-ideal baseline is more pronounced where the signal to noise ratio (SNR) is low, it gives typically an eigenvector characterised by a scores profile parallel to the baseline below the edges of the peak, and to the peak shape under its central part. In the case of zero offset of the baseline, the corresponding loadings show essentially a random fluctuation (Fig. 3a). On the other hand, the loadings associated with a sloping baseline have frequently an increasing contribution in the low UV (Fig. 3b), as under normal conditions, a baseline drift is usually caused by inadequate equalisation of the working conditions in the column and/or the detector (mobile phase composition and flow-rate, temperature), and yields therefore relatively larger spectral changes in the low UV.

Assuming that the chromatographic baseline is increasing or decreasing linearly with retention time, an appropriate background correction can be obtained by subtraction of a linear interpolation between the baseline average spectra calculated before and after the peak [22]. In the remainder of this article this type of correction is always applied.

### 4.3. Other artefacts

Fig. 4 shows the loadings and score profiles of the first five eigenvectors for homogeneous peaks of hydrochlorthiazide (Fig. 4a) and alprazolam (Fig. 4b) with a maximal absorbance of about 0.5 AU. Clearly, extra smooth eigenvectors appear, revealing the presence of non-idealities. Consequently, extra primary factors are detected by Malinowski's $F$-test (first column of Tables 2 and 3) and an incorrect peak purity assignment is obtained in WEFA. The relative amounts of the first six singular values are given in the first column of Tables 2 and 3. From the characteristics of the detector used, several sources of artefacts may be considered. In the following, the role of these possible candidates is evaluated.

#### 4.3.1. Polychromatic radiation

The need of a wider wavelength bandpass to optimise the signal-to-noise ratio in DADs explains the more limited linearity of these detectors in comparison with photomultiplier spectrophotometers

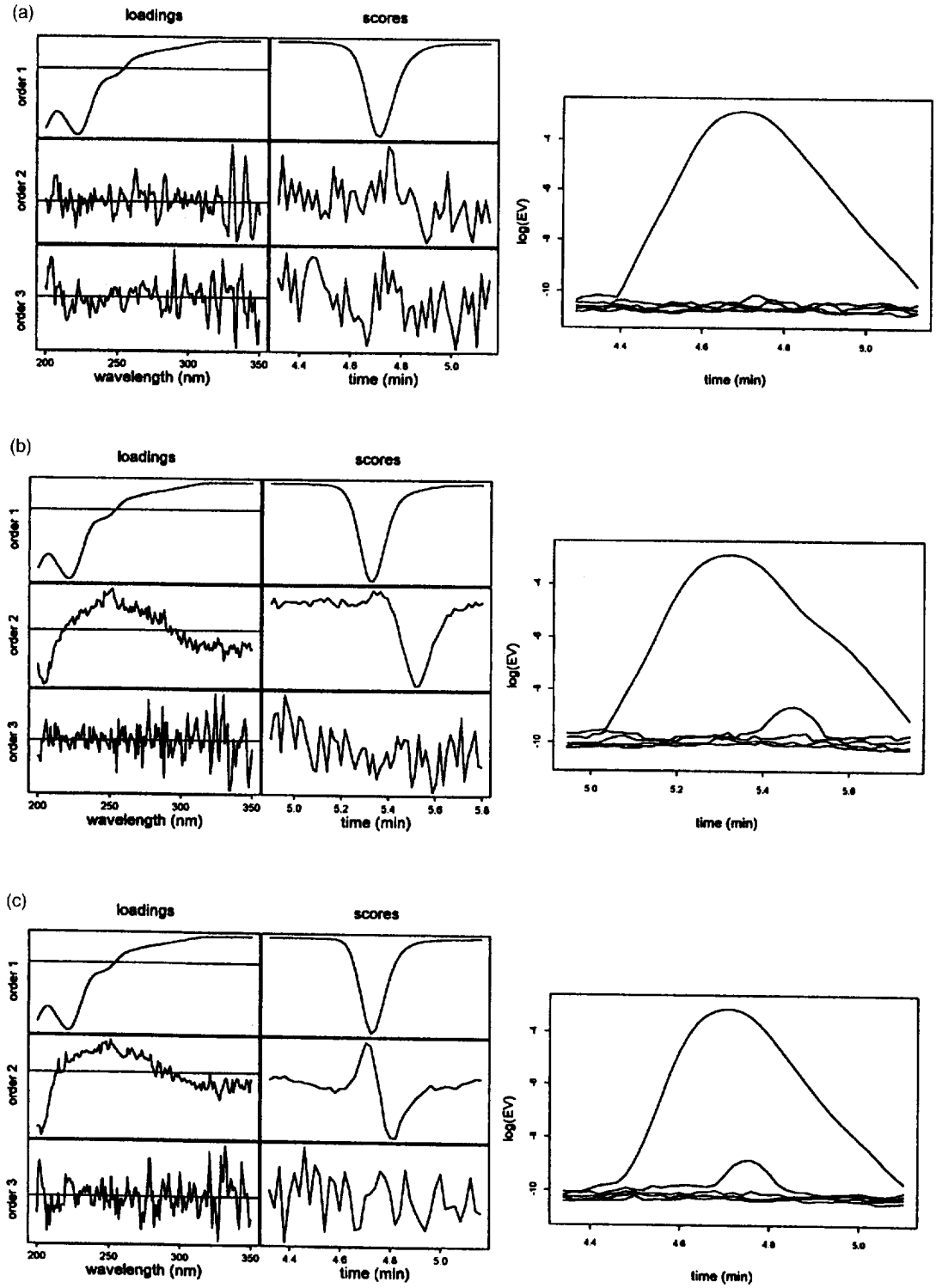


Fig. 2. Structure of the first three eigenvectors and corresponding WEFA plot (limited to the first 5 EVs) obtained for (a) an actual homogeneous alprazolam peak, (b) an actual heterogeneous alprazolam peak containing 2% of impurity with a $R_s = 0.7$ and (c) an actual heterogeneous alprazolam peak containing 2% of impurity with a $R_s = 0.3$; all with a $A_{max}$ of about 0.1 AU.

Table 1
Estimated number of primary factors by Malinowski's $F$-test ($\alpha = 5\%$) for the data of Fig. 2a–c and Fig. 3a, respectively

| With background correction | | | Without |
|---|---|---|---|
| Purity | | | |
| 100% | 98% ($R_s = 0.7$) | 98% ($R_s = 0.3$) | 100% |
| 1 | 2 | 2 | 2 |

[35]. According to Sievert and Drouen, polychromatic radiation resulting from the width of the entrance slit of the detector may be regarded as the major potential source of distortion of spectral data [36]. Dose and Guiochon modeled the DAD response as a function of the bandpass and on the shape of the absorption spectrum of the sample [29]. According to these authors, the optical band pass of a DAD is nearly rectangular and leads to the detection of a transmittance-averaged radiant power across the face of each diode [29]. The optical system behaves therefore like a boxcar filter whose width depends on the entrance slit:

$$T\lambda = \frac{1}{\Delta} \sum_{j=\lambda-\left(\frac{\Delta-1}{2}\right)}^{\lambda+\left(\frac{\Delta-1}{2}\right)} T_j^0 \qquad (4)$$

with $T_\lambda$ = transmittance measured at the nominal wavelength $\lambda$, $T_j^0$ = true transmittance value at the wavelength $j$, $\Delta$ = full width of the optical band pass.

Thus, the measured intensity does not follow the Beer–Lambert law, because the average absorbance is not the logarithm of the average intensity. Since the optical slit of the detector used is equal to 4 nm, the full width of the optical bandpass ($\Delta$) should involve five consecutive wavelengths. In practice, the true bandpass may be wider than the nominal optical slit since the resolution achieved depends not only on the optical slit but also on other characteristics of the detector such as the grating and the array resolution [36,37].

In principle, the signal could be restored by appropriately deconvolving each transmittance spectrum prior to analysis. In practice however, this turns out to be more difficult than expected the effective bandpass width has to be estimated, and because the measurement noise requires careful treatment. Our experiments with such corrections on actual spectrochromatograms led to rather disappointing results. Although hard to remove, the effect of the optical slit can be readily simulated. This can be achieved by extracting the apex spectrum and the apex chromatogram from the data, deconvolving the spectrum, recombining the deconvoluted spectrum and the chromatogram into an artificial spectrochromatogram and finally reconvolving each spectrum by transmittance averaging according to Eq. (4). In our situation, a bandpass involving seven wavelengths was found to give a good agreement between actual spectrochromatograms and data simulated according to the

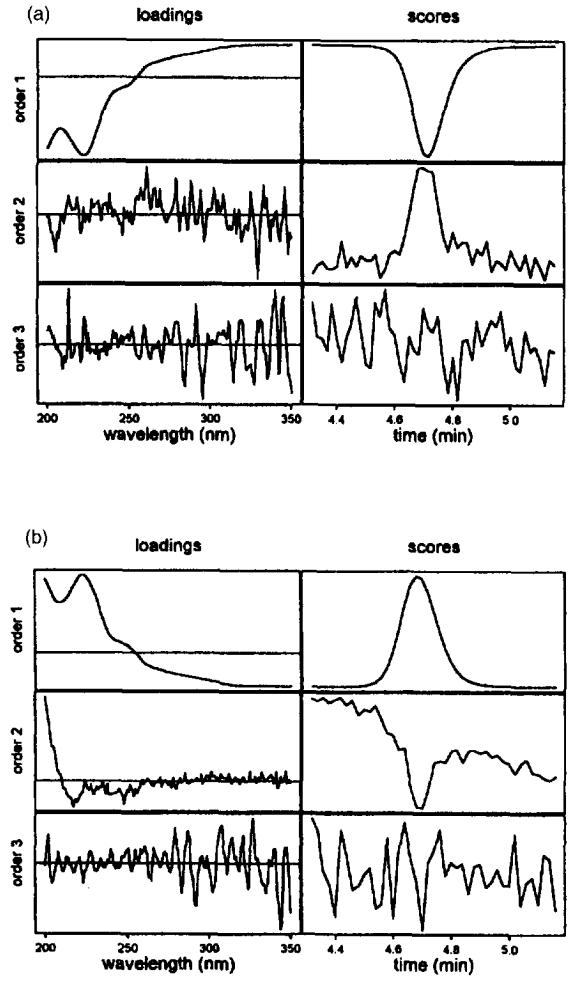


Fig. 3. Structure of the first three eigenvectors computed without background correction for (a) the data of Fig. 2a, (b) an actual homogeneous alprazolam peak with insufficient equilibration of the system after a change in the mobile phase composition.

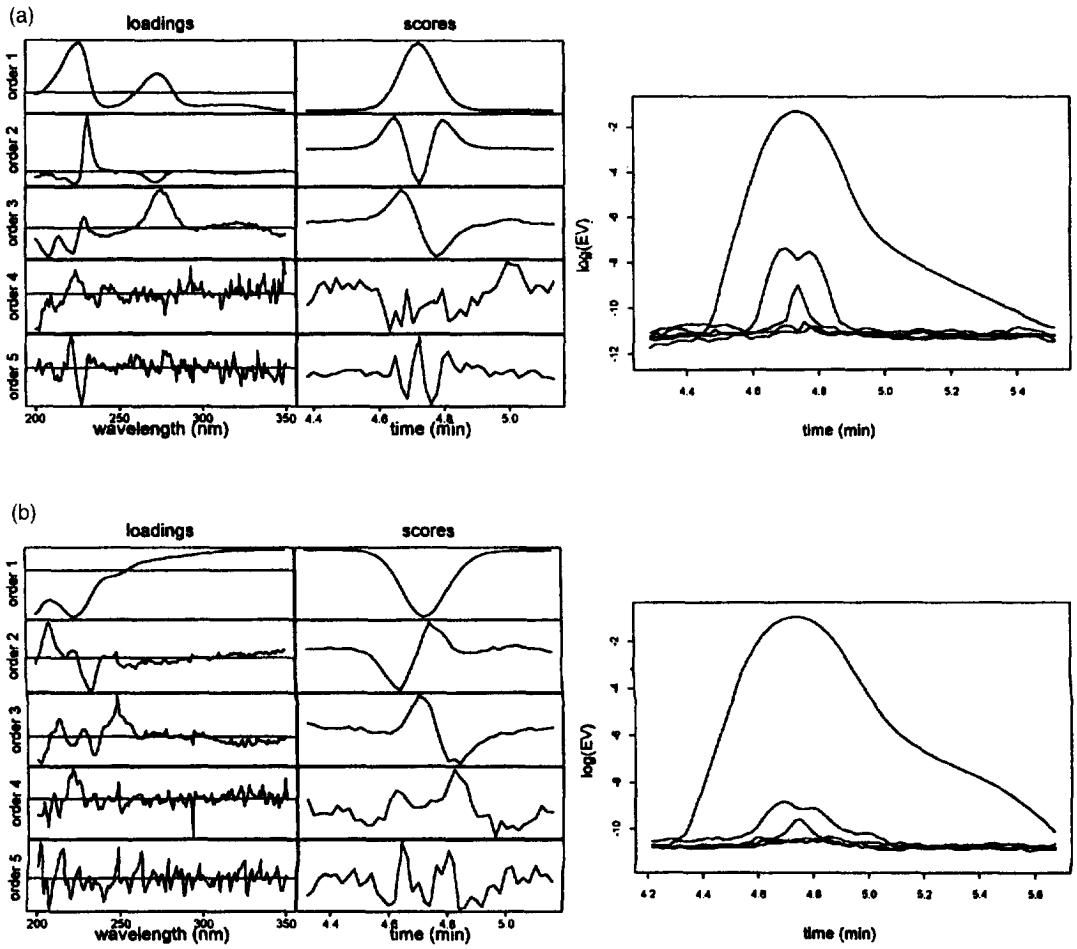


Fig. 4. Structure of the first five eigenvectors and corresponding WEFA plot obtained for (a) an actual homogeneous hydrochlorthiazide peak, and (b) an actual homogeneous alprazolam peak; both with a $A_{max}$ of about 0.5 AU.

deconvolution–reconvolution procedure. Gaussian distributed noise was subsequently superimposed to the simulated data. Since the noise level for a given instrument is not fixed but varies with the characteristics of the detector (e.g., age of the lamp, scan rate selected), the level of the noise superimposed to the theoretical data was defined by an estimation of the baseline noise of the actual data to be simulated. The so obtained semi-artificial data matrix can then be subjected to the same analyses as the actual spectrochromatograms. A comparison of the results can help to distinguish real impurity effects from optical slit effects. Note that this approach is more qualitative than the direct correction approach. In our experience, it suffers less from incomplete knowledge about the real band width and the noise structure than direct correction.

Fig. 5a and b show the structure of the first five eigenvectors calculated for simulated data sets corresponding to the actual data of Fig. 4a and b. The relative amounts of the first six singular values computed for the simulated data sets are given in the first column of Tables 4 and 5. In order to carry out relevant comparisons, the same size of the data matrix was used in SVD for both the real and the simulated data since both the eigenvalues corresponding to noise and the imbedded errors in the primary factors as defined by Malinowski [32] are a

Table 2
Relative amounts of the first six singular values (in %) and estimated number of primary factors by Malinowski's *F*-test ($\alpha = 5\%$)

| Hydrochlorthiazide | | | | |
|---|---|---|---|---|
| *Flow* | | | | |
| 1 (ml min$^{-1}$) | | | 0.5 (ml min$^{-1}$) | 2 (ml min$^{-1}$) |
| *Scan rate* | | | | |
| 32 Hz | 64 Hz | 96 Hz | 32 Hz | 32 Hz |
| 99.493 | 99.471 | 99.437 | 99.461 | 99.514 |
| 0.281 | 0.281 | 0.278 | 0.291 | 0.274 |
| 0.066 | 0.051 | 0.043 | 0.041 | 0.095 |
| 0.013 | 0.015 | 0.018 | 0.013 | 0.015 |
| 0.010 | 0.010 | 0.013 | 0.009 | 0.012 |
| 0.009 | 0.010 | 0.011 | 0.009 | 0.009 |
| *Malinowski's F-test* | | | | |
| 3 | 3 | 3 | 3 | 3 |

The relative amounts were calculated for actual homogeneous peaks of hydrochlorthiazide with a $A_{max}$ of about 0.5 AU, when the scan rate of the diode array or the flow is changed.
The first column corresponds to the data of Fig. 4a.

function of the size of the data matrix, as reported by Liang et al. [19].

Polychromatic radiation is responsible for a non-linear calibration curve which introduces a second derivative of peak pattern into the second scores vector. This pattern is clearly recognisable for the hydrochlorthiazide actual data set (Fig. 4a); moreover, the corresponding loadings display remarkable similarities with the simulated results (compare Fig. 4a and Fig. 5a). In the case of alprazolam however, the expected scores profile is not observed for the actual data set even though the loadings of the second eigenvector display a certain resemblance to the simulated one. This can be explained by assuming that a particular extra eigenvector may account for more than one non-idealities source at the same time. Indeed, if several artefacts are about the same magnitude, their contributions in SVD may be merged so that the corresponding extra eigenvectors will display an intermediate shape between the profiles induced by each artefact taken separately. For a compound whose spectrum shows a high slope/absorbance ratio such as hydrochlorthiazide, one may expect that the effect of polychromatic radiation will dominate the other non-idealities sources so that its contribution in SVD will be easily recognisable (first column of Table 4). On the other hand, if the spectrum is smoother, such as for alprazolam, the expected contribution of polychromatic radiation is lower (first column of Table 5) and can merge with the effect of other non-idealities sources as shown further.

Table 3
Relative amounts of the first six singular values (in %) and estimated number of primary factors by Malinowski's *F*-test ($\alpha = 5\%$)

| Alprazolam | | | | |
|---|---|---|---|---|
| *Flow* | | | | |
| 1 (ml min$^{-1}$) | | | 0.5 (ml min$^{-1}$) | 2 (ml min$^{-1}$) |
| *Scan rate* | | | | |
| 32 Hz | 64 Hz | 96 Hz | 32 Hz | 32 Hz |
| 99.722 | 99.652 | 99.610 | 99.725 | 99.789 |
| 0.043 | 0.041 | 0.041 | 0.043 | 0.041 |
| 0.031 | 0.024 | 0.021 | 0.028 | 0.039 |
| 0.010 | 0.011 | 0.015 | 0.011 | 0.014 |
| 0.008 | 0.010 | 0.014 | 0.009 | 0.010 |
| 0.007 | 0.010 | 0.012 | 0.007 | 0.009 |
| *Malinowski's F-test* | | | | |
| 3 | 3 | 2 | 4 | 3 |

The relative amounts were calculated for actual homogeneous peaks of alprazolam with a $A_{max}$ of about 0.5 AU, when the scan rate of the diode array or the flow is changed.
The first column corresponds to the data of Fig. 4b.

Note that the stray light will also cause deviations from linearity and its effect will be superimposed on that of polychromatic radiation. Under normal conditions, stray light is only of great concern at higher absorbance levels, typically at absorbances greater than 1.5 AU [30], but its contribution may also explain the need of a bandpass larger than the nominal optical slit in simulation.

### *4.3.2. Scan time*

Since the readout of the array sensor chip is a sequential process, the first diode in the array is read out at a different time than the last one. Consequently, non-idealities may arise if the peak absorbance changes over the period of the scan of the diode array [38]. In this case, the end of spectra will be skewed higher with respect to the beginning in the upslope of the peak, and lower in the downslope. This effect can be studied by changing the scan time or the flow-rate directly. In our detector, the array scan rate can be set to a multiple of 16 Hz between

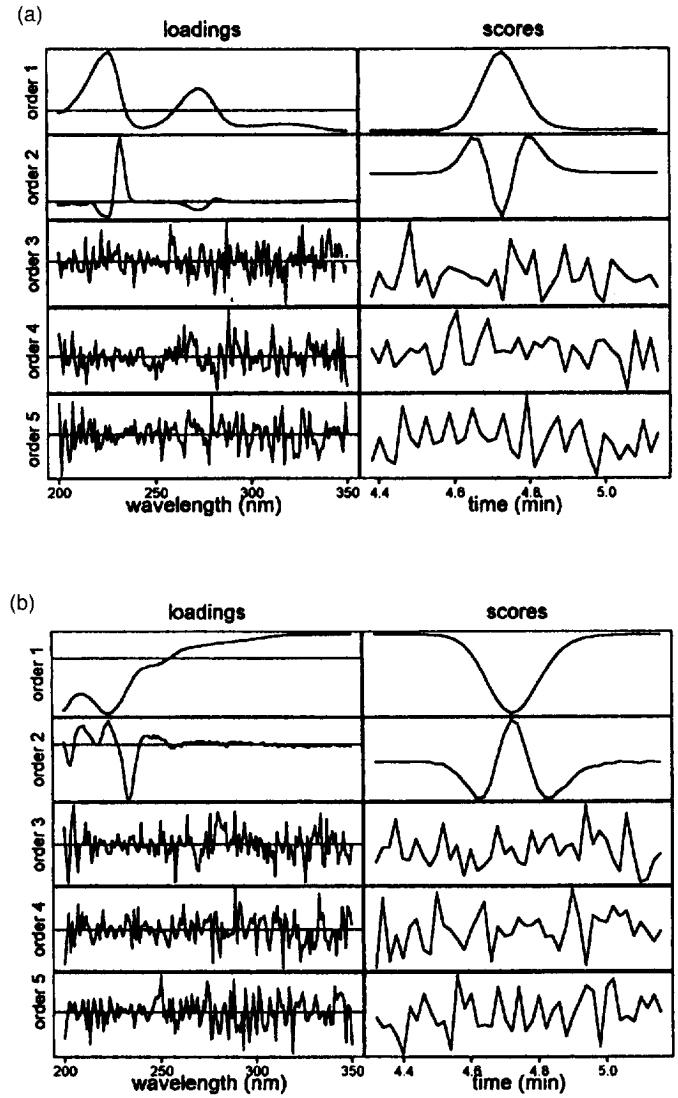


Fig. 5. Structure of the first five eigenvectors computed for simulated data, with polychromatic radiation effect introduced according to Eq. (4), and corresponding to the actual data of (a) Fig. 4a and (b) Fig. 4b.

16 and 112 Hz. However, unless over-ridden by the user, an algorithm executed automatically sets the scan rate to the slowest rate (multiple of 16 Hz) which avoids array saturation. This rate therefore depends on the lamp intensity and on the mobile phase transmittance. In the working conditions used here, a scan rate of 32 Hz (scan time of 31.25 ms) was automatically selected. The effect of changing the scan time or the flow on the relative amounts of the first six singular values is shown in Tables 2 and 3 for hydrochlorthiazide and alprazolam, respectively. In both cases, the third singular value decreases when the scan rate increases or the flow decreases, suggesting that the corresponding extra factor mostly accounts for the non-idealities induced by the scan-time. Moreover, when the scan rate increases, one

Table 4
Relative amounts of the first six singular values (in %) and estimated number of primary factors by Malinowski's $F$-test ($\alpha = 5\%$)

| Hydrochlorthiazide | | | |
|---|---|---|---|
| Polychromatic radiation | +Scan time | +Transmittance averaging error | +Heteroscedasticity |
| 99.580 | 99.528 | 99.515 | 99.498 |
| 0.278 | 0.277 | 0.278 | 0.278 |
| 0.007 | 0.059 | 0.063 | 0.063 |
| 0.007 | 0.007 | 0.015 | 0.015 |
| 0.006 | 0.006 | 0.006 | 0.009 |
| 0.006 | 0.006 | 0.006 | 0.008 |
| *Malinowski's F-test* | | | |
| 2 | 3 | 3 | 3 |

The relative amounts were calculated for simulated homogeneous peaks of hydrochlorthiazide corresponding of the data of Fig. 6a, when the various instrumental errors are successively added according to Eqs. (4)–(7).
The various columns correspond to the data of Figs. 5a, 6a, 7a, and 8a, respectively.

observes also a slight decrease in the first singular value and an increase in the last three values; this will be explained further.

The scan time effect can be introduced in simulated data according to the following equation [23]:

$$A^*_{t,\lambda} = A_{t,\lambda} - (A_{t,\lambda} - A_{t-\Delta t,\lambda}) \cdot \left(\frac{p-1}{n-1}\right) \cdot \left(\frac{t_{\text{scan}}}{\Delta t}\right) \tag{5}$$

with $A^*_{t,\lambda}$ = the distorted absorbance at time $t$ and wavelength $\lambda$, $p$ = the position of the corresponding wavelength in the spectral range investigated, $n$ = the number of wavelengths covered by the array sensor chip, $t_{\text{scan}}$ = the scan time, $\Delta t$ = the sampling interval.

The array sensor chip of the DAD used in this study consists of 512 diodes covering the range 190 to 600 nm so that $n$ equals 410. The spectra were recorded from 200 to 350 nm thus, $p$ is in the range 1 to 151. The scan time was 31.25 ms and the sampling interval was 1 s. Fig. 6a and b show the structure of the first five eigenvectors calculated when the scan rate effect is superimposed to the simulated data of Fig. 5a and b and the corre-

Table 5
Relative amounts of the first six singular values (in %) and estimated number of primary factors by Malinowski's $F$-test ($\alpha = 5\%$)

| Alprazolam | | | |
|---|---|---|---|
| Polychromatic radiation | +Scan time | +Transmittance averaging error | +Heteroscedasticity |
| 99.799 | 99.780 | 99.773 | 99.738 |
| 0.038 | 0.039 | 0.043 | 0.045 |
| 0.006 | 0.024 | 0.024 | 0.026 |
| 0.006 | 0.006 | 0.010 | 0.010 |
| 0.006 | 0.006 | 0.006 | 0.009 |
| 0.005 | 0.006 | 0.006 | 0.008 |
| *Malinowski's F-test* | | | |
| 2 | 3 | 3 | 3 |

The relative amounts were calculated for simulated homogeneous peaks of alprazolam corresponding to the data of Fig. 6b, when the various instrumental errors are successively added according to Eqs. (4)–(7).
The various columns correspond to the data of Figs. 5b, 6b, 7b and 8b, respectively.

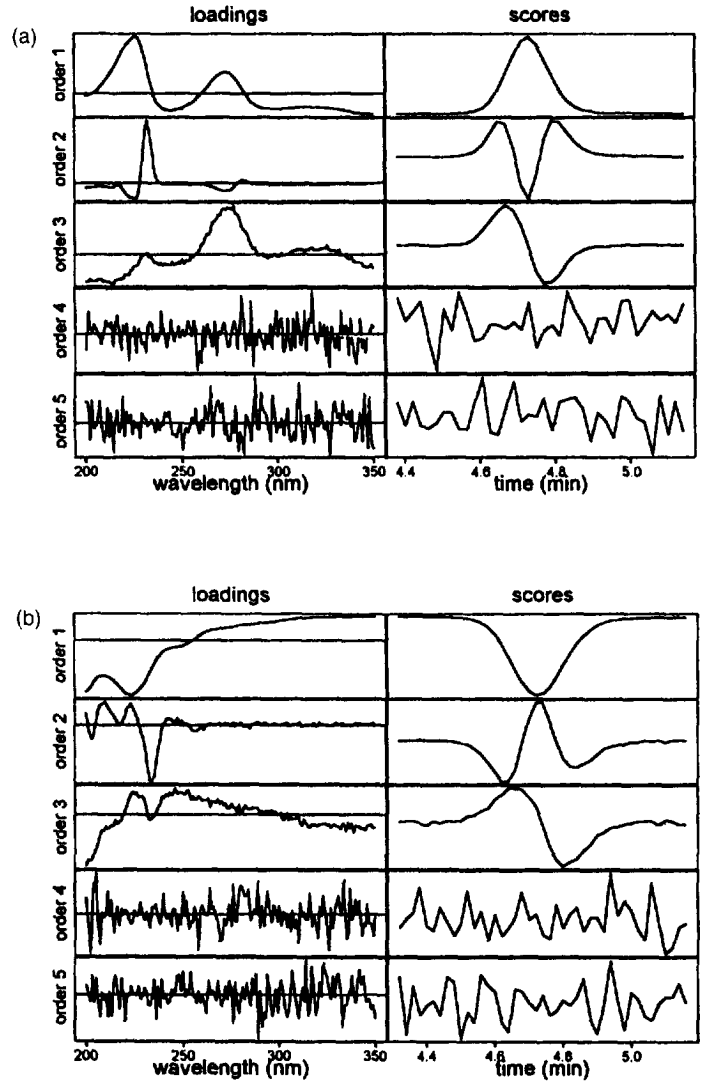


Fig. 6. Structure of the first five eigenvectors computed after addition of the scan time effect according to Eq. (5), to the simulated data of (a) Fig. 5a and (b) Fig. 5b.

sponding relative amounts of the first six singular values are given in the second column of Tables 4 and 5. The comparison with the results obtained for the actual data sets (Fig. 4a,b and first column of Tables 2 and 3) indicate an increased similarity between simulated and real data structure. In Fig. 6a and b, one can see that the scores of the extra eigenvector induced by the scan time effect, display a first derivative of peak shape, reflecting the difference of the spectral skewing between the upslope and the downslope of the peak. Hence, such a feature can be misinterpreted as the presence of an impurity (Fig. 2c).

#### *4.3.3. Transmittance averaging in the time domain*

DADs generally average several scans in transmittance before converting them into absorbances and storing them [39]. Transmittance, rather than absorbance averaging, over a time period will lead to

non-idealities if a concentration change has occurred. In the detector used in this study, absorbance is computed at 16 Hz regardless of the actual scan rate. This means that spectra collected at a faster rate are averaged over 62.5 ms intervals before a logarithm is computed. Successive absorbance scans are then averaged to obtain spectra corresponding to the sampling rate chosen (1 Hz here).

The transmittance averaging error was introduced according to:

$$A^{*}_{\frac{1}{r}\sum_{i=1}^{r} ti,\lambda} = -\log\left(\frac{\sum_{i=1}^{r} T_{ti,\lambda}}{r}\right) \quad (6)$$

with $T_{ti,\lambda}$ = transmittance at time $ti$ and wavelength $\lambda$, $r$ = the scan rate in Hz divided by 16.

The number of data points was first multiplied by $r$ through linear interpolation between the measured data and the basic noise level was accordingly multiplied by $\sqrt{r}$.

Fig. 7a and b show the structure of the first five

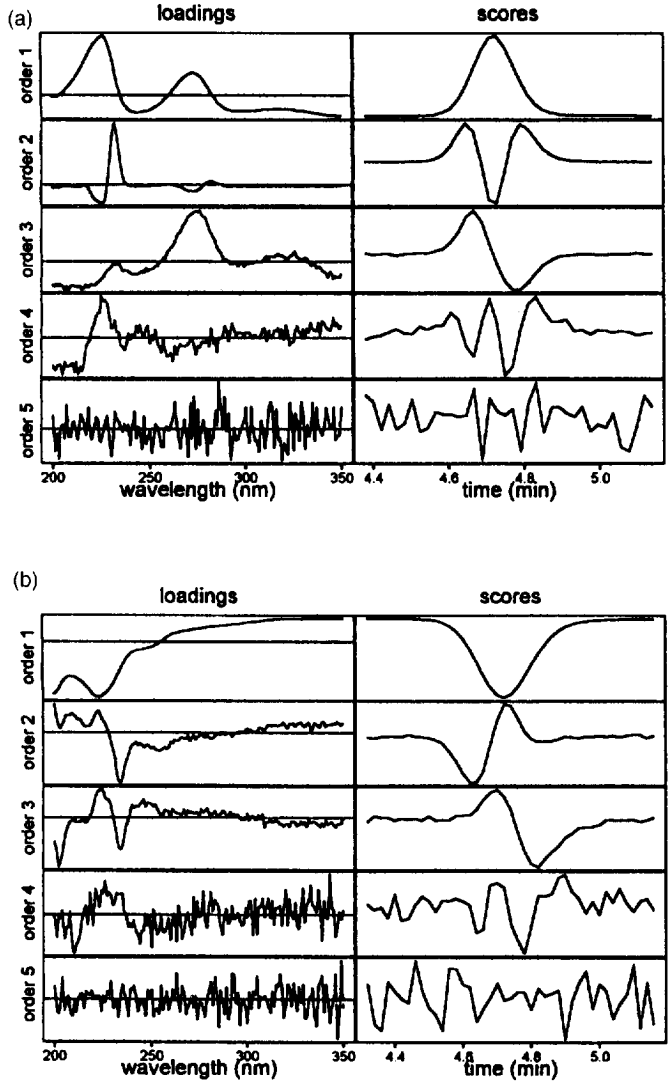


Fig. 7. Structure of the first five eigenvectors computed after addition of the transmittance averaging effect according to Eq. (6), to simulated data corresponding to those of (a) Fig. 6a and (b) Fig. 6b.

eigenvectors calculated when the transmittance averaging error is superimposed to the other non-idealities sources and the corresponding relative amounts of the first six singular values are given in the third column of Tables 4 and 5. Again the similarity between simulated and actual data is increased by the addition of this new artefact, in particular, one can notice that in the case of alprazolam, the shape of second derivative of peak for the scores of the second factor is no longer observable as in the real case. The examination of Tables 4 and 5, indicates that the transmittance averaging effect especially influences the fourth factor, this is in agreement with the increase of the relative amount of the fourth singular value observed with a faster scan rate in Tables 2 and 3. However, in these tables, the relative amounts of the fifth and sixth factors, which seem to account essentially for the noise, were also found to be larger for a faster scan rate suggesting that the noise level increases with the scan rate even if a larger number of scan are transmittance averaged. As pointed out by Jones [39], on-chip integration of the photocurrent improves the SNR proportionally to the exposure time whereas averaging the signal in memory improves the SNR only in proportion to the square root of the accumulated exposure time.

#### 4.3.4. Heteroscedasticity

Heteroscedastic noise means a measurement variance which changes with the magnitude of the signal. In UV-visible spectroscopy, it is not surprising that the noise increases with absorbance since increased absorbance is related to decreased light throughput. Proportional heteroscedastic noise can be superimposed to the noise-free simulated spectrochromatogram by the following relation [25]:

$$s_{t,\lambda} = s_0(1 + \alpha A_{t,\lambda})\xi_{t,\lambda} \quad (7)$$

with $s_{t,\lambda}$ = standard deviation of the noise at time $t$ and wavelength $\lambda$, $s_0$ = standard deviation of the baseline noise, $A_{t,\lambda}$ = signal measured at time $t$ and wavelength $\lambda$, $\alpha$ = proportionality factor and $\xi_{t,\lambda}$ = standard normal random number.

If $\alpha$ is equal to zero, homoscedastic data are simulated. For the LC–DAD system used here, an $\alpha$ of 7 was found to provide good agreement between simulated and actual heteroscedastic data [25].

Fig. 8a and b show the structure of the first five eigenvectors calculated when heteroscedastic noise is superimposed on to the other non-idealities sources. The relative amounts of the first six singular values are given in the fourth column of Tables 5 and 6. Comparison of these results with those computed for the actual data sets now indicates even stronger similarities between actual and simulated data. We can say that except for a few fine structures, the actual and the reconstructed data are equivalent.

This allows us to conclude that the four non-idealities sources invoked are sufficient to mimic most of the distortion observed in the data. Whether they can actually be identified as definite sources is more difficult to assess, but the high level of similarity achieved between real and simulated data provides strong evidence that they are effectively responsible for most of the artefacts observed. As we stated earlier, our results suggest that direct correction of spectrochromatograms, i.e., eliminating most of these artefacts, may be possible, but our latest results have been disappointing in this direction.

The procedure described above allows realistic simulations of LC–DAD data sets throughout the absorbance range generally used (Table 6). Similar results were obtained for other drug compounds well characterised for their impurity pattern.

### 4.4. Practical application

Realistic simulations can be very useful in peak purity control since they permit working at much higher than usual maximum absorbencies and thus at better signal-to-noise ratios. As an example, Fig. 9a shows the first five score and loading vectors and the WEFA traces of a spectrochromatogram of a real peak of alprazolam at a maximum absorbency of 500 mAU. The corresponding singular values and the results of Malinowski's $F$-test are presented in the first column of Table 7. Is this peak pure or not?

Clearly, there is more than one primary factor and also the WEFA traces show clear signals. However, since the maximum absorbency is quite high for DAD detection, we cannot be sure if the observed patterns are due to impurities or to artefacts. In this context, we can use realistic simulation. For this

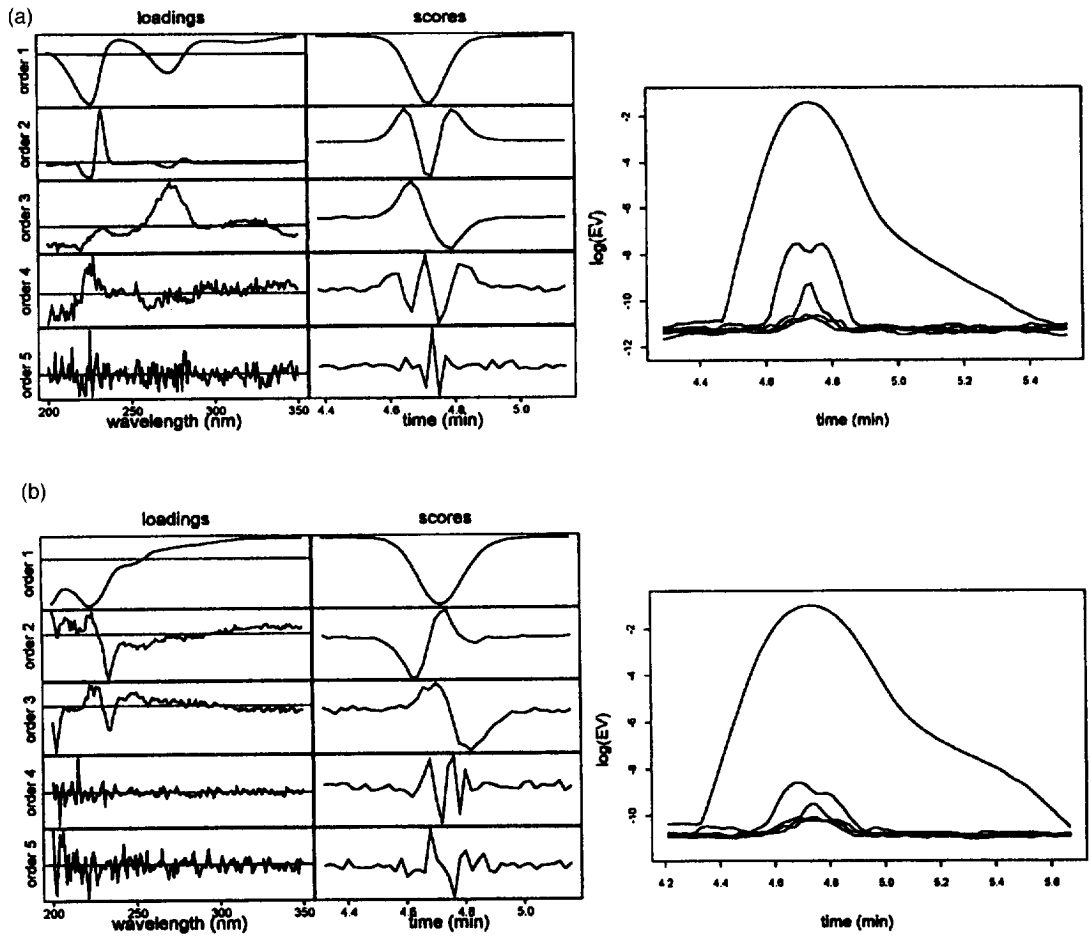


Fig. 8. Structure of the first five eigenvectors and corresponding WEFA plot obtained after addition of heteroscedastic noise according to Eq. (7), to simulated data corresponding to those of (a) Fig. 7a and (b) Fig. 7b.

Table 6
Comparison of the relative amounts of the first six singular values (in %)

| Hydrochlorthiazide | | | | | |
|---|---|---|---|---|---|
| $A_{max}$ = 0.1 AU | | $A_{max}$ = 0.3 AU | | $A_{max}$ = 0.8 AU | |
| Real | Simulated | Real | Simulated | Real | Simulated |
| 99.007 | 99.088 | 99.481 | 99.500 | 99.317 | 99.373 |
| 0.067 | 0.069 | 0.156 | 0.162 | 0.0434 | 0.427 |
| 0.058 | 0.058 | 0.077 | 0.063 | 0.067 | 0.065 |
| 0.044 | 0.036 | 0.017 | 0.015 | 0.019 | 0.021 |
| 0.040 | 0.035 | 0.016 | 0.014 | 0.008 | 0.008 |
| 0.039 | 0.034 | 0.014 | 0.012 | 0.007 | 0.007 |

The relative amounts were calculated for actual and simulated homogeneous peaks of hydrochlorthiazide with an increasing value of $A_{max}$.

purpose, we extract the apex spectrum, the $\lambda_{max}$ chromatogram, and a noise estimate at baseline. Using our system knowledge, i.e., width of optical band path, scan rate, transmittance averaging rate, heteroscedasticity model, etc., we can thus simulate what the spectrochromatogram would have looked like if a single chemical species with the (deconvolved) apex spectrum and the $\lambda_{max}$ chromatogram had eluted. This simulated spectrochromatogram can then be subjected to the same analysis as the real one. Assuming that no major source of artefact has been left out, differences in the analysis of the real and the simulated spectrochromatogram are then suggesting the presence of an impurity. Fig. 9b shows the result for our example. We see that there

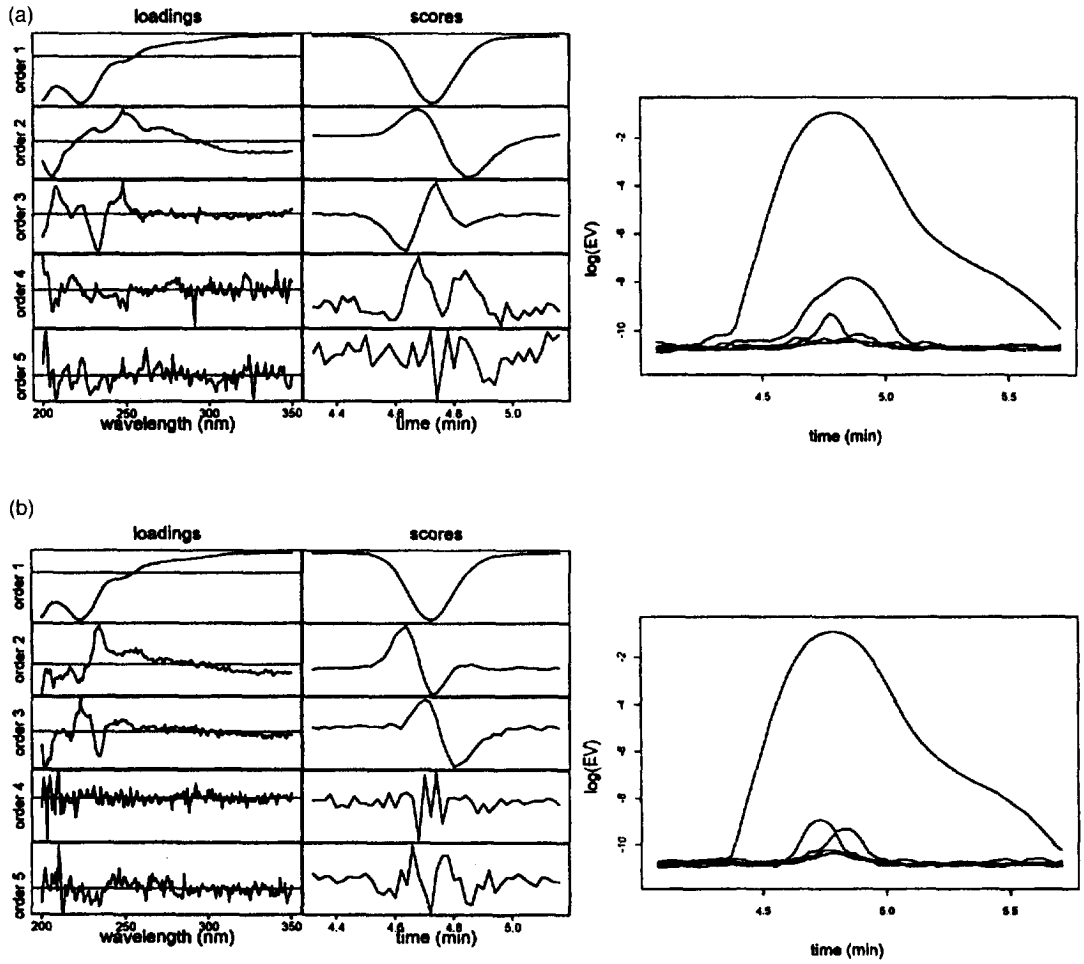


Fig. 9. Structure of the first five eigenvectors and corresponding WEFA plot obtained for (a) an actual heterogeneous alprazolam peak containing 1% of impurity with a $R_s = 0.3$ and all with a $A_{max}$ of about 0.5 AU and (b) the corresponding simulated homogeneous peak.

are clear patterns in the score and loading plots and also in the WEFA trace. However, they deviate considerably from what is shown in Fig. 9a. We thus conclude that the real peak is probably impure. This is the correct assessment, since the real peak was generated with alprazolam spiked with 1% of triazolam and eluted at a chromatographic resolution of 0.3. Note that the correct purity assessment is much more difficult at a maximum absorbency below 100 mAU.

What has been described as an informal comparison of the results of common analyses between real and realistically simulated spectrochromatograms can be made more formal and research continues in this direction. An application to ratios of dissimilarity statistics is discussed in Gilliard and Ritter [40].

Realistic simulations can also be used to anticipate the structure of the real data set and hence to select beforehand the adequate dilution of the sample which would permit the observation of a bilinear behaviour and hence would allow unequivocal interpretation of the mixture analysis results. For instance, under the working conditions used here, simulations indicate that an ideal behaviour, i.e., erratic shape for the second eigenvector in SVD and no secondary bulge visible in the WEFA plot, may be expected for a maximal absorbance below 120

Table 7
Relative amounts of the first six singular values (in %) and estimated number of primary factors by Malinowski's $F$-test ($\alpha = 5\%$)

| Real (1% imp.) | Simulated |
|---|---|
| 99.609 | 99.738 |
| 0.144 | 0.041 |
| 0.036 | 0.025 |
| 0.014 | 0.010 |
| 0.010 | 0.010 |
| 0.009 | 0.09 |
| *Malinowski's F-test* | |
| 3 | 3 |

The relative amounts were calculated for an actual heterogeneous peak of hydrochlorthiazide containing 1% impurity with a $R_s = 0.3$ and with a $A_{max}$ of about 0.5 AU and for the corresponding simulated homogeneous peak.
The columns correspond to the data of Fig. 9a and b, respectively.

mAU for alprazolam and 60 mAU for hydrochlorthiazide.

## 5. Conclusions

When applying a mixture analysis technique to LC–DAD data, one has to be aware of the possible occurrence of instrumental artefacts. Several potential ways may be explored in order to deal with this problem. Besides eventual technical improvements of the detector itself, one can concentrate on mathematical corrections aiming to return to the ideal state, or on analysis techniques which can cope with non-idealities or both together.

These approaches are however far from straightforward and a simpler method is to try to derive a simulation supposing an homogeneous peak and accounting for non-idealities, and to make comparison with the real data.

In this work, it is shown that provided the specifications of the detector are known, such realistic simulations can be carried out and successfully used for the evaluation of mixture analysis techniques. For the system used here, polychromatic radiation, scan time, transmittance average error and heteroscedasticity were found to be able to account for most of non-idealities observed in the data.

Of course, additional problems may be encountered. Mobile phase effects for instance can lead to changes in the equilibrium. Chemical factors, e.g., matrix effects or chemical equilibria, may also be of concern but are very specific for the sample studied. Moreover, environment perturbations, such as or vibrations or drastic temperature change can cause a minute shift in the wavelength dispersion of the detector which in turn can produce errors if such perturbations occur between the time an auto-zero (balance) is executed and the time spectra are computed.

However, in contrast with these last potential problems, the four instrumental factors mentioned above are of general concern and should always be taken into account. Moreover, polychromatic radiation, scan time and heteroscedasticity are widely incriminated as the main sources of non-linearities [21,22,30,36], implying that the simulation procedure described here might be portable to other LC–DAD systems with adequate adaptation of the parameters to the characteristics of the instrument (wavelength resolution, scan time, proportionality factor for heteroscedastic noise). This is supported by the work of Gerritsen et al. [20] who showed that the shape of the extra factors that were observed for pure anthracene solutions with different commercial DADs display remarkable similarities even though there are some differences in the relative amounts of the singular values between the detectors.

However, not all the instruments behave exactly in the same way as pointed out by Dose and Guiochon [29], and it is likely that the relevant sources of non-linearities would depend on the instrument used.

## Acknowledgments

The authors thank Prof. G. Guiochon, Reader in Analytical Chemistry at University of Tennessee, Knoxville, TN, USA and Mr. B.G. Archer from the Development Centre of Beckman Instruments, Fullerton, CA, USA for their assistance.

## References

[1] B.G.M. Vandeginste, W. Derks and G. Kateman, Anal. Chim. Acta, 173 (1985) 253–264.

[2] E. Sanchez and B.R. Kowalski, Anal. Chem., 58 (1986) 496–499.
[3] M. Maeder, Anal. Chem., 59 (1987) 527–530.
[4] S. Wold, P. Geladi, K. Esbensen and J. Öhman, J. Chemom., 1 (1987) 41–56.
[5] P.J. Gemperline, J. Chemom., 3 (1989) 549–568.
[6] J.C. Hamilton and P.J. Gemperline, J. Chemom., 4 (1990) 1–13.
[7] J. Öhman, P. Geladi and S. Wold, J. Chemom., 4 (1990) 135–146.
[8] O.M. Kvalheim and Y.-z. Liang, Anal. Chem., 64 (1992) 936–946.
[9] M.P.J. Gerritsen, H. Tanis, B.G.M. Vandeginste and G. Kateman, Anal. Chem., 64 (1992) 2042–2056.
[10] R. Tauler, G. Durand and D. Barcelo, Chromatographia, 33 (1992) 244–254.
[11] M.P.J. Gerritsen, H. Peters, B.G.M. Vandeginste and G. Kateman, Chemom. Intell. Lab. Syst., 18 (1993) 205–219.
[12] R. Tauler and D. Barcelo, Trac, 12 (1993) 319–327.
[13] M.F. Delaney, Anal. Chem., 56 (1984) 261R–277R.
[14] S.J. Vanslyke and P.D. Wentzell, Anal. Chem., 63 (1991) 2512–2519.
[15] H.R. Keller, D.L. Massart and J.O. De Beer, Anal. Chem., 65 (1993) 471–475.
[16] G.A. Bakken and J.H. Kalivas, Anal. Chim. Acta, 300 (1995) 173–181.
[17] T.M. Rossi and M.I. Warner, Anal. Chem., 58 (1986) 810–815.
[18] X.M. Tu, D.S. Burdick, D.W. Millican and L.B. McGown, Anal. Chem., 61 (1989) 2219–2224.
[19] Y.-z. Liang, O.M. Kvalheim, H.R. Keller, D.L. Massart, P. Kiechle and F. Erni, Anal. Chem., 64 (1992) 946–953.
[20] M.P.J. Gerritsen, N.M. Faber, M. Van Rijn, B.G.M. Vandeginste and G. Kateman, Chemom. Intell. Lab. Syst., 12 (1992) 257–268.
[21] H.R. Keller and D.L. Massart, Anal. Chim. Acta, 263 (1992) 21–28.
[22] Y.-z.Liang, O.M. Kvalheim, A. Ramani and R. Brereton, Chemom. Intell. Lab. Syst., 18 (1993) 265–279.
[23] H.R. Keller, D.L. Massart, P. Kiechle and F. Erni, Anal. Chim. Acta, 256 (1992) 125–131.
[24] H.R. Keller, D.L. Massart, Y.-z. Liang and O.M. Kvalheim, Anal. Chim. Acta, 263 (1992) 29–36.
[25] C. Ritter, J.A. Gilliard, J. Cumps and B. Tilquin, Anal. Chim. Acta, 318 (1995) 125–136
[26] M.V. Gorenstein, J.B. Li, J. Van Antwerp and D. Chapman, LC–GC, 12 (1994) 768–772.
[27] J. Schaefer, Real Time Purity, Technical note ATR 90-006, Beckman Instruments, San Ramon, CA, USA, 1990.
[28] J.A. Gilliard, J. Cumps and B. Tilquin, Chemom. Intell. Lab. Syst. Lab. Inf. Manage., 21 (1993) 235–242.
[29] E.V. Dose and G. Guiochon, Anal. Chem., 61 (1989) 2571–2579.
[30] J.-L. Excoffier, M. Joseph, J.J. Robinson and T.L. Sheehan, J. Chromatogr., 631 (1993) 15–21.
[31] H.R. Keller, D.L. Massart, Y.-z. Liang and O.M. Kvalheim, Anal. Chim. Acta, 267 (1992) 63–71.
[32] E.R. Malinowski, J. Chemom., 3 (1988) 49–60.
[33] H.R. Keller, P. Kiechle, F. Erni, J.L. Excoffier and D.L. Massart, J. Chromatogr., 641 (1993) 1–9.
[34] M. Maeder and A. Zilian, Chemom. Intell. Lab. Syst., 3, (1988) 205–213.
[35] M. Bertolin, International Laboratory, October 1991, 44–50.
[36] H.J.P. Sievert and A.C.J.H. Drouen, in L.Hubert and S.A. George (Editors), Diode Array Detection in LC, Marcel Dekker, New York, 1993, pp. 82–83.
[37] T. Alfredson and T. Sheehan, J. Chromatogr. Sci., 24 (1986) 473–482.
[38] P.J. Gemperline, Anal. Chem., 58 (1986) 2656–2663.
[39] D. Jones, Anal. Chem., 9 (1985) 1057A–1073A.
[40] J.A. Gilliard and C. Ritter, submitted to J. Chromatogr.