Journal of Pharmaceutical & Biomedical Analysis
Vol. 9, No. 8, pp. 643–654, 1991
Printed in Great Britain

0731–7085/91 $3.00 + 0.00
Pergamon Press plc

Multicomponent analysis: a case report*

MATS JOSEFSON† and LIJA TEKENBERGS-HJELTE

Department of Analytical Chemistry, AB Hässle, S-431 83 Mölndal, Sweden

Abstract: Three cases are described. Case 1: kinetic studies often need high time resolution measurements in order to follow the pattern of reactions taking place during the experiment. This is often laborious to achieve with the collection of fractions for chromatographic separation. Some tool for separation is, however, necessary in order to decompose the concentrations of reactants, products and intermediate species. The spectra of the intermediates may not be known at the time when the kinetic studies are needed. With unknown spectra there are still possibilities to use spectroscopy and multivariate techniques to obtain qualitative information. Case 2: it is possible to use Partial Least Squares (PLS) in order to describe the chromatographic profiles for the species even if the separation is insufficient for traditional peak measurement methods. This requires that mixtures are available with known concentrations of the species to be determined. Case 3: with modern diode array liquid chromatography detectors there is the possibility to capture the chromatogram and the spectra at the same time. The ability to reproduce the chromatographic profile between samples makes it possible to use the Generalized Rank Annihilation Method (GRAM) possible. Whereas PLS only treats one spectrum at a time, this method treats the full two-dimensional chromatogram as an entity. The GRAM calibration is claimed to be insensitive to interfering species which are not present in the calibration. Limitations are that GRAM requires a linear detector response and very good repeatability of the retention time. The use of GRAM for calibration with real samples is demonstrated.

Keywords: *Chemometrics; kinetic analysis; liquid chromatography; Partial Least Squares; generalized rank annihilation method; diode array detection.*

Introduction

The purpose of this paper is to describe some multivariate analysis options available for curve resolution when dealing with data arranged in two-dimensional arrays. This kind of data is obtained from a number of hyphenated techniques or several one-dimensional data sets gathered from, for example, chemical systems changing with time. The hyphenated techniques are represented by liquid chromatography followed by UV spectral detection used to determine two structural isomers. The changing chemical system examined is based on the reaction kinetics for substances similar to those used in ref. 1.

This article concentrates on three multivariate methods: Principal Components Analysis (PCA) followed by factor rotation; Partial Least Squares (PLS); and the Generalized Rank Annihilation Method (GRAM). In this order these methods put increasing demands on the amount of information known beforehand and on the treatment of data but will also yield increasing benefits by making better use of all data obtained during measurement.

PCA and related methods have been used extensively to compile information from multichannel data and generate pure spectra. One approach is to rotate the PCA solution while imposing physical constraints such as positivity, as illustrated in the classical work by Lawton and Sylvestre [2]. Other authors have further developed this scheme [3, 4]. The simplex method has also been used to find the limited solution [5]. Ratioing [6] has been used for unknown mixtures, but ratioing is dependent upon there being "flat" regions of the ratio in order to work well. Alternating Regression (AR) [7] is another approach where iterated regression combined with physical constraints resolves the mixed data.

Target Transformation Factor Analysis (TTFA) [8, 9] and Iterative Target Transformation Factor Analysis (ITTFA) [10] represent another class of techniques which requires general information about the curve form in either the spectral or the time domain in order to find a solution. This can be seen as an extended form of the physical constraints used in the previously mentioned methods.

These methods are used to extract information when nothing is known about individual concentrations in the mixture and when no pure standards are available. These

*Presented at the Symposium on "Chemometrics in Pharmaceutical and Biomedical Analysis", November 1990, Stockholm, Sweden.

†Author to whom correspondence should be addressed.

methods are sometimes hampered by rigid adherence to the imposed constraints. The constraints may, however, be interactively overridden by the use of a graphics interface. The analyst may select any rotation of three factors graphically guided by, but not forced by the constraints. Matlab [11] is used to implement this as exemplified in case 1 of this paper.

Multivariate calibration methods such as PLS can be used for concentration calculations when the concentrations are known of the individual species in a set of mixtures constituting the calibration set. A more-or-less independent variation of the species between the calibration samples is necessary in order to achieve a suitable calibration. The best calibration is obtained when the calibration set contains all sources of variation that may occur in the actual measurement. In this way PLS can be used for chromatographic data with a similar approach to that used in spectroscopy [12]. This method will only resolve the peaks in a chromatogram which are included in the calibration set, but no separation will be necessary at all. Case 2 is an example of how to deal with insufficient separation, using PLS.

The GRAM [13–15] is a calibration method that can be used directly on two-dimensional arrays. Mixture standards with known concentrations are used. As compared to PLS, a slight separation of the components is needed. Full advantage is taken of both the chromatographic profile and the set of spectra present in the two-dimensional sample array. This gives the claimed advantage [13] that the method is insensitive to other species not calibrated for, as long as the interfering substance is present either in the calibration set or in the sample but not in both. This is in contrast with the PLS calibration approach, where the amount of interfering substance must be varied in the calibration set in order to give a reliable prediction of the other substances present in the sample. In Case 3 the GRAM technique is applied to the data set from Case 2. The aim of the present work was to use the method with a non-simulated realistic data set in order to assess the benefits and drawbacks that are observed with the GRAM technique used for the first time in the author's laboratory.

Experimental

The kinetic data in Case 1 were collected at 2 nm intervals with a UV–vis spectrophotometer (HP 8450) in the wavelength range 200–400 nm. The reaction was followed by collecting a spectrum every 30 s for 20 min. Standard solutions for Cases 2 and 3 were prepared by mixing two structural isomers as a model system, to give concentrations of 90–110 μg ml^{-1}. The proportion of isomer 1 varied from 1 to 10% of the total. The chromatographic separation was performed on an analytical Polygosil C8 column, with a mobile phase consisting of acetonitrile and phosphate buffer. The resolution R_s, was estimated to be 0.7. Data were obtained with a diode-array detector (Perkin–Elmer LC-480 with LC-DES software v2.0). Spectra, from 200 to 360 nm in 4 nm steps, were collected every 1.2 s.

All software has been developed and used on IBM compatible PC systems. In Case 1 the Sirius v1.5 software (Pattern Recognition Systems a/s, Norway) was used to estimate the number of components in PCA with cross-validation according to Wold [16]. The rest of the software for Case 1 was written in Matlab by the authors.

In Cases 2 and 3 the spectra were converted from the diode array detector format to Matlab format by a conversion program written by the authors in Turbo Pascal. The selection of spectra was made in Matlab. Unscrambler v2.3 (Camo a/s, Norway) was used for the PLS calibration and the GRAM implementation was written in Matlab by Eugenio Sanchez (version: 9 September 1988).

Results and Discussion

At the primary stage of the selection and development of a new drug, very little may be known about the pure spectral shape of the degradation products and the amount of available material is usually very small. In this case it is favourable to use multi-channel analytical techniques to obtain more information per unit of available substance consumed by the analysis. The degradation of a substance in a destabilizing solvent was studied by the use of PCA. This was possible with no available standards since the spectra of the components were different. The analyst also needed a general knowledge of the kinetics for this kind of substance. A selection of the collected spectra is shown in Fig. 1. The spectrum of the starting compound was obvious but conclusions about the spectral profiles of intermediates and reactants were difficult to make

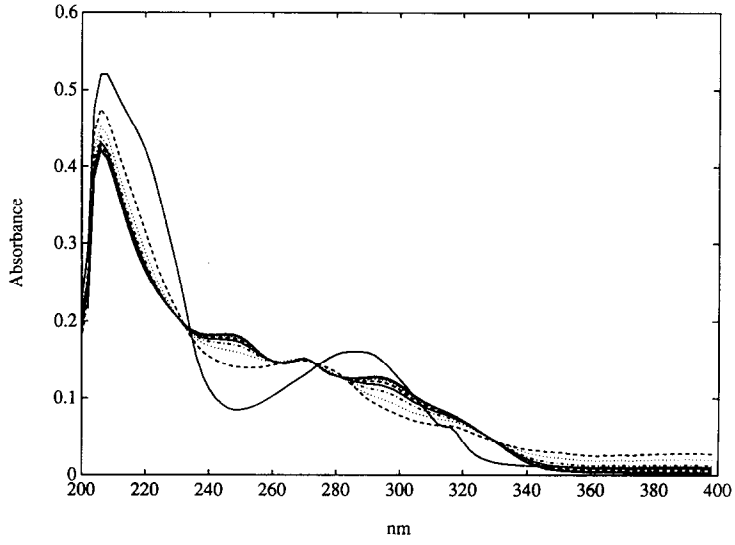

Figure 1
Spectra collected during the kinetic experiment (13 of 41).

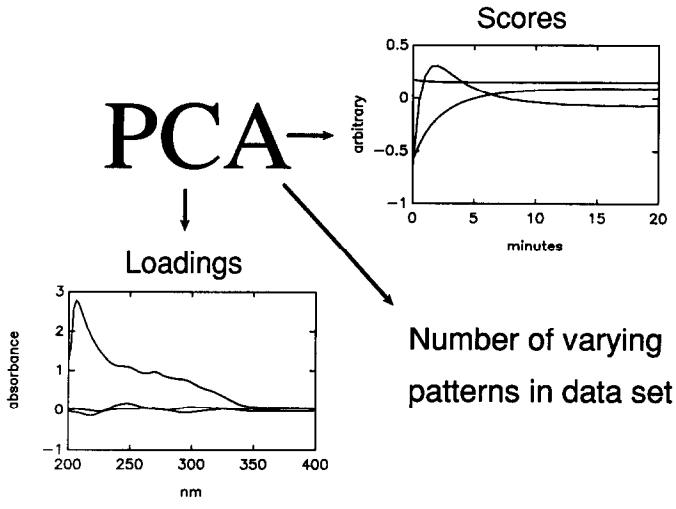

Figure 2
Output obtained from PCA with cross-validation.

from the raw multi-channel data. The information content in this series of spectra was computed by the use of PCA (Fig. 2). The cross-validation gave three significant components, which are shown in Fig. 2 as three score vectors carrying the information of the mixed kinetic profiles and three loading vectors carrying the mixed spectral information of the three discernible species. As observed in Fig. 2, the score and leading vectors covered both negative and positive values, since the PCA always gives orthogonal components with no individually clear-cut physical meaning. As described in the Introduction, many methods exist to impose a physical meaning by different methods used for the scores and the loadings. It may, however, be advantageous to handle the data in a less automated way in order to asses the full ambiguity in a situation where concentrations are not known and pure spectra are not available.

A manual mouse-driven application was developed in Matlab to fill this goal. First the score profiles of Fig. 2 were combined:

$$\text{kinetic profile} = t_1 + xt_2 + yt_3 + \text{err}, \quad (1)$$

where the values represent the profiles obtained by PCA, while x and y are the weights required to combine the three t-vectors into a

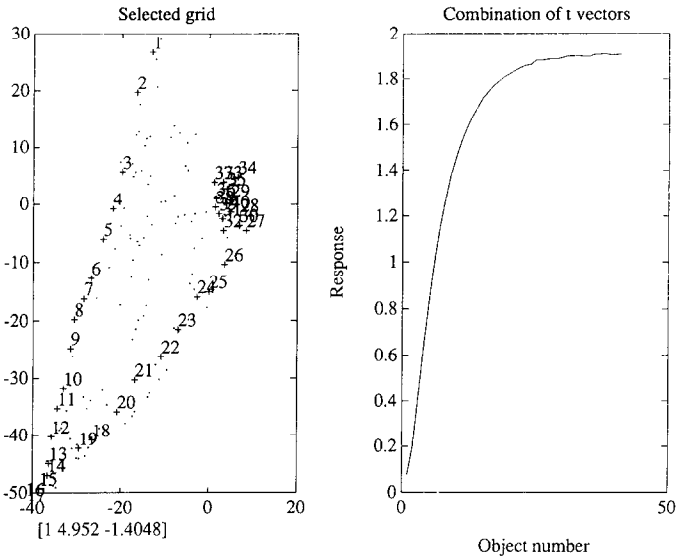

Figure 3
Screen layout of mouse-driven application. The dots in the left window represent the combination of scores resulting in positive kinetic profiles. The numbered crosses represent clicks on the mouse button to interactively generate a profile in the right window.

simulated kinetic profile. The term err is the error of the reconstruction. The Matlab random number generator was used to generate combinations of x and y which resulted in different profiles. Wherever the full profile was above the zero level, a dot was plotted at the coordinate (x,y) as shown in the left part of Fig. 3. Every dot thus represented a simulated kinetic profile, to which was attached more or less physical meaning. The scaling of x and y was adjusted to an extent where regions with no dots were visible around a triangular shaped area in order to cover all possible positive combinations. The numbered crosses show the path of clicks on the mouse. Each click resulted in a plot of the kinetic profile on the right side of the screen shown in Fig. 3. The latest combination of x and y was shown below the left window of the screen. In this manner the analyst was able to investigate different probable reaction profiles and compare them with his/her own experience in a flexible way.

Three probable profiles were found, one increasing — for the product (Fig. 3), one decreasing — for the starting material, and one intermediate which increased and decreased. Then each profile was multiplied with the raw data matrix (Fig. 4) to calculate the estimated spectra (Fig. 5) according to:

$$\text{spectrum} = Xt_c, \qquad (2)$$

where X is the raw data matrix with the spectra

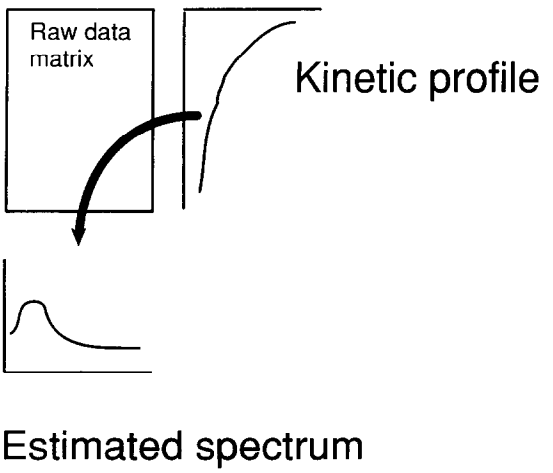

Figure 4
The spectrum of the substance following a kinetic profile was calculated by matrix vector multiplication.

as rows and t_c is the combined kinetic profile vector. This multiplication results in the vector spectrum. The main advantage with this approach is as a complement to more rigorous methods where the constraints are difficult to override in order to see what happens outside the constrained region.

The above procedure was also used for the data in Cases 2 and 3. It was not possible to simulate the chromatographic profile within the positive domain. It was, however, possible to extend the range of the (x,y) coordinates to simulate both the first and the second peak, provided that the baseline of the chromato-

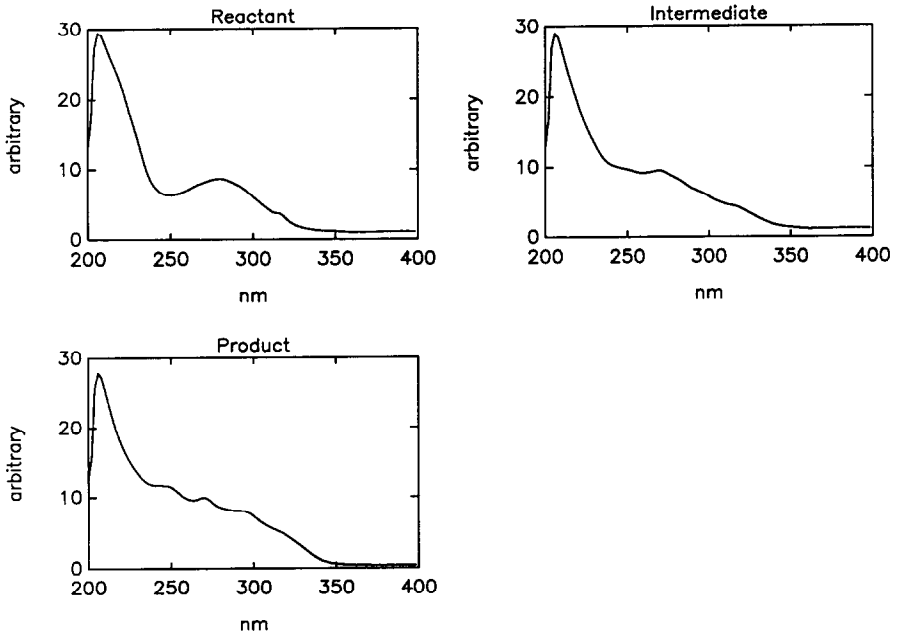

Figure 5
The spectra calculated from the estimated kinetic profiles.

gram was allowed to have a negative offset. The conclusion was that this method is useful as qualitative tool in itself, and also as a primary stage to check the feasibility of the imposed constraints with the use of more rigorous methods such as AR, ITTFA, or fitting to the applicable kinetic equations.

In Case 2, where there are samples of mixtures with known concentrations available, it is possible to use PLS. It was considered beneficial to obtain the extra information provided by the chromatographic profile. Thus it is possible to reconstruct a chromatogram for each pure constituent. The shape of each reconstructed profile and its estimated deviations becomes a calibration quality judgement. This way is a more information-rich in making use of the entire two-dimensional chromatogram, than simply calculating the concentrations from the peak maxima alone. In order to achieve this, each spectrum must be decomposed by PLS to give the pure concentrations. When the standards are injected on a chromatographic column, separation of the standards will ensue. The amount of separation could be characterized explicitly and used for the calibration. Instead, preprocessing is kept to a minimum by the use of "anti-chromatography", where a situation with no separation is simulated. This is done by addition of the spectra at the peak maxima of the standards. This method gives a way to transfer the explicit quantification of the separation to an implicit one which is soft modelled by PLS. A cleaner approach, still with minimal preprocessing, but without the advantage of chromatogram reconstruction, was also developed for comparison. Here the spectra from the peak maxima were put side by side in one object per chromatographed standard, in order to put data for the constituent combination into one object without further preprocessing.

PLS was applied to investigate the feasibility of a method for determining the concentration of two structural isomers with non-resolved chromatographic peaks. Spectra of the two isomers are shown in Fig. 6 and the chromatogram is shown in Fig. 7. The method was intended to measure the concentration of isomer 1 in peak 1 and it should not exceed 2/98 in relation to isomer 2 in peak 2. Mixtures with known concentrations of the isomers were further blended to obtain a design around the expected concentrations. Spectra were extracted from the two-dimensional chromatogram at the peak maxima and at the baseline in front of the two peaks as shown in Fig. 7. The spectra from each chromatogram at the peak maxima were: (1) anti-chromatographed (i.e. added together to obtain approximately the same situation as for the use of PLS in spectroscopy) and (2) put side by side in a

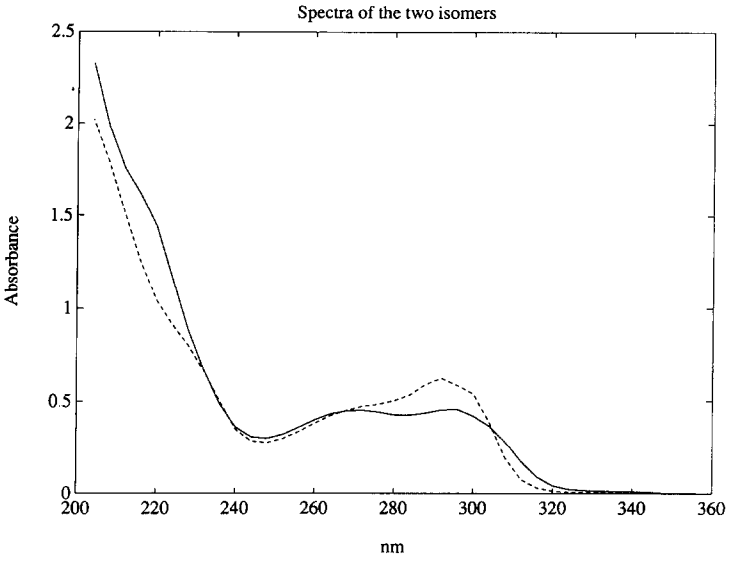

Figure 6
Spectra of the two isomers. Isomer 1, solid line; isomer 2, dashed line.

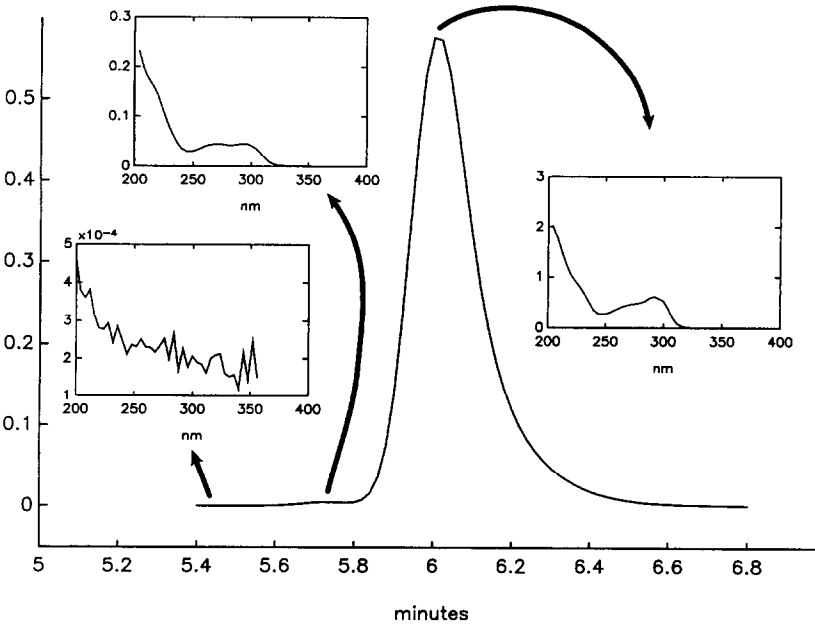

Figure 7
The section of the chromatogram used in Cases 2 and 3. Spectra for the PLS calibration were extracted as shown.

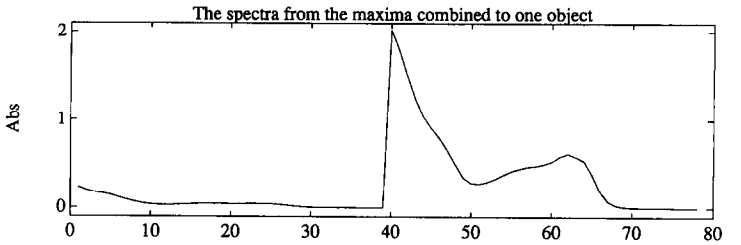

Figure 8
Spectra arranged side by side in one object for the second PLS calibration.

common object (Fig. 8). The baseline spectrum was: (1) used directly as an object and (2) two copies of the same spectrum were put side by side to fill the object size of the side by side isomer spectra. Both calibration sets were centered by subtraction of the mean spectrum and all weights for the wavelength variables were set to 1. The four corners and the center of the design (Fig. 9) were used for PLS calibration while the other standards in the design were used to test the calibration. PLS with two concurrent response variables gave three components with the Unscrambler cross-validation for the added spectra version, and two components for the side by side version of the objects.

Then the first calibration was used to calculate the concentration chromatograms for the

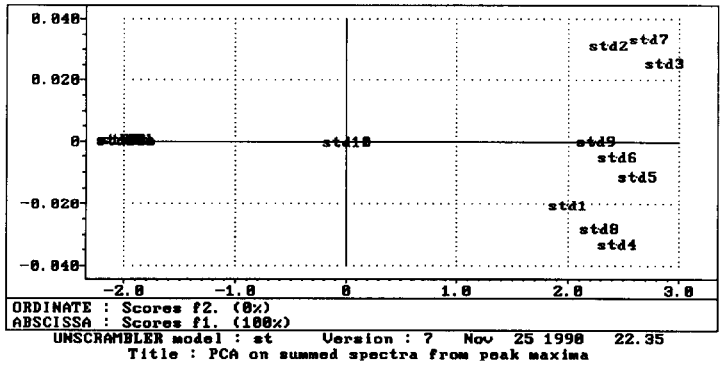

Figure 9
The two first PCA score vectors from the first PLS calibration. The shape of the square design was visible for the standards std1–9. The baseline spectra are placed in the left-most bunch. The bunch at the origin is an artefact deriving from excluded objects in the software. The point for std2 is slightly off the design indicating a small error.

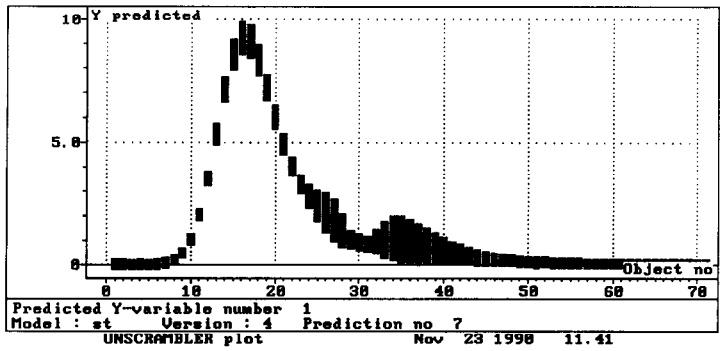

Figure 10
The predicted chromatogram of isomer 1 in concentration units ($\mu g\ ml^{-1}$) including the error estimated by Unscrambler as the height of the bars. The prediction was made by PLS calibration with added spectra.

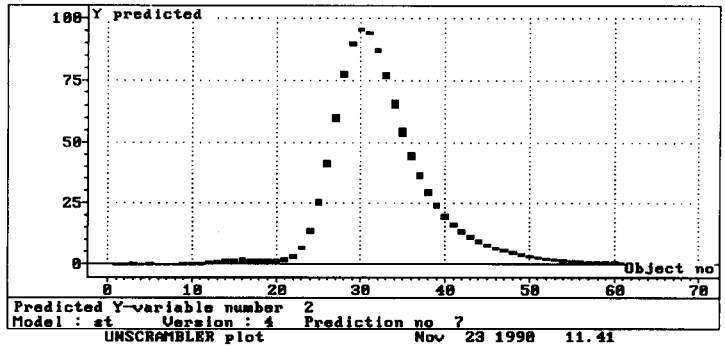

Figure 11
The predicted chromatogram of isomer 2 in concentration units ($\mu g\ ml^{-1}$) including the error estimated by Unscrambler as the height of the bars. The prediction was made by PLS calibration with added spectra.

isomers. The uncertainty in the prediction of the sample std7 was calculated by Unscrambler and shown as the heights of the bars in the Figs 10 and 11. The estimated error is 15% of the isomer 1 concentration at the isomer 1 peak maximum and 2% of the isomer 2 concentration at the isomer 2 peak maximum. An increase in the error was apparent for the concentration of isomer 1 on the upslope and downslope of the isomer 2 peak. This was to be expected, because this concentration range was not included in the design. The loading plots are shown in Fig. 12. Component 1 was interpreted as the overall change for both spectra. Component 2 was like a difference spectrum in order to adjust the first component

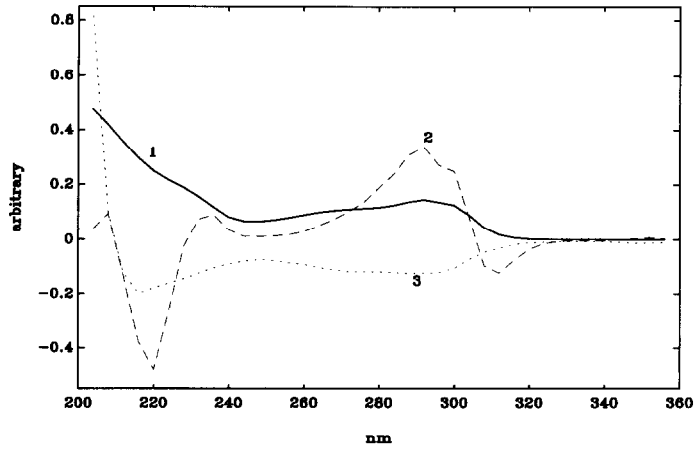

Figure 12
The loading vectors for PLS calibration with added spectra from the maxima. Component 1, solid line; components 2 and 3, dotted line. See the text for further explanation.

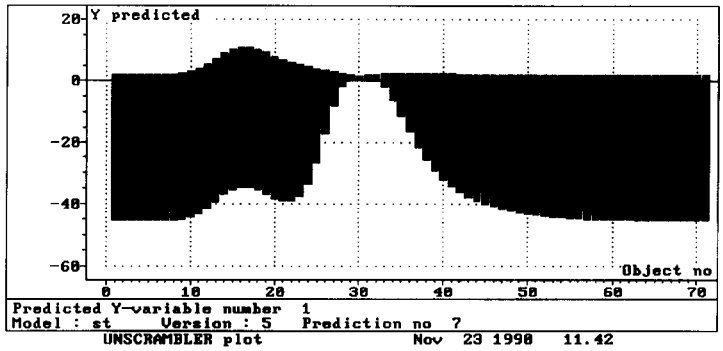

Figure 13
Prediction of the minor peak for isomer 1 from the same calibration set as in Fig. 10 but with excluded baseline spectra.

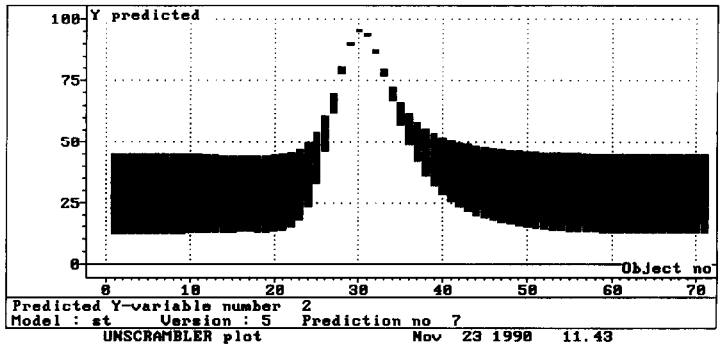

Figure 14
Prediction of the major peak for isomer 2 from the same calibration set as in Fig. 11 but with excluded baseline spectra.

by varying concentration combinations of the two isomers. Component 3 is mainly a reflection of the first component, indicating a deviation in the linearity of the measurement.

Why was the baseline spectrum necessary? Figures 13 and 14 show the increased uncertainty when the baseline spectra and their corresponding zero response values are excluded from the PLS calibration. The calibration error estimate for isomer 1 led to an impossible calibration (Fig. 13), but the calibration for isomer 2 was still usable at the peak maximum. A large bias was observed in the baseline measurement for both the isomers since the baseline was not included in the calibration.

The PLS calibration from the side by side spectra to the two responses was cross-validated to two components (see Table 1). The loading plot is shown in Fig. 15 where each original spectrum is visible. Component 1 in the variables 1–39 is the correction for tailing below the first peak. In the variables 40–78, the first component describes the main variation of the isomer 2 concentration. Component 2 has the reverse function as a main loading for isomer 1 and a correction of the tailing below the second peak in variables 40–78. This calibration needed one component less than the previous calibration for added spectra. Hence this calibration is more parsimonious and less prone to overfitting. It also gives a better understanding of the loading plot.

Both calibrations are dependent upon the quality of the chromatographic column, hence a new calibration must be provided when the column deteriorates or other conditions affecting the separation are changed. An advantage is that no pure standards are needed and that even completely unseparated species may be resolved, provided that their spectra are different. In the first calibration, the full flexibility of PLS succeeded in giving an elaborate picture of the measured signal. The price paid for this was the appearance of an extra PLS-component. The preferable calibration is that with the least number of components, which means that the anti-chromatography approach was not the optimal method for quantitation, even if it worked reasonably well. Both approaches showed the benefits of PLS as a modelling tool to deal with data containing the desired information and extract this information with minimal extra assumptions about separation and peak shape.

For Case 3, in the PLS calibration no systematic advantage was given by the chro-

Table 1
Comparison between values calculated from the balance, the integrator height, PLS with added spectra, and PLS with side by side (sbs) spectra

	Isomer 1				Isomer 2			
	Balance	PLS added	PLS sbs	Intgr. ht	Balance	PLS added	PLS sbs	Intgr. ht
Std5	5.2	4.2	4.5	4.1	99.9	99.4	100.1	100.3
Std7	9.4	9.2	9.7	10.1	94.9	95.5	96.9	96.5
Std8	1.0	1.0	1.2	1.3	94.9	95.7	95.9	96.2
Std9	5.1	4.1	4.3	4.4	89.9	91.3	91.8	91.8

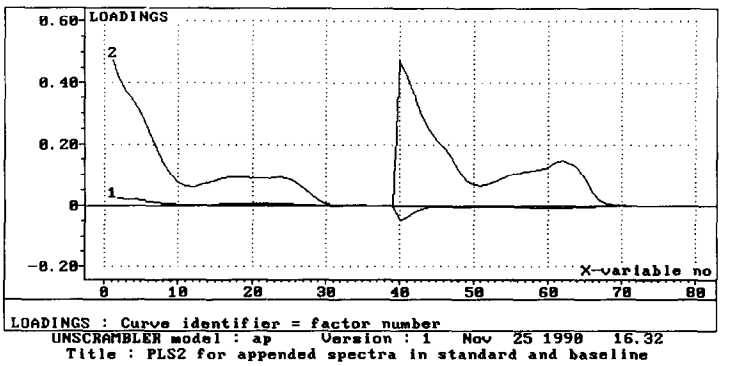

Figure 15
The loading vectors for the PLS calibration with side by side spectra from the maxima. See the text for further explanation.

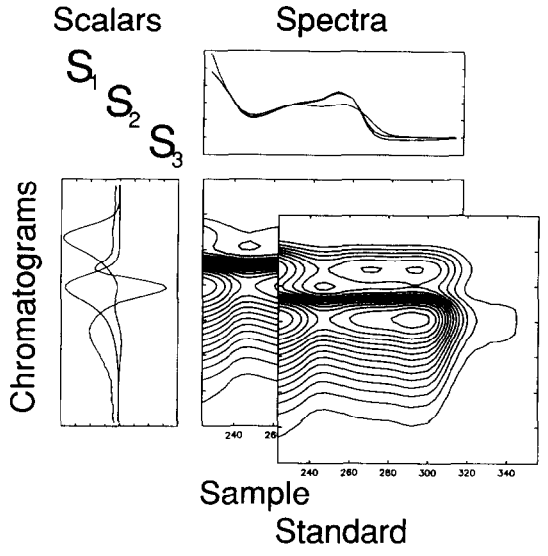

Figure 16
The scalars S_1–S_3 reflect the relations between the constituents present in both 2-D chromatograms. The deconvoluted spectra and chromatograms have arbitrary scale units, since their respective magnitudes are given by the scalars. A 2-D chromatogram for one component may be reconstructed by vector multiplication of its spectrum, its scalar, and its chromatogram.

matographic separation. A way to make use of all available information in the 2-D chromatogram is to use the GRAM technique. This method is claimed to be insensitive to interferences not calibrated for, and will give pure spectra of all constituents that are present in both of two compared samples. A scalar gives the relation between the concentration of the same constituent in the two samples (Fig. 16). Each scalar is associated with the deconvoluted spectrum and chromatogram for that constituent. In order to know which scalar measures the relationship of a substance between the two samples, it is necessary to look at the deconvoluted spectra and/or chromatograms. When one of the samples is substituted for a combination standard, it is possible to make a calibration. One drawback as compared with PLS is that GRAM requires a linear detector response. Also, the ratios of the concentrations between the standard and the sample must differ for different species. Each standard and sample must consist of a 2-D grid of values, where each substance increases and decreases once only (as in normal chromatography). The method is not suitable for 2-D spectra where the information from one species results in several peaks at different positions as in 2-D NMR.

First the 2-D chromatograms were subjected to GRAM with no pretreatment except for truncation of the wavelength range to 200–224 nm in order to avoid the nonlinearity at high absorbance values. This gave deconvolutions with negative peaks in the chromatographic profile. Then the sample 2-D chromatograms were shifted and compared with the standard chromatogram by the plot of the two chromatograms against each other at the same wavelength. This resulted in a loop-shaped plot if the chromatograms were out of

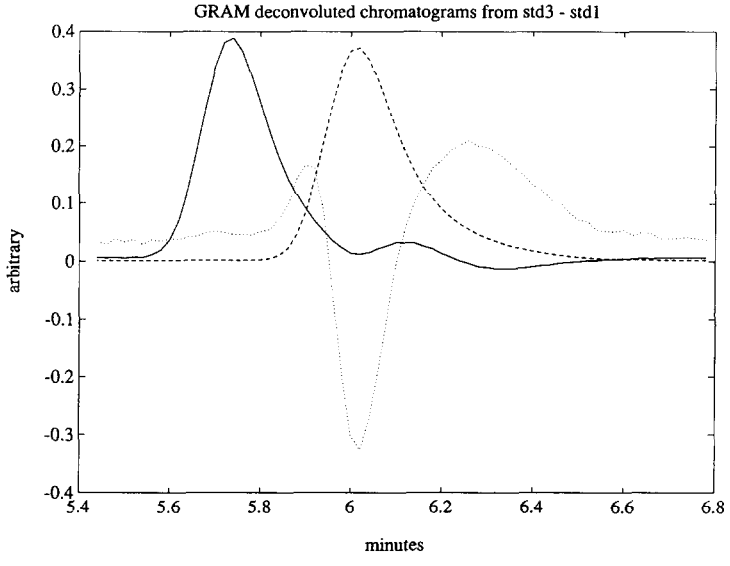

Figure 17
Chromatograms in arbitrary units generated by GRAM from std3 and std1. The dotted line is a correction for a nonlinearity in the measurement.

Table 2
A comparison of ratios calculated from balance data and by GRAM deconvolution

	Isomer 1		Isomer 2	
Combination	Balance	GRAM	Balance	GRAM
Std3–std1	10.42	10.23	1.11	1.10
Std3–std4	9.37	8.98	1.00	1.01
Std3–std7	1.01	1.02	1.05	1.05

phase and a straight line if the chromatograms were in phase. The extent of the shift needed was determined by interval halving combined with visual inspection of the plot. The samples in the order presented in Table 2 were shifted by −0.30, 0.23 and −1.68 s, respectively, vs the standard. The new values at the times corresponding to the sampling times of the standard were interpolated by the cubic spline function in Matlab. The ratios are shown in Table 2. The deconvolved chromatograms are shown in Figs 17 and 18 and the corresponding spectra are shown in Figs 19 and 20. It is obvious from the shapes of the chromatograms that this deconvolution was less than optimal. The dotted third component in Fig. 17 had the

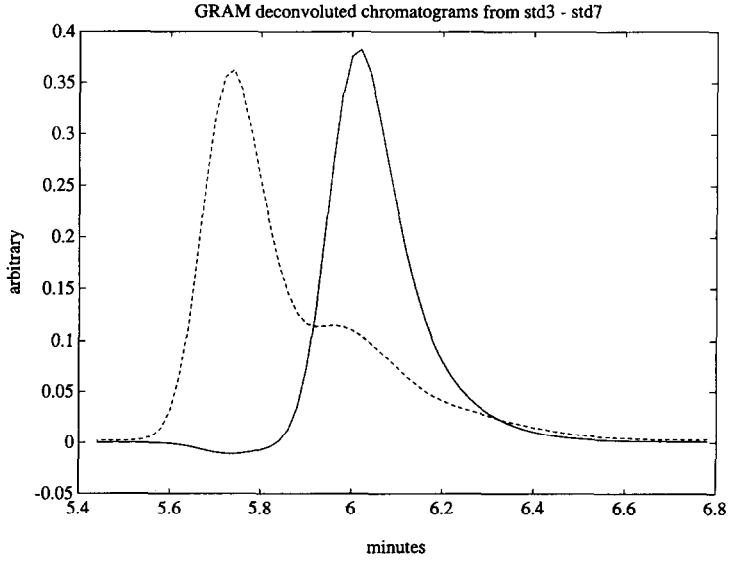

Figure 18
Chromatograms in arbitrary units generated by GRAM from std3 and std7.

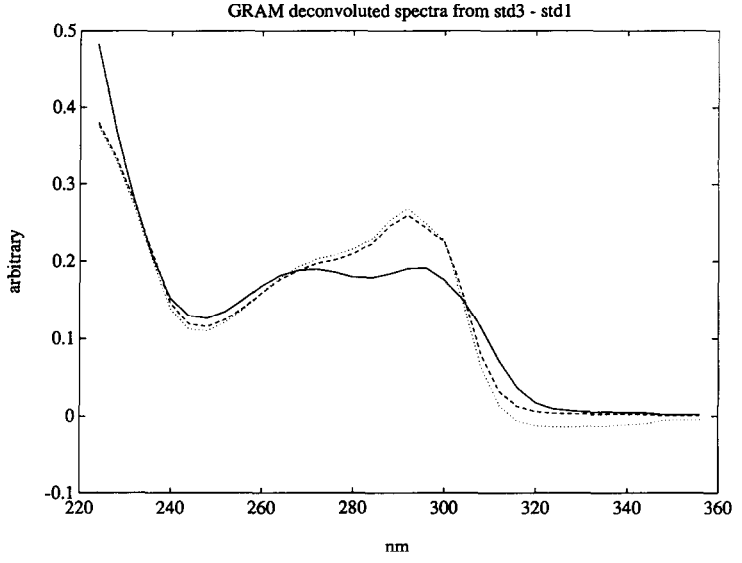

Figure 19
Spectra as deconvoluted by GRAM from std3 and std1. The line types correspond with the line types in Fig. 17.

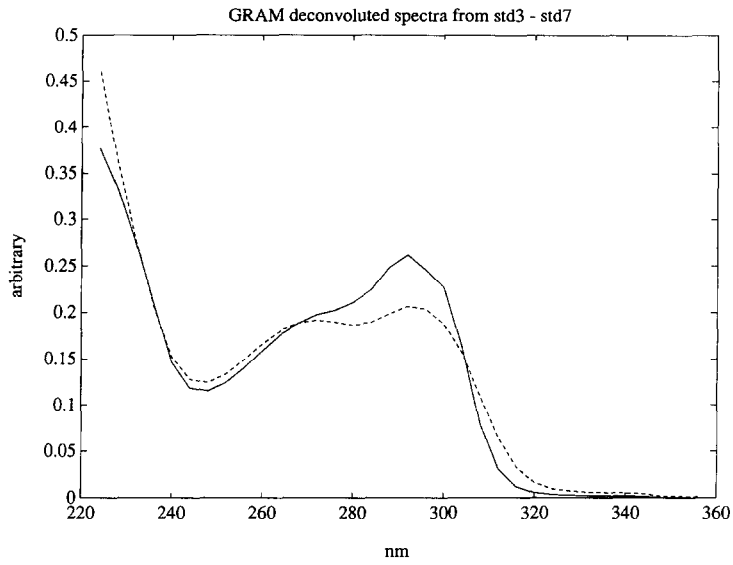

Figure 20
Spectra is deconvoluted by GRAM from std3 and std7. The line types correspond with the line types in Fig. 18.

shape of a second derivative, which may indicate nonlinearity. The method works better for std7 in Figs 18 and 20 in spite of the close similarity of the ratios of the species between sample and standard. Figure 19 shows that the nonlinear correction has approximately the same spectrum as that of isomer 2. Thus the palatable properties of the GRAM were used on non-ideal samples and the extracted spectra and chromatograms were remarkably easy to reconstruct. By inspection of the deconvoluted chromatograms a nonlinearity was identified. It was somewhat dubious to determine number of components, especially for std7. The number of attempted components was varied from 1 to 4. An indication of excessive number of components was that the spectra and quotients had imaginary parts. More work has to be done developing criteria, before the number of components can be determined on a routine basis.

References

[1] A. Brändström *et al.*, *Acta Chem. Scand.* **43**, 536–611 (1989).
[2] W.H. Lawton and E.A. Sylvestre, *Technometrics* **13**, 617–633 (1971).
[3] N. Ohta, *Anal. Chem.* **45**, 553–557 (1973).
[4] H. Martens, *Anal. Chem. Acta* **112**, 423–442 (1979).
[5] A. Meister, *Anal. Chem. Acta* **161**, 149–161 (1984).
[6] T. Hirschfeld, *Anal. Chem.* **48**, 721–723 (1975).
[7] E.J. Karjalainen, *Chemom. Intell. Lab. Syst.* **7**, 31–38 (1989).
[8] M. McCue and E.R. Malinowski, *Appl. Spectroscopy* **37**, 463–469 (1883).
[9] J.K. Strasters, H.A.H. Billet, L. de Galan, B.G.M. Vandeginste and G. Kateman, *J. Chromatogr.* **385**, 181–200 (1987).
[10] B.G.M. Vandeginste, F. Leyton, M. Gerritsen, J.W. Noor and G. Kateman, *Anal. Chem.* **57**, 971–985 (1985).
[11] T.C. O'Haver, *Chemom. Intell. Lab. Syst.* **6**, 95–103 (1989).
[12] H. Martens, *Multivariate Calibration*. Wiley, New York (1989).
[13] E. Sanchez and B.R. Kowalski, *Anal. Chem.* **58**, 496–499 (1986).
[14] E. Sanchez, L.S. Ramos and B.R. Kowalski, *J. Chromatogr.* **385**, 151–164 (1987).
[15] L.S. Ramos, E. Sanchez and B.R. Kowalski, *J. Chromatogr.* **385**, 165–180 (1987).
[16] S. Wold, *Technometrics* **20**, 397–405 (1978).

[Received for review 14 February 1991; revised manuscript received 30 May 1991]