Rank annihilation correction for the amendment of instrumental inconsistencies

Marcel Maeder [a,*], Yorck-Michael Neuhold [a], Aaron Olsen [a],
Graeme Puxty [a], Raylene Dyson [b], Arne Zilian [b]

[a] *Department of Chemistry, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia*
[b] *Solvias AG, Online Technologies, Klybeckstrasse 191, 4002 Basel, Switzerland*

Abstract

The globalisation of the analysis of a series of individual measurements often results in more robust and reliable outcomes. However, instrumental drifts that can occur between individual measurements destroy the ideal data structure and thus the advantages. A method based on rank annihilation factor analysis (RAFA) is introduced for the correction of several types of instrumental inconsistencies. It can be applied to many series of bilinear datasets. Experimental examples discussed in this paper comprise the successful correction of non-uniform retention time drifts in chromatography due to temperature or pressure changes, wavelength shifts in IR spectroscopy in an industrial control situation, and background absorption shifts in UV–VIS spectroscopy applied to equilibrium investigations.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Correction; Bilinear data; Rank annihilation; Chromatography; Baseline

1. Introduction

With the availability of rapid scanning and diode-array spectrometers, bilinear data are readily available and form the standard basis of many chemical investigations. Examples include kinetics, equilibrium studies, and chromatography. More recently, it was recognised that bilinear datasets can be linked and the global analysis of the complete set can be of significant advantage. Several soft-modelling and hard-modelling algorithms have been developed that globally analyse such series of individual measurements. Soft-modelling methods range from very general approaches with minimal demands on the structure of the data, such as Tucker [1], to methods, which rely on trilinearity, such as PARAFAC [2], GRAM [3] or DTD [4]. Hard-modelling methods, e.g. for kinetic multiwavelength analyses, where the rate law is used as a hard model, are usually more robust, but demand even more stringent prerequisites, the adherence to the chosen kinetic reaction mechanism [5].

The closer to 'ideal' the data are, the more powerful the methods that may be used for their analysis. This is particularly the case for hard-modelling analyses, which usually do not tolerate experimental inconsistencies, or if the problem is ignored, they will deliver erroneous results that are difficult to recognise. Soft-modelling algorithms, too, are much more stable and robust if ideal structures, such as trilinearity, are maintained in the data set. Unfortunately, there are several types of irregularities between the individual

---

* Corresponding author. Fax: +61-2-4921-5472.
E-mail address: chmm@cc.newcastle.edu.au (M. Maeder).

datasets, which destroy the potential advantages inherent in the structure of 3-way data. Inconsistencies exemplified in this paper include retention time shifts in chromatography due to column ageing, pressure changes or insufficient thermostating; wavelength shifts between measurements; or baseline shifts resulting from irreproducible positioning of the absorption cell. In this contribution, we propose a very generally applicable method for the detection and correction of such irregularities.

Inconsistencies of these types have been known for a long time and several different approaches for their correction have been suggested. One class of methods involves the incorporation of a correction function into a complete calibration method. Examples include Kalman filter methods [6,7] or artificial neural networks [8]. Another class attempts a correction by the addition of linear combinations of related response curves. This can be done by the addition of one or more derivatives of these curves, e.g. chromatograms [9], or by the addition of linear combinations of neighbouring response curves, basically, the neighbouring rows and/or columns of a multivariate data matrix, e.g. response curves acquired at neighbouring wavelengths [10]. Obviously only very small deviations can be corrected for by these two methods. A very straightforward idea is to adjust the chromatographic retention times by comparing peak maxima [11], or just the two extreme profiles [12].

One powerful class of methods is based on rank annihilation factor analysis (RAFA) [13,14]. It has successfully been applied to correct uniform retention time shifts in chromatography [15]. A closely related method based on orthogonal projections has been suggested for the same type of inconsistency [16]. In this contribution, we are proposing a generalisation of the RAFA algorithm. It allows the correction of inconsistencies that cannot be corrected by a simple shift (one parameter), i.e. corrections functions with any number of parameters can be developed and tested. There is no theoretical limit to the complexity of the required correction functions.

2. Rank annihilation correction (RAC)

We will introduce the concept for the example of a set of two chromatograms (chromatogram *A* and *B*) of

two solutions of the same components with different concentrations. We assume changes in the retention times for the components. This could be the effect of small temperature changes between the chromatographic runs, differences in the pressure or slow ageing of the chromatography column. The changes in retention times under such conditions are not uniform, i.e. they are individually different for each component and thus, cannot be corrected by a simple shift.

For the example, both chromatograms contain *m* components; we will discuss generalisations later. The chromatograms can be arranged in the matrices Y_A and Y_B where the rows are the absorption spectra measured during the elution and the columns are the chromatograms at the different wavelengths. According to Beer–Lambert's law, each matrix can be decomposed into the product of the matrices C_A and C_B of concentration profiles and the matrix A of molar absorption spectra.

$$Y_A = C_A A, \quad \text{and} \quad Y_B = C_B A \tag{1}$$

If nt spectra were measured at nl wavelengths, the dimensions of both matrices Y_A and Y_B are $nt \times nl$, the matrices $C_{A/B}$ have dimensions $nt \times m$, and A has dimensions $m \times nl$ where m is the number of components. In the ideal case (no inconsistencies) the columns of C_A are multiples of the corresponding columns of C_B and thus linearly dependent.

The chemical rank of both matrices Y_A and Y_B is *m*. Next, the rank of the concatenated matrix $[Y_A, Y_B]$ is investigated, (Fig. 1). We are using the MATLAB[®]
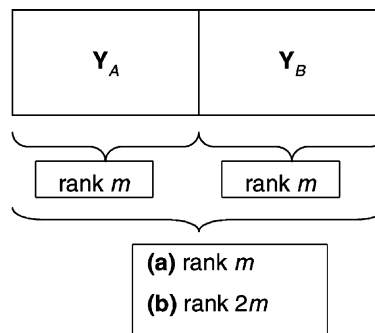

Fig. 1. The rank of two horizontally concatenated matrices Y_A and Y_B. Case (a) the vertical scalings (e.g. retention time) coincide, the rank of the $[Y_A, Y_B]$ is *m*. Case (b) the vertical scalings are different (e.g. shift in retention times), the rank of the concatenated matrix is >*m*, theoretically it is 2*m*.

notation for the two types of concatenation, $[Y_A, Y_B]$ is the horizontal concatenation (comma) and $[Y_A; Y_B]$ the vertical concatenation (semicolon). Naturally, to allow horizontal or vertical concatenation, the appropriate dimensions of the two matrices must match. That is, for horizontal concatenation the number *nt* of rows needs to be the same whereas for vertical concatenation, the number of wavelengths (*nl*) must be the same.

Under perfect conditions the concentration profiles in $C_{A/B}$ will be multiples of each other and thus the rank of $[Y_A, Y_B]$ will be the same as the rank of the individual matrices, as indicated in Fig. 1(a).

If the temperature has changed between the two chromatograms, or the column has aged measurably, the concentration matrices C_A and C_B will be different (linearly independent) and the rank of $[Y_A, Y_B]$ will be 2*m*, see Fig. 1(b). This is theoretically the case, in practice, this theoretical limit of 2*m* will hardly be reached but the rank is certainly larger than *m*.

We propose to correct one of the matrices Y_A, Y_B, say Y_B, by adjusting its time vector, with the goal of rectifying the inconsistency of the retentions. For the example, the time vector *t* at which the spectra of Y_B were measured is corrected by an appropriate function, defined by a set of parameters:

$$t_{new} = f(t_{original}, \text{parameters}) \tag{2}$$

The parameters are refined iteratively until the inconsistency is removed. This is schematically represented in Fig. 2.

Fig. 2(a) represents the concatenation of the original matrices to form $[Y_A, Y_B]$. Next, the time vector for Y_B is recalculated according to Eq. (2) to yield t_{new}. Fig. 2(b) is an attempt to represent the procedure graphically. In the example the time vector is stretched and displaced. This does not affect the number of elements in the vector, only their values. The spectra in Y_B are moved along the time axis with respect to the original time axis of the matrix Y_A. Again, the number of rows in Y_B does not change. The next step is to select the common parts of the two matrices according to t_{new}^*; we call them Y_A^*, and Y_B^*, see Fig. 2(c). The individual spectra contained in Y_A^*, and Y_B^* do not align along the time axis and therefore one of them needs to be replaced by an interpolated version, to correspond with the common time vector t_{new}^*.
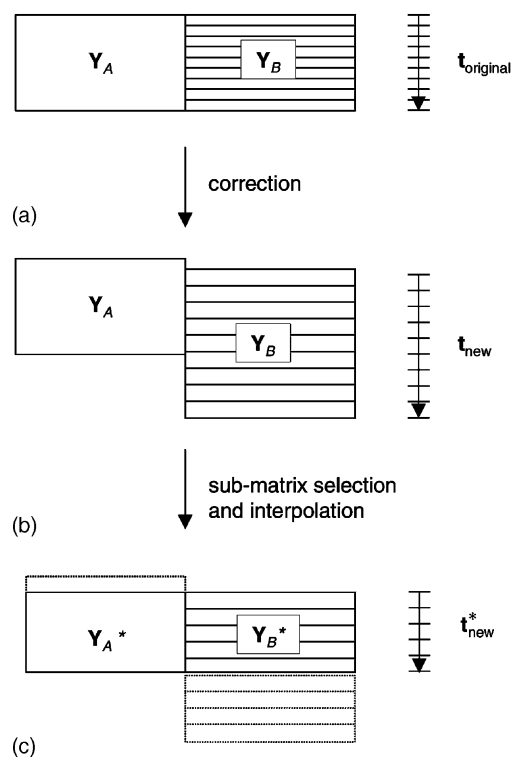

Fig. 2. (a) Original arrangement of the concatenated matrix $[Y_A, Y_B]$. (b) After correction of the vertical scaling (e.g. retention time) the alignment between Y_A and Y_B is destroyed. (c) Only the common parts, Y_A^* and Y_B^*, are concatenated.

The resulting matrix $[Y_A^*, Y_B^*]$ is then subjected to a singular value decomposition to determine its rank. If the adjustment of the time vector for Y_B^* is correct, the concentration profiles C_A^* and C_A^* will align and the rank of $[Y_A^*, Y_B^*]$ will be back to *m*. Thus, the task is to find that correction of the time vector *t* for which the resulting rank of $[Y_A^*, Y_B^*]$ is *m*.

Several aspects need closer attention: (a) correction functions, (b) rank determination, (c) interpolation, (d) mean centering, (e) minimisation algorithms, (f) types of instrumental irregularities, and (g) possible generalisations.

2.1. Correction functions

Often there are no theoretically defined functions for the appropriate correction and therefore polynomials of any degree are obvious choices. Logarithmic or exponential distortions are other possibilities.

$$\boldsymbol{t}_{\text{new}} = p_0 + p_1 \boldsymbol{t}_{\text{original}} + p_2 \boldsymbol{t}_{\text{original}}^2 + \cdots + p_i \boldsymbol{t}_{\text{original}}^j \tag{3}$$

or

$$\boldsymbol{t}_{\text{new}} = p_0 + p_1 e^{(p_2 \boldsymbol{t}_{\text{original}})} \tag{4}$$

The parameters, $p_0, p_1, \ldots$ are collected in a parameter vector $\boldsymbol{p}$.

There is no limit to the type of correction function that could be used. Naturally, the higher the number of parameters to be fitted, the more difficult and less robust the optimisation will be.

2.2. Analysis of rank as function of parameters

There is an immense literature on the determination of the rank of a matrix of measurements [18]. Limited exploration of several different algorithms showed no significant advantages for any of them. We decided to use percent residual variance (%var), which is well-recognised [15] and makes intuitively much sense. Using eigenvalues or singular values directly, as suggested originally for RAFA [13,14], is possible, however the variable amount of overlap between $\boldsymbol{Y}_A^*$ and $\boldsymbol{Y}_B^*$ during the fitting of $\boldsymbol{p}$ makes the comparison of %var more reliable, as the variable row dimension of $[\boldsymbol{Y}_A^*, \boldsymbol{Y}_B^*]$ is taken into account.

The concatenated matrix $[\boldsymbol{Y}_A^*, \boldsymbol{Y}_B^*]$ is decomposed according to the singular value decomposition [19].

$$[\boldsymbol{Y}_A^*, \boldsymbol{Y}_B^*] = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V} + \boldsymbol{R} \tag{5}$$

Fig. 3. The logarithm of the percent residual variance (%var) retaining one up to six eigenvectors (Eq. (5)). There is a distinct minimum for the inclusion of four and more eigenvectors at the parameter value of −0.149.

$\boldsymbol{R}$ is calculated for different numbers of retained vectors in $\boldsymbol{U}$, $\boldsymbol{S}$, and $\boldsymbol{V}$, and its percent residual variance is calculated for all numbers of eigenvectors and eigenvalues retained [18].

For graphical purposes, these values can be plotted as a function of the parameters, see Fig. 3 for a one-parameter plot and Figs. 4 and 9 for applications of two parameters. Figs. 3 and 4 relate to the same chromatographic data set (flow rate 1.25 ml min^{-1}, temperature 25 and 27.5 °C), which will be discussed
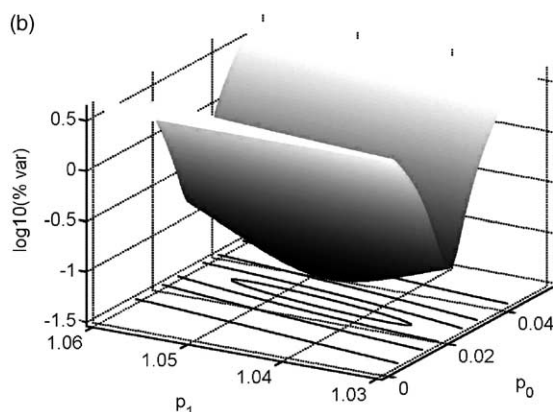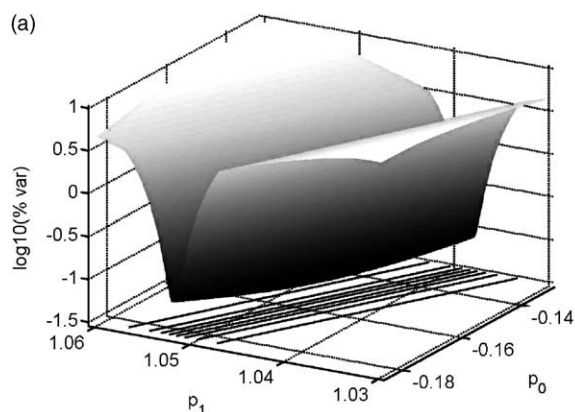

Fig. 4. The logarithm of the percent residual variance (%var) retaining four eigenvectors as a function of two parameters. (a) Without mean centering, the surface features a very narrow, steep and diagonal valley. (b) After mean centering, due to strongly reduced correlation between the parameters, the rank landscape is much more amenable to traditional optimisation routines.
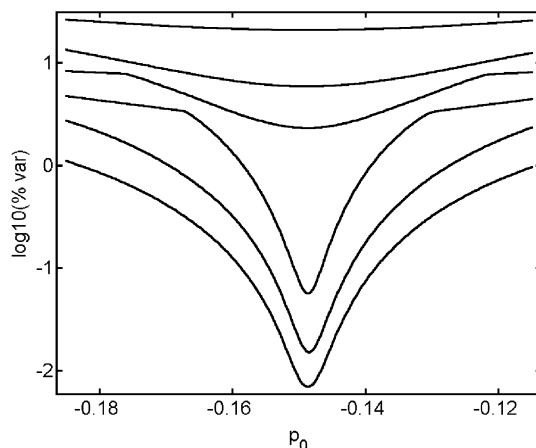
later in the experimental section. The correction function is a polynomial of degree one, $t_{\text{new}} = p_0 + p_1 t_{\text{original}}$. Fig. 3 is a one-dimensional slice through the minimum of Fig. 4a. The parameter p_1 has been fixed to its optimum value and the log (%var) is plotted as a function of the other parameter p_0. For the value of $p_0 = -0.149$ min there is a clear minimum of the rank to a value of four. This is best indicated by the minimum of the variance if four eigenvectors/eigenvalues are included ($m = 4$). Incorporation of additional eigenvectors shows essentially an identical minimum, thus we observe good robustness with respect to overestimation of the number of significant factors.

Functions of two or more parameters can be treated analogously. For two parameters the result is a surface (Figs. 4 and 9). For more parameters graphical display is impossible.

2.3. Interpolation

As indicated in Fig. 2, the time vectors for the two data matrices Y_A^* and Y_B^* must match after the correction. As continuous corrections are allowed, interpolation must be performed at all individual wavelengths for one of the two matrices to ensure identical times for the two sets of absorption spectra [12]. Interpolation is not required if only whole step shifts are permitted [15]. For all our experiments, we used polynomial interpolation of degree two or three, calculated for a wide range of points (between 5 and 21), depending on the relative density of points with respect to the structure of the data. This is essentially a Savitzky–Golay algorithm [20], used for interpolation instead of smoothing. It is important to match the degree of the polynomial with the number of data points in such a way that the resulting distortions are minimal [19]. Any alternative interpolation algorithm can be used. It is noteworthy that the computation for the interpolation steps can constitute a major time requirement of the fitting process of the parameter vector **p**. Naturally, this depends heavily on the order of the polynomial used.

2.4. Mean centering

The correlation between the parameters can be very high and, generally, mean centering of the retention time vector will reduce this correlation dramatically.

Mean centering will not affect the final result of the analysis. The effect of mean centering on the shape of the minimum as a function of the parameters is shown in Fig. 4. All calculations in this contribution have been performed on mean-centered data, however, for a more direct representation, the graph in Fig. 3 has been calculated using the original matrices.

2.5. Optimisation routines

Instead of inspection of graphical results such as displayed in Figs. 3, 4 and 9, any non-linear optimisation routine can be used to localise the optimum. This is particularly important for cases with more than two parameters, as these cannot be represented graphically. Algorithms range from the simple Simplex algorithm to more sophisticated gradient techniques. We used the MATLAB$^{\circledR}$ 'fminsearch' function [17], which is an advanced simplex algorithm. Any other non-linear optimisation algorithm could be used equally well; we have not investigated their relative merits. Initial guesses and/or boundaries for the parameters should be made sensibly. If initial guesses are too far out, most algorithms will fail to converge.

It has to be noted that there always is the possibility of one or several local minima. A typical example would be the partial alignment of only a selection of concentration profiles in chromatography. This situation is shown in Fig. 5. Graphical inspection of the result will immediately reveal such a condition.
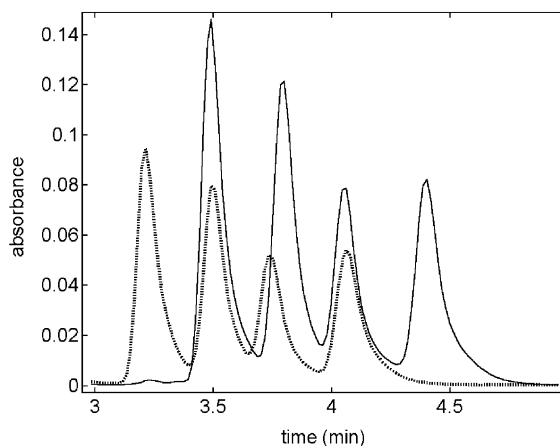

Fig. 5. Partial alignment of the wrong chromatographic peaks due to a local minimum.
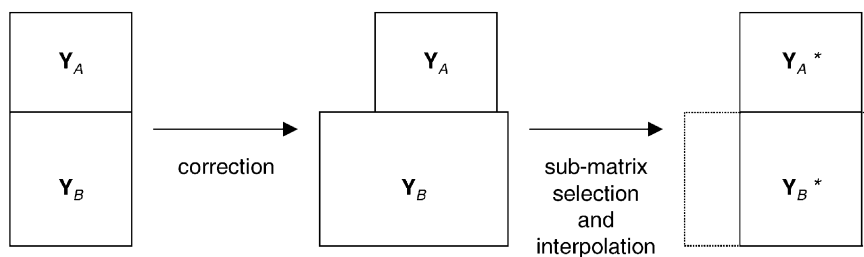
Fig. 6. Correction and interpolation for inconsistencies along the row dimension (e.g. wavelength shifts). Compare with Fig. 2.

2.6. Types of inconsistencies

Retention time distortions due to column aging or temperature drift in chromatography are typical examples of instrumental problems that can be rectified by the proposed method. They generally will require a correction that is more complex than just a constant shift of all retention times throughout the complete chromatogram [15,16].

Wavelength shifts are less commonly encountered, however, they have been observed [9] and we will discuss a real industrial example later in the experimental section. Wavelength corrections are done along the row dimension of the matrices and thus the concatenation needs to be done vertically. Fig. 6 illustrates the situation, which displays both the similarity and the difference to the one represented in Fig. 2.

Similar to retention time errors, the rank of the vertically concatenated matrix $[Y_A^*; Y_B^*]$ is equal to the sum of the two individual ranks, if there is a distortion along the wavelength scale. Correction of the wavelength axis will remove the inconsistency and the rank of $[Y_A^*; Y_B^*]$ will be the same as the rank of the individual matrices.

A very common type of inconsistency found in absorption spectroscopy is of a different nature. Small positioning errors of the absorption cell will result in small changes in the zero reading, resulting in baseline shifts [21]. This is mainly due to small differences in the reflections of the light beam at the different internal and external surfaces of the absorption cell. Usually the errors are small and are not further considered, but they can affect the trilinearity or other relevant properties of the data structures. The general equation for such a situation is

$$Y = Y + \text{shift} \tag{6}$$

If a baseline shift occurs between two independent measurements, the rank generally only increases by one. Also, there are no interpolations and other adjustments of the data required. For such a simple shift, the rank analyses are done on the concatenated matrices $[Y_A; Y_B + \text{shift}]$, Note that in this case horizontal concatenation will have an identical result. It is possible to correct for more complicated 'shifts', e.g. baseline drift with time. In such cases more complex correction functions will be applied, e.g. a linear or quadratic increase of the absorption shift with time. As before any number of parameters can be fitted.

2.7. Generalisations

The Figs. 1, 2 and 6 seem to imply that the individual matrices Y_A and Y_B need to be of identical dimensions. This is not at all the case. For horizontal concatenation, the only prerequisite is that there is a significant range of common measured and interpolated times between the matrices Y_A^* and Y_B^*, i.e. the matrix $[Y_A^*, Y_B^*]$ is covering a significant part of the overall elution. It is also not required that all components exist in both datasets. Further, there is no requirement for identical wavelength ranges (horizontal scale). In fact, the two wavelength ranges can be completely independent i.e. without any overlap. One matrix might be measured in the UV–VIS while the other is covering NIR wavelengths. This rather surprising fact can be explained in the following way: all columns of the matrices Y_A^* and Y_B^* are linear combinations of the columns of the concentration matrices C_A and C_B. After successful correction, the columns of C_A and C_B are multiples of each other ($C_A = C_B D$, where D is a diagonal matrix which even can have zero diagonal elements). Thus, all

columns of both Y_A^* and Y_B^* are linear combinations of the same m columns of the concentration matrices and thus the rank of $[Y_A^*, Y_B^*]$ is m, irrespective of the wavelengths at which they are measured.

Vertical concatenation is governed by corresponding rules. In this case the concentration profiles can be different, e.g. acquired at different temperatures and times, etc. The prerequisite here is a significant overlap along the row dimension or wavelength axis.

It is far from straightforward to define what 'significant' means in this context. It strongly depends on many different aspects such as number of components, signal to noise ratio, similarity of concentration profiles and absorption spectra. It is beyond the scope of this paper to investigate these statistical aspects thoroughly.

3. Experiments and results

3.1. Chromatography

Chromatograms were acquired on a Waters 600E HPLC system with a Waters 996 diode array detector, using a 25 cm C18 reversed phase column. Spectra were recorded at 1 s intervals from 250 to 400 nm.

In order to mimic the ageing of a chromatographic column, temperature and flow rates were varied systematically: the temperature was varied between 25 and 40 °C and the flow rate was either 1 or 1.25 ml min^{-1}. Mixtures of anthracene, pyrene, fluorene, and fluoranthene were analysed in acetonitrile. Fig. 7 shows the concentration profiles for two chromatograms measured at 25 and 27.5 °C before and after application of RAC.

The correction function was $t_{\text{new}} = p_0 + p_1 t_{\text{original}}$. Fig. 8 displays the values of the parameters p_0 and p_1 as a function of the temperature. All the analyses were done with respect to the 'standard' measurement at 25 °C and 1 ml min^{-1} flow rate. The parameter p_1 is the factor by which the time vector of the higher temperature (or flow rate) chromatograms is multiplied to relate them to the 'standard'. There is a good linear relationship between this factor and the temperature and this is independently the case for both flow rates. There is a strong negative relationship between the shift parameter p_0 and temperature and, interestingly, it is independent of the flow rate. Note, that the parameter p_0 for a non-mean centered time scale is represented.

It is not the intention to investigate here the relationship between retention time and temperature and flow rate. However, the neat relationship shown in Fig. 8 is a clear indication for the reliability and robustness of the procedure. The repetition of the 'standard' chromatogram resulted in values of 0.997 for p_1 and 0.027 min for p_0. Both are very near the ideal values of 1 and 0, and p_0 is close to the time resolution (0.017 min) of the detection system.
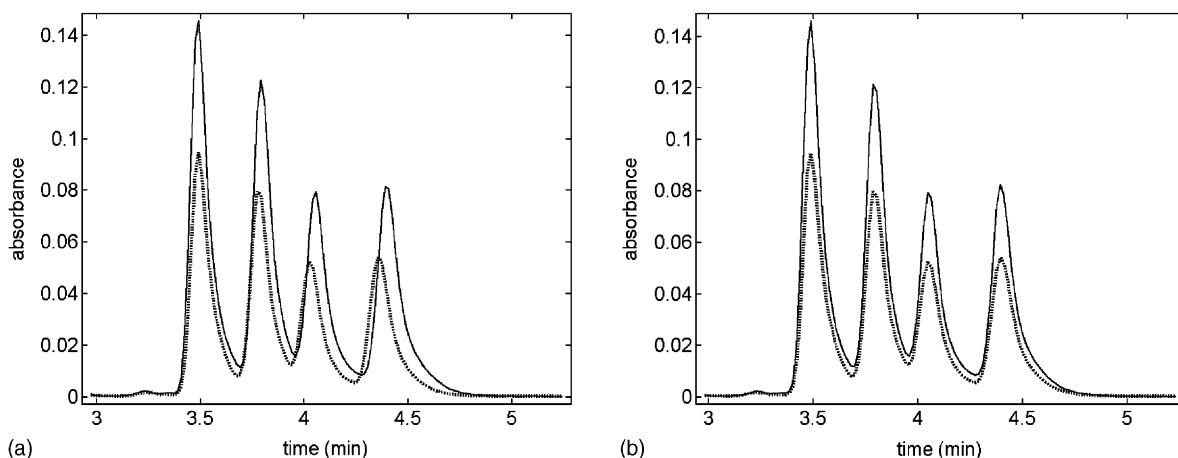

Fig. 7. The 265 nm traces of chromatograms at 25 °C (—) and 27.5 °C (···), (a) before and (b) after the chromatogram at 27.5 °C has been corrected by RAC using a first order polynomial.
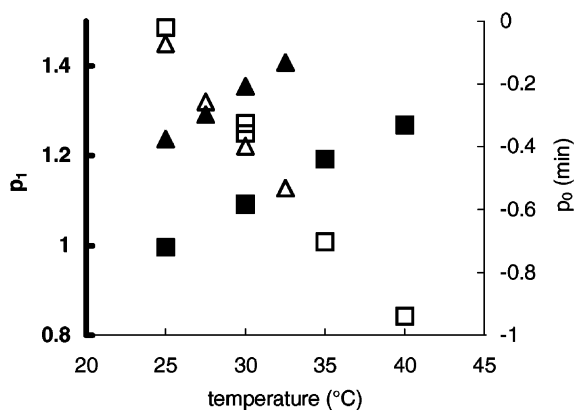
Fig. 8. The calculated stretching parameter p_1 (■, ▲) and shift parameter p_0 (□, △) versus temperature. Squares show the results for flow rates of $1 \, \text{ml min}^{-1}$, triangles the ones at $1.25 \, \text{ml min}^{-1}$.

Fig. 9. The logarithm of the percent residual variance (%var) as a function of shift parameter p_0 and the redundant stretching parameter p_1 for the distillation NIR data.

3.2. Distillation

The datasets were acquired by a custom-built thermostatted transmission cell (sapphire windows), installed into a bypass flow of an industrial distillation set-up. The NIR spectra were recorded on a Bruins Omega 20 spectrometer, attached to the flow cell via optical fibres.

Experimental inconsistencies were observed which tentatively were assigned to problems due to a lamp change in the instrument. We re-analysed two datasets, one acquired before and one after the lamp change. Ninety one spectra formed the matrices of Y_A and Y_B; they were measured at 2 min intervals in the wavenumber range of $6400–7700 \, \text{cm}^{-1}$, at $3.2 \, \text{cm}^{-1}$ intervals.

RAC, run with a simple additive wavenumber shift (polynomial of degree 0), determined a displacement of $-6.32 \, \text{cm}^{-1}$ at a percent residual variance of 8.9×10^{-5}. Attempts to fit more complex functions such as a first order polynomial did not result in significant improvements. This is clearly indicated by Fig. 9, which shows the percent residual variance for a first order polynomial with a minimum for parameters $p_1 = 0.997$ and $p_0 = -6.29 \, \text{cm}^{-1}$ at a percent residual variance of 8.8×10^{-5}.

A geometrical displacement of the source is consistent with a uniform wavenumber shift. It is feasible that a uniform shift of the wavelength instead of the wavenumber would be appropriate for theoretical rea-
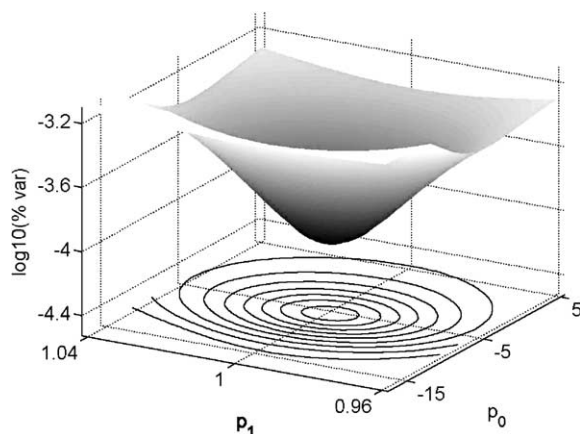
sons. Due to the narrow range of wavenumbers these two possibilities could not be distinguished.

3.3. Titrations

The effects of absorption cell positioning irregularities were investigated in a series of titrations. Cell position influences absorption readings in rather complex ways [21]. In the following, we will model the errors resulting from irreproducible cell positioning by the simple addition of a shift parameter, which is added to all absorption readings throughout the complete titration. The complexation of Cu^{2+} with dien (diethylenetriamine, or 1,4,7-triazaheptane) was investigated in aqueous solution between pH 3 and 11. The required titrations were performed directly in a standard 1 cm absorption cell equipped with pH electrode, magnetic stirrer and inlet tube for base addition delivered by an automatic burette. Burette (Metrohm 695), pH meter (Metrohm 605) and spectrophotometer (Philips-PU-8800) were all interfaced to a PC and the titrations were performed under complete computer control. For details about the titration set-up see [22]. Details on the titrations, analysis and results for this system were published elsewhere [23].

The analysis consists in the determination of the formation constants (β values) of the different complex species formed during the titrations as well as the determination of the molar absorption spectra of

Table 1
Global analyses of the complexation of Cu^{2+} by dien

Method	Species				
	CuL^{2+} $(\log\beta_{110})$	$CuLH^{+}_{-1}$ $(\log\beta_{111})$	CuL_2H^{3+} $(\log\beta_{121})$	CuL_2^{+2} $(\log\beta_{120})$	σ_Y
(a) Individual titrations	16.549(40)[a]	7.399(246)	30.430(312)	21.766(237)	1.13×10^{-3}
(b) Set of 3, local	16.548(3)	7.300(39)	30.390(22)	21.624(56)	1.34×10^{-3}
(c) Correction set of 3, local	16.548(3)	7.298(39)	30.389(22)	21.616(57)	1.34×10^{-3}
(d) Set of 3, global	16.516(43)	7.504(184)	29.921(469)	21.628(190)	1.75×10^{-2}
(e) Correction set of 3, global	16.540(7)	7.339(31)	30.312(38)	21.595(33)	2.77×10^{-3}

[a] Logarithms of the equilibrium constants are given with twice their standard errors (units of the least significant digit).

the absorbing complexes. It is the computation of the absorption spectra, which is directly affected by the influence of baseline differences between individual titrations. In order to investigate the influence of baseline errors, a series of three titrations was analysed in four different modes: (a) each titration was analysed individually; (b) the set of titrations was linked and analysed globally [23] but the molar absorption spectra were fitted individually for each titration, thus three sets of absorption spectra result ('local' mode); (c) the repetition of (b) after application of RAC; (d) global analysis but only one set of common absorption spectra was fitted to all titrations ('global' mode); and (e) repetition of (d) after application of RAC.

The results are given in Table 1 and can be summarised in the following way: the quality of fits, as represented by the σ_Y values (standard deviation of the residuals), is naturally best for the initial 'individual titrations' mode as this mode has the highest number

of variable parameters to be adjusted; they are three sets of formation constants and three sets of molar absorption spectra (mode (a) in Table 1). The disadvantage of this mode is the uncertainty of the parameters fitted to the individual titrations and this is reflected in the relatively large standard deviations for the formation constants. In the following, we will compare and relate the other results to this original analysis. The global fit with 'local' spectra (b) is marginally worse, resulting in about 15% higher σ_Y. In this mode there is one set of formation constants and three sets of absorption spectra. The non-linear parameters are essentially the same but they are much better defined, as shown by strongly reduced standard deviation (typically by a factor of 10). The individual absorption spectra are significantly different, Fig. 10(a). Prior correction with RAC in (c) does not affect the results with respect to the equilibrium constants. However, the individual molar absorption spectra now are much more
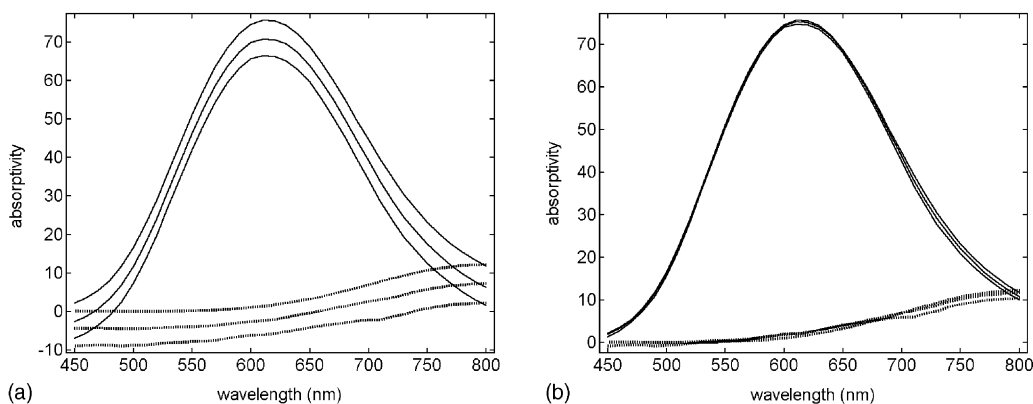

Fig. 10. Calculated local absorptivity spectra of Cu^{2+} ($\cdots$) and $[Cu(II)(dien)]^{2+}$ (—) (a) before and (b) after the application of RAC (Eq. (6)) to the measured absorbance data.

similar, Fig. 10(b). Analyses with 'global' spectra in (d) (one set of formation constants and one set of spectra) of uncorrected data result in different (wrong) formation constants, larger σ_Y (typically a factor of 10) and thus larger standard deviations. This is clearly the result of the baseline problems, which are not being accounted for in this mode. Note, that global analysis with 'global' spectra is potentially the most powerful [23] and is the preferred mode of analysis. After application of RAC in (e) the parameters are correct and their standard deviations minimal. The σ_Y values are marginally larger due to not quite perfect correction.

Fig. 10 demonstrates the efficiency of the corrections: it shows the calculated spectra for the individual titrations in 'local' mode before (mode (b) in Fig. 10(a)), and after the corrections (mode (c) in Fig. 10(b)). The result for mode (e) is not shown in the figure, it is visually indistinguishable from the result of mode (c).

4. Conclusions

A generally applicable method for the correction of instrumental inconsistencies of several types has been introduced. In contrast to most earlier algorithms which only covered a uniform shift of retention time or wavelength [15], the proposed method can deal with multiparameter corrections and thus is much more versatile. The method does not rely on the identification of individual peaks and subsequent correction based on matching the retention times of selected individual peaks [11,12]. The complete datasets are matched and thus chromatograms or any other datasets can be severely overlapped, without obvious peak positions for any of the components. It is an essential feature of factor analysis based methods, that complete matrices are analysed, rather than specifically chosen attributes (e.g. at particular wavelengths, etc.). Operator input is minimal, nevertheless, visual inspection of the results is important, as local minima can result, (Fig. 5). Computation times are acceptable, for all examples they ranged from seconds to minutes on a 500 MHz PC.

An additional strength of the method is that the chemical compositions of the samples resulting in the two datasets do not need to be identical. Relative and absolute concentrations can be completely different,

even the absence of one or several components in any of the datasets will not prevent successful analysis, provided there is still significant communality.

If there are more than two matrices that need to be corrected, all matrices are treated with respect to one pre-chosen 'standard'. Naturally this standard needs to be selected carefully, it should be the one with the least problems. It is theoretically feasible to globally adjust all matrices, however, preliminary attempts at such analyses resulted in no added robustness and thus no advantage can be gained.

There are also limitations to the applicability of RAC or any alternative correction based on a functional relationship. The individual components must show a similar behaviour with respect to the inconsistencies, such as temperature, as only one correction function is applied to all of them. Thus, for the example of chromatography, the order of elution must be maintained; also peak broadening must follow the applied correction function. To summarise, any correction will result in an improvement for the global analyses but the correction will not necessarily be perfect.

Acknowledgements

References

[1] A.K. Smilde, R. Tauler, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, Anal. Chem. 66 (1994) 3345.
[2] R. Bro, Chemometr. Intell. Lab. Syst. 38 (1997) 149.
[3] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.
[4] E. Sanchez, B.R. Kowalski, J. Chemometr. 4 (1990) 29.
[5] P. Bugnon, J.C. Chottard, J.L. Jestin, B. Jung, G. Laurenczy, M. Maeder, A.E. Merbach, A.D. Zuberbühler, Anal. Chim. Acta 298 (1994) 193.
[6] T. Rotunno, Data Handl. Sci. Technol. 15 (1995) 85.
[7] S.C. Rutan, E. Bouveresse, K.N. Andrew, P.J. Worsfold, D.L. Massart, Chemometr. Intell. Lab. Syst. 35 (1996) 199.
[8] R. Goodacre, D.B. Kell, Anal. Chem. 68 (1996) 271.
[9] H. Witjes, M. van den Brink, W.J. Melssen, L.M.C. Buydens, Chemometr. Intell. Lab. Syst. 52 (2000) 105.
[10] Y.D. Wang, B.R. Kowalski, Anal. Chem. 65 (1993) 1174.
[11] D.D. Chilcote, C.D. Scott, Anal. Chem. 45 (1973) 721.
[12] R. Andersson, M.D. Hämäläinen, Chemometr. Intell. Lab. Syst. 22 (1994) 49.

[13] C.-N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 50 (1978) 1108.

[14] A. Lorber, Anal. Chem. 57 (1985) 2395.

[15] B.J. Prazen, R.E. Synovec, B.R. Kowalski, Anal. Chem. 70 (1998) 218.

[16] B. Grung, O.M. Kvalheim, Anal. Chim. Acta 304 (1995) 57.

[17] Matlab R11.1. The Mathworks Inc., Natick, MA, 1999.

[18] E.R. Malinowski, Factor Analysis in Chemistry, Wiley Interscience, New York, 1991.

[19] W.H. Press, W.T. Vetterling, S.A. Teukolsky, B.P. Flannery, Numerical Recipes in C, Cambridge University Press, Cambridge, 1995.

[20] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627.

[21] L.D. Rothman, S.R. Crouch, J.D. Ingle Jr., Anal. Chem. 47 (1975) 1226.

[22] H. Gampp, M. Maeder, A.D. Zuberbühler, Trends Anal. Chem. 7 (1988) 111.

[23] R.M. Dyson, S. Kaderli, G.A. Lawrance, M. Maeder, A.D. Zuberbühler, Anal. Chim. Acta 353 (1997) 381.