ELSEVIER

Analytica Chimica Acta 384 (1999) 71–81

ANALYTICA CHIMICA ACTA

Two-way, unfolded three-way and three-mode partial least squares calibration of diode array HPLC chromatograms for the quantitation of low-level pharmaceutical impurities

Konstantinos D. Zissis[a], Richard G. Brereton[a,*], Samantha Dunkerley[a], Richard E.A. Escott[b]

[a]*School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK*
[b]*SmithKline Beecham Pharmaceuticals, Old Powder Mills, Near Leigh, Tonbridge, Kent TN11 9AN, UK*

Received 10 August 1998; received in revised form 16 November 1998; accepted 27 November 1998

Abstract

A series of two-way diode array chromatograms were recorded at 0.1–0.5% and 1–5% of 3-hydroxypyridine impurity co-eluting with 2-hydroxypyridine, recorded at pH 4.9 (good resolution) and 5.0 (poor resolution). Four methods for PLS calibration, namely, using summed spectral profiles, summed elution profiles, unfolded three-way PLS and true three-way PLS were applied to the datasets, both using autopredictions and cross-validation. It was found that it was possible to accurately quantify low levels of impurities. Three-way methods often performed worse than two-way methods using the summed spectral profile probably due to irreproducibility of elution times. © 1999 Published by Elsevier Science B.V. All rights reserved.

Keywords: HPLC; Partial least squares; Three-mode calibration; Impurity monitoring

1. Introduction

DAD–HPLC is a common method for impurity monitoring in chemical processes [1,2]. For example, a batch of drugs may contain a small quantity of by-product, such as an isomer. It is important not only to detect whether a peak is impure, but also to quantify the amount of such impurity. Sometimes the nature of the impurity is known, and the main aim is to determine the proportion. Quality control of products dictates upper and lower acceptable limits. A number of approaches could be employed, but when the impurity is at very low levels (less than 1%), is closely eluting and has similar spectral characteristics to the main component, chemometric approaches are an important aid.

Chromatographic data may be regarded as being multivariate. There are a variety of regression methods for the determination of concentrations of mixture [3–5], but PLS (partial least squares) is one of the most widespread. For single wavelength chromatography, the application of PLS is fairly straightforward, as a data matrix can be formed with elution time as the rows and sample number as the columns. However, when the data themselves have been obtained via coupled chromatography, the raw data are in the form of a tensor, with three dimensions, namely sample number, elution time, and spectral wavelength. This means that it is no longer a straightforward task to apply PLS. The data are commonly called three-way

*Corresponding author.

0003-2670/99/$ – see front matter
PII: S0003-2670(98)00844-7

data, and a variety of approaches such as three-way PLS [6–8], or unfolded PLS [9,10], have been proposed over the past few years to handle such data. In order to take advantage of the extra dimension such approaches are essential.

In this study we report the experimental use of a number of PLS methods for the quantification of small levels of impurities in two-component mixtures. The ability of PLS to determine the amounts of the impurities from known cross-validated calibration sets is discussed. It is important to recognise that the apparent quality of calibration models themselves is related to the quality of calibration experiments. This is especially crucial when monitoring very low levels of impurities, as often elaborate weighing and dilution schemes are required, introducing further possible experimental errors. For example, a chromatogram that is designed to contain 0.1% impurity may actually contain 0.11% impurity due to errors in sample preparation. A calibration model may predict that this sample contains 0.108% impurity. In fact, the calibration model has only 0.002% prediction error, but it may appear that the error is 0.008% or fourfold higher simply because of the method by which the samples are prepared. There is no real way of avoiding this apart from the use of plenty of replicates, including all the dilutions, and a reasonably large number of points in the calibration dataset.

This paper shows that it is possible to use PLS and chemometrics to predict the amount of an impurity to a high level of accuracy, even at very low percentages (<0.5%).

2. Experimental

2.1. Compounds and chemicals

The compounds used in this study were 2-hydroxypyridine (Acros Chemicals, 97%, New Jersey, USA) and 3-hydroxypyridine (Acros Chemicals, 98%, New Jersey, USA), respectively. Their structures are shown in Fig. 1(a) and their normalised experimental spectra (pH 5.0) in Fig. 1(b). The spectra were obtained from the chromatographic analysis of the pure compounds using a mobile phase composition of 98:2 0.05 M ammonium acetate pH 5.0 (98%, SigmaUltra, Sigma, St. Louis, USA):methanol (HPLC

2-hydroxypyridine 3-hydroxypyridine

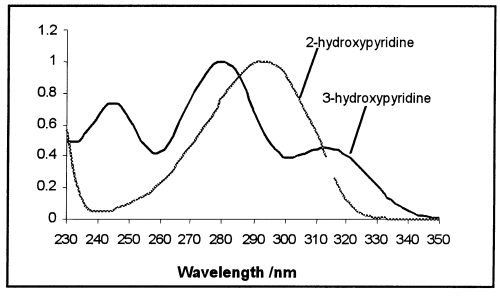

Fig. 1. (a) Structures of 2- and 3-hydroxypyridines; (b) normalised experimental spectra of 2- and 3-hydroxypyridines at pH 5.0.

grade, Rathburn Chemicals, Walkerburn, UK). The pH of the buffer solution was adjusted to the relevant value using a combination of a 25% acetic acid and a 10% ammonia solution (Sigma, Poole, UK), respectively. For the preparation of both buffer solutions and samples, deionised water was used exclusively, from a Milli-Q filtration unit (Millipore, MA, USA). Note that both chromatography and spectroscopy are pH dependent [11].

2.2. HPLC conditions

All chromatography was carried out using a Waters HPLC chromatograph (600S Controller, Model 616 pump, Model 717 autosampler) with a C_{18} reversed phase symmetry column (Waters, Milford, MA), (3.5 μm, 100 mm×4.6 mm) at ambient temperature. The mobile phase consisted of 98:2 0.05 M ammonium acetate (buffered to the pH of interest): methanol. The flow rate was set at 1 ml min^{-1} and 20 μl of the solution was injected. UV detection was performed using a Model 996 photodiode-array detector, and UV spectra were collected between the wavelength range 230–350 nm. The digital resolution was 1 s in time and 2.4 nm in wavelength. For all analyses, the data matrix for each individual run was set at 130 points in the time direction and 51 points in the wavelength direction.

2.3. Experimental design and dilution scheme

The pure compounds were run at pHs 4.9 and 5.0, respectively. At pH 4.9, 3-hydroxypyridine eluted 20 s prior to 2-hydroxypyridine (3.48 versus 3.81 min), whereas at pH 5.0, the peaks eluted very closely with each other (6 s apart), with 2-hydroxypyridine now eluting before 3-hydroxypyridine (3.38 versus 3.50 min). A schematic representation of the normalised elution profiles of a 1:1 mixture of 2- and 3-hydroxypyridines at the two pH values is given in Fig. 2. 3-Hydroxypyridine was treated as the minor impurity and was studied at two different percentage ranges which span from 0.1–0.5% in increments of 0.1%, to 1–5% in increments of 1%. All designs were based on five main concentration levels and a total number of 14 experiments. Replication was deemed important, with the central point replicated three times.

For the two compounds at each pH, stock solutions were prepared at approximately 50 mg per 100 ml for 2-hydroxypyridine (stock solution A) and 30 mg per 100 ml for 3-hydroxypyridine (stock solution B). For the 1–5% concentration range experiments, solution B was diluted 20 times by taking 5 ml and making the total volume 100 ml (solution C), using the mobile phase composition described in Section 2.1. Three ml of stock solution A was then taken and added to 1 ml of solution C (1%), 2 ml (2%), 3 ml (3%), 4 ml (4%) and 5 ml (5%). The total volume was made up to 10 ml, by adding the appropriate amount of mobile phase each time, as shown in Table 1.

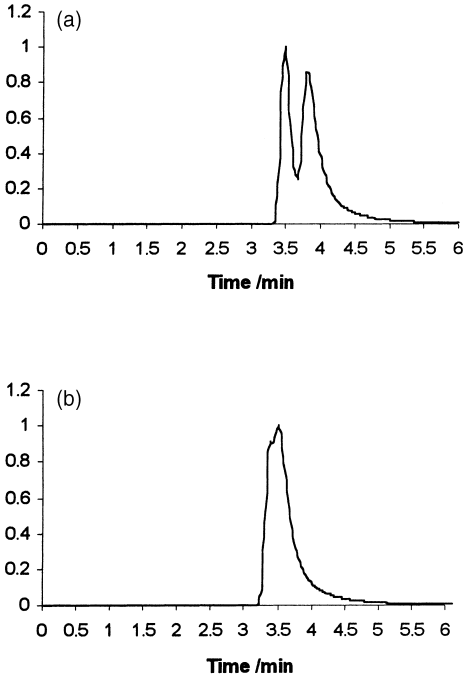

Fig. 2. (a) Normalised elution profiles of pure compounds at pH 4.9; (b) normalised elution profiles of pure compounds at pH 5.0.

All 14 samples were prepared using replicate dilutions, but no instrumental replicates were performed. Preliminary experiments suggested that instrumental reproducibility was much greater than the sample preparation reproducibility. Repeat calibration of pipettes suggested a typical precision of 0.3% using six replicated calibrations, and 0.1% for the volumetric flasks, with accuracies of 0.5% and 0.3%, respectively; in addition there is expected to be a small weighing error. However, the overall sample preparation error is under 1%. Note that when least squares regression is performed, the apparent error can be heavily influenced by one outlying sample, and the method of preparation can result in errors of a few percent that may influence the apparent prediction error. The overall number of replicates at each concentration range is given in Table 2. For the 0.1–0.5% experiments, the same sequence of dilutions was followed, but this time stock solution D was used instead of C. This was 10% the concentration of solution C, and was prepared by taking 10 ml from solution C and diluting to 100 ml in another volumetric flask.

3. Methods

Calibration is one of the most important techniques in the area of analytical chemistry [12–15]. It involves

Table 1
Dilution scheme for the 1–5% 3:2 hydroxypyridine calibration experiments

Percentage of 3-hydroxypyridine	Stock solution A (ml)	Stock solution C (ml)	Solvent (ml)
1	3	1	6
2	3	2	5
3	3	3	4
4	3	4	3
5	3	5	2

Table 2
Two different experimental designs for the minor impurity

No. of experiments	Percentage of 3-hydroxypyridine	
1	0.1	1.0
2	0.1	1.0
3	0.2	2.0
4	0.2	2.0
5	0.2	2.0
6	0.3	3.0
7	0.3	3.0
8	0.3	3.0
9	0.3	3.0
10	0.4	4.0
11	0.4	4.0
12	0.4	4.0
13	0.5	5.0
14	0.5	5.0

recording a series of responses (usually absorbances of compounds for HPLC) and relating to concentrations. A model is then built which can be subsequently used to predict concentrations of unknown compounds. Calibration can be performed in various modes namely: (i) univariate; (ii) two-way PLS; (iii) three-way unfolded PLS; and (iv) three-mode PLS. In this work, PLS models are established for the last three methods and the errors in the predictions of low-level compound concentrations were compared to determine the most efficient method. The raw data for all methods can be represented by a tensor $\underline{\boldsymbol{Z}}$, with dimensions $M \times I \times J$ (samples×time×wavelength) which can be rearranged to meet the dimensional demands of each method (Fig. 3). Following PLS decomposition, the resultant scores, $\boldsymbol{t}$, (correspond to elution profiles) and loadings, $\boldsymbol{p}$, (correspond to spectral profiles) are extracted for each technique and can be seen in Fig. 4 (K PLS components are extracted).

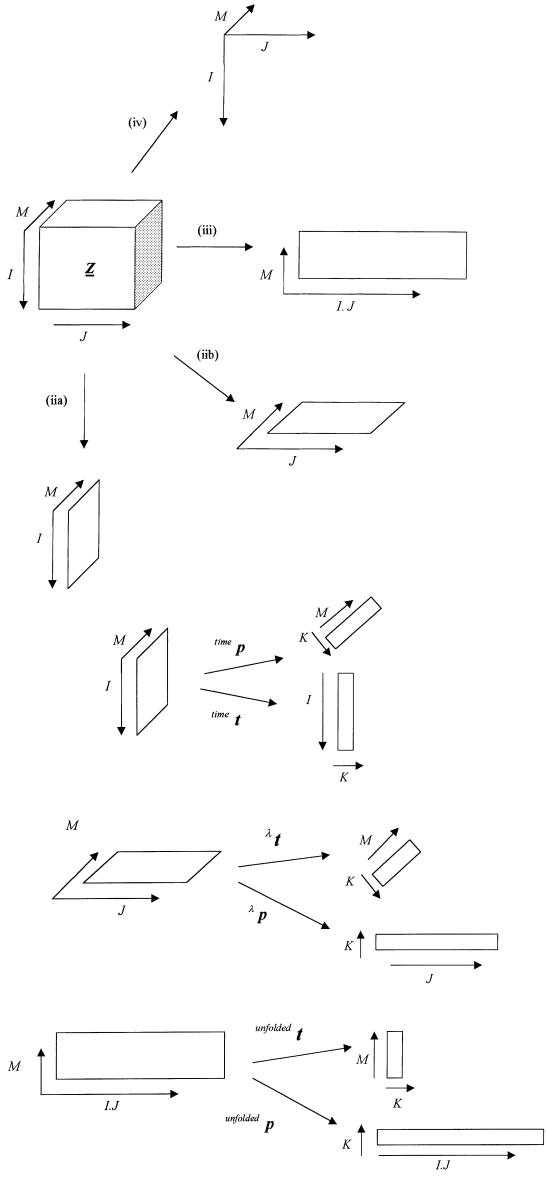

Fig. 3. Different ways of processing a three-way tensor $\underline{\boldsymbol{Z}}$.

3.1. Two-way PLS

Two-way PLS calibration has been extensively used in many areas of analytical chemistry for a substantial number of years [16–21]. Usually, an $\boldsymbol{X}$ matrix (often a matrix of absorbance values at successive time units and various wavelengths) is calibrated against a $\boldsymbol{y}$ vector (often a concentration vector for one of the compounds present in a mixture). In the case of HPLC, elution profiles in the time direction or spectral profiles in the wavelength direction of mixtures of compounds can be calibrated against compound concentration. In chemometrics, two main algorithms have been used, namely PLS1 and PLS2 [22]. The first one, described in detail in a paper by Wold [23], has been used in this study to perform all two-way PLS and unfolded three-way PLS calculations.

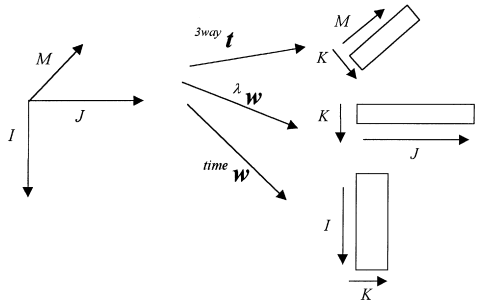

Fig. 4. Decomposition of each PLS method into scores and loadings matrices.

3.1.1. Elution profiles

The elution profile of a mixture of compounds, summed over all wavelengths, a_i, can be calibrated against compound concentration, $\boldsymbol{y}$, and

$$a_i = \sum_{j=1}^{J} x_{ij},$$

where x_{ij} is a point at time i and wavelength j. In total, I points in time are used for each time profile, and there are also J wavelengths. In this study, all wavelengths were used (J=51).

3.1.2. Spectral profiles

Alternatively, the spectral profile of a mixture of compounds, summed over all points in time, β_j, can be calibrated against compound concentration, $\boldsymbol{y}$ as

$$\beta_j = \sum_{i=I_1}^{I_2} x_{ij}.$$

In total, J points in wavelength are taken into consideration for each spectral profile. In this study, the number of points in time was restricted to 130 to make the calculations faster, and certain points before and after the elution of the peaks in the chromatograms were cut-off.

3.2. Three-way unfolded PLS

In unfolded three-way PLS calibration, tensor $\underline{\boldsymbol{Z}}$ is unfolded into a 2D $\boldsymbol{X}$ matrix. To do this, the rows of $\underline{\boldsymbol{Z}}$ are concatenated to give a row vector, so that the resultant two-way $\boldsymbol{X}$ matrix now has dimensions $M\times(IJ)$. Then, PLS is performed as in ordinary two-way PLS applications. The details of the method have been described in a previous paper [24].

3.3. Three-mode PLS

Genuine three-mode PLS calibration is a much more elaborate technique that has received a significant amount of attention in the past decade, mainly due to the availability of sophisticated modern analytical equipment. In three-mode PLS [6–8], the $\underline{\boldsymbol{Z}}$ tensor is decomposed into a set of triads, consisting of one scores vector $\boldsymbol{t}$, and two weight vectors $\boldsymbol{w}^j$ (wavelength dimension) and $\boldsymbol{w}^k$ (time dimension). The contribution to the 1st PLS component, $\hat{\boldsymbol{x}}_{ijk}$ is then expressed as $x_{ijk}=t_i w^j w^k$, and subtracted from $\boldsymbol{X}$ for subsequent PLS components. The description of the algorithm is explained in detail by Bro [25], and a simple graphical representation of the decomposition is given in Fig. 3(iv). The main advantage of three-mode PLS is that all the three vectors produced have maximum covariance with the unexplained part of the dependent variable. The algorithm also uses fewer parameters when compared to unfolded three-way PLS, which makes it much easier to visualise and interpret.

3.4. Autoprediction and cross-validation

3.4.1. Autopredictions

When a model is built from known compounds with defined concentrations, then the predictions for the concentrations of the compounds (autopredictions) are usually over-optimistic and result in a perfect fit with a minimal error. In this paper, PLS predictions (autopredictions) were first calculated for each of the 14 concentrations of 3-hydroxypyridine, by each PLS method separately, after extracting N PLS components successively. The root mean square error of all predictions, (RMSE), is subsequently calculated for each PLS method. RMSE values were calculated in mM and converted to the appropriate relative percentage error.

In total, % autoprediction errors, ap%, in concentration predictions for up to four PLS components are reported. The more the PLS components extracted, the lesser the error, since there are more terms in the model. However, it is essential to recognise that the reduction in error may be due to overfitting, so it is not

necessarily real. Hence, some form of validation is required to assess the quality of the model.

3.4.2. Cross-validation

As auto-predictions always overestimate the goodness of a model, a method needs to be applied to evaluate how well the particular model will predict the concentrations of unknown compounds. Testing, using an independent test set, as well as cross-validation, are two such methods. In this work, cross-validation has been chosen as the method of identifying how good a model is. In a previous paper [22], we advocate an independent test set because cross-validation provides an over-optimistic assessment of the model quality for poorly designed training sets. However, the case in this paper involves only one compound changing in concentration, and so there are no issues of orthogonality. Therefore, cross-validation is valid.

Several methods for cross-validation have been reported in the literature, but in this work, we focus on the method of taking one sample out at a time and using the model built by the previous samples to predict the concentration of the missing sample. Each of the M samples is then removed in turn and its concentration is predicted by the model formed by the remaining samples.

To test a model formed by a particular PLS method, a value of RMSECV was calculated. Once again, RMSECV values were calculated in mM and converted to the appropriate relative percentage, cv%.

3.5. Data pre-processing

3.5.1. Interpolation

Although the DAD instrument was set up to collect data points every 1 s for all runs, the precision of the time readings was not reproducible down to 0.1 s. For example, when one run would take down absorbance values at time points starting at 1.1, 2.1, 3.1 s at 1 s intervals, the next run may record readings starting at 1.3, 2.3 s and so on. To keep the values at the same comparable scale (to 1.0 s), they were interpolated using an in-house VBA macro.

3.5.2. Alignment

After interpolation was performed, it was found that a few chromatograms amongst each set of runs were not perfectly aligned with the rest. Alignment is an essential pre-requisite for PLS to work well, therefore, the ones that did not align precisely were corrected by appropriate shifting of the interpolated scans.

3.5.3. Unfolding and centring

As mentioned in Section 2, the raw data was stored as a three-way tensor $\underline{\boldsymbol{Z}}$ of dimensions $M \times I \times J$. To be able to extract any relevant information, $\underline{\boldsymbol{Z}}$ was unfolded to a two-way matrix $\boldsymbol{X}$ of $M \times I.J$ dimensions. For the three-way applications, $\boldsymbol{X}$ was mean centred along $\boldsymbol{M}$, and the one row-vector was subtracted from it, according to the equation

$$\overset{centred}{\underset{M,I.J}{\boldsymbol{X}}} = \underset{M,I.J}{\boldsymbol{X}} - \underset{M,I.J}{\bar{\boldsymbol{x}}}.$$

For three-mode PLS and three-way unfolded PLS, this matrix was used as the $\boldsymbol{X}$ data block. Note that for the three-way method, it is quite common to store the data in Matlab as the corresponding unfolded matrix, for reasons of simplicity. However, data-handling takes place assuming all three dimensions as described in Section 3.3.

For the two-way methods, $\underline{\boldsymbol{Z}}$ is converted to two 2D matrices with dimensions $M \times I$ and $M \times J$, respectively, which were mean centred along M, before used in PLS.

3.6. Software

The chromatographic data acquired from the Waters HPLC chromatograph were converted into the appropriate matrices using a combination of two relevant macros. The first one, the 2010 DDE Assistant for Raw Data macro (Version 2.10) was written by Waters and outputs the data as two-column vectors, the first one being the various wavelengths at successive points in time and the second being their corresponding absorbance values. The two-column matrix is then converted to an $I \times J$ matrix using a second VBA macro that was written in-house.

The unfolding and centring of tensor $\underline{\boldsymbol{Z}}$ was performed in VBA using an in-house macro, and all autoprediction and cross-validation calculations were performed in Matlab using a software provided by Dr. R. Bro. Representative results were also validated against a C++ program written by Dr. R.L. Erskine.

4. Results and discussion

4.1. Predictions at pH 5.0

4.1.1. Predictions for 1–5% concentrations of the minor impurity

The predictions for the 1–5% concentration range of the minor impurity, 3-hydroxypyridine, can be seen in Table 3, as % errors with respect to the average concentration of 3-hydroxypyridine (at 3%). In total, % errors for four PLS components are reported, although not so many components are always needed for all four PLS methods. The trend is the same for both auto-predicted and cross-validated errors, with the latter always being higher. The focus of the discussion is going to be on the cross-validated predictions, as they are known to be more realistic.

For example, for the two-way spectral profile PLS method, the cv% error is more or less constant after two PLS components are extracted, and has the low value of 3.2%. Note that the true % errors are obtained by multiplying by the average concentration (3%), so that a 3.2% error represents an overall true 0.096% error within the 1–5% range. The third PLS component shows a very slight improvement, which can also be regarded as noise, whereas in the fourth PLS component, the error starts to increase back again and the model loses confidence. Thus, two PLS components are thought to model the data quite sufficiently. After plotting true versus cross-validated predicted concentrations of 3-hydroxypyridine (values reported in mM), a straight line is obtained and is depicted in Fig. 5(a).

The rest of the methods all show very high cv% errors after extracting two PLS components. This is not surprising as there is quite a significant residual left unexplained in the $\boldsymbol{X}$ matrices, and this can be attributed to the discrepancies observed in aligning the interpolated data. This problem does not occur for the spectral profiles and is only observed for the time-

Table 3
Relative % errors for the 1–5% concentration range at pH 5.0

	No. of PLS components	Autoprediction errors (ap%)	Cross-validation errors (cv%)
Spectral two-way PLS	1	34.7	40.3
	2	2.4	3.2
	3	1.3	2.0
	4	1.1	2.5
Elution two-way PLS	1	31.6	42.1
	2	30.8	36.9
	3	9.1	13.6
	4	4.3	8.2
Three-way unfolded	1	32.4	41.5
	2	29.9	37.5
	3	2.8	4.3
	4	1.5	3.3
Three-way	1	33.9	43.7
	2	31.8	40.1
	3	2.6	4.5
	4	1.6	3.5

Note that the true % errors are obtained by multiplying by the average concentration, so that a 10% ap% error represents an overall true 0.3% error.

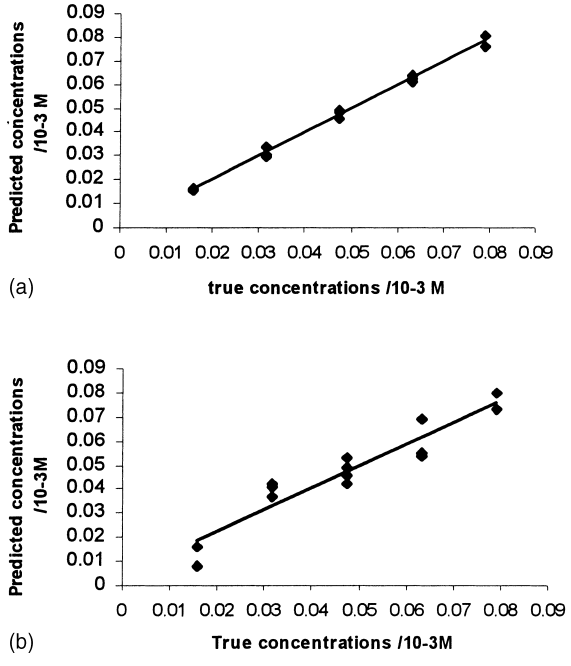

Fig. 5. (a) Predicted cross-validated versus true concentrations for the 1–5% range of 3-hydroxypyridine by spectral profile two-way PLS (two PLS components) at pH 5.0; (b) predicted cross-validated versus true concentrations for the 1–5% range of 3-hydroxypyridine by elution profile two-way PLS (three PLS components) at pH 5.0.

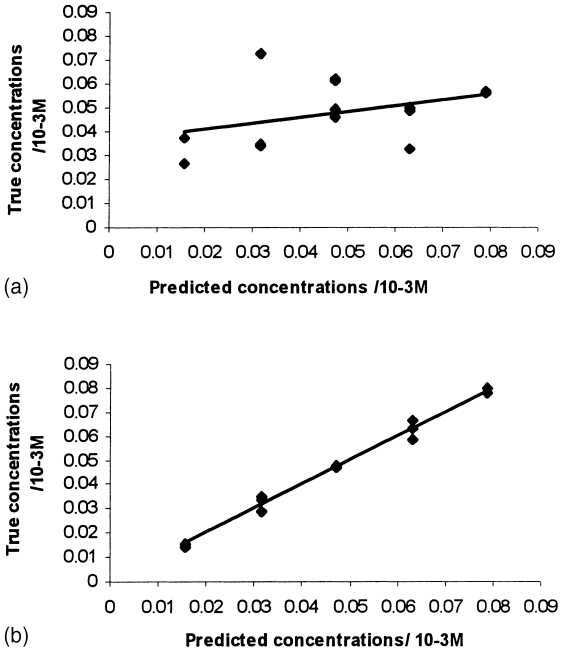

Fig. 6. (a) Predicted cross-validated versus true concentrations for the 1–5% range of 3-hydroxypyridine by three-way PLS (two PLS components) at pH 5.0; (b) predicted cross-validated versus true concentrations for the 1–5% range of 3-hydroxypyridine by three-way PLS (three PLS components) at pH 5.0.

dependent data, so a third PLS component – in two compound mixtures – is essential in getting reasonable predictions using any time-dependent method in HPLC (also observed using autopredictions). Fig. 6(a) and (b) show the difference in prediction calibration curves for the three-way PLS method after extracting 2 and 3 PLS components, respectively. The extra component problem represents the noise introduced into the system due to inexact time reproducibility. It is important to recognise that the chromatographic hardware used in this study is one of the most reproducible one in the market, so this problem will be more significant in most practical cases.

Among the three time-dependent methods examined, clearly, the three-way unfolded methods give similar errors after three PLS components (approximately 4.4%), whereas, the two-way elution profile method gives an error which is about three times as high (13.6%). This is easily understood as the three-way method incorporates some spectral information, too, whereas the two-way elution profile PLS method depends only on the time profile. As at pH 5.0 the two compounds are nearly co-eluting, the latter method gives considerably worse predictions. A fourth component can also be extracted for even better predictions. Plotting graphs of true versus predicted concentrations for the three-way methods (three PLS components) results in exactly the same graph to the one obtained with the spectral profile two-way PLS method. However, for the elution profile two-way PLS method, the graph is less linear (Fig. 5(b)).

4.1.2. Predictions for 0.1–0.5% of the minor impurity

The predictions for the 0.1–0.5% concentration range of 3-hydroxypyridine, can be seen in Table 4. From a first look at the errors reported, it is apparent that they are significantly higher, as the % range to be predicted is 10 times less than the one used above. Once again, the auto-predicted errors are less than the corresponding cross-validated. For the latter ones, all time-dependent methods appear to give errors around 34%, even after four PLS components have been

Table 4
Relative % errors for the 0.1–0.5% concentration range at pH 5.0

	No. of PLS components	Autoprediction errors (ap%)	Cross-validation errors (cv%)
Spectral two-way PLS	1	29.2	33.1
	2	14.3	20.9
	3	5.9	12.3
	4	4.3	8.1
Elution two-way PLS	1	29.7	35.4
	2	29.2	36.8
	3	23.0	41.6
	4	22.3	35.9
Three-way unfolded	1	29.7	35.4
	2	29.1	36.5
	3	22.2	41.9
	4	16.1	30.2
Three-way	1	29.7	35.4
	2	29.1	36.4
	3	23.9	42.9
	4	21.6	34.1

Note that the true % errors are obtained by multiplying by the average concentration, so that a 20% ap% error represents an overall true 0.06% error.

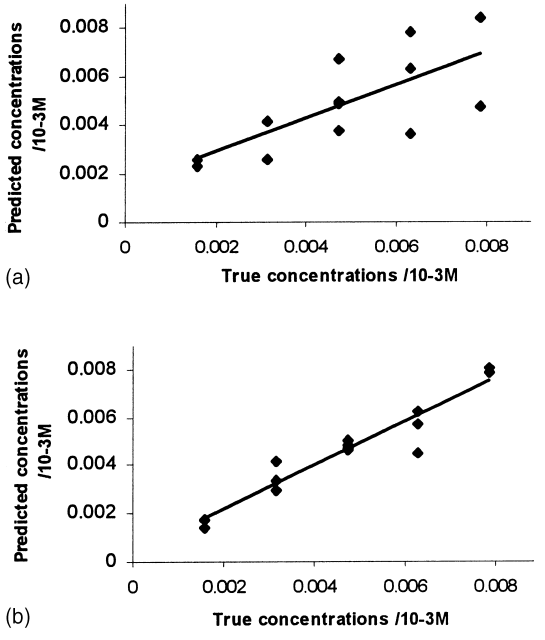

Fig. 7. (a) Predicted cross-validated versus true concentrations for the 0.1–0.5% range of 3-hydroxypyridine by three-way unfolded PLS (four PLS components) at pH 5.0; (b) predicted cross-validated versus true concentrations for the 0.1–0.5% range of 3-hydroxypyridine by spectral profile two-way PLS (three PLS components) at pH 5.0.

extracted. A typical graph of predicted cross-validated versus true concentrations for the three-way unfolded method, after four components, is shown in Fig. 7(a). What is also quite remarkable is that the autoprediction results do not seem to improve dramatically after the same number of components have been extracted and give quite high errors, too.

This is not the case, though, for the spectral profiles two-way PLS method, where both autoprediction and cross-validated errors decrease considerably with increasing number of PLS components. Thus, after two PLS components have been extracted, the average cv% prediction error is 20.9%, whereas after the third PLS is extracted this error drops to 12.2% (Fig. 7(b)). These observations are not surprising, since – as stated above – the spectral method does not depend at all on the resolution of compounds. Although the error is higher than before, it is not excessive, considering the fact that we are dealing with very low-level impurities, and more important, with co-eluting ones.

Table 5
Relative % errors for the 1–5% concentration range at pH 4.9

	No. of PLS components	Autoprediction errors (ap%)	Cross-validation errors (cv%)
Spectral two-way PLS	1	22.2	25.0
	2	4.9	6.5
	3	3.9	6.9
	4	3.3	6.7
Elution two-way PLS	1	28.9	33.9
	2	18.8	23.2
	3	5.4	8.0
	4	4.0	6.4
Three-way unfolded	1	30.3	35.5
	2	21.3	26.1
	3	5.9	8.8
	4	4.0	6.5
Three-way	1	30.7	36.0
	2	22.1	27.0
	3	6.4	9.8
	4	4.0	6.5

4.2. Predictions at pH 4.9

4.2.1. Predictions for 1–5% concentrations of the minor impurity

The predictions for the 1–5% concentration range of 3-hydroxypyridine, are given in Table 5. The basic trend is similar to the one observed for the same concentration range, described in Section 4.1.1. For the two-way spectral profiles PLS method, the cv% error is approximately constant after two PLS components are extracted, and has a value of 6.5%. For the two-way elution profile PLS method, the cv% error after three PLS components have been extracted is actually lower than that when the compounds were co-eluting, indicating that this method works better when the resolution between the peaks increases, as it is direct and only dependent on time.

However, the cv% predictions for the remaining three-way methods are slightly higher than those ones when the compounds were co-eluting. This could be attributed to the fact that there might have been a higher dilution error in any of the fourteen samples present in the dilution scheme. This would in fact

cause an increase in the reported errors, and the errors reported for the fourth PLS components could be taken as possible dilution errors. However, all four methods are seen to give approximately similar results, between 8–10% (three PLS components) for the time-dependent methods and 6.5% for the spectral method, which is still the better of the four.

To account for a possible dilution error in the sample preparation, the 1–5% set of experiments were repeated at pH 4.9. This time, the errors after three PLS components were extracted were less than those in both 1–5% set of experiments, as expected. The cv% errors for all three time-dependent methods averaged 3.3% after extraction of the third PLS component, whereas for the spectral two-way method, the cv% error was 4.7% after two PLS components and 4.1% at three PLS components. This could imply that at a higher peak resolution, time-dependent methods may work better than spectral-dependent ones. Clearly, as peaks get closer, the reverse is also true, as shown in Table 3.

4.2.2. *Predictions for 0.1–0.5% concentrations of the minor impurity*

The predictions for the 0.1–0.5% concentration range of 3-hydroxypyridine are given in Table 6. Compared with the corresponding ones at pH 5.0, they are improved. The spectral two-way PLS method gives a cv% error 7.3% less than the one at pH 5.0, after extraction of two PLS components. Further extraction of a third PLS component improves the error by approximately 4.6% (linear calibration curve seen in Fig. 8(a)), especially as we are dealing with very little impurities. The remaining time-dependent methods show a dramatic improvement in prediction at the fourth PLS component, with the elution profile two-way PLS method (Fig. 8(b)) working slightly better than the three-way methods. Once again, the spectral method appears to be working better than the rest of methods, a result which has been shown with extreme confidence for this sort of HPLC data of partially resolved impurities.

Table 6
Relative % errors for the 0.1–0.5% concentration range at pH 4.9

	No. of PLS components	Autoprediction errors (ap%)	Cross-validation errors (cv%)
Spectral two-way PLS	1	30.9	34.9
	2	10.5	13.6
	3	6.3	8.9
	4	2.6	7.5
Elution two-way PLS	1	37.9	42.2
	2	31.1	38.0
	3	14.4	36.0
	4	2.2	9.8
Three-way unfolded	1	38.0	42.3
	2	31.4	38.7
	3	14.8	39.5
	4	3.0	11.5
Three-way	1	38.0	42.3
	2	31.4	38.7
	3	14.5	39.6
	4	2.3	11.5

5. Conclusion

In this paper a comparison on the use of different approaches to PLS calibration for the quantification of a closely eluting impurity in HPLC has been made. Many conclusions can be reached.

First of all the methods are remarkably effective in quantification of low impurity levels. For example, in Table 4, a relative cross-validated error of 8.1% for spectral two-way PLS represents an average absolute error of 0.024% since the concentration range is centred on a level of 0.3% impurity, in a case where the peaks are almost completely overlapping.

The quality of predictions depends, in part, on the percentage impurity, as expected. The relationship to the peak separation is clear at the lower impurity range, especially for the time-dependent methods (see Tables 4 and 6). At higher impurity levels, this factor is not so significant because the apparent lowest cross-validated error is probably almost totally dependent on unknown sample preparation errors, and so represents an excellent prediction.

It is slightly surprising that the three-way methods perform less well than the two-way spectral methods. This is almost inevitably due to the problem of inconsistencies in elution time. Even aligning peaks

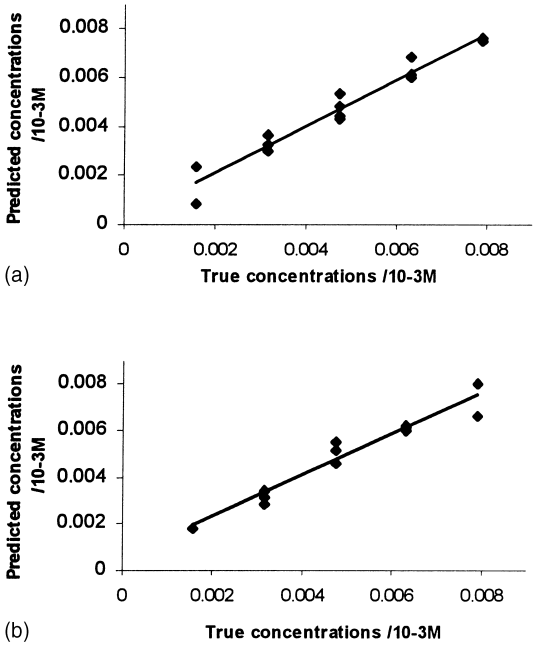

Fig. 8. (a) Predicted cross-validated versus true concentrations for the 0.1–0.5% range of 3-hydroxypyridine by spectral profile two-way PLS (three PLS components) at pH 4.9; (b) predicted cross-validated versus true concentrations for the 0.1–0.5% range of 3-hydroxypyridine by elution profile two-way PLS (four PLS components) at pH 4.9.

to their maxima does not entirely solve this problem, because the exact offsets may differ by less than one sampling time. Spectral methods are independent of this difficulty and so are not influenced by this problem. They are also largely independent of resolution. This surprising conclusion suggests that the chromatographic dimension degrades information in the case described in this paper. The results contradict simulation studies which do not take into account variability in elution time.

Despite these limitations, PLS is a very powerful approach for the quantification of small levels of impurities in partially overlapping HPLC peaks.

Acknowledgements

We thank EPSRC for financial support. Rasmus Bro is thanked for enabling access to Matlab software for PLS and Les Erskine for supplying validation software in Excel.

References

[1] T.W. Ryan, Anal. Lett. 4 (1998) 651.
[2] S. Gorog, B. Herenyi, M. Renyei, J. Pharm. Biomed. Anal. 10 (1992) 831.
[3] C. Demir, R.G. Brereton, Analyst 122 (1997) 631.
[4] C. Demir, R.G. Brereton, Analyst 123 (1998) 181.
[5] H. Martens, T. Næs, Multivariate Calibration, Wiley, New York, 1989.
[6] L. Ståhle, Chemom. Intell. Lab. Syst. 7 (1989) 95.
[7] A.K. Smilde, Chemom. Intell. Lab. Syst. 15 (1992) 143.
[8] P. Geladi, Chemom. Intell. Lab. Syst. 7 (1989) 11.
[9] K. Esbensen, S. Wold, Proceedings of the Symposium on Applied Statistics, Stavanger, 1983, p. 11.
[10] S. Wold, P. Geladi, K. Esbensen, J. Öhman, J. Chem. 1 (1987) 41.
[11] S. Dunkerley, J. Crosby, R.G. Brereton, K.D. Zissis, R.E.A. Escott, Analyst 123 (1998) 2021.
[12] B.R. Kowalski, R. Gerlachand, H. Wold, in: K. Joreskog, S. Wold (Eds.), Chemical Systems Under Indirect Observation, North-Holland, Amsterdam, 1982.
[13] S. Wold, A. Ruhe, H. Wold, W.J. Dunn III, SIAM J. Sci. Statist. Comput. 5 (1984) 735.
[14] M.F. Delaney, Chemom. Intell. Lab. Syst. 3 (1988) 45.
[15] B.K. Alsberg, M.K. Winson, D.B. Kell, Chemom. Intell. Lab. Syst. 36 (1997) 95.
[16] P.A. Araujo, D.A. Cirovic, R.G. Brereton, Analyst 121 (1996) 581.
[17] D.A. Cirovic, R.G. Brereton, P.T. Walsh, Analyst 121 (1996) 575.
[18] W. Lindberg, J. Persson, S. Wold, Anal. Chem. 55 (1983) 643.
[19] H. Martens, T. Næs, Trends in Analytical Chemistry, vol. 3, 1984, pp. 204 and 266.
[20] P. Geladi, B.R. Kowalski, Anal. Chim. Acta 185 (1986) 1.
[21] P. Geladi, B.R. Kowalski, Anal. Chim. Acta 185 (1986) 19–32.
[22] J. Sun, J. Chemom. 10 (1996) 1.
[23] S.Wold, C.Albano, W.J. Dunn III, K. Esbensen, S. Hellberg, E. Johansson, M. Sjøstrom, Pattern recognition: finding and using regularities in multivariate data, in: H. Martens, H. Russworm (Eds.), Food Research and Data Analysis, Applied Science Publishers, London, 1983.
[24] K.D. Zissis, R.G. Brereton, R. Escott, Analyst 123 (1998) 1165.
[25] R. Bro, J. Chemom. 10 (1996) 47.