Deconvolution of partially overlapping tailing peaks in diode array high performance liquid chromatography using purity ratios

Richard G. Brereton* and Samantha Dunkerley

School of Chemistry, University of Bristol, Cantock's Close, Bristol, UK BS8 1TS

Received 23rd September 1998, Accepted 17th February 1999

A new approach to deconvolution of diode array HPLC data is described. It is illustrated by the chromatograms of 2- and 3-hydroxypyridine undertaken at pH 4.8, 5.1 and 5.2 representing three different qualities of resolution from almost totally overlapping to well resolved. In all cases peaks are tailing, typical of many real situations. The first step is to select wavelengths using the normalised spectra at each point in time. Next a purity curve is obtained from these wavelengths, which is related to a real purity curve used for deconvolution. It is shown that the method is robust over the full resolvability range.

1 Introduction

There has been a huge literature on deconvolution in two way HPLC,[1–3] primarily diode array HPLC. The majority of these methods called self modelling curve resolution,[4,5] evolving factor analysis,[6,7] window factor analysis,[8,9] heuristic evolving projections,[10–12] target factor analysis,[13] and so on, have depended on determining composition one (or rank one or selective) regions in the chromatogram. These are then employed as partial estimates of spectra or elution profiles of components in the mixture, and chemometrics techniques are used to fill in the jigsaw. Most such methods make implicit assumptions about the data, for example, that the major components do exhibit regions of purity.

The majority of published chemometrics methods do not work well when peaks are totally overlapping. Most methods have been applied to cases where peakshapes are symmetrical, which often is not the case in many real-world pharmaceutical analyses.[14] In some systems it is not easy to separate out compounds completely, and often individual compounds do not have clear composition one regions. This is particularly so when peakshapes are tailing. Although the fastest eluting peak will normally have a clear composition one region, this will not necessarily be true for the slower eluting peak, if the tails are pronounced. Hence it is necessary to resolve out peaks without the assumption that each compound has a composition one region. Some approaches do, indeed, work correctly when peaks in the centre of the cluster do not have composition one regions (*e.g.*, by component stripping) but assume, nevertheless, that the components at the extreme end of the cluster have selective regions, which is not necessarily the case when peaks are severely tailing.

The theory of resolution of completely embedded peaks has been treated by Manne[15] who states that in the case where only one compound in a two component cluster has a composition one region, there is available knowledge of the spectrum of this compound, and the shape (but not actual size) of the elution profile of the other (embedded) analyte. It is pointed out that there is no unique solution to this problem, and further knowledge or assumptions about the system are required to obtain a meaningful solution. In this paper, we provide a computational approach that is tested on three independent case studies, of varying difficulties, in one of which one analyte is completely embedded in the other.

Providing the signal to noise ratio is adequate, the change in relative concentrations of compounds can be assessed by examining the change in spectral characteristics across a peak. This is well known, but has mainly been exploited for the purpose of peak purity profiling rather than deconvolution. However, the relative absorbance at any given wavelength provides a purity curve which can be used for further deconvolution. In this paper a new approach is proposed taking this information into account. The method works for a variety of situations from almost complete resolution to nearly total overlap.

Two pieces of information are assumed to be known as illustrated in Fig. 1. The first is the spectrum of the fastest eluting compound, which can be obtained from the beginning of the peak cluster, before the second component elutes. The second is the elution profile of the slower eluting peak. This is given by the peak purity curve. The spectrum of this latter peak is not known if there is no composition one region, and the elution profile of the fastest eluting peak depends on the magnitude of the profile for the second peak. Chemometrics attempts to determine two unknowns. Most chemometric methods proposed in the literature assume either that both spectra or both elution profiles are known or that all information is known about one constituent of the mixture. The case described in this paper requires assumed knowledge of the system.

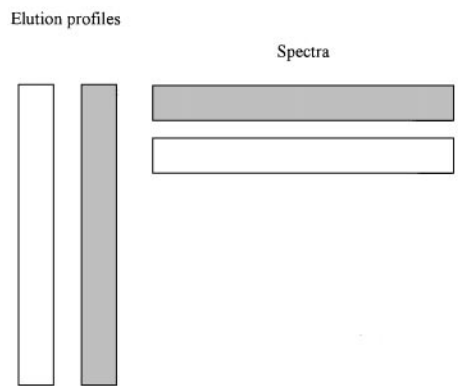

Fig. 1 Known information about compounds A and B indicated in grey.

A series of chromatograms of a mixture of 2- and 3-hydroxypyridine at three pHs is described in this paper, each representative of a different amount of resolution and overlap. Further details of the data sets are described elsewhere.[16,17]

2 Methods

2.1 Experimental

DAD-HPLC (diode array detection high-performance liquid chromatography) analysis of equimolar mixtures of 2- and 3-hydroxypyridine was performed using the conditions described previously.[16] EAS (electronic absorption spectra) were recorded every 1 s, between 200 and 400 nm at approximately 1.2 nm intervals, for a run time of 10 min. The DAD-HPLC data sets recorded at pH 4.8 (almost entirely resolved), pH 5.1 (almost completely coeluting) and pH 5.2 (partially overlapping) were selected for further study. These data sets were reduced to a wavelength region of 220–350 nm (110 wavelengths), with lower wavelengths discarded due to solvent absorption, and later ones due to the compounds not absorbing. In addition, the data were reduced to a scan range of 149 s (150 datapoints), which was selected in order to encompass the main elution region (this range differs for each data set due to the compounds' elution time changing with pH).

2.2 Choosing the region of the chromatogram

The principle of the method described in this paper is that for a two component mixture, the ratio of absorbances at any two wavelengths relates to the proportion of the two components. It is therefore essential that a region is chosen that covers primarily the eluting compounds and does not contain noise or is dominated by impurities (small percentage impurities are acceptable providing they do not dominate the chromatogram). An essential first step is to identify this region. An incorrect choice of region will result in misleading predictions. However, this can be corrected using the purity curve as described below, and new predictions obtained with a changed region.

2.3 Preprocessing and wavelength selection

The raw chromatographic data matrix $\boldsymbol{X}$ consisting of 110 columns (wavelengths) and 150 rows (points in time) is normalised as follows at each point in time

$$ {}^{n}x_{ij} = \frac{x_{ij}}{\sum_{j=1}^{J} x_{ij}} $$

where there are I points in time and J wavelengths. The reason for this is to compare the relative spectral intensities at each point in time, and is the key to the success of the method described in this paper.

Consider one compound (A) whose normalised spectrum has an intensity of 0.10 at 300 nm, and another compound (B) whose normalised spectrum has an intensity of 0.06 at the same wavelength. Then a normalised intensity of 0.09 corresponds to a 3 + 1 mixture of compounds A to B. In the case described in this paper, it is assumed that a composition one region can be found for the fastest eluting compound, even if it is a very small region, but not necessarily for the slowest eluting compound, because of peak tailing. The normalised intensity at each point in time has an upper and lower limit for each wavelength, $\max({}^{n}x_{ij}, j)$ and $\min({}^{n}x_{ij}, j)$, corresponding to the most pure regions of the chromatogram for compounds A and B.

The next step is to eliminate those wavelengths that are not useful. The reason for this is that some wavelengths may be dominated by noise, or, alternatively, the normalised spectral intensities at each wavelength may be quite similar implying that the wavelengths have very little discriminatory power. There are two criteria:

1. The first is to remove wavelengths for which $\max({}^{n}x_{ij}, j)$ and $\min({}^{n}x_{ij}, j)$ are very close in size, using a threshold accepting only those wavelengths for which $\rho = \dfrac{\max({}^{n}x_{ij}, j)}{\min({}^{n}x_{ij}, j)} > t.$

If thevalue of ρ is close to 1 this implies that the spectra of the two components in the mixture have absorbances that are very similar in magnitude, so the variation will be small over the chromatogram.

2. The second is to remove low intensity wavelengths, for which $\max({}^{n}x_{ij}, j) < s$ where s is a small number less than $1/J$ (the average intensity if all wavelengths had identical absorbances) and greater than 0. The reason for this is that if wavelengths are of low intensity, the variation will be influenced primarily by noise.

These thresholds then leave a number of suitable wavelengths. A suitable selection criterion may typically retain two thirds to three quarters of the wavelengths.

2.4 Purity curves

The next step is to calculate purity curves. These are the key to the deconvolution algorithm and show how the relative proportions of the components in the mixture vary.

These are first calculated over the wavelengths selected as described above. The first step is simply to examine the normalised chromatogram ${}^{n}x_{ij}$ at each point in time for each selected wavelength. This will take values between the maximum and minimum, and can be rescaled between 0 and 1 as follows

$$ {}^{s}r_{ij} = \frac{[{}^{n}x_{ij} - \min({}^{n}x_{ij}, j)]}{[\max({}^{n}x_{ij}, j) - \min({}^{n}x_{ij}, j)]} $$

which has a value of 1 at the maximum and 0 at the minimum.

The next phase is to determine whether a maximum or minimum corresponds to the purest point of the fastest eluting component. If $i_{j(\max)}$ corresponds to the value of i (point in time) at wavelength j for which ${}^{n}x_{ij}$ is a maximum and $i_{j(\min)}$ the corresponding minimum, then if

$$ i_{j(\max)} < i_{j(\min)} $$

$$ R_{ij} = {}^{s}r_{ij} $$

otherwise

$$ R_{ij} = 1 - {}^{s}r_{ij} $$

This ensures that the overall purity curve at each wavelength has a value of 1 at low elution times (corresponding to maximum purity of the fastest eluting compound A) and 0 at high elution times (corresponding to maximum purity of the slowest eluting compound B). This simple transformation takes account of the wavelengths at which compound A absorbs both more and less intensely than compound B.

Finally an overall purity curve is obtained

$$R_i = \frac{\sum_{i \in g} R_{ij}}{G}$$

where G acceptable wavelengths have been found (see above) to give a group g of acceptable wavelengths. This purity curve is a consensus over the relevant wavelengths. Ideally, it will vary between 1 (most pure for the fastest eluting component) to 0 (least pure for fastest eluting component). Because the observed positions of the purity maxima typically vary by a datapoint or more according to wavelength (due to noise), in practice, the upper limit is slightly less than 1 and the lower limit slightly greater than 0.

There are, of course, other possible approaches to determining these purity curves, for example, by simply viewing each spectrum in time as a vector and then determining the normalised distance between a composition one spectrum (for example one at the beginning of the elution profile) and spectra across the peak cluster. The greater the distance, the lower the value of R_i. It is not the prime purpose of this paper to compare different approaches to obtaining purity curves. However, the method in this paper works well when spectra are fairly similar because only those wavelengths that are most useful for discrimination are retained. Using conventional vector distances which use all wavelengths can lead to less informative purity curves than the method proposed in this paper.

2.5 Converting purity curves to deconvoluted profiles

The next step is to convert the purity curves to deconvoluted profiles of the two components in the mixture. A first step is to determine an overall chromatographic profile

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

which is the sum of intensity over all wavelengths (note that these now include the wavelengths that have been rejected above).

The aim is to decompose this profile into sub-profiles for each component so that

$$X_i = [p_i + (1 - p_i)]X_i$$

where p_i is the proportional contribution of component 1 to the overall profile at time i. Hence, if a given elution time corresponds entirely to pure component 1, then p_i equals 1, whereas for pure component 2, it equals 0.

The key to the algorithm is to relate values of R_i (the purity curve) to p_i—the true purity. The assumption is that for the fastest eluting peak there is a composition one region, but for the slower eluting compound this is not necessarily so. A minimum value of p_i equal to $p_{\min}$ is determined using the deconvolution method described below, and the relationship between R_i and p_i is given by

$$p_i = R_i(1 - p_{\min}) + p_{\min}$$

so that a value of R_i of 0 corresponds to a value of p_i equal to $p_{\min}$. If this latter value is also 0, the slower eluting peak also has a composition one region. This latter assumption is frequently employed in many forms of SMCR (self modelling curve resolution), where each significant component in a chromatographic mixture exhibits chromatographic selectivity. Where one peak is buried in another or the peaks are severely tailing and closely eluting this assumption is not true, and many approaches for deconvolution fall down.

It is essential to recognise that the mapping of the purity curves to the true purity depends on a number of assumptions about the data. This point is often not made explicitly clear in the literature on factor analysis in chromatography.

The profiles for each component are given by

$$C_{i1} = X_i p_i$$
$$C_{i2} = X_i(1 - p_i)$$

In this study it is assumed that the true profiles have two properties:

1. All chromatographic profiles are positive. This is a sensible assumption for chromatographic peakshapes.

2. Preferably the ratio of the maxima of each pure profile is as close as possible to 1. Note that this assumption is not valid under all circumstances.

The second assumption is an approximation, but in the absence of further information is a consequence of using an equal mixture of the two components. It is essential to recognise that in this study there is no further information that can be sensibly used. However, as shown below, this assumption provides excellent results. Other *a priori* assumptions could be made of chromatographic data and often are. Many published approaches for SMCR in chromatography will fall over if significant peaks lack either a pure variable (*e.g.*, wavelength or mass spectral ion) or chromatographic composition one point.

It is important to recognise that the second assumption is valid only for this particular data set. Even in the chromatograms described below, it is possible to include extra assumptions, such as unimodality, and a good approach might be to include as much or as little knowledge of the system as is available subject to a mathematically correct solution.

2.6 Determining true purity curves

The determination of $p_{\min}$ can be performed using a simplex algorithm.

The first step is to determine a quantitative criterion for success, denoted by q. The closer this value is to 1, the better the model. If either of the deconvoluted peaks shows any negative intensity, then q is set to 1000. This gives a heavy penalty for a negative solution which is physically meaningless. If both resolved peaks are positive q equals the ratio of the maximum height of the two deconvoluted chromatographic peaks, as determined above, calculated so that the ratio is always greater than one, dependent on which of the peaks is most intense. The closer this is to 1, the better.

The second step is to determine two starting values for $p_{\min}$ of 0 (composition one regions for both peaks) and 0.02 (which equals the step-size). Note that a fixed sized simplex algorithm is employed in this paper; the method is very fast using a Pentium II 350 MHz computer implemented using VBA in Excel. More rapid simplex or other optimisation methods could be employed but will give similar results. The method does not depend on the optimisation algorithm providing the step-size is sufficiently small.

The next step is to calculate q for the two starting values of $p_{\min}$. The closer to 1, the better. If the value corresponding to $p_{\min}$ that gives the best answer is 0.02, then test a value of 0.04. When the values of $p_{\min}$ are 0.02 and 0.04, if the value is closer to when $p_{\min} = 0.04$, test new values of 0.04 and 0.06. If not, reduce the step-size to 0.01, to test 0.04 and 0.03.

Standard simplex optimisation is performed by comparing the value of q at two values of $p_{\min}$ decreasing the step-size as relevant, until a small step-size (0.0001) is achieved, for convergence. If the value of $p_{\min}$ is less than 0, change to a value of 0 which implies composition one regions for both components (sometimes very small negative values are found at the optimum).

This step results in two positive profiles which can be expressed by the matrix $\boldsymbol{C}$ whose columns are the predicted profiles of the chromatograms.

2.7 Determining the spectra

The predicted spectra, $\boldsymbol{S}$, can be obtained using the pseudo-inverse and standard multiple linear regression as follows

$$\boldsymbol{S} = (\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{X}$$

to obtain fully deconvoluted spectra.

3 Results

3.1 Analysis at pH 5.1

The profile chromatogram at pH 5.1 is given in Fig. 2(a). Note that it is important to select an appropriate region of the overall chromatogram where at least one component elutes. The key to the method proposed in this paper is that all points in time consist of either one or the other coeluent.

The chromatogram consists of two closely overlapping peaks. Table 1 illustrates the method for wavelength selection. Two criteria are employed. The value of t was set at 1.2. This results in wavelengths for which the change in relative absorbance is small throughout the peak. Hence the wavelengths 230.84 and 232.01 nm are rejected for this reason, since the values of $\min({}^{n}x_{ij}, j)$ and $\max({}^{n}x_{ij}, j)$ are close in value. Note that had 232.01 nm been chosen erratic results would have been obtained. This is because the relative absorbance for each compound is so similar that the purity curve is influenced by noise. The minimum threshold s is chosen to be 0.1, meaning that low intensity wavelengths are likewise rejected. Hence all wavelengths higher than 327 nm are rejected.

The remaining wavelengths exhibit values of $i_{j(\max)}$ and $i_{j(\min)}$ of between 1 and 2 or between 15 and 16. This suggests that the slower eluting peak is at its purest at around datapoint 16, which is clearly reflected in the purity curve [Fig. 2(b)], whereas the first peak is at its purest at around datapoint 2. Note that the maximum value of the purity curve is not quite 1, because of the slight shift in the purest datapoint at the beginning of the elution according to wavelength. The purity curve shows a very clear and unambiguous trend, not obvious from the raw chromatographic profile, which can be used to advantage in deconvolution.

Using the simplex algorithm described above, a value of $p_{\min}$ of 0.50 is obtained, to give the profiles and spectra of Fig. 2(c) and 2(d). It is constructive to see the influence of a change in $p_{\min}$ on the quality of the deconvolution. Fig. 3 illustrates the effect of setting this parameter at 0, 0.25 and 0.75, and illustrates the effectiveness of the approach proposed in this paper. Below the optimum, the deconvoluted chromatographic peakshape for the fastest eluting compound is no longer unimodal. Above the optimum, although both chromatographic peakshapes are satisfactory, the spectrum of the slowest eluting compound is no longer satisfactory. In fact the spectrum at the optimum is close to the true spectrum at this pH.

It is very important to recognise that all the solutions of Fig. 2 and 3 provide an equally good mathematical fit to the data. Using pure regression methods it is impossible to distinguish between these solutions. The importance of refining standard statistical algorithms to obtain chemically meaningful information is insufficiently emphasised in the chemometrics literature. In fact, without extra assumptions about the data, the results of deconvolution, by any pure multivariate method, are meaningless. There is no *a priori* information as to the value of $p_{\min}$, and purely mathematical solutions will not lead to an unambiguous answer. It is necessary to put assumptions into the simplex optimisation algorithm in order to obtain a meaningful answer. The assumptions in this paper can, of course, be varied or changed, and other constraints or penalties for poor performance could be added. Again this extra *a priori* knowledge is essential to the performance of the algorithm.

Many so-called automated deconvolution algorithms do make assumptions. For example, many methods for mixture deconvolution assume that the value of $p_{\min}$ is 0, *i.e.*, that there are composition one regions for all resolved components in the mixture, and only successfully resolve out such components. Other pure variable methods common in mass spectrometry such as SIMPLISMA[18] assume that there are pure variables diagnostic of individual components in a mixture, which is sensible in mass spectrometry but less so in DAD-HPLC where it is not always possible to identify intense diagnostic wavelengths.

Providing the simplex algorithm is modified, or clear assumptions are made about the value of $p_{\min}$, the method in this paper can easily be adapted for other situations.

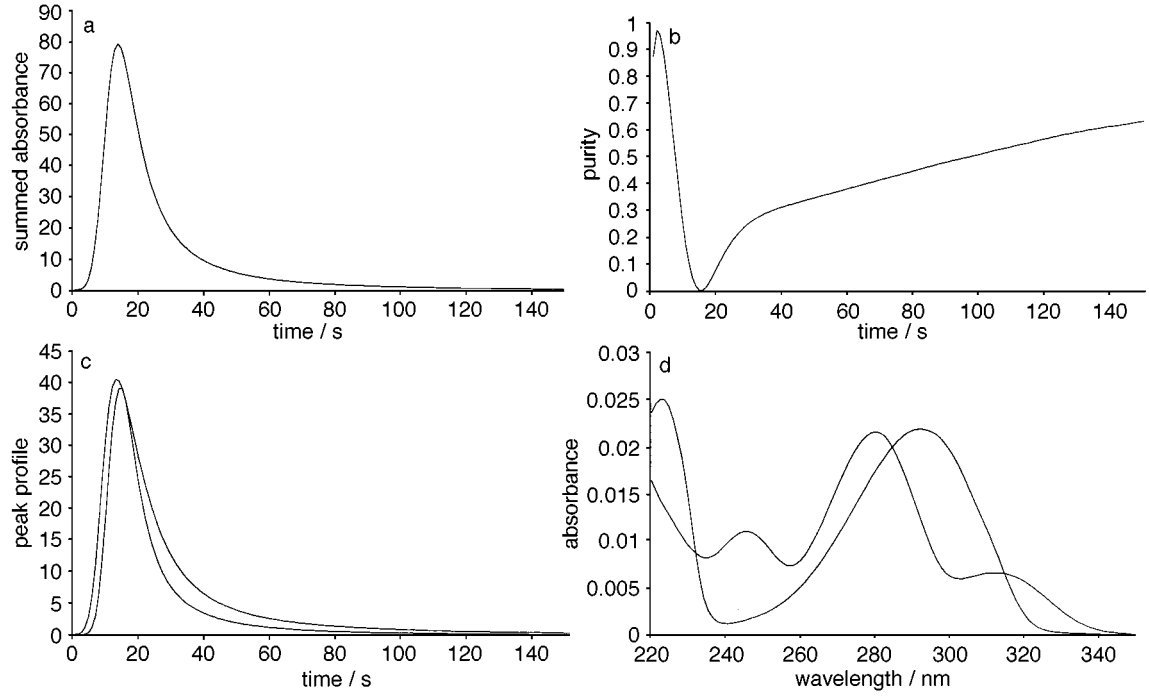

Fig. 2 (a) Summed absorbance at pH 5.1; (b) purity curve for pH 5.1; (c) optimal deconvoluted profiles at pH 5.1; (d) deconvoluted spectra at pH 5.1.

3.2 Analysis at pH 4.8 and 5.2

These two pHs have been chosen as examples where there is better resolution. The profiles, purity curves and results of deconvolution are shown in Fig. 4 and 5.

At pH 5.2, the value of p_{min} is found to be 0.302 whereas at pH 4.8 it is 0. The latter case would be easy to resolve using most conventional methods for factor analysis, but is not particularly challenging. It is nevertheless interesting to see that the purity curve increases slightly at slow elution times, suggesting that there are still very slight amounts of tailing, and that a value of 0 is an approximation.

The former case poses a greater problem. There is no significant composition one point for the slow eluting component due to the tailing peakshape, as clearly shown by the purity curve, which increases to approximately 0.5 at long elution times.

4 Conclusions

The method proposed in this paper works well. The deconvoluted spectra at the optimum values of p_{min} closely resemble the

Table 1 Wavelength selection at pH 5.1

Wavelength/nm	min($^n x_{ij}$, j)	max($^n x_{ij}$, j)	$i_{j(min)}$	$i_{j(max)}$	Accept or reject?[a]	Is $i_{j(max)} < i_{j(min)}$[a]	Wavelength/nm	min($^n x_{ij}$, j)	max($^n x_{ij}$, j)	$i_{j(min)}$	$i_{j(max)}$	Accept or reject?[a]	Is $i_{j(max)} < i_{j(min)}$[a]
220.23	0.019484	0.027069	14	1	T	F	285.31	0.017754	0.019168	16	3	F	F
221.40	0.019297	0.026904	15	1	T	F	286.50	0.017622	0.01958	16	3	F	F
222.58	0.019008	0.026243	15	1	T	F	287.69	0.01742	0.019933	16	3	F	F
223.76	0.018568	0.025068	15	2	T	F	288.88	0.017171	0.02024	16	3	F	F
224.94	0.0178	0.024132	15	2	T	F	290.07	0.016894	0.020365	16	3	T	F
226.12	0.016737	0.02245	15	2	T	F	291.26	0.016572	0.020507	16	3	T	F
227.30	0.015499	0.020464	16	2	T	F	292.45	0.016217	0.020489	16	3	T	F
228.48	0.014157	0.018112	16	2	T	F	293.64	0.015824	0.020397	16	3	T	F
229.65	0.01259	0.015429	16	3	T	F	294.83	0.015423	0.020296	16	2	T	F
230.84	0.010793	0.012466	16	2	F	F	296.02	0.015004	0.019989	16	3	T	F
232.01	0.008906	0.009227	149	3	F	F	297.21	0.014601	0.019731	16	3	T	F
233.20	0.00636	0.007502	3	15	F	F	298.40	0.014204	0.019365	16	3	T	F
234.38	0.004254	0.0065	2	14	T	T	299.59	0.013828	0.01886	16	3	T	F
235.55	0.003037	0.005992	2	15	T	T	300.78	0.013448	0.018267	16	3	T	F
236.74	0.002283	0.005845	2	16	T	T	301.97	0.013064	0.017574	16	3	T	F
237.92	0.002002	0.005938	2	16	T	T	303.16	0.012666	0.016726	16	3	T	F
239.10	0.001849	0.006175	2	16	T	T	304.35	0.012251	0.01586	16	2	T	F
240.28	0.001826	0.006486	2	16	T	T	305.54	0.011812	0.014854	16	3	T	F
241.46	0.001986	0.006805	2	16	T	T	306.74	0.011326	0.013989	16	3	T	F
242.64	0.002047	0.007094	2	16	T	T	307.93	0.010903	0.013023	16	3	F	F
243.82	0.002207	0.007322	2	16	T	T	309.12	0.010444	0.012113	16	3	F	F
245.01	0.002237	0.007473	2	16	T	T	310.31	0.009967	0.011334	16	1	F	F
246.19	0.002374	0.007528	2	16	T	T	311.51	0.009463	0.01018	16	3	F	F
247.37	0.002504	0.007484	2	16	T	T	312.70	0.008908	0.009225	18	146	F	F
248.55	0.002587	0.007345	2	16	T	T	313.89	0.00762	0.008313	1	14	F	F
249.74	0.002747	0.007128	2	16	T	T	315.08	0.006547	0.007682	1	15	F	F
250.92	0.002801	0.006851	2	16	T	T	316.28	0.005982	0.007026	3	15	F	F
252.10	0.003067	0.006541	2	16	T	T	317.47	0.004954	0.00637	2	15	T	T
253.29	0.003158	0.00623	2	15	T	T	318.66	0.004054	0.005736	3	16	T	T
254.47	0.003371	0.005959	2	15	T	T	319.86	0.003232	0.005141	3	16	T	T
255.65	0.003356	0.005764	1	15	T	T	321.05	0.002397	0.004604	2	16	T	T
256.84	0.003686	0.005676	1	15	T	T	322.25	0.001992	0.004122	3	16	T	T
258.02	0.004254	0.005713	2	15	T	T	323.44	0.001568	0.003695	2	16	T	T
259.20	0.004566	0.005891	2	15	T	T	324.63	0.00121	0.00331	2	16	T	T
260.39	0.004594	0.006208	1	15	T	T	325.83	0.000997	0.002959	2	16	T	T
261.58	0.005419	0.00665	1	15	T	T	327.02	0.000824	0.002633	3	16	F	F
262.76	0.005667	0.007186	1	15	T	T	328.22	0.000688	0.002328	3	16	F	F
263.95	0.006657	0.007799	1	15	F	F	329.41	0.000584	0.002042	3	16	F	F
265.13	0.00707	0.008472	1	15	F	F	330.61	0.000584	0.001788	3	1	F	F
266.32	0.007976	0.009202	2	15	F	F	331.80	0.000342	0.001523	2	16	F	F
267.50	0.008528	0.009978	1	15	F	F	333.00	0.000381	0.001292	2	16	F	F
268.69	0.009404	0.010784	3	15	F	F	334.20	0.000317	0.001084	3	16	F	F
269.87	0.009986	0.011597	1	15	F	F	335.39	0	0.000898	1	16	F	F
271.06	0.010618	0.012398	1	15	F	F	336.59	0	0.000737	1	16	F	F
272.25	0.011526	0.013173	1	15	F	F	337.78	0.000199	0.000598	3	16	F	F
273.43	0.01242	0.013923	2	15	F	F	338.98	0.000152	0.000482	2	16	F	F
274.62	0.012874	0.014636	1	16	F	F	340.18	0.000149	0.00066	4	1	F	F
275.81	0.013843	0.015304	2	15	F	F	341.37	0.00014	0.001018	4	1	F	F
276.99	0.014552	0.015899	1	15	F	F	342.57	9.11E-05	0.001513	137	1	F	F
278.18	0.01546	0.016434	1	15	F	F	343.77	8.64E-05	0.00273	147	1	F	F
279.37	0.015928	0.016875	1	16	F	F	344.96	4.89E-05	0.001981	131	1	F	F
280.56	0.017008	0.017414	3	145	F	F	346.16	7.22E-05	0.00154	139	1	F	F
281.75	0.017028	0.017884	1	145	F	F	347.36	3.81E-05	0.001733	139	1	F	F
282.93	0.017358	0.018344	1	150	F	F	348.56	5.91E-05	0.001871	118	1	F	F
284.12	0.017784	0.018778	16	149	F	F	349.76	4.89E-05	0.000798	133	1	F	F

[a] T, true; F, false.

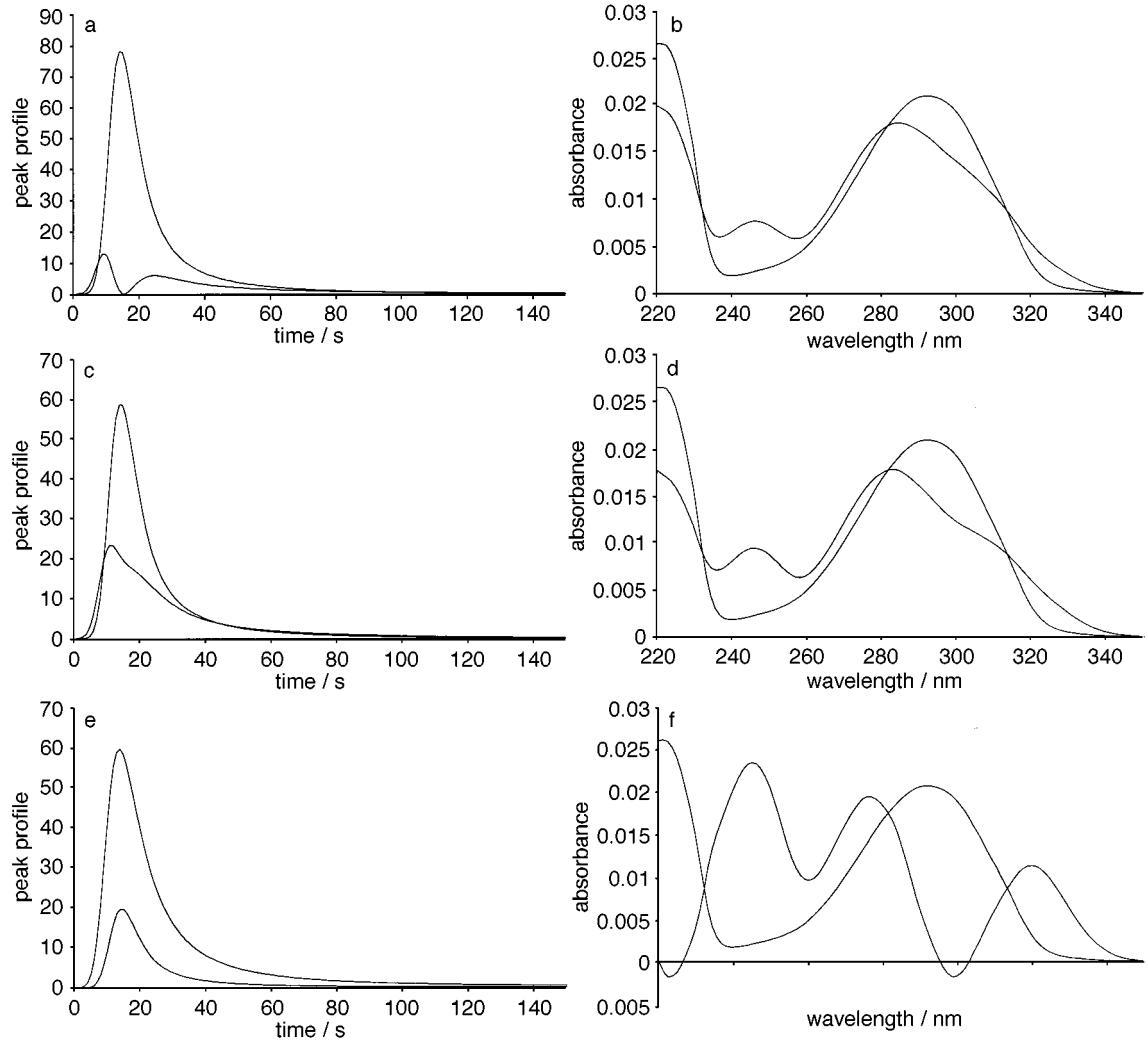

Fig. 3 (a) Deconvoluted elution profiles for p_{min} = 0 and pH 5.1; (b) deconvoluted spectra for p_{min} = 0 and pH 5.1; (c) deconvoluted elution profiles for p_{min} = 0.25 and pH 5.1; (d) deconvoluted spectra for p_{min} = 0.25 and pH 5.1; (e) deconvoluted elution profiles for p_{min} = 0.75 and pH 5.1; (f) deconvoluted spectra for p_{min} = 0.75 and pH 5.1.

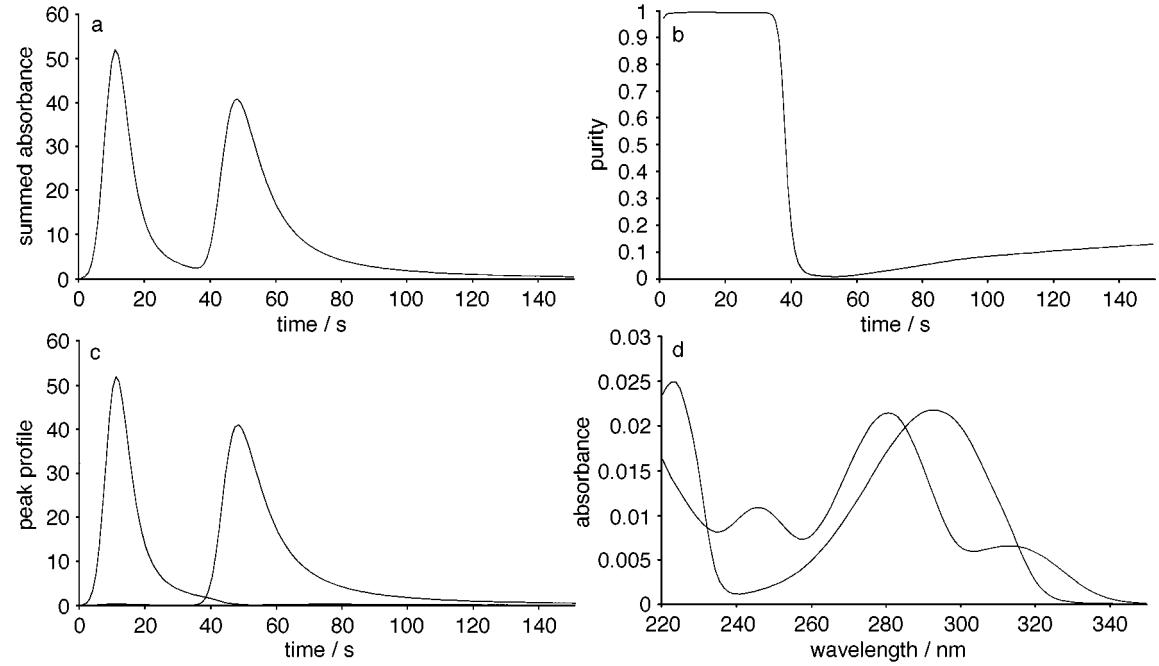

Fig. 4 (a) Summed absorbance at pH 4.8; (b) purity curve at pH 4.8; (c) deconvoluted profiles at pH 4.8; (d) deconvoluted spectra at pH 4.8.

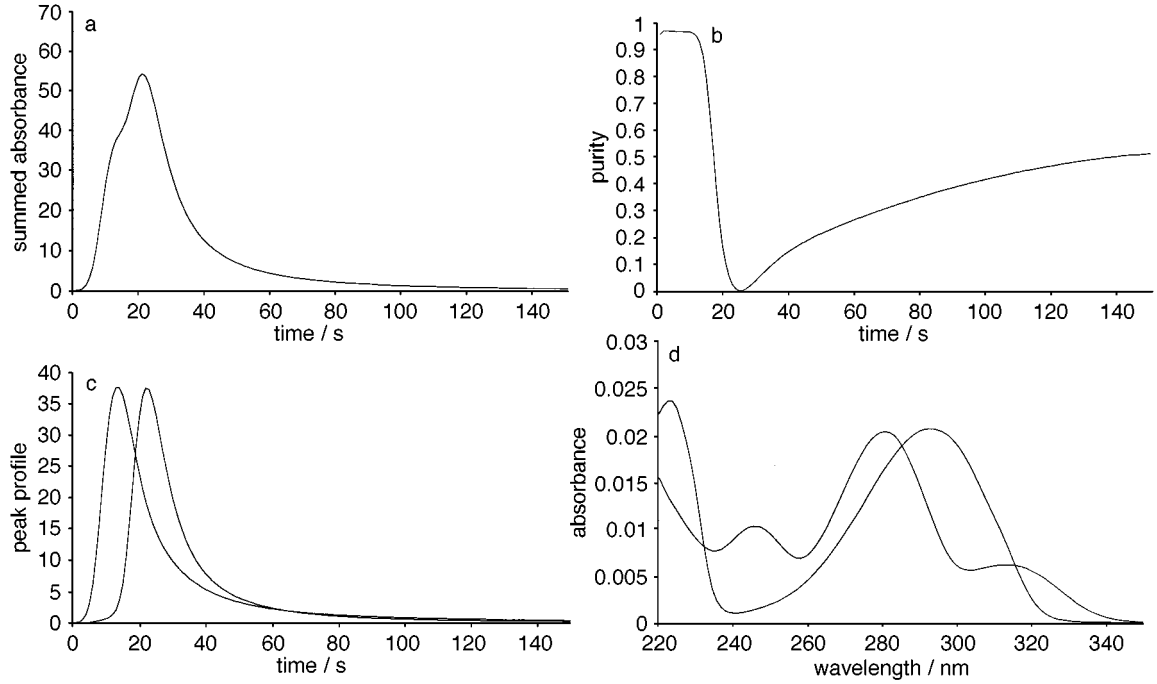

Fig. 5 (a) Summed absorbance at pH 5.2; (b) purity curve at pH 5.2; (c) deconvoluted profiles at pH 5.2; (d) deconvoluted spectra at pH 5.2.

true spectra of the compounds at each pH recorded using chromatograms of each pure component. The shapes of the elution profiles are intuitively reasonable.

Most conventional approaches to deconvolution provide disastrous results at pH 5.1, where the two peaks are almost overlapped, suggesting a weakness in most conventional multivariate approaches for factor analysis. The difficulty is that these approaches often assume either composition one points or pure variables for each significant component, which is not the case in this example. However the purity curves easily illustrate the change in relative proportions of each component, suggesting that even at pH 5.1 it is possible to deconvolute the components in the mixture.

The method proposed in this paper can easily be generalised *via* the criteria employed in the simplex algorithm used to find the optimum value of $p_{\min}$ and so can be amended for different cases and assumptions about a data set. It is even possible to ask a user for information in a number of categories (is there a known composition one region, or is there a positivity constraint, are the spectra known, and are the proportions of each component known, are the chromatographic profiles unimodal) and from this perform the deconvolution. What this paper highlights is that purely statistical solutions are insufficient for the problem described in this paper. Extra information is required to produce a physically meaningful solution.

Interestingly the mathematically simple assumption about equal heights of the deconvoluted profiles, providing the chromatographic peaks are positive, also corresponds to other assumptions such as unimodality in the chromatographic peaks, and positive spectra which are equivalent assumptions that could also be added. Obviously, the more the assumptions the more complex the algorithm and the more the penalties and weightings according to the criterion, but the simple approach outlined in this paper works well.

Acknowledgements

The EPSRC is thanked for financial support of this project.

References

1 R. G. Brereton, *Analyst*, 1995, **120**, 2313.
2 J. Toft, *Chemom. Intell. Lab. Syst.*, 1995, **29**, 189.
3 F. Cuesta Sanchez, S. C. Rutan, M. D. Gil Garcia and D. L. Massart, *Chemom. Intell. Lab. Syst.*, 1997, **36**, 153.
4 A. K. Elbergali and R. G. Brereton, *Chemom. Intell. Lab. Syst.*, 1995, **27**, 55.
5 A. K. Elbergali and R. G. Brereton, *Chemom. Intell. Lab. Syst.*, 1994, **23**, 97.
6 M. Maeder and A. D. Zuberbuehler, *Anal. Chim. Acta*, 1986, **181**, 287.
7 M. Maeder, *Anal. Chem.*, 1987, **59**, 527.
8 K. J. Schostack and E. R. Malinowski, *Chemom. Intell. Lab. Syst.*, 1993, **20**, 173.
9 R. G. Brereton, S. P. Gurden and J. A. Groves, *Chemom. Intell. Lab. Syst.*, 1995, **27**, 73.
10 Y.-Z. Liang, O. M. Kvalheim, A. Rahmani and R. G. Brereton, *J. Chemometrics*, 1993, **7**, 15.
11 O. M. Kvalheim and Y.-Z. Liang, *Anal. Chem.*, 1992, **64**, 946.
12 Y.-Z. Liang and O. M. Kvalheim, *Chemom. Intell. Lab. Syst.*, 1993, **20**, 115.
13 E. Malinowski, *Factor Analysis in Chemistry*, Wiley, New York, 2nd edn., 1991.
14 K. D. Zissis, R. G. Brereton and R. Escott, *Analyst*, 1997, **122**, 1007.
15 R. Manne, *Chemom. Intell. Lab. Syst.*, 1995, **27**, 89.
16 S. Dunkerley, J. Crosby, R. G. Brereton, K. D. Zissis and R. E. A. Escott, *Analyst*, 1998, **123**, 2021.
17 S. Dunkerley, R. G. Brereton and J. Crosby, *Chemom. Intell. Lab. Syst.*, in the press.
18 W. Windig and J. Guilment, *Anal. Chem.*, 1991, **63**, 1425.

Paper 9/01331H