FULL PAPER
THE ANALYST
www.rsc.org/analyst

Direct sampling tandem mass spectrometry (MS/MS) and multiway calibration for isomer quantitation

Cleidiane G. Zampronio,[a] Stephen P. Gurden,[a] Luis Alberto B. Moraes,[a] Marcos N. Eberlin,[a] Age K. Smilde[b] and Ronei J. Poppi*[a]

[a] *Instituto de Química, Universidade Estadual de Campinas-UNICAMP, CP 6154, 13083-970Campinas, São Paulo, Brazil*
[b] *Department of Chemical Engineering, Process Analysis and Chemometrics, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*

Received 25th March 2002, Accepted 21st May 2002
First published as an Advance Article on the web 26th June 2002

Direct sampling tandem mass spectrometry (MS/MS) was used for the quantitation of mixtures of the isomers 2-, 3- and 4-ethyl pyridine. The similarity between the analytes and the second-order nature of MS/MS data require the use of multivariate calibration techniques capable of handling multiway data. Multilinear PLS (N-PLS) was applied here, as well as the alternative technique of unfolding the data and using standard two-way PLS. Particular attention was paid to the optimal type of spectral preprocessing. Due to the presence of heteroscedastic noise the logarithmic transform of the spectra prior to calibration gives the best results. Predictions errors of the order of 10–15% were obtained, which compare well with other results found in the literature.

1. Introduction

Mass spectrometry is a common tool for the analysis of volatile and semi-volatile organic compounds.[1] One common application, which has seen increasing use in recent years, is the analysis of environmental or biological samples such as water, soil and air contaminated with organic compounds. In this area, direct sampling techniques such as membrane introduction mass pectrometry (MIMS) and hyphenated MIMS techniques[2–14] have been extensively applied. The advantage of these techniques is that they allow on-line monitoring of organic compounds with high sensitivity, yet without the need for sample preparation. This is in contrast to gas chromatography/mass spectrometry (GC/MS) and liquid chromatography/mass spectrometry (LC/MS), for which sample preprocessing is usually necessary, followed by chromatographic separation, thus increasing analysis time and effort.

Another advance in the field of mass spectrometry is the increased use of tandem mass spectrometry, also known simply as MS/MS. In this technique, a primary ion produced from the analyte is subsequently mass selected and subjected to collision-induced dissociation (CID), which produces a single ion or a series of product ions. It is known that MS/MS can provide significantly more information than standard mass spectrometry in cases where the analytes exhibit a similar primary fragmentation.[15,16] This is because product ion mass spectra can be significantly different even for similar analytes, thus providing enhanced structural information with increasing selectivity for the analytes of interest.

The use of mass spectrometry as a quantitative tool has also seen recent advances. There is a long history of using mass spectrometry for quantitation based upon finding a selective peak for an analyte of interest (selected-ion monitoring, SIM) and generating a univariate calibration model. However, for analysis of complex mixtures, *e.g.* those containing many analytes or very similar analytes such as isomers, the univariate SIM approach is very limited as finding selective peaks for each analyte of interest is not usually possible. In this case, it is necessary to apply multivariate calibration techniques to the full mass spectrum (total-ion monitoring, TIM). In multivariate calibration, a set of mixture samples of known composition is used to build a regression model, which can subsequently be used to quantify new samples of unknown composition. Whilst a number of multivariate calibration methods exist, the most commonly applied in recent chemical applications is that of partial least squares (PLS).[17–19] Some recent works describe the use of PLS and parallel factor analysis (PARAFAC) for direct sampling mass spectrometry analysis of mixtures containing analytes with similar mass spectra.[20,21] These works clearly demonstrate that multivariate calibration of full mass spectra gives superior results in comparison to the univariate approach.

In this paper, the use of direct sampling tandem mass spectrometry for the quantitative analysis of mixtures of volatile organic isomers is described. Results for multivariate calibration of an experimentally designed set of mixtures of the organic isomers 2-, 3- and 4-ethyl pyridine are presented and discussed.

Multivariate calibration of MS/MS spectra presents a different type of problem to that commonly found with other types of chemical data, such as standard mass spectrometry or near-infrared (NIR) spectroscopy. Unlike these one-dimensional techniques, MS/MS produces a two-dimensional spectrum, $\boldsymbol{X}$ ($J \times K$), where each of the K precursor ions is further fragmented to give a spectrum of J masses describing the product ions. This means that a series of samples will produce a three-way array of data, $\underline{\boldsymbol{X}}$ ($I \times J \times K$), where I is the number of mixture samples. If the concentrations of M different analytes within the mixtures are known, this information can be collected in a corresponding two-way concentration matrix $\boldsymbol{Y}$ ($I \times M$). This situation is shown in Fig. 1. With the increased occurrence of chemical instrumentation producing multiway data, new methods for multivariate calibration of this type of data have been developed. In this article, we discuss the application of some of these methods. In particular, we compare the use of N-PLS with the approach of unfolding the data and applying standard PLS. Some other second-order calibration methods are also discussed, although they were not found to be useful here for reasons discussed later.

The important issue of how to preprocess MS/MS data is also discussed. Some researchers have advocated the use of non-linear transformations such as logarithmic scaling prior to

 DOI: 10.1039/b202948k

calibration in order to combat the problem of heteroscedastic noise (*i.e.* noise of a non-uniform level).[22–24] A separate experiment to investigate the noise distribution in the instrument is described here and different types of preprocessing are discussed.

2. Experimental

The experiments were performed on an Extrel (Pittsburgh, PA) pentaquadrupole mass spectrometer that consists of three mass analyzing (Q1, Q3, Q5) and two reaction or dissociation quadrupoles (q2, q4).[25] The sequential MS/MS spectra were obtained by using the second mass analyzing quadrupole (Q3) to mass-select the precursor ions generated in the ion source by electron ionization (EI) at 70 eV. The product ions were generated in the second dissociation quadrupole (q4) by collision-induced dissociation (CID) with argon at 15 eV. The resulting ions were scanned in the third mass analyzing (Q5) using a mass range of *m*/*z* 10–120. The pressure inside the instrument was 9×10^{-5} Torr.

Each synthetic mixture sample was prepared in an ampoule that was then coupled to the instrument. Direct atmospheric sampling by the vacuum of the mass spectrometer injected the vapor sample into the mass spectrometer for analysis. The final spectrum for each sample is the mean of five successively recorded spectra, each of which is the result of multiple scans (integration time 33 ms). This procedure was used to reduce the amount of noise present in the spectra. The instrument supplies the data in binary form and it was necessary to transform it into ASCII format. The data analysis was performed on a PC using routines written in MATLAB (Version 5.3).[26] The PLS analyses were carried out using the N-way Toolbox for MATLAB.[27]

A set of 24 experimentally-designed samples of the organic isomers 2-, 3- and 4-ethyl pyridine were measured, each sample having a total volume of 100 μL. The experimental design in terms of molar fractions is given in Table 1. The volume and density were used to find the relative mass for each analyte in the mixture. This mass was used to calculate the molar fraction.

Each MS/MS spectrum consists of 11 precursor ions (*m*/*z* 39, 50, 51, 52, 65, 77, 78, 79, 92, 106 and 107) that were selected in Q3 after ionization by EI. These 11 precursor ions are subsquently dissociated to give the product ion mass spectra, scanned in Q5, with a range of *m*/*z* 10 to 120 at *m*/*z* 0.1 resolution. Thus, each spectrum has dimensions 1101×11. A typical MS/MS plot of one of the samples analyzed is given in Fig. 2. The 24 mixture samples are collected to give a three-way array, $\underline{\boldsymbol{X}}$ ($24 \times 1101 \times 11$).

In addition to the mixture samples, an extra set of spectra was measured to investigate the reproducibility and noise distribution of the instrument. A series of ten replicates of perfluorotributylamine (FC43) was measured. This calibration compound was chosen because it yields low, medium and high intensity peaks.

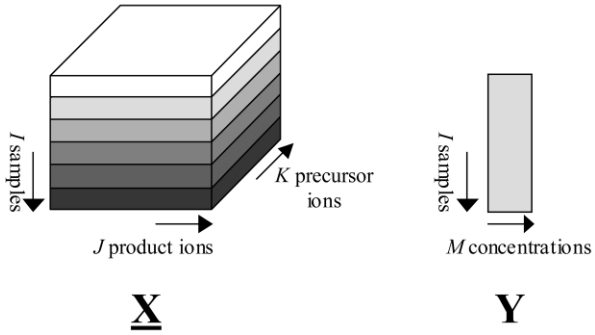

Fig. 1 A three-way array of MS/MS data, $\underline{\boldsymbol{X}}$ ($I \times J \times K$), and a corresponding two-way array of analyte concentrations, $\boldsymbol{Y}$ ($I \times M$).

3. Data preprocessing

Mass spectrometry is often used either solely for qualitative work or for simple quantitation using selective peaks. Data preprocessing is often minimal, for example a simple normalization of the mass spectrum by setting the height of the maximum peak equal to unity. However, when analyzing mixtures with a large number of components, or with very similar components such as isomers, more careful consideration of how to optimally preprocess the data prior to calibration may be required.

Mass spectrometry, and multi-quadrupole mass spectrometry in particular, is known to suffer from relatively poor reproducibility.[28,29] In addition unstable internal pressure of the equipment and contaminant signals can result in significant levels of measurement noise. This noise can be separated into two types: firstly, the overall intensity level of the measured spectrum can vary, as expressed by

$$\boldsymbol{X} = (1 + e)\underline{\boldsymbol{X}} \quad (1)$$

Table 1 Molar fractions of the mixture samples

Training samples	2-Ethyl pyridine	3-Ethyl pyridine	4-Ethyl pyridine
1	1.000	0.000	0.000
2	0.599	0.401	0.000
3	0.399	0.601	0.000
4	0.000	1.000	0.000
5	0.599	0.201	0.201
6	0.199	0.600	0.200
7	0.399	0.300	0.301
8	0.599	0.000	0.401
9	0.299	0.400	0.301
10	0.299	0.300	0.401
11	0.000	0.600	0.400
12	0.362	0.000	0.638
13	0.199	0.200	0.601
14	0.000	0.400	0.600
15	0.000	0.000	1.000
Test samples			
1	0.799	0.100	0.100
2	0.100	0.800	0.100
3	0.499	0.301	0.201
4	0.299	0.501	0.200
5	0.499	0.200	0.301
6	0.199	0.500	0.301
7	0.299	0.200	0.501
8	0.199	0.300	0.501
9	0.100	0.100	0.801

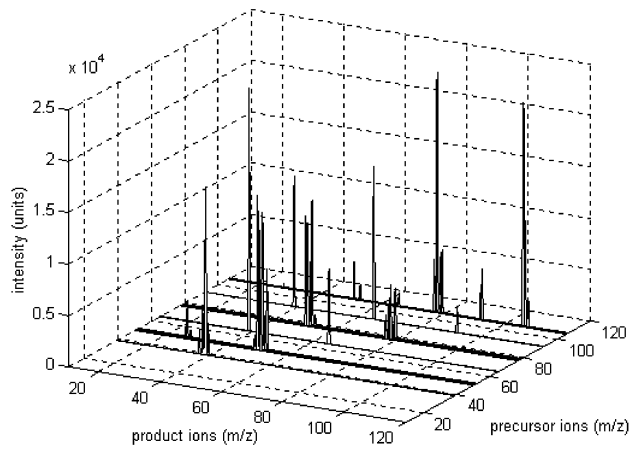

Fig. 2 Plot of the MS/MS product ion mass spectrum for sample 9.

where $\boldsymbol{X}$ is the measured and $\tilde{\boldsymbol{X}}$ is the true signal for one sample, both having dimensions $J \times K$. The scalar noise term, e, influences the overall intensity level and varies from measurement to measurement. This type of noise can be treated by applying some form of normalization, *e.g.* using an internal standard or by normalizing the spectrum to constant sum-of-squares.[22,23]

A second type of noise is variation in individual peaks, expressed by

$$x_{jk} = \tilde{x}_{jk} + e_{jk} \quad (2)$$

where x_{jk} and $\tilde{x}_{jk}$ are the measured and true signals for product ion j from precursor ion k. The noise term here, e_{jk}, may be different at each point of the spectrum. In the case where the level of noise is dependent upon the peak intensity, *e.g.* $e_{jk} = \pm\alpha\tilde{x}_{ij}$, this is known as heteroscedastic noise. As most calibration models give equal weight to the residuals at each variable, it is preferable to transform the noise to be approximately uniform across the spectral range.

As has been noted by Kvalheim *et al.*, Brakstad and others,[22–24] it is difficult to separate these two types of noise, leading to a complex situation. For example, peaks with a high intensity have most influence during spectral normalization and yet, due to heteroscedasticity, these may be the most unreliable. The result is that during the normalization step, noise in the high intensity peaks is transferred to the smaller peaks leading to distortions in the data set.[22] For this reason, it may be preferable to use a transform, such as taking logarithms, to remove heteroscedastic noise *prior* to spectral normalization. In the case of heteroscedastic noise where the noise is proportional to the signal intensity, eqn. (2) can be written as

$$x_{jk} = \tilde{x}_{jk} \pm \alpha\tilde{x}_{ij} = (1 \pm \alpha)\tilde{x}_{jk} \quad (3)$$

Taking logarithms of eqn. (3) gives

$$\log(x_{jk}) = \log((1 \pm \alpha)\tilde{x}_{jk}) = \log(1 \pm \alpha) + \log(\tilde{x}_{jk}) \quad (4)$$

It can now be seen that so long as $\alpha << 1$, $\log(1 \pm \alpha) \approx 0$ and so proportional noise is removed by the logarithmic transform. However, it should be noted that the use of non-linear transformations such as logarithms removes linearity from the data, an undesired side-effect. The general conclusion seems to be that the exact type of preprocessing to use depends upon the relative types and levels of noise within the data. This choice will clearly differ from instrument to instrument and from data set to data set.

The MS/MS spectra of ten replicates of FC43 were measured. Fig. 3 shows a plot of the standard deviation of the signal *versus* the mean signal intensity for all spectral points. It is clearly seen that heteroscedastic noise is present: signal noise increases with signal intensity. Furthermore, the signal standard deviation is approximately proportional to the signal intensity, a situation for which the logarithm transformation may be suitable.[22,23]

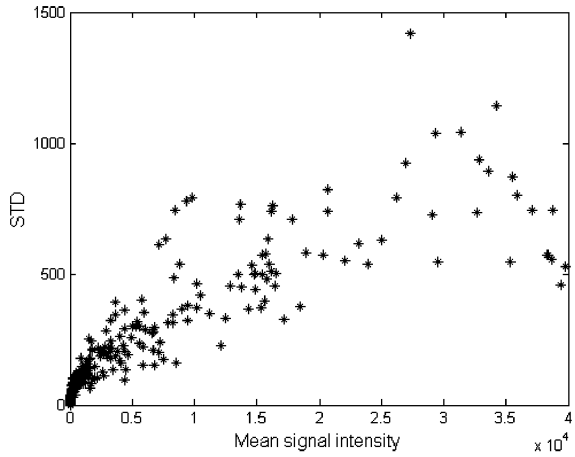

Fig. 3 Standard deviation *versus* mean signal intensity for the ten replicates of FC43.

To investigate the relative merits of different types of preprocessing, five different options were used in the calibrations discussed later: (1) No preprocessing. (2) Normalization by maximum peak (N):

$$x^*_{jk} = \frac{x_{jk}}{\max(\boldsymbol{X})} \quad (5)$$

(3) Log transformation (L):

$$x^*_{jk} = \log(x_{jk} + 1) \quad (6)$$

(4) Normalization by maximum peak followed by log transformation (N,L):

$$x^*_{jk} = \log\left(\frac{x_{jk}}{\max(\boldsymbol{X})} + 1\right) \quad (7)$$

(5) Log transformation followed by normalization by the maximum peak (L,N):

$$x^*_{jk} = \frac{\log(x_{jk} + 1)}{\max(\log(\boldsymbol{X} + 1))} \quad (8)$$

The results of the comparison are discussed in Section 5.

4. Multivariate calibration

Partial least squares (PLS) is a multivariate calibration technique well established within the chemical community.[17–19] Like all regression methods, it aims to find a relationship between a set of predictor data, $\boldsymbol{X}$, and a set of responses, $\boldsymbol{Y}$. A major difference, however, between PLS and simple multiple linear regression (MLR) (or ordinary least squares, OLS) is that PLS is able to give stable predictions even when $\boldsymbol{X}$ contains highly correlated variables (a common situation for spectrometric data). Unlike univariate techniques, PLS will work even if unknown interferents are present in new samples, provided that these interferents are also present in the data used to train the model.

In recent years, the number of regression problems within chemistry for which the predictor data has more than two dimensions has grown, due to the increase in hyphenated instrumentation and multiway spectrometry, such as fluorescence and MS/MS.[30] The original PLS algorithm was designed for two-way data (*e.g.* sample $\times$ wavelength) and a number of researchers have found ways of adapting this algorithm to handle the multiway situation.[31] For PLS, there are two main approaches:

4.1 Unfold-PLS

In the unfold-PLS (U-PLS)[32] the multiway data, $\bar{\boldsymbol{X}}$ is first rearranged ('matricized' or 'unfolded') to produce a two-way array. Thus, the three-way array, $\bar{\boldsymbol{X}}$ ($I \times J \times K$), is unfolded to give a two-way array, $\boldsymbol{X}$ ($I \times JK$). It is then possible to use the standard two-way PLS algorithm on the two-way data. For a detailed description of the PLS algorithm, the reader is referred to the literature,[17,18] the main idea is to find bilinear components in $\boldsymbol{X}$ which have maximum covariance with a univariate response, $\boldsymbol{y}$. This is expressed by

$$\max_{\boldsymbol{w}_a}\left[\operatorname{cov}(\boldsymbol{t}_a, \boldsymbol{y}) \middle| \min\left(\sum_{i=1}^{I}\sum_{jk=1}^{JK}\left(x_{i,jk} - t_{a,i}w_{a,jk}\right)^2\right)\right] \quad (9)$$

where $\boldsymbol{t}_a$ ($I \times 1$) are the PLS scores and $\boldsymbol{w}_a$ ($JK \times 1$) are the PLS weights for the ath model component.

The PLS model is used in two stages: first, a set of training samples is used to build the model. From this model, it is possible to calculate a set of regression coefficients, which can then be used to make predictions for new samples:

$$\hat{\boldsymbol{y}}_{\text{new}} = \boldsymbol{X}_{\text{new}}\boldsymbol{b}_{\text{PLS}} \quad (10)$$

In the case of multiple responses, *e.g.* concentrations for M different analytes, the most common approach is to build a separate PLS model for each response.

4.2 Multilinear PLS

Multilinear PLS (N-PLS) is an extension of the two-way PLS algorithm for cases where the predictor data, $\bar{\boldsymbol{X}}$, is an array of order higher than two.[33] Like PLS, the aim is to find components in $\bar{\boldsymbol{X}}$ which have maximum covariance with $\boldsymbol{y}$, but here multilinear components are used. In the case of a three-way $\bar{\boldsymbol{X}}$, trilinear components are used:

$$\max_{\boldsymbol{w}_a^J,\boldsymbol{w}_a^K}\left[\operatorname{cov}(\boldsymbol{t}_a,\boldsymbol{y})\left|\min\left(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(x_{ijk}-t_{ia}w_{a,j}^{J}w_{a,k}^{K}\right)^2\right)\right.\right] \quad (11)$$

where it can be seen that now two weights vectors are used for each model component: $\boldsymbol{w}_a^J$ ($J \times 1$) and $\boldsymbol{w}_a^K$ ($K \times 1$). Thus, in contrast to two-way PLS, each dimension of the three-way array is described by its own scores or weights vector. This situation is shown in Fig. 4.

N-PLS imposes a trilinear structure on the set of MS/MS data used to train the model. However, it is in no way clear to what extent a trilinear structure is actually present in the data. Certainly, the matrix produced from one sample is not bilinear as, for example, it is for other second-order instrumentation such as fluorescence spectroscopy.[34] While it is true that there are some related patterns in the fragmentations of the different precursor ions, there is no direct relationship between the various product ion spectra. However, for N-PLS, this lack of clear trilinear structure is not considered critical, as the main aim is to find a subspace of $\boldsymbol{X}$ which can be used for regression. It has been shown in the literature that multilinear regression methods can work well even on data with no strict multilinear structure.[31]

One possible advantage of N-PLS when applied to MS/MS data is that it is possible to plot separate model weights for the precursor ions (dimension K) and the product ions (dimension J). In U-PLS, the two dimensions are mixed up in the unfolding step, which could make model interpretation more complicated. Furthermore, the use of a multilinear structure for the PLS weights means than N-PLS uses fewer model parameters than U-PLS to describe the data. This may mean that the N-PLS model is more robust to the influence of noise in the data. In this paper, U-PLS and N-PLS are both used in order to investigate whether there are significant differences between the approaches.

4.3 Variable selection

In multivariate calibration, it is sometimes found that prediction accuracy can be improved by using only selected regions of the spectrum.[35–37] In particular, it is sometimes found that removing regions containing significant levels of background noise or interferent signal can improve the calibration. MS/MS data exhibits an unusual feature in that a large part of the data array may contain no information. This is because it is impossible for a precursor ion to produce product fragments with a higher mass than itself (see Fig. 2). One option, therefore, would be to immediately remove this part of the array from any analysis. However, in practice, it was found that removing this region, or even using selected peaks, did not significantly influence the results. One reason for this could be the relative absence of any background interferents or matrix effects in the synthetic mixtures. Therefore, the full spectral region is used as this enables easier plotting and interpretation of the PLS weights.

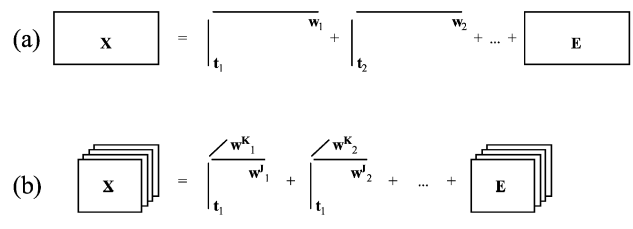

Fig. 4 Decompositions of $\bar{\boldsymbol{X}}$ for (a) U-PLS and (b) N-PLS.

4.4 Other second-order calibration methods

In addition to N-PLS, there exists a family of second-order calibration techniques which relies more strictly on the trilinear structure within the data. These techniques, such as the generalized rank annihilation method (GRAM),[38] residual bilinearization (RBL)[39] and restricted Tucker models,[40] have attracted attention because they offer the possibility of being able to handle unknown interferents in new samples (the so-called 'second-order advantage'[41]), something PLS-type methods cannot do.

Although these techniques were designed for bilinear data, some work has been reported in extending the methodology to non-bilinear data. In particular, Wang *et al.*[41] reported a limited example of the use of NBRA for the quantitation of MS/MS data. However, attempts to apply both the non-bilinear rank annihilation (NBRA) and residual bilinearization (RBL) methods to the MS/MS data described here were not successful, yielding very high prediction errors. The probable reasons for this were twofold: firstly, it is known that these techniques are quite sensitive to experimental noise;[42] secondly, the complexity of the MS/MS spectra means that a low-rank bilinear approximation of the data was not possible in this case. However, the use of these techniques remains of interest and will continue to be investigated.

5. Results and discussion

The data were split into two sets: a training set consisting of 15 samples and a test set consisting of 9 samples, as shown in Table 1. Four types of preprocessing were used and two types of calibration model, U-PLS and N-PLS. A separate calibration model was built for each analyte and in each case the number of model components was selected using leave-one-out cross-validation with a preference for a low number of model components. The calibration results for the test set are given in terms of the percentage standard error of prediction, %SEP:

$$\%\text{SEP} = \frac{100}{\bar{y}}\sqrt{\frac{\sum_{n=1}^{N}(y_n-\hat{y}_n)^2}{N}} \quad (12)$$

where y_n and $\hat{y}_n$ are respectively the true and predicted concentrations for test sample n and $\bar{y}$ is the mean test sample concentration. The results for U-PLS are given in Table 2 and those for N-PLS in Table 3.

It can be seen that quantification of 2-ethyl pyridine is easier than that for 3- and 4-ethyl pyridine. It was also found that only three PLS model components were needed for 2-ethyl pyridine, whereas five were needed for the other two analytes. Although the spectrum of 2-ethyl pyridine has no selective peaks, it is the most dissimilar of the three isomers due to the absence of a precursor ion of m/z 92.

The electron density in pyridine is relatively high at the 3-position, thus making the *m/z* 92 the most favorable carbonium ion. The fragments formed by ß-fission undergo further elimination of hydrogen cyanide resulting in *m/z* 65.[43] The 3-ethyl pyridine has a very high intensity of these ions and the 4-ethyl pyridine shows a medium intensity of the *m/z* 92 and 65. However, the ion *m/z* 92 does not exist for the 2-ethyl pyridine, because the small quantity that is present is fragmented to *m/z* 65, that has very low intensity for this isomer.

Furthermore, 2-ethyl pyridine has a higher vapor pressure (4.88 mmHg) than the 3- and 4-ethyl pyridines (2.55 mmHg and 2.22 mmHg respectively), which may mean that this isomer suffers less from variation due to the transport of the mixture sample from the ampoule into the instrument.

It is also seen that the calibration for 2-ethyl pyridine is not so dependent upon the type of preprocessing used. For 3- and 4-ethyl pyridine, the use of the logarithm transform to alleviate heteroscedastic noise is clearly advantageous. Furthermore, it is clearly better to apply logarithms before normalization (L,N), for reasons discussed earlier. It seems that for this data, the non-linearizing effect of the logarithmic transformation can be compensated for by using extra PLS components.

Of the two calibration methods, N-PLS seems to work slightly better than U-PLS when L,N scaling is used, although this result is not necessarily significant given the limited number of test samples. N-PLS does seem to be more sensitive to the use of incorrect preprocessing, perhaps because the more rigid model structure is less flexible to non-linearities in the data.

For the calibration using L,N preprocessing and N-PLS, the PLS weights describing the precursor ion dimension for the three isomers are shown in Fig. 5. It is seen that the higher *m/z* precursor ions tend to have a higher weight; this being because the heavier ions give a larger number of different product ions. To find the most important fragments, it is necessary to look at the weights for the product ions. These weights for the three N-PLS components for 2-ethyl pyridine are shown in Fig. 6. It is seen that product ions across the whole *m/z* range are significant, but the *m/z* 92 ion is the unique one that presents a high weight in the three components. It is important to comment also that the *m/z* 106 and 107 that are very important, as precursor ions in the product ions do not show high weight. The product ion weights for 3- and 4-ethyl pyridine (not shown here) exhibited a similar pattern to those for 2-ethyl pyridine and were very similar to each other. The pure spectra for these compounds are very similar, exhibiting exactly the same set of fragment ions, but at different intensities. The calibration of mixtures of isomers, which have very similar spectra, as there are no selective peaks present, is based on the differences in relative intensities of a series of product ion peaks.

When using the log transformation and N-PLS, the percentage standard errors of prediction are between 10 and 15%. This compares well with other reports of quantitation using direct sampling mass spectrometry.[20,21] In general, it is found that

Table 2 Calibration results for the five different preprocessing methods using U-PLS

	%SEP		
Preprocessing	2-Ethyl pyridine	3-Ethyl pyridine	4-Ethyl pyridine
1. No preprocessing	11.7	47.7	31.7
2. N	14.7	19.9	32.7
3. L	11.4	15.3	20.8
4. N,L	13.5	15.6	26.4
5. L,N	11.2	15.3	20.3

Table 3 Calibration results for the five different preprocessing methods using N-PLS

	%SEP		
Preprocessing	2-Ethyl pyridine	3-Ethyl pyridine	4-Ethyl pyridine
1. No preprocessing	12.6	68.0	55.8
2. N	19.9	34.8	52.8
3. L	10.9	13.4	14.3
4. N,L	18.1	26.5	42.6
5. L,N	11.1	12.3	15.0

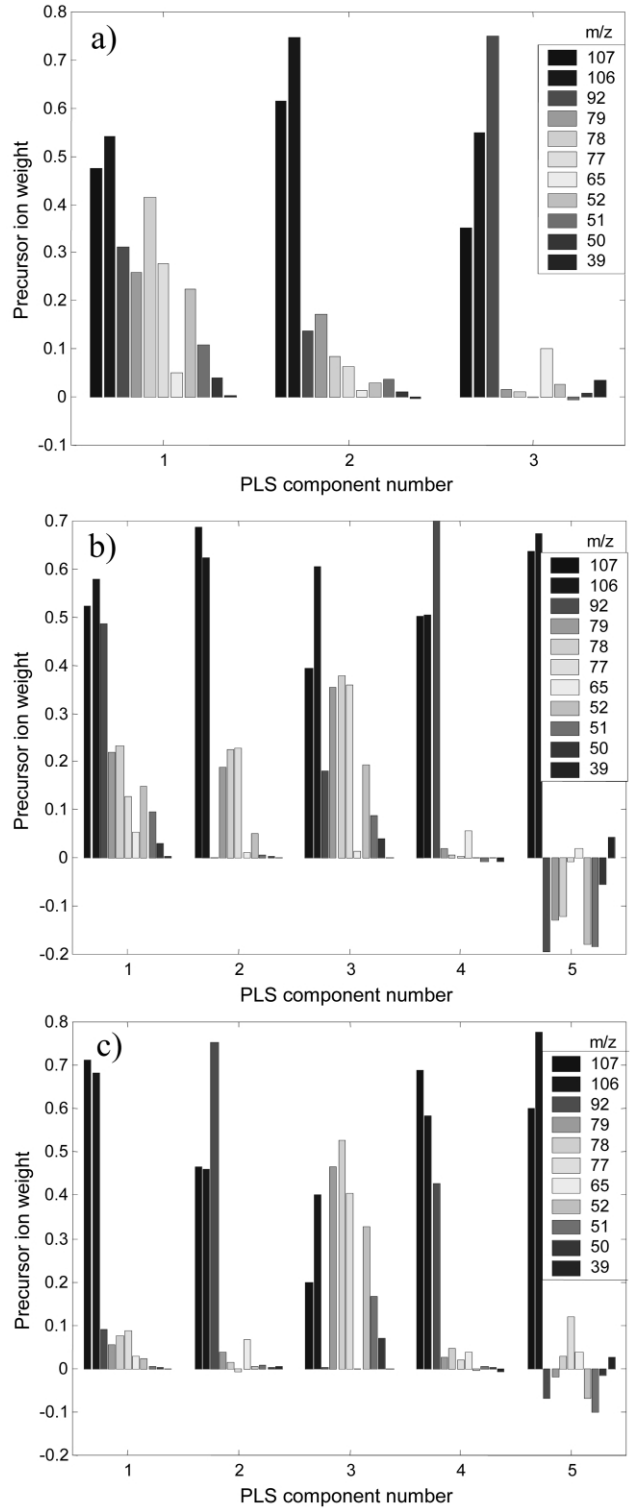

Fig. 5 N-PLS weights for the precursor ion dimension using L,N preprocessing: (a) 2-ethyl pyridine, (b) 3-ethyl pyridine and (c) 4-ethyl pyridine.

predictions for compounds at a low concentration are slightly worse than for those at a high concentration. This could be due to the use of the log transformation, which has a higher influence on linearity for low intensities. An alternative to using data transforms to counter heteroscedastic noise is the use of weighted regression methods.[44–46] However, these require detailed knowledge of the noise distribution across the spectrum, which may not always be available.

6. Conclusions

In this paper, it has been shown that the enhanced selectivity provided by direct sampling MS/MS coupled with multivariate calibration can work well for the quantitation of organic isomers. The fragment at *m/z* 92 (a product ion from precursor ions *m/z* 106 and 107) is found to be important, because this ion is present for 3- and 4-ethyl pyridine, but is not hardly produced by 2-ethyl pyridine.

Some multivariate calibration methods were applied in the MS/MS data, but they do not show better results than N-PLS. Prediction errors were of the order of 10–15%, a result which would not be possible using univariate calibration techniques for analytes with such similar spectra. It has been shown that correct preprocessing of MS/MS data is critical and that the use of the logarithmic transform prior to normalization seems to work well.

Despite the lack of a clear trilinear structure in this MS/MS data, multilinear PLS was found to give good results and provides a straightforward method of multivariate calibration for MS/MS data in which the most important product (and precursor) ions can be determined by plotting the model weights.

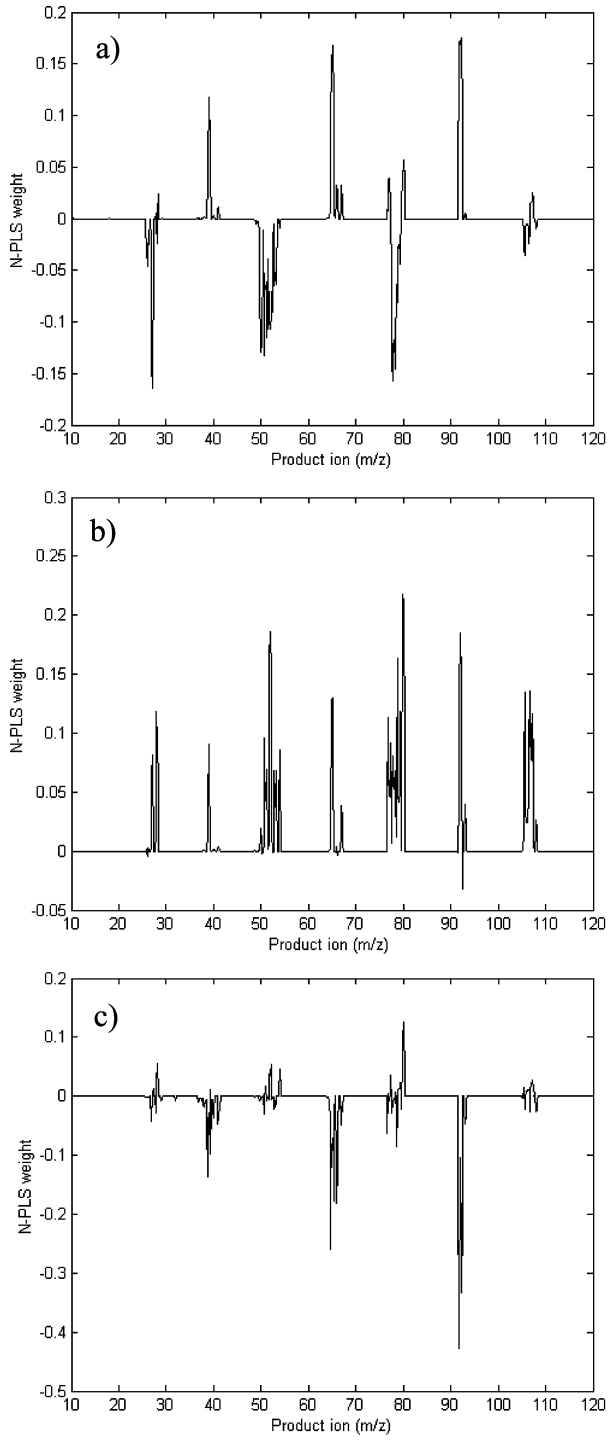

Fig. 6 N-PLS weights for components (a) 1, (b) 2 and (c) 3 for the product ion dimension using L,N preprocessing to calibrate 2-ethyl pyridine.

Acknowledgements

Financial support from the State of São Paulo Research Foundation - Brazil (FAPESP) and the Brazilian National Research Council (CNPq) is gratefully acknowledged. CGZ would like to thank the Process Analysis and Chemometrics group at the University of Amsterdam for their hospitality during the period in which some of this work was carried out.

References

1 R. A. Hites, *Biotherapy*, 1998, **11**, 77.
2 L. S. Riter, Z. Takats, L. Charles and R. G. Cooks, *Rapid Commun. Mass Spectrom.*, 2001, **15**, 1520.
3 T. Aggerholm and F. R. Lauritsen, *Rapid Commun. Mass Spectrom.*, 2001, **15**, 1826.
4 T. M. Allen, T. M. Falconer, M. E. Cisper, A. J. Borgerding and C. W. Wilkerson, *Anal. Chem.*, 2001, **73**, 4830.
5 C. Shang and E. R. Blatcheley, *Water Res.*, 2001, **35**, 244.
6 R. Haddad, M. A. Mendes, N. F. Hoehr and M. N. Eberlin, *Analyst*, 2001, **126**, 1212.
7 C. Shang and E. R. Blatchley, *Environ. Sci. Technol.*, 1999, **33**, 2218.
8 R. Kostiainen, T. Kotiaho, I. Mattila, T. Mansikka, M. Ojala and R. A. Ketola, *Anal. Chem.*, 1998, **70**, 3028.
9 R. A. Ketola, V. T. Virkki, M. Ojala, V. Komppa and T. Kotiaho, *Talanta*, 1997, **44**, 373.
10 V. Lopez-Avilla, J. Benedicto, H. Prest and S. Bauer, *Am. Lab.*, 1999, **31**, 33.
11 M. A. Mendes, R. Sparrapan and M. N. Eberlin, *Anal. Chem.*, 2000, **72**, 2166.
12 F. R. Lauritsen and R. A. Ketola, *Anal. Chem.*, 1997, **69**, 4917.
13 R. Kostiainen, T. Kotiaho, I. Mattila, M. Ojala and R. A. Ketola, *Anal. Chem.*, 1988, **70**, 3028.
14 M. A. Mendes, R. S. Pimpim, T. Kotiaho and M. N. Eberlin, *Anal. Chem.*, 1996, **68**, 3502.
15 K. L. Busch, G. L. Glish and S. A. McLuckey, *Mass Spectrometry/ Mass Spectrometry Techniques and Applications of Tandem Mass Spectrometry*, VCH, Weinheim, 1989.
16 C. G. Zampronio, L. A. B. Moraes, M. N. Eberlin and R. J. Poppi, *Anal. Chim. Acta*, 2001, **446**, 495.
17 P. Geladi and B. R. Kowalski, *Anal. Chim. Acta*, 1986, **185**, 1.
18 H. Martens and T. Næs, *Multivariate Calibration*, Wiley, Chichester, 1989.
19 P. Geladi, *Chemom. Intell. Lab. Syst.*, 2002, **60**, 211.
20 S. K. Ohorodnik, R. E. Shaffer, J. H. Callahan and S. L. R. Pehrsson, *Anal. Chem.*, 1997, **69**, 4721.
21 W. P. Gardner, R. E. Shaffer, J. E. Girard and J. H. Callhan, *Anal. Chem.*, 2001, **73**, 596.
22 O. M. Kvalheim, F. Brakstad and Y. Liang, *Anal. Chem.*, 1994, **66**, 43.
23 F. Brakstad, *Chemom. Intell. Lab. Syst.*, 1995, **29**, 157.

24 A. M. Woodward, B. K. Alsberg and D. B. Kell, *Chemom. Intell. Lab. Syst*., 1998, **40**, 101.
25 V. F. Juliano, F. C. Gozzo, M. N. Eberlin and C. Kascheres, *Anal. Chem*., 1996, **68**, 1328.
26 MATLAB v. 5.3, The Math Works Inc., 1999.
27 C. A Anderson and R. Bro, *Chemom. Intell. Lab. Syst*., 2000, **52**, 1.
28 G. Junk and H. Svec, *Anal. Chim. Acta*, 1962, **28**, 164.
29 P. Roose and U. A. T. Brinkman, *J. Chromatogr., A*, 1998, **799**, 233.
30 G. Stecher, C. W. Huck, M. Popp and G. K. Bonn, *Fresenius' J. Anal. Chem*., 2001, **371**, 73.
31 S. P. Gurden, J. A. Westerhuis, R. Bro and A. K. Smilde, *Chemom. Intell. Lab. Syst*., 2001, **59**, 121.
32 S. Wold, P. Geladi, K. Esbensen and J. Öhman, *J. Chemom*., 1987, **1**, 41.
33 R. Bro, *J. Chemom*., 1996, **10**, 47.
34 R. Bro, *Chemom. Intell. Lab. Syst*., 1997, **38**, 149.
35 F. Lindgren, P. Geladi, S. Rannar and S. Wold, *J. Chemom*., 1994, **8**, 349.
36 C. H. Spiegelman, M. J. McShane, M. J. Goetz, M. Motamedi, Q. L. Yeu and G. L. Cote, *Anal. Chem*., 1998, **70**, 35.
37 L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck and S. B. Engelsen, *Appl. Spectrosc*., 2000, **54**, 413.
38 E. Sanchez and B. R. Kowalski, *Anal. Chem*., 1986, **58**, 496.
39 Y. Wang, O. S. Borgen and B. R. Kowalski, *J. Chemom*., 1993, **7**, 439.
40 A. K. Smilde, R. Tauler, J. M. Henshaw, L. W. Burgess and B. R. Kowalski, *Anal. Chem*., 1994, **66**, 3345.
41 Y. Wang, O. S. Borgen and B. R. Kowalski, *J. Chemom*., 1993, **7**, 117.
42 N. M. Faber, *J. Chemom*., 2001, **15**, 743.
43 H. Budzikiewicz, C. Djerassi and D. H. Williams, *Mass Spectrometry of Organic Compounds*, Holden-Day, San Francisco, 1967, p. 566.
44 M. Davidian and P. D. Haaland, *Chemom. Intell. Lab. Syst*., 1990, **9**, 231.
45 M. E. Zom, R. D. Gibbons and W. C. Sonzogni, *Anal. Chem*., 1997, **69**, 3069.
46 F. Laborda, J. Medrano and J. R. Castillo, *J. Anal. At. Spectrom*., 2001, **16**, 732.