

Simulation of high- and low-resolution mass spectra for assessment of calibration methods

Phillip G. Turner^{1*}, Steve Taylor¹, John Y. Goulermas¹ and Kim Hampton²

¹Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3BX, UK

²Lidar Technologies Ltd., Arctic House, Rye Lane, Dunton Green, Sevenoaks TN14 5HD, UK

Received 13 June 2006; Revised 24 November 2006; Accepted 24 November 2006

Calibrating mixtures of residual gases in quadrupole mass spectrometry (QMS) can be difficult since low m/z ratios of molecular ions and their fragments result in overlap of signals especially in the lower mass regions. This causes problems in univariate calibration methods and encourages use of full spectral multivariate methods. Experimental assessment of regression methods has limitations since experimental sources of error can only be minimised and not entirely eliminated. A method of simulating full spectra at low and high resolution to accurate masses is described and these are then used for a calibration study of some popular linear regression methods [classical least squares regression (CLS), partial least squares (PLS), principal component regression (PCR)].
Copyright © 2007 John Wiley & Sons, Ltd.

Quadrupole mass spectrometry (QMS) is a low-cost mass spectrometric method with diverse applications. QMS is traditionally used with a front-end chromatographic separation technique (e.g. GC/MS, LC/MS). In direct sampling mass spectrometry, no separation method is applied and all components reach the analyser simultaneously.

The quadrupole mass filter is commonly used as a residual gas analyser. It consists of four electrode rods set parallel to each other in a quadupolar arrangement, to which radio-frequency (rf) and direct current (dc) voltages are applied. The resulting electromagnetic fields cause ions injected into the mass filter with a particular charge-to-mass ratio to have stable (or unstable) ion trajectories as the dc and rf voltages are scanned. Ideally, the quadrupole electrodes should exhibit a hyperbolic profile in which case the ion motion will be governed by the Mathieu equation.¹ There are a number of regions in which the ion motion is stable and these stable regions (zones) may be represented on the Mathieu stability diagram. The stability diagrams define parameters, a and q , which correspond to regions of stable trajectories. The a and q parameters are defined by:

$$a_u = a_x = -a_y = \frac{4eU}{m\omega^2r_0^2} \quad (1)$$

$$q_u = q_x = -q_y = \frac{2eV}{m\omega^2r_0^2} \quad (2)$$

where V is the peak-to-peak rf amplitude, U is the dc voltage applied between pairs of electrodes of opposite polarity, m is the ionic mass, e is the electronic charge, ω is the angular frequency of the rf voltage and r_0 is the inner radius of the quadrupole.

*Correspondence to: P. G. Turner, Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3BX, UK.

E-mail: phillturner@hotmail.com; s.taylor@liv.ac.uk

Computer simulations of quadrupole systems^{2–7} simulate individual ion motion through the analyser. Calculating the trajectories of large numbers of ions (e.g. 10^6) at each point on the mass scale then allows mass spectral peak shape and resolution variations to be theoretically simulated.

Although most commercial QMS instruments operate in zone I of the Mathieu stability diagram ($a = 0.23$ and $q = 0.71$), other zones can also be used to produce mass spectra. High resolution is achievable in zones II, III and IV.^{7–9} This study incorporates zone I and zone III simulations ($a, q \sim 3, 3$). Zone III has the ability to resolve components such as CO^+ (mass-to-charge ratio (m/z) = 27.9949) and N_2^+ (m/z 28.0061)¹⁰ and CH_3^+ (m/z 15.0235) and $^{15}\text{N}^+$ (m/z 15.0001).¹¹ Using the QMS zone IV ($a, q = 0.002, 21.3$) with an ion energy of 40 eV a resolution at half-height ($R_{1/2}$) of 13 900 for the $^{39}\text{K}^+$ ions has been achieved.⁹ In zones III and IV operation at high ion energies is usual.

Such high-resolution QMS can considerably increase the number of peaks within the spectrum, even for low molecular weight residual gases. These peaks occur due to isotopic combinations. For example, a carbon monoxide (CO) spectrum displays a single isotopic peak at mass 29 on a low-resolution QMS instrument (e.g. $R_{1/2} < 1000$); however, for $R_{1/2} > 30\,000$, two peaks would be resolved corresponding to the isotopic species combination: $^{13}\text{C}^{16}\text{O}^+$ (m/z 28.9983) and $^{12}\text{C}^{17}\text{O}^+$ (m/z 28.9991).

The additional peaks in the spectrum observed using high-resolution instruments ($R > 1000$) should improve the performance of chemometric/multivariate calibration methods by making spectra more distinguishable. A consequence of operating at higher resolution, however, is the requirement for more data points per mass unit to ensure there are enough points to characterise the high-resolution

peaks. The corresponding data file sizes are therefore significantly larger.

Experimental mass spectra will vary due to instrumental parameters, ion source settings, electron energy and, in addition, to measurement factors, gas composition, pressures, temperature, and vacuum stability. The final mass spectrum may be considered to be a linear combination of the pure component spectra, at their respective partial pressures.

In this paper we extend our simulation method used previously for the study of single mass peaks²⁻⁴ by generation of full scale mass spectra that also incorporate isotopic components and combinations. Simulations are presented for higher resolution (zone III) and normal operating region (zone I) spectra. The resulting simulation spectra are then used to assess commonly used multivariate chemometric methods. These methods are of increasing importance as QMS systems are now frequently employed as 'electronic noses' (e-noses) in a range of applications. Within this study we also apply our model to consider the improvements in calibration that may be offered when using high-resolution QMS.

CALIBRATION OF QMS SYSTEMS

Quadrupole mass spectrometers can be difficult to calibrate and there are recommended methods to be followed.^{12,13} Early chemometric methods were limited in the amounts of data that could be applied to the calibration.¹⁴ Multivariate quantitative approaches to mass spectrometry have arisen through the use of algorithms attempting to fit several calibration spectra to a single sample spectrum.

Multivariate methods are able to analyze a number of components simultaneously and to employ the entire spectrum of concentrations instead of a single concentration. Univariate methods, on the other hand, are simple models restricted to the analysis of single components. In algorithmic terms, the former type allows prediction of multiple outputs, while the latter of a single output only. Mass scales in mass spectrometry also require calibration as they are prone to mass misalignment. Mass calibration is not considered here as this work is concerned with quantifying measured gas concentrations. The impact of higher mass resolution operation on quantifying concentrations is assessed.

Multivariate methods can avoid tuning errors and eliminate the need to manually and regularly tune the mass range. In multivariate analysis the full spectrum is used in the calibration and misalignment in a sample will be identical to that in the calibration standard. Additionally, using full analogue mass spectra (typically >16 sample points for each atomic mass unit) avoids having to manually select the relevant and required masses.

In this study several traditionally used calibration methods are assessed. Principal component regression (PCR), partial least squares (PLS), classical least squares regression (CLS) and ordinary least squares regression (OLS)¹⁵ are compared and contrasted. OLS calibration models are built using only pure components. Thus OLS is often a preferred calibration method in mass spectrometry due to its simplicity and avoidance of expensive calibration mixtures. There are many publications comparing these

multivariate techniques, in the domains of NMR, IR and UV-VIS spectroscopy.¹⁶ In this paper spectra are simulated to assess these calibration methods specifically for residual gas analysis (RGA) using QMS.

PRINCIPLES OF CALIBRATION METHODS

Calibration data containing the concentrations of each component present in a mixture can be tabulated. The table of concentration values can be also expressed as a data matrix. Each row of the data matrix represents each calibration sample with the corresponding concentrations in each column. This information forms the data matrix \mathbf{X} . When a gas mixture is measured (corresponding to each row of \mathbf{X}) each spectrum is expressed as a vector y and all the vectors are combined to form data matrix \mathbf{Y} .

The two matrices are related since the signal intensities in \mathbf{Y} correspond to the concentrations in \mathbf{X} :

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E} \quad (3)$$

where matrix \mathbf{E} contains the associated errors and \mathbf{B} are the regression coefficients. An approximation can be calculated using the least squares method such that:

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (4)$$

The constituents of matrix \mathbf{X} are a set of real numbers ($x_{ij} \in \mathbb{R}$). For a given matrix \mathbf{X} , dimensions ($n \times d$) can be plotted in a d -dimensional space. In some situations d is higher than needed; for example, data on a plane in a three-dimensional space can be projected onto two dimensions with no loss in information. PLS and PCR methods¹⁷ act to reduce the dimensionality of the data represented by \mathbf{X} . Problems can occur with colinearity but PCR and PLS avoid this as the eigenvectors are constrained to be orthogonal.

PCR is based on the fact that the eigenvectors which correspond to the largest eigenvalues are the directions that explain best the data variance, while the eigenvalues signify the amount of the variance contained in those directions. Thus, by ignoring dimensions with minor variance contributions, the dimensionality of \mathbf{X} is reduced by linearly projecting the data on a subspace. This subspace defined by the orthogonal basis is taken to be the significant eigenvectors. Subsequently, the regression in (3) can be implemented in this new space with the likely data correlations removed.

PLS is also based on linear combinations of the regression inputs but, in addition to \mathbf{X} , it also uses the information in \mathbf{Y} . PLS recovers directions in the data space that have high variance of the projected \mathbf{X} and also high correlation with the response \mathbf{Y} , while PCA is only concerned with high variance in the input space. In this study, fast PLS methods have been used as a result of the large \mathbf{Y} data matrices.¹⁸ All calculations were carried out in MATLAB.¹⁹

For factor-based methods the number of factors retained will alter the prediction values. The decision on the number of factors to be retained was carried out by cross-validation.

Factor-based methods (PLS and PCR) remove redundant information in \mathbf{X} and \mathbf{Y} matrices, i.e. noise contributions and correlations. This improves the quality of regression coefficients and consequently the calibration.

The multivariate methods here rely on the mass spectrometer system behaving linearly. For each calibration, errors of prediction (RMSEP) are evaluated using the training/calibration data set X_{cal} and an independent test set X_{test} . The RMSEP value provides a performance indicator of the calibration; the lower the RMSEP value the better the calibration (using calibration data A):

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (a_i - \hat{a}_i)^2}{n}} \quad (5)$$

where \hat{a}_i are the predicted concentration values and a_i the true concentrations in the i^{th} mixture. Here n is the number of samples.

RESULTS

Isotopic ratio calculations

Any mass spectral simulation method has to be specific to a measurement situation, incorporating both instrumental factors and the measurement conditions. To assess calibrations with simulated data, the data must be representative of true mass spectrometry measurements.

The first objective in preparing the simulated spectra is to determine the position of all the spectral peaks on the mass scale and then their corresponding signal intensities. This involves including the isotopic data to determine the accurate mass for each atomic combination. This can result in complex peak patterns which are closely spaced along the mass scale. The isotopic ratios were taken from Rockwood *et al.*²⁰

The atomic composition of each molecular fragment was identified prior to calculation of the isotopic ratios. First the ratios for each atom type were calculated and then these were combined with the calculated ratios of other atom types within a given molecule or fragment. To produce simulated spectra three different scenarios have been considered and incorporated:

- (i) Ions consisting of a single atom (e.g. Ar); in this case the accurate masses of each isotope can be used directly with no further calculations required.
- (ii) Ions containing multiple atoms each with two isotopes, (e.g. CH_4 with isotope ratios: $^{12}\text{C}/^{13}\text{C}$, $^1\text{H}/^2\text{H}$); in this case the isotopic masses and abundances were calculated using the binomial expansion.
- (iii) Ions containing multiple atoms each with three or more isotopes (e.g. $^{16}\text{O}/^{17}\text{O}/^{18}\text{O}$); these were accommodated using a polynomial expansion (e.g. for O_2 in CO_2).²¹

The ratios of the fragments were deduced from NIST mass spectral data.²² The isotope calculations were then combined into a single data matrix representing the ratios of fragments and containing all the spectral peak positions on the mass scale and their corresponding signal intensities.

Characterising QMS scan lines

Our quadrupole simulation program has previously been successfully used to model single mass spectral peaks for QMS.^{2-4,6} Correct voltage ratios (U/V), electrode geometries (hyperbolic and circular electrodes), rf frequency and ion

source parameters are all incorporated into the model. To simulate gas mixtures (multiple peak mass spectra) with >500 data points per atomic mass unit (amu), a 0–60 amu (m/z 0–60) mass spectrum would require computational times of several weeks on a Pentium 4 desktop PC. Instead a method of generating spectra using a characterisation set has been used. In this approach a number of mass spectral peaks were simulated across the required mass range so that the changes in peak shape trends with increasing mass could be identified and modelled.

QMS peak shapes are dependent upon the rf (V) and dc (U) voltage ratios (scan line) applied to the quadrupole electrode rods. Changes to the $U:V$ scan line affect the peak shape, width (resolution) and peak height (sensitivity), and these parameters may also vary with increasing mass. For each different $U:V$ scan line used in the simulation, (zone I and zone III), the changes in peak width, height and shape with increasing m/z value were recorded and the resulting values fitted with a polynomial expression in MATLAB. Other mass spectral peaks in the spectrum were then calculated by this polynomial expression if the peak height (signal intensity) was known. Such polynomial characterisation of the peaks in a mass range avoids the need to simulate all peaks in every individual gas spectrum required. This method is generic and can be used to form any mass spectra for any $U:V$ scan line.

The method is illustrated in Figs. 1 and 2. In Fig. 1 both the upper and lower graphs show zone I simulated peaks at m/z values of 40 and 60. The upper figure shows the data for a

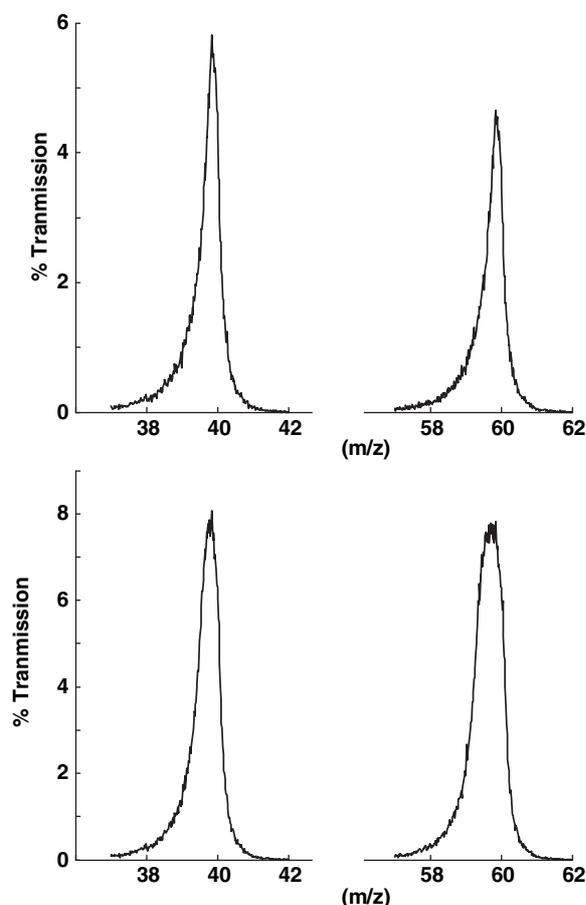


Figure 1. Simulated mass spectrum peaks at m/z 40 and 60 for two different zone I scan lines.

scan line $U = 0.16784V - 0.02$. In this case the two peaks are identical in shape, with a constant peak width, $\Delta M = 0.5 m/z$ units at 50% peak height for both peaks; however, there is resolution variation. The lower figure shows a scan line of $U = 0.166V$. In this case the resolution is constant but the peak width increases significantly with the m/z 40 and 60 peaks having ΔM values of 0.7 to 0.9 m/z units, respectively. Comparing the upper and lower graphs, the peaks in the lower figure show greater intensity. This is because the angle at which the scan line intersects the Mathieu stability region for a particular ion will determine the resultant peak height. The simulated peaks show some slight noise; this simply results from the repeated calculation of 10^5 ion trajectories at each point on the mass scale.

Mass spectral peaks do not always appear Gaussian in shape, and Fig. 2 shows an example of a skewed mass peak. In this case the peak shape may be further characterised by the ratio of the width of the low mass side to the upper mass side. The peak from the baseline to the tip of the peak is divided into 1000 equally spaced segments. For each segment the changes in width and shape are recorded. This is carried out for each of the peaks across the mass range to characterise these changes with increasing mass. The changes across the mass range are then fitted using a polynomial equation.

The simulation parameters are different for zone I (low-resolution) and zone III (high-resolution) simulations. The ion energies and scan lines used are shown in Table 1. The polynomial fits for width, shape and sensitivity can easily be combined with the calculated isotopic abundance data that contains all the mass positions and peak ratios. Analogue spectra are then generated for all the residual gas molecules required; this method applies to both low- and high-resolution data. From the generated spectra for each pure component required, the resulting spectra can be combined in any ratio to represent a mass spectrum of any given mixture of residual gases.

The resulting simulated spectra are free from ion source misalignment, detector drift and other experimental error contributions. The effects of these and baseline, interfering species, and nonlinearity can be incorporated into the

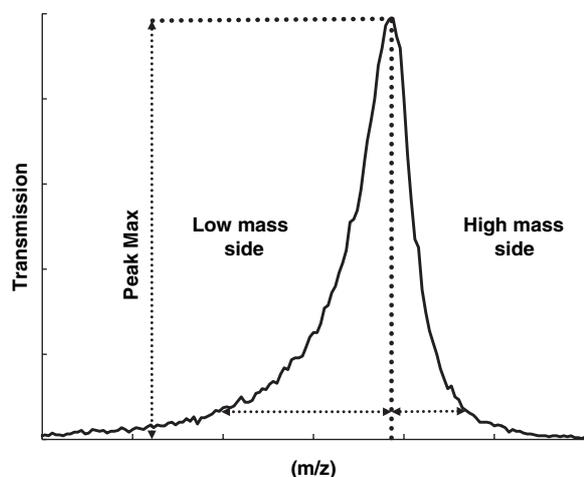


Figure 2. A mass peak with low mass side tailing and the information required to characterise changing peak trends with mass.

Table 1. QMS parameters used in the simulation of mass spectral peaks

Parameter	Settings	
Analyser length	150 mm	
RF	3.5 MHz	
r_0 (rod separation)	1 mm	
Exit plate radius	2.0 mm	
Electrode shape	Hyperbolic	
Source radius	0.2 mm	
Points per mass unit	750	
Number of ions	10^6	
	Low-resolution	High-resolution
Ions	6 eV	15 eV
Ion energy spread	0 eV	0 eV
$U:V$ ratio	$V/U = 5.958$	$V/U = 2.044$
Scan line	$U = 0.1678V$	$U = 0.4892V$

simulated spectra in a controlled manner so the multivariate calibration method adopted can then be fully assessed, for a given data set.

Generation of calibration and testing data sets

To generate the calibration data, concentrations of calibration mixtures (in X) can be fixed (to incorporate experimental design) or generated randomly. These corresponding concentration matrices are then used to generate the spectra calibration data. Here we have used two data sets: a three-component mixture (Table 2) and a ten-component mixture (Table 3). For the spectra to test the calibrations 250 target testing spectra were generated randomly for the three-component mixture and 1000 test spectra were generated for the ten-component mixture. Randomly generating all of X (ensuring that the sum of the concentrations of components summed to 100%) did not provide a suitable spread of high concentration values for the ten-component mixture. Therefore, for the ten-component mixture formed from the X matrix, the first two columns, X_1 and X_2 , are randomly generated but, since these are equal to or less than 100%, the remainders of each row X_{3-10} are generated to sum to $100 - (X_1 + X_2)$. This results in X_1 and X_2 values varying independently from those in columns X_{3-10} . The high values

Table 2. Composition of the three-component mixture of ethane, ethene and methane used to generate the calibration spectra

Spectrum no.	Molar % ethane	Molar % ethene	Molar % methane
1	50	50	0
2	50	0	50
3	70	15	15
4	15	70	15
5	15	15	70
6	100	0	0
7	100	0	0
8	0	0	100
9	33	33	33
10	24	45	31
11	0	100	0

Table 3. Composition of the ten-component mixture used to generate calibration spectra

	CO ₂	N ₂	O ₂	CO	CH ₄	Ar	C ₂ H ₆	C ₂ H ₄	C ₃ H ₆	H ₂ O
Spectrum no.	Molar % level concentration									
1	0	0.5	1	2	4.5	7	11	14	20	40
2	40	0	0.5	1	2	4.5	7	11	14	20
3	20	40	0	0.5	1	2	4.5	7	11	14
4	14	20	40	0	0.5	1	2	4.5	7	11
5	11	14	20	40	0	0.5	1	2	4.5	7
6	7	11	14	20	40	0	0.5	1	2	4.5
7	4.5	7	11	14	20	40	0	0.5	1	2
8	2	4.5	7	11	14	20	40	0	0.5	1
9	1	2	4.5	7	11	14	20	40	0	0.5
10	0.5	1	2	4.5	7	11	14	20	40	0
11	100	0	0	0	0	0	0	0	0	0
12	0	100	0	0	0	0	0	0	0	0
13	0	0	100	0	0	0	0	0	0	0
14	0	0	0	100	0	0	0	0	0	0
15	0	0	0	0	100	0	0	0	0	0
16	0	0	0	0	0	100	0	0	0	0
17	0	0	0	0	0	0	100	0	0	0
18	0	0	0	0	0	0	0	100	0	0
19	0	0	0	0	0	0	0	0	100	0
20	0	0	0	0	0	0	0	0	0	100

occur more frequently in the X matrix whereas previously the high values were suppressed.

The X matrix is then 'shuffled' (using the MATLAB function) in the row direction before preparing the test spectra, ensuring that the values are randomly sorted and that each component has an equal chance of being present at high or low concentration levels.

DISCUSSION

Comparison of a simulated spectrum with experiment

Figure 3 shows a comparison of a simulated spectrum for propane, obtained using the approach described above, with an experimentally recorded spectrum. The experimental spectrum was obtained by sampling from a fixed flow of propane, certified to 99.97% (supplied by Aldrich, Gillingham, UK). The spectrum was recorded using a commercial QMS system (Pfeiffer QS422 fitted with capillary inlet; Pfeiffer, Nashua, NH, USA). The figure also shows the mass spectrum of propane obtained from the NIST mass spectral library. NIST data is used due to the availability of the information for a large variety of molecules. The fragment ions of propane span a suitable mass range and all three spectra have been normalised so that the base peak signal intensity equals unity in each case. There is no difference between the NIST and simulated spectral intensities as the NIST data were used to provide the fragmentation ratios for the simulation. The experimental data show different fragmentation ratios from the NIST data. Furthermore, the NIST propane spectrum has peaks at mass 19 and 20, and it is unclear where these would have originated. Peaks at mass 19 and 20 are not evident in the experimental data and were therefore omitted in the simulated spectrum. The voltage scan lines used for the simulated data ensured that the peak intensities/sensitivity remained constant across the mass

range. The agreement between the simulated and experimental spectra in terms of peak shape, resolution and m/z position is good and this establishes the feasibility of the simulation model.

In the experimentally recorded spectrum shown in Fig. 3, water (m/z 16–18), oxygen (m/z 32), and argon (m/z 36 and 38) are all present and may be attributed to the background pressure in the vacuum system or to contaminants in the propane standard. Contributions for N_2^+ (m/z 28) and Ar^+ (m/z 40) will overlap with the propane peaks.

Figure 4 compares the simulation of low-resolution (zone I) and high-resolution (zone III) spectra using a four-component mixture of CO₂, CO, N₂ and propane, with each component present at 25%. The zone III higher resolution operation shows peak splitting at mass 28. This appears as a single broad peak in the low-resolution data. The components of the mass 28 peak arise from CO⁺ (m/z 27.99), C₂H₄⁺ (m/z 28.01), and N₂⁺ (m/z 28.03). Individual components are only distinguished using high-resolution spectra. Such features arising from zone III can be seen at mass 29.

The same four-component mixture at low resolution (zone I) for the 0–20 mass range with normalised transmission is shown in Fig. 5. The upper spectrum shows the resulting spectrum following addition of background and noise to the lower spectrum. The water background constituents are apparent at m/z 17 and 18. It is significant that the molecular ion of water is comparable in signal intensity to the fragment ions of the other analyte components, even though the background is added at only 10% of the total pressure (Fig. 5).

Calibration testing using simulated data

To test the different calibration models a three-component hydrocarbon mixture was considered (ethane, ethene and

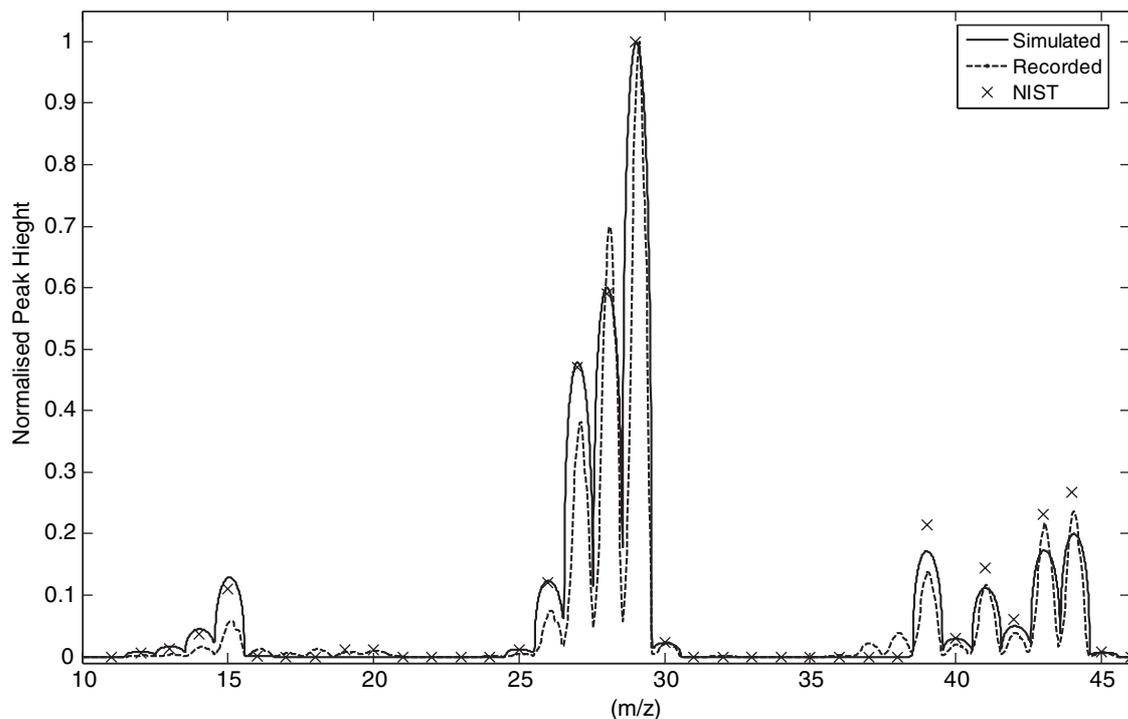


Figure 3. A simulated spectrum, experimentally recorded data and NIST library entry for propane.

methane). The spectra were all produced via the simulation process and then combined according to the concentrations shown in Table 2.

Calibrations were then performed on this data using the regression methods described earlier. Once the regression coefficients were obtained, the calibration models were tested on 250 randomly generated spectral mixtures. The errors between the predicted concentration values and the

actual concentration of the target values are expressed as RMSEP values, with lower RMSEP values indicating an improved calibration performance. The three-component mixture retained the first three factors in all instances when performing the factor-based calibrations (PCA and PLS). The RMSEP values were determined for error-free calibrations (i.e. no added noise or background) and gave a maximum value of 3.4×10^{-12} .

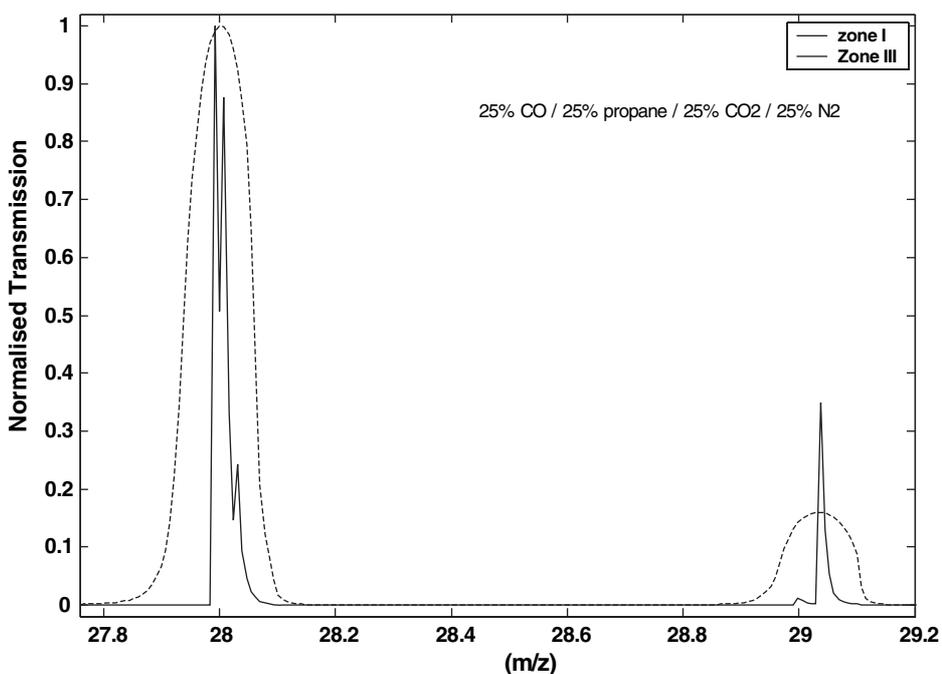


Figure 4. Comparison of the simulation of low- and high-resolution data for a four-component mixture of CO₂, CO, N₂ and propane, each component is present at 25%.

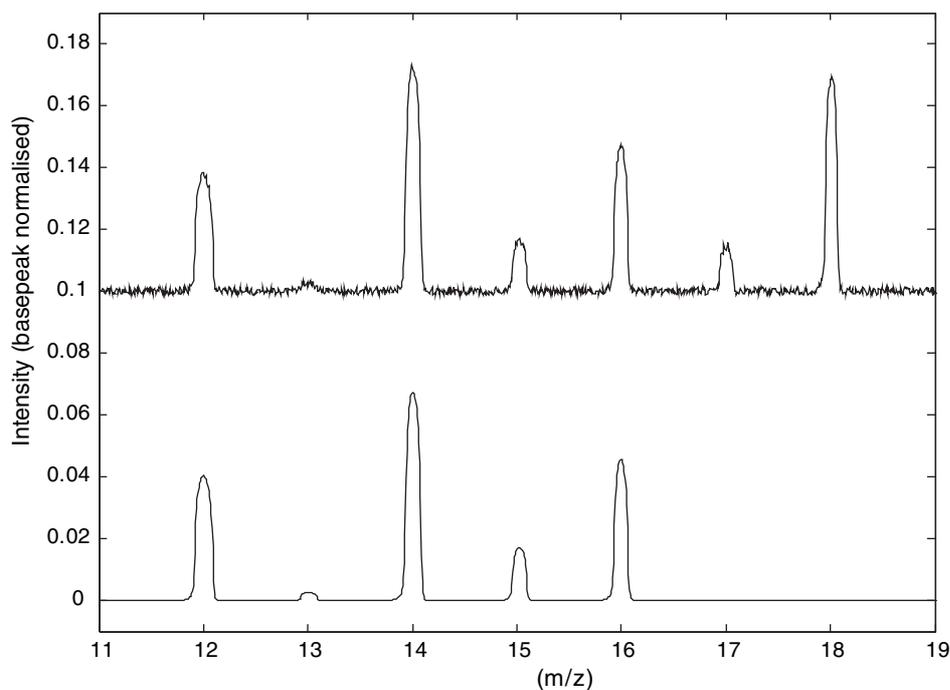


Figure 5. The lower trace shows simulated QMS spectrum. The upper trace shows the same spectrum with addition of noise and background constituents; trace is offset in *y* for clarity.

Table 4. Summary of noise effects on a ten-component calibration using low-resolution mass spectra

Data points (per <i>m/z</i>)	Method	RMSEP				
		1	16	32	64	128
1% noise	CLS	1.9814	1.0982	0.7927	0.54593	0.67990
	PLS	1.9994	1.1104	0.7950	0.54713	0.68055
	PCR	1.9785	1.0973	0.7926	0.54578	0.67981
	OLS	2.1058	1.2424	0.8672	0.62157	0.75062
0.1% noise	CLS	0.2019	0.0589	0.0408	0.0280	0.0206
	PLS	0.2021	0.0589	0.0409	0.0280	0.0205
	PCR	0.2019	0.0589	0.0408	0.0280	0.0205
	OLS	0.2123	0.0606	0.0424	0.0286	0.0207

Effects of sampling rate and noise on calibration

The effects of noise on RMSEP values for the ten-component calibrations are shown in Table 4. The sampling rates were chosen to match those available on commercially available instruments, typically a maximum of 64 data points per mass unit. From Table 4 it can be seen that the general trend from the results is that increasing the sampling rate has reduced the effects of noise on the calibration at both the 0.1% and 1% noise levels. Increasing the noise levels adversely affects the calibration. For a 1% noise level 128 data points gave no calibration improvement over 64 data points; the noise is added randomly so variations in RMSEP results can occur. Noise in the calibration data is more significant than in the test data and the effect of noise may be reduced by averaging multiple scans before determining the regression coefficients.

Background effects

The background spectra are measured spectra with no gas components added to the vacuum system. It represents the

base pressure (lowest obtainable pressure) of the vacuum system. The background can be constant or drift and is also subject to composition changes, e.g. experimentally background levels of water vary over time. Traditionally, the background pressure spectrum is subtracted from the measured sample spectra; however, this is problematic if there is drift or change in its composition. Levels drift severely following instrument start-up. Simulated background spectra were generated according to the compositions shown in Table 5.

The background spectra were added to calibration spectra and target test spectra at a chosen percentage level. The background spectra can be scaled to represent changes in base pressure; the additional background (2) (Table 5) was

Table 5. Composition of components in the background spectra

Background	N ₂	O ₂	CO ₂	Ar	H ₂ O
(1)	43.3	11.1	0.6	0.6	44.4
(2)	70.9	18.2	0.9	0.9	9.1

Table 6. Effects of background on average RMSEP values for low- and high-resolution QMS data for the three-component mixture (ethane, ethene, methane)

	Points	CLS	PLS	PCR	OLS
Changing background composition at a nondrifting base pressure (5%)					
Low-resolution	64	1.2530	1.253	1.1253	1.1273
	128	1.2363	1.2363	1.2363	1.1133
High-resolution	64	0.0956	0.0956	0.0956	0.0945
	128	0.1001	0.1001	0.1001	0.0863
Changing background composition at a drifting base pressure (20 to 5%)					
Low-resolution	64	5.1505	5.1505	5.1505	4.7416
	128	5.0917	5.0917	5.0917	4.8340
High-resolution	64	1.006	1.006	1.006	0.7271
	128	3.1064	3.1064	3.1064	2.1610

required to represent a change in the composition of the background. Background (1) has a higher level of water, generally seen in instruments shortly after start-up. Experimentally, these water levels are reduced by baking the vacuum system at elevated temperatures and by continued operation.

Constant background at constant composition showed no effect on the calibration methods. This is true when the background was present at 20% or 5% and all RMSEP errors were below 10^{-12} . The results for varying background pressure and composition are shown in Table 6 for the three-component mixture. Varying background adversely affects these calibrations. It should be noted that the higher resolution zone III operation is less sensitive to the background variations. This is due to the increased ability to differentiate between background species and the analyte gases when operating at high resolution.

An alternative to background subtraction is that the background components are incorporated into the calibration matrices **X** and **Y**. This was the case for the ten-component mixture (Table 3); constant pressures at constant composition produced RMSEP values below 10^{-12} . All calibration methods produced identical results for varying background composition and at constant 5% level the RMSEP was 0.7177 for all methods at 1, 64 and 128 data points per mass unit. For varying composition and varying pressure (20 to 5%) again all the results were identical for all sampling rates with a RMSEP of 2.6647.

With the ten-component mixture, despite increasing the number of gas components, the RMSEP values are reduced by incorporating the atmospheric components into the calibration. If the background composition and pressure remain constant the atmospheric components may then be omitted from the calibration of the mass spectrometer.

CONCLUSIONS

A convenient method for simulating full analogue quadrupole mass spectra over a given mass range has been implemented. Generation of such simulated multiple peak mass spectra from mixtures of residual gases is shown to be possible on a desktop PC. Time-consuming full mass range simulations are avoided by characterisation using a smaller number of representative peaks across the mass range and

polynomial fitting to characterise peak width, height and shape. Good agreement with an experimentally obtained spectrum (propane) is shown. The method has also been used to generate high-resolution spectra for zone III of the Mathieu stability diagrams. The ability to generate such mass spectra allows the testing of calibration procedures in experimental systems.

The simulated spectra were then used to evaluate the effects of background species. Upon calibration, nonchanging backgrounds (i.e. gas composition and pressure constant) were found to have virtually no effect on the calibration methods used for low- (zone I) or high-resolution (zone III) mass spectra. In practice these background variations will occur and must be considered when choosing a calibration model for an experimental procedure. While background and noise contributions have been investigated there is further scope for studies of noise variation with mass (or increasing rf voltage) and nonlinear pressure effects.

The multivariate calibration methods CLS, PLS, OLS and PCR were evaluated using the simulated spectra with the different errors incorporated. Factor-based methods (PCR and PLS) are generally regarded as having improved performance over the simpler least squares methods. Here, however, the multivariate regression methods were virtually indistinguishable with regard to background effects and randomly added noise, OLS performing marginally worse than the other methods.

An important conclusion of this study is the enhanced performance of calibrations when high-resolution zone III data are employed. Such high-resolution spectra perform better for nondrifting background pressure and also when both the composition and the pressure of the background vary.

REFERENCES

1. Dawson PH. *Quadrupole Mass Spectrometry and its Applications*. Elsevier: Amsterdam; 1976.
2. Gibson JR, Taylor S, Leck JH. *J. Vac. Sci. Technol. A* 2000; **18**: 237.
3. Gibson JR, Taylor S. *Rapid Commun. Mass Spectrom.* 2003; **17**: 1051.
4. Gibson JR, Taylor S. *Rapid Commun. Mass Spectrom.* 2001; **15**: 1960.
5. Ying JF, Douglas DJ. *Rapid Commun. Mass Spectrom.* 1996; **10**: 649.
6. Turner P, Taylor S, Gibson JR. *J. Vac. Sci. Technol. A* 2005; **23**: 480.

7. Du ZH, Douglas DJ, Glebova T, Konenkov NV. *Int. J. Mass Spectrom. Ion Processes* 2000; **197**: 113.
8. Dawson PH, Yu BQ. *Int. J. Mass Spectrom. Ion Processes* 1984; **56**: 25.
9. Chen W, Collings BA, Douglas DJ. *Anal. Chem.* 2000; **72**: 540.
10. Konenkov NV, Kratenko VI. *Int. J. Mass Spectrom. Ion Processes* 1991; **108**: 115.
11. Kaneko K, Hiroki S, Abe T, Murakami Y. *Vacuum* 1996; **47**: 1313.
12. Lieszkovszky, Filippelli AR, Tilford C. *J. Vac. Sci. Technol. A* 1990; **8**: 3840.
13. Basford JA, Boeckmann MD, Ellefson RE, Filippelli AR, Holkeboer DH, Lieszkovszky L, Stupak CM. *J. Vac. Sci. Technol. A* 1993; **11**: 22.
14. Tunnichiff DD, Wadsworth PA. *Anal. Chem.* 1965; **37**: 1082.
15. Malinowski ED. *Factor Analysis in Chemistry* (3rd edn). John Wiley: New York, 2002.
16. Wentzell PD, Montoto LV. *Chemom. Intell. Lab. Sys.* 2003; **65**: 257.
17. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*. Springer: New York, 2001.
18. Bhupinder, Dayal S, MacGregor JF. *J. Chemometrics* 1997; **11**: 73.
19. MATLAB programme. Available: <http://www.mathworks.co.uk/> (accessed November 2006).
20. Rockwood AL, Van Orman JR, Dearden DV. *J. Am. Soc. Mass Spectrom.* 2004; **15**: 12.
21. Yergey JA. *Int. J. Mass Spectrom. Ion Phys.* 1983; **52**: 337.
22. Stein SE. *Mass Spectra (NIST Mass Spec Data Center) NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, Linstrom PJ, Mallard WG (eds). 2005. Available <http://webbook.nist.gov/>