

Univariate and multivariate DMS calibration for a single analyte

Andrzej Szczurek^{1,*}, *Monika Maciejewska*², and *Żaneta Zajiczek*¹

¹Faculty of Environmental Engineering, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

Abstract. Differential mobility spectrometry (DMS) is a promising measurement technique. It is used in the detection of chemical warfare agents, explosives, drugs, and volatile organic compounds. The measurement principle is based on separation of gas-phase ions according to their differential mobility in alternating low and high electric fields. The DMS measurement result is a two dimensional spectrum of ion current displayed as a function of separation voltage and compensation voltage. The DMS spectral peaks, in terms of their height, location and width, are affected by gas sample composition, separation field and the gas flow rate. In this work, there is presented the calibration procedure which utilises the univariate and multivariate approach to differential ion mobility spectrum. We demonstrated the possibility of a successful retrieval of quantitative information using partial least squares regression as well as univariate linear regression. However, the multivariate approach outperformed the univariate one in terms of the quality of the model and the concentration prediction accuracy.

1 Introduction

Differential mobility spectrometry (DMS) is one of the most promising measurement techniques [1, 2]. A lot of experimental data concerning a rapidly growing number and variety of its applications had been published and discussed. DMS is a powerful tool widely used in the detection of chemical warfare agents, explosives, drugs, and volatile organic compounds [3–5]. The measurement principle of this technique is based on separation and characterization of gas-phase ions [1]. When ions move through a gas under the influence of an electric field, the combination of acceleration due to the field and deceleration due to collisions with gas molecules affects their movement. DMS separates ions based on their differential mobility in alternating low and high electric fields. The differential mobility of an ion depends on a number of properties, including the mass, shape, centre of mass, dipole moment, effects of clustering between ions and neutral gas molecules during separation.

In DMS, ions are carried by a gas between parallel or coaxial electrodes [1, 2]. The electric field alternates perpendicularly to the DMS channel. Ions oscillate between the two electrodes and may either exit the device or collide with an electrode depending on their differential mobility in high and low electric fields. Particular ions may be retained in

* Corresponding author: monika.maciejewska@pwr.edu.pl

equilibrium, inside the gap by applying a low strength constant compensation voltage (CV) to the electrodes. It corrects the migration of the ion caused by the differential mobility so that the ion can pass between the electrodes and be detected. Ions with differing drift velocities, due to their properties, can be passed through the gap at different compensation voltages. Scanning the CV produces a spectrum of the ion mixture.

The information about chemical species is included in a signal which is an ion current measured at the output of the filter gap. This current is a product of averaged ion density and carrier gas flow rate. Instrumental parameters and ion properties affect the ion density and hence the output current.

The DMS provides information about the analyzed gas in a spectral form. Differential mobility spectrum is defined as the current at the detection electrode with respect to the compensation voltage. The spectrum consists of a series of peaks, which are affected by instrumental parameters such as carrier gas flow rate, separating field amplitude and waveform, filter gap size and geometry, ion properties such as diffusion, mobility. In practice, the only parameter that is changed with time during a single analytical run is the compensation voltage. Other parameters are optimized and fixed beforehand. A spectral peak can be characterized by its position, height and width. These characteristics depend on many factors. The DMS peak height is affected by the analyte concentration, separation field and the gas flow rate. The DMS peak width (full width at half maximum) appears to be strongly dependent on the gas flow rate for the planar DMS and on the separation voltage amplitude for the coaxial DMS. The peak position depends on the concentration of an analyte. Hence, the calibration of a differential mobility spectrometer based on peak's characteristics is complex operation and it has to be done in a multidimensional space. In this work, there is presented the calibration procedure based on univariate and multivariate approach to differential mobility spectrum.

2 Experimental part

This work was based on laboratory experiments. They consisted in measurements of reference gas mixtures using differential ion mobility spectrometry. Reference gas mixtures were composed of clean air and the predefined amount of acetic acid. The examined concentration range was from 2.1 ppm to 41.7 ppm.

Measurements were realised with the dedicated setup, as shown in Fig. 1. The setup was composed of the module for preparation of clean air, module for preparation of reference gas mixtures and the measuring instrument, DMS spectrometer.

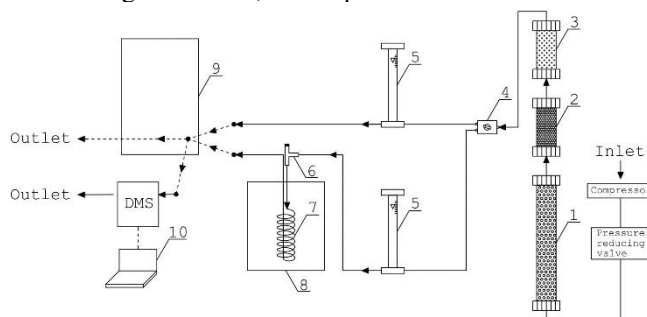


Fig. 1. Measurement setup: 1. filter (silica gel), 2. filter (soda lime and molecular sieves), 3. filter (activated charcoal), 4. switch, 5. rotameter; 6. analyte injection point, 7. coil, 8. thermostat, 9. Tedlar bag, 10. computer.

In the measurement setup, clean air was used as the carrier gas and for removing the analyte from the set-up, once the measurement was completed. The clean air was attained from indoor air passed through a set of filters. They were filled with silica gel, soda lime, molecular sieves and activated charcoal, see Fig. 1, which allowed for air drying and removal of volatile organic compounds. The air flow was enforced by the compressor. The pressure was adjusted with the pressure reducing valve.

Reference gas mixtures were prepared using an evaporation method. The module dedicated for preparation included: thermostat, steel coil with an inlet port for the liquid analyte, syringe, Tedlar bag, rotameters and gas tubing, see Fig. 1. Clean air was delivered to this module and the flow rate was controlled using rotameters. For preparing reference gas mixture, the determined amount of liquid analyte was injected, with a syringe, through the injection port to the air stream. The analyte was vaporised in the air inside the coil, which was maintained at the constant temperature. The air containing the analyte vapours was collected in the Tedlar bag. The concentration of reference gas mixture was determined based on the amount of the injected analyte, gas flow rate and the time of gas collection in the Tedlar bag. Before preparation of reference gas mixture and after the measurement, Tedlar bag was cleaned by filling it with clean air and emptying it, three times.

During measurement, the reference gas mixture from the Tedlar bag was directed to the DMS spectrometer and the ion mobility spectrum was recorded. Five measurements were done for each reference gas mixture. A single measurement lasted 30 s. During measurement, gas flow rate through the instrument was $0.5 \text{ dm}^3/\text{min}$.

The experiments consisted in measurements of reference gas mixtures composed of clean air and reference gas mixtures composed of clean air and acetic acid. The following concentrations of analyte were considered: 2.1, 4.2, 5.2, 10.4, 20.8 and 41.7 ppm. The measurement of each reference gas mixture which contained the analyte was preceded by the measurement of reference gas mixture which contained clean air. The humidity of the clean air was maintained at the level of $11\% \text{ RH} \pm 2\% \text{ RH}$ (at 323K) throughout all experiment. All reference gas mixtures were prepared using the same Tedlar bag.

The DMS spectrometer applied in experiments was constructed in the Military Institute of Chemistry and Radiometry, Poland. [3]. Gas ionization in this device is achieved by the β -emitter, nickel electrode covered with radioactive nickel ^{63}Ni . The DMS chamber has planar geometry. The drift region is formed by two electrode-plates ($5 \times 25 \text{ mm}$) separated by a 0.5 mm gap. A high amplitude, high frequency, asymmetric voltage waveform, generated by HSV generator (2 MHz) is applied to electrodes, crosswise to the direction of the carrier gas flow. The separation voltage (SV) amplitude is from 100 to 1600 V (peak to peak), which corresponds to from 6,5 to 104 Td. The compensation voltage is scanned from -30 to 8 V. At the end of the drift zone there are located two detector electrodes ($5 \times 5 \text{ mm}$), for collecting positive and negative ions respectively. Two types of ions are neutralized in parallel. The internal gas flow through the detector is 2.0 L/min . The chamber entrance temperature is 318 K and chamber temperature is 323 K.

3 Methods

The result of measurement with DMS spectrometer is a two-dimensional spectrum of ion current. Actually, two spectra are attained during a measurement, one for negative ions and one for positive ions. The ion current is recorded as a function of the separation voltage and the compensation voltage. The two voltages were discretized as follows $\text{SV}=\{100, 150, \dots, 1600\}$ and $\text{CV}=\{-30, -29.926, \dots, 7.888\}$, which resulted in a finite and countable set of combinations of (SV, CV). The ion current is recorded for each of these combinations.

Two approaches, univariate and multivariate were compared in respect of extracting the quantitative information about the analyte from DMS spectrum. The univariate approach

consists in using one independent variable as the predictor of the response variable. In this work, it was represented by simple linear regression. Simple linear regression is used for determining the calibration function in traditional analytical solutions. The multivariate quantitative approach consists in using more than one independent variable as the predictor of the response variable(s). In this work, the multivariate approach was represented by partial least squares (PLS) regression. This technique is willingly used in spectrometric calibration [6], where multivariate measurement data is obtained.

The applied univariate regression model was [7]:

$$c_{AA} = a_0 + a_1 I_{(SV,CV)} + \varepsilon \quad (1)$$

where $I_{(SV,CV)}$ is the ion current at the collection electrode (associated with the particular combination of separation voltage and compensation voltage) recorded during the measurement of reference gas mixture, containing acetic acid at the concentration c_{AA} . The model coefficients are a_0 and a_1 . The error of data fitting with the model is ε .

Partial least squares regression model consists of two equations [6]:

$$X = TP^T + E \quad (2)$$

$$Y = TQ^T + F \quad (3)$$

where X is the input data matrix (each column refers to one independent variable) and Y is the output data matrix (each column refers to one response variable). The score matrix T consists of latent variables. They are artificial variables, derived under the constraint of explaining as much as possible of the covariance between X and Y . P is the loading matrix which allows to transform loadings into independent variables. R is the loading matrix which allows to transform loadings into the dependent variables. E and F are error matrices.

We proposed that the independent variables in PLS model are ion currents associated with all compensation voltages CV and the defined separation voltage. While univariate model is based on a single element of two dimensional DMS spectrum, the multivariate model explores the one dimensional DMS spectrum.

The quality of the linear univariate models and multivariate models was evaluated using coefficient of determination (R^2). The accuracy of analyte concentration prediction with the models was determined using root mean square error (RMSE).

4 Results and discussion

The presented results are based on negative ions mobility spectra. Based on our analysis, the positive ions mobility spectra did not convey the useful information.

An important aspect of measurement is precision. The precision of DMS spectra was examined. The precision was determined as the spread of five spectra, obtained from the series of five measurements of one particular reference gas mixture. The precision was represented by the standard deviation and the coefficient of variation. As shown in Fig. 2, for all measurements the coefficient of variation was less than 3%, which indicates high precision of DMS spectra. Considering coefficient of variation, measurements of clean air were less precise than the measurements of reference gas mixtures containing analyte, while standard deviation indicated similar precision. This discrepancy could be an effect of grater ion currents during measurements of reference gas mixtures containing the analyte. Based on Fig. 2, DMS spectra for small concentrations were more precise compared with spectra for high concentrations. Clean air spectra were nearly equally precise. High precision of the readout is an important parameter of the measurements method which may be used for the quantitative assessment.

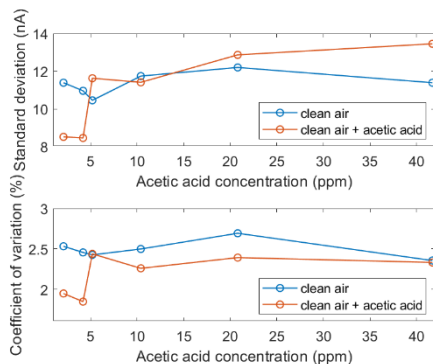


Fig. 2. Precision of DMS spectra in terms of standard deviation and coefficient of variation.

Precision was determined for spectra of reference gas mixtures containing clean air and acetic acid in clean air. Measurements of the first mixtures preceded the measurements of the second mixtures, that is why the precisions are shown in pairs.

Based on theory of ion mobility [1, 2], the analyte ion peak position and height changes depending on gas sample composition. The change of peak height is by all means valuable, as a source of information about the analyte concentration. However, the change of peak location is a difficulty which needs to be overcome while developing a calibration function. The frequently encountered cause of peak shifts is sample humidity variation. The problem, has been referred on multiple occasions [4, 5, 8]. In the current study we indicate the problem of analyte peak location (not only height) change as a consequence of the variation of analyte concentration in reference gas mixture.

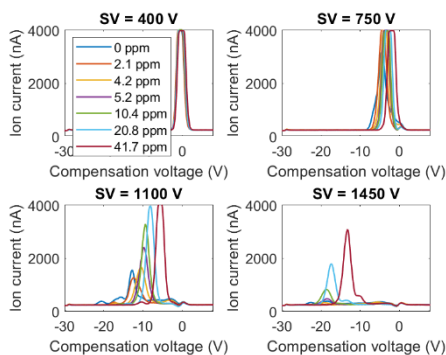


Fig. 3. Unidimensional DMS spectra for reference gas mixtures composed of clean air and acetic acid. Ion current recorded at fixed separation voltage, was shown as a function of the compensation voltage. The negative ions spectra are presented.

The overlap of peaks associated with different analyte concentrations is crucial when developing analyte quantification methods based on ion current associated with an individual combination of (SV,CV). For this approach, it would be ideal to have all concentrations peaks in the same place, and their amplitudes proportional to the analyte concentration. Although the method is still operational when all peaks partially overlap, its performance gradually decreases with the decreasing overlap. The magnitude of peaks separation is associated with the overall range of analyte concentration, which may be considered as the limiting factor for the univariate calibration method. In case of small concentration ranges, the overlap of extreme peaks is sufficient to achieve analyte quantification based on ion current associated with an individual combination of (SV,CV).

When broader concentration ranges are involved, the method becomes less effective. In such cases other multivariate spectrum analysis techniques have to be applied, which are capable to extract the quantitative information from the spectrum in spite of concentration dependent peak allocation [9].

The problem of analyte peak position change, related to analyte concentration, was illustrated in Fig. 3. We displayed one dimensional DMS spectra of reference gas mixtures containing various amounts of analyte. As shown, peaks associated with different concentrations of the same analyte have different positions in the spectrum. Additionally, the relative positions of ion current peaks depends on the magnitude of separation voltage. The degree of overlap between peaks associated with various concentrations of analyte is a function of both, compensation and separation voltages.

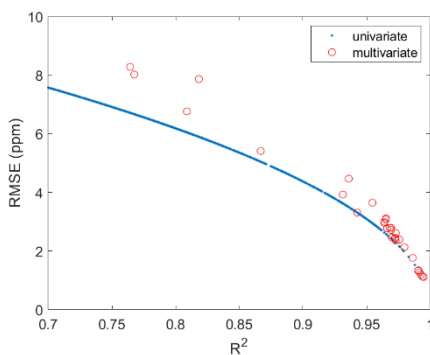


Fig. 4. Coefficient of determination (R^2) for univariate linear regression models and multivariate partial least squares regression models versus root mean squared errors (RMSE) of acetic acid concentration prediction with these models.

In Fig. 4, the model quality measure R^2 and concentration prediction error (RMSE) are shown. They characterise the developed univariate models (blue dots) and multivariate models (red circles). Based on the comparison, greatest values of coefficient of determination and smallest error values were attained in case of multivariate models.

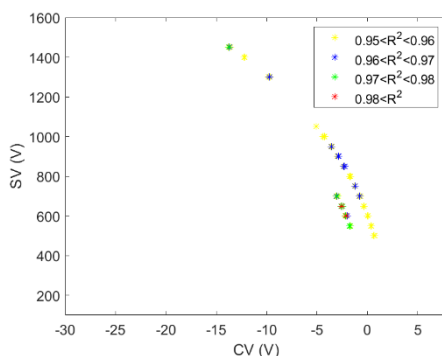


Fig. 5. Combinations of compensation voltage (CV) and separation voltage (SV). The ion current recorded at these combinations was a valuable source of information about acetic acid concentration in reference gas mixtures when using univariate linear regression.

The performance of univariate models was analysed in relation to spectral coordinates of the independent variable. The best performing univariate model was featured by $R^2 = 0.988$ and $RMSE = 1.48$ ppm. The model was associated with $SV = 650$ V and $CV = -2.5391$ V. In Fig. 5 we show (SV, CV) coordinates of independent variables of other, relatively well

performing univariate models, based on the cut off criterion, $R^2 = 0.95$. They indicate parts of spectrum which are useful for quantitative determination of analyte, using univariate models. As shown in Fig. 5 the distribution of valuable combinations (SV, CV) in two dimensional spectrum is not random. However, it would be difficult if not impossible to anticipate their location based on the visual inspection of DMS spectra, without the appropriate data analysis.

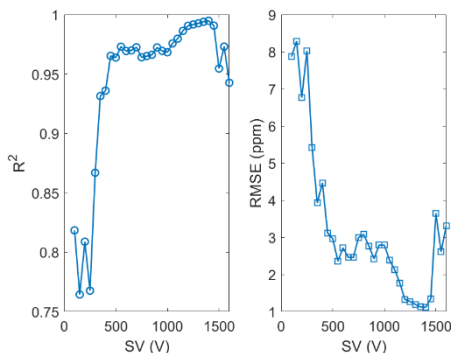


Fig. 6. Coefficient of determination (R^2) for PLS regression models and root mean squared errors (RMSE) of acetic acid concentration prediction with these models. Parameters are displayed versus separation voltages which were the source of the independent variables for the model.

The performance of the multivariate models was examined in the domain of separation voltage. As shown in Fig. 6, the indicator of quantification model performance (R^2) and the indicator of concentration prediction performance (RMSE) were dependent on the separation voltage. For small separation voltages, the model performance was poor and concentration prediction errors were high. Both parameters improved with the increase of SV. Starting from $SV=1200$ V, coefficient of determination was greater than $R^2=0.95$ and for separation voltage in a range between $SV = 1200$ V and $SV = 1450$ V, coefficient of determination was greater than $R^2 = 0.99$. The best results ($R^2 = 0.995$ and $RMSE = 1.1$ ppm) were attained for $SV = 1400$. Interestingly, for separation voltages exceeding $SV = 1450$ V model quality and prediction performance got worse.

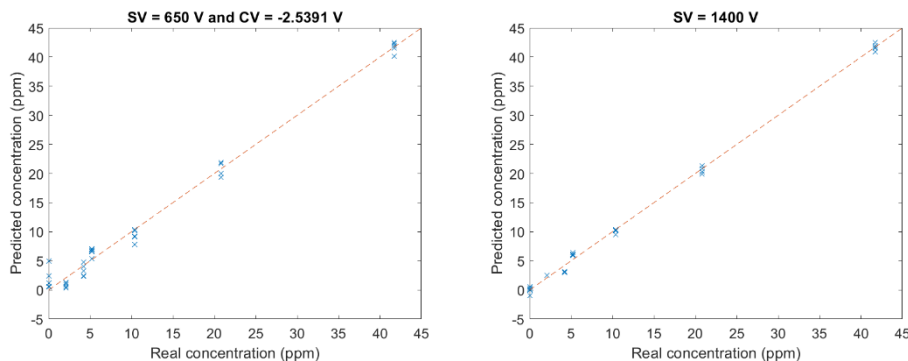


Fig. 7. Left: Diagnostic plot for the best univariate model. The independent variable was ion current recorded at the combination of compensation and separation voltage $SV = 650$ V and $CV = -2.5391$ V. Right: Diagnostic plot for the best multivariate model. The vector of independent variables consisted of ion currents recorded all compensation voltages while the separation voltage was $SV = 1400$ V.

The diagnostic plots for the best performing univariate model and the best performing multivariate model are compared in Fig. 7. The visual inspection reveals the supremacy of the multivariate model, which agrees with the evaluation based on the objective indicators R^2 and RMSE. It shall be admitted that difference between the level of indicators for best univariate and best multivariate model is not striking. However, the solution offered by multivariate approach is more reliable. There were several excellent PLS models and just one very good univariate model. It is also important that better results of multivariate model were particularly noticeable at small concentrations, which are most difficult to determine in general.

5 Conclusions

In this work we focussed on the DMS spectrum as a source of quantitative information about the analyte. There was compared the potential of the univariate and the multivariate approach to spectrum. The first utilised one explanatory variable – the ion current associated with a single combination of separation voltage and compensation voltage. The second extracted the information from multiple explanatory variables – ion currents associated with all compensation voltages and one separation voltage.

We demonstrated the possibility of a successful retrieval of quantitative information using PLS regression as well as univariate linear regression. However, the multivariate approach outperformed the univariate approach in terms of the quality of the model and the concentration prediction accuracy. The supremacy was visible at small concentrations, which are generally most difficult to determine.

The analysis allowed to identify regions of spectrum which are useful regarding quantitative determination of the analyte. The part of spectrum associated with separation voltage less than $SV=500$ V was useless for this purpose. Best performing univariate models utilised ion current recorded at smaller separation voltages than best performing multivariate models.

References

1. H. Borsdorf, G. A. Eiceman, *Appl. Spectrosc. Rev.* **41**, 323–375 (2006)
2. B. B. Schneider, E. G. Nazarov, F. Londry, P. Vouros, T. R. Covey, *Mass Spectrom. Rev.* **9999**, 1–51 (2015)
3. M. Maziejuk J. Puton M. Szyposzyńska, Z. Witkiewicz, *Talanta* **144**, 1201–1206 (2015)
4. T. N. Obee, M. Piech, J. V. Manteste, S. Dardona, *Int. J. Ion Mobilit. Spec.* **15**, 131–139 (2012)
5. N. Krylova, E. Krylov, G. A. Eiceman, J. A. Stone, *J Phys Chem A.* **107**, 19, 3648–54 (2003)
6. T. Naes, T. Isaksson, T. Fearn, T. Davies, *A User-friendly Guide to Multivariate Calibration and Classification*, NIR Publications, Chichester, UK (2004)
7. A. Szczurek, M. Maciejewska, Ż. Zajiczek, M. Maziejuk, *ASEE17, E3S Web of Conferences* **22**, 00173 (2017)
8. A. Szczurek, M. Maziejuk, M. Maciejewska, T. Pietrucha, T. Sikora, *Sens Actuators B Chem.* **240**, 1237–1244 (2017)
9. M. Maziejuk, A. Szczurek, M. Maciejewska, T. Pietrucha, M. Szyposzyńska, *Talanta* **152**, 137–146 (2016)