UNIVERSITAT DE BARCELONA

Multivariate Signal Processing for Quantitative and Qualitative Analysis of Ion Mobility Spectrometry data, applied to Biomedical Applications and Food Related Applications

Ana Verónica Guamán Novillo

FACULTAT DE FÍSICA

Departament d'Electrònica

MEMÒRIA PER OPTAR AL TÍTOL DE DOCTOR PER LA UNIVERSITAT DE BARCELONA

Doctorat en Enginyeria i Tecnologies Avançades (RD 99/2011)

Multivariate Signal Processing for Quantitative and Qualitative Analysis of Ion Mobility Spectrometry data, applied to Biomedical Applications and Food Related Applications

by

Ana Verónica Guamán Novillo

Director:

Dr. Antonio Pardo

Codirector:

Dr. Josep Samitier

Tutor:

Dr. Antonio Pardo

CHAPTER FIVE
Quantitative Analysis of IMS datasets

5.1. Introduction

Quantitative Analysis is a main part in analytical chemistry, since it is linked with high variety applications and instrumental performances. In view of volatile organic compounds (VOCs) are present at very low concentration usually in a range of ppm or even ppt -specially in biological applications, it is important to determine limit of detection and quantification of any kind of instruments. Therefore, quantitative analysis allows establishing instrumental limitations under different conditions such as temperature, humidity and surrounding parameter that might affect the sample. Furthermore, the background present in samples are usually composed by complex chemical matrix could interfere with the analyte of interest. Thus means that is necessary to perform a proper analysis in order to quantify the information in the best manner.

In this context, analytical devices such as GC/MS or electronic noses have examined in detail how overcome instrumental and sampling limitations. This understanding also covers the understanding of the signal processing strategies for solve quantitative issues. However, quantification on IMS is usually based on univariate analysis given results either overoptimistic or pessimistic. Therefore, development of multivariate strategies is required in order to make progress in IMS applicability and turn the limitations into a common process when IMS instruments are being used.

Current quantitative application of IMS is mainly focused on measurements at different ranges of concentrations in applications such as detection of explosives, illicit drugs, toxic chemicals, etc. Moreover, these studies have scarcely explored signal processing strategies and their analysis have restricted only in performing univariate calibration and univariate limit of detection estimation. Since, there is an increasing interest in the use of IMS- especially in biorelated fields, it is important to introduce new strategies and methodologies of signal processing to get all the profit of the datasets.

The content of this chapter goes from common univariate analysis to a deeper exploration of usefulness of multivariate calibration. This analysis was performed in order to determine a better and accurately quantifications in IMS. The first part consist in an exploration of univariate and multivariate calibration techniques is synthetic dataset, then two biorelated applications are studied from a quantitative perspective in which the limit of detection and quantification are calculated.

5.2. From Univariate to Multivariate Calibration in IMS using synthetic data set.

IMS performance and its non-linear behaviour and the effect of charge competition has been discussed in chapter three. These issues have a huge effect on the IMS measurements, and quantitative analysis of IMS datasets must face with them. Nonetheless, the most common method for the IMS calibration purposes is the use of univariate techniques (Zamora et al., 2011, Karpas et al., 2002b) in which the area and/or height of the peak or peaks of interest is taken for building a calibration curve.

Figure 5.1 depicts an IMS data where univariate calibration is fully successful. In this case a dataset at different concentrations (0.1 to 1.2 ppm) of Acetone were obtained. In this case, the spectra showed in Figure 5.1 depict four clearly separated peaks – two of them are reactant ion peaks and the others are linked to acetone. Since it is known the monomer of acetone appears at 8.9 ms or 1.80 $cm^2V^{-1}sec^{-1}$(the first peak at the left), height of this peak is used to build a univariate model calibration and the small peak is discarded for the analysis. In addition to that, an extra measurement at concentration 0.8 ppm was measured so that the univariate model can be tested.

The result, which is shown in Figure 5.1 (b), is quite good and the prediction concentration was 0.9 ppm with a RMSEP of 0.1ppm. In this particular case, the monomer of the acetone increases as the concentration increase and its location is well known, thus make possible the use of univariate technique. Note the RIP do not disappear during the whole experiment. In addition the intensity of the RIP decreases as soon as the concentration of acetone increases.

However, this is an easy experiment, under well controlled conditions and low level of noise. Moreover, the small peak linked to acetone (11 ms) is not consider in the calibration process, thus some information is not taking into account in the model.

In a real scenario, real sample is composed by one or more unknown compounds that would overlap with acetone. Therefore, the model will be blind to this information thereby the model will not be enough reliable. The univariate calibration main disadvantage is that the model just took into account a single peak and not the influence of other substances or other peaks of the whole spectra from the sample.

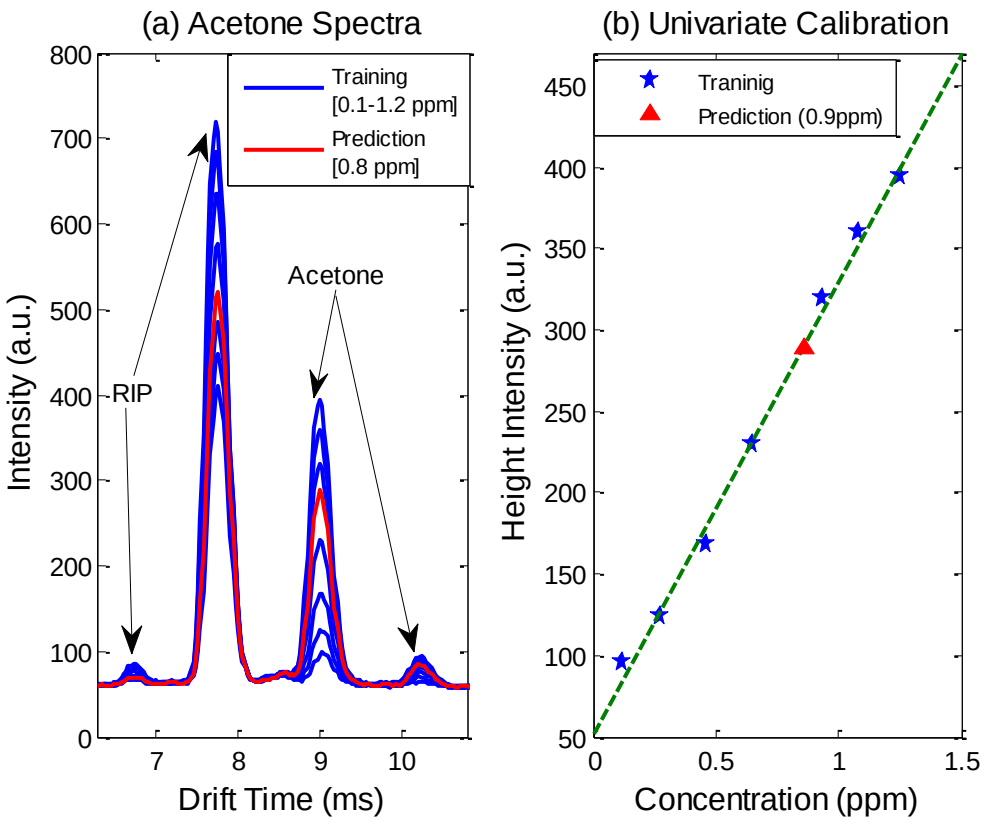


Figure 5.1 Example of univariate Calibration. (a) Raw spectra of Acetone (B) Univariate calibration obtained using peak height of acetone

In order to clarify the difficulties present in some IMS analysis, pure ethanol was measured with Ni-IMS (GDA2 Airsense (Airsense, 2012)). In principle, it is supposed to have a simple spectrum in similar way than acetone, where a model can be built using the information from a single peak. However, the ethanol presents a highger non-linear behavior than acetone in IMS. Figure 5.2 shows a set of spectrum measured with Ni-IMS at different concentration from 0.39 to 5.33 ppm of ethanol (see chapter five). In order to validate the model, three different concentrations were measurement in the same experiment.

The spectra (Figure 5.2 (a))present four peaks: two of them -with a mobility coefficient of 2.35 $cm^2V^{-1}sec^{-1}$ and 2.09 $cm^2V^{-1}sec^{-1}$, correspond to the Ni-IMS RIP, and the others, with a mobility coefficient of 1.99 $cm^2V^{-1}sec^{-1}$ and 1.83 $cm^2V^{-1}sec^{-1}$, are the monomer and dimer ionic species formed from the ethanol . In the figure, training data are represented in blue color. Higher concentrations are shown with lighter tones and lower concentrations are represented with darker tones. In addition three different samples from the validation set are represented in red.
Figure 5.2 (b) depicts peak height of the RIP, monomer and dimer of ethanol from training data. By simply visual inspection, it is clear that exist a non-linearity behaviour with the concentration. From calibration point of view, the fact of having non-linear behaviour from the same pure analyte become univariate calibration a challenge. In addition, the monomer is located near to the tail of main RIP, so small changes in the RIP tail may interfare with future predictions.

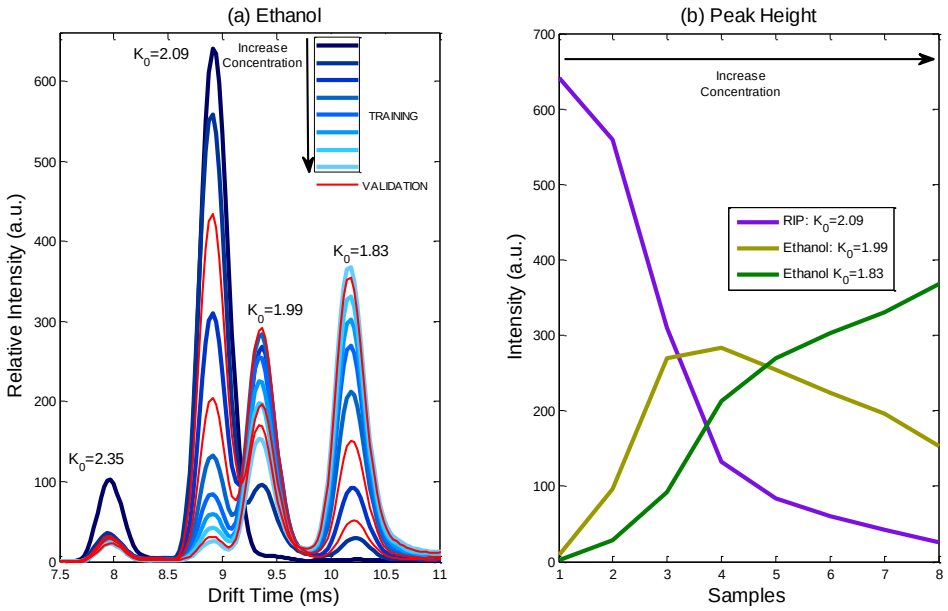


Figure 5.2 Ethanol Sample. (a) Ethanol Spectra at different concentrations. In blue is represented training set and in red is represented validation samples (b) Peak Height of RIP (K_0:2.09) and Ethanol Monomer (K_0:1.99) & Dimer (K_0:1.83)

A comparison between univariate and multivariate calibration is shown in Figure 5.3 using information from ethanol samples. In order to make easier univariate calibration procedure, monomer height peak is not taking into account when the model is built. In addition, the dimer peak is not overlapped with other peaks, thus its height is well determined and univariate calibration can be easily built as linear model. On the other hand, PLS was used as multivariate technique which use whole spectrum information to build a model and the number of latent variables was selected using a leave one out cross-validation procedure. Root mean square errors were calculated to test the final models for both training set (leave one out cross validation RMSECV) and validation set (RMSEP).

Figure 5.3 (a) shows the final univariate model and Figure 5.3(b) represent the final PLS model. Training data are shown in blue and validation data are in red color. It is noticed the error for both training and validation at least is twice smaller when PLS model is applied, and also univariate model cannot tackle the nonlinear behavior just multivariate calibration does. Furthermore, PLS calibration model use 6 latent variables which could explain the complexity of the nonlinear problem; even though a pure analyte was used in the experiments. Additional measurements were performed in a different day (Day 2), and the same concentrations levels were projected in both models to test its robustness. The results depict in Figure 5.3 (c) and (d) for univariate and multivariate calibration respectively. The error in both cases are a slightly worse than when measurements from the same day are used as validation. The instrument may be drift from day to day or external conditions changes thus the response is a little bit different. In any case, the PLS model is able to get a better fit and prediction than univariate model. This results confirm that univariate calibration is not enough precise despite peaks are well resolved as dimer of ethanol in this example (Fraga et al., 2009). Consequently, it is advisable to use multivariate calibration in most of the cases when IMS is working in order to get a better understanding of the measurements.

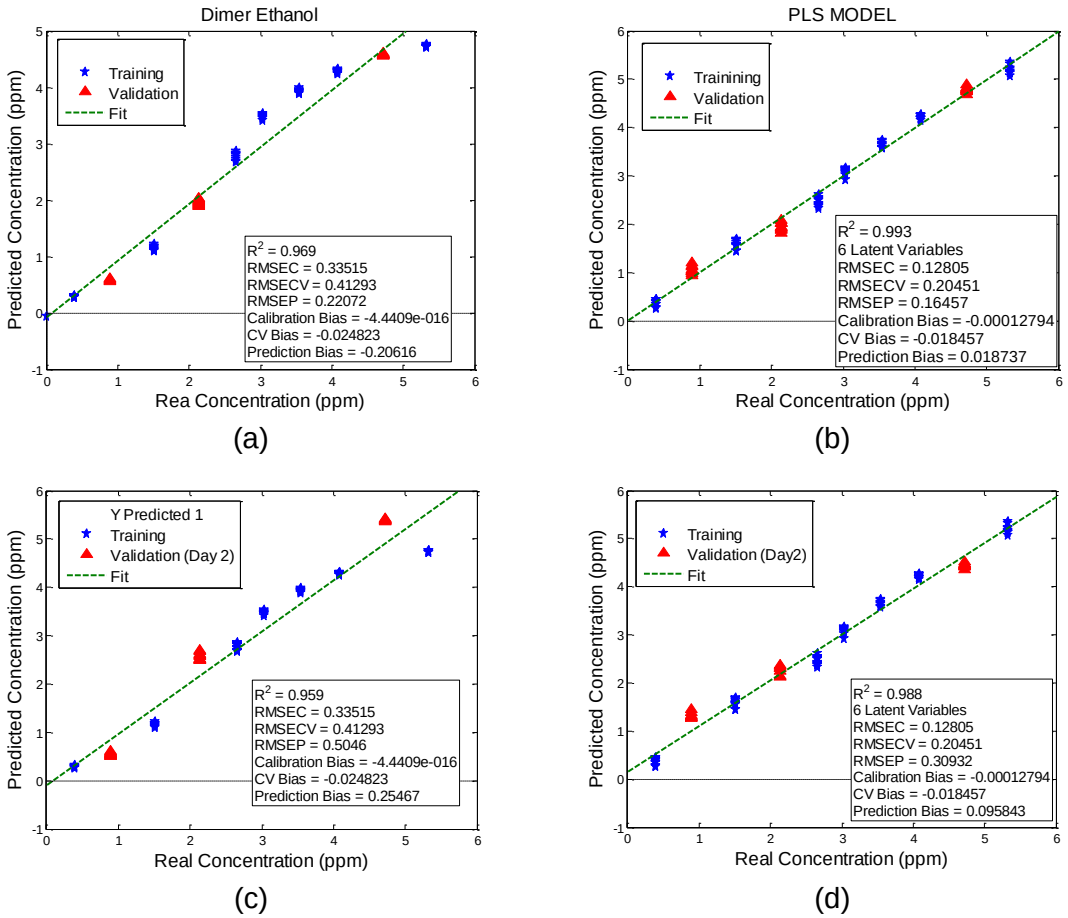


Figure 5.3 (a) Univariate Calibration and prediction using dimer information (b) Univariate Calibration and prediction of measurements done in a different day using dimer height peak (c) PLS model using whole spectra information. (d) PLS model and prediction using measurements done in a different day.

From the last results is evident that PLS model provide better results than univarite model. However, the interpreation of the model from the physic-chemical point of view is not easy. The loadings and scores of the first four latent variables of the final PLS model is shown in Figure 5.4 (a) and (b) respectively, which explains 98 % of the total variance. It is observable that the peaks has negative values which do not have any chemical meaning . Alternatively other multivariate models, such as those based on blind source separation, allow a better interpretation of the results, outperforming the PLS models performance

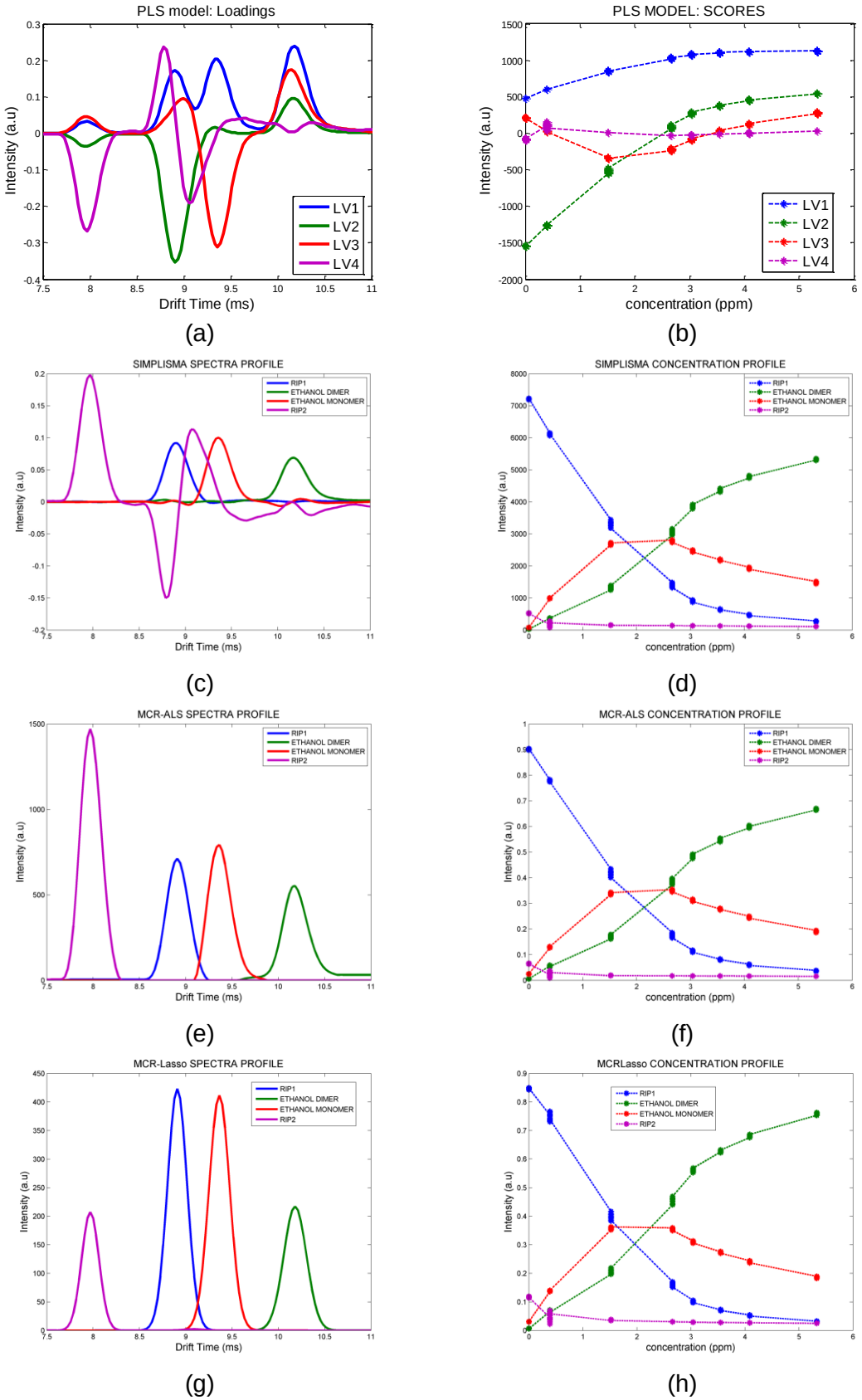


Figure 5.4 (a) Loadings of PLS model (4 Latent variables) (b) scores of PLS model (4 Latent variables) (c) Spectra profile which was obtained using SIMPLISMA using 4 pure variables (d) Concentration profile which was obtained using SIMPLISMA (e) Spectra profile which was obtained using MCR-ALS (f) Concentration Profile which was obtained using MCR-ALS (g) Spectra profile which was obtained using MCRLasso (h) Concentration Profile which was obtained using MCRLasso

In this context, models using SIMPLISMA (Harrington et al., 1997), MCR-ALS(de Juan et al., 2000) and MCRLasso (Pomareda et al., 2010) were built using the same training data of ethanol previously used, so that a comparison between them can be done. Figure 5.3 (c) and (d) shows the spectra profile and concentration profile which was got using SIMPLISMA approach that recovered 99% of the raw data. The only requirement in SIMPLISMA was to select the four peaks that was observed by visual inspection. Certaintly, there is a gain in terms of interpretation and identification of the four compounds if it is compared to PLS loadings (Figure 5.3 (a)), and specifically for the first three compounds (RIP, monomer and dimer of ethanol). It seems that the fourth pure compound (RIP: 2.35 $cm^2V^{-1}sec^{-1}$) , which only explain 1.5% of the data, are also recovering another variation of the data such as the variation of the tail between RIP (2.09 $cm^2V^{-1}sec^{-1}$) and the monomer of ethanol (1.99 $cm^2V^{-1}sec^{-1}$). Nevertheless, negative values in spectra profile do not provide any useful or real information. Furthermore, the concentration profile gives a semi-quantitative information about the concentration of the ethanol, in spite of no constraints were applied into the model like the total charge must be kept constant between all compounds.

Using SIMPLISMA results as first estimations, MCR-ALS can be built to refine the extraction of pure compounds. While, constraints cannot be directly applied in SIMPLISMA, MCR-ALS allows adding chemical or physical constraints into the model. For this purpose, the following constraints were applied to the ALS loop:

- non-negative both in spectra and concentration profiles,
- unimodality because it is supposed to have just one peak for each pure compound in spectra profile,
- closure in concentration profile due to charge must be stayed constant.

The final results is shown in Figure 5.3 (e) and (f) for respectively spectra and concentration profile in which the final model capture a 97% of the original data. As observed in the spectra profile (Figure 5.3 (e)), the pure compounds are better modeled than Figure 5.3(c) because of the use of constraints, but the tails of the third compound are not really well determined, may be, they are capturing some of the noise of the signal. The concentration profiles are normalized to 1 in order to keep equal charge between the compounds and they also represent a semi-quantitative value of the concentration.

In a similar way SIMPLISMA can be used as first estimation of MCRLasso. As MCRLasso imposes hard modeling, in this case a Gaussian model was used to fit the data. The Gaussian width of the model was related to the resolution of the instrument, in this the peak resolution (Spangler, 2002) of Ni-IMS is 32 as Eq. 3.8. Another important factor is the penalty or regularization parameter (λ) of LASOO which should be adjusted by cross validation and in this case was suited as 0.6 according to cross-validation results. Figure 5.3 (g) and (h) depicts the spectra and concentration profile respectively of the results using MCRLasso, where the final model recovered an 85% of the expected power from the raw data. It is obvious that slightly better modeled peaks were obtained than MCR-ALS and also the concentration profile was normalized imposing closure constraint which can be used as semi-quantitative values of the concentrations.

Another experiment with ethanol and 2-butanone was carried out in similar conditions (see chapter five section 5.2.1 for details). 2-butanone mean spectrum (averaged over all scans) is shown in Figure 5.5 where four main peaks can be seen in the spectrum. The first peak with a reduced mobility K0=2.10 cm^2 V s^{-1}, is related to reactant ions from the Ni^{63} ionization source and is always present. The second peak with K0=1.95 cm^2 V s^{-1}, is related to the protonated monomer of 2-butanone. The proton-bound dimmer of the analyte appears at high concentrations with K0=1.64 cm^2 V s^{-1}, and an additional third peak, whose behavior is strongly correlated with proton-bound dimmer, appears at the right of proton-bound dimmer with K0=1.55 cm^2 V s^{-1}.

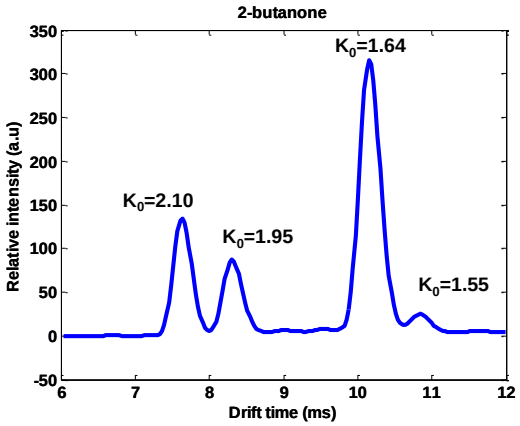


Figure 5.5 IMS mean spectrum for 2-butanone. Reduced mobility (K_0) of RIP: 2.10 cm^2 V s^{-1}, 2-butanone monomer: 1.95 cm^2 V s^{-1}, 2-butanone dimer: 1.64 cm^2 V s^{-1}

MCR-ALS (de Juan et al., 2000) was applied to 2-butanone dataset to resolve the evolution of formed species. SIMPLISMA (Cao et al., 2005, Harrington et al., 1997) was used to extract initial estimations for spectra and concentration profiles prior to MCR-ALS in which 3 components were imposed to the algorithm. Non-negativity, unimodality and closure were the constraints used within the ALS loop. The analysis was performed using the spectra region from 6ms to 12ms where relevant peaks appear.

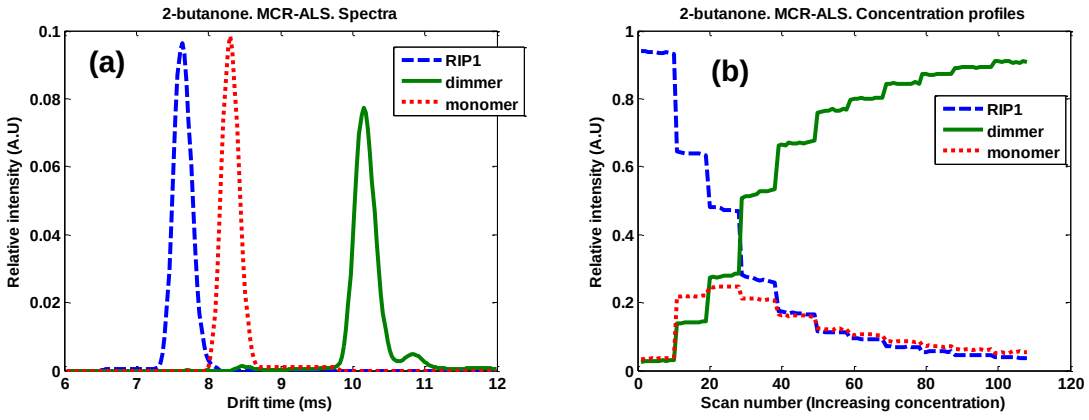


Figure 5.6 MCR-ALS results for 2-butanone spectra. (a) Spectra profile. (b) Concentration Profile.

Figure 5.6 shows the results by MCR-ALS for 2-butanone. Clearly, 2-butanone has non-linear behavior in similar way than ethanol. Moreover, since protonated-bound dimmer peaks appear at high concentrations and their behavior differs from monomer,

the concentration of the substance need to be explained using more than one component in SIMPLISMA and MCR-ALS. In the studied cases, MCR-ALS is able to resolve the different components properly. As it can be seen in Figure 5.6 and Figure 5.4 (e and f), the intensity of reactant ion peaks decreases as substance concentration increases. Although protonated monomers start increasing their intensities at low concentrations, they reach their maximum intensity in a certain substance concentration and then start to drop off. At the same time, proton-bound dimmer peaks increase their intensity when substance concentration rises further, but they reach a saturated behavior at very high concentrations. Furthermore, sometimes clustering formation takes place between the analyte and water molecules either in the reactant region or in the drift tube(Eiceman and Karpas, 2005), as a result of this chemical process a new peak could appear in the signal.

This is observed in the MCR-ALS results where a secondary peak appears in the dimmer components; the peak located at the right of the dimmers is related to a product formed by the proton-bound dimmer and a water molecule. In this case the explained variance was 99.7% that indicate MCR-ALS model is able to explain almost the total variance of the raw data and at the same time provide an easy interpretation for the different contributions. Even though, the percentage of explained variance is quite high, note that the tails of the dimer of 2-butanone is not fully well modeled. This means that the fact of having a high explained variance does not mean the peaks are properly modeled. MCR-ALS tries to explain also the noise present in the tails, but it is not able to properly solve it. An alternative can be the use of hard modeling such as MCRLasso for providing more accurate results with similar explained variance as it was seen in Figure 5.4 (g) with ethanol.

It was seen in Figure 5.3 the use of multivariate techniques such as PLS for calibration purposes gives better results than the use of univariate calibration. However MCR-ALS provides more interpretative results than the loadings of PLS model. In addition, the concentration profile of MCR-ALS can be used for building a calibration model. Therefore, MCR-ALS provides interpretation of the compounds of the sample and using a proper calibration method the quantification of the instrument. On the other hand, PLS model should be preferable when the main objective is just performing a calibration losing the possibility of interpretation.

5.2.1. Non-linear effect in IMS using synthetic dataset

Clearly, the non-linear behavior of IMS data was demonstrated in the last section. In addition, it was discussed the main issue when univariate is used for performing a calibration model, whereas multivariate techniques should be a better option for quantitative proposes. Even though, the quantitative model has significantly improved using multivariate calibration, some of the algorithm does not completely solve the strong non-linear behavior of IMS data. Thus, the challenge is to find a solution for this kind of behavior. In this thesis to use MCR algorithms combining with non-linear calibration methods for solving the non-linearity is proposed.

Figure 5.7 in section 5.2.1 explains the block diagram that was carried out for tackling this comparative analysis. The main objective of this analysis is to compare different alternatives for solving non-linarites in data analysis of IMS. The compounds that was chosen in this work present a strong nonlinear behavior as concentration increases -

especially monomers and dimmers peaks, thus polynomial PLS should be used to construct the calibration model rather than PLS or any other univariate technique. The output information from MCR-ALS concentration profiles is used to construct a new matrix (X) with dimensions M x N, where M is the number of samples and N = 2 (monomer and dimmer concentration profiles from MCR-ALS). A matrix of concentrations (Y) with dimensions M x R can also be constructed, where R=1 since we only have one substance per model. Using X and Y matrices, an optimum calibration model can be built using the cross-validation methodology explained in section 5.2.1 in chapter 5. Moreover, this cross-validation methodology can be used to assess the performance of the calibration model. In the same way, a polynomial was fit to construct a calibration model, and the same cross-validation methodology was used for set up the order of the polynomial.

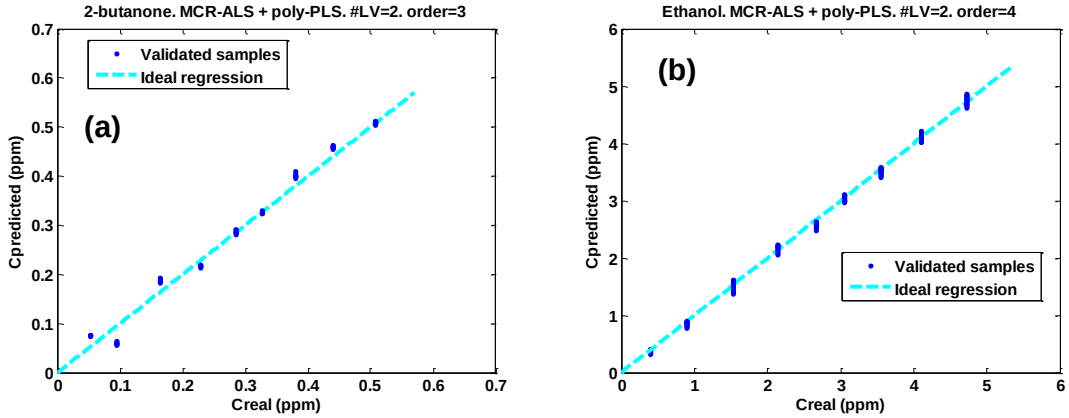


Figure 5.7 Predicted concentrations in function of substance concentrations for validation samples projected over constructed poly-PLS models. (a) Predicted 2-butanone concentrations using poly-PLS models with 2 latent variables and polynomial order =3. (b) Predicted ethanol concentrations using poly-PLS models with 2 latent variables and polynomial order =4.

Figure 5.7 shows the predicted concentrations versus the original concentrations for 2-butanone and ethanol using poly-PLS as calibration method after obtaining MCR-ALS concentration profiles. The figure only shows the validation results. Note that although the validation results are depicted on the same graph, each set of scans (belonging to a particular substance concentration) has a different calibration model (built from leave-one-block-out cross validation method). The optimum polynomial order was found to be 3 for 2-butanone data and 4 for ethanol data. The RMSECV was 5.6% (relative to full scale input range) for 2-butanone and 1.2% for ethanol (relative to full scale input range). The squared correlation coefficient was 0.98 for 2-butanone and 0.998 for ethanol. The results show that prediction accuracy is quite good using the combination of MCR-ALS and poly-PLS model.Univariate and multivariate calibration models can also be built without using MCR-ALS concentration profiles. Figure 5.8 and Figure 5.9 show predicted concentrations for ethanol and 2-butanone respectively, using the same cross-validation methodology. It can be seen that high polynomial order was needed for fitting the univariate curve as well as the number of latent variables for PLS model is quite high that means the complexity of the dataset is really meaningful. Note that univariate models are not able to deal with the non linearities of the dataset, and for the case of 2-butanone the PLS model do not provide good results either.

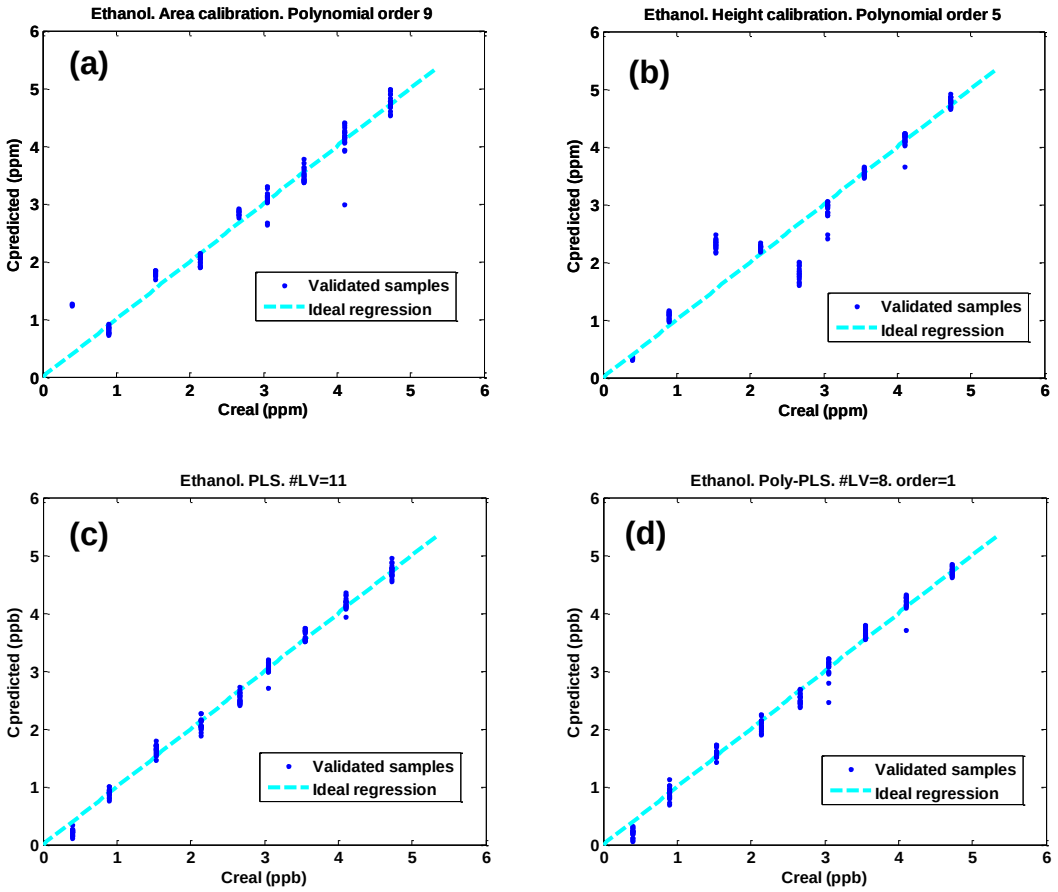


Figure 5.8 Predicted concentrations vs substance concentrations for validation samples projected over different calibration models. (a) Predicted ethanol concentrations using area calibration and fitting a polynomial of 9th order. (b) Predicted ethanol concentrations using height calibration and fitting a polynomial of 5th order. (c) Predicted ethanol concentrations using PLS models with 11 latent variables. (d) Predicted ethanol concentrations using poly-PLS models with 8 latent variables and polynomial of order 1.

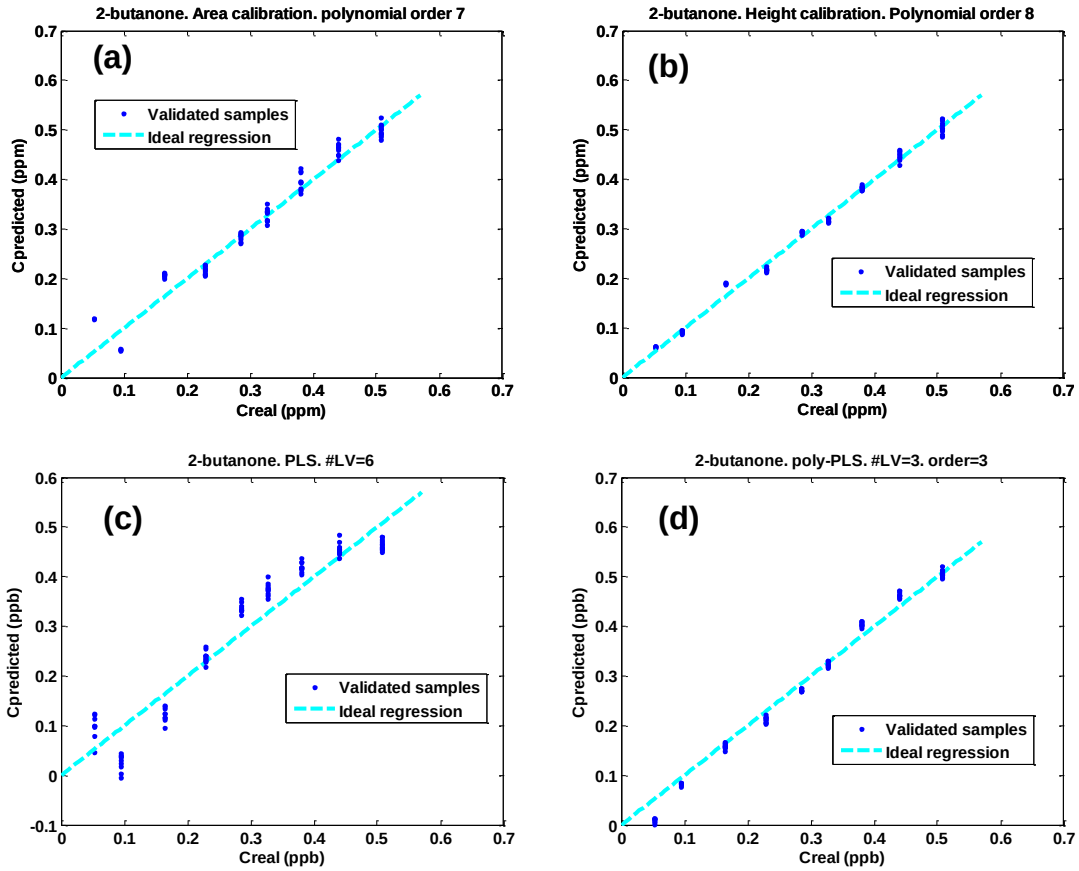


Figure 5.9 Predicted concentrations vs substance concentrations for validation samples projected over different calibration models. (a) Predicted 2-butanone concentrations using area calibration and fitting a polynomial of 7th order. (b) Predicted 2-butanone concentrations using height calibration and fitting a polynomial of 8th order. (c) Predicted 2-butanone concentrations using PLS models with 6 latent variables. (d) Predicted 2-butanone concentrations using poly-PLS models with 3 latent variables and polynomial of order 3.

Numerical results comparing univariate and multivariate techniques using and not using MCR-ALS concentration profiles are presented in Table 5.1. As it can be seen, univariate techniques can provide good results if peaks in the spectra do not appear overlapped and thus can be easily identified to calculate their area or extract their height, this is the case for 2-butanone. However, when peaks appear overlapped (case of ethanol) these techniques fail because contributions from other peaks appear in the region of the peak of interest. In situations with a high overlap between peaks the use of univariate calibration techniques can be unfeasible, unless a prior deconvolution step is carried out (e.g. using Truncated Negative Second Derivative). Using multivariate techniques, better calibration models than univariate techniques can be built as it was already proved by Fraga et. al in (Fraga et al., 2009).

Calibration method	R^2		RMSECV (% max conc)	
	2-butanone	Ethanol	2-butanone	Ethanol
Peak area (U)	*0.96*	*0.95*	*5.6*	*6.3*
Peak height (U)	*0.993*	*0.91*	*2.3*	*7.9*
PLS (M)	*0.91*	*0.993*	*7.7*	*2.3*
poly-PLS (M)	*0.992*	*0.991*	*3.0*	*2.6*
MCR-ALS + PLS (M)	*0.85*	*0.97*	*10*	*5.7*
MCR-ALS + poly-PLS (M)	*0.98*	*0.998*	*5.6*	*1.2*

Table 5.1 Comparison between different optimized calibration methods using leave-one-block-out cross validation. Results include univariate (U) and multivariate (M) methods. The best results are shown shaded.

Clearly, results from the calibration model, when PLS and poly-PLS (Wold et al., 1989) were directly applied to IMS spectra, show similar prediction accuracy. However, the number of latent variables is too high for performing an interpretation. For the PLS case, 6 and 11 latent variables for 2-butanone and ethanol respectively. For the poly-PLS case, 3 and 8 latent variables for 2-butanone and ethanol respectively. This fact hinders the qualitative interpretation of the results since many different contributions need to be taken into account in order to understand the chemical process involved in the substance behavior as concentration increases. Moreover, since no constraints are imposed to the regression coefficients, negative values which do not have any physical and chemical meaning can be found.

For instance, Figure 5.10 shows the scores and loadings from a poly-PLS calibration model with the same number of latent variables as the number of components used in MCR-ALS for 2-butanone (Figure 5.10 a and b) and ethanol (Figure 5.10 c and d). The cross-validation procedure has been applied in order to optimize the polynomial order. It is shown the difficulty to interpret the results compared to MCR-ALS solutions (Figure 5.4 and Figure 5.6) since many contributions need to be taken into account. If the optimum calibration model includes more latent variables, although prediction can be better, the interpretation of the results is even more difficult, which is the case for the results presented in Table 5.1.

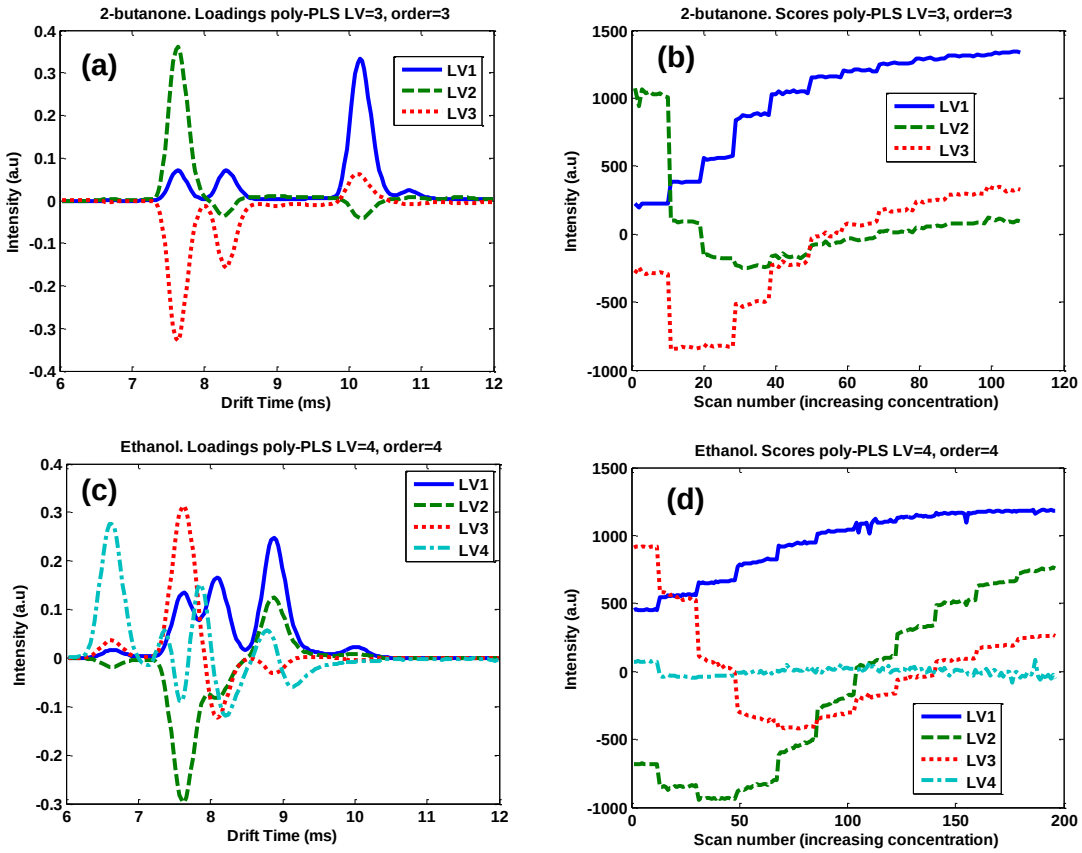


Figure 5.10 Scores and loadings from poly-PLS calibration models using the same number of latent variables as the number of components used to build MCR-ALS models. (a) Loadings for 2-butanone. (b) Scores for 2-butanone. (c) Loadings for ethanol. (d) Scores for ethanol.

In order to test the proposed methodology, different calibration model was build using measurements from two different days. The results are comparable with the results previously discussed provided the calibration and validation samples are obtained within one day. Nonetheless, instrumental drift degrades prediction accuracy which is not deeply treated in this work. This point has been already observed by different authors (Fraga et al., 2009).

The results of the different days are shown in Table 5.2 and Table 5.3. They are similar to that shown in Table 5.1. In these cases, using MCR-ALS as prior step to poly-PLS calibration provides the best results. PLS and poly-PLS directly applied to IMS spectra also provide good prediction accuracy but, as explained before, the interpretation of the chemical process is difficult since the optimum number of latent variables is too high.

Calibration method	R2		RMSECV (% max conc)	
	2-butanone	**Ethanol**	**2-butanone**	Ethanol
Peak area (U)	0.99	0.96	1.1	5.8
Peak height (U)	0.94	0.90	6.4	8.4
PLS (M)	0.99	0.99	1.2	1.8
poly-PLS (M)	0.99	0.99	2.7	2.3
MCR-ALS + PLS (M)	0.99	0.97	2.4	5.2
MCR-ALS + poly-PLS (M)	0.99	0.99	0.8	1.2

Table 5.2 Day 2. Comparison between different calibration methods using leave-one-block-out cross validation. Results include univariate (U) and multivariate (M) methods. The best results are shown shaded.

As it is shown in Table 5.1, Table 5.2 and Table 5.3, calibration models can be built within the same day and be used for prediction within the same day; however, there is a large variability in the evolution of monomer and dimmer among different days, especially in the 2-butanone case. This variability is much less in the case of ethanol. This result suggests that for some substances, calibration models constructed in one day cannot be used to predict new samples measured in a different day. These models would be only valid in the same day as it can be extracted from the results presented in Table 5.2 and Table 5.3. In the ethanol case, although calibration models constructed in one day could be used in different days, the study of the use of calibration models for prediction in different days is out of the scope of this work.

Calibration method	R2		RMSECV (% max conc)	
	2-butanone	**Ethanol**	**2-butanone**	**Ethanol**
Peak area (U)	0.87	0.96	12	6.1
Peak height (U)	0.99	0.92	2.7	7.8
PLS (M)	0.99	0.99	1.4	2.3
poly-PLS (M)	0.99	0.99	1.6	2.4
MCR-ALS + PLS (M)	0.89	0.97	8.8	5.8
MCR-ALS + poly-PLS (M)	0.99	0.99	1.2	1.3

Table 5.3 Day 3. Comparison between different calibration methods using leave-one-block-out cross validation. Results include univariate (U) and multivariate (M) methods. The best results are shown shaded.

5.2.2. Mixture effect in IMS using synthetic dataset

A second effect that was studied in IMS is the mixture effect. In this case, two biogenic amines (trimethylamine (TMA) and putrescine (PUT)) have been studied from a multivariate signal processing scope, as it was explained in section 5.2.2 in chapter 5. Note, both biogenic amines have a similar proton affinity and, a priori, their mixture should not have a big charge competitive effect (Karpas et al., 1994). However, as it is shown in Figure 5.11, there are important changes in add-mixture matrix. The first issue to notice is how the spectrum changes when a mixture is analyzed. For instance, the intensity of the TMA is lower when it is measured as pure analyte than when the same concentration is mixed with a high concentration of PUT.

On the other hand, the peak of PUT in the mixture is almost undistinguishable; indeed, just a small peak appears on the tail of the TMA peak though a high concentration of PUT was measured such as it is shown in the response of the IMS to the pure

compound (dot red line). This behavior is not only due to proton affinities, there are other factors that may contribute to it such as temperature and humidity of the spectrometer that may lead changes in the cluster formation. Nevertheless, the focus of this work is to discuss about the quantitative effect of TMA when PUT acts as interferent. This is really important since both biogenic amines have been used in the diagnosis of vaginal infections (Marcus et al., 2012, Sobel et al., 2012, Karpas et al., 2002a), and until now a ratio between the TMA and the other compounds has been calculated for the diagnosis using univariate techniques. If other biogenic amines interfere in having an accurate diagnosis, it can be really useful to tackle this problem using multivariate strategies. The study is based on the determination of the limit of detection (LOD) of TMA with and without PUT comparing the performance of the calibration between univariate and multivariate techniques (see Figure 5.8).

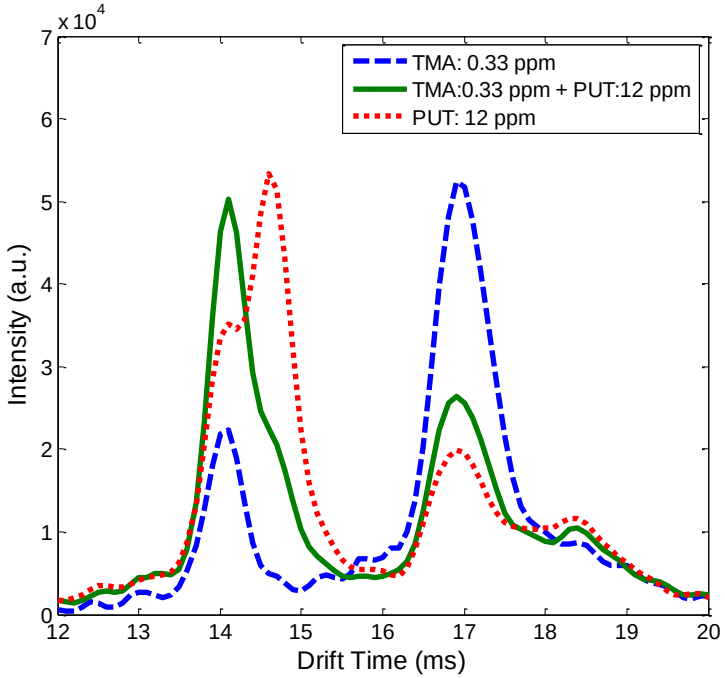


Figure 5. 11 Spectra of IMS for pure analytes and mixtures. In dashed blue line is present TMA at 0.33 ppm. In solid green line is shown the mixture of TMA and PUT, and in dot red line is shown PUT at 12 ppm.

Considering the effect of PUT on the response of TMA, the LOD of the instrument might be different than 0.1 ppm(Karpas et al., 2013) as it was calculated in section 5.1. To build a calibration curve, concentration of TMA at levels closer to the LOD was measured as pure analyte and mixed with two different concentrations of PUT. In addition, the pure analyte of PUT was measured at different concentrations and blanks of the instrument to estimate the LOD and RMSECV.

The first approach seeks to calculate the ratio between the TMA and the other compounds including TMA which was proposed elsewhere (Marcus et al., 2012, Sobel et al., 2012, Karpas et al., 2002a). The ratio, in reality, performs a normalization of the substances and it is slightly similar to do a normalization of area equal to 1. Thus, big changes of the signal are likely to be minimizes. The height of each peak was evaluated in the ten last spectra of each measurement and the mean value was taken

for building the calibration curve. The same procedure was repeated in spectra of the blanks that were split as validation set and then projected into the calibration model. Since the number of blanks is 13, the t-value that was chosen from t-table with a 95% of confidence is 2.16. This value was used in the Eq. 5.5 to calculate the LOD. Moreover, a 17-fold cross validation was applied to get the root mean square error of cross validation (Eq. 5.4).

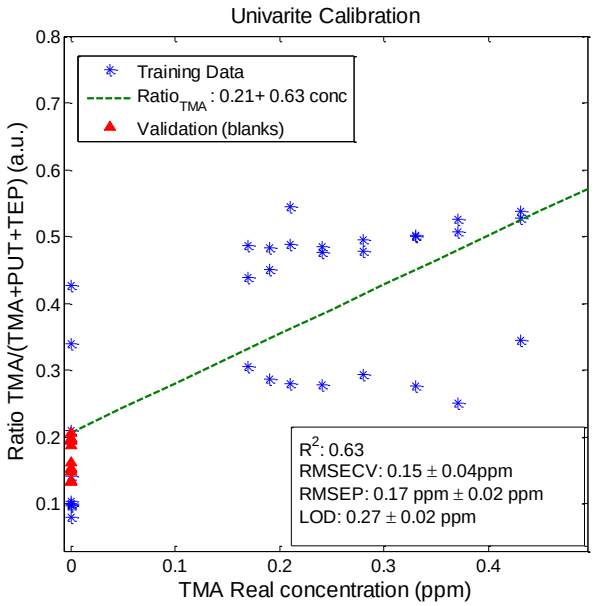


Figure 5.12 Univariate Calibration using ratio (TMA/TMA+TEP+PUT)

The results of the univariate calibration are shown in Figure 5.12. A high variation of the training data is distinguished and the same behavior is observable in the validation data. This variability is mainly due to the influence of the PUT on the TMA peak. For example, the data that is under the fit line (dashed green line) are related to measurements of pure TMA and over it are measurements of mixture of PUT and TMA. Obviously, LOD is directly affected by this influence, thus the LOD that is obtained is at least twice that the LOD expected (0.1 ppm). Nevertheless, the RMSECV and the RMSEP are quite reasonable. This results show how the ratio is really affected by the presence of other compounds and univariate techniques is not able to deal with it.

The multivariate calibration model was done using the mean spectrum over 10 spectra, as it was done with the univariate analysis. In this case the number of latent variables was set up using a 17-fold cross validation method giving as a result of 8 latent variables (LV) recovering 98% of the total variance. In Figure 5.13 (a) the loadings of the first three latent variables are shown together with these respective percentages of explained variance. Despite of the fact, the interpretability of the model is quite difficult, it can be seen that LV two and three try to compensate the mixture effect and the first LV mimics the TMA behavior.

One main advantage of the multivariate models is information of the mixture - in this case both amines - can be simultaneously obtained as it is seen in Figure 5.13 (b) and

(c). It can also be seen the accuracy of both models (TMA and PUT) are moderately different, for example both RMSECV and RMSEP for the TMA model are lower than the PUT model.

It can be notice in the TMA model, there is not a high variance between samples with or without PUT as it is depicted in Figure 5.12. Moreover, the LOD are really closer to the expected one (0.1 ppm) though the multivariate limit of detection assumes that there is an important contribution of the undesirable compounds and the same t-value was used for the calculation. If the experiment had done in different conditions such as lower levels of PUT concentration, the LOD of the PUT would have been calculated too. A preliminary conclusion is that multivariate methods allow enhancing the understanding and the quantification of compounds in presence of mixtures, though the kinetics of the instrument involves important changes in the resultant measurements or spectra.

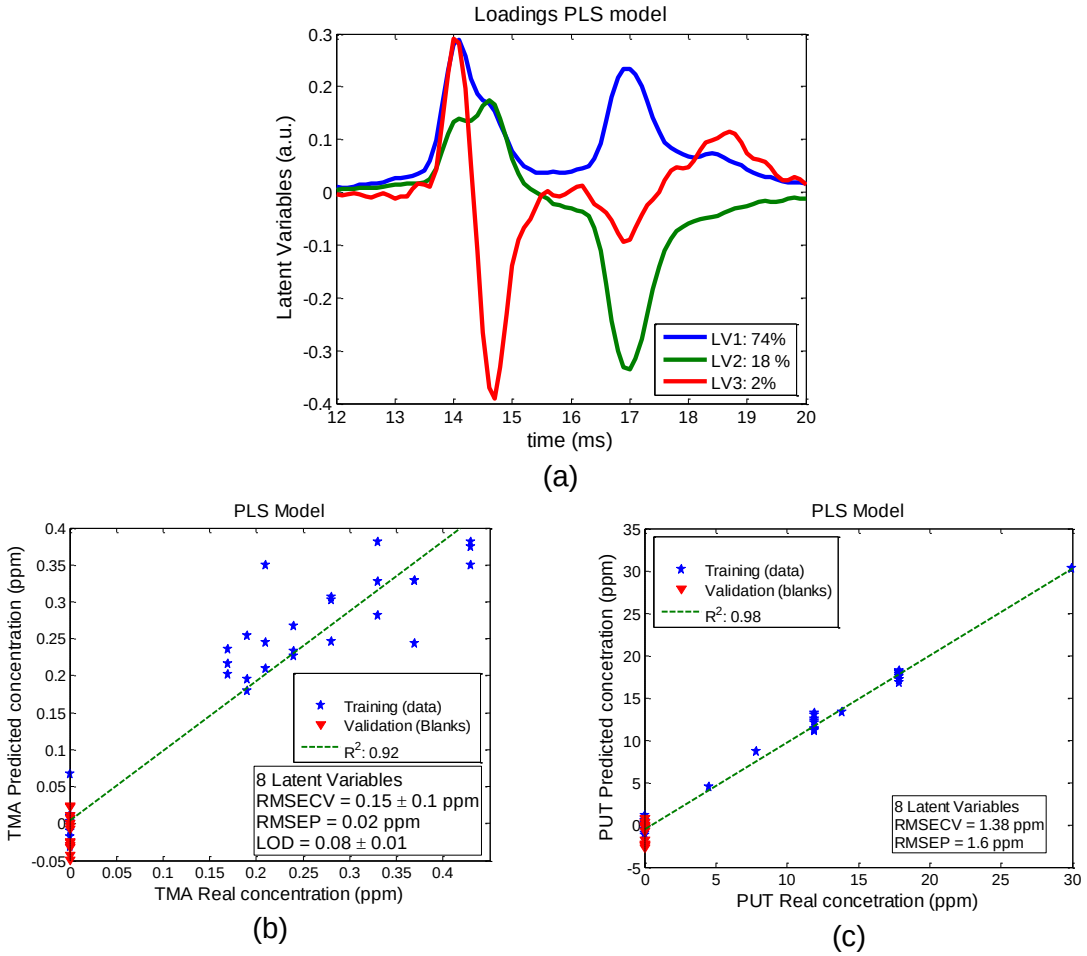


Figure 5.13 Multivariate Calibration uisng the whole spectra information. (a) Loadings of the three first latent variables of the PLS model. (b) Calibration curve of TMA (c) Calibration curve of PUT.

The third approach seeks to deconvolute the spectra and extract the pure compounds which are involve in the mixture. As it has been discussed before, the use of

techniques as hard modeling MCR-Lasso allows extracting the pure compounds leading spectra and concentration profiles. Considering that the number of compounds is known a priori, the initial estimation was settling to get three pure compounds one for each amine (TMA, TEP, and PUT). The first estimation using SIMPLISMA(Windig and Guilment, 1991) recovered a 98% of the data and the final estimation gets 95% of the data. The resolution to generate the Gaussian models was the same as the used before in section 5.1, and it was also applied closure as constraint in the concentration profile.

The final spectra and concentration profile of the MCR model is shown in Figure 5.14 (a) and (b) respectively. In the spectra profile, it can be seen how well the three compounds are modeled likewise the overlapping between TMA and PUT. On the other hand, the concentration profile, which was imposed a closure equal to one, shows a semi quantitative result of each compound in the different scenarios like blanks, pure compounds and mixture. Despite the percentage of the explained data is high, the contribution between both compounds in the mixture were not fully eliminated (see concentration profile Figure 5.14 (b)). For example, the TMA concentration profile (red line-concentration profile) has important contributions from PUT; otherwise the information can be used by its own to build a univariate calibration to estimate the LOD. Perhaps, a better modeling can be obtained, if additional information is integrated in the modeling process such as chemical information related to kinetics of the compounds involve in the measurements.

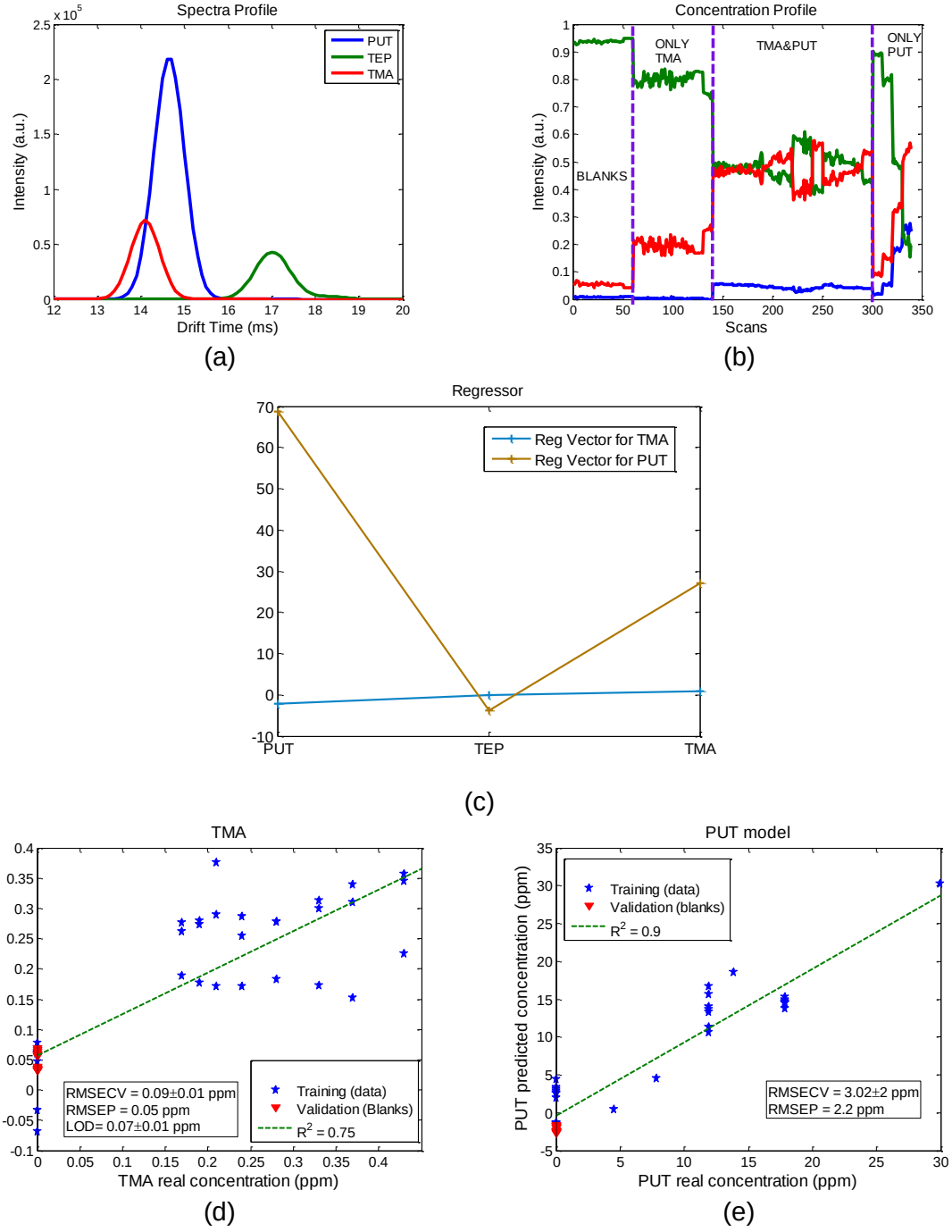


Figure 5. 14 (a) Spectra profile and (b) concentration profile as result of MCRLasso procedure. Regression model using multiple linear regression (MLR) (c) Regressors of the model, (d) TMA model and (e) PUT model.

Multiple linear regression algorithms was used for building the calibration model since the number of samples is bigger than the features. The final calibration model are depict in Figure 5.14(d) for TMA and Figure 5.14(e) for PUT. The TMA model is slightly similar to the univariate model. This is due to the inability of MCR-LASSO to completely reject the contribution of the PUT in the mixture. Nevertheless, the final figures of merits are as good as the PLS model and LOD is really closer to the expected one. The calibration model of PUT is quite worse, if it is compared with PLS model. It is due to there is an important contribution of TMA in the PUT model.

Figure 5.14 (c) shows the contributions of each pure compound in both calibration model. The first regressor linked to TMA model shows the pure compound related to TMA has more weight than the others. However, TEP and PUT is still having a slightly contribution in the model. The second regressor is associated to PUT model, in this case the pure component of PUT has a bigger importance in the model than the others. Nonetheless, it is still clear that TMA has a contribution. It can be seen that the regressor of the PUT calibration gives more importance at feature related to PUT than TMA and the other way around as it was expected.

The results of the three models are summed up in Table 5.4. An important improvement is got when a multivariate method is applied. The best performance is achieved when PLS model is used even though the lack of interpretability of the model. In contrast, univariate model gives too pessimistic results due to lack of rejection of the interferent.

TMA	Univariate Model	PLS model	MCR+MLR model
LOD (ppm)	0.27±0.02	0.08±0.01	0.07±0.01
R^2	0.63	0.92	0.75
RMSECV (ppm)	0.15±0.04	0.15±0.01	0.09±0.01
RMSEP (ppm)	0.17	0.02	0.05

Table 5.4 Limit of detection, root mean square error of prediction (RMSEP) and cross validation (RMSECV), and fit of the model (R^2) for univariate model, PLS model and MCR+MLR model.

Conclusions of non-linear and mixture analysis of IMS

In this work two main nonlinear problems about building IMS calibration models have been studied. The first one is the non-linear effect at higher concentrations and the second one is the non-linear effect introduced in mixtures. Thus, it is fundamental the study of both behaviors in controlled conditions and find a feasible solution for tackling these problems.

First of all, it is evident that univariate techniques are not able to deal with them. In the first case, the main problem was that a same compound gives rise to two peaks that come from monomer and dimer. Thus, the fact of choosing one peak or the other will drastically change the calibration results. Moreover, the evolution of both peaks regarding to concentration was really non-linear and the calibration model was not enough robust to solve it. In the other scenario, the performance of the univariate calibration was tested in the mixture and since the model does not consider the other compounds in the sample, the univariate model gives poorer results than multivariate models.

On the other hand, multivariate calibration gets better results than univariate techniques and provides interpretative benefits. The first alternative is use common and available techniques such as PLS, which benefits have been tested elsewhere(Fraga et al., 2009). And a second alternative is to combine blind source separation techniques for extract pure components of the sample and use any multivariate technique to get quantitative results.

This first approach seeks to evaluate a methodology to be applied to IMS spectra which combines the advantages of MCR-ALS for qualitative interpretation and poly-PLS for quantitative prediction of new samples which present a strong nonlinear

behavior as substance concentration increases. MCR-ALS has been demonstrated to be a suitable method to the study of ion mobility second order data. Using SIMPLISMA and MCR-ALS, IMS spectra are resolved in pure components and a qualitative estimation for the spectral and concentration profiles of these components is obtained. MCR-ALS allows the description of the chemical changes produced during run processes when concentration increases.

For the studied datasets, quantitative results show how the performances of standard multivariate calibration techniques are better than univariate techniques- especially when peaks in the spectra appear overlapped. Multivariate techniques are able to model nonlinear behaviors adding more components to the model. The datasets included strong nonlinear behaviors as substances concentration increased. While PLS is able to handle slightly nonlinear behaviors, strong nonlinear evolutions are better modeled using poly-PLS. Although prediction accuracy is similar, the results obtained from these standard techniques are often difficult to interpret, since, in order to model nonlinearities, the number of latent variables in the model is usually higher than the number of peaks. Using MCR-ALS prior to the calibration step provides a way to interpret properly the results and fix the number of latent variables, thus reducing the complexity of the calibration model.

The main goal of the second approach was to compare different quantitative strategies to deal with mixtures in samples. In fact, this issue has become a real challenge in the ion mobility spectrometry field which can occur either during the ionization of the molecules or during the transportation of the ionized molecules inside of the drift tube. Consequently, the effects in the spectra both in peaks, overlapping peaks or non-linear effects between the measured compounds can happen. These issues are the main reason why quantitative models become difficult to build.

In this study, the problem of mixtures is addressed from different perspectives starting with univariate analysis and afterwards explores the use of multivariate strategies. The main idea is to look out carefully the results and figure out the possible consequences. As main scenario, two amines, TMA and PUT, deeply used in the diagnosis of bacterial vaginosis(Marcus et al., 2012, Sobel et al., 2012, Karpas et al., 2002a) have been studied and the effect in the limit of detection of the TMA when PUT is measured at the same time. It has been seen that there is a nonlinear effect when a mixture occurs giving an increase of the intensity of the TMA and in the case of PUT a small peak is observable provided high concentration of PUT.

As it has been said above, univariate analysis provides over pessimistic results, concluding that more accurate or even reliable results can be obtained if a multivariate analysis is applied. At the same time, the LOD calculated by multivariate analysis are closer to 0.01 ppm which was obtained in a previous study when the substance was study without any mixture. Regarding, the two multivariate analyses proposed in this work, at least in terms of quantification the PLS model can deal and provide better results. Nevertheless, the main difficult is pointed out in the interpretation of the compounds involved in the measurement. On the other hand, the use of MCR allows extracting the pure compounds involved in the data-analysis and at the same time, to give a semi quantitative analysis. When the concentration profile is used to build a

multivariate calibration model, the quantification obtained are slightly worse than PLS model but enough satisfactory.

To conclude, it is remarkable that the use of multivariate techniques can provide more reliable results and tackle with complex problems present in measurements with regard to ion mobility spectrometry. Moreover, the usefulness of using blind source separation techniques in cases of extracting pure contributions of a complex matrix was proved. It seems that the percentage of recovery data is not enough figure of merit to establish if a model is well modeled by blind source separation techniques, at least in spectra of IMS. Another important conclusion the different versions of MCR such as ALS and MCRLasso shows better results than SIMPLISMA. Both MCR-ALS and MCR-Lasso returns accurately pure spectra profiles and concentration profile of the compounds. The main difference is: MCR-Lasso uses hard modeling and MCR-ALS uses a soft modeling procedure. Therefore, if the data fulfils the assumptions of MCR-Lasso, it might be a better option to use this methodology instead of the MCR-ALS. Moreover, concentration profile can be used as quantitative information as long as closure constraint is applied in both algorithms.

5.3. Feasible studies for testing IMS in real scenarios.

In the previous section, it has been demonstrated that multivariate techniques have better performance than univariate techniques for data processing of IMS measurements in lab conditions. It is a challenge to apply multivariate strategies in real scenarios with good performance and confidence on results. In order to test the multivariate strategy and the feasibility of the use of IMS in real scenarios, two scenarios have been explored.

5.3.1. Feasible study for detection of 2,4,6-tirchloroanisole (2,4,6-TCA) in wine using a portable Ni-IMS.

The methodology has been explained in detail in chapter 5 section 5.2.3 and the block diagram for the signal processing is shown in Figure 5.9. Briefly, in this case the problem to be tackled is provide a rapid screening of 2,4,6-thrichloroanisole(TCA) in wine. The main problem is TCA favor off flavor in wine that implies enormous economic losses in wine industry (ABCScience, 2013, Holmberg, 2010, iCEX, 2014). In this work, samples of TCA were measured with IMS in positive and negative modes to determine the limit of detection of IMS for this substance. Based on these results can be assessed the feasibility of using IMS as monitoring off flavor in wine.

The ion mobility spectra from the headspace vapor of 2,4,6-trichloroanisole in positive and negative modes in purified air are shown in Figure 5.15 (a) and (b), respectively, and the spectra with vapors of dichloromethane as a dopant are depicted in Figure 5.15 (c) and (d), respectively. Two peaks with reduced mobility values of 1.58 and 1.20 $cm^2V^{-1}s^{-1}$ were observed in the positive ion spectra. As an IMS-MS instrument was not available, identification of the ions and peak assignment was based on ion chemistry and drift time considerations. Thus, these peaks were assumed to arise from a TCA monomer and dimer ions, respectively, as ethers in general are known to form protonated monomers and dimers (Metro and Keller, 1973).

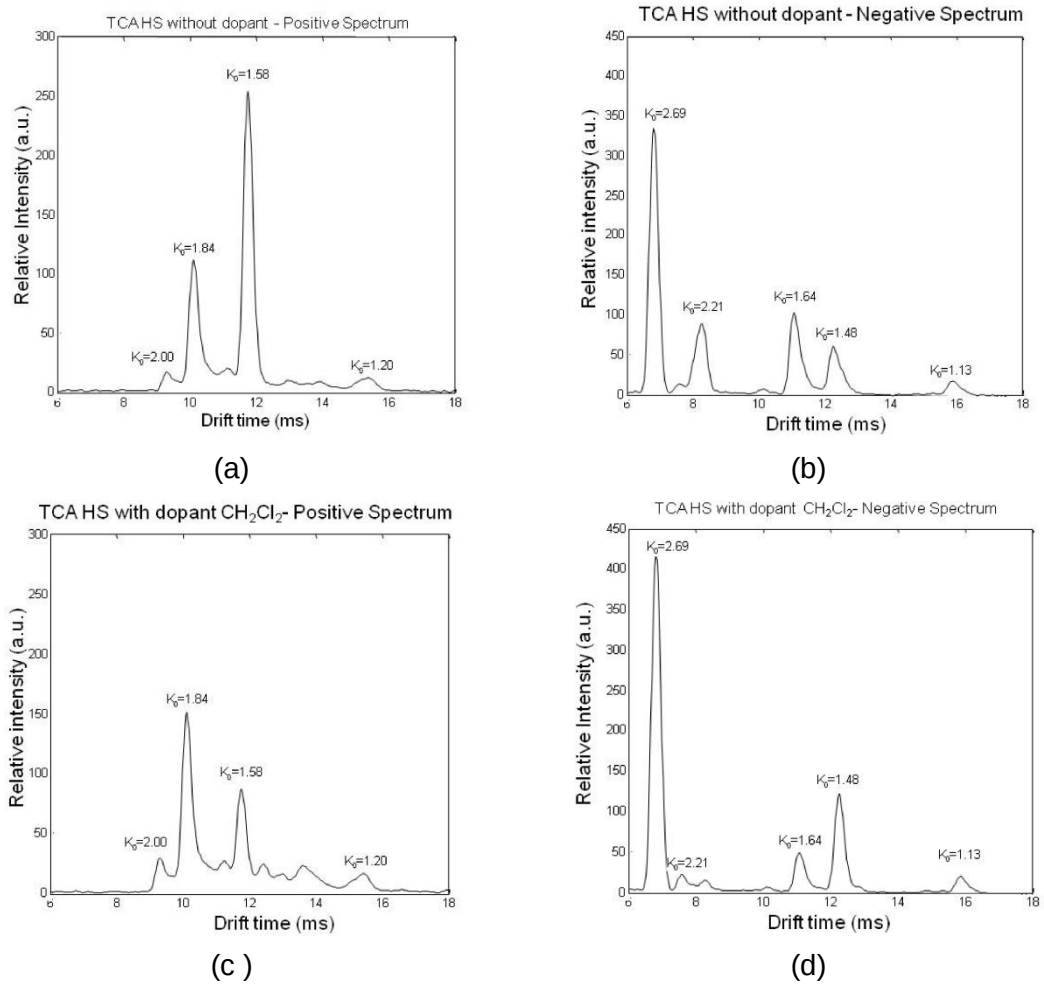


Figure 5. 15 (a) Mobility of TCA-without dopant, positive mode; (b) mobility of TCA-without dopant, negative mode; (c) Mobility of TCA with dopant, positive mode; and (d) mobility of TCA-with dopant, negative mode.

The dominant ion in the negative mobility spectrum was an ion with a reduced mobility value of 2.69 $cm^2V^{-1}s^{-1}$, identified as the chloride ion that is commonly detected in many aliphatic and aromatic chlorine compounds(Eiceman and Karpas, 2005). The ion with a reduced mobility of 1.64 $cm^2V^{-1}s^{-1}$, is quite similar to the ions reported for 2,4,6-, 2,4,5- and 2,3,5- isomers of trichlorophenol with mobility values of 1.617, 1.622 and 1.628 $cm^2V^{-1}s^{-1}$, respectively, measured at a drift tube temperature of 216ºC(Tadjimukhamedov et al., 2008). These were identified as analogous to the pheoxide ion observed in phenol, i. e. in the present work the peak at 1.64 $cm^2V^{-1}s^{-1}$was assigned to trichlorophenoxide (C6H2Cl3O-) probably formed by loss of the methyl group. Other peaks in the negative ion mobility spectra were observed with reduced mobility values of 1.48 $cm^2V^{-1}s^{-1}$and 1.13 $cm^2V^{-1}s^{-1}$. The former was assumed to be an adduct between a TCA molecule and a chloride ion and the latter a chloride bridged dimer ion. These assignments are based on the fact that aromatic compounds in general, like molecules of aromatic explosives, tend to form such adducts with negative ions under conditions that prevail in the IMS drift tube(Lawrence et al., 2001). These assignments are supported by the fact that when dichloromethane is used as a dopant the intensity of the peak at 1.48 $cm^2V^{-1}s^{-1}$ assigned to the chloride adducts increases relative to the peak at 1.64 $cm^2V^{-1}s^{-1}$ that was attributed to the phenoxide species.

Calibration of the IMS system for 2,4,6-TCA and the limit of detection

A calibration curve was prepared for 2,4,6-TCA dissolved in dichloromethane and deposited on a piece of filter paper placed in a headspace vial that was sealed and heated before measurement. The spectra were processed according to the procedure described above to improve the quality of the quantitative information. The diagram of blocks was shown in section 5.2.3., and Figure 5.9.

Figure 5.16 (a) shows raw spectra of TCA at same concentration, but measured at different days. Besides a baseline correction is needed, the misalignment is evident, especially in the peak related to chloride ion (2.69 $cm^2V^{-1}s^{-1}$). After the preprocessed methodology was applied to spectra, a better alignment of the peaks was achieved, as it is shown in Figure 5.16 (b), even though a slightly misalignment is still observed in the small peaks. The alignment of peaks is quite crucial before applying blind source separation technique due to the algorithm might consider a new compound a peak that indeed is misaligned.

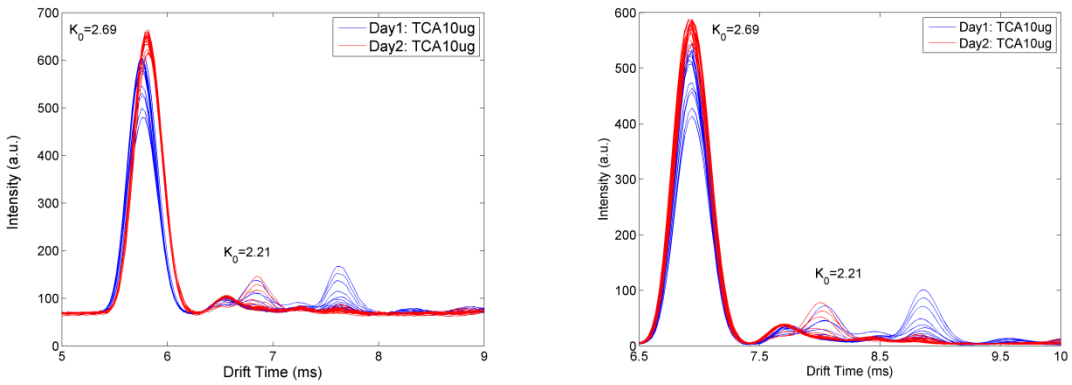


Figure 5. 16 (a) Negative raw spectra of TCA measured at two different days; and (b) Negative spectra of TCA after preprocessing strategy was applied.

Once the whole measurement was properly aligned, SIMPLISMA and MCR-LASSO were applied one after the other. The number of pure variables was selected by visual inspection so that the monomer and dimmer can be extracted from the whole matrix. At the end the number of pure compounds was set at six as it is shown Figure 5.17 (a). In solid line is shown the monomer and dimer of TCA that was recovered by the algorithm and in dashed line is shown the other compounds that the algorithm recovered such as chloride ion (dopant) and RIN. The monomer and dimer are not represented by a unique peak, it might be due to small peaks vary in the same way as TCA peaks. The concentration profile with more intensity belongs to RIN and Chloride ion, as it is shown in Figure 5.17 (b). As soon as the sample is introduced into the IMS, the intensity of RIN decrease whiles the intensity of chloride ion increase. The concentration profile of the monomer and dimer of TCA is shown in Figure 5.17(c) together with the other ions. Three different concentrations of TCA are represented in the figures. When the sample is measured, the monomer increases rapidly until reach a maximum value and then decrease. While, the dimer increase more exponentially until it reaches a stationary phase at the end of the measurement. This is the typical behavior when a monomer and dimer is presented in a sample and they are measured with IMS(Eiceman and Karpas, 2005). It is also remarkable when concentration of TCA increases the intensity of the monomer also increases, but the intensity of the dimer does not do. Maybe, it is

due to as the headspace vapor is carried from the vial to the IMS the concentration first increases, reaches a maximum after 5 to 9 seconds and then decreases as the vapor is diluted by the carrier stream. Therefore, the monomer is more sensible with the increase of concentration than the dimer.

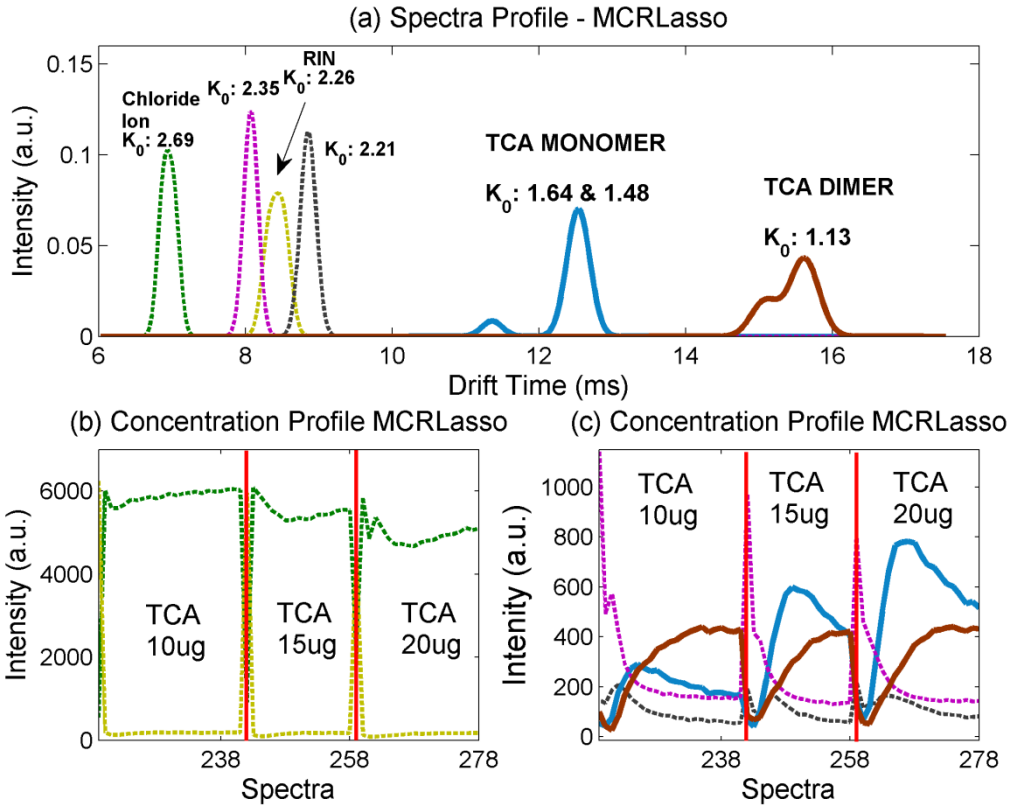


Figure 5.17 MCRLasso results of TCA samples. (a) Spectra profile of Samples; (b) Concentration profile of the RIN and Chloride Ion; and (c) Concentration profile of monomer and dimer of TCA.

As it was explained above, in order to perform the calibration, the concentration profiles from the results of the MCRLasso has to be sort in such a way of having a matrix that represent samples by the evolution of the monomer together with the evolution of the dimer of TCA. Thus the final matrix is going to have a dimension of 15 x 26(13 spectra of monomer + 13 spectra from dimer). Figure 5.18(a) depicts how the final matrix looks like after it has been sorting out. As can be seen, a new alignment is needed, so the maximum intensity of the monomer has to be around the same location in whole samples. Actually, the misalignment is due to the lack of precision when the sample is measured because it was performed manually. In addition a baseline correction is needed due to the background of the instrument change from time to time, and also smoothness of the signal is needed. In Figure 5.17 (b) depicts the final results after applying this extra processing to the signal. The shapes of both signals are cleaner than the original, and when the baseline was corrected, the changes in concentration can be seen in a better way.

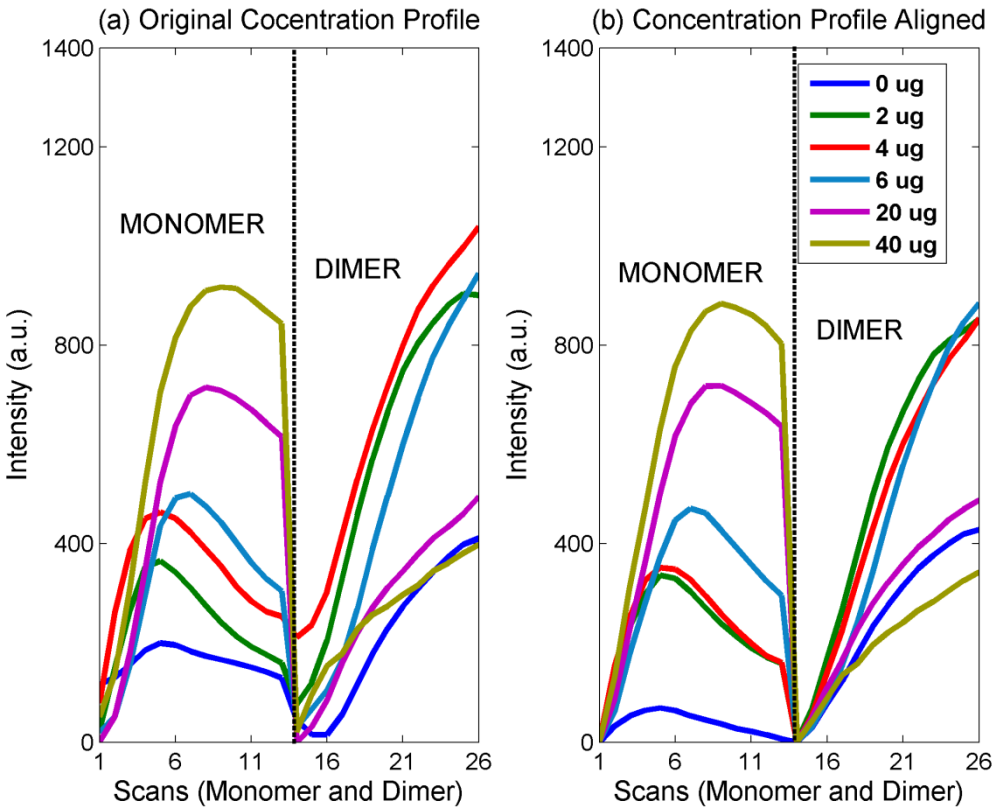
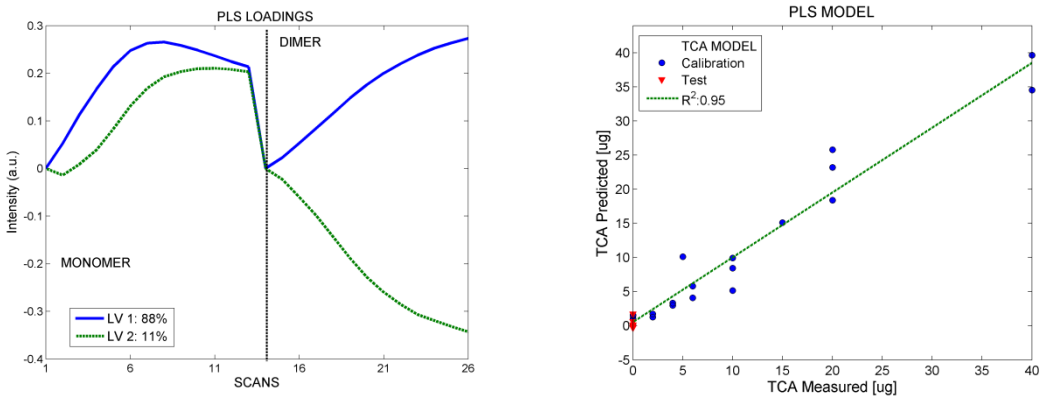


Figure 5.18 Concentration Profile for calibration. (a) Original concatenated concentration profile from MCRLasso; (b) Concentration Profile Aligned and smoothed.

PLS model was built using the concentration profile and the number of latent variables was determined using “leave on out” cross-validation procedure. The final latent variables were established at 2 that recovers 99% of the total variance of the data. The loadings of the PLS model can be seen in Figure 5.19(a) in which the evolution of both monomer and dimer change and have an effect over the model. A plot of the predicted concentrations against the real values can be observed in Figure 5.19(b). The root mean square error in cross-validation was 1.4 µg, and the R2 was 0.95.

In this case, the limit of quantification and detection was calculated using the predicted values of blanks, which were calculated projecting the blanks over the calibration curve, because there are not enough replicates to estimate the confidence band of the calibration curve accurately. At the end, the limit of quantification was 4.3µg and the limit of detection was found to be 1.7 µg of 2,4,6-TCA deposited from a dichloromethane solution on a piece of filter paper placed in a headspace vial.



(a) (b)

Figure 5.19 (a) Loadings from PLS model, and (b) Calibration curve

Relative Sensitivity for TCA dissolved in dichloromethane, ethanol and wine

The relative sensitivity of the detection system for 2,4,6-TCA dissolved in dichloro-methane, ethanol and wine can be assessed from measurements of TCA deposited on filter paper in a headspace vial. The relative signal intensities in positive and negative mode are summarized in Table 5.5, and evidently the sensitivity decreases in the order DCM>Ethanol>wine. The relatively low sensitivity for TCA in wine could be in part due to the long time allowed for drying of the sample that could have also resulted in loss of some of the TCA in the spike. It should be noted that several new peaks appear in the positive and negative mobility spectra of the blanks and spiked wine samples.

The relative recovery efficiency can be derived from these measurements. Thus, if we assume that the recovery of TCA from dichloromethane solution is unity then recovery from ethanol solution, white wine and red wine would be 56%, 7% and 9%, respectively, on average for the three main ion species.

The dichloromethane dopant increased the sensitivity of the system in negative mode and hardly affected the signal intensity in positive mode. In the present system the sensitivity is practically doubled with the addition of the dopant, which is reflected in the intensity of the signals of the ions at 1.48 and 1.13 $cm^2V^{-1}s^{-1}$.

Sensitivity (μV/μg)	**Positive Spectra at K_0:1.58**	**Positive Spectra at K_0:1.64**	**Positive Spectra at K_0:1.48**
Red wine spiked with 375 μg TCA	45(8%)	95(13%)	47(5.6%)
White wine spiked with 375 μg TCA	44(8%)	28(4%)	77(9%)
58 μg TCA in ethanol	450(78%)	470(65%)	200(24%)
60 μg TCA in CH_2Cl_2 (DCM)	580	720	840

Table 5.5 The relative sensitivity of the GDA2 to 2,4,6-trichloroanisole dissolved in dichloromethane, ethanol and wine and deposited on filter paper in a heated headspace ial. The recovery efficiency relative to TCA in dichloromethane solution is shown in parenthesis.

Conclusions

This work presents a discussion of the gas phase ion chemistry pertaining to ion mobility spectrometry measurements of 2,4,6-trichloroanisole in positive and negative modes. In positive mode two ionic species were attributed to the protonated monomer and dimer, and in negative mode a trichlorophenoxide ion as well as a monomer and dimer formed through chloride ion attachment were observed. The reduced mobility values of these ions in air at 44ºC are reported here for the first time. The experimental set up can perhaps be improved by heating the tubing between the sample vial and the IMS inlet port, although there was no evidence that absorption of TCA vapor on the tubing played a role.

An advanced signal processing technique was used to improve the quality of the data. On the one hand, MCRLasso was used to extract the pure compounds from TCA in order to get concentration profiles for subsequently calibration. In addition, MCRLasso returns a more clean spectra profile for each compound present in the sample. On the other hand, the proper pre-processing of the data allows having an accurate estimation of the limit of detection of TCA. Calibration curves were prepared and the limit of detection of the system was determined to be 1.7 µg for a sample dissolved in dichloromethane and deposited on filter paper. This limit of detection is worse by several orders of magnitude to the limit of detection reported recently (Marquez-Sillero et al., 2011a, Marquez-Sillero et al., 2011b). However, a close examination of the mobility spectra displayed in those reports shows that the calculation of the LOD was based on pre-concentration and pre-separation of the TCA and on measurement of the chloride ion while in the present work an ion species that arises specifically from the 2,4,6-TCA analyte was used for the LOD calculation and the IMS was operated as a stand-alone device.

Determination of 2,4,6-trichloroanisole in wine requires pre-concentration (enrichment) and pre-separation and a sensitive analytical device for measuring the signal intensity. The present work did not address the techniques for pre-treatment of wine samples and focused on the potential for using ion mobility spectrometry as the measurement device. The limit of detection found here would require a substantial enrichment factor, especially considering that the "off flavor" attributed to TCA is apparent at levels below 10 ng L^{-1}.

5.3.2. Feasible study for measurement potential biomarkers of prostate cancer using Ion Mobility Spectrometry.

VOCs have been studied as diagnostic and screening tool for monitoring and identification of different kinds of cancer(Sethi et al., 2013, Issaq et al., 2011, Phillips et al., 2010, Evans et al., 2009, Westhoff et al., 2009, Kind et al., 2007). According to American Cancer Society, prostate Cancer (PCa) is the second leading cause of cancer death in American men, behind only lung cancer (AmericanCancerSociety, 2014). The study of VOCs in urine has received scarce attention than breath samples. One of the major studies was undertaken by Mills and Walker (Mills and Walker, 2001) in which 103 compounds where found from 5 heterogeneous patients at different conditions.

In a preliminary study, an analysis to discriminate control subjects out of patients with PCa through urine headspace analysis in GC/MS was carried out. As preliminary result, a compound, which will be referred to as *Compound A* for the rest of de document, was found as responsible for the discrimination between control and PCa patients. This compound seems to appear mainly in patients with PA than controls as it is shown in Figure 5.20. Even though this results came from a preliminary study, it is interesting to test if IMS can be used for detecting this compound in similar conditions. Although, the GC/MS results were not quantitative, the IMS study will be carried out to establish if the spectrometer is able to measure the compound together with the limit of detection of the instrument in presence of the compound.

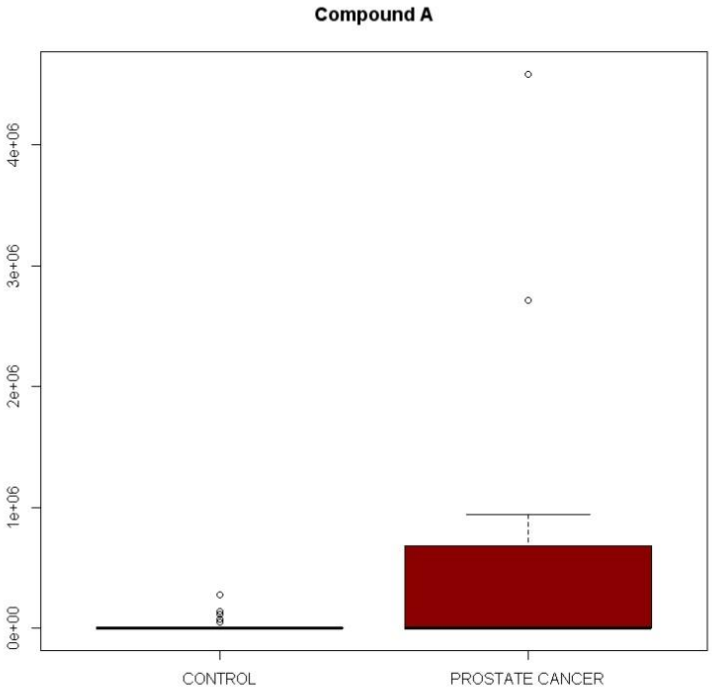


Figure 5. 20 Boxplot of Compound A of 32 control subjects and 20 patients with prostate cancer. The compound was analyzed by head-space GC/MS.

The quantitative analysis was done in three stages. The first one consisted of analyzing the pure compound at different concentration using permeation tubes for controlling the amount of the compound. Then, a solution was prepared in a permeation tube using the same range of concentrations of *compound A* mixed with water. The last one was a head-space analysis in which the compound was spiked in water to simulate a real scenario. These three stages were done using two spectrometers GDA2 and UV-IMS.

The IMS spectra of the pure *Compound A,* measured with the UV-IMS instrument are shown in Figure 5.21(a). The main peak, which reduced mobility coefficient is 1.2 $cm^2V^{-1}s^{-1}$, is the monomer of analyte because at lower concentrations is the only peak that appears. The intensity of the peak of the monomer increases as the concentration rise until a maximum value. At that point, higher concentrations, favor the formation of new peaks (dimmer or trimmers) and the intensity of the monomer decreases. As it has been explained in chapter 3 this is the typical behavior for IMS instruments.

It is well known that the sensibility of the UV-IMS decreases in presence of humidity, which is the case of the solution of Compound A with water, following the previously explained stage 2 measurements. Results of these measurements can be seen in Figure 5.21(b) where similar concentrations were measured but with added humidity. In fact, in the measurements with humidity, there is no significant difference between spectra at 0.8 ppm and 2 ppm (Figure 5.21 (b)). In contrast, in the first stage measurements, there is a significant difference between spectra at 0.5 ppm and 2 ppm in Figure 5.21 (a). This confirms that the spectrometer is going to lose sensitivity when the experiment is done under humidity conditions. Nevertheless, this can be diminished, if a humidity filter is set at the inlet of the instrument.

The response of the instrument when the experiment was done when the compound was diluted in water and analyzed by head-space is shown in Figure 5.21(c). The spectra at lower concentrations have mostly the monomer, but when the concentration increases other peaks emerges and its intensity increases further. It can be dimer or other cluster formation between the compound and water impurities.

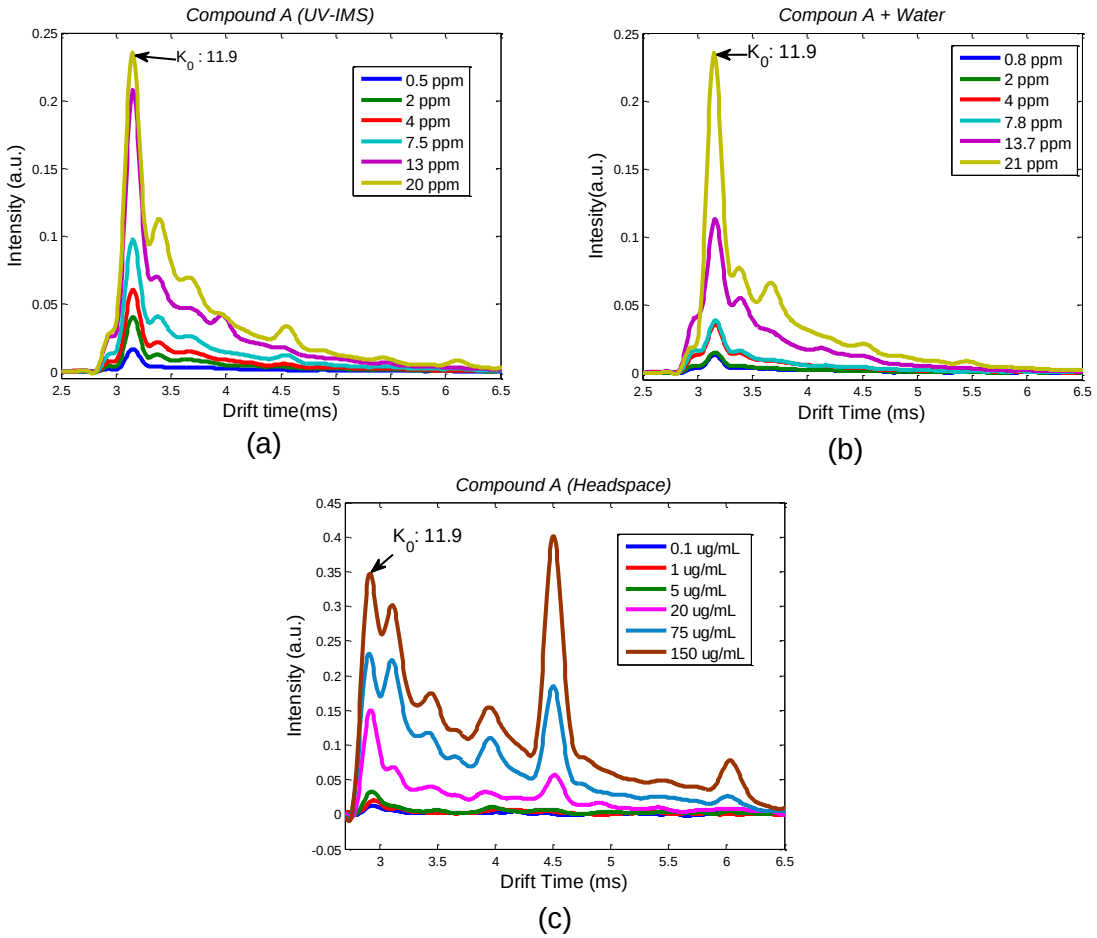
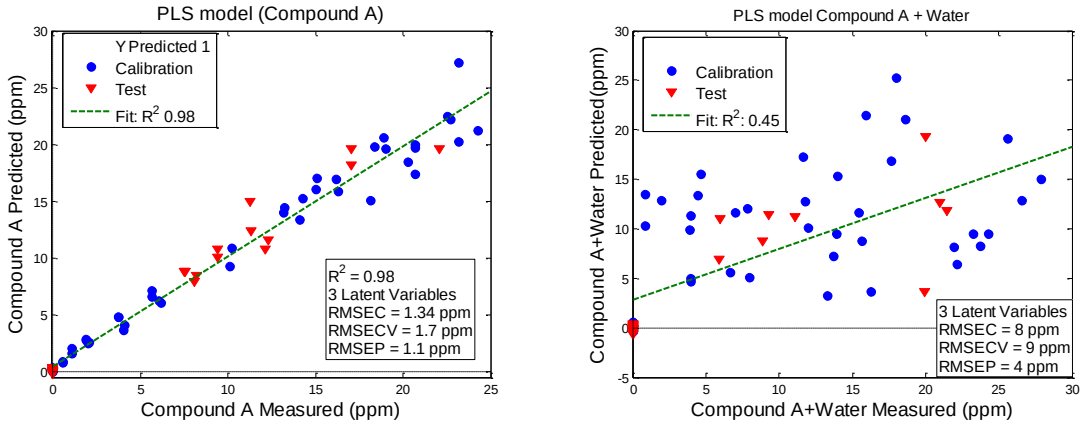


Figure 5. 21 Spectra of *Compound A* at different concentrations analyzed with UV-IMS. (a) Pure compound, (b) pure compound diluted in water, and (c) headspace analysis.

Despite of the fact that the experiment was done without any mixture, the resultant spectra was more complex than having a unique peak of the monomer, as it was initial expected. Thus, multivariate calibration models seem to be a good choice for quantitative results. A PLS model were done using the whole spectra of dataset from Figure 5.21 (a) and (b). In the case of headspace, MCRLasso was performed for

extracting the pure components and the concentration profile which was after used for building a PLS model. Leave one concentration out was used as cross validation strategy in order to determine the number of latent variables, and a subset of blanks and samples at intermediate concentration were left out for testing the model.

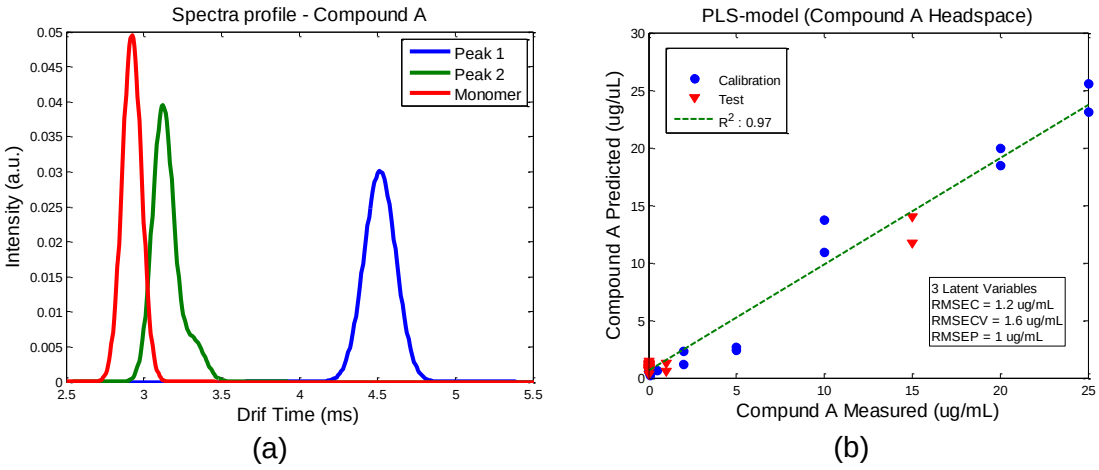
Calibration models are shown in Figure 5.22 (a) and (b) for pure compound and solution with water, respectively. The UV-IMS can detect easily small amounts of concentration in experimental conditions, as it can be seen in Table 5.6. Also, the performance get worse when the compound is mixed with water, actually, the RMSEP increases significant compared to model in Figure 5.22 (a). Thus, the LOD is 0.18 ppm in experimental conditions, but it gets larger when the compound is measured with humidity. Therefore, in a real scenario with a more complex background, it is expected that LOD will be worse.



(a) (b)

Figure 5.22 PLS models of Compound A (a) Pure compound (b) Compound diluted in water.

The spectra profile and concentration profile obtained by doing MCRLasso are shown in Figure 5.23 (a) and (c) respectively. The spectra shows three main pure compounds with a 90% of explained power, one of them is the monomer of the analyte and the other two peaks can be linked to either dimer (peak 3) or a cluster formation of the monomer (peak2). Actually, the intensity of the concentration profile of peak 2(Figure 5.23 (c)) depict an increment as the concentration rise. Moreover, it can be seen that it is need the use of the information of the three peaks in order to get more reliable results than choosing just one of them. Note, there is a small misalignment in the evolution of the peaks that can be attributable to experimental error in the sample introduction, thus it is important to perform an alignment before build any calibration model in a similar way than before (see Figure 5.18). A PLS model was built using a range of concentration from 0 to 20 µg/ml in order to avoid any over fitting in the LOD calculation.



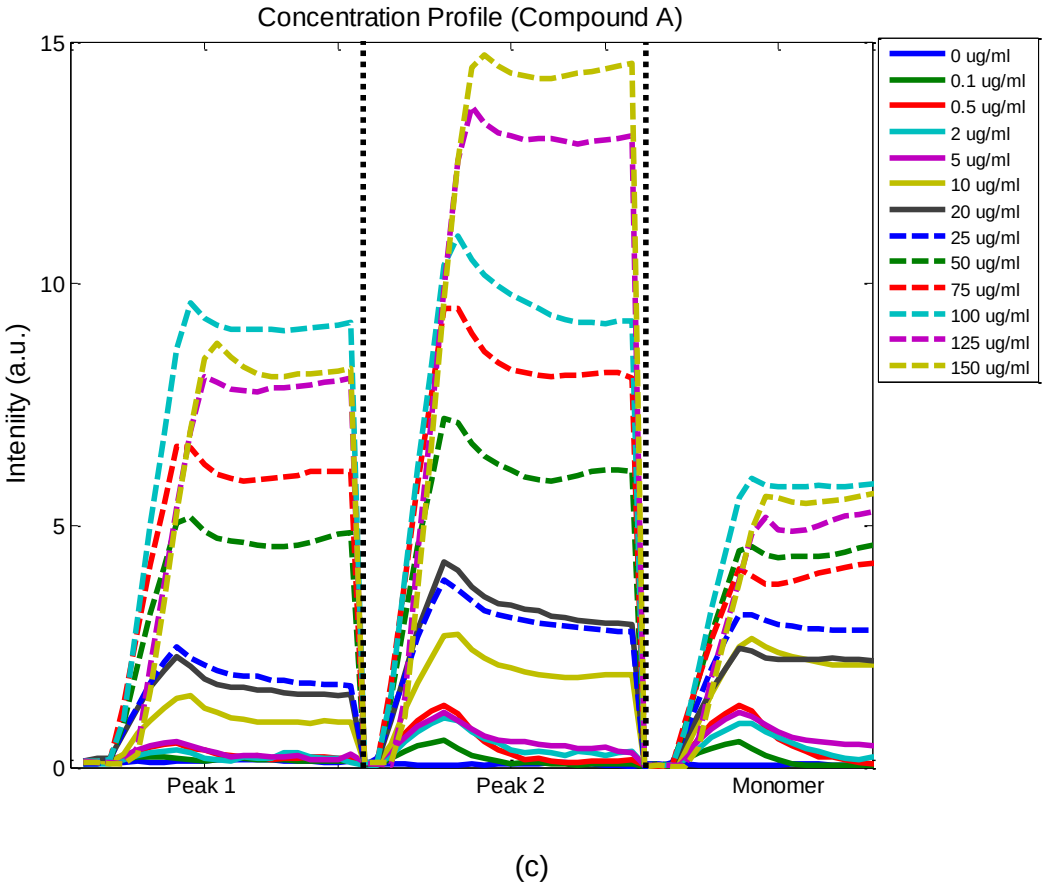


Figure 5.23 (a) Spectra profile of compound A, (b) PLS model , (c) Concentration profile.

The result of PLS model is depicting in Figure 5.23(b). A set of blanks and intermediate concentration samples were left out to the test the predictive power of the model. The LOD get in this case was 1.3 µg/ml. This LOD seems apparently to be quite high because in the monomer response (Figure 5.23 (c)) there is a difference between blank and 0.1 µg/ml. However, there are just few calibration samples in training set and the lack of reproducibility is also another factor to be taken into account when LOD is calculated. The quantitative results are summarized in Table 5.6

Quantitative results UV-IMS	R^2	RMSECV	RMSEP	LOD
Pure compound	0.98	1.7 ppm	1.1 ppm	0.18 ppm
Pure compound dilute in water	0.45	9 ppm	4 ppm	3 ppm
Head space	0.97	1.6 µg/ml	1 µg/ml	1.3 (µg/ml)

Table 5.6 Quantitative results of Compound A in UV-IMS. RMSEC: root mean square error of cross-validation. RMSEP: root mean square error of prediction.

The compound A was analyzed with GDA2 and the response of this spectrometer is shown in Figure 5.24. The analyte has a mobility coefficient of 2.07 $cm^2V^{-1}s^{-1}$. Note that the main difficulty in analyzing this analyte is the compound A appears at the tail of reactant ion peak (RIP). In fact, the peak of the analyte was observed at very high concentrations where the RIP practically disappears. This means that any change in the tail of the RIP could be associated either to this compound or any other compounds in the sample. Another option is to have a pre-concentration step before the analysis in order to enhance the sensitivity of the compound and be able to observe the peak of the analyte. It is really interesting how the reduced mobility coefficients for both spectrometers are extremely different. This can be attributed to the physical differences between these two spectrometers and/or the chemical properties of the analyte of interest. In any case, this consideration has to be taking into account for further studies.

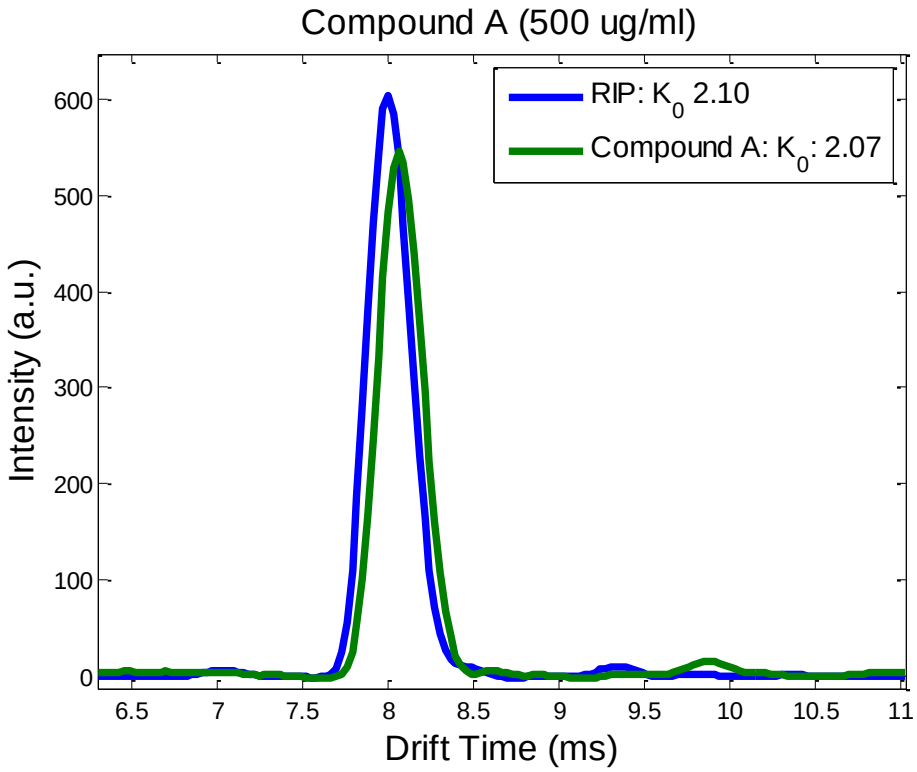


Figure 5. 24 Two spectra of RIP (blue line) and compound A (green line).

Despite of the fact that measurements using permeation tube with pure compound and solution in water was performed, the compound A peak vanishes in presence of RIP at low concentration. That is the main reason why the analysis of the LOD was not feasible to do for the two first stages in similar way than UV-IMS. In any case, the headspace analysis was done with a wider range of concentration that favors the peak formation and the subsequent analysis.

Figure 5.25 shows spectra profile and concentration profile obtained by doing MCRLasso. It can be seen that the technique was able to extract these two pure components from the matrix, despite of the high overlapping of the two peaks (RIP and compound A). The concentration profile is also shown for a specific concentration (100 µg/ml). In this case, it can be seen how there is an increment of the signal when the analyte is injected as the intensity of the RIP drops. Then the intensity of the RIP recover the previous value as the compound A decrease the signal, this behavior just take few seconds .

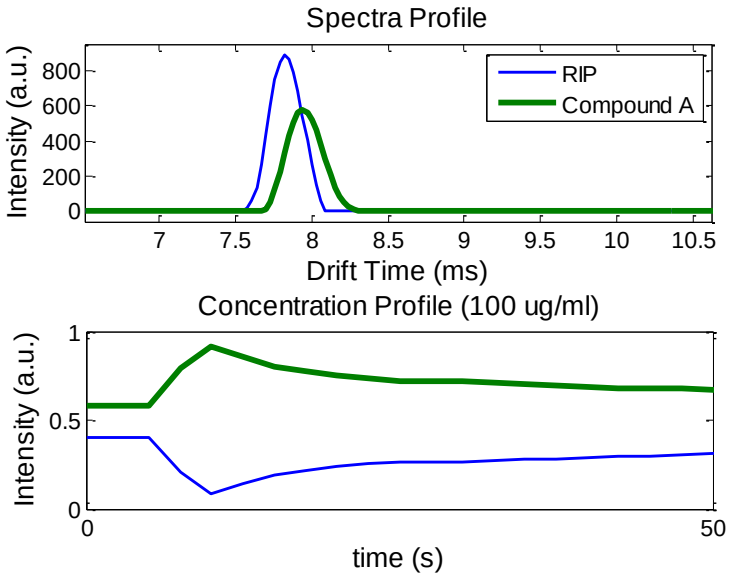
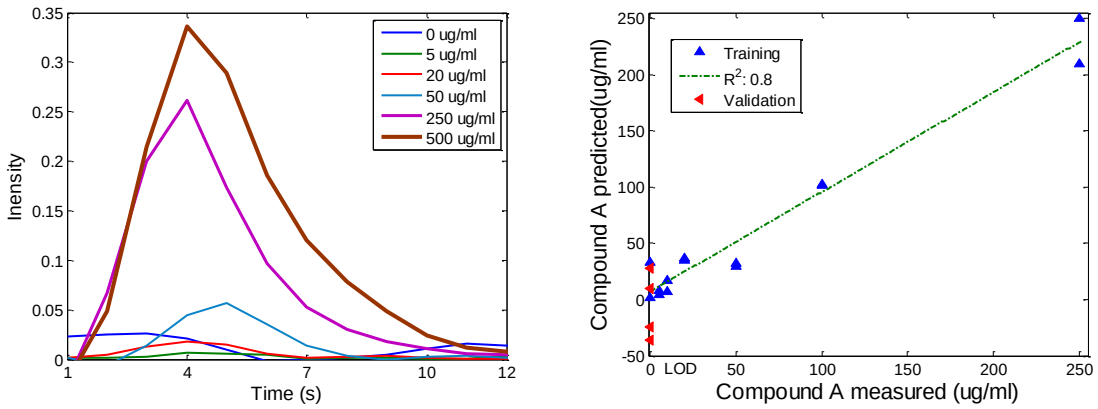


Figure 5. 25 Spectra profile and concentration profile of compound A.

In order to calculate the LOD of the IMS, the concentration profiles from different concentrations are gathered as it can be seen in Figure 5.26(a). Clearly, the lowest concentration is almost close to the noise level of the spectra and is necessary to have at least 50 µg/ml to observe a response of the instrument higher than the noise level. A pls model was built using the information of the evolution of the compound during the time. The final calibration model is shown in Figure 5.26(b) in which a set of blanks were left out to calculate the LOD of the compound A giving a result of 46 µg/ml. The number of latent variables, which was estimated by leave one concentration out methodology, was 2 with a RMSECV of 30 µg/ml and RMSEP of 40 µg/ml.



(a) (b)

Figure 5. 26 a) Concentration profile of compound A for different concentration ranges. (b) PLS model for LOD calculation.

Conclusions

If the question is if ion mobility spectrometry can be used for analyze this analyte (compound A), the answer should be positive and the best spectrometer might be the UV-IMS. However, the final goal is to test this compound under more real conditions in a more complex matrix such as the urine. It has been seen that GDA2 spectrometer has a strong limitation due to the localization of the monomer in the spectra. The fact of being in the tail of the RIP brings too much complication for the subsequent analysis. Apart from the preprocessing that is in fact quite challenging, the possibility to extract information which might not be correlated with any other compound in the matrix becomes the viability of the analysis of this compound in urine almost unfeasible. In case that this compound is representative of patients that have PA, as appears by the analysis with GC/MS, the GDA2 might be only used with a pre-concentration or pre-separation sampling technique set up before the spectrometer.

The results with UV-IMS are really promising because the limit of detection is quite reasonable. The limitation of this instrument to be used in this particular application is bound by the inner limitation of the spectrometer. The main drawback is the low performance under humidity condition, which is the case of this application, due to the sensibility of the spectrometer diminishes as the humidity increase. Other factor is that the spectrometer do not have any temperature control inside drift tube and the temperature work operation is directly related to weather conditions. Thus, the formations of cluster are not accurately controlled and the K_0 cannot be correctly established, so there is a need of a known substance that works as calibrant, which can also be useful for preprocessing steps and at least an external temperature control. The last remark to consider is the complex spectra that were obtained for a single compound, and the complexity of the spectra will be greater as the complexity of the matrix. In this case, the use of a proper signal processing will help for a better understanding and enhancing the results. The best option could be the use of MCR techniques because it allows extracting only the information of the peak of interest, and work with the concentration profile as quantitative or semi-quantitative information.

This is a preliminary work to test the viability of use IMS as analytical technique to be used as analyzer for detect a specific compound in urine samples. Despite of the fact that the compound was not tested in a real matrix, the study shows up some

challenges that should be face up to before any attempting a real analysis. There has to be consider including instrumental improvements such as pre-concentration or pre-separation techniques, temperature and humidity controls, etc. In case of the compound, which has been studied in this work, is confirmed as potential biomarker of prostate cancer, a signal processing has to be established in order to enhance the instrument selectivity and avoid interferents of the sample. Certainty, the use of multivariate technique has to be almost compulsory due to the complexity of the kind of samples, and MCR techniques would be the best option in this application.

5.4. Summary

The content of this chapter attempts to remark the importance of the use of multivariate technique in the analysis of IMS spectra. Indeed, univariate techniques has been the goal standard in IMS spectra analysis, but nowadays scientific community are paying more attention in signal processing strategies which allows a better understanding and reliable results. Moreover, the applications have diversified that implies more complex samples and spectral information to extract, thus univariate techniques are no longer suitable for this kind of applications.

This chapter has been divided into different objectives. The first one is the study of non-linear behaviours of IMS from a quantitative point of view. The use of multivariate techniques comparing with univariate techniques and the suitability of these techniques in different scenarios has been discussed. Regarding multivariate calibration models, it has been tested whether the use of whole spectra or the use of multivariate curve resolution techniques as possible solutions in quantification problems. In the last section, the same algorithms and strategies for multivariate calibration was tested in real applications for proving its usefulness.

It was confirmed that the use of univariate techniques gives poor results since do not get all profit of the relevant information. Moreover, the fact of using univariate techniques is a limitation in complex applications because the interaction with other compounds is not taking into account. This miss usually offers overoptimistic or erroneous results. Indeed, the use of multivariate techniques provide better results than univariate results, also provide a better interpretability of the results.

The main difference between using whole spectra and extracting pure compounds the ability of extracting each compound present in a complex background. Note, MCR techniques allow getting specific compounds for being used afterwards using other analytical algorithms. However, many times the analysis of the whole spectra provides slightly better quantitative results. When real applications were studied, additional issues has to be taking into account but are more related with performing an accurate pre-processing of the signal. Once these problems are solved, the use of multivariate techniques makes easier the analysis and the results get reliable and accurate.

In this thesis, different strategies have been proposed for the analysis of IMS spectra. These strategies have also point out the importance of the use of proper validation methodologies in order to confirm results.

This work has been mainly focused in the spectral analysis when IMS is used as standalone device. Nevertheless, many of this work can be extrapolate when IMS is coupled to other analytical instruments, but it is out of the scope of this thesis. The usefulness of IMS in biorelated context has been tested giving positive and promising results. Of course, each application has its own challenges, but with a proper signal processing strategy can provide reliable results.

5.5. Reference

ABCScience. 2013. ***Corked wine confuses smell receptors*** [Online]. Available: http://www.abc.net.au/science/articles/2013/09/17/3850029.htm 2014].

Airsense. 2012. *GDA2, Airsense Germany* [Online]. Available: www.airsense.com/en/products/gda-2/.

AmericanCancerSociety. 2014. *What are the key statistics about prostate cancer?* [Online]. Available: http://www.cancer.org/cancer/prostatecancer/detailedguide/prostate-cancer-key-statistics.

Cao, L. B., Harrington, P. D. & Liu, J. D. 2005. SIMPLISMA and ALS applied to two-way nonlinear wavelet compressed ion mobility spectra of chemical warfare agent simulants. *Analytical Chemistry,* 77**,** 2575-2586.

de Juan, A., Maeder, M., Martinez, M. & Tauler, R. 2000. Combining hard- and soft-modelling to solve kinetic problems. *Chemometrics and Intelligent Laboratory Systems,* 54**,** 123-141.

Eiceman, G. A. & Karpas, Z. 2005. *Ion Mobility Spectrometry,* Florida, Taylor & Francis Group.

Evans, C. A., Glen, A., Eaton, C. L., Larre, S., Catto, J. W. F., Hamdy, F. C., Wright, P. C. & Rehman, I. 2009. Prostate cancer proteomics: The urgent need for clinically validated biomarkers. *Proteomics Clinical Applications,* 3**,** 197-212.

Fraga, C. G., Kerr, D. R. & Atkinson, D. A. 2009. Improved quantitative analysis of ion mobility spectrometry by chemometric multivariate calibration. *Analyst,* 134**,** 2329-2337.

Harrington, P. D., Reese, E. S., Rauch, P. J., Hu, L. J. & Davis, D. M. 1997. Interactive self-modeling mixture analysis of ion mobility spectra. *Applied Spectroscopy,* 51**,** 808-816.

Holmberg, L. 2010. Wine Fraud. *International Journal of Wine Research,* 2**,** 105–113.

iCEX, V. 2014. *El Vino en Cifras – El Vino en Cifras –Año 2014* [Online]. Available: http://www.winesfromspain.com/icex/cma/contentTypes/common/records/mostrarDocumento/?doc=4779156 [Accessed 2014 2015].

Issaq, H. J., Waybright, T. J. & Veenstra, T. D. 2011. Cancer biomarker discovery: Opportunities and pitfalls in analytical methods. *Electrophoresis,* 32**,** 967-975.

Karpas, Z., Bell, S. E., Wang, Y. F., Walsh, M. & Eiceman, G. A. 1994. THE STRUCTURE OF PROTONATED DIAMINES AND POLYAMINES. *Structural Chemistry,* 5**,** 135-140.

Karpas, Z., Chaim, W., Gdalevsky, R., Tilman, B. & Lorber, A. 2002a. Novel application for ion mobility spectrometry: diagnosing vaginal infections through measurement of biogenic amines. *Analytica Chimica Acta,* 474**,** 115-123.

Karpas, Z., Guaman, A. V., Pardo, A. & Marco, S. 2013. Comparison of the performance of three ion mobility spectrometers for measurement of biogenic amines. *Analytica Chimica Acta,* 758**,** 122-129.

Karpas, Z., Tilman, B., Gdalevsky, R. & Lorber, A. 2002b. Determination of volatile biogenic amines in muscle food products by ion mobility spectrometry. *Analytica Chimica Acta,* 463**,** 155-163.

Kind, T., Tolstikov, V., Fiehn, O. & Weiss, R. 2007. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal Biochem,* 363**,** 185 - 195.

Lawrence, A. H., Neudorfl, P. & Stone, J. A. 2001. The formation of chloride adducts in the detection of dinitro-compounds by ion mobility spectrometry. *International Journal of Mass Spectrometry,* 209**,** 185-195.

Marcus, S., Menda, A., Shore, L., Cohen, G., Atweh, E., Friedman, N. & Karpas, Z. 2012. A novel method for the diagnosis of bacterial contamination in the anterior vagina of sows based on measurement of biogenic amines by ion mobility spectrometry: A field trial. *Theriogenology,* 78**,** 753-758.

Marquez-Sillero, I., Aguilera-Herrador, E., Cardenas, S. & Valcarcel, M. 2011a. Determination of 2,4,6-tricholoroanisole in water and wine samples by ionic liquid-based single-drop microextraction and ion mobility spectrometry. *Analytica Chimica Acta,* 702**,** 199-204.

Marquez-Sillero, I., Cardenas, S. & Valcarcel, M. 2011b. Direct determination of 2,4,6-tricholoroanisole in wines by single-drop ionic liquid microextraction coupled with multicapillary column separation and ion mobility spectrometry detection. *Journal of Chromatography A,* 1218**,** 7574-7580.

Metro, M. M. & Keller, R. A. 1973. FAST SCAN ION MOBILITY SPECTRA OF DIETHYL, DIPROPYL, AND DIBUTYL ETHERS AS DETERMINED BY PLASMA CHROMATOGRAPH. *Journal of Chromatographic Science,* 11**,** 520-524.

Mills, G. A. & Walker, V. 2001. Headspace solid-phase microextraction profiling of volatile compounds in urine: application to metabolic investigations. *J Chromatogr B Biomed Sci Appl,* 753**,** 259-68.

Phillips, M., Cataneo, R. N., Saunders, C., Hope, P., Schmitt, P. & Wai, J. 2010. Volatile biomarkers in the breath of women with breast cancer. *Journal of Breath Research,* 4**,** 8.

Pomareda, V., Calvo, D., Pardo, A. & Marco, S. 2010. Hard modeling Multivariate Curve Resolution using LASSO: Application to Ion Mobility Spectra. *Chemometrics and Intelligent Laboratory Systems,* 104**,** 318-332.

Sethi, S., Nanda, R. & Chakraborty, T. 2013. Clinical Application of Volatile Organic Compound Analysis for Detecting Infectious Diseases. *Clinical Microbiology Reviews,* 26**,** 462-475.

Sobel, J. D., Karpas, Z. & Lorber, A. 2012. Diagnosing vaginal infections through measurement of biogenic amines by ion mobility spectrometry. *European Journal of Obstetrics & Gynecology and Reproductive Biology,* 163**,** 81-84.

Spangler, G. E. 2002. Expanded theory for the resolving power of a linear ion mobility spectrometer. *International Journal of Mass Spectrometry,* 220**,** 399-418.

Tadjimukhamedov, F. K., Stone, J. A., Ppanastasiou, D., Rodriguez, J. E., Mueller, W., Sukumar, H. & Eiceman, G. A. 2008. Liquid Chromatography/electrospray ionization /ion mobility spectrometry of chlorophenols with full flow from large bore LC columns. *International Journal of Ion Mobility Spectrometry,* 11**,** 51-60.

Westhoff, M., Litterst, P., Freitag, L., Urfer, W., Bader, S. & Baumbach, J. I. 2009. Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of patients with lung cancer: results of a pilot study. *Thorax,* 64**,** 744-748.

Windig, W. & Guilment, J. 1991. INTERACTIVE SELF-MODELING MIXTURE ANALYSIS. *Analytical Chemistry,* 63**,** 1425-1432.

Wold, S., Kettanehwold, N. & Skagerberg, B. 1989. NONLINEAR PLS MODELING. *Chemometrics and Intelligent Laboratory Systems,* 7.

Zamora, D., Alcala, M. & Blanco, M. 2011. Determination of trace impurities in cosmetic intermediates by ion mobility spectrometry. *Analytica Chimica Acta,* 708**,** 69-74.