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

AGRADECIMIENTOS

"*Si caminas solo irás más rápido; si caminas acompañado llegaras más lejos" – proverbio chino.*

En este largo recorrido tengo mucho agradecimiento a todas las personas que supieron estar junto a mí para llegar donde estoy ahora. A las primeras personas que quiero agradecer es a mis Padres, Edgar y María Elena, definitivamente soy lo que soy gracias a ellos. Su apoyo siempre es incondicional, y siempre están ahí para apoyarme y animarme. Sé que este logro también es de ellos.

El gran apoyo, en este camino y espero que en muchos más, es Andrés A. Juntos nos embarcamos en un camino muy interesante y enriquecedor a todos los niveles. En todo este tiempo nos apoyamos mutuamente, muchas gracias por entenderme, por estar ahí siempre y por darme ánimo hasta el último momento. Solo los dos sabemos que esta vivencia estuvo llena de tantas aventuras, tantos recuerdos que no olvidaremos, tantas experiencias que ya son historias. Muchas gracias amor sin ti no hubiera sido lo mismo.

Mis hermanas, María Elisa y Cristina, y mi hermano Edguitar, ellos me conocen muy bien y saben estar ahí aunque no siempre físicamente. Muchas gracias por sus bromas y honestidad al decir que no saben que hago, pero sé que Uds. saben que es importante para mí y eso es suficiente. A mis cuñados, José y Vladi, muchas gracias por sus palabras de ánimo que nunca vienen mal. Y Cris, gracias por esos bellos sobrinos que tengo, me llenan de alegría y me hacen ver el mundo de otra forma. Este proceso, también me hizo entender que todo tiene un precio, simplemente muchas gracias abuelitos.

Junto con Andres, llego mi familia política. Muchas gracias Marcia, por sus palabras y por consentirme siempre. Mire tú te has convertido en una buena amiga y tu apoyo siempre es incondicional. José, Ale, Kami y Jose A., sé que nos tienen en sus oraciones y siempre es bonito saber que hay alguien cuidándonos. Sra. Bachita muchas gracias por todas sus atenciones.

Quiero agradecer a mis directores de Tesis, Dr. Antonio Pardo y Dr. Josep Samitier, los dos me han brindado todo el apoyo necesario para que esta tesis llegue a buen término.

Al grupo que me acogió durante 5 años, muchas gracias Santi y Toni, la verdad que miró hacia atrás y veo todo lo que me he superado y ha sido gracias a ustedes. Santi, tú me diste un voto de confianza que cambio mi vida profesional, espero haberte correspondido y que mi esfuerzo se haya visto reflejado en el trabajo del día a día. Toni, lo primero que pienso es "cap problema", creo que al final es cierto todo tiene una solución, y al final vimos la luz al final del túnel. Te agradezco mucho Toni por todo el apoyo y el esfuerzo que has puesto para que todo vaya bien, sobre todo cuando me quedaba estancada y pensaba que no llegábamos. Como profesional y ser humano me llevo ejemplos muy grandes a seguir.

A todos los que formaron y forman parte del ISP lab, Agusti, Sergi, Idoya, Marta, Erola, Didier, Sergio, JuanMa, Raquel, Rudys, Jordi, Sara y a todos los que seguramente me estoy olvidando. La tradición del café y los chocolates es algo que extraño. Víctor, muchas gracias por tus locuras, y creo que el viaje a Amberes no se me olvidara jamás. Luis, sabes que te tengo mucho cariño y ya te echo de menos, extraño mucho nuestras charlas y darnos ánimo para salir adelante o hundirnos hasta el fondo. Al final es verdad, "lo quiero todo y lo quiero ya". No puedo más que desearte lo mejor del mundo y sabes que puedes contar conmigo. Siempre seremos el poli bueno y la poli mala, para ser el profesor perfecto.

Cuando uno se va lejos de su casa, hay personas que te hacen la vida más sencilla y te hacen sentir en familia. Anita nunca olvidare lo que hiciste por mí cuando llegue a BCN, fuiste y eres una persona maravillosa. Virgi, tu llegaste a alegrarme la vida cuando lo necesitaba, nunca olvidare tu positivismo y alegría, supongo que debo agradecerle a Papa Dios. Victoria, siempre te sentí como mi segunda madre, tu cariño fue tan inmenso al acogernos en tu casa y en tu vida.

También, quiero agradecer a Cristina Davis y su grupo, por la acogida en Davis, California. Esos tres meses que pase en USA me sirvieron mucho en muchos aspectos. Por toda su hospitalidad e integrarme en su grupo. De la misma forma quiero agradecer al Dr. Zeev Karpas, su estadía en BCN le dio un impulso importante a mi trabajo. Muchas gracias Dr. Karpas, porque reconozco en usted una persona muy sabia y generosa por compartir todo su conocimiento. Agradezco también a Rocio y Lourdes de la Universidad de Cordoba, por esa bonita relación de trabajo que se creó y que nos enriqueció mucho.

Por ultimo quiero agradecer al Departamento de Electrónica, por todo el apoyo brindado a mi llegada a la UB, además a todos los eventos creados como calcotadas, picatronics, etc. Al IBEC, para mi fueron un soporte muy importante en todos los tramites que como extranjera tuve que vivir. Ricard, Marta, y Javi, muchas gracias por viabilizar y agilizar todos los procesos tanto internos como externos. También, gracias por los minicursos a los que tuve oportunidad de asistir y que profesionalmente me ayudaron a avanzar.

ABSTRACT

The emerging growth in the use of Volatile Organic Compounds (VOC) for screening and monitoring substances has brought a development of new analytical techniques. There are several applications where the measurement of VOC results to be useful, such as: toxic leaks, air quality measurements, explosive detection, monitoring of food and beverages quality, diagnosis of diseases, etc. Some of this applications claim for fast responses or even real time responses. In this context, there are few analytical techniques for performing gas phase analysis, among of them Ion Mobility Spectrometry (IMS).

IMS is a fast analytical device based on the time of flight of ions in a drift tube. The response of IMS lasts typically few seconds, but it can be even less than a second. This fast response has drifted its use towards novel applications, such as biomedical and food applications (bio-related applications). Nonetheless, it has also brought the need to analyze complex spectra with hundreds of compounds. In fact, tackling this disadvantage is the main focus of this thesis, where new algorithms for enhancing the IMS performance are investigated when are applied to bio-related applications.

Nonlinear behavior and charge competitions of IMS responses are important issues that need to be addressed. Both effects have a direct impact in the IMS spectra interpretation –especially when real dataset are studied. Additionally, the use of univariate spectra analysis, where peaks information is extracted manually, becomes unfeasible in bio-related applications. In this context, this work introduces multivariate methodologies focused on quantitative and qualitative analysis.

In the case of quantitative analysis, calibration models were built using univariate methodology, Partial Leas Squares (PLS) and Multivariate Curve Resolution techniques (MCR). The quantitative analysis aims tackling the main issues of IMS such as non linearities and mixture effect. Definitely, univariate techniques provides poor or overoptimistic results that minimize the impact of the IMS use. The results show a really improvement on the performance when multivariate techniques were used. Regarding the results between MCR and PLS, the main difference is the interpretability that offers MCR.

In the case of qualitative analysis, two different approaches were planned for building models for classes' discrimination. The first approach consisted on building a model through principal component analysis and linear discriminant analysis, besides of using robust cross validation methodology for obtaining reliable results. This methodology were implemented in samples of wine, where main motivation was found discrimination regarding to their origin. The results were fully satisfactory because the model was able to separate four groups with a high accuracy rate. The second approach involves the use of Multivarite Curve Resolution – Lasso algorithm for extracting pure components of samples from rats' breath and then use a feature selection technique for obtaining the most representative features subset. In this case, the objective of the application was to find a model that discriminate rats with sepsis from control rats. The results shows there were few pure components of IMS that generate a discriminatory model that means there are specific compounds in the breath linked with the disease.

Summarizing, the following proposal has as main objective resolving open issues in stand-alone IMS that are applied to the analysis of bio-related applications. Two major investigation lines were proposed in this thesis: (i) qualitative analysis and (ii) quantitative analysis. The qualitative analysis covers pre-processing algorithms and the developing of new methodologies for building models in bio-related applications. The quantitative analysis are focused on highlighting the importance of the use of multivariate techniques instead of univariate techniques. In order to reach the objectives of this thesis, a set of datasets were created, which are detailed on the content of this thesis. The results and main conclusions are deeply explained in the extended proposal.

CONTENTS

Introduction

A. Motivation 1
B. Volatile Organic Compounds (VOCs): An Overview 3
C. Volatile Organic Compounds (VOCs) Analysis 5
i. Enrichment and Sampling technique 5
ii. Analytical Techniques 6
iii. Signal Analysis 9
D. Biomedical and biological application using VOCs analysis 10
E. Summary 11
F. References 13

Objectives

Objectives 17

Chapter One: Ion Mobility Spectrometry as potential technology in biological scenarios

1.1. Introduction 19
1.2. Ion Mobility Spectrometry 19
1.2.1 Ionization source 21
1.2.2. Non-linear behavior of IMS 24
1.2.3. Proton affinity and Dopant Effect 26
1.3. Sampling introduction techniques to IMS 28
1.3.1. Main biological and biomedical applications with Stand Alone IMS 29
1.3.2. Other IMS configurations and applications 30
1.4. Summary 31
1.5. Reference 33

Chapter Two: Quantitative and Qualitative Analysis of Ion Mobility Spectrometry: from univariate to multivariate

2.1. Introduction 39
2.2. Spectral description 39
2.3. Pre-processing 41
2.3.1. Noise reduction 42
2.3.2. Baseline Removal 43
2.3.3. Misalignments 44
2.4. Qualitative Analysis 45
2.4.1. Feature extraction, feature selection and dimensionality reduction 45
2.4.2. Classifiers 47
2.4.3. Qualitative analysis used in IMS. 48
2.5. Quantitative Analysis 49
2.5.1. Univariate Calibration Model 49
2.5.2. Univariate Calibration applied to IMS 50
2.5.3. Multivariate Calibration Model 51
2.5.4. Multivariate Calibration applied to IMS 53
2.6. Self-modeling mixture analysis techniques 55
2.7. LIMIT OF DETECTION 62
2.7.1. Limit of detection applied to IMS 66
2.8. Cross validation methodologies 67
2.9. Summary 70
Reference 71

Chapter Three: Experimental Setup and Signal Processing Strategies

3.1. Introduction 79

3.2. Commercial IMS used in the present thesis. 80

3.2.1.Methods for volatile generation. 83

3.2.2.Comparative study of three IMS spectrometers 84

3.3. Data set used in this thesis: Motivation, work scenarios and signal processing methodologies. 93

3.3.1. Synthetic data: Quantitative analysis applied to linear and non-linear behavior of IMS using multivariate strategies 93

3.3.2. Synthetic dataset: Quantitative effect in the limit of detection of known analyte in presence of an interferent 99

3.3.3. Feasible study for detection of 2,4,6-tirchloroanisole (2,4,6-TCA) in wine using a portable Ni-IMS. 104

3.3.4. Qualitative analysis for discriminate wines from different “origins” 110

3.3.5. Feasible study for measurement potential biomarkers of Prostate Cancer using Ion Mobility Spectrometry 114

3.3.6. Breath analysis for detection of SEPSIS in rats using Ion Mobility Spectrometry 117

3.4. Reference 123

Chapter Four: Qualitative Analysis of IMS

4.1. Introduction 131

4.2. Pre-processing of IMS spectra 131

4.2.1. Noise reduction or smoothing 132

4.2.2. Baseline subtraction 139

4.2.3. Peak alignment 141

4.3. Discrimination of wines using Multivariate Analysis based on the information from whole spectra 145

4.4. MCR and SFFS as classification methodology: Application for detection of SEPSIS in rats. 154
4.5. Summary 161
Reference……...163

Chapter Five: Quantitative Analysis of IMS datasets

5.1. Introduction 165
5.2. From Univariate to Multivariate Calibration in IMS using synthetic data set. ... 166
5.2.1. Non-linear effect in IMS using synthetic dataset 173
5.2.2. Mixture effect in IMS using synthetic dataset 179
5.3. Feasible studies for testing IMS in real scenarios. 188
5.3.1. Feasible study for detection of 2,4,6-tirchloroanisole (2,4,6-TCA) in wine using a portable Ni-IMS. 188
5.3.2. Feasible study for measurement potential biomarkers of prostate cancer using Ion Mobility Spectrometry. 194
5.4. Summary 203
5.5. Reference 205

Conclusions

Conclusions of this thesis…….....207

Resumen en Español: Procesado de Señal Multivariante para el análisis cuantitativo y cualitativo de datos aplicados a muestras biomédicas y agroalimentarias.

I. Introducción……..…211
II. Espectrometría de movilidad de Iones…….212
i. Fuente de Ionizacion……212
ii. Compuerta eléctrica……213
iii. Tubo de deriva……..213

iv. Detector……213
v. Comportamiento no lineal del IMS……...213
vi. Afinidad protónica y efecto de adición de dopantes……..215
vii. Aplicaciones biológicas y biomédicas con el IMS……....216
viii. Descripción de datos de IMS…….…...217

III. Experimentos y metodología de análisis desarrollados en la presente tesis…….....218
i. Estudio Comparativo de los tres IMS utilizados en esta Tesis…….....219
ii. Bases de datos y metodologías de procesados de datos desarrolladas durante la tesis para análisis cuantitativos y cualitativos……..221
IV. Análisis de datos IMS en aplicaciones médicas y agroalimentarias…….225
i. Pre-procesado de Espectros de IMS……...225
ii. Análisis cualitativo de datos de IMS……....227
iii. Análisis cuantitativo de Espectros de IMS……..229

V. Conclusiones……..233
VI. Referencias…….235

List of Publications and Conferences

Publications……239
Conferences……...240
Oral Presentations…….241

LIST OF FIGURES

Introduction

FIGURE A (A) INTERACTION BETWEEN PLANTS AND INSECTS (VILLAGRA, 2014) (B) WINE TESTING (WINECLUBE, 2010) (C)DETECTION OF CANCER BASED ON DOG'S SNIFF (MEDICALDETECTIONDOGS, 2014) _____ 4

FIGURE B SCHEMATIC DIAGRAM SUMMARIZING THE ANALYSIS OF VOLATILE ORGANIC COMPOUNDS _____ 5

FIGURE C BREATH SAMPLING TECHNIQUE (PHILLIPS ET AL., 2003A) DEVELOPED BY MICHAEL PHILLIPS TO EXTRACT VOLATILE ORGANIC COMPOUNDS FROM HUMAN BREATH___ 6

FIGURE D (A) COMMERCIAL GAS CHROMATOGRAPHY MASS SPECTROMETRY (GC-MS) DEVELOPED BY THERMO SCIENTIFIC (SCIENTIFIC, 2014) (B) TWO CHROMATOGRAMS FROM TWO DIFFERENT SAMPLES SHOWING PEAKS AT DIFFERENT RETENTION TIME. 7

FIGURE E (A) CHEMICAL SENSOR ARRAYS THAT ARE USED FOR ODOR DETECTION (B) RESPONSE OF THE CHEMICAL ARRAY TO CHANGES OF SUBSTANCES BETWEEN ETHANOL AND ACETONE. (GUTIERREZ ET AL., 2011)_____ 8

FIGURE F (A) COMMERCIAL ION MOBILITY SPECTROMETER GDA2 AIRSENSE, GERMANY (B) TYPICAL RESPONSE OF IMS IN WHICH THREE DIFFERENT SUBSTANCES WAS SEPARATELY MEASURED. _____ 9

Chapter One: Ion Mobility Spectrometry as potential technology in biological scenarios

FIGURE 1.1 SCHEMATIC REPRESENTATION OF ION MOBILITY SPECTROMETER (IMS). (I) IONIZATION SOURCE REGION IN WHICH THE SAMPLE IS IONIZED, (II) DRIFT TUBE WHERE THE IONIZED MOLECULES ARE ACCELERATED BY AN ELECTRIC FIELD, (III) SHUTTER GRID ALLOWS THE IONIZED MOLECULES GO INTO DRIFT TUBE AND (IV) DETECTOR WHERE THE CHARGE OF MOLECULES ARE CONVERTED INTO A CURRENT OUTPUT. _____ 20

FIGURE 1.2 SYNTHETIC REPRESENTATION OF THE BEHAVIOR IN THE FORMATION OF PROTONATED MONOMER AND PROTON-BOUND DIMER. (A) THE INTENSITY RESPONSE OF REACTANT ION PEAK, PROTONATED MONOMER PEAK AND PROTON-BOUND DIMER. (B) SPECTRA AT DIFFERENT INSTANTS (A) JUST REACTANT ION PEAK, (B) PROTONATED MONOMER AND REACTANT ION PEAKS, AND (C) PROTON-BOUND DIMER, PROTONATED MONOMER AND REACTANT ION PEAK._____ 25

FIGURE 1 3 HYPOTHETICAL EXAMPLE ABOUT THE SELECTIVITY OF IMS UNDER DIFFERENT THRESHOLD OF PROTON AFFINITIES (EICEMAN AND KARPAS, 2005) _____ 26

Chapter Two: Quantitative and Qualitative Analysis of Ion Mobility Spectrometry: from univariate to multivariate

FIGURE 2.1 AN EXEMPLIFICATION ABOUT THE HIGH DIMENSIONALITY OF A SINGLE IMS MEASUREMENT. LEFT: AN IMAGE OF THE SCANS VS DRIFT TIME OF A SINGLE MEASUREMENT. RIGHT: DIFFERENT SPECTRA AT DIFFERENT MEASUREMENT TIME (SCANS) SHOWING THE INFORMATION VARIABILITY. 40

FIGURE 2.2 GENERA BLOCK DIAGRAM 41

FIGURE 2.3 LOW FREQUENCY NOISE COUPLED TO IMS SPECTRA. 43

FIGURE 2.4 COMMON CONSTRAINTS USED IN ITERATIVE MCR APPROACHES (DE JUAN AND TAULER, 2006) 57

FIGURE 2.5 BLOCK DIAGRAM OF MCR-ALS APPROACH 59

FIGURE 2.6 MCR-LASSO ALGORITHM'S BLOCK DIAGRAM.(POMAREDA ET AL., 2010) 61

FIGURE 2.7 CALIBRATION CURVE WITH UPPER AND LOWER CONFIDENCE LIMITS. Y_C DECISION LIMIT, X_C CRITICAL LEVEL AND X_D THE DETECTION LIMIT 63

FIGURE 2.8 CROSS VALIDATION FOR SETTING UP MODEL PARAMETERS. 68

Chapter Three: Experimental Setup and Signal Processing Strategies

FIGURE 3.1 SPECTROMETERS USED IN THE CURRENT THESIS. (A) THE HANDHELD GDA2 DEVELOPED BY AIRSENSE, GERMANY (AIRSENSE, 2012), (B) THE PORTABLE UV-IMS DEVELOPED BY GAS DORTMUND (GAS), (C) THE DESKTOP VG-TEST DEVELOPED BY 3QBD, ISRAEL (3QBD) 82

FIGURE 3.2 THE OVG CALIBRATION GAS GENERATOR DEVELOPED BY OWLSTONE (OWLSTONE, 2014) TOGETHER WITH IMS USED IN THIS PRESENT THESIS 83

FIGURE 3.3 RAW SPECTRA FROM THREE AMINES (A) GDA2 (NI-IMS) AIRSENSE (B) UV-IMS (G.A.S DURTMUND) AND (C) VG-TEST (3QBD, ISRAEL) 86

FIGURE 3.4 CALIBRATION MODEL OF TMA FOR EACH SPECTROMETER. (A) GDA, (B) UV-IMS AND (C) VG-TEST. 90

FIGURE 3.5 THE SIGNAL INTENSITY OF THE ANALYTE (TMA) AND REACTANT ION (TEP) PEAK IN HEADSPACE VIAL AS A FUNCTION OF TIME. (B) THE THEORETICAL (WITH PEFECT MIXING) DILUTION OF THE TMA HEADSPACE ANALYTE VAPOR FOR A CARRIER FLOW OF 400 ML MIN^{-1} (6.67 ML S^{-1}) AND A 20 ML VIAL VOLUME FOR VG-TEST. (C) FINAL CALIBRATION MODEL USING POLY-PLS. THE PREDICTED CONCENTRATION VS REAL CONCENTRATION IN WHICH THE FINAL MODEL HAVE A 3 LATENT VARIABLES AND A POLYNOMIAL OF ORDER 2. 91

FIGURE 3.6 SYNTHETIC REPRESENTATION OF THE EFFECT OF CONCENTRATION RESPONSE IN ION MOBILITY SPECTROMETRY. (A) CONCENTRATION PROFILES AS CONCENTRATION INCREASES. (B) SPECTRAL RESPONSES FOR TWO PARTICULAR CONCENTRATIONS 93

FIGURE 3.7 BLOCK DIAGRAM APPLIED FOR THE ANALYSIS OF NONLINEARITIES 96

FIGURE 3.8 BLOCK DIAGRAM FOR STUDYING MIXTURES OF BIOGENIC AMINES. [1] UNIVARIATE LIMIT OF DETECTION USING EQ. 3.5 . [2] MULTIVARIATE LIMIT OF DETECTION USING EQUATIONS EQ. 3.6 AND EQ. 3.7 _____ 103
FIGURE 3.9 BLOCK DIAGRAM OF SIGNAL PROCESSING FOR TCA SAMPLES _____ 108
FIGURE 3.10 (A) THE GAS CHROMATOGRAM OF THE HEADSPACE VAPOR OF 2,4,6-TRICHLOROANISOLE; (B) THE MASS SPECTRUM OF THE PEAK AT 8.32 MIN IN THE CHROMATOGRAM; (C) THE MASS SPECTRUM OF 2,4,6-TRICHLOROANISOLE (NIST DATABASE) _____ 110
FIGURE 3.11 INTRODUCTION SYSTEM OF WINE SAMPLES USING UV-IMS (GARRIDO-DELGADO ET AL., 2011A). _____ 112
FIGURE 3.12 BLOCK DIAGRAM OF THE ANALYSIS OF WINE SAMPLES FOR CLASSIFICATION PURPOSES. _____ 113
FIGURE 3.13 SAMPLE INTRODUCTION SYSTEM USED TO GENERATE AND INJECT THE SAMPLE INTO THE UV-IMS INSTRUMENT. _____ 117
FIGURE 3.14 EXPERIMENTAL SETUP OF BREATH SAMPLING (GUAMAN ET AL., 2012). _____ 120
FIGURE 3.15 BLOCK DIAGRAM OF THE ANALYSIS OF VAPORS FROM BREATH ANALYSIS FOR DETERMINING SEPSIS IN RATS. _____ 121

Chapter Four: Qualitative Analysis of IMS datasets

FIGURE 4.1 RAW SPECTRA OF A SINGLE MEASUREMENT OF TWO SPECTROMETERS. (A) GDA2 RAW SPECTRA (B) ZOOM OF THE TAIL WITHOUT PEAKS INFORMATION OF THE SAME SPECTRA AS (A), (C) UV-IMS RAW SPECTRA AND (D) ZOOM OF THE TAIL WITH NOT RELEVANT INFORMATION. _____ 132
FIGURE 4.2 (A) SMOOTHING USING SAVITZKY-GOLAY FILTER OF ORDER 2 USING DIFFERENT WIDTH SIZES, (B) SMOOTHING USING SAVITZKY-GOLAY FILTER OF ORDER 2 USING DIFFERENT WIDTH SIZES (REGION WITH NO PEAKS), (C)) SMOOTHED SPECTRA USING SAVITZKY-GOLAY FILTER OF ORDER 2 AND WIDTH OF 15, (D) SMOOTHED SPECTRA USING SAVITZKY-GOLAY FILTER OF ORDER 2 AND WIDTH OF 15 (REGION OF NO PEAKS). _____ 133
FIGURE 4.3 SAVISTZKY AND GOLAY FILTER APPLIED TO UV-IMS SPECTRA (A) FILTER OF ORDER 2 USING DIFFERENT WIDTH SIZES APPLIED TO ONE SPECTRUM (B) FILTER OF ORDER 2 USING DIFFERENT WIDTH SIZES APPLIED TO ONE SPECTRUM (REGION WITH NO PEAK INFORMATION), (C) SMOOTHED SPECTRA USING SAVITZKY-GOLAY FILTER OF ORDER 2 AND WIDTH OF 15, AND (D) SMOOTHED SPECTRA USING SAVITZKY-GOLAY FILTER OF ORDER 2 AND WIDTH OF 15 (REGION WITH NO PEAK INFORMATION) _____ 134
FIGURE 4.4 PCA USED AS FILTER. (A) LOADINGS OF PCA MODEL, (B) UV-IMS SPECTRA BEFORE AND AFTER FILTERING. _____ 136
FIGURE 4.5 ICA USED AS FILTER. (A)INDEPENDENT COMPONENTS, (B) UV-IMS SPECTRA BEFORE AND AFTER FILTERING. _____ 137
FIGURE 4.6 (A) SINGLE SPECTRUM BEFORE AND AFTER NOISE REDUCTION , (B) VG-TEST SPECTRA BEFORE AND AFTER NOISE REDUCTION (C) FAST FOURIER TRANSFORMATION OF EACH INDEPENDE COMPONENT OF ICA, AND (D) FFT OF SINGLE SPECTRUM BEFORE AND AFTER FILTERING. _____ 139
FIGURE 4.7 BASELINE SUBTRACTION (A) GDA2, (B) UV-IMS AND (C) VG-TEST. _____ 141

FIGURE 4. 8 GDA2 SPECTRA (A) MEASUREMENT OF A SINGLE ANALYTE DURING 12 MINUTES IN WHICH A SLIGHT MISALIGNMENT IS OBSERVABLE, (B) SPECTRA OF A SINGLE MEASUREMENT THAT LAST 12 MINUTES, (C) DIFFERENT MEASUREMENTS OF THE SAME ANALYTE IN WHICH A MISALIGNMENT SHOULD BE FIX IT. _____ 142
FIGURE 4. 9 ALIGNMENT OF PEAKS USING A REFERENCE PEAK (RIP). (A) MEASUREMENT OF A SINGLE ANALYTE DURING 12 MINUTES, (B) ALIGNED SPECTRA OF A SINGLE MEASUREMENT THAT LAST 12 MINUTES, (C) DIFFERENT MEASUREMENTS OF THE SAME ANALYTE. _____ 143
FIGURE 4.10 WINE SPECTRA OF DIFFERENT ORIGINS. (A) PREPROCESSED SPECTRA, (B) PREPROCESSED SPECTRA AFTER ALIGMENT. _____ 144
FIGURE 4.11 (A) MEAN SPECTRUM OF EACH WINE SAMPLE (B) SCORES OF THE PCA MODEL (C) LOADINGS OF THE PCA MODEL _____ 149
FIGURE 4.12 SCANNING PLOT OF PCA-LDA STRATEGY FROM 4 TO 20 PRINCIPAL COMPONENTS. _____ 150
FIGURE 4. 13 SCATTER PLOT FOR THE LDA OBTAINED USING 16PCS ON TRAINING SET FROM IMS DATA. MONTILLA-MORILES (RED CIRCLE), JEREZ (GREEN TRIANGLE), VALDEPEÑAS (LILIAC SQUARE) AND HUELVA (BLUE STAR). _____ 150
FIGURE 4.14 PCA MODEL OF GC DATASET._____ 152
FIGURE 4. 15 SPECTRA FROM BREATH ANALYSIS IN CONTROL AND SEPSIS RATS. (A) POSITIVE MODE IMS AND (B) NEGATIVE MODE IMS. _____ 155
FIGURE 4. 16 SPECTRA PROFILE FROM MCR-LASSO ANALYSIS. PURE COMPONENTS PEAKS (P) FROM MCR-LASSO RESULTS FOR RAT'S BREATH. EVERY COMPONENT FROM P1 TO P14 HAS ITS REDUCED MOBILITY K0 (CM2 V−1 S−1) FOR POSITIVE AND NEGATIVE MODE. FILLED PEAKS CORRESPOND TO ANESTHESIA, AIR POLLUTION AND REACTANT ION PEAK FROM IMS, AND THE OTHERS ARE RELATED TO COMPOUNDS FROM BREATH. POSITIVE SPECTRA: P1 (K0 = 2.35): RIP COMES FROM NITROGEN ION SPECIES, P2 (K0 = 2.11): RIP COMES FROM WATER ION SPECIES, P3 (K0 = 1.97): A COMPONENT FROM LABORATORY ROOM AIR, P4 (K0 = 2.04), P5 (K0 = 1.89), P6 (K0 = 1.84), P7 (K0 = 1.82), P8 (K0 = 1.79). NEGATIVE SPECTRA: P9 (K0 = 2.25): RIN, P10 (K0 = 2.11): A COMPONENT FROM LABORATORY ROOM AIR, P11 (K0 = 1.52): ANESTHESIA, P12 (K0 = 2.16), P13 (K0 = 2.01), P14 (K0 = 1.60).(GUAMAN ET AL., 2012)_____ 155
FIGURE 4.17 CONCENTRATION PROFILE OF DIFFERENT COMPOUNDS PRESENT IN BREATH SAMPLES (A) POSITIVE MODE AND (B) NEGATIVE MODE._____ 156
FIGURE 4.18 PLOT OF IMS SAMPLES USING THE THREE COMPOUNDS THAT WERE SELECTED BY SFFS ALGORITHM._____ 157
FIGURE 4. 19 CHROMATOGRAM OF BREATH SAMPLES FROM CONTROL AND SEPSIS RATS. 157
FIGURE 4.20 SCORE PLOT FROM LDA ANALYSIS _____ 160

CHAPTER FIVE: Quantitative Analysis of IMS datasets

FIGURE 5.1 EXAMPLE OF UNIVARIATE CALIBRATION. (A) RAW SPECTRA OF ACETONE (B) UNIVARIATE CALIBRATION OBTAINED USING PEAK HEIGHT OF ACETONE _____ 167
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) _____ 168
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. _____ 169
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 _____ 170
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 $cm^2\ V\ s^{-1}$ _____ 172
Figure 5.6 MCR-ALS results for 2-butanone spectra. (a) Spectra profile. (b) Concentration Profile. _____ 172
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. _____ 174
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. _____ 175
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. _____ 176
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. _____ 178
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. _____ 180
Figure 5.12 Univariate Calibration using ratio (TMA/TMA+TEP+PUT) _____ 181

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. _____ 182
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. _____ 184
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. _____ 189
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. _ 190
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. _____ 191
FIGURE 5.18 CONCENTRATION PROFILE FOR CALIBRATION. (A) ORIGINAL CONCATENATED CONCENTRATION PROFILE FROM MCRLASSO; (B) CONCENTRATION PROFILE ALIGNED AND SMOOTHED. _____ 192
FIGURE 5.19 (A) LOADINGS FROM PLS MODEL, AND (B) CALIBRATION CURVE_____ 193
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. _ 195
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. _____ 196
FIGURE 5.22 PLS MODELS OF COMPOUND A (A) PURE COMPOUND (B) COMPOUND DILUTED IN WATER. _____ 197
FIGURE 5.23 (A) SPECTRA PROFILE OF COMPOUND A, (B) PLS MODEL , (C) CONCENTRATION PROFILE._____ 198
FIGURE 5. 24 TWO SPECTRA OF RIP (BLUE LINE) AND COMPOUND A (GREEN LINE)._____ 199
FIGURE 5. 25 SPECTRA PROFILE AND CONCENTRATION PROFILE OF COMPOUND A. _____ 200
FIGURE 5. 26 A) CONCENTRATION PROFILE OF COMPOUND A FOR DIFFERENT CONCENTRATION RANGES. (B) PLS MODEL FOR LOD CALCULATION. _____ 201

Resumen en Español: Procesado de Señal Multivariante para el análisis cuantitativo y cualitativo de datos aplicados a muestras biomédicas y agroalimentarias.

FIGURA I REPRESENTACIÓN ESQUEMÁTICA DE UN ESPECTRÓMETRO DE MOVILIDAD DE IONES (IMS). (I) REGIÓN DE IONIZACIÓN DONDE LA MUESTRA ES IONIZADA, (II) TUBO DE DERIVA DONDE LOS IONES IONIZADOS SON ACELERADOS POR LA ACCIÓN DE UN CAMPO ELÉCTRICO, (III) COMPUERTA ELÉCTRICA QUE PERMITE EL PASO DE LOS IONES DE LA FUENTE DE IONIZACIÓN AL TUBO DE DERIVA, Y (IV) DETECTOR DONDE LAS CARGAS DE LAS MOLÉCULAS GENERAN UNA RESPUESTA ELÉCTRICA..... 212
FIGURA II REPRESENTACIÓN SINTÉTICA DE LA FORMACIÓN DE MONÓMEROS PROTONADOS Y DIMEROS EN UNA MUESTRA. (A) RESPUESTA DEL PICO DEL ION REACTIVO (RIP), MONÓMERO Y DÍMERO. (B) ESPECTRO DEL MISMO COMPUESTO PRESENTADO A DIFERENTES INSTANTES DE TIEMPO. (A) APARECE ÚNICAMENTE EL RIP, (B)

aparece el monómero junto con el RIP y (C) aparece el RIP, junto con el monómero y dímero de la molécula. 214
Figura III Ejemplo hipotético sobre la selectividad de IMS cuando se aplican hipotéticamente dopantes con diferentes afinidad protónica (Eiceman and Karpas, 2005). 215
Figura IV Secuencia de espectros IMS de una única medida, donde se puede observar que a medida que se incrementa el tiempo de análisis (SCANS) la información de los espectros cambian. 217
Figura V Espectrometros comerciales utilizados en la presente tesis. (a) Espectrometro de mano GDA2 desarrollado por Airsense, Alemania(Airsense, 2012) (b) El Espectrometro UV desarrollado por GAS Dortmund (GAS), (c) El espectrómetro de mesa VG-Test dearrollado por 3QBD, Israel..... 218
Figura VI Diagrama de Bloques para el analizar el efecto de no linealidad de muestras sintéticas en IMS..... 222
Figura VII Diagrama de bloques general de la presente tesis..... 225
Figura VIII Ruido de baja frecuencia acoplado al espectro de IMS (azul), Espectro suavizado utilizando savistzky-golay (rojo), Espectro filtrado utilizando PCA (magenta), Espectro filtrado utilizando ICA (verde)..... 226
Figura IX (a) Espectros desalineados que contienen un pico de referencia (b) Desalineamiento de muestras sin pico de referencia. 227
Figura X (a) Modelo final de discriminación(LDA) (tercera componente discriminante vs primera componente discriminante) (b) Tasa de clasificación luego de haber repetido el proceso 20 veces con 100 iteraciones de bootstrap..... 228
Figura XI (a) Espectros obtenidos de cada compuesto puro extraido de MCRLasso (b) Perfil de Concentraciones de cada compuesto extraido por MCRLasso 229
Figura XII Perfil de Espectros y Concentraciones al utilizar MCR-ALS en medidas de 2-butanona..... 230
Figura XIII El efecto de mezclas en IMS. TMA puro (azul), Putrecina (PUT) (rojo) y mezcla de TMA y PUT (verde)..... 230
Figura XIV (a) Perfil de Espectros después de aplicar MCR-Lasso, (b) Perfil de Concentraciones de RIN y el ion chloride (c) perfil de concentraciones de monómero y dímero de TCA 231
Figura XV Resultados de MCRLasso cuando se analiso el compuesto A relacionado con PCa (a) perfil de espectros y (b) modelo de calibración. 232

List of Tables

Chapter Two: Quantitative and Qualitative Analysis of Ion Mobility Spectrometry: from univariate to multivariate

Table 2.1 Summary about requirements of classifiers. _____ 48

Chapter Three: Experimental Setup and Signal Processing Strategies

Table 3.1 Comparison of the main operation parameters of the three IMS devices used in the present study. _____ 80
Table 3.2 The ion species observed in TMA, putrescine (PUT) and Cadaverine(CAD) and their reduced mobility values ($cm^2V^{-1}s^{-1}$) calculated relative to 2,4-lutidine(LUT) _____ 87
Table 3.3 The dependence of the response of the VG-Test, GDA2 and UV-IMS on the concentration of trimethylamine, putrescine and cadaverine. * Root mean square error of cross validation. _____ 89
Table 3.4 The limit of detection calculated on the basis of MCR-LASSO for vapors of trimethylamine, putrescine and cadaverine in air for the GDA2, GAS and VG-Test ion mobility spectrometers. Also shown is the limit of detection for TMA in headspace vapors emanate _____ 92
Table 3.5 Concentrations for 2-butanone and ethanol generated using the volatile generator system OVG (Owlstone) _____ 95
Table 3.6 Different concentrations of TMA and PUT for the mixture analysis._____ 102
Table 3.7 Range of concentrations of Compound A for both spectrometers. _____ 116

Chapter Four: Qualitative Analysis of IMS datasets

Table 4.1 Signal to noise ratio before and after using different filtering algorithms. _____ 138
Table 4.2 Summary of wine dataset analyzed by UV-IMS. _____ 147
Table 4.3 Classification performace of PCA-LDA model using boostrap validation. _____ 151

Table 4.4 Compounds analyzed by GC-FID(1) and CFS-CPS-UV-IMS(2) (Garrido-Delgado et al., 2011) 153
Table 4.5 Identification of compounds from GC dataset 159

CHAPTER FIVE: Quantitative Analysis of IMS datasets

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. 177
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. 179
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. 179
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. 185
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. 193
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. 199

Resumen en Español: Procesado de Señal Multivariante para el análisis cuantitativo y cualitativo de datos aplicados a muestras biomédicas y agroalimentarias.

Tabla I Coeficientes de Movilidad de putrecina, cadaverina y TMA para los Tres Espectrómetros 220
Tabla II Límite de Detección para TMA, PUT, CAD para GDA2, UV-IMS y VG-Test. 220