



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



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- Compartiqual 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - Compartiqual 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-ShareAlike 3.0. Spain License.**



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

RESUMEN EN ESPAÑOL

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

I. Introducción

La percepción y medida de olores y volátiles tiene un interés creciente en procesos industriales y en el campo de la medicina. Existen muchos ejemplos donde el olfato juega un papel fundamental en procesos de calidad en la industria alimentaria, entre ellos: catadores de vino, aceite, jamón, perfumes, etc., (Holmberg, 2010, wineclub, 2010). Actualmente, en medicina, existen pruebas diagnósticas bien establecidas basadas en pruebas de aliento que tratan de diagnosticar enfermedades en base a los compuestos orgánicos volátiles que son emitidos por el cuerpo debido a procesos metabólicos de la enfermedad (BBCnews, 2013, Pauling et al., 1971). La idea principal es que el olor contiene información discriminante y esta información puede resultar de gran utilidad.

Fisiológicamente hablando, los olores están formados en su mayoría por compuestos orgánicos volátiles (VOC, por sus siglas en inglés) los cuales son captados por el sistema olfativo (neuronas olfativas) y, posteriormente, son convertidos en impulsos eléctricos que luego son interpretados por el cerebro. La interpretación del olor se basa en un proceso de entrenamiento, reconocimiento e identificación del olor con respecto al aroma que desprende un objeto o un organismo natural. Utilizando el mismo concepto, se puede utilizar sensores químicos que puedan transformar información química (olores) en señal eléctrica (información) y mediante algoritmos de clasificación se pueda transformar estas señales en la percepción o discriminación de acuerdo a las necesidades del problema (Persaud and Dodd, 1982).

En esta tesis se va a utilizar la espectrometría de movilidad de iones (IMS) (Eiceman and Karpas, 2005) como instrumento de medida de VOCs. IMS es un instrumento portable, con un tiempo de respuesta en segundos y que bajo cierta configuración es capaz de analizar e identificar compuestos en el orden de ppm e incluso sub_ppm. Por otro lado, la selectividad de los instrumentos IMS es moderada, su respuesta es no-lineal y tienen baja reproducibilidad. IMS fue inicialmente utilizado para la detección de explosivos y drogas ilícitas (Eiceman et al., 1990, Eiceman, 1991, Eiceman et al., 1995, Eiceman and Stone, 2004). Sin embargo, gracias a su portabilidad y su capacidad para adaptarse a medidas en tiempo real, en los últimos años IMS se ha introducido paulatinamente en otras áreas de estudio. Estas nuevas aplicaciones, sobretudo aplicaciones en el ámbito médico y alimenticio, han traído consigo también la necesidad de desarrollar y utilizar técnicas de procesado de señal más sofisticado para extraer la información relevante de este tipo de muestras más complejas.

Los principales objetivos de esta tesis doctoral es usar y desarrollar técnicas de procesamiento de señal multivariante para abordar los problemas típicos de IMS como son: no linealidad, competencia por la carga, mejorar la selectividad, etc. Todo esto centrado en el análisis cualitativos y cuantitativos de las diferentes aplicaciones presentadas.

II. Espectrometría de movilidad de Iones.

La Espectrometría de movilidad de Iones (IMS por sus siglas en inglés), es un equipo portable, relativamente económico y fácil de usar. IMS permite detectar compuestos químicos en base a la velocidad de los sus iones en fase gaseosa cuando están siendo sometidos a un campo eléctrico a presión ambiente (Eiceman and Karpas, 2005).

Existen dos procesos fundamentales que ocurren en el IMS, la primera es la formación de iones que provienen de una muestra, y la caracterización de la movilidad de estos iones cuando son expuestos a un campo eléctrico. El IMS consiste en cuatro partes fundamentales, tal como se muestra en la Figura 1 .

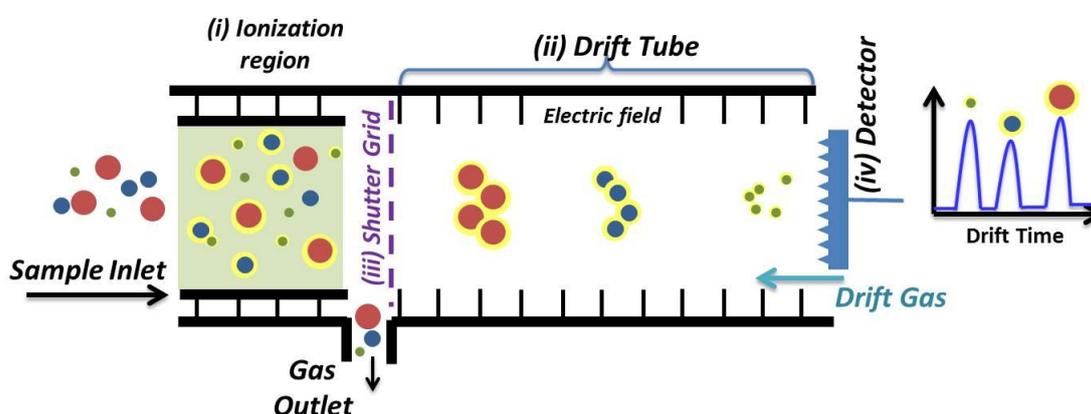


Figura 1 Representación esquemática de un espectrómetro de movilidad de iones (IMS). (i) Región de ionización donde la muestra es ionizada, (ii) Tubo de deriva donde los iones ionizados son acelerados por la acción de un campo eléctrico, (iii) compuerta eléctrica que permite el paso de los iones de la fuente de ionización al tubo de deriva, y (iv) detector donde las cargas de las moléculas generan una respuesta eléctrica.

i. Fuente de Ionización

La fuente de ionización funciona como un reservorio de iones reactivos que pueden ser positivos y/o negativos y son transportados por un gas transportador (puede ser aire o nitrógeno). Existen diferentes tipos de fuentes de ionización, entre ellas fuentes radioactivas (Kim et al., 1978, Eiceman and Karpas, 2005), fuentes ultravioleta (Siegel et al., 2005), o por descarga eléctrica (Tabrizchi et al., 2000) (Eiceman et al., 2014). De la misma manera dependiendo del tipo de fuente de ionización, el proceso puede dar paso a la formación de iones positivos y/o negativos, este es el caso de las fuentes radioactivas. Durante la ionización ocurren reacciones entre iones neutros del aire circundante y las moléculas de la muestra, que vienen ligadas a principios de termodinámica, cinética y estructura de las moléculas. Existen otros factores externos que pueden afectar estas reacciones tales como la temperatura, presión, mezcla y concentración de la muestra.

ii. Compuerta eléctrica

Una vez que las moléculas son ionizadas, estas son inyectadas al tubo de deriva mediante una compuerta electrostática cuya función principal es permitir el paso de las moléculas ionizadas de una forma sincronizada y escalada. Este es unos parámetros que necesita ser optimizado para un correcto análisis.

iii. Tubo de deriva:

Tan pronto como entran las moléculas ionizadas al tubo de deriva las moléculas son aceleradas por el efecto de la aplicación de un campo eléctrico (generalmente entre 100-300 V cm⁻¹). Efectivamente, la velocidad con la que viajan estas moléculas ionizadas implica un tiempo de vuelo en el tubo de derivas, que es lo que finalmente permite caracterizar la información química de la muestra.

La velocidad de los iones (t_d) se puede reemplazar por lo que se conoce como coeficientes de movilidad (K) (1). Como se puede observar en la ecuación (1), este coeficiente depende de la longitud del tubo de deriva (L), el campo eléctrico aplicado (E), y el tiempo de deriva (t_d) que es el tiempo en la que los iones llegan al colector (Eiceman and Karpas, 2005). Este coeficiente es lo que se utiliza para poder comparar el rendimiento entre IMS comerciales. Pero para que esto pueda suceder, se deben establecer unas condiciones experimentales equiparables, para ello existe un coeficiente normalizado llamado coeficiente de movilidad reducida (K_0) (2). Este coeficiente corrige variaciones de temperatura (T) y presión (P) del gas (Eiceman and Karpas, 2005).

$$K = \frac{L}{Et_d} \quad (1)$$

$$K_0 = K \frac{273}{T} \frac{P}{760} \quad (2)$$

iv. Detector

Al final del tubo de deriva existe un detector, usualmente es un plato de Faraday, donde la carga de los iones es convertida a corriente eléctrica. Al final se obtiene un espectro que contienen diferentes picos dependiendo de la movilidad de los iones en fase gaseosa. De esta forma, los picos con un tiempo de deriva bajo están relacionado con moléculas pequeñas y viceversa.

v. Comportamiento no lineal del IMS

Una parte muy importante de esta tesis es el estudio de las no linealidades que están presentes en el análisis de IMS. Las no linealidades presentes en el IMS abren nuevas líneas de investigación en el campo del procesamiento de señales, sobre todo cuando se requiere hacer análisis cuantitativos y cualitativos.

Intuitivamente, es de esperar que un compuesto puro dé como resultado un pico en el espectro del IMS. Sin embargo, ocurre que, dependiendo de la concentración, un compuesto puede presentar más de un pico en el espectro del IMS. La pluralidad de reacciones ion-molécula durante el proceso de ionización o la complejidad de las colisiones y fragmentaciones son algunos de los motivos que originan este comportamiento..

La Figura II contiene una representación sintética de lo que sucede cuando existe un incremento de concentración de una única molécula. En la Figura II (a) se puede observar que en un inicio únicamente hay una gran cantidad de iones reactantes, lo que significa que hay en el espectro A únicamente el pico correspondiente al pico reactante (RIP). A medida que la concentración de la molécula M incrementa se va formando el monómero de la molécula. Esto se puede observar en el espectro de la Figura II (b), el espectro B representado por el color verde, donde aparece el RIP con menor intensidad y el monómero del compuesto M con una intensidad mayor. En este ejemplo se representa que hay un incremento aún mayor de la molécula M, entonces se observa en la figura en la parte superior la intensidad del monómero empieza a disminuir y en cambio se observa que hay la aparición del dímero. De la misma forma, se observa en el espectro C que el pico RIP prácticamente desaparece, el pico del monómero aún está presente pero con una menor intensidad y aparece un nuevo pico relacionado con el dímero.

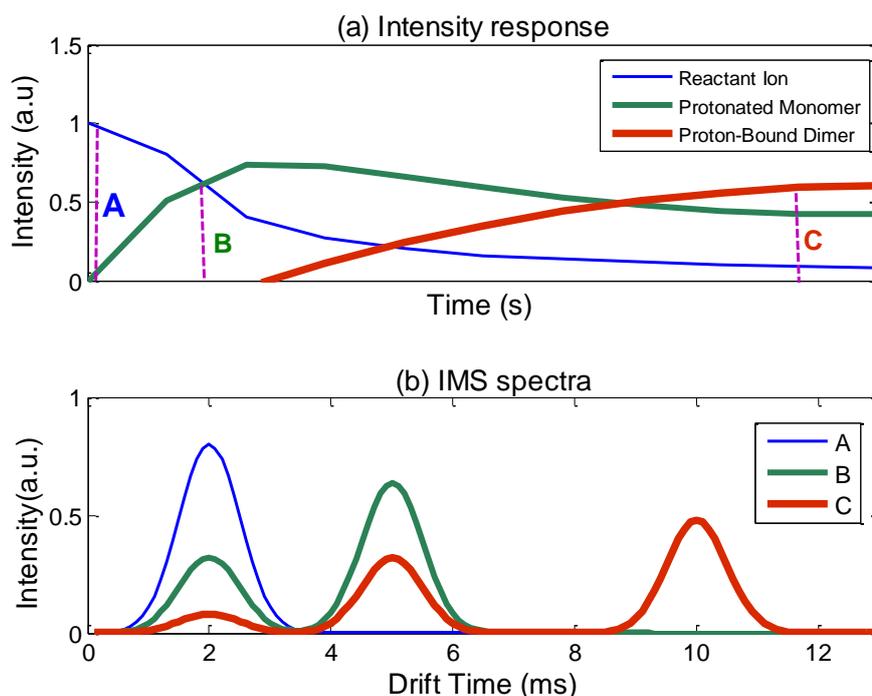


Figura II Representación sintética de la formación de monómeros protonados y dímeros en una muestra. (a) Respuesta del pico del ion reactivo (RIP), monómero y dímero. (b) Espectro del mismo compuesto presentado a diferentes instantes de tiempo. (A) Aparece únicamente el RIP, (B) aparece el monómero junto con el RIP y (C) aparece el RIP, junto con el monómero y dímero de la molécula.

Este fenómeno refleja una problemática en el análisis de los espectros de IMS, puesto que en una muestra real se desconoce la composición y la concentración de cada uno de los compuestos de la muestra. Por lo tanto es un error, en el análisis de muestras reales, pensar que un pico representa un compuesto. De la misma manera, tampoco se puede pensar que el comportamiento de un compuesto es lineal cuando su concentración se incrementa. Por lo tanto es importante determinar las técnicas apropiadas y adecuadas para poder atacar este problema y obtener resultados confiables.

vi. Afinidad protónica y efecto de adición de dopantes.

La afinidad protónica es una información que es utilizada para mejorar la selectividad del IMS y disminuir la complejidad en el análisis de las señales IMS. En este sentido la utilización de dopantes usa esta ventaja, para simplificar el análisis de los espectros IMS.

La Figura III muestra un ejemplo teórico de como la adición de dopantes con diferente afinidad protónica cambia la respuesta del IMS (Eiceman and Karpas, 2005). Los diferentes tipos de compuestos químicos tienen diferente afinidad protónica. Por ejemplo, se conoce que los alcanos tienen una afinidad protónica baja y que, por el contrario, las aminas tienen una alta afinidad protónica. Esta es la razón por la cual los IMS con fuente radioactiva, trabajando bajo el modo de operación "water chemistry", no son sensibles a alcanos puesto que el agua tiene una mayor afinidad protónica que éstos (observe la Figura III). Por lo tanto, el agua actúa como un dopante eliminando la posibilidad de poder analizar compuestos de la familia de los alcanos. Ciertamente, la adición de un dopante es útil cuando se conoce a priori el tipo de compuesto o compuestos de interés, permitiendo la eliminación o reducción del background de la muestra. Sin embargo, cuando se trata de un estudio exploratorio e inicial, donde se desconoce totalmente los compuestos involucrados durante el análisis, el uso del dopante puede ser perjudicial y un impedimento para un análisis completo.

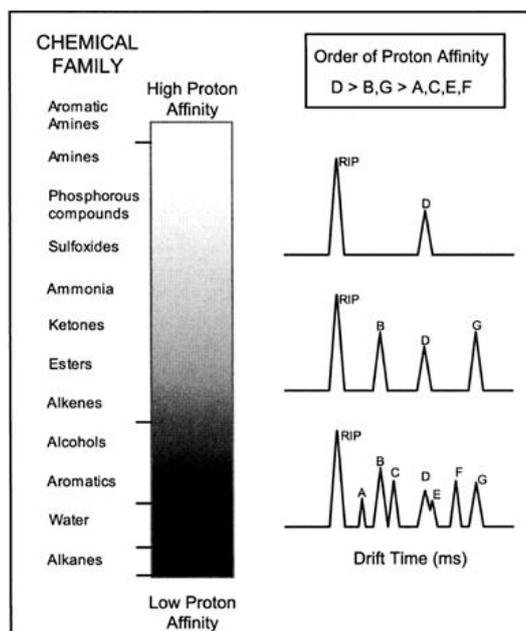


Figura III Ejemplo hipotético sobre la selectividad de IMS cuando se aplican hipotéticamente dopantes con diferentes afinidad protónica (Eiceman and Karpas, 2005).

Existe otro efecto ligado a la afinidad protónica de los compuestos químicos, conocido como competencia por la carga. Donde, durante el proceso de ionización, y dependiendo de la afinidad protónica el IMS se vuelve más selectivo a unos compuestos que a otros. Esta competencia por la carga, de hecho, puede enmascarar algunos compuestos y afectar directamente el análisis cuantitativo de los compuestos de la muestra. Esto es un problema grave en cualquier análisis, pero sobretodo en el análisis de muestras reales cuando no se conoce a priori los compuestos químicos presentes en la misma. Existen diferentes investigaciones que intentan estudiar este efecto y proveer de alguna solución al respecto, entre ellas, (Eiceman et al., 1990, Puton et al., 2008, Puton et al., 2012, Marquez-Sillero et al., 2011b).

Otro aspecto ligado a la selectividad del instrumento es la resolución pico a pico (R_{pp}) (Spangler, 2002). Este parámetro puede ser calculado mediante las ecuaciones (3) y (4), donde se t_d es el tiempo de deriva del analito y w_h es el área del pico a media altura. Este parámetro puede ser considerado en las estrategias de procesado de señal para resolver problemas de picos solapados.

$$R_{pp} = \frac{t_d}{w_h} \quad (3)$$

$$R_{pp} = \frac{2(t_{d2} - t_{d1})}{1.7(w_{h2} - w_{h1})} \quad (4)$$

vii. Aplicaciones biológicas y biomédicas con el IMS.

Inicialmente, la detección de explosivos o de drogas y estupefacientes eran las aplicaciones más comunes para IMS (Hannum et al., 2000, Buxton and Harrington, 2001, Ewing et al., 2001, Armenta and Blanco, 2012, Synder et al., 1995). Las razones de su éxito en estos ámbitos son sus ventajas: la portabilidad, la rapidez de la respuesta y la sencillez a la hora de utilizar estos equipos. Estas características han permitido que IMS sea un instrumento con potencial para ser usado en otras áreas de investigación.

En la actualidad, las aplicaciones donde IMS se está utilizando son: análisis ambiental, industria alimentaria, control de calidad de fármacos, y aplicaciones clínicas y biológicas (Armenta et al., 2011, Marquez-Sillero et al., 2011a, Borsdorf et al., 2011, Karpas, 2013).

En el campo alimenticio las aplicaciones más usuales están centradas en análisis de calidad para detectar el estado de los alimentos (Karpas et al., 2002, Bota and Harrington, 2006), detección de comida o bebidas adulteradas (Garrido-Delgado et al., 2012, Garrido-Delgado et al., 2011, Karpas et al., 2012, Alonso et al., 2008).

En el ámbito clínico se ha utilizado IMS para determinar análisis de muestras biológicas de orina, saliva, pasma y sangre, para la determinación de enfermedades o abuso de drogas (Lu et al., 2009, Alizadeh et al., 2008, Shahdousti and Alizadeh, 2011, Jafari et al., 2011). Otra área interesante donde IMS está siendo ampliamente utilizado es el análisis del aliento para determinar enfermedades como cáncer de pulmón u otras dolencias tanto en humanos (Perl et al., 2009) como en modelos animales (Guaman et al., 2012), también existen estudios para determinar los biomarcadores relacionados con enfermedades vaginales (Karpas et al., 2013, Sobel et al., 2012, Chaim et al., 2003).

viii. Descripción de datos de IMS

Un espectro de IMS representa la intensidad de corriente en función de un tiempo de deriva. Una única medida de IMS está compuesta por varios espectros, conocidos como *scans*. Un *scan*, a su vez, está conformado por el promedio de varios espectros consecutivos. El promediado tiene el objetivo de mejorar la relación señal ruido. La dimensionalidad de este tipo de datos, va ligado directamente al número de *scans* generados en cada medida. La dimensionalidad de una medida de IMS es una matriz $N \times M$, donde N corresponde al número de *scans*, y M a los puntos de deriva. En la Figura IV se puede observar un ejemplo de una medida de IMS, donde además se puede observar la respuesta dinámica y la alta dimensionalidad de datos.

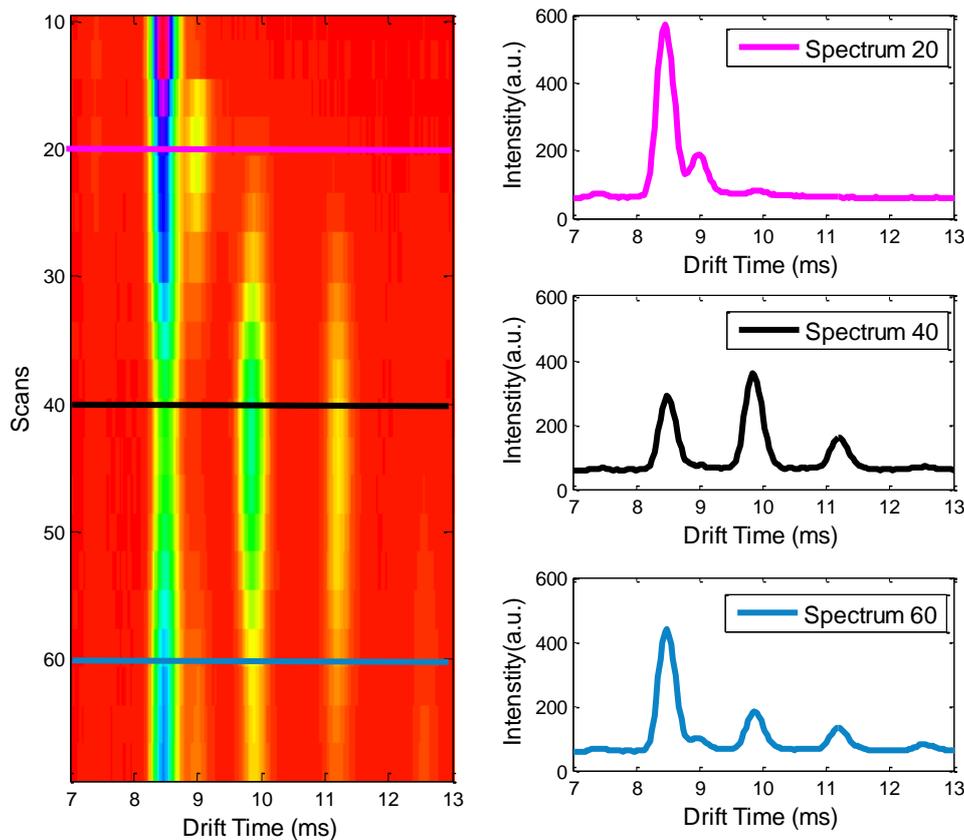


Figura IV Secuencia de espectros IMS de una única medida, donde se puede observar que a medida que se incrementa el tiempo de análisis (SCANS) la información de los espectros cambian.

III. Experimentos y metodología de análisis desarrollados en la presente tesis

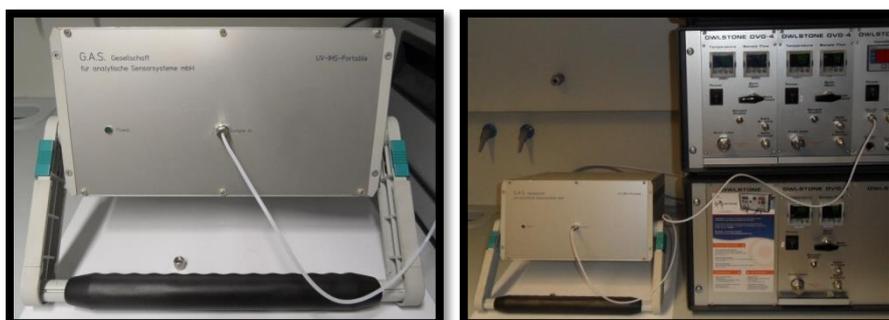
Durante el desarrollo de esta presente tesis se utilizaron diferentes espectrómetros comerciales, los mismos que son mostrados en la Figura V junto con el generador de volátiles utilizado para realizar varias pruebas sintéticas (Owlstone, 2014).



(a)



(b)



(c)

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 desarrollado por 3QBD, Israel

- **EL GDA2 (Airsense, 2012)** (Figura V (a)) es un IMS defuente radioactiva, dispone de membrana hidrófoba para evitar la humedad y el tubo de derivas dispone de un sistema de control de temperatura. La fuente radioactiva, de Ni63, es de 100MBeq. El instrumento está diseñado para trabajar sin dopante inducido, bajo la configuración conocida como “wáter chemistry”, en la que el agua funciona como dopante natural. Esto permite que una gran variedad de compuestos puedan ser analizados, siempre y cuando tengan una mayor afinidad protónica que el agua. En este sentido los alcanos, lo cuales son considerados como biomarcadores de algunas enfermedades, no pueden ser medidos por este IMS.
- **EL VG-TEST(3QBD)** (Figura V (b)) es un IMS de escritorio, cuya fuente de ionización está basado en el efecto corona. Debido a su diseño original, este IMS es usado principalmente para el análisis de muestras biomédicas. En este sentido, el IMS tiene un puerto de entrada donde se introduce un algodón con la muestra biológica a ser analizada. Además, este dispositivo permite añadir diferente dopantes para mejorar la sensibilidad del equipo, sin embargo el dopante por defecto es Triethylphosphate (TEP) para volverlo más selectivo a las amines.
- **EL UV-IMS (GAS)** (Figura V(c)) es un IMS portable cuya fuente de ionización está basado en una fuente ultravioleta de 10.6 eV. El tubo de deriva trabaja a temperatura ambiente. Este instrumento está diseñado para medir una gran variedad de moléculas incluyendo alcanos. El hecho de no poder controlar la temperatura puede favorecer la formación de nuevos iones y afectar a la identificación de los compuestos. Por lo tanto en este equipo es necesario mantener todas las variables experimentales controladas.

i. Estudio Comparativo de los tres IMS utilizados en esta Tesis.

Tomando en cuenta las diferencias en las características técnicas de los tres IMS utilizados durante esta tesis, se vio la necesidad de realizar un estudio comparativo para observar el rendimiento de los 3 equipos en unas mismas condiciones experimentales. Existen muy pocos estudios donde se realicen comparaciones entre diferentes IMS, entre ellas podemos mencionar (Borsdorf et al., 2001, Borsdorf et al., 2005, Borsdorf et al., 2009).

El estudio se basó en analizar tres aminas, putrecina (PUT), trimetilamina (TMA) y cadaverina (CAD) a diferentes niveles de concentraciones para los tres IMS. Como figura de mérito para medir el rendimiento de los tres equipos, se calculó el límite de detección y sensibilidad de los espectrómetros. Para que todo esto fuera posible, se siguió un procedimiento riguroso de preprocesamiento (véase sección IV.i) y se estableció una metodología robusta para la creación de los modelos cuantitativos (véase sección IV.iii). Para más detalles refiérase a la siguiente publicación (Karpas et al., 2013). A continuación se explican los resultados más relevantes.

En la Tabla I se puede observar los coeficientes de movilidad calculados por cada una de las aminas con respecto a la Lutidina. Es evidente que existen algunas diferencias en los coeficientes de movilidad reducida entre los espectrómetros. Esto genera una complicación en términos de transferencia de tecnología, ya que no se puede tener

una librería general para todos los espectrómetros. Estas diferencias no pueden asociarse con incertidumbres en el tiempo de deriva y escalado sino más bien con la formación de las especies de iones en cada IMS y sus diferencias en las condiciones de operación. Este resultado refleja la necesidad de realizar una calibración del IMS con la finalidad de poder identificar correctamente los compuestos de la muestra.

| Compuesto | Ion | Movilidad Reducida | | |
|--------------|--------------|--------------------|----------|---------|
| | | GDA2 | G. A. S. | VG-Test |
| Temperatura | | 44°C | 26°C | 90°C |
| 2,4-Lutidine | (LUT)H+ | 1.90 | 1.90 | 1.90 |
| TMA | (TMA)H+ | 2.22 | 2.13 | 2.10 |
| | (TMA)2H+ | 1.78 | - | - |
| Putrescine | (PUT)H+ | 1.94 | 1.99 | 2.02 |
| | (PUT)2H+ | 1.46 | - | - |
| Cadaverine | (CAD)H+ | 1.81 | 1.82 | 1.87 |
| | (CAD)2H+ | 1.36 | - | - |
| Mixed | (CAD)(PUT)H+ | 1.41 | - | - |

Tabla I Coeficientes de Movilidad de putrecina, cadaverina y TMA para los Tres Espectrómetros

El límite de detección para cada amina y cada espectrómetro fue calculado y la respuesta se muestra en la Tabla II. Los cálculos realizados por el límite de detección mediante las técnicas multivariantes muestran que el rendimiento de cada instrumento es relativamente el mismo. Por supuesto que se observan diferencias entre cada instrumento, pero tomando en cuenta la desviación para cada LOD, se puede entender que las diferencias no son estadísticamente significativas. Estos resultados son un tanto sorprendentes, tomando en cuenta las diferencias en el modo de operación de los tres espectrómetros. El presente estudio comparativo arroja resultados que tienen implicaciones directas en los posibles límites de detección que pueden ser obtenidos por IMS clásicos (sin ningún elemento preconcentrador).

| Compound | Sample type | VG-Test | GDA2 | G.A.S |
|------------|-----------------------|---------|---------|---------|
| TMA | Vapor in air (ppm) | 0.1±0.1 | 1.5±0.1 | 0.5±0.2 |
| | Headspace vapors (µg) | 0.9±0.6 | 1.9±1.2 | 0.7±1.1 |
| Putrescine | Vapor in air (ppm) | 0.5±0.3 | 0.7±0.1 | 0.7±0.7 |
| Cadaverine | Vapor in air (ppm) | 0.4±0.2 | 0.2±0.1 | 0.4±0.1 |

Tabla II Límite de Detección para TMA, PUT, CAD para GDA2, UV-IMS y VG-Test.

ii. Bases de datos y metodologías de procesados de datos desarrolladas durante la tesis para análisis cuantitativos y cualitativos

El objetivo de esta tesis es desarrollar estrategias y metodologías orientadas al análisis de datos de IMS. Los resultados de la presente tesis se dividen en estrategias para análisis cualitativos y estrategias para el análisis cuantitativo. A su vez, los resultados del análisis cuantitativo se dividen en análisis desarrollado en base de datos sintéticos (realizado en el laboratorio en entorno controlado) y base de datos de muestras reales de aplicaciones médicas y agroalimentarias.

• **Análisis Cuantitativos utilizando bases de datos sintéticas para analizar el efecto de no linealidad y efecto de mezclas:**

El objetivo principal de este conjunto de datos es establecer las diferencias entre usar métodos tradicionales de análisis cuantitativo, como lo es el análisis univariante, y utilizar dos métodos multivariantes, como lo son PLS (Wold et al., 1984) o poly-PLS (Wold et al., 1989) y MCR en su versión ALS (de Juan and Tauler, 2006) o Lasso (Pomareda et al., 2010).

A modo general, el diagrama de bloques presentado en la Figura VII representa la estrategia para analizar los datos sintéticos generados para no linealidades y efecto de competencia por la carga en mezclas. La principal diferencia en el segundo caso es que se calculó el límite de detección de los equipos en lugar del error en validación cruzada.

- *Conjunto de datos no-linealidades:* En este caso se analizaron 12 muestras de 2-Butanona y Ethanol a diferentes rangos de concentración. En este caso se aseguró que las concentraciones máximas provoquen la no-linealidad del instrumento. Para más detalle refiérase al capítulo tres de la presente tesis o al siguiente artículo (Pomareda et al., 2012).
- *Conjunto de datos para determinar el efecto cuantitativo de un compuesto en presencia de un interferente:* En este caso, se pretende comparar la diferencia del límite de detección de TMA entre analizar la amina pura y en presencia de putrecina. Para ello se generó una base de datos donde TMA fue analizado a diferentes concentraciones (0 a 0.5 ppm) cercanas al LOD establecido en (Karpas et al., 2013). Se repitió el experimento de TMA con 4 niveles diferentes de putrecina (0 a 18 ppm).

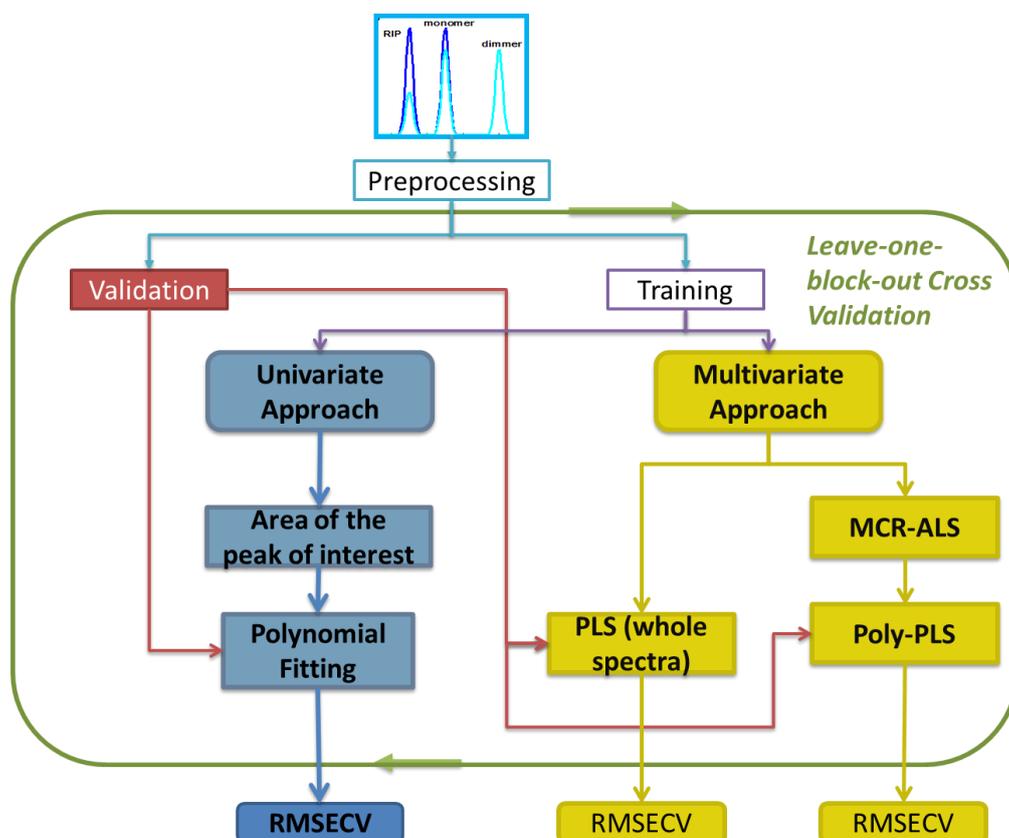


Figura VI Diagrama de Bloques para el analizar el efecto de no linealidad de muestras sintéticas en IMS.

• **Conjuntos de datos de muestras reales de aplicaciones médicas y agroalimentarias para el análisis cuantitativo.**

Con el objetivo de probar cómo las técnicas y la metodología realizada funciona con muestras sintéticas. Se utilizaron las técnicas multivariantes y una metodología similar en muestras reales, para entender la potencialidad de IMS en el análisis de muestras reales.

- *Conjunto de datos para determinar la capacidad de IMS para la detección de 2,4,6 Tricloroanisol (TCA) en vinos.*

Contextualizando, TCA es un compuesto principalmente conocido como el responsable de generar mal olor en el vino. Normalmente proviene de una contaminación bacteriana en el corcho del tapón, y de ahí puede migrar al vino. (Rubio-Coque et al., 2006, Pereira et al., 2000). El efecto directo de esta contaminación del vino afecta directamente a la economía de los grandes productores de vino y corcho mundial (wineclub, 2010). La técnica analítica estándar implica el uso SPME-GC/MS. Mediante este procedimiento se han reportado valores de detección de TCA de 1-100 ng L⁻¹ (Riu et al., 2006, Copete et al., 2009, Zalacain et al., 2004). Sin embargo, se conoce que GC/MS no puede ser utilizada como una herramienta de detección en línea, y por lo tanto no se puede aplicar directamente en procesos industriales en cadena.

El objetivo de este estudio con IMS era probar la viabilidad de utilizar IMS para detección rápida de TCA en vino. Para poder medir la viabilidad se procede a obtener el límite de detección de IMS sin preconcentrar la muestra. Para ello diferentes muestras de TCA se depositan en viales y se analizan los volátiles del espacio de cabeza. En este caso muestras de TCA desde 2 a 40 µg, son cuidadosamente colocadas en un vial, y selladas inmediatamente. En este caso

se calientan las muestras a 100 °C para promover la generación de vapor en el vial. Luego este vapor es arrastrado a 400 ml min⁻¹ al IMS, donde es analizado. Como técnica de análisis multivariante se utiliza MCR-Lasso (Pomareda et al., 2010) para obtener los compuestos puros y una técnica de calibración multivariante PLS para realizar el modelo de calibración y estimar el error de validación cruzada. Los parámetros son configurados mediante validación cruzada. Además cuatro blancos son separados para establecer el LOD del instrumento y el error de predicción del modelo. (Más detalles capítulo tres, Figure 3.9).

▪ **Conjunto de datos para determinar la viabilidad de analizar un potencial biomarcador de Cáncer de Próstata:**

El cáncer de próstata (PCa) es la tercera causa principal de muerte en hombres en la comunidad europea. El PCa puede ser tratado adecuadamente, si se detecta prematuramente. Se conoce que las células prostáticas están presente en la orina, por lo tanto realizar un diagnóstico mínimamente invasivo es hipotéticamente posible y resulta ser novedoso. En este sentido, existen investigaciones en el ámbito de la proteómica donde se han encontrado proteínas más frecuentes en pacientes con PCa que en una población control.

En un estudio preliminar la orina de pacientes con PCa y niveles altos de PSA fueron analizados mediante GC/MS. De este estudio se obtuvo que un compuesto en específico podría ser un potencial biomarcador de PCa. Por supuesto, al ser un estudio preliminar, estos resultados se deben corroborar y confirmar. En cualquier caso, y tomando en cuenta que el objetivo de esta tesis es analizar la viabilidad de IMS como técnica analítica en tiempo real, un estudio preliminar fue llevado a cabo para establecer el LOD de dos espectrómetros comerciales como lo son GDA2 y UV-IMS.

Se analizó el compuesto puro por dos métodos, por barrido de espacio de cabeza y directamente conectado al generador de volátiles (Owlstone, 2014), para los dos espectrómetros. Es de esperarse que el LOD de las muestras analizadas con el generador de volátiles sea menor, que el que se obtiene mediante barrido de espacio de cabeza. Además, para emular un entorno real se analizó el compuesto disuelto en agua. Para determinar el LOD se utilizó la técnica de MCR-Lasso y luego se construyó un modelo multivariante con validación cruzada y blancos ciegos para determinar LOD y el error de validación cruzada.

• **Conjunto de datos de muestras reales de aplicaciones médicas y agroalimentarias para el análisis cualitativo.**

En este caso el objetivo primario es realizar modelos robustos para discriminar diferentes clases. En este sentido, no existen muchos estudios donde se aborde el análisis de datos de IMS desde una perspectiva cualitativa. En este caso, se plantearon dos metodologías multivariantes (i) modelo utilizando la información del espectro y proyectarlo a un subespacio de una dimensionalidad menor a la original y (ii) utilizar MCR con alguna de sus variaciones y combinarla con técnicas de selección de características para escoger aquellos compuestos más discriminantes en la clasificación.

▪ **Conjunto de datos para la discriminación de vino blanco de diferente dominación de origen.**

La motivación de este trabajo es crear un modelo que permita clasificar muestras de vino blanco de acuerdo a su dominación de origen (más detalles en (Garrido-Delgado et al., 2011)). Esta base de datos consiste en muestras de cuatro vinos

con diferentes denominaciones de origen (i) Condado de Huelva, (ii) Jerez (iii) Valdepenas y (iv) Montilla-Moriles. Todas las muestras de vino fueron medidas mediante un dispositivo especialmente desarrollado para obtener la fase gaseosa de líquidos (Desarrollado por la Universidad de Córdoba).

El análisis multivariante llevado a cabo en este caso fue una combinación del análisis de componentes principales (PCA) con análisis discriminante de Fisher (LDA). Además, se incorporó a la metodología el método de validación conocido como bootstrap (Efron, 1979, Felsenstein, 1985).

- *Conjunto de datos para la discriminación de ratas control contra ratas con SEPSIS.* Este trabajo está enfocado en el ámbito del análisis de aliento (Di Natale et al., 2003, Pauling et al., 1971, Phillips et al., 1999, Phillips et al., 2007, Van den Velde et al., 2008). El objetivo es poder identificar los compuestos que permitan discriminar ratas con SEPSIS de ratas control. En este caso la hipótesis es que existen metabolitos expirados, totalmente relacionados con la SEPSIS (Guaman et al., 2012). Para esta base de datos el aire exhalado de 20 ratas genéticamente idénticas fue analizado por el IMS. 10 ratas control y 10 ratas fueron inyectadas para inducir los síntomas de la SEPSIS. El análisis multivariante realizado en este caso se basó en la extracción de los compuestos puros por medio de MCRLasso. Del conjunto de componentes puros extraídos se aplicó un algoritmo de selección de características conocido como sequential floating features selection (SFFS (Pudil et al., 1994)).

IV. Análisis de datos IMS en aplicaciones médicas y agroalimentarias.

El propósito de esta tesis está ejemplificado en la Figura VII. En términos generales el objetivo de esta tesis es dar respuesta y soluciones al análisis cualitativo y cuantitativo de espectros IMS de datos relacionados con aplicaciones médicas y agroalimentarias. Como se puede observar claramente, todos los datos de IMS deben pasar por un preprocesado de los mismos, y dependiendo del propósito se realizarán modelos cuantitativos o cualitativos.

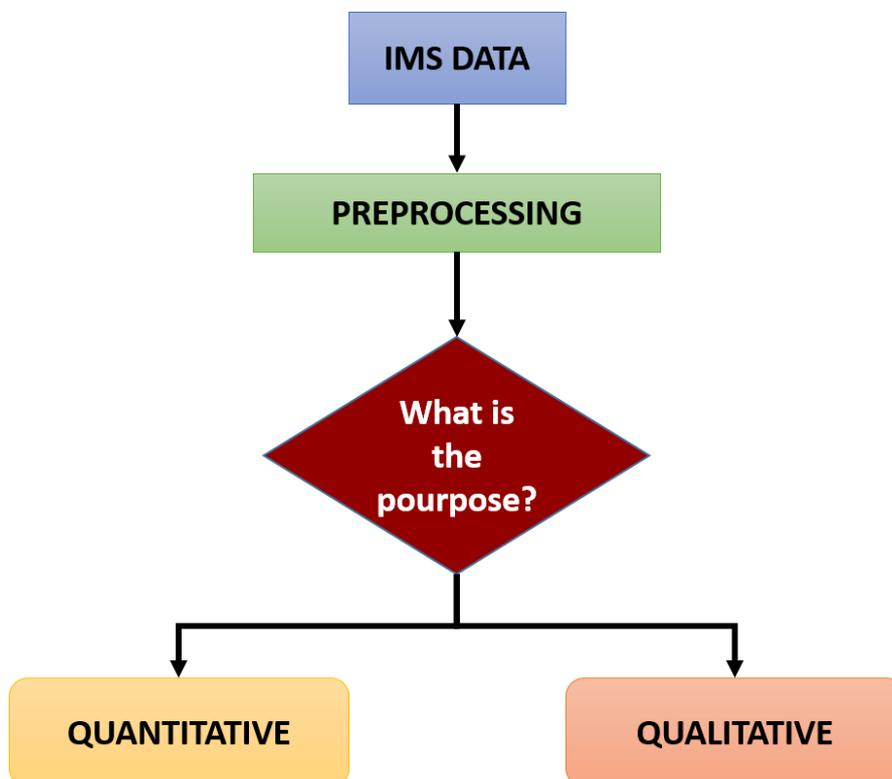


Figura VII Diagrama de bloques general de la presente tesis

i. Pre-procesado de Espectros de IMS

En general, el pre procesado se puede dividir en diferentes etapas. (i) Reducción o suavizado del ruido en el espectro de IMS, (ii) eliminación de la línea de base, (iii) alineamiento de picos y (iv) extracción de características.

Reducción de Ruido: En general una buena práctica es suavizar el ruido de los espectros en lugar de aplicar filtrado convencional. En esta tesis se utiliza el algoritmo de savitsky y golay (Savitzky and Golay, 1964). El algoritmo requiere de dos parámetros de configuración que son la anchura (o número de puntos) y el orden de polinomio. Para escoger el número de puntos se debe tomar en cuenta la anchura de pico o la resolución del espectrómetro, para evitar cualquier distorsión de los picos. Con respecto al orden del polinomio es aconsejable que sea de orden 2 o máximo 3.

En esta tesis también se desarrolló una estrategia para eliminar ruido de baja frecuencia acoplado totalmente al espectro de IMS. Este caso fue muy particular de un espectrómetro comercial, donde el ruido generado por un extractor químico quedo acoplado al espectro de IMS (Figura VIII). Debido a que las técnicas convencionales no se pueden utilizar para eliminar este ruido de baja frecuencia, puesto que atenúa y

deforma el espectro. Se propone utilizar dos metodologías de componentes principales (PCA) (Statheropoulos et al., 1999) y el método de análisis de componentes independientes (ICA) (Comon, 1994). La idea general es que el ruido de baja frecuencia puede separarse en componentes diferentes al ser proyectados los datos a un nuevo subespacio. Tal como se puede observar en la Figura VIII, la mejora es evidente al utilizar estas dos técnicas. Además se estimó la relación señal ruido antes y después de aplicar las técnicas de eliminación de ruido y se obtuvo una mejora importante de 18dB a 58 y 67 dB para PCA e ICA respectivamente.

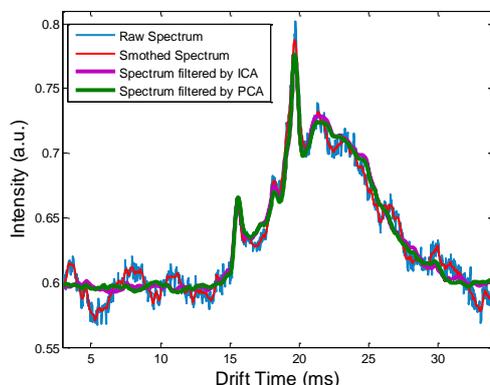


Figura VIII Ruido de baja frecuencia acoplado al espectro de IMS (azul), Espectro suavizado utilizando savitzky-golay (rojo), Espectro filtrado utilizando PCA (magenta), Espectro filtrado utilizando ICA (verde)

Eliminación de la Línea de Base: Eliminar la línea de base tiene como objetivo principal permitir una correcta determinación de información relevante. Hay que asegurar que las diferencias que se puedan observar no son debidas a cambios de la línea de base. En el caso de IMS, la línea de base se puede determinar ajustando un polinomio de un determinado orden. En este caso lo más importante es determinar los puntos para la determinación de la línea de base. En general es aconsejable utilizar puntos del espectro donde no hay picos, lo más común es ajustar la línea de base a los puntos de los extremos del espectro (valores de tiempo de deriva bajos y altos). El orden del polinomio va depender totalmente del equipo.

Alineamiento de Picos. Los espectros de IMS sufren de derivas durante el tiempo de medida o entre medidas diferentes. Este es un problema serio, tomando en cuenta que es importante identificar los compuestos presentes en la muestra y luego compararlos con los presentes en otras muestras. Al existir un desalineamiento de los picos encontrar la relación entre las muestras resulta un reto, sobretodo es crítico si se pretende utilizar métodos multivariantes para el subsecuente análisis.

Para el alineamiento se propone realizarlo en dos etapas. (i) alineamiento inicial de la muestra y (ii) alineamiento entre muestras. Además es aconsejable, en el caso de que existan muestras provenientes de un mismo origen, es preferible realizar un alineamiento intermedio bajo la hipótesis de que contienen los mismos compuestos principales. Y luego realizar el alineamiento entre muestras de diferentes orígenes.

Dependiendo del espectrómetro, podemos tener dos condiciones para el alineamiento, (i) tener un o varios picos de referencia, en el caso de GDA2 (RIP) Figura IX (a), o (ii) no tener ninguna referencia Figura IX (b). En el primer caso, los picos de referencia ayudan en verificar el desalineamiento existente y diseñar mejor la corrección a realizar. En el segundo caso, el alineamiento se vuelve un proceso más intuitivo de identificar el patrón que resulta ser un espectro de referencia. A partir de ahí, la idea es aplicar alguna técnica de corrección, como puede ser icoshift (Tomasi et al., 2011), basada en el espectro de referencia.

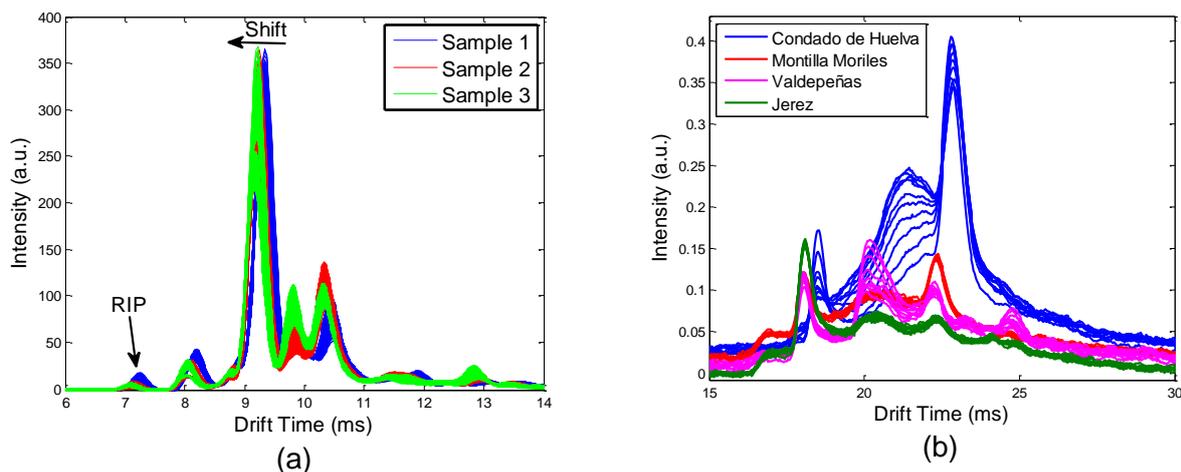


Figura IX (a) Espectros desalineados que contienen un pico de referencia (b) Desalineamiento de muestras sin pico de referencia.

ii. Análisis cualitativo de datos de IMS

• **Análisis de base de datos de vinos para discriminarlos de acuerdo a su denominación de origen.**

Después de realizar el preprocesado y eliminar los outliers, el conjunto de datos final tenía una dimensionalidad de 1632 espectros de 34 muestras x 351 puntos de deriva (variables). La propuesta inicial en este caso es combinar PCA con LDA. PCA en este caso es utilizado como técnica de reducción de dimensionalidad y LDA (Smit et al., 2007) como técnica discriminativa. En este caso, PCA proyecta los datos en un nuevo espacio de menor dimensión, y sobre ese nuevo espacio se genera una nueva proyección utilizando LDA. Al final nos queda un nuevo espacio de 1632 espectros de 34 muestras x 3 componentes discriminantes. El número de componentes principales de la proyección inicial es determinado mediante bootstrap (Efron, 1979, Felsenstein, 1985). Sobre el espacio final generado por LDA se utiliza kNN (Cover and Hart, 1967) como clasificador final.

Los datos fueron inicialmente divididos en entrenamiento (75%) y validación (25%). Los datos de entrenamiento se utilizan para construir el modelo y los datos de validación son proyectados en el modelo final para calcular el porcentaje de acierto de clasificación. En este caso se tomó en consideración que estamos clasificando muestras y no espectros, por lo tanto para calcular el porcentaje de acierto se estima mediante voto mayoritario de todos los espectros correspondientes de sus muestras. Para comprobar la robustez de la metodología, se repite 20 veces todo el procedimiento con 100 repeticiones de bootstrap, para determinar la incertidumbre de error en la clasificación.

En la Figura X se pueden observar los resultados obtenidos al aplicar la metodología descrita. En el apartado (a) se observa el modelo final donde se puede apreciar cómo quedan diferenciados las cuatro clases de vinos de acuerdo a su denominación de origen. En la Figura X (b) se observa la tasa de clasificación obtenida a medida que se va incrementando el número de componentes principales, y ejemplifica la necesidad de configurar los parámetros del modelo mediante técnicas de validación cruzada. La tasa de clasificación en este caso obtenida fue del 82% (89%-95%).

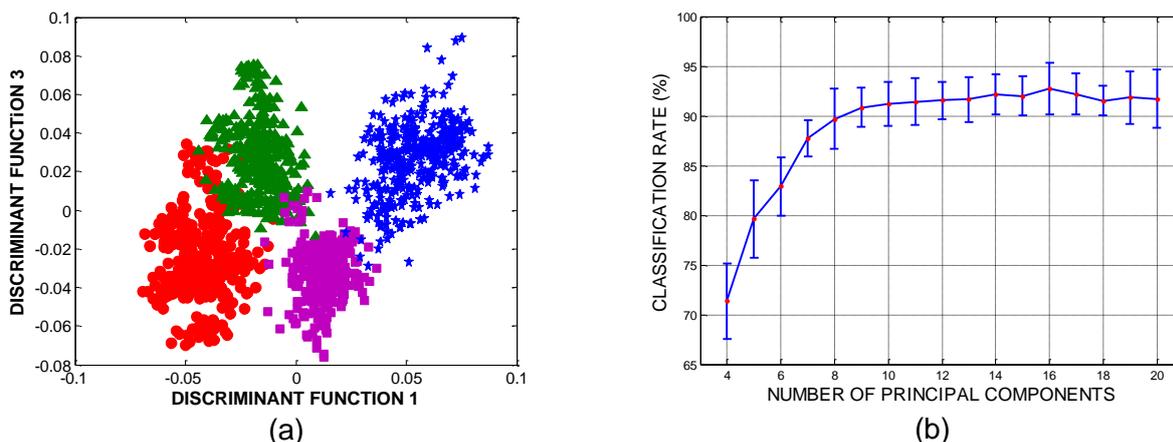


Figura X (a) Modelo final de discriminación(LDA) (tercera componente discriminante vs primera componente discriminante) (b) Tasa de clasificación luego de haber repetido el proceso 20 veces con 100 iteraciones de bootstrap.

• **Análisis de base de datos para discriminar ratas con SEPSIS de ratas control.**

En este trabajo 40 muestras fueron analizadas de 20 ratas (10 control y 10 con sepsis) con una réplica de cada una. En este caso se propuso utilizar MCRLasso (Pomareda et al., 2010) para extraer los compuestos puros del espectro de IMS. El resultado de aplicar MCRLasso es obtener el espectro de los componentes puros y la concentración asociada a cada uno de los componentes. En este caso el análisis se realizó utilizando el espectrómetro comercial GDA2, el cual genera iones positivos y negativos. Después de aplicar MCRLasso se obtuvieron 14 compuestos puros de los espectros positivos y negativos. En la Figura XI se muestra el perfil de espectros y concentraciones obtenidos del espectro positivo, en este caso 8 compuestos puros. Del total de los compuestos puros 6 compuestos fueron identificados como compuestos externos al análisis del aliento, entre ellos los asociados al RIP, Anestesia, y compuestos del entorno de medida.

Se realiza una selección de características utilizando el perfil de concentraciones de los 8 compuestos puros asociados con metabolitos propios del aire exhalado y relacionados con la enfermedad. En este caso se aplicó SFFS para obtener un subconjunto de compuestos más relevantes y asociados con la SEPSIS. Como resultado se obtuvieron 3 compuestos puros como los más significativos y potenciales biomarcadores de SEPSIS. La clasificación del modelo final es del 99% (98% - 99.7%).

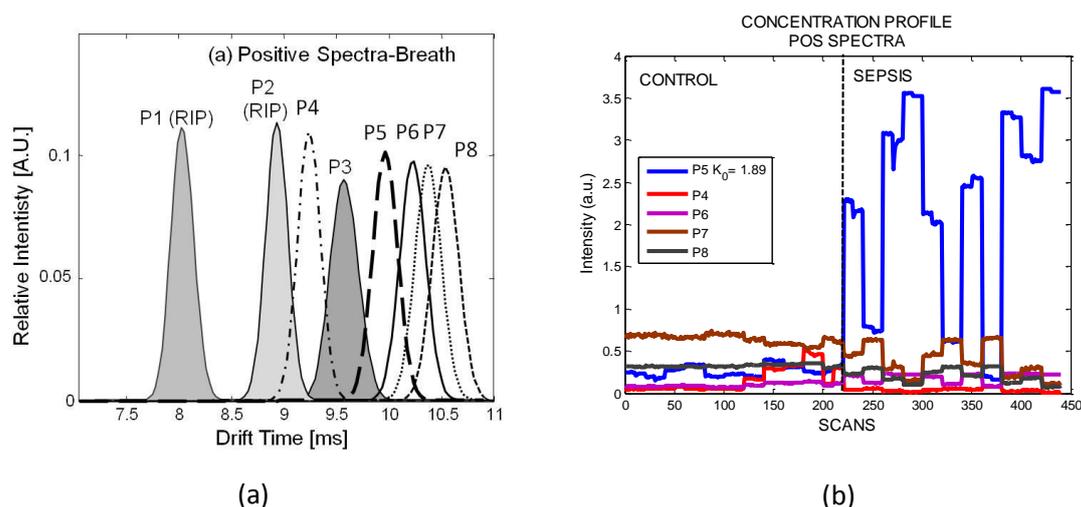


Figura XI (a) Espectros obtenidos de cada compuesto puro extraído de MCRLasso (b) Perfil de Concentraciones de cada compuesto extraído por MCRLasso

iii. Análisis cuantitativo de Espectros de IMS

- **Análisis de conjuntos de datos sintéticos para estudiar los efectos de no linealidades y competencia por la carga en mezclas**

El objetivo primordial de este análisis es entender el comportamiento no lineal de IMS y proponer soluciones en el análisis de los espectros IMS. Para este análisis se han propuesto tres metodologías: análisis univariante, siendo ésta la técnica más convencional a través del cálculo de la altura de pico, y dos técnicas multivariantes (i) PLS (Wold et al., 1984) y (ii) MCR-ALS (deJuan et al., 1997) para obtener los compuestos puros. En este último caso, utilizando la información del perfil de concentraciones, se aplicó una técnica de calibración final. Se realizó un estudio con dos compuestos etanol y 2-butanona a niveles de concentraciones suficientemente altos para provocar la no linealidad de IMS. Para ambas sustancias el espectro original presenta dos picos: la primera del monómero y la segunda del dímero.

En la Figura XII se presenta la solución al utilizar MCR-ALS para extraer los compuestos puros de 2-butanona. En este caso, se observa claramente 3 compuestos puros: el RIP (propio del GDA2), el monómero (rojo) y el dímero (verde) para 2-butanona. En el perfil de concentraciones se observa el comportamiento esperado de no linealidades, donde se observa que a medida que aumenta la concentración, aumenta la intensidad del monómero. Pero con un incremento mayor la intensidad del monómero empieza a decaer y aumenta la intensidad correspondiente al dímero.

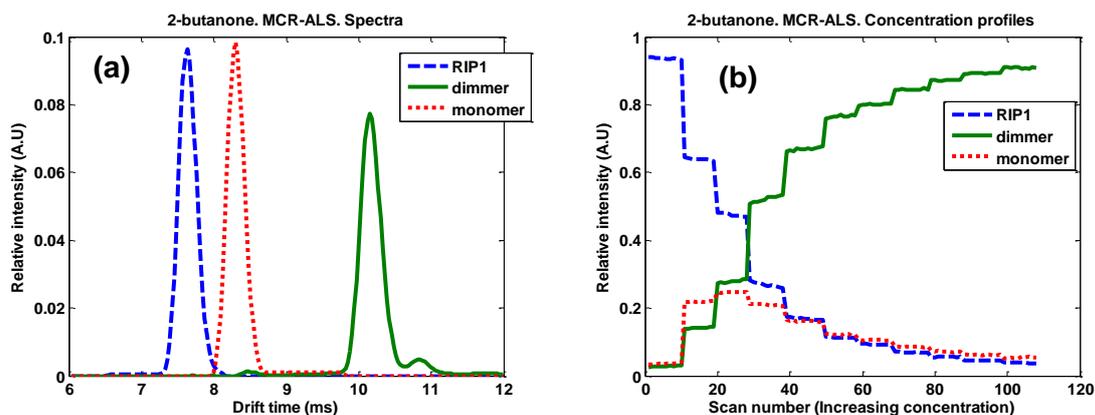


Figura XII Perfil de Espectros y Concentraciones al utilizar MCR-ALS en medidas de 2-butanona

Después de haber realizado la calibración de utilizando las tres variantes explicadas anteriormente, los resultados arrojaron claras evidencias de que el uso de técnicas multivariantes mejoran el rendimiento en el análisis cuantitativo (véase capítulo siete). Con respecto a las dos propuestas utilizadas, los resultados fueron similares con la principal diferencia que MCR-ALS ofrece una mejor interpretabilidad de los resultados.

El otro efecto que se propuso en esta tesis es evaluar el efecto en el LOD de un compuesto específico (en este caso TMA) en presencia de un interferente (putrecina). En la Figura XIII se puede observar con claridad el efecto de mezclas en IMS. Se observa cómo el espectro de TMA a una concentración determinada cambia totalmente cuando se hace una mezcla con Putrecina.

En un estudio previo (Karpas et al., 2013) se determinó que el LOD del TMA en el espectrómetro VG-TEST es 0.1 ppm. En este estudio se propone volver a establecer el LOD pero cuando existe un interferente utilizando métodos convencionales, como lo es la calibración univariante, y métodos multivariantes.

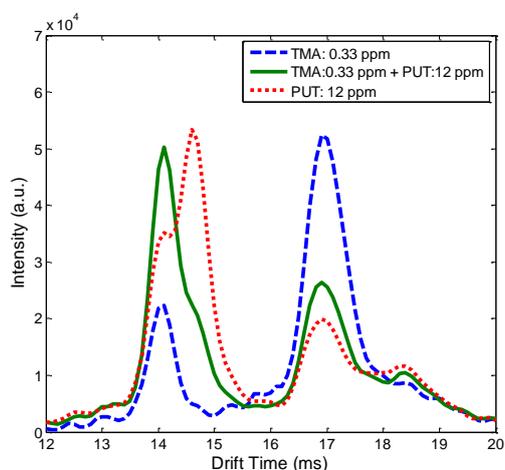


Figura XIII El efecto de mezclas en IMS. TMA puro (azul), Putrecina (PUT) (rojo) y mezcla de TMA y PUT (verde)

Utilizando técnicas univariantes el LOD del TMA obtenido fue el doble en presencia del interferente. Sin embargo al utilizar metodologías multivariantes el LOD del TMA fue similar a 0.1 ppm. En este caso el utilizar primero MCRLasso como técnica para

extraer los compuestos puros y luego una técnica de calibración mejoró sustancialmente la respuesta y por lo tanto el rendimiento de IMS.

- **Análisis de conjuntos de datos reales en aplicaciones médicas y agroalimentarias**

Sabiendo que el análisis multivariante mejora considerablemente el rendimiento de IMS se realizaron estudios de viabilidad en dos aplicaciones reales. El objetivo en ambos estudios fue determinar el LOD del IMS con la principal idea de ver la factibilidad de que este instrumento sea utilizado como una herramienta de monitoreo o diagnóstico, según sea el caso.

El primer estudio se enfocaba en determinar la viabilidad de detectar TCA mediante el IMS sin ningún método de preconcentración y extrapolar este estudio al análisis de TCA en vinos. Para ello se calculó el límite de detección del TCA primero puro, y luego en vino. Como técnica de análisis se utilizó MCR-Lasso para extraer los compuestos puros y posteriormente escoger los perfiles de concentraciones correspondientes al TCA para realizar el modelo de calibración y calcular el LOD. En la Figura XIV se puede observar los compuestos puros después de analizar TCA puro depositado en un vial mediante barrido por espacio de cabeza. Par esta aplicación el LOD encontrado fue 1.7 μg de una solución de diclorometano colocado en un vial. Cuando se realizó el análisis en el vino se concluyó que es necesario acoplar técnicas de preconcentramiento y pre separación de la muestra para mejorar la sensibilidad.

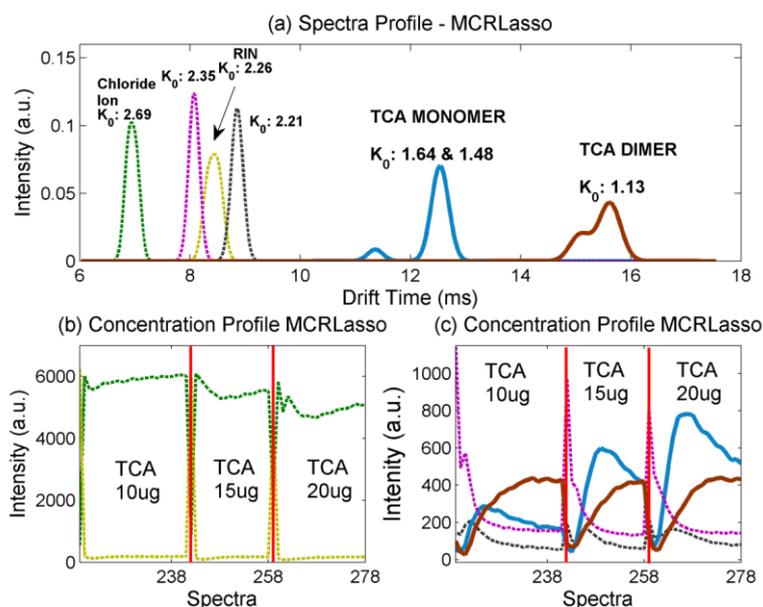


Figura XIV (a) Perfil de Espectros después de aplicar MCR-Lasso, (b) Perfil de Concentraciones de RIN y el ion chloride (c) perfil de concentraciones de monómero y dímero de TCA

El segundo trabajo estaba enfocado en analizar la factibilidad de utilizar IMS como técnica analítica de un posible biomarcador de PCa en la orina. El experimento fue llevado a cabo utilizando compuestos puros en dos espectrómetros comerciales. En el GDA2 el pico del compuesto de interés estaba totalmente solapado con el RIP, por lo que para realizar el análisis se necesitaban concentraciones en el orden de los cientos de ppm. En este sentido, el UV-IMS presenta un espectro mucho más fácil de interpretar, por lo tanto se realizaron diferentes mediciones a diferentes

concentraciones. La Figura XV (a) muestra los resultados obtenidos al utilizar MCR-Lasso en el espectro de UV-IMS y se determinó que el pico, del espectro color rojo, correspondía al monómero del compuesto de interés. La muestra a su vez presento unas impurezas, que se creen que son provenientes del agua donde se diluyo el compuesto. Utilizando el perfil de concentraciones del monómero se realizó un modelo de calibración y se determinó el LOD. El LOD obtenido para el UV-IMS del compuesto puro fue de 0.18 ppm y cuando se diluye en una matriz el LOD se ve afectado aumentando hasta 3 ppm, y cuando es medido por barrido en espacio de cabeza el LOD es 1.3 µg/ml.

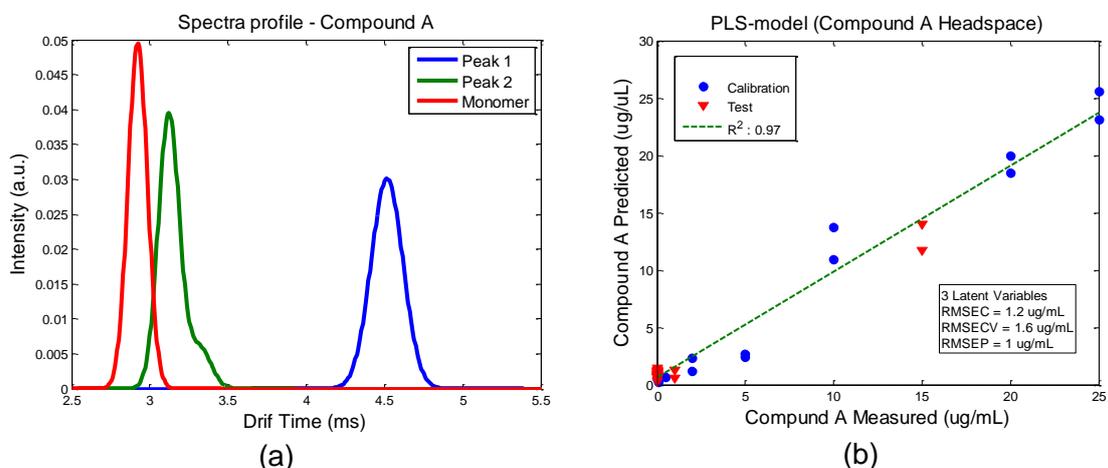


Figura XV Resultados de MCR Lasso cuando se analiso el compuesto A relacionado con PCA (a) perfil de espectros y (b) modelo de calibración.

V. Conclusiones

En esta tesis se abordaron diferentes temas, pero con un objetivo en común que es el desarrollo de nuevas metodologías en el procesado de señal multivariante en espectros IMS. A continuación un breve resumen de las conclusiones más importantes en esta tesis.

- i. Se realizó una comparación entre tres espectrómetros comerciales para evaluar su rendimiento. En este caso se realizó un estudio con 3 aminas y se determinó el límite de detección. Los resultados mostraron que los 3 espectrometros tuvieron un rendimiento similar, a pesar de que sus condiciones de operación son distintas. Se debe aclarar que se realizó una técnica robusta de preprocesado y procesado multivariante. En ningún caso se puede garantizar resultados similares al utilizar técnicas univariantes.
- ii. Se propuso una técnica específica para eliminar ruido de baja frecuencia acoplado al espectro de IMS. Donde se observó que utilizar PCA o ICA mejora notablemente la relación señal ruido si se compara con las técnicas convencionales.
- iii. Con respecto a la necesidad de alinear los picos de los espectros, se explicó los diferentes métodos para el alineamiento. Es aconsejable tener o incluir compuestos de referencia para garantizar que el proceso de alineamiento es el adecuado. En el caso de que esto no fuese posible, se aconseja realizar el alineamiento por etapas, primero un alineamiento en una misma muestra, y luego entre muestras.
- iv. Se realizaron modelos cualitativos para diferenciar o discriminar clases. Se propusieron dos modelos multivariantes con técnicas de validación cruzada. Los resultados obtenidos muestran el gran potencial de IMS como técnica diagnóstica o monitoreo.
- v. Se evaluó el rendimiento del IMS en análisis cuantitativos al utilizar métodos multivariantes y fue comparado con métodos tradicionales como lo es el cálculo de área o altura de picos. De los resultados obtenidos se observó que los modelos univariantes no son capaces de resolver comportamientos típicos de IMS como son el comportamiento no lineal y el efecto en mezclas. En este sentido las técnicas multivariantes mostraron mejores prestaciones y resultados. Se comparó la utilización de técnicas multivariantes que proyectan los datos en un nuevo subespacio como lo es PLS con técnicas de deconvolucion como lo es MCR en sus dos versiones ALS y Lasso. Los resultados obtenidos fueron bastante similares, sin embargo MCR ofrece una ventaja importante ya que permite interpretar de mejor manera los resultados.

VI. Referencias

- 3QBD. *3QBD Israel* [Online]. Available: www.3qbd.com/English/ 2012].
- Airsense. 2012. *GDA2, Airsense Germany* [Online]. Available: www.airsense.com/en/products/gda-2/.
- Alizadeh, N., Mohammadi, A. & Tabrizchi, M. 2008. Rapid screening of methamphetamines in human serum by headspace solid-phase microextraction using a dodecylsulfate-doped polypyrrole film coupled to ion mobility spectrometry. *Journal of Chromatography A*, 1183, 21-28.
- Alonso, R., Rodriguez-Estevez, V., Dominguez-Vidal, A., Ayora-Canada, M. J., Arce, L. & Valcarel, M. 2008. Ion mobility spectrometry of volatile compounds from Iberian pig fat for fast feeding regime authentication. *Talanta*, 76, 591-596.
- Armenta, S., Alcalá, M. & Blanco, M. 2011. A review of recent, unconventional applications of ion mobility spectrometry (IMS). *Analytica Chimica Acta*, 703, 114-123.
- Armenta, S. & Blanco, M. 2012. Ion mobility spectrometry for monitoring diamine oxidase activity. *Analyst*, 137, 5891-5897.
- BBCnews. 2013. *Lung infection identified using 'breath-print'* [Online]. Available: www.bbc.co.uk/news/health-20975948 2014].
- Borsdorf, H., Mayer, T., Zarejousheghani, M. & Eiceman, G. A. 2011. Recent Developments in Ion Mobility Spectrometry. *Applied Spectroscopy Reviews*, 46, 472-521.
- Borsdorf, H., Neitsch, K., Eiceman, G. A. & Stone, J. A. 2009. A comparison of the ion chemistry for mono-substituted toluenes and anilines by three methods of atmospheric pressure ionization with ion mobility spectrometry. *Talanta*, 78, 1464-1475.
- Borsdorf, H., Rammler, A., Schulze, D., Boadu, K. O., Feist, B. & Weiss, H. 2001. Rapid on-site determination of chlorobenzene in water samples using ion mobility spectrometry. *Analytica Chimica Acta*, 440, 63-70.
- Borsdorf, H., Stone, J. A. & Eiceman, G. A. 2005. Gas phase studies on terpenes by ion mobility spectrometry using different atmospheric pressure chemical ionization techniques. *International Journal of Mass Spectrometry*, 246, 19-28.
- Bota, G. M. & Harrington, P. B. 2006. Direct detection of trimethylamine in meat food products using ion mobility spectrometry. *Talanta*, 68, 629-635.
- Buxton, T. L. & Harrington, P. D. 2001. Rapid multivariate curve resolution applied to identification of explosives by ion mobility spectrometry. *Analytica Chimica Acta*, 434, 269-282.
- Chaim, W., Karpas, Z. & Lorber, A. 2003. New technology for diagnosis of bacterial vaginosis. *European Journal of Obstetrics Gynecology and Reproductive Biology*, 111, 83-87.
- Comon, P. 1994. INDEPENDENT COMPONENT ANALYSIS, A NEW CONCEPT. *Signal Processing*, 36, 287-314.
- Copete, M. L., Zalacain, A., Lorenzo, C., Carot, J. M., Esteve, M. D., Climent, M. & Salinas, M. R. 2009. Haloanisole and halophenol contamination in Spanish aged red wines. *Food Additives and Contaminants Part a-Chemistry Analysis Control Exposure & Risk Assessment*, 26, 32-38.
- Cover, T. M. & Hart, P. E. 1967. NEAREST NEIGHBOR PATTERN CLASSIFICATION. *Ieee Transactions on Information Theory*, 13.
- de Juan, A. & Tauler, R. 2006. Multivariate curve resolution (MCR) from 2000: Progress in concepts and applications. *Critical Reviews in Analytical Chemistry*, 36, 163-176.
- deJuan, A., VanderHeyden, Y., Tauler, R. & Massart, D. L. 1997. Assessment of new constraints applied to the alternating least squares method. *Analytica Chimica Acta*, 346, 307-318.
- Di Natale, C., Macagnano, A., Martinelli, E., Paollesse, R., D'Arcangelo, G., Roscioni, C., Finazzi-Agro, A. & D'Amico, A. 2003. Lung cancer identification by the analysis of breath by

- means of an array of non-selective gas sensors. *Biosensors & Bioelectronics*, 18, 1209-1218.
- Efron, B. 1979. 1977 RIETZ LECTURE - BOOTSTRAP METHODS - ANOTHER LOOK AT THE JACKKNIFE. *Annals of Statistics*, 7, 1-26.
- Eiceman, G. A. 1991. ADVANCES IN ION MOBILITY SPECTROMETRY - 1980-1990. *Critical Reviews in Analytical Chemistry*, 22, 17-36.
- Eiceman, G. A., Blyth, D. A., Shoff, D. B. & Snyder, A. P. 1990. SCREENING OF SOLID COMMERCIAL PHARMACEUTICALS USING ION MOBILITY SPECTROMETRY. *Analytical Chemistry*, 62, 1374-1379.
- Eiceman, G. A. & Karpas, Z. 2005. *Ion Mobility Spectrometry*, Florida, Taylor & Francis Group.
- Eiceman, G. A., Karpas, Z. & Hill, H. H. J. 2014. *Ion Mobility Spectrometry*, 6000 Broken Sound Parkway NW, Suite 300.
- Eiceman, G. A., Sowa, S., Lin, S. & Bell, S. E. 1995. Ion mobility spectrometry for continuous on-site monitoring of nicotine vapors in air during the manufacture of transdermal systems. *Journal of Hazardous Materials*, 43, 13-30.
- Eiceman, G. A. & Stone, J. A. 2004. Ion mobility spectrometers in national defense. *Analytical Chemistry*, 76, 390A-397A.
- Ewing, R. G., Atkinson, D. A., Eiceman, G. A. & Ewing, G. J. 2001. A critical review of ion mobility spectrometry for the detection of explosives and explosive related compounds. *Talanta*, 54, 515-529.
- Felsenstein, J. 1985. CONFIDENCE-LIMITS ON PHYLOGENIES - AN APPROACH USING THE BOOTSTRAP. *Evolution*, 39, 783-791.
- Garrido-Delgado, R., Arce, L., Guaman, A. V., Pardo, A., Marco, S. & Valcarcel, M. 2011. Direct coupling of a gas-liquid separator to an ion mobility spectrometer for the classification of different white wines using chemometrics tools. *Talanta*, 84, 471-479.
- Garrido-Delgado, R., Arce, L. & Valcarcel, M. 2012. Multi-capillary column-ion mobility spectrometry: a potential screening system to differentiate virgin olive oils. *Analytical and Bioanalytical Chemistry*, 402, 489-498.
- GAS. G.A.S GERMANY [Online]. Available: www.gas-dortmund.de/Products/UV-IMS/1.336.html.
- Guaman, A. V., Carreras, A., Calvo, D., Agudo, I., Navajas, D., Pardo, A., Marco, S. & Farre, R. 2012. Rapid detection of sepsis in rats through volatile organic compounds in breath. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences*, 881-82, 76-82.
- Hannum, D. W., Linker, K. L., Rhykerd, C. L. & Parmeter, J. E. 2000. Miniaturized explosives preconcentrators for use in man-portable explosives detection systems. *34th Annual 2000 International Carnahan Conference on Security Technology, Proceedings*, 222-227.
- Holmberg, L. 2010. Wine Fraud. *International Journal of Wine Research*, 2, 105-113.
- Jafari, M. T., Saraji, M. & Sherafatmand, H. 2011. Electrospray ionization-ion mobility spectrometry as a detection system for three-phase hollow fiber microextraction technique and simultaneous determination of trimipramine and desipramine in urine and plasma samples. *Analytical and Bioanalytical Chemistry*, 399, 3555-3564.
- Karpas, Z. 2013. Applications of ion mobility spectrometry (IMS) in the field of foodomics. *Food Research International*, 54, 1146-1151.
- Karpas, Z., Guaman, A. V., Calvo, D., Pardo, A. & Marco, S. 2012. The potential of ion mobility spectrometry (IMS) for detection of 2,4,6-trichloroanisole (2,4,6-TCA) in wine. *Talanta*, 93, 200-205.
- 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. 2002. Determination of volatile biogenic amines in muscle food products by ion mobility spectrometry. *Analytica Chimica Acta*, 463, 155-163.
- Kim, S. H., Betty, K. R. & Karasek, F. W. 1978. MOBILITY BEHAVIOR AND COMPOSITION OF HYDRATED POSITIVE REACTANT IONS IN PLASMA CHROMATOGRAPHY WITH NITROGEN CARRIER GAS. *Analytical Chemistry*, 50, 2006-2012.
- Lu, Y., O'Donnell, R. M. & Harrington, P. B. 2009. Detection of cocaine and its metabolites in urine using solid phase extraction-ion mobility spectrometry with alternating least squares. *Forensic Science International*, 189, 54-59.
- Marquez-Sillero, I., Aguilera-Herrador, E., Cardenas, S. & Valcarcel, M. 2011a. Ion-mobility spectrometry for environmental analysis. *Trac-Trends in Analytical Chemistry*, 30, 677-690.
- Marquez-Sillero, I., Cardenas, S. & Valcarcel, M. 2011b. Direct determination of 2,4,6-trichloroanisole 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.
- Owlstone, G. 2014. V-OVG Calibration Gas Generator.
- Pauling, L., Robinson, A. B., Teranish, R & Cary, P. 1971. QUANTITATIVE ANALYSIS OF URINE VAPOR AND BREATH BY GAS-LIQUID PARTITION CHROMATOGRAPHY. *Proceedings of the National Academy of Sciences of the United States of America*, 68, 2374-&.
- Pereira, C. S., Marques, J. J. F. & San Romao, M. V. 2000. Cork taint in wine: Scientific knowledge and public perception - A critical review. *Critical Reviews in Microbiology*, 26, 147-162.
- Perl, T., Carstens, E., Hirn, A., Quintel, M., Vautz, W., Nolte, J. & Junger, M. 2009. Determination of serum propofol concentrations by breath analysis using ion mobility spectrometry. *British Journal of Anaesthesia*, 103, 822-827.
- Persaud, K. & Dodd, G. 1982. ANALYSIS OF DISCRIMINATION MECHANISMS IN THE MAMMALIAN OLFACTORY SYSTEM USING A MODEL NOSE. *Nature*, 299, 352-355.
- Phillips, M., Cataneo, R. N., Condos, R., Erickson, G. A. R., Greenberg, J., La Bombardi, V., Munawar, M. I. & Tietje, O. 2007. Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis*, 87, 44-52.
- Phillips, M., Herrera, J., Krishnan, S., Zain, M., Greenberg, J. & Cataneo, R. N. 1999. Variation in volatile organic compounds in the breath of normal humans. *Journal of Chromatography B*, 729, 75-88.
- 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.
- Pomareda, V., Guaman, A. V., Mohammadnejad, M., Calvo, D., Pardo, A. & Marco, S. 2012. Multivariate curve resolution of nonlinear ion mobility spectra followed by multivariate nonlinear calibration for quantitative prediction. *Chemometrics and Intelligent Laboratory Systems*, 118, 219-229.
- Pudil, P., Novovicová, J. & Kittler, J. 1994. Floating search methods in feature selection. *Pattern Recognition Letters*, 15, 1119-1125.
- Puton, J., Holopainen, S. I., Makinen, M. A. & Sillanpaa, M. E. T. 2012. Quantitative Response of IMS Detector for Mixtures Containing Two Active Components. *Analytical Chemistry*, 84, 9131-9138.
- Puton, J., Nousiainen, M. & Sillanpaa, M. 2008. Ion mobility spectrometers with doped gases. *Talanta*, 76, 978-987.
- Riu, M., Mestres, M., Busto, O. & Guasch, J. 2006. Quantification of chloroanisoles in cork using headspace solid-phase microextraction and gas chromatography with electron capture detection. *Journal of Chromatography A*, 1107, 240-247.

- Rubio-Coque, J. J., Álvarez-Rodríguez, M. L., Goswami, M. & Feltrer-Martínez, R. 2006. *Causes and Origins of Wine Contamination by Haloanisoles (Chloroanisoles and Bromoanisoles)*, Institute of Biotechnology of Leon (INBIOTEC), León, Spain.
- Savitzky, A. & Golay, M. J. E. 1964. SMOOTHING + DIFFERENTIATION OF DATA BY SIMPLIFIED LEAST SQUARES PROCEDURES. *Analytical Chemistry*, 36, 1627-&.
- Shahdousti, P. & Alizadeh, N. 2011. Headspace-solid phase microextraction of selenium(IV) from human blood and water samples using polypyrrole film and analysis with ion mobility spectrometry. *Analytica Chimica Acta*, 684, 67-71.
- Siegel, J., Allison, J., Mohr, D. & Dunn, J. 2005. The use of laser desorption/ionization mass spectrometry in the analysis of inks in questioned documents. *Talanta*, 67, 425-429.
- Smit, S., van Breemen, M. J., Hoefsloot, H. C. J., Smilde, A. K., Aerts, J. M. F. G. & de Koster, C. G. 2007. Assessing the statistical validity of proteomics based biomarkers. *Analytica Chimica Acta*, 592, 210-217.
- 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.
- Statheropoulos, M., Pappa, A., Karamertzanis, P. & Meuzelaar, H. L. C. 1999. Noise reduction of fast, repetitive GC/MS measurements using principal component analysis (PCA). *Analytica Chimica Acta*, 401, 35-43.
- Synder, A. P., Maswadeh, W. M., Eiceman, G. A., Wang, Y.-F. & Bell, S. E. 1995. Multivariate statistical analysis characterization of application-based ion mobility spectra. *Analytica Chimica Acta*, 316, 1-14.
- Tabrizchi, M., Khayamian, T. & Taj, N. 2000. Design and optimization of a corona discharge ionization source for ion mobility spectrometry. *Review of Scientific Instruments*, 71, 2321-2328.
- Tomasi, G., Savorani, F. & Engelsen, S. B. 2011. icoshift: An effective tool for the alignment of chromatographic data. *Journal of Chromatography A*, 1218.
- Van den Velde, S., Nevens, F., Van Hee, P., van Steenberghe, D. & Quirynen, M. 2008. GC-MS analysis of breath odor compounds in liver patients. *Journal of Chromatography B- Analytical Technologies in the Biomedical and Life Sciences*, 875, 344-348.
- wineclub. 2010. *How to Conduct Comparative and Blind Wine Tastings* [Online]. Available: <http://www.wineclub.org/2010/10/comparative-blind-wine-tastings/>.
- Wold, S., Kettanehwold, N. & Skagerberg, B. 1989. NONLINEAR PLS MODELING. *Chemometrics and Intelligent Laboratory Systems*, 7.
- Wold, S., Ruhe, A., Wold, H. & Dunn, W. J. 1984. THE COLLINEARITY PROBLEM IN LINEAR-REGRESSION - THE PARTIAL LEAST-SQUARES (PLS) APPROACH TO GENERALIZED INVERSES. *Siam Journal on Scientific and Statistical Computing*, 5.
- Zalacain, A., Alonso, G. L., Lorenzo, C., Iniguez, M. & Salinas, M. R. 2004. Stir bar sorptive extraction for the analysis of wine cork taint. *Journal of Chromatography A*, 1033, 173-178.