



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA  
Ambiental – Agua

CUANTIFICACIÓN RÁPIDA DE PRODUCTOS DE LA FERMENTACIÓN EN UN  
BIORREACTOR ANAEROBIO PRODUCTOR DE HIDRÓGENO

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRO EN INGENIERÍA

PRESENTA:  
Gerardo Muñoz Montoya

TUTOR PRINCIPAL  
Dr. Alejandro Vargas Casillas, Instituto de Ingeniería

COMITE TUTOR  
Dra. Genoveva Hernández Padrón, CFATA  
Dr. Durán Moreno Alfonso, Facultad de Química

MÉXICO, D. F. septiembre de 2013

**JURADO ASIGNADO:**

Presidente: Dr. Durán Moreno Alfonso

Secretario: Dr. Rodríguez Morales Ángel Luis

Vocal: Dra. Hernández Padrón Genoveva

1<sup>er</sup> Suplente: Dr. Buitrón Méndez Germán

2<sup>do</sup> Suplente: Dr. Vargas Casillas Alejandro

Esta tesis fue realizada en el Laboratorio de Investigación en Procesos Avanzados de Tratamiento de Aguas en la Unidad Académica Juriquilla del Instituto de Ingeniería de la Universidad Nacional Autónoma de México. Se agradece al CONACYT por la beca y por la beca mixta. Se agradece también el financiamiento parcial por parte del proyecto DGAPA-PAPIIT IN116209.

**TUTOR DE TESIS:**

Alejandro Vargas Casillas

-----  
**FIRMA**

## Contenido

Resumen.....	1
Abstract .....	2
Capítulo 1. Introducción.....	3
Capítulo 2. Objetivos y Metas.....	6
2.1 Objetivo General .....	6
2.2 Objetivos Particulares .....	6
2.3 Metas .....	6
Capítulo 3. Antecedentes.....	7
3.1 Digestión anaerobia.....	7
3.2 Producción de hidrógeno a partir de ácidos grasos volátiles.....	8
3.2.1 Ácidos grasos volátiles como un indicador .....	9
3.3 Técnicas de medición .....	9
3.4 Espectroscopia infrarroja.....	11
3.4.1 Aspectos fundamentales.....	12
3.4.2 Asignación de bandas.....	15
3.4.3 Espectrómetro FT-MIR .....	16
3.4.3.1 Reflectancia Total Atenuada (ATR).....	18
3.5 Ventajas y desventajas de FT-MIR .....	19
3.6 FT-MIR en la medición de ácidos grasos volátiles .....	20
3.7 Técnicas de calibración múltiple.....	21
3.7.1 Etapas del proceso de modelado .....	22
3.7.2 Selección de muestras para la calibración .....	23
3.7.3 Validación cruzada - dejando una fuera a la vez (leave-one-out) .....	23
3.7.4 Creación del modelo de calibración .....	23

3.7.5 Validación del modelo de calibración .....	25
3.8 Regresión por mínimos cuadrados parciales ( <i>PLS</i> ).....	26
Capítulo 4. Materiales y métodos.....	29
4.1 Estrategia experimental .....	29
4.2 Toma y tratamiento de la muestra .....	30
4.2.1 Medición de AGV's por cromatografía.....	31
4.2.2 Medición de AGV's por Espectroscopia FT-MIR .....	32
4.2.2.1 Algoritmo de corrección.....	33
4.3 Calibración .....	34
4.4 Validación .....	34
Capítulo 5. Resultados y discusión .....	35
5.1 Rapidez y consumibles.....	35
5.2 Adquisición y análisis espectral .....	37
5.2.1 Representatividad del espectro.....	41
5.3 Robustez analítica .....	44
5.3.1 Valores de referencia.....	44
5.3.2 Análisis cuantitativo: modelo PLS .....	47
Capítulo 6. Conclusiones .....	56
Capítulo 7. Referencias bibliográficas .....	57

## Índice de Figuras

Figura 3.1 Etapas de digestión anaerobia.....	7
Figura 3.2 Regiones del espectro electromagnético.....	12
Figura 3.3 Representación gráfica de la ley de Beer.....	14
Figura 3.4 Frecuencias de absorción de los grupos más comunes en el MIR.....	16
Figura 3.5 Principales componentes ópticos de un equipo FT-MIR.....	17
Figura 3.6 Esquema de la técnica de reflectancia total atenuada.....	18
Figura 4.1 Estrategia experimental.....	29
Figura 4.2 Tratamiento de la muestra.....	31
Figura 4.3 Procesamiento de espectros.....	32
Figura 4.4 Algoritmo de corrección de espectros.....	33
Figura 5.1 El espectro de los ácidos grasos medidos.....	38
Figura 5.2 Espectros de muestras de ácidos grasos volátiles sin ningún filtro espectroscópico.....	39
Figura 5.3 Comparación de espectros FT-MIR.....	40
Figura 5.4 Espectros sometidos al algoritmo de corrección.....	42
Figura 5.5 Evaluación de representatividad de los espectros en $1650\text{ cm}^{-1}$ .....	43
Figura 5.6 Distribución de las concentraciones de AGV's.....	46
Figura 5.7 Error medio de predicción por el número de variables latentes y porcentaje de varianza explicada en Y (concentración) de acuerdo al número de variables latentes.....	48
Figura 5.8 Gráficas de las concentraciones medidas y centradas contra las concentraciones estimadas por cada modelo de regresión.....	49
Figura 5.9 Gráfica de <i>loadings</i> de las cinco primeras variables latentes correspondientes a la región completa de los espectros para HAc y HBU.....	51
Figura 5.10 Superposición del gráfico de <i>loadings</i> de los primeros cinco componentes y los espectros de HAc y HBU respectivamente.....	52

Figura 5.11 Conjunto espectral utilizado en la calibración del modelo PLS.....	53
Figura 5.12 Comparación de espectros por tipo de proceso en el biorreactor.....	54

## **Índice de Tablas**

Tabla 3.1 Comparación de las técnicas más utilizadas de medición de AGV's.....	10
Tabla 3.2 Ventajas y desventajas de la medición por FT- MIR de AGV's.....	19
Tabla 3.3 Algoritmo PLS-1.....	28
Tabla 4.1 Características operacionales de los reactores.....	30
Tabla 5.1 Comparación de las técnicas de cromatografía y FT-MIR.....	36
Tabla 5.2 Parámetros estadísticos descriptivos para los AGV's producidos por los biorreactores con diferentes alimentaciones.....	45
Tabla 5.3 Parámetros estadísticos descriptivos para los AGV's en el conjunto de datos de calibración.....	46
Tabla 5.4 Resumen de los resultados de acuerdo al número de variables latentes utilizado.....	50





## Resumen

La concentración de ácidos grasos volátiles es uno de los principales parámetros de seguimiento como intermediario del proceso de generación de hidrógeno en un reactor fermentativo. Tradicionalmente la concentración de ácidos grasos volátiles, se determina mediante procedimientos cromatográficos, que requieren de elevadas tasas de gases consumibles, altas temperaturas y largos tiempos de medición, calentamiento y enfriamiento. La espectroscopia infrarroja de región media con transformada de Fourier (FT-MIR) combinada con el análisis de calibración multivariada de mínimos cuadrados parciales (PLS), es una técnica versátil, eficaz y exacta para el cálculo concentraciones, ampliamente utilizada en quimiometría, características que la convierten en una alternativa de cuantificación para un rápido seguimiento del proceso.

El objetivo de este trabajo se basa en el desarrollo de un método alternativo para la predicción de la concentración de ácidos grasos volátiles empleando la espectroscopia FT-MIR y el método de calibración multivariada de mínimos cuadrados parciales (PLS). En la calibración, se utilizaron 46 muestras de efluente obtenido de cuatro reactores diferentes, dos alimentados con glucosa, una con residuos sólidos urbanos y uno más con residuos de la industria tequilera (vinaza). Cada una de las concentraciones para las muestras fue medida por cromatografía como método de referencia. Se obtuvo el espectro FT-MIR de cada una de las muestras mediante un algoritmo de corrección que asegura su repetitividad y disminuye el ruido, estandarizando y correlacionando los espectros con el valor de concentración obtenido por el cromatógrafo. Se utilizó el método PLS de calibración multivariada. Los factores que se consideraron para determinar el número óptimo de variables latentes de los modelos fueron los errores de calibración mediante la validación cruzada y la capacidad de ajuste predictivo. Un estudio puntual de cada una de las variables latentes se realizó relacionando los *loadings* con las contribuciones espectrales. El método se mostró eficaz al reducir 5 veces los tiempos de cuantificación de ácidos grasos volátiles (AGV's); sin embargo, la robustez estadística al utilizar un número pequeño de variables latentes es aún reticente.

## **Abstract**

The concentration of volatile fatty acids is one of the key monitoring parameters as an intermediary in the hydrogen generation process in a fermentation reactor. Traditionally, this concentration is determined by chromatographic methods, which require high levels of consumable gases, high temperatures and long measurement times, heating and cooling. The middle region in Fourier transform infrared spectroscopy (FT-MIR) analysis combined with multivariate calibration partial least squares (PLS) is a versatile, efficient and accurate technique for calculating concentrations, widely used in chemometrics. These characteristics make it suitable for rapid quantification in process monitoring.

The main objective in this work was to develop an alternative method for the prediction of the concentration of volatile fatty acids using FT-MIR spectroscopy and the multivariate calibration method of partial least squares (PLS). For calibration, we used 46 samples of effluent obtained from four different bioreactors, two fed with glucose, one fed with solid waste and another one fed with waste of the tequila industry (vinasse). Each concentration for the samples was measured by chromatography as a reference method. The FT-MIR spectra of each sample were obtained using a correction algorithm which ensures repeatability and reduces noise, standardizing the spectrum and correlating it with the concentration value obtained from the chromatograph. The PLS method was used for multivariate calibration. The factors considered in determining the optimum number of latent variables in the models were the calibration errors and cross validation by adjusting the predictive capability. A punctual study of each latent variable was done by relating “loadings” with the spectral contributions. The method was effective, reducing in five times the time needed for quantification of AGV's; however, statistical robustness when using a small number of latent variables is still reluctant.

## Capítulo 1. Introducción

Los nuevos procesos de producción de hidrógeno a partir de la degradación de contaminantes en aguas residuales utilizando el metabolismo de los microorganismos anaerobios, evolucionan hacia una alternativa energética prominente. (*Debabrata, 2001*) Debido a que el uso del hidrógeno como vector energético, ha mostrado alto rendimiento en comparación con los combustibles hidrocarbonados (122 kJ/g, 2.75 veces más alto que los combustibles hidrocarbonados), siendo el agua el único producto resultante de su combustión (*Han & Shin, 2004*). Sin embargo, este proceso, muestra una frágil dependencia hacia el control y el seguimiento de las variables involucradas como intermediarios en su evolución. De tal modo que puede volverse ineficaz. Así, en la búsqueda de un mejor dominio de la producción de hidrógeno por digestión anaerobia, se busca obtener información a través de estos diferentes parámetros y variables que comprometen al sistema. Los ácidos grasos volátiles (AGV's) son los principales intermediarios en un biorreactor anaerobio y por esta razón han sido utilizados como parámetros de control indicativos de la actividad de los consorcios microbianos en el desempeño de un reactor (*Madsen et al., 2011*). Las mediciones de AGV's tienen un impacto significativo en la toma de decisiones que intervienen directamente en la optimización de procesos de obtención de hidrógeno, por lo que se espera que se realicen en tiempos despreciables y cuenten con la suficiente robustez de los métodos tradicionales.

En particular este tipo de análisis se llevan a cabo por métodos cromatográficos de alta sensibilidad, pero que requieren tiempos significativos de análisis, preparación de la muestra y gran cantidad de consumibles (*Mazarevica et al., 2004*). La espectroscopia infrarroja de región media con transformada de Fourier (FT-MIR) se muestra como una técnica alternativa que ofrece grandes ventajas como: adquisición de espectros en pocos minutos, obtención de espectros con muestras de volúmenes despreciables, que requieren una preparación mínima o nula, además que puede fácilmente ser escalable a un desarrollo en línea (*Steyer et al., 2002*).

El desarrollo de este tipo de procedimientos alternativos ha sido posible gracias al avance alcanzado en los últimos años por la quimiometría. La cual es una nueva disciplina de la química analítica que emplea métodos matemáticos, estadísticos y de lógica formal para extraer información química relevante mediante un adecuado análisis de los datos experimentales. (*Dago-Morales et al., 2005*).

Entre los complejos procedimientos matemáticos y estadísticos que la quimiometría ha desarrollado se encuentran los métodos de calibración multivariada. La esencia de estos métodos consiste en modelar la relación entre una propiedad físico-química univariada, como por ejemplo: la concentración de un analito y una serie multivariada de datos experimentales del sistema objeto de estudio (*Brereton 2000*). En una etapa inicial se registran los espectros y cromatogramas de un conjunto de muestras patrones (muestras de calibración), a las cuales se les determina el valor de la propiedad físico-química de interés por valores de referencia. Posteriormente, mediante la aplicación de análisis de regresión multivariada se obtiene un modelo matemático que permite correlacionar el valor de la propiedad físico-química con los espectros o cromatogramas. Finalmente, se utiliza el modelo para predecir el valor de la propiedad físico-química o determinar las concentraciones de los analitos de muestras problemas.

La técnica de calibración robusta PLS (*Partial Least Squares*) ha sido utilizada con éxito para la predicción de la concentración de AGV's (*Uden & Sjaunja, 2009; Steyer, 2002*). Este tipo de modelo funciona como herramienta estadística que permite el análisis de mezclas en el espectro de absorción. El método PLS requiere una base de datos de calibración y una de validación para la construcción del modelo (*Adams, 2004*). Tomando la información más relevante como una combinación lineal de las bandas espectrales. Para tal fin, un número representativo de muestras son requeridas para compensar la variabilidad. Uno de los aspectos fundamentales para el correcto uso del método estadístico PLS es la selección de muestras que constituyen el grupo de calibración. Estas muestras deben ser representativas del modelo que se pretende calibrar y deben contemplar la composición química del sistema y sus posibles variaciones. Otro aspecto fundamental de la técnica es representado por los valores de referencia utilizados en el proceso de calibración. En el caso particular de mezclas complejas, resulta difícil la preparación de

patrones de referencia, por lo tanto, la calibración debe realizarse con muestras reales tomadas del propio proceso productivo.

El método PLS en comparación, tiene las ventajas de procesar grandes cantidades de datos experimentales y la posibilidad de detectar en el proceso de calibración las muestras fuera de intervalo; generalmente esto se lleva a cabo a través de procedimientos de validación cruzada. (*Gemperline, 2006*) Estas ventajas hacen posible que el método PLS logre modelos de calibración mucho más precisos, exactos y robustos (*Brereton 2000*).

Este trabajo se propuso desarrollar un modelo para la predicción del índice de AGV's en biorreactores fermentativos productores de hidrógeno mediante el empleo de la espectroscopia FT-MIR y el método PLS de calibración multivariada.

## Capítulo 2. Objetivos y Metas

### 2.1 Objetivo General

Desarrollar un método alternativo para la cuantificación de concentraciones de ácidos grasos volátiles, empleando espectroscopia FT-MIR y el método de calibración multivariada de mínimos cuadrados parciales (PLS).

### 2.2 Objetivos Particulares

- Obtener, corregir y validar de la base de datos espectral.
- Crear dos bases de datos: espectral (FT-IR) y de correlación (cromatografía).
- Establecer una correcta relación entre base de datos espectral y cromatográfica.
- Validar el método de cuantificación de AGV's propuesto.

### 2.3 Metas

- Elaboración de un algoritmo de corrección para los espectros de AGV's.
- Obtención de muestras de diferentes reactores fermentativos, productores de hidrógeno para su análisis por FT-MIR y cromatografía.
- Desarrollo del software de calibración multivariada.
- Estandarización del sistema de medición de AGV's.

## Capítulo 3. Antecedentes

### 3.1 Digestión anaerobia

La digestión anaerobia es un proceso de mediación microbiológica en ausencia de oxígeno que se caracteriza por la mineralización de la materia orgánica contenida en los biopolímeros y otros compuestos biodegradables en productos gaseosos tales como hidrógeno ( $H_2$ ), metano ( $CH_4$ ), dióxido de carbono ( $CO_2$ ) y sulfuro de hidrógeno ( $H_2S$ ). La producción de biogás se da en diferentes etapas, cada una con un nivel de complejidad particular que involucra una extensa variedad de reacciones. La Figura 3.1 muestra los cuatro procesos secuenciales, dando una descripción general del proceso de digestión.

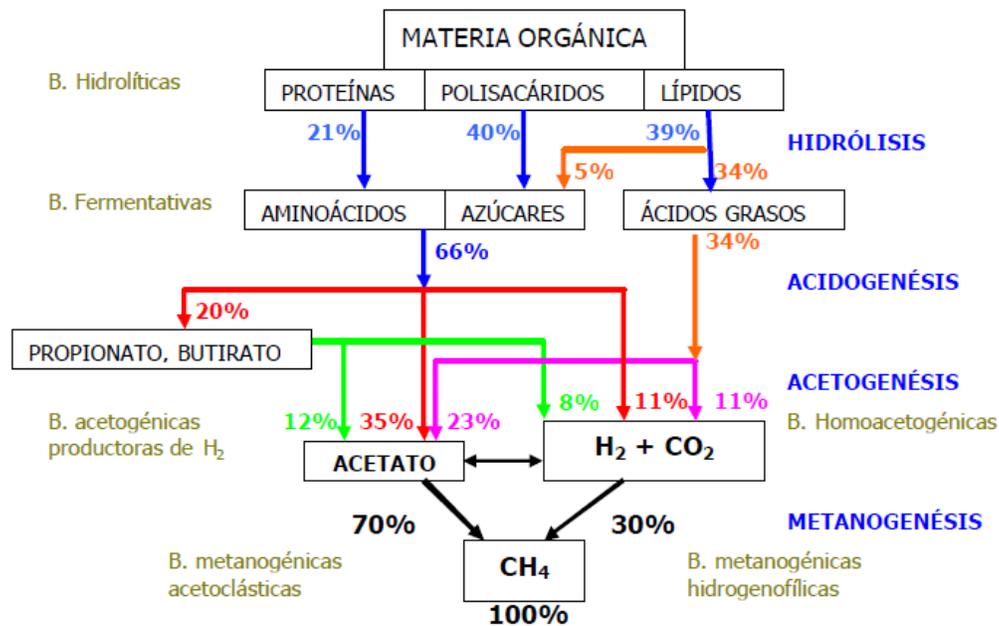


Figura 3.1 Etapas de digestión anaerobia.

El proceso se inicia con la hidrólisis de polisacáridos, proteínas y lípidos por la acción de enzimas extracelulares producidas por las bacterias hidrolíticas. Los productos de esta reacción son moléculas de bajo peso molecular como los azúcares, los aminoácidos, los ácidos grasos y los alcoholes, los cuales son transportados a través de la membrana celular; posteriormente son fermentados a ácidos grasos con bajo número de carbonos como los ácidos acético, fórmico, propiónico y butírico, así como compuestos reducidos: etanol, H<sub>2</sub> y CO<sub>2</sub>. Los productos de fermentación son convertidos a acetato, H<sub>2</sub> y CO<sub>2</sub> por la acción de las bacterias fermentativas, las cuales son conocidas como bacterias acetogénicas productoras de hidrógeno.

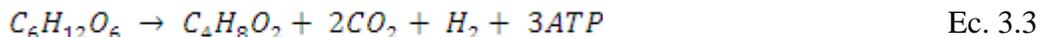
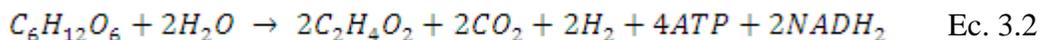
Finalmente las bacterias metanogénicas convierten el acetato a CH<sub>4</sub> y CO<sub>2</sub>, o reducen el CO<sub>2</sub> a metano. Estas transformaciones involucran dos grupos metanogénicos que son los encargados de llevar a cabo las transformaciones mencionadas anteriormente: acetotróficas e hidrógeno tróficas (*Madigan et al., 2003*).

### 3.2 Producción de hidrógeno a partir de ácidos grasos volátiles

En las fermentaciones, el equilibrio de electrones se mantiene por la producción de hidrógeno molecular (H<sub>2</sub>). Los protones derivados (H<sup>+</sup>) sirven como aceptores de electrones en la producción de H<sub>2</sub>.



. Realizando un análisis estequiométrico se llega a lo siguiente:



Las ecuaciones (2) y (3) representan la formación de biomasa, acetato y butirato, respectivamente. Este tipo de estudios han demostrado que la vía del acetato permite un mayor rendimiento de hidrógeno que la del butirato, y se ha encontrado que potenciar esta ruta metabólica mejora considerablemente su producción (*Chen et al., 2006*).

### 3.2.1 Ácidos grasos volátiles como un indicador

Aunque una medición de flujo de biogás dará una indicación de la actividad global del reactor, y frecuentemente se utiliza como tal, se busca desarrollar técnicas más específicas que permitan seguir de manera rápida el proceso, y así ayudar a optimizar el rendimiento y control del gas.

Los ácidos grasos volátiles son el intermediario más importante en el proceso de digestión anaerobia (*Vanrollegem y Lee, 2003*). Por esta razón han sido utilizados como parámetros de control indicativos de la actividad de los consorcios anaerobios en el desempeño de un reactor. Aunque su producción es inherente al sistema, no son deseables, ya que altas concentraciones de ácidos grasos volátiles son indicadores de un bajo desempeño en el proceso de producción de biogás, debido a una sobrecarga orgánica o al aumento de pH, acidificando el reactor y provocando la inhibición de las comunidades microbianas (*Madsen et al., 2011*). La concentración de ácidos grasos volátiles puede ser medida por varias técnicas de forma directa, como titulación y cromatografía, o indirectamente, correlacionando un haz infrarrojo con la concentración total de AGV's por espectroscopia (*Steyer et al., 2002*).

### 3.3 Técnicas de medición

Se han desarrollado muchos métodos y tecnologías analíticas que permiten el seguimiento de los procesos complejos de conversión en la digestión anaerobia. Debido a su gran interés y amplia aplicación, la mayoría de las técnicas han sido utilizadas en el seguimiento y

control de los procesos fermentativos; sin embargo, un reto aún mayor es el seguimiento de estos procesos en tiempos despreciables y minimizando la cantidad de reactivos involucrados en la evolución del proceso de seguimiento. Actualmente las tecnologías más utilizadas son la cromatografía y la titulación (*Ward et al., 2010*). La Tabla 1 hace una breve comparación de ambas tecnologías y evidencia sus ventajas y desventajas. Los sistemas de titulación, por ejemplo, no arrojan información detallada acerca de la abundancia relativa de los distintos ácidos grasos volátiles, sino que simplemente buscan la medida de acidez total. No es posible distinguir entre los valores de pKa y la concentración de los ácidos individuales utilizando métodos de valoración de pH. Por otro lado, la cromatografía destaca por su alta sensibilidad, su fácil adaptación a las determinaciones cuantitativas exactas, su idoneidad para la separación de especies no volátiles y su gran aplicabilidad en compuestos de gran interés en los procesos de digestión anaerobia. Sin embargo, el equipo es muy costoso y es inherente el uso de gases, la necesidad de altas temperaturas, tiempos de calentamiento y volatilización, así como el constante cambio de columnas. Estos representan un retraso en la obtención de datos (*Madsen et al., 2011*).

Cromatografía	Titulación
✓ Sensibilidad	✓ Sencillez
✓ Robustez	✓ Bajo costo
✓ Fácil identificación de especies	✓ Escalable -en línea
✗ Uso de un gases	✗ Basada sólo en pH (no en pKa)
✗ Altas presión y temperatura	✗ Identificación de especies
✗ Uso de columnas	✗ Calibración incesante

Tabla 3.1 Comparación de las técnicas más utilizadas de medición de AGV's.

La espectroscopia infrarroja, por otro lado, reúne lo mejor de los dos sistemas y se muestra como una técnica analítica alternativa para la medición de AGV's, que puede fácilmente ser adaptada a sistemas en línea debido a su rapidez y la mínima o nula preparación de la muestra, así como la reducción de espacio físico (*Spanjers et al., 2006*).

### 3.4 Espectroscopia infrarroja

Dado que los habituales métodos de análisis de AGV's resultan poco prácticos para realizar un seguimiento rápido del proceso por la cantidad de recursos necesarios y por el limitado tiempo para llevarlas a cabo, el seguimiento de los procesos de producción de hidrógeno demanda una técnica analítica rápida y versátil que dé solución a este problema.

La técnica de la espectroscopia infrarroja (IR del acrónimo en inglés *infrared*) presenta múltiples ventajas y constituye una alternativa factible a los métodos de análisis convencionales que requieren un tiempo de preparación y muestreo más largos, lo que está suscitando un creciente interés por parte de muchos grupos de investigación.

La espectroscopia infrarroja es una técnica analítica no destructiva que se emplea para determinar cualitativa y cuantitativamente los compuestos moleculares de muestras sólidas, líquidas e incluso gases (*Stuart B., 2004*). Proporciona información estructural de una gran variedad de moléculas de manera simultánea, incluso estando dentro de mezcla complejas. Esta técnica requiere una mínima preparación de muestra, lo que unido a su alta capacidad de automatización, la convierten en una técnica analítica rápida, directa y respetuosa con el medio ambiente, por lo que resulta muy ventajosa para ser utilizada en el control de procesos como análisis de rutina (*Sun D., 2009*).

### 3.4.1 Aspectos fundamentales

Dentro del espectro electromagnético, la región infrarroja como lo muestra la Figura 3.2, abarca desde las  $10 \text{ cm}^{-1}$  a los  $12\,800 \text{ cm}^{-1}$ , expresado en número de onda ( $\omega$ ), o de  $780 \text{ nm}$  (extremo rojo del espectro visible) a  $10^6 \text{ nm}$  (región de las microondas), expresado en longitudes de onda ( $\lambda$ ). La ecuación (3.4) muestra la relación entre  $\omega$  y  $\lambda$ .

$$\omega = 1/\lambda \quad \text{Ec. 3.4}$$

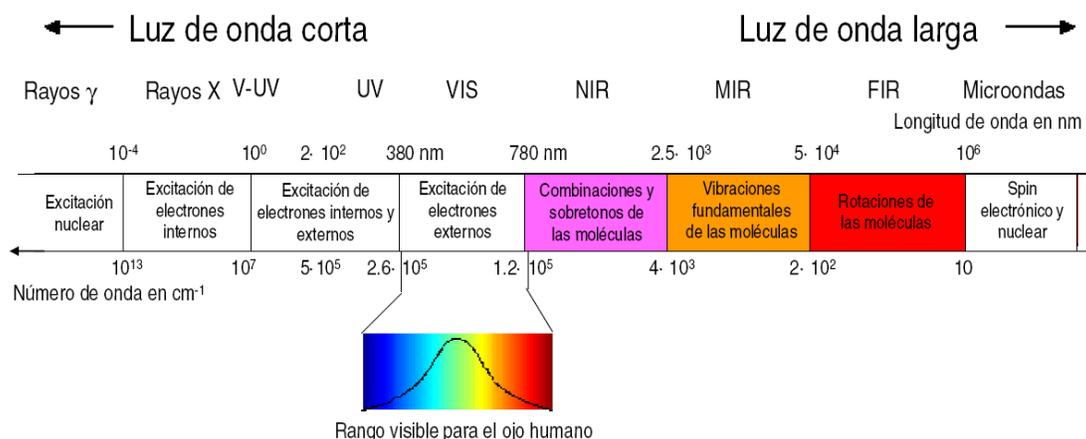


Figura 3.2 Regiones del espectro electromagnético.

La espectroscopia molecular analiza la interacción del espectro electromagnético con la materia. Dependiendo de la región del espectro en la que se mida, y, por tanto, de la energía de la radiación utilizada (caracterizada por su longitud de onda), esta interacción será de diferente naturaleza.

La vibración de los diferentes grupos funcionales en la muestra interaccionando con el haz infrarrojo es la principal responsable de la obtención de un “espectro infrarrojo” que no es más que un gráfico de la intensidad infrarroja en función del número de onda de la luz (proporcional a la frecuencia e inversamente proporcional a la longitud de onda).

Las moléculas tienen frecuencias a las cuales rotan y vibran. Los movimientos de rotación y vibración moleculares tienen niveles de energía discretos o modos vibracionales (la frecuencia a la que un sistema vibra al ser perturbado). Las vibraciones producirán bandas de absorción y dependerá del tipo de molécula de acuerdo a la cantidad de átomos y la longitud entre sí. El número de bandas de absorción infrarrojas se aumenta por las vibraciones armónicas y de combinación (*Smith, 2000*). La absorbancia de la radiación está directamente relacionada con la concentración de compuestos químicos individuales de acuerdo con la ley Bouguer-Beer-Lambert (*Griffiths, 2002*).

$$\frac{P}{P_0} = e^{-\alpha lc} \quad \text{Ec. 3.5}$$

Dónde:

$P$  y  $P_0$  son las intensidades saliente y entrante.

$\alpha$  es el coeficiente de absorción.

$c$  es la concentración.

$l$  es la longitud del trayecto óptico.

Es decir, esta ley explica que hay una relación exponencial entre la transmisión de luz a través de una sustancia y la concentración de la sustancia, así como también entre la transmisión y la longitud del cuerpo que la luz atraviesa como se esquematiza en la Figura 3.3.

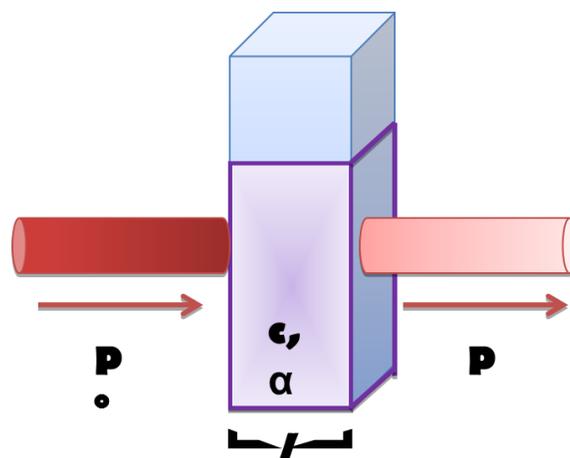


Figura 3.3 Representación gráfica de la ley de Beer.

Desde un punto de vista funcional, el espectro infrarrojo se divide en tres regiones: el infrarrojo cercano, conocida como NIR (del acrónimo en inglés *near infrared*), donde se producen las absorciones debidas a sobretonos y combinaciones de las bandas fundamentales del infrarrojo medio; el infrarrojo medio, conocida como MIR (del acrónimo en inglés *mid infrared*), donde tienen lugar las vibraciones moleculares fundamentales; y el infrarrojo lejano, conocida como FIR (del acrónimo en inglés *far infrared*), donde se producen las absorciones debidas a cambios rotacionales moleculares (*Smith 2000*). Así pues, las transiciones entre dos estados rotacionales sólo es posible observarlas con gran resolución en la región FIR para el caso de muestras gaseosas, por lo que esta región no se emplea para el análisis de muestras líquidas y sólidas. La principal utilidad de la región FIR queda restringida a la determinación estructural (análisis cualitativo) de especies inorgánicas y organometálicas con átomos pesados. Por su parte, las transiciones entre dos estados vibracionales son posibles con las técnicas espectroscópicas NIR y MIR, las cuales son útiles para la determinación cuantitativa de rutina de distintas especies de interés para la industria de los alimentos y bebidas. Ambas técnicas reciben también el nombre de espectroscopia vibracional.

En la región MIR se registran los cambios entre estados de energía vibracional de los principales grupos funcionales característicos de cada molécula, y que permiten elucidar la estructura molecular de un compuesto. Esto hace que esta región sea la más utilizada en el análisis cualitativo de muestras sólidas y líquidas. Sin embargo, la región NIR no produce absorciones específicas que proporcionen información estructural de los analitos, por lo que su aplicación se enfoca fundamentalmente hacia el análisis cuantitativo. De hecho, los espectros NIR son mucho menos nítidos y peor definidos que los MIR, ya que están llenos de bandas superpuestas debido a la existencia de sobretonos y combinaciones de las vibraciones fundamentales de la región MIR, lo que complica su interpretación y calibración (Mazarevica, 2004).

### 3.4.2 Asignación de bandas

En el espectro de absorción MIR existen dos regiones (Silverstein & Webster, 1998):

- Entre  $4000\text{ cm}^{-1}$  y  $1300\text{ cm}^{-1}$  se observan una serie de bandas asignadas a vibraciones de sólo dos átomos. En este caso, la banda de absorción se asocia únicamente a un grupo funcional, por lo que a esta región se la denomina región *de grupo funcional*. Estas vibraciones derivan de grupos que contienen hidrógeno (C-H, O-H y N-H) o grupos con dobles y triples enlaces aislados.
- Entre  $1300\text{ cm}^{-1}$  y  $900\text{ cm}^{-1}$  se producen múltiples bandas correspondientes a las vibraciones de cada enlace dentro de las moléculas. A esta región se la denomina *fingerprint* (huella dactilar o absorción característica).

La región *fingerprint* recibe este nombre porque es la región del MIR donde los compuestos orgánicos generan bandas de absorción características, dependiendo de la estructura de la molécula, permitiendo la identificación de compuestos específicos. La región *fingerprint* es específica y única para cada molécula, mientras que la región *de grupo funcional* es similar para moléculas que posean los mismos grupos funcionales. La

Figura 3.4 muestra un cuadro resumen de las frecuencias de absorción de los grupos funcionales más comunes en el MIR.

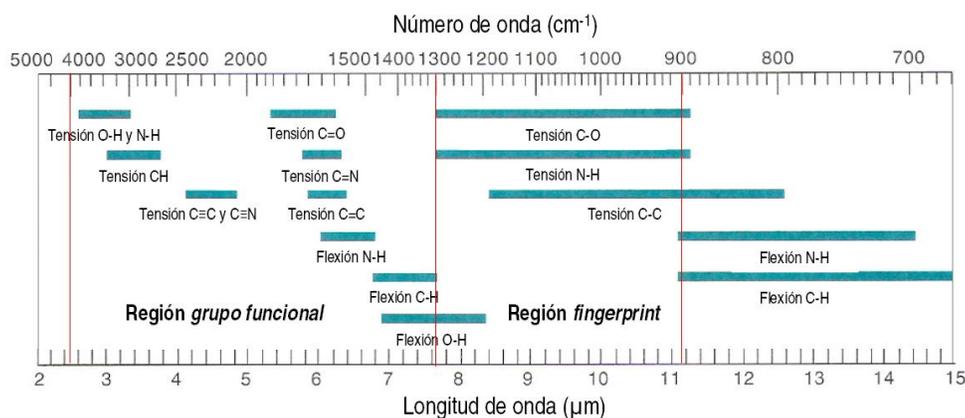


Figura 3.4 Frecuencias de absorción de los grupos más comunes en el MIR.

### 3.4.3 Espectrómetro FT-MIR

Los espectrómetros de infrarrojo con transformada de Fourier (FT-MIR) suponen un importante avance tecnológico respecto a los antiguos espectrómetros dispersivos de doble haz, ya que muestran mayor rapidez de análisis, mayor sensibilidad (relación señal/ruido), mayor precisión y menor costo. Todo esto, junto con los avances de la tecnología informática, ha hecho que la técnica FT-MIR sea reconocida como una potente herramienta analítica. Los principales componentes ópticos de los espectrómetros FT-MIR se muestran en la Figura 3.5. La fuente de radiación IR, el interferómetro, el compartimento donde se coloca la muestra y el detector.

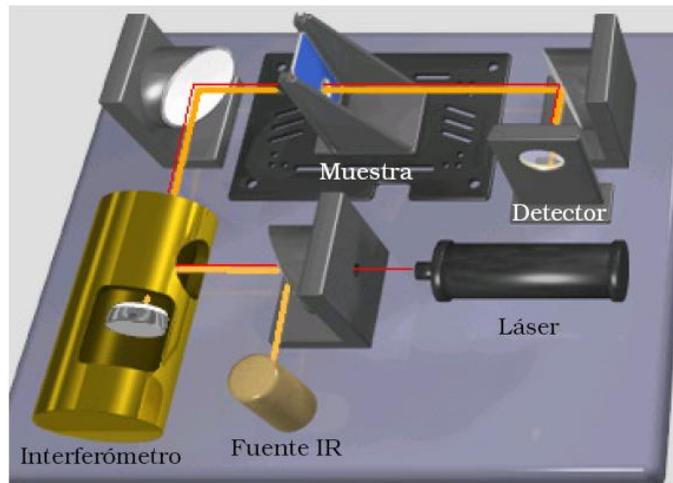


Figura 3.5 Principales componentes ópticos de un equipo FT-MIR (Fragoso 2011).

A modo de resumen, las principales ventajas de los espectrómetros FT-MIR son debidas a la presencia del interferómetro de Michelson. Las más importantes son:

- Se mejora la resolución de los espectros por aumento del recorrido del espejo móvil, lo cual es posible gracias al láser, que es el que controla su posición.
- Se obtiene una mayor sensibilidad (relación señal/ruido) como consecuencia de someter a la muestra a una irradiación simultánea con luz IR de diferentes frecuencias (y no secuencial, como ocurre en el caso de los instrumentos dispersivos), lo que produce una mayor energía del haz de luz que llega al detector, así como por el promedio de varios interferogramas de la misma muestra (barridos) (Chau 2004).

Los métodos cuantitativos de absorción en el infrarrojo proponen una mayor complejidad, debido a una mayor estrechez entre bandas de absorción y a las limitaciones instrumentales de los equipos. Sin embargo, la precisión y exactitud del análisis cuantitativo pueden verse beneficiados con equipo de transformada de Fourier (FT-IR) e instrumentos dispersivos como la reflectancia total atenuada (ATR) (Skoog *et al.*, 2000).

### 3.4.3.1 Reflectancia Total Atenuada (ATR)

La reflexión interna se produce cuando una radiación infrarroja entra en un cristal ATR transmisor y de alto índice de refracción. El cristal está diseñado para permitir una reflexión interna total que crea una onda evanescente sobre la superficie del cristal. Esta onda se extiende a la muestra que se mantiene en contacto íntimo con el cristal, registrándose un cambio de densidad que produce una refracción en el haz infrarrojo. Este fenómeno se ilustra en la Figura 3.6.

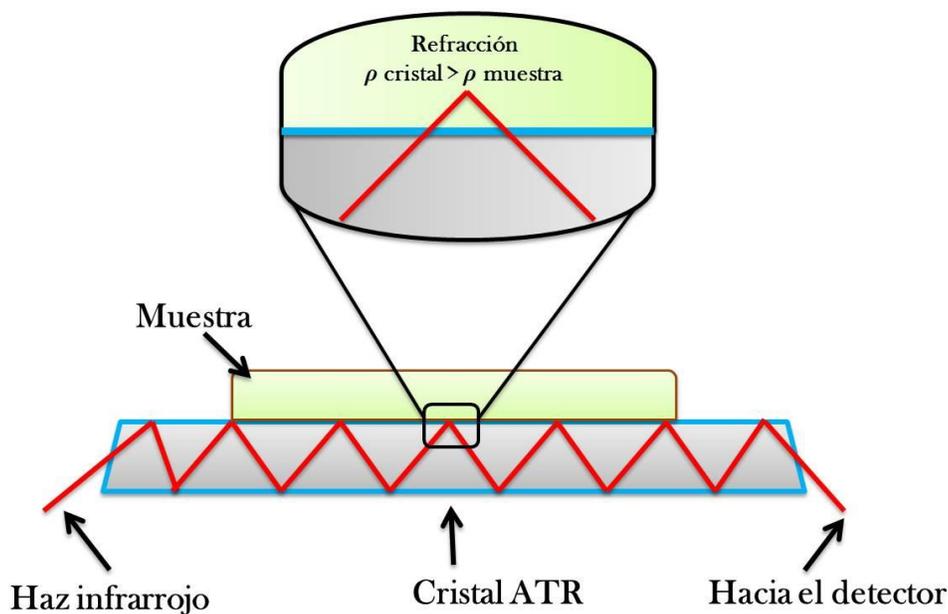


Figura 3.6 Esquema de la técnica de reflectancia total atenuada.

La profundidad de penetración de la onda evanescente dentro de la muestra es función del material del cristal y del ángulo de incidencia del haz. A menor índice de refracción y menor ángulo de incidencia, mayor penetración se consigue. La profundidad de penetración también se incrementa con la longitud de onda. Su principal ventaja práctica es la nula preparación de la muestra, lo que hace el uso de este instrumento versátil y eficiente para la medición rápida de líquidos, geles y pastas. Sin embargo, existe un ligero

desplazamiento de las bandas originales debido al uso de esta técnica. (Koulis, et. al., 2001).

### 3.5 Ventajas y desventajas de FT-MIR

El hecho de poder realizar los análisis *in-situ* mediante espectroscopia FT-MIR proporciona una serie de ventajas que se enlistan en la Tabla 3.2.

Ventajas	Desventajas
No destructiva, ni invasiva	Técnica de referencia obligada
Posibilidad de medir en estado sólido como líquido	No es posible analizar muestras que presenten una variación en la base de calibración
Manipulación mínima o nula de la muestra	No es posible la transferencia de información de equipo a equipo
Rapidez de medición	Cierta experiencia en el uso del equipo
Ausencia de reactivos	Uso de modelos matemáticos de gran complejidad
Exactitud comparable con técnicas analíticas clásicas	Costo inicial alto
Aumento de la capacidad de análisis en el laboratorio	Poca movilidad del equipo
Ahorro de espacio	
Escalable a sistemas en línea	

Tabla 3.2 Ventajas y desventajas de la medición por FT- MIR de AGV's.

La espectroscopia FT-MIR asociada con el análisis multivariable ha ido ganando aceptación en los procesos fermentativos como técnica de control rutinario, ya que una vez establecida la calibración el análisis es rápido.

### 3.6 FT-MIR en la medición de ácidos grasos volátiles

En recientes trabajos se ha demostrado el uso de la espectroscopia infrarroja en biorreactores de tipo industrial (*Steyer et al 2002*) para la determinación de carbono orgánico total (COT), demanda química de oxígeno (DQO), ácidos grasos volátiles AGV's y la alcalinidad total y parcial. La adaptación de este tipo de sistemas a la industria de la digestión anaerobia requiere del uso de fibra óptica y unidades de ultrafiltración, utilizadas para liberar de partículas suspendidas que pudiesen interferir con los instrumentos de medición.

Para este tipo de adaptación es imposible utilizar muestras sintéticas para la calibración de los instrumentos debido a que en un proceso real, las interacciones de las especies químicas forman mezclas complejas que conducen a espectros muy particulares, diferentes a los espectros simplistas de las especies químicas individuales. (*Steyer et al 2002*).

Se han reportado, además, resultados en sistemas similares, donde adaptan ligeras modificaciones en el proceso. Esto es, adicionando un filtro de 150  $\mu\text{m}$  para impedir el paso de partículas en la muestra a la unidad de medición. Este trabajo asocia una gran incertidumbre en el modelo de calibración debido a las bajas concentraciones y por tanto a las bajas abosrbencias obtenidas por los espectros infrarrojos. (*Spanjers et. al. 2006*). Por otro lado existen reportes del uso de la espectroscopia infrarroja para el monitoreo de AGV's a escala laboratorio (*Nordber et. al. 2000*) alimentado con sustrato sintético. Los datos obtenidos fueron analizados usando métodos multivariables y los autores concluyeron que la calibración de los modelos para AGV's (acetato y propionato) fue aceptable, sin embargo, no es totalmente correspondiente al método de referencia (cromatografía).

Del mismo modo varios autores han reportado el uso exitoso de espectroscopia infrarrojo cercana (NIR) (*Hansson et. al. 2002*), (*Zhang et. al. 2009*), para el seguimiento de efluentes en biorreactores productores de hidrógeno. Así como adaptaciones de la técnica a monitoreo de lodos activados (*Sarraguca 2009*) y fermentación de productos agrícolas a gran escala (*Holm-Nielsen 2007*) y (*Lomborg 2009*).

Es claro que a pesar de los pormenores y retos que la espectroscopia FT-MIR asociada con el análisis multivariable propone, esta técnica ha ido ganando aceptación en los procesos fermentativos como técnica de control rutinario, ya que una vez establecida la calibración el análisis es inmediato.

### 3.7 Técnicas de calibración multivariable

Entre todas las técnicas quimiométricas existentes, la más utilizada por el sector industrial es la de la calibración multivariable, empleada para el análisis cuantitativo. Esta técnica permite correlacionar dos tipos de medidas aplicadas sobre una misma muestra. De este modo, generalmente se correlaciona una medida que se ha obtenido de manera rápida, directa, económica y que proporciona múltiples variables a la vez, con otra más complicada de obtener, que suele ser más cara y/o más laboriosa, tanto en su preparación como en el análisis que se deriva. Esta última es la que se toma como medida de referencia. Con ambos tipos de datos se pueden construir modelos de calibración que permiten predecir las concentraciones de los analitos de interés a través de la señal analítica registrada mediante la técnica rápida (*Ferré 2006*).

Uno de los objetivos de los métodos quimiométricos es transformar la señal obtenida en el análisis instrumental (sin significación química) en información útil para el analista a través de lo que se conoce como calibración. Es por ello que la calibración, como etapa integrante del proceso analítico, es de gran importancia y sólo proporcionará buena precisión y exactitud en los resultados si se aplica, evidentemente, de forma correcta. En el ámbito de la química analítica se define calibración como el proceso que permite establecer

la relación entre la respuesta instrumental y una propiedad determinada de la muestra, que en determinaciones cuantitativas suele ser la concentración. La analogía matemática que relaciona la señal analítica con la concentración se denomina modelo o ecuación de calibración y la representación gráfica recibe el nombre de curva de calibración. (*Brereton 2000*).

Los modelos multivariantes posibilitan cuantificar un analito en una mezcla sin necesidad de conocer los otros componentes de la misma. Esto se basa en que la información contenida en las variables medidas se puede concentrar en un número menor de variables (llamadas componentes principales) sin pérdida de información relevante. La regresión no se hace sobre los datos originales, sino sobre estas nuevas variables, simplificando el modelo y la interpretación de los resultados. Por otra parte, este tipo de métodos pueden utilizar todas las variables a las que se ha registrado el espectro sin necesidad de selección previa. (*González, 2004*).

Por estas razones, la tendencia actual es la utilización de métodos de calibración basados en una reducción de variables previa al cálculo del modelo, a través de métodos de selección de variables.

### 3.7.1 Etapas del proceso de modelado

El proceso de modelado de un conjunto de datos, para determinación cuantitativa de un parámetro determinado como la concentración de AGV's, tiene como finalidad obtener un modelo capaz de predecir propiedades de nuevas muestras. Para la obtención de un modelo robusto se deben de seguir las siguientes etapas: selección de un conjunto de calibración, determinación de la propiedad o identidad por métodos de referencia adecuados, obtención de la señal analítica, cálculo del modelo, validación del modelo y aplicación del mismo. A continuación se explica de manera general cada una de las etapas implicadas en el modelado de datos.

### 3.7.2 Selección de muestras para la calibración

Para poder asegurar la bondad del modelo de calibración es necesario un proceso de validación, que consiste en aplicar el modelo a muestras de las que se conoce la propiedad estudiada y que no han sido utilizadas en la construcción del mismo. Esta validación puede ser interna o externa. La validación interna se aplica cuando el número de muestras disponibles para la calibración es relativamente pequeño, y la metodología que se sigue es la validación cruzada (*Mardia et. al., 1979*). Ésta utiliza muestras del propio conjunto de calibración.

### 3.7.3 Validación cruzada - dejando una fuera a la vez (leave-one-out)

Un caso especial de validación cruzada, que se utiliza ampliamente, es el algoritmo “dejando uno fuera” (*Brereton, 2000*) (en inglés *leave-one-out*), que consiste en emplear una de las muestras como muestra de validación sacándola del conjunto de calibración, y con el resto de muestras se recalcula el modelo. El nuevo modelo se aplica para predecir la concentración de la muestra sustraída. Este procedimiento se repite hasta que todas las muestras que componen el conjunto de calibración han sido sustraídas y predichas. Los residuales de predicción (diferencias entre los valores reales y los predichos) del total de las muestras se emplean para calcular la raíz cuadrada del error cuadrático medio de validación cruzada (en inglés *Root Mean Square Error of Cross Validation, RMSECV*) del modelo.

### 3.7.4 Creación del modelo de calibración

Para la creación de los modelos de calibración, previamente se ha debido medir la propiedad de interés a través del método de referencia en el conjunto de las muestras y haber tomado su espectro MIR.

Una vez aplicada la regresión PLS al conjunto de calibración para construir el modelo, se evalúa el ajuste del mismo a los datos a través de la raíz cuadrada del error cuadrático medio de calibración (en inglés *Root Mean Square Error of Calibration*, RMSEC), expresada en porcentaje (RMSEC %), y que puede ser definido como el error medio del modelo (Ec.3.10):

$$RMSEC\% = \frac{100}{\bar{y}} \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}, \quad \text{Ec. 3.10}$$

donde  $N$  es el número de muestras del conjunto de entrenamiento,  $y_i$  es el valor de referencia para la muestra  $i$ ,  $\hat{y}_i$  es el valor predicho para la muestra  $i$ , e  $\bar{y}$  es la media de los valores de referencia de las muestras de calibración.

Una etapa clave en la creación de los modelos de calibración es la selección del número de factores, o variables latentes, óptimas que asegure la capacidad predictiva sin que se produzca sobreajuste (*overfitting*).

Esto se debe a que los factores están jerarquizados, de manera que, cada nuevo factor que añadimos al modelo, describe menos variación sistemática en los espectros y más variación aleatoria o ruido. Por tanto, se deben incluir suficientes factores para modelar la variación sistemática importante, pero no excesivos para evitar incluir demasiado ruido.

De este modo, el sobreajuste tiene lugar cuando el modelo emplea un número de factores superior al número óptimo, describiendo más el ruido de las muestras que información relevante, lo que provoca una mala predicción.

El subajuste (*underfitting*), por el contrario, se produce cuando el número de factores empleados por el modelo es inferior al número óptimo, por lo que no son suficientes para modelar las variaciones relevantes de los espectros y, en consecuencia, no se obtienen buenas predicciones (*Mardia et. al., 1979*).

El número óptimo de factores es particular para cada problema de calibración y depende del tipo de espectro y de la propiedad a predecir. Se determina calculando modelos con distinto número de factores y validando cada modelo. El criterio más simple para la selección del número óptimo de factores de los modelos PLS, que es el que se ha seguido en el presente trabajo, consiste en determinar la mínima raíz cuadrada del error cuadrático medio de validación cruzada RMSECV (*Brereton, 2000*), representando éste frente al número de factores del modelo y tomar como óptimo el primer mínimo local o absoluto.

### 3.7.5 Validación del modelo de calibración

La raíz cuadrada del error cuadrático medio de validación cruzada (RMSECV), que se emplea para determinar el número óptimo de factores y para validar internamente el modelo por validación cruzada, como se ha comentado anteriormente, se define en porcentaje (RMSECV %) en la Ec.3.11:

$$RMSECV\% = \frac{100}{\bar{y}} \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_{-i})^2}{N_{CV}}}, \text{ Ec. 3.11}$$

donde  $N_{CV}$  es el número de muestras de validación cruzada, es decir,  $N_{CV} = N$ , siendo  $N$  el número de muestras del conjunto de calibración,  $y_i$  es la medición de referencia,  $\hat{y}_{-i}$  es el resultado estimado cuando se construye el modelo sin la muestra  $i$  e  $\bar{y}$  es la media de los valores de referencia de las muestras de calibración.

La concordancia entre los valores de referencia y los predichos por el modelo en las muestras del conjunto de validación, se realiza a través de la RMSEP, determinada en porcentaje (RMSEP %), cuando se dispone de un conjunto independiente de muestras, no utilizado en la calibración (Ec. 3.12):

$$RMSEP\% = \frac{100}{\bar{y}} \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N_{CV}}}, \quad \text{Ec. 3.12}$$

donde  $N$  es el número de muestras del conjunto de prueba,  $y_i$  es el valor de referencia para la muestra  $i$ ,  $\hat{y}_i$  es el valor predicho para la muestra  $i$  e  $\bar{y}$  es la media de los valores de referencia de las muestras de calibración. Es el método de validación más objetivo, ya que las muestras de validación no influyen en la calibración del modelo (Brereton 2000).

Tanto la RMSEP como la RMSECV proporcionan una estimación de la exactitud del modelo, facilitando la magnitud del error en las futuras muestras

### 3.8 Regresión por mínimos cuadrados parciales (PLS)

El método de mínimos cuadrados parciales fue desarrollado por Wold en 1975 (Heinen & Kaplan 2004). Este método permite utilizar regiones amplias del espectro, ya que sus algoritmos son los que se encargan de escoger un pequeño conjunto de números de ondas no correlacionados. A estas nuevas variables se las denomina factores o variables latentes, y se calculan combinando linealmente todas las variables originales (Mardia *et. al.*, 1979).

$$t = p_1 x_1 + p_2 x_2 + \dots + p_m x_m \quad \text{Ec. 3.6}$$

donde  $[p_1, p_2, \dots, p_m]$  son los *loadings* y  $[x_1, x_2, \dots, x_m]$  son los datos espectroscópicos. El vector de *scores* ( $t$ ) se define como el producto de los *loadings* ( $p$ ) y las variables originales ( $x$ ). Cuanto más parecidos sean dos espectros, y por tanto, más parecidos sean  $[p_1, p_2, \dots, p_m]$ , más similares serán sus *scores* (Mardia, *et. al.*, 1979). La virtud de los métodos de calibración multivariable PLS reside, básicamente, en cómo se calculan los *loadings*  $[p_1, p_2, \dots, p_m]$  para combinar las variables originales. Los *loadings* que calcula el

método PLS, proporcionan *scores* que están lo más correlacionado posible con las concentraciones.

En PLS se intenta contener la mayor cantidad de información para la predicción de las muestras en los primeros factores o variables latentes. Para ello el algoritmo PLS utiliza tanto la información contenida en la matriz de datos espectroscópicos (matriz **X**) como la información contenida en la matriz de datos a determinar (matriz **Y**), obteniéndose unas variables auxiliares llamadas variables latentes, factores o componentes PLS.

Las matrices **X** e **Y**, se descomponen en una suma de factores, calculando simultáneamente:

$$X = TP^t + E = \sum_{\alpha=1}^A t_{\alpha} p_{\alpha}^t + E \quad \text{Ec. 3.7}$$

$$Y = UQ^t + F = \sum_{\alpha=1}^A u_{\alpha} q_{\alpha}^t + F \quad \text{Ec. 3.8}$$

La ecuación (3.7) es la descomposición de la matriz de datos espectroscópicos **X**, siendo **T** la matriz de *scores*, **P** la de *loadings* y **E** la de residuales. El superíndice *t* representa la transpuesta de la matriz o vector. Análogamente, en la ecuación (3.8), que se refiere a la descomposición de la matriz de la propiedad a determinar (concentración) **Y**, **U** es la matriz de *scores*, **Q** la de *loadings* y **F** la matriz de residuales. Si tenemos *M* muestras, *A* factores, *K* variables y *P* analitos, la dimensionalidad de las matrices es la siguiente: **T** y **U** (*M* x *A*), **P**<sup>t</sup> (*A* x *K*) y **Q**<sup>t</sup> (*A* x *P*).

La descomposición de ambas matrices no es independiente, sino que se realiza de forma simultánea, estableciéndose una relación interna entre los scores de los bloques **X** e **Y** de forma que para cada componente *a* se cumpla:

$$\hat{u}_{\alpha} = b_{\alpha} t_{\alpha} \quad , \quad \text{Ec. 3.9}$$

donde el símbolo <sup>^</sup> indica que es una cantidad calculada y *b<sub>a</sub>* es el coeficiente de regresión para cada uno de los factores PLS. Si se desea calcular una única concentración de las presentes en la matriz **Y**, se utiliza el algoritmo llamado PLS-1, descrito en la Tabla 3.3.

Paso 1	Se calculan los pesos (W) utilizando la variabilidad restante en y expresada como $y_{a-1}$ $X_{a-1} = y_{a-1} W^T_a + E$ Escalando a continuación $W^T_a = c X^T_{a-1} y_{a-1}$ C es un factor que hace que W sea 1
Paso 2	Se estiman los scores $t_a$ utilizando los pesos obtenidos $X_{a-1} = t_a W^T_a + E$ Siendo los scores calculados $t_a = W_a X_{a-1}$
Paso 3	Con los scores calculados, se calculan los loadings $p_a$ $X_{a-1} = p_a W^T_a + E$ Siendo por mínimos cuadrados $= X^T_a t_a / t_a t_a^T$
Paso 4	Lo siguiente es calcular el loading de concentración $q_a$ $q_a = X^T_a t_a / t_a t_a^T$
Paso 5	Se crea una matriz $X_{a-1}$ nueva y una $Y_{a-1}$ sus trayendo este factor $X_{a-1}^{(nueva)} = X_{a-1} - t_a t_a^T$ $Y_{a-1}^{(nueva)} = y_{a-1} - t_a q_a$
Paso 6	Se determinan los A factores PLS y se calculan los predictores b y $b_0$ para ser utilizado en nuevas muestras $b = W(P^T W)^{-1} q$ $b_0 = \bar{y} - \bar{x}^T b$
Predicción	$y = b_0 - x^T b$

Tabla 3.3 Algoritmo PLS-1.

## Capítulo 4. Materiales y métodos

### 4.1 Estrategia experimental

Para el presente trabajo de investigación se propone una metodología encaminada a obtener un sistema de medición rápida de ácidos grasos volátiles en el proceso de producción de hidrógeno mediante fermentación en un biorreactor, utilizando espectroscopia infrarroja para la obtención de concentraciones de ácidos grasos volátiles en espectros que serán almacenados en una base de datos y calibrados bajo la técnica de mínimos cuadrados parciales. A continuación se esquematiza en bloques la metodología propuesta.

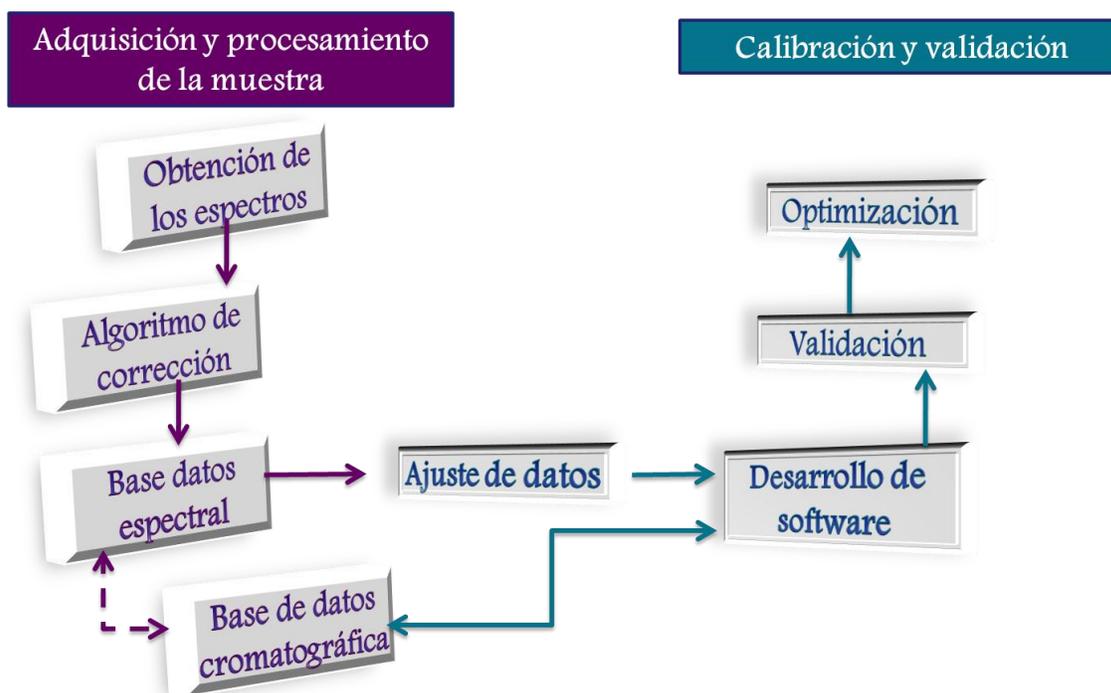


Figura 4.1 Estrategia experimental.

El proceso metodológico se seccionó en dos etapas. La primera se enfoca en encontrar las bases de datos propios a FT-MIR y de correspondencia cromatográfica. El desarrollo de un algoritmo de corrección en la toma de espectros de AGV's es necesario para la estandarización y respetabilidad de muestras.

La segunda etapa es orientada hacia el desarrollo del software y la validación de los métodos utilizando diferentes técnicas.

## 4.2 Toma y tratamiento de la muestra

La base de datos de calibración se formó a partir del efluente producido por cuatro reactores fermentativos, todos productores de hidrógeno. La Tabla 4.1 se muestran las características más importantes de operación de los reactores involucrados en la toma de muestra.

Conjunto de muestras	Tipo	Alimentación	Concentración de alimentación (g/L)	Muestreo	pH	TRH	Agitación RPM	Temperatura °C
JE (15)	Continuo	Glucosa	20 g/L	12	5.5	12	150	35
JP (22)	Continuo	Glucosa - vinaza	20 g/L-4 DQO/L	12	5.3	12	150	35
RA (22)	Continuo	Glucosa	10 g/L	12	5.5	12	Velocidad Ascendente	35
MC (17)	SBR	FORSU	5 g SV/L	12	5.5	24	180	37

Tabla 4.1 Características operacionales de los reactores.

El tratamiento dado al efluente de los reactores se describe en la Figura 4.2 Donde es necesario retirar sólidos de las muestras a través de centrifugación y filtrado. Para el análisis cromatográfico, 1.5 mL de muestra tratada se acidifican con HCl 1:1 y se conserva a 4 °C, para su posterior análisis. Por otro lado, una cantidad aproximada a 5 µL, sin ningún otro tratamiento, es necesaria para su análisis por FT-MIR.

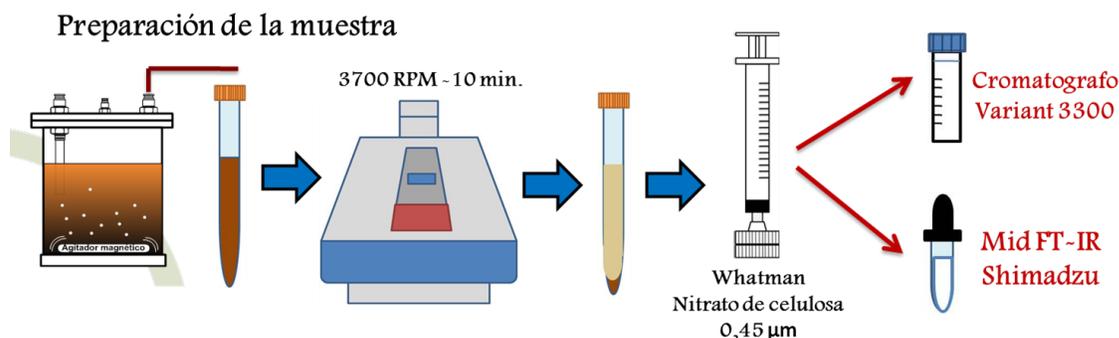


Figura 4.2 Tratamiento de la muestra.

#### 4.2.1 Medición de AGV's por cromatografía

En la determinación cuantitativa de los AGV's, se utilizó un cromatógrafo de gases (Varian 3300) equipado con un detector de ionización de flama FID y una columna capilar de sílice Zebram ZB-FFPA de 0,53 mm de diámetro, 15 m de largo y 1 µm de espesor de película. Las temperaturas del inyector y detector fueron 190 °C y 210 °C respectivamente; la temperatura inicial de horno fue de 45 °C, la cual se mantiene por 1.5 min incrementado hasta 135 °C a una tasa de 8 °C/min. Como gas portador se utilizó nitrógeno con una presión de 40 psi. Inyectando 2 µL de muestra al cromatógrafo de gases. De acuerdo con el tiempo de retención fueron identificados los AGV's de la muestra. Las áreas arrojadas por el sistema se comparan con la curva de calibración a partir de las cuales se calcula la concentración de AGV's.

#### 4.2.2 Medición de AGV's por Espectroscopia FT-MIR

El registro de los datos espectrales en la región media se llevó a cabo con el equipo FT-IR Affinty 1 (Shimadzu Instruments). La calibración del espectrómetro se verificó periódicamente con un patrón de poliestireno. Se utilizó el aditamento PIKE MIRacle™ ATR con un cristal de ZnSe (R. index 2,4-1000 w1). El análisis de referencia (*background*) se realizó con el cristal en distensión y seco. El registro espectral se realizó en el intervalo de 1800 a 900  $\text{cm}^{-1}$ , con resolución de 2  $\text{cm}^{-1}$  y un promedio de 32 barridos. En el registro y procesamiento de los espectros, se utilizó el programa IRSolution 2.1.

Una gota de muestra de efluente tratado se compacta contra el cristal, haciendo pasar varias veces el haz infrarrojo a través del cristal, penetrando la muestra en cada refracción. El procedimiento se repite con una muestra de agua.

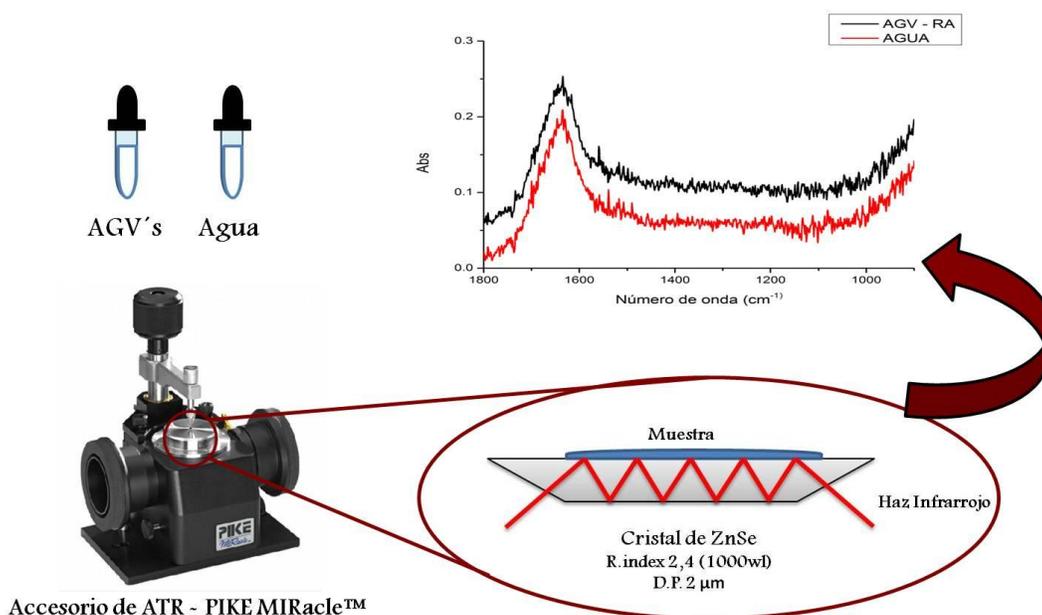


Figura 4.3 Procesamiento de espectro

### 4.2.2.1 Algoritmo de corrección

Para garantizar la confiabilidad y repetitividad en los datos espectrales, se adecuo un algoritmo de corrección semi-automático. Utilizando el software IRSolution 2.1, una serie de pasos descritos en la Figura 4.3 se aplican a todos las muestras analizadas. El beneficio del uso de este algoritmo es la adquisición de espectros con la mínima cantidad de ruido y la desaparición de la banda en  $1600\text{ cm}^{-1}$  que enmascara las vibraciones correspondientes a los AGV's.

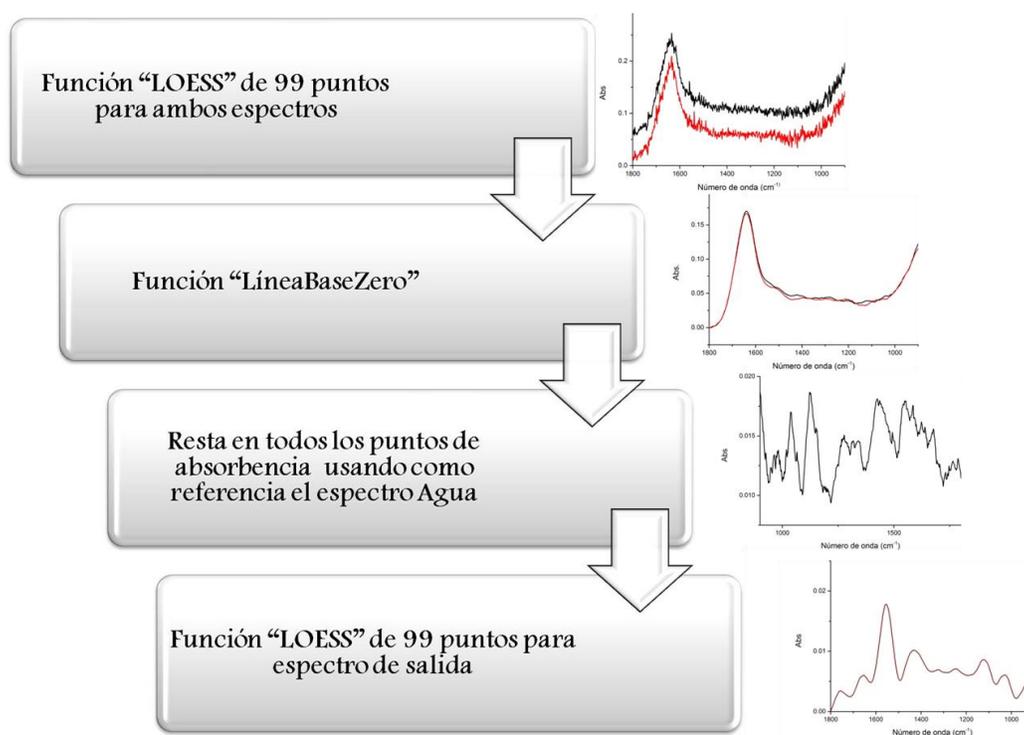


Figura 4.4 Algoritmo de corrección de espectros

### 4.3 Calibración

Una vez procesados los espectros de acuerdo al algoritmo de corrección, se crea una base de datos de calibración, utilizando la técnica de validación cruzada se descartan influencias espectrales y asegurar la independencia de cada muestra. El algoritmo empleado elimina de la calibración una muestra cada vez (leave one out) y calcula un modelo con las muestras restantes; a partir de este modelo se predice el valor de la propiedad físico-química o concentración de la muestra omitida; este proceso se repite hasta que cada una de las muestras haya sido excluida del sistema. La detección de las muestras fuera de intervalo se realizó mediante análisis de residuales y análisis estadístico de influencia espectral. El número óptimo de factores se seleccionó de la gráfica de RMSEC frente a número de variables latentes; el criterio utilizado fue el punto en el cual, RMSEC es un mínimo y el coeficiente de correlación  $R^2$  es máximo. Los cálculos de las variables latentes del modelo PLS fueron realizados en MatLab (*Statics Toolbox<sup>TM</sup>*) mediante el comando *plsregress*, el cual utiliza el algoritmo SIMPLS propuesto por (Adams, 2004).

### 4.4 Validación

Escogiendo muestras al azar (de concentración conocida) e introduciéndolas al modelo de calibración se evalúa la capacidad predictiva del sistema, descrita previamente por la validación cruzada. Los cálculos de las variables latentes del modelo PLS fueron realizados en MatLab (*Statics Toolbox<sup>TM</sup>*) utilizando el comando *crossval*.

## Capítulo 5. Resultados y discusión

En este capítulo se muestran y explican los resultados obtenidos a través de las diferentes pruebas y análisis a los datos espectrales y su correlación con sus respectivas variables dependientes. En primer lugar se incluye una valorización de los tiempos y consumibles utilizados por ambas técnicas. Posteriormente, se explica la adquisición de la base de datos de absorbancias y el análisis espectral para cada ácido. Posteriormente a través de la regresión PLS se calibra un modelo para la matriz de datos inicial, es decir los datos obtenidos directamente del espectrofotómetro y los medidos por cromatografía. Se realiza una selección de variables latentes o componentes principales más significativos en función de dos factores decisivos: el porcentaje de varianza explicada y el error cuadrático de la media. Por último el modelo PLS obtenido se somete a un estudio puntual de cotejamiento e inspección de variables.

### 5.1 Rapidez y consumibles

Varios puntos comparativos se tomaron en cuenta al evaluar en lo particular las dos técnicas utilizadas. En este apartado se valora la rapidez del método y su bajo uso de consumibles. La Tabla 5.1 muestra los gases implicados en la rutina del cromatógrafo y su consumo mientras la medición se realiza. Además del uso de columnas, el cromatógrafo requiere altas tasas de consumo de aire, hidrógeno y nitrógeno. Por otro lado, el FT-IR no requiere de consumibles en la evolución de la adquisición de la muestra. Otra ventaja del uso del espectroscopio sobre el cromatógrafo es el funcionamiento a temperatura ambiente. El cromatógrafo requiere de altas temperaturas en diferentes procesos y componentes, lo cual además consume tiempo en el calentamiento y enfriamiento, retardando la técnica, antes y después de la medición. En lo que al tiempo de medición se refiere la técnica de FT-MIR se mostró cinco veces más rápido en la adquisición de datos, consumiendo apenas 3 minutos efectivos, contra los 15 minutos de la técnica analítica cromatográfica.

	<b>Cromatografía</b>	<b>FT-MIR</b>
Consumibles en el proceso de medición	N <sub>2</sub> 2.5 ml/min H <sub>2</sub> 32.5 ml/min Aire 324 ml/min Columna	—————
Temperatura (° C)	Inyector 190 Detector 210 Columna 70 – 120	Ambiente ≈ 25
Limpieza	Acetona/Hexano	Alcohol isopropílico/Papel óptico
Tiempo de adquisición	≈ 15 minutos	≈ 3 minutos

Tabla 5.1 Comparación de las técnicas de cromatografía y FT-MIR.

Los beneficios de la cuantificación de ácidos grasos volátiles por FT-MIR son evidentes, lo que la convierte en una técnica candidata a ser empleada mediante un muestreo automatizado para hacer mediciones cuasi en-línea, ya que además de las características mostradas en la Tabla 5.1, la preparación de la muestra es relativamente sencilla, siendo muchas veces necesario tan solo la centrifugación de la misma.

## 5.2 Adquisición y análisis espectral

Cuando se emplean los métodos multivariantes como la regresión por mínimos cuadrados parciales (PLS), la selección de la gama espectral adecuada es un tema importante. El rango espectral por lo tanto debe incluir información que describe las principales características del analito, mientras que las regiones correspondientes a la matriz (si el caso) o por artefactos o ruidos deben ser excluidos (*Hussy et al. 2005*).

En un primer acercamiento al equipo de espectroscopia, es necesario la identificación de las regiones espectrales que representan las vibraciones correspondientes a los ácidos involucrados en el proceso fermentativo. Así, como referencia inicial, fueron medidos tres de los distintos ácidos producidos por un reactor digestor productor de hidrógeno en diferentes períodos de operación. Estos ácidos fueron elegidos por su concentración sobresaliente en el proceso y corresponden a ácido acético (HAc), ácido butírico (HBu) y ácido propiónico (HPr). Identificando la estructura de cada ácido y comparando con las bandas de absorción que se obtuvieron en su correspondiente espectro infrarrojo se identificaron posibles zonas de disparidad entre ellos. Estas zonas son de gran utilidad en la identificación de cada ácido para el proceso de adquisición de un modelo de calibración como el propuesto en este trabajo de investigación. Como anteriormente se explicó en el Capítulo 3, la región de compuestos específicos o *fingerprint* ubicada en el rango de  $1300-900\text{ cm}^{-1}$ , recibe este nombre porque es la región del MIR donde los compuestos orgánicos generan bandas de absorción características, dependiendo de la estructura de la molécula y permitiendo la identificación de compuestos específicos. La región de compuestos específicos o *fingerprint* es específica y única para cada molécula permitiendo una filiación entre las especies de ácidos grasos volátiles. Esta disparidad visible en la Figura 5.1, a la postre permitirá que la calibración por el método de mínimos cuadrados parciales sea más sencilla.

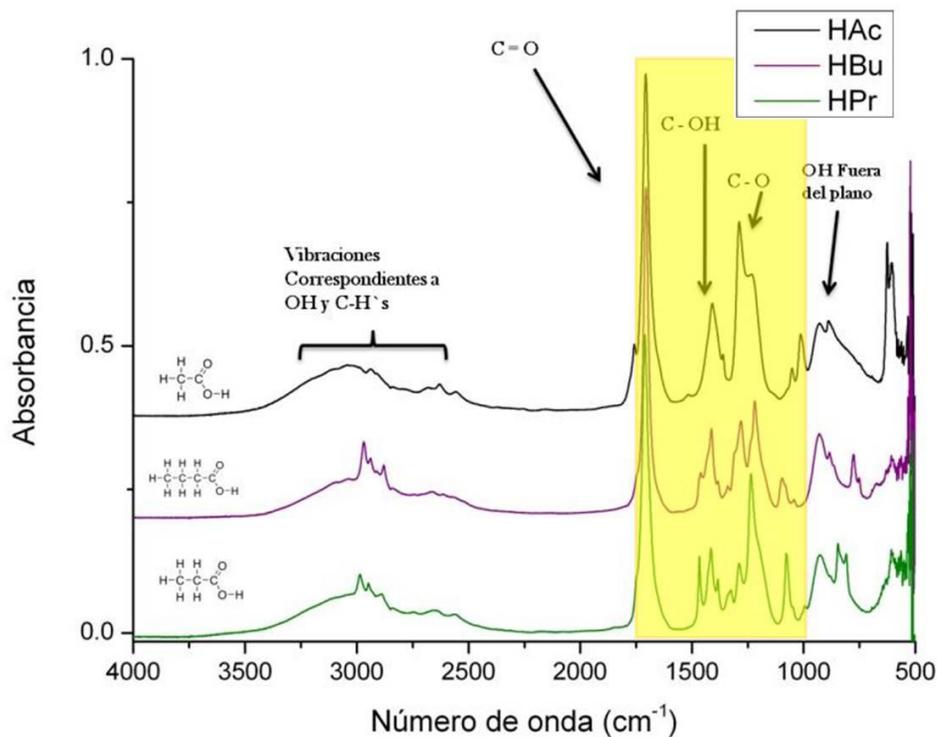


Figura 5.1 El espectro de los ácidos grasos medidos (las líneas base de cada muestra están desplazadas para mayor claridad en la presentación).

La Figura 5.1 presenta los espectros de cada ácido medido en todo el intervalo de números de onda ( $4000 - 400 \text{ cm}^{-1}$ ). Es importante mencionar que estas mediciones se hicieron en los ácidos puros, es decir, sin estar disueltos en agua, que es como se espera obtenerlos del biorreactor. En esta Figura encontramos idoneidad de identificación para los ácidos grasos volátiles en el intervalo marcado en la Figura 5.1, correspondiente a  $1800 \text{ cm}^{-1} - 900 \text{ cm}^{-1}$  debido a la disparidad de bandas correspondientes a las tensiones del grupo  $\text{C}=\text{O}$  que comparten todos los espectros de HAc, HBU y HPr. Además, esta zona presenta la conveniencia de evadir la gran banda de absorción correspondiente a la vibración del grupo OH ( $3500 \text{ cm}^{-1}$ ) perteneciente al agua, que como ya se mencionó está presente en las mediciones de las muestra obtenidas por los biorreactores, debido a que los ácidos

producidos en el proceso fermentativo para la producción de hidrógeno están disueltos a concentraciones muy bajas.

La Figura 5.2 muestra los espectros sin ningún tipo de filtro espectroscópico de un par de muestras de ácidos grasos volátiles obtenidos del biorreactor.

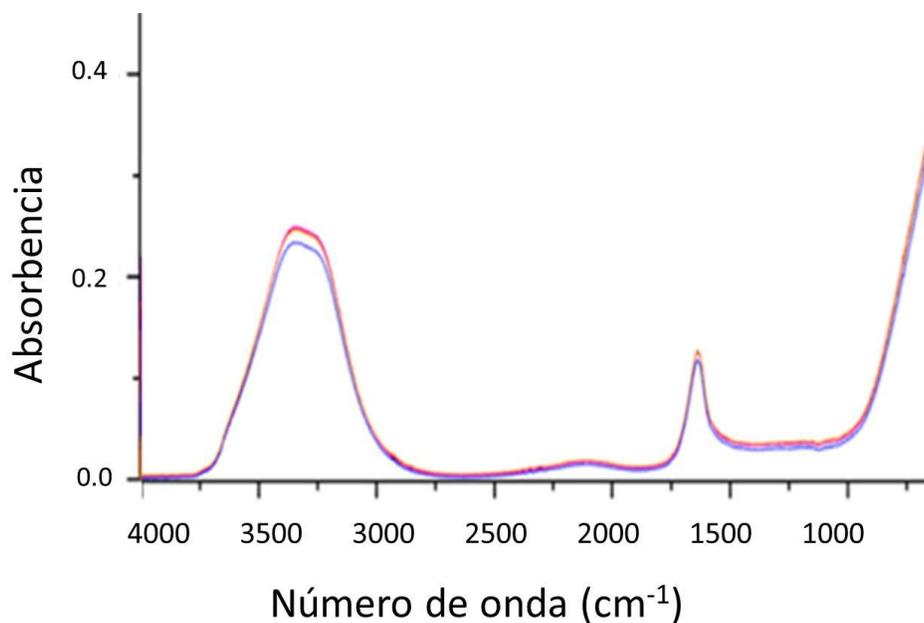


Figura 5.2 Espectros de muestras de ácidos grasos volátiles sin ningún filtro espectroscópico.

En esta Figura sólo se puede observar dos grandes bandas pertenecientes al soluto de la muestra, en este caso agua. La presencia del agua en las muestras obtenidas de los biorreactores es tan grande y las absorbencias de su molécula en la zona MIR tan fuerte que enmascara a todos los picos de ácidos grasos volátiles en los espectros de las muestras. A pesar de haber escogido la zona correspondiente al intervalo de 1800-900  $\text{cm}^{-1}$ , se puede observar, sin embargo, que la vibración a 1500  $\text{cm}^{-1}$  en el espectro del agua enmascara también las vibraciones correspondientes a los ácidos a medida que su concentración en la solución acuosa disminuye. El rango espectral por lo tanto no incluye la información que describe las principales características del analito. Ante esta complicación, el algoritmo de corrección descrito en la Sección 4.5 fue evaluado.

Bajo esta metodología de adquisición de espectros, se midieron soluciones acuosas de ácido acético (HAc) y ácido butírico (Hbu) a 500 mg/L preparadas en el laboratorio, considerando que esta concentración está en el rango de las concentraciones obtenidas por los biorreactores. Los espectros pertenecientes a estas soluciones se presentan en la Figura 5.3, donde se comparan con los espectros de HAc, HBU a 10 g/L reportados por Uden & Sjaunja (2009).

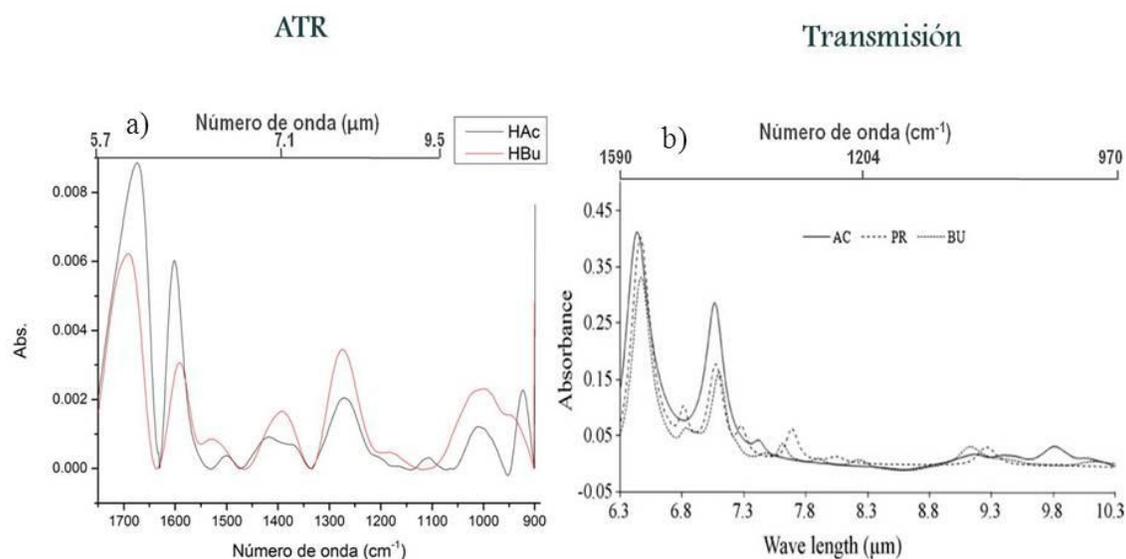


Figura 5.3 Comparación de espectros FT-MIR. a) ATR 5g/L, b) Transmisión 10g/L (Uden & Sjaunja 2009).

En la Figura 5.3 se observa una disimilitud en la nomenclatura para referirse a cada uno de los ácidos, cabe aclarar que esto se debe a que en el presente trabajo se ha propuesto desde el inicio el uso de la letra H antepuesta a las dos primeras letras de cada ácido, mientras en el artículo publicado por Uden & Sjaunja se ha preferido usar solo las dos primeras letras del ácido. Para lograr uniformidad en el presente escrito, se ha decidido no cambiar la nomenclatura para la comparación de ambos graficos en la figura.

La Figura 5.3a muestra los espectros FT-MIR obtenidos en el laboratorio por la técnica de ATR del ácido acético (HAc) y del ácido butírico (HBU) a 500mg/L y la Figura

5.3b muestra los espectros de ácido acético (AC), ácido butírico (BU) y ácido propiónico (PR) obtenidos por Uden & Sjaunja por FT-MIR pero usando una celda de transmisión. Aunque las técnicas de muestreo son diferentes, la presencia de los picos típicos correspondientes a los grupos funcionales de los ácidos carboxílicos para ambos espectros es coherente. Los picos de las mediciones registradas con la técnica de ATR se muestran muy por debajo de la magnitud de absorbencia que exhibe la celda por transmisión. Sin embargo, esto es de esperarse, pues al hacer las mediciones en diferentes equipos y bajo diferentes técnicas, las unidades de absorbencia no coincidirán, ya que estas unidades dadas de manera arbitraria por cada equipo y solo son coherentes entre sí mismas, si son obtenidas en un mismo equipo. También se observa que las contribuciones de la Figura 5.3a se muestran un poco desfasados en sus números de onda con respecto a los espectros obtenidos por transmisión Figura 5.3b, la explicación de esta fenómeno es que al utilizar un método de muestro rápido como ATR la absorción de la muestra es menor debido a la forma de interactuar con el haz infrarrojo, sacrificando un poco la magnitud y sensibilidad del espectro (*Setnicka 2000*).

Así, en ambos gráficos se distinguen las mayores contribuciones a los 6.43-6.46 o 7.06-7.08  $\mu\text{m}$ , donde los picos están cerca uno de otro, contrario a lo observado en las regiones correspondientes a números de onda más altos, 7.2-7.7  $\mu\text{m}$ , donde los picos muestran una mayor separación y son favorables para utilizar en el modelo. Para la comparación de las gráficas vale la pena recordar que el eje de las abscisas en espectros FTIR puede medirse en número de onda (wn) con unidades de  $\text{cm}^{-1}$ , o bien en longitud de onda ( $\lambda$ ) con unidades de  $\mu\text{m}$  o  $\text{nm}$ ; la relación es inversa, tal que:  $\text{wn} (\text{cm}^{-1}) = 10000 / \lambda (\mu\text{m})$ .

### 5.2.1 Representatividad del espectro

Los modelos matemáticos de calibración y predicción de espectros se basan en la representatividad que un espectro exhibe en función de su referencia. Para que modelos de

ajuste multivariables como PLS deben su éxito a la coherencia y concordancia de las variables, en este caso la concentración y la absorbancia espectral. Bajo este premisa, la Figura 5.4 muestra los espectros de muestras sintéticas que van desde los 20 a 2,5 g/L de una mezcla de ácido acético y butírico. Esta mezcla se preparó y midió por triplicado con la finalidad de evaluar la coherencia de las concentraciones conocidas de los ácidos, comparadas con su correspondiente espectral. La Figura muestra un pico representativo y perteneciente al doble enlace carbono oxígeno propio del grupo carboxilo ( $1700-1650\text{ cm}^{-1}$ ). Esta contribución es ideal para ser examinada ya que pertenece a un grupo funcional que ambos ácidos comparten. Se puede apreciar una ligera variación en las mediciones de repetición en los espectros, visualmente se podría asegurar que esta contribución efectivamente sigue la ley de Beer-Lambert y su absorbancia varía de forma consistente a la concentración. Sin embargo, la apreciación de este fenómeno es evidente en la Figura 5.5.

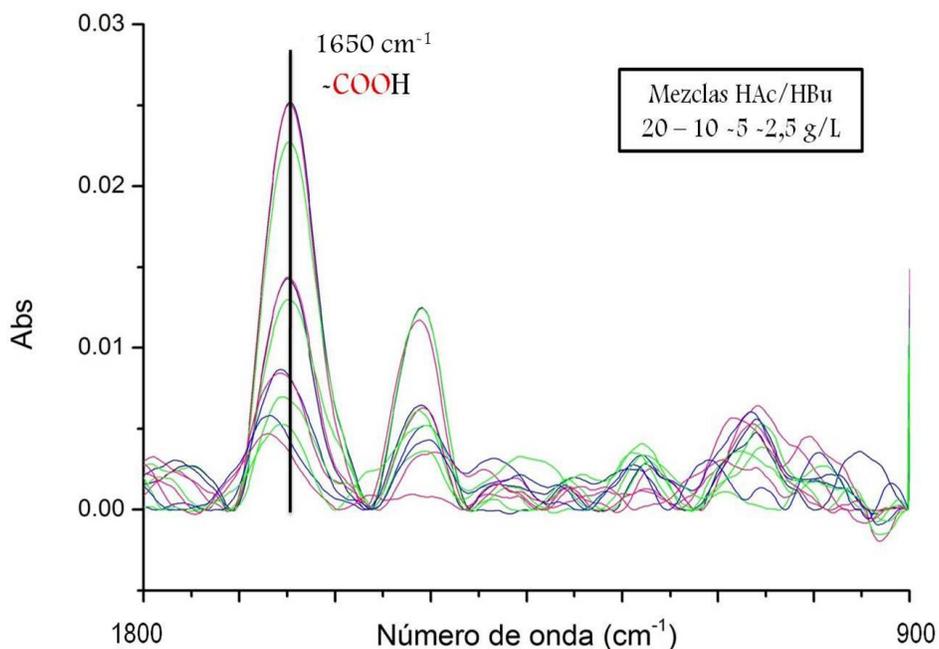


Figura 5.4 Espectros sometidos al algoritmo de corrección.

En la figura 5.5 se observan los valores de concentración teórica comparados con sus correspondientes valores de absorbancia de cada una de las repeticiones de las mezclas de ácidos en el pico de  $1650\text{ cm}^{-1}$ . La Figura muestra la relación de los datos de absorbancia de los espectros contra las concentraciones, para cada una de las soluciones. El modelo calculado a partir de estos datos muestra un coeficiente de determinación igual con 0.987 entre los valores predichos y los valores calculados de forma teórica en la preparación de las soluciones de ácidos acético y butírico.

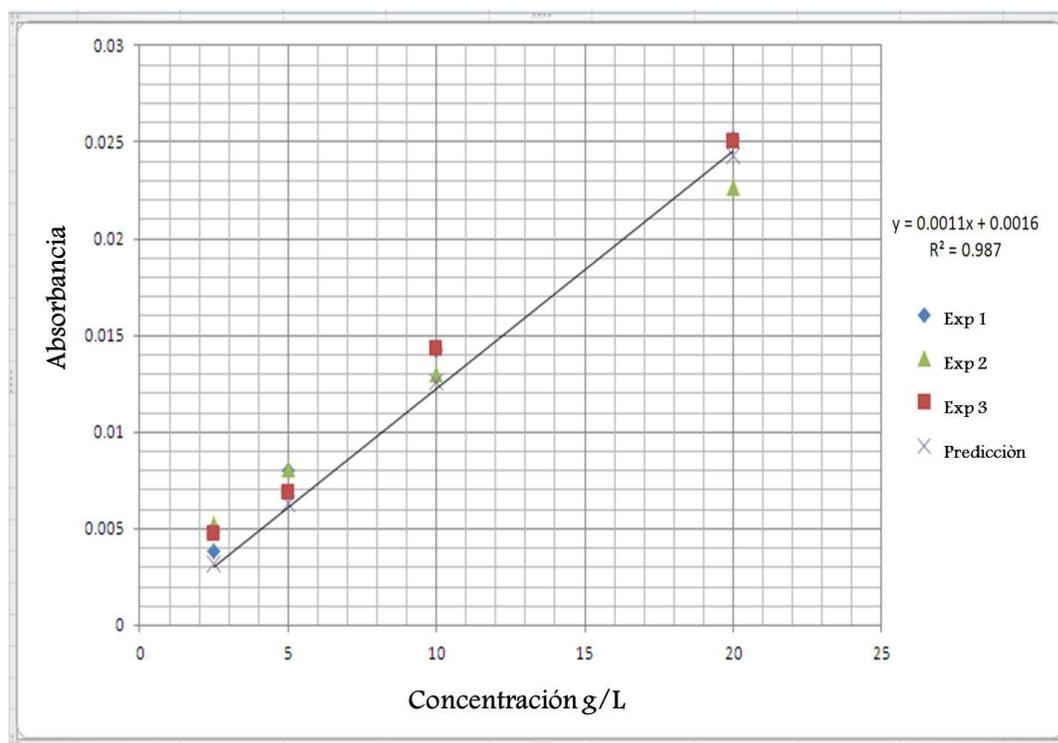


Figura 5.5 Evaluación de representatividad de los espectros en  $1650\text{ cm}^{-1}$ .

El objetivo de la calibración es determinar las concentraciones a partir de las intensidades espectrales, y no a la inversa. Sin embargo, la ecuación de calibración en esta sección implica ajustar un modelo para determinar una altura de pico a partir de una concentración conocida. Para una muestra nueva o desconocida, la concentración puede ser estimada (aproximadamente) mediante el uso de la inversa de la pendiente. Este tipo de

gráfico es siempre de gran utilidad para tener una representación visual de que la respuesta es razonable.

### 5.3 Robustez analítica

En los siguientes apartados se detallan las etapas seguidas para llevar a cabo la construcción y validación de los modelos de calibración PLS que permitieran la cuantificación de los ácidos grasos volátiles usualmente analizados en los controles del proceso de producción de hidrógeno en biorreactores fermentativos, mediante una metodología rápida de análisis.

#### 5.3.1 Valores de referencia

Para asegurar la representatividad de este estudio, se requiere un número elevado de muestras y, sobre todo, que éstas representen la máxima variabilidad de ácidos grasos volátiles producida por los biorreactores. Esto se consiguió tras un muestreo bien planificado y minuciosamente llevado a cabo de manera que, tras aplicar el método de obtención establecido a cada una de los efluentes de biorreactores, se obtuvieron 51 muestras para cada ácido. Como antes se había mencionado, los únicos ácidos que se estudiaron por su relevancia en el proceso fueron el ácido acético (HAc) y el ácido butírico (HBu). La Tabla 5.2 resume la concentración (g/L) de cada ácido en las muestras obtenidas de los reactores mediante el uso de las medidas analíticas de referencia, es decir la cromatografía, teniendo en cuenta cada grupo de alimentación para los biorreactores por separado y considerando todas las muestras analizadas semanalmente durante el período de septiembre a noviembre. En ella se puede ver, por ejemplo, que el biorreactor alimentado con residuos sólidos urbanos se caracteriza por su baja producción de AGV's, o que dentro de los reactores alimentados con glucosa (*Glucosa 1* y *Glucosa 2*) existen diferencias en los valores máximos de la producción de HAc, lo que podría representar un valor fuera de

rango en el primer reactor *Glucosa 1*. De esta manera se advirtió un comportamiento poco común en el reactor *Glucosa 1*, lo cual se manifiesta por una desviación estándar de los ácidos mucho mayor que en las otras muestras. Esta concentración representa una anomalía en el reactor *Glucosa 1*, ya que indicaría que en 24 horas se produjo un 30% más ácidos, y en las siguientes 24 horas esta concentración cayó un 50%, manteniéndose sin variación significativa hasta el fin del experimento. Sin aunar en las causas que provocaron esta singularidad en el conjunto de datos, pues rebasa los objetivos del presente trabajo, estos datos se tomarán como valores fuera de rango y serán retirados del conjunto de concentraciones.

Reactor	Muestras	AGV	Min	Max	Media	SD
Glucosa1	20	HAc	245,149	1235	616,075	334,328
		HBu	245,117	717,844	426,302	144,881
Glucosa2	10	HAc	146,203	162,632	154,737	6,745
		HBu	336,239	461,482	386,301	64,629
FORSU	11	HAc	40,296	70,283	61,272	13,527
		HBu	31,448	40,095	40,095	7,211
Vinaza	10	HAc	137,327	282,768	190,188	53,796
		HBu	177,694	330,828	263,345	59,205

Tabla 5.2 Parámetros estadísticos descriptivos para los AGV's producidos por los biorreactores con diferentes alimentaciones.

La Tabla 5.3 muestra el análisis de parámetros estadísticos del conjunto de datos de calibración del modelo PLS, una vez retiradas las muestras que por su alta concentración de HAc representan valores fuera de rango de concentración en los grupos de efluentes de biorreactores con distintas alimentaciones.

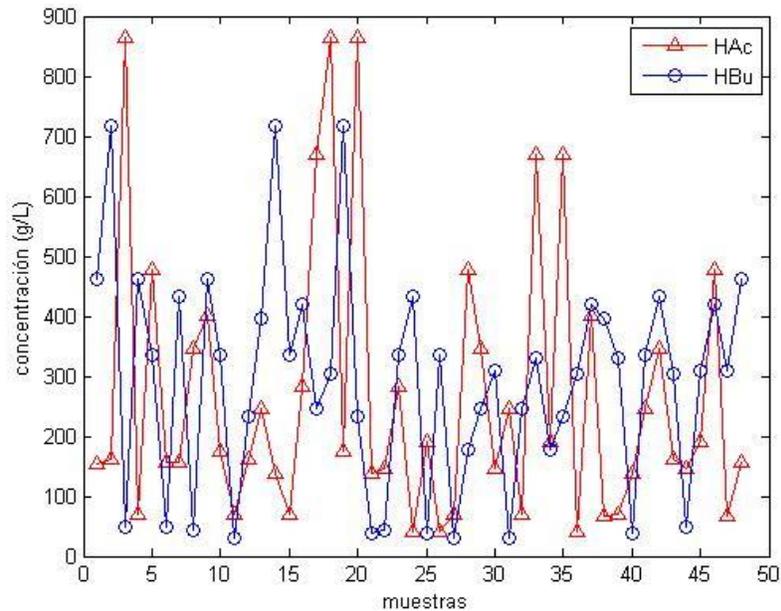


Figura 5.6 Distribución de las concentraciones de AGV's.

Datos	Muestras	AGV	Min	Max	Media	SD
		HAc	40,29	865,63	265,41	227,311
Calibración	48	HBU	31,448	717,844	293,84	181,3923

Tabla 5.3 Parámetros estadísticos descriptivos para los AGV's en el conjunto de datos de calibración.

La Figura 5.5 muestra un panorama general de la distribución de AGV's en los reactores. Las muestras están adecuadas a un proceso aleatorio de acuerdo a la toma de espectros y la Tabla 5.3 resume los parámetros estadísticos de las concentraciones de las muestras. En esta se puede observar que este conjunto de calibración presenta una amplia diferencia entre el valor mínimo y máximo para ambos ácidos; esto puede ser benéfico ya que cubre un gran porcentaje de la varianza. La desviación estándar es muy similar para ambos grupos de AGV's. La mayor parte de esta variabilidad se debe a las bajas concentraciones de AGV's presentadas por el biorreactor alimentado con residuos sólidos urbanos (FORSU).

### 5.3.2 Análisis cuantitativo: modelo PLS

Tras la metodología de selección de longitudes de onda y representatividad de valores de referencia y espectrales que se ha desarrollado para simplificar el proceso de calibración, donde se han eliminado algunas muestras no representativas del conjunto de datos, y elegido la parte del espectro correspondiente al recorrido de 882 a 1817  $\text{cm}^{-1}$  como intervalo de trabajo para el uso de las variables que construyen los espectros del ácido acético (HAc) y el ácido butírico (HBu). En esta sección se explica la adquisición del modelo PLS. Para lo cual ello es necesario recordar que este tipo de calibración descompone las variables utilizadas en sus componentes principales o variables latentes para encontrar factores subyacentes con significado físico que se pueden relacionar con un comportamiento lineal. Cabe señalar que en el presente trabajo se han utilizado un total de 44880 variables (935 números de onda por 48 muestras). Generalmente en este tipo de modelos, se dividen las muestras en dos grupos de datos: uno para la calibración y otro para la validación. Sin embargo, al utilizar un número pequeño de variables como se hace en este experimento, es recomendable el uso de la validación cruzada dejando una muestra fuera (*leave -one-out*) que como se explicó en la Sección 3.7.3 realiza tantas calibraciones como variables a evaluar, dejando una de ellas fuera del conjunto de calibración.

Una etapa clave en la creación de los modelos de calibración es la selección del número de factores, o variables latentes, óptimas que aseguren la capacidad predictiva sin que se produzca sobreajuste. Esto se debe a que los factores están jerarquizados, de manera que, cada nuevo factor que se añade al modelo, describe menos variación sistemática en los espectros y más variación aleatoria o ruido.

Como se indicó en la Sección 3.7.5, el número óptimo de variables latentes requeridas para explicar la variabilidad del modelo PLS fue realizado a partir del cálculo del error cuadrático de predicción *MSEP* usando un número definido de variables latentes. Gráficamente, la Figura 5.7 ilustra el porcentaje de varianza explicada para las concentraciones de HAc y HBu con respecto al número de variables latentes. Se aprecia la disminución del *MSEP*, conforme el número de variables latentes aumenta.

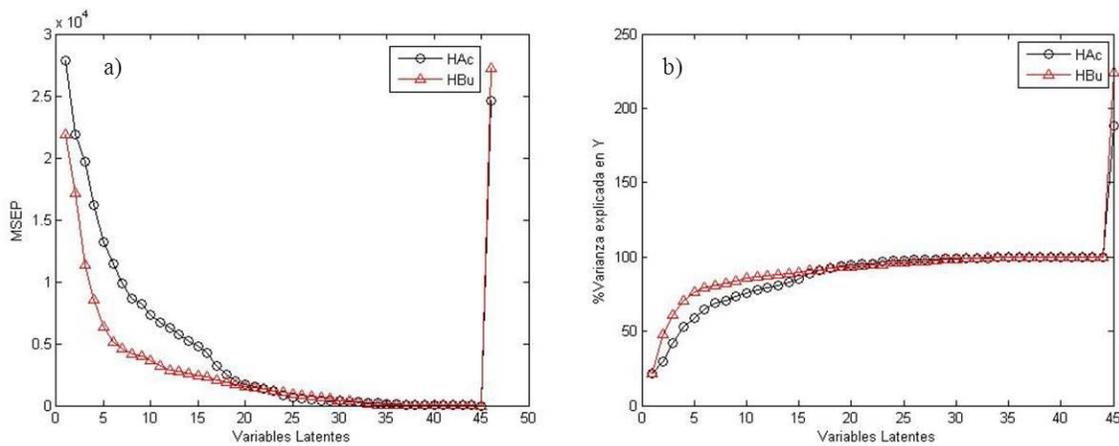


Figura 5.7 a) Error medio de predicción por el número de variables latentes b) Porcentaje de varianza explicada en Y (concentración) de acuerdo al número de variables latentes.

Los parámetros de la Figura 5.7 son decisivos para la creación del modelo de regresión por mínimos cuadrados parciales (PLS) ya que, como se observa, entre más variables latentes o componentes principales se utilicen, el MSEP disminuirá y el ajuste del coeficiente de determinación aumentará. Sin embargo, si se excede el número de variables latentes utilizado se puede llegar a un sobreajuste. Este tiene lugar cuando el modelo emplea un número de factores superior al número óptimo, describiendo más el ruido de las muestras que información relevante, lo que provoca una mala predicción. La Tabla 5.4 describe con más precisión los parámetros fundamentales para evaluar el éxito en la calibración PLS. Los parámetros de calibración analizados fueron: el número óptimo de variables latentes, el coeficiente de determinación de la calibración ( $R^2$ ), la raíz cuadrada del error cuadrático medio de calibración (RMSEC %), la raíz cuadrada del error cuadrático medio de validación cruzada (RMSEVc %).

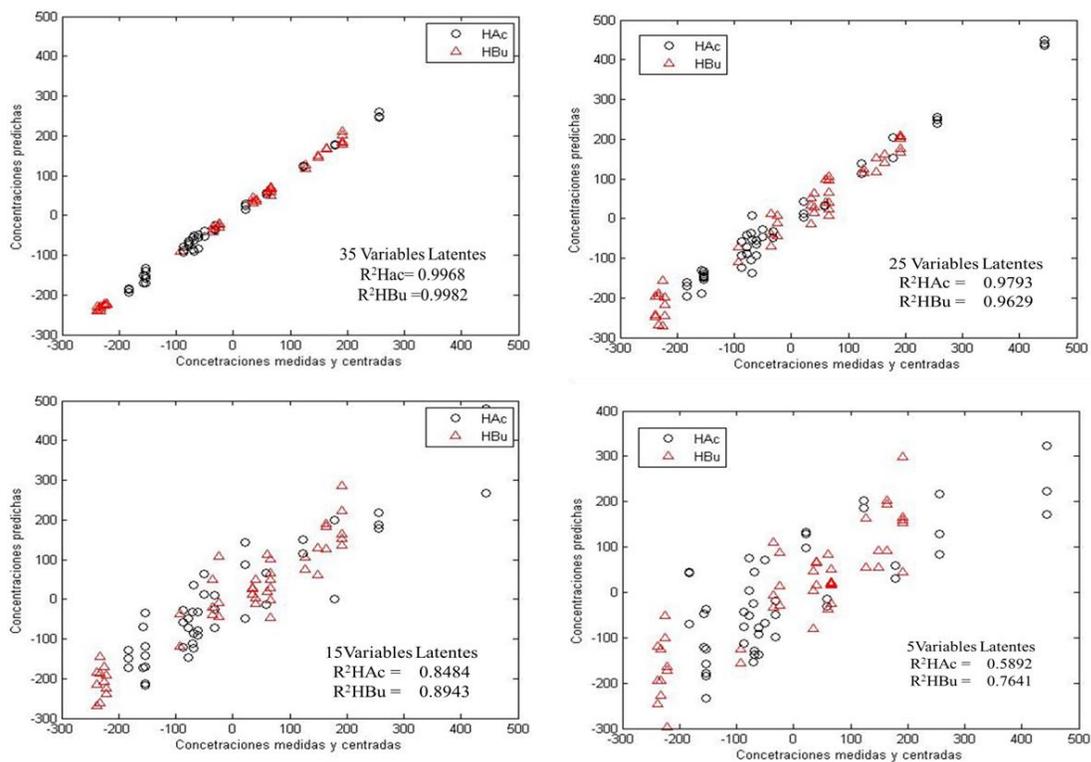


Figura 5.8 Gráficas de las concentraciones medidas y centradas contra las concentraciones predichas por cada modelo de regresión.

La Figura 5.8 es una representación gráfica de la repercusión del uso de variables latentes sobre los diferentes parámetros de evaluación en el proceso de calibración. El centrado facilita el análisis de los datos, así, aunque se tomen valores negativos, se puede observar que un aumento considerable en el uso de variables latentes conducirá a un mejor ajuste del modelo de calibración PLS, obteniendo coeficientes de correlación  $R^2$  máximos de 0.997 para el HAC y de 0.998 para el HBU con el uso de 35 variables latentes. Sin embargo, con el uso de métodos estadísticos como PLS se espera utilizar tantas variables latentes como componentes analizados. La Tabla 5.4 se muestra un análisis detallado de todos los parámetros involucrados en el proceso de calibración PLS.

	Muestras	Variables latentes	R <sup>2</sup>	RMSEC (mg/L)	RMSEV <sub>c</sub> (mg/L)
HAc			0.997	9,38	175,50
HBu	46	35	0.998	6,25	272,59
HAc			0.9793	23,99	2583
HBu	46	25	0.9629	28,51	927,47
HAc			0.8484	64,96	2150
HBu	46	15	0.8943	48,14	1619
HAc			0.5892	106,94	5315
HBu	46	5	0.7641	71,93	2450

Tabla 5.4 Resumen de los resultados de acuerdo al número de variables latentes utilizado.

Los valores presentados en la Tabla 5.4 muestran que a pesar de utilizar un gran número de variables latentes y tener determinantes de correlación elevados; los errores de validación para nuevas concentraciones bajo la metodología de validación cruzada dejando una muestra fuera (*leave-one-out*), son extremadamente grandes para todos los grupos de calibración. Esta observación es indicio de la baja o nula representatividad de las variables latentes entre contribuciones espectrales. Además, es de suma importancia resaltar que en la calibración por descomposición de factores o variables latentes, la mayor parte de la información se recoge en los primeros componentes. En este caso un menor número de componentes debería explicar el mayor porcentaje de la varianza; de lo contrario, y como es evidente en el caso de este experimento, un mayor número de variables latentes no describen la variación sistemática; más bien la correlación esta mayormente descrita por la variación aleatoria, es decir: ruido u otros componentes que no representan contribuciones significativas en el conjunto de espectros. Para probar la identidad de las contribuciones que describe cada una de las variables latentes es necesario graficar los *loadings* en el rango espectral, como se observa en la Figura 5.9

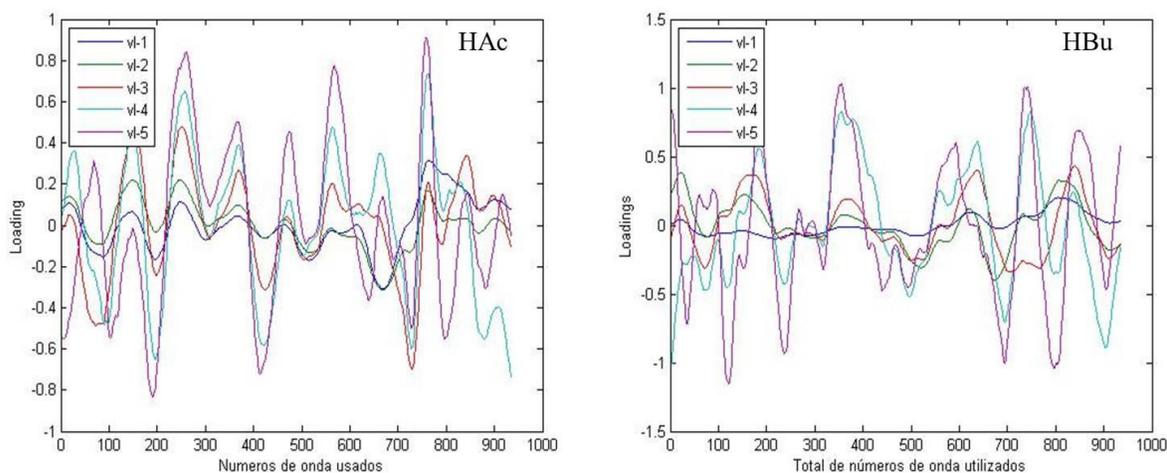


Figura 5.9 Gráfica de *loadings* de las cinco primeras variables latentes correspondientes a la región completa de los espectros para HAc y HBu.

La Figura 5.9 describe gráficamente la identidad y la explicación de la varianza entre las contribuciones espectrales. Los *loadings* del modelo PLS son las combinaciones lineales de las variables originales (espectros) que definen los componentes o variables latentes a utilizar en la calibración, es decir, describen la intensidad con la que cada variable latente en la regresión PLS depende de las variables originales, y en qué dirección. Dicho de otra forma, el gráfico de *loadings* muestra la influencia de cada número de onda en la varianza espectral. Así los *loadings* se presentan como picos de acuerdo a la relevancia que tenga cada variable latente o componente en cierto rango del espectro y la magnitud de los picos es directamente proporcional a las contribuciones de cada componente para explicar la variación en las variables de predicción. Estos picos tomarán valores positivos y negativos. Se espera que el mayor aporte de información proporcionada por los *loadings* de cada variable latente corresponda al rango espectral donde se encuentran las contribuciones primordiales de los ácidos medidos. Esto es claramente analizable al observar la Figura 5.10.

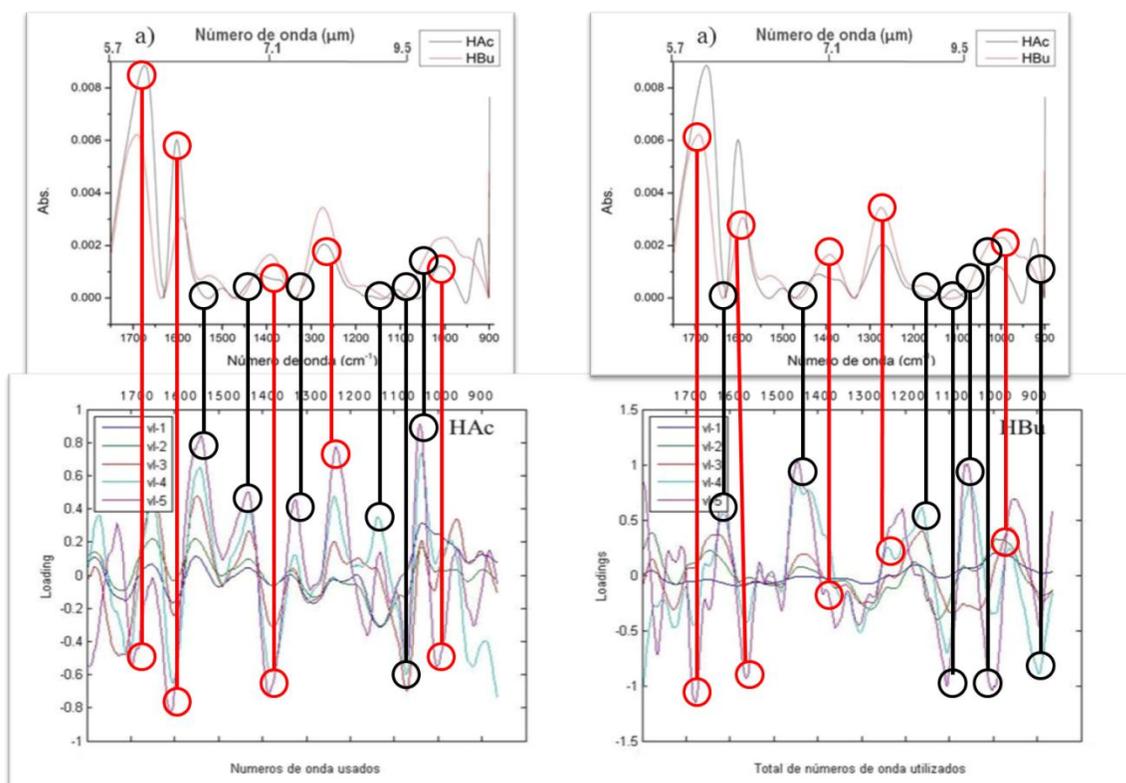


Figura 5.10 Superposición del gráfico de loadings de los primeros cinco componentes y los espectros de HAc y HBU respectivamente.

Lo anterior es fácilmente analizable al observar la Figura 5.10 que presenta una superposición de los espectros de HAc y Hbu con sus correspondientes valores de *loadings*. Los círculos rojos representan las principales contribuciones de los espectros para cada ácido y su correspondiente relevancia en el modelo de predicción. Se observa que la mayoría de las contribuciones espectrales son representadas como valles en los gráficos de *loadings*. Esto significa que existe un solapamiento de bandas entre los espectros que conforman la base de datos de calibración y difícilmente contribuirán al modelo. De igual forma, la intensidad de los picos en los gráficos de *loadings* representa la relevancia que esta contribución tiene para el modelo. Por otro lado, los círculos de color negro indican los picos correspondientes a los *loadings* de mayor relevancia, y como se puede observar al proyectarlos hacia los espectros de HAc y HBU, estos no se identifican con las principales

contribuciones que aportan los ácidos. Esto es evidencia de que, como ya se había mencionado, las contribuciones hechas por las variables latentes correlacionan en gran medida variación aleatoria o componentes que no representan las contribuciones significativas en los espectros. De acuerdo a las irregularidades encontradas hasta ahora, la Figura 5.11 representa el grupo de espectros de calibración utilizado en el desarrollo del modelo PLS.

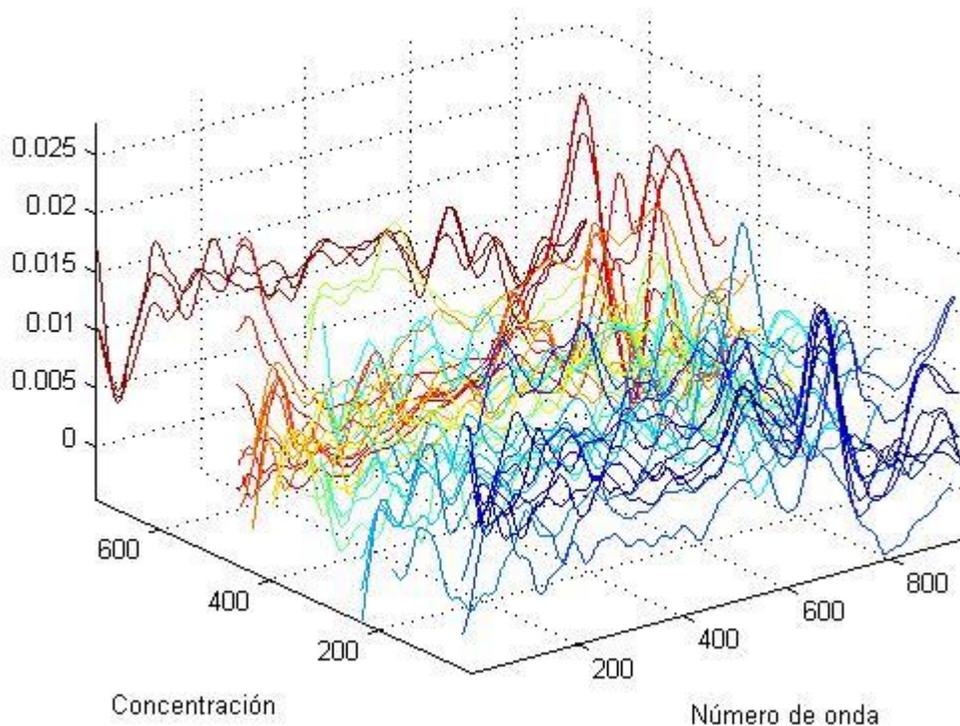


Figura 5.11 Conjunto espectral utilizado en la calibración del modelo PLS.

Podemos observar en la Figura 5.11 que las contribuciones espectrales no siguen la ley de Beer-Lambert, ya que a bajas concentraciones se distinguen picos de gran absorbancia y a altas concentraciones picos de absorbancias menores. Además en esta Figura se observa la disparidad a lo largo de las regiones espectrales. Esta inconsistencia espectral probablemente es debida a la variedad de componentes suspendidos en las

muestras obtenidas de cada reactor que interfieren con la medición de los componentes de interés (*Ward et. al. 2010*). De acuerdo al tipo de efluente que se obtenga de cada proceso fermentativo, pequeñas contribuciones interfieren con la veracidad del espectro y por supuesto con su similitud con los otros espectros obtenidos de otros efluentes. La Figura 5.11 muestra los grupos de espectros para cada reactor.

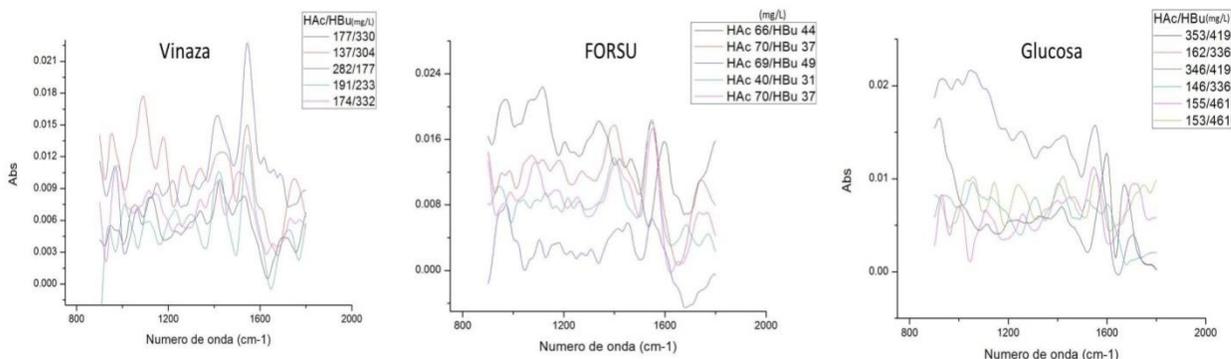


Figura 5.12 Comparación de espectros por tipo de proceso en el biorreactor.

Ante esta complejidad resulta conveniente hacer un modelo para cada reactor y de esta manera disminuir la discrepancia espectral. Además es recomendable normalizar todos los espectros hacia un pico en común que no aporte ninguna información, ya que como se puede observar en la Figura 5.12 aunque existen bandas que coinciden de manera cualitativa, es imposible establecer una correlación cuantitativa con la concentración de AGV's en el efluente si estas contribuciones no son coherentes con la ley de Beer-Lambert. La normalización hacia un mismo pico permite la homogenización de los espectros y facilita el seguimiento puntual aunando la arbitrariedad del equipo al asignar valores de absorbancia (*Arcos-Hernández et. al., 2010*). En el caso del presente trabajo es imposible utilizar la técnica de normalizado, debido a que no existe ninguna banda en común que no aporte información al modelo. Sin embargo, el uso de celdas para líquidos en vez de ATR durante la lectura de los espectros en el equipo FTIR podría garantizar el seguimiento de espectros a lo largo de todo el espectro medio del infrarrojo, facilitando así la localización de este tipo de bandas. Además, la forma de obtener información por medio de estos

dispositivos es más robusta que por medio del ATR (*Vanrolleghem & Lee, 2003*). Este tipo de técnicas y procesamiento de datos abre una nueva oportunidad de investigación para el laboratorio, dando continuidad al presente proyecto y buscando la mejora en la cuantificación de AGV's de forma rápida y confiable. Sería posible explorar en trabajos futuros diversas metodologías para la adquisición y procesamiento de datos, así como la posibilidad de individualizar algoritmos para cada uno de los conjuntos de calibración.

## Capítulo 6. Conclusiones

- El algoritmo de corrección de espectros en el FT-MIR mostró un coeficiente de correlación  $R^2$  0.987 para las muestras de AGV's sintéticos, sin embargo, para las muestras obtenidas de los reactores fermentativos no se observó ningún tipo de correlación en la contribución espectrográfica de los AGV's.
- Las mediciones de muestras de AGV's por FT-MIR se realizaron 5 veces más rápido que en el cromatógrafo.
- El modelo desarrollado a partir de PLS permitió comparar los coeficientes de regresión y los errores de calibración para los AGV's utilizando un valor bajo de variables latentes.
- Se puede afirmar que el modelo obtenido por PLS es una técnica versátil y prometedora para la cuantificación rápida de AGV's, sin embargo, este método sería provechoso de ser adecuado a las condiciones de los reactores del laboratorio.

## Capítulo 7. Referencias bibliográficas

Adams, M. (2004) *Chemometrics in analytical spectroscopy*. The Royal Society, Cambridge, UK.

Chen X., Sun Y., Xiu Z., Li X., Zhang D (2006). *Stoichiometric Analysis Of Biological hydrogen production by fermentative bacteria*. International Journal of Hydrogen Energy

Debabrata D., Nejat-Veziroglu T., (2001) *Hydrogen production by biological processes: a survey of literature*. International Journal of Hydrogen Energy.

Dago-Morales A., Fernández R., Ruiz M., Balmayor M., Laza M., Ross J.E., Otrero M., Armada S., (2006) *La espectroscopia infrarroja y el método de calibración multivariada de mínimos cuadrados parciales en la predicción del índice de octano experimental de gasolinas*. Revista CENIC. Ciencias Químicas, vol. 37, núm. 1.

Haenlein M., Kaplan M-A., (2004) *A beginner`s guide to partial least squares analysis*, Understanding Statistics, 3(4), 288-297.

Han K., Shin S., (2004). *Biohydrogen production by anaerobic fermentation of food waste*. International Journal of Hydrogen Energy.

Hussy F.R., Hawkes R., Dinsdale D.L., (2005) *Continuous fermentative hydrogen production from sucrose and sugarbeet*; International Journal of Hydrogen Energy 30 471 – 483.

Gonzalez M. F., (2004) *Desarrollo de un método de selección de variables para datos espectroscópicos en el infrarrojo cercano*. Escola Técnica Superior Enginveria. Universitat Rivora.

Madigan, M., Martinko, J., & Parker, J. (2003) *Brock Biology of Microorganisms* New Jersey: Prentice Hall Inc.

Madsen M., Jens B., Kim H., (2011) *Monitoring of anaerobic digestion processes: A review perspective*. Renewable and Sustainable Energy Reviews

Mazarevica G., Josef D., Baena J., Erwin R., Bernhard L., (2004) *On-Line Fermentation Monitoring by Mid-infrared Spectroscopy*. Applied Spectroscopy.

Setnicka V., (2000) *FT-IR Reflection Techniques; Overview*. Main Principles of Reflection Techniques.ppt

Skoog D.A., Holler F. J., (2000). *Principios de Análisis. Instrumental*, 5° ed.; Ed. McGraw-Hill.

Smith B. 2000 *Infrared Spectral Interpretation A Systematic Approach*, CRS Press.

Spanjers H, Bouvier JC, Steenweg P, Bisschops I, van Gils W, Versprille B. (2006) *Implementation of an in-line infrared monitor in full-scale anaerobic digestion process*. Water Science and Technology 53, 55–61.

Steyer J.P., Bouvier J.C, Conte T, Gras P., Harmand J. and Delgenes J.P. (2002). *On-line measurements of COD, TOC, VFA, total and partial alkalinity in anaerobic digestion processes using infra-red spectrometry*. Water Science and Technology.

Udén P., Sjaunja P.O., (2009) *Estimating volatile fatty acid concentrations in rumen samples by Fourier transform mid-IR transmission spectroscopy*. Animal Feed Science and Technology 152, 123–132.

Vanrolleghem P.A. and Lee D.S. (2003). *On-line monitoring equipment for wastewater treatment processes: state of the art*. Water Science and Technology.

Ward AJ, Hobbs PJ, Holliman PJ, Jones DL. (2008) *Optimisation of the anaerobic digestion of agricultural resources*. *Bioresource of Technology*, 99, 7928–40.