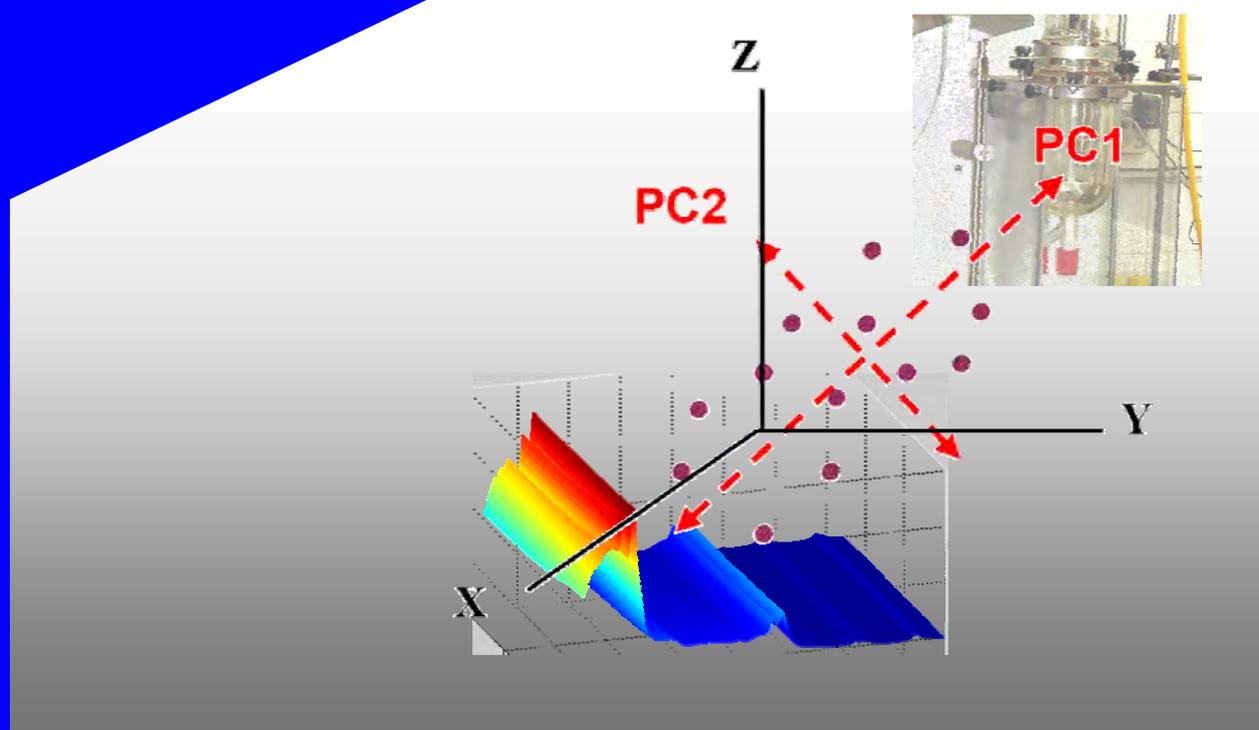


Universitat Autònoma de Barcelona
Departament de Química

Aplicación de la Espectroscopia NIR al Control Analítico de Procesos de la Industria Química

Tesis Doctoral



Autor: Miguel Castillo Martínez
Director: Marcelo Blanco
Bellaterra, 2007



Universitat Autònoma de Barcelona

Aplicación de la Espectroscopia NIR al Control
Analítico de Procesos de la Industria Química

Tesis Doctoral

Miguel Castillo Martínez

Bellaterra, 2007



**Universitat Autònoma
de Barcelona**

Departament de Química
Unitat de Química Analítica
Grup de Quimometria Aplicada

Aplicación de la Espectroscopia NIR al Control Analítico de Procesos de la Industria Química

Memoria presentada por Miguel Castillo Martínez

para conseguir el grado de Doctor en Química

Bellaterra, 2007



Universitat Autònoma de Barcelona

Dr. Marcelo Blanco Romia, catedrático de Química Analítica de la Facultad de Ciencias de la Universidad Autónoma de Barcelona,

CERTIFICA: Que el presente trabajo de investigación titulado “**Aplicación de la Espectroscopia NIR al Control Analítico de Procesos de la Industria Química**”, que constituye la Memoria presentada por Miguel Castillo Martínez, ha sido realizado bajo mi dirección y reúne las condiciones exigidas para aspirar al grado de Doctor en Química.

Y para que quede constancia, firmo el presente certificado en Bellaterra a 15 de mayo de 2007.

Dr. Marcelo Blanco Romia

AGRADECIMIENTOS

Siempre hay mucho que agradecer a la gente que nos rodea. Y más cuando se finaliza una tesis doctoral. Durante cuatro años de mucho trabajo, hasta poder escribir esta página, muchas personas han colaborado aportando ideas, ayuda, comentarios, amistad...

Quisiera agradecer al Profesor Marcelo Blanco el tiempo, la ayuda y la colaboración invertidos en esta tesis desde su inicio. Me gustaría hacer extensible este agradecimiento a los Profesores Hortensia Iturriaga, Santiago Maspoch y Jordi Coello.

A Marta Porcel por todo lo que aprendí junto a ella y por los buenos momentos que compartimos durante el primer año y medio de tesis, a Rufí Mulero por su colaboración y buen humor durante los siguientes años, a Manel Orti por su ayuda y simpatía, y a Rafael Beneyto por estos cuatro años de agradable colaboración con Uniqema (ahora Croda).

Como no, a todos los compañeros del *Grup de Quimiometria Aplicada* de la UAB, con los que he compartido momentos impagables durante 4 años, en especial a Mari, Anna, Quim, Jose, Antonio, Maneles A y B, Jordi, Miriam, Carlos, Mar, Idoia y muchos más, que aunque no los nombre están presentes en este momento.

A muchos amigos que, aunque no colaboren directamente, siempre se han interesado por el estado de salud de esta tesis (cuantas veces habré escuchado el 'y cuando la presentas...'). En especial a Marta, Elena, Adela, Esther, Josep, Salut, Vicente y Raquel.

A mis padres y a mi hermano, quienes siempre me han apoyado.

A todos, Muchas Gracias,

Miguel

ÍNDICE

Resumen	VII
Abstract	XI
Abreviaturas y notación	1
OBJETIVOS	5
INTRODUCCIÓN	
1 Control de procesos en la industria química	
1.1 Introducción general.	13
1.2 Procesos en discontinuo o <i>batch processes</i>	14
1.3 Monitorización de procesos en discontinuo	16
1.3.1 Consistencia del proceso y del producto	16
1.3.2 Alteraciones durante el proceso	17
1.3.3 Optimización del proceso	17
1.3.4 Seguridad y medio ambiente	17
1.4 Metodologías para el control de procesos	18
Referencias	20
2 Espectroscopia NIR	
2.1 Introducción	23
2.2 Fundamentos	26
2.2.1 Origen de la absorción en el infrarrojo	26
2.2.2 Origen de la absorción en el infrarrojo cercano	30
2.2.3 Instrumentación	33
2.3 Aplicación del NIR en el control de reacciones	36
2.3.1 Evolución del NIR en el control de reacciones	36
2.3.2 Ventajas del NIR en la industria	40

ÍNDICE

2.4 Aplicaciones del NIR en la industria química	41
Referencias	44
3 Quimiometría	
3.1 Introducción	51
3.2 Etapas del desarrollo de modelos de calibración multivariados	52
3.2.1 Selección del conjunto de muestras	52
3.2.2 Determinación del parámetro a modelar mediante el método de Referencia	52
3.2.3 Obtención de la señal analítica	53
3.2.4 Pretratamiento de los datos	53
3.2.5 Desarrollo y evaluación del modelo de calibración	53
3.3 Pretratamientos espectrales	54
3.3.1 Pretratamientos independientes del grupo de muestras	54
3.3.1.1 Suavizado espectral	54
3.3.1.2 <i>Standard Normal Variate</i> (SNV)	55
3.3.1.3 Derivadas	55
3.3.2 Pretratamientos dependientes del grupo de muestras	55
3.3.2.1 Promedio de espectros	55
3.3.2.2 Centrado	56
3.3.2.3 Autoescalado	56
3.3.2.4 <i>Multiplicative Scatter Correction</i> (MSC)	56
3.3.2.5 <i>Orthogonal Signal Correction</i> (OSC)	57
3.3.2.6 Señal Neta del Analito (NAS, Net Analyte Signal)	58
3.4 Técnicas de análisis multivariable	60
3.4.1 Análisis en Componentes Principales (PCA)	60
3.4.2 Regresión por Mínimos Cuadrados Parciales (PLS)	63
3.4.2.1 Introducción	63
3.4.2.2 Fundamentos teóricos	64

ÍNDICE

3.4.2.3 Construcción de un modelo PLS y evaluación de resultados	65
3.4.3 Resolución Multivariante de Curvas (MCR)	68
3.4.3.1 Introducción	68
3.4.3.2 Fundamentos	69
3.4.3.3 Algoritmo ALS (Alternating Least Squares)	71
3.4.3.4 Restricciones	72
3.4.3.5 Evaluación de resultados	74
3.4.3.6 Seguimiento de procesos a tiempo real	75
3.5 Límites de detección en calibración PLS	75
Referencias	79

METODOLOGÍA Y DISCUSIÓN DE RESULTADOS

4 Metodología experimental

4.1 Introducción	89
4.2 Procesos químicos estudiados y obtención de muestras	89
4.2.1 Escalado	89
4.2.2 Esterificación entre glicerina y mezcla de ácidos caprílico y cáprico	90
4.2.3 Esterificación entre isopropanol y ácido mirístico	92
4.2.4 Destilación de 2-etilhexanol de una matriz del éster del ácido de coco	95
4.2.5 Obtención de muestras dopadas en el laboratorio	95
4.2.5.1 Muestras para la esterificación entre glicerina y ácidos caprílico y cáprico	95
4.2.5.1 Muestras dopadas para la desodorización de 2-etilhexanol	96
4.3 Instrumentación NIR	96
4.3.1 Seguimiento <i>at-line</i> de procesos químicos	96
4.3.2 Seguimiento <i>in-line</i> de procesos químicos	98
4.4 Métodos de referencia	99

ÍNDICE

4.4.1	Determinación de índice de ácido e índice de hidroxilo	99
4.4.1.1	Índice de ácido	99
4.4.1.2	Índice de hidroxilo	100
4.4.2	Determinación de glicéridos, 2-etilhexanol e isopropanol por cromatografía de gases	102
4.4.2.1	Análisis de glicéridos y glicerina	102
4.4.2.2	Análisis de 2-etilhexanol	104
4.4.2.3	Análisis de isopropanol	104
4.4.3	Determinación de agua	104
4.5	Programas para el tratamiento quimiométrico de datos	105
	Referencias	109
5	Resultados y discusión	
5.1	Seguimiento at-line de un proceso de esterificación: reacción entre glicerina y ácidos caprílico y cáprico	113
5.1.1	Determinación de índice de ácido, índice de hidroxilo y agua mediante calibración PLS	116
5.1.2	Determinación de glicerina, ácido, mono, di, y triglicéridos y agua mediante PLS	119
5.1.3	Aplicación de la Resolución Multivariante de Curvas a la reacción de esterificación entre glicerina y ácidos caprílico y cáprico	124
5.1.3.1	Determinación de glicerina, ácido, mono, di, y triglicéridos	126
5.1.3.2	Determinación de índice de ácido e índice de hidroxilo	131
5.2	Seguimiento in-line de la esterificación entre isopropanol y ácido mirístico	138
5.2.1	Aplicación de MCR-ALS	140
5.2.2	Estimación de parámetros cinéticos	146

ÍNDICE

5.3 Determinación de bajas concentraciones de un analito mediante espectroscopia NIR	151
5.3.1 Determinación de 2-etilhexanol	153
5.3.2 Estimación de límites de detección para la determinación de 2-etilhexanol	159
Referencias	162
6 Conclusiones	165
7 Artículos	173

I) Analytical control of an esterification process by NIR spectroscopy.

II) Analytical control of an esterification batch reaction between glycerine and fatty acids by near-infrared spectroscopy.

III) Use of multivariate curve resolution to monitor an esterification reaction by near-infrared spectroscopy.

IV) Application of multivariate curve resolution to chemical process control of an esterification reaction monitored by near-infrared spectroscopy

ANEXOS

I) Study of reaction processes by in-line near-infrared spectroscopy in combination with multivariate curve resolution. Esterification of myristic acid with isopropanol.

II) Determination of low analyte concentrations by near-infrared spectroscopy: Effect of spectral pretreatments and estimation of multivariate detection limits.

El objetivo fundamental de la Tesis Doctoral es la aplicación de la espectroscopia en el infrarrojo cercano (NIR) en la monitorización y control de procesos químicos, desarrollando métodos que permiten el control analítico de parámetros de reacción y/o el seguimiento de las especies involucradas en la reacción. Para tal objetivo se han aplicado diversas técnicas quimiométricas, destacando entre ellas, PCA (*Principal Component Analysis*), PLS (*Partial Least Squares*) y MCR-ALS (*Multivariate Curve Resolution-Alternating Least Squares*).

Una reacción de un particular interés industrial es la esterificación de la glicerina con ácidos de cadena larga, que da lugar a una amplia variedad de productos según la naturaleza de los ácidos y las proporciones de ácido/glicerina que dan lugar a mezclas de mono-, di- y triglicéridos y de mezclas en diferentes proporciones de ellos. En la Memoria se ha estudiado la esterificación entre glicerina y una mezcla de ácidos caprílico y cáprico; este proceso se controla frecuentemente usando métodos volumétricos de determinación del índice de ácido (IA) y el índice de hidroxilo (IOH), siendo su determinación lenta, requieren el uso de reactivos tóxicos y solventes y con una reducida capacidad informativa. Con el objetivo de sustituir esta metódica, se prepararon modelos de calibración multivariable entre los espectros NIR de las muestras y los valores de referencia determinados por los métodos convencionales; para ello se extrajeron muestras de procesos realizados a escala de laboratorio y a escala industrial, se registró su espectro NIR *at-line* y se analizaron por los correspondientes métodos de referencia. Se construyeron modelos de calibración PLS1 para IOH, IA y agua, que proporcionan resultados comparables a los ofrecidos por los métodos de referencia y que pueden utilizarse para un control rápido de estos tres parámetros a partir únicamente del registro espectral NIR.

Cuando la reacción anterior se realiza con exceso de glicerina el producto obtenido es una mezcla de mono, di y triglicéridos, y su composición (proporción entre los diferentes componentes) es determinante para asegurar que el producto cumplirá con las especificaciones pre-establecidas. Por ello se ha estudiado el desarrollo de modelos de calibración para el seguimiento de las especies involucradas en esta reacción (glicerina, ácidos y glicéridos). Se extrajeron muestras de procesos realizados a escala industrial y de laboratorio, y además se prepararon muestras dopadas (muestras obtenidas por mezcla de muestras de síntesis con alguno de los

reactivos o productos) en el laboratorio para disminuir la correlación de concentraciones existente entre los diferentes compuestos a determinar. Se registraron los espectro NIR *at-line* y se analizaron las muestras por cromatografía de gases para la obtención de las concentraciones de referencia. Se construyeron modelos PLS1 para cada componente, que permiten la determinación de todas las especies involucradas en la reacción a partir de un solo espectro NIR y reduciendo considerablemente el tiempo de análisis en comparación a otras técnicas instrumentales. Los resultados obtenidos en la validación de los dos métodos PLS desarrollados demuestran que la calidad de los modelos es adecuada y suficiente para el control *at-line* de las diferentes especies y permiten actuar sobre el proceso para su control de forma inmediata.

Este proceso de esterificación se estudió posteriormente mediante la utilización de Resolución Multivariante de Curvas (MCR-ALS) con el objetivo de conocer la capacidad de esta técnica para el estudio de procesos en evolución. La aplicación de MCR-ALS, usando los espectros correspondientes al proceso realizado con exceso de glicerina, resultando ser un sistema deficiente de rango. Con el objeto de resolver la deficiencia de rango, se estudió el uso de restricciones de igualdad en el proceso iterativo ALS, incorporando los perfiles de concentración de las especies de este proceso. Esta estrategia permitió obtener los espectros puros de las especies involucradas en la reacción. Así se obtuvieron mediante cálculo por MCR-ALS los espectros de las especies para los que no se disponía de patrones de un adecuado nivel de pureza. Estos espectros fueron utilizados posteriormente para la estimación de los perfiles de concentración en nuevos procesos. De esta forma se requiere menos información de referencia en comparación a la construcción de un modelo PLS, lo que supone una ventaja para el desarrollo de modelos de calibración y para el estudio de procesos en evolución.

Posteriormente, se estudio el proceso con defecto de glicerina en relación al contenido de ácidos, que da lugar a la formación de triglicérido principalmente. En este caso, la deficiencia de rango se resolvió aumentando la matriz espectral con los espectros puros de los componentes del sistema, utilizándose éstos como estimaciones iniciales en el proceso iterativo, y se incluyó un mínimo de información cuantitativa como restricción de igualdad. Los perfiles de concentración obtenidos permiten establecer la evolución de cada una de las especies y también calcular los

valores de IA e IOH. Los modelos desarrollados son adecuados para la predicción de estos parámetros en nuevas muestras. Los resultados obtenidos por MCR-ALS para la predicción de muestras externas son comparables a los obtenidos por PLS en términos de RSEC/P (%), con la importante ventaja de que MCR requiere menos información de referencia.

Una metodología semejante de trabajo se siguió en el estudio, mediante el uso de espectroscopia NIR *in-line*, de la reacción de esterificación entre ácido mirístico e isopropanol, en la que éste último se añadió en continuo al medio de reacción. Con la experiencia adquirida en el uso de MCR-ALS, se estudió la aplicación de esta técnica para la obtención de los perfiles de concentración sin la necesidad de recurrir a técnicas analíticas de referencia. A semejanza del estudio anterior, la deficiencia de rango se resolvió añadiendo a la matriz de información espectral *in-line* los espectros puros de los componentes; de este modo se obtuvieron los perfiles de concentración de las especies involucradas en la reacción a partir únicamente de información espectral NIR. Los perfiles de concentración obtenidos fueron utilizados para el cálculo de constantes de velocidad, estableciendo que la reacción sigue una cinética de pseudo-primer orden, que la cinética de la reacción depende de la velocidad de adición del isopropanol y del contenido en agua de éste. El método NIR-MCR propuesto permite determinar la influencia de parámetros experimentales, de forma rápida y eficaz, sin la necesidad de análisis de referencia.

Finalmente, se utilizó la espectroscopia NIR junto con el PLS para la determinación de bajas concentraciones de un analito aplicado a un proceso de purificación de un ester mediante la eliminación de 2-etilhexanol por destilación. Se estudió el efecto de dos pretratamientos espectrales, el OSC (*Orthogonal Signal Correction*) y el NAS (*Net Analyte Signal*) para la construcción de modelos de calibración en un intervalo de concentraciones próxima al límite de detección. Ambos métodos presentan una elevada capacidad para eliminar la señal no relacionada con el analito estudiado, permitiendo una reducción del número de factores PLS necesarios para construir el modelo de calibración. Sin embargo, no mejora la capacidad, en términos de RSE (%), de predicción de muestras externas. Este estudio se ha completado con la aplicación de dos metodologías para el cálculo del límite de detección del modelo PLS desarrollado: la primera está basada en el cálculo de la varianza de predicción del modelo PLS y la segunda es una adaptación

RESUMEN

de la metodología univariante. Ambas metodologías permiten la estimación del límite de detección en modelos de calibración multivariable de forma sencilla, y confirman los resultados obtenidos anteriormente. El modelo PLS obtenido permite el seguimiento del proceso de purificación, deteniéndolo una vez se alcanza la concentración especificada, evitando aplicar un exceso de tiempo en la destilación que provoca un deterioro de las características organolépticas del producto final.

ABSTRACT

The aim of this work is the application of Near-Infrared Spectroscopy (NIR) to chemical process control, developing methods that allow the analytical control of the reaction parameters and/or the follow-up of the compounds involved in the reaction. For such an aim, different chemometrical tools have been applied, standing out between them, PCA (Principal Component Analysis), PLS (Partial Least Squares) and MCR-ALS (Multivariate Curve Resolution-Alternating Least Squares).

One reaction with industrial interest is the esterificación of glycerine with medium and long chain fatty acids that gives a wide variety of products, according to the nature of the acids and the proportions of acid/glycerine, which give place to mixtures of mono-, di- and triglycerides in different proportions. In this thesis, the esterification between glycerine and a mixture of caprylic and capric acids to obtain a triglyceride has been studied. The process is frequently controlled by using titration methods such as the Acid Value (AV) and the Hydroxyl Value (OHV), which are time-consuming, need the use of toxic reagents and solvents, and give poor information. To replace these methodologies, calibration models were built with the NIR spectra of the samples and the respectively reference values. The samples were obtained from laboratory and industrial scale esterification processes, being the NIR spectrum registered at-line. PLS1 calibration models were built for OHV, AV and water content, being the results obtained by NIR comparable to the provided ones by the reference methods, which allow the rapid control of these three parameters departing only from the NIR spectrum.

When the above reaction is done with an excess of glycerine, the obtained product it is a mixture of mono-, di- and triglycerides, and its composition is determinant to assure that the product will reach the pre-established specifications. In that case, it is interesting the construction of a PLS model for each of the compounds involved in the reaction. For this aim, samples were extracted from industrial and laboratory scale processes, and doped samples were prepared in order to reduce the concentration correlation between the different target compounds. NIR spectra were registered at-line, and the samples were analyzed by gas chromatography for to obtain the reference concentrations. PLS1 models were built for each component, which allow the determination of all the species involved in the reaction with only the spectrum NIR, reducing the time of analysis in comparison to

ABSTRACT

other instrumental techniques. The results obtained in the validation of the PLS models show that the quality of the models is adapted and sufficiently for the at-line control of the reaction, and allow to act on the process for its control.

This esterification process (esterification with a glycerine excess) was studied later by Multivariate Curve Resolution (MCR-ALS) with the aim to know the capacity of this technique for the study of reaction processes. The system studied, the spectra corresponding to the process realized with excess of glycerine, turned out to be a rank deficient system. In order to solve the rank deficiency, it was studied the use of equality constraints in the iterative process by the incorporation of the concentration profiles of the chemical species involved in the reaction; this strategy allowed to obtain the pure spectra of them. This way, the spectra of chemical species not available in high purity were obtained by MCR-ALS. These spectra were used later for the estimation of the concentration profiles in new processes; this form, less reference information is required in comparison to the construction of a PLS model, which supposes an advantage for the development of calibration models and for the study of reaction processes.

Subsequently, the reaction to obtain a triglyceride was studied. In this case, the rank deficiency was solved augmenting the data matrix with the pure spectra of the components involved in the reaction, being also used as initial estimations in the iterative process, and including a minimum of quantitative information as equality constraints. The concentration profiles obtained allow establishing the evolution of the species and, subsequently, to calculate the AV and OHV values in new samples. MCR-ALS results for the prediction of external samples are comparable to the obtained ones for PLS in terms of RSEC/P%, with the important advantage that MCR needs less reference information.

A similar methodology was applied in the study by in-line near-infrared spectroscopy of the esterification reaction between myristic acid and isopropanol, the latter being added continuously. The aim was the obtaining of the concentration profiles by MCR-ALS without the need to resort to analytical reference data. Similarly to the previous study, the rank deficiency was solved by adding to data matrix the pure spectra of the components involved in the reaction; thus, the concentration profiles were obtained departing only from spectral NIR information. The

ABSTRACT

concentration profiles obtained were used for the calculation of kinetic constants, establishing that the reaction is pseudo-first order, that the kinetics of the reaction depends on the speed of the addition of isopropanol, and also, in the water content of this one. The NIR-MCR proposed method allows determining the influence of experimental parameters in a rapid and effective way, without the need of reference analysis.

Finally, NIR spectroscopy was used in combination with PLS calibration for the determination of low analyte concentrations, applied to a purification process: the elimination of 2-ethylhexanol by distillation. The effect of two spectral pretreatments, OSC (Orthogonal Signal Correction) and NAS (Net Analyte Signal), was studied for the determination of concentrations near the detection limit. Both methods have a high capacity to reduce the NIR signal not correlated with the analyte concentration, allowing the reduction of the number of factors necessary to construct the PLS model. Nevertheless, both pretreatments do not improve the prediction results with external samples in terms of RSE%. This study has been completed by the application of two methodologies for the estimation of the detection limit of the PLS model: the first is based in the estimation of the prediction variance of the PLS model; the second is an adaptation of the univariate methodology proposed by the ISO. Both methodologies allow the estimation of the detection limit in multivariate calibration models in a simple form; and confirm that OSC and NAS do not improve the detection limit. The PLS model obtained allows monitoring the distillation process, stopping the process once the specified concentration is reached, avoiding to apply a time-out in the distillation that provokes a deterioration of the sensory properties of the final product.

A lo largo de la tesis se ha utilizado la siguiente simbología y las abreviaturas recogidas en este apartado.

Notación

n	escalar
N	dimensión de un vector o matriz
\mathbf{n}	vector
\mathbf{N}	matriz
\mathbf{X}^T	matriz transpuesta
\mathbf{X}^+	matriz pseudoinversa
\hat{y}	valor predicho
\bar{y}	valor medio

Abreviaturas

ALS	<i>Alternating Least Squares</i>
DL	<i>Detection Limit</i>
IA	Índice de Ácido
IOH	Índice de Hidroxilo
NAS	<i>Net Analyte Signal</i>
MCR	<i>Multivariate Curve Resolution</i>
MDL	<i>Multivariate Detection Limit</i>
MIR	<i>Middle Infrared</i>
MSC	<i>Multiplicative Scatter Correction</i>
MSECV	<i>Mean Square Error of Cross Validation</i>

Notación y Abreviaturas

NIPALS	<i>Non-linear Iterative Partial Least Squares</i>
NIR	<i>Near-Infrared Analysis</i>
OSC	<i>Orthogonal Signal Correction</i>
PC	<i>Principal Component</i>
PCA	<i>Principal Component Analysis</i>
PCR	<i>Principal Component Regression</i>
PLS	<i>Partial Least Square</i>
PRESS	
RSEC %	<i>Relative Standard Error of Calibration</i>
RSEP %	<i>Relative Standard Error of Prediction</i>
RMSEC	<i>Root Mean Square Error of Calibration</i>
RMSEP	<i>Root Mean Square Error of Prediction</i>
SNV	<i>Standard Normal Variate</i>



OBJETIVOS

OBJETIVOS

El desarrollo de metodologías de análisis rápidas, sencillas y fiables es una de las demandas actuales de la industria química para mejorar la eficiencia de los procesos de producción. La espectroscopia en el infrarrojo cercano reúne las características que la hacen adecuada para satisfacer estas demandas, por lo que esta técnica fue la seleccionada para el estudio de su aplicación a diferentes procesos industriales.

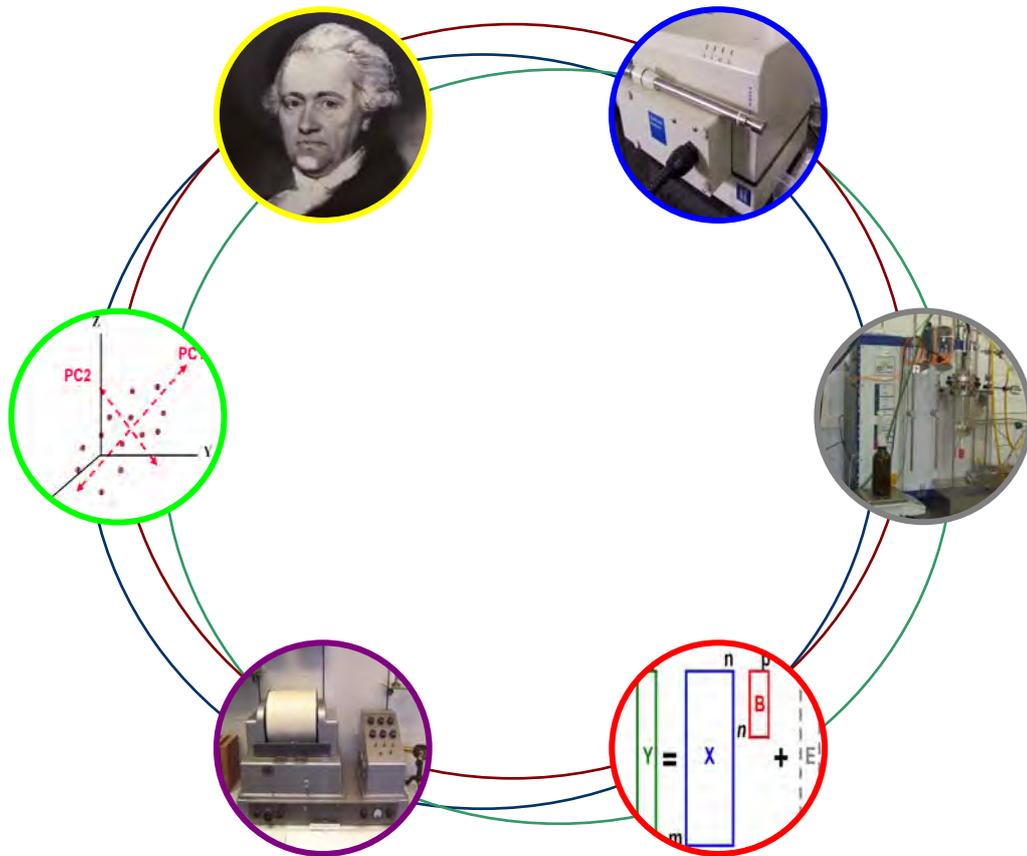
El principal objetivo de la presente tesis doctoral es el desarrollo y la aplicación de nuevas metodologías analíticas para la determinación de parámetros y especies químicas en procesos químicos industriales.

Este objetivo general se puede describir en los siguientes apartados:

1. Establecer y adquirir conocimiento de procesos químicos, así como el desarrollo de métodos analíticos de referencia para la determinación de los parámetros o especies químicas de interés y objeto del estudio.
2. Desarrollo de modelos de calibración multivariantes basados en las medidas espectroscópicas. Los modelos desarrollados han de ser robustos y con adecuada capacidad predictiva para poder ser utilizados en el control del proceso industrial.
3. Aplicación de técnicas quimiométricas de resolución de curvas a los datos espectroscópicos con el objeto de desarrollar modelos con menores exigencias de información primaria y que resulten adecuados para su aplicación a procesos industriales.

OBJETIVOS

4. Estudio comparativo de los modelos desarrollados y de sus diferentes exigencias en su aplicación al control de procesos en la industria química.



INTRODUCCIÓN

El objetivo de este capítulo es introducir conceptos del control de procesos en la industria química, así como resumir los fundamentos teóricos de la espectroscopia NIR y de las técnicas quimiométricas aplicadas.

Contenidos:

1. Control de procesos en la industria química
2. Espectroscopia NIR
3. Quimiometría



1. Control de procesos en la industria química

1. Control de procesos en la industria química

1.1. Introducción general

La química analítica de procesos se ha desarrollado durante los últimos 50 años debido a la necesidad de tener información cerca de la línea de producción, con el objetivo de ahorrar tiempo y dinero, aumentar la eficiencia y mejorar la seguridad [1]. En 1987 Callis [2] acuñó el término PAC (*Process Analytical Chemistry*), Química Analítica de Procesos, para definir las aplicaciones analíticas cuyo objetivo es proporcionar información cualitativa y cuantitativa de un proceso químico. Esta información puede ser usada no solo para monitorizar y controlar el proceso, sino también para optimizar el uso de energía, tiempo y materias primas. Actualmente, el control analítico de procesos es un campo establecido dentro de la Química Analítica, como lo evidencia el gran número de publicaciones existente. La revista *Analytical Chemistry* edita revisiones bianuales dedicadas al control de procesos [1-7], dedicando un apartado específico a las técnicas espectroscópicas, y especialmente, a la espectroscopia NIR.

Tradicionalmente, las medidas realizadas durante el proceso correspondían a temperatura, presión o velocidades de adición de reactivos. En la actualidad, el control de procesos requiere conocer información compleja a tiempo real, como puede ser la composición o concentraciones de uno o más componentes, o incluso propiedades estructurales. Esta evolución ha sido posible gracias al desarrollo de nuevas técnicas analíticas o a la adaptación de las ya existentes al análisis en proceso. Así, la miniaturización de instrumentación cromatográfica permite la determinación on-line de mezclas complejas en intervalos de tiempo inferiores a 1 minuto, permitiendo su aplicación en el control de procesos casi a tiempo real. La adaptación de técnicas espectrofotométricas en sistemas on-line con la

implementación de metodologías FIA (*Flow Injection Analysis*) o SIA (*Sequential Injection Analysis*), o la utilización de sondas de medida in-line, permiten reducir o eliminar la etapa de muestro, ampliando el uso de estas técnicas para el control de procesos. Estas innovaciones, entre otras, permiten la automatización de los sistemas de análisis, haciendo que la industria química sustituya de forma progresiva los 'analistas de laboratorio' por 'analizadores de proceso'.

1.2 Procesos en discontinuo o *batch processes*

Los procesos realizados en discontinuo, o por lotes, del término inglés *batch process*, son frecuentes en múltiples industrias químicas por su simplicidad de operación. En un proceso industrial en discontinuo los reactivos se introducen inicialmente en un reactor, equipado con un sistema de agitación para favorecer la homogeneidad de la mezcla. Posteriormente, incrementando la temperatura i/o por la adición de un iniciador, la reacción tiene lugar hasta la obtención del producto deseado. Finalmente, el reactor es vaciado para volver a comenzar el siguiente proceso. La ventaja de este tipo de reactores, además de la simplicidad de operación, es su flexibilidad, ya que generalmente están diseñados para poder realizar procesos con diferentes características, pudiendo obtenerse una amplia gama de productos en un mismo reactor. En un proceso en continuo la configuración del reactor puede ser mucho más compleja, lo que implica que generalmente son específicos para la producción de un determinado producto. La diferencia principal es que en un proceso en discontinuo la concentración de reactivos y productos evoluciona, mientras que en un proceso en continuo fluctúa alrededor de un valor en equilibrio (figura 1.1).

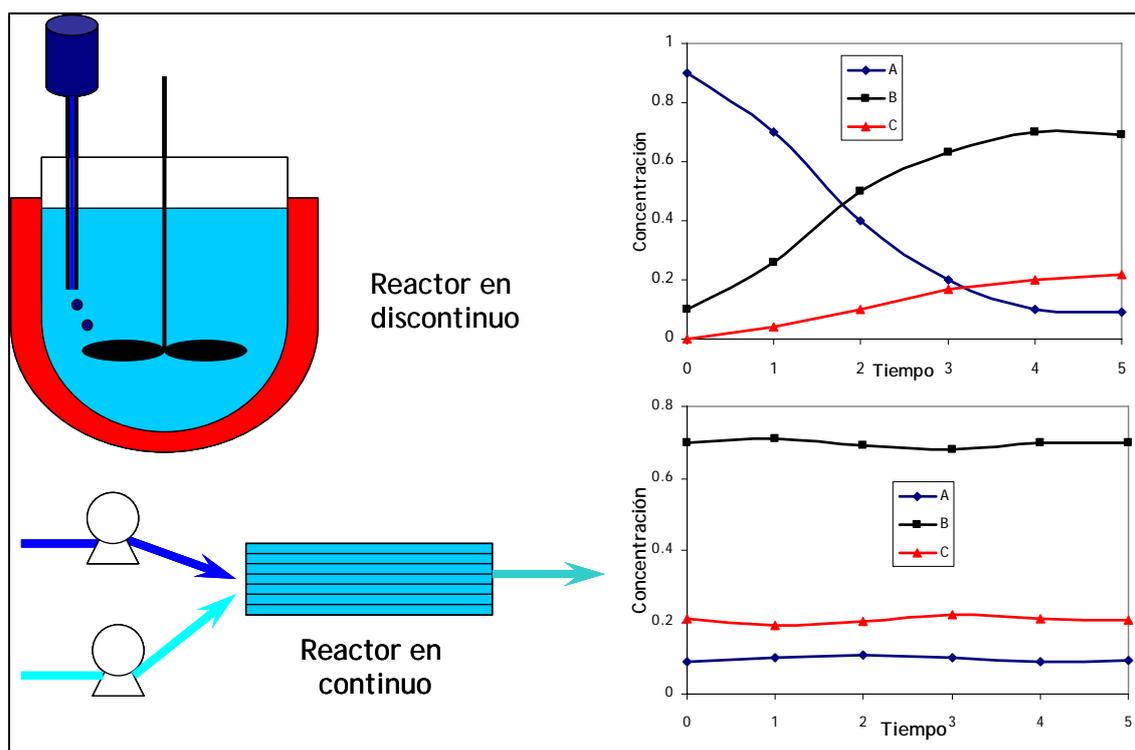


Figura 1.1. Representación de un proceso en discontinuo (arriba) y en continuo (abajo).

Las características principales de los procesos en discontinuo se detallan a continuación:

- Están establecidas las cargas de reactivos, temperaturas, presiones o velocidades de dosificación. Sin embargo, el control de estos parámetros puede no ser preciso, provocando una elevada variabilidad en el producto final.
- El tiempo de reacción es finito y dependerá del tipo de proceso.
- La conversión de reactivos generalmente es elevada (> 90%).
- Pueden trabajar en semi-batch, donde uno de los reactivos es añadido en continuo durante la reacción. Este procedimiento aporta ventajas como la minimización de reacciones secundarias, o mantener baja la concentración del reactivo para evitar su destilación.
- Son procesos de no-equilibrio. Las características deseadas en el producto

final se pueden obtener antes de llegar al estado estacionario.

A pesar de las características comentadas anteriormente, este tipo de procesos tienen una serie de limitaciones e inconvenientes importantes respecto a los procesos en continuo: tienen múltiples etapas, poseen intervalos de tiempo elevados entre operaciones (carga, descarga, limpieza), contaminación entre procesos, y la más importante, una mayor dificultad en el control analítico, ya que se trata de procesos en evolución, provocando una mayor variabilidad en el producto final.

Los procesos en discontinuo cubren una amplia variedad de procesos como son la producción de polímeros o productos de esterificación en la industria química (siliconas, poliestireno, aceites, PVC, látex, resinas, etc.), la obtención de principios activos en industria farmacéutica o la producción de cerveza en la industria alimentaria, etc.

1.3 Monitorización de procesos en discontinuo

En el apartado anterior se han comentado las desventajas de la producción por *batch*, como la reducida reproducibilidad en la fijación de las condiciones experimentales que provoca una elevada variabilidad del producto final. Por este y otros motivos, comentados a continuación, es necesario llevar a cabo un adecuado control de este tipo de procesos.

1.3.1 Consistencia del proceso y del producto

La monitorización de las variables del proceso permite comprobar si se ha seguido correctamente el procedimiento establecido para realizar el proceso. Siempre existirán desviaciones de las variables respecto a lo indicado en el procedimiento, cuanto más grandes sean estas desviaciones, mayor probabilidad existe de que el

producto final no cumpla las características deseadas. Es importante el control del proceso con el fin de obtener un producto de mayor consistencia entre los diferentes lotes y evitar lotes anómalos.

1.3.2 Alteraciones durante el proceso

Los procesos en discontinuo suelen tener múltiples etapas, provocando que sea complicado encontrar la causa que ha provocado que el producto final esté fuera de especificaciones. Los errores pueden venir desde una carga errónea del reactor, hasta un incumplimiento por el operario del procedimiento correcto.

1.3.3 Optimización del proceso

Una adecuada monitorización de un proceso en discontinuo implica el conocimiento de la evolución de los compuestos implicados en la reacción. Estudiar la variación del comportamiento de las especies involucradas en la reacción frente a cambios en las variables experimentales permite la optimización del proceso en términos de consumos de materia prima, energía y tiempo.

1.3.4 Seguridad y medio ambiente

La obtención del producto adecuado con el menor consumo de materias primas y energía es el objetivo de cualquier empresa. La obtención de lotes anómalos, además de un sobrecoste importante, es un problema de difícil solución, por el coste que implica la gestión de residuos. El endurecimiento de las leyes en temas medioambientales hace cada vez mayor la factura que han de pagar las empresas. Finalmente, la monitorización de procesos también es un factor importante para la seguridad. El control de la temperatura, la presión o la concentración de productos de elevada reactividad evita posibles situaciones de riesgo.

1.4 Metodologías para el control de procesos

Como ha sido comentado al principio de la sección 1.1, Callis [1] en 1987 estableció una clasificación de las metodologías aplicables para su realización, detalladas a continuación:

- ⇒ *Off-line*: la muestra extraída del reactor es transportada a un laboratorio donde es analizada. Esto permite que el análisis sea realizado por personal especializado, con instrumental avanzado y en un ambiente más limpio que el de una planta de producción. Sin embargo, el coste y el retardo de este tipo de análisis suponen un importante inconveniente cuando el número de análisis es grande y sobretodo, cuando el tiempo es un factor decisivo.
- ⇒ *At-line*: toma de muestra manual y proximidad de instrumentación analítica a la línea de producción. Permiten una reducción del tiempo del análisis respecto a los off-line. Sin embargo, no permiten la realización de análisis de cierta complejidad al no existir el ambiente limpio de un laboratorio, y por el uso de instrumentación sencilla y personal no especializado.
- ⇒ *On-line*: toma de muestra y transporte al instrumento automatizado. Una vez la muestra ha sido acondicionada, el análisis puede ser discontinuo (métodos, como los cromatográficos, que requieren la inyección de una alícuota) o continuo (métodos que permiten el flujo continuo de muestra, como los espectrofotométricos).
- ⇒ *In-line*: análisis *in-situ* mediante la introducción de sondas de medida dentro de la línea de proceso.
- ⇒ *Non invasive*: eliminación por completo del contacto entre el analizador y la muestra.

La elección de uno u otro sistema depende de muchos factores, intrínsecos y

extrínsecos al proceso de producción, como pueden ser la rapidez con la que se han de obtener los resultados, el instrumental necesario, el coste del análisis, el pretratamiento de la muestra, el personal necesario para el análisis, etc.

La clasificación anterior sigue siendo ampliamente vigente en la actualidad, a pesar del desarrollo de nuevas metodologías que ha provocado la creación de nuevos términos para describir la amplia aplicabilidad de los analizadores de proceso. Así términos ingleses como [2]: *automatic, fieldable, hyperspectral, in situ, near-line, nondestructive, portable, rapid, real-time* o *remote*, son cada vez más frecuentes.

Referencias

[1] J. Workman, D. J. Veltkamp, S. Doherty, B. B. Anderson, K. E. Creasy, M. Koch, J. F. Tatera, A. L. Robinson, L. Bond, L. W. Burgess, G. N. Bokerman, A. H. Ullman, G. P. Darsey, F. Mozayeni, J. A. Bamberger, M. S. Greenwood. Process analytical chemistry. *Anal. Chem.* **71** (1999) 121R-180R.

[2] J. B. Callis, D. L. Illman, B. R. Kowalski. Process Analytical Chemistry. *Anal. Chem.* **59** (1987) 624A-637A.

[3] K. R. Beebe, W. W. Blaser, R. A. Bredeweg, J. P. Chauvel, R. S. Harner, M. LaPack, A. Leugers, D. P. Martin, L. G. Wright, E. D. Yalvac. Process analytical chemistry. *Anal. Chem.* **65** 12 (1993) 199R-216R.

[4] W. W. Blaser, R. A. Bredeweg, R. S. Harner, M. A. LaPack, A. Leugers, D. P. Martin, R. J. Pell, J. Workman, L. G. Wright. Process Analytical Chemistry. *Anal. Chem.* **67** 12 (1995) 47R-70R.

[5] J. Workman, K. E. Creasy, S. Doherty, L. Bond, M. Koch, A. Ullman, D. J. Veltkamp. Process Analytical Chemistry. *Anal. Chem.* (Review) **73** 12 (2001) 2705-2718.

[6] J. Workman, M. Koch, D. J. Veltkamp. Process Analytical Chemistry. *Anal. Chem.* (Review) **75** 12 (2003) 2859-2876.

[7] J. Workman, M. Koch, D. Veltkamp. Process Analytical Chemistry. *Anal. Chem.* (Review) **77** 12 (2005) 3789-3806.



2. Espectroscopia NIR

2. Espectroscopia NIR

2.1 Introducción

La radiación NIR (*Near-InfraRed*) es la radiación del espectro electromagnético comprendida en el intervalo de 780 a 2500 nm, y está situada entre la radiación MIR (*Middle InfraRed*) y la UV-Vis (figura 2.1).

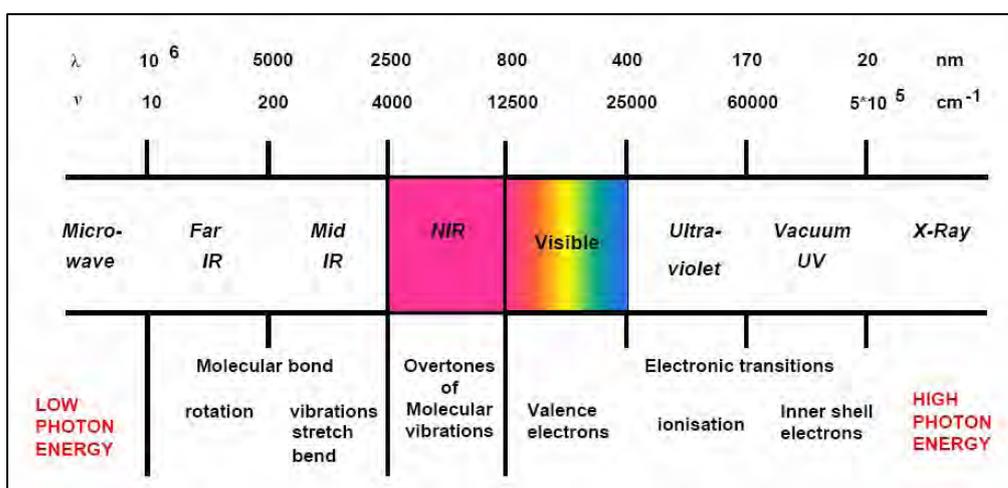


Figura 2.1. Espectro electromagnético [1]

El descubrimiento de la radiación en la región del infrarrojo cercano es debido a Sir William Herschel [2], cuando en el año 1800 demostró que existía radiación por debajo del espectro visible (figura 2.2). A pesar del temprano descubrimiento, no fue

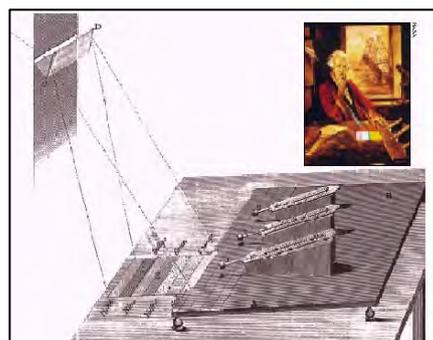


Figura 2.2. Experimento de Herschel

hasta 1881 cuando Abney y Festing [3] registraron el espectro de líquidos orgánicos en la región de 700 a 1200 nm. Sin embargo, durante las posteriores décadas el interés de los espectroscopistas estuvo centrado en la región del infrarrojo medio, convencidos del poco interés analítico de la región del infrarrojo cercano.

No fue hasta la década de los 50, con la aparición de los primeros espectrofotómetros NIR, cuando se desarrollaron las primeras aplicaciones de esta región del espectro electromagnético. W. Kaye [4] en 1954 asignó bandas espectrales de los correspondientes grupos funcionales en la región comprendida entre 700 y 3500 nm, y mostró alguna de las posibles aplicaciones analíticas de la técnica como la determinación de agua y alcoholes, y el análisis de aminas e hidrocarburos. Publicaciones posteriores describen determinaciones cuantitativas utilizando la región NIR [5,6].

En los años 60, el grupo de investigación dirigido por Karl Norris utilizó la espectroscopia NIR para el estudio de muestras complejas de origen natural [7,8], principalmente en el ámbito agroalimentario. Estos estudios abrieron un nuevo campo de aplicaciones, incrementando considerablemente el interés por esta técnica.



Figura 2.3. Norris con el NIR Cary 14. [9]

Después de esta etapa inicial, y con avances en el desarrollo de nueva instrumentación, con electrónica más estable y ópticas mejoradas, a partir de finales de los 70 se hizo posible el registro de espectros de manera rápida y reproducible (instrumentos de barrido, interferómetros) [10,11]. El desarrollo de la informática también ayudó al posterior desarrollo y aplicación de la técnica: a diferencia de otras técnicas espectroscópicas, en el infrarrojo cercano no se producen picos aislados característicos para cada compuesto, sino que hay superposición de bandas correspondientes a diferentes grupos funcionales. Para extraer la información contenida en el espectro son necesarias técnicas quimiométricas de análisis multivariable. La posibilidad de utilizar ordenadores con mayor capacidad de cálculo y memoria permitió obtener y almacenar gran cantidad de información espectral, que podía ser tratada por técnicas quimiométricas, y así poder extraer la información

analítica de interés [12-14]. El auge del NIR como técnica de análisis se ha dado en los años 80 y 90 gracias a los avances en tecnología. En 1983 Wetzel la definía como "*Sleeper among spectroscopic techniques*" [15], indicando que la técnica tenía un gran potencial, aunque hasta aquel momento no era utilizada como técnica habitual de análisis. En 1994, McClure dijo del NIR "*The giant is running strong*" [16] refiriéndose al elevado número de aplicaciones del NIR como técnica de análisis.

Actualmente, las aplicaciones de la espectroscopia NIR abarcan el análisis de productos agroalimentarios, farmacéuticos, de fibras textiles, industria química y petroquímica, e incluso aplicaciones médicas, entre otros diversos ámbitos.

Los motivos que hacen esta técnica atractiva son diversos: permite el registro de sólidos, líquidos y gases; no requiere pretratamiento de muestra; el registro espectral es rápido; es una técnica no destructiva; permite el análisis multicomponente; el análisis de rutina es rápido; es una técnica económica y además, permite el análisis en línea de producción.

Como el gran desarrollo del NIR se ha producido en los últimos 25 años, se podría esperar que, hoy por hoy, fuese una técnica madura próxima a una fase de estancamiento. Sin embargo, el desarrollo continuo de nuevas técnicas quimiométricas, la miniaturización de los componentes instrumentales, el desarrollo de la espectroscopia NIR de imagen, y la continua aplicación a nuevos campos, hacen que esta técnica esté en plena fase de expansión [17], siendo la dominadora de las técnicas espectroscópicas en el análisis a tiempo real y el análisis de proceso [18].

2.2 Fundamentos

2.2.1 Origen de la absorción en el infrarrojo

La región infrarroja del espectro electromagnético está comprendida entre 780 y 10^6 nm (figura 2.1), y está dividida en tres subregiones: FIR (*Far InfraRed*) infrarrojo lejano, MIR (*Middle InfraRed*) infrarrojo medio y NIR (*Near InfraRed*) infrarrojo cercano. La absorción en cada una de estas tres regiones es debida a interacciones de diferente naturaleza entre luz y materia (tabla 2.1).

Región	Intervalo de absorción (nm)	Origen de la absorción
FIR	$4 \cdot 10^4 - 10^6$	Rotaciones moleculares
MIR	2500 - 40.000	Vibraciones moleculares fundamentales
NIR	780 - 2500	Sobretonos y bandas de combinación de vibraciones moleculares

Tabla 2.1. Regiones del espectro electromagnético entre 190 y 10^6 nm.

Para que una molécula absorba radiación IR, la energía de la radiación incidente ha de coincidir con la diferencia de energía entre dos estados energéticos y debe producirse un cambio en el momento dipolar de la molécula debido a movimientos vibratorios. Los movimientos vibratorios que puede presentar una molécula son de dos tipos (figura 2.4), los de **estiramiento**: vibración en la que se produce un cambio en la distancia interatómica a lo largo del eje de enlace entre dos átomos; y los de **flexión**: se produce un cambio en el ángulo de enlace.

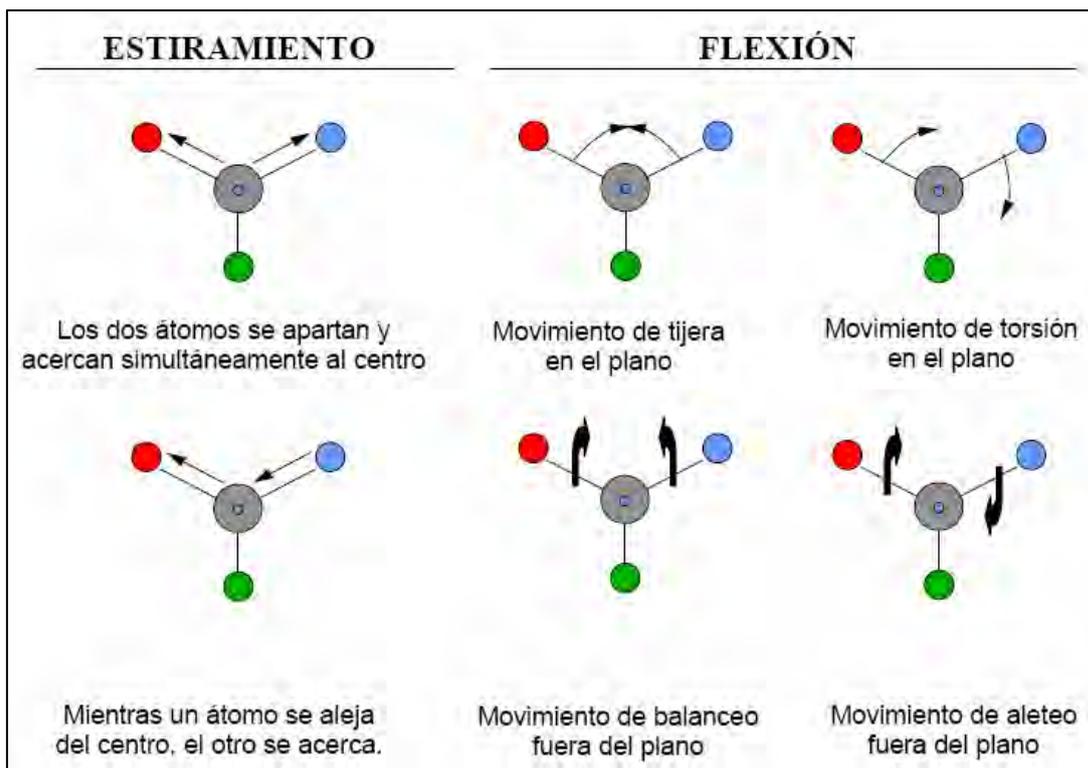


Figura 2.4. Vibraciones moleculares entre dos átomos respecto a un átomo central.

Los átomos en una molécula nunca se encuentran inmóviles, ni tan siquiera cerca del cero absoluto. El movimiento de dos átomos en un enlace se puede aproximar a un oscilador armónico simple (figura 2.5), donde el movimiento de los átomos está confinado en un pozo de potencial, la energía potencial cambia continuamente, acercándose al mínimo cuando los átomos se encuentran en la posición de equilibrio, y haciéndose máxima cuando la distancia es máxima o mínima.

La energía potencial del enlace definido por un oscilador armónico simple viene dada por la ecuación:

$$E = \frac{1}{2} kx^2 \quad [2.1]$$

Siendo k la constante de fuerza del enlace y x la distancia interatómica.

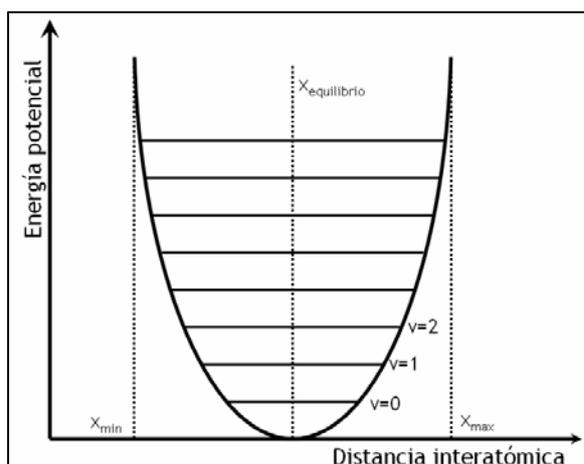


Figura 2.5. Energía potencial para el oscilador armónico

La frecuencia de vibración de un sistema formado por dos masas unidas por un muelle perfectamente elástico viene dada por:

$$\nu = \sqrt{\frac{1}{2\pi} \left(\frac{k(m_1 + m_2)}{m_1 m_2} \right)} \quad [2.2]$$

Siendo ν la frecuencia, k la constante de fuerza y m_1 y m_2 las masas de cada uno de las masas o átomos.

Como otras energías moleculares, la energía vibracional está cuantizada. Para un oscilador armónico los niveles de energía vienen dados por:

$$E_{vib} = \left(\nu + \frac{1}{2} \right) h\nu \quad [2.3]$$

Donde ν es el número cuántico vibracional que sólo puede adquirir valores enteros ($\nu = 0, 1, 2, \dots$). Según la teoría cuántica, las únicas transiciones posibles son entre niveles energéticos consecutivos, $\Delta\nu = \pm 1$. Así, cada enlace tendría una energía de absorción característica correspondiente a la transición $\Delta\nu_{0,1}$.

Este modelo vibracional es incompleto para moléculas reales, donde se ha de tener en cuenta la energía de repulsión coulombica de dos átomos cuando se aproximan, que provoca un mayor aumento de la energía potencial; y en el otro extremo, la disminución de la energía potencial cuando la distancia interatómica se acerca a aquella en la que tiene lugar la rotura del enlace. El comportamiento de las moléculas reales se acerca más al de un oscilador anarmónico (figura 2.6).

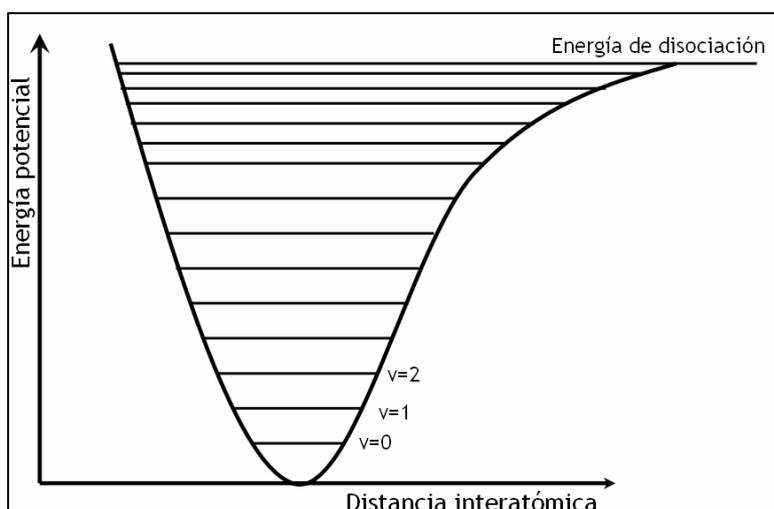


Figura 2.6. Energía potencial para el oscilador anarmónico.

La expresión de la energía vibracional (ecuación 2.3) corregida con términos que incluyen la anarmonicidad se expresa como:

$$E_{vib} = \left(\nu + \frac{1}{2}\right)h\nu - \left(\nu + \frac{1}{2}\right)^2 h\nu y - \left(\nu + \frac{1}{2}\right)^3 h\nu y y' \dots \quad [2.4]$$

donde y, y', \dots son conocidas como constantes de anarmonicidad. Para el oscilador anarmónico las transiciones entre niveles de energía consecutivos se define como:

$$\nu = \nu_{fund} \left[1 - y \left(\nu + \frac{1}{2} \right) \right] \quad [2.5]$$

En el oscilador anarmónico la frecuencia de vibración disminuye al aumentar ν , consecuencia de que los niveles no están igualmente espaciados, y pueden ser observadas transiciones $\Delta\nu = \pm 2, \pm 3, \dots$ las cuales son conocidas como sobretonos (primer y segundo sobretono respectivamente), de menor probabilidad, y que generalmente aparecen en la región del infrarrojo próximo.

2.2.2 Origen de la absorción en el infrarrojo cercano

La absorción en el infrarrojo cercano corresponde a sobretonos y bandas de combinación de las vibraciones fundamentales observadas en el infrarrojo medio. Los sobretonos, que corresponden a transiciones con $\Delta\nu \geq 2$, son aproximadamente múltiples de la frecuencia vibracional fundamental. De forma similar, las bandas de combinación son aproximadamente la suma de diferentes energías de vibración, ya sean fundamentales o sobretonos.

Es posible que las bandas de combinación correspondan a más de dos vibraciones diferentes. Sin embargo, estas vibraciones deben corresponder al mismo grupo funcional y deben tener la misma simetría. De esta manera, bandas de combinación correspondientes a dos grupos funcionales diferentes, como C-H y O-H, no son posibles [19].

Los sobretonos y las bandas de combinación tienen menor probabilidad que las transiciones fundamentales, lo que provoca que la absorción en la región NIR sea de menor intensidad que en la MIR. La intensidad de absorción en la región NIR depende de la anarmonicidad del enlace, cuanto más anarmónico sea éste, mayor probabilidad para sobretonos y bandas de combinación. Por este motivo, los enlaces C-H, N-H, y O-H, muy anarmónicos debido a la diferencia de peso atómico entre los dos átomos, son muy absorbentes en la región NIR [19]. Las bandas de grupos C=O, C-C, C-F y C-Cl, en general son de menor intensidad o no aparecen en la región NIR.

La figura 2.7 muestra los principales enlaces absorbentes en NIR, y su correspondiente región espectral.

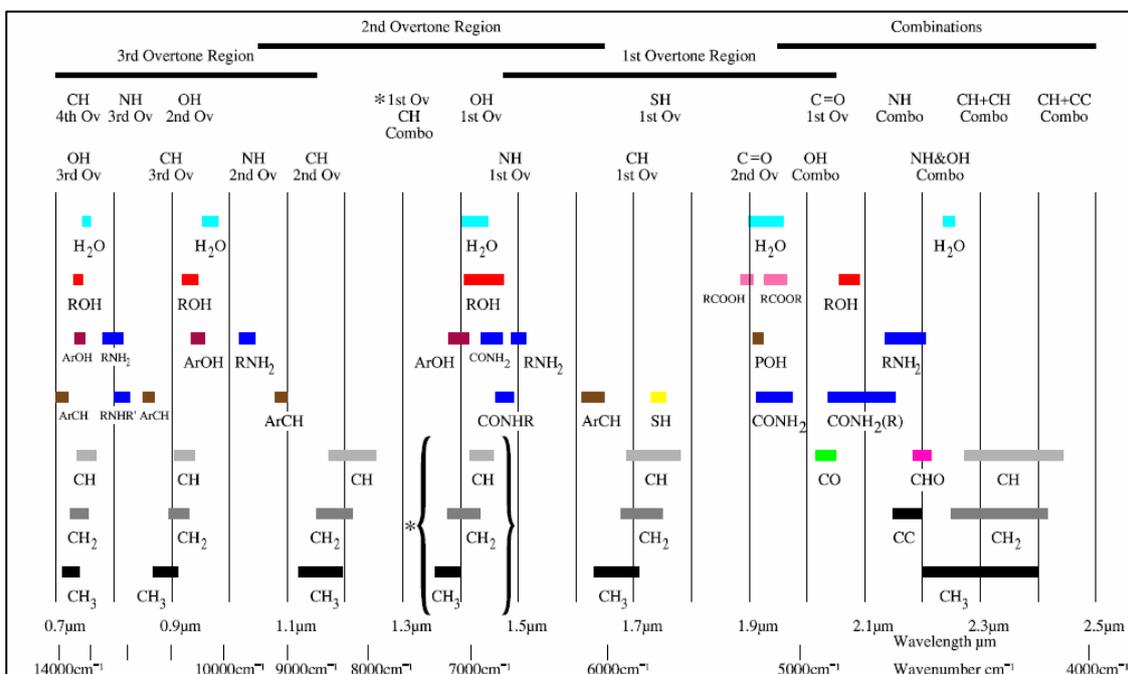


Figura 2.7. Bandas de absorción en la región NIR [1].

Los compuestos orgánicos poseen muchos modos vibracionales que provocan que el espectro NIR sea complejo. Tomando como ejemplo el cloroformo (CHCl_3), compuesto con solo un grupo funcional que absorba en la región NIR, el C-H, con una energía de vibración fundamental cuyos sobretonos aparezcan en la región NIR, y 6 modos vibracionales diferentes (tabla 2.2) [18].

Modo vibracional	V_1	V_2	V_3	V_4	V_5	V_6
Frecuencia (nm)	3,289	14,993	27,027	8,237	13,245	38,314

Tabla 2.2. Vibraciones fundamentales para la molécula de cloroformo.

Estos 6 modos vibracionales pueden producir un abundante número de sobretonos y bandas de combinación presentes en la región NIR (tabla 2.3) [19], ilustrando la complejidad del efecto de sobretonos y bandas de combinación en el espectro NIR incluso para moléculas sencillas. El solapamiento de sobretonos y

bandas de combinación genera el aspecto típico del espectro NIR, formado por bandas anchas y poco definidas.

Banda de combinación o sobretono	Posición calculada (nm)	Posición observada (nm)	Banda de combinación o sobretono	Posición calculada (nm)	Posición observada (nm)
$v_1 + v_4$	2351	2370	$2v_1 + v_4$	1404	1405
$v_1 + v_4 + v_6$	2232	2234	$2v_1 + v_4 + v_6$	1355	1357
$v_1 + v_4 + v_3$	2179	2181	$2v_1 + v_4 + v_3$	1336	1341
$4v_4$	2059	2107	$2v_1 + v_4 + v_2$	1285	1290
$v_1 + v_4 + v_2$	2047	2049	$2v_1 + v_4 + v_5$	1270	1270
$v_1 + v_4 + v_5$	2010	2008	$2v_1 + 2v_4$	1200	1210
$v_1 + 2v_4$	1841	1843	$3v_1 + v_6$	1118	-
$v_1 + 2v_4 + v_6$	1758	1757	$3v_1 + v_3$	1105	1109
$v_1 + 2v_4 + v_3$	1725	1735	$3v_1$	1097	1152
$2v_1$	1645	1692	$3v_1 + v_2$	1070	1068
$v_1 + 2v_4 + v_2$	1641	1655	$3v_1 + v_5$	1060	1059
$2v_1 + v_6$	1620	1619	$3v_1 + v_4$	1011	1018
$v_1 + 2v_4 + v_5$	1618	1618	$3v_1 + 2v_4$	903	908
$2v_1 + v_3$	1592	1590	$4v_1 + v_2$	834	830
$2v_1 + v_2$	1520	1525	$4v_1$	822	883
$v_1 + 3v_4$	1506	1517	$4v_1 + v_4$	798	800
$2v_1 + v_5$	1500	1494	$5v_1$	658	721

Tabla 2.3. Bandas de combinación y sobretonos para el cloroformo.

Otra característica de la zona del infrarrojo cercano es que las interacciones entre moléculas, como puentes de hidrógeno, afectan al espectro NIR modificando la posición de las bandas. Además, características físicas de la muestra como temperatura, tamaño de partícula o tipo de estructura cristalina también afectan en la región del infrarrojo cercano.

2.2.3 Instrumentación

El esquema básico de un instrumento NIR no difiere de cualquier otro espectrofotómetro. Sin embargo, debido a la baja intensidad de las bandas NIR, el nivel de exigencia en términos de ruido y estabilidad instrumental es mayor que en otros espectrofotómetros. Sus componentes básicos son: fuente de radiación, sistema de selección de longitudes de onda, compartimiento de muestra y detector.

➤ Fuente de radiación

La lámpara halógena de tungsteno continúa siendo la fuente de radiación más utilizada en espectroscopia NIR por su buen funcionamiento, robustez, coste y, porque proporciona un espectro continuo en la región 320 - 2500 nm. Otras fuentes de radiación son los LED (*Light Emitting Diodes*), atractivos por su eficacia, pequeño tamaño y temperatura de operación, pero que solo pueden emitir hasta los 1600 nm.

➤ Selector de longitudes de onda

A excepción de los instrumentos basados en LEDs como fuente de radiación, es necesario un sistema que descomponga el haz de luz policromática en longitudes de onda discretas, proporcionando un ancho de banda estrecho y una elevada intensidad en todo el intervalo de longitudes de onda. Los dispositivos de selección de longitudes de onda se pueden clasificar en tres tipos: filtros, redes de difracción e interferómetros.

Los instrumentos de **filtros** seleccionan las longitudes de onda interponiendo entre la fuente y la muestra materiales que permiten el paso de determinadas longitudes de onda [20]. Los más sencillos son los filtros de absorción que transmiten longitudes de onda de forma selectiva según el material con el que estén producidos,

y los de interferencia, donde las longitudes de onda transmitidas dependen del índice de refracción del material y del ángulo de incidencia de la radiación. Este último tipo de filtros

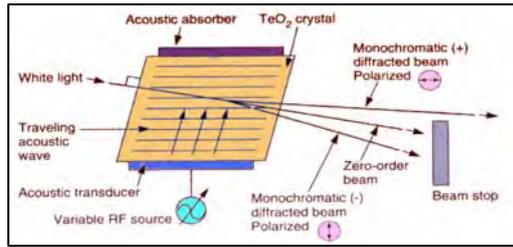


Figura 2.8. Acousto-optical tunable filter

tienen un ancho de banda de 10 ± 2 nm y transmiten aprox. un 40% de la radiación [21].

Sistemas modernos están basados en filtros optoacústicos AOTF (*Acousto-Optic Tunable Filter*), donde el índice de refracción del material (TeO_2) cambia cuando éste es atravesado por una onda acústica, generando luz monocromática polarizada. La frecuencia acústica controla la longitud de onda que atraviesa el filtro.

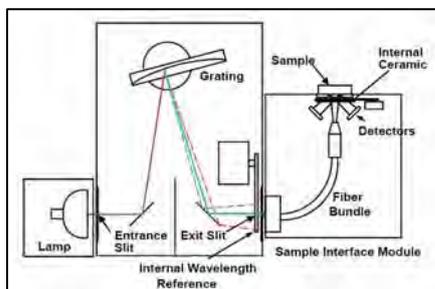


Figura 2.9. NIR con red de difracción

Las redes de difracción son dispositivos que descomponen la luz policromática en longitudes de onda monocromáticas. Cuando el haz de luz incide sobre la red de difracción, este se separa en longitudes de onda discretas.

La selección de la longitud de onda se basa en el ángulo de giro de la red de difracción, que permite que solo una longitud de onda atraviese la ranura de salida y llegue a la muestra.

Otros selectores de longitudes son los interferómetros, siendo el interferómetro de Michelson el más utilizado. Este sistema divide en dos el haz de luz, uno de los haces se refleja

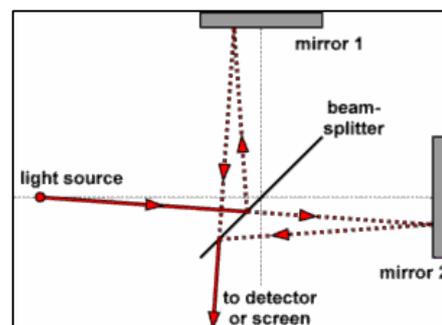


Figura 2.10. Interferómetro de Michelson

en un espejo fijo y el otro en un espejo móvil, y posteriormente, ambos haces se recombinan en el mismo punto. Variando la posición del cristal móvil se pueden crear interferencias constructivas o destructivas, obteniendo el interferograma. El espectro se obtiene del interferograma mediante transformada de Fourier.

➤ Compartimiento de muestra

El compartimiento de muestra es la parte del espectrofotómetro que más depende de la aplicación. La espectroscopia NIR tiene una gran versatilidad y adaptabilidad para el análisis de muestras de diversa naturaleza, ya sean sólidas, líquidas o gaseosas, debido a la existencia de múltiples módulos de medida adaptados a cada tipo de muestra. Así, se pueden realizar medidas *at*, *on* e *in-line* para múltiples tipos de muestra, variando el camino óptico y el sistema de medida según la necesidad, ya sea con compartimentos adaptados o mediante sondas de fibra óptica, con instrumentación portátil o fija, etc.

A pesar de la gran variedad de compartimentos de muestras, existen tres sistemas de medida (figura 2.11): transmitancia, reflectancia y transflectancia (sistema de medida donde el haz de luz atraviesa la muestra, se refleja en un reflector en contacto con la misma, y vuelve a atravesar la muestra hasta llegar de nuevo al detector).

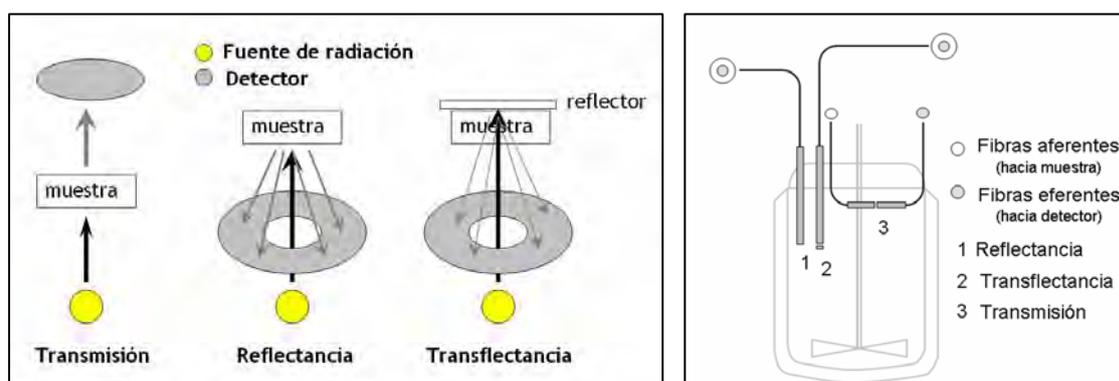
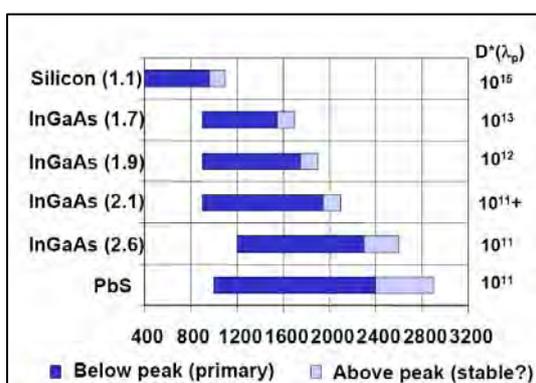


Figura 2.11. Sistemas de registro en NIR. Izq: disposición clásica. Der: sondas de fibra óptica

El modo transmitancia suele emplearse para muestras gaseosas, líquidas o semilíquidas, mientras que el modo reflectancia suele ser comúnmente empleado para sólidos. La transflectancia, caso intermedio, suele ser empleado para muestras líquidas o semilíquidas.

➤ Detectores

Los detectores empleados en espectroscopia NIR son construidos con materiales semiconductores como InGaAs, InAs, InSb, PbS o Si. El material más utilizado en la región 1100 - 2500 nm es el PbS, mientras que en la región más próxima al visible



(780 - 1100 nm) se emplea generalmente el silicio. Los detectores de InGaAs son buenas alternativas a los detectores de PbS ya que ofrecen mayor respuesta, pero con el inconveniente de que la temperatura de trabajo óptima es de -40°C [21].

Figura 2.12. Intervalos de registro para detectores en NIR

2.3 Aplicación del NIR en el control de reacciones

Las aplicaciones actuales de la espectroscopia NIR en el control de procesos abarca campos tan diferentes y diversos como el agrario, el alimentario, la industria farmacéutica, química y petroquímica, las ciencias de la tierra, la monitorización ambiental, la química clínica y médica y la biotecnología, entre otros.

2.3.1 Evolución del NIR en el control de reacciones

A partir de la década de los 80 comenzaron a ser frecuentes las publicaciones relacionadas con la aplicación del NIR al control de procesos en múltiples campos de

la química o la agricultura. Sin embargo, no fue hasta la década de los 90 cuando comenzó el auge del NIR para el control de reacciones [22]. En 1992 C. Gwosdz *et al.* [23] seleccionaron la espectroscopia NIR frente a las técnicas cromatográficas por su rapidez de análisis para la monitorización *on-line* de una reacción industrial. Poco después, Aldridge *et al.* [24-25] describen la monitorización de una reacción de polimerización mediante espectroscopia NIR. Estos artículos "precursores" de la aplicación en el control de reacciones, estaban enfocados a presentar las ventajas que aportaba el NIR y comentaban ligeramente las técnicas quimiométricas utilizadas para tratar los datos, aplicando en algunos casos calibración univariante [25]. Wise [26] comenta las aplicaciones del PLS_Toolbox para el control analítico de procesos utilizando técnicas quimiométricas como el PCA, MLR, PCR o PLS. Posteriores publicaciones muestran la aplicación del PLS, principalmente, en el control de polimerizaciones o reacciones de esterificación.

La aplicación de técnicas de resolución de curvas y de técnicas *Three-Way* al control de reacciones empezó en la década actual, con diferente evolución para ambas. Así, las técnicas basadas en la resolución multivariante de curvas (MCR, MCR-ALS, Grey Models, etc.) [27, 28-30] han demostrado una gran potencialidad para el estudio de reacciones y actualmente existen un amplio número de publicaciones. Por otro lado, la aplicación de técnicas *n-Way* [31, 32] ha sido muy tímida debido a que en la mayoría de los casos a que no se cumplen las condiciones matemáticas necesarias para su aplicación. Finalmente, las últimas aplicaciones se basan en la combinación de técnicas de resolución y modelos cinéticos empíricos para el cálculo de constantes de reacción cinéticas [33].

En la tabla 2.4 se muestran algunas de las aplicaciones de la espectroscopia NIR en el control de reacciones.

Año	Tipo de medida	Técnica quimiométrica	Aplicaciones y comentarios	Refer.
1992	on-line	-	Determinación del punto final de una reacción	23
1993	non-invasive	univariante	Monitorización de la polimerización de metil metacrilato a 870 nm	25
1997	in-line	non linear PLS	Determinación de índice de ácido e índice de hidroxilo en reacciones entre ácidos dicarboxílicos y dioles	34
1999	-	-	Método para la predicción de propiedades y rendimiento durante un proceso de cracking (patente)	35
1999	in-line	PLS	Determinación del punto final de la reacción entre D-ribosa y 2,2-dimetoxipropano	36
2000	on-line	PLS	Cuantificación de los compuestos involucrados en la esterificación de ácido acético con butanol y determinación de constantes cinéticas	37
2000	on-line	MCR	Monitorización de la transesterificación del copolímero etileno-vinilacetato	27
2001	-	-	Método para el control analítico en continuo de polímeros durante su producción o procesamiento (patente)	38
2001	in-line	PARAFAC	Estudio mediante técnicas three-way de la esterificación entre ácido acético y alcohol isoamílico	31
2002	on-line	-	Determinación de la composición en conductos de reactores para control de procesos (patente)	39
2002	on-line	PLS	Determinación del % de conversión en la polimerización del estireno	40

Tabla 2.4. Aplicación de la espectroscopia NIR para el control de reacciones

Año	Tipo de medida	Técnica quimiométrica	Aplicaciones y comentarios	Refer.
2002	in-line	PLS	Determinación del % de conversión en la copolimerización de metilmetacrilato y butilacrilato	41
2002	on-line	PLS	Determinación del peso molecular medio y del % de conversión para la polimerización del metil metacrilato	42
2002	in-line	PARAFAC	Estudio cualitativo de una reacción a partir de técnicas three-way	32
2003	in-line	OPA - MCR	Monitorización de una reacción. Obtención del nº de componentes por OPA y estudio perfiles de concentración mediante MCR	28
2003	in-line	Grey models (MCR)	Estudio cualitativo de la reacción entre diisocianato y alcohol y detección de variaciones entre lotes	29
2004	-	MCR	Estudio mediante técnicas de resolución de la reacción de curado de resinas epoxi	30
2004	on-line	-	Determinación de los componentes involucrados en un proceso químico (patente)	43
2005	in-line	PLS	Uso conjunto de NIR y ultrasonidos para la monitorización de procesos de extrusión de polietileno de alta densidad	44
2005	in-line	PLS	Monitorización de % conversión de la fotopolimerización de capas de acrilato en láminas de polímero	45
2006	in-line	MCR - Kinetic modelling	Aplicación de MCR y modelos cinéticos para el estudio de mecanismos de reacción y constantes cinéticas	33

Tabla 2.4 (continuación). Aplicación de la espectroscopia NIR para el control de reacciones

2.3.2 Ventajas del NIR en la industria

Las técnicas tradicionales de análisis químico (cromatografía, valoraciones, etc.), son ampliamente utilizadas en la industria química para el control de reacciones. Estas técnicas implican un retraso en la obtención de la información debido al tiempo necesario para realizar las diferentes etapas de la determinación tradicional, como el muestreo, el análisis y el cálculo para proporcionar un resultado. Este retraso es la causa, entre otros problemas, de que la mayoría de procesos se desarrollen bajo condiciones subóptimas.

Con la aplicación de la espectroscopia NIR se pueden desarrollar metodologías analíticas más rápidas respecto a las técnicas tradicionales de análisis, eliminando en algunos casos el uso de reactivos tóxicos y contaminantes, además de permitir el análisis multicomponente a partir de un solo espectro. Estas características aportan una serie de beneficios a la industria química como son la disminución de los productos fuera de especificaciones, la reducción del tiempo del ciclo de producción, menos horas de analistas, un menor consumo en reactivos, y además, aporta una imagen de calidad e innovación.

La evolución de la espectrofotometría NIR y, en general, de las técnicas espectrofotométricas en los últimos años, ha permitido ampliar su uso en múltiples procesos de la industria química. La evolución desde el concepto *off-line* hasta el *non-invasive* ha permitido "llevar la luz a la muestra" en vez de "llevar la muestra a la luz" [46], evitando la etapa de muestreo. La automatización de la toma de muestra o el uso de fibras ópticas sumergidas en la línea de proceso, eliminan por completo la manipulación de la muestra, obteniéndose información de forma aún más rápida. Existen diversos diseños de fibras ópticas, según el tipo de medida que sea necesaria, transmisión, reflectancia o transreflectancia.

A pesar de que el desarrollo de modelos de calibración es complejo, una vez instalada la calibración, el análisis de rutina puede ser realizado por personal no cualificado, realizando previamente un breve curso de formación. Así, en el caso de un análisis *at-line*, puede ser el mismo operario que controla el proceso en la planta de producción el que realice el análisis NIR, obtenga la información necesaria, y proceda sin demora con la siguiente etapa del proceso de producción.

En la última década, el número de publicaciones científicas relacionadas con el uso conjunto de la espectroscopia NIR y el análisis multivariable aplicado a la monitorización *in-line* o el control de procesos ha crecido paulatinamente [47], al igual que la aceptación de esta técnica como control rutinario en el mundo industrial.

La utilización de la espectroscopia NIR en el análisis y monitorización de procesos no sólo soluciona los problemas citados anteriormente, proporcionando información útil para la toma de decisiones, sino que además, es un elemento fundamental en investigación para la optimización de parámetros experimentales del proceso (temperatura, catalizador, etc.), convirtiéndose en una útil herramienta tanto a nivel de producción como de investigación [48].

2.4 Aplicaciones del NIR en la industria química

Dentro de la industria química las aplicaciones de la espectroscopia NIR no solo están cercadas al control de reacciones, si no que existen múltiples posibilidades, tal y como se muestra en la figura 2.13.

1. En una primera etapa, el NIR puede permitir el control de calidad *in-situ* de la materia prima mediante análisis cualitativo, permitiendo rechazar lotes que posteriormente puedan dar problemas en la producción. Las causas pueden

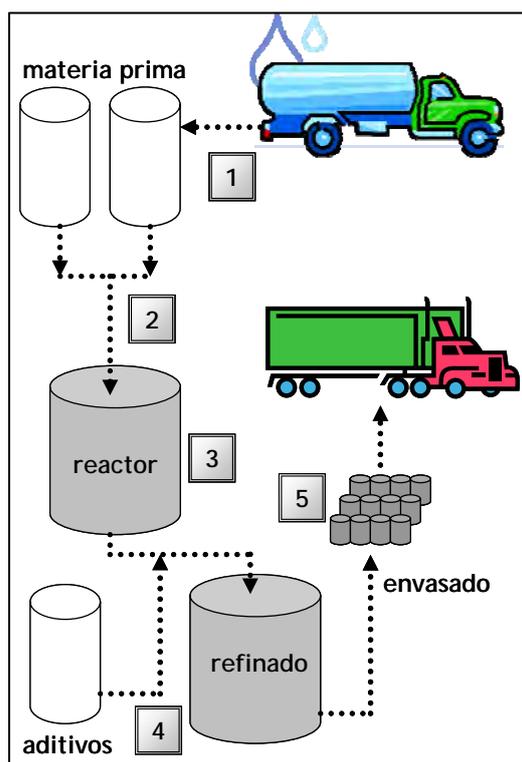


Figura 2.13. Aplicaciones del NIR en el control de un proceso químico industrial

antes de comenzar la reacción, evitando la posterior corrección del producto final, que implicaría enfriar el reactor, ajustar la carga, y volver a calentar hasta temperatura de reacción.

3. Como ya ha sido comentado extensamente, la principal aplicación es el control de la etapa de reacción.
4. Generalmente, después de la etapa de reacción, son necesarios procesos de refinado para la eliminación de excesos de reactivos y la adición de determinados aditivos. Estos procesos también pueden ser controlados mediante espectroscopia NIR.
5. Finalmente, el producto final puede ser controlado cualitativamente una vez ha sido envasado, de forma similar a como se controlaría la entrada de materia prima (punto 1).

ser varias, como por ejemplo, posibles contaminaciones debidas a una mala limpieza de la cisterna, condiciones inadecuadas durante el transporte, etc.

2. Debido a la complejidad logística que implica a gran escala la carga de productos sólidos o de elevada viscosidad, es posible que la carga del reactor no se haya realizado de forma correcta. Así, es posible controlar si la mezcla del reactor tiene la composición adecuada

Durante la realización de esta tesis, se ha aplicado la espectroscopia NIR para el control de la etapa de reacción, siendo también estudiadas las etapas 2 y 4 de la figura 2.13.

Referencias

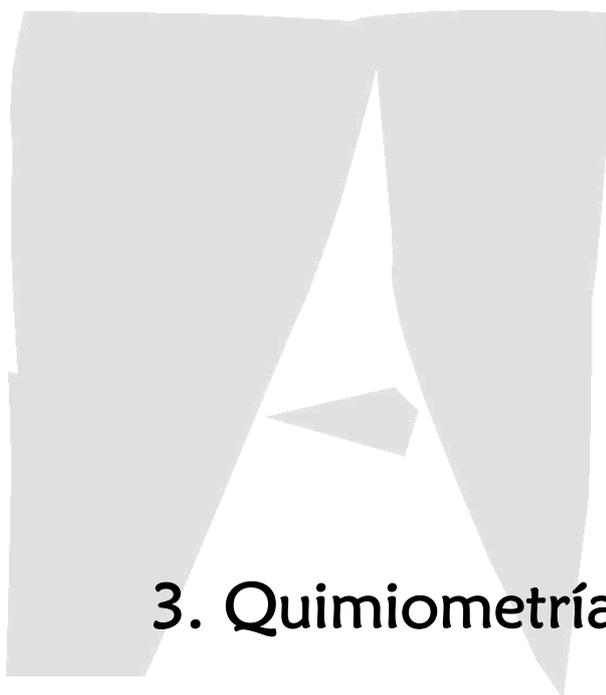
- [1] I. Murray. Scattered information: philosophy and practice of near infrared spectroscopy. Proceedings of the 11th International Conference on Near Infrared spectroscopy. *NIR Publications* (2004) 2.
- [2] W. Herschel. Investigation of the powders of the prismatic colours to heat and illuminate objects; with remarks, that prove the different refrangibility of radiant heat. To which is added, an inquiry into the method of viewing the sun advantageously, with telescopes of large apertures and high magnifying powers. *Philosophical transactions of the Royal Society* **90** (1800) 255-283.
- [3] W. Abney and E. R. Festing. On the influence of the atomic grouping in the molecules of organic bodies on their absorption in the infra-red region of the spectrum. *Philosophical transactions of the Royal Society* **172** (1881) 887-918.
- [4] W. Kaye. Near-infrared spectroscopy: I. Spectral identification and analytical applications. *Spectrochimica Acta* **6** (1954) 257.
- [5] J. A. Mitchel, C. D. Bockman and A. V. Lee Jr. Determination of acetyl content of cellulose acetate by near infrared spectroscopy. *Anal. Chem.* **29** (1957) 499-502.
- [6] G. S. Birth and K. H. Norris. An instrument using light transmittance for non-destructive measurement of fruit maturity. *Food technology* **12** (1958) 592.
- [7] K. H. Norris. *Trans. Am. Soc. Agrico. Eng.* **7** (1964) 240.
- [8] I. Ben-Gera, K. H. Norris. Direct spectrophotometric determination of fat and moisture in meta products. *Journal of food sciences* **7** (1968) 240.
- [9] F. E. Barton. Progress in near infrared spectroscopy the people, the instrumentation, the applications. *Proceedings of the 11th International Conference on Near Infrared spectroscopy*. NIR Publications (2004) 13.

- [10] D. Noble. Illuminating near-IR. *Analytical Chemistry* **67** 23 (1995) 735A-740A.
- [11] C. M. Henry. Near-IR gets the job done. *Analytical Chemistry* **71** 17 (1999) 625A-628A.
- [12] E. W. Ciurczak. *Principles of near-infrared spectroscopy. Handbook of near infrared analysis*. Marcel Dekker, 2nd edition, New Cork, pp. 7-9.
- [13] H. Mark. Chemometrics in near-infrared spectroscopy. *Anal. Chimica Acta* **223** (1989) 75-93.
- [14] K. R. Beebe and B. R. Kowalski. An introduction to multivariate calibration and analysis. *Analytical Chemistry* **59** 17 (1987) 1007A-1017A.
- [15] D. L. Wetzel. Near infrared analysis. *Anal. Chem.* **55** (1983) 1165A-1176A.
- [16] W. McClure. Near infrared spectroscopy. The giant is running strong. *Anal. Chem.* **66** (1994) 43A-53A.
- [17] T. Davies. The history of near infrared spectroscopic analysis: past, present and future from sleeping technique to the morning star of spectroscopy. *Analysis* **26** 4 (1998) 17-19.
- [18] J. Workman Jr., M. Koch and D. J. Veltkamp. Process analytical chemistry. *Anal. Chem.* **75** (2003) 2859-2876.
- [19] C. E. Miller. *Near-infrared technology in the agricultural and food industries*. P. Williams and K. Norris (editors), American association of cereal chemist, 2nd edition, Minnesota, chapter two.
- [20] D. A. Burns, E. W. Ciurczak. *Handbook of near-infrared analysis*. Marcel Dekker 2001.
- [21] E. Stark and K. Luchter. Diversity in NIR instrumentation. Proceedings of the 11th International Conference on Near Infrared spectroscopy. *NIR Publications* (2004) 13.
- [22] W. B. Wayne et al. Process analytical chemistry. *Anal. Chem.* **67** (1995) 47R-70R.

- [23] C. Gwosdz, M. S. Francis, A. Bickel. Near-infrared on-line monitoring of batch reaction completion. *Process control and quality* 4 (1992) 31-35.
- [24] P. Aldridge, J. J. Kelly, J. B. Callis and D. H. Burns. Noninvasive monitoring of bulk polymerization using short-wavelength near-infrared spectroscopy. *Anal. Chem.* 65 24 (1993) 3581.
- [25] P. K. Aldridge, D. H. Burns, J. J. Kelly, J. B. Callis. Monitoring of methyl methacrylate polymerization using non-invasive SW-NIR spectroscopy. *Process control and quality* 4 (1993) 155-160.
- [26] B. M. Wise. Using the PLS_Toolbox in PAC applications. *Process control and quality* 5 (1993) 73-85.
- [27] S. Sasic, Y. Kita, T. Furukawa, M. Watari, H. W. Siesler and Y. Ozaki. Monitoring the melt-extrusion transesterification of ethylene-vinylacetate copolymer by self-modeling curve resolution analysis of on-line near-infrared spectra. *Analyst* 125 (2000) 2315-2321.
- [28] S. Gourvéneq, C. Lamotte, P. Pestiaux and D. L. Massart. Use of orthogonal projection approach (OPA) to monitor batch processes. *Appl. Spectrosc.* 57 1 (2003) 80-87.
- [29] E. N. M. Van Sprang, H. Ramaker, J. A. Westerhuis, A. K. Smilde, S. P. Gurden and D. Wienke. Near-infrared spectroscopic monitoring of a series of industrial batch processes using a bilinear grey model. *Appl. Spectrosc.* 57 8 (2003) 1007-1019.
- [30] M. S. Larrechi and F. X. Rius. Spectra and concentration profiles throughout the reaction of curing epoxy resins from near-infrared spectroscopy and multivariate curve resolution. *Appl. Spectrosc.* 58 1 (2004) 47-53.
- [31] P. Geladi and P. Aberg. Three-way modelling of a batch organic synthesis process monitored by near infrared spectroscopy. *J. Near Infrared Spectrosc.* 9 (2001) 1-9.

- [32] P. Geladi and J. Forsström. Monitoring of a batch organic synthesis by near-infrared spectroscopy: modelling and interpretation of three-way data. *J. Chemometrics* **16** (2002) 329-338.
- [33] M. Garrido, M. S. Larrechi and F. X. Rius. Multivariate curve resolution-alternating least squares and kinetic modeling applied to near-infrared data from curing reactions of epoxy resins: mechanistic approach and estimation of kinetics rate constants. *Appl. Spectrosc.* **60** 2 (2005) 174-181.
- [34] R. A. Heikka, K. T. Immonen, P. O. Minkkinen, E. Y. O. Paatero and T. O. Salmi. Determination of acid value, hydroxyl value and water content in reactions between dicarboxylic acids and diols using near-infrared spectroscopy and non-linear partial least squares. *Anal. Chim. Acta.* **349** (1997) 287-294.
- [35] B. Descales, D. Lambert, J. R. Llinas, A. Martens and C. Granzotto. U.S. Patent 5.935.863 (1999).
- [36] C. Coffey, B. E. Cooley Jr and D. S. Walker. Real time quantitation of a chemical reaction by fiber optic near-infrared spectroscopy. *Anal. Chim. Acta.* **395** (1999) 335-341.
- [37] M. Blanco, D. Serrano. On-line monitoring and quantification of a process reaction by near-infrared spectroscopy. Catalysed esterification of butan-1-ol by acetic acid. *Analyst* **125** (2000) 2059-2064.
- [38] U. Thiele. DE Patent 19.934.349 (2001).
- [39] U. Wolf, F. K. Bruder, J. Diefendahl, A. Elschner, U. Hucks, D. Riesebeck, B. Schnittka, M. Schraut and L. Spauschus. WO Patent 2002004927 (2002).
- [40] H. H. A. Lousberg, H. F. M. Boelens, E. P. Le Comte, H. C. J. Hoefsloot and A. K. Smilde. On-line determination of the conversion in a styrene bulk polymerization

- batch reactor using near-infrared spectroscopy. *Journal of Applied Polymer Science* **84** (2002) 90-98.
- [41] R. A. M. Vieira, C. Sayer, E. L. Lima and J. C. Pinto. In-line and in situ monitoring of semi-batch emulsion copolymerizations using near-infrared spectroscopy. *Journal of Applied Polymer Science* **84** (2002) 2670-2682.
- [42] A. Cherfi, G. Fevotte and C. Novat. Robust on-line measurements of conversion and molecular weight using NIR spectroscopy during solution polymerization. *Journal of Applied Polymer Science* **85** (2002) 2510-2520.
- [43] L. Yicheng, D. Tadeusz and H. Hassan. U.S. patent 2004084623 (2004).
- [44] S. E. Barnes, E. C. Brown, M. G. Sibley, H. G. M. Edwards, I. J. Scowen and P. D. Coates. Vibrational spectroscopy and ultrasound analysis for in-process characterization of high-density polyethylene/polypropylene blends during melt extrusion. *Appl. Spectrosc.* **59** 5 (2005) 611-619.
- [45] T. Scherzer, S. Müller, R. Mehnert, A. Volland and H. Lucht. In-line monitoring of the conversion in photopolymerized acrylate coatings on polymer foils using NIR spectroscopy. *Polymer* **46** (2005) 7072-7081.
- [46] B. D. Mindel. Process interfacing of on-line NIR spectrophotometers. *Process Control and Quality*, **9** 4 (1997) 173-178.
- [47] Van Staden J. F. "Analytical aspects of chemical process control. Part 1. Fundamentals". *Pure Appl. Chem.*, **71** 12 (1999) 2303-2308.
- [48] P. J. Brimmer, F. A. DeThomas, J. W. Hall. Near infrared technology in the agricultural and food industries. 2001, Ed. American association of cereal chemist.



3. Quimiometría

3. Quimiometría

3.1 Introducción

El aumento progresivo en complejidad de la instrumentación analítica permite obtener cada vez un mayor volumen de datos. La conversión de estos datos en información útil requiere el uso de herramientas matemáticas y estadísticas, que han dado lugar a una disciplina denominada Quimiometría [1-2].

La Quimiometría fue definida en 1975 [3] como el conjunto de métodos utilizados para extraer información química útil de un conjunto de datos originales. Massart *et al.* [1] dan una definición más completa y precisa de este término: “La Quimiometría es la parte de la química que se sirve de las matemáticas, estadística y lógica formal para diseñar o seleccionar procedimientos experimentales óptimos, proporcionar información química relevante a partir del análisis de señales analíticas y, finalmente, adquirir conocimiento de los sistemas químicos”.

La complejidad de la señal NIR hace imprescindible la utilización de técnicas quimiométricas que nos permitan interpretar, entender y modelar grandes conjuntos de datos. Este apartado presenta la base teórica de las herramientas quimiométricas utilizadas en la presente tesis. En un primer lugar, se introduce el proceso de desarrollo de modelos de calibración cuantitativos. Posteriormente, se comentan los pretratamientos espectrales y de técnicas de modelado utilizadas. Finalmente, se describe el cálculo de límites de detección para la evaluación de modelos de calibración PLS.

3.2 Etapas del desarrollo de modelos de calibración multivariantes

Un modelo de calibración multivariable relaciona las N variables experimentales (por ejemplo, un espectro NIR) frente a propiedades conocidas de las muestras (por ejemplo, la concentración de un analito), siendo el objetivo la predicción de dicha propiedad en nuevas muestras. Para la construcción de un modelo de calibración es conveniente seguir las siguientes etapas:

3.2.1 Selección del conjunto de muestras

El primer paso es disponer de un número de muestras suficiente para la creación y validación del modelo. Las muestras utilizadas para el desarrollo de un modelo de calibración deben ser representativas de la variabilidad que pueda darse durante el proceso, y también entre procesos. Así, las muestras, además de cubrir todo el intervalo de la propiedad a determinar, han de incorporar la variabilidad debida a otras variables del proceso, como puede ser el origen de la materia prima o la temperatura de reacción, entre otros. Una herramienta útil para la selección de muestras es el análisis en componentes principales o PCA (sección 3.4.1).

3.2.2 Determinación del parámetro a modelar mediante el método de referencia

Mediante metodologías analíticas adecuadas se determinan las concentraciones o propiedades a modelar en las muestras. Los métodos empleados deben proporcionar resultados precisos y exactos, ya que de éstos dependerá la bondad del modelo de calibración obtenido.

3.2.3 Obtención de la señal analítica

Registro de la información analítica primaria que se va a utilizar en la construcción del modelo de calibración, siendo en nuestro caso el espectro NIR.

3.2.4 Pretratamiento de los datos

En los datos espectroscópicos pueden aparecer contribuciones no deseadas (debido al proceso de registro, a la naturaleza de la muestra o al ruido instrumental), que causan no linealidades u otros efectos que pueden afectar negativamente al desarrollo del modelo de calibración. Los pretratamientos espectrales tienen como objetivo, entre otros, minimizar las contribuciones espectrales no deseadas, simplificando los modelos y mejorando los resultados.

La selección del pretratamiento espectral adecuado suele realizarse *a posteriori* teniendo en cuenta un estudio estadístico de los resultados obtenidos en cada caso. Hacerlo *a priori* puede resultar erróneo ya que su selección no solo depende del tipo de muestra o del modo de registro espectral, sino también de contribuciones desconocidas.

Los pretratamientos espectrales utilizados en la presente memoria se comentan en la sección 3.3.

3.2.5 Desarrollo y evaluación del modelo de calibración

Una vez seleccionada la herramienta quimiométrica idónea, se construye el modelo de calibración evaluando distintos parámetros implicados en el cálculo: intervalo espectral, pretratamiento espectral, muestras anómalas, número de componentes o factores, errores de calibración, entre otros. Una vez obtenido el modelo, su capacidad predictiva es evaluada mediante un segundo grupo de muestras, de las que

se conoce la propiedad a determinar, pero que no han sido utilizadas en la etapa de construcción del modelo. Si los resultados no son satisfactorios, el modelo ha de ser recalculado.

Una vez construido y validado, el modelo de calibración puede ser aplicado para la determinación de la propiedad en nuevas muestras.

3.3 Pretratamientos espectrales

Los pretratamientos espectrales utilizados en esta tesis se han dividido en dos grupos, según su dependencia o no del grupo de muestras. Así, se entiende como pretratamientos independientes del grupo de muestras aquellos que tratan cada espectro de forma individual sin tener en cuenta el resto de muestras o datos de concentración; y como dependientes aquellos que tratan los espectros en conjunto o que requieren datos de concentraciones de referencia, por lo que el pretratamiento varía según el conjunto de datos. A pesar de que los pretratamientos más utilizados han sido los independientes del grupo de muestras, cabe destacar la importancia de los métodos dependientes para el estudio de los datos espectrales [4].

3.3.1 Pretratamientos independientes del grupo de muestras

3.3.1.1 Suavizado espectral

Los métodos de suavizado son utilizados para reducir matemáticamente el ruido aleatorio e incrementar la relación señal/ruido [2]. Estos métodos utilizan una ventana de puntos para determinar un valor central, desplazando la ventana a lo largo de todo el espectro. Los métodos más utilizados son el de Savitzky-Golay [5] y el de Transformada de Fourier [6].

3.3.1.2 *Standard Normal Variate* (SNV)

Con este pretratamiento se centra y autoescala cada espectro, restando el valor medio de la absorbancia del espectro a los valores de absorbancia de cada longitud de onda y dividiendo por la desviación estándar:

$$x'_n = \frac{x_n - \bar{x}}{s} \quad [3.1]$$

Se utiliza principalmente para corregir variaciones sistemáticas de línea base producidas por la dispersión [7], por lo que puede ser útil para sólidos o muestras líquidas con turbidez.

3.3.1.3 Derivadas

Es uno de los pretratamientos más utilizados en espectroscopia NIR por su capacidad de solventar los problemas característicos de esta técnica: el solapamiento de bandas y los desplazamientos de línea base. La primera derivada elimina desplazamientos de línea base constantes, la segunda derivada elimina desplazamientos que varían linealmente con la longitud de onda. Los métodos de derivación más utilizados son el de Savitzky-Golay [5] y el de Norris [8]. Las derivadas por Savitzky-Golay incorporan un suavizado previo, por lo que el tamaño de ventana es un parámetro crítico, si es pequeño la derivada tendrá un elevado ruido y si es grande la información espectral resultará demasiado suavizada [2].

3.3.2 Pretratamientos dependientes del grupo de muestras

3.3.2.1 Promedio de espectros

El ruido es considerado una variación instrumental aleatoria, por lo que promediando varias señales analíticas de la misma muestra, el ruido disminuye y se mejora la

relación señal/ruido.

3.3.2.2 Centrado

Consiste en centrar cada una de las N variables restando la media de todos los elementos de esa variable. Los vectores centrados tienen valor medio igual a cero, sin modificar la varianza. El centrado es un pretratamiento habitual previo al desarrollo de un modelo de calibración o de un análisis en componentes principales [2].

3.3.2.3 Autoescalado

Consiste en un centrado seguido de una normalización. Para una matriz X ($M \times N$) cada elemento de una fila se autoescala como:

$$x'_i = \frac{x_i - \bar{x}}{s} \quad [3.2]$$

donde x'_i es el valor autoescalado, \bar{x} y s son la media y la desviación estándar de los elementos de esa fila. Los datos autoescalados tienen media 0 y varianza 1.

3.3.2.4 *Multiplicative Scatter Correction* (MSC)

Este método se utiliza para compensar efectos multiplicativos o aditivos debidos al *scattering*. El MSC corrige los desplazamientos entre muestras aplicando a los espectros originales los parámetros de la recta, obtenida por mínimos cuadrados, representando los valores de absorbancia de un intervalo de longitudes de onda en el que no absorba el analito (aunque a la práctica se realiza con el espectro completo), frente a los valores de absorbancia de un espectro de referencia (generalmente el espectro medio de los datos a tratar). La corrección del espectro original se reduce

a:

$$x_i^{MSC} = \frac{x_i - a}{b} \quad [3.3]$$

donde a y b son los parámetros de la recta.

3.3.2.5 *Orthogonal Signal Correction* (OSC)

El OSC corrige los espectros originales teniendo en cuenta las concentraciones del analito, por lo que este pretratamiento es específico para el grupo de muestras seleccionado y también para el analito. Mediante OSC se elimina la información contenida en la matriz de datos espectrales (X) que es ortogonal a la matriz de concentraciones (Y).

El OSC, introducido por Wold et al. [9] utiliza un algoritmo similar al empleado en el PLS (sección 3.4.2), la diferencia es que en OSC se minimiza la matriz de covarianza entre X e Y . Posteriormente, han aparecido diferentes estrategias para mejorar el método original [10-11]. El OSC se puede definir como:

$$X_{OSC} = X - \sum_{i=1}^A T_i P_i^t \quad [3.4]$$

donde la matriz de datos OSC (X_{OSC}) corresponde a la matriz de datos original (X), a la cual se le sustrae el producto de matrices de *scores* y *loadings* (con A factores), que contienen la información no correlacionada con Y . Cuanto mayor sea el número de factores, mayor será la varianza eliminada. El sobreajuste es frecuente en OSC, normalmente son necesarios pocos factores para obtener resultados satisfactorios.

El OSC puede ser un pretratamiento útil para la determinación de analitos en baja concentración ya que permite eliminar la información no correlacionada con la concentración del analito, simplificando los modelos de calibración [12]

3.3.2.6 Señal Neta del Analito (NAS, *Net Analyte Signal*)

Esta técnica quimiométrica ha sido utilizada en el presente trabajo como pretratamiento espectral previo al desarrollo de un modelo de calibración. Sin embargo, generalmente es empleada como técnica de calibración [13-16], destacando la simplicidad de cálculo de los parámetros estadísticos del modelo como sensibilidad, selectividad y límite de detección.

El NAS fue definido por Lorber [17] como la parte del espectro de la muestra relacionada con el analito y que es ortogonal a los interferentes. El objetivo del NAS es dividir la matriz de datos \mathbf{X} en dos, una con la información espectral relacionada con el analito k , \mathbf{X}_k , y el resto de fuentes de variabilidad en la matriz \mathbf{X}_{-k} , donde estaría la contribución de los interferentes:

$$\mathbf{X} = \mathbf{X}_k + \mathbf{X}_{-k} \quad [3.5]$$

El vector NAS para una muestra se calcula proyectando su espectro (\mathbf{r}) en el espacio generado por los interferentes (matriz \mathbf{X}_{-k}), siendo el NAS la resultante ortogonal (\mathbf{r}^*), definida como (figura 3.1):

$$\mathbf{r}^* = [\mathbf{I} - \mathbf{X}_{-k}^T (\mathbf{X}_{-k}^T)^+] \mathbf{r} \quad [3.6]$$

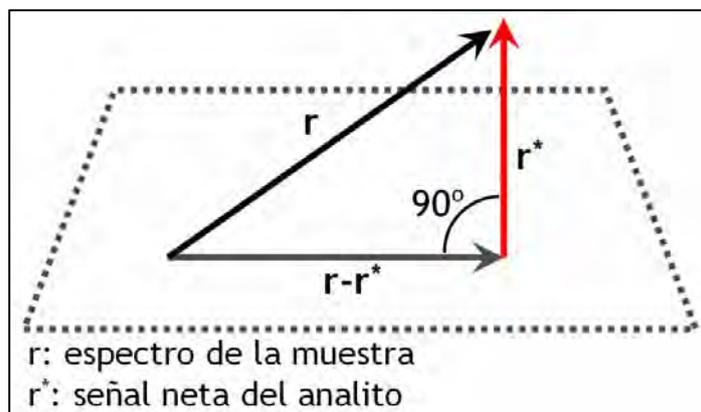


Figura 3.1. Representación geométrica del vector NAS. El vector NAS es ortogonal al espacio delimitado por los componentes de la matriz de la muestra.

Según la ecuación 3.6, para obtener la señal neta del analito es necesario conocer previamente la matriz X_{-k} , la cual se puede estimar mediante [18]:

$$X_{-k} = X - \alpha \cdot c_k \cdot r_k^t \quad [3.7]$$

donde α y c_k se definen como:

$$\alpha = \frac{1}{r_k^t \cdot X^+ \cdot c_k} \quad [3.8]$$

$$c_k = X \cdot X^+ \cdot c \quad [3.9]$$

siendo c el vector de concentraciones de las muestras de X . El vector NAS para una muestra se calcula como:

$$r_k^* = (I - X_{-k}^t (X_{-k}^t)^+) r_k \quad [3.10]$$

Una vez estimado los vectores NAS para el conjunto de muestras, posteriormente, se realiza una calibración multivariable junto con los valores de concentración. De esta manera, el NAS es aplicado como pretratamiento espectral,

obteniendo una matriz de datos simplificada que contiene únicamente la señal relativa al analito, simplificando los modelos de calibración de forma significativa.

3.4 Técnicas de análisis multivariantes

En este apartado se presenta brevemente la base teórica de las principales herramientas quimiométricas utilizadas en esta memoria para la extracción de información de los datos espectroscópicos, ya sea de forma cualitativa (PCA) o cuantitativa (métodos de calibración). Descripciones más detalladas se pueden encontrar en libros de texto donde se recopilan las técnicas quimiométricas más utilizadas [1-2, 19-24].

3.4.1 Análisis en Componentes Principales (PCA)

Debido al gran volumen de información que puede ser obtenido por las actuales técnicas de análisis químico, se han desarrollado métodos matemáticos que permiten reducir el gran volumen de datos en un número reducido de variables, para mejorar la interpretación de los datos sin pérdidas relevantes de información. El PCA, o *Principal Component Analysis* [25-27], es una herramienta que permite condensar la variabilidad presente en una matriz de datos en unos pocos parámetros representativos, denominados componentes principales (PCs) o variables latentes. El PCA es uno de los métodos más empleados para reducción de variables, además de constituir la base a partir de la cual se han desarrollado multitud de técnicas quimiométricas de análisis cualitativo y cuantitativo.

Una matriz de datos X ($M \times N$) formada por M espectros o muestras registradas a N longitudes de onda puede ser representada en un espacio de dimensión N donde cada muestra estaría representada por un solo punto. En esta situación, las características comunes entre las muestras o sus diferencias serán observables en

función de su agrupación. Debido a que no es posible representar dimensiones superiores a 3, el PCA tiene como objetivo hallar las direcciones de máxima variabilidad en que están agrupados los M puntos en el espacio de dimensión N , utilizando unos nuevos ejes llamados componentes principales para representarlos. El primer componente principal es la combinación lineal de las N variables que explica la máxima variabilidad de las muestras; el segundo se escoge de forma que sea ortogonal al primero y que explique la máxima variabilidad una vez restada la del primero, y así sucesivamente. Para definir matemáticamente el cambio de base se definen los *loadings* (P), que son los cósenos de los ángulos que forman los nuevos ejes con los originales. Los *scores* (T) son las coordenadas de las muestras en los nuevos ejes.

Matemáticamente, la matriz de datos espectrales X se descompone en el producto de las matrices de *scores* y *loadings*, más una matriz E de residuales (ecuación 3.11 y figura 3.2):

$$X = TP' + E \quad [3.11]$$

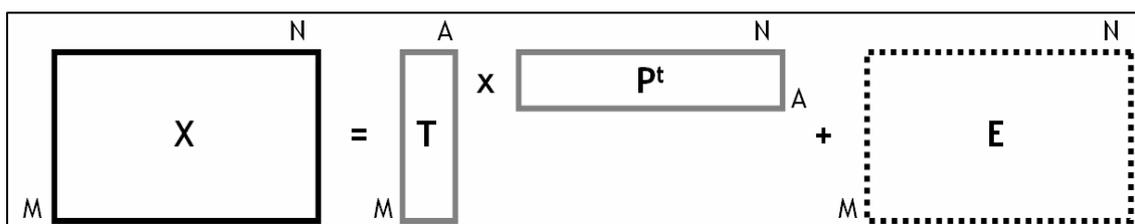


Figura 3.2. Representación gráfica de la ecuación 3.11.

Debido a la condición de ortogonalidad, los A componentes principales contienen diferente información. Los primeros describen las fuentes de variabilidad más relevantes, mientras que los últimos describen variaciones debidas al ruido, por lo que es posible reproducir la matriz de datos original con un pequeño número de componentes principales $A \ll N$, típico en datos espectroscópicos, estando limitado el número de PCs por el número de muestras, $A < M$ (si $M < N$). La matriz de datos E

contiene la información no recogida por los A componentes principales. Existen diferentes algoritmos para calcular las matrices T y P , siendo el NIPALS (*Nonlinear Iterative Partial Least Squares*) [28] uno de los más utilizados.

Previo al análisis en componentes principales es necesario un pretratamiento de los datos, siendo los más frecuentes el centrado, que se realiza siempre para que los PCs describan adecuadamente la orientación de las muestras (figura 3.3), y el autoescalado, que puede ser adecuado cuando se desea que todas las variables tengan la misma importancia independientemente de la escala de los datos. También pueden ser necesarios otros pretratamientos comentados en el apartado 3.3.

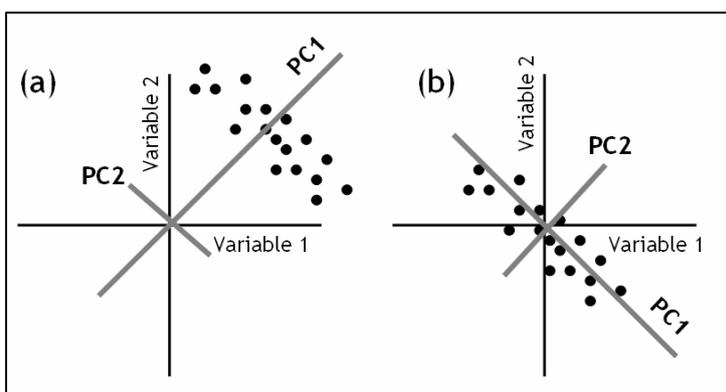


Figura 3.3. Ejemplo de PCA (a) datos sin centrar (b) datos centrados

Uno de los aspectos fundamentales del PCA es la selección del número de PCs que contienen la información relevante de la matriz de datos. La forma más habitual de realizarlo es mediante un gráfico de la varianza explicada en función del número de PCs, escogiendo en menor número de PCs para el cual no se encuentra un incremento significativo.

Una vez establecido el número de PCs, el estudio de los gráficos de *scores* y *loadings* aporta información sobre el sistema. Los *scores* son útiles para ver variaciones en los datos en un menor número de dimensiones. El gráfico de dispersión de PC's (generalmente PC1 vs. PC2 cuando ambos explican un porcentaje elevado de

la varianza explicada) revela la relación existente entre las muestras, mostrando la existencia de *clústers*, *outliers*, etc. Los *loadings* son útiles para determinar cuales son las variables originales más importantes. Definido el *loading* como el coseno del ángulo entre el PC_a y la variable *n*, cuando su valor sea próximo a $\|1\|$, indica que esa variable contribuye de forma importante a la varianza explicada por ese PC. Al contrario, cuando sea próximo a 0, ambos serán ortogonales indicando que la variable contribuye poco a la varianza explicada por ese PC. El estudio visual de los *loadings* es útil para identificar las regiones espectrales, y en consecuencia los compuestos químicos, con mayor importancia en los datos.

Además del estudio de datos, el análisis en componentes principales puede ser utilizado para filtrar el ruido en matrices de datos espectroscópicos, reconstruyendo la matriz con los *A* PCs que recogen las principales fuentes de variación, excluyendo los PCs poco significativos, que típicamente tienen una relación señal/ruido baja.

3.4.2 Regresión por Mínimos Cuadrados Parciales (PLS)

3.4.2.1 Introducción

La información espectroscópica puede correlacionarse con magnitudes físicas o químicas, obteniéndose ecuaciones o modelos de calibración. El fundamento de cualquier calibración es establecer una relación entre la señal instrumental y la propiedad a determinar, y posteriormente predecir esta propiedad a partir de la señal medida en una muestra desconocida. Es posible utilizar calibraciones univariadas cuando la señal es selectiva a la propiedad estudiada. Sin embargo, en espectroscopia NIR no siempre es posible y hay que recurrir a calibraciones multivariadas.

La regresión por mínimos cuadrados parciales (PLS) es una técnica de calibración multivariable inversa, lo que hace posible calibrar únicamente el componente deseado sin necesidad de conocer el resto de fuentes de variación. El PLS se basa en que la información contenida en las variables medidas se puede concentrar en un número menor de variables sin pérdida de información relevante, de forma similar al PCA, haciendo máxima la matriz de varianza-covarianza. La regresión se realiza sobre las nuevas variables, eliminando así los problemas de colinealidad que puedan tener los datos originales.

3.4.2.2 Fundamentos teóricos

El PLS fue desarrollado por H. Wold y colaboradores [29] entre 1975 y 1982, siendo desde entonces ampliamente contrastado, estudiado y aplicado [30-33]. En la actualidad, el PLS es probablemente el método de regresión por reducción de variables más ampliamente utilizado [20].

Durante la etapa de calibración el PLS utiliza tanto la matriz de datos espectroscópicos (X) como la información contenida en la matriz de la propiedad a determinar (Y). Ambas matrices, una vez centradas como en el caso del PCA, se descomponen en una suma de factores de forma simultánea:

$$X = TP^t + E = \sum_{a=1}^A t_a p_a^t + E \quad [3.12]$$

$$Y = UQ^t + F = \sum_{a=1}^A u_a q_a^t + F \quad [3.13]$$

donde T es la matriz de *scores*, P la de *loadings* y E la de residuales correspondientes a la matriz de datos espectroscópicos (X); U es la matriz de *scores*, Q la de *loadings* y F la de residuales para la matriz de la propiedad a determinar (Y). Si tenemos M

muestras, A componentes, M variables y K analitos, la dimensionalidad de las matrices es: T y U ($M \times A$), P^t ($A \times M$) y Q^t ($A \times K$).

La descomposición de ambas matrices se realiza de forma simultánea, estableciendo una relación interna entre los scores de los bloques X e Y , de forma que para cada componente se cumpla:

$$\hat{a}_a = b_a t_a \quad [3.14]$$

donde b_a es el coeficiente de regresión entre scores que los bloques X e Y para cada uno de los componentes. De esta manera, los nuevos ejes no coinciden con la dirección de máxima variabilidad espectral de las muestras como en el PCA, ya que están corregidos para obtener la máxima capacidad predictiva para la matriz Y . Como en el caso del PCA, el algoritmo NIPALS, es uno de los más empleados para el cálculo, en este caso adaptado al PLS.

Una vez establecido el modelo de calibración, se puede realizar la predicción de la propiedad modelada para nuevas muestras. A partir de los espectros originales, centrados o autoescalados, utilizando los valores calculados para las muestras de calibración, se calculan los *scores* de las muestras a predecir (\hat{t}), y mediante:

$$\hat{Y} = \sum_{a=1}^A b_a t_a^* q_a^t \quad [3.15]$$

Se obtiene el valor de la propiedad a determinar.

3.4.2.3 Construcción de un modelo PLS y evaluación de resultados

Durante la construcción de modelos de calibración basados en la reducción de variables es importante examinar, entre otros, los siguientes parámetros:

➤ Selección del número de factores

La selección del número de factores que configura el modelo es un punto clave en la utilización de técnicas de reducción de variables, ya que el modelo ha de incorporar la información relevante pero ha de excluir el ruido. La forma más común de seleccionar el número de factores adecuado es utilizar algún parámetro estadístico como el PRESS (ecuación 3.16) o la varianza explicada.

$$PRESS = \sum_{i=1}^N (\hat{y}_i - y_i) \quad [3.16]$$

Cuando el número de muestras es pequeño, como es el caso de los ejemplos mostrados en esta tesis, la metodología más utilizada para construir el modelo es la validación cruzada (*cross validation*) [34]. Mediante este método el conjunto de muestras se divide en segmentos, denominándose *leave-one-out* cuando cada segmento está formado por una sola muestra. Este procedimiento deja cada vez una muestra fuera y utiliza el resto de muestras para calcular el modelo, repitiendo el proceso hasta que todas las muestras han sido excluidas, y predicha su propiedad a determinar. Mediante *cross validation* es posible estimar el MSECv (*Mean Square Error of Cross Validation*):

$$MSECv = \frac{\sum_{i=1}^M (\hat{y}_i - y_i)^2}{M} \quad [3.17]$$

Así, se puede seleccionar el número de factores representando el MSECv frente al número de factores y seleccionando el mínimo o el primer mínimo local [20].

Otro criterio para establecer el número de factores es seleccionar aquel que proporcione un menor error en predicción. El RSE (*Relative Standard Error*) y el RMSE

(*Root Mean Square Error*), han sido utilizados para evaluar los errores de calibración y predicción, definidos como RSEC y RMSEC para calibración, y RSEP y RMSEP para predicción, seleccionando en número de factores en función del error de predicción.

$$RSE = 100 \cdot \sqrt{\frac{\sum_{i=1}^M (\hat{y}_i - y_i)^2}{\sum_{i=1}^M y_i^2}} \quad [3.18]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^M (\hat{y}_i - y_i)^2}{M}} \quad [3.19]$$

➤ Porcentaje de varianza explicada

Este parámetro determina cuanta información contenida en las variables originales es recogida por los nuevos factores. Es uno de los primeros parámetros a examinar, tanto para los datos espectrales (**X**) como para los datos de concentración (**Y**). Ayuda a determinar el número apropiado de factores.

➤ Gráfico de concentración predicha vs. referencia

Este gráfico se utiliza para evaluar la falta de ajuste (*lack of fit*) y el bias o sesgo, así como para detectar muestras anómalas. Las muestras han de estar distribuidas aleatoriamente alrededor de una línea con pendiente 1 y ordenada en el origen 0.

➤ Gráfico de *loadings*

Además de ser útiles para determinar las regiones espectrales con mayor importancia, los *loadings* pueden utilizarse para examinar visualmente los factores que incorporan ruido en el modelo de calibración.

➤ Gráfico de *scores*

Se utilizan para examinar similitudes o diferencias entre muestras, así como para observar pautas anómalas en los datos.

➤ Residuales de concentración vs. *leverage*

El *leverage* (h), se define para un espectro x correspondiente a una matriz de datos espectrales X , como:

$$h = x^t (X^t X)x \quad [3.20]$$

El *leverage* muestra como de diferente es el registro espectral de una muestra frente al resto. Las muestras con un *leverage* muy superior al resto se han de investigar para detectar la causa, ya que pueden ser *outliers*. Así, muestras con residuales de concentración y *leverage* elevados son posibles *outliers*.

3.4.3 Resolución Multivariante de Curvas (MCR, *Multivariate Curve Resolution*)

3.4.3.1 Introducción

La resolución multivariante de curvas (MCR) es una técnica reciente en comparación a PLS y PCA. Su desarrollo se ha producido principalmente durante finales de la década de los 90, siendo en la actualidad una técnica consolidada a partir de la cual se han desarrollado otras técnicas quimiométricas. Inicialmente, MCR estaba considerada como una técnica con capacidad para obtener información cualitativa del sistema estudiado, actualmente MCR también se utiliza con fines cuantitativos, siendo remarcable la capacidad de esta técnica para resolver cuali- y cuantitativamente sistemas químicos con un mínimo de información externa.

MCR descompone una señal instrumental compleja en las contribuciones debidas a los componentes que forman el sistema, asumiendo que los datos siguen un modelo lineal ajustado a la ley de Beer, y que se puede expresar de forma matricial de la siguiente manera [21]:

$$X = CS^t + E \quad [3.21]$$

donde X ($M \times M$) es la matriz de datos espectrales adquiridos durante un proceso en evolución o reacción química, C ($M \times A$) es la matriz relacionada con los perfiles de concentración de las A especies involucradas en la reacción, S^t ($A \times M$) es la matriz de espectros puros (o perfiles espectrales) de las especies involucradas en la reacción, y E ($M \times M$) es la matriz de residuales no explicados por las matrices C y S^t .

Una vez conocido el número de especies involucradas en la reacción y activas espectroscópicamente, el objetivo de los métodos de resolución de curvas es resolver el sistema representado en la ecuación [3.21] utilizando poca información externa, cualquier conocimiento del sistema, o modelos empíricos (*hard modelling*).

3.4.3.2 Fundamentos

El sistema representado en la ecuación [3.21] está sometido a ambigüedades, debido a que, son muchas las matrices C y S^t posibles que multiplicadas entre si puedan resolver la matriz X con un error bajo. Sin embargo, existen pocas soluciones física- y químicamente aceptables, donde S^t corresponda a los espectros de los analitos estudiados y C a sus perfiles de concentración. La tarea de los métodos de resolución es reducir el número posible de soluciones.

Existen dos tipos de ambigüedades en MCR, la rotacional y la de intensidad. La ambigüedad de intensidad, reflejada en la expresión [3.22], muestra que puede

existir cualquier escalar α y $1/\alpha$ que multipliquen a las matrices C y S^t respectivamente, obteniéndose la misma matriz X :

$$X = C S^t = (C \alpha) (1/\alpha S^t) = C^* S^{t*} \quad [3.22]$$

La ambigüedad rotacional, reflejada en la expresión [3.23], demuestra que existe un número infinito de matrices T invertibles que hacen que la descomposición no sea única. Esta ambigüedad ocurre cuando no existe selectividad para alguno de los componentes.

$$X = C S^t = (C T) (T^+ S^t) = C^* S^{t*} \quad [3.23]$$

Teniendo en cuenta que la absorción en el infrarrojo próximo está formada por bandas anchas y solapadas, donde prácticamente no existen regiones espectrales selectivas, la ambigüedad rotacional es uno de los problemas más importantes en la resolución de estos sistemas. En sistemas poco selectivos, como es nuestro caso, el uso de restricciones y de información externa a la contenida en la matriz X ayuda a solventar las ambigüedades, y en consecuencia, reducir el número de soluciones posibles de la ecuación [3.21]. En la presente tesis se han utilizado espectros registrados de las especies puras y datos cuantitativos procedentes de métodos de referencia como información externa a la matriz X . La ventaja de la resolución multivariante de curvas respecto a otras técnicas de calibración es que requieren una menor cantidad de datos cuantitativos.

Una de las etapas claves en los métodos de resolución es establecer el número de componentes que pueden ser monitorizados. En ausencia de ruido y otras causas de variabilidad, el rango de la matriz X debería ser el mismo que el número de especies químicas. Existen diferentes herramientas para determinar el número de componentes, como es el caso del SVD (*Singular Value Decomposition*),

seleccionando los valores singulares más significativos (generalmente aquellos >1) o representando los correspondientes vectores propios.

En sistemas en evolución o en sistemas pocos selectivos, como es el caso del NIR, puede existir un problema de deficiencia de rango cuando el rango de la matriz es inferior al rango químico, que dificulta obtener los perfiles de todos los analitos involucrados en el proceso. Existen diferentes metodologías para solventar la deficiencia de rango, la más común es trabajar con matrices aumentadas, utilizando en un mismo cálculo matrices de espectros correspondientes a varios procesos (figura 3.4), o aumentando la matriz de datos con los espectros puros de las especies activas espectroscópicamente.

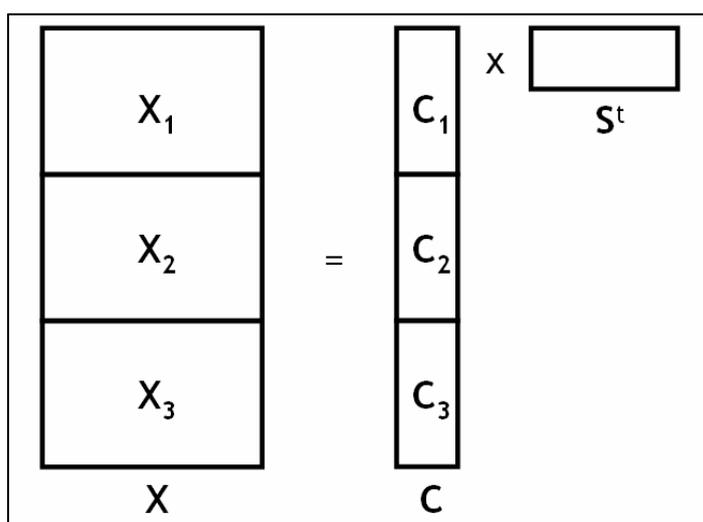


Figura 3.4. Matrices aumentadas para la aplicación de MCR-ALS

3.4.3.3 Algoritmo ALS (*Alternating Least Squares*)

La ecuación [3.21] se resuelve mediante el algoritmo de mínimos cuadrados alternados (ALS). Este método iterativo calcula en cada ciclo nuevas estimaciones de las matrices C y S^t , incorporando en cada ciclo una serie de restricciones basadas en el *knowledge* del proceso o en la información química disponible, con el objetivo de minimizar la matriz de residuales E . Para inicial el algoritmo ALS es necesario partir

de unas estimaciones iniciales de los perfiles espectrales o de los de concentración. Estas estimaciones pueden ser obtenidas a partir de métodos de resolución preeliminarios como el EFA (*Evolving Factor Analysis*) [35-37], SIMPLISIMA (*SIMPLe-to-use Iterative Self-modeling Mixture Analysis*) [38-39] o el ITTFA (*Iterative Target Transformation Factor Analysis*) [40-41].

El proceso implica las siguientes etapas:

- a) Construcción de unas estimaciones iniciales S^t .
- b) Aplicación del algoritmo ALS: (1) $C_i = X \cdot (S^t)^+$ y (2) $S_i^t = (C_i)^+ \cdot X$
- c) Reconstrucción de la nueva matriz X : $X_i = C_i \cdot S_i^t$ y estimación de la matriz de residuales $E_i = X - C_i \cdot S_i^t$.

El proceso (b-c) se repite hasta que se logra la convergencia o hasta que se alcanza un número preseleccionado de iteraciones (generalmente $i < 100$), definiéndose la convergencia como el cambio relativo del RMS (root mean square error) entre dos ciclos consecutivos.

$$RMS = \sqrt{\frac{\sum_i^M \sum_j^N e_{ij}^2}{N}} \quad [3.24]$$

3.4.3.4 Restricciones

Una de las principales características de los métodos de resolución es que no necesitan ningún tipo de información de referencia para ser aplicados. Sin embargo, toda la información disponible del sistema puede ser utilizada en forma de restricciones que se aplican durante el proceso iterativo sobre los perfiles espectrales o de concentración.

➤ Restricción de no-negatividad

Las concentraciones de las especies químicas han de ser siempre valores positivos. Los valores espectrales, salvo excepciones, son también positivos. Esta restricción se aplica forzando los valores negativos a ser cero o mediante algoritmos de mínimos cuadrados no negativos (NNLS, *Non-Negative Least Squares*).

➤ Restricción de unimodalidad

Esta restricción se puede aplicar en aquellos perfiles que se puede suponer que tienen un solo máximo. Es una situación corriente en los perfiles de concentración pero no tan corriente en los espectrales.

➤ Restricción de sistema cerrado (*closure*)

En sistemas cerrados la cantidad total de constituyentes es constante. Mediante esta restricción la suma de concentraciones de todas las especies químicas en cada punto de la reacción o del proceso es constante. Además, es posible aplicar restricciones adicionales de normalización, las cuales tienen un efecto importante sobre la ambigüedad de intensidad o escala.

➤ Restricción de igualdad (*equality constraint*)

Si se posee información respecto a los perfiles de concentración o espectrales, ya sea total o parcial, puede ser utilizada como restricción de igualdad, haciendo que las estimaciones obtenidas en cada iteración se adapten a la información suministrada [42].

De forma general, la aplicación de MCR-ALS se puede resumir con el esquema de la figura 3.5.

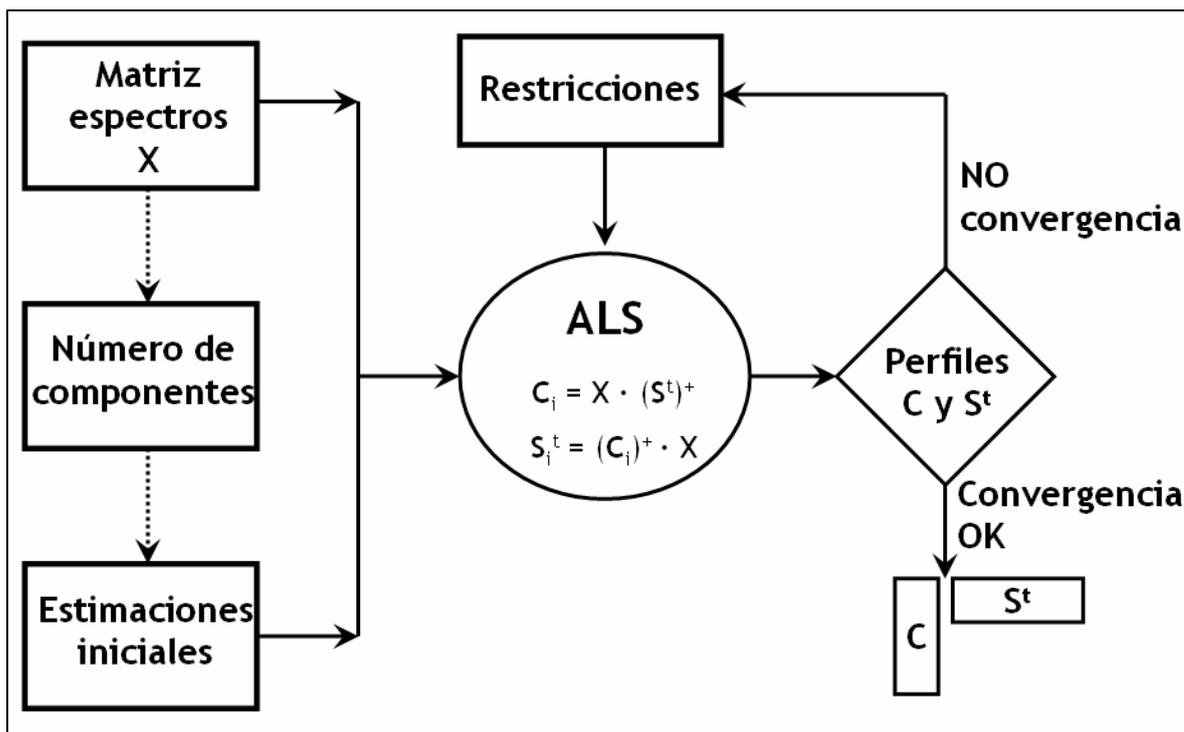


Figura 3.5. Esquema de la aplicación de MCR-ALS

3.4.3.5 Evaluación de resultados

Una vez finalizado el proceso iterativo, el ajuste del modelo obtenido se evalúa en términos de varianza capturada, definida como:

$$\%Var = \frac{\sum_i^M \sum_j^N \hat{x}_{ij}^2}{\sum_i^M \sum_j^N x_{ij}^2} \times 100 \quad [3.25]$$

Y los espectros estimados por MCR se comparan con los originales mediante coeficiente de correlación:

$$r = \cos \gamma = \frac{S_i^t \hat{S}_i}{\|S_i\| \cdot \|\hat{S}_i\|} \quad [3.26]$$

3.4.3.6 Seguimiento de procesos a tiempo real

Optimizando el cálculo de MCR para uno o varios procesos similares de forma conjunta, es posible calcular una matriz de coeficientes que permite la predicción de las concentraciones de las especies químicas en nuevos procesos. La matriz de coeficientes K se define como:

$$K = (S^t)^+ \quad [3.27]$$

El cálculo de concentraciones de un nuevo proceso se realiza:

$$\hat{C} = S \cdot K \quad [3.28]$$

Donde S es la matriz de espectros registrados para el nuevo proceso o un espectro individual.

3.5 Límites de detección en calibración PLS

El límite de detección es un valor, expresado en unidades de concentración, que describe el nivel de concentración más bajo que un analista puede determinar estadísticamente como diferente respecto a un blanco [43]. En calibración de orden cero el límite de detección es un parámetro necesario en la validación de métodos analíticos, definido por varios organismos internacionales (IUPAC, ACS). Sin embargo, en calibración multivariable el límite de detección es todavía un parámetro "difuso": existen múltiples definiciones y todavía ningún organismo internacional ha marcado una pauta.

Una de las primeras definiciones de límite de detección multivariable (MDL) en sistemas de calibración inversa, $MDL = 3 \cdot \|\varepsilon\| \cdot \|b\|$ [44], no es más que una adaptación de los sistemas de orden cero a los de primer orden donde $\|\varepsilon\|$ y $\|b\|$

corresponden a la norma euclídea del ruido instrumental y de los coeficientes de regresión del modelo respectivamente. Definiciones posteriores del MDL se basan, en su mayoría, en propagación de errores, donde el límite de detección se define en función de la varianza de la concentración predicha por el modelo de calibración:

$$MDL = \Delta(\alpha, \beta) \cdot \text{vâr}(\hat{c}_0)^{1/2} \quad [3.29]$$

donde $\Delta(\alpha, \beta)$ es el parámetro t de Student para una distribución descentrada [45].

Este método ha recibido una considerable atención en los últimos años y multitud de expresiones han sido propuestas para el cálculo de la varianza para modelos PLS [46-57]. Una de las primeras expresiones es la propuesta por Faber y Kowalski [48], que posteriormente ha sido simplificada por Faber y Bro [54] en:

$$\text{vâr}(c) = (1 + h)MSEC - \sigma_c^2 \quad [3.30]$$

donde h es el leverage de la muestra de predicción, σ_c^2 es la varianza del método de referencia y el MSEC se define como [56]:

$$MSEC = \frac{\sum_{i=1}^M (\hat{y}_i - y_i)^2}{M - df} \quad [3.31]$$

donde df se puede calcular mediante la ecuación propuesta por Van der Voet [58]:

$$df = M \left[1 - \sqrt{\frac{MSEP_{Cal}}{MSEP_{CV}}} \right] \quad [3.32]$$

definiéndose MSEP, para calibración y *cross-validation*, como:

$$MSEP = \frac{1}{M} \sum_{i=1}^M (\hat{y}_i - y_i)^2 \quad [3.33]$$

A pesar de que la ecuación 3.30 es aceptada por varios autores [54, 56-57], no existe un consenso en cual es la ecuación idónea para determinar la varianza en sistemas multivariantes, existiendo varias expresiones para el cálculo de la varianza de la ecuación 3.29. Asumiendo esta ambigüedad, y que es necesaria más investigación sobre el tema, Ortiz *et al.* [59] proponen trasladar la metodología utilizada en calibraciones de orden cero, ampliamente aceptada y definida por la ISO [60], a sistemas multivariantes. Esta solución está fundamentada en que la capacidad de detección en el dominio de las concentraciones es invariante frente a transformaciones lineales de la señal [59]. En este sentido, las ecuaciones definidas por la ISO para determinar el límite de detección pueden ser generalizadas para cualquier método de calibración multivariable lineal, como el PLS, estimándose el límite de detección a partir del gráfico de concentración predicha vs. referencia.

Según la ISO, la varianza de la señal \hat{y} , que en el caso multivariable correspondería con la concentración predicha, se define como:

$$\text{var}(\hat{y}) = w_x^2 \sigma^2 \quad [3.34]$$

donde:

$$w_x^2 = \frac{1}{r} + \frac{1}{M} + \frac{(x - \bar{x})^2}{\sum_{i=1}^M (x_i - \bar{x})^2} \quad [3.35]$$

siendo r el número de determinaciones de la muestra problema, M el número de muestras de calibración y σ^2 se calcula mediante la ecuación 3.31 con $df = 2$.

Estas dos metodologías comentadas para el cálculo de límites de detección han sido utilizados en esta tesis para estimar la capacidad de detección del NIR es sistemas donde la concentración del analito a determinar era próxima a concentraciones traza.

Referencias

- [1] D. L. Massart, B. G. M. Vandegiste, L. M. C. Buydens, S. De Jong, P. J. Lewi and J. Smeyers. *Handbook of chemometrics and qualimetrics*, Elsevier, Amsterdam (1997).
- [2] K. R. Beebe, R. J. Pell and M. B. Seasholtz. *Chemometrics. A practical guide*, John Wiley & sons, New York (1998).
- [3] B. R. Kowalski. Chemometrics: views and propositions. *Journal of Chemical Information and Computer Sciences*, 15 4 (1975) 201-203.
- [4] M. Blanco, D. Valdés. Suppressing the temperature effect in near infrared spectroscopy by using orthogonal signal correction. *J. Near Infrared Spectrosc.* 14 (2006) 155-160
- [5] A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36 (1964) 1627-1639.
- [6] G. Horlick. Digital data handling of spectra utilizing Fourier transformations. *Anal. Chem.* 44 (1972) 943-947.
- [7] R. J. Barnes, M. S. Dhanoa and S. J. Lister. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43 (1989) 772-777.
- [8] K. H. Norris and P. C. Williams. Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size. *Cereal Chemistry* 61 2 (1984) 158-165.
- [9] S. Wold, H. Antic, F. Lindgren and J. Öhman. Orthogonal signal correction of near-infrared spectra. *Chemom. Intell. Lab. Syst.* 44 1-2 (1998) 175-185.
- [10] C. A. Andersson. Direct orthogonalisation. *Chemom. Intell. Lab. Syst.* 47 (1999) 51-63.
- [11] T. Fearn. On orthogonal signal correction. *Chemom. Intell. Lab. Syst.* 50 (2000) 47-52.

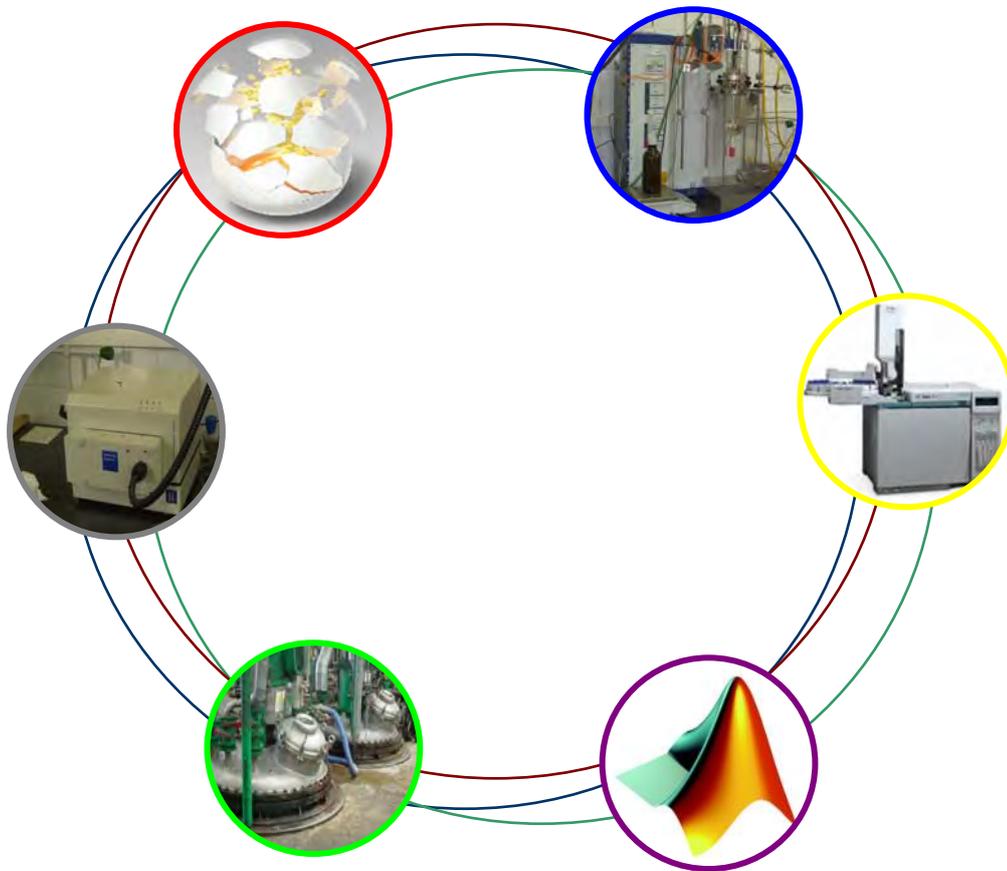
- [12] E. Bertran, H. Iturriaga, S. MasPOCH and I. Montoliu. Effect of orthogonal signal correction on the determination of compounds with very similar near infrared spectra. *Anal. Chim. Acta* **431** (2001) 303-311.
- [13] J. Ferré, S. D. Brown and F. X. Rius. Improved calculation of the net analyte signal in inverse multivariate calibration. *J. Chemometrics* **15** (2001) 537-553.
- [14] N. M. Faber. Efficient computation of net analyte signal vector in inverse multivariate calibration models. *Anal. Chem.* **70** (1998) 5108-5110.
- [15] N. M. Faber. Mean centering and computation of scalar net analyte signal in multivariate calibration. *J. Chemometrics* **12** (1998) 405-409.
- [16] M. Martínez, D. Picón, J. L. Martínez, A. Garrido, A. Espinosa, A. Muñoz and F. Salinas. Determination of carbendazim, thiabendazole and fuberidazole using net analyte signal-based method. *Talanta* **59** (2003) 1107-1116.
- [17] A. Lorber. Error propagation and figures of merit for quantification by solving matrix equation. *Anal. Chem.* **58** (1986) 1167-1172.
- [18] A. Lorber, K. Faber and B. R. Kowalski. Net analyte signal calculation in multivariate calibration. *Anal. Chem.* **69** (1997) 1620-1626.
- [19] M. A. Sharaf, D. L. Illman and B. R. Kowalski. *Chemometrics*, John Wiley & Sons, New York (1986).
- [20] H. Martens and T. Naes. *Multivariate calibration*, John Wiley & Sons, England, 1989.
- [21] R. Tauler. *Anàlisi de mesclres mitjançant resolució multivariant de corbes*, Institut d'Estudis Catalans, Barcelona (1997).
- [22] R. Kramer. *Chemometrics techniques for quantitative analysis*, Marcel Dekker, New York (1998).
- [23] M. Otto. *Chemometrics. Statistics and computer application in analytical chemistry*,

-
- Wiley-VCH, New York (1999).
- [24] A. K. Smilde, R. Bro and P. Geladi. *Multi-way analysis. Applications in the chemical sciences*, John Wiley & Sons, England (2004).
- [25] J. B. Kruskal. *Factor analysis and principal components*. International encyclopaedia of statistics, The Free Press, New York (1978).
- [26] J. E. Jackson. Principal components and factor analysis: part I - principal components. *J. of Quality Tech.* **12** (1980) 201-213.
- [27] S. Wold, K. H. Esbensen and P. Geladi. Principal component analysis. *Chemom. Intell. Lab. Syst.* **2** (1987) 37-52.
- [28] H. Wold. *Multivariate analysis*, Krishnalah, New York (1966).
- [29] H. Wold. *Soft modelling. The basic design and some extensions. In systems under indirect observation*. Eds. K. G. Jöreskog and H. Wold. Amsterdam (1982).
- [30] P. Geladi. Notes on the history and nature of partial least squares (PLS) modelling. *J. Chemometr.* **2** (1988) 231-246.
- [31] A. Höskuldsson. PLS regression methods. *J. Chemometr.* **2** (1988) 211-228.
- [32] A. Phatak and S. De Jong. The geometry of partial least squares. *J. Chemometr.* **11** (1997) 311-338.
- [33] A. J. Burnham, J. F. McGregor and R. Viveros. A statistical framework for multivariate latent variable regression methods based on maximum likelihood. *J. Chemometr.* **13** (1999) 49-65.
- [34] P. Geladi and B. R. Kowalski. An example of 2-block predictive partial least-squares regression with simulated data. *Anal. Chim. Acta* **185** (1985) 19-32.
- [35] H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbühler. Calculation of equilibrium
-

- constants from multiwavelength spectroscopic data. Model-free analysis of spectrophotometric and ESR titrations. *Talanta* **32** (1985) 1133-1139.
- [36] M. Maeder. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Anal. Chem.* **59** (1987) 527-530.
- [37] P. J. Gemperline and C. Hammlton. Evolving factor analysis applied to flow injection analysis data. *J. Chemometr.* **3** (1989) 455-461.
- [38] W. Windig. Mixture analysis of spectral data by multivariate methods. *Chemom. Intell. Lab. Syst.* **4** (1988) 201-213.
- [39] F. Cuesta and D. L. Massart. Application of SIMPLISIMA for the assessment of peak purity in liquid chromatography with diode array detection. *Anal. Chim. Acta.* **298** (1994) 331-339.
- [40] B. G. M. Vandeginste, R. Essert, T. Bosman, J. Reijnen and G. Kateman. Three-component curve resolution in liquid chromatography with multiwavelength diode array detector. *Anal. Chem.* **57** (1985) 971-985.
- [41] P. J. Gemperline. Target transformation factor analysis with linear inequality constraints applied to spectroscopic-chromatographic data. *Anal. Chem.* **58** (1986) 2656-2663.
- [42] P. J. Gemperline and E. Cash. Advantages of soft versus hard constraints in self-modelling curve resolution problems. Alternating least squares with penalty functions. *Anal. Chem.* **75** 16 (2003) 4236-4243.
- [43] G. L. Long and J. D. Winefordner. Limit of detection. A closer look at the IUPAC definition. *Anal. Chem.* **55** 7 (1983) 712A-719A.
- [44] K. S. Booksh and B. R. Kowalski. Theory of analytical chemistry. *Anal. Chem.* **66** 15 (1994) 782A-791A.

-
- [45] A. Clayton, J. W. Hines and P. D. Elkins. Detection limits with specified assurance probabilities. *Anal. Chem.* **59** (1987) 2506-2514.
- [46] A. Lorber and B. R. Kowalski. Estimation of prediction error for multivariate calibration. *J. Chemometr.* **2** (1988) 93-109.
- [47] S. De Vries and C. J. Ter Braak. Prediction error in partial least squares regression: a critique on the deviation used in The Unscrambler. *Chemom. Intell. Lab. Syst.* **30** (1995) 239-245.
- [48] K. Faber and B. R. Kowalski. Prediction error in least squares regression: further critique on the deviation used in The Unscrambler. *Chemom. Intell. Lab. Syst.* **34** (1996) 283-292.
- [49] N. M. Faber and B. R. Kowalski. Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares. *J. Chemometr.* **11** (1997) 181-238.
- [50] M. Hoy, K. Steen and H. Martens. Review of partial least squares regression prediction error in Unscrambler. *Chemom. Intell. Lab. Syst.* **44** (1998) 123-133.
- [51] N. M. Faber, D. L. Duewer, S. J. Choquette, T. L. Green and S. N. Chesler. Characterizing the uncertainty in near-infrared spectroscopic prediction of mixed-oxygenate concentrations in gasoline: sample-specific prediction intervals. *Anal. Chem.* **70** (1998) 2972-2982.
- [52] R. Boqué, M. S. Larrechi and F. X. Rius. Multivariate detection limits with fixed probabilities of error. *Chemom. Intell. Lab. Syst.* **45** (1999) 397-408.
- [53] N. M. Faber. Comparison of two recently proposed expressions for partial least squares regression prediction error. *Chemom. Intell. Lab. Syst.* **52** (2000) 123-134.
- [54] N. M. Faber and R. Bro. Standard error of prediction for multiway PLS 1. Background and a simulation study. *Chemom. Intell. Lab. Syst.* **61** (2002) 133-149.
-

- [55] A. C. Olivieri. A simple approach to uncertainty propagation in preprocessed multivariate calibration. *J. Chemometr.* **16** (2002) 207-217.
- [56] J. A. Fernández Pierna, L. Jin, F. Wahl, N. M. Faber and D. L. Massart. Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error. *Chemom. Intell. Lab. Syst.* **65** (2003) 281-291.
- [57] N. M. Faber, X. H. Song and P. K. Hopke. Sample-specific standard error of prediction for partial least squares regression. *Trends* **22** 5 (2003) 330-334.
- [58] H. Van der Voet. Pseudo-degrees of freedom for complex predictive models: the example of partial least squares. *J. Chemometr.* **13** (1999) 195-208.
- [59] M. C. Ortiz, L. A. Sarabia, A. Herrero, M. S. Sánchez, M. B. Sanz, M. E. Rueda, D. Giménez and M. E. Meléndez. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. *Chemom. Intell. Lab. Syst.* **69** (2003) 21-33.
- [60] ISO 11843-1. *Capability of detection*, Genève, Switzerland (1997).



METODOLOGÍA Y DISCUSIÓN DE RESULTADOS

Este capítulo presenta los resultados obtenidos en el transcurso de la tesis doctoral. En la primera parte se describe la metodología instrumental empleada para la obtención de muestras, y de los valores de la propiedad a determinar mediante métodos analíticos de referencia. En la segunda parte se detallan y comentan los resultados obtenidos por las técnicas de calibración multivariable ensayadas.

Contenidos:

4. Metodología experimental
5. Resultados y discusión



4. Metodología experimental

4. Metodología experimental

4.1 Introducción

En este capítulo se describen las características más relevantes de la metodología empleada en el desarrollo de los métodos de análisis propuestos: los procesos estudiados, el registro de espectros NIR, la obtención de la propiedad analítica de interés y, finalmente, el procesamiento de datos.

4.2 Procesos químicos estudiados y obtención de muestras

4.2.1 Escalado

Los procesos estudiados se han realizado en reactores industriales de producción por lotes de 5.000 - 10.000 Kg (figura 4.1), y en un reactor de laboratorio de 1 litro de capacidad (figura 4.2), reproduciendo las condiciones industriales. En ambos casos se ha utilizado materia prima del mismo origen.



Figura 4.1. Reactor industrial de Uniqema (I.C.I Spain)

El reactor de laboratorio, figura 4.2, es un LabMax (Mettler Toledo Inc.), controlado mediante el software Camile Tg 3.06 (Sagian Inc.). Este instrumento permite el control de las temperaturas de reacción y de camisa, de la velocidad de

agitación, así como el control de la adición de reactivos utilizando medidas de volumen (bomba de pistón) o peso (balanza).

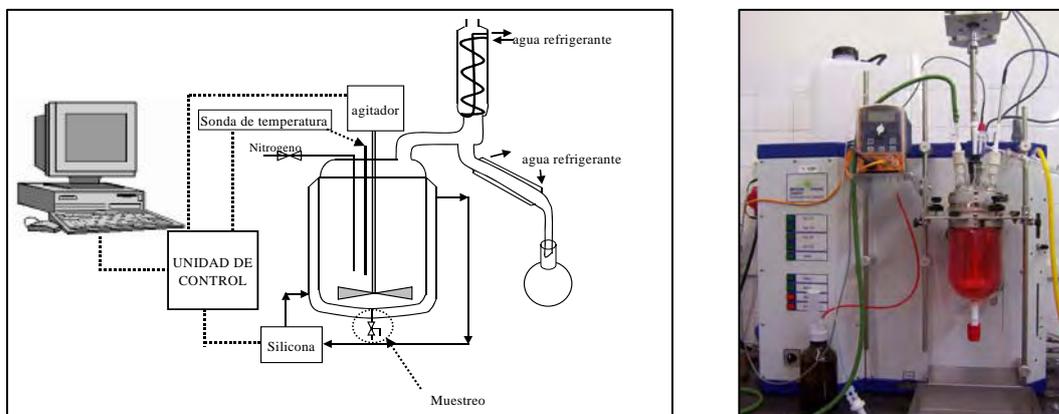


Figura 4.2. Reactor de laboratorio, esquema (izq.), fotografía (der.).

4.2.2 Esterificación entre glicerina y mezcla de ácidos caprílico y cáprico

En este proceso, una vez cargados los reactivos: glicerina (Uniqema ICI Spain, pureza 99%, $T_{fus} = 17,8^{\circ}C$, $T_{eb} = 290^{\circ}C$) y la mezcla de ácidos caprílico (ácido octanóico, Uniqema ICI Spain, pureza 99%, $T_{fus} = 16,7^{\circ}C$) y cáprico (ácido decanóico, Uniqema ICI Spain, $T_{fus} = 31,4^{\circ}C$), se incrementa la temperatura hasta $230^{\circ}C$, a presión atmosférica y con un flujo constante de nitrógeno para evitar que oxidaciones indeseadas den color al producto. Durante la reacción, tanto el agua que contiene la materia prima, como la que se forma como producto de la reacción de esterificación, se destila a causa de la elevada temperatura.

La glicerina posee tres grupos hidroxilo por molécula, por lo que puede formar mono, di y triglicéridos, y mezclas de ellos en diferentes proporciones, que además en este caso pueden ser de diferentes ácidos, caprílico o cáprico. El monoglicérido tiene 4 isómeros posibles, en la posición 1 o 2 con C8 (caprílico) y C10 (cáprico). Los diglicéridos dan dos isómeros de posición: 1,2 y 1,3 dando lugar a 4 isómeros. Las

combinaciones para el triglicérido son mayores, generándose 6 isómeros. Los diferentes compuestos que se pueden obtener se muestran en la figura 4.3.

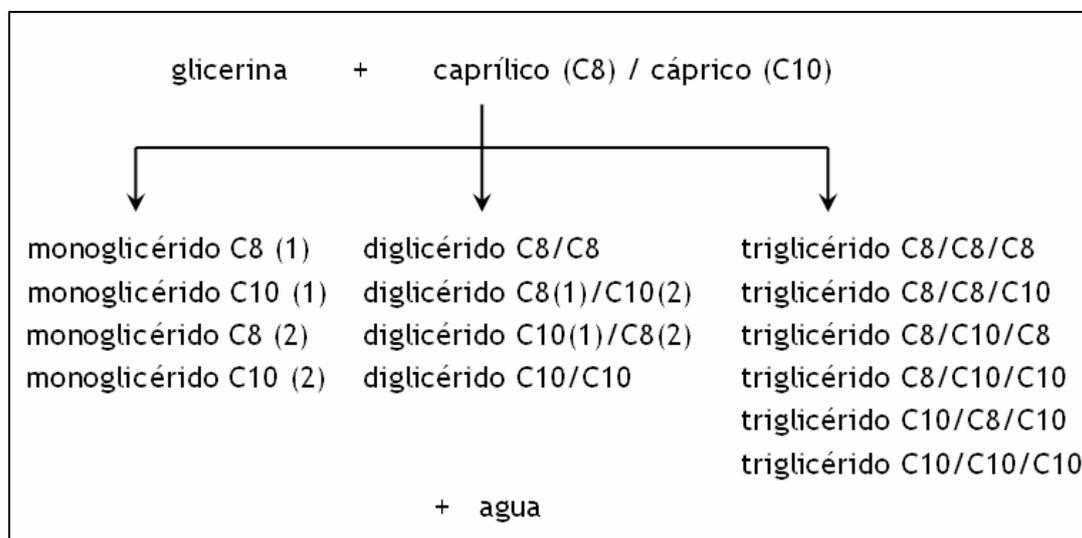


Figura 4.3. Productos de la reacción entre glicerina y ácidos C8 y C10. (1) corresponde con grupo hidroxilo del carbono 1 de la glicerina, (2) corresponde con el carbono 2.

Además de obtenerse una amplia variedad de productos, la proporción entre ellos varía en función de la proporción inicial entre glicerina y ácidos, pudiéndose obtener productos con diferentes aplicaciones y características físicas. Así, si se desea obtener principalmente triglicérido, se utiliza un exceso de ácido, con una relación molar glicerina:ácido de 1:3,4 en nuestro estudio (utilizando una proporción 60/40 de ácidos caprílico y cáprico). En este caso, una vez finalizada la reacción, se elimina el exceso de ácido por destilación aplicando vacío. Por el contrario, si se desea obtener una mezcla de glicéridos con un elevado contenido en monoglicérido, es necesario un exceso de glicerina, utilizándose en esta tesis relaciones molares de 1,1:1 y 1,35:1 glicerina:ácido (con una proporción 95/5 caprílico/cáprico). Posteriormente, para obtener el producto será necesario eliminar el exceso de glicerina mediante sucesivos lavados con agua.

Estos procesos se han realizado tanto a escala industrial como en el reactor de laboratorio, y han sido seguidos *at-line* (*off-line* en el caso de los procesos industriales) mediante espectroscopia NIR. Las muestras fueron extraídas a intervalos de 20 - 30 minutos durante la reacción, excepto en los momentos iniciales, debido a la existencia de dos fases a causa de la inmiscibilidad mutua de glicerina y mezcla de ácidos.

El número de procesos realizados y de muestras obtenidas es el siguiente:

- Procesos con exceso de glicerina:

6 procesos de laboratorio con diferente estequiometría: 3 con 1,35:1, 2 con 1,28:1, y 1 con 1,1:1.

1 proceso industrial con estequiometría 1,35:1.

- Procesos con exceso de mezcla de ácidos con una proporción inicial de reactivos 1:3,4 glicerina:ácido

2 procesos de laboratorio con 8 y 11 muestras cada uno.

6 procesos industriales con 9, 7, 7, 6, 6 y 5 muestras cada uno.

Además, se dispone de un elevado número de muestras sin clasificar de diferentes procesos.

4.2.3 Esterificación entre isopropanol y ácido mirístico

Esta reacción consiste en la esterificación directa entre ácido mirístico e isopropanol para la obtención de miristato de isopropilo. En este proceso se cargan 500 g de ácido mirístico (ácido tetradecanóico, Uniqema ICI Spain, pureza 99%, $T_{fus} = 58,5^{\circ}C$)

en el reactor de laboratorio, y se mantiene un flujo constante de nitrógeno. Una vez alcanzada la temperatura de reacción, 130°C, se añade el catalizador (mezcla de ácido p-toluensulfónico y ácido hipofosforoso) disuelto en 10 ml de isopropanol (Uniqema ICI Spain, pureza 99%, $T_{eb} = 82,5^{\circ}\text{C}$), y comienza la adición constante de isopropanol a una velocidad de 0,5 g/min. Debido a que la adición de isopropanol es lenta y constante, su concentración en el medio de reacción es baja: el isopropanol que no reacciona, destila por la elevada temperatura de reacción (superior al punto de ebullición del isopropanol). La configuración del LabMax para este proceso se muestra en la figura 4.4, siendo remarcable la adición de reactivo controlada por balanza y la sonda NIR sumergida en el medio de reacción.

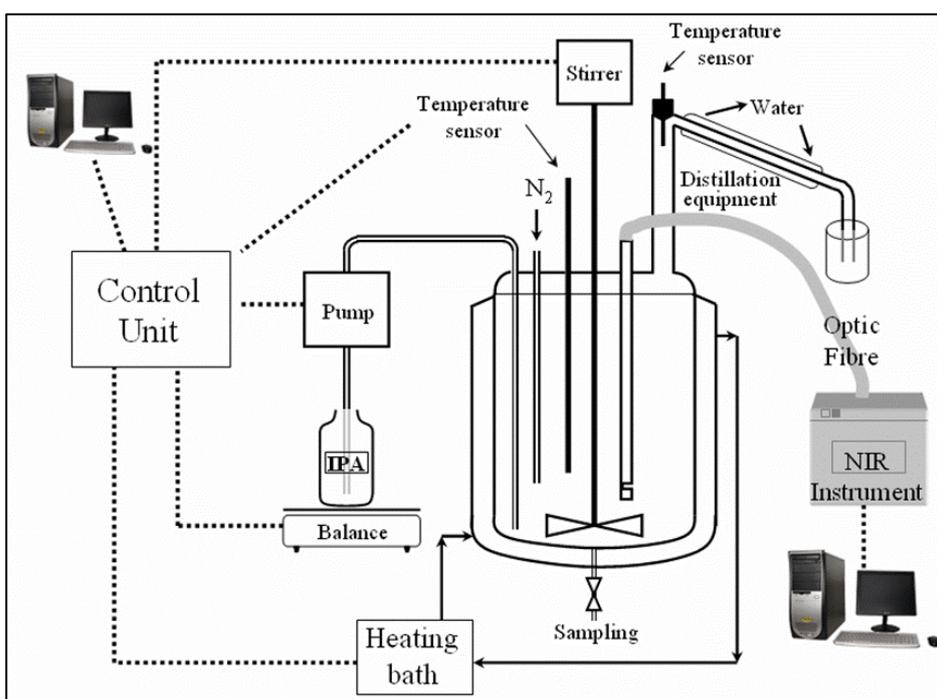


Figura 4.4. Esquema del reactor utilizado para el miristato de isopropilo.

A escala industrial el proceso se puede realizar con isopropanol de menor riqueza, o incluso se pueden reutilizar los destilados de procesos anteriores, los cuales están compuestos por mezclas de isopropanol y agua. Para comprobar el efecto del agua en la cinética de la reacción, se utilizó isopropanol con contenidos

de agua entre 0 y 25%. Se han realizado un total de 7 procesos, con 0%, 5%, 10%, 15%, 20% y 25% de agua, con una velocidad de adición de 0,5 g/min, más un proceso con el 10% de agua y una velocidad de adición de 1 g/min. Esta información está resumida en la tabla 4.1.

Los datos NIR se han obtenido mediante una sonda NIR *in-line*. Para comprobar la evolución de la reacción, aproximadamente cada 60 minutos, se extraía una muestra y se determinaba el índice de ácido (apartado 4.4.1), lo que permitía detener el proceso cuando no se observase evolución en la reacción. Para estudiar la evolución del contenido en agua en los destilados se tomaba una muestra de destilados aproximadamente cada 60 minutos. Para cada destilado se registraba el espectro NIR *at-line*.

Proceso	% agua	% isopropanol	Velocidad adición (g/min)	nº muestras	nº destilados
1	0	100	0.5	12	11
2	5	95	0.5	13	12
3	10	90	0.5	10	11
4	15	85	0.5	13	13
5	20	80	0.5	12	12
6	25	75	0.5	12	13
7	10	90	1.0	12	13

Tabla 4.1. Información de los procesos de esterificación entre ácido mirístico e isopropanol realizados en el LabMax.

4.2.4 Destilación de 2-etilhexanol de una matriz del éster del ácido de coco

Este proceso de purificación tiene por objeto la eliminación del exceso de 2-etilhexanol (2-etil-1-hexanol, Uniqema ICI Spain, pureza 99%, $T_{eb} = 184^{\circ}\text{C}$) remanente en el producto de la reacción de esterificación entre el 2-etilhexanol y el ácido graso de coco (compuesto principalmente por ácidos láurico, mirístico, palmítico y esteárico, Uniqema ICI Spain, $T_{fus} = 21 - 25^{\circ}\text{C}$). La mezcla final de reacción se trata en un reactor a una temperatura de 120°C y a vacío (presión inferior a 10 mbar) manteniendo un flujo de nitrógeno de 200 l/h, hasta que la concentración del alcohol sea inferior a 200 ppm. Este proceso ha sido realizado únicamente a escala industrial. Como el objetivo principal es eliminar el olor producido por el alcohol, este proceso es llamado desodorización.

La concentración del alcohol al inicio de la desodorización es aproximadamente un 3%, siendo el intervalo inferior a 1.000 ppm el de mayor interés, motivo por el cual la mayoría de las muestras obtenidas están dentro de este intervalo.

4.2.5 Obtención de muestras dopadas en el laboratorio

4.2.5.1 Muestras para la esterificación entre glicerina y ácidos caprílico y cáprico

Para ampliar los intervalos de concentración y disminuir la correlación entre concentraciones se decidió preparar muestras en el laboratorio. Sin embargo, debido a la imposibilidad de disponer de todos los compuestos con un grado de pureza elevado, como por ejemplo mono o diglicéridos, no es posible preparar muestras sintéticas por pesada de los productos puros. Por este motivo, se tomó como alternativa el dopaje de muestras de reacción con los productos puros disponibles:

glicerina, triglicérido, ácidos y agua. Un total de 32 muestras fueron preparadas, teniendo en cuenta los intervalos de concentración deseados para los analitos. De esta forma es posible romper la correlación existente y ampliar los intervalos de concentración inferior y superior para agua, glicerina, ácidos y triglicérido, y el intervalo inferior para mono y diglicéridos. El intervalo superior de concentraciones para estos dos compuestos se amplió realizando procesos con una mayor proporción inicial de glicerina.

4.2.5.2 Muestras dopadas para la desodorización de 2-etilhexanol

Con el objetivo de cubrir intervalos de concentración y de incrementar el número de muestras en el intervalo 0 - 1.000 ppm, muestras con baja concentración de 2-etilhexanol fueron analizadas por cromatografía de gases para determinar la concentración del alcohol, y posteriormente fueron dopadas con 2-etilhexanol. Así, alícuotas de una muestra con 56,1 ppm de 2-etilhexanol fueron dopadas a 81, 106, 155, 204, 253, 303 y 352 ppm; dos alícuotas de otra muestra con 292 ppm fueron dopadas a 556 y 1080 ppm; y finalmente dos alícuotas de otra muestra con 451 ppm fueron dopadas hasta 1060 y 1630 ppm de 2-etilhexanol.

4.3 Instrumentación NIR

Los espectros NIR se han obtenido mediante dos metodologías: *at-line*, extrayendo muestras del reactor y registrando el espectro NIR en un laboratorio anexo, e *in-line*, introduciendo una sonda en el medio de reacción.

4.3.1 Seguimiento *at-line* de procesos químicos

Para el seguimiento de la reacción entre glicerina y ácidos caprílico y cáprico, y para el control de los destilados para la reacción entre ácido mirístico e isopropanol, se ha

utilizado un espectrofotómetro FOSS NIRSystems 6500 con un módulo RCA (Rapid Content Analyzer), mostrado en la figura 5.5. El registro espectral se realiza por transflectancia (apartado 3.2.3), con una cubeta de cuarzo y un reflector de oro (figura 4.5) que limita el camino óptico a 1 mm ($2 \times 0,5$ mm). Cada espectro es el promedio de 32 scans, registrado en el intervalo 1100-2500 nm con una resolución de 2 nm.



Figura 4.5. Fotografía del modulo RCA (arriba), disposición de los 6 elementos de detección (izq.) y sistema de medida con cubeta de cuarzo y reflector de oro (der.).

Para el seguimiento de la destilación de 2-etilhexanol se ha utilizado un espectrofotómetro FOSS 5000 con un módulo VHM (Vial Heater Module), mostrado en la figura 4.6, situado en un laboratorio adjunto a la planta de producción. El registro espectral se realiza a 75°C , por transmisión, en viales de vidrio con un camino óptico de 8 mm. Cada espectro es el promedio de 32 scans, registrado en el intervalo 1100-2200 nm con una resolución de 2 nm.



Figura 4.6. Módulo VHM (izq.) y conjunto NIR-VHM (der.)

4.3.2 Seguimiento *in-line* de procesos químicos

Para el seguimiento *in-line* de la esterificación entre ácido mirístico e isopropanol se utilizó un espectrofotómetro FOSS 5000 equipado con una sonda de fibra óptica que permite realizar medidas por transflectancia, con un camino óptico de 0,4 mm (2 x 0,2 mm), ver figura 4.7. El registro espectral se realiza a intervalos de 3 minutos, siendo cada espectro el promedio de 32 scans, registrado en el intervalo 1100-2500 nm con una resolución de 2 nm.



Figura 4.7. Espectrofotómetro NIR con sonda de registro por transflectancia (sup. izq.), reactor LabMax con la sonda de registro (der.) y ampliación del camino óptico (inf. izq.).

4.4 Métodos de referencia

4.4.1 Determinación de índice de ácido e índice de hidroxilo

Los procesos de esterificación estudiados son frecuentemente controlados en la industria mediante métodos analíticos dictados por la *European Pharmacopoeia*, como el índice de ácido o el índice de hidroxilo [1]. La evolución de estos parámetros marca el estado de la reacción. Estos dos parámetros han sido utilizados para el seguimiento de la reacción entre glicerina y ácidos caprílico y cáprico.

4.4.1.1 Índice de ácido

El índice de ácido (IA) es la cantidad de hidróxido potásico necesarios para neutralizar los ácidos libres y se expresa como mg. de KOH necesarios para neutralizar un gramo de muestra. Es un método basado en una valoración ácido-base que utiliza hidróxido de potasio como agente valorante. La muestra se disuelve en etanol:xileno 1:1 y se valora con KOH 0,1 M, detectando el punto final con fenolftaleína. El índice de ácido (en mg KOH/g muestra) se calcula según la fórmula:

$$IA = \frac{56,1 \cdot V \cdot M}{W} \quad [4.1]$$

Donde:

V: volumen de KOH gastados en la valoración en ml.

M: concentración de la solución de KOH en M.

W: peso de muestra en g.

56,1: factor de conversión mol KOH/ g KOH.

Los valores de IA para los procesos estudiados están comprendidos entre 1 y 250 mg KOH/g muestra.

4.4.1.2 Índice de hidroxilo

El índice de hidroxilo (IOH) es la medida de los grupos hidroxilos presentes por gramo de muestra y se expresa como mg de KOH que se requieren para neutralizar la cantidad de ácido acético capaz de combinarse por acetilación con los grupos hidroxilo.

El método se basa en la acetilación de los grupos hidroxilo con una cantidad conocida, y en exceso, de anhídrido acético:



El anhídrido acético en exceso que no ha reaccionado se hidroliza:



El ácido acético formado por ambas reacciones se determina por valoración con KOH. El IOH se calcula teniendo en cuenta el volumen de KOH gastado para neutralizar los moles de acético generados por las reacciones [r4.1] y [r4.2] para la muestra, y el volumen de KOH necesario para neutralizar los moles de acético que se generan en un blanco (solo reacción de hidrólisis). La diferencia entre ambos volúmenes es un parámetro relacionado con la concentración de grupos OH en la muestra, ya que a mayor concentración de grupos OH, menor será la cantidad de anhídrido a hidrolizar.

Debido a que en la valoración con KOH para la muestra se determina tanto el ácido acético como cualquier otro ácido que pueda contener la muestra, es necesario determinar también el IA y tenerlo en cuenta en el cálculo del IOH (eq. [4.2]).

El procedimiento para determinar el IOH se detalla a continuación. La muestra se pesa en un matraz de fondo plano y se añaden 5 ml de anhídrido acético al 30% en piridina. Al mismo tiempo, en otro matraz se añaden 5 ml de la misma disolución de anhídrido que se utilizará de blanco. Se instala a ambos matraces una columna de reflujo y se trasladan a un baño de arena donde se calienta a 90°C durante una hora. Pasado este tiempo, se añaden 2 ml de agua y se mantienen los matraces en el baño durante 10 minutos más. Posteriormente, se sacan del baño y se limpian la columna y las paredes del matraz con etanol al 96%. Se añade una gota de fenolftaleína y se valora con KOH 0,5 M.

El índice de hidroxilo (en mg KOH/g muestra) se determina según la fórmula:

$$IOH = \frac{56,1 \cdot (V_1 - V_2) \cdot M}{W} + IA \quad [4.2]$$

dónde:

V_1 : volumen de KOH gastados en el blanco en ml.

V_2 : volumen de KOH gastados en la muestra en ml.

M: concentración de la disolución de KOH en M.

W: peso de la muestra en gramos.

IA: índice de ácido.

56,1: factor de conversión mol KOH/g KOH.

Los valores de IOH para las muestras analizadas están comprendidos entre 1 y 120 mg KOH/g muestra.

4.4.2 Determinación de glicéridos, 2-etilhexanol e isopropanol por cromatografía de gases

El cromatógrafo de gases empleado es un Hewlett-Packard 5890 Series II (Fig. 4.8) con inyector split/splitless para columnas capilares y detector de ionización en llama (FID), controlado mediante el software Chemstation v. A.09.01. En los diferentes métodos desarrollados, las temperaturas de inyección y detección son de 310°C, y el gas portador utilizado es helio. A continuación se comentan las características básicas de los métodos empleados.



Figura 4.8. Cromatógrafo de gases

4.4.2.1 Análisis de glicéridos y glicerina

Para el análisis de glicerina y glicéridos se ha empleado una columna capilar Supelco SPB™-1701 (14% cianopropilfenil/86% dimetil)siloxano de dimensiones 15m x 0.25 mm x 0.25 µm. En el análisis de estas muestras se aplica a la columna una rampa de temperatura de 7°C/min desde 120°C hasta 270°C para conseguir la separación de los componentes. Para realizar el análisis, 70 mg de muestra se diluyen en 1 ml de disolución de metilisobutilcetona, que contiene el patrón interno, y se inyecta 1 µl

con un split de 1:120. El patrón interno seleccionado fue el palmitato de metilo por su estabilidad y similitud química con los compuestos analizados. En la figura 4.9 se muestra un cromatograma de una muestra de reacción.

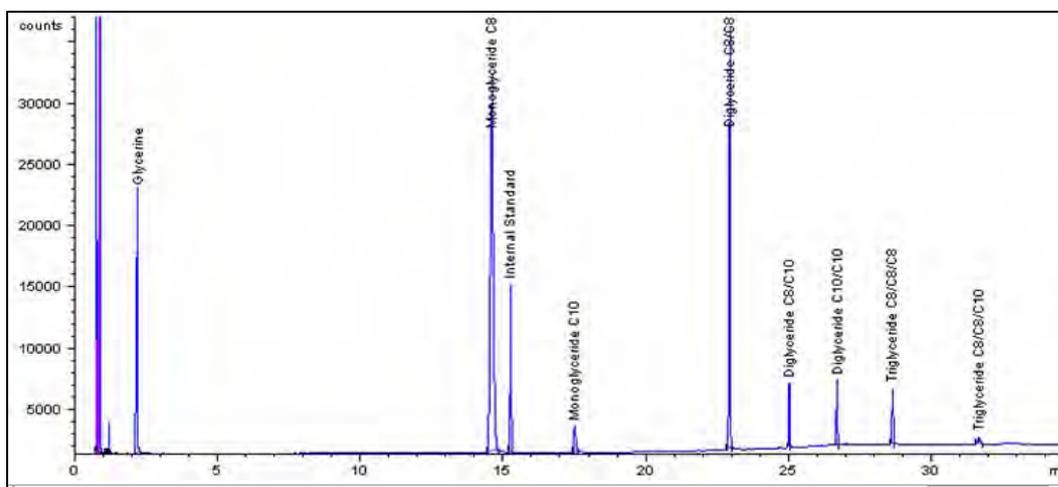


Figura 4.9. Cromatograma de una muestra de reacción entre glicerina y ácidos caprílico y cáprico.

El factor de respuesta de cada analito fue determinado utilizando la relación de áreas entre analito y patrón interno, para disoluciones de concentración conocida de glicerina, monoglicérido de C8, diglicérido C10 y triglicéridos de C8, C9 y C10.

El factor de respuesta del resto de componentes presentes en las muestras y de los cuales no se pudo disponer de producto puro para la preparación de disoluciones patrón (ya sea por su coste elevado o porque no están disponibles en estado puro como diglicéridos y triglicéridos mixtos de C8 y C10), se estimó teniendo en cuenta que para hidrocarburos la señal en un detector FID es proporcional al número de átomos de carbono (ECN) [2]. El factor de respuesta para estos compuestos se obtuvo interpolando su ECN en la gráfica que relaciona el factor de respuesta frente al ECN para los compuestos disponibles con pureza elevada [3].

4.4.2.2 Análisis de 2-etilhexanol

Para la determinación de 2-etilhexanol se ha utilizado una columna capilar Supelco SPB™-5 (5% fenil/95% dimetil)siloxano, de dimensiones 15m x 0.32 mm x 0.25 µm. En este caso se inyecta 1 µl de muestra sin dilución, con un split de 1:100. La temperatura inicial de la columna es de 90°C durante 2 min., posteriormente se aplica una rampa de 10°C/min hasta 280°C, temperatura que se mantiene durante 10 min. Se ha utilizado la técnica del patrón interno, siendo un compuesto que ya posee la muestra, con tiempo de retención de 11,5 min., el que se utilizaba como patrón. Se comprobó que la concentración de esta sustancia era constante en varias muestras correspondientes a diferentes lotes de producción.

La respuesta del 2-etilhexanol en el detector se determinó mediante adición estándar, añadiendo cantidades conocidas de 2-etilhexanol a muestras con bajo contenido de este alcohol.

4.4.2.3 Análisis de isopropanol

Para este análisis la columna empleada es un Supelcowax-10 (fase estacionaria polietilenglicol) de dimensiones 30m x 0.32 mm x 0.25 µm, aplicando un split de 1:175. La temperatura inicial del horno es de 90°C durante 2 min, posteriormente se aplica una rampa de 10°C/min hasta 250°C, temperatura que se mantiene durante 20 minutos. La sustancia utilizada como patrón interno es el 2-etilhexanol. La respuesta del isopropanol en el detector se determinó mediante una recta de calibrado.

4.4.3 Determinación de agua

La determinación de agua se realizó mediante un valorador potenciométrico Karl-Fischer Methrom mod. 716 DMS Titrino (Figura 4.10), utilizando reactivo Karl-Fischer

Hydranal (Riedel-de Haën).



Figura 4.10. Valorador Karl-Fischer.

4.5 Programas para el tratamiento quimiométrico de datos

El tratamiento de los espectros NIR y los datos de referencia se ha realizado mediante los siguientes programas informáticos:

- Vision v2.20 - 2.51 (Foss, Silver Spring, USA). Programa que permite el control del instrumento NIR y el diseño de rutinas de análisis. Ha sido utilizado para el registro de espectros NIR y para la aplicación de los modelos de calibración PLS en el análisis de rutina para el seguimiento de los procesos estudiados.

- The Unscrambler v7.0 - 9.1 (Camo Process, Trondheim, Norway). Específico para el tratamiento de datos multivariados, incorpora los algoritmos de pretratamiento de datos y de calibración multivariable más habituales. Se ha utilizado para la aplicación de pretratamientos espectrales (SNV, MSC y derivadas Savitzky-Golay, entre otros), para el análisis en componentes principales (PCA) y para el desarrollo de modelos de calibración PLS.

- Matlab v6.5 - 7.0 (The Mathworks, MA, USA). Programa base para el desarrollo de códigos o *scripts* que permiten el tratamiento de datos. En nuestro caso, ha sido

utilizado para el desarrollar códigos para el cálculo del NAS (*Net Analyte Signal*) y para la implementación de códigos y/o *software* que utilizan Matlab como base.

➤ PLS Toolbox v3.5 (Eigenvector Research, WA, USA). Conjunto de códigos implementados en entorno Matlab para el tratamiento de datos multivariable. Ha sido utilizado, principalmente, para la aplicación del pretratamiento espectral OSC, entre otros.

➤ Guipro (P. J. Gemperline, East Carolina University) [4]. Conjunto de códigos implementados en Matlab para el tratamiento de datos mediante resolución multivariante de curvas. Este software permite la aplicación de restricciones de igualdad mediante la incorporación de información externa (datos de concentraciones o espectrales). La ventana principal de Guipro, figura 4.11, muestra los pasos a seguir previos al cálculo de MCR, como son, principalmente, la selección de los intervalos espectrales y de tiempo, la selección del número de componentes, y el cálculo de estimaciones iniciales.

De entre estas opciones cabe destacar dos de ellas, la ventana correspondiente a la selección del número de componentes (figura 4.12), donde se puede ver el perfil de los *eigenvectors*, así como los *eigenvalues* correspondientes (*E'* values); y la ventana donde se pueden incorporar restricciones en el cálculo (figura 4.13). En esta última permite incorporar restricciones de forma independiente para cada componente, ya sean restricciones naturales (no-negatividad, unimodalidad, closure) o de igualdad, que permiten importar datos de referencia disponibles (espectros o concentraciones) y utilizarlos durante el proceso iterativo. Además, el software permite regular el grado de dureza con la que se aplican las restricciones de igualdad (0.01 a 20), para valores altos el método se aproxima a una técnica de *hard-modeling*, y para valores bajos a *soft-modeling*.

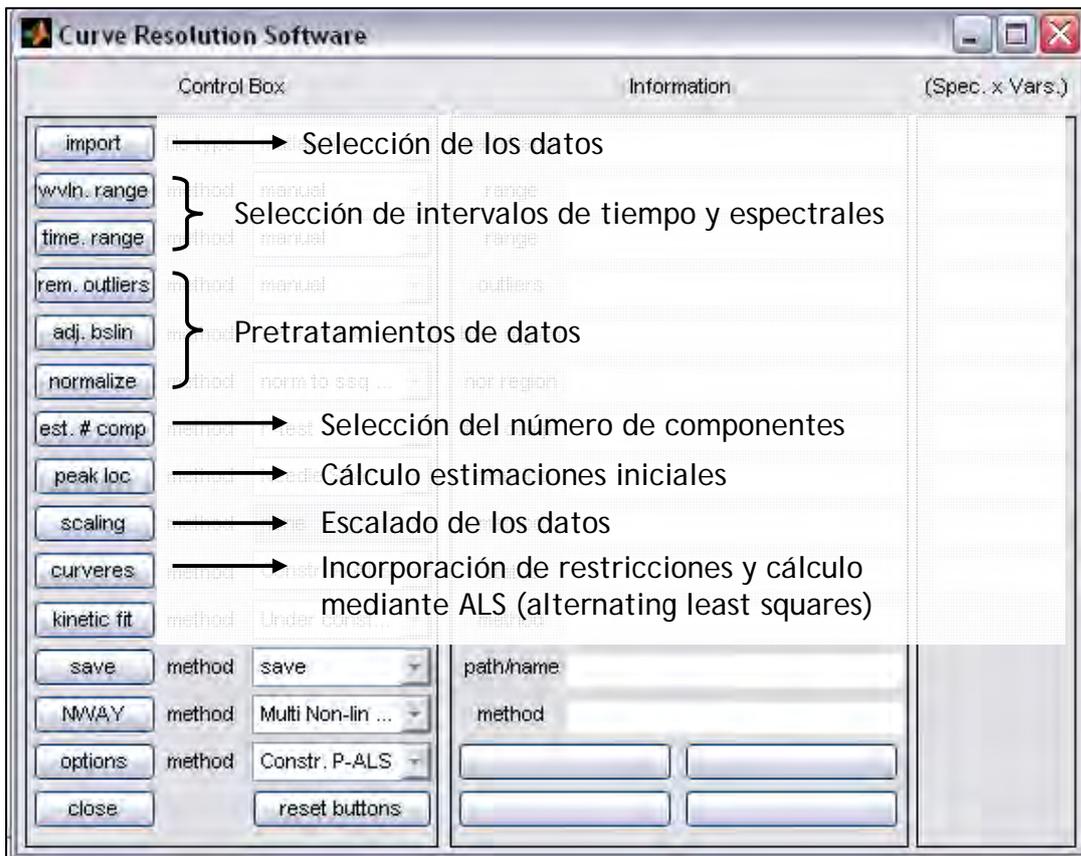


Figura 4.11. Ventana principal del software Guipro con comentarios adjuntos

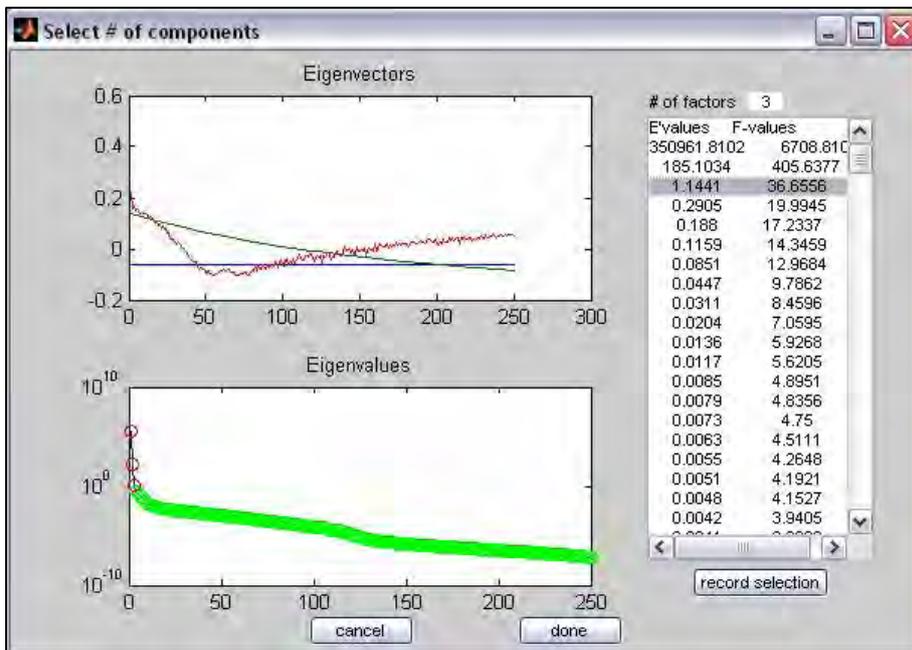


Figura 4.12. Ventana correspondiente a la selección del número de componentes

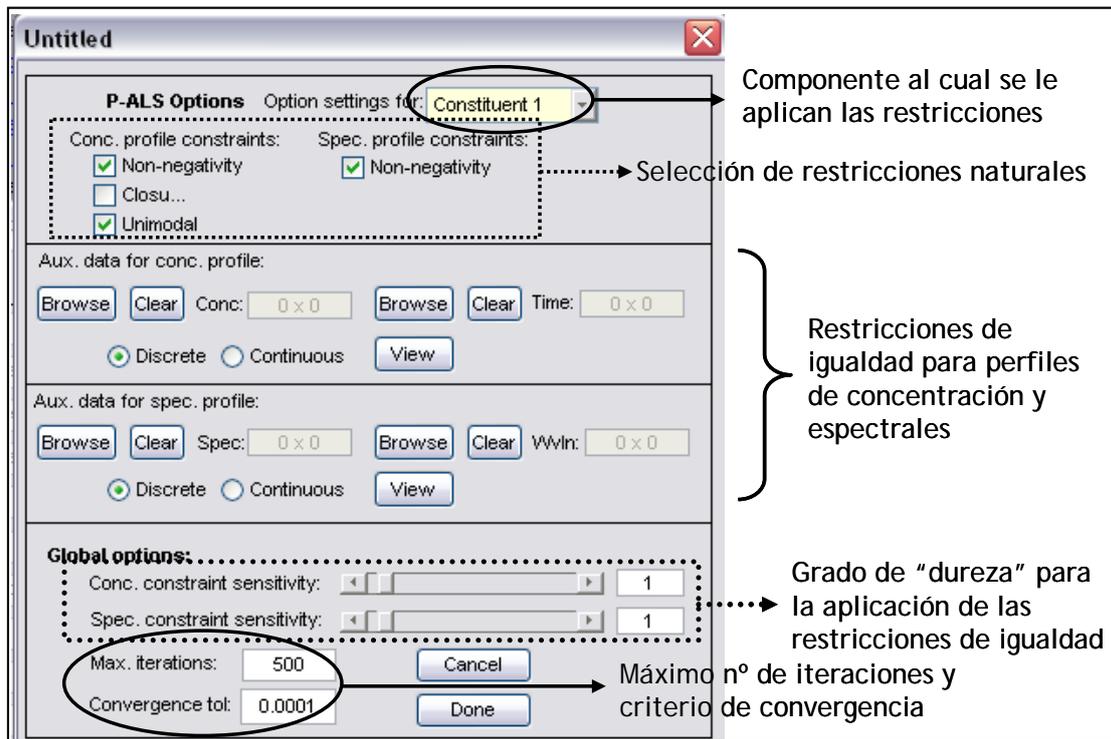


Figura 5.13. Comentarios sobre la ventana para la selección de restricciones.

Referencias

- [1] 'Medium-chain triglycerides' in *European Pharmacopoeia*, 4th Edition, Directorate for the quality of medicines. Strasbourg, France (2002).
- [2] M. Kallai, Z. Veres, J. Balla. Response of flame ionization detectors to different homologous series. *Chromatographia* 54 (2001) 511-517.
- [3] M. Blanco and M. Alcalà. Determinación rápida de ésteres industriales de polioles mediante cromatografía de gases. Pendiente de publicación.
- [4] P. J. Gemperline and E. Cash. Advantages of soft versus hard constraints in self-modeling curve resolution problems. Alternating least squares with penalty functions. *Anal. Chem.* 75 16 (2003) 4236-4243.



5. Resultados y discusión

5. Resultados y discusión

Este capítulo presenta los aspectos más relevantes de los estudios realizados durante el transcurso de la tesis doctoral y que se han agrupado en diferentes apartados para facilitar su comprensión. En el apartado 5.1 se describe el desarrollo de modelos de calibración PLS de la reacción de esterificación entre glicerina y ácidos caprílico y cáprico para la monitorización del proceso y para el control de calidad del producto final. Posteriormente, se presenta la aplicación de MCR como técnica multivariable para análisis cuali- y cuantitativo en el estudio de procesos en evolución, para el seguimiento *at-line* de esta reacción de esterificación; y para el seguimiento *in-line* y el estudio cinético de la esterificación entre isopropanol y ácido mirístico en el apartado 5.2.

En el apartado 5.3 se describe el desarrollo de modelos de calibración PLS para la determinación de bajas concentraciones por NIR, aplicado al proceso de destilación de 2-etilhexanol. En este apartado se aplican dos pretratamientos espectrales, OSC (*Orthogonal Signal Correction*) y NAS (*Net analyte Signal*) con el objetivo de mejorar los resultados obtenidos por PLS. Los resultados son comparados mediante el cálculo de límites de detección estimados a partir de metodologías univariantes y multivariantes.

5.1 Seguimiento *at-line* de un proceso de esterificación: reacción entre glicerina y ácidos caprílico y cáprico

La visualización, análisis y estudio de los espectros NIR registrados es una fase importante y útil del proceso de construcción de modelos multivariantes, ya que permite caracterizar y asociar determinadas regiones espectrales con las variaciones en concentración de los analitos estudiados. Las figuras 5.1 y 5.2 muestran el

conjunto de espectros correspondientes a dos procesos, el primero de ellos realizado con exceso de glicerina (relación molar 1,35:1 glicerina:ácido), y el segundo realizado con exceso de ácido (relación molar 1:3,4 glicerina:ácido).

Las figuras 5.1 y 5.2 muestran la variación de la absorbancia durante la progresión de la reacción. En el intervalo espectral de 1900 - 2200 nm la absorbancia es debida a bandas de combinación del agua y de los grupos hidroxilo presentes en glicerina y glicéridos. La absorbancia en la región 2000 - 2200 nm es debida principalmente a la desaparición de los grupos hidroxilo de la glicerina, y en la región 1900 - 2000 nm debido a la destilación del agua generada y/o contenida por la materia prima. Estas regiones espectrales se corresponden con la máxima absorbancia de estos dos compuestos (figura 5.3). En la región 1400-1650 nm, la absorbancia es debida al primer sobretono del enlace O-H de agua y grupos hidroxilo, donde también se produce una disminución de la absorbancia, pero sin ser tan clara y ordenada como la región anterior.

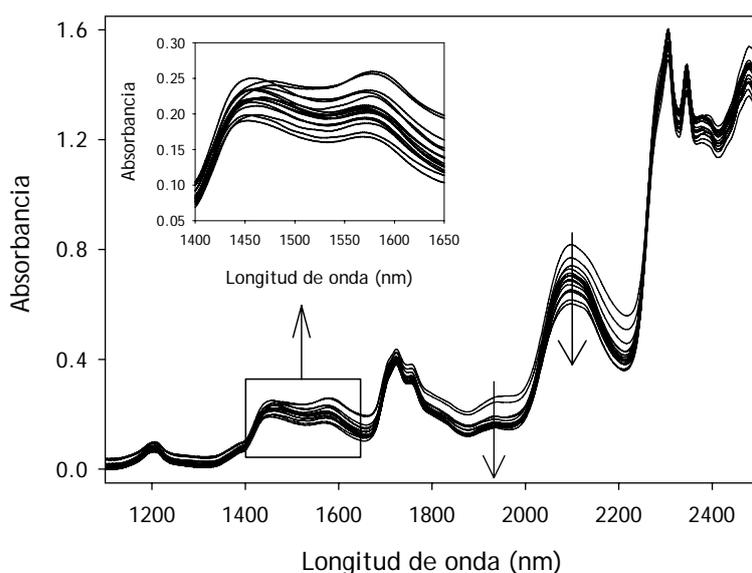


Figura 5.1. Espectros correspondientes a un proceso de esterificación con exceso de glicerina (relación molar 1,35:1 glicerina:ácido).

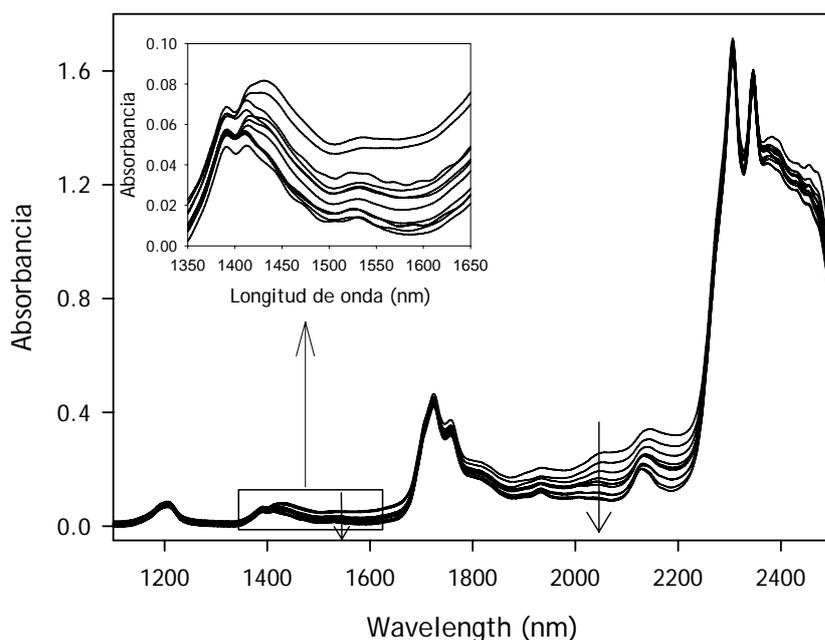


Figura 5.2. Espectros correspondientes a un proceso de esterificación con exceso de ácido (relación molar 1:3,4 glicerina:ácido).

En las regiones 1100 - 1350 nm, 1700 - 1800 nm y 2300 - 2500 nm la absorción es debida al segundo sobretono, primer sobretono y a bandas de combinación, respectivamente, de enlaces C-H.

En la figura 5.3 se recogen los espectros correspondientes a los productos que intervienen en esta reacción: glicerina, mezcla de ácidos grasos (que por su similitud química y espectral serán considerados como un único producto), triglicérido, agua, y el espectro de una mezcla de glicéridos (con una composición aproximada de 60/35/5 % de mono, di y triglicérido respectivamente), ya que no se dispone de mono y diglicérido con un grado de pureza elevado.

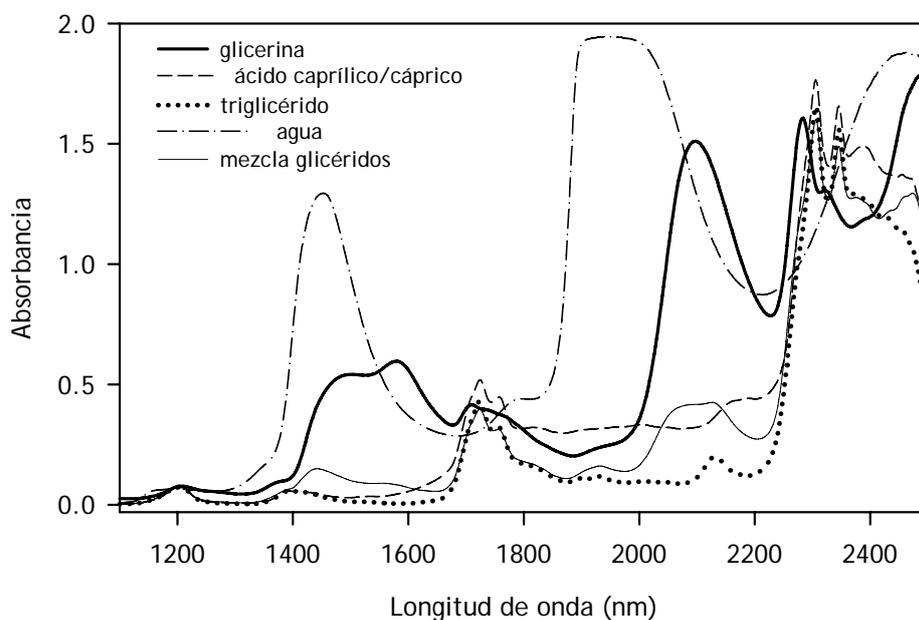


Figura 5.3. Espectros de los compuestos implicados en la esterificación entre glicerina y ácidos caprílico y cáprico.

5.1.1 Determinación de índice de ácido, índice de hidroxilo y agua mediante calibración PLS

Las reacciones de esterificación son frecuentemente monitorizadas en la industria química mediante índice de ácido (IA) e índice de hidroxilo (IOH), que son una medida de la cantidad de ácido o grupos hidroxilo sin reaccionar. Estos dos parámetros sirven para determinar del punto final de la reacción de esterificación.

La determinación de estos dos parámetros se realiza mediante procedimientos lentos y laboriosos, e implican un consumo de reactivos y solventes de toxicidad y coste elevados (especialmente en el caso del IOH). Además, reducen la capacidad de producción debido a la espera en la obtención del resultado.

Para el desarrollo de modelos de calibración PLS, se han extraído muestras de varios procesos industriales realizados en la planta de Uniqema (ICI Spain) y de dos procesos a escala de laboratorio (reactor LabMax) realizados en nuestros laboratorios, ambos correspondientes a la esterificación con exceso de ácido para la obtención de triglicérido (relación molar 1:3,4 glicerina:ácido). Los valores de referencia han sido obtenidos mediante la aplicación de los métodos comentados en el apartado de experimental para índice de ácido e índice de hidroxilo, y mediante Karl-Fischer para la determinación de agua.

El conjunto de muestras ha sido dividido en dos subconjuntos de calibración y predicción, de forma que ambos tuvieran muestras de ambas procedencias, industriales y de laboratorio, y que cubriesen el intervalo de concentración correspondiente de cada uno de los parámetros determinados. En la tabla 5.1 se muestra los pretratamientos e intervalos espectrales ensayados, el número de factores PLS que ofrecieron mejores resultados, así como los parámetros de la regresión lineal entre los valores NIR y los de referencia para los tres parámetros estudiados.

De los resultados de la tabla 5.1 cabe destacar que para agua y IOH el rango espectral 1800 - 2100 nm es el que ofrece mejores resultados, ya que es donde se encuentran las bandas más intensas para estos dos parámetros. Como indican las figuras de mérito y los RSEC/P (%), los valores NIR obtenidos son consistentes con los valores de referencia. Los RSEC/P(%) obtenidos para el agua son superiores al resto debido al bajo nivel de concentración de este analito, mientras que el hecho de que el RSEC/P(%) del IOH sea superior al del IA puede ser debido a la precisión de los respectivos métodos de referencia.

Parámetro	Índice de ácido	Índice de hidroxilo	Agua
Pretratamiento	SNV	SNV	1ª Derivada
Intervalo espectral (nm)	1100 - 2500	1800 - 2100	1800 - 2100
nº factores PLS	2	3	3
Calibración			
Intervalo concentración	1 - 150 mg KOH/g	2 - 117mg KOH/g	0,01 - 0,84 %
nº muestras	31	31	25
Pendiente	0,999 ± 0,012	0,995 ± 0,027	0,995 ± 0,029
Ordenada origen	0,06 ± 0,90	0,2 ± 1,3	0,001 ± 0,008
R ²	0,999	0,995	0,995
RSEC (%)	2,2	5,4	5,3
Predicción			
Intervalo concentración	1,5 -146 mg KOH/g	3 - 116 mg KOH/g	0,02 - 0,80 %
nº muestras	53	49	40
Pendiente	1,009 ± 0,009	1,02 ± 0,02	0,99 ± 0,04
Ordenada origen	-0,02 ± 0,57	0,2 ± 0,8	0,004 ± 0,006
R ²	0,999	0,995	0,990
RSEP (%)	2,5	6,0	9,5

Tabla 5.1. Figuras de mérito de los modelos de calibración PLS para la determinación de IA, IOH y agua.

Los parámetros recogidos en la tabla muestran que los diferentes modelos tienen una adecuada capacidad predictiva y están exentos de errores sistemáticos.

La aplicación de la espectroscopia NIR junto con la regresión por mínimos cuadrados parciales, permite el desarrollo de una metodología at-line que no requiere pretratamiento de muestra ni uso de reactivos, obteniendo el valor de IA e IOH únicamente a partir del registro espectral NIR.

5.1.2 Determinación de glicerina, ácido, mono, di, y triglicéridos y agua mediante PLS

A pesar de que algunos procesos de esterificación son monitorizados en la industria mediante índice de ácido (IA) e índice de hidroxilo (IOH), estos dos parámetros no siempre proporcionan la información necesaria para conocer el estado de la reacción. En el proceso de esterificación con exceso de glicerina (1,35:1 glicerina:ácido) el objetivo es obtener un producto con una proporción de mono, di y triglicérido dentro de unas especificaciones predeterminadas. En este caso, tanto IA como IOH no son adecuados para la monitorización de la reacción ya que no ofrecen información acerca de la composición, y no permiten determinar cuando se ha alcanzado la composición deseada.

Para monitorizar adecuadamente el proceso con exceso de glicerina es necesario determinar la composición de la mezcla de reacción, desarrollando un modelo de calibración NIR para cada uno de los analitos de interés: glicerina, ácido, agua, monoglicérido, diglicérido y triglicérido. Las muestras necesarias para desarrollar los modelos de calibración y para la predicción de los mismos se obtuvieron de 6 procesos de laboratorio y 1 industrial, y también a partir de la preparación de muestras en el laboratorio (epígrafe 4.2.5.1). Para las muestras de proceso, la concentración de ácido fue determinada por valoración (IA), la de agua mediante Karl-Fischer, y la de glicerina y glicéridos por cromatografía de gases.

La preparación de muestras en el laboratorio por dopaje con productos puros, y los procesos realizados en el laboratorio con diferente proporción de reactivos, permiten obtener un conjunto de muestras que cubran el rango de concentraciones necesario para cada componente y también reducir la correlación existente entre las concentraciones de los compuestos involucrados en la reacción. Una elevada

correlación entre concentraciones daría lugar a modelos de calibración PLS poco robustos a causa de la gran similitud espectral de los compuestos presentes.

Un total de 81 muestras fue dividido en dos conjuntos de calibración y predicción para cada compuesto a determinar. En la construcción de los modelos de calibración se ha ensayado diferentes rangos de longitudes de onda y pretratamientos espectrales. Los modelos PLS1 que proporcionaron mejores figuras de mérito para la regresión lineal de los valores NIR frente a los valores de referencia, así como los RSEC/P (%), se muestran en la tabla 5.2 para glicerina, ácido y agua, y en la tabla 5.3 para los glicéridos.

Los mejores modelos han sido obtenidos en SNV y en primera derivada. Para glicerina y agua se han obtenido buenos resultados, tanto en calibración como en predicción, con modelos de calibración con 3 factores PLS, y valores de RSEP (%) inferiores al 5%. Los resultados de calibración y predicción para ácido también son buenos, sin embargo, en este caso el RSEP (%) es superior debido a que la mayoría de muestras de predicción tienen un contenido en ácido inferior al 1 %.

Parámetro	Glicerina	Ácido	Agua
Pretratamiento	1ª Derivada	SNV	SNV
Intervalo espectral (nm)	1100 - 2150	1100 - 2500	1100 - 2500
Nº factores PLS	3	4	3
Calibración			
Intervalo concentración (%)	0,5 - 41,5	0,1 - 31,3	0,1 - 4,4
nº muestras	38	44	26
Pendiente	0,99 ± 0,04	1,00 ± 0,02	0,99 ± 0,02
Ordenada origen	0,2 ± 0,8	0,0 ± 0,2	0,00 ± 0,03
R ²	0,988	0,996	0,998
RSEC (%)	3,8	4,8	3,6
Predicción			
Intervalo concentración (%)	14,2 - 29,5	0,1 - 24,4	0,2 - 4,3
nº muestras	26	35	27
Pendiente	1,0 ± 0,1	1,04 ± 0,03	1,01 ± 0,02
Ordenada origen	0 ± 3	0,0 ± 0,02	0,00 ± 0,04
R ²	0,93	0,988	0,997
RSEP (%)	4,9	11,3	4,2

Tabla 5.2. Figuras de mérito de los modelos de calibración PLS1 para la determinación de glicerina, ácido y agua.

Los modelos para la determinación de glicéridos son más complejos debido a la similitud espectral existente entre ellos, sobretodo entre mono y diglicérido, que requieren 5 y 4 factores PLS respectivamente. A pesar de que el triglicérido es el que muestra más diferencias espectrales, los modelos para este componente son los que proporcionan peores resultados debido a su bajo nivel de concentración. Los mejores modelos para glicéridos se obtienen eliminando la región de absorción de agua y glicerina (1850-2150 nm).

Parámetro	Monoglicérido	Diglicérido	Triglicérido
Pretratamiento	1ª Derivada	SNV	SNV
Intervalo espectral (nm)	1100 - 1850, 2150 - 2400	1100 - 1850, 2150 - 2400	1100 - 2500
nº factores PLS	5	4	5
Calibración			
Intervalo concentración (%)	23,4 - 60,4	4,5 - 37,7	0,1 - 5,9
nº muestras	37	39	30
Pendiente	0,94 ± 0,07	0,96 ± 0,07	0,95 ± 0,08
Ordenada origen	2 ± 3	1 ± 2	0,1 ± 0,2
R ²	0,94	0,96	0,95
RSEC (%)	4,5	3,7	8,2
Predicción			
Intervalo concentración (%)	32,1 - 51,2	7,8 - 35,8	0,3 - 5,8
nº muestras	26	29	37
Pendiente	1,0 ± 0,2	0,9 ± 0,1	0,9 ± 0,1
Ordenada origen	1 ± 9	2 ± 3	0,1 ± 0,4
R ²	0,8	0,9	0,8
RSEP (%)	5,2	6,5	17,6

Tabla 5.3. Figuras de mérito de los modelos de calibración PLS1 para la determinación de los glicéridos.

Como se ha comentado anteriormente, para esta reacción la determinación del índice de ácido o hidroxilo no aportan información de interés. En la figura 5.4 se muestra la evolución de concentraciones de dos procesos reactivos con proporción 1,35:1 glicerina:ácido. A partir de 3 horas de reacción el perfil del ácido se muestra constante y muy próximo a cero; consecuentemente, también será constante el índice de hidroxilo. Sin embargo, se observa una evolución de los perfiles de reacción para los glicéridos que se explica por la coexistencia de reacciones de

transesterificación que producen un pequeño incremento en la concentración de monoglicérido y una leve disminución en la de di y triglicéridos.

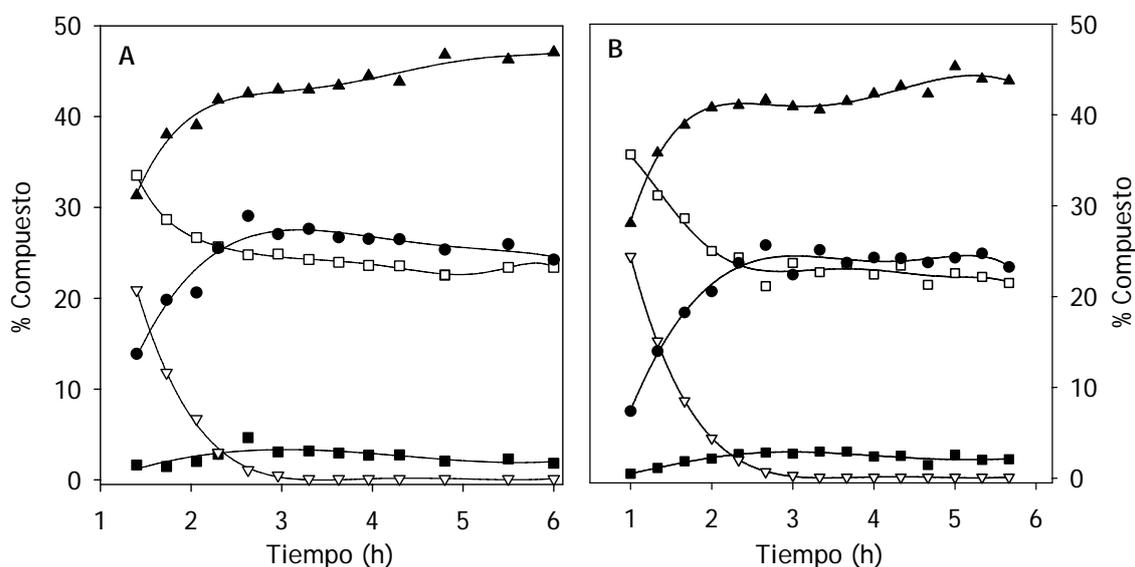


Figura 5.4. Perfiles de concentración para dos procesos con exceso de glicerina. (□) glicerina, (∇) ácido, (▲) monoglicérido, (●) diglicérido y (■) triglicérido.

Estas pequeñas variaciones en las concentraciones de glicéridos conducen a que el producto de la reacción pueda cumplir las especificaciones definidas para este producto: una vez eliminada la glicerina que no ha reaccionado, el contenido en monoglicérido ha de ser superior al 61% y el de triglicérido inferior al 4%. La figura 5.5 muestra la evolución relativa de los glicéridos obtenida por NIR para los dos procesos de la figura 5.4. A partir de las 3 horas de reacción, cuando ha reaccionado completamente el ácido, el producto no cumple las especificaciones para mono y triglicérido. Sin embargo, durante las 2 horas posteriores el contenido de monoglicérido es superior al 61%, y el de triglicérido llega a ser inferior al 4%.

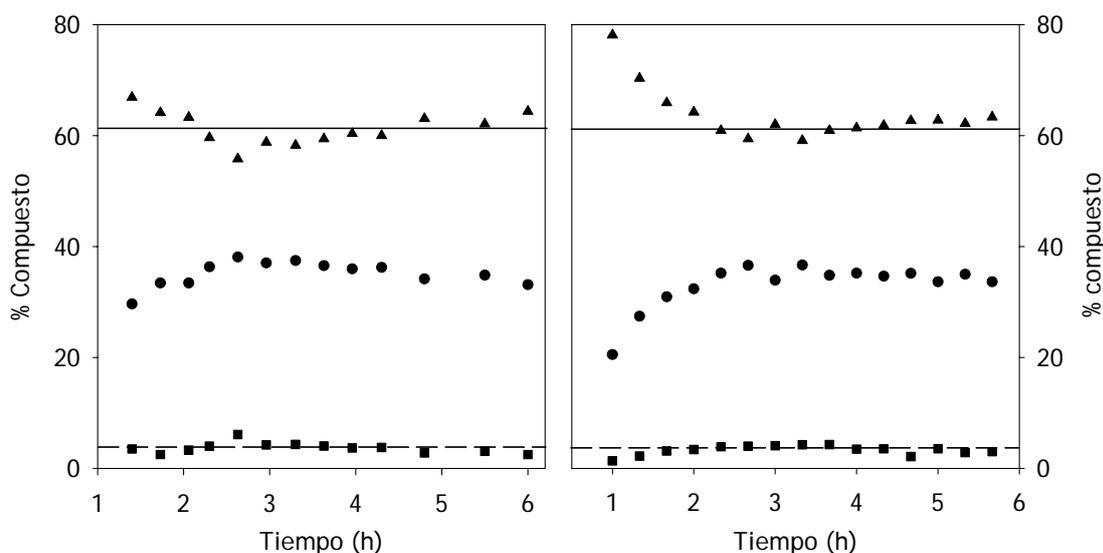


Figura 5.5. Evolución de la concentración relativa de glicéridos para los procesos con exceso de glicerina. (▲) monoglicérido, (●) diglicérido, (■) triglicérido, (—) concentración mínima de monoglicérido, (—) concentración máxima de triglicérido.

5.1.3 Aplicación de la Resolución Multivariante de Curvas a la reacción de esterificación entre glicerina y ácidos caprílico y cáprico

La aplicación de técnicas de calibración multivariable usuales (MLR, PCR, PLS) requiere el conocimiento del valor de la propiedad a determinar para cada muestra, lo que obliga a realizar mucho trabajo experimental previo a la construcción de modelos de calibración. El desarrollo de nuevas técnicas quimiométricas de análisis multivariable, como la resolución multivariante de curvas, permite la construcción de modelos de calibración con menos trabajo experimental gracias a la flexibilidad de esta técnica. El procedimiento iterativo aplicado en la resolución MCR permite la incorporación de la información disponible del sistema en forma de restricciones que facilitan la convergencia del cálculo iterativo, evitando la necesidad de disponer de las concentraciones de todas las especies a determinar correspondientes a las muestras necesarias para la construcción del modelo.

El objetivo de este trabajo es, en una primera etapa, la determinación de los perfiles de concentración de las especies involucradas en la esterificación con exceso de glicerina, y en un segundo paso, la determinación de los índices de ácido e hidroxilo para la esterificación con exceso de ácido.

Una etapa fundamental en la aplicación de cualquier algoritmo de resolución es el establecimiento y determinación del número de componentes que pueden ser monitorizados. La herramienta utilizada en nuestro caso es la descomposición en valores singulares (SVD), teniendo en cuenta los valores propios, considerando significativos los componentes con un valor propio superior a 1, y la representación gráfica de los vectores propios, cuyo perfil puede indicar la presencia o no de un componente. Para una adecuada aplicación de MCR, el rango matemático estimado mediante SVD ha de ser igual al número de componentes que se van a determinar.

Los datos espectrales NIR han sido pretratados mediante *standard normal variate* (SNV) para eliminar desplazamientos de línea base o *scattering* causados por la turbidez de algunas muestras. Se escogió este pretratamiento espectral porque ha sido uno de los que mejores resultados ofreció en la aplicación de PLS. El problema generado por este pretratamiento es que la mitad del espectro adquiere valores de intensidad negativos, lo que impide la aplicación de restricciones de no negatividad espectral en el cálculo iterativo de MCR. Para solventar este problema, se desplazaron los espectros a valores de intensidad positivos sumando un escalar (2) a cada uno de los elementos de la matriz de datos espectrales SNV, comprobándose que no afectaba a la aplicación de MCR-ALS debido a que no se modifica el perfil espectral.

5.1.3.1 Determinación de glicerina, ácido, mono, di, y triglicéridos

El objetivo es estudiar la capacidad de MCR-ALS para el estudio de los perfiles de concentración de las especies involucradas en la reacción con exceso de glicerina (1,35:1 glicerina:ácido). El rango matemático, calculado mediante SVD, de las matrices estudiadas correspondientes a los procesos de laboratorio es de 2. Incluso el rango para la matriz aumentada que incluye los espectros de 3 procesos también es 2. Este valor se corrobora observando los vectores propios: los correspondientes a los dos primeros valores singulares muestran una evolución similar para los tres procesos estudiados, mientras que el tercero no muestra una tendencia clara ni similar en ninguno de los tres procesos estudiados.

A pesar de que el rango de las matrices sea 2, el sistema químico estudiado consta de 6 componentes: glicerina, mezcla de ácidos caprílico y cáprico (se considera como un único componente), monoglicérido, diglicérido, triglicérido y agua. Esta deficiencia de rango está causada por diversos motivos: el agua está presente en baja concentración debido a que la temperatura de reacción es elevada, lo que produce la destilación del agua generada por la reacción de esterificación; el triglicérido es un componente minoritario ya que se trabaja en exceso de glicerina; y finalmente, mono y diglicéridos son dos compuestos de gran similitud química, lo que implica que sus espectros NIR serán muy similares, y por tanto serán dos compuestos de difícil diferenciación.

En un sistema con deficiencia de rango como el actual es necesario incorporar información externa, como espectros puros de los compuestos a determinar en forma restricciones de igualdad. Así, utilizando la matriz de espectros correspondientes a 16 muestras de un proceso de esterificación 1,35:1 glicerina:ácido, aplicando restricciones de no-negatividad en concentraciones y espectros, de unimodalidad en

espectros, y utilizando los espectros puros de glicerina y mezcla de ácidos como restricciones de igualdad, se resuelve el sistema obtenido los perfiles de la figura 5.6.

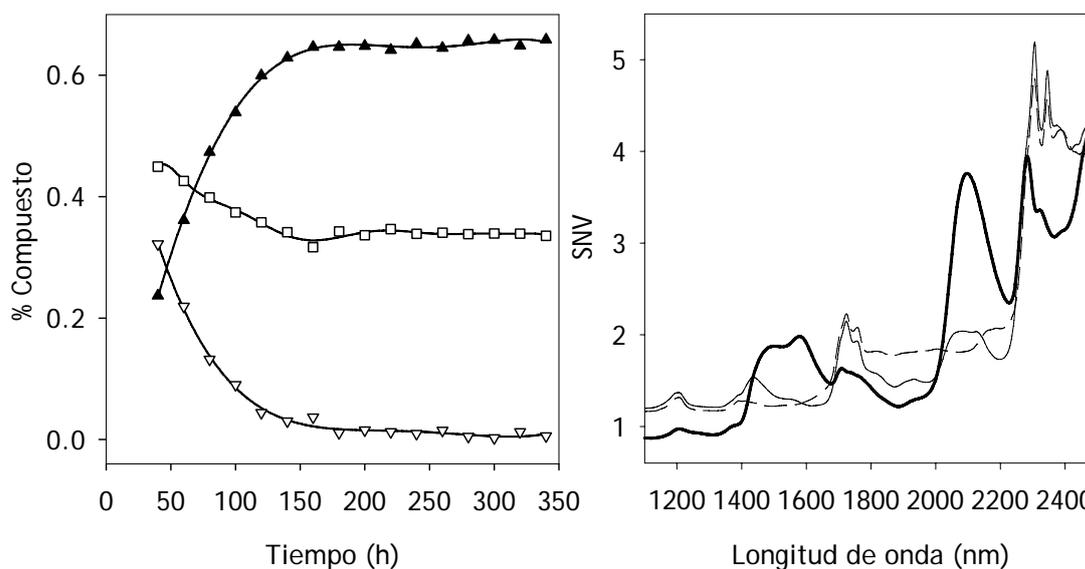


Figura 5.6. Perfiles de concentración (izq.) y espectrales (der.) obtenidos por MCR-ALS con 3 componentes: (\square) y ($\text{—}\blacktriangle\text{—}$) glicerina, (∇) y (---) ácido, (\blacktriangle) y (—) mezcla de glicéridos.

Los perfiles de la figura 5.6 muestran un comportamiento acorde con nuestro conocimiento del sistema: el componente 1 puede ser perfectamente asignado al ácido ya que su perfil desciende hasta prácticamente cero (reactivo en defecto), el componente 2 se atribuye a glicerina, su perfil desciende y se estabiliza entorno al 30 por ciento (es el reactivo en exceso) y el tercer componente indica la formación de los glicéridos, que por la similitud espectral entre ellos, están incluidos en un solo componente. La varianza capturada es del 99.99 % y el coeficiente de correlación entre los espectros calculados y los originales es superior a 0.999, tanto para el glicerol como para la mezcla de ácidos.

Se ha demostrado que es posible resolver un sistema deficiente de rango utilizando los espectros puros como restricciones de igualdad, con lo que nuestro objetivo ahora es conseguir la diferenciación de los glicéridos, que en el cálculo anterior aparecen en forma de un solo componente. Sin embargo, únicamente se dispone del espectro puro del triglicérido, por lo que no será posible diferenciar entre mono y diglicérido. Así, en este caso se plantea la estrategia de calcular en un primer paso los espectros de mono y diglicérido mediante MCR, y posteriormente, utilizar los espectros calculados, junto con los espectros puros registrados, para predecir las concentraciones de las especies involucradas en la reacción.

Para calcular los espectros de mono y diglicérido se ha utilizado una matriz de espectros correspondientes a 16 muestras de un proceso de esterificación 1,35:1 glicerina:ácido, de la cual se dispone de los valores de concentración de glicerina y glicéridos (determinados por cromatografía de gases) y de ácido (determinado por valoración). En este cálculo se han aplicado restricciones de unimodalidad en concentraciones, de no negatividad en concentraciones y espectros y se han considerado 5 componentes: glicerina, mezcla de ácidos y los tres glicéridos: mono, di y triglicérido. Como información adicional en forma de restricciones de igualdad se utilizaron los espectros registrados de glicerol, mezcla de ácidos y triglicérido, además de las concentraciones de los 5 compuestos. Los únicos parámetros no aportados son los espectros de mono y diglicérido. Los perfiles de concentración obtenidos, así como los espectros calculados se muestran en la figura 5.7.

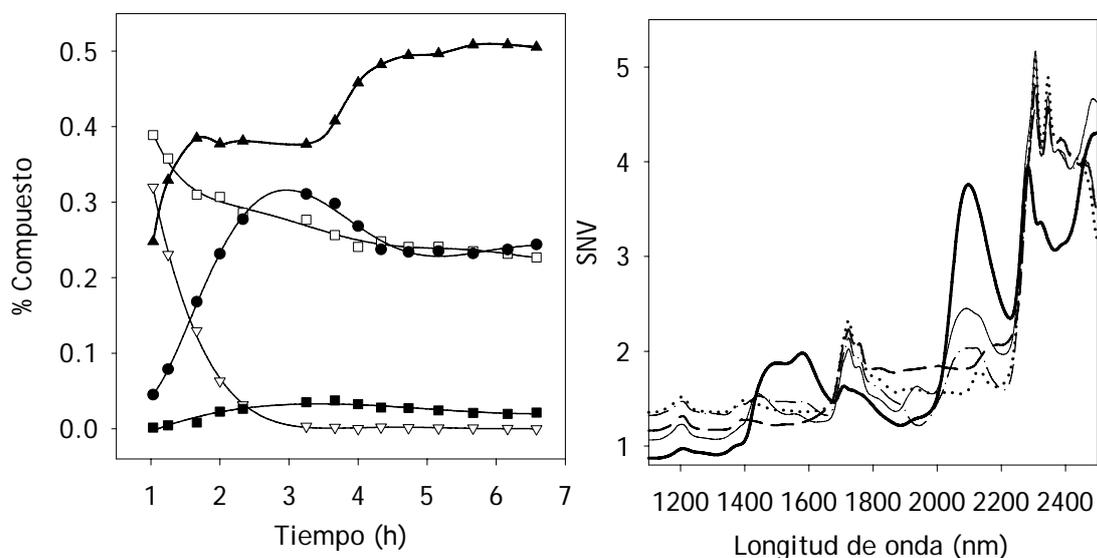


Figura 5.7. Perfiles de concentración (izq.) y espectrales (der.) obtenidos por MCR-ALS con 5 componentes: (□) y (—) glicerina, (▽) y (—) ácido, (▲) y (—) monoglicérido, (●) y (— · —) diglicérido, (■) y (· · · · ·) triglicérido.

Los perfiles espectrales estimados para mono y diglicéridos muestran una gran similitud, siendo la diferencia más notoria la de la región 2000-2200 nm donde el monoglicérido tiene una mayor absorptividad, que se explica por el mayor número de grupos -OH por molécula. Los coeficientes de correlación entre los espectros calculados y los registrados son superiores a 0.999 para glicerol, mezcla de ácidos y triglicérido.

Una vez se dispone de los espectros puros para las especies que se desean determinar, es posible solventar la deficiencia de rango inicial que impedía la aplicación de MCR, y utilizar esta técnica para determinar los perfiles de concentración. De esta forma, se aplicó MCR-ALS a dos nuevos procesos de esterificación con proporción 1,35:1 glicerina:ácido, aplicando las usuales restricciones de no negatividad en concentraciones y espectral, la restricción de unimodalidad en concentraciones, además de utilizar los espectros puros de

glicerina, ácido, monoglicérido y diglicérido como restricciones de igualdad. Los perfiles de concentración obtenidos se muestran en la figura 5.8.

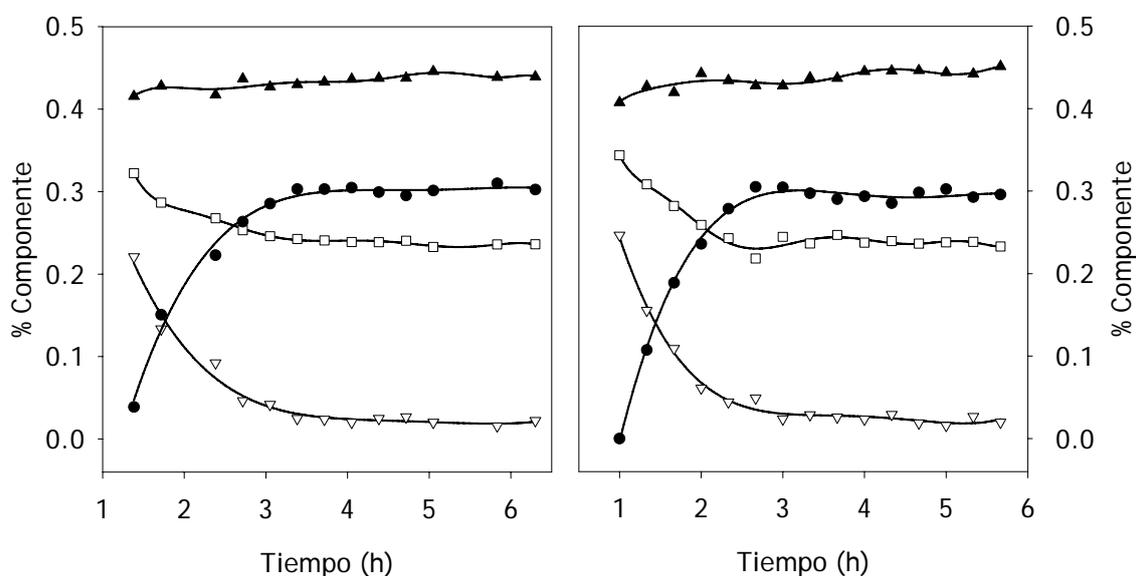


Figura 5.8. Perfiles de concentración obtenidos por MCR-ALS para dos procesos con exceso de glicerina. (□) glicerina, (▽) ácido, (▲) monoglicérido y (●) diglicérido.

Los perfiles de la figura 5.8 son similares a los obtenidos por PLS para estos dos procesos, mostrados en la figura 5.4. El ácido disminuye su concentración hasta alcanzar un contenido residual entorno a 150-200 minutos. El contenido en glicerol se estabiliza alrededor del 25% en el mismo intervalo de tiempo. Las únicas diferencias remarcables se encuentran en el contenido de mono y diglicérido de las 3 primeras muestras de reacción de ambos procesos; sin embargo, estos dos componentes muestran un perfil de reacción muy similar con los de la figura 5.4. De este modo, utilizando únicamente los espectros calculados y registrados en la iteración por ALS es posible la estimación de los perfiles de reacción para los analitos de interés en la reacción de esterificación entre glicerol y la mezcla de ácidos caprílico y cáprico.

5.1.3.2 Determinación de índice de ácido e índice de hidroxilo

El objetivo de este trabajo es estudiar la capacidad para el análisis cuantitativo de MCR-ALS, aplicado al proceso de esterificación con exceso de ácido (1:3,4 glicerina:ácido). Para este proceso se dispone de los valores de referencia de IA e IOH, por lo que éstos serán los parámetros a determinar mediante MCR, con la peculiaridad que se calcularán a partir de las concentraciones de las correspondientes especies químicas estimadas por MCR: mezcla de ácidos para el IA y glicerina y glicéridos para el IOH. El procedimiento para desarrollar el modelo de calibración será similar al desarrollado con PLS, del conjunto inicial de muestras, se crearán dos *sets*, uno para calibración, y otro para predicción, estudiando el número de componentes que se pueden extraer de los datos, y aplicando MCR utilizando los perfiles espectrales y el mínimo de información cuantitativa necesaria para resolver la ambigüedad de intensidad. Con los perfiles de concentración obtenidos se calcularán los IA e IOH, que serán comparados con los obtenidos por valoración. Con el modelo obtenido se estimarán los IA e IOH del conjunto de predicción mediante las ecuaciones 3.27 y 3.28; resultados que también serán comparados con los obtenidos por valoración.

El rango calculado mediante *singular value decomposition* (SVD) para el conjunto de espectros de un proceso es de 2 (valores propios $3,8 \cdot 10^4$, 10,2 y 0,05), inferior al teórico rango químico de 6, correspondiente a glicerol, mezcla de ácidos, mono, di y triglicérido y agua. Esta deficiencia de rango es debida a la falta de selectividad espectral y a la correlación de las concentraciones entre los diferentes componentes. Para solventar la deficiencia de rango se estudia la utilización de matrices aumentadas (apartado 3.4.3.2) con los espectros puros de los compuestos a determinar. Este procedimiento es diferente al anteriormente utilizado, consistente en la utilización de los espectros puros como restricciones de igualdad durante el

proceso iterativo. La utilización de matrices aumentadas con los espectros puros es una metodología más aceptada, ya que da mayor libertad al sistema durante el procedimiento iterativo *alternating least squares*.

Para determinar el número de componentes se utilizó una matriz correspondiente a los espectros de un proceso aumentada con los 6 espectros puros de glicerina, ácido, mono, di y triglicérido y agua. El rango matemático de esta matriz es de 6 (valores propios de $5,8 \cdot 10^4$, 751, 261, 21, 15, 1,9 y 0,3), lo que permite en principio extraer 6 componentes químicos. Aplicando MCR-ALS utilizando únicamente la restricción de no-negatividad espectral y en concentraciones, se obtienen los perfiles de concentración de la figura 5.9.

Según la figura 5.9, la concentración de agua durante el proceso reactivo es muy baja (inferior a 0,1 %). De forma similar, el diglicérido muestra un perfil de concentraciones errático y próximo a cero. Este hecho induce a pesar que estos dos componentes no son significativos y que su inclusión en el cálculo podría distorsionar los perfiles del resto de componentes: la fracción de componente puro es inferior a 1 en el resultado del cálculo MCR como se observa en la parte derecha de la figura 5.9 para casi todos los componentes, lo que indica que el ajuste obtenido no es correcto.

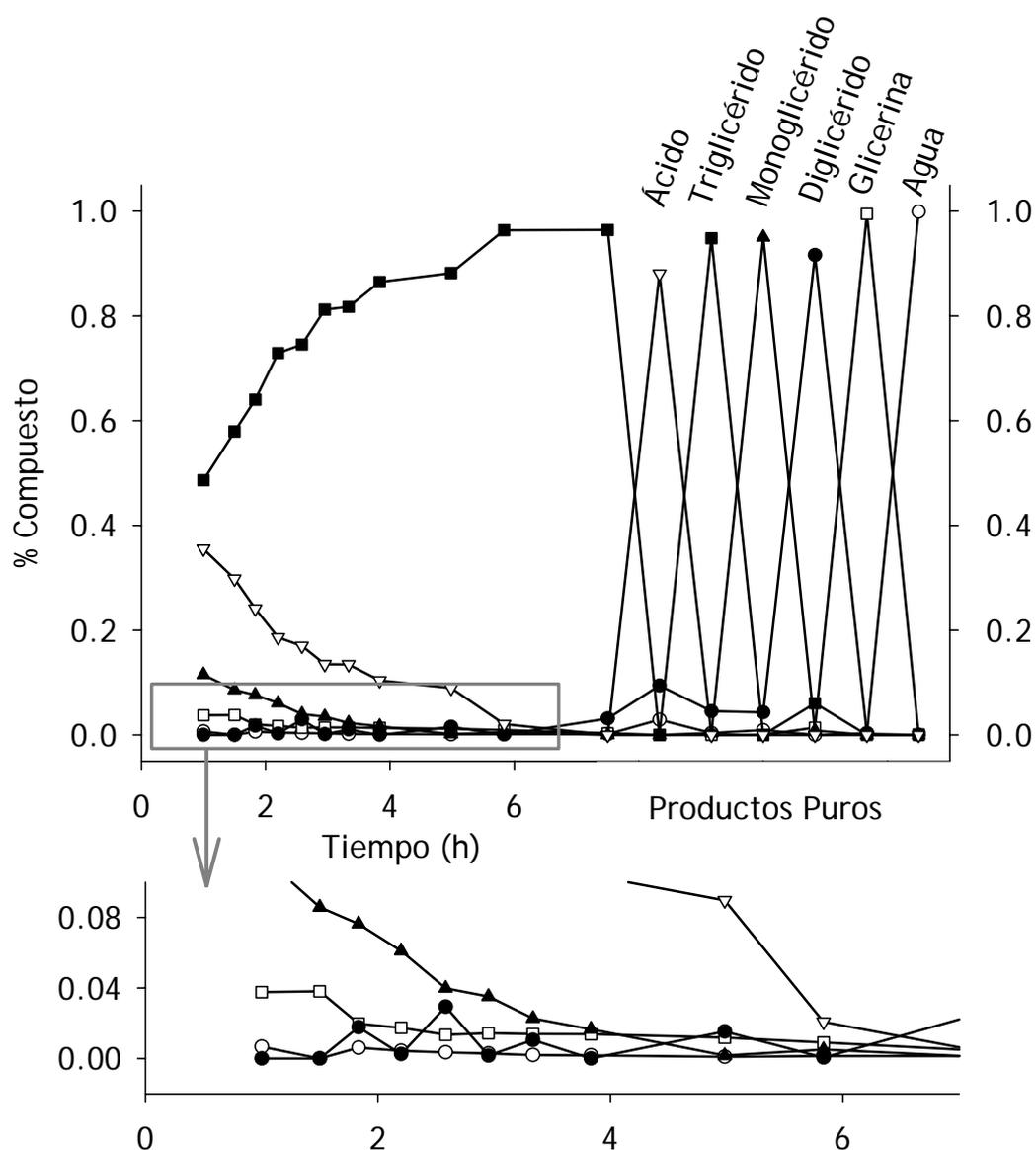


Figura 5.9. Perfiles de concentración obtenidos por MCR-ALS para un proceso con exceso de ácido, aumentado la matriz con los espectros puros. (∇) ácido, (\blacksquare) triglicérido, (\blacktriangle) monoglicérido, (\bullet) diglicérido, (\square) glicerina y (\circ) agua.

La aplicación de MCR-ALS aumentando la misma matriz de espectros únicamente con los 4 componentes significativos: glicerina, ácido, monoglicérido y triglicérido, conduce a un rango, determinado mediante SVD, de 4 (valores propios de $9,7 \cdot 10^4$, 416, 28, 16 y 0,6). Aplicando MCR-ALS a esta matriz aumentada, únicamente con la restricción de no-negatividad, conduce a los perfiles de concentración mostrados en la figura 5.10.

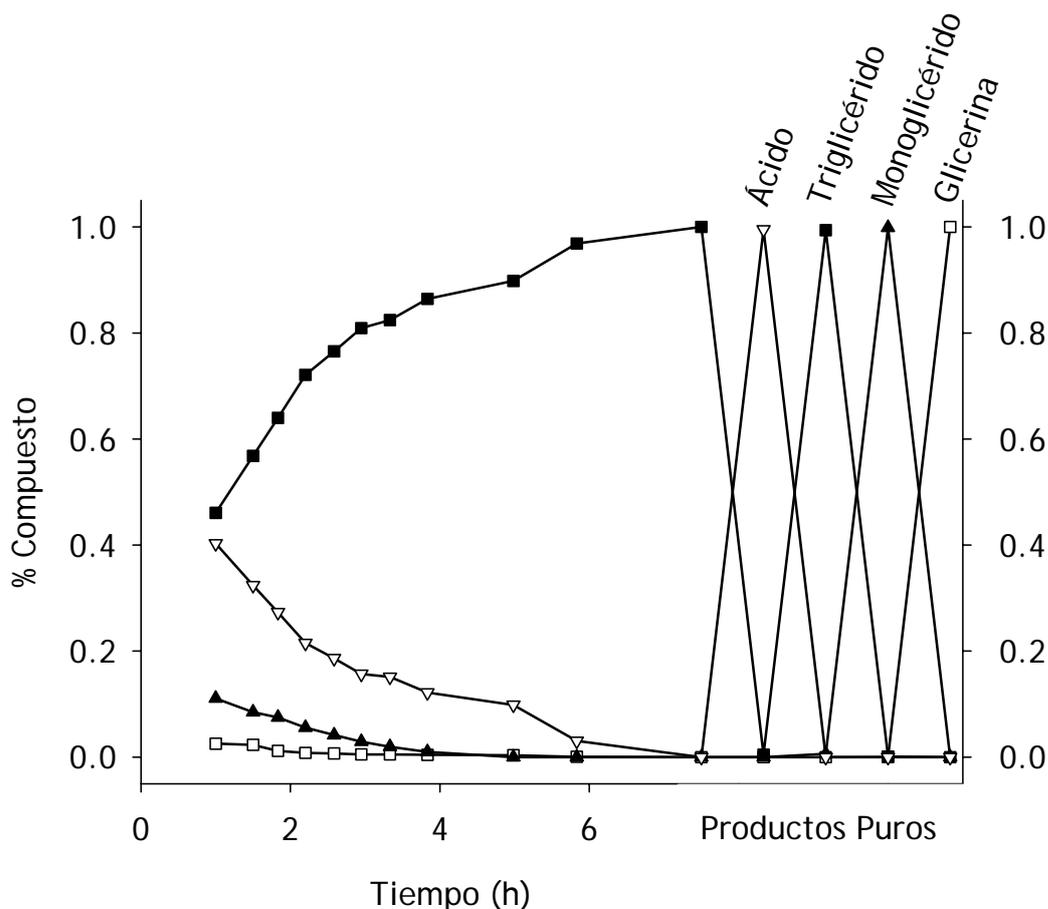


Figura 5.10. Perfiles de concentración obtenidos por MCR-ALS para un proceso con exceso de ácido, aumentado la matriz con los espectros puros de: (∇) ácido, (\blacksquare) triglicérido, (\blacktriangle) monoglicérido y (\square) glicerina.

Estos perfiles (figura 5.10) son más definidos y suaves que los de la figura 5.9, y la fracción de cada componente es prácticamente uno para todos los compuestos puros. Los perfiles de glicerina y monoglicérido muestran que su concentración se aproxima a 0 después de 4 horas de reacción, mientras que el contenido en ácido es aún elevado (entorno a un 15%) debido a que es el reactivo en exceso. A partir de las 5 horas de reacción el contenido en ácido disminuye hasta ser prácticamente 0 debido a la destilación de la mezcla de ácidos (apartado 4.2.2), aumentando el del triglicérido en proporción inversa, lo que concuerda con el conocimiento del proceso.

Aunque se ha mejorado el ajuste del sistema, al comparar los IA e IOH calculados para las muestras del proceso de la figura 5.10 con los correspondientes valores de referencia, se obtienen RSEC (%) del 15-20 %, claramente superiores a los obtenidos por PLS. Esto indica que es necesario utilizar restricciones de igualdad para solventar la ambigüedad existente en los perfiles de concentración obtenidos. Para las muestras de este proceso, se determinó la concentración de la mezcla de ácidos para su utilización como restricciones de igualdad.

Para definir la capacidad de MCR-ALS en aplicaciones cuantitativas se han preparado dos modelos de calibración. En el primero de ellos, modelo 1, se han utilizado como restricciones de igualdad los valores de concentración de mezcla de ácidos de todas las muestras de calibración. En el segundo, modelo 2, se han utilizado únicamente el valor inicial y final de la concentración de ácido de cada proceso. En ambos modelos se ha utilizado una matriz con los datos espectrales de tres procesos, con un total de 24 muestras, aumentada con los 4 espectros puros de los componentes significativos (glicerina, ácido, monoglicérido y triglicérido), y se ha aplicado la restricción de no negatividad (espectral y en concentraciones). Los perfiles espectrales y de concentración obtenidos para ambos modelos son muy similares, por lo que solo se muestran los correspondientes al modelo 1 en las figuras 5.11 y 5.12.

Para ambos modelos la varianza explicada (ecuación 3.25) es del 99,99%, y las correlaciones entre los espectros estimados y los utilizados inicialmente es superior a 0,999, lo que demuestra el buen ajuste obtenido. A partir de los valores de concentración obtenidos se han calculado los correspondientes IA e IOH para cada muestra, y se han comparado con los obtenidos por valoración. Las figuras de mérito de la regresión lineal entre el valor estimado y el de referencia, así como los RSEC/P(%) se muestran en la tabla 5.4.

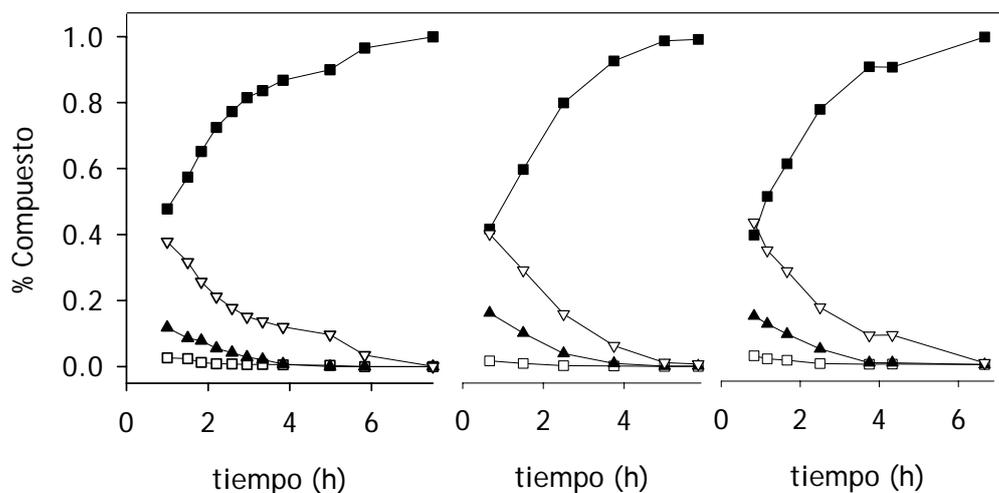


Figura 5.11. Perfiles de concentración obtenidos por MCR-ALS para tres procesos de esterificación. (∇) ácido, (■) triglicérido, (▲) monoglicérido y (□) glicerina.

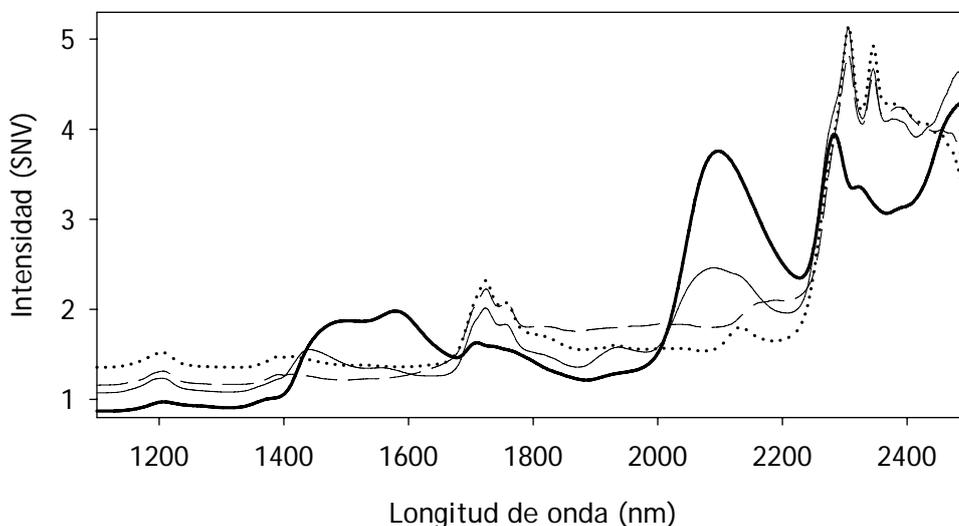


Figura 5.12. Espectros puros obtenidos por MCR-ALS correspondientes a los perfiles de concentración de la figura 6.11. (---) ácido, (.....) triglicérido, (—) monoglicérido y (—) glicerina.

En calibración, ambos modelos ofrecen pendientes próximas a 1, incluyéndose este valor dentro del intervalo de confianza, siendo ligeramente mejor el ajuste

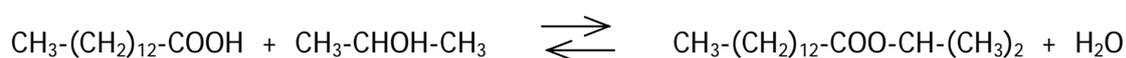
obtenido para el modelo 1. En predicción ambos modelos ofrecen resultados muy similares, y a pesar de que los valores de RSEC/P(%) obtenidos por MCR-ALS son superiores a los obtenidos por PLS (tabla 5.1), se demuestra la capacidad de MCR-ALS para el análisis cuantitativo con muy poca información cuantitativa obtenida por métodos de referencia, eliminando la necesidad de un valor de referencia para cada parámetro a determinar.

Modelo Parámetro	Modelo 1		Modelo 2	
	IA	IOH	IA	IOH
Calibración				
nº muestras	24	24	24	24
Inter. Concentr. (mg KOH/g)	0,4 - 162	1 - 117	0,4 - 162	1 - 117
Pendiente	$1,00 \pm 9 \cdot 10^{-6}$	$0,97 \pm 0,04$	$1,03 \pm 0,03$	$0,95 \pm 0,05$
Ordenada Origen	$-7 \cdot 10^{-4} \pm 7 \cdot 10^{-4}$	$0,3 \pm 2,1$	$-2,2 \pm 2,5$	$1,3 \pm 2,8$
Correlación	1	0,996	0,998	0,994
RSEC (%)	0,001	6,8	4,5	8,9
Predicción				
nº muestras	33	27	33	27
Inter. Concentr. (mg KOH/g)	1,5 - 153	3 - 116	1,5 - 153	3 - 116
Pendiente	$1,09 \pm 0,03$	$0,87 \pm 0,03$	$1,11 \pm 0,03$	$0,86 \pm 0,03$
Ordenada Origen	$-2,4 \pm 2,2$	$-0,9 \pm 1,5$	$-3,8 \pm 2,2$	$-0,1 \pm 1,5$
Correlación	0,997	0,997	0,997	0,997
RSEP (%)	11,8	16,6	12,2	16,4

Tabla 5.4. Figuras de mérito para la determinación de índice de ácido e índice de hidroxilo mediante MCR-ALS.

5.2 Seguimiento in-line de la esterificación entre isopropanol y ácido mirístico

Las reacciones de esterificación de ácidos grasos con alcoholes de bajo punto de ebullición, como el isopropanol, plantean un dilema: cuando se aplican temperaturas de reacción elevadas se favorece la destilación del alcohol, lo que da lugar a pobres rendimientos de éster, y cuando se aplican temperaturas bajas, la reacción de esterificación es muy lenta y por tanto inadecuada en la práctica industrial. Una alternativa es aplicar una temperatura elevada pero adicionar el alcohol lentamente en el medio de reacción para evitar la excesiva destilación, aumentando el rendimiento, pero sin evitar que se generen destilados de isopropanol y agua generada por la reacción. Con el objetivo de reducir el consumo de materia prima, estos destilados pueden ser reutilizados en procesos posteriores, teniendo en cuenta que un elevado contenido en agua puede hacer disminuir el rendimiento de la reacción por desplazamiento del equilibrio:



Si se desea reutilizar los destilados, es necesario establecer previamente el efecto de la concentración de agua en la cinética y en el rendimiento de la reacción. El objetivo de este trabajo es comprobar el efecto de la concentración de agua, encontrando un punto de compromiso en el que sea posible reutilizar los destilados, pero sin comprometer la velocidad de la reacción. Como se ha demostrado en el apartado anterior, es posible alcanzar este objetivo con un mínimo de análisis de referencia a partir de la aplicación de MCR-ALS a los datos espectroscópicos NIR, obtenidos mediante una sonda NIR *in-line*, registrando espectros a intervalos de 3 minutos.

Los espectros obtenidos para un proceso se muestran en la figura 5.13, donde se puede observar la disminución de la absorbancia en función del tiempo principalmente en dos intervalos: 1400 - 1650 nm, y 1800 - 2200 nm. Debido a que la concentración de isopropanol en el medio de reacción es baja, la principal causa de variabilidad espectral es la desaparición del mirístico y la formación del miristato. Esto se puede comprobar comparando los espectros de la figura 5.13 con los espectros de mirístico y miristato de la figura 5.14. Las bandas de absorción más características de estos compuestos, de forma similar a la reacción anterior, corresponden a primer sobretono y bandas de combinación de enlaces O-H y C-H.

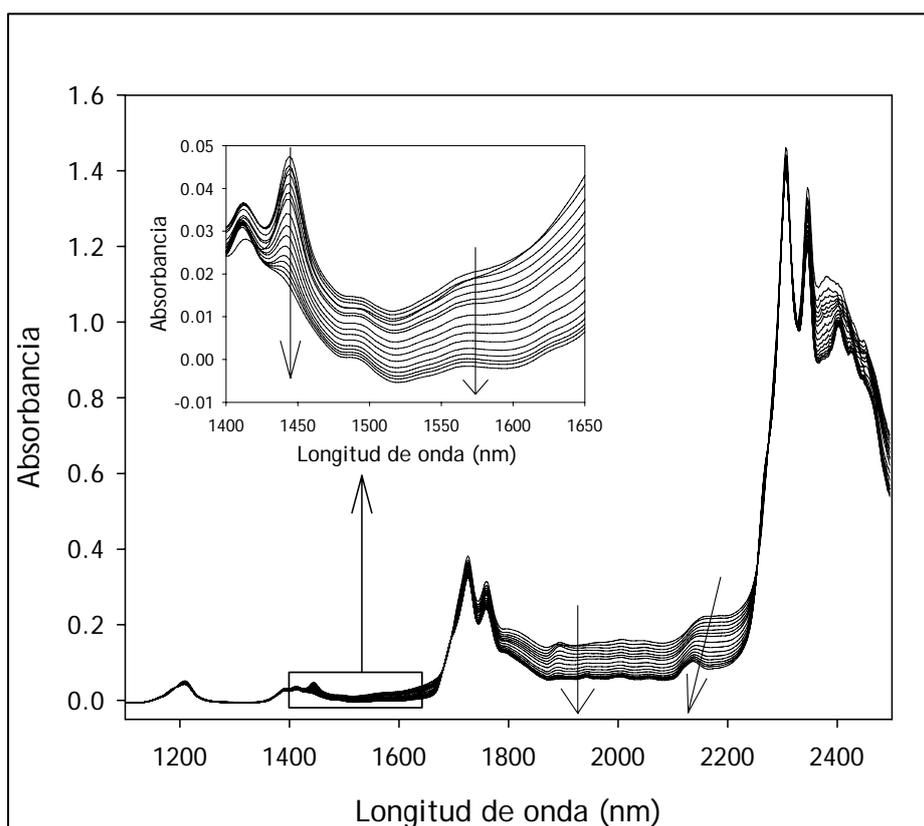


Figura 5.13. Espectros correspondientes a un proceso de esterificación entre ácido mirístico e isopropanol adicionado en continuo.

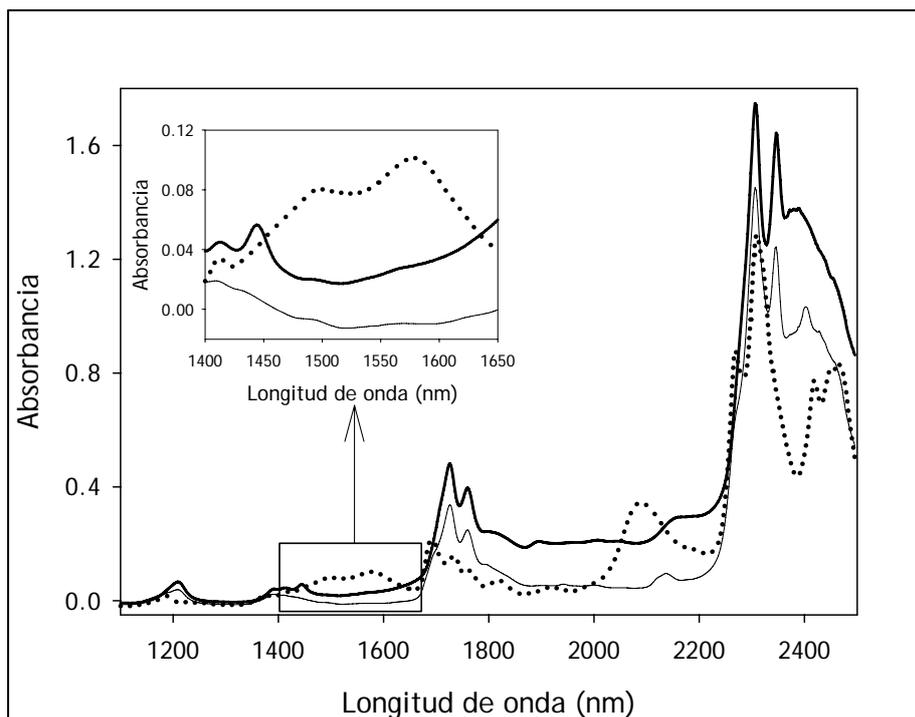


Figura 5.14. Espectros NIR de (—) ácido mirístico, (· · · · ·) isopropanol y (—) miristato de isopropilo.

5.2.1 Aplicación de MCR-ALS

Se ha aplicado a los espectros el pretratamiento MSC (*Multiplicative Scattering Correction*) para solventar problemas de *scattering*. Para aplicar la restricción de no negatividad espectral en el cálculo mediante MCR-ALS, se ha realizado una corrección de la línea de base (*baseline offset*) que desplaza los espectros a valores de absorbancia positivos.

MCR-ALS fue aplicado a los procesos 1-6 (ver tabla 4.1), realizados con diferentes proporciones de isopropanol:agua (entre el 0% y el 25% de agua a intervalos de 5 unidades), dejando el proceso 7 para un estudio posterior. Para cada una de las 6 matrices de datos, correspondientes a los espectros NIR registrados para cada proceso, se estudió el rango matemático, o número de componentes, mediante descomposición en valores propios. La figura 5.15 muestra, a modo de ejemplo, los cuatro primeros vectores propios en función del tiempo para la matriz del proceso

número 1. Los tres primeros *eigenvectors* muestran una tendencia y se asemejan a la evolución de las concentraciones de mirístico, miristato e isopropanol o agua. Sin embargo el cuarto componente no sigue ninguna tendencia definida, además de contener mucho ruido, lo que hace difícil su interpretación y asociación química. Los valores propios correspondientes a los 4 primeros *eigenvectors* son $3,5 \cdot 10^6$, 283, 1,4 y 0,7, siendo el cuarto valor inferior a 1, por lo que sigue siendo adecuada la hipótesis de que el número significativo de componentes químicos que se pueden extraer de los datos es 3.

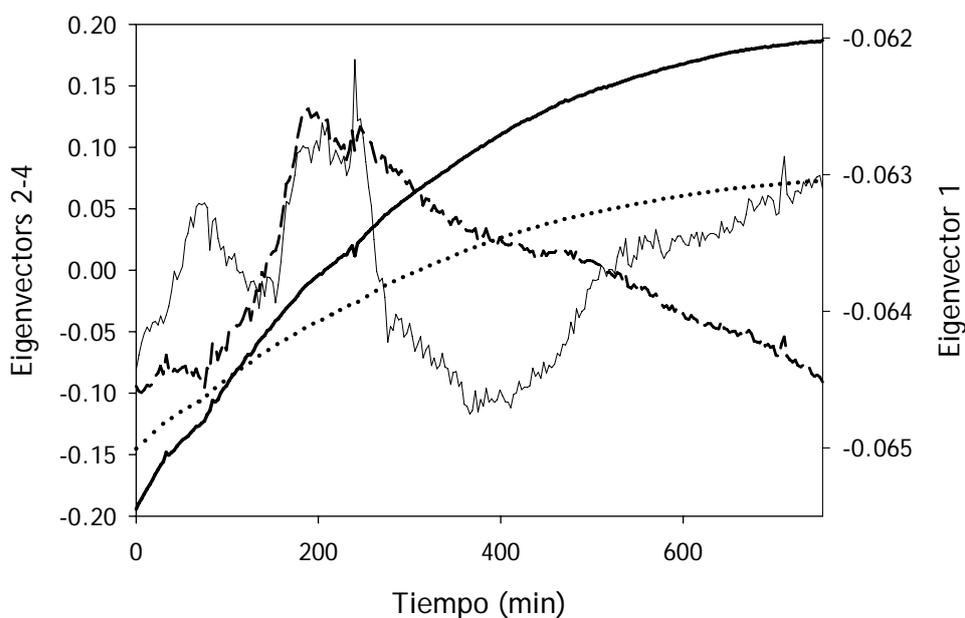


Figura 5.15. *Eigenvectors* correspondientes a una descomposición en valores singulares para la matriz del proceso 1. (—) *eigenvector* 1, (···) *eigenvector* 2, (-.-) *eigenvector* 3, (-) *eigenvector* 4.

Ácido mirístico y miristato de isopropilo son dos de los tres componentes, ya que ambos tienen una elevada concentración en algún momento de la reacción: mirístico al inicio y miristato al final. El tercer componente puede ser isopropanol o agua, aunque este último está presente en baja concentración debido a que en este

proceso únicamente era generada por la reacción y destilada por la elevada temperatura (130°C), mientras que el isopropanol es añadido en continuo a 0,5 g/min. Para corroborar que el tercer componente es isopropanol, se aplicó MCR-ALS a la matriz de datos aumentada con los espectros puros de los 4 componentes, con el objetivo de solventar la deficiencia de rango. Los perfiles de concentración obtenidos para isopropanol y agua se muestran en la figura 5.16.

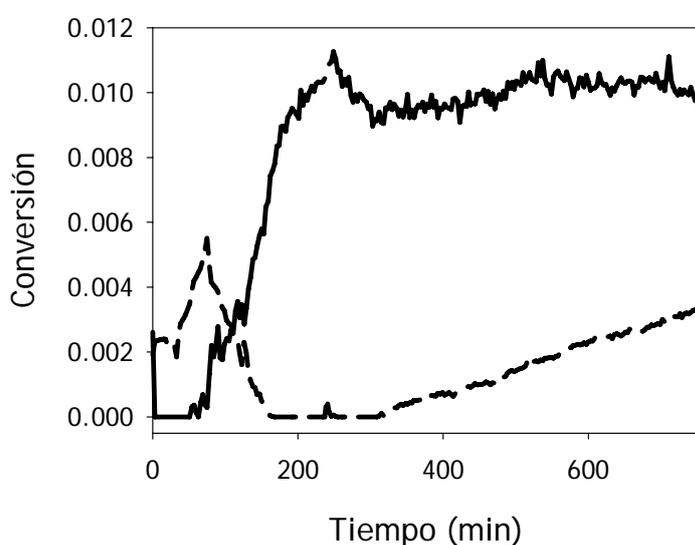


Figura 5.16. Perfiles de concentración obtenidos por MCR-ALS para (—) isopropanol y (---) agua.

Según la figura 5.16, la concentración obtenida para el agua es sensiblemente inferior a la del isopropanol en casi todo el intervalo de tiempo, lo que demuestra que el isopropanol ha de ser el tercer componente empleado en el cálculo por MCR, y no el agua.

Una vez seleccionados los tres componentes significativos, se aplicó MCR-ALS de forma individual a cada matriz de datos espectrales de los procesos 1-6. Previo al cálculo, cada matriz fue aumentada con los espectros puros registrados de mirístico,

miristato e isopropanol y se aplicó únicamente la restricción de no negatividad espectral y en concentraciones. Para los 6 procesos se obtuvo una varianza capturada (ecuación 3.25) del 100% y un coeficiente de correlación (ecuación 3.26) superior a 0,999 entre los espectros puros estimados por MCR y los registrados por NIR para mirístico, miristato e isopropanol. Estos resultados demuestran el buen ajuste espectral obtenido en el cálculo.

En la figura 5.17 se muestran los perfiles de concentración obtenidos para los 6 procesos, junto con los valores de concentración de ácido mirístico obtenidos por valoración, para las muestras de reacción extraídas a lo largo de cada proceso. Como se puede observar, la evolución de mirístico y miristato son prácticamente inversas, mientras que la concentración de isopropanol es siempre baja, entre 0,5 y 2% dependiendo del proceso, aumentando ligeramente en función del tiempo debido a que este reactivo se adiciona de forma constante en el medio, y la reacción es cada vez más lenta a medida que se consume el ácido mirístico, lo que provoca una pequeña acumulación de este reactivo en el medio.

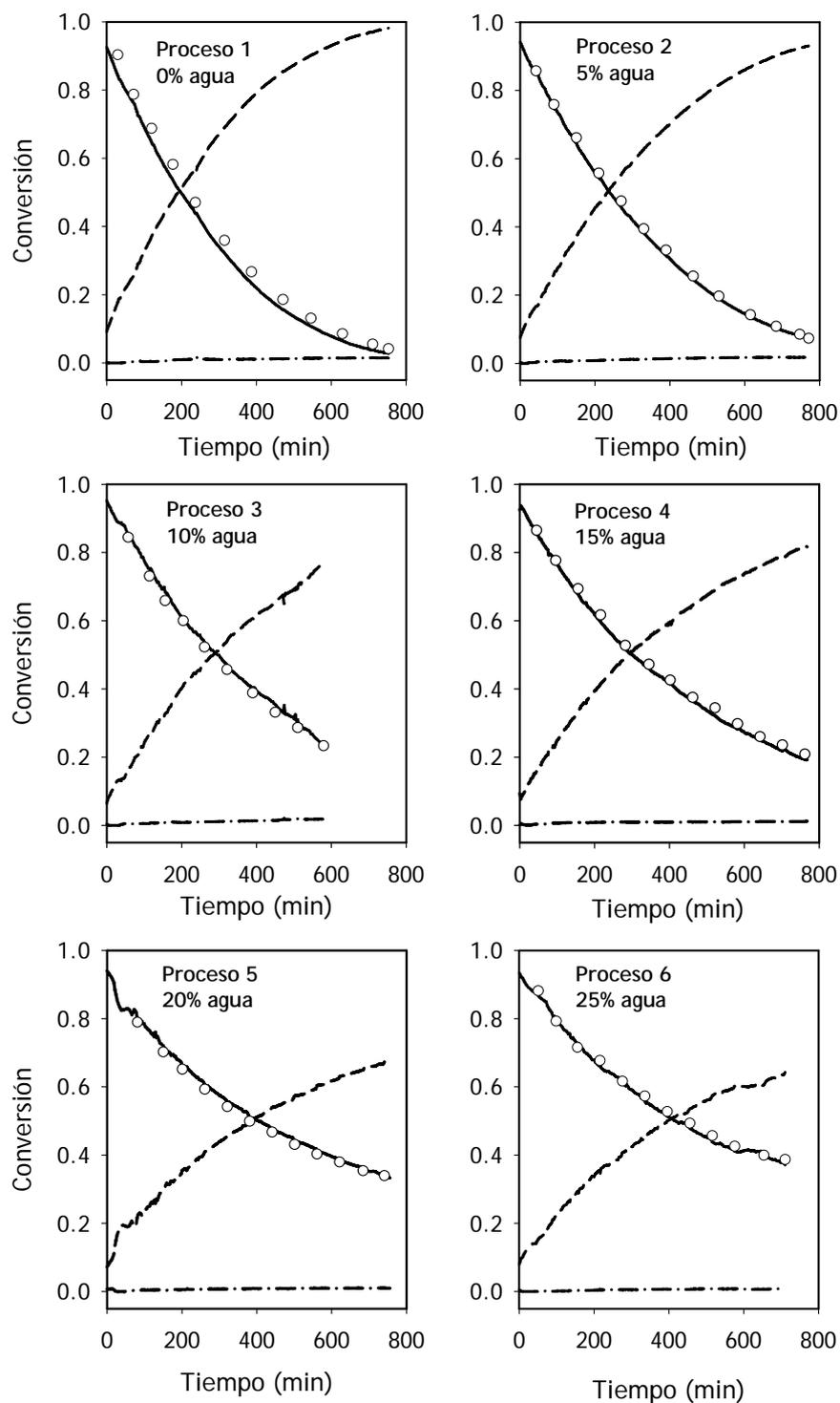


Figura 5.17. Perfiles de concentración obtenidos por MCR-ALS para 6 procesos de esterificación. (—) ácido mirístico, (---) miristato de isopropilo, (- · -) isopropanol, (○) concentración de ácido mirístico obtenida por valoración.

El grado de conversión que se alcanza en cada proceso es diferente, y depende claramente del contenido en agua del isopropanol. A mayor contenido en agua, más lenta es la velocidad de la reacción ya que la pendiente del perfil de concentración para el mirístico disminuye, y también menor es el grado de conversión que se alcanza. En la tabla 5.5 se muestra el tanto por ciento de ácido mirístico sin reaccionar al final de cada proceso; para tiempos de reacción aproximadamente similares, el proceso 6 tiene un porcentaje de ácido sin reaccionar 9 veces superior al del proceso 1.

Estos resultados parecen desaconsejar el uso de destilados con un elevado contenido en agua, ya que alargarían significativamente el tiempo de reacción y no permitirían obtener un rendimiento elevado.

Proceso	Agua (%)	Tiempo total (min.)	Ácido sin reaccionar (%)
1	0	753	4.2
2	5	771	7.3
3	10	582	23.3
4	15	768	20.9
5	20	756	33.9
6	25	711	38.7

Tabla 5.5. Tiempo de reacción y porcentaje de ácido sin reaccionar para cada proceso.

La figura 5.17 también compara el perfil de concentración para el ácido mirístico con los resultados obtenidos por valoración (no empleados en el cálculo), mostrando ambos resultados una gran similitud. La tabla 5.6 muestra la correlación y los %RSE y RMSE (ecuaciones 3.18 y 3.19 respectivamente) entre los valores de MCR y los de valoración. Los resultados ofrecen, con la excepción de los resultados del

proceso 1, un error estándar relativo inferior o entorno al 3%, así como una elevada correlación, confirmando el buen ajuste obtenido por MCR.

Proceso	RSE (%)	RMSE	R ²
1	7.2	0.034	0.9996
2	2.3	0.010	0.9998
3	3.1	0.017	0.9992
4	2.7	0.014	0.9997
5	1.8	0.010	0.9997
6	2.1	0.012	0.998

Tabla 5.6. RSE (%), RMSE y coeficiente de determinación entre los valores obtenidos por MCR y los obtenidos por valoración.

5.2.2 Estimación de parámetros cinéticos

Para poder determinar con mayor precisión el efecto del contenido en agua en la velocidad de la reacción es de utilidad la estimación de constantes cinéticas de reacción para cada uno de los 6 procesos estudiados.

Las reacciones de esterificación con catálisis homogénea son generalmente de orden 2 respecto al ácido (Ac) y al alcohol (Al):

$$v = -\frac{d[Ac]}{dt} = k \cdot [Ac][Al] \quad [5.1]$$

Sin embargo, en condiciones de elevada concentración de alcohol, o por el contrario, en un defecto constante, la reacción puede aproximarse a una cinética de pseudoprimer orden respecto del ácido:

$$v = -\frac{d[Ac]}{dt} = k' \cdot [Ac] \quad [5.2]$$

siendo $k' = k \cdot [AI]$ [5.3]

La ecuación 2 integrada es:

$$\ln[Ac] = \ln[Ac]_0 - k' \cdot t \quad [5.4]$$

donde k' es la pseudoconstante de velocidad, que depende de parámetros experimentales como la temperatura, la velocidad de adición del alcohol y el contenido en agua de éste. La constante k' puede calcularse fácilmente a partir del gráfico de $\ln[Ac]$ vs. tiempo.

La figura 5.18, correspondiente a la ecuación [5.4] para las 4 primeras horas de reacción de cada proceso, muestra como los perfiles se ajustan correctamente a una cinética de pseudoprimer orden. La linealidad obtenida en los 6 casos es alta (el coeficiente de correlación del ajuste por mínimos cuadrados es en los 6 casos superior a 0.997). En el proceso 5 se ha suprimido el intervalo 18-72 minutos debido a variaciones en el registro de los espectros provocadas por burbujas en el camino óptico. En la tabla 5.7 se muestra la pendiente, correspondiente a k' , para cada gráfico de la figura 5.18, la cual disminuye a medida que aumenta el contenido en agua del isopropanol.

Los perfiles de concentración del ácido mirístico de la figura 5.17 indican que al aumentar el contenido de agua disminuye la pendiente inicial del perfil, es decir, disminuye la velocidad de la reacción y la curva del perfil parece estabilizarse a una asíntota mayor, disminuyendo también el rendimiento de la reacción. Según la tabla 5.7, Los valores de k' , se reducen a la mitad al pasar de isopropanol puro a contenidos de agua del 20%; mientras que esta disminución es sensiblemente menor (35%) para contenidos de agua del 15%. Reutilizar los destilados con contenidos de agua del 10-15% es viable industrialmente aunque la velocidad de la reacción y el

rendimiento sean menores. Sin embargo, estos deberían ser utilizados únicamente al inicio de la reacción, finalizando la reacción con destilados con baja concentración de agua, o isopropanol puro, para obtener rendimientos elevados y realizar el proceso en menor tiempo.

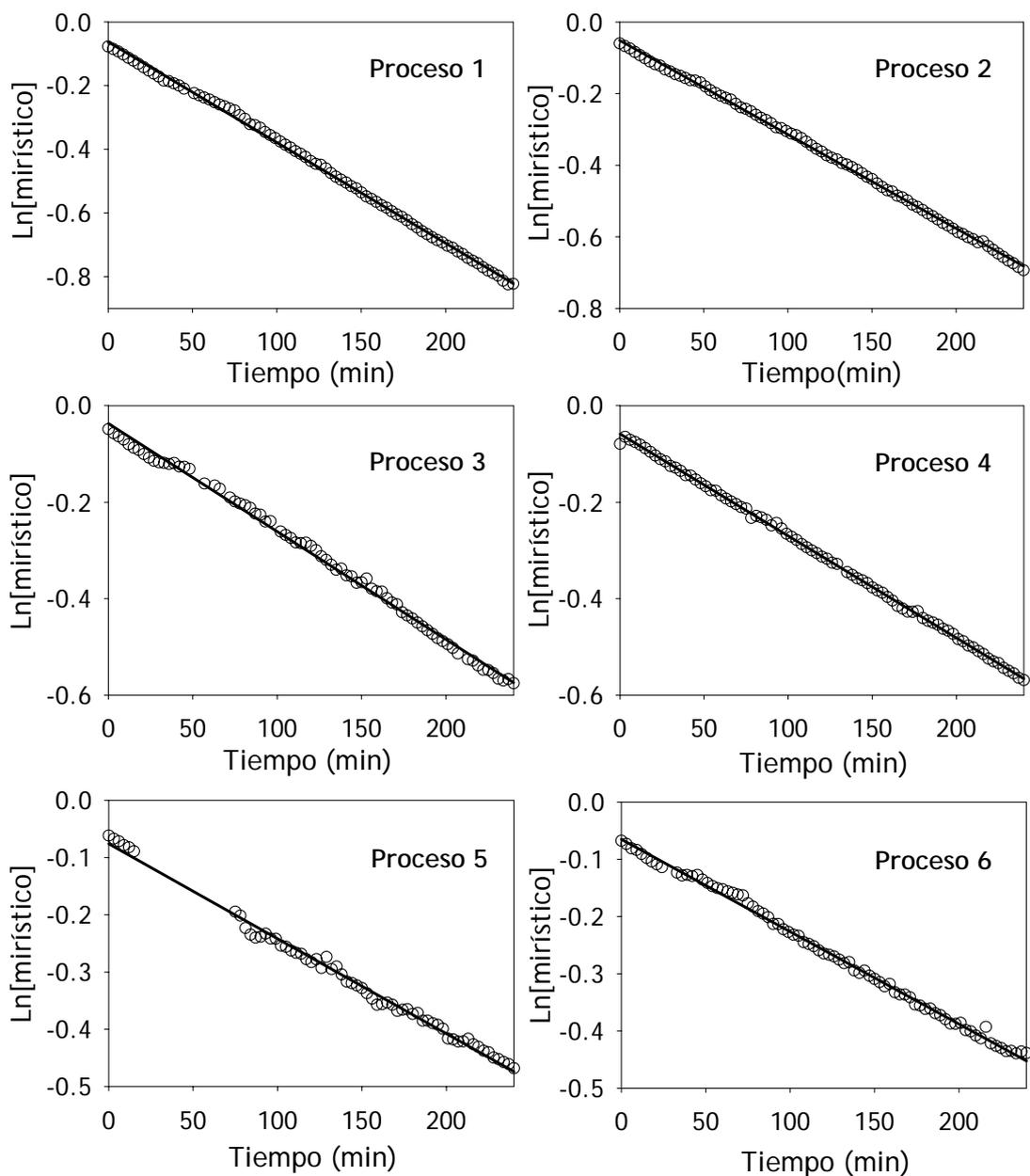


Figura 5.18. $\text{Ln}[\text{mirístico}]$ frente al tiempo para los 6 procesos estudiados. (○) valores individuales y (—) regresión lineal.

Proceso	Agua (%)	k' ($\times 10^3 \text{ min}^{-1}$)
1	0	3.16 ± 0.03
2	5	2.63 ± 0.02
3	10	2.24 ± 0.03
4	15	2.11 ± 0.01
5	20	1.66 ± 0.04
6	25	1.61 ± 0.02

Tabla 5.7. Valores de la pseudoconstante de velocidad (k').

Para poder reutilizar los destilados es necesario conocer la composición de éstos durante el proceso de reacción. Para ello, los destilados obtenidos durante los procesos reactivos han sido controlados mediante espectroscopia NIR at-line. Previamente, se desarrolló un modelo de calibración PLS a partir de 21 mezclas binarias con isopropanol entre el 50 y 100% (0-50% de agua), a intervalos de 5%. Los modelos de calibración para isopropanol y agua se construyeron en segunda derivada, con dos factores PLS y en los intervalos 1100-2498 nm para el isopropanol y 1350-2150 nm para el agua. Los RSE% obtenidos entre los valores predichos y los de referencia para las 21 muestras de calibración es de 0.6% para el isopropanol y 1.5% para el agua.

La figura 5.19 muestra la evolución en la concentración de isopropanol de los destilados obtenidos de reacción en función del tiempo para los procesos con 100%, 95%, 85% y 75% de isopropanol. Esta concentración aumenta con el tiempo y tiende hacia el valor de la mezcla de alimentación del reactor, lo que es explicable por la mayor velocidad de esterificación en los primeros minutos y una ralentización del proceso a medida que la reacción se va completando. Para los procesos del 100% y 95%, el destilado obtenido en la etapa final del proceso puede ser reutilizado directamente como mezcla de alimentación sin necesidad de ser tratado, mientras

que el primer destilado debe tratarse para reducir el contenido en agua antes de ser utilizados en procesos posteriores.

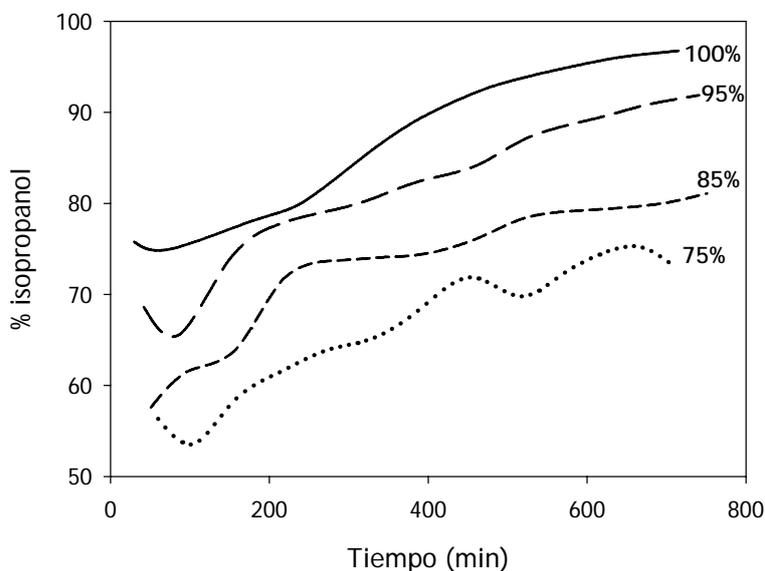


Figura 5.19. Evolución de la concentración de isopropanol en los destilados de reacción.

Otra posible alternativa para aumentar la velocidad de reacción reutilizando los destilados es aumentar la velocidad de adición del isopropanol. Para comprobar el efecto de la velocidad de adición se ha aplicado MCR-ALS a un proceso doblando la velocidad de adición (proceso 7, velocidad adición = 1 g/min) y se han comparado los perfiles obtenidos con los del proceso nº 3 (0,5 g/min), realizados ambos con una mezcla de alimentación con un 10% en agua. Los perfiles de concentración obtenidos en ambos casos se muestran en la figura 5.20, donde se puede comprobar que el aumento de la velocidad de adición conduce a una mayor velocidad de reacción, y a alcanzar concentraciones de ácido menores. Experimentalmente, una mayor velocidad de adición supone aumentar la concentración de isopropanol en el medio de reacción, aumentando así la velocidad de reacción. Esta estrategia, sin embargo, supone incrementar el volumen de destilado isopropanol-agua, aumentando el consumo de reactivos: el proceso nº 7 se finalizó con un 7,6% de ácido sin reaccionar,

similar al del proceso nº 2 (7,3%), sin embargo, el consumo de mezcla isopropanol:agua en el nº 7 fue aproximadamente el doble que en el nº 2. Por este motivo, el incremento de la velocidad de adición no es una alternativa eficiente para incrementar la velocidad y el rendimiento de la reacción.

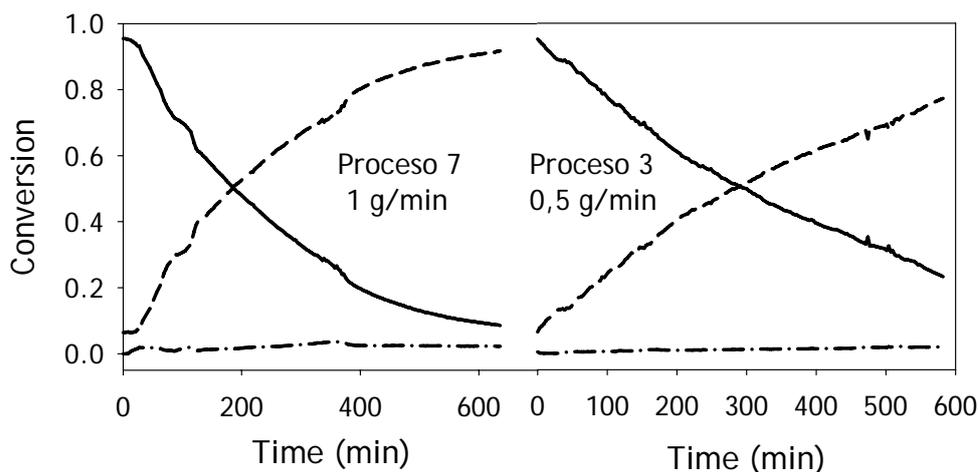


Figura 5.20. Perfiles de concentración para los procesos 3 y 7. (—) ácido mirístico, (---) miristato de isopropilo y (-·-) isopropanol.

5.3 Determinación de bajas concentraciones de un analito mediante espectroscopia NIR

La baja sensibilidad del NIR dificulta su aplicación en la determinación de analitos traza. Sin embargo, las ventajas aportadas por esta técnica pueden ser de gran utilidad en procesos de reacción donde se requiere información acerca de la composición de la mezcla en el menor tiempo posible. Para solventar la baja concentración de los analitos de interés, y poder aplicar la espectroscopia NIR para determinar analitos traza, en este trabajo se ha puesto énfasis en la metodología

experimental y en el pretratamiento de datos para intentar mejorar la determinación de 2-etilhexanol en un éster de ácido graso de coco. En un primer paso, se ha incrementado el camino óptico hasta 8 mm y se han registrado los espectros a 75°C, temperatura a la cual las muestras tienen menor turbidez, por lo que ofrecerán menos dispersión de luz. En un segundo paso, se han aplicado dos pretratamientos espectrales, NAS (Net Analyte Signal) y OSC (Orthogonal Signal Correction) para filtrar y eliminar la información espectral no relacionada con la concentración de analito, con el objetivo de mejorar los resultados obtenidos por PLS. Para comprobar el efecto de estos dos pretratamientos, se han calculado los RSEC/P (%) y se han estimado los límites de detección para cada modelo.

En el proceso estudiado, en el cual no hay reacción, la variación observable en los espectros NIR es debida a la destilación del 2-etilhexanol de una matriz compuesta por ésteres de este alcohol con ácidos grasos de origen natural. A pesar de que el proceso comienza a concentraciones de alcohol entorno al 3%, la zona de mayor interés para el control analítico se encuentra a concentraciones inferiores a 1000 ppm. La figura 5.21 muestra los espectros correspondientes a 7 muestras con concentraciones de 2-etilhexanol entre 22 ppm y 3980 ppm. Como se puede comprobar, las diferencias entre muestras son muy pequeñas, aún incluso en los gráficos ampliados, y se encuentran principalmente en los intervalos 1400 - 1650 nm y 1900 - 2200 nm, regiones donde más se diferencia el espectro del alcohol respecto al de las muestras.

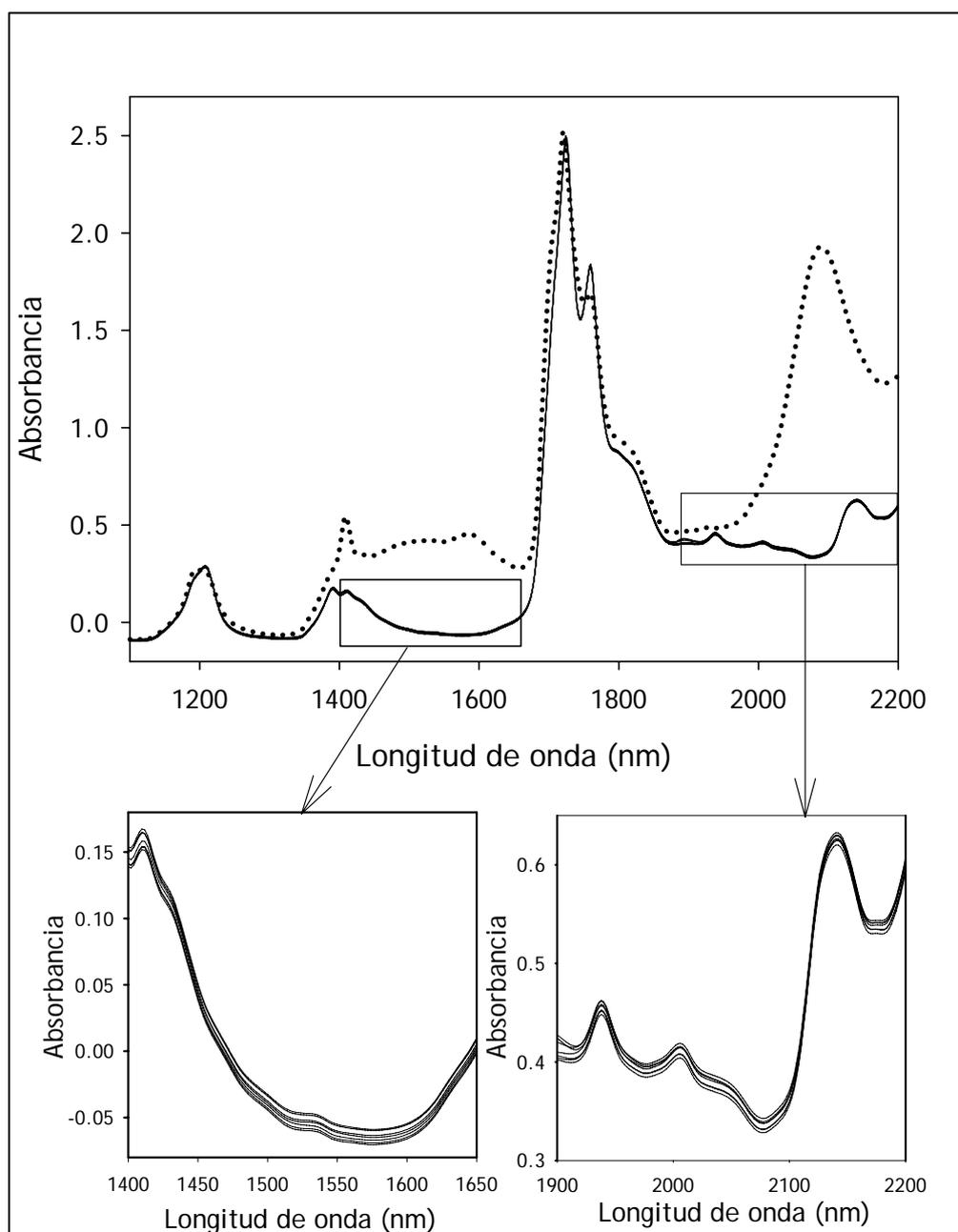


Figura 5.21. Espectros NIR de 2-etilhexanol y muestras entre 22 y 4000 ppm de 2-etilhexanol.

5.3.1 Determinación de 2-etilhexanol

Para la determinación de 2-etilhexanol se dispone de un conjunto de 33 muestras en el intervalo de concentraciones 22-1630 ppm, que se dividió en dos subconjuntos de calibración y predicción, con 20 y 13 muestras respectivamente. Se han ensayado diferentes pretratamientos y regiones espectrales, los mejores resultados se

obtuvieron con 4 factores PLS en el intervalo 1300-1500 nm y en segunda derivada Savitzky-Golay, con polinomio de segundo orden y una ventana de 11 puntos. La tabla 5.8 muestra los parámetros de la regresión de los valores NIR frente a los valores de referencia, así como los RSEC/P (%) obtenidos.

Set	Parámetros	PLS	OSC-PLS	NAS-PLS
Calibración	Factores PLS	4	2	1
	intervalo de concentración (ppm)	22 - 1630	22 - 1630	22 - 1630
	Pendiente	0,98 ± 0,07	0,99 ± 0,04	0,98 ± 0,12
	Ordenada Origen	6 ± 33	2 ± 20	6 ± 33
	r	0,991	0,997	0,991
	RSEC (%)	9,9	6,1	9,8
	Predicción	intervalo de concentración (ppm)	56 - 1060	56 - 1060
Pendiente		0,99 ± 0,12	0,96 ± 0,11	0,96 ± 0,12
Ordenada Origen		10 ± 42	42 ± 40	13 ± 43
r		0,984	0,986	0,983
RSEP (%)		12,6	14,8	13,1

Tabla 5.8. Figuras de mérito de la determinación de 2-etilhexanol mediante PLS, OSC-PLS y NAS-PLS.

Atendiendo al intervalo de concentraciones del modelo de calibración, los valores del coeficiente de correlación y RSE (%) para el modelo PLS se pueden considerar buenos, y tanto pendientes como ordenadas en el origen son próximas a 1 y 0 respectivamente, estando ambos valores incluidos dentro del intervalo de confianza.

La primera estrategia para intentar mejorar los resultados obtenidos anteriormente se basa en la aplicación de OSC (*Orthogonal Signal Correction*). Utilizando el mismo conjunto de datos, 20 muestras para calibración y 13 para predicción, se aplicó OSC previo al cálculo mediante PLS. Un solo factor OSC es suficiente para eliminar gran parte de la información espectral no correlacionada con la concentración del analito. Así, el primer factor OSC elimina un 24,9 % de la varianza espectral, mientras que dos y sucesivos factores no consiguen eliminar más de un 28,0 %. El efecto del OSC sobre los espectros se puede observar en la figura 5.22, donde se muestran los espectros en segunda derivada para las muestras de calibración, y los correspondientes espectros después de aplicar un factor OSC.

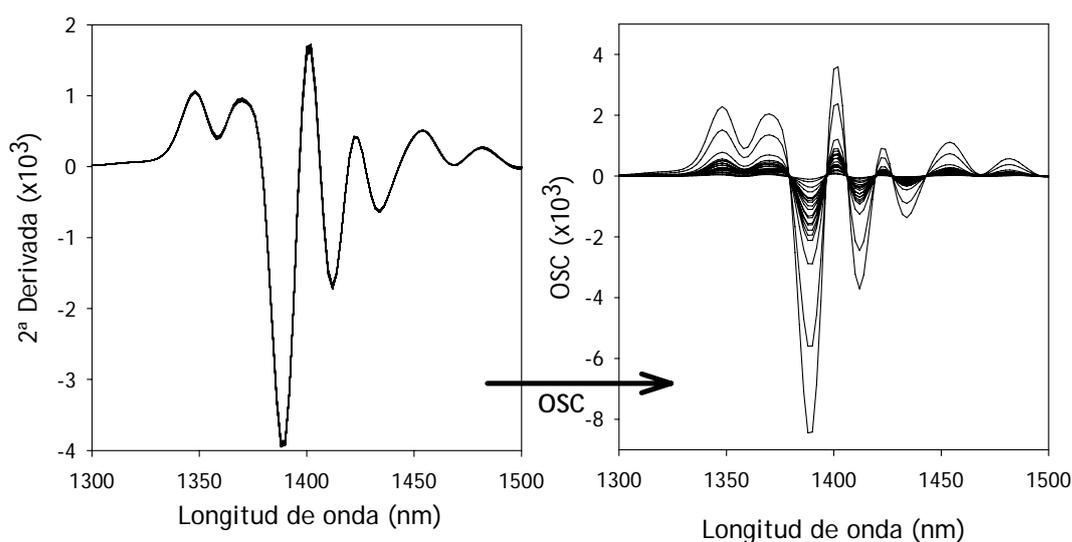


Figura 5.22. Espectros NIR en segunda derivada (izq.) y espectros filtrados mediante OSC (der.) para las muestras del conjunto de calibración.

Para seleccionar adecuadamente el número de factores OSC y evitar un sobreajuste, se estudio la capacidad predictiva utilizando muestras externas. Con 1 factor OSC, el modelo de calibración requiere 2 factores PLS (valor óptimo según el PRESS), y ofrece un RSEP (%) del 14,8. Con 2 factores OSC, el mejor modelo se obtiene con solo un factor PLS según el PRESS. La predicción de muestras externas con este modelo ofrece un RSEP (%) del 17,9, superior al anterior. Así, se considera

adecuada la utilización de un factor OSC combinado con dos factores PLS. Los parámetros de la recta de regresión de los valores predichos por NIR frente a los obtenidos por cromatografía y los RSE (%), para los conjuntos de calibración y predicción, se muestran en la tabla 5.8.

La segunda estrategia consiste en la aplicación del NAS (*Net Analyte Signal*), utilizando el mismo conjunto de datos, 20 muestras para calibración y 13 para predicción, siguiendo la metodología descrita en la figura 5.23: el NAS se aplicó a la matriz de calibración centrada y reconstruida con 7 componentes principales de un PCA. El objetivo de la reconstrucción espectral con el PCA es eliminar el ruido, ya que éste afecta al cálculo del vector. Como no existe un criterio para seleccionar el número de PC's y puede ser una decisión no trivial, se estudió a posteriori su efecto sobre el RSEC(%) y el RSEP(%), comprobando que los mejores resultados se obtienen reconstruyendo la matriz con 7 PCs, que incorporan un 99,3% de la varianza total. Una vez filtrado el ruido, se calcula el vector NAS para cada una de las muestras de calibración (ecuaciones 3.5 - 3.10). El efecto del NAS sobre los espectros se puede observar en la figura 5.24, donde se muestran los espectros en segunda derivada para las muestras de calibración, y los correspondientes vectores NAS. Aparentemente, de forma similar a los espectros OSC de la figura 5.22, este pretratamiento exalta las diferencias espectrales entre muestras, lo que podría permitir la obtención de mejores resultados en la construcción de modelos PLS.

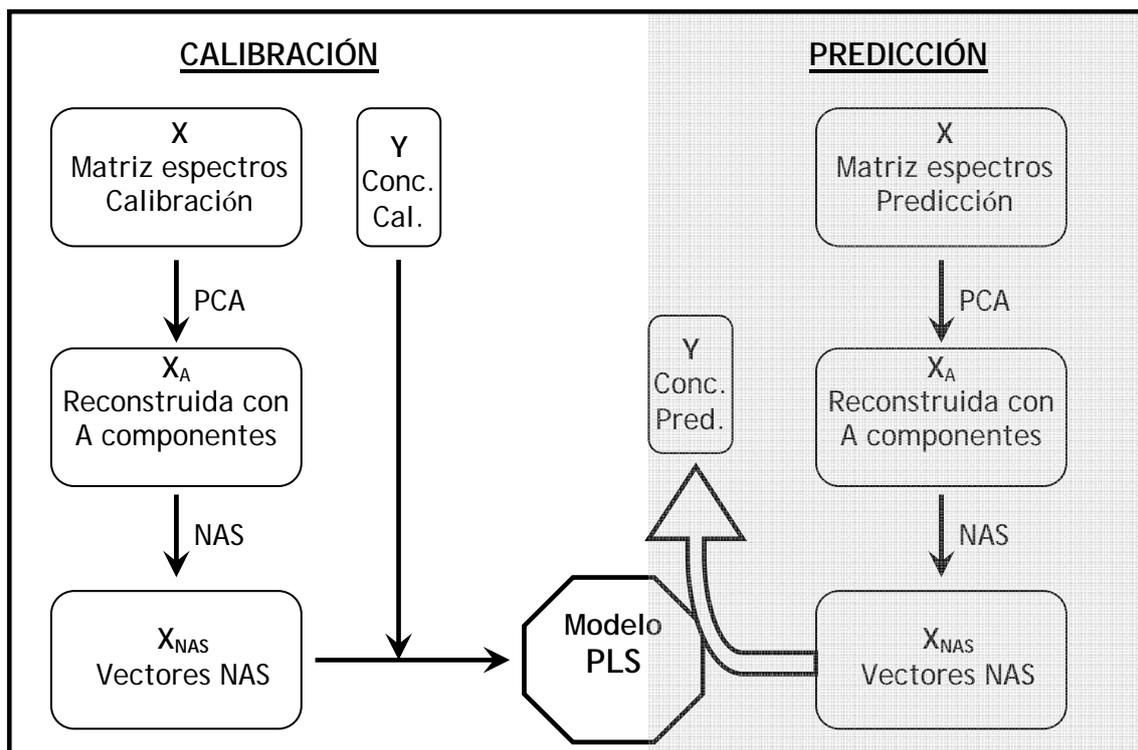


Figura 5.23. Esquema de la aplicación de la señal neta del analito (NAS) a los datos espectrales y de concentración.

Con los vectores NAS se construye el modelo de calibración PLS. Para las muestras de predicción, una vez centrado su espectro y reconstruido con 7 PC's, se calcula su vector NAS y se predice la concentración a partir del modelo PLS. En la tabla 5.8 se muestran los parámetros de la recta de regresión de los valores NIR frente a los de cromatografía y los RSE (%), para los conjuntos de calibración y predicción.

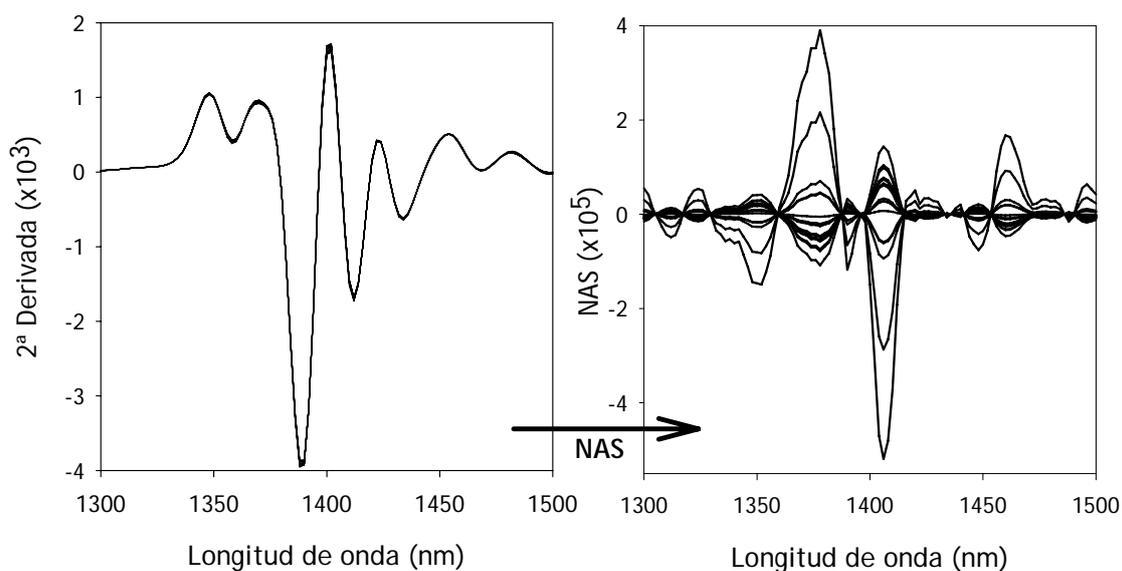


Figura 5.24. Espectros NIR en segunda derivada (izq.) y espectros filtrados mediante NAS (der.) para las muestras del conjunto de calibración.

Los resultados obtenidos por los tres modelos (PLS, OSC-PLS y NAS-PLS) son muy similares por sus figuras de mérito, tanto en calibración como en predicción. Esto supone que, para el estudio presentado, OSC y NAS no mejoran los resultados de predicción para muestras externas. La mayor ventaja que muestran OSC y NAS en este caso es la simplificación de los modelos de calibración con un menor número de factores PLS. Así, el número de factores necesarios se ve reducido de 4 a 2 para el OSC y a 1 para el NAS.

Si se comparan ambos pretratamientos, el NAS se muestra más efectivo para eliminar información espectral no correlacionada con el analito. Así, el OSC es capaz de eliminar con 1 factor un 24,9% de la información contenida en la matriz X , mientras que el NAS permite eliminar un 55,9%. Esto tiene un posterior efecto en la varianza explicada de X e Y por el modelo PLS. Como se muestra en la tabla 5.9, sin aplicar OSC ni NAS, son necesarios 4 factores PLS para explicar el 98,2 % de la varianza de Y , mientras que se puede explicar el mismo porcentaje con solo un factor en el caso del NAS, y un porcentaje ligeramente superior con 2 factores PLS en

el caso del OSC. Respecto a la varianza de la matriz X, si se comparan las varianzas explicadas por los modelos de calibración para NAS y OSC, se corrobora que el NAS tiene una mayor capacidad para filtrar la información no correlacionada con Y, ya que su varianza explicada solo con un factor es prácticamente el 100 %, muy superior a la obtenida con OSC (58,6 %).

Factores PLS	PLS		OSC - PLS		NAS - PLS	
	Var. X (%)	Var. Y (%)	Var. X (%)	Var. Y (%)	Var. X (%)	Var. Y (%)
1	53.0	66.2	58.6	94.1	100.0	98.2
2	80.7	85.4	76.0	99.3	100.0	98.5
3	87.4	96.0	85.3	99.7	100.0	98.9
4	91.9	98.2	94.4	99.7	100.0	99.1

Tabla 5.9. Varianza explicada por el modelo de calibración para las matrices X (espectros) e Y (concentraciones).

5.3.2 Estimación de límites de detección para la determinación de 2-etilhexanol

Los límites de detección para los modelos desarrollados se han estimado mediante dos metodologías, comentadas en el apartado de teoría. En la primera de ellas, se determina el límite de detección multivariante (MDL) en función de la varianza de la concentración predicha por el modelo PLS, ecuación [3.29], calculándose la varianza a partir de la ecuación [3.30] propuesta por Faber y Bro. La segunda metodología se basa en la aproximación realizada por Ortiz *et al.*, que traslada la metodología empleada en calibraciones de orden cero a sistemas multivariados. En este caso, también basada en la ecuación [3.29], y se determina la varianza a partir de las ecuaciones [3.34] y [3.35].

Para la metodología multivariante el límite de detección se determina individualmente para cada muestra, ya que está ligado a la varianza de la concentración predicha de cada muestra, mientras que para la metodología propuesta por Ortiz, el límite de detección es único para cada modelo de calibración porque esta basado en la recta de regresión de NIR vs. referencia. Ambas metodologías han sido aplicadas a las muestras de predicción.

La tabla 5.10 muestra los MDL (Faber y Bro) y DL (Ortiz *et al.*), con errores α y β ($\Delta(\alpha,\beta)$ $\alpha = \beta = 0,05$) y también solo con error α (sustituyendo $\Delta(\alpha,\beta)$ por $t(\alpha)$), determinados para el conjunto de 13 muestras de predicción. Los parámetros de las ecuaciones son los siguientes: el número de muestras de calibración $M = 20$; la varianza del método cromatográfico $\sigma_c^2 = 42,8$ ppm; los grados de libertad para MDL son 12, 16 y 18 para PLS, OSC y NAS respectivamente (estimados mediante la ecuación 4.34), y 18 ($M-2$) para DL. Debido a que los límites de detección multivariantes estimados para las 13 muestras de predicción son muy similares, en la tabla 6.10 se muestra únicamente el promedio.

Modelo	MDL (α)	MDL (α,β)	DL (α)	DL (α,β)
PLS	117 \pm 3	187 \pm 4	114	189
OSC-PLS	109 \pm 2	177 \pm 4	137	228
NAS-PLS	86 \pm 1	139 \pm 2	120	200

Tabla 5.10. Límites de detección en ppm para los tres modelos de calibración.

Como se puede observar en la tabla 5.10, OSC y NAS no ofrecen límites de detección claramente inferiores a los obtenidos por PLS, únicamente el MDL obtenido para el NAS es ligeramente menor, de forma similar a lo argumentado cuando se compararon RSEC(%) y RSEP(%). OSC y NAS eliminan la información no correlacionada con la concentración del analito, suavizando los espectros e incrementando las

diferencias espectrales entre muestras. A pesar de estas ventajas, estos pretratamientos no aportan una notable mejora de la predicción, ni tampoco de los límites de detección para muestras externas.

Respecto a las dos metodologías para el cálculo de límites de detección, si se comparan los MDL (Faber y Bro) y DL (Ortiz *et al.*) obtenidos, las dos metodologías ofrecen resultados similares, especialmente en el caso del PLS, siendo cualquiera de las dos metodologías adecuadas para el cálculo de límites de detección de modelos de calibración multivariable.

Referencias

- [1] A. De Juan, R. Tauler. Chemometrics applied to unravel multicomponent processes and mixtures. Revisiting latest trends in multivariate resolution. *Anal. Chim. Acta.* **500** (2003) 195-210.
- [2] E. R. Malinowski. Statistical F-test for abstract factor analysis and target testing. *J. Chemometr.* **3** (1988) 49-60.
- [3] R. Aafaqi, A. R. Mohamed and S. Bhatia. Kinetics of esterification of palmitic acid with isopropanol using p-toluene sulfonic acid and zinc ethanoate supported over silica gel as catalysts. *J. Chem. Technol. Biotechnol.* **79** (2004) 1127-1134.
- [4] T. Kocsisová, J. Cvengros and J. Lutisan. High-temperature esterification of fatty acids with methanol at ambient pressure. *Eur. J. Lipid Sci. Technol.* **107** (2005) 87-92.
- [5] E. Sendzikiene, V. Makareviciene, P. Janulis and S. Kitrys. Kinetics of free fatty acids esterification with methanol in the production of biodiesel fuel. *Eur. J. Lipid Sci. Technol.* **106** (2004) 831-836.



6. Conclusiones

CONCLUSIONES

Del estudio realizado de diferentes procesos químicos industriales mediante espectroscopia NIR y tratamiento de los datos espectrales con técnicas de calibración multivariable, se pueden extraer las siguientes conclusiones:

- 1 -

Se demuestra que la espectroscopia en el infrarrojo próximo es adecuada para la determinación de parámetros utilizados por la industria en el seguimiento de reacciones de esterificación, tales como el índice de ácido y el índice de hidroxilo, en sustitución de las metodologías clásicas de valoración. La información espectral se puede también aplicar a la determinación de todas las especies implicadas en el proceso de esterificación utilizando un único espectro (determinación multiparamétrica)

Ambos tipos de determinaciones se realizan aplicando modelos de regresión por mínimos cuadrados parciales (PLS).

- 2 -

La espectroscopia en el infrarrojo cercano se muestra como una técnica alternativa adecuada para la monitorización de procesos industriales. Las ventajas más significativas son: obtención rápida de información, información multiparamétrica, reducción de tiempos muertos de producción, eliminación de la necesidad de personal especializado para realizar el análisis, eliminación de la etapa de preparación de muestras, lo que implica un ahorro de costes en reactivos y una mejora medioambiental al eliminar el uso de reactivos tóxicos y disolventes.

- 3 -

La aplicación del algoritmo de resolución multivariante de curvas, MCR, a espectros NIR, registrados tanto *at-line* como *in-line*, ha permitido el seguimiento de la evolución de las concentraciones de las especies involucradas en reacciones de esterificación. Así, MCR-ALS ha permitido resolver los perfiles de concentración de las principales especies que intervienen en la reacción en dos procesos distintos: la esterificación entre glicerina y ácidos caprílico y cáprico, y la esterificación entre isopropanol y ácido mirístico.

- 4 -

Se puede aplicar MCR-ALS a sistemas deficientes de rango, el rango matemático es inferior al rango químico, introduciendo los espectros de los componentes puros como restricciones de igualdad y/o bien aumentando la matriz de datos con estos mismos espectros de las especies puras. Ambas metodologías han demostrado ser válidas para resolver la deficiencia de rango, siendo la segunda más adecuada al dejar mayor libertad de cálculo al sistema durante el proceso iterativo.

Una gran ventaja de las técnicas de resolución de curvas es su elevada flexibilidad para el cálculo de los perfiles espectrales y de concentración, utilizando diferente información del sistema químico estudiado. Así, se han podido estimar perfiles espectrales de especies para las que no se disponía de patrones, que posteriormente han sido utilizados para resolver la deficiencia de rango en otros procesos.

- 5 -

Las técnicas de resolución de curvas aventajan a otras técnicas de análisis multivariante ya que requieren menor información cuantitativa (valores de referencia) para predecir valores cuantitativos en nuevos procesos.

Los valores de RSEC/P(%) obtenidos por MCR-ALS para la determinación de las especies involucradas en ambas reacciones de esterificación son comparables a los obtenibles mediante PLS, demostrando la capacidad de MCR-ALS para el análisis cuantitativo y para la predicción en nuevas muestras.

- 6 -

Los perfiles de concentración obtenidos por MCR-ALS se han aplicado al cálculo de constantes cinéticas de reacción; este estudio se ha aplicado a la reacción de esterificación entre isopropanol y ácido mirístico y ha permitido establecer que las diferencias en la velocidad de reacción en diferentes procesos es función de la pureza del isopropanol. La utilización de las técnicas de resolución ha hecho posible la obtención de información de interés (constantes de velocidad aparentes) sin necesidad de muestreo del proceso.

- 7 -

Se ha demostrado la capacidad de la espectroscopia NIR en la determinación de analitos minoritarios (hasta aprox. 100 ppm) presentes en una matriz con elevada absorción. Se ha estudiado el efecto de dos pre-tratamientos espectrales en el

CONCLUSIONES

establecimiento de los límites de detección del analito y en la calidad de los resultados.

Del estudio de la aplicación de OSC (*Orthogonal Signal Correction*) y NAS (*Net Analyte Signal*) se concluye que con estos dos pretratamientos espectrales no conduce a una mejora de los resultados de predicción, en términos de RSE(%) o RMSE, para muestras externas al conjunto de calibración.

Sin embargo, OSC y NAS permiten la creación de modelos de calibración PLS con menor número de factores ya que filtran de forma eficiente la información espectral que no esté relacionada con la propiedad a determinar. De estos dos pretratamientos, el NAS se muestra más efectivo para eliminar información no correlacionada.

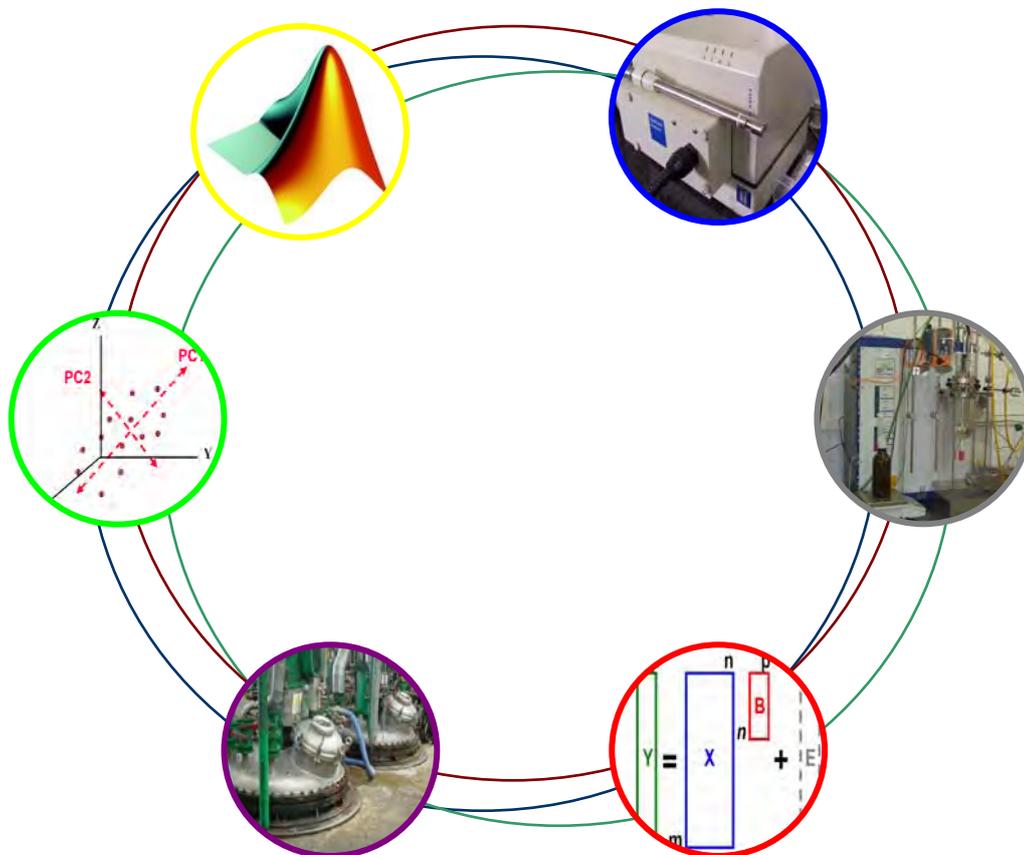
El estudio de los límites de detección para la determinación de 2-etilhexanol ha permitido confirmar la capacidad del NIR para determinar analitos minoritarios. Además, de las múltiples metodologías existentes en la bibliografía para el cálculo de límites de detección, se han seleccionado dos de ellas, que permiten estimar el límite de detección de forma sencilla. Ambas metodologías se muestran eficaces para la estimación de límites de detección en sistemas multivariantes.

- 8 -

Las metodologías desarrolladas en la tesis han aportado soluciones a situaciones reales de la industria química y se utilizan de forma satisfactoria como herramientas para el control de procesos, demostrando la capacidad de la espectroscopia en el infrarrojo próximo así como su versatilidad como técnica para el control analítico de procesos industriales.

CONCLUSIONES

La experiencia y conocimientos adquiridos durante el estudio de procesos de esterificación constituyen una sólida base para su aplicación a nuevos procesos más complejos, aplicando nuevas técnicas de registro espectral y en la utilización de nuevas técnicas de análisis multivariable que puedan ser de interés en el análisis y control de procesos.



7. Artículos

Este capítulo recoge los artículos publicados aceptados por la *Comissió de Doctorat* para la presentación de la tesis como compendio de publicaciones.

- I) Analytical control of an esterification process by NIR spectroscopy.
- II) Analytical control of an esterification batch reaction between glycerine and fatty acids by near-infrared spectroscopy.
- III) Use of multivariate curve resolution to monitor an esterification reaction by near-infrared spectroscopy.
- IV) Application of multivariate curve resolution to chemical process control of an esterification reaction monitored by near-infrared spectroscopy



ANALYTICAL CONTROL OF AN ESTERIFICATION PROCESS BY NIR
SPECTROSCOPY

M. Blanco, R. Beneyto, M. Porcel y M. Castillo

Near Infrared Spectroscopy, Proceedings of the 11th
International Conference (Córdoba, 2004), industrial section,
pp. 1021-1026. NIR Publications.

ANALYTICAL CONTROL OF AN ESTERIFICATION PROCESS BY NIR SPECTROSCOPY

M. Blanco^a, R. Beneyto^b, M. Porcel^a and M. Castillo^a

^a *Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain*

^b *Development and Quality Assurance Department, ICI Spain S.A. (Mevisa Site). Ctra Hostalric a Tossa, Km. 1.8, 08495 Fogars de la Selva, Barcelona, Spain*

Introduction

Process control in the chemical industry entails determining chemical parameters related to the species involved in the chemical reaction concerned. Such parameters continue to be determined largely by use of wet chemical methods. The esterification of glycerol by medium chain fatty acids to obtain medium chain triglycerides (MCTs) is usually monitored via the acid value (AV) and the hydroxyl value (OHV), which are measures of the amount of free, unreacted fatty acid and glycerol, respectively, remaining in the reaction medium. Both parameters are measured throughout the process and should be minimized when the product to be obtained is the corresponding triglyceride. They are also measures of appropriate reaction development, and used to identify the end-point of the reaction and to connect with the subsequent steps of the manufacturing process (refining). One other parameter to be monitored is the amount of water removed during the reaction in order to displace the equilibrium to the main reaction product.

Determining the above-described parameters entails using labour-intensive analytical (volumetric) methods; consuming some sample; using expensive, toxic reagents and solvents; and generating waste. There is thus a high interest in developing fast, reliable analytical methods circumventing these shortcomings¹. Such methods should allow one to reduce the product cycle time and avoid the production of out-of-specification batches as a result of problems arising during the reaction step.

Near-infrared spectroscopy (NIRS) constitutes a potential choice for the at-, in- and on-line monitoring of industrial chemical process²⁻⁴. This is a straightforward, rapid, non-destructive technique that requires no sample pretreatment and is applicable to all types of sample with a view to determining both chemical (water, active principles) and physical parameters (viscosity, density).

Notwithstanding its advantages, NIRS provides spectral data of such complexity that it requires the joint use of some multivariate calibration technique; especially effective in this respect are principal component regression (PCR) and partial least-squares regression (PLSR).

In this work, we developed a NIRS–multivariate calibration method for the simultaneous at-line determination of the AV, OHV and water content in an industrial process for obtaining medium chain triglycerides (MCTs). At-line analysis on the plant site can then be the first step towards continuous process control.

Experimental

Apparatus and software

NIR transreflectance spectra were recorded on a FOSS NIRSystems 6500 spectrophotometer equipped with a Rapid Content Analyser (RCA) module. The instrument was controlled via the Vision v. 2.51 software package.

The experimental set-up also included a LabMax reactor (Mettler Toledo), and the Metrohm 701 KF Titrino and 716 DMS titrators.

Multivariate calibration models were constructed using the software Unscrambler v. 7.8 from CAMO (Trondheim, Norway).

Esterification of glycerol by medium chain fatty acids

The process studied was the esterification of glycerol by a mixture of caprylic and capric acids. The process samples used to construct and validate the calibration models were obtained from 8 different manufacturing batches of Uniqema (Mevisa Site, Spain), and 2 batches from the laboratory scale reactor (LabMax). Samplings were done at different reaction times and spanned the ranges 1-150 mg KOH/g for AV, 3-120 mg KOH/g for OHV and 0.01-0.8 % for water content. All samples were subjected to the reference methods endorsed by the Pharmacopoeia Europea⁵, which are summarized in Table 1.

Table 1. Summary of the reference methods.

Parameter	Sample	Method	Solvent	Reagent(s)	Time
AV	0.1-20 g	Acid-base titration	25 ml ethanol/xylene	0.1-0.5 M KOH ethanolic solution	5 min
OHV	1-5 g	1) Acetylation	Pyridine	Acetic anhydride	90 min
		2) Acid-base titration	25 ml ethanol	0.5 M KOH ethanolic solution	
Water	1-5 g	Karl-Fischer titration	10 ml Hydranal Solvent	Hydranal composite 5	5 min

NIR spectra were recorded in the transreflectance mode at room temperature by using an optical glass cuvette and a gold reflector of 0.5 mm pathlength. Each individual spectrum was the average of 32 scans done at 2 nm intervals over the wavelength range 1100–2500 nm. The amount of sample used was 2-3 g and the reference was made by inserting the gold reflector into the empty cuvette, centred on the sample window with a centring iris. Each sample was used to record two spectra in order to obtain an average value that was used to determine the different parameters studied.

Data processing

The samples were homogeneous, but its handling could result in the formation of small air bubbles and hence in the scattering of light and alterations in the spectral baseline and slope. These effects could lead to spurious results for the parameters of interest and required the use of a mathematical treatment prior to construction the calibration models. In this work, two different types of treatment were used: the standard normal variate (SNV) and derivatives. Spectrum derivatives were obtained using the Savitzky-Golay algorithm with a second-order polynomial and a window size of 11 points.

The pre-processed data thus obtained were subjected to the PCA and PLS1 algorithms included in Unscrambler. PLS1 models were constructed by cross-validation, using as many segments as samples employed for calibration (leave-one-out procedure). The optimum number of PLS components was determined in order to minimize the sum of the squared differences between the reference and the determined parameter:

$$PRESS = \sum_{i=1}^m (\hat{y}_i - y_i)^2 \quad (1)$$

where m is the number of samples, y_i the parameter value provided by the reference method and \hat{y}_i the value calculated with the PLS1 model.

The goodness of the results provided by the different PLS1 models was assessed in terms of relative standard error:

$$RSE(\%) = \sqrt{\frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (y_i)^2}} \times 100 \quad (2)$$

and designed as RSEC(%) and RSEP(%) for calibration and prediction, respectively.

Results and discussion

Figure 1 shows the variation of the NIR spectra (following SNV pre-treatment) during the reaction step in the industrial manufacturing of an MCT. As can clearly be seen, the spectrum changed markedly in the 1900–2200 nm region, which contains the O–H combination bands for water and those for the hydroxyl groups in glycerol.

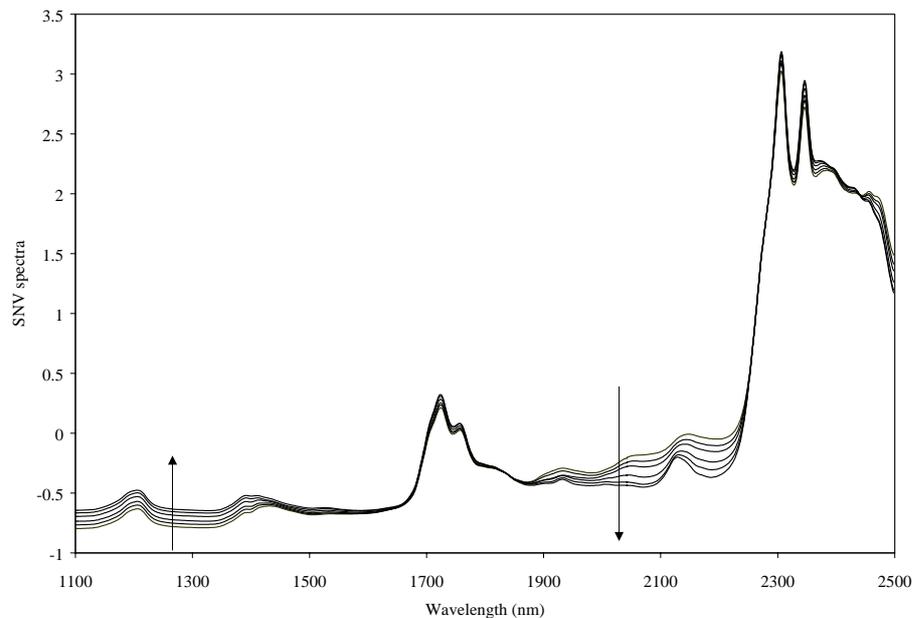


Figure 1. NIR reaction spectra as recorded in SNV mode. The arrowheads indicate the temporal evolution of the spectra.

Figure 2 shows the plot of scores of the principal component analysis (PCA) for three industrial processes, which is linked to the chemical composition variation. The arrowhead indicates the evolution of the reaction and its direction coincides with the decrease in the target parameters (AV and OHV). The samples clustered according to the reaction evolution in the first *versus* second PC plot. At the end, samples were all clustered and the point scatter in the scores plot was very low. Based on these results, principal component analysis can be an effective choice for monitoring the reaction and identifying its end-point.

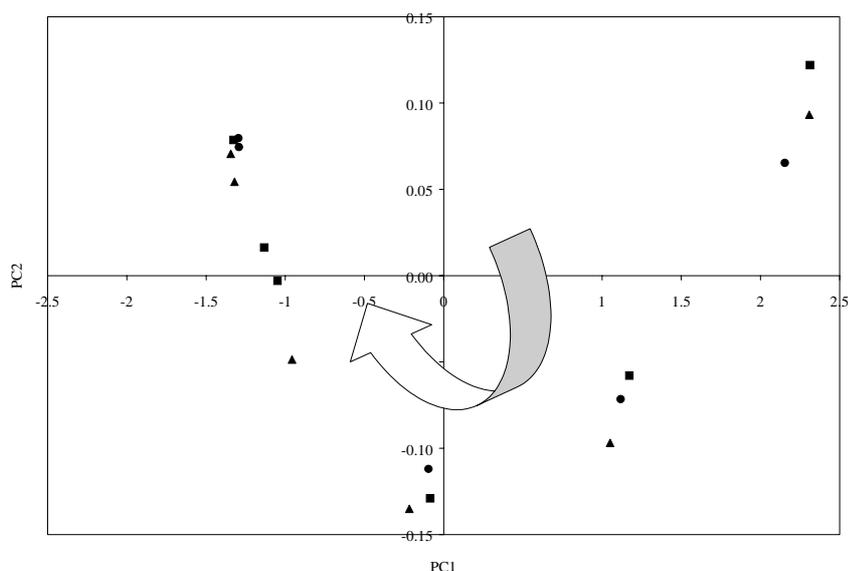


Figure 2. Plot of first *versus* second scores obtained in the PCA of three industrial processes.

Calibration models

PLS1 models were constructed by using various wavelength ranges and pre-treatments. Those providing the best results were shown in Table 2.

Table 2. Relative standard errors of calibration and prediction for the best PLS1 models for AV, OHV and water.

Parameter	Pre-treatment	Wavelength (nm)	Calibration		Prediction	
			Samples	RSEC(%)	Samples	RSEP(%)
AV ^a	SNV	Full spectrum	31	2.2	53	2.5
OHV ^b	SNV	1800-2100	31	5.4	49	6.0
Water ^b	1 st D	1800-2100	25	5.3	40	9.5

^a Number of PLS components used in the model = 2

^b Number of PLS components used in the models = 3

As can be seen, the NIR predicted values obtained with the PLS1 models were quite consistent with the reference values. The figures of merit of the models (Table 3) reveal the absence of systematic errors (slopes and intercepts were not significantly different from unity and zero, respectively).

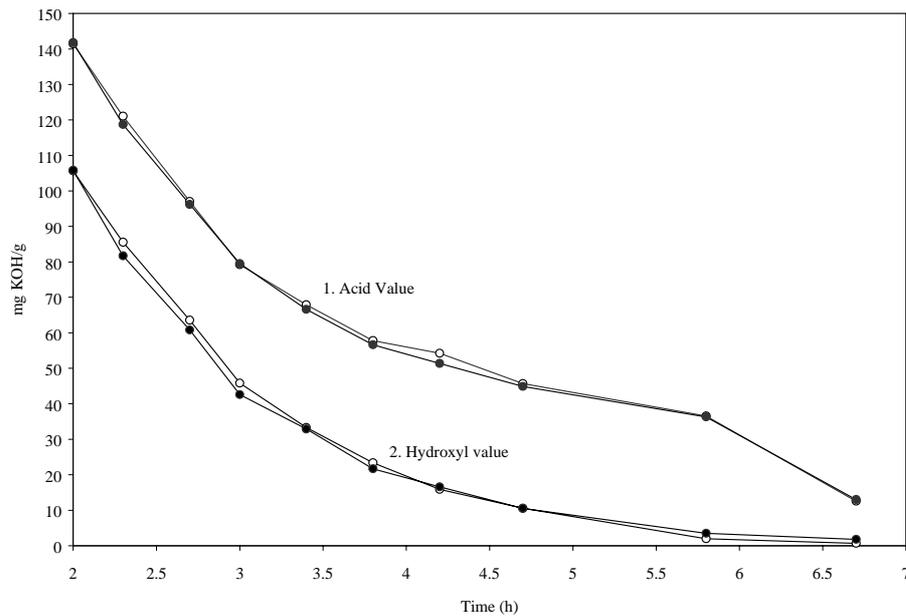
Table 3. Figures of merit and parameters ranges.

Parameter	Set	Range	Slope	Intercept	R ²
AV	Calibration	1-150 mg KOH/g	0.999 ± 0.012	0.056 ± 0.895	0.999
	Prediction	1.5-146 mg KOH/g	1.009 ± 0.009	-0.016 ± 0.573	0.999
OHV	Calibration	2-117 mg KOH/g	0.995 ± 0.027	0.164 ± 1.340	0.995
	Prediction	3-116 mg KOH/g	1.024 ± 0.021	0.173 ± 0.809	0.995
Water	Calibration	0.01- 0.84 %	0.995 ± 0.029	0.001 ± 0.008	0.995
	Prediction	0.02- 0.80 %	0.990 ± 0.036	0.004 ± 0.006	0.990

Process control

Figures 3a and 3b show the predicted profiles for one of the esterification reactions performed in the LabMax reactor. As can be seen, the predicted curves (Fig. 3a) for AV and OHV fitted the results of the volumetric reference methods very closely. On the other hand, the prediction of the water content (Fig. 3b) was not too good. This may have been the result of all the samples in the set having water contents near the lower limit of the calibration curve and hence very close to the limit of quantitation of the reference method.

a)



b)

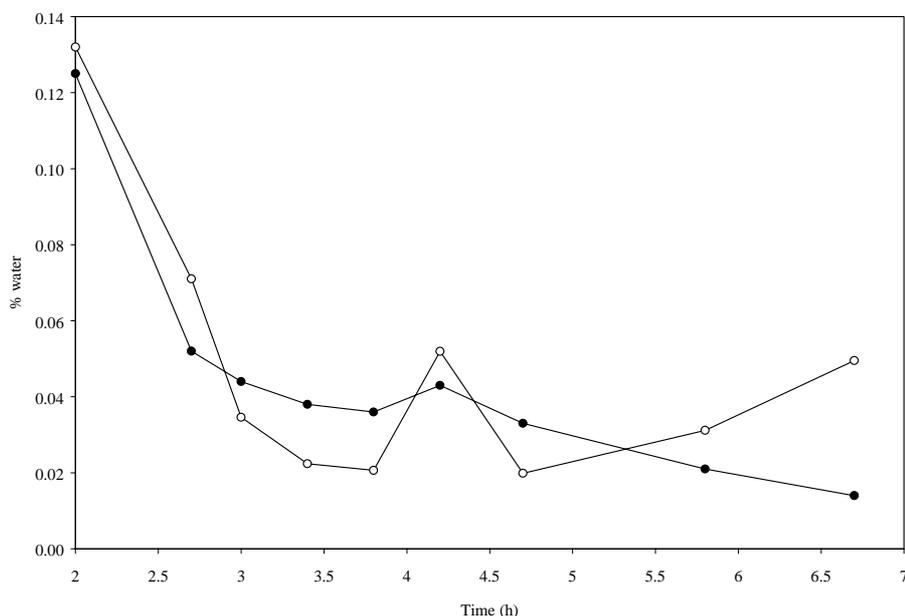


Figure 3. a) Reaction profiles for (1) AV and (2) OHV during the esterification reaction. b) Reaction profile for the water content during the esterification reaction. White (○) and black (●) circles correspond to NIR predicted values and reference values, respectively.

Conclusions

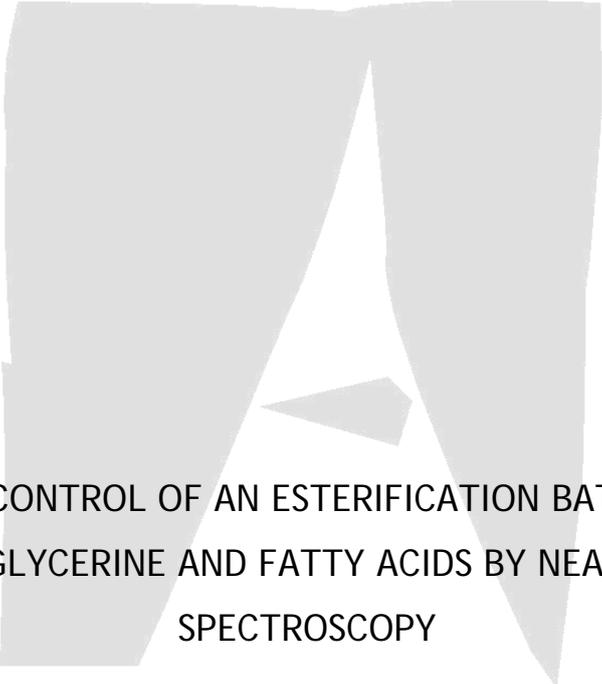
A single NIR spectrum, a small amount of sample and PLS1 calibration models allow one to determine AV, OHV and the water content in an industrial manufacturing process. The analysis time is substantially reduced (to less than 5 min), which is very important when the end of the reaction step must be identified (cycle time reduction). Finally, the proposed NIRS method avoids the use of hazardous solvents and reagents, as well as the production of waste.

Acknowledgements

The authors are grateful to Spain's DGICyT for funding this research within the framework of Project BQU 2000-0234.

References

1. M. Blanco, and I. Villarroya, *Trends Anal. Chem.*, **21**(4), 240 (2002).
2. R. A. Heikka, K. T. Immonen, P. O. Minkkinen, E. Y.O. Paatero, and T. O. Salmi, *Anal. Chim. Acta*, **349**, 287 (1997).
3. P. Geladi, and P. Åberg, *J. Near Infrared Spectrosc.*, **9**, 1 (2001).
4. D. L. Curtin, *At-Process*, **1**(2), 90 (1995).
5. "Medium-Chain Triglycerides" in *European Pharmacopoeia*, 4th Edition, Directorate for the Quality of Medicines, Strasbourg, France, (2002).



ANALYTICAL CONTROL OF AN ESTERIFICATION BATCH REACTION
BETWEEN GLYCERINE AND FATTY ACIDS BY NEAR-INFRARED
SPECTROSCOPY

M. Blanco, R. Beneyto, M. Castillo y M. Porcel

Analytica Chimica Acta 521 (2004) 143-148

Analytical control of an esterification batch reaction between glycerine and fatty acids by near-infrared spectroscopy

Marcelo Blanco^{a,*}, Rafael Beneyto^b, Miguel Castillo^a, Marta Porcel^b

^a Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

^b Development and Quality Assurance Department, ICI Spain S.A. (Mevisa Site), Ctra Hostalric a Tossa, km 1.8, 08495 Fogars de la Selva, Barcelona, Spain

Received 13 April 2004; received in revised form 1 June 2004; accepted 1 June 2004

Available online 22 July 2004

Abstract

Near-infrared spectroscopy was used to control an esterification reaction between glycerine and middle- or long-chain fatty acids performed in a laboratory-scale reactor. The process involves the initial formation of monoglycerides, which is followed by that of di- and triglycerides as well as transesterification. Establishing the end point of the process is critical with a view to ensuring that the end product will have the composition required for its intended use. PLS calibration was applied to industrial and laboratory-scale batch samples, and laboratory samples were additionally used to extend calibration ranges and avoid correlation between the concentration of the batch samples. In this way, PLS calibration models for glycerine, fatty acids, water, and mono-, di- and triglycerides, were constructed. The proposed method allows the reaction to be monitored in real time, thereby avoiding long analysis times, excessive reagent consumption and the obtainment of out-of-specification products.

© 2004 Elsevier B.V. All rights reserved.

Keywords: NIR; Control esterification process; Multivariate calibration

1. Introduction

The esterification reaction between glycerine and middle- or long-chain fatty acids is the basis for the production of esters in the food, cosmetic and pharmaceutical industry. The reaction starts with the formation of a monoglyceride, which is further esterified to a diglyceride and then to a triglyceride; the process involves simultaneous transesterification reactions and yields a mixture of esters, mono-, di- and triglycerides in a proportion that depends on that between the reactants and on the reaction conditions. When the monoglyceride is the target end product, the reaction must be performed in excess of glycerine; the process also yields di- and triglycerides, albeit in a lower proportion. The use of the end product thus obtained will depend not only on its physical properties (e.g. density, viscosity), but also on its chemical composition; its characteristics are established in terms of a series of parameter values and specifications to be met.

It is therefore essential to accurately determine the point where the reaction product possesses the required properties or composition and to be able to obtain such information in real time. To this end, production plant samples are collected for analysis throughout the process and the results obtained are used to decide whether the process must be stopped or allowed to continue. The time taken to collect such information can range from a few minutes to several hours; because the results are critical, processes are normally designed to accommodate this delay, so they involve extended production cycles.

Wet chemistry provides effective methods for process control based on classical analytical techniques; for example, the acid value [1] and the hydroxyl value [1] are frequently used to control industrial esterification reactions. However, the information obtained (viz. the variation of these two parameters) provides no clear picture of the development of the reaction; this, together with the time-consuming nature of the hydroxyl value determination, prevents the adoption of correct, timely decisions. A truly scientific approach to the problem requires the use of fast, efficient analytical techniques to monitor the process in order to avoid the production of defective batches or the need

* Corresponding author. Tel.: +34 93 581 13 67; fax: +34 93 581 2379.

E-mail address: marcel.blanco@uab.es (M. Blanco).

for corrective measures that increase costs and decrease productivity.

Spectroscopic methods constitute an effective alternative to wet chemical methods; in recent years, NIR spectroscopy has gained increasing acceptance as a choice for industrial process control. Its non-invasive, non-destructive nature, the need for no reagents, the expeditiousness with which spectra can be recorded, its low maintenance costs and its suitability to a variety of situations, among other advantages, have turned it into a major choice for the at-, on- and in-line monitoring of industrial chemical processes [2–8]. While some authors have used it to determine acid and hydroxyl values [7], applications of NIR spectroscopy in this field are still scant.

In this work, a new method based on the use of NIR spectroscopy and various PLS1 calibration models was developed for the simultaneous determination of all the ingredients of an esterification reaction (viz. glycerine, glycerides, fatty acids and water). The industrial process was reproduced in the laboratory with a view to its subsequent implementation on the production scale.

2. Experimental

2.1. Apparatus and software

A LabMax laboratory reactor from Mettler Toledo (11) was used to conduct laboratory-scale reactions (Fig. 1).

NIR transmittance spectra were recorded on a FOSS NIRSystems 6500 spectrophotometer equipped with a rapid content analyser (RCA) module. The instrument was controlled via the Vision v. 2.51 software package.

A Hewlett-Packard HP5890 Series II gas chromatograph equipped with a flame ionization detector and furnished with a Supelco SPB-1701 fused silica capillary column (15 m ×

0.25 mm, 0.25 μm film thickness) was used. The experimental set-up included a Metrohm 716 DMS Karl Fischer titrator.

Multivariate calibration models were constructed with the aid of the Unscrambler v. 8.0 software suite, from CAMO (Trondheim, Norway).

2.2. Samples

The samples were collected from esterification reactors that contain glycerine and a mixture of caprylic and capric acid in a 95:5 ratio. The samples used to construct and validate the calibration models were obtained from three different sources, however, their raw materials were all from the same source. An overall 81 samples were prepared, some of which were used for calibration and the others for external validation.

• Industrial samples

The industrial process involved the use of a 35% glycerine excess with respect to the amount theoretically needed to obtain the monoglyceride. The mixture was maintained at $(220 \pm 5)^\circ\text{C}$ under a continuous stream of nitrogen. Samples from different batches were withdrawn from the reactor throughout the process.

• Laboratory-scale samples

The industrial process was replicated in a laboratory reactor (Fig. 1). An overall six processes were conducted using different excess proportions of glycerine (see Table 1). Samples were withdrawn at different times from the reactor during each process.

• Laboratory samples

These were obtained by spiking process samples with variable, accurately weighed amounts of the available pure products (glycerine, water, triglycerides and the acid mixture). The samples thus obtained spanned the concentration range of species required to construct the models.

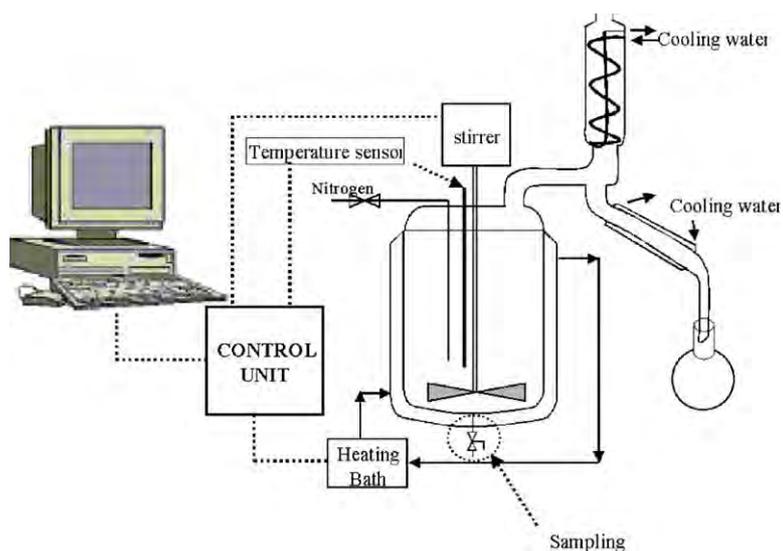


Fig. 1. Schematic drawing of lab-scale reactor.

Table 1
Composition of reaction runs in the laboratory reactor

Run	Glycerine excess (%)	Glycerine (%)	Caprylic acid (%)	Capric acid (%)
1	10	41	56	2.8
2	28	47	51	2.1
3	28	47	52	2.1
4	35	50	47	2.8
5	35	50	47	2.8
6	35	50	48	2.5

2.3. NIR spectra

NIR spectra were recorded in the transmittance mode at room temperature, using a quartz vessel and a gold reflector of 0.5 mm path length. Each individual spectrum was the average of 32 scans done at 2 nm intervals over the wavelength range 1100–2500 nm.

2.4. Reference methods

Water was determined by titration with the Karl Fischer reagent. The fatty acid content was obtained by titration with ethanolic KOH, using an 1:1 ethanol/xylene mixture and the method officially endorsed by the European Pharmacopoeia [1]. The glycerine and glyceride contents were determined by gas chromatography using methyl palmitate as internal standard and a temperature ramp of 7 °C/min from 120 to 270 °C. The percent content in each compound was obtained from the area fraction—corrected for the response factor of each analyte. The response factors for the analytes available in pure form were determined by calibration from standard solutions; those of the analytes for which no standard was available (viz. caprylic–capric mixed glycerides) were determined by interpolation of their effective carbon number into the straight line obtained by plotting the response factor against the effective number of carbon atoms for the species available in pure form [9].

2.5. Data processing

Spectral data were subjected to various treatments including the standard normal variate (SNV), and the first and second derivatives (which were obtained using the Savitzky–Golay algorithm with a second-order polynomial and a window size of 11 points).

The pre-processed data obtained with the reference methods were modelled using PLS1 algorithms included in Unscrambler. The models were constructed by cross-validation using the leave-one-out method. The optimum number of PLS components was taken to be that minimizing the sum of the squared differences between the reference and the determined parameter:

$$\text{PRESS} = \sum_{i=1}^m (\hat{y}_i - y_i)^2$$

The goodness of the results provided by the different PLS models was assessed in terms of the relative standard error (RSE) and of the root mean square error (RMSE):

$$\text{RSE} (\%) = \sqrt{\frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (y_i)^2}} \times 100$$

$$\text{RSME} = \sqrt{\frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{n}}$$

and designated RSEC (%) / RMSEC for calibration and RSEP (%) / RMSEP for external validation.

3. Results and discussion

The reactant mixture formed an upper layer consisting of an acid mixture and a lower layer initially consisting of glycerine and subsequently including glycerides; the two layers formed a single, homogeneous one, as the reaction developed. The samples studied were withdrawn from the homogeneous medium and cooled below 180 °C—where the reaction kinetics is slow—in order to stop the reaction.

As can be seen in Fig. 2, the main reaction components (viz. glycerine, water and glycerides) exhibited marked spectral differences. Thus, the spectra for glycerine esters were very similar and the greatest difference between the triglyceride and the glyceride mixture—which consisted largely of mono- and diglyceride—lay in the regions 1400–1600 and 2000–2200 nm; such regions correspond to the first overtone and the combination bands for the O–H bond, respectively, which accounts for the weak absorption of the triglyceride in them. Also, these two products and the fatty acid exhibited two absorption maxima in the 1700–1800 nm region corresponding to the first overtone of C–H bonds, which were very weak for glycerine and absent for water.

Glycerine and water were the two studied analytes with the highest absorptivity and the most characteristic and easily identified bands in the NIR region. Thus, glycerine exhibits strong absorption between 2000 and 2200 nm due to the combination bands of the OH bond, while water exhibits two absorption maxima corresponding to the first overtone (1450 nm) and combinations bands (1950 nm), respectively.

As can be seen in Fig. 3, the bands in the NIR spectrum changed during the process and appeared in the same order as the samples were withdrawn from the reactor. The greatest—and most readily identified—spectral difference was the decreased absorption in the 2000–2200 nm region, corresponding to the OH bond in glycerine. This band did not disappear altogether owing to the excess of glycerine used and the absorption of the bonds in the mono- and diglyc-

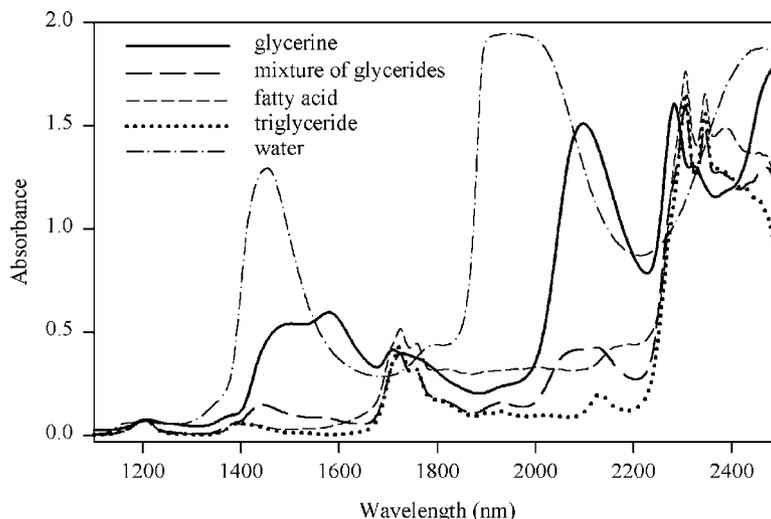


Fig. 2. NIR spectra for reagents and reaction products. Mixture of glycerides is 60/35/5% of mono-, di-, and triglyceride, respectively.

eride. This also holds for the weak band between 1550 and 1600 nm, which corresponded largely to glycerine.

The absorption between 1900 and 2000 nm corresponded to combination bands for water, which decreased in strength through constant distillation of the water produced in the esterification reaction.

The absorption in the 1100–1350 and 2300–2500 nm regions corresponded to the second overtone and the combination bands for C–H bonds, respectively.

All other absorption bands in the NIR spectrum were difficult to assign. Such was especially the case with the band between 1425 and 1500 nm, where both glycerine, water and the glycerides absorb (see Fig. 2).

3.1. Calibration and external validation

Laboratory samples were prepared by adding glycerine, triglyceride, water and fatty acids in known amounts to pre-

viously analysed glyceride mixtures obtained from laboratory syntheses. In this way, correlation between samples was reduced and the concentration range used for calibration spanned, which enabled the accurate prediction of samples departing from the values held as normal.

In order to reduce correlation between the concentrations of the different compounds to be determined, synthetic processes were conducted using variable excess amounts of glycerine and also variable—albeit to a lesser extent—acid concentration ratios. Table 1 shows the initial composition of the different sample batches.

PLS1 models were constructed by splitting the 81 initial samples into a calibration set and an external validation set; the latter was used to determine predictive capacity. The samples to be included in each set were chosen in such a way that they would span the concentration range for each analyte, so the sets were slightly different for each component.

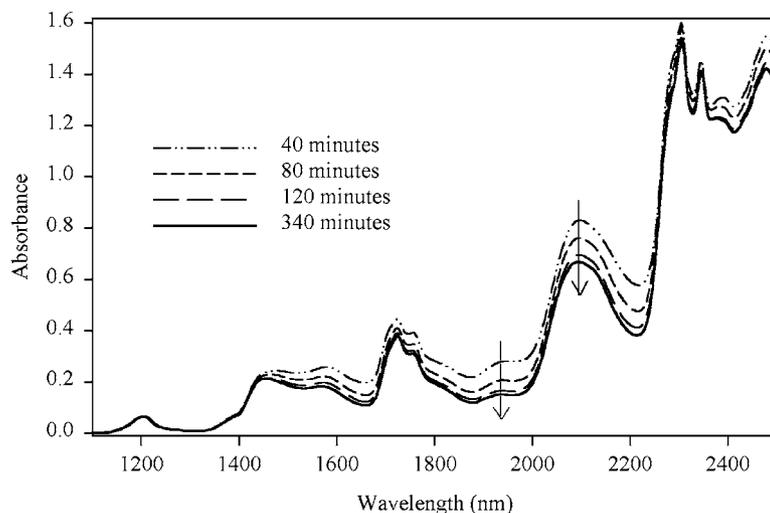


Fig. 3. NIR reaction spectra. The arrowsheads indicate the temporal evolution of the spectra.

Table 2
PLS1 models for glycerine, fatty acid, water and glycerides

Component	Spectral range (nm)	Spectral mode	PLS components	Calibration		External validation	
				RSEC(%)	RMSEC	RSEP(%)	RMSEP
Glycerine	1100–2500	SNV	3	6.1	1.1	6.1	1.4
	1850–2150	First-derivative	3	3.8	1.0	5.7	1.4
	1100–2150	First-derivative	3	3.8	1.0	4.9	1.2
Fatty acid	1100–2500	SNV	4	4.8	0.5	11.3	0.7
	1100–2500	First-derivative	4	8.0	0.8	20.4	1.2
Water	1100–2500	SNV	3	3.6	0.06	4.2	0.07
	1850–2150	First-derivative	2	4.8	0.07	5.0	0.08
	1100–2500	First-derivative	1	5.8	0.09	5.4	0.08
Monoglyceride	1100–2500	First-derivative	5	1.0	1.3	7.8	3.5
	1100–1850, 2150–2400	First-derivative	5	4.5	1.9	5.2	2.3
	1100–2500	SNV	5	6.5	1.7	7	1.8
Diglyceride	1100–1850, 2150–2400	SNV	4	3.7	1.7	6.5	1.7
	1100–1850, 2150–2400	First-derivative	3	5.7	1.8	8.15	2.1
Triglyceride	1100–2500	SNV	5	8.2	0.2	17.6	0.6
	1100–1350, 1494–1850, 2098–2500	First-derivative	6	10.7	0.3	24.7	0.7
	1100–1850, 2150–2400	First-derivative	6	7.9	0.2	24.7	0.7

Explained calibration variance of the best models is higher than 94% for each component. The best model for each component is in bold.

In constructing the calibration models, we examined the influence of the wavelength range and chose the spectral treatments resulting in the most accurate and robust models. Thus, we used the whole spectrum (1100–2500 nm) for every component and then other, specific ranges for each individual compound. We also examined various spectral treatments including SNV, and the first and second derivative. The models providing the best results and their RSE (%) and RMSE values are listed in Table 2. As can be seen, the models yielding the best results were those based on SNV or the first spectral derivative; their quality increased (i.e. their RSEP and RMSE decreased) with increasing analyte concentration and increasing absorption in the NIR region. Thus, the models for water were quite good despite its low concentration by effect of its high molar absorptivity and the low contribution of other species in its absorption region. The situation was similar for glycerine, for which the whole spectrum and its first derivative provided the best results.

The glyceride models were more complex owing to their high spectral similarity and the fact that relative errors were strongly concentration-dependent. In fact, the models for the triglyceride provided increased errors relative to the mono- and diglycerides with all spectral treatments and in all regions. The best glyceride models were those excluding the absorption region for water and glycerine (viz. 1850–2150 nm).

For evaluate the best model for each component (marked in bold in Table 2), a plot of predicted NIR values versus reference values is done. The figures of merit of the least squares regression of these plots were calculated, for calibration and validation sets (see Table 3). These results show that the different PLS models are acceptable for the analytical control of this chemical process.

3.2. Process control

The acid and hydroxyl values for industrial processes obtained with wet chemical methods provide no information of use with a view to examining the temporal course of the reaction as such values virtually level off after 3 h of reaction. In order to stop the reaction when the glyceride composition needed to obtain the desired product is reached, one must be able to control every component of the process.

The calibration models for the different components of the reaction mixture were used to monitor new synthetic processes in the LabMax that involved standard processes not employed to construct or validate any model. Fig. 4 shows the NIR predictions for such a process; although the

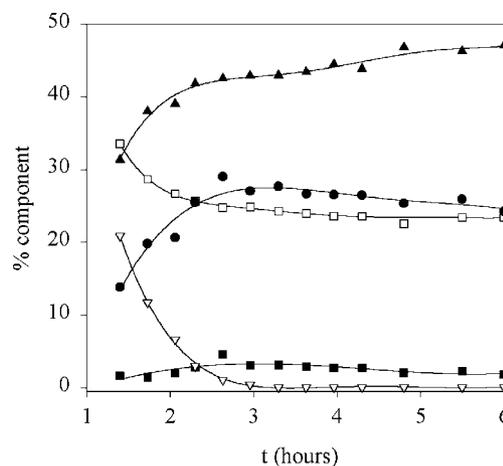


Fig. 4. Evolution of different components in a batch: (□) glycerine, (▲) monoglyceride, (●) diglyceride, (■) triglyceride and (▽) fatty acid.

Table 3

Figures of merit for regression line of predicted NIR values versus reference values for each analyte

Parameter	Set	Concentration range (%)	Slope	Intercept	R ²
Glycerine	Calibration	0.5–41.5	0.99 ± 0.04	0.2 ± 0.8	0.988
	Ex. validation	14.2–29.5	1.0 ± 0.1	0 ± 3	0.93
Fatty acid	Calibration	0.1–31.3	1.00 ± 0.02	0.0 ± 0.2	0.996
	Ex. validation	0.1–24.4	1.04 ± 0.03	0.0 ± 0.2	0.988
Water	Calibration	0.1–4.4	0.99 ± 0.02	0.00 ± 0.03	0.998
	Ex. validation	0.2–4.3	1.01 ± 0.02	0.00 ± 0.04	0.997
Monoglyceride	Calibration	23.4–60.4	0.94 ± 0.07	2 ± 3	0.94
	Ex. validation	32.1–51.2	1.0 ± 0.2	1 ± 9	0.8
Diglyceride	Calibration	4.5–37.7	0.96 ± 0.07	1 ± 2	0.96
	Ex. validation	7.8–35.8	0.9 ± 0.1	2 ± 3	0.9
Triglyceride	Calibration	0.1–5.9	0.95 ± 0.08	0.1 ± 0.2	0.95
	Ex. validation	0.3–5.8	0.9 ± 0.1	0.1 ± 0.4	0.8

Concentration range, correlation coefficient and the confidence limits for slope and intercept are also included.

glycerine and fatty acid contents actually levelled off after about 3 h, the glyceride content continued to change—to increase in the case of the monoglyceride and to slightly decrease in the case of the di- and triglyceride—beyond this point. These small variations in the glyceride content dictate whether the resulting product will meet the specifications. The specifications for this product restrict the maximum and minimum relative glyceride contents (i.e. those present after glycerine is removed from the medium); thus, the monoglyceride content should exceed 61% and the triglyceride content not exceed 4% upon removal of the glycerine. Fig. 5 shows the relative proportion of glycerides obtained from the NIR predictions for the reaction mixture samples of the new syntheses; as can be seen, the monoglyceride content obtained after the acid and glycerine contents levelled off (i.e. after 3 h) was lower than needed. It was only after about 4.5 h of reaction that such a content met the specifications and the reaction could thus be stopped (Table 3).

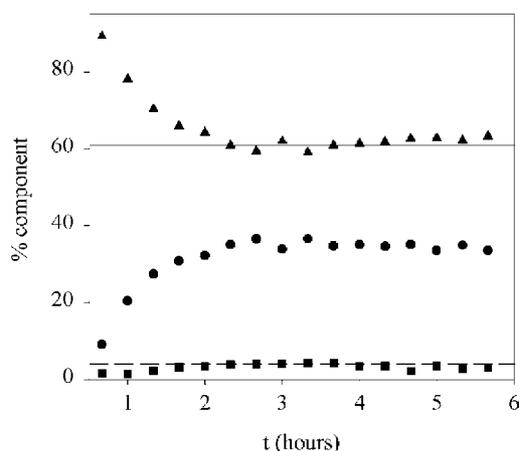


Fig. 5. Evolution of relative content of glycerides in the same batch. Straight line indicates the minimum content of monoglyceride, dash line indicates the maximum content of triglyceride: (▲) monoglyceride, (●) diglyceride and (■) triglyceride.

4. Conclusions

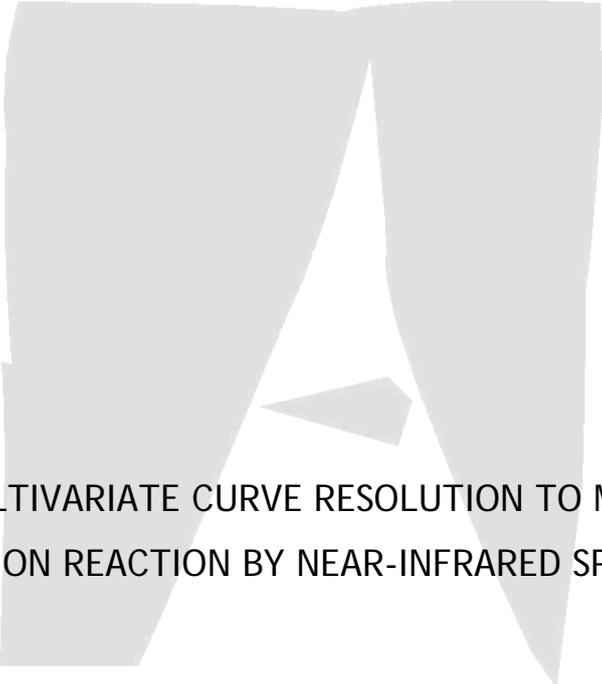
Wet chemical methods are labour-intensive and time-consuming, so they are poor choices for accurately monitoring the compounds involved in the esterification of glycerine with fatty acids. On the other hand, the joint use of NIR spectroscopy and multivariate calibration allows glycerine, fatty acid, water and glycerides to be accurately quantified throughout an industrial manufacturing process. The proposed NIR method avoids the use of hazardous solvents and reagents, and the production of waste—which are obvious environmental and safety assets—also, it substantially reduces analysis times relative to other instrumental techniques such as GC.

Acknowledgements

The authors are grateful to Spain's DGICYT for funding this research within the framework of Project BQU2000-0234. Miguel Castillo wishes to acknowledge additional support from the DURSI of the Generalitat de Catalunya in the form of a scholarship.

References

- [1] Directorate for the Quality of Medicines, Medium-Chain Triglycerides in European Pharmacopoeia, fourth ed., Strasbourg, France, 2002.
- [2] M. Blanco, D. Serrano, *Analyst* 125 (2000) 2059–2064.
- [3] C. Coffey, B.E. Cooley, D.S. Walker, *Anal. Chim. Acta* 395 (1999) 335–341.
- [4] K.L. Miller, D. Curtin, in: *Proceedings of the 49th Annual Symposium on Instrumentation for the Process Industries*, 1994, pp. 53–57.
- [5] W.D. Mockel, M.P. Thomas, in: *Proceedings of SPIE-The International Society for Optical Engineering*, 1992, p. 1681.
- [6] R.A. Heikka, K.T. Immonen, P.O. Minkkinen, E. Paatero, T.O. Salmi, *Anal. Chim. Acta* 349 (1997) 287–294.
- [7] P. Geladi, J. Forsström, *J. Chemom.* 16 (2002) 329–338.
- [8] C.A. McGill, A. Nordon, D. Littlejohn, *Analyst* 127 (2002) 287–292.
- [9] M. Kállai, Z. Veres, J. Balla, *Chromatographia* 54 (2001) 511–517.



USE OF MULTIVARIATE CURVE RESOLUTION TO MONITOR AN
ESTERIFICATION REACTION BY NEAR-INFRARED SPECTROSCOPY

M. Blanco, M. Castillo, R. Beneyto y M. Porcel

Spectroscopy Letters **38** (2005) 825-837

Use of Multivariate Curve Resolution to Monitor an Esterification Reaction by Near-Infrared Spectroscopy

Marcelo Blanco and Miguel Castillo

Department of Chemistry, Faculty of Sciences, Autonomous
University of Barcelona, Barcelona, Spain

Rafael Beneyto

Development and Quality Assurance Department, ICI Spain S.A.
(Mevisa Site), Barcelona, Spain

Marta Porcel

Department of Chemistry, Faculty of Sciences, Autonomous
University of Barcelona, Barcelona, Spain

Abstract: Multivariate curve resolution–alternating least squares (MCR-ALS) methodology was applied to near-infrared (NIR) spectra with view to estimating the reaction profile for the esterification of a mixture of caprylic and capric acids with glycerol. The reaction was conducted in excess glycerol, so the major products were the monoglyceride and diglyceride; the triglyceride was obtained in considerably lower proportions. Esterification processes were performed in a laboratory-scale reactor from which samples were withdrawn to record NIR spectra. Some samples were also analyzed by gas chromatography and acid-base titration in order to determine the composition of the reaction mixture. Concentration and spectral profiles were obtained by using the MCR-ALS algorithm. Subsequently, concentration values and the pure spectra of reagents and the triglyceride were used to refine the models. The spectra obtained were processed with MCR-ALS in new esterification batches to obtain their concentration profiles.

Keywords: Esterification reaction, MCR-ALS, NIR spectroscopy, reaction profiles

Received 13 November 2004, Accepted 27 January 2005

This paper was by special invitation as a contribution to a special issue of the journal entitled “Quantitative Vibrational Spectrometry in the 21st Century.” This special issue was organized by Professor Miguel de la Guardia, Professor of Analytical Chemistry at Valencia University, Spain.

Address correspondence to Marcelo Blanco, Department of Chemistry, Faculty of Sciences, Autonomous University of Barcelona, Bellaterra, Barcelona 08193, Spain.
E-mail: marcel.blano@uab.es

INTRODUCTION

The esterification between glycerol and middle-chain fatty acids represents a reaction of high industrial interest for the production of esters used by the food, cosmetic, and pharmaceutical industries. The reaction starts with the formation of a monoglyceride that is further esterified to a diglyceride and then to a triglyceride; the process involves simultaneous transesterifications. The industrial process can be conducted by using variable initial proportions of the raw materials; the presence of excess acid leads preferentially to the triglyceride, whereas that of excess glycerol gives rise to the mono- and diglyceride as major products and the triglyceride in lower proportions. The process studied in this work was of the latter type: the use of excess glycerol and a mixture of caprylic and capric acids yielded mono-, di-, and triglycerides in addition to residual (excess) glycerol. The monoglyceride was a mixture of esters between glycerol and both acids at positions 1 or 2 (i.e., four different compounds) in variable proportions depending on the relative concentrations of the two acids. The diglyceride consisted of a mixture of glycerol esters at positions 1,1 or 1,2 of caprylic and/or capric acid (7 compounds in all) also in variable proportions. The monoglyceride was taken to be a component consisting of several (1 or 2 caprylic or capric) esters, and so was the diglyceride. In this work, multivariate curve resolution (MCR)^[1] methodology was used to extract relevant information about the temporal evolution of the reaction products with a view to identifying the time at which the process should be stopped in order to obtain the desired ester mixture composition depending on its intended use.

As noted earlier, the fact that the chemical system involves a number of reacting chemical species makes it rather complex. Obtaining a single mixture of the mono- or diglyceride, or a sample consisting solely of monoglyceride (or diglyceride) in the same isomer ratio, is obviously impossible. This hinders the obtainment of the spectra for the two components, which would be highly useful with a view to calculating the concentration profiles by Multivariate curve resolution–alternating least squares (MCR-ALS). The high correlation between the NIR spectra for the compounds formed in the course of the reaction, the rank deficiency of the spectral data matrix, and rotational and intensity ambiguities restrict the applicability of MCR-ALS here. In the absence of noise and other variability sources such as temperature and the acid concentration ratio, the rank of the spectral data matrix should coincide with the number of chemical species involved.^[2] This assumption fails with the studied system, where two of the reaction products (the mono- and diglycerides) exhibit a high spectral correlation and similar concentration profiles. The rank of the matrix for this system is lower than the number of components involved in the reaction. This problem was circumvented by the following methods: augmented matrices,^[3] the NIR spectra available for the pure components, and the concentrations provided by reference methods as MCR-ALS constraints. Whereas the first methodology gave poor results, the latter two

provided useful spectral information for the compounds involved in the process. This methodology allowed us to estimate the near-infrared (NIR) spectra for the mono- and diglyceride. Such spectra were in turn used to determine the concentration profiles for the components involved in new reaction batches.

THEORY

Multivariate curve resolution (MCR) methodology assumes the data fit a linear pattern conforming to Beer's law, based on the following expression (1):

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (1)$$

where \mathbf{D} ($I \times J$) is the experimental data matrix, the rows of which represent spectra in a time sequence; each element d_{ij} in \mathbf{D} corresponds to the absorbance at wavelength j for the sample withdrawn at time i . \mathbf{C} ($I \times N$) contains the concentration profiles for the absorbing species involved in the reaction; each element c_{in} represents the concentration of analyte n in the sample at time i . \mathbf{S}^T ($N \times J$) contains the spectra for the absorbing species involved in the reaction; each element s_{nj} in it corresponds to the absorbance of analyte n at wavelength j . \mathbf{E} ($I \times J$) is the residuals matrix, which contains all information not explained by matrices \mathbf{C} and \mathbf{S}^T , and is included in \mathbf{D} . When the linear model is obeyed, d_{ij} is the combined absorbance at wavelength j of the N species present in sample i :

$$d_{ij} = \sum_n c_{in}s_{nj} + e_{ij} \quad (2)$$

The equation system (1) can be solved by iterative process of the ALS (alternating least squares) algorithm. Although the system has no unique solution, only a few chemically acceptable solutions exist where \mathbf{S}^T corresponds to the spectra for the analytes and \mathbf{C} to their concentration profiles.

MCR-ALS methodology is subject to two types of ambiguity, *viz.* intensity ambiguity,^[4] which means that the estimated profiles may be scaled by some unknown factor, and rotational ambiguity,^[4] which means that the estimated spectrum for any of the components may be an unknown linear combination of the real components. The introduction of the iterative process of the spectra for absorbing species involved in the reaction helps suppress rotational ambiguity^[4,5] arising from the absence of selective spectral regions—a frequent occurrence in NIR spectroscopy.^[6] The introduction of quantitative values in the iterations provides accurately scaled concentration profiles and suppresses intensity ambiguity.^[4,5] A set of feasible solutions can be obtained by using alternating least squares with a series of constraints including unimodality (when the concentration profiles possess a single maximum value), closure (when the sum of all selected components

should equal a constant value) and spectral and concentration non-negativity (when neither the concentrations nor the spectra can take negative values).

The procedure used to solve Eq. (1) therefore started with the determination of the number of components present in \mathbf{D} and with an estimation of the \mathbf{S}^T or \mathbf{C} matrix. There are different tools to estimate the number of components in a mixture, such as PCA, EFA, FSMW-EFA, or singular value decomposition (SVD). These techniques are widely commented in the references.^[4,7-9] EFA is also a useful technique in order to calculate the initial estimates of matrices \mathbf{C} or \mathbf{S}^T necessary to start the iteration process. The use of a matrix \mathbf{D} containing the spectra for more than one process, an augmented data matrix (5), can help solve the problem posed by a rank deficiency.^[3] An \mathbf{S}^T —or the spectra for the pure components, if available—or \mathbf{C} initial estimate is used to start the iterative process in order to calculate new \mathbf{C} and \mathbf{S}^T estimates in each cycle, and simultaneously, the above-described constraints being applied in accordance with the chemical information available. Iterations were stopped when matrix \mathbf{E} was minimized, that is, when the difference between the root mean squared error (RMS) between two consecutive iterations fell below a preset threshold.

$$RMS = \sqrt{\frac{\sum_i \sum_j e_{ij}^2}{n}} \quad (3)$$

The goodness of fit of the model thus obtained was assessed in terms of the percent-calculated variance:

$$\%Var = \frac{\sum_i \sum_j \hat{d}_{ij}^2}{\sum_i \sum_j d_{ij}^2} \times 100 \quad (4)$$

Finally, the quality of the MCR-ALS estimated spectra was evaluated from the similarity criterion:

$$r = \cos \gamma = \frac{S_i^T \hat{S}_i}{\|S_i\| \cdot \|\hat{S}_i\|} \quad (5)$$

EXPERIMENTAL

Apparatus and Software

Esterification processes were conducted in a 1 liter LabMax laboratory reactor from Mettler Toledo.

NIR transmittance spectra were recorded on a FOSS NIRSystems 6500 spectrophotometer equipped with a Rapid Content Analyzer (RCA) module. The instrument was controlled via the Vision v. 2.51 software package.

A Hewlett-Packard HP 5890 Series II gas chromatograph equipped with a flame ionization detector and furnished with a Supelco SPB-1701 fused silica capillary column (15 m \times 0.25 mm, 0.25 μ m film thickness) was used.

Multivariate curve resolution was used in conjunction with alternating least squares with penalty function, using the software GUIPRO (10) as implemented in Matlab.^[11]

Reagents and Samples

Samples were collected from esterification batches containing glycerol and a mixture of caprylic and capric acids in a 95:5 ratio. A glycerol excess of 35% with respect to the amount theoretically required to obtain the monoglyceride was used. The reaction mixture was heated to $220 \pm 5^\circ\text{C}$, which was held for about 6 hr, under a continuous stream of nitrogen. A distillate consisting of water and small amounts of acid was produced during the reaction. Samples from 3 batches, named as A, B, and C, were withdrawn from the reactor throughout the process after the first 40 min for the batches A and B and 80 min for batch C—the time needed for the reaction mixture to become homogeneous. The number of samples in each batch is 15 for batch A, 16 for batch B, and 13 for C.

NIR Spectra

NIR spectra for the reaction mixtures and pure products available were recorded in the transmittance mode at room temperature, using a quartz vessel and a gold reflector of 0.5-mm pathlength, which leads to absorbance units between 0 and 1.8 (inside the linear absorbance response of the instrument). Each individual spectrum was the average of 32 scans done at 2 nm intervals over the wavelength range 1100–2500 nm. Figure 1 shows the spectra for the compounds available (viz. glycerol, the acid mixture and the triglyceride).

Reference Methods

The samples withdrawn from one of the laboratory esterification batches were analyzed by using reference methods. The acid content was determined by titration with ethanolic KOH, using a 1:1 ethanol/xylene mixture as solvent.^[12] The glycerol and glyceride contents were determined by gas chromatography, using methyl palmitate as internal standard and a temperature ramp of $7^\circ\text{C}/\text{min}$ from 120°C to 270°C . The percent content in each compound was obtained by interpolation into a calibration plot of analyte-to-internal standard peak area as corrected for the response factor of each analyte. The response factors for the analytes available in pure form were determined by calibration from standard solutions; those analytes for which no standard was available (viz. caprylic–capric mixed glycerides) were

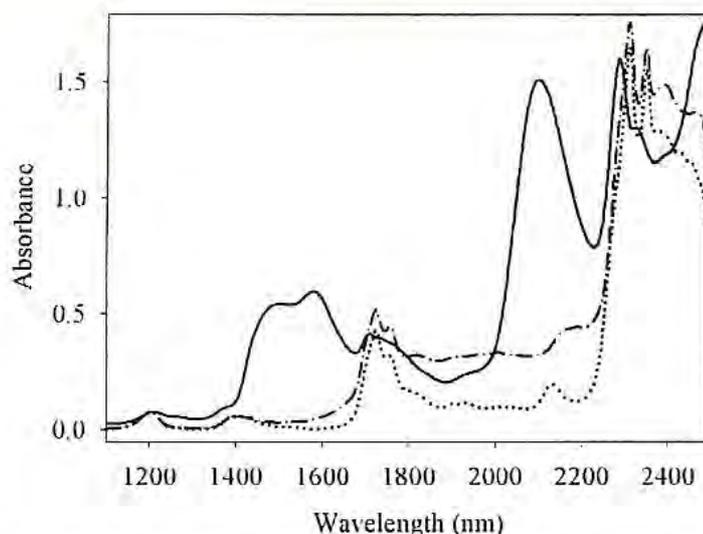


Figure 1. NIR spectra for glycerol (—), the mixture of caprylic and capric acids (---), and the triglyceride (···).

quantified by interpolation of their effective carbon number into the straight line obtained by plotting the response factors against the effective number of carbon atoms for the species available in pure form.^[13]

Data Processing

Spectra data were subjected to the standard normal variate (SNV) treatment in order to suppress potential baseline shifts or scattering caused by some turbid samples. The regions of negative absorbance provided by the SNV treatment were removed by adding a scalar to all spectra in order to eliminate negative spectral values and facilitate application of the spectral non-negativity constraint in the MCR-ALS calculations.

Experimental matrices were subjected to rank analysis using singular value decomposition (SVD) methodology; both singular values and the profile of the corresponding eigenvectors were considered in determining matrix ranks.

Use of MCR-ALS to Calculate Spectra

In this work, MCR-ALS methodology was used to obtain the estimated spectra for those analytes whose experimental spectra could not be recorded owing to the unavailability of the products in pure form. This entailed using the concentrations of the analytes involved in the reaction as determined with a reference method, as well as the NIR spectra for the analytes where available (e.g., those for glycerol, the mixture of caprylic and capric acids, and the triglyceride). This information was used as equality constraints^[10] in every iteration by the methodology included in the software GUIPRO. The initial estimates,

prior to the iteration process by ALS, were constructed by the needle search methodology^[10,14,15] implemented in the software GUIPRO.

MCR-ALS calculations were based on the concentrations for five sample components as obtained with reference methods and the recorded NIR spectra for glycerol, the acid mixture and the triglyceride. The estimated spectra were subsequently used to monitor changes in the reacting species in new esterification batches. The use of pure spectra as initial estimates or as equality constraints in calculating the concentration profiles by MCR-ALS suppresses rotational ambiguity, which in NIR spectroscopy is increased by the low spectral selectivity of the analytes present in the studied samples. The flowchart of Fig. 2 shows the different steps of the application of ALS in this study.

RESULTS AND DISCUSSION

Figure 3 shows the variation of the NIR spectra (subjected to math treatments commented in Section 3.5) during the course of the reaction. The largest spectral difference was the absorption decrease in the 2000–2200 nm region, which corresponded to the O–H bond in glycerol. This band did not disappear completely owing to the presence of excess glycerol and the absorption of O–H bonds in the mono- and diglycerides. The variation of the absorbance at 1200 nm is due to the SNV spectral pretreatment, which increases the small differences in this region.

Application of MCR-ALS to the Body of Spectral Data

The SVD calculated mathematical rank for the matrices representing the different studied batches was 2, as was that for the augmented matrix including the spectra for the three processes. The augmentation is done in a column-wise methodology,^[7] which is the most appropriate method to follow the evolution of the reaction. This rank was confirmed by the eigenvectors of the SVD: while those for the first and second singular values evolved similarly in the three processes, that for the third value exhibited no clear-cut, similar trend in any process. The SVD calculations for the augmented matrix in a column-wise methodology exhibited an identical behavior.

Although the number of factors by rank analysis of the data matrix is calculated to be 2, the chemical system studied theoretically consists of 6 components, namely: glycerol, the mixture of caprylic and capric acids—which was taken to be a single component—the mono-, di-, and triglyceride, and water. Water is present in low proportions owing to the high reaction temperature used, the constant agitation and the continuous stream of nitrogen, which caused its distillation as it is produced during esterification. Also, the triglyceride is a minor component owing to the use of excess glycerol. Therefore, the chemical system is assumed to consist of four compounds only. Mono- and

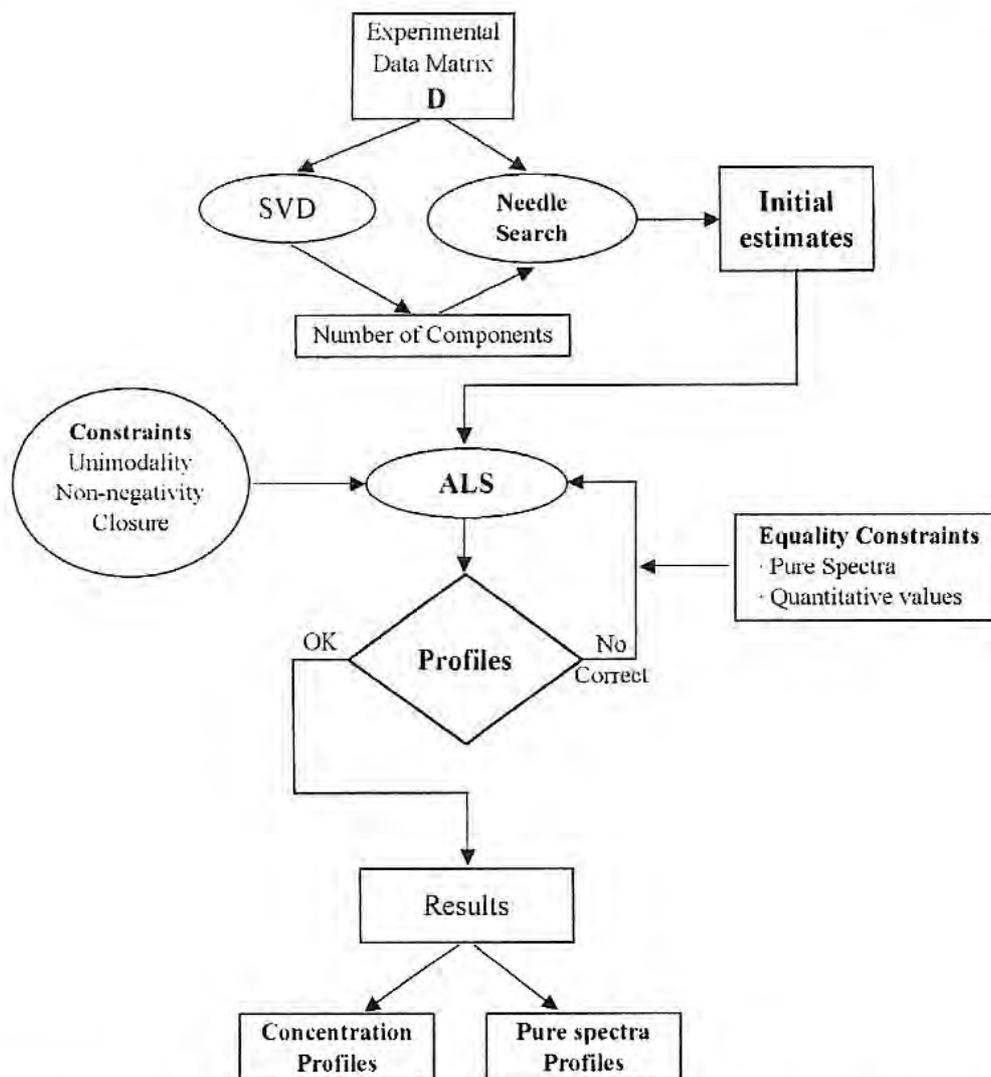


Figure 2. The flowchart shows the different steps of the application of ALS in this study.

diglycerides are chemically highly similar and exhibit very similar NIR spectra as a result. This is probably why the rank of the data matrix is much lower than the number of compounds involved. Although the esterification process involves 6 different compounds, the actual number of spectral components is 3 (viz. glycerol, the acid mixture and the glycerides), but the experimental data are consistent with a rank of only 2.

The augmentation matrix methodology did not improve the rank analysis of the data set. Therefore, we worked with a single matrix **D** corresponding to process A. Application of the MCR-ALS algorithm in combination with non-negativity and unimodality constraints for two components (coinciding with the mathematical rank of the matrix), provided the spectral and concentration profiles which afford no clear-cut conclusions: a comparison with the data matrix reveals that spectrum of component 1 coincides with the last spectrum in matrix **D** and the opposite to component 2. This analysis of

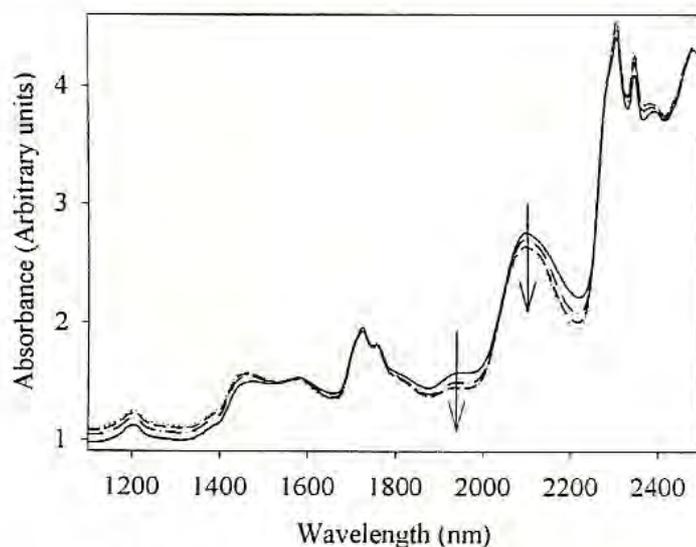


Figure 3. NIR spectra for the reaction products. The arrowheads indicate the temporal evolution of the spectra at 40 (—), 80 (---), 120 (-·-), and 340 min (···).

matrix **D** gives poor results due to the complexity of the system studied. To obtain good results the number of components needs to be near to the number of absorbing species involved in the reaction.

The analysis of matrix **D** under the assumption of 3 components resolves the second component into two; this is reflected in the fact that two of the estimated components exhibited highly similar spectral profiles. This precluded the use of a number of components exceeding the rank of the data matrix and hence resolving rotational ambiguity.

However, if the experimental data matrix is expanded in the iteration process with the spectral information available (viz. the recorded spectra for the reagents) as equality constraints, the resulting profiles exhibit a better-defined, more consistent pattern than the previous calculations. Thus, the use of MCR-ALS methodology with three components, unimodality and concentration and spectral non-negativity constraints, and the recorded NIR spectra for glycerol and the acid mixture as initial estimates in the ALS iterations, provided the profiles shown in Fig. 4. These profiles are more consistent with our knowledge^[16] of the system. Thus, compound 1 clearly coincides with the acid as its curve falls to virtually zero—the reactant is in a substoichiometric proportion. Also, component 2 coincides with glycerol as its curve initially falls and then levels off at about 30%—which is typical of an excess reactant. Finally, the third component coincides with the glycerides, which are used as a whole in the MCR-ALS calculations on account of their spectral similarity. As can be seen, the inclusion of external information (viz. the spectra for the pure reactants) allows a system with an inadequate rank to be resolved and concentration profiles consistent with available knowledge about the system to be obtained. In fact, the incorporation of the spectra for the pure reactants resolves rotational ambiguity between them,

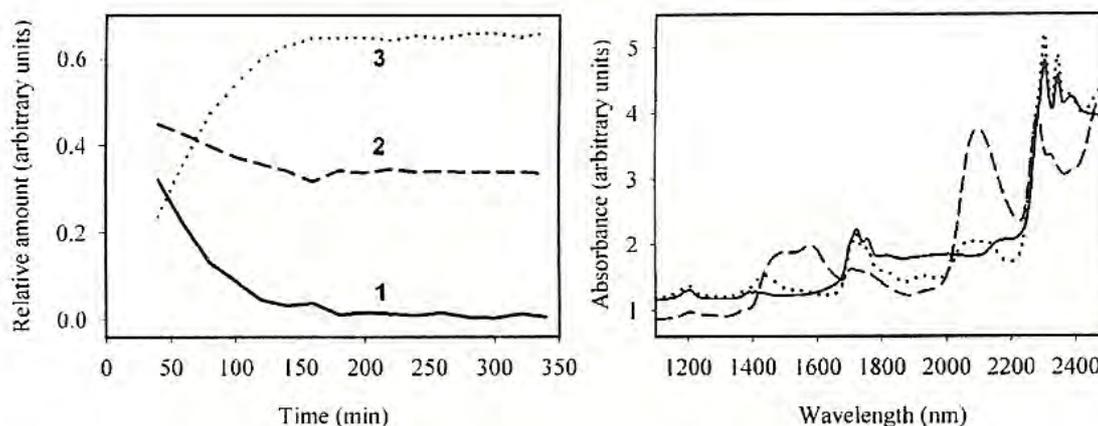


Figure 4. Concentration (left) and spectral profiles (right) obtained by using MCR-ALS with three components and the spectra for the pure components. The solid line corresponds to the acid mixture, the dashed line to glycerol, and the dotted line to glycerides. Variance captured: 99.99%.

which was one of the problems posed by the studied system. This fact reduces the correlations that decreased the rank of the experimental matrix and allowed glycerol to be accurately discriminated from the acid mixture. The captured variance for this calculation is 99.99% and the correlation coefficient between the estimated and original spectra is greater than 0.999 for both glycerol and the acid mixture.

However, the spectral information included as equality constraints in the iterative process was insufficient to solve a system of 4 components; in fact, the concentration profiles were inconsistent with available knowledge (16) about the reaction. Intensity ambiguity was poorly resolved, so, we considered the use of external quantitative values in order to obtain accurately scaled concentration profiles and hence the best approximation to the actual spectral profiles. Thus, the MCR-ALS model was expanded with the concentration values for batch samples that were previously analysed using reference methods (viz. gas chromatography and titration). Five different components were considered, namely glycerol, the acid mixture, and the three types of glyceride (mono-, di-, and tri-). Also, the recorded spectra for glycerol, the acid mixture and the triglyceride, as well as the corresponding concentrations were used as equality constraints. The only parameters not included were the spectra for the mono- and diglyceride, which must be estimated precisely by MCR-ALS. The concentration and spectral profiles thus obtained are shown in Fig. 5.

The estimated spectra profiles for the mono- and diglyceride are very similar; the strongest difference is in the 2000–2200 nm region, where the monoglyceride exhibited a higher absorbance on account of its greater number of O–H groups per molecule. The coefficients of correlation between the estimated and recorded spectra are higher than 0.999 for glycerol, the acid mixture, and the triglyceride. The estimated concentration

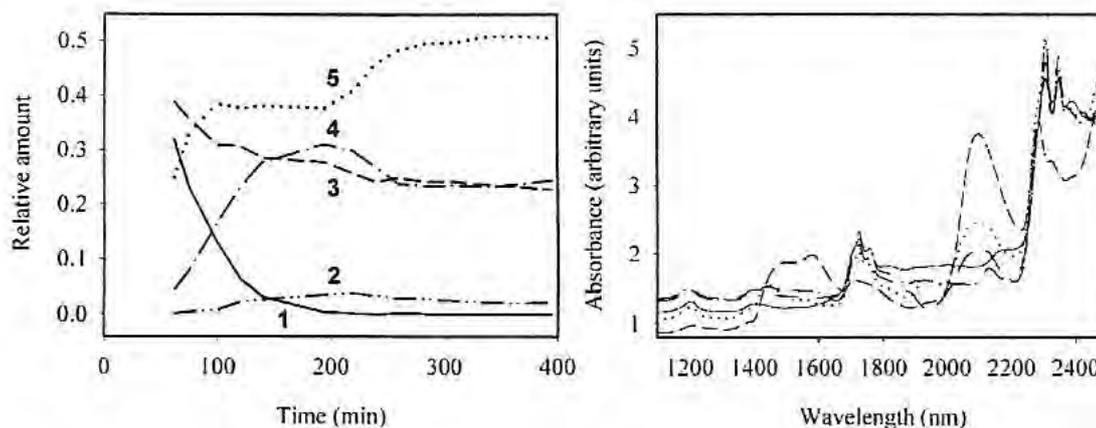


Figure 5. Concentration (left) and spectral profiles (right) obtained by using MCR-ALS with 5 components and the spectra for the pure reactants. (—) Acid mixture, (--) glycerol, (···) monoglyceride, (-·-) diglyceride, and (- - -) triglyceride. Variance captured: 99.98%.

profiles have the same as the profiles from the reference methods. After 200 min, there is equilibrium between glycerides where transesterification reactions may occur. There is easy to see that after 200 min, the concentration profiles of diglyceride, triglyceride, and glycerol fall and the profile of monoglyceride increases. At that period of time, the fatty acid has been totally consumed, so, in principle, reaction could be finished at 200 minutes. Actually, glycerol reacts with diglyceride and triglyceride to form monoglyceride. This is an important step in this industrial process.

The spectra for the mono- and diglyceride in Fig. 5 were used to estimate the corresponding concentration profiles by applying MCR-ALS to new esterification batches (B and C); only the recorded spectra for glycerol and fatty acids, and the estimated spectra for the mono- and diglyceride, in conjunction with the usual non-negativity and unimodality constraints, were used. The profiles thus obtained are shown in Fig. 6.

The concentration profiles are consistent with the expected result for both the reactants and products. Thus, the acid concentration decreases to a residual level after about 150–200 min. The glycerol content levels off at 25% over the same period. The monoglyceride content is high from the start in both processes because the sampling was done 40 min after the beginning of the reaction for the process B and approximately 80 min for C. On the other hand, the diglyceride content increases in a rapid manner similarly as in Fig. 5. The concentration profiles of the two processes are similar because they have been carried out in similar experimental conditions. Not only are the concentration profiles consistent; the concentration scale agrees with the results obtained by titration and gas chromatography.

Therefore, the use of estimated and recorded spectra as initial estimates in the ALS iterative process allows one to quantify the target analytes in the esterification reaction between glycerol and a mixture of caprylic and capric acids.

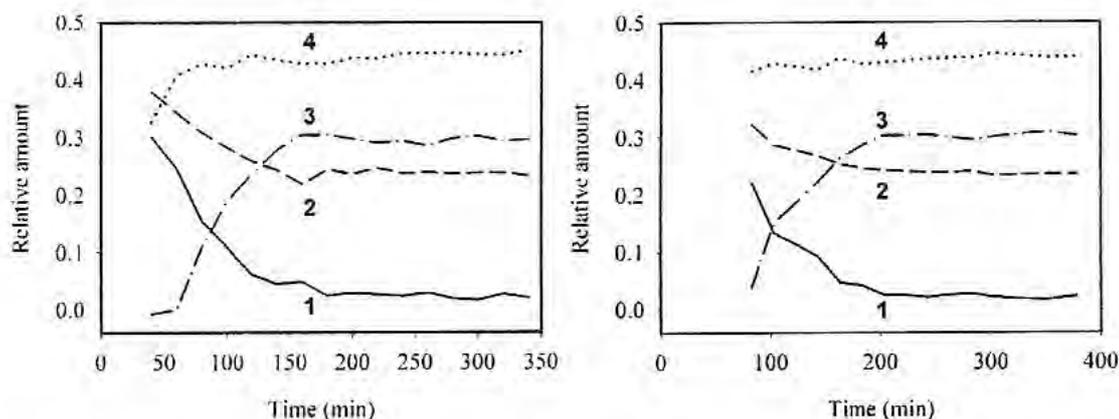


Figure 6. Concentration profiles for batches B (left) and C (right) as obtained by using MCR-ALS methodology with four components and spectral information only. (—) Acid mixture, (--) glycerol, (···) monoglyceride, and (-·-) diglyceride.

CONCLUSIONS

The application of MCR-ALS methodology to a complex chemical system provides information about the course of the reaction; however, the high correlation between the spectra for the products and the inadequate rank of the experimental data matrix considerably restrict the quality of the spectral and quantitative information thus obtained. The incorporation of spectral information (viz. the spectra for the pure reactants) helps resolve rotational ambiguity, but not intensity ambiguity. The inclusion of quantitative information (concentrations) for the components of the chemical system completely resolves both. The spectral profiles thus obtained allow the concentration profiles for components in new esterification batches to be accurately estimated.

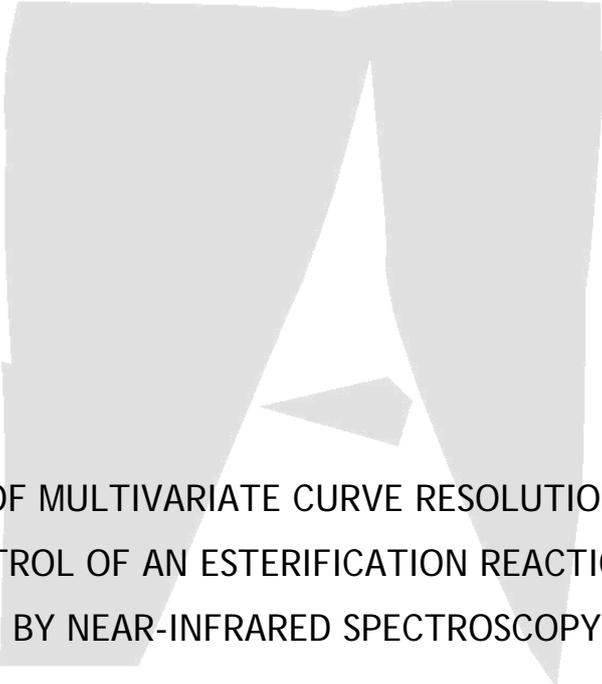
ACKNOWLEDGMENTS

The authors are grateful to Spain's MCyT for funding this research within the framework of Project BQU2003-04247. Miguel Castillo wishes to acknowledge additional support from the DURSI of the Generalitat de Catalunya in the form of a scholarship.

REFERENCES

1. Tauler, R. *Anàlisi de mescles mitjançant resolució multivariant de corbes*, 1st ed.; Institut d'Estudis Catalans: Barcelona, 1997.
2. Larrechi, M. S.; Rius, F. X. Spectra and concentration profiles throughout the reaction of curing epoxy resins from near-infrared spectroscopy and multivariate curve resolution methods. *Appl. Spectrosc.* **2004**, *58* (1), 47.

3. Amrhein, M.; Srinivasan, B.; Bonvin, D.; Schumacher, M. M. On the rank deficiency and rank augmentation of the spectral measurement matrix. *Chemom. Intell. Lab. Syst.* **1996**, *33* (1), 17.
4. Tauler, R.; Kowalski, B.; Fleming, S. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Anal. Chem.* **1993**, *65*, 2040.
5. Tauler, R.; Kowalski, B. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemom.* **1995**, *9* (1), 31.
6. Blanco, M.; Villarroya, I. NIR spectroscopy: a rapid-response analytical tool. *Trends Anal. Chem.* **2002**, *21* (4), 240.
7. De Juan, A; Tauler, R. Chemometrics to unravel multicomponent processes and mixtures Revising latest trends in multivariate resolution. *Anal. Chim. Acta.* **2003**, *500*, 195.
8. Maeder, M.; Zuberbühler, A. D. The resolution of overlapping chromatographic peaks by evolving factor analysis. *Anal. Chim. Acta.* **1986**, *181*, 287.
9. Keller, H. R.; Massart, D. L. Peak purity control in liquid chromatography with photodiode-array detection by a fixed size moving window evolving factor analysis. *Anal. Chim. Acta.* **1991**, *246*, 379.
10. Gemperline, P.; Cash, E. Advantages of soft versus hard constraints in self-modelling curve resolution problems. Alternating least squares with penalty functions. *Anal. Chem.* **2003**, *75*, 4236.
11. MATLAB, Version 6.5. The MathWorks Inc, 2002.
12. Directorate for the Quality of Medicines. Medium-chain triglycerides. In *European Pharmacopoeia*, 4th ed.; Strasbourg, France, 2002.
13. Kállai, M.; Veres, Z.; Balla, J. Response of flame ionization detectors to different homologous series. *Chromatographia* **2001**, *54*, 511.
14. Gemperline, P. J. A priori estimates of elution profiles of pure components in overlapped liquid chromatography peaks using target factor analysis. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 206.
15. Gemperline, P. J. Target transformation factor analysis with linear inequality constraints applied to spectroscopic-chromatographic data. *Anal. Chem.* **1986**, *56*, 2656.
16. Blanco, M.; Beneyto, R.; Castillo, M.; Porcel, M. Analytical control of an esterification batch reaction between glycerine and fatty acids by near-infrared spectroscopy. *Anal. Chim. Acta.* **2004**, *521*, 143.



APPLICATION OF MULTIVARIATE CURVE RESOLUTION TO CHEMICAL
PROCESS CONTROL OF AN ESTERIFICATION REACTION MONITORED
BY NEAR-INFRARED SPECTROSCOPY

M. Blanco, M. Castillo, A. Peinado y R. Beneyto

Applied Spectroscopy **60** (2006) 641-647

Application of Multivariate Curve Resolution to Chemical Process Control of an Esterification Reaction Monitored by Near-Infrared Spectroscopy

MARCELO BLANCO,* MIGUEL CASTILLO, ANTONIO PEINADO, and RAFAEL BENEYTO

Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain (M.B., M.C., A.P.); and Development and Quality Assurance Department, ICI Spain S.A. (Mevisa Site), Ctra Hostalric a Tossa km 1.8, 08495 Fogars de la Selva, Barcelona, Spain (R.B.)

Multivariate curve resolution–alternating least squares (MCR-ALS) methodology was applied to near-infrared spectroscopy (NIR) data for the esterification reaction between glycerol and a mixture of caprylic and capric acids. Batch reaction processes were conducted either at the laboratory scale or at an industrial plant, while NIR data were obtained from samples withdrawn during the reaction processes. The process has been monitored via two typical parameters for this type of reaction, namely, the acid value (AV) and the hydroxyl value (OHV). Spectral and concentration profiles were estimated by applying soft-modeling MCR-ALS to a column-wise augmented data matrix with pure spectra of the components, and concentration values for the acid were used as a soft-equality constraint. The estimated concentration profiles have been compared with the AV and OHV values, and the estimated spectral profiles were used to predict the concentration profiles for new batches. Good results have been obtained in terms of RMSE for the prediction of AV and OHV.

Index Headings: Multivariate curve resolution–alternating least squares; MCR-ALS; Near-infrared spectroscopy; NIR spectroscopy; Esterification reaction.

INTRODUCTION

The esterification reaction between glycerol and middle-chain fatty acids is the basis for the production of esters used by the food, cosmetic, and pharmaceutical industries. The reaction begins with the formation of a monoglyceride that is then further esterified to a diglyceride and, finally, to a triglyceride; the process also involves simultaneous inter-esterification reactions. These side reactions are dominant at the later stages. The reaction end product is a mixture of mono-, di-, and triglycerides in a proportion dependent on that of the reactants. Thus, using excess glycerol leads to mixtures with high mono- and diglyceride contents. On the other hand, using excess acid gives the triglyceride as the major product via intermediate mono- and diglycerides. The usefulness of the resulting end product depends not only on its physical properties (e.g., density, viscosity, solubility), but also on its chemical composition; its characteristics are defined in terms of various parameter values and target specifications.

Wet chemistry provides a number of effective methods for process control based on classical analytical techniques.¹ Thus, the acid value (AV) and the hydroxyl value (OHV) are frequently used to control industrial esterification processes. However, the time-consuming nature of OHV determinations, which typically last up to 90 min, precludes timely, correct decision-making during the process. A truly scientific approach

to the problem requires the use of fast, efficient analytical techniques to monitor the process in order to avoid the production of defective batches or the need for corrective measures, which raise costs and reduce productivity.

Spectroscopic methods constitute an effective alternative to wet chemical methods by virtue of the expeditiousness with which spectra can be recorded and their suitability to a variety of situations, among other advantages. In recent years, near-infrared (NIR) spectroscopy has gained increasing acceptance as a choice for industrial chemical process control. Monitoring has been performed at-,² on-,^{3–5} or even in-line.^{6,7} Partial least squares (PLS)^{2–7} has been the most widespread used; however, other techniques are increasingly gaining acceptance, such as orthogonal projection approach (OPA)^{8,9} or three-way methods.¹⁰

Since Tauler et al. reported the application of MCR-ALS to multiple runs of an industrial process in 1993,¹¹ few applications of multivariate curve resolution, applied to monitor laboratory-scale reactions via spectrophotometric measurements, could be found in the bibliography,^{12,13} but applications are used for descriptive purposes. However, MCR-ALS should also allow the target parameters to be predicted in new samples, with a view to applying it in the routine analysis within an industrial environment. This would enormously foster the application of MCR-ALS to the control of industrial chemical processes. In this work, spectral information derived by MCR-ALS was used to determine the target parameters in new samples. Obtaining the quantitative information required to develop a multivariate calibration model is usually a labor-intensive, time-consuming process that consumes reagents, produces waste, and involves substantial costs. The principal advantage of multivariate curve resolution over other widely used multivariate calibration chemometric techniques, such as PLS, principal component regression (PCR), or multiple linear regression (MLR), is that the former enables calibration from a small number of quantitative data; thus, MCR-ALS allows accurately scaled concentration profiles to be obtained by resolving their intensity ambiguity and using a few reference values as equality constraints during the iterative process.

Although there exists published work in which the AV and OHV are followed by using NIR and PLS,^{7,14} the feasibility of applying MCR-ALS to monitor them has not been checked due to the fact that their values depend on more than one chemical species.

The purpose of this work is to study the concentration evolution of the species involved in the reaction by using soft-modeling multivariate curve resolution with alternating least squares (MCR-ALS) on NIR spectral data, with a view to predicting the species concentration in new samples with

Received 28 November 2005; accepted 20 March 2006.

* Author to whom correspondence should be sent. E-mail: marcel.blanco@uab.es.

spectra previously estimated using MCR-ALS. To this end, we apply this methodology to monitor AV and OHV during the production of a triglyceride from caprylic and capric acids. The results thus obtained are compared with the values provided by reference methods in order to confirm the ability to develop calibration models from a few quantitative data. The estimated spectral profiles were subsequently used to predict AV and OHV in new samples.

THEORY

Multivariate curve resolution methodology assumes data to fit a linear pattern conforming to Beer's law such that¹⁵

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (1)$$

where $\mathbf{D}(I \times J)$ is the experimental data matrix, the rows of which represent spectra in a time sequence; each element d_{ij} in \mathbf{D} corresponds to the absorbance at wavelength j for the sample withdrawn at time i ; $\mathbf{C}(I \times N)$ contains the concentration profiles for the absorbing species involved in the reaction; each element c_{in} in \mathbf{C} representing the concentration of analyte n in the sample at time i ; $\mathbf{S}^T(N \times J)$ contains the spectra for the absorbing species involved in the reaction and each element s_{nj} in \mathbf{S}^T corresponds to the absorbance of analyte n at wavelength j ; and $\mathbf{E}(I \times J)$ is the residuals matrix, which contains all information not explained by matrices \mathbf{C} and \mathbf{S}^T , and is included in \mathbf{D} . When the linear model is obeyed, d_{ij} is the combined absorbance at wavelength j of the N species present in sample i :

$$d_{ij} = \sum_n c_{in}s_{nj} + e_{ij} \quad (2)$$

Equation 1 can be solved iteratively by using the alternating least squares algorithm. Although it has no unique solution, only a few chemically acceptable solutions exist where \mathbf{S}^T corresponds to the spectra for the analytes and \mathbf{C} to their concentration profiles.

Multivariate curve resolution–alternating least squares methodology is subject to two types of ambiguity,¹¹ namely, intensity ambiguity, which means that the estimated profiles may be scaled by some unknown factor, and rotational ambiguity, which means that the estimated spectrum for any of the components may be an unknown linear combination of the real components. Introducing in the iterative process the spectra for absorbing species involved in the reaction helps to suppress the rotational ambiguity¹⁶ that arises from the absence of selective spectral regions, a frequent occurrence in NIR spectroscopy.¹⁷ On the other hand, introducing quantitative values in the iterations provides accurately scaled concentration profiles and suppresses intensity ambiguity.¹⁸ One solution, within the set of feasible solutions, can be obtained by using alternating least squares methodology with a series of constraints including unimodality (when the concentration profiles possess a single maximum value), closure (when the sum of all selected components should equal a constant value), and spectral and concentration non-negativity (when neither the concentrations nor the spectra can take negative values).

The procedure used to solve Eq. 1 therefore started with the determination of the number of components present in \mathbf{D} and the estimation of matrix \mathbf{S}^T or \mathbf{C} . The number of components is estimated by using singular value decomposition (SVD). Other methodologies are available and a detailed description can be

found in the literature.^{11,18–20} In the cases in which the estimated number of components is lower than the number of chemical species in the process, there exists a rank-deficiency problem.^{21,22} There are several methodologies to cope with this problem; the most common is the use of a data matrix containing the spectra for more than one process carried out in different conditions.¹⁹ Another possibility is augmentation of the data matrix with pure spectra of species involved in the reaction.¹²

Once the number of components has been established, an initial estimate of \mathbf{S}^T or \mathbf{C} is used to start the iterations in order to calculate new \mathbf{C} and \mathbf{S}^T estimates in each cycle, the above-described constraints being simultaneously applied in accordance with the specific chemical information available. In this work, iterations were stopped when matrix \mathbf{E} is minimized, i.e., when the difference between the root mean square error (rms) between two consecutive iterations falls below a preset threshold:

$$\text{RMS} = \sqrt{\frac{\sum_i \sum_j e_{ij}^2}{n}} \quad (3)$$

The goodness of fit of the resulting model was assessed in terms of the percent calculated variance:

$$\% \text{ Var} = \frac{\sum_i \sum_j \hat{d}_{ij}^2}{\sum_i \sum_j d_{ij}^2} \times 100 \quad (4)$$

Finally, the quality of the MCR-ALS estimated spectra was evaluated by using the similarity criterion:

$$r = \cos\gamma = \frac{S_i^T \hat{S}_i}{\|S_i\| \cdot \|\hat{S}_i\|} \quad (5)$$

If the physico-chemical conditions of the process are assumed to remain constant, then one can predict the analyte concentrations at any time during the process.¹¹ To this end, one must first calculate the pseudo-inverse of the \mathbf{S}^T matrix obtained by MCR-ALS:

$$\mathbf{P} = (\mathbf{S}^T)^+ \quad (6)$$

where the “+” symbol denotes the pseudo-inverse. The concentrations for any spectrum or matrix can be calculated via matrix \mathbf{P} :

$$c_{\text{pred}} = s \cdot \mathbf{P} \quad (7)$$

where s denotes the spectrum for a new sample or a spectral matrix.

Predictions can be performed with several methodologies, such as augmenting the calibration matrix with the samples to be predicted²²; however, in this work, the selected procedure has been the employment of the pseudo-inverse matrix due to its simplicity and ability to be easily applied to routine analysis.

EXPERIMENTAL

Apparatus and Software. The laboratory processes were conducted in a 1 L LabMax reactor from Mettler Toledo; reaction variables were controlled via the software Camile Tg 3.06, from Sagian, Inc., which allows the heating fluid,

TABLE I. Summary of reference methods.

Parameter	Sample	Method	Solvent	Reagent(s)
AV	0.1–20 g	Acid–base titration	25 mL ethanol/xylene	0.1–0.5 M KOH ethanolic solution
OHV	1–5 g	(1) Acetylation (2) Acid–base titration	Pyridine 25 mL ethanol	Acetic anhydride 0.5 M KOH ethanolic solution

distillation, and reaction temperatures, in addition to various other parameters such as the stirring rate and step time, to be recorded and set throughout the reaction.

Near-infrared transmittance spectra were recorded on a FOSS NIRSystems 6500 spectrophotometer equipped with a rapid content analyzer module. The instrument was controlled via the Vision v.2.51 software package.

Multivariate curve resolution was used in combination with alternating least squares methodology with penalty function, using the software GUIPRO^{23,24} as implemented in Matlab version 7.²⁵

Esterification Reaction of Glycerol. The chemical process is the esterification of glycerol with a 60:40 mixture of caprylic and capric acids in a 10% over-stoichiometric proportion relative to that theoretically needed to obtain the triglyceride. The process was conducted at 220 ± 5 °C under a continuous stream of nitrogen. Once the reaction is finished, the unreacted acid is removed by vacuum distillation. The industrial process is monitored via AV and OHV. The process is finished when the hydroxyl value, which is due to glycerol, mono-, or diglyceride concentrations, falls to a low enough level (close to 5 mg KOH/g sample), after which the remaining acid is distilled.

The samples studied belong to six manufacturing batches from Uniqema (Mevisa Site, Spain) and two batches from the laboratory-scale reactor. Samplings were done at variable reaction times and spanned the ranges 0.4–162 mg KOH/g sample for AV and 1–117 mg KOH/g sample for OHV. No samples were withdrawn during the first hour of reaction as the reactor mixture was not yet homogeneous. All samples were subjected to the reference methods, which are summarized in Table I.

Near-Infrared Spectra. Near-infrared spectra for the reaction mixtures and the pure products where available were recorded in the transmittance mode at room temperature, using a quartz vessel and a gold reflector with a path length of 0.5 mm. Each individual spectrum is the average of 32 scans done at 2 nm intervals over the wavelength range 1100–2500 nm.

Data Processing. Spectral data were subjected to the standard normal variate (SNV) pretreatment in order to suppress potential baseline shifts or scattering caused by some turbid samples. In order to discard negative spectral values and facilitate application of the spectral non-negativity constraint in the MCR-ALS calculations, the regions of negative intensity provided by the SNV pretreatment were removed by adding the same scalar to the whole spectrum and to all the spectra used. In this situation, the name of the y -axis in Fig. 1 was changed to *intensity* instead of *absorbance*. We have studied that this pretreatment does not affect to the application of MCR-ALS because it only produces a displacement of all spectra and there are no changes in its profiles.

Experimental matrices were subjected to rank analysis by using singular value decomposition (SVD) methodology; both singular values and the profile of the corresponding eigenvectors were considered in determining ranks. The above-

mentioned pretreatment does not have an effect on the rank matrix, so, there are no changes in the rank analysis.

The goodness of the AV and OHV results was assessed in terms of the root mean square error:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{n}} \quad (8)$$

where n is the number of samples. This error is designated as RMSEC for calibration and RMSEP for prediction.

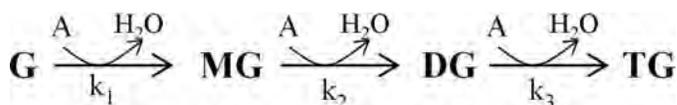
Application of Multivariate Curve Resolution–Alternating Least Squares. For determining the number of components, MCR-ALS methodology was applied to a column-wise data matrix augmented with the pure spectra of the components.

The matrix used to build the calibration model was made up of either the samples from one laboratory or two industrial processes, augmented with the spectra of the pure components. The use of a joined matrix was necessary due to the small number of samples available for each process. The matrix used in the prediction step is composed of one laboratory and four industrial processes.

The initial estimates used prior to the ALS iterative process were obtained by using needle search methodology,^{26,27} as implemented in the software GUIPRO. All calculations were done under the non-negativity constraint in both spectral and concentration data. Consistency in the ALS iterative process in the calibration step was improved by using additional external information (concentration values) as soft-equality constraints.²³

RESULTS AND DISCUSSION

The reaction involves glycerol and a mixture of caprylic and capric acids, which are dealt with as a single component on account of their chemical and spectral similarity, as reactants, and a monoglyceride, diglyceride, and triglyceride as products. Although water is also a reaction product, the reaction temperature used (220 °C) resulted in its distillation, so its content in the reaction mixture at any time can be assumed to have been very low. Under these conditions, the esterification reactions can be considered as irreversible²⁸ and can be written as the following reaction scheme:



where G corresponds to glycerol, MG to monoglyceride, DG to diglyceride, TG to triglycerides, and A to a mixture of capric/caprylic acids, and k_i are the rate constants.

Initially, the acid content of the mixture is higher than that of glycerol; as the reaction progresses, the mono- and diglyceride

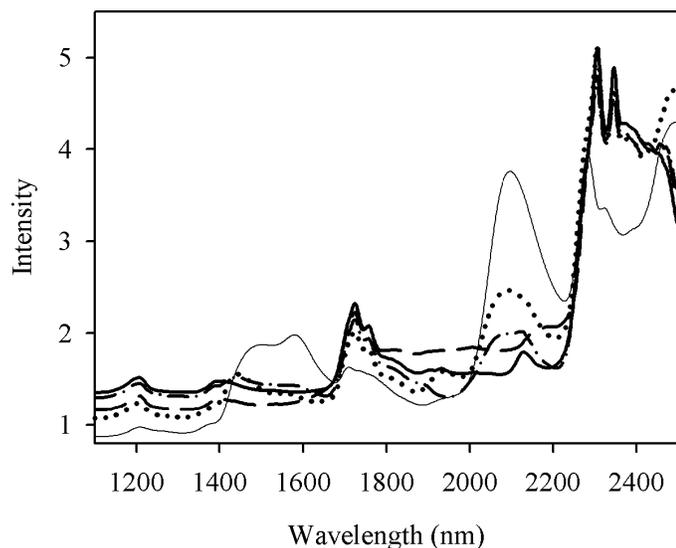


FIG. 1. SNV pretreated NIR spectra for the pure compounds. Acid (— —), triglyceride (—), monoglyceride (•••), diglyceride (— • —), and glycerol (— —).

formed react with part of the remaining acid to form the triglyceride, which is the reaction end product. Therefore, the mono- and diglyceride can be considered two reaction intermediates. As commented above, there exists a lack of homogeneity in the first hour of the reaction, which prevents proper NIR measurements. For this reason, no samples were withdrawn in this first stage of the process. Therefore, the formation of mono- and diglyceride might have gone undetected.

As can be seen in Fig. 1, the main reaction components (*viz.* glycerol, the acid mixture, and the glycerides) exhibit spectral differences. Thus, glycerol gives a strong absorption signal between 2000 and 2200 nm due to the combination bands of the O–H bond, and the acid mixture spectrum exhibits nearly constant absorption in the region 1800–2200 nm, which distinguishes it from the other reaction components. The greatest spectral difference for the glycerol esters is in the region 2000–2200 nm, where could be seen the decreased absorption of the diglyceride relative to the monoglyceride, and also the weak absorption of the triglyceride in these regions. The esters and the acid mixture additionally exhibit two absorption maxima in the region 1700–1800 nm corresponding to the first overtone of C–H bonds, which are weak for glycerol.

The spectra of glycerol, acid mixture, and triglyceride were obtained from the pure products. On the other hand, no pure mono- or diglyceride samples were available, so their spectra were taken from previous work²⁹ where they were estimated by applying MCR-ALS to NIR data obtained in a similar esterification process conducted in the presence of an excess of glycerol, which therefore led to the mono- and diglyceride as the major end products.

The evolution of the NIR spectrum (Fig. 2) in the regions 1400–1700 and 1800–2200 nm is consistent with the disappearance of O–H bonds in glycerol and the fatty acids.

Determination of the Number of Components. The SVD mathematical rank for a single-process data matrix is 2 (the first three SVD eigenvalues are 3.8×10^4 , 10.2, and 0.05 respectively). This value is quite different from the theoretical

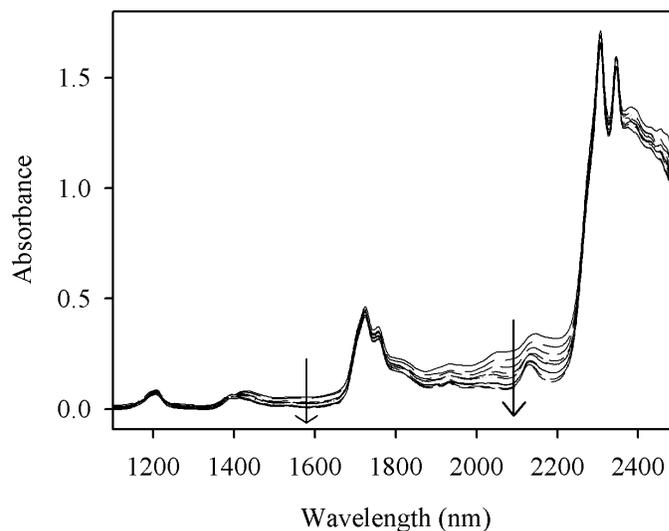


FIG. 2. NIR spectra recorded during the reaction. Arrowheads indicate temporal changes in the spectra.

number of chemical components involved in the reaction, five (*viz.* acid, glycerol, monoglyceride, diglyceride, and triglyceride), or six if water is considered. The difference between mathematical and theoretical values can be ascribed to high spectral and concentration correlations between the components, and also to the low concentrations of mono and diglycerides. The first step to build the calibration model is to determine the number of components that can be unraveled by MCR-ALS. To achieve this purpose and, due to the lack of spectral selectivity of the studied system, the use of an augmented matrix with the pure spectra of the components is required.

A matrix composed of the spectra of the samples extracted from a process was augmented with the spectra of six pure components. Obviously, the mathematical rank obtained was six (the first seven SVD eigenvalues are 5.8×10^4 , 751, 261, 21, 15, 1.9, and 0.3, respectively). After applying MCR-ALS under non-negativity constraints, the unravelled concentration profiles (Fig. 3) show the low significance profile ascribed to water; as can be seen, it is almost a horizontal line close to zero. Similarly, the profile ascribed to diglyceride content follows an erratic evolution around zero. A kinetic study of a similar chemical system, the esterification reaction between glycerol and fatty acids,²⁸ shows that the rate constant of the esterification of glycerol to monoglyceride (k_1) is higher than the rate constant of the esterification of monoglyceride to diglyceride (k_2), and the esterification of diglyceride to triglyceride (k_3) has an unknown rate constant but it could be assumed that k_3 will no longer differ from k_2 . With this hypothesis, monoglyceride becomes the most remainder intermediate product. Our estimated concentration profiles are in accordance with the calculated rate constants of this kinetic study. In the first moments of reaction, glycerol reacts fast (k_1); when the concentration of glycerol is low, after approximately 150 minutes, the content of monoglyceride is still high ($k_2 < k_1$), and the diglyceride has an erratic behavior around zero, which allows confirming that diglyceride is present in the samples in a very low concentration. These facts induce us to think that water and diglyceride could distort the other profiles; this is the reason why they were excluded from subsequent analysis.

The next step was to apply MCR-ALS to the same process

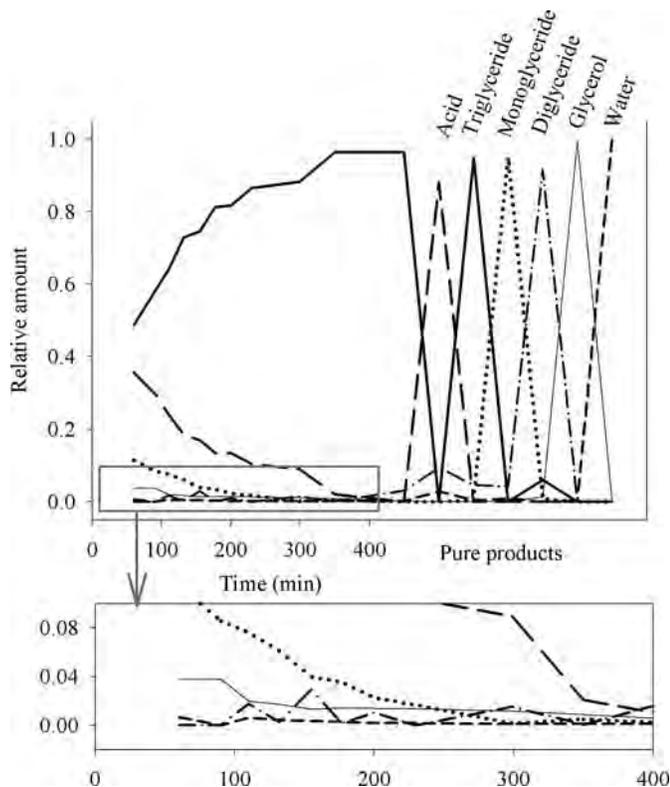


FIG. 3. MCR-ALS concentration profiles for a process matrix augmented with six components. Acid (—), triglyceride (—), monoglyceride (•••), diglyceride (—•—), glycerol (—), and water (■ ■).

matrix, but augmenting it only with the pure spectra corresponding to the four significant species (*viz.* acid, glycerol, triglyceride, and monoglyceride) (the first five SVD eigenvalues are 9.7×10^4 , 416, 28, 16, and 0.6, respectively). The results obtained are shown in Fig. 4. As can be seen, the concentration profiles reflect the virtually complete disappearance of the acid, the monoglyceride, and the glycerol, and the progressive augmentation in triglyceride content. These evolutions are consistent with acquired knowledge about the system.¹⁴

Comparing the profiles (Figs. 3 and 4) before and after the withdrawal of water and diglyceride, we can infer how the concentration profiles are softer and follow a more defined tendency in Fig. 4. To verify whether the results obtained are consistent, the matrices corresponding to the other processes were augmented in a similar way. The results obtained were similar to those mentioned above.

Model Construction. Once the number of components has been fixed, a calibration model capable of accurately predicting

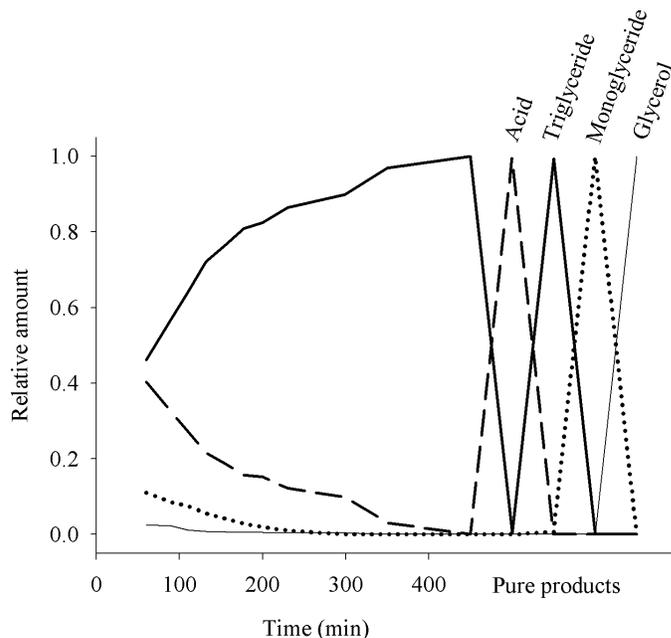


FIG. 4. MCR-ALS concentration profiles for a process matrix augmented with four components. Acid (—), triglyceride (—), monoglyceride (•••), and glycerol (—).

the evolution of the species involved was built. A calibration matrix composed of 24 sample spectra from three processes, augmented with the four pure component spectra, has been used to construct the model. The prediction set was composed of sample spectra withdrawn from five new processes (one laboratory process and four industrial processes).

The concentration profiles provided by MCR-ALS were converted into AV and OHV values, both expressed as mg KOH/g sample, because these parameters are the most common way to express the results in an industrial environment and in order to compare the MCR-ALS results with the reference values. AV is a measure of the carboxylic groups that have not reacted, and OHV is the sum of the remainder hydroxyl groups from glycerol and monoglyceride. AV and OHV values were obtained by taking into account the average molecular weights of the acids for the former, and the stoichiometry of the titration reaction with the molecular weights of glycerol and monoglyceride for the latter. The goodness of the AV and OHV results was assessed from a univariate linear regression of the MCR-ALS estimated values against values obtained from titration data.

The first step was the building of a MCR-ALS model as simple as possible in terms of external information, *i.e.*, without equality constraints. The figures of merit obtained for calibration and prediction are shown in Table II. Although the

TABLE II. Figures of merit for the regression between the MCR-ALS (without equality constraints) values against the reference values for AV and OHV.

Set Parameter	Calibration		Prediction	
	AV	OHV	AV	OHV
Samples	24		33	27
Concentration interval (mgKOH/g sample)	0.4–162	1–117	1.5–153	3–116
Slope	1.15 ± 0.05	0.87 ± 0.06	1.16 ± 0.03	0.83 ± 0.03
Offset	-3.4 ± 4.3	0.4 ± 3.2	-0.9 ± 2.3	-2.3 ± 1.4
R	0.995	0.989	0.997	0.997
RMSEC/P	11.4	8.9	9.6	9.4

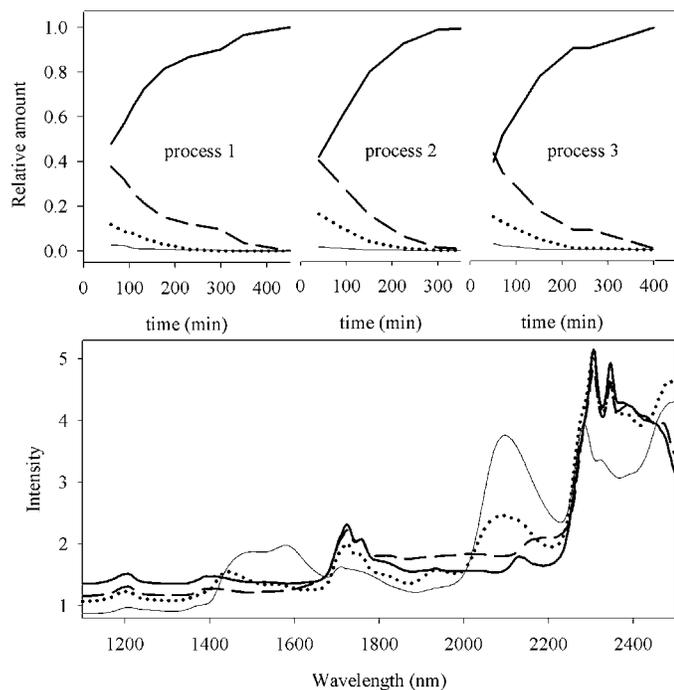


FIG. 5. (Top) Concentration profiles obtained by applying MCR-ALS to three reaction processes. (Bottom) MCR-ALS estimated spectra for pure components. Acid (—), triglyceride (—), monoglyceride (•••), and glycerol (—).

correlations for both parameters are very similar, the slopes and offsets are far from 1 for the slope and 0 for the offset. Equally, RMSEC and RMSEP values are high enough that the model cannot be successfully applied.

According to the described results, and in order to improve the calibration model, the inclusion of soft-equality constraints was tested. The individual concentrations of glycerol and monoglyceride were not available and, consequently, equality constraints could not be applied for these two analytes. The constraints were only applied to the acid concentration and, even though we had the acid concentration for all samples, we used two different combinations of providing equality constraints for this parameter: (1) In one of them (model 1) the acid concentration values for all samples were used. (2) In

the other, (model 2) only the initial and final acid concentrations for each process were used.

The results obtained for the former, in terms of concentration and spectra profiles, are shown in Fig. 5. In both models, the explained variances (Eq. 4) are 99.99% and the correlations between the pure and estimated spectra (Eq. 5) are higher than 0.999 for the four components involved in the process. The figures of merit between the estimated values provided by MCR-ALS and those provided by the reference methods are shown in Table III, for calibration and prediction.

In calibration, both models have slopes near-unity and offsets close to zero; both values fell within the confidence range. Model 1 performs slightly better judging from its closer fitting and lower RMSEC. Considering the prediction results, both models present similar figures of merit. The prediction values provided by both models are not significantly different.

These results testify to the capacity of MCR-ALS methodology to produce accurate calibration models from only a few quantitative values, and their RMSEP values show accurate and acceptable results that confirm that the proposed method is suitable to industrial process control.

CONCLUSION

Multivariate curve resolution–alternating least squares allows the unraveling and resolving of the concentration profiles for the species involved in a complex chemical system from NIR spectral data. Despite the rank-deficiency, matrix augmentation with the spectra of pure species has turned out to be a useful strategy to solve the problem.

The MCR-ALS technique allows calibration models to be constructed from a few quantitative values, which is an important advantage over other calibration techniques with a view to reducing the need for reference values and the time consumed in the model construction.

The good results obtained in predicting the target parameters in new samples from NIR spectral data allow the proposed methodology to effectively replace existing wet chemical methods for the same purpose, which are labor intensive and time consuming. This study can be the basis for future work on the development of stable and robust MCR-ALS calibration models and therefore constitutes a simple method for monitoring new chemical processes.

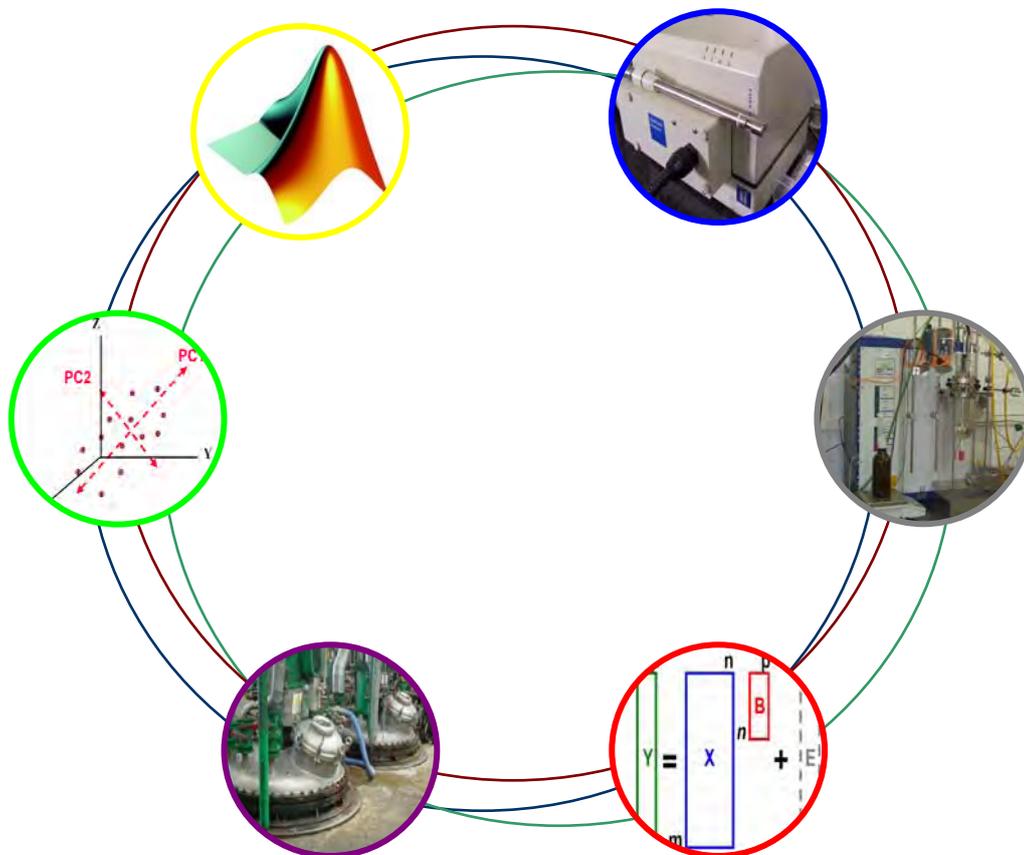
TABLE III. Calibration and prediction sample figures of merit for the regression between the MCR-ALS values against the reference values for AV and OHV.

Model Parameter	Model 1		Model 2	
	AV	OHV	AV	OHV
Calibration Samples	24			
Concentration interval (mgKOH/g sample)	0.4–162	1–117	0.4–162	1–117
Slope	$1.00 \pm 9 \times 10^{-6}$	0.97 ± 0.04	1.03 ± 0.03	0.95 ± 0.05
Offset	$-7.1 \times 10^{-4} \pm 7.2 \times 10^{-4}$	0.3 ± 2.1	-2.2 ± 2.5	1.3 ± 2.8
<i>r</i>	1	0.996	0.998	0.994
RMSEC	0.001	3.9	3.7	5
Prediction Samples	33	27	33	27
Concentration interval (mgKOH/g sample)	1.5–153	3–116	1.5–153	3–116
Slope	1.09 ± 0.03	0.87 ± 0.03	1.11 ± 0.03	0.86 ± 0.03
Offset	-2.4 ± 2.2	-0.9 ± 1.5	-3.8 ± 2.2	-0.1 ± 1.5
<i>r</i>	0.997	0.997	0.997	0.997
RMSEP	5.9	6.8	6.1	6.8

ACKNOWLEDGMENTS

The authors are grateful to Spain's MCyT for funding this research within the framework of Project BQU2003-04247. Miguel Castillo also wishes to acknowledge support from the DURSI of the Generalitat de Catalunya and European Social Fund in the form of a scholarship.

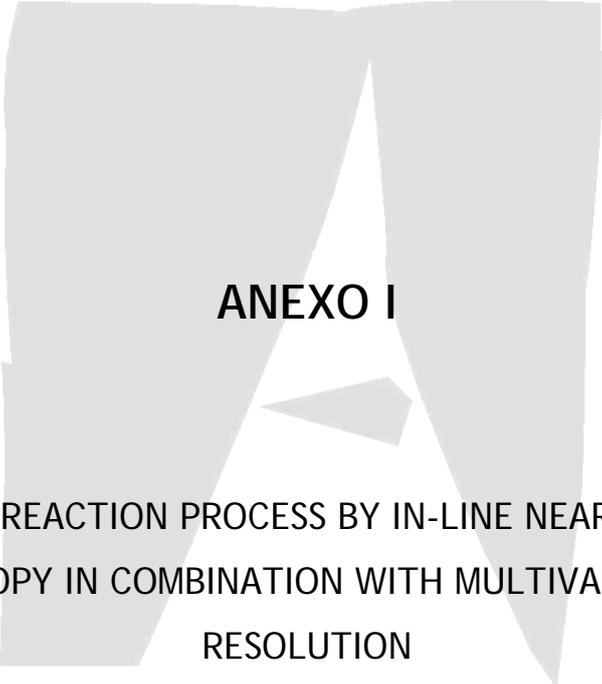
1. *Medium-Chain Triglycerides in European Pharmacopoeia* (Directorate for the Quality of Medicines, Strasbourg, 2002), 4th ed.
2. M. Blanco, R. Beneyto, M. Castillo, and M. Porcel, *Anal. Chim. Acta* **521**, 143 (2004).
3. M. Blanco and D. Serrano, *Analyst* (Cambridge, U.K.) **125**, 2059 (2000).
4. H. H. A. Lousberg, H. F. M. Boelens, E. P. Le Comte, H. C. J. Hoefsloot, and A. K. Smilde, *J. Appl. Polym. Sci.* **84**, 90 (2002).
5. A. Cherfi, G. Fevotte, and C. Novat, *J. Appl. Polym. Sci.* **85**, 2510 (2002).
6. S. E. Barnes, E. C. Brown, M. G. Sibley, H. G. M. Edwards, I. J. Scowen, and P. D. Coates, *Appl. Spectrosc.* **59**, 611 (2005).
7. R. A. Heikka, K. T. Immonen, P. O. Minkinen, E. Y. O. Paatero, and T. O. Salmi, *Anal. Chim. Acta* **349**, 287 (1997).
8. S. Gourvénec, C. Lamotte, P. Pestiaux, and D. L. Massart, *Appl. Spectrosc.* **57**, 1 (2003).
9. S. Gourvénec and D. L. Massart, *Anal. Bioanal. Chem.* **380**, 373 (2004).
10. P. Geladi and J. Forsström, *J. Chemom.* **16**, 329 (2002).
11. R. Tauler, B. Kowalski, and S. Flemming, *Anal. Chem.* **65**, 2040 (1993).
12. M. S. Larrechi and F. X. Rius, *Appl. Spectrosc.* **58**, 47 (2004).
13. S. Sasic, Y. Kita, T. Furukawa, M. Watari, H. W. Siesler, and Y. Ozaki, *Analyst* (Cambridge, U.K.) **125**, 2315 (2000).
14. M. Blanco, R. Beneyto, M. Porcel, and M. Castillo, Proceedings of the 11th International Conference on Near Infrared Spectroscopy, 1021 (2003).
15. R. Tauler, *Anàlisi de mesclres mitjançant resolució multivariant de corbes* (Institut d'Estudis Catalans, Barcelona, 1997), 1st ed.
16. R. Tauler, A. Smilde, and B. Kowalski, *J. Chemom.* **9**, 31 (1995).
17. M. Blanco and I. Villarroya, *Trac* **21**, 240 (2002).
18. A. De Juan and R. Tauler, *Anal. Chim. Acta* **500**, 195 (2003).
19. M. Maeder and A. D. Zuberbühler, *Anal. Chim. Acta* **181**, 287 (1986).
20. H. R. Keller and D. L. Massart, *Anal. Chim. Acta* **246**, 379 (1991).
21. M. Amrhein, B. Srinivasan, D. Bonvin, and M. M. Schumacher, *Chemom. Intell. Lab. Syst.* **33**, 17 (1996).
22. A. Izquierdo-Ridorsa, J. Saurina, S. Hernández-Cassou, and R. Tauler, *Chemom. Intell. Lab. Syst.* **38**, 183 (1997).
23. P. Gemperline and E. Cash, *Anal. Chem.* **75**, 4236 (2003).
24. http://personal.ecu.edu/gemperline/Research/gemper_software.html, Paul J. Gemperline, East Carolina University.
25. *MATLAB, version 7.0* (The MathWorks, Natick, MA, 2004).
26. P. J. Gemperline, *J. Chem. Inf. Comput. Sci.* **24**, 206 (1984).
27. P. J. Gemperline, *Anal. Chem.* **56**, 2656 (1986).
28. H. Szelag and W. Zwierzykowski, *Lipid* **100**, 302 (1998).
29. M. Blanco, M. Castillo, R. Beneyto, and M. Porcel, *Spectrosc. Lett.* **38**, 825 (2005).



ANEXOS

En el anexo se recogen los artículos publicados en fecha posterior a la aceptación por parte de la *Comissió de Doctorat* de la presentación de la tesis como compendio de publicaciones.

- I) Study of reaction processes by in-line near-infrared spectroscopy in combination with multivariate curve resolution. Esterification of myristic acid with isopropanol.
- II) Determination of low analyte concentrations by near-infrared spectroscopy: Effect of spectral pretreatments and estimation of multivariate detection limits.



ANEXO I

STUDY OF REACTION PROCESS BY IN-LINE NEAR-INFRARED
SPECTROSCOPY IN COMBINATION WITH MULTIVARIATE CURVE
RESOLUTION

ESTERIFICATION OF MYRISTIC ACID WITH ISOPROPANOL

M. Blanco, M. Castillo y R. Beneyto

Talanta 72 (2007) 519-525

Study of reaction processes by in-line near-infrared spectroscopy in combination with multivariate curve resolution Esterification of myristic acid with isopropanol

Marcelo Blanco^{a,*}, Miguel Castillo^a, Rafael Beneyto^b

^a Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

^b Development and Quality Assurance Department, ICI Spain S.A. (Mevisa Site), Ctra Hostalric and Tossa km 1.8, 08495 Fogars de la Selva, Barcelona, Spain

Received 27 July 2006; received in revised form 19 October 2006; accepted 8 November 2006

Available online 19 December 2006

Abstract

The acid-catalysed esterification of myristic acid with isopropanol was studied by using near-infrared spectroscopy (NIR) in combination with soft-modeling curve resolution (MCR) methodology with a view to establishing the effect of experimental variables on the reaction kinetics. The reaction was conducted at temperatures above the boiling point of the alcohol, with continuous addition of an isopropanol/water mixture to the reactor. Spectral and concentration profiles were determined by applying soft-modeling curve resolution methodology to a column-wise augmented data matrix containing the spectra for the pure components. MCR profiles were compared with reference values and found to depart from then by less than 3% as %RSE for concentrations and to exhibit correlation above 0.999 for spectra.

The reaction kinetics as estimated from the concentration profiles was found to be *pseudo*-first-order. Also, the *pseudo*-first-order rate constant was found to depend on the flow-rate of the isopropanol/water mixture and its water content; although the constant decreased with increase in the proportion of water, a content of *ca.* 15% could be used without important retarding effects on the kinetics. The proposed NIR-MCR method allows the rate constant and the influence of the initial water content to be determined with a view to minimizing consumption of the raw materials and optimizing the experimental conditions.

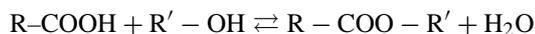
© 2007 Elsevier B.V. All rights reserved.

Keywords: Near-infrared spectroscopy; Multivariate curve resolution; In-line measurements; Monitoring; Optimization

1. Introduction

The esters of fatty acids and lower alcohols are usually obtained by direct esterification of the acid or transesterification of natural oils with an alcohol, at a relatively high temperature. The low boiling point of the alcohol requires using temperatures below such a point, high pressures in order to increase it or adding the alcohol continuously during the reaction if substantial losses through evaporation are to be avoided. Continuously adding the alcohol requires using a considerably overstoichiometric amount of this reactant and produces a large volume of alcohol/water distillate that must be recovered for subsequent use. However, the reuse of distillates containing much

water decreases the final yield to a great extent by displacing equilibrium and altering the rate of the main reaction:



It is therefore essential to determine the effect of the water content on the reaction rate and yield. By monitoring the reaction, one can identify the variation pattern of the composition of the reaction mixture in order to optimize the reaction time, alcohol consumption and distillate reuse, all with a view to reducing consumption of the alcohol and production of waste.

In this work, we studied the esterification at a high temperature (130 °C) of myristic acid with isopropanol – the latter of which was slowly added – and *p*-toluenesulphonic acid as catalyst to obtain isopropyl myristate and water. The water was distilled together with residual isopropanol, the two forming an azeotrope of boiling point 80.4 °C with a water content of 12.3%.

* Corresponding author.

E-mail address: marcel.blanco@uab.es (M. Blanco).

Esterification reactions under homogeneous catalysis are usually second-order in the acid (Ac) and alcohol (Al):

$$v = -\frac{d[\text{Ac}]}{dt} = k \cdot [\text{Ac}][\text{Al}] \quad (1)$$

In the presence of a large excess or a constant understoichiometric amount of alcohol, however, the reaction kinetics can be approximated to a *pseudo*-first-order law for the acid [1,2]:

$$v = -\frac{d[\text{Ac}]}{dt} = k' \cdot [\text{Ac}] \quad (2)$$

where

$$k' = k \cdot [\text{Al}] \quad (3)$$

Eq. (2) can be expressed in integral form as

$$\text{Ln} [\text{Ac}] = \text{Ln} [\text{Ac}]_0 - k' \cdot t \quad (4)$$

where k' is the *pseudo*-first-order rate constant, which depends of various experimental variables including temperature, and the water content and rate of addition of the alcohol, and can be readily determined from a plot of $\text{Ln} [\text{Ac}]$ versus time.

Although this type of esterification reaction under acid catalysis [1–3] has been the subject of several studies, monitoring the process entails withdrawing a large number of samples from the reaction mixture and performing an also large number of analyses. Near infrared spectroscopy (NIR) [4] provides a practical tool for monitoring chemical processes [5–9]; in combination with multivariate calibration techniques such as multivariate curve resolution-alternating least squares methodology (MCR-ALS) [10], it allows the concentration profiles for the reactants as a function of time to be obtained from spectral information for the process and little additional external information [11].

2. A brief introduction to multivariate curve resolution (MCR)

In multivariate curve resolution-alternating least squares (MCR-ALS), the spectral data matrix, \mathbf{D} ($I \times J$), is resolved into the product of a concentration matrix, \mathbf{C} ($I \times N$), and a spectrum matrix, \mathbf{S} ($N \times J$), according to:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (5)$$

where \mathbf{E} ($I \times J$) is the residual matrix which contains all information not explained by matrices \mathbf{C} and \mathbf{S}^T ; being I the number of samples, J the number of wavelengths and N the number of compounds in the samples. Eq. (5) can be resolved iteratively by using alternating least squares methodology. The process starts with the initial estimation of matrix \mathbf{C} or \mathbf{S} , each iteration producing a refined estimate, and is stopped when the difference between the root mean square error,

$$\text{RMS} = \sqrt{\frac{\sum_i \sum_j e_{ij}^2}{n}} \quad (6)$$

between two consecutive iterations falls below a preset threshold, being e_{ij} the residual between two consecutive iterations for sample i at wavelength j .

The goodness of fit of the resulting model was assessed in terms of the percent calculated variance between the estimated \mathbf{D} matrix (calculated as $\mathbf{C}\mathbf{S}^T$) and the original \mathbf{D} matrix:

$$\% \text{Var} = \frac{\sum_i \sum_j \hat{d}_{ij}^2}{\sum_i \sum_j d_{ij}^2} \times 100 \quad (7)$$

Because of rotational and intensity ambiguity [12], this multivariate methodology cannot provide a unique solution. In order to reduce the number of potential solutions and solve Eq. (5) accurately, computations are done under non-negativity, unimodality or closure constraints [10]. Augmenting the data matrix with the spectra for absorbing species involved in the reaction [11,13] and/or introducing quantitative values (concentration values) in the iterative process [10] as equality constraints, also helps suppress ambiguity.

By assuming the physico-chemical conditions between batches (such as temperature, or initial molar ratio) to remain constant, one can predict the analyte concentrations at any time during it [12] for new batches. To this end, one must first calculate the *pseudo*-inverse of the \mathbf{S}^T matrix obtained by MCR-ALS:

$$\mathbf{P} = (\mathbf{S}^T)^+ \quad (8)$$

where the plus sign denotes *pseudo*-inverse. The concentrations corresponding to any new spectrum or matrix can be calculated via matrix \mathbf{P} :

$$c_{\text{pred}} = s \cdot \mathbf{P} \quad (9)$$

where s denotes the spectrum for a new sample.

3. Experimental

3.1. Apparatus and software

Esterification processes were conducted in a LabMax Mettler Toledo laboratory reactor that was governed via the software Camile Tg 3.06 from Sagian Inc. The instrument allows the reactor, jacket and distillation temperatures, and the stirring and reactant addition rates, to be controlled (Fig. 1).

NIR in-line transmittance spectra were recorded on a FOSS NIRSystems 5000 spectrophotometer equipped with an OptiProbeTM analyser module allowing in-line spectra in the region 1100–2498 nm to be recorded by using a fiber-optic probe immersed in the reaction mixture. NIR at-line transmittance spectra were recorded on a FOSS NIRSystems 6500 spectrophotometer equipped with an RCA module in the same region as the in-line spectra. Both spectrophotometers were governed via the software Vision v. 2.51.

Multivariate curve resolution-alternating least squares computations were done by using the software GUIPRO [14,15] as implemented in Matlab v. 7 [16].

3.2. Catalysed esterification of myristic acid with isopropanol

The raw materials for the reaction were myristic acid (Uniqema, ICI Spain, 98–100% pure, AV 244–248 mg

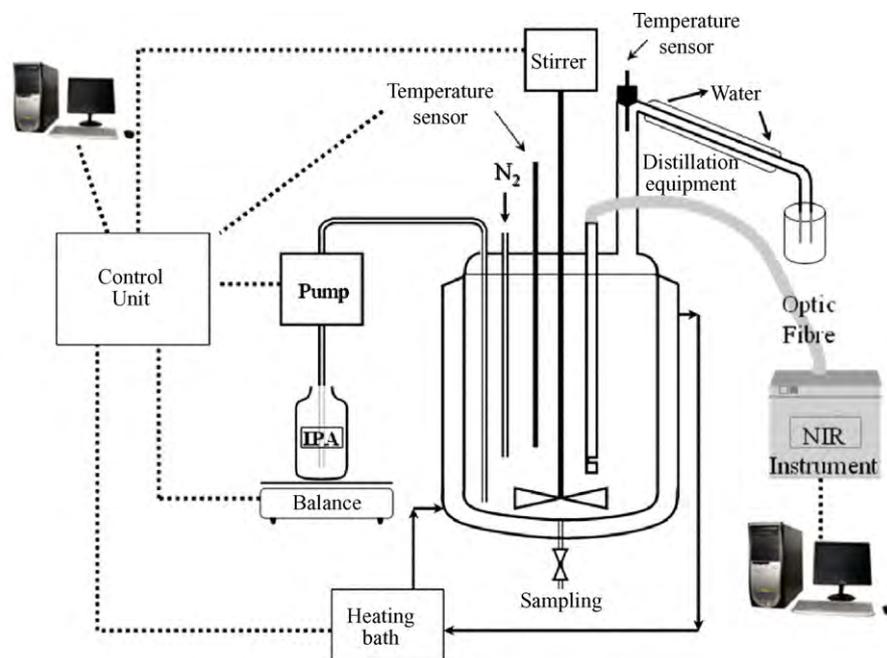


Fig. 1. Schematic drawing of the lab-scale reactor.

KOH/g), isopropanol (Uniqema, ICI Spain, 99.5% pure) and *p*-toluenesulphonic acid (50% in water). All processes were conducted in the LabMax reactor and according to the following procedure: under a continuous stream of N₂, the reactor was loaded with 500 g of acid and heated to the reaction temperature (130 °C), after which the catalyst (2.6 g) was incorporated and addition of the isopropanol/water mixture at a rate of 0.5 g/min started—by exception, a rate of 1 g/min was used in process 7. The water content in the mixtures ranged from 0 to 25%. Table 1 summarizes the characteristics of each process.

At intervals of roughly 60 min, 2–3 g samples of the reaction mixture were withdrawn in order to quantify unreacted acid by titration with 0.1 M KOH in ethanol and aliquots of the distillates were used to determine the composition of the isopropanol–water mixture by NIR spectroscopy. The process was monitored for about 700 min (before completion of the esterification process); by exception, process 4 had to be stopped after 580 min due to experimental problems.

Table 1
Experimental data for the eight processes conducted in the LabMax reactor

Process	Water in Al (%)	Al (g/min)	Unreacted acid (%)	Total time (min)
1	0	0.5	4.2	753
2	5	0.5	7.3	771
3	10	0.5	23.3	582
4	15	0.5	20.9	768
5	20	0.5	33.9	756
6	25	0.5	38.7	711
7	10	1.0	7.6	636
8	0	0.5	8.0	714

3.3. Recording of NIR spectra

Fig. 1 depicts the reactor used; as can be seen, the probe was placed far from the isopropanol and nitrogen inlets, and close to the stirrer, in order to ensure homogeneity in the mixture and facilitate the removal of any nitrogen bubbles potentially formed. Transflectance spectra were recorded at 3 min intervals by using a light path of 0.4 mm (2 mm × 0.2 mm), each spectrum being the average of 32 scans over the wavelength range 1100–2498 nm with a resolution of 2 nm.

3.4. Processing of experimental data

Spectral data were subjected to the multiplicative scatter correction (MSC) treatment in order to suppress scatter. A baseline-offset correction was also used in order to exclude negative spectral values and facilitate application of the spectral non-negativity constraint in the MCR-ALS calculations.

Experimental matrices were subjected to rank analysis by using singular value decomposition (SVD) methodology; both singular values and the profile of the corresponding eigenvectors were considered in determining ranks.

MCR-ALS methodology was applied to the individual data matrix augmented (column-wise) with the spectra for the pure components (myristic acid, isopropyl myristate and isopropanol) used as the initial estimates prior to the ALS iterative process. The sole equality constraint used during the iterative process was the concentration of the pure myristic acid scaled to 1, which allowed the MCR-ALS concentration profiles to be correctly scaled. All calculations were done under the non-negativity constraint in both spectral and concentration profiles. The goodness of the results was determined by comparing the concentration of

unreacted acid by MCR-ALS (\hat{y}) with that obtained by titration with KOH (y) in order to calculate the relative standard error,

$$\%RSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n y_i^2}} \quad (10)$$

and root mean square error (RMSE),

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (11)$$

where n is the number of samples.

4. Results and discussion

Fig. 2a shows the NIR spectra for myristic acid, isopropanol and isopropyl myristate. The most important differences between the spectra lie in the 1800–2200 nm region, where the acid absorption is virtually constant in the 1900–2100 nm region, with a slight increase between 2100 and 2200 nm. Isopropanol and the ester exhibit a band peaking at 2088 and 2136 nm, respectively. The acid and ester also exhibit some differences in the 2300–2500 nm region that are due to combination bands of C–H bonds. Fig. 2b shows the variation of the NIR spectra during a reaction process; the spectral variation is due to the disappearance of myristic acid and formation of an equivalent amount of ester.

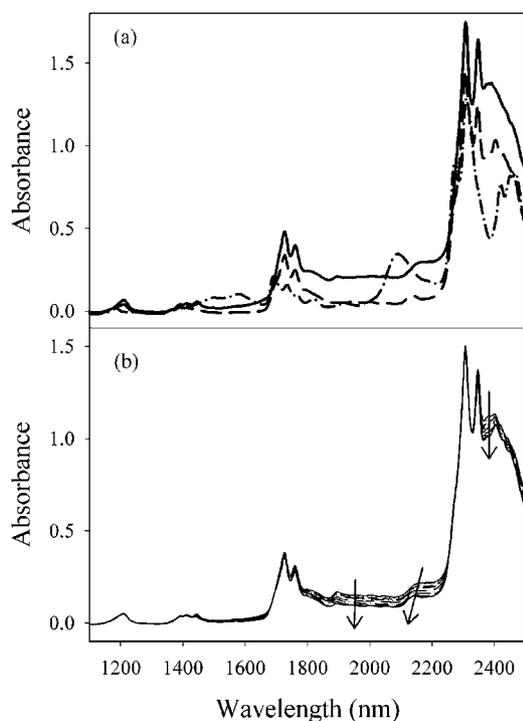


Fig. 2. (a) Spectra for the pure compounds involved in the reaction. Solid line: myristic acid; dash line: isopropyl myristate; dash-dot line: isopropanol. (b) Spectra for the evolving components. Arrowheads indicate temporal variation.

4.1. Application of multivariate curve resolution methodology

The MCR study was conducted on the first six processes shown in Table 1. For each data matrix, which contained the NIR spectra recorded in each process, we estimated the mathematical range (number of chemical components that could be extracted from the data) by examining the corresponding eigenvalues and eigenvectors, using singular value decomposition (SVD) methodology. The first four SVD eigenvalues for process 1 were 3.5×10^6 , 283, 1.4 and 0.7—those for the other processes were similar in magnitude. Since the fourth eigenvalue was less than 1, the number of significant chemical components was 3. Applying MCR required the significant components to be previously identified; myristate ester and myristic acid were clearly two such components, and isopropanol or water might be the third. The alcohol was added in a continuous manner and the water, which was added in the mixture with the alcohol or produced in the reaction, was removed by distillation (the reactor temperature was 130 °C), so its concentration was very low throughout the process. An MCR-ALS computation with the four components and a data matrix augmented with four spectra in order to offset the rank deficiency allowed the methodology described in Section 3.4 to provide the concentration profiles for the four components. The concentration profile for water was near-zero, which suggests a very low concentration of this component in the reaction medium—much lower than that of isopropanol; also, the MCR concentration profile for myristic acid was inconsistent with the concentration determined by titration, with errors (%RSE) of 15% for processes 4–6 (*viz* those involving the highest water contents in the isopropanol–water mixture, and hence the highest concentration of water in the reaction medium). Therefore, water was not a significant component of the reaction medium, so the three components finally chosen for the MCR-ALS study were myristic acid, isopropyl myristate and isopropanol.

With three components, MCR-ALS was implemented by augmenting the spectral matrix for the process with the spectra of the three pure compounds for processes 1–6. The variance accounted for was 100% and the coefficient of determination, R^2 , of the spectral profiles estimated for each process using the corresponding spectra for the pure compounds exceeded 0.999, in all six cases; this testifies to the good fitting of the spectral data. As can be seen from the concentration profiles of Fig. 3, myristic acid and isopropyl myristate evolved virtually inversely. On the other hand, the isopropanol concentration was always low (0.5–2% depending on the particular process) and increased slightly as the reaction developed in some cases by effect of this reactant being continuously added to the medium and the net reaction rate decreasing with time.

Fig. 3 compares the concentration profile for myristic acid with its reference concentrations as determined by titration—which were not used in the calculations; as can be seen, the two were quite consistent. As can be seen from Table 2, the %RSE and RMSE values obtained from the MCR and titration data were accurate—by exception, process 1 gave some spurious results that were ascribed to inappropriate standardiza-

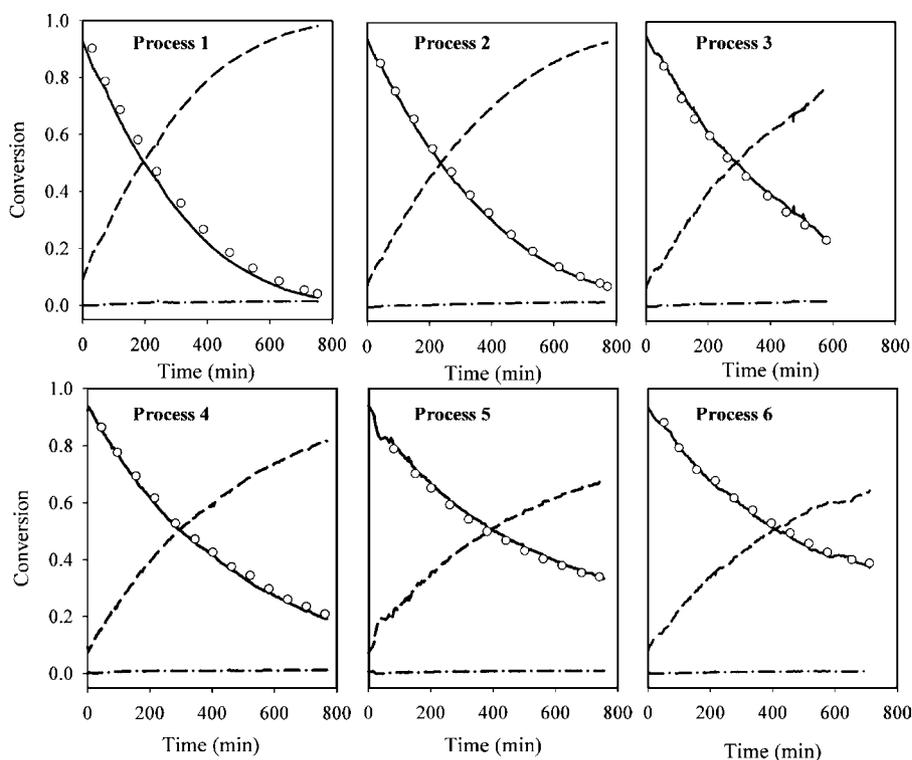


Fig. 3. MCR-ALS estimated concentration profiles for processes 1–6. Solid line: myristic acid; dash line: isopropyl myristate; dash-dot line: isopropanol; (O): acid concentrations obtained by titration.

tion of the KOH solution. The goodness of fit of the MCR-ALS results is reflected in the high correlation between the MCR spectral profiles and the spectra for the pure components ($R^2 = 0.99$), and also in the low %RSE values for the concentration profiles.

4.2. Estimation of kinetic constants

A plot of the natural logarithm of the MCR concentration profile for the acid as a function of time, Eq. (4), was used to calculate the reaction rates for processes 1–6. As can be seen from Fig. 4, the reaction profiles obtained during the first four hours fitted a *pseudo*-first-order law. Linearity was very good in the six processes—the correlation coefficient of the least-squares fit exceeded 0.997 in the six. The interval from 18 to 72 min in process 5 was excluded owing to oscillations in the spectral recording possibly caused by the presence of bubbles in

the light path. Table 3 shows the slope values (k') obtained, which decreased with increasing water content in the isopropanol mixture.

Constant k can be calculated by applying Eq. (3), where $[AI]$ is the alcohol concentration in the reaction medium. Assuming that we do not know the alcohol concentration in the reaction medium, we can estimate it as the average of the concentration profile obtained by MCR-ALS for this compound. Also, in order to check whether the isopropanol MCR-ALS values were accurate, we analysed 7 samples from processes 1 and 7 by gas chromatography. The average concentration (g/g) by chromatography for process 1 was 0.97×10^{-2} (1.03×10^{-2} with MCR) and that for process 7 was 1.66×10^{-2} (2.14×10^{-2} with MCR); therefore, the average MCR values were quite accurate. Table 3 lists k' , average isopropanol concentrations and k values ($k = k'/[AI]$) for each process. The k values obtained for processes 2–5 were very similar to one another, which confirm the accuracy of the proposed methodology for estimating reaction

Table 2
Correlation, relative standard error and root mean square error between MCR-ALS and reference values for the myristic acid concentration

Process	Concentration		
	R^2	%RSE	RMSE
1	0.9996	7.2	0.034
2	0.9998	2.3	0.010
3	0.9992	3.1	0.017
4	0.9997	2.7	0.014
5	0.9997	1.8	0.010
6	0.998	2.1	0.012

Table 3
Kinetic data for processes 1–6

Process	Water (%)	$k' (\times 10^3 \text{ min}^{-1})$	[Isopropanol] ($\times 10^3$ mean)	$k (\text{min}^{-1})$
1	0	3.16 ± 0.03	10.3 ± 0.5	0.31
2	5	2.63 ± 0.02	12.4 ± 0.6	0.21
3	10	2.24 ± 0.03	11.4 ± 0.7	0.20
4	15	2.11 ± 0.01	9.4 ± 0.3	0.22
5	20	1.66 ± 0.04	7.3 ± 0.3	0.23
6	25	1.61 ± 0.02	5.2 ± 0.3	0.31

k' and [isopropanol] include confidence interval (95%)

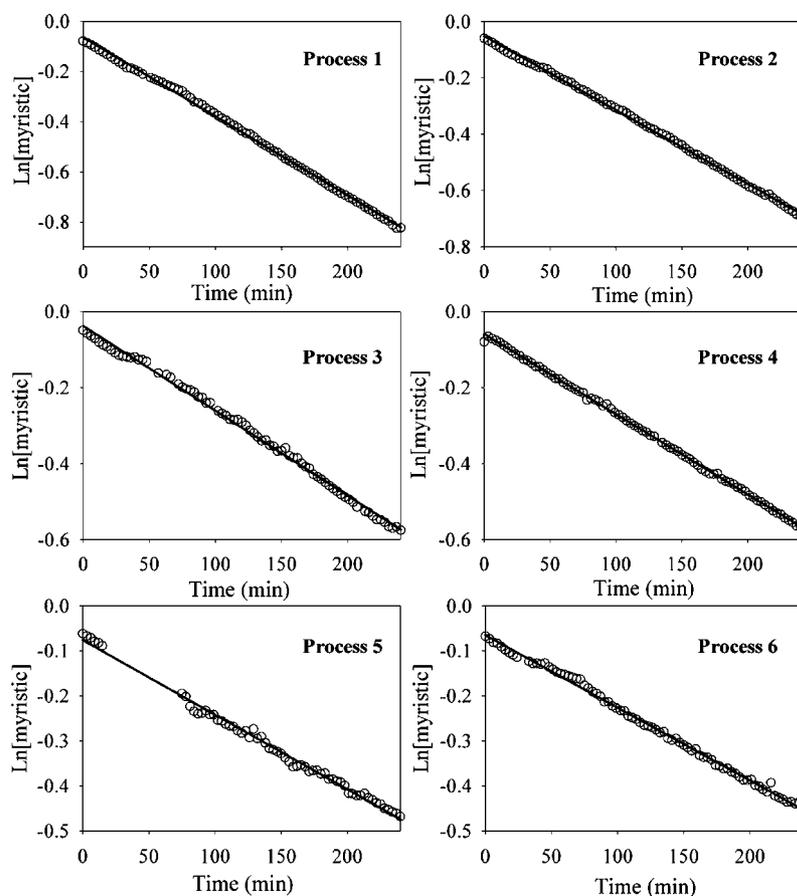


Fig. 4. Plot of Ln [myristic] vs. time for processes 1–6. Solid line: regression line.

constants; spurious estimates of k as seen for processes 1 and 6 could be caused by deviations in the estimated isopropanol concentration.

4.3. Study of the esterification process

Once concentration profiles and kinetic data were obtained, the process was evaluated in terms of yield and reaction rate as a function of the proportion of water in the alcohol mixture. A comparison of the concentration profiles of myristic acid in Fig. 3 reveals that the reaction rate decreases and the profile appears to level off at a higher asymptote, all of which results in a decreased reaction yield. The effective rate constant (k' , Table 3) halved from pure isopropanol to a mixture containing 20% water, the decrease being only 35% in the presence of less water (15%). Therefore, reusing distillates with substantial water contents is industrially viable, albeit at the expense of a decreased reaction rate and yield. This requires the knowledge of the composition of the distillates during the reaction, which was monitored by NIR spectroscopy here. Previously, a PLS calibration model was constructed from 21 laboratory samples containing 50–100% isopropanol (and 0–50% water) at 5% intervals. The calibration models for isopropanol and water were constructed from second-derivative values, using two PLS factors over the wavelength range 1100–2498 nm for isopropanol and 1350–2150 nm for water. The %RSE between the predicted

and reference values for the 21 calibration samples was 0.6% for isopropanol and 1.5% for water.

Fig. 5 shows the variation of the isopropanol concentration in the distillates of processes 1, 2, 4 and 6; as can be seen, it increased with time, to the level in the mixture fed to the reactor. This can be ascribed to an increased esterification rate during the first few minutes of the process that subsequently decreased as the reaction completed. The distillates obtained at the end of the process can be directly reused untreated as feed mixtures in subsequent processes; those collected at an

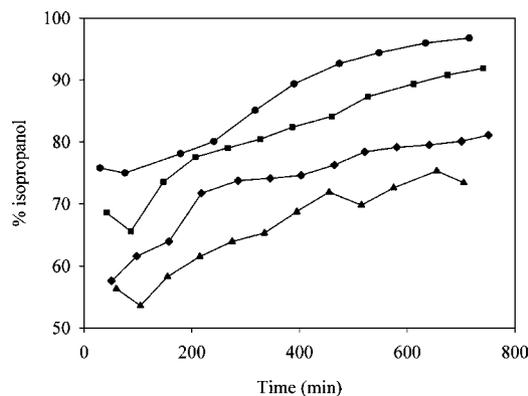


Fig. 5. Concentration of isopropanol in distilled samples from process 1 (●), 2 (◆), 4 (■) and 6 (▲).

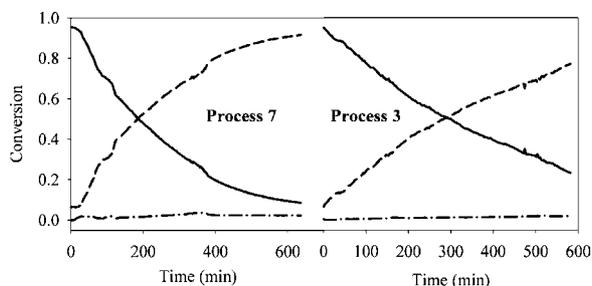


Fig. 6. MCR-ALS estimated concentration profiles for processes 3 and 7. Solid line: myristic acid; dash line: isopropyl myristate; dash-dot line: isopropanol.

early stage, however, must previously be re-distilled in order to concentrate the alcohol.

One other way of increasing the reaction rate with reused distillates is by increasing the rate of addition of isopropanol. In order to assess the effect, we applied MCR-ALS methodology to a process involving an increased rate of addition of the alcohol-water mixture (process 7, addition rate = 1 g/min) and compared the resulting profiles with those for process 3, which was conducted at 0.5 g/min; both used a feed mixture containing 10% water. As can be seen from the concentration profiles of Fig. 6, the increased addition rate resulted in an increased reaction rate and in decreased acid concentrations. Experimentally, increasing the addition rate raises the isopropanol concentration in the medium, and hence the reaction rate. This approach entails increasing the volume of isopropanol-water distillate and reactant consumption as a result. Thus, the proportion of unreacted acid at the end of process 7 (7.6%) was similar to that for process 2 (7.3%); however, the amount of isopropanol-water mixture consumed in process 7 was roughly twice that used in process 2.

The ability to predict the concentration profiles for new processes allows their monitoring in real time. Thus, by using Eqs.

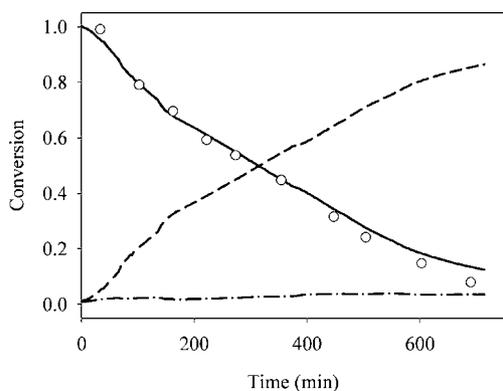


Fig. 7. Predicted concentration profiles for process 8. (—) myristic acid; (---) isopropyl myristate; (-.-) line: isopropanol; (○): acid concentrations obtained by titration.

(8) and (9) in combination with the estimated spectral profiles for process 1 (that shows the high reacted acid), we obtained the concentration profiles for process 8, which was not used in previous calculation. Fig. 7 shows the resulting concentration profiles and the acid concentrations obtained by titration; as can be seen, the fit was similarly good to that of Fig. 3 (%RSE = 6.8).

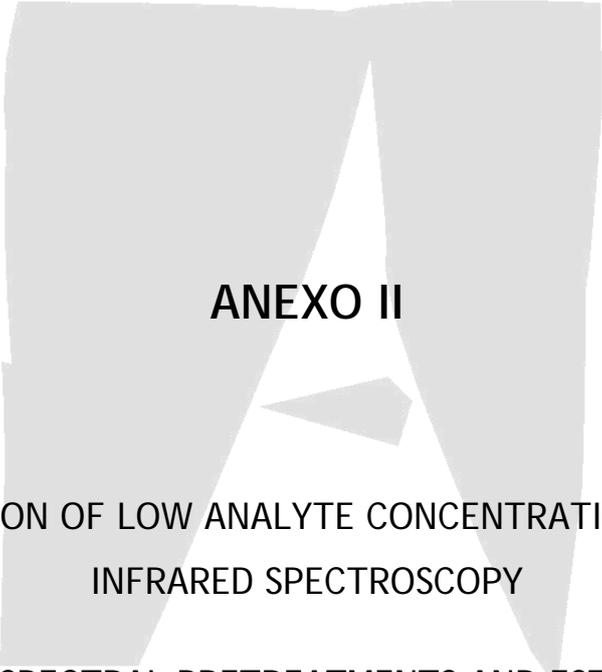
5. Conclusions

The application of soft-modeling MCR-ALS methodology to in-line NIR data allows estimating the variation of the reactant concentrations from spectra recorded during the process; the sole source of external information is the pure spectra of the species involved in the reaction. This avoids the need to withdraw and analyse samples of the reaction mixture in order to elucidate the kinetics of the process and/or develop calibration models. Our results in terms of %RSE and RMSE are as good as those provided by other chemometric calibration techniques such as PLS or PCR, with the advantage that the proposed alternative involves less laboratory work.

Estimating kinetic constants from MCR concentration profiles allows one to establish the effect of experimental variables on the reaction rate and yield; also, the information thus obtained allows one to take steps towards minimizing consumption of raw materials and optimizing parameters such as temperature or the rate of addition of the reactants. Although its industrial implementation would require a previous upscaling study at a pilot plant, the process methodology has proved viable for monitoring industrial processes.

References

- [1] T. Kocsisová, J. Cvengros, J. Lutisan, *Eur. J. Lipid Sci. Technol.* 107 (2005) 87–92.
- [2] R. Aafaqi, A.R. Mohamed, S. Bhatia, *J. Chem. Technol. Biotechnol.* 79 (2004) 1127–1134.
- [3] E. Sendzikiene, V. Makareviciene, P. Janulis, S. Kitrys, *Eur. J. Lipid Sci. Technol.* 106 (2004) 831–836.
- [4] M. Blanco, I. Villaroya, *Trends Anal. Chem.* 21 (4) (2002) 240–240.
- [5] A. Cherfi, G. Févotte, *Macromol. Chem. Phys.* 203 (2002) 1188–1193.
- [6] C. Pasquini, S.H.F. Scafi, *Anal. Chem.* 75 (2003) 2270–2275.
- [7] S.E. Barnes, E.C. Brown, M.G. Sibley, H.G.M. Edwards, I.J. Scowen, P.D. Coates, *Appl. Spectrosc.* 59 (5) (2005) 611–619.
- [8] M. Watari, Y. Ozaki, *Appl. Spectrosc.* 59 (5) (2005) 600–610.
- [9] M. Blanco, D. Serrano, *Analyst* 125 (2000) 2059–2064.
- [10] A. De Juan, R. Tauler, *Anal. Chim. Acta.* 500 (2003) 195–210.
- [11] R. Tauler, A. Smilde, B. Kowalski, *J. Chemometr.* 9 (1) (1995) 31–58.
- [12] R. Tauler, B. Kowalski, S. Flemming, *Anal. Chem.* 65 (1993) 2040–2047.
- [13] M.S. Larrechi, F.X. Rius, *Appl. Spectrosc.* 58 (1) (2004) 47–53.
- [14] P. Gemperline, E. Cash, *Anal. Chem.* 75 (2003) 4236–4243.
- [15] http://personal.ecu.edu/gemperlinep/Research/gemper_software.html. P. Gemperline, East Carolina University.
- [16] MATLAB, version 7.0, The MathWorks, 2004.



ANEXO II

DETERMINATION OF LOW ANALYTE CONCENTRATIONS BY NEAR-
INFRARED SPECTROSCOPY

EFFECT OF SPECTRAL PRETREATMENTS AND ESTIMATION OF
MULTIVARIATE DETECTION LIMITS

M. Blanco, M. Castillo, A. Peinado y R. Beneyto

Analytica Chimica Acta 581 (2007) 318-323

Determination of low analyte concentrations by near-infrared spectroscopy: Effect of spectral pretreatments and estimation of multivariate detection limits

Marcel Blanco^{a,*}, Miguel Castillo^a, Antonio Peinado^a, Rafael Beneyto^b

^a *Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*

^b *Development and Quality Assurance Department, ICI Spain S.A. (Mevisa Site),
Ctra Hostalric a Tossa km 1.8, 08495 Fogars de la Selva, Barcelona, Spain*

Received 18 May 2006; received in revised form 20 July 2006; accepted 10 August 2006

Available online 18 August 2006

Abstract

Near infrared spectroscopy (NIRS) was used in combination with partial least squares (PLS) calibration to determine low concentrated analytes. The effect of the orthogonal signal correction (OSC) and net analyte signal (NAS) pretreatments on the models obtained at concentrations of analyte near its detection limit was studied. Both pretreatments were found to accurately resolve the analyte signal and allow the construction of PLS models from a reduced number of factors; however, they provided no substantial advantage in terms of %RSE for the prediction samples. Multiple methodologies for the estimation of detection limits could be found in the bibliography. Nevertheless, detection limits were determined by a multivariate method based on the sample-specific standard error for PLS regression, and compared with the univariate method endorsed by ISO 11483. The two methods gave similar results, both being effective for the intended purpose of estimating detection limits for PLS models. Although OSC and NAS allow isolating the analyte signal from the matrix signal, they provide no substantial improvement in terms of detection limits. The proposed method was used to determine 2-ethylhexanol at concentrations from 20 to 1600 ppm in an industrial ester. The detection limit obtained, round 100 ppm, testifies to the ability of NIR spectroscopy to detect low concentrated analytes.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Near infrared spectroscopy; Detection limit; Orthogonal signal correction; Net analyte signal; Low concentration analytes

1. Introduction

In recent years, the advantages of near infrared spectroscopy over other analytical techniques have increased its acceptance in various industrial fields [1]. However, this technique has the disadvantage that the typically low molar absorptivity of NIR signals considerably reduces its sensitivity. The growing demand for industrial quality has fostered the development of sensitive, expeditious techniques capable of determining trace amounts of analytes as it is often trace species that dictate the quality of products and their fitness for purpose.

The substantial instrumental advances in sensitivity and precision of recent years have allowed the concentration levels

that can be determined by NIR spectroscopy to be significantly reduced. However, NIR signals are complex [1] and require the use of multivariate calibration methods to construct models with an adequate predictive ability. The use of signal filtering algorithms to reduce noise or suppress extraneous contributions (*i.e.* isolate the analyte signal) such as the orthogonal signal correction (OSC) [2–4] or net analyte signal (NAS) pretreatments [5–6] may help simplify calibration models and improve the results when amounts of analyte near its detection limit are to be determined.

The knowledge of the detection or determination limit for an analyte can be of great help with a view to constructing accurate models for determining low analyte contents. There are various methods for calculating detection limits by reverse multivariate calibration that can be applied to the body of calibration samples and the prediction set. Some such methods are briefly described in the Theory section below.

* Corresponding author. Tel.: +34 935811367; fax: +34 935811367.

E-mail address: marcel.blanco@uab.es (M. Blanco).

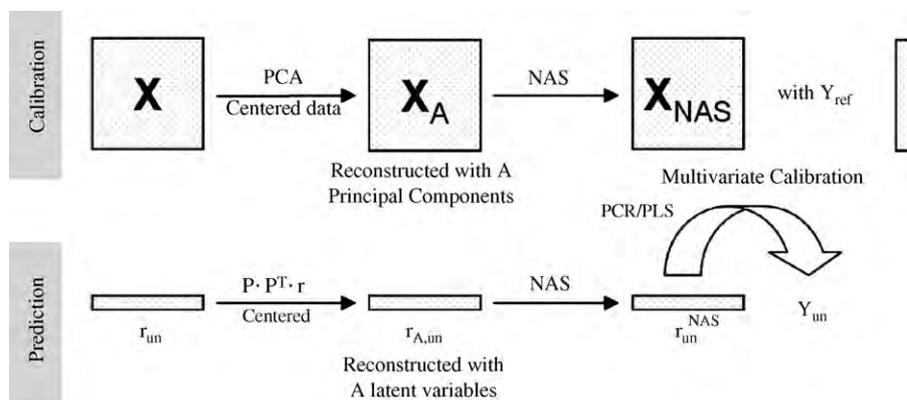


Fig. 1. Scheme of the methodology for NAS calculation proposed in this paper.

In this work, we determined the content of a residual alcohol in an industrial ester. The presence of the alcohol alters the sensory properties of the final product. Thus, for the end-product to possess the desired properties, its content in residual alcohol should be very low. For this reason, the alcohol is removed by distillation to a concentration below its sensory detection limit (*ca.* 200 ppm). The alcohol was determined by constructing PLS models from both raw data and data filtered using the OSC and NAS algorithms. The detection limit obtained with each pretreatment was also determined.

2. Theory

Below are briefly described the chemometric techniques used in this work. A more detailed description and the process used to derive the mathematical equations used can be found in the references cited.

2.1. Orthogonal signal correction (OSC)

The OSC algorithm [4–6] corrects spectral data on the basis of analyte concentrations. Thus, it removes bilinear components from matrix \mathbf{X} that are orthogonal (unrelated) to \mathbf{Y} . This algorithm is similar to the ordinary PLS algorithm, the difference being that OSC minimizes the covariance between \mathbf{X} and \mathbf{Y} . In principle, OSC could be useful for quantifying low concentrated analytes provided it can efficiently remove major extraneous contributions to the analyte signal [7].

2.2. Net analyte signal (NAS)

The net analyte signal was defined by Lorber [5] as that part of the sample spectrum, which is related to the analyte and orthogonal to the interferences. The NAS algorithm splits the data matrix \mathbf{X} into two containing spectral information pertaining to the analyte k , \mathbf{X}_k , and to all other variability sources, \mathbf{X}_{-k} , including the contribution of the interferences ($\mathbf{X} = \mathbf{X}_k + \mathbf{X}_{-k}$). The NAS vector for a given sample is calculated by projecting spectrum \mathbf{r} onto the space defined by the interferences (matrix \mathbf{X}_{-k}), the NAS being the orthogonal resultant (\mathbf{r}^*), which is defined as

$$\mathbf{r}^* = [\mathbf{I} - \mathbf{X}_{-k}^T (\mathbf{X}_{-k}^T)^+] \mathbf{r} \quad (1)$$

A more detailed description of NAS and the procedure used to calculate matrices \mathbf{X}_k and \mathbf{X}_{-k} can be found elsewhere [6–10]. Usually, the NAS algorithm is implemented as a univariate calibration technique [8–11]. In this work, however, we used it as a spectral pretreatment in order to obtain the net analyte signal for subsequent multivariate calibration.

The procedure, which is depicted in Fig. 1, involved calculating matrix \mathbf{X}_A , which was reconstructed from A principal components (PCA) with centring of the original data. The purpose of the PCA was to filter noise in the data matrix prior to calculating \mathbf{X}_{-k} in order to avoid spurious NAS calculations. Once the matrix was constructed, the NAS vector for each sample was calculated according to Lorber et al. [6] and improvements in Ferré et al. [8], after which the NAS vectors and concentration values were used for PCR or PLS calibration. The spectra for the prediction samples were centred by subtracting the average spectrum for the calibration samples and reconstructed from A latent variables ($\mathbf{r}' = \mathbf{P}\mathbf{P}^T\mathbf{r}$, where \mathbf{P} is the matrix consisting of A loadings from a PCA), its NAS vector being calculated and used in the corresponding PCR or PLS model to predict the unknown concentration.

2.3. Detection limits for PLS models

The multivariate detection limit (MDL) for an inverse calibration system can be calculated easily from $\text{MDL}_k = 3 \cdot \|\varepsilon\| \cdot \|b_k\|$ [12], where $\|b_k\|$ and $\|\varepsilon\|$ denote the Euclidean norm of the regression coefficients of the PLS model and instrumental noise, respectively. However, it is desirable to have a quantitative measure of the detection limit in the predictions obtained when applying a multiway calibration model [13]. In that situation, the detection limit is defined as a function of the variance of the concentration predicted by the model:

$$\text{MDL}_k = \Delta(\alpha, \beta) \hat{\text{var}}(c_{0,k})^{1/2} \quad (2)$$

where $\Delta(\alpha, \beta)$ is a parameter of a non-central Student's t distribution [14] with v degrees of freedom and considers the likelihood of making errors α and β . A former empirical expression for calculating the variance in Eq. (2) is that reported by Faber and Kowalski [15]:

$$\hat{\text{var}}(c_{\text{un}}) = (I^{-1} + h_{\text{un}}) [\sigma_c^2 + \|b_k\|^2 \sigma_R^2] + \|b_k\|^2 \sigma_{r_{\text{un}}}^2 \quad (3)$$

where I is the number of samples in the calibration model; σ_c^2 the variance of the concentrations (in the reference method); σ_R^2 and σ_{un}^2 are the variances of the method – the NIR error in this case – for the calibration samples and unknown sample, respectively; $\|b_k\|$ the Euclidean norm of the vector of regression coefficients; and h_{un} is the leverage of the unknown sample in the calibration space. Subsequent papers introduce modifications in Eq. (3): thus, Boqué and Rius [16] incorporated the hypothesis test by calculating the spectrum for the sample in the absence of analyte, and Olivieri [17] incorporated the corresponding variances and confidence limits of the pre-processing method (NAS or OSC). However, Eq. (3) has been further simplified by Faber and Bro [13]:

$$\hat{v}\hat{r}(c_{un}) = [(1 + h)\text{MSEC} - \sigma_c^2] \quad (4)$$

where MSEC is defined as [18]:

$$\text{MSEC} = \frac{\sum_{i=1}^I (\hat{y}_i - y_i)^2}{I - \text{d.f.}} \quad (5)$$

d.f. could be estimated by the equation proposed by Van der Voet [19]. Eqs. (3) and (4) are compared by Fernández et al. [18]. In this study Eq. (4) is selected for the estimation of multivariate detection limits.

Taking into account the complexity and the variety of equations for the estimation of detection limits in multivariate terms, a univariate approximation proposed by Ortiz et al. [20] has been used. When the analyte signal is univariate, ISO [21] recommends constructing a signal *versus* concentration plot. The variance of the signal, \hat{y} , can thus be defined as [22]:

$$\text{V}\hat{r}(\hat{y}) = w_x^2 \hat{\sigma}^2 \quad (6)$$

where:

$$w_x^2 = \frac{1}{r} + \frac{1}{I} + \frac{(x - \bar{x})^2}{\sum_{i=1}^I (x_i - \bar{x})^2} \quad (7)$$

r being the number of determinations of the test sample and I that of calibration samples and $\hat{\sigma}^2$ is calculated by Eq. (5) with d.f. = 2.

Detection limit can be defined as:

$$\text{DL}(\alpha, \beta) = \frac{\Delta(\alpha, \beta)(\text{V}\hat{r}(\hat{y}))^{1/2}}{\hat{b}} \quad (8)$$

where \hat{b} is the slope of the calibration curve.

Ortiz et al. converted the above-described univariate method into a multivariate one by assuming the detection limit to be insensitive to any linear transformation applied to the signal (*i.e.* that the detection limit would be the same even if the spectroscopic data were modified by application of a linear PLS variable reduction algorithm). This allowed the ISO-endorsed method to be generalized to any multivariate calibration method and detection limits to be determined by applying Eq. (8) to the linear regression of the PLS predicted *versus* reference values.

3. Experimental and methods

3.1. Instrumental and software

NIR spectra were recorded on a Foss NIRSystems 5000 near infrared spectrophotometer equipped with a Vial Heater Module (VHM). Spectra were recorded in the transmittance mode at 75 °C, using tightly closed glass vials of 8 mm light path. Each individual spectrum is the average of 32 scans done at 2 nm intervals over the wavelength range 1100–2200 nm.

A Hewlett-Packard HP 5890 Series II gas chromatograph equipped with a flame ionization detector and furnished with a Supelco SPB-5 fused silica capillary column [poly(5% phenyl/95% dimethyl)siloxane, 15 m × 0.25 mm, 0.25 μm film thickness] was used to determine analyte concentrations.

The NAS was calculated in Matlab 7.0 (The Mathworks, Inc., Natick, MA), using home-made code, and the OSC with the software PLS_Toolbox 3.5 (Eigenvector Research, Inc., Manson, WA) as implemented in Matlab 7.0. PLS calibration models were constructed with The Unscrambler v. 9.2 (Camo Process, Oslo, Norway).

3.2. Samples

The samples, which are liquid at room temperature, consist of mixtures of esters of 2-ethylhexanol with middle- and long-chain fatty acids obtained by direct esterification in the presence of appropriate catalysts. Following esterification, residual 2-ethylhexanol was removed by distillation and steam sweeping. Samples were withdrawn from different batches only during the last step. In order to have the sample set span an appropriate range of evenly distributed concentrations, samples containing low 2-ethylhexanol concentrations were spiked with substantial amounts of the alcohol. This allowed a set of 33 samples spanning the alcohol content range 22–1630 ppm to be obtained.

3.3. Gas chromatographic analysis

The 2-ethylhexanol content in industrial samples was experimentally determined by gas chromatography, using helium as carrier gas and a split ratio of 1:100. Samples were injected undiluted into the chromatograph. Column temperature, initially 90 °C for 2 min, raised at 10 °C min⁻¹ to 280 °C, final temperature held for 10 min to facilitate elution of low volatile compounds, with injector and detector temperatures of 310 °C. The response factor for the target concentration range was determined by adding an accurately weighed amount of 2-ethylhexanol to aliquots of sample with a low content in the alcohol.

3.4. NIR data processing

Spectral data were subjected to second-derivative treatments based on the Savitzky–Golay algorithm with a second-order polynomial and a window size of 11 points.

NAS computations were done following the scheme of Fig. 1 and the rules proposed by Lorber et al. [6] and Ferré et al. [8].

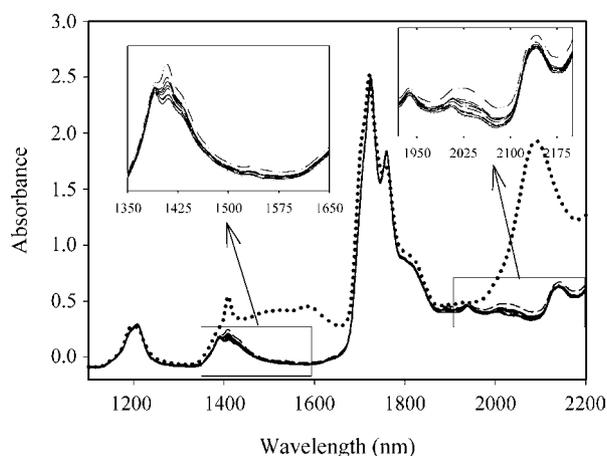


Fig. 2. Effect of 2-ethylhexanol concentration in NIR spectra for samples between 22 ppm and 3.6%. Dotted line is the spectrum of 2-ethylhexanol.

OSC calculations were done with the mean-centred matrix, using the `osccalc.m` and `oscapp.m` files in PLS_Toolbox 3.5. The number of OSC factors was selected in terms of the spectral variance accounted for by successive factors and the %RSEP for external samples.

PLS models were constructed by cross-validation, using the leave-one-out method. The optimum number of PLS factors was taken to be that minimizing the relative standard error of prediction, %RSEP, for the samples in the validation set.

4. Results and discussion

As can be seen from Fig. 2, the spectra for a set of samples with variable 2-ethylhexanol contents exhibited a slight absorbance decrease in the regions 1350–1650 and 1900–2200 nm. 2-ethylhexanol absorbs maximally in the regions 1300–1600 and 2000–2200 nm, which correspond to the first overtone and combination bands of the O-H bond, respectively.

PLS calibration was done over the 22–1630 ppm range, using a total of 33 samples that were split into calibration and prediction sets consisting of 20 and 13 samples, respectively. The samples to be included in each set were chosen from the PC1 *versus* PC2 scatter plot of a PCA, in such a way that they span the concentration range and belong to different production batches. The best results were obtained by using second-derivative spectra of the 1300–1500 nm region in combination with four PLS factors. Table 1 shows the figures of merit of the regression

line, and the corresponding RSEC and RSEP values. The slopes and intercepts were close to 1 and 0, respectively, and both values were included in the confidence interval. The %RSE values obtained can be deemed good for the concentration range spanned by the calibration model.

The spectra for the previous 33 samples, with 20 in a calibration set and 13 in a prediction set, were subjected to the OSC and NAS pretreatments prior to PLS calibration. With OSC as applied to the mean-centred calibration matrix, a single factor sufficed to remove most of the spectral variance not related to the analyte concentration. Thus, one OSC factor suppressed 24.9% of the spectral variance and two 28.0%. In order to select the most effective number of OSC factors while avoiding overfitting, we examined the predictive ability of the model with external samples. With one OSC factor, the PLS calibration model required two factors – which provided the optimum PRESS – and gave an %RSEP of 14.8. With two OSC factors, the best model required a single PLS factor and predicted the external samples with an %RSEP of 17.9 (*i.e.* there was overfitting, which is a common occurrence with this spectral pretreatment). Therefore, the best model was taken to be that involving a single OSC factor and modelling the remaining information with two PLS factors. Table 1 shows the figures of merit of the regression of NIR predicted *versus* GC experimental values and the %RSE values for the calibration and prediction sets.

The NAS pretreatment was applied to a mean-centred calibration matrix reconstructed from seven factors of a PCA, which efficiently removed noise in the NIR spectra. There is no universally accepted criterion for choosing the number of PCs to be used to reconstruct the matrix in this case, which is quite a consequential decision. For this reason, we examined the influence of the number of PCs used on %RSEC and %RSEP and found the best results to be those provided by reconstructing the matrix with seven PCs, which jointly accounted for 99.3% of the total variance. Then, we calculated the NAS vector for each calibration sample using the procedure described in the Theory section. Fig. 3 shows the initial second-derivative spectra, which expose the very slight differences between samples, and the NAS vectors, which exhibit greater differences. The NAS vector calculated for the prediction samples following centring and reconstruction of their spectra from seven PCs was used to estimate the corresponding analyte concentrations by PLS. Table 1 shows the figures of merit of the regression of NIR predicted *versus* GC experimental values, and the %RSE values for the calibration and prediction sets.

Table 1

Figures of merit and %RSE for PLS model and for OSC-PLS and NAS-PLS models, for calibration and external validation samples

Parameter	PLS model		OSC (1 factor)-PLS		NAS-PLS	
	Calibration	External validation	Calibration	External validation	Calibration	External validation
PLS components	4		2		1	
Concentration range (ppm)	22–1630	56–1060	22–1630	56–1060	22–1630	56–1060
Slope	0.98 ± 0.07	0.99 ± 0.12	0.99 ± 0.04	0.96 ± 0.11	0.98 ± 0.07	0.96 ± 0.12
Intercept	6 ± 33	10 ± 42	2 ± 20	42 ± 40	6 ± 33	13 ± 43
<i>r</i>	0.991	0.984	0.997	0.986	0.991	0.983
%RSEC/P	9.9	12.6	6.1	14.8	9.8	13.1

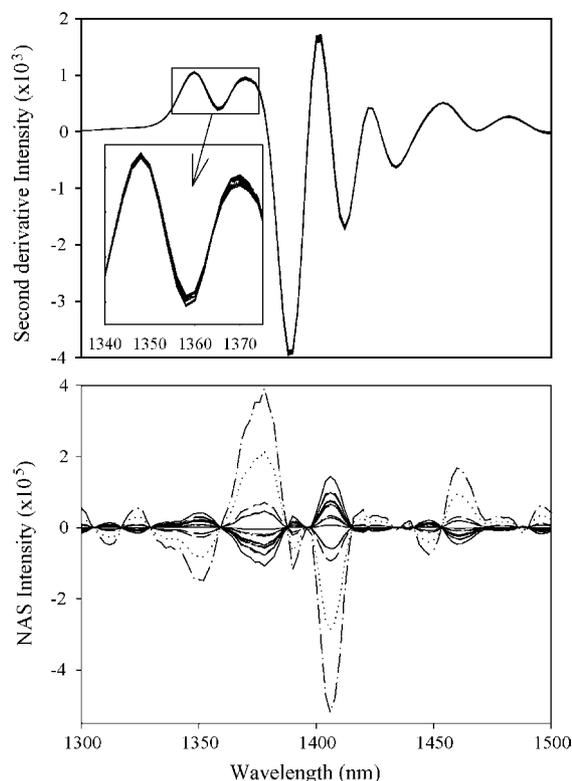


Fig. 3. 2nd derivative spectra of calibration samples (2-ethylhexanol concentration in the range 20–1630 ppm) and their corresponding NAS vectors.

As can be seen from Table 1, the calibration and prediction results provided by the models were very similar as regards figures of merit. Therefore, neither OSC nor NAS improved the predictions for external samples, as previously found by other authors [7,23]. The greatest advantage of OSC and NAS is that they provide simpler models with a reduced number of PLS factors. Thus, the number of factors required was reduced to two with OSC and one with NAS. Of the two, NAS proved more efficient in removing spectral information not related to the analyte. Thus, OSC with a single factor successfully removed 24.9% of the information contained in matrix \mathbf{X} , whereas NAS removed as much as 55.9%. This influenced the variances accounted for by the PLS model. As can be seen in Table 2, four PLS factors were required to account for 98.2% of the variance in \mathbf{Y} ; by contrast a single PLS factor explained an identical proportion of the total variance with NAS, and two PLS factors accounted for an even higher proportion with OSC. A comparison of the spectral variance explained by the NAS and OSC calibration models

Table 3

Detection limits (in ppm) for the three PLS models

Pretreatment	MDL(α) (Eq. (4))	MDL(α, β) (Eq. (4))	DL(α) (Eq. (8))	DL(α, β) (Eq. (8))
–	117 ± 3	187 ± 4	114	189
OSC	109 ± 2	177 ± 4	137	228
NAS	86 ± 1	139 ± 2	120	200

Best results for each methodology are in bold.

confirmed that the former was more efficient in filtering information not correlated to the target parameter; thus, the variance accounted for by a single factor was virtually 100% and much higher than that explained with OSC (58.6%).

4.1. Calculation of detection limits

Eqs. (4) and (8) described under Theory were used with the three calibration models constructed over the concentration range 22–1630 ppm to calculate detection limits. Table 3 shows the multivariate detection limits MDLs with α and β errors ($\Delta(\alpha, \beta)$, $\alpha = \beta = 0.05$) and also with only α error (substituting $\Delta(\alpha, \beta)$ by $t(\alpha)$). MDLs were calculated by applying Eqs. (4) and (8) to the 13 prediction samples. Parameters of the equations were as follows: number of calibration samples, $I = 20$; variance of the chromatographic method as the calculated variance for 17 samples containing variable amounts of analyte and analysed in triplicate, $\sigma_c^2 = 42.8$ ppm. The degrees of freedom for Eq. (4) were calculated according the equation proposed by Van der Voet [19]; and for Eq. (8) were $(I - 2)$. Finally, for Eq. (4) the detection limit is the mean value of the 13 samples, because this equation allows obtaining a detection limit for each of them.

Although Eqs. (4) and (8) are obtained from different sources, the results provided by both methodologies are in agreement. Results are similar for OSC and PLS, and there are little differences for NAS, due to the fact that leverage values for this pretreatment are lower than those provided by OSC and PLS. In that sense, both equations could be applied for the estimation of the detection limit for multivariate calibration models.

In a prior study of multivariate detection limits by applying Eq. (3) to the same data, no agreement was found with the results obtained with Eq. (8). Specifically, Eq. (3) provided very optimistic detection limits for OSC and NAS (~50 ppm) and, in the contrary, similar values to those shown in Table 3 were obtained for the PLS model. Olivieri [17] consider modifying Eq. (3) for including not only the effect of calibration concentra-

Table 2
Explained variances for \mathbf{X} and \mathbf{Y} of three proposed calibration models after pre-treatments

PLS factor	–		OSC		NAS	
	Var. X (%)	Var. Y (%)	Var. X (%)	Var. Y (%)	Var. X (%)	Var. Y (%)
1	53.0	66.2	58.6	94.1	100.0	98.2
2	80.7	85.4	76.0	99.3	100.0	98.5
3	87.4	96.0	85.3	99.7	100.0	98.9
4	91.9	98.2	94.4	99.7	100.0	99.1

In bold, selected number of factors for each model.

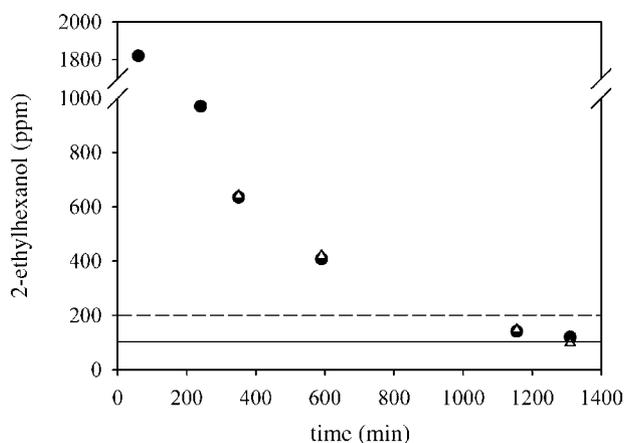


Fig. 4. Evolution of 2-ethylhexanol concentration during a deodorization process. (●) NIR values and (△) gas chromatography values. (Footnote: dash line corresponds to sensorial detection limit and solid line corresponds to the NIR detection limit with α error).

tion and signal errors but also of errors introduced by the OSC and NAS pre-processing technique. However, this increases the complexity of the calculation of multivariate detection limits and involves the application of different equations for different pretreatments. Eqs. (4) and (8) are easy-to-use and allowed the detection limits to be calculated properly independently of the pretreatment applied in the calibration step.

As can be seen from Table 3, OSC and NAS do not provide lower detection limits than those obtained with the PLS model, only the NAS detection limit estimated by Eq. (4) is slightly lower. Therefore, as noted in discussing the %RSEC and %RSEP values above, NAS and OSC fail to substantially improve the predictions for external samples.

4.2. Application to industrial process control

The proposed PLS model is used for the routine control of the industrial process. Fig. 4 shows the variation of the NIR predicted concentration during the deodorization process. These values are in agreement with gas chromatography values for samples with low content in 2-ethylhexanol. The calibration model allowed the process to be monitored approximately down to 100 ppm of the alcohol, which roughly coincides with the detection limit obtained considering only an α error (see Table 3). Application of the proposed method to other deodorization processes provided similar results, which confirms that it can be effectively used with a view to assuring the required quality in the manufactured product.

5. Conclusions

Calibration models constructed from spectra subjected to an OSC or NAS pretreatment provide no substantial improvement

in the prediction samples in terms of %RSEP. The sole advantage gained by using these chemometric techniques is greater simplicity in the resulting models. Similar conclusions could be obtained by comparing detection limits for the three calibration models.

Two equations are tested to estimate the multivariate detection limit (Eqs. (4) and (8)) that allow estimate the detection limit in a simple manner, using expressions that are widely known among analytical chemists.

According to the results obtained in this paper, NIR spectroscopy allows detect minor analytes (concentration round 100 ppm) and can be applied, with successful, to the quality control of industrial products.

Acknowledgements

The authors are grateful to Spain's MCyT for funding this research within the framework of Project BQU2003-04247. Miguel Castillo also wishes to acknowledge the support from the DURSI of the Generalitat de Catalunya and European Social Fund in the form of a scholarship.

References

- [1] M. Blanco, I. Villarroya, *Trac* 21 (2002) 240.
- [2] S. Wold, H. Antti, F. Lindgren, J. Öhman, *Chemometr. Intell. Lab. 44* (1998) 175–178.
- [3] J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, S. Wold, *Chemometr. Intell. Lab. 44* (1998) 229–244.
- [4] T. Fearn, *Chemometr. Intell. Lab. 50* (2000) 47–52.
- [5] A. Lorber, *Anal. Chem.* 58 (1986) 1167–1172.
- [6] A. Lorber, K. Faber, B. Kowalski, *Anal. Chem.* 69 (1997) 1620–1626.
- [7] E. Bertran, H. Iturriaga, S. MasPOCH, I. Montoliu, *Anal. Chim. Acta* 431 (2001) 303–311.
- [8] J. Ferré, S.D. Brown, F.X. Rius, *J. Chemometr.* 15 (2001) 537–553.
- [9] N.M. Faber, *Anal. Chem.* 70 (1998) 5108–5110.
- [10] N.M. Faber, *J. Chemometr.* 12 (1998) 405–409.
- [11] M. Martínez, D. Picón, J.L. Martínez, A. Garrido, A. Espinosa, A. Muñoz, F. Salinas, *Talanta* 59 (2003) 1107–1116.
- [12] K.S. Booksh, B.R. Kowalski, *Anal. Chem.* 66 (1994) 15.
- [13] N.M. Faber, R. Bro, *Chemometr. Intell. Lab.* 61 (2002) 133–149.
- [14] A. Clayton, J.W. Hines, P.D. Elkins, *Anal. Chem.* 59 (1987) 2506–2514.
- [15] K. Faber, B.R. Kowalski, *Chemometr. Intell. Lab.* 34 (1996) 283–292.
- [16] R. Boqué, M.S. Larrechi, F.X. Rius, *Chemometr. Intell. Lab.* 45 (1999) 397–408.
- [17] A.C. Olivieri, *J. Chemometr.* 16 (2002) 207–217.
- [18] J.A. Fernández, L. Jin, F. Wahl, N.M. Faber, D.L. Massart, *Chemometr. Intell. Lab.* 65 (2003) 281–291.
- [19] H. Van der Voet, *J. Chemometr.* 13 (1999) 195–208.
- [20] M.C. Ortiz, L.A. Sarabia, A. Herrero, M.S. Sanchez, M.B. Sanz, M.E. Rueda, D. Jiménez, M.E. Meléndez, *Chemometr. Intell. Lab.* 69 (2003) 21–33.
- [21] ISO 11843-2, Capability of detection, Genève, Switzerland. 2000.
- [22] A. Clayton, J.W. Hines, P.D. Elkins, *Anal. Chem.* 59 (1987) 2506–2514.
- [23] H.C. Goicoechea, A.C. Olivieri, *Chemometr. Intell. Lab.* 56 (2001) 73–81.



Grup de Quimiometria Aplicada

