



TESIS DOCTORAL

Desarrollo de nuevas metodologías espectrales para el control analítico de productos y procesos petroquímicos y farmacéuticos

Anna Palou Garcia

Directores: Prof. Marcelo Blanco Romía y Dr. Manel Alcalà Bernàrdez

PROGRAMA DE DOCTORADO DE QUÍMICA

Departamento de Química

Facultad de Ciencias

2014

Memòria presentada per aspirar al Grau de Doctor per **Anna Palou Garcia**

Anna Palou Garcia

Vist-i-plau,

Prof. Marcelo Blanco Romía

Dr. Manel Alcalà Bernàrdez

Bellaterra, 3 de Novembre de 2014

AGRADECIMIENTOS

Esta tesis doctoral se ha realizado gracias a las siguientes ayudas institucionales:

Ministerio de Ciencia e Innovación (MICINN)

Beca pre-doctoral FPI (BES-2010-030395) y beca de estancia breve (EEBB-I-12-04068)

Proyecto CTQ2009-08312 “Desarrollo de nuevas metodologías espectrales para el control analítico de productos y procesos”

Ministerio de Economía y Competitividad (MINECO)

Proyecto CTQ2012-34392 “Productos y procesos mediante técnicas espectroscópicas moleculares combinadas con tratamiento de datos multivariados”

Departament d’Innovació, Universitats i Empresa (Generalitat de Catalunya)

SGR2009-1470 i SGR2014-249

Grup de Quimiometria Aplicada (GQA)

Departament de Química, Universitat Autònoma de Barcelona (UAB)

Bellaterra, Barcelona

Esta tesis doctoral se ha realizado gracias a las siguientes ayudas privadas:

Compañía Española de Petróleos S.A.U. (CEPSA),

Refinería de San Roque, Cádiz

Esteve Farma,

Planta de Martorelles, Barcelona

VTT Technical Research Centre of Finland

Kuopio, Finland

Fer una tesi no és una feina fàcil. Com carregar (i descarregar) un castell, una tesi no s'aconsegueix únicament amb un tronc sòlid, sinó que requereix de molta altra gent perquè l'anxaneta arribi a fer l'aleta. En primer lloc, requereix d'un equip tècnic que guii els assajos i dirigeixi l'execució del castell per arribar a aconseguir els objectius establerts. En aquest equip, el Prof. Marcelo Blanco i el Dr. Manel Alcalà com a cap i sots cap de colla, juntament amb la resta de professors integrants del grup de quimiometria aplicada de la UAB (Santi, Jordi i Hortència). També dins d'aquest equip, en Jukka Rantanen, el meu tutor a Copenhaguen, així com la resta de components del grup que dirigeix, el grup de Tecnologia Farmacèutica de la Universitat de Copenhaguen (Dinamarca), on vaig fer una estança durant la tesi.

Però una tesi necessita de tota una pinya que li doni força, que li doni suport, sobretot en aquells moments més difícils o més delicats, evitant que defalleixi. Mans, laterals, vents, agulles, falques... són els meus companys de laboratori que dia a dia m'han ajudat a donar forma a aquesta tesi. Des dels companys dels meus principis al laboratori, quan ni tan sols sabia on em ficava, passant pels que m'han aguantat i m'han aconsellat dia a dia durant gran part del procés i pels més nous que algun dia no gaire llunyà també es trobaran escrivint unes línies similars a aquestes (Anna, Rubens, Ricardo, Julen, David, Juan, Sergio, Vanessa, Diego, Judit, Dong); aquells companys més efímers, però no per això menys importants (Anne, Jenny, Hjalte, Liisi, Korbi...); el Dani, el meu tipògraf personal i, sobretot, al Jordi, el millor primeres i el principal culpable d'aquesta tesi.

A la pinya també hi ha gent que ningú veu, però que estan col·locats estratègicament i són els principals pilars del castell. Es tracta de baixos, crosses i contraforts, aquells que discretament treballen de valent perquè tot surti bé i el tronc no es desestabilitzi. Els meus agraïments no són suficients per el meu germà Xavier, pels meus pares, Josep i Carme, i pel Sergi, que sempre hi són pel que faci falta.

Tots vosaltres heu fet possible aquesta tesi.

Anna

ÍNDICE

LISTA DE ACRÓNIMOS.....	1
RESUMEN.....	7
OBJETIVOS.....	17
INTRODUCCIÓN.....	25
1. INDUSTRIA PETROQUÍMICA Y FARMACÉUTICA	27
1.1. Industria petroquímica.....	29
1.1.1. Composición de los crudos.....	30
1.1.2. El proceso de refino.....	31
1.1.3. Regulación y control de calidad de productos y procesos petroquímicos	37
1.1.4. Productos refinados	38
1.2. Industria farmacéutica	45
1.2.1. Producción de comprimidos.....	46
1.2.2. Producción de liofilizados	50
1.2.3. Regulación de la industria farmacéutica.....	54
1.2.4. Control de calidad de procesos y productos farmacéuticos	55
REFERENCIAS	59
2. ESPECTROSCOPIA E IMAGEN QUÍMICA NIR.....	65
2.1. Fundamentos de la absorción NIR	67
2.2. Espectroscopia NIR.....	73
2.2.1. Modos de adquisición	76
2.2.2. Instrumentación NIR	78
2.3. Imagen química NIR (NIR-CI)	87
2.3.1. Instrumentación NIR-CI.....	89
2.4. Aplicaciones industriales del NIR y el NIR-CI	94
2.4.1. Aplicaciones NIR en la industria petroquímica.....	95
2.4.2. Aplicaciones NIR y NIR-CI en la industria farmacéutica	98
REFERENCIAS	104

3. QUIMIOMETRÍA.....	121
3.1. Introducción	123
3.1.1. Orígenes de la quimiometría	123
3.1.2. Notación.....	124
3.2. Métodos cualitativos	125
3.2.1. Análisis en Componentes Principales (PCA)	126
3.2.2. Coeficientes de correlación (CC).....	129
3.3. Métodos cuantitativos	131
3.3.1. Calibración univariante	133
3.3.2. Calibración multivariante.....	134
3.4. Etapas del proceso de modelado.....	139
3.4.1. Diseño, selección y preparación de muestras.....	140
3.4.2. Métodos de referencia.....	141
3.4.3. Obtención de la señal analítica	142
3.4.4. Preprocesamiento y extracción de datos de imágenes químicas.....	143
3.4.5. Pretratamientos espectrales.....	145
3.4.6. Selección del grupo de calibración/validación.....	153
3.4.7. Construcción del modelo de calibración	155
3.4.8. Evaluación del modelo	156
3.4.9. Elección del número de factores	158
3.4.10. Aplicación del modelo e interpretación de los resultados.....	159
REFERENCIAS	160
INDUSTRIA PETROQUÍMICA.....	171
4. MONITORIZACIÓN DE DISTINTAS PROPIEDADES FISICOQUÍMICAS DE GASÓLEOS MEDIANTE NIR.....	173
4.1. Introducción	175
4.2. Experimental.....	177
4.2.1. Condiciones experimentales del registro de espectros NIR.....	177
4.2.2. Muestras	177

4.2.3. Datos de referencia	179
4.2.4. Tratamiento de datos.....	179
4.3. Resultados y discusión.....	182
4.3.1. Selección de muestras	182
4.3.2. Construcción de los modelos PLS	190
4.3.3. Monitorización de los modelos	193
4.4. Conclusiones	196
REFERENCIAS	197

5. CARACTERIZACIÓN DE LA COMPOSICIÓN DE PARAFINAS PARA SU APLICACIÓN INDUSTRIAL.....	201
5.1. Introducción	203
5.2. Experimental.....	205
5.2.1. Muestras	205
5.2.2. GC/MS.....	206
5.2.3. NMR	206
5.2.4. Métodos de referencia.....	207
5.2.5. NIR	207
5.3. Resultados y discusión.....	213
5.3.1. GC/MS con ionización por impacto electrónico	213
5.3.2. GC/MS con ionización química.....	213
5.3.3. NMR	218
5.3.4. Análisis de parafinas	220
5.3.5. Determinación por NIR.....	221
5.4. Conclusiones	225
REFERENCIAS	226

INDUSTRIA FARMACÉUTICA	229
-------------------------------------	------------

6. DETERMINACIÓN DE LA DISTRIBUCIÓN DE API, EXCIPIENTES Y RECUBRIMIENTO EN COMPRIMIDOS FARMACÉUTICOS MEDIANTE NIR-CI	231
---	------------

6.1. Introducción	233
6.2. Experimental.....	235
6.2.1 Las muestras	235
6.2.2. Instrumentación	236
6.2.3. Análisis de datos	238
6.3 Resultados y discusión.....	241
6.3.1 Cuantificación de API y excipientes.....	242
6.3.2. Cuantificación del recubrimiento	247
6.4. Conclusiones	252
REFERENCIAS	253
7. DETECCIÓN DE CAMBIOS INDUCIDOS POR LA HUMEDAD EN FORMULACIONES LIOFILIZADAS MEDIANTE NIR-CI.....	255
7.1. Introducción	257
7.2. Experimental.....	259
7.2.1. Composición de las muestras.....	259
7.2.2. Freeze-drying.....	259
7.2.3. NIR-CI	261
7.2.4. Valoraciones Karl Fischer (KF)	262
7.2.5. Difracción de rayos X en polvo (XRPD).....	262
7.2.6. Análisis de datos	262
7.2.7. Determinación de la humedad.....	263
7.2.8. Análisis de la sacarosa cristalizada	264
7.3. Resultados y discusión.....	265
7.3.1. Visualización del contenido de agua después del freeze-drying.....	265
7.3.2. Absorción de agua en distintas condiciones de almacenamiento	266
7.3.3. Cristalización de la sacarosa.....	268
7.4. Conclusiones	272
REFERENCIAS	273
CONCLUSIONES	277

ANEXOS

- I) Robust PLS models for monitoring physico-chemical properties of diesel oils by NIR spectroscopy
- II) Characterization of paraffin waxes composition towards industrial applications
- III) Determination of drug, excipients and coating distribution in pharmaceutical tablets using NIR-CI
- IV) NIR imaging for high-throughput screening of moisture induced changes in freeze-dried formulations

LISTA DE ACRÓNIMOS

Acrónimo	Inglés	Castellano
AEMPS	----	Agencia Española de Medicamentos y Productos Sanitarios
ANN	<i>Artificial Neural Networks</i>	Redes Neuronales Artificiales
AOTF	<i>Acousto-optical Tunable Filters</i>	Filtros Optoacústicos Sintonizables
API	<i>Active Pharmaceutical Ingredient</i>	Principio Activo Farmacéutico
AS	<i>Autoscale</i>	Autoescalado
ASTM	<i>American Society for Testing and Materials</i>	Sociedad Americana para Pruebas y Materiales
B	<i>Background reference</i>	Blanco instrumental
BSA	<i>Bovine Serum Albumin</i>	Albúmina de Suero Bovino
CC	<i>Correlation Coefficient</i>	Coefficiente de Correlación
CEPSA	----	Compañía Española de Petróleos, S.A.U.
CFPP	<i>Cold Filter Plugging Point</i>	Punto de obstrucción de Filtro en Frío
CFR	<i>Code of Federal Regulations</i>	Código de Regulaciones Federales
CI	<i>Cetane Index</i>	Índice de Cetano
CLS	<i>Classical Least Squares</i>	Regresión clásica por mínimos cuadrados
CN	<i>Cetane Number</i>	Número de Cetano
Cov	<i>Covariance</i>	Covarianza
CP	<i>Cloud Point</i>	Punto de Nube
CV	<i>Cross Validation</i>	Validación Cruzada
CWS	<i>Calibration WorkShop Software</i>	----
D	<i>Dark reference</i>	Referencia oscura
D 95%	<i>95% distillation temperature</i>	Temperatura de destilación del 95%
Dens.	<i>Density at 15 °C</i>	Densidad a 15 °C
EMA	<i>European Medicine Agency</i>	Agencia Europea del Medicamento
EMSC	<i>Extended MSC</i>	Variante Extendida del MSC
EN	<i>European Standard</i>	Norma Europea
FAME	<i>Fatty Acid Methyl Ester</i>	Éster Metílico de Ácido Graso

FCC	<i>Fluid Catalytic Cracking</i>	Craqueo Catalítico en Lecho Fluidizado
FDA	<i>Food and Drug Administration</i>	----
FIR	<i>Far-InfraRed</i>	Infrarrojo Lejano
FP	<i>Flash Point</i>	Punto de Inflamabilidad
FPA	<i>Focal Plane Array</i>	Matriz de Plano Focal
FT	<i>Fourier transform</i>	Transformada de Fourier
FT-NIR	<i>Fourier Transform Near-Infrared instrument</i>	Equipo de Infrarrojo Cercano por Transformada de Fourier
GC	<i>Gas Chromatography</i>	Cromatografía de Gases
GC/MS	<i>Gas Chromatography/Mass Spectrometry</i>	Cromatografía de Gases/Espectrometría de Masas
GMP	<i>Good Manufacturing Practices</i>	Buenas Prácticas de Fabricación
HPLC	<i>High-Performance Liquid Chromatography</i>	Cromatografía Líquida de Alta Eficacia
ICH	<i>International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use</i>	Conferencia Internacional sobre Armonización de los Requisitos Técnicos para el Registro de Productos Farmacéuticos para Uso Humano
ICS	<i>International Chemometrics Society</i>	Sociedad Internacional de Quimiometría
ILS	<i>Inverse Least Squares</i>	Regresión Inversa por Mínimos Cuadrados
IR	<i>InfraRed</i>	Infrarrojo
ISO	<i>International Standardization Organization</i>	Organización Internacional de Normalización
KF	<i>Karl Fischer</i>	Karl Fischer
KNN	<i>K-Nearest Neighbor</i>	K-Vecino más Cercano
LCTF	<i>Liquid Crystal Tunable Filter</i>	Filtros Sintonizables de Cristal Líquido
LDA	<i>Linear Discriminant Analysis</i>	Análisis Discriminante Lineal
LED	<i>Light-Emitting Diodes</i>	Diodos Emisores de Radiación
LPG	<i>Liquefied Petroleum Gas</i>	Gas Licuado del Petróleo
LWR	<i>Locally Weighted Regression</i>	Regresión Local Ponderada
MC	<i>Mean Centering</i>	Centrado a la Media
MCR-ALS	<i>Multivariate Curve Resolution-</i>	Resolución multivariante de curvas y

	<i>Alternating Least Squares</i>	mínimos cuadrados alternos
MCT	<i>Mercury-Cadmium-Telluride</i>	Mercurio-Cadmio-Telurio
MEK	<i>Methyl Ethyl Ketone</i>	Metil Etil Cetona
MIR	<i>Mid-InfraRed</i>	Infrarrojo Medio
MLR	<i>Multiple Linear Regression</i>	Regresión Lineal Múltiple
MSC	<i>Multiplicative Scatter Correction</i>	Corrección del Efecto Multiplicativo de Dispersión
MW	<i>Molecular Weight</i>	Peso Molecular
NaN	<i>Not a Number</i>	No es un Número
NIR	<i>Near-InfraRed</i>	Infrarrojo Cercano
NIR-CI	<i>Near-InfraRed Chemical Imaging</i>	Imagen Química en el Infrarrojo Cercano
NIRS	<i>Near InfraRed Spectroscopy</i>	Espectroscopia del Infrarrojo Cercano
NMR	<i>Nuclear Magnetic Resonance</i>	Resonancia Magnética Nuclear
OSC	<i>Orthogonal Signal Correction</i>	Corrección de la Señal Ortogonal
PA	<i>Process Analysis</i>	Análisis de Proceso
PAC	<i>Process Analytical Chemistry</i>	Química Analítica de Proceso
PAT	<i>Process analytical Technology</i>	Tecnología Analítica de Proceso
PC	<i>Principal Component</i>	Componente principal
PCA	<i>Principal Component Analysis</i>	Análisis en Componentes Principales
PCR	<i>Principal Component Regression</i>	Regresión de Componentes Principales
PDA	<i>Photodiode Array</i>	Matriz de Fotodiodos
Ph. Eur.	<i>European Pharmacopoeia</i>	Farmacopea Europea
PLS	<i>Partial Least Squares</i>	Regresión Parcial por Mínimos Cuadrados
PLS-DA	<i>PLS-Discriminant Analysis</i>	Análisis PLS discriminante
PRESS	<i>Predicted REsidual Sum of Squares</i>	Suma de los Cuadrados de los Residuales Predichos
PRM	<i>Pattern Recognition Methods</i>	Métodos de Reconocimiento de Pautas
QDA	<i>Quadratic Discriminant Analysis</i>	Análisis Discriminante Cuadrático
QNP	<i>Quadruple Nucleus Probe</i>	Sonda de Núcleo Cuadrupolar
R	<i>Reflectance data matrix</i>	Matriz de datos en reflectancia
RGB	<i>Red Green Blue</i>	Rojo Verde Azul

RH	<i>Relative Humidity</i>	Humedad Relativa
RI	<i>Refractive Index</i>	Índice de Refracción
RMSE	<i>Root Mean Square Error</i>	Error Cuadrático Medio
RMSEC	<i>Root Mean Square Error of Calibration</i>	Error Cuadrático Medio de Calibración
RMSECV	<i>Root Mean Square Error of Cross Validation</i>	Error Cuadrático Medio de Validación Cruzada
RMSEP	<i>Root Mean Square Error of Prediction</i>	Error Cuadrático Medio de Predicción
RSE	<i>Relative Standard Error</i>	Error Estándar Relativo
RSEC	<i>Relative Standard Error of Calibration</i>	Error Estándar Relativo de Calibración
RSECV	<i>Relative Standard Error of Cross Validation</i>	Error Estándar Relativo de Validación Cruzada
RSEP	<i>Relative Standard Error of Prediction</i>	Error Estándar Relativo de Predicción
S	<i>Sample</i>	Muestra
SC	<i>Sulphur Content</i>	Contenido en Azufre
SEM	<i>Scanning Electron Microscopy</i>	Microscopía Electrónica de Barrido
SIM	<i>Single Ion Monitoring</i>	Seguimiento Individual de Iones
SNV	<i>Standard Normal Variate</i>	Variable Normal Estándar
SVM	<i>Supported Vector Machines</i>	Máquinas de Soporte Vectorial
TBP	<i>Total Boiling Point</i>	Punto de Ebullición Total
USDA	<i>United States Department of Agriculture</i>	Departamento de Agricultura de los Estados Unidos
USP	<i>United States Pharmacopeia</i>	Farmacopea de los Estados Unidos
UV	<i>Ultraviolet</i>	Ultravioleta
UV-Vis	<i>Ultraviolet-Visible</i>	Ultravioleta-Visible
VHM	<i>Vial Heater Module</i>	Módulo Calentador de Viales
WAT	<i>Wax Appearance Temperature</i>	Apariencia de la Cera en función de la Temperatura
WT	<i>Wavelet Transform</i>	Transformación por ondas
XRPD	<i>X-Ray Powder Diffractometry</i>	Difracción de Rayos X en Polvo

RESUMEN

ABSTRACT

RESUMEN

La química analítica tiene el reto de aportar información de interés de una manera más amplia y completa sobre los productos de interés, para asegurar la calidad exigida por consumidores y organismos reguladores en el menor tiempo posible. En esta memoria se proponen nuevas soluciones analíticas basadas en la espectroscopia del infrarrojo cercano (NIR) y la imagen química NIR (NIR-CI) para la resolución de problemas reales de la industria farmacéutica y petroquímica. La rapidez y el carácter no destructivo y no invasivo de estas técnicas son las características determinantes para su aplicación en el control de calidad industrial, mientras que su gran versatilidad es la clave para su adaptación a las características y necesidades de cada empresa.

En la industria petroquímica, se desarrollan estrategias capaces de reducir tiempo y costes de los análisis, mejorando tanto la exactitud como la reproducibilidad de los resultados y la representatividad de las muestras analizadas. La espectroscopia NIR constituye una alternativa ventajosa respecto a los distintos métodos analíticos de referencia establecidos para abordar la gran variedad de análisis que requieren los productos petroquímicos para asegurar su calidad final. En el contexto de la industria petroquímica se han realizado dos estudios orientados a la determinación de parámetros de interés de gasóleos y parafinas respectivamente:

- 1) Se han desarrollado modelos cuantitativos para la monitorización en planta (*on-line*) de siete parámetros fisicoquímicos de gasóleos. La construcción de los distintos modelos de regresión por mínimos cuadrados parciales (PLS) se lleva a cabo utilizando únicamente espectros NIR de muestras *post-blending* registrados en planta. La dificultad que conlleva la construcción de estos modelos es la gran variabilidad intrínseca de las muestras de gasóleos, debida principalmente al origen y composición de los crudos de los que proceden y a las distintas corrientes de fabricación que se mezclan durante el proceso. Esta variabilidad se refleja directamente en el espectro NIR de las muestras dificultando la construcción de modelos robustos y con capacidad predictiva a largo plazo. Con el objetivo de recoger toda la variabilidad natural de las muestras se

seleccionaron grupos de calibración específicos para cada parámetro a determinar siguiendo tres criterios: rango de valor de referencia, variabilidad y naturaleza de los espectros y origen de los crudos. Los modelos PLS desarrollados presentaron buena capacidad predictiva en todos los casos y la mantienen en las predicciones de control hasta la fecha, confirmando su robustez.

- 2) Se ha determinado el contenido de aceite (isoalcanos) en parafinas industriales (mayoritariamente *n*-parafinas). El estudio se inicia con la caracterización de los componentes principales de las parafinas (parafina desaceitada y aceite) por cromatografía de gases acoplada a espectrometría de masas (GC/MS) y resonancia magnética nuclear (NMR). Los resultados obtenidos facilitan el desarrollo de dos nuevos métodos para la determinación del contenido de aceite en parafina: el primero permite determinar el contenido total de isoparafinas mediante GC, mientras que el segundo se basa en la espectroscopia NIR para la cuantificación del contenido de aceite extraíble por metil etil cetona (MEK) según la norma ASTM D721. En este último caso se requiere el pretratamiento de los espectros y la selección de datos adecuados, que permitan la construcción de modelos PLS capaces de distinguir y cuantificar el aceite contenido en la parafina a pesar de la elevada similitud química y espectral que existe entre ellos. Las predicciones PLS son concordantes con las del método oficial (ASTM D721), por lo que el método NIR representa una alternativa real al método ASTM, con la ventaja de ser mucho más rápido y reproducible.

En la industria farmacéutica, se proponen nuevos métodos analíticos capaces de dar una visión rápida y general del estado real de las muestras a lo largo de un proceso. Para ello, se ha utilizado el NIR-CI que permite extraer, en muy poco tiempo, una gran cantidad de información y presentarla de una forma sencilla y visual. Los modelos desarrollados evidencian los efectos de los tratamientos y las condiciones a los que han sido sometidas las muestras durante el proceso de estudio. Esto permite el ajuste del diseño de nuevos fármacos y su proceso de producción,

pero también posibilita el control del futuro proceso de producción, asegurando la calidad del producto. Estos conceptos se incluyen en lo que se conoce como tecnología analítica de procesos (PAT).

- 3) Se han desarrollado nuevos métodos para la determinación de concentración y distribución del contenido de principio activo (API) y excipientes en comprimidos farmacéuticos, así como del grosor y distribución de su recubrimiento mediante NIR-CI. La extracción de los datos de interés de las imágenes y la obtención de información tanto cualitativa como cuantitativa se ha conseguido con la aplicación de los algoritmos PLS e Isys-PLS1. Se han desarrollado modelos Isys-PLS1 para determinar la concentración y distribución de los distintos componentes, tanto en las caras interiores como exteriores del comprimido (antes y después del recubrimiento). Las diferencias entre resultados en función del grosor de recubrimiento exigen su determinación, para lo que se construyeron modelos PLS. En ambos casos, la proximidad de los resultados obtenidos con los valores de referencia confirma la utilidad del NIR-CI en este tipo de análisis.

- 4) Se ha llevado a cabo el cribado rápido y a gran escala de distintas muestras liofilizadas compuestas por sacarosa, manitol y proteína en distintas proporciones. El objetivo es la identificación de aquellas formulaciones más apropiadas para ser liofilizadas según su absorción/desorción de agua y otros fenómenos relacionados en distintas condiciones ambientales. Después de ser liofilizadas, las muestras fueron almacenadas a dos niveles de humedad relativa y registradas por NIR-CI diariamente durante diez días. El seguimiento de la humedad se consigue mediante un modelo univariante. Además, con la aplicación de coeficientes de correlación a las imágenes, se confirma la idoneidad del NIR-CI para la localización de transformaciones del estado sólido, como la cristalización de la sacarosa que, igual que la humedad, también afecta negativamente a la estabilidad y la actividad de la proteína.

Las metodologías desarrolladas constituyen alternativas reales a métodos utilizados convencionalmente, mejorando la rapidez y los costes de análisis y confirmandolas posibilidades del control de procesos en tiempo real, sin disminuir la calidad de los resultados analíticos. Los resultados obtenidos mediante la aplicación de estas técnicas, son indicativos del potencial del NIR y NIR-CI en el control de calidad industrial.

ABSTRACT

The challenge of analytical chemistry is to provide information of interest in a wider and more comprehensive way about the products of interest. It must allow ensuring the quality demanded by consumers and regulatory organisations in the shortest time possible. In this thesis new analytical solutions based on near-infrared spectroscopy (NIR) and NIR chemical imaging (NIR-CI) for solving real problems in the pharmaceutical and petrochemical industry are proposed. The expeditiousness and non-destructive and non-invasive character of these techniques are decisive in their application to the quality control of industrial products, while their versatility is crucial to adapt them to the features and needs of each company.

In the petrochemical industry field, strategies capable of reducing time and costs of analysis, improving both accuracy and reproducibility of the results and the representativeness of the analysed samples have been developed. NIR spectroscopy is an advantageous alternative regarding the analytical reference methods established to address the wide variety of analysis that petrochemicals require to ensure the final quality. In the context of the petrochemical industry two studies focused in determining parameters of interest in diesel and paraffin have been carried out:

- 1) Quantitative models for monitoring on-line seven physicochemical parameters of gas oils have been developed. Only NIR spectra of post-blending plant samples have been used for the construction of partial least squares (PLS) regression models. The high variability of diesel samples is the main difficulty in the models development. It is mainly due to their crude oils origin and composition and the mixed streams during the manufacturing processes. All this variability is directly reflected in the NIR spectra of the samples, complicating the achievement of robust and long-term predictive models. In order to collect all the natural variability of the samples, specific calibration groups for each parameter to be determined were selected. It was performed according to three criteria: reference value range, spectral nature and variability

and crude oils origin. The developed PLS models showed good predictive ability in all cases and still maintain it, confirming their robustness.

- 2) The oil (isoalkanes) content in industrial paraffins (mainly *n*-alkanes) has been determined. Firstly, the major components of paraffins (deoiled paraffin and oil) were characterized by gas chromatography-mass spectrometry (GC/MS) and nuclear magnetic resonance (NMR). The obtained information facilitates the development of two new methods for the determination of paraffin oil content. The first method determines the total isoparaffins content by GC, while the second allows the quantification of the methyl ethyl ketone (MEK) removable oil (according to ASTM D721 standard) by NIR spectroscopy. In the latter case, spectral pretreatment and selection of appropriate data were needed to construct the model in order to face the high chemical and spectral similarity between oil and paraffin. PLS results are consistent with those of the official method (ASTM D721), positioning NIR as a real alternative with the advantage of being much faster and reproducible.

In the pharmaceutical industry, new analytical methods able to give a quick overview of the actual state of the samples along the process are proposed. NIR-CI has been used to quickly extract a lot of information of the sample and present it in a simple and visual way. The sample treatments and conditions during the study are reflected in the developed models. It allows adjusting the design of new drugs and their production process and also enables the control of the future production processes ensuring product quality. All these possibilities are comprised in the Process Analytical Technology (PAT) concept.

- 3) New methods for determining the concentration and distribution of active pharmaceutical ingredient (API) and excipients as well as the thickness and distribution of the coating have been developed by using NIR-CI. The extraction of useful data from images has been achieved with the application of PLS and Isys-PLS1 algorithms, which allow to obtain both qualitative and quantitative information from the sample. Isys-PLS1 models were developed to determine the

concentration and distribution of each component in inner and outer tablet faces (before and after coating). The differences observed in the results depending on the tablets coating thickness, demands for its determination. Consequently, PLS models were built for coating thickness determination. The proximity of the results obtained to the reference values confirms the usefulness of NIR-CI in this type of analysis.

- 4) A high throughput screening of samples consisting of lyophilized sucrose, mannitol and protein in varying proportions have been performed. The aim of this study was to identify the most suitable formulations for lyophilisation, according to their water absorption/desorption and related phenomena in different environmental conditions. After freeze-drying, samples were stored at two levels of relative humidity NIR images recorded daily during ten days. Moisture monitoring was achieved by using an univariate model. Furthermore, the application of correlation coefficients to the images, confirmed the suitability of NIR-CI for locating solid state transformations, such as sucrose crystallization which negatively affects protein stability and activity.

The developed methodologies are demonstrated to be real alternatives to conventionally used methods. They improve the speed and cost of analysis as well as the possibilities for process control in real time, without decreasing quality of analytical results. The results obtained by using these techniques, joined to their advantages are indicative of the potential of NIR and NIR-CI in industrial quality control.

OBJETIVOS

OBJECTIVES

OBJETIVOS

Garantizar la calidad de los productos elaborados es un reto primordial para cualquier industria, sea cual sea su sector. Para ello, cada producto debe cumplir con unas normativas cada vez más estrictas, en cuanto a número de parámetros y rangos de especificación, antes de poder ser comercializado. Esto implica un mayor número de análisis y su aplicación en distintos puntos del proceso, lo que para una técnica tradicional se traduce en un aumento considerable en el tiempo de análisis, número de analistas necesario, consumo de disolventes, destrucción de muestras, generación de residuos, entre otros. En estas condiciones, el cumplimiento de los requisitos de calidad es cada vez más difícil de compaginar con el carácter competitivo de las empresas, haciéndose evidente la necesidad de encontrar nuevas metodologías analíticas rápidas y versátiles, que no requieran preparación, ni destrucción de las muestras. Técnicas como la espectroscopia del infrarrojo cercano (NIRS) o la imagen química NIR (NIR-CI), son la respuesta a estas necesidades.

El objetivo general de esta tesis es el desarrollo de nuevos métodos analíticos basados en NIRS y NIR-CI, para la resolución de problemas reales de la industria farmacéutica y petroquímica. Se trata de aprovechar la versatilidad de estas técnicas para adaptar el análisis a cada situación (tipo de muestra, condiciones de registro, objetivo del análisis) y combinarlas con los algoritmos quimiométricos adecuados para extraer la información de interés en cada caso y poderla interpretar. Todo este proceso de desarrollo de modelos de calibración para la determinación de parámetros químicos y físicos, ha conllevado la definición de otros objetivos más específicos en cada caso de estudio:

- Definición de estrategias para la obtención de modelos de calibración NIR robustos y efectivos a largo plazo cuando se trabaja con muestras con gran variabilidad.
- Caracterización de productos y desarrollo de modelos NIR capaces de cuantificar ciertos componentes a pesar de presentar elevada similitud química y espectral entre ellos.

- Determinación de la distribución de componentes (en forma de concentración o grosor) a lo largo de la superficie de una muestra mediante imágenes NIR-CI y evaluación del efecto de un recubrimiento en su predicción.
- Seguimiento de un gran número de muestras simultáneamente mediante NIR-CI y cribado en función de su comportamiento en distintas condiciones ambientales.

OBJECTIVES

Ensuring the quality of manufactured products is a major challenge for industries. Increasingly stringent regulations in terms of number of parameters and specification ranges must be complied for each product before marketing. It supposes a great number of analyses per product and its application in different points of the production process. For a traditional technique, it results in a considerable increase in the analysis time, number of analysts, solvent consumption, samples destruction, wastes, among others. In these conditions, the fulfillment of quality requirements is increasingly difficult to reconcile with the competitive nature of business and the need for new fast and versatile analytical methodologies that avoid sample preparation and destruction became evident. Techniques such as near-infrared spectroscopy (NIRS) or NIR chemical imaging (NIR-CI) are the answer to these needs.

The general aim of this thesis is the development of new analytical methods based on NIRS and NIR-CI to solve real problems of pharmaceutical and petrochemical industries. The versatility of both techniques is used to adapt any analysis to the situation (sample type, recording conditions, objective). The obtained data is combined with appropriate chemometric algorithms to extract the information of interest in each case and to facilitate their interpretation. The whole process of development of calibration models for the determination of chemical and physical parameters suppose the definition of more specific objectives in each case of study:

- Definition of strategies for obtaining robust NIR calibration models with long-term effectiveness when working with large variability samples.
- Characterization of products and development of NIR models capable of quantifying components with high chemical and spectral similarity.
- Determination of the distribution of components (as a concentration or thickness) along the surface of a sample by NIR-CI images and evaluation of the effect of the coating on predictions.

- High throughput samples monitoring and screening according to their behavior in different environmental conditions by using NIR-CI.

INTRODUCCIÓN

1.INDUSTRIA PETROQUÍMICA Y FARMACÉUTICA

1.1. Industria petroquímica.....	29
1.1.1. Composición de los crudos.....	30
1.1.2. El proceso de refino.....	31
1.1.3. Regulación y control de calidad de productos y procesos petroquímicos	37
1.1.4. Productos refinados	38
1.2. Industria farmacéutica	45
1.2.1. Producción de comprimidos.....	46
1.2.2. Producción de liofilizados	50
1.2.3. Regulación de la industria farmacéutica.....	54
1.2.4. Control de calidad de procesos y productos farmacéuticos	55
REFERENCIAS	59

1. INDUSTRIA PETROQUÍMICA Y FARMACÉUTICA

Las industrias petroquímica y farmacéutica son los sectores de trabajo de esta tesis. Se trata de dos sectores muy distintos con problemáticas analíticas específicas, que confluyen en un objetivo común: la necesidad de encontrar nuevos métodos para poder asegurar los estándares de calidad sin afectar su competitividad.

1.1. Industria petroquímica

La industria petroquímica es aquella dedicada a obtener derivados químicos del petróleo y los gases asociados. El petróleo es una de las sustancias más consumidas en las sociedades modernas. Esto es debido a que, además de proporcionar materias primas para plásticos, gomas, cauchos, ceras, fibras sintéticas y otros muchos productos, los fueles derivados del petróleo suplen casi la mitad de las necesidades energéticas mundiales [1].

La palabra petróleo proviene de los términos latinos *petra* y *oleum*, lo que literalmente significa ‘aceite de piedra’. Esta definición se refiere a la compleja mezcla de hidrocarburos saturados en estado sólido, líquido o gaseoso que se encuentran en yacimientos naturales [2,1]. El petróleo tal como se ha descrito se conoce como ‘crudo’ y no tiene ninguna utilidad directa, sino que su valor aparece después de ser sometido al proceso industrial del refinado [3,4]. Durante este proceso, el crudo se somete a distintas destilaciones y reacciones químicas. Todas estas operaciones se llevan a cabo en las refinerías, instalaciones enormes con decenas de kilómetros de cañerías, diseñadas para el trabajo en continuo. Estas instalaciones suelen estar dotadas de puntos de control, que permiten la monitorización de distintos parámetros fisicoquímicos, aportando información en continuo, para el ajuste del proceso productivo. Esto permite que, a pesar de tratarse de un proceso en continuo, su curso pueda ser modificado en función de las transformaciones químicas que requiera el crudo original para generar fracciones y productos de interés comercial. Una característica muy conveniente teniendo en cuenta la gran variabilidad de crudos, en

cuanto origen, composición y características, que pueden entrar como materia prima en una misma refinería.

En este apartado, se detalla en primer lugar la composición de los crudos del petróleo (apartado 1.1.1), la base de la industria petroquímica. Seguidamente, se proporciona una descripción de las etapas generales del refino (apartado 1.1.2) y los controles de calidad y regulaciones a los que se someten los productos petroquímicos antes, durante y después de este proceso (apartado 1.1.3). Al final de este apartado (apartado 1.1.4) se aborda la descripción de los dos productos refinados utilizados en la parte experimental de esta tesis: los diésels y las parafinas. Se detallan sus características y aplicaciones, así como el marco de regulaciones y los análisis a los que deben someterse para ser comercializados.

1.1.1. Composición de los crudos

Los crudos están constituidos mayoritariamente (>95%) por una compleja mezcla de hidrocarburos, desde el metano (C1), hasta especies complejas (C50 o más), no destilables sin previa descomposición [4]. Por lo que respecta a la estructura de los hidrocarburos se puede distinguir entre parafinas lineales saturadas (parafinas) o insaturadas (olefinas), ramificadas (isoparafinas), cicladas (cicloparafinas o naftenos), aromáticas, así como estructuras mucho más complejas [4]. Otros componentes minoritarios pueden ser compuestos de azufre, oxígeno y nitrógeno, así como distintos elementos metálicos en cantidades traza [1,2,4,5]. Los compuestos de azufre (SH₂, tiofenos, mercaptanos, disulfúros...) son los más significativos, no sólo por su cantidad (0.5-5%), sino también por su corrosividad, olor y otras propiedades indeseables que obligan a su eliminación mediante procesos de desulfuración durante el refino [2].

La composición de cada crudo es única, ya que las proporciones relativas, tanto de los distintos tipos de hidrocarburos como de las impurezas presentes, varían según el origen del yacimiento y la profundidad de la extracción [5]. Esto afecta a las propiedades del crudo, pero también a su valor, al proceso de refino posterior y a los

productos finales [1,2]. Las concentraciones de los diferentes tipos de hidrocarburos, su proporción de carbono (C/H) y la cantidad de heteroelementos en un crudo, determinan el rendimiento y la calidad de los productos refinados finales y, por tanto, el valor económico del crudo [4]. En algunos casos, para conseguir las cantidades requeridas y la calidad necesaria de los productos finales, se recurre a la mezcla de distintos crudos [6]. Dependiendo de la composición del crudo/mezcla de crudos a tratar, las cantidades de las distintas fracciones obtenidas en las destilaciones varían [2] y, en función de ello y de los productos finales requeridos, se definen los tratamientos de refino posteriores [3].

1.1.2. El proceso de refino

El término refino se aplica al conjunto de procesos llevados a cabo en las refinerías y que presenta distintas etapas: el fraccionamiento, el craqueo, el reformado y el refino final. Los productos obtenidos del refino son, en muchos casos, sometidos también al mezclado o *blending* antes de su liberación.

▪ **Fraccionamiento del crudo**

El primer paso en el proceso de refino es la destilación fraccionada, que permite separar los hidrocarburos que componen el crudo de partida en distintas fracciones (o cortes de destilación), en función de su punto de ebullición (T_{eb}). Cada fracción está compuesta por multitud de hidrocarburos con punto de ebullición dentro del mismo rango [4]. A mayor punto de ebullición, más pesadas son las fracciones, lo que implica hidrocarburos con mayor número de carbono, peso molecular (MW, *Molecular Weight*) y viscosidad.

El crudo de partida se calienta y se utiliza para alimentar una primera columna de destilación, donde se lleva a cabo la **destilación atmosférica**. Como su nombre indica, esta primera destilación se realiza a presión atmosférica y elevadas temperaturas, sin sobrepasar los 370-380°C para evitar llegar a condiciones de craqueo [3,6], *i. e.* ruptura de los componentes de alto MW. Las distintas fracciones obtenidas

INTRODUCCIÓN

en esta destilación por orden de volatilidad son: la fracción de cabeza (fracción ligera), que incluye gas y gasolina, las fracciones medias, formadas por queroseno y gasóleo, y el residuo o cola de la columna, que contiene las fracciones más pesadas. Las propiedades y principales aplicaciones de cada fracción se resumen en la Tabla 1.1. Como se ha comentado anteriormente, las proporciones obtenidas de las distintas fracciones varían de forma significativa en función de la composición del crudo de partida.

Las fracciones pesadas obtenidas como residuo atmosférico son destiladas de nuevo en un sistema análogo al anterior, pero trabajando a bajas presiones (50-100 mmHg) [3]. Este segundo fraccionamiento se conoce como **destilación al vacío** y permite la separación de los componentes menos volátiles del petróleo a temperaturas inferiores a su punto ebullición ($<400^{\circ}\text{C}$), evitando así la descomposición de parte del residuo. Como productos de esta segunda destilación, se obtiene de nuevo gasóleo (fracción ligera), además de lubricantes (fracciones centrales) y el residuo, que normalmente se emplea como combustible o asfalto [2].

Tabla 1.1. Propiedades y principales aplicaciones de las fracciones obtenidas en el destilado atmosférico [2,6].

Fracción	Producto	Intervalo de T_{eb} ($^{\circ}\text{C}$)	Nº de carbonos	Aplicaciones
Fracciones ligeras	Gases de refinería	<20	C1-C4	Combustible para refinería, calefacción, Gas Licuado del Petróleo (LPG)
	Gasolina	40-150	C5-C9	Carburante automobiles
	Nafta	150-200	C10-C12	Materia prima para prod. químicos, disolventes
Fracciones medias	Queroseno	170-250	C13-C17	Carburante para turborreactores, lámparas de alumbrado,
	Gasóleo	250-320	C17-C20	Carburante motores Diesel Calefacción doméstica

Fracciones pesadas (residuo atmosférico)	Fuel-oil ligero	340-400		Combustible marinos, locomotoras
	Fuel-oil pesado	400-500	C20-C35	Materia prima para lubricantes, ceras y parafinas, cremas, aceites esenciales
	Asfaltos	> 500	> C35	Pavimentación, techado, impermeabilización Abrasivos, electrodos

▪ Conversión por craqueo

El objetivo del craqueo es la conversión de las fracciones de menor demanda en el mercado en otras de mayor demanda, es decir, que las fracciones más pesadas se convierten en fracciones más ligeras, como gasolinas y gasóleos, que son, en general, más valuosas. Para ello, se requiere la ruptura de los componentes de alto MW, que se puede llevar a cabo de forma térmica o catalítica. En el **craqueo térmico**, la descomposición de los componentes de alto MW se lleva a cabo elevando temperatura y presión, en función de los componentes a craquear [3]. Aunque este método ha sido mayoritariamente reemplazado por el craqueo catalítico, aun conserva ciertas aplicaciones, como el *visbreaking* y la coquización. El *visbreaking* tiene como objetivo la reducción de la viscosidad de los residuos pesados de las destilaciones. Dado que en este caso las reacciones térmicas no se llegan a completar, las condiciones de trabajo son más suaves que en el craqueo térmico [2,3]. Para la coquización, en cambio, se aplican unas condiciones más severas para convertir los componentes no volátiles en productos destilables [2,3].

En el caso del **craqueo catalítico**, la descomposición se lleva a cabo mediante calentamiento y reacción mediada por catalizadores, mayoritariamente aluminosilicatos cristalinos (zeolitas) o tamices moleculares en lecho fluidizado (FCC, *Fluid Catalytic Cracking*) [3]. Las principales ventajas del craqueo catalítico respecto al térmico es la mejora el rendimiento de la conversión y el aumento de la calidad de las gasolinas al producir mayor ramificación de las cadenas (mayor octanaje) [2,3].

Además, el perfeccionamiento de este sistema permite trabajar a temperatura y presión menores [2].

Otra alternativa es el **hidrocraqueo**, un proceso de craqueo catalítico con incorporación simultánea de hidrógeno a presión [2,3]. Esta reacción da lugar a hidrocarburos de menor tamaño y saturados, lo que es de gran utilidad cuando se requiere generación de gasóleo [6] y aceites lubricantes [7]. Además, permite eliminar parte del azufre u otros heteroátomos no deseados [3].

▪ **Reformado**

Los procesos de reformado son en muchos casos similares a los de craqueo, ya que se llevan a cabo mediante la aplicación de calor y presión y, en algunos casos, en presencia de un catalizador. La diferencia entre ambos procesos radica en el objetivo: mientras el proceso de craqueo incrementa la cantidad de gasolina y gasóleo obtenida en la destilación, el reformado se ocupa de aumentar la calidad de la gasolina (mejora del octanaje), modificando la estructura de las cadenas carbonadas [2,3]. De nuevo, el proceso catalítico es la opción predominante, dado que es más efectivo y económico que el térmico y los carburantes obtenidos son de mayor calidad [3].

Algunas reacciones llevadas a cabo en los procesos de reformado son: alquilaciones, isomerizaciones, polimerizaciones, esterificaciones o deshidrogenaciones, todas ellas destinadas a aumentar la relación C/H de los hidrocarburos y/o el número de ramificaciones.

▪ **Refino final**

El refino final consiste en la aplicación de distintos tratamientos de depuración, con el objetivo de eliminar aquellos compuestos indeseables, principalmente compuestos sulfurados. Los procesos utilizados para la desulfuración dependen del tipo y la

cantidad de compuestos de azufre presentes y de las especificaciones del producto final.

El proceso más común es el **hidrotratamiento**, un proceso de hidrogenación catalítica a presión, parecido al hidrocrackeo, pero llevado a cabo en condiciones más suaves. En este caso, no se modifica la longitud de las cadenas carbonadas, sino que el hidrógeno sustituye los heteroátomos (azufre, oxígeno y nitrógeno) presentes en los hidrocarburos y se combina con ellos dando lugar a ácido sulfhídrico (H_2S), amoníaco (NH_3) y agua [3,4]. Este proceso también implica la saturación de compuestos aromáticos y olefinas [7].

Otro tipo de tratamiento es el **refino químico**, que consiste en aplicar una solución ácida (mayoritariamente ácido sulfúrico) o básica (NaOH) sobre la fracción a purificar. Con el cambio de medio, se desencadenan ciertas reacciones químicas que implican la disociación de las impurezas de los hidrocarburos para formar nuevos compuestos, que precipitan o se disuelven en el medio [2,3].

Finalmente, se describen los métodos de **refino físico**, que se aplican principalmente a fracciones pesadas, como aceites y residuos de la destilación atmosférica, con objetivos distintos. Entre los métodos de refino físico se distinguen [2,3,7]:

- **Desasfaltado:** tiene como objetivo separar los materiales asfálticos de alto MW de los aceites lubricantes. Se realiza por adición de un disolvente de la fracción hidrocarbonada (normalmente propano líquido), mientras que los asfaltos precipitan. Los aceites procedentes del desasfaltado se conocen como '*bright stock*'.
- **Extracción con disolventes:** se trata de una eliminación selectiva de hidrocarburos aromáticos, que afectan negativamente las propiedades reológicas de los aceites lubricantes. El disolvente, mayoritariamente furfural, disuelve preferentemente las especies de carácter aromático y nafténico, dejando insolubles las de carácter parafínico, que aportan mejores propiedades lubricantes.

- **Desparafinado:** consiste en separar las fracciones de parafina contenidas en los aceites, asegurando así la fluidez y los bajos puntos de congelación ($\leq -10^{\circ}\text{C}$) exigidos a los lubricantes. El proceso de desparafinado también suele llevarse a cabo mediante extracción por disolvente. El aceite se disuelve en el disolvente y la mezcla se enfría para que la cera cristalice y precipite. El disolvente más comúnmente utilizado es una mezcla de tolueno, que disuelve aceite y parafina y reduce su viscosidad, y metil etil cetona (MEK, *Methyl Ethyl Ketone*), que actúa como agente precipitante de la parafina. La separación de las dos fases formadas se lleva a cabo mediante filtros de rotación al vacío.
- **Decoloración:** se emplea como paso de acabado en aceites lubricantes y ceras, ya que permite eliminar materiales asfálticos y otros componentes reactivos, como olefinas y compuestos de azufre que dan colores y olores no deseados. Aunque en algunos casos se utiliza la percolación con bauxita, el tratamiento de decoloración más habitual se lleva a cabo con tierras decolorantes. En este último caso, las parafinas se hacen pasar, por percolación o en discontinuo, a través de estas tierras, que adsorben las impurezas responsables de su color. La separación de las parafinas se consigue mediante la subsiguiente filtración y centrifugación.

▪ Mezclado

La mayoría de los flujos provenientes de la refinación suelen mezclarse en distintas proporciones, con el objetivo de conseguir que los productos finales cumplan las especificaciones industriales y gubernamentales necesarias para poder ser comercializados, con un coste mínimo. En general, la gasolina es una mezcla de entre 6 y 10 componentes, mientras que el combustible diésel contiene entre 4 y 6 [4]. Las mezclas incluyen flujos con propiedades distintas debido al origen del crudo y, en especial, a la etapa de refino de la que provienen. En algunos casos, también se añaden distintos productos a la mezcla para mejorar o añadir ciertas propiedades a los gasóleos, como puede ser la adición de queroseno al diésel para mejorar sus

propiedades de fluidez a bajas temperaturas. Otro ejemplo muy común es la adición de cierta cantidad de biodiéselos a los diéselos. Los biodiéselos están constituidos por ésteres metílicos de ácidos grasos (FAME, *Fatty Acid Methyl Ester*) aportando carácter renovable y biodegradable al combustible, sin necesidad de cambiar los motores.

1.1.3. Regulación y control de calidad de productos y procesos petroquímicos

El control de calidad tiene una gran importancia en la industria petroquímica. La calidad del producto final (p. ej. fueles compatibles con el sistema que los consume y que produzcan la máxima energía en el proceso de combustión), la eficiencia de los procesos de refino o la liberación de contaminantes son distintas facetas del proceso de refino que pueden ser optimizadas con un buen control de calidad [8]. Para ello, la calidad debe garantizarse desde las materias primas (crudos) hasta el producto final, cubriendo todo el proceso de refino, mediante la monitorización de parámetros críticos a tiempo real.

La caracterización de los crudos permite determinar su composición y propiedades y, por ende, su uso potencial. Lo mismo ocurre con las distintas fracciones de destilados o demás corrientes de refino, el análisis de las cuales permite llevar a cabo la monitorización del proceso. En ambos casos, la información recopilada es crítica a la hora de tomar decisiones sobre las condiciones de operación de las etapas posteriores del proceso de refino. Por lo que respecta al producto final, el control de calidad se centra en el análisis de ciertas características físico-químicas de interés en función del tipo de producto y su aplicación comercial posterior. Para cada producto existen un gran número de parámetros físico-químicos a determinar, los resultados de los cuales deben cumplir con las especificaciones técnicas correspondientes para poder ser comercializados. El valor de estas especificaciones viene reforzado por normas legales que aseguran que el producto cumple con los límites mínimos de los estándares de calidad. Para la determinación de cada uno de estos parámetros, existen métodos estándar, como los de la *International Standardization Organization* (ISO) y la *American Society for Testing and Materials* (ASTM), definidos por las normativas correspondientes según su ámbito de aplicación.

1.1.4. *Productos refinados*

La industria petroquímica produce un gran número de productos refinados, la mayoría de los cuales se enumeran en la Tabla 1.1. De entre todos ellos, es preciso destacar los gasóleos y las parafinas, objetos de estudio en esta tesis. Las principales características y aplicaciones de ambos productos se detallan a continuación.

▪ **Gasóleos**

El gasóleo, también conocido como diésel o petrodiesel, es un derivado refinado del petróleo, constituido por una mezcla de compuestos parafínicos, naftalénicos y aromáticos, con puntos de ebullición que oscilan entre los 149 y 371 °C [9]. Por lo que respecta a la longitud de las cadenas carbonadas, los gasóleos suelen incluir cadenas entre 8 y 22 C, aunque la mayoría tienen entre 17 y 20 C y son de carácter parafínico. Como se ha descrito en el apartado anterior, el gasóleo proviene principalmente de las fracciones medias de la destilación atmosférica, aunque también se recupera en las fracciones ligeras de la destilación al vacío. Pero no todos los gasóleos se consiguen por destilación directa. En muchos casos, otras fracciones más pesadas son sometidas a tratamientos de FCC, hidrocrqueo, *visbreaking* y coquización, para conseguir gasóleo. De hecho, el FCC suele ser la fuente mayoritaria de gasóleo en las refinerías [4]. Gran parte del gasóleo obtenido se somete posteriormente a hidrotreatmento, con el objetivo de eliminar ciertas impurezas. En especial, se pretende eliminar los compuestos de azufre, que son corrosivos y provocan deposiciones en el motor, contribuyendo a su desgaste. Además, también generan emisiones perjudiciales para el medioambiente, motivo por los que sus niveles están regulados [9].

Aunque los gasóleos también se utilizan como combustible para calefacción doméstica, su uso principal es como carburante en motores de autocombustión o diésel, tanto de transporte pesado y maquinaria, como de transporte ligero. Algunas de las condiciones generales que debe cumplir un gasóleo para ser empleado como carburante son la ausencia de elementos corrosivos (como los compuestos de azufre),

una volatilidad adecuada para evitar los humos en el escape, unos niveles de viscosidad que permitan una inyección rápida, y un retraso de encendido bajo [2,9]. Esta última característica se explica por el funcionamiento de los motores diésel. En este caso, en lugar de una fase de explosión como se da en los motores de gasolina, tiene lugar una fase de combustión debida a la inyección de gasóleo a altas presiones y su contacto con el aire caliente comprimido en el cilindro. Si el retraso es muy grande, el rendimiento del motor baja, produciendo más humos y aumentando el consumo de carburante. Como medida del retraso de encendido del carburante se utiliza el índice de cetano (CI, *Cetane Index*), descrito en la Tabla 1.2. Cuanto más elevado es el número de cetano, menor es el retraso de la ignición y mejor la calidad de combustión [9]. El CI depende principalmente de la composición química del gasóleo. Los crudos parafínicos dan lugar a un gasóleo muy apreciado, mientras que los gasóleos obtenidos de crudos aromáticos presentan muy mal rendimiento como carburantes, por lo que normalmente suelen craquearse y se destinan a otros usos o se les adicionan pequeñas cantidades de nitrato y nitrito de etilo para mejorar sus características de inflamación [2]. Además, una baja relación de C/H de los hidrocarburos de un gasóleo, favorece que estos ardan completamente, sin dejar residuos carbonosos, perjudiciales para la conservación del motor [2].

Cada refinería produce distintos grados de diésel, que se distinguen principalmente por sus contenidos de azufre, CI, densidad y otras propiedades físicas. Por lo general, cada grado es una mezcla de 3 a 5 componentes producidos en la refinería [4]. Además en muchos casos, también se incluye biodiésel a la mezcla. Esto se debe a la similitud entre las propiedades físicas de ambos carburantes, lo que posibilita el uso del biodiésel puro o mezclado con diésel, sin ninguna modificación del motor ni del sistema de almacenaje y distribución del carburante [10,11,12].

En los últimos años, los biodiésels han ganado especial atención debido a su carácter de energía renovable y su biodegradabilidad. Su síntesis se lleva a cabo mediante reacción catalítica de transesterificación de aceites vegetales o grasas animales con alcoholes de cadena corta (metanol o etanol), para formar los metil/etil ésteres de ácidos grasos, conocidos como FAME (biodiésel) y glicerina. Además de aceites vegetales, también se pueden usar los aceites de freír como materia prima para

la formación de biodiésel, reutilizando así residuos domésticos e industriales [11,12]. Como ventajas, los biodiésel no contienen compuestos sulfurados, ni aromáticos, presentan un elevado CI y mayores propiedades lubricantes. Además, su uso reduce las emisiones de gases invernadero, partículas y azufre. Por lo que respecta a las emisiones de CO₂ generadas por el uso de diésel y biodiésel, estas son comparables, pero la obtención de biodiésel reduce las emisiones de producción, ya que se basa en el cultivo de plantas [13].

La producción y comercialización tanto de gasóleos, como de mezclas de gasóleos y biodiésel, exige el cumplimiento de especificaciones de numerosos parámetros físico-químicos para asegurar su calidad. En el caso de los gasóleos y las mezclas de estos con biodiésel, la normativa europea EN 590 [14] y la de la *American Society for testing and Materials*, ASTM D975 [15] constituyen los referentes a seguir. Cada una de estas normativas define las especificaciones para los distintos parámetros a controlar y los métodos de referencia a aplicar para la determinación de cada uno de ellos. Estos parámetros incluyen desde propiedades fisicoquímicas como la densidad, el CI o la inflamabilidad, típicas de cualquier petrodiesel, hasta el contenido de ciertos componentes de la mezcla, como los FAME) [16]. Algunos de estos parámetros han sido determinados en esta tesis y se describen en la Tabla 1.2. En la misma tabla también se listan los métodos de referencia para su determinación. Como se puede ver, se requiere un equipo específico para cada parámetro, lo que encarece en gran medida el control de calidad. Se trata mayoritariamente de equipos que trabajan *at-/off-line*, por lo que requieren muestreo previo a cada análisis. Aunque el muestreo se lleva a cabo unas tres veces al día, en cada ensayo se analizan solamente unos pocos mililitros de muestra, por lo que es dudosa la representatividad de las fracciones analizadas en la producción en continuo de una planta petroquímica, que puede producir varias toneladas de producto por minuto. Por lo que respecta al tiempo de análisis, en general, cada ensayo puede durar desde unos pocos minutos hasta más de media hora, dependiendo del ensayo y del tipo de muestra. A este tiempo, se le añade el tiempo de preparación de la muestra y el equipo, lo que también conlleva consumo de disolventes, como mínimo, para la limpieza del instrumento). También es importante mencionar que, en algunos de los análisis

(como la destilación o el método Pensky-Martens) se trabaja a altas temperaturas, lo que supone un consumo de energía elevado y un riesgo añadido para los analistas.

Tabla 1.2. Definición y metodología analítica utilizada para determinar algunos de los parámetros normativos en gasóleos y mezclas de gasóleos con biodiésels [9,14,15,16].

Parámetro	Definición	Test		Técnica
		EN 590	ASTM	
Densidad a 15°C (Dens)	Unidad de masa por unidad de volumen a 15 °C y 101,3 kPa	EN ISO 3675	D 4052	Areómetro (o densímetro)
		EN ISO 12185	---	Método del tubo en U oscilante
Índice de cetano (CI)	Porcentaje de cetano en una mezcla referencia con igual punto de inflamación que el carburante sometido a prueba	EN ISO 4264	D 4737	Ecuación de las cuatro variables
Contenido en azufre (SC)	Contenido de compuestos sulfurados del carburante	EN ISO 20846	D 5453	Fluorescencia en el ultravioleta
		EN ISO 20884	D 7039	Espectrometría por fluorescencia de rayos X de longitud de onda dispersiva
Destilado del 95% (D 95%)	Temperatura a la cual se ha destilado el 95% del carburante	EN ISO 3405	D 86	Destilación manual o automática a presión atmosférica
Punto de nube (CP)	Temperatura a la cual empiezan a precipitar ciertos compuestos del combustible	EN ISO 23015	D 2500	Baño de enfriamiento de temperatura constante
Punto de inflamabilidad (FP)	Temperatura más baja a la que un material volátil se enciende bajo efecto de una fuente de ignición	EN ISO 2719	D 93	Método Pensky-Martens en vaso cerrado

Contenido en FAME	Reflejo del contenido en biodiésel	EN 14078	D 7371	Espectroscopia del infrarojo
-------------------	------------------------------------	----------	--------	------------------------------

▪ Parafinas

Las parafinas industriales son mezclas de hidrocarburos saturados (alcanos) que se obtienen como coproducto en la fabricación de aceites lubricantes en la industria petroquímica. Como se ha descrito anteriormente, la materia prima para la obtención de estos aceites es el residuo de la destilación atmosférica del crudo. Las parafinas representan alrededor de un 15% en peso de este residuo y han de ser separadas del aceite para evitar su cristalización a bajas temperaturas [7].

El residuo de la destilación atmosférica alimenta la unidad de destilación al vacío que separa el crudo en cortes de destilación crecientes, que implican MW, viscosidad y punto de fusión también crecientes. Pero esta separación también afecta a la composición y estructura del producto: a fracciones más pesadas, mayor nº de ramificaciones y menor contenido de cadenas lineales, lo que dificulta la formación de estructuras cristalinas grandes típicas de las fracciones más ligeras [7]. Las parafinas, concretamente, proceden de las fracciones ligeras y medias del destilado al vacío, lo que se conoce como ceras macrocristalinas. Químicamente, se trata de alcanos, con un 80-90% de cadenas lineales (*n*-parafinas) con longitudes medias entre C20 y C30 principalmente, acompañadas de alcanos ramificados (isoparafinas) e incluso cíclicos (cicloparafinas). Sus puntos de fusión varían entre 50-70°C y solidifican en forma de cristales grandes y bien definidos [17,18].

Tanto los destilados obtenidos en estas fracciones, como el *bright stock*, se someten al proceso de desparafinado para separar el aceite de las parafinas en la medida deseada. Para conseguir parafinas con los niveles de aceite requeridos para la mayoría de aplicaciones (entre 0 y 1,5 %), es necesario aplicar el proceso de extracción tres veces, con los mismos disolventes, pero a temperaturas cada vez mayores [7].

El uso principal de la parafina es la fabricación de velas y productos relacionados, pero su uso también se extiende a la industria textil, papelera, maderera

y del caucho, como impermeabilizante y protector; en la industria cosmética para la fabricación de cremas; y en la industria alimentaria, como recubrimiento de papel parafinado, adhesivo en envases, conservante en dulces, frutas y quesos, etc. Para la mayoría de estas aplicaciones, el color de las parafinas obtenidas del desparafinado es demasiado intenso por lo que son sometidas a procesos de acabado con tierras decolorantes o percolación con bauxita. Para las aplicaciones alimentarias, las especificaciones son especialmente rigurosas y requieren la eliminación de la mayoría de impurezas además del color y el olor. Para ello se someten a los procesos de refinado con ácido sulfúrico e hidrogenación a alta presión. Las parafinas refinadas que cumplen los ensayos de la *Food and Drug Administration* (FDA), en 21 CFR 172.886 (para uso en alimentos) [19] o 21 CFR 178.3710 (para uso en contacto con alimentos) [20], y de la Farmacopea Europea [21], pueden ser utilizadas para estar en contacto directo con alimentos. Según la normativa de la FDA [20], para tener calidad alimentaria, la parafina debe ser completamente blanca en estado sólido e incolora y transparente en estado fundido, con niveles de azufre por debajo de 5 ppm y eliminación total de hidrocarburos aromáticos policíclicos (PAHs), potencialmente tóxicos y carcinógenos [22,23].

La combinación de las propiedades físicas (punto de fusión/congelación, dureza/penetrabilidad, viscosidad...), funcionales (impermeabilidad, flexibilidad, adhesión...) y químicas (contenido de aceite, color, olor, estabilidad frente a la luz y el calor...) de una parafina es lo que determina su aplicación última. La mayoría de estas propiedades se determina mediante análisis rutinarios definidos por distintas normas ASTM [18,24].

Cualquier diferencia observada en las propiedades de dos parafinas responderá a diferencias en su composición química, es decir, al origen y la composición de los crudos y al corte de destilación del que proceden. Dado que las parafinas refinadas están constituidas por hidrocarburos saturados, la proporción de cadenas lineales y ramificadas determinará en gran parte las características del producto y su aplicación posterior. Gran parte de los hidrocarburos ramificados que constituyen la parafina purificada son las isoparafinas, también conocidas como aceite. Un exceso de aceite tiende a exudarse de la parafina, dándole un tacto grasiento indeseable. Además, puede

INTRODUCCIÓN

tener efectos negativos sobre propiedades como la dureza, la adhesión, la consistencia, la flexibilidad, la resistencia al rozamiento, el punto de fusión, el olor, el color, etc. El aceite es muy difícil de eliminar totalmente de la parafina, por lo que su determinación es de especial importancia en la caracterización industrial de parafinas. Esta determinación se lleva a cabo rutinariamente mediante un proceso de extracción con MEK, descrito en la norma ASTM D721 [25], muy similar al proceso de desparafinado industrial descrito anteriormente (apartado 1.1.2). Este método de análisis es lento y poco preciso, ya que implica operaciones como disolución, cristalización, filtración, etc. a bajas temperaturas. Sin embargo, el resultado se utiliza como indicativo de calidad de la parafina en cuestión, repercutiendo directamente en su aplicabilidad.

1.2. Industria farmacéutica

La industria farmacéutica se encarga de la manufactura de productos medicinales para la prevención y el tratamiento enfermedades. Los fármacos constan principalmente de uno o más principios activos (API, *Active Pharmaceutical Ingredients*) que llevan a cabo la acción terapéutica, y los excipientes, que constituyen la mayor parte del fármaco. Los excipientes son componentes farmacéuticamente inertes, pero indispensables para el correcto funcionamiento del fármaco. Mejoras en las condiciones de administración (sabores, anestésicos según la vía de administración), el aseguramiento de las dosis exactas de API (diluyentes o *bulking agents*) y la absorción mayoritaria por su diana en el organismo (disgregantes, recubrimientos entéricos), el mantenimiento de la calidad durante la vida útil del fármaco (estabilizantes, antioxidantes) o la simplificación del proceso de manufactura (lubricantes, aglutinantes) son algunos de los beneficios que aportan los excipientes. Pero son las propiedades del API las que dictan su elección, así como la forma farmacéutica del medicamento y su vía de administración.

Actualmente, existe una gran cantidad de formas farmacéuticas distintas, entre las que se incluyen preparaciones sólidas (polvos, granulados, comprimidos, cápsulas), semisólidos (cremas, geles, supositorios) y líquidos (jarabes, colirios, aerosoles). Cada una de estas formas presenta distintas propiedades que determinan la vía de aplicación y, por supuesto, distintos procesos de producción. A pesar de ello, los distintos procesos de producción de la industria farmacéutica presentan ciertos puntos en común. Las materias primas para la manufactura de fármacos son muy diversas, pero en todos los casos se trata de materiales muy puros y de alta calidad. La formulación de las materias primas, que da lugar al fármaco final, suele ser un proceso multietapa, que involucra operaciones unitarias muy técnicas y que trabajan mayoritariamente en *batch*. Esto implica que las operaciones se llevan a cabo en discontinuo, sobre una cantidad de material llamada lote. Además, los procesos de producción de la industria farmacéutica están altamente optimizados para conseguir que el producto final se adecúe a las estrictas especificaciones a las que se encuentra sometido este sector, lo que implica que la variabilidad del producto final sea muy limitada [26].

En la primera parte de este apartado, se describen las principales características de dos formas farmacéuticas concretas tratadas en esta tesis: los comprimidos recubiertos (apartado 1.2.1) y los liofilizados (apartado 1.2.2), profundizando principalmente en las etapas que siguen sus respectivos procesos productivos. La segunda parte del apartado se centra en dos aspectos muy característicos de la industria farmacéutica: la alta y estricta regulación a la que se encuentra sometida (apartado 1.2.3) y las consiguientes estrategias desarrolladas para llevar a cabo los controles de calidad (apartado 1.2.4).

1.2.1. Producción de comprimidos

Los comprimidos son la forma farmacéutica más frecuente [27,28]. Su popularidad es debida a las ventajas que ofrece, ya que es una forma estable, fácil transportar y de administrar (mayoritariamente por vía oral), permite controlar el tiempo de la acción farmacológica y su proceso de producción es bastante rápido y automatizado [27,29]. A grandes rasgos, este proceso consiste en la compresión de mezclas en polvo o granuladas, pero esto conlleva una serie de etapas (Figura 1.1) en las que la materia prima se somete a distintos tratamientos mayoritariamente físicos. A continuación, se describen las etapas básicas de la fabricación de comprimidos.



Figura 1.1. Etapas del proceso de producción de comprimidos farmacéuticos.

▪ Mezclado

La primera etapa básica de la fabricación de la mayoría de productos farmacéuticos es el mezclado de las materias primas que, en el caso de los comprimidos, suelen encontrarse en forma de polvo [30,31]. La finalidad última del mezclado es conseguir una distribución homogénea de todos los componentes en el fármaco [32,33], necesaria para asegurar la uniformidad de contenido entre los comprimidos individuales finales.

Existen una gran variedad de equipos para el mezclado de materiales pulvulentos, que se pueden clasificar en dos tipos generales [30]: los mezcladores móviles, en los que la carcasa, que puede tener distintas formas, rota alrededor de un eje; y los agitadores de carcasa fija, equipados con bandas en espiral, hélices o paletas en su interior, que se encargan de cizallar la mezcla, rompiendo los posibles aglomerados, y arrastrar parte de ella de un extremo al otro del contenedor para homogeneizar toda la mezcla. Los tiempos de mezclado óptimos dependen de la eficiencia del mezclador utilizado, pero también de los materiales a analizar [33]. Factores como las diferencias de tamaño de partícula y densidad entre componentes o tiempos de mezclado mayores al óptimo, favorecen la segregación, afectando negativamente a la uniformidad de la mezcla [27].

▪ Granulación

La producción de comprimidos comporta, en la mayoría de los casos, una etapa de granulación, que implica la adhesión de las partículas en polvo iniciales para formar estructuras mayores (gránulos). La granulación mejora las propiedades de deslizamiento de la mezcla, consiguiendo un llenado de matrices más uniforme y sin aire, esencial para la compactación. Además, previene la segregación de los componentes, muy típica de mezclas en polvo, debido a las diferencias de tamaño de partícula y densidad de los componentes [29,33].

La granulación se puede llevar a cabo por vía seca o por vía húmeda [29,33], siendo esta última la más común [27]. La **granulación por vía húmeda** implica el amasado de la mezcla en polvo con una dispersión del excipiente aglutinante, que garantiza la adhesión de las partículas una vez seco. Este proceso puede llevarse a cabo en mezcladores/ granuladores de alta velocidad, similares a los agitadores de palas descritos en el apartado anterior, y que permiten llevar a cabo mezclado, amasado y granulación, en el mismo equipo y de forma muy rápida. Otra alternativa es la granulación en lecho fluido, en la cual, se inyecta aire caliente por la parte inferior de la mezcla, formando una suspensión de partículas en polvo, a la vez que se pulveriza el fluido de granulación por la parte superior. De esta forma, las partículas se adhieren por colisión, formándose los gránulos, que se secan gracias al aire caliente. A parte del secado por lecho fluido, los gránulos obtenidos también pueden secarse en hornos de bandejas o por aplicación de microondas. Cuando los fármacos no son estables en las condiciones de la granulación húmeda, es decir, contacto con soluciones acuosas y/o calor, se utiliza la **granulación por vía seca** [29]. En este caso, las partículas de polvo se agregan en prensas de alta presión o mediante compresión por rodillo y la tableta resultante se fragmenta de nuevo mediante molienda para obtener el material granular.

En todos los casos, los gránulos obtenidos se tamizan para conseguir la fracción del tamaño deseado (normalmente 0.25-0.70 mm, rompiendo los aglomerados y eliminando el material demasiado fino [27]. Posteriormente, los gránulos se mezclan con otros excipientes necesarios para la formulación, como lubricantes, que facilitan el desmoldado del comprimido después de la compactación, y desintegrantes, que promueven la desintegración del comprimido cuando entran en contacto con agua [27,29].

▪ **Compactación/compresión** [29]

La mezcla de todos los ingredientes del fármaco, ya sea en polvo o en forma de gránulos, se compacta para dar lugar a los núcleos, *i. e.* comprimidos no recubiertos de geometría definida. La máquina de compactación consta de una matriz, que se

rellena con un volumen concreto del material a compactar. Un troquel superior aplica la presión de compactación de forma mecánica. Esta presión debe estar bien definida para que los comprimidos sean suficientemente duros como para no fragmentarse durante su manipulación y, a la vez, sean capaces de disgregarse en el organismo en el tiempo establecido [32,34]. Finalmente, el comprimido se desmolda por presión ascendente de un troquel situado en la parte inferior de la matriz. Esta operación suele agilizarse mediante el uso de prensas rotativas, con un gran número de moldes que se someten al ciclo de compactación de forma secuencial [35].

▪ **Recubrimiento**

La etapa final del proceso de producción de comprimidos es el recubrimiento de los núcleos, en el que se aplica un material de cobertura sobre el exterior de la formulación con el objetivo de aportar ciertas ventajas sobre los sistemas no recubiertos. Las funciones principales de cualquier recubrimiento son: preservar la unidad y el aspecto del comprimido de las condiciones ambientales (aire, humedad y luz) y emascarar olores, sabores o apariencias desagradables. El recubrimiento también puede facilitar la deglución del comprimido, mejorar su apariencia o aumentar la resistencia a la manipulación. En muchos casos, también se utiliza para controlar la velocidad de liberación del fármaco, lo que se conoce como recubrimientos entéricos. Esto permite modular la duración del efecto farmacológico, así como retrasar su liberación hasta el lugar adecuado del organismo, lo que protege el fármaco de una degradación prematura y previene posibles efectos locales indeseados.

Aunque existen distintos tipos de recubrimiento, en la mayoría de casos, estos consisten en una fina película polimérica (polímeros acrílicos o celulósicos) que puede ir acompañada de otros componentes como pigmentos, que pueden facilitar la identificación de los comprimidos y mejorar su aspecto, o plastificantes que aportan flexibilidad [29,36]. Las características de los componentes del recubrimiento son las que dictan su uso posterior.

INTRODUCCIÓN

Para su aplicación, los polímeros se disuelven o dispersan en una solución acuosa u orgánica antes de ser pulverizados sobre los núcleos. El proceso tiene lugar en contenedores rotatorios, que rotan en un eje inclinado, o en lecho fluido, de forma que los comprimidos se mantienen en movimiento continuo y facilita el recubrimiento de la superficie de todos los comprimidos con la solución de recubrimiento vaporizada [36]. La evaporación de los disolventes, que normalmente comporta la aplicación de calor sobre el contenedor de recubrimiento, deja un depósito fino de material de recubrimiento alrededor de cada núcleo. Esta fina capa se adhiere al comprimido, manteniendo la forma y las inscripciones originales del núcleo y dando lugar a los comprimidos recubiertos [34].

1.2.2. Producción de liofilizados

En el caso de los productos liofilizados, suele tratarse de fármacos con componentes biológicos, como proteínas, que son termolábiles y/o inestables en solución acuosa. Por este motivo, las soluciones o suspensiones de estos fármacos se someten al tratamiento de liofilización, también conocido como *freeze-drying*, que permite eliminar la mayor parte del agua a bajas temperaturas. De esta forma, se consiguen fármacos sólidos suficientemente estables como para soportar los procesos de distribución y almacenaje y que pueden ser posteriormente reconstituidos por la



Figura 1.2. Etapas del proceso de liofilización.

adición de agua u otras soluciones específicas en el momento de su administración [29].

La liofilización es un proceso costoso, ya que requiere de equipos caros y supone un consumo energético elevado. Además, es mucho mucho más largo, ya que consta de distintas etapas que pueden durar días o incluso semanas si no están bien optimizadas (Figura 1.2). La primera de las fases es la **congelación**, en la que se convierte la mayoría del agua de la solución farmacéutica en hielo. Este hielo, se elimina por sublimación en la etapa de **secado primario**, mientras que el resto de agua no congelada se elimina por desorción durante el **secado secundario**. A continuación, se describe cada una de las etapas del proceso con más detalle.

▪ Etapa de congelación

La congelación es el primer paso en el proceso de liofilización y aquel en el que más agua se elimina. Al final de esta etapa, que suele durar varias horas, sólo un 1% del agua total inicial permanece en el fármaco [37,38].

Como se ha descrito arriba, durante la congelación, la mayoría del agua de la solución se convierte en hielo, mientras que los solutos, o cristalizan, o bien se transforman en un sistema de sólidos amorfos, como ocurre con las proteínas. Esto se consigue con un descenso de la temperatura, que suele llegar alrededor de los -50°C . En muchos casos, este proceso incluye un paso de *annealing*, que implica el mantenimiento de la temperatura por encima de la temperatura final de congelación durante cierto tiempo para asegurar la cristalización del *bulking agent*. Este agente, se utiliza como relleno inerte en los productos liofilizados, donde la concentración del API es muy baja [38]. En muchos casos, el *bulking agent* de elección es el manitol, ya que su forma cristalina es fácilmente liofilizable, proporciona consistencia y elegancia al producto final y evita problemas de colapso [39]. Se entiende como colapso de una muestra la pérdida de su estructura macroscópica, un fenómeno indeseado que puede tener lugar durante o después del proceso de liofilización. Si el manitol no cristaliza durante la congelación, puede provocar problemas de estabilidad en etapas

posteriores, dado que su cristalización implica la liberación de agua al medio y puede conllevar la rotura del vial [40,41].

El proceso de congelación comporta muchos cambios, que pueden afectar a la estabilidad de las proteínas [37,42]: el incremento de la concentración de proteína, que favorece su agregación; los cambios de pH debidos a la cristalización de las sales tamponadoras; la reducción de las interacciones hidrofóbicas de la proteína nativa con su entorno, debidas a la deshidratación por cristalización; la formación de grandes interfases hielo-agua y el incremento de la fuerza iónica. Estos cambios se pueden minimizar optimizando la formulación del fármaco (escogiendo los tampones adecuados y reduciendo su concentración, incrementando la concentración de proteína, usando surfactantes, crio- y lioprotectores) y la velocidad de enfriamiento. [39,37,43,44].

En la mayoría de los casos, también se añaden estabilizantes a la formulación (típicamente sacarosa y trehalosa), para mantener la estabilidad del liofilizado [39]. Los estabilizantes presentan elevada viscosidad después de la etapa de congelación, lo que ralentiza enormemente la desnaturalización de las proteínas, evitando que tenga lugar en el marco del proceso de liofilización [37]. Es por este motivo que, contrariamente a lo que ocurre con el *bulking agent*, la cristalización de los agentes estabilizadores de la proteína debe evitarse, para mantener el entorno de la proteína y evitar su desestabilización [39].

▪ Secado primario

Una vez la muestra está solidificada, empieza el secado primario, en el que se elimina el hielo por sublimación al vacío. Para ello, se aumenta la temperatura de la cámara hasta llegar a la temperatura adecuada (normalmente $\leq -15^{\circ}\text{C}$), que debe mantenerse a lo largo de toda la etapa, además de disminuir la presión para acelerar la sublimación [37]. Dado que esta etapa es la más larga del proceso, debe de estar bien optimizada para minimizar el tiempo de trabajo y asegurar la máxima calidad del producto final. En este sentido, la temperatura de trabajo es uno de los factores más determinantes:

cuanto más elevada sea, menor será el tiempo de secado primario, pero mayor el riesgo de colapso de la muestra [37]. De nuevo, el colapso se contraresta con un exceso de *bulking agent* (manitol o glicina).

El final de esta etapa tiene lugar cuando el hielo se ha eliminado, por lo que en muchos casos, es importante monitorizar el proceso para poder establecer su punto final.

▪ Secado secundario

Después del secado primario, el producto aún contiene agua residual, que supone entre un 5 y un 20% del total del sólido [37]. El objetivo del secado secundario es reducir esta humedad residual hasta niveles menores al 1% [29,37]. Dado que, en este caso, la eliminación de la humedad se consigue por desorción, la temperatura de trabajo es mucho mayor que en las etapas anteriores (40-50 °C), pero no suficiente como para desnaturalizar las proteínas ‘secas’. El aumento de temperatura se lleva a cabo lentamente (0.1-0.15 °C·min⁻¹) para prevenir el colapso de los sólidos amorfos de la muestra. La temperatura final se mantiene durante unas 3-6 h, dependiendo de la concentración de sólidos. La presión, en cambio, suele mantenerse igual que en el secado primario.

▪ Almacenaje

Aunque no son estrictamente parte del proceso de *freeze-drying*, durante la extracción del material liofilizado de la cámara de liofilización y su posterior almacenamiento a largo plazo, las proteínas liofilizadas también pueden perder parte de su actividad. Existe una gran variedad de mecanismos que pueden favorecer la inestabilidad proteica, entre los que se incluyen procesos físicos (agregación) y diferentes procesos de degradación química (desamidación, oxidación, hidrólisis...), que dependen de factores como la temperatura de almacenamiento, el pH, la naturaleza de los excipientes o la humedad residual [42,45]. Además, el contenido de humedad de un

producto liofilizado puede variar significativamente durante el almacenamiento debido al agua de los tapones y las fugas o a la liberación de agua de excipientes en formas hidratadas o cristalización de amorfos [42]. La reabsorción de agua, así como el agua residual remanente después de la liofilización, favorecen la desestabilización de los productos liofilizados [46], que ya de por sí son proclives a la degradación durante almacenajes a largo plazo. En algunos, casos los productos se almacenan en condiciones de refrigeración para minimizar la degradación, pero en cualquier caso, la formulación y las condiciones del proceso son determinantes para garantizar la estabilidad del producto liofilizado durante el almacenamiento [39,43,44,46].

En general, el proceso óptimo de *freeze-drying* es aquel que consigue mayor calidad del fármaco a menor coste, lo que requiere no sólo la optimización de la formulación y los distintos parámetros del proceso [37,38], sino también la monitorización del estado del fármaco en cada momento y la determinación del punto final de cada una de las etapas [47,48,].

1.2.3. Regulación de la industria farmacéutica

Una de las características más destacadas de la industria farmacéutica es que se encuentra altamente regulada por distintos organismos nacionales e internacionales, como la FDA en los EUA, la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS) en España o la *European Medicine Agency* (EMA) en la Unión Europea. Estas instituciones obligan a cualquier industria farmacéutica al cumplimiento de las buenas prácticas de fabricación (GMP, *Good Manufacturing Practices*), unas guías con requerimientos básicos para asegurar la calidad de los productos y evitar riesgos para el consumidor. Además, las nuevas GMPs desarrolladas por la FDA en el 2004 y conocidas como cGMP (current GMP) del siglo 21, también promueven la innovación en el sector. En cualquier caso, las normativas GMP constituyen sólo una guía general para las industrias, de modo que su interpretación detallada y el cumplimiento en consonancia, son responsabilidad individual de cada compañía.

Los requerimientos cualitativos y cuantitativos para la composición de los medicamentos, los ensayos a llevar a cabo en materias primas, fármacos y demás material utilizado en la producción vienen definidos en las farmacopeas. Estos documentos sirven como instrucciones técnicas para la manufactura y el control de calidad de medicamentos y pueden presentar diferentes ámbitos de alcance y repercusión legal. En la mayoría de los casos, una farmacopea es de obligado cumplimiento para la comercialización de productos farmacéuticos en todo su ámbito de aplicación. Algunas de las farmacopeas más prestigiosas son la europea, *European Pharmacopoeia* (Ph. Eur.), y la americana, *United States Pharmacopea* (USP), los métodos y especificaciones de las cuáles son, en muchos casos, adoptadas y/o adaptadas por otras farmacopeas. Las divergencias entre requerimientos técnicos de los distintos países dificultaba la globalización de los mercados, ya que implicaba la duplicación de ensayos y procedimientos muy largos y costosos para reunir las condiciones específicas de cada país. Con el objetivo de armonizar los requerimientos técnicos de la industria farmacéutica, en 1990 se creó la *International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use* (ICH), una organización que reúne las autoridades regulatorias en la materia de la Unión Europea, Japón y los Estados Unidos para discutir los aspectos técnicos y científicos del registro de fármacos, que se plasman en las guías ICH.

1.2.4. Control de calidad de procesos y productos farmacéuticos

El control de calidad constituye un elemento fundamental en los procesos de producción farmacéutica. Esto implica someter todos los productos (materias primas, intermedios de producción y productos acabados, pero también productos en desarrollo o productos liberados hasta su caducidad) a una gran variedad de determinaciones de características físicas y químicas. Parámetros como la identificación de componentes, la riqueza y la uniformidad del contenido en API, la determinación de impurezas y humedad o la velocidad de disolución de cápsulas y comprimidos son algunos de los ensayos requeridos habitualmente por las

farmacopeas. El planteamiento tradicional del control de calidad implica la medida de estos parámetros mediante ensayos en el laboratorio: técnicas cromatográficas o espectroscopia UV mayoritariamente, aunque también se utilizan técnicas como la valoración por Karl Fischer (KF) para la determinación del contenido en agua o baños específicos para los ensayos de disolución. Este tipo de ensayos conllevan la retirada de parte del material de producción, ya sea materia prima, producto acabado o producto intermedio, y el consiguiente análisis en el laboratorio, que puede situarse lejos del proceso (*off-line*) o en la misma planta de producción (*at-line*). Asimismo, todas estas técnicas requieren preparación de la muestra que, junto con el tiempo de análisis, supone entre media hora (UV) y varias horas (en algunos test de disolución [49]) de trabajo, además de un elevado consumo de disolventes. El gran número de muestras a analizar para maximizar la representatividad de los análisis, así como el gran número de parámetros a determinar en cada muestra para asegurar la calidad del producto final en todos los sentidos, implican una gran inversión en tiempo y dinero. Esto implica que el planteamiento tradicional del control de calidad tenga un impacto negativo sobre la productividad y la competitividad. La necesidad de compatibilizar la demanda de mayor productividad con menores costes de producción, cumpliendo a la vez con unos estándares de calidad cada vez más exigentes, son las causas principales del cambio de enfoque en el control de calidad de la industria farmacéutica. En el nuevo enfoque, la calidad del producto final se consigue identificando aquellos parámetros que tienen más impacto en el proceso y utilizando los métodos adecuados para su seguimiento a tiempo real (*on-line* o *in-line*) durante todo el proceso. En la aplicación *in-line*, la interficie del analizador está directamente en contacto con el proceso, mientras que cuando la aplicación se lleva a cabo de forma *on-line*, se necesita recirculación de la muestra para que esta pueda llegar al analizador. En cualquier caso, ambos métodos implican una mejora respecto a los procesos *off-* y *at-line*, ya que no requieren la extracción de la muestra y los análisis se pueden llevar a cabo de forma mucho más automatizada y en tiempo real.

Este nuevo enfoque conocido como tecnología analítica de procesos (PAT, *Process Analytical Technology*) es una iniciativa impulsada por la FDA, basada en los principios de la química analítica de procesos (PAC, *Process Analytical Chemistry*) o

análisis de procesos (PA, *Process Analysis*). La misma FDA lo define como “un sistema para diseñar, analizar y controlar la producción mediante la medida de los parámetros de calidad oportunos, tanto en las materias primas, como en los materiales en proceso de elaboración, con el objetivo de asegurar la calidad del producto final” [50].

Además de mantener las características de seguridad y calidad esperadas en el producto final, la aplicación del PAT también mejora la eficiencia de manufactura. La monitorización de los parámetros críticos a lo largo del proceso proporciona información en cuanto a la situación en tiempo real del producto/proceso. Esta información puede aprovecharse para la detección y corrección prácticamente inmediatas de cualquier problema, lo que se traduce en una reducción del tiempo y los costes de producción respecto al enfoque tradicional. La iniciativa PAT promueve, además, la utilización de esta información para retroalimentar el propio proceso de producción, corrigiendo cualquier variable que pueda mejorar la calidad del producto final y posibilitando la implementación continua de mejoras en el proceso.

Aunque el alcance del PAT no se limita a la industria farmacéutica, la mayoría las aplicaciones publicadas (~70%) se desarrollan en este sector, debido a las exigencias de los organismos reguladores [51]. Un 25% de estas aplicaciones se focalizan en la identificación y cuantificación de principio activo, uno de los parámetros críticos durante el proceso de manufactura y en el producto acabado [51]. Las aplicaciones PAT, se pueden diferenciar según si se llevan a cabo antes, durante el proceso de producción. El primer caso, incluye todos los controles rutinarios de materias primas, usados para identificar y confirmar la calidad satisfactoria de cada lote. En muchos casos estos análisis se llevan a cabo en el mismo almacén, mediante sondas de fibra óptica y a través de las mismas bolsas de polietileno del embalaje. En el segundo caso, el control de los parámetros de calidad es más complejo debido a las condiciones de la planta de producción, mucho menos reproducibles que las del laboratorio. La monitorización de ciertos parámetros para la determinación del punto final de algunas etapas, como el mezclado o alguno de los pasos del liofilizado, son aplicaciones características llevadas a cabo durante el proceso de manufactura. Una vez terminado el proceso de fabricación, el objetivo no es determinar las propiedades físico-químicas del producto final en detalle, sino confirmar que todos los lotes

presentan la misma calidad, mediante ensayos de liberación en tiempo real que reemplacen los análisis del producto final.

A pesar de las ventajas del PAT, su implementación en la industria es lenta debido a la fuerte inversión inicial que comporta y la inseguridad de la aceptación de los nuevos métodos por parte de las autoridades regulatorias. Aunque este último argumento está revirtiendo y ya existen procesos totalmente PAT aceptados por las agencias reguladoras, la implementación de nuevas tecnologías de monitorización y control sigue siendo larga y costosa, ya que implica la capacitación de personal y la cualificación de la instrumentación (descrita por las GMP). Además, en este sector industrial tan regulado, cualquier método analítico, tanto cualitativo como cuantitativo, debe ser aprobado por las autoridades regulatorias antes de ser implementado en rutina. Para demostrar que un método cumple con las condiciones necesarias para el uso para el cual ha sido desarrollado, es imprescindible validarlo. Según los requisitos de la ICH [52] y la Ph. Eur. [53], la validación de métodos cualitativos requiere la evaluación de la robustez y la especificidad, mientras que para los modelos cuantitativos, los análisis de exactitud, linealidad y precisión (repetitividad y precisión intermedia) se añaden a los ya mencionados.

Uno de los puntos más críticos de la aplicación de esta nueva filosofía de trabajo es la elección de la técnica de análisis. Las singulares condiciones de medida dictadas por el PAT, exigen analizadores rápidos, y sencillos y aplicables *on-* o *in-line*, que permitan reflejar la situación del producto/proceso en tiempo real. La mayoría de analizadores aplicados como soluciones PAT son instrumentos espectroscópicos, integrados en los procesos, mediante celdas de flujo, ventanas de cuarzo y sondas de inmersión [51].

REFERENCIAS

- 1 Speight, J.G. (2014) *The Chemistry and technology of petroleum*. 5^a ed, CRC Press, Taylor & Francis Group, Boca Raton, FL.
- 2 Magdalena, M., López, B. (1996) *El petróleo y el gas natural, materias primas para la industria química*. En: Vian, A. *Introducción a la química industrial*. 2^a ed., Editorial Reverté, Barcelona.
- 3 Speight, J.G. (2005) *Petroleum Refinery Processes*. En: *Kirk-Othmer Encyclopedia of Chemical Technology* [en línea]. Wiley Online Library [Consultado en agosto de 2014]. Disponible en: <http://onlinelibrary.wiley.com/doi/10.1002/0471238961.1805060919160509.a01.pub2/pdf>
- 4 MathPro (2011) *Introducción a la refinación del petróleo y producción de gasolina y diésel con contenido ultra bajo de azufre*. The International Council on Clean Transportation (ICCT), Bethesda, MD.
- 5 Hidajat, K., Chong, S.M. (2000) Quality characterization of crude oils by partial least square calibration of NIR spectra profiles. *J. Near Infrared Spec.* **8**, 53-59.
- 6 Devold, H. (2013) *Oil and gas production handbook. An introduction to oil and gas production, transport, refining and petrochemical industry*. 3^a ed. ABB Oil and Gas, Oslo.
- 7 Sánchez, J. (2003) *Purificación de parafinas de petróleo por hidrogenación catalítica*. Rodríguez, F. (dir.) Tesis doctoral, Universidad Complutense de Madrid.
- 8 Khanmohammadi, M., Garmarudi, A.B., de la Guardia, M. (2012) Characterization of petroleum based products by infrared spectroscopy and chemometrics. *TrAC-Trend. Anal. Chem.* **35**, 135-149.
- 9 UFA Co-operative Limited (2009) [sitio web]. [Consultado en agosto de 2014]. Disponible en: http://www.ufa.com/petroleum/resources/fuel/diesel_fuel_resources.html
- 10 Monteiro, M.R., Ambrozina, A.R.P., Lião, L.M., Ferreira, A.G. (2008) Critical review on analytical methods for biodiesel characterization. *Talanta* **77**, 593-605.

- 11 Balabin, R.M., Safieva, R.Z. (2011) Near-infrared (NIR) spectroscopy for biodiesel analysis: Fractional composition, iodine value and cold filter plugging point from one vibrational spectrum. *Energ. Fuel.* **25**, 2373-2382.
- 12 Balabin, R.M., Smirnov, S.V. (2011) Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. *Anal. Chim. Acta* **692**, 63-72.
- 13 Komorizono, I., de Carvalho W.F., Poppi R.J. (2009) Application of near infrared spectroscopy and multivariate control charts for monitoring biodiesel blends. *Anal. Chim. Acta* **642**, 217-221.
- 14 UNE-EN 590:2009+A1:2010 (2010) Automotive fuels - Diesel - Requirements and test methods. *Comité Européen de Normalisation (CEN)*, Bruselas.
- 15 ASTM Standard D 975-12a (2012) Standard Specification for Diesel Fuel Oils. *ASTM International*, West Conshohocken, PA.
- 16 BOE-A-2010-13704 (2010) Boletín Oficial del Estado, N° 215. *Ministerio de Industria, Turismo y Comercio*, Madrid.
- 17 Leray, C. (2006) Waxes. En: *Kirk-Othmer Encyclopedia of Chemical Technology* [en línea]. [Consultado en agosto de 2014]. Disponible en:

<http://onlinelibrary.wiley.com/doi/10.1002/0471238961.2301240503152020.a01.pub2/pdf>
- 18 The International Group, Inc. (2012) [sitio web]. [Consultado en agosto de 2014]. Disponible en: <http://igiwax.com/reference/waxbasics.html>
- 19 21CFR172.886 (2013) Food additives permitted for direct addition to food for human consumption. *U.S. Food and Drug Administration (FDA)*, Silver Spring, MD.
- 20 21CFR178.3710 (2013) Indirect food additives: adjuvants, production aids, and sanitizers. *U.S. Food and Drug Administration (FDA)* Silver Spring, MD.
- 21 European Pharmacopoeia (2011). 7^a ed., *EDQM Council of Europe*, Estrasburgo.
- 22 Toxicological profile for polycyclic aromatic hydrocarbons (PAHs) (1995), Agency for Toxic Substances Disease Registry (ATSDR), U.S. *Department of Health and Human Services*, Atlanta, GA.

- 23 Woo, O.T., Chung, W.K., Wong, K.H., Chow, A.T., Wong, P.K. (2009) Photocatalytic oxidation of polycyclic aromatic hydrocarbons: Intermediates identification and toxicity testing. *J. Hazard. Mater.* **168**, 1192-1199.
- 24 ASTM International (2014) [sitio web]. West Conshohocken, PA. [Consultado en agosto de 2014]. Disponible en: www.astm.org
- 25 ASTM D721-06 (2006) Método de Ensayo Normalizado para Contenido de aceite en Parafinas del Petróleo, *ASTM International*, West Conshohocken, PA,.
- 26 Blanco, M., Cueva-Mestranza, R., Peguero, A. (2011) NIR analysis of pharmaceutical samples without reference data: Improving the calibration. *Talanta* **85**, 2218-2225.
- 27 Armstrong, N.A. (2007) Tablet Manufacture. En: Swarbrick, J. (ed.) *Encyclopedia of pharmaceutical technology*. Vol. 1., 3ª ed.; Informa Healthcare USA, Inc. New York, NY, pp. 3653-3672.
- 28 Maurer, L., Leuenberger, H. (2009) Terahertz pulsed imaging and near infrared imaging to monitor the coating process of pharmaceutical tablets. *Int. J. Pharmaceut.* **370**, 8-16,
- 29 Rzeszotarski, W.J. (2005) Pharmaceuticals. En: *Kirk-Othmer Encyclopedia of Chemical Technology*. [en línea]. Wiley Online Library [Consultado en agosto de 2014]. Disponible en:
<http://onlinelibrary.wiley.com/doi/10.1002/0471238961.1608011826011415.a01.pub2/pdf>
- 30 Aiache, J.M., Beyssac, E. (2007) Powders as Dosage Forms. En: Swarbrick, J. (ed.) *Encyclopedia of pharmaceutical technology*. Vol. 1., 3ª ed.; Informa Healthcare USA, Inc. New York, NY.
- 31 O'Connor, R.E., Schwartz, J.B., Felton, L.A. (2005) Powders. En: *Remington. The science of pharmacy and practice*, 21ª ed.; Lippincott Williams & Wilkins, Philadelphia, PA.
- 32 Peguero, A. (2010) *La espectroscopia NIR en la determinación de propiedades físicas y composición química de intermedios de producción y productos acabados*. Blanco, M. (dir.) Tesis doctoral, Universitat Autònoma de Barcelona.
- 33 Ciurczak, E.W., Drennen, J.K. (2002) *Pharmaceutical and medical applications of near-infrared spectroscopy*. Marcel Dekker, New York, NY.

- 34 Niazi, F.K. (2004) Handbook of pharmaceutical manufacturing formulations. Compressed solid products. Vol. 1; , CRC Press, Boca Raton, FL.
- 35 Bogda, M.J. (2007) Tablet Compression: Machine Theory, Design, and Process Troubleshooting. En: Swarbrick, J. (ed.) *Encyclopedia of pharmaceutical technology*. Vol. 1., 3^a ed.; Informa Healthcare USA, Inc. New York, NY.
- 36 Felton, L.A. (2007) Film Coating of Oral Solid Dosage Forms. En: Swarbrick, J. (ed.) *Encyclopedia of pharmaceutical technology*. Vol. 1., 3^a ed.; Informa Healthcare USA, Inc. New York, NY.
- 37 Tang, X.(C.), Pikal, M.J. (2004) Design of freeze-drying processes for pharmaceuticals: Practical advice. *Pharmaceutical Research* 21(2), 191-200.
- 38 Franks, F. (1998) Freeze-drying of bioproducts: putting principles into practice. *Eur. J. Pharm. Biopharm.* 45, 221-229.
- 39 Pikal, M.J. (2007) Freeze Drying. En: Swarbrick, J. (ed.) *Encyclopedia of pharmaceutical technology*. Vol. 1., 3^a ed.; Informa Healthcare USA, Inc. New York, NY.
- 40 Williams, N.A., Dean, T. (1991) Vial breakage by frozen mannitol solutions: correlation with thermal characteristics and effect of stereoisomerism, additives, and vial configuration. *J. Parent. Sci. Techn.* 45(2), 94-100.
- 41 Jiang, G., Akers, M., Jain, M., Guo, J., Distler, A., Swift, R., Wadhwa, M.V., Jameel, F., Patro, S., Freund, E. (2007) Mechanistic studies of glass vial breakage for frozen formulations. I. Vial breakage caused by crystallizable excipient mannitol. *PDA J. Pharm. Sci. Technol.* 61(6), 441-451.
- 42 Wang, W. (2000) Lyophilization and development of solid protein pharmaceuticals. *Int. J. Pharm.* 203(1-2), 1-60.
- 43 Sarciaux, J.M., Mansour, S., Hageman, M.J., Nail, S.L. (1999) Effects of buffer composition and processing conditions on aggregation of bovine IgG during freeze-drying. *J. Pharm. Sci.* 88(12), 1354-1361.
- 44 Abdelwahed, W., Degobert, G., Stainmesse, S., Fessi, H. (2006) Freeze-drying of nanoparticles: Formulation, process and storage considerations. *Adv. Drug Deliver. Rev.* 58(15), 1688-1713.

- 45 Ford, A.W., Dawson, P.J. (1994) Effect of type of container, storage temperature and humidity on the biological activity of freeze-dried alkaline phosphatase. *Biologicals* **22**(2), 191-197.
- 46 Pikal, M.J., Dellerman, K., Roy, M.L. (1992) Formulation and stability of freeze-dried proteins: effects of moisture and oxygen on the stability of freeze-dried formulations of human growth hormone. *Dev. Biol. Stand.* **74**, 21-37.
- 47 De Beer, T.R.M., Alleso, M., Goethals, F., Coppens, A., Vander Heyden, Y., Lopez de Diego, H., Rantanen, J., Verpoort, F., Vervaet, C., Remon, J.P., Baeyens, W.R.G. (2007) Implementation of a process analytical technology system in a freeze-drying process using Raman spectroscopy for in-line process monitoring. *Anal. Chem.* **79**, 7992-8003.
- 48 De Beer, T.R.M., Vercruyssen, P., Burggraeve, A., Quinten, T., Ouyang, J., Zhang, X., Vervaet, C., Remon, J.P., Baeyens, W.R.G. (2009) In-line and real-time process monitoring of a freeze-drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools. *J. Pharm. Sci.* **98**, 3430-3444.
- 49 Mattes, R., Root, D., Zhou, Y., Xiao, C., Johnson, J., Shukla, A. (2009) NIR prediction of solid dosage form dissolution profiles. *NIR News* **20**(5), 10-11.
- 50 Guidance for Industry. PAT – a framework for innovative pharmaceutical development, manufacturing, and quality assurance(2004), *U.S. Food and Drug Administration (FDA)*, Silver Spring, MD.
- 51 Pomerantsev, A.L., Rodionova, O.Y. (2012) Process analytical technology: A critical view of the chemometricians. *J. Chemometr.* **26**, 299-310.
- 52 Q2(R1) - Validation of analytical procedures: Text and methodology (1996) [en línea]. International conference on harmonisation of technical requirements for registration of pharmaceuticals. [Consultado en agosto de 2014]. Disponible en:
http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf.
- 53 European Pharmacopoeia 8.0 (2014). *EDQM Council of Europe*, Estrasburgo.

2. ESPECTROSCOPIA E IMAGEN QUÍMICA NIR

2.1. Fundamentos de la absorción NIR	67
2.2. Espectroscopia NIR.....	73
2.2.1. Modos de adquisición	76
2.2.2. Instrumentación NIR	78
2.3. Imagen química NIR (NIR-CI)	87
2.3.1. Instrumentación NIR-CI.....	89
2.4. Aplicaciones industriales del NIR y el NIR-CI	94
2.4.1. Aplicaciones NIR en la industria petroquímica.....	95
2.4.2. Aplicaciones NIR y NIR-CI en la industria farmacéutica	98
REFERENCIAS	104

2. ESPECTROSCOPIA E IMAGEN QUÍMICA NIR

2.1. Fundamentos de la absorción NIR

La radiación del infrarrojo cercano (NIR, *Near-InfraRed*) fue descubierta en 1800 por William Herschel [1], en su afán por descubrir el color asociado al calor emitido por la luz del sol. Al descomponer la luz blanca, observó que el aumento de temperatura se producía en el color rojo, justo al final del espectro visible, y que seguía más allá de este (infrarrojo), en forma de radiación invisible capaz de transportar calor. Años después, se demostraría que la única diferencia entre la luz y la radiación infrarroja es su longitud de onda y ambas serían incluidas en lo que hoy se conoce como espectro electromagnético (Figura 2.1).

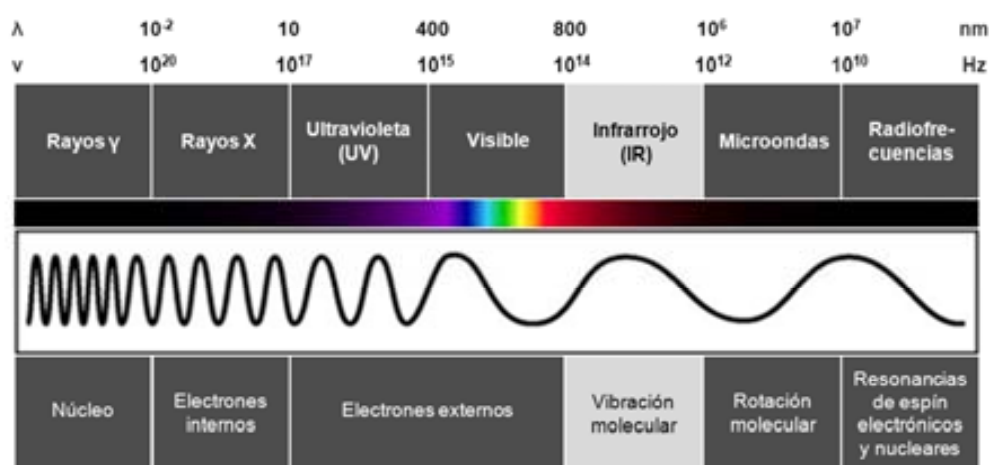


Figura 2.1. Espectro electromagnético, con los fenómenos asociados a cada región. Destacada en gris, la zona del infrarrojo [2].

El infrarrojo (IR, *InfraRed*) es la región del espectro electromagnético comprendida entre los 780 y los 10^6 nm ($12800-10$ cm^{-1} , en número de onda), *i. e.*, entre el visible y las microondas en el espectro electromagnético. La radiación infrarroja no es lo suficientemente energética como para provocar transiciones electrónicas como sucede en el ultravioleta-visible (UV-Vis) o los rayos X, por lo que las bandas que aparecen en esta zona corresponden principalmente a cambios en los estados vibracionales y rotacionales de la molécula (Figura 2.2) [3,4].

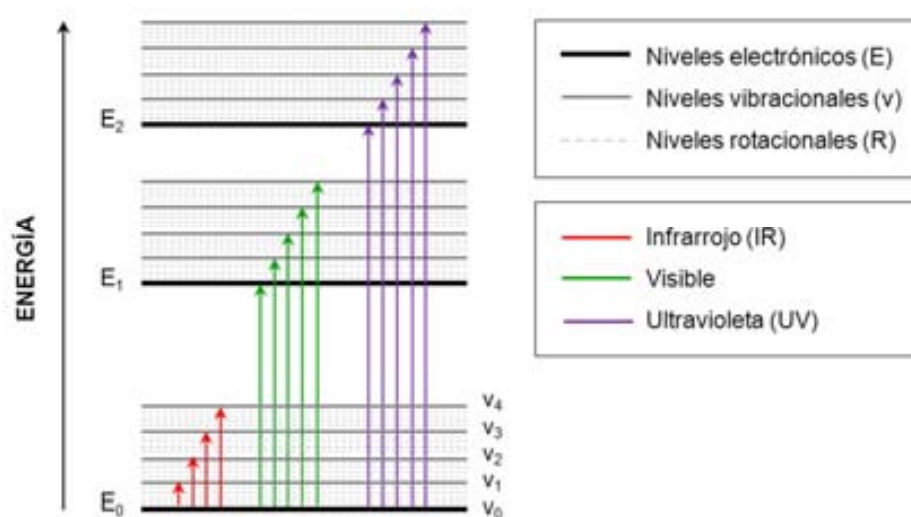


Figura 2.2. Diagrama de los niveles de energía de un átomo o molécula. Trazadas en colores, se observan distintas transiciones que pueden producirse por absorción de radiación IR, visible y UV. Para el IR (en rojo) sólo se muestran las transiciones puramente vibracionales, entre las que se distinguen (de izquierda a derecha): banda fundamental, 1º, 2º y 3er sobretono. Las bandas rotacionales no se muestran en la figura [4].

En función de la energía de la radiación, las características de su interacción con la materia y los requerimientos instrumentales, se pueden distinguir tres subzonas en la zona IR (de mayor a menor energía): infrarrojo cercano (NIR, *Near-InfraRed*), infrarrojo medio (MIR, *Mid-InfraRed*) e infrarrojo lejano (FIR, *Far-InfraRed*), las características de las cuáles se describen en la Tabla 2.1.

Tabla 2.1. División de la región del infrarrojo [2,3].

	Región	Intervalo de long. onda (nm)	Intervalo de n° onda (cm ⁻¹)	Origen de la absorción (tipo de transición)
+	NIR	780 - 2500	12800 - 4000	Sobretonos y bandas de combinación de las vibraciones moleculares fundamentales (MIR)
-	MIR	2500 - 50000	4000 - 200/400 ^a	Vibraciones y rotaciones moleculares fundamentales

FIR	50000 - 10 ⁶	200/400 ^a - 10	Rotaciones moleculares y vibración esqueleto
-----	-------------------------	---------------------------	--

^a *Existen discrepancias en los valores límite entre las zona MIR y FIR.*

En la zona del FIR, las bandas de absorción corresponden principalmente a rotaciones moleculares. La absorción en el MIR, en cambio, es debida a transiciones entre los estados vibracionales de más baja energía, es decir, entre el nivel vibracional fundamental ($v=0$) y el primer nivel vibracional excitado ($v=1$), lo que corresponde a la banda fundamental en la Figura 2.2. Estas bandas son las más probables y por ende, las más intensas de la zona infrarroja [5]. Las transiciones rotacionales se dan simultáneamente con las vibracionales en esta zona, pero son muy restringidas en sólidos y líquidos, por lo que suelen aparecer solapadas con las bandas vibracionales [3]. En el caso del NIR, las bandas de absorción corresponden a los sobretonos (Figura 2.2) y las bandas de combinación de vibraciones fundamentales MIR. Los primeros implican transiciones entre el nivel vibracional fundamental ($v=0$) y el segundo estado excitado o superiores ($v=\pm 2, \pm 3, \dots$), mientras que las segundas se dan cuando la radiación excita simultáneamente más de un modo de vibración en la molécula, es decir, cuando son dos enlaces los que absorben un cuanto de energía.

Para que haya absorción infrarroja, es necesario que una molécula experimente un cambio neto en su momento dipolar como consecuencia de un movimiento de vibración o rotación. En estas circunstancias, el campo eléctrico de la radiación puede interactuar con la molécula. Si la frecuencia de la radiación coincide con la frecuencia natural de vibración/rotación de la molécula, tiene lugar una transferencia neta de energía que aumenta la amplitud del movimiento, dando lugar a la absorción de la radiación [3].

El momento dipolar de un enlace depende en gran parte de la magnitud de la diferencia de carga entre los átomos que lo forman [3]. Es por este motivo que las especies homonucleares (N_2 , O_2 , H_2 ...) no absorben en el infrarrojo y que enlaces como $C=O$, $C-C$, $C-Cl$ o $C-F$ se observan de forma muy débil. Por el contrario, las

bandas más intensas en esta región vienen dadas por grupos funcionales con enlaces heteronucleares X-H, presentes en la mayoría de compuestos orgánicos, donde X= O, N, C y S [6]. Por esta razón, la espectroscopia NIR, además de permitir la cuantificación de los diferentes componentes de una mezcla (información química), es sensible a interacciones intermoleculares, como los puentes de hidrogeno, que proporcionan información sobre la estructura de la muestra (información física).

Otro factor que afecta el momento dipolar es la distancia entre los átomos que forman el enlace [3], un factor que se altera constantemente debido a las fluctuaciones moleculares normales. Las vibraciones moleculares, en particular, afectan a esta distancia, ya se trate de vibraciones de tensión (*stretching*), debidas a cambios en la longitud de los enlaces o vibraciones de flexión (*bending*), debidas a cambios en el ángulo entre enlaces [2,3]. A pesar de que la distancia entre los átomos del enlace se ve afectada en todos los casos, no todos los modos de vibración son activos en el infrarrojo para todas las moléculas, sino que esto depende de la estructura de la molécula en cuestión [2]. Sólo en aquellos casos donde la vibración implica un cambio en el momento dipolar de la molécula se da la absorción infrarroja.

Se considera que la vibración de enlaces moleculares diatómicos sigue el modelo del oscilador armónico (ley de Hooke), donde las características de la vibración se aproximan a las que seguirían dos masas unidas por un resorte. Cuando las masas se encuentran en reposo en la posición de equilibrio, su energía potencial es cero, mientras que cuando el resorte se comprime o se extiende, su energía potencial aumenta tal como se observa en el gráfico de la Figura 2.3 [3]. En este modelo, los distintos niveles energéticos son equidistantes y la transición sólo se permite entre niveles de energía vecinos ($\Delta v = \pm 1$), lo que se conoce como regla de selección [3].

Esto explica el fenómeno de absorción en el MIR, debido a las vibraciones fundamentales del enlace [2], pero no tiene en cuenta ciertas fuerzas de atracción y repulsión electrostática entre las moléculas ni la posible disociación de los enlaces [5,7]. Para explicar el origen de la absorción NIR se aplica la teoría del oscilador anarmónico. Como se observa en la Figura 2.3, las curvas armónica y anarmónica son muy similares a energías potenciales bajas, por lo que el comportamiento de los

enlaces es armónico alrededor de la posición de equilibrio (banda fundamental), pero pierde la simetría al alejarse de esta [3]. De este modo, los niveles de energía dejan de ser equidistantes: a niveles de energía más elevados, la diferencia energética entre niveles consecutivos es menor. Además, las bandas de vibración no siguen la regla de selección, sino que el incremento del número cuántico vibracional adopta valores $|\Delta v| > 1$, normalmente $\Delta v = \pm 2$ o ± 3 (primer y segundo sobretono, respectivamente). Los sobretonos aparecen en la región NIR a longitudes de onda entre los 780 y los 2000nm según su orden y la naturaleza del enlace en cuestión (Figura 2.4) [7]. Al tratarse de transiciones menos frecuentes que las vibraciones fundamentales del MIR, su intensidad es menor: entre 10-100 veces menor para el primer sobretono, dependiendo del enlace en cuestión [6,7] y su probabilidad decrece a medida que aumenta el número cuántico vibracional (v) [8].

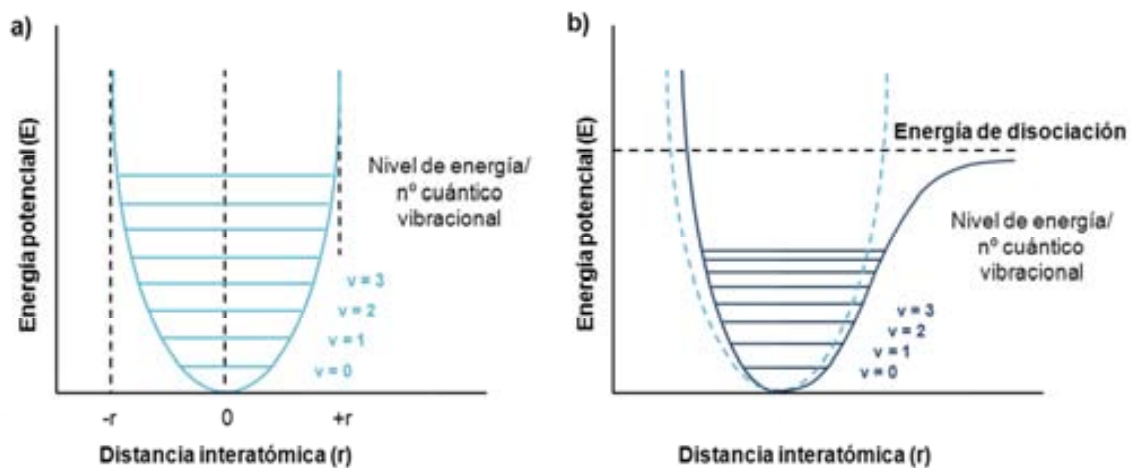


Figura 2.3. Diagrama de energía de los modelos vibracionales de las moléculas: (a) modelo del oscilador armónico y (b) modelo del oscilador anarmónico.

En el caso de las moléculas poliatómicas, el espectro vibracional adquiere mayor complejidad debido al mayor número de átomos y enlaces, y pueden producirse bandas de combinación (descritas arriba). Estas bandas, adoptan valores equivalentes a combinaciones lineales de las frecuencias de las bandas fundamentales que interaccionan (suma o diferencia de múltiples de las frecuencias interaccionantes). Su intensidad también es baja y suelen aparecer entre los 1900 y los 2500nm (Figura 2.4) [7,8].

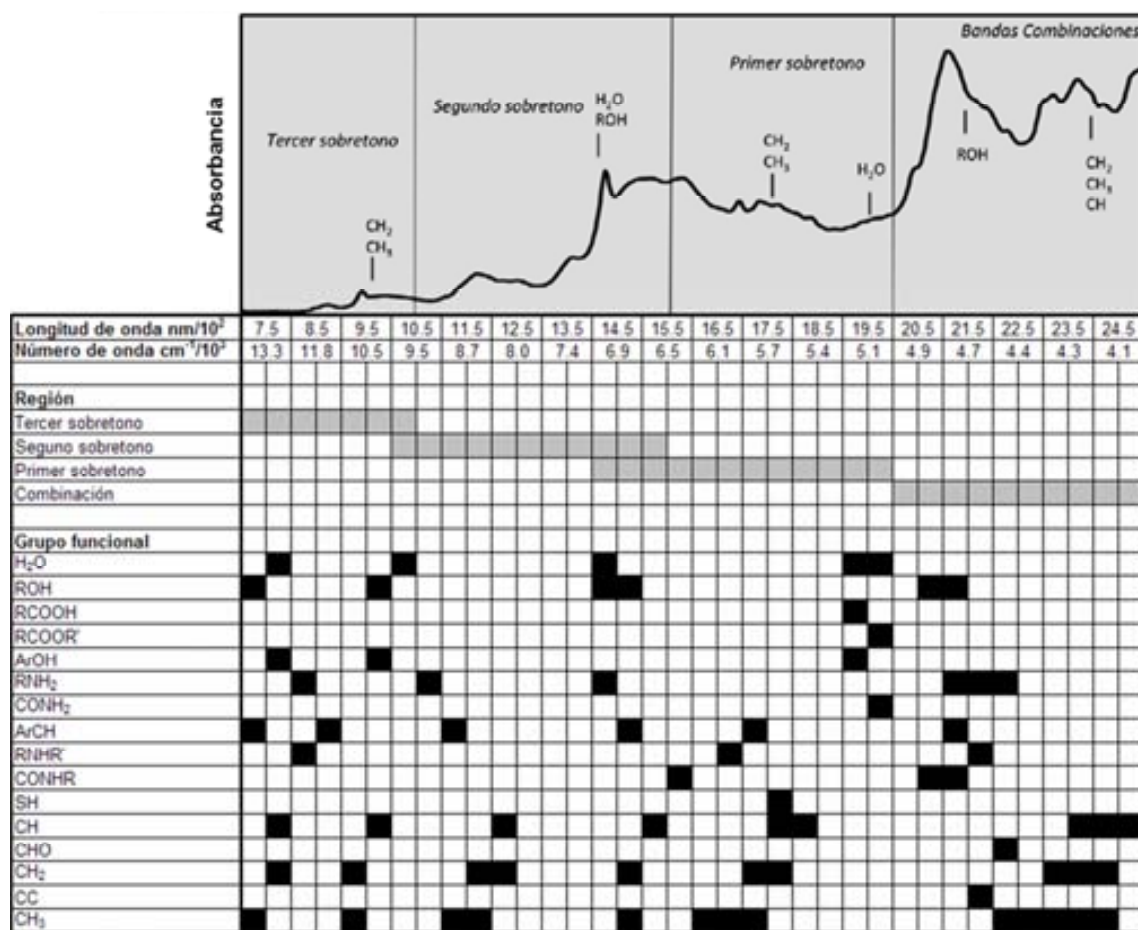


Figura 2.4. Asignación de bandas de grupos funcionales en la región NIR [11,12,13].

2.2. Espectroscopia NIR

Dado que las señales en el infrarrojo implican intercambio de energía entre la radiación electromagnética y la materia, el infrarrojo es considerado como una técnica espectroscópica [2]. Existen numerosas técnicas espectroscópicas. En todos los casos se trata de métodos rápidos y la mayoría son ampliamente utilizados en los laboratorios de análisis de rutina, pero cada una de ellas presenta características diferenciales con las demás. Dependiendo de la zona del espectro electromagnético, la radiación tiene un nivel de energía concreto, que determina la naturaleza de su interacción con la materia.

Las espectroscopias asociadas a la zona infrarroja del espectro electromagnético se conocen como espectroscopias vibracionales, dado el fenómeno físico al que están asociadas, y moleculares, dado que las vibraciones implican más de un átomo. Este grupo de espectroscopias incluye distintas técnicas, que destacan por su rapidez y versatilidad, ya que permiten el análisis de materiales en cualquier estado (sólido, líquido y gas) y presentan una gran cantidad de aplicaciones tanto en el laboratorio, como en la industria, donde unidas a la fibra óptica, permiten cada vez una mejor integración en los procesos de producción. De entre las técnicas de espectroscopia vibracional destacan el Raman, una técnica muy reciente basada en la dispersión de la radiación, y el MIR y el NIR, como técnicas establecidas, ampliamente utilizadas en aplicaciones tanto cualitativas como cuantitativas. Mientras que el MIR ha sido la espectroscopia vibracional más usada a lo largo del tiempo [3], el NIR, a pesar de su temprano inicio, ha ido ganando popularidad mucho más lentamente [14].

En 1881 (80 años después del descubrimiento del NIR), Abney y Festing registraron los primeros espectros NIR de líquidos orgánicos mediante placas fotográficas sensibles a la radiación NIR (1000-1200 nm) [15]. En 1905, Coblentz publicaba cientos de espectros de distintos compuestos orgánicos, con los que consiguió caracterizar un gran número de grupos moleculares y concluyó que el espectro NIR es único para cada compuesto [16]. Pero a pesar de las buenas perspectivas de la técnica, esta quedó relegada debido a la complejidad de los espectros NIR y la dificultad de interpretación, comparados con los del MIR o el UV-Vis.

INTRODUCCIÓN

Como se ha comentado en el apartado anterior, los fenómenos asociados a la absorción NIR (sobretonos y bandas de combinación), son fenómenos poco probables. Es por este motivo, que la absorción en esta zona es débil y por consiguiente, las bandas correspondientes son poco intensas. Además, los solapamientos entre sobretonos y bandas de combinación dan lugar a bandas anchas y poco definidas. Estas características limitan la sensibilidad y la selectividad del NIR y también restringieron su aplicabilidad durante décadas. Fue en los años 60, cuando un grupo de investigación del *United States Department of Agriculture* (USDA), liderado por Karl Norris, promovió realmente la utilización de la técnica NIR. Otro impulsor de esta técnica es Phil Williams que, en los años 70, introdujo los análisis NIR a gran escala en la *Canadian Grain Commission*. A raíz de sus estudios cuantitativos en el ámbito agroalimentario [17,18,19], la aceptación y el interés por el NIR crecieron considerablemente y, desde entonces, la técnica ha ido evolucionando rápidamente.

La mejora de los equipos NIR, junto con el avance de los sistemas informáticos capaces de procesar la información de los espectros NIR, así como la aparición del análisis de datos y los métodos de calibración fueron mejorando poco a poco las posibilidades de esta técnica [12]. La complejidad de los espectros se compensa con la aplicación de un tratamiento quimiométrico adecuado. Esto permite extraer la información espectral relevante que relaciona los datos espectroscópicos con el parámetro de interés, pudiendo ser este analizado cualitativa o cuantitativamente. Asimismo, el bajo coeficiente de absorción del NIR supone ahora una ventaja sobre el resto de espectroscopias vibracionales. En primer lugar, permite trabajar con muestras muy gruesas o concentradas, sin tener que diluir o disminuir el camino óptico. Además, aumenta la penetrabilidad de la radiación en la muestra, que puede llegar a ser de varios milímetros [10]. De este modo, se expone más material, reduciéndose los errores de muestreo y los problemas de heterogeneidad de la muestra, disminuyendo el efecto de posibles interferencias, como la contaminación de superficies y posibilitando el registro de las muestras a través de envases, con una contribución mínima [10,20].

A pesar de las desventajas que presenta el NIR (Tabla 2.2), su uso se ha ido expandiendo a muchos campos hasta convertirse en la técnica analítica de elección

para muchos sectores (industria química, petroquímica, agroalimentaria, farmacéutica, etc.). Esto responde al avance de las técnicas quimiométricas de procesamiento de datos y la mejora de los equipos NIR [12]. Pero el creciente interés en la técnica NIR se debe en gran parte a las ventajas que ofrece sobre otras técnicas analíticas [7] (Tabla 2.2).

Tabla 2.2. Ventajas y desventajas de la espectroscopia NIR.

Ventajas	Desventajas
<p>Es una técnica no invasiva, ni destructiva</p>	<p>Es una técnica poco sensible por lo que la cuantificación de componentes muy minoritarios puede presentar dificultades. A pesar de ello, el NIR se ha usado en componentes con concentraciones por debajo del 0.1% [21,22].</p>
<p>Es muy versátil, ya que permite registrar el espectro de muestras en cualquier estado (sólido, líquido y gaseoso, aunque las aplicaciones a gases son muy limitadas).</p>	<p>Presenta una baja selectividad, por lo que se requiere tratamiento quimiométrico posterior para modelar los datos y extraer la información relevante, un proceso que requiere tiempo y un gran número de muestras para incorporar todas las fuentes de variabilidad.</p>
<p>El pretratamiento de la muestra es mínimo o inexistente en la mayoría de los casos. Esto reduce el tiempo de análisis, así como el uso de reactivos, lo que disminuye tanto el coste como el impacto del análisis sobre el medio ambiente. Además, dado que la manipulación de la muestra y el uso de reactivos son mínimos, minimizan los errores humanos durante los análisis y se mejora la seguridad de los analistas.</p>	<p>Se trata de una metodología relativa por lo que requiere un conocimiento de los valores de referencia del parámetro a cuantificar.</p>
<p>Los espectros NIR contienen información tanto química, como física de la muestra. Se puede determinar simultáneamente más de un analito, así como ciertas propiedades físicas con un solo espectro.</p>	
<p>El registro de los espectros es rápido y permite obtener información de la muestra en tiempo real.</p>	
<p>La fibra óptica permite realizar los análisis en la misma cadena de producción (<i>on-/in-line</i>).</p>	
<p>El coste de los análisis es relativamente bajo. En el mercado se pueden adquirir equipos sencillos de sobremesa por menos de 3000€.</p>	

Estas ventajas posicionan el NIR como una herramienta clave en el control de calidad industrial y el monitoreo de parámetros físicoquímicos durante los procesos de producción, lo que supone la sustitución gradual de técnicas analíticas lentas, destructivas e invasivas (cromatografía de gases y líquidos) y técnicas de control de procesos no específicas (temperatura, presión, pH...).

2.2.1. Modos de adquisición

Existen tres sistemas de registro del espectro NIR: transmitancia, reflectancia y un tercer modo, intermedio de los dos anteriores, denominado transflectancia (Figura 2.5). El modo apropiado de registro en cada caso viene dictado por las propiedades de la muestra. Para el análisis de gases y líquidos se ha utilizado clásicamente el modo de transmitancia [23], aunque el desarrollo de detectores de mayor sensibilidad ha posibilitado su aplicación al análisis de muestras sólidas [24]. Sin embargo, el modo más habitual para el registro de muestras sólidas y en polvo es la reflectancia, mientras que la transflectancia, se utiliza normalmente en líquidos y semilíquidos [25].

▪ Modo transmitancia

En el modo de transmitancia o transmisión se trabaja con la fracción de radiación transmitida (no absorbida) por la muestra, es decir, la radiación que llega al detector después de atravesar la muestra. La absorción (A) de la radiación en este caso sigue la ley de Lambert-Beer:

$$A = \log \frac{1}{T} = \log \frac{I_0}{I} = abc \quad \text{Ecuación 2.1}$$

donde, T representa la transmitancia, I_0 la intensidad de la radiación incidente, I la intensidad de la radiación transmitida, a la absortividad molar, b el camino óptico y c la concentración de analitos absorbentes. Pueden producirse desviaciones de esta ley debido a causas diversas como concentraciones elevadas, que aumentan la absorción

de la radiación, pero también la reflexión y la dispersión de la radiación en distintas direcciones, lo que se conoce como transmitancia difusa.

▪ Modo reflectancia

La reflectancia estudia la radiación reflejada por la muestra irradiada, de la cual se pueden distinguir dos componentes: la especular y la difusa. La primera, regida por las leyes de Fresnel, da nombre a la radiación que se refleja en el mismo ángulo de incidencia sin penetrar en la muestra, por lo que no contiene ninguna información sobre ella. La contribución de la reflectancia especular (únicamente en forma de ruido) suele minimizarse con una posición apropiada de los detectores respecto a la muestra y la fuente de radiación. La componente difusa en cambio, es el resultado de la absorción parcial de la radiación por la muestra y su dispersión en todas las direcciones (Figura 2.5). La radiación que ha superado la primera capa de partículas difunde a través de reflexiones, refracciones, dispersiones o nuevas absorciones, de manera que puede interactuar con la muestra, recogiendo información en diferentes puntos y profundidades. La señal analítica medida en este caso se expresa en forma de absorbancia aparente (A_{ap}):

$$A_{ap} = \log \frac{1}{R} = \log \frac{R_{ref}}{R_m} = a'c \quad \text{Ecuación 2.2}$$

donde R es la reflectancia relativa, R_{ref} la reflectancia de referencia, R_m la reflectancia de referencia, a' la constante de proporcionalidad y c la concentración de analitos absorbentes.

▪ Modo transreflectancia

En el registro por transreflectancia, la muestra, se coloca en una cubeta junto con el reflector. Parte de la radiación incidente es reflejada por la muestra, mientras que el resto la atraviesa hasta llegar al reflector, que marca el final del camino óptico, reflejando la luz restante hacia el detector.

En todos los casos, lo que se mide es la absorción de la radiación por la muestra. Dado que la región NIR se caracteriza por una baja absorptividad, la obtención de información en la mayoría de los casos se debe a la mayor penetración de la radiación en la muestra. Esta baja absorptividad suele combinarse con un alto efecto de dispersión de la radiación al interactuar con la materia, de forma que los fotones incidentes siguen un camino impredecible a través de la muestra gobernado por la reflexión y la dispersión [26]. En esta situación, es cuando se habla de reflectancia y/o transmitancia difusas (Figura 2.5) [10]. Esta dualidad absorción-dispersión, intrínseca de la señal NIR implica la presencia de información tanto química (composición) como física (tamaño y morfología de las partículas, densidad, índice de refracción [27,28,29]) de la muestra [8,24]. Este hecho puede ser ventajoso en algunos casos, ya que permite el análisis de parámetros de ambos tipos en una sola medida, pero, a la vez, los efectos de la dispersión son considerados como interferencias cuando sólo se requiere la información química [8].

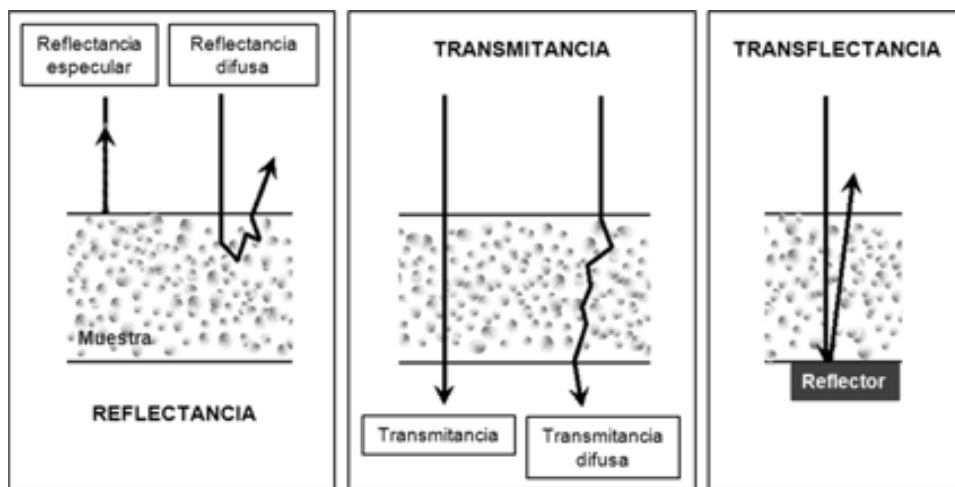


Figura 2.5. Esquema del recorrido de la radiación NIR en los distintos modos de registro.

2.2.2. Instrumentación NIR

En los años 40, la espectrofotometría empezó a ganar importancia como técnica analítica, lo que propició la proliferación de equipos comerciales [30]. En muchos de ellos, la región NIR ya se encontraba disponible, ofrecida como extensión a las longitudes de onda del UV-Vis o del MIR, pero su interés era mínimo debido a las

peculiares características de esta zona (apartado 2.2) que restringían su aplicabilidad [28,31]. La baja intensidad de las bandas en la zona NIR implica que cualquier cambio en una muestra se refleje de forma muy sutil en el espectro NIR. Es por este motivo, que el nivel de exigencia de los espectrofotómetros NIR, en términos de ruido permisible y estabilidad instrumental, debe ser mucho mayor que en otros instrumentos, sobre todo si se pretende aplicar con carácter cuantitativo [29]. Esto retrasó la aparición de los primeros espectrofotómetros NIR, la evolución de los cuales ha estado siempre ligada al avance tecnológico. No fue hasta los años 70 cuando las prestaciones de los equipos empezaron a mejorar: se desarrollaron nuevos sistemas de selección de longitudes de onda y detección, que permitían el registro de espectros NIR de forma rápida y reproducible. El avance de la informática y la quimiometría, la aparición de las sondas de fibra óptica y la miniaturización de los equipos también han sido factores clave en la evolución de la espectroscopia NIR.

Los espectrofotómetros NIR son esencialmente iguales que los equipos empleados en otras regiones del espectro electromagnético, compuestos por una fuente de radiación, un selector de longitudes de onda, un compartimento de la muestra y un detector. Pero además, pueden incorporar una gran variedad de accesorios dependiendo de las características de la muestra y las condiciones y necesidades analíticas particulares de cada caso, lo que hace del NIR una técnica muy flexible [7].

▪ Fuente de radiación

La fuente de radiación más empleada en los equipos NIR son las lámparas halógenas de tungsteno con ventana de cuarzo. Se trata de lámparas baratas y de larga duración (unos 12 meses de trabajo continuo), capaces de proporcionar espectros continuos en la región de 320 a 2500nm, cubriendo toda la zona NIR. Otra alternativa son los *Light Emitting Diodes* (LEDs), que a diferencia de las lámparas halógenas, son de espectro reducido y cubren rangos alrededor de los 50 nm [2,8]. Esto implica que sean útiles solamente en aplicaciones con analitos que absorben en regiones muy concretas.

El rango específico de longitudes de onda en el que emiten los LEDs varía dependiendo de su composición, pero nunca supera los 1650nm.

▪ Selector de longitudes de onda

Los únicos equipos que prescinden de este elemento son los **equipos basados en LEDs** como fuente de radiación [32]. Dado que los LEDs trabajan en rangos estrechos de longitudes de onda, estos instrumentos no requieren selector de longitudes de onda. Esto implica que no incorporan partes móviles en su mecanismo, por lo que son equipos directos y robustos, muy adecuados para uso en equipos portátiles [7].

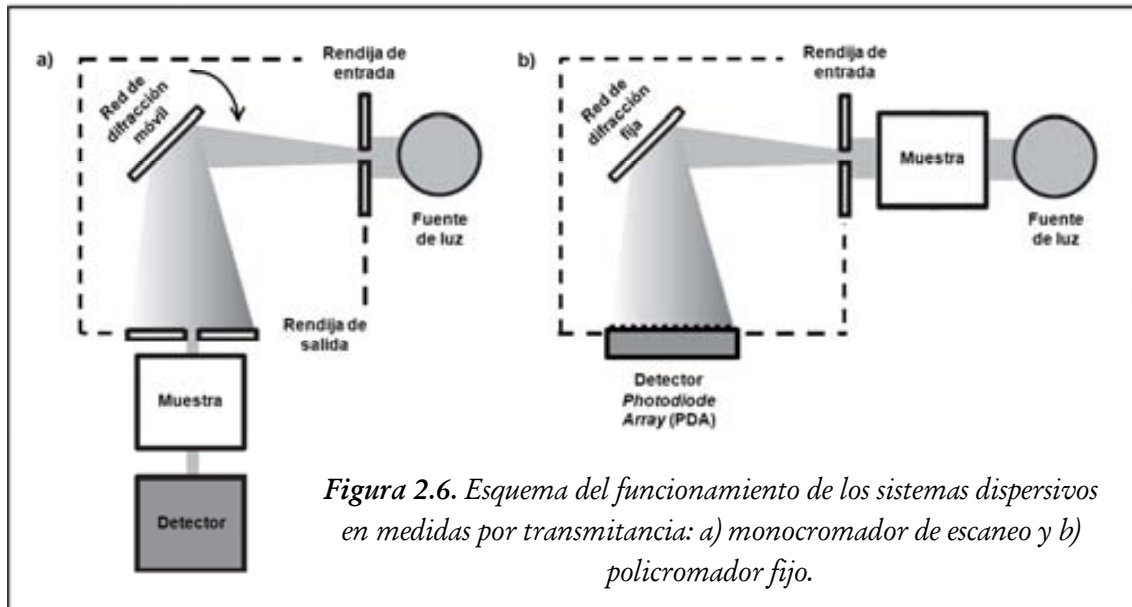
Exceptuando los espectrofotómetros de LEDs, en todos los demás casos, es necesario un sistema de selección de longitudes de onda. Según el sistema empleado, los equipos se pueden clasificar en dispersivos y no dispersivos.

EQUIPOS DISPERSIVOS

Los sistemas dispersivos o de barrido se basan en los monocromadores como selectores de longitudes de onda, que descomponen la luz policromática en longitudes de onda discretas. La radiación proveniente de la muestra entra al monocromador por una rendija en forma de haz estrecho y es dividida en las distintas longitudes de onda mediante un elemento dispersante. Los elementos dispersantes de la radiación más utilizados actualmente en espectroscopia NIR son las redes de difracción, que gracias a sus estrías difractan cada longitud de onda en un ángulo diferente.

En el caso de los equipos con **monocromadores de escaneo** [2,33], el barrido de longitudes de onda discretas se realiza moviendo la red mediante un motor, de manera que en cada momento una longitud de onda discreta es enfocada sobre la rendija de salida del monocromador (Figura 2.6a). Esto implica que para completar un barrido se requiera un tiempo elevado. La resolución espectral en este tipo de equipos viene determinada por el tamaño de las rendijas de entrada y salida del

monocromador. Si se reduce su tamaño, mejora la resolución, pero disminuye la relación señal/ruido. En general, se trata de instrumentos baratos y robustos, con buena relación señal/ruido, pero su velocidad y precisión vienen limitadas por la calidad del sistema mecánico de selección de longitud de onda [24,31].



La extensión de los equipos de los monocromadores de escaneo son los **policromadores con red de difracción fija** [10,30], donde la rendija de salida del monocromador se sustituye por un *Photodiode Array* (PDA) (Figura 2.6b). Cada longitud de onda dispersada por la red de difracción llega a uno de los elementos detectores del PDA, pudiéndose medir simultáneamente distintas longitudes de onda por distintos canales. La resolución espectral y el rango de longitudes de onda dependen del número de elementos detectores del PDA. La velocidad de respuesta de estos equipos es mucho mayor que en los anteriores, aunque también es más difícil asegurar su precisión y evitar el calentamiento de la muestra [31]. Los policromadores con PDA precinden de partes móviles por lo que son robustos y tolerantes a las vibraciones y su tamaño también puede ser más reducido.

EQUIPOS NO DISPERSIVOS

Los equipos no dispersivos más sencillos son los **equipos de filtros** [2,28]. En este caso, la selección de longitudes de onda se lleva a cabo interponiendo materiales semitransparentes entre la fuente de radiación y la muestra, permitiendo únicamente el paso de determinadas longitudes de onda. Los filtros más sencillos son los filtros de absorción, en los que las longitudes de onda transmitidas dependen del material del que están hechos. Otro tipo de filtros usados en espectroscopia NIR son los filtros de interferencia (o de Fabry-Perot), en los que las longitudes de onda transmitidas dependen del índice de refracción del material y del grosor del filtro.

Los espectrofotómetros equipados con estos dispositivos suelen ser robustos y de bajo coste y se utilizan para aplicaciones específicas en las que los analitos a determinar absorben en intervalos espectrales determinados. Una alternativa puede ser el montaje de una rueda giratoria con varios filtros (Figura 2.7a), lo que permite realizar medidas en distintas zonas del espectro [28,30].

Otro tipo de equipos no dispersivos son los basados en **filtros optoacústicos** (AOTF, *Acousto-Optical Tuneable Filter*) [34,35]. Estos equipos aprovechan las propiedades de los materiales birrefringentes (normalmente cristales de TeO_2), que cambian su índice de refracción cuando son atravesados por una onda acústica. De esta forma, se consigue que una de las longitudes de onda de la luz policromática incidente sea difractada por el material y dirigida hacia la muestra, mientras que el resto atraviesan el cristal de TeO_2 , transparente a la radiación NIR (Figura 2.7b). Las ondas acústicas se consiguen por transformación de señales de radiofrecuencia mediante transductores piezoeléctricos, de manera que la selección de longitudes de onda se realiza modulando la señal de radiofrecuencia inicial. Esto implica que el escaneo de longitudes de onda es muy rápido y permite seleccionar sólo aquellas longitudes de onda necesarias en cada análisis. Además, no requiere de ninguna parte móvil, por lo que se trata de una técnica muy robusta y de elevada repetibilidad, especialmente adecuada para trabajar en condiciones agresivas, como son las plantas de industriales [10].

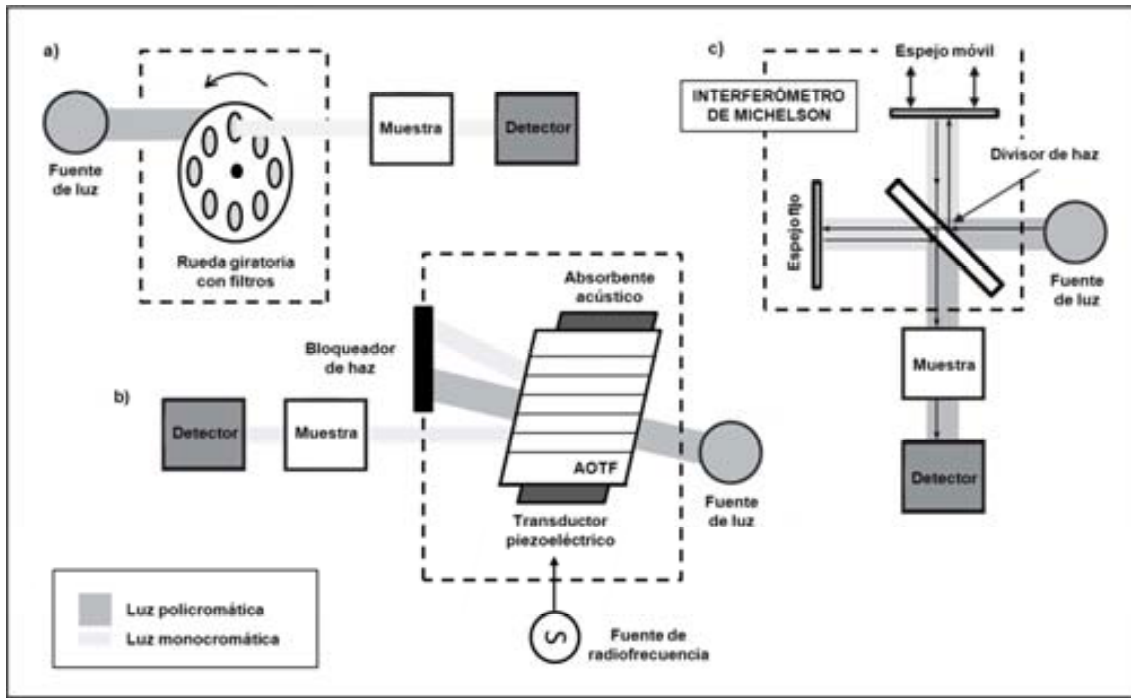


Figura 2.7. Esquema básico del funcionamiento de: a) un equipo de filtros, b) un equipo AOTF y c) un FT-NIR, todos ellos en modo transmitancia.

Por su parte, los **equipos de transformada de Fourier (FT, *Fourier transform*)** [2,36] se basan en la división de la luz incidente en dos haces desfasados recombinándolos posteriormente (Figura 2.7c). El dispositivo más comúnmente utilizado para conseguir la división y el desfase de los haces es el interferómetro de Michelson que consta de un divisor de haz, un espejo fijo y un espejo móvil. El divisor de haz divide la radiación policromática en dos haces, uno de los cuales se refleja en el espejo fijo y el otro en el espejo móvil. El cambio de posición del espejo móvil provoca que ambos haces estén desfasados como consecuencia de la diferencia de las trayectorias ópticas y que, al combinarse de nuevo, puedan interferir constructiva o destructivamente. La señal obtenida se denomina interferograma y es una representación de la intensidad de la señal en función del tiempo. Con la aplicación del algoritmo de FT, el dominio del tiempo se convierte en el de la frecuencia, obteniendo el espectro NIR. La principal ventaja de los FT-NIR es que son equipos muy precisos y reproducibles debido al uso de láseres HeNe como referencia interna, pero son sensibles a vibraciones [5,31].

ELECCIÓN DEL SISTEMA

Como se ha visto, existe un gran número de configuraciones de selección de longitud de onda, cada una de ellas, con características diferenciales. La selección del método apropiado se basa en las características de la aplicación: naturaleza de la muestra, sensibilidad requerida, estabilidad, rapidez,... [8]

Los equipos que no requieren de partes móviles, como los de LEDs o los de filtros, son compactos y fácilmente miniaturizables, lo que los hace ideales para medidas de campo [2]. Pero, a la vez, sus aplicaciones son mucho más específicas y restringidas que en el caso de los instrumentos de espectro entero (equipos dispersivos y FT-NIR), la aplicabilidad de los cuales es mucho más flexible. Esta flexibilidad es muy útil para equipos dirigidos al trabajo *off-line* y *at-line*, que pueden ser utilizados en gran variedad de análisis, ya sea en control de calidad, investigación o laboratorios en planta. En este tipo de medidas, la rapidez no es caudal, sino que son más importantes factores como la sensibilidad, la resolución y la reproducibilidad de los equipos [8,24]. La demanda de resolución y reproducibilidad elevadas, limitadas en los instrumentos con monocromadores de escaneo, es una de las razones principales por las que estos espectrofotómetros han sido desplazados por los de transformada de Fourier [2].

Para los análisis en proceso, en cambio, se precisa de medidas *in-line* y *on-line* para la obtención de información en tiempo real y en condiciones de trabajo duras. Es por ello que se requieren instrumentos rápidos, robustos y estables a las variaciones ambientales, como los AOTF o los policromadores basados en PDA [8,24]. Estos equipos posibilitan un gran número de adquisiciones por segundo y son poco sensibles a las vibraciones debido a la ausencia de partes móviles en su mecanismo.

▪ Compartimento de la muestra

La espectroscopia NIR tiene una gran versatilidad y adaptabilidad para el análisis de muestras de diversa naturaleza (sólida, líquida o gaseosa) sin necesidad de

pretratamiento. Por este motivo, el compartimiento de la muestra es la parte del espectrofotómetro NIR que más variaciones presenta.

Como se ha comentado anteriormente, el estado y las propiedades ópticas de la muestra determinan el modo de registro más apropiado en cada caso. A nivel de instrumentación, la diferencia básica entre los distintos modos de registro (reflectancia, transmitancia o transflectancia), es la posición de la muestra en el instrumento y el camino óptico (Figura 2.8). Pero la naturaleza de la muestra puede tener muchas otras implicaciones en lo que al compartimiento de la muestra se refiere. Cuando se trabaja con muestras líquidas o en polvo es necesario el uso de cubetas, normalmente de cristal o cuarzo, que son materiales transparentes a la radiación NIR. En algunos casos las muestras deben ser calentadas o enfriadas mientras se realiza el registro del espectro, ya que la temperatura afecta el espectro NIR [24]. Para solucionar estas situaciones se han diseñado accesorios que pueden ser acoplados a los espectrofotómetros de manera que se pueda llevar a cabo el registro con la mínima manipulación de la muestra. Otro tipo de muestras son aquellas que presentan formas y/o tamaños que imposibilitan el registro en compartimentos convencionales dentro de un espectrofotómetro o muestras que no pueden ser transportadas. En estos casos el registro se lleva a cabo con instrumentación portátil, muy útil para las medidas de campo, o mediante la aplicación de fibra óptica, una estrategia utilizada también para medidas en continuo *on-* e *in-line*, en la misma línea de producción.

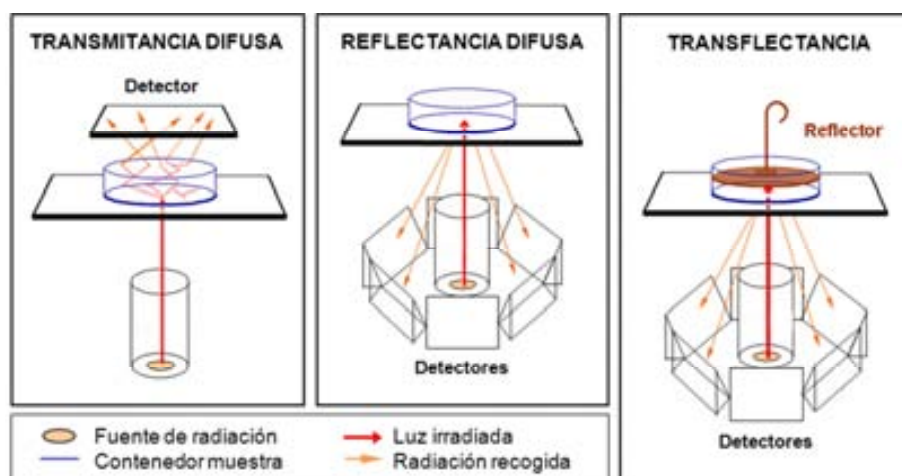


Figura 2.8. Esquema de la situación de la muestra y recorrido de la radiación NIR en equipos con distintos modos de registro (adaptación [37]).

▪ **Detector**

Por lo que respecta a los detectores, los más comunes son los basados en semiconductores como Si, PbS, PbSe, InSb o InGaAs. Su funcionamiento se basa en el hecho que la energía de los fotones que llega al semiconductor promociona ciertos electrones a estados libres, incrementando la conducción eléctrica del material [2]. Los detectores de Si destacan por su rapidez, tamaño reducido y bajo ruido, pero su sensibilidad sólo abarca hasta los 1100nm. Los de PbS y PbSe son los más populares ya que, a pesar de ser más lentos, cubren el rango entre 1100 y 2500nm, presentan buena relación señal/ruido y son baratos. El más caro es el InGaAs, que combina la rapidez y el tamaño reducido de los de Si, pero alcanza un rango de longitudes de onda similar al de los de PbS en su versión extendida [2,30]. Otro detector usado en la región NIR es el de Mercurio-Cadmio-Telurio (MCT), pero este requiere ser refrigerado con nitrógeno líquido para mantener los niveles de sensibilidad [2,34].

2.3. Imagen química NIR (NIR-CI)

La imagen química NIR (NIR-CI, *Near-Infrared Chemical Imaging*) es la combinación de la espectroscopia NIR convencional y el análisis por imagen digital. Los mismos fundamentos y características de la espectroscopia NIR descritos en los apartados 1.1 y 1.2 son aplicables al NIR-CI, donde se aprovechan para generar imágenes. Una imagen digital es la representación gráfica de un objeto o una escena real en dos dimensiones espaciales (M y N). Una característica de las imágenes es que requieren algún tipo de contraste para diferenciar las regiones de interés en un campo de visión. El contraste puede estar basado en cualquier calidad medible de la muestra que se pueda expresar en función del espacio y constituye la tercera dimensión de la imagen. Cuantas más variables se midan en cada píxel, considerando un píxel como la fracción más pequeña de la imagen que contiene información, mejor será el contraste observado. El ejemplo más conocido son las imágenes RGB (*Red Green Blue*), basadas en la mezcla de los colores rojo, verde y azul, es decir, en tres variables. Por consiguiente, las imágenes resultantes están constituidas por cubos de datos tridimensionales $\underline{\mathbf{X}}(M \times N \times 3)$. En el caso de las imágenes NIR, el contraste se consigue gracias a la espectroscopia NIR. En este caso, la información viene dada por la absorbancia de la muestra a todas las longitudes de onda irradiadas (ℓ), por lo que los cubos de datos resultantes son $\underline{\mathbf{X}}(M \times N \times \ell)$. Estos cubos de datos se conocen como **imágenes hiperespectrales** y pueden entenderse como:

- a) Una imagen de $M \times N$ píxeles, donde cada píxel contiene toda la información de un espectro completo a todas las longitudes de onda registradas (Figura 2.9a)
- b) Una serie de imágenes planas consecutivas de $M \times N$ píxeles, cada una registrada a una misma longitud de onda para todos los píxeles (Figura 2.9b).

La información contenida en la dimensión λ se representa en la imagen bidimensional según una escala de colores, de forma que el color de una zona de la imagen es el reflejo de la absorción media de todas las λ en la zona correspondiente de la muestra. Esta imagen en sí misma no desvela información química sobre la muestra, pero es posible conseguirla con la aplicación de tratamientos y algoritmos quimiométricos (capítulo 3). Las imágenes tratadas quimiométricamente se conocen

INTRODUCCIÓN

con el nombre de **imágenes químicas** porque de ellas se puede extraer información química (composición, concentración, humedad...) y, dado que se trata de NIR, también se obtiene información física (estructura, tamaño de partícula, densidad, viscosidad...) de la muestra. Todo ello, se determina a lo largo de toda la imagen, proporcionando información sobre la distribución espacial de cada propiedad en la superficie de la muestra.

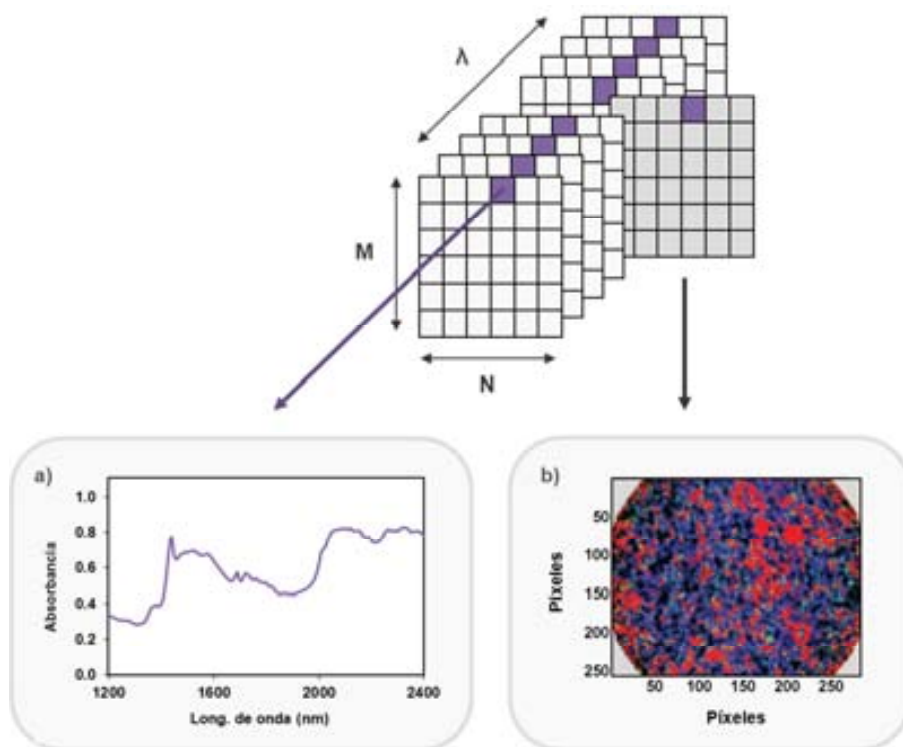


Figura 2.9. Relación de los datos espectrales y espaciales en el cubo de datos hiperespectral ($M \times N \times \lambda$): a) espectro asociado a un píxel y b) distribución espacial de la absorbancia a una longitud de onda concreta.

Con el NIR-CI se consigue aprovechar las ventajas que tiene el NIR como técnica analítica (apartado 2.2) y aplicarlas a toda la superficie de la muestra: mientras que la espectroscopia NIR convencional solamente proporciona un espectro promedio de la superficie de las muestras analizadas, el NIR-CI permite registrar un espectro completo por cada píxel de la imagen adquirida, obteniendo simultáneamente información espectral y espacial. Considerando una imagen espectral de tamaño convencional (320 x 256 píxeles), se obtiene un total de 81920 píxeles, i.e. 81920 espectres NIR con información física y química de la muestra, que multiplicado por

el número de longitudes de onda de registro (normalmente >100) son más de 8×10^6 datos. Esta capacidad de adquisición de datos, unida a la amplia variedad de herramientas para el tratamiento de las imágenes obtenidas, es la base del gran potencial de la técnica.

2.3.1. Instrumentación NIR-CI

Un equipo NIR-CI está diseñado para obtener la información tanto espectral (λ), como espacial ($M \times N$) de la muestra. La información espectral se consigue con los componentes básicos de un espectrofotómetro NIR (apartado 2.2.2): una fuente de luz, un sistema de selección de longitudes de onda, un compartimento de la muestra y un detector. Además de estos elementos, para la obtención de la información espacial se requiere de un sistema óptico configurable en el equipo, que permita seleccionar la resolución espacial de la imagen final [38,39], es decir, el tamaño de píxel.

En la Figura 2.10 se muestra un ejemplo de como se disponen estos elementos en un equipo NIR-CI.

De la misma forma que en los espectrómetros NIR, la fuente de luz suele tratarse de lámparas halógenas de tungsteno, que suelen ir acompañadas de filtros de polarización para minimizar los reflejos. Como selector de longitudes de onda suelen utilizarse los sistemas de FT, redes de difracción o filtros sintonizables: AOTF o *Liquid Crystal Tunable Filter* (LCTF). Con excepción de los LCTF, los mecanismos

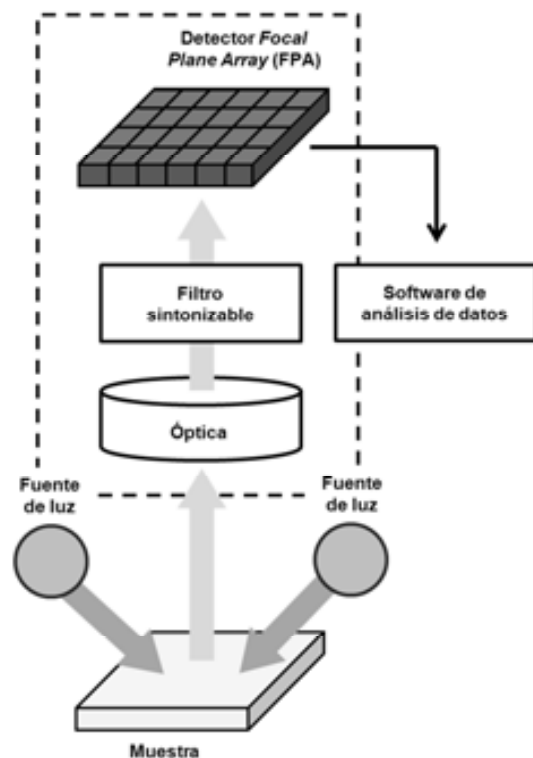


Figura 2.10. Esquema básico del funcionamiento de un equipo NIR-CI con detector FPA en modo reflectancia difusa.

del resto de sistemas se describen detalladamente en el apartado 2.2.2. Los LCTF son filtros sintonizables que, a diferencia de los AOTF, utilizan elementos de cristal líquido controlados electrónicamente para transmitir solamente ciertas longitudes de onda. Por lo que respecta al resto de elementos del espectrofotómetro, los mismos principios descritos para el compartimento de la muestra y semiconductores usados como detectores en NIRS son aplicables a los equipos NIR-CI. Sin embargo, en ambos casos su configuración puede variar en función de la forma de adquisición de la información espacial de la imagen hiperespectral. La información espacial se puede obtener utilizando dos modos básicos de registro: el **registro por barrido** (más conocido como *mapping* o *pushbroom system*) o el **registro por planos** (*global imaging* o *staring imager system*).

▪ Registro por barrido

El registro por barrido consiste en la medida secuencial de regiones adyacentes de una muestra (Figura 2.11), ya sean puntos concretos (registro por puntos) o líneas de puntos (registro por líneas). Históricamente, este fue el primer método usado para la adquisición de imágenes hiperespectrales en espectroscopia vibracional [40]. Sus inicios se encuentran en una publicación de la revista *Nature* del año 1949 [41], donde se describía el uso de un microscopio acoplado a un espectrómetro IR. No se trataba de un aparato de imagen, ya que sólo registraba espectros en un solo punto, pero fue un primer paso en la adquisición de datos en puntos localizados en el espacio. Los primeros mapas químicos fueron publicados en 1988 por Harthcock y Atkin [42] y estaban registrados usando un microscopio con FT-IR acoplado a una plataforma móvil para la muestra. Los sistemas de imagen química que registraban información espectral y espacial a la vez empezaron a ser disponibles a principios de los años 90 [36]. A pesar de las mejoras exponenciales, el principio de funcionamiento de estos equipos sigue siendo el mismo: el movimiento sistemático de la muestra en dos dimensiones después de cada registro, permite la adquisición de espectros de zonas adyacentes de la muestra hasta cubrir toda la superficie deseada. El área de la muestra

a analizar y el tamaño de píxel son definidos previamente y determinan el número de píxeles de la imagen final [6].

El registro por barrido permite la obtención de imágenes de alta resolución espacial, pero requiere un desplazamiento muy preciso en cada adquisición. Esto implica que el proceso de análisis sea bastante lento, limitando tanto el tamaño de la superficie a registrar, como su aplicación *in-/on-line* [38].

▪ Registro por planos

A diferencia del barrido, el registro por planos se basa en el uso de los detectores *Focal Plane Array* (FPA), compuestos por miles de detectores individuales. Esto significa que pueden adquirir miles de píxeles de una sola vez. Las primeras matrices de detectores bidimensionales fechan de los años 70 [43] y fueron desarrollados con aplicaciones militares, como detección de misiles o aparatos de visión nocturna [44,45]. Progresivamente los FPA se han ido aplicando a la espectroscopia, lo que ha fomentado el desarrollo de detectores cada vez más baratos y de mayor formato [42,43], aumentando su popularidad en el campo de la imagen química.

La mayor ventaja del registro por planos es que no requiere de partes móviles, sino que se registran imágenes de todo el plano de la muestra a cada longitud de onda deseada (Figura 2.11). Es decir, que el FPA registra simultáneamente la intensidad de la absorción NIR en todos los píxeles de la imagen y lo repite para todas las longitudes de onda deseadas, que se van sucediendo secuencialmente gracias a los selectores de longitud de onda, mayoritariamente filtros sintonizables (Figura 2.10). El número de píxeles viene dado por las características del FPA, normalmente de 256x320 píxeles, mientras que la medida del píxel se determina con el sistema óptico [6]. La imagen hiperespectral final se obtiene de la combinación de las imágenes a todas las longitudes de onda registradas.

Esta técnica permite reducir considerablemente el tiempo de análisis, sin embargo, puede conllevar una disminución de la resolución espacial con respecto al registro por barrido.



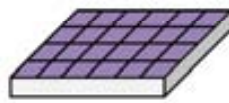
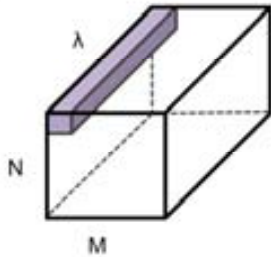
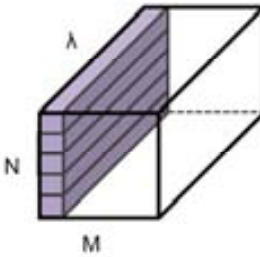
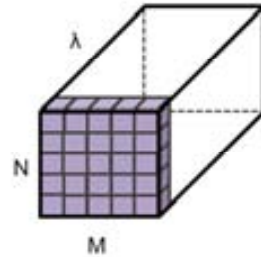
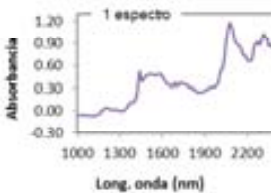
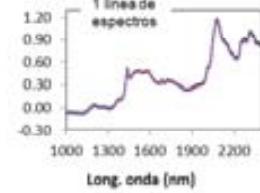
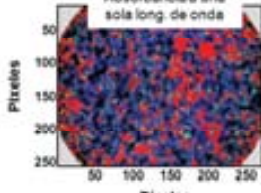
	REGISTRO POR BARRIDO		REGISTRO POR PLANOS
	Registro por puntos	Registro por líneas	
Muestra			
Cubo de datos			
Información obtenida			

Figura 2.11. Parte de la muestra analizada e información obtenida con una sola adquisición de los distintos modos de registro en equipos NIR-CI.

ELECCIÓN DEL SISTEMA

En ambos tipos de registro (registro por barrido y por planos), se obtiene como resultado la imagen hiperespectral, pero existen importantes diferencias entre los dos modos de adquisición [6]. De nuevo, la elección de la técnica dependerá de la aplicación [36].

El registro por barrido se basa normalmente en espectrómetros de FT [6] por lo que proporciona gran precisión de longitud de onda y mayor resolución espectral, dando lugar a datos espectrales de mayor calidad. Además, la iluminación directa de zonas pequeñas (píxeles o líneas de píxeles) mejora la calidad de la imagen, aumentando la relación señal/ruido y la uniformidad de la iluminación y reduciendo los efectos de mezclado entre píxeles [6]. Esto junto con la gran resolución espacial, hacen del registro por barrido un método muy adecuado para investigación y trabajos

con muestras desconocidas, donde se requiere una mayor información espectral. El estudio de la distribución de componentes minoritarios, como puede ser la detección de impurezas [42] o el análisis de superficies irregulares [36] donde es muy conveniente la posibilidad de modular el enfoque en cada posición espacial, son ejemplos de aplicaciones donde este modo de registro es ventajoso. En cambio, en el caso de muestras grandes o con geometrías irregulares su utilidad es mucho más limitada [6].

En cambio, cuando se requiere una caracterización rápida y general de una muestra conocida, donde la relación señal/ruido no es tan importante, el modo de registro por planos es más apropiado [36]. Esto se debe principalmente al tiempo de adquisición que, al no depender del área analizada, es mucho menor que en los sistemas por barrido (<5 min), lo que lo hace muy adecuado para el análisis de muchas muestras en poco tiempo [36]. Otra ventaja de este modo de registro es que presenta una gran flexibilidad en cuanto a campo visual se refiere: cambiando las lentes utilizadas, se puede aumentar o disminuir la zona de análisis, adaptándola a las necesidades de cada caso en particular [6]. Esto facilita el análisis de muestras grandes o irregulares, así como el estudio simultáneo de múltiples muestras a la vez [46]. Además, como se ha comentado anteriormente, el registro por planos no requiere de partes móviles por lo que los equipos son más simples y robustos y, por ende, más adecuados para su uso en ambientes industriales y aplicaciones *in-* y *on-line* [6].

Por el contrario la ausencia de partes móviles implica que toda la superficie de la muestra tiene que ser iluminada durante el tiempo de adquisición, lo que puede provocar el calentamiento de la muestra [42]. Por este motivo, esta configuración es poco adecuada para muestras con componentes lábiles al calor o con bajo punto de fusión.

2.4. Aplicaciones industriales del NIR y el NIR-CI

El interés por la espectroscopia NIR en el sector industrial se debe principalmente a las ventajas que ofrece esta técnica, listadas en el apartado 2.2, para el control de calidad y monitoreo de procesos de gran variedad de productos. Esto, junto con las cada vez más arraigadas políticas industriales de análisis de procesos [47], también conocidas como PAC o, más recientemente PAT (apartado 1.2.4), que abordan gran variedad de problemas analíticos en términos de tiempo real, han favorecido y siguen favoreciendo la instauración de las técnicas NIR y NIR-CI, combinadas con la quimiometría, en gran variedad de sectores y situaciones. Prueba de ello, son el número creciente de referencias bibliográficas sobre estas técnicas (Figura 2.12), así como la publicación reciente de dos normas ASTM dirigidas al establecimiento de las prácticas estándares para el uso de las espectroscopias MIR y NIR en análisis cualitativo, ASTM 1790 (2010) [48] y cuantitativo, ASTM 1655 (2012) [49].

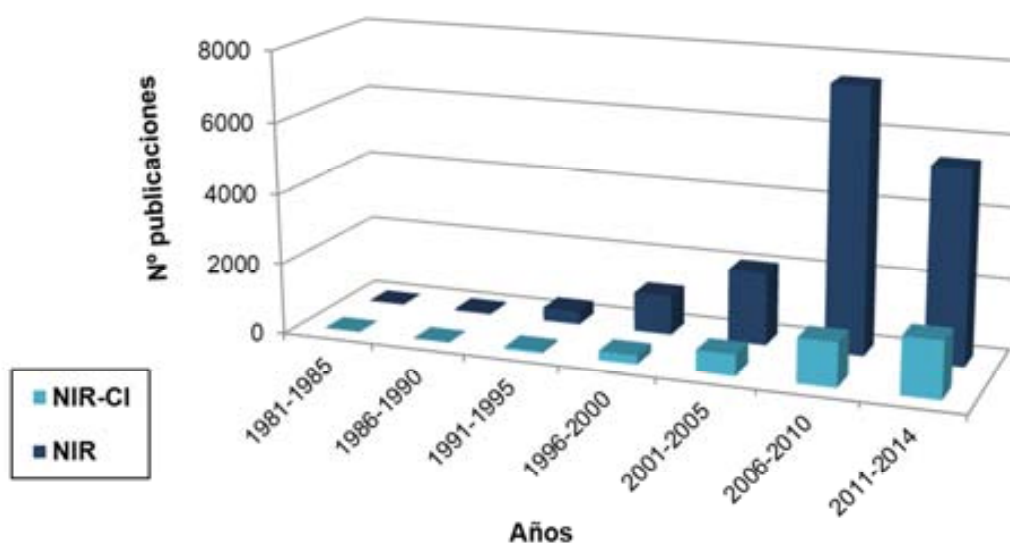


Figura 2.12. Evolución del número de publicaciones relacionadas con las técnicas NIR y NIR-CI, según datos extraídos de [50].

Las primeras aplicaciones industriales de la espectroscopia NIR fueron impulsadas por Karl Norris en los años 60, en el campo agroalimentario [15,51]. Estos

estudios iban dirigidos al análisis nutricional y de calidad de cereales: determinación de humedad, contenido en fibras, grasas y proteínas... La eficacia de estos métodos hizo que se extendieran rápidamente a otros productos agrolimentarios de interés como la soja, las semillas oleaginosas o los forrajes [52,53,54], así como la leche o la carne [16,52]. Actualmente, existen aplicaciones de la espectroscopia NIR para la mayoría de productos alimentarios y bebidas (carne, pescado, cereales, leche huevos, frutas, verduras, zumos, vinos, cervezas) [55], destinados principalmente a la calidad de los productos [56,57] y que incluyen análisis tan dispares como la identificación de transgénicos [58], la diferenciación entre alimentos frescos y congelados [59] o la detección de contaminación por bacterias [60].

Siguiendo el ejemplo de las aplicaciones en la agricultura, muchas otras áreas han ido adoptando también el NIR como técnica analítica con distintas aplicaciones. La industria química [49,61,62] y petroquímica [49,63,64], la farmacéutica [20,49,65,66], el sector de la biomedicina [20,49,67] o la biotecnología [68,69] son algunos sectores donde más se ha implementado esta técnica. En concreto, la industria farmacéutica y petroquímica son dos áreas en las que el NIR ha encontrado una gran aplicabilidad [70]. Dada la importancia de ambos sectores en esta tesis, en los apartados 2.4.1 y 2.4.2 se hace una revisión más detallada de las aplicaciones NIR desarrolladas en estos campos.

En el caso del NIR-CI, su implementación es mucho más reciente, pero sus aplicaciones también se han extendido rápidamente a distintos sectores, entre los que destacan el sector agroalimentario [71,72,73], la clasificación de residuos [74,75,76] y, en especial, la industria farmacéutica [36,42,77] (apartado 2.4.2).

2.4.1. Aplicaciones NIR en la industria petroquímica

Como se describe en el capítulo 1, los métodos de referencia en la industria petroquímica vienen definidos por los estándares de la ISO y la ASTM. Ambos, son ampliamente aceptados y reconocidos para el análisis de productos petroquímicos, pero en muchos casos presentan grandes inconvenientes: son muy lentos, requieren

volúmenes de muestra grandes y el uso de solventes, tienen un coste muy elevado y no permiten el trabajo *on-line* [78,79]. La gran cantidad de especificaciones que deben cumplir los productos refinados para poder ser comercializados, juntamente con la aplicación del PA o PAC para el seguimiento de parámetros críticos del proceso, implican la necesidad de metodologías más rápidas y efectivas. Se requieren técnicas que se adapten a la producción en continuo de la industria petroquímica y permitan la obtención de información del producto/proceso en tiempo real. Una metodología que cumple con estos requisitos es la espectroscopia NIR que, combinada con la quimiometría, permite analizar de forma rápida y *on-site* distintas propiedades físicas y químicas, con precisión [62,77,80] y repetibilidad [76,78] comparables a las de los métodos de referencia. Por esta razón, el NIR ha sido adoptado como herramienta analítica en la industria petroquímica con gran variedad de aplicaciones, tanto por lo que respecta a productos, como a etapas del proceso, cubriendo tanto el análisis y caracterización de las materias primas, como la monitorización y supervisión del proceso o el aseguramiento de la calidad del producto final. Algunos ejemplos de aplicaciones NIR a la industria petroquímica se muestran en la Tabla 2.3, ordenados según el tipo de muestra que se somete a análisis. A pesar del gran número de aplicaciones desarrolladas para ciertos productos petroquímicos, como gasóleos y gasolinas, el uso del NIR no es homogéneo en todo el sector. Los estudios destinados al análisis de parafinas, por ejemplo, son muy limitados y se reducen a su detección y análisis como parte de la caracterización de crudos [81,82] u otras corrientes del proceso de refino [83,84], con el objeto de prevenir los problemas que causa su cristalización y deposición en cañerías durante el transporte y refino.

Tabla 2.3. Ejemplos de algunas aplicaciones de la espectroscopia NIR a la industria petroquímica. En verde, aplicaciones a materia prima, en azul, a productos intermedios y, en lila, a producto acabado.

Ref.	Año	Muestra	Parámetros analizados	Aplicación
[85]	2000	Crudos	TBP, densidad	Caracterización de materia prima
[79]	2009		CP, deposición ceras	
[80]	2003		WAT, tiempo de fusión	

[77]	1999	Destilado medio	RI, densidad, %H y aromáticos y CN	
[81]	2004	Destilado medio	Composición: parafinas, naftas y aromáticos	
[78]	2007	Gasolinas FCC hidrotratadas	Diolefinas conjugadas	Control en proceso
[82]	2009	Corrientes del reformado	Composición: parafinas, isoparaf., naftas y aromáticos	
[86]	2012	Diésel no hidrotratado	Punto anilina, CI, densidad, T dest	
[87]	2008	Gasolinas	Clasificación por origen y calidad	Control calidad y detección de adulteraciones
[88]	2014		Etanol	
[89]	2002	Nafta / polipropileno	Hidrocarburos y densidad / etileno y viscosidad	Control de calidad
[90]	2003	Diésel	SC	Control de calidad
[91]	2009	Mezclas diésel-biodiésel	Contenido biodiésel, aceite vegetal y nafta	Control de calidad y detección de adulteraciones
[92]	2010		Densidad, SC y T dest	
[93]	2011		Clasificación según cont. biodiésel y aceite vegetal	
[94]	2012		Contenido biodiésel y aceite vegetal	
[95]	2008	Biodiésels	Composición FAME	Control de calidad
[96]	2008		Iodine value, CFPP, viscosidad y densidad	
[97,98]	2011		Densidad, viscosidad, metanol y agua	
[99]	2011		Composición fraccional, iodine value, CFPP	
[100]	2011	Aceite de motor	Clasificación según origen y viscosidad	Control de calidad y detección de adulteraciones
[101]	2012	Lubricantes	Viscosidad y composición	Control de calidad
[102]	2001	Asfaltos	Penetración, viscosidad y densidad	Control de calidad
[103]	2004		Penetración, viscosidad y FP	

2.4.2. Aplicaciones NIR y NIR-CI en la industria farmacéutica

Como se ha comentado en el capítulo 1, las exigencias de los organismos reguladores en los Estados Unidos (FDA) y la Unión Europea (EMA) han propiciado la implementación del PAT en la industria farmacéutica. Para ello, es necesaria la implementación de métodos analíticos que permitan reflejar la situación del proceso de manufactura en tiempo real [104]. Se requieren técnicas rápidas, aplicables *on-* o *in-line*, que sustituyan algunos métodos comúnmente usados como la cromatografía líquida (HPLC, *High-Performance Liquid Chromatography*) o de gases (GC, *Gas Chromatography*), que son destructivas, requieren mucho tiempo de análisis y producen una elevada cantidad de residuos. De entre todas las técnicas adaptables a las condiciones requeridas por el PAT destacan las técnicas espectroscópicas. Las ventajas generales de las espectroscopias vibracionales (NIR, MIR y Raman) en cuanto a rapidez, versatilidad y análisis remoto se han descrito anteriormente (apartado 2.1), con especial énfasis en los atributos del NIR, la clave para que esta técnica ocupe el primer lugar en cuanto a aplicaciones PAT se refiere [102,105].

Actualmente, la mayoría de farmacopeas ya han adoptado el NIR como técnica analítica. Tanto la versión europea [106] como la americana [107] incluyen un capítulo que describe sus características principales y aborda algunas de las cuestiones más importantes a tener en cuenta para su aplicación en la industria farmacéutica. Esta misma información se incluye también en otras farmacopeas regionales como la española [108] o la británica [109]. Esto representa un voto de confianza hacia estas técnicas, lo que disminuye la inseguridad y favorece la implementación de esta técnica, a pesar del coste económico que ello implica y de la necesidad de validación y aceptación de los métodos por las autoridades (apartado 1.2.4). También la EMA [110] considera la espectroscopia NIR como una técnica con una gran utilidad en la industria farmacéutica para la identificación y cualificación de materias primas, así como distintos ensayos para la determinación de distintas características fisicoquímicas en productos intermedios y acabados.

Todo ello ha impulsado el desarrollo de un gran número de aplicaciones de la espectroscopia NIR a la industria farmacéutica. La versatilidad del NIR favorece la

disparidad de estas aplicaciones, entre las cuales se pueden diferenciar análisis cualitativos y cuantitativos; *off*-, *at*-, *on*- o *in-line*; para determinación de propiedades tanto químicas (identificaciones y determinación de uniformidades, composición...), como físicas (polimorfismos, dureza, perfiles de disolución, tamaño de partícula...); de fármacos en cualquier punto del proceso productivo (materia prima, producto intermedio y acabado), así como del mismo proceso (monitorización y determinación del punto final de operaciones como mezclado, granulación, secado, liofilización, recubrimiento). Algunas de estas aplicaciones se listan en la Tabla 2.4 y otras muchas se describen en revisiones generales sobre NIR y su empleo en la industria farmacéutica [111,112].

Tabla 2.3. Ejemplos de algunas aplicaciones de la espectroscopia NIR a la industria farmacéutica. En verde, aplicaciones a materia prima, en azul, a productos intermedios y, en lila, a producto acabado.

Ref.	Año	Muestra	Parámetros analizados	Aplicación
[113,114]	2001 2003	Materia prima	Identidad	Identificación materia prima
[115]	2014	Soluciones acuosas	Conc. de soluto y ratio entre polimorfos	Cristalización
[116]	2001	Mezclas en polvo	Homogeneidad de distrib. de comp. y punto final de mezclado	Mezclado (<i>Blending</i>)
[117]	2009		Concentraciones de API y excipientes	
[118,119]	2012 2013		Punto final de mezclado	
[120]	2006	Granulados	Tranformaciones polimórficas	Granulación
[121]	2007		Composición en API y excipientes	
[122]	2010		Identific. y monitoriz. de las etapas del proceso. Determinación de humedad, distrib. del tamaño de partícula y densidad	
[123,124]	2011 2014		Contenido de agua y punto final del proceso	

INTRODUCCIÓN

[125]	2009	Formulaciones liofilizadas	Monitoriz. de las etapas del proceso: control de sublimación, nucleación, cristaliz., liberación de agua	Liofilización (<i>Freeze-drying</i>)
[126]	2011		Desnaturalización proteica y puentes de hidrógeno proteína-lioprotector	
[127]	2014	Mezclas <i>post-blending</i>	Uniformidad de contenido (segregación)	Llenado matriz de compactación
[128]	2011	Pellets	Grosor de recubrimiento	Recubrimiento
[129]	2008	Núcleos/ Comprimidos	Grosor de recubrimiento	Recubrimiento
[130]	2008		Grosor de recubrimiento y perfiles de disolución	
[131]	2010		Grosor de recubrimiento y punto final del proceso	

CONTROL DE PRODUCTO ACABADO	[132]	2004		Contenido en humedad y composición para diferenciación del lugar de manufactura	Control de calidad y detección de falsificaciones
	[133]	2007	Comprimidos	Porosidad, dureza y contenido en API	
	[134]	2007		Dureza, uniformidad de contenido y perfil de disolución	
	[135,136]	2004 2009		Perfiles de disolución	
	[137]	2010		Cuantificación API y excipientes	
	[138]	2005		Composición y uniformidad química y física para detección de falsific.	
	[139]	2008	Agregados	Distribución del tamaño de partícula	Control de calidad
	[140,141,142]	2002 2008 2010	Formulaciones liofilizadas	Contenido de agua	Control de calidad
	[143]	2010		Clasificación en función de la composición, contenido en	

agua y estado sólido				
[144]	2010		Identificación de polimorfos	
[145]	2008	Geles	Concentración de API y excipientes	Control de calidad
[146]	2011		Concentraciones de API y excipientes, viscosidad y pH	

En el caso del NIR-CI, a pesar de que su aparición es mucho más reciente, la capacidad para extraer gran cantidad de información de superficies de forma rápida y simple ha promovido rápidamente su aplicación en la industria farmacéutica [147,109]. Destaca su uso en el aseguramiento de la calidad de productos farmacéuticos, mayoritariamente de productos sólidos y en aplicaciones *at-line*, algunas de las cuales se listan en la Tabla 2.5a. Como principal ventaja respecto el NIRS, el NIR-CI proporciona información sobre la distribución de los parámetros analizados en toda la superficie de la muestra; además, el campo de visión que ofrece, permite llevar a cabo análisis de alto rendimiento, *i. e.* de varias muestras a la vez, incluso cuando ya están empaquetadas en blíster [44,148,149]. El NIR-CI también es muy útil en el desarrollo de nuevos productos y entendimiento de los procesos (Tabla 2.5b). En este sentido, existen muchos estudios donde se evalúan los efectos de la modificación de formulaciones [150] o distintos parámetros del proceso [151,152,153,154] con el objetivo de discriminar las mejores condiciones, para mejorar el diseño y el proceso de producción, así como solucionar problemas ya existentes [155].

Tabla 2.5. Ejemplos de algunas aplicaciones de la espectroscopia NIR-CI a la industria farmacéutica. En azul, aplicaciones a productos intermedios, en lila, a productos acabado y, en naranja, a productos en desarrollo.

	Ref.	Año	Muestra	Parámetros analizados	Aplicación
CONTROL EN PROCESO	[156]	2008	Mezclas en polvo	Patrones de distrib. de comp., punto final de mezclado y tamaño de dominios de API	Mezclado (<i>Blending</i>)
	[114,157]	2001		Homogeneidad de distrib.	

INTRODUCCIÓN

		2012		de comp. y punto final de mezclado	
	[158]	2008	Granulados	Composición química	Granulación
	[159]	2009	Núcleos/ Comprimidos	Crecimiento y distribución del recubrimiento y diferencias inter e intracomprimido	Recubrimiento
	[160]	2011	Pellets	Distribución de comp. en el núcleo y las capas de recubrimiento y su grosor	Extrusión/esferoniz.
CONTROL DE PRODUCTO ACABADO	[161]	2008		Detección de defectos y contaminantes en superf.	
	[162,163]	2007 2011		Uniformidad de contenido	
	[164,165]	2009 2012		Cuantificación y distrib. de componentes	
	[136,166,167]	2005 2009 2010	Comprimidos	Identific., cuantific. y distrib. de comp. para detección de falsific.	Control de calidad y detección de falsificaciones
	[168]	2013		Monitorización de la distrib. de comp. durante la disolución	
	[169,170]	2011 2014		Distribución y cuantificación de formas polimórficas	
	[171]	2006	Formulaciones liofilizadas	Homogeneidad de mezclas proteína-azúcar	Control de calidad
	[172]	2010		Distribución de componentes	
	[146]	2001	Comprimidos en blíster	Determinación de API y humedad	Análisis de alto rendimiento
	[44]	2006		Uniformidad de contenido	
[147]	2002	Cápsulas de gelatina	Detección degradación		
PRODUCTO Y ENTENDIMIENTO	[149]	2008	Mezclas en polvo	Número y tamaño de agregados	Mezclado (<i>Blending</i>)
	[148]	2014	Granulados	Mezclado y distrib. del líquido de granulación y distrib. del tamaño de partícula	Granulación
	[150]	2008	Extrudatos y núcleos	Homogeneidad de distrib.de componentes	Extrusión

[153]	2004		Identificación y tamaño de dominios de distintos comp.	
[151]	2008	Comprimidos	Uniformidad de distrib. de API	Producto acabado
[152]	2008		Distribución densidad	

REFERENCIAS

- 1 Herschel, W. (1800) Investigation of the powers 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. *Philos. Trans. R. Soc. Lond.* **90**, 255-283.
- 2 Maspoch, S. (2012) El procesado de la señal espectroscópica. En: Ríos, A., Moreno, M.C., Simonet, B.M. (coords.) *Técnicas espectroscópicas en química analítica*. Vol. I; Editorial Síntesis, Madrid, pp. 103-152.
- 3 Skoog, D.A., Holler, F.J., Nieman, T.A. (2001) *Principios de análisis instrumental*, 5ª ed. McGraw-Hill, Madrid.
- 4 Toxicología ambiental y salud pública [sitio web]. Departamento de Biología Aplicada, Universidad Miguel Hernández, Elche [Consultado en mayo de 2014]. Disponible en: [//repositorio.innovacionumh.es/Proyectos/P_22CursoMateriales/Miguel_Angel_Sogorb/Wimba/index.htm](http://repositorio.innovacionumh.es/Proyectos/P_22CursoMateriales/Miguel_Angel_Sogorb/Wimba/index.htm)
- 5 Siesler, H.W. (2008) Basic principles of near-infrared spectroscopy. En: Burns, D.A., Ciurczak, E.W. *Handbook of Near-Infrared Analysis*. 3ª ed. CRC Press, Boca Raton, FL, pp. 7-19.
- 6 Ravn, C. (2009) *Near-infrared chemical imaging in formulation development of solid dosage forms*. Tesis Doctoral, University of Copenhagen & Novo Nordisk A/S.
- 7 Blanco, M., Villaroya, I. (2002) Spectroscopy: A rapid-response analytical tool. *TrAC-Trend. Anal. Chem.* **21**, 240-250.
- 8 Reich, G. (2005) Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv. Drug Deliver. Rev.* **57**, 1109-1143.
- 11 Rosas, J.G. (2013) *La Tecnología Analítica de Procesos (PAT) en la industria farmacéutica. Utilización de la espectroscopia NIR y técnicas de imagen química NIR-CI*. Publicia, Omniscryptum GmbH & Go. KG, Saarbrücken.
- 12 Simpson, M.B. (2010) Near-infrared spectroscopy for process analytical technology: Theory, technology and implementation. En: Bakeev K.A. (ed.) *Process analytical Technology*. 2ª ed.; Wiley, Chichester, pp. 107-155.

- 13 Alcalà, M. (2006) *Utilización de la espectroscopia NIR en el control analítico de la industria farmacéutica*. Blanco, M. (dir.) Tesis doctoral, Universitat Autònoma de Barcelona.
- 14 Hindle, P.H. (2008) Historical Development. En: Burns, D.A., Ciurczak, E.W. *Handbook of Near-Infrared Analysis*. 3^a ed. CRC Press, Boca Raton, FL, pp. 3-6.
- 15 Abney, W., Festing, E.R. (1981) On the influence of the atomic grouping in the molecules of organic bodies on their absorption in the infra-red region of the spectrum, *Philos. Trans. R. Soc. Lond.* **172**, 887-918.
- 16 Coblenz, W. W. (1905) Investigations of inf-rared spectra. Part 1. Infra-red absorption spectra [en línea]. Publication No. 35 of Carnegie Institute of Washington, The Henry E. Wilkens Printing Co., Washington [Consultado en mayo de 2014]. Disponible en: <https://archive.org/details/investigationsi03coblgoog>
- 17 Hart, J.R., Norris, K.H., Golumbic, C. (1962) Determination of the moisture content of seeds by near-infrared spectrophotometry of their methanol extracts. *Cereal Chem.* **39**(2), 94-99.
- 18 Ben-Gera, I., Norris, K.H. (1968) Direct spectrophotometric determination of fat and moisture in meat products. *J. Food Sci.* **33**(1), 64-67. Referencia encontrada en [14].
- 19 Williams P.C. (1975) Application of near infrared reflectance spectroscopy to analysis of cereal grains and oilseeds. *Cereal Chem.* **52**, 561-576. Referencia encontrada en [55].
- 20 Todd, S. (2009). *NIR and Raman: Complementary techniques for raw material identification*. Technical note: 51768, Thermo Fischer Scientific Inc., Madison, WI.
- 21 McClure, W.F. (1994) The giant is running strong. *Anal. Chem.* **66**(1), 43A-53A.
- 22 Ciurczak, E.W., Drennen, J.K. (2002) *Pharmaceutical and medical applications of near-infrared spectroscopy*. Marcel Dekker, New York, NY.
- 23 Chen, J.Y., Iyo, C., Kawano, S., Terada, F. (1999) Development of calibration with sample cell compensation for determining the fat content of unhomogenised raw milk by a simple near infrared transmittance method, *J. Near Infrared Spec.* **7**, 265-273.
- 24 Broad, N.W., Jee, R.D., Moffat, A.C., Smith, M.R. (2001) Application of transmission near-infrared spectroscopy to uniformity testing of content of intact steroid tablets, *Analyst* **126**, 2207-2211.

- 25 Blanco, M., Romero, M.A. (2002) Near infrared transfectance spectroscopy determination of dexketoprofen in a hydrogel. *J. Pharm. Biomed. Anal.* **30(3)**, 467-472.
- 26 Cogdill, R.P., Drennen, J.K. (2006) Near-infrared spectroscopy. En: Brittain, H.G. (ed.) *Spectroscopy of pharmaceutical solids*, Taylor & Francis Group, New York, NY, pp. 313-412.
- 27 Chen, Y.X., Thosar, S.S., Forbess, R.A., Kemper, M.S., Rubinovitz, R.L., Shukla, A.J. (2001) Prediction of drug content and hardness of intact tablets using artificial neural network and near-infrared spectroscopy. *Drug Dev. Ind. Pharm.* **27(7)**, 623-631.
- 28 Otsuka, M. (2004) Comparative particle size determination of phenacetin bulk powder by using kubelka-munk theory and principal component regression analysis based on near-infrared spectroscopy. *Powder Technol.* **141(3)**, 244-250.
- 29 Gupta, A., Peck, G.E., Miller, R.W., Morris, K.R.. (2004) Nondestructive measurements of the compact strength and the particle-size distribution after milling of roller compacted powders by near-infrared spectroscopy. *J. Pharm. Sci.* **93(4)**, 1047-1053.
- 30 Workman, J.J., Burns, D.A. (2008) Commercial NIR Instrumentation. En: Burns, D.A., Ciurczak, E.W. *Handbook of Near-Infrared Analysis*. 3^a ed. CRC Press, Boca Raton, FL, pp. 67-78.
- 31 Siesler, H.W. Analytical Services of the Department of Physical Chemistry [sitio web]. University of Duisburg-Essen, Essen [Consultado en mayo de 2014]. Disponible en: <http://www.nir-spektroskopie.de/indexe.htm>
- 32 Sablinskas, V. (2003) Instrumentation. En: Gauglitz, G., Vo-Dinh, T. *Handbook of Spectroscopy*, Wiley-VCH, Weinheim, pp. 47-69.
- 33 Andersen, H.V., Wedelsbäck, H., Hansen, P.W. (2013) *A white paper from FOSS: NIR spectrometer technology comparison*. P/N 1026672 (1), FOSS, Hilleroed.
- 34 Tran, C.D. (1992) Acousto-optic devices: Optical elements for spectroscopy. *Anal. Chem.*, **64(20)**, 971A-981A.
- 35 Bei, L., Dennis, G.I., Miller, H.M., Spaine, T.W., Carnahan, J.W. (2004) Acousto-optic tunable filters: Fundamentals and applications as applied to chemical analysis techniques. *Prog. Quant. Electron.* **28(2)**, 67-87.

- 36 McCarthy, W.J., Kemeny, G.J. (2008) Fourier transform spectrophotometers in the near-infrared. En: Burns, D.A., Ciurczak, E.W. *Handbook of Near-Infrared Analysis*. 3^a ed. CRC Press, Boca Raton, FL, pp. 79-91.
- 37 Peinado, A.C. (2005) *Aplicación y desarrollo de técnicas quimiométricas para el seguimiento de fermentaciones alcohólicas*. Blanco, M. (dir.) Tesis Doctoral, Universidad Autónoma de Barcelona.
- 38 Gendrin, C., Roggo, Y., Collet, C. (2008) Pharmaceutical applications of vibrational chemical imaging and chemometrics: A review. *J. Pharm. Biomed. Anal.* **48**, 533-553.
- 39 Amigo, J.M. (2010) Practical issues of hyperspectral imaging analysis of solid dosage forms. *Anal. Bioanal. Chem.* **398**(1), 93-109.
- 40 Roggo, Y., Ulmschneider, M. (2008) Chemical Imaging and Chemometrics: Useful tools for process analytical technology. En: Cox, S. (ed.) *Pharmaceutical Manufacturing Handbook: Regulations and quality*. Wiley, Hoboken, NJ, pp. 411-431.
- 41 Barer, R., Cole, A.R.H., Thompson, H.W. (1949) Infra-red spectroscopy with the reflecting microscope in physics, chemistry and biology. *Nature*, **163**, 198-200. Referencia encontrada en [45].
- 42 Harthcock, M.A., Atkin, S.C. (1988) Imaging with functional group maps using infrared microspectroscopy. *Appl. Spectrosc.* **42**(3), 449-455.
- 43 Sprafke, T., Beletic, J.W. (2008) High-performance infrared focal plane arrays for space applications, *Opt. Photonics News* **19**(6), 22-27.
- 44 Lewis, E.N., Schoppelrei, J.W., Makein, L., Kidder L.H., Lee, E. (2010) Near-infrared chemical imaging for product and process understanding. En: Bakeev K.A. (ed.) *Process analytical Technology*. 2^a ed.; Wiley, Chichester, pp. 245-279.
- 45 Lewis, E.N., Schoppelrei, J., Lee, E. (2004) Near-infrared chemical imaging and the PAT initiative. *Spectroscopy* **19**(4), 26-36.
- 46 Lee, E., Huang, W.X., Chen, P., Lewis, E.N., Vivilecchia, V. (2006) High-throughput analysis of pharmaceutical tablet content uniformity by near-infrared chemical imaging. *Spectroscopy* **21**, 24-32.
- 47 Guenard, R., Thurau, G. (2010) Implementation of Process Analytical Technologies. En: Bakeev K.A. (ed.) *Process analytical Technology*. 2^a ed.; Wiley, Chichester, pp. 107-155.

- 48 ASTM E1790-04 (2010) Standard Practice for Near Infrared Qualitative Analysis. *ASTM International*, West Conshohocken, PA, Estados Unidos.
- 49 ASTM E1655-05 (2012) Standard Practices for Infrared Multivariate Quantitative Analysis. *ASTM International*, West Conshohocken, PA, Estados Unidos.
- 50 Web of Science (WOS) [base de datos online]. Thomson Reuters, New York, NY [Consultado en septiembre de 2014]. Disponible mediante licencia en: <https://www.accesowok.fecyt.es/>
- 51 Burns, D.A., Ciurczak, E.W. (2008) *Handbook of Near-Infrared Analysis. Part IV. Applications*. 3ª ed. CRC Press, Boca Raton, FL, pp. 7-19.
- 52 Hymowitz, T., Dudley, J.W., Collins, F.I., Brown, C.M. (1974) Estimations of protein and oil concentration in corn, soybean, and oat seed by near-infrared light reflectance. *Crop Sci.* **14**, 713-715.
- 53 Norris, K.H., Barnes, R.F., Moore, J.E., Shenk, J.S. (1976) Predicting forage quality by infrared reflectance spectroscopy. *J. Anim. Sci.* **43**, 889-897.
- 54 Watson, A.C. (1977) Near infrared reflectance spectrophotometric analysis of agricultural products. *Anal. Chem.* **49**(9), 835A-840A.
- 55 Sun D.W. (2009) *Infrared spectroscopy for food quality analysis and control. Part II: Applications*. Academic Press, Elsevier, New York, NY.
- 56 Cen, H., He, Y. (2007) Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Tech.* **18**(2), 72-83.
- 57 Plans, M., Simó, J., Casañas, F., Sabaté, J. (2012) Near-infrared spectroscopy analysis of seed coats of common beans (*Phaseolus vulgaris L.*): A potential tool for breeding and quality evaluation. *J. Agric. Food Chem.* **60**, 706-712.
- 58 Alishahi, A., Farahmand, H., Prieto, N., Cozzolino, D. (2010) Identification of transgenic foods using NIR spectroscopy: A review. *Spectrochim. Acta A* **75**(1), 1-7.
- 59 Uddin, M., Okazaki, E., Turza, S., Yumiko, Y., Fukuda, Y., Tanaka, M. (2005) Non-destructive visible/NIR spectroscopy for differentiation of fresh and frozen-thawed fish. *J. Food Sci.* **70**, C506-C510.

- 60 Rodriguez-Saona, L.E., Khambaty, F.M., Fry, F.S., Dubois J., Calvey E.M. (2004) Detection and identification of bacteria in a juice matrix with Fourier transform-near infrared spectroscopy and multivariate analysis. *J. Food Protect.* **67(11)**, 2555-2559.
- 61 Garrido, M., Rius, F.X., Larrechi, M.S. (2008) Multivariate curve resolution–alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. *Anal. Bioanal. Chem.* **390(8)**, 2059-2066.
- 62 Brearley, A.M., Foulk, S.J. (2010) Near-infrared spectroscopy (NIR) as a PAT tool in the chemical industry: Added value and implementation challenges. En: Bakeev K.A. (ed.) *Process analytical Technology*. 2^a ed.; Wiley, Chichester, pp. 493-520.
- 63 Monteiroa, M.R., Ambrozina, A.R.P., Lião, L.M., Ferreira, A.G. (2008) Critical review on analytical methods for biodiesel characterization. *Talanta* **77**, 593-605.
- 64 Felizardo, P., Menezes, J.C., Neiva-Correia, M.J. (2011) Process analytical technology use in biofuels manufacturing. En: Undey, C., Low, D., Menezes, J.C., Koch, M. (eds.) *PAT applied in biopharmaceutical process development and manufacturing: An Enabling Tool for Quality-by-Design*, CRC Press, Boca Raton, FL, pp. 201-221.
- 65 Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., Jent, N. (2007) A review of near-infrared spectroscopy and chemometrics in pharmaceutical technologies. *J. Pharm. Biomed. Anal.* **22**, 683-700.
- 66 Bakeev K.A. (2010) *Process analytical Technology*. 2^a ed.; Wiley, Chichester.
- 67 Aziz, S.M., Khambatta, F., Vaithianathan, T., Thomas, J.C., Clark, J.M., Marshall, R. (2010) A near infrared instrument to monitor relative hemoglobin concentrations of human bone tissue in vitro and in vivo. *Rev. Sci. Instrum.* **81**, 043111.1-043111.7.
- 68 Scarff, M., Arnold, S.A., Harvey, L.M., McNeil, B. (2006) Near infrared spectroscopy for bioprocess monitoring and control: Current status and future trends. *Crit. Rev. Biotechnol.* **26**, 17-39.
- 69 Undey, C., Low, D., Menezes, J.C., Koch, M. (2011) *PAT applied in biopharmaceutical process development and manufacturing: An enabling tool for quality-by-design*, CRC Press, Boca Raton.
- 70 Pasquini, C. (2003) Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* **14(2)**, 198-219.

- 71 Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., Frias, J.M. (2007) Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends Food Sci. Tech.* **18**, 590-598.
- 72 Salzer, R., Siesler, H.W. (2009) *Infrared and Raman spectroscopic imaging. Part three. Agricultural applications.* Wiley-VCH, Weinheim.
- 73 Lorente, O., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O.L., Blasco, J. (2011) Recent advantages and applications of hyperspectral imaging for fruit and vegetable quality. *Food Bioprocess Tech.* **5**(4), 1121-1142.
- 74 Kulcke, A., Gurschler, C., Spöck, G., Leitner, R., Kraft, M. (2003) On-line classification of synthetic polymers using near infrared spectral imaging. *J. Near Infrared Spec.* **11**(1), 78-81.
- 75 Kobori, H., Yonenobu, H., Noma, J., Tsuchikawa, S. (2008) Discriminant analyzing system for wood wastes using a visible-near-infrared chemometric imaging technique. *Appl. Spectrosc.* **62**(8), 854-859.
- 76 Wienke, D., Van Den Broek, W., Melssen, W., Buydens, L., Feldhoff, R., Huth-Fehre, T., Kantimm, T., Winter, F., Cammann, K. (1996) Near-infrared imaging spectroscopy (NIRIS) and image rank analysis for remote identification of plastics in mixed waste. *Anal. Bioanal. Chem.* **354**(7-8), 823-828.
- 77 Salzer, R., Siesler, H.W. (2009) *Infrared and Raman spectroscopic imaging. Part four. Polymers and pharmaceuticals.* Wiley-VCH, Weinheim.
- 78 Khanmohammadi, M., Garmarudi, A.B., de la Guardia, M. (2012) Characterization of petroleum based products by infrared spectroscopy and chemometrics. *TrAC-Trend. Anal. Chem.* **35**, 135-149.
- 79 Zanier-Szydłowski, N., Quignard, A., Baco, F., Biguerd, H., Carpot, L., Wahl, F. (1999) Control of refining processes on mid-distillates by near infrared spectroscopy. *Oil Gas Sci. Technol.* **54**(4), 463-472.
- 80 López-García, C., Biguerd, H., Marchal-George, N., Schildknecht-Szydłowski, N. (2007) Near infrared monitoring of low conjugated diolefins content in hydrotreated FCC gasoline streams. *Oil Gas Sci. Technol.* **62**(1), 57-68.

- 81 Leontaritis, K.J., Leontaritis, J.D. (2003) Cloud point and wax deposition measurement techniques. Conference paper. *International Symposium on Oilfield Chemistry*, 5-7 February, 2003, Houston, TX. Society of Petroleum Engineers (SPE).
- 82 Paso, K., Kallevik, H., Sjöblom, J. (2009) Measurement of wax appearance temperature using near-infrared (NIR) scattering. *Energ. Fuel.*, 23(10), 4988-4994.
- 83 Aji, S., Schildknecht-Szydlowski, N., Faraj, A. (2004) Partial least square modelling for the control of refining processes on mid-distillates by near infrared spectroscopy. *Oil Gas Sci. Technol.* 59(3), 303-321.
- 84 Bueno, A., Baldrich, C.A., Molina D. (2009) Characterization of catalytic reforming streams by NIR spectroscopy. *Energ. Fuel.* 23(6), 3172-3177.
- 85 Hidajat, K., Chong, S.M. (2000) Quality characterisation of crude oils by partial least square calibration of NIR spectral profiles. *J. Near Infrared Spec.* 8, 53-59.
- 86 Alves, J.C.L., Poppi, R. J. (2012) Diesel oil quality parameter determinations using support vector regression and near infrared spectroscopy for hydrotreating feedstock monitoring. *J. Near Infrared Spec.* 20, 419-425.
- 87 Balabin, R.M., Safieva, R.Z. (2008) Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. *Fuel* 87, 1096-1101.
- 88 Lutz, O.M.D., Bon, G.K., Rode, B.M., Huck, C.W. (2014) Reproducible quantification of ethanol in gasoline *via* a customized mobile near-infrared spectrometer. *Anal. Chim. Acta* 826, 61-68.
- 89 Macho, S., Larrechi, M.S. (2002) Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *TrAC-Trend. Anal. Chem.* 21, 799-806.
- 90 Breitzkreitz, M.C., Raimundo, I.M., Rohwedder, J.J.R., Pasquini, C., Filho, H.A.D., José, G.E., Araújo, M.C.U. (2003) Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration. *Analyst* 128, 1204-1207.
- 91 Komorizono, I., de Carvalho W.F., Poppi R.J. (2009) Application of near infrared spectroscopy and multivariate control charts for monitoring biodiesel blends. *Anal. Chim. Acta* 642, 217-221.

- 92 Bezerra de Lira, L.F., de Vasconcelos, F.V.C., Pereira, C.F., Paim, A.P.S., Stragevitch, L., Pimentel, M.F. (2010) Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel* **89**, 405-409.
- 93 Pontes, M.J., Pereira, C.F., Pimentel, M.F., Vasconcelos, F.V., Silva, A.G. (2011) Screening analysis to detect adulteration in diesel/biodiesel blends using near infrared spectrometry and multivariate classification. *Talanta* **85**, 2159-2165.
- 94 De Vasconcelos, F.V.C., de Souza, P.F.B., Pimentel, M.F., Pontes, M.J.C., Pereira, C.F. (2012) Using near-infrared overtone regions to determine biodiesel content and adulteration of diesel/biodiesel blends with vegetable oils. *Anal. Chim. Acta* **716**, 101-107.
- 95 Baptista, P., Felizardo, P., Menezes, J.C., Correia, M.J. (2008) Multivariate near infrared spectroscopy models for predicting the methyl esters content in biodiesel. *Anal. Chim. Acta* **607(2)**, 153-159.
- 96 Baptista, P., Felizardo, P., Menezes, J.C., Correia, J.N. (2008) Multivariate near infrared spectroscopy models for predicting the iodine value, CFPP, kinematic viscosity at 40 °C and density at 15 °C of biodiesel. *Talanta* **77**, 144-151.
- 97 Balabin, R.M., Smirnov, S.V. (2011) Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. *Anal. Chim. Acta* **692**, 63-72.
- 98 Balabin, R.M., Lomakina, E.I., Safieva, R.Z. (2011) Neural network (ANN) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. *Fuel* **90**, 2007-2015.
- 99 Balabin, R.M., Safieva, R.Z. (2011) Near-infrared (NIR) spectroscopy for biodiesel analysis: Fractional composition, iodine value and cold filter plugging point from one vibrational spectrum. *Energ. Fuel.* **25**, 2373-2382.
- 100 Balabin, R.M., Safieva, R.Z., Lomakina, E.I. (2011) Near-infrared (NIR) spectroscopy for motor oil classification: From discriminant analysis to support vector machines. *Microchem. J.* **98**, 121-128.
- 101 Zamora, D., Blanco, M., Bautista, M. (2012) An analytical method for lubricant quality control by NIR spectroscopy. *Talanta* **89**, 478-483.

- 102 Blanco, M., Maspocho, S., Villaroya, I. Peralta, X., González, J.M., Torres, J. (2001) Determination of physical properties of bitumens by use of near-infrared spectroscopy with neural networks. Joint modelling of linear and non-linear parameters. *Analyst* **126**, 378-382.
- 103 Lima, F.S.G., Leite, L.F.M. (2004) Determination of asphalt cement properties by near infrared spectroscopy and chemometrics. *Pet. Sci. Technol.* **22(5-6)**, 589-600.
- 104 Pomerantsev, A.L., Rodionova, O.Y. (2012) Process analytical technology: A critical view of the chemometricians. *J. Chemometr.* **26**, 299-310.
- 105 Aaltonen, J., Gordon, K.C., Strachan, C.J., Rades, T. (2008) Perspectives in the use of spectroscopy to characterise pharmaceutical solids. *Int. J. Pharmaceut.* **364**, 159-169.
- 106 European Pharmacopoeia 8.0 (2014). Section 2.2.40: Near-infrared spectroscopy. *EDQM Council of Europe*, Estrasburgo, pp. 62-67.
- 107 United States Pharmacopeia, USP26NF21 (2003). Chapter 1119: Near-infrared spectrophotometry. *U.S. Pharmacopeial Convention*, Rockville, MD, pp. 2388-2391.
- 108 Real Farmacopea Española (2002) Sección 2.2.40: Espectrofotometría en el infrarojo cercano. 2ª ed., Agencia Española del Medicamento, Ministerio de Sanidad y Consumo, Madrid, pp. 57-58.
- 109 British Pharmacopoeia (2008) Appendix II A, Infrared Spectrophotometry: Near-infrared Spectrophotometry. British Pharmacopoeia Commission, *Medicines and Healthcare Products Regulatory Agency (MHRA)*, London, pp. A142-A146.
- 110 Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations (2004). *European Medicines Agency (EMA)*, London.
- 111 Reich, G. (2005) Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv. Drug Deliver. Rev.* **57**, 1109-1143.
- 112 Cogdill, R.P., Drennen, J.K. (2006) Near-infrared spectroscopy. En: Brittain, H.G. (ed.) *Spectroscopy of pharmaceutical solids*, Taylor & Francis Group, New York, NY, pp. 313-412.

- 113 Blanco, M., Romero, M.A. (2001) Near infrared libraries in the pharmaceutical industry: A solution for identity confirmation. *Analyst* **126**, 2212-2217.
- 114 Kemper, M.S., Luchetta, L.M. (2003) A guide to raw material analysis using near infrared spectroscopy. *J. Near Infrared Spec.* **11**(3), 155-174.
- 115 Simone, E., Saleemi, A.N., Nagy, Z.K. (2014) Raman, UV, NIR, and mid-IR spectroscopy with focused beam reflectance measurement in monitoring polymorphic transformations. *Chem. Eng. Technol.*, **37**(8), 1305-1313.
- 116 El-Hagrasy, A., Morris, H.R., D'Amico, F., Lodder, R.A., Drennen III, J.K. (2001) Near-Infrared Spectroscopy and Imaging for the Monitoring of Powder Blend Homogeneity. *J. Pharm. Sci.*, **90**(9), 1298-1307.
- 117 Wu, H., Tawakkul, M., White, M.M., Khan, M.A. (2009) Quality-by-Design (QbD): An integrated multivariate approach for the component quantification in powder blends. *Int. J. Pharmaceut.* **372**(1-2), 39-48.
- 118 Flaten, G.R., Ferreira, A.P., Bellamy, L., Frake, P. (2012) PAT within the QbD framework: Real-time end point detection for powder blends in a compliant environment. *J. Pharm. Innov.* **7**, 38-45.
- 119 Rosas, J.G., Blanco, M., Santamaría, F., Alcalà, M. (2013) Assessment of chemometric methods for the non-invasive monitoring of solid blending processes using wireless near infrared spectroscopy. *J. Near Infrared Spec.* **21**, 97-106.
- 120 Blanco, M., Alcalà, M., González, J.M., Torras, E. (2006) Near infrared spectroscopy in the study of polymorphic transformations. *Anal. Chim. Acta* **567**, 262-268.
- 121 Li, W., Cunningham, J., Rasmussen, H., Winsted, D. (2007) A qualitative method for monitoring of nucleation and granule growth in fluid bed wet granulation by reflectance near-infrared spectroscopy. *J. Pharm. Sci.* **96**(12), 3470-3477.
- 122 Alcalà, M., Blanco, M., Bautista, M., González, J.M. (2010) On-line monitoring of a granulation process by NIR spectroscopy. *J. Pharm. Sci.* **99**(1), 336-345.
- 123 Peinado, A., Hammond, J., Scott, A. (2011) Development, validation and transfer of a Near Infrared method to determine in-line the end point of a fluidised drying process for commercial production batches of an approved oral solid dose pharmaceutical product. *J. Pharm. Biomed. Anal.* **54**, 13-20.

- 124 Fonteyne, M., Arruabarrena, J., De Beer, J., Hellings, M., Van Den Kerkhof, T., Burggraeve, A., Vervaet, C., Remon, J.P., De Beer, T. (2014) NIR spectroscopic method for the in-line moisture assessment during drying in a six-segmented fluid bed dryer of a continuous Tablet production line: Validation of quantifying abilities and uncertainty assessment. *J. Pharm. Biomed. Anal.* **100**, 21-27.
- 125 De Beer, T.R.M., Vercruyssen, P., Burggraeve, A., Quinten, T., Ouyang, J., Zhang, X., Vervaet, C., Remon, J.P., Baeyens, W.R.G. (2009) In-line and real-time process monitoring of a freeze-drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools. *J. Pharm. Sci.* **98**, 3430-3444.
- 126 Pieters, S., De Beer, T., Kasper, J.C., Boulpaep, D., Waszkiewicz, O., Goodarzi, M., Tistaert, C., Friess, W., Remon, J.P., Vervaet, C., Van Der Heyden, Y. (2012) Near-Infrared Spectroscopy for In-Line Monitoring of Protein Unfolding and Its Interactions with Lyoprotectants during Freeze-Drying. *Anal. Chem.* **84**, 947-955.
- 127 Wahl, P.R., Fruhmann, G., Sacher, S., Straka, G., Sowinski, S., Khinast, J.G. (2014) PAT for tableting: Inline monitoring of API and excipients via NIR spectroscopy. *Eur. J. Pharm. Biopharm.* **87(2)**, 271-278.
- 128 Lee, M.J., Seo, D.Y., Lee, H.E., Wang, I.C., Kim, W.S., Jeong, M.Y., Choi, G.J. (2011) In line NIR quantification of film thickness on pharmaceutical pellets during a fluid bed coating process. *Int. J. Pharmaceut.* **403**, 66-72.
- 129 Römer, M., Heinämäki, J., Strachan, C., Sandler, N., Yliruusi, J. (2008) Prediction of tablet film-coating thickness using a rotating plate coating system and NIR spectroscopy. *AAPS PharmSciTech.* **9(4)**, 1047-1053.
- 130 Tabasi, S.H., Fahmy, R., Bensley, D., O'Brien, C., Hoag, S.W. (2008) Quality by Design, Part II: Application of NIR Spectroscopy to Monitor the Coating Process for a Pharmaceutical Sustained Release Product. *J. Pharm. Sci.* **97(9)**, 4052-4066.
- 131 Lee, M.J., Park, C.R., Youngkim, A., Soookwon, B., Bang, K.H., Cho, Y.S., Jeong, M.Y., Choi, G.J. (2010) Dynamic calibration for the in-line NIR monitoring of film thickness of pharmaceutical tablets processed in a fluid-bed coater. *J. Pharm. Sci.* **99(1)**, 325-335.

- 132 Yoon, W.L., Jee, R.D., Charvill, A., Lee, G., Moffat, A.C. (2004) Application of near-infrared spectroscopy to the determination of the sites of manufacture of proprietary products. *J. Pharm. Biomed. Anal.* **34**,933-944.
- 133 Shah, R., Tawakul, M., Khan, M.A. (2007) Process analytical technology: Chemometric analysis of Raman and near infra-red spectroscopic data for predicting physical properties of extended release matrix tablets. *J. Pharm. Sci.* **96**(5), 1356-1365.
- 134 Blanco, M., Alcalà, M., González, J.M., Torras, E. (2006) A process analytical technology approach based on near infrared spectroscopy: Tablet hardness, content uniformity and dissolution test measurements of intact tablets. *J. Pharm. Sci.* **95**(1), 2137-2144.
- 135 Donoso, M., Ghaly, E.S. (2004) Prediction of drug dissolution from tablets using near-infrared diffuse reflectance spectroscopy as a nondestructive method. *Pharm. Dev. Technol.* **9**(3), 247-263.
- 136 Mattes, R., Root, D., Zhou, Y., Xiao, C., Johnson, J., Shukla, A. (2009) NIR prediction of solid dosage form dissolution profiles. *NIR News* **20**(5), 10-11.
- 137 Blanco, M., Peguero, A. (2010) Analysis of pharmaceuticals by NIR spectroscopy without a reference method. *TrAC-Trend. Anal. Chem.* **29**(10), 1127-1136.
- 138 Westenberger, B.J., Ellison, C.D., Fussner, A.S., Jenney, S., Kolinski, R.E., Lipe, T.G., Lyon, R.C., Moore, T.W., Reville, L.K., Smith, A.P., Spencer, J.A., Story, K.D., Toler, D.Y., Wokovich, A.M., Buhse, L.F. (2005) Quality assessment of internet pharmaceutical products using traditional and non-traditional analytical techniques. *Int. J. Pharmaceut.* **306**(1-2), 56-70.
- 139 Blanco, M., Peguero, A. (2008) An expeditious method for determining particle size distribution by near infrared spectroscopy: Comparison of PLS2 and ANN models. *Talanta* **77**, 647-651.
- 140 Stokvold, A., Dyrstad, K., Libnau, F.O. (2002) Sensitive NIRS measurement of increased moisture in stored hygroscopic freeze dried product. *J. Pharm. Biomed. Anal.* **2**, 867-873.
- 141 Zheng, Y., Lai, X., Bruun, S.W., Ipsen, H., Larsen, J.N., Løwenstein, H., Søndergaard, I., Jacobsen, S. (2008) Determination of moisture content of lyophilized allergen vaccines by NIR spectroscopy. *J. Pharm. Biomed. Anal.* **46**, 592-596.

- 142 Grohganz, H., Gildemyn, D., Skibsted, E., Flink, J.M., Rantanen, J. (2010) Towards a robust water content determination of freeze-dried samples by near-infrared spectroscopy. *Anal. Chim. Acta* **676**, 34-40.
- 143 Grohganz, H., Fonteyne, M., Skibsted, E., Falck, T., Palmqvist, B., Rantanen, J. (2010) Classification of lyophilised mixtures using multivariate analysis of NIR spectra. *Eur. J. Pharm. Biopharm.* **74**, 406-412.
- 144 Grohganz, H., Gildemyn, D., Skibsted, E., Flink, J.M., Rantanen, J. (2011) Rapid solid-state analysis of freeze-dried protein formulations using NIR and Raman spectroscopies. *J. Pharm. Sci.* **100(7)**, 2871-2875.
- 145 Blanco, M., Alcalà, M., Bautista, M. (2008) Pharmaceutical gel analysis by NIR spectroscopy. Determination of the active principle and low concentration of preservatives. *Eur. J. Pharm. Sci.* **33**, 409-414.
- 146 Rosas, J.G., Blanco, M., González, J.M., Alcalà, M. (2011) Quality by Design Approach of a Pharmaceutical Gel Manufacturing Process, Part 2: Near Infrared Monitoring of Composition and Physical Parameters. *J. Pharm. Sci.* **100(10)**, 4442-4451.
- 147 Gowen, A.A., O'Donnell, C.P., Cullen, P. J., Bell, S.E.J. (2007) Recent applications of Chemical Imaging to pharmaceutical process monitoring and quality control. *Eur. J. Pharm. Biopharm.* **69**, 10-22.
- 148 Malik, I., Poonacha, M., Moses, J., Lodder, R.A. (2001) Multispectral imaging of tablets in blister packaging. *AAPS PharmSciTech.* **2(2)**, 38-44.
- 149 Hamilton, S., Lodder, R.A. (2002) Hyperspectral imaging technology for pharmaceutical analysis. *Proc. SPIE 4626, Biomedical Nanotechnology Architectures and Applications*, **136**.
- 150 Vercruyssen, J., Toiviainen, M., Fonteyne, M., Helkimo, N., Ketolainen, J., Juuti, M., Delaet, U., Van Assche, I., Remon, J.P., Vervaet, C., De Beer, T. (2014) Visualization and understanding of the granulation liquid mixing and distribution during continuous twin screw granulation using NIR chemical imaging. *Eur. J. Pharm. Biopharm.* **86(3)**, 383-392.
- 151 Li, W., Woldu, A., Kelly, R., McCool, J., Bruce, R., Rasmussen, H., Cunningham, J., Winstead, D. (2008) Measurement of drug agglomerates in powder blending simulation samples by near infrared chemical imaging. *Int. J. Pharmaceut.* **350**, 369-373.

- 152 Gendrin, C., Roggo, Y., Spiegel, C., Collet, C. (2008) Monitoring galenical process development by near infrared chemical imaging: One case study. *Eur. J. Pharm. Biopharm.* **68(3)**, 828-837.
- 153 Hilden, L.R., Pommier, C.J., Badawy, S.I.F., Friedman, E.M. (2008) NIR chemical imaging to guide/support BMS-561389 tablet formulation development. *Int. J. Pharmaceut.* **353(1-2)**, 283-290.
- 154 Ellison, C.D., Ennis, B.J., Hamad, M.L., Lyon, R.C. (2008) Measuring the distribution of density and tableting force in pharmaceutical tablets by chemical imaging. *J. Pharm. Biomed. Anal.* **48(1)**, 1-7.
- 155 Clarke, F. (2004) Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. *Vib. Spectrosc.* **34(1)**, 25-35.
- 156 Ma, H., Anderson, C. A. (2008) Characterization of pharmaceutical powder blends by NIR chemical imaging. *J. Pharm. Sci.* **97**, 3305-3320.
- 157 Rosas, J.G., Blanco, M. (2012) A criterion for assessing homogeneity distribution in hyperspectral images. Part 1: Homogeneity index bases and blending processes. *J. Pharm. Biomed. Anal.* **70**, 680-690.
- 158 Šašić, S. (2008) Chemical imaging of pharmaceutical granules by Raman global illumination and near-infrared mapping platforms. *Anal. Chim. Acta* **611(1)**, 73-79.
- 159 Maurer, L., Leuenberger, H. (2009) Terahertz pulsed imaging and near infrared imaging to monitor the coating process of pharmaceutical tablets. *Int. J. Pharmaceut.* **370**, 8-16.
- 160 Sabin, G.P., Breitzkreitz, M.C., de Souza, A.M., da Fonseca, P., Calefe, L., Moffa, M., Poppi, R.J. (2011) Analysis of pharmaceutical pellets: An approach using near-infrared chemical imaging. *Anal. Chim. Acta* **706**, 113-119.
- 161 Cairós, C., Amigo, J.M., Watt, R., Coello, J., Maspocho, S. (2008) Implementation of enhanced correlation maps in near infrared chemical images: Application in pharmaceutical research. *Talanta* **79**, 657-664.
- 162 Gendrin, C., Roggo, Y., Collet, C. (2007) Content uniformity of pharmaceutical solid dosage forms by near infrared hyperspectral imaging: A feasibility study. *Talanta* **73**, 733-741.

- 163 Cruz, J., Blanco, M. (2011) Content uniformity studies in tablets by NIR-CI. *J. Pharm. Biomed. Anal.* **56**(2), 408-412.
- 164 Cruz, J., Bautista, M., Amigo, J. M., Blanco, M. (2009) NIR-chemical imaging study of acetylsalicylic acid in commercial tablets. *Talanta* **80**, 473-478.
- 165 Rosas, J.G., Blanco, M. (2012) A criterion for assessing homogeneity distribution in hyperspectral images. Part 2: Application of homogeneity indices to solid pharmaceutical dosage forms. *J. Pharm. Biomed. Anal.* **70**, 691-699.
- 166 Lopes, M.B., Wolff, J.C., Bioucas-Dias, J.M., Figueiredo, M.A. (2009) Determination of the composition of counterfeit Heptodin tablets by near infrared chemical imaging and classical least squares estimation. *Anal. Chim. Acta* **641**, 46-51.
- 167 Puchert, T., Lochmann, D., Menezes, J.C., Reich, G. (2010) Near-infrared chemical imaging (NIR-CI) for counterfeit drug identification—A four-stage concept with a novel approach of data processing (Linear Image Signature). *J. Pharm. Biomed. Anal.* **51**, 138-145.
- 168 Ishikawa, D., Murayama, K., Awa, K., Genkawa, T., Komiyama, M., Kazarian, S.G., Ozaki, Y. (2013) Application of a newly developed portable NIR imaging device to monitor the dissolution process of tablets. *Anal. Bioanal. Chem.* **405**, 9401-9409.
- 169 De Carvalho Rocha, W.F., Sabin, G.P., Março, P.H., Poppi, R.J. (2011) Quantitative analysis of piroxicam polymorphs pharmaceutical mixtures by hyperspectral imaging and chemometrics. *Chemometr. Intell. Lab.* **106**(2), 198-204.
- 170 Brondi, A.M., Terra L.A., Sabin, G.P., Garcia, J.S., Poppi, R.J., Trevisan, M.G. (2014) Mapping the polymorphic forms of fexofenadine in pharmaceutical tablets using near infrared chemical imaging. *J. Near Infrared Spec.* **22**(3), 211-220.
- 171 Jovanovic, N., Gerich, A., Bouchard, A., Jiskoot, W. (2006) Near-infrared imaging for studying homogeneity of protein-sugar mixtures. *Pharm. Res.* **23**(9), 2002-2013.
- 172 Zidan, A., Spinks, C., Fortunak, J., Habib, M., Khan, M. (2010). Near-infrared investigations of novel anti-HIV tenofovir liposomes. *AAPS J.* **12**(2), 202-214.

3. QUIMIOMETRÍA

3.1. Introducción	123
3.1.1. Orígenes de la quimiometría	123
3.1.2. Notación.....	124
3.2. Métodos cualitativos	125
3.2.1. Análisis en Componentes Principales (PCA)	126
3.2.2. Coeficientes de correlación (CC).....	129
3.3. Métodos cuantitativos	131
3.3.1. Calibración univariante	133
3.3.2. Calibración multivariante.....	134
3.4. Etapas del proceso de modelado	139
3.4.1. Diseño, selección y preparación de muestras.....	140
3.4.2. Métodos de referencia.....	141
3.4.3. Obtención de la señal analítica	142
3.4.4. Preprocesamiento y extracción de datos de imágenes químicas.....	143
3.4.5. Pretratamientos espectrales.....	145
3.4.6. Selección del grupo de calibración/validación.....	153
3.4.7. Construcción del modelo de calibración	155
3.4.8. Evaluación del modelo	156
3.4.9. Elección del número de factores	158
3.4.10. Aplicación del modelo e interpretación de los resultados.....	159
REFERENCIAS	160

3. QUIMIOMETRÍA

3.1. Introducción

La información analítica contenida en los espectros NIR, típicamente con bandas anchas y solapadas es poco selectiva y muy influenciada por un gran número de factores físicos, químicos y estructurales, por lo que las diferencias espectrales debidas a las variables de interés pueden ser muy ligeras y difíciles de observar a simple vista. Estas dificultades, multiplicadas por la gran cantidad de espectros que pueden ser obtenidos en poco tiempo con los instrumentos actuales de NIRS y en especial de NIR-CI, hacen de la interpretación de este tipo de datos una tarea muy compleja. Es por esta razón, que el éxito del NIR no hubiera sido posible sin su combinación con la quimiometría [1], que permite procesar de forma rápida y eficiente grandes conjuntos de datos.

3.1.1. Orígenes de la quimiometría

El término quimiometría fue acuñado por Svante Wold en 1971 [2,3]. Solo tres años después se formaba la *International Chemometrics Society* (ICS), que describe la quimiometría como “la disciplina que relaciona las medidas realizadas en sistemas o procesos químicos con el estado del sistema mediante la aplicación de métodos matemáticos o estadísticos”. Pero, de entre la gran cantidad de definiciones del término quimiometría que se pueden encontrar hoy día en la bibliografía [4,5,3], la más completa y también más popular es la propuesta por Massart *et al.* [6] que considera la quimiometría como “la disciplina química que utiliza métodos matemáticos, estadísticos y otros métodos basados en la lógica formal para: (1) diseñar o seleccionar procedimientos de medida y experimentos óptimos; y (2) proporcionar el máximo de información química relevante mediante el análisis de datos químicos”. Esta definición engloba métodos, principalmente multivariados, aplicables a distintas etapas del modelado de datos, un proceso que se describe en los apartados que siguen.

3.1.2. Notación

Para facilitar el entendimiento de este apartado, a continuación se puntualizan algunos detalles sobre la notación utilizada. Los escalares se representan en minúsculas en cursiva (p. ej. x), mientras que los vectores, y las matrices se simbolizan en minúscula y mayúscula negritas respectivamente (p. ej. x y \mathbf{X}). Los datos espectrales (variables dependientes) se denotan por defecto con la letra “ x ”, mientras que la “ y ” se utiliza para las respuestas (variables independientes). Las dimensiones de las matrices de datos espectrales serán en todos los casos $\mathbf{X}(M \times N)$, donde M corresponde al número de filas de la matriz (equivalente al número de muestras o espectros de trabajo) y N representa el número de columnas de la matriz, es decir, las variables, que en el caso del NIR siempre son longitudes o números de onda. Por lo que respecta a la matriz de respuestas, sus dimensiones corresponden con $\mathbf{Y}(M \times I)$, donde I es el número de respuestas o clases del sistema. Otros símbolos ampliamente utilizados son: \bar{x} para denotar el promedio, σ como símbolo de la desviación estándar y superíndices como una “ t ” o un -1 (\mathbf{X}^t o \mathbf{X}^{-1}) para expresar la transpuesta de una matriz y su inversa respectivamente. El resto de nomenclatura utilizada se describe en el apartado que corresponde.

3.2. Métodos cualitativos

La identificación (reconocimiento inequívoco) y la caracterización (reconocimiento de algunos atributos específicos) de una muestra a partir de su espectro son ejemplos de los análisis cualitativos más comunes [7; **Error! Marcador no definido.**]. Las técnicas quimiométricas utilizadas para llevarlos a cabo se conocen como métodos de reconocimiento de pautas (PRM, *Pattern Recognition Methods*) y funcionan estableciendo un criterio matemático que permite expresar cuantitativamente la similitud entre una muestra y una referencia o entre una muestra y una clase [1]. Los PRM pueden clasificarse en “supervisados” y “no supervisados” dependiendo de si se tiene conocimiento previo o no de las muestras. En la Figura 3.1 se citan algunos de los métodos cualitativos más utilizados de cada tipo. Los métodos no supervisados suelen utilizarse como un primer análisis exploratorio de muestras cuya clasificación o agrupación es desconocida *a priori*. Los métodos supervisados, en cambio, permiten la construcción de ciertas reglas de clasificación en base a las características observadas en el grupo de calibración y aplicarlas posteriormente a muestras desconocidas. Una de las grandes aplicaciones de los PRM supervisados es la creación de bibliotecas espectrales para identificación y cualificación de muestras desconocidas [8]. A partir de espectros de muestras conocidas se crean clases representativas para cada identidad química (p. ej. celulosa, lactosa, etc.) y características concretas (p. ej. tipo de polimorfo, tamaño de partícula, etc.). La identificación de una muestra desconocida se lleva a cabo por comparación de su espectro con cada una de las clases de la biblioteca.

A parte de llevar a cabo análisis cualitativos por sí mismos, en muchos casos los análisis cualitativos se aplican a modo exploratorio, para la obtención de información sobre el sistema, como paso previo al desarrollo de un modelo cuantitativo.



Figura 3.1. Representación jerárquica de algunos de los PRM utilizados en análisis cualitativos por NIR [8,9].

3.2.1. Análisis en Componentes Principales (PCA)

El análisis en componentes principales (PCA, *Principal Component Analysis*) es una de las técnicas multivariantes más utilizadas en quimiometría [10,11], por lo que se encuentra descrita en gran cantidad de bibliografía [7,10,12,13,14]. Se trata de un método de reducción de variables, pero sin pérdida de información de interés, una característica muy útil dada la cantidad de variables que se manejan en los análisis NIR y su elevada colinealidad. Por defecto, el PCA suele venir precedido por el pretratamiento de centrado a la media (apartado 3.4.5) [7,14], como se representa en la Figura 3.2a. La compresión de los datos se consigue por descomposición matemática de las variables correlacionadas originales (absorbancia a muchas longitudes de onda, \mathbf{X}) en un conjunto mucho menor de variables latentes (\mathbf{A}) no correlacionadas, denominadas componentes principales (PC, *Principal Components*). La matriz de datos espectrales $\mathbf{X}(M \times N)$ se descompone en sus PCs mediante la Ecuación 3.1:

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} \quad \text{Ecuación 3.1}$$

donde $T(M \times A)$ es la matriz de *scores* de los PCs, $P(A \times N)$, la matriz de *loadings* y $E(M \times N)$, la matriz de residuales del modelo. Los *scores* son las intensidades de las nuevas variables A para cada muestra, los *loadings* describen las nuevas variables A en términos de las variables originales y los residuales corresponden a la variabilidad de los datos originales no explicada por el modelo (Figura 3.2b). Por lo que respecta a los PCs, se trata de combinaciones lineales de las variables originales, ortogonales entre ellas, que actúan como nuevos ejes de coordenadas del sistema. La ortogonalidad de los PCs es lo que permite eliminar la correlación intrínseca de las variables originales. Otra característica de los PCs es que se escogen en función de la dirección que recoge mayor variabilidad del conjunto de muestras. Así, los componentes se ordenan según la varianza que explican: el primer PC acumula la máxima variabilidad espectral del conjunto de muestras, el segundo recoge una variabilidad inferior al primero (pero superior al tercero) y así sucesivamente. De esta forma, el estudio de las causas de variabilidad de un grupo de muestras es mucho más fácil e intuitivo, ya que los primeros PCs explican los efectos sistemáticos de los datos y, a partir de cierto número de componentes, la varianza se debe principalmente al ruido. La selección del número de PCs es un equilibrio entre la incorporación de la máxima variabilidad de las muestras al modelo y el mínimo ruido, lo que normalmente suele conseguirse alrededor del 95% de la varianza explicada. En la Ecuación 3.2 se muestra la matriz resultante de escoger a componentes:

$$\mathbf{X} = \mathbf{T}_1\mathbf{P}_1^t + \mathbf{T}_2\mathbf{P}_2^t + \dots + \mathbf{T}_a\mathbf{P}_a^t + \mathbf{E} \quad \text{Ecuación 3.2}$$

Toda la variación no descrita por los primeros a PCs del modelo se encuentra en la matriz de residuales (\mathbf{E}). La elección del número de PCs no siempre es obvia dado que cada componente no se corresponde con un solo analito o propiedad. Además, en casos donde la variabilidad del parámetro de interés supone una parte muy baja de la varianza total del modelo, su información puede confundirse con ruido.

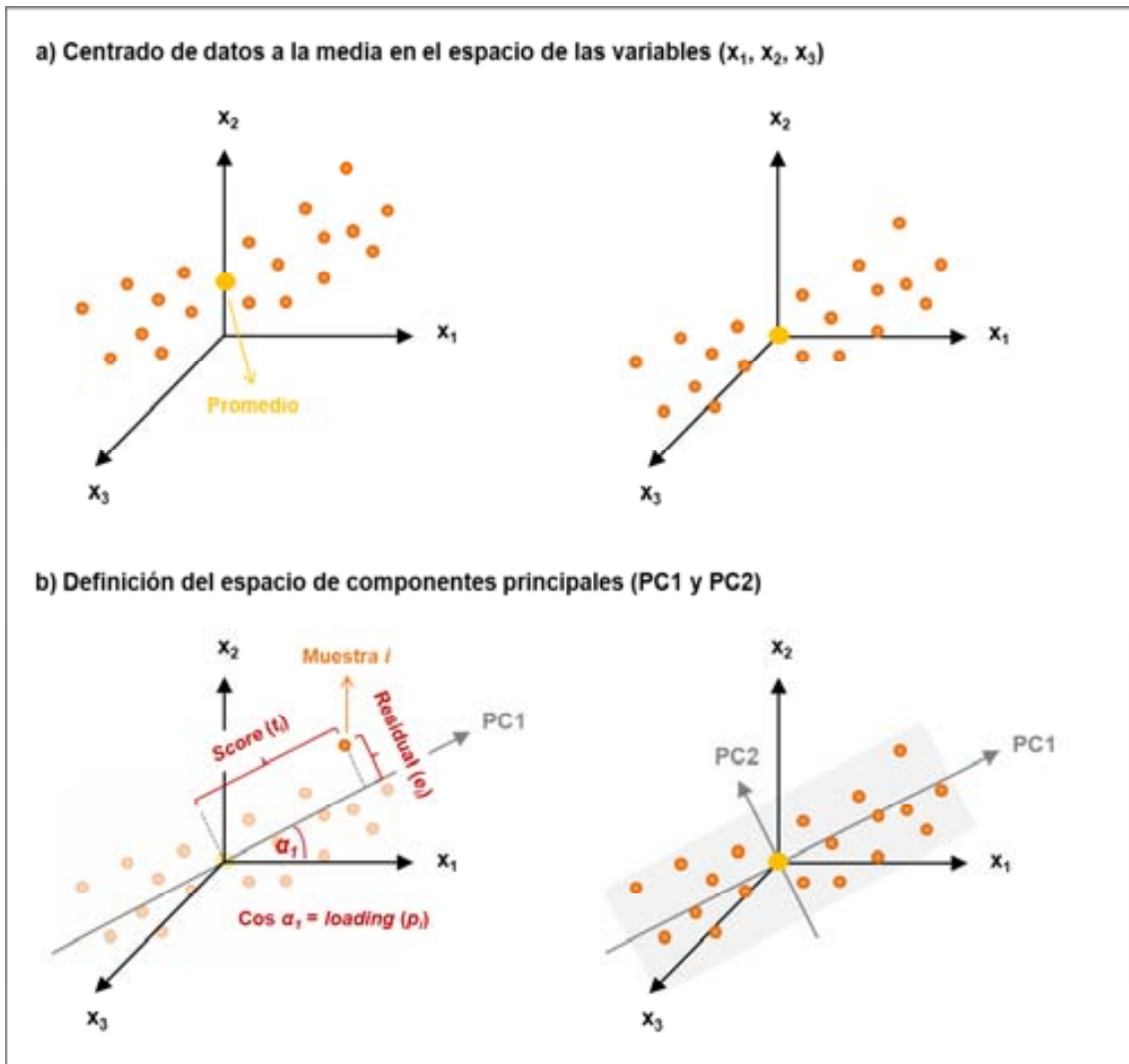


Figura 3.2. Aplicación del algoritmo PCA (2 PCs) a un conjunto de datos definido por 3 variables (x_1 , x_2 y x_3).

Los objetivos principales del PCA son simplificar y explorar los datos, lo que se consigue mediante los gráficos de *scores* y *loadings* del PCA. La exploración de las relaciones entre muestras según las tendencias y agrupaciones que presentan y su posterior clasificación [10,12,14,15] se lleva a cabo a partir de una representación de los *scores* de las muestras de un PC frente a los *scores* de otro/s componente/s. Si existe una relación entre muestras, estas aparecerán agrupadas en el gráfico de *scores*. Por el contrario, muestras alejadas indican baja similitud, una característica que se

utiliza para la detección de posibles *outliers*¹ espectrales [14]. Este mismo gráfico también se utiliza para la selección de conjuntos de calibración [13], dado que escogiendo un grupo de muestras que cubran todo el espacio de los *scores* se puede asegurar una selección de muestras representativas del conjunto global (apartado 3.4.6). Otra aplicación del gráfico de *scores* es la monitorización de procesos [10,16,17], que también se puede llevar a cabo representando los *scores* de un solo PC frente al tiempo [15,17]. Además, los valores de los *scores* de distintos PCs se han utilizado como variables de partida para otros algoritmos de modelado [18]. Por lo que respecta a los *loadings*, la representación de los valores que toman en un PC frente a otro, posibilita el análisis de las relaciones entre variables [12,14]. En este caso, las variables más próximas entre ellas tendrán comportamientos correlacionados y las más dispersas, tendencias inversamente correlacionadas, mientras que la distancia al origen de las coordenadas, indica el impacto que tienen en el modelo: a mayor distancia, mayor es el impacto [14]. En cambio, cuando se representan los *loadings* de un PC respecto a las distintas variables se obtienen los pesos que toma cada variable, pudiendo determinar qué tipo de información contiene cada PC del modelo [16,17]. La superposición de ambos gráficos (*scores* y *loadings*) en un *bi-plot*, muestra qué variables tienen más peso en la disposición de las muestras [10].

Para identificar y/o caracterizar nuevas muestras, estas pueden proyectarse en un modelo PCA existente para predecir su posición en el gráfico de *scores* y determinar su parecido con las muestras incluidas en el modelo.

3.2.2. Coeficientes de correlación (CC)

El coeficiente de correlación (CC) es un índice estadístico que mide el grado de relación entre dos variables cuantitativas y se define como la relación entre la covarianza (Cov) y las desviaciones estándar (σ) de cada una de las variables (Ecuación 3.3) [19]. En espectroscopia NIR, el CC se utiliza como medida de similitud espectral

¹ Los *outliers* son muestras con errores de registro (*outliers* espectrales), errores en el valor de referencia o con características muy distintas del resto de muestras analizadas.

entre una muestra y una referencia. En este caso, su ecuación se puede escribir como se indica en la Ecuación 3.4, donde x_j e y_j son los valores de absorbancia a la longitud de onda j para el espectro de la muestra, $x = (x_1, \dots, x_j, \dots, x_n)$, y el espectro de referencia, $y = (y_1, \dots, y_j, \dots, y_n)$, respectivamente. El CC puede tomar valores entre $|0-1|$, indicando mayor similitud espectral cuanto más cercano a uno está.

$$CC = \frac{Cov(x,y)}{\sigma_x \sigma_y} \quad \text{Ecuación 3.3}$$

$$CC = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2 \sum_{j=1}^n (y_j - \bar{y})^2}} \quad \text{Ecuación 3.4}$$

El CC es esencialmente un parámetro cualitativo utilizado como criterio discriminante en bibliotecas espectrales con fines identificativos [20,21]. En este caso, se construyen clases con los espectros seleccionados. Para determinar si una muestra problema pertenece a una clase concreta, se calcula el CC entre el espectro de la muestra y el centro de calibración de la clase y se compara el resultado obtenido con el límite de discriminación establecido previamente. Pero el CC también se ha aprovechado en otro tipo de aplicaciones, principalmente en NIR-CI, en lo que se conoce como mapas de correlación [22,23], pero también en NIRS [24] e, incluso, en aplicaciones cuantitativas [25].

3.3. Métodos cuantitativos

Los métodos cuantitativos establecen relaciones cuantitativas entre los datos de medida con un instrumento y los valores conocidos de alguna propiedad de la muestra (datos de referencia). Estas relaciones deben permitir la predicción de esta propiedad cuantitativamente y de forma exacta y precisa en muestras desconocidas (sin valor de referencia), usando solamente su espectro NIR [26].

El principio básico que rige las aplicaciones espectroscópicas cuantitativas es el hecho que la magnitud medida es proporcional a la concentración del analito y que las constantes de proporcionalidad deben ser determinadas experimentalmente [7]. Este proceso, conocido como calibración, requiere la intervención de dos tipos de variables, las independientes (absorbancias medidas a las distintas longitudes de onda, \mathbf{X}) y las dependientes (parámetro o propiedad a predecir, \mathbf{Y}). La ecuación matemática resultante es el modelo de calibración cuantitativo que será utilizado para la determinación cuantitativa de nuevas muestras.

A pesar de que la espectroscopia actual se caracteriza por la obtención simultánea de información de un gran número de variables independientes (\mathbf{X}), en algunos casos es suficiente utilizar sólo algunas de ellas. En función del número de variables utilizadas para construir el modelo, este será univariante (una sola variable) o multivariante (más de una variable) y los algoritmos matemáticos utilizados diferirán. En la Tabla 3.1 se comparan algunos de los algoritmos univariantes y multivariantes lineales más comunes. Pero la relación entre la señal NIR y el parámetro de interés no siempre es completamente lineal. Las causas más comunes de esta no linealidad pueden ser la respuesta del detector, derivas de la fuente de luz, así como factores químicos o físicos de la muestra (p. ej. el tamaño de partícula o interacciones entre componentes que modifican la absorción en función de la concentración) o externos a ella (como presión y temperatura ambientales), que fomentan el desplazamiento de bandas del espectro [27]. En algunos casos, una ligera no linealidad se puede corregir con ciertos pretratamientos o añadiendo un factor de más a un modelo lineal, pero cuando la tendencia es fuertemente no lineal, requiere del uso de modelos no lineales [9] como las redes neuronales artificiales (ANN,

Artificial Neural Networks) [28], la regresión local ponderada (LWR, *Locally Weighted Regression*) [29], las máquinas de soporte vectorial (SVM, *Supported Vector Machines*) [30] o versiones no lineales del PLS [31].

Tabla 3.1. Comparativa de los aspectos principales de algunos de los métodos cuantitativos lineales más comunes.

Algoritmo	Tamaño de datos	Datos de partida	Referencia y grupo de calibración	Objetivo	Ventajas	Limitaciones
Univariante [7]	1 sola variable	Valor de absorbancia a una sola long de onda	Sí	Determinación cuantitativa de un parámetro de interés	Rápido y sencillo	Requiere que el analito a cuantificar sea conocido y presente una banda no solapada
Regresión lineal múltiple (MLR)^a [12,13]	□ n° variables (< n° muestras)	Espectros mezcla	Sí	Determinación cuantitativa de un parámetro de interés	Rápido y sencillo	N° de variables limitado No admite correlación Tendencia al sobreajuste cuando hay ruido
Regresión clásica por mínimos cuadrados (CLS)^b [32,33]	□ n° variables	Espectros puros de todos los comp.	NO	Determinación de conc.	Permite multideterminaciones simultáneas	Es necesario conocer todos los comp. de la muestra y disponer de su espectro puro No admite correlación ni comportamientos no lineales
Regresión de componentes principales (PCR) [12,13]	□ n° variables (trabaja con variables reducidas)	Espectros mezcla	Sí	Determinación cuantitativa de un parámetro de interés	Posible detección de <i>outliers</i>	Sólo tiene en cuenta la varianza en X para la reducción de var. Hay que elegir n°factores
Regresión parcial por mínimos cuadrados (PLS)^c [14,34]	□ n° variables (trabaja con variables reducidas)	Espectros mezcla	Sí	Determinación cuantitativa de uno (PLS1) o más parámetros de interés (PLS2)	Modela tanto la varianza de X como de Y Posible detección de <i>outliers</i>	Requiere un buen conjunto de calibración Hay que elegir n° factores
Resolución multivariante de curvas (MCR-ALS) [35,36]	□ n° variables (trabaja con variables reducidas)	Admite información de ≠ tipos: espectros puros, conc., estimac.,	NO	Estimación de los espectros puros de la mezcla y seguimiento de sus conc.	□ información inicial necesaria (no requiere info. de todos los comp.)	Método iterativo lento sin solución única No se puede escoger el número de comp. (estimado por el algoritmo)

		restricciones.				
Isys-PLS1 [37]	∞ n° variables	Espectros puros de los comp.a cuantificar	NO	Determinación de conc.	No es necesario conocer todos los comp.	Requiere crear una biblioteca espectral con los comp. a cuantificar

^a El MLR también se denomina como *regresión inversa por mínimos cuadrados (ILS, Inverse Least Squares)* [13]

^b El CLS equivale a la variante *PLS-class del PLS discriminante (PLS-DA, PLS-Discriminant Analysis)* [33]

^c El PLS también se conoce como *proyección de estructuras latentes (PLS, Projection to Latent Structures)*

En este apartado se describen aquellos algoritmos utilizados para el desarrollo de los distintos modelos presentados en esta tesis, lo que incluye algoritmos univariantes y multivariantes, pero lineales en todos los casos.

3.3.1. Calibración univariante

Es el algoritmo de calibración más sencillo, ya que se basa en una única variable, lo que es de gran utilidad en técnicas que proporcionan una única señal analítica. En el caso del NIR, a pesar del gran número de variables que presentan los espectros, los métodos univariantes son muy útiles cuando el analito o propiedad de interés son conocidos y se identifican con una banda específica que no se superpone con otros picos del espectro. Pero dada la baja selectividad de NIR esto es poco común, de forma que la variable utilizada para la regresión no suele tratarse del valor de absorbancia a una sola longitud de onda. Normalmente suelen emplearse otro tipo de valores discretos que maximizan la selectividad, como puede ser la diferencia o la relación entre absorbancias o intensidades a dos longitudes de onda distintas [38] o el CC (apartado 3.2.2) del espectro de la muestra respecto un espectro de referencia [25]. El modelo se obtiene relacionando los valores de la variable elegida para cada muestra i (x_i) con sus valores de referencia correspondientes (y_i) mediante regresión por mínimos cuadrados:

$$y_i = b_0 + b_1x_i + e_i \quad \text{Ecuación 3.5}$$

donde b son los coeficientes de regresión que corresponden con la ordenada al origen (b_0) y el pendiente (b_1), y e es el error asociado a la medida. De esta forma se determinan aquellos valores de b que minimizan el sumatorio al cuadrado de los residuales [7,12]. La predicción de nuevas muestras se consigue por interpolación de los nuevos valores de absorbancia en la recta de calibración. A pesar de su sencillez, los modelos univariantes constituyen una alternativa que, en algunos casos, puede funcionar incluso mejor que los métodos multivariantes [38].

3.3.2. Calibración multivariante

En la mayoría de los casos, es difícil encontrar una sola variable independiente NIR que explique toda la variación de y y no contenga interferencias de otros analitos o propiedades. En estos casos, es necesario el uso de más de una variable para compensar la baja selectividad del NIR y modelar adecuadamente el parámetro de interés. La calibración multivariante permite resolver ciertos problemas de interferencias y no linealidades, además, en muchos casos, posibilita la detección de muestras anómalas.

- **Partial Least Squares regression (PLS)**

El algoritmo de regresión parcial por mínimos cuadrados (PLS, *Partial Least Squares*) fue introducido en el campo de la economía por Herman Wold en 1975 [39], pero su popularidad llegó gracias a las aplicaciones en quimiometría, llevadas a cabo por Svante Wold y Harald Martens [14,40]. Fueron ellos mismos quienes modificaron ligeramente el algoritmo y lo rebautizaron con el nombre de ‘proyección de estructuras latentes’ (PLS, *Projection to Latent Structures*), que responde a las mismas siglas [14,34]. Actualmente, el PLS (independientemente de los términos que conforman sus siglas), es uno de los métodos de calibración multivariante más usado

en quimiometría [11,36,41,42,43], por lo que se encuentra ampliamente referenciado [7,12,13,14,34,40] y tiene aplicaciones en numerosos campos [143440,44,45,46].

El PLS es un método de regresión que, como el PCA (apartado 3.2.1), se basa en la reducción de variables, una característica muy útil en el tratamiento de grandes cantidades de variables, aunque estas sean ruidosas y/o espectralmente correlacionadas [13,14,34]. La diferencia entre ambos algoritmos radica en el hecho que, mientras el PCA solo modela los datos espectrales (\mathbf{X}), el PLS modela las relaciones de estos datos \mathbf{X} con los valores de referencia del parámetro o parámetros a determinar (\mathbf{Y}). En el caso concreto en que se determina un solo parámetro, el algoritmo se conoce como PLS1, mientras que si se pretende determinar más de uno, se denomina PLS2.

En la aplicación del PLS, los datos suelen ser previamente centrados a la media, tal como se representa en la Figura 3.2a, y autoescalados a la unidad [7,13,14]. La descomposición de las matrices \mathbf{X} ($M \times N$) e \mathbf{Y} ($M \times I$) se realiza de forma simultánea, como se describe en las Ecuaciones 3.15 y 3.16, intentando maximizar la covarianza entre ambas [7,13]:

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} \quad \text{Ecuación 3.6}$$

$$\mathbf{Y} = \mathbf{UQ}^t + \mathbf{F} \quad \text{Ecuación 3.7}$$

donde $\mathbf{T}(M \times A)$ y $\mathbf{U}(M \times A)$ son las matrices de *scores* para los bloques \mathbf{X} e \mathbf{Y} respectivamente, $\mathbf{P}(A \times N)$ y $\mathbf{Q}(A \times I)$, sus matrices de *loadings* y $\mathbf{E}(M \times N)$ y $\mathbf{F}(M \times I)$, las matrices de residuales del modelo, para ambos bloques. Por su parte, A corresponde al número de variables reducidas utilizadas, conocidas como variables latentes (LVs, *Latent Variables*) o factores/componentes PLS. De la misma forma que en los PCs del PCA, el PLS concentra la máxima información en unos pocos factores: el primer componente siempre explica la mayor parte de la varianza, mientras que los componentes sucesivos presentan la mayor varianza posible de la restante. A diferencia de los PCs, los primeros factores también concentran la mayor capacidad predictiva. Esto se debe al hecho que el PLS busca para cada componente la dirección en el espacio de las X que explique la máxima varianza tanto en X como en Y , es decir, aquellos componentes en X que mejor predican en Y . Esta capacidad predictiva

se consigue gracias a la relación interna que se establece entre las nuevas variables (los *scores* de los bloques **X** e **Y**):

$$\mathbf{U} = \mathbf{bT} \quad \text{Ecuación 3.8}$$

donde **b** es el coeficiente de regresión para cada uno de los factores del modelo. Esta relación se aprovecha posteriormente para llevar a cabo el paso de regresión [34], que al realizarse sobre las nuevas variables, elimina los problemas de colinealidad que puedan tener los datos originales. Posteriormente, la predicción del valor **Y** de muestras desconocidas se puede realizar utilizando la Ecuación 3.9 [40], donde **G** son los nuevos residuales de **Y**.

$$\mathbf{Y} = \mathbf{bTQ}^t + \mathbf{G} \sim \mathbf{TQ}^t + \mathbf{G} \quad \text{Ecuación 3.9}$$

La correcta predicción de las nuevas muestras depende de la capacidad predictiva del modelo de calibración. Esta capacidad viene, en parte, definida por el número de factores PLS utilizados. El número óptimo de factores es aquel que recoge la máxima información del parámetro de interés, evitando aquella relativa al ruido u otros parámetros, por lo que depende de la naturaleza de las muestras y el parámetro a cuantificar. La elección de los factores se lleva a cabo como se describe en el apartado 3.4.9.

Una de las ventajas del método PLS es que permite resolver algunos de los problemas típicos de los datos espectroscópicos como la colinealidad, que se elimina en la reducción de variables, o el ruido, que al concentrarse en componentes PLS elevados se puede eludir fácilmente [13,34]. Además, la baja selectividad del NIR también se minimiza al incluir la varianza de **Y** en el modelado. Se reduce así el impacto de grandes variaciones en los datos espectroscópicos (**X**), irrelevantes para el modelo [7,13], como puede ser la absorción de interferentes. Otra característica del PLS, es que como bien se indica en la Tabla 3.1, este algoritmo parte de espectros de mezclas. Esto es debido al hecho que se trata de una técnica de calibración

multivariable inversa (respecto a la ley de Lambert-Beer²) [12], lo que hace posible modelar el componente o propiedad deseados sin necesidad de conocer el resto de componentes y/o fuentes de variación. A su vez, esto implica que requiere un conjunto de muestras de calibración suficientemente grande y completo (apartado 3.4.6) como para poder construir un modelo de calibración adecuado a las muestras a analizar [47].

▪ Isys-PLS1

Otro algoritmo utilizado en esta tesis con fines cuantitativos es la variante del PLS de Isys 5.0 (Malvern Instruments, Malvern, UK), un *software* específico para tratamiento de imagen. La información proporcionada por Malvern sobre este algoritmo es mínima, pero se entiende que se trata de una variación del PLS-Class y la regresión clásica de mínimos cuadrados (CLS, *Classical Least Squares*). Según Amigo et al. [33], el PLS-Class es una aplicación específica del PLS discriminante (PLS-DA, *PLS-Discriminant Analysis*), aplicada mayoritariamente a imagen hiperespectral, cuyos resultados son equivalentes a los obtenidos mediante CLS. El objetivo de ambos algoritmos (PLS-Class y CLS) es el de calcular por método directo la concentración de nuevas muestras a partir de los espectros puros de los componentes [48,49]. Para ello es indispensable disponer del espectro puro de todos los componentes de las muestras, ya que se asume que el espectro de cada muestra es la combinación lineal de los espectros de los distintos componentes puros, ponderados por su concentración. Esto se consigue mediante la Ecuación 3.10:

$$\mathbf{Y} = \mathbf{XS}^t(\mathbf{SS}^t)^{-1} \quad \text{Ecuación 3.10}$$

² La ley de Lambert-Beer implica que la absorbancia = concentración \times absorptividad. A diferencia de lo que ocurre en esta ecuación, en el PLS la información espectral se considera como variable independiente y la información relativa a la concentración (u otro parámetro de interés) es la variable dependiente a predecir.

donde $\mathbf{X}(M \times N)$ es la matriz de espectros problema, $\mathbf{S}(I \times N)$ es la matriz de espectros puros e $\mathbf{Y}(M \times I)$ es la matriz de concentraciones obtenida. A diferencia de estos métodos, el Isys-PLS1 no requiere el espectro puro de todos los componentes de la formulación, sino que su construcción únicamente precisa la adquisición del espectro/imagen química de los componentes puros a cuantificar [37]. Los espectros puros se utilizan para la creación de una biblioteca sobre la cual se construyen los modelos de calibración (uno para cada componente), sin necesidad de conjunto de calibración. Los modelos se definen con la elección del número de factores apropiado y el *threshold* a aplicar. Un aumento en el número de factores permite incrementar la varianza explicada en Y por el modelo, pero, a la vez, también incrementa el riesgo de sobreajuste, por lo que es necesario encontrar un punto de compromiso óptimo. El *threshold* permite ajustar los límites de discriminación entre clases, por lo que su elección depende del número de factores escogido previamente y de la similitud entre el espectro puro del componente a cuantificar y los espectros puros de los demás componentes de la muestra. Al aplicar el modelo obtenido a una matriz, este procede ponderando los datos espectrales entre 0 y 1 en función de su similitud con los espectros puros del componente a cuantificar almacenados la biblioteca. Esta escala se aplica para cada espectro/píxel de la imagen y corresponde a una concentración del 0 al 100% del componente en cuestión. En el caso de las imágenes hiperespectrales el resultado se presenta en forma de mapa de concentraciones para cada uno de los componentes presentes en la librería.

3.4. Etapas del proceso de modelado

El uso de la quimiometría permite compensar la complejidad de los espectros NIR y analizar la gran cantidad de información que contienen, tanto cualitativa como cuantitativamente. Pero el proceso de modelado de los datos no es obvio, sino que precisa de una serie de etapas básicas que van desde el diseño y preparación de muestras hasta la aplicación del modelo de calibración desarrollado [9]. Por ello, en este apartado se presentan las etapas generales de un proceso de modelado de datos espectroscópicos (Figura 3.3).



Figura 3.3. Esquema del proceso de modelado de datos espectroscópicos.

3.4.1. Diseño, selección y preparación de muestras

Se parte de la premisa que para desarrollar un modelo de calibración de calidad es necesario disponer de un conjunto de muestras que sea representativo de todas las fuentes de variabilidad que puedan afectar al espectro de las muestras actuales y/o de futuras muestras de predicción [50]. Como ya se ha comentado anteriormente, esta variabilidad puede ser de origen químico (composición, humedad...) o físico (tamaño y morfología de las partículas, densidad, compactación..., normalmente asociadas al proceso de manufactura) y puede afectar tanto las variables independientes (espectro) como las dependientes (parámetro de interés) [13].

En el desarrollo de los modelos cuantitativos, es de especial importancia que las muestras de calibración abarquen un rango de magnitud del parámetro de interés suficientemente ancho como para poder conseguir un modelo robusto que prediga con precisión incluso aquellas muestras fuera de la normalidad, sin tener que extrapolar [13]. Además, en la mayoría de los casos, donde el objetivo del modelo es la predicción de muestras industriales, el grupo de calibración también debe contener toda la variabilidad del proceso de producción [51]. El uso de muestras de distintos lotes para incorporar toda la variabilidad muestral puede ser suficiente en algunos casos, pero en otros, conseguir un conjunto de muestras adecuado es mucho más complejo. El motivo es que los procesos de producción industriales están optimizados para la obtención de un producto final de concentraciones muy concretas, lo que implica que la variabilidad del producto final sea muy limitada, cubriendo un rango de concentraciones demasiado estrecho para poder construir un modelo de calibración con garantías [9]. La extensión del rango de concentraciones se puede conseguir fácilmente preparando muestras en el laboratorio, pero estas muestras difícilmente contendrán la variabilidad del proceso de producción industrial. En la bibliografía se recogen distintas estrategias que permiten la preparación de muestras que cumplan ambos requisitos [9,51,52]:

- **Muestras preparadas en planta piloto**, simulando los procesos de producción industriales reales [53,54,55] o siguiendo los mismos pasos en el laboratorio [56,57]. Este método es laborioso y requiere de equipamiento específico, pero

permite incorporar toda la variabilidad del proceso y el rango de concentraciones deseado.

- **Muestras dopadas**, que se consiguen añadiendo a las muestras de producción ciertas cantidades de alguno de los componentes de la mezcla para conseguir el rango deseado de concentraciones [58,59]. Las cantidades añadidas deben ser pequeñas para que las características físicas de la muestra no varíen significativamente.
- **Muestras sintéticas de laboratorio**, preparadas por pesada, a las cuales suelen añadirse cierto número de muestras de producción para incorporar la variabilidad física del proceso de producción [58,59,60,61]. Aunque no se ha usado en esta tesis, una alternativa para la adición de esta variabilidad es la adición del **espectro de proceso** (diferencia entre el espectro NIR de las muestras de producción y muestras con concentraciones nominales preparadas en el laboratorio) a los espectros de las muestras sintéticas de laboratorio [51,52,62].

Los resultados de las distintas estrategias son bastante similares [51,52,58,59] y su elección depende del equipamiento disponible en cada laboratorio, así como de las características físicas de las muestras de producción [9]. En función de la estrategia elegida, se realiza el diseño de las muestras a preparar, que deberán cubrir homogéneamente todo el rango de concentraciones establecido del parámetro de interés. Además, el diseño de las muestras debe minimizar la colinealidad, un problema muy común en los datos espectroscópicos, que puede afectar a la calidad de los modelos [63,64]. Sobre todo cuando la correlación espectral entre distintos componentes de la muestra es elevada, es importante minimizar la correlación entre concentraciones para mejorar la robustez del modelo de calibración [51].

3.4.2. Métodos de referencia

Las técnicas relacionadas con el NIR son técnicas relativas, ya que las señales analíticas resultantes no son debidas a un solo analito/propiedad de la muestra. Es por este

motivo que la información del parámetro de interés no se obtiene de forma directa, sino que es necesario establecer relaciones empíricas entre los espectros NIR del conjunto de muestras de calibración/validación y sus valores de referencia para el parámetro de interés. De aquí la necesidad de disponer de métodos de referencia para poder desarrollar modelos NIR cuantitativos y, aunque existen algunos ejemplos de métodos cuantitativos desarrollados sin método de referencia [51,56], no es la estrategia más habitual.

La elección del método de referencia se realiza en función del parámetro a determinar y su precisión y exactitud afectan directamente la calidad de los modelos desarrollados [7]. Dada la diversidad de métodos de referencia usados en el desarrollo de esta memoria, su descripción particular se encuentra en la parte experimental de cada trabajo.

3.4.3. Obtención de la señal analítica

Consiste en la adquisición de la información analítica de las muestras, en este caso, sus espectros o imágenes NIR. Previamente al registro, es necesaria la adquisición del blanco, una medida de la respuesta del equipo u otros interferentes (como el material de la cubeta) en ausencia de la muestra. En la mayoría de los casos, el blanco instrumental (**B**, *Background reference*) se lleva a cabo registrando una placa cerámica del 99% de reflectancia, ya sea manual o automáticamente. En el caso de los equipos NIR-CI, por planos se mide también una referencia oscura (**D**, *Dark reference*), normalmente, un espejo. Una vez adquirida la muestra (**S**, *Sample*), los datos obtenidos se usan para corregir los espectros en modo de reflectancia (**R**), aplicando la Ecuación 3.11 [65]. Este cálculo se aplica a todos los píxeles de cada una de las imágenes registradas.

$$\mathbf{R} = (\mathbf{S} - \mathbf{D}) / (\mathbf{B} - \mathbf{D}) \quad \text{Ecuación 3.11}$$

Los posibles modos de registro por NIR ya han sido descritos anteriormente (apartado 2.2.1), así como su transformación a unidades de absorbancia, necesaria en todos los casos, sea cual sea el modo de registro utilizado. La única restricción en este

sentido es que todas las muestras, tanto las usadas para la construcción del modelo, como las muestras a predecir posteriormente, sean registradas utilizando el mismo modo de registro y, a ser posible, también con el mismo equipo, para que su variabilidad sea incluida en el modelo.

3.4.4. *Preprocesamiento y extracción de datos de imágenes químicas*

La mayor ventaja del NIR-CI respecto el NIRS es la cantidad de información obtenida de la muestra. Sin embargo el tratamiento de la gran cantidad de datos que representa una imagen hiperespectral y su estructura tridimensional en forma de cubo (apartado 2.3) suponen un obstáculo que debe resolverse de forma eficaz para no limitar las aplicaciones de la técnica.

▪ **Preprocesamiento espacial**

El preprocesamiento espacial de las imágenes hiperespectrales permite mejorar la calidad de las imágenes, así como reducir el número de datos a analizar, simplificando el subsiguiente análisis de la imagen [66]. A pesar de que existen un gran número de estrategias descritas en la bibliografía [10,67], en este apartado solamente se describen brevemente los métodos utilizados en este trabajo.

SELECCIÓN DE LAS ZONAS DE INTERÉS

Uno de los pretratamientos más utilizado es la segmentación, que consiste en separar unidades estructurales de las imágenes, como puede ser la figura de interés y el fondo, *i. e.* la parte descartable, sin interés analítico [67]. Una forma de conseguirlo es mediante *thresholding*, un proceso que consiste en definir un valor umbral de alguna propiedad de los píxeles de la imagen (normalmente absorbancia a una longitud de onda o media de absorbancias) que permita discernir la imagen de interés del fondo. El resultado es una imagen binaria de ceros (fondo) y unos (imagen de interés), que se

utiliza como máscara de la imagen original. Normalmente los ceros suelen convertirse en NaN (*Not a Number*), con el objetivo que los píxeles irrelevantes del fondo puedan eluir los análisis y tratamientos posteriores, reduciendo el número de datos a analizar y por ende, el tiempo de análisis.

ELIMINACIÓN DE LOS PÍXELES DEFECTUOSOS

En algunas ocasiones, la imagen presenta píxeles vacíos en que suelen ser debidos a problemas con la adquisición de datos: imperfecciones de los detectores, iluminación no homogénea, ángulo incorrecto [67]. Estos píxeles son fácilmente eliminables mediante el uso de filtros espaciales, donde el píxel defectuoso se substituye por la media de los píxeles vecinos. Esta misma estrategia con matices también se utiliza en la disminución del ruido. En este caso, el tratamiento se aplica a todos los píxeles de la imagen y el valor del píxel central también se incluye en el cálculo de la media [10]. A pesar de la disminución del ruido, este tratamiento reduce la resolución espacial de la imagen, pudiéndose perder detalles de estructuras pequeñas y imposibilitando su discriminación [67].

▪ **Desdoblamiento de la imagen**

Dado que la mayoría de algoritmos y pretratamientos espectrales están desarrollados para matrices de datos bidimensionales, el cubo de datos hiperespectral debe ser desdoblado antes de aplicar cualquier pretratamiento espectral o de cualquier algoritmo para la extracción de la información. El desdoblamiento, en inglés *unfolding*, consiste en la reorganización de los datos tridimensionales obtenidos en el registro, $\mathbf{X}(M \times N \times \lambda)$, en una matriz bidimensional, $\mathbf{X}(MN \times \lambda)$, donde cada fila se corresponde con el espectro de un píxel de la imagen (Figura 3.4).

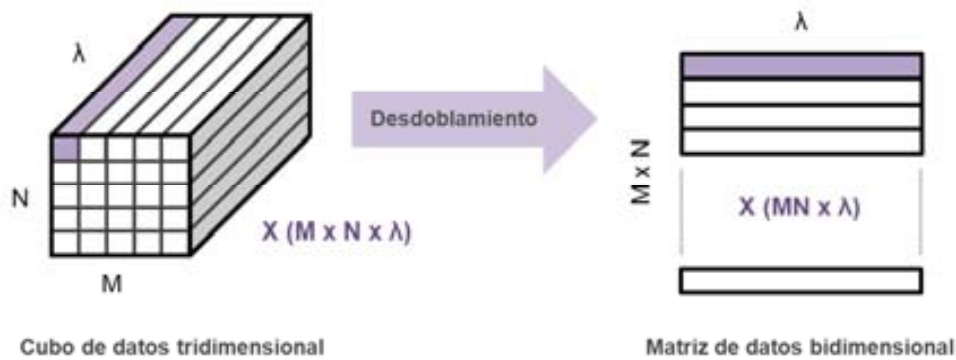


Figura 3.4. Esquema del proceso de desdoblamiento de un cubo de datos hiperespectral.

A la matriz de datos bidimensional se le aplican los pretratamientos y algoritmos necesarios para aprovechar el contraste de la imagen y convertirlo en información de interés. Los tratamientos utilizados en esta tesis se definen en los apartados que siguen y se utiliza la notación $X(M \times N)$ como sinónimo de $X(MN \times \lambda)$. Reorganizando de nuevo los datos en forma cubo tridimensional (*refolding*), se consigue reproducir esta información en forma de imagen o mapa químico, mucho más visual e interpretable.

3.4.5. Pretratamientos espectrales

En ciertos casos, los espectros NIR presentan contribuciones no relacionadas con el analito de interés. Por este motivo, previamente a la construcción de modelos de calibración, es importante la aplicación de pretratamientos a los espectros y/o la selección del rango espectral de trabajo más adecuado. Los pretratamientos son correcciones matemáticas que se aplican con el objetivo de mejorar la relación señal/ruido y eliminar información no relevante de los espectros, facilitando el posterior desarrollo de métodos de calibración más simples y robustos [1,12].

Los principales fenómenos que afectan la calidad de los espectros NIR son el ruido instrumental y las derivas de la línea base [68]. El ruido instrumental se debe a la relación señal/ruido del espectrofotómetro y suele tratarse de una variación aleatoria, eliminable mediante pretratamientos como el promediado de espectros o las

técnicas de suavizado (*smoothing*), entre otros. Por lo que respecta a las derivas de línea base, estas son debidas a la dispersión de la luz, más conocida como *scattering*. El *scattering* es consecuencia de ciertas características físicas de la muestra (morfología, distribución y tamaño de partícula, densidad, grado de compactación, rugosidad de la superficie) o ligeras variaciones en las condiciones de registro (temperatura, ángulo, interferencia de luces externas, iluminación no uniforme de la muestra en imagen). A pesar de que este tipo de información puede ser muy útil en algunas aplicaciones, en la mayoría de casos, las contribuciones espectrales de origen no químico pueden enmascarar información de interés, afectando negativamente al objetivo final del análisis [69]. La contribución del *scattering* puede ser minimizada en la mayoría de los casos con una adecuada presentación de la muestra, ajustando las características de adquisición de los espectros y/o aplicando los pretratamientos matemáticos adecuados [12]. Algoritmos como la variable normal estándar (SNV, *Standard Normal Variate*), la corrección del efecto multiplicativo de dispersión (MSC, *Multiplicative Scatter Correction*) y su variante extendida (EMSC, *Extended MSC*) y las derivadas, son algunos de los pretratamientos más comunes en estos casos. Además, en el caso del NIR-CI, el uso de pretratamientos también permite minimizar ciertas distorsiones espaciales como pueden ser los desplazamientos de la muestra en el registro por barrido, que pueden afectar al enfoque de la imagen [68].

Como se puede ver, existe una gran variedad de pretratamientos (Figura 3.5), todos ellos descritos y comparados en la bibliografía [7,12,70,71,72,73,74], con características diferenciales y efectos distintos sobre los espectros. La selección del algoritmo (o combinación de algoritmos) más apropiado en cada caso es una etapa crucial en el modelado de los datos: la aplicación del pretratamiento adecuado facilita el modelado posterior, mientras que en el caso contrario, el mismo pretratamiento puede introducir artefactos en los datos o incluso eliminar parte de la información de interés [68,75]. Aunque la visualización de los espectros, el conocimiento de la muestra y las condiciones de registro pueden servir como guía, la selección del pretratamiento es un proceso mayoritariamente experimental.

A continuación se describen brevemente las características de los pretratamientos utilizados en el desarrollo de este trabajo.



Figura 3.5. Representación jerárquica de algunos de los pretratamientos espectrales utilizados en NIR [12,43].

▪ Promediado de espectros

Es un pretratamiento simple basado en el promediado de espectros (o por filas) de una misma muestra, de forma que la señal del analito no se ve afectada, pero el ruido aleatorio se reduce en un factor igual a la raíz cuadrada del número de espectros promediados [7]. Este pretratamiento se aplica rutinariamente en la mayoría de equipos comerciales, ya que permite suavizar el ruido instrumental.

Aunque el promediado por filas es el más habitual, es importante mencionar la existencia del promediado por columnas, en el que se promedian absorbancias a distintas longitudes de onda de un mismo espectro. En este caso también se consigue suavizar el ruido instrumental pero, a la vez, se pierde cierta resolución espectral.

▪ Centrado a la media

El centrado a la media (MC, *Mean Centering*) [12] es una técnica que consiste en sustraer a cada espectro individual el espectro medio del conjunto de datos. Para ello,

se resta la media de cada variable (\bar{x}_j) al valor de absorbancia correspondiente de cada fila ($x_{i,j}$), tal como se muestra en la Ecuación 3.12.

$$x_{i,j}^{MC} = x_{i,j} - \bar{x}_j \quad \text{Ecuación 3.12}$$

Este pretratamiento permite eliminar la información de intensidad absoluta de cada variable, asegurando que los datos son interpretados en función de su variación respecto de la media. El centrado es aplicado por defecto por algunos *softwares* en algoritmos tan comunes como el PCA o el PLS (descritos más adelante) y su uso sólo se desaconseja en casos donde el valor absoluto tiene significado para el análisis.

▪ Autoescalado

El autoescalado (AS, *AutoScale*), también conocido como estandarización, es una técnica que modifica las influencias relativas de las distintas variables, ponderando su peso en el modelo. Consiste en aplicar primero un centrado de los espectros, sustrayendo la media de cada variable (\bar{x}_j) al valor de absorbancia correspondiente de cada fila ($x_{i,j}$). Posteriormente, se divide el resultado por la desviación estándar de la variable (σ_j), como se indica en la Ecuación 3.13, de forma que se iguala la varianza de todas las variables [12]. La particularidad de los datos autoescalados es que cada variable tiene una media igual 0 y una desviación estándar de 1.

$$x_{i,j}^{AS} = \frac{x_{i,j} - \bar{x}_j}{\sigma_j} \quad \text{Ecuación 3.13}$$

El autoescalado es de gran utilidad en aquellos casos en que las unidades no son las mismas para todas las variables de trabajo o en que las variables que presentan mayor varianza no son las más relevantes para el parámetro a modelar. En estos casos, este pretratamiento permite igualar el nivel de contribución de las distintas variables al modelo, independientemente de su rango de magnitud, evitando el enmascaramiento de las variables con valores/varianzas inferiores.

▪ Variable Normal Estándar

El pretratamiento de la variable normal estándar (SNV, *Standard Normal Variate*) propuesto por Barne et al. [76] es uno de los pretratamientos más utilizados para la eliminación del *scattering*, dada su simplicidad y efectividad en la corrección de variaciones espectrales debidas a efectos multiplicativos [12]. Estos efectos pueden ser fácilmente confundidos con los efectos multiplicativos debidos a cambios de concentración, por lo que su minimización es de especial interés en análisis cuantitativos.

Matemáticamente, el SNV equivale a un autoescalado por filas, ya que opera individualmente sobre cada espectro, centrándolo a una intensidad media de valor cero y desviación estándar igual a uno. Para conseguirlo, cada valor de absorbancia ($x_{i,j}$), se transforma de acuerdo con la Ecuación 3.14, donde (\bar{x}_i) es la absorbancia media para cada espectro y σ_i es su desviación estándar.

$$x_{i,j}^{SNV} = \frac{x_{i,j} - \bar{x}_i}{\sigma_i} \quad \text{Ecuación 3.14}$$

▪ Suavizado

Los métodos de suavizado se utilizan para reducir matemáticamente el ruido aleatorio e incrementar la relación señal/ruido. Todos ellos se basan en el uso de una ventana móvil, de un número fijo de puntos ($n+1$), que debe ser impar y ≥ 3 [7]. El método más simple es el **suavizado por promedio de columns**, en el que el valor original del punto central de la ventana se substituye por el valor promedio de todos los puntos de la ventana. El mismo proceso se repite sucesivamente, avanzando la ventana punto a punto hasta el final del espectro. Este tipo de suavizado también se conoce como **filtro de ventana móvil rectangular**, dado que todos los puntos de la ventana tienen el mismo peso en el cálculo del promedio [7]. Otras aproximaciones proceden ponderando los pesos de los distintos puntos de la ventana, lo que se conoce como **filtro de ventana móvil ponderada**. El caso más común es el **suavizado de Savitzky-**

Golay [77] que utiliza cálculos polinómicos de convolución para asignar los pesos a cada punto de la ventana.

Una consecuencia inevitable del uso de la ventana móvil es la pérdida $n/2$ puntos al inicio y al final del espectro. Además, es importante una correcta optimización del tamaño de ventana: si el tamaño es demasiado pequeño no suavizará el ruido correctamente, mientras que si es demasiado grande, puede conllevar la pérdida de información relevante.

▪ Derivadas

Las derivadas son uno de los pretratamientos más comunes en espectroscopia NIR, debido a su capacidad de solventar los problemas característicos de esta técnica: los desplazamientos de línea base y el solapamiento de bandas. Matemáticamente, la derivada corresponde a la pendiente de una curva. Dado que la pendiente no se ve afectada por efectos aditivos de la línea base, la 1ª derivada es muy eficiente en su eliminación. Esto también implica que los máximos del espectro original (pendiente nula) se convierten en cero en la 1ª derivada. Esto cambia al aplicar la 2ª derivada, donde los máximos del espectro original se corresponden con los mínimos. La 2ª derivada es una medida del cambio de pendiente de los espectros originales por lo que es muy efectiva en la supresión de los desplazamientos de línea base que varían linealmente con la longitud de onda.

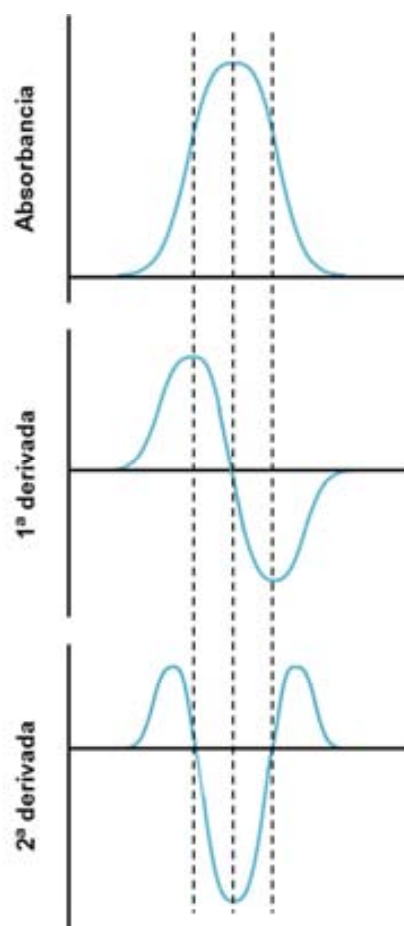


Figura 3.6. Efecto de la aplicación de la primera y la segunda derivada sobre un pico gaussiano.

Así, una derivada de orden n , presentará $n+1$ máximos (positivos y negativos) por cada pico original (Figura 3.6) lo que implica la reducción del ancho de banda y la mejora de la resolución entre picos superpuestos [7]. Esta tendencia puede mejorar la resolución de picos de interés, pero a su vez presenta ciertas desventajas, ya que a medida que crece el orden de la derivada se incrementa la complejidad de la señal y se amplifica el ruido. Este hecho ha limitado el uso de derivadas de órdenes elevadas, siendo la 1ª y 2ª, las más utilizadas. Otra alternativa es la combinación de las derivadas con algoritmos de suavizado [69,71], o el uso de métodos de derivación que llevan implícito el suavizado espectral, como el caso de Savitzky-Golay [77]. Otro algoritmo existente para el cálculo de derivadas es el de Norris [78], también conocido como *gap derivative* o *gap-segment* [79]. En ambos casos, la aplicación de los algoritmos a los espectros se lleva a cabo por el método de ventana móvil descrito más arriba, con la consiguiente pérdida de datos en los extremos de los espectros [72] y la necesidad de optimizar el tamaño de ventana según la mejora de la resolución requerida y la amplificación del ruido espectral que conlleva.

▪ Corrección Ortogonal de la Señal

La corrección ortogonal de la señal (OSC, *Orthogonal Signal Correction*), es un pretratamiento especialmente desarrollado para datos espectroscópicos NIR [80]. Se basa en el hecho que la mayor parte de la varianza comprendida por un conjunto de datos NIR no tiene valor analítico [70,80], por lo que su eliminación facilita su posterior modelado. La mayoría de pretratamientos existentes tienen por objetivo eliminar este tipo de contribuciones no deseadas para mejorar la señal, pero en ningún caso aseguran que sólo se elimina aquella información no relacionada con la propiedad a determinar [81,82]. El OSC, en cambio, elimina del conjunto de datos espectrales (\mathbf{X}) solamente aquella varianza ortogonal con el parámetro de interés (\mathbf{Y}). Existen diferentes algoritmos para la aplicación del OSC [72,80,81,82,83], con resultados similares en todos los casos [82], pero el objetivo común de todos ellos es el de facilitar la obtención de modelos de calibración más simples, es decir, con menos factores.

Una característica del OSC es que precisa de un conjunto de calibración con datos de referencia del parámetro de interés, la precisión de los cuales es crucial para su buen funcionamiento. Esto implica que cada modelo OSC es aplicable únicamente para modelar aquel parámetro para el que ha sido optimizado y para las características particulares de aquel conjunto de muestras seleccionado. Además, la construcción de modelos OSC también requiere optimización del número de factores: cuantos más factores se elijan, mayor será la reducción de la varianza ortogonal, pero también mayor la posibilidad de sobreajuste del modelo a los datos del conjunto de calibración [80]. Una vez construido el modelo OSC, los nuevos espectros se corrigen por proyección, dando lugar a espectros de apariencia poco semejante a los originales y de intensidad menor, pero más fáciles de modelar.

▪ **Combinación de pretratamientos**

La combinación de pretratamientos es una estrategia muy usual en la mejora de la señal espectroscópica NIR. Un ejemplo es, la ya mencionada, aplicación secuencial del suavizado y las derivadas que limita la amplificación del ruido. Otro caso muy común es la combinación de las derivadas y el SNV. El aumento de la resolución conseguido con las derivadas, junto con la corrección del efecto multiplicativo del *scattering* propia del SNV, hacen de esta opción una alternativa muy útil y con muy buenos resultados. En este caso es importante mencionar que el orden de aplicación de los pretratamientos afecta al resultado [75,84]. Según [84] de la misma forma que el resultado de la desviación estándar no es el mismo antes y después de aplicar la derivada, el efecto del SNV también cambia. La mayor eliminación del efecto de *scattering* se consigue aplicando primero derivadas y después SNV, por lo que este orden será preferible al combinar ambos pretratamientos.

▪ **Truncación de los espectros**

Aunque no se puede considerar como un pretratamiento, la truncación de los espectros es una alternativa simple y efectiva para disminuir ruido y eliminar información espectral no relacionada con el analito/propiedad de interés [75,85], lo

que favorece la obtención de modelos más simples. La selección de variables puede llevarse a cabo por visualización de los espectros, pero debido a las características del NIR, es aconsejable disponer de conocimiento previo de las zonas de interés en el espectro, para poder eliminar las zonas espectrales menos informativas o afectadas por la absorción de interferentes. Otra alternativa es la aplicación del algoritmo *jack-knife* [86] que recalcula el modelo añadiendo y eliminando distintos rangos espectrales para proporcionar una indicación de las longitudes de onda con mayor influencia en el modelo. Aunque no se describen en esta tesis, existen un gran número de algoritmos desarrollados con esta misma finalidad y descritos en la bibliografía [85,87].

3.4.6. Selección del grupo de calibración/validación

De todas las muestras disponibles para el desarrollo del modelo, una parte se utilizarán para la construcción y el resto, para su validación. Para asegurar que los resultados del modelo a desarrollar sean representativos, es importante que las muestras del grupo de calibración cumplan estrictamente los requisitos mencionados en el apartado 3.2.1, es decir, que incluyan toda la variabilidad de las muestras a predecir [13]. Esto incluye el rango del parámetro a predecir, pero también otras fuentes de variabilidad química (composición, humedad, formas polimórficas) o física (densidad, tamaños de partícula) que pueden afectar al espectro. Además hay que evitar la inclusión de *outliers*, que podrían afectar fuertemente la tendencia del modelo.

Dado que los valores de referencia y el diseño experimental son conocidos, es fácil seleccionar muestras que cubran todo el rango de valores de la magnitud a cuantificar e incluyan la variabilidad debida al proceso de producción/preparación. Aún así, es difícil asegurar que el grupo de calibración incluye toda la variabilidad espectral. La visualización de los espectros a simple vista puede ser útil en esta selección y en el descarte de *outliers* muy claros, pero en la mayoría de los casos, las diferencias entre espectros NIR son tan sutiles que son difíciles de detectar a simple vista. Las herramientas utilizadas en esta tesis para incluir toda la variabilidad

espectral en el proceso de selección del grupo de calibración son el análisis en componentes principales (PCA) y el algoritmo de Kennard & Stone.

▪ **Análisis en Componentes Principales (PCA)**

El PCA identifica las fuentes de mayor variabilidad espectral del conjunto de muestras para crear nuevos ejes de coordenadas. El funcionamiento de este algoritmo se ha desarrollado con detalle en el apartado 3.2.1, ya que se considera como un método cualitativo. El resultado es la redistribución de todas las muestras en estos nuevos ejes (gráfico de *scores*), lo que permite visualizar la disposición de todas las muestras en el espacio de trabajo. Su uso como método de selección de muestras se basa en el hecho que un buen conjunto de calibración debe tener muestras que cubran todo el espacio de trabajo, lo que en algunos casos puede llevar a incluir algún *outlier* en la calibración [13]. La aplicación del **algoritmo de Hotelling T²** [14,88] puede servir de ayuda en la discriminación entre muestras extremas y *outliers* espectrales. Se trata de una generalización multivariante del test-T de Student que se basa en el hecho que los *scores* de las muestras se distribuyen de forma normal si todo funciona correctamente. Aplicado sobre un gráfico de *scores*, este estadístico define el área con el porcentaje de confianza elegido (normalmente 95% o 99%), lo que se conoce como elipse de Hotelling T². Las muestras incluidas en la elipse no son *outliers* y deben ser tomadas en cuenta en el desarrollo del modelo de calibración.

▪ **Algoritmo de Kennard & Stone**

El algoritmo propuesto por Kennard & Stone [89,90] permite la selección automática de las muestras de calibración, de forma que cubran uniformemente toda la región experimental. En primer lugar, el algoritmo elige la muestra con valor más cercano a la media del conjunto de muestras, aquella supuestamente más representativa. A partir de aquí, se procede escogiendo secuencialmente aquellas muestras proyectadas a mayor distancia en el espacio de diseño respecto a aquellas previamente seleccionadas. Para calcular las distancias entre las muestras $p = (p_1, \dots, p_j, \dots, p_n)$, y $q = (q_1, \dots, q_j, \dots, q_n)$, se utilizan normalmente los valores al

cuadrado de la distancia Euclídea, definidos por la Ecuación 3.15. Este proceso se repite hasta que se llega al número deseado de muestras de calibración.

$$d_E^2(p, q) = \sum_{j=1}^n (p_j - q_j)^2 \quad \text{Ecuación 3.15}$$

3.4.7. Construcción del modelo de calibración

La información analítica contenida en los espectros NIR se puede extraer mediante métodos calibración, que permiten establecer relaciones empíricas entre la señal analítica (absorbancia) y ciertas propiedades de las muestras. Como ya se ha comentado, la calibración NIR es relativa ya que, en la mayoría de los casos, requiere del conocimiento de valores de referencia del parámetro a predecir [1].

Los métodos de calibración pueden ser cualitativos (apartado 3.2) o cuantitativos (apartado 3.3), según si caracterizan/clasifican muestras o determinan ciertos parámetros, pero su objetivo final es siempre la construcción de modelos capaces de predecir acuradamente características y/o propiedades de muestras desconocidas. El método de calibración de elección dependerá del objetivo del análisis, la naturaleza de la muestra, el conocimiento *a priori* del sistema y los datos disponibles [9]. Una vez, seleccionada la herramienta quimiométrica idónea para modelar los datos, se construye el modelo de calibración utilizando las muestras seleccionadas previamente (apartado 3.4.6).

Los mismos métodos utilizados para NIR convencional son aplicados a datos hiperespectrales NIR-CI previamente desdoblados (apartado 3.2.4). En este caso, tanto los métodos cualitativos como los cuantitativos suelen conocerse en imagen química como métodos de clasificación, ya que permiten identificar regiones con características espectrales similares, dando información sobre la distribución y concentración de propiedades físicas y químicas de la muestra a lo largo de su superficie [65,91]. Es importante mencionar que, aunque no son tan comunes, también existen métodos *three-way* o *n-way* útiles para el modelado de imágenes hiperespectrales. Estos métodos no requieren el desdoblado de la imagen, por lo que tienen en cuenta las correlaciones entre píxeles, negligidas en los métodos 2D [66].

3.4.8. Evaluación del modelo

Una vez construido el modelo de calibración, se evalúa el ajuste de los datos al mismo. Esto se lleva a cabo mediante el gráfico de valores predichos *vs* valores de referencia. En un modelo correcto, las muestras deben estar distribuidas uniformemente a lo largo de la recta de calibración, de pendiente y coeficiente de correlación (R^2) próximos a 1 y ordenada al origen, próxima a 0.

Posteriormente, el modelo debe ser validado para asegurar la calidad de su capacidad predictiva sobre nuevas muestras. Una primera evaluación del modelo es la **validación interna**, que valora la capacidad predictiva del modelo sobre las mismas muestras de calibración. Este tipo de validación es muy útil cuando se dispone de pocas muestras, ya que permite prescindir del grupo de validación, pero también se emplea para simular los efectos de cambios en la composición del grupo de calibración/validación. La validación interna se lleva a cabo por validación cruzada (CV, *Cross Validation*) [92], que divide el conjunto de muestras de calibración en varios bloques o segmentos. El modelo se construye tantas veces como número de segmentos se ha elegido, utilizando cada vez un segmento distinto como bloque de validación y el resto para construir el modelo [12,14,93]. En el caso concreto en que el número de segmentos es igual al número de muestras de calibrado, es decir, cada segmento tiene una sola muestra, este procedimiento se conoce como *leave-one-out* [12,93,94].

Cuando se dispone de suficientes muestras para crear un conjunto de validación independiente del de calibración (apartado 3.2.6), se lleva a cabo la **validación externa** del modelo, que consiste en aplicar el modelo construido a las muestras de validación [12,64]. Estas muestras presentan la misma naturaleza que las de calibración y su valor de referencia también es conocido, pero no han sido utilizadas para la construcción del modelo.

En ambos casos, la estimación del poder predictivo del modelo se consigue comparando los valores predichos con los reales, graficando uno respecto el otro y/o mediante parámetros como el error cuadrático medio (RMSE, *Root Mean Square Error*) o el error estándar relativo (RSE, *Relative Standard Error*) [9,13]. En el primer

caso se trata de una medida absoluta que mantiene las mismas unidades que el parámetro predicho, mientras que en el segundo, la medida es relativa y se expresa en porcentaje. El RMSE y el RSE vienen definidos por las Ecuaciones 3.6 y 3.7 respectivamente:

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{m}} \quad \text{Ecuación 3.16}$$

$$RSE (\%) = \sqrt{\frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m y_i^2}} \times 100 \quad \text{Ecuación 3.17}$$

donde \hat{y}_i es el valor predicho por el modelo para la muestra i , y_i es el valor de referencia para la misma muestra y m , el número de muestras predichas. En el caso de la validación interna estos parámetros se denominan error cuadrático medio de la validación cruzada/calibración (RMSECV/RMSEC, *Root Mean Square Error of Cross Validation/Calibration*) y error estándar relativo de la validación cruzada/calibración (RSECV/RSEC, *Relative Standard Error of Cross Validation/Calibration*), mientras que en el caso de la validación externa, así como en las sucesivas predicciones de nuevas muestras, se conocen como error cuadrático medio de la predicción (RMSEP, *Root Mean Square Error of Prediction*) y error estándar relativo de la predicción (RSEP, *Relative Standard Error of Prediction*). Otros estadísticos básicos, utilizados en esta tesis con la misma finalidad fueron el promedio de los residuales, el bias (Ecuación 3.18), la suma de los cuadrados de los residuales predichos (PRESS, *Predicted Residual Sum of Squares*) (Ecuación 3.19) y el test-T de los residuales de muestras apareadas.

$$Bias = \sum_{i=1}^m (\hat{y}_i - y_i) \quad \text{Ecuación 3.18}$$

$$PRESS = \sum_{i=1}^m (\hat{y}_i - y_i)^2 \quad \text{Ecuación 3.19}$$

3.4.9. Elección del número de factores

En el caso de los modelos multivariantes con reducción de variables es necesario elegir el número de factores/componentes óptimos. Para ello, los métodos de evaluación de modelos descritos en el apartado anterior se aplican a los resultados obtenidos con distinto número de factores o componentes, con el objetivo de elegir aquel que presente mejor capacidad predictiva. Cuanto mayor sea el número de factores, mayor será el porcentaje de información incluida en el modelo (var-Y) y menores los errores de calibración (RMSEC/RMSECV). Sin embargo, la complejidad del modelo también aumenta con el número de factores y se puede llegar al sobreajuste, lo que implica una disminución de la capacidad predictiva del modelo (RMSEP) por modelado del ruido [93]. Es por ello, que se debe llegar a una solución de compromiso.

La tendencia de acumulación de la varianza explicada en Y (var-Y) [64] a medida que aumenta el número de factores puede servir como guía en esta tarea, pero no es un parámetro definitivo. Uno de los métodos más populares es la elección del número de factores con menor error de predicción. Para ello se grafican los resultados del cálculo de RMSEP o RSEP respecto al número de factores [64]. A medida que aumenta el número de factores, el valor de ambos parámetros disminuye hasta llegar a un mínimo o una disminución relativa significativa, fenómenos indicativos del número óptimo de componentes [13]. En algunos casos es difícil establecer el punto donde se encuentra esta inflexión, por lo que es útil la aplicación de algoritmos para determinar en qué punto la disminución del error deja de ser significativa. El test-F propuesto por Haaland y Thomas [94] permite comparar el modelo con menor valor de PRESS con un modelo con menor número de factores. Si la diferencia entre los valores de PRESS de los modelos comparados es suficientemente significativa se mantendrá el modelo inicial (con menor PRESS), por el contrario, una diferencia no significativa indica que se puede prescindir de este/os factor/es y escoger el modelo alternativo con menos factores.

Si a pesar de la optimización del número de factores, los resultados de la validación no son satisfactorios, el modelo debe ser recalculado, replanteando la elección en cada una de las etapas del proceso de modelado.

3.4.10. *Aplicación del modelo e interpretación de los resultados*

Cuando los resultados de validación de un modelo son satisfactorios, este puede ser aplicado para la determinación del parámetro de interés en nuevas muestras. En el caso del NIR convencional, se dispone de un solo espectro por muestra, por lo que cada predicción corresponde a una muestra. En el caso de las imágenes químicas, en cambio, lo que se obtiene es la predicción para cada uno de los píxeles de la imagen. Lo que interesa en este caso es aprovechar el contraste entre píxeles obtenido mediante las técnicas de clasificación y convertirlo en un mapa de distribución del analito o propiedad de interés, una forma mucho más sencilla y visual de interpretar la información. Para ello, se lleva a cabo el **procesado de la imagen**, mediante el cual, se convierte el rango de predicciones obtenidas en una escala de grises o colores, en función de la cual se mapean los resultados a lo largo de la imagen. Otra opción muy común para la interpretación de resultados de imágenes químicas son los histogramas, donde se representa el número de píxeles frente al valor del parámetro predicho (contraste), lo que proporciona información sobre la homogeneidad de la propiedad medida en la muestra [11,65]. La principal desventaja de este método es que se pierde la información espacial.

REFERENCIAS

- 1 Blanco, M., Villaroya, I. (2002) Spectroscopy: A rapid-response analytical tool. *TrAC-Trend. Anal. Chem.* **21**, 240-250.
- 2 Hopke, P.K. (2003) The evolution of chemometrics. *Anal. Chim. Acta* **500**, 365-377.
- 3 Kiralj, R., Ferreira, M.M.C. (2006) The past, present, and future of chemometrics worldwide: Some etymological, linguistic, and bibliometric investigations. *J. Chemometr.* **20**, 247-272.
- 4 Wold, S. (1995) Chemometrics; what do we mean with it, and what do we want from it? *Chemometr. Intell. Lab.* **30**, 109-115.
- 5 Kowalski, B.R. (1975) Chemometrics: Views and propositions. *J. Chem. Inform. Comput. Sci.* **15**, 201-203.
- 6 Massart, D.L., Vandeginste, B.G.M., Deming, S.M., Michotte, Y., Kaufman, L. (2003) *Chemometrics: A textbook*. 1^a ed., 5^a reimpresión. Elsevier Science B. V., Amsterdam.
- 7 MasPOCH, S. (2012) El procesado de la señal espectroscópica. En: Ríos, A., Moreno, M.C., Simonet, B.M. (coords.) *Técnicas espectroscópicas en química analítica*. Vol. I; Editorial Síntesis, Madrid, pp. 103-152.
- 8 Blanco, M., Romero, M.A. (2001) Near infrared libraries in the pharmaceutical industry: A solution for identity confirmation. *Analyst* **126**, 2212-2217.
- 9 Blanco, M., Alcalà, M. (2010) NIR spectroscopy in pharmaceutical analysis: off-line and at-line PAT applications. En: Bakeev, K.A. (ed.) *Process analytical Technology*. 2^a ed.; Wiley, Chichester, pp. 463-491
- 10 Brereton, R.G. (2007) *Applied Chemometrics for Scientists*. Wiley, Chichester.
- 11 Gendrin, C., Roggo, Y., Collet, C. (2008) Pharmaceutical applications of vibrational chemical imaging and chemometrics: A review. *J. Pharm. Biomed. Anal.* **48**, 533-553.
- 12 Miller, C.E. (2010) Chemometrics in Process Analytical Technology (PAT). En: Bakeev, K.A. (ed.) *Process analytical Technology*. 2^a ed.; Wiley, Chichester, pp. 353-438
- 13 Blanco, M., Alcalà, M. (2009) Multivariate calibration for quantitative analysis. En: Sun, D.W. (ed.), *Infrared spectroscopy for food quality analysis and control*. Academic Press, Elsevier, New York, NY, pp. 51-82.

- 14 Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikström, C., Wold, S. (2006) *Multi- and megavariate data analysis. Part I, Basic principles and applications*. 2^a ed., Umetrics, Umea.
- 15 Felizardo, P., Menezes, J.C., Neiva-Correia, M.J. (2011) Process analytical technology use in biofuels manufacturing. En: Undey, C., Low, D., Menezes, J.C., Koch, M. (eds.) *PAT applied in biopharmaceutical process development and manufacturing: An Enabling Tool for Quality-by-Design*, CRC Press, Boca Raton, FL, pp. 201-221.
- 16 De Beer, T.R.M., Alleso, M., Goethals, F., Coppens, A., Vander Heyden, Y., Lopez de Diego, H., Rantanen, J., Verpoort, F., Vervaet, C., Remon, J.P., Baeyens, W.R.G. (2007) Implementation of a process analytical technology system in a freeze-drying process using Raman spectroscopy for in-line process monitoring. *Anal. Chem.* **79**, 7992-8003.
- 17 De Beer, T.R.M., Vercruyse, P., Burggraeve, A., Quinten, T., Ouyang, J., Zhang, X., Vervaet, C., Remon, J.P., Baeyens, W.R.G. (2009) In-line and real-time process monitoring of a freeze-drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools. *J. Pharm. Sci.* **98**, 3430-3444.
- 18 Kemsley, E.K. (1996) Discriminant analysis of high-dimensional data: A comparison of principal components analysis and partial least squares data reduction methods. *Chemometr. Intell. Lab.* **33**, 47-61.
- 19 Asuero, A.G., Sayago, A., González, A.G. (2006) The correlation coefficient: An overview. *Crit. Rev. Anal. Chem.* **36**, 41-59.
- 20 Yoon, W.L., Jee, R.D., Moffat, A.C., Blackler, P.D., Yeung, K., Lee, D.C. (1999) Construction and transferability of a spectral library for the identification of common solvents by near-infrared transfectance spectroscopy. *Analyst* **124**, 1197-1203.
- 21 Blanco, M., Eustaquio, A., González, J.M., Serrano, D. (2000) Identification and quantitation assays for intact tablets of two related pharmaceutical preparations by reflectance near-infrared spectroscopy: Validation of the procedure. *J. Pharm. Biomed. Anal.* **22**, 139-148.
- 22 Jovanovic, N., Gerich, A., Bouchard, A., Jiskoot, W. (2006) Near-infrared imaging for studying homogeneity of protein-sugar mixtures. *Pharm. Res.* **23**, 2002-2013.

- 23 Cairós, C., Amigo, J.M., Watt, R., Coello, J., MasPOCH, S. (2009) Implementation of enhanced correlation maps in near infrared chemical images: Application in pharmaceutical research. *Talanta* **79**, 657-664.
- 24 Rosas, J.G., de Waard, H., de Beer, T., Vervaet, C., Remon, J.P., Hinrichs, W.L.J., Frijlink, H. W., Blanco, M. (2014) NIR spectroscopy for the in-line monitoring of a multicomponent formulation during the entire freeze-drying process. *J. Pharm. Biomed. Anal.* **97**, 39-46.
- 25 Blanco, M., Cruz, J., Bautista, M. (2008) Development of a univariate calibration model for pharmaceutical analysis based on NIR spectra. *Anal. Bioanal. Chem.* **392**, 1367-1372.
- 26 Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., Jent, N. (2007) A review of near-infrared spectroscopy and chemometrics in pharmaceutical technologies. *J. Pharm. Biomed. Anal.* **22**, 683-700.
- 27 Bertran, E., Blanco, M., MasPOCH, S., Ortiz, M.C., Sánchez, M.S., Sarabia, L.A. (1999) Handling intrinsic non-linearity in near-infrared reflectance spectroscopy. *Chemometr. Intell. Lab.* **49**, 215-224.
- 28 Long, J.R., Gregoriou, V.G., Gemperline, P.J. (1990) Spectroscopic calibration and quantitation using artificial neural networks. *Anal. Chem.* **62**, 1791-1797.
- 29 Cleveland, W.S., Devlin, S.J. (1988) Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**, 596-610.
- 30 Thissen, U., Pepers, M., Üstün, B., Melssen, W.J., Buydens, L.M.C. (2004) Comparing support vector machines to PLS for spectral regression applications. *Chemometr. Intell. Lab.* **73**, 169-179.
- 31 Baffi, G., Martin, E.B., Morris, A.J. (1999) Non-linear projection to latent structures revisited: The quadratic PLS algorithm. *Comput. Chem. Eng.* **23**, 395-411.
- 32 Mark, H., Workman, J. (2010) Classical least squares, Part I: Mathematical theory. *Spectroscopy-US* **25**, 16-21.
- 33 Amigo, J.M., Ravn, C., Gallagher, N.B., Bro, R. (2009) A comparison of a common approach to partial least squares-discriminant analysis and classical least squares in hyperspectral imaging. *Int. J. Pharmaceut.* **373**, 179-182.

- 34 Wold, S., Sjöström, M., Eriksson, L. (2001) PLS-regression: A basic tool of chemometrics. *Chemometr. Intell. Lab.* **58**, 109–130.
- 35 Tauler, R., Kowalski, B. (1993) Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Anal. Chem.* **65**, 2040-2047.
- 36 Tauler, R., de Juan, A. (2006) Multivariate Curve Resolution (MCR) from 2000: Progress in concepts and applications, *Crit. Rev. Anal. Chem.* **36**, 163-176.
- 37 Puchert T., Lochmann D., Menezes J.C., Reich G. (2010) Near-infrared chemical imaging (NIR-CI) for counterfeit drug identification - a four-stage concept with a novel approach of data processing (Linear Image Signature). *J. Pharm. Biomed. Anal.* **51**, 138-145.
- 38 Patel, A.D., Luner, P.E., Kemper, M.S. (2000) Quantitative analysis of polymorphs in binary and multi-component powder mixtures by near-infrared reflectance spectroscopy. *Int. J. Pharmaceut.* **206**, 63-74.
- 39 Wold, H. (1982) Soft Modeling. The Basic Design and Some Extensions. En: Jöreskog, K.G., Wold, H. (eds.), *Systems Under Indirect Observation*. North Holland, Amsterdam. Referencia encontrada en [34].
- 40 Abdi, H. (2007) Partial Least Squares Regression. PLS-Regression, en: Salkind, N. (ed.) *Encyclopedia of Measurement and Statistics*. SAGE Publications, Thousand Oaks, CA.
- 41 Ballabio, D., Todeschini, R. (2009) Multivariate Classification for Qualitative Analysis, En: Sun, D.W. (ed.), *Infrared spectroscopy for food quality analysis and control*. Academic Press, Elsevier, New York, NY, pp. 83-104.
- 42 Rodionova, O.Y., Pomerantsev, A.L. (2006) Chemometrics: Achievements and prospects. *Russ. Chem. Rev.* **75**, 271-287.
- 43 Rosas, J.G. (2013) *La Tecnología Analítica de Procesos (PAT) en la industria farmacéutica. Utilización de la espectroscopia NIR y técnicas de imagen química NIR-CI*. Publicia, Omniscryptum GmbH & Co. KG, Saarbrücken.
- 44 Sun D.W. (2009) *Infrared spectroscopy for food quality analysis and control. Part II: Applications*. Academic Press, Elsevier, New York, NY.
- 45 Einax, J.W., Zwanziger, H.W., Geiß, S. (1997) *Chemometrics in environmental analysis. Part B: Case studies*. Wiley-VCH Verlag GmbH, Weinheim.

- 46 Bakeev, K.A. (2010) *Process analytical Technology*. 2^a ed.; Wiley, Chichester.
- 47 Ravn, C., Skibsted, E., Bro, R. (2008) Near-Infrared Chemical Imaging (NIR-CI) on pharmaceutical solid dosage forms - Comparing common calibration approaches. *J. Pharm. Biomed. Anal.* **48**, 554-561.
- 48 Chevallier, S., Bertrand, D., Kohler, A., Courcoux, P. (2006) Application of PLS-DA in multivariate image analysis. *J. Chemometr.* **20**, 221-229.
- 49 Gendrin, C., Roggo, Y., Collet, C. (2007) Content uniformity of pharmaceutical solid dosage forms by near infrared hyperspectral imaging: A feasibility study. *Talanta* **73**, 733-741.
- 50 Macho, S., Larrechi, M.S. (2002) Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *Trends Anal. Chem.* **21**, 799-806.
- 51 Blanco, M., Pequero, A. (2010) Analysis of pharmaceuticals by NIR spectroscopy without a reference method. *TrAC-Trend. Anal. Chem.* **29**, 1127-1136.
- 52 Cárdenas, V., Blanco, M., Alcalà, M. (2014) Strategies for selecting the calibration set in pharmaceutical near infrared spectroscopy. A comparative study. *J. Pharm. Innov.* DOI 10.1007/s12247-014-9192-3.
- 53 Eustaquio, A., Graham, P., Jee, R.D., Moffatt, A.C., Trafford, A.D. (1998) Quantification of paracetamol in intact tablets using near-infrared transmittance spectroscopy. *Analyst* **123**, 2303-2306.
- 54 Broad, N.W., Jee, R.D., Moffat, A.C., Smith, M.R. (2001) Application of transmission near-infrared spectroscopy to uniformity of content testing of intact steroid tablets. *Analyst* **126**, 2207-2211.
- 55 Xiang, D., Konigsberger, M., Wabuyele, B., Hornung, K., Cheney, J. (2009) Development of robust quantitative methods by near-infrared spectroscopy for rapid pharmaceutical determination of content uniformity in complex tablet matrix. *Analyst* **134**, 1405-1415.
- 56 Blanco, M., Bautista, M., Alcalá, M. (2008) Preparing calibration sets for use in pharmaceutical analysis by NIR spectroscopy. *J. Pharm. Sci-US* **97**, 1236-1245.

- 57 Blanco, M., Alcalà, M., González, J.M., Torras, E. (2006) A process analytical technology approach based on near infrared spectroscopy: Tablet hardness, content uniformity and dissolution test measurements of intact tablets. *J. Pharm. Sci.-US* **95**, 2137-2144.
- 58 Blanco, M., Coello, J., Iturriaga, H., Maspoch, S., de la Pezuela, C. (1997) Strategies for constructing the calibration set in the determination of active principles in pharmaceuticals by near infrared diffuse reflectance spectrometry. *Analyst* **122**, 761-765.
- 59 Blanco, M., Coello, J., Iturriaga, H., Maspoch S., Pou, N. (2001) Influence of the procedure used to prepare the calibration sample set on the performance of near infrared spectroscopy in quantitative pharmaceutical analyses. *Analyst* **126**, 1129-1134.
- 60 Blanco, M., Coello, J., Eustaquio, A., Iturriaga, H., Maspoch, S. (1999) Analytical control of pharmaceutical production steps by near infrared reflectance spectroscopy. *Anal. Chim. Acta* **392**, 237-246.
- 61 Dubois, P., Martinez, J.R., Levillain, P. (1987) Determination of five components in a pharmaceutical formulation using near infrared reflectance spectrophotometry. *Analyst* **112**, 1675-1679.
- 62 Blanco, M., Cueva-Mestranza, R., Peguero, A. (2011) NIR analysis of pharmaceutical samples without reference data: Improving the calibration. *Talanta* **85**, 2218-2225.
- 63 Kleinbaum, D., Kupper, L., Nizam, A., Muller, K. (2008) Applied Regression Analysis and Other Multivariable Methods, 4^a ed., Thomson Higher Education, Belmont, CA.
- 64 Esbensen, K.H. (2002) *Multivariate Data Analysis - in Practice. An Introduction to Multivariate Data Analysis and Experimental Design*. 5^a ed, Camo Process AS, Oslo.
- 65 Lewis, E.N., Schoppelrei, J.W., Makein, L., Kidder L.H., Lee, E. (2010) Near-infrared chemical imaging for product and process understanding. En: Bakeev K.A. (ed.) *Process analytical Technology*. 2^a ed.; Wiley, Chichester, pp. 245-279.
- 66 Roggo, Y., Ulmschneider, M. (2008) Chemical Imaging and Chemometrics: Useful tools for process analytical technology. En: Cox, S. (ed.) *Pharmaceutical Manufacturing Handbook: Regulations and quality*. Wiley, Hoboken, NJ, pp. 411-431.
- 67 Russ, J.C. (1999) *Image processing handbook*. 3^a ed. CRC Press, Boca Raton, FL.

- 68 Amigo, J.M. (2010) Practical issues of hyperspectral imaging analysis of solid dosage forms. *Anal. Bioanal. Chem.* **398**(1), 93-109.
- 69 Reich, G. (2005) Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv. Drug Deliver. Rev.* **57**, 1109-1143.
- 70 Boysworth, M.K., Booksh, K.S. (2008) Aspects of multivariate calibration applied to near-infrared spectroscopy. En: Burns, D.A., Ciurczak, E.W. *Handbook of Near-Infrared Analysis*. 3^a ed. CRC Press, Boca Raton, FL, pp. 207-229.
- 71 Candolfi, A., de Maesschalck, R., Jouan-Rimbaud, D., Hailey, P.A., Massart, D.L. (1999) The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. *J. Pharm. Biomed. Anal.* **21**, 115-132.
- 72 Rinnan, Å., Nørgaard, L., van den Berg, F., Thygesen, J., Bro, R., Engelsen, S.B. (2009) Data Pre-processing. En: Sun, D.W. (ed.), *Infrared spectroscopy for food quality analysis and control*. Academic Press, Elsevier, New York, NY, pp. 29-50.
- 73 Azzouz, T., Puigdomènech, A., Aragay, M., Tauler, R. (2003) Comparison between different data pre-treatment methods in the analysis of forage samples using near-infrared diffuse reflectance spectroscopy and partial least-squares multivariate calibration method. *Anal. Chim. Acta* **484**, 121-134.
- 74 Pizarro, C., Esteban-Díez, I., Nistal, A.J., González-Saiz, J.M. (2004) Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Anal. Chim. Acta* **509**, 217-227.
- 75 Huang, J., Romero-Torres, S., Moshgbar, M. (2010) Practical considerations in data pre-treatment for NIR and Raman spectroscopy [en línea]. *Am. Pharm. Rev.* **13** [Consultado en mayo de 2014]. Disponible en: <http://www.americanpharmaceuticalreview.com/>
- 76 Barnes, R.J., Dhanoa, M.S., Lister, S.J. (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **43**, 772-777.
- 77 Savitzky, A., Golay, M.J.E. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627-1639.
- 78 Norris, K.H., Williams, P.C. (1984) 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 Chem.* **61**, 158-165.

- 79 Hopkins, D.W. (2001) What is a Norris derivative? *NIR News* **12**, 3-5.
- 80 Wold, S., Antti, H., Lindgren, F., Öhman, J. (1998) Orthogonal signal correction of near-infrared spectra. *Chemometr. Intell. Lab.* **44**, 175-185.
- 81 Fearn, T. (2000) On orthogonal signal correction. *Chemometr. Intell. Lab.*, **50**, 47-52.
- 82 Svensson, O., Kourti, T., MacGregor, J.F. (2002) An investigation of orthogonal signal correction algorithms and their characteristics. *J. Chemometr.* **16**, 176-188.
- 83 Sjöblom, S., Svensson, O., Josefson, M., Kullberg, H., Wold, S. (1998) An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemometr. Intell. Lab.* **44**, 229-244.
- 84 Fearn, T. (2008) The interaction between standard normal variate and derivatives. *NIR news* **19(7)**, 16-17.
- 85 Xiaobo, Z., Jiewen, Z., Povey, M.J.W., Holmes, M., Hanpin, M. (2010) Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* **667**, 14-32.
- 86 Martens, H., Martens, M. (2000) Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual. Prefer.* **11**, 5-16.
- 87 Balabin, R.M., Smirnov, S.V. (2011) Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. *Anal. Chim. Acta* **692**, 63-72.
- 88 Wikström, C., Albano, C., Eriksson, L., Fridén, H., Johansson, E., Nordahl, Å., Rännar, S., Sandberg, M., Kettaneh-Wold, N., Wold, S. (1998) Multivariate process and quality monitoring applied to an electrolysis process: Part I. Process supervision with multivariate control charts. *Chemometr. Intell. Lab.* **42**, 221-231.
- 89 Kennard, R.W., Stone, L.A. (1969) Computer aided design of experiments. *Technometrics*; **11**, 137-148.
- 90 Daszykowski, M., Walczak, B., Massart, D.L. (2002) Representative subset selection. *Anal. Chim. Acta* **468**, 91-103.
- 91 Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Bell, S.E.J. (2008) Recent applications of chemical imaging to pharmaceutical process monitoring and quality control. *Eur. J. Pharm. Biopharm.* **69**, 10-22.

- 92 Wold, S. (1978) Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 20, 397-405.
- 93 Coello, J., Maspoch, S. (2007) Calibración Multivariable. En: Blanco, M., Cerdà, V. (eds.) *Temas avanzados de quimiometría*. Universitat de les Illes Balears, Palma, pp 233-280.
- 94 Haaland, D.M., Thomas, E.V. (1988) Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* 60, 1193-1202.

INDUSTRIA PETROQUÍMICA

4. MONITORIZACIÓN DE DISTINTAS PROPIEDADES FISICOQUÍMICAS DE GASÓLEOS MEDIANTE ESPECTROSCOPIA NIR

4.1. Introducción	175
4.2. Experimental	177
4.2.1. Condiciones experimentales del registro de espectros NIR.....	177
4.2.2. Muestras	177
4.2.3. Datos de referencia.....	179
4.2.4. Tratamiento de datos	179
4.3. Resultados y discusión.....	182
4.3.1. Selección de muestras	182
4.3.2. Construcción de los modelos PLS.....	190
4.3.3. Monitorización de los modelos	193
4.4. Conclusiones	196
REFERENCIAS	197

4. MONITORIZACIÓN DE DISTINTAS PROPIEDADES FÍSICOQUÍMICAS DE GASÓLEOS MEDIANTE ESPECTROSCOPIA NIR

4.1. Introducción

La producción y comercialización de productos industriales va siempre sujeta al cumplimiento de especificaciones de numerosos parámetros. Esto permite establecer y asegurar unos niveles de calidad, así como obtener la confianza de los consumidores. Como se ha mencionado en el capítulo 1, la calidad de los gasóleos y las mezclas de estos, viene regulada por la normativa europea EN 590 [1] y la de la *American Society for testing and Materials*, ASTM D 975 [2]. Estas normativas definen los parámetros a evaluar, los valores límites y los métodos de referencia a aplicar para cada uno de ellos. En general, son muchos los parámetros a controlar (19 según el BOE [3]) y los métodos analíticos designados para ello suelen ser largos y tediosos, por lo que no son adecuados para la obtención rápida de información y el establecimiento de la situación del producto/proceso en tiempo real.

Dadas las ventajas de la espectroscopia NIR (capítulo 2), esta técnica permite llevar a cabo el control, tanto del proceso como de los parámetros de calidad establecidos en las normativas, en muy poco tiempo y con resultados similares o incluso mejores que los métodos de referencia por lo que respecta a precisión [4,5,6] y repetibilidad [6,7]. La idoneidad de esta técnica en el campo petroquímico se demuestra con el gran número de aplicaciones existentes, así como por la existencia de dos normas ASTM dirigidas al uso de la espectroscopia MIR y NIR para el análisis cualitativo, ASTM 1790 (2010) [8] y cuantitativo, ASTM 1655 (2012) [9].

En este estudio se propone la utilización de la espectroscopia NIR para el control *on-line* e *in-situ* de la calidad de gasóleos resultantes del proceso de mezclado. Estos gasóleos son mezclas muy complejas de destilados procedentes de diferentes crudos, de origen y composición distintos, y diferentes corrientes de fabricación, además de cierta cantidad de biodiésel. Todo ello incrementa de forma muy

importante la variabilidad de las mezclas resultantes y por ende, la variabilidad de sus espectros NIR, complicando notablemente la construcción de modelos de calibración adecuados. Este problema suele corregirse con la construcción de modelos independientes para cada tipo de gasóleo, pero esta estrategia causa dificultades para su aplicación en rutina por exigir la selección del origen de los crudos antes de aplicar el modelo [6,10]. Además, es muy difícil controlar todas las fuentes de variabilidad de las muestras en el proceso y disponer de rangos amplios para cada parámetro de interés, lo que requiere el registro de muestras durante largos períodos de tiempo antes del desarrollo del modelo [10].

En este capítulo se describe el proceso de desarrollo de modelos PLS para la monitorización de siete propiedades fisicoquímicas claves para asegurar la calidad final de los gasóleos. Se desarrolla un único modelo para cada parámetro independientemente del origen de la muestra. La utilización de distintos PCA permite recoger todas las fuentes de variación de las muestras en el grupo de calibración, lo que constituye la base para poder desarrollar modelos sencillos y suficientemente robustos, que soporten mejor la gran variabilidad de las muestras de gasóleo y mantengan una buena capacidad predictiva a largo plazo.

4.2. Experimental

4.2.1. Condiciones experimentales del registro de espectros NIR

Para el desarrollo de este trabajo, se utilizó un equipo FT-NIR Matrix-F de Bartec Benke GmbH (Reinbek/Hamburg, Alemania). Los espectros NIR fueron registrados *on-line* en modo de transmitancia, mediante una sonda integrada en un conducto de recirculación del flujo de gasóleo proveniente del proceso de mezclado. Por este motivo, la temperatura de registro fue aproximadamente la misma que en el proceso de mezclado, unos 37 °C.

Para la obtención de cada espectro se realizaron 10 *scans*, cubriendo el rango entre 1000 y 2200 nm, con una resolución de 2 nm. Para el registro de los espectros se utilizaron los *softwares* OPUS-NT y Calibration Workshop (CWS) v2.10.001 de SensoLogic GmbH (Norderstedt, Alemania).

4.2.2. Muestras

Todas las muestras analizadas en este trabajo son muestras de gasóleos *post-blending* de la planta de San Roque de la Compañía Española de Petróleos, S.A.U. (CEPSA), producidas a partir de noviembre del 2011. Se dispuso de un total de 278 muestras para la construcción, validación y evaluación de los modelos (muestras hasta abril de 2012). Las muestras posteriores se utilizaron para el control en planta del funcionamiento de dichos modelos.

La composición de las muestras incluía pequeñas cantidades añadidas de queroseno, gasóleos atmosféricos y atmosféricos pesados provenientes de distintos procesos de refinado como destilación directa, FCC, *visbreaking*, coquización e hidrotratamiento. Los espectros de todas las muestras de calibración y validación se muestran en la Figura 4.1a. La fuente de variabilidad más importante es la composición de los gasóleos que se debe esencialmente a la procedencia de los crudos de origen y, a simple vista, es indetectable en los espectros. Usualmente, cada gasóleo es el resultado de la mezcla de destilados de entre 2 y 5 crudos de origen variable, así como de distintas corrientes provenientes de distintos procesos de

refino. Durante la construcción y validación de los modelos, se trabajó con gasóleos procedentes de mezclas de 12 crudos (C1-C12) de distintos orígenes geográficos más otro, consistente en una mezcla de los anteriores. Otros 11 crudos adicionales fueron incorporados en planta posteriormente, durante el período de validación de los modelos.

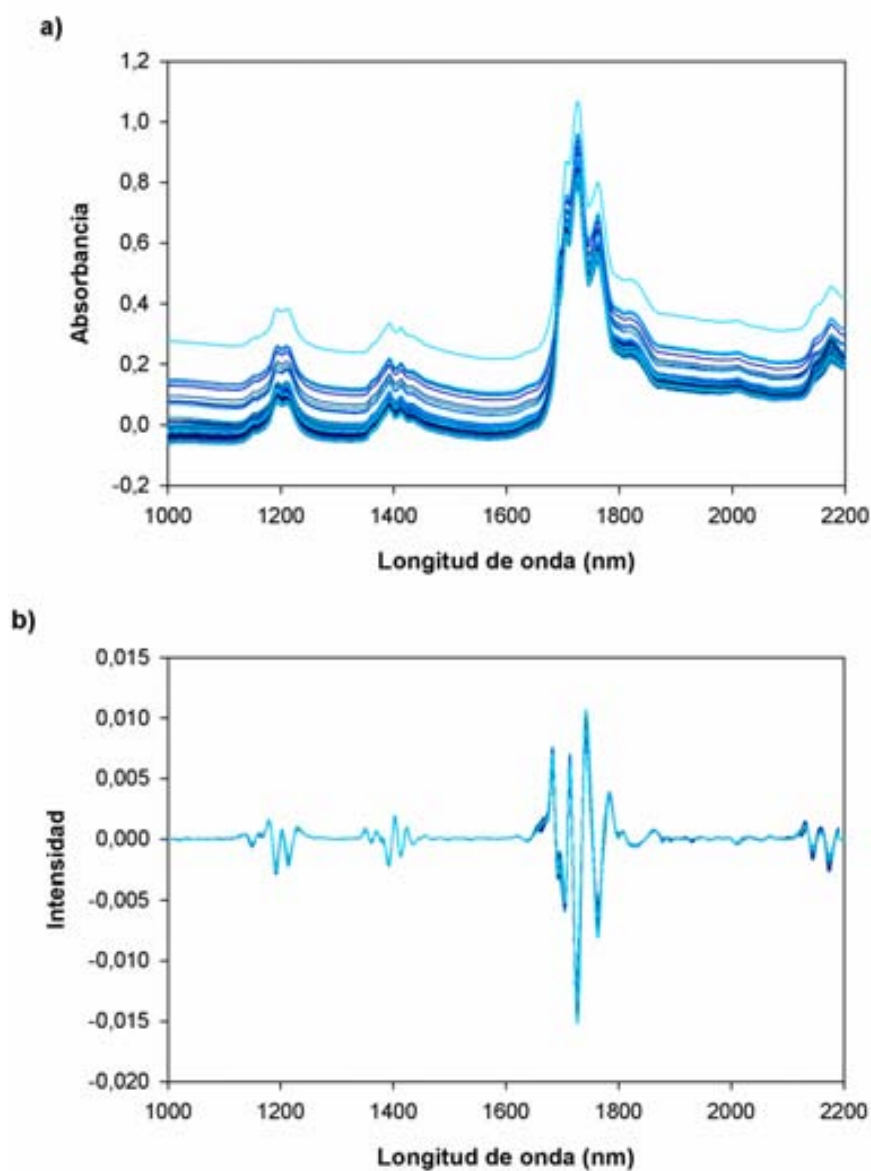


Figura 4.1. Espectros de todas las muestras de calibración y validación: a) en absorbancia y b) en 2ª derivada.

4.2.3. Datos de referencia

Las muestras se recogieron directamente del oleoducto tres veces al día y fueron referenciadas en el laboratorio. Cada referencia se asocia con el espectro registrado a la misma hora. Para la mayoría de las muestras mencionadas en el apartado anterior se dispone de valores de referencia de los parámetros físico-químicos a modelar: densidad a 15°C (Dens, g/cm³), índice de cetano (CI, *Cetane Index*, adimensional), contenido en azufre (SC, *Sulphur*, mg/kg) y ésteres metílicos de ácido graso (FAME, *Fatty Acid Methyl Ester*, % v/v), temperatura de destilación del 95% (D 95%, °C), punto de nube (CP, *Cloud Point*, °C) e inflamabilidad (FP, *Flash Point*, °C). Estos valores de referencia fueron obtenidos mediante los métodos establecidos por las normas ASTM y EN-ISO. En la Tabla 4.1 se recogen los valores de repetitividad y reproducibilidad de cada uno de ellos, calculados según establece la norma ASTM D86 [11].

Tabla 4.1. Precisión de los métodos de referencia.

Parámetro	Método	Repetitividad	Reproducibilidad
Dens (g/cm ³)	EN ISO 12185	0.0002	0.0005
CI	EN ISO 5165	1	4
FAME (% v/v)	EN 14078	0.3	1.4
CP (°C)	ASTM D 23015	2	4
D 95% (°C)	ASTM D 86	2.1–2.9	4.3–6.3
FP (°C)	EN ISO 2719	1.62–2.03	3.98–4.97
SC (mg/kg)	EN ISO 20846	0.66–1.16	1.34–2.35

4.2.4. Tratamiento de datos

▪ Pretratamiento espectral

Previamente a la construcción de los modelos de calibración, se procedió al pretratamiento de los espectros con el objetivo de eliminar las contribuciones

espectrales no deseadas y facilitar el posterior desarrollo de métodos de calibración más simples y robustos. Se ensayaron diversos pretratamientos, consiguiendo los mejores resultados con la aplicación de la 2ª derivada de Savitzky-Golay, con suavizado de 11 puntos de ventana móvil y polinomio de 2º grado. Con ello se logró aumentar la resolución de las bandas anchas y solapadas, propias de la espectroscopia NIR sin aumentar significativamente el ruido. Para los modelos de algunos parámetros ha sido necesaria la aplicación posterior de otro pretratamiento, el SNV, que permite corregir intensidades y desviaciones de la línea de base debidas al efecto de dispersión de la luz. En el caso de las muestras de gasóleo, esta dispersión de la luz es posiblemente causada por la presencia de microgotas de líquidos inmiscibles, burbujas o fluctuaciones de densidad en la muestra. Todos los pretratamientos espectrales se aplicaron directamente sobre los espectros en absorbancia mediante el programa The Unscrambler v9.7 de CAMO (Trondheim, Norway) y CWS.

▪ Construcción de los modelos de calibración

La determinación cuantitativa de los distintos parámetros de interés se consiguió mediante modelos de regresión PLS. Como ya se ha comentado en el capítulo 3, la construcción de los modelos PLS requiere un conjunto de muestras de calibración que incluya todas las posibles fuentes de variación espectral [10] y sea representativo, tanto de las muestras conocidas, como de las futuras muestras a predecir. La construcción de los modelos se llevó a cabo por CV y la elección del número óptimo de factores se realizó teniendo en cuenta el promedio residual de las predicciones, RMSEP, y la aplicación del criterio estadístico test-t de Student de los residuales. En caso de duda, se aplicó el test-F propuesto por Haaland y Thomas [12] con este mismo objetivo.

El *software* utilizado para los análisis en planta (CWS) está diseñado para la aplicación de modelos y la obtención de resultados, por lo que no es apropiado para la construcción de modelos de calibración. Por este motivo, los modelos PLS fueron contruidos, optimizados y validados originalmente con The Unscrambler v.9.7,

más adecuado y versátil para esta tarea y recalculados posteriormente con el *software* CWS. Este recálculo fue necesario dado que los pretratamientos y algoritmos aplicados no son exactamente iguales en ambos programas, lo que implica ciertos cambios respecto al modelo PLS original.

▪ Métodos de selección de muestras

Para la selección del grupo de calibración, se han combinado distintas estrategias con el fin de asegurar la selección de aquellas muestras más representativas para el modelo, limitando la información redundante y los niveles de ruido.

Con el objetivo de maximizar la variabilidad espectral, se aplicó el algoritmo de Kennard y Stone [13] que, de entre todas las muestras, elige secuencialmente aquellas menos semejantes espectralmente. La aplicación de este método se ha llevado a cabo mediante MATLAB v 7.0 de The MathWorks (Natick, MA). Esta elección se complementó considerando los valores de referencia y asegurando que las muestras del set de calibración se repartieran homogéneamente en todo el rango de trabajo para cada parámetro. Finalmente, se analizaron los datos mediante PCA. Dado que las nuevas variables, más conocidas como PCs, se ordenan según la varianza que explican, el estudio de las causas de variabilidad de un grupo de muestras es mucho más fácil e intuitivo. En este trabajo, se ha aplicado el PCA utilizando tres tipos de datos distintos: espectros (con distintos pretratamientos), datos de referencia de los parámetros a cuantificar y composición de crudos de cada muestra. Los modelos PCA se construyeron mediante los *softwares* The Unscrambler v.9.7 y The Unscrambler X v.10.2 de CAMO (Trondheim, Norway). La información conseguida con los PCA, se utilizó también para la detección de *outliers* (en calibración y predicción) y es de gran utilidad para el control de la variabilidad de futuras muestras de predicción, un punto crucial para el entendimiento del proceso [14].

4.3. Resultados y discusión

4.3.1. Selección de muestras

La selección de las muestras de calibración es un punto crucial en el desarrollo de modelos PLS robustos y con buena capacidad predictiva, especialmente cuando se trabaja con una elevada variabilidad en las muestras. En este caso, el origen y las características específicas de cada crudo son importantes fuentes de variabilidad que afectan las características de los diésels resultantes y causan pequeñas variaciones en los espectros NIR. La incorporación de todas estas variaciones es determinante para garantizar la calidad y robustez de los modelos finales, por lo que hay que asegurar la representación del máximo número de crudos en los diferentes conjuntos de calibración. Para ello, se han combinado los criterios habituales de selección de muestras (información espectral y valores de referencia) con la información de composición de crudos de cada muestra para crear un grupo de calibración específico para cada parámetro, es decir, para cada modelo.

▪ Información espectral

La información espectral de las muestras se ha analizado mediante PCA. El gráfico de *scores* de este análisis permite determinar el grado de semejanza espectral de las muestras entre ellas y situarlas dentro del espacio de trabajo. Esto permite detectar potenciales *outliers* espectrales usando como baremo la elipse de Hotelling T^2 y ver qué efecto tiene el pretratamiento sobre ellos. En la Figura 4.2 se muestra el gráfico de *scores* para PC1 y PC2 del PCA de los espectros en absorbancia (Figura 4.2a) y en 2ª derivada (Figura 4.2b). Se observa que aquellas muestras que caen fuera de la elipse de Hotelling en la Figura 4.2a (posibles *outliers* espectrales), quedan claramente corregidas en la Figura 4.2b, lo que indica que la aplicación de la 2ª derivada ha eliminado las diferencias existentes y las muestras no deben ser descartadas.

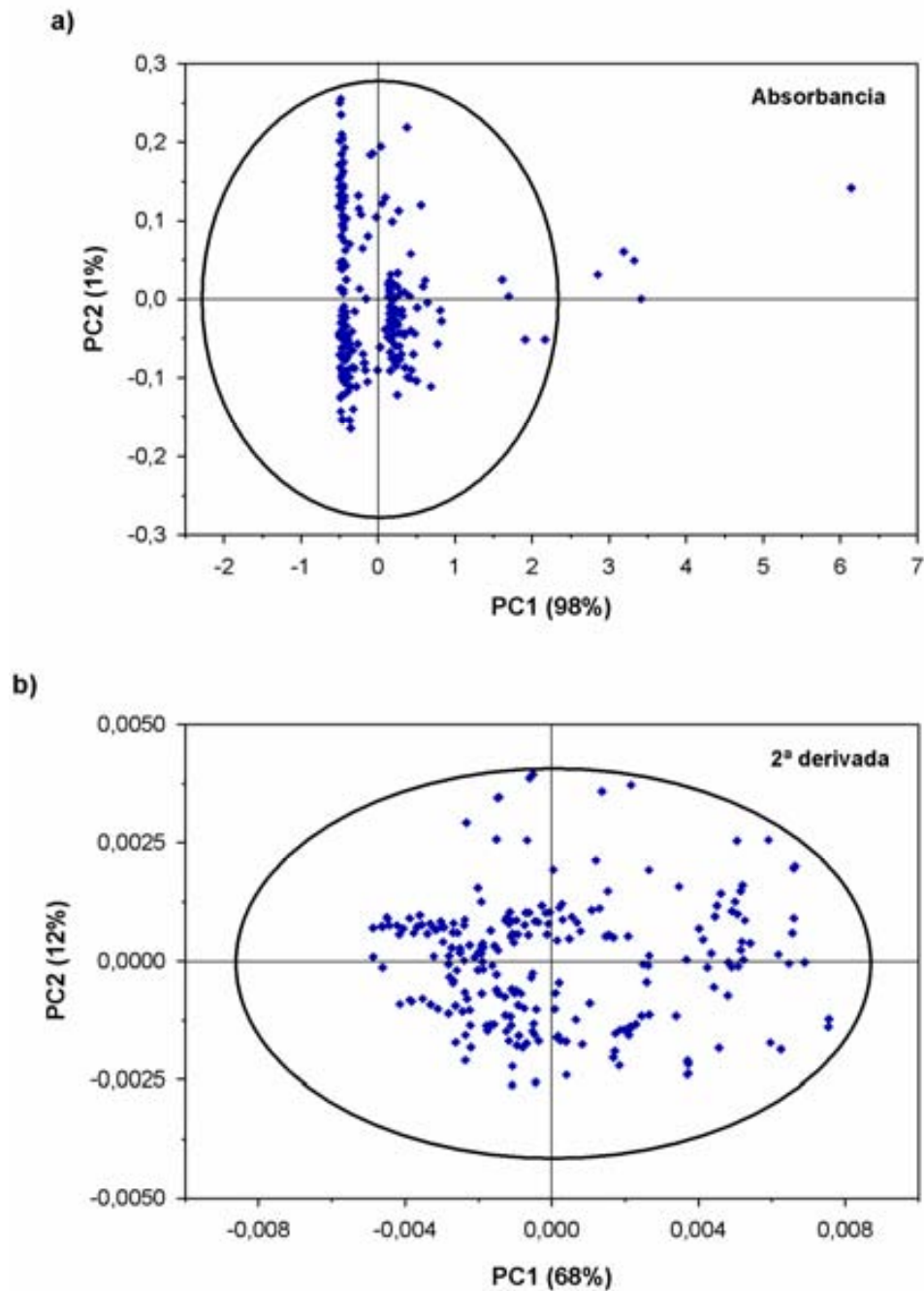


Figura 4.2. Scores de los PCA de los espectros de todas las muestras de calibración/validación: a) en absorbancia y b) en 2ª derivada. El círculo corresponde a la elipse de Hotelling T^2 calculada para el 99% de la varianza en ambos casos.

Cuando se grafican los scores de los PCA espectrales respecto a los valores de referencia de cada parámetro a determinar, se puede detectar la existencia de posibles *outliers* por referencia incorrecta, así como anticipar si el parámetro analizado será o no fácil de modelar y qué pretratamiento será más favorable para ello. A modo de

ejemplo, en la Figura 4.3a se representan los valores de los *scores* del PC1, el PC que explica mayor varianza, del PCA de la Figura 4.2a (espectros en absorbancia) coloreados según los valores de referencia para CI. No se observa ninguna relación aparente: los *scores* toman valores independientemente de los valores de referencia del CI. La Figura 4.3b es el gráfico homólogo a la anterior, pero con los *scores* del primer PC del PCA de la Figura 4.2b (espectros en 2ª derivada). En este caso sí se observa una clara tendencia: a medida que aumenta el valor de los *scores*, disminuye el valor de referencia para CI. Este hecho es indicativo de que este parámetro será modelable con este pretratamiento. Un comportamiento similar se observa para los parámetros densidad y FAME. Para el resto de parámetros, la relación es más difícil de detectar a simple vista, lo que indica que probablemente se trata de parámetros menos relacionados con el espectro y, por consiguiente, más difíciles de modelar.

Una vez analizados, los espectros se utilizan para una primera selección del grupo de calibración mediante el algoritmo de Kennard y Stones: se escogen 60 espectros para el grupo de calibración, dejando el resto para validación.

▪ Valores de referencia

Se ha realizado un análisis PCA de los valores de referencia de los distintos parámetros del conjunto de muestras disponibles para construir los modelos. Para conseguir una contribución equilibrada de cada uno de los parámetros independientemente de su magnitud, los valores se han autoescalado a varianza unidad. El gráfico de dispersión de *scores* de este PCA, nos proporciona una idea del comportamiento del conjunto de valores de todos los parámetros de una muestra con respecto a las demás, mientras que el gráfico de *loadings*, muestra las relaciones entre los distintos parámetros. En la Figura 4.4 se observa el gráfico *bi-plot* (*scores* y *loadings* superpuestos), que permite establecer el grado de influencia de cada parámetro sobre las muestras. En este ejemplo, los *scores* evidencian que las muestras se separan claramente en dos grupos, mientras que los *loadings* permiten deducir que esta segregación es debida al parámetro 9 (% FAME), el único claramente relacionado con uno de los dos grupos. El *cluster* superior contiene las muestras con

FAME y el *cluster* inferior, aquellas que no lo contienen. Este resultado fue contrastado con los valores de FAME de referencia y sugiere que el modelo para FAME será fácil de construir y que este tipo de PCA discriminará claramente si una muestra nueva contiene FAME o no.

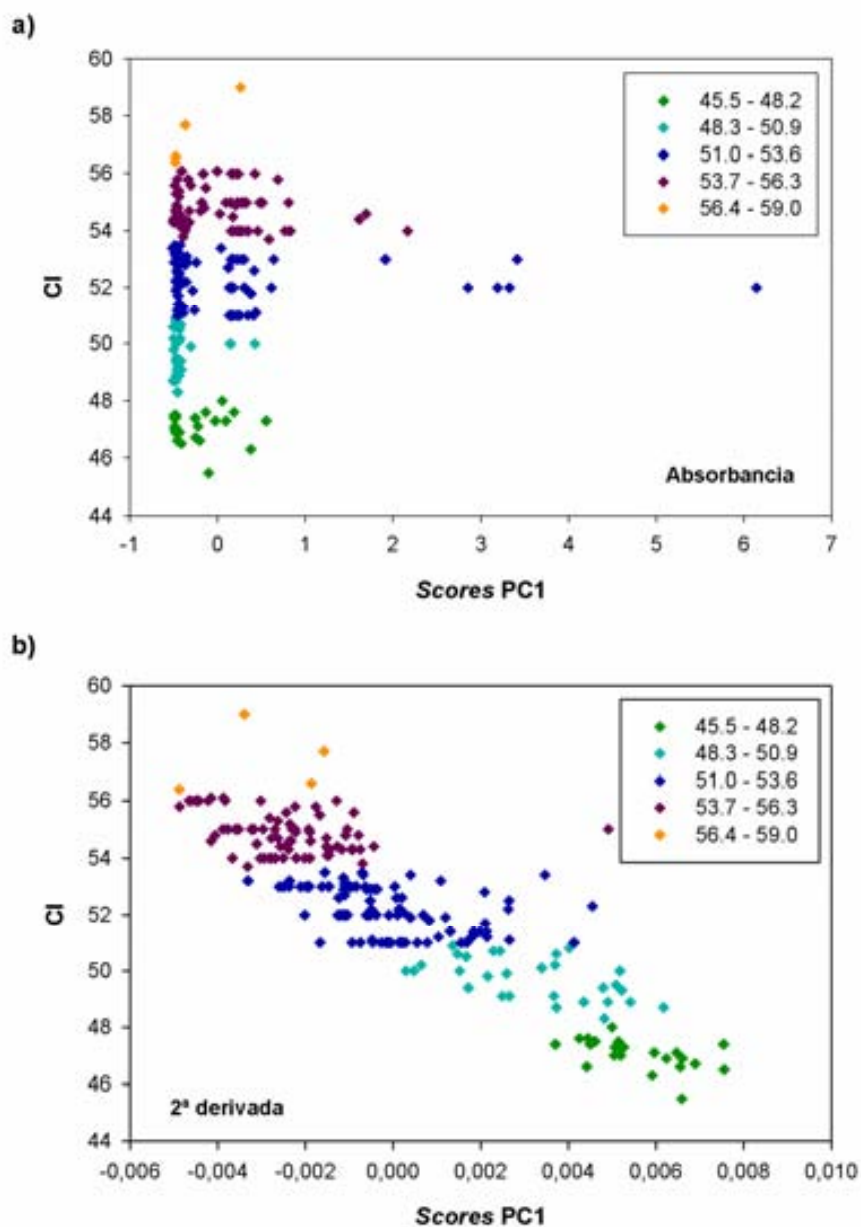


Figura 4.3. Gráficos de scores de los PC1 de la Figura 4.2 frente los valores de referencia de CI para todas las muestras de calibración/validación: a) en absorbancia y b) en 2ª derivada. Las muestras se representan en distintos colores en función de su valor de referencia para CI (ver leyenda).

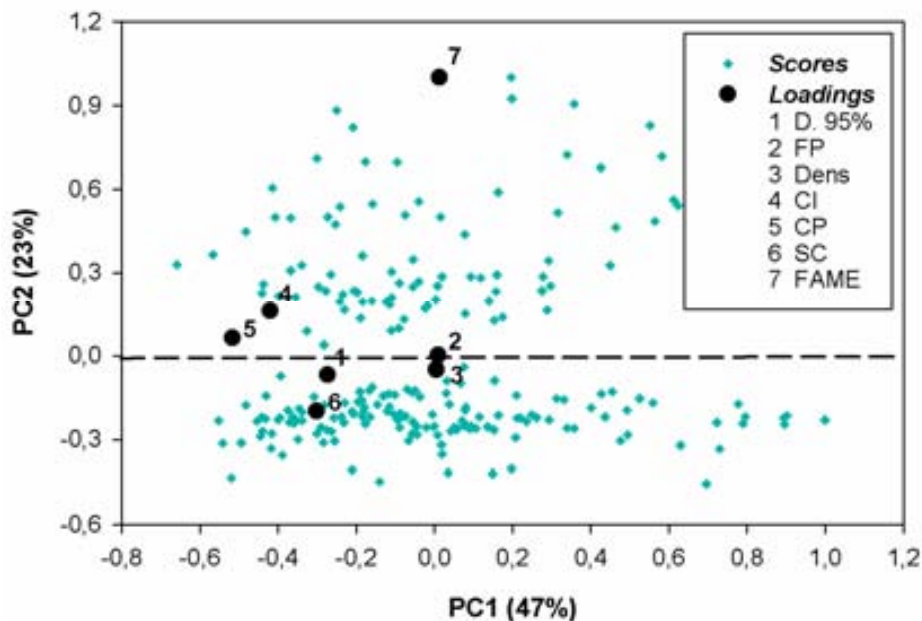


Figura 4.4. Gráfico bi-plot del PCA de valores de referencia para todas las muestras de calibración/validación. Las muestras de la parte superior (por encima de la línea discontinua) contienen FAME, mientras que las de la parte inferior, no.

Los rangos cubiertos por las muestras para cada parámetro se pueden ver en la Tabla 4.2 y los histogramas de su distribución se muestran en la Figura 4.5. Se observa que, en la mayoría de los casos, el rango de valores por parámetro es muy estrecho y que la mayoría de valores se acumulan en un extremo del rango. La falta de datos uniformemente distribuidos a lo largo del rango de trabajo, limita la obtención de modelos de calibración adecuados y reduce su capacidad predictiva. Los ejemplos más claros de este comportamiento son el del SC y el FAME. En el caso del azufre, más del 95% de las muestras se encuentra en sólo tres clases centrales de las 10 que tiene el histograma. En esta situación, los pocos valores extremos de cada parámetro toman una importancia crucial a la hora de construir los modelos PLS, mientras que los valores centrales tendrán que ser cuidadosamente seleccionados para evitar la repetición de la información e incorporación de ruido al modelo. Por lo que respecta al FAME, más del 55% de las muestras se encuentran en el extremo inferior del rango, mientras que no hay ninguna muestra que presente valores centrales. Esta es la razón de la segregación de muestras en el PCA de la

Figura 4.4 e implica que habrá que elegir adecuadamente las muestras de calibración para que se repartan por todo el rango de trabajo.

Todas aquellas muestras con valores de referencia extremos o poco representados se añaden a los grupos de calibración seleccionados inicialmente por Kennard y Stone.

Tabla 4.2. Rangos cubiertos por las muestras de gasóleo de calibración/validación para cada parámetro.

	Dens (g/cm ³)	CI	FAME (% v/v)	CP (°C)	D 95% (°C)	FP (°C)	SC (mg/kg)
Mín.	0.8262	45.5	0	-12.4	336.5	56	2
Máx.	0.8617	59.0	14.9	2.2	384.9	71	22

▪ Composición de crudos

La información de la composición de crudos de cada muestra (si la muestra contiene o no un crudo determinado) también se ha usado para la construcción de modelos PCA. La singularidad de este análisis es el hecho de trabajar con variables cualitativas. Por este motivo, se ha generado un código binario que permite identificar si una muestra contiene un determinado crudo: si es así, se asigna un 1, en caso contrario, se le asigna un 0. Dado que se dispone de 12 crudos, cada muestra contendrá 12 variables con valores 0 o 1 para la construcción del PCA.

La proximidad entre muestras en el gráfico de *scores* de este tipo de PCA, es indicativa de la semejanza en su combinación de crudos: dos muestras superpuestas están compuestas por los mismos crudos. La proximidad entre crudos (*loadings*) implica que estos suelen coincidir en las mezclas para formar gasóleos: dos crudos superpuestos están presentes exactamente en las mismas muestras de gasóleo.

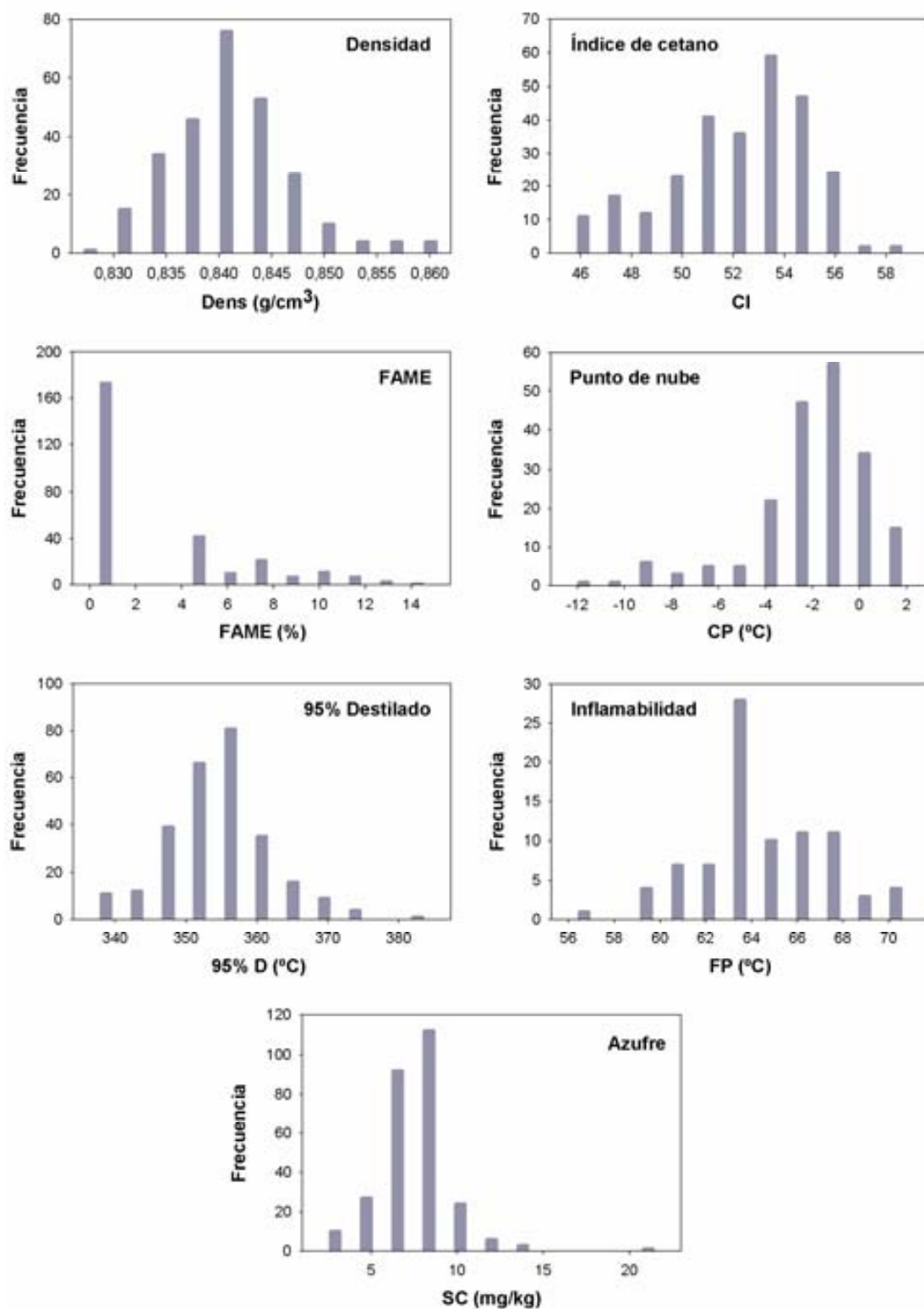


Figura 4.5. Histogramas ilustrando la distribución de los valores de referencia de las muestras de calibración/validación para cada parámetro de interés.

Cuando se superponen *scores* y *loadings* (gráfico *bi-plot*), las muestras quedarán más cercanas de aquellos crudos que contienen. En el gráfico *bi-plot* de la Figura 4.6, construido con el mismo número de muestras que el de la Figura 4.4, las muestras de la misma composición aparecen superpuestas y próximas a los *loadings* que indican su composición. En esta misma figura, se advierte un grupo de muestras claramente separadas del resto, delimitado por una elipse negra. Este grupo está constituido por 42 muestras, agrupadas en 4 composiciones distintas y muy cercanas a los crudos C5 y C12. Se ha comprobado que todas las muestras de este grupo contienen ambos crudos y que estos no están presentes en ninguna otra muestra. Es por este motivo, que se ha prestado especial atención en incluir muestras de este *cluster* en los grupos de calibración de todos los parámetros.

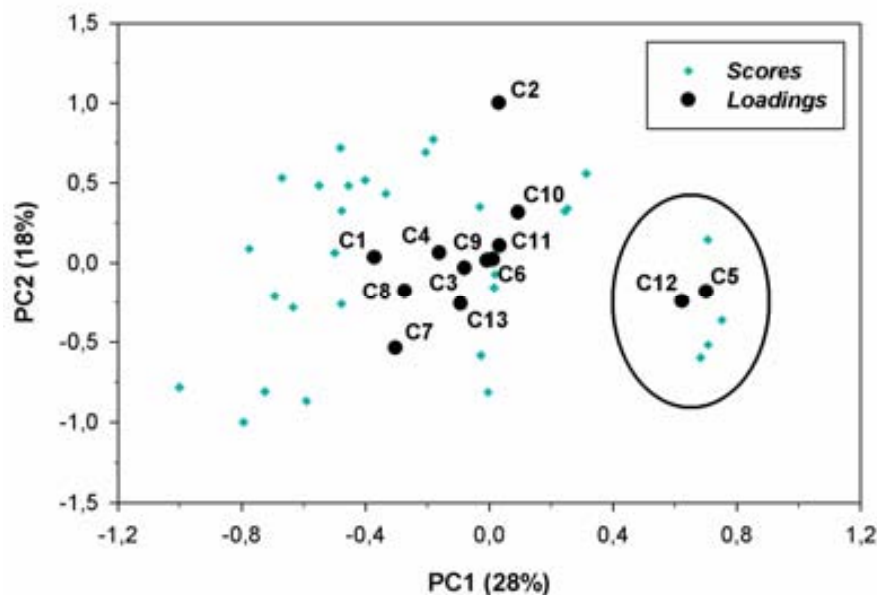


Figura 4.6. Gráfico bi-plot del PCA de composición para todas las muestras de calibración/validación. El círculo delimita todas aquellas muestras que contienen los crudos C5 y C12.

Una forma rápida de determinar qué composiciones incluye un grupo de calibración es proyectando las muestras que lo componen sobre el PCA de composición de todas las muestras disponibles, como se indica en el ejemplo de la Figura 4.7. En este caso las muestras proyectadas (circunferencias negras) se superponen a otras muestras en todos los casos. Las muestras no superpuestas

simbolizan aquellas composiciones no representadas en el grupo de calibración proyectado. Con el objetivo de incluir la variabilidad de todos los crudos/composiciones de crudos, las muestras necesarias serán añadidas a los grupos de calibración ya creados.

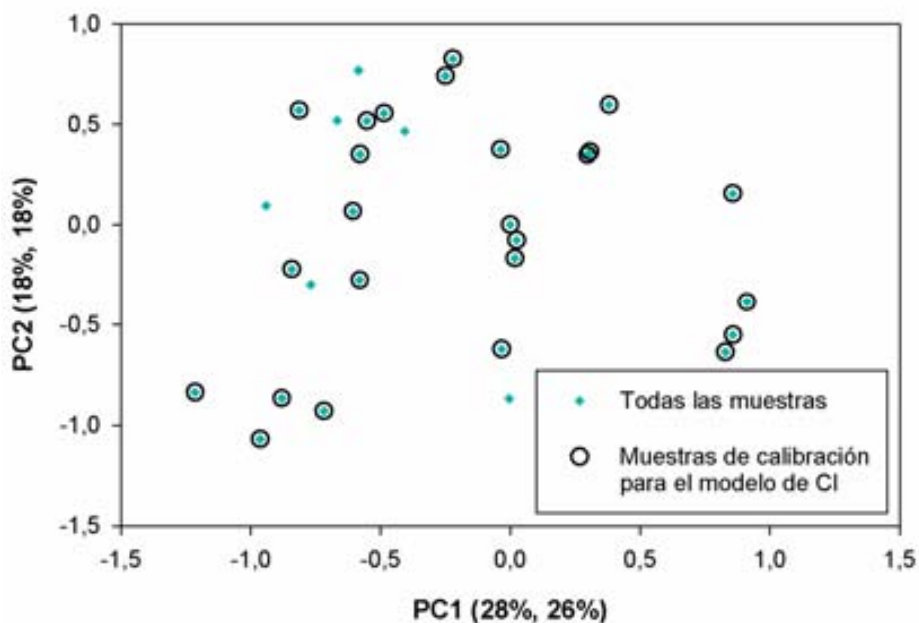


Figura 4. 7. Proyección de las muestras de un supuesto grupo de calibración sobre el PCA de composición de todas las muestras disponibles (ver Figura 4.6). Las muestras proyectadas se superponen a las muestras originales con idéntica composición de crudos.

4.3.2. Construcción de los modelos PLS

▪ Modelos originales

Para la determinación cuantitativa en planta de las propiedades de interés, se construyeron y optimizaron 7 modelos de calibración PLS, uno para cada propiedad a determinar (apartado 4.2.3). En la Tabla 4.3 se muestran las características y los estadísticos más relevantes de los modelos construidos con el programa The Unscrambler, la valoración de los cuáles se llevó a cabo en términos de promedios de los residuales y RMSEP.

Se observa que todos los modelos superan el 95% de varianza explicada con un número de factores que oscila entre 2 (para el CI) y 8 (para el destilado). Los

promedios residuales se mantienen en niveles aceptables en todos los casos y los valores de RMSEP son próximos a los de repetitividad de los métodos de referencia. Destaca la calidad de los estadísticos para los modelos de los tres primeros parámetros de la Tabla 4.3 (Dens, CI y FAME): todos ellos explican una varianza mayor del 97% sin superar los 4 factores en ninguno de los casos y presentan promedios residuales y valores de RMSEP muy bajos. El hecho de que estos tres parámetros se modelen mejor que los demás confirma la hipótesis obtenida de los análisis por PCA: se trata de propiedades muy relacionadas con el espectro y, por tanto, más fáciles de modelar.

Tabla 4.3. Características y estadísticos principales de los modelos construidos con The Unscrambler v.9.7.

		Dens (g/cm ³)	CI	FAME (%)	CP (°C)	D 95% (°C)	FP (°C)	SC (mg/kg)
Pretratamiento espectral		2ª Der. + SNV	2ª Der.	2ª Der.	2ª Der.	2ª Der. + SNV	2ª Der.	2ª Der. + SNV
Calibración^a	Factores	4	2	4	7	8	7	5
	Var-Y (%)	97.23	97.82	98.13	96.46	97.61	96.88	95.69
	RMSEC	0.0012	0.45	0.63	0.53	1.18	0.57	0.48
Validación externa^a	Promedio residuales	0.0000	-0.06	0.14	0.11	0.26	-0.01	0.04
	RMSEP	0.0016	1.02	0.67	1.22	5.39	3.72	2.56
Predicción^b	Promedio residuales	-0.0002	0.06	0.29	0.32	0.23	1.61	-0.66
	RMSEP	0.0016	0.7	0.51	1.15	4.03	2.90	1.88

^a Muestras recogidas hasta marzo de 2012

^b Muestras recogidas desde marzo hasta abril de 2012

Es también importante señalar la bondad de los estadísticos para SC, dado que el NIR es insensible a moléculas sin momento dipolar como es el caso del azufre.

Además, las cantidades de azufre en las muestras de gasóleo se encuentran por debajo del límite de detección del NIR (200 ppm). La bondad de los resultados se atribuye a la modificación de ciertas propiedades de la muestra en presencia de S.

▪ Modelos recalculados

El *software* instalado en el equipo de planta (CWS) no permite realizar los mismos pretratamientos espectrales aplicados con The Unscrambler e incluso el resultado de un mismo pretratamiento (e. g. derivadas) puede diferir entre ambos programas. Por esta razón, los modelos construidos con The Unscrambler no eran equivalentes a los desarrollados mediante CWS y proporcionaban resultados distintos en la predicción de los distintos parámetros. Para resolver este problema se estudiaron los diferentes pretratamientos del *software* CWS y se eligieron aquellos que proporcionan los resultados más semejantes a los de The Unscrambler: los modelos PLS construidos originalmente con 2ª derivada de Savitzky-Golay (11 puntos ventana y polinomio de 2º grado) se recalcularon en CWS con 2ª derivada *gap-segment* (*Gap* = 2 y *Segments* = 3). Aunque los resultados no son exactamente los mismos, las diferencias no son muy relevantes. Este proceso se ha realizado con todos y cada uno de los parámetros estudiados. En la Tabla 4.4 se evalúan los modelos recalculados con CWS. La valoración de estos modelos se llevó a cabo de la misma forma que para los modelos originales para que los valores sean comparables. Los pretratamientos se han mantenido en todos los casos y sólo se varía el número de factores en algunos casos con el fin de mantener los estadísticos de calibración lo más parecidos posible.

Tabla 4.4. Características y estadísticos principales de los modelos recalculados con CWS.

		Dens (g/cm ³)	CI	FAME (%)	CP (°C)	D 95% (°C)	FP (°C)	SC (mg/kg)
Pretratamiento espectral		2ª Der. + SNV	2ª Der.	2ª Der.	2ª Der.	2ª Der. + SNV	2ª Der.	2ª Der. + SNV
Calibración	Factores	4	4	5	5	8	3	5
	Var-Y (%)	97.16	97.35	98.28	92.62	97.15	86.45	95.83

	RMSEC	0.0012	0.50	0.60	0.77	1.29	1.19	0.47
Validación externa ^a	Promedio residuales	0.0001	-0.03	0.00	0.04	1.05	-0.18	-0.08
	RMSEP	0.0016	1.02	0.52	1.39	5.30	3.68	2.59
Predicción ^b	Promedio residuales	-0.0004	0.09	0.34	0.66	1.86	1.84	-0.78
	RMSEP	0.0012	0.64	0.57	1.41	4.79	2.12	1.89

^a Muestras recogidas hasta marzo de 2012

^b Muestras recogidas desde marzo hasta abril de 2012

Se observa que para los tres primeros parámetros (Dens, CI y FAME) y para el D 95%, los modelos mantienen los estadísticos de los modelos originales, aunque en algunos casos, ha habido que aumentar el número de factores. Para el resto de parámetros (SC, CP, y FP), los estadísticos han empeorado ligeramente respecto a los originales, pero sus predicciones y el valor de RMSEP revelan que los modelos son eficaces y predicen correctamente. Por lo que respecta al número de factores, este se ha mantenido en el caso del S, mientras que ha disminuido en los modelos para CP y FP. En todos los casos se ha elegido el número de factores óptimo para CWS.

4.3.3. Monitorización de los modelos

Los modelos construidos con el *software* CWS se aplicaron a la predicción en planta y se realizó el seguimiento periódico de las predicciones para comprobar que su eficacia se mantenía a lo largo del tiempo. Periódicamente se predijeron nuevas muestras de producción y se evaluaron los resultados obtenidos, analizando el valor de los residuales individuales de cada muestra para asegurar que la magnitud de los errores se mantenía y se distribuía de forma uniforme en el tiempo. En la Figura 4.8 se representan los residuales del parámetro densidad de todas las muestras registradas des de abril de 2012 hasta febrero de 2013. Se observa que los valores de residuales se

mantiene constantes y próximos a cero la mayor parte del tiempo, exceptuando tres muestras en que los valores absolutos de los residuales son claramente superiores. Estas desviaciones se observan también en los residuales de los parámetros restantes, excepto para punto de nube (resultados no mostrados), y coinciden temporalmente con paradas del equipo para solventar ciertos problemas con la sonda, registrados en julio de 2012. Este gráfico indica que el modelo mantiene la robustez y permite detectar derivas en las predicciones debidas a problemas en el equipo NIR. El promedio de residuales y el RMSEP para estas muestras, descartando aquellas afectadas por las derivas, se muestran en la Tabla 4.5.

Esta metodología ha demostrado su validez para predecir los siete parámetros de interés y detectar cualquier problema durante el proceso de registro. Cualquier nueva muestra puede ser analizada utilizando de nuevo los mismos criterios utilizados y métodos descritos para la selección de muestras (valores de referencia, datos espectrales y composición de muestras). La introducción de nuevos espectros en el modelo de calibración se valora mediante la proyección de los nuevos espectros en un gráfico PCA de los conjuntos de calibración/validación. Esto permite establecer fácilmente el origen del problema y decidir qué tipo de muestras hay que incluir en caso de requerir una actualización del modelo de calibración.

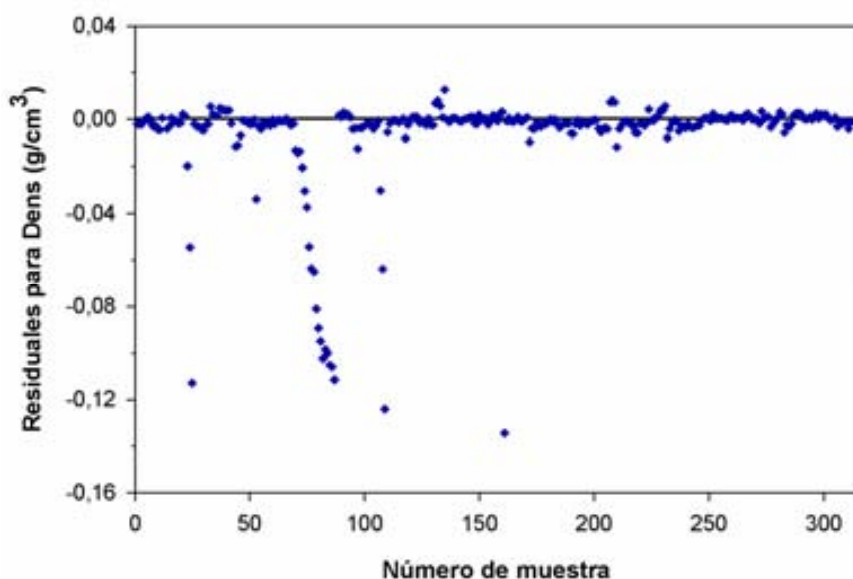


Figura 4.8. Representación gráfica de las nuevas muestras (ordenadas según el tiempo de producción) en función de sus residuales para el parámetro densidad.

Tabla 4.5. Residuales de las predicciones obtenidas mediante CWS para muestras recogidas des de mayo de 2012 hasta febrero de 2013.

Estadístico	Dens (g/cm ³)	CI	FAME (%)	CP (°C)	D. 95% (°C)	CP (°C)	SC (mg/kg)
Promedio residuales	-0.0017	-0.13	0.71	0.17	0.031	1.34	-1.47
RMSEP	0.0090	1.24	1.80	2.52	5.20	3.58	2.94

4.4. Conclusiones

Mediante el trabajo descrito en este capítulo, se evaluó la capacidad de la espectroscopia NIR para la monitorización en planta de siete propiedades fisicoquímicas de los gasóleos: densidad, índice de cetano, contenido en azufre y FAME, temperatura de destilación del 95%, punto de nube e inflamabilidad.

Partiendo únicamente de los espectros NIR registrados en planta de forma *on-line* se consiguió desarrollar modelos PLS eficaces para la determinación de cada uno de los parámetros de interés. Dada la gran variabilidad de las muestras de gasóleo, el buen funcionamiento de estos modelos se debe en gran parte a la cuidadosa selección de muestras llevada a cabo teniendo en cuenta tres criterios: rangos de valores de referencia, datos espectrales y composición de muestras durante un tiempo suficientemente largo como para acumular la variabilidad necesaria.

Actualmente, los modelos siguen en fase de control en planta para garantizar su funcionamiento *in situ* a lo largo del tiempo y asegurar su robustez. Los resultados obtenidos hasta el momento y la versatilidad y rapidez en el registro de muestras, demuestran que el NIR es una técnica idónea para el control de calidad de gasóleos en tiempo real, mejorando notablemente la funcionalidad de las técnicas de referencia usadas hasta el momento.

REFERENCIAS

- 1 UNE-EN 590:2009+A1:2010 (2010) Automotive fuels - Diesel - Requirements and test methods. *Comité Européen de Normalisation (CEN)*, Bruselas.
- 2 ASTM Standard D 975 - 12a (2012) Standard specification for diesel fuel oils. *ASTM International*, West Conshohocken, PA.
- 3 BOE-A-2010-13704 (2010), Boletín Oficial del Estado, N° 215 [en línea]. *Ministerio de Industria, Turismo y Comercio*, Madrid [Consultada en agosto de 2012]. Disponible en: http://www.boe.es/diario_boe/txt.php?id=BOE-A-2010-13704
- 4 Felizardo, P., Menezes, J. C., Neiva-Correia, M. J. (2011) Process analytical technology use in biofuels manufacturing, in: *PAT applied in biopharmaceutical process development and manufacturing: An enabling tool for quality-by-design*, Undey, C., Low, D., Menezes, J.C., Koch, M. (Eds.), CRC Press, Boca Raton, FL.
- 5 Zanier-Szydłowski, N., Quignard, A., Baco, F., Biguerd, H., Carpot, L., Wahl, F. (1999) Control of refining processes on mid-distillates by near infrared spectroscopy. *Oil Gas Sci. Technol.* **54**(4), 463-472.
- 6 López-García, C., Biguerd, H., Marchal-George, N., Schildknecht-Szydłowski, N. (2007) Near infrared monitoring of low conjugated diolefins content in hydrotreated FCC gasoline streams. *Oil Gas Sci. Technol.* **62**(1), 57-68.
- 7 Khanmohammadi, M., Garmarudi, A.B., de la Guardia, M. (2012) Characterization of petroleum based products by infrared spectroscopy and chemometrics. *TrAC-Trend. Anal. Chem.* **35**, 135-149.
- 8 ASTM E1790 - 04 (2010) Standard Practice for Near Infrared Qualitative Analysis. *ASTM International*, West Conshohocken, PA.
- 9 ASTM E1655 - 05 (2012) Standard Practices for Infrared Multivariate Quantitative Analysis. *ASTM International*, West Conshohocken, PA.
- 10 Macho, S., Larrechi, M.S. (2002) Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *TrAC-Trends Anal. Chem.* **21**, 799-806.
- 11 ASTM Standard D 86 - 99aE1 (1999) Standard test method for distillation of petroleum products at atmospheric pressure. *ASTM International*, West Conshohocken, PA.

- 12 Haaland, D.M., Thomas, E.V. (1988) Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **60**, 1193-1202.
- 13 Kennard, R.W., Stone, L.A. (1969) Computer aided design of experiments. *Technometrics* **11**, 137-148.
- 14 Monteiro, M.R., Ambrozin, A.R.P., Lião, L.M., Ferreira, A.G. (2008) Critical review on analytical methods for biodiesel characterization. *Talanta* **77**, 593-605.

5. CARACTERIZACIÓN DE LA COMPOSICIÓN DE PARAFINAS PARA SU APLICACIÓN INDUSTRIAL

5.1. Introducción	203
5.2. Experimental	205
5.2.1. Muestras	205
5.2.2. GC/MS.....	206
5.2.3. NMR	206
5.2.4. Métodos de referencia.....	207
5.2.5. NIR	207
5.3. Resultados y discusión	213
5.3.1. GC/MS con ionización por impacto electrónico	213
5.3.2. GC/MS con ionización química.....	213
5.3.3. NMR	218
5.3.4. Análisis de parafinas	220
5.3.5. Determinación por NIR.....	221
5.4. Conclusiones	225
REFERENCIAS	226

5. CARACTERIZACIÓN DE PARAFINAS PARA SU APLICACIÓN INDUSTRIAL

5.1. Introducción

La parafinas industriales son co-productos en la fabricación de aceites lubricantes en la industria petrolífera, con multitud de aplicaciones en distintos sectores industriales (ver capítulo 1). Estas aplicaciones se establecen a partir de sus propiedades físicas, químicas y funcionales, que dependen directamente de su composición química. De ahí la importancia de la caracterización de parafinas industriales, que se ha llevado a cabo en este trabajo. Para ello, se aplicó la cromatografía de gases acoplada a espectrometría de masas (GC/MS, *Gas Chromatography/Mass Spectrometry*) que permite establecer la longitud media de cadena hidrocarbonada [1,2,3,4], así como la proporción de *n*- e isoparafinas [2,3,5,6]. También se han llevado a cabo análisis por resonancia magnética nuclear (NMR, *Nuclear Magnetic Resonance*) [4,5]: ^1H -NMR, que indica el grado de ramificación de las cadenas hidrocarbonadas y ^{13}C -NMR, para la detección de hidrocarburos olefínicos y aromáticos.

Todos estos análisis se llevaron a cabo con el objetivo último de poder cuantificar el contenido de aceite (isoparafinas), un requisito en la caracterización industrial de parafinas. Estas isoparafinas, constituidas por una mezcla compleja de hidrocarburos ramificados, son muy difíciles de eliminar totalmente de la parafina y afectan negativamente a su aspecto y propiedades, limitando su aplicabilidad (apartado 1.1.4). Su separación industrial (desparafinado) se lleva a cabo por extracción con una mezcla de tolueno y MEK. Este método de separación industrial es la base del método propuesto por la *American Society for Testing and Materials*, ASTM D721 [7] para su determinación rutinaria mediante extracción con MEK a bajas temperaturas. Sin embargo, este método de análisis es lento y poco preciso, ya que implica operaciones como disolución, cristalización, filtración, etc. a temperaturas del orden de -30°C . Otro método analítico comúnmente utilizado es la NMR, aplicada sobre la base de que el aceite contiene un mayor número de grupos $-\text{CH}_3$ (mayor ramificación) que las cadenas lineales de las parafinas [8]. En este capítulo, además de la caracterización

de parafinas industriales mediante GC/MS y NMR, se proponen dos métodos nuevos (alternativos al análisis ASTM) para la determinación rápida y precisa del contenido de aceite en parafinas. El primero permite determinar el contenido total de isoparafinas mediante GC [1,9]; mientras que el segundo se basa en la espectroscopia NIR para la construcción de modelos PLS que permitan cuantificar el contenido de aceite extraíble con MEK. Como se ha descrito en el capítulo 2, el NIR es una técnica rápida y no destructiva que no requiere pretratamiento de la muestra, pero en este caso presenta un obstáculo claro: la similitud química y espectral del aceite (isoparafinas) y las *n*-parafinas que componen las muestras de parafina.

5.2. Experimental

5.2.1. Muestras

Las muestras son parafinas industriales hidrogenadas con bajo contenido en aceite (<2%). Proviene de distintos crudos de petróleo y han sido obtenidas en la refinería CEPSA (San Roque, España).

Se trabajó con tres tipos de muestras: A) muestras originales de producción, B) muestras de producción desaceitadas y C) muestras desaceitadas dopadas con cantidades conocidas de aceite en el laboratorio. Las parafinas desaceitadas proceden del mismo proceso industrial que el resto de parafinas y posteriormente son desaceitadas de nuevo en el laboratorio, siguiendo el método establecido en la norma ASTM D721 [7] (descrito en el capítulo 1). El proceso se repite hasta que la extracción de aceite es nula. Aunque la extracción de aceite nunca es completa, la parafina resultante se considera como parafina desaceitada.

El aceite obtenido en este proceso de extracción se ha utilizado para la preparación de las muestras desaceitadas dopadas (tipo C) con diferentes concentraciones de aceite. Para ello, se prepararon diluciones intermedias (6, 8, 10 y 12% en peso de aceite) añadiendo pequeñas cantidades de aceite a distintas muestras de parafina desaceitada (tipo B). Posteriormente se mezclaron estas diluciones intermedias con parafinas desaceitadas en proporciones previamente establecidas, para conseguir muestras con un contenido en aceite que cubre el rango de concentraciones entre 0 y 1.5% (m/m). En este proceso se utilizaron dos aceites procedentes de crudos distintos.

Una muestra de parafina desaceitada (tipo B) y otra de aceite fueron caracterizadas por GC/MS y NMR. El análisis mediante GC/MS se ha aplicado también a varias muestras de parafina de los tipos A y B, en concentraciones de 2000 ppm (m/m) en hexano. Estas muestras incluyen parafinas con contenidos de aceite comprendidos entre 0.00 y 1.51% según el método ASTM D721. Para la construcción de modelos por NIR se han utilizado 43 muestras de tipo B y C sin diluir. Todas ellas

producidas entre 2012 y 2013 y con concentraciones de aceite entre 0.00 y 1.50% (valores de referencia obtenidos por pesada).

5.2.2. GC/MS

Los análisis se realizaron en un cromatógrafo de gases HP 6890 Series II GC-System, con una columna capilar HP-5MS (5% fenil metilpolisiloxano) de 30 m de longitud x 0.25 mm x 0.25 μm de Agilent Technologies (Santa Clara, CA). En todos los casos, se tomaron alícuotas de 2 μl de la muestra diluida en hexano, que se inyectaron en el sistema a 310°C. El gradiente de temperatura aplicado fue de 2 °C/min entre 150-320 °C. Para facilitar la identificación de compuestos se inyectó un patrón $\text{C}_{36}\text{H}_{74}$ en las mismas condiciones. La detección se realizó mediante un espectrómetro de masas HP 5973 Mass Selective Detector acoplado al cromatógrafo. Se utilizaron dos métodos distintos de ionización: impacto electrónico a 70 eV e ionización química con metano como gas ionizante.

La adquisición de los cromatogramas y los espectros de masa se llevó a cabo mediante el *software* Enhanced MSD Chemstation versión E.02.00.493 de Agilent Technologies.

5.2.3. NMR

Las muestras, diluidas en deuterocloroformo, se analizaron por ^1H -NMR y ^{13}C -NMR. En ambos casos se utilizó un equipo de Bruker (Bruker Corporation) acoplado a una sonda de núcleo cuadrupolar (QNP, *Quadruple Nucleus Probe*) $^1\text{H}/^{13}\text{C}/^{19}\text{F}/^{31}\text{P}$ y trabajando a una frecuencia 250MHz. Para cada muestra se adquirieron 85 *scans*.

5.2.4. *Métodos de referencia*

Los valores de referencia de contenido de aceite en muestras de parafina se han obtenido aplicando dos métodos de referencia distintos según el tipo de muestra. En las muestras industriales tipo A, el contenido de aceite se ha determinado mediante el proceso descrito en la norma ASTM D721 [7], que consiste en disolver la muestra en MEK y enfriarla a -32°C , lo que produce la precipitación de las parafinas quedando el aceite en disolución. Posteriormente se filtra la parafina en frío y se evapora el disolvente; el contenido de aceite de la parafina se determina por pesada del residuo. Este método no permite cuantificar el aceite total de la muestra, sino solamente el aceite extraíble con MEK y además es poco preciso.

Para las muestras tipo C (dopadas en el laboratorio), el valor de referencia se ha establecido directamente por pesada del aceite añadido a parafinas desaceitadas (tipo B). A las parafinas desaceitadas se les asigna un valor de 0% m/m en aceite, aunque en realidad retienen cierta cantidad residual de aceite debido a la solubilidad del mismo en la parafina.

5.2.5. *NIR*

▪ **Características del equipo NIR**

Los espectros NIR se registraron mediante un espectrofotómetro FOSS NIRSystems 5000 de FOSS NIRSystems, Inc (Silver Spring, MD), equipado con un módulo calentador de viales (VHM, *Vial Heater Module*) (Figura 5.1), que mantiene las muestras a una temperatura constante de 60°C , en la que las parafinas se encuentran en estado líquido. Los registros se hicieron en modo transmitancia, a través de viales cilíndricos de 8x43 mm (6 mm de diámetro interno). Los viales vacíos, se utilizaron como blanco.



Figura 5.1. Detalle del módulo VHM, acoplado al espectrofotómetro NIR.

La adquisición de espectros se realizó mediante el *software* VISION v. 2.51 de FOSS NIRSystem. Los espectros, promedio de 32 *scans*, se registraron en el rango de 1100-2500 nm, con una resolución de 2nm. Por encima de 2200 nm la señal se satura y no es apropiada para la construcción de modelo de calibración. Cada muestra se ha registrado por triplicado, utilizando para la construcción y validación del modelo la media de los 3 espectros obtenidos.

▪ **Análisis de los datos espectrales**

La correlación espectral NIR entre los componentes de la muestra, parafina y aceite, es muy elevada (0.98 para el rango 1100-2200nm; Figura 5.2a) y las concentraciones de aceite en las muestras muy bajas (<2%), por lo que la contribución espectral debida al contenido de aceite en las muestras de parafina es muy pequeña (Figura 5.2b). Adicionalmente se observan pequeñas diferencias entre muestras de parafinas desaceitadas causadas por el distinto origen de los crudos de los que proceden. Por este motivo, es de especial interés eliminar o reducir cualquier tipo de contribución espectral no debida al contenido de aceite y que afecte a la construcción de los modelos de calibración. Se ha propuesto la aplicación del OSC [10,11], un

pretratamiento especialmente desarrollado para datos espectroscópicos NIR, que tiene por objeto eliminar aquella información espectral de la matriz \mathbf{X} no relacionada (ortogonal) con \mathbf{Y} (contenido de aceite).

Para la construcción de los modelos PLS, se ensayaron distintos pretratamientos (SNV, derivadas y OSC), siendo los más adecuados la 1ª derivada y el OSC. El algoritmo utilizado para la obtención de la 1ª derivada de los espectros fue el de Savitzky-Golay con una ventana de 11 puntos móviles y polinomio de segundo grado. El modelo OSC se creó con los espectros del conjunto de calibración y se aplicó a todas las muestras. En este caso, los mejores resultados se obtienen aplicando el modelo OSC sobre los espectros en 1ª derivada. El conjunto de calibración seleccionado es el mismo que para el modelo PLS y la elección de las muestras que lo componen se describe en el siguiente apartado. Por lo que respecta a los componentes del modelo, los mejores resultados se obtienen con 2 factores OSC lo que concuerda con lo descrito en la bibliografía [10]: el primer componente OSC corrige principalmente efectos de línea base, mientras que aumentar el número de componentes conlleva un gran riesgo de sobreajuste. Los resultados de la aplicación del modelo OSC se muestran en la Figura 5.3 y se comentan con más detalle en el apartado 5.3.5.

Ambos pretratamientos se han aplicado mediante el programa The Unscrambler de CAMO (Trondheim, Norway) y se han ensayado utilizando distintos rangos espectrales.

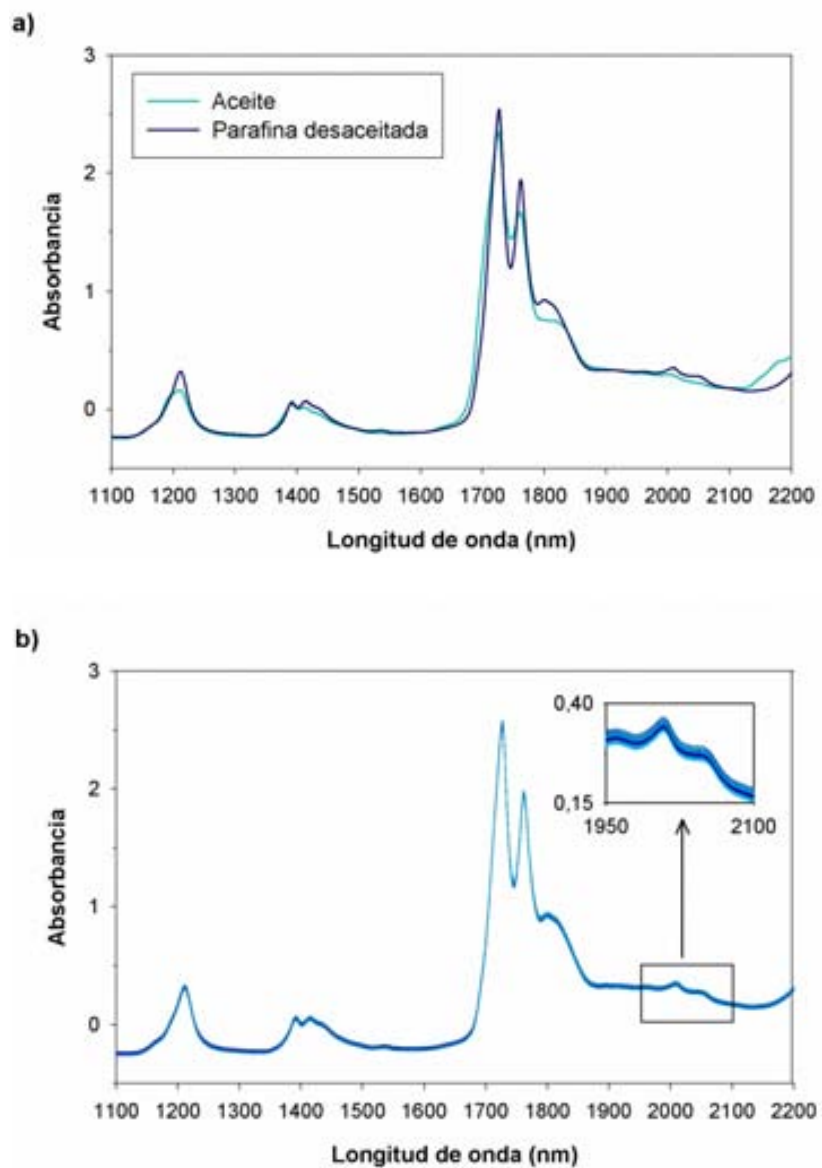


Figura 5.2. Espectros en absorbancia para (a) aceite y parafina desaceitada y (b) todas las muestras de parafina disponibles (0.0-1.5 % aceite).

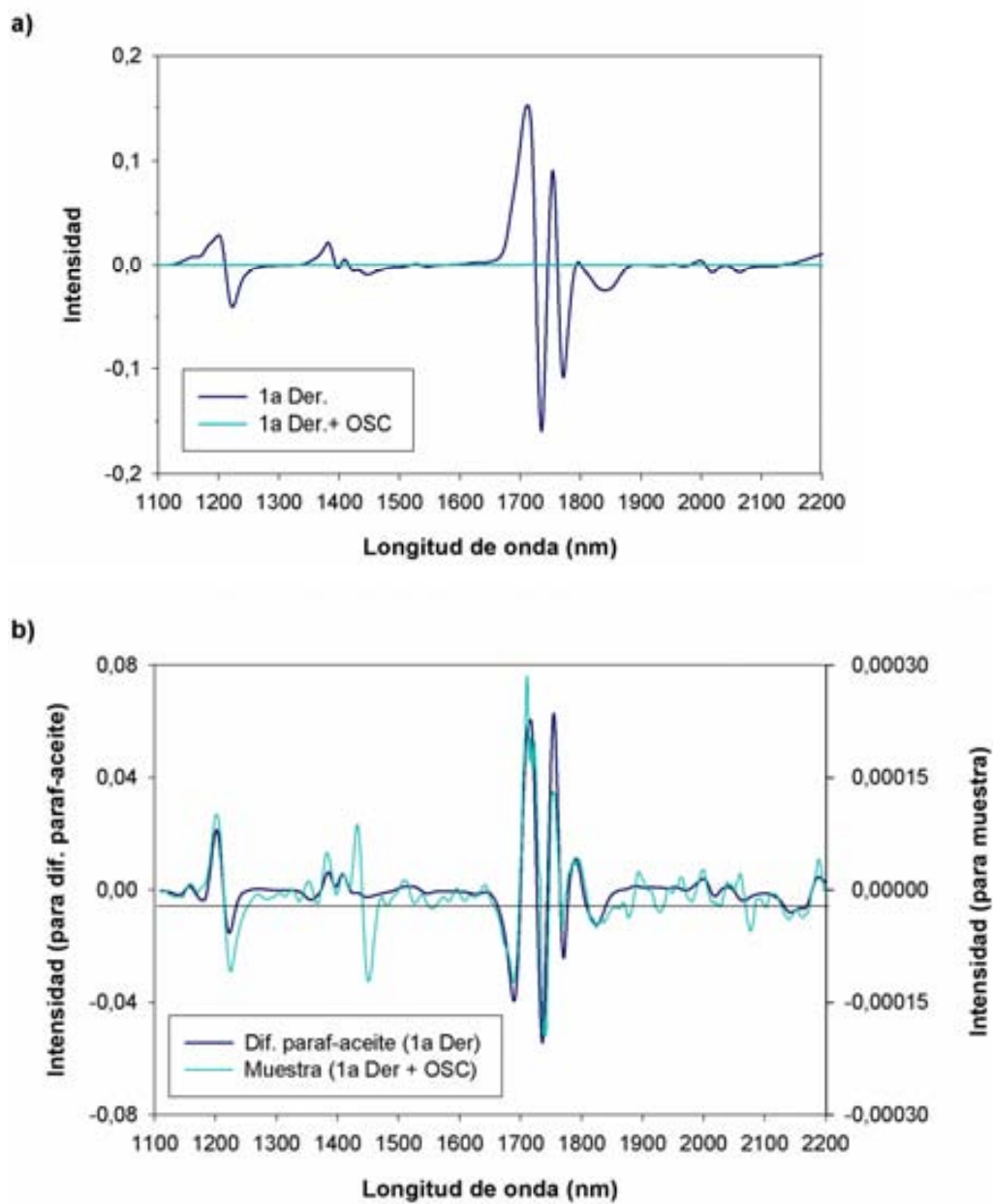


Figura 5.3. Efecto del pretratamiento OSC: (a) espectro en 1ª derivada para una muestra, antes y después de la aplicación del OSC; (b) espectro diferencia obtenido de restar el espectro del aceite del de la parafina (ambos en 1ª derivada,) después de la aplicación de OSC (rescalado).

▪ Modelos PLS

La determinación cuantitativa del aceite en parafinas se ha realizado construyendo modelos de regresión basados en el algoritmo PLS, ampliamente descrito en la bibliografía [12]. Este tipo de modelos requieren un conjunto de muestras de calibración cuyos valores de referencia sean conocidos. El conjunto de calibración debe cubrir todo el rango de la magnitud a determinar y ser representativo, tanto de las muestras de *training*, como de las muestras a predecir, incluyendo todas las posibles fuentes de variación espectral [13] (origen del crudo, proceso de refino, tipo de muestra). Estas consideraciones son cruciales para el desarrollo de modelos PLS robustos y con adecuada capacidad predictiva. Por ello, la elección del grupo de calibración se llevó a cabo manualmente, analizando los gráficos de *scores* del PCA de las muestras para cubrir todo el espacio espectral y asegurando que los valores de referencia cubrieran todo el rango de trabajo.

La construcción de los modelos se llevó a cabo mediante CV y la elección del número óptimo de factores se ha realizado tomando el menor valor del promedio residual de las predicciones, *i.e.* menor RMSEP. Para ello se empleó el programa The Unscrambler.

5.3. Resultados y discusión

5.3.1. GC/MS con ionización por impacto electrónico

El primer paso en la caracterización de las muestras de parafina y aceite fue el análisis por GC/MS con ionización por impacto electrónico. En la Figura 5.4 se muestran superpuestos los cromatogramas correspondientes a una muestra de parafina desaceitada y a otra de aceite. El cromatograma de parafina muestra picos bien resueltos y estrechos a intervalos de tiempo similares, que se atribuyen a *n*-parafinas (alcanos lineales); entre los picos de *n*-parafinas, aparecen otros picos pequeños que se atribuyen a isoparafinas (alcanos ramificados). Contrariamente, el cromatograma del aceite muestra una banda ancha no resuelta que induce a pensar en una mezcla de hidrocarburos de composición mucho más variada y heterogénea.

Los intervalos de elución para las dos muestras son muy semejantes, por lo que se puede afirmar que, la longitud de las cadenas de la parafina y del aceite es similar. Mediante esta técnica es prácticamente imposible establecer la longitud exacta de estas cadenas debido a la elevada fragmentación molecular y la consiguiente ausencia de pico molecular en el espectro de masas (los espectros de masas de los diferentes picos son muy semejantes). En la misma Figura 5.4 se muestra también el cromatograma de una *n*-parafina patrón, C₃₆H₇₄, obtenido en las mismas condiciones que las muestras. El hecho que el tiempo de retención de este patrón sea superior a los de las parafinas es indicativo de que la longitud de la cadena de las mismas es inferior a 36 átomos de carbono. Los espectros de masas de los picos cromatográficos para parafinas y aceites (no mostrados) son muy diferentes entre sí pero en ambos se detectan picos separados 14 unidades de masa entre señales consecutivas, correspondiente a sucesivas pérdidas de CH₂ e indicativos de hidrocarburos saturados.

5.3.2. GC/MS con ionización química

El MW de los componentes de las parafinas fue determinado mediante el análisis de las mismas muestras por GC/MS con ionización química, usando CH₄ como gas

ionizante para obtener iones $(M-H)^+$. Los cromatogramas de la parafina desaceitada y el aceite son idénticos a los obtenidos por impacto electrónico, pero incluyendo información sobre el MW de cada uno de los picos.

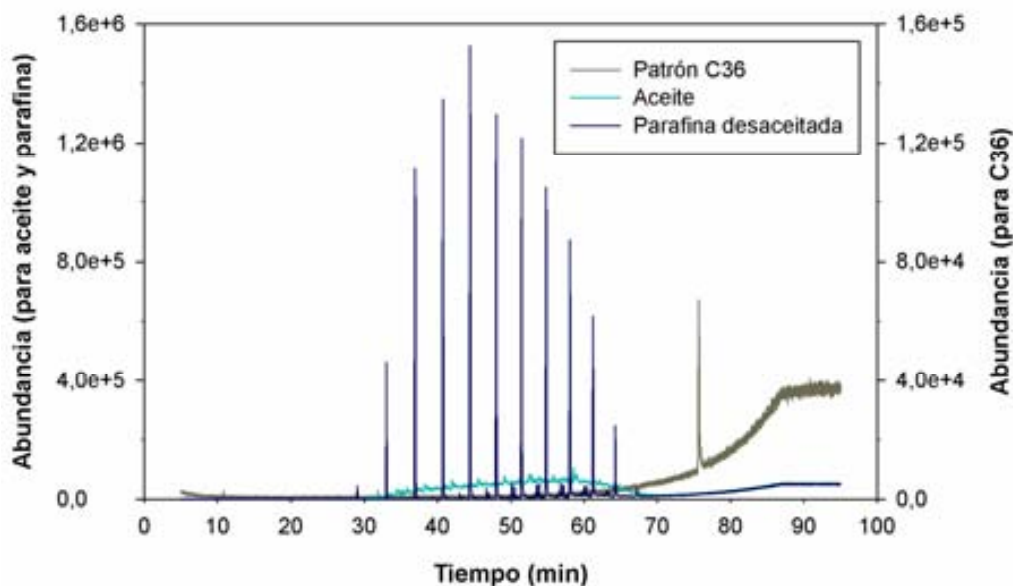


Figura 5.4. Cromatogramas superpuestos del patrón $C_{36}H_{74}$ (escala de la derecha), la muestra de aceite y la muestra de parafina desaceitada (escala de la izquierda).

▪ Parafina desaceitada

Los picos más intensos del cromatograma de la parafina dan lugar, en todos los casos, a espectros de masas similares al de la Figura 5.5a, correspondiente al pico de ~ 51.5 min en la Figura 5.4. Se observa que todas las señales del espectro presentan pérdidas de 14 unidades entre ellas y que la relación m/z para el pico base ($M-1$) es de 393. Este valor de $M-1$ corresponde al peso molecular de un hidrocarburo alifático de 28 átomos de carbono ($MW=394 \text{ g}\cdot\text{mol}^{-1}$). De la misma forma, se han analizado los espectros de masa para el resto de picos en el cromatograma de la parafina. Los MW de los picos base, incluyendo el patrón C_{36} , muestran una relación lineal con los tiempos de retención, tal como se describe en la bibliografía [14,15] para series homologas de compuestos. En la Tabla 5.1 se muestran los datos obtenidos. La regresión lineal entre los tiempos de retención (R_i) y los MW experimentales y

teóricos, dan lugar a las rectas $y=0.251x-48.149$ ($r^2=0.998$) e $y=0.245x-46.286$ ($r^2=0.997$) respectivamente. Estos resultados se utilizaron para establecer la composición de las parafinas desaceitadas (ver Tabla 5.1).

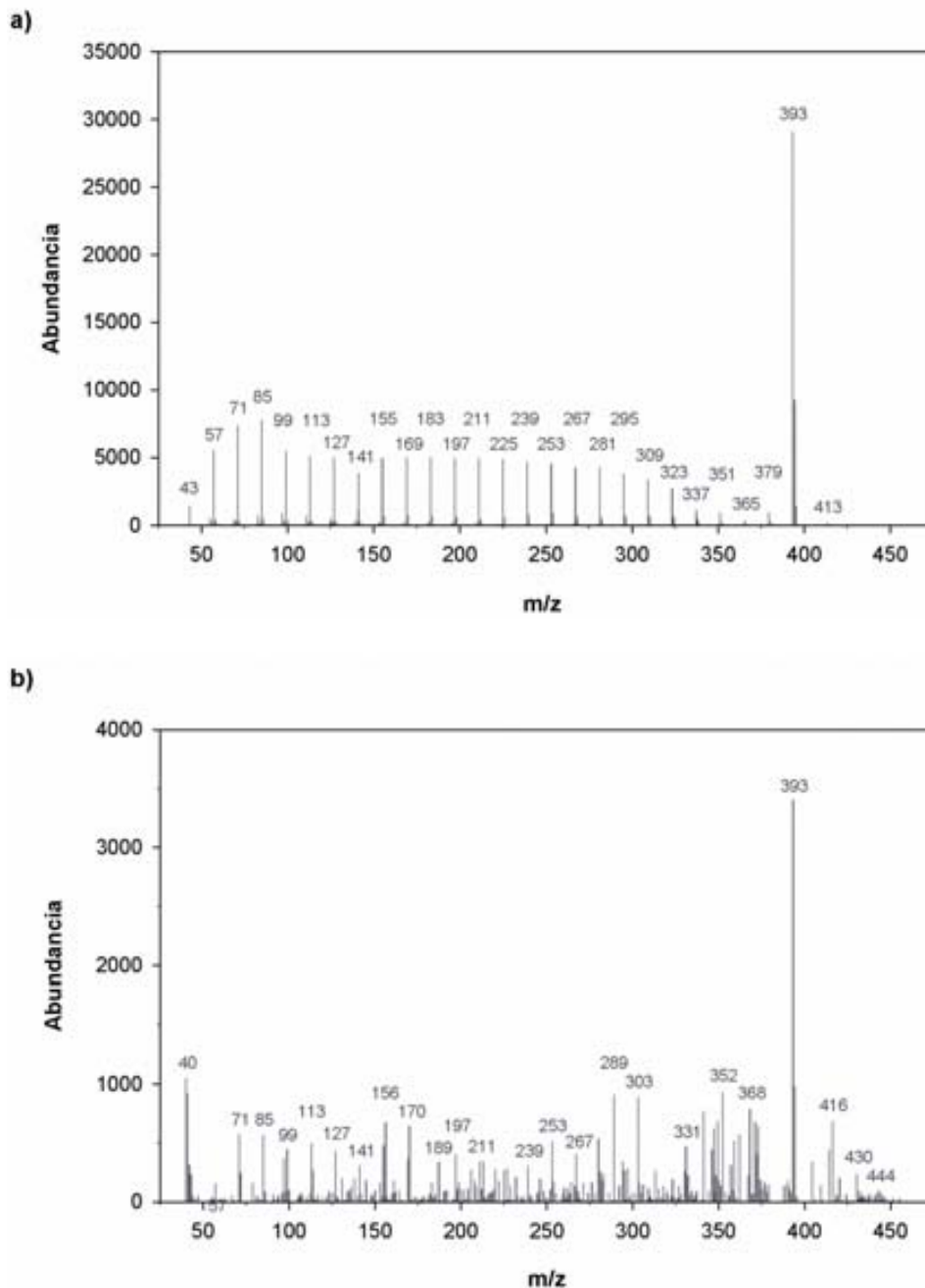


Figure 5.5. Espectros de masas por ionización química para el pico de ~ 51.5 min en el cromatograma de (a) la muestra de parafina y (b) la muestra de aceite.

Se ha procedido a la extracción de cromatogramas en la masa de los picos base de los espectros de masas, usando el modo *Single Ion Monitoring* (SIM). En la Figura 5.6 se muestra, a título de ejemplo, el cromatograma extraído para el pico de $m/z=393$ (similar al de los demás), donde se observa un pico mayoritario precedido por dos picos menos intensos. En la columna usada, los hidrocarburos alifáticos de mismo MW eluyen en el siguiente orden: isoparafinas > *n*-parafinas > cicloparafinas [1,2]. Se puede concluir que las dos señales más pequeñas detectadas a la izquierda del pico mayoritario corresponden a isómeros estructurales de este (*i. e.*, cadenas alifáticas isoméricas, que se pueden asignar a isoparafinas). Los picos débiles del final del cromatograma se atribuyen a fragmentos de parafinas de mayor longitud de cadena con el mismo MW.

Tabla 5.1. Tiempos de retención y números de carbono de los picos cromatográficos para las muestras de parafina desaceitada y el patrón C₃₆H₇₄ (última fila).

R_t (min) ^a	Exp. m/z ^b	MW teórico	Átomos C	Átomos H
24.7	---	296	21	44
28.7	309	310	22	46
32.7	323	324	23	48
36.6	337	338	24	50
40.4	351	352	25	52
44.1	365	366	26	54
47.7	379	380	27	56
51.2	393	394	28	58
54.5	407	408	29	60
57.6	421	422	30	62
60.9	435	436	31	64
63.9	449	450	32	66
66.8	-	464	33	68

69.7	-	478	34	70
72.5	-	492	35	72
75.7	-	496	36	74

^a Ver Figura 5.4

^b Obtenido del espectro de masas

▪ Aceite

El perfil cromatográfico del aceite está pobremente resuelto, por lo que no es posible obtener los espectros de masas de compuestos individuales. El espectro de masas (Figura 5.5b) también muestra diferencias de 14 unidades de masa entre algunas de las masas más relevantes, aunque aparecen otras muchas masas indicativas de la complejidad química (mayor isomería) de la muestra.

El cromatograma para un pico del espectro de masas extraído por SIM muestra una señal no resuelta mucho más ancha que en el caso de la parafina. El tiempo de retención del cromatograma para el pico de $m/z=393$ (Figura 5.6), es ligeramente menor para el aceite, que para la parafina; esto unido a la amplia banda de elución, confirma que el aceite está constituido por una mezcla compleja de hidrocarburos saturados ramificados (isoparafinas) con una gran variedad de isomerías.

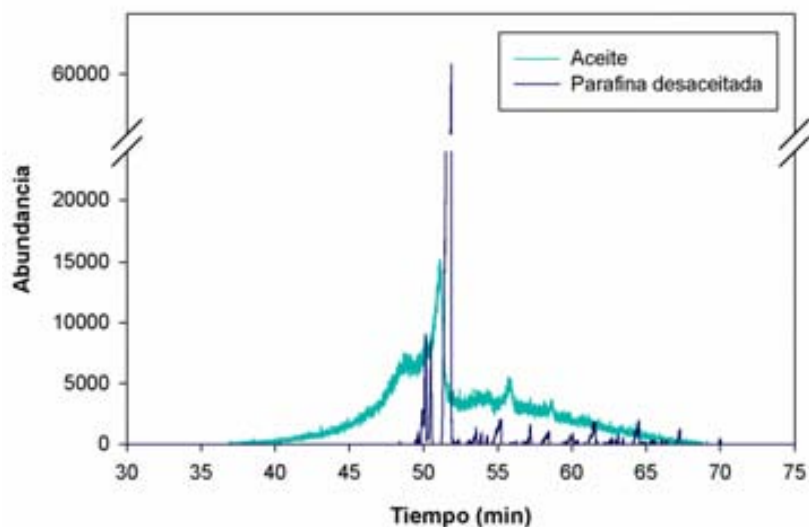


Figura 5.6. Cromatogramas para el pico localizado en $m/z = 393$ en el espectro de masas para el aceite (Figura 5.5b) y la parafina desaceitada (Figura 5.5a).

5.3.3. NMR

Ambos tipos de muestras (parafina desaceitada y aceite) han sido estudiados mediante $^1\text{H-NMR}$ (Figura 5.7) con el objeto de establecer el tipo de hidrocarburos presentes en la muestra y el contenido relativo de cada uno de ellos. El espectro NMR de la parafina muestra la existencia de dos tipos de protones que, por su desplazamiento químico ($\sim 1.25\text{ppm}$ y $\sim 0.9\text{ppm}$), corresponden a los grupos CH_2 y CH_3 respectivamente [5,16]. La integral de ambas señales indica una relación entre ellos de 8.82, lo que concuerda con una longitud media de cadena lineal de 28 carbonos; el valor verdadero podría ser ligeramente superior a causa de la presencia de isoparafinas en la muestra).

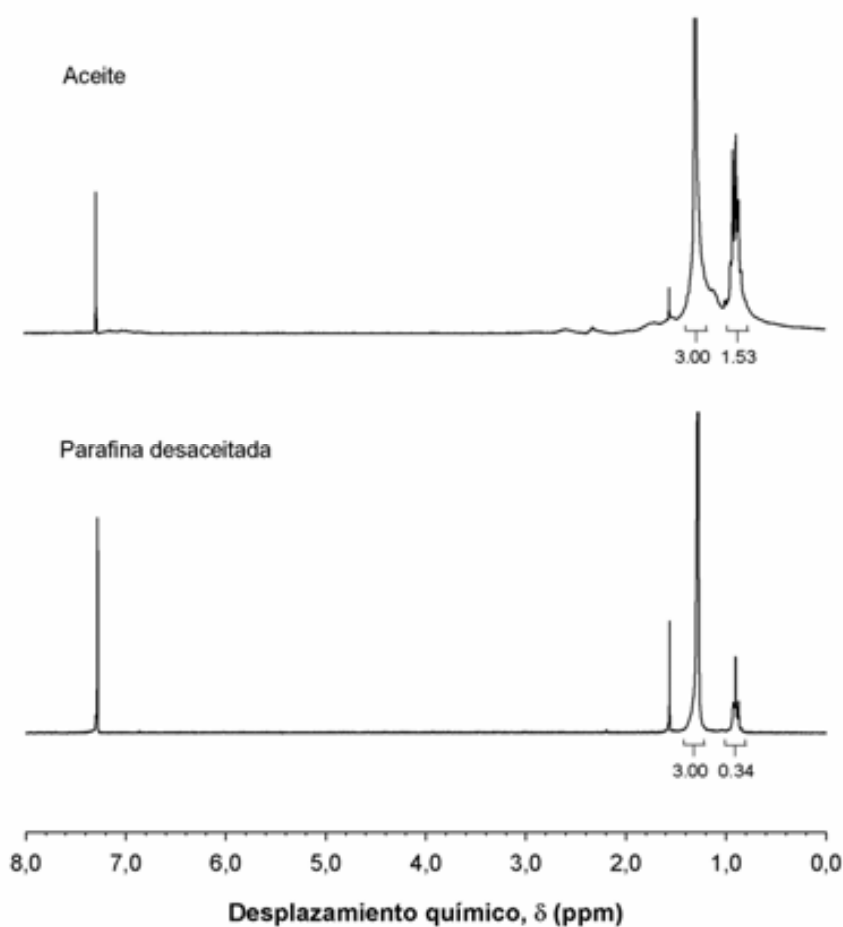


Figura 5.7. Espectro $^1\text{H-NMR}$ con los resultados de integración de los picos mayoritarios del aceite (arriba) y la parafina desaceitada (abajo). En ambos casos, el pico situado a $\sim 7.25\text{ppm}$ se asigna al solvente, deuterocloroformo.

En el espectro NMR del aceite la relación entre las integrales de las señales CH_2 y CH_3 es de 1.96. Para una cadena 28 carbonos, esta relación corresponde a 21 grupos CH_2 y CH (que se solapan en el espectro NMR) y 7 grupos CH_3 , indicativo del importante grado de ramificación de las cadenas hidrocarbonadas que componen la muestra de aceite.

En ninguno de los dos casos, se observa la presencia de protones aromáticos (6.5-9.2 ppm [4,16]), lo que confirma la eficiencia de su eliminación durante los procesos de refino.

También se han registrado los espectros de ^{13}C -NMR que permiten determinar el tipo de átomos de carbono presentes en la muestra, dando información sobre la longitud de las cadenas hidrocarbonadas y el grado de ramificación. Estos espectros muestran claramente que la muestra contiene únicamente cadenas saturadas, sin la menor señal de hidrocarburos olefínicos ni aromáticos.

5.3.4. Análisis de parafinas

Se ha realizado un análisis mediante GC de muestras de parafina de los diferentes tipos. Teniendo en cuenta la relación lineal entre MW y tiempo de retención, se puede establecer que la amplitud máxima de las cadenas de alcanos en las muestras es de 21 a 35 átomos de carbono, aunque la mayoría está comprendida entre 22 y 33 átomos de carbono; estas ligeras diferencias se atribuyen al distinto origen de los crudos. Los picos de los alcanos lineales muestran una distribución casi gaussiana, siendo el alcano C_{26} el más abundante. Suponiendo una respuesta igual de los diferentes alcanos lineales, el área de cada pico será proporcional a su concentración en la parafina y de ello se puede deducir que el MW medio de la parafina es de valor ligeramente superior al MW del alcano C_{27} (ver Tabla 5.2).

Tabla 5.2. Resumen de los resultados obtenidos de los datos cromatográficos para varias muestras de parafina.

Tipo de muestra	<i>n</i>	Contenido	Contenido	Átomos C en alcano	
		aceite (%)	isoparafinas (%) ^a	Rango máximo	Rango promedio
		<i>Según método ref. ASTM D721</i>	<i>Según integración picos cromatogr.</i>		
A	5	0.3–1.5	10.4–10.9	21–35	27.2–27.4
B	6	0.0	7.8–8.9	22–34	27.3–27.5

^a El resto corresponde a *n*-parafinas

No existen métodos analíticos para la determinación de aceite en parafinas, con excepción de los indicados por la ASTM para definir productos de interés industrial. En este trabajo se desarrolla un método basado en la relación de áreas de ambos componentes: alcanos ramificados y lineales. Dado que la mayoría de isoparafinas en parafinas purificadas forman parte de la porción de aceite, se puede considerar todos los alcanos ramificados de la muestra como contenido en aceite. Asumiendo que la respuesta del detector es proporcional a la concentración de cada componente, la relación entre áreas será también la proporción relativa de los componentes en la parafina industrial. Los resultados obtenidos para muestras de parafinas de los tipos A y B (Tabla 5.2) son congruentes con el conocimiento que tenemos de ellas. Las muestras de tipo A muestran un contenido de aceite comprendido entre 10.4 y 10.9% de aceite con un valor medio de 10.7%; mientras que las muestras que han sido desaceitadas en el laboratorio (tipo B) muestran un contenido de aceite comprendido entre 7.8 y 8.9% con un valor medio de 8.3%. Este amplio rango de concentraciones de aceite se explica por la variabilidad propia del procedimiento usado en la eliminación del aceite (ASTM D721). No se piensa que las diferencias en la distribución de los *n*-alcanos (debida al distinto origen de las parafinas industriales) sean responsables de esta variabilidad.

5.3.5. Determinación por NIR

El uso y las aplicaciones de las parafinas industriales se establecen a partir de parámetros definidos en las normas ASTM. En este trabajo se propone un nuevo método, alternativo al de la norma ASTM D721 y basado en la espectroscopia NIR, para la determinación del aceite remanente del proceso industrial.

El método requiere la construcción de modelos PLS que permitan predecir el contenido de aceite de las muestras industriales (tipo A). El rango de concentraciones que abarcan las muestras industriales es muy estrecho, por lo que resulta insuficiente para construir un modelo de calibración robusto. Para ampliar el rango de

concentraciones de trabajo hasta el 1.5%, se han preparado en el laboratorio muestras dopadas tipo C (como se describe en el apartado 5.2.1).

Las muestras tipo C y las muestras desaceitadas (tipo B) constituyen los grupos de calibración y validación. La selección de las muestras, así como la construcción y evaluación de los modelos PLS se llevaron a cabo como se describe en el apartado 5.2.5. El conjunto de calibración está formado por 27 muestras, mientras que las 16 muestras restantes se utilizaron para la validación externa del modelo. Las muestras industriales (tipo C) se reservaron para la predicción.

Los mejores resultados se han conseguido aplicando como pretratamientos 1ª derivada y 1ª derivada + OSC (apartado 5.2.5), con distintos rangos espectrales. Como se ha comentado anteriormente, la aplicación del OSC permite seleccionar únicamente aquella información relacionada con el parámetro de interés, en este caso, el contenido de aceite. Dado que los espectros de los componentes de las muestras están altamente correlacionados, la mayor parte de la matriz X es eliminada por el OSC y la información espectral remanente es muy reducida. Esto se constata en la Figura 5.3a donde se muestran, en la misma escala de intensidad, el espectro en 1ª derivada antes y después de aplicar el OSC. En la Figura 5.3b se muestra el mismo espectro después de aplicar el OSC (con escala ajustada) y se observa que las zonas con mayor intensidad coinciden con aquellas zonas del espectro con mayores diferencias entre parafina y aceite. Según el modelo OSC, las zonas más útiles para modelar el contenido de aceite no son las más parecidas al espectro del aceite, sino aquellas que recogen las mayores diferencias entre los espectros de parafina y aceite.

En la Tabla 5.3, se muestran las características y los estadísticos de los dos mejores modelos conseguidos con cada pretratamiento. En ambos casos, uno de los modelos está construido con el rango completo de longitudes de onda (1100-2200 nm), mientras que el otro se construyó con un intervalo espectral más reducido. En todos los modelos, la varianza explicada sobrepasa el 97% con un número de factores que oscila entre 5 y 8. El número de factores óptimo para los modelos disminuye cuando se añade el pretratamiento OSC a los espectros de las muestras. Lo que indica

que el OSC permite simplificar los modelos PLS. Se observa el mismo efecto al reducir el intervalo espectral utilizado, sea cual sea el pretratamiento. En ambos casos, el rango espectral que mejor funciona corresponde a la zona final del espectro NIR, con longitudes de onda entre 1800 y 2200.

Las predicciones para los grupos de calibración (RMSEC (B+C)) y validación externa (RMSEP (B+C)), presentan buenos resultados en todos los casos. Los valores de RMSEP aumentan ligeramente en la predicción de muestras industriales (RMSEP (A)). Esta diferencia es más acusada en términos de bias, lo que indica que el método de extracción y la imprecisión de este último ($R=0.23$) tienen un efecto determinante en la calidad de los resultados.

Tabla 5.3. Características y estadísticos de los modelos de calibración PLS.

Modelos PLS		1	2	3	4	
Pretrat. esp.		1ª Der		1ª Der + OSC (2F)		
Características modelos PLS	Rango (nm)	1100-2200	2000-2200	1100-2200	1800-2200	
	Factores PLS	8	6	7	5	
	Var-Y (%)	98.68	97.11	98.07	98.81	
<hr/>						
Calibración (B+C) (0.00-1.50% aceite)		RMSEC	0.05	0.08	0.07	0.05
<hr/>						
Valid. ext. (B+C) (0.10-1.50% aceite)		RMSEP	0.19	0.15	0.18	0.14
		Bias	0.05	0.00	0.06	0.03
<hr/>						
Predicción (A) (0.10-0.49% aceite)		RMSEP	0.22	0.82	0.56	0.79
		Bias	0.20	0.52	0.44	0.44
<hr/>						

Los modelos se simplifican al recortar el rango espectral y/o aplicar OSC, pero los estadísticos no mejoran significativamente, sino que los valores de RMSEP y bias aumentan ligeramente. El modelo PLS número 1, a pesar de requerir un mayor número de factores, es el más sencillo de todos en cuanto a pretratamiento (1ª derivada y rango completo) y presenta la mejor capacidad predictiva para muestras industriales de parafina.

5.4. Conclusiones

Se han caracterizado exhaustivamente muestras industriales de parafina y aceite mediante GC/MS y NMR y se han propuesto dos métodos nuevos para la determinación del contenido de aceite en parafinas.

Los resultados muestran que ambas muestras (parafina desaceitada y aceite) están constituidas por hidrocarburos saturados. En el caso de las parafinas, se trata principalmente de alcanos lineales (*n*-parafinas) con longitud de cadena comprendida entre 21 y 35 átomos de carbono y peso molecular medio correspondiente a hidrocarburos C₂₇. El aceite, en cambio, está constituido por una mezcla de hidrocarburos ramificados (isoparafinas) altamente isomerizada de número de átomos de carbono similar a los de la parafina. La separación conseguida en el cromatograma de GC/MS permite obtener el contenido total de aceite en las parafinas a partir de la integración de los picos.

El segundo método se basa en la espectroscopia NIR para la construcción de modelos PLS con espectros de muestras obtenidas por dopaje de muestras de parafina desaceitada. Los resultados obtenidos están en buena concordancia con los que proporciona la norma oficial (ASTM D721) y las ventajas que aporta en términos de rapidez y reproducibilidad hacen de este método una buena alternativa para la determinación de aceite en parafinas industriales.

REFERENCIAS

- 1 O'Connor, J.G., Burow, F.H., Norris, M.S. (1962) Determination of normal paraffins in C₂₀ to C₃₂ paraffin waxes by molecular sieve adsorption. molecular weight distribution by gas-liquid chromatography. *Anal Chem.* **34**, 83-85.
- 2 Levy, E.J., Doyle, R.R., Brown, R.A., Melpolder, F.W. (1961) Identification of components in paraffin wax by high temperature gas chromatography and mass spectrometry. *Anal. Chem.* **33**, 698-704.
- 3 ASTM D5442-93 (2008) Standard test method for analysis of petroleum waxes by gas chromatography. *ASTM International*, West Conshohocken, PA.
- 4 Espada, J.J., Coutinho, J.A.P., Peña, J.L. (2010) Evaluation of methods for the extraction and characterization of waxes from crude oils. *Energ. Fuel.* **24**, 1837-1843.
- 5 Musser, B.J., Kilpatrick, P.K. (1998) Molecular characterization of wax isolated from a variety of crude oils. *Energ. Fuel.* **12**, 715-725.
- 6 O'Connor, J.G., Norris, M.S. (1960). Molecular sieve adsorption: Application to hydrocarbon type analysis. *Anal. Chem* **32**, 701-706.
- 7 ASTM D721-06 (2006) Método de Ensayo Normalizado para Contenido de aceite en Parafinas del Petróleo, *ASTM International*, West Conshohocken, PA.
- 8 Sánchez, J. (2003) *Purificación de parafinas de petróleo por hidrogenación catalítica*. Rodríguez, F. (dir.) Tesis doctoral, Universidad Complutense de Madrid.
- 9 Coto, B., Coutinho, J.A.P., Martos, C., Robustillo, M.D., Espada, J.J., Peña, J.L. (2011) Assessment and improvement of n-paraffin distribution obtained by HTGC to predict accurately crude oil cold properties. *Energ. Fuel.* **25**, 1153-1160.
- 10 Wold, S., Antti, H., Lindgren, F., Öhman, J. (1998) Orthogonal signal correction of near-infrared spectra. *Chemometr. Intell. Lab.* **44**, 175-185.
- 11 Fearn, T. (2000) On orthogonal signal correction. *Chemometr. Intell. Lab.* **50**, 47-52.
- 12 Breton, R.G. (2007) *Applied Chemometrics for Scientists*. Wiley, Chichester.

- 13 Macho, S., Larrechi, M.S. (2002) Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *TrAC-Trend. Anal. Chem*, **21**, 799-806.
- 14 Destailats, H., Spaulding, R.S., Charles, M.J. (2002) Ambient air measurement of acrolein and other carbonyls at the Oakland-San Francisco Bay Bridge toll plaza. *Environ. Sci. Technol.* **36**, 2227-2235.
- 15 Craig, B.M., Murty, N.L. (1958) The separation of saturated and unsaturated fatty acid esters by gas-liquid chromatography. *Can. J. Chemistry* **38**, 1297-1301.
- 16 Pretsch, E., Bühlmann, P., Affolter, C., Herrera, A., Martínez, R. (2000 ed. original) *Determinación estructural de compuestos orgánicos*, Masson, Barcelona.

INDUSTRIA FARMACÉUTICA

6. DETERMINACIÓN DE LA DISTRIBUCIÓN DE API, EXCIPIENTES Y RECUBRIMIENTO EN COMPRIMIDOS FARMACÉUTICOS MEDIANTE NIR-CI

6.1. Introducción	233
6.2. Experimental.....	235
6.2.1 Las muestras	235
6.2.2. Instrumentación	236
6.2.3. Análisis de datos	238
6.3 Resultados y discusión.....	241
6.3.1 Cuantificación de API y excipientes.....	242
6.3.2. Cuantificación del recubrimiento.....	247
6.4. Conclusiones	252
REFERENCIAS	253

6. DETERMINACIÓN DE LA DISTRIBUCIÓN DE API, EXCIPIENTES Y RECUBRIMIENTO EN COMPRIMIDOS FARMACÉUTICOS MEDIANTE NIR-CI

6.1. Introducción

Debido a su estabilidad y facilidad de manipulación y administración, los comprimidos son la forma farmacéutica más utilizada actualmente [1,2]. El control de calidad de este tipo de productos implica la monitorización de un gran número de parámetros a lo largo del proceso de producción (capítulo 1), entre ellos la identificación y cualificación de la materia prima, así como la determinación de ciertas propiedades fisicoquímicas del producto elaborado.

Una técnica que está tomando importancia en el análisis de productos farmacéuticos a lo largo de todo el proceso de producción es el NIR-CI. El interés creciente de la industria farmacéutica por el NIR-CI se explica por su potencial como método PAT, que responde a la rapidez y simplicidad con la que el NIR-CI adquiere gran cantidad de información de toda la superficie de la muestra sin alterarla. El incremento en el uso del NIR-CI queda reflejado en la diversidad de estudios relacionados con el control de calidad de comprimidos farmacéuticos llevados a cabo en los últimos años: análisis de la homogeneidad de mezclas en polvo [3], determinaciones del tamaño de partícula [4], estudios de composición de fármacos [5], análisis de la distribución de componentes en comprimidos sólidos [6], determinación de la uniformidad de contenido [7]. Entre los análisis más relevantes para la calidad de los comprimidos, destacan aquellos relacionados con la identificación y cuantificación de API [8], como es el aseguramiento de la uniformidad de contenido en los comprimidos finales. La determinación de este parámetro mediante NIR-CI permite además visualizar la distribución de los distintos componentes en la muestra. El inconveniente de este tipo de análisis no destructivos es que la contribución de los componentes de estudio a la señal analítica puede quedar enmascarada por la laca de recubrimiento.

El recubrimiento es la etapa final del proceso de producción de los comprimidos farmacéuticos y consiste en la aplicación de una fina película que se adhiere al comprimido, sin modificar su forma original. Las funciones y características generales del recubrimiento se describen en el apartado 1.2.1. En cualquier caso, la utilidad del recubrimiento depende en gran parte de su grosor y de su distribución en la superficie del comprimido. Un recubrimiento demasiado fino o demasiado grueso puede afectar a su funcionalidad, haciendo el núcleo más susceptible a factores externos o modificando la velocidad de liberación del API, por ejemplo. Es por este motivo, que el recubrimiento también constituye un aspecto importante en el control de calidad de fármacos sólidos. Para ello, no es suficiente con determinar el grosor medio de lacado de un comprimido, sino que es necesario disponer de información más detallada acerca de su distribución en toda la superficie del mismo para asegurar su homogeneidad. En este sentido, el uso del NIR-CI es ventajoso, ya que permite el análisis del grosor de recubrimiento en toda la superficie de la muestra. Aunque existen algunas referencias que tratan este tema [9,2], la aplicación del NIR-CI al estudio del recubrimiento no está tan extendida como las mencionadas más arriba.

En este capítulo se pretende analizar cuantitativamente la concentración y distribución de los distintos componentes de núcleos (comprimidos sin recubrir) y comprimidos farmacéuticos con distintos grosores de recubrimiento, así como determinar el grosor y la distribución de la capa de laca del recubrimiento. Todo ello se lleva a cabo a partir de imágenes NIR-CI. La extracción de los datos de interés y la construcción de los modelos de calibración para cada componente se consigue con la aplicación de los algoritmos PLS e Isys-PLS1.

6.2. Experimental

6.2.1 Las muestras

Las muestras usadas para este trabajo son comprimidos farmacéuticos sin recubrimiento (núcleos) y comprimidos recubiertos con distintos grosores de capa. Todos los núcleos presentan la misma composición en API y excipientes, con valores nominales de concentración del 35% de API y 65% de la suma de excipientes, de los cuáles: 40% de excipiente #1, 20% de excipiente #2, 2% de excipiente #3, 1.5% de excipiente #4 y 1.5% de excipiente #5. El API y los dos excipientes mayoritarios representan un 95% del total del comprimido. En el caso de los comprimidos recubiertos, la presencia de la laca implica un cambio en las proporciones de API y excipientes: los porcentajes de los diferentes componentes disminuyen ligeramente a medida que aumenta el grosor del recubrimiento. Las variaciones para los componentes mayoritarios en los distintos niveles de recubrimiento se recogen en la Tabla 6.1.

Tabla 6.1. Composición (% m/m) de API y los dos excipientes mayoritarios para comprimidos con distintos grosores de recubrimiento.

Nivel de recubrimiento	Composición muestra (% m/m)		
	API	E #1	E #2
0x	35.0	40.0	20.0
0.5x	34.4	39.3	19.7
1x	33.9	38.8	19.4
2x	33.0	37.7	18.8
3x	32.0	36.6	18.3

Por lo que respecta al nivel de recubrimiento, este se establece en función del tiempo empleado para ello. Se considera que el proceso de recubrimiento es constante y reproducible con el tiempo, de modo que si el tiempo de recubrimiento estándar en muestras de producción da un grosor de recubrimiento de x , el doble o el triple de este tiempo dará lugar a grosores de recubrimiento de $2x$ y $3x$. De la misma forma, se deduce que los núcleos presentan un grosor de recubrimiento de $0x$. Así, los

comprimidos utilizados presentan un rango de grosores de recubrimientos de entre $0x$ y $3x$, divididos en 5 niveles: $0x$, $0.5x$, $1x$, $2x$, $3x$ (Tabla 6.1).

El grosor de la película de recubrimiento de los comprimidos fue calculado midiendo la superficie aproximada del comprimido (1.92 cm^2), la diferencia de masa de un mismo comprimido antes y después de recubrir (de 290 a 317 mg) y la densidad real del polímero de recubrimiento ($1.19 \text{ g}\cdot\text{cm}^{-3}$). El resultado de esta medida fue de $39.4 \text{ }\mu\text{m}$ de grosor para el tiempo de recubrimiento estándar (x), $21.9 \text{ }\mu\text{m}$ para $0.5x$, $78.8 \text{ }\mu\text{m}$ para $2x$ y $118.82 \text{ }\mu\text{m}$ para $3x$. La medida de grosor estándar obtenida ($39.4 \text{ }\mu\text{m}$) es comparable a los $40 \text{ }\mu\text{m}$ obtenidos mediante microscopía electrónica de barrido (SEM, *Scanning Electron Microscopy*). Esta comparación no fue posible para el resto de niveles de recubrimiento, por lo que, en este trabajo todos los grosores de recubrimiento se refieren al tiempo estándar de recubrimiento ($x \sim 40 \text{ }\mu\text{m}$).

Además de los comprimidos, también se han utilizado muestras en polvo de todos los componentes puros de la formulación. Las muestras en polvo se utilizaron para la construcción del modelo de cuantificación de API y excipientes, que posteriormente se aplicó a comprimidos de todos los niveles de recubrimiento. Por lo que respecta al modelo para determinación del grosor de recubrimiento, se crearon conjuntos de calibración y validación con comprimidos, abarcando en ambos casos todo el rango de grosores de laca ($0-3x$). El modelo obtenido se aplicó a las muestras de producción, el objeto de este estudio.

6.2.2. Instrumentación

Para la adquisición de las imágenes se utilizó la cámara hiperespectral SyNIRgy™ Chemical Imaging System de Malvern Instruments (Malvern, UK) (Figura 6.1), equipada con un detector de tipo InSb *FPA (stirling cooled)* de 320×256 píxeles y controlada por el *software Pixys*® 1.1, también de Malvern. Las muestras se colocaron en la pletina y se iluminaron con 4 lámparas halógenas situadas alrededor en un ángulo apropiado ($45^\circ-0^\circ-45^\circ$), para maximizar la señal de la reflectancia difusa. La distancia del objetivo, dotado de filtros polarizadores para evitar reflejos, se ajustó la

según el tamaño del objeto. En las condiciones seleccionadas, las imágenes tienen una resolución espacial de $40 \times 40 \mu\text{m}$ por píxel. El registro se llevó a cabo en modo de reflectancia difusa, cubriendo un rango de longitudes de onda comprendido entre 1200-2400 nm, a una resolución espectral de ~ 10 nm. El tiempo de adquisición de una imagen, en estas condiciones fue de unos 3 minutos aproximadamente.



Figura 6.1. Cámara hiperespectral SyNIRgyTM (Malvern Instruments).

Se registraron 6 imágenes de cada núcleo: 4 de cada una de las caras exteriores (A-D) y 2 de las caras interiores (E-F) obtenidas por partición de los comprimidos, como se indica en la Figura 6.2a. El análisis de las caras interiores se llevó a cabo para reafirmar la homogeneidad de las muestras, aunque el método analítico no requiere la destrucción de las muestras. En el caso de las muestras de producción, se registraron sólo las 4 caras exteriores de los comprimidos, denominadas de la misma forma que en los núcleos (A-D). Además, para todas las muestras, se han registrado 2 imágenes de cada comprimido, pertenecientes a 2 caras exteriores opuestas. En este último caso, cada imagen incluye 2 caras homólogas de 2 comprimidos de iguales características (mismo grosor de lacado), tal y como se muestra en la Figura 6.2b. Dado que estas imágenes se usaron para la construcción del modelo de calibración para determinación del grosor de recubrimiento, no fueron consideradas en su totalidad. Sólo se han tenido en cuenta porciones rectangulares interiores de cada comprimido con el fin de evitar las distorsiones producidas por la luz en los bordes de los

comprimidos de la imagen, que podrían afectar a la calidad del modelo desarrollado posteriormente.

Las imágenes de los compuestos puros en polvo se obtuvieron colocando una cantidad suficiente de los mismos en una cubeta de vidrio y aplanando la superficie antes de adquirir la imagen.

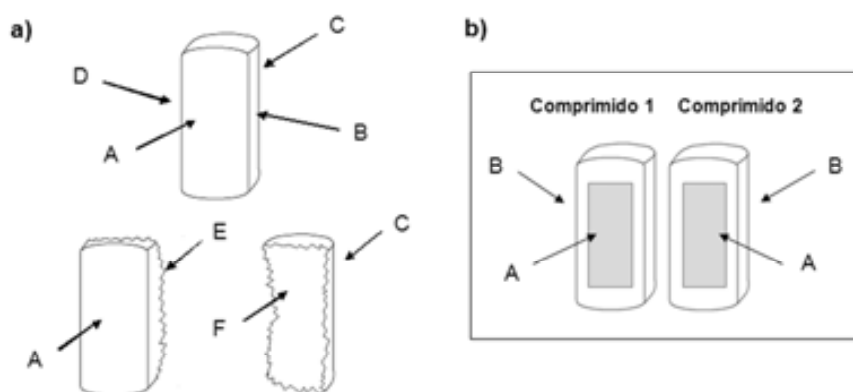


Figura 6.2. Esquema de las distintas caras analizadas para: a) núcleos y comprimidos de producción y b) todos los comprimidos. El área sombreada en gris en b) corresponde al área seleccionada para el análisis.

6.2.3. Análisis de datos

Previamente a la aplicación de cualquier pretratamiento espectral o de cualquier algoritmo para la extracción de la información el cubo de datos tridimensional de la imagen hiperespectral se desdobló, obteniendo una matriz bidimensional. Una vez desdobladas, las matrices de datos obtenidas para cada imagen fueron pretratadas mediante el algoritmo SNV.

Para la cuantificación de API y excipientes, todos los pretratamientos espectrales utilizados, así como los algoritmos de modelado se aplicaron mediante el *software* Isys 5.0 de Malvern. A la matriz de datos resultante del pretratamiento, se le aplicó el algoritmo PLS de Isys (Isys-PLS1). La información proporcionada por

Malvern sobre este algoritmo es mínima, pero se entiende que se trata de una variación del PLS-DA que permite cuantificar individualmente cada uno de los componentes de la muestra, usando únicamente su espectro puro. A diferencia del CLS, este algoritmo no requiere el espectro puro de todos los componentes de la formulación, sino que su construcción únicamente precisa la adquisición del espectro puro de los componentes a cuantificar [10]. Los espectros puros fueron utilizados para la creación de la biblioteca, sobre la cual se construyeron los modelos de calibración Isys-PLS1. Estos modelos se acaban de definir con la elección del número de factores apropiado y el *threshold* a aplicar. Un aumento en el número de factores permite incrementar la varianza explicada en Y por el modelo, pero, a la vez, también incrementa el riesgo de sobreajuste, por lo que es necesario encontrar un punto de compromiso óptimo. El *threshold* permite ajustar los límites de discriminación entre clases, por lo que su elección depende del número de factores escogido previamente y de la similitud entre el espectro puro del componente a cuantificar y los espectros puros de los demás componentes de la muestra. Al aplicar el modelo obtenido a una imagen, este procede ponderando los datos espectrales entre 0 y 1 en función de su similitud con el espectro puro almacenado la biblioteca del componente a cuantificar. Esta escala se aplica para cada píxel de la imagen y corresponde a una concentración del 0 al 100% del componente en cuestión. El resultado se presenta en forma de mapa de concentraciones de cada imagen para cada uno de los componentes presentes en la librería.

Para la determinación del grosor y la distribución del film de recubrimiento, en cambio, las rutinas usadas para el desdoblamiento y la reorganización de los datos fueron programadas en código MATLAB con MATLAB v 7.0 de The MathWorks (Natick, MA), mientras que la aplicación de los pretratamientos y la construcción de los modelos se llevó a cabo con The Unscrambler v.9.8 de CAMO (Trondheim, Norway). El algoritmo de regresión utilizado en este caso fue el PLS [11] de The Unscrambler, que utiliza como datos espectrales X los espectros medios del área seleccionada de cada imagen y como Y , los valores de referencia de grosor del recubrimiento. El modelo PLS construido, se validó por CV aplicando el algoritmo *jack-knife* para determinar el rango de longitudes de onda más correlacionado con la

propiedad a modelar. Posteriormente es necesaria la elección del número óptimo de factores que se llevó a cabo siguiendo los mismos criterios descritos anteriormente. La plasmación de los resultados en los mapas de distribución se consiguió reorganizando los resultados de las predicciones PLS de la misma forma que en la imagen original.

6.3 Resultados y discusión

El objeto de este estudio es la aplicación del NIR-CI para el control de la distribución de los distintos componentes de una nueva formulación farmacéutica, aprovechando esta información para su aplicación en las etapas de desarrollo del producto. A partir de los datos obtenidos mediante NIR-CI se han desarrollado métodos cuantitativos para la determinación de API, excipientes y también recubrimiento, así como su distribución en los comprimidos farmacéuticos.

En la Figura 6.3 se muestran los espectros en absorbancia de los distintos componentes puros (API y los 5 excipientes). Se observan diferencias claras entre la mayoría de los espectros, corroboradas por los CC entre pares de componentes puros mostrados en la Tabla 6.2. Destaca la elevada similitud espectral entre los excipientes #2 y #4 (CC = 0.984), lo que dificulta el posterior desarrollo de los modelos de calibración para estos dos componentes.

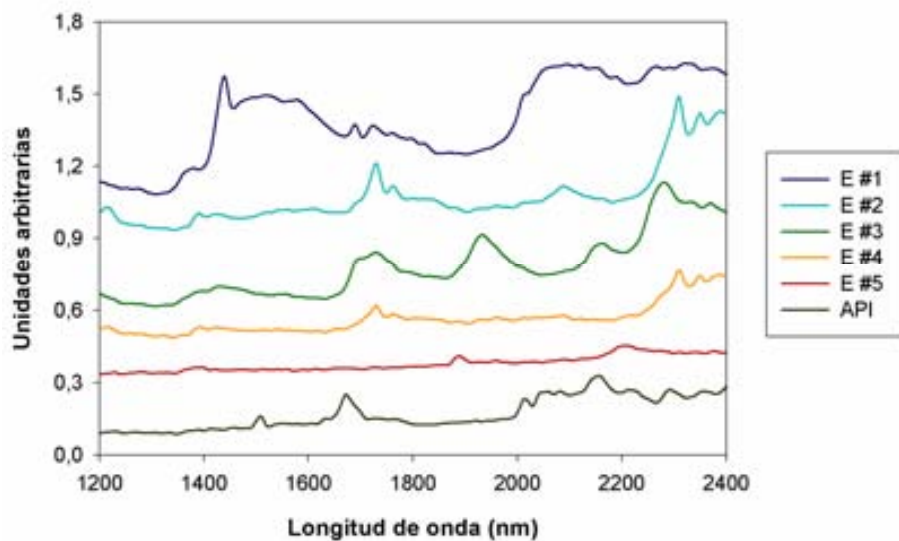


Figura 6.3. Espectros NIR en absorbancia de todos los componentes puros de la formulación (API y excipientes). Los espectros han sido obtenidos promediando el espectro de cada píxel de un área de la imagen hiperespectral. Cada espectro ha sido desplazado de forma arbitraria para facilitar su visualización.

Por lo que respecta al recubrimiento, la contribución espectral introducida por la capa de laca es baja. Esto se aprecia en la Figura 6.4, donde se comparan los

espectros promedio de un área de las imágenes registradas para un núcleo y un comprimido lacado. Se observa que la laca provoca un ligero aumento de la línea base del espectro medio.

Tabla 6.2. Coeficientes de correlación entre los espectros NIR promedio en absorbancia de todos los componentes puros de la formulación (API y excipientes) en todo el rango de trabajo.

Coeficientes de correlación						
	API	E #1	E #2	E #3	E #4	E #5
API	1.000					
E #1	0.810	1.000				
E #2	0.621	0.601	1.000			
E #3	0.693	0.603	0.849	1.000		
E #4	0.640	0.588	0.984	0.903	1.000	
E #5	0.823	0.693	0.669	0.847	0.740	1.000

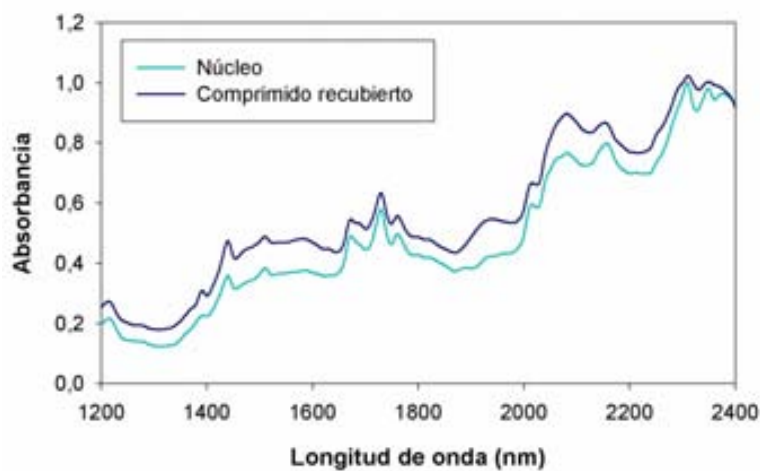


Figura 6.4. Espectros NIR de un comprimido recubierto y uno sin recubrir (núcleo). Los espectros han sido obtenidos promediando el espectro de cada píxel de un área de la imagen hiperespectral.

6.3.1 Cuantificación de API y excipientes

▪ Núcleos (0x)

Como se ha descrito anteriormente, para determinar la concentración de API y los 5 excipientes presentes en los comprimidos, se construyeron distintos modelos (uno

para cada analito), usando en todos los casos el algoritmo PLS del *software* Isys 5.0. En este apartado, se presentan los resultados de su aplicación a núcleos, para evaluar los resultados eludiendo el efecto del recubrimiento. Los modelos utilizados presentan un número de factores variable, comprendido entre 3 y 8, según el analito a modelar: a concentraciones de analito menores en la muestra, mayor es el número de factores necesario para la construcción del modelo. Los valores de concentración obtenidos para cada componente en cada una de las caras de un núcleo se muestran en la Tabla 6.3.

En todas las caras de los núcleos analizados, las concentraciones predichas por el modelo son muy próximas al valor nominal, tanto para el API y los excipientes mayoritarios como para los excipientes minoritarios. Las mayores desviaciones estándar se observan para los dos componentes mayoritarios (API y excipiente #1) lo que es indicativo de una distribución homogénea incompleta. Los componentes minoritarios, en cambio, muestran valores de desviación estándar muy bajos en todos los casos, indicativo de la buena homogeneidad de su distribución. Destaca el hecho que la suma de los porcentajes de todos los componentes para cada cara, se aproxima al 100% en todos los casos, lo cual es un indicativo de la exactitud y la capacidad predictiva de los modelos de calibración.

Tabla 6.3. Concentraciones de API y excipientes (% m/m) obtenidas aplicando el modelo Isys-PLS1 a las diferentes caras (nombradas según se indica en la Figura. 1a) de un núcleo.

	Composición del núcleo (% m/m)						Suma
	API	E #1	E #2	E #3	E #4	E #5	
Cara A	35.51	38.93	20.15	1.97	1.48	1.47	99.51
Cara B	35.45	39.48	21.12	1.98	1.48	1.51	101.00
Cara C	37.11	38.45	19.69	1.99	1.53	1.53	100.30
Cara D	37.45	40.38	20.14	1.95	1.53	1.51	102.96
Cara E	31.61	44.16	18.62	1.95	1.48	1.51	99.38
Cara F	31.47	43.38	19.31	1.97	1.55	1.53	99.21
Media	34.78	40.79	19.84	1.97	1.51	1.51	100.40

Desv. Est.	2.62	2.41	0.85	0.02	0.03	0.02	5.95
------------	------	------	------	------	------	------	------

Sin embargo, se detectan pequeñas diferencias de composición entre las caras del comprimido para los tres componentes mayoritarios. La concentración de API y excipientes #1 y #2 en las caras interiores (E y F) de los núcleos muestran desviaciones respecto a los valores de concentración nominales y del resto de las caras. Este comportamiento se repite en los distintos núcleos analizados y puede ser explicado por una mayor rugosidad en la superficie resultante del corte lo cual puede tener cierta influencia en la reflectancia, durante el proceso de adquisición de las imágenes y, por consiguiente, en las concentraciones calculadas.

En la Figura 6.5 se pueden observar los mapas de concentración de los distintos componentes obtenidos con los mismos modelos Isys-PLS1 para una de las caras exteriores de un núcleo. Como se puede ver, la distribución del excipiente #2 y del API son muy similares, con áreas de mayor concentración localizadas en las mismas zonas de la superficie del comprimido. Por el contrario, las áreas de mayor concentración del excipiente #1 son complementarias a las del API. Este patrón de distribución se puede atribuir al hecho que el API y el excipiente #2 son mezclados en una primera etapa, antes de añadir el resto de excipientes. Por lo que respecta a los excipientes minoritarios, el reparto es uniforme en toda la imagen. La distribución de API y excipientes observada para el resto de caras del mismo núcleo y otros núcleos analizados no presenta diferencias significativas respecto a la cara descrita anteriormente y sigue las mismas tendencias que ésta.

En la misma Figura 6.5 se muestran también los histogramas para cada uno de los componentes del núcleo. Para los excipientes minoritarios, los histogramas presentan una estrecha distribución de los valores de concentración y una forma simétrica, cercana a la normalidad. En cambio, los histogramas para los excipientes mayoritarios presentan bases más anchas, hecho indicativo de la presencia de zonas del comprimido con valores de concentración más alejadas del valor nominal. En ningún caso, estas distribuciones pueden considerarse como normales: los histogramas del API y del excipiente #2 presentan una forma muy similar (debido a su mezclado

previo), con colas a la izquierda, mientras que las colas observadas en el histograma del excipiente #1 son a la derecha.

▪ **Comprimidos lacados (0-3x)**

La determinación de API y excipientes también se realizó en los comprimidos con distintos grosores de lacado. Para ello, se construyeron modelos de calibración Isys-PLS1, que se aplicaron a dos caras, A y B, de un mismo comprimido, para cada nivel de recubrimiento. Los resultados correspondientes a la media de los valores obtenidos para los dos comprimidos de una misma imagen (comprimidos con igual porcentaje de recubrimiento) se muestran en la Tabla 6.4. No se incluyen los resultados de los componentes minoritarios porque presentan desviaciones importantes respecto a los valores de referencia (30-50% de los valores nominales). Para los tres componentes mayoritarios, los modelos utilizados presentan un número de factores que oscila entre 3 y 6 dependiendo del analito a modelar y del grosor de recubrimiento de la muestra. Las concentraciones predichas son muy próximas a los valores nominales (Tabla 6.1) y, siguiendo la misma tendencia que estos, disminuyen gradualmente con el aumento del grosor del lacado. Es evidente que las desviaciones respecto los valores de referencia aumentan notablemente en los comprimidos con mayor grosor de recubrimiento (2x y 3x) y, sobretodo, en el caso del excipiente #2, el menos abundante de los tres.

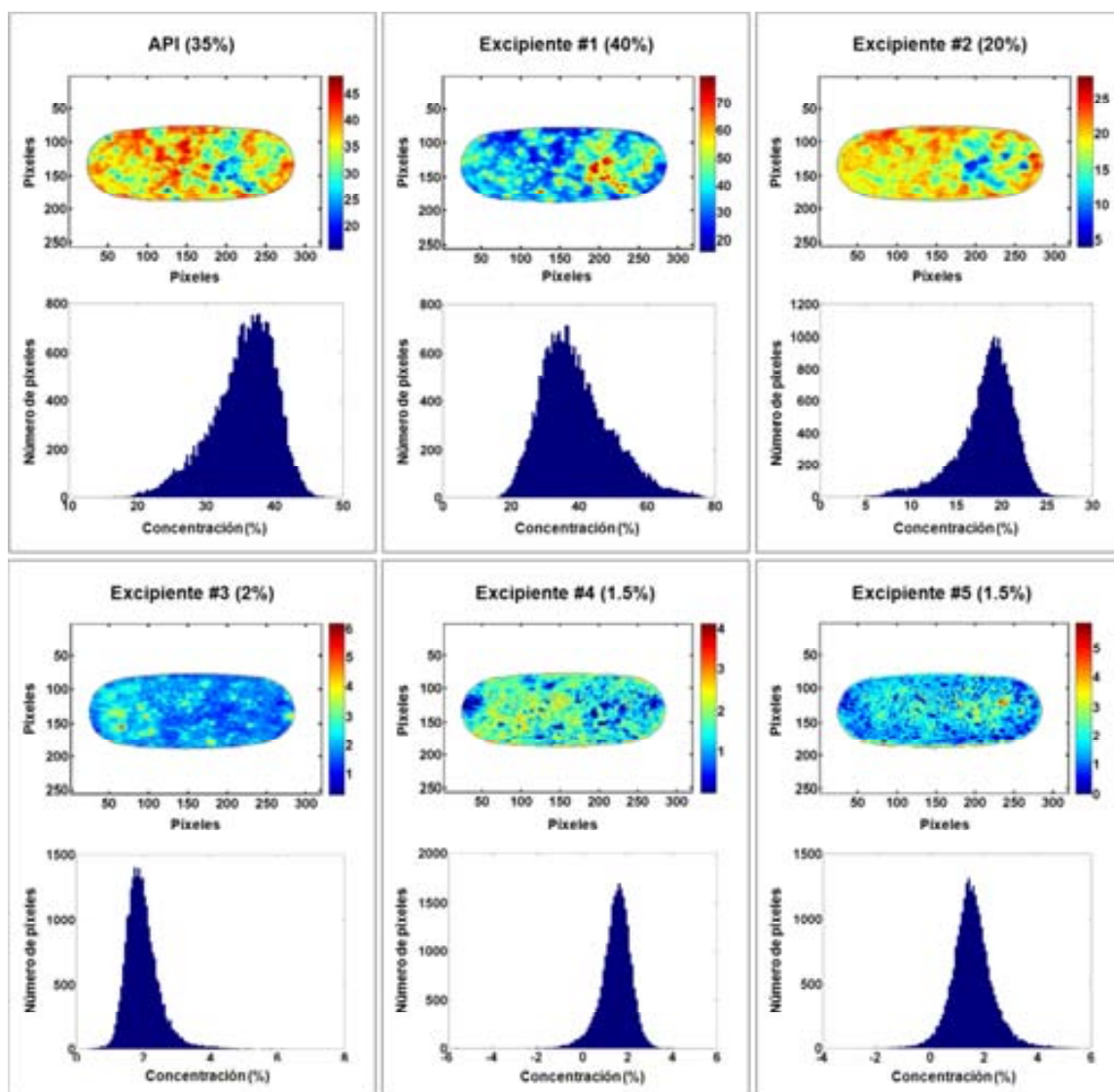


Figura 6.5. Mapas de concentración e histogramas de API y excipientes calculados mediante Isy-PLS1 para una de las caras exteriores de un núcleo. Entre paréntesis se muestran las concentraciones nominales (% m/m) para cada componente.

De estos resultados se puede concluir que las desviaciones observadas en los cálculos de concentración son debidas principalmente al recubrimiento, que afecta de manera importante a la determinación de la concentración de los componentes del comprimido, en especial, a la de los componentes minoritarios. Este hecho se puede relacionar con la absorción de la radiación incidente por parte de la laca del recubrimiento, lo que reduce la contribución de los componentes del núcleo a la señal espectroscópica.

Tabla 6.4. Concentraciones (% m/m) de API y los dos excipientes mayoritarios en comprimidos con distinto grosor de recubrimiento.

Nivel de recubrimiento	Cara	Composición (% m/m)		
		API	E #1	E #2
0x	A	35.25	40.74	20.20
	B	35.67	39.46	21.34
0.5x	A	35.85	40.30	19.67
	B	35.31	40.27	20.36
1x	A	34.56	40.80	20.79
	B	35.20	40.18	21.64
2x	A	33.15	39.15	17.95
	B	32.95	42.40	15.31
3x	A	30.85	38.51	12.65
	B	29.65	39.78	14.03

6.3.2. Cuantificación del recubrimiento

Determinar el espesor de la capa de laca de recubrimiento es un parámetro importante ya que puede afectar a la velocidad de disolución del comprimido en cuestión y también a la calidad de los resultados de cuantificación de los diferentes componentes. En este trabajo, se determinó el grosor del recubrimiento en muestras de producción usando un modelo PLS construido con los espectros promedio de cada una de las caras de los comprimidos frente al grosor de recubrimiento. En la Figura 6.6 se muestran los espectros NIR en SNV para cada nivel de recubrimiento (de 0x a 3x). A lo largo del espectro se pueden observar varias bandas que presentan diferencias en función del nivel de recubrimiento. Destaca especialmente el incremento de intensidad de la banda entre 1900 y 2000 nm al aumentar el grosor de recubrimiento.

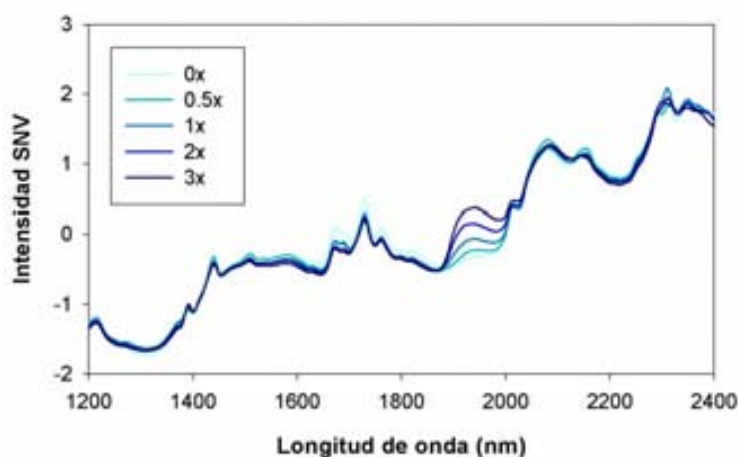


Figura 6.6. Espectros NIR en SNV de comprimidos con distintos niveles de recubrimiento. Los espectros han sido obtenidos promediando el espectro de cada píxel de un área de la imagen hiperespectral.

Para la construcción del modelo, se utilizaron 10 imágenes de comprimidos lacados, dos para cada nivel de recubrimiento pertenecientes a las dos caras exteriores opuestas de los mismos comprimidos. Además, cada imagen incluye dos comprimidos de iguales características, tal y como se describe en la Figura 6.2b. Se promediaron los espectros del área seleccionada en cada imagen, dando lugar a un total de 4 espectros por nivel de recubrimiento (2 comprimidos por imagen y 2 caras por comprimido). De los 20 espectros finales, 3 se descartaron como *outliers*, 12 se usaron para construir el modelo de calibración y 5, para validarlo. A cada espectro promedio del conjunto de calibración se le asignó un grosor de recubrimiento: 0x, 0.5x, 1x, 2x o 3x. Con el fin de conseguir el modelo que mejor se ajusta al conjunto de validación, se ensayaron distintos pretratamientos (derivadas de Savitzky-Golay y SNV) y rangos espectrales. Los mejores resultados se obtuvieron pretratando los datos con SNV y trabajando con el rango espectral completo. La selección de un rango espectral que coincidiera con la principal banda de absorción de la laca de recubrimiento (1900-2000 nm) no mejoraba la capacidad predictiva del modelo. El modelo PLS conseguido con el rango espectral entero solamente necesita un factor para explicar el 99.3% de la varianza. Las características de este modelo se describen con más profundidad en la Tabla 6.5.

Tabla 6.5. Características del modelo de calibración PLS usado para la determinación del grosor de recubrimiento.

Características del modelo de calibración PLS	
Nº muestras de calibración	12
Rango espectral (nm)	1200-2400
Pretratamiento espectral	SNV
Nº factores PLS	1
Y-Var explicada (%)	99.31
Nº muestras de validación	5
RMSEP	14.04
RSEP (%)	8.32

La aplicación del modelo de calibración a una imagen NIR-CI permite cuantificar el grosor promedio del recubrimiento y visualizar la distribución de la laca en la superficie del comprimido. En la Figura 6.7 se muestran los mapas de concentración de recubrimiento para las imágenes de las cuatro caras exteriores de un mismo comprimido de producción (Figura 6.2a) a las que se les ha aplicado el modelo PLS descrito. De la comparación de los cuatro mapas de concentración se puede deducir que el material de recubrimiento no se distribuye de forma homogénea en todas las caras, ya que se observan puntos dónde la concentración predicha es superior al resto de la superficie. Este efecto puede ser debido a la presencia de agregados de laca producidos en el mismo proceso de recubrimiento, lo que indicaría un esprayado irregular dentro del tambor de recubrimiento, o a colisiones entre comprimidos durante el mismo proceso de recubrimiento. En la misma Figura 6.7 se muestran también los histogramas de los mapas de concentración para facilitar el análisis de la distribución de la capa de lacado.

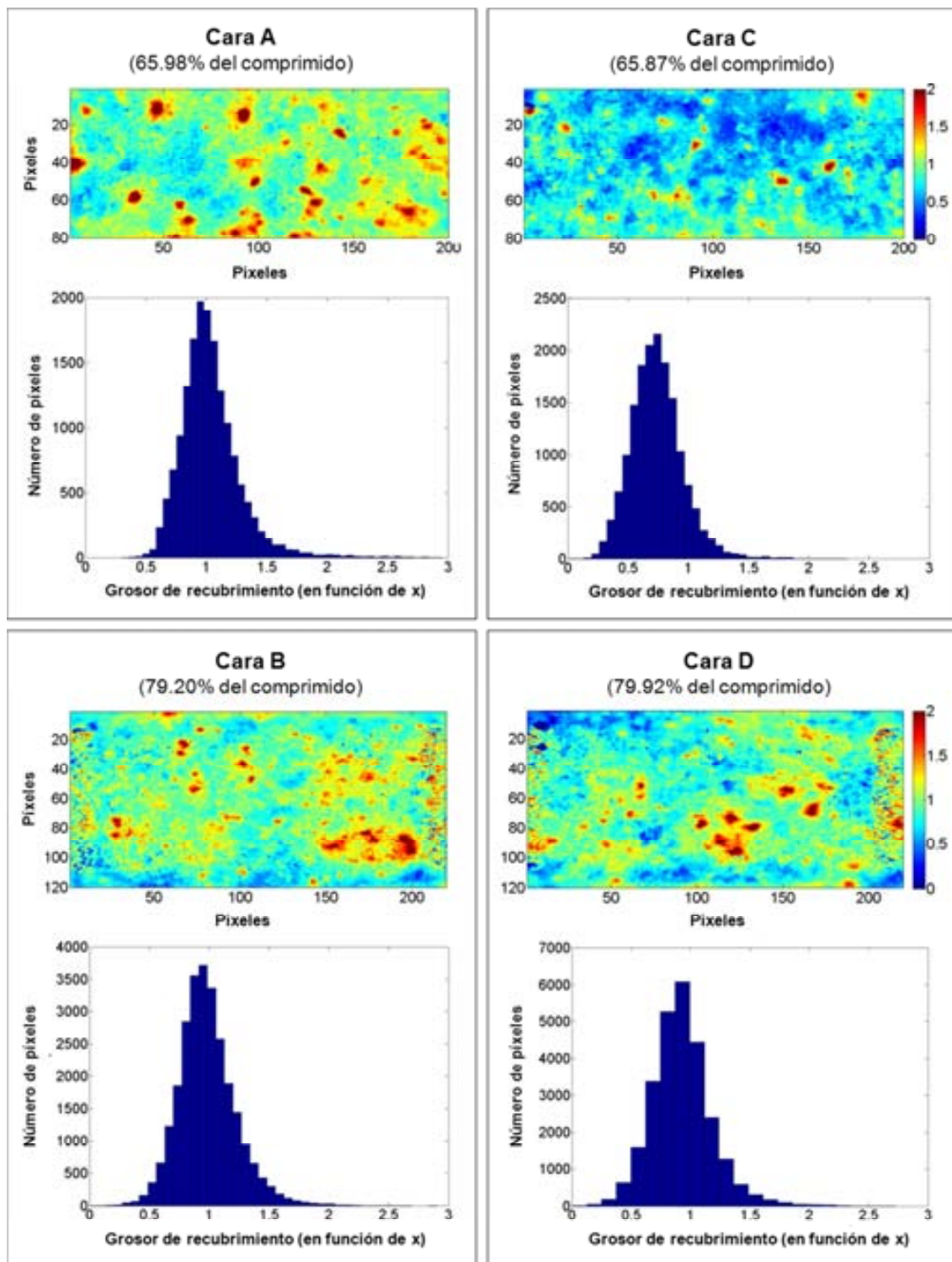


Figura 6.7. Mapas de distribución de recubrimiento e histogramas para las cuatro caras de un comprimido de producción calculado mediante el modelo PLS. Entre paréntesis se indica el porcentaje de imagen que representa el área mostrada, con respecto a la imagen del comprimido entero.

En la Tabla 6.6 se muestran los valores medios de recubrimiento para cada una de las caras exteriores de dos comprimidos de producción, obtenidos mediante el modelo PLS descrito anteriormente. Ambos comprimidos presentan diferencias respecto el valor nominal de hasta el 10% ($0.1x$). No obstante, el valor promedio de las 4 caras del comprimido, da resultados muy cercanos a los nominales para todas las muestras estudiadas.

Tabla 6.6. Grosos de recubrimiento (en función de x) en dos comprimidos de producción, obtenidos aplicando el modelo PLS descrito en la Tabla 6.5.

Grosor de recubrimiento (en función de x)		
	Comprimido 1	Comprimido 2
Cara A	1.035	1.122
Cara B	0.984	1.177
Cara C	0.739	1.035
Cara D	0.925	1.223
Promedio (4 caras)	0.921	1.139

6.4. Conclusiones

En este trabajo, se ha evaluado la capacidad de la técnica NIR-CI para la determinación cuantitativa de la composición de distintos comprimidos farmacéuticos, recubiertos y sin recubrir, y el análisis de la distribución de los distintos componentes. También se ha demostrado la capacidad de los algoritmos PLS e Isys-PLS1 para obtener información cualitativa y cuantitativa de calidad a partir de imágenes hiperespectrales sin necesidad de una calibración con valores de referencia.

Un tratamiento convencional de los datos espectrales de la imagen permite determinar el grosor de la capa de recubrimiento de los comprimidos y también su distribución en las diferentes caras del comprimido, por lo que contribuye en la predicción de ciertas propiedades del comprimido relacionadas con el recubrimiento.

La rapidez de los análisis realizados y la fiabilidad en la predicción de resultados, hacen del NIR-CI una técnica muy adecuada para la determinación de la composición cuantitativa de un comprimido farmacéutico, así como para evaluar la homogeneidad de distribución de los componentes y el recubrimiento superficial de los comprimidos. Estas características demuestran el gran potencial de esta técnica para su aplicación en ensayos *at-line* para el control de procesos de producción en la industria farmacéutica.

REFERENCIAS

- 1 Armstrong, N.A. (2007) Tablet manufacture. En: Swarbrick, J. (ed.) *Encyclopedia of pharmaceutical technology*. Vol. 1., 3^a ed.; Informa Healthcare USA, Inc. New York, NY, pp. 3653-3672.
- 2 Maurer, L., Leuenberger, H. (2009) Terahertz pulsed imaging and near infrared imaging to monitor the coating process of pharmaceutical tablets. *Int. J. Pharmaceut.* **370**, 8-16.
- 3 Ma, H., Anderson, C. A. (2008) Characterization of pharmaceutical powder blends by NIR chemical imaging. *J. Pharm. Sci.* **97**, 3305-3320.
- 4 Li, W., Woldu, A., Kelly, R., McCool, J., Bruce, R., Rasmussen, H., Cunningham, J., Winstead, D. (2008) Measurement of drug agglomerates in powder blending simulation samples by near infrared chemical imaging. *Int. J. Pharmaceut.* **350**, 369-373.
- 5 Lopes, M.B., Wolff, J.C., Bioucas-Dias, J.M., Figueiredo, M.A. (2009) Determination of the composition of counterfeit Heptodin tablets by near infrared chemical imaging and classical least squares estimation. *Anal. Chim. Acta* **641**, 46-51.
- 6 Cruz, J., Bautista, M., Amigo, J. M., Blanco, M. (2009) Nir-chemical imaging study of acetylsalicylic acid in commercial tablets. *Talanta* **80**, 473-478.
- 7 Gendrin, C., Roggo, Y., Collet. C. (2007) Content uniformity of pharmaceutical solid dosage forms by near infrared hyperspectral imaging: A feasibility study. *Talanta* **73**, 733-741.
- 8 Pomerantsev, A.L., Rodionova, O.Y. (2012) Process analytical technology: A critical view of the chemometricians. *J. Chemometr.* **26**, 299-310.
- 9 Cairós, C., Amigo, J.M., Watt, R., Coello, J., MasPOCH, S. (2008) Implementation of enhanced correlation maps in near infrared chemical images: Application in pharmaceutical research. *Talanta* **79**, 657-664.
- 10 Puchert, T., Lochmann, D., Menezes, J.C., Reich, G. (2010) Near-infrared chemical imaging (NIR-CI) for counterfeit drug identification—A four-stage concept with a novel approach of data processing. *J. Pharm. Biomed. Anal.* **51**, 138-145.
- 11 Ravn, C., Skibsted, E., Bro, R. (2008) Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—Comparing common calibration approaches. *J. Pharm. Biomed. Anal.* **48**, 554-561.

7. DETECCIÓN DE CAMBIOS INDUCIDOS POR LA HUMEDAD EN FORMULACIONES LIOFILIZADAS MEDIANTE NIR-CI

7.1. Introducción	257
7.2. Experimental	259
7.2.1. Composición de las muestras.....	259
7.2.2. Freeze-drying.....	259
7.2.3. NIR-CI	261
7.2.4. Valoraciones Karl Fischer (KF)	262
7.2.5. Difracción de rayos X en polvo (XRPD).....	262
7.2.6. Análisis de datos	262
7.2.7. Determinación de la humedad.....	263
7.2.8. Análisis de la sacarosa cristalizada	264
7.3. Resultados y discusión	265
7.3.1. Visualización del contenido de agua después del freeze-drying.....	265
7.3.2. Absorción de agua en distintas condiciones de almacenamiento	266
7.3.3. Cristalización de la sacarosa.....	268
7.4. Conclusiones	272
REFERENCIAS	273

7. DETECCIÓN DE CAMBIOS INDUCIDOS POR LA HUMEDAD EN FORMULACIONES LIOFILIZADAS MEDIANTE NIR-CI

7.1. Introducción

El creciente número de productos biofarmacéuticos ha favorecido la implantación del *freeze-drying* como método para prolongar su preservación y como operación unitaria cada vez más común en la industria farmacéutica [1]. Es por ello, que ha aumentado también el interés en la optimización de los procesos de liofilización, con el objetivo de conseguir una mayor calidad de los fármacos finales con el menor coste posible. Así, cuando se formula un nuevo producto basado en proteínas hay que evaluar cuidadosamente un gran número de atributos para asegurar su calidad [2,3]. Uno de los parámetros más importantes es el contenido de agua residual en el producto liofilizado, ya que afecta directamente la estabilidad de la proteína. El contenido de agua óptimo no es siempre el más bajo posible [4], sino que viene definido por la capa de hidratación de la proteína en cuestión y los excipientes estabilizantes, que substituyen los enlaces de la proteína con el agua [5,6]. Estos excipientes también pueden verse afectados por la captación de agua en muestras liofilizadas: la humedad promueve la cristalización de los estabilizantes amorfos como la sacarosa y la trehalosa [6,7], lo que tiene un efecto negativo en la estabilidad de las proteínas, ya que implica la pérdida de interacciones estabilizante-proteína y la liberación de agua.

La humedad en los productos liofilizados se puede controlar en gran medida optimizando el proceso. La elección de los excipientes y sus concentraciones (formulación) son determinantes de la humedad residual de los liofilizados, pero también afectan a su higroscopicidad [8], una propiedad muy crítica durante el almacenamiento del producto. De hecho, a pesar de que los viales se cierran al vacío o después de ser purgados con nitrógeno, se han reportado cambios en el contenido de humedad residual debido a la liberación de agua por el tapón o por posibles fugas en los viales [9,10,11]. Es por este motivo que entender las causas de la humedad residual

en la muestra y su comportamiento de absorción/desorción del agua es crucial para el desarrollo de nuevas formulaciones de productos liofilizados.

El método estándar para la determinación de humedad es la valoración por KF. Aunque se trata de un método adecuado para el análisis de muestras liofilizadas, es lento, consume reactivos y genera residuos. Además el hecho de que sea un método destructivo, impide el seguimiento de una misma muestra a lo largo del tiempo, especialmente si los volúmenes de muestra de los que se dispone son pequeños. Un método alternativo para la determinación de agua es el NIR [12], las ventajas del cual y su aplicabilidad como técnica PAT ya han sido descritas en capítulos anteriores. En el campo del *freeze-drying*, el NIR ha sido ampliamente usado para la cuantificación de la humedad [12,13,14,15,16], así como para detectar distintas formas polimórficas de los sólidos [17,18] y detectar cambios estructurales en las proteínas [19,20]. Por lo que respecta al NIR-CI, también es comúnmente utilizado como herramienta analítica en la industria farmacéutica, pero sus aplicaciones a productos liofilizados son, hasta el momento, escasas [21,22].

El objetivo de este estudio es el de evaluar el uso del NIR-CI como herramienta para el cribado de formulaciones idóneas para liofilizar. Este cribado se basa en el seguimiento de la absorción/desorción del agua y su distribución, así como la detección de cambios de estado sólido relacionados con la humedad en distintas formulaciones liofilizadas. Para ello, se prepararon las distintas formulaciones modelo a evaluar y se liofilizaron en placas de pocillos. Las placas se almacenaron durante 10 días a distintos niveles de humedad relativa (RH, *Relative Humidity*) y temperatura ambiente y se registraron diariamente para detectar los cambios inducidos por la humedad como efecto de la composición de la formulación. Para la rápida visualización de la distribución de agua en las imágenes de las muestras liofilizadas, se construyó un modelo univariante basado en los valores de referencia obtenidos por KF.

7.2. Experimental

7.2.1. Composición de las muestras

Las muestras utilizadas son mezclas de manitol, sacarosa, lisozima y albúmina de suero bovino (BSA, *Bovine Serum Albumin*), todo de Sigma-Aldrich (St. Louis, MO). La liofilización por *freeze-drying* se llevó a cabo en placas metálicas de 96 pocillos (Figura 7.1), de 3 mm de fondo y 7 mm de diámetro cada uno. Cada pocillo se llenó con 100 μl de la muestra correspondiente. Para poder rellenar los pocillos fue necesario sellar previamente uno de los lados de las placas, lo que se consiguió aplicando film transparente multicapa de R&G Faserverbundwerkstoffe (Waldenbuch, Alemania) con el espray adhesivo Spray Mount™ de 3M (St. Paul, MN).

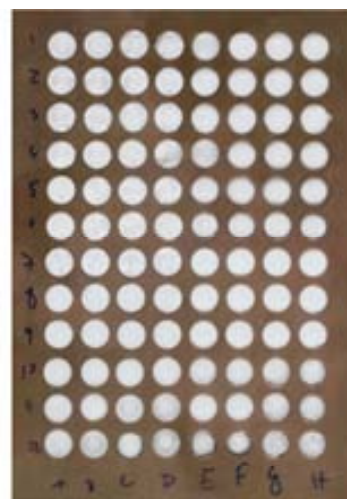


Figura 7.1. Imagen de una de las placas de 96 pocillos utilizadas en los experimentos.

En total se prepararon cuatro placas, dos correspondientes a la formulación mostrada en la Tabla 7.1 (placas 1 y 3) y dos correspondientes a la formulación de la Tabla 7.2 (placas 2 y 4). Las formulaciones fueron diseñadas con el objetivo de determinar el efecto del contenido en sólido, azúcar amorfo y dos proteínas distintas, así como el contenido y la distribución de la humedad.

7.2.2. Freeze-drying

El proceso de liofilización se llevó a cabo por *freeze-drying*, mediante el liofilizador Lyostar II de SP Scientific (Stone Ridge, NY). El ciclo de *freeze-drying* aplicado empieza con un proceso de enfriamiento de la muestra a una velocidad de $1\text{ }^{\circ}\text{C}\cdot\text{min}^{-1}$ hasta $-50\text{ }^{\circ}\text{C}$, seguido por una fase de 15 min de mantenimiento de la temperatura ($-50\text{ }^{\circ}\text{C}$). Posteriormente, se aumenta de nuevo la temperatura $1\text{ }^{\circ}\text{C}\cdot\text{min}^{-1}$ hasta llegar a los $-10\text{ }^{\circ}\text{C}$, temperatura a la que se lleva a cabo el *annealing*, que se prolonga 2 h. Seguidamente, se aplica un nuevo proceso de enfriamiento hasta $-50\text{ }^{\circ}\text{C}$, la etapa de

secado primario a $-20\text{ }^{\circ}\text{C}$ y 0.1 mbar durante 8 h y, finalmente, el secado secundario a $20\text{ }^{\circ}\text{C}$ y a la misma presión durante 3 h. Las condiciones de esta última etapa (una temperatura más baja de lo habitual durante un tiempo más corto) se escogieron deliberadamente con el fin de resaltar las posibles variaciones de humedad causadas por las diferencias en la formulación de las distintas muestras.

Tabla 7.1. Formulaciones para las placas 1 y 3.

Fila / columna	Parámetro variable	A	B	C	D	E	F	G	H
1-3 (5% sólido)	Sacarosa (%)	0	5	10	15	20	30	40	50
4-6 (5% sólido, 5 % sac)	Lisozima ($\text{mg}\cdot\text{ml}^{-1}$)	0	1	2.5	5	10	20	30	40
7-9 (5% sólido, 5 % sac)	BSA ($\text{mg}\cdot\text{ml}^{-1}$)	0	1	2.5	5	10	20	30	40
10-12 (100% man)	Sólido (%)	2	4	6	8	10	12	14	16

Los porcentajes de sacarosa y manitol (% m/m) corresponden a la cantidad de sacarosa o manitol respecto la suma de ambos.

El porcentaje de sólido (% m/v) se refiere a la masa de manitol y sacarosa en el volumen total de la muestra.

Tabla 7.2. Formulaciones para las placas 2 y 4.

Fila / columna	Parámetro variable	A	B	C	D	E	F	G	H
1-3 (5 % sólido, 30 % sac)	BSA ($\text{mg}\cdot\text{ml}^{-1}$)	0	1	2.5	5	10	20	30	40
4-6 (30 % sac)	Sólido (%)	2	4	6	8	10	12	14	16
7-9 (10 % sólido)	Sacarosa (%)	10	20	30	40	50	60	70	80
10-12 (5 % sólido)	Sacarosa (%)	10	20	30	40	50	60	70	80

Los porcentajes de sacarosa y manitol (% m/m) corresponden a la cantidad de sacarosa o manitol respecto la suma de ambos.

El porcentaje de sólido (% m/v) se refiere a la masa de manitol y sacarosa en el volumen total de la muestra.

7.2.3. NIR-CI

Las imágenes espectrales de las muestras liofilizadas se registraron en modo reflectancia difusa. Para ello se utilizó la cámara hiperespectral SWIR de Specim Ltd. (Oulu, Finland) mostrada en la Figura 7.2, equipada con un detector MCT enfriado a $-70\text{ }^{\circ}\text{C}$ gracias a un sistema Peltier de cuatro plataformas. Este equipo funciona por el método de barrido por líneas, de modo que permite registrar una fila de 320 píxeles de una sola vez. Para cada píxel, el rango espectral cubierto abarca des de 970 hasta 2500 nm, a una resolución espectral de $\sim 10\text{ nm}$.

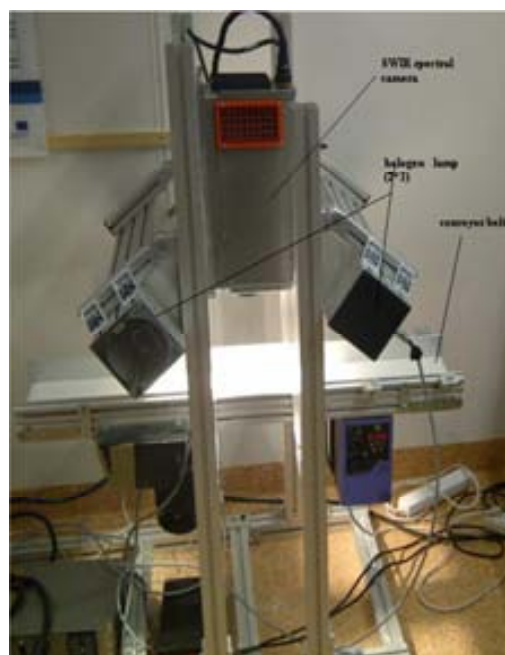


Figura 7.2. Cámara hiperespectral SWIR (Specim Ltd.).

Las placas de pocillos con las muestras se colocaron en una cinta transportadora Mini de ENP Industrisystem AB (Hjälteby, Sweden) que se movía a una velocidad constante de $3.12\text{ cm}\cdot\text{s}^{-1}$. La cámara se situó justo encima de la cinta a la distancia necesaria para que cada línea cubriera una longitud de 10cm en la superficie de la muestra. Las muestras en la cinta se iluminaron mediante dos filas de 3 lámparas halógenas de Specim Ltd. de 75 W dispuestas en una geometría $45^{\circ}\text{-}0^{\circ}\text{-}45^{\circ}$, con el fin de minimizar la detección de la reflectancia especular y maximizar la contribución de la reflectancia difusa. La resolución espacial de la cámara en las condiciones descritas y trabajando a la máxima frecuencia de captura ($100\text{ capturas}\cdot\text{s}^{-1}$) es de $312 \times 312\text{ }\mu\text{m}$ por

píxel. Para cada imagen hiperespectral se tomaron 500 capturas, con lo que el tamaño final de la imagen es de 500 x 320 píxeles y el tiempo de adquisición, de unos 5 s. El área de muestra cubierta en cada imagen es de 10 x 16cm, lo que asegura que la imagen comprende la placa de pocillos entera.

7.2.4. Valoraciones Karl Fischer (KF)

Los valores de referencia para el contenido de agua en muestras liofilizadas se obtuvieron mediante valoración KF. Para ello se utilizó un valorador Coulómetro C30 KF Compact de Mettler Toledo (Greifensee, Switzerland). Las muestras se pesaron justo antes de la valoración, asegurando una exposición mínima al aire ambiental.

7.2.5. Difracción de rayos X en polvo (XRPD)

Las medidas de difracción de rayos X en polvo (XRPD, *X-ray powder diffractometry*) se llevaron a cabo mediante un difractómetro X'Pert PRO de PANalytical (Almelo, The Netherlands), equipado con un detector PIXcel. Los difractogramas se obtuvieron usando una fuente de radiación Cu K α (45 kV, 40 mA, $\lambda = 1.54187 \text{ \AA}$). Las muestras se midieron de forma automática gracias a la aplicación de un lector de pocillos de la placa. Los difractogramas fueron adquiridos en modo de transmisión des de 9 hasta 26 $^{\circ}2\theta$ con un resolución de 0.026 $^{\circ}2\theta$ y una velocidad de escaneo de 0.027 $^{\circ}2\theta/s$.

7.2.6. Análisis de datos

Previamente a la aplicación del pretratamiento y el modelado de los datos, las imágenes hiperespectrales fueron preprocesadas. En primer lugar, se les aplicó una máscara binaria en forma de placas de 96 pocillos diseñada especialmente para analizar únicamente la información de las muestras en el interior de los pocillos, obviando el resto de la placa. Posteriormente, las imágenes se desdoblaron para pasar del cubo de

datos tridimensional a una matriz bidimensional, facilitando así el trabajo posterior. Después del modelado, los datos volvieron a reorganizarse de la misma forma que en la imagen original con el fin de poder visualizar claramente los resultados en forma de mapa de concentraciones/ correlaciones de la muestra. Todas las transformaciones descritas, así como la aplicación de los pretratamientos espectrales se llevaron a cabo mediante rutinas desarrolladas en MATLAB v 7.11.0 The MathWorks (Natick, MA).

7.2.7. Determinación de la humedad

Con el fin de determinar de forma cualitativa la distribución de la humedad en las muestras liofilizadas se creó un modelo univariante. El modelo se basa en la banda del agua del espectro NIR, localizada alrededor de los 1940 nm (Figura 7.3). El valor de la absorbancia en el punto máximo de este pico fue corregido con el valor de la señal en el mínimo a 1875 nm para evitar contribuciones debidas a diferencias de línea base. La elección del mínimo se hizo con el objetivo de minimizar la contribución espectral del resto de componentes, los espectros NIR de los cuáles se muestran en la.

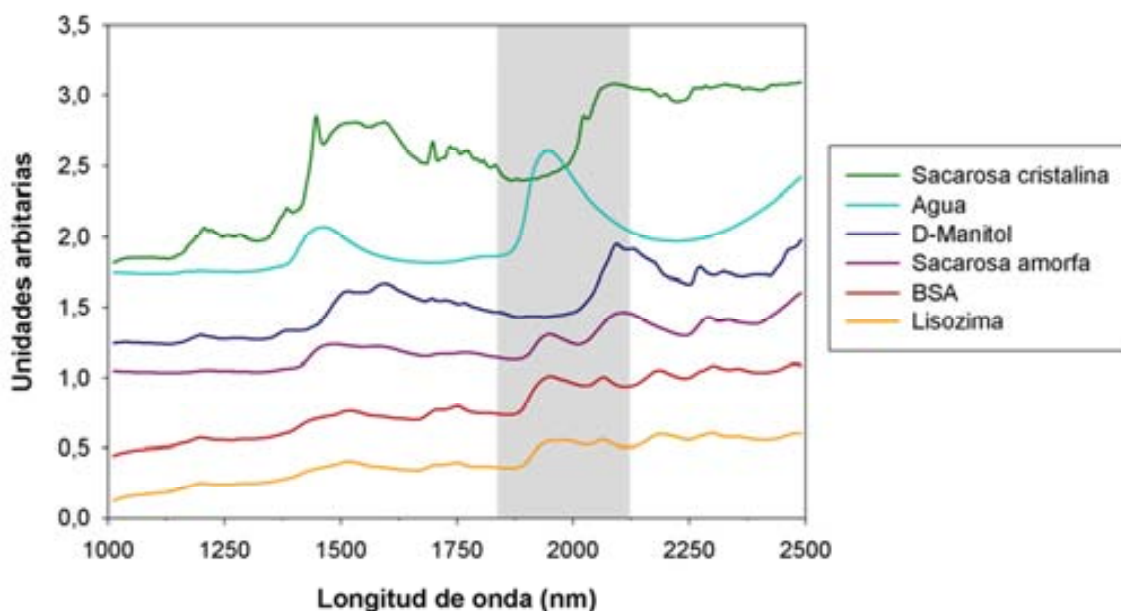


Figura 7.3. Espectros NIR de todos los componentes puros incluidos en las muestras (sacarosa cristalina y amorfa, agua, manitol, BSA y lisozima). Los espectros han sido desplazados de forma arbitraria para facilitar su visualización. La zona sombreada corresponde a la banda del agua a 1940 nm utilizada en la construcción del modelo de calibración.

Las imágenes de las placas de pocillos con muestras utilizadas para la calibración y la validación del modelo, fueron adquiridas justo antes de ser sometidas a la valoración KF. Se minimizan así las diferencias de humedad en la muestra entre el momento del registro por NIR-CI y el de la obtención del valor de referencia por KF. Los espectros de estas muestras se consiguieron promediando los espectros de todos los píxeles de la imagen que conforman cada muestra, *i. e.* todos los píxeles de un pocillo. Todos los espectros fueron pretratados mediante el algoritmo SNV con el objetivo de minimizar las contribuciones del *scattering*. Para la construcción del modelo se usaron un total de 62 muestras de calibración, con valores de referencia comprendidos entre 0.72 y 5.60% de humedad. El RMSEP del modelo fue de 0.76, un valor muy parecido al obtenido para la calibración (RMSEC = 0.68), indicativo de la buena capacidad predictiva del modelo.

7.2.8. *Análisis de la sacarosa cristalizada*

Con el propósito de visualizar la cristalización de la sacarosa en las muestras, se calculó el CC entre el espectro puro de la sacarosa cristalina y el espectro de la muestra en cada píxel del pocillo. Los resultados de este cálculo se usaron como una medida de la similitud espectral entre la muestra y la sacarosa cristalina.

Como se muestra en la Figura 7.3, el espectro de la sacarosa cristalina pura está altamente correlacionado con la mayoría de analitos de las muestras. Su correlación con el manitol es de 0.967, mientras que para el resto, exceptuando el agua, las correlaciones superan en todos los casos el 0.900. Para evitar la contribución del resto de componentes de la muestra en el cálculo del CC de la sacarosa cristalina, se probaron distintos rangos espectrales y pretratamientos (SNV y derivadas de Savitzky-Golay). Los mejores resultados se obtuvieron con el rango 2140-2500 nm de los espectros pretratados en 1ª derivada, con 3 puntos de ventana móvil y polinomio de segundo grado.

7.3. Resultados y discusión

7.3.1. Visualización del contenido de agua después del freeze-drying

Con el fin de investigar el potencial del NIR-CI para la rápida determinación de humedad residual y distribución del agua después de la liofilización, las placas de pocillos fueron registradas inmediatamente después de ser sometidas al proceso de *freeze-drying*. Los mapas de distribución de la humedad calculados a partir de los cubos de datos resultantes se muestran en la Figura 7.4. Como se advierte en la figura, el contenido en agua de las muestras varía entre el 2 y el 6%. Estos valores relativamente elevados de humedad son debidos a la aplicación de una etapa corta de secado secundario a solamente 20 °C (descrita en el apartado 7.2.2).

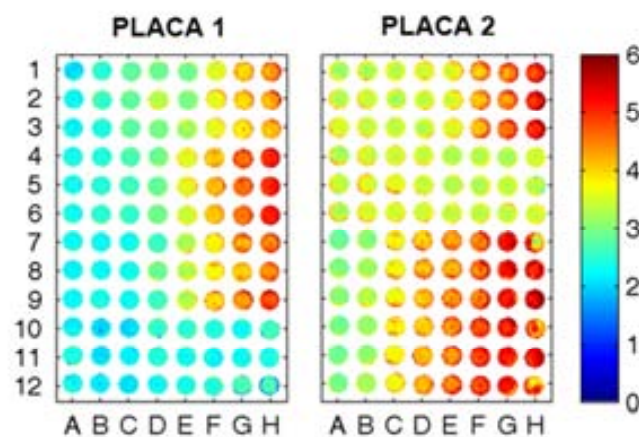


Figura 7.4. Mapas de distribución de la humedad en las placas de pocillos: placa 1 (izquierda) y placa 2 (derecha) adquiridas inmediatamente después de la liofilización. La escala de color representa el contenido de agua en porcentaje.

En la misma imagen se observa que las filas 10-12 en la placa 1 y 4-6 en la placa 2, con contenidos del 2% (izquierda) al 16% (derecha) de sólido en ambos casos, muestran cambios mínimos o inexistentes en el contenido de agua. Este comportamiento pone de manifiesto que el contenido de sólido, independientemente del contenido de sacarosa amorfa, no afecta a la humedad residual de las muestras después del liofilizado. Contrariamente, cuando se aumenta el contenido de componente amorfo en la matriz se observa un incremento de humedad en las muestras después de la liofilización. Este efecto sucede tanto para la sacarosa (filas 7-12

de la placa 2), como para la proteína (filas 4-9 de la placa 1) en la Figura 7.4. En este último caso no se advierten grandes diferencias en la retención de humedad entre muestras con BSA y lisozima, aunque parece que ésta es ligeramente inferior en las muestras con BSA.

Por lo que respecta a la distribución de agua, es en general uniforme en los pocillos de las placas de la Figura 7.4, aunque hay algunas excepciones. Los valores de humedad son ligeramente superiores en los bordes de los pocillos, lo que probablemente sea debido a un artefacto experimental causado por un efecto de las sombras. Los pocillos de la columna H en las filas 7-12 de la placa 2 presentan una distribución claramente desigual de la humedad. Estas muestras contienen en todos los casos un alto porcentaje de sacarosa (80%), lo que puede llegar a colapsar las muestras, dando lugar a una distribución heterogénea en estos pocillos.

7.3.2. Absorción de agua en distintas condiciones de almacenamiento

Para estudiar la absorción de humedad en las distintas formulaciones preparadas, las placas de pocillos se almacenaron a dos niveles de RH y fueron registradas diariamente para monitorizar los cambios. En la Figura 7.5 se representa la evolución del contenido de agua en dos placas con la misma formulación, pero distintas condiciones de almacenamiento: la placa 1 almacenada al 11% de RH y la placa 3, al 43%. Se observa que el contenido de agua en formulaciones que sólo contienen manitol (filas 10-12) disminuye durante el almacenamiento a ambos niveles de RH, un efecto que se advierte des del segundo día. Una posible explicación para esta disminución en el contenido de agua, incluso cuando se almacena al 43% de RH, es que una pequeña parte del manitol permanece en forma amorfa después de la liofilización y la cristalización de este manitol amorfo puede liberar el agua ligada a la matriz amorfa. Para probar esta hipótesis, se evaluaron los difractogramas XRPD de las muestras antes y después del almacenamiento. No se encontró ningún cambio de intensidad de reflexión significativo referente al manitol, lo que indica que este efecto se debe simplemente a la baja capacidad de absorción de agua de las formulaciones puras con manitol.

En la misma Figura 7.5 se observa que en los pocillos con concentraciones del 5-10% de sacarosa (filas 1-3, columnas B y C) la retención de humedad es muy parecida entre las placas reservadas a alta y baja RH. Sin embargo, para muestras con más del 30% de sacarosa (filas 1-3, columnas F-H) la retención de humedad es claramente mayor en la placa almacenada al 43% de RH (placa 3) ya des del día 2. Al aumentar la concentración de sacarosa de la formulación, aumenta también su higroscopicidad, con lo que la retención de humedad depende del nivel de humedad del ambiente.

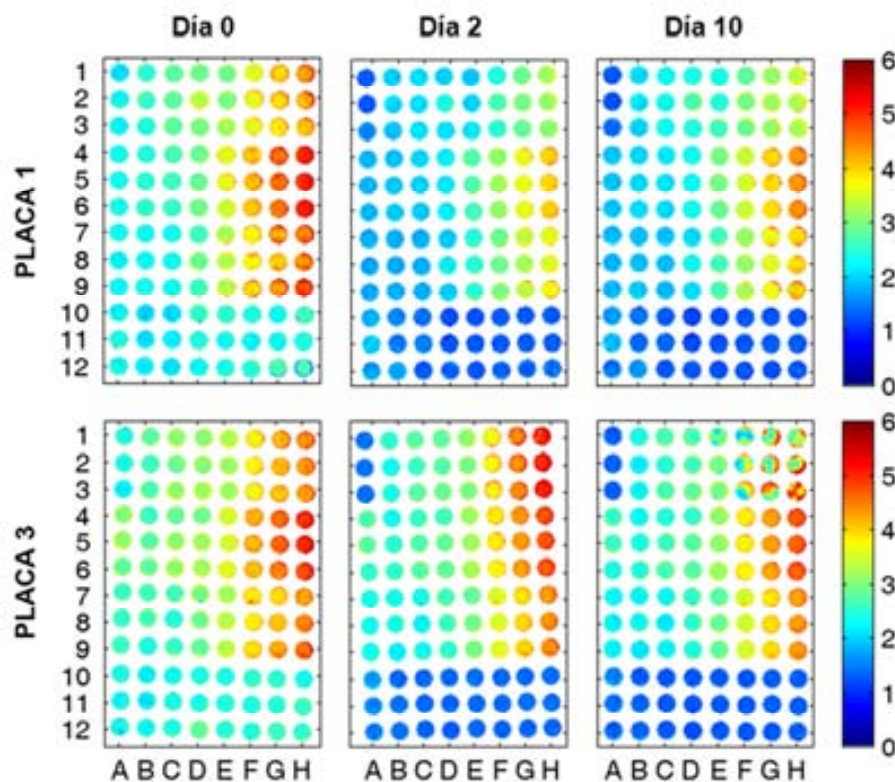


Figura 7.5. Distribución del agua en las placas de pocillos: placas 1 (arriba) y 3 (abajo) en los días 0, 2 y 10 después del freeze-drying. La composición es la misma para ambas placas, mientras que las condiciones de almacenamiento fueron del 11% de RH para la placa 1 y del 43% de RH para la placa 3.

Un detalle interesante es el hecho que estas muestras (muestras con >30% de sacarosa) que absorbieron gran cantidad de humedad durante los dos primeros días de almacenamiento, parece que han perdido agua después de 10 días de almacenamiento a alta RH. Además, la distribución del agua que presentan es claramente desigual, con algunas áreas con apariencia de manchas secas con sólo un 2% de humedad, mientras

el resto llega a más del 5%. Este efecto no se advierte para las mismas muestras almacenadas al 11% de RH, donde la distribución de la humedad en el interior de los pocillos se mantiene uniforme durante los 10 días de almacenamiento. Estos resultados sugieren que la distribución heterogénea del agua depende del nivel de RH durante el tiempo de almacenamiento. Este efecto tampoco tiene lugar en las muestras con altas concentraciones de proteína (filas 4-9, columnas F-H), a pesar de que alcanzan contenidos de humedad similar cuando se almacenan a altos niveles de humedad.

Con el objetivo de investigar con más detalle la absorción de agua a altos niveles de humedad ambiental, se evaluó la evolución de las muestras de la placa 4 a lo largo de los 10 días de almacenamiento (Figura 7.6). El contenido de agua de las muestras con más de un 80% de sacarosa (filas 7-12, columna H) disminuye ya el primer día de almacenamiento, mostrando además colapso pronunciado. En las muestras restantes, la distribución del agua es relativamente uniforme hasta el tercer día, cuando empiezan a aparecer pequeñas manchas secas en la imagen. Estas manchas van creciendo a lo largo de los 10 días de almacenamiento, hasta ocupar casi la totalidad del pocillo en algunas de las formulaciones. Una observación interesante es que la pérdida de humedad localizada sólo tiene lugar en muestras con más del 20% de sacarosa, por lo que puede relacionarse con la cristalización de la sacarosa amorfa. La consiguiente liberación de agua previamente unida a la matriz amorfa explicaría tanto la pérdida de humedad localizada, como la distribución desigual de agua en los pocillos [23].

7.3.3. *Cristalización de la sacarosa*

Para verificar que las manchas secas mencionadas anteriormente están realmente relacionadas con la cristalización de la sacarosa, se evaluaron también los difractogramas XRPD de las mismas placas. En la Figura 7.7 se muestran los difractogramas para la muestra G7 de la misma placa (placa 4). Las reflexiones características para la sacarosa aparecen claramente a partir del cuarto día de almacenamiento y aumentan hasta el día 8. La confirmación de los resultados por

XRPD demuestra que el NIR-CI en combinación con un buen tratamiento de datos es capaz de detectar las transformaciones en estado sólido en muestras liofilizadas.

Dada la cristalización de la sacarosa en algunas de las muestras liofilizadas, se pretende determinar si el NIR-CI es una técnica adecuada para el seguimiento de este fenómeno. Para ello, se evaluaron los espectros NIR de la sacarosa cristalina y amorfa (Figura 7.3). Las diferencias entre ambos espectros son evidentes y pueden aprovecharse para la visualización del proceso de cristalización a partir de los datos de las imágenes NIR.

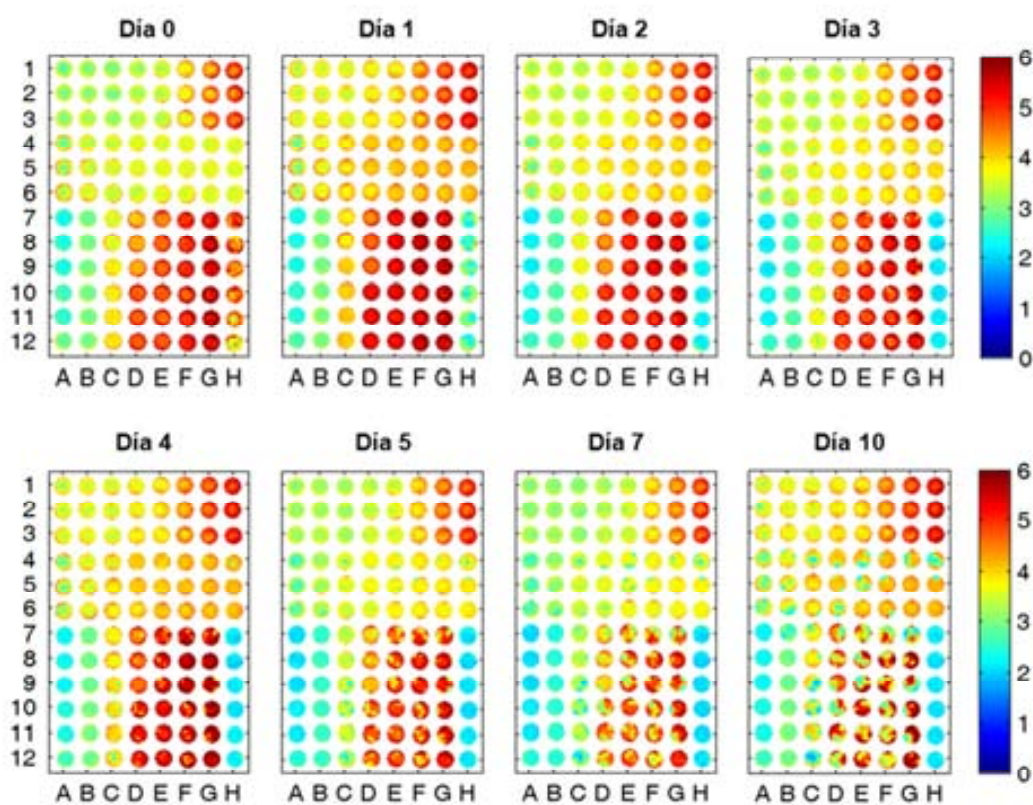
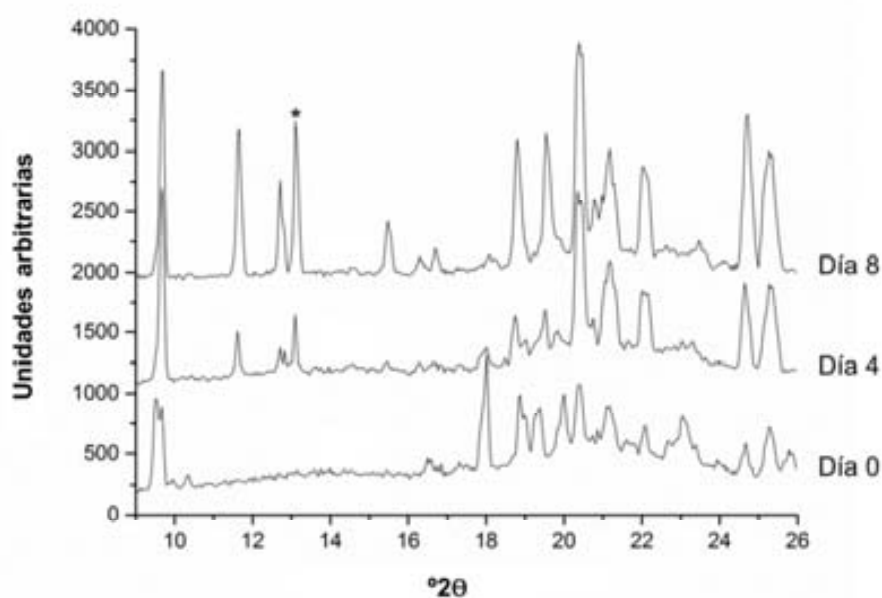


Figura 7.6. Evolución de la distribución del agua en la placa 4 durante 10 días de almacenamiento al 43% de RH.

Las imágenes NIR se transformaron en mapas de similitud espectral de la muestra con la sacarosa cristalina pura. Esto se consiguió mediante el cálculo de los CC entre ambos espectros para cada píxel de los pocillos. Los mapas resultantes (Figura 7.8) muestran la formación de pequeños dominios de sacarosa cristalina a

partir del cuarto día de almacenamiento (filas 7-12 principalmente, columnas C-H). Durante los días siguientes, los núcleos de cristalización van aumentando en tamaño hasta cubrir la mayor parte de los pocillos el décimo día, en las muestras con mayor contenido en sacarosa. Coincidiendo con las observaciones anteriores (Figura 7.6), la mínima concentración de sacarosa en la que apareció cristalización en los mapas de CC fue del 30%. El contenido en % sólido de las formulaciones tampoco tiene ningún impacto significativo en este caso, ya que se ha encontrado sacarosa cristalizada en pocillos con contenidos del 2 al 16% de sólido. Por el contrario, el contenido de proteína inhibe la cristalización de la sacarosa, dado que ninguna de las muestras con proteína mostró manchas en la Figura 7.6 o cristalización en la Figura 7.8. El hallazgo que la proteína inhibe la cristalización de la sacarosa coincide con estudios previos en sistemas que sólo contenían sacarosa y proteína [6]. Hay que destacar que la mínima concentración de proteína usada en este trabajo es de $2.5 \text{ mg}\cdot\text{ml}^{-1}$, una concentración muy elevada para formulaciones de fármacos, en los que las cantidades de proteína utilizadas podrían no ser suficientes para inhibir la cristalización de la sacarosa.



*Figura 7.7. Resultados del análisis por XRPD de la muestra G7 de la placa 4 en los días 0, 4 y 8 de almacenamiento. El manitol hemihidrato se convierte a delta manitol al desaparecer la reflexión a $17.9^\circ 2\theta$, mientras que la cristalización de la sacarosa se puede deducir de la aparición de las reflexiones a $11.7^\circ 2\theta$, $12.7^\circ 2\theta$ y $13.1^\circ 2\theta$ (reflexión específica de la sacarosa *).*

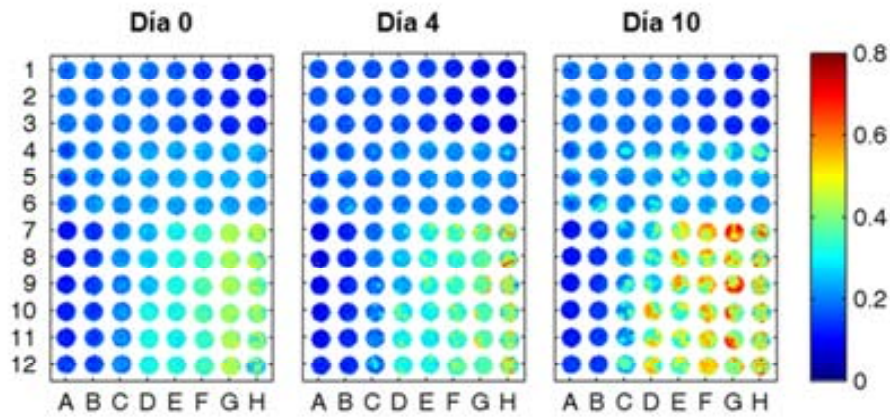


Figura 7.8. Mapas de correlación espectral para la sacarosa cristalina. Los mapas corresponden a las imágenes de la placa 4 para los días 0, 4 y 10 después de la liofilización. Los valores altos (en rojo) indican elevada correlación del espectro NIR de la muestra con la sacarosa cristalina.

Si se trata del contenido de humedad en la matriz, de la concentración de sacarosa o una combinación de ambos lo que determina el punto de inicio de la cristalización, no se puede deducir de estos datos. En cualquier caso, tal y como se advierte en las Figuras 7.6 y 7.8, los niveles críticos de humedad y sacarosa encontrados experimentalmente en este trabajo están alrededor del 4% de humedad y el 30% de sacarosa.

7.4. Conclusiones

La combinación de la liofilización por *freeze-drying* en placas de pocillos y la rapidez de los análisis por NIR-CI permiten el cribado de un gran número de formulaciones distintas en poco tiempo. En sólo 5 s se puede registrar la imagen de una placa con 96 formulaciones distintas, adquiriendo información tanto espectral, como espacial de cada una de las muestras. Esta información fue tratada quimiométricamente para la determinación de agua en las muestras. Además, se comprobó que el NIR-CI también es adecuado para el seguimiento de las transformaciones en estado sólido en las muestras liofilizadas, lo que permite la detección de la cristalización de sacarosa, un fenómeno muy pronunciado en muestras con más del 30% de sacarosa cuando se almacenan al 43% de humedad relativa.

La rápida determinación de la distribución de la humedad por NIR-CI, junto con la posibilidad de análisis del estado sólido de las muestras liofilizadas, hacen de esta técnica una herramienta muy versátil en el campo de la liofilización, con aplicaciones potenciales como el cribado de formulaciones o el control de producto.

REFERENCIAS

- 1 Kasper, J.C., Winter, G., Friess, W. (2013) Recent advances and further challenges in lyophilization. *Eur. J. Pharm. Biopharm.* **85(2)**, 162-169.
- 2 Awotwe-Otoo, D., Agarabi, C., Wu, G.K., Casey, E., Read, E., Lute, S., Brorson, K.A., Khan, M.A., Shah, R.B. (2012) Quality by design: Impact of formulation variables and their interactions on quality attributes of a lyophilized monoclonal antibody. *Int. J. Pharmaceut.* **438(1-2)**, 167-175.
- 3 Kauppinen, A., Toiviainen, M., Aaltonen, J., Korhonen, O., Järvinen, K., Juuti, M., Pellinen, R., Ketolainen, J. (2013) Microscale freeze-drying with Raman spectroscopy as a tool for process development. *Anal. Chem.* **85(4)**, 2109-2116.
- 4 Greiff, D. (1971) Protein structure and freeze-drying: The effects of residual moisture and gases. *Cryobiology* **8(2)**, 145-152.
- 5 Hsu, C., Ward, C.A., Pearlman, R., Nguyen, H., Yeung, D., Curley, J.G. (1992) Determining the optimum residual moisture in lyophilized protein pharmaceuticals. *Dev. Biol. Stand.* **74**, 255.
- 6 Costantino, H.R., Curley, J.G., Wu, S., Hsu, C.C. (1998) Water sorption behavior of lyophilized protein-sugar systems and implications for solid state interactions. *Int. J. Pharm.* **166(2)**, 211-221.
- 7 Izutsu, K., Yoshioka, S., Kojima, S. (1994) Physical stability and protein stability of freeze-dried cakes during storage at elevated temperatures. *Pharm. Res.* **11(7)**, 995-999.
- 8 Fakes, M.G., Dali, M.V., Haby, T.A., Morris, K.R., Varia, S.A., Serajuddin, A.T.M. (2000) Moisture sorption behavior of selected bulking agents used in lyophilized products. *PDA J. Pharm. Sci. Technol.* **54(2)**, 144-149.
- 9 Wang, W. (2000) Lyophilization and development of solid protein pharmaceuticals. *Int. J. Pharm.* **203(1-2)**, 1-60.
- 10 Stokvold, A., Dyrstad, K., Libnau, F.O. (2002) Sensitive NIRS measurement of increased moisture in stored hygroscopic freeze dried product. *J. Pharm. Biomed. Anal.* **28(5)**, 867-873.

- 11 Ford, A.W., Dawson, P.J. (1994) Effect of type of container, storage temperature and humidity on the biological activity of freeze-dried alkaline phosphatase. *Biologicals* **22**(2), 191-197.
- 12 Reich, G. (2005) Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv. Drug. Deliv. Rev.* **57**(8), 1109-1143.
- 13 Grohganz, H., Fonteyne, M., Skibsted, E., Falck, T., Palmqvist, B., Rantanen, J. (2009) Role of excipients in the quantification of water in lyophilised mixtures using NIR spectroscopy. *J. Pharm. Biomed. Anal.* **49**(4), 901-907.
- 14 Jones, J.A., Last, I.R., MacDonald, B.F., Prebble, K.A. (1993) Development and transferability of near-infrared methods for determination of moisture in a freeze-dried injection product. *J. Pharm. Biomed. Anal.* **11**(11-12), 1227-1231.
- 15 Lin, T.P., Hsu, C.C. (2002) Determination of residual moisture in lyophilized protein pharmaceuticals using a rapid and non-invasive method: Near infrared spectroscopy. *PDA J. Pharm. Sci. Technol.* **56**(4), 196-205.
- 16 Kauppinen, A., Toiviainen, M., Korhonen, O., Aaltonen, J., Järvinen, K., Paaso, J., Juuti, M., Ketolainen, J. (2013) In-line multipoint near-infrared spectroscopy for moisture content quantification during freeze-drying. *Anal. Chem.* **85**(4), 2377-2384.
- 17 Grohganz, H., Fonteyne, M., Skibsted, E., Falck, T., Palmqvist, B., Rantanen, J. (2010) Classification of lyophilised mixtures using multivariate analysis of NIR spectra. *Eur. J. Pharm. Biopharm.* **74**(2), 406-412.
- 18 Grohganz, H., Gildemyn, D., Skibsted, E., Flink, J.M., Rantanen, J. (2011) Rapid solid-state analysis of freeze-dried protein formulations using NIR and Raman spectroscopies. *J. Pharm. Sci.* **100**(7), 2871-2875.
- 19 Bai, S., Nayar, R., Carpenter, J.F., Manning, M.C. (2005) Non-invasive determination of protein conformation in the solid state using near infrared (NIR) spectroscopy. *J. Pharm. Sci.* **94**(9), 2030-2038.
- 20 Izutsu, K., Fujimaki, Y., Kuwabara, A., Hiyama, Y., Yomota, C., Aoyagi, N. (2006) Near-infrared analysis of protein secondary structure in aqueous solutions and freeze-dried solids. *J. Pharm. Sci.* **95**(4), 781-789.

- 21 Zidan, A., Spinks, C., Fortunak, J., Habib, M., Khan, M. (2010). Near-infrared investigations of novel anti-HIV tenofovir liposomes. *AAPS J.* **12(2)**, 202-214.
- 22 Jovanovic, N., Gerich, A., Bouchard, A., Jiskoot, W. (2006) Near-infrared imaging for studying homogeneity of protein-sugar mixtures. *Pharm. Res.* **23(9)**, 2002-2013.
- 23 Jouppila, K., Roos Y.H. (1994) Water sorption and time-dependent phenomena of milk powders. *J. Dairy Sci.* **77**, 1798-1808.

CONCLUSIONES

CONCLUSIONS

CONCLUSIONES

En esta tesis se han desarrollado nuevas soluciones analíticas basadas en las técnicas NIRS y NIR-CI para la resolución de problemas reales de la industria farmacéutica y petroquímica. Para ello, se ha aprovechado la gran versatilidad de ambas técnicas espectroscópicas, adaptando cada análisis a la coyuntura del problema (naturaleza y propiedades de la muestra, condiciones de registro y objetivos del análisis), y se ha combinado con los algoritmos quimiométricos adecuados para extraer la información de interés en cada caso y poderla interpretar.

De los estudios realizados se extraen las siguientes conclusiones:

- 1) Monitorización de distintos parámetros fisicoquímicos de gasóleos
 - a) Las condiciones de registro de los espectros NIR se adaptan al registro de muestras líquidas, como son los gasóleos, mediante la adquisición de espectros por transmitancia. El NIR se adapta también a las condiciones de trabajo continuo en planta de producción en continuo, con la aplicación de una sonda integrada en un conducto de recirculación del flujo que permite el trabajo *on-line*.
 - b) Los modelos PLS desarrollados permiten la determinación simultánea de los siete parámetros fisicoquímicos de interés en gasóleos: densidad a 15°C, índice de cetano, contenido en azufre y ésteres metílicos de ácido graso, temperatura de destilación del 95%, punto de nube e inflamabilidad. Además, las predicciones de estos modelos obtenidos permiten la monitorización del proceso productivo en tiempo real.
 - c) La estrategia de selección de muestras propuesta, basada en la aplicación de distintos análisis en PCA, permite asegurar la representatividad de las fuentes de variabilidad estudiadas en toda su magnitud en los distintos grupos de calibración. Su efectividad queda demostrada con los buenos resultados obtenidos para la predicción de los distintos parámetros de los gasóleos a largo plazo, utilizando un solo modelo por cada parámetro a pesar de la elevada variabilidad de las muestras.

- d) Los resultados obtenidos son indicativo del potencial de la espectroscopia NIR como alternativa a los métodos de referencia convencionales (ISO y ASTM) para la determinación de los parámetros fisicoquímicos de los gasóleos mencionados anteriormente.
- 2) Caracterización de la composición de parafinas para su aplicación industrial
- a) La caracterización exhaustiva de muestras industriales de parafina y aceite mediante GC/MS y NMR demuestra que ambas muestras están constituidas por hidrocarburos saturados: alcanos lineales (*n*-parafinas) con longitud de cadena comprendida entre 21 y 35 átomos de carbono en el caso de la parafina e hidrocarburos ramificados (isoparafinas) altamente isomerizados en el caso del aceite. El peso molecular medio de las cadenas se encuentra alrededor de C₂₇ en ambos casos.
- b) Se proponen dos nuevos métodos (alternativos al análisis ASTM D721) para la determinación rápida y precisa del contenido de aceite en parafinas: mediante GC y NIR respectivamente. La separación de hidrocarburos conseguida mediante GC/MS permite obtener el contenido total de aceite en las parafinas a partir de la integración de los picos.
- c) Los modelos PLS construidos con los espectros NIR de las muestras permiten cuantificar el contenido de aceite extraíble con MEK. Para ello, el uso del NIR se ha adaptado a la naturaleza de las muestras parafínicas, trabajando en transmitancia a elevadas temperaturas. Los buenos resultados, son indicativos de la capacidad del NIR en la determinación de componentes de mezclas complejas, a pesar de existir elevada similitud química y espectral entre ellos, como es el caso del aceite (isoparafinas) en una mezcla de *n*-parafinas. En esta situación, la elección del pretratamiento y el rango espectral adecuados es crucial.
- d) Las ventajas que aporta el método NIR en términos de rapidez y reproducibilidad hacen de esta técnica una buena alternativa para la determinación de aceite en parafinas industriales.

3) Determinación cuantitativa y distribución de API, excipientes y recubrimiento en comprimidos farmacéuticos

- a) En este caso, el análisis NIR aplicado en forma de imagen posibilita la obtención de información en toda la superficie de la muestra. La adquisición de los espectros se lleva a cabo en modo reflectancia difusa, de acuerdo con el tipo de muestra analizado: comprimidos farmacéuticos.
- b) Se determina simultáneamente la concentración de los distintos componentes (API y excipientes) de núcleos y comprimidos recubiertos, así como el grosor de recubrimiento de los comprimidos. En todos los casos, se analiza también la distribución del componente a lo largo de la superficie del comprimido.
- c) Se demuestra la capacidad de los algoritmos PLS e Isys-PLS1 para la extracción de los datos de interés y la obtención de información cualitativa y cuantitativa de imágenes hiperespectrales sin necesidad de una calibración con valores de referencia.
- d) Los resultados evidencian la contribución del recubrimiento en el espectro NIR y su efecto negativo en la predicción de las concentraciones de API y excipientes.

4) Detección de cambios inducidos por la humedad en formulaciones liofilizadas

- a) El análisis mediante NIR-CI es especialmente útil para el seguimiento simultáneo de un gran número de muestras en muy poco tiempo. Concretamente, en este caso se consigue registrar la imagen de una placa con 96 formulaciones distintas en sólo 5 s. De nuevo, la adquisición de los espectros se lleva a cabo por reflectancia difusa, el modo más adecuado para el registro de muestras sólidas.

- b) La información de las imágenes NIR se trata quimiométricamente para poder extraer la información relevante. La determinación de agua en las muestras se consigue mediante la construcción de un modelo univariante basado en los datos de KF, mientras que simultáneamente, la cristalización de la sacarosa se detecta mediante la aplicación del coeficiente de correlación del espectro de la sacarosa cristalina respecto a cada muestra. En ambos casos, las imágenes NIR permiten la rápida visualización de la distribución del analito de interés en las muestras.
- c) Se monitorizan los cambios inducidos por la humedad como efecto de la composición de la formulación y las condiciones ambientales: la absorción/desorción de agua y las transformaciones del estado sólido en las muestras liofilizadas. Con ello se evidencia la capacidad del NIR-CI para el cribado de las formulaciones y las condiciones adecuadas para liofilizar.

Las estrategias desarrolladas en esta memoria aportan soluciones rápidas, sencillas y eficaces a problemas analíticos reales de las industrias petroquímica y farmacéutica. Con ello se demuestra el potencial de la espectroscopia NIR y NIR-CI en el control de calidad de gran variedad de productos industriales y en distintas etapas de sus procesos de desarrollo y producción.

CONCLUSIONS

In this thesis, new analytical solutions based on NIRS and NIR-CI techniques for solving real problems in the pharmaceutical and petrochemical industries have been developed. The versatility of both spectroscopic techniques has been used to adapt each analysis to the problem features (sample nature and properties, recording conditions and objectives of the analysis). The obtained data have been combined with appropriate chemometric algorithms for extracting the information of interest in each case in a way that results can be easily interpreted.

From the realized studies, the following conclusions can be drawn:

- 1) Monitoring various physicochemical parameters of gas oils
 - a) The recording conditions for NIR spectra are fit to register liquid samples such as diesel, by using transmittance acquisition mode. NIR is also adapted to the continuous workflow in production plant with the application of a probe embedded in a flow recirculation pipe allowing the work online.
 - b) The PLS models developed allow the simultaneous determination of seven physicochemical parameters of interest in gas oils: density at 15 °C, cetane number, sulphur and fatty acid methyl esters content, 95% distillation temperature, cloud point and flammability. The predictions obtained allow the monitoring of the production process in real time.
 - c) The proposed samples selection strategy, based on the application of different PCA analysis, ensures the representativeness of the studied sources of variability in the whole range in the calibration groups. Its effectiveness is demonstrated by the good results obtained for the long-term prediction of different gas oil parameters and the use of one model for each parameter despite the high variability of the samples.
 - d) The obtained results are indicative of the potential of NIR spectroscopy as an alternative to standard reference methods (ISO and ASTM) for determining the physicochemical gas oils parameters listed above.

2) Characterization of paraffins composition for their industrial application

- a) The thorough characterization of industrial paraffin samples and oil by GC/MS and NMR shows that both samples are composed of saturated hydrocarbons: linear alkanes (*n*-paraffins) with chain lengths between 21 and 35 carbon atoms in the paraffin case and branched hydrocarbons (isoparaffins) highly isomerized in the oil case. The average molecular weight of chains is about C₂₇ in both cases.
- b) Two methods (alternative to ASTM D721) for fast and accurate determination of oil content in paraffins are proposed, by using GC and NIR respectively. GC/MS hydrocarbon separation allows paraffin total oil content determination by peak integration.
- c) PLS models constructed with samples NIR spectra allow the MEK removable oil quantification. For this application, the use of NIR has been adapted to the nature of the paraffin samples, working on transmittance at high temperatures. Good results are indicative of the ability of NIR in determining components in complex matrices, such as oil (isoparaffins) in a mixture of *n*-paraffins, despite their high chemical and spectral similarity. In this situation, the choice of suitable pretreatment and spectral range is crucial.
- d) The advantages of the NIR method in terms of speed and reproducibility make this technique a good alternative for the determination of oil in industrial paraffins.

3) Quantitative determination and distribution of API, excipients and coating of pharmaceutical tablets

- a) In this case, the NIR image analysis allows obtaining information from the whole surface of the sample. The spectra acquisition is performed in diffuse

reflectance mode according to the type of sample analyzed: pharmaceutical tablets.

- b) The concentration of different components (API and excipients) in cores and coated tablets, as well as the coating thickness of the tablets, are simultaneously determined. The distribution of each component along the tablet surface is also analyzed.
 - c) The ability of the PLS and Isys-PLS1 algorithms for extracting the data of interest and obtaining qualitative and quantitative information from hyperspectral images without calibration with reference values is demonstrated.
 - d) The results show the contribution of the coating in NIR spectrum and its negative effect on the prediction of API and excipients concentrations.
- 5) Detection of moisture-induced changes in lyophilized formulations
- a) NIR-CI is especially useful for simultaneous monitoring of a large number of samples in a short time, *i.e.* high throughput analysis. Concretely, in this case the image of a plate with 96 different formulations is achieved in only 5 s. The acquisition of the spectra is performed again by diffuse reflectance, the most suitable mode for recording solid samples.
 - b) The relevant information in the NIR images is extracted by using chemometrics. Water determination in samples is achieved by building a univariate model based on KF data. The crystallization of sucrose is simultaneously detected by applying the correlation coefficient between crystalline sucrose and each analyzed sample spectrum. In both cases, the NIR images allow the rapid visualization of distribution of the analyte of interest in the samples.

- c) Moisture induced changes as an effect of the formulation composition and storage conditions are monitored: absorption/desorption of water and solid-state transformations in the lyophilized samples. Thus, the capacity of the NIR-CI for the screening of suitable formulations and conditions for freeze-drying are evidenced.

The strategies developed in this thesis provide quick, easy and effective solutions to real analytical problems of the petrochemical and pharmaceutical industries. The provided solutions demonstrate NIRS and NIR-CI potential in the quality control of a wide variety of industrial products and different stages of their development and production processes.

ANEXOS

ANEXOS

- I) Robust PLS models for monitoring physico-chemical properties of diesel oils by NIR spectroscopy
- II) Characterization of paraffin waxes composition towards industrial applications
- III) Determination of drug, excipients and coating distribution in pharmaceutical tablets using NIR-CI
- IV) NIR imaging for high-throughput screening of moisture induced changes in freeze-dried formulations

ANEXO I

Robust PLS models for monitoring physico-chemical properties of diesel oils by NIR spectroscopy

Anna Palou, Marcelo Blanco, Rafael Larraz, José Francisco Gómez,
Teresa

Martínez, Josep Maria González, Manel Alcalà

Artículo enviado

Robust PLS models for monitoring physico-chemical properties of diesel oils by NIR spectroscopy

Anna Palou^a, Marcelo Blanco^a, Rafael Larraz^b, José Francisco Gómez^c, Teresa Martínez^d, Josep Maria González^d, Manel Alcalà^{a}*

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

^b Centro de Investigación CEPSA, Avda. Punto Com 1, 28805 Alcalá de Henares (Madrid), Spain

^c Refinería Gibraltar – San Roque CEPSA, Puente Mayorga, s/n, 11360 San Roque (Cádiz), Spain

^d CEPSA, Campo de las Naciones, Avda. del Partenón 12, 28042 Madrid, Spain

KEYWORDS: Diesel oil physico-chemical properties, Near-infrared spectroscopy, Principal component analysis, Partial least-squares regression.

ABSTRACT

Regulations on gas oil marketing require producers to fulfil stringent specifications for a number of physico–chemical parameters. The analytical methods typically used to determine such parameters are labour-intensive and time consuming, so they preclude the expeditious obtainment of the information needed to release a product and assure quality in its production. In this work, we used near-infrared (NIR) spectroscopy to determine seven physico–chemical properties of gas oils obtained at an industrial production plant. The products studied differed widely in crude origin and composition, and their differences reflected in their NIR spectra. The fact that the samples came from various production streams using different fatty acid methyl esters (FAMEs) and from different transformation processes increased their variability even further. Special attention is required to include all this variability in the calibration set to ensure that the ensuing chemometric models would be robust enough to retain a good predictive ability in the long term. Samples were selected according to three different criteria, namely: the range spanned by the reference values for the target parameters, the nature and variability of the spectra, and the origin of the crudes used to produce the gas oils. A specific calibration set was assigned to each target parameter. The ensuing partial least-squares (PLS) models exhibited a good predictive ability in all instances and continue to provide good predictions, which confirm their robustness.

1. INTRODUCTION

The production and marketing of industrial products is always subject to control tests involving the determination of specific parameters to assure quality in the products and earn consumers' confidence. EN 590¹ and ASTM D 975² are the European and USA reference standards, respectively, for quality assurance of gas oils and their mixtures (biodiesels, petrodiesel blends) by the petrochemical industry. These standards establish a limiting value for each parameter to be controlled and the reference method to be used for its determination.

Assuring quality in commercial gas oil entails determining a number of parameters (15–17 depending on the particular standard). These parameters include physico–chemical properties such as density, cloud point and flash point, which are typical of each petrodiesel, and the contents in some mixture components such as fatty acid methyl esters (FAME)³.

The reference analytical methods typically used to determine physical properties of gas oils involve labour-intensive, time-consuming laboratory procedures⁴ using specific instruments for each determination. On the other hand, near infrared (NIR) spectroscopy allows both processes and quality-related parameters to be characterized as accurately as or even more so than existing reference methods for this purpose^{5–8} but in a much more expeditious manner.

The NIR technique is rapid and non-destructive; also, it requires no sample pretreatment, which makes it a suitable choice for analyses in real time⁹. In addition, it can be implemented with fibre optics to enable in-line/on-line process monitoring for

the analysis and characterization of raw materials¹⁰, monitoring and surveillance of processes⁸, and quality assurance of end-products¹¹. The suitability of NIR spectroscopy for these purposes is supported by its large number of uses in the petrochemical industry and by two ASTM norms currently in force, namely: ASTM 1655 (2012) for quantitative analysis by mid infrared and near infrared analysis, and ASTM 1790 (2010) for qualitative analysis with the two infrared techniques⁴.

A near-infrared spectrum can contain large amounts of information on any physical¹² or chemical property¹³ associated to absorption in this region of the electromagnetic spectrum. Efficiently extracting it requires using a suitable chemometric tool to relate spectral variability with information of interest about the target sample. The most commonly used chemometric tools for NIR spectra include principal component analysis (PCA) for qualitative analysis and partial least squares (PLS) regression for quantitative analysis. PCA and PLS are two highly powerful algorithms applicable in variable forms to different situations the combined use of which other chemometric tools enhances their modelling and calibration capacities^{4,7,12,13}.

In this work, we used NIR spectroscopy for the on-line, *in-situ* monitoring of quality-related parameters in gas oils obtained by blending. Gas oil blends are highly complex mixtures of distillates from crudes of variable composition and properties depending on their origin and extraction depth¹⁰, which increases the variability of the blends. Oil crudes are multicomponent mixtures consisting mainly of hydrocarbons (naphthenes, paraffins and aromatic substances) in addition to sulphur and nitrogen compounds, and metal impurities¹⁰. The differences in crude origin and composition, and the changes caused by blending in various refining processes, reflect in the NIR spectra for gas oil samples and make it rather complicated to develop effective models

for determining the required parameters. Biodiesels additionally contain FAME of also variable origin and composition. The problem is usually addressed by constructing a separate model for each gas oil type; however, routine application of the model requires the prior selection of crude origin^{8,14}. In addition, it is rather difficult to control all potential sources of variability in the samples and processes, and to have a wide enough range of variation for each target parameter—which requires acquiring sample data over long periods prior to constructing the model¹⁴.

The primary aim of this work was to develop effective PLS models for monitoring various physico–chemical properties of gas oils that govern quality in the end-product by using a single model for each parameter irrespective of sample origin. To this end, the samples used to construct the models were chosen in such a way as to maximize variability. PCA revealed that the calibration set contained all sources of variability present in the sample set and hence that the ensuing model was robust enough for application to very large numbers of widely variable gas oil samples.

2. EXPERIMENTAL SECTION

2.1. NIR operating conditions. Near-infrared spectra were recorded on a Bartec Benke GmbH FT-NIR Matrix-F spectrophotometer. Spectra were on-line acquired in the transmittance mode, using a probe inserted in the gas oil recycling stream from the blending line. All spectra were recorded at the temperature of the blending process (about 37 °C).

Each spectrum was acquired in 10 scans of the 1000–2200 nm wavelength range at 2 nm intervals. Spectra were recorded with the aid of the software OPUS-NT and

Calibration Workshop (CWS) v. 2.10.001 from SensoLogic GmbH (Norderstedt, Germany).

2.2. Samples. All samples studied were post-blending gas oils produced at the CEPSA plant in San Roque (Spain) after November 2011. A total of 278 samples obtained until April 2012 were used to construct, validate and assess models. Subsequent samples were used to assess performance of the models on site.

The samples composition included the addition of small amounts of kerosene, atmospheric and heavy atmospheric gas oils from a number of refining processes such as fluid catalytic cracking (FCC), direct distillation, coking and visbreaking, and hydrotreatment. The spectra of all calibration and validation samples are showed in Fig. 1a. Their greatest source of variability was their composition, which was essentially dependent on the origin of the crudes and undetectable having a glance at their spectra. Usually, hydrocarbon streams employed to formulate the gas oil can be derived from 2 – 5 crudes of variable origin and streams from various refining processes. Models were constructed and validated by using biodiesels consisting of mixtures of 12 crudes (C1-C12) from 10 different geographical origins plus one of no specific origin (a mixture of the previous ones). Eleven crudes were incorporated at a later time during the validation of the models.

2.3. Reference data. Samples were collected directly from the pipeline three times per day and referenced in the laboratory. Each sample is subsequently associated to the spectrum registered at the same time. Most of the previous samples had reference values

for the target physico–chemical parameters, namely: density at 15 °C (Dens, g/cm³), cetane index (CI, dimensionless), sulphur content (S, mg/kg), content in fatty acid methyl esters (FAME, % v/v), 95% distilled (D. 95%, °C), cloud point (CP, °C) and flash point (FP, °C). The precision data of the reference methods¹⁵ are listed in Table 1.

2.4. Chemometrics and data pre-processing. This section describes the specific procedures used for chemometric treatment and pre-processing of the data, and their most relevant characteristics for the intended purpose. The specific algorithms used, however, are described in detail elsewhere¹⁶⁻¹⁷.

2.4.1. Spectral pretreatment. Spectra are usually subjected to some pretreatment prior to constructing calibration models based on them. Spectral pretreatments are intended to suppress unwanted contributions due to various factors including physical properties of the sample, slight differences in the recording conditions or, simply, instrumental noise, in order to improve the signal-to-noise ratio and facilitate the development of more simple and robust calibration models.

In this work, we assessed various pretreatments and chose the second spectral derivative with smoothing (2nd Der.), using an 11-point window and the Savitzky–Golay algorithm (Fig. 1b). This removed baseline effect and improved resolution between the typically broad, overlapped NIR bands without significantly increasing spectral noise. Also, some parameters were modelled by subsequently applying the standard normal variate (SNV), an algorithm that autoscales (i.e. centres at a zero mean intensity and unity standard deviation) each individual spectrum. In this way, SNV corrects

intensities and baseline deviations due to light scattering possibly generated by microdrops of non-miscible liquids, bubbles or density fluctuations in the sample. The previous pretreatments were all applied to absorbance spectra, using the software The Unscrambler v. 9.7 from CAMO (Trondheim, Norway) and the above-mentioned CWS.

2.4.2. Development of calibration models. The target parameters were quantified by using PLS regression models. The models were constructed from a calibration set consisting of samples of the same nature as those to be used for prediction, having known reference values and spanning the whole range of the quantity to be measured. Therefore, the calibration set included every potential source of spectral variability¹⁴ and was representative of both known samples and samples to be predicted.

The models were constructed by cross-validation. The optimum number of factors to be used was selected in terms of the average residual of the predictions, root mean square error of prediction (RMSEP) and Student's *t*-value for the residuals (or the *F*-value of Haaland and Thomas¹⁸ in case of uncertainty). The *F*-test of Haaland and Thomas allows the model with the lowest predicted residual sum of squares (PRESS) to be replaced with one containing fewer factors. If the difference between the PRESS values is significant enough, then the initial model (viz. that with the lowest PRESS) is retained; otherwise, the alternative model is selected.

The software used for on-site analyses, CWS v. 2.10.001, is not appropriated to construct calibration models, since is designed to apply them and provide results. Therefore, all models were constructed, optimized and validated by using the software The Unscrambler v. 9.7, which is more adequate and versatile for building models, and subsequently recalculated with CWS v. 2.10.001. Recalculation was imposed by the fact

that the pretreatments and algorithms used were not exactly identical with the two software packages and hence led to differences from the original PLS model.

2.4.3. Sample selection. The samples for the calibration set, described in section 2.2, were selected in various ways in order to ensure inclusion of the most representative for the model concerned while suppressing redundant information and minimizing noise.

Thus, spectral variability was maximized by using the algorithm of Kennard and Stones¹⁹. The algorithm selected the sample with the closest value to the mean for the set (i.e. the presumably most representative sample). Then, it sequentially selected those samples projected at the greatest distance from the previously selected samples in the design space. Calculations for this purpose were done with the software MATLAB v. 7.0 from The MathWorks (Natick, MA, USA).

The calibration samples were chosen in such a way as to the reference values uniformly span the whole operating working for each target parameter.

The results were subjected to PCA in order to condense the information contained in the initial data and project it on as few latent variables or principal components (PCs) as possible. PCs were ranked according to proportion of variance explained in order to facilitate identification of sources of variability in the samples. In this work, we applied PCA to three different types of data, namely: variably pretreated spectra, reference data for the target parameters and crude composition. PCA models were constructed by using The Unscrambler v. 9.7 and The Unscrambler X v. 10.2, both from CAMO (Trondheim, Norway). The information thus obtained was additionally used to detect

calibration and prediction outliers, and was highly useful for assessing variability in future prediction samples—which is crucial with a view to accurately characterizing a process⁴.

3. RESULTS AND DISCUSSION

3.1. Sample selection. As stated in Section 2.4.2, appropriate selection of the calibration samples is crucial in the development of robust PLS models with a good predictive ability (especially when dealing with highly variable samples).

In this work, crude origin and composition, the two major sources of variability in the diesels, resulted in slight differences between NIR spectra that were the keys to developing the required calibration models. This required including as many crude types in the calibration sets as possible in order to incorporate the greatest possible variability with a view to maximizing quality and robustness in the ensuing models. To this end, we used a combination of the typical sample selection criteria (viz. spectral information and reference values) and crude composition to construct a specific calibration set for each target parameter (i.e. for each individual model).

3.1.1. Spectral information. The spectral information obtained was subjected to PCA. A scores plot allows the degree of similarity or difference between samples to be assessed in order to place them in the working space. This facilitates the detection of potential spectral outliers by using Hotelling's T^2 statistic to assess their response to a pretreatment. Figure 2 shows the scores plot for PC1 vs PC2 of the PCA on the spectral data in the absorbance (Fig. 2a) and the 2nd Der. mode (Fig. 2b). As can be seen, those

samples falling outside the Hotelling ellipse in Fig. 2a (i.e. potential spectral outliers) were clearly corrected in Fig. 1b. Therefore, the 2nd Der. treatment suppressed differences and avoided the need to discard outliers.

Plotting the scores of a spectral PCA against the reference values for each target parameter allowed potential outliers with an erroneous reference value to be identified and whether the parameter concerned would be easily modelled to be anticipated —and, if so, the most suitable treatment for this purpose to be selected. Figure 3a shows the scores of PC1 —the component explaining the highest proportion of the variance in the PCA of the spectral data (Fig. 2a). As can be seen, there was no apparent relationship: the scores were independent of the reference value for the CI. Figure 3b is the counterpart of Fig. 2 for the 2nd Der. spectra. Unlike before, a clear trend was apparent: the reference value for CI decreased with increase in the scores. This indicates that CI can be accurately modelled with this pretreatment. A similar trend was observed for the parameters density and FAME. The relationships for the remaining parameters were not so apparent, which suggests that they were less closely related to the spectral characteristics of the samples and hence more difficult to model than the previous ones.

After analysis, the spectra were used for a preliminary selection of samples to be included in the calibration set by using the algorithm of Kennard and Stones (see Section 2.4.3). Sixty of the absorbance spectra were included in the calibration set and the remainder in the validation set.

3.1.2. Reference values. The reference values of the target parameters for the whole set of samples were subjected to PCA in order to facilitate construction of the models. Previously, the values were autoscaled to unity variance in order to avoid an excessive

influence of the parameters with the greatest values and hence an unbalanced contribution to the models. The scatter plot of the PC scores provided a rough depiction of the behaviour of the body of values for the target parameters of a sample in relation to all others; also, the loadings plot exposed the relationships between parameters. The bi-plot of Fig. 4 (superimposed scores and loadings) reflects the influence of each parameter on the samples. Thus, the scores allowed two distinct sample clusters to be distinguished and the loadings revealed that the separation was due to parameter 9 (FAME). This was the sole parameter clearly related to one of the clusters: the upper cluster, including all samples containing FAME, and the lower cluster, consisting of all other samples. This result was confirmed by checking the FAME values and suggested that the FAME model would be easy to construct and also that the graph would clearly reflect whether a sample contained any FAME.

Table 2 lists the ranges spanned by the samples and Fig. 5 shows the corresponding distribution histograms. As can be seen, most of the ranges were very narrow and the majority of values fell at one end. The lack of data uniformly distributed across the working range hindered the development of accurate calibration models and reduced their predictive ability. This was especially so with the parameters S and FAME. Thus, more than 95% of the samples had S values falling in one of three central classes of the ten in the histogram. In his situation, the few extreme values of the parameters were crucial in constructing the PLS models and the central values required very careful selection in order to avoid using redundant information and incorporating noise into the model. On the other hand, more than 55% of samples had FAME values falling at the lower end of the range and no sample fell in the central region. This led to sample segregation in the PCA (see Fig.4) and required careful selection of the calibration samples in order to span the whole working range.

All those samples with extreme or poorly represented reference values were added to the calibration sets initially constructed by using the Kennard–Stones algorithm.

3.1.3. Crude composition. The information about crude composition for each sample (viz. whether it contained a given crude) was used to construct PCA models, again with exclusion of spectral information. This analysis was special in that it used qualitative variables that were converted into quantitative variables by using a binary code: 1 if the crude concerned was present and 0 otherwise. Because a total of 12 crudes were considered, each sample was assigned 12 values equal to 1 or 0 for the PCA.

The closeness between samples (scores) in the PC plot was a measure of their similarity as regards crude composition; thus, two overlapped samples had an identical composition. The closeness between crudes (loadings) was an indication of their coinciding in a particular combination in the gas oils; thus, two overlapped crudes were present in the same gas oil samples.

In a scores and loadings bi-plot, each sample would fall closest to the crudes it contained. As can be seen in Fig. 6, constructed from the same number of samples as Fig. 4, the samples of identical composition were superimposed and close to the loadings reflecting their composition. Also, a well-defined cluster of samples separated from the others was observed. The cluster, defined by the black ellipse in the figure, consisted of 42 samples having 4 different compositions and very close to the crudes C5 and C12 (in red). All samples in the cluster were confirmed to contain both crudes. This led us to include some samples in the cluster in the calibration sets for all parameters.

All those samples containing crudes not represented in the calibration set were incorporated into the previously established calibration sets. One easy way of identifying the specific compositions to be included in the calibration set was by projecting its samples onto the composition PCA plot for all samples. As can be seen from Fig. 7, the projected samples superimposed with others in all cases. The non-superimposed samples were those whose composition was not represented in the projected calibration set.

3.2. Construction of PLS models. *3.2.1. Original models.* On-site quantitative determinations of the seven target parameters were performed by using an optimized PLS calibration model for each (see Section 2.4.2). Table 3 shows the figures of merit of the models constructed with the software The Unscrambler as assessed in terms of average residuals and RMSEP.

As can be seen, all models explained more than 95% of the total variance with a number of PCs ranging from 2 for CI to 8 for the distillate. The average residuals were acceptable in all cases and RMSEP was similar to the repeatability of the reference methods. The statistics for the Dens, CI and FAME models were especially good; thus, they explained more than 97% of the total variance with 4 factors or fewer and exhibited very low average residuals and RMSEP values. The fact that these three parameters were modelled better than the others confirms the hypotheses established from the PCA results. In fact, the three properties are closely related to the spectral characteristics of the samples and hence easy to model.

It is also important to note the positive statistics of S, since the sulphur lack of dipole moment makes NIR an insensitive technique to detect it and their content in the

gas oil samples (see Table 3) is lower than NIR detection limit (200ppm). The goodness of the results can be ascribed to the modification of certain sample properties when S varies.

3.2.2. Recalculated models. The software governing operation of the plant, CWS, excluded some of the spectral pretreatments present in The Unscrambler; also, some pretreatments (e.g. derivative algorithms) differed between both softwares. As a result, their models were different and so were their predictions of the target parameters. In order to solve the problem, we examined the different pretreatments in CWS leading to similar results in The Unscrambler. Thus, the PLS models constructed with Unscrambler and 2nd Der. with the Savitzky–Golay algorithm (11-point window, second-order polynomial) were recalculated by using 2nd Der. gap–segment (2 gaps, 3 segments) in CWS. Although the results were not exactly identical, the differences were acceptable. This procedure was then applied to all target parameters. Table 4 shows the figures of merit of the CWS-recalculated models as assessed identically with the original models to ensure comparability. The pretreatments used were the same in all instances and only the number of factors used was varied in some cases to obtain as similar as possible calibration statistics.

As can be seen, the models for the first three parameters (Dens, CI and FAME) and D. 95% retained the goodness of the statistics for the original models —some, however, required using a greater number of factors. The statistics for the other parameters (S, CP and FP) were slightly worse than those for the original models; based on their results and RMSEP values, however, they were acceptable for predictive purposes. Whereas the recalculated model for S used the same number of factors as the original model, those for cloud point and flash point used fewer factors. In all cases, the optimum number of factors for CWS was used.

3.3. Monitoring of model performance. The models constructed with CWS were used to predict the target parameters on site and the resulting predictions monitored to ensure whether the models would retain their effectiveness over time. It has been done by periodically predicting the parameter values for new samples and checking their accuracy in terms of the individual residuals for each sample to ensure that the magnitude of the errors remained constant and uniformly distributed in time.

Fig. 8 shows the residuals of the parameter density for all samples analysed from April 2012 to February 2013. As can be seen, the residuals were constant and near-zero in most cases. By exception, three samples had markedly increased values. These deviations were also observed in all other parameters except the cloud point (results not shown) and coincided with stops of the spectrophotometer to solve a problem with the probe in July 2012. The graph indicates that the model retained its robustness and detected deviations in the predictions due to problems with the NIR equipment. Table 5 lists the average residuals and RMSEP values for the samples —those subject to the previous deviations excluded.

The proposed methodology therefore proved effective for predicting the seven target parameters and detecting any problem arising during the data acquisition process. New samples can be analysed with the same criteria and methods for sample selection (reference value, spectra and composition). The decision of incorporating new spectra into the calibration model is made on the basis of a PCA projection of the new spectra on the calibration/validation set, which allows one to easily identify the problem and choose what type of sample to include if the calibration model is to be updated.

4. CONCLUSIONS

In this work, we assessed the ability of NIR spectroscopy for the on-site monitoring of seven physico-chemical properties of gas oils, namely: density, cetane index, sulphur content, FAME content, 95% distilled, cloud point and flash point.

PLS models exclusively constructed from on-site recorded NIR spectra allowed the accurate determination of each target parameter. Given the high variability of gas oil samples, the good performance of these models is to a great extent a result of careful selection of samples in terms of their reference values, spectral properties and composition over a long enough periods for the required amount of variability to accumulate.

At present, the models continue to be used on-site in order to check their robustness and accuracy over time. However, the results obtained so far, and the ease and expeditiousness with which samples can be measured, demonstrate that the NIR technique is a highly suitable choice for the quality control of gas oils in real time in a much more practical manner than with existing reference techniques for this purpose.

FIGURES

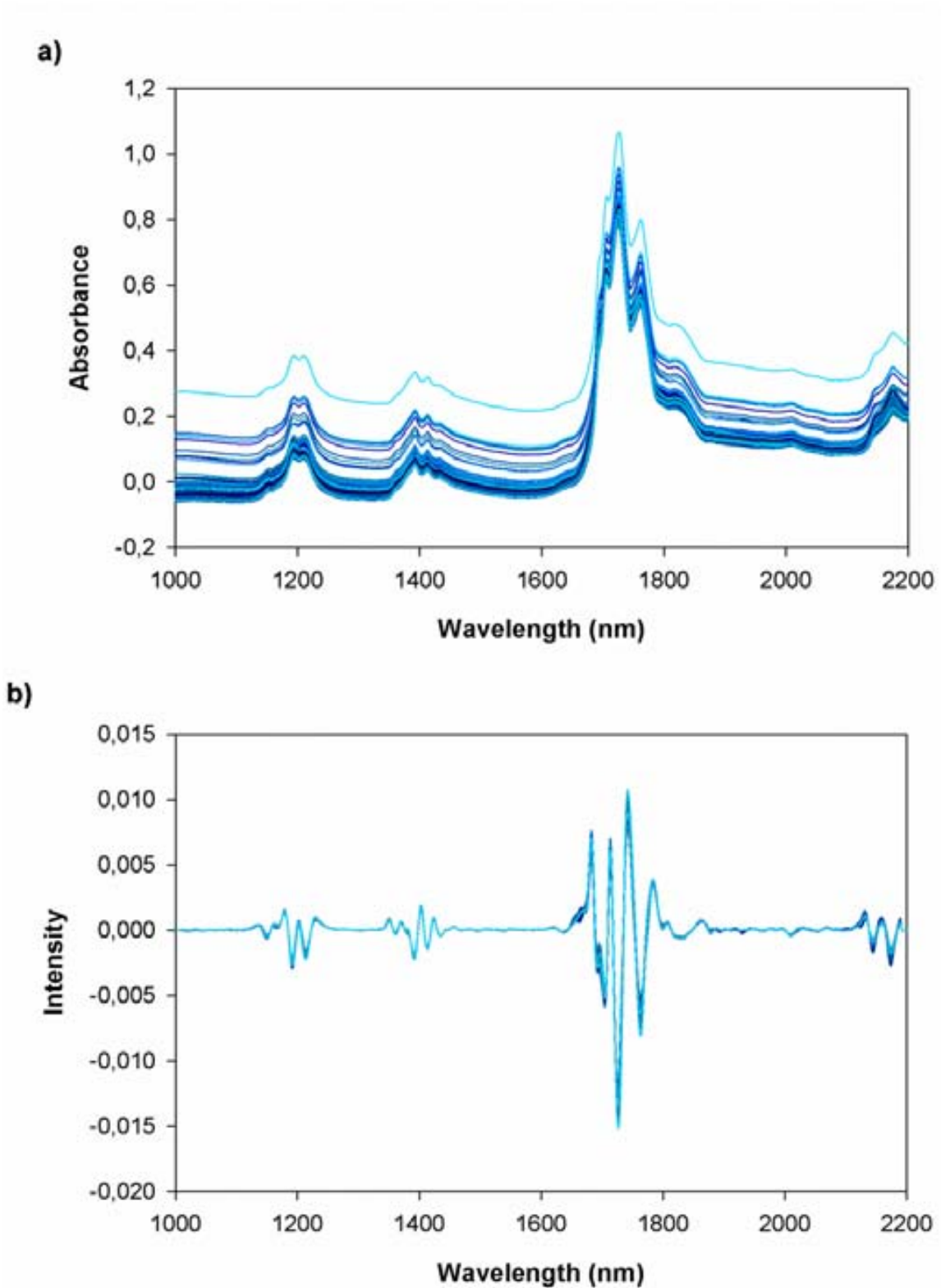


Figure 1. NIR spectra of calibrations and validation samples: (a) Absorbance; (b) Second derivative pretreatment.

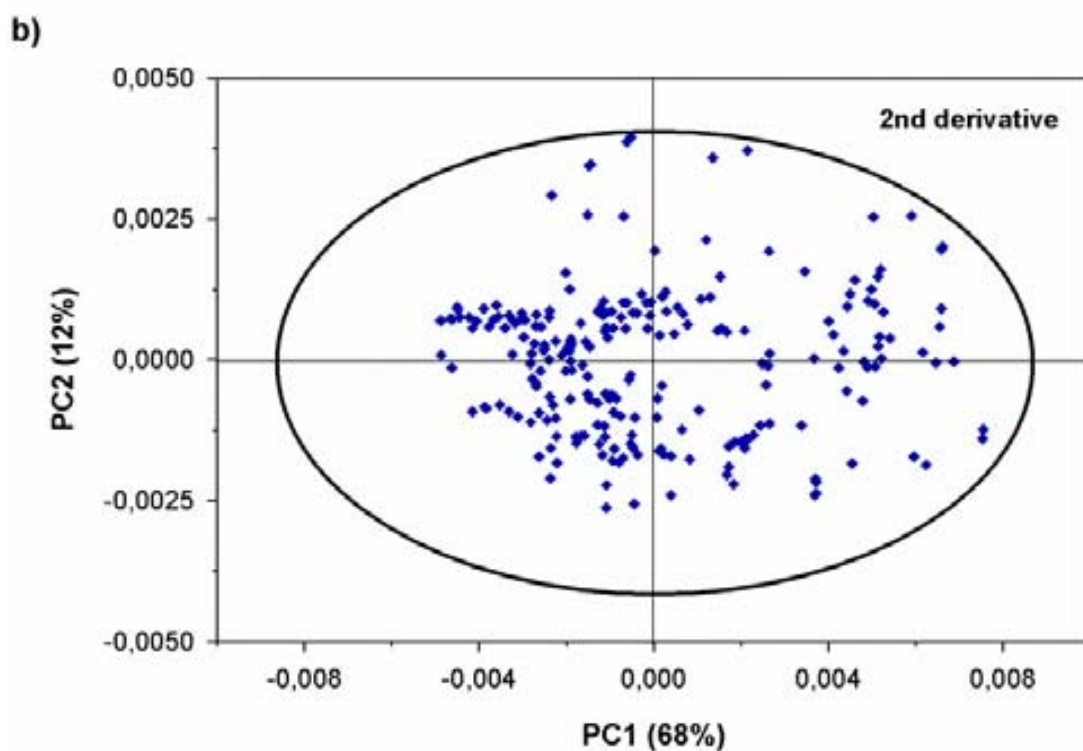
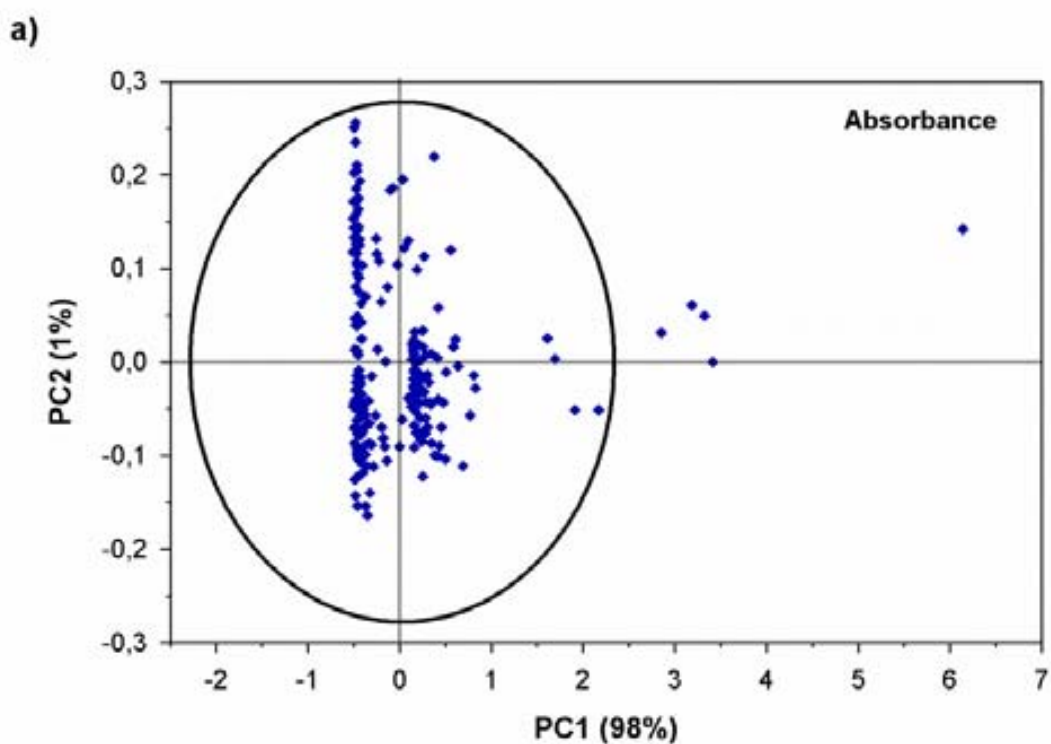


Figure 2. Principal component analysis (PCA) scores plot of all calibration and validation samples: (a) absorbance; (b) second derivative. Hotelling's T^2 ellipse calculated for 99% significance.

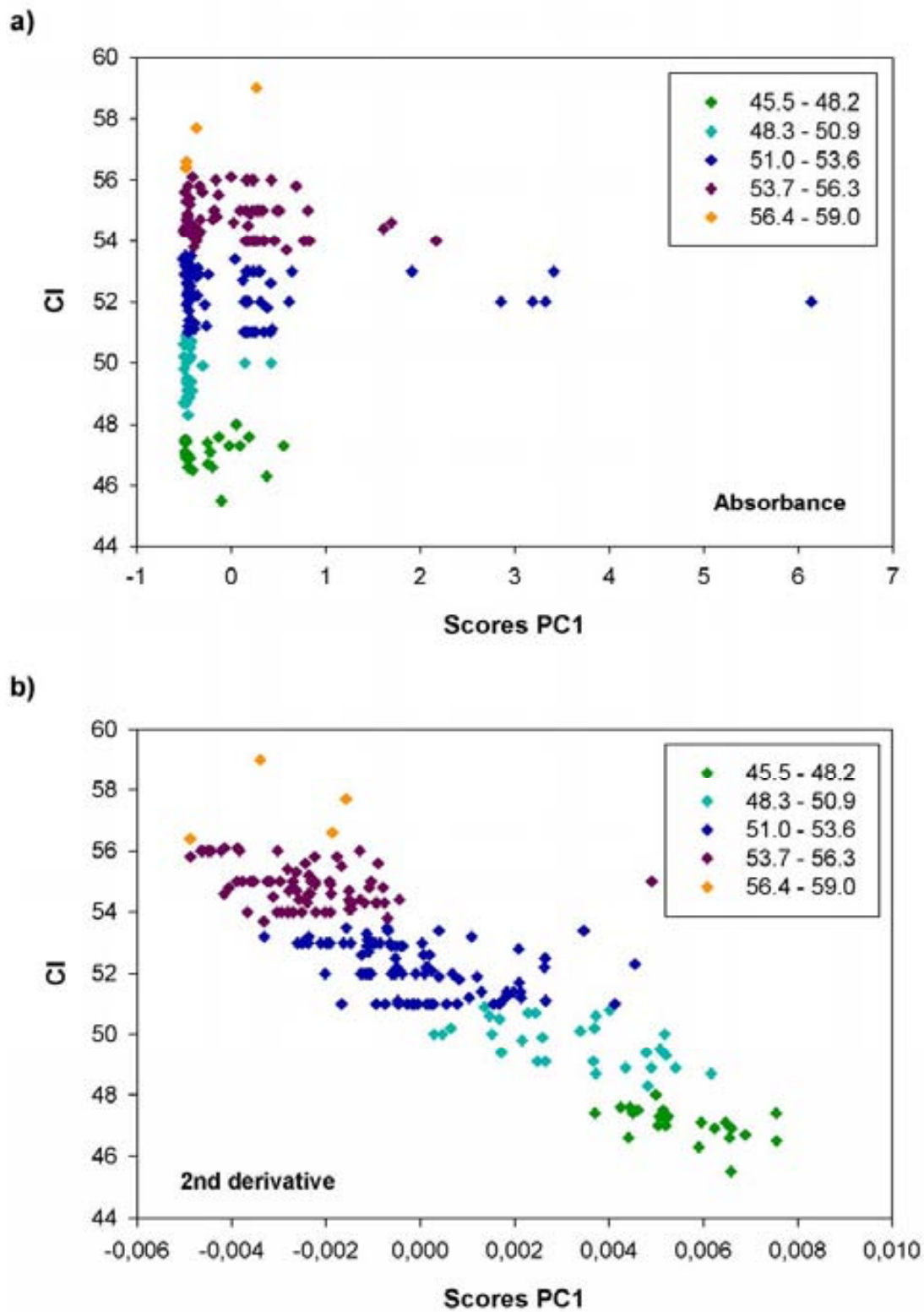


Figure 3. PC1 scores plot of Figure 1 vs cetane index (CI) reference values for all calibration/validation. Color group indicates range cetane index reference values: (a) of absorbance spectra; (b) of second derivative spectra.

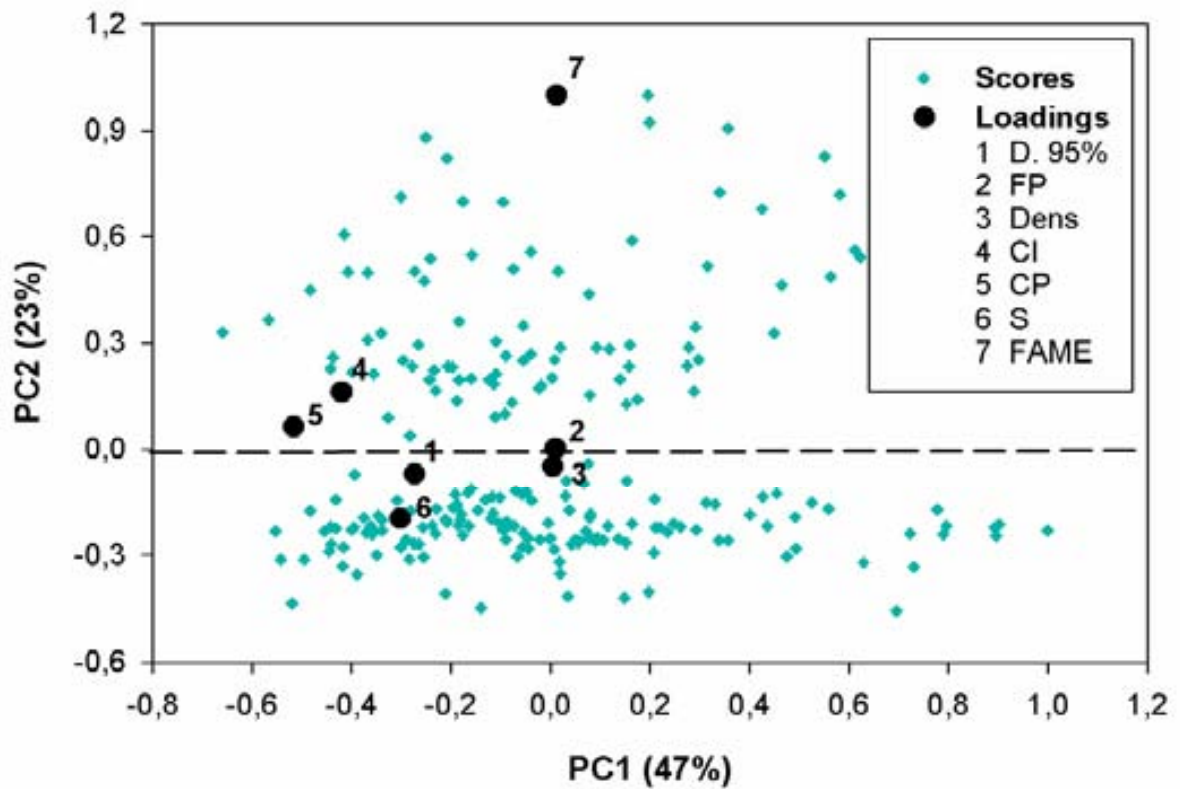


Figure 4. PCA scores and loadings bi-plot of reference values for all calibration/validation samples. Scores (samples) are shown in blue and loadings (parameters) in red. Samples in the upper (orange) circle contained FAME, whereas those in the lower (green) circle did not.

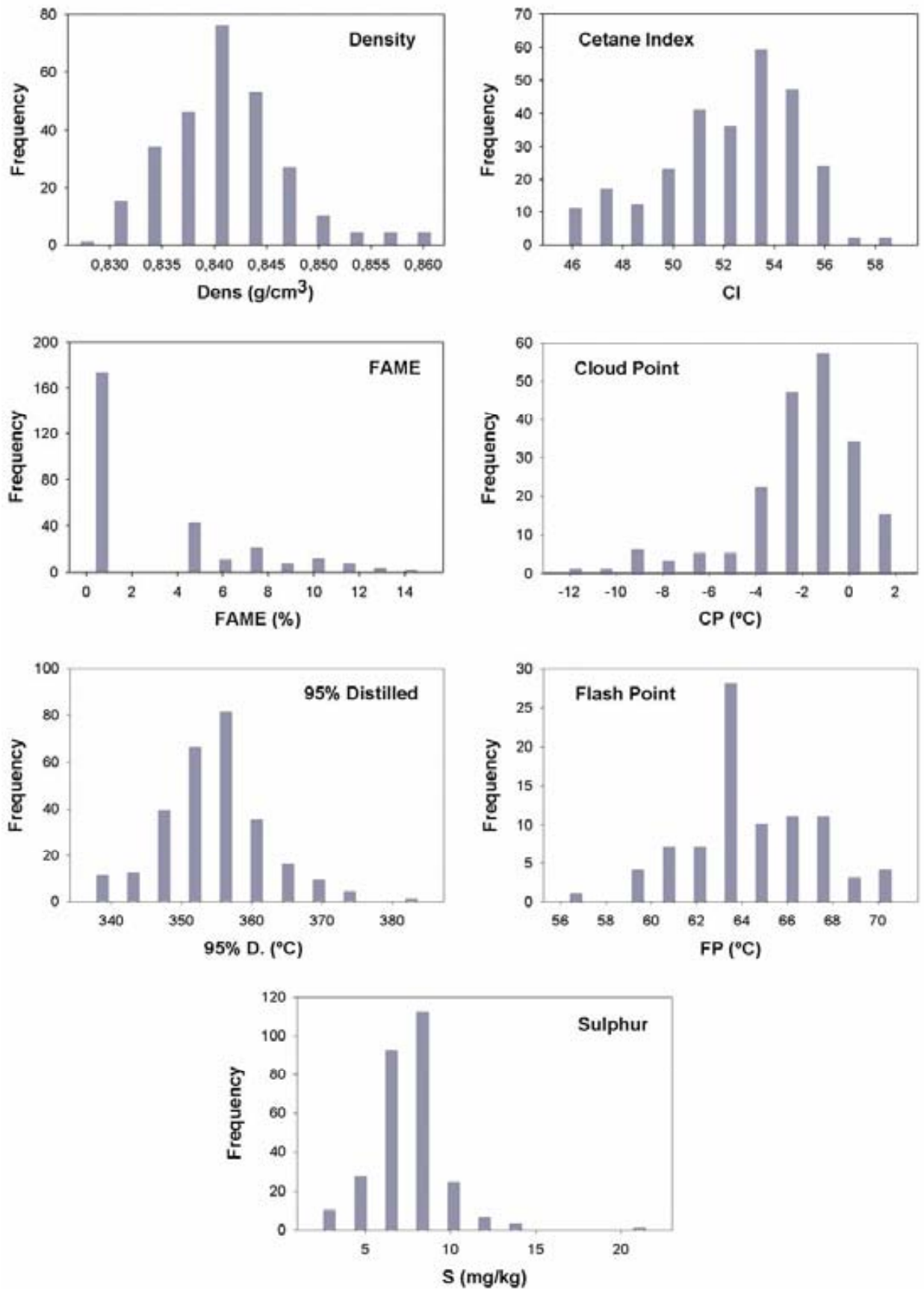


Figure 5. Distributions of reference values for each parameter.

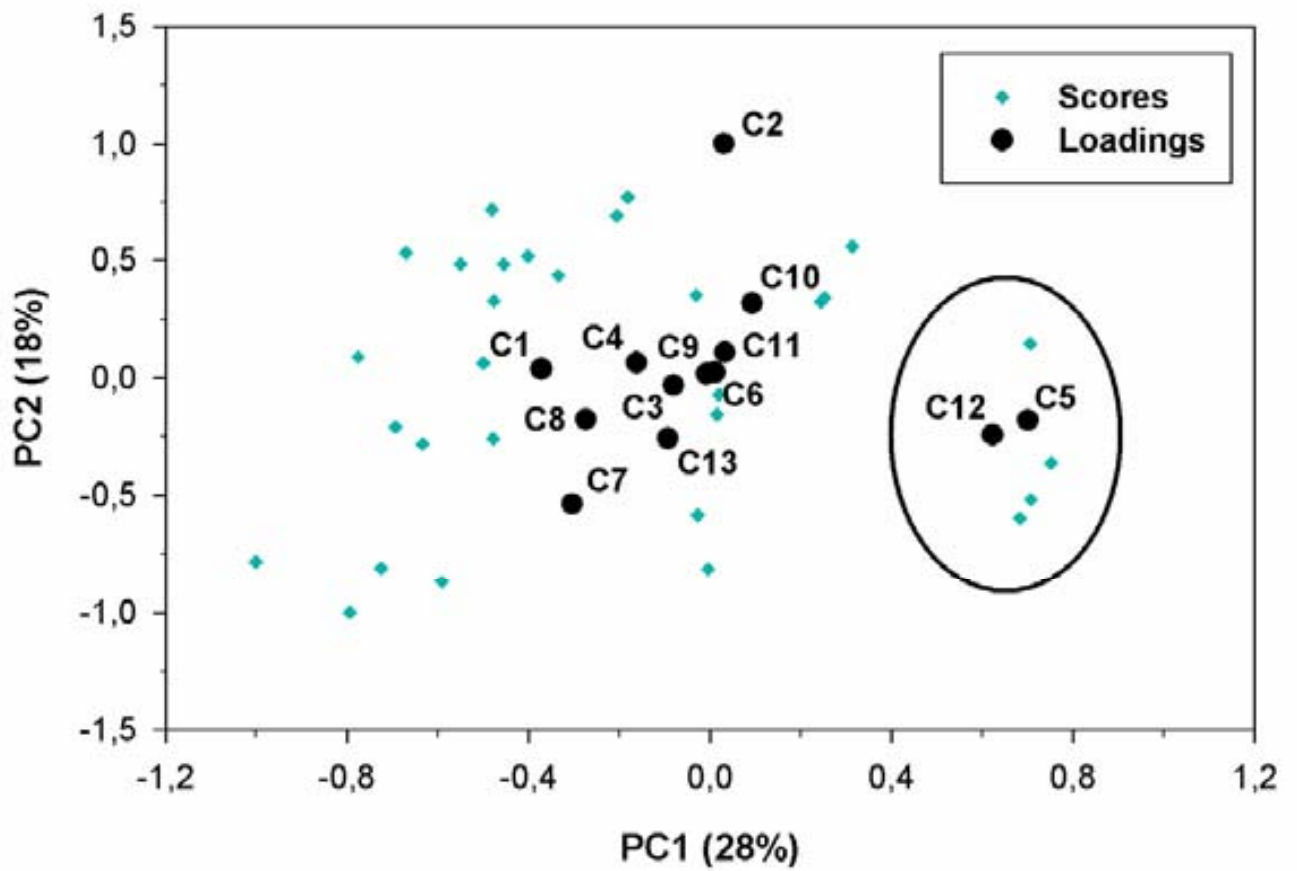


Figure 6. PCA scores and loadings bi-plot of the crude composition for all calibration/validation samples. Scores (samples) are shown in blue and loadings (crude types) in red. The circle encompasses all samples containing the crudes C5 and C12.

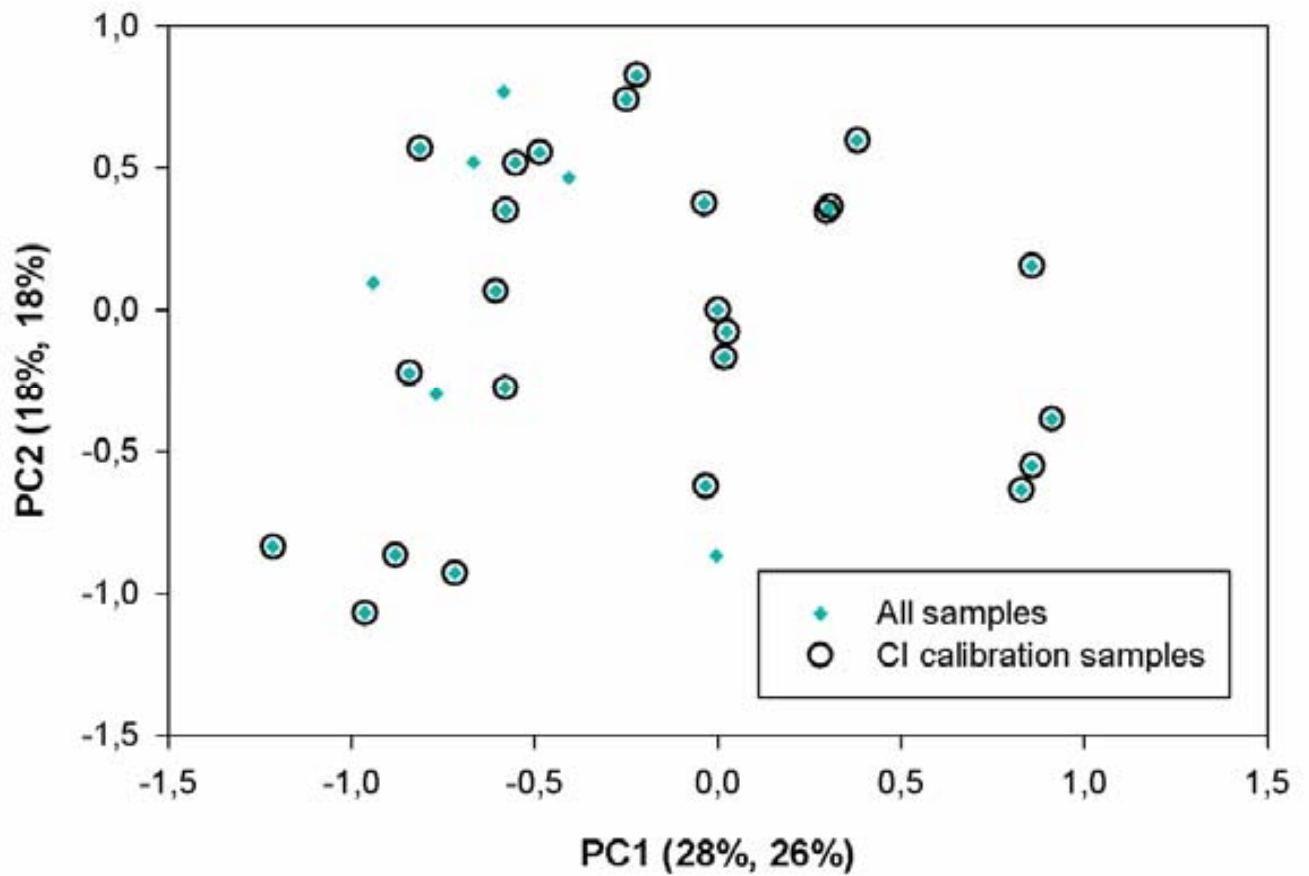


Figure 7. PCA projected scores plot of the cetane index calibration group over the PCA of composition of all samples (Figure 5). Projected samples (circled green dots) are overlapped on the original samples (blue dots) with an identical crude composition.

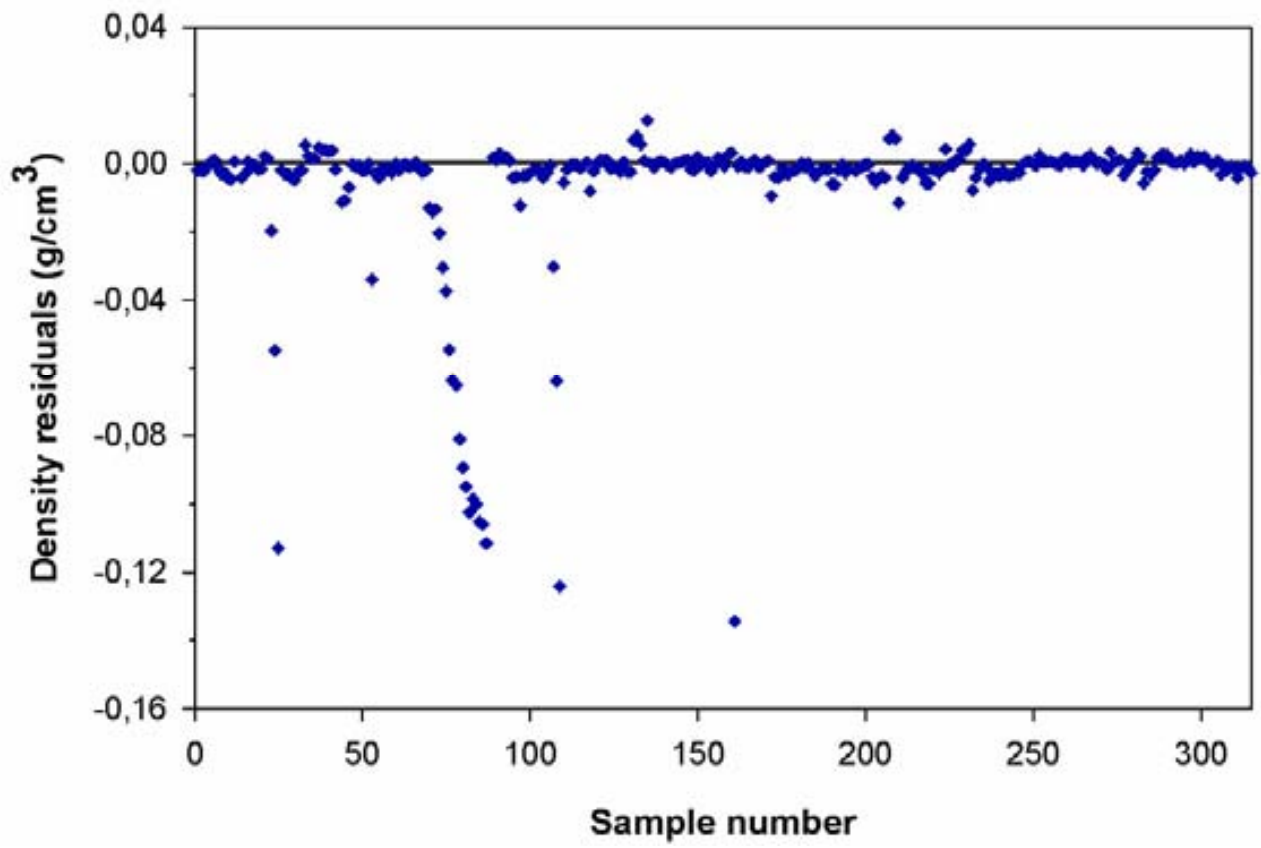


Figure 8. Calculated NIR residuals of density vs time.

TABLES

TABLE 1. Precision of the reference methods.

Parameter	Method	Repeatability (<i>r</i>)	Reproducibility (<i>R</i>)
Dens (g/cm³)	EN ISO 12185	0.0002	0.0005
CI	EN ISO 5165	1	4
FAME (% v/v)	EN 14078	0.3	1.4
CP (°C)	ASTM D 23015	2	4
D. 95% (°C)	ASTM D 86	2.1–2.9	4.3–6.3
FP (°C)	EN ISO 2719	1.62–2.03	3.98–4.97
S (mg/kg)	EN ISO 20846	0.66–1.16	1.34–2.35

TABLE 2. Parameter ranges spanned by the gas oil calibration and external validation samples.

	Dens (g/cm³)	CI	FAME (% v/v)	CP (°C)	D. 95% (°C)	FP (°C)	S (mg/kg)
Min.	0.8262	45.5	0	-12.4	336.5	56	2
Max.	0.8617	59.0	14.9	2.2	384.9	71	22

TABLE 3. Statistical figures of merit of the models constructed with The Unscrambler v. 9.7.

		Dens (g/cm ³)	CI	FAME (%)	CP (°C)	D. 95% (°C)	FP (°C)	S (mg/kg)
Spectral pretreatment		2 nd Der. + SNV	2 nd Der.	2 nd Der.	2 nd Der.	2 nd Der. + SNV	2 nd Der.	2 nd Der. + SNV
Calibration ^a	Factors	4	2	4	7	8	7	5
	Var-Y (%)	97.23	97.82	98.13	96.46	97.61	96.88	95.69
	RMSEC	0.0012	0.45	0.63	0.53	1.18	0.57	0.48
Validation (ext. set) ^a	Aver. res.	0.0000	-0.06	0.14	0.11	0.26	-0.01	0.04
	RMSEP	0.0016	1.02	0.67	1.22	5.39	3.72	2.56
Prediction ^b	Aver. res.	-0.0002	0.06	0.29	0.32	0.23	1.61	-0.66
	RMSEP	0.0016	0.7	0.51	1.15	4.03	2.90	1.88

^a Samples collected until March 2012

^b Samples collected from March to April 2012

TABLE 4. Statistical figures of merit of the CWS recalculated models.

		Dens (g/cm ³)	CI	FAME (%)	CP (°C)	D. 95% (°C)	FP (°C)	S (mg/kg)
Spectral pretreat.		2 nd Der. + SNV	2 nd Der.	2 nd Der.	2 nd Der.	2 nd Der. + SNV	2 nd Der.	2 nd Der. + SNV
Calibration^a	Factors	4	4	5	5	8	3	5
	Var-Y (%)	97.16	97.35	98.28	92.62	97.15	86.45	95.83
	RMSEC	0.0012	0.50	0.60	0.77	1.29	1.19	0.47
Validation (ext. set)^a	Aver. res.	0.0001	-0.03	0.00	0.04	1.05	-0.18	-0.08
	RMSEP	0.0016	1.02	0.52	1.39	5.30	3.68	2.59
Prediction^b	Aver. res.	-0.0004	0.09	0.34	0.66	1.86	1.84	-0.78
	RMSEP	0.0012	0.64	0.57	1.41	4.79	2.12	1.89

^a Samples collected until March 2012

^b Samples collected from March to April 2012

TABLE 5. Monitoring of the residual statistics of CWS predictions with samples collected from May 2012 to February 2013.

Statistic	Dens (g/cm³)	CI	FAME (%)	CP (°C)	D. 95% (°C)	CP (°C)	S (mg/kg)
Aver. res.	-0.0017	-0.13	0.71	0.17	0.031	1.34	-1.47
RMSEP	0.0090	1.24	1.80	2.52	5.20	3.58	2.94

AUTHOR INFORMATION

Corresponding author

* Tel./Fax: +34 93 581 4899. E-mail: manel.alcala@uab.cat

ACKNOWLEDGEMENTS

The authors are grateful to Spain's Ministry of Economy and Competitiveness (MINECO) for funding this research within the framework of Project CTQ2012-34392.

REFERENCES

- (1) UNE-EN 590:2009+A1:2010. *Automotive fuels – Diesel – Requirements and test methods*; Comité Européen de Normalisation (CEN): Brussels, Belgium, 2010.
- (2) ASTM Standard D975-12a, *Standard Specification for Diesel Fuel Oils*; American Standard for Testing and Materials (ASTM): West Conshohocken, PA, 2012.
- (3) BOE-A-2010-13704. Boletín Oficial del Estado, no. 215; Ministerio de Industria, Turismo y Comercio: Madrid, Spain, 2010. Available at: http://www.boe.es/diario_boe/txt.php?id=BOE-A-2010-13704
- (4) Monteiro, M. R.; Ambrozin, A. R. P.; Lião, L. M.; Ferreira, A. G. *Talanta* **2008**; *77*: 593–605.
- (5) Felizardo, P.; Menezes, J. C.; Neiva-Correia, M. J. *Process Analytical Technology Use in Biofuels Manufacturing*. In: Undey, C.; Low, D.; Menezes, J. C.; Koch, M. editors. *PAT Applied in Biopharmaceutical Process Development and*

Manufacturing: An Enabling Tool for Quality-by-Design; CRC Press: Boca Raton, FL, 2011.

- (6) Zanier-Szydłowski, N.; Quignard, A.; Baco, F.; Biguerd, H.; Carpot, L.; Wahl, F. *Oil Gas Sci. Technol.* **1999**, *54*, 463–472.
- (7) Aji, S.; Schildknecht-Szydłowski, N.; Faraj, A. *Oil Gas Sci. Technol.* **2004**, *59*, 303–321.
- (8) López-García, C.; Biguerd, H.; Marchal-George, N.; Schildknecht-Szydłowski, N. *Oil Gas Sci. Technol.* **2007**, *62*, 57-68.
- (9) Blanco, M.; Villaroya, I. *Trends Anal. Chem.* **2002**, *21*, 240–250.
- (10) Hidajat, K.; Chong, S. M. *J. Near Infrared Spectrosc.* **2000**, *8*, 53-59.
- (11) Balabin, R. M.; Safieva, R.Z. *Fuel* **2008**, *87*, 1096–1101.
- (12) Baptista, P.; Felizardo, P.; Menezes, J. C.; Neiva-Correia, M. J. *Talanta* **2008**, *77*, 144–151.
- (13) Baptista, P.; Felizardo, P.; Menezes, J. C.; Neiva-Correia, M. J. *Anal. Chim. Acta* **2008**, *607*, 153–159.
- (14) Macho, S.; Larrechi, M. S. *Trends Anal. Chem.* **2002**, *21*, 799–806.
- (15) ASTM Standard D86-99a^{E1}, *Standard Test Method for Distillation of Petroleum Products at Atmospheric Pressure*; American Standard for Testing and Materials (ASTM): West Conshohocken, PA, 1999.
- (16) Burns, D.A.; Ciurczak, E.W. *Handbook of Near Infrared Analysis*, 3rd ed.; Marcel Dekker: New York, 1992.
- (17) Brereton, R.G. *Applied Chemometrics for Scientists*; John Wiley & Sons, Ltd.: Chichester, U.K., 2003.
- (18) Haaland, D. M.; Thomas, E. V. *Anal. Chem.* **1988**, *60*, 1193–1202.
- (19) Kennard, R. W.; Stone, L. A. *Technometrics* **1969**, *11*, 137–148.

ANEXO II

Characterization of paraffin waxes composition towards industrial applications

Anna Palou, Jordi Cruz, Marcelo Blanco, Rafael Larraz, Juana Frontela,
César M. Bengoechea, Josep M. González, Manel Alcalà

Energy & Fuels (2014), 28(2), 956-963.

Characterization of the Composition of Paraffin Waxes on Industrial Applications

Anna Palou,[†] Jordi Cruz,[‡] Marcelo Blanco,[†] Rafael Larraz,[§] Juana Frontela,[§] César M. Bengoechea,[§] Josep M. González,[§] and Manel Alcalà^{*,†}

[†]Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

[‡]Escola Universitària Salesiana de Sarrià (EUSS), Passeig Sant Joan Bosco 74, 08017 Barcelona, Spain

[§]Compañía Española de Petróleos, S.A.U. (CEPSA), Campo de las Naciones, Avenida del Partenón 12, 28042 Madrid, Spain

ABSTRACT: The use of industrial paraffin depends upon the properties that are strongly influenced by the composition but also affected by the oil content. Because paraffin is a byproduct of lubricant oils in the petrochemical industry, there is inherently a certain amount of oil that is difficult to remove completely. The determination of the oil content in paraffin is described by the ASTM method, which involves the oil extraction with methyl ethyl ketone (MEK) at a very low temperature ($-32\text{ }^{\circ}\text{C}$). However, this method is slow and scarcely precise. In this work, we characterized the main components of industrial paraffin (de-oiled paraffin and extracted oil) by gas chromatography in combination with mass spectrometry (GC/MS) and nuclear magnetic resonance (NMR) spectroscopy. Also, we developed two methods for the determination of the oil content in paraffin. The first method measures total isoparaffin content by GC. The second method uses near-infrared (NIR) spectroscopy to quantify the content of MEK-removable oil. The results obtained with the NIR-based method were quite consistent with those of the official, extraction-based method. NIR spectroscopy therefore provides an effective alternative to the ASTM method with the added advantage of substantially greater expeditiousness and reproducibility.

1. INTRODUCTION

Industrial paraffin is a mixture of saturated hydrocarbons (alkanes) obtained as a byproduct of lubricant oils in the petrochemical industry. Paraffin constitute about 15% by weight of the crude and require isolation from the oil to avoid crystallization at low temperatures. Paraffin components are solid at room temperature (melting points of $50\text{--}70\text{ }^{\circ}\text{C}$), whereas oils must have a freezing point below $-10\text{ }^{\circ}\text{C}$ to be useful as lubricants.¹ Oil and paraffin are basically a mixture of saturated hydrocarbons. Therefore, the main chemical differences between them depend upon the length and type of these hydrocarbons. Paraffin consists of alkanes containing around 80–90% linear chains (*n*-paraffin) with 20–30 carbons,^{2,3} and oil is basically branched chains (isoparaffin).

Paraffin is recovered from the light/medium fractions of the vacuum distillate in the lubricant oil production process. Then, paraffin is processed to remove aromatic hydrocarbons and isolate oil from paraffin as efficiently as possible in a de-oiling/dewaxing process, which commonly involves solvent extraction (i.e., solvent dewaxing). The oil is extracted in an appropriate solvent, and the mixture is cooled to have the wax crystallize and precipitate. Usually, the solvent is a mixture of toluene, which dissolves the oil and paraffins and reduces the viscosity, and methyl ethyl ketone (MEK), which precipitates paraffins.^{4,5} The two phases obtained are separated using rotary filters *in vacuo*, and the resulting product contains 20–30% oil. Reducing the oil content to 0–1.5% by weight usually requires repeating the extraction process 3 times at increasing temperatures. The paraffins thus obtained occasionally exhibit a yellowish or brown color because of the presence of sulfur, oxygen, and nitrogen compounds and polycyclic aromatic hydrocarbons (PAHs), which are potentially toxic and carcinogenic.^{6,7} The

contents in these compounds depend upon the composition of the originating crude⁸ and fractionation section.

Paraffins are used mainly to prepare candles and related products but are also useful as a protective and water-proofing agent in the textile, paper, wood, and rubber industries, as a cream ingredient in the cosmetic industry, and as a paper-coating agent, a packaging adhesive, or a cake, fruit, and cheese preservative in the food industry. The paraffins obtained by solvent dewaxing are too dark for most of these uses and must be decolorized by clay treatment or percolated through bauxite.² The specifications for food applications are especially stringent and require other additional treatments to fulfill the requirements of the U.S. Food and Drug Administration (FDA) standards 21 CFR 172.886 for use in foods⁹ and 21 CFR 178.3710 for direct contact with foods¹⁰ or the European Pharmacopoeia applicable standards.¹¹

Ultimately, the uses of paraffins are dictated by a combination of their physical (melting or freezing point, hardness or penetrability, and viscosity), functional (impermeability, flexibility, and adhesiveness), and chemical properties (color, odor, light, and heat stability). Most of these properties are assessed by following routine procedures established by ASTM International.^{1,3} A difference between the properties of two paraffins is usually the result of one in chemical composition (*viz.* the origin and composition of the crude). Because refined paraffins consist of saturated hydrocarbons, their characteristics and potential uses essentially depend upon their proportion of linear and branched chains.

Received: November 4, 2013

Revised: January 3, 2014

Published: January 5, 2014

The contents in *n*-paraffins and isoparaffins can be determined in various ways. Some methods are based on the relationships of the different types of carbon chain with physical properties, such as melting and freezing points, viscosity or refractive index, and use of an abacus or a specific equation to correlate them.¹ Others use separation across molecular sieves to isolate *n*-paraffins^{12,13} or separation of adducts formed between *n*-paraffins and urea;¹⁴ still others use gas chromatography (GC) alone¹⁵ or in combination with mass spectrometry (GC/MS)^{12,16} to isolate and identify different types of hydrocarbons. These methods, which are slow, destructive, and irreproducible, are typically used to detect and determine *n*-paraffins present in crudes and responsible for clogging of pipes and equipment in petrochemical industries.

The presence of a high content of isoparaffin tends to be exudate from paraffin, resulting in an undesirable oily texture on the surface. Also, this high content of isoparaffin can negatively affect some properties (hardness, adhesiveness, consistency, flexibility, friction resistance, melting point, odor, or color). The quality and potential use of the paraffin depends upon the isoparaffin content. Because the de-oiling process is difficult to be completely achieved, the determination of oil in paraffin is mandatory for the characterization of industrial paraffin. The oil content of industrial paraffin is usually determined by extraction with MEK according to ASTM D721.¹⁷ The procedure is sluggish and poorly precise because it involves dissolution, crystallization, filtration, and various other operations at a low temperature. One alternative method uses nuclear magnetic resonance (NMR) spectroscopy on the assumption that the oil will contain a greater number of methyl groups than the linear chains in the paraffin.¹

In this work, we first characterized the main components (de-oiled paraffin and extracted oil) of industrial paraffin with a view to quantify further their oil content in an expeditious, accurate manner. De-oiled paraffin and oil were chemically characterized using GC/MS to determine the average length of the hydrocarbon chain^{13,16,18,19} and the proportions of *n*-paraffins and isoparaffins.^{8,12,16} Also, proton nuclear magnetic resonance (¹H NMR) spectroscopy was used to assess chain branching, and carbon nuclear magnetic resonance (¹³C NMR) spectroscopy was used to identify olefins and aromatic hydrocarbons.^{8,19} Second, we developed two alternatives to the ASTM analytical method for this purpose. One determines the total content in isoparaffins by GC,^{13,15} and the other quantifies MEK-removable oil by near-infrared (NIR) spectroscopy. This is a rapid, non-destructive technique requiring no sample preparation, which, however, is subject to a marked hindrance: the chemical and spectral similarity of the oil and de-oiled paraffin in paraffin samples.

2. EXPERIMENTAL SECTION

2.1. Samples. The samples were hydrogenated industrial paraffins obtained from oil crudes at the Compañía Española de Petróleos, S.A.U. (CEPSA) refinery in San Roque (Spain), containing small amounts of oil (<2%). We examined three different types of samples, namely, (A) original production samples, (B) de-oiled production samples, and (C) de-oiled samples doped with known amounts of oil in the laboratory. De-oiled paraffins were obtained in the same industrial process as all others and, subsequently, further de-oiled according to ASTM D721¹⁷ in the laboratory (see section 2.4). The process was repeated until no oil was extracted. Although the oil cannot be completely extracted, the resulting material was taken to be de-oiled paraffin.

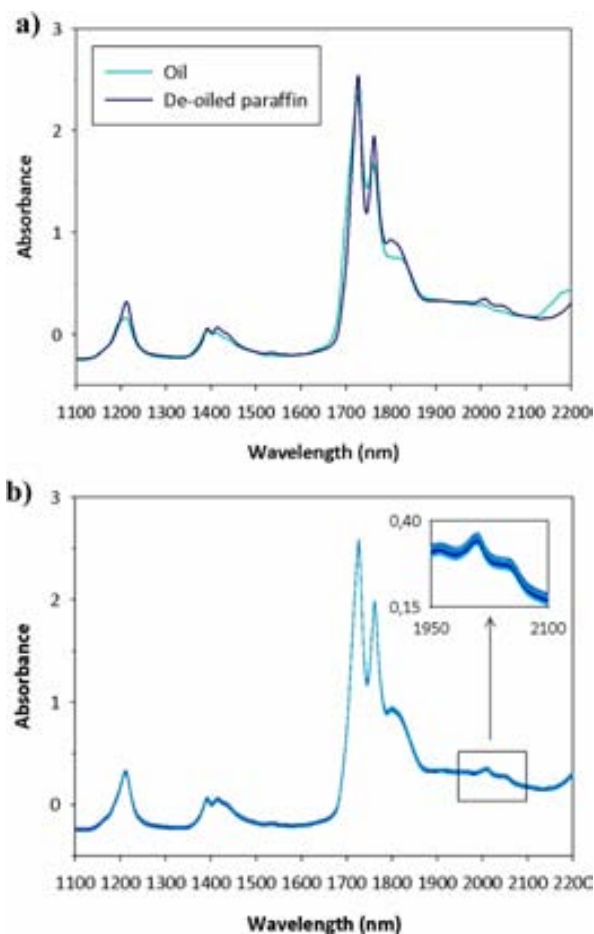


Figure 1. Absorbance spectra for (a) oil and de-oiled paraffin and (b) all paraffin samples (0.0–1.5% oil).

Extracted oil was used to prepare the de-oiled, doped samples (type C) with different oil concentrations. To this end, samples of de-oiled paraffin (type B) were diluted with small amounts of oil (6, 8, 10, and 12%), and the resulting solutions were mixed with de-oiled paraffins in preset proportions to obtain samples containing 0–1.5 wt % oil. Two different oils obtained from two also different crudes were used in the process.

One de-oiled paraffin (type B) sample and another sample of extracted oil were characterized by GC/MS (dissolved in hexane) and also by NMR spectroscopy (dissolved in deuterated chloroform). The GC technique was also applied to various paraffin samples of the A and B types containing a 2000 ppm (w/w) concentration in hexane. The samples contained 0.00–1.51% oil as determined by the ASTM D721 endorsed method.

NIR models were constructed using 43 undiluted samples of the B and C types. All were produced between 2012 and 2013 and contained oil in proportions of 0.00–1.50% that were measured by weighing.

2.2. GC/MS. GC/MS analyses were performed on a HP 6890 series II GC system equipped with a HP-5MS capillary column (5% phenyl methylpolysiloxane, 30 m long × 0.25 mm inner diameter, 0.25 μm film thickness) from Agilent Technologies. Aliquots of 2 μL in hexane were injected into the system at 310 °C, and a temperature gradient of 2 °C min⁻¹ from 150 to 320 °C was applied. Identification of compounds was facilitated by injecting a C₃₆H₇₄ standard under identical conditions. Compounds were detected using a HP 5973 mass selective detector interfaced to the gas chromatograph. The detector was operated in the electron ionization mode at 70 eV and in the chemical ionization mode with CH₄ as ionizing gas.

Chromatograms and mass spectra were acquired with the software Enhanced MSD ChemStation, version E.02.00.493, from Agilent Technologies, Inc. (Santa Clara, CA).

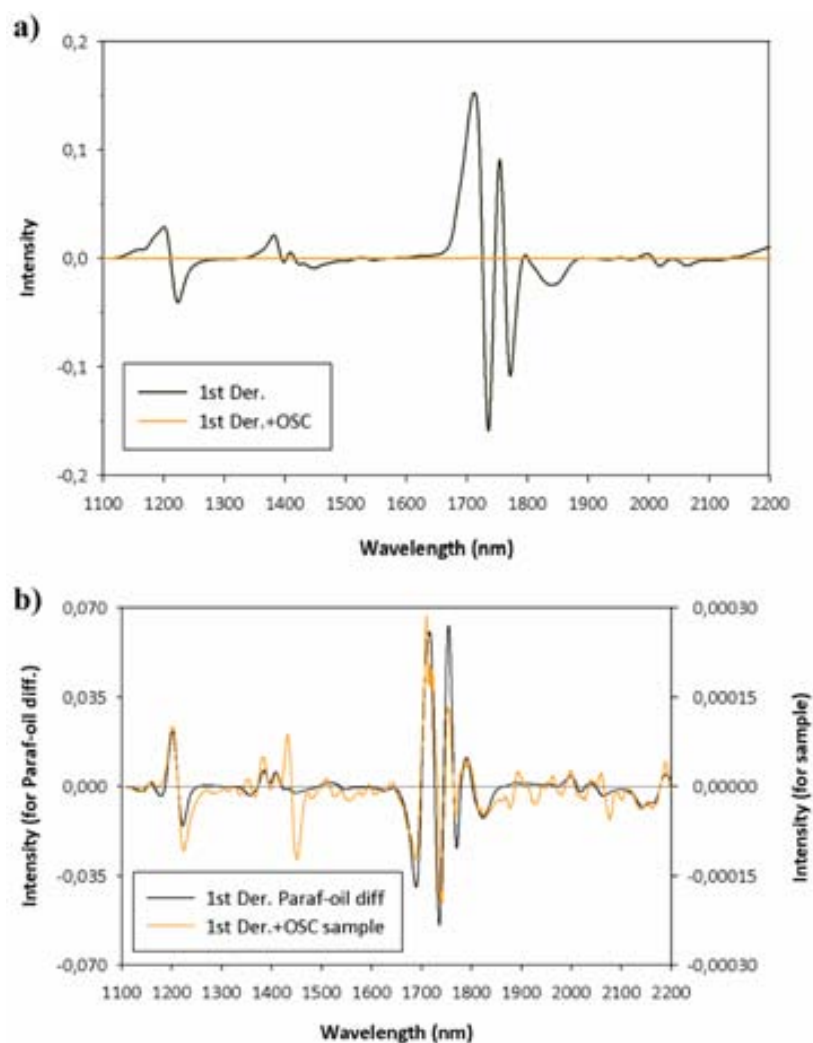


Figure 2. Effect of the OSC pretreatment: (a) first-derivative spectrum for a sample before and after OSC application and (b) first-derivative paraffin–oil difference spectrum and the sample after OSC application (rescaled).

2.3. NMR Spectroscopy. Following dilution in deuterated chloroform, samples were analyzed by ^1H and ^{13}C NMR spectroscopy on a Bruker spectrometer equipped with a $^1\text{H}/^{13}\text{C}/^{19}\text{F}/^{31}\text{P}$ quadruple nucleus probe (QNP) and operated at a frequency of 250 MHz. Each sample was scanned 85 times.

2.4. Reference Methods. The reference oil contents of the paraffin samples were obtained by analysis with two reference methods. Thus, the oil content of type A samples was determined according to ASTM D721,¹⁷ which involves dissolution in MEK and cooling at $-32\text{ }^\circ\text{C}$ to precipitate paraffins, followed by cold filtering and evaporation of the solvent to quantify oil in the resulting residue by weighing. This method does not allow for all oil in the samples but only MEK-extracted oil to be quantified; also, it is poorly accurate.

The reference oil contents of type C (doped laboratory) samples were determined by weighing the amount of oil added to de-oiled (type B) samples. Although the de-oiled samples were assigned a 0% (w/w) oil content, they contained a residual amount of oil, owing to its solubility in paraffins.

2.5. NIR Spectroscopy. **2.5.1. NIR Features.** NIR spectra were recorded on a FOSS NIRSystems 5000 spectrophotometer equipped with a vial heater module (VHM) to maintain samples at a constant temperature of $60\text{ }^\circ\text{C}$, where paraffins are liquid. Spectra were recorded in the transmittance mode, using $8 \times 43\text{ mm}$ cylindrical vials of 6.5 mm inner diameter. Empty vials were used for blank measurements.

Spectra were acquired with the aid of the software VISION, version 2.51, from FOSS NIRSystems. Each spectrum was the average of 32 scans performed at 2 nm intervals over the wavelength range of 1100–2500 nm. Wavelengths above 2200 nm gave saturated signals that were useless for calibration. Each sample was measured in triplicate and the average of the three spectra used for the construction and validation of the model.

2.5.2. Spectral Data Analysis. NIR spectra can contain unwanted contributions from physical properties of the sample, slight differences in recording conditions, or simply, instrumental noise. These contributions can be suppressed to facilitate construction of calibration models using an appropriate spectral pretreatment, such as derivatives or the standard normal variate (SNV), and the most suitable spectral range.

Spectral correlation between sample components (paraffin and oil) was very high (0.98 over the range of 1100–2200 nm; Figure 1a), and the oil contents of the samples were very low (<2%). As a result, the contribution of the oil in the spectra was very small (Figure 1b). Also, there were slight differences between de-oiled paraffin samples arising from differences in the originating crudes. This required suppressing or at least reducing any spectral contribution not a result of the oil affecting the accuracy of the calibration models. To this end, we used orthogonal signal correction (OSC),^{20,21} a pretreatment method specially devised for NIR spectroscopic data, to remove spectral information in matrix X unrelated (orthogonal) to matrix Y (the oil content matrix).

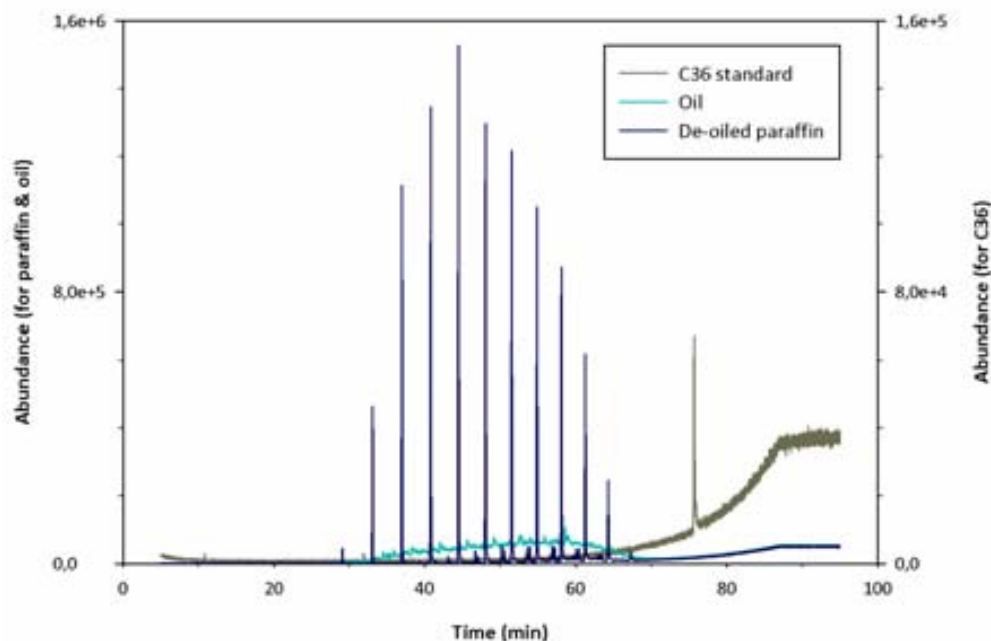


Figure 3. Superimposed GC chromatograms for $C_{36}H_{74}$ standard (right scale) and oil and de-oiled paraffin (left scale).

Derivatives and SNV and OSC spectral pretreatments were tested in the partial least-squares (PLS) calibration model construction. First-derivative and OSC models were found to be the most suitable. First-derivative spectra were obtained using the Savitzky–Golay algorithm with an 11 moving point window and a second-order polynomial. On the other hand, the OSC model was constructed from the whole calibration set and applied to all samples. The best results were obtained by applying OSC to first-derivative pretreated spectra. The calibration set used was identical to that for the PLS model, and its samples were selected as described in the following section. The best results were obtained with two OSC factors; this is consistent with previous reports:²⁰ the first component corrects baseline effects and, using a great number of components, increases the risk of overfitting. The results obtained with OSC are shown in Figure 2 and discussed in detail in section 3.5.

Both pretreatments were tested over various spectral ranges and applied using the software The Unscrambler (CAMO, Trondheim, Norway).

2.5.3. PLS Model. The quantitation of oil in paraffins was based on the use of regression models constructed using the PLS algorithm, which is widely documented.²² PLS models use a set of calibration samples with known reference values. The calibration set must span the whole range of the target quantity and be representative of both the training samples and those to be predicted; also, it should provide for any potential sources of spectral variability (crude origin, refining method, and sample type)²³ if a robust, adequately predictive PLS model is to be obtained. This led us to select the samples for inclusion in the calibration set by inspecting the scores plot for a principal component analysis (PCA) of the samples to use the whole spectral range and ensure that the reference values would span the entire working range.

Models were constructed by cross-validation, and their optimum number of factors was selected in terms of the root-mean-square error of prediction (RMSEP) as determined with The Unscrambler.

3. RESULTS AND DISCUSSION

3.1. GC/MS Electron Ionization. The first step to characterize the paraffin and oil samples involved using GC/MS in the electron ionization mode. Figure 3 shows the chromatograms for a sample of de-oiled paraffin and another of oil. As seen, the chromatogram for the paraffin exhibited well-

defined, sharp peaks at similar intervals that can be ascribed to *n*-paraffins (linear alkanes) and intervening, smaller peaks as a result of isoparaffins (branched alkanes). In contrast, the chromatogram for the oil exhibited a single, unresolved broad band, suggestive of a much more heterogeneous and complex mixture of hydrocarbons.

Because the elution intervals for both samples were very similar, the hydrocarbon chain lengths in both must also be very similar. This analytical technique does not allow one to determine chain lengths, owing to the high fragmentation and consequent absence of a molecular peak in the mass spectrum; the mass spectra for the different peaks were very similar. Figure 3 also shows the chromatogram for a *n*-paraffin standard ($C_{36}H_{74}$) obtained under identical conditions as those for the samples. The fact that the retention time for the standard exceeded those for the paraffins indicates that their chain length was smaller than 36 carbon atoms. The mass spectra for the chromatographic peaks of the paraffin and oil (results not shown) differed markedly; both, however, contained peaks 14 mass units apart between consecutive signals corresponding to successive losses of methylene groups and indicating the presence of saturated hydrocarbons.

3.2. GC/MS Chemical Ionization. The molecular weight (MW) of the paraffin components was determined by GC/MS in the chemical ionization mode, using CH_4 as ionizing gas to obtain $(M - H)^+$ ions. The chromatograms for the de-oiled paraffin and oil were identical to those obtained in the electron ionization mode but contained information about MW for each peak.

3.2.1. De-oiled Paraffin. The strongest peaks in the chromatograms for the paraffin gave mass spectra similar to that of Figure 4a, corresponding to the peak at ca. 51.5 min in Figure 3. As seen, all consecutive spectral signals exhibited 14 unit losses and the m/z ratio for the base peak ($M - 1$) was 393, which corresponds to an aliphatic hydrocarbon of 28 carbon atoms ($MW = 394 \text{ g mol}^{-1}$). The molecular weight of the base peak in the mass spectra for the chromatographic peaks of the paraffin, C_{36} standard included, exhibited a linear

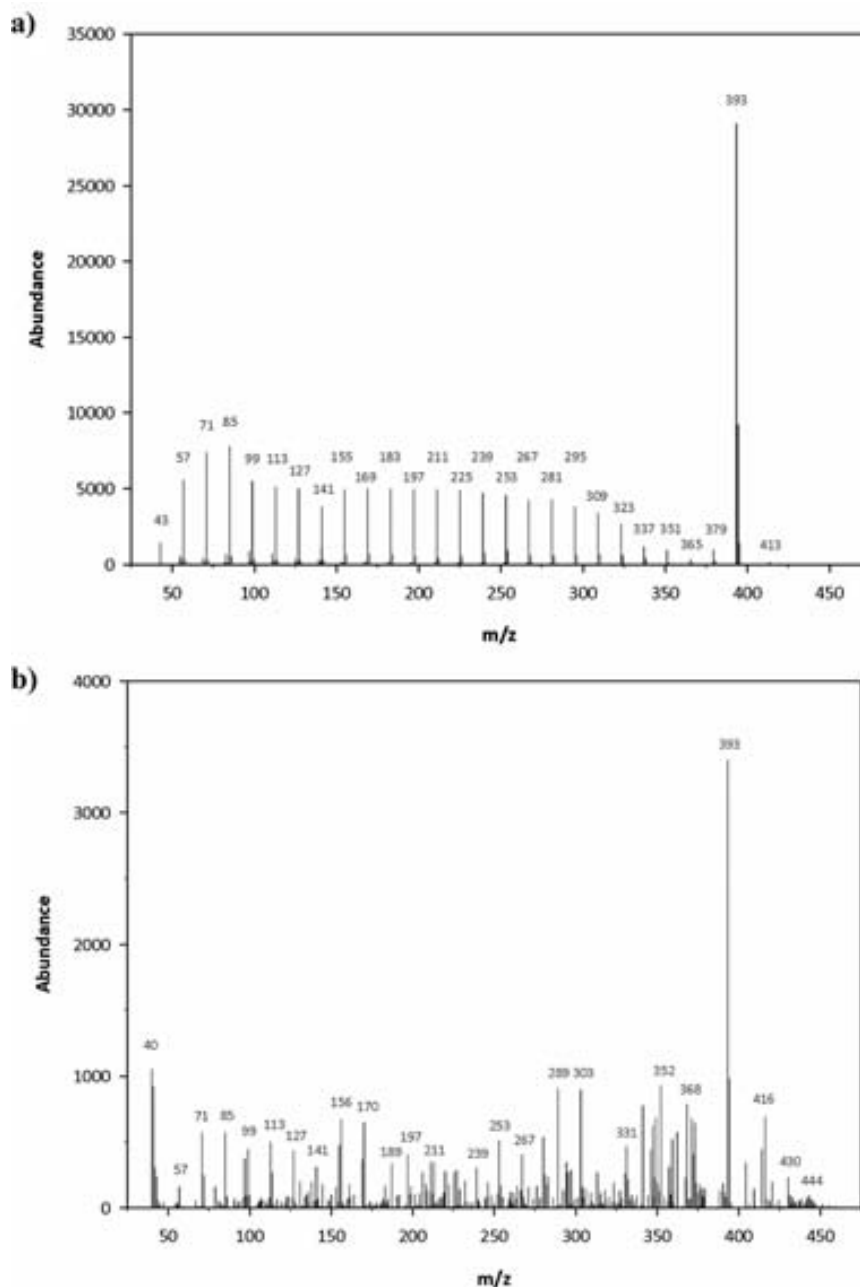


Figure 4. Chemical ionization mass spectra for the peak at ~51.5 min in the chromatograms for (a) paraffin and (b) oil.

relationship with the hydrocarbon retention times, as stated in references for the homologue series.^{24,25} Table 1 shows obtained data, and the obtained linear least-squares regressions for both experimental and theoretical data are $y = 0.251x - 48.149$ ($r^2 = 0.998$) and $y = 0.245x - 46.286$ ($r^2 = 0.997$), respectively. These results were used to establish the composition of the de-oiled paraffin samples (see Table 1).

The next step was to extract chromatograms at the mass of the base peak in the mass spectrum using the single ion monitoring (SIM) mode. By way of example, Figure 5 shows the chromatogram for the peak at m/z 393, similar to the others, which contains a major peak preceded by two smaller peaks. In the column used, aliphatic hydrocarbons of the same MW elute in the following sequence: isoparaffins > n -paraffins > cycloparaffins.^{13,16} One can therefore conclude that the two smaller signals on the left of the major peak correspond to

structural isomers of the compound (i.e., to isomeric aliphatic chains that can be assigned to isoparaffins). The weak peaks at the end of the chromatogram can be assigned to longer paraffin fragments with an identical MW.

3.2.2. Oil. As noted earlier, the chromatographic profile for the oil was poorly resolved and precluded recording mass spectra for individual compounds as a result. The mass spectrum (Figure 4b) also exhibited differences of 14 mass units between some signals, but the spectrum was complicated by the presence of many other mass values reflecting the high chemical complexity (increased isomerism) of the sample.

The chromatogram for a SIM-selected mass peak exhibited an unresolved signal much broader than that for the paraffins. The retention time of the chromatogram for the peak at m/z 393 (Figure 5) was slightly shorter for the oil than for the paraffins; this, together with the broad elution band obtained,

Table 1. Retention Times and Carbon Number Ratios of the Chromatographic Peaks for the De-oiled Paraffin Samples and C₃₆H₇₄ Standard (Last Row)

R _t (min) ^a	experimental <i>m/z</i> ^b	theoretical MW	C atoms	H atoms
24.7		296	21	44
28.7	309	310	22	46
32.7	323	324	23	48
36.6	337	338	24	50
40.4	351	352	25	52
44.1	365	366	26	54
47.7	379	380	27	56
51.2	393	394	28	58
54.5	407	408	29	60
57.6	421	422	30	62
60.9	435	436	31	64
63.9	449	450	32	66
66.8		464	33	68
69.7		478	34	70
72.5		492	35	72
75.7		496	36	74

^aSee Figure 3. ^bObtained from MS spectra.

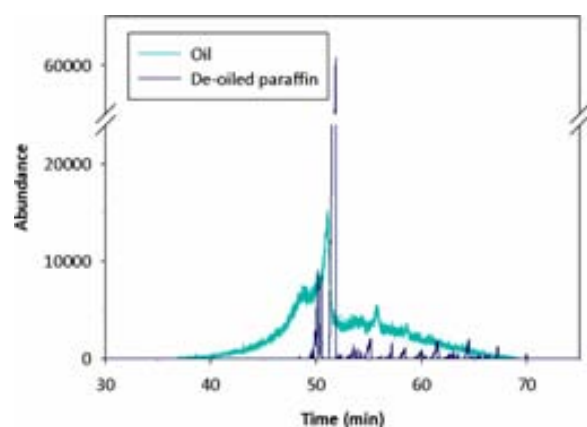


Figure 5. Chromatograms for the peak at *m/z* 393 in the mass spectra for oil (Figure 4b) and de-oiled paraffin (Figure 4a).

confirms that the oil consisted of a complex mixture of isomers of branched saturated hydrocarbons (isoparaffins).

3.3. NMR Spectroscopy. De-oiled paraffin and oil samples were also examined by ¹H NMR spectroscopy (Figure 6) to identify the types of hydrocarbons present and their relative contents. The spectrum for the paraffin revealed the presence of two types of protons corresponding to CH₂ groups ($\delta = \sim 1.25$ ppm) and CH₃ groups ($\delta = \sim 0.9$ ppm).^{8,26} Integrating the two signals provided a ratio of 8.82 between them, which is consistent with an average linear chain length of 28 carbons; the actual value might be slightly greater, owing to the presence of isoparaffins in the paraffin sample.

The CH₂/CH₃ signal ratio in the ¹H NMR spectrum for the oil was 1.96. For a 28 carbon chain, such a ratio corresponds to 21 CH₂ and CH groups, which are overlapped in the spectrum, and 7 CH₃ groups and reflects highly branched hydrocarbon chains in the oil samples.

Neither spectrum contained the typical signals for aromatic protons (6.5–9.2 ppm^{19,26}), which confirms the efficient removal of aromatic hydrocarbons in the refining process.

The ¹³C NMR spectra also recorded allowed for the types of carbon atoms present in the samples, the length of hydrocarbon

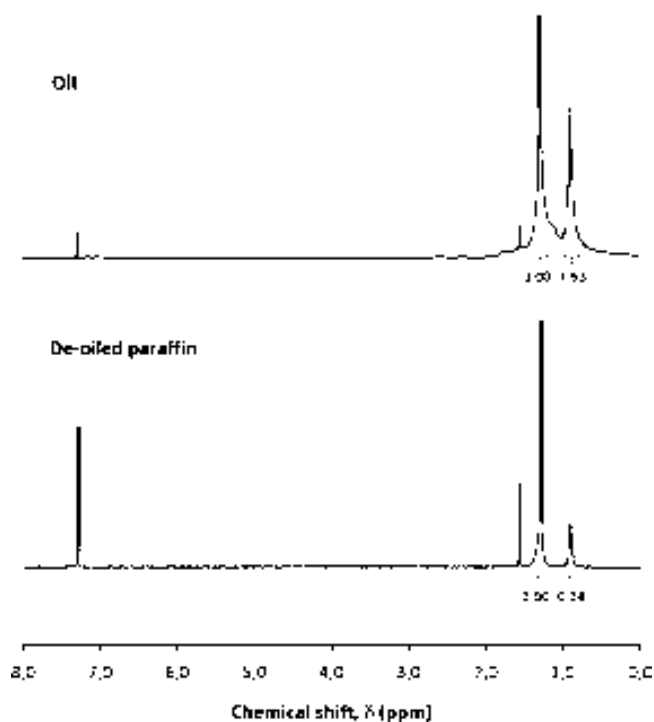


Figure 6. ¹H NMR spectra with peak integration results from oil (above) and de-oiled paraffin (below). In both cases, the peak at ~ 7.25 ppm can be assigned to the solvent deuterated chloroform.

chains, and their degree of branching to be assessed. On the basis of them, the samples contained saturated chains only, with no sign of olefins or aromatics.

3.4. Analysis of Paraffins. A GC analysis of the paraffin samples based on the linear relationship between MWs and retention times revealed that the maximum number of carbon atoms in the alkane chains was 21–35 but mostly ranged from 22 to 33, with differences ascribed to the origin of the crude. The peaks for the linear alkanes exhibited a near-Gaussian distribution, with the C₂₆ compound as the most abundant. If an identical response from each linear alkane is assumed, then the area under each peak should be proportional to the content of the compound concerned and the average MW of the paraffin samples should slightly exceed that for the C₂₇ alkane (Table 2).

No analytical methods for determining oil in paraffins, except those of the ASTM standard, have been developed for industrial products. In this work, we developed one approximation based on the area ratio for two components: branched alkanes and linear alkanes. Because most isoparaffins in purified paraffins are a part of the oil portion of paraffins, we can consider all branched alkanes in the sample as oil content. If one assumes the detector response to be proportional to the concentration of each component, then the ratio between the two areas will coincide with the relative proportions of the components of industrial paraffin. The results thus obtained for type A and B samples (Table 2) are consistent with available knowledge. Thus, type A samples contained 10.4–10.9% oil (average of 10.7%), and laboratory de-oiled (type B) samples contained 7.8–8.9% oil (average of 8.3%). Such wide oil concentration ranges are a result of the variability inherent in the oil removal procedure (ASTM D721) rather than differences in the distribution of *n*-alkanes by the effect of the different origins of industrial paraffins.

Table 2. Summary of the Results for Various Paraffin Samples as Obtained from Chromatographic Data

sample type	<i>n</i>	oil content (%)		isoparaffin content (%) ^a		C atoms in alkane	
		reference method (ASTM D721)	chromatographic peak integration	maximum range	average range for samples		
B	6	0.0	7.8–8.9	22–34	27.3–27.5		
A	5	0.3–1.5	10.4–10.9	21–35	27.2–27.4		

^aThe remainder corresponds to *n*-paraffins.

Table 3. Figures of Merit and Statistics of the PLS Calibration Models

PLS models		1	2	3	4
PLS model features	spectral pretreatment	first derivative		first derivative + OSC (2F)	
	spectral range (nm)	1100–2200	2000–2200	1100–2200	1800–2200
	PLS factors	8	6	7	5
	variance of <i>Y</i> (%)	98.68	97.11	98.07	98.81
calibration (B + C) (0.00–1.50% oil)	RMSEC	0.05	0.08	0.07	0.05
external validation (B + C) (0.10–1.50% oil)	RMSEP	0.19	0.15	0.18	0.14
	bias	0.05	0.00	0.06	0.03
prediction (A) (0.10–0.49% oil)	RMSEP	0.22	0.82	0.56	0.79
	bias	0.20	0.52	0.44	0.44

3.5. NIR Determination. The specific uses and applications of industrial paraffins are established from the parameters defined in ASTM standards. In this work, we developed an alternative method to ASTM D721 based on NIR spectroscopy for determining residual oil from the industrial process.

The method uses PLS regression models to predict the oil content of industrial (type A) samples. The oil concentration range spanned by the industrial samples used was inadequately wide to construct a robust enough model and required preparing doped (type C) laboratory samples, as described in section 2.1, to expand the operating range to concentrations up to 1.5%.

Type C and de-oiled (type B) samples were used for calibration and validation. Samples were selected, and PLS models were constructed and assessed, as described in section 2.5.3. The calibration set consisted of 27 samples, and the validation set consisted of the remaining 16 samples. The industrial (type C) samples were used for prediction.

The best results were obtained using first-derivative and first-derivative + OSC models (see section 2.5.2) over various spectral ranges. As noted earlier, OSC allowed for information exclusively related to the target parameter (oil content) to be extracted. Because the spectra for the sample components were highly correlated, the OSC model eliminated most of matrix *X* and retained very little information. This is quite apparent from Figure 2a, which shows the first-derivative spectrum obtained before and after applying the OSC pretreatment, both on the same intensity scale. Figure 2b shows the spectrum after application of OSC but rescaled; as seen, the strongest signals coincided with those exhibiting the greatest differences between the paraffin and oil spectra. On the basis of the results of the OSC model, the most suitable zones for modeling the oil content were not the most similar to the oil spectrum but rather those exhibiting the greatest differences between the paraffin and oil spectra.

Table 3 shows the characteristics and statistics of the two best models (regarding the shown statistics) obtained with each spectral pretreatment. One model in each pair was constructed from data of the whole spectral range (1100–2200 nm), and the other was constructed using a restricted range. The variance explained by all individual models, which used 5–8 factors, exceeded 97%. Applying the OSC pretreatment reduced the

number of factors needed for optimal results and, hence, enabled the construction of simpler PLS models. Restricting the wavelength range used had the same effect, regardless of the spectral pretreatment. The optimum range in both cases was one at the end of the NIR spectral zone (1800–2200 nm).

Predictions for the samples in the calibration set [root-mean-square error of calibration (RMSEC)] and external validation set [RMSEP (B + C)] were all good, and their statistics were comparable. RMSEP was slightly greater for the industrial (type A) samples. The fact that the difference in bias was more marked suggests that the extraction method and its decreased reproducibility (0.23) have a decisive impact on the quality of the results.

Simplifying the starting models by restricting the spectral range and/or applying OSC resulted in no significantly better results; rather, it increased RMSEP and bias to some extent. Despite its using a greater number of factors, the PLS 1 model was the simplest with regard to spectral pretreatment (first derivative with the whole wavelength range); also, it exhibited the highest predictive ability for industrial paraffin samples.

4. CONCLUSION

Industrial paraffin and oil samples were thoroughly characterized using GC/MS and NMR spectroscopy, and two new methods for determining the oil content of paraffins were developed. On the basis of the results, both types of samples (de-oiled paraffin and oil) consist of linear saturated hydrocarbons (*n*-paraffins). The paraffins have a chain length of 21–35 carbon atoms and an average molecular weight corresponding to a C₂₇ hydrocarbon, whereas the oil consists of a mixture of highly isomeric branched hydrocarbons (isoparaffins) containing a number of carbon atoms similar to that for paraffins. The GC/MS resolution achieved allows for the total oil content of paraffins to be easily calculated by integrating the corresponding peaks.

The other method uses NIR spectroscopy to construct PLS models from the spectra for samples prepared by doping de-oiled paraffin samples. This method provides results on par with those of the officially endorsed method (ASTM D721) and, hence, an advantageous alternative by virtue of its greater expeditiousness and reproducibility.

AUTHOR INFORMATION

Corresponding Author

*Telephone/Fax: +34-93-581-4899. E-mail: manel.alcala@uab.cat.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are grateful to Spain's Ministry of Economy and Competitiveness (MINECO) for funding this research within the framework of Project CTQ2012-34392.

REFERENCES

- (1) Sánchez, J. Purification of paraffin oil by catalytic hydrogenation. Ph.D. Thesis, Universidad Complutense de Madrid, Madrid, Spain, 2003.
- (2) Waxes. *Kirk–Othmer Encyclopedia of Chemical Technology*; John Wiley and Sons, Inc.: Hoboken, NJ, Dec 4, 2000; <http://onlinelibrary.wiley.com/doi/10.1002/0471238961.2301240503152020.a01/pdf> (accessed Oct 7, 2013).
- (3) The International Group, Inc. (IGI). <http://igiwax.com/reference/waxbasics.html> (accessed Oct 7, 2013).
- (4) Petroleum refinery processes. *Kirk–Othmer Encyclopedia of Chemical Technology*; John Wiley and Sons, Inc.: Hoboken, NJ, Aug 19, 2005; <http://onlinelibrary.wiley.com/doi/10.1002/0471238961.1805060919160509.a01.pub2/pdf> (accessed Oct 7, 2013).
- (5) SET Laboratories, Inc. <http://www.setlaboratories.com/solvent/tabid/108/Default.aspx> (accessed Oct 7, 2013).
- (6) Agency for Toxic Substances Disease Registry (ATSDR), U.S. Department of Health and Human Services. *Toxicological Profile for Polycyclic Aromatic Hydrocarbons (PAHs)*; ATSDR, U.S. Department of Health and Human Services: Atlanta, GA, 1995.
- (7) Woo, O. T.; Chung, W. K.; Wong, K. H.; Chow, A. T.; Wong, P. K. *J. Hazard. Mater.* **2009**, *168*, 1192–1199.
- (8) Musser, B. J.; Kilpatrick, P. K. *Energy Fuels* **1998**, *12*, 715–725.
- (9) U.S. Food and Drug Administration (FDA). Food additives permitted for direct addition to food for human consumption. *Code of Federal Regulations*; FDA: Silver Spring, MD, 2013; Part 172, Title 21, Vol. 3.
- (10) U.S. Food and Drug Administration (FDA). Indirect food additives: Adjuvants, production aids, and sanitizers. *Code of Federal Regulations*; FDA: Silver Spring, MD, 2013; Part 178, Title 21, Vol. 3.
- (11) European Directorate for the Quality of Medicines and HealthCare (EDQM) Council of Europe. *European Pharmacopoeia*, 7th ed.; EDQM Council of Europe: Strasbourg, France, 2011.
- (12) O'Connor, J. G.; Norris, M. S. *Anal. Chem.* **1960**, *32*, 701–706.
- (13) O'Connor, J. G.; Burow, F. H.; Norris, M. S. *Anal. Chem.* **1962**, *34*, 83–85.
- (14) Zimmerschild, W. J.; Dinerstein, R. A.; Weitkamp, A. W.; Marschner, R. F. *Ind. Eng. Chem.* **1950**, *42*, 1300–1306.
- (15) Coto, B.; Coutinho, J. A. P.; Martos, C.; Robustillo, M. D.; Espada, J. J.; Peña, J. L. *Energy Fuels* **2011**, *25*, 1153–1160.
- (16) Levy, E. J.; Doyle, R. R.; Brown, R. A.; Melpolder, F. W. *Anal. Chem.* **1961**, *33*, 698–704.
- (17) ASTM International. *ASTM Standard D721-06: Standard Test Method for Oil Content of Petroleum Waxes*; ASTM International: West Conshohocken, PA, 2006.
- (18) ASTM International. *ASTM D5442-93, Standard Test Method for Analysis of Petroleum Waxes by Gas Chromatography*; ASTM International: West Conshohocken, PA, 2008.
- (19) Espada, J. J.; Coutinho, J. A. P.; Peña, J. L. *Energy Fuels* **2010**, *24*, 1837–1843.
- (20) Wold, S.; Antti, H.; Lindgren, F.; Öhman, J. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175–185.
- (21) Fearn, T. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 47–52.
- (22) Brereton, R. G. *Applied Chemometrics for Scientists*; John Wiley and Sons, Inc.: Hoboken, NJ, 2003.

(23) Macho, S.; Larrechi, M. S. *TrAC, Trends Anal. Chem.* **2002**, *21*, 799–806.

(24) Destailats, H.; Spaulding, R. S.; Charles, M. J. *Environ. Sci. Technol.* **2002**, *36*, 2227–2235.

(25) Craig, B. M.; Murty, N. L. *Can. J. Chem.* **1958**, *38*, 1297–1301.

(26) Pretsch, E.; Bühlmann, P.; Affolter, C.; Herrera, A.; Martínez, R. *Determinación Estructural de Compuestos Orgánicos*; Masson: Barcelona, Spain, 2005.

ANEXO III

Determination of drug, excipients and coating distribution in pharmaceutical tablets using NIR-CI

Anna Palou, Jordi Cruz, Marcelo Blanco, Jaume Tomàs, Joaquín de los
Ríos, Manel Alcalà

Journal of Pharmaceutical Analysis (2012), 2(2), 90-97.



ORIGINAL ARTICLE

Determination of drug, excipients and coating distribution in pharmaceutical tablets using NIR-CI

Anna Palou^a, Jordi Cruz^b, Marcelo Blanco^a, Jaume Tomàs^c,
Joaquín de los Ríos^c, Manel Alcalà^{a,*}

^aUnitat de Química Analítica, Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

^bEscola Universitària Salessiana de Sarrià, Passeig Sant Joan Bosco, 74, 08017 Barcelona, Spain

^cEsteve Farma, c/Sant Martí s/n, 08197 Martorelles, Barcelona, Spain

Received 7 October 2011; accepted 15 November 2011

Available online 22 November 2011

KEYWORDS

Near infrared Chemical Imaging (NIR-CI);
Hyperspectral imaging;
Component distribution;
Tablet coating distribution;
Partial Least Squares (PLS) regression

Abstract The growing interest of the pharmaceutical industry in Near Infrared-Chemical Imaging (NIR-CI) is a result of its high usefulness for quality control analyses of drugs throughout their production process (particularly of its non-destructive nature and expeditious data acquisition). In this work, the concentration and distribution of the major and minor components of pharmaceutical tablets are determined and the spatial distribution from the internal and external sides has been obtained. In addition, the same NIR-CI allowed the coating thickness and its surface distribution to be quantified. Images were processed to extract the target data and calibration models constructed using the Partial Least Squares (PLS) algorithms. The concentrations of Active Pharmaceutical Ingredient (API) and excipients obtained for uncoated cores were essentially identical to the nominal values of the pharmaceutical formulation. But the predictive ability of the calibration models applied to the coated tablets decreased as the coating thickness increased.

© 2011 Xi'an Jiaotong University. Production and hosting by Elsevier B.V. All rights reserved.

*Corresponding author. Tel./fax: +34 935814899.

E-mail address: manel.alcala@uab.cat (M. Alcalà)



1. Introduction

Near Infrared-Chemical Imaging (NIR-CI) is a technique based on conventional NIR spectroscopy with the added advantage that it affords recording of a large amount of both spectral and spatial information in a single image. While the conventional NIR technique only provides the average spectrum for the surface of each sample, NIR-CI gives one spectrum per pixel in each acquired image and hence much more information about the whole sample surface [1]. The easiness to obtain this information through with NIR-CI has aroused interest in NIR-CI in many fields of study. This is

particularly applicable in the analysis of pharmaceutical products, which is made easy by NIR-CI by virtue of its ability to acquire a vast amount of information in an expeditious manner, all without altering the sample. The increasing interest aroused by NIR-CI in the pharmaceutical field is evident from the variety of studies based on this technique reported in recent years. Such studies include the homogeneity of powder samples [2], particle size determinations [3], product composition [4], the determination of the concentrations and distribution of components in solid tablets [5] and content uniformity [6], among others. Also, the many uses of NIR-CI in the pharmaceutical industry have been the subject of two interesting reviews [1,7].

Film coatings on commercial tablets have been the subject of some study [8,9], but not so much as pharmaceutical components despite their significance. The coating film applied to a drug tablet is primarily intended to improve its esthetics and function. Thus, the coating allows the unpleasant taste of some APIs to be masked and facilitates swallowing of tablets, in addition to preserving their integrity and giving them a uniform appearance. Some coatings, however, are intended to facilitate the controlled or enteric release of the dosage form of a drug [10]. In any case, the exact function of a coating depends largely on its thickness and distribution on the tablet surface. In fact, too thin or too thick a coating can alter the effectiveness of a tablet by making the core more vulnerable to external factors or diminishing the API release rate. Simply measuring the average thickness of the lacquer film on a tablet is inadequate to ensure that the tablet will meet the specifications; rather, this requires obtaining more information about the way the coating is distributed throughout the tablet surface. The capabilities of the NIR-CI technique in this respect make it a suitable choice for quality control in tablet production processes.

However, the vast amount of information contained in a hyperspectral image requires the use of an effective procedure to extract it. In fact, hyperspectral data form a three-way cube that must be unfolded and processed with appropriate two-way multivariate algorithms in order to extract the target information.

In those cases where some components of a pharmaceutical are unknown or a spectrum unavailable, algorithms requiring no calibration set are especially useful; such algorithms include Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) [11] and its augmented form [12]. However, the process involved in extracting information from chemical images can be slow and cumbersome. One alternative choice in wide use for the quantitative analysis of CI data is Partial Least Squares (PLS) [13]. The Isys software included with the NIR-CI instrument allows the easy application of PLS for a rapid and accurate determination of components distribution in a sample. This algorithm requires a series of data distributed over a wide enough range of specific values of the target property to allow the construction of a suitable calibration model for the samples to be analyzed. Each pixel of the image is assigned to a response matrix based on the similarity of its spectra with those for the pure components [14].

In this work we intend to evaluate the distribution of the components in a tablet and quantify the thickness and surface distribution of the film coating during the development of a new pharmaceutical formulation. Individual calibration models for the components were used to obtain their distribution

map and a calibration model derived from the NIR-CI hyperspectral image was used to establish the distribution map for the coating film.

2. Materials and methods

2.1. Samples

We studied three different types of samples, namely: (a) cores (uncoated tablets), (b) tablets with a coating of variable thickness and (c) tablets with the standard coating. The coated tablets (b) included samples with film thicknesses between $0x$ and $3x$, where x is considered the standard coating thickness of tablets (c).

We consider the coating process is constant and reproducible along time. The standard time encompasses a coating thickness namely x . The double or triple of this standard time are named $2x$ and $3x$. The reference coating thickness has been calculated measuring the approximate tablet surface (1.92 cm^2), the tablet weight difference before and after coating (from 290 mg to 317 mg) and the real density of the coating polymer (1.19 g/cm^3). This measurement was $39.4 \mu\text{m}$ of coating thickness for a x standard coating time, $21.9 \mu\text{m}$ for $0.5x$, $78.8 \mu\text{m}$ for $2x$ and $118.2 \mu\text{m}$ for $3x$. The estimated measurement for a standard x coating time ($39.4 \mu\text{m}$) was quite similar to the coating thickness obtained by scanning electron microscopy (SEM, $40 \mu\text{m}$). However, we could not compare the calculated thickness and the SEM results for all the coating levels. In this work, all the coating thickness results are referred to the x standard coating time instead of the microns measurements, in order to simplify and better understand the conclusions.

The uncoated cores had the same API and excipient composition, namely: 35% API and 65% excipients. The excipient mixture contained 40% of excipient #1, 20% of excipient #2, 2% of excipient #3, 1.5% of excipient #4 and 1.5% of excipient #5. The API and the two major excipients in combination accounted for 95% of the tablet content. The presence of the lacquer in the coated tablets reduced the proportions of API and excipients to an extent dependent on the coating thickness. Table 1 shows the resulting changes in the major components.

The samples in groups *a* and *b* were used to determine the concentrations of the tablet components (API and excipients).

Table 1 Composition (% w/w) of API and the two major excipients at different coating levels.

Coating level	Sample composition (% w/w)		
	API	E #1	E #2
0x	35.0	40.0	20.0
0.5x	34.4	39.3	19.7
1x	33.9	38.8	19.4
2x	33.0	37.7	18.7
3x	32.0	36.6	18.3

0x means uncoated core and 1x means nominal standard coating.

Those in group *b* were used as calibration and validation sets for the PLS models employed to characterize the coating thickness, which were subsequently applied to the group *c* tablets.

2.2. Instrumentation

Images were acquired with a SyNIRgyTM Chemical Imaging System hyperspectral camera from Malvern Instruments (Malvern, UK). The camera was equipped with an InSb detector with a focal plane array of 320×256 pixels and controlled via the software Pixys[®] 1.1., also from Malvern. Samples were placed on the plate and lighted with halogen lamps arranged at appropriate angles around the plate, the lens distance being adjusted according to the object size. The images obtained under the selected conditions had a resolution of $40 \mu\text{m}$ (each pixel covered an area of $40 \mu\text{m} \times 40 \mu\text{m}$). The acquisition time for a reflectance image spanning the wavelength range 1200–2400 nm was about 3 min. Prior to acquisition of the image for the sample (*S*), the camera was used to record one for the background (*B*, a reference ceramic plate of 99% reflectance) and another for a dark reference (*D*, a mirror). The image for the sample, *S*, was converted into reflectance data, *R*, using the equation $R = (S - D)/(B - D)$. This calculation was done at every pixel in each image. The data thus obtained were then converted into absorbance units.

Each core was used to record 6 different images: 4 of each outer side (A–D) and 2 of the interior sides (E–F) which were established by breaking the tablets as shown in Fig. 1a. The analysis of the interior sides was performed to check the homogeneity of the samples. The analytical method does not require breaking the samples as it is considered nondestructive. The tablets group *c* (standard *x* coated) were used to study the four outer sides only, which were designated identically with those of the cores (A–D). On the other hand, each of the two tablets in sample group *b* was used to record the images, one from each of two opposite outer sides. Fig. 1b shows the image of two tablets with identical characteristics (coating thickness). The images were not studied in their entirety, but rather in selected rectangular areas on the inner sides of each tablet in order to avoid image distortions caused by light impinging on tablet edges.

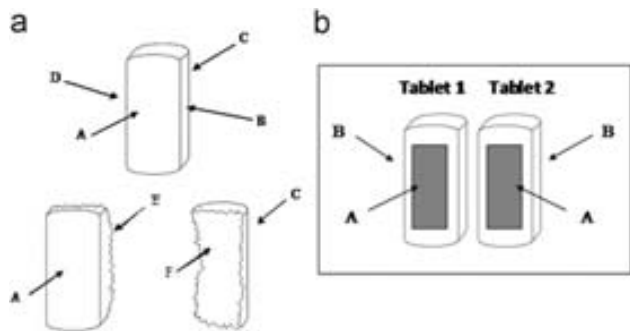


Figure 1 Schema of the tablet sides analyzed by hyperspectral imaging: (a) cores and production tablets and (b) tablets with different coating times. The selected area for analysis of samples (B) is highlighted with dark gray color.

The images for the pure compounds were obtained by placing an adequate powder amount of each in a glass cell and gently pressing its surface prior to recording.

2.3. Data processing

The data in a hyperspectral image are arranged in a three-way data cube $X(M \times N \times \lambda)$ two dimensions in which ($M \times N$) hold spatial information while the third (λ) contains spectral information.

Prior to application of any spectral treatment or algorithm to extract the required information, the data cube must be unfolded into a two-dimensional matrix since most existing treatments and algorithms have been developed for two-way data. The data matrices thus obtained here from each image were processed with the Standard Normal Variate (SNV) algorithm. All spectral quantitative treatments used were applied using the software Isys 5.0, from Malvern, with routines developed in MATLAB code (MATLAB v 7.0, The MathWorks, Natick, MA). This software only required the previous acquisition of the pure spectrum for each component to be determined [15]. The pure spectra were used to construct a library and then quantify the concentration of each component. The calibration model was completed by selecting an appropriate thresholds and number of factors. Using an increased number of factors increased the variance *Y* explained by the model, but also the risk of overfitting; a compromise was therefore needed for an optimal choice. The threshold used, which allows one to adjust the limits for discrimination between classes, is a function of the similarity between the spectrum for each pure component to be quantified and those for the other components in addition to the number of factors previously selected. Application of the resulting model weighted the spectral data in image based on a 0 to 1 scale. This scale corresponds from 0 to 100% concentration of each component at every pixel of the image. The result was delivered as a concentration map and mean concentration for each component present in the library. This algorithm has been applied for the calculation of the calibration models for every component of the formulation. The information provided by Malvern about the PLS algorithm included in the Isys software is minimum. However we understand this algorithm is a variation of the discriminant PLS, that allows to individually quantify every component of the sample only using its pure spectrum without the need of the other spectra (it does not act as the classical least squares algorithm that strictly needs all the pure spectra components of a formulation).

The thickness and distribution of the coating were quantified using PLS calibration models calculated with the software The Unscrambler v. 9.8 from CAMO (Trondheim, Norway). The PLS regression algorithm [15] uses as *X*-spectral data the mean spectrum of the selected area of every image. The *Y*-variable corresponded to the thickness reference value (Table 1). The PLS model thus constructed was validated by cross-validation applying jack-knifing algorithm in order to identify the wavelength range most strongly correlated with such property. Then, the optimum number of factors was selected under the above-described criteria. The analysis of a NIR-CI image by this PLS model was performed after the unfolding of the three-way data cube into a two-way data set.

The distribution map of tablet coating was performed refolding the two-way data set of PLS prediction results into a three-way data cube.

3. Results and discussion

The goal of this study is the application of NIR-CI to better understand the API and excipients distribution of a novel pharmaceutical formulation and eventually feedback this information to the development stages. Quantitative methods for the determination of tablet components and their distribution in the core and its coating from NIR-CI image have been developed. Fig. 2 shows the spectra for the different pure components of the studied pharmaceutical (viz. the API and 5 excipients). The spectra exhibited marked differences that are confirmed by the pairwise correlation coefficients shown in Table 2. The high spectral similarity of the excipients #2 and #4 (correlation coefficient of 0.984) is one of the main difficulties during the development of the calibration models of these two components. The coating film does not provide a significant absorption compared to the cores spectra. Fig. 3 compares the average spectra of selected area of an uncoated core and a coated tablet. The spectral differences were minimal as a result of the subtle contribution of the coating

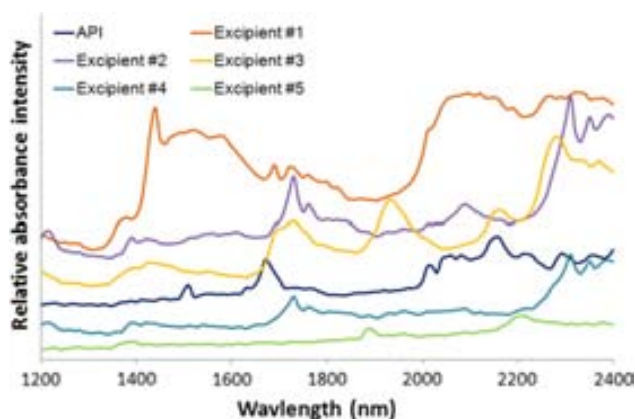


Figure 2 NIR spectra of pure API and excipients. Each spectrum has been baseline shift to avoid overlapping. The NIR spectra have been obtained averaging the spectra of each pixel of an area of the hyperspectral image.

Table 2 Correlation coefficient between the average NIR spectra of the pure tablet components (API and excipients).

	Correlation coef cients					
	API	E #1	E #2	E #3	E #4	E #5
API	1.000					
E #1	0.810	1.000				
E #2	0.621	0.601	1.000			
E #3	0.693	0.603	0.849	1.000		
E #4	0.640	0.588	0.984	0.903	1.000	
E #5	0.823	0.693	0.669	0.847	0.740	1.000

The NIR spectra have been obtained averaging the spectra of each pixel of the hyperspectral image.

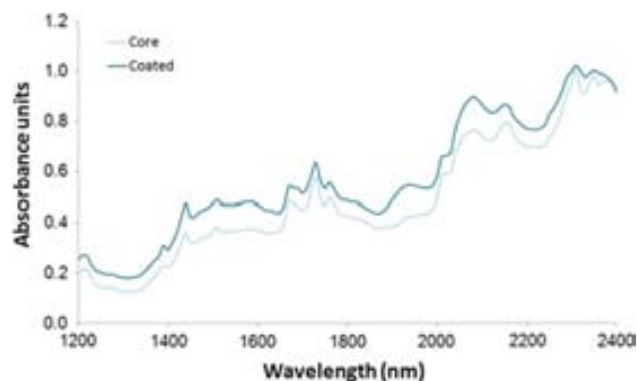


Figure 3 NIR spectra of an uncoated core and a coated tablet. The NIR spectra have been obtained averaging the spectra of each pixel of an area of the hyperspectral image.

Table 3 API and excipients concentration (% w/w) from different sides of the same core.

Core side	Core composition (% w/w)						Sum
	API	E #1	E #2	E #3	E #4	E #5	
Side A	35.51	38.93	20.15	1.97	1.48	1.47	99.51
Side B	35.45	39.48	21.12	1.98	1.48	1.51	101.00
Side C	37.11	38.45	19.69	1.99	1.53	1.53	100.30
Side D	37.45	40.38	20.14	1.95	1.53	1.51	102.96
Side E	31.61	44.16	18.62	1.95	1.48	1.51	99.38
Side F	31.47	43.38	19.31	1.97	1.55	1.53	99.21
Mean	34.78	40.79	19.84	1.97	1.51	1.51	100.40
St. dev.	2.62	2.41	0.85	0.02	0.03	0.02	5.95

Results calculated with the PLS calibration model. Sides A to F according to Fig. 1.

layer and the lacquer produces a slightly increased baseline of the average spectrum.

3.1. Quantitation of the API and excipients

We constructed a model to determine the concentration of the active principle and five excipients present in the cores by calculating a PLS calibration model for each analyte. The models were constructed by compiling a library containing the spectra for the pure components to be quantified and applying the Isys 5.0 algorithm with a variable number of factors from 3 to 8 depending on the particular analyte. The number of PLS factors needed for each calibration model increases as the concentration of the analyte decreases. The concentrations of the six components thus obtained for each core side are listed in Table 3. The model-predicted concentrations for each core side were very similar to the nominal values for the API and the major and minor excipients. The greatest standard deviation corresponds to the two major components (API and Excipient #1) that indicate an incomplete homogenous distribution. However, the standard deviation for the minor excipients is low, that is a confirmation of their homogeneous distribution. Worth special note is the fact that the combined

proportions of the components on each side were close to 100% in all cases, which suggest that the calibration models used provide accurate predictions.

However, there are slight differences in composition between sides for the three major components. Thus, the concentrations of the three major components (API and excipients #1 and #2) on the inner sides (E and F) in the cores departed from both the nominal values and the values for the other sides. These differences can be ascribed to the increased roughness of the cut surface, which may influence the reflectance of the acquired images and consequently it can also influence the calculated concentrations.

Fig. 4 shows the concentration maps for the components on one of the outer sides of a core as obtained with the Isys 5.0 models. The distributions of excipient #2 and the API were very similar, with coinciding areas of increased concentrations. On the other hand, the high-concentration areas for excipient #1 were complementary with those for the API. This distribution pattern can be ascribed to the fact that the API and excipient #2 are mixed together before the other components

are added. By contrast, the minor excipients were uniformly distributed across the images. The API and excipient distributions observed on the other sides and cores exhibited no significant differences from the previous side and exhibited identical patterns.

Fig. 4 also includes the histogram for each core component. The histograms for the minor excipients exhibited a narrow distribution of concentration values. The histograms for the API and excipient #2 were very similarly shaped, with tails on the left. By contrast, the tails in the histogram for excipient #1 were on the right, which suggests an incomplete homogenous distribution of this excipient with the blend of API and excipient #2. In any case, the distributions cannot be considered as normal shaped.

3.2. Quantitation of coated tablets

The API and excipients were also determined in tablets coated with a lacquer film of variable thickness. We used PLS models

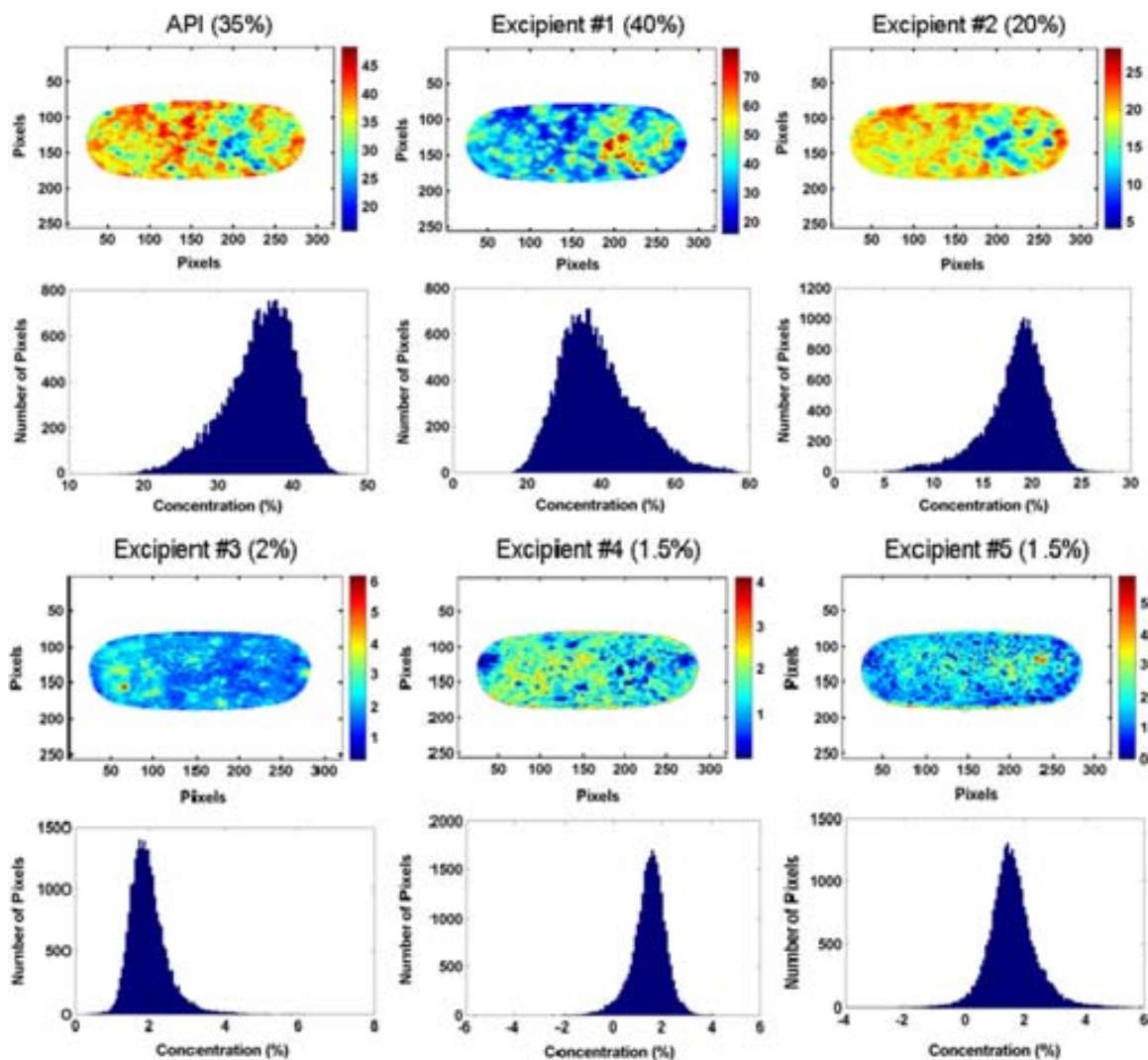


Figure 4 Concentration distribution maps and histograms for API and excipients for side A of a core calculated by PLS. In parenthesis is shown the nominal concentration (% w/w) of each component.

Table 4 Concentrations (% w/w) of API and two major excipients in tablets with different coating thickness.

Coating level	Tablet side	Tablet components (%w/w)		
		API	E #1	E #2
0x	A	35.25	40.74	20.20
	B	35.67	39.46	21.34
0.5x	A	35.85	40.30	19.67
	B	35.31	40.27	20.36
1x	A	34.56	40.80	20.79
	B	35.20	40.18	21.64
2x	A	33.15	39.15	17.95
	B	32.95	42.40	15.31
3x	A	30.85	38.51	12.65
	B	29.65	39.78	14.03

Results obtained with the PLS calibration model.

based on an also variable number of factors (3–6) depending on the particular analyte and coating thickness. Table 4 shows the results obtained for two sides (A and B) of the same tablet but coated to a different thickness with a lacquer film. The results are the averages for two tablets coated with an identical coating time and recorded in a single image. The values for the minor components have been omitted because they exhibited high deviations (50–30%) from the nominal values. The calculated concentrations for the three major components are close to their respective nominal values (Table 1) and decrease gradually with increasing coating thickness. Obviously, deviations from the nominal values are the highest in the tablets with the thickest coatings (levels 2x and 3x); this is particularly so for excipient #2, which has the low concentration.

Based on the previous results, the deviations in the calculated concentrations are mainly the result of the lacquer coating, which affects the determination of the concentrations of the major components and, especially, the minor ones. This fact can be related with a higher absorption of the incident radiation or a certain reflective scattering of the coating that reduces the signal of the core components.

3.3. Quantitation of the thickness of coating

The thickness of the coating is an important variable inasmuch as it can influence the rate of dissolution of a tablet and also the quality of components quantitation. In this work, tablet thickness was determined using a PLS model constructed from the average spectrum for each tablet side as a function of coating time. Fig. 5 shows the mean spectra for each coating level (from 0x to 3x). The band between 1800 and 2000 nm increased their intensity with the coating time. Also there are other bands along the whole range that showed differences according to the coating time.

A total of 10 images of tablets coated to a variable thickness (two tablets per image) were obtained. The spectra of the selected area of image were averaged to obtain 4 spectra per each thickness (2 tablets per image and 2 sides per tablet). Three of the 20 spectra thus obtained were discarded as

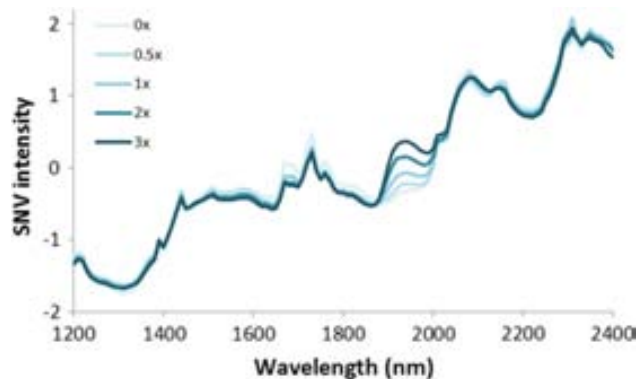


Figure 5 NIR spectra of tablets coated at different levels. The level 1x corresponds to the coating thickness of production tablets.

outliers, 12 were used to construct the calibration set and 5 to validate it. Each average spectrum in the calibration set was assigned a time of coating of 0x, 0.5x, 1x, 2x or 3x. These coating thickness were determined from the increase in weight of the tablets upon coating. Different spectral pretreatments were tested during the development of the PLS calibration models (derivation, Savitzky–Golay, SNV) and also different wavelength ranges were selected in order to ensure optimal fitting to the validation set. The best results were obtained by processing data with the SNV algorithm and using the entire spectral range. The calibration model calculated using the spectral range where the coating presented a significant absorption did not improve the predictive ability compared to the model without this range. Under these conditions, a model with a single factor accounted for 99.3% of the total variance. Such a PLS model, which is described in detail in Table 5, was adopted for application to tablets c.

The application of the calibration model to a NIR-CI allowed to quantify the mean coating thickness and its distribution over the surface. Fig. 6 shows the concentration maps for the tablet coating on the four outer sides of a tablet c (Table 1), which were processed with the previous PLS model. A comparison of the four concentration maps reveals that the coating material is not homogeneously distributed on all sides. There were some points where the predicted concentration was higher than those on the remainder of the surface. This may have resulted from the formation of lacquer aggregates during the coating process and indicate an irregular spraying inside the coating container drum. Fig. 6 additionally shows the histograms for easier analysis of the heterogeneous coating distribution.

Table 6 shows the average coating values for each of the four studied sides on two tablets c calculated with the above-described PLS model. Both tablet 1 and 2 exhibited differences of up to 10% (0.1x) in coating content from the nominal value between sides; however, the average for the four tablet sides was very similar to the nominal value in all samples.

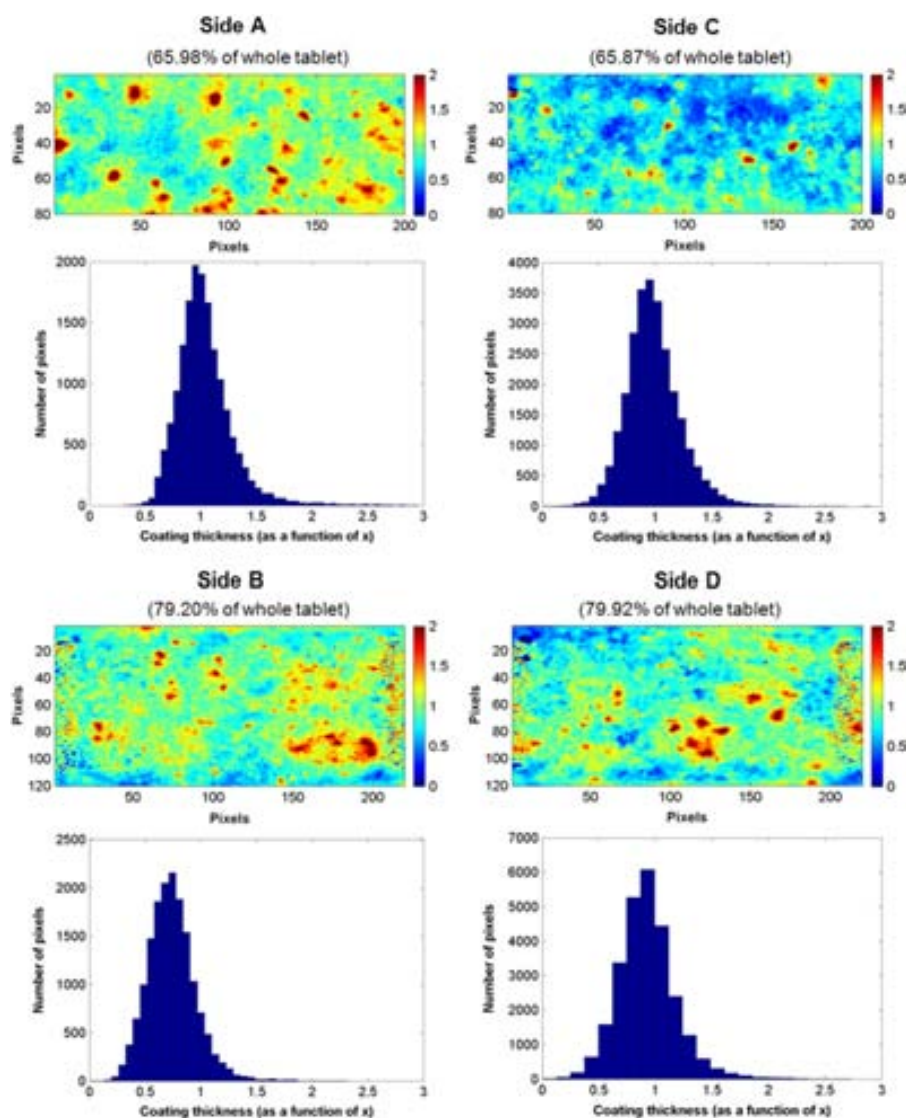
4. Conclusions

In this work, we assessed the potential for the NIR-CI technique for determining the composition of coated and uncoated drug

Table 5 Figures of merit for the quantification of coating thickness obtained with the PLS calibration model.

PLS calibration model features							
Number of calibration samples included	Spectral range (nm)	Spectral pretreatment	Number of PLS factors	Y-variance (%)	Number of validation samples	RMSEP ^a	RSEP (%) ^a
12	1200–2400	SNV	1	99.31	5	14.04	8.32

RSEP: Relative Standard Error of Prediction.

^aRMSEP: Root Mean Square Error of Prediction.**Figure 6** Tablet coating distribution maps and histograms for the four sides of one production tablet calculated by the PLS calibration model. The percentage of whole tablet image that represents the shown area is provided by the valued between parenthesis.

tablets, and the distribution of their components. We found the PLS algorithm to be effective towards extracting quality quantitative information from hyperspectral images without the need for calibration against reference values, and also towards establishing the distribution of tablet components. Conventional processing of the spectral data contained in an image allows one

to determine the thickness of the coating layer with a view to predicting some pharmaceutical properties of the tablet. The expeditiousness with which it allows tablets to be analyzed, and the accuracy of its predictions, make the NIR-CI technique a highly suitable choice for the determination of the quantitative composition of pharmaceutical tablets, and for assessing

Table 6 Coating thickness (as a function of x) in two production tablets obtained applying the described PLS calibration model.

Tablet side	Coating thickness (as a function of x)	
	Tablet 1	Tablet 2
Side A	1.035	1.122
Side B	0.984	1.177
Side C	0.739	1.035
Side D	0.925	1.223
Mean (4 sides)	0.921	1.139

x = nominal thickness (approximately 40 μm).

uniformity in the distribution of its components and surface coating. These advantages testify to the potential of NIR-CI for use in at-line tests for the control of pharmaceutical production processes.

Acknowledgments

The authors are grateful to Spain's MICINN for funding this research within the framework of Project CTQ2009-08312.

References

- [1] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, et al., Recent applications of chemical imaging to pharmaceutical process monitoring and quality control, *Eur. J. Pharm. Biopharm.* 69 (2008) 10–22.
- [2] H. Ma, C.A. Anderson, Characterization of pharmaceutical powder blends by NIR chemical imaging, *J. Pharm. Sci.* 97 (2008) 3305–3320.
- [3] W. Li, A. Woldu, R. Kelly, et al., Measurement of drug agglomerates in powder blending simulation samples by near infrared chemical imaging, *Int. J. Pharm.* 350 (2008) 369–373.
- [4] M.B. Lopes, J.C. Wolff, J.M. Bioucas-Dias, et al., Determination of the composition of counterfeit Heptodin tablets by near infrared chemical imaging and classical least squares estimation, *Anal. Chim. Acta* 641 (2009) 46–51.
- [5] J. Cruz, M. Bautista, J.M. Amigo, et al., Nir-chemical imaging study of acetylsalicylic acid in commercial tablets, *Talanta* 80 (2009) 473–478.
- [6] C. Gendrin, Y. Roggo, C. Collet, Content uniformity of pharmaceutical solid dosage forms by near infrared hyperspectral imaging: a feasibility study, *Talanta* 73 (2007) 733–741.
- [7] G. Reich, Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications, *Adv. Drug Delivery Rev.* 57 (2005) 1109–1143.
- [8] C. Cairós, J.M. Amigo, R. Watt, et al., Implementation of enhanced correlation maps in near infrared chemical images: application in pharmaceutical research, *Talanta* 79 (2008) 657–664.
- [9] L. Maurer, H. Leuenberger, Terahertz pulsed imaging and near infrared imaging to monitor the coating process of pharmaceutical tablets, *Int. J. Pharm.* 370 (2009) 8–16.
- [10] J. Hogan, Diseño y fabricación de las formas farmacéuticas, in: M.E. Aulton (Ed.), *Farmacia la ciencia del diseño de las formas farmacéuticas*, Elsevier, Madrid, 2004, pp. 441–448.
- [11] L. Zhang, M.J. Henson, S. Sonja, Multivariate data analysis for Raman imaging of a model pharmaceutical tablet, *Anal. Chim. Acta* 545 (2005) 262–278.
- [12] J.M. Amigo, J. Cruz, M. Bautista, et al., Study of pharmaceutical samples by NIR chemical image and multivariate analysis, *Trends Anal. Chem.* 27 (2008) 696–713.
- [13] C. Ravn, E. Skibsted, R. Bro, Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—comparing common calibration approaches, *J. Pharm. Biomed. Anal.* 48 (2008) 554–561.
- [14] L.J. Makein, L.H. Kidder, E.N. Lewis, et al., Non-destructive evaluation of manufacturing process changes using near infrared chemical imaging, *NIR News* 19 (2008) 11–15.
- [15] T. Puchert, D. Lochmann, J.C. Menezes, et al., Near-infrared chemical imaging (NIR-CI) for counterfeit drug identification—a four-stage concept with a novel approach of data processing, *J. Pharm. Biomed. Anal.* 51 (2010) 138–145.

ANEXO IV

NIR imaging for high-throughput screening of moisture induced changes in freeze-dried formulations

Hjalte Trnka, Anna Palou, Pierre Emanuel Panouillot, Ari Kauppinen,
Maunu Toiviainen, Holger Grohgan, Manel Alcalà, Mikko Juuti,
Jarkko Ketolainen, Jukka Rantanen

Journal of Pharmaceutical Sciences (2014), 103(9), 2839-2846.

Near-Infrared Imaging for High-Throughput Screening of Moisture Induced Changes in Freeze-Dried Formulations

HJALTE TRNKA,¹ ANNA PALOU,² PIERRE EMANUEL PANOUILLOT,³ ARI KAUPPINEN,⁴ MAUNU TOIVIAINEN,³ HOLGER GROHGANZ,¹ MANEL ALCALÀ,² MIKKO JUUTI,³ JARKKO KETOLAINEN,⁴ JUKKA RANTANEN¹

¹Department of Pharmacy, University of Copenhagen, Copenhagen 2100, Denmark

²Applied Chemometrics Group, Analytical Chemistry Unit, Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona 01893, Bellaterra, Barcelona, Spain

³Optical Measurement Technologies, VTT Technical Research Centre of Finland, Kuopio, FI 70211, Finland

⁴School of Pharmacy, Promis Centre, Faculty of Health Sciences, University of Eastern Finland, Kuopio, FI 70211, Finland

Received 9 December 2013; revised 3 February 2014; accepted 24 February 2014

Published online 24 March 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jps.23948

ABSTRACT: Evaluation of freeze-dried biopharmaceutical formulations requires careful analysis of multiple quality attributes. The aim of this study was to evaluate the use of near-infrared (NIR) imaging for fast analysis of water content and related physical properties in freeze-dried formulations. Model formulations were freeze-dried in well plates. Samples were imaged with a NIR hyperspectral camera after freeze-drying and upon storage. On the basis of Karl Fischer titration reference values, a univariate quantification model was constructed and used to visualize the distribution of water within freeze-dried samples. Differences observed between samples stored at 11% and 43% relative humidity (RH) were found to be related to the amount of amorphous component in the sample. When stored at 43% RH, the moisture content in samples with high sucrose content increased within 2 days and some degree of localized drying was observed within the samples after 3 days of storage. Further investigations with X-ray powder diffraction confirmed this local drying to be related to crystallization of sucrose. The combination of fast analysis of water content and spatial solid-state information makes NIR imaging a powerful tool for formulation development of freeze-dried samples. © 2014 Wiley Periodicals, Inc. and the American Pharmacists Association *J Pharm Sci* 103:2839–2846, 2014

Keywords: High throughput technologies; Freeze-drying; Protein formulation; NIR imaging; Water sorption; Solid state; X-ray diffraction; Formulation screening; Sucrose crystallization; Preformulation

INTRODUCTION

The increasing number of biopharmaceutical products has made freeze-drying a central unit operation in the pharmaceutical industry.¹ When formulating new protein-based products and safeguarding the quality of them in a manufacturing environment, several critical quality attributes the need to be analyzed.^{2,3} One of the most important attributes is the residual moisture in the freeze-dried cake. It has long been accepted that the optimal moisture content for protein formulations, unlike small molecules, is not always the lowest possible.⁴ Several authors agree that there seems to be an optimal range of residual moisture where protein stability is the highest.^{4–7} Further complicating the matter, the unique nature of each protein can make prediction of optimal moisture content challenging. It is however generally accepted that the hydration shell of the protein should be preserved for optimal stability.^{8,9}

The moisture level in the freeze-dried product not only affects the protein itself, but also the excipients added to stabilize the product can be affected. Increasing moisture increases the mobility and lowers the glass transition temperature of amorphous excipients, and this can in turn lead to decreased stability of the system.^{6,10–12} Moisture uptake in freeze-dried samples has even been shown to promote crystallization of amorphous

stabilizers such as sucrose and trehalose.¹² This is thought to have a pronounced effect on proteins' stability, as this crystallization results in phase separation of the glass designed to stabilize the protein and releases water that can further induce protein degradation.^{9,12}

The residual moisture content in freeze-dried products can, to a large extent, be controlled by process optimization. The formulation can however affect not only the residual moisture but especially the hygroscopicity of the sample.¹³ Although vials are typically stoppered under vacuum or after nitrogen flushing, the residual moisture content has been reported to change because of both moisture released from the stopper, moisture transport through the stopper and leakage.^{7,14–16} Understanding the effect of residual moisture and the moisture-sorbing behavior is thus crucial when developing new freeze-dried biopharmaceutical formulations.

The standard method for determining moisture is Karl Fischer (KF) titration. Although this method is accurate and typically well suited for freeze-dried samples, it is relatively slow and more importantly a destructive method. Following the same sample over time or establishing a direct relationship of moisture and protein stability of a sample is thus not possible, at least for small sample volumes. Another widely used method for moisture determination is near-infrared (NIR) spectroscopy.^{17–21} NIR has the advantage of being fast and non-destructive and has thus also been applied as a process analytical technology (PAT) tool for controlling the freeze-drying process.²² To obtain a good quantification model, some data

Correspondence to: Jukka Rantanen (Telephone: +45336585; E-mail: jukka.rantanen@sund.ku.dk)

Journal of Pharmaceutical Sciences, Vol. 103, 2839–2846 (2014)

© 2014 Wiley Periodicals, Inc. and the American Pharmacists Association

Table 1. Formulation Compositions of Plates 1 and 3

Row/Column	A	B	C	D	E	F	G	H
1–3 (5% Solid)	0% Suc	5% Suc	10% Suc	15% Suc	20% Suc	30% Suc	40% Suc	50% Suc
4–6 (5% Solid, 95:5 mannitol–sucrose)	No protein	1 mg/mL lysozyme	2.5 mg/mL lysozyme	5 mg/mL lysozyme	10 mg/mL lysozyme	20 mg/mL lysozyme	30 mg/mL lysozyme	40 mg/mL lysozyme
7–9 (5% Solid, 95:5, mannitol–sucrose)	No protein	1 mg/mL BSA	2.5 mg/mL BSA	5 mg/mL BSA	10 mg/mL BSA	20 mg/mL BSA	30 mg/mL BSA	40 mg/mL BSA
10–12 (100% Mannitol)	2% Solid	4% Solid	6% Solid	8% Solid	10% Solid	12% Solid	14% Solid	16% Solid

The basic formulation of each row is given on the left, whereas the variation within the row is specified for each column. All samples were prepared in triplicate, A1–3 are thus identical.
Suc, sucrose.

Table 2. Formulation Compositions of Plates 2 and 4

Row/Column	A	B	C	D	E	F	G	H
1–3 (5% Sugar solid, 30% sucrose)	No protein	1 mg/mL BSA	2.5 mg/mL BSA	5 mg/mL BSA	10 mg/mL BSA	20 mg/mL BSA	30 mg/mL BSA	40 mg/mL BSA
4–6 (30% Sucrose)	2% Solid	4% Solid	6% Solid	8% Solid	10% Solid	12% Solid	14% Solid	16% Solid
7–9 (10% Solid)	10% Suc	20% Suc	30% Suc	40% Suc	50% Suc	60% Suc	70% Suc	80% Suc
10–12 (5% Solid)	10% Suc	20% Suc	30% Suc	40% Suc	50% Suc	60% Suc	70% Suc	80% Suc

The basic formulation of each row is given on the left, whereas the variation within the row is specified for each column. All samples were prepared in triplicate, A1–3 are thus identical.
Suc, sucrose.

pretreatment are needed, as well as reliable reference values are required. These values are typically obtained from KF titration. Different types of quantification models have been used in the literature, both simple models based on the baseline corrected height of the water band at 1940 nm and multivariate models such as partial least squares models.²³ In the freeze-drying field, NIR has been widely studied for quantification of moisture.²⁴ Apart from detecting water, NIR has also been shown to be a fast tool for detecting solid-state characteristics in freeze-dried samples.^{25,26} Furthermore, NIR has even been shown to be useful for detecting structural changes in proteins.^{27,28}

In the recent years, NIR imaging has received increasing focus because of the spatial information generated. In the food and paper industry, NIR imaging is well established for determining distribution of both water and other analytes.^{29–33} In the pharmaceutical field, NIR imaging has also been accepted as a flexible analytical tool, among others to determine blend uniformity, to identify counterfeit products, and to evaluate the distribution of API in tablets.^{34–38} In the freeze-drying field, NIR imaging has however only been sparsely used. Zidan et al.³⁹ investigated the use of NIR imaging for the evaluation of drug loading in freeze-dried liposomes, whereas Jovanović et al.³⁶ used NIR imaging to study uniformity of freeze-dried protein–sugar samples.

The aim of this study was to evaluate the use of NIR imaging as a formulation development tool for the screening of moisture sorption and desorption, moisture distribution, and detecting related solid-state changes in freeze-dried formulations. This was performed by freeze-drying small volumes of different formulations in well-plates, and storing them at 11% relative humidity (RH) and 43% RH, respectively, at room temperature. Images of the plates were taken by an NIR camera daily over 10 days to detect moisture-induced changes as an effect of formulation composition. A robust univariate model based on KF reference values was applied to attain a fast visualization of

the water distribution in the NIR images of the freeze-dried samples.

MATERIALS AND METHODS

Sample Composition

Samples consisted of mannitol, sucrose, lysozyme, and bovine serum albumin (BSA), all from Sigma–Aldrich (St. Louis, Missouri). Freeze-drying was carried out in custom-made brass 96-well plates with wells 3 mm deep and 7 mm in diameter. Spray mount glue from 3M (St. Paul, Minnesota) was used to seal the plates on one side with transparent multilayered composite film (R&G Faserverbundwerkstoffe, Waldenbuch, Germany). One-hundred microliters of sample solution was filled into each well. A total of four well plates were prepared, two with formulations according to Table 1 (plates 1 and 3) and two according to Table 2 (plates 2 and 4). The formulations were chosen to get an insight into solid content, amorphous sugar content, protein concentration, and the effect of different proteins on moisture content and distribution.

Freeze-Drying

The freeze-drying was carried out in a Lyostar II freeze dryer (SP Scientific, Stone Ridge, New York). The freeze-drying cycle consisted of freezing at 1°C/min to –50°C, holding for 15 min, heating to –10°C at 1°C/min, annealing at –10°C for 2 h, freezing to –50°C, primary drying at –20°C for 8 h at 0.1 mbar, and finally secondary drying at 20°C for 3 h at the same pressure.

Near-infrared (NIR) Imaging

Spectral images of the freeze-dried cakes were collected using a line-scanning (pushbroom) hyperspectral camera (SWIR; Specim Ltd., Oulu, Finland). The camera detects a row of 320 spatial pixels at a time, and it disperses the incoming light from each pixel in the spectral range 970–2500 nm onto one column

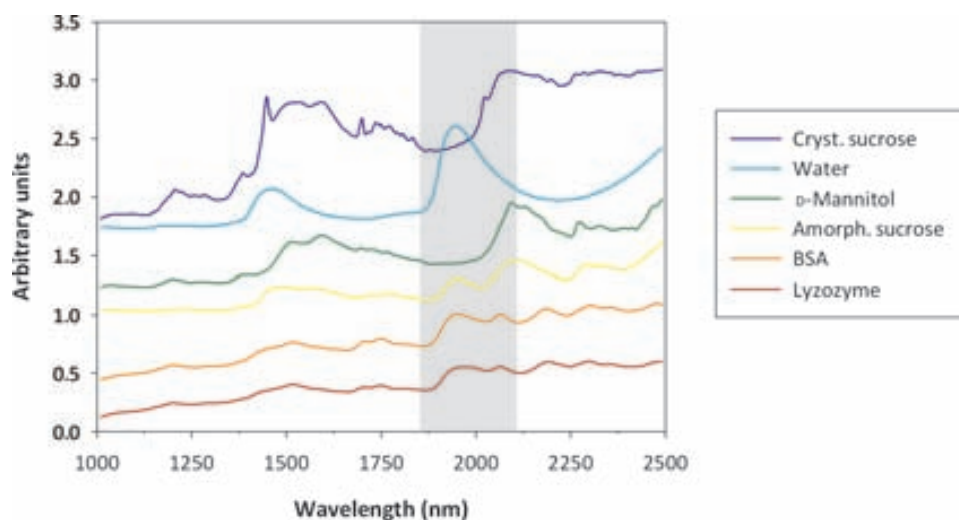


Figure 1. Near-infrared spectra [$\log(1/R)$] of crystalline sucrose, water, mannitol, amorphous sucrose, BSA, and lysozyme. The spectra have been shifted in order to avoid overlapping. The water band is highlighted in gray.

on the 320×256 -pixel mercury–cadmium–telluride (MCT) detector (14-bit readout, cooled to -70°C with a four-stage Peltier system). The well plate containing 96 samples was placed onto a conveyor belt (Mini; ENP, Hjärteby, Sweden), which was moving at the steady speed of 3.12 cm/s. The camera was placed above the belt and the camera-to-sample distance was set such that the imaged line had the length of 10 cm on the sample surface, and the speed of the conveyor belt was adjusted such that the geometry was preserved—that is, spatial pixels were squares of the size $312 \times 312 \mu\text{m}^2$ on the sample surface—at the maximum frame rate permitted by the camera (100 frames/s). The samples on the conveyor belt were illuminated with two rows of three 75-W halogen lamps (Specim Ltd.), and the measurement was conducted in diffuse reflectance mode in the $45^\circ\text{--}0^\circ\text{--}45^\circ$ geometry. Each collected spectral image consisted of 500 frames (a 5-s measurement, 500×320 spatial pixels, 256-element spectrum at each spatial pixel), which corresponds to an area of $10 \times 16 \text{ cm}^2$ on the conveyor belt, which ensured that the entire well plate fitted in the image.

Karl Fischer (KF) Titration

Water determination for calibration samples was performed on a Mettler Toledo C30 coulometric KF titrator (Mettler-Toledo GmbH, Greifensee, Switzerland). Samples were weighed just before titration with minimal exposure to the surrounding air.

X-ray Powder Diffractometry

X-ray powder diffractometry (XRPD) measurements were performed using an X-ray powder diffractometer from PANalytical X'Pert PRO with a PIXcel detector (PANalytical B.V., Almelo, The Netherlands). Diffractograms were obtained using a Cu $K\alpha$ radiation source (45 kV, 40 mA, $\lambda = 1.54187 \text{ \AA}$). Samples were measured in high throughput mode using a well-plate reader. The diffractograms were collected in transmission mode from 9° to $26^\circ 2\theta$ with a step size of $0.026^\circ 2\theta$ and a scan speed of $0.027^\circ 2\theta/\text{s}$.

Data Treatment

Before pretreatment or model/equation application, the NIR spectral images were transformed. A 96-well plate binary mask was designed for each image in order to analyze only the sample wells. The masked images were further transformed from a 3D data cube to a 2D to facilitate further work. After model/equation application, the data matrices were reshaped in the original dimensions in order to visualize the results. All these transformations, as well as the spectral pretreatments, were performed in Matlab® 7.11.0 (MathWorks, Natick, Massachusetts).

Water Content Determination

A univariate model was created for the qualitative analysis of the water distribution in freeze-dried samples. The model was based on the water band located at around 1940 nm. This value was baseline corrected with the 1875 nm signal for each sample. The choice of the minimum was performed according to contribution from other sample components as shown in Figure 1.

The model was constructed using a NIR spectrum averaged from all pixels within the calibration well imaged shortly before KF titration, giving the reference water value. All the spectra were previously pretreated by standard normal variate (SNV) correction in order to avoid light scattering contribution. A total of 62 freeze-dried samples were used to calibrate the model with KF reference values between 0.72% and 5.60% moisture. The root-mean-squared error of prediction of the model was 0.8%.

Crystallized Sucrose Analysis

In order to visualize crystallization of sucrose, the correlation coefficient (CC) between pure crystalline sucrose and sample spectra was calculated for each pixel in the wells. CC was subsequently used as a measure of the similarity between sample spectra and crystalline sucrose. CC was calculated using Eq. (1), where x_i and y_i are the absorbance values at each wavelength i for the target spectrum, $x = (x_1, \dots, x_i, \dots, x_n)$, and the pure crystalline sucrose spectrum $y = (y_1, \dots, y_i, \dots, y_n)$,

respectively. (1)

$$CC = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

As illustrated in Figure 1, the spectrum of pure crystalline sucrose is highly correlated with most of the other analytes in the samples. Mannitol correlation reaches 0.967, whereas the rest, except water, are all higher than 0.900. To avoid the contribution of the other sample analytes in the crystalline sucrose CC calculation, different spectral ranges and pretreatments (SNV and Savitzky–Golay derivatives) were tested. The best results were achieved by using the 2140–2500 nm spectral range of the spectra transformed by the first derivative (three-point spectral window and second polynomial order).

RESULTS AND DISCUSSION

Visualization of Water Content After Freeze-Drying

In order to investigate the potential of NIR imaging for fast determination of residual moisture and distribution of water (moisture maps) after freeze-drying, the well plates were imaged immediately after freeze-drying. The moisture maps calculated from the resulting hyperspectral datacubes are shown in Figure 2. The moisture content in the samples varied between 2% and 6%. These relatively high moisture contents were caused by the deliberate use of a relatively short secondary drying step at only 20°C in order to highlight the possible variation in moisture contents caused by the choice of formulation.

Looking further into the data, it is apparent that the solid content, regardless of amorphous sucrose content, does not affect the residual moisture after freeze-drying because rows 10–12 in plate 1 and rows 4–6 in plate 2 show no or only minimal change in moisture going from 2% solid (left) to 16% solid (right). Increasing the amount of amorphous component in the dried matrix on the contrary clearly shows increased moisture content after freeze-drying as anticipated. The effect of amor-

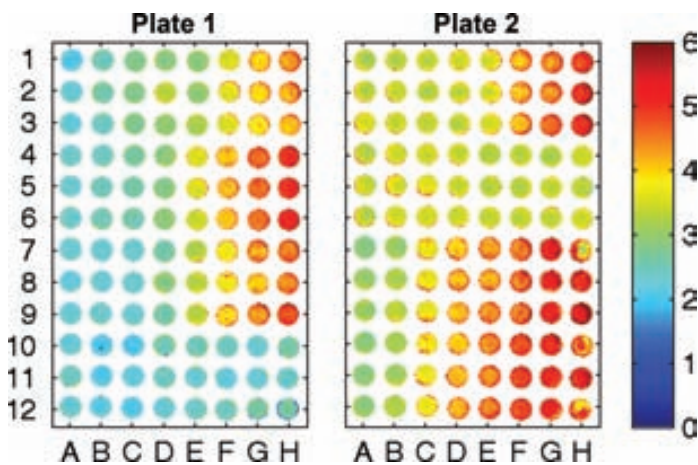


Figure 2. Distribution of water in well plates: plate 1 (left) and plate 2 (right) imaged immediately after freeze-drying. Color scale represents moisture content in percentage.

phous component increasing moisture content is seen for both sucrose (rows 7–12 in plate 2) and for protein (rows 4–9 in plate 1) in Figure 2. It is also seen that there is no major difference in moisture retention between samples containing BSA and lysozyme. BSA samples seem to have slightly lower moisture content compared with lysozyme samples with the same concentration.

The distribution of water in each single well seems generally uniform in Figure 2. Slightly higher values are noted on the edges of the well, which is however most likely because of an experimental artifact, caused by a shadow effect. A clearly uneven distribution inside a well is seen in samples in column H, rows 7–12 in plate 2. These samples have a high sucrose fraction, which seemingly causes uneven moisture distribution. The uneven moisture distribution is attributed to some degree of collapse observed in these samples.

Water Uptake at Different Storage Conditions

To study the moisture-sorbing behavior of the different formulations, the well plates were stored at two different relative humidities and imaged daily to monitor changes. Figure 3 shows the effect of storage at 11% RH (plate 1) and 43% RH (plate 3). It was observed that the water content of formulations containing only mannitol (rows 10–12) decreased during both high and low humidity storage, an effect that already occurred on day 2. One possible explanation for this decrease in the moisture content even at 43% RH is that a small part of the mannitol remains amorphous after freeze-drying and crystallization of this amorphous mannitol can release the water bound in the amorphous matrix. To investigate whether this was true, XRPD diffractograms were evaluated, to see whether increasing reflection intensity upon storage could be identified. However, no significant change in intensity was found, indicating that the pure mannitol formulations simply have very low moisture absorption capacity.

Another observation from Figure 3 is that adding only 5%–10% sucrose (rows 1–3, b and c) results in the retention of moisture at a surprisingly small difference between storage at high and low humidity. Adding more sucrose to the formulation makes the cake more hygroscopic, as seen from the higher water contents obtained when storing formulations with more than 30% sucrose at 43% RH (columns F–H, rows 1–3). Storage at lower humidity decreased the water content of these formulations to approximately 3%. Interestingly, the samples with 30%–50% sucrose that absorbed moisture during the first 2 days of storage appeared to contain less water after storage for 10 days at high humidity. The distribution of water was clearly uneven, with some areas containing only about 2% water, whereas other areas of the same sample contained as much as 5.5% water. This effect was not noted for the same samples stored at 11% RH, where moisture distribution in the samples remained uniform. This suggests that the observed uneven distribution is dependent on the storage humidity. The effect of spot-wise drying is however not observed in samples with high protein concentrations, although these samples have similar moisture content when stored at high humidity.

In order to investigate the water sorbing at high humidity further, the storage behavior of plate 4 shown in Figure 4 was evaluated. It can be observed that the water content of samples containing 80% sucrose (column H, rows 7–12) decreased already after 1 day of storage and these samples also showed

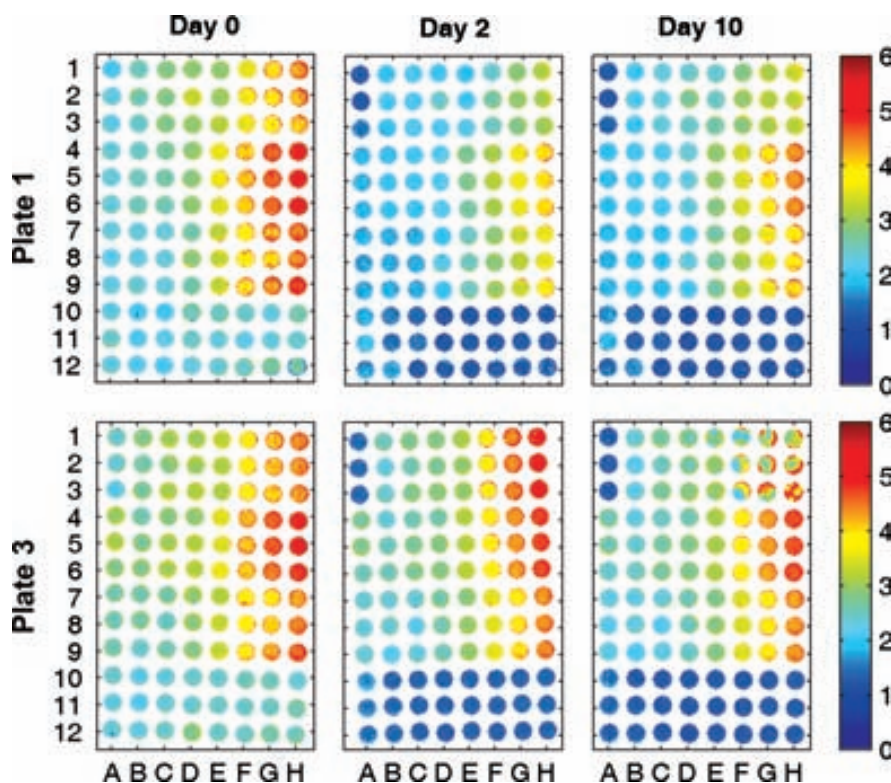


Figure 3. Distribution of water in well plates: plates 1 (top) and 3 (bottom) at days 0, 2, and 10 after freeze-drying. The composition is the same for both plates, but the storage humidity condition was 11% RH in plate 1 and 43% RH in plate 3.

pronounced collapse. In the remaining samples, distribution of water was relatively uniform until day 3 where small dry spots appeared. These dry spots then grew until they covered most of the well in some formulations by day 10. An interesting observation is that the loss of moisture only happens in samples containing more than 20% sucrose in the dry matrix. This observation can be related to the localized loss of moisture caused by the crystallization of amorphous sucrose. The release of water previously bound in the amorphous matrix would explain both the drying and the nonuniform distribution of water, as this can be caused by crystallization.⁴⁰

Sucrose Crystallization

In order to determine whether NIR imaging was capable of elucidating the speculated sucrose crystallization, the NIR spectra of amorphous and crystalline sucrose in Figure 1 were evaluated. The NIR spectrum of crystalline sucrose shows several features that are not observed in the spectrum of the amorphous sample. These features were exploited to visualize the crystallization process by calculating the CC for each pixel in the NIR images of the wells indicating similarity with the NIR spectra of crystalline sucrose. The resulting images (Fig. 5) showed the formation of small crystalline sucrose domains by day 4 (columns C–H, rows 7–12 among other). During the following days, the crystallization proceeds across the well finally covering most of the well by day 10 in samples with high sucrose content. As previously observed in the moisture plots, the minimum sucrose concentration where crystallization appeared was 30% according to the CC maps. Furthermore, the solid content of the formulation does not have any substantial impact, as sucrose was found to crystallize in concentrations

from 2% to 16%. Protein content on the contrary inhibited the crystallization of sucrose, as none of the samples containing protein showed any sign of either spot-wise drying in Figure 4 or sucrose crystallization in Figure 5. The finding that protein inhibits sucrose crystallization is in good agreement with previous findings in systems containing only sucrose and protein.⁹ It should however be noted that the lowest protein concentration in this study is 2.5 mg/mL. If developing formulations for potent drugs, the low protein concentration might not be sufficient to inhibit the possible sucrose crystallization. Whether it is the moisture content in the matrix, the sucrose concentration, or a combination that determines the 30% sucrose onset point cannot be deduced from these data. If the moisture content is the critical parameter, Figure 4 shows that the critical moisture level is around 4%.

In order to verify that the observed drying was related to the crystallization of sucrose, XRPD diffractograms were evaluated. Figure 6 shows the diffractograms of sample G7. The characteristic reflections of sucrose clearly appeared at day 4 and further increased until day 8. The confirmation of the solid-state results demonstrates that NIR imaging in combination with the right data treatment is indeed capable of detecting important solid-state transformations in freeze-dried samples.

PERSPECTIVES

The combination of fast determination of moisture distribution and solid-state analysis makes NIR imaging a very versatile tool in the field of freeze-drying. Apart from the formulation screening possibilities evaluated in this study, the high measuring speed makes it feasible to use NIR imaging as a product

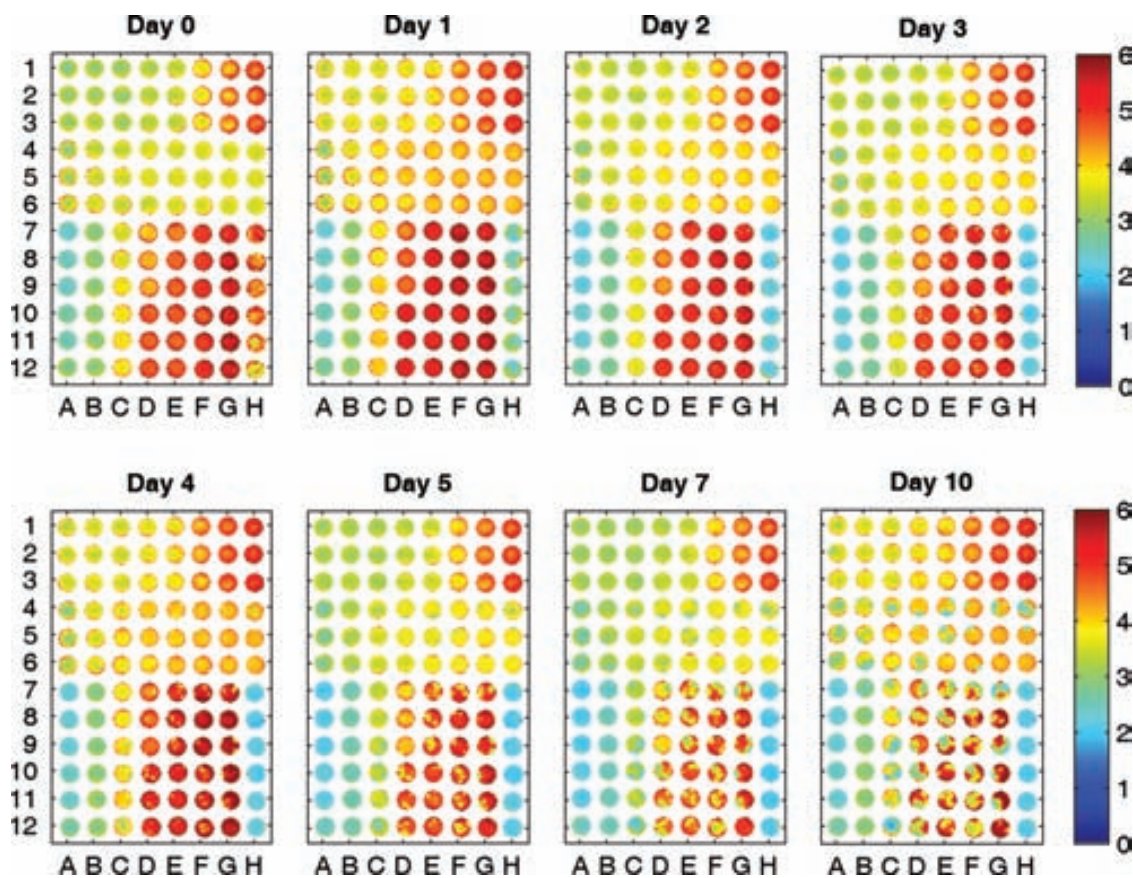


Figure 4. Distribution of water in well plate: plate 4 evolution during 10 days of storage at 43% RH.

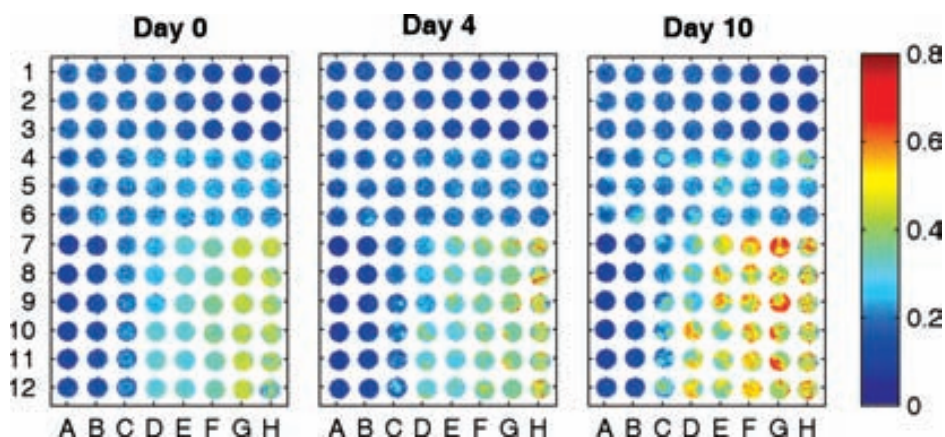


Figure 5. The spectral correlation maps for the crystalline sucrose. The maps corresponds to plate 4 images from days 0, 4, and 10 after freeze-drying. Higher values (red color) indicate higher correlation of the NIR spectra to that of crystalline sucrose.

control tool. Imaging of freeze-dried vials in production settings passing a NIR camera in combination with fast spectral image analysis could make real-time removal of vials with unwanted characteristics possible. Such solutions are needed when implementing the PAT tools for control purposes and when considering real-time release testing.

CONCLUSIONS

The combination of well plate freeze-drying and fast NIR imaging was feasible for the fast analysis of a high num-

ber of formulations. Taking only 3 s to measure a well plate with 96 formulations, NIR imaging was found to be a powerful tool for fast determination of water having applications both as a formulation screening tool and potentially for product control. Furthermore, NIR imaging provided spatial information about both moisture distribution and solid-state transformations in freeze-dried samples. Spatial solid-state information enabled the detection of sucrose crystallization, a phenomenon found to be pronounced in samples containing 30% sucrose or more when stored at 43% RH.

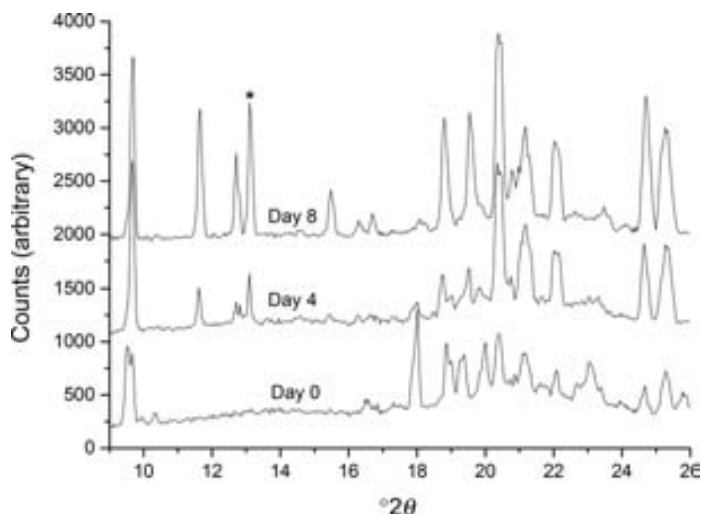


Figure 6. X-ray powder diffractometry analysis of the sample G7 at days 0, 4, and 8. Hemihydrate converts to delta mannitol (reflection at $17.9^{\circ}2\theta$ disappears), whereas sucrose crystallizes as seen from the reflections at $11.7^{\circ}2\theta$, $12.7^{\circ}2\theta$, and $13.1^{\circ}2\theta$. *The sucrose-specific reflection at $13.1^{\circ}2\theta$.

ACKNOWLEDGMENTS

The authors would like to acknowledge the grant from Lundbeckfonden for the purchase of the X-ray powder diffractometer (grant decision 479/06). Funding from the Danish Council for Independent Research, Medical Sciences, is acknowledged (HT) and Spain's Ministry of Economy and Competitiveness (MINECO) for the BES-2010-030395 fellowship (in the CTQ 2009-08312 project) (AP) is also acknowledged.

REFERENCES

- Kasper JC, Winter G, Friess W. 2013. Recent advances and further challenges in lyophilization. *Eur J Pharm Biopharm* 85(2):162–169.
- Awotwe-Otoo D, Agarabi C, Wu GK, Casey E, Read E, Lute S, Brorson KA, Khan MA, Shah RB. 2012. Quality by design: Impact of formulation variables and their interactions on quality attributes of a lyophilized monoclonal antibody. *Int J Pharm* 438(1–2):167–175.
- Kaappinen A, Toiviainen M, Aaltonen J, Korhonen O, Järvinen K, Juuti M, Pellinen R, Ketolainen J. 2013. Microscale freeze-drying with Raman spectroscopy as a tool for process development. *Anal Chem* 85(4):2109–2116.
- Greiff D. 1971. Protein structure and freeze-drying: The effects of residual moisture and gases. *Cryobiology* 8(2):145–152.
- Breen ED, Curley JG, Overcashier DE, Hsu CC, Shire SJ. 2001. Effect of moisture on the stability of a lyophilized humanized monoclonal antibody formulation. *Pharm Res* 18(9):1345–1353.
- Roy ML, Pikal MJ, Rickard EC, Maloney AM. 1992. The effects of formulation and moisture on the stability of a freeze-dried monoclonal antibody-vinca conjugate: A test of the WLF glass transition theory. *Dev Biol Stand* 74:323–339.
- Wang W. 2000. Lyophilization and development of solid protein pharmaceuticals. *Int J Pharm* 203(1–2):1–60.
- Hsu C, Ward CA, Pearlman R, Nguyen H, Yeung D, Curley JG. 1992. Determining the optimum residual moisture in lyophilized protein pharmaceuticals. *Dev Biol Stand* 74:255.
- Costantino HR, Curley JG, Wu S, Hsu CC. 1998. Water sorption behavior of lyophilized protein–sugar systems and implications for solid-state interactions. *Int J Pharm* 166(2):211–221.
- Franks F. 1992. Freeze-drying: From empiricism to predictability. The significance of glass transitions. *Dev Biol Stand* 74:9–18, discussion 19.
- Ikeda K, Yoshioka S, Kojima S. 1994. Physical stability and protein stability of freeze-dried cakes during storage at elevated temperatures. *Pharm Res* 11(7):995–999.
- Roos Y, Karel M. 1991. Phase transitions of mixtures of amorphous polysaccharides and sugars. *Biotechnol Prog* 7(1):49–53.
- Fakes MG, Dali MV, Haby TA, Morris KR, Varia SA, Serajuddin ATM. 2000. Moisture sorption behavior of selected bulking agents used in lyophilized products. *PDA J Pharm Sci Technol* 54(2):144–149.
- Stokvold A, Dyrstad K, Libnau FO. 2002. Sensitive NIRS measurement of increased moisture in stored hygroscopic freeze dried product. *J Pharm Biomed Anal* 28(5):867–873.
- Lueckel B, Helk B, Bodmer D, Leuenberger H. 1998. Effects of formulation and process variables on the aggregation of freeze-dried interleukin-6 (IL-6) after lyophilization and on storage. *Pharm Dev Technol* 3(3):337–346.
- Ford AW, Dawson PJ. 1994. Effect of type of container, storage temperature and humidity on the biological activity of freeze-dried alkaline phosphatase. *Biologicals* 22(2):191–197.
- Carpenter JF, Prestrelski SJ, Dong A. 1998. Application of infrared spectroscopy to development of stable lyophilized protein formulations. *Eur J Pharm Biopharm* 45(3):231–238.
- Grohganz H, Fonteyne M, Skibsted E, Falck T, Palmqvist B, Rantanen J. 2009. Role of excipients in the quantification of water in lyophilized mixtures using NIR spectroscopy. *J Pharm Biomed Anal* 49(4):901–907.
- Jones JA, Last IR, MacDonald BF, Prebble KA. 1993. Development and transferability of near-infrared methods for determination of moisture in a freeze-dried injection product. *J Pharm Biomed Anal* 11(11–12):1227–1231.
- Lin TP, Hsu CC. 2002. Determination of residual moisture in lyophilized protein pharmaceuticals using a rapid and non-invasive method: Near infrared spectroscopy. *PDA J Pharm Sci Technol* 56(4):196–205.
- Kaappinen A, Toiviainen M, Korhonen O, Aaltonen J, Järvinen K, Paaso J, Juuti M, Ketolainen J. 2013. In-line multipoint near-infrared spectroscopy for moisture content quantification during freeze-drying. *Anal Chem* 85(4):2377–2384.
- De Beer TRM, Verduyn P, Burggraef A, Quinten T, Ouyang J, Zhang X, Vervaet C, Remon JP, Baeyens WRG. 2009. In-line and real-time process monitoring of a freeze drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools. *J Pharm Sci* 98(9):3430–3446.
- Reich G. 2005. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv Drug Deliv Rev* 57(8):1109–1143.
- Maltesen M, Bjerregaard S, Hovgaard L, Havelund S, Weert M, Grohganz H. 2011. Multivariate analysis of phenol in freeze-dried and spray-dried insulin formulations by NIR and FTIR. *AAPS PharmSciTech* 12(2):627–636.
- Grohganz H, Fonteyne M, Skibsted E, Falck T, Palmqvist B, Rantanen J. 2010. Classification of lyophilized mixtures using multivariate analysis of NIR spectra. *Eur J Pharm Biopharm* 74(2):406–412.
- Grohganz H, Gildemyn D, Skibsted E, Flink JM, Rantanen J. 2011. Rapid solid-state analysis of freeze-dried protein formulations using NIR and Raman spectroscopies. *J Pharm Sci* 100(7):2871–2875.
- Bai S, Nayar R, Carpenter JF, Manning MC. 2005. Noninvasive determination of protein conformation in the solid state using near infrared (NIR) spectroscopy. *J Pharm Sci* 94(9):2030–2038.
- Ikeda K, Fujimaki Y, Kuwabara A, Hiyama Y, Yomota C, Aoyagi N. 2006. Near-infrared analysis of protein secondary structure in aqueous solutions and freeze-dried solids. *J Pharm Sci* 95(4):781–789.
- ElMasry G, Sun D-W, Allen P. 2012. Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *J Food Eng* 110(1):127–140.

30. Fernández Pierna JA, Vermeulen P, Amand O, Tossens A, Dardenne P, Baeten V. 2012. NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed. *Chemometr Intell Lab* 117(0):233–239.
31. Iqbal A, Sun D-W, Allen P. 2013. Prediction of moisture, color and pH in cooked, pre-sliced turkey hams by NIR hyperspectral imaging system. *J Food Eng* 117(1):42–51.
32. Kamruzzaman M, ElMasry G, Sun D-W, Allen P. 2013. Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. *Food Chem* 141(1):389–396.
33. Whitworth MB, Millar SJ, Chau A. 2010. Food quality assessment by NIR hyperspectral imaging. Conference proceeding: Sensing for Agriculture and Food Quality and Safety II 767605.
34. Awa K, Okumura T, Shinzawa H, Otsuka M, Ozaki Y. 2008. Self-modeling curve resolution (SMCR) analysis of near-infrared (NIR) imaging data of pharmaceutical tablets. *Anal Chem Acta* 619(1):81–86.
35. Dubois J, Wolff J-C, Warrack JK, Schoppelrei J, Lewis EN. 2007. NIR chemical imaging for counterfeit pharmaceutical products analysis. *Spectroscopy* 22(2):40.
36. Jovanović N, Gerich A, Bouchard A, Jiskoot W. 2006. Near-infrared imaging for studying homogeneity of protein–sugar mixtures. *Pharm Res* 23(9):2002–2013.
37. Ma H, Anderson CA. 2008. Characterization of pharmaceutical powder blends by NIR chemical imaging. *J Pharm Sci* 97(8):3305–3320.
38. Schonbichler SA, Bittner LK, Weiss AK, Griesser UJ, Pallua JD, Huck CW. 2013. Comparison of NIR chemical imaging with conventional NIR, Raman and ATR–IR spectroscopy for quantification of furosemide crystal polymorphs in ternary powder mixtures. *Eur J Pharm Biopharm* 84(3):616–625.
39. Zidan A, Spinks C, Fortunak J, Habib M, Khan M. 2010. Near-infrared investigations of novel anti-HIV tenofovir liposomes. *AAPS J* 12(2):202–214.
40. Jouppila K, Roos YH. 1994. Water sorption and time-dependent phenomena of milk powders. *J Dairy Sci* 77(7):1798–1808.